Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique

Lionel Berthet

▶ **To cite this version:**

Lionel Berthet. Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique. Hydrologie. AgroParisTech, 2010. Français. NNT : 2010AGPT0017 . pastel-00529652

HAL Id: pastel-00529652

https://pastel.hal.science/pastel-00529652

Submitted on 26 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Doctorat ParisTech

École doctorale *Géosciences et Ressources naturelles*

T H È S E

pour obtenir le grade de docteur délivré par

l'Institut des Sciences et Industries du Vivant et de l'Environnement

AgroParisTech

Spécialité : Hydrologie

présentée et soutenue publiquement par

Lionel BERTHET

le 19 février 2010

Prévision des crues au pas de temps horaire :

pour une meilleure assimilation de l'information de débit dans un modèle hydrologique

Directeur de thèse : **Vazken ANDRÉASSIAN**

Jury		
Pr. Denis DARTUS	INP, Toulouse	Président
Pr. András BÁRDOSSY	Institut für Wasserbau, Stuttgart	Rapporteur
Dr. Thierry LEVIANDIER	ENGEES, Strasbourg	Rapporteur
Dr. André BACHOC	SCHAPI, Toulouse	Examinateur
M. Rémy GARÇON	EDF – DTG, Grenoble	Examinateur
Dr. Cyril KAO	AgroParisTech, Paris	Examinateur
Dr. Éric PARENT	AgroParisTech, Paris	Examinateur

AgroParisTech
Cemagref, unité de recherche *Hydrosystèmes et Bioprocédés*
Parc de Tourvoie – BP 44 – 92 163 Antony cedex

Ce document a été rédigé sous LaTeX 2$_\varepsilon$ (Knuth, 1984, Lamport, 1994, LaTeX 3 Project Team, 1997). Les calculs ont été réalisés sous Fortran, Scilab et R (R Development Core Team, 2006). Les graphes ont été tracés avec R, de nombreuses cartes avec les S.I.G. *OpenJump* et *QGIS*.

Ces logiciels sont libres, sous licences *LaTeX Project Public License* pour le premier, *CeCILL* pour Scilab et *GNU General Public License* pour les autres.

Résumé

La modélisation hydrologique Pluie – Débit compte parmi les outils incontournables pour prévoir les crues, car elle permet d'atteindre des horizons plus lointains que de nombreuses autres méthodes. Pour tenter d'améliorer la qualité des prévisions, les hydrologues ont proposé de nombreuses approches d'assimilation de données.
Cette thèse s'intéresse à un modèle Pluie – Débit fonctionnant au pas de temps horaire. Elle propose une comparaison de nombreuses méthodes de mise-à-jour de ce modèle par l'assimilation de la donnée de débit observé à l'exutoire du bassin versant, dans le but de fournir des prévisions de crue à des horizons allant de quelques heures à quelques jours. Les mises-à-jour étudiées portent sur les paramètres du modèle, ses états et ses sorties. Certaines sont déterministes, d'autres probabilistes.
Ce travail a permis de mettre en avant l'interaction qui existe entre la structure du modèle hydrologique et les mises-à-jour. Ces deux éléments se caractérisent par des dynamiques temporelles différentes. Nos résultats plaident pour que le modèle soit considéré comme un tout et non comme la juxtaposition d'un modèle hydrologique et d'une procédure de mise-à-jour. Il nous paraît préférable d'optimiser la structure et de caler le modèle en tenant compte des mises-à-jour qui seront appliquées.
Le modèle donnant les meilleures prévisions associe une structure hydrologique très proche de celle obtenue au pas de temps journalier et deux mises-à-jour du modèle, l'une modifiant un état crucial du modèle, l'autre corrigeant ses sorties par un réseau de neurones artificiels. Cette combinaison a été évaluée à l'aide d'une sélection de critères de performance adaptés à l'objectif du modèle, sur un vaste ensemble de bassins versants français. Ses performances sont robustes et se comparent favorablement à d'autres modèles de prévision.

Abstract

Rainfall-Runoff modelling belongs to the major tools for flood forecasting, as it allows to reach longer lead times than by many other methods. The hydrological community has developed numerous data assimilation techniques in order to improve on the forecast quality.
In this study, we compared many methods which aim at updating a Rainfall-Runoff model at an hourly time step with lead times from one hour to a few days. All the tested techniques assimilate the streamflow measured at the catchment outlet. They update the model's states, parameters or outputs. Some are deterministic while others are probabilistic.
This study highlighted the interaction between the structure of the hydrological model and the updates, both characterized by different dynamics. Hence considering the model as a whole seems more effective than considering it as the juxtaposition of a hydrological model followed by an updating procedure. We therefore advocate taking into account the updating techniques to be applied to the model when identifying its structure and calibrating it.
The model which provides the best forecasts combines a hydrological structure and two updating methods. The flowchart is almost identical to the one optimized for the daily time step. The first update modifies one key model state, while the second one is an artificial neural network which predicts the remaining model error. This combination has been tested over a large data set of French watersheds. Its predictive performances are robust and compare favourably with those of other flood forecasting models.

Table des matières

Contexte 1

1 Introduction 3
- 1.1 Première formulation des objectifs . . . 5
- 1.2 Modélisation Pluie – Débit . . . 5
- 1.3 Assimilation de données et mise-à-jour du modèle . . . 9
- 1.4 Reformulation de nos objectifs . . . 11
- 1.5 Organisation de ce mémoire . . . 12

I Méthodologie et données 15

2 Méthodologie 17
- 2.1 Construire un modèle (Pluie, Débit)→ Débit . . . 19
- 2.2 Faut-il être au plus près des conditions opérationnelles ? . . . 26
- 2.3 Multiplier les expériences à défaut de pouvoir les reproduire . . . 30
- 2.4 Au-delà de la recherche de l'amélioration des performances . . . 35
- 2.5 Présentation graphique des résultats . . . 35

3 Critères de performance 37
- 3.1 Faut-il employer la fonction de vraisemblance ? . . . 39
- 3.2 Plusieurs familles de critère . . . 40
- 3.3 Splendeurs et misères des critères quadratiques . . . 46
- 3.4 Degré de signification de la différence de valeurs d'un critère . . . 48
- 3.5 Des critères pour juger des améliorations du modèle et pour évaluer le modèle 50
- 3.6 Est-il nécessaire d'employer une multitude de critères ? . . . 50
- 3.7 Bilan . . . 52

4 Un vaste échantillon de bassins versants 53
- 4.1 Données disponibles sur chaque bassin . . . 55
- 4.2 Caractériser la variété hydrologique de l'ensemble de 1024 bassins versants . 58
- 4.3 Définition et caractérisation de groupes de bassins versants homogènes . . . 62
- 4.4 Limiter les biais : passer de l'ensemble à l'échantillon . . . 66
- 4.5 Bilan et conclusions . . . 72

II Passer du pas de temps journalier au pas de temps horaire 73

5 Influence de la résolution temporelle sur le modèle 75
- 5.1 Comparaison des hydrogrammes journalier et horaire . . . 77
- 5.2 Influence du pas de temps sur le fonctionnement et les performances du modèle 81
- 5.3 Bilan . . . 87

6 Adapter la structure du modèle pour le pas de temps horaire 89
- 6.1 Modification des fonctions non conservatives . . . 91
- 6.2 Distribution de la fonction de production . . . 95
- 6.3 Modifications de la forme de l'hydrogramme unitaire . . . 95

6.4 Ajout d'une seconde branche d'écoulement au routage 98
6.5 Bilan 102

III Mises-à-jour du modèle 103

7 Mise-à-jour des états 105
7.1 Quel(s) état(s) mettre à jour ? 107
7.2 Partir d'une mise-à-jour simple 109
7.3 Faire remonter la mise-à-jour 110
7.4 Emploi du filtre de Kálmán étendu 116
7.5 Mise-à-jour incomplète du niveau du réservoir de routage 129
7.6 Bilan 131

8 Un essai d'évaluation des incertitudes : le filtre particulaire 133
8.1 Rapide état de l'art 135
8.2 Du cadre bayésien au filtre particulaire 138
8.3 Associer le filtre particulaire et le modèle déterministe GRP 139
8.4 Expérimentations et discussion 144
8.5 Bilan 154

9 Mise-à-jour des paramètres 155
9.1 Mettre à jour un invariant temporel ? 157
9.2 Notre point de départ : la méthode de Yang et Michel 160
9.3 Proposition d'une méthode alternative 163
9.4 Bilan 172

10 Mise-à-jour des sorties 173
10.1 L'erreur du modèle hydrologique n'est pas du bruit blanc 176
10.2 Peut-on débiaiser les sorties du modèle ? 176
10.3 Régressions simples 180
10.4 Correction par modèles ARIMA 183
10.5 Correction par réseaux de neurones artificiels 188
10.6 Bilan 200

IV Influence des mises-à-jour sur le fonctionnement du modèle 201

11 Initialisation du modèle GRP 203

12 Influence des mises-à-jour sur le calage du modèle 219
12.1 Le calage est influencé par les mises-à-jour 221
12.2 Faut-il tenir compte des mises-à-jour lors du calage ? 222
12.3 Stabilité des paramètres avec l'horizon de prévision 227
12.4 Bilan 231

V Évaluation du modèle de prévision 233

13 Évaluation du modèle sur un vaste échantillon de bassins versants 235
13.1 Évaluation du modèle GRP sur 846 bassins « indépendants » 238

13.2 Où ? Quand ? Comment ? . . . 245
13.3 Évaluations complémentaires . . . 251
13.4 Pertinence de la réduction de l'échantillon de bassins versants . . . 252
13.5 Bilan . . . 255

14 Comparaison à d'autres stratégies de prévision Pluie – Débit 257
14.1 PQb . . . 259
14.2 Les réseaux de neurones . . . 262
14.3 Bilan . . . 266

15 Robustesse du modèle 269
15.1 Influence de lacunes dans les observations . . . 272
15.2 Influence du bruit dans les données . . . 275
15.3 Bilan et discussion . . . 288

Bilan et perspectives 290

Bibiographie 297

Annexes 315

A Glossaire 317

B Notations 327

C Structure du modèle GRP 337
C.1 Intégration de la loi de vidange d'un réservoir de type « Puissance » . . . 339
C.2 Structure hydrologique du modèle . . . 340
C.3 Mises-à-jour proposées par Tangara (2005) . . . 344
C.4 Valeurs des paramètres fixes . . . 345
C.5 Équations du réservoir de production . . . 345

D Compléments sur les critères de performance 349
D.1 Lien entre critère de persistance et critère de Nash-Sutcliffe . . . 351
D.2 Usons de critères vraiment différents . . . 351

E Combien d'années pour obtenir une valeur sûre d'un critère quadratique ? 361

F Les critères quadratiques sont-ils vraiment continus ? 393

G Données et graphes décrivant l'ensemble des bassins versants 421
G.1 Analyse de la variété hydrologique de l'ensemble des bassins versants décrite par six indices caractérisant le débit . . . 423
G.2 Données et graphes annexes au chapitre 4 . . . 427

H Influence des données douteuses sur l'évaluation des modèles de prévision de crues 475
H.1 Quantification des crues qualifiées de douteuses . . . 477
H.2 Comparaison des performances obtenues avec et sans ces données douteuses 479

H.3 « Conclusion » 482

I Classification ascendante hiérarchique et analyse en composantes principales 483
I.1 Classification ascendante hiérarchique 485
I.2 Analyse en composantes principales 486

J Distribution de la fonction de production de GRP 489
J.1 Principes 491
J.2 Données de précipitation employées 492
J.3 Comparaison des flux en sortie de la fonction de production 493
J.4 Influence sur les performances en prévision 498
J.5 Conclusions 501

K Compléments au chapitre 6 503
K.1 Inversion de la fonction de vidange du réservoir de routage 505
K.2 Mise-à-jour de la pluie efficace PRT_t 506
K.3 Représentation graphique du fonctionnement de GRP 509
K.4 Implémentation du filtre de Kálmán étendu appliqué à la fonction de routage de GRP 509
K.5 Interprétation de la mise-à-jour « incomplète » du niveau du réservoir de routage 517

L Compléments au chapitre 8 521
L.1 Principes et fonctionnement du filtre particulaire 523
L.2 Évaluation des estimation des incertitudes de prévision 528

M Premiers tests d'emploi des prévisions d'ensemble ARPEGE 533
M.1 Données disponibles 535
M.2 Essais réalisés 535
M.3 Résultats et discussion 536

N Durées de vie des mises-à-jour et efficacité du modèle quand peu de données sont disponibles 539

O Compléments au chapitre 10 549
O.1 Traduction de l'épigraphe du chapitre 10 551
O.2 Autocorrélation des bassins à forts régimes de base 551
O.3 Faiblesse de la correction régressive multiplicative pour des horizons lointains 552
O.4 Modèles ARIMA : valeurs admissibles des paramètres 554
O.5 Configuration des réseaux de neurones 555
O.6 Calage des réseaux de neurones : algorithme de Levenberg-Marquardt avec régulation bayésienne 561

P Cartes des performances du modèle GRP 569

Q Méthode des plus proches voisins 599
Q.1 Les mêmes causes entraînent les mêmes effets 601
Q.2 Résultats 602
Q.3 Conclusions 603

Table des figures

1.1 Transformation de la pluie en débit 6

2.1 Illustration de la description formelle d'un modèle Pluie – Débit 22
2.2 Principes des mises-à-jour variationnelles et séquentielles 23
2.3 Horizon maximal sans influence du scénario de pluies futures 32
2.4 Autocorrélation du débit et horizon de prévision H_{max} 33
2.5 Comparaison des hydrogrammes de bassins lents et rapides 34
2.6 Lecture d'une boîte à moustache 36

3.1 Exemple d'hydrogramme avec chevelus de prévision 40
3.2 Évolution des différents critères quadratiques avec l'horizon de prévision . . 42
3.3 Le critère de retard moyen au dépassement de seuil 47
3.4 Exemple de perte de robustesse et gain en performance médiane 50

4.1 Exemple d'hydrogramme étêté 57
4.2 Emplacement des 1 024 bassins versants français. 59
4.3 Plans principaux de l'A.C.P. sur les 10 indices décrivant le débit et la pluie. 64
4.4 Partition de l'échantillon de 1 024 bassins en 4 groupes homogènes par C.A.H., représentée dans les 3 principaux plans de l'A.C.P. 65
4.5 Répartition géographique des quatre groupes homogènes de bassins versants définis à partir de l'échantillon de 1 024 bassins versants par la C.A.H. calculée sur les 10 indices décrivant les régimes hydrographiques et pluvieux. 66
4.6 Comparaison de l'ensemble initial et de l'échantillon restreint de 178 bassins 71

5.1 Comparaison des hydrogrammes horaires et journaliers 78
5.2 Rapport des quantiles extrêmes de débit, aux pas de temps journalier et horaire 79
5.3 Écart relatif des hydrogrammes horaires et journaliers en France 80
5.4 Schémas des structures des modèles de prévision et simulation au pas de temps horaire 82

6.1 Hydrogrammes asymétriques 97
6.2 Introduction d'une seconde branche de routage 99

7.1 Chroniques des états internes de GRP 108
7.2 Quatre stratégies de mise-à-jour de l'hydrogramme unitaire 112
7.3 Performances du modèle mis-à-jour par le filtre de Kálmán 124
7.4 Performances du modèle mis-à-jour par le filtre de Kálmán 125
7.5 Intervalles de confiance estimés par le filtre de Kálmán 128
7.6 Mise-à-jour incomplète du niveau du réservoir de routage 130

8.1 Schéma général de fonctionnement du filtre particulaire avec GRP 143
8.2 Schéma des premiers essais d'implémentation du filtre particulaire 145
8.3 Fiabilité de la combinaison GRP – Filtre particulaire quand seules les entrées de pluie sont bruitées 147
8.4 Exemple de chroniques des estimations des incertitudes par filtre particulaire 150
8.5 Exemple de chroniques des estimations des incertitudes par filtre particulaire 151

8.6 Influence de la réactivité des bassins sur le calage de σ_Q et σ_M 152

9.1 Principe de la mise-à-jour de Yang et Michel (2000) 161
9.2 Alternative à la mise-à-jour des paramètres proposés par Yang 165
9.3 Gains de performance dûs à la mise-à-jour des paramètres du modèle GRP . 166
9.4 Trois définitions des sous-ensembles de calage 168
9.5 Dispersion des paramètres calés 172

10.1 Illustration des notations principales des débits prévus et observés 176
10.2 Autocorrélations des erreurs de prévision 177
10.3 Hétéroscédasticité des erreurs de prévision 178
10.4 Biais de prévision en fonction du débit prévu 180
10.5 Gains de performance dus à des corrections de l'erreur par régression 182
10.6 Performances de la correction par modèles ARIMA 187
10.7 Schéma d'un perceptron 192
10.8 Performances de la correction d'erreurs par réseaux de neurones 195
10.9 Exemples de correction d'erreurs par réseaux de neurones 198

12.1 Influence des mises-à-jour sur les valeurs calées des paramètres 223
12.2 Influence des mises-à-jour sur les valeurs calées des paramètres (2) 224
12.3 Évolution des distances entre jeux de paramètres avec l'horizon de prévision 225
12.4 Évolution des paramètres de GRP en fonction de l'horizon de prévision . . . 229
12.5 Évolution des paramètres de GRP en fonction de l'horizon de prévision . . . 230
12.6 Stabilité des performances en contrôle en fonction de l'horizon au calage . . 232

13.1 Évaluation des performances du modèle sur 846 bassins : critère de persistance 239
13.2 Évaluation des performances du modèle sur 846 bassins : critère de persistance 241
13.3 Évaluation des performances du modèle sur 846 bassins : retard à la montée 242
13.4 Évaluation des performances du modèle sur 846 bassins : taux de fausses alertes 244
13.5 Évaluation des performances du modèle sur 846 bassins : indice de succès critique 246
13.6 Influence des performances en simulation sur les performances du modèle . . 250
13.7 Comparaison des performances du modèle calculées sur 178 ou sur 1024 bassins 254

14.1 Distributions des performances du modèle PQb 261
14.2 Comparaison géographique des performances de PQb et GRP 262
14.3 Distributions des performances des RNA employés seuls en prévision 265

15.1 Influence des lacunes de pluie sur les performances du modèle GRP 273
15.2 Influence de l'incertitude des observations de pluie sur les performances du modèle GRP 276
15.3 Influence de l'incertitude des prévisions de pluie sur les performances du modèle GRP 278
15.4 Influence du format temporel des prévisions de pluie sur les performances du modèle GRP 280
15.5 Influence du format temporel des prévisions de pluie sur les performances du modèle GRP 282
15.6 Impact d'erreurs dans les observations de débit 284
15.7 Impact d'erreurs dans les observations de débit : avec le filtre de Kálmán . . 285

15.8 Impact d'erreurs dans les observations de débit : avec une mise-à-jour incomplète . . . 287
15.9 Une du *Cri de Paris* du dimanche 13 février 1910 (après la crue centennale de la Seine à Paris). . . . 296

A.1 Régions hydrographiques françaises et leurs principaux cours d'eau. . . . 325

C.1 Structure du modèle GRP. . . . 341

D.1 A.C.P. réalisée sur 26 critères de performance . . . 354
D.2 A.C.P. réalisée sur 26 critères de performance . . . 356
D.3 A.C.P. réalisée sur 26 critères de performance . . . 358

G.1 Plan principal de l'A.C.P. sur les 6 indices décrivant le débit. . . . 424
G.2 Répartition de l'échantillon de 1 024 bassins en 4 groupes homogènes au sens des 6 indices décrivant le débit. . . . 424
G.3 Répartition géographique des quatre groupes homogènes de bassins versants définis à partir de l'échantillon de 1 024 bassins versants par la C.A.H. calculée sur les 6 indices décrivant les régimes hydrographiques. . . . 425
G.4 Répartition de l'échantillon de 1 024 bassins en 8 groupes homogènes au sens des 6 indices décrivant le débit. . . . 426
G.5 Caractéristiques physiques des 1 024 bassins versants. . . . 428
G.6 Histogrammes des 6 indices décrivant le débit. . . . 429
G.7 Histogrammes des 4 indices décrivant les pluies. . . . 430
G.8 Corrélations entre indices : densité locale 2 à 2. . . . 431
G.9 Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325. . . 434
G.10 Influence de différentes propriétés topographiques sur la répartition des bassins dans le plan principal de l'A.C.P. . . . 438
G.11 Influence du régime pluvieux sur la répartition des bassins dans le plan principal de l'A.C.P. . . . 440
G.12 Comparaison des 6 indices décrivant le régime hydrographique et des critères de performance sur les échantillons restreints obtenus par C.A.H. sur ces 6 indices . . . 442
G.13 Comparaison des 10 indices décrivant les régimes hydrographique et pluvieux et des critères de performance sur les échantillons restreints obtenus par C.A.H. sur ces 10 indices . . . 451
G.14 Distribution du critère de Nash-Sutcliffe en contrôle représentée dans le plan principal des A.C.P. . . . 463
G.15 Distribution du critère de Nash-Sutcliffe en contrôle représentée en fonction de la distance à l'origine dans l'espace des descripteurs . . . 465
G.16 Variété intra-groupe du critère de Nash-Sutcliffe en contrôle sur les échantillons restreints . . . 466

H.1 Taux de données douteuses en crue déterminé par un algorithme sup-seuil . 478
H.2 Localisation des bassins présentant des données non validées en hautes eaux 480
H.3 Comparaison des performances de GRP avec et sans la prise en compte des données douteuses . . . 481
H.4 Comparaison des performances de GRP-RNA avec et sans la prise en compte des données douteuses . . . 482

I.1 Exemple de dendrogramme 486

J.1 Fonction de production distribuée 492
J.2 Fonction de production globale 493
J.3 Différence entre sorties des fonctions de production globale et distribuée . . 495
J.4 Comparaison des différences de flux des fonctions de production globale et distribuée pour un événement sur l'Allier 496
J.5 Prévisions faites avec les fonctions de production globale et distribuée sur l'Allier 499
J.6 Différence maximales de performance des modèles avec des fonctions de production globale et distribuée 501

K.1 Effets d'une mise-à-jour des pluies efficaces sur R 508
K.2 Fonctionnement des principaux éléments de GRP 510
K.3 Fonction de routage du modèle GRP 512
K.4 Indépendance de l'observation et de la simulation conditionnellement à la réalité. 518

L.1 Propagation des particules avec perturbation sur les entrées 525
L.2 Mise-à-jour des poids des particules 526
L.3 Rééchantillonnage des particules 527
L.4 Le CRPS correspond à une mesure (par la norme $\|\cdot\|_2^2$ de l'aire (bleu clair) entre la distribution des prévisions obtenues (rouge) et la prévision parfaite (bleue). 529

M.1 Exemple de prévisions d'ensemble à partir des pluies ARPEGE 536
M.2 Exemple de diagramme de Talagrand obtenu à partir des pluies ARPEGE . . 537

O.1 Influence de γ sur la correction finale de Tangara (2005). 553

P.1 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 (horizon : 1 heure) 571
P.2 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 (horizon : 6 heures) 572
P.3 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 (horizon : 24 heures) 573
P.4 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 (horizon : 48 heures) 574
P.5 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 (horizon : 1 heure) 575
P.6 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 (horizon : 6 heures) 576
P.7 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 (horizon : 24 heures) 577
P.8 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 (horizon : 48 heures) 578
P.9 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en montée (horizon : 1 heure) 579
P.10 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en montée (horizon : 6 heures) 580

P.11 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en montée (horizon : 24 heures) 581
P.12 Carte de France des $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en montée (horizon : 48 heures) 582
P.13 Carte de France des retards moyens à la montée de crue (horizon : 1 heure) 583
P.14 Carte de France des retards moyens à la montée de crue (horizon : 6 heures) 584
P.15 Carte de France des retards moyens à la montée de crue (horizon : 24 heures)585
P.16 Carte de France des retards moyens à la montée de crue (horizon : 48 heures)585
P.17 Carte de France des horizons atteignables 586
P.18 Carte de France des indices de succès critiques (horizon : 1 heure) 587
P.19 Carte de France des indices de succès critiques (horizon : 6 heures) 588
P.20 Carte de France des indices de succès critiques (horizon : 24 heures) 589
P.21 Carte de France des indices de succès critiques (horizon : 48 heures) 590
P.22 Carte de France des probabilités de détection (horizon : 1 heure) 591
P.23 Carte de France des probabilités de détection (horizon : 6 heures) 592
P.24 Carte de France des probabilités de détection (horizon : 24 heures) 593
P.25 Carte de France des probabilités de détection (horizon : 48 heures) 594
P.26 Carte de France des taux de fausses alertes (horizon : 1 heure) 595
P.27 Carte de France des taux de fausses alertes (horizon : 6 heures) 596
P.28 Carte de France des taux de fausses alertes (horizon : 24 heures) 597
P.29 Carte de France des taux de fausses alertes (horizon : 48 heures) 598

Liste des tableaux

3.1 Significativité d'une différence moyenne de deux modèles sur plusieurs bassins (test de Wilcoxon et Mann-Whitney) . . . 49

4.1 Distributions des descripteurs physiques de nos 1 024 bassins versants . . . 60
4.2 Statistiques des indices décrivant les régimes hydrographiques et pluvieux sur les 1 024 bassins versants. . . . 62
4.3 Valeurs propres et vecteurs propres associés de la matrice de covariance des indices de débit et de pluie. . . . 63
4.4 Moyennes des indices pour les différents échantillons . . . 68
4.5 Écarts-types des indices pour les différents échantillons . . . 68
4.6 Comparaison des groupes définis par C.A.H. calculée sur 6 et sur 10 indices 70

5.1 Sensibilité au paramètre fixe A . . . 83
5.2 Sensibilité au paramètre fixe B . . . 84
5.3 Sensibilité au paramètre fixe α . . . 85
5.4 Sensibilité à la loi de vidange du réservoir de routage . . . 85
5.5 Sensibilité au paramètre fixe γ . . . 86
5.6 Performances moyennes de GR4P ($C_{2M,P}$) : l'exposant de la correction est un paramètre à caler.Le gain est défini par rapport au modèle GR3P avec un exposant γ de correction des sorties fixe. . . . 86

6.1 Les différentes combinaisons de fonctions non conservatives testées . . . 93
6.2 Performances moyennes des combinaisons de fonctions non conservatives . . 94
6.3 Performances médianes de GRP avec un hydrogramme asymétrique . . . 97
6.4 Performances moyennes du modèle GRP avec un hydrogramme unitaire décalé. 98
6.5 Performances liées à l'ajout d'une seconde branche . . . 101
6.6 Performances liées à l'ajout d'une seconde branche avec mise-à-jour de la répartition de la pluie efficace . . . 101

7.1 Importance de la mise-à-jour du réservoir de routage . . . 109
7.2 Gains de performance obtenus par mise-à-jour de l'hydrogramme unitaire . 114
7.3 Hypothèses du filtre de Kálmán . . . 118
7.4 Notations employées dans les sections consacrées au filtre de Kálmán . . . 119
7.5 Distribution des performances de GRP mis-à-jour avec le filtre de Kálmán . 126

8.1 Précision des prévisions de GRP avec filtre particulaire quand seules les entrées de pluie sont bruitées . . . 145

9.1 Performances médianes des différentes mises-à-jour des paramètres testées . 171

10.1 Caractéristiques des essais de débiaisage entrepris . . . 179
10.2 Revue de littérature concernant l'emploi des réseaux de neurones en hydrologie 190
10.3 Configuration des réseaux de neurones pour la correction de l'erreur . . . 194

13.1 Évaluation des performances du modèle sur 846 bassins : critères quadratiques 238
13.2 Évaluation des performances du modèle sur 846 bassins : critères temporels 243
13.3 Évaluation des performances du modèle sur 846 bassins : critères d'alerte . . 245

13.4 Évaluation des performances du modèle sur 846 bassins avec des pluies futures nulles 251

14.1 Configurations des RNA testés en prévision Pluie – Débit 263

15.1 Influence des lacunes de pluie sur les performances du modèle GRP 272
15.2 Impact d'erreurs dans les observations de débit 283

C.1 Récapitulatif des paramètres de la structure du modèle GRP. 341

D.1 Critères de performance utilisés dans l'A.C.P. 352
D.2 Valeurs propres cumulées de l'A.C.P. réalisée sur les 26 critères de performance 355
D.3 Valeurs propres cumulées de l'A.C.P. réalisée sur 19 critères de performance 355
D.4 Valeurs propres cumulées de l'A.C.P. réalisée sur 19 critères de performance 357

G.1 Valeurs propres et vecteurs propres associés de la matrice de covariance des indices de débit. 423

H.1 Répartition des hautes eaux et fortes pluies sur les données non validées . . 479

J.1 Différence des performances mesurées par le critère $C_{2M,P}$. La troisième décimale, non significative, n'est donnée que pour illustrer la faiblesse des résultats. 500
J.2 Différence des performances mesurées par le critère $C_{2M,P}$ au dessus d'un débit seuil. 500

Remerciements

Comment s'étaient-ils rencontrés ? Par hasard, comme tout le monde. Comment s'appelaient-ils ? Que vous importe ? D'où venaient-ils ? Du lieu le plus prochain. Où allaient-ils ? Est-ce que l'on sait où l'on va ? Que disaient-ils ?

Diderot, *Jacques le Fataliste*

Je remercie d'abord MM. Bárdossy, Leviandier, Bachoc, Dartus, Garçon, Kao et Parent qui ont accepté d'évaluer l'intérêt de ce travail.

Ensuite, ma gratitude va à Vazken Andréassian et Charles Perrin pour – et j'en oublie forcément – leur enthousiasme, les opportunités qu'ils m'ont ouvertes, leur patience et la qualité du suivi quotidien de ce travail.

L'ambiance qui règne dans l'équipe *Hydrologie* et plus globalement au sein de l'unité de recherche *Hydrosystèmes et Bioprocédés* a été un terreau très appréciable pour ce travail. Je dois beaucoup à toute l'équipe, notamment Julien Lerat et Nicolas Le Moine pour les discussions passionnées et « philosophiques », Cécile Loumagne pour son encadrement bienvaillant, Pierre Javelle qui m'a beaucoup aidé à comprendre les besoins des services opérationnels, Audrey Valery pour avoir partagé les principales étapes de ce travail en simultané, Mamoutou Tangara et Thibault Mathevet sans qui cette thèse ne serait pas, Marine Riffard et une certaine idée partagée de l'humour, Jean-Luc Payan parti courir en Nouvelle-Zélande, Laurent Coron pour sa mise en place opérationnelle du modèle et à qui je souhaite bonne chance pour sa thèse qui débute, Maria-Helena Ramos et Annie Randrianasolo pour une collaboration en arpèges, Jean-Louis Rosique et Julien Peschard pour leur patience inépuisable et la qualité de leurs cartes, Adnan Tahir et Rémi Lamblin pour avoir réalisé de beaux stages de master malgré leur maître de stage débutant : Adnan tout particulièrement [1] pour avoir essuyé les plâtres et m'avoir appris comment rédiger une bonne page de remerciements. Je remercie en outre Xavier Litrico et Simon Munier pour leur conseils kálmánesques précieux.

Au delà du Cemagref, je tiens à saluer la gang hydrologique québécoise. Tout d'abord Vincent Fortin qui m'a accueilli dans son laboratoire et ouvert les portes de sa maison, fait découvrir et aimer son pays avec talent, ainsi que l'inférence bayésienne (avec talent également mais cela va sans dire). François Anctil ensuite, pour ses conseils neuronaux et Thomas-Charles Fortier-Filion pour la bonne humeur et l'efficacité qu'il a déployées au Cemagref et sa contribution décisive aux Service de Prévision des Crues français et québécois.

Des S.P.C., je remercie particulièrement mes futurs collègues de Paris et d'Orléans – notamment Yan Lacaze et David Goutx – qui m'ont permis par des échanges nourris, de mieux comprendre leurs problématiques et d'affiner la traduction que j'ai pu en faire en termes de recherche.

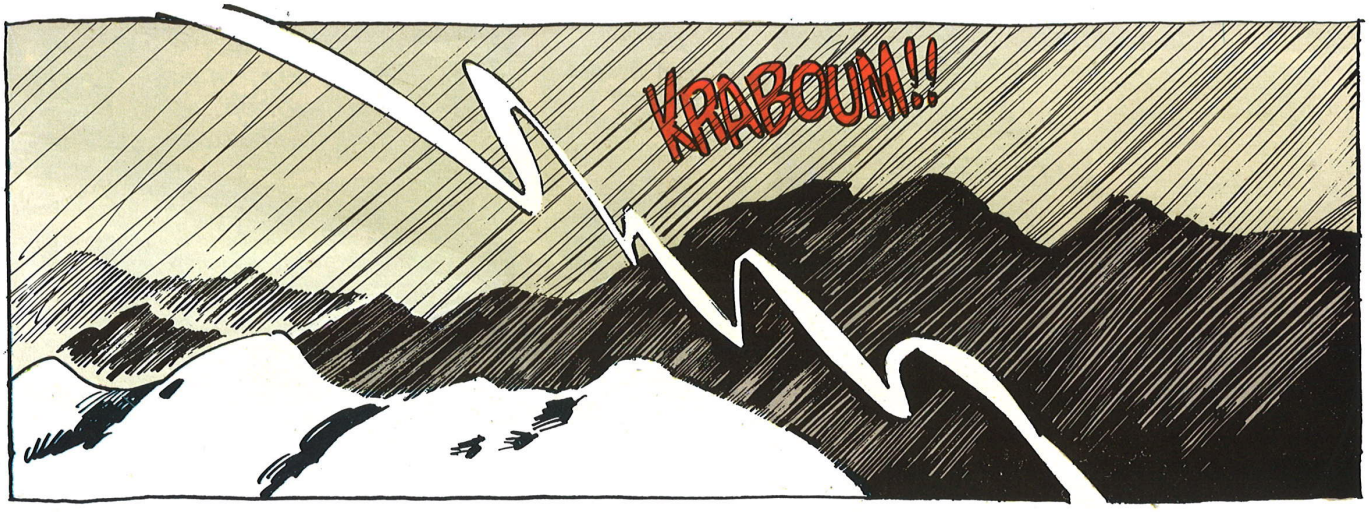
La recherche ne se fait pas ou se fait nettement moins bien sans l'appui de nombreuses personnes. Au premier rang, je souhaite remercier Michel Penel, Pierrick Givone et Jean-Philippe Torterotot pour leur suivi, Valérie Dansin, Sophie Morin et Catherine Geoffroy qui ont pallié avec le sourire mon amnésie récurrente concernant toute procédure administrative. Cette thèse n'aurait pas pu être réalisée sous cette forme sans la réactivité et la

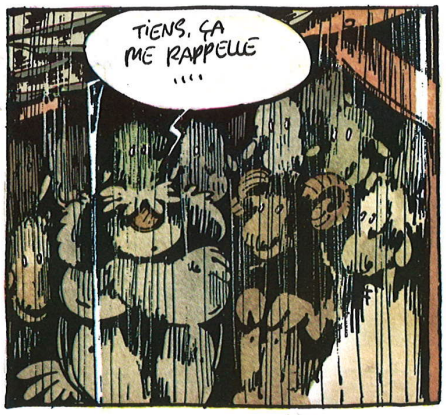
1. Rémi, moins téméraire, a prudemment travaillé avec trois encadrants dont deux confirmés.

fiabilité de Gilles Bonnet qui maintient le calculateur du Cemagref. Le service de documentation et de manière plus générale, les services d'apppui à la recherche du Cemagref créent un environnement de travail privilégié ; je les en remercie vivement.

Je profite de cette page pour saluer chaleureusement l'équipe d'Ultimate Frisbee du Cemagref qui m'a fait découvrir ce sport. Après trois ans de pratique, j'hésite encore entre les *Empirical Distributed Modellers* et les *Lumped Physically-based modellers*.

Enfin – ou aurais-je dû commencer par là ? – un grand merci à ma famille pour m'avoir soutenu et parfois ces derniers temps supporté. Mes amis m'ont également apporté leur regard curieux et leur aide, notamment Marion Ghibaudo et Sébastien Zimmer dont l'expertise en LaTeX 2_ε m'a fait gagner bien du temps.


KRABOUM!!


TIENS, ÇA ME RAPPELLE ....

JE NE VOUS AI JAMAIS RACONTÉ L'HISTOIRE DE ...
?
TIENS! NON!!

...

Contexte

Les crues : premières catastrophes naturelles en France

Les inondations figurent au premier rang des catastrophes naturelles dans le monde : elles causeraient environ 20 000 décès par an [2]. Cette primauté se retrouve également en Europe et en France :

De 1998 à 2007, *les inondations survenues en Europe se sont soldées par quelque 700 morts, par le déplacement d'environ un demi-million de personnes et par 25 milliards d'euros au moins de pertes économiques couvertes par les assurances* selon l'Agence européenne pour l'Environnement.

La France compte, quant à elle, 13 000 communes inondables et plus de 2 millions de personnes vivent dans des zones inondables. On dénombre 72 morts dues aux dix plus importantes inondations de ces dix dernières années. 80 % du coût des dommages imputables aux risques naturels sont causés par les inondations, soit en moyenne 250 millions d'euros par an.

Les politiques publiques fixent des objectifs de plus en plus ambitieux en matière de lutte contre les inondations, qui se sont traduits récemment par la directive 2007/60/CE du Parlement européen et du Conseil (qui rappelle, dès le premier paragraphe de l'exposé des motifs, que *les inondations constituent une menace susceptible de provoquer des pertes de vies humaines et le déplacement de populations, de nuire à l'environnement, de compromettre gravement le développement économique et de saper les activités économiques de la Communauté*). En France, le passage récent des Services d'*Annonce* des Crues aux Services de *Prévision* des Crues (2002) et la création du SCHAPI [3] (2003) illustrent cette volonté.

Intérêt de la prévision

Les mesures de protection structurelle (ouvrages) sont loin de pouvoir protéger intégralement les populations et les infrastructures humaines (Thielen *et al.*, 2009). La commission d'enquête de l'Assemblée nationale (2001) consacrée aux causes des inondations répétitives ou exceptionnelles rapporte que *l'importance des masses d'eau concernées rend assez illusoire de vouloir lutter contre les inondations* (pages 23 et 33).

Les analyses coût-bénéfice montrent dans de nombreux cas l'intérêt socio-économique de la prévision. Ainsi, le National Hydrologic Warning Council (2002) évalue la réduction des dommages de crues obtenue par les services de prévision hydrologique du *National Weather Service* à 1,62 milliards de dollars (USD) quand la moyenne des dommages annuels s'établit

2. http://www.prim.net, site portail thématique du ministère de l'Écologie, du Développement et de l'Aménagement durables dédié à la prévention des risques majeurs

3. Service central d'Hydrométéorologie et d'Appui à la Prévision des Inondations.

à 4,3 milliards de dollars sur la période 1981 – 2000. Mieux, cette même étude estime à 766 millions de dollars les gains à attendre d'une amélioration de ces services.

Approches opérationnelles communément employées

La prévision s'appuie d'abord sur l'expertise humaine et la connaissance des bassins versants. Les prévisionnistes disposent également d'approches empiriques simples, comme des régressions portant sur les débits.

De nombreux outils hydrauliques exploitent les propriétés de la propagation de l'onde de crue, comme par exemple les réglettes de Bachet (1934). Différents outils plus ou moins complexes existent, allant jusqu'à une modélisation hydraulique détaillée du réseau hydrographique. Quand elles ne sont pas couplées à des modèles hydrologiques, ces méthodes ne peuvent fonctionner qu'en aval de points de mesure et leurs horizons de prévision sont limités par la dynamique des crues (temps de propagation dans le réseau hydrographique). Si elles sont utiles sur les crues des grands cours d'eau, elles sont inadaptées aux crues éclairs les plus violentes.

Des méthodes d'analyse statistique du risque de crue en fonction des prévisions probabilistes de précipitation permettent d'étendre les alertes (e.g. Lavabre, 2005, Norbiato *et al.*, 2008) mais ne renseignent que très peu sur l'intensité des événements à venir.

Afin d'allonger les horizons de prévision, les prévisionnistes font de plus en plus appel à la modélisation hydrologique Pluie–Débit qui permet d'atteindre des horizons plus lointains sous réserve de connaissance de la pluie et donc de gagner du temps pour prévenir les riverains et prendre les mesures de protection nécessaires. Cela se traduit par une coopération plus forte entre services de prévision météorologique et de prévision hydrologique, notamment dans les structures organisationnelles : le SCHAPI est implanté auprès de Météo-France, les *River Flood Forecast centers* américains dépendent du *National Weather Service*, les hydrologues et les météorologues suédois du *Sveriges meteorologiska och hydrologiska institut* (SMHI) partagent les mêmes locaux et le système européen d'alerte aux crues (EFAS) s'appuie sur les prévisions du centre européen pour les prévisions météorologiques à moyen terme (ECMWF).

Ces outils restent largement perfectibles et de nombreuses difficultés demeurent, tant au niveau de la modélisation elle-même qu'en ce qui concerne la coordination avec les prévisions de précipitations quantitatives. Des outils opérationnels sont développés qui visent à optimiser la combinaison de la modélisation hydrologique actuelle et les données disponibles en temps réel. C'est dans cette perspective que s'inscrit ce travail de thèse.

1
Chapitre

Introduction

Il y a des autheurs, desquels la fin c'est dire les evenemens. La mienne, si j'y sçavoye advenir, seroit dire sur ce qui peut advenir.

Montaigne, *Essais*, livre I, chapitre 21 : *De la Force de l'Imagination*

Objectifs

- ▷ Définir notre objectif opérationnel.
- ▷ Introduire les concepts communs à l'ensemble de ce travail et destinés à répondre à cet objectif.
- ▷ Présenter les problématiques liés à ces concepts...
- ▷ ... et en déduire une nouvelle formulation de nos objectifs.

Sommaire

1.1 Première formulation des objectifs 5
1.2 Modélisation Pluie – Débit 5
1.2.1 Objectifs de simulation et objectifs de prévision 5
1.2.2 Typologie rapide des modèles existants 6
1.2.3 Niveau de complexité requis 7
1.2.4 Limites de la modélisation Pluie – Débit 8
1.3 Assimilation de données et mise-à-jour du modèle 9
1.3.1 Définitions 9
1.3.2 Problématiques soulevées 10
1.4 Reformulation de nos objectifs 11
1.5 Organisation de ce mémoire 12
1.5.1 Corps principal 12
1.5.2 Ce dont nous ne parlerons pas ou très peu 13
1.5.3 Conventions de lecture 13

1.1 Première formulation des objectifs

L'objectif initial de ce travail est d'améliorer les prévisions de crue faites par un modèle hydrologique de type Pluie – Débit sur les bassins versants français pour lesquels une prévision doit être établie (stations réglementaires). Le modèle fourni doit impérativement être performant, c'est-à-dire donner les prévisions les meilleures possible et se comparer favorablement à des modèles alternatifs. Il doit en outre répondre à un cahier des charges permettant son usage dans les services opérationnels. Nous le voulons (autant que possible) :

1. robuste : le modèle doit fonctionner à partir des données communément mesurées par les services opérationnels et ne pas être complètement dépourvu face à des situations de fonctionnement dégradées (données manquantes ou mal estimées par exemple) ;
2. simple d'emploi : un modèle qui exige le contrôle de nombreuses données et de nombreuses opérations de la part du prévisionniste est inutilisable en situation de crise ;
3. capable de fournir une représentation du fonctionnement du bassin versant compréhensible par l'hydrologue : ce dernier pourra alors construire son expertise du modèle, accroître sa confiance en lui, voire anticiper une partie de ses erreurs ;
4. non spécifique à un site particulier mais pouvant s'adapter à différents bassins versants.

À première vue, ces objectifs concernent l'ingénieur. Il ne s'agit pas d'accroître les connaissances théoriques en hydrologie, mais plutôt de s'appuyer sur les connaissances hydrologiques pour atteindre les meilleures performances possible. On imagine cependant mal améliorer des prévisions de façon robuste sans comprendre ce qui les améliore.

Dans le reste de cette introduction, nous allons explorer les facettes de cette première formulation des objectifs pour aboutir à une reformulation plus précise.

1.2 Modélisation Pluie – Débit

1.2.1 Objectifs de simulation et objectifs de prévision

La modélisation Pluie – Débit a pour but de rendre compte de la transformation des précipitations sur une aire donnée (le bassin versant[g], [1]) en le débit qui en résulte à l'exutoire[g] de ce bassin (Fig. 1.1). Il s'agit au départ d'un exercice de simulation répondant à un objectif de compréhension scientifique (i.e., expliciter les différents flux et phénomènes associés, voire les réduire à des phénomènes physiques plus fondamentaux). La simulation hydrologique peut également être motivée par un objectif pratique : rendre compte des séries temporelles de débits.

1. Le supscript [g] indique que le mot ainsi annoté est défini dans le glossaire, page 318.

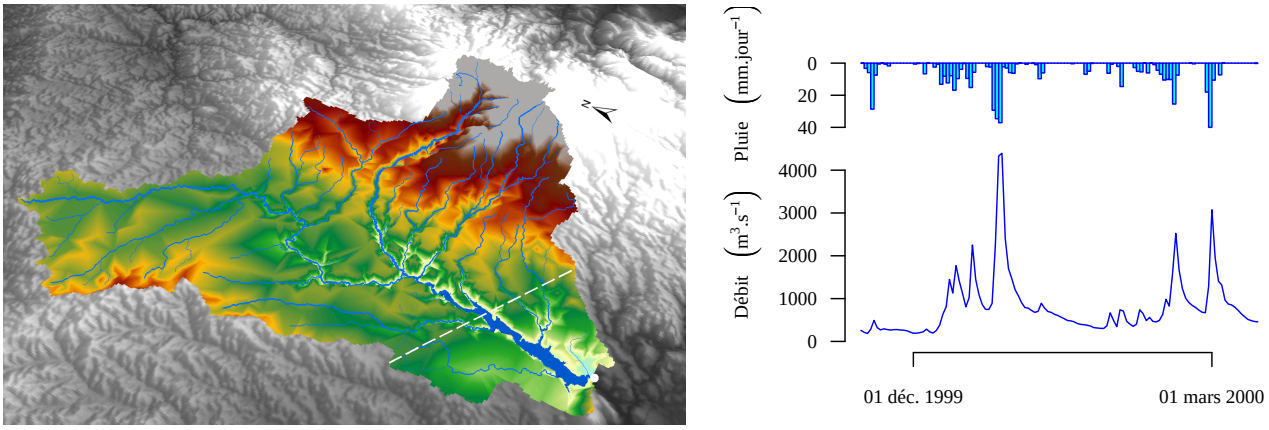


FIGURE 1.1 – À gauche, le bassin versant (de surface) de la Dordogne à Bort-Les-Orgues (1 010 km^2). Son contour suit les lignes de crête. Le réseau hydrographique est coloré en bleu. Le segment pointillé blanc mesure approximativement 14 km. L'objectif de la modélisation Pluie – Débit est de rendre compte de la transformation des précipitations sur le bassin versant en son débit à l'exutoire (signalé par le point blanc). Le graphique de droite illustre cette transformation sur ce bassin versant de la Dordogne pour l'hiver 1999 – 2000.

Cette modélisation peut également être employée à des fins de prévision. Il s'agit d'un objectif très différent des précédents. La simulation s'intéresse de façon égale à tous les pas de temps passés ou éventuellement à tous les pas de temps répondant à certaines caractéristiques (crues, étiages...) et n'utilise en entrée que les données causales du débit. La prévision se concentre sur un petit nombre de pas de temps suivant l'instant de prévision, démarcation entre le passé et le futur, et s'autorise l'emploi de toute donnée passée explicative (mais non nécessairement causale).

1.2.2 Typologie rapide des modèles existants

Face à cette diversité d'objectifs, les hydrologues ont adopté différents angles d'attaque (Sivapalan *et al.*, 2003a) et développé une multitude de modèles différents (e.g. Singh, 1995, Singh et Frevert, 2002, etc). Cette apparente profusion ne doit cependant pas nous faire oublier que de très nombreux modèles partagent les mêmes concepts et sont finalement très proches (Moore *et al.*, 2005). Parmi beaucoup d'autres, Wagener *et al.* (2003) ont proposé une classification de ces modèles basée sur leur approche (allant du réductionnisme à des processus physiques élémentaires à l'empirisme) et sur leur complexité : ils distinguent ainsi les modèles distribués « à base physique »(g) des modèles conceptuels(g) et des modèles empiriques(g).

Les modèles se distinguent également par leur résolution spatiale : globaux(g), ils considèrent le bassin versant comme une unité spatiale. Semi-distribués(g) ou distribués(g), ils dé-

coupent le bassin versant en sous-unités hydrologiques (bassins versants) ou en mailles géométriques. Enfin, on distingue classiquement fonctionnements événementiel(g) ou continu(g), selon le mode d'initialisation des modèles choisi.

1.2.3 Niveau de complexité requis

La gamme des niveaux de complexité des modèles hydrologiques existants est large. Il est donc nécessaire d'estimer la complexité que nous sommes capables d'introduire raisonnablement dans un modèle Pluie – Débit (Tetzlaff *et al.*, 2008). Cette question est centrale en modélisation hydrologique (Sivapalan, 2003, Blöschl, 2006, McDonnell *et al.*, 2007, Cho et Olivera, 2009, Micovic et Quick, 2009, parmi beaucoup d'autres).

Dawdy (2007) rappelle que l'hydrologie a d'abord émergé comme une discipline de l'ingénieur en proposant des approches très simples. Cet auteur souligne la dichotomie actuelle entre boîtes noires (modélisation des séries temporelles) et équations physiques semi-empiriques (compréhension des processus sous-jacents), et appelle à une jonction entre ces deux stratégies.

Beven (2001) se demande à quel point un modèle peut être approximatif et conserver un élément de « réalisme » dans ses explications et prédictions, le réalisme étant le plus souvent traduit en une quête réductionniste vers des niveaux (plus) fondamentaux d'explication. Des modèles plus complexes peuvent en principe être espérés plus réalistes mais seront plus difficiles à réfuter car ils présentent plus de degrés de liberté. Cependant selon les arguments présentés par Popper (1934), plus le modèle compte de degrés de liberté, moins il est explicatif.

De fait, ces mêmes auteurs relèvent l'absence d'implication logique entre profondeur explicative et pouvoir prédictif. Ce qui n'est pas sans rappeler le mathématicien et épistémologue René Thom (1993) qui illustrait la différence entre prédire et expliquer par l'exemple des inondations : « *On arrive souvent à comprendre des situations sans pouvoir agir : c'est le cas du monsieur qui, victime d'une inondation, monte sur son toit lorsque le niveau de l'eau monte !* »

Bergström (1991), Lindström *et al.* (1997) ont discuté la notion de *complexité optimale* du modèle au regard des données disponibles et des objectifs du modèle. Vingt-sept ans plus tard, Fenicia *et al.* (2008) rappellent encore qu'augmenter la complexité du modèle n'améliore pas nécesserairement ses performances et mettent en avant la recherche d'un équilibre entre une modélisation trop complexe qui est inapplicable et une modélisation simplifiée à l'excès dans le but de s'adapter aux données, l'empêchant de représenter la complexité du bassin versant. Gupta *et al.* (2008) insistent par ailleurs sur la nécessaire adéquation de la complexité du modèle à la quantité d'information contenue dans les données disponibles. Dans le prochain chapitre de ce mémoire, nous aurons à choisir le type de modèle en fonction de notre objectif particulier : la prévision des crues.

1.2.4 Limites de la modélisation Pluie – Débit

La modélisation Pluie – Débit rencontre un certain nombre de difficultés, parmi lesquelles se trouvent :

Les difficultés d'expérimentation limitant la corroboration des modèles
L'hydrologie souffre d'un mal commun à plusieurs sciences de la Terre : il est très difficile de réaliser une expérimentation contrôlée et on ne peut pas assurer la reproductibilité des « expériences » au sens strict. Les sciences du laboratoire mettent au point des expériences destinées à tester une théorie. Ces expériences consistent à placer l'objet étudié dans les N conditions $(X_i = x_i)_{i\in[\![1,N]\!]}$ qui, selon la théorie, permettent d'affirmer alors qu'on observera $Y = y$. Tout expérimentateur averti de la théorie peut monter la *même* expérience, c'est-à-dire placer un même objet dans les mêmes N conditions. Mieux, tout expérimentateur peut modifier les conditions X_i et vérifier que l'observation Y variera bien comme la théorie le prévoit. Cette multiplication des expériences permet d'atteindre un *degré de corroboration* (au sens de Popper) de plus en plus élevé... ou de réfuter la théorie.

En comparaison, le bassin versant est un objet difficilement isolable de son environnement (Sivakumar, 2008). L'hydrologue ne peut pas y convoquer la pluie pour monter ou reproduire une expérience. Il ne peut pas plus faire varier les conditions X_i à sa guise. Il ne dispose que de *mesures de terrain*, c'est-à-dire des mesures effectuées pendant une certaine durée (généralement courte) des grandeurs Z_i qui sont (plus ou moins bien) reliées aux grandeurs X_i permettant de déterminer Y selon la théorie. Il doit se contenter des séries de données disponibles dont la durée dépasse très rarement cinquante ans et compte peu d'événements de crues fortes. Aussi, les modèles hydrologiques ne peuvent pas atteindre de haut degré de corroboration sur un seul bassin versant.

Les limites des capacités d'expérimentation amoindrissent les possibilités de réfutation des modèles. Aussi, bien que les modèles hydrologiques soient très défectueux (Beven, 2006), il est souvent possible d'attribuer un échec du modèle au manque de données pertinentes ou à leurs incertitudes, ou encore à l'absence d'un jeu de paramètres adéquat qui en découle.

La question de la généralité des modèles
Les phénomènes physiques élémentaires commandant la transformation de la pluie en débit sont évidemment supposés universels. Mais leur combinaison varie de bassin en bassin. Aussi, certains hydrologues prônent la modélisation de chaque bassin afin de tenir compte de ses spécificités (voir par exemple Savenije, 2009).

Cependant, Linsley (1982) classe la généralité au même rang des propriétés à exiger de nos modèles que la précision ou l'applicabilité. Il lui paraît très difficile de tester nos hypothèses et d'améliorer nos connaissances théoriques si *a new model for every application eliminates the opportunity for learning that comes with repeated applications of the same model.* Pour

Clarke (2008), la reproduction (*replication*) des expériences de modélisation sur plusieurs bassins est une des deux conditions principales nécessaires à une bonne expérimentation.

La faiblesse des modèles

Selon Welles *et al.* (2007), relayé en cela par Kerr (2007), peu de progrès, si ce n'est aucun, ont été accomplis en prévision du débit à moyen terme (pour des horizons de prévision supérieurs à 3 jours) au court des vingt dernières années. Sans être aussi pessimiste, il nous faut reconnaître que les capacités de prévision restent largement perfectibles. Il existe plusieurs sources majeures d'incertitude qui limitent nos capacités prédictives : les estimations des précipitations (sur le bassin versant), les prévisions de précipitation future et le modèle lui-même. Voyons donc maintenant comment chercher à améliorer la qualité des modèles de prévision.

1.3 Assimilation de données et mise-à-jour du modèle

Face à la faiblesse des modèles hydrologiques et aux sources de larges incertitudes mentionnées précédemment, l'hydrologue est amené à améliorer ses prévisions de débit par des méthodes complémentaires. Liu et Gupta (2007) indiquent trois voies à explorer : l'amélioration des modèles hydrologiques, celles de l'instrumentation permettant d'obtenir des données nouvelles et / ou plus informatives et enfin l'assimilation de données. C'est à cette dernière proposition que nous allons nous attacher.

1.3.1 Définitions

La définition de l'assimilation de données présente différentes facettes. Selon McLaughlin (1995), l'origine de l'emploi de ce concept en hydrologie réside dans les sciences atmosphériques. Liu et Gupta (2007) définissent le problème de l'assimilation de données comme celui de la jonction entre le modèle et les données (*merging models with data*). Il s'agit d'employer *au mieux* deux informations de natures différentes et toutes deux entachées d'incertitude : celles provenant du modèle et celles contenues dans les observations (Walker et Houser, 2005). Ces techniques conduisent soit à une meilleure connaissance de l'état présent du système, soit à formuler une correction sur les sorties du modèle.

McLaughlin (2002) distingue trois familles d'assimilation de données : l'interpolation, le lissage et le filtrage. C'est cette dernière catégorie qui nous intéresse particulièrement dans le cadre de la prévision : elle vise à caractériser l'état actuel du système grâce à des observations passées et présentes[2].

2. En comparaison, le lissage est une réanalyse permettant d'obtenir *a posteriori* la série temporelle la plus probable à partir des séries des observations et des simulations. L'interpolation est l'exercice équivalent dans le domaine spatial.

L'assimilation de données permet de *mettre à jour* le modèle, c'est-à-dire de modifier un ou plusieurs de ses éléments pour que la représentation qu'il donne du système (i.e., le bassin versant) étudié combine au mieux l'information provenant des observations et la connaissance affirmée dans le modèle. Nous décrirons les différentes approches possibles dans le prochain chapitre.

1.3.2 Problématiques soulevées

Comme nous le verrons dans le chapitre suivant, des approches de modélisation et / ou de mises-à-jour du modèle très variées et bien différentes sont proposées afin de prévoir les crues. Il y a loin d'avoir unanimité au sein de la communauté hydrologique sur le type d'outils à employer et peu de comparaisons poussées (sur de nombreuses approches et sur un grand échantillon de bassins versants) ont été réalisées. Cette thèse a pour but de réaliser une telle comparaison, et par là, d'apporter des éléments de réponse à certains des nombreux défis de recherche soulevés par l'usage de techniques d'assimilation de données (Walker et Houser, 2005). Parmi ceux-ci, on compte :

La représentation des incertitudes
L'incertitude est au cœur du projet de l'assimilation de données. En effet, c'est pour lutter contre les incertitudes du modèle ou de ses entrées que l'assimilation de données est employée. De fait, on trouve de nombreuses définitions de l'assimilation de données en termes probabilistes (e.g. McLaughlin, 2002). En cela, ces approches rencontrent les aspirations de nombreux hydrologues (e.g. Houdant, 2004, Pappenberger et Beven, 2006). L'un des éléments clefs de l'assimilation de données est alors le modèle d'erreur (Walker et Houser, 2005).

Prise en compte des incertitudes des observations ?
Cependant, nous verrons dans les chapitres suivants qu'il existe de nombreuses méthodes de mise-à-jour des modèles qui sont formulées de façon déterministe. Est-ce contradictoire ? Walker et Houser (2005) font tout d'abord remarquer qu'il est nécessaire que les incertitudes de mesure soient inférieures à celles de la modélisation.

Ensuite, les trois sources majeures d'incertitudes mentionnées plus haut créent une incertitude de modélisation très large. Au point qu'on peut se demander si l'incertitude sur les observations ne peut pas être considérée comme négligeable face à elles (à condition que les données assimilées soient rigoureusement contrôlées comme Walker et Houser, 2005, le préconisent). Dans les méthodes d'assimilation présentées dans ce mémoire, certaines seront « déterministes » dans le sens où l'incertitude sur les observations assimilées ne sera pas prise en compte. D'autres méthodes l'expliciteront et nous comparerons leurs résultats.

Faut-il adapter le modèle aux méthodes d'assimilation ?
Les interactions entre le modèle et les méthodes d'assimilation de données sont cruciales. McLaughlin (1995) compare l'intérêt de méthodes générales d'assimilation et des méthodes spécifiques à un modèle particulier. Les premières sont susceptibles d'être très efficaces (voire optimales) mais leurs qualités ne sont démontrées qu'avec des hypothèses qui sont rarement vérifiées pour les modèles hydrologiques. Les secondes sont moins puissantes *a priori* mais adaptées aux modèles existants. Le débat sur l'opportunité de construire des modèles Pluie – Débit moins performants mais parfaitement adaptés à telle ou telle technique d'assimilation de données reste ouvert (un exemple emblématique est donné par Young, 2002).

Cela amène la question suivante : faut-il considérer le modèle et les méthodes d'assimilation comme deux entités différentes, deux étapes de la modélisation dédiée à la prévision ? Ou bien est-ce un tout qui forme l'outil de prévision ? Cette question sera présente en filigrane dans de nombreuses parties de ce mémoire, à commencer par le chapitre 2 portant sur la méthodologie adoptée pour ce travail.

Le niveau de complexité requis
McLaughlin (2002) fait remarquer qu'il existe peu de cas où les problèmes d'assimilation optimale présentent une solution analytique. Il est donc nécessaire, ici encore, de rechercher un compromis entre efficacité et robustesse.

Il nous faut donc choisir la complexité du modèle hydrologique pour la tâche assignée, mais également celle des méthodes d'assimilation de données en fonction de notre objectif et de l'adéquation au modèle hydrologique lui-même. Par exemple, une méthode d'assimilation comptant plusieurs dizaines de paramètres a-t-elle un sens si le modèle hydrologique est parcimonieux ?

1.4 Reformulation de nos objectifs

Ce travail a pour objectif d'améliorer la qualité des prévisions de débits de crue sur des horizons de prévision allant d'une heure à quelques jours à l'aide d'un modèle fonctionnant aussi bien que possible sur un grand nombre de bassins versants français. Nous étudierons plus particulièrement l'amélioration de la qualité des prévisions que l'on peut attendre de la mise-à-jour du modèle. Ainsi défini, notre objectif est essentiellement pratique. N'oublions cependant pas que les méthodes d'assimilation de données peuvent constituer une information sur le fonctionnement du modèle : notre objet d'étude sera le modèle.

Nous comparerons différentes familles de méthodes d'assimilation de données appliquées à un même modèle hydrologique. Nous examinerons à travers ces comparaisons plusieurs problématiques :

1. Quelles sont les interactions entre mises-à-jour et modèle hydrologique ? Comment les mises-à-jour affectent-elles le modèle hydrologique et son interprétation ?
2. Ces interactions nous conduisent-elles à adapter le modèle hydrologique ? Est-il préférable de considérer modèle hydrologique et mises-à-jour comme deux modules indépendants ou doit-on les voir comme un tout ?
3. Quel est le niveau de complexité requis pour obtenir une assimilation de données efficace ? Réciproquement quel niveau de complexité peut-on intégrer dans les méthodes de mise-à-jour ?
4. Est-il toujours nécessaire de prendre en compte l'incertitude des données assimilées face aux incertitudes de modélisation ? Comment employer les mises-à-jour pour évaluer les incertitudes de prévision ?

1.5 Organisation de ce mémoire

1.5.1 Corps principal

Après ce chapitre introductif, ce mémoire se poursuit par cinq grandes parties. La première traite de la méthodologie de ce travail et compte trois chapitres qui s'intéressent successivement aux choix généraux de méthodologie (chapitre 2), aux critères de performance permettant d'évaluer nos modèles (chapitre 3) et enfin aux données que nous allons employer (chapitre 4).

La deuxième partie présente le passage du pas de temps journalier au pas de temps horaire. Le cinquième chapitre évalue l'influence du pas de temps sur la représentation des crues et sur la structure du modèle GR3P (travaux de thèse de Tangara, 2005, dont le modèle est le point de départ de ce travail). Le chapitre suivant traite la modification de la structure du modèle GR3P pour s'adapter au pas de temps horaire.

Vient ensuite le cœur de ce mémoire avec la troisième partie dédiée à l'amélioration des méthodes de mise-à-jour du modèle. Le septième chapitre s'intéresse aux mises-à-jour des états du modèle. Le huitième explore l'emploi d'une méthode bayésienne de mise-à-jour de ces états : le filtre particulaire. Les deux chapitres suivants traitent de la mise-à-jour des paramètres et des sorties du modèle.

La quatrième partie discute l'impact des mises-à-jour sur le fonctionnement du modèle, en commençant par l'initialisation du modèle (chapitre 11, rédigé sous forme d'un article) puis en évaluant l'influence de la mise-à-jour sur le calage du modèle (chapitre 12).

Enfin, la dernière partie évalue le modèle de prévision obtenu, d'abord sur un millier de bassins versants français (chapitre 13), puis par comparaison à d'autres approches de modélisation Pluie – Débit (chapitre 14). Le dernier chapitre examine la robustesse du modèle face à des conditions dégradées.

1.5.2 Ce dont nous ne parlerons pas ou très peu

Bien qu'ils soient cruciaux pour la prévision des crues par des modèles hydrologiques, nous ne traiterons pas (ou si peu) de plusieurs sujets afférents. Le premier d'entre eux est le calcul et l'emploi des prévisions quantitatives des pluies futures. Ce thème est un sujet de recherche à part entière.

La question du calage des modèles hydrologiques est fondamentale et fait l'objet de publications abondantes. Nous ne l'examinerons qu'à travers la question des interactions avec les méthodes d'assimilation de données et de mise-à-jour des paramètres.

Enfin, la question des bassins non jaugés ne sera pas examinée. Cette question épineuse fait l'objet d'un large effort de recherche à travers la décennie PUB (Sivapalan *et al.*, 2003b). Elle est encore plus difficile ici puisque l'assimilation appelle à l'emploi de plus de données que ne le fait le modèle Pluie – Débit seul. Nous verrons cependant dans l'annexe N que nos travaux présentent encore un intérêt dans les bassins occasionnellement jaugés.

1.5.3 Conventions de lecture

Les deux premières annexes (situées en pages 318 et 328) ont pour but de faciliter la lecture de ce mémoire. La première est un glossaire proposant des définitions des termes scientifiques ou techniques qui apparaîtront dans le texte. La présence d'un mot au glossaire est indiqué par le supscript $^{(g)}$.

La deuxième annexe rappelle la signification de toutes les notations employées au cours de ces pages. Très classiquement, les symboles minuscules gras indiquent des vecteurs et les symboles majuscules gras des matrices.

Première partie

Méthodologie et données

2
Chapitre

Méthodologie

Pluralitas non est ponenda sine necessitate. / There are many methods for predicting the future. For example, you can read horoscopes, tea leaves, tarot cards, or crystal balls. Collectively, these methods are known as 'nutty methods'. Or you can put well-researched facts into sophisticated computer models, more commonly referred to as 'a complete waste of time'.

Guillaume d'Occam / Scott Adams, *The Dilbert Future*

Objectifs

- ▷ Préparer le cadre général des expériences présentées dans la suite de ce travail.
- ▷ Pourquoi et comment choisir un grand nombre de bassins versants ?
- ▷ Comment travailler avec des bassins versants différents ?

Sommaire

2.1 Construire un modèle (Pluie, Débit)→ Débit 19
2.1.1 Complexité optimale du modèle 19
2.1.2 Quelle donnée assimiler ? 20
2.1.3 Différentes approches de mise-à-jour du modèle 21
2.1.4 Le modèle est-il un ou deux ? 24
2.1.5 Une démarche itérative par comparaison de modèles 25
2.1.6 Quelles améliorations proposer et retenir ? 26
2.2 Faut-il être au plus près des conditions opérationnelles ? 26
2.2.1 Tester le modèle sur l'inconnu 27
2.2.2 Quels horizons de prévision ? 28
2.2.3 Quels scénarios de pluies futures ? 29
2.2.4 Modélisation continue ou événementielle ? 29
2.3 Multiplier les expériences à défaut de pouvoir les reproduire . 30
2.3.1 Motivations 30
2.3.2 L'échantillon de bassins versants disponibles 30
2.3.3 Tous les bassins n'ont pas la même dynamique temporelle 31
2.3.4 Tester les capacités d'extrapolation du modèle 33
2.4 Au-delà de la recherche de l'amélioration des performances . . 35
2.5 Présentation graphique des résultats 35

Nous avons décrit dans l'introduction les objectifs vers lesquels nous voulons tendre. Nous présentons ici nos choix de méthodologie pour les atteindre. Tout ce chapitre sera parcouru par la dualité soulevée dans le chapitre précédent : fournir un outil pratique qui améliore les prévisions de débit ; améliorer notre compréhension du fonctionnement du modèle et par là, peut-être, du bassin versant. Ces deux buts, aussi différents soient-ils, ne sont pas opposés et peuvent être utiles l'un à l'autre. Ils demandent d'examiner la plupart des questions de méthodologie sous au moins deux angles d'attaque différents.

2.1 Construire un modèle (Pluie, Débit)→ Débit

2.1.1 Complexité optimale du modèle

Quel type de modèle ?

Comme nous l'avons vu dans l'introduction, les modèles hydrologiques existants couvrent une large gamme de complexité. Il nous faut donc choisir le type de modèle adapté à notre application. En 1993, Brath et Rosso font remarquer que des modèles de type *boîte noire* sont préférés à des modèles conceptuels pour la prévision de crue. Depuis, de nombreux pays ont opté pour des modèles conceptuels relativement simples (O'Connor, 2006). Des modèles distribués à base physique sont également utilisés pour la prévision des crues : Marine en est un exemple (Estupina Borrell *et al.*, 2005).

Outre les qualités pratiques énumérées dans l'introduction, le modèle devra pouvoir être mis-à-jour efficacement. McLaughlin (2002) pointe la nécessité de travailler avec des modèles suffisamment parcimonieux pour cela.

Nous choisissons de travailler sur un modèle global hybride mi-conceptuel mi-empirique [1]. Le développement des modèles de la famille GR du Cemagref a été réalisé avec le souci constant de la juste complexité de la modélisation. Ils sont adaptés à l'assimilation de données. Plus précisément, le modèle GR3P développé par Tangara (2005) pour la prévision au pas de temps journalier se caractérise par :

- **ses qualités fonctionnelles**
  - ▷ **Performance** : GR3P a été testé sur plusieurs centaines de bassins versants et obtient des performances honorables, en comparaison aux résultats d'autres modèles publiés dans la littérature.
  - ▷ **Frugalité** : il n'utilise en entrées que les données de pluie de bassin (observée et prévue) et de débit observé. Il emploie en outre une estimation de la moyenne interannuelle de l'évapotranspiration potentielle.

1. Classe de modèles qualifiés de *hybrid metric-conceptual models* par Wagener *et al.* (2004). Ces modèles présentent des fondements conceptuels mais leur développement est largement empirique.

▷ **Parcimonie** : ce modèle compte trois paramètres libres, ce qui est peu en comparaison de la plupart des modèles de prévision [2].

– **ses qualités opérationnelles** : il s'agit d'un modèle conceptuel proposant une représentation suffisamment simple pour permettre le développement d'une expertise par des opérateurs.

Choix de la résolution temporelle

Comme nous le verrons au chapitre 5, les crues sur de nombreux bassins présentent des dynamiques rapides demandant un pas de temps sub-journalier pour rendre compte des débits de pointe. Les données horaires de pluie et de débit sont maintenant disponibles en de très nombreuses stations françaises. Nous travaillerons principalement au pas de temps horaire. Tahir (2008) montre que le modèle résultant peut néanmoins fonctionner à des pas de temps intermédiaires (entre une heure et un jour).

2.1.2 Quelle donnée assimiler ?

La mise-à-jour du modèle est un élément-clef pour obtenir des prévisions utiles. Voyons quelles données nous pouvons assimiler.

Choix de la donnée assimilée

Le prévisionniste opérationnel dispose d'assez peu d'informations en temps réel [3]. Des études décrivant l'assimilation de différentes observations à un modèle hydrologique ont été publiées. Parmi les premières données assimilées, on trouve l'humidité du sol (e.g. Aubert *et al.*, 2003, Walker et Houser, 2005, Crow et Ryu, 2008). Malheureusement dans la pratique, l'humidité du sol n'est pas mesurée sur tous les bassins versants qui nous intéressent. En outre, elle n'est connue que de façon indirecte, le plus souvent par des mesures ponctuelles qu'il est difficile de relier avec un ou plusieurs états du modèle.

C'est pourquoi nous nous sommes tournés vers une autre observation classiquement assimilée pour la prévision des crues : le débit à l'exutoire du bassin versant. En France, cette donnée est mesurée sur la grande majorité des stations réglementaires, c'est-à-dire aux exutoires des bassins versants qui intéressent les Services de Prévision des Crues. Sa mesure n'est bien sûr pas parfaite : elle est le plus souvent indirecte (à travers des courbes de tarage qui ne sont pas toujours stationnaires et souvent extrapolées dans les domaines de

2. Plusieurs études ont insisté sur le fait que le calage des modèles ne permet de déterminer la valeur que d'un nombre limité de paramètres (par exemple Mein et Brown, 1978, Jakeman et Hornberger, 1993, Perrin *et al.*, 2001) : au-delà d'une demi-douzaine de paramètres, soit certains paramètres demeurent peu sensibles au calage (ils peuvent prendre n'importe quelle valeur dans une certaine gamme sans modifier la valeur de la fonction-objectif), soit on a de (gros) problèmes d'équifinalité (plusieurs jeux de paramètres bien différents conduisant à des niveaux de performances semblables)

3. Certes l'argument ne tient pas à long terme. La mise en place et l'entretien d'un réseau de mesures se décident sur, entre autres, les résultats de travaux de recherche qui aident à évaluer la pertinence économique de différents réseaux de mesures.

hautes eaux). Cette mesure est cependant considérée comme plutôt précise en comparaison à d'autres mesures hydrologiques.

Conséquences sur la modélisation

Dans ce cadre, l'assimilation de données ne consiste pas simplement à rajouter des entrées à notre modèle. Cette approche introduit un glissement dans la modélisation. On passe d'une modélisation causale d'un phénomène (la transformation de la pluie en débit vue comme l'estimation du débit en fonction de ses causes : la pluie, l'humidité du sol, la géologie, la couverture du sol...) à la prévision de débits, connaissant tout ce qu'il est possible de connaître à la date où l'on émet la prévision, sans forcément définir complètement une relation causale (phénoménologique) entre l'information assimilée et la sortie du modèle.

Le modèle doit dès lors être pensé comme un modèle (Pluie, Débit)→ Débit : il n'a plus seulement un fonctionnement séquentiel « au fil de l'eau ». Le débit devient une information primordiale dans certains cas : il peut être plus important que les entrées de pluie. Les améliorations apportées à un modèle de simulation Pluie – Débit peuvent alors ne plus être très utiles pour le prévisionniste (au moins pour les horizons courts).

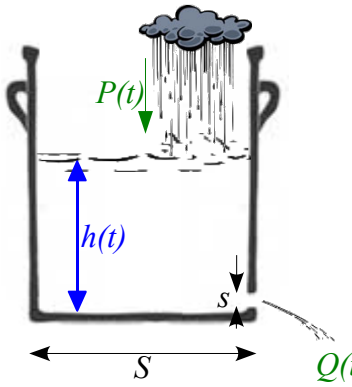
Une faiblesse de ce choix vient du fait que la même donnée, le débit, sert à la fois à la mise-à-jour du modèle (via son assimilation) et à son évaluation. En conséquence, si la mesure du débit est affectée d'une erreur systématique (par exemple due à la courbe de tarage), l'erreur lors de l'assimilation sera en apparence bénéfique pour l'évaluation des performances : les seules erreurs dans la mesure de la donnée assimilée dont nous pourrons tenir compte sont les erreurs aléatoires (accidentelles).

2.1.3 Différentes approches de mise-à-jour du modèle

Un modèle hydrologique peut être décrit dans une vision systémique (Liu et Gupta, 2007) comme étant un opérateur f, caractérisé par un jeu de paramètres(g) $\boldsymbol{\Theta}$ invariant dans le temps. L'état du système est caractérisé par un vecteur d'état $\mathbf{x}_t$. Le modèle f prend alors en compte des entrées $\mathbf{e}_t$ (pour nous les pluies et l'évapotranspiration potentielle et les autres entrées éventuelles du modèle) pour modifier le vecteur des états(g) et déterminer les sorties du modèle $\mathbf{y}_t$ (vecteur qui se réduit souvent aux débits pour nous). La mise-à-jour vise à modifier l'un de ces éléments en se basant sur l'information tirée d'observations $\mathbf{z}_t$. La figure 2.1 illustre cette description formelle dans le cas d'un modèle Pluie – Débit volontairement simpliste se réduisant à un réservoir unique.

Refsgaard (1997) distingue quatre types de mises-à-jour :

1. **Mise-à-jour des entrées** (pluie). Cette mise-à-jour se justifie par la médiocrité de notre connaissance des précipitations sur le bassin. Elle vise à améliorer les prévisions futures en modifiant les précipitations passées de façon à obtenir *a posteriori* de meilleures prévisions sur le passé.



Entrée	$P(t)$, volume d'eau précipitée par unité de temps.
États	$h(t)$, niveau d'eau dans le réservoir.
Paramètres	S, section du réservoir ;
	s, section de l'évacuation.
Sortie	$Q(t)$, débit sortant.

Fonctionnement : $\frac{dh}{dt} = \frac{P(t) - Q(t)}{S}$ avec $Q(t) = s\sqrt{2\,g\,h(t)}$

FIGURE 2.1 – Illustration de la description formelle d'un modèle Pluie – Débit par l'exemple d'un modèle simpliste se réduisant à un réservoir unique dont l'écoulement est régi par le théorème de Bernoulli.

2. **Mise-à-jour des paramètres**. Les paramètres(g) du modèle pour un bassin donné sont théoriquement des invariants temporels de la transformation Pluie – Débit et doivent décrire le fonctionnement du bassin. Ces paramètres sont calés sur une chronique finie [4] : même en faisant abstraction de l'erreur intrinsèque du modèle [5], il y a une incertitude sur les paramètres. Aussi, il est compréhensible de chercher à modifier (légèrement) les paramètres pour adapter le modèle aux conditions hydrologiques des derniers pas de temps. Le modèle fonctionne alors en mode *adaptatif*.
3. **Mise-à-jour des états**. Cette approche revient à supposer que du fait des incertitudes sur les entrées, sur les paramètres et de la fausseté du modèle, certains états n'ont pas la valeur qu'ils devraient avoir à la date d'émission de la prévision. Cette mise-à-jour vise à corriger leurs valeurs.
4. **Mise-à-jour des sorties ou correction de l'erreur**. Ici, on prend acte du dysfonctionnement du modèle sans chercher particulièrement à l'expliquer. La connaissance hydrologique n'étant pas suffisante pour établir de bonnes prévisions, on décide d'adjoindre au modèle conceptuel un modèle purement empirique de traitement de ses sorties (généralement un modèle de nature statistique (Troutman, 1985) ou de traitement du signal) pour tenir compte des propriétés des erreurs observées au calage.

Dans l'exemple simpliste présenté précédemment à la figure 2.1, la mise-à-jour peut consister à modifier les pluies précédentes P_{t-i} (mise-à-jour des entrées) ou bien à changer le niveau dans le réservoir h_t pour obtenir un meilleur accord des simulations aux observations (mise-à-jour des états) ou encore à modifier la taille du trou s pour moduler le débit

4. et généralement assez courte (quelques années), pouvant ne compter que peu d'événements de crue.
5. L'incertitude lié à la modélisation pourrait d'ailleurs être supérieure à celle des autres sources d'incertitude sur les paramètres. Le raisonnement pose ici un modèle parfait, ce qui est une hypothèse aventureuse.

simulé ou à changer la section du réservoir S pour adapter la hauteur d'eau dans le réservoir et donc influer sur le débit sortant (mise-à-jour des paramètres) ou enfin à opérer une correction statistique sur les débits sortants (mise-à-jour des sorties).

Walker et Houser (2005) différencient en outre approches séquentielles[(g)] et approches variationnelles[(g)] (Fig. 2.2) :

– L'**approche séquentielle** vise à modifier un ou plusieurs éléments du modèle en fonction des observations, à chaque fois que de nouvelles observations sont disponibles (il s'agit généralement de corriger l'estimation des états du modèle). Le filtre de Kálmán relève typiquement de cette approche.

– L'**approche variationnelle** consiste à faire varier une ou plusieurs variables (par exemple des états des pas de temps passés ou présent) juste avant d'émettre la prévision, de façon

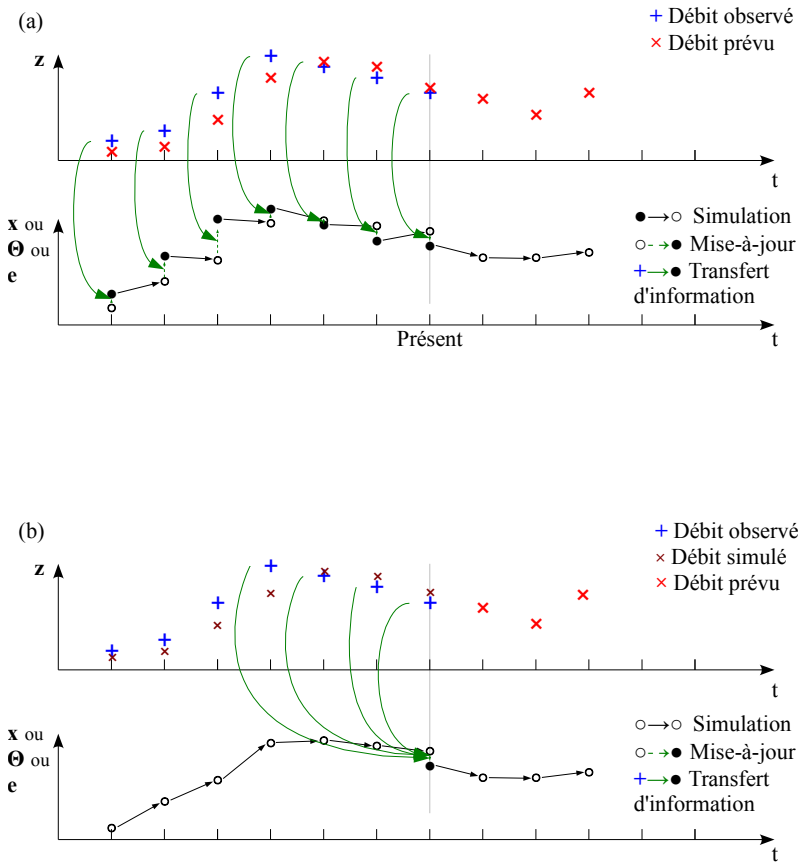


Figure 2.2 – (a) L'approche séquentielle met à jour le modèle à chaque fois qu'elle dispose d'information (observations) à assimiler, ce jusqu'à la date d'émission de la prévision. $\mathbf{z}_t$ est le vecteur des observations qui sont assimilées pour mettre à jour les états du modèle $\mathbf{x}_t$ ou ses paramètres $\boldsymbol{\Theta}$ ou encore les entrées $\mathbf{e}_t$.
(b) La mise-à-jour variationnelle emploie l'information sur les derniers pas de temps précédant l'instant de prévision pour mettre à jour le modèle à la date d'émission de la prévision.

à minimiser une fonction de coût (estimation de l'erreur) calculée sur les derniers pas de temps. Le recalage des paramètres peut être qualifié de variationnel (Brath et Rosso, 1993).

L'existence de plusieurs classifications possibles montre que les mises-à-jour sont elles-même des objets de recherche dont le fonctionnement et les conséquences doivent être analysés sous plusieurs angles. En conséquence, il nous faudra tester de nombreuses méthodes différentes d'assimilation du débit.

Parmi les approches de mise-à-jour les plus populaires, on retrouve la mise-à-jour des états (préconisée par Moore *et al.*, 2005) et les corrections des sorties. La mise-à-jour des états est le plus souvent réalisée par des méthodes de filtrage, le plus connu restant le filtre de Kálmán (1960). Les corrections des sorties sont très fréquemment basées sur des modèles auto-régressifs (Box et Jenkins, 1976) ou des réseaux de neurones artificiels (Maier et Dandy, 2000). Les comparaisons de méthode de mise-à-jour portant sur des éléments différents (par exemple, les états et les sorties du modèle) sont encore assez rares.

2.1.4 Le modèle est-il un ou deux ?

Deux approches de fonctionnement des mises-à-jour sont possibles. Dans une première approche, le modèle hydrologique se distingue nettement des mises-à-jour. Le modèle hydrologique est d'abord calé seul (i.e. en *mode simulation*). Dans un second temps seulement, les mises-à-jour sont introduites (et éventuellement calées).

Une approche plus intégrée est possible. Elle considère le modèle et ses mises-à-jour comme un tout indissociable. L'ensemble est calé directement (en une seule fois) sur un objectif de prévision. Le modèle de prévision se distingue alors plus nettement du modèle de simulation (outil de la connaissance de la transformation Pluie – Débit) dont il est issu. Il affirme clairement que son but premier n'est pas la compréhension mais l'obtention des meilleures prévisions possibles. Le meilleur calage n'est plus celui qui rend le mieux compte de la transformation Pluie – Débit mais celui qui *in fine* conduit aux meilleures prévisions : ce calage doit donc tenir compte des mises-à-jour. Tangara (2005) a montré que cette démarche permettait dans certains cas d'obtenir de meilleures prévisions. Cette approche présente cependant plusieurs limites. D'abord sur un plan pratique, le nombre de paramètres (modèle hydrologique et mises-à-jour) à caler en une seule fois augmente, ce qui peut rendre le calage plus difficile et moins robuste. Ensuite, les paramètres du modèle hydrologique peuvent être calés à des valeurs qui ont *a priori* moins de « signification » hydrologique du fait d'interactions inévitables avec les mises-à-jour [6]. Ceci n'est pas forcément un défaut en soi mais peut rendre une analyse hydrologique du calage (et donc du fonctionnement du modèle) plus difficile pour les usagers. Enfin, une telle approche peut être périlleuse en cas

6. Typiquement, ces paramètres peuvent dépendre (un peu) de l'horizon de prévision.

de données de débit incertaines ou manquantes : si l'opérateur « éteint » les mises-à-jour, le modèle résultant qui a été optimisé pour l'usage avec les mises-à-jour peut se révéler moins efficace que le modèle optimisé pour la simulation.

2.1.5 Une démarche itérative par comparaison de modèles

Les modèles hydrologiques Pluie – Débit sont des théories scientifiques et doivent donc, selon Popper (1934), être réfutables. Malheureusement dès qu'on utilise un critère de réfutation (binaire) un peu strict, ils sont réfutés : les prédictions du modèle ne concordent pas avec les observations avec toute la précision souhaitée. Conscients que tous les modèles sont « faux » dans cette optique exigeante (Beven, 2006), de nombreux hydrologues adoptent une démarche visant à rechercher le modèle le « moins faux » dans certaines limites d'acceptabilité (Beven, 2008a) pour continuer d'améliorer le modèle.

Comme le rappelle Kuhn (1969) dans son onzième chapitre, *toutes les théories ayant une importance historique ont été d'accord avec les faits, mais seulement plus ou moins. Et il n'y pas de réponse plus précise à donner à qui voudrait savoir si et dans quelle mesure une théorie particulière concorde avec les faits. Par contre, si les théories sont examinées collectivement ou même par paires, c'est une question utile que de se demander laquelle de deux ou plusieurs théories concurrentes s'adapte « le mieux » aux faits.* Ainsi, comparer des modèles devient la seule chose possible. Cette opération abandonne les critères binaires de réfutation pour un critère quantitatif d'adéquation (ou de fausseté) permettant d'identifier le modèle le moins déméritant.

Dans ce cadre, la construction du modèle est donc nécessairement itérative : il s'agit de tester des hypothèses de nature soit empirique soit théorique, pour établir le modèle atteignant les meilleures valeurs de ce critère.

Pour l'évaluation finale de notre modèle, nous chercherons à le confronter à plusieurs modèles de prévision. La comparaison entre différents modèles permet d'évaluer les points forts et les points faibles de ces modèles (Askew, 1989, WMO, 1992, Perrin *et al.*, 2001, Smith *et al.*, 2004). Nous utiliserons des modèles de natures assez différentes de celle de notre modèle car, comme le fait remarquer Moore *et al.* (2005), le nombre pléthorique des modèles Pluie – Débit conceptuels empiriques (dits modèles à réservoir) cache assez mal le fait que ces modèles sont en fait très similaires dans leurs conception et fonctionnement : ils peuvent être vus comme autant de maxima locaux situés sur le plateau (en performance) des meilleurs modèles de cette famille. Nous comparerons notre modèle à plusieurs autres utilisés opérationnellement (en France) tels que PQb, GR3H ainsi qu'à des modèles de natures radicalement différentes comme les plus proches voisins et les réseaux de neurones (Chap. 14).

2.1.6 Quelles améliorations proposer et retenir ?

Qui dit démarche itérative, dit point de départ. Le modèle GR3P [7], conçu pour le pas de temps journalier, sera cette structure initiale. Ce modèle est décrit dans l'annexe C (page 339). La première étape sera donc d'adapter le modèle au pas de temps horaire.

Un des charmes (et défauts) de l'approche empirique est qu'il existe une infinité de variantes possibles. Il faut donc limiter *a priori* le champ d'investigation aux essais jugés prioritaires. Pour assurer que notre modèle continue à respecter le cahier des charges que nous avons défini, nous proposons d'effectuer des trajectoires « quasi fermées » dans l'espace des modifications possibles : nous partirons des variantes les plus simples (conceptuellement et pour la mise en œuvre pratique) pour aller vers des approches plus compliquées en ne retenant ces dernières que si elles présentent un intérêt en comparaison des premières. Puis quand cela est possible, nous chercherons à simplifier les solutions retenues en éliminant leurs complexités inutiles (dans notre cas).

On peut d'ores et déjà entrevoir plusieurs axes de recherche d'améliorations :
- amélioration de la structure du modèle et comparaison avec les modèles de simulation Pluie – Débit,
- essais de nouveaux points d'entrée pour les mises-à-jour présentes dans le modèle de Tangara (2005),
- essais de mises-à-jour de natures et de complexités différentes.

2.2 Faut-il être au plus près des conditions opérationnelles ?

La question de la distance aux conditions opérationnelles à laquelle il faut se placer pour atteindre nos objectifs est au cœur de ce travail. Elle est liée à la dualité de nos objectifs : d'une part répondre à un besoin opérationnel et d'autre part comprendre, au moins partiellement, le fonctionnement des béquilles fournies au modèle (les mises-à-jour), qui peuvent révéler des défauts du modèle hydrologique (e.g. Fenicia *et al.*, 2008).

Deux voies potentiellement divergentes s'ouvrent devant nous. La première engrange les améliorations sans forcément se soucier d'apporter des progrès scientifiques ; on prend alors le risque de s'acheminer peu à peu vers un modèle de type *boîte noire* : les améliorations constatées deviennent de moins en moins explicables et le modèle peut perdre, outre son intérêt scientifique (ce qui n'est pas en soi très grave pour l'utilisateur), la possibilité d'être amélioré par les avancées de la compréhension hydrologique à venir. L'opérateur peut également perdre la capacité à développer son expertise et sa confiance dans le modèle. Par exemple, ceci peut arriver quand on utilise des méthodes statistiques d'une grande com-

7. Nous distinguons dans tout le texte, le modèle de prévision GR3P proposé par Tangara (2005) qui compte trois paramètres, du modèle GRP étudié dans ce mémoire et évoluant durant le travail présenté, qui compte un nombre variable de paramètres selon les mises-à-jours (paramétriques) proposées.

plexité (ayant de nombreux degrés de liberté) pour mettre à jour le modèle ou ses sorties, comme les réseaux de neurones. La seconde voie n'accepte au contraire des améliorations que si elles sont bien explicables par une théorie hydrologique, ce qui en limite drastiquement le nombre et peut conduire à des améliorations intéressantes sur le plan explicatif, mais faiblement utiles opérationnellement.

Dans la première optique, on veillera à se mettre dans les conditions le plus proches possible de celles du prévisionnistes (temps réel simulé en utilisant exactement les mêmes données que celles dont il dispose...) tandis que dans la seconde approche, on privilégiera les conditions qui permettent d'améliorer le seul modèle de prévision (et non le système complet de prévision) en laissant à d'autres le soin d'améliorer les conditions d'application du modèle pour les prévisionnistes.

Trop près des conditions opérationnelles, nous risquons de passer à côté d'une compréhension des causes expliquant les succès relatifs de nos différents essais ; trop loin, nous pourrions éventuellement produire un résultat intéressant intellectuellement mais inadapté à la réalité des opérations de prévision. Nous visons une voie intermédiaire qui définit ce qui nous paraît indispensable pour que le modèle soit bien adapté aux conditions du temps réel tout en laissant à d'autres, spécialistes en leurs domaines, le soin d'améliorer les conditions de fonctionnement du système de prévision dans lequel s'inscrit le modèle (par exemple, la prévision des pluies futures).

2.2.1 Tester le modèle sur l'inconnu

Les modèles que nous testons comportent des paramètres(g), c'est-à-dire des grandeurs spécifiques au bassin versant traité et considérées comme invariantes dans le temps. Si nous ne pouvons pas affirmer un contenu « physique » (c'est-à-dire reliable à des mesures de telle ou telle quantité physique sur le bassin) de ces paramètres, ni de connaissances *a priori*, leurs valeurs ne peuvent être estimées que par le calage du modèle : une première confrontation du modèle à des données.

Pour connaître la valeur d'un modèle hydrologique (pour un exercice de simulation), il est nécessaire de le tester sur des données qui n'ont pas été utilisées lors de sa mise au point. Les performances au calage d'un modèle ne sont pas suffisantes pour évaluer ses qualités prédictives (Refsgaard et Henriksen, 2004). Cela est encore plus vrai pour évaluer les capacités de prévision d'un modèle. Reproduire, même parfaitement, une série de données passées par le calage du modèle n'assure en aucun cas que le modèle sera capable de réaliser de bonnes prévisions sur d'autres séries.

Nous utilisons donc (très classiquement) le premier test de Klemeš (1986), dit de *calage-contrôle* (*split-sample test*). Il s'agit d'utiliser un premier ensemble de données pour caler le modèle (i.e., estimer les valeurs des paramètres qui permettent à notre modèle de rendre

compte au mieux des observations), puis de tester le modèle ainsi calé sur un nouvel ensemble de données, indépendantes des premières. Dans la pratique, nous scinderons la chronique des données disponibles en deux périodes. Nous commencerons par caler le modèle avec la première période P_1 pour obtenir un jeu de paramètres $\mathbf{\Theta}_1$ puis contrôlerons la qualité du modèle avec ce jeu de paramètres sur la seconde période. Dans une seconde expérience, nous utiliserons la seconde période P_2 pour caler le modèle que nous contrôlerons sur P_1 avec le jeu de paramètres $\mathbf{\Theta}_2$. Dans la suite de cette thèse, nous examinerons, sauf mention contraire, les performances en contrôle.

Nous n'appliquerons pas les trois autres tests définis par Klemeš. Le test de proximité spatiale est liée à la question des bassins non jaugés. Celui du *differential split-sample test* (qui consiste à choisir des périodes de calage et de contrôle contrastées) sera discuté plus loin quand nous nous intéresserons aux capacités d'extrapolation de nos modèles (page 33).

2.2.2 Quels horizons de prévision ?

Nous cherchons à établir un modèle fonctionnant aux pas de temps fins (heure ou quelques heures). Un modèle de prévision vise à décrire ce qui se passe dans le futur proche. Il ne cherche pas à être performant à tous les ordres de grandeur temporels : par exemple, il n'a pas pour objectif premier de fournir de bons bilans (volumiques) mensuels ou annuels. En conséquence, sa description de la transformation Pluie – Débit n'a plus à être exhaustive et s'intéresse aux phénomènes dont la dynamique correspond peu ou prou à l'horizon sélectionné. Le choix du pas de temps se base également sur les processus dont on veut rendre compte (e.g. Obled *et al.*, 2008). Les horizons de prévision ne devraient donc pas être très grands face au pas de temps. En outre, on verra dans la suite que les mises-à-jour possèdent leurs dynamiques propres (certaines pouvant être indépendantes des dynamiques du modèle et du bassin versant) : leur efficacité diminue quand le nombre de pas de temps augmente, ce qui plaide là encore pour des horizons d'un ordre de grandeur proche de celui du pas de temps.

Nous testerons chaque variante du modèle sur les horizons de 1, 3, 6, 12, 18, 24, 36, 48 et parfois 72 heures. L'horizon de 1 heure peut paraître d'utilité faible. Cependant, outre l'intérêt que le modélisateur peut y trouver pour comprendre le fonctionnement de son modèle, certains prévisionnistes (National Research Council (NRC), 2002) font remarquer qu'une bonne prévision à un horizon aussi réduit peut tout de même permettre de sauver des vies et de diminuer les dommages matériels jusqu'à 10 %. Au delà de 72 heures, il nous paraît raisonnable d'utiliser un modèle fonctionnant à un pas de temps plus long [8].

8. La résolution temporelle des prévisions de pluie est de toute façon inférieure à la résolution horaire à cette échéance : on choisira donc un pas de temps plus long.

2.2.3 Quels scénarios de pluies futures ?

Il est évident que pour évaluer les capacités de prévision de nos modèles, nous devons les faire tourner « comme si » nous ne connaissions pas les données après la date t à laquelle la prévision est établie. Aussi, le modèle (et ses mises-à-jour) ne doi(ven)t pas disposer des valeurs des débits postérieurs à la date t lors des tests. Cependant, le modèle a besoin de valeurs pour les précipitations entre les dates t et $t + H$ (date pour laquelle on prévoit le débit).

En conditions opérationnelles, le prévisionniste ignore les quantités d'eau qui vont effectivement précipiter entre l'instant où il émet sa prévision et l'horizon qu'il souhaite atteindre. Au mieux dispose-t-il de prévisions quantitatives de précipitations (PQP [9]), qui présentent une forte incertitude.

Nous disposons de très peu d'archives de PQP sur nos bassins. Plusieurs scénarios peuvent être employés à la place des PQP. Comme notre but est l'amélioration du modèle hydrologique de prévision (et non tout le système de prévision), nous préférons travailler avec les observations (non disponibles en conditions réelles lors de la prévision) des précipitations au-delà de l'instant d'émission de la prévision. Ce scénario très optimiste permet de ne pas inclure les précipitations futures comme source d'incertitude lors de la comparaison de deux modèles (ou tout au moins d'en limiter très fortement l'impact). Lors de l'évaluation finale du modèle, d'autres scénarios seront employés (notamment celui de précipitations futures nulles).

2.2.4 Modélisation continue ou événementielle ?

L'initialisation du modèle est une question pratique pouvant se révéler cruciale pour obtenir de bonnes performances en simulation. En effet, de nombreux modèles Pluie – Débit sont très sensibles aux conditions initiales et l'initialisation (événementielle) est souvent délicate à réaliser. En prévision, les mises-à-jour du modèle peuvent s'interpréter comme une réinitialisation partielle du modèle. Il est donc intéressant d'examiner la sensibilité du modèle de prévision à la stratégie d'initialisation. Cette étude sera menée au chapitre 11.

9. Souvent appelées QPF pour l'anglais *Quantitative Precipitation Forecasts*.

2.3 Multiplier les expériences à défaut de pouvoir les reproduire

2.3.1 Motivations

Comme nous l'avons vu dans l'introduction, nous cherchons à établir un modèle qui fonctionne aussi bien que possible sur différents bassins versants pour des raisons pratiques et scientifiques.

Un objectif pratique

Il nous paraît plus utile de proposer un modèle qui, s'il n'est pas optimal pour tel ou tel bassin versant, fournira des prévisions d'une qualité acceptable[10] sur la plupart des bassins versants d'un territoire. En outre, comme le rappelle Rémy Garçon[11], la confrontation d'un même modèle à différents bassins versants permet au prévisionniste d'enrichir considérablement son expertise en ce modèle.

Justifications méthodologiques

Comme nous l'avons vu dans l'introduction, l'hydrologue doit faire face à la faible reproductibilité de ses « expériences » sur chaque bassin versant. Or les chroniques généralement disponibles pour un bassin versant contiennent finalement très peu de signal (en particulier si on s'intéresse aux crues). Aussi, même si nous souscrivions complètement à la thèse selon laquelle il est illusoire de chercher à identifier un modèle idéal unique pour tous les bassins (e.g. Savenije, 2009), il est nécessaire de tester le modèle sur de nombreux bassins (même s'ils sont « hydrologiquement » similaires) pour multiplier les possibilités de réfutation : la probabilité que les performances du modèle soient acceptables par hasard sera plus faible et les conclusions plus robustes. Il pourrait s'agir, selon Linsley (1982), de la seule voie permettant de détecter les erreurs de modélisation et de les corriger. Au final, nous espérons qu'un tel modèle « contienne » plus d'hydrologie qu'un modèle ne pouvant reproduire correctement les débits que sur un seul bassin ou seulement pour des bassins très semblables (Andréassian *et al.*, 2006).

2.3.2 L'échantillon de bassins versants disponibles

Nous disposons d'un échantillon de chroniques au pas de temps horaire sur un millier de bassins versants (collectées et préparées par Le Moine, 2008), c'est-à-dire des séries temporelles de pluie, de débit et d'évapotranspiration potentielle sur ces bassins versants. Ces données seront décrites et analysées au chapitre 4.

10. En comparaison aux performances qu'obtiendraient d'autres modèles sur ces bassins versants.
11. Remarque orale lors du colloque *Prévisions hydrométéorologiques* organisé par la Société hydrotechnique de France à Lyon, les 18 et 19 novembre 2008.

2.3.3 Tous les bassins n'ont pas la même dynamique temporelle

L'exercice de prévision se distingue de celui de simulation par la prise en compte d'une dimension temporelle supplémentaire à travers l'horizon de prévision. Cette durée doit être comparée à un *temps caractéristique* de la relation Pluie – Débit sur chaque bassin. En effet, prévoir à 6 heures est un exercice relativement simple pour les crues de la Somme (crues de nappe) mais est difficile sur les Gardons. Il est donc nécessaire de tenir compte de l'existence de bassins *rapides* et de bassins *lents* au sein de notre échantillon (Fig. 4.2, Chap. 4, page 59).

Les difficultés apparaissent quand on veut définir précisément les mots *rapides*(g) et *lents*(g). Plusieurs approches sont possibles. On peut d'abord s'intéresser à la dynamique de la transformation de la pluie en débit, par exemple en recherchant un temps de réponse aux pluies caractéristique du bassin. Le paramètre (calé) X_3 du modèle GR3P (Tangara, 2005), qui contrôle la durée de l'hydrogramme unitaire, peut être lu comme un indicateur de cette dynamique. Mais choisir cet indice impose une vision du bassin versant : celle du modèle dont la qualité est, reconnaissons-le, toute relative. Il nous paraît dangereux de qualifier une grandeur caractéristique du bassin par la vision qu'en a notre modèle. Une autre approche serait d'étudier les séries de pluie et de débit en utilisant la durée Δt qui maximise la corrélation entre les séries $\{P_t\}_t$ et $\{Q_{t+\Delta t}\}_t$. Cet indice semble pertinent pour les deux bassins italiens étudiés par Solomatine *et al.* (2008), il ne l'est pas pour notre échantillon de bassins : les courbes des corrélations en fonction de Δt sont assez plates. L'indice est pertinent quand le rendement du bassin est élevé, c'est-à-dire quand le ruissellement est prédominant (Solomatine *et al.*, 2008). Dans les autres cas, le stockage de l'eau dans le sol et les écoulements souterrains brouillent le signal et les deux grandeurs ne sont pas corrélées, ce qui rend l'indice inopérant. Nous pourrions alors être tentés d'introduire une *pluie efficace* et donc une fonction de production sous-jacente, ce qui nous emmenerait vers les approches de Duband *et al.* (1993) : ce serait réintroduire de nombreuses hypothèses, réintroduire un modèle, ce que nous cherchions justement à éviter.

Une approche employée avec Pierre Javelle (Berthet *et al.*, 2008) consiste à calculer pour chaque bassin versant, les performances du modèle pour deux scénarios de pluie future bien différents (par exemple, le scénario de pluie future connue (observées a posteriori) et le scénario de pluies futures nulles). Ces performances sont égales pour les (très) petits horizons de prévision : en effet, l'eau précipitée entre t et $t + H$ n'a pas encore atteint l'exutoire. L'horizon de prévision H_{max} à partir duquel la différence des performances obtenues avec les deux scénarios devient significative nous informe sur le temps de début de réaction du bassin aux pluies... mais tel que vu par le modèle et le critère de performance choisi (Fig. 2.3).

Dans un souci de simplicité et pour éviter un trop grand nombre d'*a priori*, nous avons choisi une autre voie. Nous avons étudié l'autocorrélation des débits pour un décalage tem-

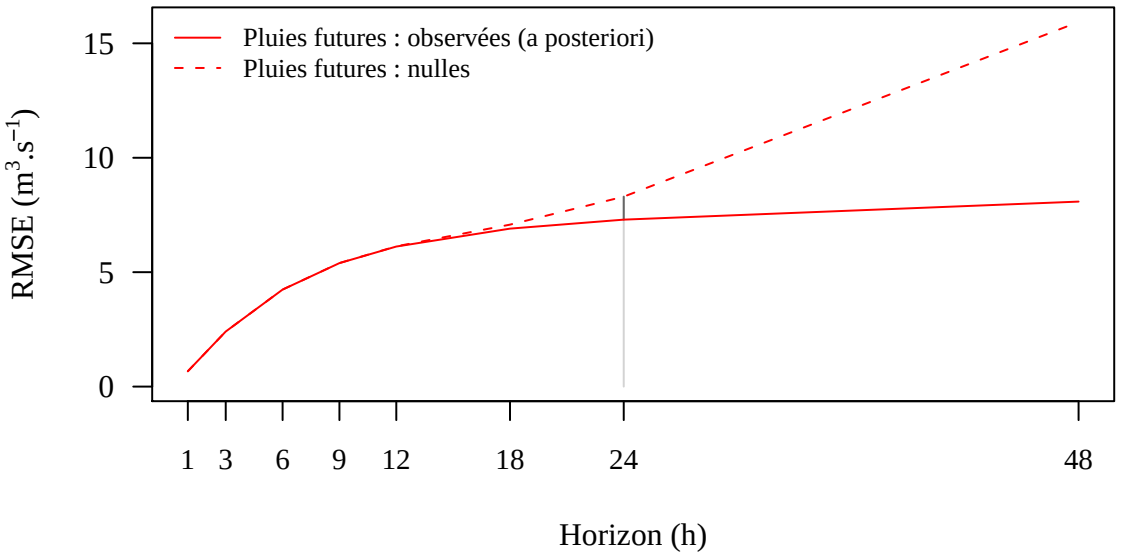


FIGURE 2.3 – Exemple d'estimation de H_{max} : les performances du modèle GR3P sont très peu dépendantes du scénario de pluies futures choisi pour des horizons inférieurs à 24 heures sur le bassin de la Grande Sauldre à Brinon sur Sauldre (K6332510, 594 km^2).

porel égal à l'horizon de prévision $\rho_H(Q)$. Une autocorrélation forte indiquera un signal lisse, donc probablement un bassin plutôt lent. Au contraire, une autocorélation faible sera la signature d'un signal saccadé, probablement lié à des événements intenses et brefs (le bassin lissant peu dans le temps les précipitations). Le passage de la caractérisation lisse / saccadé à une typologie lent / rapide n'est pas direct et probablement faux sur certains bassins. Cependant, nous avons comparé cet indice à l'indice H_{max} : il existe une assez bonne corrélation entre ces deux indices (Fig. 2.4), plus marquée pour les bassins où le modèle obtient ses meilleures performances. Il est possible de trouver quelques bassins présentant un débit lisse (forte autocorrélation) mais perçus comme rapides (H_{max} petit) : les hydrogrammes de ces bassins se caractérisent par un signal relativement lent avec un bruit rapide assez intense qui peuvent être expliquées par des manœuvres d'ouvrages (la présence d'un ouvrage se décèle dans l'exemple donné dans la figure 2.5.c) ou par un fort débit de base accompagné d'écoulements de surface rapides. Il n'y a pas par contre de bassins présentant un débit saccadé et vus comme lents par l'indice H_{max} (Fig. 2.4) ! L'indice $\rho_H(Q)$ est en outre relativement stable quand on le calcule sur deux périodes différentes. Le classement des bassins selon cet indice est également stable pour différents horizons de prévision.

Notre indice restant relativement grossier, nous ne l'emploierons pas directement. Nous définissons n classes (typiquement 4) de bassins définies par les quantiles de l'autocorrélation du débit : le bassin versant i est rangé dans la classe j si l'autocorrélation de son débit

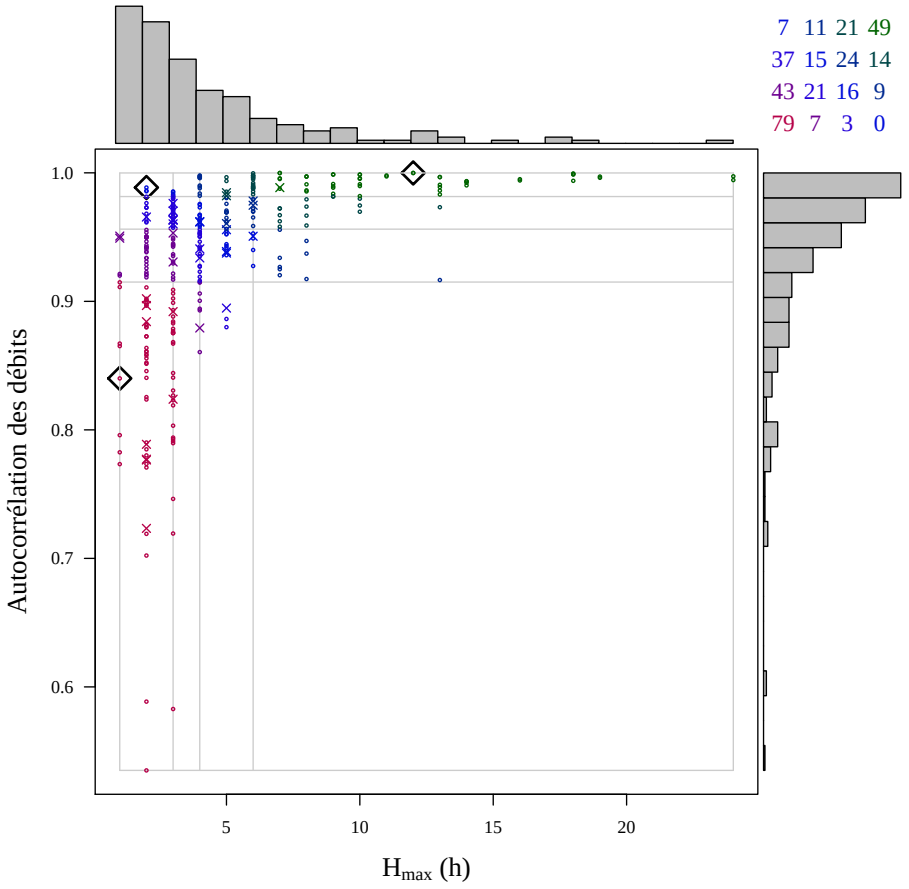


Figure 2.4 – Comparaison à l'horizon de prévision 6 heures de l'autocorrélation du débit $\rho_{H=6\ h}(Q)$ et de l'horizon de prévision à partir duquel le scénario de pluie future joue sur les performances du modèle. Les traits gris indiquent les quartiles des deux indices. Le décile des bassins sur lesquels le modèle obtient ses meilleures performances est représenté par des croix, les autres bassins par des petits points : la corrélation est nettement meilleure pour les croix, ce qui nous pousse à penser que l'indice H_{max} n'est valable que lorsque le modèle est performant. Les losanges indiquent les bassins dont les hydrogrammes sont représentés à la figure 2.5.

$\rho_H^{(i)}(Q)$ vérifie :

$$\rho_H^{(i)}(Q) \in \left[q_{\frac{j-1}{n}}(\rho_H(Q)), q_{\frac{j}{n}}(\rho_H(Q))\right]$$

2.3.4 Tester les capacités d'extrapolation du modèle

Les services opérationnels ont pour mission de prévoir des crues de n'importe quelle ampleur : connue ou encore jamais vécue. Malheureusement, les modèles ne sont pas tous conçus pour prévoir ce qui n'a jamais (encore) été observé. Les modèles purement empiriques comme les méthodes des plus proches voisins, les régressions diverses jusqu'aux

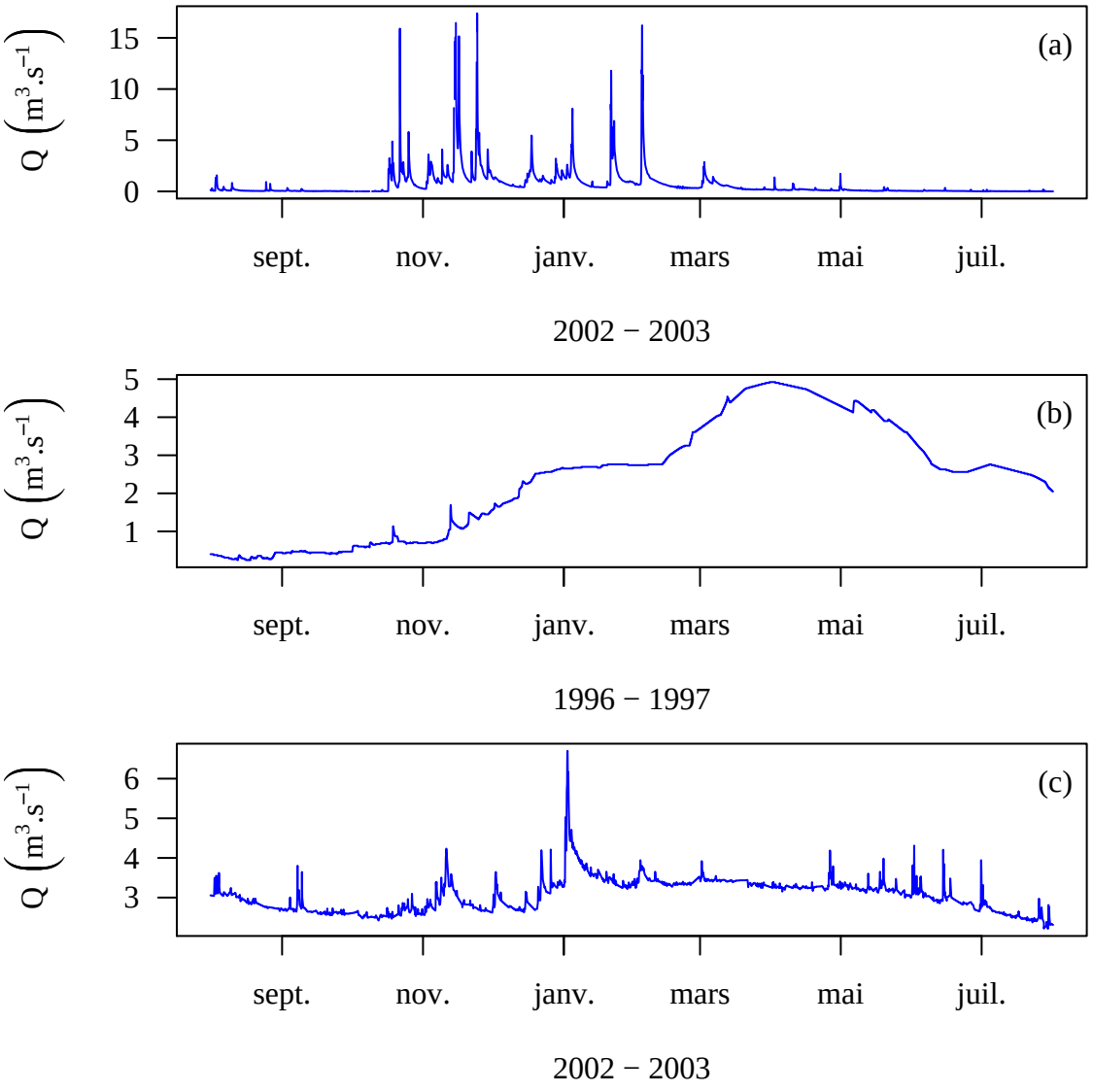


FIGURE 2.5 – Exemple d'hydrogrammes enregistrés sur des bassins réagissant différemment : (a) L'Aire à Saint-Julien-en-Genevois (40 km^2, V0245610), le bassin présentant (à la fois) de très faibles valeurs de H_{max} et d'autocorrélation des débits ; (b) La Suippe à Orainville (802 km^2, H6313020), un bassin présentant à la fois de très hautes valeurs de H_{max} et d'autocorrélation des débits : il s'agit d'une rivière de la Champagne crayeuse en forte interaction avec la nappe ; (c) Le Cailly à Notre-Dame de Bondeville (210 km^2, H9913020) : l'un des bassins présentant une valeur de H_{max} faible et une auto-corrélation des débits élevée : cette rivière est également alimentée par la nappe (ici la nappe de la Craie) mais réagit vivement aux précipitations.

réseaux de neurones ne sont pas construits pour traiter des informations très différentes de celles avec lesquelles ils ont été calés. Au contraire, les modèles purement conceptuels et les modèles réductionnistes sont basés sur le principe de la conservation de la masse et de l'énergie : on peut espérer, dans une certaine mesure, qu'ils puissent traiter des précipitations d'une intensité nouvelle.

Quasiment tous les modèles demandent une étape de calage qui permet de déterminer les valeurs des paramètres les plus *utiles*[12] (au sens de la fonction-objectif employée) pour rendre compte des observations durant une période finie (les observations sont donc limitées à une gamme bornée d'intensité). Si le modèle est bon, si les observations sont de bonne qualité, si la méthode de calage est pertinente... alors ces paramètres calés peuvent avoir un sens. Autrement, on risque de retomber dans les travers des modèles empiriques dépeints par Kirchner (2006) : *Such models are often good mathematical marionettes ; they often can dance to the tune of the calibration data.*

Sans aller jusqu'au *differential split-sample test* complet de Klemeš (1986), nous nous proposons d'analyser les différences de performance (en contrôle) du modèle sur les deux périodes P_1 et P_2 selon que le contrôle comporte des événements de crue d'une intensité non présentée lors du calage. Pour cela, nous utiliserons un indice d'extrapolation défini au chapitre suivant.

2.4 Au-delà de la recherche de l'amélioration des performances

Le bassin versant et les phénomènes qui y ont lieu ne constituent pas notre principal objet d'étude. Ce dernier s'intéresse aux relations entre la représentation du bassin versant faite par le modèle de simulation et les béquilles que nous lui fournissons (i.e., les procédures de mise-à-jour). L'objectif étant de pouvoir fournir les prévisions les plus utiles possible, cet objet englobe également les conditions d'emploi (futures) du modèle. En effet, définir l'utilité du modèle dépend fortement de sa praticité et de ses pré-requis pour fonctionner.

Une des approches permettant cette compréhension compare les dynamiques temporelles du fonctionnement du bassin versant et des assimilation de données. On cherchera donc des *temps caractéristiques* des méthodes d'assimilation des données qui devront être comparées aux horizons de prévision.

2.5 Présentation graphique des résultats

Des représentations graphiques très différentes seront employées pour illustrer nos comparaisons de modèles.

12. ... et non pas les plus *réalistes* pour autant qu'on puisse rattacher les paramètres à une quantité physique caractérisant le bassin.

Nous représenterons d'abord les comparaisons des performances « globales » de nos modèles. Quand une comparaison se limitera à deux alternatives, nous utiliserons principalement les nuages de points. S'il s'agit de comparer plus de variantes, nous emploierons des *boîtes à moustaches* pour rendre compte des distributions des performances de nos modèles : les minima et maxima sont indiqués par les ronds, les moustaches représentent les quantiles 5 et 95 %, les boîtes rendent compte des quantiles 25, 50 (médiane) et 75 %, enfin les croix donnent les moyennes (Fig. 2.6).

Nous utiliserons également des représentations plus événementielles. Il s'agira alors surtout d'hydrogrammes accompagnés des chevelus de prévision (cf. Fig. 3.1, page 40).

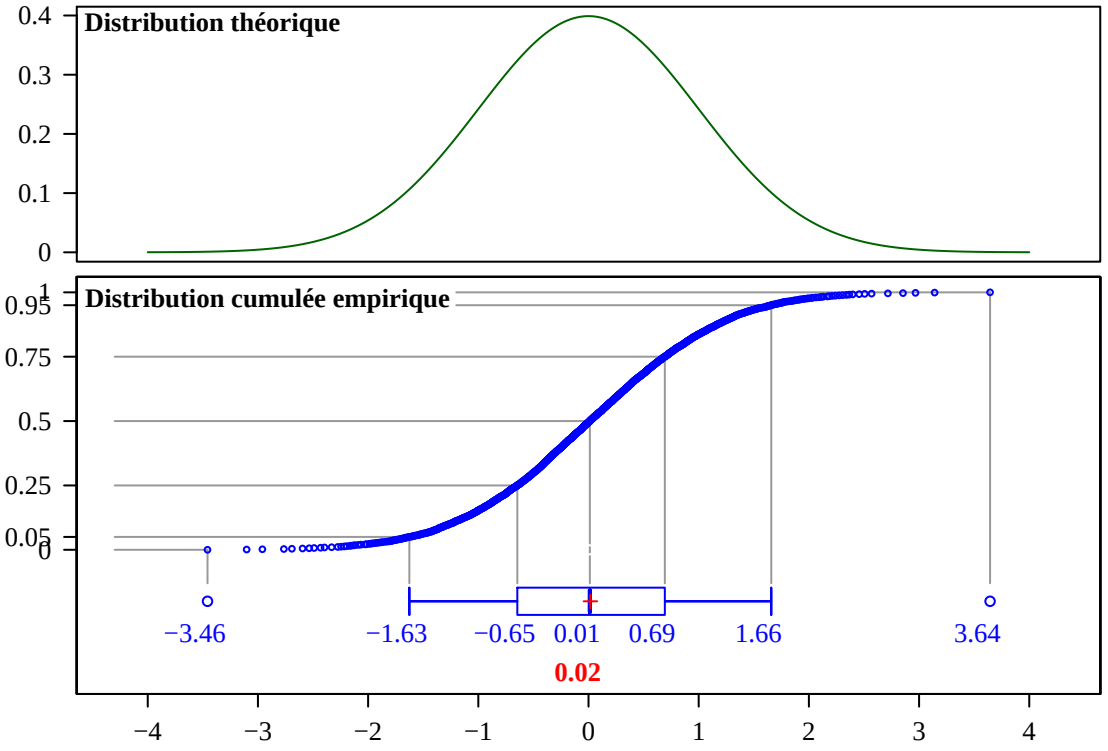


FIGURE 2.6 – Lecture d'une boîte à moustaches. Nous avons tiré 10 000 points dans une loi normale (en haut). La distribution cumulée empirique est donnée en bas, accompagnée de la boîte à moustaches résumant la distribution. Les segments gris indiquent les correspondances (de quantiles). La croix rouge représente la moyenne (dont la valeur est indiquée en rouge également).

3
Chapitre

Critères de performance

Homer – People can come up with statistics to prove anything, Kent. Forfty percent of all people know that.

John Swartzwelder, *Homer the Vigilante*,
The Simpsons (Saison 5, épisode 11)

Objectifs

- ▷ Proposer plusieurs critères permettant d'évaluer les performances des modèles hydrologiques de prévision.
- ▷ Évaluer le raffinement nécessaire de la définition de ces critères.

Sommaire

3.1 Faut-il employer la fonction de vraisemblance ? 39
3.2 Plusieurs familles de critère 40
3.2.1 Critères visuels 40
3.2.2 Des critères quadratiques 41
3.2.3 Critères liés à des tables de contingence 43
3.2.4 Critères ciblés, exemple des critères de retard 44
3.3 Splendeurs et misères des critères quadratiques 46
3.4 Degré de signification de la différence de valeurs d'un critère . 48
3.5 Des critères pour juger des améliorations du modèle et pour évaluer le modèle 50
3.6 Est-il nécessaire d'employer une multitude de critères ? 50
3.7 Bilan 52

Nous avons vu dans l'introduction que dès 1982, Linsley déplorait l'accroissement pléthorique du nombre de modèles qui a pour effet de limiter les possibilités de comparaison des résultats des expériences de modélisation. On peut malheureusement faire un constat similaire au sujet des critères de performance : les hydrologues choisissent très souvent des critères qui leur sont propres (ou raffinent des critères utilisés largement par la communauté), ce qui entraîne la même conséquence. Dawson *et al.* (2007) proposent un ensemble varié de critères pour l'évaluation de modèles hydrologiques.

Les critères d'évaluation doivent permettre de répondre aux deux questions formulées précédemment :

1. Le modèle est-il valable sur un plan scientifique ?
2. Qu'attendons-nous de la prévision ? En quoi nous est-elle *utile* ?

Là encore, ces deux familles de questions ne sont pas « orthogonales ». Néanmoins, elles incitent à examiner les résultats de façons assez différentes.

Comparer deux hydrogrammes est un exercice complexe. Il n'existe malheureusement pas de critère qui permette de rendre compte de toutes les caractéristiques d'une prévision et qui soit pertinent dans toutes les situations et pour tous les bassins. Le choix est d'abord subjectif et dépend de la question posée au prévisionniste : seules certaines caractéristiques de l'hydrogramme peuvent être importantes pour répondre à cette question. Aussi, nous travaillerons avec plusieurs critères choisis pour rendre compte de ce qui se passe sur les hautes eaux.

Dans ce chapitre, nous choisissons un ensemble de critères assez classiques pour évaluer nos modèles et discutons l'influence de ces choix en cherchant à répondre à la question : qu'auraient été nos résultats si nous avions choisi telle ou telle version raffinée de nos critères ?

3.1 Faut-il employer la fonction de vraisemblance ?

Utiliser la fonction de vraisemblance peut constituer le premier choix pour estimer la qualité de nos modèles. La fonction de vraisemblance est la fonction qui associe au modèle (éventuellement à un jeu de paramètres) la probabilité des observations conditionnellement à ce modèle (respectivement, ces paramètres). En hydrologie et plus particulièrement dans le cadre de ce travail, cet outil présente plusieurs défauts.

La vraisemblance dérive d'une théorie bien construite : pour pouvoir être interprétée dans ce cadre, il est nécessaire d'expliciter un modèle d'erreur et de le vérifier. Cette étape de validation est souvent ardue en hydrologie où la structure des erreurs est assez complexe (e.g. Beven, 2008a). Aussi, les critères du type « moindres carrés » sont très abusivement interprétés en terme de maximum de vraisemblance. Des vraisemblances formelles plus

complexes (et *a priori* plus adaptées aux problèmes hydrologiques) ont été proposées et là encore, leurs tests ont souvent montré que les hypothèses sous-jacentes n'étaient pas toutes toujours valides (Yang *et al.*, 2007). Ainsi ces outils ne sont pas forcément plus riches en enseignements que des critères ne se basant pas sur une théorie mathématique complexe.

En outre, la vraisemblance du modèle ne répond pas (directement) à l'objectif premier du prévisionniste. Ce qu'il vise est plutôt l'*utilité* de ses prévisions.

3.2 Plusieurs familles de critère

3.2.1 Critères visuels

Comparer visuellement l'hydrogramme observé avec les prévisions établies à différents instants et pour différents horizons est sans doute la meilleure méthode pour étudier un cas donné (Fig. 3.1). Outre son aspect chronophage (spécialement sur un grand échantillon de bassins versants), cette comparaison demande une certaine expertise et reste subjective. Si on ne peut pas s'en passer pour évaluer le modèle, se tourner vers des indices numériques plus synthétiques est nécessaire.

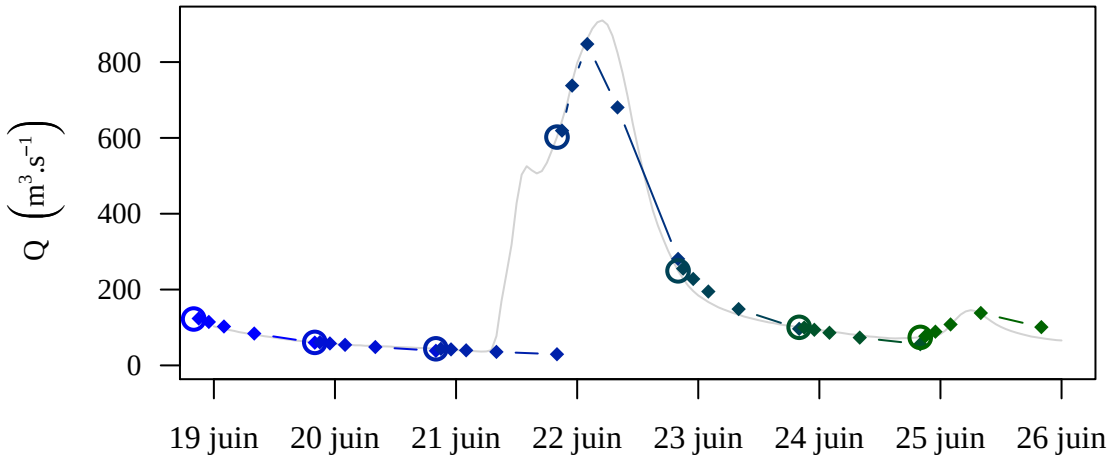


FIGURE 3.1 – Exemple d'hydrogramme permettant d'évaluer les prévisions établies par le modèle, ici sur l'Illinois River à Siloam Springs (1489 km^2) du 19 au 26 juin 2000. L'hydrogramme observé est représenté par la courbe continue grise. Des chevelus de différentes couleurs relient chaque instant de prévision (**o**) et les prévisions faites à plusieurs horizons (◆) jusqu'à 24 heures. On voit ici que la prévision (effectuée sans connaissance des pluies futures) du 21 juin rate complètement le départ de la crue mais que le modèle donne par la suite des prévisions informatives.

3.2.2 Des critères quadratiques

Historiquement, les critères quadratiques comptent parmi les critères les plus populaires dans les publications en hydrologie.

L'erreur quadratique moyenne

$$MSE(H) = \frac{1}{N}\sum_{t=1}^{N}\left(Q_{t+H} - \widehat{Q}_{t+H|t}\right)^2 \quad \in [0, +\infty[$$

où H est l'horizon de prévision, Q_{t+H} le débit qui sera observé à la date $t+H$ et $\widehat{Q}_{t+H|t}$ est la prévision établie à la date t pour la date $t+H$. L'erreur quadratique moyenne faite par le modèle sur un événement ou une plage temporelle n'est pas adimensionnelle et ne permet pas de comparer les performances d'un modèle sur deux périodes avec des caractéristiques très différentes ou sur deux bassins différents : ce critère n'est pas adapté au travail sur un large échantillon de bassins. Sa racine carrée (RMSE) est plus communément employée : elle a la dimension du débit et permet donc une interprétation plus directe.

Le critère de Nash-Sutcliffe

$$NSE(H) = 1 - \frac{\sum_t \left(Q_{t+H} - \widehat{Q}_{t+H|t}\right)^2}{\sum_t \left(Q_{t+H} - \overline{Q}\right)^2} \quad \in]-\infty, 1] \quad \text{où } \overline{Q} = \frac{1}{n}\sum_{t=1}^{n} Q_{t+H}$$

Le critère de Nash et Sutcliffe (1970) compte parmi les critères les plus utilisés par la communauté hydrologique. Il s'agit d'une normalisation du précédent. Adimensionnel, il peut être employé pour comparer les performances de plusieurs modèles sur un groupe de bassins différents [1]. Il peut se comprendre comme une comparaison du modèle à un modèle de référence (Seibert, 2001) « naïf » qui donnerait pour prévision le débit moyen. Il est de peu d'utilité en prévision car ce modèle naïf l'est un peu trop et les performances de tout modèle utilisant l'information contenue dans le dernier débit dépassent aisément 0,95 aux horizons les plus courts [2].

Critère de persistance

$$PI(H) = 1 - \frac{\sum_t \left(Q_{t+H} - \widehat{Q}_{t+H|t}\right)^2}{\sum_t \left(Q_{t+H} - Q_t\right)^2} \quad \in]-\infty, 1]$$

1. Le modèle de référence peut être plus ou moins bien adapté aux différents bassins : ce type de critères donne toujours une évaluation relative des performances.

2. Ce qui peut donner à certains la fausse impression d'une excellente performance en comparaison de l'intervalle des valeurs possibles $]-\infty, 1]$.

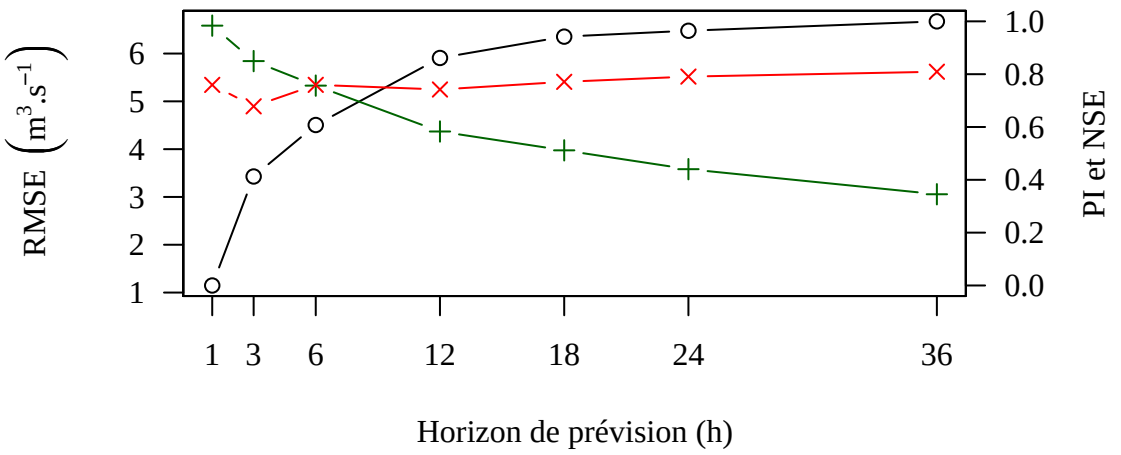


FIGURE 3.2 – Évolution des différents critères quadratiques avec l'horizon de prévision sur le bassin du Guiers Mort à Saint-Laurent-du-Pont (89 km^2, V1504010). La RMSE (ronds noirs) augmente avec l'horizon de prévision ; le critère de Nash-Sutcliffe (+ verts) décroît en conséquence. La valeur du critère de persistance (× rouges) dépend de la comparaison de deux modèles dont les erreurs sont toutes deux fonctions de l'horizon de prévision, ce qui rend son interprétation en fonction de l'horizon de prévision délicate.

Le critère de persistance (Kitanidis et Bras, 1980b) remédie au défaut du précédent en utilisant un modèle de référence, dit modèle de persistance, qui dispose des mêmes informations sur le débit que le modèle testé : ce modèle naïf donne comme prévision, le dernier débit observé.

Les critères de persistance et de Nash-Sutcliffe sont liés par l'autocorrélation des débits $\rho_H(Q)$ (démonstration donnée en annexe D.1, page 351) :

$$\frac{1 - NSE(H)}{1 - PI(H)} = 2 \cdot (1 - \rho_H(Q))$$

Si le modèle de persistance est plus pertinent en prévision que celui du critère de Nash-Sutcliffe, il présente par contre quelques difficultés d'emploi : ses performances dépendent de l'horizon de prévision. En conséquence, la comparaison des performances du modèle testé à celles du modèle naïf, effectuée par le critère de persistance, dépend également de l'horizon de prévision : ce critère ne permet pas de comparer les performances à deux horizons de prévision différents, contrairement au critère de Nash-Sutcliffe (Fig. 3.2).

Au chapitre précédent, nous avions proposé une caractérisation de la réactivité des bassins versants en fonction de l'autocorrélation des débits (page 31). Son but est de discuter les performances des modèles ou de techniques de mise-à-jour sur différents bassins en fonction de la réactivité de ces derniers. Or, le critère de persistance gomme partiellement les effets de la réactivité des bassins versants sur l'évaluation des performances des modèles. En effet, les performances du modèle naïf de ce critère dépendent elles-mêmes de la réactivité : à un

horizon de prévision donné, les performances de ce modèle seront bien supérieures sur un bassin lent que sur un bassin plus rapide. Il est donc possible qu'un modèle obtienne des valeurs du critère de persistance similaires sur deux groupes de bassins, les premiers caractérisés comme « lents », les seconds comme plus « réactifs », même si l'intuition s'attendait à ce que le modèle testé soit plus performant sur les bassins d'une classe particulière de réactivité.

Critère « maison »

$$C_{2M,P} = \frac{PI}{2 - PI} \quad \in]-1, 1]$$

Ce critère est le cousin en prévision du critère C_{2M} proposé par Mathevet *et al.* (2006) pour la simulation. Il s'agit d'un homéomorphe du critère de persistance qui le rend borné : en ce sens, il s'agit du même critère. Ce critère est d'une lecture moins directe que ceux de Nash-Sutcliffe et de persistance (qui indiquent que le modèle testé présente une erreur quadratique moyenne de x % plus petite que celle du modèle naïf). Nous en préférons néanmoins l'usage car il évite des valeurs très négatives sur le petit nombre de bassins où le modèle testé est nettement moins bon que le modèle de persistance. Ce critère limite ainsi une influence lourde de quelques valeurs sur le calcul de statistiques comme la moyenne.

En outre, il donne des valeurs des performances inférieures à celles du critère de persistance et génère donc moins de confiance indue dans le modèle (des valeurs de 0,9 sur une gamme allant de $-\infty$ à 1 pouvant entraîner un optimisme excessif).

3.2.3 Critères liés à des tables de contingence

Une famille très riche de critères peut être développée dans le cadre de prévisions binaires (ou par extension, de prévisions par classes discrètes). La prévision porte alors sur la réalisation d'un événement : quatre cas de figure sont possibles selon que l'événement a été ou non observé et selon qu'il avait été ou non prévu. Ces quatre situations sont classiquement dénombrées dans la *table de contingence* :

	Réalisé	Non réalisé
Prévu	Alerte valable (AV)	Fausse alerte (FA)
Non prévu	Alerte manquée (AM)	Non alerte valable (NAV)

Cette approche intéresse particulièrement le prévisionniste qui doit fournir certaines réponses binaires : y aura-t-il ou non inondation ? Le seuil d'alerte sera-t-il dépassé ? Faut-il

prévenir les services préfectoraux ?

À partir des tables de contingence qui comptabilisent le nombre d'occurrences des quatre cas de figure mentionnés ci-dessus, de nombreux critères peuvent être définis (Bontron, 2004, chapitre 3). Ces critères sont largement employés par les prévisionnistes (par exemple au Centre européen de Prévisions des crues : Ramos *et al.*, 2007) et en météorologie.

L'emploi de ces critères soulève pour nous une difficulté pratique : il nous faut définir l'événement (binaire) qui nous intéresse. Or nous avons vu que la qualité d'une prévision est une chose complexe et difficilement « objectivable » : la qualité dépend naturellement de l'usage que nous faisons de la prévision. Aussi, la définition de l'événement à prévoir est fortement conventionnelle et peut refermer des hypothèses *ad hoc* permettant de présenter de bons résultats. Cela peut concerner le choix du seuil définissant un événement de crue. Pour éviter au maximum ces travers, nous éviterons toute définition complexe et ne traiterons que la question du dépassement de seuil défini par un quantile de débit.

Nous utiliserons trois critères classiques :
– la probabilité de détection : $POD = \dfrac{AV}{AV + AM}$
– l'indice de succès critique (également connu sous l'appellation de *threat score*) :

$$CSI = \frac{AV}{AV + AM + FA}$$

– le taux de fausse alertes : $FAR = \dfrac{FA}{AV + FA}$

Même si ces trois indices décrivent différents aspects des performances de prévision, ils ne sont pas indépendants (Schaefer, 1990).

Nous emploierons en outre une légère modification de l'indice de succès critique,

$$CSI'(\nu) = \frac{AV}{AV + AM + \nu FA} \quad \text{avec } 0 \leq \nu \leq 1$$

où ν permet de donner une importance moins forte aux fausses alertes (en comparaison aux alertes manquées). Cet indice est intermédiaire entre l'indice de succès critique et la probabilité de détection. Certains le préfèrent car il donne un poids plus faible aux fausses alertes, considérées comme moins pénalisantes que les alertes manquées.

3.2.4 Critères ciblés, exemple des critères de retard

Les critères quadratiques évaluent globalement la distance qui sépare la courbe des observations de celles des prévisions. Dans une optique plus opérationnelle, on peut décider de s'intéresser explicitement au comportement du modèle à certaines dates seulement (pour nous, sur les périodes de crues). Dans la pratique, beaucoup d'arbitraire ou d'hypothèses

ad hoc peuvent entrer dans la définition de ce qui est une crue et ce qui ne l'est pas. Sur un large échantillon de bassins versants, on se contentera d'étudier les périodes de *hautes eaux*, c'est-à-dire celles dont les débits sont supérieurs à un quantile donné $q_x(Q)$. Quand la définition des critères le permet, on peut également prendre en compte les dates d'événements « potentiels », c'est-à-dire les dates où de fortes pluies n'ont pas engendré de crue.

Un exemple classique de critère événementiel est l'erreur sur les débits de pointe mais il s'agit encore d'une erreur exprimée en termes de débits donc qui n'apporte pas foncièrement de nouveauté par rapport aux critères précédents. Au contraire, les critères événementiels temporels portent *a priori* un regard différent sur les performances du modèle.

Retard de prévision des pointes de crue

Le critère temporel le plus simple est le retard[3] de prévision sur la pointe de crue : si on note t_{max}, l'instant de l'observation du débit maximal sur la crue et $\hat{t}^H_{max}$, l'instant pour lequel on prévoit la pointe de crue, ce critère s'écrit pour la crue i :

$$R_1^{(i)} = \hat{t}^H_{max} - t_{max}$$

Retard de prévision des plus hautes eaux

Dans certains cas, l'instant du débit maximal est imprécis et importe peu aux prévisionnistes. Il peut y avoir par exemple un plateau de débits très proches du débit de pointe. Ce qui nous intéresse alors est plutôt le retard de l'annonce de la période de débit maximal. Il suffit de modifier le critère précédent en remplaçant t_{max} par $t_{Q=\alpha Q_{max}(i)}$ où α est un nombre légèrement inférieur à 1 (par exemple 0,8 ou 0,9), et $\hat{t}^H_{max}$ par $t_{\widehat{Q}=\alpha \widehat{Q}^H_{max}(i)}$. Ce critère sera noté $R^{(i)}_{1bis}(\alpha)$.

Retard moyen de prévision du dépassement de seuil

Les critères de retard présentés jusqu'ici ne s'intéressent qu'à la pointe de crue. Cette information est très lisible et aisément communicable mais elle n'épuise pas forcément les demandes faites aux services opérationnels. Ces demandes concernent souvent des dépassements de seuils à la montée (« *Quand le cours d'eau va-t-il sortir de son lit ? Quand va-t-il atteindre le collège ?* » ...). Sur un échantillon d'un millier de bassins versants, il est bien sûr très difficile d'évaluer le modèle sur des seuils pertinents localement. On se contentera d'évaluer le retard moyen au dépassement de seuil à partir d'un débit minimum Q_m (choisi de façon arbitraire comme étant un quantile des débits observés) lors d'un événement E.

Nous calculons lors de la montée de l'événement E, le retard $\Delta t\,(q) = \widehat{t}(q) - t(q)$ pour tout débit q supérieur à Q_m, c'est-à-dire la différence entre $t(q)$, l'instant où le débit q

3. Le plus souvent, la prévision de la montée et de la pointe de crue est en retard sur les observations. Cependant dans les cas contraires, on aura un retard négatif.

est observé pour la première fois et $\widehat{t}(q)$ l'instant pour lequel le débit q est prévu pour la première fois (Fig. 3.3.a). Le critère de retard sur l'événement est alors la moyenne sur tous les débits compris entre Q_m et $Q_M = min\left(\underset{E}{max}(Q), \underset{E}{max}\left(\hat{Q}\right)\right)$.

En pratique, ce critère s'écrit

$$R_2 = \frac{1}{Q_M - Q_m} \cdot \int_{Q_m}^{Q_M} t^{Prev}(Q) - t^{Obs}(Q) dQ = \frac{1}{Q_M - Q_m} \cdot \int_{t_0}^{t_1} \widetilde{Q}^{Obs}(t) - \widetilde{Q}^{Prev}(t) dt$$

avec $\begin{cases} \widetilde{Q}_i^{Obs} = min\left(Q_M, max\left(Q_m, max\left(Q_i^{Obs}, \widetilde{Q}_{i-1}^{Obs}\right)\right)\right) \\ \widetilde{Q}_i^{Prev} = min\left(Q_M, max\left(Q_m, max\left(Q_i^{Prev}, \widetilde{Q}_{i-1}^{Prev}\right)\right)\right) \end{cases}$

Cette réécriture du critère illustre la non indépendance des critères sur les différences de débit et ceux caractérisant les erreurs temporelles : comme le souligne Madsen (2000), les critères de performance classiques calculés sur les débits (comme le critère de persistance) prennent en compte la qualité temporelle du modèle. En fait, il parait illusoire de définir une erreur de phase indépendamment de l'erreur d'amplitude.

Critère d'avance

Jusqu'ici, nous avons cherché à définir le retard de l'hydrogramme prévu sur l'hydrogramme observé de la manière la plus adaptée à la prévision. Nous pouvons renverser le point de vue en cherchant à évaluer l'avance (moyenne) que peut avoir le prévisionniste sur un événement.

Pour une classe d'événements (définie par exemple par le dépassement d'un seuil), nous évaluons la fraction de bonne alerte en fonction de l'horizon de prévision H. Il s'agit tout simplement de $1 - FAR(H)$. Puis nous examinons la décroissance de cet indice en fonction de H pour déterminer la valeur maximale H_M de l'horizon pour laquelle l'indice reste supérieur à un seuil d'acceptabilité x (par exemple 90 %). On peut alors affirmer qu'en moyenne, on peut prévoir les événements de cette classe avec un horizon de prévision au moins égal à H_M dans x % des cas.

3.3 Splendeurs et misères des critères quadratiques

Savoir si les critères quadratiques sont adaptés à l'évaluation de modèles hydrologiques est une question régulièrement débattue. Cependant leur usage reste très commun.

Les modèles peuvent produire une prévision à chaque pas de temps. Est-il nécessaire de tenir compte de toutes les erreurs pour évaluer le modèle ? Notamment de celles faites en dehors d'événements de crue ? À l'inverse, certains auteurs critiquent l'importance trop forte que les critères quadratiques donnent aux erreurs faites sur les plus forts événements. Dans le cadre de l'atelier de la *Cour des miracles de l'hydrologie* tenu à Paris du 18 au 20

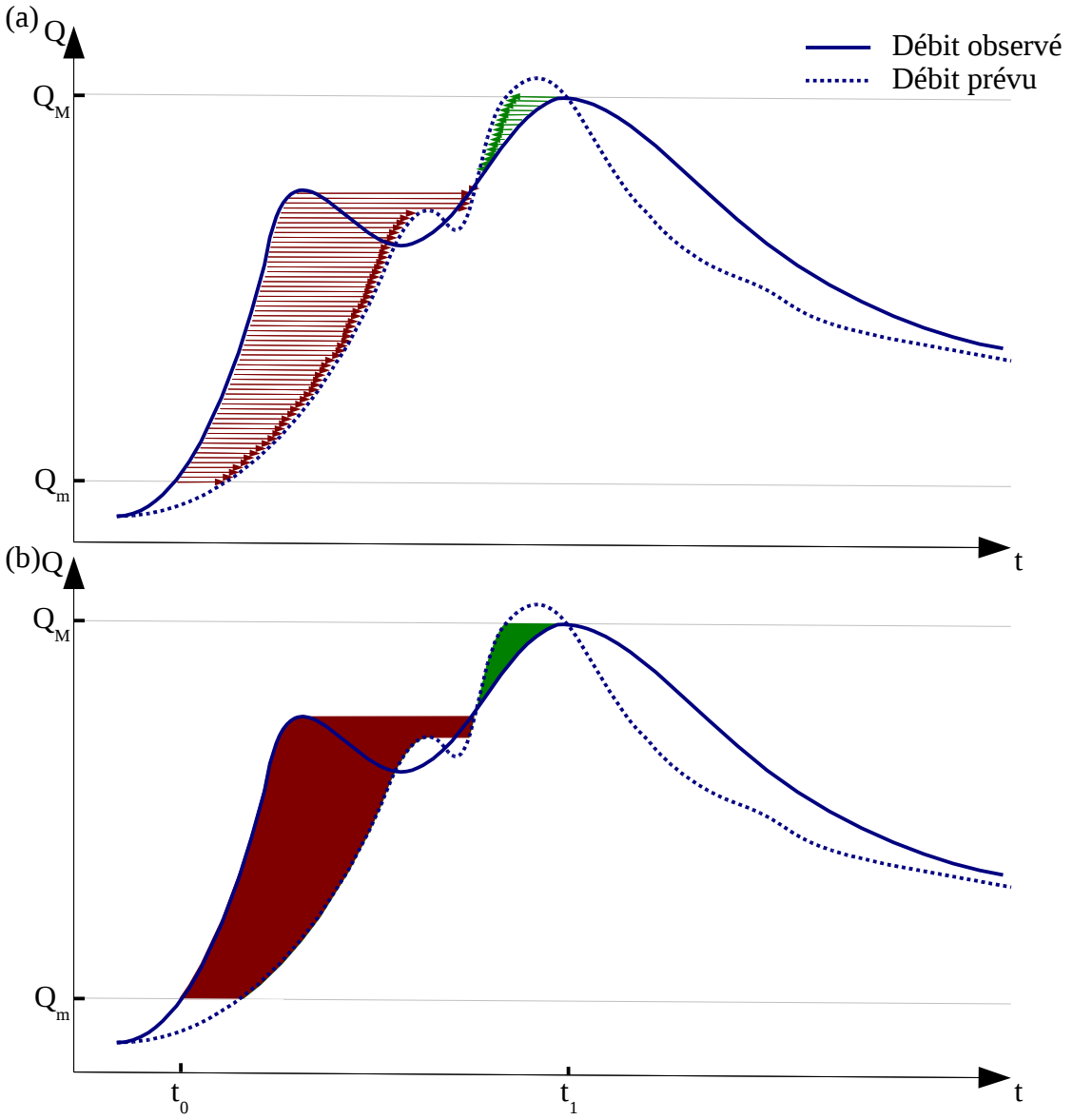


Figure 3.3 – (a) Le critère de retard moyen au dépassement de seuil consiste à prendre la moyenne des différences temporelles entre l'observation du dépassement de tout seuil Q et l'instant où ce dépassement est prévu, pour tout débit Q supérieur à un débit minimal Q_m. (b) Mathématiquement, cela revient à calculer l'aire entre les deux courbes (observations et prévisions) pour les débits compris entre Q_m et Q_M (une intégrale de Lebesgue en quelque sorte) et à la diviser par la différence $Q_M - Q_m$.

juin 2008, nous avons présenté un travail traitant de l'évaluation par un critère quadratique des modèles de prévision des crues.

Nous cherchions à déterminer si les critères quadratiques évaluent vraiment de façon continue les modèles et dans le cas contraire, quels sont les points effectivement pris en compte par un critère quadratique. En prévision, un très petit nombre de pas de temps où l'erreur du modèle est importante contribue pour l'essentiel de l'erreur quadratique totale (MSE). Il s'agit des pas de temps pour lesquels il y a la plus grosse variation de débit, donc ceux pour lesquels le modèle est susceptible d'avoir les plus grosses difficultés. Ce sont également les pas de temps qui intéressent le plus les prévisionnistes. Dans notre cas d'étude, les mises-à-jour du modèle rendent le critère plus événementiel qu'en simulation : un plus petit nombre de pas de temps porte l'essentiel de l'erreur totale du modèle.

Une seconde question, plus théorique, a été abordée. Quel que soit le critère, l'évaluation du modèle est dépendante dans un premier temps de la chronique sur laquelle cette évaluation est faite. Certains critères sont définis de façon à ce qu'ils convergent quand la chronique s'allonge : l'évaluation devient « absolue » (indépendante des données). Si notre critère prend peu de points en compte et si ces points sont définis relativement aux événements présents dans la chronique, on peut se demander s'il doit converger si on utilise une chronique infiniment longue. Si le critère quadratique testé converge quand la quantité de données d'évaluation tend vers l'infini, la vitesse de convergence est très lente : le nombre de pas de temps d'évaluation nécessaires pour assurer une bonne estimation de l'espérance de l'erreur quadratique moyenne est très élevé. À partir de là, le critère de persistance n'a qu'un sens très relatif pour les séries temporelles dont nous disposons : il ne donne qu'une évaluation relative du niveau d'erreur qu'on peut attendre du modèle sur un bassin donné. Ceci nous encourage encore plus à travailler sur de nombreux bassins.

Ces deux questions ont fait l'objet d'articles soumis pour le numéro spécial du Journal des Sciences hydrologiques devant rassembler les différentes contributions à cet atelier. Ces articles sont proposés dans les annexes E et F (pages 363 et 395) dans leurs versions soumises au printemps 2009.

3.4 Degré de signification de la différence de valeurs d'un critère

Notre travail consiste à comparer des modèles (le plus souvent très proches les uns des autres) sur un grand nombre de bassins versants. Dans la plupart des cas, la variante A sera meilleure que la variante B sur un sous-groupe de bassins, le contraire se produisant sur un autre sous-groupe tandis que les deux variantes obtiendront des performances similaires sur les bassins restants.

Examiner les distributions peut conduire à des choix subjectifs. On peut notamment se

demander à partir de quelle différence de valeur pour un critère de performance donné, une variante est significativement meilleure qu'une autre (Mathevet, 2005). Pour répondre à cette question, nous éviterons les tests statistiques demandant trop d'hypothèses sur la nature des distributions comme le test de Fisher : nous préférerons un test non paramétrique comme le test de Wilcoxon et Mann-Whitney (utilisé par exemple par Andréassian *et al.*, 2003, pour détecter des changements graduels de comportements hydrologiques). Nous allons réaliser un calcul « de coin de table » pour fixer des ordres de grandeur.

Un des points importants dans le test de Wilcoxon et Mann-Whitney est la forme des distributions comparées. Partons de la série réellement observée des performances de notre modèle de départ GR3P (fonctionnant au pas de temps horaire). Ajoutons-y un bruit sans couleur particulière (un bruit gaussien de moyenne μ) et observons les variations de la probabilité que la série originale et la série bruitée soient issues de la même distribution selon le test (Tab. 3.1).

μ	1040 bassins	178 bassins
$\mathbf{0,5\ 10^{-2}}$	0,36	0,71
$\mathbf{1,0\ 10^{-2}}$	0,09	0,49
$\mathbf{1,5\ 10^{-2}}$	0,01	0,31
$\mathbf{2,0\ 10^{-2}}$	10^{-3}	0,19
$\mathbf{2,5\ 10^{-2}}$	$6\ 10^{-5}$	0,10
$\mathbf{3,0\ 10^{-2}}$	10^{-6}	0,05

Table 3.1 – *p-values*[4] pour l'hypothèse H_0 : la série de référence et la série bruitée par une loi normale de moyenne μ et d'écart-type 10^{-2}, sont issues de la même distribution.

Pour un échantillon de 178 bassins[5], on considérera que deux variantes donnent des performances significativement différentes dès que les distributions seront éloignées en moyenne de 3 points (0,03) sur le critère de persistance. On pourrait se contenter de 1 à 1,5 points sur l'échantillon total.

Examiner la distribution complète des différences de performance reste cependant nécessaire. Nous vérifierons notamment pour quels bassins une variante apporte des améliorations. S'agit-il des bassins sur lesquels le modèle est déjà performant ? Ce gain général s'accompagne-t-il de dégradations pour les bassins pour lesquels le modèle est faible ? C'est-à-dire : ce gain de performance est-il obtenu au prix d'une perte en robustesse du modèle (Fig. 3.4) ?

4. Probabilité d'obtenir une valeur de la statistique du test au moins aussi grande que celle calculée avec les données. La *p-value* se compare au risque de première espèce α_{seuil} choisi (probabilité de rejeter l'hypothèse H_0 alors qu'elle est valable), par exemple $\alpha_{seuil} = 0,05$.

5. L'échantillon que nous définirons au prochain chapitre compte 178 bassin versants.

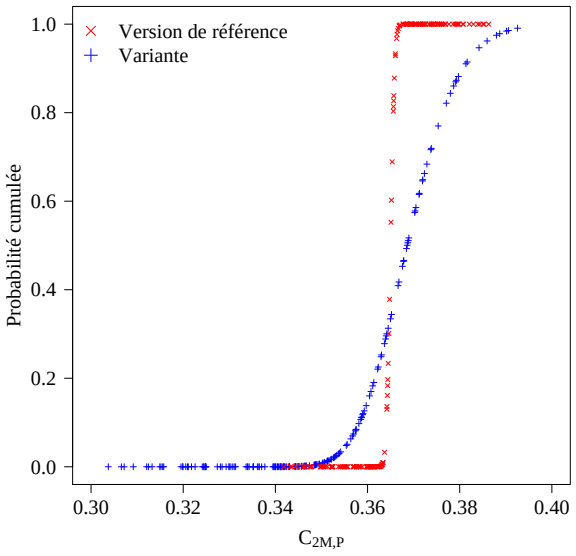


FIGURE 3.4 – Ces distributions (fictives) de performance de deux variantes sont telles que la variante améliore la médiane des performances au prix d'une perte en robustesse (queue basse de la distribution).

3.5 Des critères pour juger des améliorations du modèle et pour évaluer le modèle

L'amélioration empirique du modèle est une démarche itérative où à chaque étape, les performances du modèle *en l'état* sont comparées à celles d'une ou plusieurs variantes. Cette comparaison ne permet d'utiliser qu'un nombre restreint de critères. En effet, utiliser plusieurs critères conduirait facilement à des fronts de Pareto où une version x du modèle est meilleure que la version y pour le critère A mais moins bonne pour le critère B.

Il est donc préférable de n'utiliser qu'un nombre réduit de critères lors de cette démarche et de définir un ordre de priorité parmi eux : on évaluera les performances de plusieurs variantes selon un premier critère, puis si ce dernier ne permet pas de les différencier de façon significative, on utilise un second critère... Nous travaillerons principalement avec un critère « global » : la persistance car il s'agit d'un critère décrivant l'ensemble des événements. Si cela se révéle nécessaire, nous emploierons un critère de retard (R_2).

3.6 Est-il nécessaire d'employer une multitude de critères ?

Il sera également nécessaire d'évaluer le plus complètement possible le modèle que nous proposerons à l'issue de nos travaux, afin de se faire l'idée la plus juste de ses capacités prédictives. Cela suppose d'examiner le modèle sous des angles différents, c'est-à-dire par

des critères de performance complémentaires.

Cependant, on peut multiplier à l'envie les critères en raffinant indéfiniment les critères « principaux » présentés précédemment. Outre le fait que cette démarche diminue les capacités de communication entre hydrologues, il est intéressant de savoir si cela apporte réellement de l'information nouvelle : si plusieurs critères sont employés, il est nécessaire d'évaluer les corrélations entre ces critères (Węglarczyk, 1998).

Pour répondre à cette question, nous avons multiplié nos critères de base en nous appuyant sur les nombreuses suggestions qui nous ont été faites par les services opérationnels et par d'autres équipes d'hydrologues, puis nous avons analysé la variabilité des valeurs de ces critères obtenues par notre modèle de départ à l'aide d'une analyse en composantes principales (A.C.P., voir l'annexe I.2, page 486).

L'annexe D.2 présente une étude visant à évaluer le degré de raffinement utile des critères de performance (page 351). Elle porte sur 26 critères qui peuvent être regroupés en trois grandes familles : critères de persistance, critères de retard et critères d'alerte (CSI, CSI' et POD). Nous avons raffiné chacun de ces trois critères de base pour obtenir différentes variantes qui sont *a priori* plus adaptées à l'évaluation de prévisions. Le tableau D.1 (page 352) énumère ces 26 critères.

Le principal résultat de cette étude est que les trois premiers axes principaux « expliquent » entre 69 % et 75 % de la variabilité observée (selon les horizons de prévision). Ces fractions montent de 80 à 84 % si on élimine les 7 critères avec scénario de pluies futures nulles. Sur les représentations de l'A.C.P., on observe que les critères de chaque famille sont très resserrés entre eux et éloignés des autres familles.

Si on examine la seule famille des critères de persistance (avec un même scénario de pluies futures), on observe que la première composante principale explique de 80 % (horizon de 48 heures) à 90 % (horizon de prévision d'une heure) de la variabilité. En outre, tous les critères de persistance ont la même projection sur cette composante : en d'autres termes, ces critères parlent essentiellement de la même chose, ce qui ne va pas contre l'intuition.

Ainsi, plutôt que de raffiner un critère, il paraît plus pertinent d'employer un petit nombre de critères bien différents. Autant prendre les critères les plus simples (et les plus communs au sein de la communauté hydrologique) parmi chaque famille de critères : ils seront plus facilement interprétables et permettront plus aisément les comparaisons avec d'autres études.

3.7 Bilan

Ce chapitre nous a permis de présenter les critères qui nous serviront à évaluer les performances de notre modèle. Ces indices peuvent être rangés en trois grandes familles :

1. critères quadratiques : principalement, les critères de persistance,
2. critères temporels : critères de retard et d'avance,
3. critères événementiels d'alerte.

Pour l'élaboration du modèle, il apparaît difficile de ne pas se restreindre à un petit nombre de critères hiérarchisés pour évaluer si les variantes testées apportent une amélioration par rapport au modèle initial.

L'évaluation finale du modèle peut être réalisée avec un grand nombre de critères mais ceux-ci ne sont pas indépendants les uns des autres. Une étude des corrélations entre ces critères nous suggère qu'il n'est pas forcément nécessaire d'utiliser une multitude de critères ou de les raffiner *ad libitum* mais qu'il est plutôt préférable d'employer quelques critères bien différents les uns des autres.

4

Chapitre

Un vaste échantillon de bassins versants

Nothing is so commonplace as the wish to be remarkable.

Oliver Wendell Holmes

Objectifs

- ▷ Décrire les données dont nous disposons.
- ▷ Déterminer si l'ensemble de bassins versants dont nous disposons est varié,
- ▷ si nous pouvons caractériser sa variété hydrologique,
- ▷ si l'ensemble des bassins versants est idéal,
- ▷ si nous avons besoin de l'ensemble complet pour exprimer cette variété.

Sommaire

4.1 Données disponibles sur chaque bassin 55
4.1.1 Données de précipitation 55
4.1.2 Évapotranspiration potentielle 56
4.1.3 Débit 57
4.1.4 Qualité des données validées 58
4.2 Caractériser la variété hydrologique de l'ensemble de 1024 bassins versants 58
4.2.1 Indices géographiques et topographiques 58
4.2.2 Indices décrivant les débits 60
4.2.3 Indices caractérisant les pluies 61
4.2.4 Premiers résultats statistiques 62
4.3 Définition et caractérisation de groupes de bassins versants homogènes 62
4.3.1 Analyse portant sur les indices de pluie et de débits 63
4.3.1.1 Résultats de l'analyse en composantes principales . . . 63
4.3.1.2 Résultats de la classification ascendante hiérarchique . 64
4.4 Limiter les biais : passer de l'ensemble à l'échantillon 66
4.4.1 Méthode proposée 66
4.4.2 Tests réalisés 67
4.4.3 Résultats 67
4.4.3.1 Distributions des indices sur des échantillons restreints 67
4.4.3.2 Distributions des critères de performance sur des échantillons restreints 69
4.4.3.3 Variété du critère de performance au sein des groupes . 69
4.4.4 Réduction de l'échantillon 70
4.5 Bilan et conclusions 72

Nous disposons de séries de données horaires de pluie, d'évapotranspiration potentielle et de débit sur un ensemble initial de 1 070 bassins versants, tous situés en France métropolitaine (incluant la Corse). Ces bassins correspondent aux stations hydrométriques présentant des données exploitables dans la banque de données Hydro sur la période 1995 – 2005 (Le Moine, 2008). Quarante-six bassins présentant trop de lacunes de données ont été retirés pour aboutir à un ensemble de 1 024 bassins.

Face à cette masse de données, plusieurs questions se posent à nous de prime abord. La première s'intéresse à la variété hydrologique de l'ensemble des bassins versants pour lesquels nous disposons de données : l'ensemble est-il varié ? Pouvons-nous caractériser cette diversité ?

Ensuite, cet ensemble ne constitue pas *a priori* un échantillon « idéal » mais est le résultat de l'extraction des données exploitables de la banque Hydro pour les travaux menées au sein de l'unité de recherche. L'échantillon présente-t-il un biais ? Certains types de bassins versants sont-ils très représentés quand d'autres ne le sont quasiment pas ? Il est utile de chercher à identifier les « redondances hydrologiques » au sein de cet ensemble de bassins versants pour obtenir un échantillon réduit plus équilibré. En outre, cette opération nous fera gagner par la suite un temps de calcul précieux.

Ce chapitre débute par une courte description des données dont nous disposons sur ces bassins. Le chapitre se poursuit par une présentation des indices permettant de décrire les bassins versants et leurs régimes hydroclimatiques. Ces indices sont ensuite employés pour définir et caractériser des groupes de bassins homogènes au sein des 1 024 afin de répondre aux questions soulevées ci-dessus. Enfin, les mêmes méthodes nous permettent de réduire le nombre de bassins versants à conserver pour obtenir un échantillon de bassins moins dépendant des contraintes pratiques de l'obtention des données.

Pour ne pas surcharger ce chapitre, de très nombreuses figures ont été placées en annexe G qui débute page 423.

4.1 Données disponibles sur chaque bassin

La plus grande partie des données ont été rassemblées par Le Moine (2008). Nous décrivons ici brièvement la méthodologie qu'il a employée.

4.1.1 Données de précipitation

Nous diposons des données horaires d'environ 600 pluviomètres répartis sur le territoire métropolitain. Les pluviomètres dont les données horaires présentent plus de 15 % de lacunes ont été écartés. 580 pluviomètres ont ainsi été conservés sur la France. Les précipitations neigeuses sont prises en compte à travers leur équivalent en eau.

Pour chaque poste pluviographique, les lacunes ont été complétées en s'appuyant sur les données SAFRAN horaires corrigées (e.g. Le Moigne, 2002). Soit D, l'ensemble des dates (horaires) d pour lesquelles le pluviomètre donne une information : $P_{\text{Pluviomètre},d} \geq 0$. Les cumuls des pluies mesurées par ce pluviomètre sur l'ensemble des dates D sont calculés, ainsi que les cumuls des pluies de la maille SAFRAN correspondante sur ce même ensemble :

$$P_{\text{SAFRAN},tot} = \sum_{d \in D} P_{\text{SAFRAN},d} \qquad P_{\text{Pluviomètre},tot} = \sum_{d \in D} P_{\text{Pluviomètre},d}$$

Le rapport $\zeta = \dfrac{P_{\text{Pluviomètre},tot}}{P_{\text{SAFRAN},tot}}$ permet de compléter les lacunes des données du pluviomètres sans introduire de biais par

$$P_{\text{Pluviomètre},d \notin D} = \zeta \cdot P_{\text{SAFRAN},d}$$

Les pluies de bassins sont alors estimées par une moyenne pondérée des pluies mesurées aux pluviomètres. La pondération se fait par la méthode de Thiessen : les coefficients sont les aires des surfaces définies par l'intersection du bassin-versant par les polygones de Thiessen définis par les emplacements des postes pluviométriques.

Le réseau sur lequel s'appuie l'interpolation SAFRAN étant beaucoup plus dense que celui des pluviomètres dont nous disposons, on peut espérer que le bilan en eau soit meilleur à long terme avec les pluies données par SAFRAN qu'avec celles mesurées au pluviomètre. Aussi, l'estimation de la pluie de bassin $P_{BV,d}$ applique les propriétés à long terme des données SAFRAN aux dynamiques des pluies de bassin évaluées par le réseau pluviométrique $P_{BV,\text{Pluviomètre},d}$ (Berthet, 2006) :

$$P_{BV,d} = \frac{1}{\zeta} \cdot P_{BV,\text{Pluviomètre},d}$$

4.1.2 Évapotranspiration potentielle

La donnée d'évapotranspiration potentielle n'est pas cruciale en temps de crue : elle est alors le plus souvent négligeable face aux pluies et débits en jeu. Par contre, elle reste indispensable pour le calcul des conditions initiales de l'humidité du bassin. En raison de la sensibilité limitée du modèle à cette variable à court terme, nous utilisons les moyennes interannuelles journalières des évapotranspirations potentielles estimées par la formule de Oudin *et al.* (2005a,b), que nous désagrégeons au pas de temps horaire suivant une sinusoïde.

4.1.3 Débit

Les données brutes dont nous disposons sont les débits instantanés de la banque Hydro accompagnés d'un code qualité permettant de distinguer les données validées des données douteuses ou des données dont la qualité est ignorée. Un code de chaînage indique la possibilité d'interpoler linéairement entre deux instants de mesure, ou au contraire la présence de lacunes.

Le Moine (2008) a calculé les séries de débit horaire à partir des données validées. Ce choix est dicté par la nécessité de bâtir nos modèles sur des données de bonne qualité. Toutefois, les données de débit non validées apparaissent le plus souvent sur des événements de crue pour deux raisons. En effet, certains opérateurs renseignent un code « qualité douteuse » dès qu'ils estiment la courbe de tarage imprécise. En outre, les pannes techniques des appareils de mesure ou de transmission arrivent le plus fréquemment lors d'événements de crue. Cela peut aboutir à des hydrogrammes étêtés (Fig. 4.1).

Différentes approches pour quantifier l'importance des données douteuses lors des passages de hautes eaux sont présentées dans l'annexe H (page 477). Cette étude montre que se restreindre aux données validées (sur un grand nombre de bassins) n'a qu'une influence limitée sur l'évaluation du modèle de prévision des crues.

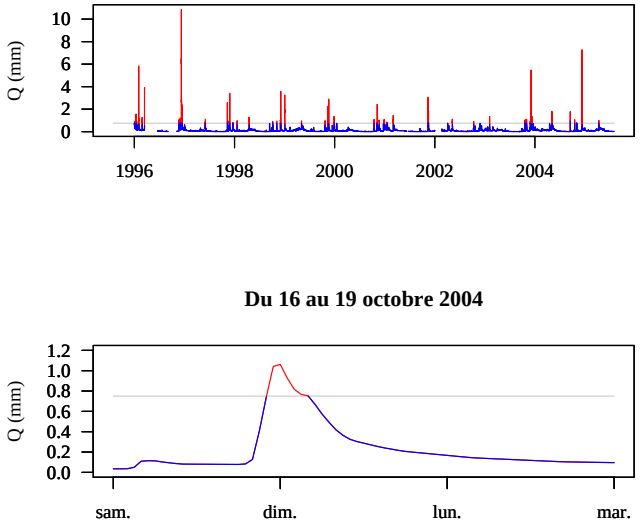


Figure 4.1 – Exemple d'hydrogramme étêté, ici le Fium-Orbo à Ghisoni (Sampolo), 115 km^2. Tous les débits instantanés au-dessus d'un seuil se situant aux alentours de 0.75 mm.h^{-1} sont codés comme douteux (points rouges). Les débits validés sont représentés en bleu.

4.1.4 Qualité des données validées

Cet échantillon est représentatif des données que l'on peut trouver sur les bases de données nationales. La qualité n'en est pas parfaite : il existe vraisemblablement des erreurs malgré le soin porté par les producteurs à les valider. Pour l'exercice de cette thèse et après les travaux de Le Moine (2008) sur ces données, nous jugeons leur qualité suffisante pour ne pas perturber nos conclusions.

Notons en outre que le temps disponible pour valider les données dans le cadre de la prévision opérationnelle est nécessairement limité. Il est illusoire de faire fonctionner le modèle avec des données « parfaites ». Pour cette raison, nous n'avons pas poussé plus loin la critique des données.

4.2 Caractériser la variété hydrologique de l'ensemble de 1024 bassins versants

Nous allons étudier l'ensemble de bassins versants dont nous disposons à travers le prisme d'indices décrivant leur topographie, puis leur régime d'écoulement et enfin leur régime pluviométrique.

4.2.1 Indices géographiques et topographiques

Une première source d'information sur les bassins versants concerne leurs géographies et leurs topographies. Nous disposons principalement de leurs localisations, de leurs superficies et de leurs altitudes (à travers leurs courbes hypsométriques(g)).

La répartition spatiale de nos bassins sur le territoire (Figure 4.2), permet de penser que nous rencontrons à travers cet échantillon la plus grande partie de la variété climatique française. On note cependant une (relativement) faible représentation des bassins versants du pourtour méditerranéen quand le Nord-Est de la France est fortement présent dans l'ensemble des bassins versants.

L'ensemble de 1 024 bassins compte principalement des petits bassins versants (Tab. 4.1, Fig. G.5, page 428). La très grande majorité de ces bassins s'étendent sur une superficie inférieure à 1 000 km^2. Ils concernent donc principalement des petits cours d'eau en amont des grandes rivières et fleuves français. On notera d'ailleurs que sur la carte 4.2, la plupart des zones d'absence de nos bassins (en blanc) correspond au tracé des fleuves.

La pente est une caractéristique locale. Il est difficile de définir la « pente » d'un bassin versant dans sa globalité. Nous avons choisi ici de nous appuyer sur les courbes hypsométriques en définissant une variante de l'indice de pente IPB (e.g. Le Moine, 2008) :

$IPB = \frac{z_{0.95} - z_{0.05}}{\sqrt{S}}$ où z_x désigne le quantile x d'altitude sur le bassin et S sa superficie.

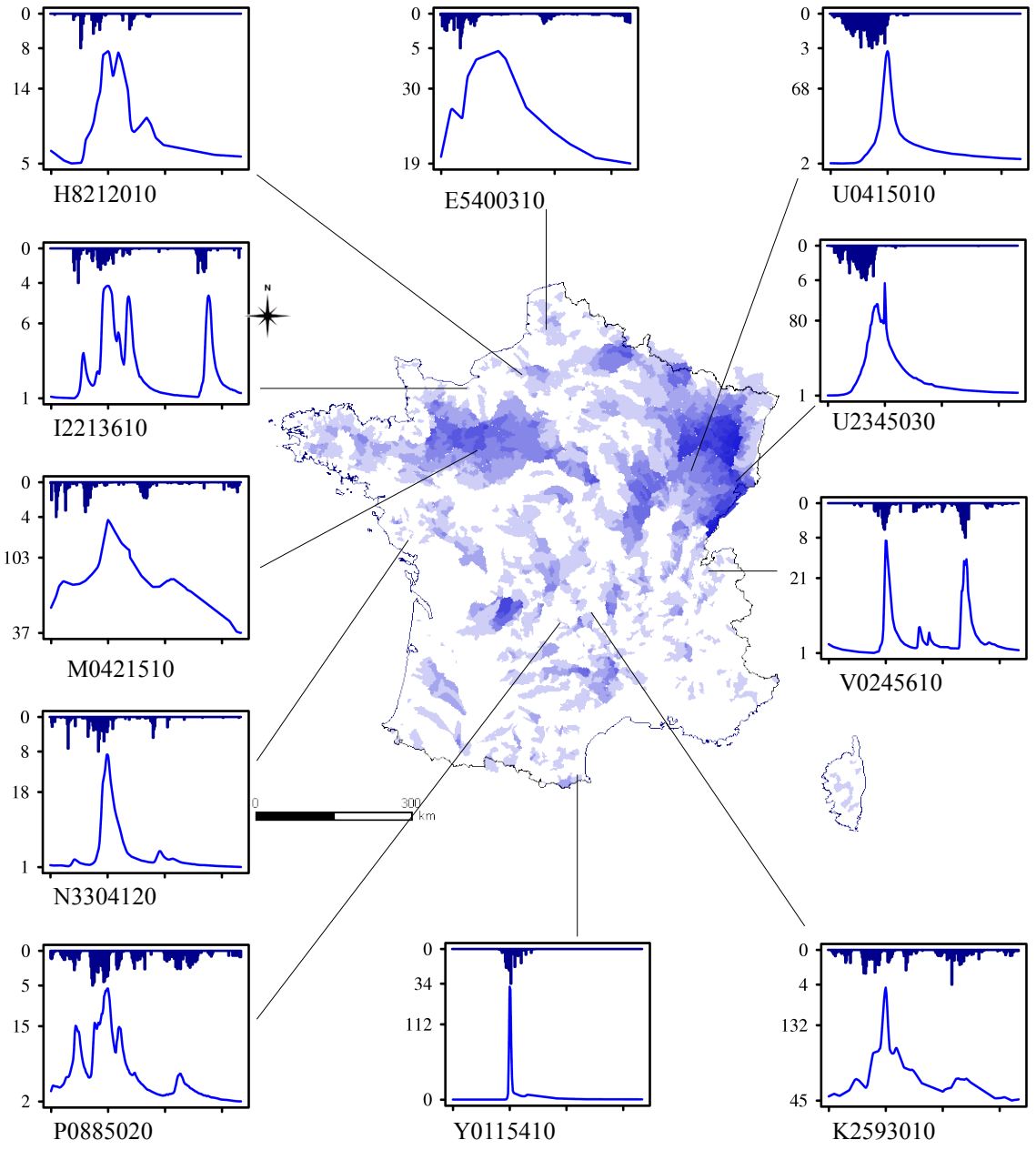


Figure 4.2 – 1 024 bassins versants sur la France (l'intensité de la couleur d'une aire dénote le nombre de bassins de notre basse de données la recouvrant). Les hydrogrammes entourant la carte illustrent très partiellement la variété des crues de nos bassins : ils correspondent aux plus fortes crues de onze bassins versants sur la période 1995 – 2005. Ces bassin sont la Canche à Brimeux (894 km^2, E5400310), l'Andelle à Vascoeuil (371 km ^{2}, H821201), l'Ancre à Cricqueville-en-Auge (57 km^2, I2213610), l'Alagnon à Lempdes (984 km^2, K2593010), l'Huisne à Montfort-le-Gesnois (1890 km^2, M0421510), le Marillet à Saint-Florent-des-Bois (48 km^2, N3304120), le Mars à Bassignac (117 km^2, P0885020), le Breuchin à la Proiselière-et-Langle (123 km^2, U0415010), la Savoureuse à Belfort (141 km^2, U2345030), l'Aire à Saint-Julien-en-Genevois (40 km^2, V0245610) et la Massane à Argelès-sur-Mer (17 km^2, Y0115410). Les pluies (en bleu foncé, en haut) sont données en mm, les débits (bleu marine, en bas) en m^3.s^{-1}. Tous les graphes représentent une durée de dix jours.

Quantiles	0	0.10	0.25	0.50	0.75	0.90	1	Moyenne
Superficie (km^2)	9	37	76	157	384	1035	9423	461
Altitude de l'exutoire (m)	2	36	80	194	338	586	1589	260
Indice de pente IPB (‰)	0.0	2.9	5.7	12	27	55	257	22

TABLE 4.1 – Résumés des distributions statistiques de quelques descripteurs physiques des 1 024 bassins versants.

Cet indice nous permet de ne pas prendre en compte les parties les plus basses (et plates) ni les parties les plus élevées (et escarpées) des bassins versants. Si quelques bassins présentent un indice de pente assez fort, la majorité d'entre eux se révèlent beaucoup plus plats (Tab. 4.1 et Fig. G.5).

Les altitudes à l'exutoire s'échelonnent de 2 m (cote NGF) à 1 589 m. La répartition s'équilibre entre bassins de plaines et bassins plus montagneux (Fig. G.5) : la plupart des bassins ne sont ni très pentus ni d'altitude très élevée ; cependant l'échantillon compte quelques bassins de montagne (notamment dans les Alpes et le Jura).

Ainsi, les indices que nous venons de présenter montrent que les différents régimes climatiques français sont représentés par l'ensemble de 1 024 bassins dont nous disposons. Voyons maintenant quelle est la variété hydrométéorologique de ces bassins versants. Comme nous cherchons à construire un modèle de prévision de la transformation de la pluie en débit, nous allons employer des indices décrivant les régimes hydrologiques et pluvieux.

4.2.2 Indices décrivant les débits

Les indices décrivant le débit ont été choisis parmi les descripteurs mentionnés par Michel (1991, Chap. 4) et employés par Tangara (2005). Ceux décrivant la pluie sont pour la plupart calqués sur ces indices.

Indices caractérisant le régime d'écoulement

Deux indices sont employés pour décrire le régime d'écoulement : il s'agit de l'écoulement annuel moyen (LAM) et de l'irrégularité saisonnière (IND) :

$$LAM = \frac{\overline{Q_{annuel}}}{S} \quad \text{et} \quad IND = \frac{Q_{Mmax} - Q_{Mmin}}{\overline{Q_{annuel}}}$$

L'indice LAM, exprimé en mm, est la lame d'eau correspondant au débit annuel moyen $\overline{Q_{annuel}}$. L'indice IND, exprimé en pourcentage, décrit la différence entre le mois présentant l'écoulement le plus fort et le mois présentant l'écoulement le plus faible, rapporté au débit annuel moyen. Les mois présentant les plus fort et plus faible débits mensuels moyens sont sélectionnés chaque année. Une moyenne de ces débits sur l'ensemble des années complètes permet de calculer Q_{Mmax} et Q_{Mmin}.

Indices décrivant les valeurs extrêmes

Nous employons trois indices pour caractériser les étiages et les crues : ce sont le débit spécifique d'étiage (ETI), l'étiage modulaire (ESM) et un indice basé sur le coefficient de crue biennale (CRU) :

$$ETI = \frac{DCN_{10,2}}{S} \ , \quad ESM = \frac{10 \cdot ETI}{LAM} \quad \text{et} \quad CRU = \frac{Q_{jx2}}{S^{0,8}}$$

Les étiages sont classiquement décrits par le dixième débit journalier biennal $DCN_{10,2}$: le dixième débit le plus faible de chaque année est retenu et la médiane de ces débits est calculée. Pour permettre la comparaison entre bassins, ce débit est exprimé en lame d'eau. L'indice ESM correspond à l'indice précédent rapporté au débit annuel moyen.

Enfin, les crues sont caractérisées de façon similaire par le débit journalier maximal biennal Q_{jx2} : chaque année, le débit le plus fort est retenu ; Q_{jx2} est la médiane de ces débits. Le facteur 0,8 est dû au fait que les débits de crues augmentent généralement proportionnellement non pas à S mais à $S^{0,8}$ (cf. la formule Crupedix).

Indice définissant le débit de base

L'indice d'écoulement de base BFI est notre dernier indice caractérisant le régime des débits. La chronique des débits est divisée en une suite de quintuplets (au pas de temps journalier), dont on prend les débits minimaux pour définir une suite finie $\left(Q_i^{[5]}\right)_{i \in [\![1,N]\!]}$. On sélectionne ensuite parmi eux les débits $Q_i^{[5]}$ tels que $Q_i^{[5]} \leq 0,9 \cdot min(Q_{i-1}^{[5]}, Q_{i+1}^{[5]})$. Ces débits reliés définissent l'écoulement de base. Pour quelques bassins présentant un régime d'écoulement très stable (notamment en Artois-Picardie), la valeur 0,9 était trop restrictive et a été remplacée par 1. L'indice BFI est le rapport de le volume de l'écoulement de base par le volume écoulé total.

4.2.3 Indices caractérisant les pluies

Les trois premiers indices proposés pour décrire les pluies sont calqués sur les indices précédents :

$$PAM = \frac{\overline{P_{annuelle}}}{S} \ , \quad IRR = \frac{P_{max} - P_{min}}{\overline{P_{annuelle}}} \quad \text{et} \quad PMX = \frac{P_{jx2}}{S}$$

La pluie annuelle moyenne PAM, exprimée en mm, correspond à l'indice LAM. L'indice IRR, équivalent de l'indice IND, mesure la différence entre le mois de l'année le plus pluvieux et le mois le plus sec, rapporté à la pluviométrie moyenne. Les pluies les plus intenses sont décrites par l'indice PMX construit sur le modèle de l'indice CRU.

Enfin, pour prendre en compte l'irrégularité des pluies et l'existence de jours de pluie nulle, nous utilisons l'indice SEC qui est la fraction annuelle moyenne des jours non pluvieux.

Indice	Signification	Unité	Moyenne	Écart-type
LAM	Écoulement annuel moyen	mm	432	305
IND	Irrégularité saisonnière des écoulements	%	257	85,7
ETI	Débits spécifiques d'étiage (basé sur $DCN_{10,2}$)	$dL.s^{-1}.km^{-2}$	23,8	24,7
ESM	Étiage modulaire	%	182	155
CRU	Coefficient de crue biennale	$L.s^{-1}.km^{-2}$	272	660
BFI	Indice d'écoulement de base	%	57,9	16,1
PAM	Précipitation annuelle moyenne	mm	1 049	282
IRR	Irrégularité saisonnière des précipitations	%	195	22,4
PMX	$10^{ème}$ Pluie la plus intense biennale	mm^{-2}	39,2	17,7
SEC	Fraction annuelle moyenne des jours secs	%	34,3	8,11

TABLE 4.2 – Statistiques des indices décrivant les régimes hydrographiques et pluvieux sur les 1 024 bassins versants.

4.2.4 Premiers résultats statistiques

Les moyennes et écart-types des indices sont donnés dans le tableau 4.2. Les valeurs calculées pour les indices décrivant les débits sont proches de celles obtenues par Tangara (2005) sur un échantillon de 188 bassins versants composés de 130 bassins français et de bassins américains, ivoiriens, brésiliens et australiens. Les distributions statistiques des indices caractérisant les débits sont présentées par la figure G.6, en page 429. Les distributions des indices décrivant les pluies sont données en figure G.7. Ces distributions ressemblent fortement à celles obtenues par Tangara (2005). Or les deux échantillons sont à peu près indépendants (ils comptent 56 bassins en commun). Ces similarités nous poussent à penser que chaque échantillon représente plutôt bien la variété des régimes hydrologiques possibles de la réunion de ces deux ensembles... Et des zones géographiques couvertes.

Cependant, les indices sont partiellement corrélés, comme nous pouvons le constater sur la figure G.8, à partir de la page 431. Du fait de cette corrélation, il n'est pas possible de représenter clairement la variété de l'échantillon à partir de ces indices. Il nous faut donc aller plus loin dans l'analyse de nos données pour caractériser les bassins versants et pour décrire la variété hydrologique de l'échantillon. C'est le sujet des prochaines sections.

4.3 Définition et caractérisation de groupes de bassins versants homogènes

Nous employons deux méthodes classiques d'analyse des échantillons pour définir et caractériser des sous-groupes de bassins versants : la classification ascendante hiérarchique et l'analyse en composantes principales. Ces deux techniques sont présentées en annexe I à partir de la page 485.

Elles ont pour but de proposer une meilleure description de la variété de nos données et un échantillonnage présentant moins de tendances artificielles et indésirables.

Deux analyses ont été menées : la première ne concerne que les six indices décrivant le débit, la seconde s'intéresse aux dix indices caractérisant la pluie et le débit. Au cours de chacune de ces analyses, nous avons examiné l'influence potentielle des caractéristiques topographiques sur la définition des sous-groupes. Nous ne présentons ici que la seconde analyse car la suite du chapitre se base sur elle. Le lecteur intéressé trouvera le compte-rendu de l'analyse sur les six indices décrivant le débit en annexe G à partir de la page 423.

4.3.1 Analyse portant sur les indices de pluie et de débits

4.3.1.1 Résultats de l'analyse en composantes principales

Les valeurs propres et vecteurs propres associés obtenus lors de l'A.C.P. sont donnés dans le tableau 4.3. Le nombre de degrés de liberté ayant augmenté par rapport à la première analyse, la somme des deux plus grandes valeurs propres est de 62 % : la représentation de la variété de l'échantillon dans le plan principal de l'A.C.P. sera donc moins bonne. Si on se restreint aux trois premiers axes, on peut représenter (en 3D) 76 % de la variété de l'échantillon.

Valeurs propres		**3,31**	**2,87**	**1,41**	**0,87**	**0,48**	**0,38**	**0,30**	**0,19**	**0,14**	**0,05**
Fraction cumulée (%)		33,1	61,8	75,9	84,6	89,4	93,2	96,2	98,1	99,5	100
	LAM	0,39	-0,35	-0,18	0,07	0,32	0,19	-0,03	-0,32	-0,14	0,66
	IND	0,23	0,45	-0,02	-0,25	0,45	0,12	-0,01	0,57	0,31	0,21
	ETI	0,05	-0,53	0,01	0,05	0,51	0,03	-0,27	0,31	-0,23	-0,48
	ESM	-0,29	-0,41	0,21	-0,03	-0,06	-0,56	-0,12	0,27	0,35	0,41
Vecteurs	CRU	0,45	-0,09	-0,20	-0,16	0,11	-0,53	0,57	-0,13	0,14	-0,26
propres	BFI	-0,29	-0,34	0,25	-0,33	0,05	0,50	0,58	0,02	0,19	0,00
	PAM	0,42	-0,26	-0,13	0,09	-0,40	0,30	-0,26	0,05	0,62	-0,18
	IRR	0,22	0,08	0,61	-0,50	0,11	-0,09	-0,34	-0,42	0,06	-0,10
	PMX	0,43	-0,09	0,37	-0,03	-0,44	0,01	0,15	0,45	-0,48	0,13
	SEC	0,11	0,11	0,54	0,73	0,24	0,01	0,21	-0,07	0,20	-0,04

Table 4.3 – Valeurs propres et vecteurs propres associés de la matrice de covariance des indices de débit et de pluie.

Le premier axe principal est fortement correlé avec les indices LAM et PAM, ainsi qu'avec les indices PMX et CRU : cet axe permet de décrire la variété de l'échantillon en termes de quantités d'eau tombée et écoulée, sur une échelle de temps annuelle et / ou de quelques jours (plus fortes précipitations, crues).
Le deuxième axe principal se caractérise par une corrélation positive marquée avec l'indice

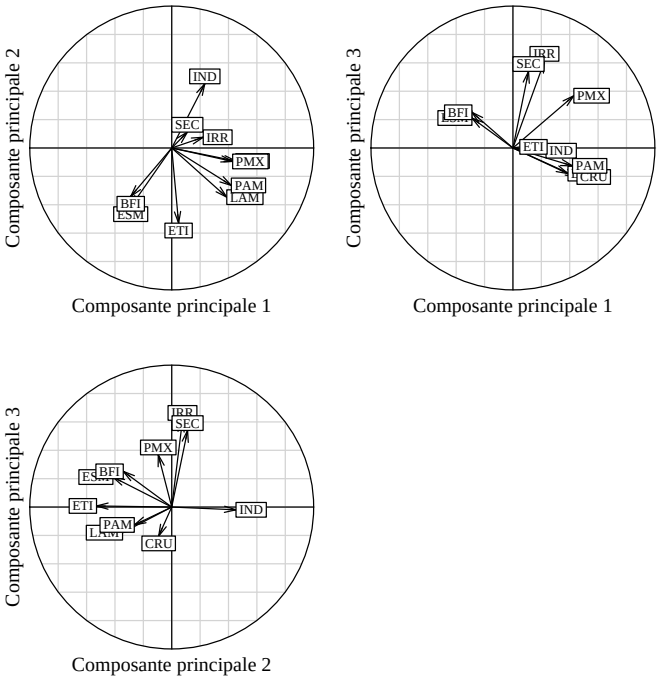


FIGURE 4.3 – Plans principaux de l'A.C.P. sur les 10 indices décrivant le débit et la pluie.

IND et une corrélation négative avec l'indice ETI. Il est également corrélé dans une moindre mesure négativement avec l'indice ESM. Cet axe différencie les bassins à débits irréguliers des bassins à débits réguliers. Il est à noter que cet axe est très peu corrélé avec les indices décrivant la pluie (Fig. 4.3).
Au contraire, le troisième axe principal n'est fortement corrélé qu'avec des indices de pluie : positivement avec l'indice IRR, puis avec l'indice SEC. Cet axe caractérise l'irrégularité des pluies.

4.3.1.2 Résultats de la classification ascendante hiérarchique

Nous avons utilisé la C.A.H. pour définir 4 puis 8 groupes homogènes de bassins au sens de nos 10 indices. Intéressons-nous à la partition de l'échantillon en 4 groupes (Fig. 4.4). Les groupes sont assez peu différenciés par le deuxième axe principal, mais plus selon une diagonale du plan (1,3) : les groupes se distinguent principalement par la quantité d'eau annuelle qui les traversent et les forts précipitations et débits. Il est plus difficile de décrire cette partition en 4 groupes que pour l'analyse ne portant que sur les débits (Ann. G, page 423) : il y a trop de degrés de liberté. La discussion sur la partition plus fine en 8 groupes n'est guère plus informative.

La répartition géographique est par contre plus intéressante (Fig. 4.5). Nous pouvons distinguer un premier groupe de bassins situés en zone montagneuse (Pyrénées, Alpes, Massif

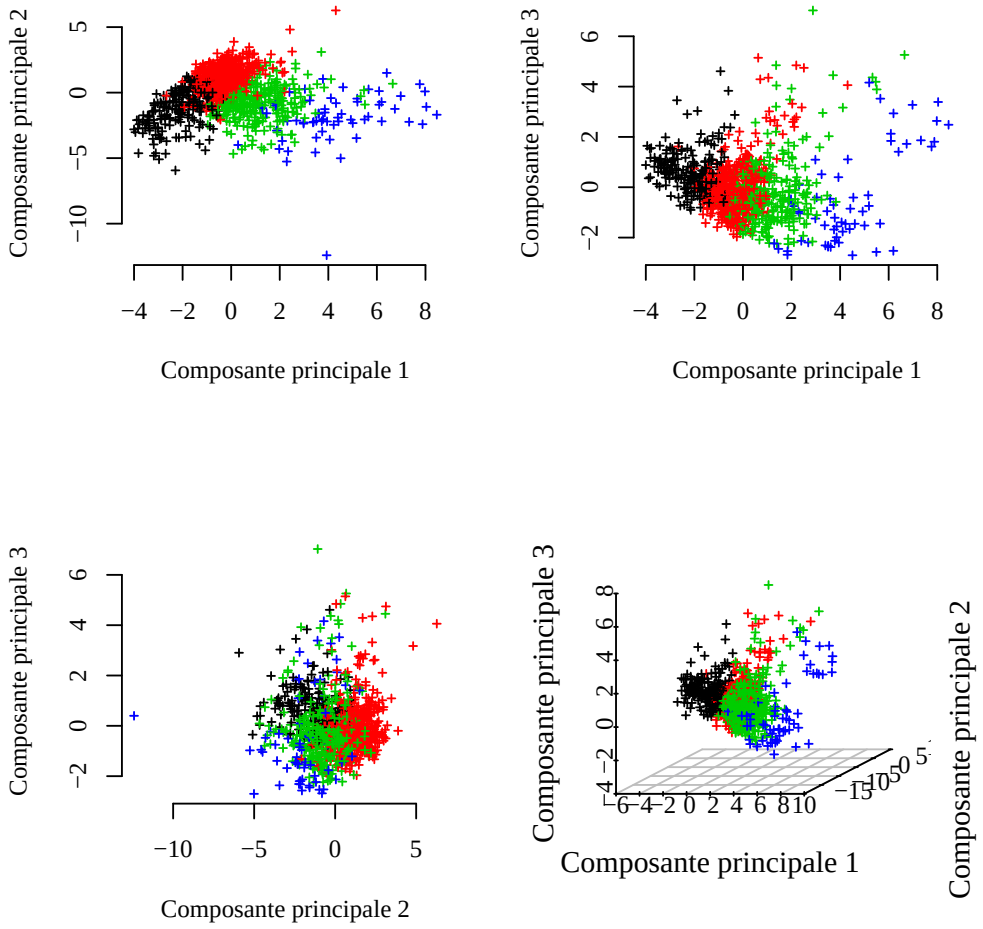


Figure 4.4 – Partition de l'échantillon de 1 024 bassins en 4 groupes homogènes par C.A.H., représentée dans les 3 principaux plans de l'A.C.P.

central, Massif armoricain...), un groupe de bassins du Nord (Artois, Picardie, bassin aval de la Seine), un groupe de bassins de plaine et enfin un groupe petit réduit de bassins versants moins facilement caractérisables.

Ces deux analyses et les graphes décrivant la variété de différents indices présentés en annexe G montrent que l'échantillon contient des bassins versants présentant de nombreux régimes hydrométéorologiques différents présents en France (métropolitaine).

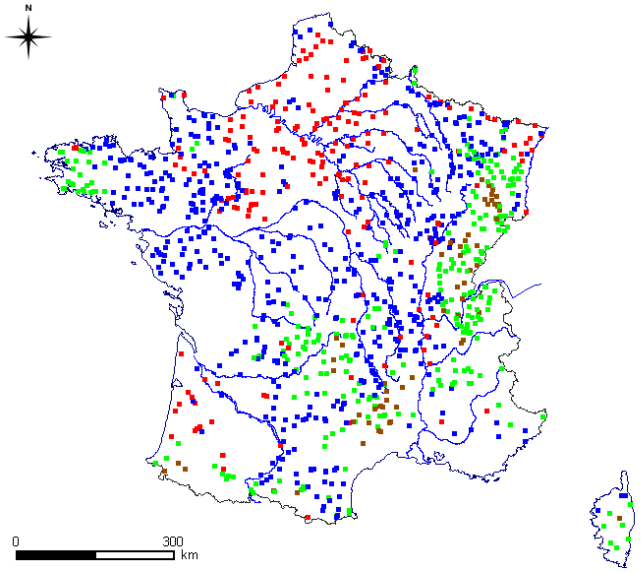


FIGURE 4.5 – Répartition géographique des quatre groupes homogènes de bassins versants définis à partir de l'échantillon de 1 024 bassins versants par la C.A.H. calculée sur les 10 indices décrivant les régimes hydrographiques et pluvieux.

4.4 Limiter les biais : passer de l'ensemble à l'échantillon

Il est tout à fait possible, voire probable, que l'ensemble initial compte parmi ses 1 024 bassins, des individus présentant une certaine similarité hydrologique et sur lesquels le modèle se comporte de façon très semblable. Cela introduit un déséquilibre dans notre évaluation du modèle en défaveur des bassins les plus différents. En outre, tester le modèle sur des bassins similaires est une redondance moins informative mais coûteuse en temps de calcul. Nous cherchons ici à réduire ce biais (et le temps de calcul des tests) tout en conservant la variété hydrologique du jeu de données initial de 1 024 bassins versants.

4.4.1 Méthode proposée

Nous nous appuyons sur les résultats de la classification ascendante hiérarchique (C.A.H.). Une fois le dendrogramme construit comme expliqué dans l'annexe I.1, nous le descendons partiellement à partir de son sommet pour arriver à des niveaux comptant déjà un nombre significatif de groupes de bassins versants mais nettement moins que l'ensemble initial. Par la construction de la C.A.H., ces groupes sont le plus homogènes possible au sens des indices utilisés par la C.A.H. Nous déterminons alors au sein de chaque groupe, le bassin versant qui est le plus proche des autres (au sens de la distance euclidienne pour rester cohérent avec l'algorithme utilisé lors de la C.A.H.). Nous sélectionnons ainsi un bassin par groupe pour obtenir un échantillon comptant entre 10 et 3 fois moins de bassins versants que l'échantillon initial, tout en conservant un maximum de la variété initiale.

4.4.2 Tests réalisés

Pour tester cette méthode, nous avons repris les résultats des deux C.A.H. qui ont été réalisées précédemment (la première sur les indices décrivant le débit uniquement, la seconde sur les indices caractérisant la pluie et le débit). Nous sommes descendus de 5, 6 et 7 niveaux de noeuds dans le dendrogramme pour définir 107, 190 et 322 groupes (respectivement 116, 221 et 384). Nous avons ainsi déterminé les échantillons restreints que nous nommerons pour la suite Q-107, Q-190, Q-322 (respectivement PQ-116, PQ-221 et PQ-384).

Le but de nos tests est de comparer la variété hydrologique des échantillons restreints à celle de l'ensemble complet. Nous avons donc tracé les distributions expérimentales des indices caractérisant la pluie et le débit et les avons comparées aux distributions obtenues sur l'échantillon complet. Pour obtenir des distributions comparables, les bassins des échantillons restreints sont pondérés par le nombre de bassins que comptent les groupes dont ils sont issus.

L'objectif *in fine* de cette méthode est d'utiliser des échantillons restreints pour décrire les performances de modèles. Il est donc important d'observer comment réagissent les distributions d'un critère de performance quand on restreint l'échantillon par notre méthode. Pour cela, nous avons comparé les distributions des critères de Nash-Sutcliffe NSE obtenus pour les différents échantillons avec le modèle de simulation GR4J en validation sur deux périodes distinctes P_1 et P_2 (Le Moine, 2008). Nous nous intéressons également à la variété du critère de Nash-Sutcliffe à l'intérieur des groupes.

4.4.3 Résultats

4.4.3.1 Distributions des indices sur des échantillons restreints

Les distributions des 6 indices caractérisant le débit sur les échantillons restreints construits par C.A.H. sur ces indices sont présentées en annexe G, à partir de la page 442, par la figure G.12. Les distributions des 10 indices caractérisant le débit et la pluie des échantillons restreints basés sur ces 10 indices sont données à la suite (Fig. G.13).

Nous observons que même les échantillons les plus parcimonieux donnent déjà une bonne idée des distributions de tous les indices. Naturellement, plus l'échantillon est large, plus les distributions sont proches des distributions obtenues avec l'échantillon complet. Pour quantifier les écarts, nous avons calculé les moyennes et écarts-types pour chaque indice et pour chaque échantillon. Ils sont donnés dans les tableaux 4.4 et 4.5.

Pour un même niveau de descente dans le dendrogramme (échantillons Q-107 et PQ-116, Q-190 et PQ-221, Q-322 et PQ-384), les distributions calculées sur les échantillons restreints obtenus par C.A.H. sur les 10 indices caractérisant la pluie et le débit sont nettement plus proches des distributions calculées sur l'échantillon complet que celles calculées sur les

	Tous	PQ-384	PQ-221	PQ-116	*Q-322*	*Q-190*	*Q-107*
LAM	**432**	516	535	570	*534*	*565*	*600*
IND	**257**	235	240	249	*249*	*248*	*244*
ETI	**23.8**	33.2	34.3	35.5	*30.8*	*34.7*	*38.8*
ESM	**182**	239	241	235	*212*	*229*	*243*
CRU	**272**	321	346	378	*351*	*382*	*395*
BFI	**0.579**	0.611	0.613	0.616	*0.577*	*0.588*	*0.593*
PAM	**1049**	1111	1112	1113	–	–	–
IRR	**195**	197	196	198	–	–	–
PMX	**39.2**	42.8	42.7	44.0	–	–	–
SEC	**34.3**	34.4	34.2	34.1	–	–	–
NSE P_1	**76.1**	77.4	78.8	79.5	*77.4*	*76.0*	*77.4*
NSE P_2	**79.9**	81.4	82.6	84.4	*82.1*	*81.8*	*81.0*

TABLE 4.4 – Moyennes des indices pour les différents échantillons

	Tous	PQ-384	PQ-221	PQ-116	*Q-322*	*Q-190*	*Q-107*
LAM	**305**	390	422	467	*422*	*464*	*521*
IND	**85**	104	114	123	*110*	*121*	*129*
ETI	**24.7**	32.4	36.4	42.2	*34.4*	*39.6*	*46.9*
ESM	**155**	192	202	203	*202*	*218*	*235*
CRU	**236**	312	351	386	*352*	*399*	*405*
BFI	**0.161**	0.192	0.200	0.210	*0.217*	*0.240*	*0.255*
PAM	**283**	320	315	319	–	–	–
IRR	**22.4**	27.1	26.4	27.3	–	–	–
PMX	**17.7**	23.1	22.0	24.8	–	–	–
SEC	**8.11**	8.21	8.14	7.76	–	–	–
NSE P_1	**31.8**	31.4	31.2	37.2	*31.7*	*39.5*	*50.1*
NSE P_2	**18.3**	17.3	16.7	16.7	*16.2*	*20.4*	*21.2*

TABLE 4.5 – Écarts-types des indices pour les différents échantillons

échantillons obtenus par C.A.H. sur les seuls indices caractérisant le débit. Cela confirme, si cela était nécessaire, que l'hydrologie du bassin-versant ne peut pas être décrite uniquement par des indices caractérisant les débits.

4.4.3.2 Distributions des critères de performance sur des échantillons restreints

Les distributions des critères de Nash-Sutcliffe calculés en validation avec le modèle de simulation GR4J sur les échantillons construits par C.A.H. sur les 6 indices décrivant le débit sont données en figure G.12 (page 444). Celles correspondant aux échantillons basés sur les 10 indices sont représentées par la figure G.13 (page 454).

Les distributions des critères de Nash-Sutcliffe obtenus avec GR4J sont également bien conservées lors des réductions d'effectifs de l'échantillon testé. Ici encore, on note une meilleure représentation de la distribution de l'échantillon complet obtenue avec les échantillons restreints déterminés par la C.A.H. appliquée aux indices caractérisant la pluie et le débit. On peut observer la disparition dans les échantillons restreints des bassins donnant les plus mauvaises performances. Ceci est quelque peu intriguant car cette disparition signifie que ces bassins appartiennent à des sous-groupes de bassins « hydrologiquement proches » les uns des autres qui comptent des bassins donnant de nettement meilleurs résultats. Les indices choisis ne permettent donc pas de rendre compte de toute la variété hydrologique des bassins.

4.4.3.3 Variété du critère de performance au sein des groupes

Enfin, nous avons tracé les distributions du critère de Nash-Sutcliffe à l'intérieur des groupes correspondant aux échantillons restreints Q-107, Q-190, PQ-116 et PQ-221 [1] dans la figure G.16 (à partir de la page 466). Les groupes définis par C.A.H. dans l'espace des indices de pluie et de débit (ou seulement de débit) sont homogènes au sens de ces indices mais également au niveau des performances obtenues par un modèle global Pluie–Débit.

Sur tous ces tests, les groupes contenant le plus de bassins obtiennent des critères de Nash-Sutcliffe très stables et élevés. Les bassins présentant les critères de performance les plus bas sont seuls dans leurs groupes ou inclus dans des groupes comptant un faible effectif : ils ressemblent moins aux autres. On observe de surcroît que les bassins versants sur lesquels le modèle hydrologique GR4J présentent ses meilleures performances sont situés statistiquement au centre du nuage de points dans le plan principal de l'A.C.P. Ces observations peuvent être interprété ainsi : le modèle GR4J a été optimisé sur un nombre important de bassins versants et la variété de ces bassins versants était similaire à

1. Les échantillons Q-322 et PQ-384 comptant plus de bassins, ces graphes devenaient fastidieux et illisibles car un très grand nombre de groupes ne contiennent qu'un seul bassin.

celle de notre échantillon : plus un bassin est proche du bassin « moyen », plus le modèle y obtient de bons résultats.

4.4.4 Réduction de l'échantillon

En pratique, il semble possible de tester la robustesse des modèles sur des échantillons moins fournis que l'ensemble initial, qui conserveraient la plus grande partie de sa variété. Nous pouvons utiliser les échantillons construits par C.A.H. dans ce but.

Quelle C.A.H. employer ?

Les groupes définis par les deux C.A.H. sont assez différents, comme on peut le constater dans le tableau 4.6. Les observations précédentes permettent de conclure que la variété du critère de Nash-Sutcliffe est globalement plus faible à l'intérieur des groupes définis par C.A.H. sur les 10 indices que pour les groupes définis à partir des 6 indices décrivant le débit. En outre, la distribution du critère est mieux reconstituée avec les échantillons restreints définis par C.A.H. sur les 10 indices. Il est donc préférable d'utiliser un échantillon restreint obtenu par C.A.H. sur les indices décrivant les régimes des pluies et des débits.

		C.A.H. - Q			
		1	2	3	4
CAH - PQ	1	118	20	155	2
	2	102	64	192	0
	3	49	0	8	79
	4	164	7	47	37

TABLE 4.6 – Table d'intersection des deux partitions de l'ensemble des 1 024 bassins versants en quatre groupes définis par la C.A.H. calculée sur les six indices décrivant le régime des débits et celles en quatre groupes déterminés sur les dix indices : chaque case indique le nombre de bassins versants dans l'intersection d'un groupe de l'une des C.A.H. et un groupe de l'autre C.A.H.

Quelle pondération ?

Jusqu'ici, nous avons pondéré chaque bassin versant des échantillons restreints par le nombre de bassins contenus par le groupe défini par la C.A.H. dont il est issu. Cette pondération avait pour but de pouvoir comparer de façon similaire les échantillons restreints entre eux et à l'échantillon complet.

Cette pondération n'est liée qu'à l'ensemble initial des données. Or celui-ci n'est lui-même pas un échantillon idéal mais est le résultat de contraintes pratiques (disponibilité des données) qui n'est pas justifié sur le plan hydrologique. Pour réduire la part du hasard et celle des contraintes pratiques dans notre base de données, nous allons abandonner cette pondération et calculer les distributions des critères de performance sur l'échantillon restreint en

donnant un poids égal à chaque bassin sélectionné qui est alors censé représenter un type de régime hydrologique.

Définition de l'échantillon servant à l'ensemble des tests de cette thèse

Nous avons basé l'ensemble du travail présenté dans ce mémoire de doctorat sur un ensemble initial de bassins versants légèrement différent de celui sur lequel s'est basé cette étude (pour des problèmes de disponibilité au moment où l'échantillon a été calculé). La même méthode – C.A.H. calculée sur les 10 indices définis précédemment – a été employée pour obtenir un échantillon de 178 bassins qui nous servira pour toutes les expériences numériques qui sont décrites dans la suite de ce document.

Au début de ce chapitre, nous avions constaté que le Nord-Est de la France était très représenté au sein des 1 024 bassins versants alors que le pourtour méditerranéen est quasiment absent. La méthode que nous avons présentée ne permet évidemment pas d'obtenir plus de bassins versants méditerranéens mais nous constatons sur la figure 4.6 que l'échantillon restreint est nettement plus équilibré (au moins géographiquement parlant) que l'ensemble initial.

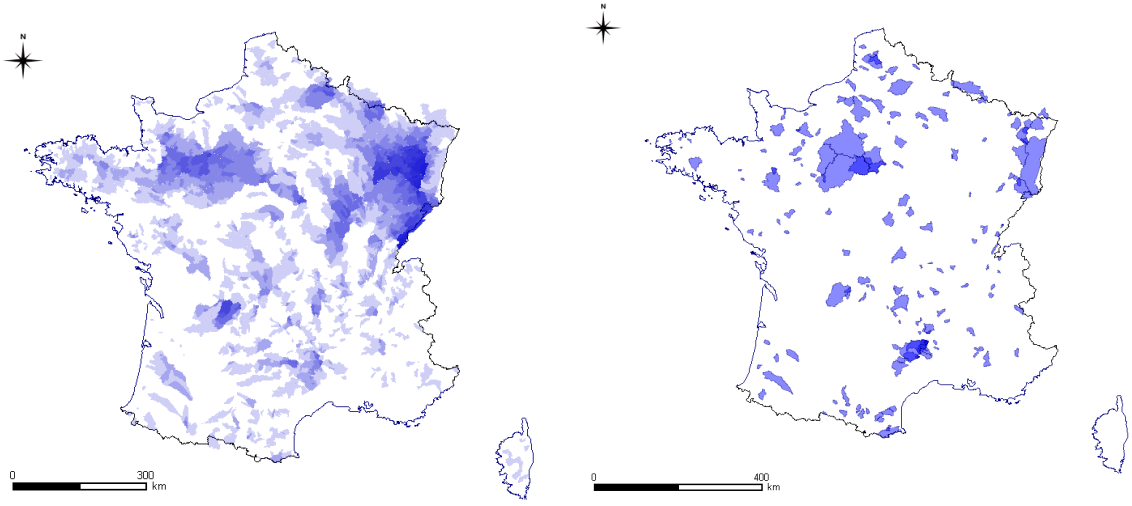


Figure 4.6 – Ensemble initial des 1 024 bassins versants (à gauche) et échantillon restreint à 178 bassins choisis pour contenir la diversité hydrologique de l'ensemble initial (à droite). L'intensité des couleurs indique le nombre de bassins superposés.

Employer cette classification pour comparer les résultats de nos modèles ?

Dans le chapitre 2 consacré à la méthodologie de nos tests, nous avons montré que les bassins versants ne réagissent pas tous de la même façon. Nous avons adopté l'autocorrélation des débits pour caractériser la réactivité de nos bassins et enrichir l'analyse des comparaisons de différents modèles sur notre échantillon de bassins. Est-il possible de compléter cette analyse en se basant sur la classification présentée dans ce chapitre ? Cela nous

paraît difficile car cette classification est basée sur de nombreux indices à la différence de celle uniquement établie sur l'autocorrélation des débits. Si la classification compte peu de groupes, caractériser ces derniers sera complexe ; si elle en compte beaucoup, l'analyse des résultats des modèles en fonction de ces groupes sera ardue et difficile à interpréter en pratique.

4.5 Bilan et conclusions

Nous disposons de chroniques de données horaires de pluie, évapotranspiration potentielle et débit sur un ensemble de 1 024 bassins versants. La relation Pluie – Débit sur ces bassins versants est décrite par 10 indices calculés sur les données de précipitation et de débit. Une classification ascendante hiérarchique des bassins versants calculée sur ces indices permet de classer les bassins en groupes homogènes que nous pouvons caractériser par une analyse en composantes principales. Cette étude montre que l'ensemble des bassin versants est varié et présente la plus grande partie des régimes hydrologiques de la France métropolitaine.

La classification ascendante hiérarchique nous permet en outre de définir des échantillons comportant moins de bassins mais renfermant la variété de l'échantillon initial, au sens des indices que nous avons définis. Les distributions des indices calculées sur ces échantillons restreints sont raisonnablement proches des distributions calculées sur l'ensemble des bassins versants. En outre, ces échantillons permettent de représenter la distribution d'un critère de performance d'un modèle de la famille de celui que nous voulons élaborer, alors que ce critère n'a pas servi à définir les échantillons.

Enfin, nous avons pu constater grâce à cette approche que l'ensemble des 1 024 bassins n'est pas idéal dans le sens où il comprend un grand nombre de bassins très similaires (toujours au sens des indices définis). Cette démarche permet d'enlever les « redites » contenues dans l'ensemble initial et de réduire ce déséquilibre : la distribution des critères de performance calculée sur ces échantillons sera moins dépendante des *alea* liés à la disponibilité des données. Dans une vision pratique, réduire ainsi le nombre de bassins conduit à diviser les temps de calcul d'un facteur de l'ordre de 5 à 6. Nous travaillerons désormais sur un échantillon de 178 bassins versants ainsi sélectionnés.

Deuxième partie

Passer du pas de temps journalier au pas de temps horaire

5
Chapitre

Influence de la résolution temporelle sur le modèle

Objectifs

- Définir la résolution temporelle appropriée pour la modélisation dédiée à la prévision des crues.
- Caractériser l'influence du pas de temps sur le fonctionnement du modèle.

Sommaire

5.1 Comparaison des hydrogrammes journalier et horaire 77

5.2 Influence du pas de temps sur le fonctionnement et les performances du modèle 81

5.2.1 Influence sur le fonctionnement : optimisation des valeurs des paramètres fixes 81

5.2.2 Quel pas de temps pour quel horizon ? 87

5.3 Bilan 87

Le modèle GRP, comme l'immense majorité des modèles hydrologiques, ne simule pas les débits instantanés, mais estime le débit moyenné sur un pas de temps. L'une des premières étapes de la modélisation est le choix de la résolution temporelle. Nous nous intéressons ici au pas de temps du modèle. Son choix est d'abord fonction du phénomène dont nous voulons rendre compte. Le pas de temps doit être suffisamment fin pour permettre une bonne description temporelle du phénomène, mais ne pas être inutilement court pour éviter des calculs sans intérêt. Les différentes phases des hydrogrammes présentent des temps caractéristiques bien différents. Face au dilemme du choix d'un pas de temps de modélisation approprié qui ne soit ni trop fin (inutile la plupart du temps), ni trop grossier (inadapté aux événements les plus importants), Hughes (1993) choisit une approche basée sur un pas de temps variable qui évolue en fonction des événements (de crue). Comme nous nous intéressons à la prévision des crues, ce choix sera d'abord guidé par la dynamique de ce phénomène.

Le pas de temps de la modélisation est également fonction de la résolution des données disponibles : il n'est pas inférieur à celui des données d'entrée (observations et prévisions de pluie). Il arrive fréquemment qu'on ne dispose de chroniques suffisamment longues pour permettre un calage que pour des pas de temps relativement longs. Nalbantis (1995) propose une méthode permettant d'exploiter des séries temporelles longues à un pas de temps grossier pour apporter de l'information à des pas de temps plus fins pour lesquels on ne dispose pas de séries suffisamment longues pour caler le modèle : il cale son modèle au pas de temps grossier puis transfère l'information obtenue sur les paramètres au pas de temps fin.

Enfin, en prévision, le pas de temps doit être choisi en tenant également compte des horizons qui nous intéressent et qui nous sont accessibles. Ces horizons sont des multiples du pas de temps.

Ce chapitre débute par la comparaison des hydrogrammes de crue représentés aux pas de temps journalier et horaire. Dans une seconde partie, nous examinerons l'influence du pas de temps sur le fonctionnement et les performances du modèle GRP proposé par Tangara (2005).

5.1 Comparaison des hydrogrammes journalier et horaire

Le temps de réponse à un fort événement de pluie est inférieur à 24 heures sur de nombreux bassins en France. Aussi, le pas de temps journalier pourrait ne pas être adapté à de nombreuses situations de crues.

Nous commençons par explorer les différences entre les informations contenues par les hydrogrammes journalier et horaire. L'agrégation temporelle au pas de temps journalier a pour effet de lisser fortement le signal. La figure 5.1 illustre la différence entre ces deux

hydrogrammes sur les 9 événements de crue les plus forts de la période 1995 – 2005 sur la Creuse à Felletin (L4010710, 165 km^2). Nous constatons sans surprise que les composantes les plus rapides et les plus fortes (notamment au niveau des pointes de crue) sont éliminées ou très fortement atténuées par l'hydrogramme journalier. Il n'est pas possible d'évaluer les pointes de crue avec précision à l'aide d'un tel hydrogramme.

Cette propriété est naturellement liée à la réactivité du bassin versant. Le débit de la

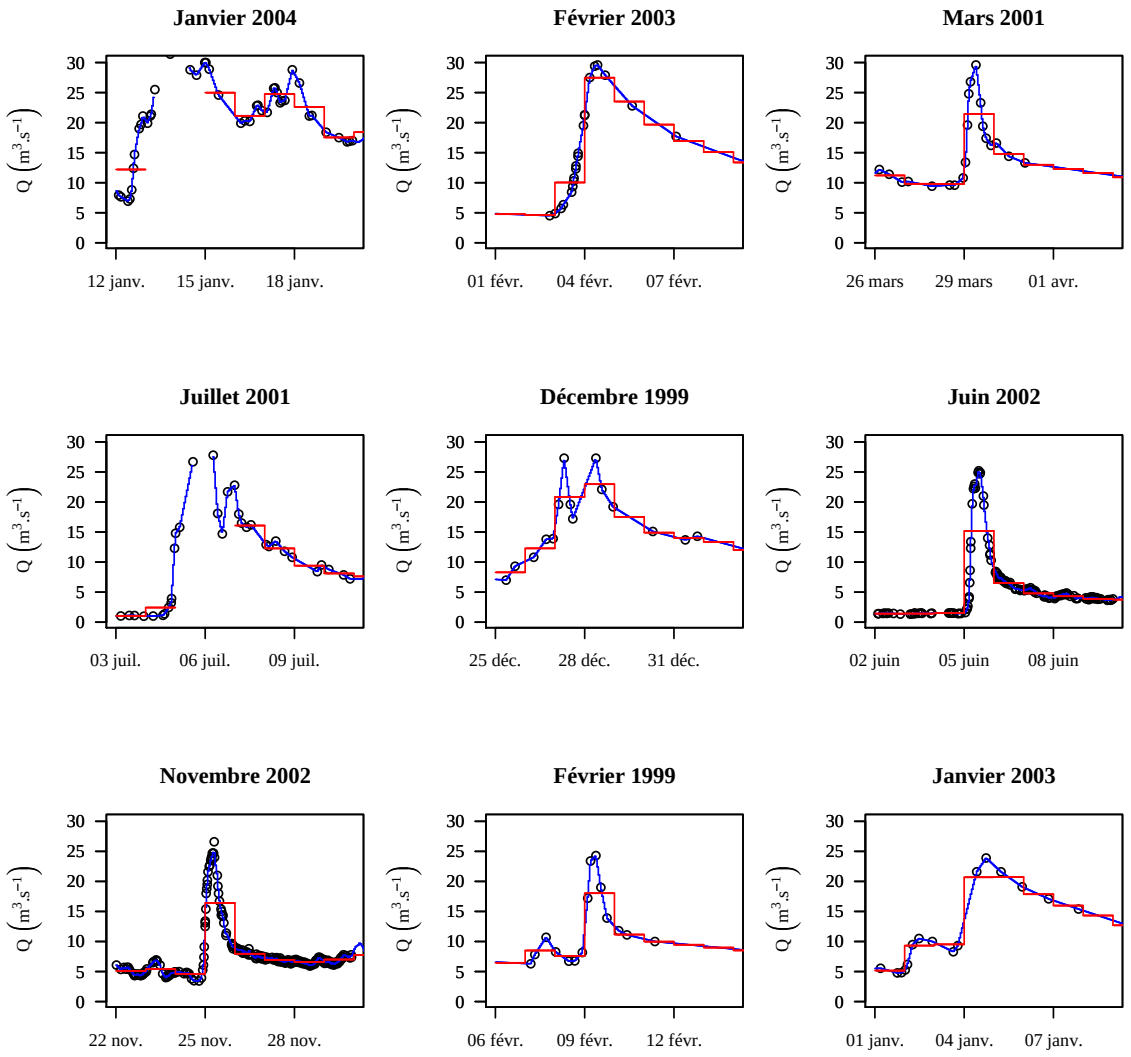


FIGURE 5.1 – Comparaison des hydrogrammes horaires (en bleu) et journaliers (en rouge) sur les neuf plus forts événements recensés sur la Creuse à Felletin (165 km^2) de 1995 à 2005. L'hydrogramme instantané est donné en arrière plan par les cercles noirs : on observe que l'hydrogramme horaire rend presque parfaitement compte des mesures instantanées. Les hydrogrammes sont classés par valeur décroissante des débits de pointe observés.

Creuse à Felletin se caractérise par une autocorrélation se situant autour de la médiane des autocorrélations du débit [1], calculée sur notre échantillon de 178 bassins versants : cette rivière n'est ni particulièrement rapide, ni particulièrement lente, comparée aux autres bassins français (de notre échantillon).

Nous avons évalué sur notre échantillon, le rapport des débits horaire et journalier pour différents quantiles. Nous constatons sans surprise qu'en dehors des périodes de hautes eaux (quantiles inférieurs à 0.95), les rapports sont très proches de 1 : le pas de temps horaire apporte peu d'information supplémentaire, par rapport au pas de temps journalier (Fig. 5.2). Cela change pour les débits élevés et de crue : les différences deviennent significatives et nous observons des différences médianes de 5 % pour le quantile 0.99, 15 % pour le quantile 0.999 et 30 % pour les débits maximaux. La situation présentée pour la Creuse n'est pas un cas isolé : le pas de temps journalier sous-estime fortement le débit de pointe sur une grande majorité de bassins versants. C'est probablement le cas également pour le

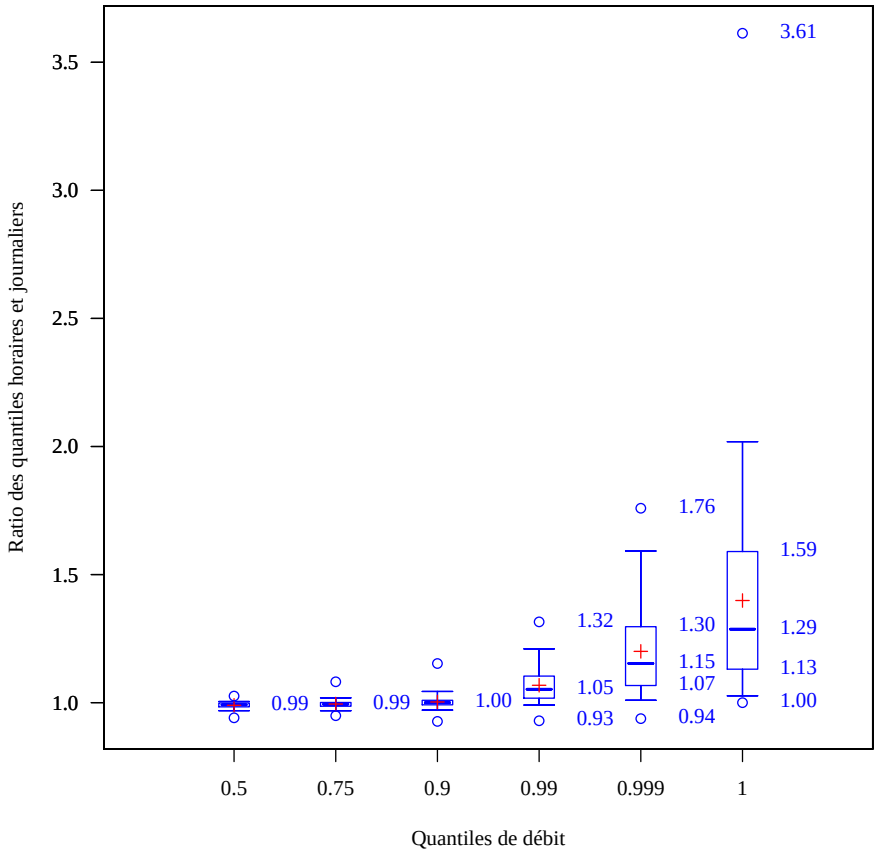


Figure 5.2 – Distribution des rapports entre débit horaire et débit journalier sur 178 bassins en France, pour les quantiles 0.5 (médiane), 0.75, 0.90, 0.99, 0.999, 1 (débit maximal).

1. Vérifié pour des décalages temporels compris entre 1 et 48 heures.

pas de temps horaire sur certains bassins, mais bien entendu dans une moindre mesure.

Nous avons finalement comparé à chaque pas de temps (horaire) la différence relative entre le débit horaire et le débit journalier correspondant sur les périodes de hautes eaux (débit supérieur au quantile 0.95). La figure 5.3 représente cet écart relatif maximal sur une carte de France. Il n'est pas étonnant de constater que cet écart maximal est loin d'être négligeable pour un grand nombre de bassins situés dans le Sud de la France et les Cévennes... On retrouve d'ailleurs une corrélation (négative) entre cet écart et l'autocorrélation des débits.

Les hydrogrammes de crue aux pas de temps journalier et horaires peuvent donc être très dissemblables sur un grand nombre de bassins versants français et contenir des informations différentes, principalement sur les événements de crue. Il est très possible que cette différence d'information joue sur le fonctionnement du modèle et éventuellement sur ses performances : c'est ce que nous allons explorer maintenant.

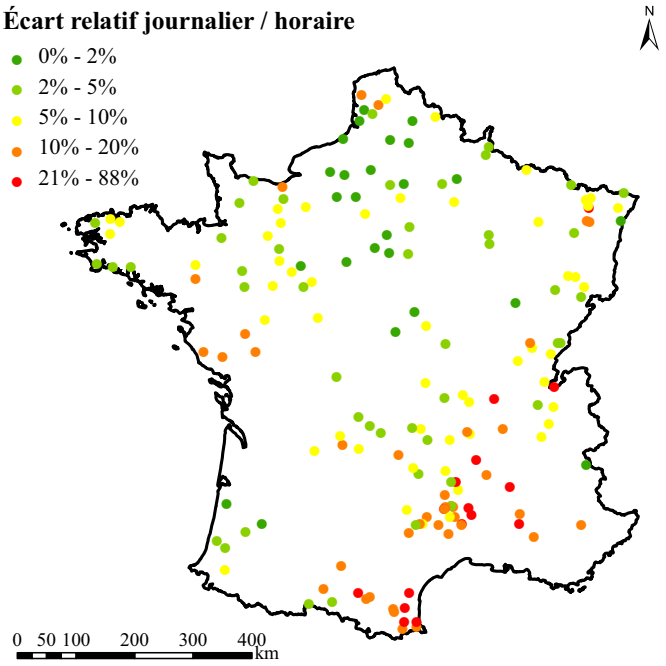


FIGURE 5.3 – Ecart relatif maximal de l'hydrogramme horaire à l'hydrogramme journalier sur 178 bassins versants français. Les écarts n'ont été calculés qu'en périodes de hautes eaux (au-dessus du quantile 0.95 du débit) sur une période de 10 ans (1995 – 2005).

5.2 Influence du pas de temps sur le fonctionnement et les performances du modèle

5.2.1 Influence sur le fonctionnement : optimisation des valeurs des paramètres fixes

Wang *et al.* (2009) ont montré que la valeur de certains paramètres d'un modèle conceptuel dépendaient de la résolution temporelle. Nalbantis (1995) avait d'ailleurs proposé des règles de passage d'un pas de temps à un autre. Nous nous intéressons ici à l'optimisation des paramètres fixes[(g)] pour établir et illustrer l'influence du pas de temps sur le fonctionnement du modèle. Les paramètres fixes sont des paramètres auxquels le modèle est suffisamment peu sensible pour que leurs valeurs puissent être fixées pour tous les bassins versants (éventuellement sur une zone géographique restreinte : la France en ce qui nous concerne) sans perte de performance. Le modèle GR3P [2] en possède cinq (Fig. 5.4, page suivante) :

- A, la capacité du réservoir de production (en mm),
- B, le coefficient de la fonction de percolation (sans dimension),
- α, l'exposant commandant la forme de l'hydrogramme unitaire (sans dimension),
- la loi de vidange du réservoir de routage,
- γ, l'exposant de la correction des débits, autorégressive multiplicative proposée par Tangara (sans dimension).

Comme nous le verrons, certains dépendent analytiquement du pas de temps.

Nous avons déterminé les valeurs optimales de ces paramètres fixes au pas de temps horaire pour l'échantillon de 178 bassins versants (voir également Tahir, 2008, pour l'évolution des valeurs de ces paramètres fixes à différents pas de temps compris entre 1 et 24 heures, dont les résultats sont reportés dans l'annexe C.4). Les performances du modèle avec différentes valeurs de ces paramètres fixes ont été calculées pour différents horizons de prévision (1, 6 et 24 heures). Les valeurs choisies tiennent compte du comportement du modèle à ces horizons. Les résultats sont présentés et discutés dans les paragraphes suivants.

Capacité du réservoir de production A

Le modèle est particulièrement peu sensible à la valeur de ce paramètre fixe (Tab. 5.1). Différentes valeurs comprises entre 100 et 900 mm ont été testées. Les différences de performance entre ces valeurs ne sont pas significatives. La « meilleure » valeur se situe entre 250 et 300 mm quand Tangara (communication personnelle) optimise ce paramètre à une valeur entre 300 et 350 mm au pas de temps journalier. La capacité de ce réservoir sera

2. La structure du modèle telle que proposée par Tangara (2005) est détaillée dans l'annexe C.2, page 340.

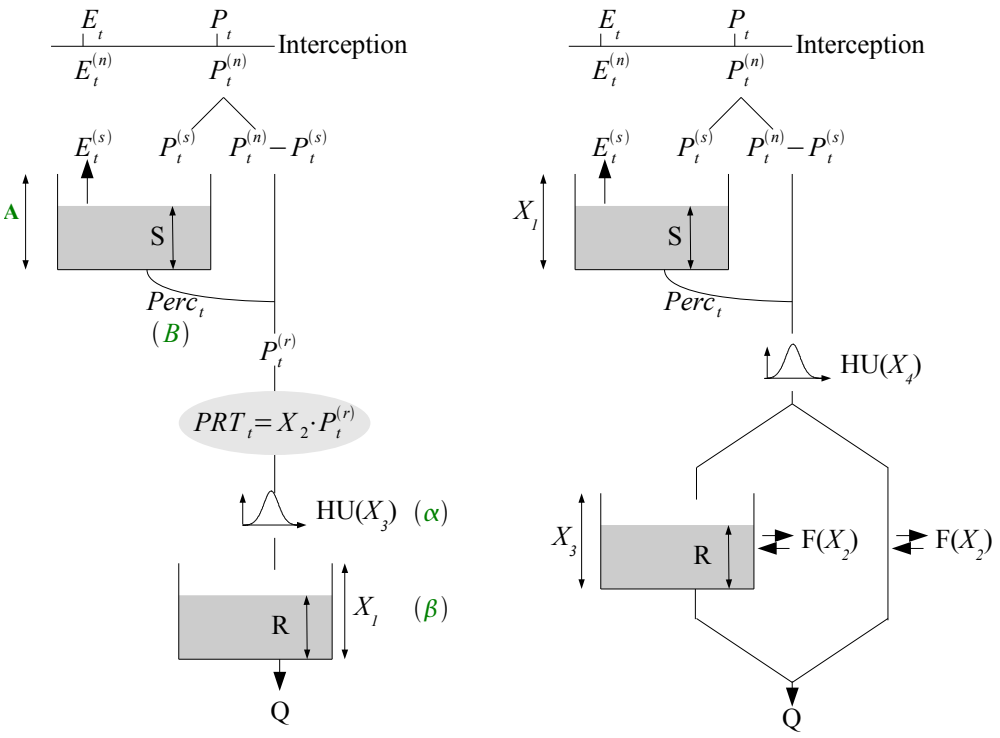


FIGURE 5.4 – Structures du modèle de prévision GR3P (à gauche) et du modèle de simulation GR4H (Mathevet, 2005, à droite) fonctionnant au pas de temps horaire. Sur le schéma de gauche, les paramètres fixes concernant la structure du modèle sont indiqués en vert.

fixée pour la suite à 275 mm au pas de temps horaire.

Cette insensibilité au pas de temps était attendue car le paramètre A qui intervient dans les équations décrivant le fonctionnement du réservoir de production sur un pas de temps est en fait le paramètre de la loi instantanée de comportement de ce réservoir (Ann. C.5, page 345) et ne dépend donc pas du pas de temps.

Coefficient de la fonction de percolation B

La percolation de la fonction de production des modèles GR (aux pas de temps journalier et horaire) est définie par la fonction de vidange d'un réservoir de type « puissance 5 ». Dans le cadre de cette définition, la percolation s'écrit (Mathevet, 2005, annexe 8, page 417 ; repris dans l'annexe C.2 de ce mémoire, page 340) :

$$Perc_t = S_t^+ \cdot \left(1 - \left(1 + \left(\frac{S_t^+}{K}\right)^4\right)^{-\frac{1}{4}}\right) \quad \text{avec } K = \left(k\,(\alpha - 1)\,\Delta t\right)^{\frac{1}{1-\alpha}}$$

où S_t^+ est le niveau du réservoir de production après incorporation (d'une partie) de l'eau précipitée au pas de temps t, k est le paramètre de la loi instantanée, Δt est le pas de

A	1 heure	6 heures	24 heures
100	0.225	0.250	0.376
200	0.225	0.256	0.390
225	0.225	0.257	0.391
250	**0.226**	**0.257**	0.392
275	0.226	0.2597	0.393
300	0.225	0.256	**0.393**
325	NR	NR	0.393
350	0.225	0.256	0.393
400	0.224	0.255	0.392
500	0.223	0.253	0.389
550	0.223	0.252	0.386
600	0.222	0.251	0.384
650	0.222	0.249	0.381
700	0.221	0.249	0.378
750	0.221	0.248	0.377
800	0.221	0.247	0.375
900	0.220	0.245	0.371

Table 5.1 – Performances moyennes ($C_{2M,P}$) sur 178 bassins de GR3P utilisé au pas de temps horaire pour différentes valeurs du paramètre A. La troisième décimale n'est donnée que pour illustrer la faiblesse des variations.

temps et α est l'exposant de la loi instantanée (ici égal à 5).

Dans la pratique, K est écrit sous la forme $K = A/B$ où B est un paramètre fixe. Le modèle est sensible à la valeur prise par B (Tab. 5.2). La valeur optimale au pas de temps de ce paramètre fixe est 5.25. Si la percolation fonctionne bien selon une loi de vidange de type puissance avec un exposant égal à 5, les valeurs de K au pas de temps horaire et journalier sont reliés par un facteur $24^{1/4}$. Connaissant les valeurs optimisées de A et B aux pas de temps horaires et journaliers, nous pouvons comparer les valeurs expérimentales de K :

$$\frac{K_H}{K_J} = \frac{275/5.25}{300/2.25} \simeq \frac{1}{2.56} \quad \text{à comparer avec} \quad 24^{1/4} \simeq 2.21$$

On retrouve l'ordre de grandeur attendu.

Exposant de l'hydrogramme unitaire α

L'exposant α commande la forme de l'hydrogramme unitaire : plus il est élevé, plus l'hydrogramme sera piqué. Ce paramètres fixe prend des valeurs positives ; la valeur limite 0 correspond à un hydrogramme parfaitement plat. Nous avons testé des valeurs de α allant de 1 à 5 (ainsi que la valeur limite 0). Les performances globales du modèle sont un peu sensibles à cette valeur (Tab. 5.3). L'ensemble des bassins se comporte de la même façon aux horizons de prévision courts. La valeur optimale de α ne dépend pas de l'horizon

B	1 heure	6 heures	24 heures
2.00	0.174	0.176	0.269
2.50	0.198	0.215	0.322
3.00	0.211	0.236	0.356
3.50	0.218	0.246	0.374
3.75	0.221	0.249	0.379
4.00	0.222	0.249	0.381
4.25	0.223	**0.250**	**0.383**
4.50	**0.224**	**0.250**	**0.383**
4.75	**0.224**	**0.250**	0.382
5.00	**0.224**	0.249	0.381
5.25	**0.224**	0.249	0.380
5.50	**0.224**	0.248	0.379

TABLE 5.2 – Performances moyennes ($C_{2M,P}$) sur 178 bassins de GR3P utilisé au pas de temps horaire pour différentes valeurs du coefficient de la fonction de percolation. La troisième décimale n'est donnée que pour illustrer la faiblesse des variations autour de l'optimum.

de prévision (dans la gamme testée). Comme chez Mathevet (2005) et Le Moine (2008), la valeur optimale obtenue au pas de temps horaire est de 1.25 : l'hydrogramme le plus performant dans ce modèle est plus arrondi au pas de temps horaire qu'au pas de temps journalier, conduisant logiquement à un étalement plus uniforme de la pluie efficace sur plus de pas de temps.

Régime du réservoir de routage

Différentes lois de vidange ont été testées pour le réservoir de routage : linéaire, puissance et exponentielle (Tab. 5.4). Un réservoir de type linéaire entraîne des difficultés de calage sur un certain nombre de bassins versants. Les performances globales sont moins bonnes avec ce type de réservoir (comme observé au pas de temps journalier par Tangara, 2005). Un réservoir exponentiel conduit également à des performances globales nettement inférieures. Cependant, un nombre non négligeable de bassins obtiennent leurs meilleures performances avec un réservoir linéaire ou exponentiel.

De façon générale, les meilleures performances sur notre échantillon de bassins sont obtenues avec un réservoir suivant une loi de vidange puissance avec un exposant β de l'ordre de 2 (horizons courts) à 2.5 (horizons plus longs).

Exposant de la correction des sorties γ

Le paramètre auquel le modèle est le plus sensible est l'exposant γ de la correction des sorties (Tab. 5.5). La valeur optimale de ce paramètre décroit avec l'horizon de prévision : elle est de 0.475 pour un horizon de 1 heure, 0.45 à 6 heures et 0.3 à 24 heures. Cette décroissance s'explique par le caractère moins informatif du dernier débit observé sur le

α	1 heure	6 heures	24 heures
0.00	0.181	0.231	0.376
1.00	0.228	0.257	0.383
1.10	0.230	0.257	0.383
1.20	0.230	0.258	0.383
1.25	**0.230**	**0.258**	**0.383**
1.30	0.230	0.258	0.383
1.40	0.230	0.257	0.383
1.50	0.230	0.257	0.383
2.00	0.225	0.253	0.382
2.25	0.224	0.251	0.381
2.50	0.222	0.249	0.381
2.75	0.219	0.248	0.381
3.00	0.217	0.246	0.381
3.50	0.213	0.243	0.380
4.00	0.209	0.241	0.380
5.00	0.203	0.238	0.379

Table 5.3 – Performances moyennes ($C_{2M,P}$) sur 178 bassins de GR3P utilisé au pas de temps horaire pour différentes valeurs de l'exposant de l'hydrogramme unitaire. La troisième décimale n'est donnée que pour illustrer la faiblesse des variations.

Loi	β	1 heure	6 heures	24 heures
Linéaire		0.165	0.182	0.317
	1.5	0.179	0.194	0.311
	2	**0.222**	**0.249**	0.382
	2.5	0.220	0.248	**0.384**
Puissance	3	0.212	0.245	0.383
	4	0.213	0.238	0.379
	5	0.210	0.232	0.376
Exponentiel		0.134	0.207	0.351

Table 5.4 – Performances moyennes ($C_{2M,P}$) sur 178 bassins de GR3P utilisé au pas de temps horaire pour différents réservoirs de routage (performances du modèle avec le réservoir de routage linéaire calculées pour 1033 bassins). La troisième décimale n'est donnée que pour illustrer la faiblesse des variations.

débit à venir, à mesure que l'horizon de prévision s'éloigne.

Aux horizons courts, la grande majorité des bassins obtiennent leurs meilleures perfor-

γ	1 heure	6 heures	24 heures
0	0.136	0.240	0.380
0.05	0.166	0.244	0.381
0.1	0.195	0.247	0.381
0.15	0.222	0.249	0.381
0.2	0.246	0.252	**0.382**
0.25	0.268	0.254	**0.382**
0.275	NR	NR	**0.382**
0.3	0.285	0.256	**0.382**
0.4	0.304	**0.258**	0.381
0.425	0.306	**0.258**	NR
0.45	**0.307**	**0.258**	NR
0.475	**0.307**	0.258	NR
0.5	0.306	0.258	0.381
0.6	0.286	0.254	0.379

TABLE 5.5 – Performances moyennes ($C_{2M,P}$) sur 178 bassins de GR3P utilisé au pas de temps horaire pour différentes valeurs de l'exposant de la correction du débit. La troisième décimale n'est donnée que pour illustrer la faiblesse des variations.

mances pour les valeurs optimales indiquées. La situation est plus nuancée aux horizons plus lointains. Comme nous le verrons par la suite, les résultats du modèle sont moins sensibles à cette correction des sorties aux horizons de prévisions lointains.

Libération des paramètres fixes

Nous avons cherché à « libérer » ces paramètres fixes, c'est-à-dire à donner au modèle des degrés de liberté supplémentaires en le laissant les caler. Nous avons libéré un paramètre fixe à la fois. Pour tous les paramètres fixes sauf l'exposant de la correction des sorties γ, le gain en performances obtenu lors de ces essais est plus que négligeable : moins d'un point de $C_{2M,P}$. Pour γ, le gain moyen sur l'ensemble des bassins est de deux points à un horizon d'une heure (Tab. 5.6) ; ce gain décroît à mesure que l'horizon de prévision s'éloigne.

γ	1 heure	6 heures	24 heures
Performances	0.328	0.268	0.382
Gain	0.021	0.001	0.000
Exposant de correction moyen	0.49	0.64	0.43

TABLE 5.6 – Performances moyennes de GR4P ($C_{2M,P}$) : l'exposant de la correction est un paramètre à caler.Le gain est défini par rapport au modèle GR3P avec un exposant γ de correction des sorties fixe.

Bilan : quelle sensibilité du modèle aux paramètres fixes ?
Nous avons pu constater dans cette étude que les paramètres fixes auxquels le modèle est le plus sensible sont ceux qui sont situés le plus en aval du modèle (et en interaction avec les mises-à-jour). Inversement, le modèle est particulièrement peu sensible aux paramètres fixes en amont comme la capacité du réservoir de production. Cette observation est un indice du rôle prépondérant joué par les procédures de mise-à-jour du modèle aux horizons courts. En présence de ces mises-à-jour (notamment de la mise-à-jour du réservoir de routage), la fonction de production du modèle aura un rôle amoindri et on peut s'attendre à une sensibilité plus faible du modèle à ses entrées (calcul de la pluie de bassin) et au niveau initial du réservoir de production qu'en mode « Simulation ». Une mise-à-jour des entrées ou du réservoir de production concomitante à une mise-à-jour du réservoir de routage ne devrait pas avoir une grande influence sur le modèle.

5.2.2 Quel pas de temps pour quel horizon ?

Les horizons de prévision qui nous sont accessibles sont limités par les précisions des prévisions quantitatives de précipitation (Habets *et al.*, 2004) et les temps de réaction des bassins. Vues les dynamiques de crue relevées dans la partie précédente, ces horizons vont de quelques heures à quelques jours.

Tahir (2008) a évalué les performances du modèle fonctionnant à différents pas de temps (de 1 à 24 heures) pour prévoir les débits à des horizons allant de 1 à 72 heures. Il observe que pour un pas de temps d'évaluation donné Δt_{Ev}, les meilleures performances sont obtenues pour le modèle fonctionnant à un pas de temps égal ou légèrement inférieur à Δt_{Ev} pour un scénario de pluies futures égales aux observations faites *a posteriori*. L'étude ne permet pas de conclure de façon claire quant à l'adaptation du pas de temps à l'horizon de prévision : il est préférable de travailler à la résolution temporelle la plus élevée dans la gamme des horizons testés, c'est-à-dire au pas de temps horaire.

Comme la plupart des services de prévision des crues (si ce n'est pas tous) disposent de données de pluie et de débit au pas de temps horaire (voire à des pas de temps plus fins pour certaines rivières), il est pertinent de développer un modèle de prévision fonctionnant au pas de temps horaire.

5.3 Bilan

Nous avons constaté que les hydrogrammes de crue au pas de temps horaire apportent une information nettement différente de celle contenue dans les mêmes hydrogrammes au pas de temps journalier pour la plupart des bassins versants de notre étude.

Cette différence de contenu influe sur le fonctionnement du modèle et sur ses performances.

Une analyse de sensibilité des performances au pas de temps (en fonction de l'horizon de prévision) a confirmé qu'il est préférable de travailler au pas de temps le plus fin.

Par ailleurs, nous avons mis en évidence que les valeurs de certains paramètres fixes du modèle journalier doivent être adaptées au pas de temps horaire, confirmant ainsi des résultats obtenus sur les modèles de simulation. Cette adaptation conduit à des améliorations, certes limitées, mais non négligeables des performances du modèle.

Il est probable que le pas de temps influe sur la structure du modèle elle-même. Mouelhi (2003) a réalisé une étude de l'influence du pas de temps sur les modèles GR de simulation, pour les pas de temps pluri-annuel à journalier, qui va dans ce sens. La première étape de notre travail consiste donc à passer d'un modèle de prévision (i.e. en tenant compte de ses mises-à-jour) au pas de temps journalier à un modèle au pas de temps horaire : dans le prochain chapitre, nous examinerons si la structure proposée par Tangara (2005) au pas de temps journalier peut être améliorée pour le pas de temps horaire.

6
Chapitre

Adapter la structure du modèle pour le pas de temps horaire

Il semble que la perfection soit atteinte non quand il n'y a plus rien à ajouter, mais quand il n'y a plus rien à retrancher.

Antoine de Saint-Exupéry, *Terre des hommes*

Objectifs

- ▷ Proposer et tester différentes modifications de la structure de GRP au pas de temps horaire.
- ▷ Évaluer si la meilleure structure en prévision est également celle obtenue en simulation.
- ▷ Vérifier le renforcement de l'influence des éléments les plus en aval du modèle, par les mises-à-jour.

Sommaire

6.1 Modification des fonctions non conservatives 91
6.1.1 Différences entre modèles GR de prévision et de simulation . . . 92
6.1.2 Tests réalisés 92
6.1.3 Résultats 93
6.2 Distribution de la fonction de production 95
6.3 Modifications de la forme de l'hydrogramme unitaire 95
6.3.1 Asymétrisation de l'hydrogramme unitaire 96
6.3.2 Introduction d'un décalage temporel 98
6.4 Ajout d'une seconde branche d'écoulement au routage 98
6.4.1 Première étape : répartition fixe des flux 100
6.4.2 Deuxième étape : répartition calée des flux 100
6.4.3 Troisième étape : répartition variable des flux 100
6.5 Bilan 102

Nous avons vu au chapitre précédent que l'information contenue dans les séries temporelles des pluies et des débits est fonction du pas de temps. En conséquence, le pas de temps influence le fonctionnement du modèle et la structure optimale dépend du pas de temps (Mouelhi, 2003). Dans ce chapitre, nous cherchons à optimiser la structure du modèle GRP pour le pas de temps horaire, en partant de la structure proposée par Tangara (2005) qui est détaillée en annexe C.2 (pages 340 et suivantes).

Comme nous travaillons dans le cadre de la prévision, les modifications de la structure que nous proposerons, seront testées en mode prévision, c'est-à-dire avec les mises-à-jour du modèle (il s'agira ici de celles définies par Tangara). Au pas de temps journalier, ce dernier avait constaté que la structure lui permettant d'obtenir les meilleures performances en prévision différait de la structure du modèle fonctionnant en simulation au même pas de temps. Est-ce toujours le cas au pas de temps horaire ? Nous allons tester des structures se rapprochant de celle du modèle de simulation fonctionnant au pas de temps horaire (GR4H de Mathevet, 2005, ou GR5H de Le Moine, 2008).

Les résultats du chapitre précédent indiquent que les mises-à-jour, basées sur l'assimilation du débit et situées très en aval du modèle, rendent les sorties de ce dernier très dépendantes des états situés en aval et peu sensibles aux composants plus en amont du modèle (comme la fonction de production). Nous chercherons néanmoins à modifier les différentes parties du modèle :

- au niveau de la fonction de production :
  - modification des fonctions non conservatives,
  - distribution de la fonction de production,
- au niveau de la fonction de routage :
  - modification de la forme de l'hydrogramme unitaire,
  - ajout d'une seconde branche au routage.

6.1 Modification des fonctions non conservatives

Les entrées et sorties du modèle GRP sont la pluie P, l'évapotranspiration potentielle E sur le bassin versant (de surface) et le débit Q à l'exutoire. Il est connu qu'on ne peut pas espérer fermer le bilan en eau à l'aide de ces seules variables : l'évapotranspiration réelle doit remplacer l'ETP, les échanges entre bassins hydrogéologiques doivent être pris en compte (Le Moine, 2008)... C'est pourquoi le modèle GRP, comme les autres modèles GR (aux pas de temps annuel, mensuel, journalier et horaire), comporte des fonctions non conservatives qui sur le long terme permettent d'obtenir un débit cumulé bien différent de la somme des différences entre pluie et évapotranspiration potentielle.

Les fonctions non conservatives de GRP sont très différentes de celles présentes dans les mo-

dèles de simulation GR4J, GR4H ou GR5H. Voyons si nous pouvons tirer une amélioration de GRP de la confrontation entre modèles de prévision et de simulation.

6.1.1 Différences entre les fonctions non conservatives des modèles GR de prévision et de simulation

Le modèle de simulation au pas de temps journalier présente une *fonction d'échange* censée représenter les transferts entre bassins (de surface). Elle dépend du niveau R du réservoir de routage :

$$F = X_2 \cdot \left(\frac{R}{X_3}\right)^{\frac{7}{2}} \quad \text{où } X_3 \text{ est la capacité du réservoir.} \tag{6.1}$$

La structure du modèle GR3P (conçu pour le même pas de temps) ne contient qu'une opération non conservative : l'ajustement multiplicatif de la pluie efficace pour donner le flux entrant dans la fonction de routage $PRT = X_2 \cdot P^{(r)}$. Ces fonctions non conservatives sont très différentes tant sur le plan de leurs formulations que pour leurs niveaux d'action au sein de la structure du modèle.

Le modèle de prévision est construit en tenant compte des opérations de mise-à-jour. Ces dernières sont par nature non conservatives. Elles peuvent donc expliquer pourquoi les solutions empiriques obtenues par Perrin (2000) en simulation et Tangara (2005) en prévision diffèrent.

6.1.2 Tests réalisés

Le Moine (2008) a montré qu'une formule linéaire à deux paramètres permettait d'obtenir de meilleurs résultats en simulation au pas de temps horaire :

$$F = X_a \cdot \left(\frac{R}{X_3} - X_b\right) \tag{6.2}$$

Cette formule peut dans certains cas (et avec toutes les réserves habituelles à ce type d'interprétation) être reliée à des flux darcéens (échanges hydrogéologiques). De fait, cette expression des échanges améliore significativement la simulation des étiages, ce qui n'est pas directement le plus utile pour la prévision des crues. Est-il intéressant d'introduire ce type de fonctions non conservatives pour la prévision au pas de temps horaire ?

Nous avons testé différentes combinaisons de fonctions non conservatives en faisant varier leurs emplacements au sein du modèle (correction de la pluie efficace à la fin de la fonction de production et/ou fonction d'échange lors du routage) et leurs formulations. Dans tous les cas, le niveau du réservoir de routage est mis-à-jour de la même façon que dans GR3P, ce qui peut fortement interagir avec les fonctions d'échange basées sur ce niveau. Le tableau 6.1 présente les variantes testées. Certaines variantes (NoEch2 et NoEch3) visent à relier l'ajustement de la pluie efficace au niveau du réservoir de routage (mémoire courte

du système) quand d'autres (NoEch4 et NoEch5) supposent cet ajustement fonction de l'intensité des précipitations. Enfin, un essai non mentionné dans le tableau a consisté en une correction multiplicative de la pluie (ou de l'ETP) en entrée du modèle et non de la pluie efficace.

Variante	Ajustement de la pluie efficace	Fonction d'échange
NoEch 0	Aucune	Aucune
NoEch 1	$PRT = X_2 \cdot P^{(r)}$	Aucune
NoEch 2	$PRT = max\left(0, \left(1 + X_2 \cdot \frac{R}{X_1}\right) \cdot P^{(r)}\right)$	Aucune
NoEch 3	$PRT = max\left(0, \left(1 + X_2 \cdot \frac{R}{X_1}\right)^2 \cdot P^{(r)}\right)$	Aucune
NoEch 4	$PRT = max\left(0, \left(1 + X_2 \cdot \frac{P^{(r)}}{\overline{P}}\right) \cdot P^{(r)}\right)$	Aucune
NoEch 5	$PRT = \begin{cases} P^{(r)} & \text{pour } P^{(r)} < X_2 \cdot \overline{P} \\ \alpha \cdot P^{(r)} & \text{sinon} \end{cases}$	Aucune
Ech 0	Aucune	$F(R) = X_4 \cdot \left(\frac{R}{X_1} - X_5\right)$
Ech 1	$PRT = X_2 \cdot P^{(r)}$	$F(R) = X_4 \cdot \left(\frac{R}{X_1} - X_5\right)$
Ech 2	$PRT = max\left(0, \left(1 + X_2 \cdot \frac{R}{X_1}\right) \cdot P^{(r)}\right)$	$F(R) = X_4 \cdot \left(\frac{R}{X_1} - X_5\right)$

Table 6.1 – Les différentes combinaisons de fonctions non conservatives testées. La variante NoEch 5 a été testée pour différentes valeurs du paramètre fixe α allant de 0.72 à 1.15.

6.1.3 Résultats

La version actuelle de GRP (NoEch 1) est la solution qui présente globalement (pour les 178 bassins versants testés et en considérant différents horizons de prévision) les meilleures performances. Les moyennes des critères de performance sur l'échantillon de bassins sont données dans le tableau 6.2.

L'examen des distributions montre que sur certains bassins, la variante Ech 0 (correspondant au modèle de simulation GR5H) obtient des performances significativement supérieures. Nous avons cherché à caractériser ces bassins. Notons d'abord qu'ils sont répartis sur tout le territoire. Comme la formulation (6.2) a permis d'obtenir de meilleurs résultats en simulation sur les périodes d'étiage principalement, nous avons examiné si ce groupe de bassins se caractérisait par un régime hydrologique marqué par des réactions lentes. C'est tout le contraire : ces bassins présentent des débits dont l'autocorrélation est inférieure à la moyenne de celles des débits des bassins de l'échantillon complet. De même, leurs indices de débit de base (BFI) sont inférieurs à ceux de l'échantillon complet. Ces bassins ne sont

Horizons	1 h	3 h	6 h	12 h
No Ech 0	0.36	0.25	0.26	0.31
No Ech 1	*0.37*	*0.27*	*0.29*	*0.36*
No Ech 2	0.36	0.25	0.26	0.31
No Ech 3	0.36	0.25	0.26	0.31
No Ech 4	0.35	0.24	0.26	0.32
No Ech5 - 0.72	0.36	0.25	0.27	0.32
No Ech5 - 0.77	0.36	0.25	0.27	0.32
No Ech5 - 0.82	0.36	0.25	0.27	0.32
No Ech5 - 0.87	0.36	0.25	0.27	0.32
No Ech5 - 0.92	0.36	0.25	0.26	0.31
No Ech5 - 0.95	0.36	0.25	0.26	0.31
No Ech5 - 0.98	0.36	0.25	0.26	0.31
No Ech5 - 1.02	0.36	0.25	0.26	0.31
No Ech5 - 1.05	0.36	0.25	0.26	0.31
No Ech5 - 1.08	0.36	0.25	0.25	0.30
No Ech5 - 1.10	0.36	0.25	0.25	0.30
No Ech5 - 1.12	0.36	0.24	0.25	0.30
No Ech5 - 1.15	0.36	0.24	0.25	0.30
Ech 0	0.37	0.26	0.27	0.33
Ech 1	0.34	0.21	0.22	0.27
Ech 2	0.37	0.26	0.27	0.32

TABLE 6.2 – Performances moyennes ($C_{2M,P}$) des différents essais menés sur des fonctions non conservatives du modèle de prévision GRP au pas de temps horaire. La ligne en italique indique la variante initiale (GR3P).

donc pas des bassins particulièrement lents ou marqués par des débits de base élevés : le gain de performance sur ces bassins ne peut pas s'expliquer ainsi dans notre étude (en prévision).

Combiner deux fonctions non conservatives, l'une ajustant la pluie efficace, l'autre fonction du niveau du réservoir de routage (variantes Ech 1 et Ech 2) s'est révélé peu robuste : cela a conduit aux plus mauvaises performances, possiblement en raison d'un surparamétrage.

L'option d'une correction multiplicative des pluies en entrée est très légèrement moins performante que celle sur les pluies efficaces. On note cependant dans ce cas, que pour la très grande majorité des bassins versants, le facteur multiplicatif (X_2) est plus proche de 1 (sa valeur neutre). Quant à la correction multiplicative de l'ETP, elle donne des résultats significativement moins bons aux grands horizons de prévision (et légèrement moins bons pour les horizons de prévision les plus courts).

En conclusion, les différents essais menés n'ont pas permis de dégager une formulation des fonctions non conservatives de GRP conduisant à de meilleures performances.

6.2 Distribution de la fonction de production

La sensibilité des modèles hydrologiques (de prévision comme de simulation) à la donnée de pluie est une question abondamment discutée dans la littérature. Comme nous le verrons au chapitre 15 étudiant la robustesse du modèle face à des conditions dégradées, la qualité de la donnée de pluie revêt une forte importance. Les différentes sources de données (pluviomètres au sol, estimations radar, données satellitales, etc.) ont fait et font toujours l'objet de comparaison en ce qui concerne leur impact sur les prévisions hydrologiques (Anquetin *et al.*, 2005, Berthet, 2006, Werner et Cranston, 2009, Valeriano *et al.*, 2009, parmi beaucoup d'autres) sans qu'il soit toujours possible de distinguer clairement une meilleure source de données.

La variabilité spatiale des pluies est également souvent questionnée. Bourqui (2008) analyse par exemple la sensibilité d'un modèle hydrologique global à la variabilité spatiale des pluies sans noter de fort impact. Cependant, un des arguments phares des partisans des modèles distribués est lié à la variabilité spatiale des pluies et de la prise en compte de l'humidité du sol (Bergström et Graham, 1998).

Nous avons mené une étude sur la distribution de la fonction de production du modèle GRP qui évalue l'influence de la prise en compte de la distribution spatiale des précipitations sur la fonction de production du modèle, puis sur les performances en prévision. Ce travail, présenté dans l'annexe I.2 (page 491), a été mené avec deux types d'estimation des pluies qui diffèrent par leurs variabilités spatiales : les pluies Safran et celles mesurées par le réseau de pluviomètres fonctionnant au pas de temps horaire.

Cette étude montre que sur l'immense majorité des pas de temps, les flux simulés en sortie des fonctions de production globales et distribuées sont très proches. L'impact sur la qualité des prévisions est négligeable en dehors d'un petit nombre de bassins versants, situés sur l'arc méditerranéen.

Il serait intéressant d'étudier une distribution du modèle plus complète incluant le routage. Ceci renvoie aux travaux menés en simulation par Lerat (2009) ou aux premiers essais en prévision présentés par Munier (2009). Cependant, cette approche demande au préalable d'avoir traité la question de la mise-à-jour d'un modèle distribué par l'assimilation d'une donnée « intégrée » : le débit à l'exutoire du bassin. Lamblin (2009) a menée une étude exploratoire sur la prévision aux confluences, réelles ou chimériques (Andréassian *et al.*, 2004), mais le sujet reste à approfondir.

6.3 Modifications de la forme de l'hydrogramme unitaire

Le Moine (2008) a mené plusieurs essais de modification de la forme de l'hydrogramme unitaire (HU) en simulation au pas de temps horaire allant au delà de la seule modification

de l'exposant de la formulation de l'hydrogramme unitaire. Ses essais n'ont pas montré une grande sensibilité des performances du modèle à l'HU. Cependant, l'optimisation du paramètre fixe α commandant la forme de l'HU, réalisée au chapitre précédent, montre que cette sensibilité n'est pas totalement négligeable. Nous testons donc ici deux modifications : le premier essai porte sur une asymétrisation de la forme de l'hydrogramme unitaire tandis que le second introduit un décalage temporel.

6.3.1 Asymétrisation de l'hydrogramme unitaire

Plusieurs formes d'hydrogramme unitaire de GRP ont été testées sur un petit nombre de bassins versants sans prouver un grand intérêt (Berthet, 2006). Nous reprenons ici l'essai qui nous paraît *a priori* le plus intéressant : l'asymétrisation de la forme de l'HU qui répond à l'intuition « physique ».

Nous introduisons un paramètre κ dans la fonction de l'hydrogramme unitaire cumulé (voir l'annexe C.2, en page 343) permettant d'asymétriser l'hydrogramme unitaire (Fig. 6.1) :

$$\begin{cases} UHC(i)=0 & \text{si } i \leq 0 \\ UHC(i)=\dfrac{i^{\alpha}}{i^{\alpha}+\kappa \cdot (X_3-i)^{\alpha}} & \text{si } 0 < i < X_3 \\ UHC(i)=1 & \text{si } i \geq X_3 \end{cases} \tag{6.3}$$

Dans GR3P, κ est égal à 1, ce qui correspond à un hydrogramme unitaire symétrique.

Deux possibilités s'ouvrent à nous : laisser κ libre (introduisant ainsi un quatrième paramètre à caler) ou lui assigner une valeur fixe.

Quand κ est calé par le modèle, la médiane et la moyenne de ses valeurs sont très proches de 1 (entre 0.97 et 1) aux trois horizons de prévisions testés (1, 6 et 12 heures). Cependant, ce constat cache que le paramètre est souvent calé à l'une des extrémités de la gamme autorisée ($[0.4, 1.4]$) : 10 % environ des calages aboutissent à une valeur de κ égale à 0.4 tandis que 25 % environ conduisent à la borne supérieure 1.4 (qui correspond à une forme contre-intuitive de l'hydrogramme unitaire). Si les valeurs calées aux bornes sont exclues de l'analyse, la moyenne et la médiane des distributions se situent entre 0.85 (horizon d'une heure) à 0.95 (horizon de 12 heures). Enfin, on n'observe pas de lien entre la valeur obtenue au calage (notamment en butée) et les performances du modèle. La différence des performances obtenues par cette variante du modèle et GRP n'est pas significative.

Quand κ est fixé, la différence de performance n'est pas plus significative (Tab. 6.3). On observe que les « meilleures » performances sont obtenues pour des valeurs de κ inférieures à 1 tout en restant proche de cette valeur (HU symétrique). Complexifier la forme de l'hydrogramme unitaire par cette approche n'apporte pas d'amélioration significative.

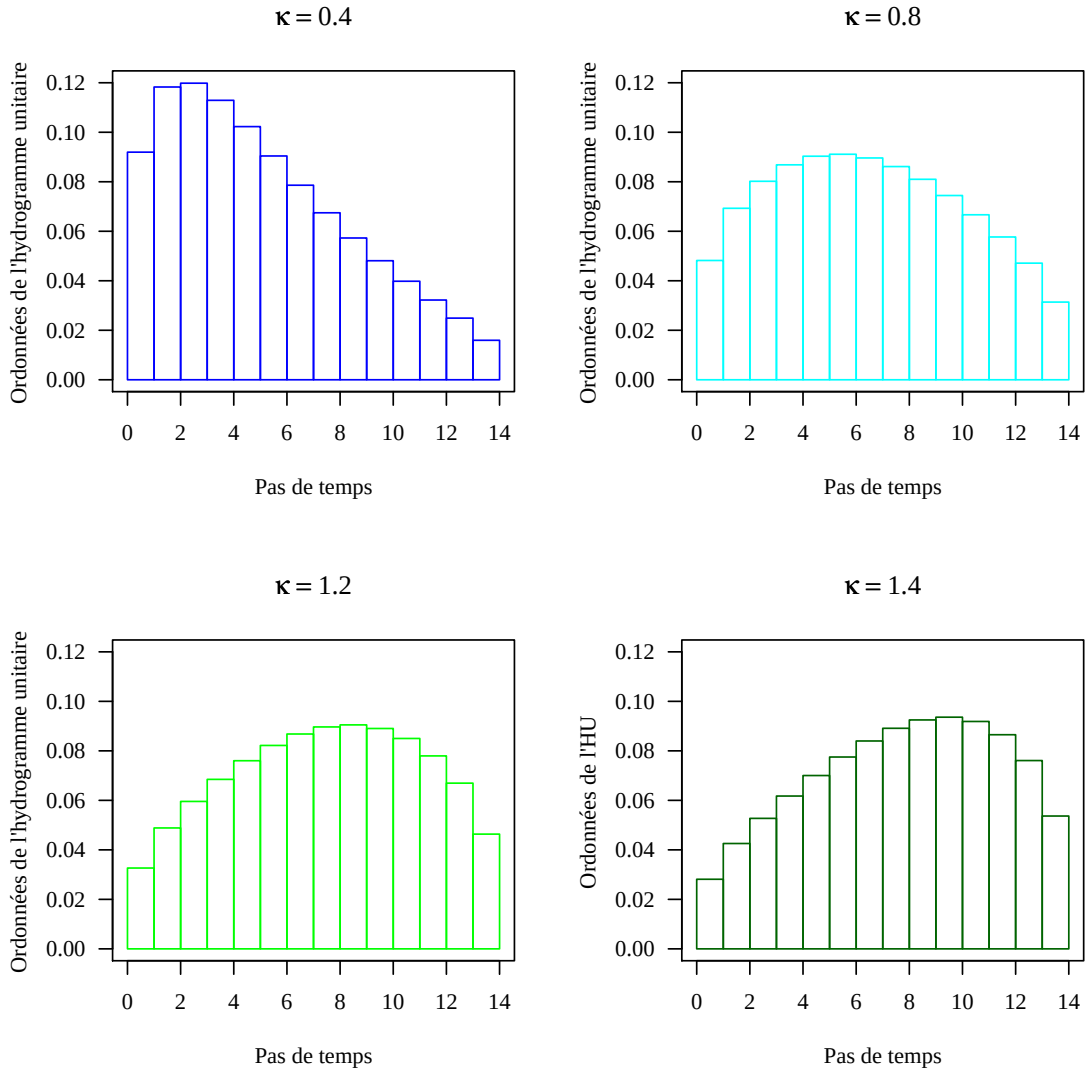


Figure 6.1 – Hydrogrammes asymétrisés par la formule (6.3). Ces hydrogrammes ont été calculés pour un temps de base égal à 14 pas de temps.

	κ	κ fixe								
Horizon	*libre*	0.4	0.5	0.6	0.7	0.8	0.9	*1.0*	1.1	1.2
1 heure	*0.407*	0.408	0.408	0.407	0.409	0.408	**0.411**	*0.406*	0.409	0.407
6 heures	*0.292*	0.291	0.291	0.291	0.293	**0.294**	0.291	*0.290*	0.293	0.292
12 heures	***0.375***	0.372	0.374	0.374	0.374	0.374	**0.375**	*0.374*	0.374	0.374

Table 6.3 – Médianes des distributions (sur 178 bassins) de performance ($C_{2M,P}$) obtenues par le modèle avec un hydrogramme unitaire asymétrique. La colonne en italique indique la variante initiale (GR3P). La troisième décimale non significative n'est indiquée que pour rendre compte de la faiblesse des différences entre les variantes testées.

6.3.2 Introduction d'un décalage temporel

Au pas de temps horaire, on peut supposer que la pluie n'a pas (ou peu) atteint l'exutoire de certains bassins lors des premiers pas de temps suivant la précipitation. Aussi, la(es) première(s) ordonnées de l'hydrogramme unitaire devrai(en)t être nulles. Nous proposons donc de tester l'introduction d'un décalage temporel au niveau de l'entrée dans l'hydrogramme unitaire. Ici encore, nous avons deux possibilités : laisser le modèle caler la durée de ce décalage ou fixer une durée pour tous les bassins versants. Dans le second cas, nous testons des valeurs multiples du temps de base X_3 de l'hydrogramme unitaire (le facteur allant de 0.1 à 1.5).

Si la durée de ce décalage est laissée libre au calage, elle se fixe majoritairement à des durées relativement courtes : la médiane de la durée est de l'ordre de 3 heures pour les trois horizons de prévision testés (1, 6 et 12 heures). Pour 88 % des bassins versants, les performances sont dégradées à un horizon d'une heure (80 % à 6 heures et 71 % à 12 heures). Les performances sont (un peu) améliorées sur un tout petit nombre de bassins pour lesquels le modèle obtenait des performances très faibles (caractérisées par un critère de persistance négatif). Fixer la durée du décalage libre à une fraction du temps de base de l'hydrogramme unitaire dégrade encore plus les performances du modèle (Tab. 6.4).
Ces résultats sont cohérents avec le changement de forme de l'hydrogramme unitaire observé lors du passage du pas de temps journalier vers le pas de temps horaire : l'hydrogramme prend une forme plus arrondie. Ainsi non seulement, l'hydrogramme « étale » la pluie efficace sur plus de pas de temps (qu'au pas de temps journalier), mais il le fait dès le premier pas de temps.

	Aucun	Décalage fixe								Décalage
Horizon	*décalage*	0.1	0.2	0.3	0.5	0.8	1.0	1.2	1.5	libre
1 heure	*0.41*	0.16	0.16	0.16	0.15	0.14	0.13	0.13	0.12	0.31
6 heures	*0.29*	0.27	0.27	0.26	0.26	0.25	0.25	0.24	0.24	0.25
12 heures	*0.37*	0.35	0.35	0.35	0.34	0.34	0.34	0.33	0.33	0.33

TABLE 6.4 – Distribution des performances du modèle GRP quand un décalage temporel est introduit « en amont » de l'hydrogramme unitaire. La ligne en italique indique la variante initiale (GR3P).

6.4 Ajout d'une seconde branche d'écoulement au routage

Les modèles de simulation GR aux pas de temps journalier (Perrin, 2000) et horaire (Mathevet, 2005, Le Moine, 2008) présentent une fonction de routage composée de deux branches d'écoulement parallèles séparant la pluie efficace PRT en un flux rapide et en un flux lent (Fig. 5.4). *A contrario*, la solution préconisée par Tangara (2005) en prévision au pas de

temps journalier ne compte qu'un seul flux. Or l'existence d'un flux rapide peut paraître souhaitable *a priori* au pas de temps horaire. Nous testons donc ici une variante du modèle de prévision dont la structure est beaucoup plus proche de celle du modèle de simulation GR4H.

Nous avons travaillé sur deux structures. La première, dite *directe*, se caractérise par une branche de routage conduisant directement une fraction $1 - \lambda$ de la pluie efficace vers le débit calculé (Fig. 6.2). La deuxième branche de la fonction de routage de la seconde structure, appelée *semi-directe*, est un hydrogramme unitaire qui ralentit cette fraction de la pluie efficace.

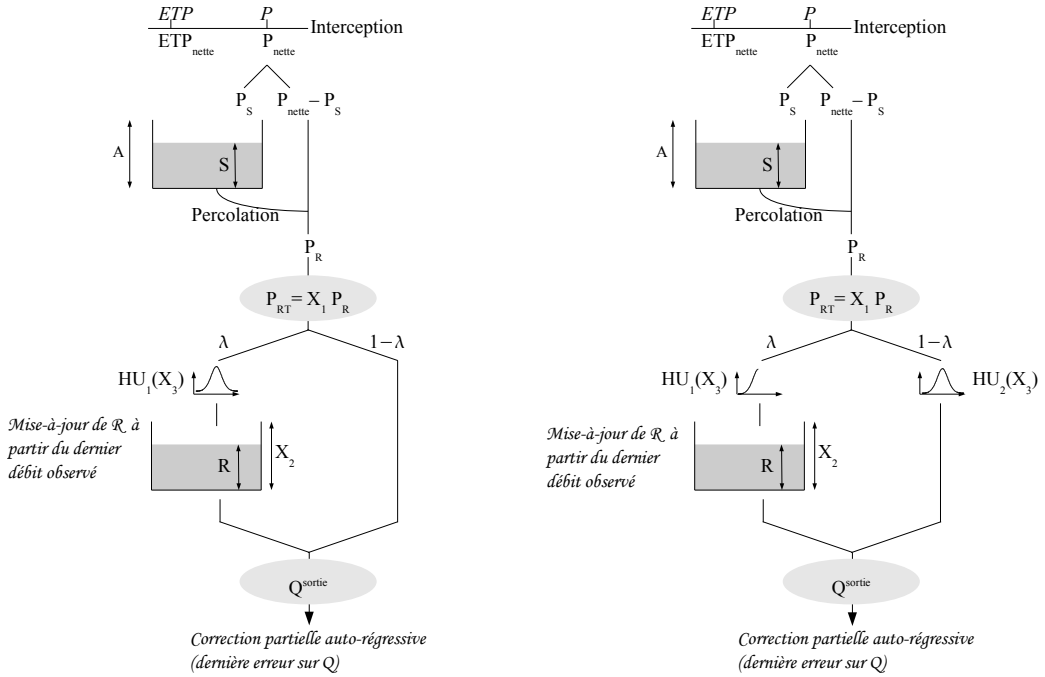


Figure 6.2 – Structures des variantes testées dont la fonction de routage contient une seconde branche d'écoulement : directe (à gauche) ou semi-directe (à droite).

Dans cette configuration à deux branches, seul le niveau du réservoir de routage (de la branche lente) sera mis-à-jour. Pour ce faire, il suffit de déterminer le flux passant au pas de temps t dans la branche rapide et par soustraction, le débit attendu en sortie du réservoir de routage. On en déduit alors le niveau du réservoir mis-à-jour.

Pour tester de façon le plus exhaustive possible l'intérêt de l'ajout d'une seconde branche dans la fonction de routage, nous allons procéder en trois étapes en laissant le modèle de plus en plus libre de diriger la pluie efficace vers la seconde branche qui est un écoulement direct.

6.4.1 Première étape : répartition fixe des flux

Dans ce premier essai, le facteur λ répartissant la pluie efficace vers la branche lente et vers la branche rapide de la fonction de routage est fixe et commun à tous les bassins. Nous l'avons fixé à 0.9 (comme dans les modèles GR4J et GR4H).

La distribution des performances du modèle semi-direct est très proche de celle des performances du modèle GRP monobranche à tous les horizons de prévision (les médianes de ces distributions sont données dans le tableau 6.5, 3^{e} colonne). Ce n'est pas du tout le cas pour la structure directe pour laquelle on observe des chutes de performance significatives (5^{e} colonne du même tableau). Ce résultat confirme ceux de travaux antérieurs sur les modèles GR de simulation aux pas de temps journalier et horaire qui n'ont jamais réussi à améliorer les performances en introduisant un écoulement direct.

6.4.2 Deuxième étape : répartition calée des flux

Pour tenter de remédier à la situation précédente, nous décidons ici de laisser le calage décider (pour chaque bassin) la valeur du facteur λ.

Quand λ est calé, les distributions des performances obtenues par les deux structures testées sont quasiment identiques à celles du modèle GRP (4^{e} et 6^{e} colonnes du tableau 6.5). Les distributions (pour les 178 bassins) des valeurs du paramètre λ ont été étudiées (non montrées ici). Le mode des distributions est proche de 1 pour les deux structures testées. Cependant, la distribution est beaucoup plus « piquée » en 1 pour la structure directe que pour la structure semi-directe. On retrouve ici un résultat obtenu dès le premier test : le modèle laissé libre d'employer ou non une seconde branche de la fonction de routage, refuse catégoriquement de le faire dans le cas d'une branche directe. Par contre, il l'emploie pour faire transiter une petite quantité d'eau s'il s'agit d'une branche semi-directe (exactement comme dans GR4J par exemple).

6.4.3 Troisième étape : répartition variable des flux

Dans un dernier essai, nous imaginons que si la plupart du temps, ajouter une seconde branche à la fonction de routage n'améliore pas les performances du modèle, il se peut que sur certains pas de temps, cela puisse être bénéfique. Nous allons donc mettre à jour le facteur λ en fonction du dernier débit observé : on recherche la valeur de λ comprise entre 0 et 1 permettant d'expliquer le dernier débit observé. Cette approche est bien entendu compatible avec la mise-à-jour classique du niveau du réservoir de routage de GRP.

L'examen des distributions de performance (dont les médianes sont données dans le tableau 6.6) indique de façon générale que cette procédure détériore les performances du modèle dans tous les cas. Comme on pouvait s'y attendre, la structure directe réalise les plus mauvaises performances. Nous avons testé deux cas : dans le premier, les paramètres du

Structure	*Monobranche*	semi-directe		directe	
Répartition λ	–	fixe	calée	fixe	calée
1 heure	*0.41*	0.39	0.41	0.30	0.40
6 heures	*0.29*	0.29	0.29	0.05	0.29
12 heures	*0.37*	0.37	0.37	0.03	0.37
24 heures	*0.47*	0.46	0.46	0.01	0.45
48 heures	*0.54*	0.54	0.54	0.01	0.54

Table 6.5 – Médiane des performances (mesurées en $C_{2M,P}$) des deux premiers essais de rajout d'une seconde branche à la fonction de routage de GRP pour différents horizons de prévision. Deux structures, l'une semi-directe, l'autre directe, sont testées en laissant fixe ou en calant le facteur de répartition de la pluie efficace entre les deux branches de la fonction de routage. La colonne en italique indique les performances de la variante initiale (GR3P).

modèle sont calés avec un facteur de répartition de la pluie efficace λ fixe ; dans le second cas, la mise-à-jour de λ est réalisée même au calage. Les performances sont supérieures quand λ est mis-à-jour durant le calage sans toutefois atteindre les performances obtenues par le modèle GRP monobranche. Cela confirme les observations de Tangara (2005) sur l'importance de caler le modèle dans ses conditions d'emploi en prévision, c'est-à-dire en intégrant ses mises-à-jour.

Structure	*Monobranche*	semi-directe		directe	
λ mis-à-jour au calage	–	oui	non	oui	non
1 heure	*0.41*	0.35	0.32	0.15	-0.89
6 heures	*0.29*	0.09	0.07	0.18	-0.23
12 heures	*0.37*	0.08	0.06	0.25	0.01
24 heures	*0.47*	0.08	0.07	0.33	0.12
48 heures	*0.54*	0.11	0.10	0.38	0.17

Table 6.6 – Médiane des performances (mesurées en $C_{2M,P}$) de l'essai de rajout d'une deuxième branche à la fonction de routage de GRP avec mise-à-jour du facteur de répartition λ de la pluie efficace entre les deux branches pour différents horizons de prévision. Deux structures, l'une semi-directe, l'autre directe sont testées. La mise-à-jour de λ est testée uniquement en contrôle ou lors du calage et du contrôle. La colonne en italique indique les performances de la variante initiale (GR3P).

Au final, les fonctions de routage testées comportant deux branches d'écoulement n'améliorent pas les performances du modèle, quelque soit la manière de répartir la pluie efficace entre ces deux branches. Nous conservons donc pour le pas de temps horaire la fonction de routage proposée par Tangara (2005).

6.5 Bilan

De nombreuses variantes du modèle GRP ont été testées. Au final, c'est la structure initiale, optimisée par Tangara (2005) pour le pas de temps journalier, qui permet d'obtenir les meilleures performances de prévision. On peut noter un parallèle avec la modélisation globale dédiée à la simulation où les différences entre GR4H ou GR5H (au pas de temps horaire) et GR4J au pas de temps journalier sont plutôt mineures, en comparaison avec les structures optimisées pour des pas de temps plus larges.

Parmi les variantes proposées, nombreuses sont celles qui s'inspiraient du modèle GR de simulation au pas de temps horaire. Employer la structure des modèles GR de simulation ne permet pas d'améliorer les performances en prévision. Comme au pas de temps journalier, la présence de mises-à-jour du modèle influence sa structure.

Il semble qu'il n'y ait pas beaucoup de gains de performance à attendre d'autres modifications de la structure du modèle. Nous allons donc nous concentrer dans la prochaine partie de ce mémoire, à l'amélioration de la mise-à-jour du modèle par assimilation des observations de débit.

Troisième partie

Mises-à-jour du modèle

7

Chapitre

Mise-à-jour des états

En matière d'État, il faut tirer profit de toutes choses, et ce qui peut être utile ne doit jamais être méprisé.

Cardinal de Richelieu *Maximes d'État*

Objectifs

- ▷ Comparer plusieurs stratégies de mise-à-jour des états.
- ▷ Étudier l'intérêt d'une mise-à-jour « incomplète » des états, non basée sur une confiance totale dans les observations de débit.

Sommaire

7.1 Quel(s) état(s) mettre à jour ? 107
7.2 Partir d'une mise-à-jour simple 109
7.2.1 Une mise-à-jour « complète » et non-paramétrique 109
7.2.2 Performances 109
7.3 Faire remonter la mise-à-jour 110
7.3.1 Mise-à-jour « pas à pas » de l'hydrogramme unitaire 111
7.3.2 Mise-à-jour de la pluie efficace PRT_{t-i} 115
7.3.3 Bilan des essais de mise-à-jour de l'hydrogramme unitaire 116
7.4 Emploi du filtre de Kálmán étendu 116
7.4.1 Principe 116
7.4.2 Usages en hydrologie 120
7.4.3 Mise-à-jour des états de GRP par le filtre de Kálmán 121
7.4.4 Tests réalisés et résultats 123
7.4.5 Discussion 127
7.5 Mise-à-jour incomplète du niveau du réservoir de routage . . . 129
7.5.1 Tests et résultats 129
7.5.2 Discussion 129
7.6 Bilan 131

Les états d'un modèle sont les variables qui caractérisent l'objet modélisé à un instant donné en fonction des différents forçages auxquels il est soumis. Ce sont donc des variables susceptibles de varier rapidement quand les forçages varient. Il est donc très logique de chercher à modifier la valeur de ces variables lors d'une mise-à-jour du modèle.

Dans ce chapitre, nous allons commencer par nous demander quels sont les états du modèle GRP qu'il faut chercher à mettre à jour. Nous partirons d'une méthode déterministe à la fois extrêmement simple et tout à fait efficace de mise-à-jour d'un seul état, proposée par Tangara (2005), que nous chercherons à complexifier, tout d'abord empiriquement puis en employant une des méthodes de mise-à-jour des états le plus souvent discutées : le filtre de Kálmán.

Alors que les premières méthodes que nous testerons ont une confiance aveugle dans l'observation de débit à assimiler, le filtre de Kálmán ne lui accorde qu'une confiance relative. Nous examinerons l'impact de ce choix à travers cette approche et une méthode simple présentant la même propriété.

Sauf mention contraire, la notation t indiquera dans ce chapitre, l'instant de prévision qui sera également supposé être l'instant de la dernière observation disponible de débit. Pour tout état x, nous distinguerons l'état calculé par le modèle au pas de temps t (en partant d'états antérieurs éventuellement mis-à-jour) $x_{t|t-1}$ de l'état mis-à-jour grâce à une observation faite à ce pas de temps t, $x_{t|t}$.

7.1 Quel(s) état(s) mettre à jour ?

Une des premières questions concernant la mise-à-jour des états est : sur quel(s) état(s) doit-on effectuer la mise-à-jour ? Différents arguments peuvent être avancés pour étayer une réponse : Seo *et al.* (2003) préfèrent ne pas toucher à l'humidité du sol car c'est un composant à mémoire longue et que la fenêtre temporelle sur laquelle se fait l'assimilation de données est courte. Wöhling *et al.* (2006) au contraire privilégient ce compartiment du modèle car ce dernier, très en amont au sein du modèle, lui assure un effet sur plusieurs pas de temps.

Les états du modèle GRP sont organisés *en cascades* de l'amont (niveau du réservoir de production) vers l'aval (niveau du réservoir de routage). *A priori*, plus un état est situé « en aval » du modèle, c'est-à-dire proche du débit simulé, plus il est susceptible d'être corrélé avec le débit (Fig. 7.1). Aussi, dans le cas où seule l'information des débits antérieurs observés peut être employée pour mettre à jour les états du modèle, il est logique de chercher en priorité à mettre à jour les états les plus corrélés aux débits, c'est-à-dire le plus souvent les états les plus « en aval » du modèle.

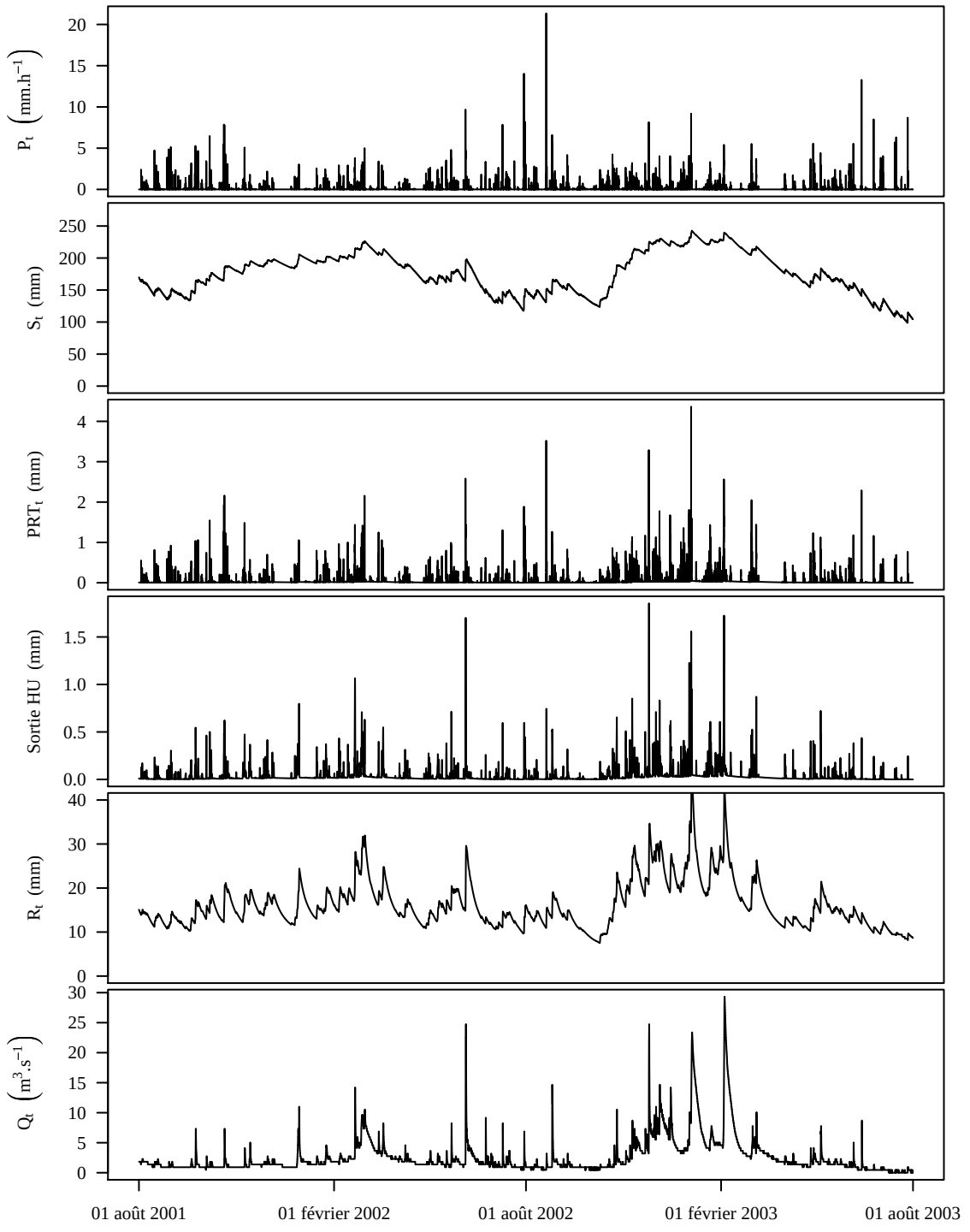


FIGURE 7.1 – Exemple de chroniques des états internes de GRP sur le bassin de la Creuse à Felletin (L4010710, 165 km^2) sur deux années hydrologiques. Les graphes donnent successivement les pluies, le niveau du réservoir de production, la pluie efficace PRT, la sortie de l'hydrogramme unitaire, le niveau du réservoir de routage et le débit observé.

7.2 Partir d'une mise-à-jour simple

7.2.1 Une mise-à-jour « complète » et non-paramétrique

Tangara (2005) propose la mise-à-jour « complète » du niveau du réservoir de routage, qui est l'état le plus « en aval » du modèle GRP. Ici, l'adjectif « complet » signifie que la sortie $\widehat{Q}_{t|t}$ du modèle ainsi mis-à-jour sera exactement égale à l'observation Q_t.

Comme rappelé dans l'annexe C, la vidange de ce réservoir est une fonction puissance, caractérisée par l'équation différentielle $dR = -k \cdot R^{\alpha}\, dt$ dont la résolution donne accès à la fonction $f : R_t^+ \mapsto Q_t$ reliant le débit au niveau du réservoir (intégré sur un pas de temps). Cette fonction f est bijective. La mise-à-jour de Tangara (2005) consiste à employer la réciproque de cette fonction pour la dernière observation de débit pour s'assurer que le débit simulé correspondant est égal à l'observation :

$$R_{t|t}^+ = f^{-1}\left(Q_t\right) \quad \text{où } R_{t|t}^+ \text{ est le niveau du réservoir mis-à-jour (avant sa vidange)}$$

Dans le cas d'un réservoir quadratique, l'inversion de la fonction f est analytique. Pour des puissances α non entières ou supérieures à 3, une inversion numérique est possible [1].

7.2.2 Performances

Gains en performance associés à cette mise-à-jour

Tangara (2005) a montré l'importance des effets de cette mise-à-jour. Au pas de temps horaire, cette mise-à-jour seule assure des critères de persistance nettement positifs aux horizons de prévision les plus courts (Tab. 7.1).

	$C_{2M,P}$			C_{2M}		
	Simulation	MàJ R	Corr.	Simulation	MàJ R	Corr.
$H = 1$ h	-0.97	0.15	-0.92	0.54	0.99	0.81
$H = 6$ h	-0.55	0.28	-0.19	0.54	0.91	0.75
$H = 24$ h	0.15	0.46	0.36	0.54	0.75	0.67
$H = 48$ h	0.37	0.54	0.49	0.54	0.70	0.65

Table 7.1 – Comparaison des performances globales (mesurées en $C_{2M,P}$, critère de persistance borné et en C_{2M}, critère de Nash-Sutcliffe borné) obtenues par le modèle GRP (a) sans aucune mise-à-jour, (b) avec la mise-à-jour complète du niveau du réservoir de routage et (c) sans cette mise-à-jour mais avec la correction des erreurs en sortie du modèle (proposée par Tangara, 2005).

1. Et peu coûteuse en temps de calcul : il est en effet possible de tabuler la fonction f indépendamment du paramètre X_3 (capacité du réservoir de routage), voir l'annexe K.1.

Limites de l'approche : les erreurs de mesure

Cette mise-à-jour est purement déterministe et « complète ». Elle ne recherche pas une estimation « optimale » tenant compte de l'incertitude sur la dernière observation de débit. Cela suppose que les observations de débit télétransmises ne sont pas injectées automatiquement dans le modèle mais font l'objet d'une critique du prévisionniste (notamment pour écarter les valeurs absurdes). La prise en compte d'une observation fortement entachée d'erreur pourrait avoir des conséquences catastrophiques sur la qualité de la prévision.

Nous étudierons, dans les sections 7.4 et 7.5 de ce chapitre, deux méthodes ne présentant pas ces défauts : le filtre de Kálmán et une variante de la méthode de Tangara qui ne fournit pas une mise-à-jour « complète ».

Durées d'influence de cette mise-à-jour

La mise-à-jour du réservoir de routage a une influence qui dure plusieurs centaines de pas de temps (Berthet *et al.*, 2009b, reproduit dans l'annexe N, page 541). Nous y définissons la *durée de vie* d'une mise-à-jour du modèle comme étant le laps de temps suivant la mise-à-jour pendant lequel son effet reste significatif. Une telle durée de vie signifie que cette mise-à-jour est efficace pour les horizons de prévision qui nous intéressent dans le cadre de cette thèse. Par contre, la mise-à-jour étant complète, elle fait « oublier » au modèle toute l'information qui était contenue dans les observations de débit antérieures à la dernière observation.

7.3 Faire remonter la mise-à-jour

La partie précédente a montré l'efficacité de la mise-à-jour du niveau du réservoir de routage. Nous allons maintenant rechercher des mises-à-jour empiriques des états situés plus en amont du réservoir de routage, à savoir les stocks d'eau en transit dans l'hydrogramme unitaire. La mise-à-jour du réservoir de routage précédente modifiait le niveau du réservoir (avant vidange du dernier pas de temps observé) pour que le débit simulé soit égal à l'observation. Tous les essais auront pour principe commun de modifier les stocks de l'hydrogramme unitaire de façon à ce que le niveau du réservoir atteigne ce niveau $\widehat{R}_t$, ce qui permet de conserver l'effet (bénéfique) de la mise-à-jour de Tangara (2005). Nous testerons deux approches principales : d'abord la remontée à travers l'hydrogramme unitaire en mettant à jour les états de l'hydrogramme les uns après les autres en partant de l'aval, puis une mise-à-jour de l'ensemble de l'hydrogramme unitaire par l'amont.

7.3.1 Mise-à-jour « pas à pas » de l'hydrogramme unitaire

Le principe de base de cette approche consiste à reporter la différence de volume nécessaire à la mise-à-jour du réservoir de routage $\delta R_t = R^+_{t|t} - R^+_t = f^{-1}(Q_t) - R^+_t$ sur le dernier compartiment de l'hydrogramme unitaire et éventuellement si besoin sur les suivants [2].

Tests réalisés

Quatre stratégies ont été définies (illustrées par la figure 7.2) :

1. La différence δR_t est entièrement répercutée sur le niveau du premier pas de temps en attente de l'hydrogramme unitaire. Cela signifie que l'hydrogramme unitaire peut contenir des quantités négatives d'eau en attente.
2. La différence δR_t est répartie sur l'ensemble des niveaux des pas de temps en attente de l'hydrogramme unitaire. La répartition est proportionnelle aux niveaux dans l'hydrogramme en régime permanent :

$$\delta HU_t(i) = \sum_{k=i}^{\eta} \gamma_k \delta R_t$$

   où $\eta = \lfloor X_3 \rfloor$ est la taille de l'hydrogramme unitaire (en nombre de pas de temps) et γ_k les ordonnées de l'hydrogramme.
3. La différence δR est répercutée sur le premier compartiment en attente tant que celui-ci contient de l'eau : si de l'eau doit être retirée de ce compartiment et si son contenu ne suffit pas, le restant est prélevé dans le deuxième compartiment, puis dans le troisième si nécessaire et ainsi de suite. Dans le cas où on doit retirer la quantité A d'eau de l'hydrogramme :
   (a) Si $\exists\ n \leq \eta / \sum_{i=1}^{n} HU_t(i) \geq A$ alors
   $$\forall\ i \in [\![1, n-1]\!], HU_{t|t}(i) = 0 \text{ et } HU_{t|t}(n) = HU_t(n) - \left(A - \sum_{i=1}^{n} HU_t(i)\right)$$
   (b) Si $A > \sum_{i=1}^{\eta} HU_t(i)$ alors $\forall\ i \in [\![1, \eta]\!], HU_{t|t}(i) = 0$

   Dans le second cas (l'hydrogramme unitaire ne contient pas assez d'eau pour réaliser la mise-à-jour), le réservoir de routage est mis-à-jour de façon exacte (i.e. il n'y a pas conservation du bilan en eau car cette mise-à-jour du réservoir de routage ne correspond pas à la quantité d'eau qui lui est apportée).
4. Le quatrième essai est identique au précédent, sauf dans le cas où l'hydrogramme unitaire ne contient pas assez d'eau pour réaliser la mise-à-jour. Dans cet essai, l'hydrogramme unitaire est vidé comme dans l'essai précédent mais le réservoir de routage ne reçoit que l'eau contenue dans l'hydrogramme unitaire, il n'y a donc pas mise-à-jour exacte du réservoir de routage.

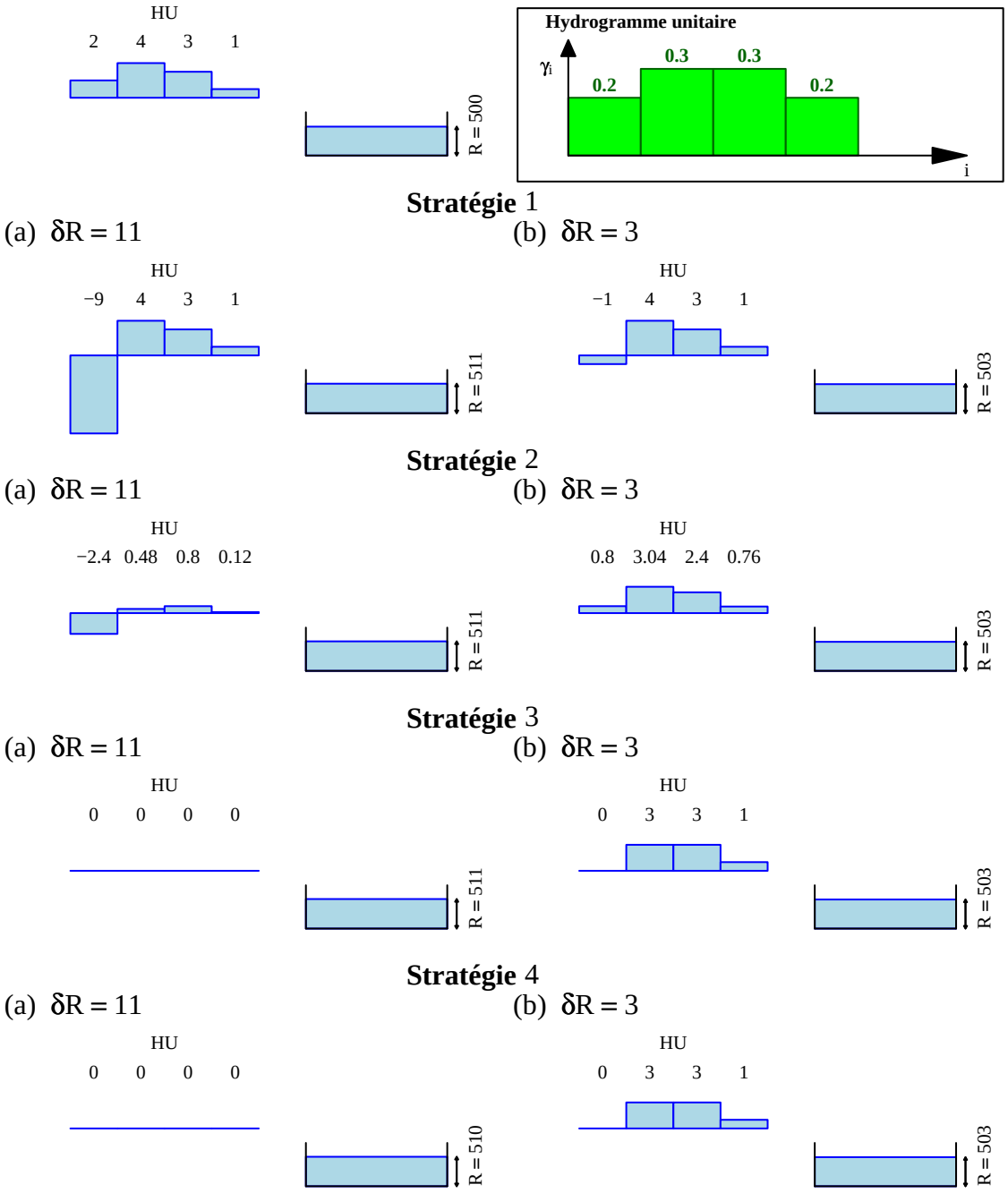


FIGURE 7.2 – Deux exemples fictifs de l'action des quatre stratégies testées de mise-à-jour de l'hydrogramme. Dans tous les cas, le niveau du réservoir de routage est initialement de 500 mm. L'hydrogramme comporte quatre compartiments et ses ordonnées sont 0.2, 0.3, 0.3 et 0.2. Avant mise-à-jour, les stocks d'eau en attente dans l'hydrogramme unitaire sont de 2, 4, 3 et 1 mm. Dans le premier exemple (colonne de gauche (a)), il est nécessaire d'apporter 11 mm au réservoir de routage tandis que dans le second exemple (colonne de droite (b)), δR n'est que de 3 mm.

Pour chacune de ces stratégies, il est possible de conserver ou non les modifications sur les états induites par ces mises-à-jour d'un pas de temps de prévision à l'autre. Nous avons donc comparé les quatre stratégies précédentes avec l'une ou l'autre option, soit huit variantes (évaluées pour huit horizons de prévision).

Résultats globaux

Dans toutes les configurations présentées ci-dessus, la mise-à-jour de l'hydrogramme unitaire ne permet pas d'améliorer les performances globales du modèle (que ce soit pour l'ensemble des pas de temps ou uniquement au-dessus d'un seuil en débit). Les tendances en fonction des différents horizons de prévision ne sont pas claires comme on peut le constater dans le tableau synoptique 7.2. La comparaison des différentes stratégies testées permet néanmoins quelques remarques utiles pour d'autres approches ; ces dernières sont détaillées dans les paragraphes suivants.

Influence de l'enregistrement de la mise-à-jour d'une date de prévision à l'autre

Quelle que soit la variante testée, il semble indifférent, voire néfaste d'enregistrer les mises-à-jour d'une date de prévision à l'autre pour les faibles horizons de prévision. A contrario, plus les horizons de prévisions deviennent longs, plus l'enregistrement de la mise-à-jour se révèle bénéfique. On peut l'expliquer ainsi : aux courts horizons de prévision, l'effet de la dernière mise-à-jour se fait sentir très fortement et donc suffit seule à assurer des bonnes performances. A plus long terme, le modèle se rapproche de la simulation et les sorties dépendent plus des états autres que le niveau du réservoir de routage. Les mises-à-jour de l'hydrogramme unitaire développées ici paraissent donc « hydrologiquement » raisonnables même si elles ne permettent pas d'améliorer les performances.

Influence de l'existence de réserves négatives

La variante 1 peut conduire à des étrangetés où l'hydrogramme unitaire a certains compartiments remplis négativement. Ceci est corrigé dans les essais 3 à 4. On observe que cette particularité, apparemment absurde, n'a pas d'influence notable sur le fonctionnement du modèle.

Influence de la conservation des masses d'eau dans la mise-à-jour du routage

La comparaison entre GRP et les différentes variantes montre que la conservation de l'écoulement n'est pas nécessaire pour réaliser de bonnes prévisions et que la version initiale – sans conservation de l'eau – conduit à de meilleures performances. Les essais 3 et 4 qui ne diffèrent que par l'exactitude de la mise-à-jour du réservoir en aval de l'hydrogramme confirment que la conservation de l'eau n'est pas une clef du problème.

2. Les notations employées sont celles décrites dans l'annexe C, page 339 et suivantes.

H = 1 h → H = 48 h	Sans mémorisation de la mise-à-jour d'une date de prévision à l'autre	Avec mémorisation de la mise-à-jour d'une date de prévision à l'autre
1. Report sur le premier compartiment de l'hydrogramme unitaire.	GRP bien meilleur GRP un peu meilleur	GRP bien meilleur Égalité
2. Report proportionnel sur tous les compartiments de l'hydrogramme unitaire.	GRP bien meilleur GRP un peu meilleur mais il existe des BV pour lesquels la mise-à-jour de l'HU améliore les performances	GRP bien meilleur Égalité sauf pour quelques BV sur lesquels GRP est meilleur
3. Report successif sur les compartiments de l'hydrogramme unitaire avec mise-à-jour exacte du niveau du réservoir de routage.	Égalité (1h) GRP meilleur (12 h, 18 h) GRP un peu meilleur (48 h)	GRP meilleur (1 h) GRP bien meilleur (3 h, 6 h) Égalité (24 h, 36 h ,48h)
4. Report successif sur les compartiments de l'hydrogramme unitaire sans mise-à-jour exacte du niveau du réservoir de routage.	GRP un tout petit peu meilleur GRP bien meilleur (12 h, 18 h) GRP un peu meilleur (48 h)	GRP meilleur GRP bien meilleur (3 h, 6 h) GRP un peu meilleur (48 h)

TABLE 7.2 – Gains de performance obtenus par mise-à-jour de l'hydrogramme unitaire selon la stratégie de mise-à-jour employée. Les quatre lignes correspondent aux quatre stratégies présentées page 111. Les deux colonnes indiquent si la mise-à-jour de l'hydrogramme est conservée d'un pas de temps de prévision à l'autre ou non. L'évolution des gains de performance en fonction de l'horizon de prévision est représentée sur la diagonale de chaque case. Les indications reportées dans le tableau correspondent à un examen de la distribution sur 178 bassins des performances globales. Elles correspondent aux conventions suivantes : pour un gain médian inférieur à 3 points de $C_{2M,P}$, il y a *égalité* (le gain n'est pas considéré significatif), pour un gain inférieur à 5 points, le gain est faible (*un peu meilleur*), pour un gain supérieur à 8 points, le gain est important (*bien meilleur*).

7.3.2 Mise-à-jour de la pluie efficace PRT_{t-i}

Jusqu'ici, nous avons considéré un seul jeu de variables permettant de décrire l'état du système modélisé à un instant t donné. Il s'agit des niveaux des deux réservoirs S_t et R_t et des stocks d'eau de l'hydrogramme unitaire $HU_t(i)$. D'autres vecteurs d'état issus de la combinaison de ce premier jeu d'états sont possibles. Un autre vecteur d'état intéressant est composé des niveaux des deux réservoirs S_t et R_t et des dernières pluies efficaces $PRT_{t-\eta}$, ..., PRT_t.

Nous allons maintenant chercher à mettre à jour l'hydrogramme unitaire à travers ce vecteur d'état. Pour cela, nous déterminons la perturbation δPRT_{t-k} à appliquer sur la pluie efficace antérieure PRT_{t-k} qui permet d'obtenir le niveau du réservoir de routage R_t correspondant à la mise-à-jour complète de ce niveau. Une fois la mise-à-jour de PRT_{t-k} connue, les stocks d'eau de l'hydrogramme unitaire s'en déduisent directement. Les calculs sont détaillés dans l'annexe K.2, page 506.

Nous avons réalisé plusieurs tests pour comparer les performances obtenues avec cette mise-à-jour pour différentes valeurs de k (de 0 à 12), c'est-à-dire en jouant sur l'ancienneté de l'état mis-à-jour.

Résultats

Quelle que soit la valeur de k choisie, les performances globales sont moins élevées que celles du modèle GRP (où la mise-à-jour des états ne modifie que le niveau du réservoir de routage). Il n'y pas de tendance claire, en fonction de la durée de l'hydrogramme unitaire.

Les pertes de performance globale (mesurées en $C_{2M,P}$) sont assez logiquement modérées à l'horizon d'une heure : de 5 points à moins d'un point de diminution. Elles augmentent avec l'horizon de prévision : les pertes médianes atteignent 10 à 15 points de $C_{2M,P}$ pour un horizon de 12 heures. La mise-à-jour de l'hydrogramme unitaire qui a pour but d'améliorer la représentation des états impliqués dans les prévisions à plus d'un pas de temps est ici un échec. Cette approche par l'amont de la mise-à-jour de l'hydrogramme unitaire est nettement moins efficace que l'approche par l'aval présentée précédemment.

Une explication possible réside dans la forme de l'hydrogramme unitaire. Les ordonnées de ses tout derniers compartiments sont faibles. La détermination de la mise-à-jour de la pluie efficace antérieure est donc peu robuste :

$$\delta PRT_{t-k} = \frac{1}{\gamma_{k+1}} \delta R_t \quad \text{où } \gamma_k \text{ est l'ordonnée du } k^{\text{e}} \text{ compartiment de l'HU.}$$

La moindre erreur sur l'observation Q_t entraîne une erreur sur δPRT_{t-k} supérieure à l'erreur de mise-à-jour sur le niveau du réservoir δR_t. Cette erreur sera ensuite répercutée sur les niveaux futurs du réservoir de routage (jusqu'à atteindre l'horizon de prévision fixé) avec une amplitude potentiellement supérieure à celle de l'erreur initiale.

7.3.3 Bilan des essais de mise-à-jour de l'hydrogramme unitaire

Nous avons vu dans le chapitre 6 que la modification de la forme de l'hydrogramme unitaire avait des conséquences plutôt faibles sur les performances du modèle. Nous observons ici que nous n'avons pas obtenu d'amélioration de ces performances par des mises-à-jour empiriques des états de l'hydrogramme. De plus, les mises-à-jour proposées peuvent affaiblir la robustesse du modèle de prévision.

Ces résultats nous incitent à ne pas faire remonter plus en amont les mises-à-jour sur les états et à ne pas mettre à jour les entrées. Il est probable que la robustesse serait encore plus compromise sans obtenir de gains significatifs de performance. Ces résultats corroborent ceux obtenus par Tangara (2005) au pas de temps journalier.

Voyons maintenant ce que nous pouvons obtenir par une méthode classique d'assimilation de données, le filtre de Kálmán.

7.4 Emploi du filtre de Kálmán étendu

Le filtre de Kálmán est une méthode séquentielle très répandue dans le domaine de l'assimilation de données. Comme nous le verrons, elle permet de réaliser une mise-à-jour « partielle » des états qui se base sur les confiances relatives que nous avons dans le modèle et dans les observations. Nous commencerons par en présenter le principe et les variantes. Toutes les notations sont rassemblées dans le tableau 7.4 (page 119).

7.4.1 Principe

L'espace des états

Le formalisme employé ici est très souvent utilisé dans les techniques séquentielles. Il décrit le modèle et le système dans l'*espace des états*. Le système y est complètement décrit par un vecteur $\mathbf{x}_t$. Ces états ne sont pas (toujours) observables ; par contre, le système peut être observé à travers d'autres grandeurs rassemblées dans le vecteur $\mathbf{z}_t$. Les observations sont reliées aux états par l'*équation d'observation* :

$$\mathbf{z}_t = h\left(\mathbf{x}_t\right) + \mathbf{v}_t \tag{7.1}$$

où h est l'opérateur d'observation et $\mathbf{v}_t$ du bruit correspondant à l'incertitude d'observation.

La dynamique du système est rendue par l'*équation de propagation* :

$$\mathbf{x}_{t+1} = f\left(\mathbf{x}_t, \mathbf{e}_t\mathbf{u}_t, \mathbf{w}_t\right) \tag{7.2}$$

où f est un opérateur (le modèle), $\mathbf{e}_t$ est le vecteur des entrées accompagnées de leurs erreurs rassemblées dans $\mathbf{u}_t$ et $\mathbf{w}_t$ est le vecteur des erreurs dues au modèle (dans le passage du pas de temps t au pas de temps $t+1$).

Le filtre de Kálmán original

Ce filtre a été présenté initialement par Rudolf E. Kálmán (1960) qui s'intéressait aux processus stochastiques modélisés par la réponse d'un système dynamique connu à des entrées aléatoires[3]. Cette classe de problèmes est très importante en télécommunications et en automatique (contrôle). Le but de cet article était alors de trouver une estimation optimale de l'état $\mathbf{x}_t$ du système (non mesurable) à partir du modèle et des observations $\mathbf{z}_0, \ldots, \mathbf{z}_t$.

Pour décrire ces problèmes, Kálmán pose un certain nombre d'hypothèses (résumées dans le tableau 7.3). Toutes partent de l'idée d'un bruit blanc en entrée du système ou à l'œuvre à l'intérieur du système. Aussi, les erreurs du modèle et d'observation sont considérées indépendantes (dans le temps) et gaussiennes de moyennes nulles. Elles sont donc complètement caractérisées par leurs matrices de variance-covariance $\mathbf{Q}_t$ et $\mathbf{R}_t$ qui vérifient :

$$\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_{t'}^T\right) = \delta_{t,t'}\mathbf{Q}_t \quad \text{et} \quad \mathbb{E}\left(\mathbf{v}_t\mathbf{v}_{t'}^T\right) = \delta_{t,t'}\mathbf{R}_t$$

Les variables d'état $\mathbf{x}_t$ sont également considérées gaussiennes de moyenne nulle. Pour que cela reste vrai pour tout t, deux hypothèses sont rajoutées. La première décrit les observations comme des réalisations de variables aléatoires gaussiennes. La seconde pose que le système est linéaire ainsi que l'observation : les opérateurs f et h sont décrits par les matrices $\mathbf{F}$ et $\mathbf{H}$.

La meilleure estimation du vecteur aléatoire gaussien $\mathbf{x}_t$ correspond à la minimisation d'une fonction de coût quadratique :

$$\min_{\phi} J(\phi) = \min_{\phi} \mathbb{E}\left(\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2\right)$$

Dans ce cadre, Kálmán montre que l'estimation optimale $\mathbf{x}_{t|t}$, estimation de l'état au pas de temps t en utilisant au moins une observation à ce pas de temps, peut être obtenue de manière récursive et est une combinaison linéaire du vecteur des états estimé par le modèle $\mathbf{x}_{t|t-1}$ et de la différence entre observations et prédictions du système :

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t\left(\mathbf{z}_t - \mathbf{H}\mathbf{x}_{t|t-1}\right) \tag{7.3}$$

3. C'est par exemple le cas des effets macroscopiques d'un système constitué d'un grand nombre d'entités microscopiques en physique statistique.

où $\mathbf{K}_t$ est appelée *matrice de gain*. Son calcul est détaillé dans l'annexe K.4 (pages 509 et suivantes).

L'équation (7.3) permet en pratique d'implémenter le filtre par la répétition de deux étapes :
- **Propagation** : une estimation $\mathbf{x}_{t|t-1}$ de l'état à la date t est obtenue à partir de l'état $\mathbf{x}_{t-1|t-1}$ au pas de temps précédent, grâce à l'équation (7.2).
- **Mise-à-jour** : l'observation z_t est employée pour obtenir l'état $\mathbf{x}_{t|t}$ grâce aux équations (7.1) et (7.3).

Les erreurs sur les états étant gaussiennes (et de moyenne nulle), il suffit de calculer leur matrice de variance-covariance $\mathbf{P}_t$ pour les décrire entièrement. Cette matrice entre dans le calcul de la matrice de gain et est estimée à chaque itération (voir l'annexe K.4, page 509).

H_1	Les erreurs dues au modèle sont des variables aléatoires gaussiennes indépendantes.
H_2	Les erreurs d'observation sont des variables aléatoires gaussiennes indépendantes.
H_3	Les erreurs (initiales) sur les états sont des variables aléatoires gaussiennes.
H_4	Le modèle est linéaire.
H_5	La meilleure estimation $\mathbf{x}_{t\|t}$ correspond à la minimisation de l'erreur quadratique.

TABLE 7.3 – Hypothèses du filtre de Kálmán (original). Les hypothèses H_1 à H_4 permettent de conclure que les erreurs sur les états restent gaussiennes à tous les pas de temps et l'hypothèse H_5 est une définition de l'optimalité de l'estimation, cohérente avec les hypothèses précédentes.

Les variantes

Le filtre de Kálmán est devenu par la suite très populaire et a été employé dans des domaines éloignés des questions pour lesquels il avait été conçu initialement. Les hypothèses présentées ci-dessus – et en premier chef, la linéarité supposée du système – n'étaient pas toujours validées dans ces nouveaux domaines d'application. Plusieurs auteurs ont proposé des variantes pour adapter le filtre à ses nouvelles conditions d'emploi.

Face à une non-linéarité du modèle, l'approche la plus simple est de linéariser (là où cela est nécessaire) les opérateurs f et h des équations (7.1) et (7.2), autour de l'état $\mathbf{x}_{t|t}$. Il s'agit du *filtre de Kálmán étendu* (EKF). Cette méthode reste assez simple mais peut néanmoins être pénible à mettre en œuvre si le calcul du jacobien de f (à chaque pas de temps) est long. En outre, plusieurs auteurs ont pointé des limites à cette approche. Ljung (1979) rapporte que cette méthode peut donner des estimations biaisées si le modèle d'erreur n'est pas correct ou si les valeurs des paramètres ne sont pas exactes. Il montre en outre que l'EKF n'est pas toujours stable et peut diverger en cas de fortes non linéarités. Miller *et al.* (1994), Evensen (1994) et Vrugt *et al.* (2005) expliquent cette instabilité par le fait que les moments statistiques d'ordre supérieurs à 3 sont négligés.

$\mathbf{x}_{t\|t-1}$	le vecteur des états après propagation,
$\mathbf{x}_{t\|t}$	le vecteur des états après mise-à-jour (i.e. l'estimation obtenue grâce à l'assimilation de l'observation $\mathbf{z}_t$),
$\mathbf{z}_t$	le vecteur des observations,
$\mathbf{e}_t$	le vecteur des entrées,
$\mathbf{u}_t$	les erreurs sur les entrées, de matrice de variance-covariance $\mathbf{S}_t$,
$\mathbf{v}_t$	les erreurs sur les observations, de matrice de variance-covariance $\mathbf{R}_t$,
$\mathbf{w}_t$	les erreurs sur états, de matrice de variance-covariance $\mathbf{Q}_t$,
$\mathbf{P}_{t\|t-1}$	la matrice de variance-covariance des erreurs des états estimés $\mathbf{x}_{t\|t-1}$,
$\mathbf{P}_{t\|t}$	la matrice de variance-covariance des erreurs des états mis-à-jour $\mathbf{x}_{t\|t}$,
$\mathbf{Q}_t$	la matrice de variance-covariance des erreurs dues au modèle,
$\mathbf{R}_t$	la matrice de variance-covariance des erreurs d'observation,
$\mathbf{S}_t$	la matrice de variance-covariance des erreurs sur les entrées,
f_t	l'opérateur représentant le modèle (i.e. la dynamique du système),
h_t	l'opérateur d'observation (lien entre les états et les observations),
$\mathbf{F}_t$	la matrice de f dans le cas linéaire, sa matrice jacobienne sinon,
$\mathbf{A}_t$	la matrice de prise en compte des entrées,
$\mathbf{H}_t$	la matrice de h dans le cas linéaire, sa linéarisation sinon,
$\mathbf{K}_t$	la matrice de gain.

Table 7.4 – Principales notations employées dans les sections consacrées au filtre de Kálmán et à ses dérivés (filtre de Kálmán étendu et filtre de Kálmán d'ensemble). Ces notations s'inscrivent dans le formalisme de l'espace des états. L'indice t n'est ajouté aux matrices, dans le corps principal, que dans les situations où ces matrices varient dans le temps.

Evensen (1994, 2003) a développé une approche, dite *filtre de Kálmán d'ensemble* (EnKF). Son but est d'employer le cadre conceptuel du filtre de Kálmán étendu sans linéariser les opérateurs pour calculer les matrices de gain et de variance-covariance des erreurs sur les états. Pour cela, le système n'est plus représenté par un vecteur d'état mais par un ensemble de vecteurs d'états. Chaque membre emploie l'équation de propagation (7.2) pour passer d'un pas de temps à l'autre. Lors de la mise-à-jour, la matrice de variance-covariance des erreurs de prévision $\mathbf{P}_{t|t-1}$ peut être estimée directement sur l'ensemble des états prédits $\mathbf{x}_{t|t-1}$ (par rapport à un état particulier ou par rapport à la moyenne de l'ensemble). La matrice de variance-covariance des erreurs de mise-à-jour $\mathbf{P}_{t|t}$ est calculable de la même façon, ce qui n'oblige pas à négliger les moments statistiques d'ordres supérieurs à 2. Finalement, le calcul de ces matrices n'est pas nécessaire : la matrice de gain s'obtient directement à partir de l'ensemble. Le filtre de Kálmán d'ensemble peut être plus économique en calculs que le filtre de Kálmán étendu pour les systèmes où le nombre d'états est supérieur à

l'effectif de l'ensemble, ce qui ne sera pas le cas pour nous. Nous n'emploierons donc pas cette variante.

Signification des matrices de variance-covariance

Il est nécessaire de spécifier les matrices de variance-covariance des erreurs du modèle $\mathbf{Q}$ et des observations $\mathbf{R}$ (ainsi que celle des entrées, le cas échéant). Il s'agit d'hyperparamètres(g) qu'il est nécessaire de préciser avec soin (Georgakakos, 1986). L'utilisateur doit donc expliciter un modèle d'erreur (restreint au cadre d'erreurs gaussiennes en théorie). Il indique ainsi les niveaux de confiance relatifs pour les observations et pour le modèle. Si les valeurs contenues dans $\mathbf{Q}$ sont très nettement supérieures à celles contenues dans $\mathbf{R}$, l'utilisateur indique au filtre une confiance bien plus forte dans les observations que dans le modèle, ce qui entraînera une mise-à-jour quasi-complète des états pour que les sorties du modèle « collent » aux observations passées.

7.4.2 Usages en hydrologie

L'emploi du filtre de Kálmán en hydrologie et dans des sciences connexes a suscité de nombreuses contributions scientifiques. En météorologie, Gauthier *et al.* (1993) emploient le filtre de Kálmán pour assimiler des observations satellitaire de vents ; Storm *et al.* (1989) évaluent les conséquences des incertitudes sur la pluie de bassin sur les débits simulés par le modèle hydrologique NAM incorporé dans un filtre de Kálmán. Cette technique d'assimilation de données est souvent utilisée pour mettre à jour des modèles hydrauliques (e.g. Madsen et Skotner, 2005, Neal *et al.*, 2007). En hydrologie, plusieurs auteurs ont employé le filtre étendu pour assimiler l'humidité du sol dans un modèle sol-végétation-atmosphère (Reichle *et al.*, 2002) et dans un modèle Pluie – Débit (Aubert *et al.*, 2003). D'autres approches assimilent les observations de débit dans un modèle Pluie – Débit : une version linéarisée par morceaux ou simplifiée du modèle de Sacramento SAC-SMA (Kitanidis et Bras, 1980a,b), un modèle hydrométéorologique de prévision temps réel (des précipitations et des débits, Georgakakos, 1986) ou encore le modèle GR4J (Aubert *et al.*, 2003). Dans ces trois cas, la mise-à-jour améliore les performances mais ces dernières restent faibles. Le filtre de Kálmán est appliqué à la partie linéaire de IHACRES (le routage) par Schreider *et al.* (2001) et Ouachani *et al.* (2007) emploient le filtre pour mettre à jour les états « inférieurs » du modèle HBV (Bergström et Forsman, 1973, Lindström *et al.*, 1997) (ici encore, grosso modo la fonction de routage de ce modèle). Enfin, plusieurs auteurs ont cherché à mettre à jour les états et les paramètres. Moradkhani *et al.* (2005b) emploient successivement deux filtres de Kálmán d'ensemble à cet effet. Vrugt *et al.* (2005) cherchent à affiner la distribution des paramètres tout en mettant à jour les états du modèle par la combinaison de l'algorithme de calage SCEM-UA et d'un filtre de Kálmán d'ensemble.

Plusieurs auteurs ont comparé le filtre de Kálmán à d'autres méthodes de mise-à-jour du modèle. Da Ros et Borga (1997) d'une part et Seo *et al.* (2003) d'autre part l'ont

comparé à des méthodes variationnelles : les premiers le trouvent plus performant pour leur modèle tandis que les seconds trouvent que le modèle mis-à-jour par le filtre de Kálmán est moins performant et peut même se révéler inférieur au modèle de persistance (même observation chez Aubert *et al.*, 2003). Refsgaard (1997) compare différentes mises-à-jour du modèle hydrologique NAM et observe que le filtre de Kálmán est un peu plus performant que l'ajout d'un module hydraulique mis-à-jour. En revanche, Khu *et al.* (2001) trouvent qu'une correction très simple des erreurs (dont la formule est obtenue par programmation génétique, i.e. une forme automatisée d'essais-erreurs) du même modèle NAM obtient de bien meilleurs résultats que le filtre de Kálmán.

De nombreuses contributions relèvent le non respect des hypothèses liées à l'emploi du filtre de Kálmán tant sur la structure des erreurs que sur la linéarité du modèle (e.g. Kitanidis et Bras, 1980b, Gordon *et al.*, 1993, Fortin *et al.*, 1997, Wang et Cai, 2008). Ces auteurs en déduisent que les efforts (parfois longs) consentis pour employer le filtre de Kálmán n'assurent pas l'optimalité du filtre : Moore *et al.* (2005) en concluent que les « *computations based on such solutions can be highly complex and there is no guarantee that they will provide better forecasts than ones based on the simple empirical scheme described above* » (voir également Xiong et O'Connor, 2002, Moore, 2007). Les résultats indiqués au paragraphe précédent le montrent assez. Cependant, certains auteurs estiment que les filtres de Kálmán sont relativement simples à mettre en œuvre (comparés au calcul de l'adjoint d'un modèle...). Le Dimet et Talagrand (1986), quant à eux, indiquent que les méthodes séquentielles optimales peuvent ne pas être appropriées aux événements extrêmes (qui nous intéressent !).

D'autres auteurs pointent le fait que l'adaptation du modèle au filtre agit comme un lit de Procuste et peut modifier en profondeur le modèle, notamment s'il présente de fortes non linéarités ou des effets de seuil (Kitanidis et Bras, 1980a, Seo *et al.*, 2003). Tucci et Clarke (1980) mettent en garde : « *it would be a pity if hydrological principle were to be sacrified in order to facilitate the mathematics* ». De son côté, Young (2002) développe des modèles *boîtes noires* parfaitement adaptés mathématiquement pour être mis-à-jour à l'aide d'un filtre de Kálmán. D'autre part, Pauwels et De Lannoy (2009) indiquent que l'une des tâches les plus difficiles (et subjectives) de l'implémentation du filtre de Kálmán réside dans l'écriture des opérateurs d'observation h quand il n'y a pas de lien direct entre les états et les observations, ce qui peut survenir en hydrologie, notamment en cas de problématiques de changement d'échelle.

7.4.3 Mise-à-jour des états de GRP par le filtre de Kálmán

Nous cherchons ici à mettre à jour les états du modèle hydrologique GRP par l'emploi du filtre de Kálmán. Les observations assimilées seront uniquement les observations passées du débit à l'exutoire. Commençons par étudier la validité des hypothèses du filtre. Nous

examinerons ensuite la sensibilité des sorties de GRP à ses états internes pour finalement choisir quels états mettre à jour.

Validité des hypothèses

Le modèle GRP n'est clairement pas linéaire. Cependant, les non linéarités ne sont pas très fortes. La fonction de routage est assez facilement linéarisable : elle combine une partie linéaire (convolution de l'hydrogramme unitaire) et une partie assez faiblement non linéaire (le réservoir de routage) facilement linéarisable (Fig. K.2, en annexe K.3, 510). La fonction de production est plus délicate à linéariser au niveau de la prise en compte des entrées dont la dérivée n'est pas continue autour de $E - P = 0$. Cela ne concerne pas de grands volumes d'eau mais un très grand nombre de pas de temps.

Nous n'avons pas d'information sur la structure statistique des erreurs dues au modèle. Les erreurs sur les entrées du modèle (principalement les pluies) sont complexes et nous ne pouvons pas garantir qu'elles soient gaussiennes et indépendantes dans le temps. Les erreurs sur les observations de débits sont clairement non gaussiennes. Une partie de ces erreurs n'est pas stochastique mais liée aux erreurs de la courbe de tarage : les erreurs ne seront pas indépendantes dans le temps.

En conséquence, les erreurs sur les états du modèle nous sont inconnues. Par contre, nous avons une certaine connaissance de la structure des erreurs sur les sorties du modèle (à l'incertitude sur les mesures de débit près). Ces erreurs sont loin d'être gaussiennes et indépendantes dans le temps : elles sont le plus souvent hétéroscédastiques et autocorrélées (voir le chapitre 3). En conséquence, la meilleure estimation des états n'est pas obtenue par une minimisation de l'erreur quadratique (sur les états) au sens de la fonction de vraisemblance.

Les hypothèses du filtre de Kálmán qui permettent d'affirmer l'optimalité de l'estimation ne sont pas (du tout) vérifiées dans le cadre de la mise-à-jour des états du modèle hydrologique GRP. Le filtre est techniquement applicable mais rien ne garantit, à ce niveau de connaissance, qu'il donnera de meilleurs résultats qu'une autre méthode.

Sensibilité des sorties de GRP à ses états

Les parties 7.2 et 7.3 nous ont montré que les sorties de GRP sur les premiers pas de temps futurs $t+1, \ldots, t+H$ sont très sensibles à la valeur du niveau du réservoir de routage puis qu'ils sont très nettement moins sensibles aux états « plus en amont » de l'hydrogramme unitaire. On en déduit que les sorties du modèle ne seront sensibles à une mise-à-jour du réservoir de production que si deux conditions sont réunies : la modification du niveau du réservoir devra être très importante et l'effet ne se verra qu'après un certain nombre de pas de temps (en fonction de la durée de l'hydrogramme unitaire).

Stratégie adoptée

Nous nous limitons à une mise-à-jour des états de la fonction de routage pour deux raisons. D'une part, ce choix nous permet d'employer facilement le filtre de Kálmán étendu (EKF) car la non-linéarité de la fonction de routage n'est pas très forte. D'autre part, nous sommes principalement intéressés par des effets à court terme répondant au choix des horizons de prévision courts.

Les états mis-à-jour seront donc les η volumes d'eau en attente dans l'hydrogramme unitaire ($\eta = \lfloor X_3 \rfloor$) et le niveau du réservoir de routage. Nous ajouterons à cela un état « fantôme » q pour tenir compte de l'autocorrélation des débits (par la technique de l'augmentation des états décrite par exemple par Reichle *et al.*, 2002, où le débit sortant du modèle sera la somme du débit simulé et d'un terme dont l'erreur rendra compte de l'autocorrélation du débit).

Une description détaillée de l'implémentation du filtre et le calcul des matrices impliquées se trouvent dans l'annexe K.4.

7.4.4 Tests réalisés et résultats

39 essais ont porté sur le choix des matrices de variance-covariance des erreurs sur les observations et des erreurs dues au modèle, $\mathbf{R}_t$ et $\mathbf{Q}_t$. Dans ce cadre, nous avons évalué les performances globales de la mise-à-jour, ainsi que la qualité des estimations de la variance sur les états.

Performances globales du filtre de Kálmán étendu

Sur l'ensemble de ces essais, pas un ne permet d'obtenir des performances significativement meilleures que celles du modèle GRP sur l'ensemble des bassins versants, quel que soit l'horizon de prévision. La mise-à-jour par filtre de Kálmán est relativement meilleure sur les bassins les plus lents que sur les bassins les plus réactifs. Cependant, sur ces bassins lents, nous n'avons pas mis en évidence une combinaison de matrices de variance-covariance $\mathbf{R}_t$ et $\mathbf{Q}_t$ permettant d'obtenir des performances supérieures à celles de la mise-à-jour de Tangara (2005). Il apparaît donc nécessaire de « caler » les matrices de variance-covariance sur chaque bassin.

Nous considérons maintenant pour chaque bassin, la combinaison $\{\mathbf{Q}_t, \mathbf{R}_t\}$ obtenant les meilleures performances parmi les 39 essais réalisés. Même dans ces conditions, les performances du modèle GRP mis-à-jour par le filtre de Kálmán ne sont pas significativement supérieures à celles du modèle dont seul le niveau du réservoir de routage est mis à jour (Fig. 7.3). Cette approche ne correspond pas à des conditions opérationnelles puisqu'il ne s'agit pas d'un calage : nous considérons la meilleure paramétrisation du filtre de Kálmán sur la période de « contrôle ». Il s'agit d'une borne supérieure des performances atteignables en conditions opérationnelles (si on se limite à nos 39 essais pour définir la paramétrisation

des matrices $\mathbf{Q}_t$ et $\mathbf{R}_t$, ce qui ne peut pas s'apparenter à un calage fin).

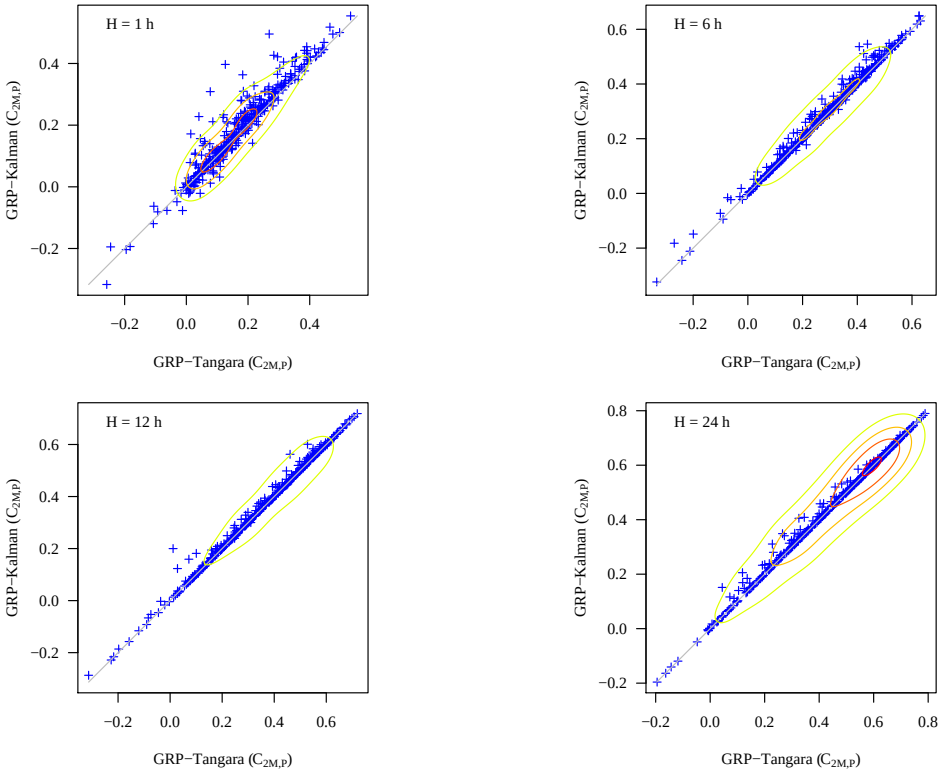


FIGURE 7.3 – Comparaison des performances mesurées par le critère $C_{2M,P}$ du modèle GRP dont seul le niveau du réservoir de routage est mis-à-jour et du modèle GRP dont les états de la fonction de routage sont mis-à-jour par le filtre de Kálmán. Pour chaque bassin × période, la meilleure paramétrisation au contrôle du filtre de Kálmán sur cette période a été retenue.

Si malgré la remarque précédente, nous effectuons un pseudo-calage basé sur les 39 essais de paramétrisation du filtre de Kálmán, les résultats sont assez similaires avec une bonne stabilité des performances sur la majorité des bassins versants. Nous constatons toutefois un défaut de robustesse de la méthode sur une dizaine de bassins versants, se classant parmi les bassins sur lesquels le modèle GRP est le plus faible : le modèle mis à jour par le filtre de Kálmán y présente des performances très nettement inférieures à celles obtenues avec la meilleure paramétrisation pour la période considérée (Fig. 7.4 et Tab. 7.5).

Quelle est la meilleure paramétrisation de la matrice $\mathbf{Q}_t$?

Parmi les 39 essais de paramétrisation de la matrice $\mathbf{Q}_t$, nous pouvons distinguer deux classes de matrices : les matrices constantes et les matrices variables. Les matrices constantes ne contiennent que des valeurs indépendantes du temps. Une autre approche (e.g. Georgakakos, 1986, Moradkhani *et al.*, 2005b) propose de faire évoluer la matrice en fonction

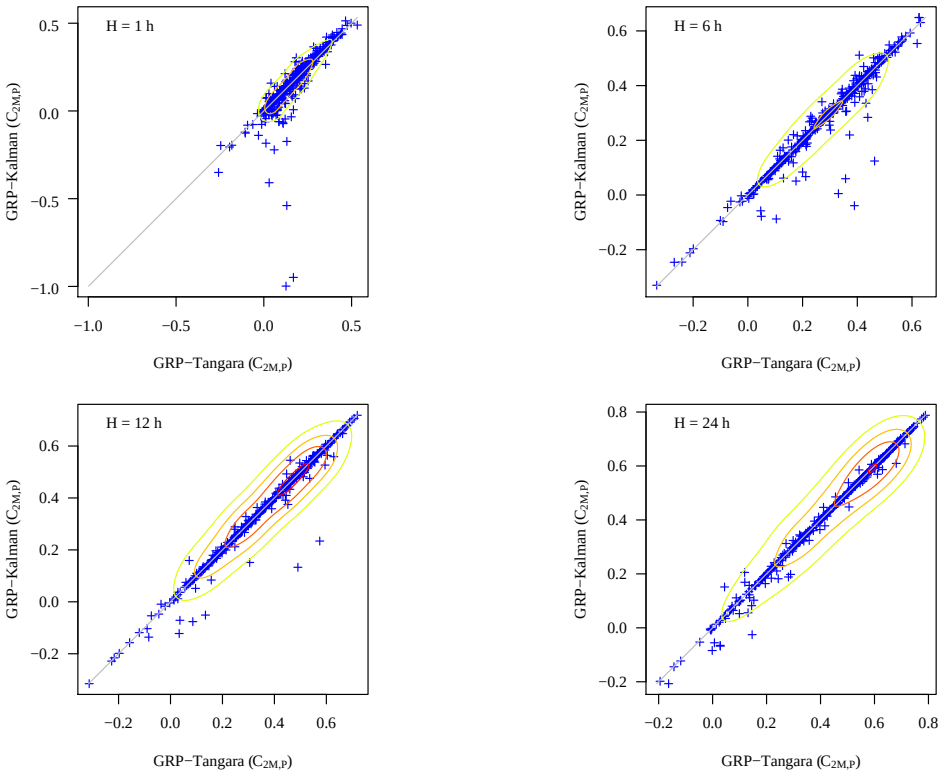


Figure 7.4 – Comparaison des performances mesurées par le critère $C_{2M,P}$ du modèle GRP dont seul le niveau du réservoir de routage est mis-à-jour et du modèle GRP dont les états de la fonction de routage sont mis-à-jour par le filtre de Kálmán. Pour chaque bassin × période, la meilleure paramétrisation au calage du filtre de Kálmán sur cette période a été retenue.

du débit pour s'adapter à l'hétéroscédasticité des erreurs dues au modèle. Nous avons simplement écrit chaque élément $q_{i,j}$ de ces matrices $\mathbf{Q}_t$ sous la forme :

$$q_{i,j} = Cov\left(X_i, X_j\right) = \chi_i \chi_j X_i X_j \quad \text{avec } \chi_i \text{ et } \chi_j \text{ constants}$$

Les essais de matrices constantes dans le temps ont tous conduit à des performance nettement plus faibles que celles du modèle GRP dans la version proposée par Tangara (2005). Il est donc nécessaire de tenir compte de l'hétéroscédasticité des erreurs du modèle.

La mise-à-jour par le filtre de Kálmán est principalement sensible à la valeur de la variance des erreurs sur le niveau du réservoir de routage (l'élément D de la matrice $\mathbf{Q}_t$ de l'annexe K, page 516). Les meilleurs résultats sont obtenus pour des valeurs de D nettement supérieures à la valeur de la variance sur les observations de débit (matrice $\mathbf{R}_t$) : nous sommes donc amenés à réaliser une mise-à-jour quasi-complète du niveau du réservoir de routage... c'est-à-dire une mise-à-jour très proche de celle proposée par Tangara (2005).

	GRP-Tangara					GRP-Kálmán				
Horizon	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
1 heure	0.00	0.07	0.16	0.24	0.39	-0.07	0.06	0.15	0.26	0.41
6 heures	0.02	0.16	0.28	0.40	0.53	0.00	0.14	0.28	0.39	0.53
12 heures	0.03	0.22	0.37	0.51	0.62	0.01	0.21	0.37	0.50	0.62
24 heures	0.04	0.29	0.46	0.60	0.70	0.03	0.28	0.46	0.60	0.70

TABLE 7.5 – Quantiles 0.05, 0.25, 0.50 (médiane), 0.75 et 0.95 des distributions des performances (critère $C_{2M,P}$) du modèle GRP avec mise-à-jour de Tangara (2005) et avec mise-à-jour de sa fonction de routage par filtre de Kálmán aux horizons de prévision 1, 6, 12 et 24 heures.

La méthode est peu sensible aux autres éléments de la matrice $\mathbf{Q}_t$, notamment aux variances des erreurs faites sur les états de l'hydrogramme unitaire. Pour une même valeur de la variance des erreurs sur le niveau du réservoir de routage, les différences de performance sont assez modestes. Ceci est très cohérent avec les résultats de la troisième partie de ce chapitre où nous avons constaté que le modèle était peu sensible à nos différents essais de mise-à-jour de l'hydrogramme unitaire. Ces résultats font écho aux conclusions de Pauwels et De Lannoy (2009) qui cherchaient à mettre à jour l'hydrogramme unitaire du modèle HBV : cette approche modifie les états sur plusieurs pas de temps alors que cela n'a qu'un impact très limité sur le débit calculé. Cette situation conduit à une faible robustesse : dans la plupart des cas, les incertitudes dues au modèle domine les incertitudes des états initiaux et il n'y a pas d'améliorations liées à l'assimilation du débit.

Qualité des estimations de l'incertitude par le filtre de Kálmán

Une des hypothèses du filtre de Kálmán est le caractère gaussien des erreurs. Celles-ci sont donc entièrement décrites par leurs moyennes (supposées nulles) et leur matrice de variance-covariance $\mathbf{P}_t$. Dans ces conditions, il est trivial de calculer un intervalle de confiance de la valeur du niveau du réservoir de routage pour n'importe quelle probabilité. Comme la relation f qui lie ce niveau au débit simulé est monotone, nous obtenons directement un intervalle de confiance sur le débit simulé.

Nous avons estimé à chaque pas de temps les intervalles de confiance à 50, 80 et 90 % du débit simulé par cette méthode. Sur chaque période, nous avons calculé la fréquence d'appartenance du débit observé à ces intervalles de confiance. Si ceux-ci sont fiables, les fréquences associées devraient être proches de 50, 80 et 90 %.

Nous en sommes loin. À l'horizon d'une heure, les fréquences d'appartenance sont nettement supérieures à ces valeurs cibles (Fig. 7.5) : le filtre de Kálmán surestime grandement les incertitudes sur les débits sortants, ce qui est dû, entre autres choses, à la nature non gaussienne et non indépendante (dans le temps) des erreurs faites par le modèle. Nous notons d'ailleurs que les intervalles de confiance sont d'autant plus grands que nous avons

choisi une variance des erreurs sur le niveau du réservoir de routage élevée (conduisant à une mise-à-jour quasi complète de cet état) : il y a ici une sorte de compromis entre qualité de la prévision déterministe (associée au vecteur des états $\mathbf{x}_{t|t}$) et fiabilité de l'estimation des incertitudes (associée à la matrice $\mathbf{P}_{t|t}$).

À mesure que l'horizon de prévision s'éloigne, la fréquence d'appartenance aux intervalles de confiance diminue. Elle passe sous les valeurs cibles (50, 80 ou 90 %) dès l'horizon de 6 heures. Ceci s'explique par le fait que cette estimation se restreint à une seule source d'incertitude, celle des états du modèle. Cette dernière devient négligeable face aux autres sources d'incertitude à mesure que l'horizon de prévision s'éloigne.

7.4.5 Discussion

L'un des grands avantages théoriques du filtre de Kálmán est qu'il n'a pas une confiance aveugle dans les observations de débit servant à la mise-à-jour. Nos tests montrent que cet avantage théorique ne conduit pas à de meilleures performances de la mise-à-jour : le filtre de Kálmán calé fournit une mise-à-jour quasi complète du niveau du réservoir de routage.

Ceci s'explique d'abord par le fait que nos données sont de meilleure qualité que celles dont le prévisionniste peut disposer en temps réel : nos données ont été examinées et validées en temps différé. Ces données contiennent donc peu d'erreurs aléatoires, c'est-à-dire celles qui sont prises en compte par le filtre de Kálmán. Seules restent les erreurs systématiques (e.g. dues aux incertitudes sur la courbe de tarage) qui non seulement ne sont pas prises en compte par les méthodes employées mais en outre interviennent lors du calcul de performance. Nous examinerons dans le chapitre 15 (page 269), la sensibilité du modèle à des erreurs dans les observations de débit servant à la mise-à-jour des états du modèle et auront de nouveau l'occasion de comparer ces différentes méthodes.

Doit-on conserver le filtre de Kálmán ?

Le filtre de Kálmán est une méthode de mise-à-jour beaucoup plus complexe que la méthode de Tangara (2005). Cependant, il ne permet pas d'obtenir de meilleures performances. Dans notre cadre de travail où nous n'assimilons que le débit observé à l'exutoire du bassin versant, il ne paraît pas utile de l'employer. Une potentialité du filtre que nous n'avons pas employée est sa capacité à assimiler simultanément différentes observations. L'étude menée par Simon Munier et Xavier Litrico (Perrin *et al.*, 2009, page 53) a permis de montrer sur un bassin versant qu'une variante du modèle GRP mis-à-jour par le filtre de Kálmán en assimilant le débit à l'exutoire et deux débits en amont obtenait de meilleurs résultats que le modèle GRP. Cela confirme l'intuition : il est probablement très intéressant d'employer de l'information provenant de l'amont et à ce titre, il apparaît nécessaire d'adapter les travaux de Lerat (2009) à la prévision des crues. Dans ce cadre, l'emploi du filtre de Kálmán peut être une solution intéressante et il serait alors utile de l'inclure parmi les approches à tester.

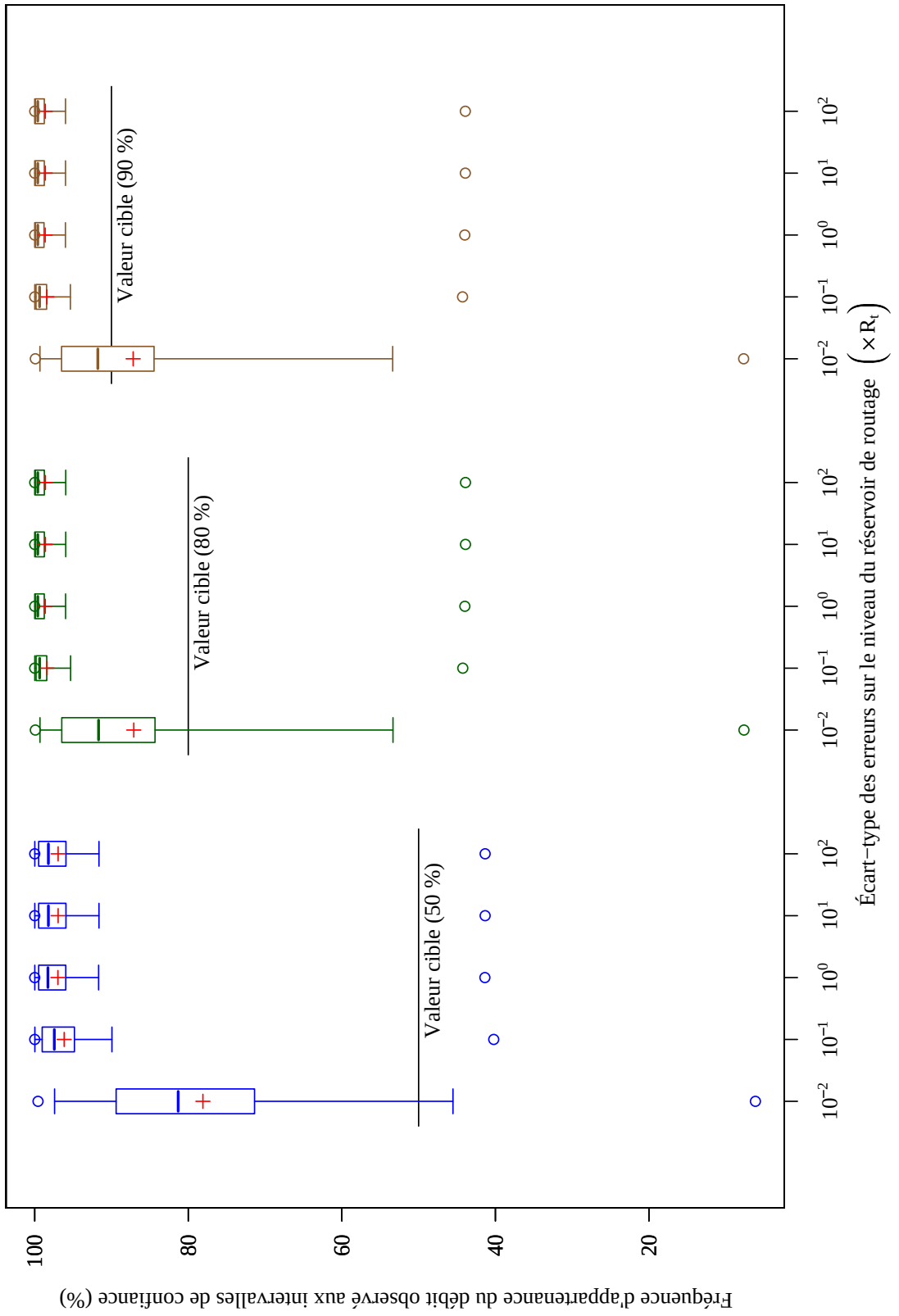


FIGURE 7.5 – Distribution (sur 178 bassins versants × 2 périodes) des fréquences d'appartenance du débit observé aux intervalles de confiance à 50, 80 et 90 % estimés par le filtre de Kálmán, pour 5 valeurs différentes de la variance des erreurs dues au modèle sur le niveau du réservoir de routage (paramètre de la matrice $\mathbf{Q}_t$). Le paramètre variance sur les observations de débit est choisi égal à $0.15^2 Q_t^2$. L'horizon de prévision est égal à 1 heure.

7.5 Mise-à-jour incomplète du niveau du réservoir de routage

Au contraire du filtre de Kálmán, la mise-à-jour du niveau du réservoir de routage proposée par Tangara (2005) est « complète » dans le sens où elle vise à obtenir un débit simulé à l'instant de prévision égal à l'observation (cf. la deuxième section de ce chapitre).

Nous proposons ici une mise-à-jour « incomplète » de ce niveau du réservoir plus simple à mettre en œuvre et moins paramétrisée que le filtre de Kálmán. Il suffit de ne plus réaliser la mise-à-jour uniquement à partir du débit observé :

$$R_{t|t}^{+} = f^{-1}\left(Q_t\right)$$

mais à partir d'une valeur du débit $\widetilde{Q}_t$ fonction du débit observé et du débit simulé par le modèle :

$$R_{t|t}^{+} = f^{-1}\left(\widetilde{Q}_t\right)$$

Nous proposons une simple moyenne pondérée $\widetilde{Q}_t = \xi Q_t + (1-\xi)\,\widehat{Q}_{t|t-1}$. Le paramètre ξ est fonction de notre confiance relative dans les observations et dans le modèle. Cette formulation très simple peut s'interpréter de façon probabiliste (Annexe K, page 517). Examinons les performances du modèle pour différentes valeurs de ξ.

7.5.1 Tests et résultats

Des valeurs de ξ allant de 0.5 (« confiances » égales dans les observations Q_t et les débits calculés $\widehat{Q}_{t|t-1}$) à 1 (confiance absolue dans les observations) ont été testées. La valeur du paramètre ξ permettant d'obtenir les meilleures performances est 1 (Fig. 7.6), et ce pour tous les horizons de prévision.

Ce résultat ne dépend pas de la réactivité des bassins versants. On note toutefois qu'aux horizons de prévision les plus courts, les performances sont un peu (mais cependant de façon significative) améliorées par cette mise-à-jour incomplète sur un petit nombre de bassins (jusqu'à 5 % de notre échantillon).

Nous observons également une forte perte de performances sur quelques autres bassins, même avec des valeurs de ξ très proche de 1.

7.5.2 Discussion

Les résultats présentés sont parfaitement cohérents avec ceux obtenus par le modèle GRP dont les états sont mis-à-jour par le filtre de Kálmán : dans les deux cas, une mise-à-jour quasi complète est nécessaire pour obtenir les meilleures performances.

Dans les deux cas, nous avons pu constater que sur un petit nombre de bassins, les performances peuvent être considérablement dégradées par une mise-à-jour « incomplète » (même

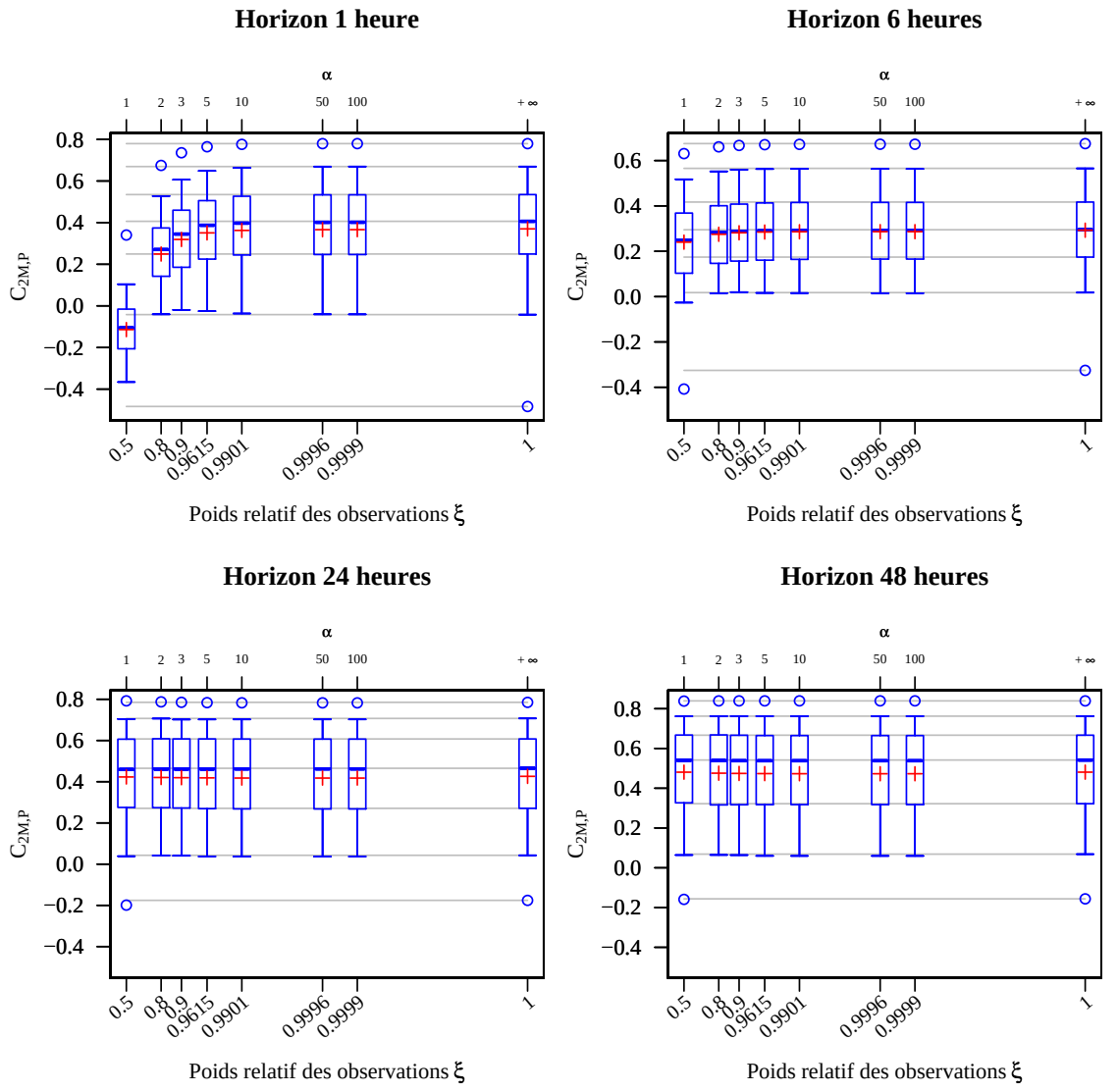


FIGURE 7.6 – Distribution (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP dont le niveau du réservoir de routage est mis-à-jour de façon « incomplète ». Les résultats sont présentés en fonction du poids relatif des observations ξ (axes des abscisses principal). La correspondance avec le rapport α des variances des distributions de probabilité du débit réel conditionnellement à l'observation et à la simulation dans la cas gaussien (cf. Annexe K, page 517) est indiquée sur l'axe des abscisses secondaires.

très peu). Cela révèle des problèmes de robustesse, et une faiblesse de la méthode à un paramètre que nous avons proposée : même si les procédures de mise-à-jour peuvent prendre en compte la possibilité d'erreurs dans les données de débit, une validation préalable reste nécessaire avant d'employer ces données pour mettre à jour le modèle.

7.6 Bilan

Nous avons testé plusieurs méthodes de mise-à-jour des états du modèle GRP. Les premières sont complètes (dans le sens où elles accordent une confiance absolue dans les observations assimilées), les secondes partielles.

Les essais réalisés n'ont pas permis de définir une stratégie de mise-à-jour meilleure que la mise-à-jour « complète » du niveau du réservoir de routage (élément le plus en aval du modèle) proposée par Tangara (2005). En particulier, le modèle se révèle peu sensible à la mise-à-jour des états de l'hydrogramme unitaire. Deux raisons peuvent expliquer ce constat. Le premier est que nous mettons les états à jour à partir de l'information de débit. L'état le mieux corrélé à ce signal est sans conteste ce niveau du réservoir de routage.

Mais cela ne suffit pas. Un deuxième facteur explicatif réside dans la dynamique de la prévision des crues et la brièveté des horizons de prévision qui nous intéressent. Il ne nous est pas utile de chercher à mettre à jour les états plus en amont dans le modèle comme le niveau du réservoir de production pour ces horizons. Une telle mise-à-jour pourrait cependant intéresser l'hydrologue cherchant à prévoir les étiages.

Les méthodes de mise-à-jour incomplète du débit testées sont le filtre de Kálmán (étendu) et une approche simple qui « probabilise » la mise-à-jour de Tangara (2005). Nous n'avons pas trouvé de paramétrisation du filtre de Kálmán permettant d'obtenir des performances supérieures à celles du modèle GRP avec la mise-à-jour de Tangara. Les différents essais de ces deux méthodes conduisent aux mêmes conclusions que les essais des méthodes « complètes » : seule la mise-à-jour du niveau du réservoir de routage est cruciale, le modèle est peu sensible aux mises-à-jour des autres états (de la fonction de routage). En outre, la mise-à-jour de ce niveau doit être quasi-complète pour permettre au modèle d'obtenir ses meilleures performances. Ce dernier point est probablement dû à la qualité des données de débit validées en temps différé que nous employons. Le chapitre 15 (page 269) sera l'occasion de tester la sensibilité de la mise-à-jour de cet état aux erreurs sur les observations de débit.

8

Chapitre

Un essai d'évaluation des incertitudes : le filtre particulaire

Quand la statistique n'est pas fondée sur des calculs rigoureusement vrais, elle égare au lieu de diriger. L'esprit se laisse prendre aisément aux faux airs d'exactitude qu'elle conserve jusque dans ses écarts, et il se repose sans trouble sur des erreurs qu'on revêt à ses yeux des formes mathématiques de la vérité.

Alexis de Tocqueville, *de la Démocratie en Amérique*

Objectifs

- ▷ Réaliser une étude exploratoire d'une méthode d'évaluation des incertitudes de prévision.
- ▷ Examiner les interactions entre mises-à-jour du modèle et évaluation des incertitudes

Sommaire

8.1 Rapide état de l'art 135
8.1.1 Écoles fréquentiste et bayésienne 135
8.1.2 Différentes sources d'incertitude 136
8.1.3 Quelles sources d'incertitudes prendre en compte ? 137
8.1.4 Implicite ou explicite ? 137
8.2 Du cadre bayésien au filtre particulaire 138
8.2.1 Éléments clefs du filtre particulaire 138
8.2.2 Avantages et limites du filtre particulaire 139
8.3 Associer le filtre particulaire et le modèle déterministe GRP . 139
8.3.1 Problématiques étudiées 139
8.3.2 Évaluation des incertitudes et des performances 140
8.3.3 Mise en place pratique 140
8.4 Expérimentations et discussion 144
8.4.1 Premières expériences : peut-on ne tenir compte que de l'incertitude sur les pluies ? 144
8.4.2 Perturbation des entrées de pluie et de débit 148
8.4.3 Discussion et tests complémentaires 150
8.5 Bilan 154

Chercher à mettre à jour un modèle revient à reconnaître une incertitude non négligeable dans ses sorties. C'est pourquoi les techniques de mise-à-jour, qu'elles soient déterministes (i.e., elles modifient tel ou tel élément du modèle pour lui donner une voire sa « meilleure » valeur étant données les observations) ou probabilistes, sont intrinsèquement liées au concept d'incertitude (Liu et Gupta, 2007). Dans ce mémoire de doctorat, nous présentons principalement des méthodes de mise-à-jour déterministe. Certaines sont accompagnées d'une interprétation probabiliste (par exemple, le filtre de Kálmán présenté au chapitre 7).

Dans ce chapitre, nous nous intéressons à la question de l'estimation des incertitudes de prévision : ce n'est plus la mise-à-jour d'un ou plusieurs composants du modèle déterministe qui est en jeu, mais la mise-à-jour de notre estimation de ses incertitudes.

L'étude des estimations des incertitudes en hydrologie est un vaste domaine de recherche, auquel de nombreux travaux sont consacrés (Montanari *et al.*, 2009). Ce chapitre se limite à une étude exploratoire d'une méthode, le filtre particulaire, employée avec le modèle GRP. Nous nous pencherons en particulier sur les interactions entre estimation des incertitudes et mise-à-jour déterministe du modèle.

Nous entamerons ce chapitre par un état de l'art rapide et forcément incomplet des approches probabilistes d'évaluation des incertitudes employées en hydrologie. Nous présenterons ensuite l'approche bayésienne d'estimation des incertitudes des états du modèle pour introduire le fonctionnement du filtre particulaire. La troisième section se concentrera sur les questions soulevées par l'application de cette méthode au modèle GRP. Le chapitre se poursuivra enfin par la présentation des essais que nous avons réalisés et par des éléments de réponse à ces questions.

8.1 Rapide état de l'art

L'évaluation des incertitudes d'un modèle a fait l'objet d'un très grand nombre de contributions. En rendre compte dépasse largement le sujet qui nous intéresse. Nous nous contentons ici de mettre en exergue quelques points qui nous paraissent particulièrement importants pour la prévision des crues.

8.1.1 Écoles fréquentiste et bayésienne

Deux grandes familles d'approches existent : il s'agit de l'école fréquentiste et l'école bayésienne (Gallagher et Doherty, 2007). Ces deux approches partent de deux représentations conceptuelles bien différentes :

Approche fréquentiste
Les paramètres sont des entités fixes dont la valeur « vraie » nous est inconnue pour diverses

raisons. Il est nécessaire de prendre en compte une incertitude sur leurs valeurs due au bruit dans les données, aux erreurs de modélisation, etc. Des exemples de méthodes fréquentistes sont l'approximation au premier ordre dont les limites sont discutées par Kuczera (1988), ou la méthode basée sur le modèle méta-gaussien proposée par Montanari et Grossi (2008).

Approche bayésienne

Les paramètres sont représentés par des variables aléatoires et sont donc décrits par des densités de probabilité. Le théorème de Bayes permet de faire évoluer notre connaissance de cette densité en partant d'une connaissance antérieure à une ou plusieurs observations (le *prior*) pour obtenir une densité *a posteriori* grâce à l'extraction de l'information contenue dans ce jeu de données. Comme nous le verrons dans les pages suivantes, de nombreuses méthodes bayésiennes ou s'inspirant de la démarche bayésienne existent. Parmi les plus connues, on peut citer les méthodes de Monte-Carlo par chaîne de Markov (e.g. Kuczera et Parent, 1998), BaRE (Thiemann *et al.*, 2001) ou encore la méthode GLUE (Beven et Binley, 1992, Beven et Freer, 2001).

8.1.2 Différentes sources d'incertitude

Les sources « primaires »

La modélisation hydrologique comporte plusieurs sources d'incertitudes « primaires ». En premier lieu, les données d'entrée sont souvent mal estimées : aux erreurs de mesure, il faut rajouter les problèmes de non représentativité des mesures due par exemple aux changements d'échelle et de spatialisation des entrées. La pluie de bassin est par exemple, une variable difficile à estimer. En outre, les prévisions de pluie future sont encore plus incertaines.

À ces incertitudes, se rajoutent celles liées au modèle : erreurs de conceptualisation, simplification des équations... Ces erreurs sont évidemment mal connues et donc mal décrites (Krzysztofowicz, 1999, Kavetski *et al.*, 2006a).

Les erreurs de mesure des sorties (pour nous le débit) participent également à l'incertitude des sorties. En effet, ces erreurs expliquent une partie de l'incertitude sur les paramètres du modèle quand ceux-ci sont calés (d'autres sources d'incertitude sur les paramètres sont les erreurs de conceptualisation, l'emploi d'une fonction-objectif incorrecte au sens d'une fonction de vraisemblance, etc.).

Conséquences sur l'incertitude des prévisions

Le modélisateur doit donc tenir compte de l'incertitude sur les sorties qui en résulte. Il peut le faire à plusieurs niveaux : sur les entrées du système mais également par une description des incertitudes engendrées en interne au sein du modèle sur ses paramètres, ou encore sur ses états.

8.1.3 Quelles sources d'incertitudes prendre en compte ?

Plusieurs approches très différentes sont proposées pour quantifier les incertitudes de prévision. La première étape réside souvent dans l'identification des sources d'incertitude dominantes. Celles-ci sont fonction de l'horizon de prévision (e.g. Velázquez *et al.*, 2009). Certains auteurs pointent les incertitudes sur les prévisions de pluie (e.g. Garçon *et al.*, 2008). D'autres se concentrent sur l'incertitude de modélisation.

De nombreuses approches font porter toute l'incertitude des sorties du modèle sur une seule source. Celle-ci peut être les paramètres (c'est le cas des méthodes GLUE ou BaRE par exemple), ou les entrées du modèle comme le font les méthodes ensemblistes. Ces approches sont souvent limitées par le fait que considérer une seule source n'est pas suffisant : les résultats ne sont pas cohérents sauf à surestimer l'unique source d'incertitudes prise en compte (cf. la critique formulée par Krzysztofowicz, 1999, concernant les méthodes ensemblistes – e.g. de Roo *et al.*, 2003). En outre, la source choisie n'est pas forcément la source d'incertitudes dominante (Kuczera, 1988).

Krzysztofowicz (1999) puis Krzysztofowicz et Maranzano (2004) proposent de considérer deux sources d'incertitude : la première concerne les entrées du modèle (source que ces auteurs estiment dominante, dans le cas d'un modèle de simulation Pluie – Débit) quand la seconde correspond à l'agrégation des autres sources d'incertitudes dans un « processeur des incertitudes hydrologiques ».

Au contraire, d'autres auteurs préfèrent expliciter toutes les sources d'incertitude prises en compte. Il s'agit de l'approche défendue par la méthode BATEA (Kavetski *et al.*, 2006a). Certains cherchent à rendre compte simultanément de l'incertitude sur les paramètres et sur les états du modèle (Vrugt *et al.*, 2005, Moradkhani *et al.*, 2005a). Ces méthodes peuvent être grevées par des problèmes d'interaction entre les différentes sources d'incertitude considérées. La question des dynamiques d'évolution de ces objets (paramètres et états) se pose également : comment mettre à jour simultanément les densités de probabilité d'objets invariants (ou variant très peu dans le temps) et celles d'objets variant rapidement dans le temps ?

8.1.4 Implicite ou explicite ?

La prise en compte des différentes sources d'incertitude peut être implicite ou explicite. Les tenants d'une expression explicite des sources d'incertitude s'appuient sur la rigueur de l'approche et la possibilité de réfuter les hypothèses précises faites sur le modèle d'erreur (Kavetski *et al.*, 2006b). Les avocats d'une prise en compte (au moins partiellement) implicite rappellent l'extrême difficulté qu'il y a à valider un modèle d'erreur et à fournir une fonction de vraisemblance exacte. Ils plaident donc pour une approche plus simple pouvant être mise en œuvre en pratique. Ce débat est récurrent (il apparaît par exemple comme

un des points discutés autour de la méthode GLUE par Mantovan et Todini, 2006, Beven *et al.*, 2008, parmi d'autres).

8.2 Du cadre bayésien au filtre particulaire

Nous ne disposons pas de séries homogènes de prévisions de précipitation sur de longues périodes. Nous allons donc nous concentrer sur l'estimation de l'incertitude sur les états initiaux du modèle pour la prévision, c'est-à-dire sur les états à l'instant de prévision. C'est le but de nombreuses méthodes comme le filtre de Kálmán ou le filtre particulaire. C'est vers ce dernier que nous nous tournons. Cette approche ne présente pas certains défauts (en hydrologie) des filtres de Kálmán : elle pose moins d'hypothèses sur les distributions des états. Cette approche a été de plus en plus étudiée en hydrologie au cours de ces dernières années (Moradkhani *et al.*, 2005a, Weerts et El Serafy, 2006, Smith *et al.*, 2008, Salamon et Feyen, 2009).

Le filtre particulaire est une méthode d'inférence bayésienne. Nous ne présentons pas le principe et le fonctionnement de cette approche en détail ici, mais préférons nous concentrer sur son application au modèle GRP. Le lecteur intéressé peut se reporter à l'annexe L (à partir de la page 523) pour une présentation détaillée du filtre particulaire.

8.2.1 Éléments clefs du filtre particulaire

L'approche bayésienne décrit les états du modèle par des densités de probabilité. Aussi, la mise-à-jour porte sur l'incertitude (ici des états) à travers ces densités et non sur leur valeur la plus probable. Comme toutes les méthodes de mise-à-jour présentées jusqu'ici dans ce mémoire, le filtre particulaire fonctionne en deux temps :

1. propagation : emploi du modèle GRP pour passer d'un pas de temps au suivant ;
2. mise-à-jour : assimilation de données à l'aide du théorème de Bayes.

Les équations décrivant ce cycle (Ann. L) ne sont que très rarement solubles analytiquement. Le filtre particulaire est une solution numérique : les densités de probabilité des vecteurs des états sont approximées par un ensemble fini de vecteurs d'état auxquels sont associés des poids. La distribution (discrète) de ces états est proche de la distribution continue recherchée. On peut montrer que la solution fournie par le filtre particulaire converge vers la solution analytique de l'inférence bayésienne quand le nombre de particules tend vers l'infini.

L'étape de propagation est réalisée sur chaque particule de l'ensemble (évolution du vecteur des états de chaque particule d'un pas de temps à l'autre), tandis que la mise-à-jour concerne les poids. Le poids d'une particule est modifié en fonction de l'accord de cette particule à l'observation : à une particule qui « colle » aux observations, un poids élevé, à

une particule éloignée des observations, un poids faible.

Dans la pratique, le filtre particulaire souffre d'un mal récurrent : la dégénérescence. Après un certain nombre d'itérations, il reste un faible nombre de particules dont les poids ne sont pas négligeables. En conséquence, la distribution discrète représente très mal la densité de probabilité recherchée. Une méthode, appelée rééchantillonnage, permet de lutter contre la dégénérescence. Cette technique peut par contre conduire à un appauvrissement de l'échantillon, c'est-à-dire à un échantillon comptant un grand nombre de fois les mêmes particules. Cependant, la prise en compte des incertitudes sur les entrées à travers une perturbation de celles-ci lors de l'étape de propagation réduit considérablement les risques d'appauvrissement.

8.2.2 Avantages et limites du filtre particulaire

Pour Moradkhani *et al.* (2005a), les principaux avantages du filtre particulaire résident dans sa capacité à représenter de façon plutôt complète (non limitée aux deux premiers moments statistiques) les densités de probabilité et à bien prendre en compte les non linéarités : à la différence du filtre de Kálmán d'ensemble, il s'agit d'une inférence bayésienne complète.

Parmi ses défauts, il nous faut mentionner qu'il est très difficile de déterminer le nombre minimal de particules permettant d'assurer une approximation acceptable de la densité *a posteriori* par le filtre (Gordon *et al.*, 1993). Nous mettrons en évidence d'autres difficultés plus spécifiques à la prévision hydrologique lors de la présentation de la mise en pratique du filtre sur le modèle GRP (suite de ce chapitre).

8.3 Associer le filtre particulaire et le modèle déterministe GRP

Nous nous sommes basés sur la littérature présentée précédemment (ainsi que dans l'annexe L) et en particulier sur les travaux de Gordon *et al.* (1993) et Weerts et El Serafy (2006) pour réaliser cette association.

8.3.1 Problématiques étudiées

Les essais réalisés ont pour but d'étudier principalement deux questions liées à l'estimation des incertitudes sur les débits prévus :

Quelle interaction entre mise-à-jour déterministe du modèle GRP et estimation des incertitudes ?
L'un des éléments clefs du modèle GRP est l'assimilation du débit, conduisant notamment à la mise-à-jour du niveau du réservoir de routage. En ce sens, le modèle est plus un modèle (Pluie, Débit) – Débit qu'un modèle Pluie – Débit. La mise-à-jour contraint fortement un

des états principaux du modèle. Elle réduit donc fortement les incertitudes sur les états dues aux incertitudes sur les entrées de précipitation à l'instant de prévision... Et par conséquent, les incertitudes sur les débits en sortie. Nous examinerons donc à quelle date passée, il vaut mieux faire cette mise-à-jour. Cette question est cruciale en ce qui concerne le modèle GRP car, selon l'instant choisi pour effectuer la mise-à-jour, la précision des prévisions ainsi que leur fiabilité n'atteindront pas du tout les mêmes niveaux.

Quelles incertitudes sur les entrées devons-nous prendre en compte ?
De nombreuses méthodes d'estimation des incertitudes se basent uniquement sur les incertitudes sur les entrées du modèle de simulation (i.e., les précipitations passées). Est-ce suffisant ou faut-il tenir compte des incertitudes sur le débit qui est assimilé par la méthode ?

8.3.2 Évaluation des incertitudes et des performances

L'évaluation du filtre particulaire ne peut se faire que de façon indirecte. En effet, cette approche vise à estimer la densité de probabilité des états à l'instant de prévision t. Or les états ne nous sont pas accessibles par l'observation. Nous pourrions examiner l'incertitude des débits simulés à t par le modèle mais l'objectif pratique reste d'estimer les incertitudes de prévision du débit $\widehat{Q}_{t+H|t}$ prévu à un horizon H. Nous allons donc évaluer la méthode testée par l'estimation des incertitudes de prévision.

Deux aspects de la prévision probabiliste sont évalués. Tout d'abord, la précision(g) des prévisions (i.e. de la médiane ou de la moyenne de l'ensemble), sera comparée à la précision du modèle déterministe (avec la mise-à-jour des états à la date t) à travers le critère de performance $C_{2M,P}$. Le CRPS (e.g. Bontron, 2004, également présenté dans l'annexe L, page 523) sera en outre comparé à l'erreur absolue moyenne du membre non perturbé.

La fiabilité(g) sera ensuite examinée à travers la fréquence d'appartenance du débit observé à six intervalles de confiance : à 50 %, 80 % et 90 % calculés à tous les instants de prévision disponibles, puis calculés sur les instants de prévision précédés d'un cumul de pluie sur Δt supérieur à son quantile 0.95 pour évaluer la fiabilité sur les événements pluvieux, qui est souvent moins bonne.

Enfin, nous n'examinerons pas un autre critère classique : la perfection de l'ensemble (critère défini dans l'annexe L, page 523, qui indique la fiabilité de l'ensemble en fonction de l'horizon de prévision). En effet, nos résultats ne sont pas suffisamment encourageants pour étudier cet aspect des prévisions.

8.3.3 Mise en place pratique

Le filtre particulaire a été appliqué au modèle GRP fonctionnant sans les corrections en sortie. Le niveau du réservoir de routage est mis-à-jour par la méthode de Tangara (2005)

dans la plupart des essais réalisés (détaillés ci-dessous). Nous avons travaillé sur notre échantillon de 178 bassins versants malgré l'importance du temps de calcul associé au nombre élevé de particules [1]. Nous avons dû par contre nous restreindre pour la plupart des essais à un seul horizon que nous avons choisi égal à 6 heures.

Perturbation des entrées

Les entrées du modèle pour chaque particule peuvent être bruitées afin de rendre compte de l'incertitude sur ces entrées.

Le bruitage des pluies de la i^{e} particule a été réalisé en multipliant les pluies sur une plage $[t - \Delta t, t]$ par un facteur $\gamma_{i,t}$ tiré dans une loi Γ, pour tout instant de prévision t. La loi Γ a été choisie de moyenne 1 (seul un paramètre de cette loi est laissé libre pour jouer sur sa dispersion). La durée Δt est fixée égale au temps de base calé (paramètre X_3) de l'hydrogramme unitaire du modèle GRP. L'approche est apparentée à celle des *storm multipliers* proposée par Kavetski *et al.* (2006a).

Dans certains essais, le débit Q_t assimilé pour mettre à jour les états du modèle, est également bruité. Nous avons opté pour une loi multiplicative très simple : une loi normale de moyenne 1 et d'écart-type σ_Q à évaluer (les rares valeurs négatives sont ramenées à 0).

Faut-il perturber des pluies futures ?

Comme nous l'avons vu, le filtre particulaire vise à estimer l'incertitude à l'instant de prévision : il s'intéresse donc aux entrées précédant cet instant. Cependant, les pluies futures (précipitées entre l'instant de prévision t et le pas de temps cible $t + H$) peuvent peser fortement sur l'incertitude de prévision. Nous ne disposons pas de prévisions quantitatives de pluie sur nos 178 bassins versants mais nous pourrions chercher en pratique à prolonger les perturbations générées sur les pluies passées aux précipitations à venir, bien que nous sachions que les incertitudes sur les pluies futures sont par nature bien supérieures à celles sur les pluies observées. Nous testerons en outre des prévisions de précipitation d'ensemble dont nous disposons sur 28 bassins.

Tenir compte de cette source d'incertitude pourrait amoindrir la robustesse des estimations car une fois la densité de probabilité des états estimée à l'instant de prévision t (à travers le calcul des poids des particules), rien ne vient évaluer la pertinence de ces perturbations et donc nous perdons l'assurance fournie par le cadre bayésien. Ne pas le faire participerait au resserrement des incertitudes estimées en sortie, dès que l'horizon de prévision est suffisamment lointain pour que les pluies passées (ou dit autrement, l'état du modèle à l'instant de la prévision) aient une influence mineure sur les sorties du modèle à l'horizon de prévision. Les deux stratégies seront testées ici.

1. À ce sujet, je renouvelle ici mes remerciements à Vincent Fortin d'Environnement Canada, non seulement pour son accueil et la collaboration scientifique qui a permis la mise en place de cette étude, mais également pour les capacités de calcul québécoises que nous avons pu employer.

Nombre de particules

Les essais où seules les entrées de précipitation sont bruitées comptent 50 particules, auxquelles s'ajoute le membre de contrôle (non perturbé). Quant aux essais où les pluies et les débits sont perturbés, ils s'appuient sur 601 particules correspondant à 20 perturbations du débit, 30 perturbations des pluies et au membre de contrôle.

Calcul des poids et rééchantillonnage

À chaque pas de temps, les particules sont rééchantillonnées selon la procédure du rééchantillonnage résiduel (RR) décrite en annexe L. Les poids sont calculés par l'équation (Weerts et El Serafy, 2006) :

$$w_t^{(i)} \propto \mathrm{e}^{-\frac{\left(\widehat{Q}_t^{(i)} - Q_t\right)^2}{2 \cdot \sigma_W^2}} \quad \text{avec } \sigma_W = \min\left(0.05\overline{Q}, \sigma_M \cdot Q\right) \tag{8.1}$$

Cette forme est loin d'être une fonction de vraisemblance valable pour les sorties du modèle. Nous avons également testé une forme basée sur les transformées logarithmiques des débits pour lutter contre l'hétéroscédasticité mais nous n'avons pas pris en compte la dépendance temporelle des erreurs sur les débits observés et simulés.

Sélection des particules pour la prévision

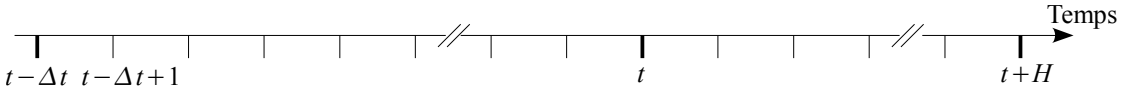
Diverses approches empiriques ont été testées pour sélectionner les particules : prise en compte de toutes les particules ou des particules de plus forts poids dont le cumul atteint 0.9 (pour éliminer les particules les moins probables).

Paramétrisation

Le temps de calcul prohibitif nous a amené à employer les paramètres du modèle GRP (fonctionnant sans module de correction des sorties) calés au préalable. Le filtre particulaire ajoute trois paramètres : un paramètre de la loi Γ bruitant les entrées de précipitation jouant sur la dispersion de cette loi, l'écart-type σ_Q des perturbations sur les débits assimilés par le modèle et la variance du calcul des poids σ_M^2. Une étude de sensibilité nous permettra de ne considérer que les paramètres ayant une réelle importance, au regard de nos objectifs. Ces paramètres ne bénéficieront pas d'un réel calage mais nous explorerons l'espace de ces paramètres par quelques dizaines d'exécutions.

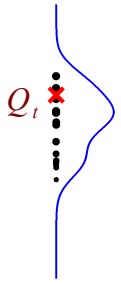
Résumé de l'implémentation du filtre particulaire

La figure 8.1 présente le fonctionnement général du filtre. t est l'instant de prévision et H l'horizon de prévision. Les points correspondent aux particules ; les tailles des points sont proportionnelles aux poids des particules. Les courbes représentent les densités de probabilité approximées par les ensembles de particules. La croix rouge (étape 2) indique

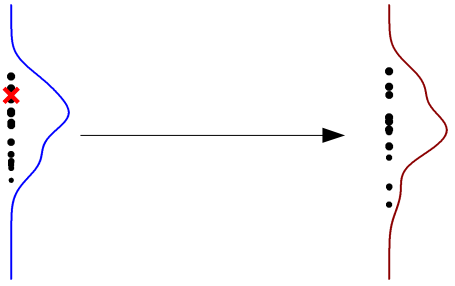


1. Propagation des états (avec perturbations des entrées de pluie)

2. Mise-à-jour de la densité de probabilité des états (calcul des poids)



3. Prévision



4. Rééchantillonnage et passage au pas de temps suivant

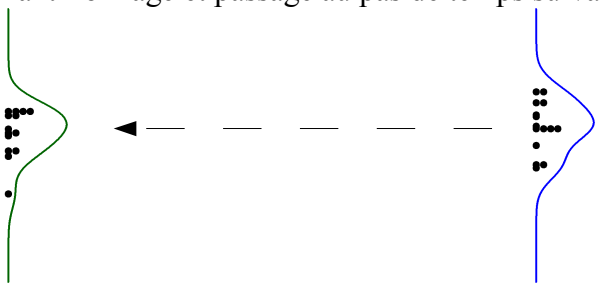


Figure 8.1 – Schéma général de fonctionnement du filtre particulaire avec GRP. La légende et le fonctionnement sont détaillés dans le paragraphe intitulé *Résumé de l'implémentation du filtre particulaire* page 142.

l'observation Q_t. Les flèches continues représentent la propagation des particules, c'est-à-dire l'évolution de chaque particule par l'opérateur f (le modèle). Enfin, la flèche pointillée représente la sélection des particules au pas de temps $t-\Delta t+1$ conformément au rééchantillonnage effectué au pas de temps t.

En partant d'un échantillon de particules équiprobables approximant la densité de probabilité des états au pas de temps $t-\Delta t$, le fonctionnement du filtre particulaire peut être détaillé en quatre étapes représentées sur la figure 8.1 :
(1) Les particules de cet échantillon évoluent sur la plage $[t-\Delta t, t]$ selon le modèle GRP dont les entrées sont bruitées. L'échantillon des particules à l'instant $t-\Delta t+1$ est conservé en mémoire.
(2) Au pas de temps t, la dernière observation de débit Q_t (la croix rouge) est employée pour mettre à jour la densité de probabilité des états : les poids $w_t^{(i)}$ sont calculés.
(3) Les particules de cet échantillon mis-à-jour évoluent jusqu'à la date $t+H$ pour laquelle la prévision est émise. La densité de probabilité du débit Q_{t+H} est estimée (à partir des débits calculés pour chaque particule et des poids estimés à l'étape précédente).
(4) L'échantillon de particules mis-à-jour à la date t est rééchantillonné : on aboutit à un nouvel échantillon de particules équiprobables approximant la même densité. Les particules de l'échantillon à la date $t-\Delta t+1$ correspondantes sont sélectionnées pour créer l'échantillon rééchantillonné et mis-à-jour permettant de démarrer le même processus au pas de temps suivant.

8.4 Expérimentations et discussion

8.4.1 Premières expériences : peut-on ne tenir compte que de l'incertitude sur les pluies ?

Dans un premier temps, seules les entrées de précipitation passées sont bruitées. Les perturbations multiplicatives sont issues d'une loi Γ de moyenne 1 et de facteur d'échelle 5 (ce qui donne un écart-type de 0.45).

La difficulté majeure dans ce cadre réside dans le choix de l'instant de la mise-à-jour des états du modèle par rapport à l'instant de prévision où sont calculés les poids des différentes particules. Trois tests sont menés (cf. Fig. 8.2) :

(a) la mise-à-jour est réalisée juste avant le calcul des poids (sur $\widehat{Q}_{t|t-1}$) ;
(b) la mise-à-jour est réalisée au début de la période de perturbations des pluies, c'est-à-dire à $t-\Delta t$: cette approche permet aux particules de « s'imprégner » pleinement des perturbations sur les entrées ;
(c) la mise-à-jour n'est réalisée qu'après le calcul des poids et n'influence donc pas l'estimation des poids des particules mais contraint fortement les prévisions.

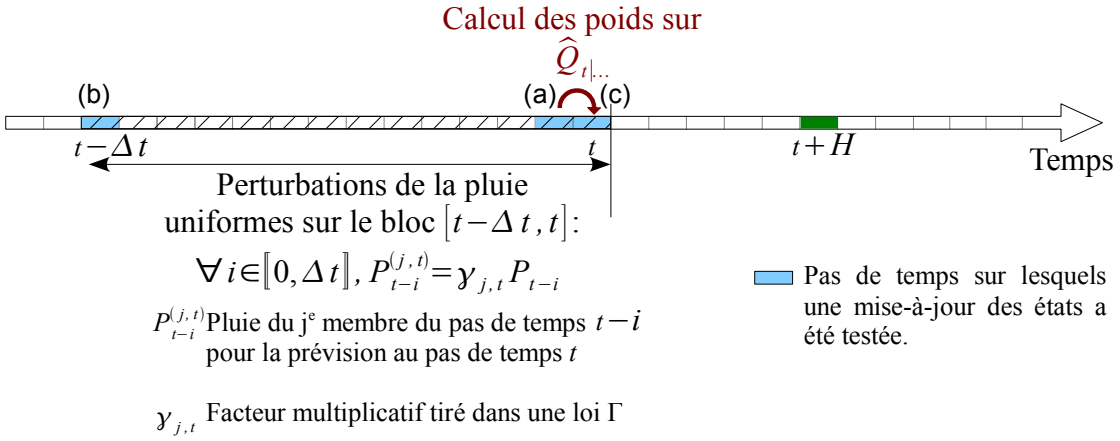


Figure 8.2 – Premiers essais d'implémentation du filtre particulaire : seules les entrées de pluie sont bruitées. Différents instants de mise-à-jour des états sont testés : (a) juste avant le calcul des poids des particules, (b) en $t-\Delta t$: au démarrage des perturbations des pluies et (c) après le calcul des poids.

Précision de l'ensemble

Quand la mise-à-jour du niveau du réservoir de routage est réalisée juste avant le calcul des poids des particules (a), les performances de la moyenne ou de la médiane de l'ensemble mesurées par le critère $C_{2M,P}$ sont légèrement mais significativement inférieures à celles du modèle déterministe (qui réalise la mise-à-jour du niveau du réservoir de routage à l'instant de prévision). Ce n'est plus le cas si la mise-à-jour est réalisée à l'instant de prévision, après le calcul des poids (c). Enfin, dans le troisième test (b) où la mise-à-jour est placée avant la plage de bruitage des entrées de pluie, nous observons sans surprise une très forte diminution de la précision de l'ensemble (Tab. 8.1). Ces résultats sont cohérents avec les études présentées plus loin dans ce mémoire : les performances du modèle GRP reposent en grande partie sur l'assimilation du dernier débit observé (Annexe N, page 541 et chapitre 15, page 269).

$\Delta C_{2M,P}$	Cas(a)	Cas (b)	Cas (c)
Moyenne de l'ensemble	-0.08	-1.19	-0.02
Médiane de l'ensemble	-0.08	-1.19	-0.02

Table 8.1 – Différences des performances mesurées à l'aide du critère $C_{2M,P}$ de GRP avec filtre particulaire et de GRP « déterministe » quand seules les entrées de pluie sont bruitées. Trois cas sont testés : (a) la mise-à-jour des états du modèle est réalisée juste avant le calcul des poids des particules ; (b) la mise-à-jour est réalisée au début de la plage de bruitage des données de pluie ; (c) la mise-à-jour est réalisée à l'instant de prévision (après le calcul des poids). L'horizon de prévision est ici de 6 heures et l'écart-type σ_M est fixé à 0.15.

L'essai (a) a été répété à différents horizons de prévision (1, 12, 24 et 48 heures) sans qu'on observe des différences significatives. De même, ce test a été réalisé avec différentes variances σ_M^2. Une certaine sensibilité apparaît : les résultats fléchissent de 5 points de $C_{2M,P}$ en moyenne si la valeur choisie est trop faible ($\sigma_M \leq 0.05$), c'est-à-dire si on sous-estime le bruit dans les observations de débit.

Dans les expériences (a) et (c), le CRPS des ensembles est légèrement inférieur aux erreurs absolues moyennes de la médiane de l'ensemble ou de sa moyenne ou encore du modèle déterministe : la différence est de l'ordre de 5 % de cette erreur absolue moyenne. La situation change du tout au tout dans le cas de la configuration (b) : le CRPS est nettement inférieur à l'erreur absolue moyenne de la médiane ou de la moyenne de l'ensemble mais est logiquement très supérieur à l'erreur absolue moyenne du modèle déterministe. L'ensemble est plus informatif qu'une prévision déterministe faite par le modèle GRP avec une mise-à-jour des états établie à $t - \Delta t$ mais nous renseigne nettement moins que le modèle GRP avec mise-à-jour à l'instant de prévision t.

Fiabilité

Quelque soit la configuration testée (a), (b) ou (c), la méthode employée sous-estime grandement les incertitudes du modèle (Fig. 8.3). Par exemple dans la première configuration testée, la médiane des fréquences d'appartenance à l'intervalle de confiance à 90 % est de 6 % quand elle devrait être 15 fois plus grande ! On note que les incertitudes sont un peu moins sous-estimées sur les événements pluvieux : ceci est attendu car seules les incertitudes sur les pluies sont prises en compte ici. La configuration (b), dans laquelle les perturbations sur les entrées ne sont pas du tout effacées par une mise-à-jour des états, conduit à de meilleures estimations des incertitudes sur les bassins où ces estimations étaient les moins mauvaises sans que les résultats ne soient satisfaisants.

Si la stratégie de mise-à-jour du modèle choisie influe sur la fiabilité des prévisions, il ne s'agit pas du principal facteur expliquant l'échec de ces premiers essais. Il apparaît nécessaire de prendre en compte les incertitudes sur les débits en bruitant également les données de débit qui seront assimilées au sein du modèle GRP. Cette conclusion est cohérente avec notre connaissance de l'importance de la donnée de débit au sein du modèle GRP.

Influence du niveau de bruit dans les entrées de précipitation

Le schéma (c) a été exécuté avec un bruitage des pluies plus fort : des lois Γ de facteur d'échelle 3 et 2 (correspondant à des écarts-types de 0.58 et 0.71). Plus le bruit sur les données de pluie est fort, plus les fréquences d'appartenance aux intervalles de confiance sont élevées. Cependant, malgré un bruit intense sur les pluies, les résultats restent proches des précédents : les incertitudes restent largement sous-estimées.

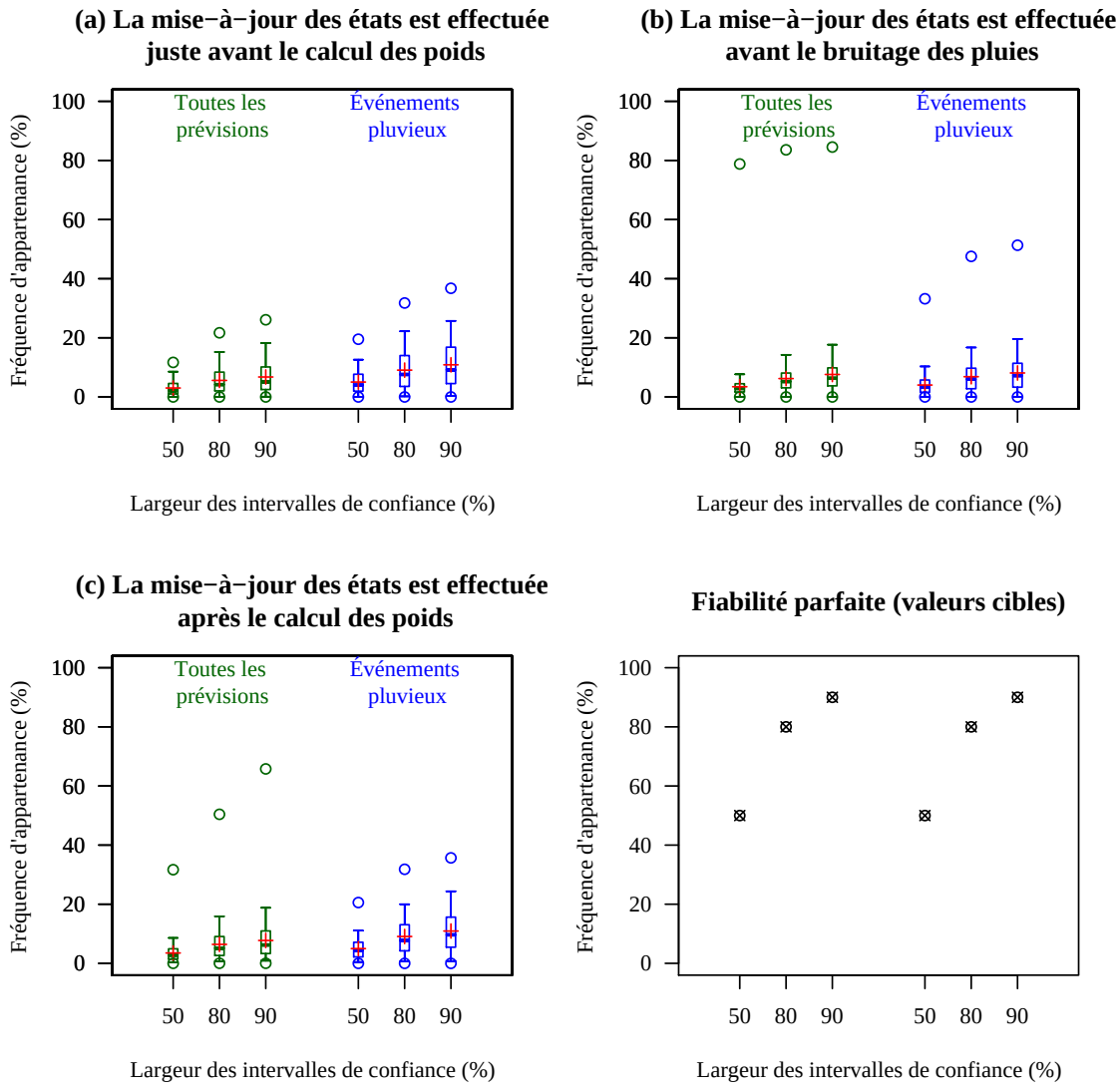


FIGURE 8.3 – Distribution (sur 178 bassins versants × 2 périodes) des fréquences d'appartenance aux intervalles de confiance à 50, 80 et 90 % calculées sur l'ensemble des chroniques de prévision ou uniquement sur les événements fortement pluvieux dans les trois configurations (a), (b) et (c) testées. L'horizon est de 6 heures et σ_M est fixée à 0.2. Le graphe en bas à droite représente la situation idéale où les distributions sont concentrées sur les valeurs cibles : 50, 80 et 90 %.

8.4.2 Perturbation des entrées de pluie et de débit

Cette série d'essais se caractérise par une perturbation des entrées de pluie (identique aux précédentes) et des entrées de débit assimilées.

Tests réalisés

Dans un premier temps, pour que les perturbations sur les entrées de pluie aient le temps d'agir, on place la mise-à-jour au début de la période de rétroaction (configuration (b)).

Des tests préliminaires sur différents niveaux de perturbation des entrées de précipitation ont montré que le modèle était assez peu sensible à ce niveau. Nous avons décidé de nous concentrer sur les valeurs de σ_Q (caractérisant le niveau de bruit sur les entrées de débit) et σ_M (variance du calcul des poids). Un calage manuel (et grossier) par essais-erreurs a été réalisé en explorant le plan de ces deux paramètres : les valeurs de σ_Q sont comprises entre 0,05 et 0,70 quand celles de σ_M sont choisies entre 0,15 et 1,5. Pour chaque bassin, nous prenons le meilleur jeu de paramètres (σ_Q, σ_M) calés sur la période P_1 pour contrôler sur la période P_2 (et vice-versa).

Enfin, des essais ont été réalisés avec quelques valeurs du couple de paramètres en modifiant le calcul des poids : ce dernier se fait sur les transformées logarithmiques des débits, afin de limiter l'impact de l'hétéroscédasticité des débits et de se rapprocher d'une fonction de vraisemblance correcte.

Résultats du calage

Deux « calages » ont été réalisés. Le premier est basé sur la fiabilité des intervalles de confiance à 50, 80 et 90 % sur tous les pas de temps. Le second ne s'intéresse qu'aux pas de temps précédés d'un fort cumul pluvieux (supérieur au quantile 0.95). Dans le premier cas, l'écart-type du bruit sur les débits assimilés σ_Q est calé à des valeurs supérieures à 0.4 pour la plupart des bassins versants, ce qui paraît peu réaliste au regard de la qualité des données (validées) dont nous disposons. Il est le plus souvent admis que les erreurs aléatoires relatives de mesure du débit sont plutôt de l'ordre de 10 à 20 % (voir le chapitre 15). Le paramètre σ_M le plus souvent calé autour de 0.85 : la fonction de vraisemblance est loin d'être piquée autour de l'observation. Ceci peut signifier que le filtre particulaire a besoin de quasiment toutes les particules pour rendre compte de l'incertitude et donc que les incertitudes prises en compte en entrée, bien que fortes, ne suffisent pas à rendre compte des incertitudes de prévision. Trois facteurs peuvent expliquer ces observations : l'importance des erreurs structurelles, le rôle de la mise-à-jour qui contraint trop fortement le niveau du réservoir de routage (car cet état conditionne fortement les prévisions à horizon court) et enfin, l'absence de prise en compte de l'incertitude sur les pluies futures. Ces résultats sont cependant loin de permettre une conclusion car le calage a été plus que « parcellaire ».

Quand on ne considère que les événements de pluie, ces valeurs grimpent encore : le filtre particulaire se tord pour obtenir des intervalles de confiance suffisamment larges pour être fiables.

Précision du modèle

Nous avons choisi de mettre à jour le niveau du réservoir de routage uniquement au pas de temps $t - \Delta t$ afin de ne pas « effacer » partiellement les effets du bruit sur les pluies. Les performances du modèle sont donc très nettement inférieures à celles du modèle GRP quand cet état est mis-à-jour à l'instant de prévision t. La précision des prévisions médianes ou moyennes de l'ensemble est significativement inférieure à celle du modèle de persistance sur 95 % des bassins : ces prévisions ne sont clairement pas informatives. L'option que nous avons choisie n'est pas satisfaisante pour obtenir un ensemble précis.

Fiabilité des estimations des incertitudes de prévision

Prendre en compte l'incertitude des débits permet par contre d'obtenir des intervalles de confiance relativement fiables dans la plupart des cas : les médianes des distributions de fréquence d'appartenance aux intervalles de confiance sont proches de leurs valeurs cibles. En revanche, les distributions de ces valeurs sur nos 178 bassins sont larges : les incertitudes sont nettement sous-estimées sur un nombre significatif de bassins versants.

Calculer les poids sur les transformées logarithmiques des débits (pour affaiblir l'hétéroscédasticité des séries) ne modifie que très peu ces résultats.

Études de cas

Nous avons examiné plusieurs chroniques d'estimation des incertitudes sur des bassins pour lesquels les estimations sont fiables globalement et d'autres pour lesquels ces estimations sous-estiment grandement les incertitudes.

Nous observons un comportement général sur toutes les chroniques que nous avons examinées. L'estimation des incertitudes est assez réactive sur les montées de crue : les intervalles de confiance s'élargissent très rapidement sur ces passages de l'hydrogramme unitaire. Par contre, l'estimation est peu réactive sur les décrues : les intervalles de confiance sont très lents à redevenir aussi étroits qu'avant la crue (Fig. 8.4).

Selon l'option choisie pour l'optimisation des paramètres σ_M et σ_Q du filtre particulaire, la largeur des intervalles de confiance varie considérablement. Dans le premier cas (i), tous les pas de temps sont pris en compte pour optimiser la fiabilité des prévisions. Dans le second cas (ii), seuls les pas de temps précédés d'un cumul de pluie important (supérieur au quantile 0.95) sont considérés. Pour tous les bassins examinés, on observe que les intervalles de confiance dans le second cas sont bien plus larges. En particulier sur la figure 8.4 (ii), la proportion de points appartenant à l'intervalle de confiance à 50 % est très largement

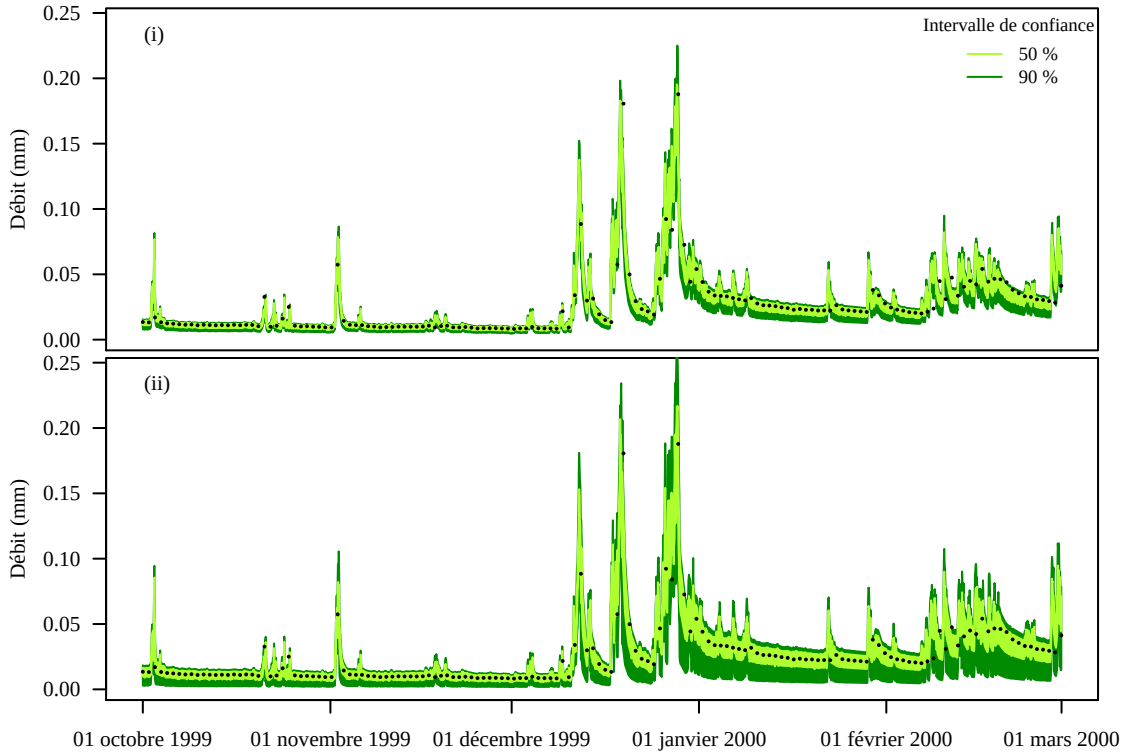


FIGURE 8.4 – Chroniques des estimations des incertitudes de prévision à 6 heures par filtre particulaire, sur le bassin du le Cubry à Pierry (99 km^2, H5224310) lors de l'hiver 1999 – 2000. Les niveaux de verts indiquent les intervalles de confiance à 50 et 90 %. Les points indiquent les débits observés (une fois par jour). (i) Les paramètres σ_M et σ_Q sont optimisés sur l'ensemble des pas de temps. (ii) Ces paramètres ne sont optimisés que sur les pas de temps précédés d'un événement pluvieux important.

supérieure à sa valeur attendue (de l'ordre de 90 %). Par contre, sur les événements de crue, les intervalles de confiance de la figure 8.5 (i) sont mal estimés alors qu'ils sont bien meilleurs en (ii).

8.4.3 Discussion et tests complémentaires

À ce point, quelques tests complémentaires ont été proposés pour affiner l'interprétation de ces résultats.

Sensibilité de la méthode à l'intensité des perturbations sur les pluies
Ici encore, tirer les perturbations sur les pluies passées dans une loi Gamma plus large ne nous a pas permis d'améliorer significativement la fiabilité des prévisions.

Comment expliquer les très fortes valeurs de σ_M et σ_Q ?
Une interprétation proposée de ces valeurs inattendues réside dans la non prise en compte de l'incertitude des pluies futures. L'examen des distributions des paramètres σ_Q et σ_M

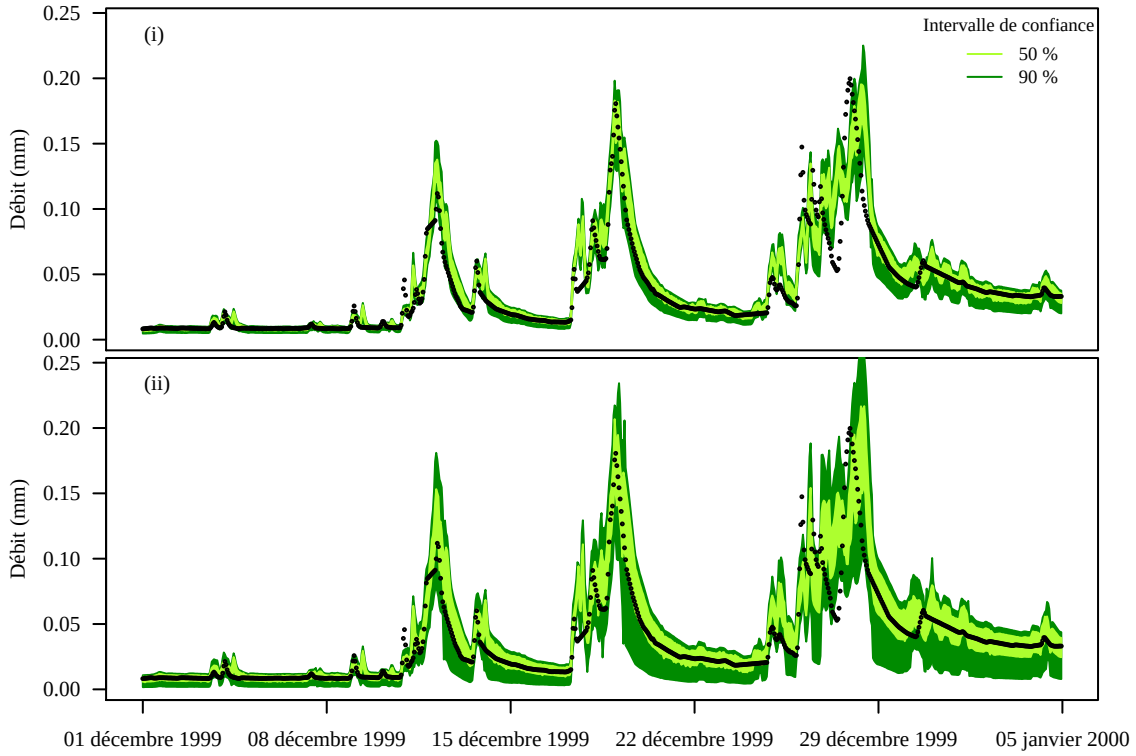


Figure 8.5 – Chroniques des estimations des incertitudes de prévision à 6 heures par filtre particulaire, sur le bassin du le Cubry à Pierry (99 km^2, H5224310) sur trois événements de crue. Les niveaux de verts indiquent les intervalles de confiance à 50 et 90 %. Les points indiquent les débits observés (à toutes les heures). (i) Les paramètres σ_M et σ_Q sont optimisés sur l'ensemble des pas de temps. (ii) Ces paramètres ne sont optimisés que sur les pas de temps précédés d'un événement pluvieux important.

en fonction de l'autocorrélation des débits montre une tendance très nette : le calage des paramètres σ_M et σ_Q sur les bassins les plus réactifs (c'est-à-dire ceux pour lesquels les pluies futures, et donc leurs incertitudes, jouent le plus grand rôle) conduit à des valeurs comparativement supérieures à celles obtenues sur des bassins plus lents (Fig. 8.6). Les pluies futures jouent bien un rôle clef dans l'estimation des incertitudes.

Rôle de l'incertitude des pluies futures

Nous avons donc tenté de perturber les pluies futures : pour chaque particule, les précipitations entre l'instant de prévision t et le pas de temps cible $t + H$ sont perturbées par le facteur $\gamma_{i,t}$ multiplicatif déjà à l'œuvre sur les précipitations observées (entre $t - \Delta t$ et t). Cet essai s'est soldé par une chute de la précision de l'ensemble (-10 points de $C_{2M,P}$). Le CRPS de l'ensemble devient supérieur à la MAE. Les intervalles de confiance ne sont que légèrement élargis : prendre en compte l'incertitude sur les pluies futures de cette manière ne résoud pas le problème.

Une deuxième approche possible pour incorporer les incertitudes sur les pluies futures se

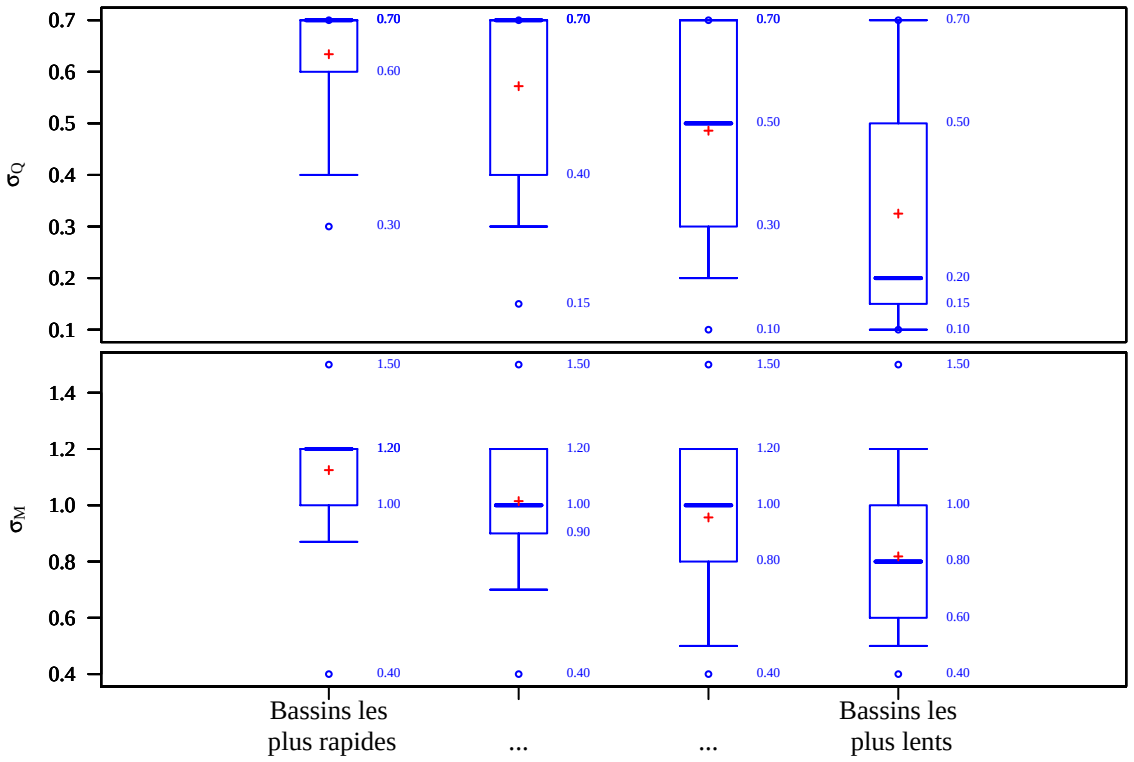


FIGURE 8.6 – Distributions (sur 4 × 89 bassins × 2 périodes) des valeurs calées des paramètres σ_M (en haut) et σ_Q (en bas) sur quatre sous-groupes de bassins définis par classe d'autocorrélation des débits : les 89 bassins estimés les plus réactifs constituent le premier groupe quand les 89 les plus lents forment le quatrième groupe.

base sur les prévisions ensemblistes de précipitation. Nous avons réalisé un essai restreint à partir des prévisions de pluie d'ensemble fournies par le modèle ARPEGE de Météo-France. Ces premiers résultats, calculés sur peu de données (environ 5 mois sur 28 bassins) montrent une forte sous-estimation des incertitudes de prévision quand on ne tient compte que de l'incertitude sur les pluies futures (telles qu'évaluées par ARPEGE, Ann. M, page 535).

Ces résultats peuvent faire écho à une observation d'Evensen (1994) : les approches de filtre d'ensemble (EnKF, filtre particulaire...) sont souvent présentées sur des systèmes météorologiques ou océanographiques. On y observe des comportements chaotiques : deux particules partant d'états initiaux très voisins peuvent donner des prévisions tout à fait différentes. Bien connaître l'incertitude à la date t et appliquer le modèle peut suffire à estimer l'incertitude à la date $t + H$. Je n'ai jamais observé de tels comportements d'un modèle hydrologique où nous exploitons au contraire la stabilité de nos modèles face à des perturbations (par exemple pour les initialiser) : l'incertitude liée aux erreurs de modélisation joue sans doute un rôle majeur. Sans la prise en compte de l'incertitude des précipitations futures (par des scénarios de pluie future différents), il paraît peu probable

qu'un modèle conceptuel simple comme GRP puisse fournir des estimations correctes de l'incertitude de prévision [2].

Rôle de la mise-à-jour déterministe

Dans l'essai d'application du filtre particulaire que nous présentons, le modèle est mis-à-jour bien avant l'instant de prévision et les données de débit servant à cette mise-à-jour sont bruitées. Cela entraîne, comme nous l'avons vu, une chute brutale de la précision des prévisions, qui est accentuée par l'intensité inattendue du bruit sur le débit. Pour échapper à ce défaut majeur, nous avons essayé deux modifications.

La première passe par une voie plus classique où les états du modèle ne sont pas mis-à-jour mais où le paramètre σ_M contrôlant le calcul des poids est beaucoup plus faible, ce qui réduit fortement les poids des particules dont le débit $\widehat{Q}_t^{(i)}$ est éloigné de l'observation Q_t. Cet essai réalisé en imposant $\sigma_M = 0.2$ n'a pas permis d'améliorer la fiabilité de l'ensemble tout en faisant chuter la précision : cette voie n'a pas été explorée plus avant.

La seconde approche consiste à réduire fortement la durée Δt pendant laquelle les particules incorporent des pluies perturbées. En réduisant cette durée à deux pas de temps, nous ne modifions que très peu la fiabilité de l'ensemble tout en améliorant considérablement sa précision : un gain moyen supérieur à 20 points de $C_{2M,P}$. Las, la fraction de bassins sur lesquels la moyenne des valeurs de $C_{2M,P}$ de l'ensemble est significativement inférieure à celle du modèle déterministe est de 74 %.

Nous sommes ici dans ce qui nous paraît être une impasse : la mise-à-jour des états du modèle entraîne un appauvrissement de l'échantillon de particules. Mais bruiter les données servant à cette mise-à-jour entraîne une forte chute de précision en comparaison avec le modèle déterministe.

La délicate écriture de la fonction de vraisemblance

Un dernier facteur pouvant expliquer cet échec provient de l'écriture de la fonction de vraisemblance : une pierre d'achoppement commune à l'ensemble des approches bayésiennes (cf. notamment Liu et Gupta, 2007). Il est nécessaire de fournir un bon modèle d'erreur (comme le soulignent Weerts et El Serafy, 2006) et pourtant ici comme dans les études qui nous ont inspiré, nous avons employé une fonction de vraisemblance gaussienne que nous savons invalidée. Malheureusement l'écriture d'un modèle d'erreur validé reste un exercice difficile (Yang *et al.*, 2007) qui limite toujours nos capacités en prévision des incertitudes par la voie bayésienne.

2. Un résultat similaire aurait probablement été obtenu avec un modèle plus complexe.

8.5 Bilan

L'estimation des incertitudes de prévision est un sujet de recherche central tant sur le plan pratique que pour accroître notre compréhension théorique. Ce chapitre présente une étude exploratoire de l'application du filtre particulaire au modèle GRP. Cette première étude est loin d'être un franc succès et appelle un travail beaucoup plus complet, qui n'a pas été réalisé dans le cadre de cette thèse. L'incertitude prédictive que nous obtenons se compare assez bien avec celle présentée par Salamon et Feyen (2009) dont les intervalles de confiance sont moins larges mais clairement sous-estimés. Les intervalles de confiance calculés par Moradkhani *et al.* (2005a) sont nettement plus resserrés mais ratent une fraction assez élevée de pointes de crue. Cependant, ils sont nettement plus raisonnables que les nôtres sur les décrues : on peut donc penser que leurs valeurs des hyperparamètres sont plus raisonnables. En l'état, nos résultats ne sont pas exploitables opérationnellement : les intervalles de confiance fournis, bienqu'à peu près fiables, ne sont que très peu informatifs.

Cette étude a permis de mettre l'accent sur plusieurs difficultés. La première est le dilemme entre la course à la meilleure précision possible et la recherche d'une estimation fiable des prévisions. Nous mettons à jour les états du modèle pour améliorer la précision mais cette démarche contraint le modèle au point de restreindre l'estimation des incertitudes. Une autre difficulté rencontrée réside dans la différence de calage des paramètres du filtre particulaire selon que nous cherchons à être fiables sur tous les pas de temps ou principalement sur les événements pluvieux.

Un point clef que nous n'avons pas pu traiter en profondeur à cause du manque de données, est la prise en compte des incertitudes sur les prévisions de pluie future. D'après nos premiers résultats (et en accord avec de nombreux auteurs), elle est primordiale. Cependant, les premiers essais avec les prévisions d'ensemble d'ARPEGE conduisent également à une sous-estimation des incertitudes : il semble donc nécessaire de traiter à la fois les incertitudes sur l'état du modèle à l'instant de prévision et celle des pluies futures (comme préconisé par Krzysztofowicz, 1999). Dans ce cadre, le filtre particulaire serait employé (et calé) pour estimer les incertitudes à l'instant de prévision [3] ; l'incertitude sur les prévisions due à l'incertitude sur les pluies futures serait prise en compte dans un second temps. Cela pourrait permettre de caler les paramètres du filtre particulaire à des valeurs plus raisonnables, conduisant à des incertitudes sur les états à l'instant de prévision plus petites et par là échapper à Charybde et à Scylla : obtenir *in fine* des prévisions à la fois plus précises et aussi fiables.

3. Dans notre essai, nous avons cherché à évaluer directement les incertitudes à l'instant prévu $(t+H)$. Cette situation peut être un facteur d'explication des valeurs élevées de σ_M et de σ_Q : le principe du filtre particulaire conduisant plutôt à l'évaluation des incertitudes à l'instant d'application du théorème de Bayes.

9
Chapitre

Mise-à-jour des paramètres

Objectifs

- ▷ Justifier la mise-à-jour des paramètres.
- ▷ Établir un état de l'art des méthodes les plus couramment employées.
- ▷ Étudier la méthode de Yang et Michel, qui se caractérise par sa recherche de robustesse.
- ▷ Proposer une alternative et en évaluer les performances.

Sommaire

9.1 Mettre à jour un invariant temporel ? 157
9.1.1 Les paramètres du modèle peuvent être mal définis et mal connus 157
9.1.2 Rapide état de l'art 158
9.2 Notre point de départ : la méthode de Yang et Michel 160
9.2.1 Principe 160
9.2.2 Performances 162
9.2.3 Limites 162
9.3 Proposition d'une méthode alternative 163
9.3.1 Principe 163
9.3.2 Tests réalisés et résultats 164
9.3.3 Interprétation 170
9.4 Bilan 172

Parmi les grandes approches de mise-à-jour d'un modèle hydrologique, Refsgaard (1997) mentionne la mise-à-jour des paramètres du modèle. Dans ce chapitre, nous examinerons tout d'abord ce qui peut justifier de mettre à jour des paramètres. Une revue rapide des méthodes proposées dans la littérature nous amènera ensuite à étudier plus en détails la méthode de Yang et Michel (2000) puis à proposer une méthode alternative de mise-à-jour des paramètres dont les performances seront examinées sur notre échantillon de bassins versants.

9.1 Mettre à jour un invariant temporel ?

9.1.1 Les paramètres du modèle peuvent être mal définis et mal connus

Les paramètres peuvent recevoir plusieurs définitions (Troutman, 1985). Ils sont souvent considérés comme étant des caractéristiques du fonctionnement du bassin versant ne variant pas dans le temps, par opposition aux états qui varient en fonction des flux de masse, d'énergie entrant ou sortant du système constitué par le bassin versant. Ils peuvent également être définis comme des degrés de liberté laissés au modèle sans plus d'interprétation (pour Mantovan et Todini, 2006, ce sont des *rather temporary "dummy" or "convenient" quantities of uncertain nature*) ; cela est notamment le cas pour les modèles *boîtes noires* et de façon plus générale pour tous les modèles calés numériquement. Mettre à jour les paramètres peut paraître incohérent avec leur nature d'invariants selon la première définition (mais pas pour la deuxième). Cependant, même dans le cadre de cette première définition, plusieurs limites pratiques autorisent cette stratégie.

Tout d'abord, les modèles hydrologiques sont loin d'être une représentation parfaite de la relation Pluie – Débit. Les modèles s'appuyant sur une représentation de la physique du point ou du (petit) élément représentatif peuvent éventuellement définir des paramètres mesurables de façon précise mais en si grand nombre que les mesures (de terrain ou par un appareillage éloigné, de type satellitaire) ne peuvent pas fournir avec toute la précision (numérique, spatiale ou temporelle) nécessaire. Inversement, les modèles conceptuels plus simples emploient des paramètres dont l'interprétation est imprécise (en termes physiques) et indirecte (Wagener et McIntyre, 2005) ; ces paramètres sont difficilement liés à une mesure (de façon directe ou indirecte). En conséquence, les paramètres sont soit des objets mal connus, soit mal définis. Dans la pratique, plusieurs auteurs (par exemple Gupta *et al.*, 1998, Madsen, 2000, Wagener *et al.*, 2001) estiment qu'un jeu de paramètres unique ne permet pas aux modèles hydrologiques actuels de reproduire convenablement les différents régimes de débits et mettent ainsi en doute l'existence d'un jeu de paramètres *vrais* (qu'il soit ou non accessible par calage). Sorooshian et Gupta (1983) rappellent que les paramètres de nombreux modèles varient selon la saison. Certains auteurs (comme Wagener *et al.*, 2003) proposent même des approches d'estimation dynamique des paramètres.

Dans la majorité des cas, les paramètres sont estimés, soit par régionalisation (par régressions comme chez Sefton et Howarth, 1998, Merz et Blöschl, 2004 ou par des techniques de partitionnement (*clustering*) comme chez Burn et Boorman, 1993) soit par calage. Quelle que soit la manière dont ils sont obtenus (mesure, régionalisation, calage), ces paramètres dépendent de façon plus ou moins directe de mesures caractérisées par un certain niveau d'incertitude qui génère une incertitude sur la valeur de ces paramètres. En outre, la régionalisation des paramètres reste une approche très incertaine malgré les efforts de recherche engagés dans le cadre de la décennie *Predictions in Ungauged Basins* (PUB, Sivapalan *et al.*, 2003b). Le calage est, quant à lui, un procédé délicat à mener, appelant à de nombreux choix subjectifs (quelle fonction objectif ? Quel algorithme ? Quelle période de données (disponible) ?) et est sujet à plusieurs difficultés (surface ponctuée de nombreux optima locaux équivalents (Duan *et al.*, 1992), ...) dont vont dépendre plus ou moins largement les valeurs calées des paramètres.

Ainsi, la mise-à-jour des paramètres qui manque *a priori* de cohérence vis-à-vis de la première définition des paramètres (comme invariants du système) est tout-à-fait justifiée par la méconnaissance que nous en avons.

9.1.2 Rapide état de l'art

Comme pour les mises-à-jour portant sur les états, on peut distinguer deux approches conceptuellement différentes pour la mise-à-jour des paramètres. La première est déterministe et vise à identifier le jeu de paramètres qui permettra la meilleure prévision possible à l'instant t. La seconde, probabiliste, cherche à affiner la distribution des paramètres grâce à l'information contenue dans les données nouvellement disponibles.

Approches variationnelles

De nombreuses méthodes de mise-à-jour des paramètres sont variationnelles(g). Tucci et Clarke (1980) s'intéressent aux difficultés du *recalage* d'un modèle hydrologique quand suffisamment de nouvelles données sont disponible et proposent une méthode basée sur une durée de rétroaction(g)égale à la durée de mémoire du bassin M (et le recalage est effectué sur tous les points de $[\![t-M,t]\!]$). Plus récemment, Yang et Michel (2000) ont présenté une méthode alternative encadrant plus rigoureusement le recalage que ne le fait une méthode variationnelle standard. Seo *et al.* (2003) proposent une mise-à-jour variationnelle de certains états et de deux paramètres du modèle Sacramento employé dans les centres de prévision des crues du *National Weather Service* américain.

Approches séquentielles

La mise-à-jour des paramètres peut également être de nature séquentielle(g). Par exemple, un filtre de Kálmán peut être appliqué aux paramètres, conjointement ou parallèlement

aux états (Liu et Gupta, 2007). Ces méthodes duales [1] ont la tâche délicate de déterminer ce qui relève de la mise-à-jour des paramètres et ce qui relève de la mise-à-jour des états ; elles peuvent se révéler instables.

La méthode GLUE (pour *Generalized Likelihood Uncertainty Estimation*, Beven et Binley, 1992, Beven et Freer, 2001) est une des approches séquentielles probabilistes les plus connues. Ses auteurs introduisent la notion d'*équifinalité*(g) : différentes combinaisons structure de modèle – jeu de paramètres peuvent être considérées comme atteignant un même degré de vraisemblance. La méthode emploie de façon itérative le théorème de Bayes pour mettre à jour la distribution des paramètres à l'aide d'une *vraisemblance informelle*. Cette dernière particularité fait l'objet d'une controverse (Thiemann *et al.*, 2001, Beven, 2006, Mantovan et Todini, 2006, Beven *et al.*, 2008). Une autre approche bayésienne a été proposée plus récemment par Thiemann *et al.* (2001) ; il s'agit de l'estimation bayésienne récursive (BaRE). Elle s'appuie sur un cadre bayésien formel (Chap. 8 et Ann. L) et est obligée de faire des hypothèses très fortes sur la structure de l'erreur. Si les hypothèses sur la structure de l'erreur sont correctes, son approche itérative permet de réduire progressivement les incertitudes sur les paramètres.

Approches liant le jeu de paramètres à un autre facteur

Certains auteurs proposent de relier le jeu de paramètres à un facteur extérieur à la modélisation hydrologique. Ainsi Oudin *et al.* (2006a) emploient deux paramétrisations d'un même modèle correspondant aux débits faibles et élevés ; à chaque pas de temps, une paramétrisation est sélectionnée en fonction d'un indice climatique ou du niveau d'humidité du sol modélisé (via un module SMA [2]). Hay *et al.* (2009) proposent également une approche à plusieurs jeux de paramètres (calés) ; le jeu de paramètres à employer à une date t est déterminé par l'état de l'atmosphère (via la carte des altitudes de pression 700 hPa).

Difficultés liées à la mise-à-jour des paramètres

La mise-à-jour des paramètres hérite dans une certaine mesure des difficultés liées au calage du modèle. De nombreux auteurs (comme Mein et Brown, 1978, Jakeman et Hornberger, 1993, Perrin *et al.*, 2001) rappellent que le calage ne peut ordinairement déterminer qu'un nombre très limité de paramètres. Comment, dans ce contexte, ajouter un petit nombre d'observations, même très proches, peut-il permettre la détermination robuste d'un jeu de paramètres assurant une meilleure prévision ? De plus, la surface de réponse des fonctions objectifs peut également être « accidentée » et présenter de nombreux optima locaux (voir par exemple à ce sujet l'approche ROPE développée par Bárdossy et Singh, 2008, qui vise à se placer dans la zone optimale de l'espace des paramètres plutôt que rechercher l'optimum). Ces difficultés incitent à examiner le degré de liberté à accorder à la procédure

1. qui mettent à jour et les états et les paramètres.
2. Module de suivi de l'humidité du sol : *Soil Moisture Accounting*.

de mise-à-jour pour faire varier le jeu de paramètres.

Les incertitudes sur les prévisions proviennent de diverses sources : entrées, structure du modèle, paramétrage... La mise-à-jour des paramètres ne peut pas corriger les incertitudes provenant de toutes ces sources. Un équilibre doit notamment être atteint entre la mise-à-jour des états et mise-à-jour des paramètres. Les états sont (au moins pour certains) censés varier rapidement tandis que les paramètres sont beaucoup moins variables. Cela pose la question de la fréquence de la mise-à-jour (pour une méthode séquentielle) ou de la profondeur (en nombre de pas de temps) du calcul de la fonction de coût (pour une méthode variationnelle).

9.2 Notre point de départ : la méthode de Yang et Michel

Comme nous avons très peu de connaissance *a priori* concernant les valeurs des paramètres et leurs incertitudes, une approche variationnelle semble plus indiquée qu'une méthode séquentielle. En outre, elle peut s'interpréter assez facilement en termes de *recalage* du modèle. Yang et Michel (2000) ont mis l'accent sur plusieurs difficultés liées à la méthode variationnelle et ont proposé une solution cherchant à les dépasser. Nous partons de leur mise-à-jour des paramètres.

9.2.1 Principe

Calage des paramètres

Les paramètres sont calés au préalable en *simulation* sur une chronique de façon tout à fait ordinaire. Le jeu de paramètres de base Θ_0 ainsi obtenu est censé représenter le comportement moyen à long terme du bassin (tel que vu par le modèle), en supposant que la phase de calage se passe convenablement, que le modèle ait des performances acceptables, que la chronique de calage soit de longueur suffisante, etc.

Laisser du temps au modèle pour adapter ses états

Un petit changement du jeu de paramètres à un instant t ne modifie pas nécessairement les sorties du modèle de façon significative dès cet instant. En effet, les débits calculés sur les premiers pas de temps suivant l'instant t sont largement conditionnées par les états *en cours* du modèle, eux-mêmes fonctions du jeu de paramètres et des états précédents. La mise-à-jour doit donc tenir compte de ce délai à travers une *période de rétroaction* d'une durée d_{RA}.

À chaque instant de prévision (noté t sur la figure 9.1), la mise-à-jour part de l'état en *simulation* du modèle au pas de temps $t - d_{RA}$. La procédure de mise-à-jour recherche un jeu de paramètres Θ_t autour du jeu de paramètres Θ_0 (algorithme de recherche locale) qui minimise la dernière erreur quadratique ε_t lorsqu'il est employé par le modèle pour simuler

les états (et le débit) aux pas de temps $t - d_{RA}$ à t. Le modèle effectue alors la prévision à l'aide de ce jeu de paramètres Θ_t et de l'état du système obtenu au pas de temps t en utilisant le jeu de paramètres de base Θ_0 pour les pas de temps précédant $t - d_{RA}$ et le jeu de paramètres mis-à-jour Θ_t pour les pas de temps au-delà.

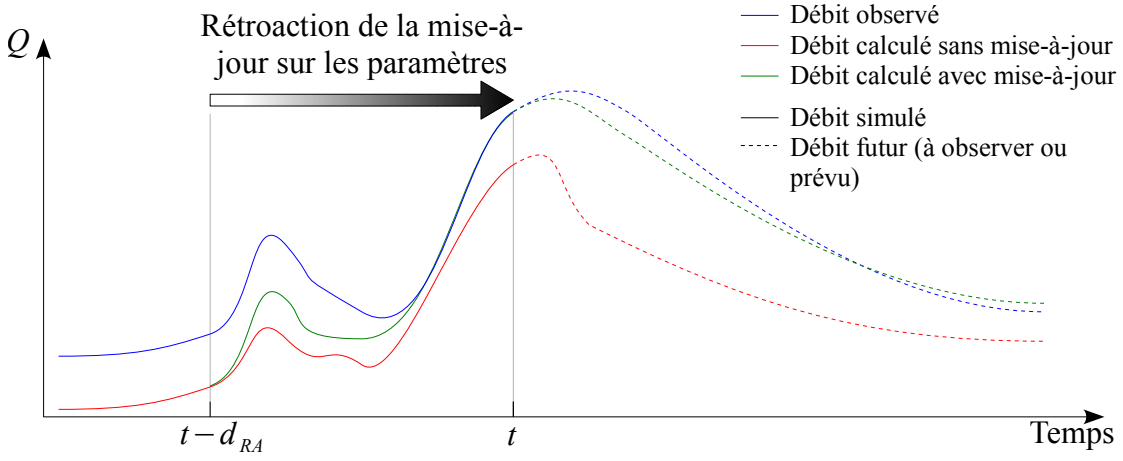


Figure 9.1 – Principe de la mise-à-jour proposée par Yang et Michel (2000). Les paramètres sont modifiés d_{RA} pas de temps avant l'instant de prévision t de façon à minimiser l'erreur quadratique sur les tout derniers pas de temps avant l'instant de prévision.

Toujours partir des états simulés par le modèle sans mis-à-jour

Dans la procédure décrite ci-dessus, la mise-à-jour effectuée pour réaliser la prévision au pas de temps $t - 1$ n'est pas conservée au pas de temps t : la procédure de mise-à-jour part pour chaque prévision de l'état du modèle fonctionnant avec le jeu de paramètre Θ_0 représentant le comportement moyen du bassin.

Limiter l'écart entre les jeux de paramètres de base et mis-à-jour

Cette méthode de mise-à-jour est pensée pour corriger les erreurs sur la structure du modèle et sur notre méconnaissance d'un jeu de paramètres *vrais* (dans la mesure où il existe). Elle ne doit pas être appliquée pour tordre le modèle et le forcer à corriger n'importe quelle erreur. Une erreur sur les entrées ou sur la mesure du débit notamment pourrait entraîner l'adoption d'un jeu de paramètres non pertinent et mettre en péril la robustesse des prévisions. Aussi, l'écart entre le jeu de paramètres de base Θ_0 et le jeu de paramètres mis-à-jour Θ_t est limité : une étude de sensibilité du débit aux paramètres permet de définir l'écart maximal sur chaque paramètre comme étant un multiple de l'écart-type de ce paramètre (vu comme une variable aléatoire).

9.2.2 Performances

Yang et Michel (2000) comparent cette méthode à la correction de l'erreur à l'aide d'un modèle autorégressif additif d'ordre 1 pour le modèle GR4J sur le bassin versant de la Seine à Paris et trois sous-bassins. La mise-à-jour des paramètres surclasse la correction de l'erreur pour les quatre bassins versants étudiés, quelque soit le scénario de pluies futures employé [3].

La méthode de Yang a également été appliquée au pas de temps journalier par Tangara (2005, page 70) au modèle GR4J et à une version simplifiée de TOPMODEL (Beven et Kirkby, 1979, Michel *et al.*, 2003) et testée sur 188 bassins versants à un horizon d'un jour (un pas de temps). Les performances des deux modèles mis à jour par cette méthode variationnelle se sont révélées décevantes, comparées à celles obtenues par les mêmes modèles corrigés par report de l'erreur. Dans le cas de la version simplifiée de TOPMODEL, la moyenne du critère de persistance était significativement négative. Cependant, la méthode avait un effet positif sur les prévisions : les performances (même inférieures à celles du modèle de persistance) étaient nettement supérieures à celles obtenues sans aucune mise-à-jour du modèle.
Dans la comparaison menée par Tangara (2005), la mise-à-jour variationnelle des paramètres était appliquée en prenant les nombreuses modalités d'application (durée de rétro-action, écart maximal toléré du jeu de paramètres mis-à-jour à Θ_0, paramètres mis-à-jour, etc) identiques pour tous les bassins, ce qui peut expliquer en partie la faiblesse des résultats. Cependant, nous avons vérifié (au pas de temps horaire) que la grande faiblesse de ces résultats (persistance moyenne négative pour un modèle et performances inférieures au simple report d'erreur) s'explique également par le fait que la mise-à-jour des paramètres était appliquée à l'exclusion de toute autre mise-à-jour, notamment de celle des états.

9.2.3 Limites

Cette méthode est sujette à certaines des difficultés mentionnées dans la première partie de ce chapitre :

Robustesse ou efficacité ?
Comme le montrent Yang et Michel (2000), si le jeu de paramètres mis-à-jour peut s'éloigner significativement du jeu de paramètres de base, la robustesse de la mise-à-jour est compromise. Il suffit d'une erreur forte dans les chroniques des entrées ou du débit pour que le modèle se torde pour minimiser une erreur absurde, mettant en danger la pertinence des prévisions futures. Inversement, brider la mise-à-jour en s'assurant que le jeu de paramètres mis-à-jour ne s'éloigne que peu du jeu de paramètres de base limite fortement

3. La correction de l'erreur obtient des performances légèrement supérieures sur un bassin pour un scénario de pluies futures nulles, mais la différence est trop faible pour être significative.

ses effets : certaines erreurs qu'il serait nécessaire de prendre en compte seront faiblement traitées et les prévisions peu améliorées. De plus, la robustesse est mise à rude épreuve par le nombre de degrés de liberté laissés à la méthode pour s'adapter aux différents bassins et modèles.

Faire corriger toutes les erreurs par une seule méthode ?
La difficulté mentionnée au paragraphe précédent provient du fait que la méthode proposée cherche à corriger des erreurs liées à différentes sources d'incertitudes. Les paramètres sont supposés être invariants dans le temps (ou ne varier que lentement) au contraire des états qui varient (plus) rapidement. Il semble donc plus pertinent de faire porter la mise-à-jour des paramètres sur des erreurs variant lentement et la mise-à-jour des états sur des erreurs moins autocorrélées (au bémol près que certains paramètres contrôlent les réponses rapides). Liu et Gupta (2007) font remarquer que la mise-à-jour des seuls paramètres (ou des seuls états) peut conduire à des prévisions biaisées. L'absence de combinaison avec la mise-à-jour des états dans la procédure proposée ici ne paraît pas être la solution la plus efficace.

Signification du jeu de paramètres mis-à-jour
Le jeu de paramètres mis-à-jour n'est finalement issu que de deux éléments : la toute dernière erreur observée et le jeu de paramètres de base Θ_0 (qui sert de point de départ au recalage). L'erreur quadratique du modèle calculée sur un seul pas de temps, fût-il le dernier, est très peu informative sur la qualité du modèle et des prochaines prévisions. D'ailleurs, les méthodes visant de caler le modèle sur un petit nombre de points restent encore approximatives (e.g., Perrin *et al.*, 2007, Seibert et Beven, 2009). Si le jeu de paramètres Θ_t est assez éloigné du jeu de paramètres de base, on peut s'interroger sur la signification à lui donner, ainsi que sur la confiance qu'on peut lui accorder.

Temps de calcul
La méthode proposée passe par un recalage (relativement) court effectué à chaque instant de prévision. Cette opération qui ne doit pas prendre plus de quelques minutes, n'est pas rédhibitoire en temps réel. Elle devient très difficile à gérer quand il s'agit de la tester sur plusieurs dizaines de bassins versants et sur plusieurs années (au pas de temps horaire).

9.3 Proposition d'une méthode alternative

9.3.1 Principe

Nous conservons les principes de la méthode de Yang et modifions la méthode sur deux points. La mise-à-jour des paramètres sera combinée avec une mise-à-jour des états. Les jeux

de paramètres mis-à-jour ne pourront prendre qu'un nombre fini de valeurs, déterminées par un calage préalable dans l'esprit du calage discret proposé par Perrin *et al.* (2008).

Coupler mise-à-jour des états et mise-à-jour des paramètres

Nous avons vu que mettre à jour uniquement les paramètres du modèle n'était pas judicieux. Cela ne correspond ni à la multiplicité des sources d'incertitudes ni aux dynamiques des erreurs. Le niveau du réservoir de routage sera donc mis-à-jour par la méthode validée au chapitre 7. Cela correspond à une mise-à-jour forte et contraignante des états. En comparaison, la mise-à-jour des paramètres sera moins intense et n'aura pas pour but d'éliminer complètement la dernière erreur de prévision observée.

Une bibliothèque de jeux de paramètres calés au préalable

Il est difficile d'interpréter un jeu de paramètres « recalés » sur un (très) petit nombre de pas de temps et de lui accorder une grande confiance. Plusieurs études (Yapo *et al.*, 1996, Perrin *et al.*, 2007, avec le modèle GR4J, Merz *et al.*, 2009) montrent que quelques années sont suffisantes pour caler un modèle parcimonieux [4]. Nous disposons d'au moins cinq années pour caler les modèles sur tous les bassins. Il nous est donc possible de partitionner cette période de calage en n sous-ensembles (non nécessairement continus dans le temps). Le calage sera réalisé $n+1$ fois : au i^{e} calage ($1 \leq i \leq n$), seules les prévisions faites aux pas de temps appartenant au i^{e} sous-ensemble seront prises en compte lors du calcul de la fonction objectif et permettront d'obtenir le i^{e} jeu de paramètres. Le dernier calage sera réalisé sur l'ensemble de la période de base : il s'agit du jeu de paramètres Θ_0.

Pour réaliser une prévision $\widehat{Q}_{t+H|t}$ au pas de temps t, seuls les $n+1$ jeux de paramètres calés précédemment seront testés sur la période de rétroaction (Fig. 9.2). Le jeu de paramètres sélectionné pour la prévision sera celui minimisant une fonction de coût sur les derniers pas de temps de cette période de rétroaction (qui forment la *fenêtre d'assimilation*). Ainsi, tous les jeux de paramètres employés lors des mises-à-jour seront issus d'un calage sur un nombre important de pas de temps ; nous espérons qu'ils auront plus de contenu hydrologique (au sens des données de pluie et de débit) qu'un jeu de paramètres obtenu par un recalage sur un tout petit nombre de pas de temps (comme cela peut être le cas pour une méthode variationnelle classique).

9.3.2 Tests réalisés et résultats

De façon générale, nous observons un gain de performance relativement élevé par rapport au modèle dont seul le niveau du réservoir de routage est mis à jour, pour les horizons de

4. Cela reste fortement dépendant du modèle : Yapo *et al.* (1996) indiquent une durée de 8 ans pour le modèle NWSRFC-SMA et Merz et Blöschl (2004) montrent des différences significatives entre deux jeux de paramètres du modèle HBV calé sur des périodes de 11 ans différentes. Le facteur crucial n'est pas tant la longueur la chronique disponible pour le calage que la présence d'années humides et d'événements pluvieux (Gan *et al.*, 1997, Anctil *et al.*, 2004b).

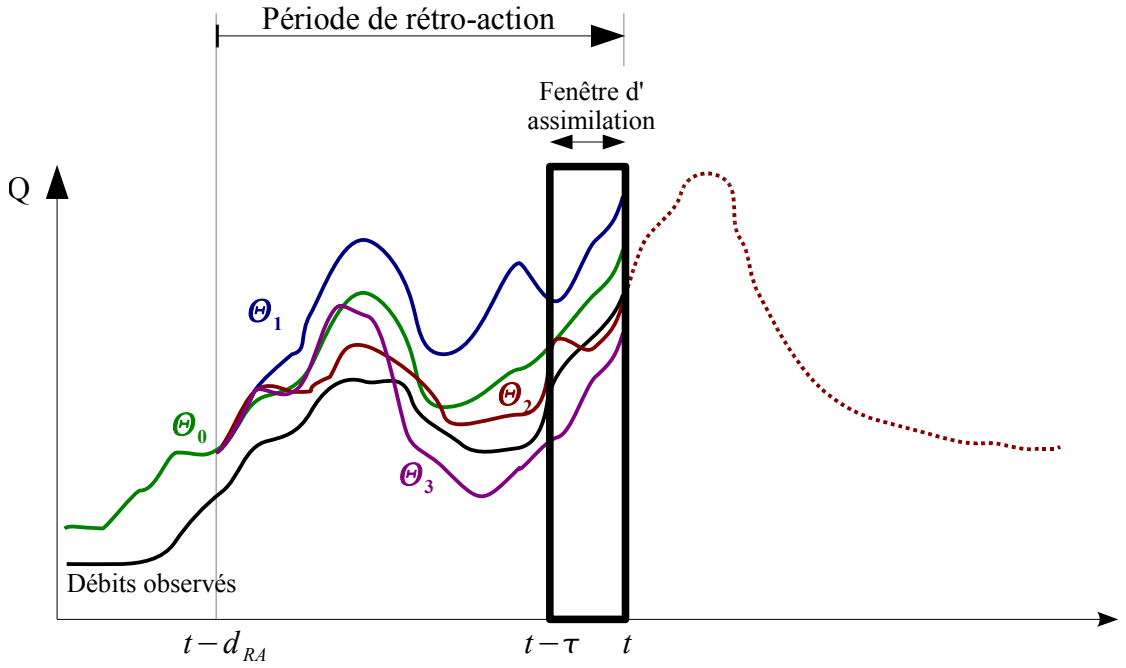


Figure 9.2 – Alternative à la mise-à-jour des paramètres proposés par Yang. L'état du modèle au début de la période de rétroaction est celui obtenu avec le jeu de paramètres de base Θ_0. Les jeux de paramètres Θ_0, Θ_1, ..., Θ_n (ici $n = 3$) sont employés sur la période de rétroaction pour réaliser des prévisions $\widehat{Q}_{t-i|t-i-1}$. Le jeu de paramètres Θ_2 permet d'obtenir les meilleures prévisions sur les tout derniers pas de temps (la fenêtre d'assimilation $[\![t-\tau,\ldots,t]\!]$) : il est donc sélectionné pour réaliser la prévision $\widehat{Q}_{t+H|t}$.

prévision les plus courts : le gain médian (et moyen) est de l'ordre de six points de $C_{2M,P}$ à une heure et s'amenuise pour devenir non significatif pour des horizons supérieurs à 12 heures (Fig. 9.3). La meilleure version testée se base sur une bibliothèque de $10+1$ jeux de paramètres calés sur des sous-ensembles aléatoires (et imbriqués). La durée de rétroaction est de 72 heures. Tous les paramètres y sont mis-à-jour et la fenêtre d'assimilation est réduite au seul dernier pas de temps.

La méthode proposée présente plusieurs degrés de liberté que nous allons explorer pour évaluer ses performances. Le tableau 9.1, situé à la fin de cette section (page 171) rassemble et résume les résultats détaillés dans les prochains paragraphes. Pour pouvoir comparer ces essais à la version horaire du modèle GR3P (Tangara, 2005), les modèles testés bénéficient de la correction des erreurs présente dans le modèle initial (voir le chapitre suivant).

Comment définir les sous-ensembles du calage ?

Il existe un très grand nombre de façons de scinder la période de calage en n sous-ensembles (non nécessairement continus) permettant d'obtenir différents jeux de paramètres. Nous en avons étudié trois (Fig. 9.4).

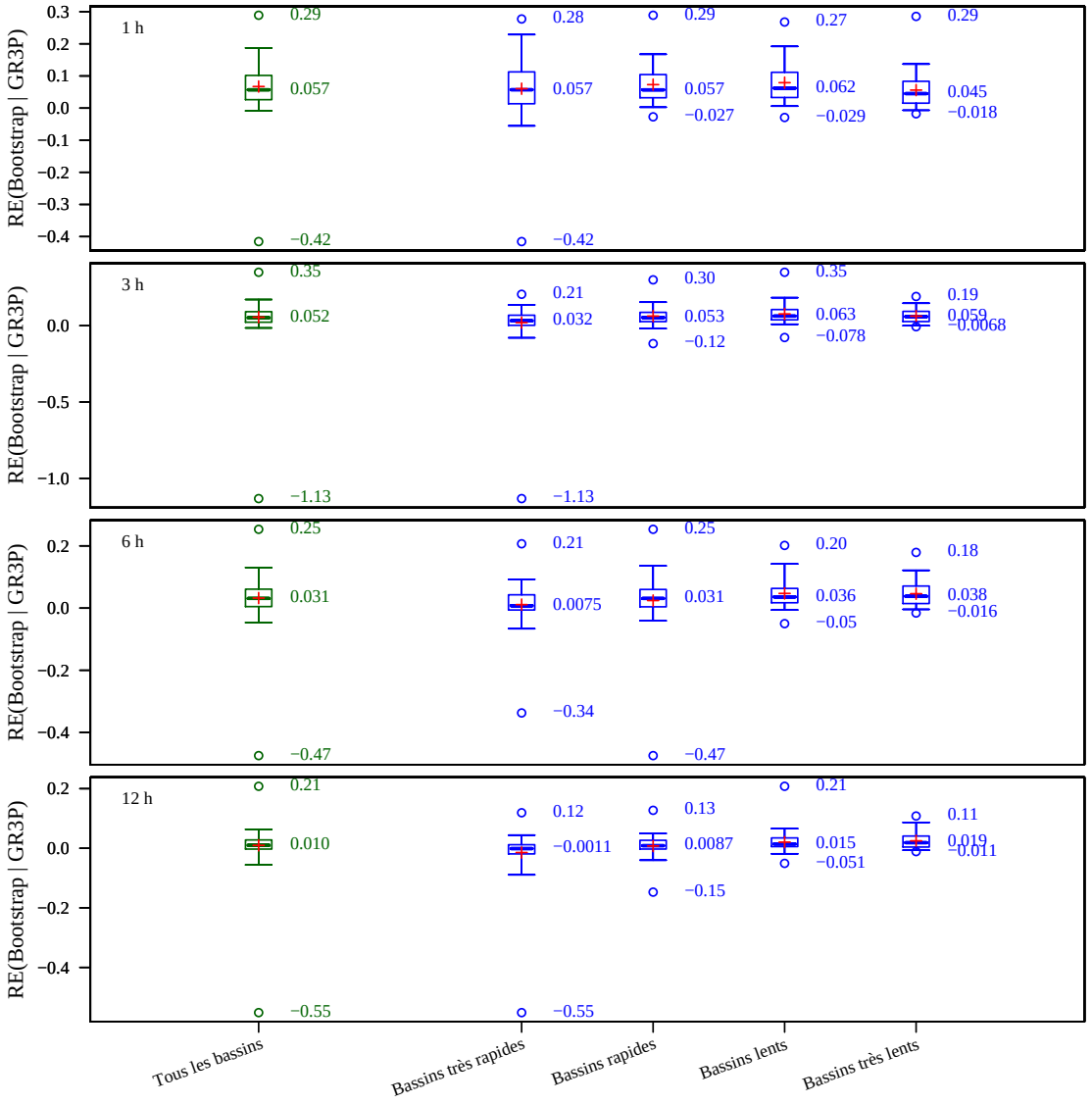


FIGURE 9.3 – Distribution des gains de performance dûs à la mise-à-jour des paramètres du modèle GRP pour quatre horizons de prévision (1, 3, 6 et 12 heures). Le critère est la diminution relative de la RMSE du modèle obtenue par l'emploi du bootstrap. La meilleure version testée de la mise-à-jour des paramètres a été employée ici. La boîte à moustache de gauche représente la distribution des gains sur l'échantillon de 178 bassins versants. L'échantillon a ensuite été scindé en quatre sous-groupes correspondant à quatre classes d'auto-corrélation du débit. Les gains en performance par classe sont donnés sur les quatre boîtes à moustache de droite.

– La première partition de la période de calage est la plus simple : la période est découpée en n périodes continues de même durée.
– Nous avons vu que de nombreux auteurs (e.g. Gupta *et al.*, 1998) font remarquer que les modèles hydrologiques actuels ne sont pas capables de bien simuler ou prévoir le débit dans ses différents régimes avec un unique jeu de paramètres. Nous testons donc une partition de la période de calage par classes de débit bornées par des quantiles régulièrement espacés (approche comparable à ce que propose Khan, 1989).
– Enfin, une partition aléatoire est employée : chaque pas de temps est affecté aléatoirement à un sous-ensemble avec une probabilité uniforme pour tous les sous-ensembles. Chaque sous-ensemble contient ainsi un ou plusieurs pas de temps de tous les événements de crue.

Aux horizons de prévision les plus courts (inférieurs à 6 heures), scinder la période de calage en sous-périodes continues ou réaliser une partition aléatoire de cette période permettent d'obtenir les meilleures performances (avec un avantage non significatif pour la première méthode). Aux horizons plus lointains, on observe un décrochement des performances pour la première méthode. Réaliser des classes de débits conduit dans tous les cas à de moins bonnes performances. Ce dernier résultat rappelle les observations faites en simulation : caler le modèle sur une seule classe de débit conduit souvent à un jeu de paramètres moins robuste.

Durée de la période de rétroaction

Une période de rétroaction est nécessaire pour que les états du modèle au pas de temps t et la prévision $\widehat{Q}_{t|t-1}$ aient pu s'adapter et profiter de la mise-à-jour des paramètres. On pourrait considérer que plus cette période est longue, mieux cela vaut. Cependant, nous chercherons à en limiter sa durée car cette procédure alourdit la tâche de prévision et nécessite de garder en mémoire les entrées du modèle sur une longue plage de données.
L'examen de la sensibilité des performances du modèle à la durée d_{RA} de la période de rétroaction (de 12 à 720 h) montre que les performances du modèle ainsi mis-à-jour y sont peu sensibles. Il n'y a pas d'intérêt marqué (même pour les bassins les plus lents) à recaler le modèle par cette méthode sur une longue période de rétroaction.

Taille de la bibliothèque de jeux de paramètres

Les services de prévision ne disposent (dans le meilleur des cas) que d'une chronique limitée permettant de caler le modèle. Il est donc nécessaire d'atteindre un équilibre entre le nombre de jeux de paramètres disponibles pour les prévisions et le « contenu » hydrologique de chaque jeu de paramètres (qui sera d'autant plus faible que le nombre de jeux de paramètres sera grand puisque chaque jeu sera calé sur un plus faible nombre de points). Nous disposons de périodes de calage d'environ 5 ans sur nos bassins et testons des bibliothèques comptant de 5 à 50 jeux de paramètres.

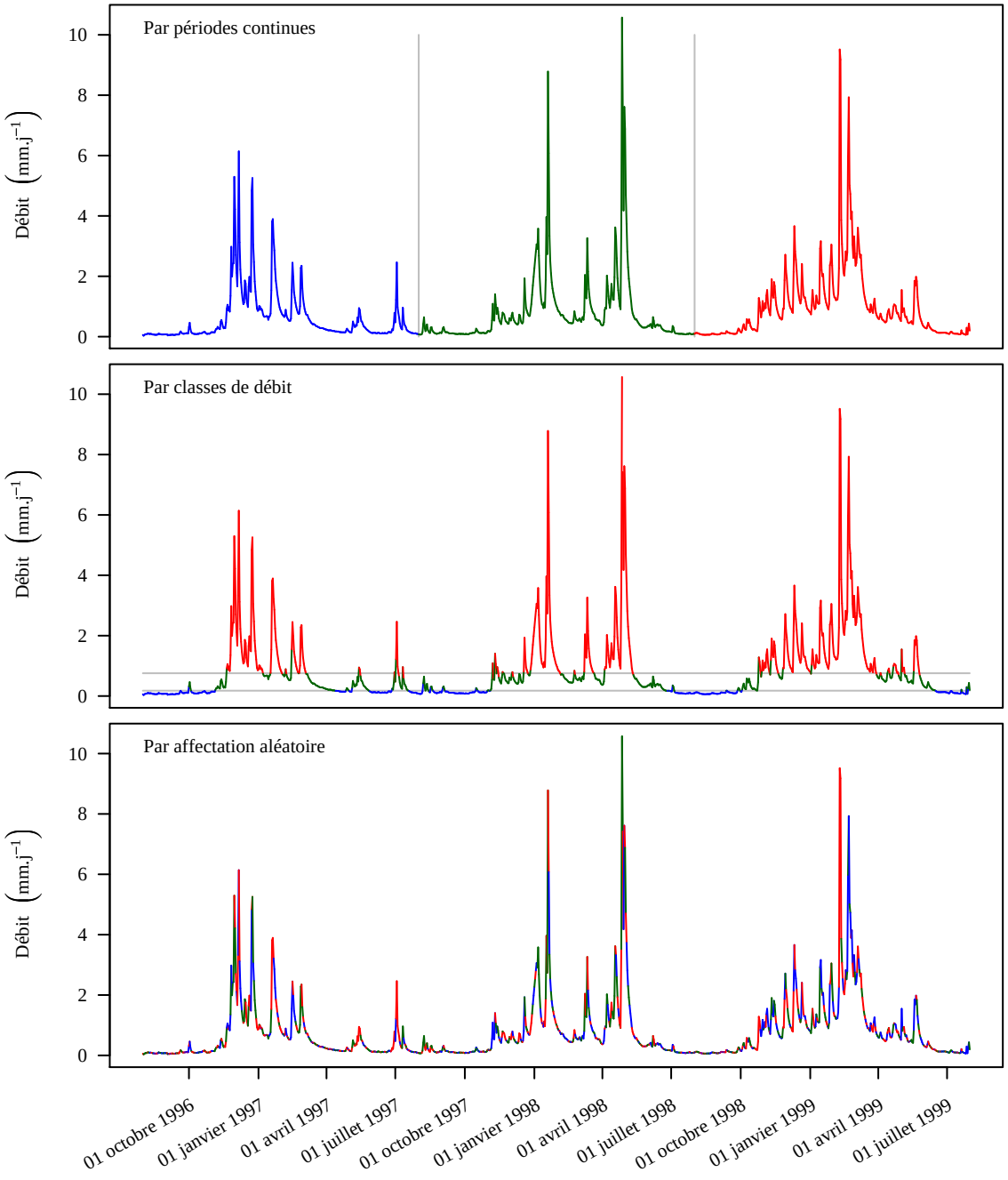


FIGURE 9.4 – Trois définitions des sous-ensembles de calage ont été testés (sur le bassin versant de la Petite Creuse à Fresselines).

Au delà d'un nombre minimal de jeux de paramètres, il n'y pas grand intérêt à augmenter ce nombre. Aux horizons de prévision courts, la bibliothèque comptant 20 jeux de paramètres obtient de meilleures performances que les autres sans que le coût en calcul du passage de 10 à 20 jeux de paramètres se justifie pleinement. À des horizons plus lointains, la différence est encore plus faible, voire en faveur de 10 jeux de paramètres.
Employer un très grand nombre de jeux de paramètres (50) peut détériorer nettement les performances du modèle : - 10 points en médiane à un horizon d'une heure. Cet affaiblissement, attendu, est constaté aux horizons de prévision les plus courts et est d'autant plus intense que les bassins sont rapides. On notera cependant que pour les bassins les plus lents (ceux dont le débit est le plus autocorrélé), employer 50 jeux de paramètres améliore globalement les performances.

Choix des paramètres mis-à-jour

Il est possible qu'il ne soit pas pertinent de mettre à jour tous les paramètres. Nous essayons donc de mettre à jour le paramètre de la fonction de production, ou les paramètres de la fonction de routage uniquement.
Restreindre le nombre de paramètres mis-à-jour diminue légèrement les performances dans les deux cas testés et pour tous les horizons de prévision.

Largeur de la fenêtre d'assimilation

Nous avons également testé le nombre d'erreurs $\varepsilon_{t-i|t-i-1}$ prises en compte pour choisir le jeu de paramètres employé pour la prévision $\widehat{Q}_{t+H|t}$. Seo *et al.* (2003) indiquent que cette profondeur d'assimilation peut influencer la pertinence d'une méthode variationnelle classique (recalage). Cette fenêtre est progressivement élargie pour passer d'une assimilation initialement calculée sur un seul pas de temps à un calcul du critère effectué sur 1, 2, 3, 6 et 12 heures.
Là encore, la sensibilité des performances à ce facteur est faible. Sur l'ensemble des bassins versants, les meilleures performances sont obtenues pour une fenêtre d'assimilation réduite à un pas de temps. Comme on pouvait s'y attendre, la largeur optimale de la fenêtre d'assimilation dépend de la réactivité du bassin : les performances pour les bassins les plus rapides sont pénalisées par une assimilation sur plus de 6 pas de temps, chose qui n'est pas observée pour les bassins les plus lents de notre échantillon.

À titre de comparaison : une approche multi-modèle

Pour évaluer l'intérêt d'utiliser les états obtenus avec le jeu de paramètres Θ_0 de base pour les pas de temps antérieurs à la période de rétroaction, censés représenter le comportement à long terme du modèle [5], une approche multi-modèle a été testée. Son calage est identique à celui de la méthode proposée et aboutit à la même bibliothèque de n jeux de paramètres. Le

5. En intégrant la mise-à-jour du niveau du réservoir de routage...

même modèle GRP est ensuite employé en parallèle avec les différents jeux de paramètres. À chaque instant de prévision t, le modèle (i.e. le jeu de paramètres) minimisant les toutes dernières erreurs de prévision observées $\varepsilon_{t-i|t-i-1}$ est sélectionné pour assurer la prévision $\widehat{Q}_{t+H|t}$.
L'approche multimodèle obtient des performances légèrement moins bonnes que l'approche proposée à 1 heure (i.e. l'horizon pour lequel le gain constaté est le plus élevé) puis obtient des performances équivalentes à la méthode précédente mais au prix d'un temps de calcul extrêmement élevé associé à la nature multi-modèle (prohibitif pour nos tests, pas forcément pour le temps réel) qui nous incite à abandonner l'approche.
Avant cet abandon, l'influence de la largeur de la fenêtre d'assimilation, c'est-à-dire le nombre de pas de temps sur lesquels la fonction de coût est calculée pour déterminer le meilleur modèle, a également été testée pour cette approche. Ici encore, les meilleurs résultats globaux ont été obtenus pour une fenêtre réduite à un pas de temps.

Aucun lien avec la réactivité des bassins
Nous n'avons trouvé aucun lien entre l'autocorrélation des débits sur nos différents bassins et le gain de performance dû à cette mise-à-jour des débits. Pas plus de corrélation entre la réactivité des bassins et les performances n'a été détectée pour interpréter les différences de performance entre les variantes testées.

9.3.3 Interprétation

Le résultat le plus frappant est sans doute l'insensibilité des performances à la durée de rétroaction, ce qui va à l'encontre de ce que Yang et Michel (2000) avaient trouvé. Ceci s'explique très probablement par le fait que la capacité du réservoir de production est fixe dans GRP (à la différence du modèle GR4J employé par Yang et Michel, 2000) tandis que le réservoir de routage est mis-à-jour, ce qui annule partiellement l'effet de la mise-à-jour du paramètre X_1 (capacité du réservoir). Comme nous l'avons montré pour l'initialisation du modèle (Berthet *et al.*, 2009a, reproduit au chapitre 11), la mise-à-jour du niveau du réservoir de routage efface très vite l'influence des états précédents (liés à la valeur du paramètre fixant la capacité du réservoir durant la période de rétroaction).

Nous ne savons pas expliquer de façon satisfaisante la perte d'influence de cette méthode de mise-à-jour pour les horizons dépassant 12 h. Nous avons constaté que plus l'horizon de prévision était grand, plus la diversité des paramètres autour du jeu Θ_0 s'amenuise (Fig. 9.5), ce qui conduit finalement à des prévisions très proches à un horizon lointain quelque soit le jeu de paramètres sélectionné. Ceci peut être lié à la diminution de l'influence de la mise-à-jour du niveau du réservoir de routage quand l'horizon augmente.

Notons enfin que le modèle préfère la méthode définissant aléatoirement les sous-ensembles de calage, qui lui assure les jeux de paramètres les plus proches : ces jeux de paramètres

Variante	1 h	3 h	6 h	12 h	24 h
« **Version optimale** » (10 jeux de paramètres calés sur des sous-ensembles aléaoires ; durée de rétroaction égal à 12 heures ; tous les paramètres sont mis-à-jour ; fenêtre d'assimilation réduite au dernier pas de temps)	**47.3**	**31.9**	**32.0**	**37.6**	**45.6**
Sous-ensembles continus	47.4	32.3	28.7	30.0	34.6
Classes de débit	46.5	29.7	21.5	23.7	32.3
$d_{RA} = 24$ h	46.9	32.7	30.4	37.0	45.6
$d_{RA} = 168$ h	47.9	32.1	30.9	37.1	45.2
$d_{RA} = 720$ h	47.3	31.9	32.0	37.5	44.2
5 jeux de paramètres	40.8	30.8	30.4	36.8	44.6
20 jeux de paramètres	49.0	32.8	31.7	36.5	44.5
Sur la production uniquement	44.7	30.4	29.9	36.3	43.6
Sur le routage uniquement	45.5	31.5	30.8	37.3	44.7
Assimilation calculée sur 2 heures	NA	31.3	30.8	37.0	44.3
Assimilation calculée sur 3 heures	NA	29.9	30.3	36.7	45.1
Approche multi-modèle	45.3	32.1	32.5	38.2	46.6
Sans mise-à-jour des paramètres	*40.6*	*27.8*	*29.1*	*37.4*	*46.5*

Table 9.1 – Médiane des performances globales (exprimées en $C_{2M,P}$) des variantes testées les plus intéressantes, calculées sur l'échantillon de 178 bassins versants. Les performances de la combinaison obtenant les meilleurs résultats sont présentées en premier et sont suivies des performances des variantes où un seul facteur est modifié. Ces facteurs sont la définition des sous-ensembles sur lesquels les jeux de paramètres sont calés, la durée de la période de rétroaction, le nombre de jeux de paramètres calés contenus dans la bibliothèque, les paramètres mis-à-jour et la taille de la fenêtre d'assimilation. Les deux dernières lignes donnent les performances du modèle sans cette mise-à-jour et dans l'approche multi-modèle pour comparaison.

auraient été plus divers s'ils avaient été calés par classes de débit par exemple. Trop de diversité dans les jeux de paramètres peut alors amoindrir les performances du modèle. Il ne sert donc à rien de chercher à augmenter la variabilité des jeux de paramètres proposés au modèle.

La méthode a été construite sur un certain nombre d'idées *a priori* mais ces hypothèses restent difficiles à valider. En effet, plusieurs degrés de liberté apparaissent dans cette approche et les résultats des études de sensibilité menées ont pu fournir une solution globalement optimale mais restent difficilement interprétables.

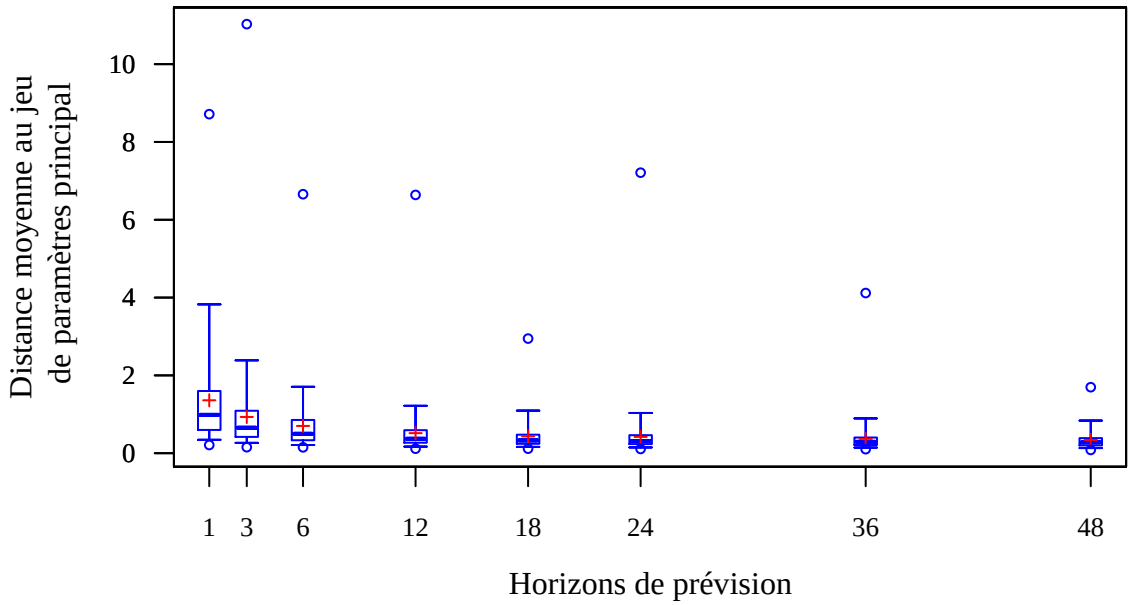


FIGURE 9.5 – Distribution sur 178 bassins de la distance euclidienne moyenne des $n = 10$ jeux de paramètres calés sur des sous-ensembles de la période de calage au jeu de paramètres de base Θ_0 (calé sur la période de calage entière). Les distances sont calculées pour les paramètres transformés : la transformation sert au calage pour définir la sensibilité (le pas de recherche) en fonction de la zone de l'espace des paramètres et encadre chaque paramètres dans le segment $[-10, 10]$, ce qui permet ici de traiter chaque paramètre de façon équitable. La distance moyenne diminue avec l'horizon de prévision : à horizon court, la bibliothèque offre plus de possibilités de mise-à-jour forte qu'à des horizons plus lointains.

9.4 Bilan

Même si les paramètres sont définis comme des invariants temporels, la faiblesse de la connaissance que nous en avons explique le besoin de les mettre à jour à mesure que de nouvelles observations sont disponibles. De nombreuses méthodes de mise-à-jour des données ont été proposées, tant variationnelles que séquentielles (notamment probabilistes).

Mettre à jour les paramètres (surtout quand il s'agit de la seule mise-à-jour du modèle) peut amoindrir la robustesse des prévisions. La méthode développée par Yang et Michel (2000) se caractérise par une recherche de robustesse. Malheureusement, elle n'obtient pas des performances encourageantes. Nous avons proposé une variante de cette méthode qui limite les jeux de paramètres à un ensemble restreint de jeux préalablement calés et qui est combinée à la mise-à-jour d'un état du modèle. Cette approche permet des gains de performance intéressants mais son action reste difficilement interprétable.

10
Chapitre

Mise-à-jour des sorties

¿ qué causa tiene para volverse loco ? ¿ Qué dama le ha desdeñado, o qué señales ha hallado que le den a entender que la señora Dulcinea del Toboso ha hecho alguna niñería con moro o cristiano ?

-Ahí esta el punto -respondió don Quijote- y ésa es la fineza de mi negocio ; que volverse loco un caballero andante con causa, ni grado ni gracias : el toque está desatinar sin ocasión y dar a entender a mi dama que si en seco hago esto, ¿qué hiciera en mojado ? Cuanto más, que harta ocasión tengo en la larga ausencia que he hecho de la siempre señora mía Dulcinea del Toboso ; que, como ya oíste decir a aquel pastor de marras, Ambrosio : quien está ausente todos los males tiene y teme. Así que, Sancho amigo, no gastes tiempo en aconsejarme que deje tan rara, tan felice y tan no vista imitación. Loco soy, loco he de ser hasta tanto que tú vuelvas con la respuesta de una carta que contigo pienso enviar a mi señora Dulcinea ; y si fuere tal cual a mi fe se le debe, acabarse ha mi sandez y mi penitencia ; y si fuere al contrario, seré loco de veras, y, siéndolo, no sentiré nada.
Ansí que, de cualquiera manera que responda, saldré del conflito y trabajo en que me dejares, gozando el bien que me trujeres, por cuerdo, o no sintiendo el mal que me aportares, por loco.

Miguel de Cervantes Saavedra, *L'Ingénieux Hidalgo Don Quijote de la Manche*, (traduction page 551)

Objectifs

- ▷ Faire le lien entre modélisation de l'erreur et mise-à-jour des sorties.
- ▷ Présenter plusieurs méthodes de mise-à-jour des sorties du modèle hydrologique.
- ▷ En évaluer les performances prédictives.
- ▷ Établir leur potentiel dans un cadre opérationnel.

Sommaire

10.1 L'erreur du modèle hydrologique n'est pas du bruit blanc . . . 176
10.2 Peut-on débiaiser les sorties du modèle ? . . . 176
10.2.1 Essais réalisés . . . 177
10.2.2 Résultats . . . 179
10.2.3 Discussion . . . 180
10.3 Régressions simples . . . 180
10.3.1 Report d'erreur . . . 180
10.3.2 Correction multiplicative régressive . . . 181
10.4 Correction par modèles ARIMA . . . 183
10.4.1 Principe . . . 183
10.4.2 Démarche expérimentale . . . 185
10.4.3 Résultats . . . 186
10.5 Correction par réseaux de neurones artificiels . . . 188
10.5.1 Qu'est-ce qu'un réseau de neurones ? . . . 188
10.5.2 Structure et configuration du réseau de neurones . . . 191
10.5.3 Résultats . . . 194
10.5.4 Cette méthode est-elle opérationnelle ? . . . 197
10.6 Bilan . . . 200

Les mises-à-jour étudiées jusqu'ici avaient toutes en commun de modifier tel ou tel élément du modèle hydrologique sur la base de nouvelles observations de débit : reconnaissant l'imperfection de notre connaissance du système hydrologique et des processus en jeu, nous utilisions ces nouvelles données pour améliorer notre représentation du bassin versant. La mise-à-jour des sorties du modèle (ou correction de l'erreur) vise quant à elle à employer l'information sur la structure statistique de l'erreur pour modifier les prévisions sans modifier le modèle hydrologique lui-même : il s'agit d'utiliser un modèle d'erreur des sorties du modèle hydrologique (ou une ébauche de modèle).

Les corrections des sorties comptent parmi les approches de mise-à-jour les plus populaires car elles sont simples à mettre en place (Moore, 2007). Leur implémentation est indépendante de celle du modèle, ce qui permet de créer des outils modulaires de prévision comportant différents modèles hydrologiques et différentes procédures de correction (e.g. O'Connor, 2006). Parmi les méthodes les plus souvent employées, on compte les fonctions de transfert (e.g Shamseldin, 2006), les modèles ARIMA (Box et Jenkins, 1976), les réseaux de neurones artificiels (e.g. Hsu *et al.*, 1995) et la programmation génétique (Khu *et al.*, 2001). Plusieurs auteurs (par exemple Xiong et O'Connor, 2002, Goswami *et al.*, 2005, Madsen *et al.*, 2000) ont proposé des comparaisons des performances de différentes approches dont les conclusions ne sont pas toujours concordantes. Leurs résultats dépendent de l'horizon de prévision, des bassins versants et probablement du modèle hydrologique employé.

Nous entamerons ce chapitre en caractérisant rapidement la nature de l'erreur en sortie du modèle puis nous proposerons et comparerons différentes méthodes permettant de réduire cette erreur. La mise-à-jour du réservoir de routage étant apparue comme cruciale pour obtenir de bonnes prévisions, le modèle « de base » ici sera le modèle GRP avec mise-à-jour de cet état.

Dans tout ce chapitre, nous noterons (Fig. 10.1) :

H	l'horizon de prévision (exprimé en nombre de pas de temps),
Q_t	le débit observé au pas de temps t,
$\widehat{Q}_{t+H\|t}$	le débit prévu par le modèle (dont les états sont mis-à-jour) au pas de temps t pour le pas de temps $t+H$ avant toute correction de l'erreur en sortie du modèle,
$\widehat{\widehat{Q}}_{t+H\|t}$	le débit prévu par le modèle après correction de l'erreur,
$\varepsilon_{t+H\|t} = Q_{t+H} - \widehat{Q}_{t+H\|t}$	l'erreur de prévision à l'horizon H faite au pas de temps t par le modèle avant correction de l'erreur,
$\widehat{\varepsilon}_{t+H\|t} = \widehat{\widehat{Q}}_{t+H\|t} - \widehat{Q}_{t+H\|t}$	la correction de l'erreur de prévision faite au pas de temps t pour l'horizon H.

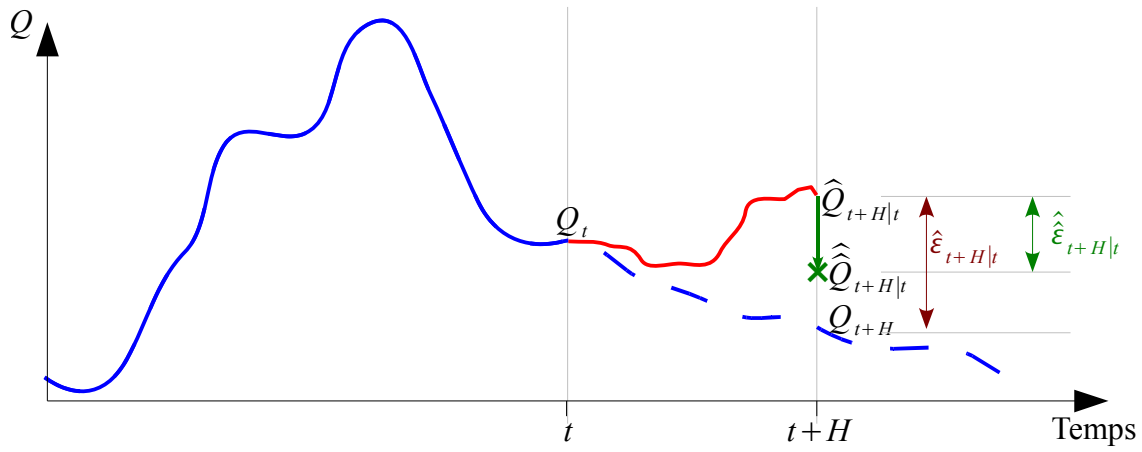


FIGURE 10.1 – Notations employées dans ce chapitre. t est l'instant de prévision.

10.1 L'erreur du modèle hydrologique n'est pas du bruit blanc

Nous avons vu au chapitre 3, que l'erreur des modèles hydrologiques était, le plus souvent, loin d'être du bruit blanc. Établir et valider un modèle d'erreur est loin d'être chose facile (e.g. Beven, 2008b). Yang *et al.* (2007) ont, par exemple, proposé un modèle d'erreur basé sur deux grandes caractéristiques de l'erreur des modèles hydrologiques : l'hétéroscédasticité(g)et la dépendance temporelle. Ils ont montré la pertinence de ce modèle, sauf sur les plus fortes erreurs. Berthet *et al.* (2009) ont constaté que ce modèle d'erreur n'était pas complètement validé pour le modèle hydrologique GRP (en annexe F).

10

Commençons par mettre en évidence les deux caractéristiques du modèle d'erreur pointées précédemment pour les erreurs en sortie du modèle après la mise-à-jour de ses états. Ces erreurs se caractérisent d'abord par une dépendance temporelle assez forte, révélée par les courbes d'autocorrélation des erreurs (Fig. 10.2). La figure 10.3 illustre l'hétéroscédasticité des erreurs des modèles hydrologiques : elle représente l'écart-type des erreurs faites par le modèle pour différentes gammes de débits sur l'ensemble des bassins. Il apparaît clairement que l'écart-type des erreurs dépend du débit.

Nous avons donc vérifié que le bruit en sortie de GRP (avec mise-à-jour des états) est loin d'être blanc : il contient de l'information qui est *a priori* encore valorisable pour réduire les incertitudes de prévision.

10.2 Peut-on débiaiser les sorties du modèle ?

Il semble naturel *a priori* de commencer par débiaiser les sorties du modèle en se basant sur les erreurs observées sur la période de calage. Comme les erreurs sont hétéroscédastiques, il

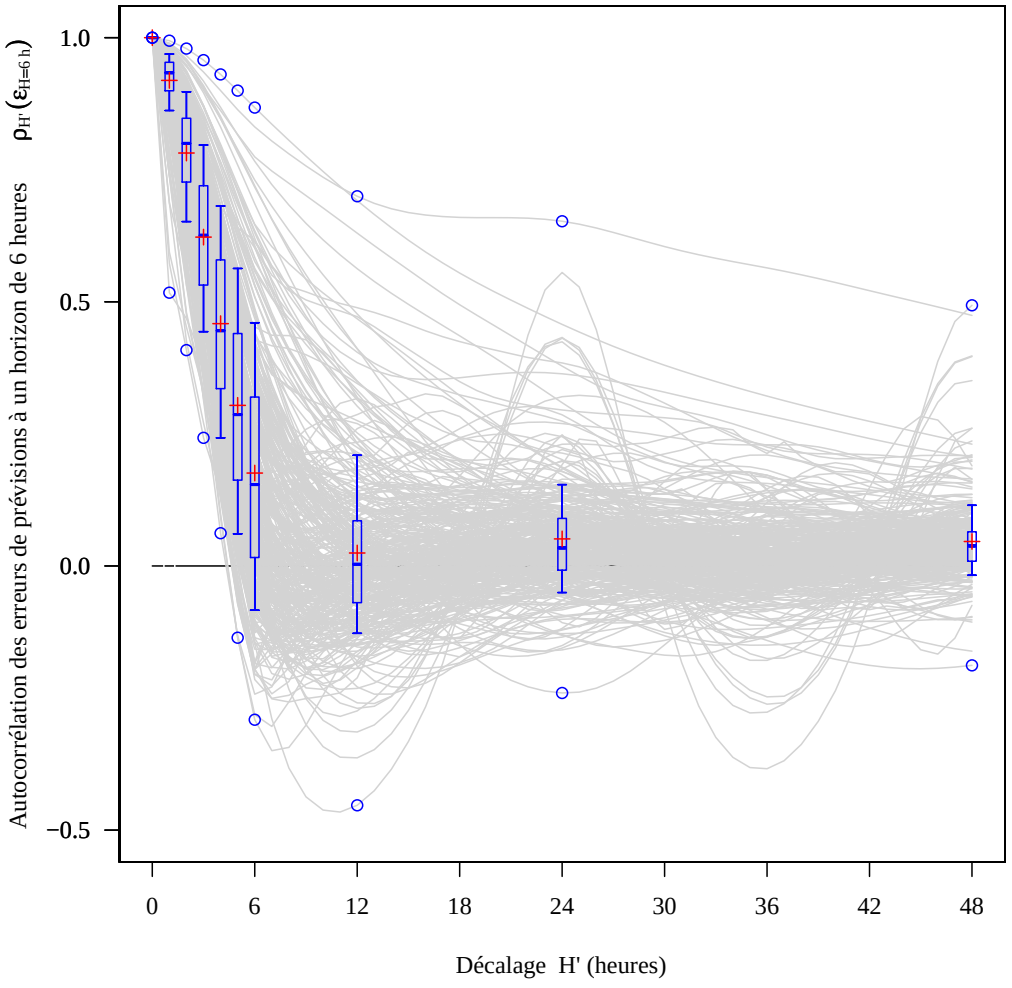


Figure 10.2 – Trace des autocorrélations des erreurs de prévision pour un horizon de 6 heures pour nos 178×2 bassins-périodes. Les boîtes à moustaches indiquent les distributions des autocorrélations pour plusieurs décalages. On observe pour un petit nombre de bassins, un comportement pseudo-périodique expliqué dans l'annexe O.2, page 551.

est nécessaire de réaliser ce débiaisage pour un débit donné ou plutôt pour une gamme de débits donnée (pour avoir un nombre significatif de points). Pour chaque gamme de débit, le biais multiplicatif (entre débits observé et prévu sans mise-à-jour des sorties du modèle) est calculé lors du calage, puis appliqué aux débits correspondants lors du contrôle.

10.2.1 Essais réalisés

Définition des gammes de débit

Plusieurs manières de définir les gammes de débit prévu au calage ont été testées. Une première méthode est de couper l'intervalle des débits prévus sur la période de calage en

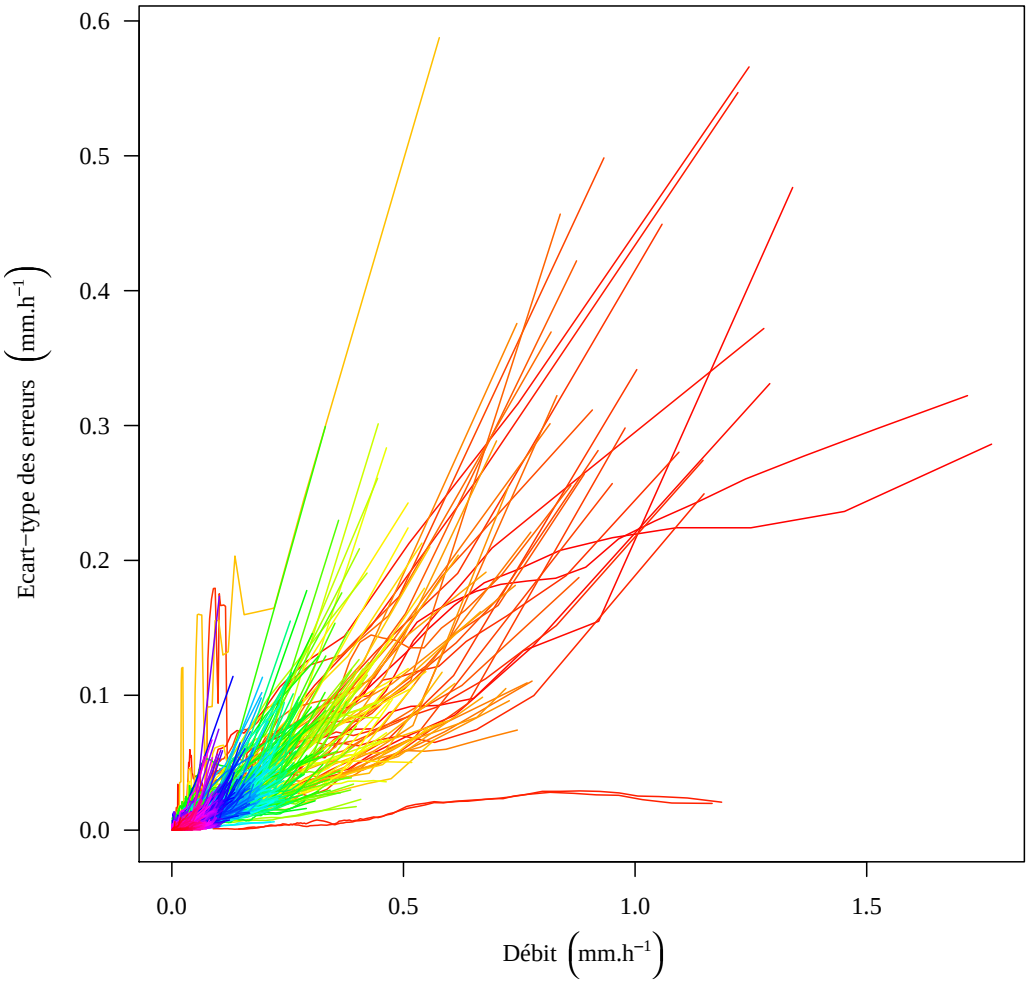


FIGURE 10.3 – Illustration de l'hétéroscédasticité des erreurs. Chaque courbe représente l'évolution de l'écart-type des erreurs faites sur les débits appartenant à une fenêtre glissante pour un des 178×2 bassins-périodes. Les fenêtres glissantes correspondent à la plage de débits délimitée par les quantiles $q_{\frac{i}{100}}(Q)$ et $q_{\frac{i+5}{100}}(Q)$ pour i allant de 0 à 95.

segments de longueurs égales. Une seconde méthode découpe cet intervalle au niveau de quantiles de débit prévus $q_{\frac{i}{N}}\left(\widehat{Q}\right)$. La première méthode a l'avantage d'avoir des gammes de dimensions égales, mais avec des effectifs très variables (certains segments comptant très peu voire aucun point), ce qui peut poser un problème statistique lors du calcul du biais. La seconde méthode permet d'assurer un nombre suffisant de points dans tous les segments mais au prix de segments très larges (dans la queue de distribution).

Méthodes de débiaisage

Nous avons d'abord distingué deux méthodes selon que nous autorisions ou non un débiaisage en extrapolation (c'est-à-dire pour des débits non observés sur la période de calage) en

prenant alors le biais du segment des débits les plus élevés observés au calage. Nous avons également exploré la distinction entre montée et descente de crue. Enfin, certains essais ont été repris en ne définissant plus les segments de débits en fonction du débit prévu $\widehat{Q}_{t+H|t}$ mais en fonction du dernier débit observé Q_t. Toutes les combinaisons d'essais possibles n'ont pas été réalisées (les premières étant non fructueuses), le tableau 10.1 résume les essais entrepris.

N°	Gamme	Facteur	Segments	Distinction Montée / Descente
1	$[0, \infty]$	$\widehat{Q}_{t+H\|t}$	Longueur identique	Non
2	$\left[0, \widehat{Q}_{lim}\right]$	$\widehat{Q}_{t+H\|t}$	Longueur identique	Non
3	$\left[0, \widehat{Q}_{lim}\right]$	$\widehat{Q}_{t+H\|t}$	Longueur identique	Oui
4	$[0, Q_{lim}]$	Q_t	Longueur identique	Non
5	$\left[0, \widehat{Q}_{lim}\right]$	$\widehat{Q}_{t+H\|t}$	Sur quantiles de Q_{Calage}	Non
6	$\left[0, \widehat{Q}_{lim}\right]$	$\widehat{Q}_{t+H\|t}$	Sur quantiles de Q_{Calage}	Oui
7	$[0, Q_{lim}]$	Q_t	Sur quantiles de Q_{Calage}	Non

Table 10.1 – Caractéristiques des essais de débiaisage entrepris

10.2.2 Résultats

Globalement aucun essai de débiaisage des prévisions n'a donné de résultats significativement positifs. L'essai le plus intéressant est l'essai N° 6 (découpage des segments selon les quantiles du débit prévu sur la période de calage et distinction entre biais à la montée et biais à la descente). Cet essai dépasse le modèle non débiaisé sur plusieurs horizons de prévision et sur plusieurs classes de bassins (rapides à lents) sans que le gain n'atteigne des niveaux significatifs. Cette variante reste cependant très légèrement moins bonne que le modèle non débiaisé sur l'ensemble des horizons de prévision (là encore les différences sont non significatives).

La vitesse de réaction du bassin, vue à travers l'autocorrélation des débits semble être un facteur explicatif du gain obtenu par le débiaisage (pour l'essai N°6). Alors que pour les horizons de prévision les plus courts (inférieurs ou égaux à 6 heures), on n'observe pas de différence sur le gain en performance par le débiaisage selon les classes de bassins, le débiaisage n'est utile pour les horizons plus lointains (à partir de 12 heures) que pour les bassins les plus lents (ceux dont le débit est le plus fortement autocorrélé). A ces horizons, le débiaisage dégrade les performances du modèle sur les bassins les plus rapides, parfois de façon significative (plus de 2 ou 3 points de $C_{2M,P}$).

On observe pour tous les essais (sauf l'essai N° 6) une dégradation très forte des perfor-

mances les plus faibles, ce qui met en question la robustesse du débiaisage et la pertinence de l'approche quand le débiaisage ne s'appuie pas sur la bonne variable.

10.2.3 Discussion

Deux constats peuvent expliquer la difficulté à débiaiser les erreurs en se basant sur les erreurs observées au calage. Tout d'abord, les biais constatés sur la période de calage ne sont pas identiques à ceux de la période de contrôle sur un nombre important de bassins versants (Fig. 10.4). Pour un horizon de 12 heures par exemple, les biais sont de signes opposés dans 31 % des cas. En outre, le débiaisage en extrapolation est difficile à justifier, or ce sont les erreurs sur les plus gros débits que nous avons le plus besoin de corriger...

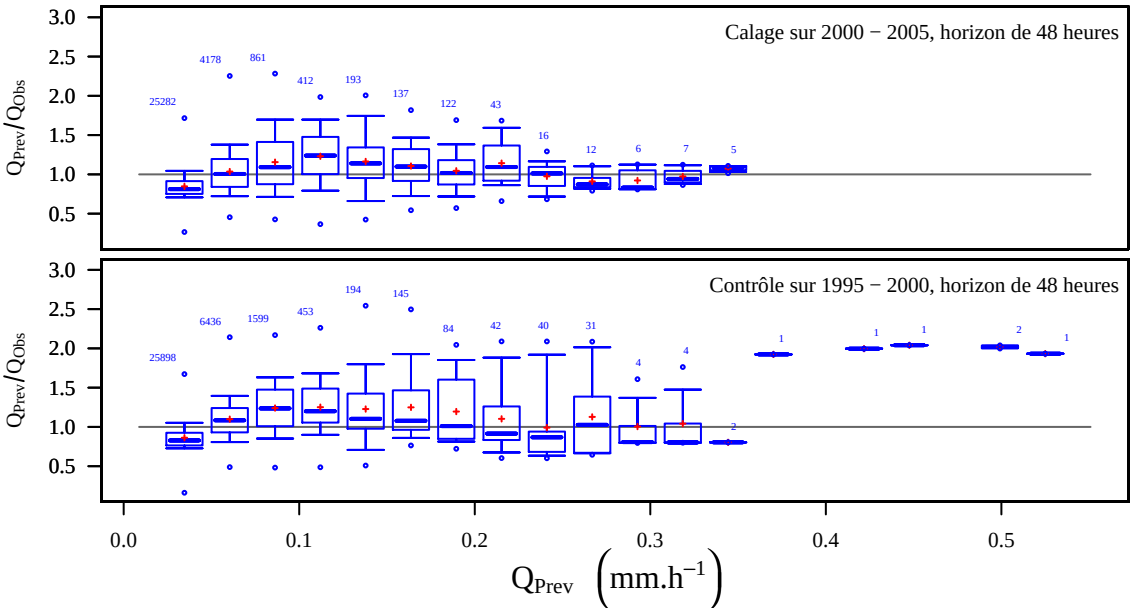


FIGURE 10.4 – Tracé des statistiques de biais des prévisions à 48 heures sur le bassin de la Moder à Schweighouse-sur-Moder (622 km^2, A3301010) : on observe que les deux gammes de débits ne correspondent pas d'une période à l'autre et les gammes de ratio sont étendues (on peut trouver des gammes beaucoup plus étendues).

10.3 Régressions simples

10.3.1 Report d'erreur

La correction d'erreur la plus simple (et l'une des plus couramment utilisée) est le report de l'erreur constatée au pas de temps d'émission de la prévision :

$$\widehat{\varepsilon}_{t+H|t} = \varepsilon_{t|t-H'} \quad \text{soit } \widehat{\widehat{Q}}_{t+H|t} = \widehat{Q}_{t+H|t} + \left(Q_t - \widehat{Q}_{t|t-H'}\right)$$

où H est l'horizon de prévision et H' un nombre entier de pas de temps.

Cas $\mathbf{H' = 1}$

Le plus souvent, H' est choisi égal à une heure : la toute dernière erreur faite est reportée. Cette méthode apporte des gains significatifs aux horizons de prévisions courts (Fig. 10.5), puis son effet s'amenuise à mesure que l'horizon de prévision s'éloigne. Ce comportement s'explique aisément par le fait que les erreurs du modèle dont les états sont mis-à-jour croissent avec l'horizon de prévision : le report d'erreur est donc inefficace aux horizons de prévisions lointains (supérieurs à 12 heures).

Les observations corroborent ici l'intuition : plus l'autocorrélation des débits est élevée (i.e., plus le bassin est « lent »), plus la correction par report d'erreur apporte un gain élevé en performance et plus l'horizon de prévision au delà duquel l'effet de la correction est nul, est éloigné.

Cas $\mathbf{H' = H}$

Pour tenter de remédier au défaut mentionné dans le test précédent, nous choisissons maintenant H' égal à l'horizon. Cette approche ne permet pas de corriger ce défaut (Fig. 10.5). Au contraire, les performances aux horizons lointains sont dégradées.

Amplification de l'erreur à 1 pas de temps

Le test précédent a montré qu'il était plus bénéfique d'employer la dernière erreur $\varepsilon_{t|t-1}$ plutôt que l'erreur $\varepsilon_{t|t-H}$ plus lointaine mais d'ordre de grandeur plus convenable. Nous avons également testé une amplification de la dernière erreur :

$$\widehat{\widehat{Q}}_{t+H|t} = \widehat{Q}_{t+H|t} + \left(Q_t - \widehat{Q}_{t|t-H'}\right) \cdot \frac{\overline{\left|\varepsilon_{k+H|k}\right|}^{\text{Calage}}}{\overline{\left|\varepsilon_{k+1|k}\right|}^{\text{Calage}}}$$

où l'erreur est amplifiée du rapport des moyennes des erreurs (avant correction) aux horizons de H et 1 heures mesurées au calage.

Cette méthode dégrade significativement les performances dès l'horizon de trois heures. L'amplification des erreurs est clairement une approche dangereuse.

10.3.2 Correction multiplicative régressive

Tangara (2005) propose une correction de l'erreur non paramétrique [1] qui cherche à exploiter les deux caractéristiques décrites précédemment des erreurs du modèle. Il s'agit d'un

1. Son paramètre γ a été fixé pour tous les bassins et pour tous les horizons, pour un pas de temps donné (Chap. 5 et Ann. C).

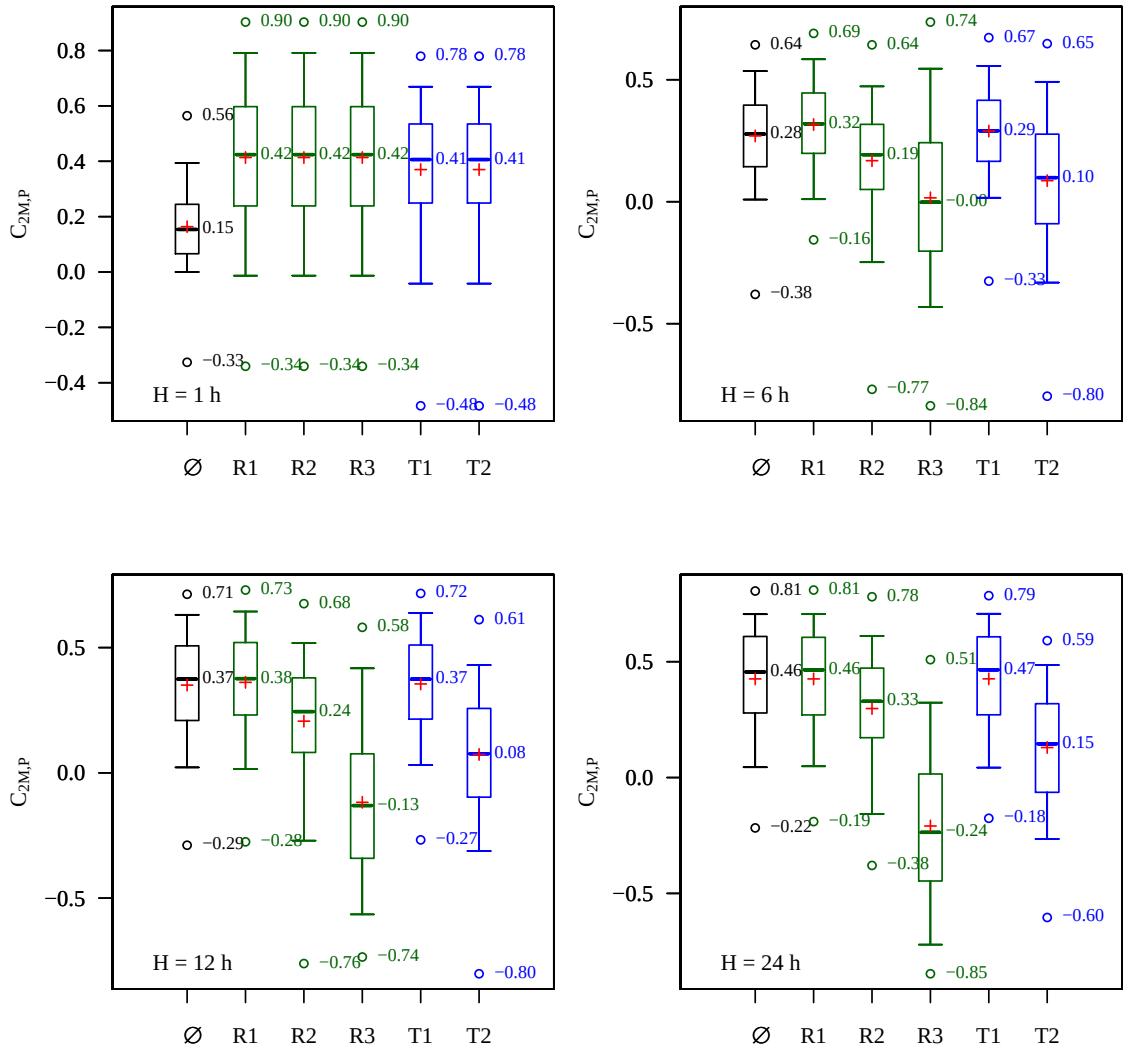


FIGURE 10.5 – Comparaison des performances de prévision de GRP avec 5 corrections d'erreur par régressions simples aux horizons de prévision de 1, 6, 12 et 24 heures. Le premier essai, noté ∅, a été réalisé sans correction de l'erreur du modèle hydrologique. R1 désigne le report de l'erreur $\varepsilon_{t|t-1}$, R2 est le report de l'erreur $\varepsilon_{t|t-H}$ et R3, le report de l'erreur $\varepsilon_{t|t-1}$ amplifiée. T1 correspond à la correction de Tangara prenant en compte l'erreur $\varepsilon_{t|t-1}$ tandis que T2 est cette même correction avec l'erreur $\varepsilon_{t|t-H}$.

modèle auto-régressif portant sur une transformation logarithmique de l'erreur :

$$\widehat{\widehat{Q}}_{t+H|t} = \widehat{Q}_{t+H|t} \cdot \left(\frac{Q_t}{\widehat{Q}_{t|t-1}} \right)^{\gamma}$$

Gains en performance

Cette correction apporte des gains en performance du même ordre de grandeur que le report d'erreur, légèrement inférieurs aux horizons de prévision les plus courts, légèrement supérieurs pour les horizons lointains (Fig. 10.5).

Peut-on adapter le coefficient α aux différents bassins

Nous avions vu lors de l'adaptation du modèle GRP (Tangara, 2005) du pas de temps journalier au pas de temps horaire qu'un léger gain (peu significatif) pouvait être obtenu en calant le paramètre (ici fixe) α. Nous avons tenté de lier ce paramètre à des caractéristiques du bassin – l'autocorrélation des débits ou celle des erreurs – en choisissant une fonction f de $[0,1]$ dans $[0,1]$, telle que $\alpha = f\left(\rho_H\left(Q\right)\right)$ ou $\alpha = f\left(\rho_H\left(\varepsilon(Q)\right)\right)$. Plusieurs fonctions f ont été testées : des fonctions puissance et des arcs de cercle (sur le carré $[0,1]^2$). Aucune n'a permis d'obtenir un gain significatif permettant de relier l'intensité de la correction à une caractéristique du bassin.

Limites d'emploi

L'effet de la correction multiplicative régressive s'éteint aux horizons lointains, pour des raisons similaires à celles de l'extinction de l'effet du report d'erreur[2]. Nous avons donc réalisé une expérience similaire à celles menées pour le report de l'erreur en modifiant la correction multiplicative pour qu'elle prenne en compte l'erreur faite H pas de temps auparavant ($\varepsilon_{t|t-H}$) et non l'erreur à 1 heure ($\varepsilon_{t|t-1}$). Ici encore, nous observons que les performances sont moins bonnes (Fig. 10.5) : l'erreur la plus récente est plus informative que l'erreur antérieure faite H pas de temps avant. Cela est d'autant plus vrai que l'autocorrélation des débits est faible (bassins rapides). De manière générale, ces expériences montrent qu'il est préférable de s'intéresser aux erreurs les plus récentes pour mettre à jour le modèle, même si elles n'ont pas la bonne variance. Nous allons donc nous tourner vers une méthode plus générale prenant en compte les toutes dernières erreurs.

10.4 Correction par modèles ARIMA

10.4.1 Principe

Les modèles auto-régressifs à moyenne mobile constituent une famille de modèles stochastiques linéaires employés dans de très nombreux domaines. Ils traitent principalement la

2. Une démonstration est proposée dans l'annexe O.3, page 552.

dépendance temporelle de la variable stochastique. Nous les emploierons dans ce but pour réduire les erreurs en sortie du modèle.

Formulation à un pas de temps

La formalisation présentée par Box et Jenkins (1976) définit ces modèles comme la somme d'un modèle auto-régressif (AR) d'ordre p et d'un modèle à moyenne mobile (MA) d'ordre q appliqués à la différentiation discrète (I) d'ordre d de la variable stochastique z à modéliser. Le modèle AR(p) valorise l'autocorrélation du signal tandis que le modèle MA(q) rend compte de la partie aléatoire de ce signal :

$$\widehat{w}_{t|t-1} = \underbrace{-\sum_{i=1}^{p} \varphi_i w_{t-i}}_{AR(p)} + \underbrace{\sum_{j=1}^{q} \theta_j a_{t-j}}_{MA(q)} + a_t \quad \text{où} \quad \underbrace{w_t = \nabla^d z_t}_{I(d)} \tag{10.1}$$

où

- $\widehat{z}_t$ est la prédiction du modèle ARIMA (ici le débit simulé par le modèle hydrologique ou l'erreur de ce modèle),
- a_t est un bruit blanc (i.e. un « choc » tiré dans une loi gaussienne),
- ∇ est l'opérateur *différentiation* : $\nabla z_t = z_t - z_{t-1}$;
- φ_i sont les paramètres de la partie auto-régressive du modèle (AR),
- θ_j sont les paramètres de la partie « moyenne mobile » du modèle (MA).

Un modèle ARIMA est donc caractérisé par le triplet d'entiers $\{p, d, q\}$.

Sous conditions sur les paramètres (voir annexe O.4, page 554), ces modèles ont deux propriétés importantes pour nous :

- la **stationnarité**(g),
- l'**inversibilité**(g), qui assure que la valeur de la variable z_t ne dépend que de la valeur de cette variable à quelques pas de temps précédents ($t-1$, $t-2$, ...).

Nous appliquerons ce modèle à l'erreur $\varepsilon_{t|t-1} = Q_t - \widehat{Q}_{t|t-1}$ en sortie du modèle GRP, considérée ici comme étant une variable stochastique : $z_t = \varepsilon_{t|t-1} - \overline{\varepsilon_1}$ où $\overline{\varepsilon_1}$ est la moyenne des erreurs à un pas de temps.

Généralisation à H pas de temps

Pour atteindre des horizons plus lointains, l'équation (10.1) est employée de façon récursive (Moore, 2007) pour déterminer $\widehat{z}_{t+1|t}$, $\widehat{z}_{t+2|t}$, ..., $\widehat{z}_{t+H|t}$:

$$\widehat{w}_{t+H|t} = -\sum_{i=1}^{p} \varphi_i \widehat{w}_{t+H-i|t} + \sum_{j=1}^{q} \theta_j a_{t+H-j|t} + a_{t+H} \quad \text{où } w_t = \nabla^d \left(\varepsilon_{t|t-1} - \overline{\varepsilon_1}\right) \tag{10.2}$$

avec

$$a_{t+H-j|t} = \begin{cases} 0 & \text{si } H-j>0 \\ a_{t+H-j|t+H-j-1} = w_{t+H-j} - w_{t+H-j|t+H-j-1} & \text{sinon} \end{cases}$$

$$\widehat{w}_{t+H-i|t} = w_{t+H-i} \text{ pour } H-i \leq 0$$

Ici, les chocs a_t sont égaux à leurs espérances : 0 pour les pas de temps futurs et à leurs réalisations pour les pas de temps passés. De même, les valeurs de la variable modélisée (différentiée) w sont prises égales aux réalisations pour les pas de temps passés et égales aux prédictions pour les pas de temps futurs.

Dans le cas où $d=0$ et où l'erreur moyenne est nulle, l'équation (10.2) se simplifie en :

$$\widehat{\varepsilon}_{t+H|t} = -\sum_{i=1}^{p} \varphi_i \widehat{\varepsilon}_{t+H-i|t} + \sum_{j=1}^{q} \theta_j a_{t+H-j|t} + a_{t+H} \tag{10.3}$$

où

$$a_{t+H-j|t} = \begin{cases} 0 & \text{si } H-j>0 \\ a_{t+H-j|t+H-j-1} = Q_{t+H-j} - \widehat{Q}_{t+H-j|t+H-j-1} & \text{sinon} \end{cases}$$

$$\widehat{\varepsilon}_{t+H-i|t} = \begin{cases} Q_{t+H-i} - \widehat{Q}_{t+H-i|t} & \text{pour } H-i \leq 0 \\ \widehat{\widehat{Q}}_{t+H-i|t} - \widehat{Q}_{t+H-i|t} & \text{sinon} \end{cases}$$

10.4.2 Démarche expérimentale

Nous avons employé ces modèles pour corriger les erreurs du modèle GRP avec sa mise-à-jour du réservoir de routage (Tangara, 2005). Box et Jenkins (1976) conseillent d'employer le modèle pour des triplets $\{p,d,q\}$ petits (entre 0 et 2) : nous avons donc testé les 24 configurations possibles[3].

Les mêmes auteurs indiquent que la différentiation est surtout utile dans le cas où la variable stochastique modélisée présente une autocorrélation non négligeable pour des décalages temporels H' supérieurs à quelques pas de temps. Avant de tester les modèles ARIMA, nous avons donc examiné les autocorrélations des erreurs $\rho_{H'}\left(\varepsilon_{t+H|t}\right)$ et des différentielles $\rho_{H'}\left(\nabla^d\left(\varepsilon_{t+H|t}\right)\right)$. Nous avons constaté que les erreurs sont assez corrélées entre elles : l'autocorrélation diminue assez lentement. Plus l'horizon de prévision H est élevé, plus l'autocorrélation est élevée et diminue lentement. Si on prend une différentielle d'ordre 1 ou 2, les autocorrélations diminuent beaucoup plus vite en fonction de H', même pour des

3. C'est-à-dire les 3^3 combinaisons auxquelles on retire les combinaisons caractérisées par $(p,q)=(0,0)$.

horizons de prévision H élevés. Il paraît donc nécessaire de tester les modèles ARIMA pour des degrés de différentiation non nuls.

10.4.3 Résultats

Comparaisons des performances

Les performances globales de la méthode de correction par modèle ARIMA sont très nettement supérieures à celles des méthodes précédentes : aux horizons de prévision les plus courts (1 et 3 heures), le gain médian est de l'ordre de 10 points de $C_{2M,P}$ et plus de 90 % des bassins versants voient leurs performances s'améliorer ou rester du même ordre de grandeur (Fig. 10.6). Au-delà le gain diminue jusqu'à devenir négligeable en suivant le rythme de diminution de l'efficacité de la correction de Tangara (2005).

La composante autorégressive AR semble nécessaire et plus importante que la composante à moyenne mobile MA. Les ordres p et q n'ont pas d'influence forte sur l'ensemble des bassins.

Les meilleurs résultats aux horizons de prévision courts (i.e. inférieurs à 12 heures) sont obtenus par les modèles ARMA ($d = 0$). Les modèles ARIMA avec $d = 1$ sont très proches de la correction de Tangara (2005) pour un horizon de prévision d'une heure, sont meilleurs que cette correction pour les autres horizons inférieurs à 12 heures, puis lui sont équivalents. Les essais réalisés avec une double différentiation ($d = 2$) semblent plus instables et n'obtiennent généralement pas de bons résultats.

Aucun lien avec les performances sur le bassin, ni avec l'autocorrélation des débits (caractérisant la réactivité du bassin versant) n'a pu être établi.

L'usage qui a été fait jusqu'ici de ces modèles s'est focalisé sur les propriétés temporelles des erreurs du modèle et ne s'est pas intéressé à leur hétéroscédasticité. Nous avons également testé les modèles ARIMA sur les transformées logarithmiques du débit. Les performances des meilleurs modèles sont légèrement inférieures à celles des modèles ARIMA calculés sur les débits. De plus, ces modèles sont moins robustes dès que le signal est différencié (dès $d = 1$). De façon générale, on retrouve un résultat similaire à la comparaison du report d'erreur additif et de la correction de Tangara : la prise en compte simple de l'hétéroscédasticité des erreurs par une transformation logarithmique n'est pas efficace.

Peut-on déterminer la meilleure configuration ARIMA ?

Nous avons étudié 24 configurations possibles du modèle ARIMA. Dans la pratique, comment le prévisionniste peut-il choisir la configuration qui lui convient ? Pour plus de 90 % des bassins versants, la différence entre les performances en contrôle de la meilleure configuration du modèle ARIMA et les performances en contrôle de la configuration ayant obtenu les meilleures performances au calage, est faible (et probablement non significative). En pratique, il est donc possible de s'appuyer sur la comparaison des résultats du calage d'une

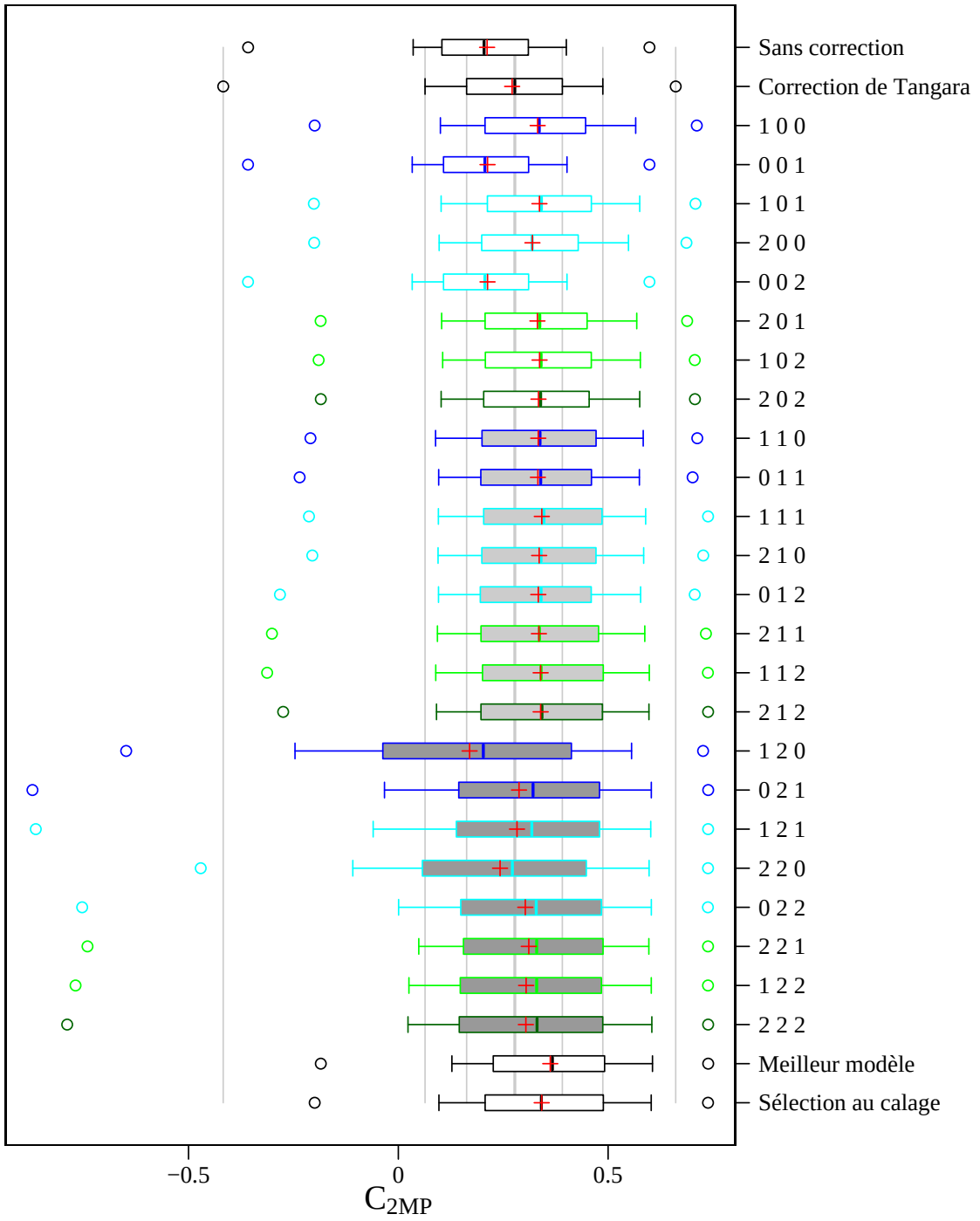


Figure 10.6 – Performances globales mesurées à l'aide du critère $C_{2M,P}$ pour les 24 configurations de la correction d'erreur par modèle ARIMA testées (repérées par le triplet $\{p, d, q\}$). Les couleurs des boîtes correspondent au niveau de complexité des configurations (nombre de paramètres $p + q$) tandis que l'intensité du fond gris indique le degré de différentiation d. Les deux dernières boîtes représentent les distributions des meilleures performances (au contrôle) et des performances au contrôle du modèle ayant obtenu le meilleur score au calage (pour chaque bassin). L'horizon est de trois heures.

batterie de configurations différentes (le calcul est relativement court) pour sélectionner celle à utiliser. On peut également procéder à une validation croisée, ce qui demande bien sûr plus de données.

Limites

L'efficacité des modèles auto-régressifs que nous avons présentés est limitée par l'horizon de prévision. Les paramètres φ_i étant choisis de module inférieur à 1 pour assurer la stationnarité du processus, les corrections apportés par ces modèles ne peuvent qu'être très faibles pour des horizons supérieurs à quelques pas de temps : c'est bien ce qu'on observe dès des horizons de 12 heures alors que l'ordre de grandeur des erreurs à des horizons plus lointains augmente ! Nous avons tenté sans succès de dépasser ce problème en appliquant le modèle ARIMA aux erreurs de prévision à des horizons plus lointains : $z_t = \varepsilon_{t|t-H} - \overline{\varepsilon_H}$.

10.5 Correction par réseaux de neurones artificiels

10.5.1 Qu'est-ce qu'un réseau de neurones ?

Origine et principe

Les réseaux de neurones artificiels (RNA) ont pour origine des travaux en sciences cognitives et en calcul parallèle. Ils constituent une famille d'outils pour la modélisation [4]. Ces outils sont des réseaux d'unités élémentaires. Chaque unité (appellée *neurone*) « contient » une information (une variable nommée *état*(g)). Elle reçoit des informations d'autres unités, réalise une opération simple (via sa *fonction d'activation*(g)) à partir de ces informations pour déduire son état, information qu'elle transmet à son tour à d'autres unités.

La puissance des RNA ne provient pas des opérations élémentaires réalisées par chaque unité mais de la très grande flexibilité liée à sa structure en réseau : ces outils contiennent très peu d'information a priori sur le phénomène qu'on souhaite leur faire modéliser. Ce sont par essence des outils « boîtes noires » dont la flexibilité peut conduire à très bonnes performances... mais aussi à des échecs retentissants.

Quels usages en fait-on ?

Les réseaux de neurones artificiels ont des emplois très variés : leur grande flexibilité leur permet de réaliser des régressions (non linéaires) sur des séries d'entrées – sorties de natures (statistiques) très différentes. Sarle (1994) montre d'ailleurs que de nombreux modèles statistiques peuvent être ré-écrits sous la forme d'un réseau de neurones. Quand les réseaux de neurones sont bien configurés et correctement calés, ce sont des outils de régression très puissants. Ils peuvent notamment être d'excellents interpolateurs capables de dégager les

4. Ce sont des structures artificielles. Ils ne sont pas directement un modèle d'un phénomène (sauf en sciences cognitives où ils miment de façon grossière les interactions entre neurones biologiques, d'où ils tirent leur nom et leur vocabulaire) mais plutôt un outil destiné à des tâches variées.

tendances dessinées par les données et non pas uniquement établir une très bonne corrélation entre les sorties et les exemples proposés lors du calage (à la différence d'interpolations polynomiales par exemple). Ils sont également employés pour la reconnaissance de « motifs » dans des données (mise en évidence d'une organisation implicite des données).

En hydrologie, de nombreux auteurs ont employé les réseaux de neurones pour réaliser des régressions non-linéaires. Certains les cataloguent même comme des méthodes opérationnelles qui ne sont pas destinées à améliorer nos connaissances hydrologiques (Abrahart *et al.*, 2007). Ils ont d'abord été utilisés pour modéliser la transformation Pluie – Débit, essentiellement en prévision (Campolo *et al.*, 1999, Coulibaly *et al.*, 1999), plus rarement en simulation, puis pour corriger les erreurs en sortie d'un modèle hydrologique (Anctil *et al.*, 2003). Zealand *et al.* (1999) leur ont demandé d'établir le débit hebdomadaire (contrôlé par les fontes nivales) en entrée d'un lac. Ils sont également employés (réseaux de Kohonen) pour identifier des motifs (*pattern*) de l'humidité du sol (Lauzon *et al.*, 2004). Les réseaux de neurones constituent aussi une méthode pour combiner les sorties de différents modèles hydrologiques (dans une approche multi-modèles) mais Shamseldin *et al.* (1997) ont trouvé que la combinaison des sorties de plusieurs modèles par un RNA n'est pas foncièrement meilleure qu'une moyenne pondérée de ces sorties, sur un échantillon de 11 bassins.

L'abondance de la littérature sur ce sujet permet de comparer les résultats. Ces derniers peuvent être très divergents d'une étude à l'autre. Si Goswami *et al.* (2005) trouvent que les RNA se situent parmi les meilleures solutions testées pour la prévision hydrologique ou en correction d'un modèle de prévision (en employant des données exogènes de pluies pour la correction), d'autres auteurs obtiennent des résultats inverses : Madsen *et al.* (2000) comparent différentes techniques de correction de l'erreur d'un modèle hydrologique dont les perceptrons[(g)](cf.infra), des modèles linéaires ou génétiques ; à longue échéance, ce sont les réseaux de neurones qui donnent les plus mauvais résultats lors de ce test. Xiong et O'Connor (2002) trouvent que la correction des sorties d'un modèle hydrologique par un perceptron ne donne pas de meilleurs résultats qu'un modèle auto-regressif simple. Enfin Anctil *et al.* (2003) discutent l'utilité du perceptron pour corriger les sorties d'un modèle conceptuel : le réseau de neurones semble vraiment utile quand les prévisions initiales sont de faible qualité.

Ces résultats ne sont cependant pas directement comparables entre eux. Il faut tenir compte du pas de temps et des horizons de prévision testés. Le tableau 10.2 donne ces informations pour les articles concernés. Les résultats restent cependant divergents, même à pas de temps, objectif du réseau et horizon donnés. Ainsi Madsen *et al.* (2000) et Toth et Brath (2007) ont des conclusions opposées quant à l'intérêt d'employer des perceptrons pour la correction d'un modèle conceptuel au pas de temps horaire à des horizons supérieurs à 6 heures. Toujours pour la correction, Xiong et O'Connor (2002) et Goswami *et al.* (2005) aboutissent à des résultats très différents pour le pas de temps journalier.

Article	Pas de temps	Horizons	Tests sur
Abrahart *et al.* (2007)	6 heures	6 et 24 heures	1 bassin versant
Anctil *et al.* (2003)	journalier	1 et 3 jours	1 bassin versant
Anctil *et al.* (2004a)	journalier	1 jour	2 bassins versants
Anctil *et al.* (2004b)	journalier	1 jour	1 bassin versant
Anctil et Rat (2005)	journalier	1 jour	47 bassins versants
Campolo *et al.* (1999)	horaire	1 à 10 heures	20 événements
Elshorbagy *et al.* (2000)	annuel	–	1 bassin versant
Firat (2008)	journalier	1 jour	2 bassins versants
Gaume et Gosset (2003)	journalier	Simulation	1 bassin versant
Goswami *et al.* (2005)	journalier	1 à 6 jours	1 bassin versant
Hsu *et al.* (1995)	journalier	1 jour	1 bassin versant
Jain et Kumar (2009)	6 heures et journalier	–	3 bassins versants
Kim et Kim (2008)	horaire	1 à 5 heures	1 bassin versant
Lekkas (2008)	horaire	10 heures	1 bassin versant
Lin et Chen (2008)	horaire	simulation ?	1 bassin versant
Madsen *et al.* (2000)	horaire	1 à 9 heures	1 bassin versant
Minns et Hall (1996)	horaire (?)	Simulation	Données synthétiques
Shamseldin (1997)	journalier	Simulation	6 bassins versants
Toth et Brath (2007)	horaire	1 à 12 heures	2 bassins versants
Xiong et O'Connor (2002)	journalier	1 à 3 jours	11 bassins versants
Zealand *et al.* (1999)	journalier	1 jour	1 bassin versant

TABLE 10.2 – Pas de temps et horizons des cas d'étude présentés dans les articles auxquels il est fait référence dans cette note. Seuls les RNA employés pour une modélisation Pluie – Débit sont présentés.

La diversité (pour ne pas dire les divergences) des résultats présentés dans la littérature provient sans doute du caractère « boîte noire » de la méthode et de son manque de robustesse : une régression aussi compliquée soit-elle peut très bien fonctionner sur un bassin versant et pas du tout pour un autre (or la plupart de ces études ont été menées sur un seul ou sur quelques bassins versants). Avec la question de la structure du réseau, celle de l'apprentissage[g]est primordiale : comme le font remarquer Anctil *et al.* (2004b), le perceptron utilisé pour la prévision hydrologique (transformation Pluie – Débit) a besoin d'une période d'apprentissage (calage) plus longue qu'un modèle conceptuel. Ceci est lié d'une part au fait que le réseau de neurones ne contient pas en lui-même de connaissance hydrologique (ainsi que le suggérent les tenants des modèles hydrologiques conceptuels ou des partisans des modèles dits « à base physique ») mais aussi et peut-être surtout au fait que les réseaux de neurones présentent un grand nombre de degrés de liberté (paramètres qu'il faut caler). Un apprentissage non adapté aux situations rencontrées en contrôle ou trop court conduit donc à des échecs, ce qui conduit Kirchner (2006) à les critiquer – avec les autres modèles « boîtes noires » – : *Such models are often good mathematical marionettes ; they often can dance to the tune of the calibration data. However, their predictive*

validity is often in doubt..

Il apparaît donc particulièrement nécessaire de tester ce type d'approche sur un grand nombre de cas (de bassins versants). Une telle étude a été menée par Anctil et Rat (2005) en prévision directe au pas de temps journalier : ils ont montré qu'il était possible d'obtenir des prévisions à 1 jour acceptables avec un réseau simple employé de façon automatisée. Nous allons la mener au pas de temps horaire pour la correction de l'erreur d'un modèle conceptuel global de prévision (puis en prévision directe, au chapitre 14).

10.5.2 Structure et configuration du réseau de neurones

La famille des réseaux de neurones est extrêmement vaste. Le modélisateur dispose de nombreuses possibilités pour construire le réseau de neurones adéquat afin de réaliser la tâche qu'il s'est assignée. Selon les choix qu'il fera, ce modélisateur pourra obtenir des résultats très différents (Maier et Dandy, 2000). Il importe donc de maîtriser le choix des principales caractéristiques et ne pas considérer les réseaux de neurones comme des outils « clefs en main » auxquels on soumettrait de façon automatique et aveugle n'importe quel problème, tendance que dénoncent Sarle (1994) ou Flood et Kartam (1994).

Les différents éléments permettant de configurer les réseaux de neurones sont détaillés en annexe O.5 (page 555). Nous nous concentrons ici sur les points mentionnés comme étant les plus importants dans la revue de la littérature (et notamment par le panorama dû à Maier et Dandy, 2000). Nous sommes loin d'avoir exploré tout le champ des possibles mais les choix documentés que nous avons faits, nous permettent de penser que les résultats que nous avons obtenus donnent une bonne indication des potentialités des réseaux de neurones pour le travail qui nous intéresse. Nous présentons ici nos principaux choix de configuration pour la tâche spécifique qu'est la correction des erreurs en sortie d'un modèle de prévision hydrologique. Le tableau 10.3 situé à la fin de cette partie résume nos choix.

Sélection de la structure du réseau

Nous avons opté pour des perceptrons multicouches, c'est-à-dire des réseaux organisés en couches de neurones sans boucle de rétroaction (Fig. 10.7) dont la fonction d'activation des neurones cachés est la composition semi-linéaire d'une intégration des entrées et d'une fonction logistique (l'une des fonctions d'activation les plus couramment employées). L'état X du réseau est donc relié à ses entrées E_i par :

$$X = f\left(\sum_{i=0}^{N_{\text{Entrées}}} w_{E_i \to X} \cdot E_i + b_X\right) \quad \text{avec} \quad f : x \mapsto \frac{1}{1+e^{-x}}$$

Ce choix est guidé par la démonstration que de tels réseaux de neurones dont les fonctions d'activation sont des fonctions bornées, continues par morceaux et non polynomiales constituent une classe d'approximateurs universels de l'ensemble des fonctions continues

par morceaux (Hornik *et al.*, 1989). Cela signifie qu'on peut trouver pour toute fonction continue par morceaux, un tel réseau de neurones l'approximant avec la précision voulue (Leshno *et al.*, 1993).

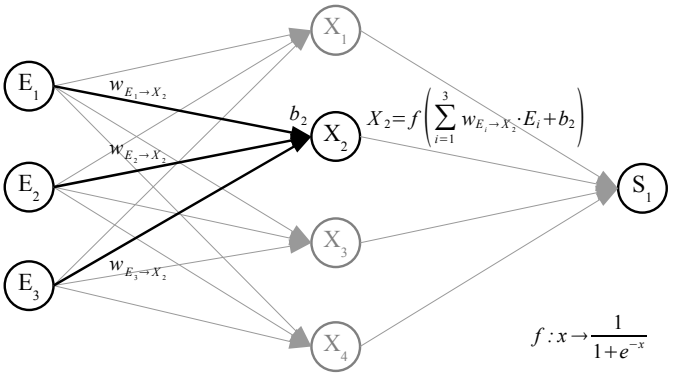


FIGURE 10.7 – Perceptron à trois entrées et une couche cachée comptant 4 neurones.

Nous nous limitons à une seule couche cachée car la tâche que nous réalisons ne devra de toute façon pas nécessiter un grand nombre de neurones cachés (ce nombre peut et doit rester de l'ordre du nombre de neurones de la couche d'entrée). Le nombre de neurones cachés sera étudié par essais-erreurs : les valeurs testées vont de 2 à 5.

Nous utilisons des réseaux complètement connectés : toutes les connexions possibles entre couches sont établies. La procédure de calage peut, le cas échéant, éliminer des connexions inutiles en leur assignant un poids très faible.

10

Choix des entrées

Le choix des données en entrée est crucial pour le bon fonctionnement du réseau. De nombreuses études (annexe O.5, page 555) montrent la grande sensibilité des performances des réseaux à leurs entrées.

Les entrées nous paraissant *a priori* pertinentes pour corriger les sorties du modèle hyrologique sont :

- les erreurs faites aux pas de temps précédents $\varepsilon_{t|t-1}$, $\varepsilon_{t-1|t-2}$, $\varepsilon_{t-2|t-3}$ $\cdots$
- le dernier débit observé Q_t,
- le cumul des pluies futures (de $t+1$ à $t+H$), susceptible d'expliquer les variations du débit qui ne sont pas encore inscrites dans les chroniques des débits observés Q_t et des erreurs de prévision précédentes,
- l'intensité maximale des pluies futures (entre $t+1$ et $t+H$).

Nous comparons ici plusieurs méthodes de choix des entrées. Nous procédons d'abord par essais-erreurs (sur l'échantillon des bassins versants). Puis, nous examinons si nous obtenons

des gains de performance additionnels en ne sélectionnant que les entrées jugées les plus pertinentes par une étude de sensibilité pour chaque bassin. La sensibilité de l'entrée E_i est définie à partir de la somme des dérivées partielles des sorties en cette entrée :

$$S_i = \sum_{t=1}^{N_P} \left(\frac{\partial \widehat{\varepsilon}_{t+H|t}}{\partial E_i(t)} \right)^2$$

Si la dérivée est très faible, la valeur de l'entrée correspondante modifiera peu la sortie et vice-versa. La procédure de calage consiste alors en :

1. le calage du réseau avec toutes les entrées considérées pertinentes *a priori*,
2. le calcul des sensibilités S_i sur la période de calage,
3. la sélection des entrées auxquelles le réseau est le plus sensible et qui expliquent 80, 90, 95 ou 99 % de la sensibilité totale,
4. le recalage avec ces seules entrées sur la même période.

Calage (*apprentissage* du réseau)

Le calage est réalisé par l'algorithme de Levenberg-Marquardt avec une régulation bayésienne pour limiter les risques de surcalage, l'un des grands écueils à éviter (rendu possible par le grand nombre de degrés de liberté au sein du réseau). En conséquence, nous ne pratiquons pas de validation croisée (une autre méthode classique employée pour éviter le surcalage). Le cadre théorique de cette méthode et l'algorithme sont détaillés dans l'annexe O.6. Nous prenons un nombre de périodes(g) égal à 1 : l'ensemble des exemples (des données employées en calage) est présenté au modèle pour le calcul de chaque modification du jeu de paramètres.

La fonction-objectif sera la somme des erreurs quadratiques car nous sommes intéressés comme précédemment par les plus grosses erreurs (situées lors des crues). De plus cette fonction-objectif se prête bien à la régulation bayésienne.

Le calage sera répété plusieurs fois (5 répétitions sont généralement suffisantes[5]) et le jeu de paramètres conduisant aux sorties les plus proches des médianes des sorties obtenues avec tous les jeux de paramètres sera sélectionné, ce qui est un autre facteur de lutte contre le surcalage.

Les capacités de calcul nécessaires au calage des réseaux de neurones nous incitent à ne travailler que sur les situations où la correction des sorties du modèle hydrologique conceptuel est importante : les hautes eaux (que nous définissons par rapport à un quantile de débit). Cela réduit le domaine à explorer, ce qui devrait faciliter la tâche des réseaux de neurones.

Comme nous ne réalisons pas de validation croisée, les séries de données sont scindées en

5. Julien Lerat, communication personnelle.

deux périodes : l'une sert au calage, l'autre au contrôle. Les données sont normalisées (par leur moyenne / variance).

Architecture	Perceptron multicouche.
Degré de connexion	maximal (entre couches).
Nombre de couches cachées	1.
Nombre de neurones cachés	faible et déterminé par essais-erreurs.
Choix des entrées	Essais-erreurs ou analyse de sensibilité.
Traitement des entrées	Normalisation (moyenne / variance).
Fonction d'activation	Sigmoïde (fonction logistique)
Calage	Algorithme de Levenbeg-Marquardt à régulation bayésienne (LMBR).
Nombre de répétitions du calage	5.
Nombre de périodes	1.
Nombre maximal d'itérations	30 par défaut.
Fonction-objectif	Somme des erreurs quadratiques.
Régulation	Somme des carrés des paramètres.

TABLE 10.3 – Principales options sélectionnées pour l'emploi de réseaux de neurones en correction des sorties d'un modèle hydrologique conceptuel.

10.5.3 Résultats

Gains en performances

Les performances obtenues avec les différentes configurations testées partagent les mêmes caractéristiques. Dans cette section, les résultats présentés correspondent à la meilleure configuration détaillée au cours des paragraphes suivants.

Côté pile : les réseaux de neurones obtiennent des performances très nettement supérieures aux autres méthodes de correction de l'erreur sur la très grande majorité des bassins (Fig. 10.8). De plus, cette méthode est la seule à rester efficace pour des horizons supérieurs à 12 heures : à titre d'exemple, pour un horizon de 48 heures, la correction par réseau de neurones améliore significativement les performances du modèle sur plus de 30 % des bassins, tandis qu'elle les dégrade significativement sur 15 % des bassins.

Côté face : la distribution des performance est beaucoup plus étalée. La correction par réseau de neurones s'avère catastrophique pour un petit nombre de bassins. Sur la plupart de ces bassins, le modèle initial obtenait des performances faibles (persistance légèrement positive, voire négative). Cependant, il ne s'agit pas d'une condition suffisante : sur certains bassins, la correction par RNA peut « sauver » le modèle obtenant des performances très faibles.

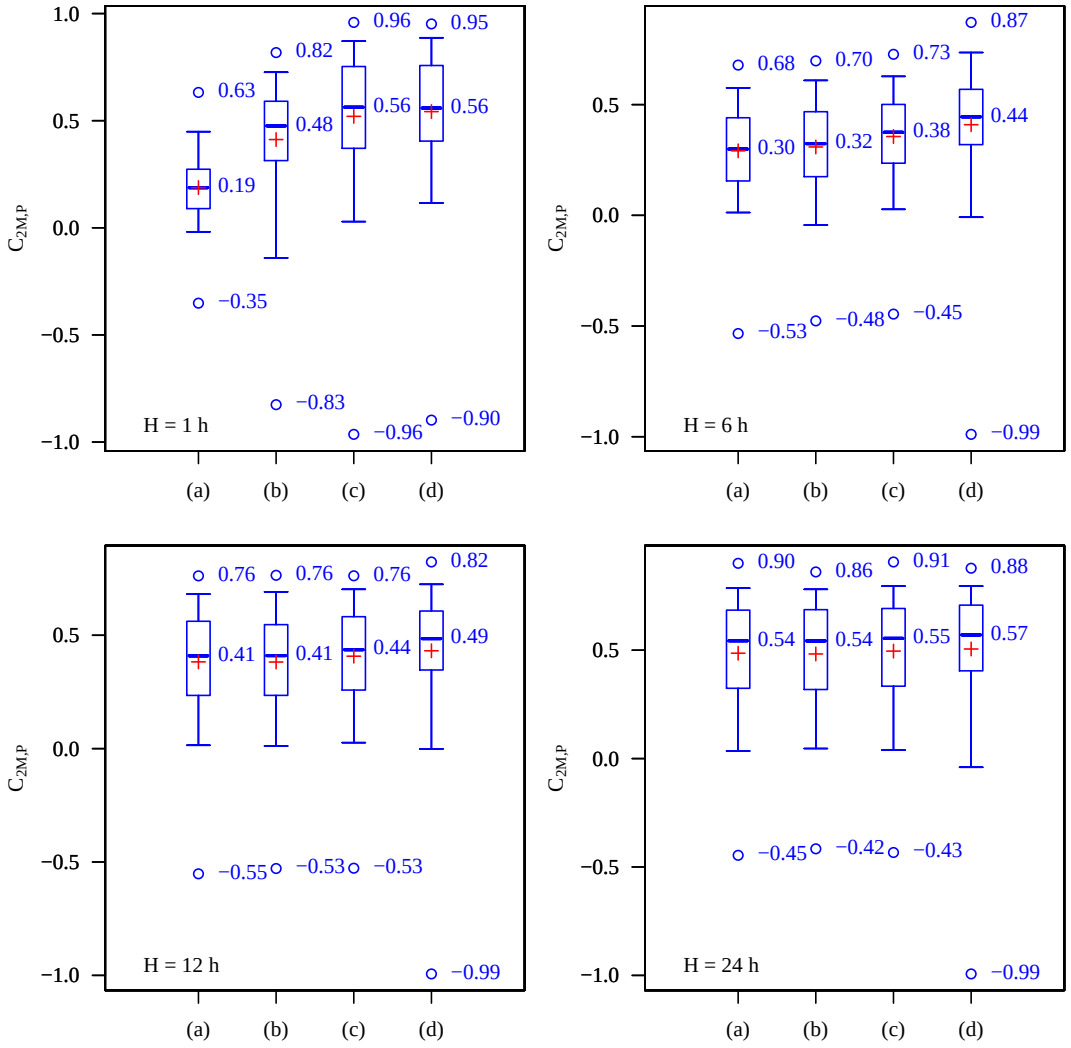


Figure 10.8 – Distributions des performances globales du modèle GRP mesurées par le critère $C_{2M,P}$ sans correction de l'erreur (a), avec la correction de Tangara (b), avec mise-à-jour des sorties par un modèle ARIMA (c) ou par réseau de neurones (d). Les comparaisons sont présentées pour les horizons de 1, 6, 12 et 24 heures. Tous les critères de performance sont calculés ici sur les périodes de hautes eaux sur lesquelles les performances des réseaux de neurones ont été estimées.

Plus le débit est autocorrélé, plus la différence entre les performances avec la correction par réseau des neurones et les performances sans correction est élevée. Ceci est également vrai quand on compare la correction par RNA et la correction de Tangara. Pour les horizons de prévision testés les plus élevés (de 24 à 48 heures), il n'y a plus de gain de performance pour le quart des bassins qui présentent les plus faibles valeurs d'autocorrélation du débit.

Choix des entrées par essais-erreur

Les performances de réseaux de neurones ayant comme entrées diverses combinaisons des variables sélectionnées (de 2 à 5 éléments) ont été comparées sur nos 178 bassins versants. La meilleure combinaison globale compte trois entrées : les deux dernières erreurs et le dernier débit observés. Rajouter d'autres entrées n'améliore plus les performances globales (sur l'échantillon de bassins versants).

La combinaison optimale des entrées varie d'un bassin à l'autre

Cependant, nous avons observé lors des essais-erreurs que la combinaison optimale des entrées pouvait varier d'un bassin à l'autre. Il est donc potentiellement intéressant d'employer une analyse de sensibilité pour déterminer les entrées les plus pertinentes pour chaque bassin.

Les performances obtenues quand on ne sélectionne comme entrées du réseau de neurones que les entrées auxquelles les sorties sont sensibles sont légèrement supérieures à celles obtenues avec l'ensemble des entrées : s'il existe des entrées inutiles, du moins perturbent-elles peu le réseau.

Aux horizons de prévision courts, l'entrée la plus souvent sélectionnée est la dernière erreur observée, viennent ensuite les erreurs précédentes ; la dernière observation de débit est rarement employée par le modèle. Cette situation très tranchée aux horizons de prévision courts disparaît peu à peu à mesure que l'horizon s'éloigne : la dernière observation de débit devient une information à laquelle la sortie est de plus en plus sensible et finit même par supplanter la dernière erreur observée. Les entrées apportant de l'information sur les pluies futures sont surtout sélectionnées pour les horizons lointains (elles le sont d'autant plus que le bassin réagit vite).

Sur les bassins les plus lents, l'information à laquelle le réseau est le plus sensible est la dernière erreur observée. Pour les bassins plus rapides, le réseau de neurones aura tendance à utiliser des combinaisons plus riches de variables d'entrée pour effectuer la correction aux horizons de prévisions un peu lointains.

Nombre de neurones cachés

Les meilleurs résultats sont obtenus pour deux neurones dans la couche cachée. Les autres variantes obtiennent des performances légèrement inférieures (elles restent très proches :

moins d'un point de $C_{2M,P}$ de différence).

Ce nombre de neurones cachés peut paraître étonnamment faible, surtout en comparaison d'études proposant des réseaux de plusieurs dizaines de neurones. Cependant, Gaume et Gosset (2003) ont obtenu un résultat très similaire sur deux cas d'études apparentés de prévision au pas de temps journalier : avec un réseau de neurones modélisant une relation débit – débit et avec un réseau modélisant une relation pluie – débit. Dans une étude comparant huit méthodes de lutte contre le surcalage, Giustolisi et Laucelli (2005) ont trouvé que le nombre optimal de neurones cachés (pour les deux techniques jouant sur ce paramètre) était également de cet ordre sur deux cas d'étude. La configuration optimale ainsi obtenue compte trois neurones d'entrée et deux neurones dans la couche cachée, soit 11 paramètres.

Peut-on simplifier la structure finale obtenue ?
11 paramètres, employés uniquement pour la correction de l'erreur, est un nombre élevé au regard d'études sur la complexité des modèles hydrologiques (e.g. Mein et Brown, 1978, Jakeman et Hornberger, 1993, Perrin *et al.*, 2001). Nous avons donc cherché à réduire le nombre de paramètres en examinant d'abord si certains étaient toujours calés à la même valeur : ce n'était pas le cas.

Nous avons ensuite recherché des relations simples (linéaires) entre ces paramètres par A.C.P.(g)(Ann. I). Pour l'horizon d'une heure, la somme cumulée des 6 plus grandes valeurs propres (sur 11) permet d'expliquer plus de 90 % de la variabilité des paramètres : on peut donc espérer réduire le nombre de paramètres calés sans trop de dégâts sur les performances à cet horizon. Il n'existe cependant pas de combinaison linéaire fortement explicative entre paramètres pour les autres horizons de prévision.

10.5.4 Cette méthode est-elle opérationnelle ?

À la vue des gains de performance mesurés par le critère $C_{2M,P}$, il peut paraître surprenant de poser la question de l'emploi des réseaux de neurones en conditions opérationnelles pour la correction des sorties du modèle hydrologique GRP. Au-delà de la méfiance que peut provoquer la nature terriblement « boîte noire » de ces outils, plusieurs éléments doivent encore être discutés pour évaluer la qualité opérationnelle de la méthode.

Conséquences de la nature événementielle de la correction
La correction d'erreur par réseaux de neurones est purement événementielle : bien que les entrées de débit puissent être proches d'un pas de temps à l'autre (et les entrées d'erreurs en partie communes), la correction peut varier grandement d'un pas de temps à l'autre, conduisant à un hydrogramme prévu quelque peu erratique autour de l'hydrogramme observé (cf. Fig. 10.9). Ceci peut devenir gênant sur certains événements. Ce défaut peut

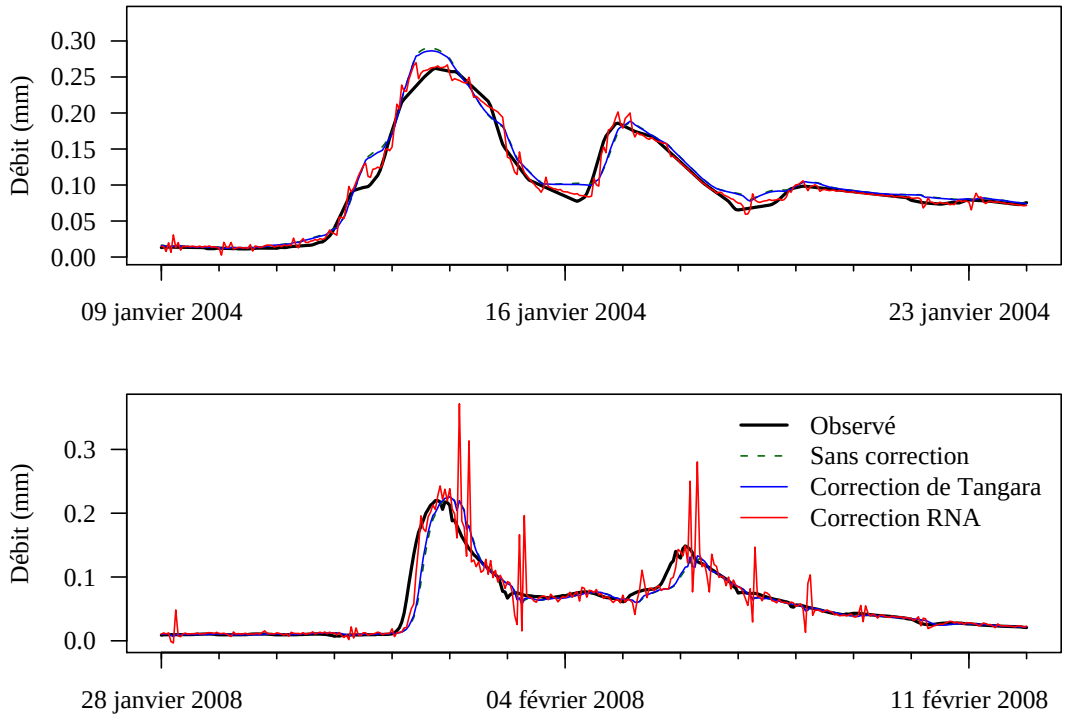


FIGURE 10.9 – Prévisions réalisées par le modèle GRP avec différentes mises-à-jour des sorties sur le bassin de la Bezonde à Pannes (339 km^2) lors de deux événements hivernaux. L'horizon de prévision est de 6 heures. On note que la correction de Tangara a un effet très limité ici. La correction par RNA est bien meilleure en montée mais oscille autour du débit qui sera réellement observé, voire annonce quelques prévisions facilement identifiables comme absurdes.

être interprété comme un signe du surcalage du réseau de neurones malgré la régulation bayésienne appliquée à l'algorithme de calage.

De plus, cette méthode de correction est réalisée pour un horizon donné, ce qui peut gêner l'établissement d'un *chevelu* de prévision (prévision de $t+1$ à $t+H$) présentant une cohérence interne : les corrections pour deux horizons intermédiaires H_1 et H_2 pouvant ne pas concorder.

Ces problèmes ne sont pas rédhibitoires et cette méthode de correction des sorties paraît utile pour les services opérationnels. Les sorties brutes du modèle hydrologique corrigées par RNA doivent cependant être expertisées par le prévisionniste.

Robustesse

Comme nous l'avons vu précédemment, la correction de neurones peut conduire à de clairs échecs pour quelques événements et quelques bassins. La probabilité de tels échecs est plus grande quand les performances avant correction des erreurs ne sont pas très bonnes mais

il est difficile d'annoncer *a priori* le risque d'échec. En outre, les performances au calage ne renseignent que très peu sur les performances au contrôle (ce qui constitue ici encore un indice de surcalage).

Capacité à extrapoler

Comme la grande majorité des approches basées sur les données (*data-driven approaches*), les réseaux de neurones peuvent conduire à des catastrophes en extrapolation. Il est donc important de leur fournir au calage des exemples couvrant le champ des possibles, ce qui est par nature impossible en hydrologie. Aussi pour une utilisation pratique, il serait nécessaire de mentionner la plage des « événements » explorés lors du calage. Mais les entrées étant de natures différentes (débits, erreurs de prévision...), les événements vus au calage doivent être caractérisés par plusieurs variables, ce qui rend la tâche ardue. Il est possible d'obtenir un bon fonctionnement de la méthode pour des événements d'intensité non présente dans les données de calage et d'obtenir des résultats désastreux pour un événement d'intensité moyenne mais présentant des erreurs du modèle hydrologique d'un niveau jamais rencontré par le modèle lors de son calage.
Enfin, il n'est possible de travailler que sur des phénomènes stationnaires.

Emploi opérationnel

Les réseaux de neurones constituent déjà un outil performant pour réduire les erreurs en sortie du modèle hydrologique de prévision et ont à ce titre un intérêt opérationnel. Cependant, leur robustesse n'est pas assurée et il est nécessaire d'employer en complément d'autres méthodes plus robustes mais aux potentiels plus modestes (comme les modèles ARIMA). Comme l'indiquaient Flood et Kartam (1994), ces outils ne peuvent pas être fournis « clefs en main » et le prévisionniste doit acquérir une certaine expertise lui permettant de déceler les sorties du modèle qui lui paraissent douteuses.

Peut-on ouvrir la boîte noire ?

Les RNA sont typiquement des boîtes noires. Cette caractéristique peut rendre délicate la prise en main de ces outils par les prévisionnistes et amoindrir la confiance qu'ils peuvent avoir en eux. Plusieurs auteurs ont présenté des approches permettant de donner une interprétation (partielle) du fonctionnement des RNA. Jain et Kumar (2009) cherchent par exemple à attribuer un régime de débit à chaque neurone caché. Une autre méthode consiste à proposer au RNA des entrées correspondant à des caractéristiques précises de l'hydrogramme, par exemple une décomposition du signal par ondelettes (Satyaji Rao et Krishna, 2009). À terme, de telles approches pourraient rendre l'emploi des RNA plus simples et leur fonctionnement interne plus compréhensible.

10.6 Bilan

L'erreur en sortie du modèle GRP (comme de très nombreux autres modèles hydrologiques) n'est pas du bruit blanc : il est donc judicieux de chercher à tirer parti de la structure statistique de l'erreur pour améliorer la qualité de nos prévisions. Néanmoins, la structure de l'erreur est un objet complexe et encore mal connu. Dans ce chapitre, nous avons proposé le test de plusieurs méthodes de correction de l'erreur.

Le débiaisage de l'erreur en sortie du modèle paraît très difficile à réaliser pour de multiples raisons : l'hétéroscédasticité des erreurs, la gestion de l'extrapolation... Les tests réalisés n'ont pas apporté de conclusions positives.

Plusieurs régressions classiques de complexité croissante ont été comparées. Toutes améliorent les performances du modèle significativement pour les horizons de prévision les plus courts mais se révèlent sans effet pour des horizons dépassant une douzaine de pas de temps. Parmi elles, la méthode la plus performante est celle des modèles ARIMA, outils qui sont en outre tout à fait robustes.

Enfin, les réseaux de neurones permettent des gains de performance remarquables sur la très grande majorité des bassins versants et restent effectifs à des horizons de prévision plus lointains. Malheureusement, ces outils, plus complexes, peuvent dans un petit nombre de cas, conduire à des échecs retentissants. D'autres travaux (notamment sur les méthodes de lutte contre le surcalage, Giustolisi et Laucelli, 2005) sont nécessaires pour les rendre plus robustes. Cette méthode n'est pas employable sans une certaine expertise de la part du prévisionniste.

Quatrième partie

Influence des mises-à-jour sur le fonctionnement du modèle

11

Chapitre

Initialisation du modèle GRP

Seigneur, tant de prudence entraîne trop de soin :
Je ne sais pas prévoir les malheurs de si loin.

Jean Racine, *Andromaque* (acte I. scène 4)

Objectifs

- ▷ Évaluer la durée de « chauffe » nécessaire au modèle GRP.
- ▷ Analyser l'influence des mises-à-jour du modèle sur son initialisation et période de chauffe.

Cet article propose une comparaison des fonctionnement événementiel et continu pour le modèle de prévision Pluie – Débit GRP dont un état est mis-à-jour. Il étudie l'influence de l'initialisation du niveau du réservoir de production (qui rend compte de l'humidité du bassin versant) sur les prévisions. Comme en simulation, l'approche continue permet d'obtenir de meilleurs résultats en prévision. Cependant, comme on pouvait s'y attendre, la mise-à-jour du réservoir de routage amoindrit considérablement la différence de performance entre les deux approches d'initialisation.

Hydrol. Earth Syst. Sci., 13, 819–831, 2009
www.hydrol-earth-syst-sci.net/13/819/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribution 3.0 License.

Hydrology and Earth System Sciences

How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments

L. Berthet[1,2], **V. Andréassian**[1], **C. Perrin**[1], **and P. Javelle**[3]

[1]Cemagref, Hydrosystems and Bioprocesses Research Unit Antony, France
[2]AgroParisTech ENGREF, 19 avenue du Maine, 75732 Paris, France
[3]Cemagref, Hydrology and Hydraulic Works Research Unit, Aix-en-Provence, France

Received: 17 February 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 5 March 2009
Revised: 27 May 2009 – Accepted: 2 June 2009 – Published: 18 June 2009

Abstract. This paper compares event-based and continuous hydrological modelling approaches for real-time forecasting of river flows. Both approaches are compared using a lumped hydrologic model (whose structure includes a soil moisture accounting (SMA) store and a routing store) on a data set of 178 French catchments. The main focus of this study was to investigate the actual impact of soil moisture initial conditions on the performance of flood forecasting models and the possible compensations with updating techniques. The rainfall-runoff model assimilation technique we used does not impact the SMA component of the model but only its routing part. Tests were made by running the SMA store continuously or on event basis, everything else being equal. The results show that the continuous approach remains the reference to ensure good forecasting performances. We show, however, that the possibility to assimilate the last observed flow considerably reduces the differences in performance. Last, we present a robust alternative to initialize the SMA store where continuous approaches are impossible because of data availability problems.

1 Introduction

1.1 Continuous vs. event-based approaches to modelling

From the catchment point of view, the hydrological cycle is a sequence of wetting and drying periods. On a given date, the moisture state of a catchment is the consequence of the past sequence of meteorological conditions. The initial moisture conditions at the beginning of a rainfall event have a major influence on a catchment's hydrological response. Therefore the set-up (as defined by Refsgaard and Henriksen, 2004) of a hydrological model requires choosing the initial conditions. Depending on how this is done, hydrological models will be categorized as *continuous* or *event-based*.

The initialization through a continuous approach consists in running the model during a *warm-up* period in order to let the model states reach values that no longer depend on arbitrarily chosen initial values. The duration of this warm-up depends on the catchment (its memory of past conditions) and on the model and may last a few months (Kitanidis and Bras, 1980b). A climatic cycle (i.e., one year) is often used, although it has been shown that some catchments (especially those where large aquifers feed streamflow) need up to several years (Le Moine, 2008). In an operational forecasting perspective, the major drawback of the continuous approach lies in its data requirements: long continuous precipitation time series up to the day of interest are difficult to provide (data gaps occur frequently because of real-time data repatriation difficulties). Another consequence is the necessity of having to gather a long enough data series before issuing the first forecast at new locations.

In contrast, event-based models require a separate method to derive the initial values of model states. Numerous methods exist. If the model states reliably represented measurable physical quantities, recent measurements or values based on climatology would be solutions. For example, Brocca et al. (2009) showed that assimilating soil moisture measurements into the event-based SCS-CN model can be useful for flow

Correspondence to: L. Berthet
(lionel.berthet@cemagref.fr)

simulation on a small catchment. However, these results should be generalized, as mentioned by the authors.

Continuous approaches have been recommended to modellers for many years (e.g. Kitanidis and Bras, 1980a; Linsley, 1982) as a rigorous solution to the estimation of initial conditions. However, we must acknowledge that event-based approaches are still often preferred in real-time operational applications (Lamb and Kay, 2004). Event-based models may be simpler because they often do not need to include all the processes necessary in a continuous model. This means more limited data requirements which may ease model implementation and use. Another reason lies in the difficulty maintaining and validating automatic measurement networks over a long period in many countries. This is a frequent situation when looking for high time resolution series. To bypass this obstacle, Nalbantis (1995) suggests relying on coarser data series (e.g. daily) to estimate fine (hourly) initial conditions. The problem may also be cultural. Some end-users, who traditionally use hydraulic propagation methods, are culturally in favor of an event-based approach. Despite all the good reasons advanced by hydrologists for using continuous approaches, practitioners often continue using event-based models and see them as the only solution. Another reason in favor of the event-based modelling is raised when the model is not used only for flood forecasting (e.g., for other uses like torrential flood modelling).

1.2 Sensitivity of hydrological models to the initialization procedure

The report of the National Research Council (NRC) (2002) identified as crucial the question of initial conditions. There is a wide consensus among hydrologists that hydrological models' outputs are very sensitive to initial conditions, especially soil moisture or catchment wetness (e.g. Refsgaard et al., 1999; Moore et al., 2006; Vivoni et al., 2007). Event-based models can lead to very different outputs when run with different initial conditions (Da Ros and Borga, 1997). As hydrological processes are essentially non linear, even a slight uncertainty on initial conditions can lead to dramatic uncertainty on streamflow (Zehe and Blöschl, 2004).

Many authors have studied the effects of initialization on the response of models that seek to reproduce physical processes. Already at the inception of the Soil Conservation Service (SCS) Curve Number (CN) formula, modulating the CN value according to the antecedent moisture conditions has been found to be necessary (Ogrosky and Mockus, 1964). More recently, Noto et al. (2008) showed that the degree of sensitivity to the initialization procedure depends on other factors, such as the intensity of precipitation or the catchment's physical properties. Vieux et al. (2004) demonstrated that the sensitivity of the model is lower when the catchment is already very wet.

1.3 The real-time forecasting specificities

The sensitivity of hydrological models to initial conditions is of prime importance for operational forecasting. For example, Norbiato et al. (2008) showed that initial conditions (antecedent soil moisture) are essential for efficient flash flood alerts.

Real-time forecasting systems most often use a data assimilation method to improve short-range prediction accuracy (Shamseldin, 2006). Among the different assimilation techniques, state updating is quite popular (Refsgaard, 1997; Moore, 2007). This method estimates state variables depending on the very last observed discharges. Consequently, the question of initial conditions appears to be less important if some (or even all) states are re-estimated by this updating technique (Nalbantis, 2000; Aubert et al., 2003; Moore et al., 2005).

State updating, when performed by assimilation of a small number of measured inputs (e.g. only discharge or discharge and soil moisture) compared to the number of internal states, leads to uncertainties which combine with the uncertainty on the initial values. Indeed, trying to update several model states simultaneously may endanger model robustness and it leaves the modeller in the uncomfortable situation where there are more unknowns than equations to solve.

The uncertainties due to the initial conditions may also be taken into account by using an ensemble forecast whose members differ in their initial conditions (e.g. Dietrich et al., 2008). However, this issue is not within the scope of this paper and we will focus on deterministic forecasting methods for the sake of simplicity.

1.4 Scope of the paper

This paper has both a theoretical and an applied objectives. The theoretical one is to contribute to a more general answer to the relative merits of continuous and event-based approaches for flood forecasting, through the comparison of different initialization approaches for the very same flood forecasting model. Indeed, although this issue has long been in the forefront, the literature does not provide any clear answer to this question. In addition, we investigate the possible interplay between the updating techniques and the initialization impact: the applied objective is to determine whether we can define simple initialization schemes which allow issuing forecasts without running the model over a long pre-forecast period. Initialization strategies are tested on a set of 178 French catchments.

Several authors compared different models with different initialization strategies. For example, Amengual et al. (2008) compared the performances of two different models – one being continuous, the other one event-based – to hindcast a flash flood event and found little difference between them. Instead, we choose to use the very same model in order to focus exclusively on the initialization strategies.

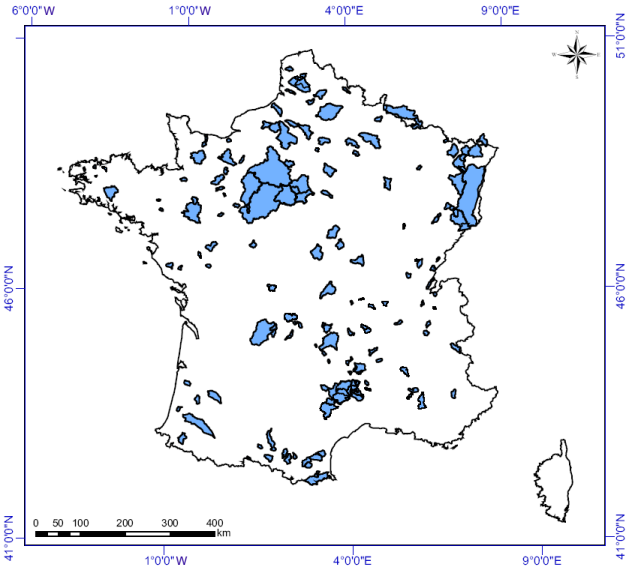


Fig. 1. Locations of the 178 French catchments used in this study.

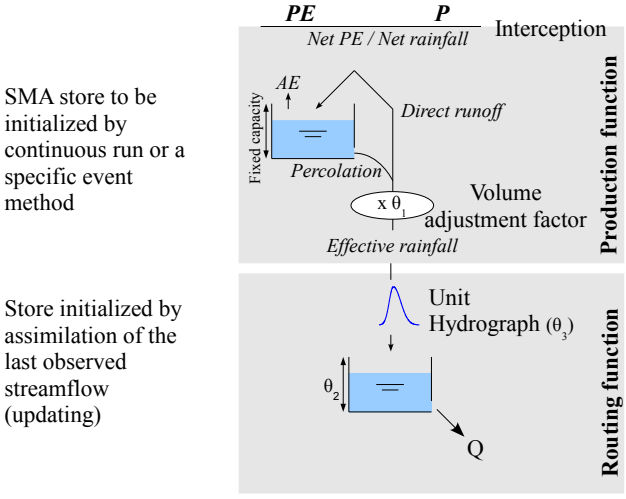


Fig. 2. Structure of the GRP model. The SMA store has to be initialized while the routing store level is updated when used in forecasting mode. *PE* and *AE* are potential and actual evapotranspiration. θ_i are model parameters: a volume adjustment factor, the capacity of the routing store and the base time of the unit hydrograph.

The remainder of the article is organized as follows: first the data and the model are described as well as the assessment criteria we employed. Then Sect. 3 details the methodology. The results are shown and discussed in Sect. 4. Finally a number of conclusions are drawn.

2 Catchments set, model and assessment criteria

2.1 Test set of 178 French catchments

The comparison is based on 178 French unregulated catchments (Fig. 1), chosen to represent the hydroclimatic variability encountered in the country (note, however, that no selected catchment is significantly snow-affected and requires a snowmelt module to ensure a proper modelling). Catchment areas range from 10 to 5940 km^2 (average of 354 km^2; median of 125 km^2). By working on various catchments, we aim to ensure more general conclusions to our study (Andréassian et al., 2006). Our data set covers a varied range of hydrological behaviours: some Mediterranean catchments experience flash floods, whereas others typically have slow floods. We used continuous hourly precipitation, discharge and potential evapotranspiration (PE) data series from 1995 to 2005. PE values were computed using the formula proposed by Oudin et al. (2005), based on temperature and extra-atmosphere global radiation (Morton, 1983). The quality of data used in this study was controlled and is supposed higher than the quality of data used in real-time operations.

2.2 Forecasting model

Our objective was to compare different initialization modes using the very same model structure. We deliberately used a simple model (GRP) in order to unambiguously analyze the effects of different initializations. It is an efficient operational model, one of those used to forecast river flows in real time on the Seine basin upstream from Paris (Cemagref, 2005). Detailing the structure of the forecasting model is not within the scope of this paper; therefore, only a brief description follows.

GRP is a hybrid metric-conceptual lumped parsimonious model, designed specifically for flood forecasting (see Tangara, 2005, for mathematical details). Its structure was derived from the structure of the GR4J model (Perrin et al., 2003). GRP can classically be described as a production function followed by a routing function (Fig. 2). The production function consists in a non-linear "soil moisture accounting" (SMA) reservoir and a volume adjustment coefficient which determine the runoff ratio. The SMA store requires either a specific initialization or a continuous running mode. The routing function is composed of a unit hydrograph (UH) and a non-linear routing store.

The forecasting model GRP uses a combination of two assimilation (updating) functions for flood forecasting. The first exploits the last observed discharge information to update the state of the model routing store, while the second draws information from the last model error to update the model's output through a multiplicative ARIMA model (Box and Jenkins, 1976). We do not use the Kalman filter (or one of its heirs) because we found it could lead to performance

losses during flood events when it assimilates streamflow alone (Aubert et al., 2003). The important thing to note here is that the level of the SMA store is not updated in the model and its initialization will be the main focus of the tests presented hereafter.

The model includes two main state variables: the levels of the production store and of the routing store. We do not consider the internal states of the unit hydrograph since they are very transient states: their values no longer depend on their initial values after a finite number of time steps because the model UH has a finite number of ordinates.

2.3 Assessment criteria

For a given lead time L, the overall evaluation of the forecasts is based on the persistence index PI (Kitanidis and Bras, 1980a):

$$PI(L) = 1 - \frac{\sum_t \left(\widehat{Q}_{t+L|t} - Q_{t+L}\right)^2}{\sum_t \left(Q_t - Q_{t+L}\right)^2} \quad (1)$$

where Q_t and Q_{t+L} are the observed discharge at time step t and $t+L$, respectively, while $\widehat{Q}_{t+L|t}$ is the forecast issued at time step t for time step $t+L$. A PI value of 1 indicates a perfect fit between forecasted and observed discharges. A positive value means that the root mean square error (RMSE) of the assessed model is lower than the RMSE of the persistence model which gives the last observed discharge as a prediction for the future time steps. A negative value implies that the model is less efficient than the persistence model. The criterion value mostly reflects performances during floods since it is a quadratic criterion. The PI is a well-suited quadratic criterion to assess forecasting models, since it compares the tested model to a naive one that uses the same information of observed discharge.

In addition, we used a time criterion to assess the time difference between the observed and forecasted flood events. We are interested in the time-to-peak delay (or the time to a fraction – say 90% – of the peak). Therefore we considered the mean time-to-peak delay for the identified flood events.

Last, the visual comparison of observed and forecasted hydrographs for significant events will complete our analysis.

3 Methodology

3.1 Modus operandi

Since only the effects of the initialization procedure are to be assessed, we used the same model and the same sets of parameters for the various initialization methods we tested. For each catchment, the model was calibrated by an automatic algorithm in a continuous mode (over a continuous series of 5 years). The PI was used as the objective function for parameter calibration. Even if the effects of the calibration approach (continuous or event-based) are not insignificant for the overall performance of the model (see e.g. Tan et al., 2008), we do not discuss this aspect here for the sake of brevity.

Flood forecasting requires future precipitation scenarios. In real-time conditions, some quantitative precipitation forecasts (QPF) may be available. In our study, we adopted a perfect foreknowledge scenario: this scenario corresponds to observed precipitations for the future time steps. While this is clearly not a realistic scenario in real-time conditions (it is overly optimistic), we selected this approach because we wish to focus the analysis on the effect of the initial conditions without adding other sources of uncertainty.

We used a classical split-sample test scheme (Klemeš, 1986) to assess the model's versions. The complete 10-year record available on every catchment was split into two 5-year sub-periods that alternatively served for calibration and validation. Only the results obtained on validation periods are shown.

Different catchments may have very different responses. To check whether different initialization solutions could fit different types of catchments, we divided the complete set into four subsets (of equal sizes) depending on streamflow autocorrelation, an index which provides information on the way catchments behave: catchments with flash floods have a low streamflow autocorrelation, while catchments with slow variations of streamflow present a much higher autocorrelation. For the same reason, forecasts are issued for different lead times: we assessed our results for 1-, 3-, 6-, 12-, 18-, 24-, 36- and 48-h lead times.

3.2 Tested continuous approach

We first tested the continuous approach: the model runs continuously for 1 year of warm-up (to obtain model states that are independent of the initial conditions) then continuously for the 4-year validation period. A previous analysis of our catchments (not shown here) demonstrated that a year of warming up is sufficient to reach states that no longer depend on the initial values for the studied catchments.

3.3 Tested event-based approaches

As the routing store of the selected model is updated using observed flow, the difference between the continuous and event-based approaches lies in the need to initialize the SMA store when working on an event basis. Three different simple event-based initializations are tested. For all of them, the performance criteria are computed over the same 4-year validation period so that they can be directly compared to the values obtained by the continuous approach.

3.3.1 Poor-man's initialization

The simplest initialization of the production store level is to choose an arbitrary value and then to run the model on a very short pre-forecast period (5 days) before issuing the forecast.

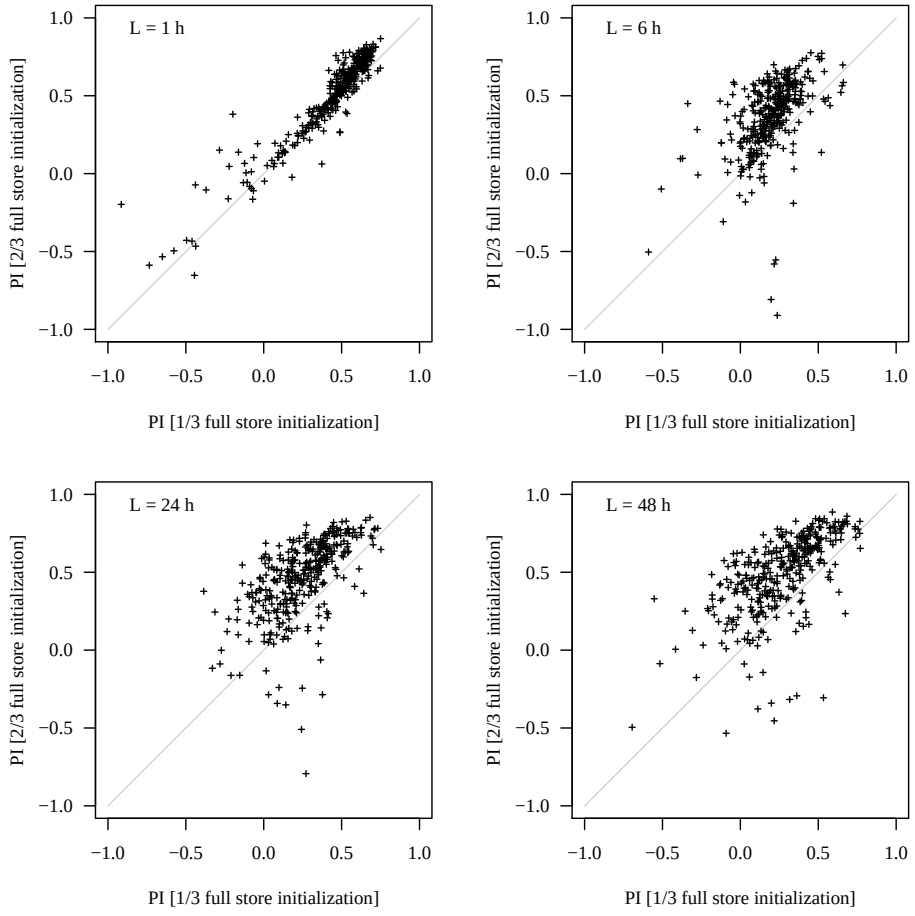


Fig. 3. Persistence indexes for two Poor-man's initializations at different initial values: the level of the SMA store is initially set at one-third or two-thirds of its capacity. Different initializations lead to very different performances.

Different values of the SMA store level (zero, one, two and three thirds of its capacity) are tested. This option makes it possible to check that the forecasts are indeed sensitive to the initialization of this store level.

3.3.2 Climatic initialization

The second tested approach consists in initializing the level of the production store at its pluri-annual average value (as calculated on the calibration period, i.e. over 5 years) at the beginning of a very short simulation period preceding the date of the forecast issue. Different simulation period lengths (from 5 to 15 days) are considered.

3.3.3 Antecedent moisture conditions initialization

The third initialization procedure is more elaborate. It looks in the calibration period archive for the time step that has the most similar antecedent precipitation index (API) value to the API value of the time step t at which we issue the forecast. The API (see e.g. Kohler and Linsley, 1951) is computed as follows:

$$\mathrm{API}(t) = \sum_{i=0}^{N_\alpha} \alpha^i P_{t-i} \tag{2}$$

where α is a decay rate, N_α is the number of antecedent time steps taken into account and P_{t-i} is the precipitation at time step $t-i$. N_α is chosen to ensure that $\alpha^i P_{t-i}$ would be negligible compared to any precipitation P_t for any i higher than N_α. Different values of α from $1-10^{-1}$ to $1-10^{-5}$ were tested. An α value of $1-10^{-3.5}$ was selected here as it gave the best results. Detailed results are not shown for the sake of simplicity.

4 Results and discussion

4.1 Results on the whole catchment set

First the Poor-man's initialization showed wide performance differences depending on the initial conditions for the tested

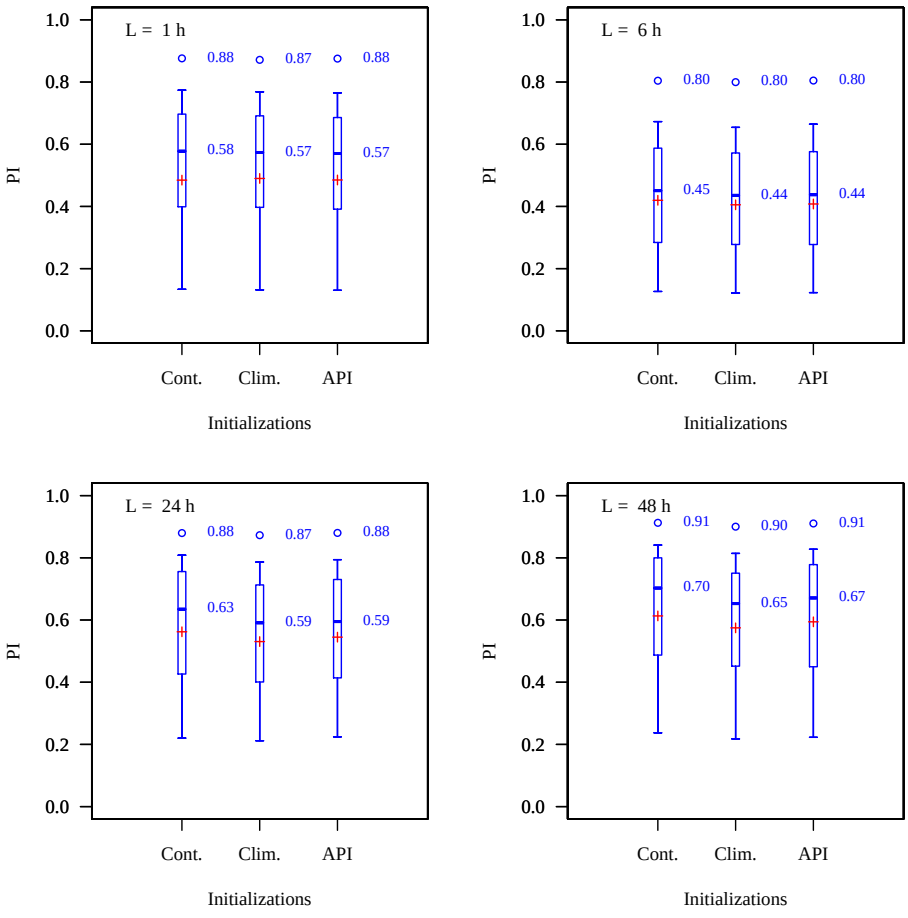


Fig. 4. Performance (persistence index) obtained by the model on 178 catchments, according to the initialization modes: the continuous strategy is depicted on the left and two event-based strategies (climatic and API strategies) are on the depicted right. The results are displayed for lead times ranging from 1 to 48 h. The pre-forecast period for event-based strategies lasts 120 h (5 days). Boxplots give minimum and maximum values (dots) as well as 0.05 and 0.95 quantiles (whiskers), 0.25, 0.50, 0.75 quantiles (boxes). Mean values are indicated by crosses.

model and our catchments (Fig. 3). Thus, at the 1-h lead time, the persistence differences for different (arbitrary) initial values are greater than 0.03 (which is a significant difference) on more than 75% of the catchments; for the 48-h lead time, this difference is greater than 0.14 for more than 90% of the catchments.

The results clearly show that the continuous approach gives the best results (see Table 1 and Figs. 4 and 5), and that the longer the lead time, the greater the difference in performance. Our interpretation is the following: model states do not reflect reality directly but are distorted representations of the real world as seen by the model. It is a better choice to initialize the model states as if they were seen from a given reality (continuous approaches) by the model rather than to impose values derived from measurements that do not correspond to the inner model logic. These results can be compared to those presented by Anctil et al. (2004a) and Merz and Blöschl (2009): antecedent soil moisture is a better control on the runoff coefficient ratio than antecedent precipitation depth and the state of a conceptual rainfall-runoff model can give valuable information on catchment moisture state.

The model has small time-to-peak errors: even for a 48-h lead time, the time-to-peak errors of the model on more than 90% of the catchments are smaller than 5 h. No significant difference in time-to-peak errors can be noted, whatever initialization method is considered (Table 2). Event-based initialization can even lead to very slightly smaller time-to-peak error than the continuous approach. Indeed, the initialization of the SMA store has a much more important influence on the simulated flood volume than on its timing. This can be seen on an example (Tarn River at Millau, spring 2004 floods, Fig. 6c): the different initializations lead to very different discharge magnitudes, but they all have the same temporal behaviour.

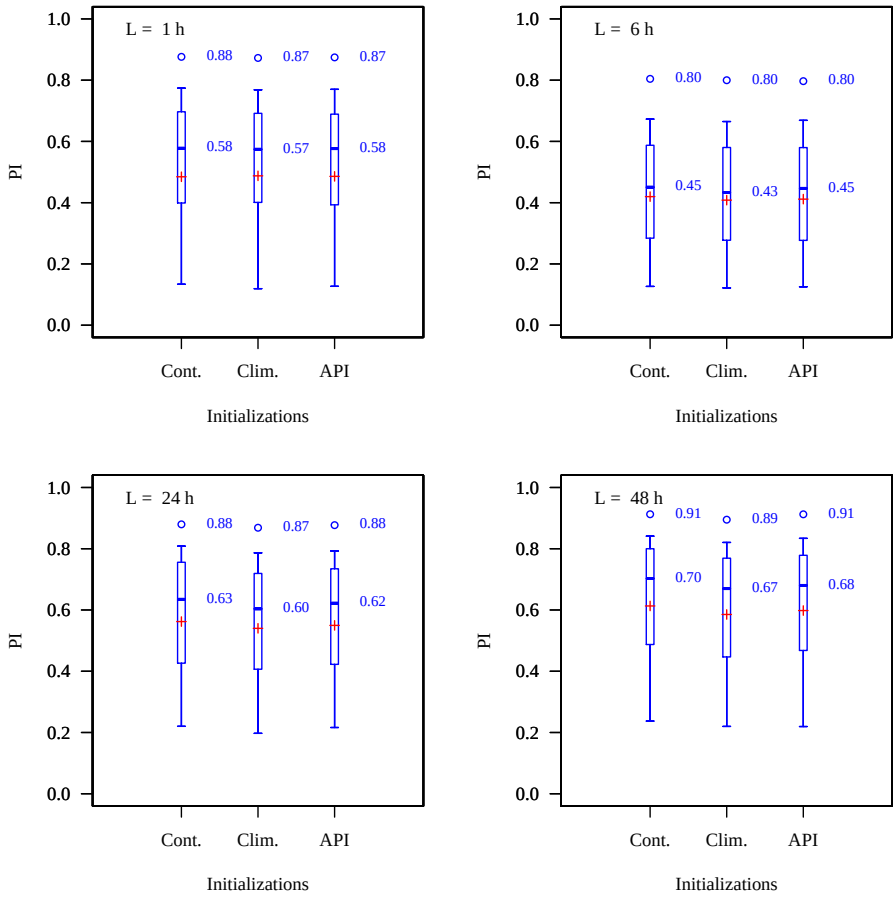


Fig. 5. Performance (persistence index) obtained by the model on 178 catchments, according to the initialization modes: continuous or event-based (climatic and API) strategies with a pre-forecast period lasting 360 h. Results are displayed for lead times ranging from 1 to 48 h.

Table 1. Medians of persistence index values obtained using the GRP model on the 178 catchment set with different initialization approaches for 1-, 6-, 24- and 48-h lead times.

Initialization	1-h lead time	6-h lead time	24-h lead time	48-h lead time
Continuous	*0.58*	*0.45*	*0.63*	*0.70*
Best Poor man's	0.56	0.40	0.50	0.64
Best climatic	0.57	0.44	0.61	0.67
Best API-based	0.58	0.45	0.62	0.68

As expected, the event-based initialization strategies lead to poorer forecasting performances. However the good news (from an operational point of view) is that the performance loss due to the use of a simple event-based initialization strategy is not large for most catchments (see Figs. 4 and 5). This result must be moderated by the fact that we used controlled *offline* data which are supposed to be high quality compared to the data used in real-time operations. Event-based initialization can not always be done with sufficiently controlled data and therefore might encounter more errors in rainfall while the continuous approach may do a better use of controlled data.

The event-based initialization strategies we tested ranked in a quite logical manner: the best one is not surprisingly the API method, which is the most informative approach concerning the catchment initial moisture conditions. Then comes the climatic solution, which provides little information. The strategy that leads to the lowest performances is the Poor-man's approach.

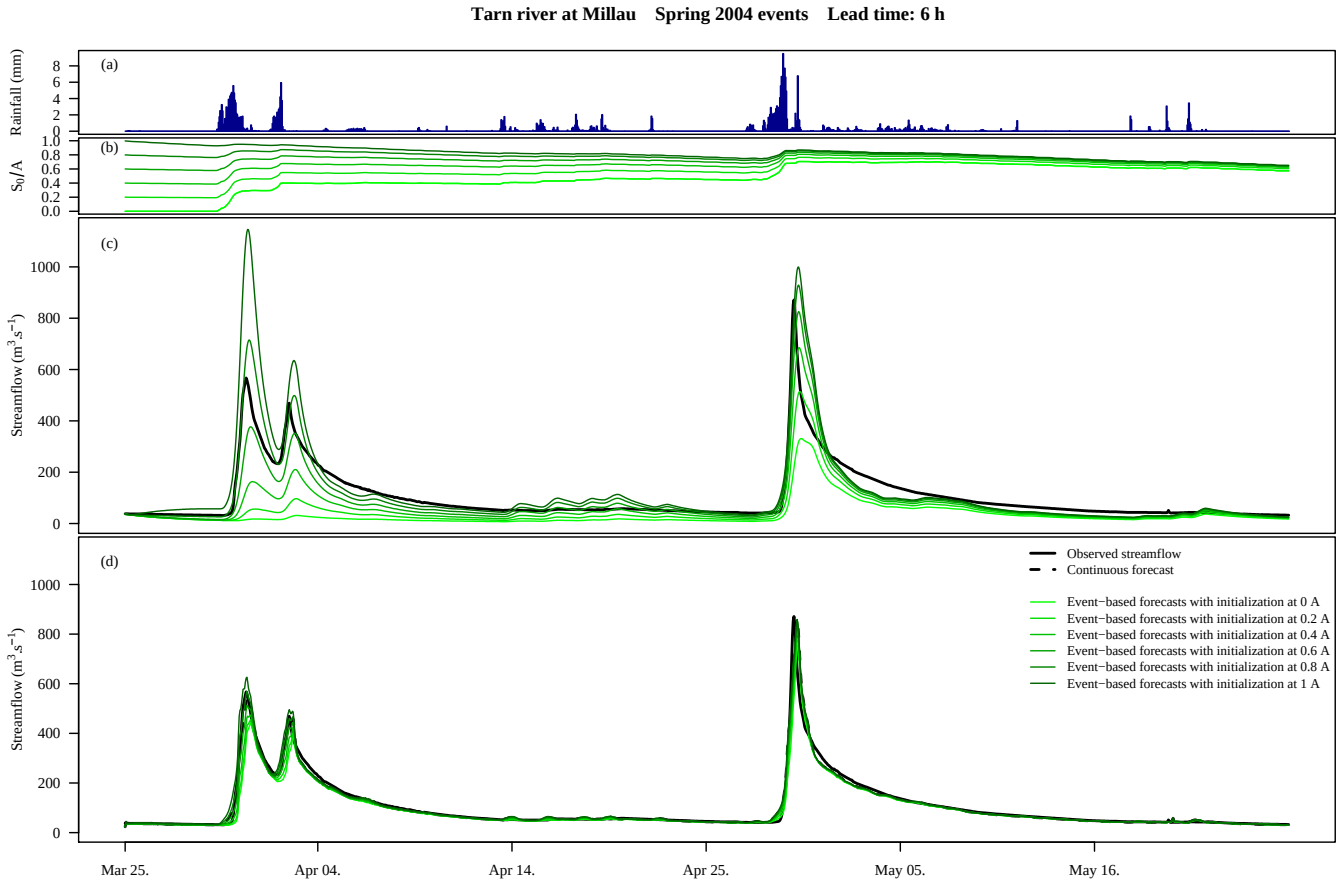


Fig. 6. **(a)** Precipitation during spring 2004 events. **(b)** Changes in the production store level after its initialization at different values on the 25 March 2004: there is convergence mainly during showers. Different initializations lead to very different forecasts when no updating technique is applied **(c)**, whereas forecasts depend much less on the initial production store content when the model is updated **(d)**.

Table 2. Mean time-to-peak errors obtained with the different initialization methods (expressed in hours)

Initialization	1-h lead time	6-h lead time	24-h lead time	48-h lead time
Continuous	*0.23*	*1.39*	*0.69*	*−0.09*
Best Poor man's	0.25	1.42	0.76	−0.05
Best climatic	0.24	1.43	0.73	−0.09
Best API-based	0.24	1.45	0.71	−0.01

Those results are indeed model-dependent: they depend actually on the model structure used and on the choice of the updating methods. However, Moore (2007) pointed out that many models present strong structural and behavioural similarities; consequently the conclusions may remain valid for many conceptual models. Conversely, the choice of the updating method is crucial: updating the SMA store leads to very poor performance with the same model structure (Javelle and Berthet, 2008; Tangara, 2005).

4.2 Use of a pre-forecast period: a compromise approach?

Actually, many so called event-based approaches are not purely event-based since they use a short pre-forecast period on which the model is run before issuing the forecast: the model is initialized at time step $t-\tau$ and run during the pre-forecast period from $t-\tau$ to t before being used to issue a forecast at t of the runoff values at $t+1, \ldots, t+L$. Many event-based approaches consider initial conditions to

Table 3. Medians of persistence index values obtained by the selected model with the climatic initialization approach using pre-forecast period of different lengths for 1-, 6-, 24- and 48-h lead times. HU length ranges from 1 to 64 h depending on the catchment.

Pre-forecast length	1-h lead time	6-h lead time	24-h lead time	48-h lead time
HU length	0.57	0.42	0.57	0.64
5 days	0.57	0.44	0.59	0.65
10 days	0.57	0.44	0.61	0.66
15 days	0.57	0.43	0.60	0.67
Continuous	*0.58*	*0.45*	*0.63*	*0.70*

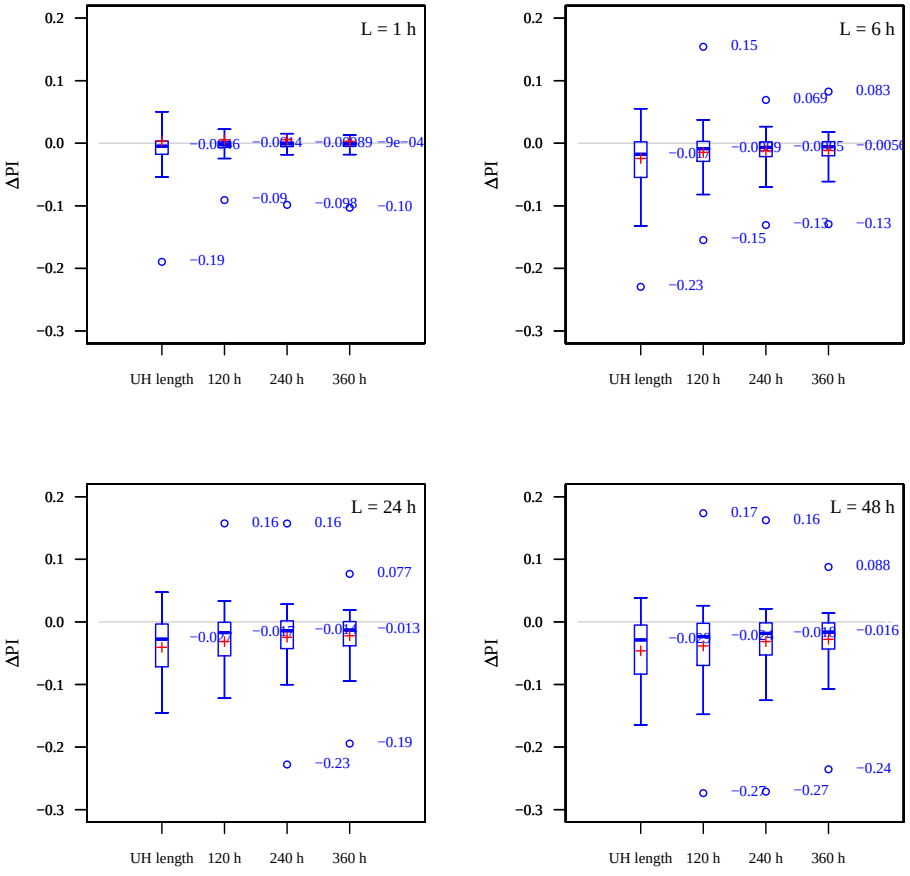


Fig. 7. Differences between the performance (persistence index) obtained by the model on 178 catchments in continuous mode and using an event-based climatic initialization, according to the pre-forecast period length: the longer the pre-forecast period, the smaller the differences.

be parameters; this requires pre-forecast period data to calibrate the initial conditions. These models can not really be considered as purely event-based. Thus, some modellers, e.g. Merz and Bárdossy (1998) and Sheikh et al. (2009), chose to use an inter-event model to initialize the most sensitive states. Another example is given by Anctil et al. (2004a), who used artificial neural networks (ANN) for flow forecasting on two catchments: they showed that a long-term soil moisture index derived from a continuous model is a valuable input which improves forecasts. Here again the resulting models are not purely event-based: they belong to a continuum between event-based and continuous models.

In this paper, our initialization strategies also use a short pre-forecast period to come closer to a continuous approach: the resulting initial conditions are a mix of assumptions implied by the initial choice and of the model's internal representation of the catchment behaviour.

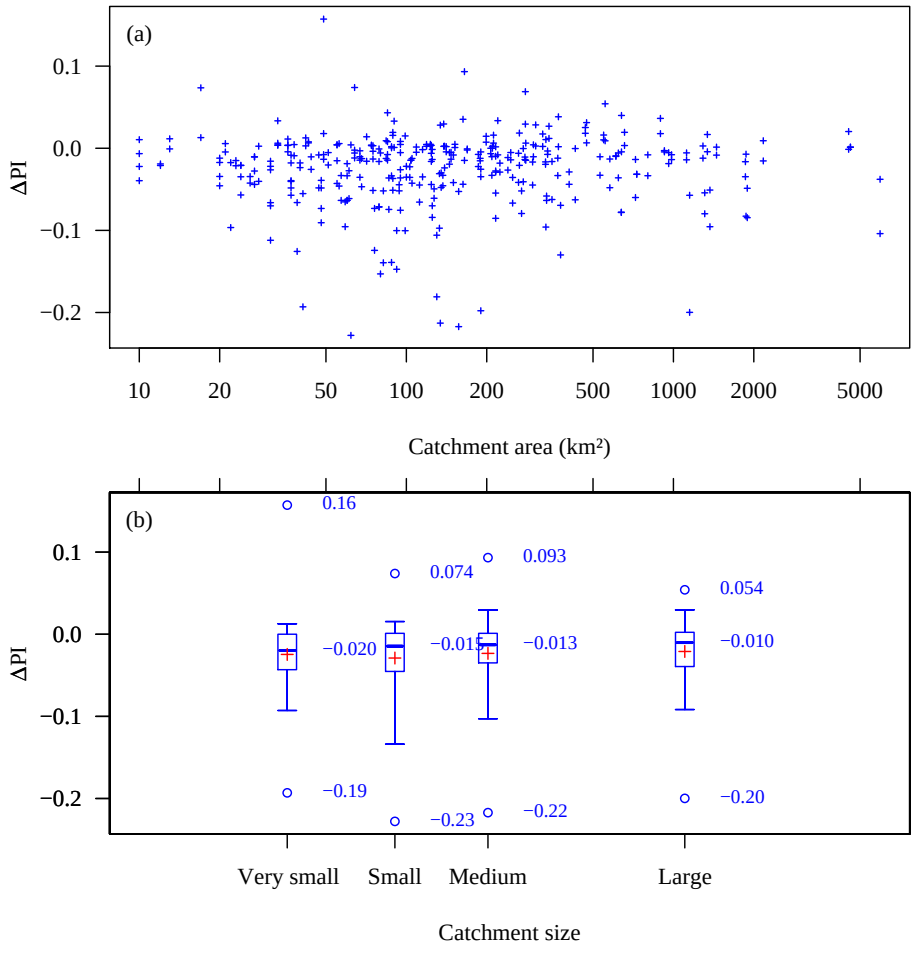


Fig. 8. Differences between the performance (persistence index) obtained by the model on 178 catchments in continuous mode and using an event-based climatic initialization approach, depending on the catchments' areas **(a)**. Here lead time is 6 h. The pre-forecast period for the event-based strategy lasts 240 h. Four classes of catchments are defined: the first class groups the 25% smallest catchments, the second class, the following 25%, etc. Distributions of the difference in performance for every class are displayed **(b)**.

In the strategies we tested, this pre-forecast period lasts at least the length of the UH (to obtain proper values in the UH) and was tested up to 15 days. It is clear for all event-based initializations that the longer the pre-forecast period, the better the performance (Fig. 7). For the climatic initialization, a pre-forecast period of 5 days leads to performance significantly lower than the performance obtained with a continuous initialization. However, a 15-day pre-forecast period allows performances close to what is given by the continuous approach (see Table 3).

4.3 Do results depend on catchment size?

Figure 8 shows the difference in performances obtained by the same model running in continuous mode and in event-based mode (with climatic initialization) depending on the catchment area. No clear trend can be detected from these analyses. It is interesting to note moreover that the model's performance does not decrease as the catchment size increases: the catchment behaves as a low-pass filter (Oudin et al., 2005) and as the catchment size increases, forecasting in fact becomes an easier task (for a given lead time).

4.4 Do results depend on catchment reactivity?

We found no relationship between catchment response time and the impact of choosing a continuous or an event-based strategy. Figure 9 shows the difference in performances obtained by the same model running in continuous mode and in event-based mode (with climatic initialization) depending on the catchments discharge autocorrelation. No trend was detected from these analyses.

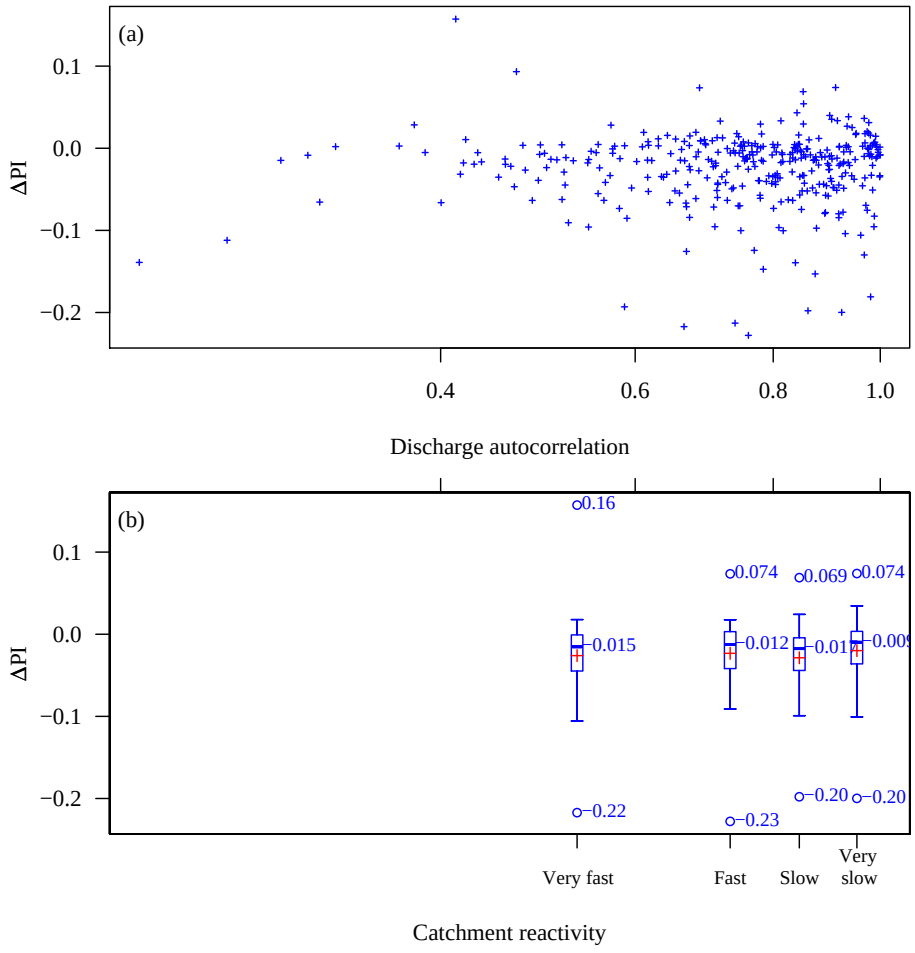


Fig. 9. Differences between the performance (persistence index) obtained by the model on 178 catchments in continuous mode and using an event-based climatic initialization approach, depending on the discharge autocorrelation **(a)**. Here the lead time is 6 h. The pre-forecast period for the event-based strategy lasts 240 h. Four classes of catchments are defined: the first class groups the 25% of catchments with the lowest discharge autocorrelation, the second class, the following 25%, etc. Distributions of the difference in performance for every class are displayed **(b)**.

4.5 Impacts of the updating procedure

When used to issue discharge forecasts, hydrological models are most often updated (Refsgaard, 1997). In practice, this means that the discharge forecast no longer depends on forcing variables only (e.g. precipitation, evapotranspiration, etc.) but also on the information contributed by the data assimilation process. The discharge forecast is constrained by data assimilation and consequently it may depend (much) less on the internal states and so on their initialization.

We used the GRP model with and without updating techniques to compare the influence of initialization of the model's SMA store (which is never updated) on forecasts in both cases. Figure 6 shows an example of spring floods for the Tarn River at Millau (2170 km^2). We chose different initial production store levels from 0 (empty store) to its maximum capacity A (full store). From those initial values, the production store levels converge slowly; convergence is mostly achieved during showers (Fig. 6a, b). The differences in production store levels lead to dramatic differences in discharge forecasts when no updating technique is applied (Fig. 6c), whereas the 6-h forecasts are much more constrained with data assimilation (Fig. 6d). Similar results were obtained for the other catchments of our data set.

Thus, the updating procedure used in forecasting models does limit the impact of crude initialization procedures, in comparison with simulation models: this explains why a rather simple procedure with a short pre-forecast period gives results close to those obtained when using a continuous initialization.

5 Synthesis and conclusions

Initial conditions are known to be of crucial importance for hydrological models. In this paper, we compared different initialization strategies of the soil moisture component of a rainfall-runoff forecasting model. The continuous mode was compared to several event-based approaches for the same model. The main conclusion is that the best results were obtained when the model was run in a continuous mode. This corroborates the results of previous studies (e.g., Anctil et al., 2004b). However, we showed that one of our tested event-based initialization strategies (i.e., the API-based initialization) could lead to performances rather close to what is obtained with the continuous approach, provided that the model can be run on a short pre-forecast period.

Indeed, the sensitivity of the model outputs to the initial conditions is much lower in *forecasting mode* than in *simulation mode* (i.e. without updating through the assimilation of measured streamflow): the output is considerably constrained by the information contributed during the observed flow assimilation process, which partly compensates for the errors in initial values.

Given the large and varied data set used here, we believe that these results are not catchment-dependent (in particular we found no relation to catchment size or reactivity). The results may remain to some extent model-dependent. However, we expect that the behaviours we observed can also be found for many forecasting models, since they all have to use efficient data assimilation.

The loss in performance when running the model using event-based strategies is not substantial: indeed, in most cases, the difference is not really significant. This means that if an efficient assimilation of the last observed streamflow is possible, event-based strategies can be efficiently used for operational purposes when and where it is impossible to run a model continuously.

Acknowledgements. This work is partly funded by a research grant from the French Ministry of Agriculture. The authors gratefully acknowledge the assistance of Nicolas Le Moine (Cemagref) for data collection and Météo France for providing meteorological data as well as French and Quebec operational Flood Forecasting Services for fruitful discussions.The authors also thank Uwe Ehret and Julien Lerat and two anonymous reviewers for their comments which led to improvements in the article.

Edited by: N. Verhoest

References

Amengual, A., Diomede, T., Marsigli, C., Martín, A., Morgillo, A., Romero, R., Papetti, P., and Alonso, S.: A hydrometeorological model intercomparison as a tool to quantify the forecast uncertainty in a medium size basin, Nat. Hazards Earth Syst. Sci., 8, 819–838, 2008, http://www.nat-hazards-earth-syst-sci.net/8/819/2008/.

Anctil, F., Michel, C., Perrin, C., and Andréassian, V.: A soil moisture index as an auxiliary ANN input for stream flow forecasting, J. Hydrol., 286, 155–167, 2004a.

Anctil, F., Perrin, C., and Andréassian, V.: Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models, Environmental Modelling & Software, 19, 357–368, 2004b.

Andréassian, V., Hall, A., Chahinian, N., and Schaake, J.: Introduction and Synthesis: Why should hydrologists work on a large number of basin data sets?, IAHS-AISH Publication, 307, 1–5, 2006.

Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall - Runoff model, J. Hydrol., 280, 145–161, 2003.

Box, G. E. P. and Jenkins, G. M.: Time Series Analysis: Forecasting and Control, Holden Day Inc., Oakland, California, USA, 575 pp., 1976.

Brocca, L., Melone, F., Moramarco, T., and Singh, V. P.: Assimilation of Observed Soil Moisture Data in Storm Rainfall-Runoff Modeling, J. Hydrol. Eng., 14, 153–165, doi:10.1061/(ASCE) 1084-0699(2009)14:2(153), 2009.

Cemagref: Inventory and diagnosis of simple existing flood forecasting models on the Seine River basin (in French), Final report 133 pages, Hydrosystems and Bioprocesses Research Unit, DIREN Île-de-France, Antony, France, 2005.

Da Ros, D. and Borga, M.: Adaptive use of a conceptual model for real time flood forecasting, Nord. Hydrol., 28, 169–188, 1997.

Dietrich, J., Trepte, S., Wang, Y., Schumann, A. H., Voß, F., Hesser, F. B., and Denhard, M.: Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event, Nonlinear Proc. Geophys., 15, 275–286, 2008.

Javelle, P. and Berthet, L.: Inter-comparaison des modèles de prévision des crues développés au Cemagref: GR3H et GR3P, convention 2008 Cemagref / Ministère de l'Écologie et du Développement durable – service central d'hydrométéorologie et d'appui la prévision des inondations, Rapport Cemagref, 23 pp. (in French), 2008.

Kitanidis, P. and Bras, R.: Real-time forecasting wih a conceptual hydrologic model. 1. Analysis of uncertainty, Water Resour. Res., 16, 1025–1033, 1980a.

Kitanidis, P. and Bras, R.: Real-time forecasting with a conceptual hydrologic model. 2. Applications and results., Water Resour. Res., 16, 1034–1044, 1980b.

Klemeš, V.: Operational testing of hydrologic simulation models, Hydrol. Sci. J., 31, 13–24, 1986.

Kohler, M. A. and Linsley, R. K. J.: Predicting runoff from storm rainfall, Res. Paper, 34, US Weather Bureau, Washington DC, USA, 1951.

Lamb, R. and Kay, A.: Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain, Water Resour. Res., 40, W07501,

doi:10.1029/2003WR002428, 2004.

Le Moine, N.: Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles Pluie – Débit?, Ph.D. thesis, Université Pierre et Marie Curie (Paris VI), 2008.

Linsley, R.: Proceedings of the international symposium on rainfall-runoff modelling, in: Proceedings of the international symposium on rainfall-runoff modelling, edited by: Singh, V., 3–22, Water Resources Publications, Littleton, CO, USA, 1982.

Merz, B. and Bárdossy, A.: Effects of spatial variability on the rainfall runoff process in a small loess catchment, J. Hydrol., 212–213, 304–317, 1998.

Merz, R. and Blöschl, G.: A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria, Water Resour. Res., 45, W01405, doi:10.1029/2008WR007163, 2009.

Moore, R., Cole, S., Bell, V., and Jones, D.: Issues in flood forecasting: Ungauged basins, extreme floods and uncertainty, in: IAHS-AISH Publication no. 305, 103–122, 2006.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483–499, 2007, http://www.hydrol-earth-syst-sci.net/11/483/2007/.

Moore, R. J., Bell, V. A., and Jones, D. A.: Forecasting for flood warning, Comptes Rendus Geosciences, 337, 203–217, http://www.sciencedirect.com/science/article/B6X1D-4F0199D-6/2/c3a587ebec49a67b3ff4217a8dfab536, 2005.

Morton, F.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, J. Hydrol., 66, 1–76, 1983.

Nalbantis, I.: Use of multiple-time-step information in rainfall-runoff modeling, J. Hydrol., 165, 135–159, 1995.

Nalbantis, I.: Real-time flood forecasting with the use of inadequate data, Hydrol. Sci. J., 45, 269–284, 2000.

National Research Council (NRC): Report of a Workshop on Predictability & Limits-To-Prediction in Hydrologic Systems, 0-309-08347-8, National Academic Press, Washington DC, USA, 118 pp., 2002.

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., and Anquetin, S.: Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, J. Hydrol., 362, 274–290, 2008.

Noto, L., Ivanov, V., Bras, R., and Vivoni, E.: Effects of initialization on response of a fully-distributed hydrologic model, J. Hydrol., 352, 107–125, 2008.

Ogrosky, H. O. and Mockus, V.: in Handbook of Applied Hydrology, chap. Hydrology of Agricultural Lands, 21:1–21:97, McGraw-Hill inc., USA, 1964.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, J. Hydrol., 303, 290–306, 2005.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, 2003.

Refsgaard, J. C.: Validation and Intercomparison of Different Updating Procedures for Real-Time Forecasting, Nord. Hydrol., 28, 65–84, 1997.

Refsgaard, J. C. and Henriksen, H. J.: Modelling guidelines – terminology and guiding principles, Adv. Water Resour., 27, 71–82, 2004.

Refsgaard, J. C., Thorsen, M., Jensen, J., Kleeschulte, S., and Hansen, S.: Large scale modelling of groundwater contamination from nitrate leaching, J. Hydrol., 221, 117–140, 1999.

Shamseldin, A. Y.: River Basin Modelling for Flood Risk Mitigation, chap. Real-time river flow forecasting, Taylor & Francis/Balkema, Leiden, the Netherlands, 181–195, 2006.

Sheikh, V., Visser, S., and Stroosnijder, L.: A simple model to predict soil moisture: Bridging Event and Continuous Hydrological (BEACH) modelling, Environ. Model. Softw., 24, 542–556, 2009.

Tan, S., Chua, L., Shuy, E., Lo, E.-M., and Lim, L.: Performances of rainfall-runoff models calibrated over single and continuous storm flow events, J. Hydrol. Eng., 13, 597–607, 2008.

Tangara, M.: Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global, Ph.D. thesis, École pratique des hautes études de Paris, 2005.

Vieux, B., Cui, Z., and Gaur, A.: Evaluation of a physics-based distributed hydrologic model for flood forecasting, J. Hydrol., 298, 155–177, 2004.

Vivoni, E., Entekhabi, D., Bras, R., and Ivanov, V.: Controls on runoff generation and scale-dependence in a distributed hydrologic model, Hydrol. Earth Syst. Sci., 11, 1683–1701, 2007, http://www.hydrol-earth-syst-sci.net/11/1683/2007/.

Zehe, E. and Blöschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resour. Res., 40, W10202, doi:10.1029/2003WR002869, 2004.

12
Chapitre

Influence des mises-à-jour sur le calage du modèle

Il [...] se cachoyt en l'eau pour la pluye.

François Rabelais, *Gargantua*, Chapitre XI : De l'adolescence de Gargantua

Objectifs

- ▷ Évaluer l'impact des mises-à-jour sur le calage du modèle.
- ▷ Déterminer s'il est préférable de caler le modèle avec ou sans les mises-à-jour.

Sommaire

12.1 Le calage est influencé par les mises-à-jour 221
12.2 Faut-il tenir compte des mises-à-jour lors du calage ? 222
12.3 Stabilité des paramètres avec l'horizon de prévision 227
12.3.1 Stabilité des paramètres 227
12.3.2 Stabilité des performances 231
12.4 Bilan 231

Nous avons déjà rencontré la question du paramètrage du modèle et des difficultés du calage au chapitre 9. Il y était question de mise-à-jour des paramètres, une approche souvent employée en prévision (e.g. Corradini *et al.*, 1987, Tucci et Clarke, 1980, Yang, 1993, Calvo et Savi, 2009). Dans ce chapitre, la question du paramètrage sera discutée sous un angle différent : quelle est l'influence des mises-à-jour du modèle (hors mise-à-jour des paramètres) sur le calage ?

Dans ce contexte, nous nous intéresserons à nouveau plus à l'aspect « degrés de liberté » des paramètres qu'à leurs liens potentiels à des caractéristiques physiques du bassin ou des processus en jeux. Nous examinerons dans un premier temps dans quelle mesure la présence de mises-à-jour du modèle pendant le calage influence les valeurs calées des paramètres. Puis nous chercherons à savoir si cette approche, où peu de sens physique est donné aux paramètres, est justifiable : employer un jeu de paramètres calés en tenant compte des mises-à-jour, permet-il une amélioration des performances en contrôle ?

Si les paramètres ne sont compris que comme étant des degrés de liberté, leurs valeurs peuvent changer en fonction de l'horizon de prévision utilisé au calage. Cela peut entraîner des difficultés pratiques. Nous analyserons dans une troisième partie, la sensibilité des valeurs calées des paramètres à l'horizon de prévision.

12.1 Le calage est influencé par les mises-à-jour

Nous avons comparé les paramètres obtenus par le calage du modèle GRP avec trois stratégies de mise-à-jour différentes :

(a) aucune mise-à-jour (c'est-à-dire un calage en simulation),

(b) mise-à-jour directe du niveau du réservoir de routage (Tangara, 2005),

(c) mise-à-jour du niveau du réservoir de routage et correction des débits en sortie proposée par Tangara (2005).

Lors du calage, les paramètres ne sont pas calés directement tels qu'ils sont définis dans le modèle mais à travers des transformations numériques : X_1, capacité du réservoir de routage par une transformation logarithmique, X_2, facteur d'ajustement multiplicatif de la pluie efficace par une transformation quadratique et X_3, temps de base de l'hydrogramme unitaire par une transformation linéaire. Cela permet de caler les valeurs de ces paramètres transformés sur une même gamme [1] $[-10, 10]$ et d'adapter la sensibilité des paramètres (transformés) au pas de recherche de l'algorithme de calage. Les différences entre jeux de paramètres seront étudiées du point de vue du modèle, c'est-à-dire à travers les valeurs transformées des paramètres. De plus, cela permet de donner au modèle une sensibilité à peu près égale aux trois paramètres, donc de mettre les trois paramètres sur un pied

1. Le paramètre X_1 transformé peut dépasser la valeur maximale 10 mais ne s'en éloigne guère.

d'égalité et de pouvoir calculer une distance entre jeux de paramètres.

L'examen des distributions des paramètres montre qu'il y a de fortes différences entre les résultats du calage du modèle sans aucune mise-à-jour et du modèle avec la mise-à-jour du niveau du réservoir de routage. Les figures 12.1 et 12.2 en donnent un exemple pour un horizon de prévision de 6 heures. Les différences sont beaucoup plus faibles entre le calage du modèle avec la seule mise-à-jour du réservoir de routage et celui du modèle avec cette mise-à-jour et la correction des débits en sortie. Cette observation renvoie à un résultat de la partie III consacrée aux mises-à-jour : la mise-à-jour des états (et plus particulièrement celle du niveau du réservoir de routage) conduit à une forte amélioration des performances ; la correction des sorties du modèle proposée par Tangara (2005) permet en complément d'augmenter les performances dans une moindre mesure.

Nous avons ensuite examiné l'évolution des distances entre paramètres obtenus par calage du modèle avec les différentes stratégies de mise-à-jour (Fig. 12.3) en fonction de l'horizon de prévision. Les distances entre jeux de paramètres obtenus avec différentes mises-à-jour décroissent à mesure que l'horizon de prévision s'éloigne, ce qui est dû à la diminution de l'action de la mise-à-jour faite à la date t sur la prévision au pas de temps $t+H$ quand H augmente. Cette décroissance est lente quand on compare les versions (a) sans mise-à-jour et (b) avec mise-à-jour du réservoir de routage : la distribution des distances entre jeux de paramètres est loin d'être écrasée en un pic en 0 à l'horizon le plus lointain testé (48 heures). A contrario, les distances entre jeux de paramètres calés avec la mise-à-jour du réservoir de routage et (b) avec ou (c) sans la correction des débits en sortie du modèle tendent rapidement vers 0. On retrouve ici les ordres de grandeurs des durées de vie des différentes mises-à-jour (question discutée dans la partie précédente et par Berthet *et al.*, 2009b, reproduit dans l'annexe N, page 541) : l'effet de la mise-à-jour du réservoir de routage persiste pendant plusieurs centaines de pas de temps, celui des corrections des sorties est beaucoup plus évanescent.

12

12.2 Faut-il tenir compte des mises-à-jour lors du calage ?

Dans une approche déterministe « classique » (ou bien encore si les incertitudes sont évaluées par une méthode fréquentiste [2]), le modèle est calé dans un premier temps, puis employé sur de nouvelles données. Il est possible de se demander s'il est préférable ou non de mettre-à-jour le modèle pendant la phase de calage.

Mises-à-jour non paramétriques

Comme cela a été montré dans la section précédente, les mises-à-jour du modèle pendant le

2. Dans une approche bayésienne où l'incertitude sur les paramètres est traduite par des distributions de probabilité mises-à-jour à mesure que de nouvelles données sont disponibles, la question ne se pose pas en ces termes.

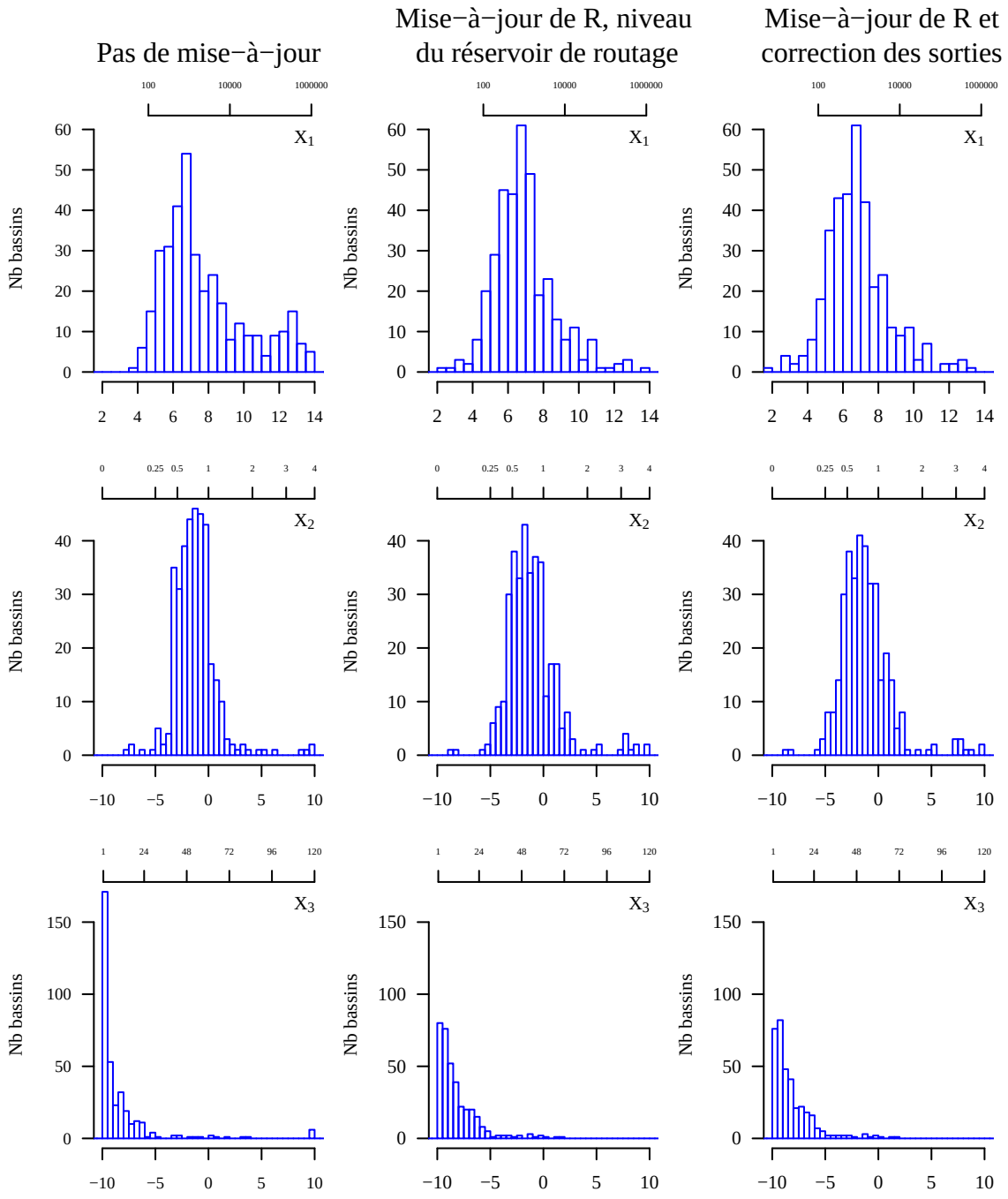


Figure 12.1 – Distributions (sur 178 bassins versants × 2 périodes) des trois paramètres X_1, X_2 et X_3 pour trois versions du modèle : sans aucune mise-à-jour (colonne de gauche), avec la mise-à-jour du niveau du réservoir de routage (colonne centrale), avec cette mise-à-jour et la correction des débits en sortie du modèle (dans la version de Tangara, 2005, colonne de droite). Les paramètres sont calés pour un horizon de 6 heures. Les axes principaux des abscisses (sous les figures) donnent l'échelle des valeurs transformées des paramètres tandis que les axes secondaires (au-dessus des figures) indiquent leurs valeurs réelles (X_1 en mm, X_2 sans dimension et X_3 en heures).

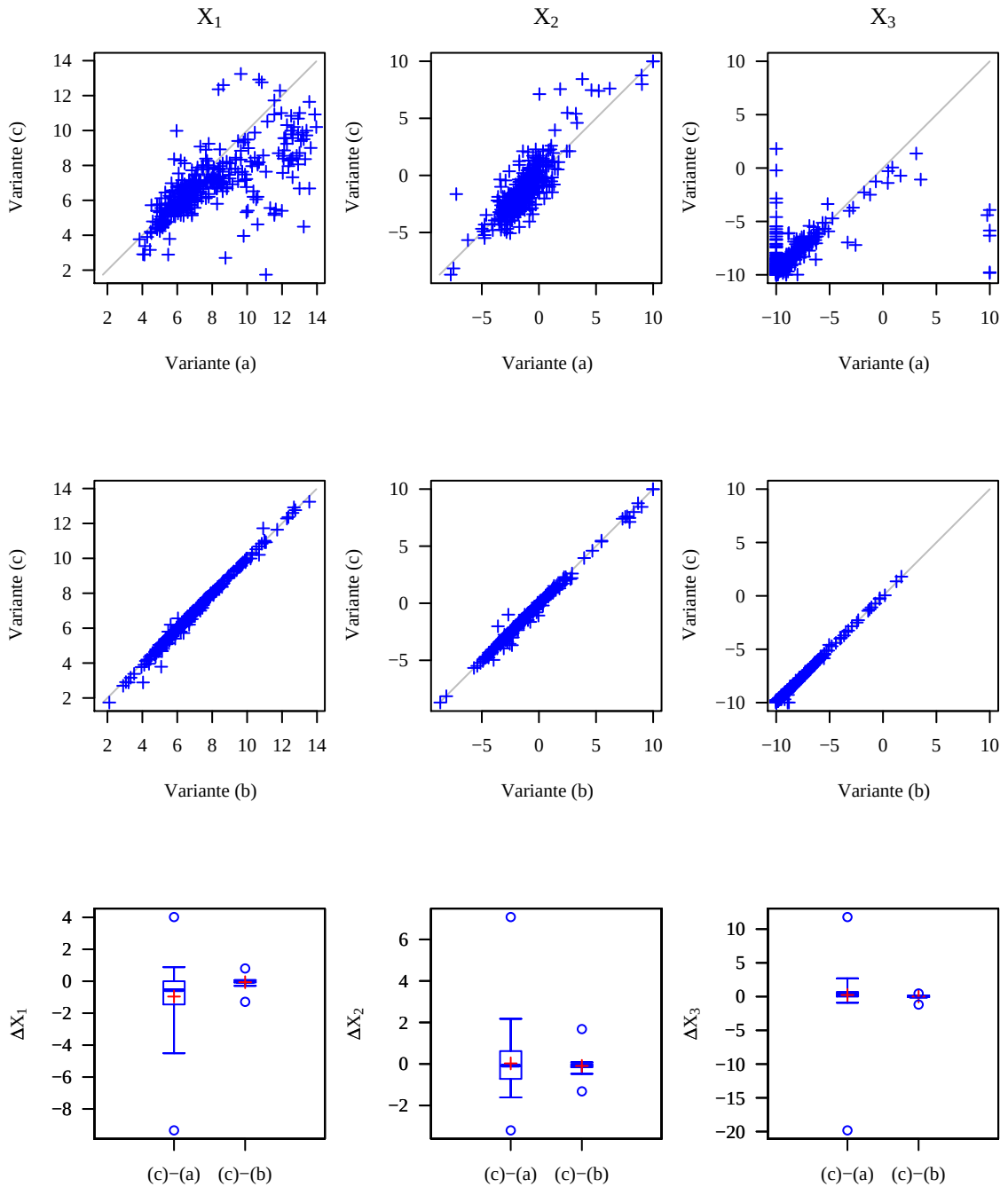


FIGURE 12.2 – Différences entre les valeurs calées des paramètres X_1, X_2 et X_3 pour trois schémas différents de mise-à-jour du modèle : (a) sans aucune mise-à-jour, (b) avec la mise-à-jour du niveau du réservoir de routage et (c) avec cette mise-à-jour et la correction des débits en sortie du modèle (dans la version de Tangara, 2005). Les paramètres sont calés pour un horizon de 6 heures.

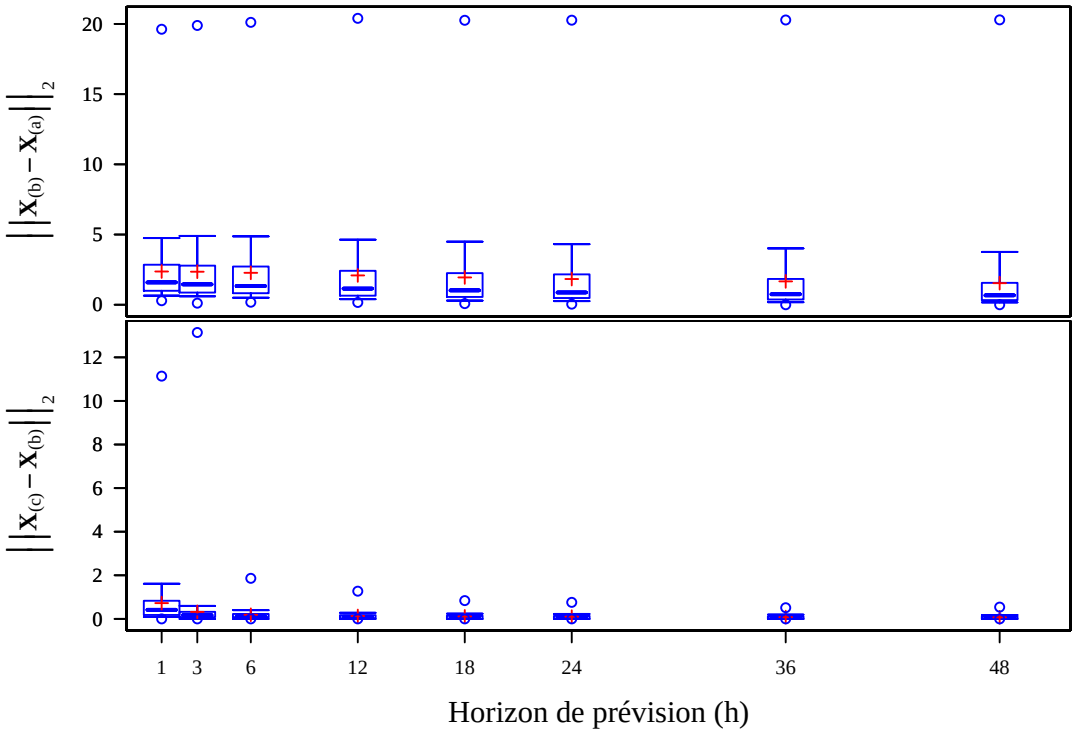


Figure 12.3 – Évolution des distributions (sur 178 bassins × 2 périodes) de la distance entre jeux de paramètres calés pour trois schémas différents de mise-à-jour du modèle : (a) sans aucune mise-à-jour, (b) avec la mise-à-jour du niveau du réservoir de routage et (c) avec cette mise-à-jour et la correction des débits en sortie du modèle (dans la version de Tangara, 2005). Les paramètres sont calés pour un horizon de 6 heures. La distance choisie est la distance euclidienne dans l'espace des paramètres transformés (pour que les paramètres aient une même gamme de variation).

calage déplacent le jeu de paramètres optimal (dans l'espace des paramètres). Même si on souhaite donner aux jeux de paramètres une signification hydrologique (et s'il est possible de le faire, ce qui suppose une confiance relativement élevée dans les différents élements du modèle : s'il est alimenté avec des données exactes, il simule bien le débit, pour de bonnes raisons au sens de Kirchner (2006)), les modifications dans le jeu de paramètres calé, induites par les mises-à-jour, ne sont pas forcément synonymes de détérioration de ce sens hydrologique. En effet, cela peut signifier que les stratégies de mise-à-jour sont adaptées aux erreurs (dans les entrées, les états ou encore les sorties), ce qui permet aux paramètres d'avoir des valeurs plus « réalistes ». Cela peut également être lié à une perte de sensibilité du calage vis à vis de certains paramètres, l'information du débit assimilé masquant alors leurs rôles.

Pour répondre à cette question dans une optique purement pratique, il suffit de comparer les performances du modèle en contrôle avec deux jeux de paramètres calés sur une même période : le premier obtenu alors que le modèle était mis-à-jour, le second sans mises-à-jour.

Tangara [3] a montré au pas de temps journalier qu'il valait mieux caler le modèle avec ses mises-à-jour.

Un test similaire a été réalisé au pas de temps horaire et conduit aux mêmes conclusions (Ann. F, page 395) : les performances du modèle calé en mode « Simulation » sont moindres que celles obtenues par le modèle calé en mode « Prévision » sur une grande majorité des bassins. On observe tout de même le contraire pour quelque 20 % des bassins, ce qui est loin d'être négligeable. Les bassins pour lesquels le calage en simulation conduit à de meilleures performances au contrôle (en prévision) que le calage en prévision ne présentent pas une gamme de valeurs d'autocorrélation du débit particulière et peuvent être des bassins sur lesquels le modèle obtient de bonnes performances aussi bien que de mauvaises. Les différences de performances sont significatives pour un grand nombre de bassins versants.

Nous avons recherché une tendance de la différence entre performances du modèle calé en simulation et en prévision en fonction de la réactivité des bassins versants (vue à travers le prisme de l'autocorrélation des débits). Une majorité des bassins dont l'autocorrélation du débit est la plus faible (bassins les plus rapides) obtiennent des performances similaires quand ils sont calés en prévision ou en simulation pour des horizons de l'ordre d'un à deux jours. Au contraire, pour une majorité des bassins appartenant à la classe des bassins les moins réactifs, les performances du modèle calé en prévision restent significativement meilleures de celles du modèle calé en simulation, même à un horizon de prévision de 5 jours (120 heures).

Pour quelques bassins, le calage en mode « Simulation » est très pénalisant mais les bassins pour lesquels la différence est le plus fortement négative sont le plus souvent des bassins sur lesquels le modèle obtient de toute façon de piètres performances.

Mises-à-jour paramétriques

Si les mises-à-jour sont paramétriques, le problème est différent. Des compensations entre paramètres « hydrologiques » (i.e., les paramètres X_1, X_2 et X_3 de la structure du modèle) et paramètres des mises-à-jour sont à craindre s'ils sont calés simultanément. Cela peut souvent conduire à des paramètres « hydrologiques » très peu interprétables. Cette situation devrait être d'autant plus probable que le nombre de paramètres des méthodes de mise-à-jour serait élevé (pour des modèles ARIMA un peu complexes, et pour tous les réseaux de neurones). Pour éviter ce problème et d'autres difficultés pratiques (temps de calcul, fonctionnement de l'algorithme de calage), il paraît préférable de caler le modèle hydrologique sans les mises-à-jour paramétriques, puis dans un second temps de caler ces méthodes de mise-à-jour.

Pour illustrer ce point, nous avons comparé deux stratégies de calage d'une variante du modèle GRP. Dans la version de Tangara (2005), l'exposant γ de la correction des erreurs

3. Communication personnelle.

en sortie du modèle est un paramètre fixe. Il sera ici un quatrième paramètre à caler. Deux calages sont étudiés :

- Calage en mode « Prévision » : les quatre paramètres sont calés ensemble pour chaque horizon de prévision.
- Calage en mode « Simulation » : les trois paramètres X_1, X_2 et X_3 sont calés dans un premier temps sans les mises-à-jour. Puis le paramètre γ est calé pour chaque horizon.

Les jeux de paramètres ainsi calés ont été comparés aux valeurs obtenues par calage en mode « Prévision » des trois paramètres de GRP (pour lequel γ était fixe). Comme on s'y attendait, la première option permet d'obtenir des résultats légèrement meilleurs que la seconde option en contrôle. Par contre, pour quelques bassins (de 4 à 5 % aux horizons courts), on observe des changements très significatifs des valeurs de la capacité du réservoir de production X_1 et du temps de base de l'hydrogramme unitaire X_3 (les différences entre valeurs transformées des paramètres sont supérieures à 5). Il est à noter qu'un changement de valeur du paramètre X_1 n'entraîne pas un changement pour le paramètre X_3 (et vice-versa). Enfin, ces bassins ne se caractérisent pas par des performances faibles. Ainsi, sur un nombre réduit de bassins (mais nous n'avions ici qu'un seul paramètre additionnel), on note des variations fortes des valeurs des paramètres de la structure hydrologique.

12.3 Stabilité des paramètres avec l'horizon de prévision

L'influence des mises-à-jour sur le modèle décroît quand l'horizon de prévision augmente. Comme la prise en compte des mises-à-jour lors du calage modifie le jeu de paramètres calé, il faut s'attendre à ce que le calage aboutisse à des résultats différents pour des horizons de prévision différents. Cette situation peut avoir des conséquences pratiques désagréables. Les prévisionnistes ne cherchent pas à prévoir les débits à un seul horizon H donné mais si possible sur la plage horaire $[t, t+H]$ (t étant le pas de temps présent). Si le paramétrage n'est performant que pour établir la prévision $\widehat{Q}_{t+H|t}$ alors il faudrait disposer de tous les paramétrages correspondant aux horizons H' entre 1 et H (évidemment faisable mais peu pratique). Nous examinons dans cette section la stabilité des jeux de paramètres et la stabilité des performances du modèle en fonction de l'horizon de prévision employé au calage.

12.3.1 Stabilité des paramètres

Avec mise-à-jour des états et correction des débits en sortie

Le modèle GRP (dans la version de Tangara, 2005) a été calé aux horizons de prévision 1, 3, 6, 12, 18, 24, 36 et 48 heures. Nous avons examiné l'évolution des distributions de chaque paramètre avec l'horizon de prévision et comparé ces jeux de paramètres avec le jeu de paramètre obtenu par un calage en simulation (sans aucune mise-à-jour : cf. supra).

Les distributions des paramètres calés en mode « Prévision » sont stables avec une évolution légère (Fig. 12.4). La moyenne des écarts des valeurs des paramètres aux différents horizons de prévision par rapport à l'horizon 1 heure reste toujours très proche de 0 (ce qui est cohérent avec la stabilité des paramètres). La distribution de ces écarts s'étale doucement à mesure que l'horizon de prévision s'éloigne. Pour la très grande majorité des bassins, ces écarts restent contenus à $\pm$ 1 (pour les paramètres transformés).

Les valeurs des paramètres tendent vers la valeur obtenue en simulation quand l'horizon de prévision s'éloigne. Cette observation était attendue et correspond à l'atténuation de l'effet des mises-à-jour pour les horizons de prévision lointains. Les distributions des paramètres calés pour le plus grand horizon testé (48 h) n'ont pas encore rejoint les distributions des paramètres calés en simulation : les mises-à-jour sont toujours actives.

Enfin dans la même logique, on observe que la vitesse de « convergence » du jeu de paramètres calé en mode « Prévision » vers le jeu de paramètres calé en mode « Simulation » en fonction de l'horizon de prévision est, en moyenne, plus rapide pour les bassins fortement réactifs que pour les bassins les plus lents.

Par contre, aucune tendance de l'évolution des paramètres n'a pu être détectée en fonction des performances du modèle ou encore de leur valeur calée en simulation.

Impact des différentes mises-à-jour du modèle

Détaillons maintenant l'impact des différentes mises-à-jour. Nous avons d'abord vérifié qu'en mode « Simulation », le calage est indépendant de l'horizon de prévision : ce sont bien les mises-à-jour qui sont à l'origine de l'influence de l'horizon sur le calage. La figure 12.5 présente l'évolution des valeurs des paramètres du modèle GRP avec une stratégie de mise-à-jour différente : seul le niveau du réservoir de routage est mis-à-jour. Le calage des paramètres (notamment X_3) est nettement moins sensible à l'horizon de prévision que le calage du modèle GRP avec ses deux mises-à-jour. Cela s'explique par la durée de vie élevée de la mise-à-jour du niveau du réservoir de routage : l'effet de cette mise-à-jour perdure longtemps et est similaire pour tous les horizons de prévision qui nous intéressent.

La correction des débits en sortie du modèle (dans sa version non-paramétrique Tangara, 2005) est donc le principal responsable de l'évolution du calage en fonction de l'horizon de prévision. Ce résultat est en accord avec la brièveté des durées de vie de ce type de mise-à-jour.

Ainsi, pour les corrections paramétriques de l'erreur en sortie du modèle, le calage en deux étapes (des trois paramètres de la structure X_1, X_2 et X_3 dans un premier temps ; puis des paramètres de la méthode de mise-à-jour) permet d'avoir des jeux de paramètres "hydrologiques" plus stables. Par contre, les valeurs des paramètres de la méthode de

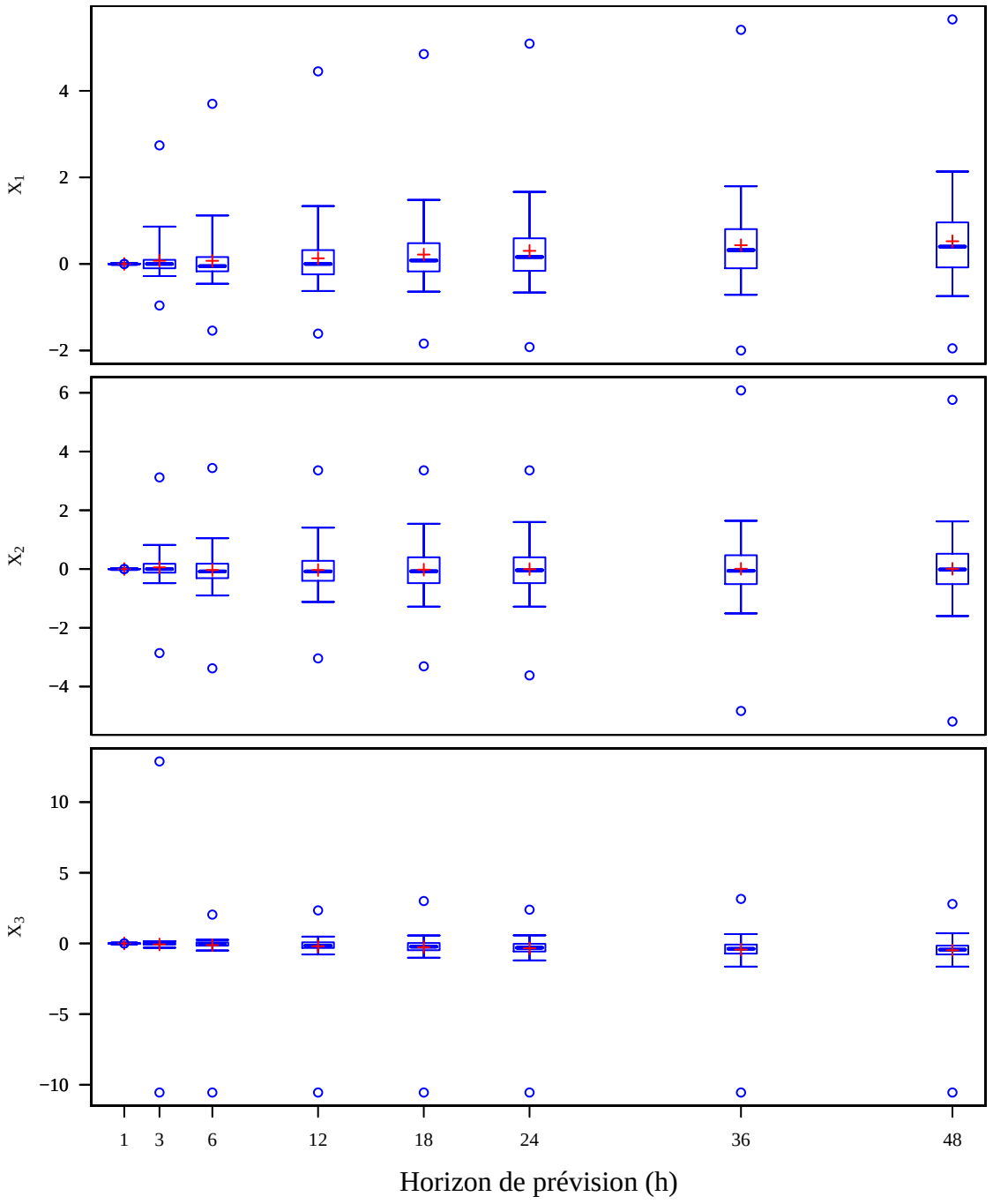


Figure 12.4 – Distributions (sur 178 bassins versants × 2 périodes) des différences entre les valeurs des paramètres du modèle GRP calé à à l'horizon de prévision h et celles des paramètres calés à un horizon d'une heure, en fonction de l'horizon h. Ici le modèle fonctionne avec ses deux mises-à-jour : celle du niveau du réservoir de routage et la correction des débits simulés.

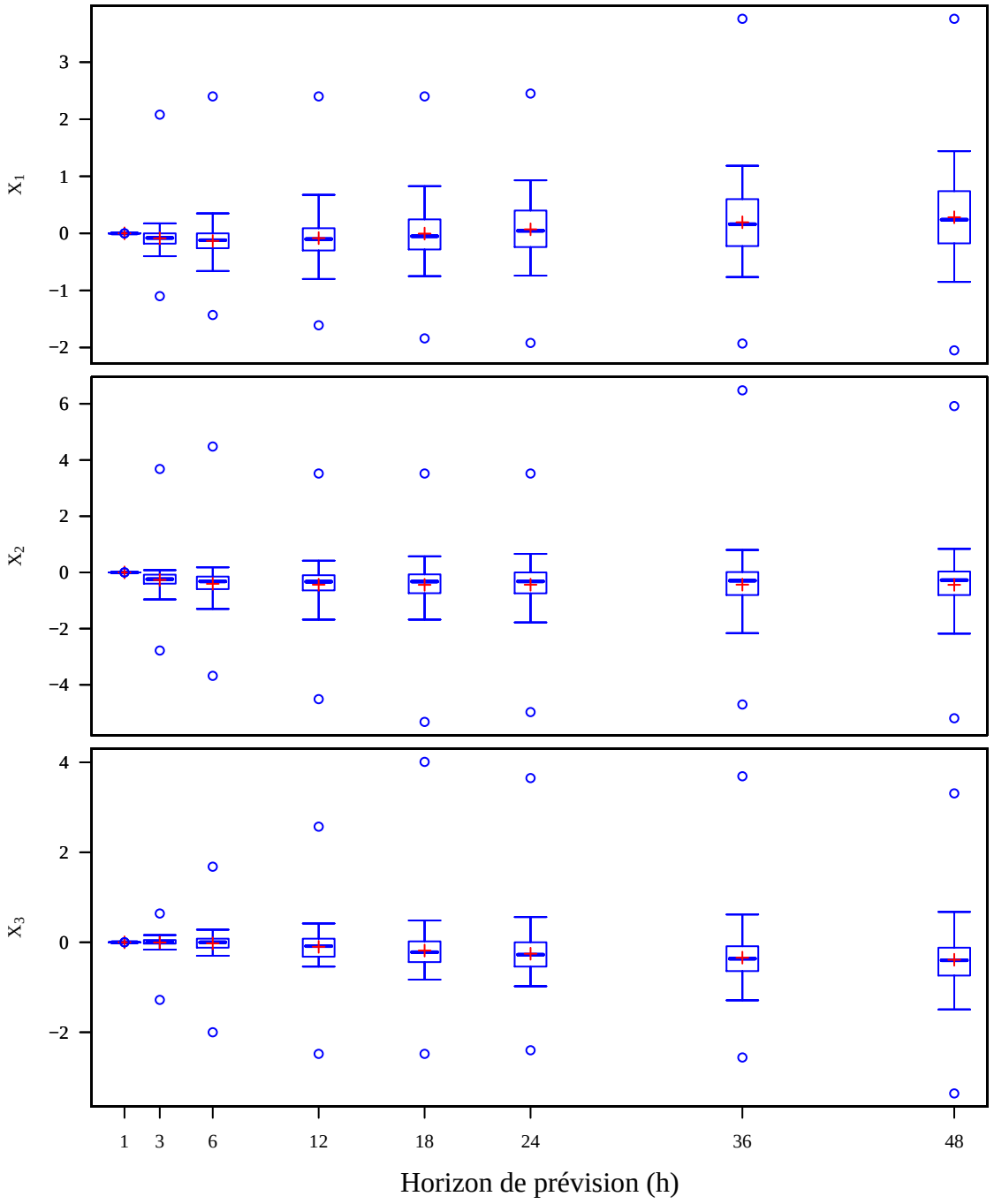


FIGURE 12.5 – Distributions (sur 178 bassins versants) des différences entre les valeurs des paramètres du modèle GRP calé à à l'horizon de prévision h et celles des paramètres calés à un horizon d'une heure, en fonction de l'horizon h. Seul le niveau du réservoir de routage est mis-à-jour ici.

correction des erreurs en sortie (ARIMA ou réseaux de neurones artificiels) peuvent être (très) dépendants de l'horizon de prévision.

12.3.2 Stabilité des performances

À la différence de la section précédente, le modèle (dans sa version non paramétrique de Tangara, 2005) est calé au préalable aux horizons de prévision H_C 1, 3, 6, 12, 24 et 48 heures en mode « Prévision ». Les performances au contrôle (mesurées par $C_{2M,P}$) sont calculées aux horizons de prévision h_V 1, 3, 6, 12, 24 et 48 heures.

Pour obtenir les meilleures performances à un horizon de prévision h_V, il faut avoir calé le modèle à cet horizon de prévision (Fig. 12.6). On constate ensuite que les distributions des performances en validation obtenues avec les paramètres calés aux différents horizons de prévision restent proches et se dégradent lentement à mesure que H_C s'éloigne de H_V.

Si on s'intéresse maintenant aux distributions des écarts de performances, on constate que pour des horizons de prévision en calage H_C et en validation H_V assez proches (moins d'une vingtaine de pas de temps), les performances restent proches sur la majorité des bassins versants (écarts absolus inférieurs à 3 points de $C_{2M,P}$, la différence est non significative, Fig. 12.6). Plus les horizons H_C et H_V diffèrent, plus les écarts extrêmes se creusent. On observe également que les écarts les plus grands sont le plus souvent obtenus pour des bassins pour lesquels les performances à l'horizon h_V sont faibles (inférieures ou proches de 0).

Enfin, plus l'horizon de prévision choisi en validation est grand, plus il est important que l'horizon de prévision employé au calage soit proche de cet horizon : il vaut mieux caler à 48 heures pour prévoir à 1 heure que caler à 1 heure pour prévoir à 48 heures (Fig. 12.6). Ceci est logique avec le fait que caler à 48 heures signifie montrer au modèle la résultante de ce qui se passe de 1 à 48 heures, tandis que caler à 1 heure ne fait explorer qu'un pas de temps au modèle.

12.4 Bilan

Comme on devait s'y attendre, les mises-à-jour du modèle influencent le calage et conduisent à l'obtention de jeux de paramètres différents de jeux qui auraient été calés en mode « Simulation ». Les variations du jeu de paramètres calés sont essentiellement dues à la mise-à-jour du niveau du réservoir de routage.

Cependant, pour une stratégie de mise-à-jour donnée, les jeux de paramètre évoluent assez peu en fonction de l'horizon de prévision employé au calage et il est possible de caler les paramètres X_1, X_2 et X_3 à un seul horizon pour effectuer les prévisions sur une gamme d'horizons d'un à deux jours. Les différences entre jeux de paramètres calés en mode « Prévision » à deux horizons différents sont principalement dues à la correction des débits

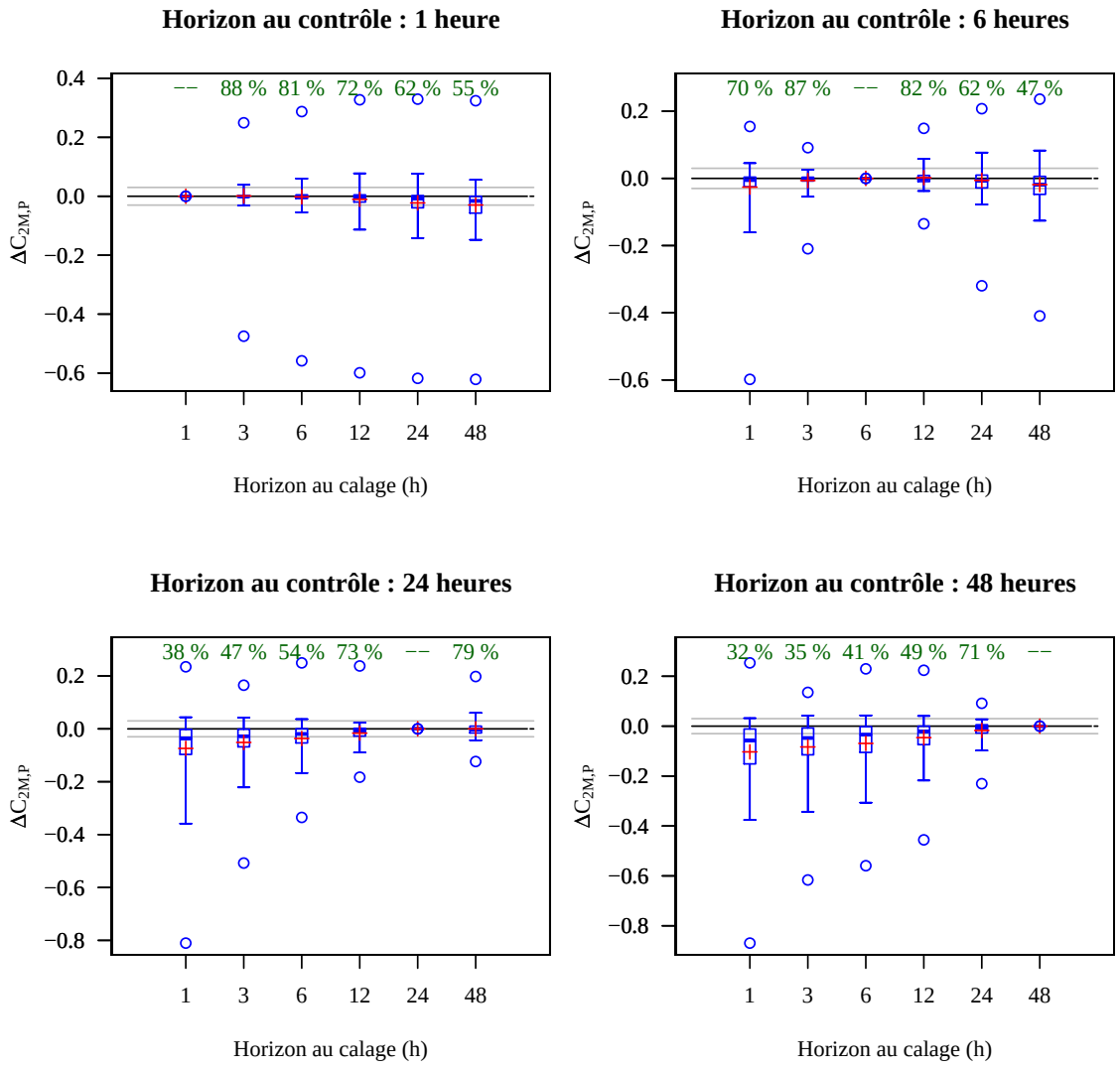


FIGURE 12.6 – Exemple de comparaison des distributions des écarts de performance en contrôle, en fonction de l'horizon de prévision employé au calage. Les pourcentages au-dessus des boîtes à moustaches indiquent les fractions de l'échantillon de bassins pour lesquelles la différence de performance est inférieure (en valeurs absolues) à 0.03 (lignes grisées horizontales).

en sortie du modèle. Les meilleures performances (en contrôle) à un horizon donné sont obtenues en calant le modèle à cet horizon cible. Si une gamme d'horizon intéresse le prévisionniste, il est préférable de caler sur un horizon plus grand que plus petit pour obtenir les meilleures performances sur cette gamme.

Enfin, comme au pas de temps journalier, il a été montré qu'il valait mieux inclure les mises-à-jour (non paramétriques) lors du calage des paramètres de la structure du modèle.

Cinquième partie

Évaluation du modèle de prévision

13

Chapitre

Évaluation du modèle sur un vaste échantillon de bassins versants

for as for geography, 'tis of absolute use to him ; he must be acquainted intimately with every country and its boundaries where his profession carries him ; he should know every town and city, and village and hamlet, with the canals, the roads, and hollow ways which lead up to them ; there is not a river or a rivulet he passes, Trim, but he should be able at first sight to tell thee what is its name – in what mountains it takes its rise – what is its course – how far it is navigable – where fordable – where not.

Lawrence Sterne, *The Life and Opinions of Tristram Shandy, Gentleman* Vol. VIII, Chap. XIX

Objectifs

- ▷ Évaluer le modèle sur un large échantillon de bassins versants :
- ▷ à travers des critères variés,
- ▷ sur des bassins non utilisés pour établir le modèle.
- ▷ Évaluer l'impact de la sélection de 178 bassins versants censés représenter la variabilité hydroclimatique d'un millier, sur l'évaluation des performances.

Sommaire

13.1 Évaluation du modèle GRP sur 846 bassins « indépendants » . 238
13.1.1 Critères quadratiques . . . 238
13.1.2 Critères temporels . . . 240
13.1.3 Critères événementiels d'alerte . . . 243
13.2 Où ? Quand ? Comment ? . . . 245
13.2.1 Influence de la taille des bassins versants . . . 247
13.2.2 Influence de la réactivité des bassins versants . . . 247
13.2.3 Influence de la performance du modèle en simulation . . . 248
13.2.4 Influence de la localisation des bassins versants . . . 249
13.3 Évaluations complémentaires . . . 251
13.3.1 Avec un scénario de pluies futures nulles . . . 251
13.3.2 Extrapolation . . . 252
13.4 Pertinence de la réduction de l'échantillon de bassins versants 252
13.5 Bilan . . . 255

Le travail présenté dans les chapitres précédents a permis d'établir un modèle de prévision des crues au pas de temps horaire, c'est-à-dire la combinaison d'une structure et d'une ou plusieurs stratégies de mise-à-jour s'appuyant sur l'assimilation des dernières observations de débit. Ces résultats sont basés sur une phase de tests réalisés sur les séries de données de pluie et de débit pour 178 bassins versants. Il est maintenant nécessaire de tester ce modèle sur des bassins versants qui n'ont pas été employés lors de sa mise au point.

Nous disposons d'un ensemble d'un millier de bassins versants français (Le Moine *et al.*, 2007). Nous commencerons par évaluer le modèle sur le sous-ensemble des 846 bassins n'ayant pas servi aux tests qui ont permis d'aboutir à notre modèle. Ces 846 bassins ne sont toutefois pas totalement indépendants de l'échantillon de 178 bassins versants. En effet, ces derniers ont été sélectionnés pour représenter la variabilité de l'ensemble complet (Chap. 4). Dans la première section, les valeurs des critères de performance définis dans le chapitre 3 (page 39) seront calculées sur ces 846 bassins. Ces critères s'intéressent à différents aspects des prévisions :

- des critères quadratiques évaluant globalement l'adéquation de l'hydrogramme prévu à l'hydrogramme observé,
- des critères de décalage temporels : retard sur les pics, retard moyen lors de la montée de crue et avance de prévision,
- des critères événementiels d'alerte (au dépassement de seuil).

Nous chercherons par la suite à caractériser où, quand et comment le modèle se révèle bon ou mauvais (au sens relatif du terme) à travers l'influence de divers facteurs (telles les superficies des bassins, leurs localisations, leurs réactivités ou les performances obtenues par le modèle en mode « Simulation ») sur les valeurs des différents critères. Dans une troisième section, nous comparerons les performances du modèle obtenues sur l'ensemble des bassins et sur l'échantillon de 178 bassins à travers quelques critères de performance clefs.

Nous comparerons plusieurs stratégies de mise-à-jour du modèle. La première stratégie notée GRP-T correspond aux résultats de Tangara (2005) : le niveau du réservoir de routage est mis-à-jour de façon complète en fonction de la dernière observation du débit et les débits calculés sont corrigés en sortie du modèle par un modèle autorégressif multiplicatif. Puis viennent les deux stratégies qui apparaissent comme étant les plus intéressantes au pas de temps horaire : la correction des débits calculés est réalisée par un modèle ARIMA (version GRP-ARIMA) ou par un réseau de neurones (GRP-RNA). Enfin, à titre de comparaison, la version sans aucune correction des débits en sortie, GRP-0, sera évaluée. Comme les calculs de correction par réseaux de neurones n'ont pu être réalisés que sur les pas de temps présentant les 10 % des plus forts débits, toutes les variantes seront évaluées sur ces mêmes pas de temps (hautes eaux).

Il nous faut enfin mentionner que, sauf mention contraire, tous les tests ont été réalisés

avec un scénario de pluies futures égales aux observations postérieures, de façon à évaluer le modèle hydrologique, plus ou moins indépendamment de la qualité de ses entrées.

13.1 Évaluation du modèle GRP sur 846 bassins « indépendants »

Cet échantillon correspond à l'ensemble d'un millier de bassins versants auquel ont été soustraits les 178 bassins ayant servi à établir le modèle que nous testons.

13.1.1 Critères quadratiques

Six critères quadratiques de la famille $C_{2M,P}$ (Chap. 3 page 41) ont été calculés pour trois seuils de débit correspondant aux quantiles 0.90 (i.e., sur l'ensemble des chroniques dont nous disposons), 0.95 et 0.99. Pour chacun de ces seuils, nous avons pris en compte soit l'ensemble des pas de temps où le débit dépasse ces seuils, soit uniquement les phases de montée de crue.

Les ordres de grandeur des valeurs de $C_{2M,P}$ calculées sur l'ensemble des chroniques pour les différents modèles sont identiques à ceux obtenus avec l'échantillon de 178 bassins. Les mêmes conclusions sont obtenues : il y a un avantage net à employer la correction des débits en sortie par réseaux de neurones (Tab. 13.1 et Fig. 13.1).

Quand on s'intéresse aux pointes (débits au dessus du quantile 0,99), les réseaux de neu-

Critères	1 h	3 h	6 h	12 h	18 h	24 h	36 h	48 h
GRP-ARIMA								
$C_{2M,P}\left(Q \geq q_{0.90}\left(Q\right)\right)$	0.63	0.48	0.45	0.48	0.54	0.57	0.61	0.63
$C_{2M,P}\left(Q \geq q_{0.95}\left(Q\right)\right)$	0.64	0.49	0.46	0.50	0.54	0.57	0.61	0.62
$C_{2M,P}\left(Q \geq q_{0.99}\left(Q\right)\right)$	0.64	0.48	0.44	0.48	0.52	0.55	0.58	0.59
$C_{2M,P}\left(Q \geq q_{0.90}\left(Q\right)\right)$ en montée	0.65	0.50	0.45	0.48	0.51	0.54	0.57	0.58
$C_{2M,P}\left(Q \geq q_{0.95}\left(Q\right)\right)$ en montée	0.66	0.51	0.45	0.49	0.53	0.55	0.58	0.59
$C_{2M,P}\left(Q \geq q_{0.99}\left(Q\right)\right)$ en montée	0.67	0.49	0.44	0.49	0.54	0.56	0.58	0.59
GRP-RNA								
$C_{2M,P}\left(Q \geq q_{0.90}\left(Q\right)\right)$	0.67	0.54	0.50	0.53	0.56	0.59	0.65	0.69
$C_{2M,P}\left(Q \geq q_{0.95}\left(Q\right)\right)$	0.69	0.57	0.51	0.53	0.56	0.58	0.63	0.66
$C_{2M,P}\left(Q \geq q_{0.99}\left(Q\right)\right)$	0.68	0.55	0.51	0.51	0.52	0.54	0.56	0.57
$C_{2M,P}\left(Q \geq q_{0.90}\left(Q\right)\right)$ en montée	0.69	0.56	0.50	0.48	0.50	0.50	0.53	0.56
$C_{2M,P}\left(Q \geq q_{0.95}\left(Q\right)\right)$ en montée	0.70	0.58	0.52	0.50	0.51	0.51	0.53	0.55
$C_{2M,P}\left(Q \geq q_{0.99}\left(Q\right)\right)$ en montée	0.69	0.57	0.52	0.52	0.54	0.54	0.56	0.55

TABLE 13.1 – Médiane des distributions (sur 846 bassins versants × 2 périodes) des critères de persistance calculés pour les deux stratégies de mise-à-jour retenues : GRP-ARIMA et GRP-RNA.

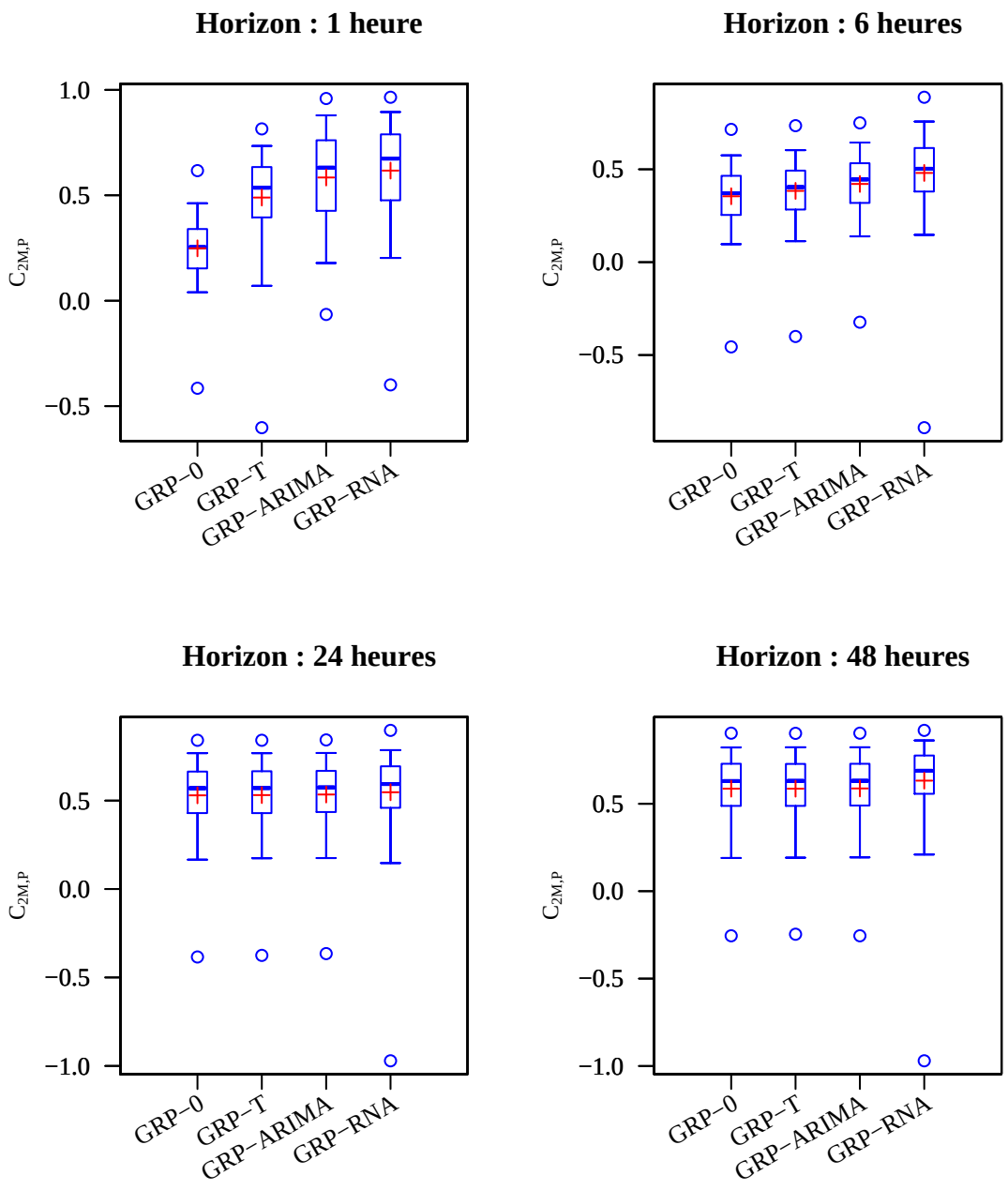


FIGURE 13.1 – Distributions (sur 846 bassins versants × 2 périodes) des valeurs du critère $C_{2M,P}$ calculées en contrôle sur tous les pas de temps de débit supérieurs au quantile 0.90 (i.e. sur toute la gamme des débits ayant servi au calage des RNA) pour 846 bassins, des modèles GRP-0, GRP-Tangara, GRP-ARIMA et GRP-RNA.

rones maintiennent leur avantage aux horizons de prévision courts (inférieurs à 24 heures, Fig. 13.2). Aux horizons de prévision plus lointains, les réseaux de neurones n'apportent plus de gain en performance sur les pointes, voire dégradent légèrement les prévisions (de façon faiblement significative).

Les observations sont similaires quand on examine les performances uniquement sur les phases de montée de crue.

Enfin, comme nous l'avions constaté au chapitre 10, le modèle GRP-RNA obtient des performances très faibles sur un très petit nombre de bassins versants que nous nous efforcerons de caractériser dans la prochaine section.

13.1.2 Critères temporels

Les critères temporels ont été calculés sur des événements de crue. Nous avons défini d'abord pour chaque bassin et pour chacune des deux périodes de données (5 ans), les événements en prenant pour seuil le débit, le quantile 0.90. Parmi eux, nous avons sélectionné les 20 événements d'une durée minimale de 12 heures (pour pouvoir réaliser les calculs) présentant les plus hauts débits. S'il y avait moins de 20 événements d'une durée supérieure ou égale à 12 heures, ils ont été tous pris en compte pour établir les statistiques qui suivent.

Critères sur le pic

Les trois variantes du critère sur le pic testées (R_1, $R_{1bis}(0.8)$ et $R_{1bis}(0.9)$ définis page 44) nous apportent les mêmes informations : le modèle GRP-RNA présente un retard plus faible sur le pic que les autres modèles, ce à tous les horizons de prévision (Tab. 13.2). On ne peut guère être plus précis car le fait que ces critères prennent des valeurs entières rend la lecture de leurs distributions difficile. À certains horizons de prévision, seule la moyenne permet de distinguer les modèles, les uns des autres.

Critère sur le retard à la montée R_2

Nous avions constaté lors de l'examen des hydrogrammes de prévision du modèle GRP-RNA (voir les exemples donnés au chapitre 10, page 198) que dans de nombreux cas, les prévisions par ce modèle étaient plus proches des observations pendant les montées de crue que celles fournies par les autres modèles testés. Ces observations sont confirmées par le critère R_2 sur 846 nouveaux bassins : le modèle GRP-RNA présente un retard à la montée bien inférieur à celui des autres modèles (4^e et 8^e lignes du tableau 13.2 et Fig. 13.3). Ce critère qui prend des valeurs réelles (non entières) est beaucoup plus lisible que les précédents et permet une meilleure comparaison des différents modèles.

Critère d'avance

Enfin, nous avons étudié l'horizon maximal permettant sur chaque bassin de prévoir le

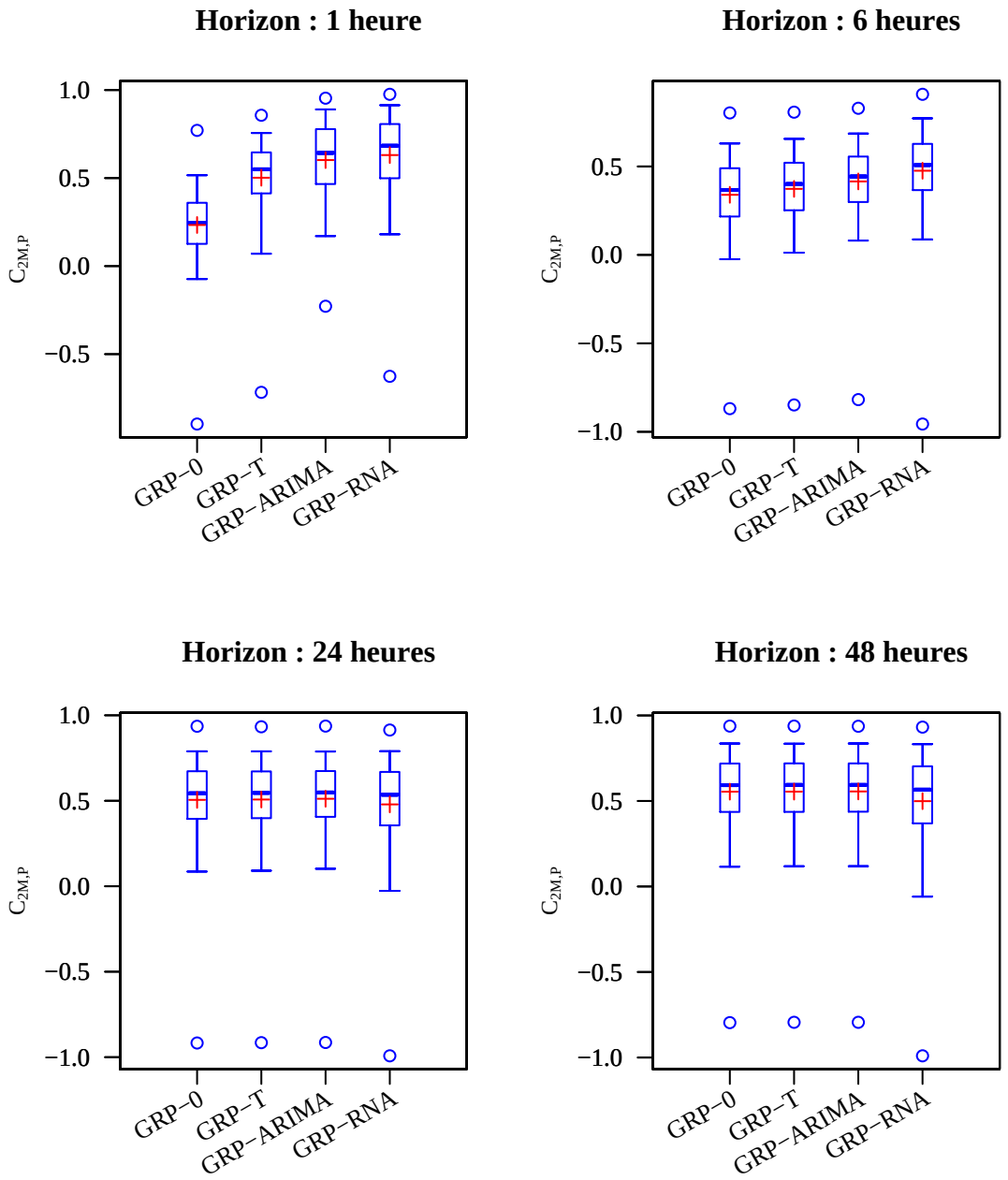


FIGURE 13.2 – Distributions (sur 846 bassins versants × 2 périodes) des valeurs du critère $C_{2M,P}$ calculées en contrôle sur tous les pas de temps de débit supérieurs au quantile 0.99 pour 846 bassins, des modèles GRP-0, GRP-Tangara, GRP-ARIMA et GRP-RNA.

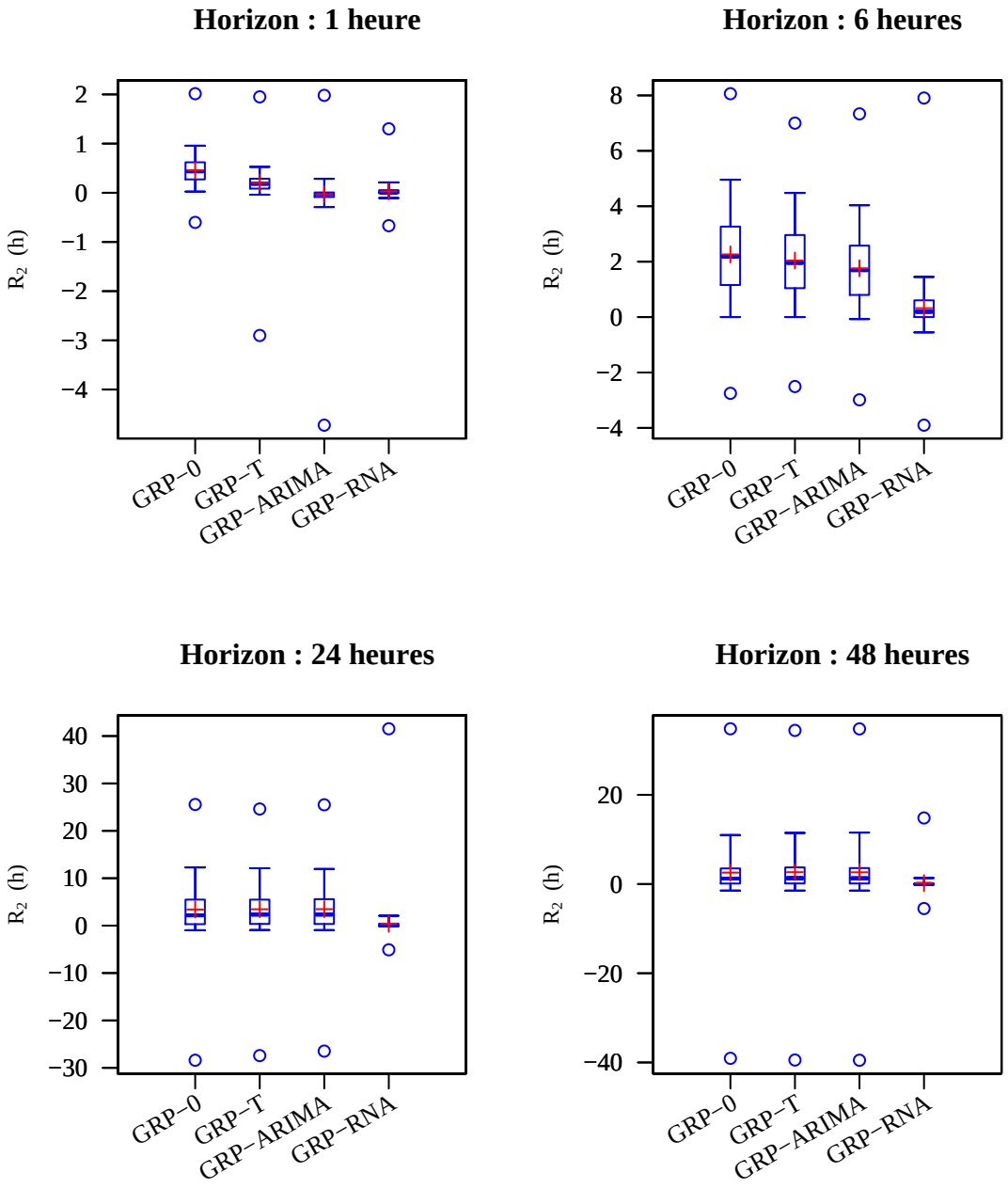


FIGURE 13.3 – Distributions (sur 846 bassins versants × 2 périodes) des valeurs du critère R_2 de retard à la montée, calculées sur 846 bassins pour les modèles GRP-0, GRP-Tangara, GRP-ARIMA et GRP-RNA.

Critères	1 h	3 h	6 h	12 h	18 h	24 h	36 h	48 h
				GRP-ARIMA				
R_1	1.00	2.00	4.00	5.00	4.50	2.50	1.00	0.50
$R_{1bis}(0.9)$	0.00	1.50	3.00	4.00	2.50	1.00	0.00	0.00
$R_{1bis}(0.8)$	0.00	1.00	2.00	2.50	1.00	0.00	0.00	0.00
R_2	-0.02	0.65	1.69	2.84	2.92	2.34	1.65	1.30
				GRP-RNA				
R_1	0.00	1.50	2.00	2.00	1.00	0.50	0.00	0.00
$R_{1bis}(0.9)$	0.00	1.00	1.50	1.00	0.00	0.00	0.00	0.00
$R_{1bis}(0.8)$	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
R_2	0.00	0.07	0.19	0.20	0.10	0.04	0.00	0.00

Table 13.2 – Médiane des distributions (sur 846 bassins versants × 2 périodes) des critères de retard calculés pour les deux stratégies de mise-à-jour retenues : GRP-ARIMA et GRP-RNA. Les retards sont exprimés en heures.

dépassement du seuil $q_{99}(Q)$ avec une probabilité de 75 et 90 % (page 46) avec le modèle GRP-RNA. Pour une probabilité de 75 %, cet horizon est supérieur ou égal à 12 heures pour un bassin versant sur deux et il reste supérieur ou égal à 24 heures sur un tiers des bassins. Si le seuil de probabilité exigé est élevé à 90 %, l'horizon décroît : il n'est supérieur ou égal à 12 heures que dans 10 % des cas (Fig. P.17, page 586).

13.1.3 Critères événementiels d'alerte

Les critères d'alerte suivants ont été calculés pour un seuil d'alerte correspondant au quantile 0.99 des débits (soit environ quatre jours dans l'année). Ils sont calculés à chaque pas de temps de débit supérieur au quantile 0.90, indépendamment de l'occurrence d'une alerte aux pas de temps précédents.

Taux de fausses alertes (FAR)

Les observations sont très cohérentes avec celles faites sur les critères quadratiques (Tab. 13.3 et Fig. 13.4). Le modèle GRP-RNA présente des taux de fausses alertes inférieurs à ceux des modèles GRP-ARIMA, eux-mêmes inférieurs à ceux du modèle employant la correction de Tangara (2005). Au-delà d'un horizon de 12 heures, tous les modèles sauf celui avec correction par RNA sont équivalents. Ce dernier continue à produire moins de fausses alertes que les autres approches pour un horizon de 48 heures.

Probabilité de détection (POD)

À un horizon d'une heure, les modèles GRP-ARIMA et GRP-RNA obtiennent des scores équivalents. Mais dès l'horizon de 3 heures, le modèle GRP-RNA se révèle meilleur (Tab. 13.3). La différence s'accentue à mesure que l'horizon de prévision s'éloigne.

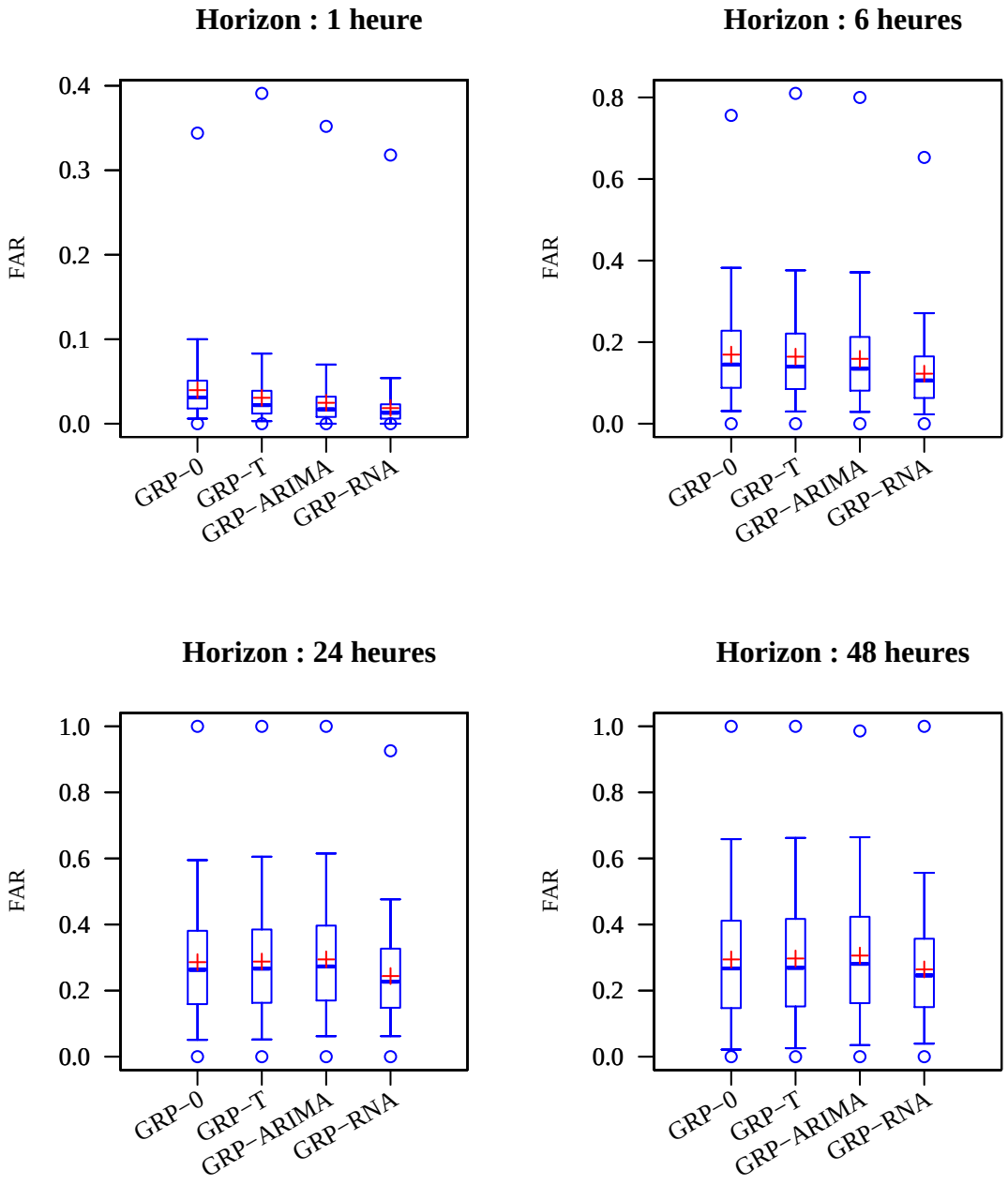


FIGURE 13.4 – Distributions (sur 846 bassins versants × 2 périodes) des valeurs du taux de fausses alertes (FAR), calculées sur 846 bassins pour les modèles GRP-0, GRP-Tangara, GRP-ARIMA et GRP-RNA.

Indice de succès critique (CSI)

Les mêmes observations sont faites quant au critère CSI (Tab. 13.3 et Fig. 13.5), ce que nous attendions du fait de la relation existant entre les critères CSI, FAR et POD (e.g. Schaefer, 1990).

La modification proposée du critère CSI en CSI' (Chap. 3, page 44), visant à amoindrir le poids des fausses alertes, revient à se rapprocher de l'indice POD. Comme l'examen des distributions de ces deux critères CSI et POD conduit aux mêmes observations, l'emploi de la version modifiée CSI' n'est pas plus informative.

Critères	1 h	3 h	6 h	12 h	18 h	24 h	36 h	48 h
				GRP-ARIMA				
CSI	0.97	0.86	0.73	0.57	0.50	0.46	0.42	0.40
$CSI'(0.8)$	0.97	0.87	0.75	0.59	0.52	0.48	0.44	0.41
$CSI'(0.5)$	0.98	0.89	0.78	0.62	0.55	0.51	0.46	0.43
$CSI'(0.2)$	0.98	0.91	0.81	0.66	0.59	0.54	0.49	0.47
POD	0.98	0.92	0.83	0.69	0.62	0.57	0.52	0.49
FAR	0.02	0.07	0.14	0.22	0.26	0.27	0.28	0.28
				GRP-RNA				
CSI	0.97	0.89	0.78	0.65	0.60	0.56	0.54	0.53
$CSI'(0.8)$	0.97	0.90	0.80	0.67	0.62	0.59	0.56	0.55
$CSI'(0.5)$	0.98	0.91	0.82	0.71	0.66	0.63	0.60	0.59
$CSI'(0.2)$	0.98	0.93	0.85	0.75	0.70	0.68	0.65	0.64
POD	0.98	0.94	0.87	0.78	0.73	0.71	0.68	0.68
FAR	0.01	0.05	0.11	0.18	0.21	0.23	0.24	0.25

TABLE 13.3 – Médiane des distributions (sur 846 bassins versants × 2 périodes) des critères d'alerte calculés pour les deux stratégies de mise-à-jour retenues : GRP-ARIMA et GRP-RNA.

13.2 Où ? Quand ? Comment ?

Nous avons employé quatre indices permettant de caractériser les bassins versants :

- leur superficie,
- l'autocorrélation des débits, comme indicateur de la réactivité des bassins (Chap. 2, page 31),
- les performances du modèle en mode « Simulation »,
- leur localisation.

Pour chacun des trois premiers indices, les 846 bassins versants ont été répartis en 8 sous-ensembles définis par les quantiles $i/8$ des valeurs de ces indices : chaque sous-ensemble compte le même nombre de bassins versants (à l'arrondi près). Nous avons ensuite comparé

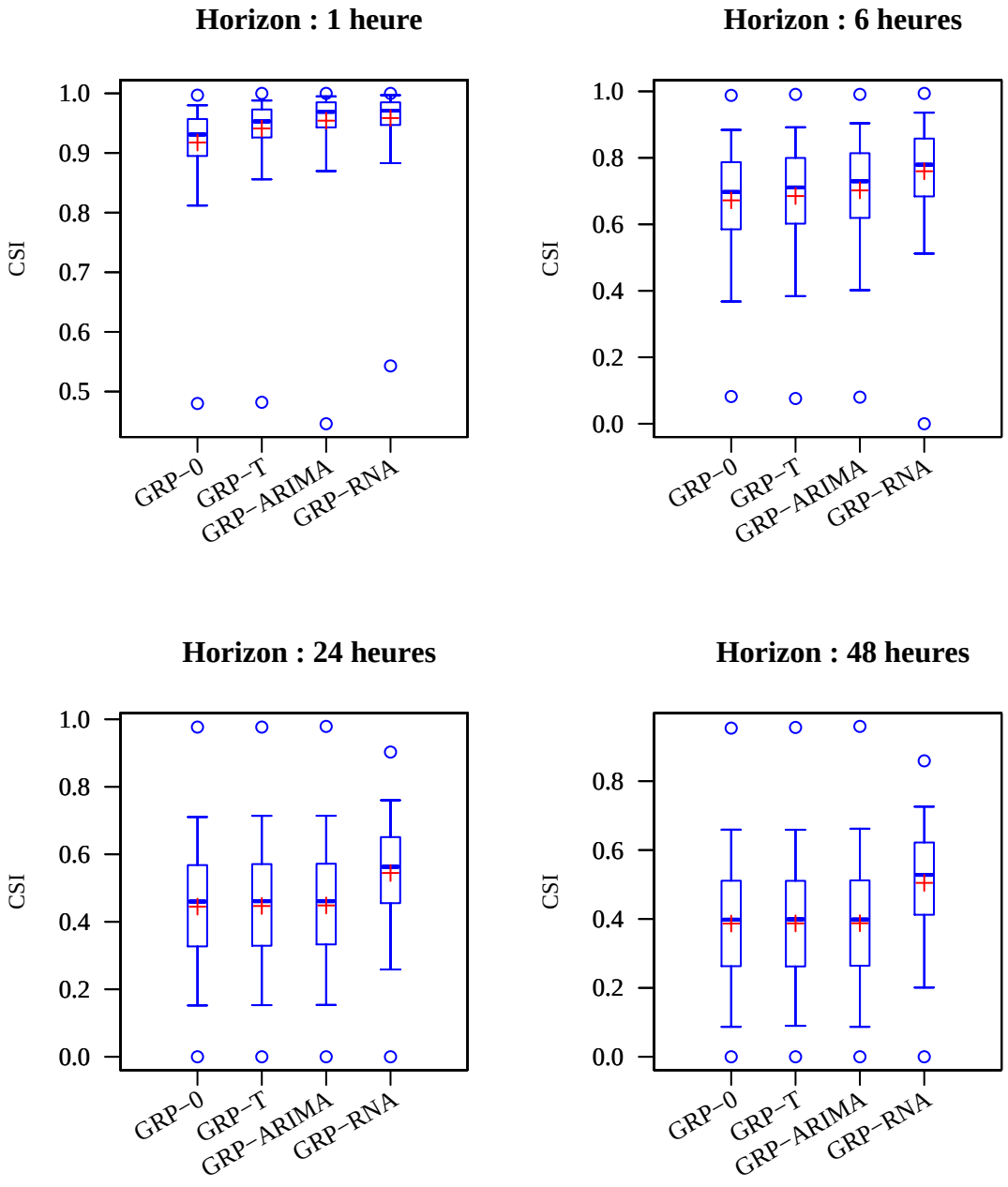


FIGURE 13.5 – Distributions (sur 846 bassins versants × 2 périodes) des valeurs de l'indice de succès critique (CSI), calculées sur 846 bassins pour les modèles GRP-0, GRP-Tangara, GRP-ARIMA et GRP-RNA.

les distributions des valeurs des différents critères de performance aux différents horizons de prévision sur ces sous-ensembles.

13.2.1 Influence de la taille des bassins versants

Critères de persistance

Les critères de performance quadratiques employés augmentent tous avec la taille du bassin (pour tous les modèles et à tous les horizons de prévision testés). Cette tendance est moins nette quand on se concentre sur les montées de crue ou sur les plus forts débits (au-dessus du quantile 0.99). Les écarts entre les différents modèles augmentent également avec la taille du bassin versant aux horizons de prévision courts (1 et 6 heures), tandis qu'ils se réduisent aux horizons les plus lointains : les corrections des erreurs en sortie du modèle sont d'autant plus efficaces que le bassin est grand. Nous n'observons pas de changement d'ordre entre les performances des différents modèles testés.

Critères temporels

Les critères R_1 et R_{1bis} sur les décalages temporels entre pics observés et prévus ne permettent pas de tirer une conclusion quelconque. Cela est probablement dû au fait qu'ils ne prennent que des valeurs entières. Le critère R_2 est un peu plus instructif et dévoile une légère tendance à diminuer quand la taille du bassin augmente.

Critères événementiels

Les indices CSI et POD augmentent quand la superficie augmente. Dans le même temps, le taux de fausses alertes diminue. Nous n'observons pas de changement d'ordre entre les performances des différents modèles testés.

L'ensemble des critères concorde : la montée de crue sera statistiquement mieux prévue sur un grand bassin que sur un petit. Les alertes y seront en probabilité meilleures. Ces résultats sont en accord avec les tests antérieurs réalisés en simulation sur le modèle GR4J, ainsi qu'avec les résultats présentés récemment par Merz *et al.* (2009). Le problème de la variabilité de la pluie, plus difficilement caractérisable sur des petits bassins versants, et le rôle intégrateur (filtre passe-bas, Oudin *et al.*, 2004) des grands bassins versants pourraient expliquer en partie ce constat.

13.2.2 Influence de la réactivité des bassins versants

Les tendances observées ici sont souvent similaires à celles liées à la superficie du bassin versant. En effet, la taille du bassin a une influence directe sur sa réactivité.

Critères de persistance

Le modèle naïf du critère de persistance est d'autant plus difficile à dépasser que le bassin versant est lent. On observe tout de même que les performances quadratiques (à travers le critère de persistance) des différents modèles s'accroissent avec l'autocorrélation du débit : plus un bassin réagit lentement, plus les modèles sont susceptibles d'y fournir de bonnes prévisions. On retrouve les mêmes tendances concernant les écarts entre les performances des modèles GRP-ARIMA et GRP-RNA. Là encore, on n'observe aucun changement dans l'ordre des performances des différents modèles. On retrouve ici un résultat cohérent avec ceux obtenus en simulation.

Critères temporels

Ici, encore les critères R_1 et R_{1bis} ne montrent aucune tendance. Le retard à la montée R_2 décroît légèrement quand l'autocorrélation augmente pour les horizons courts (inférieurs à la durée de vie des corrections des erreurs en sortie du modèle) : ces corrections sont d'autant plus efficaces que le bassin réagit lentement.

Critères événementiels

Les indices CSI et POD augmentent quand l'autocorrélation des débits augmente. Dans le même temps, le taux de fausses alertes diminue. Nous n'observons pas de changement d'ordre entre les performances des différents modèles testés. Ainsi, les alertes seront en probabilité meilleures sur bassin lent que sur un bassin plus réactif.

13.2.3 Influence de la performance du modèle en simulation

Nous employons le critère de performance C_{2M} calculé sur les débits supérieurs au quantile 0.90 (i.e., tous les pas de temps dont nous disposons) du modèle GRP fonctionnant sans mise-à-jour, comme indice de la performance du modèle en simulation.

Critères de persistance

Comme on pouvait s'y attendre, les valeurs des critères de persistance augmentent nettement avec la performance (quadratique) du modèle en simulation. Cependant, tous les modèles ne se comportent pas de façon identique : la tendance est significativement plus forte pour le modèle GRP-RNA. Ainsi la correction par réseaux de neurones fonctionne encore mieux que les autres modèles sur les bassins dont le comportement est déjà plutôt bien simulé. On notera en outre que les quelques bassins sur lesquels le modèle GRP-RNA obtient de très mauvaises performances appartiennent aux classes des bassins les moins biens simulés. Inversement, le modèle GRP-RNA obtient des valeurs de persistance nettement supérieures à celles des autres modèles (dont GRP-ARIMA) sur tous les bassins appartenant à la classe des bassins les mieux simulés (Fig. 13.6). Cet effet se renforce à mesure que l'horizon de prévision s'éloigne.

Si on s'intéresse aux plus forts débits (supérieurs au quantile 0.99), les tendances précédentes ne sont plus significatives.

Critères temporels
Les retards des modèles de prévision diminuent avec l'augmentation de la performance du modèle de simulation.

Critères événementiels
L'indice de succès critique augmente significativement avec le niveau de performance en simulation. Il en va de même pour POD. Parallèlement, on constate une décroissance du taux de fausses alertes. La différence des valeurs de CSI obtenues par les différents modèles mis-à-jour décroit, sauf pour GRP-RNA qui reste nettement en tête ; ce n'est pas le cas pour FAR et POD. Globalement comme on pouvait s'y attendre, les alertes sont de meilleures qualités quand le modèle en simulation est déjà bon.

13.2.4 Influence de la localisation des bassins versants

Pour cette analyse, nous avons représenté les valeurs des différents critères de performance sur une carte de France et comparé les différents modèles. Les cartes sont données en annexe P (page 570).

Critères de persistance
Nous retrouvons des résultats assez proches de ceux établis par Le Moine (2008) en simulation. Les régions où le modèle de prévision est statistiquement le meilleur sont l'Ouest de la France et le Nord-Est. Le modèle est moins bon sur le pourtour méditerranéen (en incluant les Cévennes), ainsi que sur le bassin de la Somme (les crues de 2001 n'y sont peut-être pas pour rien) et le Nord du bassin de la Seine pour des horizons supérieurs à 24 heures. Le modèle GRP-RNA est particulièrement meilleur que le modèle GRP-ARIMA dans les zones montagneuses tandis que le second l'emporte sur le premier sur le bassin de la Somme et dans les Pyrénées orientales.

Critères temporels
Nous n'avons employé que le critère de retard R_2 et le critère d'avance de prévision. Les retards observés se répartissent de façon relativement uniforme sur la France.

Critères événementiels
Les résultats concordent avec les observations concernant les critères quadratiques. Les critères POD, CSI et FAR conduisent aux mêmes conclusions. Les zones où les alertes sont les mieux données sont l'Ouest de la France (Bretagne) et l'Est (Alsace, Lorraine, bassin de la Saône). Les alertes sont les plus difficiles à réaliser dans les zones montagneuses. Le

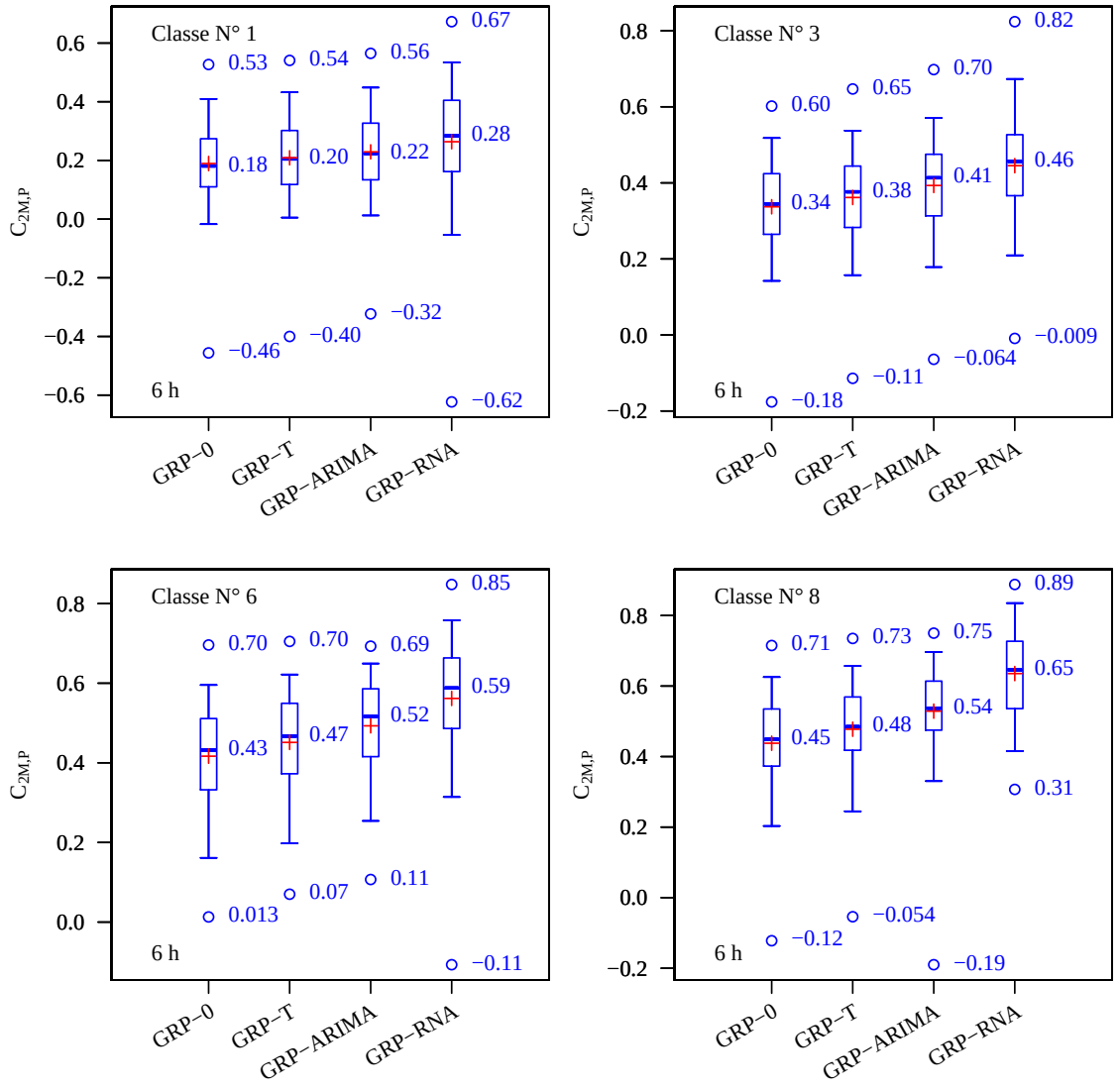


FIGURE 13.6 – Exemple de regroupement des bassins versants en classes en fonction d'un facteur potentiellement explicatif des performances des différents modèles. Ici, l'ensemble de 846 bassins a été scindé en 8 classes de même effectif selon les performances ($C_{2M,P}$) du modèle en simulation. Les distributions des valeurs du critère $C_{2M,P}$ calculé pour tous les débits supérieurs au quantile 0.90 à un horizon de prévision de 12 heures, sont données pour les quatre modèles sur les classes N° 1 (bassins pour lesquels le critère en simulation obtient ses plus mauvaises valeurs), 3, 6 et 8 (bassins pour lesquels le critère en simulation obtient ses meilleures valeurs).

modèle GRP-RNA y est cependant nettement meilleur que les autres modèles. *A contrario*, le modèle GRP-ARIMA est meilleur sur le bassin de la Seine (pour la probabilité de détection POD).

13.3 Évaluations complémentaires

Nous avons présenté jusqu'ici des résultats généraux sur les performances du modèle dans de très bonnes conditions de prévision. Voyons maintenant l'influence du scénario de pluies futures puis les performances du modèle en extrapolation.

13.3.1 Avec un scénario de pluies futures nulles

Nous ne disposons pas de données de prévision quantitative de précipitation en quantités suffisantes pour réaliser un test général les employant (Ann. M, page 535). Nous nous contentons ici de tester le modèle avec un scénario peu avantageux pour le modèle sur les 846 bassins versants : les pluies futures seront supposées nulles. Le chapitre 15 sera l'occasion de tester le modèle confronté à des données de prévision de précipitation bruitées.

Les performances des quatre variantes GRP-0, GRP-T, GRP-ARIMA et GRP-RNA, évaluées à travers le critère $C_{2M,P}$ et le critère de retard R_2, ont été comparées (Tab. 13.4). Nous constatons d'abord que le changement de scénario des pluies futures ne modifie pas l'ordre des performances des différentes variantes testées.

Critères	1 h	3 h	6 h	12 h	24 h	48 h
Critère quadratique $C_{2M,P}$						
GRP-0	0.24	0.29	0.31	0.26	0.19	0.15
GRP-T	0.53	0.37	0.33	0.26	0.19	0.15
GRP-ARIMA	0.59	0.46	0.37	0.26	0.17	0.14
GRP-RNA	0.66	0.53	0.44	0.34	0.29	0.31
Critère de retard R_2 (heures)						
GRP-0	0.4	1.2	2.6	5.6	12.8	22.4
GRP-T	-0.1	-1.0	-2.5	-6.0	-13.6	-22.8
GRP-ARIMA	0.1	-0.7	-2.4	-6.0	-13.9	-23.5
GRP-RNA	-0.0	-0.2	-0.6	-1.3	-2.0	-2.2

TABLE 13.4 – Médiane des distributions (sur 846 bassins versants × 2 périodes) des critères $C_{2M,P}$ et R_2 obtenus avec le modèle GRP et ses quatre variantes de correction des sorties.

La comparaison de ces résultats aux tableaux 13.1 et 13.2 montre bien que l'importance du scénario de pluies futures croît avec l'horizon de prévision : la baisse de $C_{2M,P}$ est quasi-négligeable à l'horizon d'une heure mais de presque de quarante points à 48 heures.

L'annonce du pic de crue est évidemment faite très en avance quand le scénario de pluies futures est nul sauf pour le modele GRP-RNA [1].

13.3.2 Extrapolation

La comparaison des performances du modèle sur les périodes P_i où il est en extrapolation (le débit maximal de la période de contrôle $\underset{P_i}{max}(Q)$ dépasse le débit maximal de la période de calage $\underset{P_{3-i}}{max}(Q)$) et des performances globales sur l'ensemble des données ne permet pas de montrer une faiblesse en extrapolation : les performances sur les périodes contenant des passages en extrapolation peuvent aussi bien être supérieures à celles en calage que l'inverse.

Pour affiner ce résultat, nous avons testé plusieurs indices du degré d'extrapolation de la période de contrôle par rapport à la période de calage. Certains sont basés sur une comparaison des débits sur les périodes de calage et de contrôle, d'autres sur les périodes de retour estimées des maxima sur ces périodes. Nous avons sélectionné l'indice suivant :

$$I_{Ex} = \frac{1}{N_{\text{Contrôle}}} \sum_{t=1}^{N} \frac{max\left(0, Q_t - Q^M_{\text{Calage}}\right)}{Q^M_{\text{Calage}}}$$

où N est le nombre de pas de temps de la période de contrôle et Q^M_{Calage}, le débit maximal sur la période de calage. Nous avons recherché en vain une relation entre cet indice et les critères de persistance et de retard.

Cette apparente indifférence à l'extrapolation peut s'expliquer par la nature non totalement empirique du modèle, ainsi que par le fait que la correction apportée aux sorties ne se base pas tant sur les débits que sur les erreurs sur les débits. Notons toutefois que Imrie *et al.* (2000) et Cigizoglu (2003) ont montré que des réseaux de neurones artificiels appliqués à la prévision Débit – Débit étaient capables d'extrapoler. Enfin, notre indicateur d'extrapolation peut tout simplement ne pas être pertinent et ne pas nous permettre de tirer des conclusions fermes.

13

13.4 Pertinence de la réduction de l'échantillon de bassins versants

La problématique que nous abordons dans la dernière partie de ce chapitre est bien différente de ce qui précède. Nous nous intéressons maintenant aux conséquences de l'échantillonnage des bassins versants sur l'évaluation des performances des différents modèles testés.

1. Ce résultat est très étonnant et il est à considérer avec suspicion. Malgré une vérification de notre code, nous n'avons pas trouvé d'erreurs qui pourraient expliquer ces valeurs.

Les 178 bassins versants employés pour établir le modèle de prévision au pas de temps horaire, ont été sélectionnés pour représenter la diversité hydroclimatique de l'ensemble d'un millier de bassins versants (Chap. 4). Dans cette dernière partie, nous souhaitons revenir sur cette procédure pour évaluer la différence entre les distributions des performance obtenues par les différents modèles sur l'ensemble total de bassins versants et sur l'échantillon de 178 bassins.

Nous avons sélectionné quatre critères de performance : le critère quadratique $C_{2M,P}$ calculé sur les débits supérieurs au quantile 0.90, le critère de retard à la montée R_2, l'indice de succès critique (CSI) et le taux de fausses alertes (FAR). Pour chacun de ces critères et pour chaque modèle testé, nous avons comparé les distributions des valeurs obtenues sur l'ensemble des bassins versants et sur l'échantillon de 178 bassins.

Critère de persistance

Lors du passage de 1024 à 178 bassins versants, le cœur de la distribution des valeurs du critère $C_{2M,P}$ (les valeurs comprises entre les quantiles 0.05 et 0.95) est assez bien conservé. On note cependant que de façon générale (pour tous les modèles et pour tous les horizons de prévision), la distribution calculée sur 178 bassins versants est légèrement moins élevée que celle calculée sur 1024 bassins (la différence de l'ordre de 4 à 5 points de $C_{2M,P}$ est significative). Cette observation est en accord avec les remarques précédentes (Chap. 4) : le passage de 1024 à 178 bassins versants rééquilibre l'échantillon et compte relativement moins de bassins versants du Nord-Est, or nous avons précédemment que le modèle obtient le plus souvent de bonnes performances sur ces bassins.

Enfin, les valeurs extrêmes de la distribution sont plutôt bien conservées. C'est notamment le cas pour le modèle GRP-RNA : les quelques bassins qui conduisent à des performances catastrophiques avaient une probabilité non négligeable d'être éliminés (Fig. 13.7 en haut) mais sont bien conservés dans l'échantillon de 178 bassins versants.

Critère de retard à la montée de crue

Le cœur de la distribution calculée sur 1024 bassins est très bien reproduit par l'échantillon de 178 bassins. Par contre, les valeurs extrêmes sont très mal reproduites pour certains horizons de prévision (un exemple en Fig. 13.7 en bas). À la différence du critère précédent, la sélection de 178 bassins versants ne conserve pas tout le temps (loin s'en faut) les valeurs extrêmes.

Critères d'alerte

En ce qui concerne le CSI, les distributions calculées sur 178 et sur 1024 bassins sont toujours très proches. C'est également le cas pour le taux de fausses alertes, sauf pour

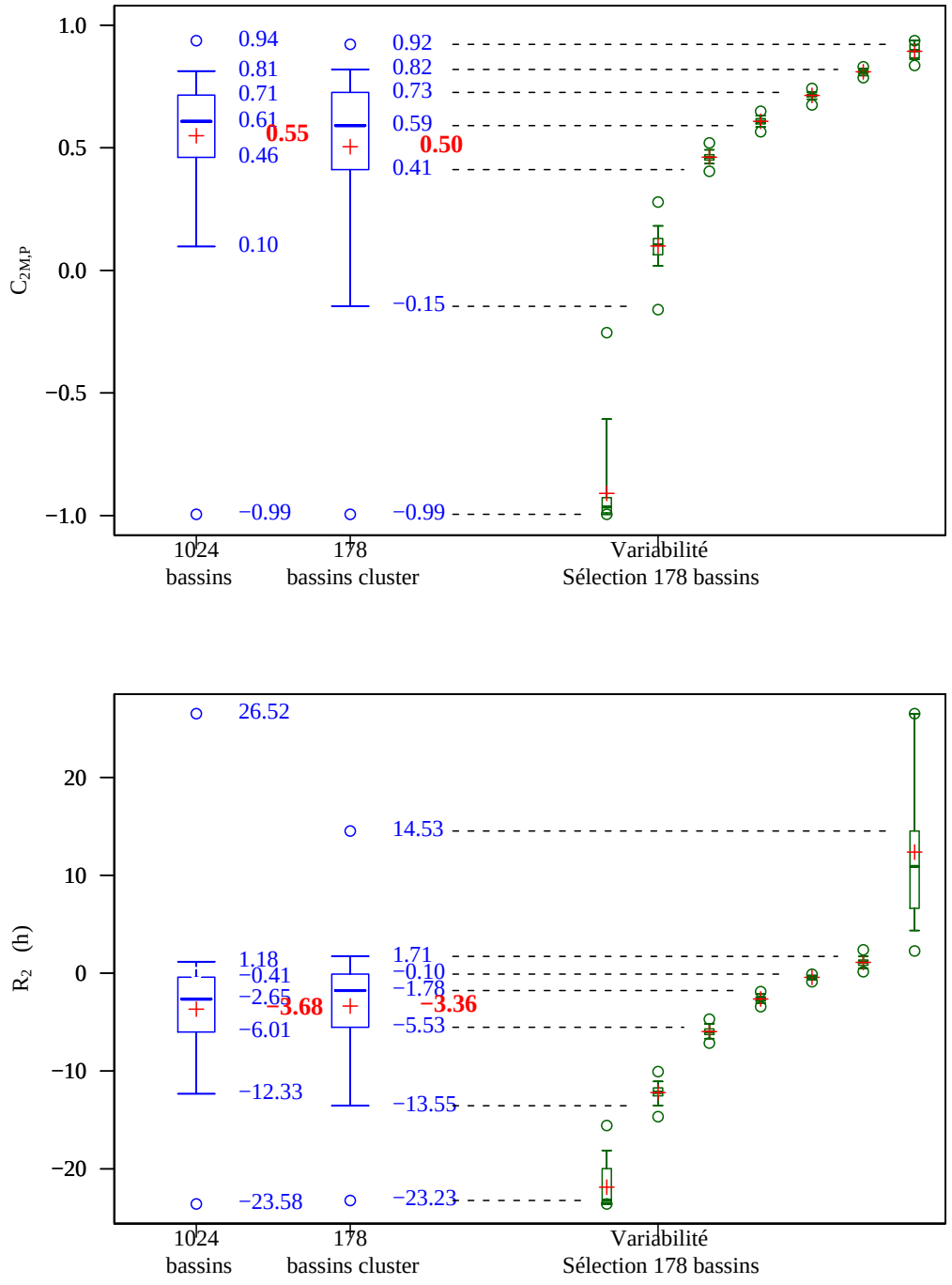


FIGURE 13.7 – Exemples de comparaison des distributions des valeurs de critères de performance sur l'ensemble de 1024 bassins et sur l'échantillon de 178 bassins. Les boîtes à moustaches vertes (droite) indiquent la variabilité attendue sur la distribution de 178 bassins tirés aléatoirement parmi 1024. En haut, le critère de performance est le critère $C_{2M,P}$, le modèle employé ici est GRP-RNA et l'horizon est de 24 heures. On observe que la probabilité d'obtenir une valeur minimale sensiblement plus élevée que la valeur minimale sur 1024 bassins est très forte. En bas, nous nous intéressons au critère de retard R_2 pour le modèle GRP-ARIMA, également pour l'horizon de prévision de 24 heures.

quelques horizons de prévision où une valeur extrême est « perdue » lors de la réduction de 1024 à 178 bassins versants.

13.5 Bilan

Les performances du modèle GRP avec trois stratégies de mise-à-jour ont été évaluées à travers une batterie de critères sur un ensemble de 846 bassins versants qui n'avaient pas été employés pour établir le modèle.

L'analyse de l'influence de plusieurs facteurs potentiellement explicatifs a montré que les prévisions (et notamment celles en montée de crue) sont statistiquement meilleures sur un grand bassin versant que sur un plus petit. Moins le bassin est réactif, plus le modèle est susceptible d'obtenir de bonnes performances. Enfin, les performances en prévision sont (positivement) corrélées aux performances en simulation.

Les valeurs du critère de persistance et des critères de retard ainsi obtenues confirment les résultats des chapitres précédents. Les autres critères de performance permettent de comparer plus finement les deux modèles GRP-RNA et GRP-ARIMA. Si le premier est globalement meilleur et se caractérise par des retards à la montée de crue nettement plus faibles, il peut aboutir à des échecs fracassants dans quelques rares cas ; le modèle GRP-ARIMA est plus robuste. Cependant, les échecs du modèle GRP-RNA sont partiellement « prévisibles » : ils sont limités à certaines zones géographiques et à des bassins pour lesquels le modèle en simulation fournit d'assez mauvais résultats.

Enfin, nous avons comparé les distributions des différents critères de performance calculées sur 1024 bassins et sur l'échantillon réduit de 178 bassins. Cet échantillon a été conçu pour représenter la variété hydroclimatique de l'ensemble des 1024 bassins versants. Nous constatons que les performances enregistrées sur cet échantillon représentent également assez bien les performances calculées sur l'ensemble des bassins.

14
Chapitre

Comparaison à d'autres stratégies de prévision Pluie – Débit

Les progrès du progrès vont de progrès en progrès.

Alexandre Vialatte, *Dernières Nouvelles de l'Homme*

Objectifs

▷ Comparer le modèle conceptuel GRP à d'autres stratégies opérationnelles de prévision.

Sommaire

14.1 PQb 259
14.1.1 Formulations et paramètres 259
14.1.2 Performances 260
14.2 Les réseaux de neurones 262
14.2.1 Mise en œuvre et choix des entrées 262
14.2.2 Résultats 263
14.2.3 Discussion 264
14.2.4 Dernier essai 266
14.3 Bilan 266

Pour compléter notre évaluation du modèle de prévision au pas de temps horaire, nous avons comparé ses performances à celles de deux autres modèles. Plus qu'une comparaison de structures de modèles hydrologiques (comme chez Perrin, 2000), nous avons cherché à tester des approches éloignées de la modélisation conceptuelle. Le premier modèle, PQb, est inclus dans le logiciel Sophie [1] employé au Schapi [2] et dans de nombreux services de prévision des crues (SPC) français. L'autre est un modèle purement empirique : les réseaux de neurones artificiels [3].

14.1 PQb

14.1.1 Formulations et paramètres

Formulation originale de Fabret

PQb est défini par la formule :

$$\widehat{Q}_{t+\Delta t|t} = \frac{\alpha - 1}{\alpha} \cdot Q\left(t\right) + \frac{S \cdot b}{\alpha \cdot \Delta t'} \cdot \Pi \tag{14.1}$$

avec

- α le coefficient de décrue (sans dimension, doit être supérieur à 1),
- Δt l'horizon de prévision (h),
- S la surface du bassin versant (km^2),
- b le coefficient de réduction des pluies ; rapport entre la pluie tombée et la pluie ruisselée : sa valeur est donc comprise entre 0 et 1 (sans dimension),
- Π la pluie tombée entre $t - \Delta t'$ et t (mm).

Le modèle comporte donc trois paramètres à ajuster : α, b et $\Delta t'$.

Variantes

Dans la formule présentée dans le paragraphe précédent, PQb n'a connaissance que des pluies passées. Nous avons également testé une variante permettant de placer ce modèle dans les mêmes conditions que GRP. Cette variante permet de tenir compte des pluies futures :

$$\widehat{Q}_{t+\Delta t|t} = \frac{\alpha - 1}{\alpha} \cdot Q\left(t\right) + \frac{S \cdot b}{\alpha \cdot \Delta t'} \cdot \Pi' \tag{14.2}$$

où Π' est la pluie tombée entre $t + \Delta t - \Delta t'$ et $t + \Delta t$.

1. Système Ouvert de Prévisions Hydrologiques Informatisé avec Expertise.
2. Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations.
3. Les réseaux de neurones artificiels ont été employés pour corriger les sorties du modèle GRP dans le chapitre 10. Ici, nous les utilisons pour réaliser la prévision Pluie – Débit complète.

Enfin, une variante auto-régressive sera testée :

$$\begin{cases} \widehat{Q}_{t+1|t} = \dfrac{\alpha-1}{\alpha} \cdot Q\left(t\right) + \dfrac{S \cdot b}{\alpha \cdot \Delta t'} \cdot \Pi_t \\ \widehat{Q}_{t+1+i|t} = \dfrac{\alpha-1}{\alpha} \cdot \widehat{Q}_{t+i|t} + \dfrac{S \cdot b}{\alpha \cdot \Delta t'} \cdot \Pi_{t+i} \text{ pour } i \text{ dans } [\![1, H-1]\!] \end{cases} \tag{14.3}$$

14.1.2 Performances

Les différentes versions présentées ci-dessus ont été testées sur notre échantillon de 178 bassins. Les essais ont été menés pour les horizons de prévision de 1, 3, 6, 12, 18, 24, 36 et 48 heures. Les performances ont été comparées, bassin par bassin aux performances de GRP. Les distributions des paramètres ont été examinées pour s'assurer que PQb fonctionnait correctement au pas de temps horaire. Les essais ont été menés avec deux scénarios de pluie future : les pluies observées *a posteriori* et des pluies futures supposées nulles.

Formulation originale

GRP présente des performances nettement supérieures à PQb pour tous les horizons de prévision, même quand PQb utilise des pluies futures connues (Fig. 14.1). En moyenne, les performances sont supérieures d'environ 20 points de $C_{2M,P}$ pour GRP, ce qui est très largement significatif. Cet écart se réduit néanmoins quelque peu quand l'horizon de prévision s'accroît.

Au calage, les paramètres de PQb n'atteignent que rarement les limites de leurs gammes autorisées (critère nécessaire mais non suffisant pour affirmer que le calage fonctionne correctement).

Une représentation des performances des deux modèles PQb et GRP sur une carte de France (Fig. 14.2) ne permet pas de mettre en évidence des zones géographiques où PQb serait (statistiquement) meilleur que GRP. Les bassins pour lesquels PQb est meilleur sont pour la plupart des petits bassins. Les performances de GRP y sont le plus souvent modestes, voire médiocres.

Performances des variantes

PQb profite grandement de l'information contenue dans les données de pluie future (logiquement d'autant plus que l'horizon de prévision est grand).

La variante auto-régressive ne permet pas d'améliorer les performances de PQb, quel que soit l'horizon de prévision.

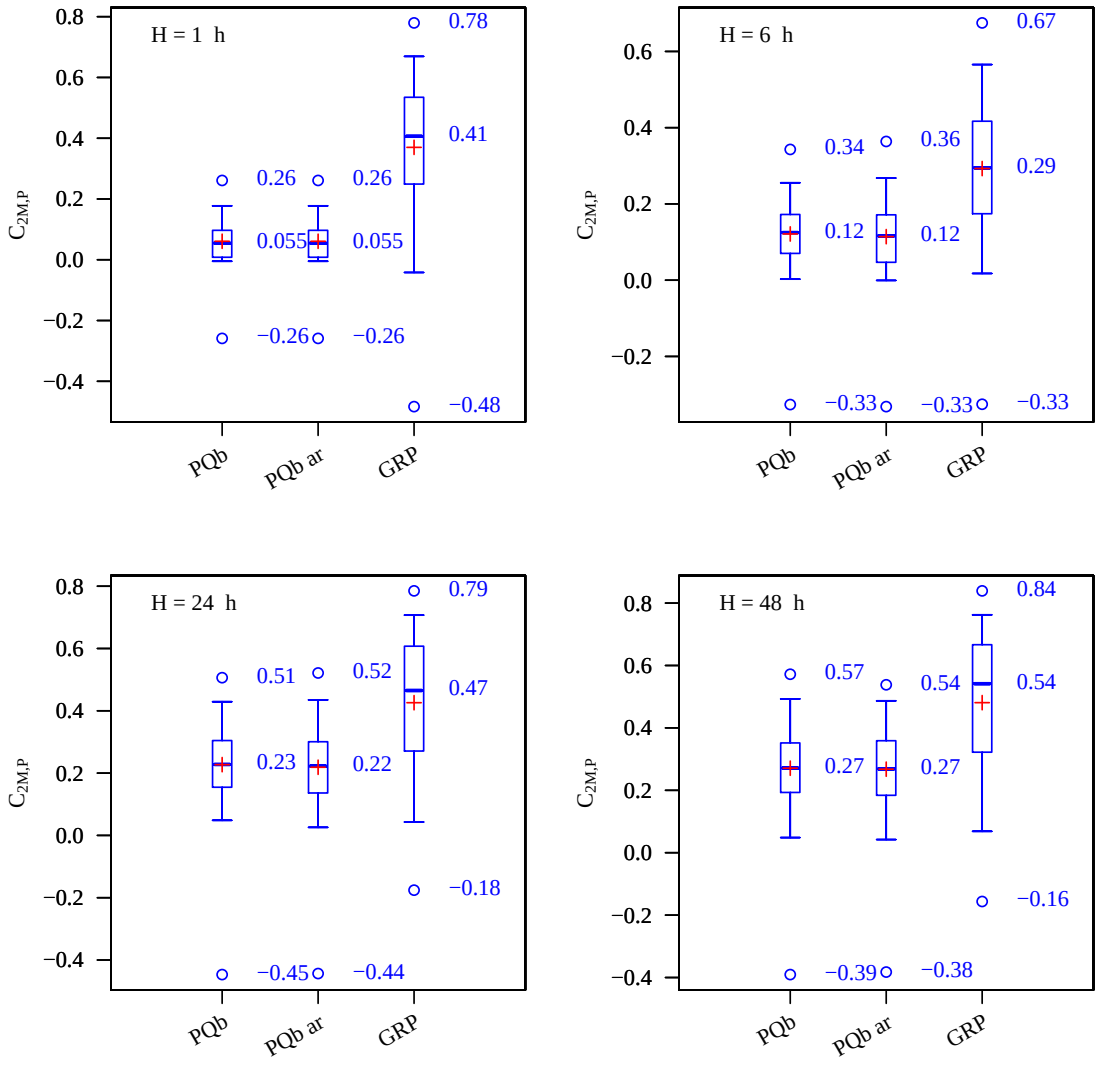


FIGURE 14.1 – Distributions des performances (mesurées par le critère $C_{2M,P}$ sur 178 bassins versants × 2 périodes) du modèle PQb dans sa formulation originale et dans la variante auto-régressive (PQb ar), ainsi que du modèle GRP (avec les mises-à-jour définies par Tangara, 2005), aux horizons de prévision 1, 6, 24 et 48 heures. Le scénario de pluies futures employé ici est celui constitué des observations de précipitation faites *a posteriori*.

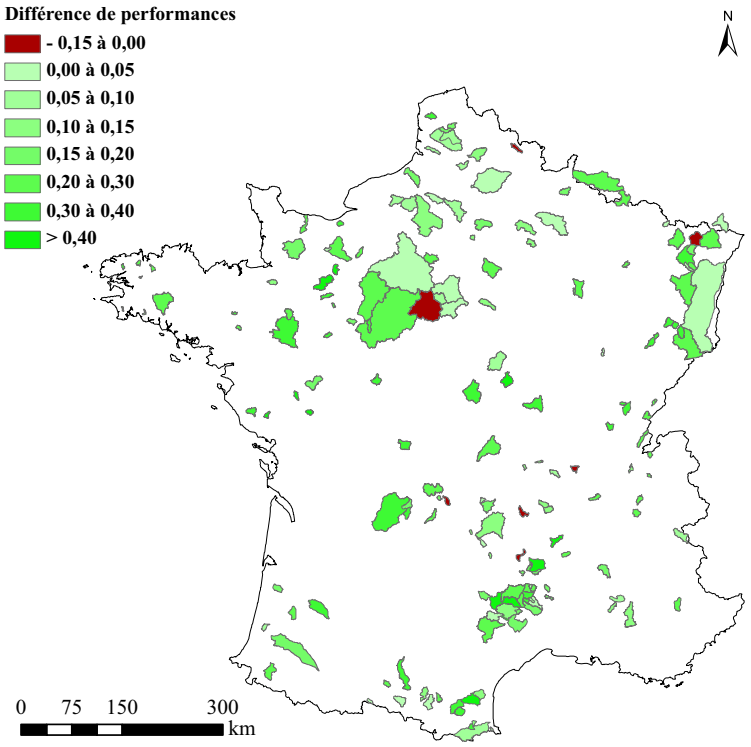


FIGURE 14.2 – Comparaison des performances (mesurées à l'aide du critère $C_{2M,P}$) de PQb et de GRP pour un horizon de 6 heures sur notre échantillon de 178 bassins. Les bassins pour lesquels PQb est meilleur que GRP sont colorés en rouge.

14.2 Les réseaux de neurones

Nous avons montré aux chapitres 10 et 13 que les réseaux de neurones artificiels (RNA) sont des outils très intéressants pour corriger les erreurs en sortie du modèle GRP... « en aval » du modèle hydrologique. En inversant la perspective, nous pouvons nous interroger sur la nécessité d'un modèle conceptuel « en amont » du réseau de neurones : ce dernier est-il capable seul d'utiliser les entrées de la combinaison formée par le modèle conceptuel et le réseau de neurones pour obtenir des performances aussi bonnes ? Autrement dit, la conceptualisation apportée par le modèle hydrologique est-elle nécessaire au bon fonctionnement de la combinaison ? De nombreux auteurs ont testé les réseaux de neurones utilisés seuls pour la prévision hydrologique avec des succès très divers (e.g. Anctil et Rat, 2005, Toth et Brath, 2007).

14.2.1 Mise en œuvre et choix des entrées

Pour répondre à cette question, nous nous appuyons sur les résultats du chapitre 10 et sur la littérature explorée à cette occasion. Nous nous basons sur un perceptron(g) à une seule

couche cachée. Les réglages du RNA sont ceux choisis lors du chapitre 10.

Campolo *et al.* (1999) ont montré l'importance des dernières données de débit. Pour choisir les entrées du modèle, nous avons testé la méthode des plus proches voisins pour la prévision Pluie – Débit (résumée en annexe Q, page 601) dont le calage nous a permis d'évaluer l'importance relative des différentes données de pluie et de débit. Il est apparu (comme on pouvait s'y attendre) que les toutes dernières observations de débit sont les données les plus indispensables à la prévision. Les données de pluie sont moins utiles mais leur utilité augmente avec l'horizon de prévision. Nous avons constaté qu'il était pertinent d'employer des données de pluie sur une plus vaste plage temporelle que celle des données de débit utiles. Enfin, les situations peuvent se révéler très différentes d'un bassin à l'autre. Aussi, nous allons tester plusieurs configurations de réseaux de neurones artificiels :

(a) Le réseau le plus simple compte trois entrées. Toutes sont des observations de débit : Q_t, Q_{t-1} et Q_{t-2}. La couche cachée est composée de trois neurones.

(b) Le deuxième réseau introduit des observations de pluie et un scénario de pluies futures. Ses quatre entrées sont Q_t, $Q_{t-\Delta t}$, le cumul des pluies entre $t-\Delta t$ et l'instant de prévision t pour Δt valant 1, 3, 6, 12 et 24 heures, et le cumul des pluies futures (de $t+1$ à $t+H$). Il y a trois neurones dans la couche cachée.

(c) Le troisième réseau est plus complexe : il distingue une histoire récente (entre $t-\Delta t_1$ et t) et une histoire plus lointaine du bassin (entre $t-\Delta t_2$ et $t-\Delta t_1$). Ses six entrées sont Q_t, $Q_{t-\Delta t_1}$, le cumul des pluies entre $t-\Delta t_1$ et t, $Q_{t-\Delta t_2}$, le cumul des pluies entre $t-\Delta t_2$ et $t-\Delta t_1$ avec les valeurs de Δt_1 et Δt_2 répertoriées dans le tableau 14.1, et le cumul des pluies futures (de $t+1$ à $t+H$). La couche cachée compte trois neurones.

Au total, nous testons donc 14 réseaux sur chacun de nos bassins.

Δt_1	1	3	1	3	6	3	6	12
Δt_2	6	6	12	12	12	24	24	24

Table 14.1 – Valeurs des paramètres Δt_1 et Δt_2 de la troisième série d'essais de réseaux de neurones employés comme modèles de prévision. Les valeurs sont données en heures.

14.2.2 Résultats

Même si pour chaque horizon, il y a quelques bassins pour lesquels les réseaux de neurones sont nettement meilleurs que le modèle GR3P, le modèle conceptuel se révèle nettement plus sûr (distribution des performances beaucoup plus resserrée et ne présentant pas d'échec complet) et bien meilleur que les réseaux de neurones pour la très grande majorité des bassins versants (pour plus de 90 % d'entre eux). Aux horizons de prévision les plus courts,

les réseaux de neurones ne sont guère meilleurs que le modèle naïf pour une petite moitié des bassins versants.

L'étude des séries d'essais (b) et (c) montrent que le réseau de neurones obtient de meilleurs résultats quand on donne au réseau des informations sur les tout derniers débits. Les réseaux n'ont globalement pas su prendre en compte l'information complémentaire apportée par les cumuls de pluie (telle que nous la leur fournissons ici).
Rajouter deux entrées pour simuler la connaissance de données de flux « rapides » (entre $t - \Delta t_1$ et t) et de flux « lents » (entre $t - \Delta t_2$ et $t - \Delta t_1$) n'améliore pas les résultats sur les bassins pour lesquels les réseaux de neurones étaient le moins mauvais.

Utiliser l'information des trois derniers débits est la meilleure option étudiée pour l'horizon de prévision le plus court mais naturellement plus l'horizon de prévision s'éloigne, moins ce réseau de neurones est performant. Ce réseau compte néanmoins parmi les réseaux testés les plus performants.

Une analyse des performances des RNA en fonction de l'autocorrélation des débits (indice de la réactivité des bassins versants) n'a pas permis de circonscrire un sous-groupe de bassins obtenant de meilleures performances : que ce soit sur un bassin réagissant rapidement ou sur un bassin lent, les RNA obtiennent statistiquement des performances inférieures à celles du modèle GRP.

14.2.3 Discussion

Les performances présentées ici sont le plus souvent très faibles. Cela peut provenir du fait que nous avons travaillé de façon générique sur un grand nombre de bassins versants (même schéma des données d'entrée, mêmes constructions des réseaux). Cette étude sur la prévision directe par réseau de neurones est de plus sommaire (en comparaison aux préconisations de Maier et Dandy, 2000). Cependant, même en sélectionnant pour chaque bassin, le meilleur réseau, les performances restent faibles. Utiliser un réseau de neurones comptant peu d'entrées (et de neurones cachés) pour la prévision hydrologique au pas de temps horaire [4] en prenant en entrée des pluies et des débits semble donc être une approche peu pertinente. Cette tâche est peut-être trop complexe et exigerait alors des réseaux plus complexes. Ce qui signifie également plus de difficultés pour définir la bonne structure et les bonnes entrées (avec des combinaisons différentes pour différents bassins), mais implique aussi un calage plus délicat à réaliser, un temps de calcul nettement plus long... Toutes choses qui n'encouragent pas à utiliser cette méthode dans des conditions opérationnelles.

Ces résultats sont en contradiction avec ceux de Toth et Brath (2007), obtenus sur deux bassins au pas de temps horaire pour des horizons de prévision d'une à douze heures, qui

4. Les résultats sont bien différents au pas de temps journalier (voir par exemple Anctil et Rat, 2005). Le Pr. Anctil m'a indiqué dans une communication personnelle qu'il considérait la prévision hydrologique au pas de temps horaire avec un réseau de neurones très difficile.

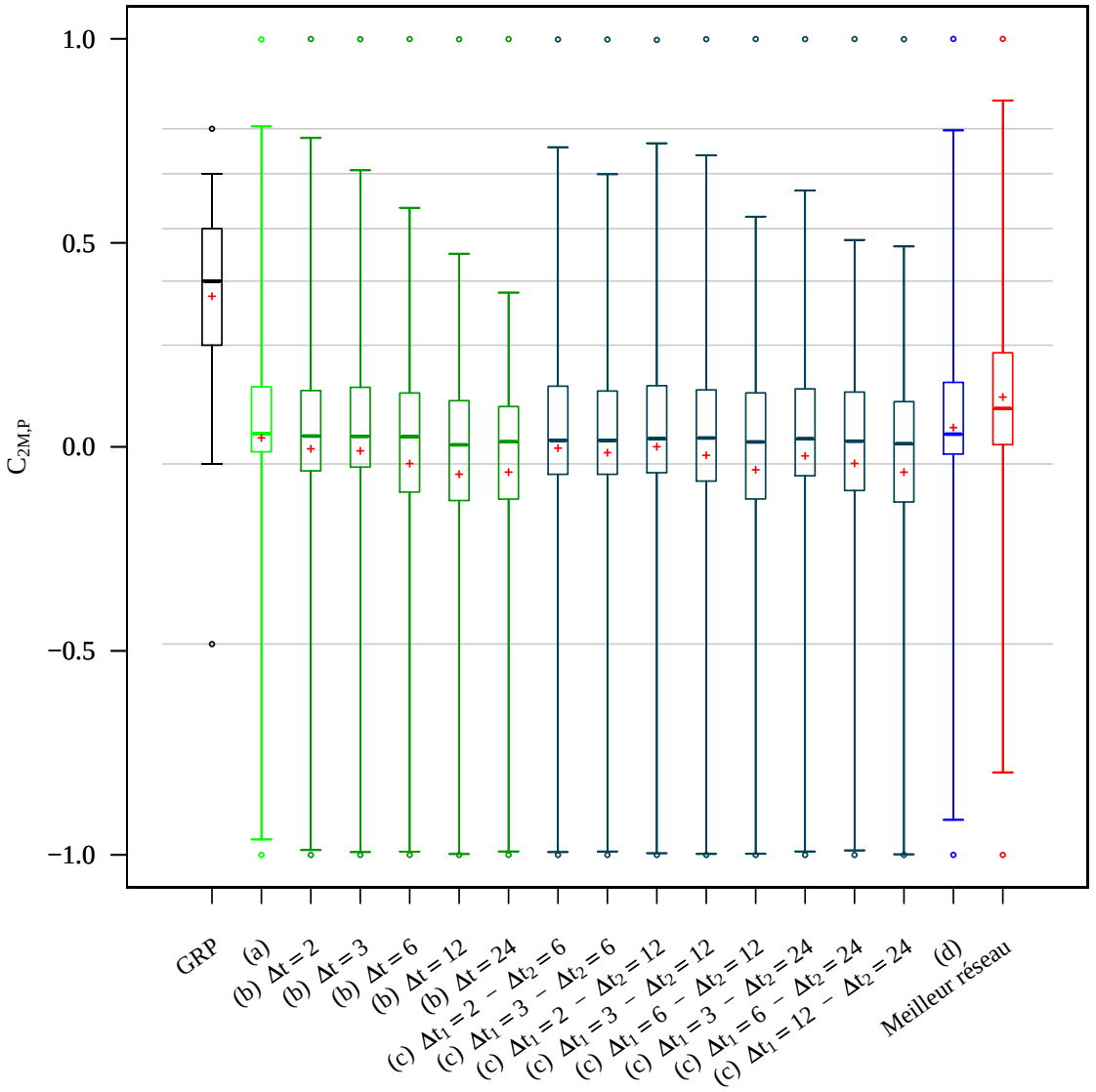


Figure 14.3 – Distributions des performances ($C_{2M,P}$) calculées sur 178 bassins versants × 2 périodes du modèle GRP (avec les mises-à-jour définies par Tangara, 2005) et de quinze réseaux de neurones artificiels. Les quatorze premiers RNA ((a) à (c)) sont ceux présentés en page 263. Le quinzième (d) correspond au dernier essai introduit page 266. La dernière boîte à moustache correspond à la distribution où on aurait choisit pour chaque bassin, le RNA donnant les meilleures performances (en contrôle). Les deux configurations les plus fréquemment sélectionnées sont (a) (21 %) et (d) (13 %). L'horizon de prévision est ici de 6 heures.

montraient que si la combinaison du modèle conceptuel et d'une correction des erreurs par réseau de neurones était la plus efficace [5], le réseau de neurones employé seul pour la prévision des débits s'en tirait très honorablement (mieux que le modèle conceptuel avec une correction basique, le report d'erreur).

Pourquoi nos réseaux de neurones présentent-ils de tels échecs quand ils sont employés directement en prévision Pluie – Débit ? Quelques explications peuvent être avancées :

1. Le nombre de neurones cachés est trop faible par rapport à la complexité de la tâche demandée : ce nombre de trois neurones qui était suffisant au pas de temps journalier ne le serait plus au pas de temps horaire ?
2. Il serait plus facile de chercher à prédire les différences de débits plutôt que les débits eux-mêmes (Anctil et Rat, 2005) car ils présentent une distribution plus symétrique et non bornée à 0 (un peu comme dans le cas de la prévision d'erreurs).
3. Le choix des entrées n'est pas pertinent : utiliser des cumuls de pluie introduit des relations non univoques entre les entrées et les pluies observées (en plus des relations non univoques entre séries de pluie et séries de débits...).
4. Enfin, il est possible qu'on ait besoin d'une entrée comme l'état du bassin (décrit par exemple par les sorties de la fonction de production du modèle GRP), mais on se redirige alors vers la combinaison d'un modèle hydrologique (certes incomplet) et d'un réseau de neurones (Anctil *et al.*, 2004a).

14.2.4 Dernier essai

Pour tester ces idées (en bloc), nous avons réalisé un dernier essai (noté (d) sur la figure 14.3) où le réseau compte 10 neurones cachés et doit déterminer les variations de débit en prenant des entrées non cumulatives : les trois derniers débits et pluies observés et les pluies de $t+H-5$ à $t+H$ (en éliminant les doublons pour les horizons de prévision les plus courts). Ce réseau se révèle meilleur que les autres réseaux de neurones à l'horizon d'une heure tout en restant très nettement insuffisant (comparé au modèle de persistance et encore plus au modèle GRP). Aux horizons de prévision supérieurs, il obtient des performances similaires à celles des réseaux précédents.

14.3 Bilan

Nous avons testé deux [6] méthodes de prévision très différentes de notre approche semi-conceptuelle, semi-empirique. Ces deux modèles sont PQb, outil opérationnel disponible

5. Dans une idée similaire, Lekkas (2008) montre que la combinaison d'un réseau de neurones (pour la prévision débit – débit) à un modèle empirique de correction d'erreurs (autorégressif à moyenne mobile) améliorait sensiblement les performances du premier.

6. Trois, si nous comptons la méthode des plus proches voisins (annexe Q, page 601).

pour les Services de Prévision des Crues français, et les réseaux de neurones artificiels. Les performances de ces approches sont nettement inférieures à celles du modèle GRP.

Nous n'avons certes pas exploré l'ensemble des approches empiriques, dites *data-driven* (Voir, par exemple, Khu *et al.*, 2001, Young, 2002, Solomatine *et al.*, 2008). Mais l'échec des approches testées et notamment celui des RNA nous interrogent sur la capacité de telles approches à prévoir le débit au pas de temps horaire. Est-il alors nécessaire d'incorporer un niveau minimal d'approches conceptuelles (à la manière de Anctil *et al.*, 2004a) ? L'importance croissante des mises-à-jour que nous avons mise en évidence dans la troisième partie de ce mémoire ne doit pas nous faire oublier que ces mises-à-jour s'appuient sur une structure hydrologique mi-conceptuelle, mi-empirique et que les avancées en termes de performance de notre modèle reposent non seulement sur les mises-à-jour mais également sur la structure hydrologique du modèle : peut-on mettre à jour efficacement un outil de toute façon impropre à la tâche qui lui est assignée ?

Mentionnons enfin que nous avons comparé le modèle GRP au modèle GR3H (e.g. Michel, 1991) dont la structure est beaucoup plus proche de celle de GRP, mais dont le fonctionnement événementiel rappelle ceux de PQb ou des réseaux de neurones artificiels. Cette comparaison avait avant tout un objectif opérationnel car le modèle GR3H est implémenté dans Sophie. Le modèle GRP obtient des performances significativement supérieures à celles de GR3H sur la très grande majorité des 1 070 bassins versants employés pour cette étude (Javelle, 2008).

15
Chapitre

Robustesse du modèle face à des conditions dégradées

Objectifs

- ▷ Caractériser le comportement du modèle en conditions dégradées.
- ▷ Estimer la sensibilité du modèle à différents types d'erreur dans les données.
- ▷ Comparer les impacts de ces différentes erreurs.

Sommaire

15.1 Influence de lacunes dans les observations 272
15.1.1 Lacunes de données de pluie (passées) 272
15.1.2 Lacunes de débit 274
15.1.3 Synthèse 274
15.2 Influence du bruit dans les données 275
15.2.1 Cas des observations de pluie 275
15.2.2 Sensibilité du modèle au scénario de pluies futures 277
15.2.2.1 Erreurs aléatoires dans les prévisions de pluies futures . 277
15.2.2.2 Effet de la répartition temporelle des prévisions de pluie 279
15.2.3 Sensibilité des mises-à-jour aux erreurs sur les mesures de débit . 281
15.2.3.1 Mise-à-jour complète du niveau du réservoir de routage 281
15.2.3.2 Emploi du filtre de Kálmán 283
15.2.3.3 Mise-à-jour partielle du niveau du réservoir de routage 286
15.2.3.4 Discussion 286
15.3 Bilan et discussion 288

Jusqu'ici nous avons testé le modèle dans des conditions « idéales » : avec des données validées et sans lacune. En outre, les pluies futures étaient prises égales aux données observées *a posteriori*. Cela ne signifie pas que les données étaient exemptes d'erreur, loin s'en faut, mais que leur qualité est en moyenne très nettement supérieure à celle des informations dont les services disposent en temps réel. Ceci est d'autant plus vrai que les conditions de mesure et d'acquisition de données peuvent être très dégradées. Dans ce chapitre, nous allons examiner le comportement du modèle dans des situations où la donnée est de faible qualité.

La qualité des données peut également influencer le calage des paramètres. Cet aspect n'est pas spécifique à la prévision et a été bien documenté en ce qui concerne les modèles hydrologiques de simulation (Troutman, 1985, Andréassian *et al.*, 2001, Oudin *et al.*, 2006b, Bárdossy et Singh, 2008, parmi beaucoup d'autres). D'autre part, le calage du modèle est une opération qui ne demande pas forcément à être faite en temps réel. Le plus souvent, le calage peut être réalisé avec des données validées en temps différé [1]. C'est pourquoi les erreurs dans les données qui nous paraissent les plus dangereuses sont celles qui interviennent lors du calcul de prévisions en temps réel. Nous limitons donc notre étude aux conséquences d'une mauvaise qualité des données au moment d'émettre la prévision.

Le modèle emploie deux données principales : la pluie et le débit [2]. Les deux peuvent être entachées de fortes incertitudes : erreurs systématiques (Aitken, 1973) et erreurs aléatoires. Certaines erreurs systématiques peuvent être compensées par le calage (e.g. Dawdy et Bergmann, 1969). De plus, les erreurs systématiques de débit interviennent également lors de l'évaluation du modèle, ce qui peut perturber l'estimation de leur impact sur les performances. Les erreurs les plus gênantes en temps réel pour des prévisions à court terme sont les erreur aléatoires. Nous allons explorer deux types d'erreurs aléatoires. En premier lieu, des lacunes liées à des problèmes de rapatriement ou à des appareils en panne. Les lacunes peuvent éventuellement être complétées par l'imputation de valeurs probables (Kalteh et Hjorth, 2009) mais cette opération est difficile et peut, bien sûr, engendrer des erreurs. Ensuite, des valeurs erronées qui peuvent être dues à des dysfonctionnements des instruments de mesure, des erreurs de transcription, etc.

En simulation, le modèle est surtout sensible aux erreurs sur les pluies (Paturel *et al.*, 1995, Andréassian *et al.*, 2001, Moulin *et al.*, 2009). La situation est différente en prévision car la mise-à-jour du modèle rend les erreurs sur les données de débit importantes. Dans un premier temps, nous allons étudier l'impact de lacunes dans les observations de pluie, puis de débit. La seconde partie de ce chapitre est consacrée à l'influence d'erreurs aléatoires

1. Il n'est cependant pas évident que le calage fournissant les « meilleures » prévisions en temps réel soit celui obtenu avec des données validées d'une qualité trop différente de celle des données usuellement disponibles en conditions opérationnelles.

2. L'évapotranspiration potentielle est une variable de peu d'importance en prévision des crues et peut donc être estimée par des moyennes interannuelles, valeurs toujours disponibles qui ne présentent pas d'erreurs aléatoires.

dans les données sur les performances du modèle : nous examinerons successivement les conséquences d'erreurs dans les observations des précipitations, puis dans les prévisions quantitatives de pluie et enfin dans les observations de débit servant à la mise-à-jour du modèle. Notre objectif est d'établir des ordres de grandeur des impacts et d'évaluer les importances relatives de ces erreurs.

15.1 Influence de lacunes dans les observations

15.1.1 Lacunes de données de pluie (passées)

La première expérience vise à simuler l'effet de lacunes dans les observations de pluie passée sur les performances du modèle sur nos 178 bassins versants. Pour chaque instant de prévision t de la série de données dont nous disposons, les données de pluie sur la plage $[\![t-\Delta t, t]\!]$ sont mises à zéro pour effectuer la prévision $\widehat{Q}_{t+H|t}$. Différentes longueurs de lacune Δt sont testées : 1, 3, 6, 12, 18, 24, 36, 48 et 72 heures.

Une chute significative de performances

La présence de lacunes, même courtes, fait chuter assez fortement les performances de prévision (mesurées à l'aide du critère $C_{2M,P}$) du modèle GRP (Tab. 15.1). L'effet est logiquement d'autant plus fort que l'horizon de prévision est court (les prévisions de pluie futures ne sont pas affectées). Les performances du modèle de prévisions restent cependant supérieures à celles du modèle de persistance sur une grande majorité des bassins.

	Durées des lacunes (h)						
Horizon	0	1	6	12	24	48	72
1 heure	0.41	0.40	0.30	0.28	0.27	0.26	0.26
6 heures	0.29	0.29	0.18	0.12	0.09	0.09	0.08
12 heures	0.37	0.37	0.30	0.24	0.21	0.21	0.21
24 heures	0.47	0.46	0.43	0.39	0.38	0.37	0.36
48 heures	0.54	0.55	0.54	0.52	0.51	0.50	0.49

TABLE 15.1 – Médiane (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP (avec les mises-à-jour de Tangara, 2005) sans lacune dans les observations de pluie (première colonne) puis avec des lacunes dans les observations de pluie précédant immédiatement l'instant de prévision.

Une augmentation mesurée de l'erreur quadratique moyenne

L'erreur quadratique moyenne du modèle GRP augmente avec la durée des lacunes de pluie, de façon cohérente avec ce qui précède (Fig. 15.1). On observe trois phases. Dans un premier temps pour des lacunes très courtes (de l'ordre d'une heure), l'erreur augmente peu : le modèle n'a pas eu le temps d'être gêné par la lacune de pluie. Puis la détérioration des

prévisions est assez rapide (pour des durées de lacune variant de 1 à 12 heures). Au-delà, la croissance de l'erreur quadratique s'assagit : le modèle devient insensible aux lacunes trop lointaines comme on pouvait s'y attendre (à cause de la mise-à-jour du modèle intervenant à l'instant de prévision).

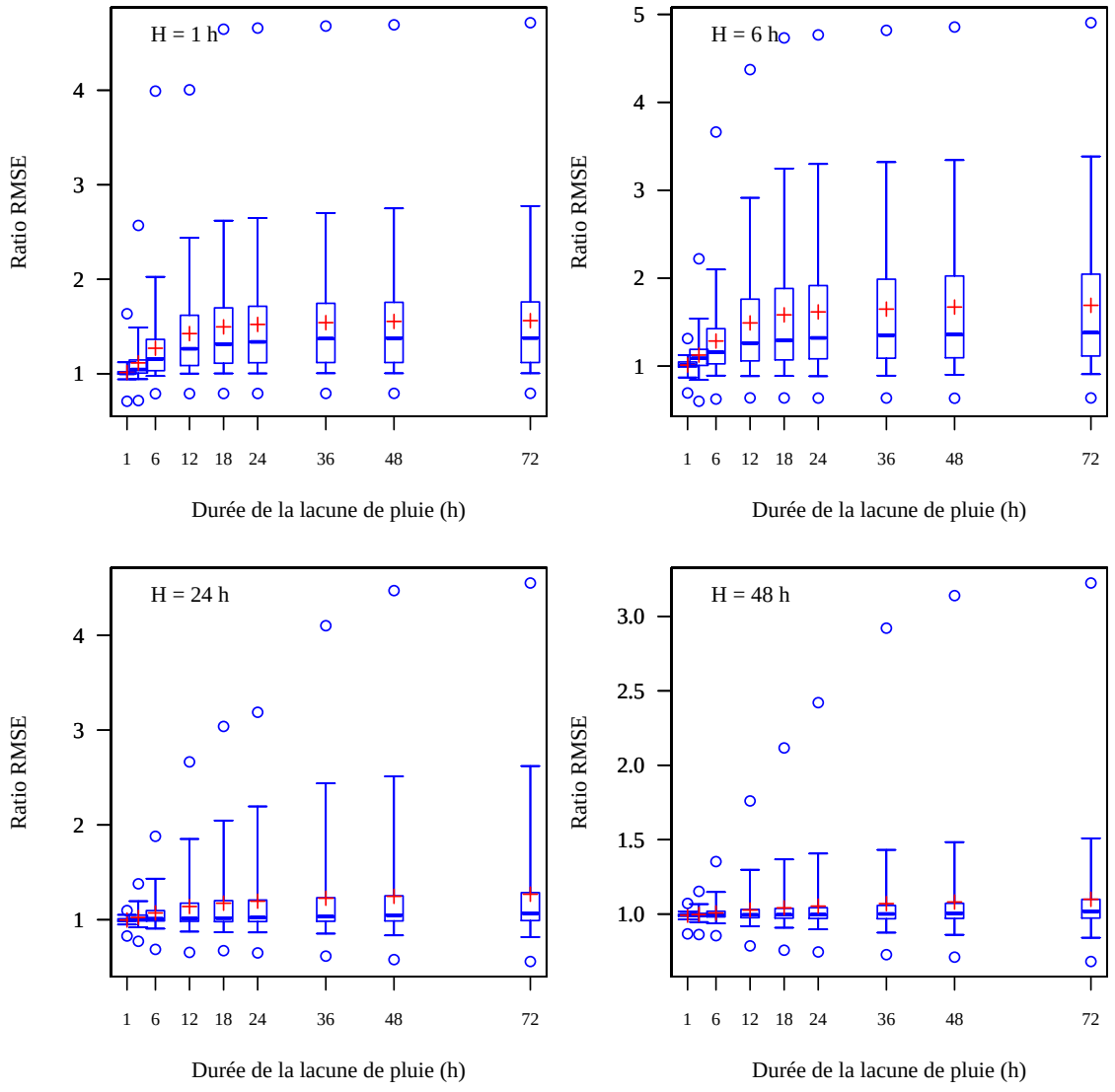


Figure 15.1 – Distributions (sur 178 bassins versants × 2 périodes) de l'erreur quadratique moyenne du modèle GRP (avec les mises-à-jour de Tangara, 2005) avec des lacunes dans les observations de pluie précédant immédiatement l'instant de prévision, normalisée par l'erreur quadratique moyenne du même modèle fonctionnant sans lacune de pluie.

Influence de la réactivité des bassins versants

La détérioration des performances due à des lacunes de pluie est nettement contrôlée par la réactivité du bassin versant : plus le bassin réagit vite, moins il sera sensible à une lacune dans les pluies passées (à la double condition que le modèle soit mis-à-jour grâce à l'observation de débit et que nous disposions des prévisions de pluie future). Ainsi, à un horizon de prévision de 6 heures, la chute des performances du modèle GRP due à l'absence des données de pluie sur les 72 dernières heures était environ de 9 points de $C_{2M,P}$ (en médiane) pour les 45 bassins versants les plus réactifs (le quart de notre échantillon présentant la plus faible autocorrélation des débits) tandis qu'elle était de 18 points sur les 45 bassins les moins réactifs (présentant l'autocorrélation des débits la plus élevée).

15.1.2 Lacunes de débit

L'impact de l'absence de données de débit pour mettre à jour le modèle a fait l'objet d'une étude présentée lors de l'assemblée générale de l'Association internationale des Sciences hydrologiques (Berthet *et al.*, 2009b, reproduit en annexe N, page 541). Une lacune de débit sur N pas de temps revient à déplacer l'instant de prévision au dernier pas de temps où la mise-à-jour est possible et donc à établir une prévision à l'horizon $H' = H + N$ où H est l'horizon réel (le scénario de pluies « futures » est alors composée de pluies observées et de pluies à venir).

Les performances du modèle sont très sensibles à de telles lacunes. L'étude montre notamment une augmentation rapide et forte de l'erreur quadratique moyenne pour des lacunes courtes (multiplication médiane par 6 pour H' égal à 6 heures). L'erreur quadratique moyenne continue à augmenter pour des lacunes plus longues mais beaucoup plus lentement à partir de 24 heures (ordre de grandeur moyen, les situations pouvant évidemment être très différentes d'un bassin à l'autre). Cependant, même pour des lacunes de 5 jours, employer le dernier débit observé pour mettre le modèle à jour reste significativement bénéfique : les performances sont encore bien supérieures à celles du modèle fonctionnant en mode « simulation » (c'est-à-dire sans aucune mise-à-jour).

15.1.3 Synthèse

La comparaison des deux premières séries d'essais montre que le modèle est nettement plus sensible à une lacune dans les observations de débit. Il s'agit d'un autre indice du fait que la donnée de débit est, pour ce modèle, une donnée primordiale pour son bon fonctionnement. Nous observons également un comportement identique face à ces deux types de lacune : une détérioration rapide des performances suivie d'un ralentissement quand les lacunes deviennent très longues.

15.2 Influence du bruit dans les données

15.2.1 Cas des observations de pluie

Dans cet essai, les pluies passées sont bruitées par des erreurs multiplicatives tirées dans une loi lognormale de moyenne 1 (pas de biais). Ce choix se justifie par l'hétéroscédasticité des erreurs d'observations des pluies. Nous avons testé différentes valeurs de l'écart-type σ du bruit ajouté aux observations de pluie : 0.3, 0.5, 0.8, 1 et 1.2 : ces valeurs ont été choisies pour encadrer très largement les estimations des incertitudes des estimations radar de pluie réalisées par Habib *et al.* (2008). Pour chacun de ces essais et pour chaque bassin, 10 séquences aléatoires ont été tirées : les distributions des critères de performance sont calculées sur 178 bassins × 2 périodes × 10 tirages. Nous avons également vérifié que les distributions des critères de performance étaient semblables d'un tirage à l'autre.

Bien que validées, les données initiales ne sont pas exemptes d'erreurs. Le modèle est imparfait. Il est donc possible que l'ajout d'un bruit (qu'on espère léger) puisse améliorer les performances du modèle dans certains cas. Sur l'ensemble de nos essais (178 bassins, 2 périodes et 10 tirages de pluies), les performances n'étaient significativement meilleures que dans 3 % des cas, dans la configuration la plus favorable à des gains de performance (les prévisions de débit à un horizon d'une heure, basées sur les perturbations testées les plus légères). Il est donc possible de comparer les performances du modèle employant des observations de pluie non bruitées à celles du même modèle avec des observations perturbées pour caractériser la dégradation des performances face aux erreurs dans les données de pluie.

Un impact plus que limité

Bien que les erreurs sur les observations de pluie que nous avons testées aient des écarts-types assez élevés, les chutes de performance (mesurées par le critère $C_{2M,P}$) sont assez limitées (Fig. 15.2). L'erreur quadratique moyenne varie de moins de 10 % dans la majorité des cas quand l'écart-type de la perturbation est de 0.5, de moins de 25 % s'il est égal à 1 (pour chaque horizon de prévision testé). Nous vérifions que les différences de performance sont d'autant plus importantes que l'horizon de prévision est court.

La réactivité du bassin versant, un facteur clef

L'intuition se vérifie ici encore : la réactivité du bassin explique l'intensité de la détérioration des performances du modèle avec la perturbation des pluies observées. Pour les bassins les moins réactifs (un quart de l'échantillon), les pluies observées (passées) continuent de jouer un rôle important sur les débits à venir et les performances chutent de 7 points de $C_{2M,P}$ (médiane et moyenne) à un horizon d'une heure (pour une perturbation caractérisée par un écart-type de 1). Les performances du modèle sur les bassins les plus réactifs (un quart

de l'échantillon) ne baissent que de 3 points (médiane et moyenne) au même horizon et avec les mêmes perturbations.

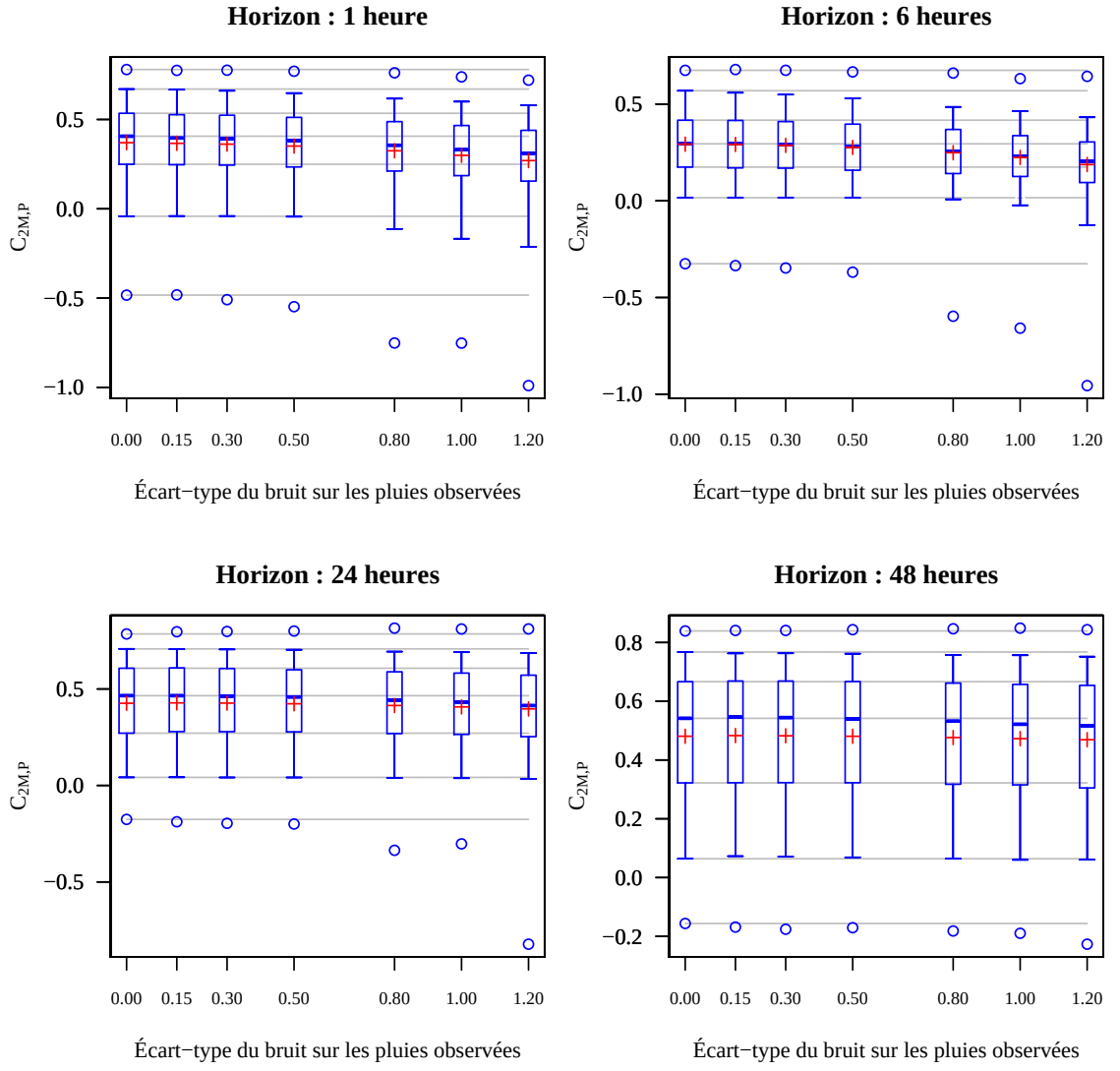


FIGURE 15.2 – Distributions (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP (avec les mises-à-jour de Tangara, 2005) lorsque les observations de pluie sont bruitées par des perturbations tirées dans une loi lognormale de moyenne 1 et d'écart-type égal à 0.15, 0.3, 0.5, 0.8, 1 ou 1.2. Quand cet écart-type est de de 1, le bruit généré est supérieur à 50 % des observations dans 18 % des cas, et supérieur à 100 % dans 10 % des cas.

15.2.2 Sensibilité du modèle au scénario de pluies futures

Employer les sorties des modèles numériques météorologiques pour alimenter nos scénarios de pluie future permet d'améliorer la qualité des prévisions opérationnelles de débit (comme le montrent Habets *et al.*, 2004, pour des grands bassins versants). Cependant les prévisions quantitatives de pluie restent encore très incertaines. De nombreux services météorologiques proposent d'ailleurs des ensembles de prévision de précipitations qui peuvent être injectés dans les modèles hydrologiques (parmi beaucoup d'autres, Palmer *et al.*, 2006, Rodwell, 2006, Ramos *et al.*, 2007, Collier, 2007). Une étude similaire au cas des erreurs aléatoires dans les observations de pluie est donc menée. Une seconde difficulté réside dans la différence de pas de temps entre la prévision des précipitations et le modèle hydrologique. Les prévisions de pluie de Météo-France, par exemple, sont communiquées par les bulletins de prévision au pas de temps journalier ou au pas de temps tri-horaire par les bulletins RR3 (*Rainfall Rate* au pas de temps 3 h). Sans employer ces fichiers (dont nous ne disposons pas), nous testerons l'effet d'une répartition uniforme (i.e. sans information supplémentaire) des prévisions de pluie future données sur des pas de temps supérieurs à une heure.

15.2.2.1 Erreurs aléatoires dans les prévisions de pluies futures

Nous avons reproduit l'expérience précédente en perturbant les pluies prévues entrant dans le modèle. Les perturbations multiplicatives, tirées dans les mêmes lois lognormales, sont appliquées aux pluies observées *a posteriori* entre les pas de temps $t+1$ et $t+H$ (t désignant l'instant de prévision et H l'horizon).

Un impact du même ordre de grandeur aux horizons courts... en moyenne
Aux horizons de prévision les plus courts (jusqu'à trois heures en moyenne, à moduler selon la réactivité du bassin), l'impact des perturbations des pluies futures sur les performances du modèle sont du même ordre de grandeur que celui des perturbations sur les observations de pluie (Fig. 15.3). Pour des horizons plus lointains, l'impact devient naturellement de plus en plus important. Pour un horizon de 24 heures par exemple, l'erreur quadratique moyenne augmente de plus de 30 % sur la moitié des bassins versants quand l'écart-type des perturbations multiplicatives des pluies futures est égal à 1.

Mais des conséquences plus fortes
À niveau d'incertitude égal (sur les pluies observées et les pluies futures), les conséquences sur les incertitudes des débits prévus sont comparables. Mais les incertitudes sur les pluies futures sont bien plus grandes. Garçon *et al.* (2008) présentent un exemple sur le bassin du barrage de Vouglans (1 120 km^2) où la variance des erreurs sur les débits prévus (à 6 jours) est imputable à 83 % aux pluies prévues.

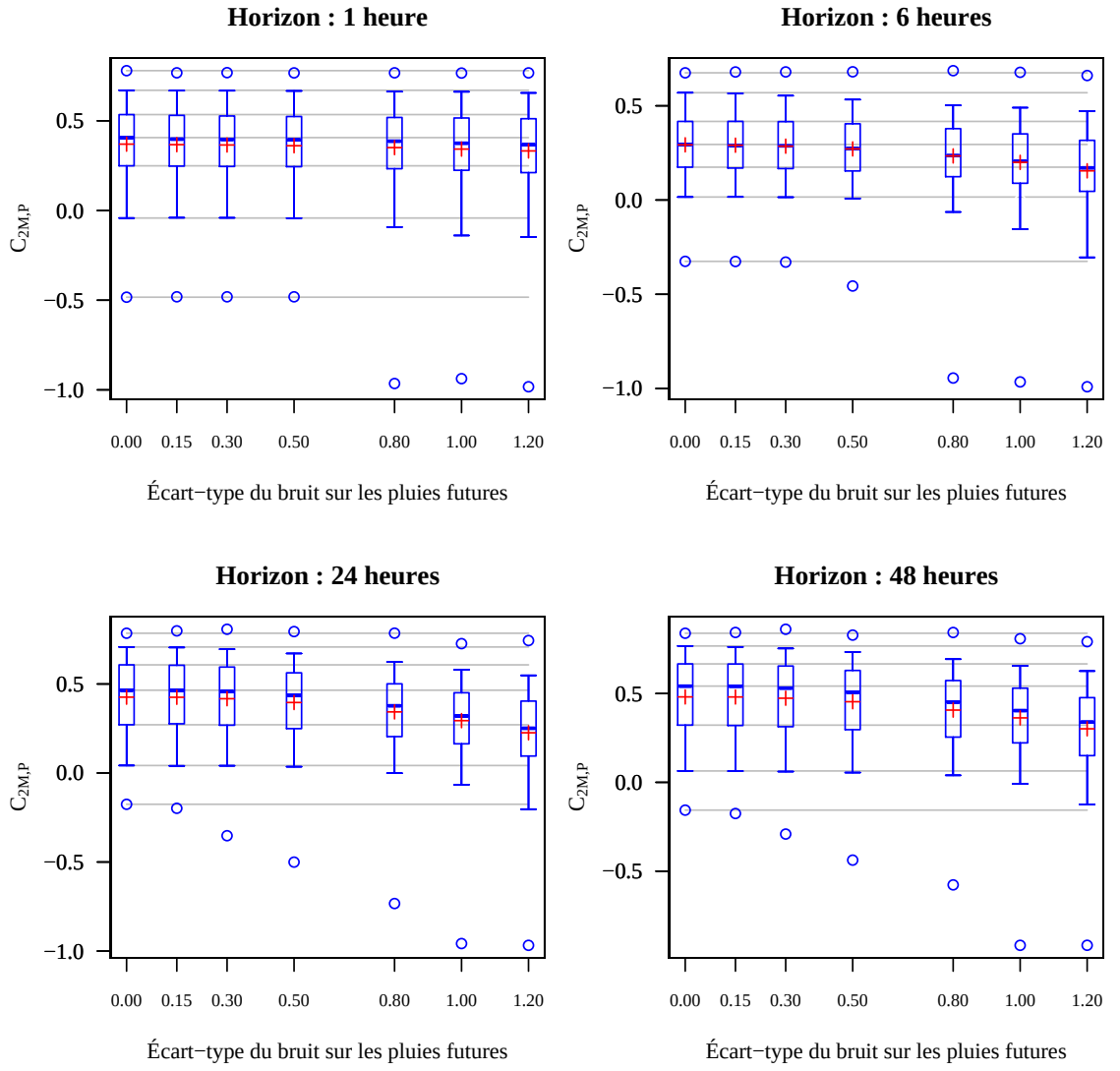


FIGURE 15.3 – Distributions (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP (avec les mises-à-jour de Tangara, 2005) lorsque les prévisions de pluie future sont bruitées par des perturbations tirées dans une loi lognormale de moyenne 1 et d'écart-type égal à 0.15, 0.3, 0.5, 0.8, 1 ou 1.2.

Une variabilité plus grande

Cette augmentation de l'impact des perturbations en moyenne avec l'horizon de prévision s'accompagne également d'une variabilité plus forte. À l'horizon de 48 heures, l'augmentation de l'erreur quadratique moyenne est supérieure à 100 % sur plus de 10 % des bassins (pour une valeur de σ égale à 1). Cela s'explique par le fait que les perturbations sur les observations des pluies (passées) sont partiellement « effacées » par la mise-à-jour des états du modèle effectuée à l'instant de prévision. Au contraire des perturbations sur les pluies prévues que rien ne vient éclipser.

Les bassins les plus réactifs sont les plus sensibles

Résultat attendu puisque les bassins réactifs sont ceux pour lesquels les pluies passées et la mise-à-jour des états effectuée juste avant la prévision ont le moins d'influence à un horizon de prévision donné, les bassins sont d'autant plus sensibles aux perturbations sur les pluies futures que l'autocorrélation des débits est faible.

15.2.2.2 Effet de la répartition temporelle des prévisions de pluie

Cette fois, les cumuls de pluie future sur une durée τ sont supposés égaux aux cumuls des observations *a posteriori*. Le prévisionniste ne dispose que de ces cumuls. Nous testons la répartition uniforme de ces cumuls sur les pas de temps horaires. Différentes valeurs de τ sont testées : 2, 3, 6, 12 et 24 heures.

Un impact limité

L'effet sur les distributions de performance (mesurée par le critère $C_{2M,P}$) est très faible, si les prévisions de pluie sont données par cumuls d'une durée inférieure ou égale à 6 heure. Au-delà, les performances se dégradent de façon significative certes, mais très nettement moins que pour les études de sensibilité précédentes (Fig. 15.4) : le format temporel des prévisions de pluie future n'est clairement pas son aspect le plus critique.

Un impact variable selon l'horizon de prévision

Ici encore, trois phases sont observées. Pour les horizons de prévision les plus courts, les pluies futures ont un rôle très limité sur les débits en sortie du modèle : la durée du pas de temps de prévision des pluies joue donc un rôle quasi nul. Pour des horizons intermédiaires, on note une dégradation significative mais limitée. Pour des horizons plus lointains, la question du pas de temps des prévisions apparaît moins cruciale (à la condition que les cumuls soient corrects).

L'effet dépend fortement de la réactivité des bassins versants

Les performances du modèle sont naturellement plus sensibles à la résolution temporelle des prévisions de pluie pour les bassins qui réagissent le plus vite : l'influence de cette

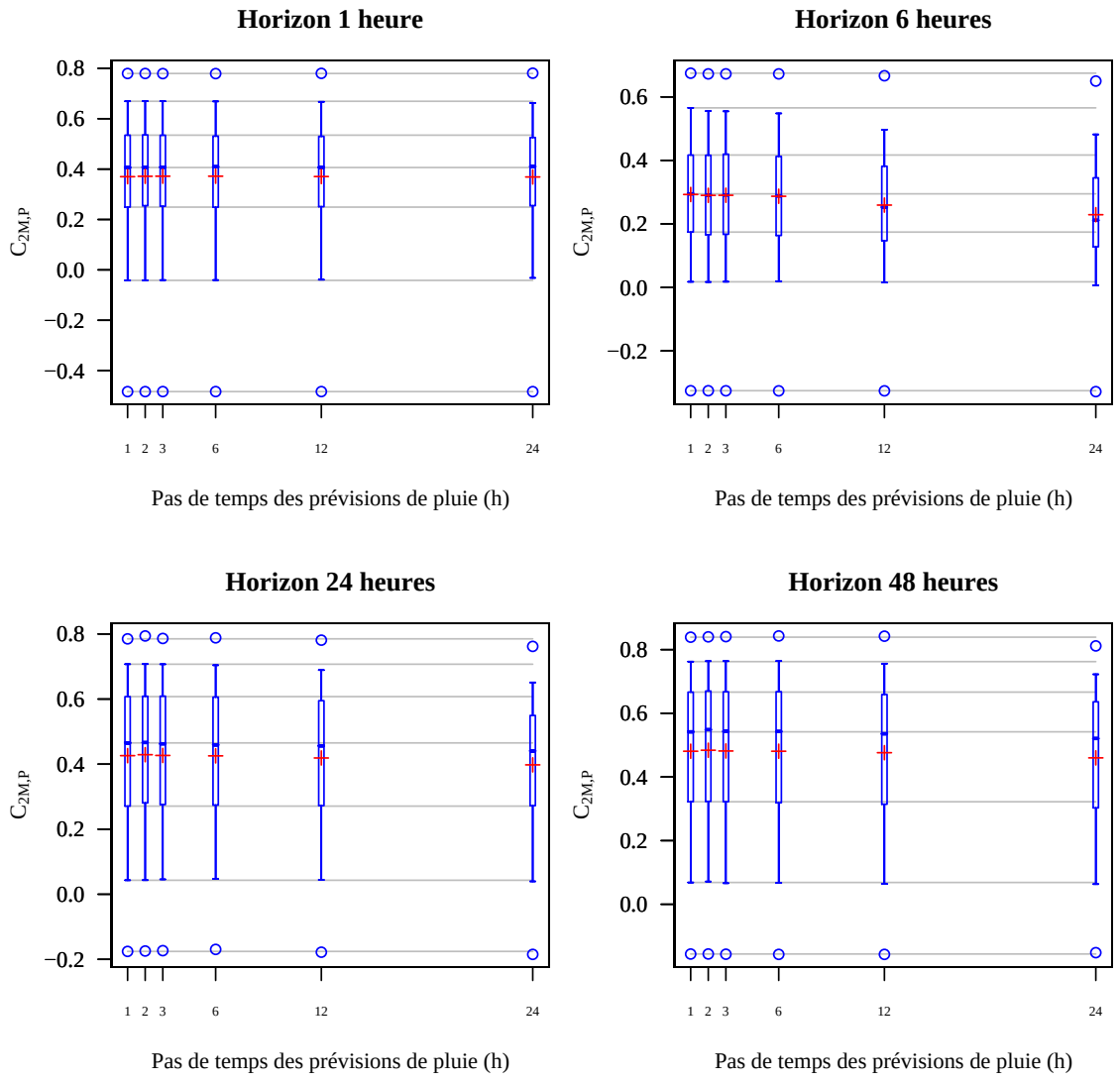


FIGURE 15.4 – Distributions (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP (avec les mises-à-jour de Tangara, 2005) lorsque les prévisions de pluie future sont fournies sur des pas de temps de 1, 2, 3, 6, 12 et 24 heures.

résolution temporelle est faible sur les bassins présentant la plus forte autocorrélation des débits, beaucoup plus élevée sur les bassins dont les débits sont le moins autocorrélés (Fig. 15.5). Cette influence de la réactivité des bassins versants sur la sensibilité aux durées de cumul des prévisions de pluie future est également d'autant plus forte que l'horizon de prévision est court.

15.2.3 Sensibilité des mises-à-jour aux erreurs sur les mesures de débit

Les états du modèle de prévision GRP sont mis-à-jour grâce à la dernière observation du débit. Le modèle devient donc sensible à la qualité de cette observation. Nous examinerons tout d'abord la réaction du modèle quand le niveau du réservoir de routage est mis-à-jour de façon s'appuyant « complète » (mise-à-jour proposée par Tangara, 2005, et toujours considérée comme la plus efficace sur des données validées). Puis nous comparerons cette approche à deux mises-à-jour partielles : l'une réalisée par une approche simple dérivée de la mise-à-jour précédente, l'autre par le filtre de Kálmán.

15.2.3.1 Mise-à-jour complète du niveau du réservoir de routage

La mise-à-jour du niveau du réservoir de routage proposée par Tangara (2005) est la plus efficace de toutes les méthodes de mise-à-jour des états testées au chapitre 7 avec des données validées. Un défaut potentiel de cette méthode réside dans sa sensibilité à la qualité des données qu'elle emploie car il s'agit d'une mise-à-jour s'appuyant « complètement » sur l'observation assimilée. Nous examinons donc l'impact d'erreurs dans les données de débit sur les performances du modèle GRP employant cette mise-à-jour.

Tests réalisés

Les incertitudes sur les observations de débit sont souvent évaluées autour de 10 à 15 % du débit (e.g. Sorooshian et Gupta, 1995). Des erreurs systématiques, liées à la courbe de tarage, supérieures à cet ordre de grandeur peuvent cependant survenir lors de crues. Nous introduirons dans un premier temps des perturbations multiplicatives [3] issues d'une loi lognormale, de moyenne 1 et d'écart-type σ égal à 0.05, 0.10, 0.15, 0.20 ou 0.50.

Résultats

Quand la mise-à-jour du niveau du réservoir de routage est complète, le modèle est très sensible aux erreurs faites sur les observations de débit assimilées (Fig. 15.6). Aux horizons de prévision les plus courts, les performances médianes sont même inférieures à celles du modèle de persistance : le modèle n'est plus du tout utile sur plus de la moitié des bassins versants ! Cela correspond à de très fortes augmentations de l'erreur quadratique moyenne

3. La mise-à-jour du niveau du réservoir de routage effaçant complètement les effets des mises-à-jour précédentes, il n'est pas utile ici de prendre en compte une autocorrélation des erreurs sur le débit.

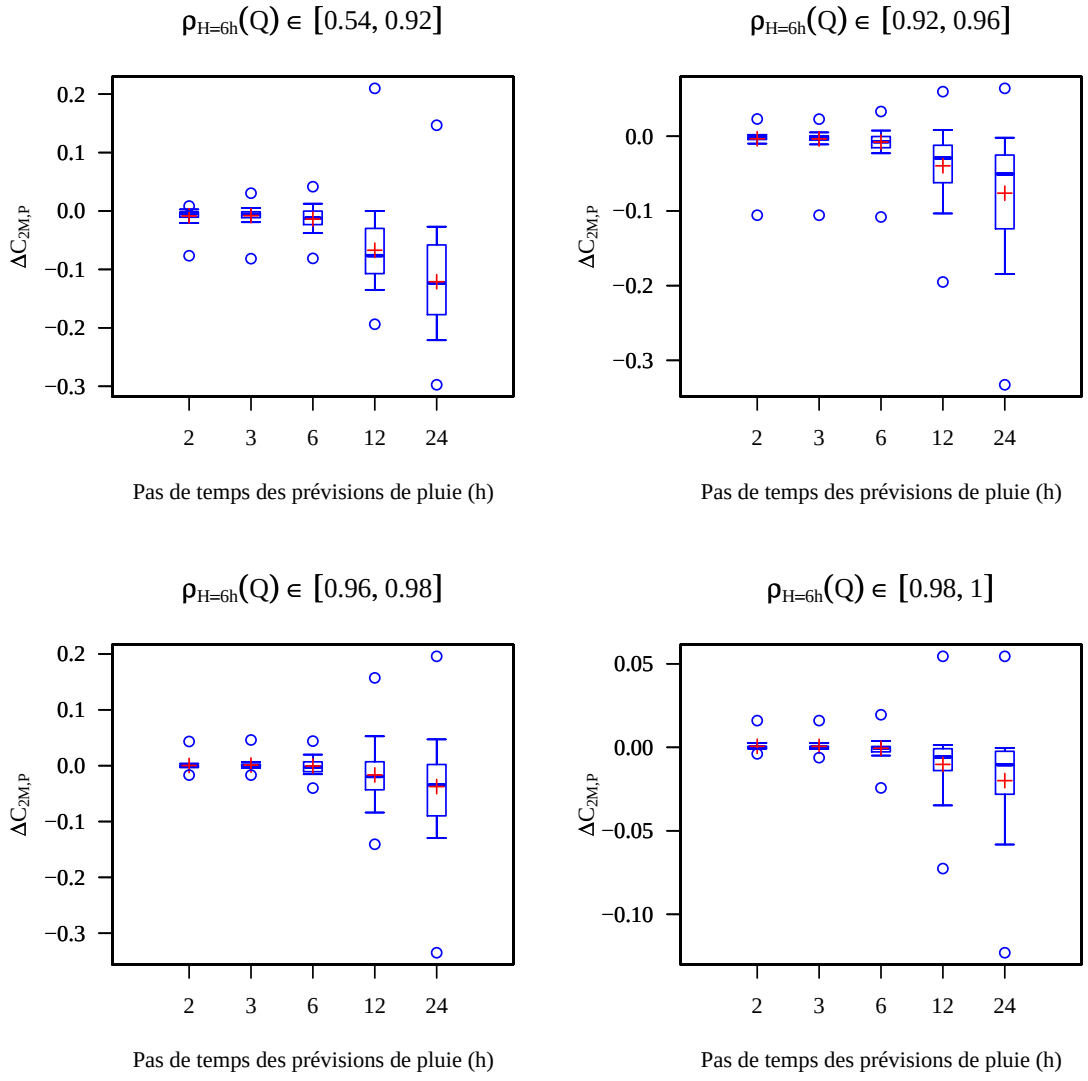


FIGURE 15.5 – Distributions (sur 178 bassins versants × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP (avec les mises-à-jour de Tangara, 2005) lorsque les prévisions de pluie sont fournies sur des pas de temps de 2, 3, 6, 12 et 24 heures, auxquelles on a soustrait la valeur du critère pour des prévisions de pluie horaire. L'horizon est de 6 heures. Chacun des quatre graphiques correspond à une classe de réactivité des bassins versants correspondant à une gamme de l'autocorrélation des débits : les bassins les plus réactifs sont situés en haut à gauche ; les bassins les plus lents en bas à droite. Les échelles des ordonnées diffèrent avec les classes de réactivité.

faite par le modèle (Tab. 15.2).

Cette sensibilité aux erreurs dans les données de débit décroît naturellement avec l'horizon de prévision mais cette décroissance est lente : à un horizon de 24 heures, l'erreur quadratique moyenne est plus que doublée sur trois quarts des bassins versants par l'introduction d'un bruit sur les données de débit.

	Quantiles				
Horizon	0.05	0.25	0.50	0.75	0.95
1 heure	83	230	392	588	992
6 heures	67	122	164	248	352
24 heures	92	105	120	173	333
48 heures	98	103	112	146	354

Table 15.2 – Distributions (sur 178 bassins × 2 périodes) des rapports (exprimés en %) des erreurs quadratiques moyennes faites par le modèle quand les données de débit à assimiler sont bruitées (écart-type égal à 0.15 fois le débit observé) et quand elles ne le sont pas.

15.2.3.2 Emploi du filtre de Kálmán

Pour remédier à ce défaut, le filtre de Kálmán a été étudié dans le chapitre 7 consacré à la mise-à-jour des états. Nous avions alors montré que le filtre ne permettait pas d'obtenir de meilleures performances que la mise-à-jour de Tangara lorsque les données de débit assimilées étaient validées (de bonne qualité). Qu'en est-il lorsque la donnée de débit est plus incertaine ?

Tests

Dans un premier temps, le protocole expérimental est identique au précédent, en ce qui concerne la nature et l'intensité des perturbations introduites sur les données de débit assimilées. Les mêmes 10 tirages aléatoires sont employés pour vérifier que nos résultats ne dépendent pas fortement du tirage aléatoire. Dans ce cadre, 9 paramétrisations des matrices de variance-covariance ont été testées.

Nous avons ensuite testé une perturbation différente : l'erreur introduite est additive, normale de moyenne nulle et d'écart-type proportionnel au débit observé. Deux valeurs de cet écart-type ont été testées : une erreur faible (2 %) et une erreur forte (20 %). 10 matrices de variance-covariance ont été proposées au filtre.

Une sensibilité à la nature des erreurs ?

Les résultats sont très contrastés selon la nature des erreurs (additive normale ou multiplicative log-normale). Dans le premier cas (Fig. 15.7), le filtre de Kálmán bien paramétré permet d'améliorer légèrement les prévisions par rapport aux prévisions obtenues avec le

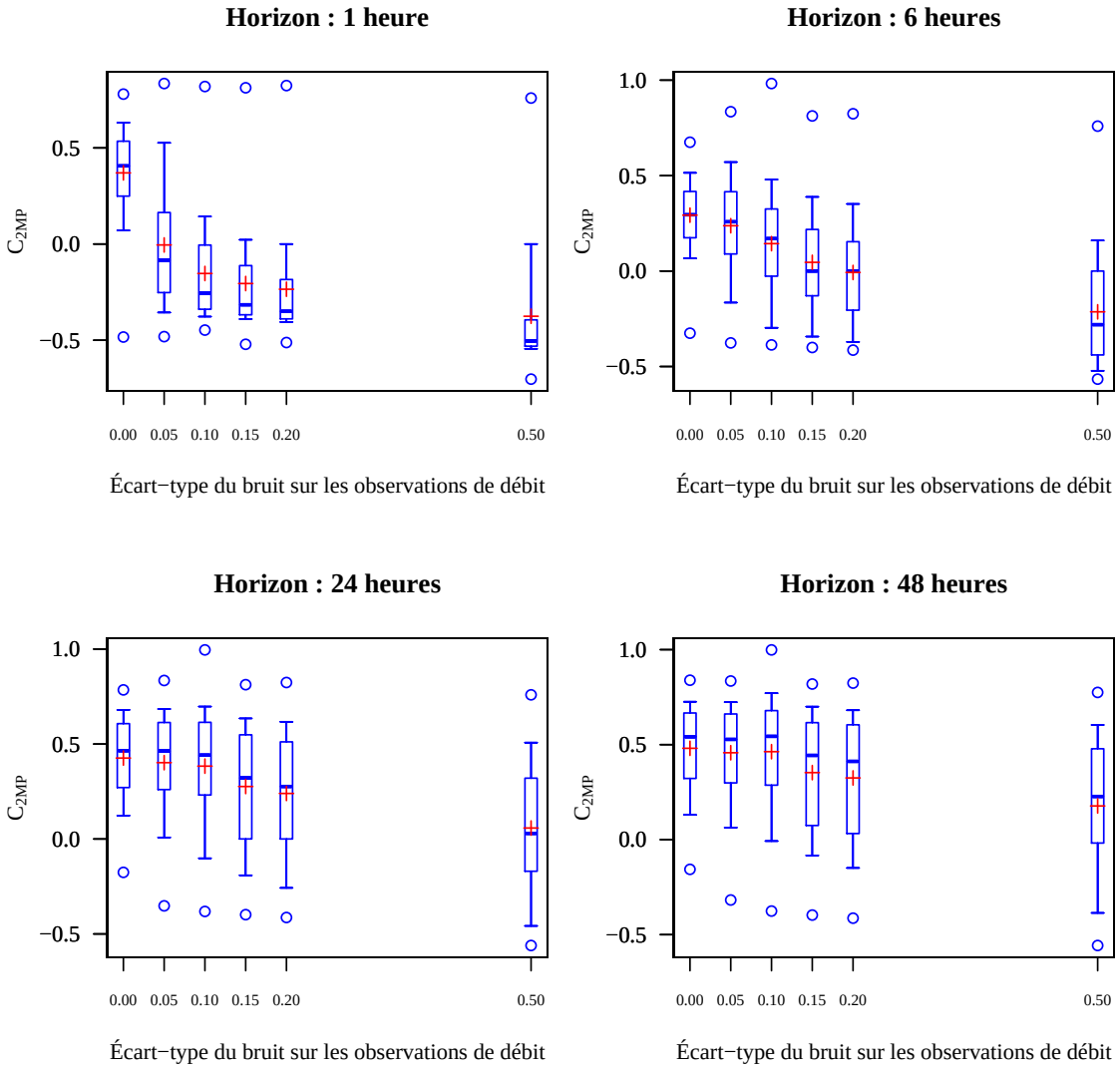


FIGURE 15.6 – Distributions (sur 178 bassins × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP avec la mise-à-jour complète proposée par Tangara (2005), quand le débit assimilé est perturbé par un facteur aléatoire d'écart-type 0.05, 0.10, 0.15, 0.02 ou 0.50.

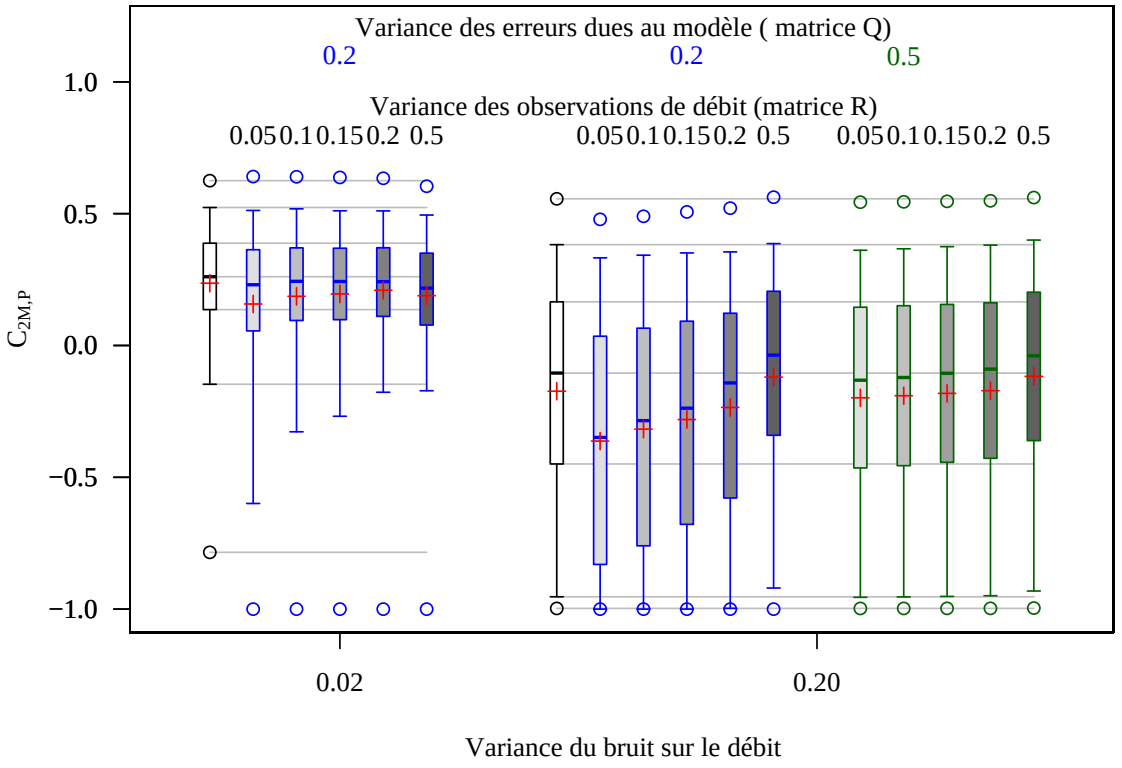


Figure 15.7 – Distributions (sur 178 bassins × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP mis-à-jour par le filtre de Kálmán, quand le débit assimilé est perturbé par un bruit gaussien d'écart-type $0.02 \cdot Q_t$ (à gauche) ou $0.20 \cdot Q_t$ (à droite). L'horizon est de 6 heures. Pour chaque niveau de bruit, différentes paramétrisations des matrices de variance-covariance $\mathbf{Q}_t$ et $\mathbf{R}_t$ sont testées. Nous indiquons la valeur de la variance des erreurs d'observation (matrice $\mathbf{R}_t$) en noir et la valeur de la variance des erreurs sur le niveau du réservoir de routage (matrice $\mathbf{Q}_t$) en couleur. La boîte à moustache dessinée en noir correspond au cas où la mise-à-jour est complète.

modèle GRP mis-à-jour par la méthode de Tangara (2005), lorsque le bruit introduit est assez important (20 %). Pour obtenir ce résultat, nous avons dû opter pour une variance des erreurs des observations plus grandes que celle des erreurs du modèle : nous affichions une confiance faible dans ces observations. Le gain est relativement modeste, comparé à la chute de performance associée au bruit sur les débits (plusieurs dizaines de points de $C_{2M,P}$, Fig. 15.6) : dans le meilleur cas (qui a lieu pour l'horizon de 3 heures), le gain médian est de 10 points de $C_{2M,P}$. La configuration permettant d'obtenir les plus forts gains médians est paramétrée par une variance des erreurs sur le modèle faible par rapport à celle des erreurs sur les observations. Cette configuration entraîne une perte de robustesse sur un petit nombre de bassins. Une solution plus équilibrée permet d'éviter cette situation mais conduit à des gains de performances légèrement moins élevés.

Dans le deuxième test (distribution du bruit log-normale), aucune des paramétrisations

testées du filtre de Kálmán n'a permis d'obtenir des performances supérieures à celles obtenues par le modèle dont le niveau du réservoir de routage est mis-à-jour de façon complète.

Cette observation nous interroge sur la sensibilité du filtre de Kálmán à la nature des erreurs sur les observations. Le filtre fait l'hypothèse de distributions gaussiennes. Si le filtre se révélait très sensible à la forme de la distribution des erreurs sur les observations, il serait nécessaire de réaliser une transformation des débits pour obtenir une distribution (approximativement gaussienne), ce qui est loin d'être toujours évident (e.g. Yang *et al.*, 2007).

15.2.3.3 Mise-à-jour partielle du niveau du réservoir de routage

À titre de comparaison, nous avions proposé au chapitre 7 (page 129) une mise-à-jour « incomplète » plus simple du niveau du réservoir de routage : le débit servant à la mise-à-jour est une moyenne pondérée du débit simulé et du débit observé à l'instant de la précision. La pondération est fonction de notre confiance relative dans les simulations et observations.

Tests réalisés

Nous nous sommes replacés dans le cas d'un bruit multiplicatif lognormal traité précédemment. Seuls trois niveaux de confiance relative ξ sont testés : 0.5, 0.9 et 0.96 (correspondant respectivement à des valeurs du rapport des variance α de 1, 3 et 5, cf. Annexe K, page 517). Il ne s'agit pas ici de tirer des conclusions complètes sur cette méthode mais plutôt d'illustrer son potentiel.

Résultats

Cette méthode permet d'améliorer les performances du modèle GRP dans cette situation (Fig. 15.8). Cette approche est donc capable de traiter des erreurs non gaussiennes. Elle ne compte qu'un seul paramètre à caler, ce qui la rend beaucoup plus facilement utilisable.

15.2.3.4 Discussion

Les essais des deux mises-à-jour partielles des états de GRP n'ont pas été conduits assez loin pour permettre d'établir une solution préférentielle. À ce stade, l'approche à un paramètre est loin de paraître moins bien armée que le filtre de Kálmán pour « limiter la casse » en cas de données erronées dans les observations de débit.

Ces deux méthodes ne sont probablement pas bien adaptées au problème posé par les erreurs dans les données de débit. Comme nous l'avons mentionné auparavant, ces erreurs ne vérifient pas le plus souvent les hypothèses communément faites par les méthodes

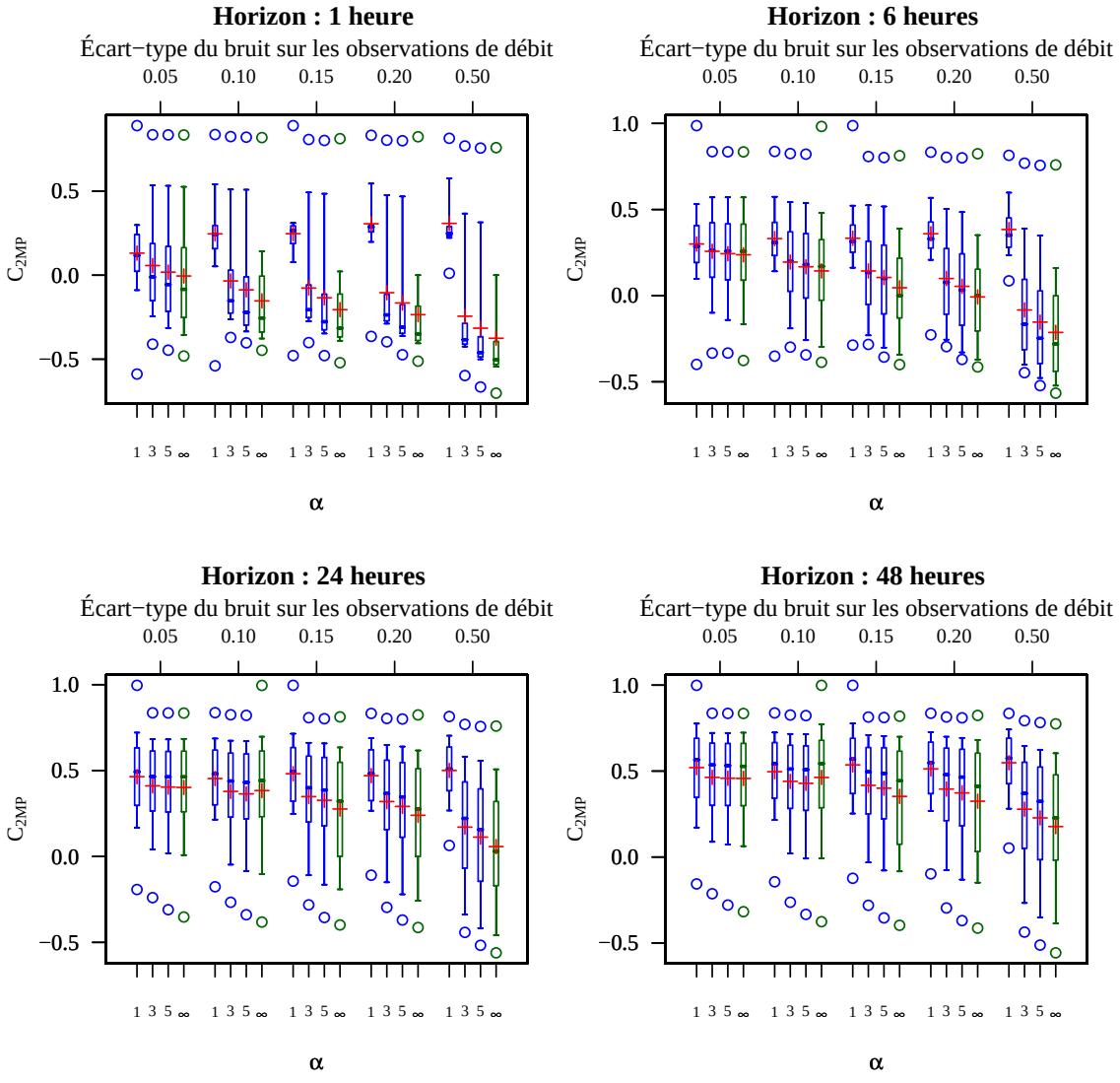


Figure 15.8 – Distributions (sur 178 bassins × 2 périodes) des valeurs du critère de performance $C_{2M,P}$ obtenues par le modèle GRP avec la mise-à-jour complète proposée par Tangara (2005), quand le débit assimilé est perturbé par un facteur aléatoire d'écart-type 0.05, 0.10, 0.15, 0.20 ou 0.50. Pour chaque niveau de bruit, trois niveaux de confiance ξ sont testés : 0.50, 0.90 et 0.960 qui correspondent aux valeurs de α de 1, 3 et 5 (cf. Annexe K, page 517). La boîte à moustache dessinée en vert (pour $\alpha = +\infty$) correspond au cas où la mise-à-jour est complète.

classiques : elles peuvent être très autocorrélées et contiennent sans doute une partie systématique non négligeable liée à la courbe de tarage. Il reste un travail conséquent à effectuer pour caractériser la structure de ces erreurs et proposer les outils adaptés.

C'est pourquoi il nous paraît important d'insister sur la nécessité de ne pas assimiler des observations à l'aveuglette, mais au contraire d'examiner attentivement et de valider ces données. Un premier traitement algorithmique peut être réalisé, mais le regard du prévisionniste est encore nécessaire. Dans tous les cas, l'outil de prévision doit indiquer simultanément la dernière observation sur laquelle il s'est basé et la prévision, pour que le prévisionniste puisse l'analyser.

15.3 Bilan et discussion

Nous avons testé la sensibilité du modèle GRP à des erreurs aléatoires de ses différentes entrées : observations de pluie, prévisions de pluie future et mesures de débit. Deux situations ont été examinées : dans le premier cas, les données ne sont pas disponibles (lacunes) ; dans le second, des valeurs erronées sont disponibles. Nous avons cherché à tester des écarts aux valeurs « réelles » qui peuvent être rencontrés par les prévisionnistes.

Les erreurs auxquelles le modèle est le plus sensible, sont sans conteste celles qui affectent la donnée de débit. Cela correspond au défaut majeur au revers d'une des qualités fondamentales de l'approche GRP : des gains de performance substantiels ont été obtenus en faisant un meilleur usage de l'information de débit lors de la mise-à-jour du modèle. Faire disparaître l'information de débit ou la bruiter, ruine ces avancées.

Les incertitudes sur les pluies observées sont supérieures aux incertitudes sur les mesures de débit. Cependant leur impact est moindre. Le modèle est relativement peu sensible aux erreurs sur les pluies observées au contraire des modèles de simulation. Kobold et Brilly (2006) relèvent que le modèle HBV est très sensible à l'information de pluie et une incertitude de 10 % sur les pluies peut générer des incertitudes de 17 % sur les pointes. Moulin *et al.* (2009) montrent que l'incertitude d'estimation des pluies explique une grande partie de l'incertitude des débits simulés sur un petit bassin versant tandis qu'elle n'est en cause que d'une faible part de l'erreur de simulation faite sur des bassins plus grands. Plus que le caractère intégrateur du bassin versant (Oudin *et al.*, 2004), c'est bien la mise-à-jour du modèle qui assure cette relative robustesse du modèle de prévision face aux erreurs dans les données de précipitations.

Les prévisions de pluie future sont bien plus incertaines, plus en raison de l'incertitude sur les cumuls prévus que de la résolution temporelle des prévisions de précipitation. L'impact de cette incertitude est supérieur à celle des observations de pluie mais ne dépasse pas celle associée aux débits pour des horizons courts (inférieurs à 2 jours).

Bilans et perspectives

I know very well, how little reputation is to be got by writings which require neither genius nor learning, nor indeed any other talent, except a good memory, or an exact journal. I know likewise, that writers of travels, like dictionary-makers, are sunk into oblivion by the weight and bulk of those who come last, and therefore lie uppermost. And it is highly probable, that such travellers, who shall hereafter visit the countries described in this work of mine, may, by detecting my errors (if there be any), and adding many new discoveries of their own, justle me out of vogue, and stand in my place, making the world forget that ever I was an author. This indeed would be too great a mortification, if I wrote for fame.

Jonathan Swift, *Gulliver's travels*

Bilan

La prévision des crues reste un défi tant scientifique qu'opérationnel. De nombreuses approches sont proposées pour améliorer la prévision des débits et l'alerte. Dans ce travail, nous nous sommes appuyés sur la riche expérience de modélisation hydrologique existant au Cemagref et avons cherché à affiner la prévision des débits au pas de temps horaire par l'emploi de méthodes d'assimilation du débit mesuré à l'exutoire. Quel bilan pouvons-nous en tirer ?

Sur un plan méthodologique

Nous avons respecté l'une des marques de fabrique de l'équipe *Hydrologie* de l'unité de recherche *Hydrosystèmes et Bioprocédés* du Cemagref (Antony), à savoir travailler sur un très grand nombre de bassins versants. Toutefois, nous avons proposé une approche alternative : au lieu de tester le modèle sur tous les bassins pour lesquels des données sont disponibles en quantité et qualité suffisante, nous avons évalué les similarités hydrométéorologiques des bassins de cet ensemble (à l'aide d'une mesure très imparfaite bien sûr mais que nous pensons tout de même informative), pour définir un échantillon plus restreint que nous espérons plus équilibré.

De même, nous avons étudié les corrélations existant entre un grand nombre de critères de performance pour choisir le niveau de raffinement utile dans la définition des critères et pour sélectionner ceux qui nous ont servi à évaluer notre modèle.

Confirmation des résultats de Tangara (2005) obtenus au pas de temps journalier

Plusieurs résultats acquis au pas de temps journalier par Tangara (2005) dont les travaux constituaient le point de départ de cette thèse se sont trouvés confirmés au pas de temps horaire. Parmi ceux-ci, on compte :

- La structure hydrologique du modèle de prévision est plus simple que celle du modèle de simulation au même pas de temps. De façon plus générale, la grande similarité des structures des modèles de simulation aux pas de temps journalier (Perrin *et al.*, 2001) et horaire (Mathevet, 2005, Le Moine, 2008) se retrouve en prévision : nous conservons une structure de modèle ne comptant que trois paramètres libres. La fonction de routage ne comporte qu'une seule branche d'écoulement.
- Quand la donnée de débit est assimilée, celle-ci prend une importance primordiale pour assurer la qualité des prévisions à (relativement) courte échéance. Le modèle devient un modèle (Pluie, Débit) – Débit.

- En conséquence, il est préférable, au pas de temps horaire comme au pas de temps journalier, de caler les paramètres de la structure hydrologique en mode « Prévision », c'est-à-dire avec les mises-à-jour non paramétriques du modèle.
- La mise-à-jour des états se concentre sur la mise-à-jour du niveau du réservoir de routage, c'est-à-dire l'état le plus en aval du modèle (i.e., l'état le plus proche du débit). Cette mise-à-jour assure une forte amélioration de la précision des prévisions. *A contrario*, son absence conduit systématiquement à des performances nettement plus faibles, quelle que soit l'approche de mise-à-jour utilisée.
- La mise-à-jour des paramètres apporte un gain de performance mais celui-ci est nettement plus modeste que le précédent.

Des gains significatifs des performances du modèle par assimilation du débit...

Les expérimentations menées au pas de temps horaire ont permis d'améliorer significativement la précision des prévisions de débit. Ces progrès sont dus pour l'essentiel à l'emploi de meilleures méthodes d'assimilation des dernières observations de débit à l'exutoire.

Les meilleurs résultats ont été obtenus par la combinaison de la mise-à-jour du réservoir de routage assimilant entièrement la donnée de débit et d'une correction des sorties par réseaux de neurones artificiels (méthode de loin la plus performante) ou par modèle ARIMA (approche la plus robuste).

Nous avons cherché à mettre au point et à tester des méthodes adaptées aux conditions opérationnelles. Les mises-à-jour du modèle proposées se caractérisent par des « durées de vie » qui les rendent utiles aux prévisionnistes. Elles ne se basent que sur des données disponibles pour le prévisionniste. Leurs robustesses ont été établies.

Cet outil est applicable à un grand nombre de bassins versants et les travaux présentés dans le cadre de cette thèse font d'ores et déjà l'objet de transfert vers des outils logiciels destinés aux services opérationnels de prévision de crues (notamment les SPC et le SCHAPI). Le logiciel GRP bénéficie des avancées proposées par Tangara (2005) et de certaines de celles présentées dans cette thèse.

... Ou des résultats en trompe-l'œil ?

Cet aspect positif de nos résultats ne doit pas nous faire oublier leurs limites tant théoriques que pratiques :

Le modèle assimile les données de débit à l'exutoire pour prévoir le débit à l'exutoire. Ses performances sont évaluées au niveau du débit à l'extoire. Est-ce raisonnable ? Les données assimilées ont été validées par des hydrologues. Il reste cependant la possibilité d'erreurs fortes, notamment systématiques (par exemple liées aux courbes de tarage) alors que les

progrès que nous avons faits se concentrent sur la réduction des erreurs aléatoires.

Les améliorations des prévisions en termes de précision s'appuient quasi uniquement sur un meilleur emploi des dernières observations de débit. Dussent-elles manquer ou être entachées d'erreur, ces progrès seront immédiatement remis en cause. Aussi, l'emploi de ces résultats doit impérativement s'accompagner de la validation humaine des données à assimiler.

Quelques perspectives

Il reste un grand nombre de questions non résolues qui laissent ouvertes des perspectives passionnantes de recherche et des améliorations potentielles de nos outils de prévision. En effet, si ce travail espère apporter des améliorations à la prévision des crues par la modélisation hydrologique, il est loin de proposer une solution suffisante à l'ensemble des difficultés que peut rencontrer le prévisionniste. Nous pouvons mettre en avant quelques pistes de travail qui nous semblent particulièrement importantes à traiter.

La question du pas de temps

Cette question peut ne pas sembler être de première importance. Pourtant, elle reste très intriguante. D'abord fonction de la disponibilité des données, désormais peu influencée par les capacités de calcul, elle est aussi une problématique de la modélisation. Nous avons vu que les débits de pointe sont nettement mieux rendus au pas de temps horaire qu'au pas de temps journalier sur un grand nombre de bassins[1]. Cependant, l'étude, très exploratoire menée avec Tahir (2008), ne permet pas de montrer la supériorité de la modélisation au pas de temps horaire sur celle à des pas de temps multi-horaire. Le pas de temps influe sur deux aspects de la modélisation : d'abord sur les processus (ou plus prosaïquement sur les aspects du signal de débit) dont nous pouvons espérer rendre compte, ensuite sur l'efficacité des mises-à-jour. Comme nous l'avons vu, l'efficacité de certaines méthodes de correction des sorties du modèle dépend fortement du nombre de pas de temps nécessaire pour atteindre l'horizon de prévision. Il paraît donc nécessaire de chercher à définir un équilibre entre un pas de temps suffisamment fin pour décrire le phénomène qui nous intéresse et un pas de temps assez long pour permettre l'usage de méthodes de correction des sorties.

Assimiler d'autres données

Assimiler le débit à l'exutoire apparaît de prime abord comme étant l'action la plus naturelle pour améliorer nos prévisions. Si cela permet de fortes améliorations des prévisions, cette approche présente des limites et il apparaît nécessaire de chercher à assimiler d'autres

1. Sur certains bassins, particulièrement rapides, le pas de temps horaire n'est pas suffisant pour capturer les pointes de crue.

données.

Des travaux exploratoires menés par Simon Munier et Xavier Litrico (U.M.R. *G-Eau*, Cemagref, Montpellier) sur le bassin du Serein montrent que l'assimilation par GRP de débits en amont présente un fort intérêt. Cette approche n'a cependant rien d'évident et il est nécessaire de définir finement la combinaison entre les aspects hydrologiques, les aspects hydrauliques et l'assimilation de données. Il s'agit en quelque sorte d'établir la jonction entre les travaux de Lerat (2009) en simulation et ceux réalisés en prévision (dans cette thèse). On pourra s'appuyer sur le « versant » hydraulique de cette question proposé par Munier (2009) pour en explorer le « versant » hydrologique. Cette approche, bien que prometteuse, ne concerne que les seuls bassins pour lesquels il existe une station de jaugeage en amont ou en aval dont le signal est informatif.

Pour tous les bassins et notamment les petits bassins éloignés de toute autre station de jaugeage, il est nécessaire de chercher à assimiler d'autres types de données. Il peut s'agir par exemple d'observations en relation avec l'humidité du bassin versant. Pour l'instant, cela pose deux problèmes : la disponibilité opérationnelle et la qualité des observations d'abord, la relation entre ces observations et les états du modèle ensuite.

Rendre compte de l'incertitude des prévisions

Si faire une conjecture a ici un sens, il s'agit peut-être de la question la plus difficile et la plus centrale concernant l'assimilation de données en hydrologie. Se tourner vers l'assimilation, c'est d'abord reconnaître les limites de la modélisation et assumer une incertitude non négligeable. Incertitude qu'il nous faut comparer à l'incertitude des observations. Cette comparaison demande d'établir un modèle des erreurs d'observation et un modèle des erreurs de simulation. C'est une tâche délicate qui fait l'objet de nombreuses contributions mais il s'agit surtout d'un véritable nœud de recherche pour toutes les méthodes qui, comme l'inférence bayésienne, reposent sur l'expression d'une fonction de vraisemblance. Nous avons vu que les méthodes les plus simples d'évaluation des incertitudes (comme le filtre de Kálmán) échouent. Les essais d'une méthode plus complexe que nous avons menés en collaboration avec Vincent Fortin d'Environnement Canada, ne nous ont pas permis d'obtenir des résultats entièrement convaincants.

Une autre difficulté en est la multiplicité des sources d'incertitude à prendre en compte et la dépendance de l'importance de ces sources à l'horizon de prévision : incertitudes de l'état initial, incertitudes des prévisions de pluie future... Déterminer à quelles sources d'incertitude il faut s'intéresser et à quel horizon de temps, est une question largement explorée par les hydrologues mais reste encore un vrai défi de recherche. Il apparaît nécessaire de caractériser correctement l'incertitude sur l'état à l'instant de la prévision puis d'évaluer les conséquences de l'incertitude sur les pluies futures (par exemple à travers des prévisions de précipitation d'ensemble). Mais de même que la complexité d'un modèle n'est pas la

somme de la complexité de ses éléments, l'incertitude de ses sorties n'est pas la somme des incertitudes de ses différents composants... Il faut tenir compte des interactions entre ces sources d'incertitudes et le fonctionnement du modèle.

Liens entre mises-à-jour et modèles hydrologiques

Au cours de cette thèse, nous ne nous sommes finalement appuyés que sur un seul modèle hydrologique (alors que nous avons employé un grand nombre de bassins versants...). Nous avons proposé des méthodes qui étaient spécifiques à cette structure et d'autres, comme les corrections des sorties du modèle, qui sont susceptibles de s'appliquer à tout modèle. On retrouve parmi les mises-à-jour les plus efficaces des méthodes appartenant à ces deux catégories. Nos conclusions sont cependant spécifiques à la structure hydrologique de GRP et il ne nous est pas encore possible d'affirmer qu'une approche est plus prometteuse que l'autre. Il nous paraît d'abord important d'évaluer plus avant l'influence de ces mises-à-jour sur le fonctionnement du modèle. Nous avons examiné leurs impacts sur l'initialisation et le calage du modèle. Il faut explorer ces questions sur d'autres structures hydrologiques.

Et si nous recommencions ?

Deux points me sautent aux yeux à la relecture de ces dernières pages. La première est que les travaux présentés doivent énormément à toutes les personnes, au sein de mon équipe de recherche comme à l'extérieur, qui m'ont apporté leur aide et m'ont accordé du temps. La seconde est l'écho de l'introduction qui résonne dans ces perspectives. Nous pouvons faire nôtres les paroles de M. l'Ingénieur (Fig. 15.9, page suivante).

FIGURE 15.9 – Une du *Cri de Paris* du dimanche 13 février 1910 (après la crue centennale de la Seine à Paris).

Bibliographie

An article a day keeps the Doctor away.

Abrahart, R., Heppenstall, A. et See, L. (2007). Timing error correction procedure applied to neural network rainfall-runoff modelling. *Hydrological Sciences Journal*, 52(3): 414–431.

Aitken, A. (1973). Assessing systematic errors in rainfall-runoff models. *Journal of Hydrology*, 20(2):121–136.

Anctil, F., Michel, C., Perrin, C. et Andréassian, V. (2004a). A soil moisture index as an auxiliary ANN input for stream flow forecasting. *Journal of Hydrology*, 286(1-4):155–167.

Anctil, F., Perrin, C. et Andréassian, V. (2003). ANN output updating of lumped conceptual rainfall/runoff forecasting models. *Journal of the American Water Resources Association*, 39(5):1269–1279.

Anctil, F., Perrin, C. et Andréassian, V. (2004b). Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. *Environmental Modelling & Software*, 19:357 – 368.

Anctil, F. et Rat, A. (2005). Evaluation of neural network streamflow forecasting on 47 watersheds. *Journal of Hydrologic Engineering*, 10(1):85–88.

Andréassian, V., Hall, A., Chahinian, N. et Schaake, J. (2006). Introduction and synthesis : Why should hydrologists work on a large number of basin data sets ? *IAHS-AISH Publication*, n° 307:1–5.

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C. et Loumagne, C. (2004). Impact of a spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models : a theoretical study using chimera watersheds. *Water Resources Research*, 40(5):1 – 9.

Andréassian, V., Parent, E. et Michel, C. (2003). A distribution-free test to detect gradual changes in watershed behavior. *Water Resources Research*, 39(9):SWC101–SWC1011.

ANDRÉASSIAN, V., PERRIN, C., MICHEL, C., USART-SANCHEZ, I. et JACQUES, L. (2001). Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology*, 250:206 – 223.

ANQUETIN, S., YATES, E., DUCROCQ, V., SAMOUILLAN, S., CHANCIBAULT, K., DAVOLIO, S., ACCADIA, C., CASAIOLI, M., MARIANI, S., FICCA, G., GOZZINI, B., PASI, F., PASQUI, M., GARCIA, A., MARTORELL, M., ROMERO, R. et CHESSA, P. (2005). The 8 and 9 september 2002 flash flood event in France : A model intercomparison. *Natural Hazards and Earth System Science*, 5(5):741–754.

ARULAMPALAM, M. S., MASKELL, S., GORDON, N. et CLAPP, T. (2002). A tutorial on particle fillers for on-line non-linear/non-gaussian bayssian tracking. *IEEE Transactions on Signal Processing*, 50(2):174 – 188.

ASKEW, A. J. (1989). Real-time intercomparison of hydrological models. *In IAHS Symposium, Baltimore, IAHS Publication n° 181, 125–132.*

AUBERT, D., LOUMAGNE, C. et OUDIN, L. (2003). Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall – runoff model. *Journal of Hydrology*, 280(1-4):145–161.

BACHET (1934). Note sur la propagation et l'annonce des crues. Rapport technique, Annales des Ponts et Chaussées iii(34), 409-465.

BÁRDOSSY, A. et SINGH, S. (2008). Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences Discussions*, 5(3):1641–1675.

BERGSTRÖM, S. (1991). Principles and confidence in hydrological modelling. *Nordic Hydrology*, 22:123–136.

BERGSTRÖM, S. et FORSMAN, A. (1973). Development of a conceptual deterministic rainfall-runoff model. *Nordic Hydrology*, 4:147–170.

BERGSTRÖM, S. et GRAHAM, L. P. (1998). On the scale problem in hydrological modelling. *Journal of Hydrology*, 211:253–265.

BERTHET, L. (2006). Quelle donnée de précipitation pour un modèle global de prévision pluie – débit ? Mémoire de D.E.A., Master Hydrologie – Hydrogéologie, AgroParisTech.

BERTHET, L., ANDRÉASSIAN, V., PERRIN, C. et JAVELLE, P. (2009a). How crucial is it to account for the antecedent moisture conditions in flood forecasting ? comparison of event-based and continuous approaches on 178 catchments. *Hydrology and Earth System Sciences*, 13(6):819–831.

BERTHET, L., PERRIN, C., ANDRÉASSIAN, V., TANGARA, M., LOUMAGNE, C., RAMOS, M.-H., MOINE, N. L., LERAT, J. et JAVELLE, P. (2008). Comment passer d'un modèle hydrologique à un système de prévision des crues ? écueils liés à la structure des modèles et aux échelles d'espace et de temps. *In Prévisions hydrométéorologiques*, pages 135 – 144, 119e colloque de la Société hydrotechnique de France, Lyon (France).

BERTHET, L., RAMOS, M., PERRIN, C., ANDRÉASSIAN, V. et LOUMAGNE, C. (2009b). Can discharge assimilation methods be used to improve flood forecasting when few data are available ? *IAHS-AISH Publication*, N° 333:94 – 100.

BEVEN, K. (2001). How far can we go in distributed hydrological modelling ? *Hydrology and Earth System Sciences*, 5(1):1 – 12.

BEVEN, K. (2006). On undermining the science ? *Hydrological Processes*, 20(14):3141–3146.

BEVEN, K. (2008a). On doing better hydrological science. *Hydrological Processes*, 22(17): 3549–3553.

BEVEN, K. et BINLEY, A. (1992). The future of distributed models : model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298.

BEVEN, K. et FREER, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1-4):11–29.

BEVEN, K. et KIRKBY, M. (1979). Physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin – Bulletin des Sciences hydrologiques*, 24(1):43–69.

BEVEN, K., SMITH, P. et FREER, J. (2008). So just why would a modeller choose to be incoherent ? *Journal of Hydrology*, 354(1-4):15–32.

BEVEN, K. J. (2008b). *Environmental Modelling : An Uncertain Future ?* Routledge, London, New York.

BLÖSCHL, G. (2006). Hydrologic synthesis : Across processes, places, and scales. *Water Resources Research*, 42(3):W03S02.

BONTRON, G. (2004). *Prévision quantitative des précipitations : adaptation probabiliste par recherche d'analogues. Utilisation des réanalyses NCEP / NCAR et application aux précipitations du Sud-Est de la France.* Thèse de doctorat, INPG.

BOURQUI, M. (2008). *Impact de la variabilité spatiale des pluies sur les performances des modèles hydrologiques.* Thèse de doctorat, ENGREF (Paris) et Cemagref (Antony).

BOX, G. E. P. et JENKINS, G. M. (1976). *Time Series Analysis : Forecasting and Control.* Holden Day Inc., Oakland, California.

BRATH, A. et ROSSO, R. (1993). Adaptive calibration of a conceptual model for flash flood forecasting. *Water Resources Research*, 29(8):2561–2572.

BURN, D. et BOORMAN, D. (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology*, 143(3-4):429–454.

CALVO, B. et SAVI, F. (2009). Real-time flood forecasting of the Tiber river in Rome. *Natural Hazards*, 50(3):461–477.

CAMPOLO, M., ANDREUSSI, P. et SOLDATI, A. (1999). River flood forecasting with a neural network model. *Water Resources Research*, 35(4):1191–1197.

CELLULE D'ANIMATION SANDRE ET GROUPE RÉFÉRENTIEL RNDE (2002). Référentiel hydrologique. Rapport technique, Ministère de l'Aménagement du territoire et de l'Environnement. http ://sandre.eaufrance.fr/ftp/sandre/francais/document/inter_th/ddd/sandre_present_refhydrographique_v2002-1.pdf.

CHO, H. et OLIVERA, F. (2009). Effect of the spatial variability of land use, soil type, and precipitation on streamflows in small watersheds. *Journal of the American Water Resources Association*, 45(3):673–686.

CIGIZOGLU, H. (2003). Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrological Sciences Journal*, 48(3):349–362.

CLARKE, R. (2008). Issues of experimental design for comparing the performance of hydrologic models. *Water Resources Research*, 44(1):W01409.

COLLIER, C. (2007). Flash flood forecasting : What are the limits of predictability ? *Quarterly Journal of the Royal Meteorological Society*, 133(622):3–23.

COMMISSION D'ENQUÊTE DE L'ASSEMBLÉE NATIONALE (2001). *Rapport de la commission d'enquête sur les causes des inondations répétitives ou exceptionnelles et sur les conséquences des intempéries afin d'établir les responsabilités, d'évaluer les coûts ainsi que la pertinence des outils de prévention, d'alerte et d'indemnisation.* Numéro 3386.

CORRADINI, C., MELONE, F. et UBERTINI, L. (1987). Some remarks on real time flood forecasting by simplified semi-distributed models. *In IAHS Publication (International Association of Hydrological Sciences)*, numéro 164, pages 121–127, IRPI, Perugia, Italy, IRPI, Perugia, Italy.

COULIBALY, P., ANCTIL, F. et BOBÉE, B. (1999). Prévision hydrologique par réseaux de neurones artificiels : état de l'art. *Canadian Journal of Civil Engineering*, 26(3):293–304.

CROW, W. T. et RYU, D. (2008). A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals. *Hydrol. Earth Syst. Sci. Discuss.*, 5(4):2005–2044.

DA ROS, D. et BORGA, M. (1997). Adaptive use of a conceptual model for real time flood forecasting. *Nordic Hydrology*, 28(3):169–188.

DAWDY, D. (2007). Prediction versus understanding (the 2006 Ven Te Chow lecture). *Journal of Hydrologic Engineering*, 12(1):1–3.

DAWDY, D. et BERGMANN, J. (1969). Effect of rainfall variability of streamflow simulation. *Water Resources Research*, 5(5):958–966.

DAWSON, C., ABRAHART, R. et SEE, L. (2007). HydroTest : A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, 22(7):1034–1052.

DE ROO, A. P., GOUWELEEUW, B., THIELEN, J., BARTHOLMES, J., BONGIOANNINI-CERLINI, P., TODINI, E., BATES, P. D., HORRITT, M., HUNTER, N., BEVEN, K., andErdmann HEISE, F. P., RIVIN, G., HILS, M., HOLLINGSWORTH, A., HOLST, B., KWADIJK, J., REGGIANI, P., DIJK, M. V., SATTLER, K. et SPROKKEREEF, E. (2003). Development of a european flood forecasting system. *International Journal River Basin Management*, 1(1):49 – 59.

DUAN, Q., SOROOSHIAN, S. et GUPTA, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4):1015 – 1031.

DUBAND, D., OBLED, C. et RODRIGUEZ, J. (1993). Unit hydrograph revisited : an alternate iterative approach to UH and effective precipitation identification. *Journal of Hydrology*, 150(1):115–149.

EDIJATNO (1991). *Mise au point d'un modèle élémentaire Pluie–Débit au pas de temps journalier*. Thèse de doctorat, Université Louis Pasteur de Strasbourg et CEMAGREF.

ELSHORBAGY, A., SIMONOVIC, S. et PANU, U. (2000). Performance evaluation of artificial neural networks for runoff prediction. *Journal of Hydrologic Engineering*, 5(4):424–427.

ESTUPINA BORRELL, V., CHORDA, J. et DARTUS, D. (2005). Flash-flood anticipation. *Comptes Rendus - Geoscience*, 337(13):1109–1119.

EVENSEN, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143 – 10162.

EVENSEN, G. (2003). The ensemble Kalman filter : Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367.

FENICIA, F., SAVENIJE, H., MATGEN, P. et PFISTER, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, 44(1):W01402, doi :10.1029/2006WR005563.

FIRAT, M. (2008). Comparison of artificial intelligence techniques for river flow forecasting. *Hydrology and Earth System Sciences*, 12(1):123–139.

FLOOD, I. et KARTAM, N. (1994). Neural networks in civil engineering. I : Principles and understanding. *Journal of Computing in Civil Engineering*, 8(2):131–148.

FORESEE, F. et HAGAN, M. T. (1997). Gauss-Newton approximation to Bayesian learning. *In IEEE International Conference on Neural Networks - Conference Proceedings*, volume 3, pages 1930–1935.

FORTIN, V., OUARDA, T. B. M. J., RASMUSSEN, P. F. et BOBÉE, B. (1997). Revue bibliographique des méthodes de prévision des débits. *Revue des Sciences de l'Eau*, 10(4):461–487.

GALLAGHER, M. et DOHERTY, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software*, 22(7):1000–1020.

GAN, T. Y., DLAMINI, E. M. et BIFTU, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, 192:81 – 103.

GARÇON, R., HOUDANT, B., GARAVAGLIA, F., MATHEVET, T., PAQUET, E. et GAILHARD, J. (2008). Expertise humaine des prévisions hydrométéorologiques et communication de leurs incertitudes dans un contexte décisionnel. *In Prévisions hydrométéorologiques*, pages 115 – 126, 119^{e} colloque de la Société hydrotechnique de France, Lyon (France).

GAUME, E. et GOSSET, R. (2003). Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology ? *Hydrology and Earth System Sciences*, 7(5):693–706.

GAUTHIER, P., COURTIER, P. et MOLL, P. (1993). Assimilation of simulated wind Lidar data with a Kalman filter. *Monthly Weather Review*, 121(6):1803–1820.

GEORGAKAKOS, K. P. (1986). Generalized stochastic hydrometeorological model for flood and flash-flood forecasting : 2. case studies. *Water Resources Research*, 22(13):2096–2106.

GIUSTOLISI, O. et LAUCELLI, D. (2005). Improving generalization of artificial neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal*, 50(3):439–457.

GORDON, N., SALMOND, D. et SMITH, A. (1993). Novel approach to nonlinear/non-gaussian Bayesian state estimation. *IEE Proceedings, Part F : Radar and Signal Processing*, 140(2):107–113.

GOSWAMI, M., O'CONNOR, K., BHATTARAI, K. et SHAMSELDIN, A. (2005). Assessing the performance of eight real-time updating models and procedures for the Brosna River. *Hydrology and Earth System Sciences*, 9(4):394–411.

GUPTA, H., SOROOSHIAN, S. et YAPO, P. (1998). Toward improved calibration of hydrologic models : Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4):751–763.

GUPTA, H., WAGENER, T. et LIU, Y. (2008). Reconciling theory with observations : Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22(18): 3802–3813.

HABETS, F., LEMOIGNE, P. et NOILHAN, J. (2004). On the utility of operational precipitation forecasts to served as input for streamflow forecasting. *Journal of Hydrology*, 293(1-4):270–288.

HABIB, E., ADUVALA, A. et MESELHE, E. (2008). Analysis of radar-rainfall error characteristics and implications for streamflow simulation uncertainty. *Hydrological Sciences Journal*, 53(3):568–587.

HAY, L., MCCABE, G., CLARK, M. et RISLEY, J. (2009). Reducing streamflow forecast uncertainty : Application and qualitative assessment of the upper Klamath river basin, Oregon. *Journal of the American Water Resources Association*, 45(3):580–596.

HORNIK, K., STINCHCOMBE, M. et WHITE, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

HOUDANT, B. (2004). *Contribution à l'amélioration de la prévision hydrométéorologique opérationnelle. Pour l'usage des probabilités dans la communication entre acteurs*. Thèse de doctorat, École Nationale du Génie Rural, des Eaux et Forêts.

HSU, K.-L., GUPTA, H. et SOROOSHIAN, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31(10):2517–2530.

HUGHES, D. (1993). Variable time intervals in deterministic hydrological models. *Journal of Hydrology*, 143(3-4):217–232.

IMRIE, C., DURUCAN, S. et KORRE, A. (2000). River flow prediction using artificial neural networks : Generalisation beyond the calibration range. *Journal of Hydrology*, 233(1-4):138–153.

JAIN, A. et KUMAR, S. (2009). Dissection of trained neural network hydrologic models for knowledge extraction. *Water Resources Research*, 45(7):W07420.

JAKEMAN, A. J. et HORNBERGER, G. M. (1993). How much complexity is warranted in a rainfall-runoff model ? *Water Resources Research*, 29(8):2637 – 2649.

JAVELLE, P. (2008). Inter-comparaison des modèles de prévision des crues développés au Cemagref : GR3H et GR3P. Convention 2008 Cemagref / Ministère de l'Écologie et du Développement Durable - Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations, Rapport Cemagref, 23 pages.

KALTEH, A. M. et HJORTH, P. (2009). Imputation of missing values in a precipitation-runoff process database. *Hydrology Research*, 40(4):420–432.

KARLSSON, M. et YAKOWITZ, S. (1987). Nearest-neighbor methods for nonparametric rainfall runoff forecasting. *Water Resources Research*, 23(7):1300–1308.

KAVETSKI, D., KUCZERA, G. et FRANKS, S. (2006a). Bayesian analysis of input uncertainty in hydrological modeling : 1. theory. *Water Resources Research*, 42(3):W03407.

KAVETSKI, D., KUCZERA, G. et FRANKS, S. (2006b). Bayesian analysis of input uncertainty in hydrological modeling : 2. application. *Water Resources Research*, 42(3):W03408.

KERR, R. A. (2007). River-level forecasting shows no detectable progress in 2 decades. *Science*, 316(5831):1555.

KHAN, M. (1989). Evaluation of rainfall-discharge models with discrimination. *Journal of Hydrology*, 108(C):63–78.

KHU, S., LIONG, S.-Y., BABOVIC, V., MADSEN, H. et MUTTIL, N. (2001). Genetic programming and its application in real-time runoff forecasting. *Journal of the American Water Resources Association*, 37(2):439–451.

KIM, S. et KIM, H. S. (2008). Uncertainty reduction of the flood stage forecasting using neural networks model. *Journal of the American Water Resources Association*, 44(1):148 – 165.

KIRCHNER, J. (2006). Getting the right answers for the right reasons : Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3):W03S04, doi :10.1029/2005WR004362.

KITANIDIS, P. et BRAS, R. (1980a). Real-time forecasting wih a conceptual hydrologic model. 1. analysis of uncertainty. *Water Resources Research*, 16(6):1025–1033.

KITANIDIS, P. et BRAS, R. (1980b). Real-time forecasting with a conceptual hydrologic model. 2. applications and results. *Water Resources Research*, 16(6):1034–1044.

KLEMEŠ, V. (1986). Operational testing of hydrologic simulation models. *Hydrological Sciences Journal*, 31(1):13 – 24.

KÁLMÁN, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

KNUTH, D. E. (1984). *The TEXbook, Volume A, de Computers and Typesetting*. Addison-Wesley, Reading, Mass., 2e édition. ISBN 0-201-13448-9.

KOBOLD, M. et BRILLY, M. (2006). The use of HBV model for flash flood forecasting. *Natural Hazards and Earth System Science*, 6(3):407–417.

KRZYSZTOFOWICZ, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9):2739–2750.

KRZYSZTOFOWICZ, R. et MARANZANO, C. (2004). Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology*, 293(1-4):57–73.

KUCZERA, G. (1988). On the validity of first-order prediction limits for conceptual hydrologic models. *Journal of Hydrology*, 103(3-4):229–247.

KUCZERA, G. et PARENT, E. (1998). Monte-Carlo assessment of parameter uncertainty in conceptual catchment models : The metropolis algorithm. *Journal of Hydrology*, 211(1-4):69–85.

KUHN, T. (1962, 1969). *The Structure of Scientific Revolutions*. University of Chicago Press.

LAMBLIN, R. (2009). Prévision des crues aux confluences : impact de la spatialisation sur les performances des modèles hydrologiques de prévision. Mémoire de D.E.A., INP Toulouse et Cemagref.

LAMPORT, L. (1994). *LATEX : A Document Preparation System*. Addison-Wesley, Reading, Mass., 2e édition. ISBN 0-201-52983-1.

LATEX 3 PROJECT TEAM (1997). LATEX2ε for authors.

LAUZON, N., ANCTIL, F. et PETRINOVIC, J. (2004). Characterization of soil moisture conditions at temporal scales from a few days to annual. *Hydrological Processes*, 18(17): 3235–3254.

LAVABRE, J. (2005). AIGA : un dispositif d'alerte des crues sur l'ensemble du réseau hydrographique. *In Exposé au Conseil scientifique et technique du Cemagref.*

LE DIMET, F.-X. et TALAGRAND, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus, Series A*, 38 A(2):97–110.

LE MOIGNE, P. (2002). Description de l'analyse des champs de surface sur la france par le système SAFRAN. Note technique de Météo France.

LE MOINE, N. (2008). *Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles Pluie – Débit ?* Thèse de doctorat, Université Pierre et Marie Curie (Paris VI).

LE MOINE, N., ANDRÉASSIAN, V., PERRIN, C. et MICHEL, C. (2007). How can rainfall-runoff models handle intercatchment groundwater flows ? theoretical study based on 1040 French catchments. *Water Resources Research*, 43:W06428.

LEKKAS, D. F. (2008). Using complementary methods for improved flow forecasting. *Hydrological Sciences Journal*, 53(4):696 – 705.

LERAT, J. (2009). *Quels apports hydrologiques pour les modèles hydrauliques ? Vers un modèle intégré de simulation des crues.* Thèse de doctorat, Université Pierre et Marie Curie, Paris.

LESHNO, M., LIN, V. Y., PINKUS, A. et SCHOCKEN, S. (1993). Multilayer feedforward networks with a non polynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867.

LIN, G.-F. et CHEN, G.-R. (2008). A systematic approach to the input determination for neural network rainfall-runoff models. *Hydrological Processes*, 22(14):2524–2530.

LINDSTRÖM, G., JOHANSSON, B., PERSSON, M., GARDELIN, M. et BERGSTRÖM, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201:272–288.

LINSLEY, R. (1982). Rainfall – Runoff models – an overview. *In* SINGH, V., éditeur : *Rainfall – Runoff relationship,proceedings of the international symposium on Rainfall – Runoff modeling*, pages 3 – 22, Littleton, CO. Water Resources Publications.

LIU, Y. et GUPTA, H. (2007). Uncertainty in hydrologic modeling : Toward an integrated data assimilation framework. *Water Resources Research*, 43(7):W07401.

LJUNG, L. (1979). Asymptotic behavior of the Extended Kalman Filter as a parameter estimator for linear systems. *IEEE Trans Autom Control*, AC-24(1):36–50.

MACKAY, D. J. C. (1992a). Bayesian interpolation. *Neural Computation*, 4:415–447.

MACKAY, D. J. C. (1992b). A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4:448–472.

MADSEN, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, 235(3-4):276–288.

MADSEN, H., BUTTS, M., KHU, S. et LIONG, S. (2000). Data assimilation in rainfall-runoff forecasting. *In Proceedings of the 4th Hydroinformatics Conference*. Iowa, USA. IAHR.

MADSEN, H. et SKOTNER, C. (2005). Adaptive state updating in real-time river flow forecasting - a combined filtering and error forecasting procedure. *Journal of Hydrology*, 308(1-4):302–312.

MAIER, H. et DANDY, G. (2000). Neural networks for the prediction and forecasting of water resources variables : A review of modelling issues and applications. *Environmental Modelling and Software*, 15(1):101–124.

MANTOVAN, P. et TODINI, E. (2006). Hydrological forecasting uncertainty assessment : Incoherence of the GLUE methodology. *Journal of Hydrology*, 330(1-2):368–381.

MATHEVET, T. (2005). *Quels modèles Pluie-Débit globaux au pas de temps horaire ? Développements empiriques et intercomparaison de modèles sur un large échantillon de bassins versants.* Thèse de doctorat, École nationale du Génie rural, des Eaux et des Forêts (ENGREF).

MATHEVET, T., MICHEL, C., ANDRÉASSIAN, V. et PERRIN, C. (2006). A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS-AISH Publication*, n° 307:211–219.

MCDONNELL, J., SIVAPALAN, M., VACHÉ, K., DUNN, S., GRANT, G., HAGGERTY, R., HINZ, C., HOOPER, R., KIRCHNER, J., RODERICK, M., SELKER, J. et WEILER, M. (2007). Moving beyond heterogeneity and process complexity : A new vision for watershed hydrology. *Water Resources Research*, 43(7):W07301.

MCLAUGHLIN, D. (1995). Recent developments in hydrologic data assimilation. *Reviews of Geophysics*, 33(SUPPL. 2):977–984.

MCLAUGHLIN, D. (2002). An integrated approach to hydrologic data assimilation : Interpolation, smoothing, and filtering. *Advances in Water Resources*, 25(8-12):1275–1286.

MEIN, R. et BROWN, B. (1978). Sensitivity of optimized parameters in watershed models. *Water Resources Research*, 14(2):299–303.

MERZ, R. et BLÖSCHL, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1-4):95–123.

MERZ, R., PARAJKA, J. et BLÖSCHL, G. (2009). Scale effects in conceptual hydrological modeling. *Water Resources Research*, 45:W09405.

MICHEL, C. (1991). *Hydrologie appliquée aux petits bassins ruraux*. Cemagref.

MICHEL, C., PERRIN, C. et ANDRÉASSIAN, V. (2003). The exponential store : A correct formulation for rainfall-runoff modelling. *Hydrological Sciences Journal*, 48(1):109–124.

MICHEL, C., PERRIN, C., ANDRÉASSIAN, V., OUDIN, L. et MATHEVET, T. (2006). Has basin-scale modelling advanced beyond empiricism ? *IAHS-AISH Publication*, n° 307: 211–219.

MICOVIC, Z. et QUICK, M. C. (2009). Investigation of the model complexity required in runoff simulation at different time scales. *Hydrological Sciences Journal*, 54(5):872 – 885.

MILLER, R., GHIL, M. et GAUTHIEZ, F. (1994). Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056.

MINNS, A. W. et HALL, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal*, 41:399–417.

MONTANARI, A. et GROSSI, G. (2008). Estimating the uncertainty of hydrological forecasts : A statistical approach. *Water Resour. Res.*, 44:W00B08.

MONTANARI, A., SHOEMAKER, C. A. et van de GIESEN, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology : An overview of issues and challenges. *Water Resources Research*, 45:W00B00.

MOORE, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1):483–499.

MOORE, R. J., BELL, V. A. et JONES, D. A. (2005). Forecasting for flood warning. *Comptes Rendus Geosciences*, 337(1-2):203–217.

MORADKHANI, H., HSU, K.-L., GUPTA, H. et SOROOSHIAN, S. (2005a). Uncertainty assessment of hydrologic model states and parameters : Sequential data assimilation using the Particle Filter. *Water Resources Research*, 41(5):1–17.

MORADKHANI, H., SOROOSHIAN, S., GUPTA, H. et HOUSER, P. (2005b). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2):135–147.

MOUELHI, S. (2003). *Modélisation pluie-débit aux pas de temps mensuel, annuel et pluri-annuel*. Thèse de doctorat, Géosciences et Ressources naturelles, ENGREF.

MOULIN, L., GAUME, E. et OBLED, C. (2009). Uncertainties on mean areal precipitation : assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.*, 13(2):99–114.

MUNIER, S. (2009). *Modélisation intégrée des écoulements pour la gestion en temps réel d'un bassin versant anthropisé*. Thèse de doctorat, Cemagref.

NALBANTIS, I. (1995). Use of multiple-time-step information in rainfall-runoff modeling. *Journal of Hydrology*, 165(1-4):135–159.

NASH, J. E. et SUTCLIFFE, J. V. (1970). River Flow Forecasting through conceptual models : part I - A : discussion of principles. *Journal of Hydrology*, 10:282–290.

NATIONAL HYDROLOGIC WARNING COUNCIL (2002). Use and benefits of the National Weather Service River and Flood Forecasts. Rapport technique, EASPE, Inc.

NATIONAL RESEARCH COUNCIL (NRC) (2002). *Report of a Workshop on Predictability & Limits-To-Prediction in Hydrologic Systems*. Numéro 0-309-08347-8. National Academic Press, Washington, D.C.

NEAL, J., ATKINSON, P. et HUTTON, C. (2007). Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. *Journal of Hydrology*, 336(3-4):401–415.

NORBIATO, D., BORGA, M., DEGLI ESPOSTI, S., GAUME, E. et ANQUETIN, S. (2008). Flash flood warning based on rainfall thresholds and soil moisture conditions : An assessment for gauged and ungauged basins. *Journal of Hydrology*, 362(3-4):274–290.

OBLED, C., ZIN, I. et HINGRAY, B. (2008). Choix des pas de temps et d'espace pour des modélisations parcimonieuses en hydrologie des crues. *In Prévisions hydrométéorologiques*, 119e colloque de la Société hydrotechnique de France, Lyon (France).

O'CONNOR, K. M. (2006). *River Basin Modelling for Flood Risk Mitigation*, chapitre The Galway Real-Time River Flow Forecasting System (GFFS). Taylor & Francis / Balkema, Leiden, the Netherlands.

OUACHANI, R., BARGAOUI, Z. et OUARDA, T. (2007). Intégration d'un filtre de Kalman dans le modèle hydrologique HBV pour la prévision des débits. *Hydrological Sciences Journal*, 52(2):318–337.

OUDIN, L., ANDRÉASSIAN, V., MATHEVET, T., PERRIN, C. et MICHEL, C. (2006a). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42(7):W07410.

OUDIN, L., ANDRÉASSIAN, V., PERRIN, C. et ANCTIL, F. (2004). Locating the sources of low-pass behavior within rainfall-runoff models. *Water Resources Research*, 40(11): W1110101–W1110114.

OUDIN, L., HERVIEU, F., MICHEL, C., PERRIN, C., ANDRÉASSIAN, V., ANCTIL, F. et LOUMAGNE, C. (2005a). Which potential evapotranspiration input for a lumped rainfall-runoff model ? part 2 – towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, 303:290 – 306.

OUDIN, L., MICHEL, C. et ANCTIL, F. (2005b). Which potential evapotranspiration input for a lumped rainfall-runoff model ? part 1 – can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs ? *Journal of Hydrology*, 303:275 – 289.

OUDIN, L., PERRIN, C., MATHEVET, T., ANDRÉASSIAN, V. et MICHEL, C. (2006b). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology*, 320(1-2):62–83.

PALMER, T., BUIZZA, R., HAGEDOEN, R., LAWRENCE, A., LEUTBECHER, M. et SMITH, L. (2006). Ensemble prediction : A pedagogical perspective. *ECMWF Newsletter*, 106:10–07.

PAPPENBERGER, F. et BEVEN, K. (2006). Ignorance is bliss : Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(5):W05302.

PATUREL, J. E., SERVAT, E. et VASSILIADIS, A. (1995). Sensitivity of conceptual rainfall-runoff algorithms to errors in input data - case of the GR2M model. *Journal of Hydrology*, 168:111 – 125.

PAUWELS, V. R. N. et DE LANNOY, G. J. M. (2009). Ensemble-based assimilation of discharge into rainfall-runoff models : A comparison of approaches to mapping observational information to state space. *Water Resour. Res.*, 45:W08428.

PERRIN, C. (2000). *Vers une amélioration d'un modèle Pluie–Débit au travers d'une approche comparative*. Thèse de doctorat, INPG.

PERRIN, C., ANDRÉASSIAN, V., ROJAS SERNA, C., MATHEVET, T. et LE MOINE, N. (2008). Discrete parameterization of hydrological models : Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research*, 44:W08447.

PERRIN, C., CORON, L., FORTIER-FILION, T.-C., MUNIER, S., LITRICO, X., BERTHET, L., LERAT, J., PESCHARD, J., JAVELLE, P. et ANDRÉASSIAN, V. (2009). Analyse et modélisation des sous-bassins amont du bassin de la seine, rapport de fin d'étude. Rapport technique 76 p., Cemagref, UR HBAN (Antony) et UMR G-Eau (Montpellier), DIREN Ile-de-France.

PERRIN, C., MICHEL, C. et ANDRÉASSIAN, V. (2001). Does a large number of parameters enhance model performance ? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242:275 – 301.

PERRIN, C., OUDIN, L., ANDRÉASSIAN, V., ROJAS-SERNA, C., MICHEL, C. et MATHEVET, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. *Hydrological Sciences Journal*, 52(1):131–151.

POPPER (1934). *Logik der Forschung*. Wien.

R DEVELOPMENT CORE TEAM (2006). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

RAMOS, M.-H., BARTHOLMES, J. et POZO, J. T.-d. (2007). Development of decision support products based on ensemble forecasts in the European flood alert system. *Atmospheric Science Letters*, 8(4):113–119.

RANDRIANASOLO, A., RAMOS, M., THIREL, G., ANDRÉASSIAN, V. et MARTIN, E. (2010). Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmospheric Science Letters*, pages accepté, en attente de publication.

REFSGAARD, J. C. (1997). Validation and intercomparison of different updating procedures for real-time forecasting. *Nordic Hydrology*, 28(2):65 – 84.

REFSGAARD, J. C. et HENRIKSEN, H. J. (2004). Modelling guidelines - terminology and guiding principles. *Advances in Water Resources*, 27:71 – 82.

REICHLE, R., MCLAUGHLIN, D. et ENTEKHABI, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114.

RODWELL, M. J. (2006). Comparing and combining deterministic and ensemble forecasts : How to predict rainfall occurrence better. *ECMWF Newsletter*, 106:17–23.

RUMELHART, D. E., HINTON, G. E. et WILLIAMS, R. J. (1986). *Parallel distributed processing : explorations in the microstructure of cognition, vol. 1 : foundations*, volume 1, chapitre Learning internal representations by error propagation, pages 318–362. MIT Press, Cambridge, MA, USA.

SALAMON, P. et FEYEN, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, 376(3-4):428–442.

SARLE, W. (1994). Neural networks and statistical models. *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, pages 1538–1550.

SATYAJI RAO, Y. et KRISHNA, B. (2009). Modeling hydrological time series data using wavelet neural network analysis. *In* YILMAZ, K. K., YUCEL, I., GUPTA, H. V., WAGENER, T., YANG, D., SAVENIJE, H., NEAL, C., KUNSTMANN, H. et POMEROY, J., éditeurs : *New Approaches to Hydrological Prediction in Data-sparse Regions*, numéro 333 de IAHS-AISH Publication, pages 101 – 111.

SAVENIJE, H. H. G. (2009). HESS Opinions "The art of hydrology". *Hydrology and Earth System Sciences*, 13(2):157–161.

SCHAEFER, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and Forecasting*, 5(4):570–575.

SCHREIDER, S., YOUNG, P. et JAKEMAN, A. (2001). An application of the Kalman filtering technique for streamflow forecasting in the Upper Murray Basin. *Mathematical and Computer Modelling*, 33(6-7):733–743.

SEFTON, C. et HOWARTH, S. (1998). Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology*, 211(1-4):1–16.

SEIBERT, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes*, 15:1063–1064.

SEIBERT, J. et BEVEN, K. (2009). Gauging the ungauged basin : how many discharge measurements are needed ? *Hydrology and Earth System Sciences Discussions*, 6(2):2275–2299.

SEO, D.-J., KOREN, V. et CAJINA, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology*, 4(3):627–641.

SHAMSELDIN, A. (1997). Application of a neural network technique to rainfall-runoff modelling. *Journal of Hydrology*, 199(3-4):272–294.

SHAMSELDIN, A. et O'CONNOR, K. (1996). A nearest neighbour linear perturbation model for river flow forecasting. *Journal of Hydrology*, 179(1-4):353–375.

SHAMSELDIN, A., O'CONNOR, K. et LIANG, G. (1997). Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, 197(1-4):203–229.

SHAMSELDIN, A. Y. (2006). *River Basin Modelling for Flood Risk Mitigation*, chapitre Real-time river flow forecasting, pages 181 – 195. Taylor & Francis / Balkema, Leiden, the Netherlands.

SINGH, V. P., éditeur (1995). *Computer Models of Watershed Hydrology*. Water Resources Publications.

SINGH, V. P. et FREVERT, D. K., éditeurs (2002). *Mathematical Models of Small Watershed Hydrology and Applications*. Water Resources Publications.

SIVAKUMAR, B. (2008). Undermining the science or undermining Nature ? *Hydrological Processes*, 22(6):893–897.

SIVAPALAN, M. (2003). Process complexity at hillslope scale, process simplicity at the watershed scale : is there a connection ? *Hydrological Processes*, 17(5):1037–1041.

SIVAPALAN, M., BLÖSCHL, G., ZHANG, L. et VERTESSY, R. (2003a). Downward approach to hydrological prediction. *Hydrol. Process.*, 17:2101 – 2111.

SIVAPALAN, M., TAKEUCHI, K., FRANKS, S., GUPTA, V., KARAMBIRI, H., LAKSHMI, V., LIANG, X., MCDONNELL, J., MENDIONDO, E., O'CONNELL, P., OKI, T., POMEROY, J., SCHERTZER, D., UHLENBROOK, S. et ZEHE, E. (2003b). IAHS decade on Predictions in Ungauged Basins (PUB), 2003-2012 : Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6):857–880.

SMITH, A. et GELFAND, A. (1992). Bayesian statistics without tears : A sampling-resampling perspective. *The American Statistician*, 46(2):84–88.

SMITH, M. B., GEORGAKAKOS, K. P. et LIANG, X. (2004). The distributed model intercomparison project (DMIP) : motivation and experiment design. *Journal of Hydrology*, 298:4–26.

SMITH, P., BEVEN, K. et TAWN, J. (2008). Detection of structural inadequacy in process-based hydrological models : A particle-filtering approach. *Water Resources Research*, 44(1):W01410.

SOLOMATINE, D., MASKEY, M. et SHRESTHA, D. (2008). Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes*, 22(2): 275–287.

SOROOSHIAN, S. et GUPTA, V. K. (1983). Automatic calibration of conceptual rainfall-runoff models : the question of parameter observability and uniqueness. *Water Resources Research*, 19(1):260 – 268.

SOROOSHIAN, S. et GUPTA, V. K. (1995). *Computer Models of Watershed Hydrology*, chapitre Model Calibration, pages 23 – 68. Water Resources Publications.

STORM, B., JENSEN, K. H. et REFSGAARD, J. C. (1989). Estimation of catchment rainfall uncertainty and its influence on runoff prediction. *Nordic Hydrology*, 19:77–88.

SUGARAWA, M. (1993). On the weights of precipitation stations. *In Advances in theorical hydrology*. O'Kane J.P. editor.

TAHIR, A. (2008). Flood forecasting at short and large lead times : how to choose the best adapted model ? Mémoire de D.E.A., Master Hydrologie – Hydrogéologie, Université *Pierre et Marie Curie* (Paris VI).

TANGARA, M. (2005). *Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global*. Thèse de doctorat, École pratique des hautes études de Paris.

TETZLAFF, D., MCDONNELL, J. J., UHLENBROOK, S., MCGUIRE, K. J., BOGAART, P. W., NAEF, F., BAIRD, A. J., DUNN, S. M. et SOULSBY, C. (2008). Conceptualizing catchment processes : simply too complex ? *Hydrological Processes*, 22(11):1727–1730.

THIELEN, J., BARTHOLMES, J., RAMOS, M.-H. et DE ROO, A. (2009). The European Flood Alert System - Part 1 : Concept and development. *Hydrology and Earth System Sciences*, 13(2):125 – 140.

THIEMANN, M., TROSSET, M., GUPTA, H. et SOROOSHIAN, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, 37(10):2521–2535.

THOM, R. (1993). *Prédire n'est pas expliquer*. Flammarion, Paris.

TOTH, E. et BRATH, A. (2007). Multistep ahead streamflow forecasting : Role of calibration data in conceptual and neural network modeling. *Water Resources Research*, 43(11): W11405.

TROUTMAN, B. (1985). Errors and parameter estimation in precipitation-runoff modeling 1. theory. *Water Resources Research*, 21(8):1195–1213.

TUCCI, C. et CLARKE, R. (1980). Adaptive forecasting with a conceptual rainfall-runoff model. *Hydrological forecasting. Proc. Oxford symposium, April 1980, (International Association of Hydrological Sciences, Washington DC ; IAHS-AISH Publication 129)*, pages 445–454.

VALERIANO, O. C. S., KOIKE, T., YANG, D., NYUNT, C. T., van KHANH, D. et CHAU, N. L. (2009). Flood simulation using different sources of rainfall in the Huong River, Vietnam. *Hydrological Sciences Journal*, 54(5):909–917.

VELÁZQUEZ, J. A., PETIT, T., LAVOIE, A., BOUCHER, M.-A., TURCOTTE, R., FORTIN, V. et ANCTIL, F. (2009). An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting. *Hydrology and Earth System Sciences*, 13(11):2221 – 2231.

VRUGT, J., DIKS, C., GUPTA, H., BOUTEN, W. et VERSTRATEN, J. (2005). Improved treatment of uncertainty in hydrologic modeling : Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41(1):1–17.

WAGENER, T., BOYLE, D., LEES, M., WHEATER, H., GUPTA, H. et SOROOSHIAN, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1):13–26.

WAGENER, T. et MCINTYRE, N. (2005). Identification of rainfall-runoff models for operational applications. *Hydrol. Sci. J.*, 50(5):735–751.

WAGENER, T., MCINTYRE, N., LEES, M., WHEATER, H. et GUPTA, H. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling : Dynamic identifiability analysis. *Hydrological Processes*, 17(2):455–476.

WAGENER, T., WHEATER, H. et GUPTA, H. V., éditeurs (2004). *Rainfall-runoff Modelling in Gauged and Ungauged Catchments*. Imperial College Press.

WALKER, J. et HOUSER, P. (2005). *Advances in Water Science Methodologies*, chapitre Hydrologic data assimilation, Chapter 2, pages 25 – 48. Éditeur : Taylor and Francis, The Netherlands.

WANG, D. et CAI, X. (2008). Robust data assimilation in hydrological modeling - a comparison of Kalman and H-infinity filters. *Advances in Water Resources*, 31(3):455–472.

WANG, Y., HE, B. et TAKASE, K. (2009). Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge. *Hydrological Sciences Journal*, 54(5):886 – 898.

WEERTS, A. et EL SERAFY, G. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, 42(9):W09403.

WĘGLARCZYK, S. (1998). The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology*, 206(1-2):98–103.

WEILER, M. et NAEF, F. (2003). Simulating surface and subsurface initiation of macropore flow. *Journal of Hydrology*, 273(1-4):139–154.

WELLES, E., SOROOSHIAN, S., CARTER, G. et OLSEN, B. (2007). Hydrologic verification : A call for action and collaboration. *Bulletin of the American Meteorological Society*, 88(4):503–511.

WERNER, M. et CRANSTON, M. (2009). Understanding the value of radar rainfall nowcasts in flood forecasting and warning in flashy catchments. *Meteorological Applications*, 16(1): 41–55.

WHEATER, H., JAKEMAN, A. et BEVEN, K. (1993). *Modelling change in environmental systems*, chapitre Progress and directions in rainfall-runoff modelling, pages 101 – 132. John Wiley & sons, Chichester.

WÖHLING, T., LENNARTZ, F. et ZAPPA, M. (2006). Technical note : Updating procedure for flood forecasting with conceptual HBV-type models. *Hydrology and Earth System Sciences*, 10(6):783–788.

WMO (1992). Simulated real-time intercomparison of hydrological models. Operational Hydrology Report 38, World Meteorological Organization, Geneva, Switzerland.

XIONG, L. et O'CONNOR, K. (2002). Comparison of four updating models for real-time river flow forecasting. *Hydrological Sciences Journal*, 47(4):621–640.

YANG, J., REICHERT, P., ABBASPOUR, K. et YANG, H. (2007). Hydrological modelling of the Chaohe Basin in China : Statistical model formulation and Bayesian inference. *Journal of Hydrology*, 340(3-4):167–182.

YANG, X. (1993). *Mise au point d'une méthode d'utilisation d'un modèle Pluie–Débit conceptuel pour la prévision des crues en temps réel.* Thèse de doctorat, Ecole nationale des Ponts et Chaussées.

YANG, X. et MICHEL, C. (2000). Flood forecasting with a watershed model : A new method of parameter updating. *Hydrological Sciences Journal*, 45(4):537–546.

YAPO, P. O., GUPTA, H. V. et SOROOSHIAN, S. (1996). Automatic calibration of conceptual rainfall-runoff models : sensitivity to calibration data. *Journal of Hydrology*, 181:23 – 48.

YOUNG, P. (2002). Advances in real-time flood forecasting. *Philosophical Transactions : Mathematical, Physical and Engineering Sciences (Series A)*, 360(1796):1433–1450.

ZEALAND, C., BURN, D. et SIMONOVIC, S. (1999). Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*, 214(1-4):32–48.

Annexes

A

Annexe

Glossaire

Littérature – Occupation des oisifs

Flaubert, *Dictionnaire des idées reçues*

Les lectures variées et les discussions qui ont enrichi ce travail tout au long de ces trois dernières années m'ont permis de constater que de nombreuses incompréhensions qui persistent entre les différents acteurs de la prévision ou entre scientifiques, ne sont dues qu'à des définitions différentes (Refsgaard et Henriksen, 2004).

Pour limiter les pertes de temps, je donne ici quelques définitions qui ont cours dans ce document. Ces définitions reflètent certaines convictions, voire certains *a priori* : je ne prétends en aucune manière qu'elles soient meilleures que d'autres ; elles ne sont données que pour faciliter la lecture de ce document. On trouvera des définitions de référence dans le *dictionnaire français d'hydrologie* [1] proposé par la commission de terminologie du Comité national français des Sciences hydrologiques (CNFSH).

A

Apprentissage. Pour les réseaux de neurones artificiels (RNA), terme qui désigne le calage du réseau. L'apprentissage peut être *supervisé* ou *non supervisé* selon que le RNA a pour tâche de simuler une sortie ou de rendre compte de régularités ou de motifs dans les données.

Assimilation de données. De façon générale, méthode visant à faire la jonction entre les données et le modèle (Liu et Gupta, 2007), c'est-à-dire entre nos connaissances pratiques du système étudié et nos connaissances théoriques. Plus particulièrement ici, emploi de données différentes des données d'entrée pour affiner les prévisions établies par le modèle, que ce soit par modification de la représentation du système faite par ce modèle ou par correction de ses sorties.

B

Bassin versant. Zone géographique drainée par le réseau hydrographique vers un point bas, nommé exutoire(g), qui le définit. Toute goutte d'eau précipitée à l'intérieur du bassin versant finit par atteindre l'exutoire (si elle ne s'évapore pas ni ne s'infiltre). Le bassin versant de surface est donc délimité par les lignes de crête. Il peut différer du bassin versant hydrogéologique : c'est le cas si les limites des aquifères ne coïncident pas avec les lignes de crête.

C

Calage (du modèle). Processus d'estimation des valeurs des paramètres(g) (vus alors comme des degrés de liberté) par confrontation du modèle avec des données. L'évolution des paramètres peut, le cas échéant, être limitée dans une gamme finie, fonction de l'interprétation physique ou conceptuelle des paramètres.

Conditions initiales. États du modèle au démarrage d'une simulation. En prévision, les conditions initiales sont les états à l'instant de prévision.

1. En ligne sur http://91.121.162.160/glu/indexdic.htm.

Continu (fonctionnement). Fonctionnement permettant de s'affranchir de la détermination des conditions initiales (à l'instant de prévision). Ces dernières sont fournies par l'état du modèle aux pas de temps antérieurs. Pour que ces conditions initiales soient indépendantes de celles spécifiées au premier temps de modélisation, il est nécessaire que le modèle ait fonctionné pendant une assez longue période (dite période de chauffe ou *warm-up*) précédant l'instant de prévision. Le fonctionnement continu s'oppose au fonctionnement événementiel[g].

Contrôle. Test du modèle sur des données qui n'ont pas été employées pour sa mise au point ou pour son calage. Le contrôle ne conduit pas à une « vérification » du modèle (Refsgaard et Henriksen, 2004) mais offre la possibilité d'évaluer sa pertinence et / ou, le cas échéant, de le réfuter.

Correction des sorties. Une des approches de mise-à-jour du modèle définies par Refsgaard (1997). Elle emploie les propriétés statistiques des erreurs du modèle pour les réduire.

Courbe hypsométrique. Courbe donnant la fraction de la surface d'un bassin située en dessous d'une altitude z en fonction de z.

D

Données d'entrée du modèle. Variables externes au modèle (i.e., qui ne constituent pas des états de celui-ci) dont la prise en compte par ce dernier modifie ses états (internes).

Distribué (modèle). Modèle dont la résolution spatiale est plus fine que celle du bassin versant (à la différence du modèle global[g]) et s'appuie sur une maille géométrique ou sur des *unités de réponse hydrologique*, considérées comme homogènes. Les entrées du modèle sont adaptées à cette maille et les flux entre les mailles doivent être explicités.

E

Équifinalité. Situation dans laquelle les performances de différents modèles (ou combinaison de modèles et de paramètres) ne peuvent pas être ordonnées avec l'emploi des données disponibles et du critère de performance (ou le modèle d'erreur) choisi : il n'est pas possible d'affirmer qu'un modèle est meilleur que les autres.

États (du modèle). Quantités internes au modèle variables au cours du temps. Certains états peuvent être liés à des grandeurs physiques sur le bassin versant comme l'humidité du sol mais expliciter ce lien peut être délicat pour nombre d'états de modèles hydrologiques. De ce fait, les états caractérisent d'abord le fonctionnement du modèle et parfois le fonctionnement du bassin versant.

État (d'une unité au sein d'un réseau de neurones). Variable (le plus souvent dans $\mathbb{R}$) caractérisant tout neurone d'un réseau. L'état d'un neurone est fonction des informations entrantes.

Événementiel (fonctionnement). Fonctionnement requérant que les conditions initiales du modèle (par exemple, l'humidité du bassin versant) soient précisées par le modélisateur, ce qui, le plus souvent, n'est pas une tâche aisée. Le fonctionnement événementiel s'oppose au fonctionnement continu(g).

Exutoire (d'un bassin versant de surface). Point du bassin versant par lequel s'écoule (sous forme de débit) toute l'eau collectée sous forme de précipitation sur le bassin versant qui ne s'est ni évaporée, ni infiltrée.

F

Fiabilité (d'une prévision par ensemble). Incapacité d'un observateur à distinguer entre observations[2] et réalisations d'un membre particulier (quelconque) de l'ensemble sur un grand nombre de pas de temps. La fiabilité caractérise la pertinence des intervalles de confiance calculés sur l'ensemble des prévisions.

Fonction d'activation. Pour les réseaux de neurones, la fonction déterminant l'état d'un neurone à partir de ses entrées.

Forçages. Variables mesurées pouvant être internes ou externes au modèle dont la prise en compte modifie ses états (terme plus souvent employé en météorologie).

G

Global (modèle). Un modèle global a pour unité spatiale le bassin versant (le plus souvent de surface). Il ne prend en entrée que les flux agrégés entrant et sortant de cette unité de travail. Cela ne lui interdit pas de définir des compartiments internes (conceptuels) et d'expliciter les flux entre eux (par exemple, les écoulements de surface, de subsurface, souterrrains...).

H

Hétéroscédasticité. Propriété statistique d'une variable aléatoire dont la variance n'est pas constante sur son domaine de définition. En hydrologie, les erreurs du modèle sont très souvent hétéroscédastiques : leur variance est fonction du débit. Les erreurs sur les débits les plus élevés ont une variance supérieure à celle des erreurs faites sur les étiages. La propriété opposée (homoscédasticité) constitue une hypothèse nécessaire à de nombreux modèles statistiques, c'est pourquoi les calculs statistiques sur les erreurs (dont les critères de performance) sont souvent effectués non pas avec les erreurs sur les débits mais avec les erreurs sur des transformées des débits (comme les transformations de Box-Cox).

2. Pour rester cohérent et garder le sens physique des modèles qui y prétendent, les observations sont rapportées à la résolution spatiale à laquelle le modèle fonctionne. Par exemple, il s'agit de la moyenne des pluies sur un bassin pour un modèle global.

Horizon de prévision. Durée H séparant l'instant t auquel la prévision est émise du pas de temps $t + H$ pour lequel cette prévision est donnée. On peut s'intéresser à tous les débits prévus sur cette plage temporelle ou uniquement à celui du pas de temps $t + H$.

Hyperparamètres. Paramètres du modèle d'erreur ou des procédures de mise-à-jour appelés ainsi pour les distinguer des paramètres du modèle hydrologique.

I

Inversibilité (d'un processus stochastique). Un processus stochastique est dit inversible si la réalisation x_t ne dépend que très faiblement (à la limite, pas) des réalisations passées au-delà d'un délai m : cette réalisation dépend essentiellement des réalisations x_{t-1}, $\ldots$, x_{t-m}.

L

Lent. Caractérise un bassin versant qui n'est pas rapide [(g)]. Si un horizon de prévision particulier est considéré, caractérise un bassin dont le temps de réaction est plutôt supérieur à cet horizon.

M

Mise-à-jour du modèle. Mise en œuvre de l'assimilation de données. Technique visant à modifier la valeur d'un élément du modèle (entrées, variables d'état, paramètres ou sorties) sur la base d'observations de nature différente de celle des entrées.

Modèle hydrologique Pluie – Débit. Modèle simulant la transformation de la pluie en débit sur le bassin versant dans un but scientifique ou pratique (comme la prédétermination des débits de crues extrêmes ou la prévision des crues et des étiages).

Modèles dits *à base physique*. Également appelés modèles *fondés sur la physique*. Modèle se basant sur une approche réductionniste vers des volumes élémentaires sur lesquels on peut employer les lois de la physique. Cette approche vise à une grande *profondeur explicative* (Beven, 2006). Elle doit faire face à la très grande hétérogénéité qui peut se rencontrer dans un bassin versant. En outre, cette approche pose la question des échelles pertinentes en fonction des objectifs du modélisateur, entre le bassin versant et le centimètre cube de terre dans lequel, par exemple, les vers de terre creusent des macropores, lieu d'écoulement préférentiel étudiés par certains hydrologues (e.g. Weiler et Naef, 2003). De nombreux processus hydrologiques sont non linéaires : l'agrégation des processus ne rend pas forcément compte de l'ensemble et il n'est pas suffisant de « valider » (ou de ne pas réfuter) chaque hypothèse prise séparément pour « valider » le modèle complet (Michel *et al.*, 2006).

Modèles conceptuels. Modèles se basant sur un certain nombre de principes de base (plus ou moins physiques) comme la conservation de la masse et de l'énergie, ou la description du bassin versant par fonctions (de production et de routage). La modélisation conceptuelle cherche à éviter les problèmes d'échelle des modèles à base physique en se concentrant sur les processus jugés dominants par les hydrologues et en employant des volumes de contrôle sur lesquels les paramètres et les variables d'état sont moyennés temporellement et spatialement (Kuczera et Parent, 1998). Selon ces auteurs, un modèle dont un ou plusieurs paramètres doivent être calés contre l'une des sorties du modèle, est conceptuel.

Modèles empiriques. Modèles dont la mise au point se base quasi uniquement sur l'expérimentation et la confrontation aux données sans employer d'affirmations théoriques préalables. Dans la pratique, de nombreux modèles conceptuels relèvent dans une large mesure de l'empirisme : l'expression des principes de base en équations est définie empiriquement, ce qui conduit Wheater *et al.* (1993) à définir une classe mixte : les *hybrid metric conceptual models*.

Modèle d'erreur. Ensemble des hypothèses sur les propriétés statistiques des erreurs du modèle (hydrologique). Il est nécessaire d'établir un modèle d'erreur pour proposer une fonction de vraisemblance (destinée à évaluer les performances du modèle).

Modèle hydrologique de prévision. Outil basé sur ou dérivant d'un modèle hydrologique utilisant les entrées de ce dernier (la pluie) et tout autre donnée utile comme les débits observés à l'exutoire, pour donner une estimation du débit prévu à un horizon de prévision donné.

P

Paramètres du modèle. Dans ce travail, grandeurs du modèle caractéristiques du fonctionnement du modèle sur un bassin versant et invariantes dans le temps. Ces grandeurs ne sont pas forcément aisément reliables à des quantités physiques ponctuelles ou volumiques. Dans cette définition, les conditions initiales ne sont pas des paramètres, mêmes pour des modèles où elles peuvent être calées. Nous distinguerons les paramètres libres qui doivent être calés pour chaque bassin, des paramètres fixes pour lesquels une valeur commune à l'ensemble des bassins peut être identifiée sans détériorer les performances du modèle.
La notion de paramètres est riche de plusieurs autres facettes et la définition qu'on en donne dépend dans une certaine mesure de la philosophie de modélisation choisie. La définition que nous avons proposée provient d'une approche systémique. Dans un autre cadre, Sivapalan *et al.* (2003a) définissent les paramètres comme étant des propriétés du « substrat » (*landscape properties*) qui gouvernent la réponse du bassin versant. Dans les modèles de type boîte-noire (ainsi que dans une moindre mesure pour tous les modèles qui admettent nécessiter une phase de calage), les paramètres jouent un rôle de degré de liberté du modèle, ce qui en limite le contenu interprétatif (physique).

Perceptrons. Réseaux de neurones organisés en couches de neurones, directs (i.e. sans boucle de rétro-actions) dont la fonction d'activation des neurones cachés est la composition d'une fonction logistique et d'une intégration des entrées :

$$X = f\left(\sum_{i=0}^{N_{\text{Entrées}}} w_{E_i \to X} \cdot E_i + b_X\right) \quad \text{avec} \quad f : x \mapsto \frac{1}{1+e^{-x}}$$

où les entrées du perceptron sont notées E_i. $w_{E_i \to X}$ désigne le poids de connexion du i^{e} neurone d'entrée vers le neurone caché X et b_X est le biais de ce neurone.

Performance(s). Mesure de l'adéquation des prédictions du modèle aux observations. Notion indéfinissable plus avant : s'il est vrai que tout hydrologue « a » son modèle, il a souvent plusieurs définitions de la performance de celui-ci.

Périodes. Pour les réseaux de neurones, le nombre de périodes (*epochs*) est le nombre de modifications des paramètres lors du calage du réseau pour une présentation de l'ensemble des données utilisées au calage. Par exemple, si l'ensemble des données de calage est divisée en deux et si les paramètres sont modifiés après une itération de recherche du meilleur jeu de paramètres Θ_{i+1} autour de Θ_i sur la première moitié de cet ensemble, puis après une itération avec la seconde moitié des données de calage, le nombre de périodes est égal à 2.

Précision (d'une prévision par ensemble). Distance entre une mesure de la tendance centrale de l'ensemble (mode, moyenne, médiane...) et la valeur observée (ou sa distribution si on tient compte d'une erreur de mesure qu'on peut expliciter.

Prédiction. Caractérise la situation d'un modèle travaillant sur des données qu'il n'a pas utilisées lors de son calage. Capacité à simuler le signal recherché sans forcément chercher à l'expliquer.

Prévoir. Désigne ici l'action d'émettre une prévision à une date t au sujet du débit (non encore observé à la date $t+H$. À cet effet, le prévisionniste ne peut utiliser que les données observées jusqu'à la date t et des scénarios / prévisions des entrées du modèle au-delà.

R

Rapide. Caractérise un bassin qui n'est pas lent(g). Si un horizon de prévision particulier est considéré, caractérise un bassin dont le temps de réaction est plutôt inférieur à cet horizon.

Régions hydrographiques. Premier niveau du découpage hydrographique qui est la partition du territoire français en fonction du réseau hydrographique (Cellule d'animation SANDRE et Groupe référentiel RNDE, 2002) : chaque région correspond au bassin versant d'un grand fleuve ou d'une grande rivière, ou encore à une zone côtière (Fig. A.1).

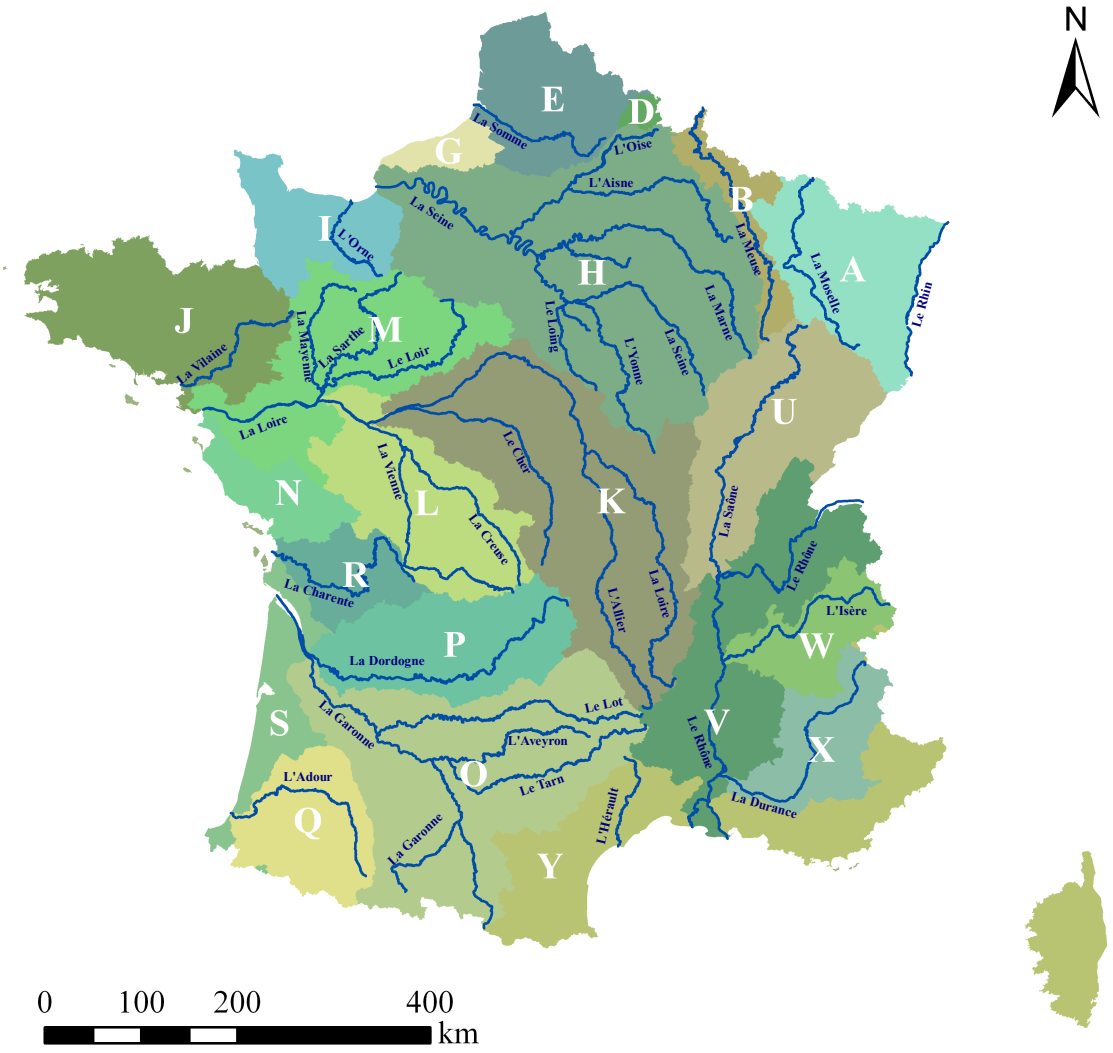


Figure A.1 – Régions hydrographiques françaises et leurs principaux cours d'eau.

Rétroaction (période de). Période précédant l'instant de prévision au début de laquelle la mise-à-jour d'un élément du modèle (entrées ou paramètres ou états) est réalisée. L'effet d'une mise-à-jour n'étant pas instantanée, la période de rétroaction permet au modèle de s'adapter à cette mise-à-jour et d'en tirer pleinement profit.

S

Séquentielle (approche). Approche qui met le modèle à jour à chaque fois que de nouvelles données peuvent être assimilées. Le plus souvent, la mise-à-jour n'emploie alors que les dernières observations.

Stationnarité. Un processus stochastique est dit stationnaire si la fonction de densité jointe des variables x_t, $f\left(x_t, x_{i+1}, \cdots, x_{t+j}\right)$, est indépendante de t. En pratique, on considérera stationnaire tout processus tel que l'espérance et la variance de la variable x_t ne dépendent pas du temps t.

Semi-distribué (modèle). Modèle qui détaille le bassin versant en sous-bassins versants considérées comme homogènes. Les entrées du modèle sont spécifiées pour chaque unité et les flux entre ces unités doivent être explicités.

V

Variationnelle (approche). Approche consistant à mettre à jour le modèle à l'instant de précision en employant alors toutes les données jugées pertinentes (i.e., éventuellement sur plusieurs pas de temps passés). Cette méthode tire son nom du fait qu'elle passe souvent par la minimisation d'une fonction de coût, opération réalisée en faisant varier les variables qu'on cherche à mettre à jour (par exemple, un recalage).

Vraisemblance (fonction de). Fonction qui associe au modèle (éventuellement à un jeu de paramètres) la densité de probabilité des observations conditionnellement à ce modèle (respectivement, ces paramètres).

Annexe

Notations

Cette annexe renferme l'ensemble des notations employées au fil de cette thèse. Certaines notations peuvent renvoyer à différentes grandeurs ou notions selon le contexte. Dans ce cas, les différentes significations sont indiquées avec si besoin des références aux chapitres où elles sont employées.

Notations	Significations	Où ?
A	Capacité du réservoir de production de GRP (voir l'annexe C, pages 340 et suivantes).	
$\mathbf{A}_t$	Matrice de prise en compte des entrées (dans le cadre du filtre de Kálmán).	Chap. 7
AM	Nombre d'alertes manquées (voir page 43).	
AV	Nombre d'alertes valables (voir page 43).	
a_t	chocs (variables aléatoires gaussiennes) dans les modèles ARIMA.	Chap. 10
B	Paramètre fixe de la fonction de percolation de GRP (voir l'annexe C, pages 340 et suivantes).	
	Opérateur *arrière* : $B\,z_t = z_{t-1}$.	Chap. 10
b	Paramètre du modèle PQb : coefficient de réduction des pluies.	Chap. 14
b_X	Biais du neurone X dans les réseaux de neurones.	Chap. 10
BFI	Indice d'écoulement de base.	Chap. 4
$C_{2M,P}$	Transformation du critère de persistance (voir page 43).	
$CRPS$	Critère de rang continu (*Continuous Rank Probability Score*, voir page 528).	
CRU	Coefficient de crue biennale.	Chap. 4
CSI	l'indice de succès critique (voir page 44).	
CSI'	Modification de l'indice de succès critique (voir page 44).	
$\mathcal{D}$	ensemble des exemples servant au calage (d'un réseau de neurones).	Ann. O
d	Degré de différentiation des modèles ARIMA.	Chap. 10
d_{RA}	Durée de la période de rétroaction pour la mise-à-jour des paramètres par la méthode de Yang (originale ou modifiée).	Chap. 9
E_i	i^{e} entrée du neurone X dans les réseaux de neurones.	Chap. 10
E_D	Fonction-objectif (somme quadratique des erreurs).	Ann. O
E_W	Somme quadratique des poids (paramètres) d'un réseau de neurones.	Ann. O
E_t	Évapotranspiration potentielle (également notée ETP) cumulée sur le pas de temps t.	
$E_t^{(n)}$	Évapotranspiration potentielle nette (variable interne de GRP, après la phase d'interception, voir pages 340 et suivantes).	
$E_t^{(s)}$	Évapotranspiration réelle calculée (variable interne de GRP, voir pages 340 et suivantes).	
$\mathbf{e}_t$	Vecteur des $\mathbf{e}_t$ du modèle au pas de temps t (formalisme de l'espace des états).	Chap. 7
ESM	Étiage modulaire.	Chap. 4
ETI	Débits spécifiques d'étiage (basé sur $\text{DCN}_{10,2}$).	Chap. 4

Notations	Significations	Où ?
$\mathbf{F}_t$	Matrice de l'opérateur f quand celui-ci est linéaire dans le strict cadre du filtre de Kálmán, ou matrice jacobienne de f calculée en $x_{t-1\|t-1}$ dans le cadre du filtre de Kálmán étendu.	Chap. 7
FA	Nombre de fausses alertes (voir page 43).	
FAR	Taux de fausses alertes (voir page 44).	
f	Opérateur du modèle propageant les états $\mathbf{x}_t$ vers les états $\mathbf{x}_{t+1}$.	Chap. 7
	Simplification de la précédente : fonction reliant R_t^+ au débit $\widehat{Q}_t$.	Chap. 7
H	Horizon de prévision.	
H_{max}	Horizon de prévision maximal sur un bassin donné tel que le scénario de pluies futures employé ne modifie pas les performances du modèle.	
$\mathbf{H}_t$	Matrice de l'opérateur h quand celui-ci est linéaire dans le strict cadre du filtre de Kálmán, ou linéarisation de h calculée en $x_{t\|t-1}$ dans le cadre du filtre de Kálmán étendu.	Chap. 7
$\mathbf{H}(\mathbf{w}, \alpha, \beta)$	matricienne hessienne de la fonction-objectif. (pour l'algorithme LMBR, page 561)	Ann. O
$HU_t(i)$	Stock d'eau en attente dans le i^{e} compartiment de l'hydrogramme unitaire de GRP.	
h	Opérateur d'observation reliant les états $\mathbf{x}_t$ aux observations $\mathbf{z}_t$ (formalisme de l'espace des états).	Chap. 7
$\mathbf{I}_k$	matrice identité d'ordre k.	Ann. O
IND	Irrégularité saisonnière des écoulements.	Chap. 4
IPB	Indice de pente.	Chap. 4
IRR	Irrégularité saisonnière des précipitations.	Chap. 4
$\mathbf{J}$	matrice jacobienne de la fonction-objectif (fonction des paramètres).	Ann. O
$K = A/B$	Coefficient de la fonction de percolation du modèle (voir l'annexe C, pages 340 et suivantes).	
$\mathbf{K}_t$	Matrice de gain (mise-à-jour des états par filtrage).	Chap. 7
k	Constante de vitesse dans les lois instantanées de vidange des réservoirs de type « Puissance ».	
	Nombre de paramètres libres du réseau de neurones.	Ann. O
LAM	Écoulement annuel moyen.	Chap. 4
MAE	Erreur absolue moyenne.	
MSE	Erreur quadratique moyenne (voir page 41).	
N	Nombre de particules contenues dans l'échantillon approximant la densité de probabilité des états.	Chap. 8
	Nombre d'exemples présentés au réseau de neurones lors de l'apprentissage.	Ann. O
NAV	Nombre de non alertes valables (voir page 43).	
NSE	Critère de Nash-Sutcliffe (voir page 41).	

Notations	Significations	Où ?
P_t	Précipitation cumulée sur le pas de temps t (dans GRP, pluie de bassin).	
$P_t^{(n)}$	Précipitation nette (variable interne de GRP, après la phase d'interception, voir pages 340 et suivantes).	
$P_t^{(s)}$	Part de la pluie dirigée vers le réservoir de production (variable interne de GRP, voir pages 340 et suivantes).	
$P_t^{(i,t)}$	Donnée de précipitation perturbée du i^{e} membre du filtre particulaire.	Chap. 8
$P_{\text{SAFRAN},d}$	Pluie estimée par le modèle météorologique SAFRAN pour le jour d.	Chap. 4
$P_{\text{SAFRAN},tot}$	Cumul de précipitation estimé par le modèle météorologique SAFRAN pour la période entière.	Chap. 4
$P_{\text{Pluviomètre},d}$	Cumul de précipitation mesuré par le réseau de pluviomètres au sol pour le jour d.	Chap. 4
$P_{\text{Pluviomètre},tot}$	Cumul de précipitation mesuré par le réseau de pluviomètres au sol pour la période entière.	Chap. 4
P_1	Première période de données.	
P_2	Deuxième période de données.	
$\mathbf{P}_{t\|t-1}$	Matrice de variance-covariance des erreurs des états estimés $\mathbf{x}_{t\|t-1}$.	Chap. 7
$\mathbf{P}_{t\|t}$	Matrice de variance-covariance des erreurs des états mis-à-jour $\mathbf{x}_{t\|t}$.	Chap. 7
p	Nombre de paramètres de la partie autorégressive des modèles ARIMA.	Chap. 10
	Densité de probabilité.	Chap. 8
PAM	Précipitation annuelle moyenne.	Chap. 4
$Perc_t$	Percolation du réservoir de production dans le modèle GRP (voir l'annexe C, pages 340 et suivantes).	
PI	Critère de persistance (voir page 41).	
PMX	10^{e} Pluie la plus intense biennale.	Chap. 4
POD	Probabilité de détection (voir page 44).	
PRT_t	Pluie efficace de GRP entrant dans la fonction de routage (voir pages 340 et suivantes).	
Q_m	Débit minimal au-dessus duquel le critère R_2 est calculé.	Chap. 2
Q_M	Débit maximal sur l'événement Ev lors du calcul du critère R_2 : $Q_M = min\left(\underset{Ev}{max}(Q), \underset{Ev}{max}\left(\widehat{Q}\right)\right)$	Chap. 2
Q_{max}	Sur une crue, le débit horaire maximal (débit de pointe).	
Q_t	Débit observé moyen sur le pas de temps t.	
$\widehat{Q}_{t+H\|t}$	Débit prévu à l'instant de prévision t à un horizon H. Si cette notation est employée quand on utilise un modèle sans et avec une fonction d'assimilation des données, elle réfère au débit prévu par le modèle sans mise-à-jour.	Chap. 10
$\widehat{\widehat{Q}}_{t+H\|t}$	Débit prévu à l'instant de prévision t à un horizon H par un modèle avec mise-à-jour.	

Notations	Significations	Où ?
$\widetilde{Q}^{Obs}$	Fonction croissante du débit observé qui lui est égale autant que possible, entrant dans le calcul du critère de retard R_2 (voir page 46).	
$\widetilde{Q}^{Calc}$	Fonction croissante du débit simulé qui lui est égale autant que possible, entrant dans le calcul du critère de retard R_2 (voir page 46).	
$Q^{(o)_t}$	Débit observé.	
$Q_t^{(p)}$	Débit prévu.	
$Q_t^{(r)}$	Débit réel (inconnu).	
$\mathbf{Q}_t$	Matrice de variance-covariance des erreurs dues au modèle $\mathbf{w}_t$ (formalisme de l'espace des états).	Chap. 7
q	Nombre de paramètres de la partie « moyenne mobile » des modèles ARIMA	Chap. 10
$q_x(Y)$	Quantile de niveau x (entre 0 et 1) de la grandeur Y.	
$\mathcal{R}$	Un réseau de neurones (en tant que modèle).	Ann. O
R_1	Critère événementiel de retard sur le pic de crue (voir page 45).	
$R_{1bis}(\alpha)$	Critère événementiel de retard sur le débit de crue $Q = \alpha Q_{max}$ (voir page 45).	
R_2	Critère événementiel de retard au dépassement de seuil en montée (voir page 46).	
R_t	Niveau du réservoir de routage après vidange au pas de temps t (voir pages 340 et suivantes).	
R_t^+	Niveau du réservoir de routage avant vidange au pas de temps t (voir pages 340 et suivantes).	
$R_{t\|t}^+$	Niveau du réservoir de routage avant vidange au pas de temps t mis-à-jour à ce pas de temps (voir pages 340 et suivantes).	
$\mathbf{R}_t$	Matrice de variance-covariance des erreurs d'observation $\mathbf{v}_t$ (formalisme de l'espace des états).	Chap. 7
S	Surface du bassin versant.	
S_i	Sensibilité des sorties du réseau de neurones à la i^{e} entrée : $S_i = \sum_{t=1}^{N_P} \left(\frac{\partial \widehat{\varepsilon}_{t+H\|t}}{\partial E_i(t)} \right)^2$.	Chap. 10
S_t	Niveau du réservoir de production à la fin du pas de temps t (voir pages 340 et suivantes).	
S_t^+	Niveau du réservoir de production intermédiaire, au pas du temps t (voir pages 340 et suivantes).	
$S_{t\|t}^+$	Niveau du réservoir de production intermédiaire, au pas du temps t mis-à-jour à ce pas de temps (voir pages 340 et suivantes).	
$\mathbf{S}_t$	Matrice de variance-covariance des erreurs sur les entrées $\mathbf{u}_t$ (formalisme de l'espace des états).	Chap. 7
SEC	Fraction annuelle moyenne des jours secs.	Chap. 4
Tr	Opérateur linéaire *Trace*.	

Notations	Significations	Où ?
t	Le temps. Souvent l'instant auquel la prévision est élaborée. Un pas de temps.	
t_{max}	Sur une crue, le pas de temps où l'observation du débit est maximale.	Chap. 3
$\hat{t}^H_{max}$	Sur une crue, le pas de temps pour lequel on prévoit (à un horizon H) le débit le plus élevé : $\widehat{Q}_{\hat{t}^H_{max}\mid\hat{t}^H_{max}-H} = \underset{t}{max}\left(\hat{Q}_{t\mid t-H}\right)$.	Chap. 3
$\mathbf{u}_t$	Vecteur des erreurs sur les entrées $\mathbf{e}_t$ du modèle au pas de temps t (formalisme de l'espace des états).	Chap. 7
$\mathbf{v}_t$	Vecteur du bruit sur les observations au pas de temps t (formalisme de l'espace des états).	Chap. 7
w_t	Variable différenciée (dans les modèles ARIMA) : $w_t = \nabla^d z_t$.	Chap. 10
$w_{E_i \to X}$	Poids de la connexion entre l'entrée E_i et le neurone X dans les réseaux de neurones.	Chap. 10
$w_t^{(i)}$	Poids associé à la i^{e} particule au pas de temps t (filtre particulaire).	Chap. 8
$\mathbf{w}$	Vecteur des poids (paramètres) d'un réseau de neurones.	Ann. O
$\mathbf{w}^{MP}$	Vecteur des poids (paramètres) optimisant la fonction-objectif.	Ann. O
$\mathbf{w}_t$	Vecteur des erreurs dues au modèle au pas de temps t (formalisme de l'espace des états).	Chap. 7
X_1	Paramètre du modèle GR3P contrôlant la capacité du réservoir de routage (mm, positif).	
X_2	Paramètre du modèle GR3P multiplicatif de la sortie de la fonction de production (sans dimension, positif).	
X_3	Paramètre du modèle GR3P contrôlant la durée de l'hydrogramme unitaire exprimée en nombre de pas de temps (ici en heures).	
$\mathbf{x}_t$	Vecteur des états du modèle au pas de temps t (formalisme de l'espace des états).	Chap. 7
$\mathbf{x}_{t\mid t-1}$	Vecteur des états calculés par le modèle au pas de temps t à partir des états au pas de temps précédents. Ces derniers peuvent éventuellement avoir été mis-à-jour grâce à des observations réalisés aux pas de temps $t-1$ ou antérieurs (formalisme de l'espace des états).	Chap. 7
$\mathbf{x}_{t\mid t}$	Vecteur des états estimé au pas de temps t à partir d'au moins une observation au pas de temps t, et éventuellement des états calculés par le modèle $\mathbf{x}_{t\mid t-1}$ (formalisme de l'espace des états).	Chap. 7
$\mathbf{x}_t^{(i)}$	Vecteur des états de la i^{e} particule (filtre particulaire).	Chap. 8
$\mathbf{z}_t$	Vecteur des observations au pas de temps t (formalisme de l'espace des états).	Chap. 7& 8

Notations	Significations	Où ?
α	Exposant commandant la forme de l'hydrogramme unitaire de GRP (voir pages 340 et suivantes).	
	Paramètre du modèle PQb : coefficient de décrue.	Chap. 14
	Un des deux paramètres de la régulation de l'algorithme de Levenberg-Marquardt, déterminés par inférence bayésienne.	Ann. O
	Rapport des variances des incertitudes des observations et des simulations du modèle.	Chap. 7 et 15
β	Exposant de la loi de vidange (de type « Puissance ») du réservoir de routage du modèle GRP (voir pages 340 et suivantes).	
	Un des deux paramètres de la régulation de l'algorithme de Levenberg-Marquardt, déterminés par inférence bayésienne.	Ann. O
γ	Exposant de la correction des erreurs de Tangara.	
γ_i	i^{e} ordonnée de l'hydrogramme unitaire du modèle GRP (i de 1 à $\lceil X_3 \rceil$).	
$\gamma_{i,t}$	Facteur multiplicatif de perturbation des pluies en entrée de GRP pour la i^{e} particule au pas de temps t (filtre particulaire).	Chap. 8
Δt	Durée de la période de perturbation des pluies en entrée de GRP par le filtre particulaire.	Chap. 8
	L'horizon de prévision (modèle PQb).	Chap. 14
$\delta(x)$	Distribution de Dirac (centrée en 0).	
δR_t	Modification de la valeur de l'état R_t pour sa mise-à-jour.	Chap. 7
$\delta HU(i)$	Modification de la valeur de l'état $HU(i)$ pour sa mise-à-jour.	Chap. 7
δPRT	Modification de la valeur de l'état PRT pour sa mise-à-jour.	Chap. 7
$\varepsilon_{t\|t'}$	Erreur de la prévision du débit à la date t, réalisée à la date t' : $\varepsilon_{t\|t'} = \widehat{Q}_{t\|t'} - Q_t$.	
ζ	Biais entre les pluies estimées par le modèle météorologique SAFRAN et les estimations basées sur les mesures faites par les pluviomètres au sol.	Chap. 4
η	Partie entière du temps de base de l'hydrogramme unitaire dans GRP : $\eta = \lfloor X_3 \rfloor$.	
Θ	Opérateur de la partie « moyenne mobile » des modèles ARIMA : $\Theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ soit $\Theta(B)(a_t) = 1 - \sum_{j=1}^{q} \theta_j a_{t-j}$.	Chap. 10
$\boldsymbol{\Theta}$	Vecteur des paramètres du modèle (formalisme de l'espace des états).	
Θ_0	Jeu de paramètres « de base » obtenu par calage du modèle sur tous les pas de temps de la période de calage.	Chap. 9
θ_j	Paramètres de la partie « moyenne mobile » des modèles ARIMA.	Chap. 10

Notations	Significations	Où ?
κ	Coefficient d'asymétrisation de l'hydrogramme unitaire.	
λ	Paramètre de répartition des flux entre les deux branches de la fonction de routage.	Chap. 6
ν	Paramètre du critère de performance CSI'.	Chap. 3 et 13
ξ	Niveau de confiance relative dans les observations et dans le modèle	Chap. 7
Π	Pluies précipitées entre $t - \Delta t'$ et t (formulation du modèle PQb).	Chap. 14
$\rho_{\Delta t}(X)$	Autocorrélation de la variable X pour un décalage temporel Δt.	
σ_Q	Écart-type des perturbations rajoutées par le filtre particulaire aux observations de débit assimilées par GRP.	Chap. 8
σ_M	Écart-type dans la fonction de vraisemblance permettant le calcul des poids des particules pour le filtre particulaire.	Chap. 8
Φ	Opérateur de la partie auto-régressive des modèles ARIMA : $\Phi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$ soit $\Phi(B)(z_t) = 1 - \sum_{i=1}^{p} \varphi_i z_{t-i}$.	Chap. 10
φ_i	Paramètres de la partie autorégressive des modèles ARIMA.	Chap. 10
$\omega_t^{(i)}$	Bruit associé à la i^{e} particule.	Chap. 8
∇	Opérateur de différentiation discrète : $\nabla z_t = z_t - z_{t-1}$.	

C

Annexe

Structure du modèle GRP

Le modèle GRP présente la même structure aux pas de temps journaliers et horaires (voir le chapitre 6). Nous distinguons ici la structure du modèle (*flowchart*) que nous présentons en premier, des mises-à-jour (nous détaillerons celles qui sont proposées par Tangara, 2005, par la suite). La dernière partie rappelle les valeurs des paramètres fixes pour différents pas de temps (de 1 à 24 heures).

C.1 Intégration de la loi de vidange d'un réservoir de type « Puissance »

Avant de présenter la structure du modèle GRP, rappelons la loi de vidange d'un réservoir de type « Puissance ». Cette loi intervient dans le modèle, au niveau de la percolation et du réservoir de routage.

En notant R le niveau du réservoir, la sortie instantanée Q d'un tel réservoir est décrite par

$$Q = -dR = k \cdot R^a \, dt \quad \text{où } a \neq 1 \tag{C.1}$$

On considère que le réservoir est alimenté au tout début du pas de temps t pour atteindre le niveau R_t^+ puis est vidangé selon la fonction de vidange précédente. Intégrons cette relation sur le pas de temps Δt :

$$\begin{aligned}
-\frac{dR}{R^a} &= k \; dt \\
\left[\frac{R^{1-a}}{a-1}\right]_t^{t+\Delta t} &= k \; \Delta t \\
\left[\frac{\left(R_t^+ - Q_t\right)^{1-a}}{a-1} - \frac{{R_t^+}^{1-a}}{a-1}\right] &= k \; \Delta t \\
\frac{1}{\left(R_t^+ - Q_t\right)^{a-1}} - \frac{1}{{R_t^+}^{a-1}} &= (a-1)\, k \; \Delta t \\
\frac{1}{\left(1 - \dfrac{Q_t}{R_t^+}\right)^{a-1}} &= 1 + (a-1)\, k \; \Delta t \cdot {R_t^+}^{a-1} \\
1 - \frac{Q_t}{R_t^+} &= \frac{1}{\left(1 + (a-1)\, k \; \Delta t \cdot {R_t^+}^{a-1}\right)^{\frac{1}{a-1}}} \\
Q_t &= R_t^+ \cdot \left(1 - \frac{1}{\left(1 + (a-1)\, k \; \Delta t \cdot {R_t^+}^{a-1}\right)^{\frac{1}{a-1}}}\right)
\end{aligned}$$

Soit en posant X_1, paramètre de dimension [L], défini par $X_1^{a-1} = \dfrac{1}{k\,(a-1)\,\Delta t}$

$$Q_t = f_{X_1,a}\left(R_t^+\right) = R_t^+ \cdot \left(1 - \frac{1}{\left(1 + \left(\frac{R_t^+}{X_1}\right)^{a-1}\right)^{\frac{1}{a-1}}} \right) \tag{C.2}$$

C.2 Structure hydrologique du modèle

Le modèle GRP est un modèle global. Ses entrées sont la pluie de bassin cumulée sur le pas de temps P_t et l'évapotranspiration (ETP) du bassin sur le pas de temps E_t. Dans la pratique, une ETP interannuelle suffit. Son fonctionnement peut se comprendre comme étant l'enchaînement d'une fonction de production et d'une fonction de routage. Nous décrivons la succession des opérations ayant lieu au pas de temps t. La figure C.1 représente la structure du modèle GRP.

C.2.1 Fonction de production

La fonction de production vise à calculer la *pluie efficace*. Il s'agit d'un module de suivi de l'humidité du bassin (approche connue sous le sigle SMA, *soil moisture accounting*).

Interception

Le modèle débute par une phase d'*interception* très simple qui calcule la pluie nette $P_t^{(n)}$ et l'évapotranspiration nette $E_t^{(n)}$:

$$\begin{cases} P_t^{(n)} = P_t - E_t \;\text{ et }\; E_t^{(n)} = 0 \;\text{ si }\; P_t \geq E_t \\ E_t^{(n)} = E_t - P_t \;\text{ et }\; P_t^{(n)} = 0 \;\text{ si }\; E_t \geq P_t \end{cases} \tag{C.3}$$

Fonctionnement du réservoir de production

Le flux net interagit avec le réservoir de production. Si la pluie nette est positive, une partie $P_t^{(s)}$ est dirigée vers le réservoir de production tandis que le reste $P_t^{(n)} - P_t^{(s)}$ s'écoule vers la fonction de routage (cf. infra). La fraction de la pluie nette stockée dans le réservoir est fonction du niveau de remplissage de ce réservoir (du taux d'humidité du bassin) :

$$P_t^{(s)} = A \cdot \frac{\left(1 - \left(\frac{S_{t-1}}{A}\right)^2\right) \cdot tanh\left(\frac{P^{(n)}}{A}\right)}{1 + \frac{S_{t-1}}{A} \cdot tanh\left(\frac{P_t^{(n)}}{A}\right)} \tag{C.4}$$

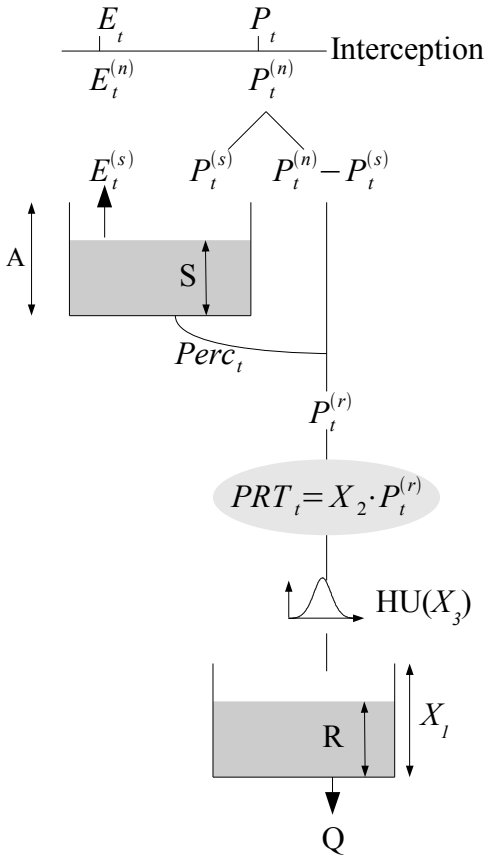


Figure C.1 – Structure du modèle GRP.

X_1	Capacité du réservoir de routage (mm)	libre
X_2	Facteur de l'ajustement multiplicatif de la pluie efficace (sans dimension)	libre
X_3	Temps de base de l'hydrogramme unitaire (h)	libre
A	Capacité du réservoir de production (mm)	fixe
K	Coefficient de la fonction de percolation (mm, peut également se mettre sous la forme $K = A/B$, où B est un paramètre fixe adimensionnel)	fixe
α	exposant commandant la forme de l'hydrogramme unitaire (sans dimension)	fixe
β	exposant de la loi de vidange du réservoir de routage (sans dimension)	fixe

Table C.1 – Récapitulatif des paramètres de la structure du modèle GRP.

où A est la capacité du réservoir de production, S_t son niveau au pas de temps t. Aucun volume ne quitte le réservoir par évapotranspiration : $E_t^{(s)} = 0$.

Dans le cas contraire où l'ETP nette est positive, une quantité d'eau $E_t^{(s)}$ s'échappe du réservoir de routage (et est perdue pour le modèle) :

$$E_t^{(s)} = \frac{S_{t-1} \cdot \left(2 - \frac{S_{t-1}}{A}\right) \cdot tanh\left(\frac{E_t^{(n)}}{A}\right)}{1 + \left(1 - \frac{S_{t-1}}{A}\right) \cdot tanh\left(\frac{E_t^{(n)}}{A}\right)} \tag{C.5}$$

Le réservoir n'enregistre aucune entrée : $P_t^{(s)} = 0$.

Les équations (C.4) et (C.5) correspondent à l'intégration de lois de remplissage et vidange instantanées du réservoir de production, quadratiques en S (Edijatno, 1991, annexe 31, pages 329 à 332 ; calculs également reportés dans la partie C.5 de cette annexe, page 345).

Le niveau du réservoir de production est modifié en conséquence :

$$S_t^+ = S_{t-1} - E_t^{(s)} + P_t^{(s)}$$

Ainsi, le réservoir de production n'est pas un réservoir de transit qui ralentirait le flux d'eau précipitée. Son action est plutôt celle d'un « compteur » mémoire de l'humidité du bassin qui module l'écoulement direct.

Percolation

Le réservoir perd de l'eau vers la fonction de routage, suivant une loi de puissance 5 : $dS = -k \cdot S^a \, dt$ avec $a = 5$. Cette fonction est traditionnellement appelée *percolation* sans que nous associons *stricto sensu* cette composante à la percolation des sols. Une fois cette loi intégrée sur le pas de temps (équation (C.2)), on a

$$Perc_t = S_t^+ \cdot \left(1 - \left(1 + \left(\frac{S_t^+}{K}\right)^4\right)^{-\frac{1}{4}}\right) \tag{C.6}$$

où K est lié au pas de temps par la relation

$$K = \left(k \, (a-1) \, \Delta t\right)^{\frac{1}{1-a}} \tag{C.7}$$

Ce coefficient s'écrit utilement (pour les modèles où la capacité du réservoir de production est un paramètre libre) sous la forme $K = A/B$ où B est un paramètre fixe du modèle.

Le réservoir de production atteint finalement le niveau :

$$S_t = S_t^+ - Perc_t \tag{C.8}$$

Ajustement du flux direct

L'eau percolée rejoint le flux direct pour subir un ajustement multiplicatif qui donne la pluie efficace entrant dans la fonction de routage :

$$PRT_t = X_2 \cdot \left(P_t^{(n)} - P_t^{(s)} + Perc_t \right) \tag{C.9}$$

où le coefficient d'ajustement du volume X_2 est un paramètre libre du modèle.

C.2.2 Fonction de routage

La fonction de routage va maintenant retarder la sortie de la pluie efficace PRT_t sur les prochains pas de temps. Cette fonction enchaîne un routage linéaire (par un hydrogramme unitaire) et un routage non linéaire (via un réservoir de routage).

Hydrogramme unitaire

Le flux direct PRT_t est convolué par un hydrogramme unitaire symétrique dont les ordonnées γ_i sont calculées par $\gamma_i = UHC(i) - UHC(i-1)$ où UHC est la fonction de l'hydrogramme unitaire cumulé définie par :

$$\begin{cases} UHC(i)=0 & \text{si } i \leq 0 \\ UHC(i)=\dfrac{i^\alpha}{i^\alpha + (X_3 - i)^\alpha} & \text{si } 0 < i < X_3 \\ UHC(i)=1 & \text{si } i \geq X_3 \end{cases} \tag{C.10}$$

où le temps de base de l'hydrogramme unitaire X_3 est un paramètre libre du modèle et l'exposant α est un paramètre fixe.

La sortie de l'hydrogramme unitaire s'écrit donc

$$quh_t = \sum_{i=1}^{\lceil X_3 \rceil} \gamma_i \cdot PRT_{t-i+1} \tag{C.11}$$

Cette quantité d'eau entre dans le réservoir de routage dont le niveau devient

$$R_t^+ = R_{t-1} + quh_t \tag{C.12}$$

Sortie du réservoir de routage

La sortie du réservoir de routage suit une loi de vidange de type puissance :

$$\widehat{Q} = -dR = k' \cdot R^\beta \, dt \tag{C.13}$$

Chez Tangara (2005), le réservoir est quadratique ($\beta = 2$) et la loi de vidange (équation (C.2)) donnant le débit sortant du modèle $\widehat{Q}_t$, se simplifie en :

$$\widehat{Q}_t = f_{X_1,\beta=2}\left(R_t^+\right) = \frac{{R_t^+}^2}{R_t^+ + X_1} \tag{C.14}$$

où la capacité du réservoir de routage X_1 est un paramètre libre du modèle.

Le niveau du réservoir de production devient :

$$R_t = R_t^+ - \widehat{Q}_t \tag{C.15}$$

C.3 Mises-à-jour proposées par Tangara (2005)

Tangara (2005) propose la combinaison de deux mises-à-jour. La première exploite la dernière observation du débit pour mettre à jour le niveau du réservoir de routage quand la seconde utilise la dernière erreur observée pour corriger la prévision à venir (en sortie du modèle).

C.3.1 Mise-à-jour du réservoir de routage

Il s'agit d'une mise-à-jour directe et « complète », discutée dans la partie 7.2.1 (page 109). Elle calcule le niveau du réservoir $R_{t|t}^+$ donnant le dernier débit observé Q_t. Pour un réservoir quadratique, il vient très facilement

$$R_{t|t}^+ = f_{X_1,\beta=2}^{-1}\left(Q_t\right) = \frac{\sqrt{Q_t^2 + 4X_1 \cdot Q_t} + Q_t}{2} \tag{C.16}$$

Le niveau du réservoir mis-à-jour après vidange s'en déduit directement :

$$R_{t|t} = f_{X_1,\beta=2}^{-1}\left(Q_t\right) = \frac{\sqrt{Q_t^2 + 4X_1 \cdot Q_t} - Q_t}{2} \tag{C.17}$$

Si le réservoir n'est pas quadratique, il est possible de réaliser la mise-à-jour de façon numérique (annexe K.1, page 505).

C.3.2 Correction finale du débit

La correction du débit proposée par Tangara (2005) est auto-régressive multiplicative :

$$\widehat{\widehat{Q}}_{t+H|t} = \widehat{Q}_{t+H|t} \cdot \left(\frac{Q_t}{\widehat{Q}_{t|t-1}}\right)^{\gamma} \tag{C.18}$$

où $\widehat{Q}_{t+H|t}$ est la prévision faite au pas de temps t pour un horizon H et $\widehat{\widehat{Q}}_{t+H|t}$ est cette prévision corrigée. γ est un paramètre fixe.

C.4 Valeurs des paramètres fixes à différents pas de temps

Le tableau suivant donne les valeurs des paramètres fixes optimisées par Tahir (2008) pour six pas de temps différents. Ce dernier a montré que les paramètres fixes du modèle GRP sont à peu près indépendants de l'horizon de prévision.

Δt	1 h	2 h	3 h	6 h	12 h	24 h
Capacité du réservoir de production A	275	275	275	275	300	325
Coefficient de la fonction de percolation B	5,25	4,75	3,50	3,00	2,50	2,50
Exposant de l'hydrogramme unitaire α	1,25	1,50	1,75	1,75	1,75	2,25
Exposant de la fonction de vidange du réservoir de routage β	2	–	–	2	–	2,5
Exposant de la correction finale γ	0,450	0,400	0,325	0,175	0,125	0,100

C.5 Intégration des équations du réservoir de production

Les équations (C.4) et (C.5) correspondent à l'intégration de lois instantanées de vidange et de remplissage du réservoir de production caractérisant un fonctionnement quadratique en S (Edijatno, 1991, annexe 3.1, pages 329 à 332).

C.5.1 Cas où la pluie nette est positive

La répartition de la pluie nette suit une loi instantanée quadratique en S :

$$dS = \left(1 - \left(\frac{S}{A}\right)^2\right) P^{(n)}\, dt \quad \text{et} \quad dP^{(s)} = \left(\frac{S}{A}\right)^2 P^{(n)}\, dt \tag{C.19}$$

Notons $p = \dfrac{P^{(n)}}{A}$ et $s = \dfrac{S}{A}$. L'équation précédente s'intègre sur le pas de temps t :

$$\begin{aligned} ds &= \left(1 - s^2\right) p\, dt \\ \frac{ds}{1 - s^2} &= p\, dt \\ \frac{ds}{1 - s} + \frac{ds}{1 + s} &= 2p\, dt \\ \ln\left(\frac{1 + s_t^+}{1 + s_{t-1}}\right) - \ln\left(\frac{1 - s_t^+}{1 - s_{t-1}}\right) &= 2p_t \end{aligned}$$

$$
\begin{aligned}
\frac{1+s_t^+}{1-s_t^+} &= \frac{1+s_{t-1}}{1-s_{t-1}} \cdot \mathrm{e}^{2p_t} = X \\
1+s_t^+ &= X - s_t^+ \cdot X \\
s_t^+ \cdot (1+X) &= X - 1 \\
s_t^+ &= \frac{\frac{1+s_{t-1}}{1-s_{t-1}}\mathrm{e}^{2p_t} - 1}{\frac{1+s_{t-1}}{1-s_{t-1}}\mathrm{e}^{2p_t} + 1} \\
s_t^+ &= \frac{(1+s_{t-1})\,\mathrm{e}^{2p_t} - (1-s_{t-1})}{(1+s_{t-1})\,\mathrm{e}^{2p_t} + (1-s_{t-1})} \\
s_t^+ &= \frac{\left(\mathrm{e}^{2p_t}-1\right) + s_{t-1}\left(\mathrm{e}^{2p_t}+1\right)}{\left(\mathrm{e}^{2p_t}+1\right) + s_{t-1}\left(\mathrm{e}^{2p_t}-1\right)} \\
s_t^+ &= \frac{tanh\,(p_t) + s_{t-1}}{1 + tanh\,(p_t)\,s_{t-1}}
\end{aligned}
$$

La formule (C.4) s'en déduit directement via $P_t^{(s)} = A \cdot \left(s_t^+ - s_{t-1}\right)$.

C.5.2 Cas où l'évapotranspiration nette est positive

La loi d'évolution instantanée du niveau du réservoir de production s'écrit dans ce cas :

$$
dS = -\frac{S}{A}\left(2 - \frac{S}{A}\right) E^{(n)}\, dt \tag{C.20}
$$

soit en notant $s = S/A$ et $e = E^{(n)}/A$:

$$
ds = -s\,(2-s)\,e\,dt
$$

qui s'intègre en

$$
\begin{aligned}
\frac{ds}{s\,(2-s)} &= -e\,dt \\
\frac{ds}{s} + \frac{ds}{2-s} &= -2e\,dt \\
\ln\left(\frac{s_t^+}{s_{t-1}}\right) - \ln\left(\frac{2-s_t^+}{2-s_{t-1}}\right) &= -2e_t \\
\frac{s_t^{+1}}{s_{t-1}} &= \frac{2-s_t^+}{2-s_{t-1}} \cdot \mathrm{e}^{-2e_t} \\
s_t^+ \cdot \left(\frac{1}{s_{t-1}} + \frac{\mathrm{e}^{-2e_t}}{2-s_{t-1}}\right) &= \frac{2}{2-s_{t-1}} \cdot \mathrm{e}^{-2e_t}
\end{aligned}
$$

$$
\begin{aligned}
s_t^+ &= \frac{\frac{2}{2-s_{t-1}} \cdot \mathrm{e}^{-2e_t}}{\frac{1}{s_{t-1}} + \frac{1}{2-s_{t-1}} \cdot \mathrm{e}^{-2e_t}} \\
s_t^+ &= \frac{2}{\frac{2-s_{t-1}}{s_{t-1}} \cdot \mathrm{e}^{2e_t} + 1} \\
s_t^+ &= \frac{2s_{t-1}}{s_{t-1} + (2-s_{t-1}) \cdot \mathrm{e}^{2e_t}}
\end{aligned}
$$

On va en déduire la différence $E_t^{(s)} = A \cdot \left(s_{t-1} - s_t^+\right) = A \cdot \dfrac{f\left(s_{t-1}\right)}{g\left(s_{t-1}\right)}$.

Le numérateur f s'écrit

$$
\begin{aligned}
f\left(s_{t-1}\right) &= s_{t-1}^2 + s_{t-1}\left(2-s_{t-1}\right)\mathrm{e}^{2e_t} - 2s_{t-1} \\
&= s_{t-1}^2 \cdot \left(1-\mathrm{e}^{2e_t}\right) + 2s_{t-1}\mathrm{e}^{2e_t} - 2s_{t-1} \\
&= 2s_{t-1}\left(\mathrm{e}^{2e_t}-1\right) - s_{t-1}^2\left(\mathrm{e}^{2e_t}-1\right) \\
&= s_{t-1}\left(2-s_{t-1}\right)\left(\mathrm{e}^{2e_t}-1\right) \\
&= s_{t-1}\left(2-s_{t-1}\right) tanh(e_t)\left(1+\mathrm{e}^{2e_t}\right)
\end{aligned}
$$

Le dénominateur g nous donne

$$
\begin{aligned}
\frac{g\left(s_{t-1}\right)}{1+\mathrm{e}^{2e_t}} &= \frac{s_{t-1} + (2-s_{t-1}) \cdot \mathrm{e}^{2e_t}}{1+\mathrm{e}^{2e_t}} \\
&= 1 + \frac{s_{t-1} + (2-s_{t-1}) \cdot \mathrm{e}^{2e_t} - 1 - \mathrm{e}^{2e_t}}{1+\mathrm{e}^{2e_t}} \\
&= 1 + \frac{(s_{t-1}-1) + (1-s_{t-1}) \cdot \mathrm{e}^{2e_t}}{1+\mathrm{e}^{2e_t}} \\
&= 1 + (1-s_{t-1})\, tanh(e_t)
\end{aligned}
$$

L'équation (C.5) en découle directement.

Annexe

Compléments sur les critères de performance

D.1 Lien entre critère de persistance et critère de Nash-Sutcliffe

Ces deux critères décrivent tous deux l'erreur quadratique du modèle mais ne sont pas équivalents puisqu'ils comparent les performances du modèle à celles de deux modèles naïfs différents qui ne réagissent pas de façon similaire sur tous les bassins versants. Il existe une relation entre les deux critères et l'auto-corrélation des débits du bassin versant considéré (qui « compare » les deux modèles naïfs) :

$$\frac{1-NS}{1-Pers} = \frac{\sum_i \left(Q_{i+H} - Q_i\right)^2}{\sum_i \left(Q_{i+H} - \overline{Q}\right)^2}$$

Or

$$\begin{aligned}\sum_i \left(Q_{i+H} - Q_i\right)^2 &= \sum_i \left(Q_{i+H} - \overline{Q} + \overline{Q} - Q_i\right)^2 \\ &\approx 2\sum_i \left(Q_i - \overline{Q}\right)^2 + 2\sum_i \left(Q_i - \overline{Q}\right)\left(Q_{i+H} - \overline{Q}\right)\end{aligned}$$

d'où il ressort que

$$\frac{1-NS}{1-Pers} \approx 2 - 2\cdot\frac{cov\left(Q_i, Q_{i+H}\right)}{Var(Q)} \approx 2\cdot\left(1-\rho_H(Q)\right)$$

où $\rho_H(Q)$ est l'auto-corrélation des débits observés à H pas de temps.

D.2 Usons de critères vraiment différents

L'étude présentée ici vise à évaluer l'intérêt de raffiner les critères de performance présentés au chapitre 3 pour évaluer au mieux les performances en prévision.

D.2.1 Méthodologie employée

L'étude se base sur des analyses en composantes principales (annexe I.2, page 486) réalisées sur tout ou partie de l'ensemble des 26 critères répertoriés dans le tableau D.1. Ces critères appartiennent à trois familles :

- critères de persistance,
- critères de retards,
- critères d'alerte CSI'.

Au sein d'une même famille, les critères diffèrent par les pas de temps pris en compte pour leur calcul et par les scénarios de pluie future employés.

N°	Critère de	calculé sur...	Scénario de pluies futures
1	Persistance	toute la période de contrôle	Observations *a posteriori*
2	Persistance	toute la période de contrôle	Pluies futures nulles
3	Retard	le franchissement de seuil en montée (R_2)	Observations *a posteriori*
4	Retard	le franchissement de seuil en montée (R_2)	Pluies futures nulles
5	Retard	le pic de crue (R_1)	Observations *a posteriori*
6	Retard	le pic de crue (R_1)	Pluies futures nulles
7	Retard	le dépassement de 90 % du débit de pointe (R_{1bis})	Observations *a posteriori*
8	Retard	le dépassement de 90 % du débit de pointe (R_{1bis})	Pluies futures nulles
9	Retard	le dépassement de 80 % du débit de pointe (R_{1bis})	Observations *a posteriori*
10	Retard	le dépassement de 80 % du débit de pointe (R_{1bis})	Pluies futures nulles
11	Persistance	les hautes eaux ($Q(t) \geq q[90](Q)$)	Observations *a posteriori*
12	Persistance	les hautes eaux ($Q(t) \geq q[98](Q)$)	Observations *a posteriori*
13	Persistance	les hautes eaux ($Q(t+H) \geq q[90](Q)$)	Observations *a posteriori*
14	Persistance	les hautes eaux ($Q(t+H) \geq q[98](Q)$)	Observations *a posteriori*
15	Persistance	les pluies fortes ($CP(t+H) \geq q[90](CP)$)	Observations *a posteriori.*
16	Persistance	les pluies fortes ($CP(t+H) \geq q[90](CP)$)	Observations *a posteriori*
17	Persistance	les pluies fortes ($CP(t) \geq q[90](CP)$)	Pluies futures nulles
18	Persistance	les pluies fortes ($CP(t) \geq q[98](CP)$)	Pluies futures nulles
19	Persistance	les montées de crues ($Q(t) > Q(t-1)$) au dessus du seuil $q[95](Q)$)	Observations *a posteriori*
20	Persistance	les montées de crues ($Q(t+H) > Q(t+H-1)$) au dessus du seuil $q[95](Q)$)	Observations *a posteriori*
21	Alerte $CSI'(\nu)$	le seuil $q[90](Q)$ avec $\nu = 0$ (POD)	Observations *a posteriori*
22	Alerte $CSI'(\nu)$	le seuil $q[98](Q)$ avec $\nu = 0$ (POD)	Observations *a posteriori*
23	Alerte $CSI'(\nu)$	le seuil $q[90](Q)$ avec $\nu = 0,5$	Observations *a posteriori*
24	Alerte $CSI'(\nu)$	le seuil $q[98](Q)$ avec $\nu = 0,5$	Observations *a posteriori*
25	Alerte $CSI'(\nu)$	le seuil $q[90](Q)$ avec $\nu = 1$ (CSI)	Observations *a posteriori*
26	Alerte $CSI'(\nu)$	le seuil $q[98](Q)$ avec $\nu = 1$ (CSI)	Observations *a posteriori*

TABLE D.1 – Critères de performance utilisés dans l'A.C.P.

Chaque vecteur V_i des performances mesurées par le critère i est normé :

$$\widetilde{V_i} = \frac{V_i - \overline{V_i}}{\sigma_{V_i}}$$

De cette façon, les gammes des valeurs prises par les performances des différents critères sont ramenées à un ordre de grandeur semblable, ce qui évite d'avoir une variance écrasée par un critère qui varierait sur une gamme de valeurs beaucoup plus large que celles des autres critères. En outre, la variance totale à expliquer vaut le nombre de composantes testées.

Trois A.C.P. ont été réalisées : la première sur tous les critères (26), la deuxième sur les critères calculés avec un scénario de pluies futures connues (19) et la dernière sur les critères de persistance uniquement.

D.2.2 Résultats

Tous les critères

Les trois premières composantes principales expliquent de 69 % à 75% de la variance totale (selon l'horizon de prévision, Tab. D.2). La première composante est celle des critères de persistance sans qu'il soit fait de distinction entre les critères calculés sur toute la période de contrôle et ceux calculés sur des événements (Fig. D.1). La deuxième composante correspond aux critères d'alertes aux horizons de prévision les plus courts, puis aux critères de retard pour les horizons de prévision plus loitains. À l'inverse, la troisième composante est corrélée avec les critères de retard puis avec les critères d'alerte à mesure que l'horizon de prévision s'éloigne. On voit apparaître des différences de corrélation selon le scénario de pluies futures utilisées. Les composantes suivantes sont plus difficiles à déchiffrer : elles expliquent moins de variance tout en s'appuyant sur des distinctions plus fines à l'intérieur des familles de critères. La représentation des critères dans les plans principaux montre bien qu'à l'intérieur de chaque grande famille, les critères sont très proches.

Critères calculés avec un scénario de pluies futures connues

Quand on restreint l'A.C.P. aux critères calculés avec un scénario de pluies futures égales aux observations *a posteriori*, les trois premières composantes principales expliquent de 80 % à 84 % de la variance totale (Tab. D.3). Les corrélations entre critères et premières composantes principales sont plus fortes (mais la base compte sept vecteurs de moins) et correspondent mieux aux grandes familles de critères. La première composante principale est clairement celle des critères de persistance calculés sur toute la période ou sur des événements sans qu'on puisse faire de distinction nette entre eux. Elle est légèrement corrélée aux critères d'alerte. La deuxième composante représente nettement la famille des critères d'alerte. La troisième composante rend compte de la variété introduite par les critères de

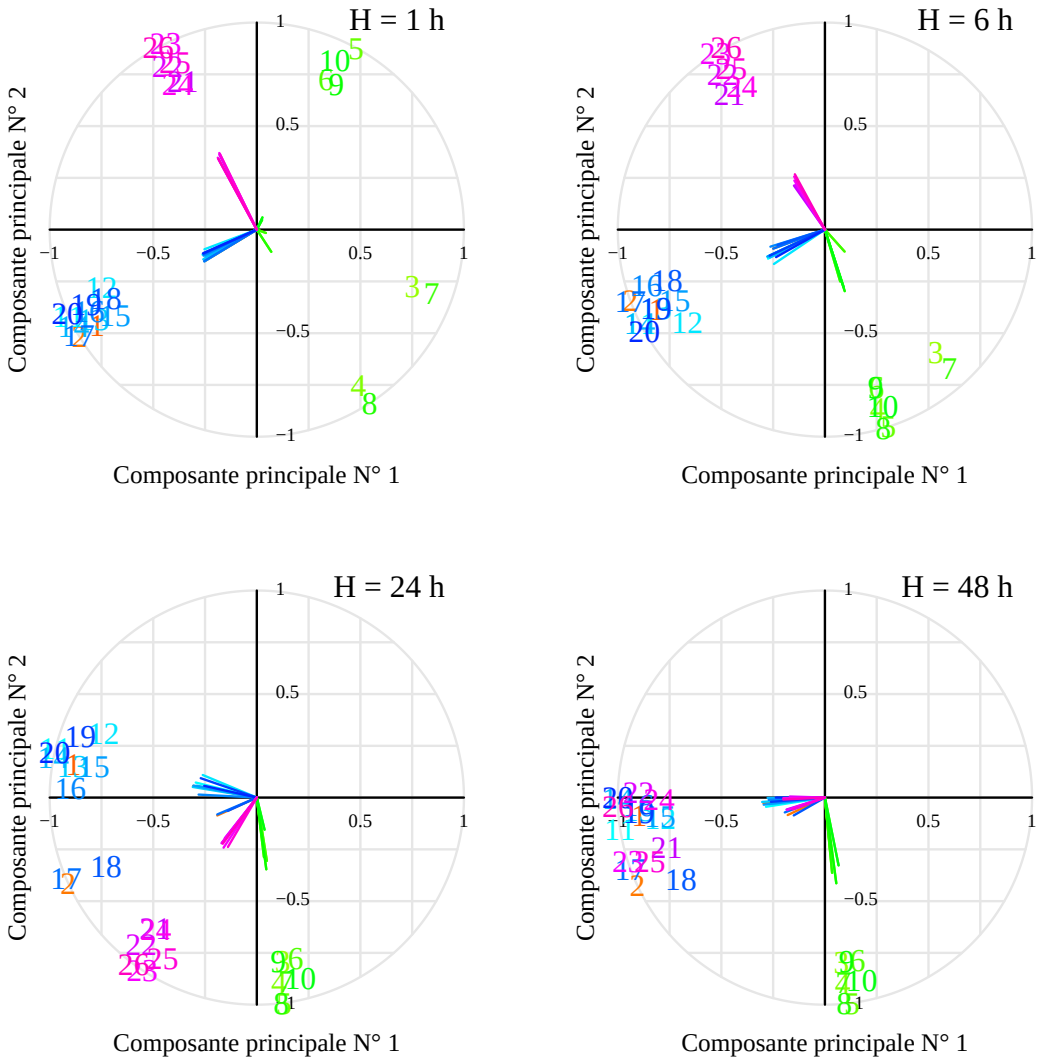


FIGURE D.1 – Représentation de l'A.C.P. des 26 critères de performance dans le plan des deux premières composantes principales, pour quatre horizons de prévision différents : 1, 6, 24 et 48 heures.

	Composantes principales					
Horizon	1	2	3	4	5	6
1 heure	0.51	0.65	0.75	0.84	0.90	0.95
6 heures	0.43	0.61	0.76	0.83	0.87	0.91
12 heures	0.38	0.56	0.73	0.80	0.86	0.91
24 heurs	0.36	0.54	0.69	0.77	0.84	0.88
48 heures	0.38	0.57	0.70	0.78	0.84	0.88

Table D.2 – Valeurs propres cumulées des 6 vecteurs propres les plus significatifs de l'A.C.P. réalisée sur l'ensemble des 26 critères de performance. Les valeurs propres ont été normées (leur somme vaut 1) ; il s'agit donc de la fraction de la variance des valeurs des critères expliquée par les n premiers axes (ici n va de 1 à 6).

retard. La représentation des critères dans les plans principaux confirme les résultats de la première A.C.P. (de façon renforcée, Fig. D.2).

	Composantes principales					
Horizon	1	2	3	4	5	6
1 heure	0.56	0.74	0.85	0.95	0.97	0.98
6 heures	0.48	0.70	0.84	0.90	0.94	0.96
12 heures	0.47	0.70	0.83	0.89	0.93	0.96
24 heurs	0.45	0.68	0.82	0.88	0.92	0.95
48 heures	0.47	0.65	0.81	0.86	0.91	0.94

Table D.3 – Valeurs propres cumulées des 6 vecteurs propres les plus significatifs de l'A.C.P. réalisée sur l'ensemble des 19 critères de performance calculés avec un scénario de pluies futures égales aux observations *a posteriori*. Les valeurs propres ont été normées (leur somme vaut 1). Il s'agit de la fraction de la variance des valeurs des critères expliquée par les n premiers axes.

Critères de persistance

Nous restreignons cette fois l'analyse aux critères de persistance, qu'ils soient calculés sur toute la période de contrôle ou sur des événements. Deux A.C.P. sont réalisées, la première sur l'ensemble de ces douze critères, la seconde sur les neuf critères calculés avec un scénario de pluies futures connues seulement. Le tableau D.4 donne les fractions de variance des neuf critères calculés avec un scénario de pluies futures connues, expliquées par les n premières composantes de l'A.C.P., selon l'horizon de prévision.

Dans les deux cas, les deux premières composantes principales rendent compte de la plus grande partie de la variance. On observe dans le plan principal de l'A.C.P. que les critères sont moins groupés (on explique moins de variété, mais de façon plus fine). Cependant, on peut observer que la première compostante principale explique de 80 % à 90 % de la variété à expliquer et que les projections de chacun des critères de persistance sur cette

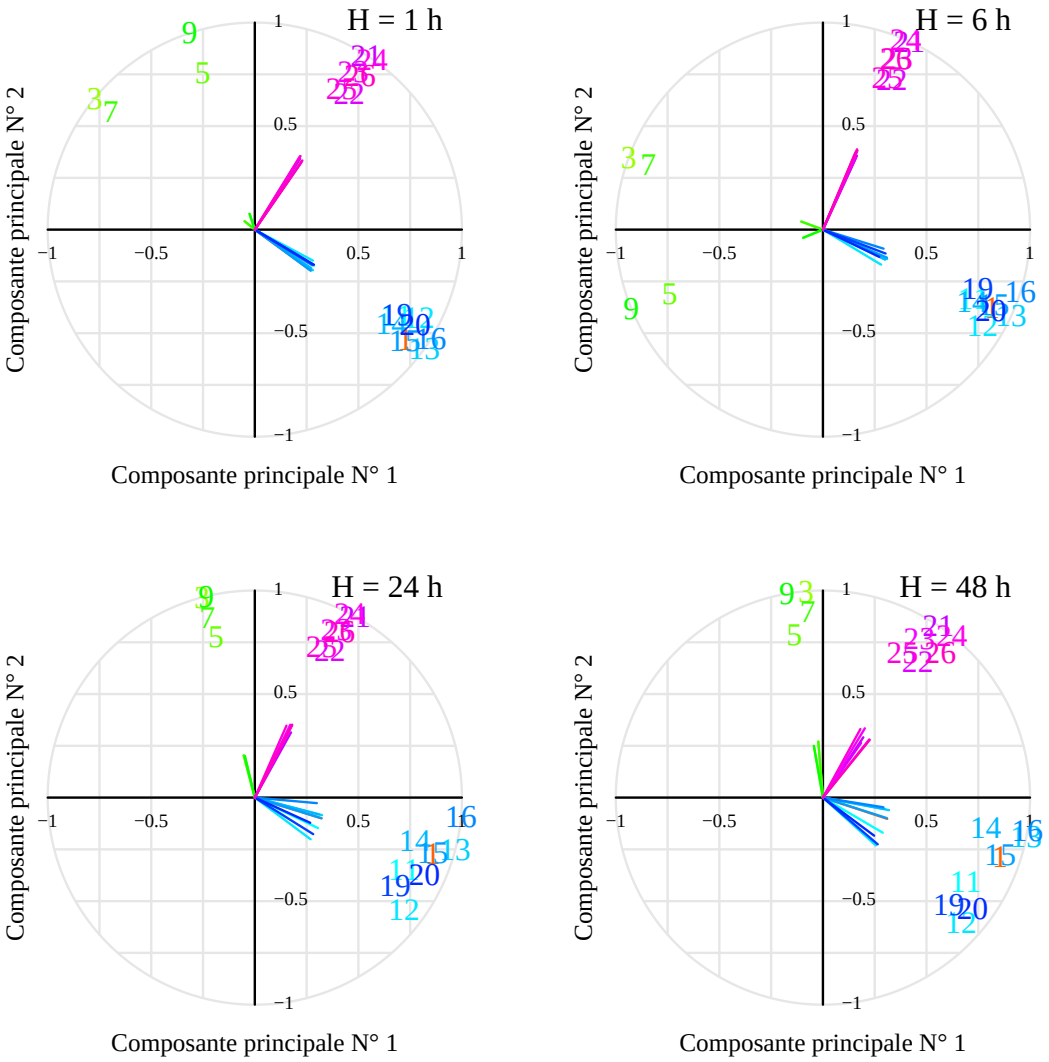


FIGURE D.2 – Représentation de l'A.C.P. des 19 critères de performance calculés avec un scénario de pluies futures connues dans le plan des deux premières composantes principales, pour quatre horizons de prévision différents : 1, 6, 24 et 48 heures.

	Composantes principales					
Horizon	1	2	3	4	5	6
1 heure	0.94	0.97	0.98	0.99	1.00	1.00
6 heures	0.89	0.93	0.96	0.98	0.99	0.99
12 heures	0.88	0.93	0.96	0.98	0.99	0.99
24 heurs	0.85	0.92	0.96	0.98	0.99	0.99
48 heures	0.83	0.92	0.95	0.97	0.98	0.99

Table D.4 – Valeurs propres cumulées des 6 vecteurs propres les plus significatifs de l'A.C.P. réalisée sur l'ensemble des 9 critères de persistance calculés avec un scénario de pluies futures égales aux observations *a posteriori*. Les valeurs propres ont été normées (leur somme vaut 1). Il s'agit de la fraction de la variance des valeurs des critères expliquée par les n premiers axes.

première composante principale sont identiques (Fig. D.3) : tous les critères contribuent de la même façon à la première composante principale qui décrit la plus grande partie de la variété explorée. Il paraît donc assez indifférent d'utiliser le critère de persistance sur toute la période de contrôle ou sur des événements de hautes eaux, de fortes pluies ou de montée des eaux. Cette conclusion est surtout validée pour les horizons de prévision les plus courts. Ceci peut s'expliquer par le fait que le critère de persistance compare en fait les erreurs faites par le modèle à celles réalisées par le modèle de persistance. Comme notre modèle emploie l'information du dernier débit, il est probable qu'il y ait une corrélation entre les erreurs du modèle testé et celles du modèle de persistance : cette étude ne donnerait peut-être pas les mêmes résultats en simulation. L'hétéroscédasticité des erreurs est également un facteur pouvant expliquer cette relative indifférence au critère de persistance choisi.

Dans le cas de l'A.C.P. réalisée sur les critères calculés avec un scénario de pluies futures connues, la première composante principale est écrasante. On observe néanmoins pour les horizons de prévision supérieurs à 6 heures, l'apparition de deux groupes de critères dans la projection dans le plan formé par les deux composantes principales. Le premier groupe est constitué du critère calculé sur toute la période de contrôle et les critères calculés sur des événements de fortes pluies (critères 1 et 13 à 16), le second groupe correspond aux critères calculés sur des événements basés sur des seuils de débits (critères 11, 12, 19 et 20).

Dans le second cas, on observe une seconde composante plus représentative ; deux groupes bien distincts de critères se forment dans la projection dans le plan formé par les deux composantes principales, le premier groupe correspond aux critères calculés en scénario de pluies futures connues, le second aux critères calculés en scénario de pluies futures nulles (critères 2, 17 et 18) : la variance principale présente dans le groupe de critères est donc liée au scénario de pluies.

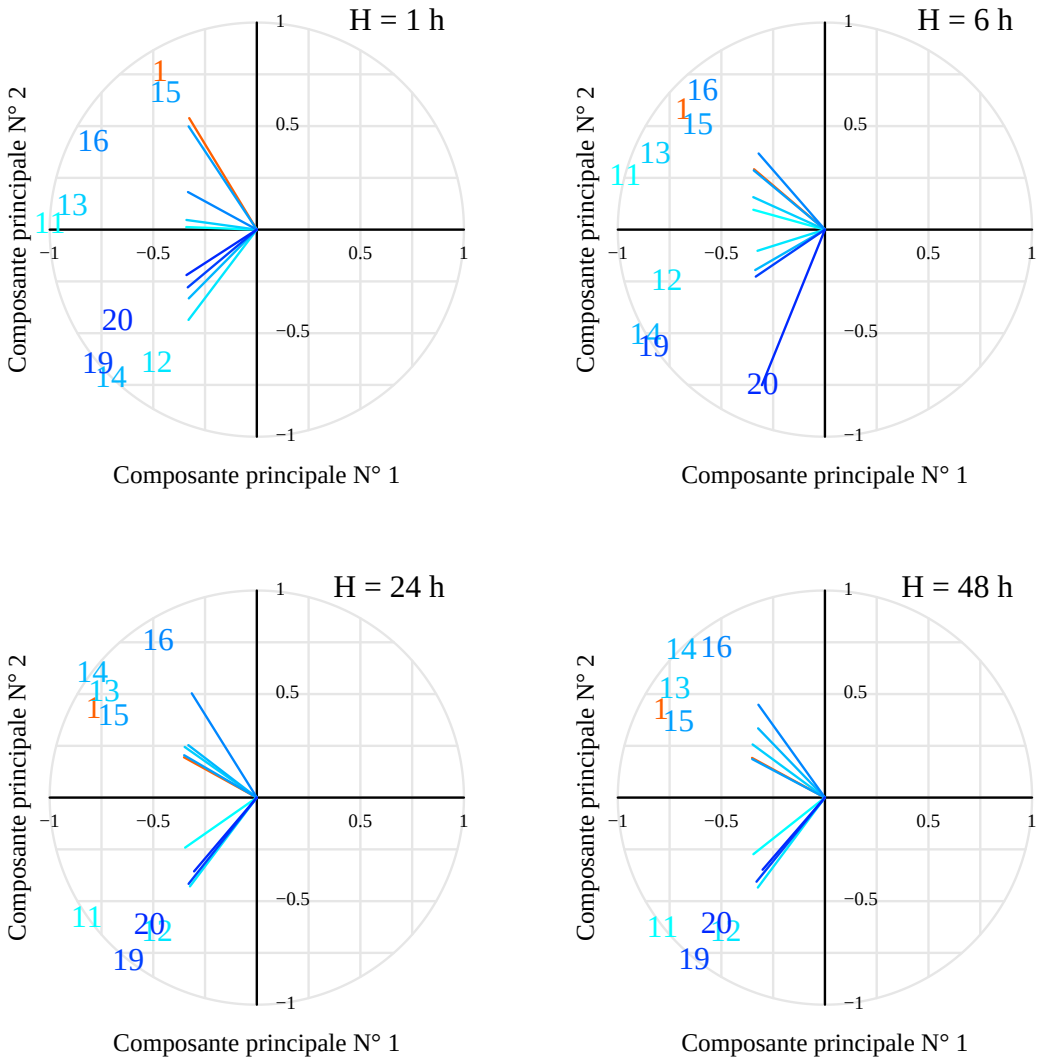


FIGURE D.3 – Représentation de l'A.C.P. des 9 critères de persistance dans le plan des deux premières composantes principales, pour quatre horizons de prévision différents : 1, 6, 24 et 48 heures.

Influence de l'horizon de prévision

Les corrélations sont légèrement moins évidentes pour les horizons de prévision les plus lointains (faible diminution de la variance expliquée par les premières composantes principales) et critères un peu moins groupés. Il s'agit d'une observation « de second ordre » par rapport aux grandes tendances décrites ci-dessus.

D.2.3 Discussion

Les A.C.P. réalisées sur tous les critères (en pluies futures parfaitement connues) permettent d'affirmer d'une part que 3 ou 4 critères (définis comme les composantes principales les plus influentes) permettent de rendre compte de la plus grande partie de l'information apportée par l'ensemble des critères puisque la somme des valeurs propres associées dépassent les 80 %, voire les 90 % de la variété explorée.

D'autre part, comme les critères sont très nettement groupés (en familles de critères) dans les plans principaux de l'A.C.P., chaque critère apporte la même information au sein de la i^e composante principale (sa projection sur l'axe i). Une autre façon de le voir est de constater que les groupes de critères se situent quasiment à angle droit dans les plans principaux. Aussi, la variété entre familles de critères est très largement supérieure à celle existant à l'intérieur de chaque famille.
Nous avons ainsi montré que les critères d'alerte CSI' donnent une information assez indépendante du coefficient ν portant sur le nombre de fausses alertes. Les critères d'alerte donnent par contre une information assez différente et donc complémentaire de celle apportée par le critère de persistance.
De même, l'étude confirme les résultats du travail mené lors de l'élaboration de critères de retards : ces derniers sont relativement similaires entre eux (mais moins que les critères d'alerte par exemple) et apportent une information différente de l'information contenue dans le critère de persistance.

L'A.C.P. confirme que les scénarios de pluie future jouent un rôle fort dans l'évaluation du modèle. Il est donc nécessaire d'évaluer le modèle sous différents scénarios.

Aussi, plus que de raffiner la définition de critères de persistance, de retard ou d'alerte, il importe prioritairement d'utiliser un critère de chacune de ces familles.

E

Annexe

Combien d'années pour obtenir une valeur sûre d'un critère quadratique ?

Cet article a été soumis au journal des sciences hydrologiques (*Hydrological Sciences Journal*) pour son numéro spécial consacré à l'atelier *Cour des miracles de l'hydrologie*, qui s'est tenu à Paris du 18 au 20 juin 2008. Il est actuellement accepté et en attente de publication du numéro spécial (*in press*).

Quel que soit le critère, l'évaluation du modèle est dépendante dans un premier temps de la chronique sur laquelle cette évaluation est faite. Certains critères sont définis de façon à ce qu'ils convergent quand la chronique s'allonge : l'évaluation devient « absolue » (indépendante des données). Si notre critère prend peu de points en compte et si ces points sont définis relativement aux événements présents dans la chronique, on peut se demander s'il doit converger si on utilise une chronique infiniment longue. Si le critère quadratique testé converge quand la quantité de données d'évaluation tend vers l'infini, la vitesse de convergence est très lente : le nombre de pas de temps d'évaluation nécessaires pour assurer une bonne estimation de l'espérance de l'erreur quadratique moyenne est très élevé. À partir de là, le critère de persistance n'a qu'un sens très relatif pour les séries temporelles dont nous disposons : il ne donne qu'une évaluation relative du niveau d'erreur qu'on peut attendre du modèle sur un bassin donné. Ceci nous encourage encore plus à travailler sur de nombreux bassins.

How significant are quadratic criteria? Part 1: How many years are necessary to ensure the data independence of a quadratic criterion value?

LIONEL BERTHET[1,2], VAZKEN ANDRÉASSIAN[1] , CHARLES PERRIN[1], & CECILE LOUMAGNE[1]

[1] Cemagref, *Hydrosystems and Bioprocesses* Research Unit, PB 44, 92163 Antony cedex, France. E-mail: lionel.berthet @cemagref.fr
[2] AgroParisTech ENGREF, 19 avenue du Maine, F 75732 Paris

ABSTRACT

Quadratic criteria are widely used to assess the performance of hydrological models. However, the statistical nature of the errors makes the interpretation of these criteria tricky. In this paper, root mean square error (RMSE) values were computed for a hydrological model over a set of 178 varied catchments on two 5-year data series: the computed RMSE values can actually reflect the content of the data series with which they are calculated as much as the intrinsic skills of the model. The error model proposed by Yang *et al.* (2007) is used to assess a lower bound of the RMSE confidence interval width, depending on the length of the data series used for the assessment. Our analysis indicates that the data series would have to be longer than several decades to ensure that computed RMSEs are close to their statistical expectation. The practical consequences of this result are raised and discussed.

Keywords: Model assessment; skill scores; quadratic criterion

INTRODUCTION

Assessing the skills of hydrological models is of crucial importance for scientists to improve their understanding of hydrological processes and to build better models. It is also important for end-users because it helps them to select adequate tools for the operational tasks they need to perform. Over a given catchment, model assessment aims to estimate the model's predictive accuracy, i.e. the level of error that one can expect with this model when run on this catchment outside the periods used for model set-up and calibration. The visual inspection of the computed hydrographs, in spite of its primary importance, remains time-consuming and clearly subjective (Houghton-Carr, 1999). Instead, numerical performance criteria are apparently more objective, and many have been introduced in the literature (see e.g. Nash and Sutcliffe, 1970; Sorooshian *et al.*, 1983; Sorooshian and Gupta, 1983; Mathevet *et al.*, 2006; Dawson *et al.*, 2007).

Quadratic criteria (i.e. based on square residuals) are among the most widely used, even though several papers have shown some of their drawbacks in assessing the performance of hydrological models. Some authors consider these criteria to be inadequate in a formal statistical framework (like the maximum likelihood). Martinec and Rango (1989) and Legates and McCabe (1999) show that all errors are not treated identically: indeed, quadratic criteria give much more importance to the greatest errors, which occur most frequently during flood events, than to the smallest errors. Furthermore, model errors are most often correlated in time (Yang *et al.*, 2007; Beven, 2008). Consequently, these criteria remain difficult to interpret consistently in spite of their numerical nature.

Can we interpret performance criteria values as expected criteria values?

Models are calibrated or tested over finite data series that encompass a set of significant events: it is only a small sample of all the possible events that may occur on a given catchment. Consequently, the computed value of any performance criterion depends on the choice of the data series. For model assessment to be truly *meaningful,* the criterion value should not depend too heavily on the data series, i.e. this value should represent the level of error that can be *expected* using the model on any given data series (for a given catchment). Otherwise, the value reflects more the content of the data series than the skill of the tested model.

Most often, our assessment is uncertain because models are assessed on short time series (decades-long time series for precipitation and runoff are seldom available) or even on a small number of events only (see e.g., Tan *et al.*, 2008).

Scope of the paper: when are the criterion values significant?

Despite the drawbacks of quadratic criteria, some hydrologists may still choose to use them if they fit well with the model's application: for example, a modeller interested in flood management may positively consider the strong dependence of the quadratic criteria values on the errors made during flood events. In such a case, it is necessary to check the significance of the criterion value.

In this paper, we aim at verifying whether the value of the criterion computed on a given series actually reflects the intrinsic model skill (for a given catchment) or depends mostly on the data series content. Our objective is to determine at least a lower bound of the length of a time series that ensures that the value of the RMSE is close to its statistical

69 expectation (and therefore independent of the input data used). In the companion paper,
70 we will analyse the role of unusual model errors on the value of quadratic criteria.

71

72 **DATA AND MATERIALS**

73 **Catchment set and hydrological data**

74 The study was conducted on a set of 178 French unregulated catchments (Fig. 1), part of
75 the set collected by Le Moine *et al*. (2007). The set is representative of the hydroclimatic
76 variability encountered in the country: from catchments experiencing Mediterranean flash
77 floods to much slower basins. No high-elevation or snow-affected catchment is included
78 in this set. Catchment areas range from 10 to 5,940 km² (354 km² on average). Working
79 on various catchments ensures more general and robust conclusions to our study
80 (Andréassian *et al.*, 2006).

81 Available data consist of hourly areal precipitation, potential evapotranspiration (PE) and
82 discharge from 1995 to 2005. PE values were computed using the formula proposed by
83 Oudin *et al.* (2005), based on temperature and extra-atmospheric global radiation.

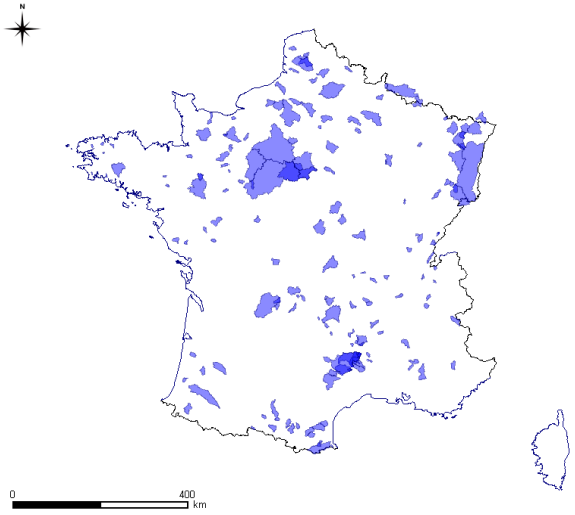


Fig. 1 Location of the 178 catchments used in this study.

The GRP rainfall-runoff model

The GRP model is a continuous, lumped, hybrid metric-conceptual model (Tangara, 2005; Berthet *et al.*, 2009; Berthet, 2010). It is one of the operational models used to forecast river flows in real time on French catchments, including the Seine river basin upstream of Paris. Detailing its structure is beyond the scope of this paper because the study conducted in this paper does not depend on a particular hydrological model. The model totals three free parameters (that require calibration), hereafter named θ_1, θ_2 and θ_3. More details are provided in the companion paper (Berthet *et al.*, 2010, this issue).

METHODOLOGY

What can a short data series teach us?

We started this study with real (and quite short) data series. We assessed the GR3P model over the set of 178 catchments with the split-sample test procedure described by Klemeš

(1986): the model was calibrated on a period P1 and validated on a non overlapping period P2; then the roles of periods P1 and P2 for calibration and validation were reversed. This was done over the catchment set with 10-year hourly data series. For each catchment, the computed root mean square errors (RMSE) obtained for periods P1 and P2 in validation were compared.

Use of an error model to compute longer synthetic error series

Since it is practically speaking impossible to have 100-year or longer data series for many catchments, we had to generate longer synthetic model error series to explore the changes in the RMSE values when computed over longer periods. This was done in two steps: first, we generated a number of synthetic runoff series using a bootstrap-like sampling method, and then we used the error model proposed by Yang *et al.* (2007) to provide synthetic error series based on the simulated runoff series.

Finally, we used those synthetic series to assess the minimal length t_m after which all error simulations provide stable and similar RMSE values for every catchment. Details for each step follow.

Synthetic runoff simulation

We used the 10-year hourly time series of runoff to simulate 100-year time series using an iterative method based on the chronological analogue method. This method provides a synthetic series that is statistically coherent with the 10-year observed runoff series:

Every time step *t* of the observed runoff series (1995–2005) is characterised by the vector of the *m* previous observations of runoff: $\mathbf{X}_t = (Q_t, Q_{t-1}, ..., Q_{t-m+1})$. Here *m* was chosen equal to 4 in order to allow the test of different time-dependence components within the

error model (introduced in the following section). Let l_i be the synthetic series length under construction at the i^{th} iteration of the generation process. The method is iterative:

1. The vector $\hat{\mathbf{X}}_i = \left(\hat{Q}_{l_i}, \hat{Q}_{l_i-1}, ..., \hat{Q}_{l_i-m+1}\right)$ is calculated and we determine the subset of the segments $\mathbf{X}_t$ (in the observation series) that are the closest to the vector $\hat{\mathbf{X}}_i$.
2. We randomly select one segment $\mathbf{X}_k = (Q_k, Q_{k-1}, ..., Q_{k-m+1})$ in this subset.
3. An integer p is drawn from a convenient probability density function (here a Poisson distribution with an average equal to 48 h was chosen: p ranges from 1 to 72 hours).
4. The segment $\left(Q_{k+1}, Q_{k+2}, ..., Q_{k+p}\right)$ is then added to the synthetic series under construction.

The generation process ends when the series is long enough for our purposes. For this paper, we generated a total of 100 series, each 100-year long.

Since the process cannot extrapolate (it does not produce events of a magnitude outside the range of the observations), we randomly added flood events, using a Gumbel law for annual maximum values (quite well validated at the daily time step on most of the catchments) to calibrate their magnitudes.

Error series simulation

The synthetic runoff series described in the previous section were generated in order to simulate error series longer than the real series available, based on an error model.

The error structure of a hydrological model is a complex mathematical object, due to the numerous sources of uncertainty (Beven, 2008b), including input uncertainty, model

143 structure and parameter uncertainty, and output uncertainty. Few papers have reported
144 tentative validation of an error model in hydrology.

145 Several authors (see e.g. Beven and Freer, 2001; Yang *et al.*, 2007) pointed out two key
146 features characterising the statistical structure of the errors made by a hydrological
147 model:

148 1. Heteroscedasticity: log-transformed runoff values were used to account for this
149 property;
150 2. Lack of temporal independence: an auto-regressive AR(1) model (Box and
151 Jenkins, 1976) was added.

152 Thus, Yang *et al.* (2007) modelled the conditional probability densities of the individual
153 errors as:

154 $$f_{E_t|E_{t-1}=\varepsilon_{t-1},Q_t=q_t,Q_{t-1}=q_{t-1}}(\varepsilon)=\frac{1}{\sigma\cdot(q_t+\varepsilon_t)\sqrt{2\pi\cdot(1-\rho^2)}}\cdot e^{\frac{((\log(q_t)-\log(q_t+\varepsilon_t))-\rho\cdot(\log(q_{t-1})-\log(q_{t-1}+\varepsilon_{t-1})))^2}{2\sigma^2\cdot(1-\rho^2)}} \quad (1)$$

155 where ε_t is the error on time step *t*, q_t is the observed discharge, σ^2 is the variance of the
156 multiplicative error and ρ is the one-time-step-ahead autocorrelation.

157 Yang *et al.* (2007) tested their model on one case study on which their proposed structure
158 was quite well validated. However, they identified several limitations: the autocorrelation
159 of the residuals significantly varied from dry to wet periods. This implies that their error
160 model will underestimate the expectation of errors during periods of rapid variations in
161 the catchment state (i.e. storm events). We selected this error model for the present study
162 because its weaknesses were not limiting for our objectives (as discussed in the next
163 section).

Using the model of errors proposed by Yang *et al.* (2007) and the synthetic runoff series, we simulated one hundred 100-year time series of errors. The Yule-Walker equations were used to derive the AR(p) model from the autocorrelation function (see e.g. Box and Jenkins, 1976).

Assessment of the data series length required for a meaningful assessment

For each catchment, we computed the partial *RMSE(t)* for the j^{th} synthetic error series, i.e. the RMSE computed from time step 1 to time step *t*:

$$RMSE_j(t) = \frac{1}{t}\sum_{k=1}^{t} \varepsilon_j(k)^2 \quad (j \text{ from 1 to 100}). \tag{2}$$

Then we assessed the distribution of partial RMSE for different values of *t* until it becomes stable and close enough to a single value. This property was evaluated through a relative spread index defined by:

$$\Delta_t = \frac{q_{95}(RMSE(t)) - q_5(RMSE(t))}{q_{50}(RMSE(t))} \tag{3}$$

where $q_x(RMSE(t))$ is the empirical quantile *x* of the criterion. This index converges to 0 when *t* tends towards infinity (see the appendix); for each catchment *i*, we computed the minimum assessment duration T_i such as for any $t > T_i$, $\Delta_t \leq 0.1$. This duration is used to characterise the time series length that allows computing a RMSE value acceptably close to its statistical expectation on catchment *i*.

Sensitivity to the number of simulations

The sensitivity of the results presented in the next sections to the number of simulations was tested to ensure that the set of 100 simulated model error series was sufficient to represent rare events and rare errors.

185

186 **ASSESSMENT OF THE ERROR MODEL**

187 Since we planned to work with synthetic data, the validity of the model used to generate
188 these data had to be tested. A full validation of the error model is far beyond the scope of
189 the paper. Since Yang *et al.* (2007) have already pointed out a number of deficiencies of
190 the error model they proposed, we first had to check whether the error model behaviour
191 was consistent enough with our intended application.

192 For every catchment, we used the set of 100 error series as an ensemble to obtain the 90%
193 confidence interval for the RMSE calculated over 5 years. This confidence interval was
194 then compared to the empirical RMSE values calculated with the two observed 5-year
195 data series (Fig. 2). In particular, we computed the percentages (for 178 catchments$\times$two
196 periods) of the empirical RMSE values within the 90% confidence interval and of values
197 smaller than the interval's lower bound or greater than its upper bound. The error model
198 proposed by Yang *et al.* (2007) underestimates the dispersion of the RMSE since only
199 40% of the empirical RMSE values fall within the 90% confidence interval: this falsifies
200 the error model used (Fig. 2). However, 7% of the values are smaller than the lower
201 bound of the confidence interval, which is close to the expected 5%. This may indicate
202 that the lower bound of the confidence interval is correctly assessed. Conversely, the
203 large percentage of empirical values greater than the upper bound of this interval shows
204 that this upper bound is greatly underestimated (53%). This seems in accordance with the
205 discussion by Yang *et al.* (2007), who concluded that a constant autocorrelation
206 hypothesis is inappropriate for wet seasons and particularly for flood events.
207 Consequently, the largest errors (which occur mostly during flood events) are not well

represented. This explains why the lower bound (quantile 0.05 obtained with synthetic data series encompassing only small errors, thus featuring comparatively small events) is correctly estimated while the upper bound (quantile 0.95 obtained with series encompassing large errors and consequently comparatively stronger flood events) is underestimated.

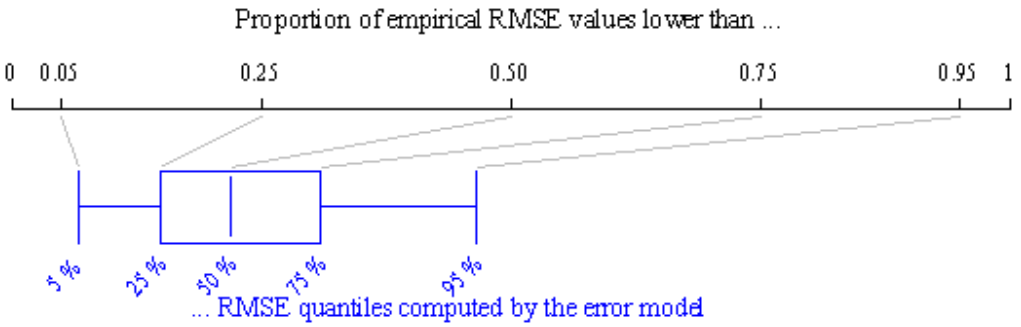


Fig. 2 Proportions of the catchments with empirical RMSE lower than RMSE quantiles computed by the error model. For example, the empirical RMSE is lower than the quantile 0.75 computed by the error model for 31% of the catchments.

Note that we also tested a slightly different error model featuring an AR(3) component (e.g. Beven *et al.*, 2008) instead of the AR(1), but it did not give better results (not shown here for the sake of brevity). Consequently, we used the model proposed by Yang *et al.* (2007) for the rest of the study but restricted its use to the estimation of a lower bound to the value of the minimal assessment duration T_i, which ensures an acceptable uncertainty for the RMSE values since the error model underestimates the relative spread Δ_t.

RESULTS

Uncertainty of the RMSE values computed over 5-year data series

Here we analysed the errors actually made by the GRP model (in validation mode) on periods when flow observations were available. We observed that the empirical RMSE values could significantly differ from period P1 to period P2 (Fig. 3): the differences between the two computed values are higher than 20% of the smallest one for 50% of the catchments. This illustrates that the computed value of the RMSE does indeed depend on the period on which it is calculated.

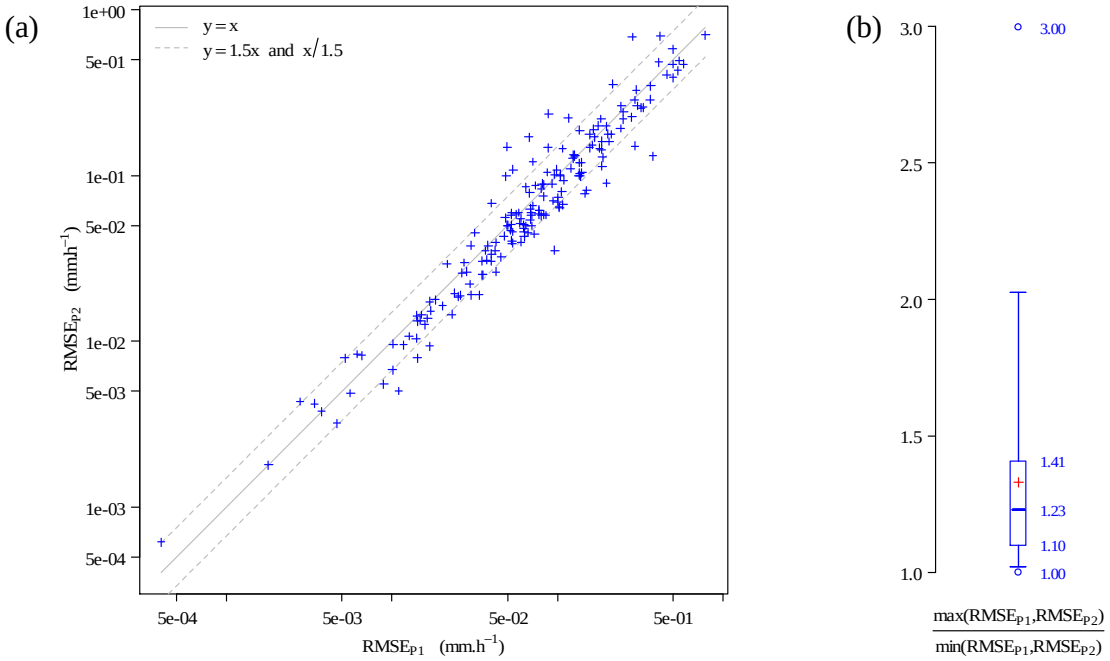


Fig. 3 Root mean square errors (RMSE) were computed on two independent periods for 178 catchments. (a) Scatterplot of the RMSE values computed on periods P1 and P2. The dashed lines represent an error ratio of 50%. (b) Distribution (boxplot) of the ratio of the highest RMSE value to the lowest value for each catchment.

Actually, it is well known that the quadratic criteria are overly influenced by the largest errors, which occur mostly on flood events (Legates and McCabe, 1999). Then we checked the stability of the empirical RMSE computed over a 5-year series depending on

241 the magnitude of the flood events observed in the 5-year series. For each catchment and
242 for each period, we ranked the errors in decreasing order and selected the subset of the
243 largest errors accounting for 80% of the total RMSE. Then we computed the range of the
244 runoff observations corresponding to this error subset (see the companion paper for
245 further details). For each catchment, we characterised the differences between the two
246 periods by computing an overlapping index of the runoff intervals $\left[Q_{\min}^{P1}, Q_{\max}^{P1}\right]$ and
247 $\left[Q_{\min}^{P2}, Q_{\max}^{P2}\right]$ (on periods P1 and P2, respectively) as the length of their intersection divided
248 by the length of their union. An index value of 1 indicates that the ranges are equal on
249 periods P1 and P2 while a value of 0 denotes that the ranges do not overlap at all. Fig. 4
250 displays these index values depending on the ratio of the RMSE computed on periods P1
251 and P2. It is clear that when periods P1 and P2 exhibit runoffs of the same magnitude, the
252 probability of having close RMSE values is high. Conversely, if the periods are
253 characterised by runoffs of different magnitudes, then it is much more likely that the two
254 RMSE values would differ considerably. This observation strengthens the concerns of
255 some authors who suggest assessing the model's performance for bands of flow rates
256 (Khan, 1989).

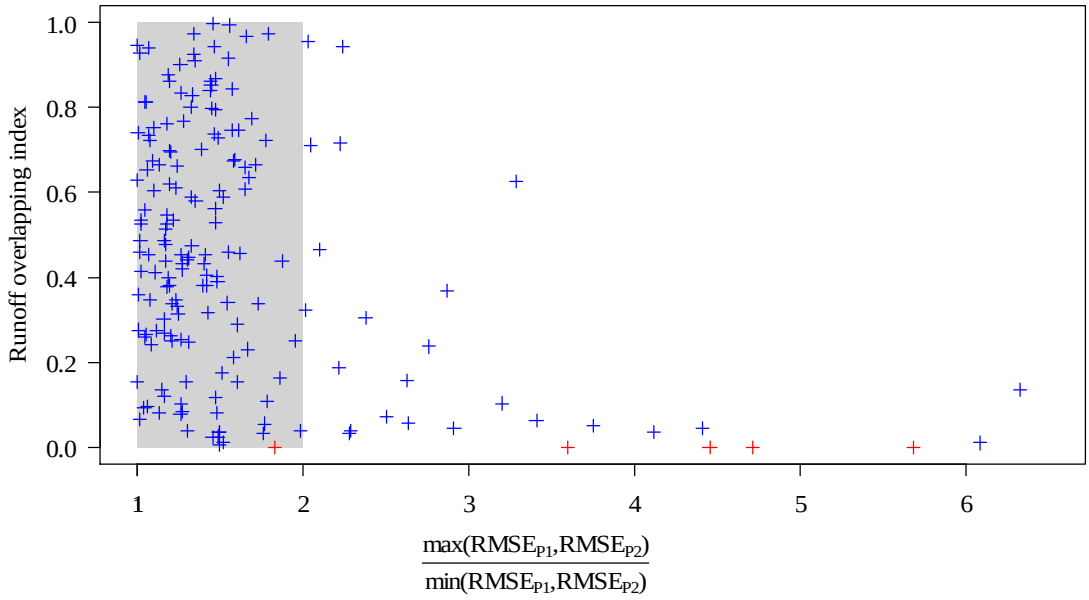


Fig. 4 Runoff overlapping index (indicating whether or not the runoff observations associated with the largest errors on periods P1 and P2 are of the same magnitude: an index value of 1 indicates that the ranges are equal on periods P1 and P2 while a value of 0 denotes that the ranges do not overlap at all) depending on the ratio of the RMSE computed over periods P1 and P2.

Can a minimal assessment duration be defined?

Fig. 5 displays a typical example of how the distribution of the simulated partial $RMSE(t)$ evolves and of the relative spread index Δ_t (eq. 3) for one catchment (Alagnon River at Lempdes). The spread is very wide during the first 10 years but decreases quickly. Then it decreases at a much slower rate. The median of the distribution becomes stable after several years (e.g. 10 years), meaning that it can be taken as an estimate of the expectation of the RMSE.

Fig 6. extends the analysis to the whole set of catchments. The same pattern is observed for almost all of the catchments. Consequently, the distribution of the relative spread indexes with the assessment duration evolves similarly.

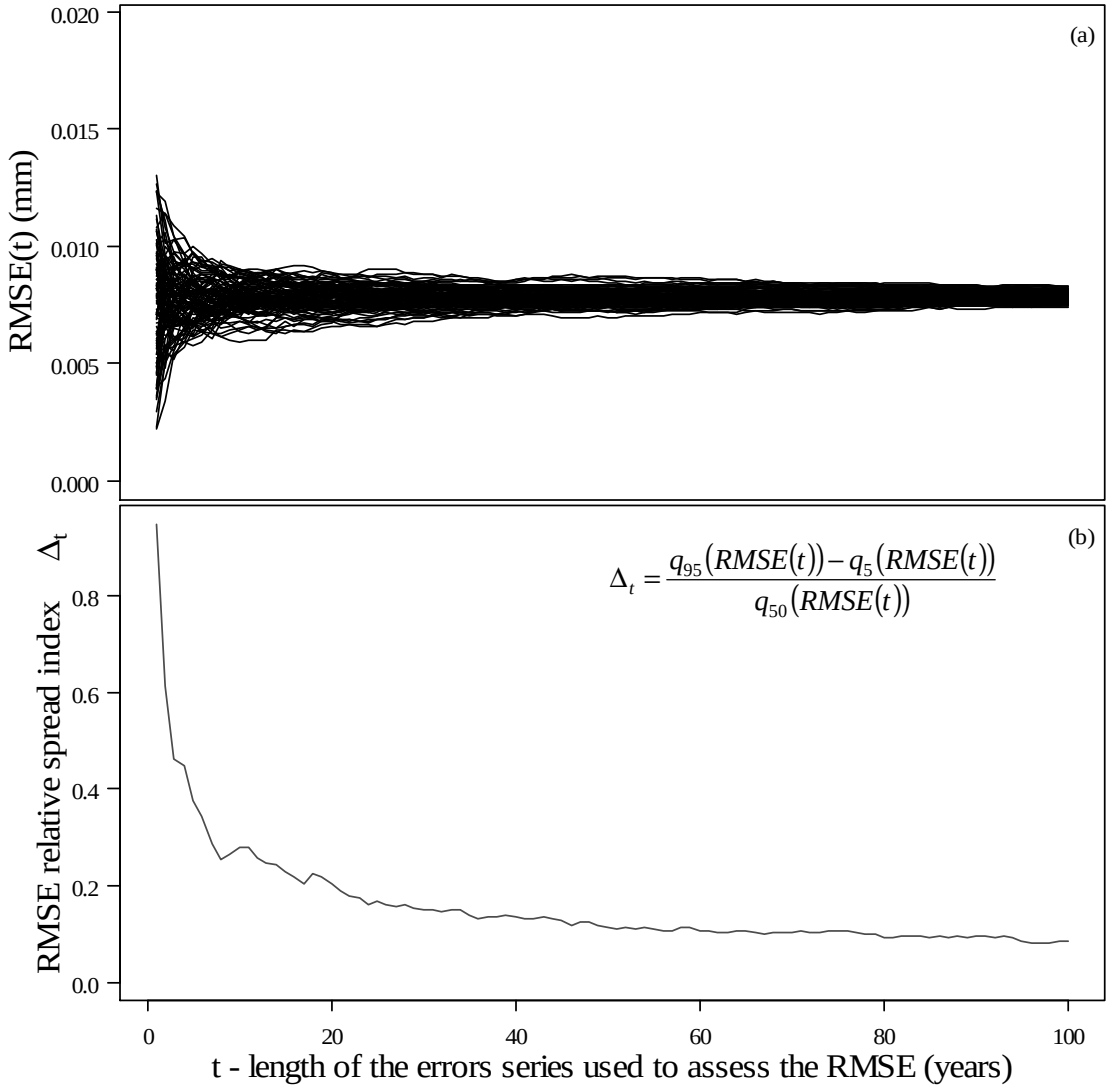


Fig. 5 Spread of the distribution of simulated partial RMSE(t) depending on the length of the synthetic error series used to assess the RMSE for the Alagnon River at Lempdes (984 km²). (a) The 100 RMSE values computed on synthetic data depending on the assessment duration *t*. (b) The relative spread index depending on the assessment duration *t*.

The relative spread index remains high for many catchments even when the partial RMSE(t)s are assessed over several decades. For a period of 50 years, the relative spread index drops below 0.1 for only 14% of the catchments. Thus the minimal duration to assess the RMSE expectation is much longer in many cases. Fig. 7 shows the percentage of catchments for which the assessment duration is long enough to get a relative spread index lower than a chosen threshold (according to the error model that was selected). In many cases, the minimum assessment duration is longer than several decades, i.e. much more than the data series that are usually available. Note that due to the previously underlined limitations of the error model, these values are lower bounds; true minimal assessment durations are probably even longer. Consequently, the data series length has to be far longer than 10 years in order to compute a meaningful RMSE for most catchments.

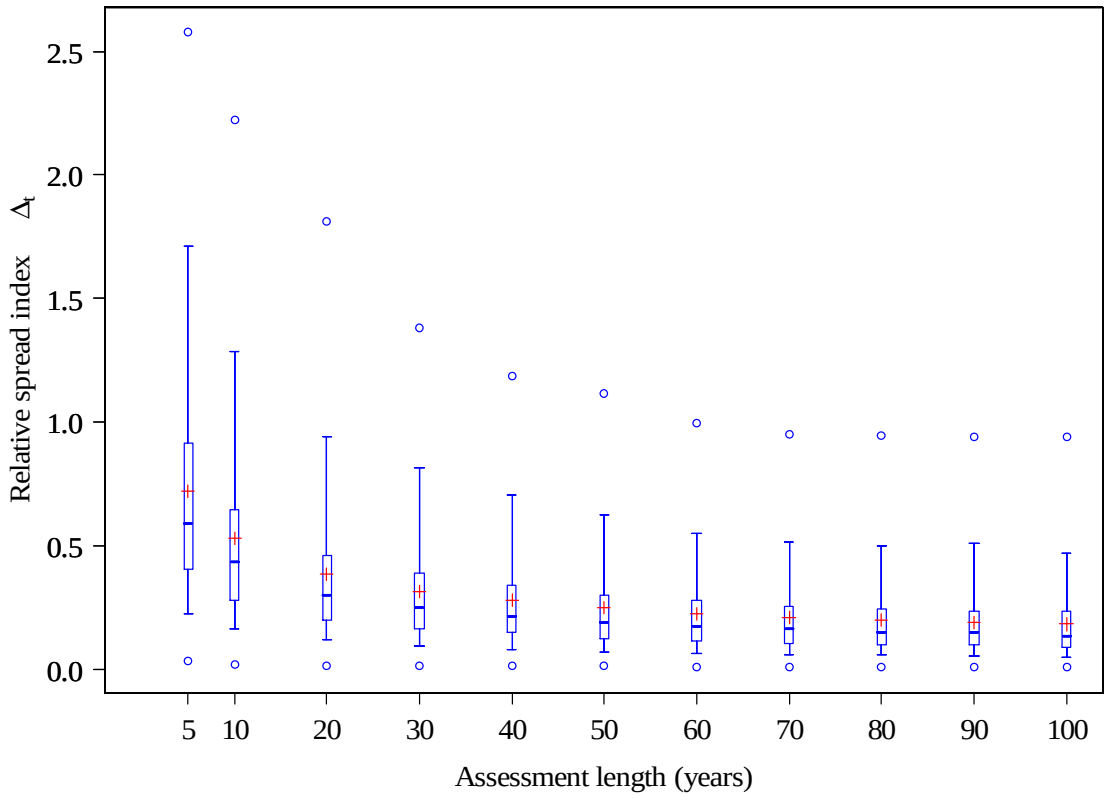
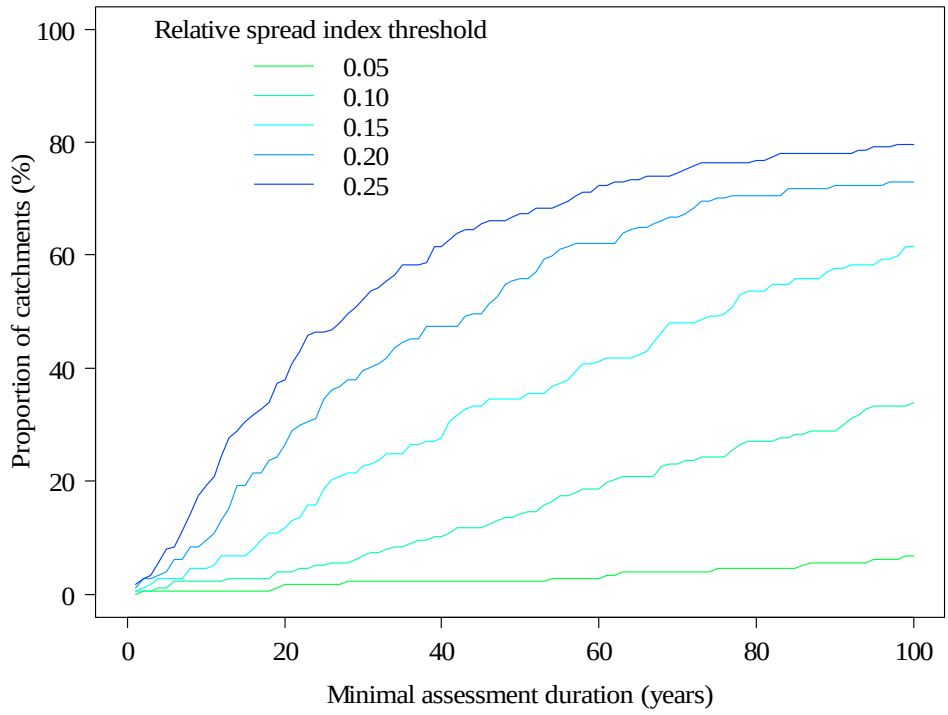


Fig. 6 Distributions (over the set of 178 catchments) of the RMSE relative spread index Δ_t depending on the assessment duration *t*.



297

298 **Fig. 7** Cumulative distribution of the minimal assessment durations, with respect to the chosen
299 accuracy (relative spread index threshold). For example, according to the error model, with a 60-
300 year time series, the relative spread index will be lower than 0.1 for only 20% of the catchments.

301

302 **DISCUSSION AND PERSPECTIVES**

303 The results shown in the previous section have a number of consequences on our ability
304 to assess a model on a single catchment over a rather short period with a quadratic
305 criterion, as is commonly done.

306 **Consequences for model validation over a single catchment**

307 The results show that we simply cannot confidently assess the intrinsic value of a model
308 on a single catchment (i.e. its level of predictive accuracy) using a RMSE value
309 computed over a few years. It provides little information on the model's skill and the
310 confidence interval can often be very large. It is therefore not possible to conclusively

311 judge the model's performance when assessed with RMSEs on a short time period. Is
312 there a way out of this apparent dead-end? Two partial solutions to this problem are well
313 known.

314 An absolute assessment of models based on RMSE values seems difficult; a relative one
315 may be more accurate. Some authors (e.g. recently Seibert, 2001; Perrin *et al.*, 2006;
316 Schaefli and Gupta, 2007) advocate using a benchmark model as done in the Nash-
317 Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970) or the persistence index (Kitanidis
318 and Bras, 1980) to normalise the criterion: the RMSE of the tested model is compared to
319 the RMSE obtained with a reference model. This makes interpreting the criterion value
320 easier. Besides, this approach is valuable in that it reduces the influence of the data series
321 content (of the selected period) on the criterion value. Since the comparison of the RMSE
322 of the tested model to the RMSE of a benchmark model is achieved with the same data
323 set, we can expect the result of this comparison computed on a few years to be closer to
324 its own expectation than the RMSE of a single model to its expectation. The result of the
325 comparison between the RMSE of two models is likely to be most often stable (when
326 computed on different periods). We tested this hypothesis over a set of different rainfall-
327 runoff models (some had close structures and are expected to have close results, some
328 were much more different; the differences are not shown here for the sake of brevity)
329 assessed on the same periods P1 and P2. The results of the comparison over periods P1
330 and P2 were in agreement for 91% of the catchment set (this fraction was nearly the same
331 for all the model couples tested), i.e. when a model is better than another one on one
332 period, it tends to remain better on another period. Thus, this relative RMSE-based
333 evaluation is much more stable from one period to another than the simple RMSE.

However, this approach may not be a completely satisfactory solution since there is no evidence that both models would react in the same way when they face the same data series: the data series content may still have some influence on the comparison output.

If a sufficient amount of data is available, another test is to divide the data series into a few subperiods (as done in the split-sample test framework; Klemeš, 1986) and to compare the value of the RMSE obtained on the entire data series to the distribution of the RMSE values computed on the subseries rather than simply using their mean or median values. If the distribution is extended or / and if the RMSE value computed over the entire data set differs much from the central value of those computed over the subseries, then little confidence should be given to the computed RMSE.

Are there any consequences for model calibration with a quadratic criterion?

Calibration may be defined as a process that determines the parameter set(s) that provide(s) the best performance of the model according to specified criteria(on) (the objective function(s)). Since parameters are assumed to be time-invariant or to slowly vary, the calibrated parameter set should ideally provide the best performance for any data series used later in practical applications or in validation. Therefore, the calibration process should ideally result in the parameter set(s) which optimise(s) the expectation of the objective function. Note that aiming at determining the best parameter set for the calibration may even be counter-productive; for some models – especially those with a large number of parameters, such as artificial neural networks – we may have to avoid overfitting, i.e. a case where the calibrated parameter set provides an excellent adjustment over the calibration period but is poorly predictive because the calibration process takes

into account insignificant specificities of the calibration data series (an example of a manner to avoid overfitting is Bayesian regulation, see MacKay, 1992).

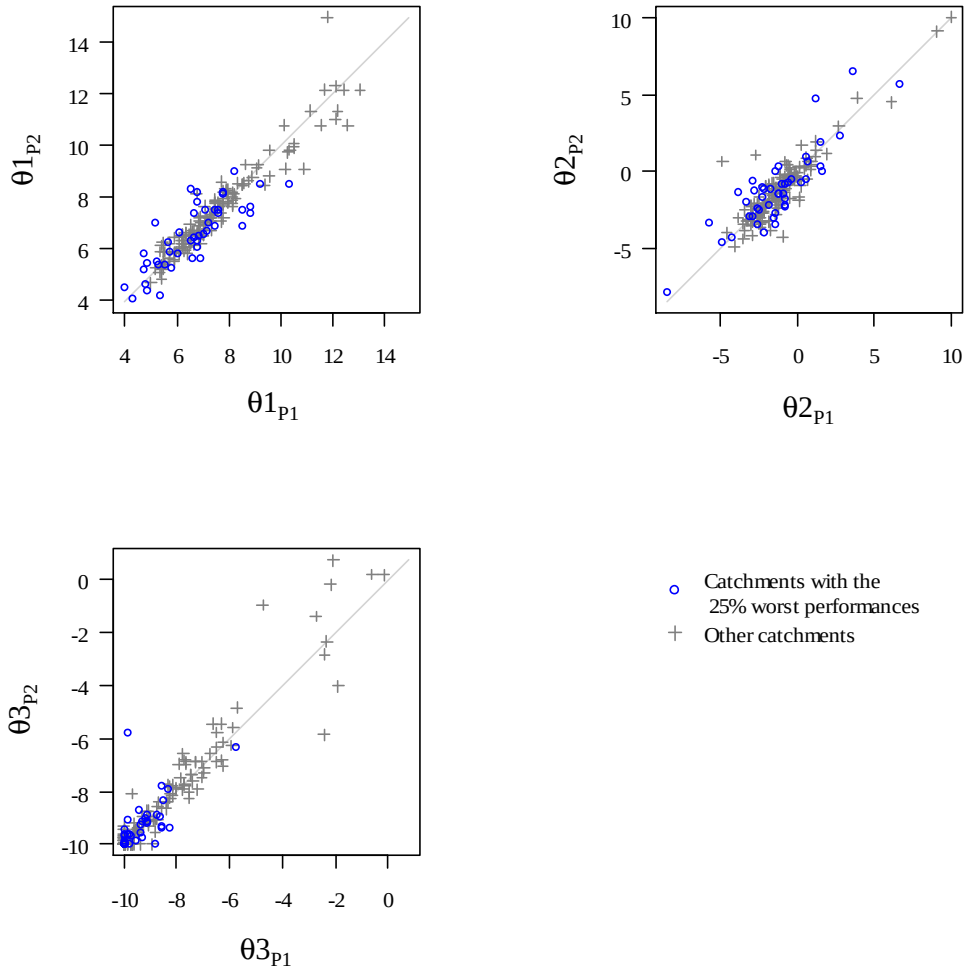
The calibrated parameter set depends to some extent on the choice of the objective function (which is commonly the RMSE). If the objective function values computed during the calibration process are not good estimates of their expectation (i.e. if the computed values reflect the calibration data series content too closely), then there is no logical reason why the model would achieve good performance (at least the best) with the calibrated parameter set over new data. The parameter sets calibrated over two different periods might be quite different. It should be noted, however, that several other reasons may explain why the calibrated parameter set differs, e.g. algorithmic problems (local-search algorithm trapped in local optima) and the fact that many current models cannot accurately predict low and high flows with the same parameter set (e.g. Wagener, 2003; Oudin *et al.*, 2006).

The real calibration process can be described as a comparison process of a number of models (that share the same structure and differ from their parameter sets) over the same data series. Actually, we are not interested in the RMSE expectations of each combination of model structure and parameters, but rather in the expectations of the comparisons between them. The comparison of two models is similar to the comparison of a model to a benchmark model discussed in the previous section. Therefore, even though the RMSE values computed for each model may be highly dependent on the data series content (as much as in the validation phase discussed previously), the comparison of those RMSE values is less dependent on the series and is more stable from one data series to another. Nonetheless, the calibration may often result in a parameter set

379 influenced to some extent by the data used; it is well known that a parameter set
380 calibrated over a rather dry period may lead to poor results on wetter periods.

381 Many authors (Jakeman and Hornberger, 1983; Sefton and Howarth, 1988; Yapo *et al.*,
382 1996) have shown that a rather short data series can be sufficient to obtain calibrations
383 that are relatively insensitive to the selected period. Perrin *et al.* (2007) and Seibert and
384 Beven (2009) have even proposed methods to calibrate a model with a very small number
385 of runoff measurements in the context of poorly gauged catchments. Conversely, Merz
386 and Blöschl (2004) showed significant differences between HBV parameter values when
387 calibrated over two distinct 11-year periods.

388 To illustrate this effect on our 178-catchment set, we compared the parameter sets
389 calibrated over two 5-year periods (called P1 and P2) for every catchment. In most cases,
390 the parameter sets are close to each other. However, for a significant number of
391 catchments, the distance between the two parameter sets is not insignificant and this
392 occurs even when the model provides good results (Fig. 8). This stability is, however,
393 better achieved for parsimonious models since they have fewer degrees of freedom to
394 adapt to the specificities of a given data series than for over-parameterised catchments
395 (Perrin *et al.*, 2001).



396

397 **Fig. 8** Comparison of the parameter values obtained by calibration on periods P1 and P2
398 (logarithm transformed values are shown here).

399

400 **Can the number of catchments compensate for the short duration of the data series?**

401 Finally, some authors (e.g. Andréassian *et al.*, 2007) advocate testing the model over a
402 large number of catchments in order to ensure acceptable model robustness. Here, the
403 model is assessed on each catchment within a set; the distribution of the criterion values
404 is scrutinised or statistics are computed from this distribution.

405 In this paper, we have shown that it is not possible to obtain a reliable assessment of the
406 RMSE expectation of a hydrological model for a given catchment when computed on a
407 short data series. Can we then rely on the distribution of RMSE over a catchment set?
408 How can we interpret the distribution of the RMSE computed on different catchments
409 over short periods? To answer this question, we randomly drew 50 distributions of the
410 RMSE computed with the synthetic data: for each draw, we randomly selected one of the
411 100 synthetic errors for each catchment and computed the corresponding RMSE simply
412 normalised by the average runoff (in order to compare RMSE over different catchments).
413 We computed the 0.05, 0.25, 0.5, 0.75 and 0.95 quantiles of each normalised RMSE
414 distribution drawn. It is worth noting that the RMSE distribution over a large number of
415 catchments is stable much sooner than the RMSE value on a single one (Fig. 9). If we
416 except the highest quantile, all the quantiles do not depend on the draw after a few years.
417 Thus, rather short data series may adequately describe the distribution of the RMSE over
418 a set of catchments (if we except the highest quantiles). This is, in our opinion, a further
419 argument for assessing models on large data sets.

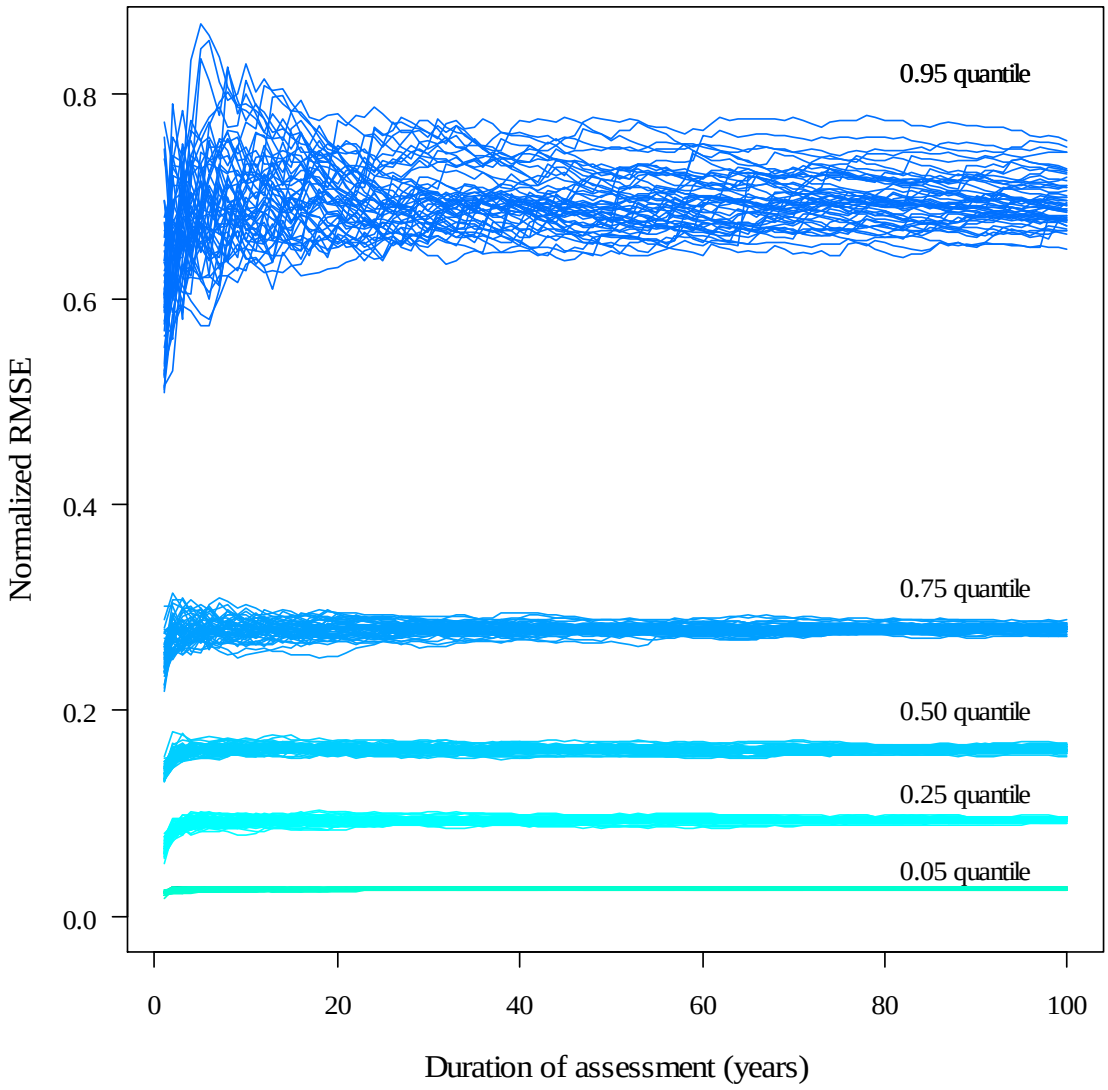


Fig. 9 0.05, 0.25, 0.50, 0.75 and 0.95 quantiles of 50 simulated distributions of RMSE (normalised by the average runoff) computed with the synthetic error series for 178 catchments: a distribution is obtained by randomly drawing one synthetic error series per catchment.

CONCLUSIONS

This paper has illustrated the large uncertainty associated with the interpretation of the RMSE value computed when assessing a hydrological model over a data series of a few years to a few decades. Here, a criterion value computed over a given data series is

considered *meaningful* only if it gives a close estimate of the criterion expectation, i.e. the level of intrinsic accuracy of the model on the catchment studied. Since we were limited by the length of the observed rainfall and runoff time series, we used the error model proposed by Yang *et al.* (2007) to assess the lower bound of the confidence interval width of the RMSE value computed over an *N*-year data series. We showed that a RMSE value of a hydrological model computed over a short data series reflects not only the actual skill of the model to reproduce the hydrological behaviour of a given catchment, but also (sometimes much more) the data series content. For most catchments, our analysis showed that data series over several decades are necessary to compute RMSE values that are marginally dependent on the data series content. Consequently, extreme caution should be taken in the use of RMSE or quadratic criteria values computed on usually short time series as estimates of the predictive model level of error.

This result is another reason to advocate the use of relative criteria instead of absolute criteria, which implicitly means using a benchmark model when assessing a hydrological model with a quadratic criterion over a single catchment. This approach helps to lessen the influence of the data series content on the computed criterion values.

When the model is assessed over many catchments, the distributions of the normalised RMSE over the catchment set turns out to be less influenced by the data series content of each catchment. Data series covering only a few years may suffice to achieve an intrinsic assessment of the model skills over the catchment set.

Acknowledgements

This work is partly funded by a research grant from the French Ministry of Agriculture. The authors gratefully acknowledge the assistance of Nicolas Le Moine (Cemagref) for

data collection and Météo France for providing meteorological data. Fruitful discussions with Nicolas Le Moine and Éric Parent helped to improve this work. This article greatly benefited from the comments of two anonymous reviewers.

REFERENCE LIST

Andréassian, V, Hall, A., Chahinian, N. & Schaake, J. (2006) Introduction and Synthesis: Why should hydrologists work on a large number of basin data sets? In: *Large Sample Basin Experiments for Hydrological Model Parameterization* (ed. by V. Andréassian, A. Hall, N. Chahinian & J. Schaake) 1–5 , IAHS Publ. 307, IAHS Press, Wallingford, UK.

Andréassian, V., Lerat, L., Loumagne, C., Mathevet, T., Michel, C., Oudin, L. & Perrin, C. (2007) What is really undermining hydrologic science today? *Hydrol. Proc.*, **21**, 2819-2822

Berthet, L. (2010). Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique. PhD Thesis, AgroParisTech, Paris, Cemagref, Antony, France, 603 pp.

Berthet, L., Andréassian, V., Perrin, C. & Javelle, P. (2009) How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. *Hydrol. Earth Syst. Sci.* **13**(6), 819–831.

Berthet, L., Andréassian, V., Perrin, C. & Loumagne, C. (2010). How significant are quadratic criteria? Part – 2: on the relative contribution of unusual flood events to the overall value of a quadratic criterion. *Hydrol. Sci. J.* (accepted).

Beven, K. and Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249**, 11–29.

Beven, K., Smith, P., and Freer, J. (2008) So just why would a modeller choose to be incoherent ? *J. Hydrol.* **354**, 15–32.

Beven, K. J. (2008b) *Environmental Modelling: An Uncertain Future?* Routledge, London, New York.

Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control.* Holden Day Inc., Oakland, California, U.S.A.

Dawson, C., Abrahart, R. & See, L. (2007) Hydrotest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Env. Mod. & Soft.* **22**(7), 1034–1052.

Houghton-Carr, H. (1999) Assessment criteria for simple conceptual daily rainfall-runoff models. *Hydrol. Sci. J.* **44**(2), 237–262.

489 Jakeman, A. J. & Hornberger, G. M. (1993) How much complexity is warranted in a
490 rainfall-runoff model ? *Wat. Resour. Res* **29**(8), 2637– 2649.
491 Khan, M. (1989) Evaluation of rainfall-discharge models with discrimination. *J. Hydrol.*
492 **108**, 63–78.
493 Kitanidis, P. & Bras, R. (1980) Real-time forecasting with a conceptual hydrologic
494 model. 2. applications and results. *Wat. Resour. Res.* **16**(6),1034–1044.
495 Klemeš, V. (1986) Operational testing of hydrologic simulation models. *Hydrol. Sci. J.*
496 **31**(1), 13 – 24.
497 Le Moine, N., Andréassian, V., Perrin, C. & Michel, C. (2007) How can rainfall-runoff
498 models handle intercatchment groundwater flows? theoretical study based on 1040
499 french catchments. *Wat. Resour. Res.* **43**(6), W06428
500 Legates, D. & McCabe Jr., G. (1999) Evaluating the use of 'goodness-of-fit' measures in
501 hydrologic and hydroclimatic model validation. *Wat. Resour. Res.* **35**(1), 233–241.
502 MacKay, D. J. C. (1992) Bayesian interpolation. *Neur. Comput.* **4**, 415– 447.
503 Martinec, J. & Rango, A. (1989) Merits of statistical criteria for the performance of
504 hydrological models. *Wat. Resour. Bull.* **25**(2), 421– 432.
505 Mathevet, T., Michel, C., Andréassian, V. & Perrin, C. (2006) A bounded version of the
506 nash-sutcliffe criterion for better model assessment on large sets of basins. In: *Large*
507 *Sample Basin Experiments for Hydrological Model Parameterization* (ed. by V.
508 Andréassian, A. Hall, N. Chahinian & J. Schaake) 211–219 , IAHS Publ. 307, IAHS
509 Press, Wallingford, UK.
510 Merz, R. & Blöschl, G. (2004) Regionalisation of catchment model parameters *J. Hydrol.*
511 **287**, 95–123.
512 Nash, J. E. & Sutcliffe, J. V. (1970) River Flow Forecasting through conceptual models:
513 part I - A: discussion of principles. *J. Hydrol.* **10**, 282–290.
514 Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. & Loumagne,
515 C. (2005) Which potential evapotranspiration input for a lumped rainfall-runoff
516 model? Part 2 – towards a simple and efficient potential evapotranspiration model for
517 rainfall-runoff modelling. *J. Hydrol.* **303**, 290 – 306.
518 Oudin, L., Andréassian, V., Mathevet, T., Perrin, C. & Michel, C. (2006) Dynamic
519 averaging of rainfall-runoff model simulations from complementary model
520 parameterizations. *Wat. Resour. Res.* **42**, W07410.
521 Perrin, C., Andréassian, V. & Michel, C. (2006) Simple benchmark models as a basis for
522 model efficiency criteria. *Large Rivers* **17**, 221–244.
523 Perrin, C., Michel, C. & Andréassian, V. (2001) Does a large number of parameters
524 enhance model performance? Comparative assessment of common catchment model
525 structures on 429 catchments. *J. Hydrol.* **242**, 275 – 301.
526 Perrin, C., Oudin, L., Andréassian, V., Rojas-Serna, C., Michel, C. & Mathevet, T. (2007)
527 Impact of limited streamflow data on the efficiency and the parameters of rainfall-
528 runoff models. *Hydrol. Sci. J.*, **52**(1), 131–151.
529 Schaefli, B. & Gupta, H. (2007) Do Nash values have value? *Hydrol. Proc.* **21**(15),
530 2075–2080.
531 Sefton, C. & Howarth, S. (1998) Relationships between dynamic response characteristics
532 and physical descriptors of catchments in England and Wales. *J. Hydrol.* **211**, 1–16.

Seibert, J. (2001) On the need for benchmarks in hydrological modelling. *Hydrol. Proc.* **15**, 1063–1064.

Seibert, J. & Beven, K. (2009) Gauging the ungauged basin: how many discharge measurements are needed? *Hydrol. Earth Syst. Sci.* **13**(5): 883–892.

Sorooshian, S. & Gupta, V. K. (1983) Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. *Wat. Resour. Res.* **19**(1), 260 – 268.

Sorooshian, S., Gupta, V. K. & Fulton, J. L. (1983) Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. *Wat. Resour. Res.* **19**, 251– 259.

Tan, S., Chua, L., Shuy, E., Lo, E.-M. & Lim, L. (2008) Performances of rainfall-runoff models calibrated over single and continuous storm flow events. *J. Hydrol. Eng.* **13**(7), 597–607.

Tangara, M. (2005) Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global, Ph.D. Thesis, Ecole pratique des hautes études de Paris & Cemagref.

Wagener, T. (2003) Evaluation of catchment models. *Hydrol. Proc*. **17**(16), 3375–3378.

Yang, J., Reichert, P., Abbaspour, K. & Yang, H. (2007) Hydrological modelling of the chaohe basin in china: Statistical model formulation and bayesian inference. *J. Hydrol*. **340**, 167–182.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *J. Hydrol.* **181**, 23 – 48.

APPENDIX

In this section, we show that the expectation of the RMSE exists in the error model by Yang *et al.* (2007).

Let F_i be the logarithm of the multiplicative error: $F_i = \log\left(\frac{\hat{Q}_i}{Q_i}\right)$ where $\hat{Q}_i$ is the model output for time step I while Q_i is the observed runoff. In the error model, Fi can be simulated by the recurrence relation $F_i = \rho \cdot F_{i-1} + I_i$ where I_i is the innovation drawn in a normal distribution $\mathbf{N}\left(0, \sigma^2 \cdot \left(1-\rho^2\right)\right)$; the first F_0 is drawn in a normal distribution $\mathbf{N}\left(0, \sigma^2\right)$.

565 Using an induction, it is easy to show that any F_i is drawn in a normal distribution
566 $\mathbf{N}\left(0,\sigma^2\right)$, since $\rho \cdot F_{i-1}$ and I_i are two independent variables drawn in two normal laws
567 $\mathbf{N}\left(0,\rho^2\sigma^2\right)$ and $\mathbf{N}\left(0,\sigma^2 \cdot \left(1-\rho^2\right)\right)$.

568 Therefore, the variable $\hat{Q}_i$ follows a lognormal distribution and its conditional
569 expectation is $\mathbb{E}\left(\hat{Q}_i | Q_i = q\right) = q \cdot e^{\frac{\sigma^2}{2}}$. Thus the expectation of the quadratic error is:

570 $$\mathbb{E}\left(\left(\hat{Q}_i - Q_i\right)^2\right) = \int_{q \in \mathbb{R}^+} q \cdot \left(e^{\frac{\sigma^2}{2}} - 1\right) \cdot g(q) dq$$

571 where g is the probability density function of the runoff.

572 This integral exists if g does not exhibit an overly heavy tail. For tails following a
573 Gumbel law or an exponential law, the expectation for the quadratic error exists.

$\mathcal{F}$

Annexe

Les critères quadratiques sont-ils vraiment continus ?

Cet article a été soumis au journal des sciences hydrologiques (*Hydrological Sciences Journal*) pour son numéro spécial consacré à l'atelier *Cour des miracles de l'hydrologie*, qui s'est tenu à Paris du 18 au 20 juin 2008. Il est actuellement accepté et en attente de publication du numéro spécial (*in press*).

Nous cherchons à déterminer si les critères quadratiques évaluent vraiment de façon continue les modèles et dans le cas contraire, quels sont les points effectivement pris en compte par un critère quadratique. En prévision, un très petit nombre de pas de temps où l'erreur du modèle est importante contribue pour l'essentiel de l'erreur quadratique totale (MSE). Il s'agit des pas de temps pour lesquels il y a la plus grosse variation de débit, donc ceux pour lesquels le modèle est susceptible d'avoir les plus grosses difficultés. Ce sont également les pas de temps qui intéressent le plus les prévisionnistes. Dans le cas d'étude présenté dans cet article, les mises-à-jour du modèle rendent le critère plus « événementiel » qu'en simulation : un plus petit nombre de pas de temps porte l'essentiel de l'erreur totale du modèle.

How significant are quadratic criteria? Part 2: on the relative contribution of large flood events to the overall value of a quadratic criterion.

LIONEL BERTHET[1,2], VAZKEN ANDRÉASSIAN[1] , CHARLES PERRIN[1], & CECILE LOUMAGNE[1]

[1] Cemagref, *Hydrosystems and Bioprocesses* Research Unit, PB 44, 92163 Antony cedex, France. E-mail: lionel.berthet @cemagref.fr
[2] AgroParisTech ENGREF, 19 avenue du Maine, F 75732 Paris

ABSTRACT

Quadratic criteria (i.e. based on squared residuals) are widely used to assess the performance of hydrological models. However, the largest errors have a relatively strong influence on the final criteria values, which may be considered a drawback for a complete assessment. This paper studies the case of updated models used for real-time forecasting. It is shown that the fraction of the data series actually impacting the final criterion value is small on many catchments and corresponds to the time steps characterised by the greatest runoff variations. In fact, model updating makes the error distribution more peak-shaped, giving even more relative importance to the time steps with the largest errors. Therefore, assessing the performance of an updated model with a quadratic criterion emphasises that these criteria focus more on the most difficult time steps to model (and the most interesting ones in the case of short-term flood forecasting).

Keywords: Model assessment; skill scores; quadratic criteria; flood forecasting; persistence index; heteroscedasticity.

INTRODUCTION

The number of scientific articles dedicated to the assessment of hydrological models shows that it is still considered an open and crucial issue. Indeed, many methodological questions must be discussed in order to design a proper assessment procedure, e.g. the way the procedure ensures aleation and repetition (Clarke, 2008). Among these questions, the measure of the distance between observations and simulation is of primary importance.

The choice of this measure depends fundamentally on the aim underlying the assessment (Wagener and McIntyre, 2005). End-users use measures which are far from being a formal (or informal) likelihood: for example, if the model is used only for flood forecasting, an end-user may choose a measure that focuses on high flow errors. Conversely, to diagnose model deficiencies and to improve the model, Wagener *et al.* (2001) suggest a multi-criteria assessment based on pattern signatures. Indeed, using one measure does not allow a comprehensive assessment of the model and no criterion assesses the current models for the entire range of runoff from low flows to floods (Gupta *et al.*, 1998; Wagener, 2003). Finally, these criteria are also a tool for communication (Schaefli and Gupta, 2007; Pappenberger and Beven, 2006), which implies that they should be interpretable unambiguously not only by modellers, but also by end-users.

Deficiencies of quadratic criteria

Quadratic criteria (i.e., criteria based on square residuals) are among the most popular numerical criteria, even if they are known to suffer from many deficiencies. Among their drawbacks, it is worth pointing out that all errors are not treated identically due to the quadratic nature of these criteria: much more importance is given to the largest errors

(Martinec and Rango, 1989; Legates and McCabe, 1999). These errors occur most frequently on flood peaks. However, few papers discuss quantitatively the impact of high flow events on the quadratic criteria computation. In the companion paper (Berthet *et al.*, 2010), we showed that the interpretation of quadratic criteria in terms of predictive accuracy expectation is far from being straightforward.

These properties make it difficult to interpret the values of a quadratic criterion. Most often, a benchmark model is required, which is an easy-to-understand model whose performance is compared to the performance of the model tested. Choosing this benchmark is difficult since a given benchmark model does not perform identically on different catchments (Martinec and Rango, 1989; Legates and McCabe, 1999; Seibert, 2001; Perrin *et al.*, 2006; Schaefli and Gupta, 2007) and with different uses of the model. For flood forecasting, the benchmark model embedded in the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970) is far too naïve (it "forecasts" a constant flow equal to the mean observed flow) and a specific benchmark for flood forecasting is needed: Kitanidis and Bras (1980) suggested using the persistence model as the benchmark.

Different approaches have been proposed to facilitate the interpretation of a quadratic criterion. Murphy (1988) and Pebesma *et al.* (2007) exhibit different decompositions of the root mean square errors in a number of key elements. Gupta *et al*. (2009) used a decomposition to highlight several limitations of the quadratic criteria. Since quadratic criteria do not treat all errors identically; it is necessary to understand which errors have a real impact on the criteria values. Krause *et al*. (2005) proposed an approach to detect when the value of a criterion evolves, which helps to understand which information is really embedded by a given criterion.

74

75

76 **Impacts of data assimilation procedures**

77 When used for flood forecasting, hydrological models are usually updated, through some
78 data assimilation scheme. Data assimilation (model updating) modifies the distribution of
79 errors in a non homogeneous way (Liu and Gupta, 2007). This can lead to changes in the
80 subset of the errors in a series which truly account for the value of a quadratic criterion.
81 Thus, the interpretation of a given criterion may have to evolve if the model is updated
82 and the quadratic criteria might be less well adapted to assessing hydrological models
83 dedicated to flood forecasting.

84 **Scope of the paper**

85 In this paper, we pursue two objectives. Our first aim is to assess the fraction of an error
86 series (i.e. how many time steps?) that is truly significant in the computation of a
87 quadratic criterion used to evaluate a hydrological forecasting model (i.e. for the RMSE,
88 the fraction of the error series that represents the largest part of the total quadratic error).
89 Our second objective is to assess the extent to which model updating (very often used in
90 real-time applications) changes the error distribution and to determine whether these
91 changes strengthen or weaken the influence of the largest errors on the assessment by
92 quadratic criteria.

93

DATA AND MATERIALS

The study was conducted on a set of 178 French unregulated catchments, described in the companion paper (Berthet et al., 2010).

The GRP rainfall-runoff forecasting model and its updating techniques

The GRP model is a continuous, lumped, hybrid metric-conceptual model designed for flood forecasting (Tangara, 2005; Berthet, 2010). It is one of the operational models used to forecast river flows in real time on French catchments, including the Seine river basin upstream of Paris.

The GRP model's structure was derived from the structure of the GR4J model (Perrin *et al.*, 2003). It can classically be described as the combination of a production function (which computes the effective rainfall) followed by a routing function (Fig. 1). The former consists in a non linear soil moisture accounting (SMA) store and a volume-adjustment coefficient that determines the runoff ratio. The routing function is composed of a unit hydrograph and a nonlinear routing store.

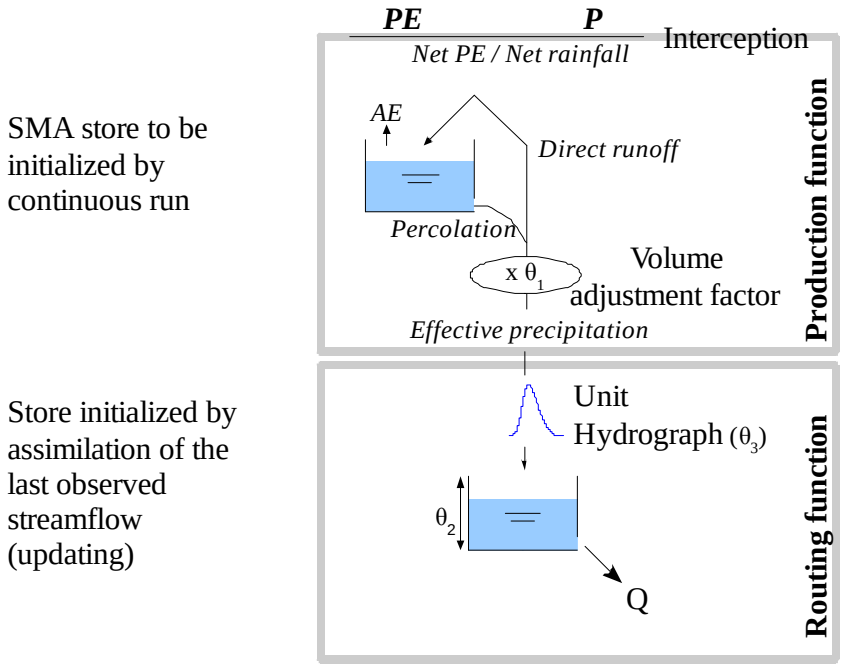


Fig. 1 Structure of the GRP model. PE, potential evapotranspiration; P, rainfall; Q, runoff; θ_i are the parameters.

The GRP model is parsimonious. Three parameters have to be calibrated: one in the production function (θ_1, the volume adjustment factor) and two for the routing function (θ_3, the base time of the unit hydrograph and θ_2, the total capacity of the routing store).

Different data assimilation approaches have been tested for this structure to obtain more accurate forecasts at lead times ranging from 1 h to a few days. This preliminary analysis (not shown here) indicated that the most efficient updating strategy for the GRP model is a combination of a direct updating of the routing store level (using the last observed discharge measurement) and an output updating based on the last observed forecast error (see Berthet, 2010, for more details). Hereafter, only this most efficient updating technique combination for the GRP model is used.

Quadratic criteria

Many quadratic criteria are available in the literature to assess the performance of a hydrological model. Since we are interested here in flood forecasting, we used the persistence index (Kitanidis and Bras, 1980), given by:

$$PI_L = 1 - \frac{MSE^{TM}}{MSE^{PM}} = 1 - \frac{\sum_t \left(\hat{Q}_{t+L|t} - Q_{t+L}\right)^2}{\sum_t \left(Q_t - Q_{t+L}\right)^2} \quad (1)$$

where L is the lead time of the forecast, Q_t is the observed runoff at time step t and $\hat{Q}_{t+L|t}$ is the forecast of runoff made at time step t for the time step $t+L$; TM and PM stand for tested model and persistence model, respectively.

The persistence index compares the forecast mean square error made by the tested model to the forecast mean square error of a naïve model – the persistence model – which gives the last observed runoff as a forecast. Its values range from $-\infty$ to 1: the higher the value, the better the performance. A positive value indicates that the tested model performs better than the persistence model.

As we were interested in some specific time steps, we also used the partial persistence index (PPI), which is defined as follows: let us sort the error series in some order O:

1. $O = O_C$: chronologically order ; the PI is computed from time step 1 to time step n;
2. $O = O_{DD}$: discharge descending order to study the impact of flow magnitude;
3. $O = O_{RV}$: runoff variation descending order to study the impact of discharge variations.

4. $O = O_{ME}$: model errors descending order to study the impact of the modelling over the quadratic criteria assessment.

The partial persistence index with respect to the ranking order O is the persistence index computed over the n first elements of the sorted series $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\}$:

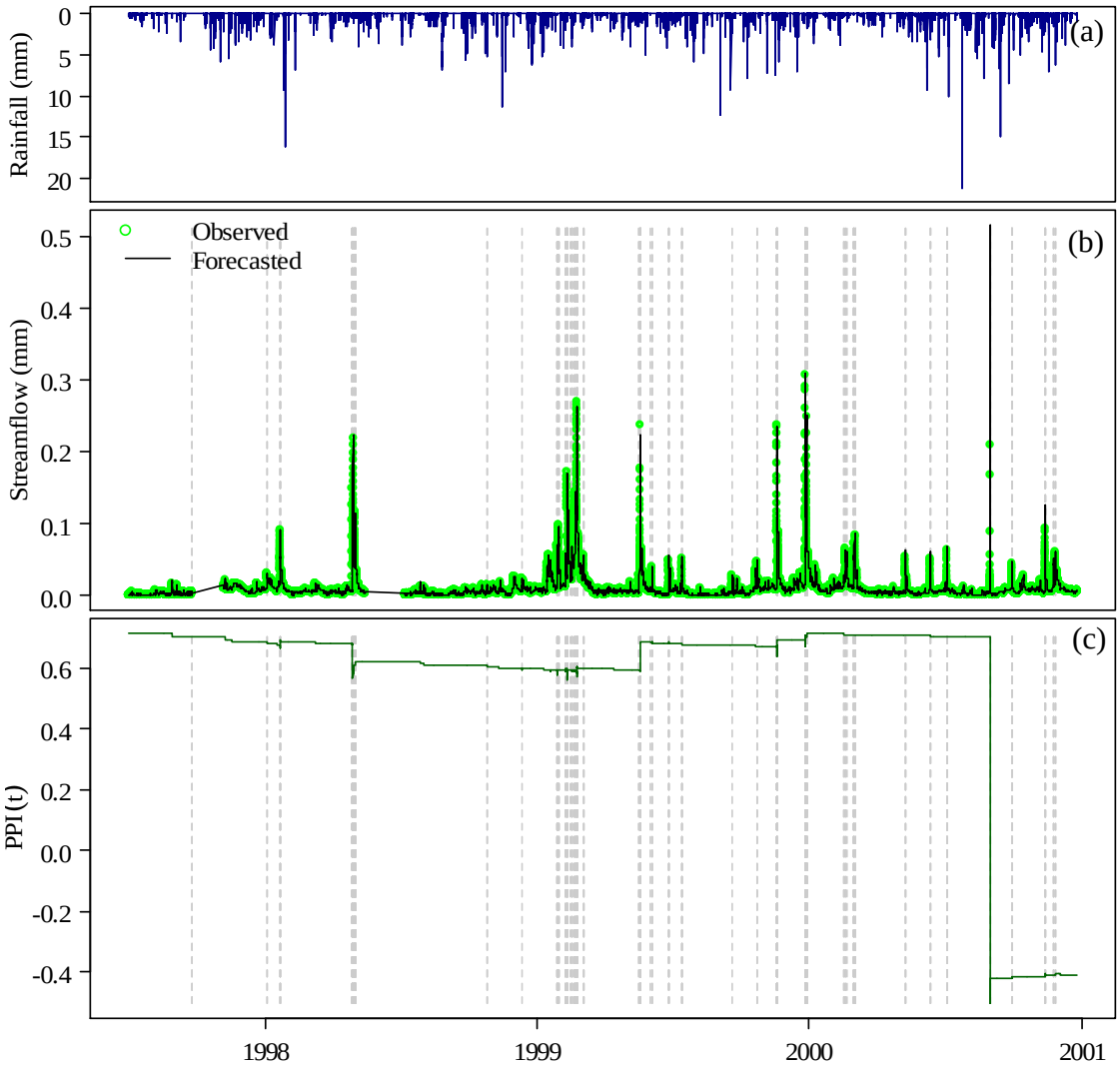
$$PPI_{L,O}(n) = 1 - \frac{PMSE_{L,O}^{TM}(n)}{PMSE_{L,O}^{PM}(n)} = 1 - \frac{\sum_{t=1}^{n}\left(\hat{Q}_{t+L|t} - Q_{t+L}\right)^2}{\sum_{t=1}^{n}\left(Q_t - Q_{t+L}\right)^2} \quad (2)$$

In this paper, we also used the partial mean square errors of the tested model ($PMSE_{L,O}^{TM}$ in equation (2)) and of the persistence model ($PMSE_{L,O}^{PM}$):

$$PMSE_{L,O}^{TM}(n) = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{Q}_{t+L|t} - Q_{t+L}\right)^2 \quad \text{and} \quad PMSE_{L,O}^{PM}(n) = \frac{1}{n}\sum_{t=1}^{n}\left(Q_t - Q_{t+L}\right)^2 \quad (3)$$

THE OUDAN RIVER CASE STUDY: WHEN A SINGLE FLOOD EVENT OVER 5 YEARS DETERMINES THE CRITERION VALUE

Let us start our analysis with a case study illustrating the influence of a single flood event on the assessment of the flood forecasting model: the catchment of the Oudan River at Riorges (Fig. 2). The hydrological model is run (in validation mode) over the period 1997–2001 (preceded by a 1-year warm-up period over which the criterion is not computed).



156

157 **Fig. 2** Case study of the influence of one flood event (30 August 2000) on the Oudan River at
158 Riorges (27 km²). Observed runoff is denoted by light circles while the solid line represents the
159 forecasts (for a 1-h lead time). Grey vertical dashed lines indicate the time steps with the 5%-
160 highest variation of flow. (a) Rainfall. (b) Observed and forecasted runoff. (c) Partial persistence
161 index (PPI) with respect to the chronological order O_C.

162

163 Fig. 2b shows how the PPI evolves with respect to the chronological order (i.e. the
164 persistence index is calculated from 1 October 1997 to t). The PPI_{L,O_C} value changes

significantly at a limited number of time steps. Between these time steps, the PPI_{L,O_C} decreases very slowly if it is positive or increases very slowly if negative (an explanation for this behaviour is proposed in the appendix). The time steps with significant changes in the PPI_{L,O_C} value coincide almost perfectly with the time steps exhibiting the largest variations of streamflow.

The final value of the PPI_{L,O_C} (i.e. the PI value over the 1997–2001 period) is mostly determined by one event: the flood of 30 August 2000. On this specific event, the model performs poorly, whereas other floods had been satisfactorily forecasted (it is not within the scope of this paper to discuss the many possible reasons for the model's failure on this event).

The partial persistence index computed with respect to different ranking orders is displayed in Fig. 3. When the data series are ranked with respect to the decreasing order of the runoff variations O_{RV}, less than 3% of the whole data series is necessary to obtain a $PPI_{L,O_{RV}}$ value that is very close to the final (overall) PI value ($\forall x \geq 3\%, \left| PPI_{L,O\,RV}(x) - PI \right| \leq 0.03$).

This case study illustrates an extreme behaviour: even if the flood which led to this behaviour is not an exceptional one, the modelled response is fortunately quite unusual. Is this an exception? In the following, on the basis of a large set of catchments, we assess to which extent the supposedly continuous assessment of a forecasting model can be event-based.

METHODOLOGY

Two different approaches are proposed to answer the questions raised in the introduction and illustrated in the Oudan River case study: the assessment of the fraction of the error series that is actually taken into account in the final value of a quadratic criterion and the influence of model updating in forecasting models.

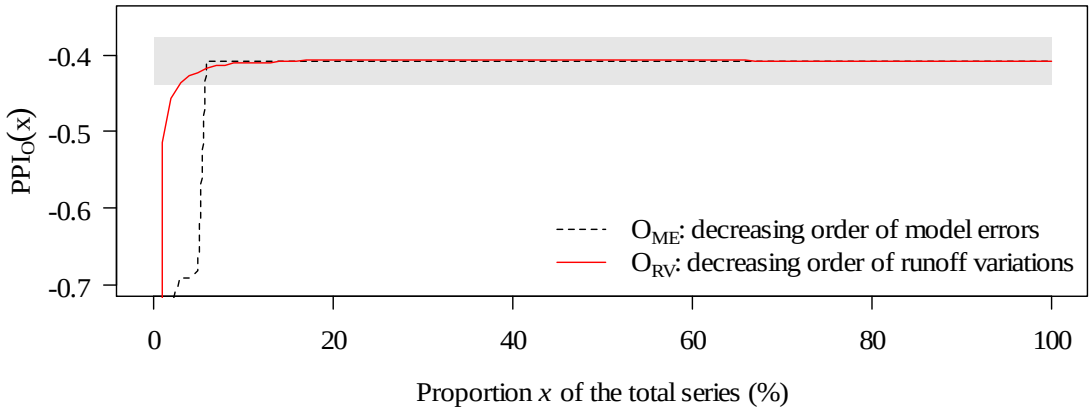


Fig. 3 Partial persistence indexes computed over a percentage x of the 5-year data series, with respect to different ranking order O for the Oudan River. Dashed line: data series are ranked following the decreasing order of the tested model errors. Solid line: data series are ranked following to the decreasing order of the runoff variations. The grey area denotes the final value of the PPI_O (i.e. the PI value computed over the complete data series) ± 0.03.

First step: assessment of the fraction of the time series actually impacting the criterion value

The analysis made for the previous case study (Fig. 3) was carried out for the 178 selected catchments. For each catchment, we assessed the partial persistence index $x \mapsto PPI_{L,O_{RV}}(x)$ with respect to the decreasing order of the runoff variations O_{RV}. Therefore, for any x value, $PPI_{L,O_{RV}}(x)$ is the partial persistence index computed over

204 the *x* % of the time steps showing the highest runoff variations. Then, for each catchment,
205 we determined the fraction of the data series actually influencing the persistence index
206 value. It was defined as the minimum *x′* such that the partial persistence index computed
207 for any *x* higher than *x′* is close enough to the persistence index:
208 $x' = \min\left(\left\{x / \forall y \geq x, \left|PPI_{L,O\,RV}(y) - PI\right| \leq 0.03\right\}\right)$.

209 The distribution of the *x′* fractions is finally scrutinised.

210 **Second step: how much do the updating techniques modify the error distribution?**

211 We assessed the changes in the error distribution attributable to the data assimilation of
212 the flow forecasting GRP model. The model was run with and without updates for
213 different lead times: 1, 6, 24, 48 and 72 h. Since the quadratic criteria are mostly
214 influenced by the errors of relatively high magnitude compared to the magnitude of the
215 common errors, we were interested in the distribution shape, i.e. the distribution of the
216 errors normalised by the range of the errors. The error series for 178 catchments are
217 compared visually and using an index reflecting the distribution spread that is their
218 standard deviation normalised by their range. This index characterises the shape (for a
219 unimodal distribution); it ranges from 0 (Dirac's delta) to 0.5 (two equi-weighted Dirac's
220 delta located at the bounds of the interval); a uniform distribution obtains an index value
221 of $\sqrt{1/12}$ (approximately 0.29). The index contributes information on the weight of the
222 extreme error values compared to the usual values.

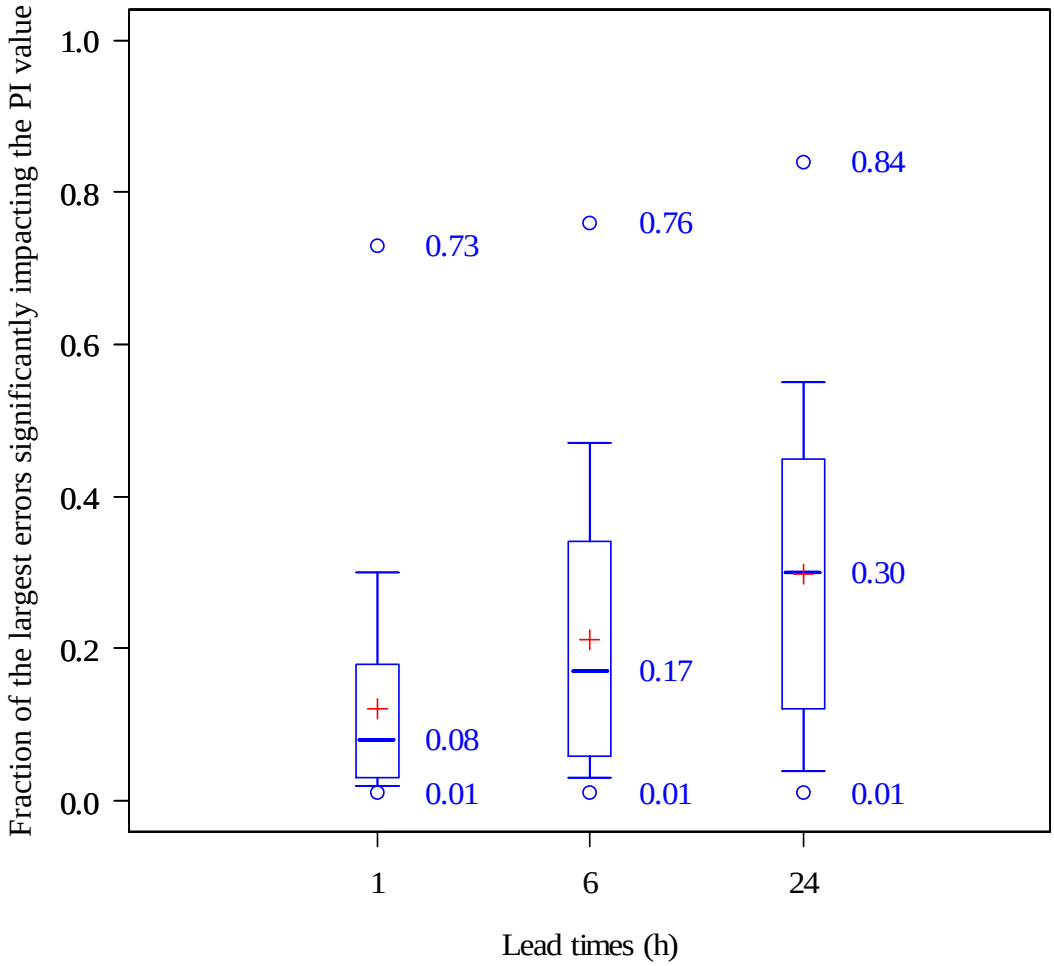
223 Finally, to further describe the impact of this change in the error distribution, on the
224 quadratic criteria, we assessed the partial mean square error $x \mapsto \mathrm{PSME}^{TM}_{L,O_{ME}}(x)$ with
225 respect to the decreasing order of the square errors O_{ME} for each catchment. Similarly to

the first step, we determined the minimum x for which the partial MSE is higher than 0.99 of the total MSE. An analysis of the sensitivity to different factors (catchment size, catchment reactivity) is provided.

RESULTS AND DISCUSSION

What fraction of the data series actually influences the PI value?

The distribution (over the set of 178 catchments) of the data series fractions that have a significant impact on the persistence index value is displayed in Fig. 4. For the shortest lead times (i.e. when the influence of the updates is the strongest), the fraction is actually very low: for a majority of catchments, less than 8% of the data series has a significant impact on the persistence index. For a small number of catchments, a single event has a determining impact. For longer lead times, the PI value is influenced by a much wider fraction of the data series. This is why we will later address the influence of model updating on the error distribution.



240

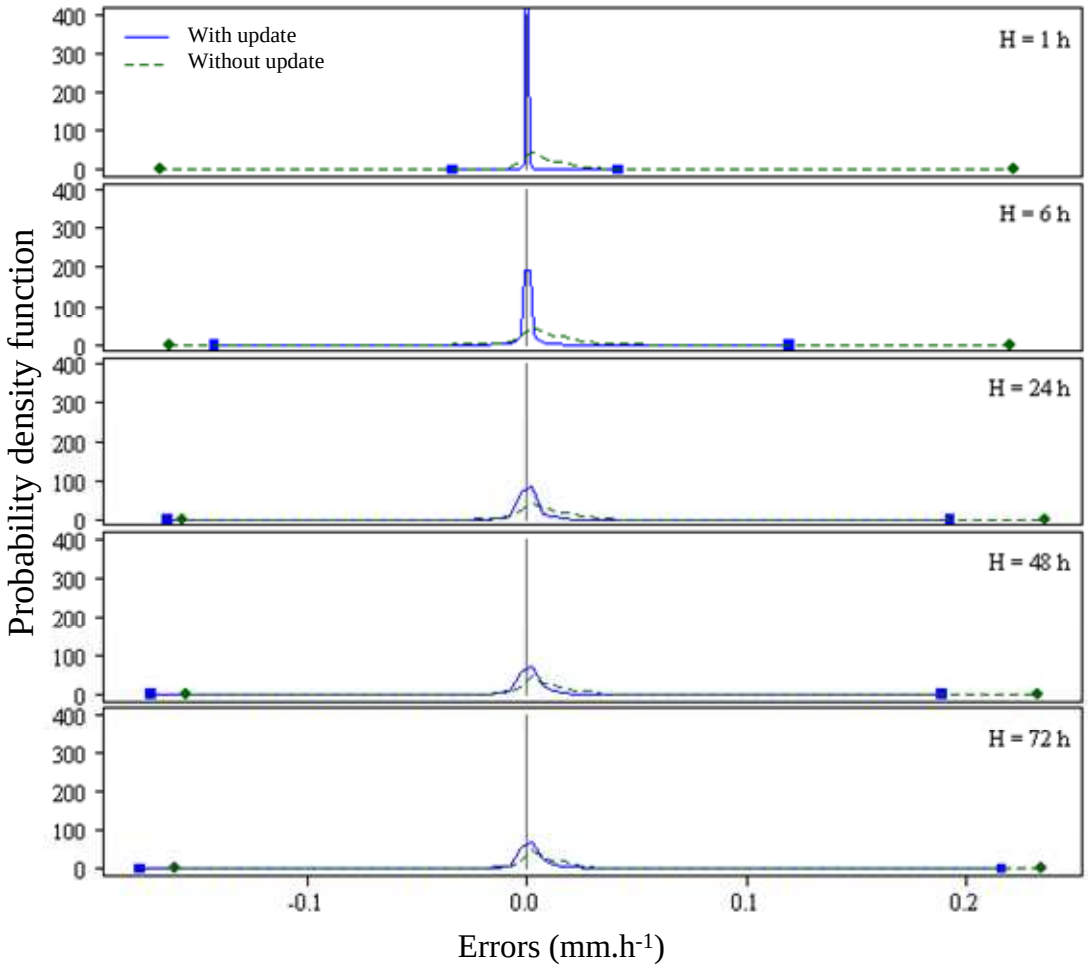
241 **Fig. 4** Distribution (over the set of 178 catchments) of the fractions of the data series which have
242 a significant impact on the persistence index value, depending on the lead time. Quantiles
243 displayed are 0 (minimum value), 0.05, 0.25, 0.5, 0.75, 0.95 and 1. The cross indicates the mean
244 of the distribution.

245

246 **Changes in the shape of the error distribution due to the updates**

247 Fig. 5 displays the error distributions for one example (The Vire River at Tessy-sur-Vire,
248 639 km²). As expected for short lead times, the error range is smaller when the model is
249 updated than when the model is not. When the lead time increases, the range of the errors

250 becomes comparable, which was expected since the effects of the updating process
251 vanish with time (Berthet et al., 2009a,b). However, what really matters here is how the
252 shape of the distribution contributes, in order to understand what fraction of the data
253 series influences the criterion. We observe that the error distribution is much closer to a
254 peak shape when the model is updated than when it is not.



255

256 **Fig. 5** Example of probability density functions of the errors made by the model with or without
257 updates for forecasts at different lead times *L* (Vire River at Tessy-sur-Vire, 639 km²).

258

259 The situation is similar for a large majority of catchments. This is summarised in Fig. 6,
260 which provides the distribution of the spread index for different lead times. The spread

index is an estimate of the relative width of error distribution. The shorter the lead time, the more peaky the normalised error distribution for the updated model is: when the model is updated, the common errors are concentrated within a much smaller range compared to the extreme errors' magnitude. The updates strengthen the heteroscedasticity of the hydrological model. For longer lead times, the influence of the updates decreases and the spread index comes closer to the value of this index computed with the model without updates. However, this influence does not completely vanish even for a 72-h lead time.

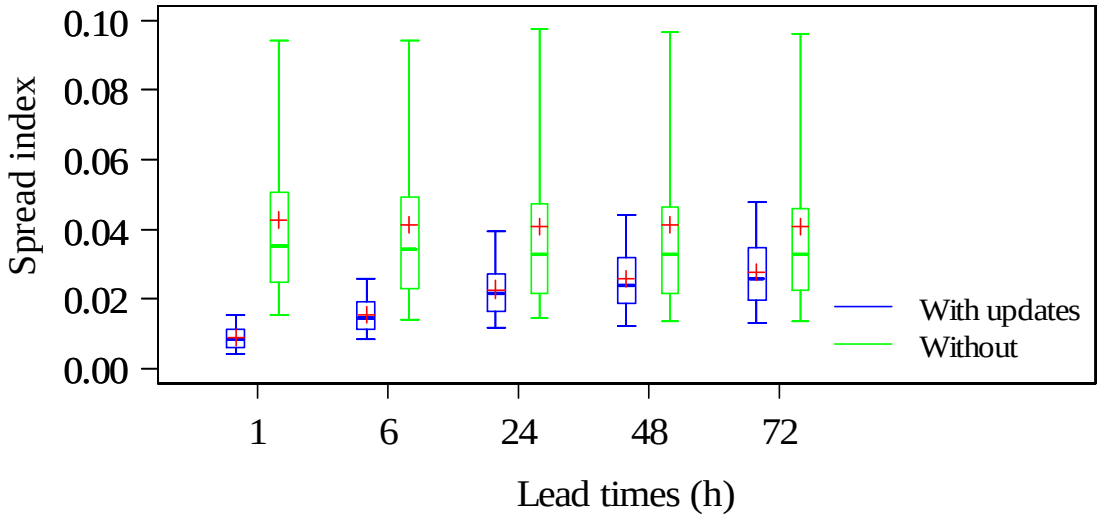


Fig. 6 Distributions over the 178-catchment set of the spread index of the error distribution (standard deviation normalised by the range) computed with the forecasting model with and without updates. The index is 0 for a Dirac delta but 0.29 for a uniform distribution. Quantiles shown are 0.05, 0.25, 0.5, 0.75, and 0.95. The crosses indicate means of the distributions.

Influence of model updating through the assimilation of the last observed runoff on the error distribution

Model updating has a clear influence on the shape of the error distribution: the extreme errors are larger compared to the range of usual errors when the model is updated than

when it is not. Quadratic criteria are mostly impacted by the largest errors; we assessed the fraction of the largest errors necessary to compute a PMSE value close to the MSE. Fig. 7 shows the partial MSE with respect to the decreasing order of the square errors O_{ME} for one catchment (The Vire River at Tessy-sur-Vire, 639 km²). As expected, no significant change with the lead time is observed when the model is not updated. The error distributions of the updated model get closer to the distributions obtained with a model that has not been updated when the lead time increases. When the model is updated, a much smaller fraction x of the data series is sufficient to accurately assess the final value of the MSE than without updates.

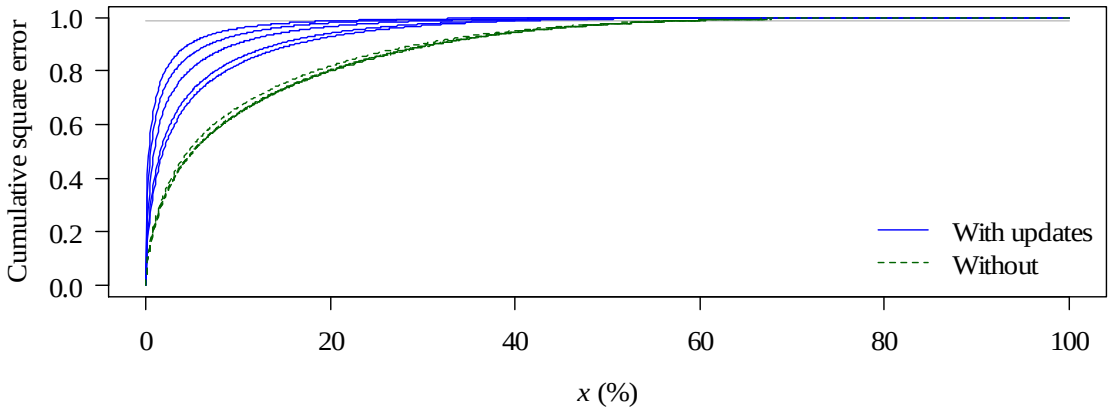
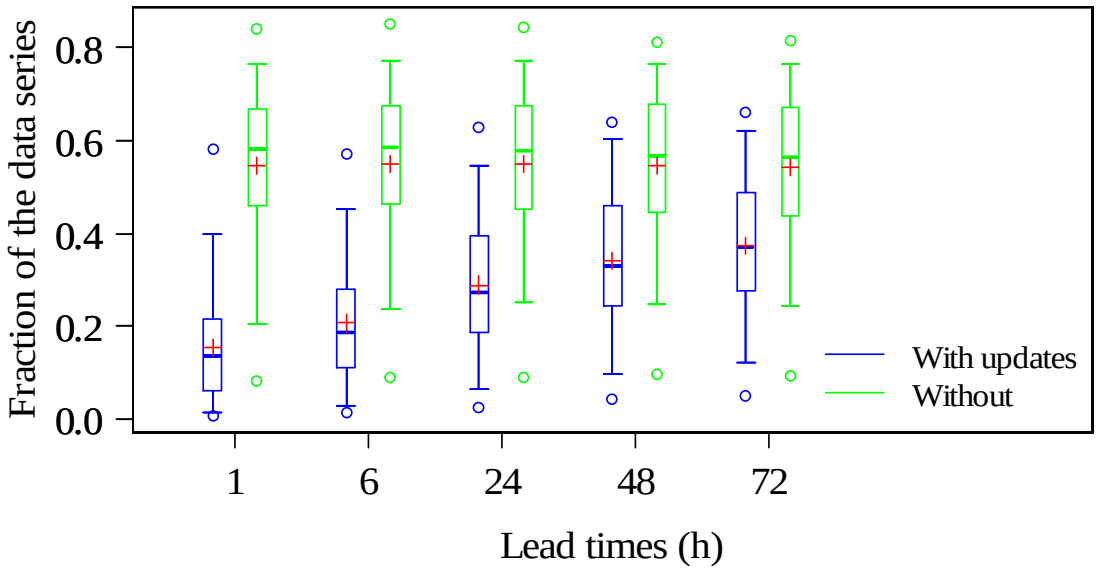


Fig. 7 Normalised cumulative square errors (square errors are ranked by decreasing order O_{ME}) obtained by the model for forecasts on the Vire River at Tessy-sur-Vire (639 km²) at 1-h, 6-h, 24-h, 48-h and 72-h lead times. Solid lines: the model is updated (the longer the lead time, the lower the cumulative square error). Dashed lines: the model is not updated. The light grey horizontal solid line indicates the 0.99 threshold.

The results over the entire set of 178 catchments are very similar to the results shown for the previous example (Fig. 8). For a 1-h lead time, less than 20% of the data series really influences the MSE values for 75% of the catchments.



296

297 **Fig. 8** Distribution over 178 catchments of the data series fractions necessary to compute a MSE
298 value greater than 0.99 of the total MSE (calculated over the whole data series). Quantiles shown
299 are 0 (minimum value), 0.05, 0.25, 0.5, 0.75, 0.95 and 1 (maximum value). The cross indicates
300 the mean of the distribution.

301

302 **What if the criteria are only computed on high flows?**

303 When the model is dedicated to flood forecasting, some users may choose to assess its
304 performance only on high flows or on a set of selected flood events. The distribution of
305 the high-flow errors differs from the complete error distribution. To check whether this
306 modifies the above results, we applied the same methodology but only for time steps with
307 the 10% highest streamflow values. The results were quite similar to the previous results:
308 for example, half of the high-flow time steps have a real impact on the MSE value (i.e.
309 99% of the total high-flow MSEs are achieved with 50% of the high-flow errors) for most
310 of the catchments (at a lead-time of 6 h).

311

Does this behaviour depend on the catchment size or reactivity?

To check whether the previous results depend on the physical or behavioural properties of the catchments (e.g. catchment size), the 178 catchments were ranked by increasing value of the property of interest and the catchment set was split into four subsets of equal size. The distributions of the spread index values of the error distributions were scrutinised for the four subsets.

For the catchment area, the distributions of the spread index values are very similar for the four subsets: there is no apparent dependency on catchment size.

We used the same technique in order to detect a potential dependence on a behavioural characteristic of the catchment: its reactivity. To characterise catchment reactivity, we used the discharge autocorrelation. Here again, all the groups obtained very similar distribution shapes and no dependence could be detected.

Dependence of the results on the hydrological model

The previous results were obtained using a single model and therefore may be model-dependent. However, many hydrological models present error distributions characterised by the same properties: temporal dependence, heteroscedasticity, etc. (Sorooshian *et al.*, 1983; Beven and Binley, 1992; Yang *et al.*, 2007); therefore the conclusions drawn here would remain valid to some extent for other models. Our results may also depend strongly on the data assimilation schemes. Nonetheless, it remains that all efficient updating methods will drastically modify the error distribution obtained in simulation and therefore the questions raised here still apply. A desirable property of the updating techniques is also to reduce the largest errors significantly (even if they are the most difficult to handle): it is indeed valuable to get a peak-shaped error distribution centred on

0 but the largest errors also have to be proportionally more reduced than common small errors, which is not the case in this study.

Can we calibrate flood forecasting models in *simulation mode*?

Updating may be understood as a process where our knowledge of the errors made by the model is used to reduce the prediction uncertainty (e.g. Liu and Gupta, 2007). In a probabilistic framework, model updating aims at obtaining more informative distributions (e.g. by Bayesian approaches) of the parameter values, the state values, etc. Unfortunately, error distribution is complex and still only partially understood: showing a formal likelihood is currently a difficult task.

For operational real-time applications, many updating methods remain deterministic: they do not provide the probability distributions of the model's components or outputs but only an updated value, even if some of these methods are based on a probabilistic approach, such as the Kálmán filter (if an exception is made of the ensemble Kálmán filter – see Evensen, 2003) or ARIMA output updating (Box and Jenkins, 1976). Within this framework, the model may interact with the updating procedure: the calibrated model parameters for the model without updating may differ from the parameters calibrated with the updated model and then be sub-optimal.

We calibrated the model in both modes (with and without updating it) and we examined the distance between the two parameters set in the parameter space for each catchment. In most cases, they were close but not identical (not shown here for the sake of brevity). We also quantified the difference of the quadratic criterion assessed on a nonoverlapping validation period (Fig. 9). For a large majority of catchments, model performance

(assessed by the persistence index) is better when the model had been calibrated using its updating procedure than when calibrated in simulation mode.

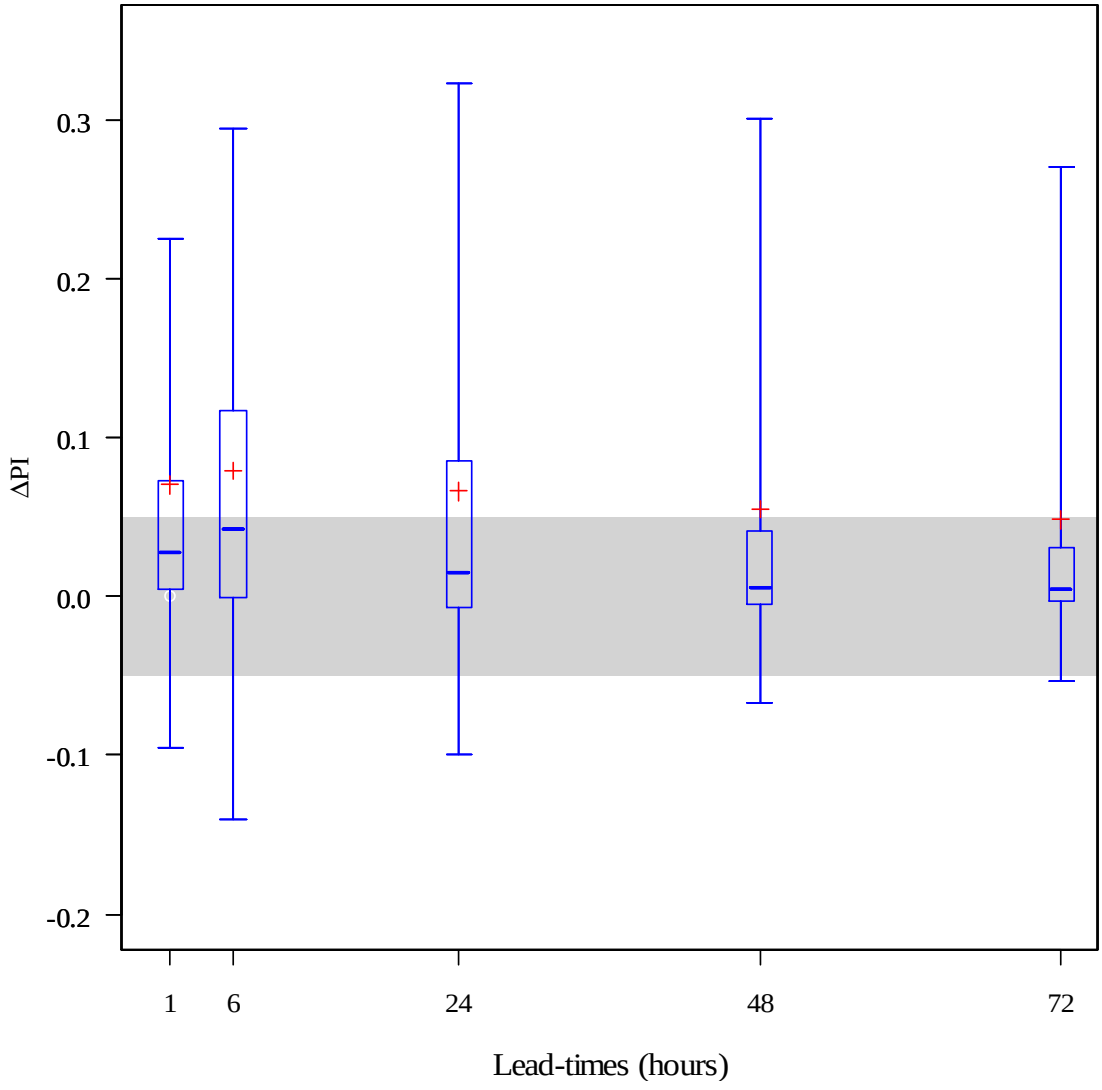


Fig. 9 Distributions of the PI values obtained by the model on a validation period when using the parameter sets obtained by calibration of the model with and without updating. Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 are displayed. The cross indicates the mean. A positive difference implies that the model performed better with the parameters set calibrated with the updated model. The grey zone ($|\Delta PI| < 0.05$) denotes what we consider to be an insignificant PI difference.

CONCLUSIONS

This paper has illustrated with an example that the value of the persistence index – a quadratic criterion commonly used to assess the performance of forecasting hydrological models – computed over a continuous data series is mostly influenced by a very small number of time steps characterised by a high runoff variation.

Then we showed that the fraction of the data series which actually influences the value of the persistence index is very small on many catchments (from 10% to 15% depending on the lead time in forecasting mode). This result would be the same for any quadratic criterion, such as Nash and Sutcliffe efficiency. For a given catchment and a given model, the smaller the lead time is, the smaller this fraction is. This fact indicates that model updates influence this fraction.

This dependence on a limited number of time steps is well known for the quadratic criteria. Here, however, model updating (a technique applied to many flood forecasting hydrological models) substantially modifies the error distribution shape: the updating techniques used here make the distribution much more peak-shaped; when the model is updated, the magnitude of the extreme errors is relatively high compared to the magnitude of common errors. The direct consequence is that a much smaller number of time steps has an impact on the computed value of a quadratic criterion. The well-known drawbacks of quadratic criteria are even more pronounced in the case of updated models. Thus, assessing an updated hydrological model with a quadratic criterion mostly comes down to assessing it on the time steps of large runoff variations (i.e. the most valuable time steps in assessing the model's skill for flood forecasting).

Acknowledgements

This work is partly funded by a research grant from the French Ministry of Agriculture. The authors gratefully acknowledge the assistance of Nicolas Le Moine (Cemagref) for gathering hydrological data and Météo France for providing meteorological data. Two anonymous reviewers are also thanked for their useful comments that helped to clarify the text.

REFERENCE LIST

Andréassian, V, Hall, A., Chahinian, N. & Schaake, J. (2006) Introduction and Synthesis: Why should hydrologists work on a large number of basin data sets? In: *Large Sample Basin Experiments for Hydrological Model Parameterization* (ed. by V. Andréassian, A. Hall, N. Chahinian & J. Schaake) 1–5 , IAHS Publ. 307, IAHS Press, Wallingford, UK.

Berthet, L. (2010). Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique. PhD Thesis, AgroParisTech, Paris, Cemagref, Antony, France, 603 pp.

Berthet, L., Andréassian, V., Perrin, C. & Javelle, P. (2009a) How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. *Hydrol. Earth Syst. Sci.* **13**(6), 819–831.

Berthet, L., Ramos, M., Perrin, C., Andréassian, V. & Loumagne, C.(2009b). Can discharge assimilation methods be used to improve flood forecasting when few data are available? *IAHS-AISH Publication n° 33*, 94 – 100.

Berthet, L., Andréassian, V., Perrin, C. & Loumagne, C. (2010) How significant are quadratic criteria? Part – 1: How many years are necessary to ensure the data-independence of a quadratic criterion value? *Hydrol. Sci. J.* (accepted).

Beven, K. & Binley, A. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.* **6**(3), 279– 298.

Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*. Holden Day Inc., Oakland, California, U.S.A.

Clarke, R. (2008) Issues of experimental design for comparing the performance of hydrologic models. *Wat. Resour. Res.* **44**, W01409.

Evensen, G. (2003) The ensemble Kálmán filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53**(4), 343–367.

Gupta, H., Sorooshian, S. & Yapo, P. (1998) Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Wat. Resour. Res*. **34**, 751–763.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez-Baquero, G. F. (2009) Decomposition of the mean squared error & nse performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **377**, 80–91.

Kitanidis, P. & Bras, R. (1980) Real-time forecasting with a conceptual hydrologic model. 2. applications and results. *Wat. Resour. Res.* **16**(6),1034–1044.

Krause, P., Boyle, D. & Bäse, F. (2005) Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **5**, 89–97.

Legates, D. & McCabe Jr., G. (1999) Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Wat. Resour. Res.* **35**(1), 233–241.

Liu, Y. & Gupta, H. (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Wat. Resour. Res.* **43**, W07401.

Martinec, J. & Rango, A. (1989) Merits of statistical criteria for the performance of hydrological models. *Wat. Resour. Bull.* **25**(2), 421– 432.

Murphy, A. (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.* **116**(12), 2417–2424.

Nash, J. E. & Sutcliffe, J. V. (1970) River Flow Forecasting through conceptual models: part I - A: discussion of principles. *J. Hydrol.* **10**, 282–290.

Pappenberger, F. & Beven, K. (2006) Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Wat. Resour. Res.* **42**, W05302.

Pebesma, E., Switzer, P. & Loague, K. (2007) Error analysis for the evaluation of model performance: Rainfall-runoff event summary variables. *Hydrol. Proc.* **21**(22), 3009–3024.

Perrin, C., Michel, C. & Andréassian, V. (2003) Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* **279**, 275 – 289.

Perrin, C., Andréassian, V. & Michel, C. (2006) Simple benchmark models as a basis for model efficiency criteria. *Large Rivers* **17**, 221–244.

Schaefli, B. & Gupta, H. (2007) Do Nash values have value? *Hydrol. Proc.* **21**(15), 2075–2080.

Seibert, J. (2001) On the need for benchmarks in hydrological modelling. *Hydrol. Proc.* **15**, 1063–1064.

Sorooshian, S., Gupta, V. K. & Fulton, J. L. (1983) Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. *Wat. Resour. Res.* **19**, 251– 259.

Tangara, M. (2005) Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global, Ph.D. Thesis, Ecole pratique des hautes études de Paris & Cemagref.

Wagener, T. (2003) Evaluation of catchment models. *Hydrol. Proc.* **17**(16), 3375–3378.

Wagener, T., Boyle, D., Lees, M., Wheater, H., Gupta, H. & Sorooshian, S. (2001) A framework for development and application of hydrological models. *Hydrol. Earth Syst. Sci.* **5**(1), 13–26.

Wagener, T. & McIntyre, N. (2005) Identification of rainfall-runoff models for operational applications. *Hydrol. Sci. J.* **50**(5), 735–751.

Yang, J., Reichert, P., Abbaspour, K. & Yang, H. (2007) Hydrological modelling of the chaohe basin in china: Statistical model formulation and bayesian inference. *J. Hydrol.* **340**, 167–182.

APPENDIX: BEHAVIOUR OF THE PERSISTENCE CRITERION OVER TIME STEPS WITH SMALL RUNOFF VARIATIONS

Let us assume that runoff slowly varies from time step t to time step $t+L$: $\Delta Q_t = \left|Q_{t+L} - Q_t\right| << Q_t$. Therefore, the persistence model gives a good forecast for this time step. The forecasting model also provides a good forecast if it has been updated by assimilation of the last observation of runoff Q_t. Let us assume, moreover, that the errors of the persistence model and of the tested model are on the same order of magnitude (which has been verified in practice).

How does the partial persistence index (PPI) evolve if the PPI is already calculated over a large number of time steps t to ensure that ΔQ_t is much lower than the partial RMSE of the persistence model and of the naïve model ($PMSE_{TM}(t-1)$ and $PMSE_{PM}(t-1)$)?

Let x be the ratio between ΔQ_t^2 and $PMSE_{PM}(t-1)$: $x << 1$. The partial persistence index at time step t is:

$$\begin{aligned} PPI_{L,O}(t) &\approx 1 - \frac{PMSE_{TM}(t-1) + \Delta Q_t^2}{PMSE_{PM}(t-1) + \Delta Q_t^2} = 1 - \left(\frac{PMSE_{TM}(t-1)}{PMSE_{PM}(t-1)} + x \right) \cdot \frac{1}{1+x} \\ &\approx 1 - \left(\frac{PMSE_{TM}(t-1)}{PMSE_{PM}(t-1)} + x \right) \cdot \left(1 - x + o(x)\right) = \left(1 - \frac{PMSE_{TM}(t-1)}{PMSE_{PM}(t-1)} \right) \cdot (1-x) + o(x) \\ &\approx PPI_{L,O}(t-1) \cdot (1-x) + o(x) \end{aligned} \quad (485)$$

where $o(x)$ is a Landau notation.

Thus, the absolute value of the partial persistence index will decrease very slowly (at a rate of $1-x$) when the runoff variation is much lower than the runoff and lower than the square root of the partial MSE (i.e. when the criterion is assessed over a large number of time steps). This pattern can be seen in the example illustrated in Fig. 2.b.

G
Annexe

Données et graphes décrivant l'ensemble des bassins versants

Figures

G.1 Plan principal de l'A.C.P. sur les 6 indices décrivant le débit. 424
G.2 Répartition de l'échantillon de 1 024 bassins en 4 groupes homogènes au sens des 6 indices décrivant le débit. 424
G.3 Répartition géographique des quatre groupes homogènes de bassins versants définis à partir de l'échantillon de 1 024 bassins versants par la C.A.H. calculée sur les 6 indices décrivant les régimes hydrographiques. 425
G.4 Répartition de l'échantillon de 1 024 bassins en 8 groupes homogènes au sens des 6 indices décrivant le débit. 426
G.5 Caractéristiques physiques des 1 024 bassins versants. 428
G.6 Histogrammes des 6 indices décrivant le débit. 429
G.7 Histogrammes des 4 indices décrivant les pluies. 430
G.8 Corrélations entre indices : densité locale 2 à 2. 431
G.9 Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325. 434
G.10 Influence de différentes propriétés topographiques sur la répartition des bassins dans le plan principal de l'A.C.P. 438
G.11 Influence du régime pluvieux sur la répartition des bassins dans le plan principal de l'A.C.P. 440
G.12 Comparaison des 6 indices décrivant le régime hydrographique et des critères de performance sur les échantillons restreints obtenus par C.A.H. sur ces 6 indices 442
G.13 Comparaison des 10 indices décrivant les régimes hydrographique et pluvieux et des critères de performance sur les échantillons restreints obtenus par C.A.H. sur ces 10 indices 451
G.14 Distribution du critère de Nash-Sutcliffe en contrôle représentée dans le plan principal des A.C.P. 463
G.15 Distribution du critère de Nash-Sutcliffe en contrôle représentée en fonction de la distance à l'origine dans l'espace des descripteurs 465
G.16 Variété intra-groupe du critère de Nash-Sutcliffe en contrôle sur les échantillons restreints 466

Cette annexe débute par l'analyse de la variété hydrologique de nos 1 024 bassins versants décrite par les six indices caractérisant le débit introduits dans le chapitre 4. La seconde partie regroupe les données et les graphes permettant l'exploration de l'échantillon présenté dans le chapitre 4, qui ne sont pas inclus dans le texte.

G.1 Analyse de la variété hydrologique de l'ensemble des bassins versants décrite par six indices caractérisant le débit

G.1.1 Résultats de l'analyse en composantes principales

Les valeurs propres et vecteurs propres associés obtenus lors de l'A.C.P. sont donnés dans le tableau G.1. La somme des deux plus grandes valeurs propres atteint 81 % : présenter l'échantillon dans le plan défini par les deux vecteurs propres associés permet donc de représenter plus des quatre cinquièmes de la variété de l'échantillon.

Valeurs propres		**2,85**	**2,02**	**0,44**	**0,34**	**0,28**	**0,06**
Fraction cumulée (%)		47	81	88	94	99	100
Vecteurs propres	LAM	0,07	0,67	0,07	-0,30	-0,24	0,62
	IND	-0,52	0,00	0,55	-0,04	0,62	0,22
	ETI	0,43	0,41	0,10	-0,41	0,43	-0,53
	ESM	0,53	-0,09	-0,25	0,38	0,54	0,46
	CRU	-0,18	0,59	0,01	0,74	-0,02	-0,27
	BFI	0,47	-0,15	0,79	0,22	-0,29	0,01

Table G.1 – Valeurs propres et vecteurs propres associés de la matrice de covariance des indices de débit.

Le premier axe principal est fortement corrélé positivement avec BFI et les indices décrivant les étiages (ESM et ETI) et négativement avec IND. Ce premier axe est donc associé avec les phases d'écoulement de faible débit.
Le deuxième axe principal est, quant à lui, corrélé positivement en priorité avec les indices LAM et CRU. Il permet de décrire le débit annuel total et la contribution des périodes de plus fort débit (cf. Fig. G.1).

G.1.2 Résultats de la classification ascendante hiérarchique

Nous sommes descendus de 2 puis 3 niveaux dans le dendrogramme obtenu lors de la C.A.H. sur les 6 indices décrivant le débit (Fig. I.1, page 486) pour définir 4 puis 8 groupes homogènes. Ce dendrogramme est bien « équilibré » dans le sens où les groupes ainsi définis comptent des effectifs du même ordre de grandeur.

Le premier groupe est le plus fourni (433 bassins). Il se situe le plus au centre du plan

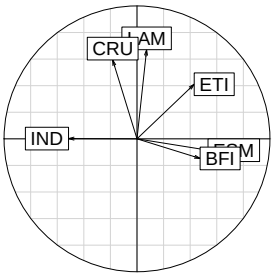


FIGURE G.1 – Plan principal de l'A.C.P. sur les 6 indices décrivant le débit.

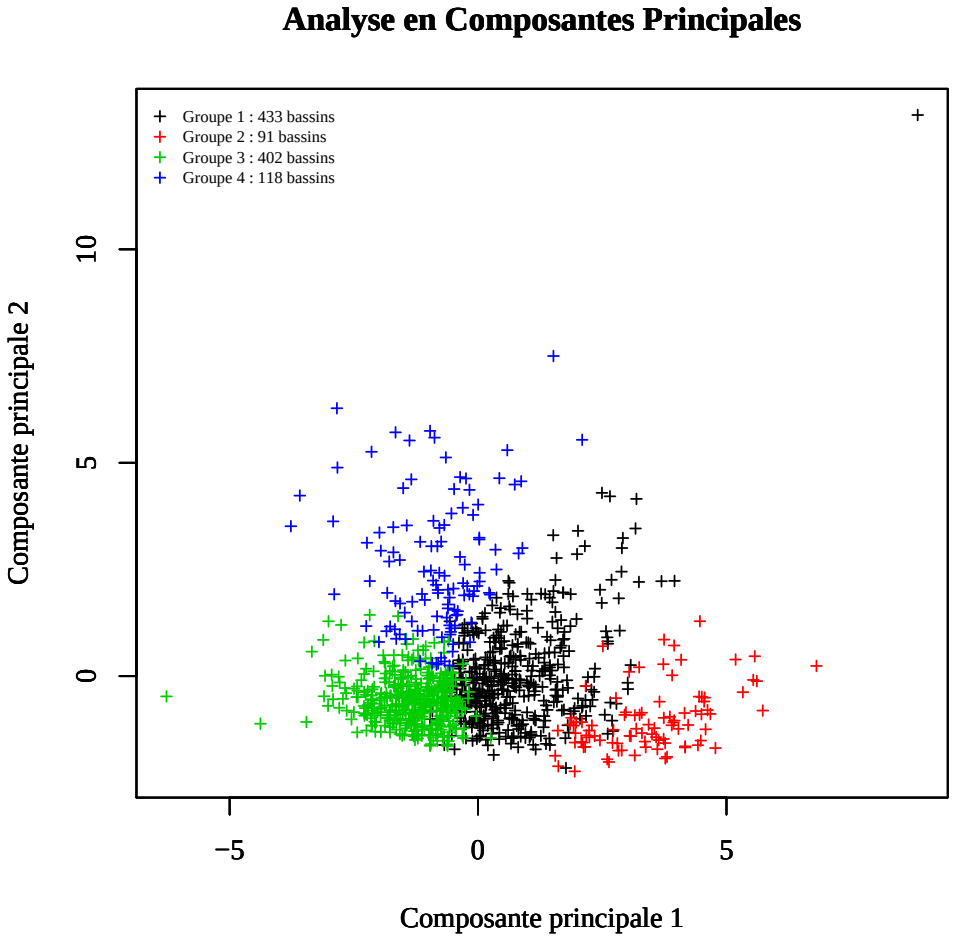


FIGURE G.2 – Répartition de l'échantillon de 1 024 bassins en 4 groupes homogènes au sens des 6 indices décrivant le débit.

principal (cf. Fig. G.2). Il est donc constitué des bassins présentant les écoulements les plus « moyens » tant en terme de quantité annuelle que pour les extrêmes (crues et surtout étiages). Le second groupe (91 bassins) est situé dans une zone définie par une abscisse significativement positive et une ordonnée négative ou faiblement positive ; il s'agit des bassins présentant les étiages les plus forts et parmi les plus faibles débits annuels. Le troisième groupe (402 bassins très regroupés dans ce plan) se caractérise par des abscisses et des ordonnées faibles : ce sont donc des bassins de débits les plus stables (sans étiage ni crue très forts). Enfin le quatrième groupe (118 bassins) est placé dans une zone d'abscisses négatives mais d'ordonnées positives : ces bassins présentent un fort débit annuel ou des crues importantes mais pas d'étiages sévères.

La répartition géographique des 4 groupes de bassins versants est donnée en figure G.3. Nous observons que plusieurs groupes sont géographiquement cohérents : les bassins indiqués en rouge sont quasiment tous situés dans le bassin de la Seine ou dans le Nord ; les bassins montagneux appartiennent majoritairement au groupe coloré en marron.

La partition en 8 groupes homogènes permet une description plus fine (Fig. G.4).

G.1.3 Facteurs liés à la variété

Nous nous intéressons ici à l'influence que pourraient avoir divers facteurs (localisation, altitude, surface, pente, forêt...) sur le régime d'écoulement. Le but est de déterminer si certaines caractéristiques permettent de classer au moins partiellement les bassins versants.

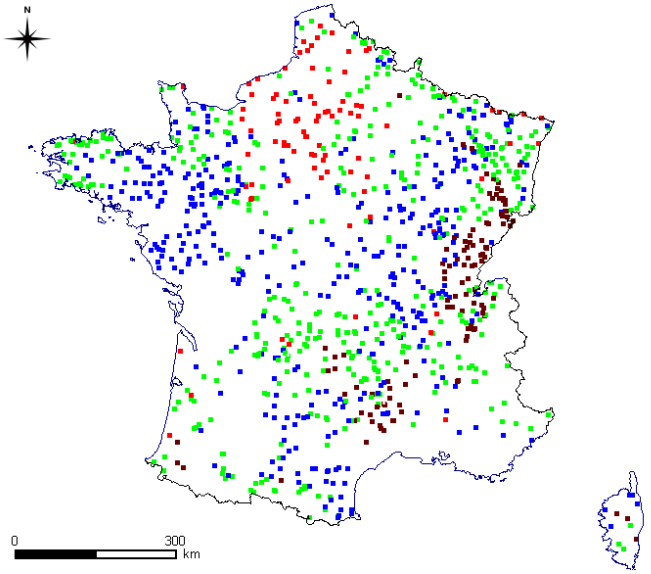


Figure G.3 – Répartition géographique des quatre groupes homogènes de bassins versants définis à partir de l'échantillon de 1 024 bassins versants par la C.A.H. calculée sur les 6 indices décrivant les régimes hydrographiques.

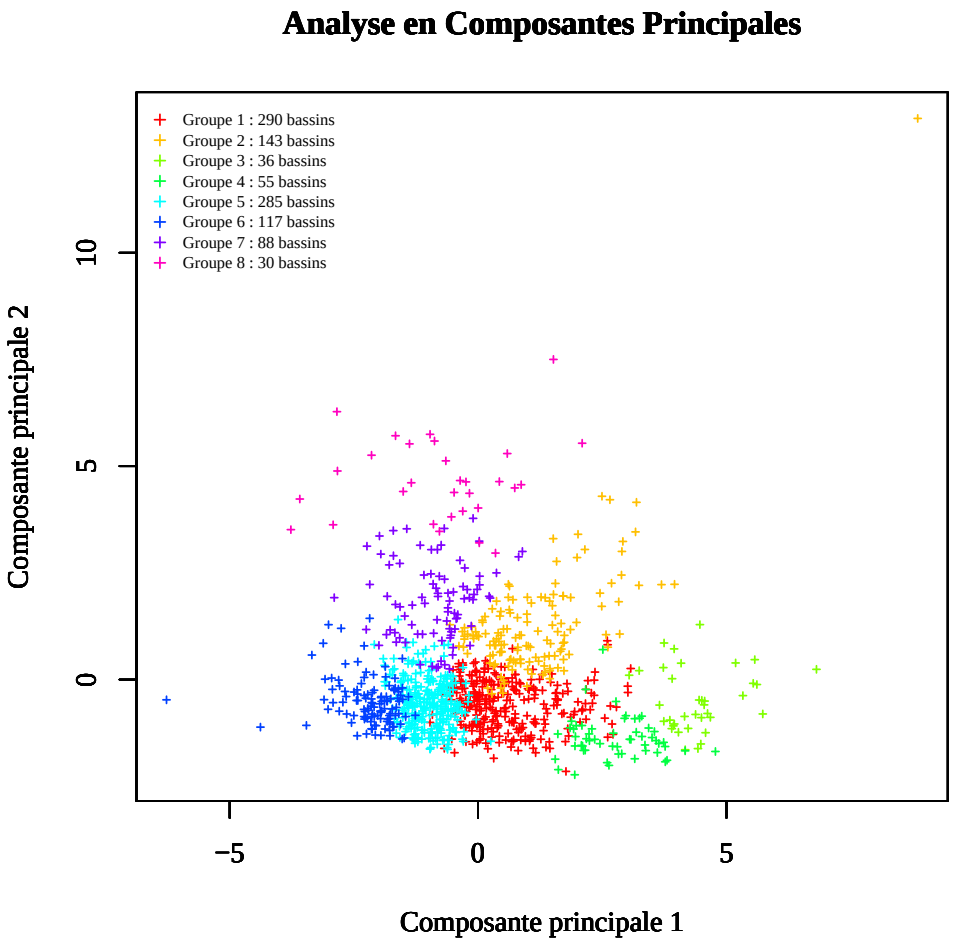


FIGURE G.4 – Répartition de l'échantillon de 1 024 bassins en 8 groupes homogènes au sens des 6 indices décrivant le débit.

Commençons par la localisation (limitée aux régions hydrographiques représentées dans le glossaire, page 324). La figure G.9, à partir de la page 434, donne les répartitions des bassins versants dans le plan principal de l'A.C.P. en fonction de leurs régions hydrographiques. Les bassins de certaines régions hydrographiques (comme par exemple l'Artois-Picardie ou la Normandie) sont fortement regroupés, ce qui suggère que ces bassins se ressemblent fortement. D'autres régions hydrographiques sont au contraire présentes dans tout le nuage de points : elles renferment des bassins versants bien différents (comme les régions de la Saône ou de l'Adour). Ainsi, la région hydrographique peut dans certains cas donner une information hydrologique mais ceci est loin d'être généralisable à toutes les régions hydrographiques françaises.

Nous avons également représenté les distributions des bassins en fonction de quelques caractéristiques physiques dans le plan principal de l'A.C.P. (Fig. G.10, à partir de la page 438). Nous observons une absence de corrélation forte avec la superficie du bassin versant (néanmoins, on peut noter que plus la surface est grande, plus le bassin aura tendance à être proche de l'origine). Il est difficile de définir des influences nettes pour les autres caractéristiques investiguées.

Enfin, il est intéressant d'observer l'influence du régime pluvieux (décrits par les indices PAM, IRR, PMX et SEC). La figure G.11 (à partir de la page 440) montre une certaine corrélation entre le régime des pluies et celle des débits, comme on pouvait s'y attendre. Cependant, la corrélation est trop floue (notamment concernant l'irrégularité de la pluie très peu corrélée avec le débit) : comme on pouvait s'y attendre, la description des régimes hydrologiques ne peut pas être réalisée en se basant uniquement sur une description du débit.

G.2 Données et graphes annexes au chapitre 4

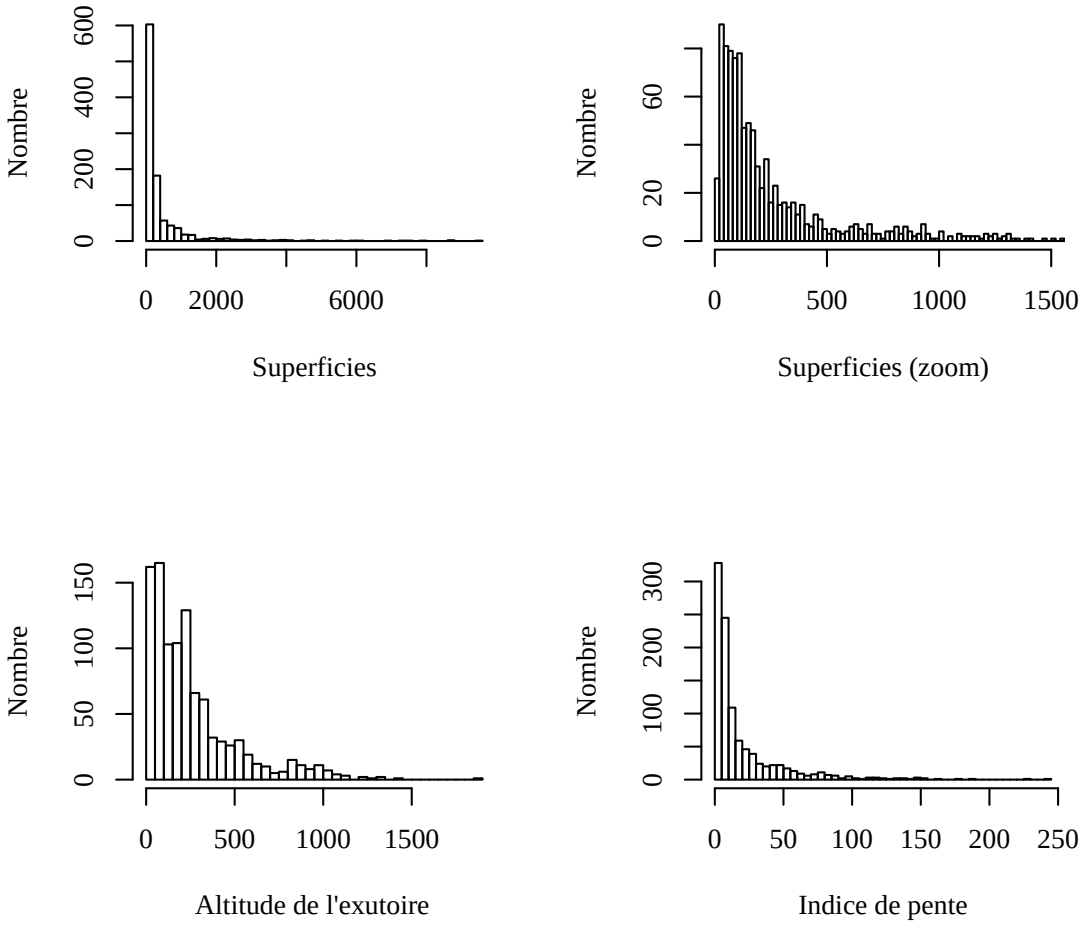


FIGURE G.5 – Caractéristiques physiques des 1 024 bassins versants.

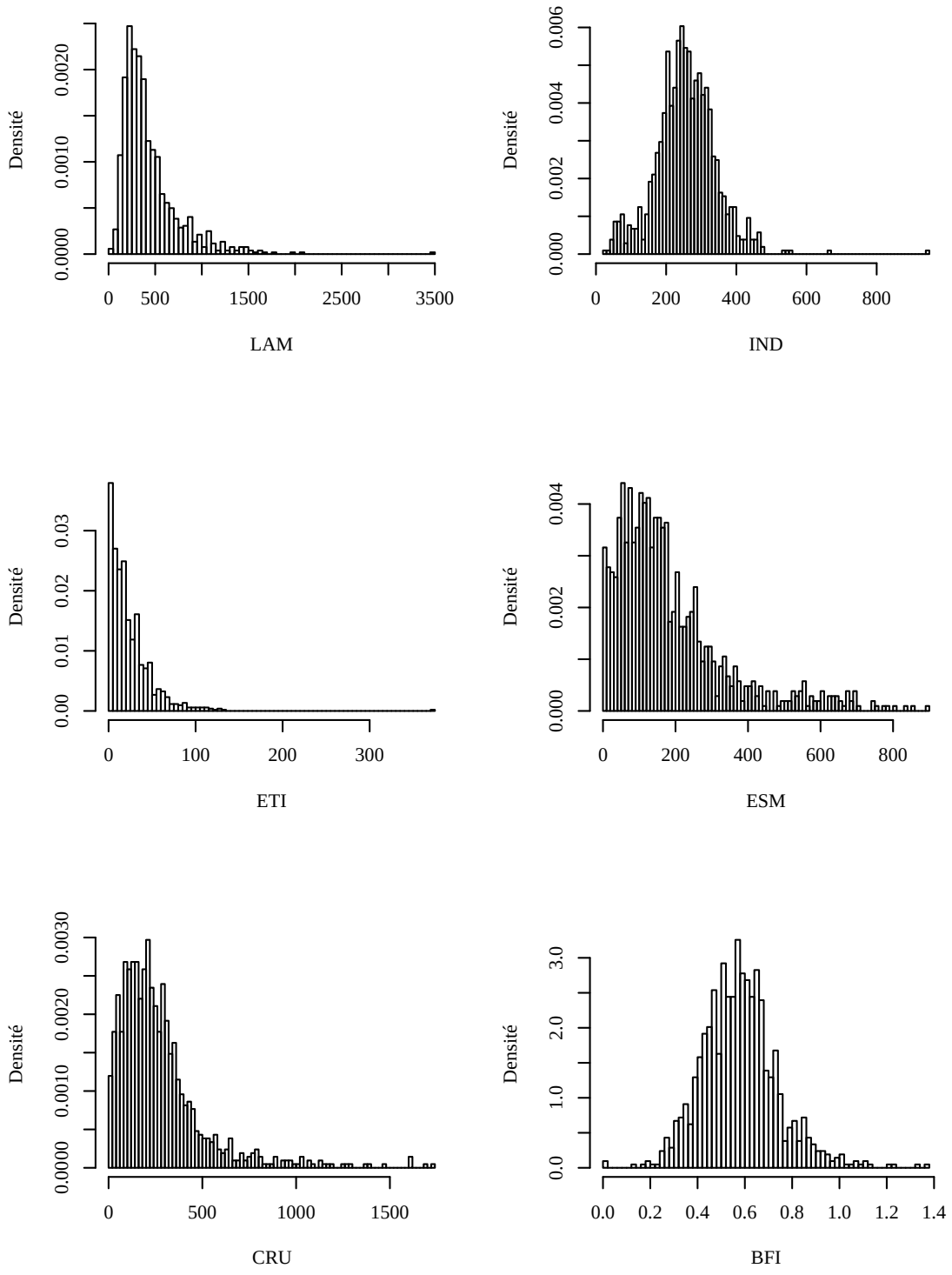


Figure G.6 – Histogrammes des 6 indices décrivant le débit.

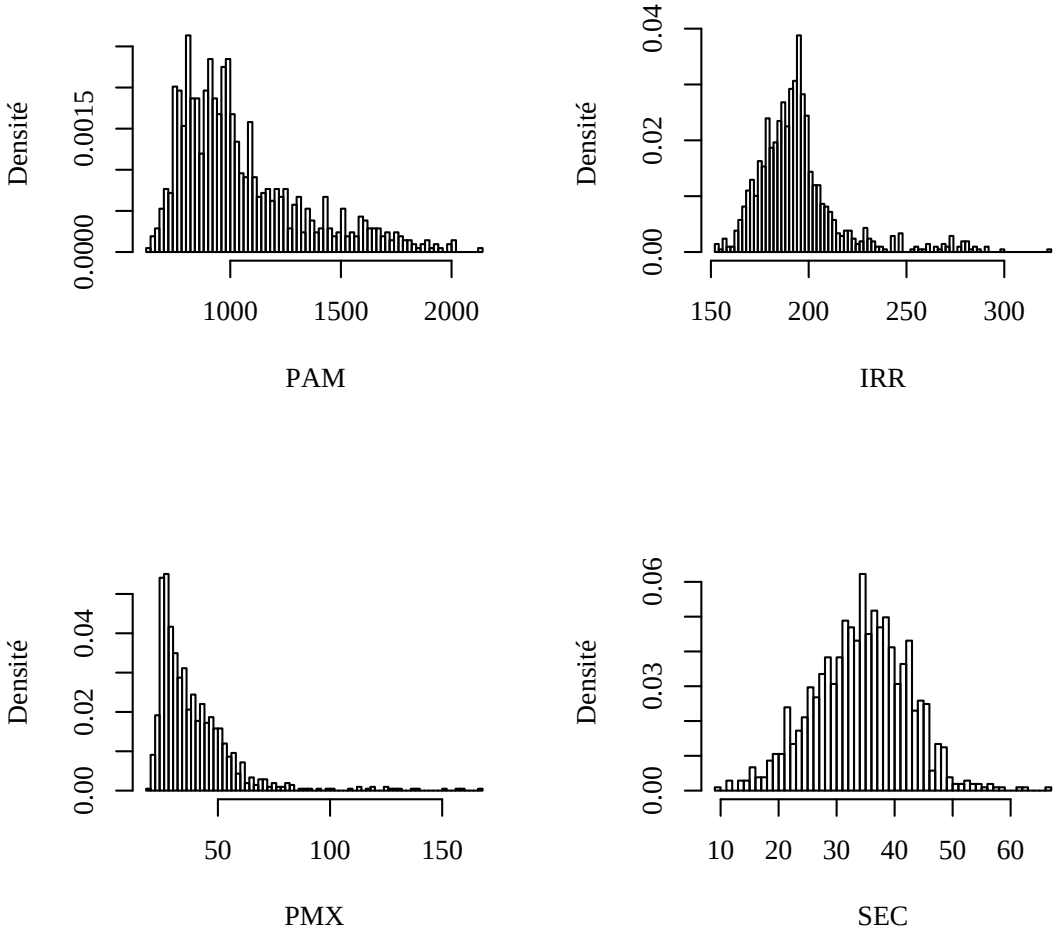


FIGURE G.7 – Histogrammes des 4 indices décrivant les pluies.

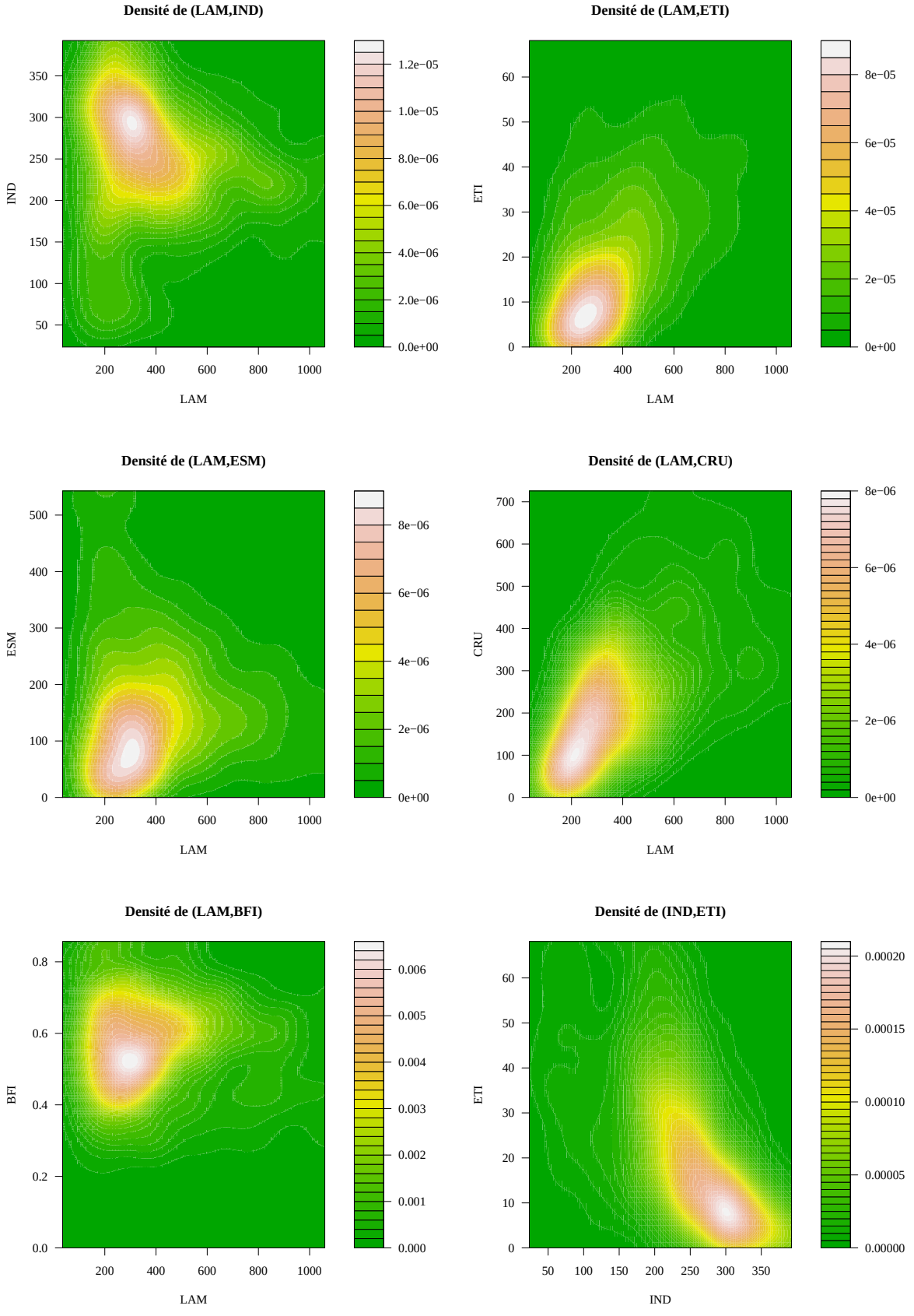


FIGURE G.8 – Corrélations entre indices : densité locale 2 à 2.

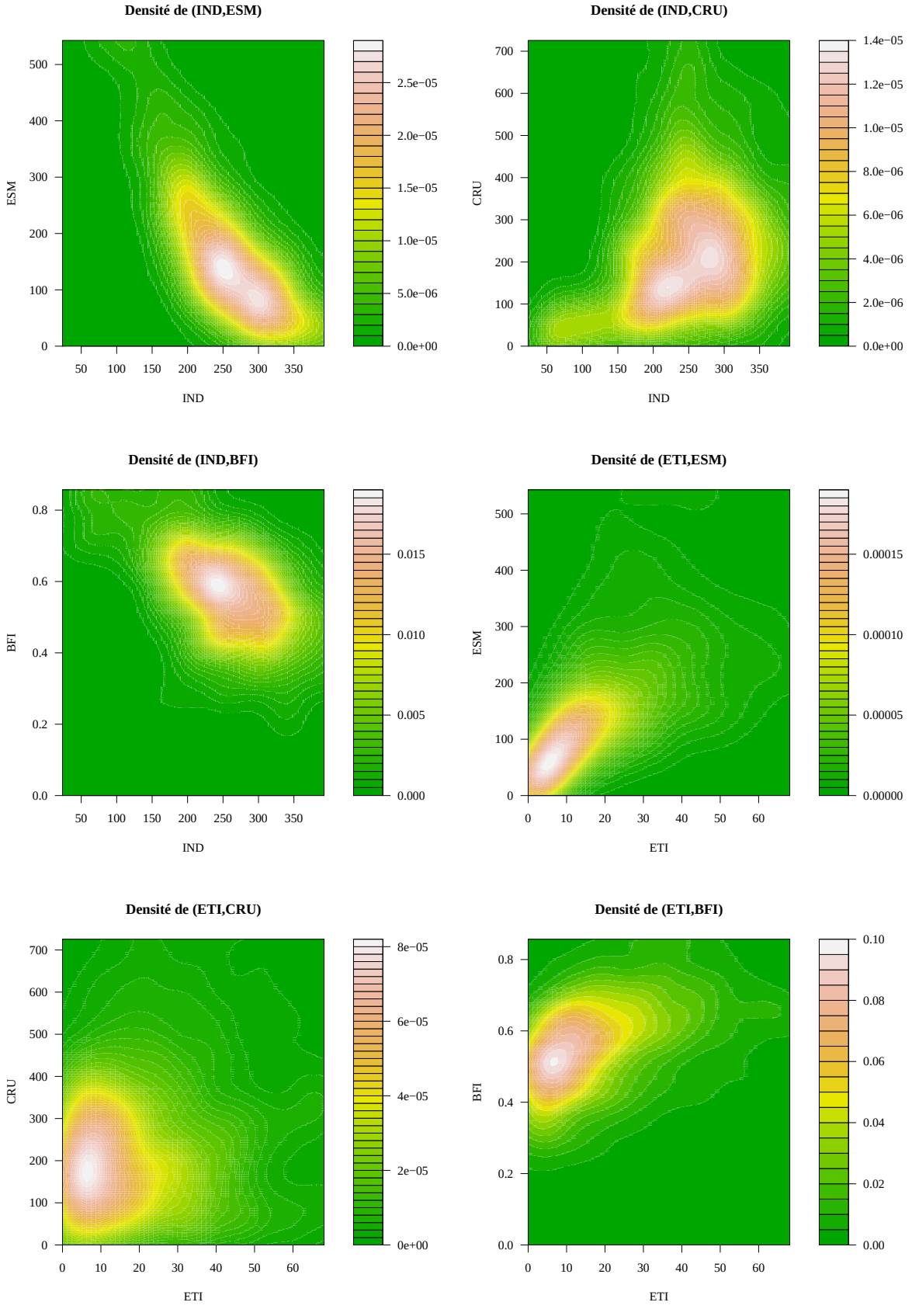


FIG. G.8: Suite – Corrélation entre indices : densité locale 2 à 2.

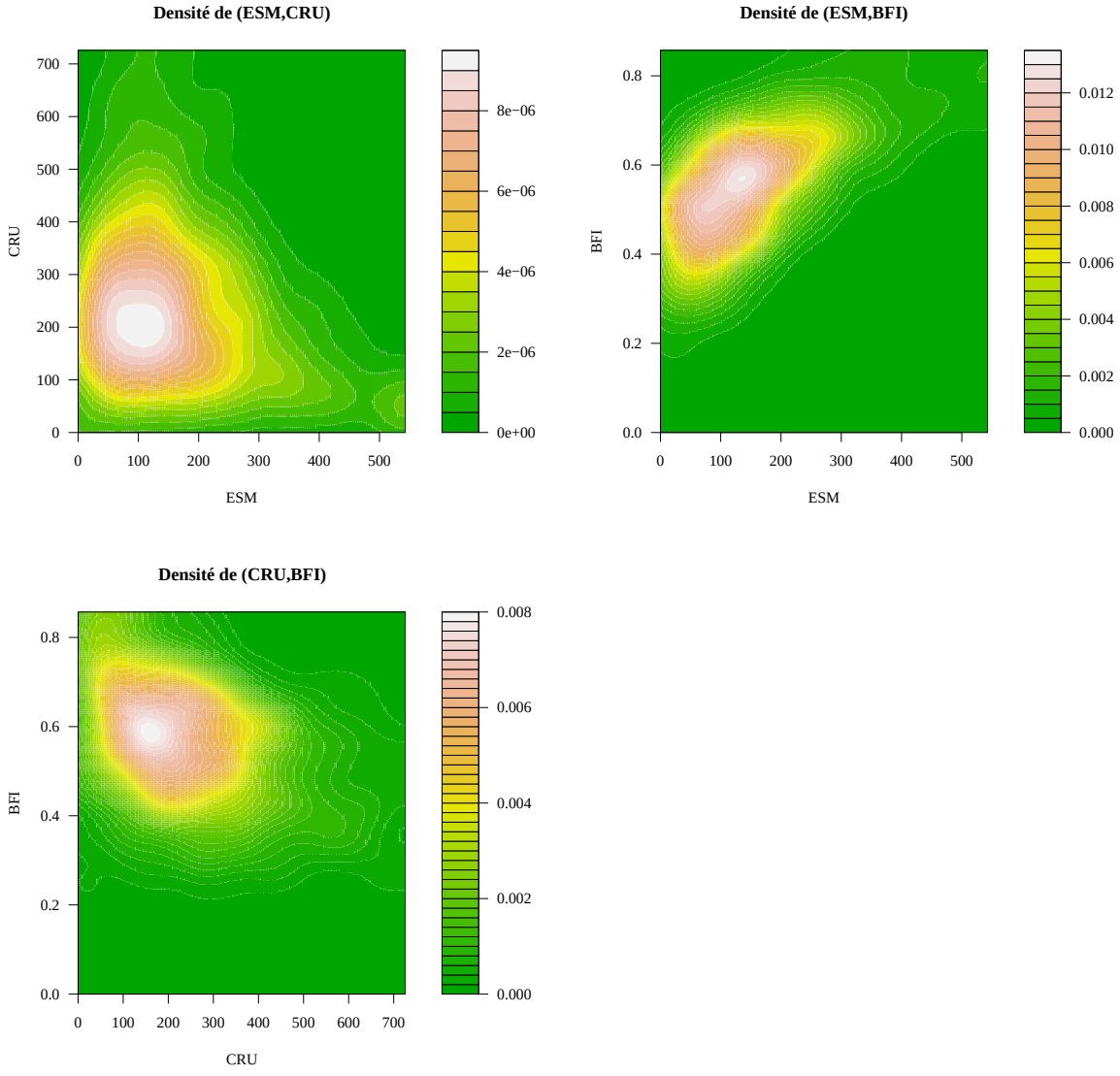


Fig. G.8: Suite et fin – Corrélation entre indices : densité locale 2 à 2.

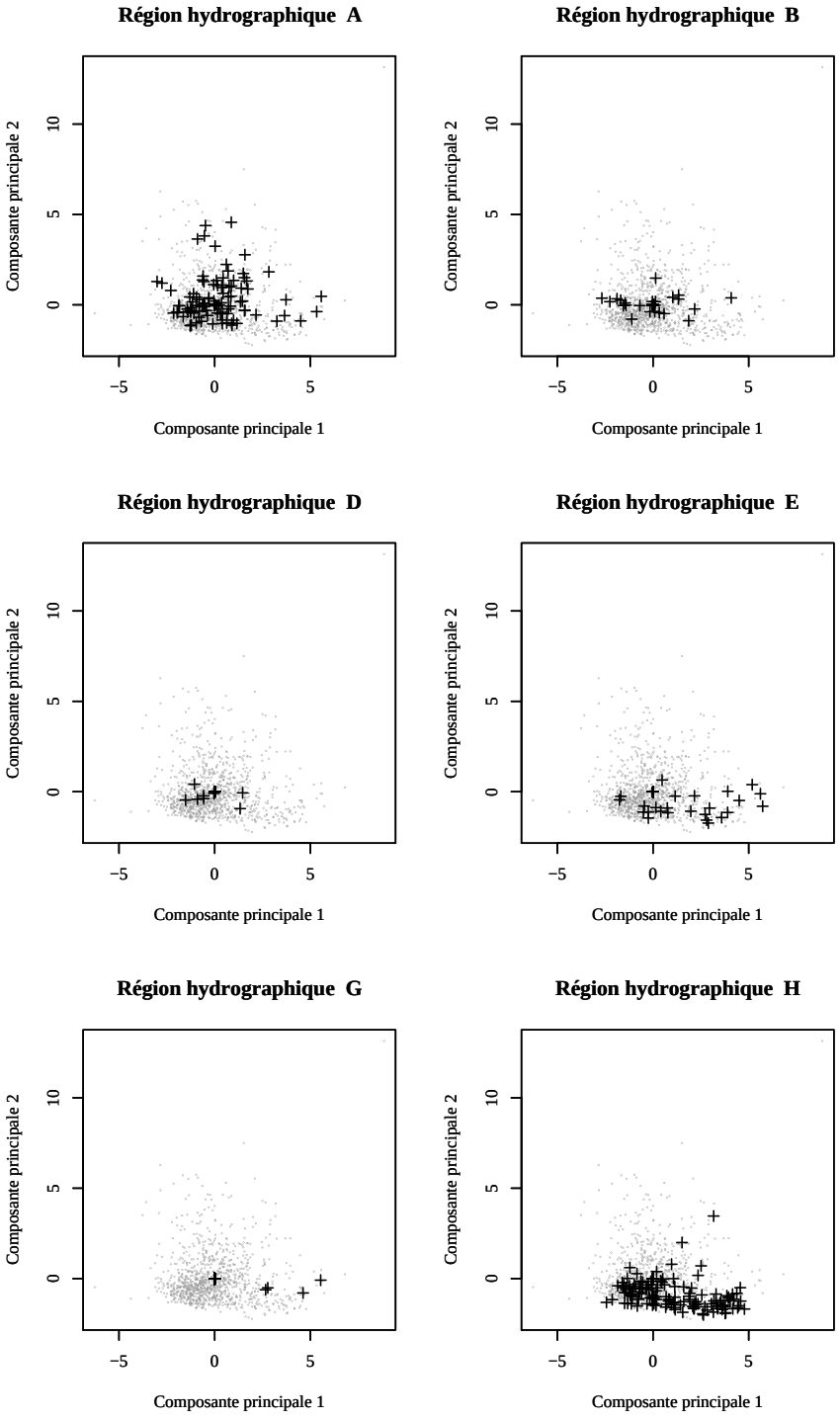


FIGURE G.9 – Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325.

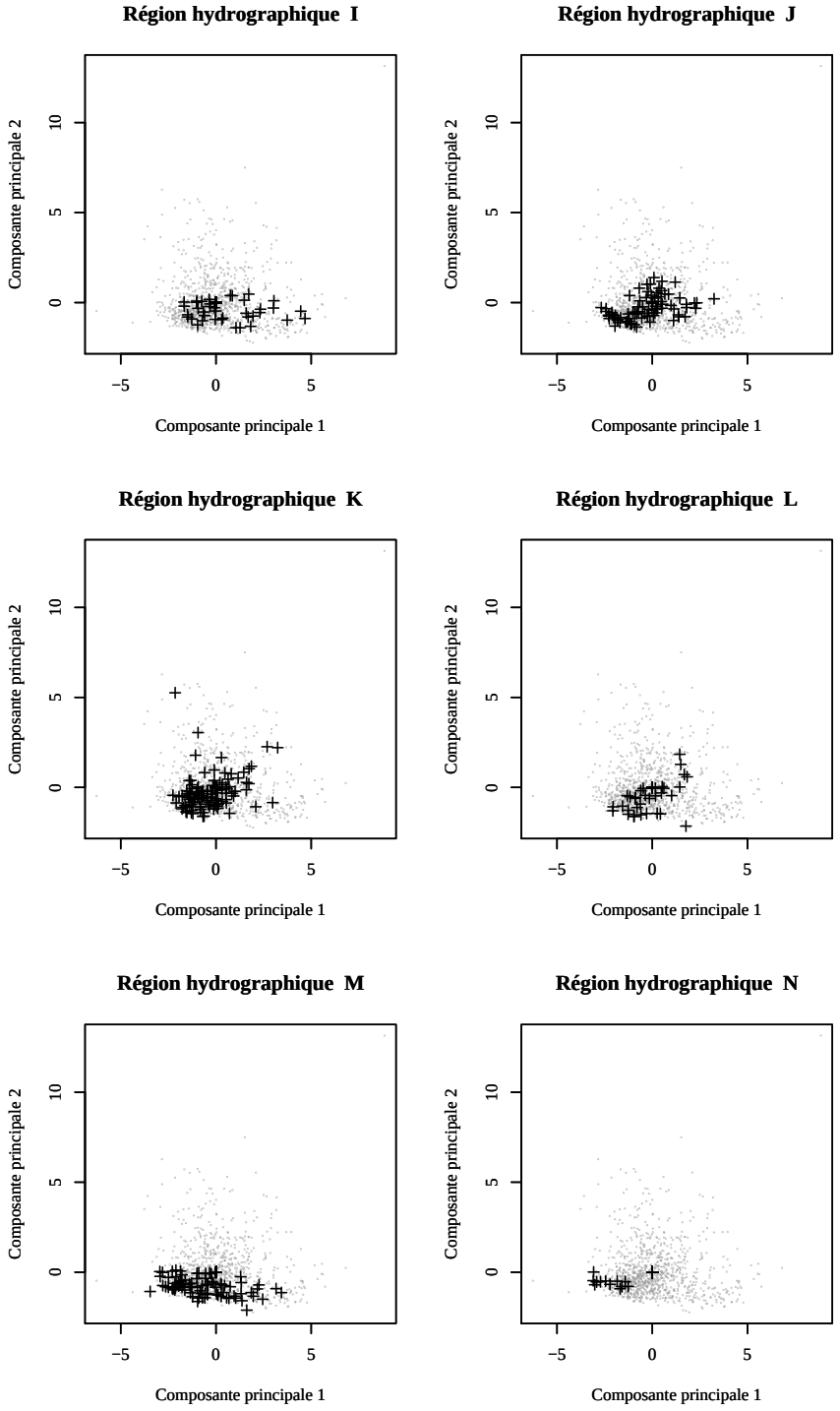


Fig. G.9: Suite – Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325.

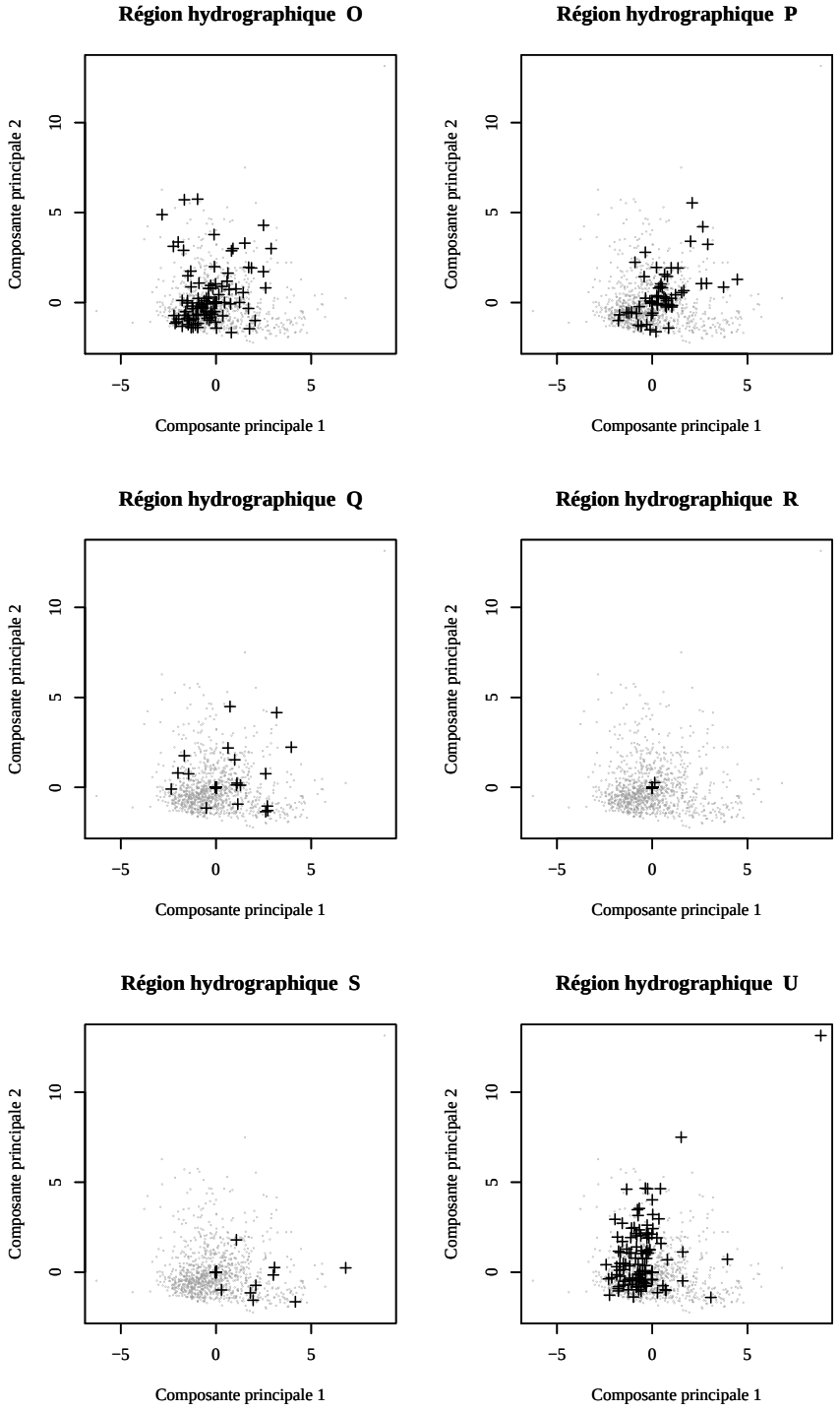


FIG. G.9: Suite – Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325.

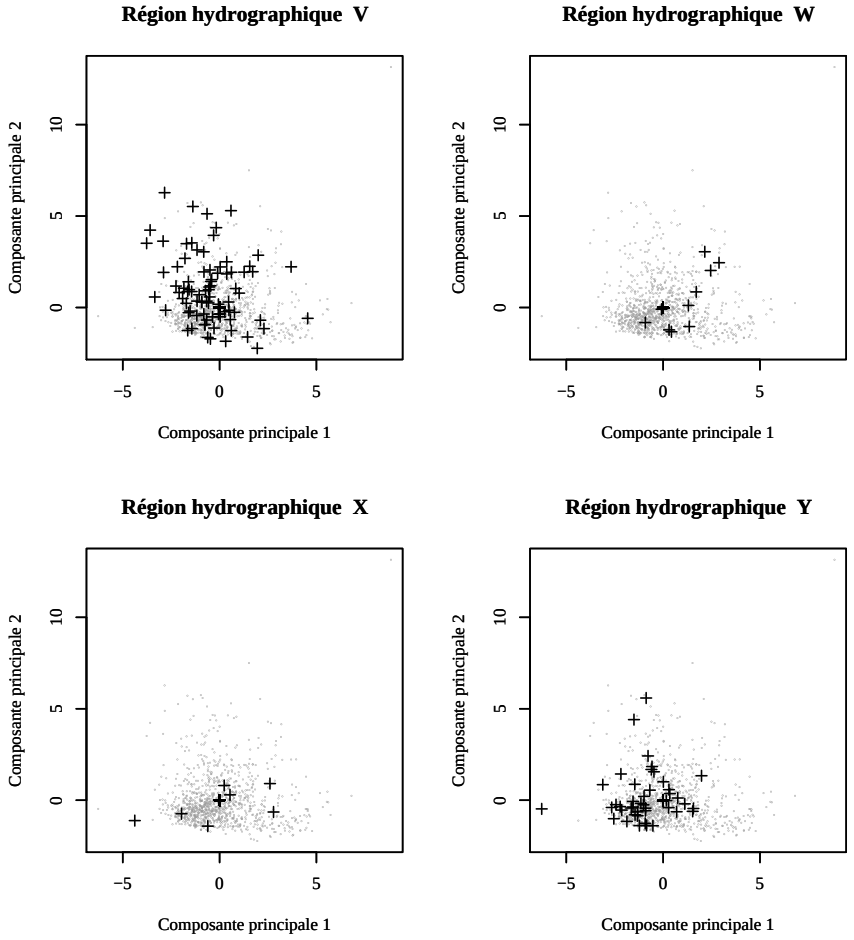


Fig. G.9: Suite et fin – Répartition des bassins versants selon leurs régions hydrographiques. Une carte des régions hydrographiques est présentée dans le glossaire, page 325.

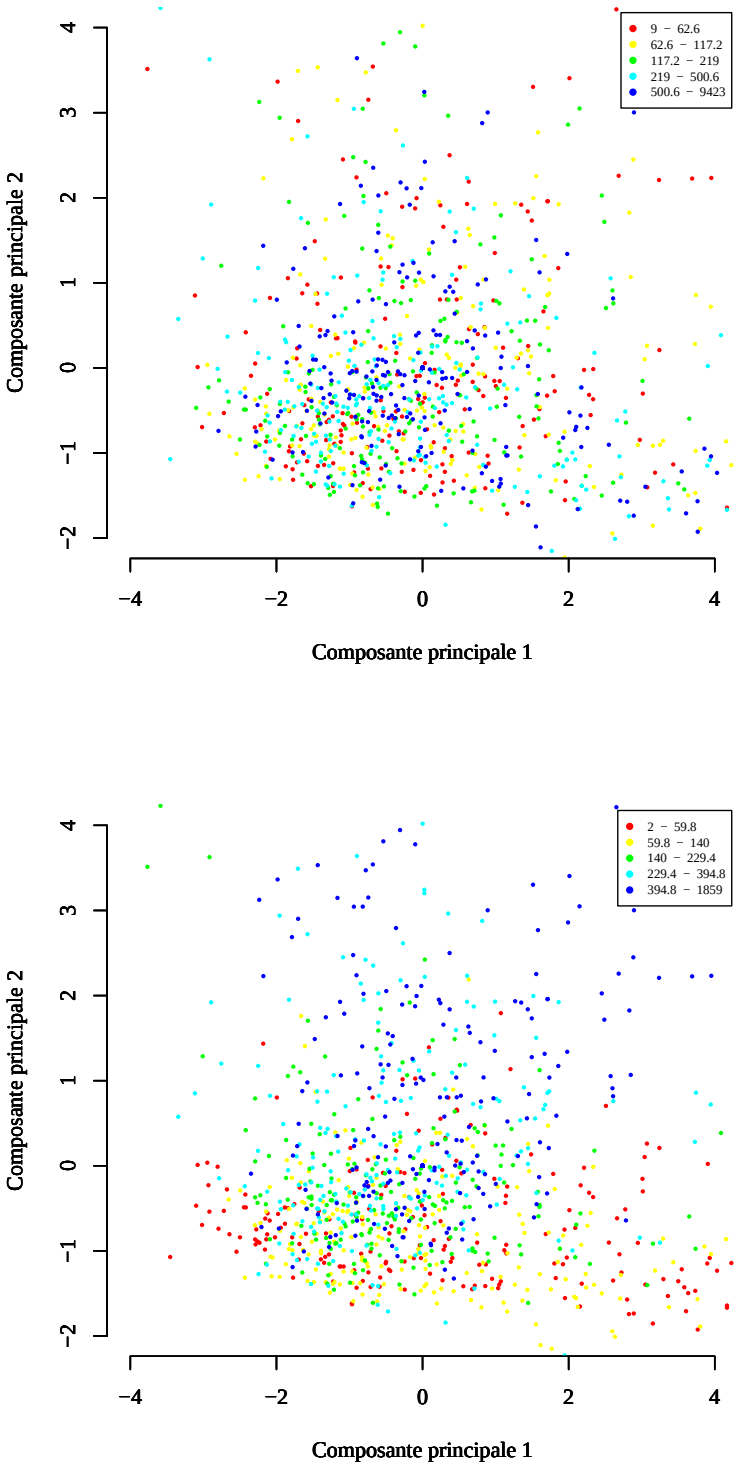


FIGURE G.10 – Influence de différentes propriétés topographiques sur la répartition des bassins dans le plan principal de l'A.C.P. : la surface (en haut) et l'altitude à l'exutoire (en bas).

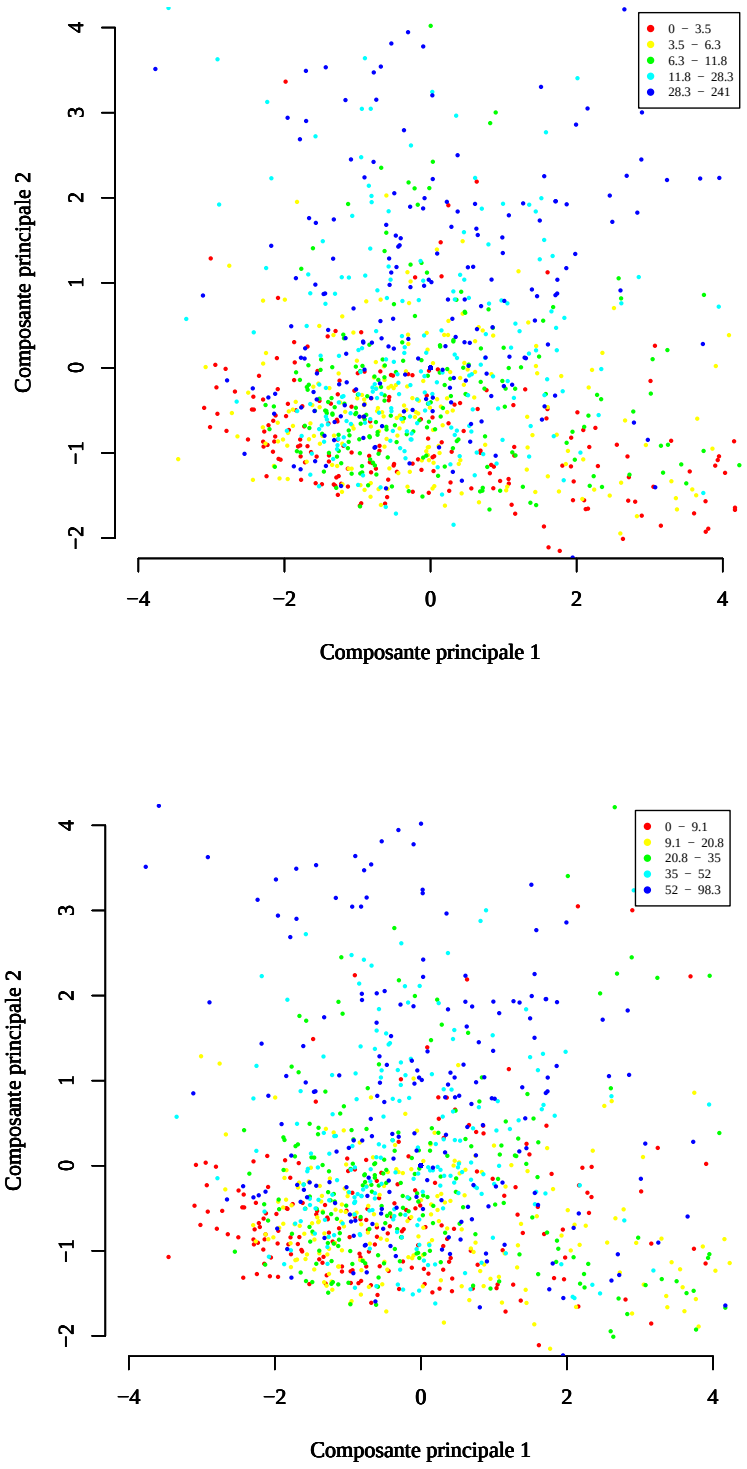


FIG. G.10: Suite – Influence de différentes propriétés topographiques sur la répartition des bassins dans le plan principal de l'A.C.P. : la pente (à travers l'indice IPB, en haut) et la forêt (fraction du bassin recouverte, en bas).

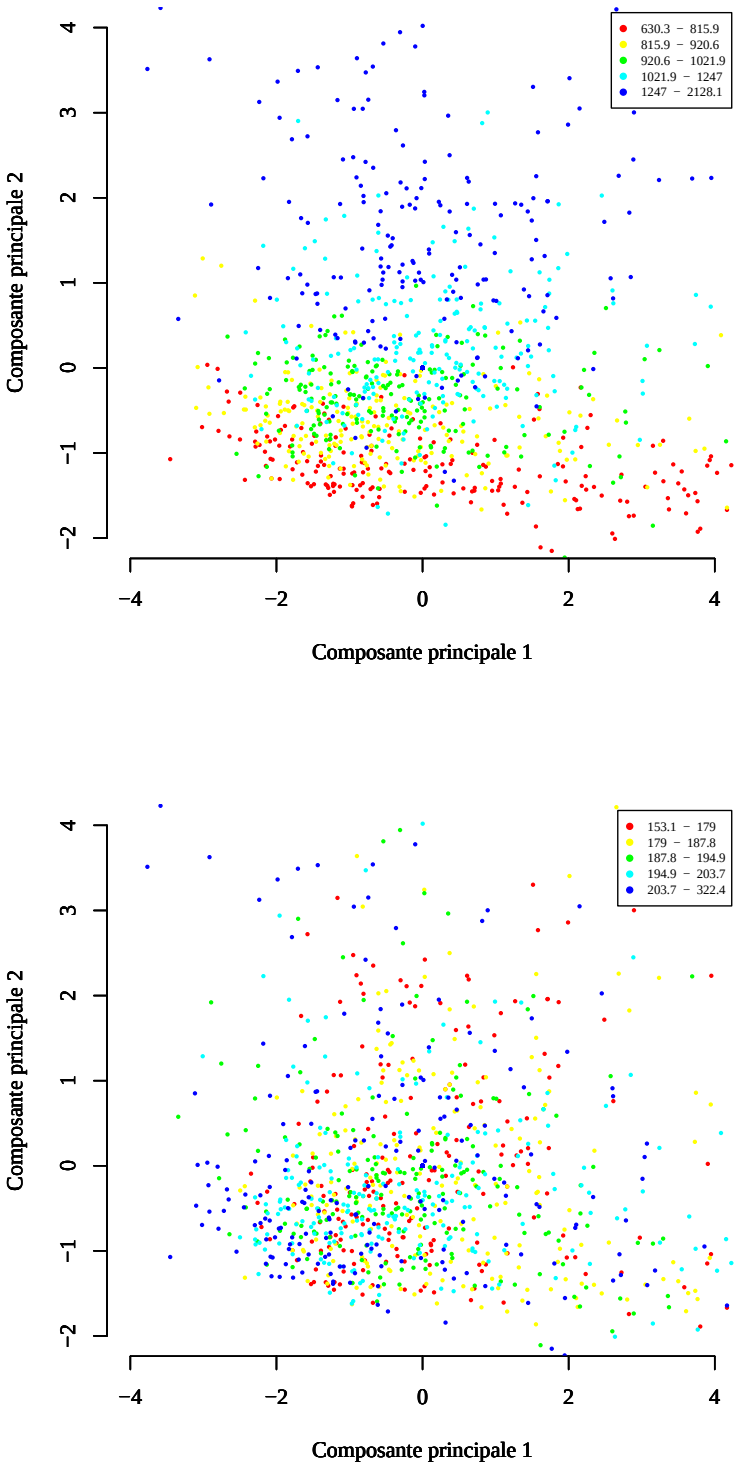


FIGURE G.11 – Influence du régime pluvieux sur la répartition des bassins dans le plan principal de l'A.C.P. : la pluie annuelle moyenne (en haut) et l'irrégularité mensuelle des pluies IRR (en bas).

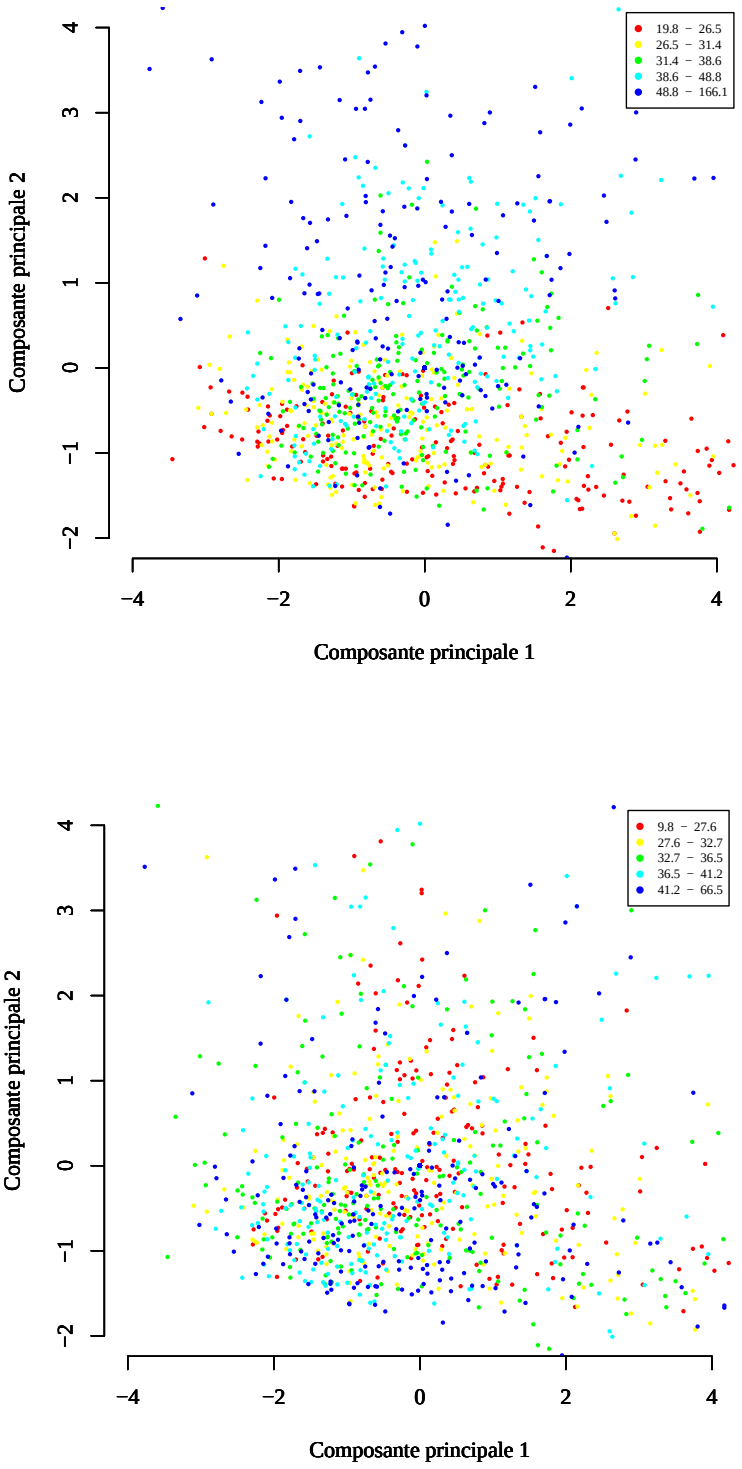


FIG. G.11: Suite – Influence du régime pluvieux sur la répartition des bassins dans le plan principal de l'A.C.P. : indice des pluies intenses PMX (en haut) et fraction de jours secs SEC (en bas).

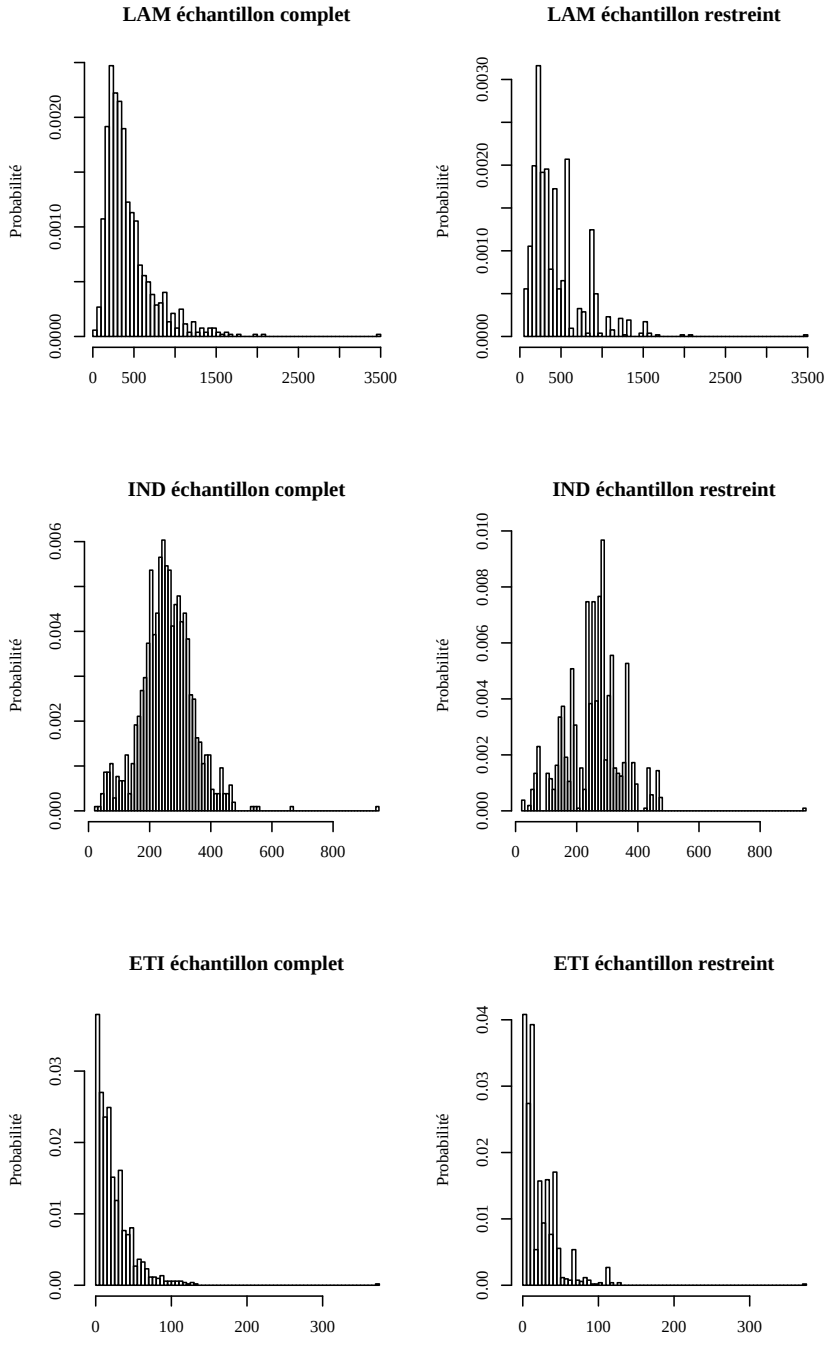


FIGURE G.12 – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 107 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit

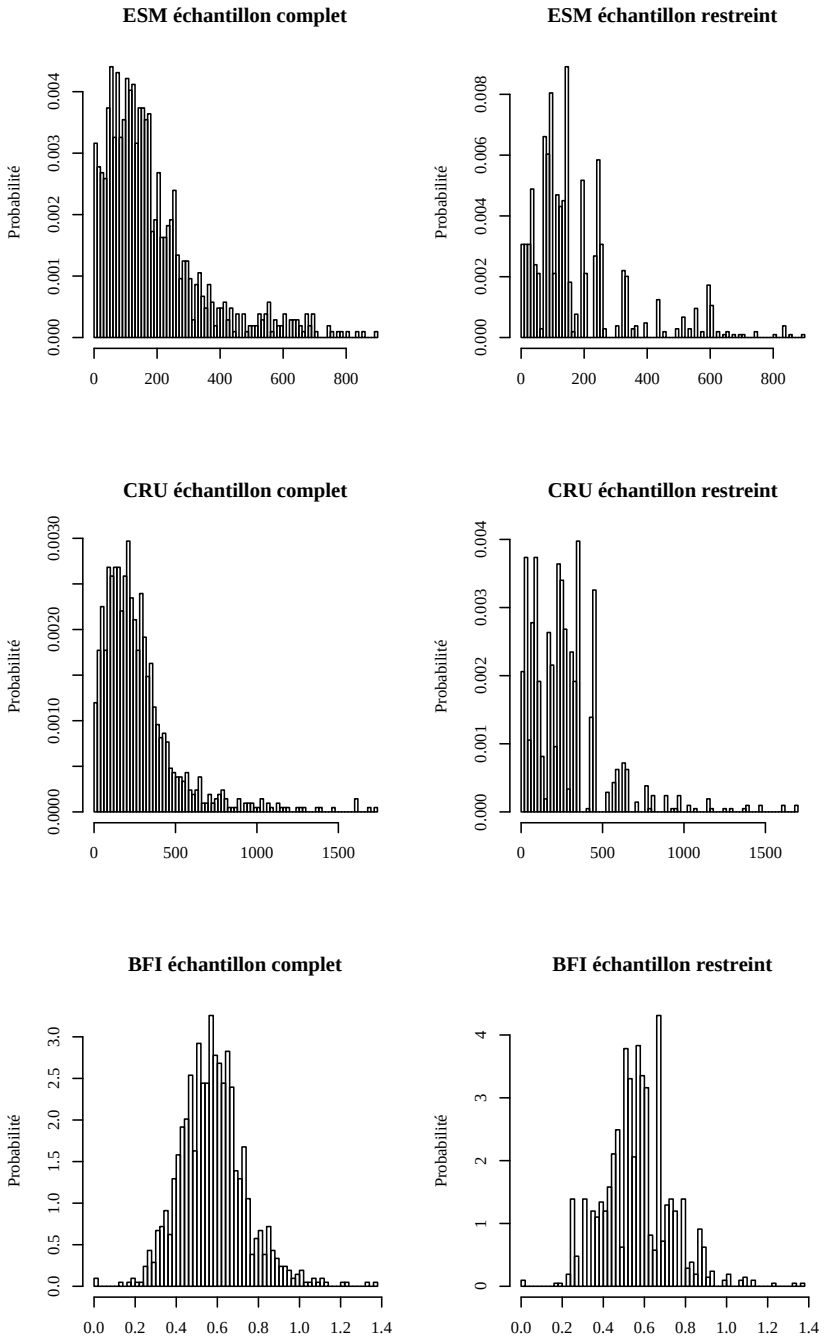


Fig. G.12: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 107 bassins obtenu par C.A.H. sur ces 6 indices caractérisant le débit.

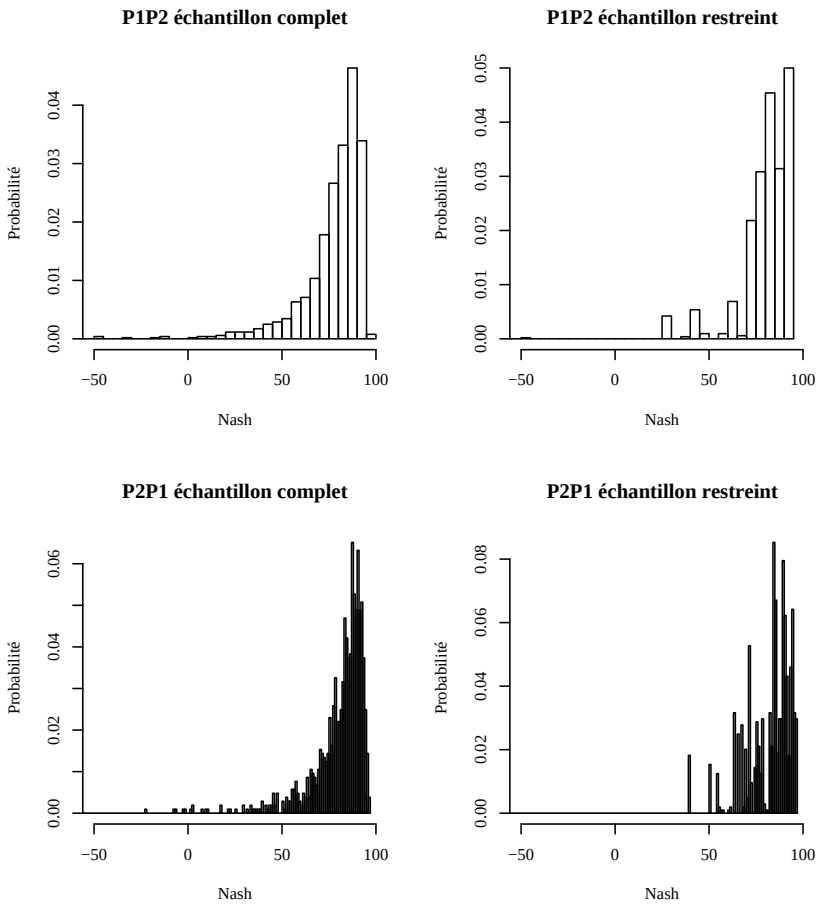


FIG. G.12: Suite – Comparaison des distributions des critères de Nash-Sutcliffe ($\times$ 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 107 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

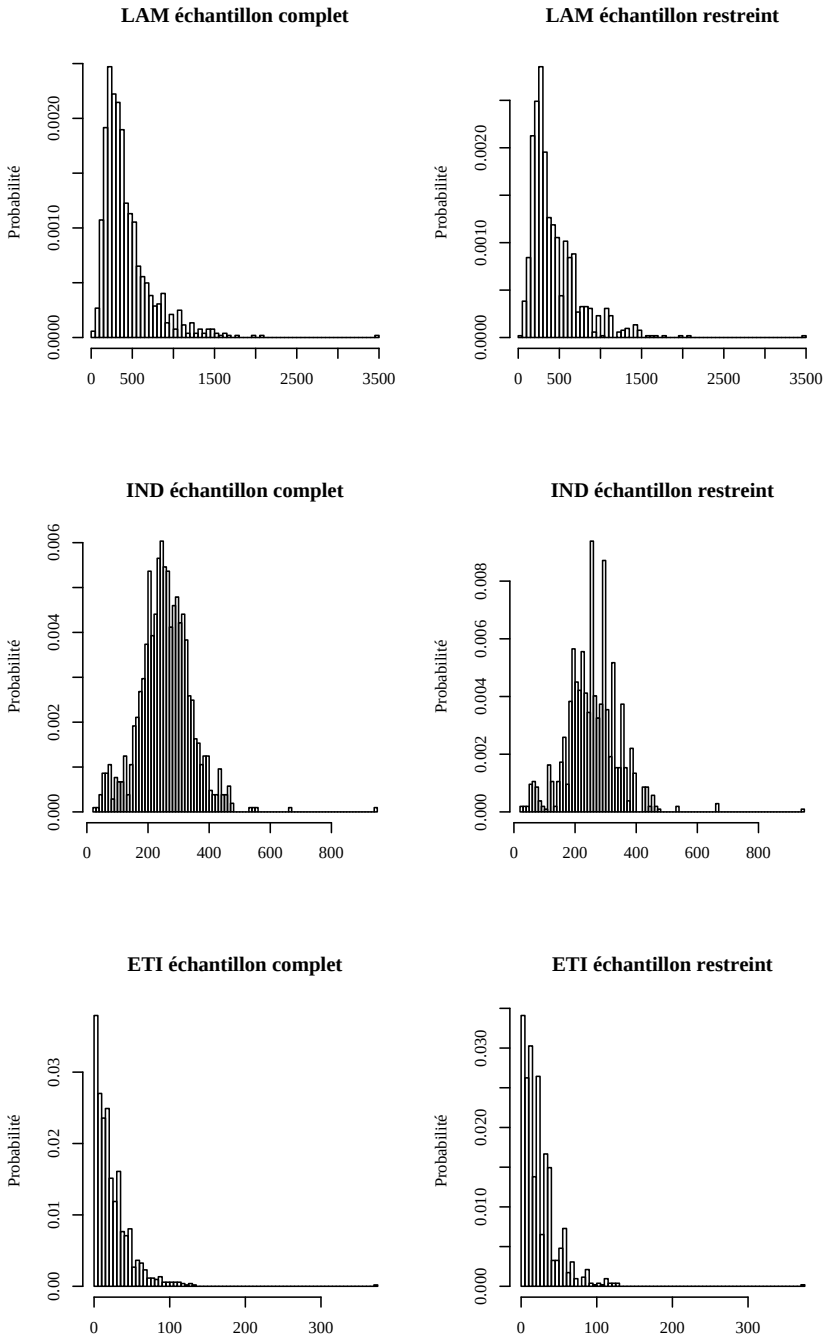


FIG. G.12: Suite – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 190 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

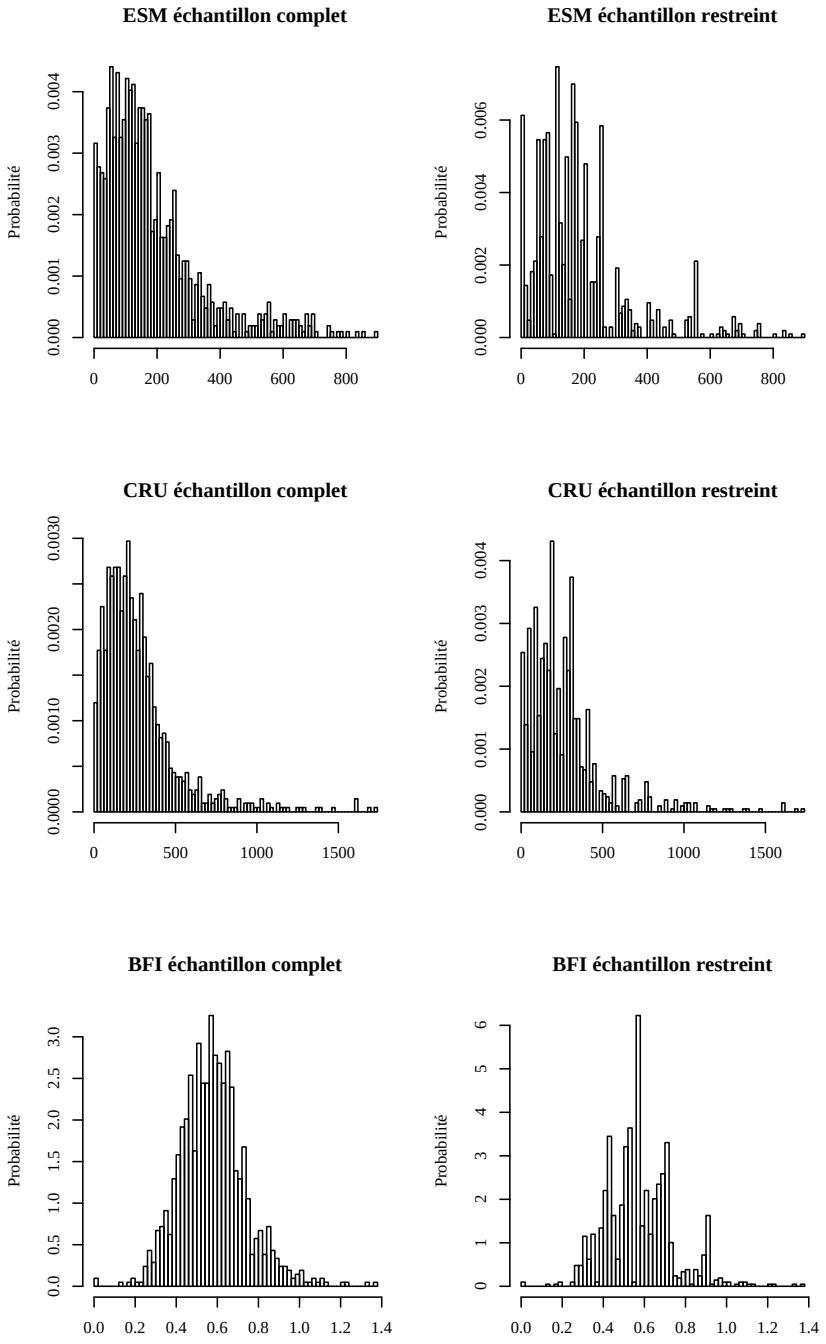


FIG. G.12: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 190 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

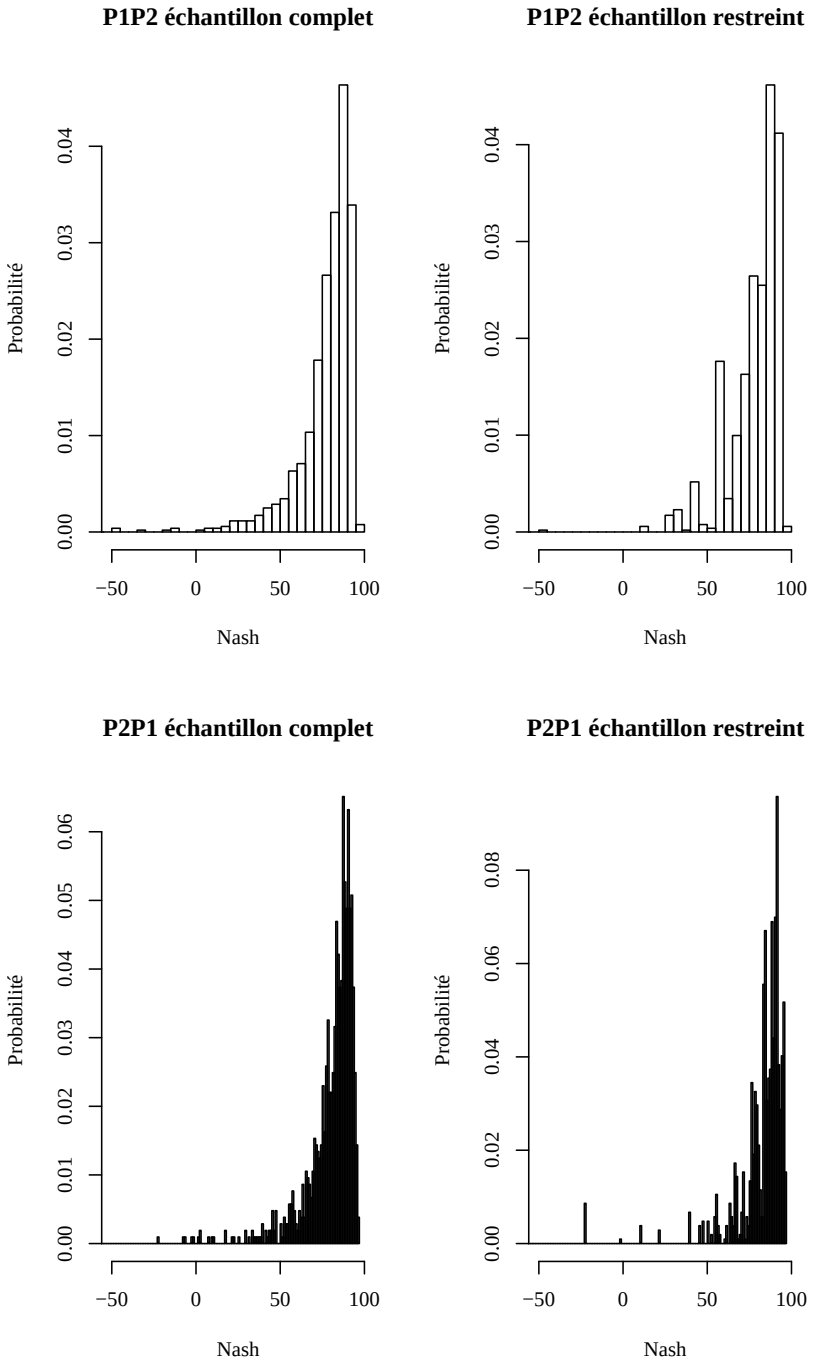


FIG. G.12: Suite – Comparaison des distributions des critères de Nash-Sutcliffe (× 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 190 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

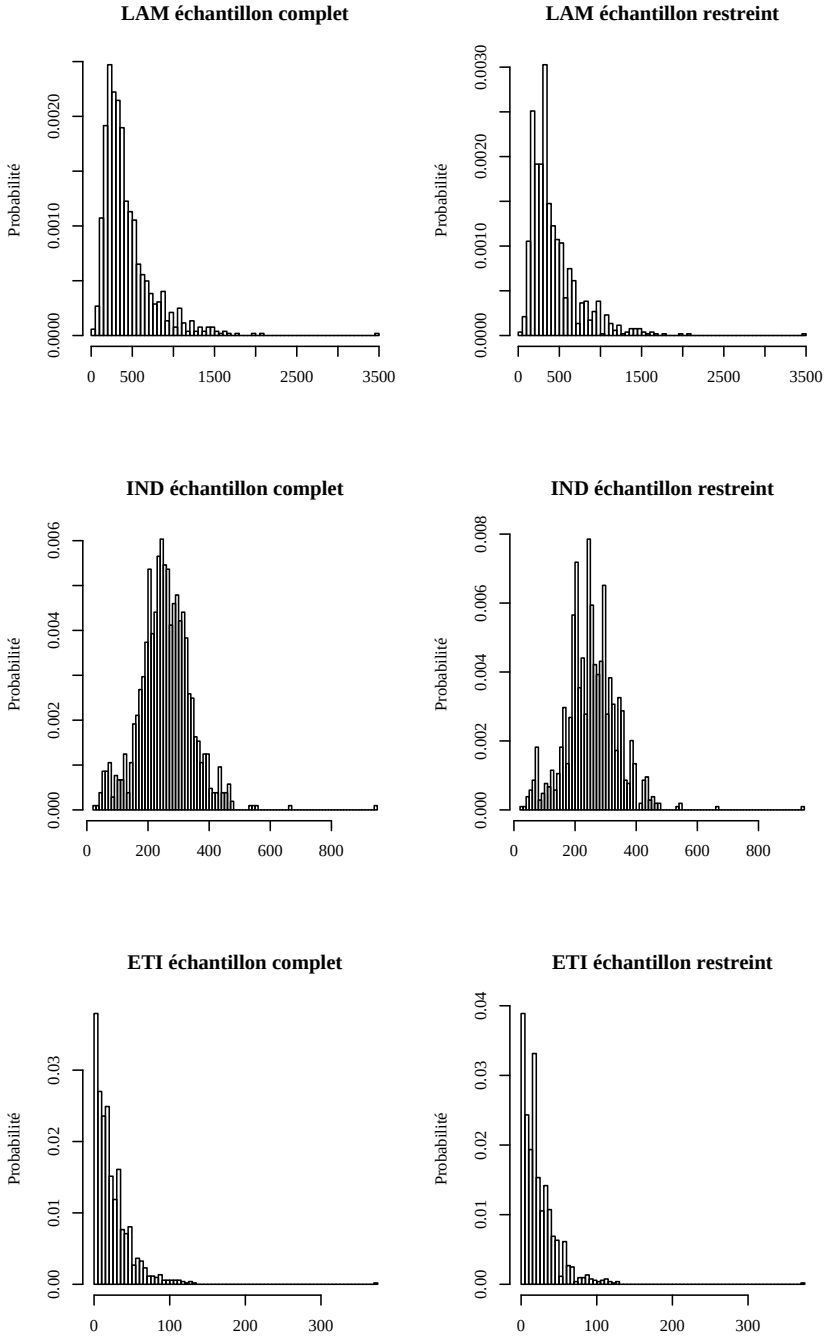


FIG. G.12: Suite – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 322 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

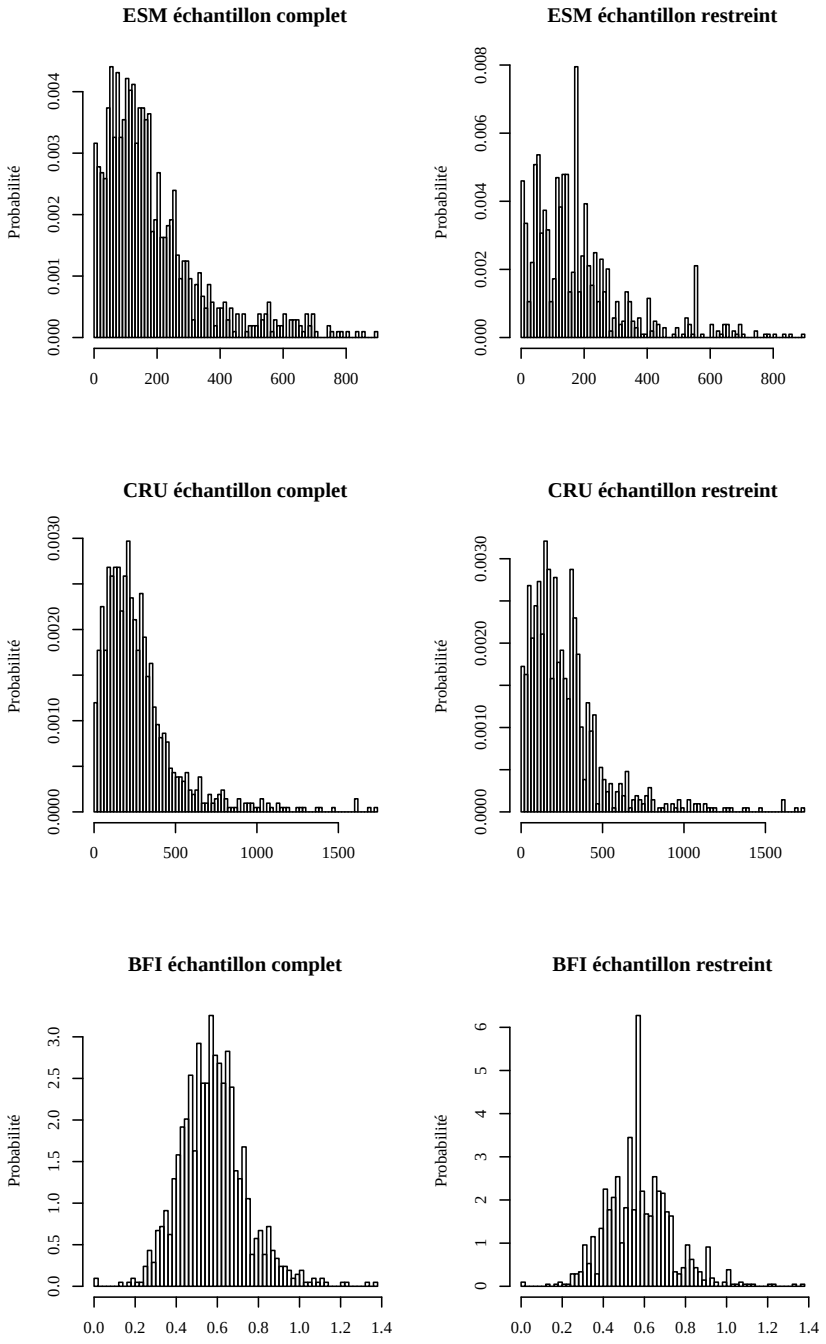


Fig. G.12: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 322 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

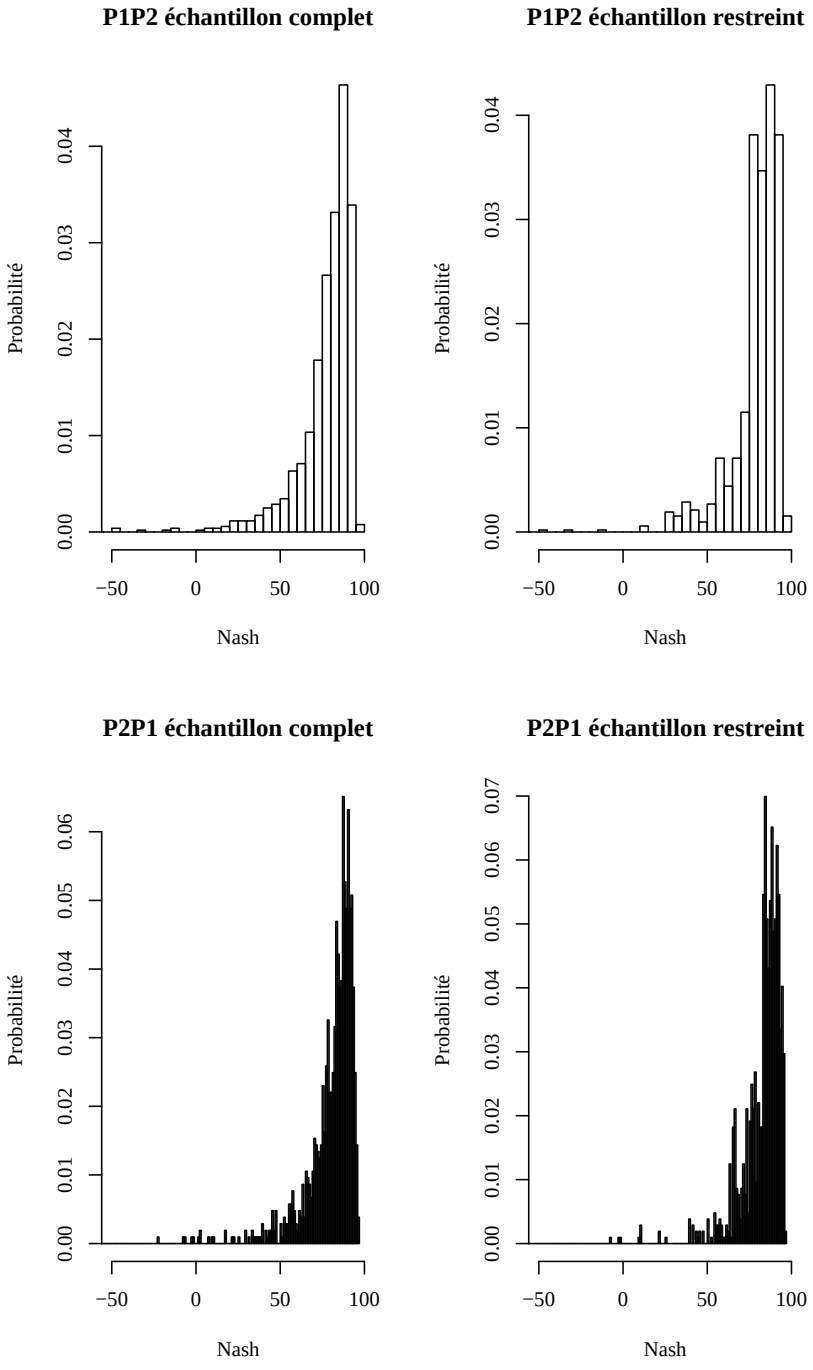


FIG. G.12: Suite – Comparaison des distributions des critères de Nash-Sutcliffe (× 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 322 bassins obtenu par C.A.H. sur les 6 indices caractérisant le débit.

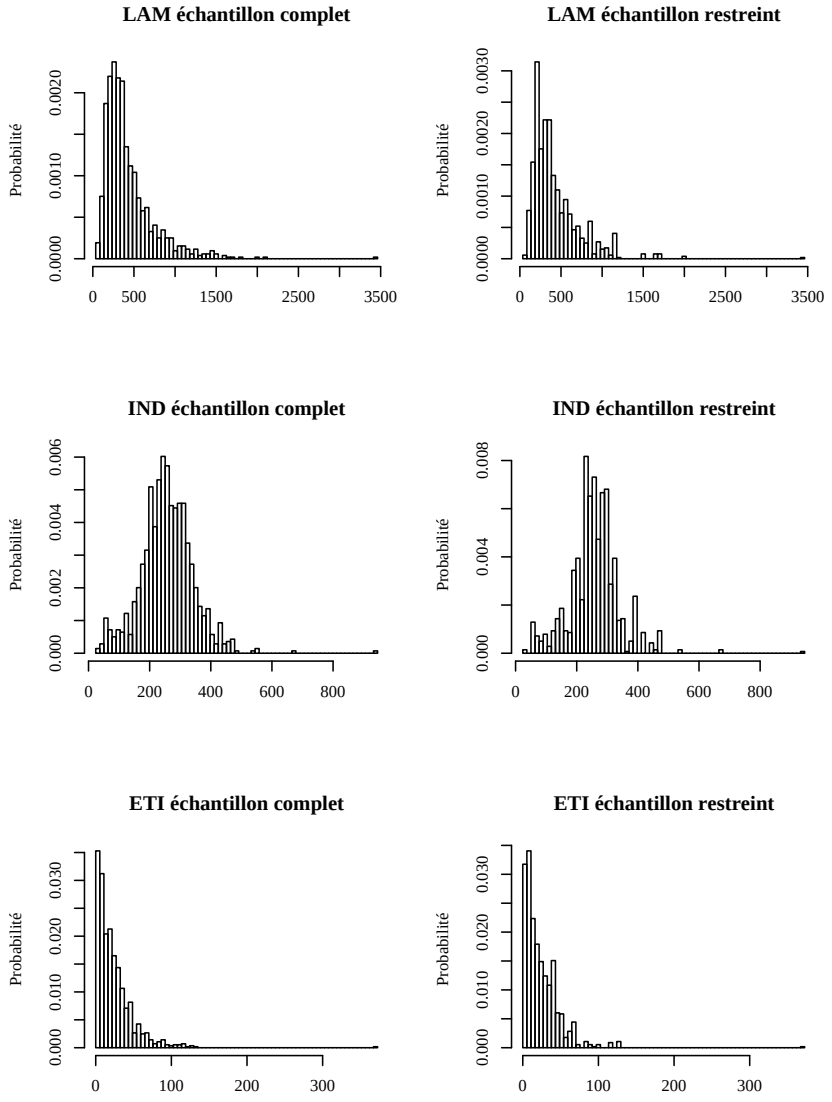


FIGURE G.13 – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 116 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

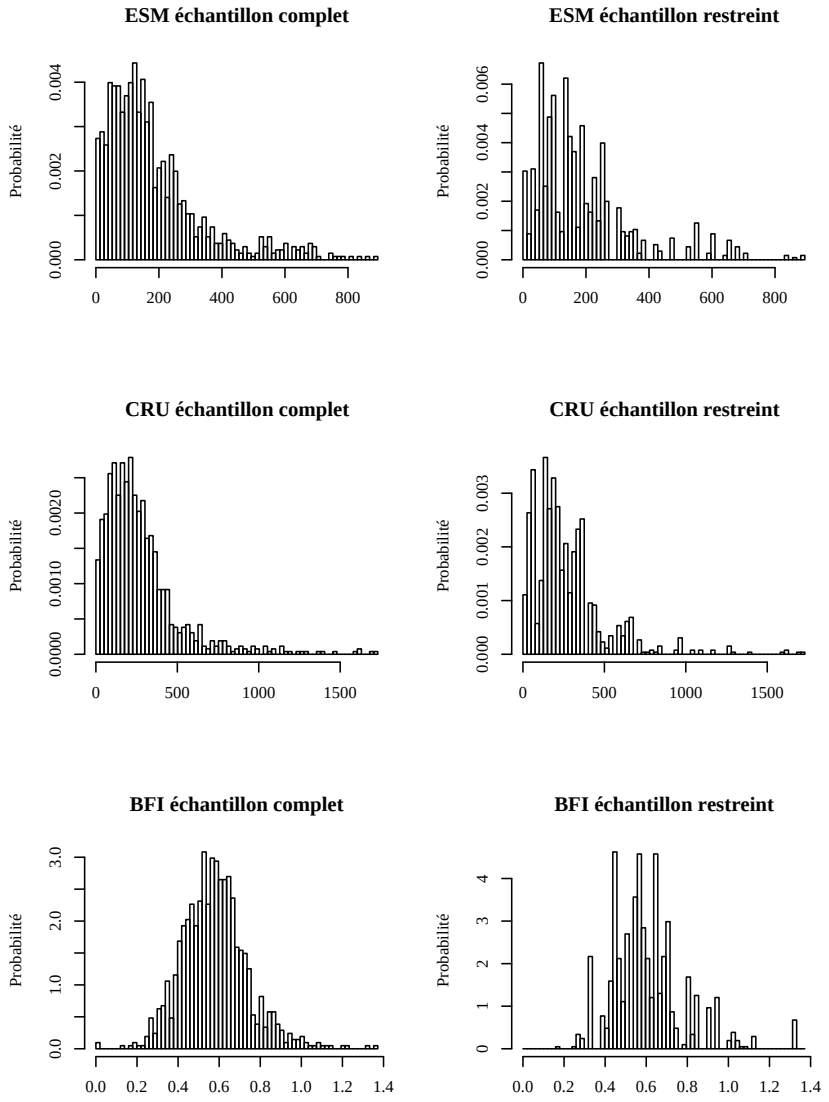


FIG. G.13: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 116 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

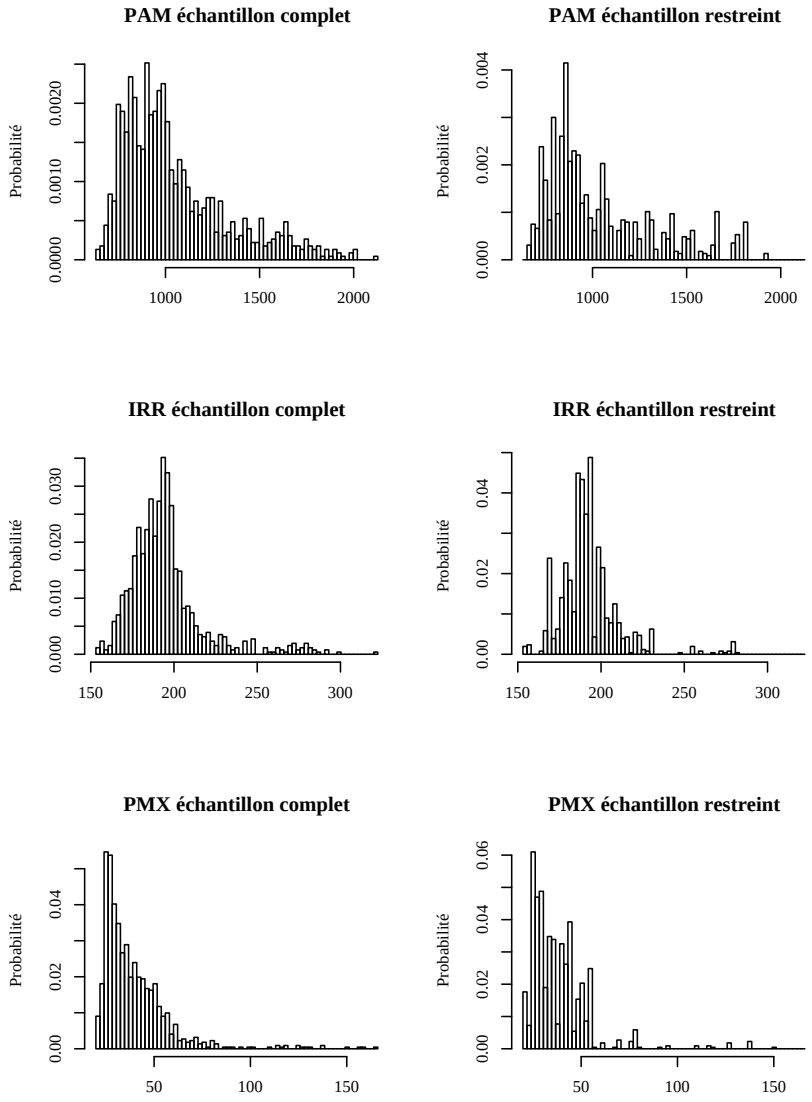


FIG. G.13: Suite – Comparaison des distributions des indices PAM, IRR et PMX de l'échantillon complet et de l'échantillon restreint comptant 116 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

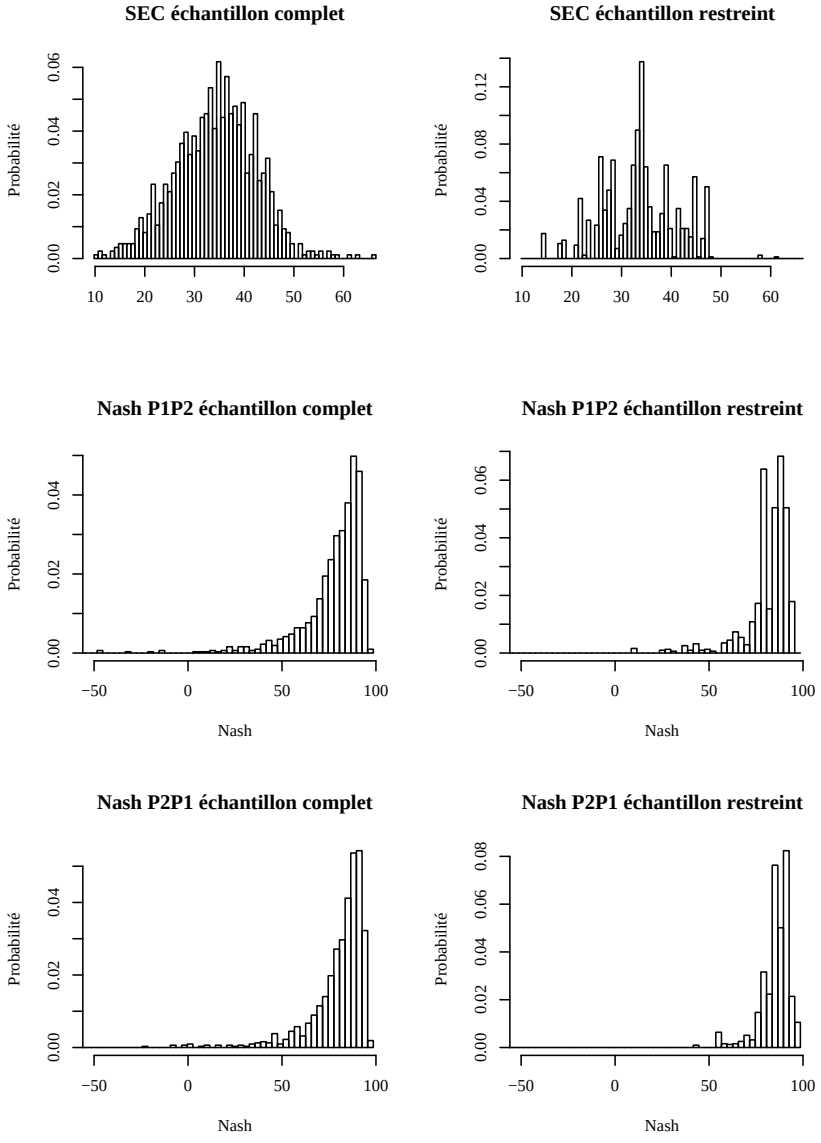


FIG. G.13: Suite – Comparaison des distributions de l'indice SEC et des critères de Nash-Sutcliffe (× 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 116 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

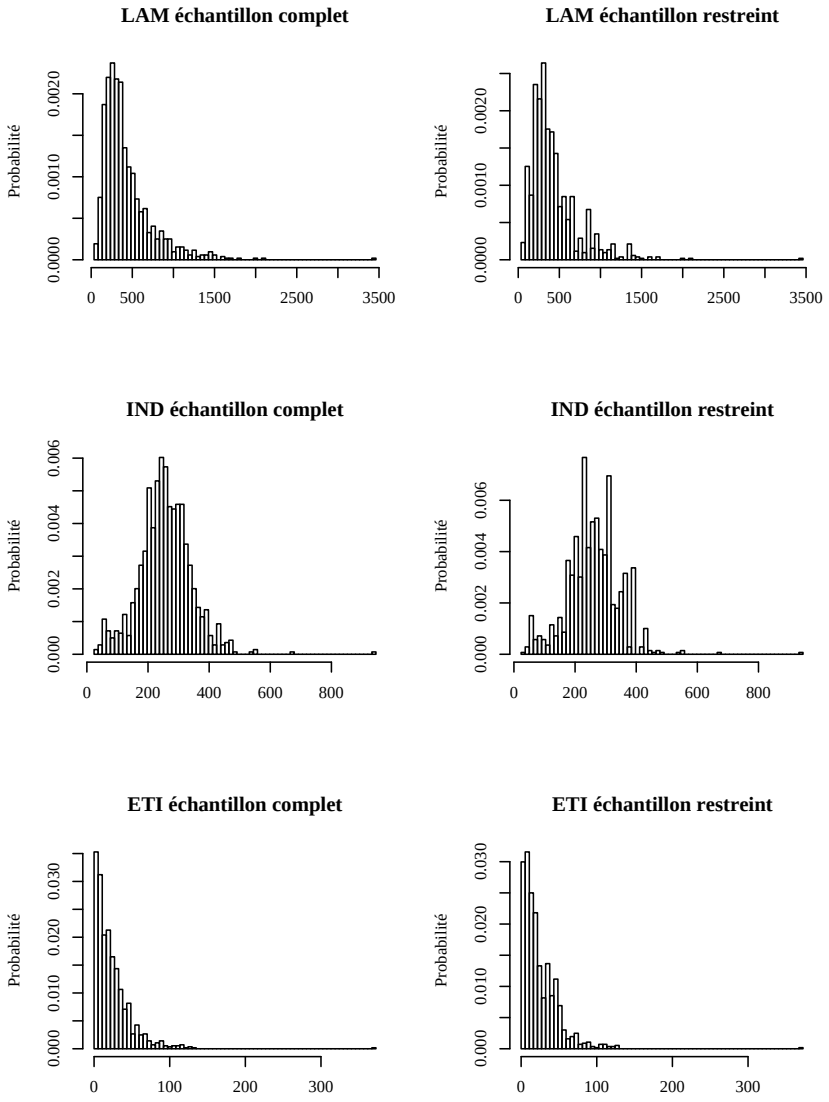


FIG. G.13: Suite – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 221 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

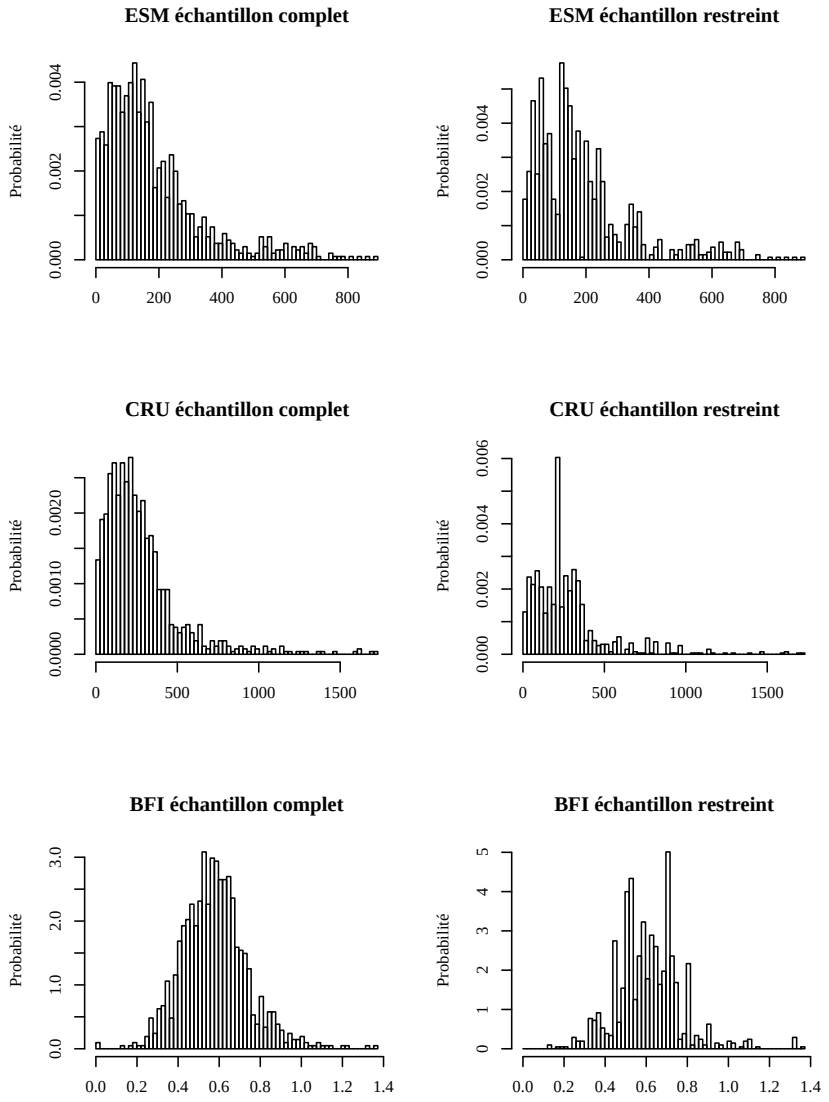


FIG. G.13: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 221 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

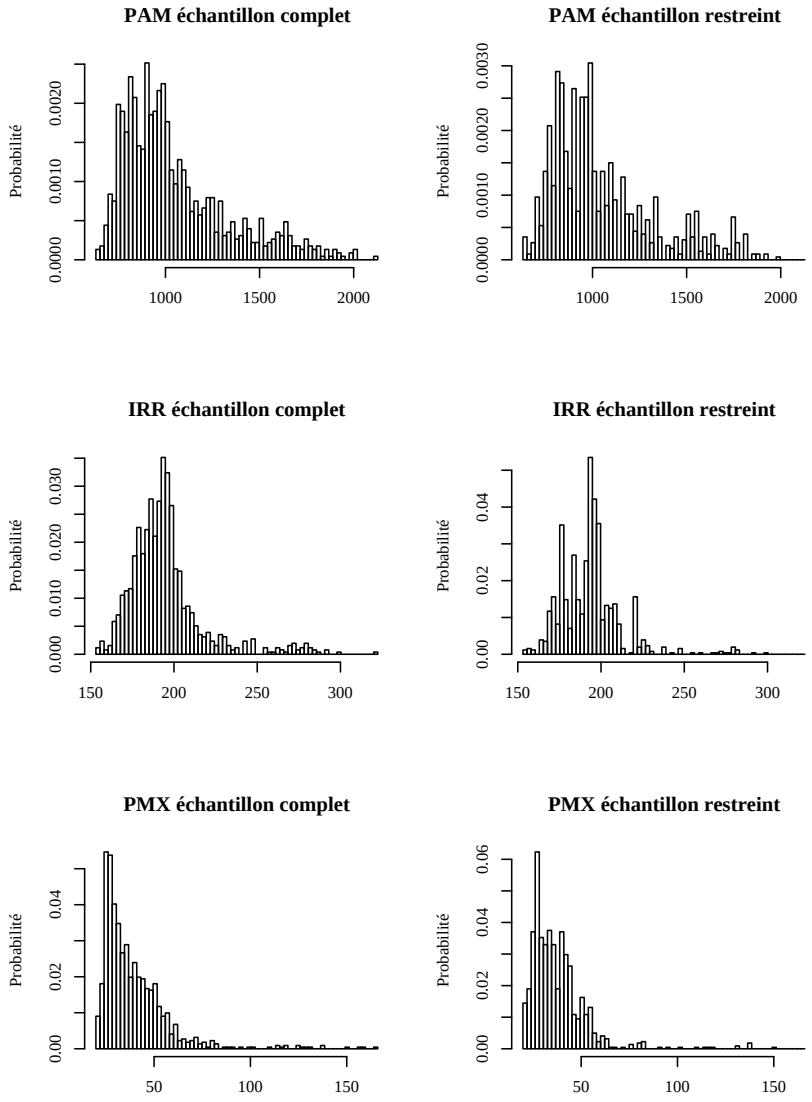


FIG. G.13: Suite – Comparaison des distributions des indices PAM, IRR et PMX de l'échantillon complet et de l'échantillon restreint comptant 221 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

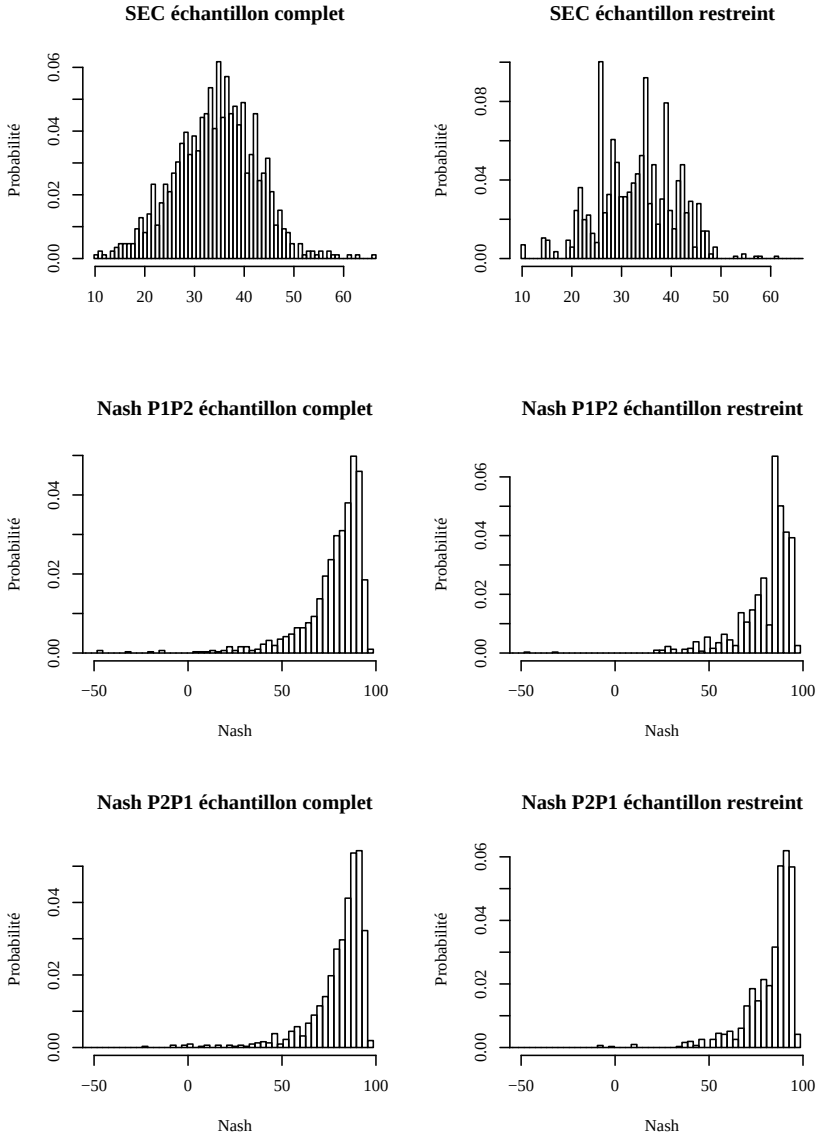


FIG. G.13: Suite – Comparaison des distributions de l'indice SEC et des critères de Nash-Sutcliffe (× 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 221 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

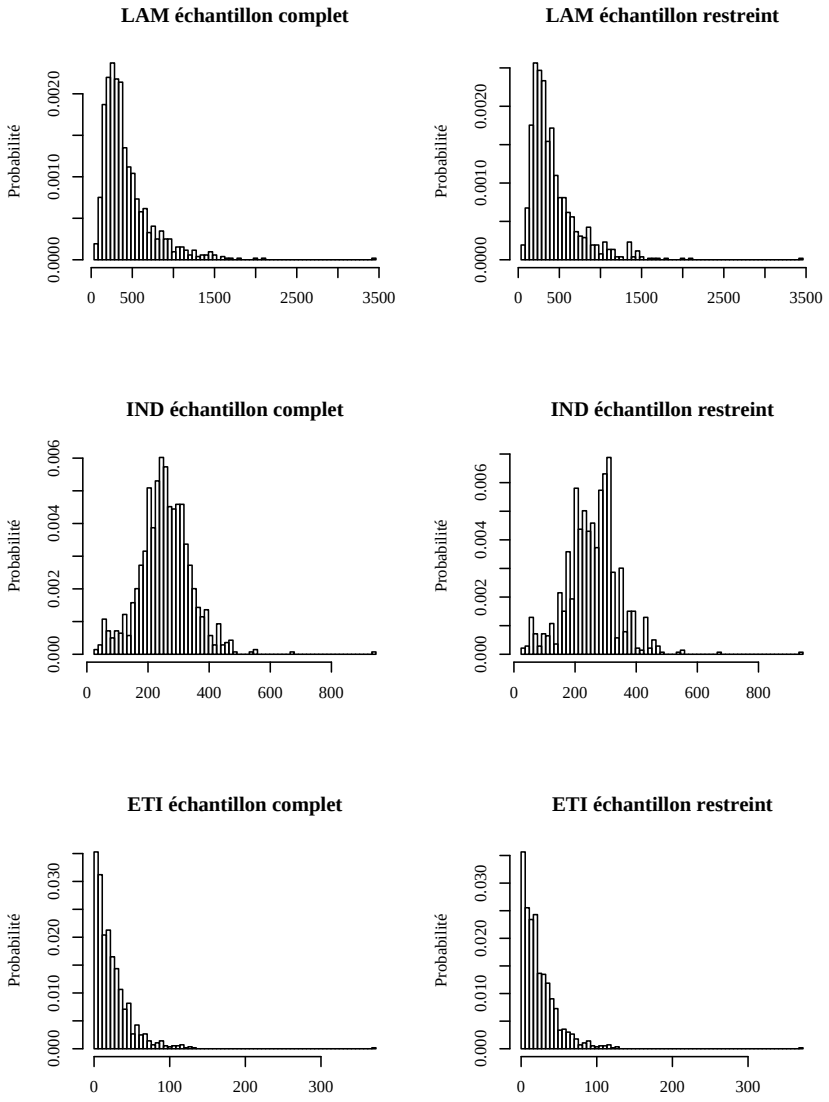


FIG. G.13: Suite – Comparaison des distributions des indices LAM, IND et ETI de l'échantillon complet et de l'échantillon restreint comptant 384 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

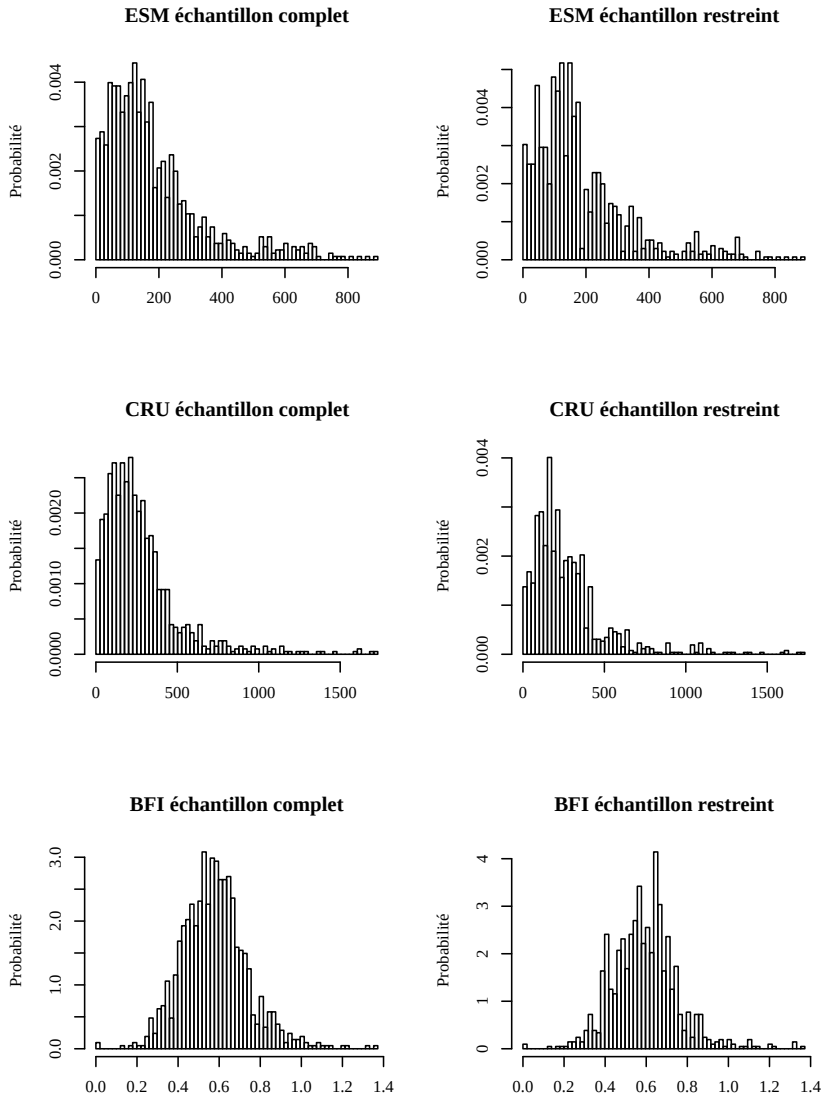


FIG. G.13: Suite – Comparaison des distributions des indices ESM, CRU et BFI de l'échantillon complet et de l'échantillon restreint comptant 384 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

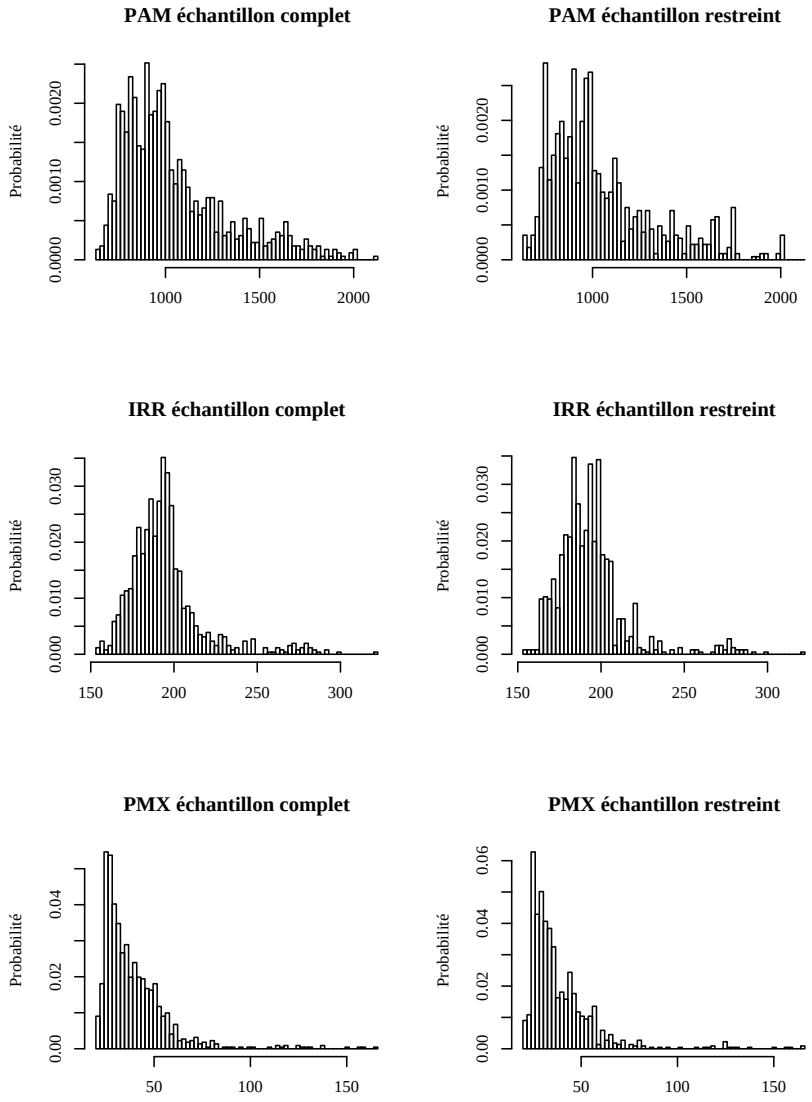


FIG. G.13: Suite – Comparaison des distributions des indices PAM, IRR et PMX de l'échantillon complet et de l'échantillon restreint comptant 384 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.

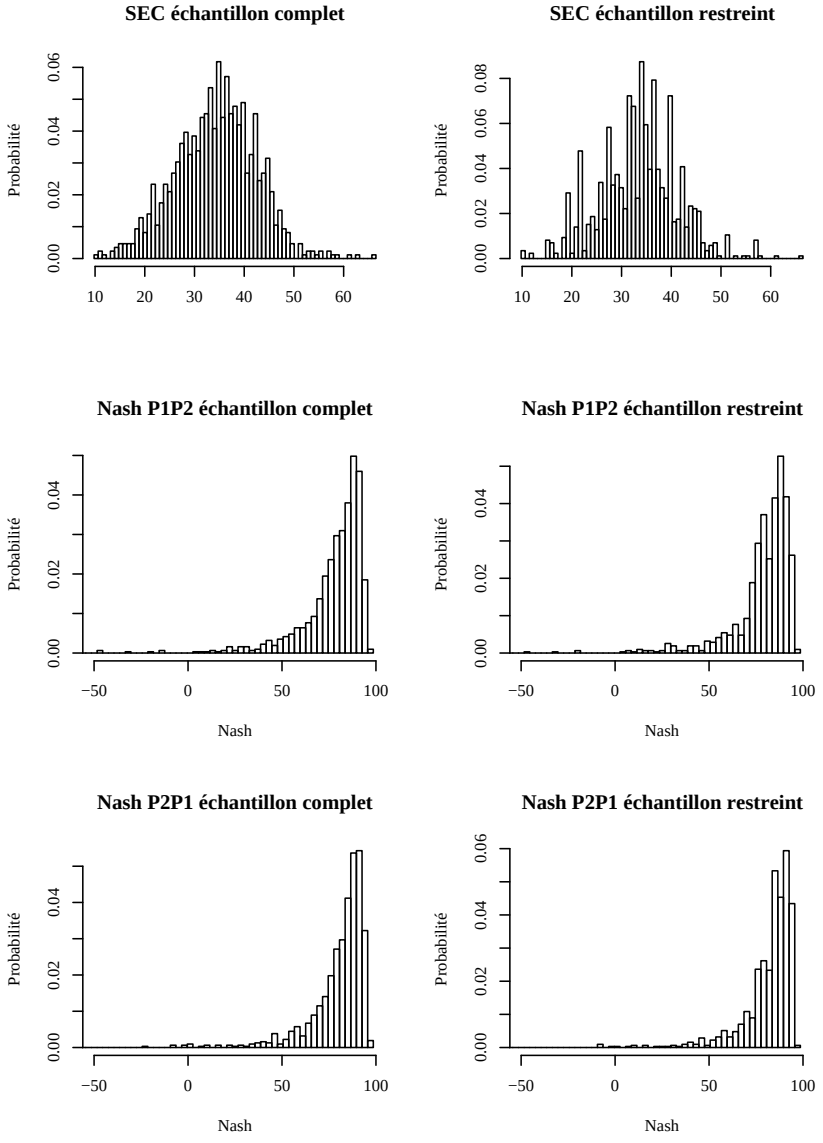
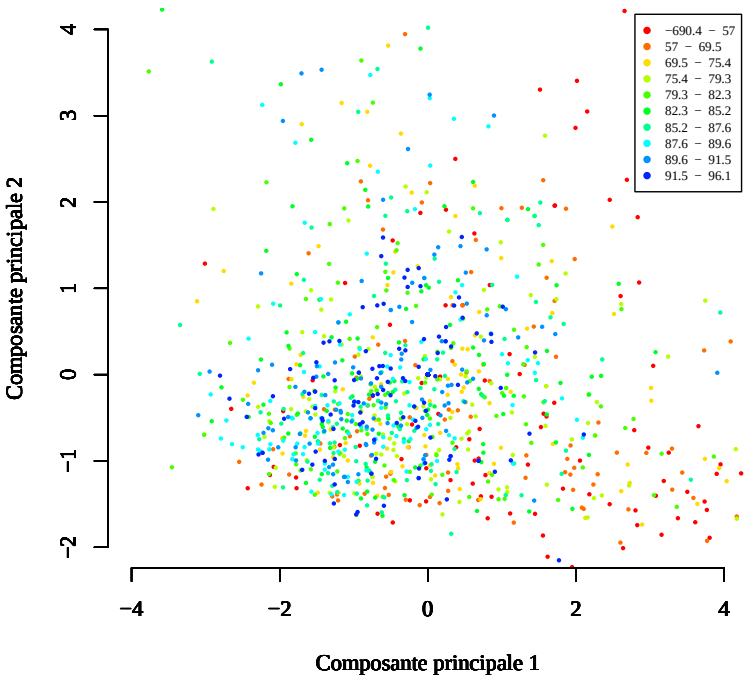


FIG. G.13: Suite – Comparaison des distributions de l'indice SEC et des critères de Nash-Sutcliffe (× 100) calculés en contrôle avec GR4J sur l'échantillon complet et sur l'échantillon restreint comptant 384 bassins obtenu par C.A.H. sur les 10 indices caractérisant la pluie et le débit.



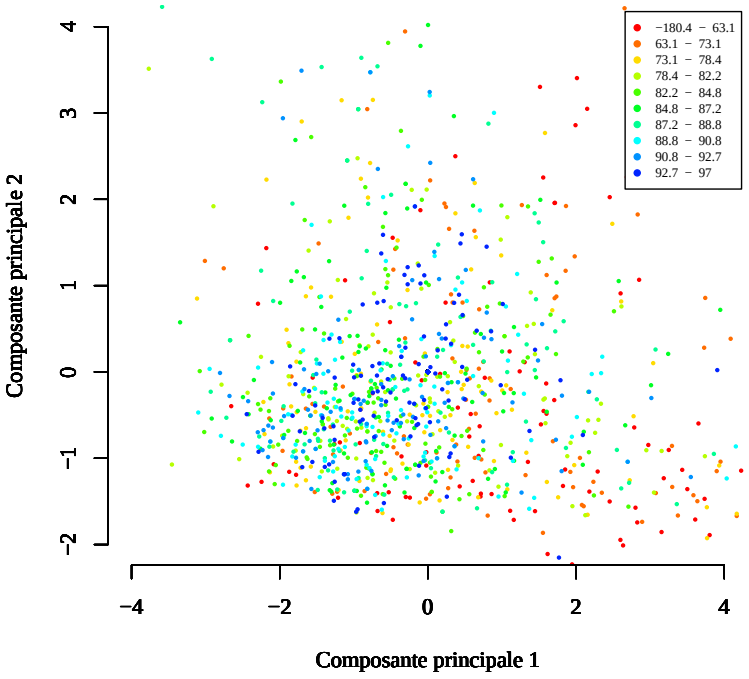


FIGURE G.14 – Distribution du critère de Nash-Sutcliffe (× 100) en contrôle représentée dans le plan principal de l'A.C.P. basée sur les 6 indices décrivant le débit : sur la période P_1 (en haut) et sur la période P_2 (en bas).

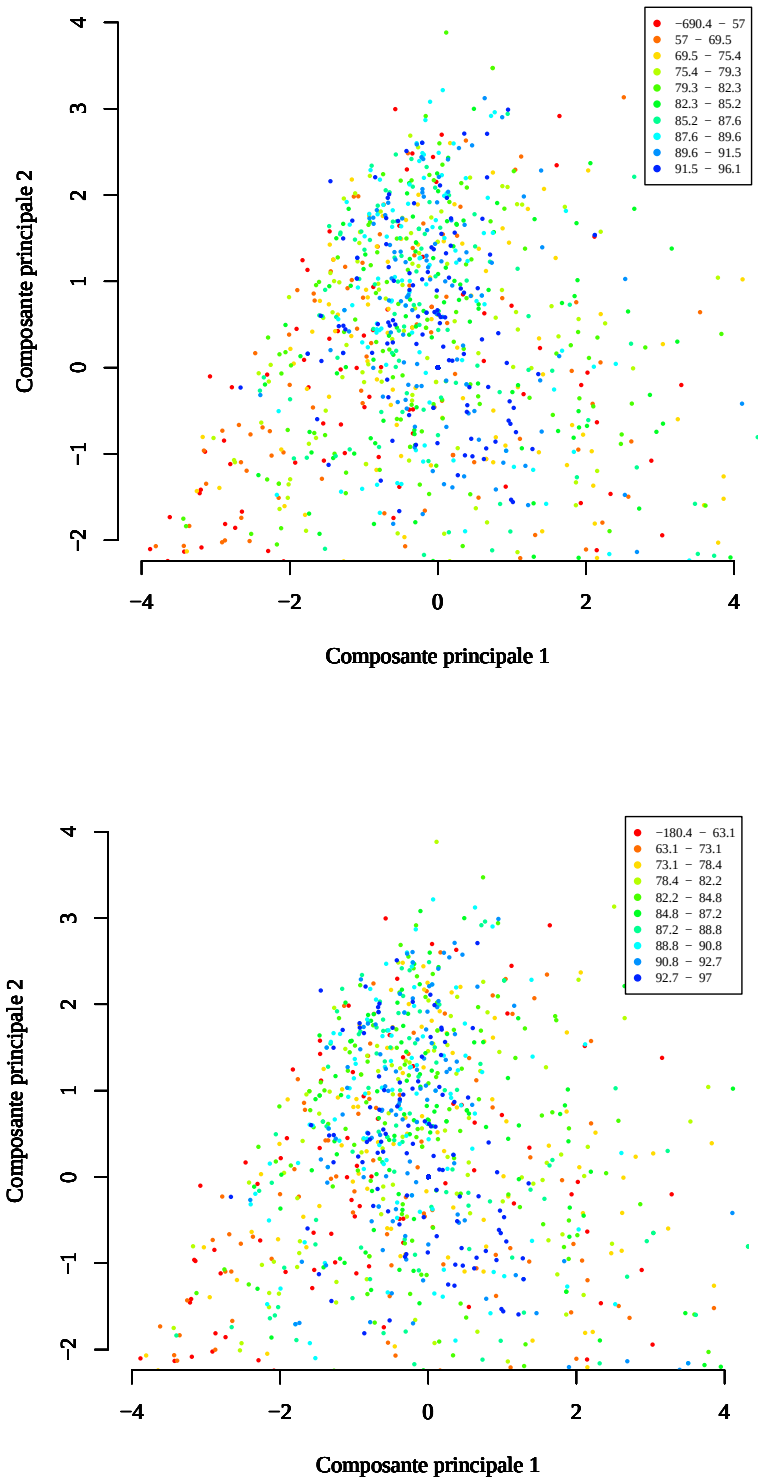


FIG. G.14: Suite – Distribution du critère de Nash-Sutcliffe (× 100) en contrôle sur la seconde période représentée dans le plan principal de l'A.C.P. basée sur les 10 indices décrivant le débit et la pluie : sur la période P_1 (en haut) et sur la période P_2 (en bas).

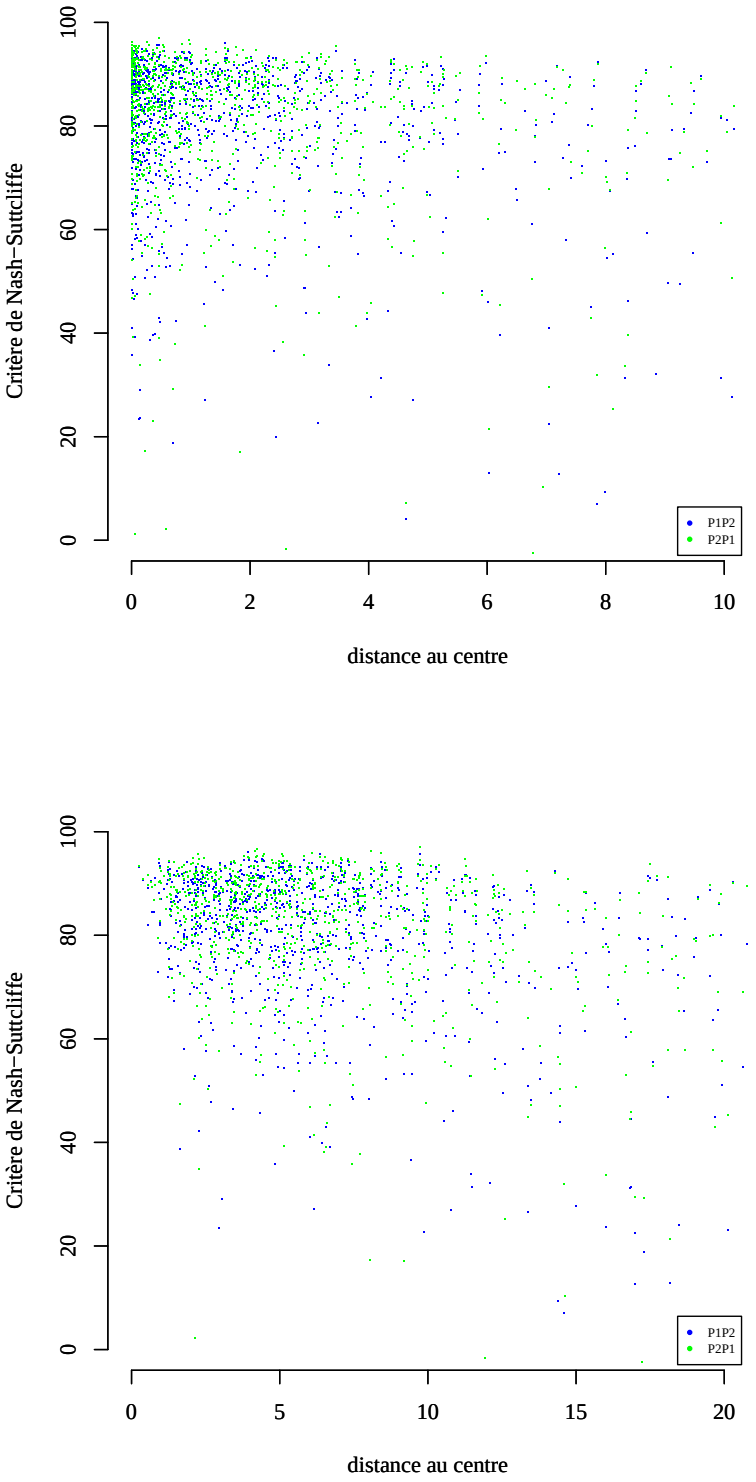


FIGURE G.15 – Distribution du critère de Nash-Sutcliffe ($\times$ 100) en contrôle représentée en fonction de la distance à l'origine dans l'espace des 6 indices décrivant le débit (en haut) et dans l'espace des 10 indices décrivant les régimes hydrographiques et pluvieux (en bas).

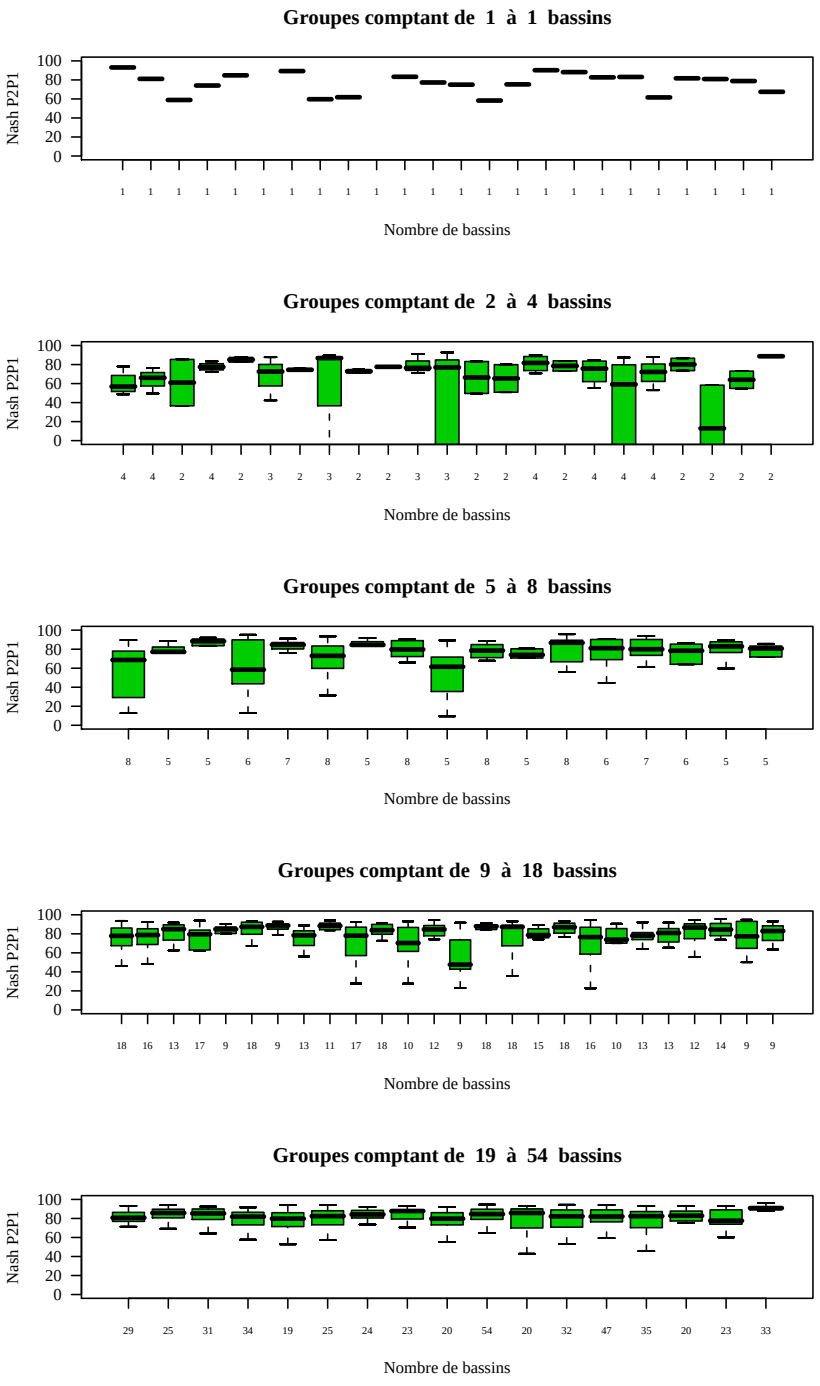


FIGURE G.16 – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la première période dans l'échantillon restreint Q-107.

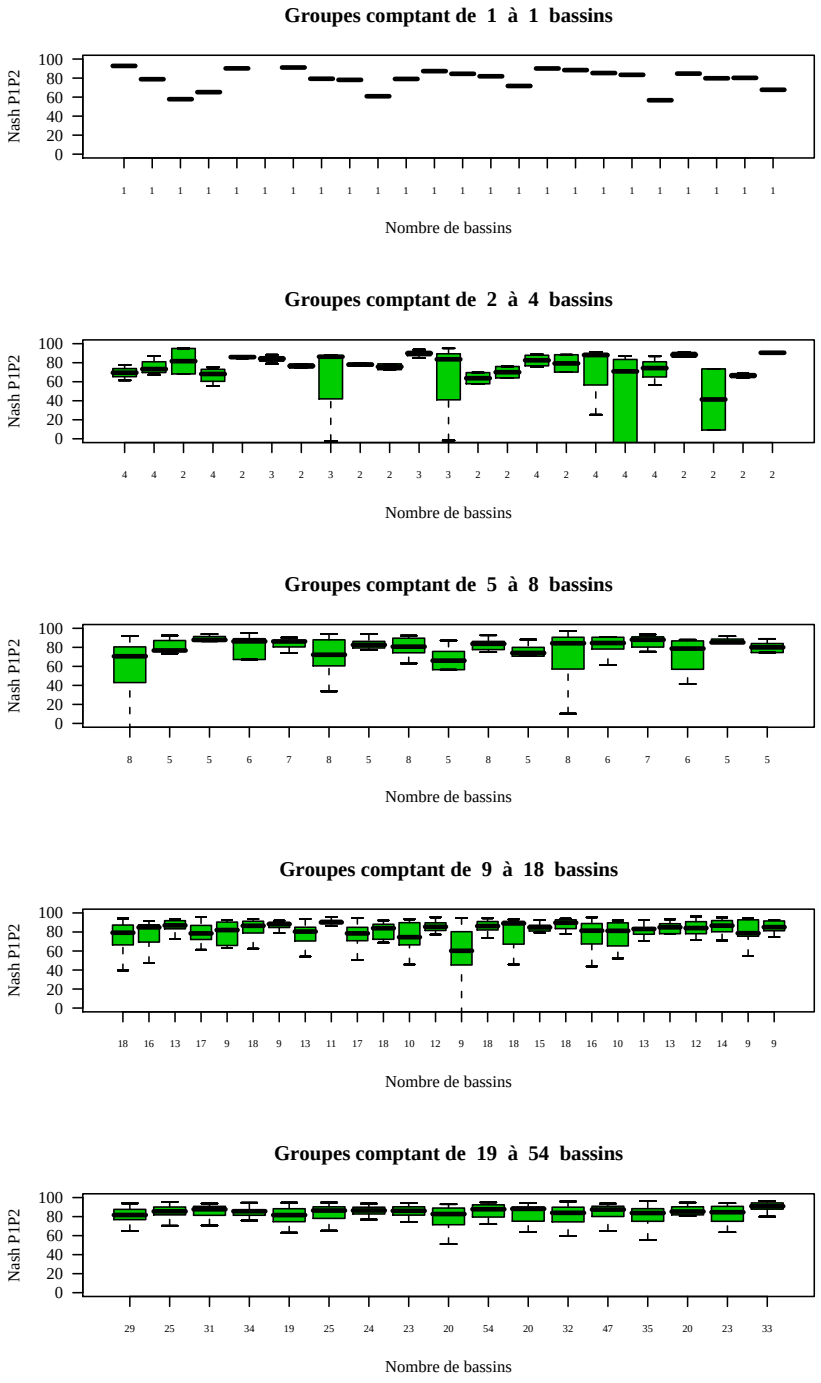


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la seconde période dans l'échantillon restreint Q-107.

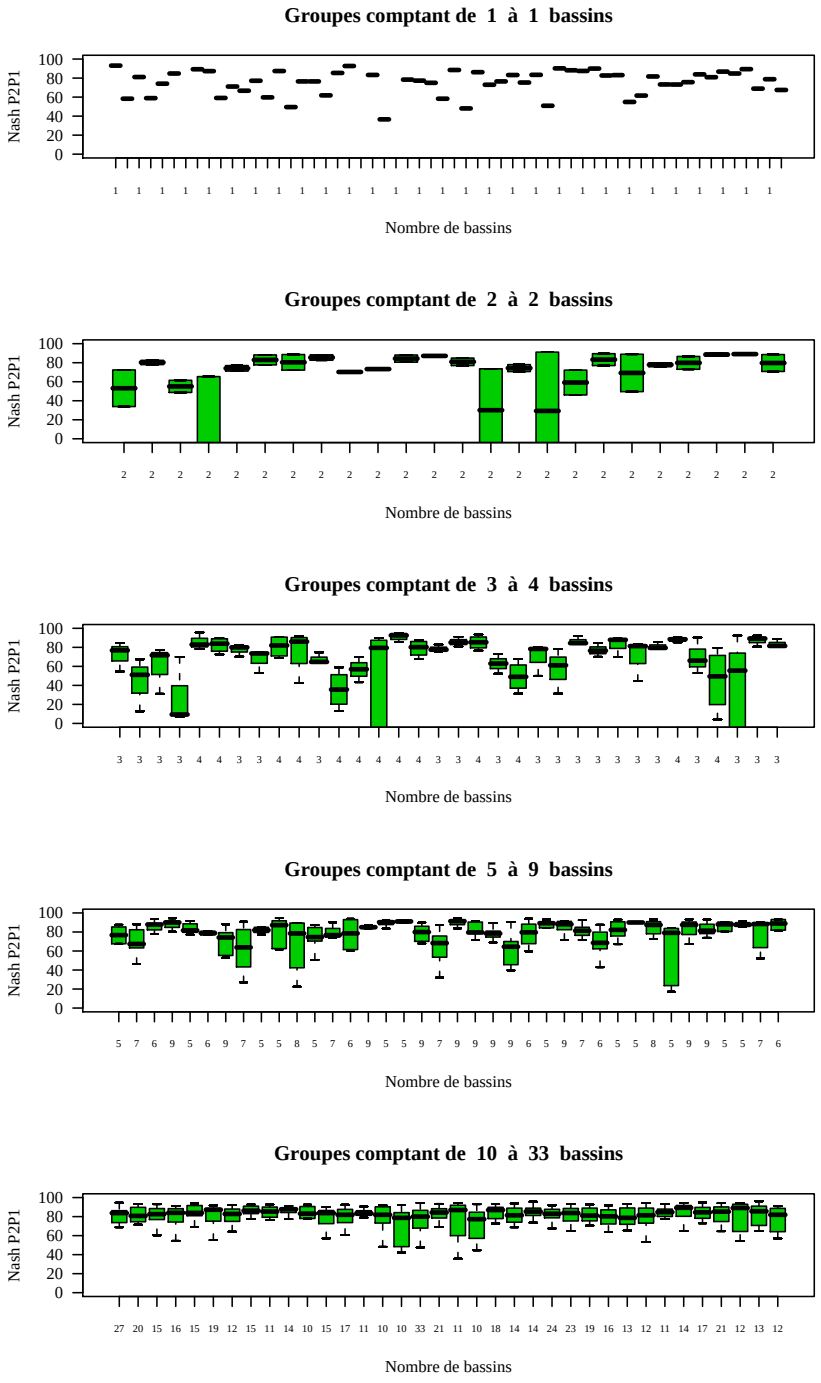


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la première période dans l'échantillon restreint Q-190.

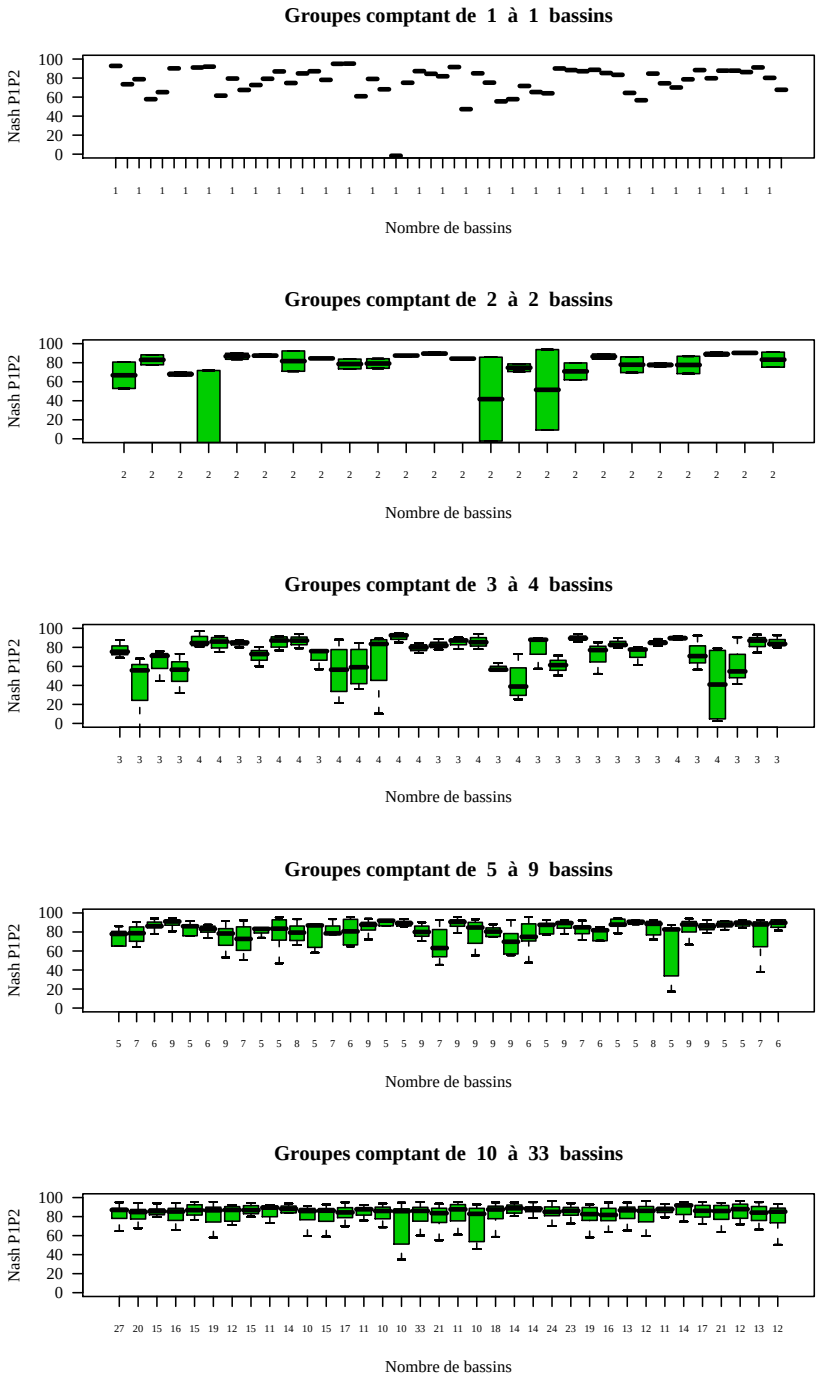


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la seconde période dans l'échantillon restreint Q-190.

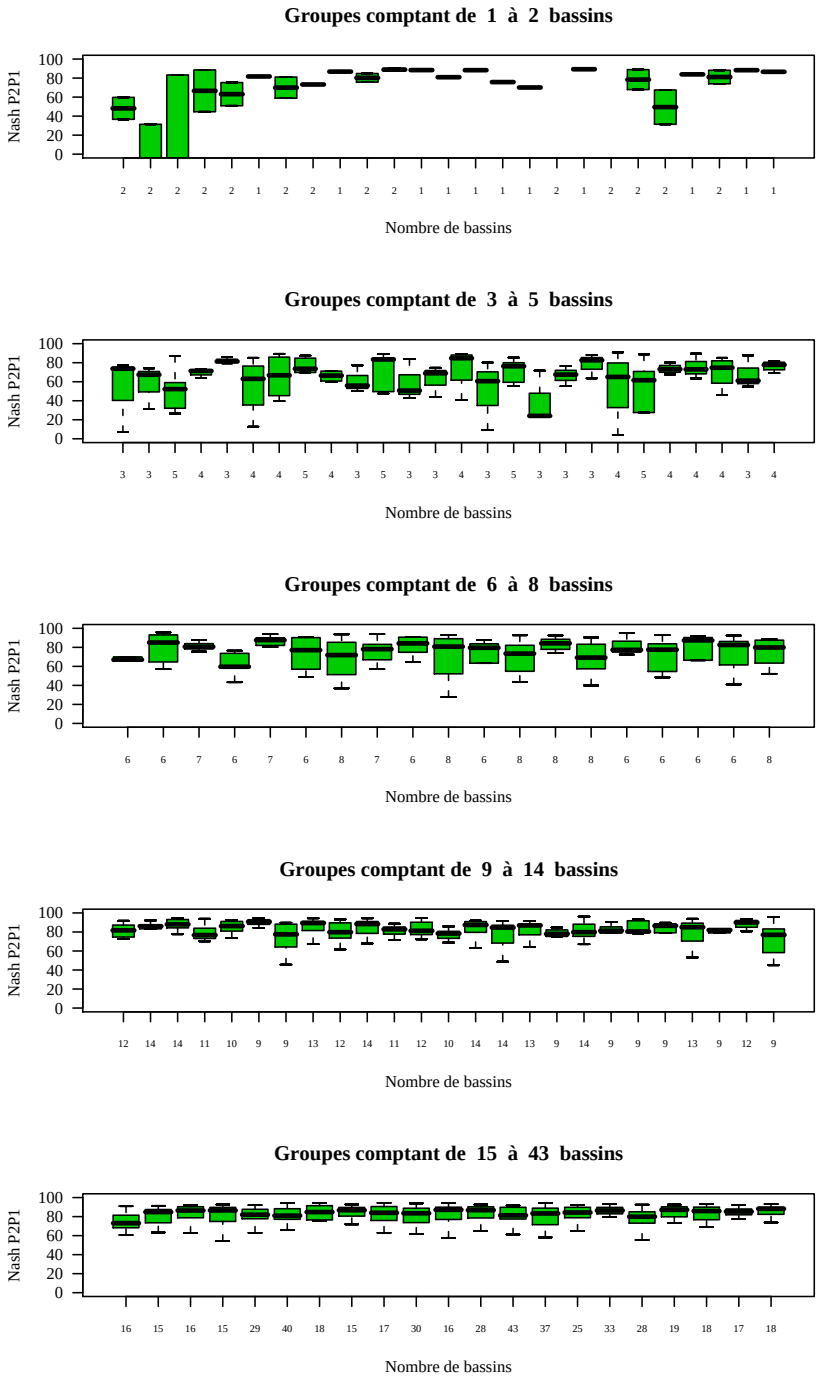


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la première période dans l'échantillon restreint PQ-116.

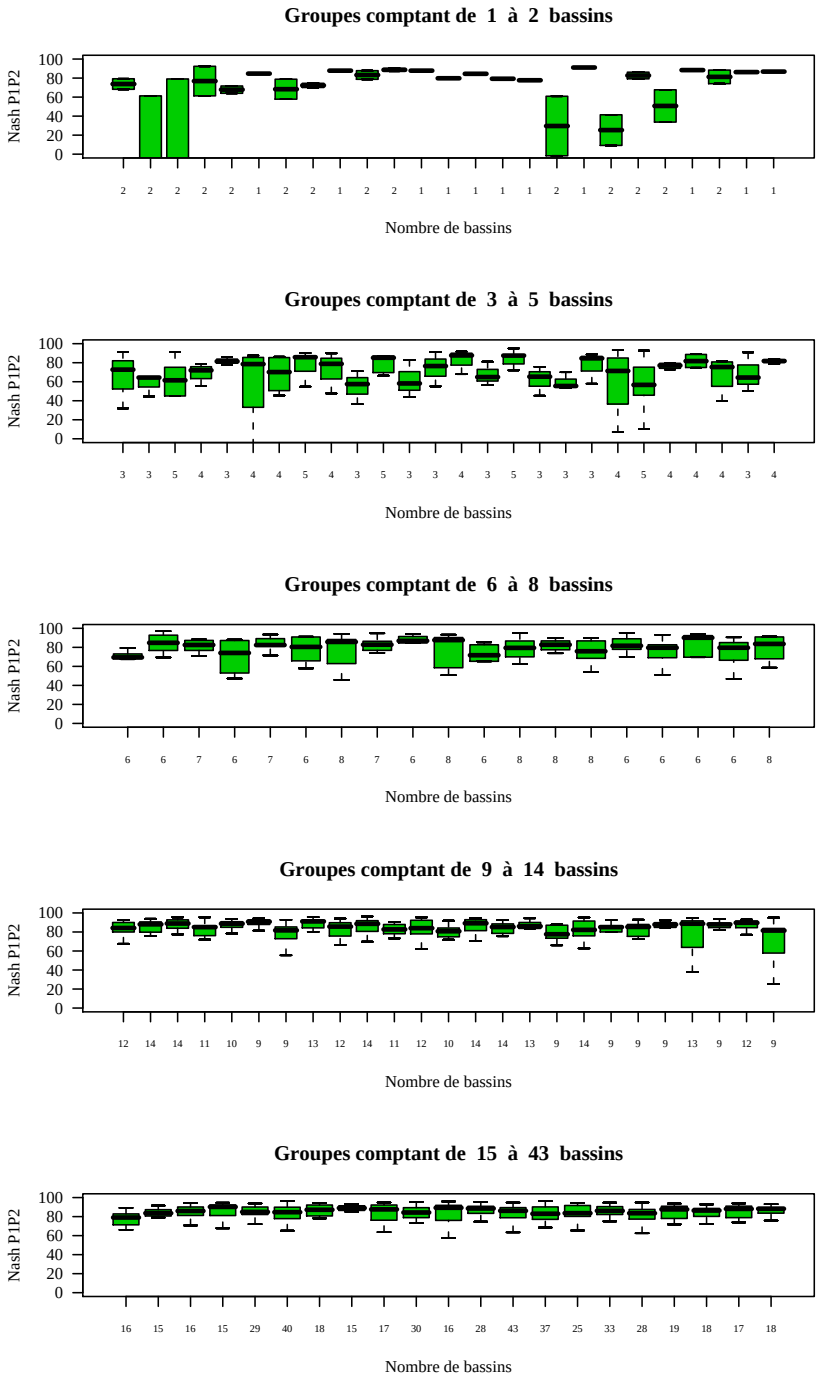


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la seconde période dans l'échantillon restreint PQ-116.

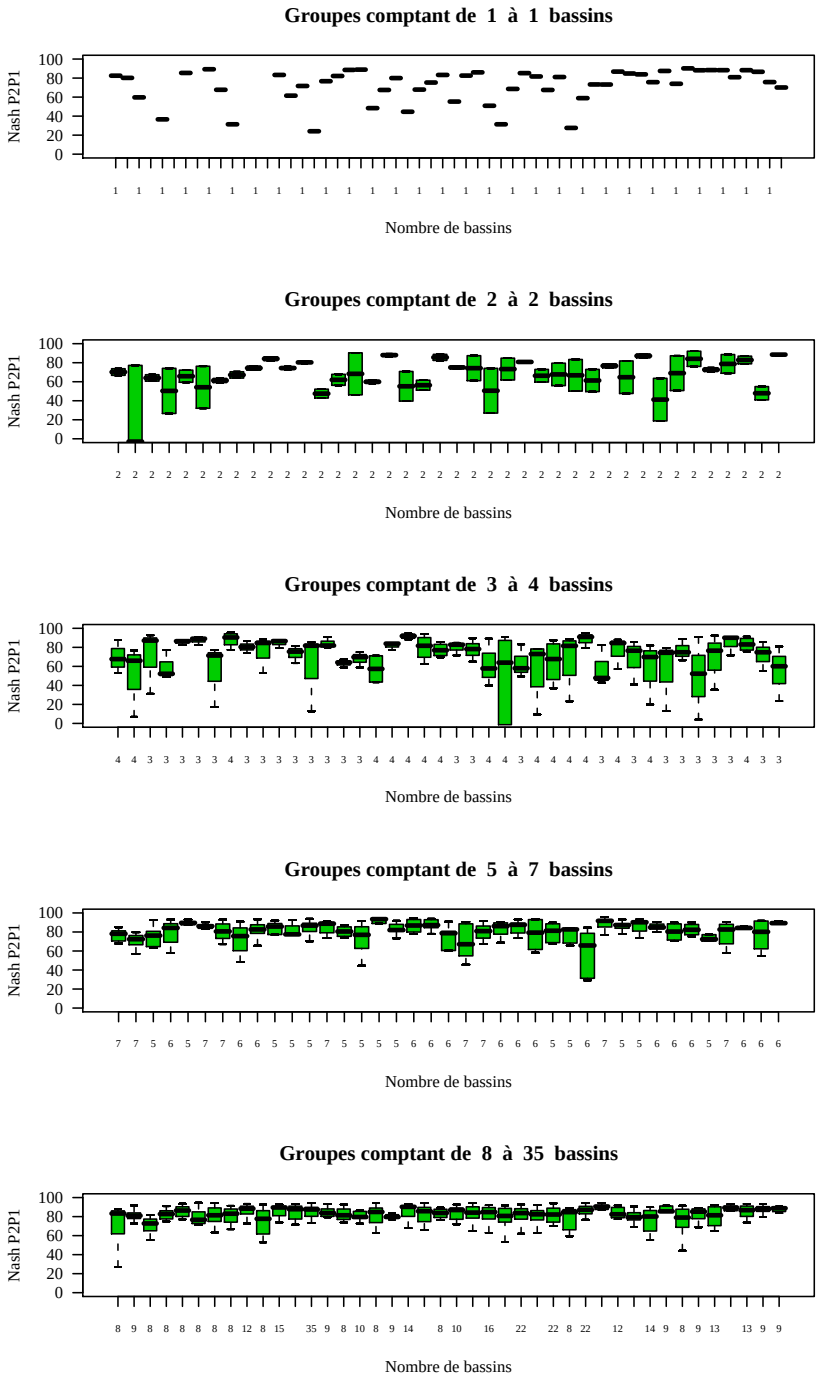


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la première période dans l'échantillon restreint PQ-221.

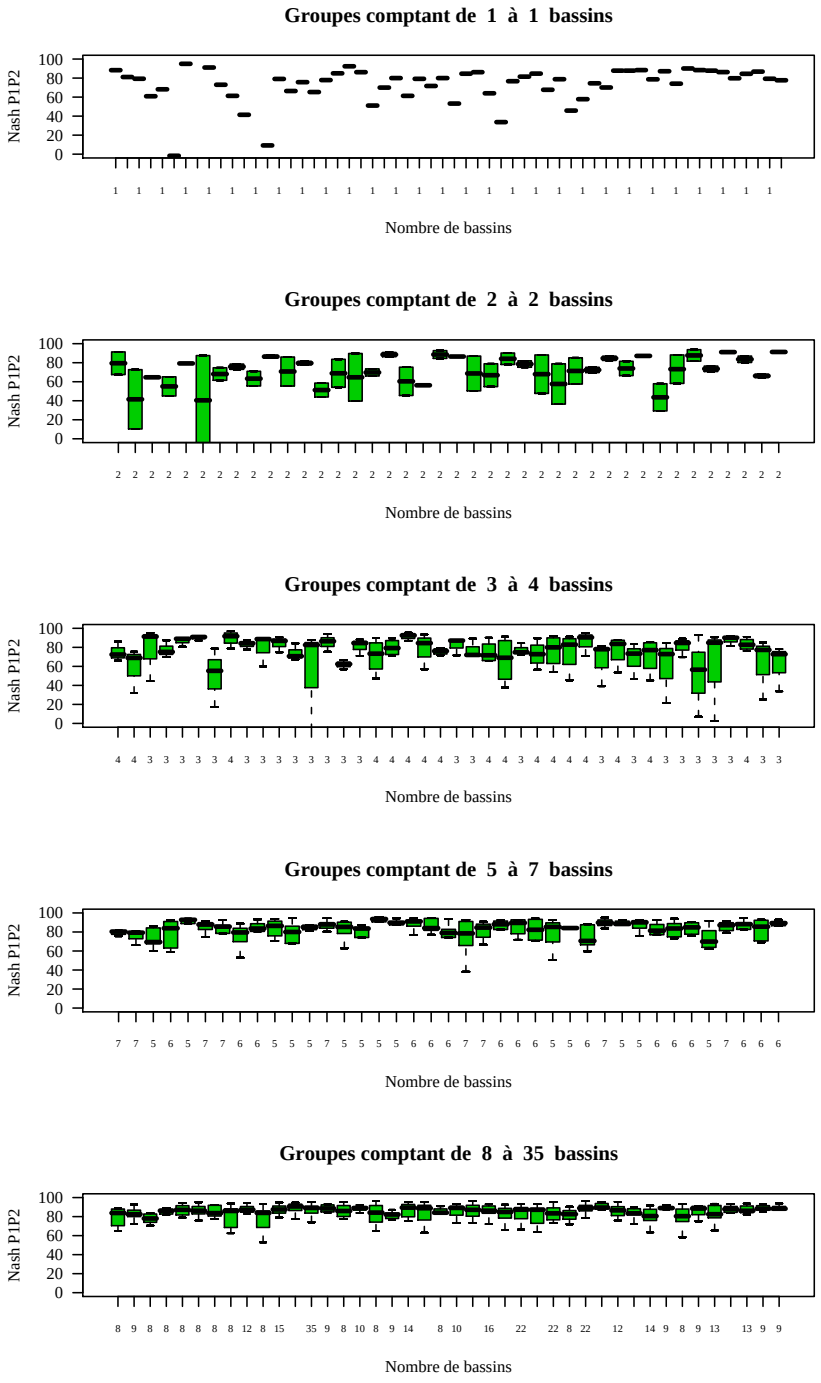


FIG. G.16: Suite – Variété intra-groupe du critère de Nash-Sutcliffe ($\times$ 100) en contrôle sur la seconde période dans l'échantillon restreint PQ-221.

$\mathcal{H}$

Annexe

Influence des données douteuses sur l'évaluation des modèles de prévision de crues

Comme nous l'avons vu au chapitre 4, page 57, de nombreuses données de débit en hautes eaux sont qualifiées de douteuses dans la banque Hydro. Cela peut gêner l'évaluation d'un modèle de prévision des crues. Dans cette étude, nous allons quantifier par plusieurs méthodes l'importance des données ainsi qualifiées lors des passages de hautes eaux puis nous estimerons l'impact de la non prise en compte de ces données dans l'évaluation d'un modèle Pluie – Débit.

H.1 Quantification des crues qualifiées de douteuses

Trois méthodes ont été comparées en collaboration avec Pierre Javelle [1].

H.1.1 Approche par un algorithme sup-seuil

Pierre Javelle a employé un algorithme sup-seuil sur l'ensemble des données (validées et douteuses) pour calculer le débit de période de retour 6 mois. Il a dénombré ensuite le nombre de pics de crues au-dessus de ce seuil dans les données validées uniquement (Fig. H.1). On observe que pour un très grand nombre de bassins, beaucoup de crues sont tout simplement non prises en compte dans les seules données validées.

L'algorithme sup-seuil fixe initialement son seuil au débit maximal observé sur la chronique. Il diminue progressivement son seuil en comptant le nombre de pics sans lacunes qui sont au-dessus. L'algorithme se termine quand le nombre de pics de crue correspond à la moyenne attendue (ici 2 crues par an). Lors de la descente du débit seuil, le nombre de pics situés au-dessus commence par augmenter. Si le seuil entre dans la gamme des débits fréquents, ce nombre peut se stabiliser voire décroître. L'algorithme ne fonctionne correctement dans la phase où le nombre de pics augmente. En présence de lacunes ou de variations interannuelles, il peut échouer. C'est une explication possible pour le petit nombre de bassins versants où l'algorithme n'a pas su définir une valeur correcte pour le débit seuil.

Malgré ces défauts, on peut considérer le nombre moyen de pics de crue par an dans les seules données validées comme un indice de la présence de crue dans les données validées.

H.1.2 Ratio des quantiles en hautes eaux

Une autre approche consiste à calculer le rapport ou la différence entre les quantiles de débits $q_x(Q)$ calculés sur les données validées uniquement ou sur l'ensemble des données validées et douteuses, pour un quantile élevé (périodes de crues). On choisit ici de déterminer les rapports des quantiles 100, 99.5, 99, 98, 95 et 90 % (pour ce dernier, les points sont largement éloignés des pics de crue).

1. Alors au groupement d'Antony du Cemagref, actuellement au groupement d'Aix-en-Provence.

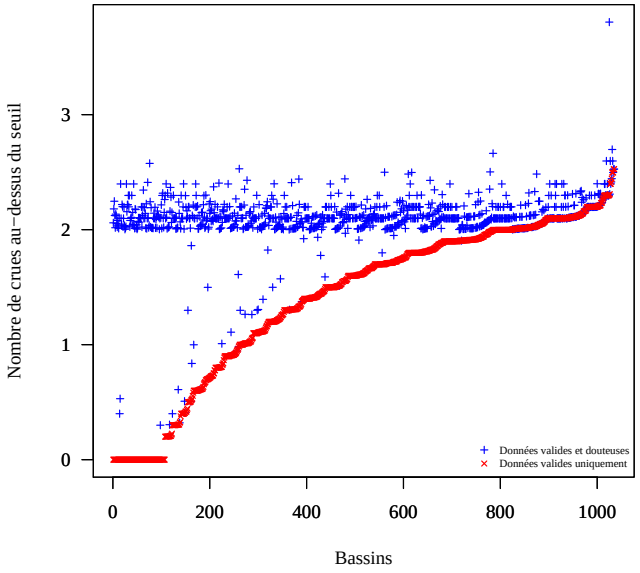


FIGURE H.1 – Nombre de pics de crues par an au dessus du débit de période de retour 6 mois $Q^{[6\ mois]}$, selon que les données douteuse sont prises en compte (+) ou non (×). Les débits $Q^{[6\ mois]}$ ont été calculés par l'algorithme sup-seuil à partir de toutes les données disponibles pour chaque bassin. Les bassins sont classés par le nombre de pics au dessus du débit $Q^{[6\ mois]}$ quand seules les données validées sont prises en compte.

Cet indice est assez bien corrélé avec le précédent, ce qui est rassurant quant à leurs significations. Cependant quand le nombre de pics validés au-dessus du seuil déterminé par l'algorithme sup-seuil tend vers 0, le ratio des quantiles en hautes eaux peut prendre n'importe quelle valeur (supérieure à 1).

H.1.3 Répartition des événements de fortes pluies et de hauts débits

Nous avons calculé les distributions des cumuls de pluie (sur trois heures) sur la période entière puis nous avons sélectionné les cumuls précédant les débuts des passages où les données de débit ne sont pas validées et avons examiné leur répartition dans cette distribution (Tab. H.1). De même, la distribution des débits validés a été calculée et nous avons détaillé la répartition des débits validés juste avant ou juste après les passages de débit non validé. Dans les trois cas, nous observons une très forte surreprésentation des plus forts cumuls de pluie ou des plus forts débits : un cinquième à un quart des débits bordant ces passages sont supérieurs au quantile q_{99}, c'est-à-dire 20 à 25 fois plus que s'il n'y avait pas concentration des non validation des données de débit sur les périodes de très hautes eaux. La comparaison des résultats présentés dans le tableau H.1 montre que ce sont moins les

Quantiles des distributions	0	0,8	0,9	0,95	0,98	0,99	1
Répartition attendue (%)	*80,0*	*10,0*	*5,0*	*3,0*	*1,0*	*1,0*	
Répartition des cumuls de pluie (%)	48,9	14,4	5,0	6,4	4,7	11,6	
Répartition des débits antérieurs (%)	66,5	3,6	3,6	5,3	4,0	17,0	
Répartition des débits postérieurs (%)	64,8	2,7	2,3	3,4	3,4	23,4	

Table H.1 – Répartition des cumuls de pluie antérieurs aux passages de données non validées et débits antérieurs et postérieurs par rapport aux distributions calculées sur les données de débit validées.

événements climatiques que le niveau du débit qui joue : le facteur principal expliquant cette concentration des données non validées en hautes eaux est donc probablement lié à la limitation de la courbe de tarage.

La carte H.2 ne montre pas de région qui soit particulièrement affectée par le phénomène. Ce sont plutôt les petits bassins qui sont concernés (ce qui est assez logique) mais le faible nombre de grands bassins disponibles dans notre ensemble de données ne permet pas de conclure avec certitude.

H.2 Comparaison des performances obtenues avec et sans ces données douteuses

Nous avons comparé l'évaluation des performances de deux modèles selon que nous tenions compte des données douteuses ou non pour calculer le critère de performance. Ces modèles sont GR3P (Tangara, 2005) et GRP avec correction des sorties par réseaux de neurones artificiels.

H.2.1 Avec le modèle GRP

Nous avons comparé les performances du modèle GRP dans la version proposée par Tangara (2005) en employant les seules données validées puis les données validées complétées des données douteuses. Les performances ont été mesurées avec le critère de persistance classique et le critère de persistance calculé sur les transformations logarithmiques des débits.

Dans tous les cas (quel que soit le critère de performance employé et quel que soit l'horizon de prévision choisi), la différence entre les performances est non ou faiblement significative pour une majorité de bassins (Fig. H.3). Il y a quelques bassins pour lesquels la différence est très forte mais la différence va aussi bien à la hausse qu'à la baisse. On n'observe pas de comportements particuliers par classes de réactivité des bassins.

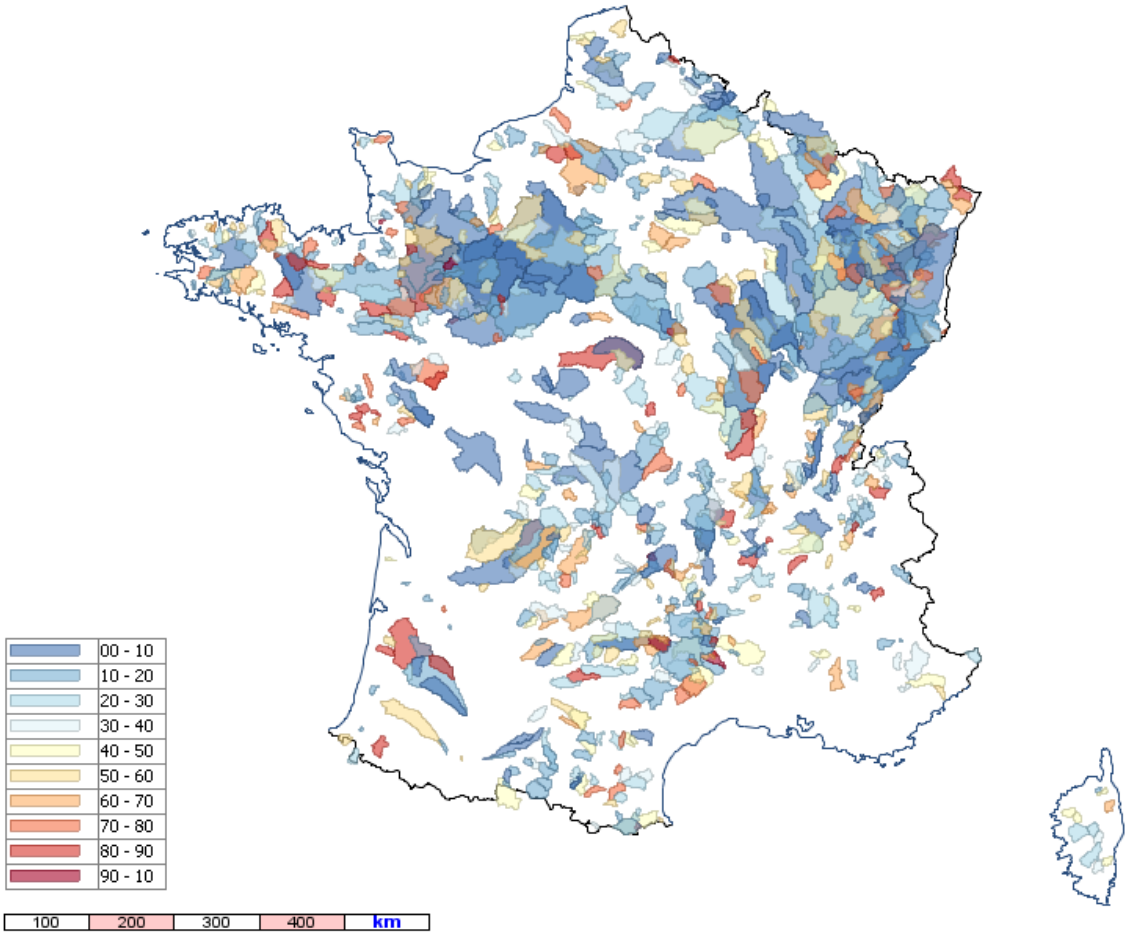


FIGURE H.2 – Taux de données de débit non validées en hautes eaux (définies par $Q_{Fin} \geq q_{99}(Q_{Obs})$) : ratio du nombre de données non validées en hautes eaux sur le nombre total de données non validées (par bassin).

Quand l'horizon de prévision s'accroît, le modèle employant les données douteuses obtient des résultats légèrement meilleurs en moyenne que le modèle n'utilisant que les seules données validées : cela n'interdit pas de penser que la plupart des données douteuses ne sont pas d'une qualité très inférieure à celle des données validées... Il semble donc possible d'utiliser les données douteuses sans grand dommage pour évaluer le modèle GRP. On peut aussi utiliser sans regret les seules données validées pour tester le modèle GRP et ses évolutions.

H.2.2 Avec le modèle GRP-RNA

Nous avons ensuite réalisé la même comparaison avec le modèle GRP dont les sorties sont corrigées par réseaux de neurones artificiels (Chap. 10) car il s'agit d'une de nos meilleures alternatives et car cette approche n'est définie qu'en hautes eaux pour des raisons de

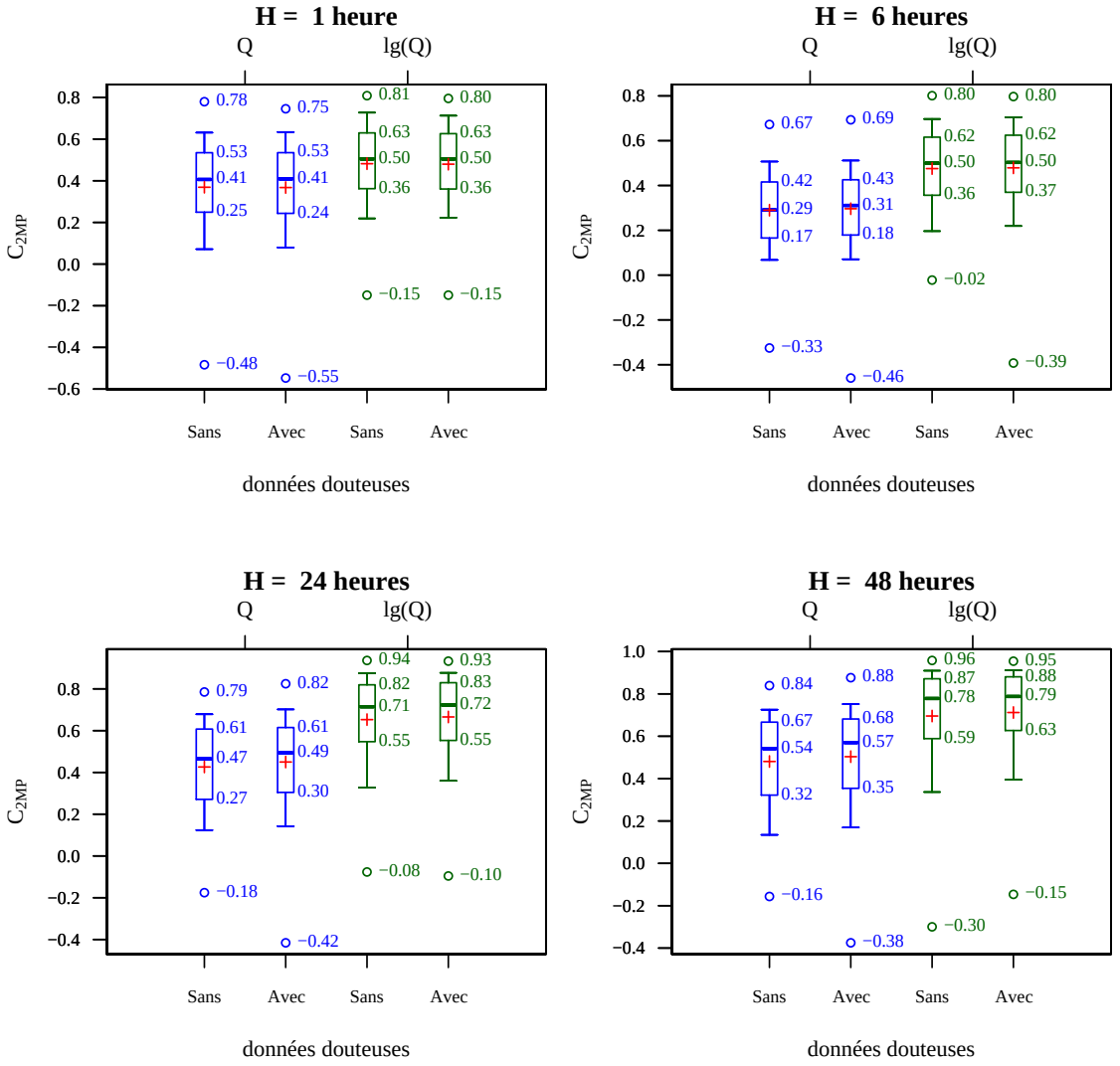


Figure H.3 – Distribution des critères de performance $C_{2M,P}$ obtenus par le modèle GRP sur 178 bassins versants × 2 périodes avec et sans la prise en compte des données douteuses. Deux critères de comparaison sont représentés : le critère $C_{2M,P}$ classique (en bleu) et celui calculé sur les transformations logarithmiques des débits (en vert).

capacités de calcul. Nous souhaitons donc savoir si ces bonnes performances ne sont pas dues au fait qu'elles sont calculées sur une gamme étroite de débit, qui de plus ne serait pas représentative des crues réelles.

Les distributions des performances obtenues avec et sans les données douteuses ne sont pas très différentes (la correction est légèrement meilleure quand on emploie les données douteuses) mais on observe des différences ponctuelles qui peuvent être assez fortes, sans pouvoir le relier à l'indice sup-seuil de manière convaincante (Fig. H.4).

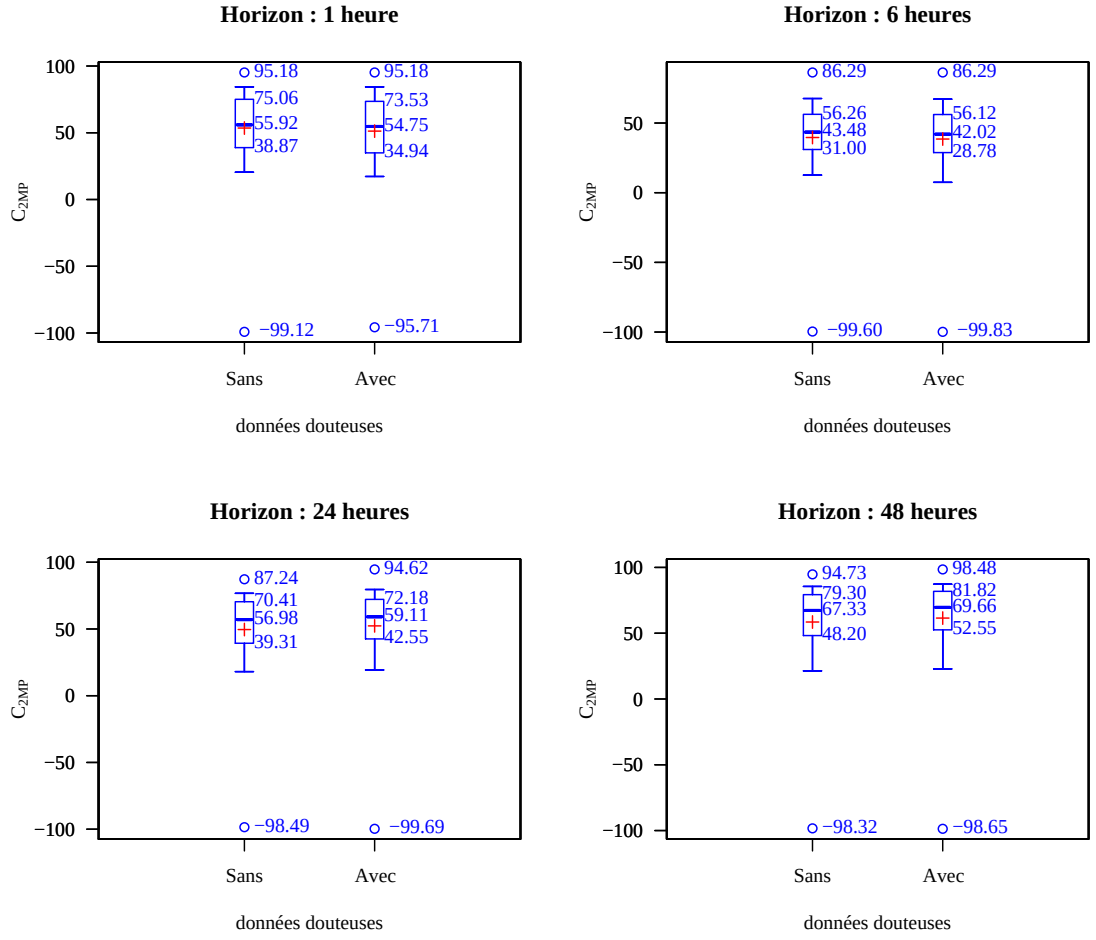


FIGURE H.4 – Distribution (sur 178 bassins versants × 2 périodes) des critères de performance $C_{2M,P}$ obtenus par le modèle GRP avec correction par réseau de neurones artificiels avec et sans la prise en compte des données douteuses.

H.3 « Conclusion »

Sur la base de ces deux comparaisons, nous estimons que ne pas employer les données douteuses lors de l'évaluation des différentes versions du modèle n'est pas un problème majeur.

I

Annexe

Classification ascendante hiérarchique et analyse en composantes principales

Cette annexe présente le principe de deux méthodes d'analyse classique de la variété d'un échantillon :
– la classification ascendante hiérarchique (C.A.H.),
– l'analyse en composantes principales (A.C.P.).

Dans les deux cas, nous considérons un échantillon d'objets (par exemple des bassins versants) caractérisés par n indices. On associe au i^{e} objet le vecteur $\mathbf{x}_i$ rassemblant les n valeurs des indices caractérisant cet objet $x_{i,1}, \ldots, x_{i,n}$.

I.1 Classification ascendante hiérarchique

La classification ascendante hiérarchique (C.A.H.) a pour objectif de définir des groupes « homogènes » au sein de l'échantillon, au sens des indices retenus pour décrire les éléments de l'échantillon.

Au départ, chaque entité de l'échantillon constitue une classe. Une distance d est calculée entre toutes les classes ; les deux classes les plus proches au sens de la distance d sont alors regroupées pour former une nouvelle classe. On recommence jusqu'à ce qu'il ne reste plus qu'une seule classe. On construit ainsi un dendrogramme (ou arbre de classification).

Plusieurs distances peuvent être utilisées. Nous avons choisi le critère de Ward qui définit la distance d entre deux classes C_1 et C_2 par

$$d\left(C_1, C_2\right) = \frac{n_1 \cdot n_2}{n_1 + n_2} \|\mathbf{m}_1 - \mathbf{m}_2\|_2^2$$

où n_i est l'effectif de la classe i et $\mathbf{m}_i$ son barycentre. Ce critère permet de limiter l'augmentation de l'inertie intra-classe, c'est-à-dire d'optimiser l'homogénéité des groupes construits par la C.A.H. au sens de la distance euclidienne entre deux éléments d'un groupe.

Pour mettre tous les indices sur un pied d'égalité, nous ne travaillons pas directement sur les indices, mais sur les indices centrés réduits :

$$x'_{i,j} = \frac{x_{i,j} - \mu_{x_{.,j}}}{\sigma_{x_{.,j}}}$$

où $\mu_{x_{.,j}}$ est la moyenne des valeurs prises par l'indice j sur l'échantillon et $\sigma_{x_{.,j}}$ est son écart-type.

En redescendant dans le dendrogramme de quelques niveaux, nous déterminons quelques grands groupes (4, 8 ...) homogènes nous permettant de donner une première description de l'échantillon. En descendant plus dans le dendrogramme, nous définissons des groupes d'effectifs réduits qui présentent tous une inertie intra-classe (i.e., une variété au sens des indices choisis) majorée grâce au critère de Ward.

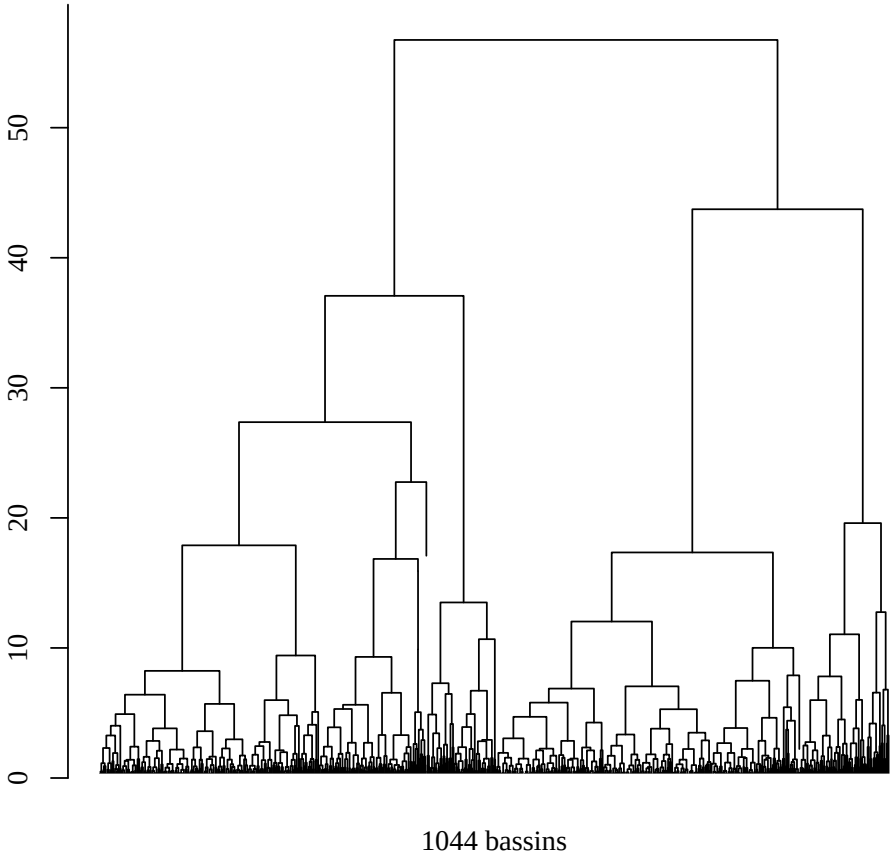


FIGURE I.1 – Exemple de dendrogramme. Celui représenté ici a été obtenu lors de la C.A.H. sur l'ensemble de 1 024 bassin versants caractérisés par les 6 indices décrivant le régime hydrographique (Chap. 4, page 423).

I.2 Analyse en composantes principales

L'analyse en composantes principales (A.C.P.) est une méthode complémentaire à la C.A.H. qui permet (entre autres choses) de représenter au mieux dans un espace de dimension réduite (idéalement 2) l'échantillon et les groupes définis lors de la C.A.H. Elle permet donc de caractériser ces groupes.

La méthode consiste à partir des n indices plus ou moins corrélés pour aboutir à n variables non-corrélées, combinaisons linéaires des précédents. Ces combinaisons linéaires sont choisies pour leur représentation de la variété de l'échantillon, c'est-à-dire qu'elles maximisent la variance de l'échantillon. Elles sont appelées *composantes principales*.

Les composantes principales sont obtenues en diagonalisant la matrice de variance-covariance des indices, c'est-à-dire en déterminant des axes combinaisons linéaires non corrélées (orthogonales) des indices : les vecteurs propres $\overrightarrow{V_1}, \ldots, \overrightarrow{V_n}$. La variance de l'échantillon sur chaque composante principale $\overrightarrow{V_i}$ est alors donnée par la valeur propre associée v_i. La somme des valeurs propres est égale à la variance totale. Ainsi, sélectionner un sous-espace de base $\left\{\overrightarrow{V_1}, \ldots, \overrightarrow{V_j}\right\}$ (avec $j \leq N$) permet de représenter la fraction de la variété de l'échantillon suivante :

$$\frac{\sum_{i=1}^{j} v_i}{\sum_{i=1}^{N} v_i}$$

Il reste alors à classer par ordre décroissant ces valeurs propres et les vecteurs propres associés pour définir les sous-espaces de dimension réduite (donc de représentation facile : le plan, l'espace 3D) qui permettent de présenter le maximum de la variété totale de l'échantillon.

Ici encore, pour que les variétés de tous les indices soient décrites avec la même importance, nous ne travaillons pas directement avec les indices mais avec les indices centrés réduits.

J

Annexe

Distribution de la fonction de production de GRP

Dans cette étude, la fonction de production de GRP est distribuée de façon à tenir compte de la variabilité spatiale des pluies sur le bassin versant. Nous étudions l'influence de cette distribution sur le flux entrant dans la fonction de routage du modèle puis sur la qualité des prévisions réalisées par le modèle GRP.

J.1 Principes

La fonction de production des modèles GR [1] est définie par un réservoir de production dont le niveau détermine, pour une pluie nette $P^{(n)}$ donnée, la quantité d'eau $P^{(s)}$ alimentant le réservoir de production et le flux $F = P^{(n)} - P^{(s)}$ dirigé vers la fonction de transfert. Le réservoir se vide également par une fonction de percolation (ce flux s'ajoute à celui se dirigeant vers la fonction de transfert, il est le plus souvent faible devant le premier lors d'averses).

Pour prendre en compte la distribution spatiale des pluies, nous proposons de définir un réservoir de profondeur A_i et de niveau $S_i(t)$ pour chaque maille, de calculer le flux F_i résultant de ce réservoir alimenté par la pluie nette de la maille P_i. L'ensemble des flux partiels F_i est rassemblé ensuite vers un flux unique

$$F_{Distribué} = \frac{\sum_{i=1}^{N} \alpha_i F_i}{\sum_{i=1}^{N} \alpha_i}$$

où N est le nombre de mailles et α_i est la fraction de la i^{e} maille appartenant au bassin versant. Ce flux est comparé au flux obtenu par la fonction de production globale (classique) qui prend en entrée la pluie nette

$$P_{Globale} = \frac{\sum_{i=1}^{N} \alpha_i P_i}{\sum_{i=1}^{N} \alpha_i}$$

Les figures J.1 et J.2 illustrent cette alternative.

Nous réalisons cette comparaison sur nos 178 bassins versants français. Dans un but de simplicité, nous supposons dans le cas distribué que tous les réservoirs de production ont la même profondeur A. Cette hypothèse peut paraître brutale en comparaison avec le modèle PDM (Moore, 2007) dont la fonction de production est constituée par un ensemble de réservoirs de capacités variables mais correspond parfaitement à notre modèle de prévision où la taille du réservoir de production est un paramètre fixe (indépendant du bassin).

1. ou tout au moins, de la plupart d'entre eux.

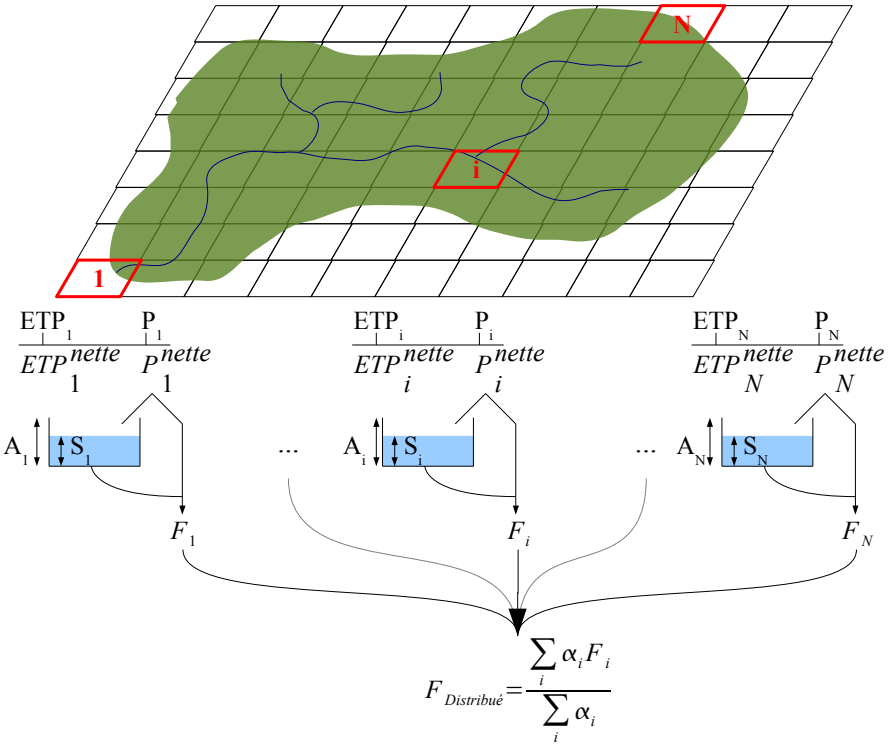


FIGURE J.1 – Fonction de production distribuée

J.2 Données de précipitation employées

J.2.1 Pluies SAFRAN

Les pluies SAFRAN sont données sur une maille carrée de 64 km^2. Elles sont obtenues au pas de temps journalier en supposant la pluie homogène au sein de chaque zone climatique Symposium ; cette pluie de zone Symposium est modulée selon l'altitude de la maille. La désagrégation temporelle (au sein d'une journée) est basée sur un modèle d'atmosphère.

J.2.2 Réseau de pluviomètres au pas de temps horaire

L'étude sera reproduite en employant cette fois les pluviomètres horaires de Météo France pour une analyse de sensibilité : on s'attend à ce que les données mesurées au sol soient plus hétérogènes que la pluie SAFRAN.

En effet, le réseau des pluviomètres horaires est beaucoup plus lâche que le réseau des pluviomètres journaliers : on en compte environ 700 sur le territoire métropolitain dont quelque 500 exploitables pour nos calculs (à comparer aux 3 000 à 4 000 postes journaliers). On peut supposer qu'utiliser les données provenant de ce réseau pour un bassin donné peut

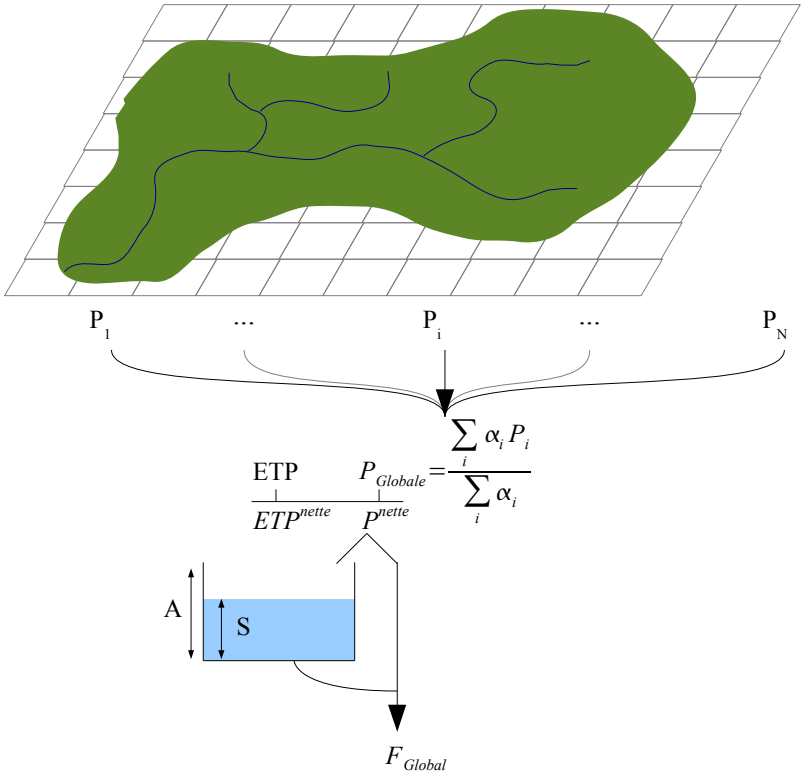


Figure J.2 – Fonction de production globale

correspondre à rechercher de l'information pluviométrique plus éloignée du bassin versant que l'emploi des pluies Safran si les zones Symposium sont globalement plus petites que les polygones de Thiessen. Les pluies Safran devraient donc être plus homogènes que les pluies du réseau de mesures au sol interpolées de façon grossière.

De plus, nous utilisons une interpolation de Thiessen pour calculer la précipitation sur chaque maille. Cette méthode, bien que maladroite (e.g. Sugarawa, 1993), a deux avantages pour cette étude de sensibilité : être aisée à mettre en œuvre et accentuer les hétérogénéités des données de précipitations. C'est une autre raison pour que les pluies issues d'une interpolation de mesures au sol soient plus hétérogènes que les pluies Safran.

J.3 Comparaison des flux en sortie de la fonction de production

J.3.1 Avec les pluies Safran

Pour chaque bassin versant, nous réalisons le calcul des deux flux sortant des fonctions de production sur la période 1995 – 2005. Les écarts absolus et relatifs (définis par l'écart

absolu divisé par la moyenne) sont analysés sur la période 1996 – 2005 (la première année servant de période de mise-en-route du modèle).

Très faible écart sur l'immense majorité des pas de temps

On observe un écart faible sur l'écrasante majorité des pas de temps. Les distributions (sur 178 bassins) des écarts entre les flux sont tracés sur la figure J.3. Les différences les plus grandes sont inférieures au millimètre sur quasiment tous les bassins. De plus, pour tous les bassins versants, la différence entre les sorties des deux fonctions de production est nettement inférieure au centième de millimètre (non significatif) sur l'immense majorité des pas de temps.

Écarts les plus forts

Les écarts les plus forts sont très ponctuels et se manifestent naturellement lors de fortes averses durant lesquelles on observe de fortes différences de pluies d'une maille à l'autre (La figure J.4 présente la différence entre les sorties des fonctions de production du modèle pour un événement assez fort sur le bassin de l'Allier à Langogne). Les pics de la différence des flux se manifestent lors d'événements forts où le réservoir de production se remplit brutalement, ce qui traduit la force de l'averse). Ces écarts maximaux sont de l'ordre du dixième de millimètre, ce qui est faible face à la quantité de pluie qui tombe lors de ces averses. On observe que l'hypothèse d'égalité du niveau des réservoirs entre mailles se vérifie souvent assez bien dans le cas des pluies SAFRAN.

Ces écarts les plus forts sont observés le plus souvent en début d'événements et sont systématiquement du même signe : le flux global est plus faible que le flux distribué sur ces écarts. Nous proposerons une explication à cette observation dans la prochaine section.

Un lien avec la taille du bassin versant ?

Nous n'avons pas trouvé de corrélation nette entre les écarts des sorties des fonctions de production globale et distribuée, et la taille du bassin versant : ce n'est pas parce qu'un bassin a une grande superficie qu'il est nécessaire de prendre en compte la variabilité des pluies dans notre modélisation Pluie – Débit. Inversement, de très petits bassins peuvent avoir des différences de flux assez importantes (relativement aux autres).

Cette absence de corrélation ne nous permet pas une conclusion définitive pour au moins deux raisons. La première tient à l'échantillon de bassins versants : malgré la sélection par C.A.H. permettant d'éliminer des bassins trop similaires, les bassins les plus réactifs (typiquement de l'arc méditerranéen) sont sous-représentés ; les bassins de tailles moyennes ou grandes de l'échantillon sont le plus souvent situés dans la partie Nord de la France. La deuxième raison est liée à notre fonction de production qui a été conçue dans un esprit de modélisation globale et non pour rendre compte de la production sur des mailles géométriques ne pouvant être définies comme des bassins versants.

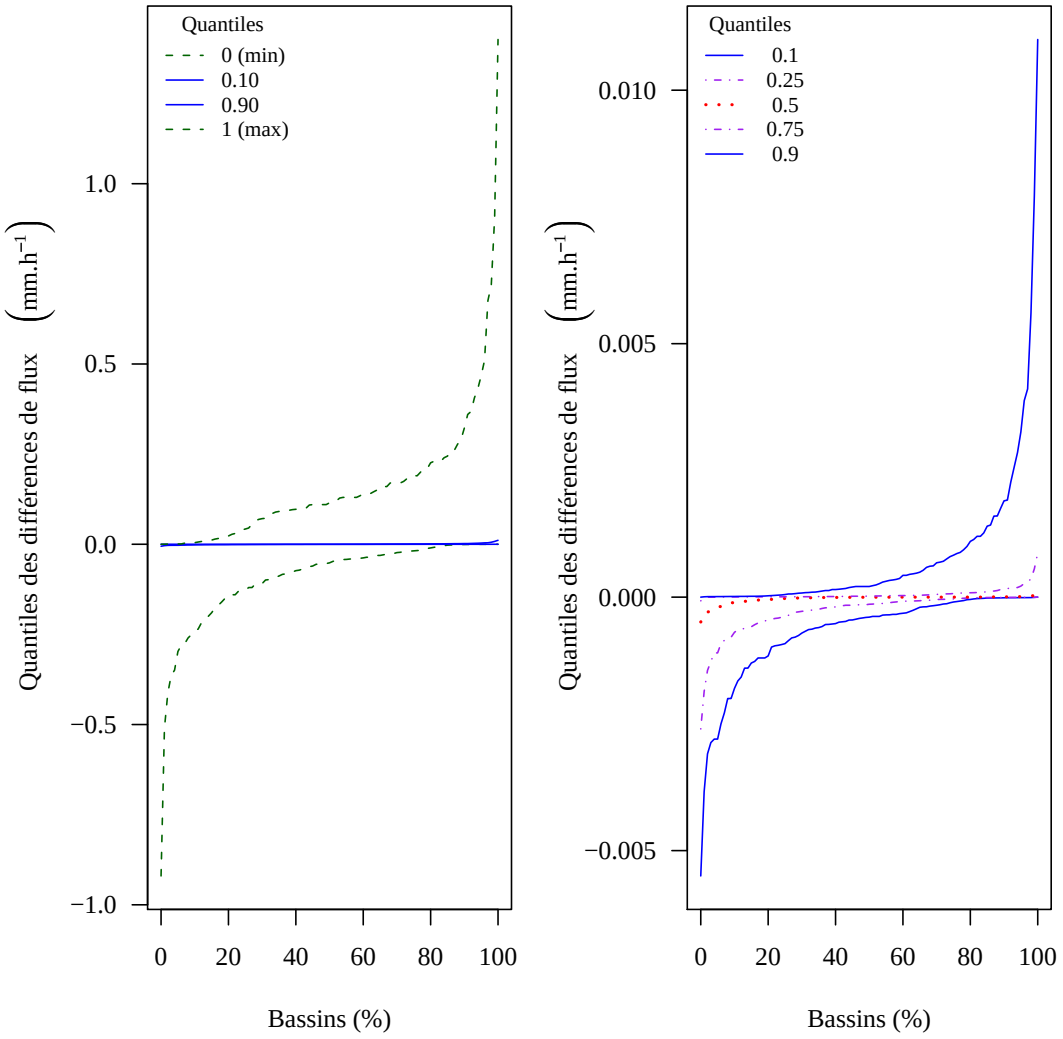


Figure J.3 – Distributions (sur 178 bassins versants) des différences entre les sorties des fonctions de production globale et distribuée. Le graphe de gauche présente les distributions des différences minimales et maximales enregistrées sur chaque bassin (durant 11 ans) et les distributions des quantiles 0.10 et 0.90 de ces différences. Le graphe de droite zoome et permet de voir les quantiles 0.25, 0.5 (distribution des différences médianes) et 0.75. Le nombre de pas de temps où la pluie nette est nulle varie entre 28 et 44 % sur l'ensemble de nos bassins ; on peut donc considérer que les courbes continues représentent environ les quantiles 0.17 et 0.83 des différences de flux quand ils sont calculés sur les pas de temps de pluie nette non nulle.

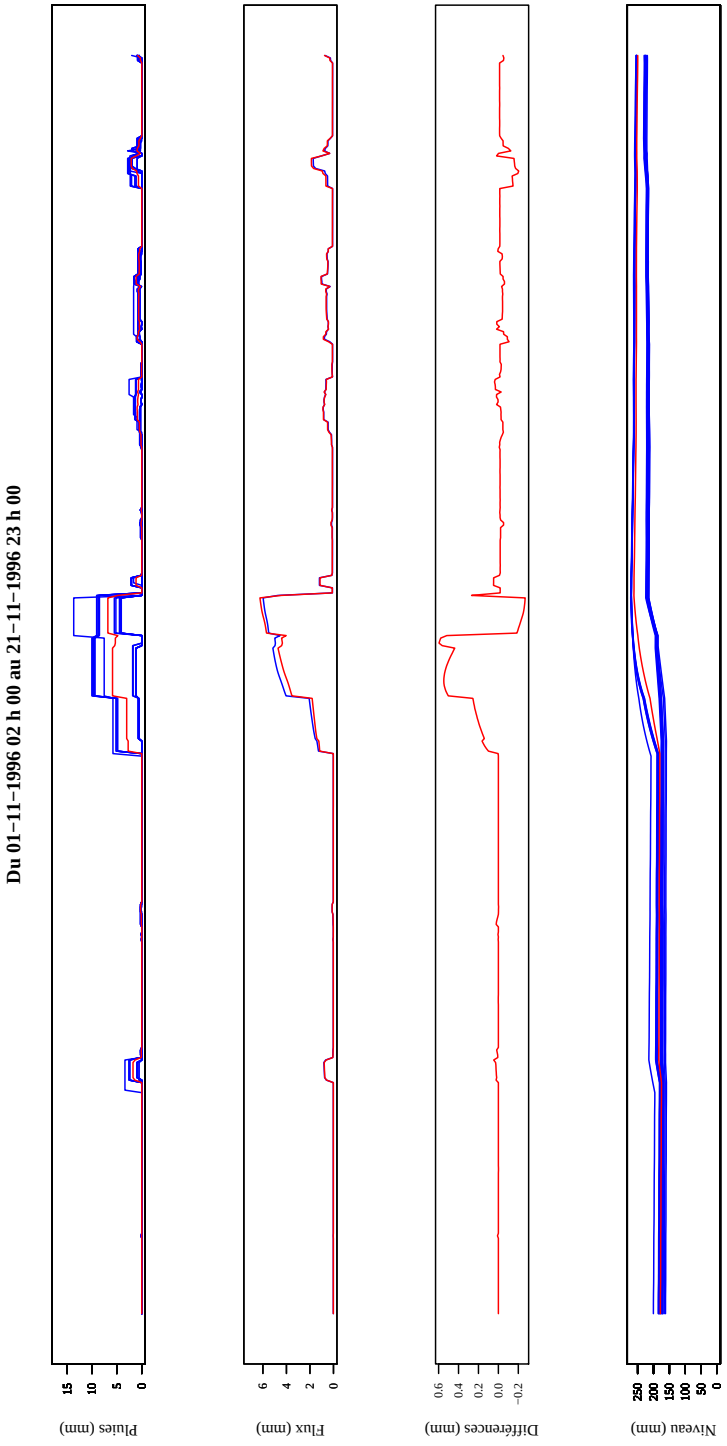


FIGURE J.4 – Chroniques sur un événement des différentes variables de la fonction de production pour l'Allier à Langogne (K2070810) qui compte avec le Gardon de Saint-Jean à Corbès (V7135010) parmi les bassins présentant les plus fortes différences de flux bien qu'ils soient tous deux de tailles assez modestes et ne s'étendent respectivement que sur 12 et 14 mailles. Le premier graphe présente les séries de pluies (les pluies de chaque maille sont tracées en bleu, la pluie globale, en rouge). Les chroniques des flux de production sont ensuite tracées (en bleu, par la fonction distribuée ; en rouge, par la fonction globale). Dans le troisième graphe, la différence entre les deux flux de production est représentée et enfin, la dernière fenêtre donne les variations du niveau du réservoir de production global.

On notera cependant que l'absence de corrélation de la nécessité de distribuer la pluie au sein de la fonction de production avec la superficie du bassin est corroborée par les résultats de Bourqui (2008). Cette absence de corrélation peut-être imputée (outre à la qualité de la donnée !), au rôle intégrateur des grands bassins versants.

La question de l'influence de l'hétérogénéité de la pluie sur les bassins sera examinée de façon indirecte dans la quatrième section.

J.3.2 Tests avec les pluies mesurées au sol

On observe en premier lieu que les pluies sont moins homogènes et conduisent à des différences entre flux globaux et flux distribués plus importantes que précédemment. Il reste cependant exact que les différences entre ces flux sont très faibles sur une grande majorité des pas de temps. Les écarts maximaux, intervenant sur des événements fortement pluvieux, peuvent atteindre quelques 5 mm, soit jusqu'à 10 % du flux.

Sur les graphes de chroniques étudiés (pour les mêmes bassins que pour les pluies Safran), on observe très logiquement que les niveaux des réservoirs sont plus dispersés et que les différences de flux (et leurs signes) sont moins prédictibles.

J.3.3 Une proposition d'explication

Nous avons observé que le flux global est le plus souvent plus faible que le flux distribué. Cela peut s'expliquer [2] par la convexité de la fonction de production $f : P^{(n)} \longmapsto F = P^{(n)} - P^{(s)}$: pour toute famille $\{\alpha_i\}_{i\in[\![1,N]\!]}$ de réels positifs de somme 1,

$$f\left(\sum_{i=1}^{N}\alpha_i x_i\right) \leq \sum_{i=1}^{N}\alpha_i f\left(x_i\right)$$

Si dans le cas de la fonction de production distribuée, tous les réservoirs de production ont des niveaux égaux (ce qu'on peut supposer en début d'épisode), l'équation précédente nous conduit directement à $F_{Global} \leq F_{Distribué}$ sur les premiers pas de temps de l'averse.

La fonction de production étant conservative, si le flux distribué est plus fort en début d'averse que le flux global alors le réservoir global se remplit plus vite que les réservoirs de maille de la description distribuée. Aussi, en fin d'averse, la percolation du mode global sera plus forte que celles du mode distribué. Si de nouvelles pluies tombent, alors le flux global peut devenir supérieur au flux distribué car le réservoir global acceptera moins d'eau. En conséquence, la première tendance peut s'inverser et nous obtiendrons des flux globaux légèrement plus élevés que les flux distribués.

2. En négligeant la percolation qui est le plus souvent faible devant F lors des événements pluvieux.

J.4 Influence sur les performances en prévision

Les flux de production calculés précédemment sont envoyés vers la fonction de transfert du modèle GRP, pour réaliser des prévisions aux horizons de 1, 6 et 24 heures. Nous comparons ici, les performances de ce modèle (en calage – contrôle).

J.4.1 Avec des pluies SAFRAN

Statistiques sur l'ensemble des 178 bassins

Nous n'observons sur nos 356 bassins-périodes (178 bassins × 2 périodes) qu'un seul cas où les performances sont significativement différentes, et ce pour le seul horizon d'une heure.

Ce bassin, l'Auzon à Mormoiron (V6125010), est un petit bassin (75 km^2) situé dans le Sud-Est de la France en amont d'un petit affluent du Rhône. On n'observe pas sur la chronique de prévision d'écart très élevé dans les prévisions (en tout cas pas plus élevé que pour les bassins explorés dans la partie suivante). Seuls quelques événements diffèrent légèrement en débits prévus.

Comme nous avons vu que le flux de production ne diffère que sur quelques événements de fortes pluies, nous avons examiné les critères de performances événementiels (persistance, écart sur les débits de pointe, retards sur les événements de crue). Nous n'avons pas plus de différence entre l'usage d'une fonction de production distribuée et celui d'une fonction globale que par l'évaluation par un critère de performance continu.

Illustrations sur des cas particuliers

Nous avons examiné plus en détail les chroniques de prévision pour trois bassins versants. Les deux premiers sont l'Allier à Langogne (324 km^2, K2070810) et le Gardon de Saint-Jean à Corbès (263 km^2, V7135010) qui ont les plus fortes différences de flux. Le troisième, la Somme à Péronne (1 290 km^2, E6351410) compte parmi les bassins présentant les écarts les plus faibles entre les sorties des fonctions de production globale et distribuée. Un exemple pour l'Allier à Langogne est donné en figure J.5.

Pour l'Allier et le Gardon de Saint-Jean, nous observons que l'impact de ces différences de flux est nul quasiment tout le temps (ce qui est logique, la différence de flux est différente de 0 pour quelques événements seulement). La fonction de transfert étale la réponse et donc la différence de flux de production : les différences de prévisions sont inférieures à 0.3 mm.h^{-1} sur l'événement le plus fort ; autant dire qu'elles sont négligeables. On observe que ces différences de prévisions diminuent avec l'accroissement de l'horizon de prévision, ainsi que l'apparition de « bruit » de plus en plus fort quand l'horizon de prévision s'éloigne.

Pour le bassin de la Somme, la différence sur les débits prévus ne dépasse pas les 10^{-3} mm.h^{-1}, c'est-à-dire en deçà de la limite de signification numérique, eu égard à la précision numérique de la donnée de pluie.

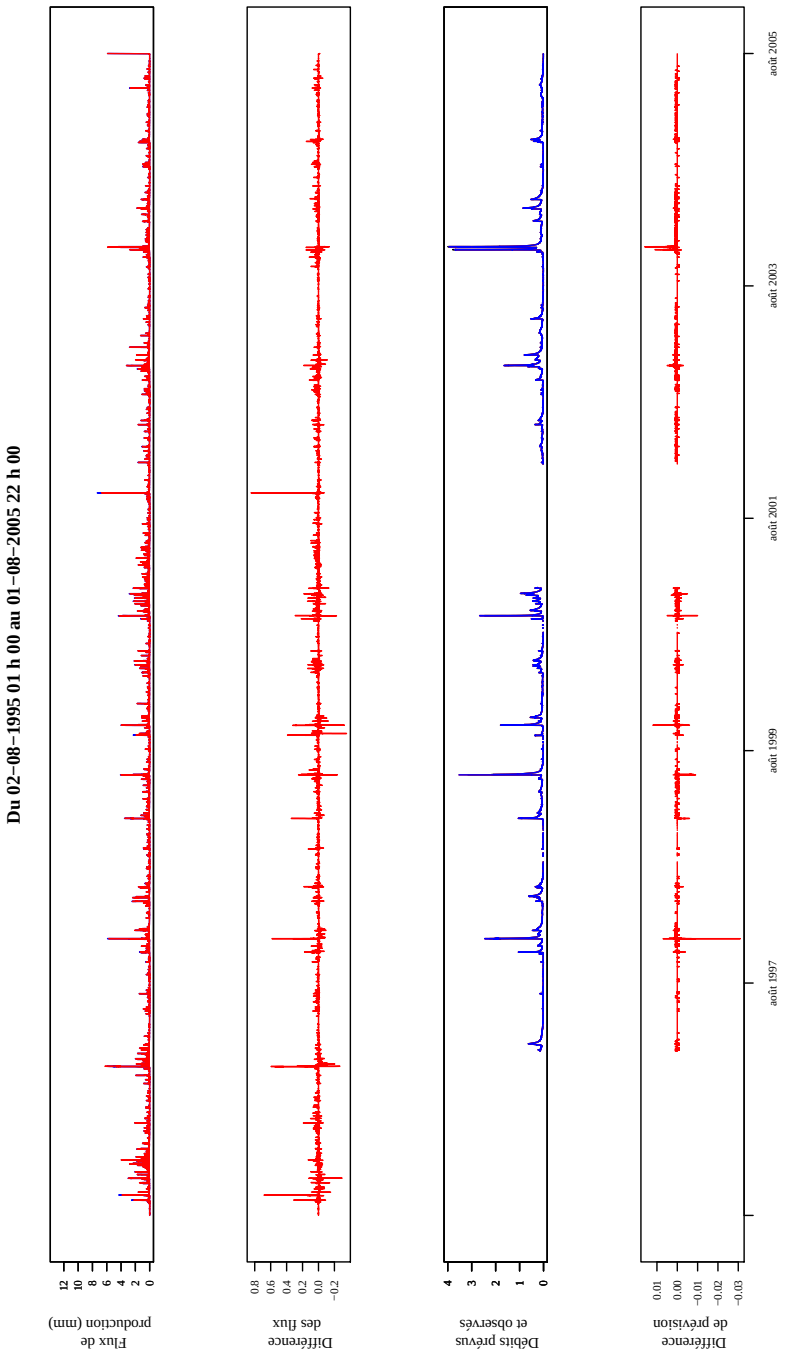


Figure J.5 – Prévisions à 6 heures sur l'Allier à Langogne. Les graphes représentent successivement les sorties des fonctions de production globale et distribuée (en mm), puis la différence (en mm) de ces sorties. Viennent ensuite les débits prévus et les différences entre débits prévus.

J.4.2 Avec des données issues des pluviométriques

Les différences de performances sont plus importantes que lors de l'expérience réalisée avec les pluies SAFRAN. Ces écarts restent minimes sur la très grande majorité des bassins, mais peuvent être importants sur quelques bassins. Les tableaux J.1 et J.2 donnent la distribution des différences de performances mesurées par $C_{2M,P}$ de façon continue puis au-dessus d'un seuil de débit. On observe que pour plus de 90 % des bassins versants, la différence de performance reste non significative quel que soit l'horizon de prévision. Parmi les 10 % des bassins sur lesquels la différence est significative, on compte une proportion plus que non négligeable de bassins où l'approche globale donne de meilleurs résultats que la distribution de la fonction de production.

Horizon	0	5 %	25 %	50 %	75 %	95 %	100 %
1 heure	-0,143	-0,006	-0,001	0,000	0,000	0,008	0,060
6 heures	-0,053	-0,001	-0,000	0,000	0,001	0,008	0,117
24 heures	-0,112	-0,021	-0,002	0,000	0,002	0,016	0,098

TABLE J.1 – Différence des performances mesurées par le critère $C_{2M,P}$. La troisième décimale, non significative, n'est donnée que pour illustrer la faiblesse des résultats.

Horizon	0	5 %	25 %	50 %	75 %	95 %	100 %
1 heure	-0,114	-0,014	-0,001	0,000	0,001	0,018	0,079
6 heures	-0,061	-0,021	-0,003	0,000	0,002	0,018	0,151
24 heures	-0,085	-0,028	-0,003	0,000	0,002	0,018	0,158

TABLE J.2 – Différence des performances mesurées par le critère $C_{2M,P}$ au dessus d'un débit seuil.

Nous avons relevé les bassins présentant des écarts de performance (en $C_{2M,P}$ au-dessus d'un seuil) supérieurs à 5 points. La première observation est qu'en dehors d'un bassin récurrent, les autres bassins peuvent présenter une forte différence de performance pour un horizon de prévision mais pas pour les autres. Parmi ces bassins, on trouve quelques petits bassins (U425210, V5445010, I3103010), des bassins de la Loire amont (K2010820, 49 km^2 et K2593010, 984 km^2), des bassins de la Garonne (O3011010, 67 km^2, 03141010, 925 km^2 et O3401010, 2170 km^2) et la quasi-totalité des bassins de l'arc méditerranéen (Y0325010, 31 km^2, Y0624020, 216 km^2, Y0655010, 305 km^2, Y2035010, 332 km^2 et Y02840601 729 km^2). S'intéresser à la distribution de la fonction de production se justifie principalement sur l'arc méditerranéen et sur les Cévennes (Fig. J.6).

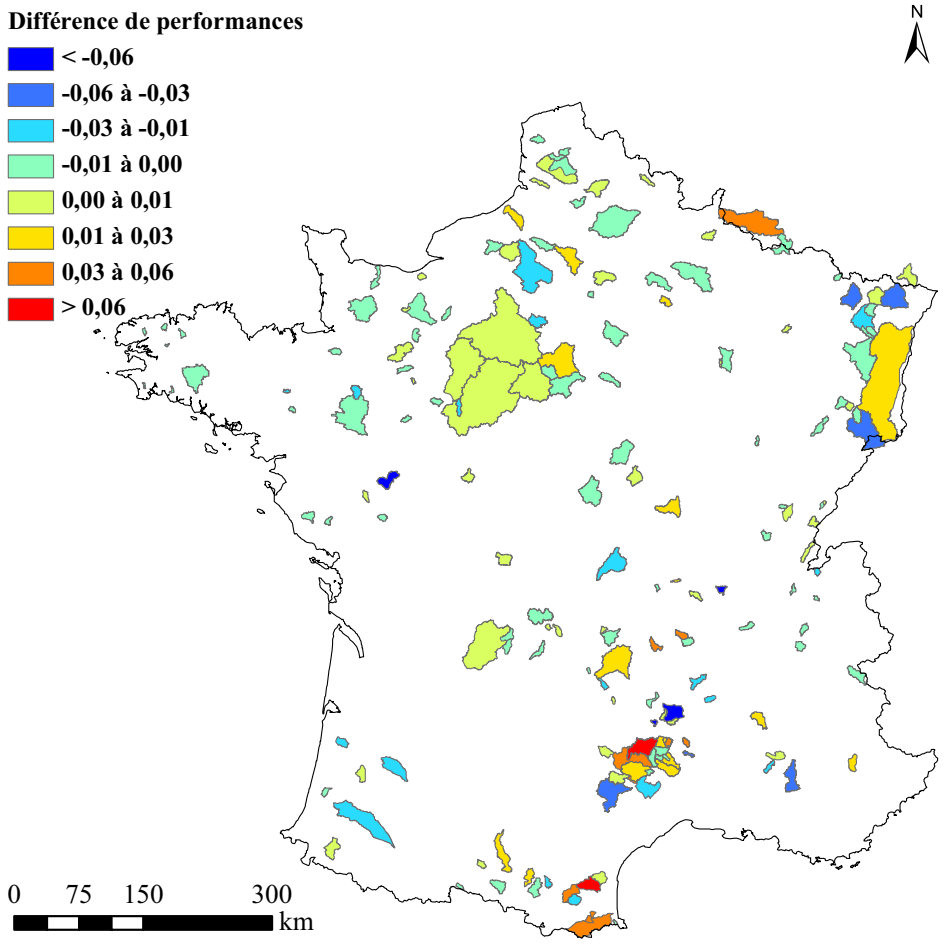


Figure J.6 – Maxima – sur les trois horizons de prévision et sur les deux périodes – des différences de performances (mesurées en $C_{2M,P}$) entre le modèle ayant une fonction de production distribuée et le modèle complètement global.

J.5 Conclusions

Nous pouvons conclure que la distribution de la fonction de production (seule) est strictement inutile pour rendre compte de la variabilité spatiale de la pluie dans notre modélisation Pluie – Débit, quand on utilise les pluies Safran.

Ces conclusions restent valables pour les pluies mesurées sur les pluviomètres, sauf pour quelques bassins, notamment ceux situés sur l'arc méditerranéen.

Enfin, les différences observées entre les deux données de pluies sont le plus souvent faibles et de peu de conséquence. Néanmoins, elles peuvent être significatives sur certains bassins. Cette observation met l'accent sur la nécessité de prendre en compte la nature de la donnée de pluie dans nos études. En outre, cela peut signifier (vue la faible densité du réseau

des pluviomètres horaires) que les conclusions de cette étude concernent plus les données employées que le modèle : si nous disposions d'une carte exacte de la pluie, les conclusions seraient peut-être à nuancer. Cette remarque « de précaution » est tout de même amoindrie par plusieurs observations :

- On retrouve l'importance relative de la distribution de la pluie sur l'arc méditerranéen.
- Si on considère des pluies stratiformes (plus grande partie de la France), le fait que le réseau pluviométrique soit lâche fait que nos données par interpolations de Thiessen sont sans doute plus hétérogènes que la pluie réelle. Ceci n'est pas exact pour des cellules pluvieuses convectives.

K

Annexe

Compléments au chapitre 6

Soit x un état du modèle. Dans toute cette annexe, les notations distinguerons les états calculés par le modèle au temps t (à partir d'états antérieurs éventuellement mis-à-jour) x_t des états mis-à-jour grâce à une observation faite au pas de temps t, $x_{t|t}$.

K.1 Inversion de la fonction de vidange du réservoir de routage

Les réservoirs de type « Puissance » sont définis par

$$dR = -k \cdot R^{\alpha} dt \quad \text{avec } \alpha \neq 1$$

Le débit discrétisé est donné par l'équation (C.2) en annexe C.1 (page 340) :

$$Q_t = f_{X_1,\alpha}\left(R_t^+\right) = R_t^+ \cdot \left(1 - \frac{1}{\left(1 + \left(\frac{R_t^+}{X_1}\right)^{\alpha-1}\right)^{\frac{1}{\alpha-1}}}\right) \tag{K.1}$$

où R_t^+ est le niveau du réservoir avant sa vidange (au pas de temps t) et X_1 sa capacité. Le cas $R_t^+ > X_1$ est autorisé mais le volume R_t après vidange est inférieur ou égal à la capacité du réservoir X_1.

L'inversion analytique est évidente dans le cas quadratique. Elle reste possible (mais plus pénible) dans le cas $\alpha = 3$. Elle est impossible (à ma connaissance) dans les autres cas. Il est cependant tout-à-fait possible de tabuler la fonction f pour réaliser une inversion numérique. Il est plus pratique de réécrire l'équation (K.1) en transformant f de façon à ne plus avoir à se préoccuper du paramètre X_1.

Soient x_t le niveau relatif du réservoir de routage et y_t, le débit relatif en sortie :

$$x_t = \frac{R_t^+}{X_1} \in [0, x_{max}] \quad \text{et} \quad y_t = \frac{Q_t}{X_1}$$

Il vient immédiatement $y_t = g_\alpha\left(x_t\right)$ avec

$$g_\alpha : x \mapsto x \cdot \left(1 - \frac{1}{\left(1 + x^{\alpha-1}\right)^{\frac{1}{\alpha-1}}}\right)$$

Il suffit donc de tabuler g_α sur le segment $[0, x_{max}]$ pour pouvoir inverser la fonction de vidange. x_{max} ne peut pas être très supérieur à 1 (vu que R_{t-1} – après vidange – ne peut

être pas être supérieur à la capacité du réservoir X_1). Dans la pratique au pas de temps horaire, on peut choisir $x_{max} = 1.5$.

K.2 Mise-à-jour de la pluie efficace PRT_t

Nous détaillons ici les calculs des essais de mise-à-jour de la pluie efficace (partie 7.3.2, page 115).

Connaissant la dernière observation de débit Q_t, on sait mettre à jour le niveau du réservoir de routage $R^+_{t|t} = f_\alpha^{-1}(Q_t)$ où f_α est la fonction de vidange du réservoir de routage (réservoir de type « Puissance » α). L'exposant $+$ indique le niveau du réservoir avant vidange.

K.2.1 Effets d'une mise-à-jour de la pluie efficace PRT_{t-i} sur les niveaux à venir du réservoir de routage

Notre but est de déduire la mise-à-jour à appliquer à un état PRT_{t-i} (pour $i \in [\![0, X_3 - 1]\!]$).

Définissons les pertubations (i.e. mises-à-jour) suivantes :
- la pertubation sur le niveau du réservoir de routage observé au pas de temps t avant vidange : $\delta R^+_t = R^+_{t|t} - R^+_t$,
- la pertubation sur le k^{e} compartiment de l'hydrogramme unitaire à la date t : $\delta HU_t(k) = HU_{t|t}(k) - HU_t(k)$. Cette perturbation va se propager jusqu'au 1er compartiment de l'hydrogramme unitaire à la date $t + k - 1$: $\delta HU_{t+k-1}(1) = \delta HU_{t|t}(k)$.
- la pertubation sur l'état PRT à la date $t - i$: $\delta PRT_{t-i} = PRT_{t-i|t} - PRT_{t-i}$.

En remontant le cheminement de l'eau dans le modèle, nous avons :

$$\begin{cases} R^+_t &= R_{t-1} + HU_t(1) \\ HU_t(1) &= \sum_{i=1}^{E(X_3)} \alpha_i \cdot PRT_{t-i+1} \end{cases}$$

où $HU_\tau(n)$ est le niveau du n^{e} compartiment de l'hydrogramme unitaire au pas de temps τ et α_i est le coefficient associé à ce compartiment (la somme des coefficients valant 1).

Si on veut mettre à jour un seul des états PRT_{t-i}, il suffit de prolonger la mise-à-jour $R^+_{t|t}$:

$$\begin{cases} R^+_{t|t} &= R_{t-1} + HU_{t|t}(1) \\ HU_{t|t}(1) &= \sum_{\substack{k=1 \\ k \neq i+1}}^{E(X_3)} \alpha_k \cdot PRT_{t-k+1} + \alpha_{i+1} PRT_{t-i|t} \end{cases}$$

On en déduit :

$$\delta R_t^+ = \delta HU_t(1) = \alpha_{i+1} \cdot \delta PRT_{t-i} \quad \text{pour } 0 \leq i \leq E(X_3) - 1 \tag{K.2}$$

Il nous faut ensuite répercuter cette mise-à-jour sur tous les niveaux de l'hydrogramme unitaire à la date t [1] :

$$HU_{t|t}(k) = HU_t(k) + \alpha_{i+k} \cdot \delta PRT_{t-i} \quad \text{pour } 1 \leq i + k \leq X_3$$

soit

$$\delta HU_t(k) = \frac{\alpha_{i+k}}{\alpha_{i+1}} \cdot \delta R_t^+$$

Nous sommes intéressés par l'effet de la mise-à-jour sur le pas de temps $t + H$ où H est l'horizon de prévision. Le débit sera d'abord modifié par les modifications de R_t à R_{t+H} liées à la rémanence des apports au réservoir de routage entre t et $t + H - 1$:

$$\delta R_{t+k}^+ = \begin{cases} \delta HU_{t+k}(1) = \delta HU_t(k+1) = \dfrac{\alpha_{i+k+1}}{\alpha_{i+1}} \cdot \delta R_t^+ & \text{si } \ 0 \leq i \leq E(X_3) - 1 - k \\ 0 & \text{sinon.} \end{cases}$$

Le débit sera enfin directement modifié par l'apport à la date $t + H$:

$$\delta R_{t+H}^+ = \begin{cases} \dfrac{\alpha_{i+H+1}}{\alpha_{i+1}} \cdot \delta R_t^+ & \text{si } \ 0 \leq i \leq E(X_3) - H - 1 \\ 0 & \text{sinon.} \end{cases}$$

K.2.2 Conséquences de cette mise-à-jour

Examinons le rapport $\beta_{i,k}$ (où i dénote le choix de l'état PRT_{t-i} mis-à-jour) pour k compris entre 1 et l'hrorizon de prévision H :

$$\beta_{i,k} = \begin{cases} \dfrac{\alpha_{i+k+1}}{\alpha_{i+1}} & \text{si } 0 \leq i \leq E(X_3) - k - 1 \\ 0 & \text{sinon} \end{cases}$$

1. On atteint cette formule en appliquant successivement à l'équation (K.2) les transformations $t \mapsto t + k - 1$ et $i \mapsto i + k - 1$, puis en laissant la mise-à-jour se propager dans le temps :
$\delta HUt + k - 1(1) = \alpha_{i+1} \cdot \delta PRT_{t+k-i-1}$ pour $0 \leq i \leq E(X_3) - 1$
$\delta HUt + k - 1(1) = \alpha_{i+k} \cdot \delta PRT_{t-i}$ pour $1 \leq i + k \leq E(X_3)$
$\delta HUt(k) = \alpha_{i+k} \cdot \delta PRT_{t-i}$ pour $1 \leq i + k \leq E(X_3)$

On peut également obtenir cette équation en repartant de $t - i$, puis propager la mise-à-jour dans le temps et employer la transformation $k - i \mapsto k$:
$\delta HU_{t-i}(k) = \alpha_k \delta PRT_{t-i}$ pour $1 \leq k \leq E(X_3)$
$\delta HU_t(k - i) = \alpha_k \delta PRT_{t-i}$ pour $1 \leq k - i \leq E(X_3) - i$
$\delta HU_t(k) = \alpha_{k+i} \delta PRT_{t-i}$ pour $1 \leq k \leq E(X_3) - i$

Quelle que soit la valeur de k (pas de temps futur), plus on choisit i grand (i.e. plus on fait reporter la mise-à-jour loin dans le passé), plus $\beta_{i,k}$ sera faible (cf. fig. K.1). Aussi, plus i sera grand, moins l'information contenue dans δR_t^+ sera transférée dans δR_{t+k}^+. Cependant, pour i trop petit devant X_3, on obtient des valeurs de $\beta_{i,k}$ significativement supérieures à 1, ce qui peut amoindrir la robustesse de la mise-à-jour face à des erreurs d'observations.

L'hydrogramme unitaire étant symétrique, on peut remarquer que $\beta_{i,k}$ sera égal ou proche

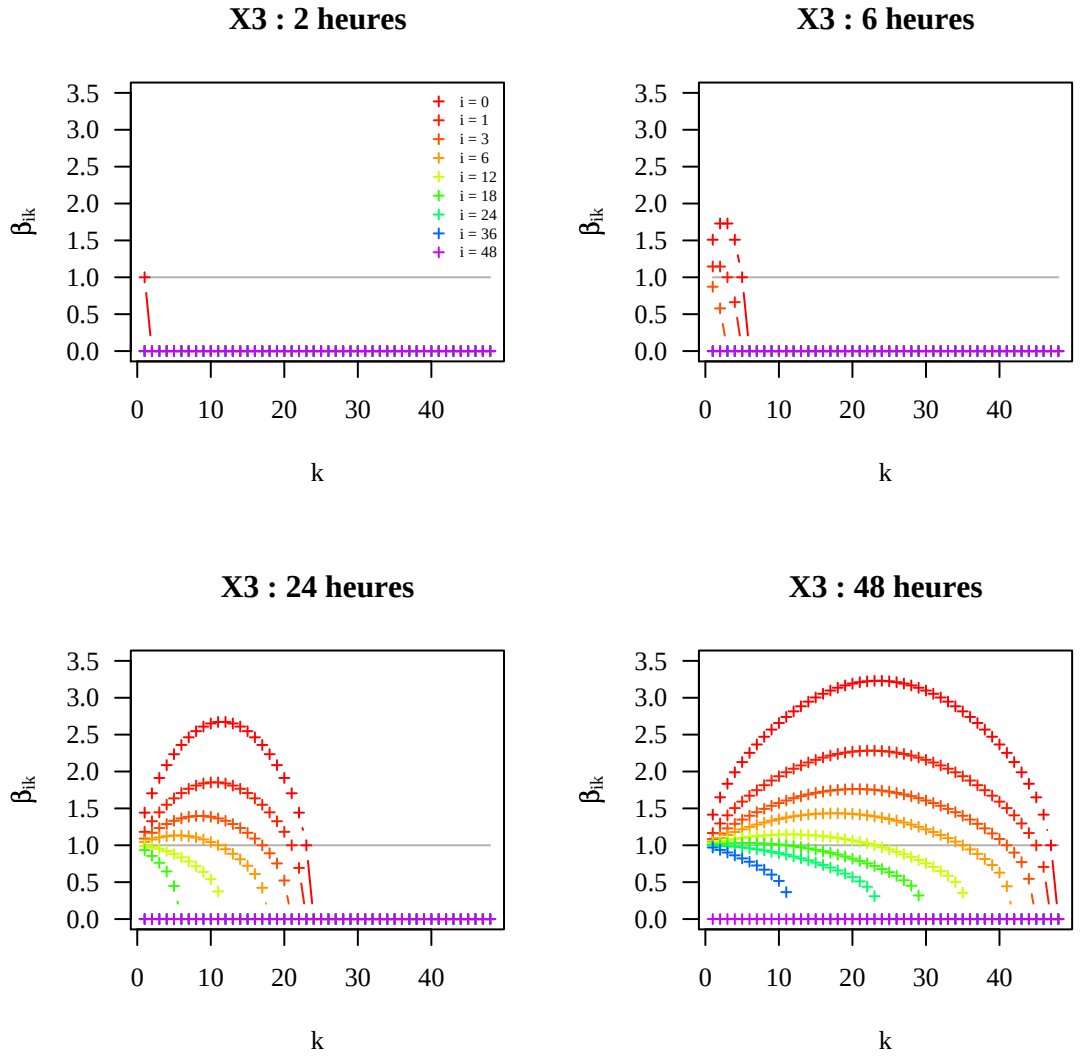


FIGURE K.1 – Effets d'une mise-à-jour sur une pluie efficace PRT_{t-i} sur les niveaux du réservoir de routage entre $t+1$ et $t+H$: $\beta_{i,k}$ en fonction de k pour différentes valeurs de i et de X_3.

de 1 si $\alpha_{i+k+1} = \alpha_{i+1} = \alpha_{X_3-i-1}$ soit pour

$$i = \frac{\lfloor X_3 \rfloor - k}{2} - 1$$

Enfin, on peut remarquer que pour chaque pas de temps $t+k$, si on modifie plusieurs états PRT_{t-i}, ces modifications se combinent linéairement pour donner δR_{t+k}^{Av}. Mais bien sûr les coefficients linéaires ne sont pas les mêmes entre deux pas de temps $t+k$ et la fonction de vidange n'est pas linéaire : l'effet complet n'est donc pas linéaire.

K.3 Représentation graphique du fonctionnement de GRP

La figure K.2 représente graphiquement le fonctionnement des principaux éléments de GRP en fonction de leurs entrées. L'opérateur d'observation h reliant le niveau du réservoir de routage R_t et le débit Q_t est également tracé. Tous les états sont pris à la fin de chaque pas de temps. Ces représentations permettent de déterminer le degré de non linéarité du modèle et les erreurs à attendre d'une linéarisation dans le cadre du filtre de Kálmán étendu.

K.4 Implémentation du filtre de Kálmán étendu appliqué à la fonction de routage de GRP

K.4.1 Fonctionnement du filtre de Kálmán étendu

Nous présentons ici l'implémentation du filtre de Kálmán étendu. Le filtre fonctionne de façon séquentielle : chaque itération compte une phase de propagation (prévision par le modèle) et une phase de mise-à-jour. Nous employons dans cette partie les notations classiques pour le filtre de Kálmán (répertoriées dans le tableau 7.4, page 119). Le fonctionnement du cycle est décrit par les équations de propagation (7.2) et d'observation (7.1) :

$$\begin{cases} \mathbf{x}_t = f\left(\mathbf{x}_{t-1}, \mathbf{e}_t, \mathbf{u}_t, \mathbf{w}_t\right) & \text{équation d'état} \\ \mathbf{z}_t = h\left(\mathbf{x}_t, \mathbf{v}_t\right) & \text{équation d'observation} \end{cases}$$

linéarisées pour donner :

$$\begin{cases} \mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{A}\mathbf{e}_t + \mathbf{u}_t + \mathbf{w}_t \\ \mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \end{cases}$$

Mise-à-jour des états grâce à l'observation $\mathbf{z}_t$

(a) Calcul de la matrice $\mathbf{H}_t$, matrice jacobienne de h calculée en $\mathbf{x}_{t|t-1}$.

(b) Calcul du gain $\mathbf{K}_t = \mathbf{P}_{t|t-1}{}^t\mathbf{H}_t \left(\mathbf{H}_t\mathbf{P}_{t|t-1}{}^t\mathbf{H}_t + \mathbf{R}_t\right)^{-1}$

(c) Mise-à-jour des états $\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t\left(\mathbf{z}_t - h\left(\mathbf{x}_{t|t-1}\right)\right)$

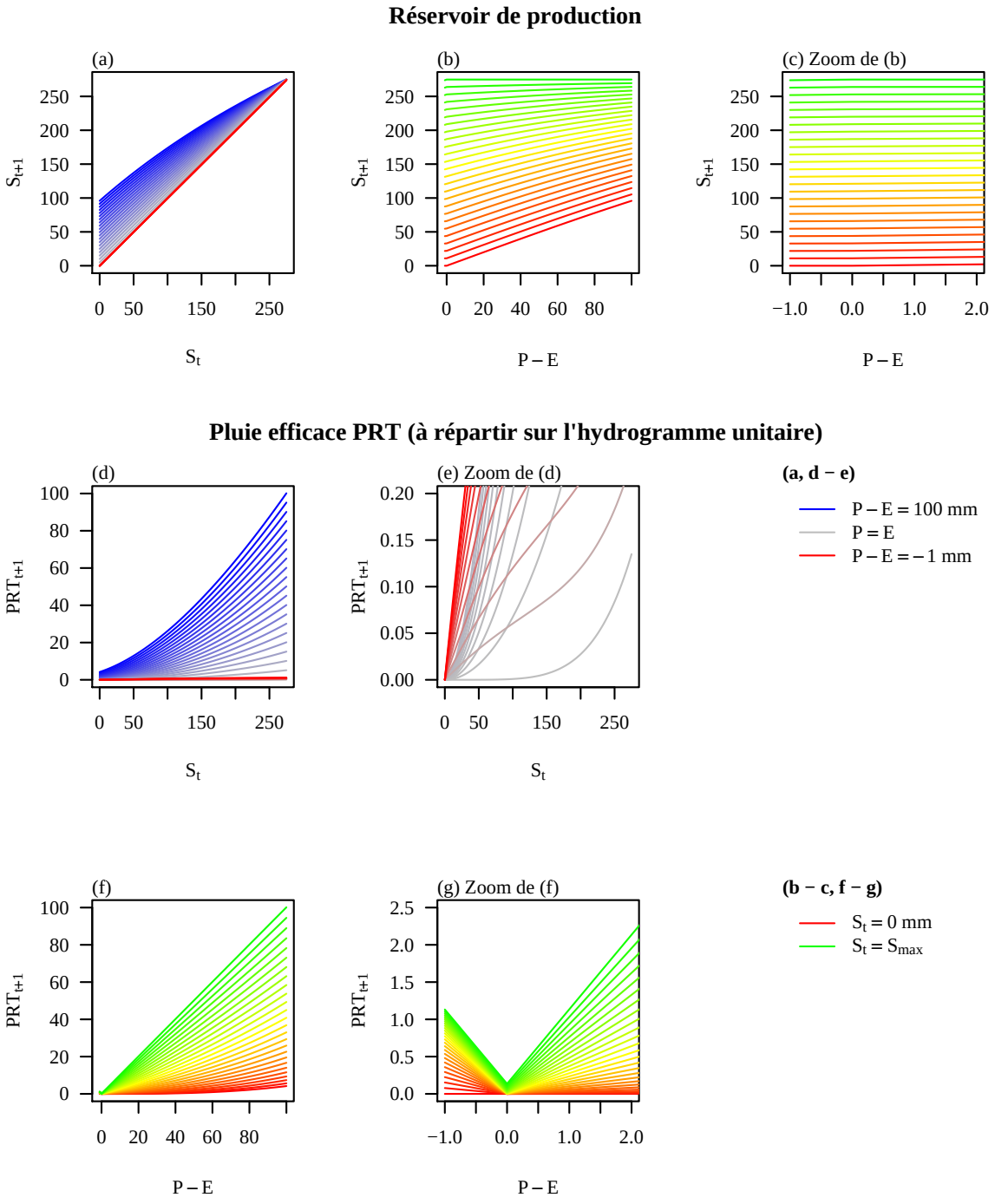


FIGURE K.2 – Fonctionnement des principaux éléments de GRP. Niveau du réservoir de production en fonction de sa valeur au pas de temps précédent (a), et de la pluie nette $P - E$ (b –c). Pluie efficace en sortie de la fonction de production en fonction du niveau du réservoir de routage (d –e) et de la pluie nette $P - E$ (f – g).

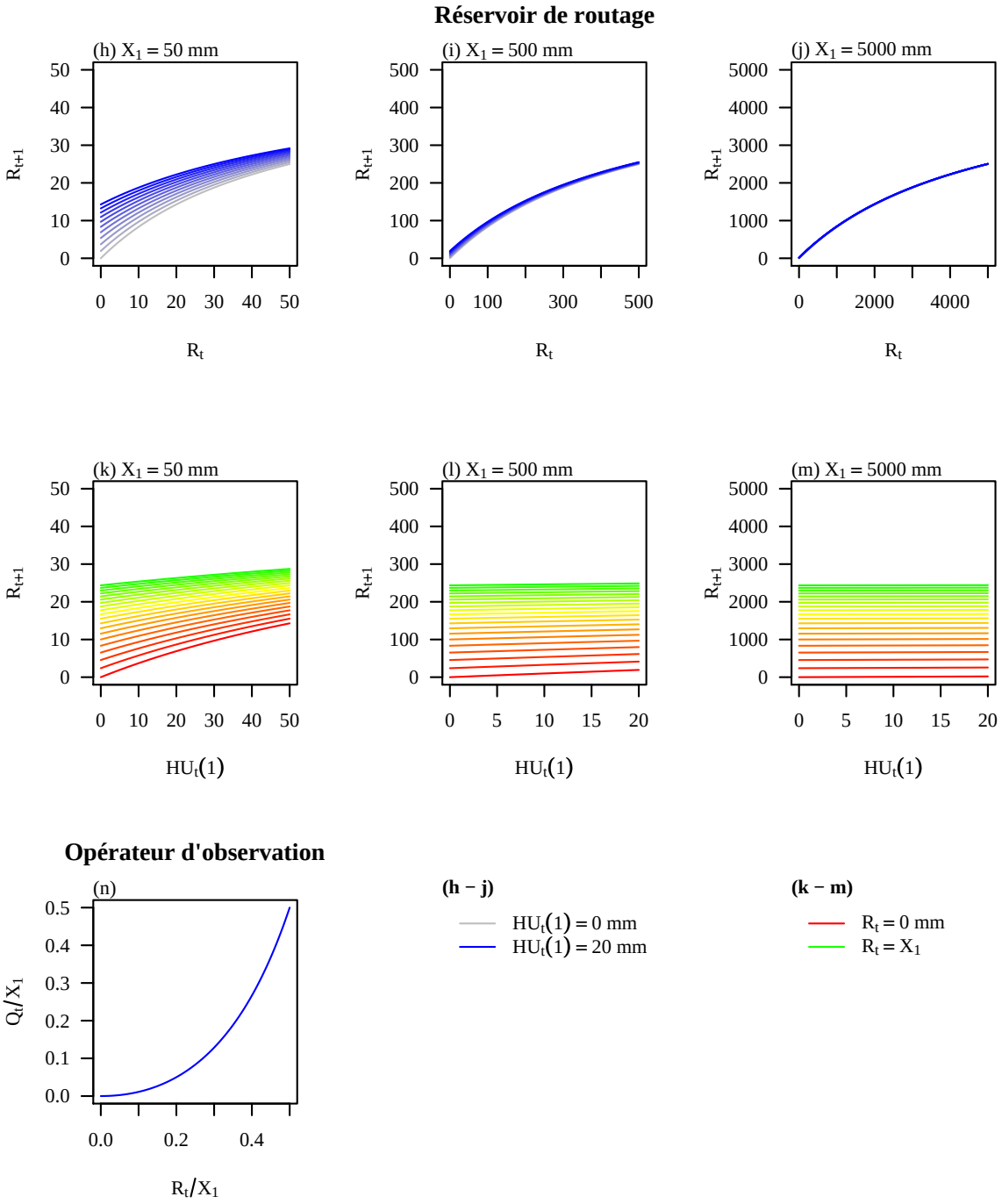


Fig. K.2: suite – Fonctionnement des principaux éléments de GRP. Niveau du réservoir de production en fonction de sa valeur au pas de temps précédent (h – j) et de son entrée, le stock d'eau en attente dans le premier compartiment de l'hydrogramme unitaire (k – l). Opérateur d'observation reliant le niveau du réservoir de routage à la fin du pas de temps au débit écoulé au même pas de temps (n).

(d) Mise-à-jour de la matrice de variance-covariance des erreurs sur les états
$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\,\mathbf{P}_{t|t-1}$

Passage du pas de temps t au pas de temps $t+1$

(e) Passage du modèle $\mathbf{x}_{t+1|t} = f\left(\mathbf{x}_{t|t}\right)$

(f) Calcul de $\mathbf{F}_{t+1}$, matrice jacobienne de f calculée en $\mathbf{x}_{t|t}$ et de la matrice de prise en compte des observations $\mathbf{A}_t$.

(g) Propagation des erreurs $\mathbf{P}_{t+1|t} = \mathbf{F}_{t+1}\mathbf{P}_{t|t}{}^t\mathbf{F}_{t+1} + \mathbf{A}_{t+1}\mathbf{S}_t{}^t\mathbf{A}_{t+1} + \mathbf{Q}_t$

K.4.2 Application à la fonction de routage de GRP

Le filtre de Kálmán étendu est appliqué à la fonction de routage du modèle GRP qui est composée d'un hydrogramme unitaire et d'un réservoir quadratique. Nous noterons (Fig. K.3) :

η	le nombre de compartiments de l'hydrogramme unitaire,
$HU_t(i)$	le niveau de l'eau en attente dans le i^{e} compartiment de l'hydrogramme unitaire à la fin du pas de temps t (i de 1 à η),
PRT_t	la pluie efficace entrant dans l'hydrogramme unitaire,
R_t^+	le niveau du réservoir de routage pendant le pas de temps t, après la verse du premier compartiment de l'hydrogramme unitaire mais avant la vidange du réservoir de routage,
R_t	le niveau du réservoir de routage à la fin du pas de temps t,
$\widehat{Q}_t$	le débit calculé par le modèle, sortant au pas de temps t.

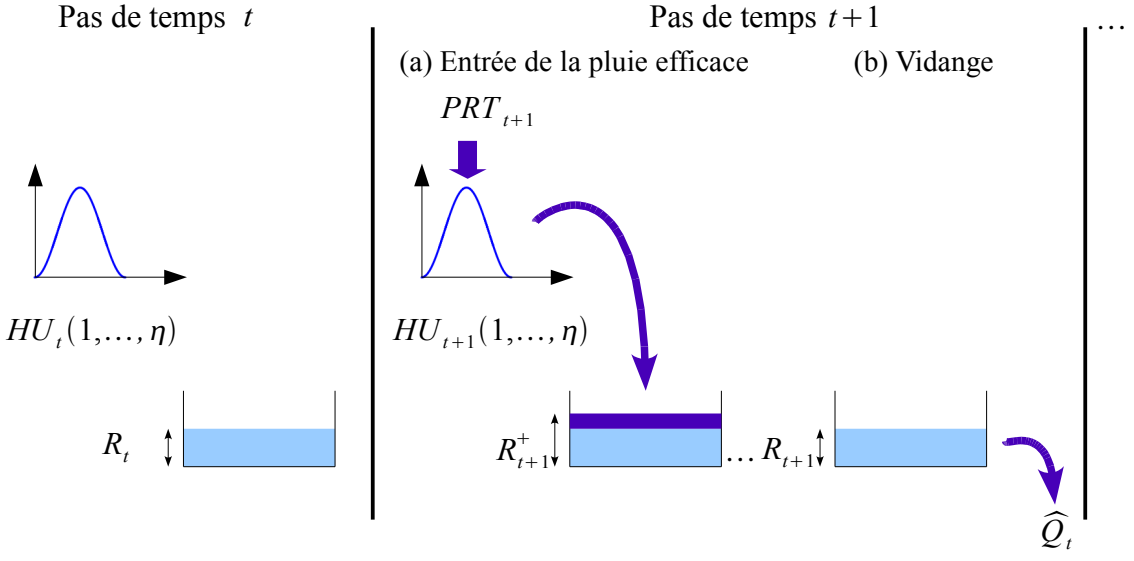


FIGURE K.3 – Fonction de routage du modèle GRP

Équations de la fonction de routage

Le passage du pas de temps t au pas de temps $t+1$ se fait en trois étapes :

1. Distribution de la pluie efficace dans l'hydrogramme unitaire :

$$\begin{cases} \forall i \in [\![1, \eta - 1]\!], HU_{t+1}(i) = HU_t(i+1) + \gamma_i PRT_{t+1} \\ HU_{t+1}(\eta) = \gamma_\eta PRT_{t+1} \end{cases}$$

   où les coefficients γ_i sont les ordonnées de l'hydrogramme unitaire.

2. Le dernier compartiment de l'hydrogramme unitaire se vide dans le réservoir de routage :

$$R^+_{t+1} = R_t + HU_{t+1}(1) = R_t + HU_t(2) + \gamma_1 PRT_{t+1}$$

3. Vidange du réservoir de routage pour donner le débit calculé :

$$\begin{cases} \widehat{Q}_{t+1} = \dfrac{\left(R^+_{t+1}\right)^2}{R^+_{t+1} + X_1} \\ R_{t+1} = R^+_{t+1} - \widehat{Q}_{t+1} \end{cases}$$

   où X_1 est la capacité du réservoir de routage.

Vecteur des états

Le routage du modèle GRP compte $\eta + 1$ états :
– les η stocks d'eau en attente dans l'hydrogramme unitaire,
– le niveau dans le réservoir de routage R_t.
Nous rajoutons un état fantôme q_t pour prendre en compte l'existence d'erreurs systématiques sur les débits. Ces $\eta + 2$ états sont rassemblés dans le vecteur des états :

$$\mathbf{x}_t = {}^t\left(HU_t(1), \ldots, HU_t(\eta), R_t, q_t\right)$$

Tous les états sont pris à la fin du pas de temps.

Linéarisation du modèle

Seule la dernière étape du routage est non linéaire. La matrice jacobienne de f,

$\mathbf{F}_{t+1,[i,j]} = \dfrac{\partial f_{t+1}}{\partial x_{[i]}}\left(\widehat{x}_{t,[j]}\right)$ s'écrit

$$\mathbf{F}_{t+1} = \left[\begin{array}{cccccccc|cc}
0 & 1 & 0 & \cdots & & & \cdots & 0 & 0 & 0 \\
\vdots & 0 & 1 & 0 & & & & \vdots & \vdots & \vdots \\
 & & \ddots & \ddots & \ddots & & & & & \\
 & & & & & & & & & \\
 & & & & & & & & & \\
 & & & & \ddots & \ddots & \ddots & \vdots & & \\
 & & & & & 0 & 1 & 0 & & \\
\vdots & & & & & & 0 & 1 & \vdots & \vdots \\
0 & \cdots & & & & & \cdots & 0 & 0 & 0 \\
\hline
0 & \alpha_{t+1} & 0 & \cdots & & & \cdots & 0 & \alpha_{t+1} & 0 \\
0 & \cdots & & & & & \cdots & 0 & 0 & 1
\end{array}\right]$$

La première partie $(\eta \times \eta)$ de la matrice jacobienne se déduit directement. Reste à calculer α_{t+1}. Comme $\dfrac{\partial R_{t+1}^{+}}{\partial R_t} = \dfrac{\partial R_{t+1}^{+}}{\partial HU_t(2)} = 1$, on a :

$$\alpha_{t+1} = \frac{\partial R_{t+1}}{\partial R_t} = \frac{\partial R_{t+1}}{\partial R_{t+1}^{+}} = \frac{\partial R_{t+1}}{\partial HU_t(2)}$$

Or $R_{t+1} = R_{t+1}^{+} - \widehat{Q}_{t+1}$; on en déduit :

$$\begin{aligned}
\alpha_{t+1} = 1 - \frac{\partial \widehat{Q}_{t+1}}{\partial R_{t+1}^{+}} &= 1 - \frac{2R_{t+1}^{+} \cdot \left(R_{t+1}^{+} + X_1\right) - \left(R_{t+1}^{+}\right)^2}{\left(R_{t+1}^{+} + X_1\right)^2} \\
&= 1 - \frac{\left(R_{t+1}^{+}\right)^2 + 2R_{t+1}^{+}X_1}{\left(R_{t+1}^{+}\right)^2 + 2R_{t+1}^{+}X_1 + X_1^2} = \frac{X_1^2}{\left(R_{t+1}^{+} + X_1\right)^2} \\
&= \frac{X_1^2}{\left(R_{t+1} + \widehat{Q}_{t+1} + X_1\right)^2}
\end{aligned}$$

Dans ces équations, R_{t+1} et $\widehat{Q}_{t+1}$ sont les valeurs de R et de $\widehat{Q}$ calculées au pas de temps $t+1$ en partant d'un niveau du réservoir de routage mis à jour au pas de temps t, c'est-à-dire fonction du débit observé Q_t.

La prise en compte de la seule entrée dans le modèle est linéaire. La matrice $\mathbf{A}_{t+1}$ s'écrit

directement :

$$\mathbf{A}_{t+1,[i,j]} = \frac{\partial f_{t+1}}{\partial e_{t,[i]}}\left(x_{t+1,[j]}\right) \quad \text{et } \mathbf{A} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \\ \vdots \\ \gamma_\eta \\ 0 \\ 0 \end{bmatrix}$$

Linéarisation de l'observation

L'opérateur des observations h est non linéaire : il relie R_t au débit $\widehat{Q}_t$ via :

$$\begin{cases} \widehat{Q}_t = \dfrac{\left(R_t^+\right)^2}{R_t^+ + X_1} \\ R_t = R_t^+ - \widehat{Q}_t \end{cases}$$

Il en découle

$$\begin{aligned} \widehat{Q}_t &= \frac{\left(R_t + \widehat{Q}_t\right)^2}{R_t + \widehat{Q}_t + X_1} \\ 0 &= R_t^2 + R_t\widehat{Q}_t - X_1\widehat{Q}_t \end{aligned}$$

soit

$$\widehat{Q} = h\left(\mathbf{x}_t\right) = \frac{R_t^2}{X_1 - R_t}$$

La linéarisation de cet opérateur aboutit à la matrice $\mathbf{H}_t$, définie par $\mathbf{H}_{t,[i,j]} = \dfrac{\partial h}{\partial x_{[i]}}\left(x_{t,[j]}\right)$, qui s'écrit :

$$\mathbf{H}_t = [0, \ldots, 0, hq_t, 0]$$

avec

$$\begin{aligned} hq_t = \frac{\partial \widehat{Q}_t}{\partial R_t} &= \frac{2R_t\left(X_1 - R_t\right) + R_t^2}{\left(X_1 - R_t\right)^2} \\ &= \frac{2R_tX_1 - R_t^2}{\left(X_1 - R_t\right)^2} = \frac{X_1^2 - \left(X_1 - R_t\right)^2}{\left(X_1 - R_t\right)^2} \\ &= \frac{X_1^2}{\left(X_1 - R_t\right)^2} - 1 \end{aligned}$$

K.4.3 Choix des matrices de variance-covariance

Différentes combinaisons des matrices de variance-covariance des erreurs du modèle et d'observation ont été testées. Les matrices $\mathbf{R}_t$ et $\mathbf{S}_t$ sont de dimension 1×1. La matrice $\mathbf{Q}_t$ est de dimension $(\eta+2) \times (\eta+2)$; nous avons utilisé la structure suivante pour cette matrice :

$$\mathbf{Q}_t = \left[\begin{array}{ccccccccc|cc}
A_1 & B_{1,2} & 0 & \cdots & & & & \cdots & 0 & C & 0 \\
B_{1,2} & A_2 & B_{2,3} & 0 & \cdots & & & \cdots & 0 & 0 & 0 \\
0 & B_{2,3} & A_3 & B_{3,4} & 0 & & & \cdots & 0 & \vdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & & \vdots & & \\
 & & & & & & & & & & \\
 & & & & & & & & & & \\
 & & & & & & & & & & \\
 & & & & & \ddots & \ddots & \ddots & 0 & & \\
\vdots & & & & & 0 & B_{\eta-2,\eta-1} & A_{\eta-1} & B_{\eta-1,\eta} & \vdots & \vdots \\
0 & \cdots & & & & \cdots & 0 & B_{\eta-1,\eta} & A_\eta & 0 & 0 \\
\hline
C & 0 & \cdots & & & & & \cdots & 0 & D & 0 \\
0 & \cdots & & & & & & \cdots & 0 & 0 & 0
\end{array}\right]$$

K.4.4 Récapitulatif des variables employées avec leurs dimensions

Vecteur des états	$\mathbf{x}_t$	$\eta+2$
Vecteur des entrées	$\mathbf{e}_t$	1
Vecteur des observations	$\mathbf{z}_t$	1
Matrice de covariance des erreurs sur les états	$\mathbf{Q}_t$	$(\eta+2) \times (\eta+2)$
Matrice de covariance des erreurs d'observation	$\mathbf{R}_t$	1×1
Matrice de covariance des erreurs sur les entrées	$\mathbf{S}_t$	1×1
Matrice jacobienne du modèle	$\mathbf{F}_t$	$(\eta+2) \times (\eta+2)$
Matrice d'incorporation des entrées	$\mathbf{A}_t$	$(\eta+2) \times 1$
Matrice d'observation	$\mathbf{H}_t$	$1 \times (\eta+2)$
Matrice de propagation des erreurs	$\mathbf{P}_t$	$(\eta+2) \times (\eta+2)$
Matrice de gain	$\mathbf{K}_t$	$(\eta+2) \times 1$

K.5 Interprétation de la mise-à-jour « incomplète » du niveau du réservoir de routage

Au chapitre 7 (page 129), nous avons dérivé une mise-à-jour « incomplète » de la mise-à-jour du niveau du réservoir de routage proposée par Tangara (2005). Au lieu d'employer le débit observé à l'instant de prévision t pour mettre à jour ce niveau, nous avons employé une moyenne pondérée $\widetilde{Q}_t$ des débits observé et simulé à t :

$$\widetilde{Q}_t = \xi Q_t + (1-\xi)\,\widehat{Q}_{t|t-1}$$

Nous proposons ici une interprétation probabiliste de ce choix.

K.5.1 Interprétation probabiliste

La valeur réelle du débit $Q_t^{(r)}$ nous est inconnue. Nous avons plus ou moins confiance dans l'observation du débit $Q_t^{(o)}$ et dans la valeur prévue $Q_t^{(p)} = \widehat{Q}_{t|t-1}$. Cette confiance relative peut être décrite par des distributions de probabilité conditionnelles. Calculons la loi de probabilité de la valeur réelle $Q_t^{(r)}$ en fonction de la valeur observée et de la valeur prévue du débit. Le théorème de Bayes nous permet d'écrire :

$$p\left(Q_t^{(r)} \middle| Q_t^{(o)}, Q_t^{(p)}\right) \propto p\left(Q_t^{(o)} \middle| Q_t^{(r)}, Q_t^{(p)}\right) p\left(Q_t^{(r)} \middle| Q_t^{(p)}\right)$$

Or l'indépendance conditionnelle (Fig. K.4) de l'observation et de la simulation nous permet de simplifier :

$$p\left(Q_t^{(o)} \middle| Q_t^{(r)}, Q_t^{(p)}\right) = p\left(Q_t^{(o)} \middle| Q_t^{(r)}\right)$$

Ce qui nous donne :

$$\begin{aligned} p\left(Q_t^{(r)} \middle| Q_t^{(o)}, Q_t^{(p)}\right) &\propto p\left(Q_t^{(o)} \middle| Q_t^{(r)}\right) p\left(Q_t^{(r)} \middle| Q_t^{(p)}\right) \\ &\propto p\left(Q_t^{(r)} \middle| Q_t^{(o)}\right) p\left(Q_t^{(o)}\right) p\left(Q_t^{(r)} \middle| Q_t^{(p)}\right) \end{aligned} \tag{K.3}$$

où $p\left(Q_t^{(o)}\right)$ est la connaissance *a priori* sur les observations (uniforme ou climatologie ...).

Ce qui nous intéresse principalement est la relation de proportionalité suivante :

$$p\left(Q_t^{(r)} \middle| Q_t^{(o)}, Q_t^{(p)}\right) \propto p\left(Q_t^{(r)} \middle| Q_t^{(o)}\right) p\left(Q_t^{(r)} \middle| Q_t^{(p)}\right) \tag{K.4}$$

Si nous nous limitons à la recherche d'une seule valeur de débit Q_t à assimiler lors de la mise-à-jour du modèle, alors il s'agit de celle qui maximise $p\left(Q_t^{(r)} \middle| Q_t^{(o)}, Q_t^{(p)}\right)$.

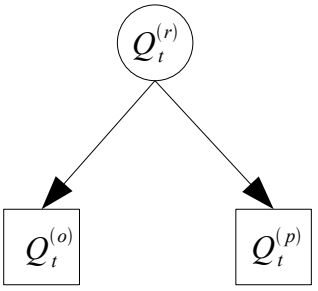


FIGURE K.4 – Indépendance de l'observation et de la simulation conditionnellement à la réalité.

K.5.2 Cas gaussien

Pour commencer, on peut supposer deux lois gaussiennes des erreurs d'observation et de simulation (comme au sein du filtre de Kálmán) :

$$\begin{cases} p\left(Q_t^{(r)} \middle| Q_t^{(o)} = q_t^{(o)}\right) = e^{-\frac{\left(q_t^{(r)} - q_t^{(o)}\right)^2}{2\sigma_o^2}} \\ p\left(Q_t^{(r)} \middle| Q_t^{(p)} = q_t^{(p)}\right) = e^{-\frac{\left(q_t^{(r)} - q_t^{(p)}\right)^2}{2\sigma_p^2}} \end{cases}$$

Chercher la valeur qui maximise la probabilité de $Q_t^{(r)}$ conditionnellement à $Q_t^{(o)}$ et $Q_t^{(p)}$ revient à minimiser la fonction f :

$$f : q \mapsto \frac{\left(q - q_t^{(o)}\right)^2}{\sigma_o^2} + \frac{\left(q - q_t^{(p)}\right)^2}{\sigma_p^2}$$

qui se dérive facilement :

$$\frac{df}{dq}(q) = 2\frac{\left(q - q_t^{(o)}\right)}{\sigma_o^2} + 2\frac{\left(q - q_t^{(p)}\right)}{\sigma_p^2}$$

pour donner :

$$\frac{df}{dq}(\widehat{q}_t) = 0 \Leftrightarrow \widehat{q}_t = \frac{\sigma_o^2}{\sigma_p^2 + \sigma_o^2} q_t^{(p)} + \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2} q_t^{(o)}$$

soit

$$\widehat{q}_t = \left(1 + \alpha^2\right)^{-1} q_t^{(p)} + \left(1 + \frac{1}{\alpha^2}\right)^{-1} q_t^{(o)} \tag{K.5}$$

En notant α, le rapport de notre incertitude concernant la simulation sur notre incertitude sur les observations :

$$\alpha = \frac{\sigma_p}{\sigma_o}$$

Nous retrouvons notre moyenne pondérée en posant :

$$\xi = \left(1 + \frac{1}{\alpha^2}\right)^{-1}$$

K.5.3 Transformation logarithmique des débits

Si nous employons la classique transformation logarithmique des débits visant à limiter l'hétéroscédasticité des sorties du modèle et des observations, nous avons :

$$\begin{cases} p\left(Q_t^{(r)} \middle| Q_t^{(o)} = q_t^{(o)}\right) = e^{-\frac{\left(\lg\left(q_t^{(r)}\right) - \lg\left(q_t^{(o)}\right)\right)^2}{2\sigma_o^2}} \\ p\left(Q_t^{(r)} \middle| Q_t^{(p)} = q_t^{(p)}\right) = e^{-\frac{\left(\lg\left(q_t^{(r)}\right) - \lg\left(q_t^{(p)}\right)\right)^2}{2\sigma_p^2}} \end{cases}$$

La valeur $q_t^{(r)}$ qui maximise la probabilité de $Q_t^{(r)}$ conditionnellement à $Q_t^{(o)}$ et $Q_t^{(p)}$ s'obtient de façon très similaire en minimisant la fonction g :

$$g : q_t^{(r)} \mapsto \frac{\left(\lg\left(q_t^{(r)}\right) - \lg\left(q_t^{(o)}\right)\right)^2}{2\sigma_o^2} + \frac{\left(\lg\left(q_t^{(r)}\right) - \lg\left(q_t^{(p)}\right)\right)^2}{2\sigma_p^2}$$

Ce qui nous donne :

$$\widehat{q}_t = q_t^{(p)\,\frac{1}{1+\alpha^2}}\, q_t^{(o)\,\frac{1}{1+\frac{1}{\alpha^2}}}$$

Soit une forme bien différente de celle que nous avons employée. Cependant, cette forme possède les mêmes propriétés aux limites de α en 0 et en $+\infty$.

Tenir compte d'une dépendance temporelle conduit à des formes plus complexes de la valeur du débit qui maximise sa probabilité conditionnellement à l'observation et à la simulation.

L

Annexe

Compléments au chapitre 8

Cette annexe présente les principes et le fonctionnement du filtre particulaire employé au chapitre 8. Elle détaille ensuite l'évaluation des estimations des incertitudes de prévision.

L.1 Principes et fonctionnement du filtre particulaire

Le filtre particulaire, employé au chapitre 8 est une méthode d'estimation récursive bayésienne.

L.1.1 L'estimation bayésienne récursive

L'assimilation de données par inférence bayésienne fonctionne d'une manière similaire à toutes les méthodes de mise-à-jour que nous avons vues jusqu'ici. Elle s'appuie sur un cycle à deux temps : le modèle hydrologique est d'abord employé pour estimer les états du système et ses sorties, puis une information extérieure (pour nous, les débits observés [1]) est valorisée pour mettre-à-jour notre connaissance du système (ici, les densités de probabilité des états). Voyons comment cela se traduit dans le formalisme bayésien.

Premier temps : propagation des états

Nous nous plaçons ici dans l'espace des états (comme pour le filtre de Kálmán, page 116) : l'état du système est représenté par un vecteur aléatoire $\mathbf{x}_t$. Nous supposons que ce vecteur évolue de façon récursive (processus markovien d'ordre 1) :

$$\mathbf{x}_t = f\left(\mathbf{x}_{t-1}, \mathbf{e}_t, \omega_k\right) \tag{L.1}$$

où $\mathbf{e}_t$ est le vecteur des entrées du modèle au pas de temps t et ω_t un bruit gaussien (on suppose qu'il n'y a plus d'information à retirer du bruit introduit par cette étape).

Le système nous est connu par un certain nombre d'observations (vecteurs $\mathbf{z}_t$). Traduisons l'équation (L.1) dans le formalisme employé par les approches bayésiennes, c'est-à-dire en termes de densités de probabilité conditionnellement aux observations :

$$p\left(\mathbf{x}_t \left| \{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]} \right.\right) = \int p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1} \right.\right) p\left(\mathbf{x}_{t-1} \left| \{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]} \right.\right) d\mathbf{x}_{t-1} \tag{L.2}$$

Deuxième temps : mise-à-jour des densités de probabilité des états

Le théorème de Bayes nous permet de mettre à jour la densité de probabilité des états grâce aux observations $\mathbf{z}_k$, c'est-à-dire de passer de la densité *a priori* des états $p\left(\{\mathbf{x}_k\}_{k\in[\![0,t]\!]}\right)$ à la densité *a posteriori* $p\left(\{\mathbf{x}_k\}_{k\in[\![0,t]\!]} \left| \{\mathbf{z}_k\}_{k\in[\![1,t]\!]} \right.\right)$ à l'aide de la fonction de vraisemblance

1. Pour la plupart de nos essais, il ne s'agit pas d'une information complètement extérieure puisque les débits observés sont employés par la mise-à-jour du niveau du réservoir de routage.

$p\left(\{\mathbf{z}_k\}_{k\in[\![1,t]\!]}\middle|\{\mathbf{x}_k\}_{k\in[\![0,t]\!]}\right)$:

$$p\left(\{\mathbf{x}_k\}_{k\in[\![0,t]\!]}\middle|\{\mathbf{z}_k\}_{k\in[\![1,t]\!]}\right)=\frac{p\left(\{\mathbf{z}_k\}_{k\in[\![1,t]\!]}\middle|\{\mathbf{x}_k\}_{k\in[\![0,t]\!]}\right)p\left(\{\mathbf{x}_k\}_{k\in[\![0,t]\!]}\right)}{p\left(\{\mathbf{z}_k\}_{k\in[\![1,t]\!]}\right)} \tag{L.3}$$

La nature récursive du modèle (équation (L.1)) permet de simplifier (L.3) pour ne plus tenir compte que de la dernière observation :

$$p\left(\mathbf{x}_t\middle|\{\mathbf{z}_k\}_{k\in[\![1,t]\!]}\right)=\frac{p\left(\mathbf{z}_t\,|\mathbf{x}_t\right)p\left(\mathbf{x}_t\middle|\{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]}\right)}{p\left(\mathbf{z}_t\middle|\{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]}\right)} \tag{L.4}$$

La constante de normalisation pouvant s'écrire :

$$p\left(\mathbf{z}_t\middle|\{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]}\right)=\int p\left(\mathbf{z}_t\,|\mathbf{x}_t\right)p\left(\mathbf{x}_t\middle|\{\mathbf{z}_k\}_{k\in[\![1,t-1]\!]}\right)d\mathbf{x}_t$$

Les équations (L.2) et (L.4) définissent l'estimation bayésienne récursive. Elles permettent d'obtenir les densités de probabilité des états conditionnellement aux observations à chaque pas de temps par itération du cycle :

1. simulation (équations (L.1) et (L.2)),
2. mise-à-jour (équation (L.4)).

L.1.2 Une implémentation pratique : le filtre particulaire

Un système rarement soluble analytiquement

Le système développé ci-dessus n'est soluble analytiquement que dans un très petit nombre de cas (Gordon *et al.*, 1993). Un cas bien connu que nous avons examiné dans le chapitre 7 est celui du modèle linéaire avec des densités de probabilité toutes gaussiennes : on aboutit alors au filtre de Kálmán. Thiemann *et al.* (2001) emploient, quant à eux, d'autres hypothèses sur la nature et les paramètres des densités de probabilité pour obtenir un cas soluble ; ils construisent ainsi une estimation bayésienne récursive (BaRE) analytique des paramètres.

La représentation particulaire

Dans les autres cas, une approche classique emploie les méthodes de Monte-Carlo séquentielles, autre nom du filtre particulaire. Elle consiste à approximer les densités de probabilité par des ensembles de « particules ». Dans ce cadre, la densité de probabilité d'un vecteur aléatoire $\mathbf{x}_t$ est représentée par un échantillon (fini) de valeurs $\left\{\mathbf{x}_t^{(i)}\right\}_{i\in[\![1,N]\!]}$ auxquelles

sont associés des poids $w_t^{(i)}$ de façon à ce que la distribution basée sur cet échantillon $\mathbf{x} \mapsto \sum_{i=1}^{N} w_t^{(i)} \delta\left(\mathbf{x} - \mathbf{x}_t^{(i)}\right)$ approxime la densité de probabilité de $\mathbf{x}_t$.

Propagation des états

Le filtre particulaire met en œuvre le cycle simulation – mise-à-jour de l'estimation bayésienne récursive. La première étape, la propagation des états, obtient simplement les états $\mathbf{x}_t^{(i)}$ par l'opérateur f, i.e. le modèle hydrologique, à partir des états $\mathbf{x}_{t-1}^{(i)}$ au pas de temps précédent :

$$\mathbf{x}_{t-1}^{(i)} \longrightarrow \mathbf{x}_t^{(i)} = f\left(\mathbf{x}_{t-1}^{(i)}, \mathbf{e}_t, \omega_t^{(i)}\right) \tag{L.5}$$

Les incertitudes sur les entrées sont prises en compte (Gordon *et al.*, 1993, Moradkhani *et al.*, 2005a) à travers un bruitage des entrées au niveau de chaque particule (Fig. L.1).

Mise-à-jour des densités de probabilité

La seconde étape met à jour non pas les états décrits par chaque particule mais la densité de probabilité des états. Il s'agit ici de la mise-à-jour des poids des particules permettant d'approximer la distribution *a posteriori* des états $\mathbf{x}_t$ (Fig. L.2). Une méthode classique se base sur l'échantillonnage préférentiel séquentiel (*Sequential Importance Sampling*, SIS). Cette technique vise à répondre à une difficulté majeure : l'impossibilité pratique d'échantillonner dans une densité de probabilité (ici la densité *a posteriori* des états). Smith et Gelfand (1992), Arulampalam *et al.* (2002) montrent qu'il est possible de générer un échantillon E_2 approximant une densité de probabilité D_2 dans laquelle on ne sait pas tirer directement, si on dispose d'une densité de probabilité D_1 (dite densité préférentielle) dans laquelle on peut tirer un échantillon $E_1 = \{\theta_1, \ldots, \theta_n\}$ et si on peut calculer les valeurs $D_2(\theta_i)$.

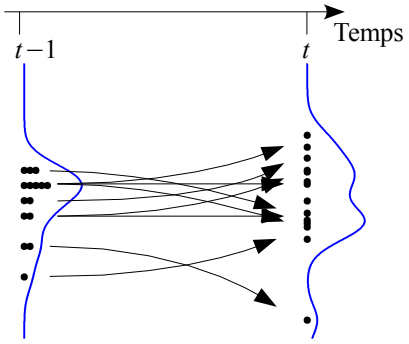


Figure L.1 – Sur ce schéma, les points noirs représentent l'échantillon de particules tandis que la densité de probabilité des états correspondant à cet échantillon est représentée par la courbe bleue. Dans la première étape de l'inférence bayésienne par filtrage particulaire, chaque particule évolue du pas de temps $t-1$ au pas de temps t, conformément au modèle (équation L.1). Les entrées étant bruitées, des particules identiques au pas de temps $t+1$ diffèrent au pas de temps t.

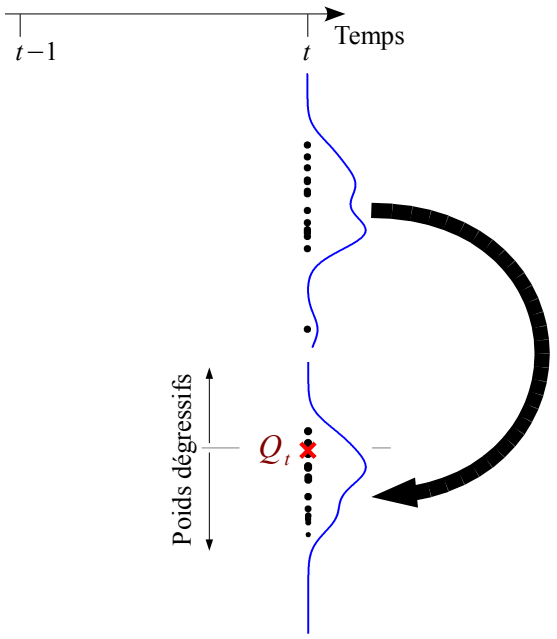


FIGURE L.2 – L'observation du débit Q_t permet d'estimer la densité *a posteriori* des états à travers la mise-à-jour des poids des particules. Sur ce schéma, les poids sont proportionnels à la taille des particules : plus les particules donnent un débit proche de Q_t, plus leurs poids relatifs seront élevés.

En pratique, Arulampalam *et al.* (2002) ont montré que, dans un cadre séquentiel, si on choisissait le prior $p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1}^{(i)} \right.\right)$ comme distribution préférentielle, alors les poids pouvaient s'écrire de façon récursive :

$$w_t^{(i)} \propto w_{t-1}^{(i)} p\left(\mathbf{z}_t \left| \mathbf{x}_t^{(i)} \right.\right) \tag{L.6}$$

Dégénérescence des échantillons

La mise-à-jour SIS que nous venons de présenter, rencontre très souvent un problème majeur : la dégénérescence de l'échantillon. Après un certain nombre de pas de temps, il ne reste que quelques particules dont les poids associés ne sont pas négligeables. L'échantillon de particules approxime alors très mal la densité de probabilité qu'elle est censée représenter. Cette situation est inextricable car la variance des poids ne peut qu'augmenter dans le temps (Arulampalam *et al.*, 2002).

Rééchantillonnage par importance : SIR

Cette difficulté peut être résolue par le couplage de SIS à un rééchantillonnage, comme Smith et Gelfand (1992) l'ont proposé. La méthode se nomme alors *Sampling Importance Resampling* (SIR).

Le rééchantillonnage consiste à multiplier les particules de plus forts poids au sein de l'échantillon et à éliminer aléatoirement certaines particules de poids trop peu significatifs,

en tirant N particules (tirage avec remise) parmi les N particules selon une loi de probabilité discrète définie par les poids $w_{t-1}^{(i)}$. Le rééchantillonnage aboutit à un ensemble de particules équiprobables où les particules de l'ensemble initial qui présentaient les plus forts poids sont très représentées (Fig. L.3).

En partant d'une population équiprobable, l'équation (L.6) devient :

$$w_t^{(i)} \propto p\left(\mathbf{z}_t \left| \mathbf{x}_t^{(i)} \right.\right) \tag{L.7}$$

Cette méthode n'est pas exempte de défaut : elle peut conduire à des échantillons contenant un grand nombre de fois des particules similaires (appauvrissement de l'échantillon). Ce problème est particulièrement aigu quand il y a peu de bruit dans les entrées (Moradkhani *et al.*, 2005a). La prise en compte des incertitudes dans les entrées réintroduit de la variabilité dans les particules (cf. supra) et devrait limiter l'appauvrissement.

Rééchantillonnage résiduel : RR

Weerts et El Serafy (2006) emploient une amélioration du SIR, dite rééchantillonnage résiduel (*Residual Resampling*, RR). Les poids $w_t^{(i)}$ sont d'abord calculés conformément à l'algorithme SIR. Puis le rééchantillonnage est réalisé en deux étapes. Le poids de chaque

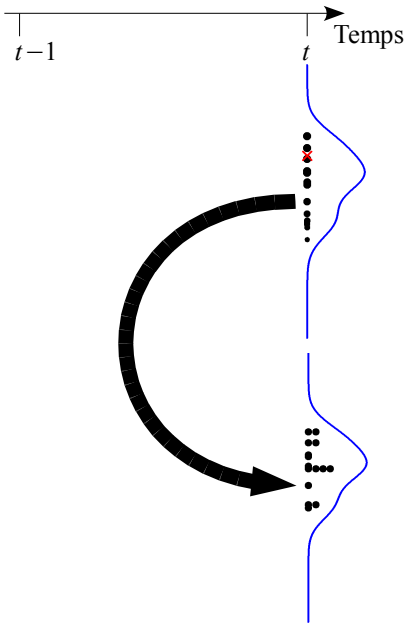


Figure L.3 – La dernière étape du filtre particulaire consiste à rééchantillonner les particules au pas de temps t pour obtenir des particules à nouveau équiprobables en dupliquant les particules de forts poids et en éliminant aléatoirement les particules de poids faibles.

particule est d'abord réévalué par [2] :

$$w_{1,t}^{(i)} = \frac{\left\lfloor N w_t^{(i)} \right\rfloor}{\sum_{i=1}^{N} \left\lfloor N w_t^{(i)} \right\rfloor} \tag{L.8}$$

pour tirer $\sum_{i=1}^{N} \left\lfloor N w_{1,t}^{(i)} \right\rfloor$ particules. L'échantillon est ensuite complété par un tirage de $N - \sum_{i=1}^{N} \left\lfloor N w_{1,t}^{(i)} \right\rfloor$ particules dans une densité définie par les poids :

$$w_{2,t}^{(i)} = \frac{N w_t^{(i)} - \left\lfloor N w_t^{(i)} \right\rfloor}{N - \sum_{i=1}^{N} \left\lfloor N w_t^{(i)} \right\rfloor} \tag{L.9}$$

Nous emploierons cette méthode car elle conduit à une variance des poids plus faible (Salamon et Feyen, 2009).

L.2 Évaluation des estimation des incertitudes de prévision

L.2.1 Le critère CRPS

Principe

Le CRPS – *Continuous rank probability score* – permet de comparer différentes prévisions probabilistes en terme de distance à l'observation. Il calcule la distance entre une prévision probabiliste et la prévision probabiliste parfaite qui serait une distribution de Dirac centrée sur l'observée [3] (cf. Fig L.4).

Considérons les distributions de probabilité cumulée de :

1. notre prévision probabiliste obtenue par un modèle hydrologique : $F(Q)$,
2. la prévision parfaite (dont la distribution de probabilité cumulée est une distribution de Heaviside : 0 avant la valeur observée, 1 après) : $H\left(Q - Q_{Obs}\right)$

Le CRPS s'écrit alors

$$CRPS = \left\langle \int_{\mathbb{R}^+} \left(F(Q) - H\left(Q - Q_{Obs}\right)\right)^2 dQ \right\rangle$$

où $\langle X \rangle$ est la moyenne des réalisations de X sur un grand nombre de pas de temps.

2. La notation $\lfloor \cdot \rfloor$ désigne la partie entière.
3. On ne tient pas compte ici de l'erreur d'observation. Le critère peut s'adapter à une telle situation où on spécifierait la distribution de l'erreur d'observation ; mais les calculs seraient plus difficiles à réaliser.

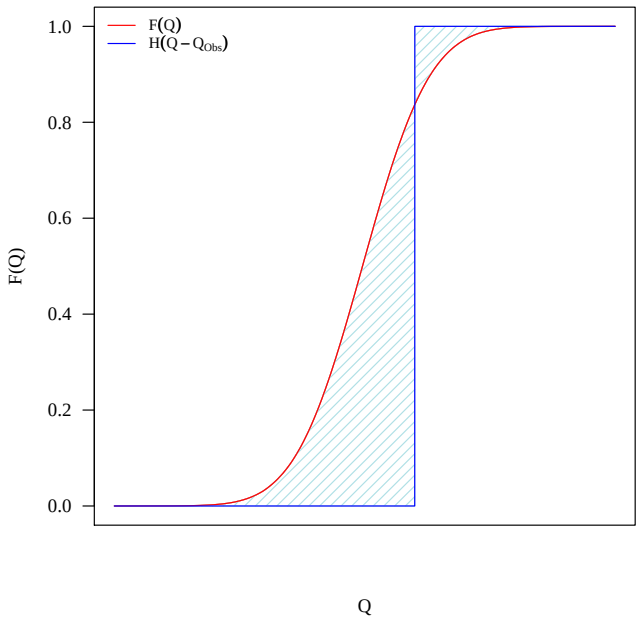


Figure L.4 – Le CRPS correspond à une mesure (par la norme $\|\cdot\|_2^2$ de l'aire (bleu clair) entre la distribution des prévisions obtenues (rouge) et la prévision parfaite (bleue).

Comparaison à des prévisions déterministes

Si on compare de la même façon une prévision déterministe à la prévision parfaite, le CRPS revient à mesurer la distance entre deux distributions de Heaviside :

$$\left\langle \int_{\mathbb{R}^+} \left(H\left(Q - Q_{Prev}\right) - H\left(Q - Q_{Obs}\right)\right)^2 \right\rangle$$

Comme $0^2 = 0$ et $1^2 = 1$, ce calcul revient à calculer l'erreur absolue moyenne

$$\langle |Q_{Prev} - Q_{Obs}| \rangle$$

Ainsi, on peut comparer le CRPS de l'ensemble de prévisions à l'erreur absolue moyenne d'une prévision déterministe, indépendante de l'ensemble (par exemple celle issue d'un modèle déterministe non perturbé) ou dépendante de l'ensemble (sa moyenne, sa médiane ou son mode [4]).

4. Le mode d'une distribution est le point qui présente la plus forte probabilité. Ce choix peut être peu robuste en cas de distributions multimodales (i.e., qui présentent plusieurs maxima de probabilité).

L.2.2 Ensembles parfaits

Notations

$X_e(J)$	Prévision du membre J de l'ensemble,
X_T	Valeur « vraie » de X (qu'on estimera par son observation),
$\langle X_e \rangle$	moyenne des membres de l'ensemble (pour un pas de temps donné),
$\|\ldots\|$	distance euclidienne dans l'espace des X (pour les débits, c'est la différence),
$\overline{X}$	espérance temporelle de X.

Propriétés de l'ensemble parfait

Un ensemble parfait est parfaitement fiable. De cette définition, il vient que pour le premier moment d'un ensemble parfait,

$$\overline{X_e(J)} = \overline{\langle X_e \rangle} = \overline{X_T} \tag{L.10}$$

De même pour le deuxième moment de la distribution d'un ensemble parfait, on a pour tout membre J,

$$\overline{\|X_e(J) - \langle X_e \rangle\|^2} = \overline{\|X_T - \langle X_e \rangle\|^2} \tag{L.11}$$

Ceci étant vérifié pour tout membre J, il vient :

$$\overline{\left\langle \|X_e - \langle X_e \rangle\|^2 \right\rangle} = \overline{\|X_T - \langle X_e \rangle\|^2} \tag{L.12}$$

Cette dernière équation (L.12) est moins forte que l'équation (L.11) : elle en découle mais ne lui est pas équivalente.

Comment le vérifier ?

Il « suffit » de tracer en fonction de l'horizon de prévision,

1. l'erreur de la moyenne de l'ensemble (comparée à l'observée),
2. la variance de l'ensemble (variabilité) autour de sa moyenne.

Pour un ensemble parfait, les deux courbes doivent se superposer. Les « ensemblistes » se satisfont d'ensembles pour lesquels ces deux courbes se superposent (à peu près), bien qu'il ne s'agit pas d'une condition suffisante pour prouver la perfection de l'ensemble.

Quelques propriétés

Avec les mêmes notations que précédemment,

$$\begin{aligned}\overline{\left\langle \|X_e - X_T\|^2 \right\rangle} &= \overline{\left\langle \|(X_e - \langle X_e \rangle) + (\langle X_e \rangle - X_T)\|^2 \right\rangle} \\ &= \overline{\left\langle \|X_e - \langle X_e \rangle\|^2 \right\rangle + \left\langle \|X_T - \langle X_e \rangle\|^2 \right\rangle + 2\left\langle (X_e - \langle X_e \rangle) \cdot (\langle X_e \rangle - X_T) \right\rangle}\end{aligned}$$

L'équation (L.12) nous indique que les deux premiers termes sont identiques. Le troisième terme est nul car il peut se réécrire :

$$\overline{\langle (X_e - \langle X_e \rangle) \cdot (\langle X_e \rangle - X_T) \rangle} = \overline{\langle X_e - \langle X_e \rangle \rangle \cdot (\langle X_e \rangle - X_T)} = 0$$

On a donc :

$$\overline{\left\langle \|X_e - X_T\|^2 \right\rangle} = 2\overline{\left\langle \|\langle X_e \rangle - X_T\|^2 \right\rangle} = 2\overline{\left\langle \|\langle X_e \rangle - X_e\|^2 \right\rangle} \tag{L.13}$$

L'incertitude de la prévision (l'erreur moyenne qu'on ferait en se basant sur un seul membre de l'ensemble) est égale à deux fois la variabilité des membres de l'ensemble autour de leur moyenne. Dans la démonstration, il faut que tous les membres soient définis de la même façon, pour qu'ils soient interchangeables.

Cette propriété peut être interprétée par un processus (imaginaire) où la formation de l'ensemble se ferait en deux étapes :

1. on tire la moyenne de l'ensemble de façon aléatoire,
2. on tire les membres de l'ensemble (et l'observée) dans une loi de distribution centrée sur la moyenne.

Ce qui précède peut s'écrire :

$$Var\,(X_e - X_T) = 2Var\,(X_e - \langle X_e \rangle) = 2Var\,(X_T - \langle X_e \rangle)$$

Cette propriété suppose que les membres de l'ensemble soient interchangeables : on ne peut pas avoir de construction des membres de l'ensemble qui conduisent systématiquement pour un membre donné à des débits inférieurs à l'observation (par exemple, en sous-estimant tout le temps les entrées de pluie).

M

Annexe

Premiers tests d'emploi des prévisions d'ensemble Arpege

Cette annexe présente les premiers essais d'emploi au pas de temps horaire des prévisions d'ensemble de précipitation fournies par le modèle Arpege (Action de Recherche à Petite Échelle et Grande Échelle) de Météo-France.

M.1 Données disponibles

Le modèle Arpege fournit des ensembles de prévision de précipitation constitués de onze membres : la meilleure prévision accompagnée de dix prévisions aux conditions initiales perturbées par recherche de vecteurs singuliers. Le Cemagref dispose de ces prévisions sur 19 mois de mars 2005 à septembre 2006 [1]. Il s'agit de prévisions établies une fois par jour sur un horizon de 60 heures au pas de temps tri-horaire.

Avec ces données, nous ne pouvons pas réaliser un test complet (notamment en calage – contrôle sur de longues périodes) de l'intérêt des prévisions d'ensemble de précipitation pour notre modèle. Nous menons ici un test préliminaire avec les données recoupant nos données de débit : soit sur 5 mois de mars à juillet 2005 et sur 28 bassins. Nous ne nous intéressons ici qu'à la fiabilité des prévisions.

M.2 Essais réalisés

Nous avons effectué les prévisions d'ensemble hydrologiques basées sur ces prévisions d'ensemble de pluie. Les hydrogrammes de prévision ont été tracés en série et examinés.

Nous avons en outre tracé pour nos 28 bassins, les diagrammes de Talagrand, c'est-à-dire l'histogramme de répartition du rang du débit observé par rapport aux onze membres prévus. L'histogramme comporte douze classes : la première classe correspond aux pas de temps où le débit observé est inférieur aux onze débits prévus ; la i^{e} classe rassemble les pas de temps pour lesquels le débit observé est supérieur à $i-1$ débits prévus et inférieurs aux autres. Des prévisions fiables conduisent à un diagramme de Talagrand plat (les classes sont équiprobables). Un diagramme en forme de U dénote une sous-dispersion. Les diagrammes de Talagrand ont été réalisés pour les pas de temps qui présentent un cumul moyen de précipitations prévues (par Arpege) supérieur à un seuil (1, 5 ou 10 mm) pour éliminer les cas où tous les membres de l'ensemble des pluies Arpege contiennent des scénarios de pluie nulle... et égale.

1. Les données de prévision d'ensemble PEARP ont été gracieusement fournies dans le cadre du protocole d'expérimentation sur les prévisions hydrologiques d'ensemble au Schapi, signé en mai 2008, entre la Direction de la Production de Météo-France, le Schapi et le Cemagref. Ces données sont exploitées dans les recherches menées par l'équipe du Cemagref d'Antony portant sur la prévision hydrologique d'ensemble, en coopération avec Météo-France/CNRM. Des études menées avec GR3P au pas de temps journalier sont présentées par Randrianasolo *et al.* (2010).

M.3 Résultats et discussion

L'examen visuel des hydrogrammes prévus montre que le plus souvent la gamme des prévisions hydrologiques ne contient pas le débit observé (Fig. M.1). Cela peut être du soit à une sous-dispersion, soit à un biais des prévisions. Nous n'avons pas observé de biais très nets sur notre échantillon. Ce résultat est confirmé par tous les diagrammes de Talagrand qui présentent une forme en U très marquée (Fig. M.2). L'interprétation des diagrammes de Talagrand est toutefois limitée par le faible nombre de pas de temps pris en compte (surtout pour les seuils les plus élevés).

Cette sous-estimation des incertitudes est plus marquée que dans les expériences récentes réalisées à partir d'un autre modèle météorologique (celui d'Environnement Canada) par Velázquez *et al.* (2009) sans toutefois être en contradiction.

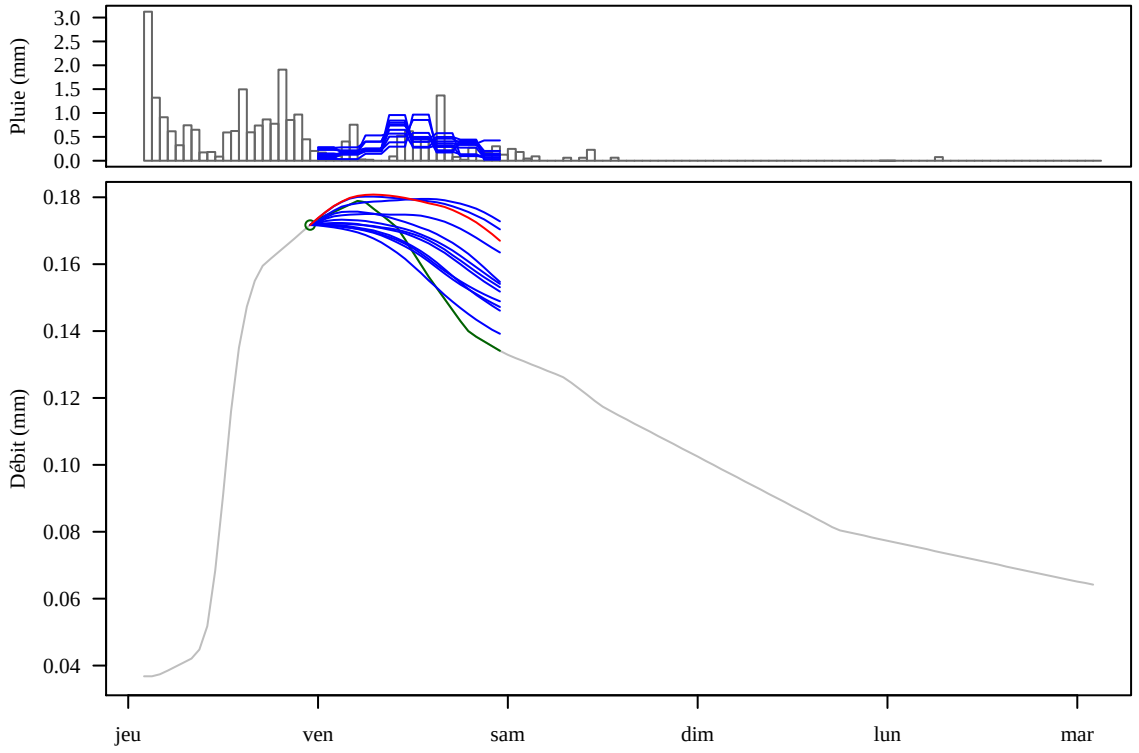


Figure M.1 – Prévision hydrologique d'ensemble basée sur les prévisions de précipitation d'ensemble Arpege pour l'Isle à Bassilac (1860 km^2, P7001510), le 8 avril 2005. En haut, le hyétogramme observé (en gris) et les onze prévisions de pluie (en bleu). En bas, l'hydrogramme observé (en gris et en vert sur la période de prévision). Les prévisions des onze membres de l'ensemble sont tracées en bleu ; la prévision faite avec le scénario de pluies « futures » contenant les observations *a posteriori* est indiquée en rouge. Cet exemple illustre une situation fréquente de sous-estimation des incertitudes : l'ensemble de 11 membres ne contient pas l'observation.

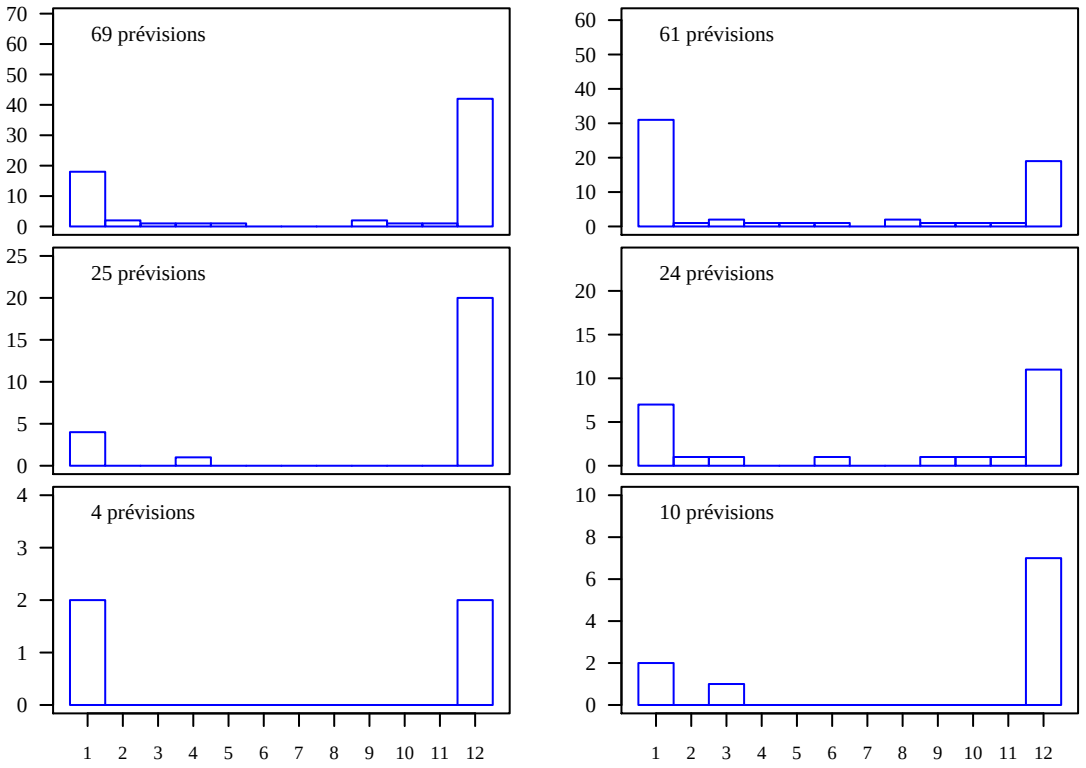


Figure M.2 – Exemples de diagrammes de Talagrand obtenus sur 4,5 mois de prévision des débits à partir des prévisions d'ensemble de précipitation fournies par le modèle Arpege. Les prévisions sont ici calculées sur le bassin de la Semoy à Haulmé (1 340 km ^{2}, B6111010) à gauche et sur le bassin de l'Isle à Bassilac (1 860 km^2, P7001510) à droite. Nous n'avons retenu que les pas de temps pour lesquels le cumul des pluies futures prévues (sur 24 heures) dépassaient un seuil de 1, 5 ou 10 mm (de haut en bas).

De nombreuses prévisions d'ensemble sont connues pour être sous-dispersées. Randrianasolo *et al.* (2010) observe une sous-dispersion des prévisions d'ensemble d'Arpege. Elles le sont particulièrement sur les 24 premières heures (c'est-à-dire les prévisions que nous employons). Cela peut expliquer la forte sous-dispersion des prévisions hydrologiques que nous constatons.

Ici, nous ne nous sommes intéressés qu'à la fiabilité des prévisions. Les travaux menés au sein de notre équipe de recherche s'attachent également à caractériser la précision de ces pluies (par rapport aux observations) et les performances du modèle GR3P alimenté avec ces données pour annoncer le dépassement de seuils d'alerte. Les premiers résultats de ces travaux sont encourageants. De plus, Annie Randrianasolo [2] m'a indiqué que les 28 bassins sur lesquels j'ai dû travailler compte parmi ceux sur lesquels elle obtient ses plus mauvais résultats (en termes de probabilité de détection et de biais).

2. Communication personnelle. Annie Randrianasolo mène actuellement une thèse sur l'utilisation des prévision d'ensemble Arpege par le modèle GR3P au pas de temps journalier.

Nous ne pouvons donc pas établir de conclusions fortes à partir de si peu de données ; cependant, il apparaît assez nettement que ne prendre en compte que les incertitudes sur les pluies futures n'est pas suffisant pour estimer l'incertitude des prévisions hydrologiques (comme le mentionne Krzysztofowicz, 1999).

N

Annexe

Durées de vie des mises-à-jour et efficacité du modèle quand peu de données sont disponibles

Cette annexe est la présentation réalisée lors de la conférence de l'Association internationale des Sciences hydrologiques, qui s'est tenue en septembre 2009. Il s'agit d'une étude de cas visant à déterminer les capacités en prévision du modèle GRP sur des bassins pour lesquels peu de données de débit (permettant une mise-à-jour) sont disponibles.

La présentation débute par une comparaison des durées de vie de différentes mises-à-jour du modèle. La *durée de vie* est définie ici comme étant la durée minimale Δt telle que les prévisions du débit pour le pas de temps $t + \Delta t$, réalisées avec et sans mises-à-jour à la date t ($\widehat{Q}_{t+\Delta t|t}$ et $Q_{t+\Delta t}$) ne sont plus significativement différentes.

Une fois ces durées de vie estimées, la présentation se poursuit avec l'estimation de l'influence de la raréfaction des données de débit disponibles (en partant du cas idéal où la donnée est disponible pour tous les pas de temps) sur les erreurs de prévision.

New Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009.

Can discharge assimilation methods be used to improve flood forecasting when few data are available?

LIONEL BERTHET[1,2], MARIA-HELENA RAMOS[1], CHARLES PERRIN[1], VAZKEN ANDRÉASSIAN[1] & CÉCILE LOUMAGNE[1]

1 *Cemagref, Hydrosystems and Bioprocesses Research Unit, BP 44, F-92163 Antony Cedex, France*
lionel.berthet@cemagref.fr

2 *AgroParisTech ENGREF, 19 Avenue du Maine, F-75732 Paris, France*

Abstract Forecasting floods is a major issue for public safety all over the world. Due to the difficulties inherent in the flood forecasting exercise, data assimilation techniques have been developed to cope with model errors. Unfortunately, these techniques require recent (real or near real-time) observations which may not be readily available in regions lacking automatic measurements networks. This paper investigates the impact of data assimilation techniques on discharge forecasts and model performance when few (but not zero) discharge measurements are available for the data assimilation. A parsimonious rainfall–runoff model is applied to a set of 178 French catchments. We explore the time properties of different discharge data assimilation schemes. Life times of the updates and model performance are assessed as a function of the time between the last available discharge observation and the forecast. State updating proves to have an added value to the forecasting system, even when data availability is limited.

Key words flood forecasting; data availability; rainfall–runoff modelling; data assimilation

INTRODUCTION

A wide range of modelling strategies is available for flood forecasting. Among them, rainfall–runoff (RR) modelling has demonstrated good performance and potential for flood forecasting. Nonetheless, RR models are far from being perfect tools. They often need to integrate additional updating techniques to be kept "on track" when used in real-time operational forecasting.

Depending on the modelling approach, different updating techniques can be used (Refsgaard, 1997). They can be applied to update model inputs, internal states, parameters or outputs. Algorithms commonly used in flood forecasting are: Kalman filtering (see e.g. Da Ros & Borga, 1997; Aubert *et al*., 2003; Evensen, 2003), autoregressive models (ARIMA, see Box & Jenkins, 1976), and artificial neural networks (ANN, Maier & Dandy, 2000; Anctil *et al*., 2003). These techniques are based on real-time data assimilation. Their use implies that additional data are provided to the RR model beyond its usual input (i.e. mainly precipitation) to improve the forecasts. In most flood forecasting real-time contexts, these additional data can only be discharge observations. Operationally, as discharge measurements become available, they are introduced into the model to adjust it and/or correct forecast errors.

A need for robust and long-memory data assimilation approaches

Updating can significantly improve the accuracy of the forecasts, but it also increases the complexity of the system. Different studies have compared the potential of different data assimilation schemes for flood forecasting (see e.g. WMO, 1992; Madsen *et al*., 2000; Moore, 2007). One of the main difficulties arises from the availability of real-time or near real-time data at appropriate modelling space–time scales (data scarcity problems), as well as from the retrieving and quality-controlling of field data in real time (processing and management of data sets). Under these conditions, forecasting RR-based models require robust updating procedures, i.e. the models have to remain efficient when the operational conditions are far from "ideal" laboratory conditions. The way streamflow information is incorporated in flood forecasting models must remain coherent with the characteristics of such additional information.

In continuous model simulation, it is well known that initial states tend to dissipate after some time of simulation. Similarly, the effects of updating in forecasting models are limited in time. One can therefore say that any update is characterized by a "life time", defined as the time beyond

Copyright © 2009 IAHS Press

which the additional information used at the start of the forecasts no longer has a significant impact on the forecasts issued. The impact of an updating procedure progressively declines from "High", immediately after the updating to "Low" (Fig. 1) for some delay after the updating. This delay may depend on various factors, among them the modelling approach and the physical characteristics of the catchment being modelled. If this life time is lower than the operational forecast length $\Delta t = N + L$ (where N is the lag time, expressed as a number of time steps, between the last available discharge observation and the time step at which the forecast is issued, and L is the lead time), the update is no longer effective.

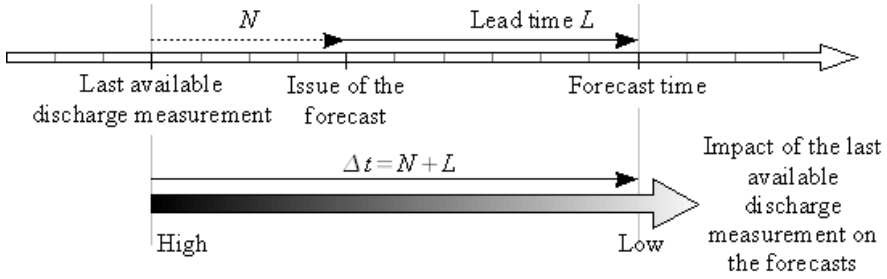


Fig. 1 Impact of an updating procedure: life times of updates are compared to the operational forecast length (i.e. duration between the last available discharge measurement and the forecast time step).

Scope of the paper

In this study, we address the following question: What is the maximum delay for which the performance of a given updating procedure still has a significant positive impact on the forecast? In other terms, at which minimum rate should discharge measurements be available to guarantee a real improvement of forecasting performances through flow data assimilation?

A two-step methodology was adopted. First, a preliminary assessment of the "life times" of an updated routing store, when considering variable updating lengths (theoretical evaluation), and of auto-regressive updated outputs (empirical evaluation) was performed. Second, we evaluate model performance when forecasts are issued on the basis of updating techniques that assimilate discharge observations available at different time steps before the forecast is issued. Data and methods are presented in the next sections, followed by results and conclusions.

DATA AND MODEL

Catchment set and hydrological data

The study is carried out on a set of 178 French unregulated catchments (Fig. 2). The set is representative of the hydroclimatic variability encountered in the country: from catchments experiencing Mediterranean flash floods to much slower catchments. Catchments in high-elevation zones are not considered, since the RR model used does not include a snow-accounting module. Catchment areas range from 10 to 5940 km^2 (354 km^2 on average). Working on various catchments ensures more general and robust conclusions to our study (Andréassian *et al*., 2006).

Available data consist of hourly areal precipitation, potential evapotranspiration (PE) and discharge from 1995 to 2005. PE values were computed using the formula proposed by Oudin *et al*. (2005), based on temperature and extraterrestrial radiation.

The GRP forecasting rainfall–runoff model and its updating techniques

The GRP model is a continuous, lumped, hybrid metric-conceptual model, designed specifically for flood forecasting (Tangara, 2005). It is one of the operational models used to forecast river flows in real time on French catchments, including the Seine River basin upstream of Paris.

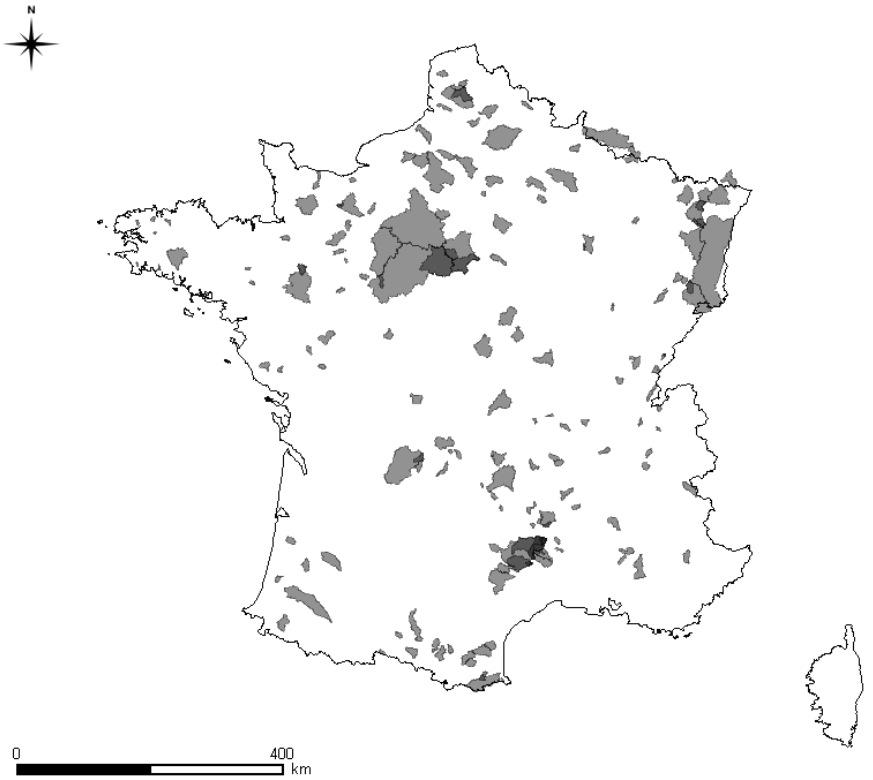


Fig. 2 Location of the 178 catchments used in this study.

Detailing the structure of the GRP model is beyond the scope of this paper; only a brief description follows.

The GRP structure was derived from that of the GR4J model (Perrin *et al*., 2003). It can classically be described as the combination of a production function followed by a routing function. The former consists of a nonlinear soil moisture accounting (SMA) store and a volume-adjustment coefficient that determines the runoff ratio. The routing function is composed of a unit hydrograph and a nonlinear routing store.

A preliminary analysis (not shown here) indicated that the most efficient updating strategy for the GRP model is a combination of a direct updating of the routing store (using the last observed discharge measurement) and an output updating based on the last observed forecast error. For the latter, simple multiplicative regressions, as well as ARIMA corrections and artificial neural networks (ANN) output updating, lead to valuable performance gains. Only these most efficient updating techniques for the GRP model are studied and compared hereafter.

METHODOLOGY

A two-step approach was chosen. First, we studied each updating technique separately, to assess typical magnitudes of their life times by theoretical or empirical means. Then, we evaluated the losses in the performance of the GRP model when the delay N between the last available discharge observation and the moment at which the forecast is issued increases. The methodology applied is described below.

Assessment of the life times of the state updates

The life times of the routing store updates are closely linked to the dynamic of this store. The routing store of the GRP model is a quadratic one: the output discharge is controlled by its level and its total capacity.

At every time step, we can represent the behaviour of the store level by equation (1):

$$S_t = f(S_{t-1} + I_t) = S_{t-1} + I_t - O_t \quad (1)$$

where S_t is the level of the store at time step t, I_t is its input and O_t its output.

After a given number (m) of time steps following an update δS_T (at time step T), the updated store reaches the level it would have without any updating (equation (2)). The life time of the update δS_T can be defined as the minimum number of time steps, m necessary for this level to be reached without updating.

$$\underbrace{f\left(I_{T+m} + f\left(I_{T+m-1} + f\left(\ldots + f\left(I_T + S_T + \delta S\right)\right)\right)\right)}_{\text{Evolution of the store level, } m \text{ time steps after the update}} \approx \underbrace{f\left(I_{T+m} + f\left(I_{T+m-1} + f\left(\ldots + f\left(I_T + S_T\right)\right)\right)\right)}_{\text{Corresponding evolution of the store level without any update}} \quad (2)$$

For each catchment, we located the time steps corresponding to the 50 largest updates δS performed by the forecasting system during the simulation period (1995–2005). From these time steps, the model was run with and without the updates until the difference between both, updated and non-updated, simulated routing store levels was negligible. Life times were thus evaluated and statistics were calculated over the 50 available values.

Assessment of the life times of the error correction updates

Output corrections are based on the use of the information contained in the model error (i.e. the error between forecast and observed discharges). They most often use a statistical relationship linking observed and forecast flows: $\varepsilon_{t+L|t} = g(Q_{t1}, Q_{t2}, ..., \hat{Q}_{t1}, \hat{Q}_{t2}, ...)$, where $\varepsilon_{t+L|t}$ is the error made on the forecast issued at time step t for a lead time L; $Q_{t1}, Q_{t2}, ...$ are observed discharges at time steps $t1$, $t2$ prior to t; and $\hat{Q}_{t1}, \hat{Q}_{t2}, ...$ are the corresponding forecast discharges. The analysis of the calibrated parameters of the different output updating techniques allowed us to estimate their life times for every catchment. For example, we assessed ARIMA updating life times as the minimum value of l such that $\rho^l \cdot \bar{\varepsilon}_1$ is negligible compared to $\bar{\varepsilon}_l$ where ρ is the calibrated autocorrelation parameter; $\bar{\varepsilon}_1$ and $\bar{\varepsilon}_l$ are the average forecast errors for lead times 1 and l hours.

Assessment of the loss in model performance when discharge data availability for updating decreases

The GRP model (using the routing store updating and an ARIMA output updating) was run on our set of 178 catchments with updating performed only with the discharge values observed at N time steps prior to the time at which the forecast is issued (Fig. 1). Forecasts for different lead times L and for different availabilities of past discharge observations, characterized by the delay N, were assessed. In our study, lead times ranged from 1 to 48 h, while N values ranged from 0 (ideal case: current discharge data are available at the time step the forecast is issued) to a maximum lag of 72 h.

We used *a posteriori* observed precipitations for the future precipitation scenarios rather than quantitative precipitation forecasts (which usually drive operational RR forecasting models), as we wanted to focus on the impact of low availability of discharge observations. In this way, no bias was introduced due to the typical decrease of skill in precipitation forecasts when increasing lead times.

The performance of the model was assessed by the root mean square error (RMSE) between observed and forecasted discharges, normalized by the RMSE computed when $N = 0$ and $L = 1$. This ratio represents the error multiplicative factor when discharge observation availability decreases and when the lead time increases. Following classical procedures advocated by Klemeš (1986), data were divided in two periods for model calibration and validation and only performances on validation data are shown.

RESULTS AND DISCUSSION

Assessments of the life times of the GRP model updates

Table 1 shows the median values (computed over the 178 catchments) of the life times obtained when considering three classes of the routing store updates among the 50 largest updates detected on the validation period: a series with the smallest updates over all catchments, a series with their median values and a series with the largest updates. On events that have long-lasting impacts, the life times of the updating procedure are typically much larger than the considered forecast lead times: for a lead time of 24 h, for instance, the median life time of the updating procedure is of 197 h (Table 1), i.e. approximately eight times greater than the forecast lead time. When considering the smallest updates, their life times are still significant. These results indicate that the updating technique considered (routing store updating) can still be useful even if the availability of the discharge observations is low (i.e. if the last assimilation of a discharge observation takes place on average long before the forecast time).

In contrast, the results from the assessment of life times of error correction updates show that the impacts of simple regression and ARIMA corrections become very limited after a small number of time steps due to the auto-regressive nature of these updating techniques: typically, after 12 time steps (hours), the correction is nil on most of the 178 catchments. The limited impact of these techniques make them not very useful for forecasting systems where data retrieval can only be done once a day.

Table 1 Median life times m (in hours) for the smallest, the median and the largest updates among the 50 largest updates on every catchment.

Lead times (h)	Smallest update	Median update	Largest update
1	44	85	166
6	44	90	175
24	51	103	197
48	57	110	224

Losses in model performance

Since we used *a posteriori* observed precipitations as future precipitation scenarios, the performances of the model do not depend directly on N and L, but only on the operational forecast length $\Delta t = N + L$. Figure 3 shows the evolution of the error multiplicative factor with Δt. This criterion first increases very rapidly with Δt. Then, the performance losses increase more slowly and become much more stable after approximately $\Delta t = 3$ days.

These results are in accordance with the previous results: the initial fast performance loss corresponds to the fast extinction of the output updating. Once the output updating is totally ineffective, the performance losses increase slowly because the extinction of the routing store updating is much slower (as shown in Table 1).

Regarding the plateau observed after $\Delta t = 3$ days, the associated performances do not correspond to the performances obtained by the model with no updating at all (shown as $N = +\infty$ on Fig. 3). In fact, even for the maximum operational forecast length ($\Delta t = 120$ h), the performance of the updated model is significantly better than those obtained with no updating on more than 75% of the studied catchments (Fig. 4). This indicates that the life time of the routing store updating is greater than five days on most of the cases: this data assimilation scheme improves the forecasts significantly, even when performed only every five days.

CONCLUSIONS

In this paper, we evaluated to what extent data assimilation techniques can efficiently improve flow forecasting models when few discharge data are available. Different updating techniques for

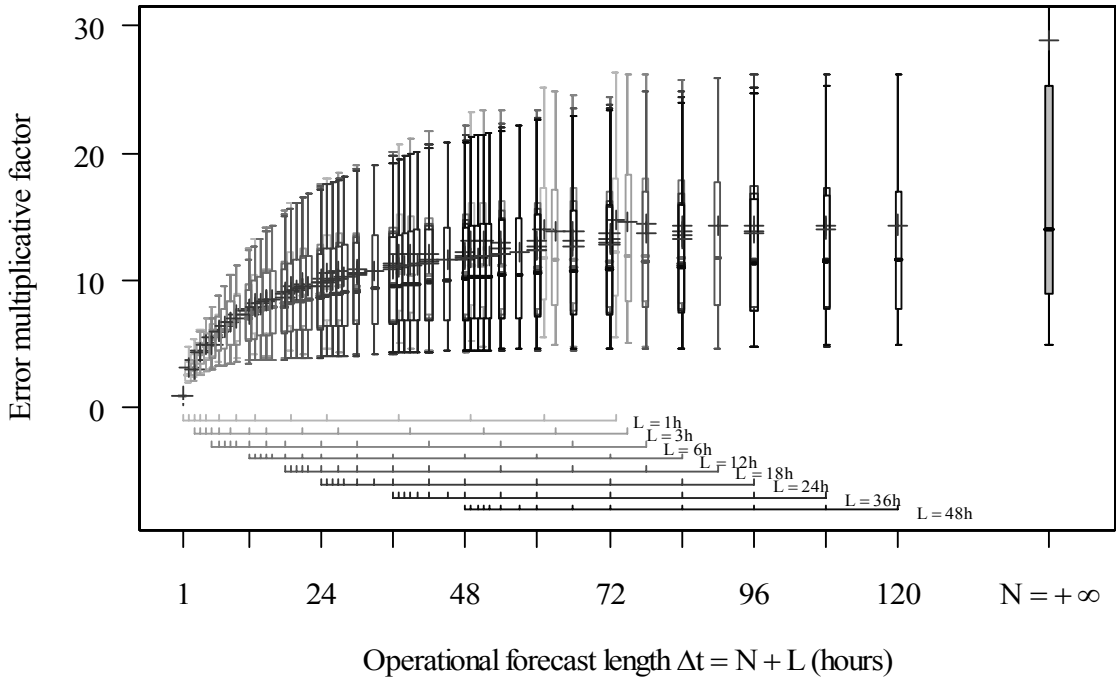


Fig. 3 Distributions (box plots) of the error multiplicative factor as a function of the operation forecast length. The box plots depict the quantiles 0.05, 0.25, 0.5 (median), 0.75 and 0.95. Crosses represent the mean of each distribution.

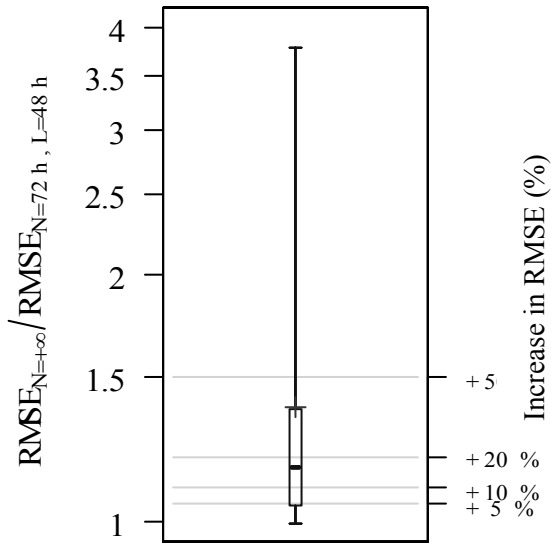


Fig. 4 Distribution (box plot) of the ratio of the RMSE obtained by the model with no updating to the RMSE obtained for Δt =120 h over the set of 178 catchments (left y-axis) and the corresponding increase of the RMSE from the case (Δt =120 h) to the situation with no assimilation at all (Δt =+∞), expressed as a fraction of the RMSE obtained for Δt =120 h (right *y*-axis). The box plot shows the quantiles 0.05, 0.25, 0.5, 0.75 and 0.95. For example, for almost 50% of the catchments, the increase of RMSE is higher than 20% of the RMSE obtained at Δt =120 h.

a simple rainfall–runoff model were compared over a set of 178 catchments: state updating proved to be effective for a longer time than output updating for the flood forecasting model studied. This updating method brings valuable improvements to the forecasts even when discharge data is available only once every five days. Further work is needed to apply this methodology on different

models and on more updating techniques. In operational forecasting, the approach presented can be useful to evaluate life times of local updates and provide guidance to data monitoring strategies in real time.

REFERENCES

Anctil, F., Perrin, C. & Andréassian, V. (2003) Ann output updating of lumped conceptual rainfall/runoff forecasting models. *J. Am. Water Resour. Assoc.* **39**, 1269–1279.

Andréassian, V., Hall, A., Chahinian, N. & Schaake, J. (2006) Introduction and synthesis: why should hydrologists work on a large number of basin data sets? In: *Large Sample Basin Experiments for Hydrological Model Parameterization* (ed. by V. Andréassian, A. Hall, N. Chahinian & J. Schaake), 1–5. IAHS Publ. 307, IAHS Press, Wallingford, UK.

Aubert, D., Loumagne, C. & Oudin, L. (2003) Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. *J Hydrol.* **280**, 145–161.

Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control.* Holden Day Inc., Oakland, California, USA.

Da Ros, D. &. Borga M. (1997) Adaptive use of a conceptual model for real time flood forecasting. *Nordic Hydrol.* **28**, 169–188.

Evensen, G. (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53**, 343–367.

Klemeš, V. (1986) Operational testing of hydrologic simulation models. *Hydrol. Sci. J.* **31**, 13–24.

Maier, H. & Dandy, G. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Software* **15**, 101–124.

Madsen, H., Butts, M., Khu, S. & Liong, S. (2000) Data assimilation in rainfall–runoff forecasting. In: *Proc. Fourth Hydroinformatics Conference*. International Association of Hydraulic Engineering and Research, University of Iowa, USA.

Moore, R. J. (2007) The PDM rainfall–runoff model. *Hydrol. Earth System Sci.* **11**, 483–499.

Perrin, C., Michel C. & Andréassian, V. (2003) Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* **279**, 275–289.

Refsgaard, J. C. (1997) Validation and intercomparison of different updating procedures for real-time forecasting. *Nordic Hydrol.* **28**, 65–84.

Tangara, M. (2005) Nouvelle méthode de prévision de crue utilisant un modèle pluie–débit global. PhD Thesis, Ecole Pratique des Hautes Etudes de Paris & Cemagref, France.

WMO (1992) Simulated real-time intercomparison of hydrological models. Operational Hydrology Report no. 38, World Meteorological Organization, Geneva, Switzerland.

O

Annexe

Compléments au chapitre 10

O.1 Traduction de l'épigraphe du chapitre 10

Quelle raison avez-vous de devenir fou ? quelle dame vous a rebuté ? ou quels indices avez-vous trouvés qui fissent entendre que ma dame Dulcinée du Toboso ait fait quelque enfantillage avec More ou chrétien ?
- Eh ! par Dieu, voilà le point, répondit don Quichotte ; et c'est là justement qu'est le fin de mon affaire. Qu'un chevalier errant devienne fou quand il en a le motif, il n'y a là ni gré ni grâce ; le mérite est de perdre le jugement sans sujet, et de faire dire à ma dame : « S'il fait de telles choses à froid, que ferait-il donc à chaud ? » D'ailleurs, n'ai-je pas un motif bien suffisant dans la longue absence qui me sépare de ma dame et toujours maîtresse Dulcinée du Toboso ? car, ainsi que tu l'as entendu dire à ce berger de l'autre jour, Ambroise : Qui est absent, tous les maux craint ou ressent. Ainsi donc, ami Sancho, ne perds pas en vain le temps à me conseiller que j'abandonne une imitation si rare, si heureuse, si inouïe. Fou je suis, et fou je dois être jusqu'à ce que tu reviennes avec la réponse d'une lettre que je pense te faire porter à ma dame Dulcinée. Si cette réponse est telle que la mérite ma foi, aussitôt cesseront ma folie et ma pénitence ; si le contraire arrive, alors je deviendrai fou tout de bon, et, l'étant, je n'aurai plus nul sentiment.
Ainsi, de quelque manière qu'elle réponde, je sortirai de la confusion et du tourment où tu m'auras laissé, jouissant du bien que tu m'apporteras, à la faveur de ma raison, ou cessant de sentir le mal, à la faveur de ma folie.

O.2 Autocorrélation pseudo-périodique des bassins à forts régimes de base

La figure 10.2 (page 177) du chapitre 8 présente l'autocorrélation des erreurs pour différents décalages temporels Δt pour nos 178 bassins. Quelques bassins se caractérisent par une auto-corrélation fortement pseudo-périodique ; parmi eux, on compte l'Essonne à Ballancourt-sur-Essonne (H4042010), la Maye à Airy (E6498315) ou encore la Durance à Val-des-Prés (X0010010). On constate en traçant ces autocorrélations pour des erreurs à différents horizons que la pseudo-période correspond au double de l'horizon de prévision H. Cette observation appelle à une interprétation.

Les bassins ayant un très fort régime de base voient leur débit varier plutôt lentement et régulièrement : sur de très nombreux pas de temps, $Q_{t+H} - Q_t \simeq Q_{t+2H} - Q_{t+H}$.

En outre, si le modèle rend bien compte du débit de ces bassins, la mise-à-jour du réservoir de routage doit assurer que $\widehat{Q}_{t+H|t} \simeq \widehat{Q}_{t|t} = Q_t$. Nous avons vérifié graphiquement que, pour les bassins mentionnés, les débits prévus $\widehat{Q}_{t+H|t}$ étaient très proches des débits observés Q_t.

Il s'en suit :

$$\begin{aligned}\rho_{\Delta t=H}\left(\varepsilon_H\right) &= corr\left(\varepsilon_H\left(t+H\right),\varepsilon_H\left(t\right)\right)\\ &= corr\left(Q_{t+H}-\widehat{Q}_{t+H|t},Q_{t+2H}-\widehat{Q}_{t+2H|t+H}\right)\\ &\simeq corr\left(Q_{t+H}-Q_t,Q_{t+2H}-Q_{t+H}\right)\end{aligned}$$

Ceci explique que la valeur de l'autocorrélation des erreurs de ces bassins pour des décalages temporels Δt égaux à H, $2H$,$3H$... soit assez forte et de signe alternatif.

O.3 Faiblesse de la correction régressive multiplicative pour des horizons lointains

La correction des sorties du modèle journalier de prévision (Tangara, 2005) est un modèle autorégressif multiplicatif du premier ordre :

$$\widehat{\widehat{Q}}_{t+H|t}=\widehat{Q}_{t+H|t}\cdot\left(\frac{Q_t}{\widehat{Q}_{t|t-1}}\right)^{\gamma}\quad\text{avec}\quad\gamma\in\left]0,1\right]$$

A priori, l'erreur de prévision (du modèle sans la correction finale) a tendance à croître avec l'horizon de prévision pour différentes raisons (incertitude plus grande sur les pluies futures, effet de la mise-à-jour du réservoir de routage qui s'estompe [1] ...). On en déduit l'existence d'un horizon de prévision (suffisamment grand) tel que :

$$\left|1-\frac{Q_{t+1}}{\widehat{Q}_{t+1|t}}\right|\ll\left|1-\frac{Q_{t+H}}{\widehat{Q}_{t+H|t}}\right|$$

Dans ce cas, les ordres de grandeurs des erreurs des prévisions à 1 heure effectuées à t et $t+1$ sont similaires (comparées aux erreurs à H heures) et on a :

$$\left|1-\frac{Q_{t}}{\widehat{Q}_{t|t-1}}\right|\ll\left|1-\frac{Q_{t+H}}{\widehat{Q}_{t+H|t}}\right|$$

L'erreur de prévision se décompose en deux termes :

$$\widehat{\widehat{Q}}_{t+H|t}-Q_{t+H}=\left(\widehat{\widehat{Q}}_{t+H|t}-\widehat{Q}_{t+H|t}\right)+\left(\widehat{Q}_{t+H|t}-Q_{t+H}\right)$$

1. On peut également considérer possible que l'amplitude de l'erreur de prévision ne croît plus quand l'effet de la mise-à-jour du réservoir de routage a disparu (retour à la simulation, ceci ne tenant qu'en supposant connues les pluies futures). L'explication théorique d'atténuation de l'effet de la correction finale de l'erreur suppose que cette atténuation a lieu plus rapidement que celle de l'effet de la mise-à-jour du réservoir de routage, ce qu'on observe expérimentalement.

Le premier terme s'écrit :

$$\Delta_t^{Corr} = \widehat{\widehat{Q}}_{t+H|t} - \widehat{Q}_{t+H|t} = \widehat{Q}_{t+H|t} \cdot \left(\left(\frac{Q_t}{\widehat{Q}_{t|t-1}} \right)^{\gamma} - 1 \right)$$

tandis que le second terme est :

$$\Delta_t^{MàJ} = \widehat{Q}_{t+H|t} - Q_{t+H} = \widehat{Q}_{t+H|t} \cdot \left(1 - \frac{Q_{t+H}}{\widehat{Q}_{t+H|t}} \right)$$

Comme $\gamma \in]0, 1]$, on a (cf. Fig. O.1) :

$$\left| \left(\frac{Q_t}{\widehat{Q}_{t|t-1}} \right)^{\gamma} - 1 \right| \leq \left| \frac{Q_t}{\widehat{Q}_{t|t-1}} - 1 \right|$$

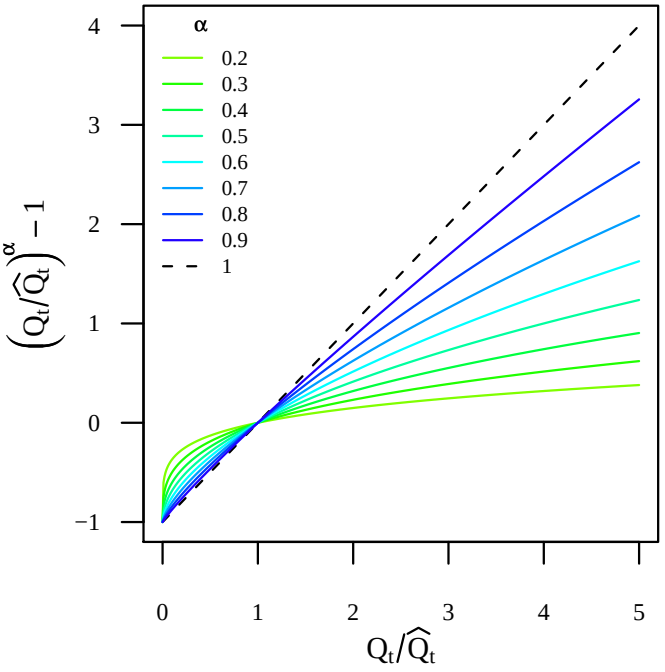


Figure O.1 – Influence de γ sur la correction finale de Tangara (2005).

D'où on déduit que $\left|\Delta_t^{Corr}\right| \ll \left|\Delta_t^{MàJ}\right|$ pour H suffisamment grand : la correction finale modifie alors très peu la sortie du modèle en comparaison de l'erreur présente dans la prévision.

Un des facteurs pouvant expliquer la plus ou moins grande rapidité de la croissance des erreurs avec l'horizon de prévision est l'autocorrélation des débits : plus les débits sont autocorrélés, plus les erreurs du modèle doivent rester petites (à un horizon donné) donc

plus H devra être grand pour observer l'effet décrit ci-dessus. Il ne s'agit que d'un facteur explicatif parmi sans doute beaucoup d'autres.

O.4 Modèles ARIMA : valeurs admissibles des paramètres

L'équation (10.1) est souvent formulée sous une écriture plus compacte :

$$\Phi(B)\widehat{w}_t = \Theta(B)a_t$$

où B est l'opérateur *arrière* défini par $B_t = z_{t-1}$ et

$$\begin{cases} \Phi(B)=1-\sum_{i=1}^{p}\varphi_i B^i \\ \Theta(B)=1-\sum_{j=1}^{q}\theta_j B^j \end{cases}$$

O.4.1 Assurer la stationnarité du modèle ARIMA

Le processus AR $\Phi(B)z_t = a_t$ est stationnaire si la série $\Psi(B) = \Phi^{-1}(B)$ converge pour $|B| \leq 1$. Cela nous impose :
pour $p = 1$, $-1 < \varphi_1 < 1$,
pour $p = 2$, $-1 < \varphi_2 < 1$ et $\varphi_2 - 1 < \varphi_1 < 1 - \varphi_2$.

Les processus MA sont par définition stationnaires quels que soient leurs paramétrages.

O.4.2 Assurer l'inversibilité du modèle ARIMA

L'inversibilité d'un processus (Box et Jenkins, 1976, p. 49 – 51) assure que z_t dépend principalement des valeurs passées proches. Un processus MA $z(t) = \Theta(B)a_t$ peut également s'écrire $\Pi(B)z_t = a_t$ où $\Pi(B) = 1 - \sum_{i=0}^{+\infty}\pi_i B^i = \Theta^{-1}(B)$. Il est inversible si la série Π converge pour $|B| \leq 1$. Cela nous impose :
pour $q = 1$, $-1 < \theta_1 < 1$,
pour $q = 2$, $-1 < \theta_2 < 1$ et $\theta_2 - 1 < \theta_1 < 1 - \theta_2$.

Les processus AR sont inversibles quels que soient leurs paramétrages.

O.5 Configuration des réseaux de neurones

Cette partie présente plus en détail les configurations possibles des réseaux de neurones.

O.5.1 Structure

Un réseau de neurones est composé d'un ensemble d'unités parmi lesquelles on distingue trois classes :
– les neurones d'entrée ont pour états les entrées du modèle,
– les neurones cachés ont des états *a priori* sans signification hydrologique ou physique,
– les neurones de sortie ont pour états les sorties du modèle.

Architecture

Il existe un très grand nombre de manières d'établir un réseau entre des unités élémentaires. Une des caractéristiques majeure est la présence ou non de boucles de connexions (par exemple, l'existence de connexions A $\rightarrow$ B, B $\rightarrow$ E et E $\rightarrow$ A). Dans ce cas, le réseau sera dit récurrent (ou réseau de Elman). S'il n'y a aucune boucle, le réseau sera qualifié de direct (*feed-forward*). Pour l'étude de séries temporelles, les réseaux récurrents peuvent présenter des effets de mémoire intéressants mais ils n'ont cependant pas prouvé une réelle supériorité sur les réseaux directs (Maier et Dandy, 2000).

Les réseaux sont le plus souvent organisés en couches de neurones successives (mais cela n'a rien d'obligatoire). On distingue alors une couche d'entrée, puis éventuellement une ou plusieurs couches cachées et enfin une couche de sortie. la couche d'entrée contient autant de neurones que le modèle compte d'entrées : les états de ces neurones sont les entrées du modèle. Dans le cas des réseaux directs, les neurones d'une couche C sont uniquement connectés vers des neurones de la couche suivante $C + 1$. Au contraire, si le réseau est récurrent, certains neurones de la couche C sont également connectés vers des neurones de couches précédentes $C - i$ (avec i positif ou nul).

De nombreuses structures de réseaux ont été testées pour des usages en hydrologie. Les plus fréquentes sont les perceptrons (c'est-à-dire des réseaux organisés en couches de neurones sans boucles de rétroaction, avec une fonction d'activation logistique pour les neurones cachés, cf. infra). Des structures alternatives comme les réseaux à régression généralisée ou les réseaux de neurones avec inférence floue ont été proposées ; Firat (2008) trouve que ces structures peuvent obtenir de meilleurs résultats que les « simples » perceptrons. Kim et Kim (2008), de même que Lekkas (2008), ont quant à eux comparé perceptrons et réseaux récurrents pour la prévision des crues (en prenant en entrées des hauteurs d'eau en amont de l'exutoire). Les performances ne semblent pas significativement différentes au regard de ces tests.

Paramétrisation et degré de connexions

Il découle des descriptions précédentes que les réseaux de neurones comptent un très grand nombre de paramètres et que ce nombre croît (beaucoup) plus vite que le nombre de neurones. Prenons l'exemple d'un réseau direct multicouche entièrement connecté (chaque neurone d'une couche C a pour entrées tous les neurones de la couche précédente $C-1$), cas très fréquent comme nous allons le voir. Si on note N_C, le nombre de neurones de la couche C, le nombre de paramètre par neurone est $N_{C-1}+1$ (les poids des neurones de la couche précédente et le biais) ; ce nombre doit être multiplié par N_C pour obtenir le nombre total de paramètres de la couche C. A titre d'exemple, si on considère un réseau ne contenant qu'une seule couche cachée, le nombre total de paramètres sera

$$N_\Theta = (N_{\text{Entrées}}+1)\cdot N_{\text{Couche cachée}} + (N_{\text{Couche cachée}}+1)\cdot N_{\text{Sortie}}$$

Soit pour le réseau de la figure 10.7 (page 192) comptant 3 entrées, 4 neurones dans la couche cachée et 1 sortie, un total de $(3+1)\cdot 4+(4+1)\cdot 1 = 21$ paramètres.

Si on ne limite pas drastiquement la structure du réseau, le nombre total de paramètres s'envole et dépasse largement le nombre de paramètres dans une modélisation conceptuelle : le réseau de neurones doit donc gérer de lourds problèmes de surparamétrisatisation. Aussi, les réseaux de neurones employés en hydrologie comptent généralement très peu de neurones : on est donc infiniment loin du fonctionnement d'un cerveau, même de celui d'un hydrologue au sortir de la nouvelle année socialement hydrologique (NASH). Il convient donc de choisir avec soin les entrées, le nombre de neurones cachés et la méthode d'apprentissage (i.e., de calage, cf. infra).

Certaines architectures sont optimisées par une réduction automatisée du nombre des connexions. En partant d'un réseau complètement connecté (comme dans l'exemple précédent), les algorithmes de taille (*pruning algorithms*) éliminent les connexions les moins pertinentes. A l'inverse, les approches *constructives* partent du plus petit réseau possible (i.e. sans couche cachée) et le développent peu à peu. Ces deux familles d'algorithmes d'optimisation des connexions peuvent se perdre dans des minima locaux (Maier et Dandy, 2000).

Nombre de neurones cachés et de couches cachées

Plutôt que de toucher au degré de connexion, on peut chercher à limiter le nombre de neurones d'un réseau entièrement connecté. Le nombre de neurones cachés dépend de la complexité de la tâche qui est demandé au réseau de neurones. Les réseaux les plus petits (parcimonieux ?) ont souvent de meilleures capacités de généralisation, tandis que les réseaux plus grands apprennent plus rapidement et peuvent mieux définir des surfaces de décision complexes. Mais un trop grand nombre de neurones cachés peut entraîner un surcalage préjudiciable aux performances en contrôle ! La relation entre nombre d'exemples

apportés au calage et le nombre de poids de connexion est également importante.

Utiliser plus d'une couche cachée améliore la flexibilité (la non-linéarité) du réseau et peut permettre l'emploi d'un nombre moins élevé de neurones cachés pour réaliser une tâche de complexité donnée. Cependant si la première couche cachée permet d'accroître fortement le potentiel des réseaux de neurones, nous verrons dans quelques instants, que l'ajout de couches cachées supplémentaires ne donne pas d'avantage théorique additionnel.

En général, le nombre de neurones cachés est optimisé par essais-erreurs.

O.5.2 Fonctionnement

Quelles entrées ?

Shamseldin (1997) compare des perceptrons prenant différentes entrées pour réaliser une prévision hydrologique. Il observe une grande sensibilité aux entrées. Le réseau de neurones disposant de toutes les entrées disponibles semble plus robuste mais ne donne pas les meilleures performances. Ce dernier résultat est confirmé par Lin et Chen (2008) qui proposent une méthode pour déterminer les entrées pertinentes d'un réseau de neurones en éliminant les entrées non pertinentes qui apportent plus de bruit que d'information au réseau de neurones.

Limiter les entrées à celles qui sont les plus pertinentes présente deux avantages : tout d'abord contenir le nombre de paramètres, ensuite ne pas introduire de bruit inutile dans le signal d'entrée. Le choix des entrées est d'abord affaire de connaissances *a priori* du modélisateur mais il est possible d'utiliser l'information du réseau de neurones pour déterminer les entrées les plus pertinentes.

Il est très difficile de déterminer directement l'importance relative des entrées par une lecture des poids calés. En effet, l'existence de la couche cachée conduit à une recombinaison de l'information jusqu'au neurone de sortie. De nombreuses études procèdent par essais-erreurs. Certains auteurs sélectionnent automatiquement les entrées une par une : au départ, des réseaux à une seule entrée (parmi une famille définie a priori) sont testés. L'entrée dont le réseau obtient les meilleurs résultats est conservée, on détermine l'entrée qui conduira aux meilleurs résultats en combinaison avec l'entrée précédente, et ainsi de suite. Cette méthode peut se révèler contre-productive car elle détermine la combinaison séquentielle des meilleures entrées et non la meilleure combinaison des entrées. Une autre approche est une étude de sensibilité conduite sur un réseau prenant toutes les entrées *a priori* potentiellement pertinentes (Chap. 8).

Dans le cadre d'un réseau direct utilisé pour l'étude de séries temporelles, se pose également la question des délais : pour prévoir la sortie $Y(t)$ quels délais Δt_i doit-on appliquer aux entrées X_i, c'est-à-dire quelles entrées $X_i(t - \Delta t_i)$ doit-on sélectionner ? Anctil *et al.* (2004a) étudient l'intérêt d'une entrée basse fréquence du réseau de neurone en simulation

hydrologique. Abrahart *et al.* (2007) pointent la non prise en compte des erreurs temporelles par les réseaux de neurones utilisés en prévision hydrologique et propose une méthode neuro-évolutive pour corriger ce défaut.

Prétraitement des données

Pour que les variables d'entrée puissent être traitées sur un pied d'égalité par le réseau de neurones (i.e. pour que le réseau puisse déterminer la pertinence des différentes combinaisons des informations fournies par les entrées), il est nécessaire de normaliser les entrées. Nous verrons dans le paragraphe suivant que les fonctions d'activation des neurones cachés ou des neurones de sortie sont généralement bornées (le plus souvent entre 0 et 1) ; il est d'usage de normaliser les entrées sur le segment $[0, 1]$, pour que les paramètres des connexions partant des neurones d'entrées soient dans une gamme similaire aux autres paramètres (voire sur $[0.1, 0.9]$ ou $[0.2, 0.8]$ pour permettre une meilleure utilisation des valeurs en dehors de la gamme des entrées normalisées).

Fonction d'activation

La fonction d'activation permet de définir l'état d'un neurone en fonction des états des N_E neurones qui lui sont connectés comme entrées : il s'agit d'une fonction de $\mathbb{R}^{N_E}$ dans $\mathbb{R}$. De nombreuses fonctions d'activation sont semi-linéaires : ce sont des compositions d'une intégration des entrées (somme pondérée) et d'une fonction f pouvant être non linéaire (cas le plus fréquent). L'une des fonctions d'activation les plus populaires est directement inspirée des fonctions d'activation des neurones biologiques : il s'agit d'une fonction semi-linéaire pour laquelle f est une sigmoïde. Ainsi, pour un neurone A ayant N_E neurones en entrée, d'états $\{X_i\}_{i\in[\![1,N_E]\!]}$ et de poids $\{w_{A,i}\}_{i\in[\![1,N_E]\!]}$, la sortie du neurone sera donnée par :

$$X_A = f\left(\sum_{i=1}^{N_E} w_{A,i}\cdot X_i + b_A\right) \quad \text{avec} \quad f(x) = \frac{1}{1+\mathrm{e}^{-x}}$$

où b_A est le biais du neurone A.

Des réseaux directs organisés en couches de neurones utilisant cette fonction d'activation pour les neurones cachés et où les neurones sont connectés vers tous les neurones de la couche suivante sont nommés *perceptrons*, ou *perceptrons multicouches* s'ils possèdent une ou plusieurs couches cachées (cf. Fig 10.7). Ces réseaux jouent un rôle très important car ils constituent une famille d'approximateurs universels des fonctions continues par morceaux (Chap. 10).

Les fonctions *à base radiale* forment une autre famille de fonctions d'activation plutôt utilisées dans la reconnaissance de motif (ou de « clustering »). Les neurones les utilisant sont caractérisés non pas par un biais et des poids de connexion mais par un vecteur dit central $\mathbf{X}_0$ de même dimension que le vecteur $\mathbf{X}$ des entrées du neurone. La sortie du

neurone sera alors fonction de la distance de l'entrée $\mathbf{X}$ au vecteur central $\mathbf{X}_0$:

$$X_A = g\left(\|\mathbf{X}_0 - \mathbf{X}\|\right)$$

où g est une fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$, décroissante et de limite nulle en ∞.

Bien d'autres fonctions d'activation existent : linéaires, polynomiales, exponentielles...

Il est à noter qu'un nombre non négligeable de réseaux présentent des neurones n'ayant pas tous le même type de fonctions d'activation (fonctions logistiques et linéaires par exemple).

O.5.3 Apprentissage

Nature de l'apprentissage

Dans le vocabulaire spécifique des réseaux de neurones (provenant des sciences cognitives), l'*apprentissage* (*training*) désigne tout simplement la méthode permettant d'assigner des valeurs aux paramètres. Quand le réseau de neurones doit établir une régression (non linéaire) entre des variables, l'apprentissage est qualifié de *supervisé* : il s'agit d'un calage.

Quand le réseau de neurones est utilisé pour repérer des motifs dans les données ou réaliser un « clustering », le calage a lieu sur des sorties qui sont également les données d'entrées (Sarle, 1994). L'apprentissage est alors dit *non supervisé*.

Algorithmes de calage

Du fait de la surparamétrisation des RNA, le calage est une étape cruciale et délicate : la surface de réponse peut être complexe (nombreux optima locaux potentiels). Différentes méthodes ont été proposées pour caler les réseaux de neurones ; certaines procèdent à une exploration locale de la surface de réponse, d'autre sont globales.

Les réseaux de neurones ont pu devenir populaires grâce à l'algorithme proposé par Rumelhart *et al.* (1986) : la rétro-propagation de la règle généralisée du delta. Il combine une règle itérative d'incrément des paramètres (couche après couche en partant de la couche de sortie pour atteindre la couche d'entrée : la rétro-propagation) où l'incrément est calculé classiquement d'après la méthode de la descente du gradient au niveau de chaque neurone (règle généralisée du delta) : les modifications sur la couche $C+1$ sont utilisées pour calculer les modifications de la couche C. Cet algorithme est très performant en temps de calcul mais peut se perdre dans des optima locaux (comme toute approche locale).

Initialement, Rumelhart *et al.* (1986) proposaient une modification des paramètres après la présentation de chaque exemple. Maier et Dandy (2000) introduisent la notion de *nombre de périodes* (*epochs*), nombre de fois où les paramètres sont modifiées lors d'une présentation de l'ensemble des exemples (du calage).

L'algorithme de Levenberg-Marquardt (cf. l'annexe O.6) ou ses (nombreuses) variantes

constituent une alternative souvent utilisée. Il s'agit d'un algorithme local qui est bien adapté aux réseaux de neurones car il s'appuie sur le jacobien de la transformation étudiée, qui est calculable analytiquement dans le cas des réseaux de neurones (pour les fonctions d'activation classiques qui sont choisies dérivables).

Le calage peut être réalisé plusieurs fois en prenant différents points de départ dans l'espace des paramètres pour éviter les extrema locaux. Les performances du réseau à la fin de chaque calage sont comparées pour déterminer la meilleure paramétrisation au final.

Lutte contre le surcalage

Comme nous l'avons vu, les réseaux de neurones emploient un grand nombre de paramètres, c'est-à-dire largement plus que le nombre maximal de paramètres communément admis pour la modélisation globale Pluie – Débit (Mein et Brown, 1978 ; Jakeman et Hornberger, 1993 ; Perrin *et al.*, 2001). Ils sont donc enclins au surcalage, c'est-à-dire à un calage où le réseau profite de ses trop nombreux degrés de liberté pour coller parfaitement aux exemples qui lui sont présentés lors du calage au prix d'une surface de réponse non représentative avec les tendances et motifs du phénomène étudié. Un modèle surcalé conduira à des résultats médiocres sur des données nouvelles, comme lors d'un contrôle : il aura une faible capacité de généralisation (que ce soit en interpolation comme en extrapolation, c'est-à-dire pour des données dans ou en dehors du domaine qui a été exploré lors du calage).

Il est donc nécessaire d'obliger le calage à « ne pas en faire trop ». Plusieurs techniques sont disponibles (Giustolisi et Laucelli, 2005). La première est le calage partiel (*early training stop*) où le calage est interrompu avant que n'ait été atteint un minimum (local) de la fonction-objectif. Cette approche est assez dangereuse car l'opérateur a peu de moyens de contrôler ce qu'il fait. Giustolisi et Laucelli (2005) montrent que cette technique de lutte contre le surcalage est loin d'être optimale. La deuxième technique est la validation croisée (Maier et Dandy, 2000) où l'opérateur dispose d'un jeu de données indépendantes de celles du calage (et de celles du contrôle) avec lesquelles il teste l'avancée du calage toutes les n itérations : il peut donc arrêter le calage avant que n'ait été découvert le meilleur jeu de paramètres (au sens des données du calage) si les performances sur les données du deuxième jeu ne s'améliorent plus. Une troisième approche est la régulation : elle consiste à pénaliser les jeux comptant des paramètres trop grands (qui conduisent à des surfaces de réponses plus chaotiques que des paramètres plus petits). Pour cela, on rajoute à la fonction-objectif un terme régulateur fonction de la magnitude des paramètres.

O.6 Calage des réseaux de neurones : algorithme de Levenberg-Marquardt avec régulation bayésienne

Pour éviter le surcalage, l'algorithme de Levenberg-Marquardt sous approximation de Gauss-Newton avec régulation bayésienne a été utilisé. Cette annexe présente la construction de cet algorithme. Nous allons présenter ici ce qui se cache derrière chacune des lettres de son acronyme : LMBR. Procédons à l'envers et commençons par le R.

O.6.1 R pour régulation

La régulation est introduite pour lutter contre le surcalage. Elle consiste à forcer le calage à choisir un jeu de paramètres plutôt petits assurant une surface de réponse assez lisse en pénalisant au sein de la fonction-objectif, les jeux de paramètres trop élevés. Nous utiliserons classiquement les moindres carrés comme fonction-objectif (de base). Avec la régulation, cette fonction-objectif devient :

$$M(\mathbf{w},\alpha,\beta) = \frac{\beta}{2}\cdot \underbrace{\sum_{t=1}^{N}\left(Q_{t+H}-\widehat{\widehat{Q}}_{t+H|t}\right)^2}_{\substack{E_D(\mathcal{D}|\mathbf{w},\mathcal{R})\\ \text{Somme des erreurs quadratiques}}} + \frac{\alpha}{2}\cdot \underbrace{\sum_{j=1}^{k} w_j^2}_{\substack{E_W(\mathbf{w}|\mathcal{R})\\ \text{Régulation}}} \tag{O.1}$$

avec :

- $\mathcal{R}$ le réseau de neurones (le modèle qu'on se choisit),
- $\mathbf{w}$ vecteur des paramètres libres du réseau de neurones,
- k nombre de paramètres libres du réseau de neurones,
- α coefficient de la régulation (à déterminer),
- β coefficient de la régulation (à déterminer),
- $\mathcal{D}$ ensemble des exemples servant au calage $\left\{Q_{t+H},\widehat{\widehat{Q}}_{t+H|t}\right\}_{t\in[\![1,N]\!]}$,
- N nombre d'exemples présentés au réseau de neurones lors de l'apprentissage.

O.6.2 B pour bayésien

Il est nécessaire de déterminer les valeurs de α et β, paramètres de la fonction-objectif. Plutôt que de les choisir a priori, ceci est fait par inférence bayésienne. La suite de cette partie est directement inspirée des travaux de MacKay (1992b).

Dans une optique bayésienne, l'emploi précédent de la somme des erreurs quadratiques comme fonction-objectif est compris comme l'utilisation de la vraisemblance en supposant que les erreurs sont indépendantes et tirées dans une loi gaussienne [2]. Nous noterons désor-

2. On a vu plus tôt dans ce mémoire ce qu'il faut penser d'une telle hypothèse en hydrologie sans transformation des variables.

mais E_D ce terme. La régulation quadratique E_W revient à introduire un *a priori* gaussien centré en 0 sur les paramètres.

L'approche bayésienne vise à inférer la distribution de probabilité de α et de β a posteriori (après avoir pris connaissance des données $\mathcal{D}$ présentées au calage). Nous nous contenterons de déterminer leurs valeurs les plus probables. Le théorème de Bayes nous donne :

$$P(\alpha, \beta|\mathcal{D}, \mathcal{R}) = \frac{P(\mathcal{D}|\alpha, \beta, \mathcal{R}) \cdot P(\alpha, \beta)}{P(\mathcal{D}|\mathcal{R})} \tag{O.2}$$

En l'absence d'*a priori* sur les coefficients α et β, le prior $P(\alpha, \beta)$ est uniforme[3] et nous devons chercher à maximiser $P(\mathcal{D}|\alpha, \beta, \mathcal{R})$.

Pour α et β connus, ce terme nous est à nouveau donné par le théorème de Bayes appliqué à la distribution de probabilité a posteriori du jeu de paramètres $\mathbf{w}$:

$$P(\mathbf{w}|\mathcal{D}, \alpha, \beta, \mathcal{R}) = \frac{P(\mathcal{D}|\mathbf{w}, \beta, \mathcal{R}) \cdot P(\mathbf{w}|\alpha, \mathcal{R})}{P(\mathcal{D}|\alpha, \beta, \mathcal{R})} \tag{O.3}$$

dont les autres termes sont facilement calculables pour E_D et E_W quadratiques. En effet,

$$P(\mathbf{w}|\alpha, \mathcal{R}) = \frac{\mathrm{e}^{-\alpha \cdot E_W(\mathbf{w}|\mathcal{R})}}{Z_W(\alpha)}$$

où la constante de normalisation se calcule simplement pour E_W quadratique :

$$Z_W(\alpha) = \int_{\mathbb{R}^k} \mathrm{e}^{-\alpha \cdot E_W(\mathbf{w}|\mathcal{R})} d^k\mathbf{w} = \left(\frac{2\pi}{\alpha}\right)^{\frac{k}{2}}$$

On obtient exactement de la même façon,

$$P(\mathcal{D}|\mathbf{w}, \beta, \mathcal{R}) = \frac{\mathrm{e}^{-\beta \cdot E_D(\mathcal{D}|\mathbf{w}, \mathcal{R})}}{Z_D(\beta)} \quad \text{avec} \quad Z_D(\beta) = \int_{\mathbb{R}^N} \mathrm{e}^{-\beta \cdot E_D(\mathcal{D}|\mathbf{w}, \mathcal{R})} d^N\mathbf{Q} = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}}$$

Enfin, on évalue la probabilité a posteriori du jeu de paramètres par :

$$P(\mathbf{w}|\mathcal{D}, \alpha, \beta, \mathcal{R}) = \frac{\mathrm{e}^{-M(\mathbf{w}, \alpha, \beta)}}{Z_M(\alpha, \beta)} \quad \text{avec} \quad Z_M(\alpha, \beta) = \int_{\mathbb{R}^k} \mathrm{e}^{-M(\mathbf{w}, \alpha, \beta)} d^k\mathbf{w}$$

3. Hypothèse *ad hoc*, mais comment s'en passer ? Ces hypothèses appellent à un contrôle de la méthode *a posteriori*.

M étant une fonction quadratique de $\mathbf{w}$, nous pouvons l'écrire sous la forme :

$$M\left(\mathbf{w},\alpha,\beta\right)=M\left(\mathbf{w^{MP}},\alpha,\beta\right)+\frac{1}{2}\left(\mathbf{w}-\mathbf{w}^{MP}\right)^{T}\mathbf{H}\left(\mathbf{w}^{MP},\alpha,\beta\right)\left(\mathbf{w}-\mathbf{w}^{MP}\right)$$

où $\mathbf{w}^{MP}$ est le jeu de paramètres minimisant M pour un couple (α,β) donné et $\mathbf{H}\left(\mathbf{w}^{MP},\alpha,\beta\right)$, la matrice hessienne [4] de $\mathbf{w}\mapsto M\left(\mathbf{w},\alpha,\beta\right)$ calculée en $\mathbf{w}^{MP}$. On en déduit la constante de normalisation :

$$Z_{M}\left(\alpha,\beta\right)=\mathrm{e}^{-M\left(\mathbf{w^{MP}},\alpha,\beta\right)}\sqrt{\frac{\left(2\pi\right)^{k}}{\det\ \mathbf{H}\left(\mathbf{w}^{MP},\alpha,\beta\right)}}$$

L'équation (O.3) nous donne donc :

$$P\left(\mathcal{D}|\alpha,\beta,\mathcal{R}\right)=\frac{\frac{\mathrm{e}^{-\alpha\cdot E_{W}\left(\mathbf{w}|\mathcal{R}\right)}}{Z_{W}\left(\alpha\right)}\cdot\frac{\mathrm{e}^{-\beta\cdot E_{D}\left(\mathcal{D}|\mathbf{w},\mathcal{R}\right)}}{Z_{D}\left(\beta\right)}}{\frac{\mathrm{e}^{-M\left(\mathbf{w},\alpha,\beta\right)}}{Z_{M}\left(\alpha,\beta\right)}}=Z_{M}\left(\alpha,\beta\right)\cdot\left(\frac{2\pi}{\beta}\right)^{-\frac{N}{2}}\cdot\left(\frac{2\pi}{\alpha}\right)^{-\frac{k}{2}}\qquad\text{(O.4)}$$

Soit

$$\begin{aligned}\lg\ P\left(\mathcal{D}|\alpha,\beta,\mathcal{R}\right)=&-\beta\cdot E_{D}\left(\mathcal{D}|\mathbf{w}^{MP},\mathcal{R}\right)-\alpha\cdot E_{W}\left(\mathbf{w}^{MP}|\mathcal{R}\right)\\&-\frac{1}{2}\cdot\lg\left(\det\ \mathbf{H}\left(\mathbf{w}^{MP},\alpha,\beta\right)\right)+\frac{k}{2}\cdot\lg\,\alpha+\frac{N}{2}\cdot\lg\,\beta\\&-\frac{N}{2}\lg\left(2\pi\right)\end{aligned}\qquad\text{(O.5)}$$

En dérivant (O.5) par rapport à α, nous avons au maximum de $P\left(\mathcal{D}|\alpha,\beta,\mathcal{R}\right)$,

$$0=-E_{W}\left(\mathbf{w}^{MP}|\mathcal{R}\right)-\frac{1}{2}\frac{d}{d\alpha}\lg\left(\det\ \mathbf{H}\left(\mathbf{w}^{MP},\alpha,\beta\right)\right)+\frac{k}{2\alpha}$$

4. La matrice hessienne est la matrice des dérivées partielles secondes :

$$H\left(\mathbf{w},\alpha,\beta\right)_{i,j}=\left.\frac{\partial M}{\partial w_{i}\partial w_{j}}\right|_{\mathbf{w},\alpha,\beta}$$

Or la matrice hessienne de $E_W(\mathbf{w}|\mathcal{R})$ est trivialement[5] l'identité $\mathbf{I}_k$, on en déduit que $\mathbf{H}(\mathbf{w}^{MP}, \alpha, \beta) = \alpha\mathbf{I}_k + \beta\nabla\nabla E_D(\mathcal{D}|\cdot, \mathcal{R})$, d'où il découle :

$$\begin{aligned}\frac{d}{d\alpha}\lg\left(\det\ \mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)\right) &= \mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\frac{d\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)}{d\alpha}\right)\\ &= \mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\mathbf{I}_k\right)\\ &= \mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\right)\end{aligned}$$

Il vient :

$$2\widehat{\alpha}\cdot E_W\left(\mathbf{w}^{MP}|\mathcal{R}\right) = k - \widehat{\alpha}\cdot\mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \widehat{\alpha}, \widehat{\beta}\right)^{-1}\right) \tag{O.6}$$

De même en annulant la dérivée de (O.5), nous obtenons :

$$0 = -E_D\left(\mathcal{D}|\mathbf{w}^{MP}, \mathcal{R}\right) - \frac{1}{2}\frac{d}{d\beta}\lg\left(\det\ \mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)\right) + \frac{N}{2\beta}$$

La dérivée du deuxième terme s'obtient comme précédemment par :

$$\begin{aligned}\frac{d}{d\beta}\lg\left(\det\ \mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)\right) &= \mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\frac{d\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)}{d\beta}\right)\\ &= \mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\nabla\nabla E_D\left(\mathcal{D}|\cdot, \mathcal{R}\right)\right)\\ &= \frac{1}{\beta}\mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right) - \alpha\mathbf{I}_k\right)\right)\\ &= \frac{1}{\beta}\cdot\left(k - \alpha\cdot\mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \alpha, \beta\right)^{-1}\right)\right)\end{aligned}$$

D'où il suit que :

$$2\widehat{\beta}\cdot E_D\left(\mathcal{D}|\mathbf{w}^{MP}, \mathcal{R}\right) = N - k + \widehat{\alpha}\cdot\mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \widehat{\alpha}, \widehat{\beta}\right)^{-1}\right) \tag{O.7}$$

La quantité $\gamma = k - \widehat{\alpha}\cdot\mathrm{Tr}\left(\mathbf{H}\left(\mathbf{w}^{MP}, \widehat{\alpha}, \widehat{\beta}\right)^{-1}\right)$ varie entre 0 et k, et indique le nombre de paramètres bien définis (MacKay, 1992a). Les équations (O.6) et (O.7) peuvent également s'écrire :

$$2\widehat{\alpha}\cdot E_W\left(\mathbf{w}^{MP}|\mathcal{R}\right) = \gamma \quad\text{et}\quad 2\widehat{\beta}\cdot E_D\left(\mathcal{D}|\mathbf{w}^{MP}, \mathcal{R}\right) = N - \gamma \tag{O.8}$$

MacKay (1992b) fait remarquer que pour les réseaux de neurones, une seule valeur de α

5. $\dfrac{\partial^2}{\partial w_i \partial w_j}\left(\dfrac{1}{2}\sum_{l=1}^{k} w_l^2\right) = \delta_{i,j}$

peut ne pas être optimale car les poids des différentes connexions peuvent être très différents (notamment entre deux paires différentes de couches). Le prior devient alors irréaliste. On peut améliorer les résultats en proposant plusieurs valeurs de α (par exemple, une valeur par couche) et en choisissant pour prior

$$E_W = \frac{1}{2} \sum_{j=1}^{c} \alpha_j \sum_{i=1}^{k_j} w_{j,i}^2$$

Cependant, les résultats qu'il présente avec ce nouveau prior ne sont pas très différents des précédents et cette modification augmente encore un peu la complexité de la démarche (et la quantité de calculs). Comme notre travail ne cherche pas à obtenir absolument le meilleur réseau de neurones possible pour la tâche que nous nous sommes assignée, nous garderons le prior « simple ».

O.6.3 LM pour Levenberg-Marquardt

Une fois la fonction-objectif bien définie, il nous reste à choisir un algorithme de calage. Un algorithme local puissant est celui proposé par Marquardt à la suite de travaux de Levenberg. Pour en comprendre son fonctionnement, partons de l'algorithme de Newton.

Algorithme de Newton

L'algorithme de Newton est un algorithme de calage local du 2e ordre. Partant d'un jeu de paramètres initial $\mathbf{w}_0$, il modifie de façon itérative ce jeu de paramètres d'un incrément ($\mathbf{w}_{l+1} = \mathbf{w}_l + \delta\mathbf{w}_l$). Cet incrément est calculé en approximant la fonction à minimiser par une fonction quadratique[6] :

$$\mathbf{H}\left(\mathbf{w}_l, \alpha, \beta\right) \delta\mathbf{w}_l = -\nabla M\left(\mathbf{w}_l, \alpha, \beta\right) \tag{O.9}$$

où $\mathbf{H}\left(\mathbf{w}_l, \alpha, \beta\right)$ est la matrice hessienne de la fonction-objectif M calculée en $\mathbf{w}_l$.

Amélioration pratique : l'algorithme de Levenberg-Marquardt

L'algorithme de Newton a un avantage majeur sur les méthodes d'optimisation locales du premier ordre : sa vitesse de convergence est bien supérieure... quand il converge ! En effet, s'il y a convergence alors ce sera bien vers un minimum (local) de M, mais la convergence n'est pas assurée : on peut observer des oscillations. Il n'y aura convergence que si le point de départ $\mathbf{w}_0$ est suffisamment proche d'un minimum local.

L'algorithme de Levenberg-Marquardt allie les avantages des algorithmes des premier et deuxième ordres : loin d'un optimum local l'algorithme se comporte comme une descente

6. Cela correspond à déterminer un 0 de la différentielle de M par la méthode de Newton de résolution des équations non linéaires.

du gradient (algorithme du premier ordre dont la convergence est assurée mais qui est moins rapide) tandis qu'à l'approche d'un optimum local, il adopte de plus en plus le comportement de l'algorithme de Newton. Cette propriété est assurée en utilisant une modification de l'équation (O.9) pour déterminer l'incrément :

$$\left(\mathbf{H}\left(\mathbf{w}_l, \alpha, \beta\right) + \mu \mathbf{I}_k\right) \delta \mathbf{w}_l = -\nabla M\left(\mathbf{w}_l, \alpha, \beta\right) \tag{O.10}$$

où $\mathbf{I}_k$ est la matrice identité de dimension k et μ est un facteur qui est modifié automatiquement au cours de la recherche pour privilégier l'un ou l'autre mode de recherche : loin de l'optimum, μ sera choisi grand et l'algorithme de Levenberg-Marquardt s'approchera de la descente du gradient ; à l'approche de l'optimum, μ décroîtra et l'algorithme se comportera comme celui de Newton.

Approximation de Gauss-Newton

Un autre défaut de l'algorithme de Newton est qu'il emploie une approximation quadratique : le calcul de la matrice hessienne peut être douloureux pour les systèmes présentant un grand nombre de paramètres, en effet si k est le nombre de paramètres, la complexité du calcul de son déterminant est $O\left(k^3\right)$. Comme cette matrice intervient également dans l'inférence bayésienne donnant α et β, on utilise une approximation classique dans l'emploi pratique de l'algorithme de Levenberg-Marquardt pour des fonctions-objectifs quadratiques : la matrice hessienne $\mathbf{H}$ est remplacée par le produit $\mathbf{J_M}^{\mathbf{T}}\mathbf{J_M}$ où $\mathbf{J_M}$ est la matrice jacobienne de M. Plus exactement, comme nous savons que le second terme de M (Eq. (O.1)) présente une matrice hessienne évidente (cf. O.6.2), on approxime la matrice hessienne de M par :

$$\begin{aligned}\mathbf{H}\left(\mathbf{w}, \alpha, \beta\right) &= \nabla\nabla M \\ &= \alpha \cdot \nabla\nabla E_D + \beta \cdot \nabla\nabla E_W \\ &\simeq \alpha \cdot \mathbf{J}^{\mathbf{T}}\mathbf{J} + \beta \cdot \mathbf{I}_k\end{aligned}$$

où $\mathbf{J}$ est la matrice jacobienne de E_D :

$$\mathbf{J}_{1i} = \frac{\partial E_D}{\partial w_i}$$

L'incrément du jeu de paramètres dans l'algorithme de Levenberg-Marquardt s'écrira donc :

$$\begin{aligned}\Delta\mathbf{w} &= \left(\alpha \cdot \mathbf{J}^{\mathbf{T}}\mathbf{J} + (\beta + \mu) \cdot \mathbf{I}_k\right)^{-1} \nabla M \\ &= \left(\alpha \cdot \mathbf{J}^{\mathbf{T}}\mathbf{J} + (\beta + \mu) \cdot \mathbf{I}_k\right)^{-1} (\alpha \nabla E_D + \beta \mathbf{w})\end{aligned}$$

O.6.4 Implémentation de l'algorithme

Le calage est réalisé plusieurs fois par l'algorithme d'optimisation partant de jeux de paramètres initiaux différents.

L'implémentation que nous utilisons (routine `ANN_LMBR` de la contribution Scilab HydroGR[7] due à Julien Lerat) suit presque entièrement les préconisations de Foresee et Hagan (1997) :

1. Choisir les valeurs initiales des paramètres : $\mathbf{w}_0$.
2. En déduire les valeurs initiales de α et β :
$$\alpha_0 = \frac{k}{2E_W(\mathbf{w}_0)} \quad \text{et} \quad \beta_0 = \frac{N-k}{2E_D(\mathbf{w}_0)}$$
3. Itérer la recherche de l'optimum selon l'équation (O.10) :
   (a) Calculer l'approximation de la matrice hessienne $\mathbf{H}(\mathbf{w}_l, \alpha_l, \beta_l)$ puis déterminer l'incrément $\delta\mathbf{w}_l$.
   (b) Si cet incrément conduit à une diminution de M, conserver le nouveau jeu de paramètres et diminuer μ (se rapprochant ainsi de l'algorithme de Gauss-Newton).
   Sinon, garder l'ancien jeu de paramètres et augmenter μ jusqu'à trouver un incrément qui diminue M.
   (c) Une fois un nouveau jeu de paramètres obtenu, estimer les valeurs de γ_{l+1}, α_{l+1} et β_{l+1} par
$$\gamma_{l+1} = k - 2\alpha_l \cdot \mathrm{Tr}\left(\mathbf{H}(\mathbf{w}_l, \alpha_l, \beta_l)^{-1}\right)$$
$$\alpha_{l+1} = \frac{\gamma_{l+1}}{2E_W(\mathbf{w}_{l+1})} \quad \text{et} \quad \beta_{l+1} = \frac{N-\gamma_{l+1}}{2E_D(\mathbf{w}_{l+1})}$$
   (d) Arrêter s'il y a convergence ou si μ devient trop grand ou trop petit.

Cette implémentation peut ne pas paraître découler directement de la discussion de l'inférence bayésienne : l'estimation de α et β est ici faite à chaque itération de l'algorithme de Levenberg-Marquardt alors qu'elle n'est valable que proche d'un minimum (local). Cependant, à chaque nouvelle estimation du jeu de paramètres $\mathbf{w}_l$ se rapprochant de l'optimum, l'estimation qu'on peut en déduire de α et β est plus précise, ce qui nous pousse à la réaliser à chaque itération (Foresee et Hagan, 1997).

7. Disponible et documentée sur `http://lerat.julien.free.fr/CONT_FR/HYDROGR_FR.html`

P

Annexe

Performances du modèle GRP sur 1024 bassins versants français : cartes

Cette annexe regroupe les cartes représentant les valeurs des différent critères de performance obtenues par les modèles GRP-RNA et GRP-ARIMA sur 1024 bassins français (Chap. 13). Ces critères sont

- Critères $C_{2M,P}$: pour les débits supérieurs au quantile 0.90, pour les débits supérieurs au quantile 0.99, pour les débits supérieurs au quantile 0.90, en phase de montée,
- Critère de retard à la montée de crue,
- Critères événementiels : indice de succès critique (CSI), probabilité de détection (POD) et taux de fausses alertes (FAR) pour le dépassement du seuil $q_{99}(Q)$.

Les cartes de performance sont données pour quatre horizons de prévision différents : 1, 6, 24 et 48 heures. Trois cartes sont disposées sur chaque page : elles correspondent à un critère de performance et à un horizon de prévision donnés. Les cartes en haut de la page donnent les performances du modèle GRP-ARIMA (à gauche) et GRP-RNA (à droite). La troisième carte représente les différences entre les valeurs du critère obtenues par les modèles GRP-RNA et GRP-ARIMA.

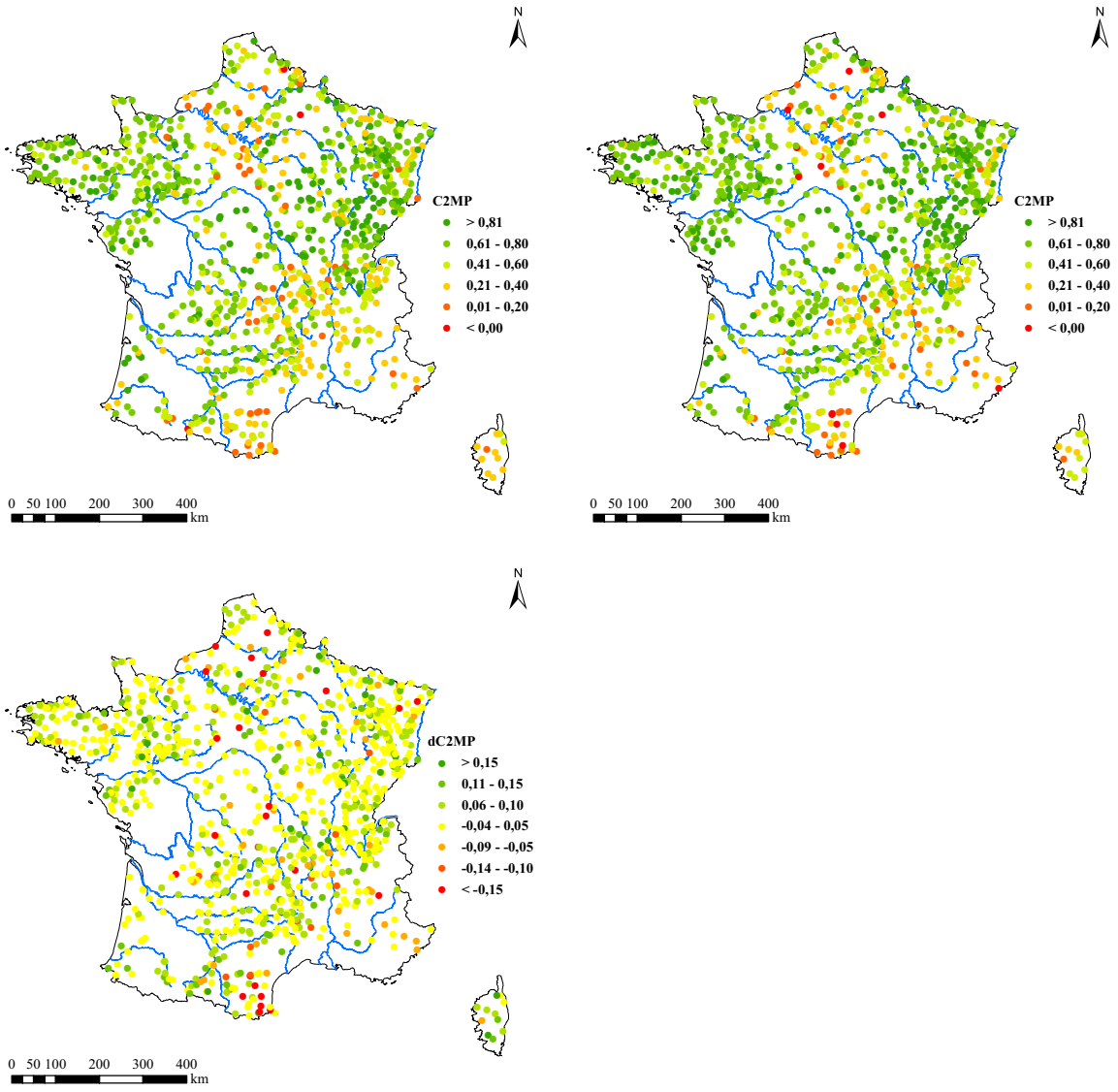


FIGURE P.1 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 ; Horizon : 1 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

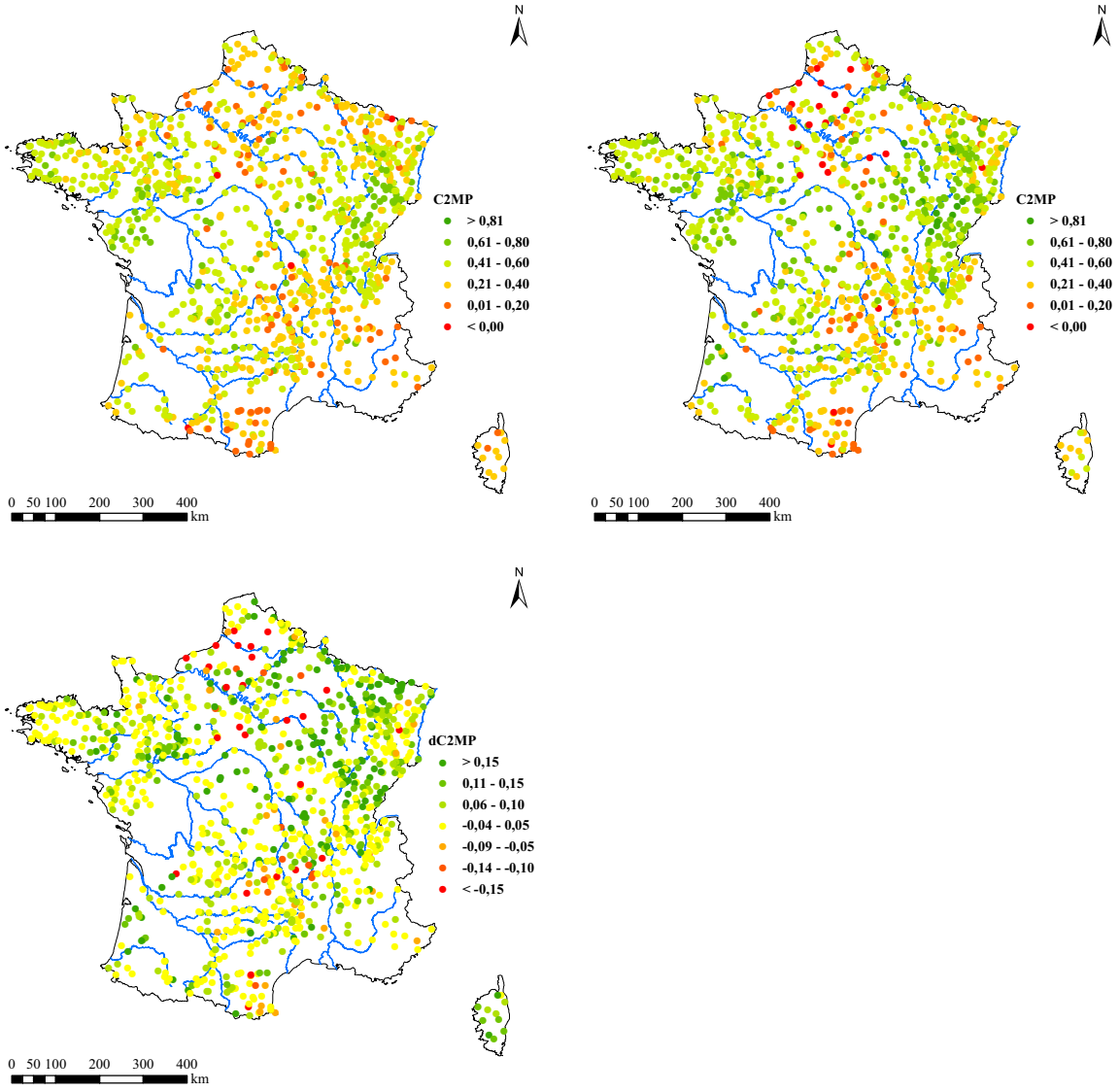


FIGURE P.2 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 ; Horizon : 6 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

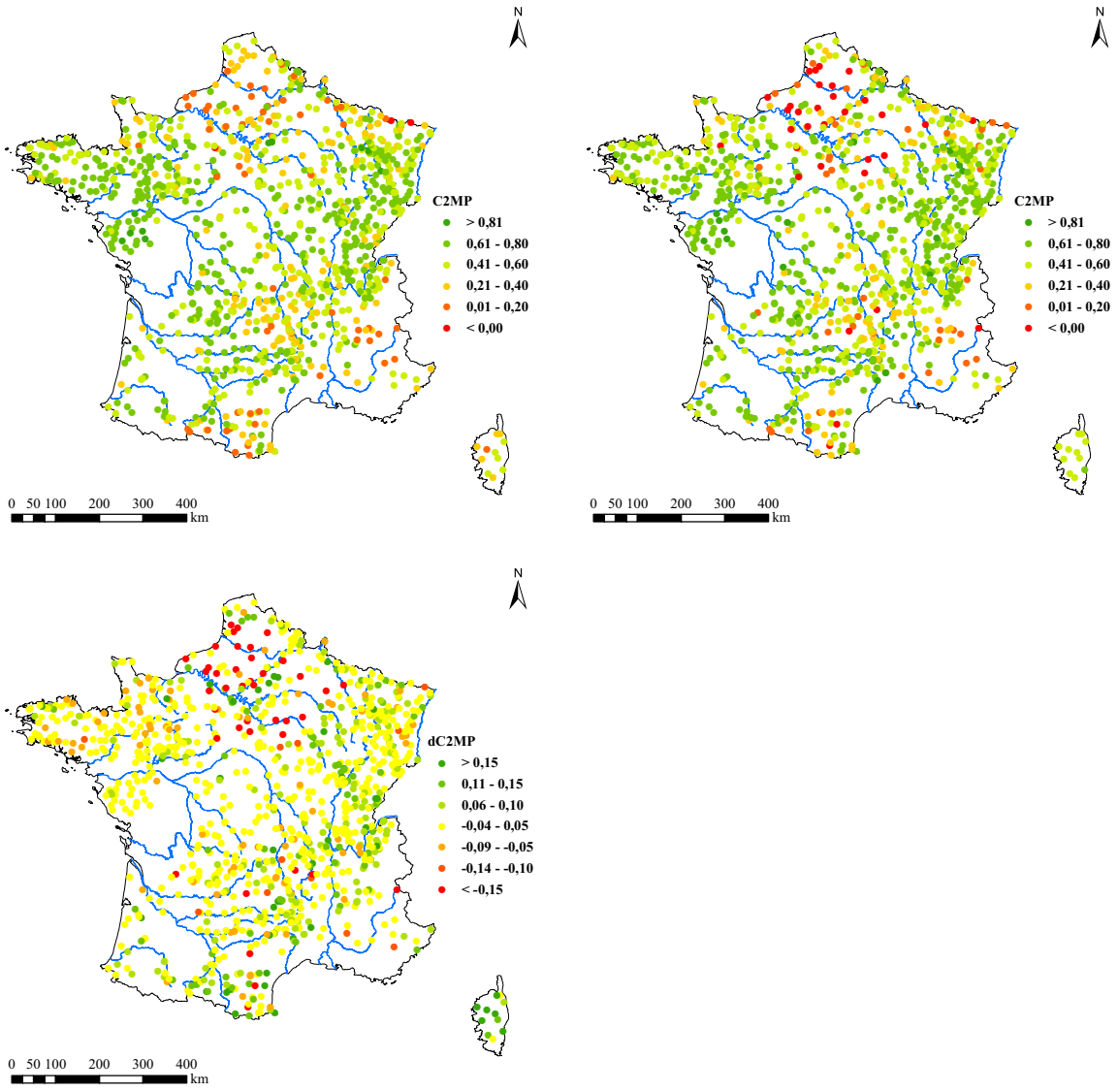


FIGURE P.3 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 ; Horizon : 24 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

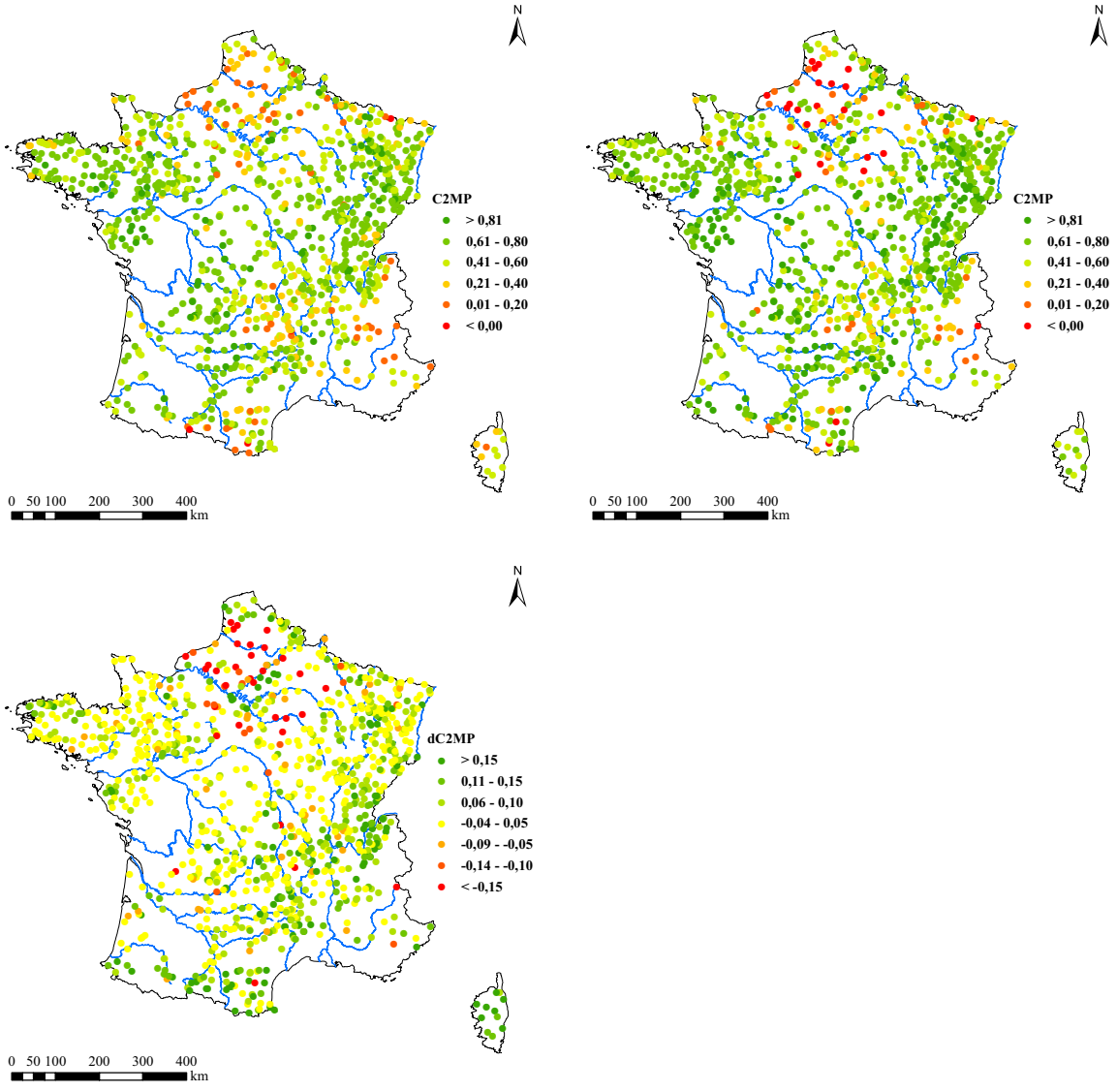


FIGURE P.4 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 ; Horizon : 48 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

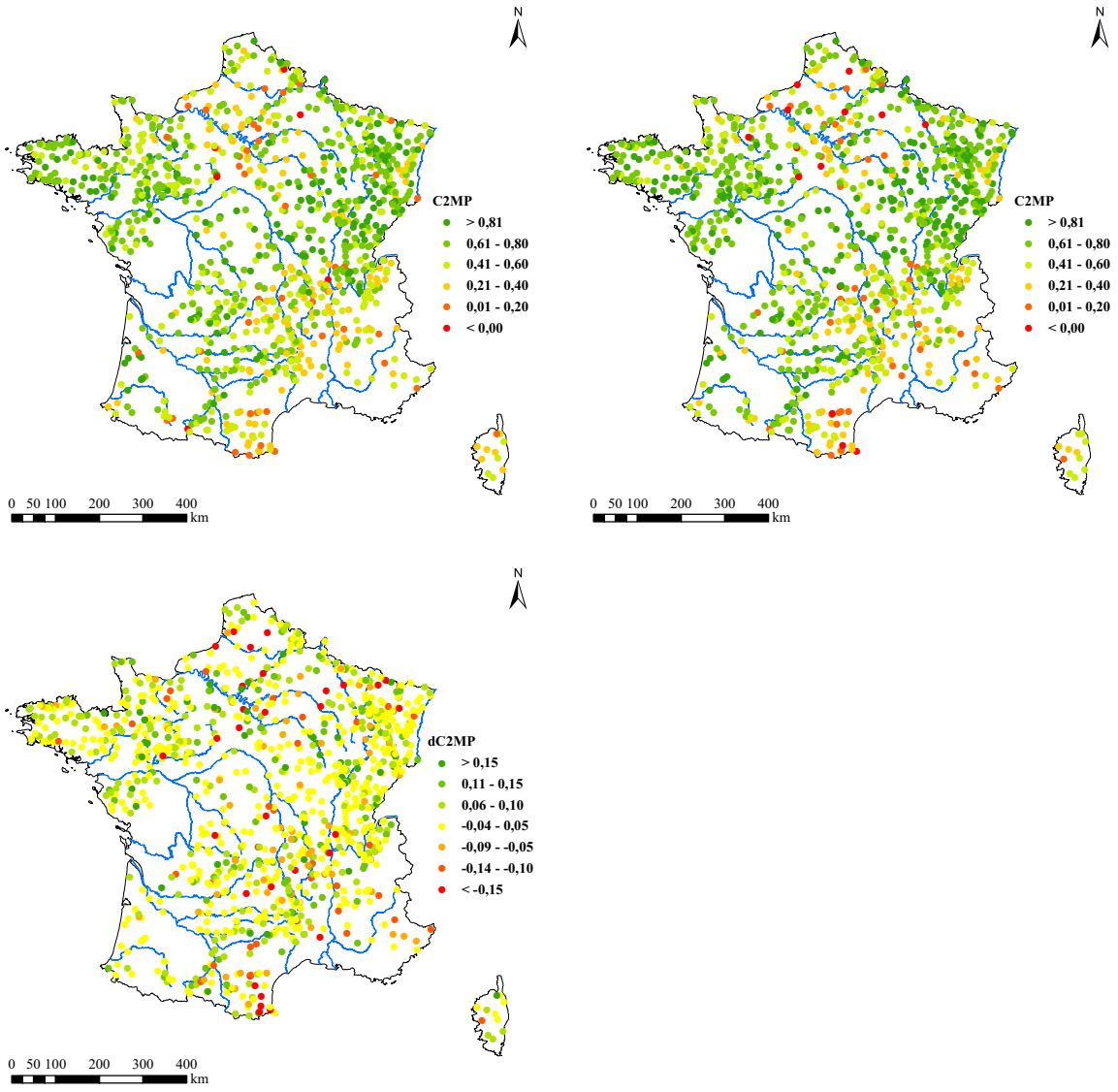


FIGURE P.5 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 ; Horizon : 1 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

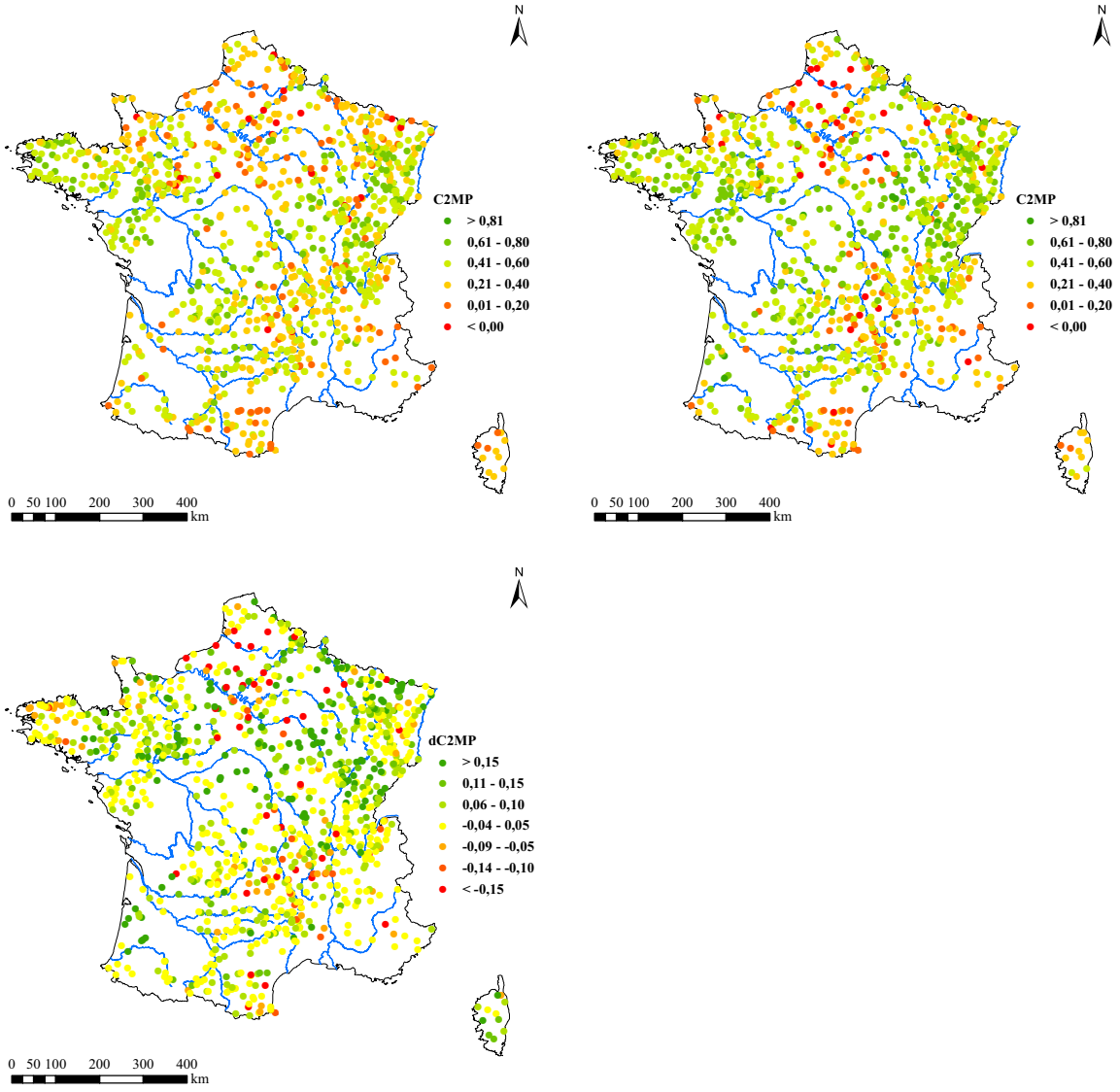


FIGURE P.6 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 ; Horizon : 6 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

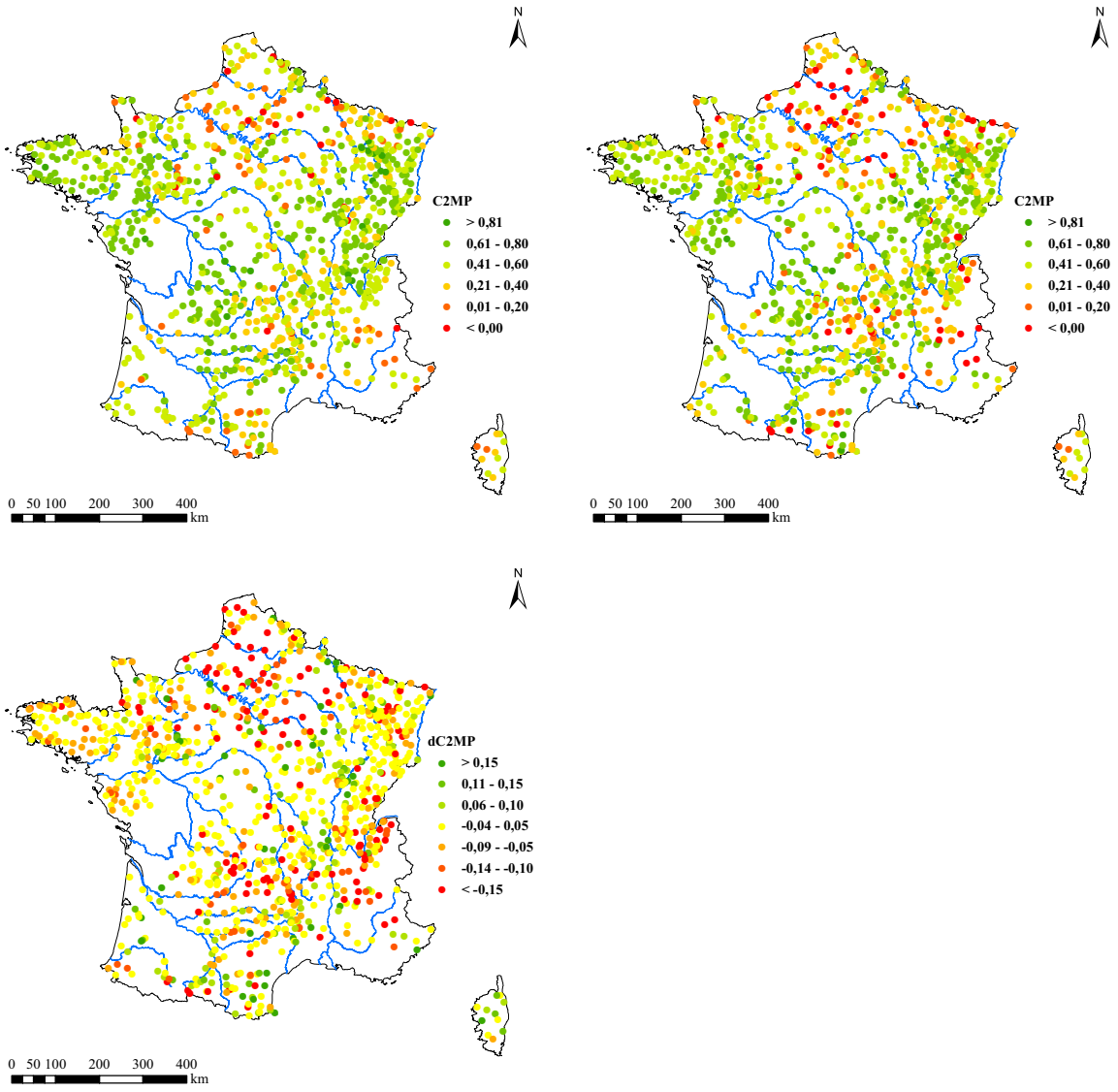


Figure P.7 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 ; Horizon : 24 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

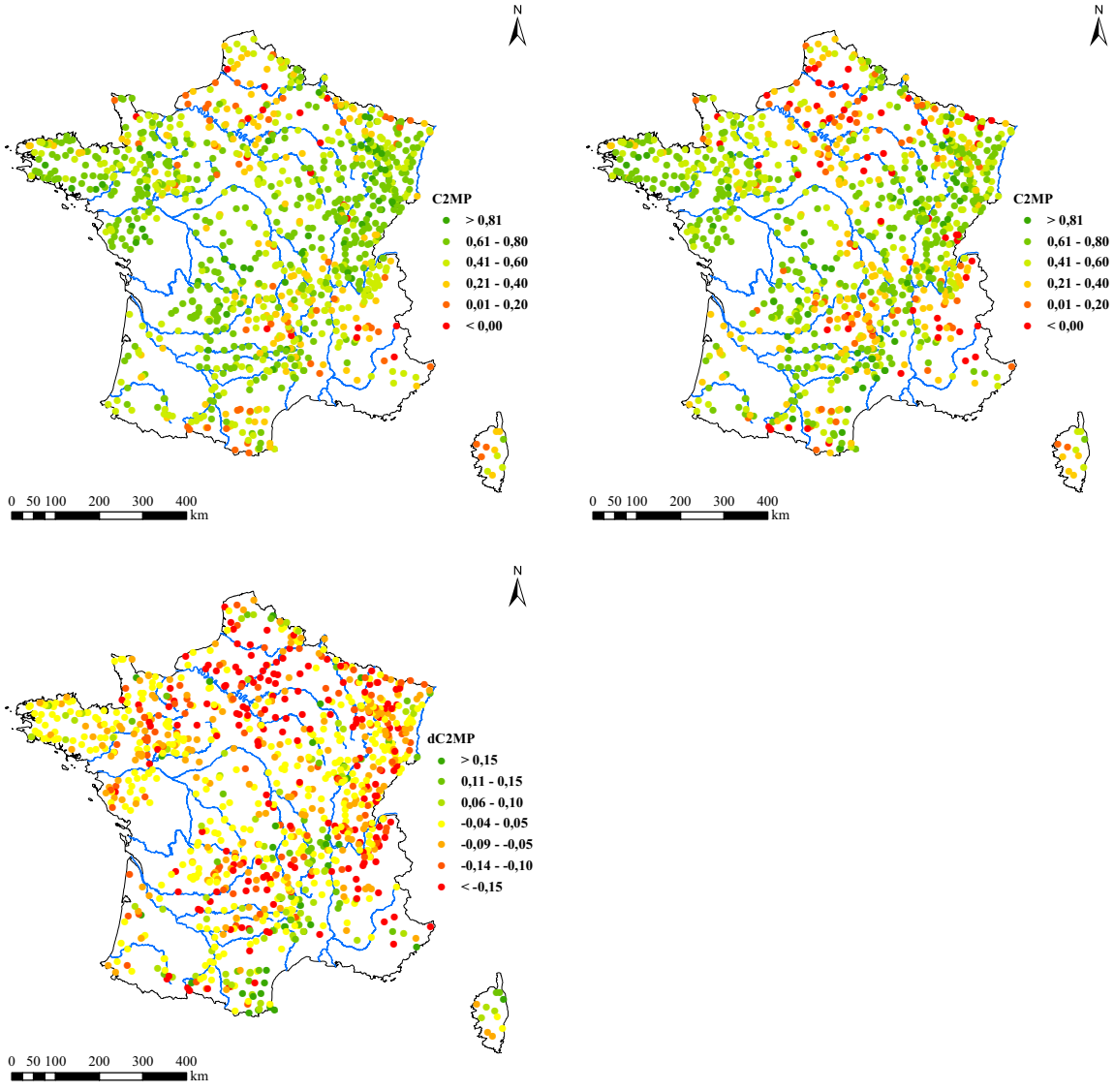


FIGURE P.8 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.99 ; Horizon : 48 h. En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

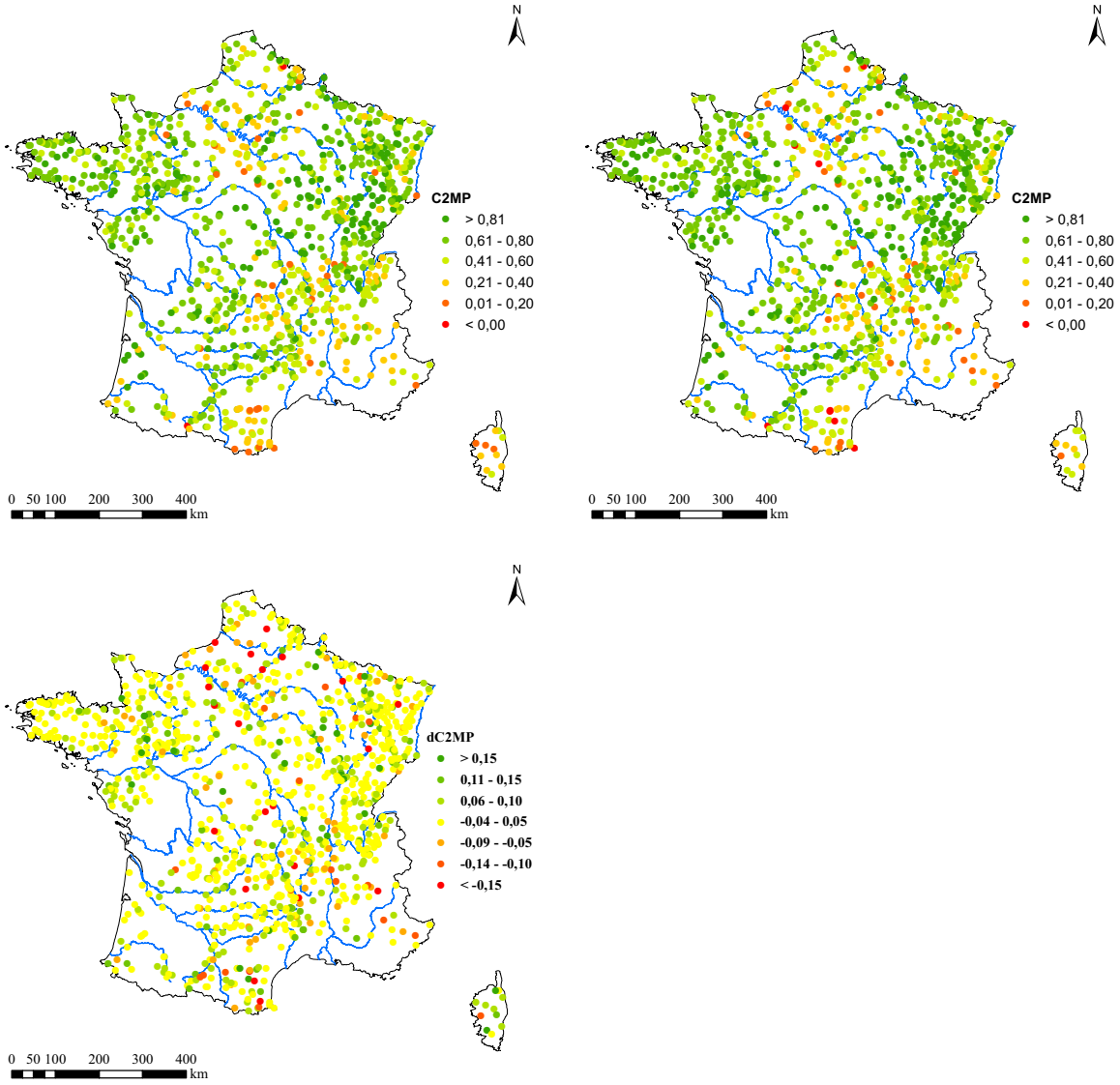


FIGURE P.9 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en phase de montée ; Horizon : 1 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

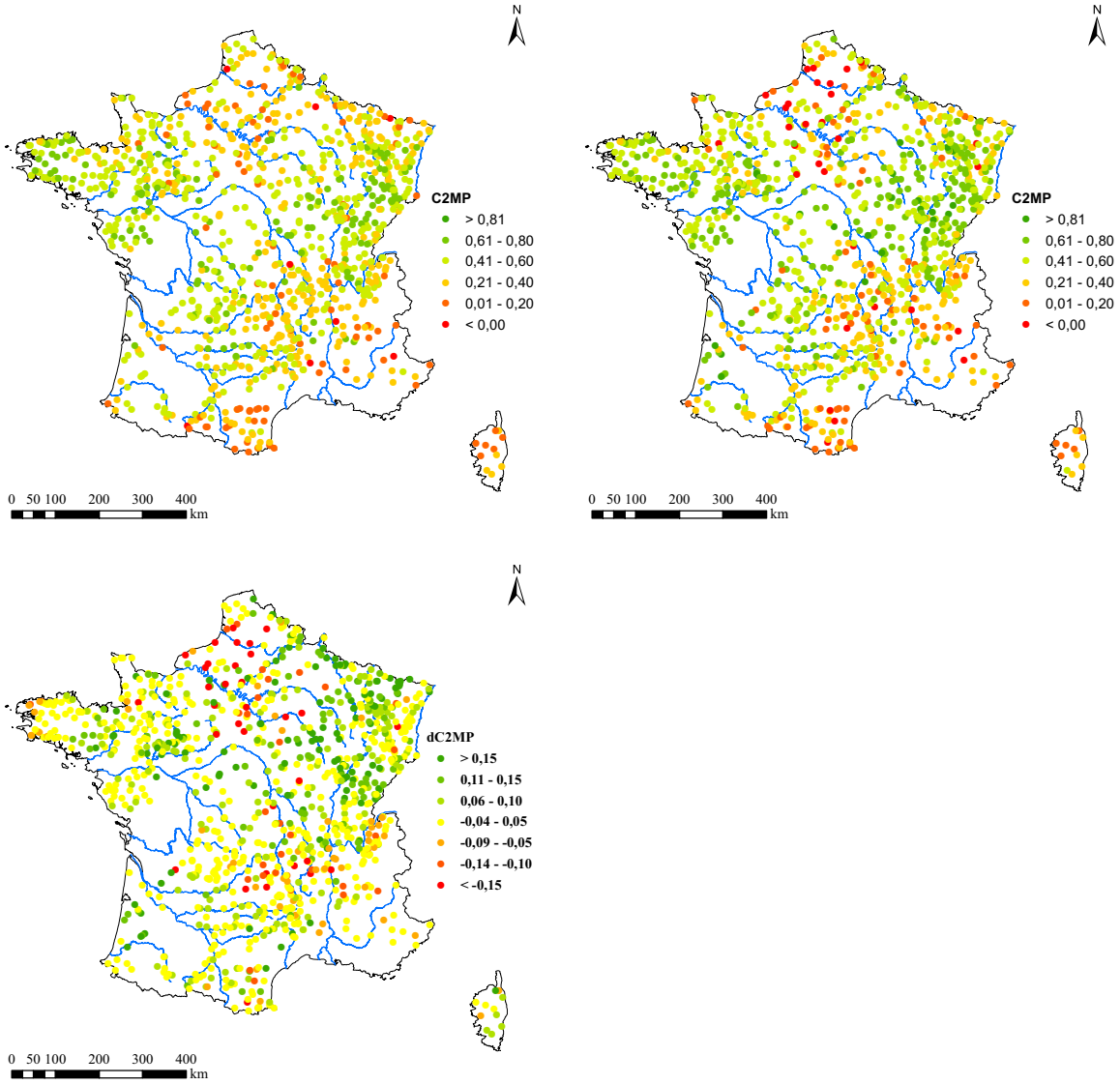


FIGURE P.10 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en phase de montée ; Horizon : 6 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

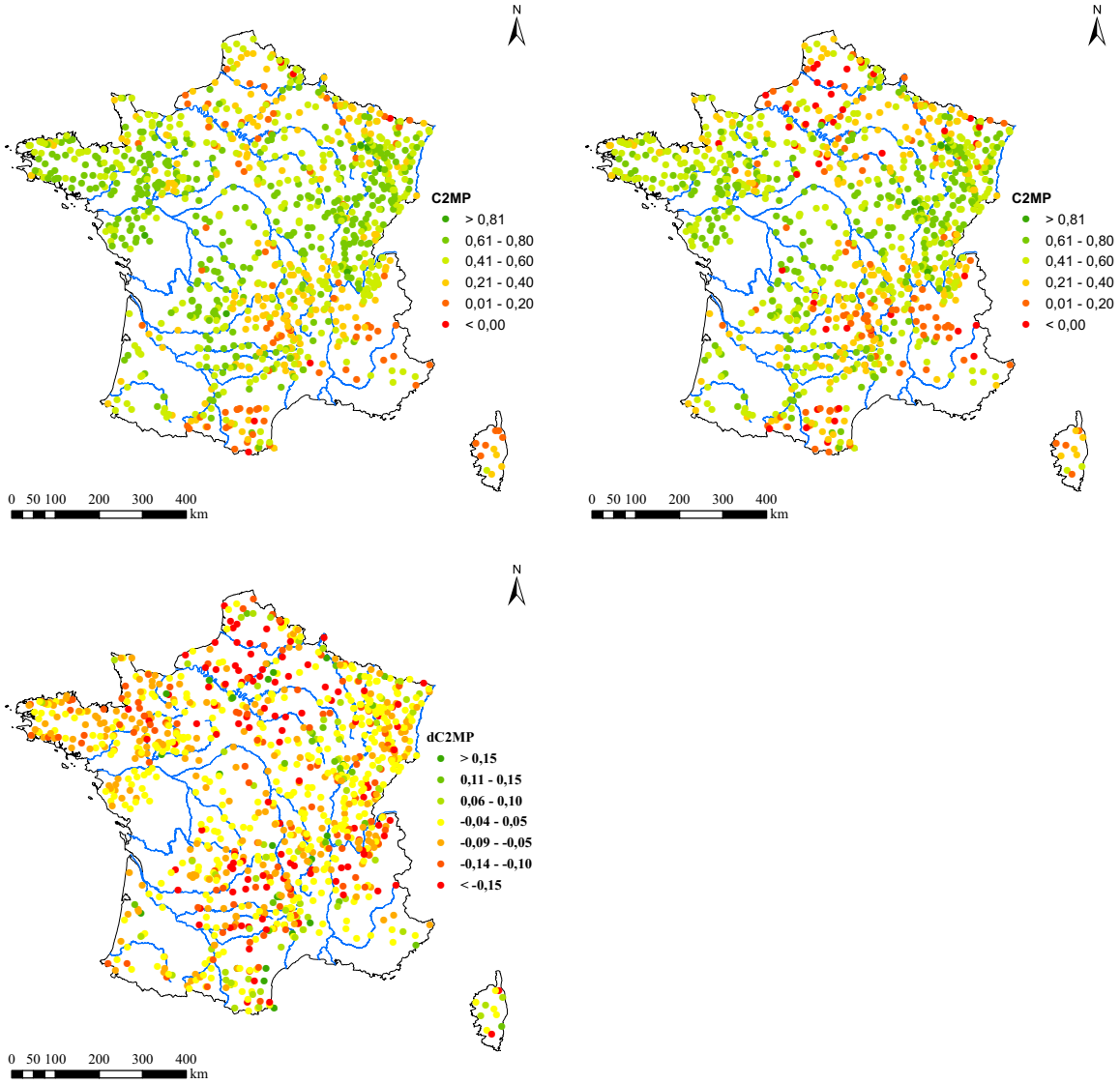


FIGURE P.11 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en phase de montée ; Horizon : 24 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

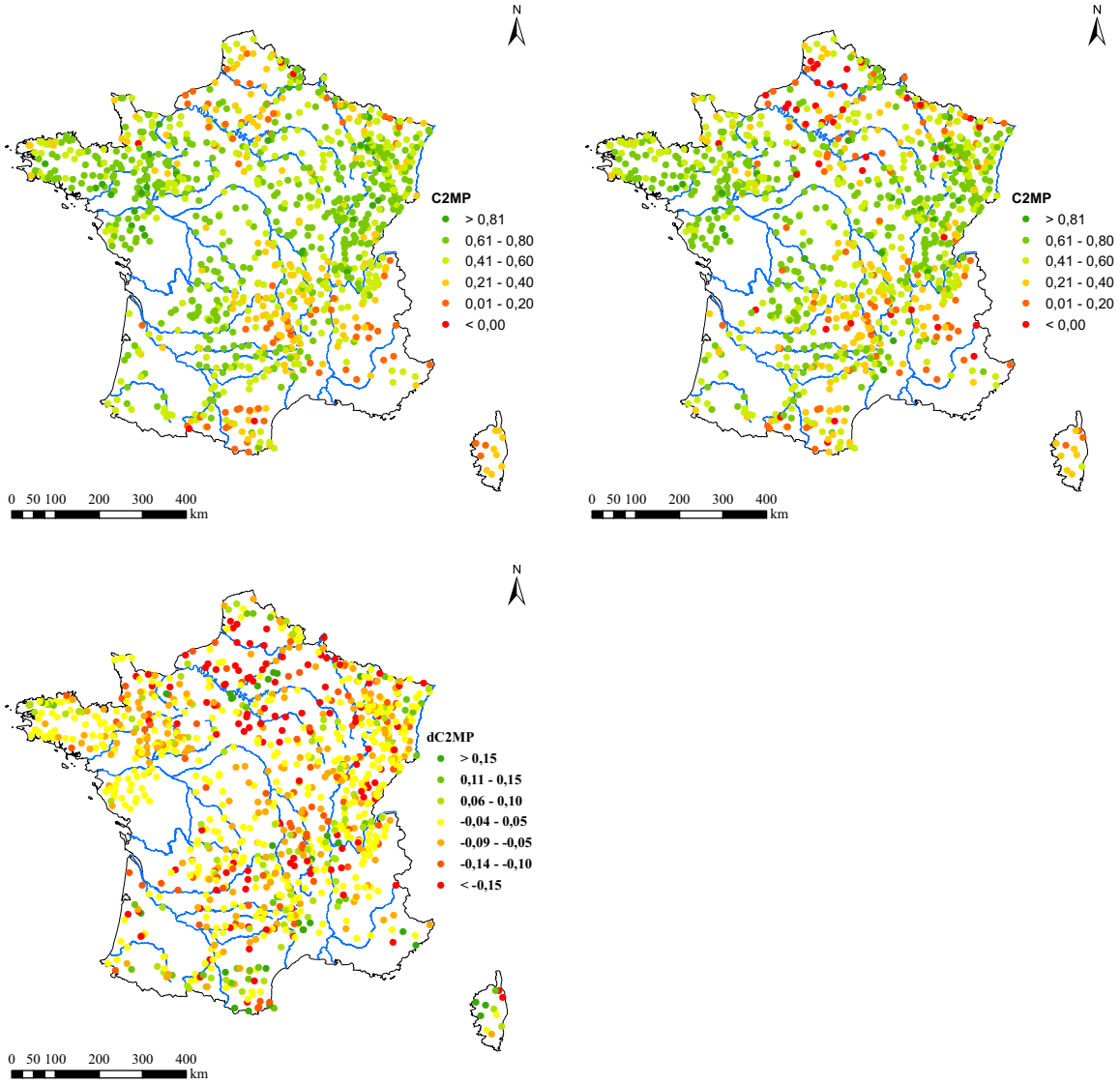


FIGURE P.12 – Critère : $C_{2M,P}$ pour les débits supérieurs au quantile 0.90 en phase de montée ; Horizon : 48 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

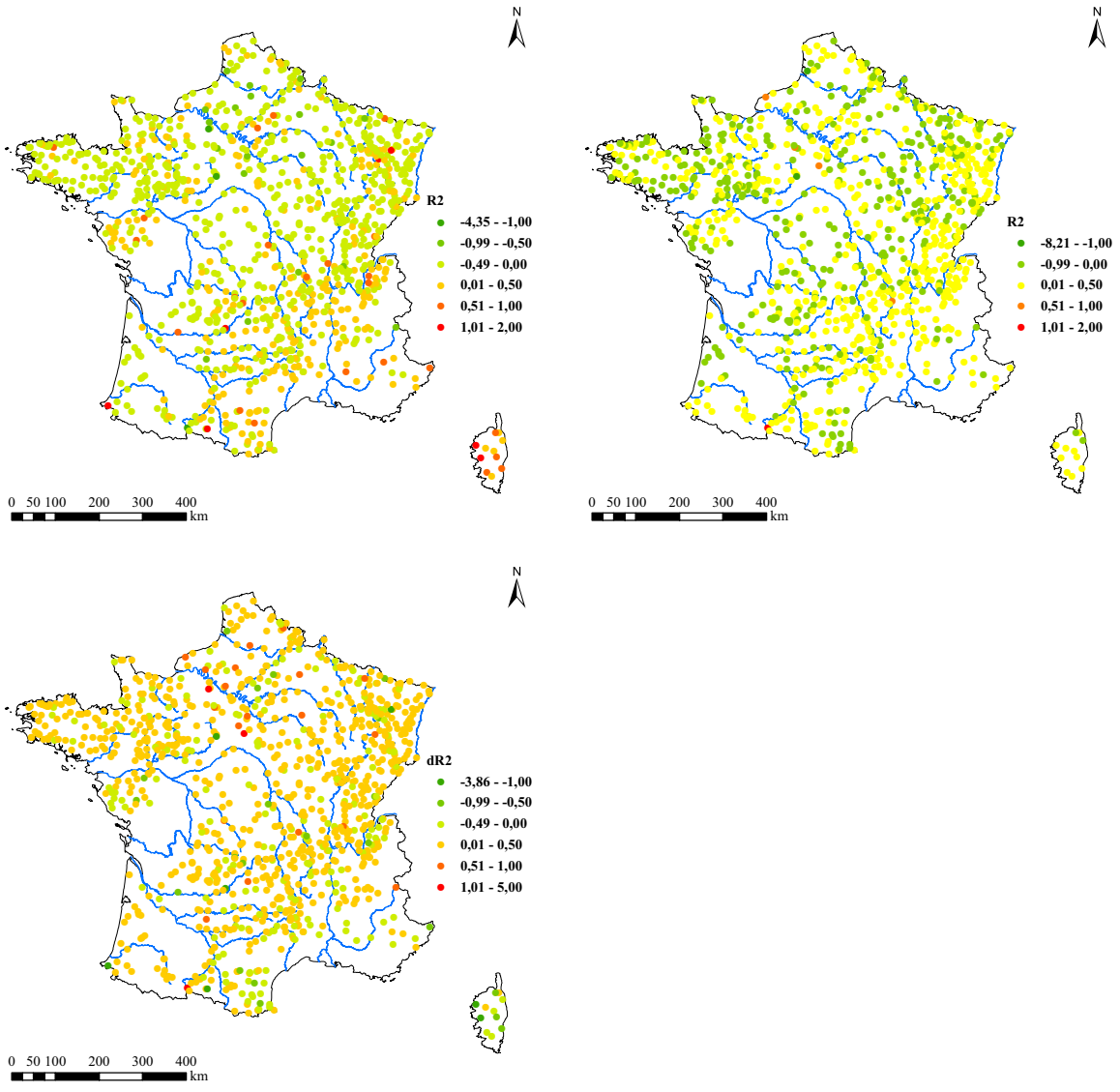


FIGURE P.13 – Critère : retard moyen à la montée de crue R_2 ; Horizon : 1 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

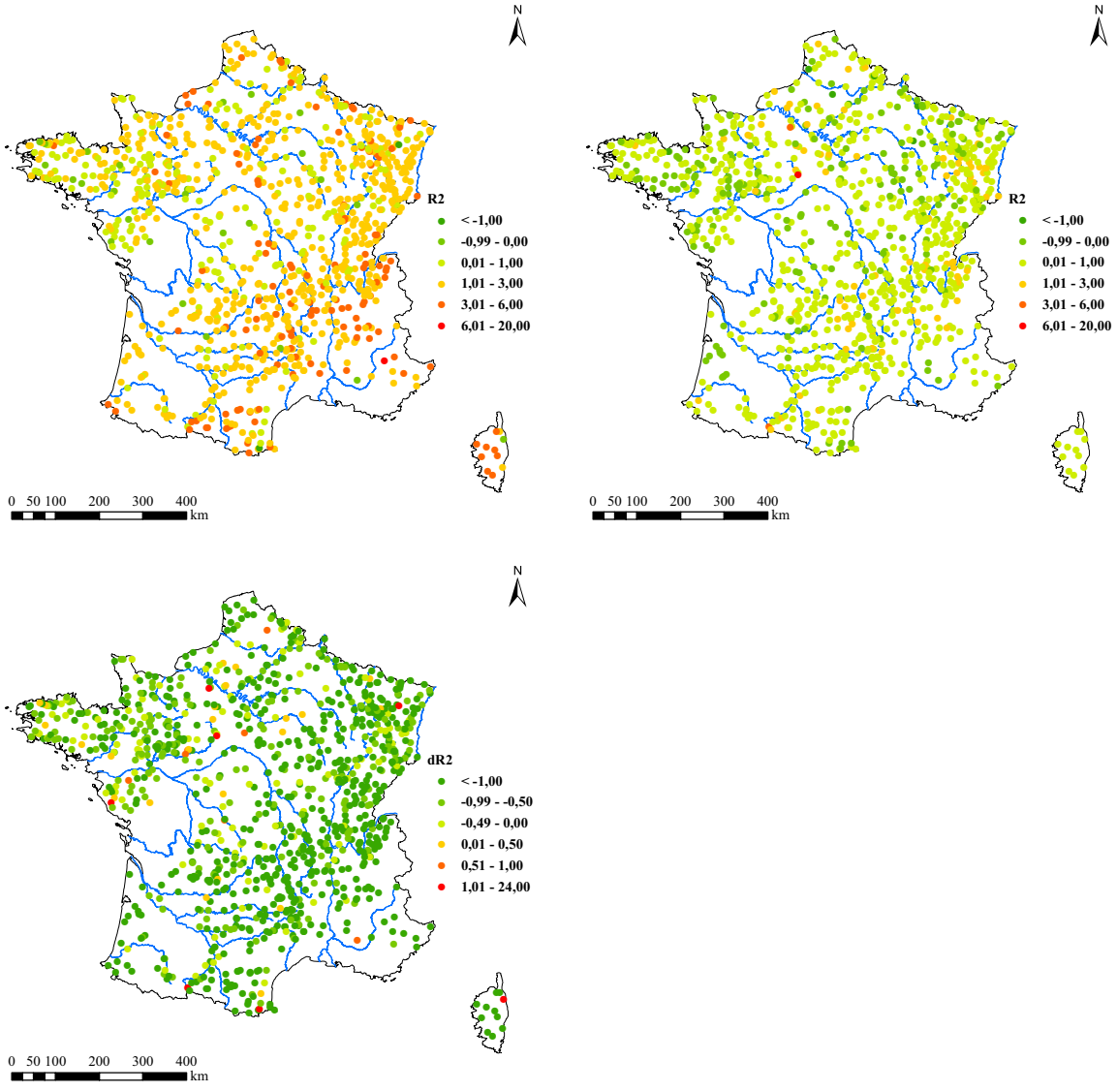


FIGURE P.14 – Critère : retard moyen à la montée de crue R_2 ; Horizon : 6 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

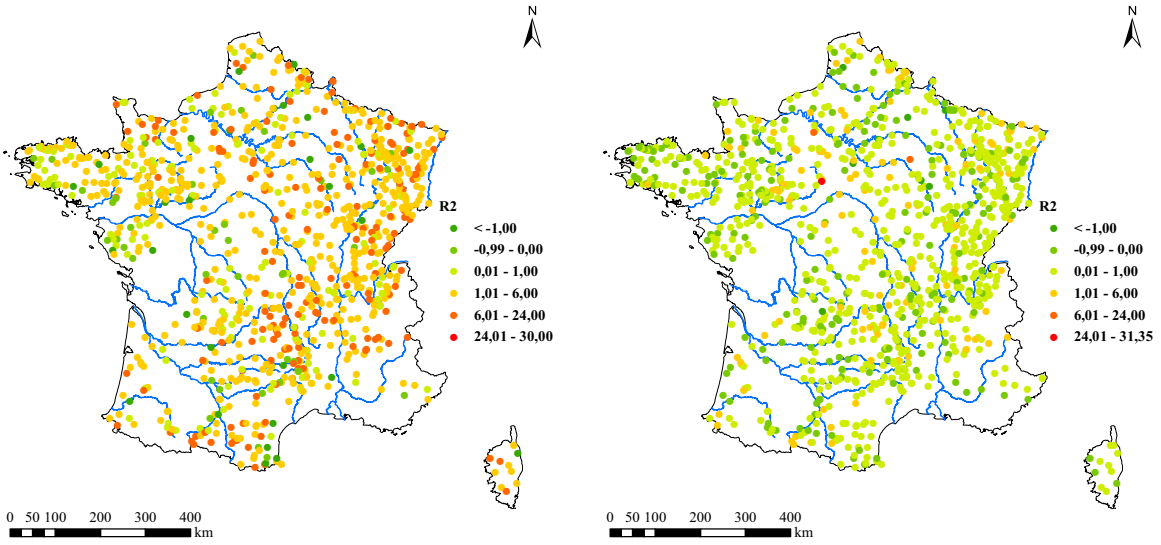


Figure P.15 – Critère : retard moyen à la montée de crue R_2 ; Horizon : 24 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA.

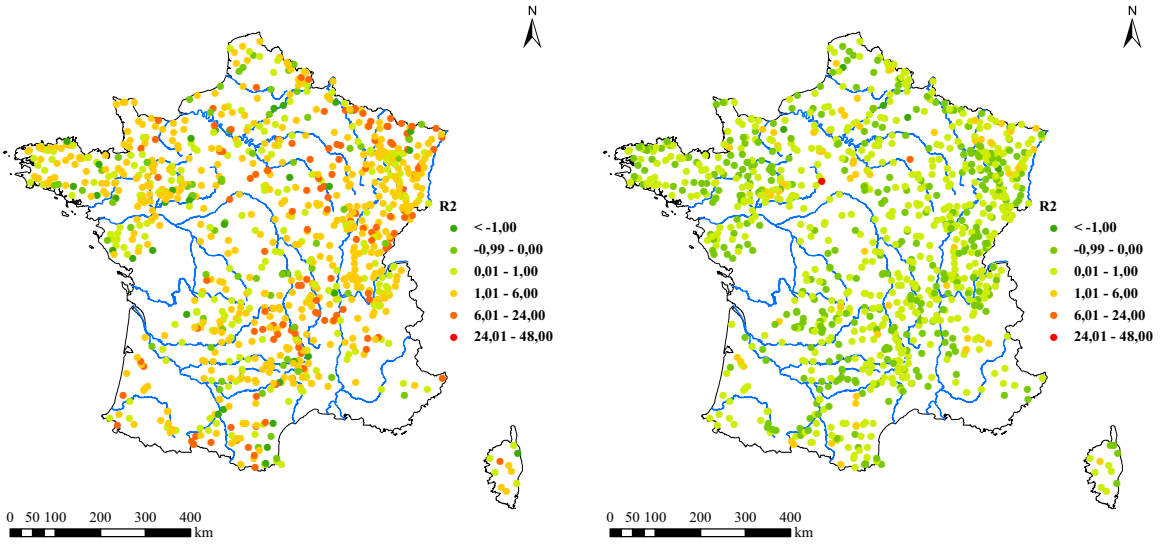


Figure P.16 – Critère : retard moyen à la montée de crue R_2 ; Horizon : 48 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA.

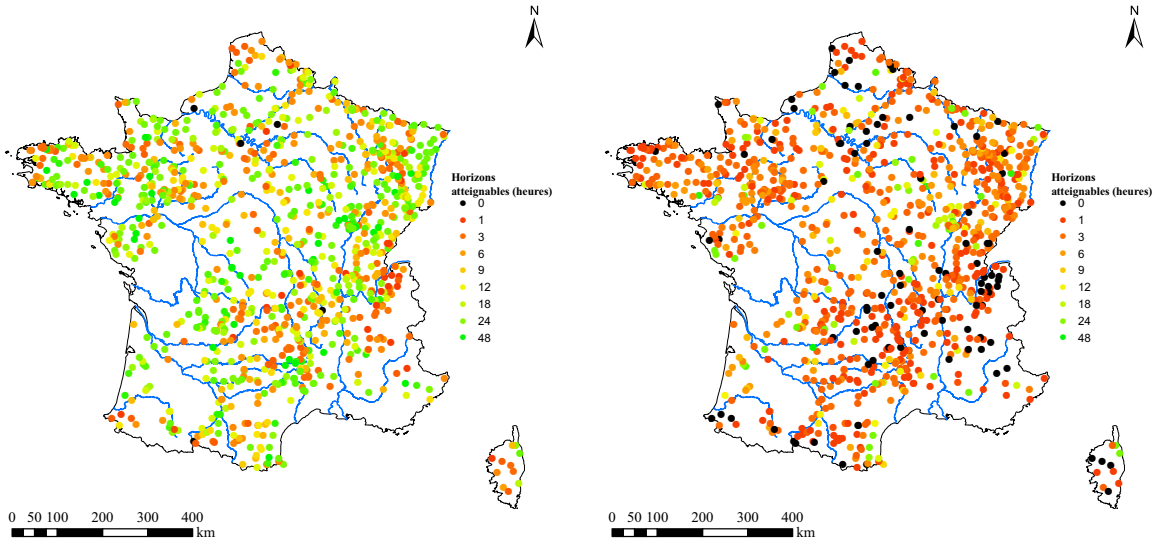


FIGURE P.17 – Critère : horizon atteignable pour prévoir le dépassement du seuil $q_{99}(Q)$ avec une probabilité au moins égale à 0.75 (gauche) ou 0.90 (droite) avec le modèle GRP-RNA.

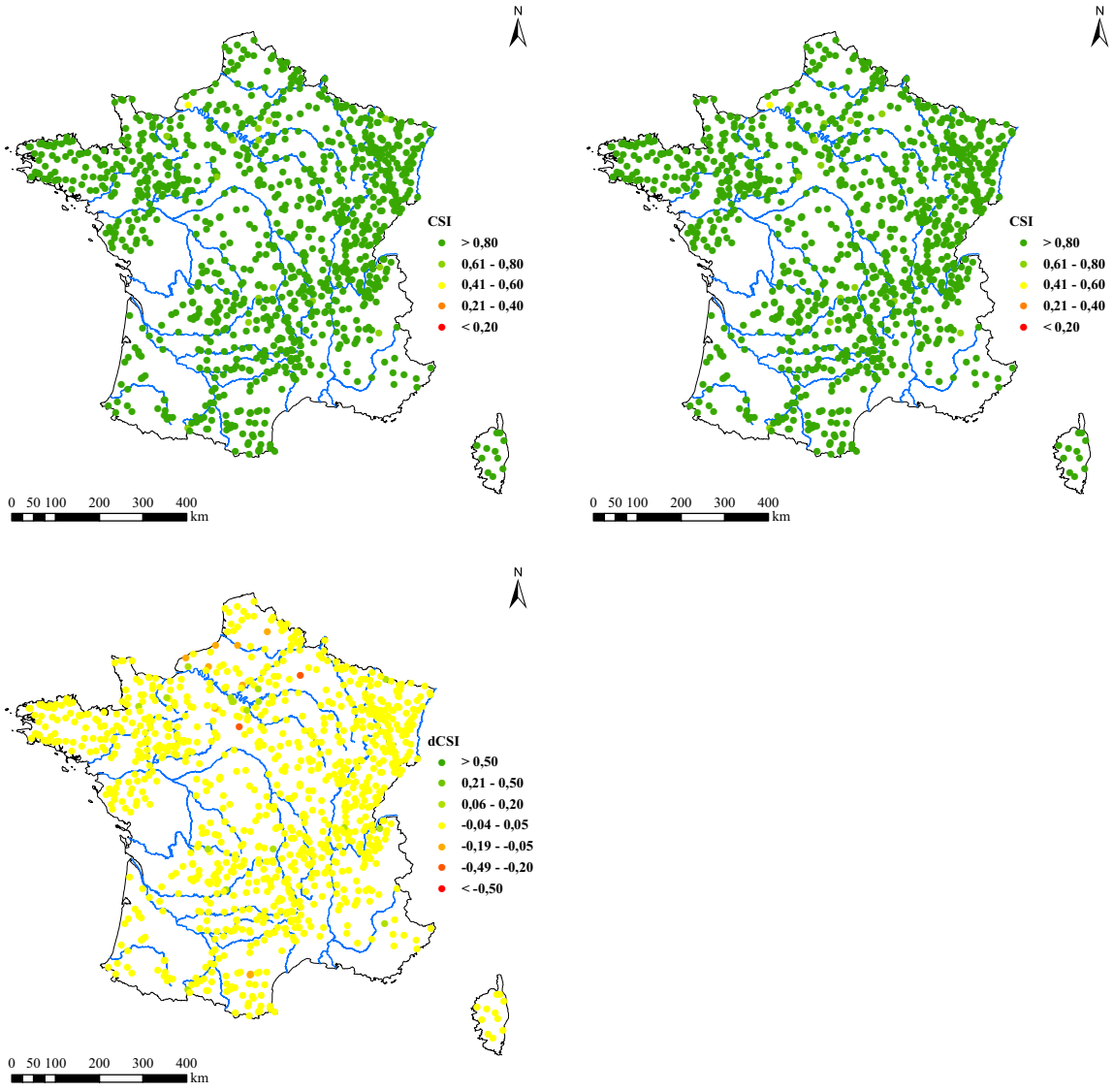


Figure P.18 – Critère : indice de succès critique (CSI) ; Horizon : 1 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

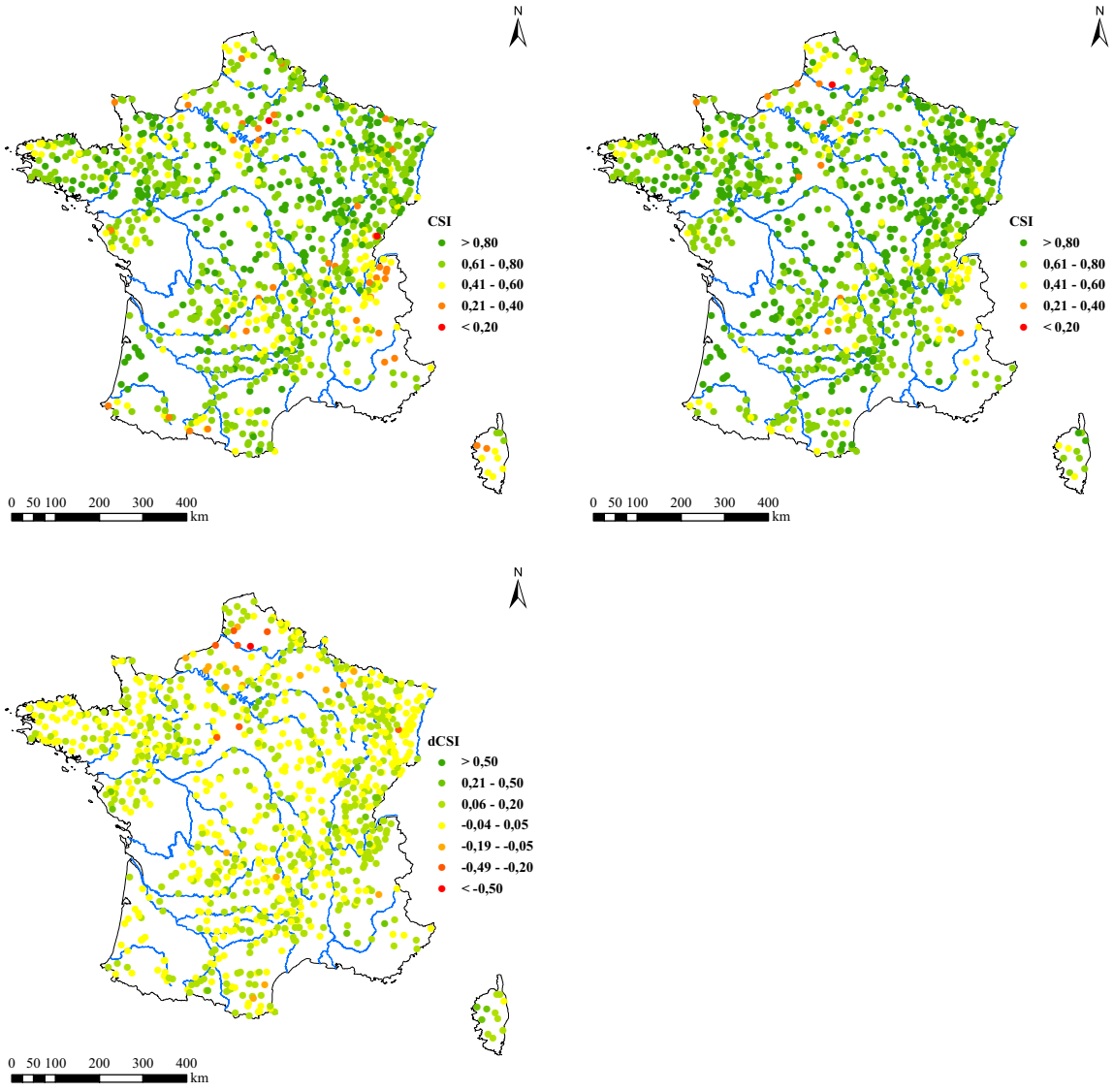


FIGURE P.19 – Critère : indice de succès critique (CSI) ; Horizon : 6 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

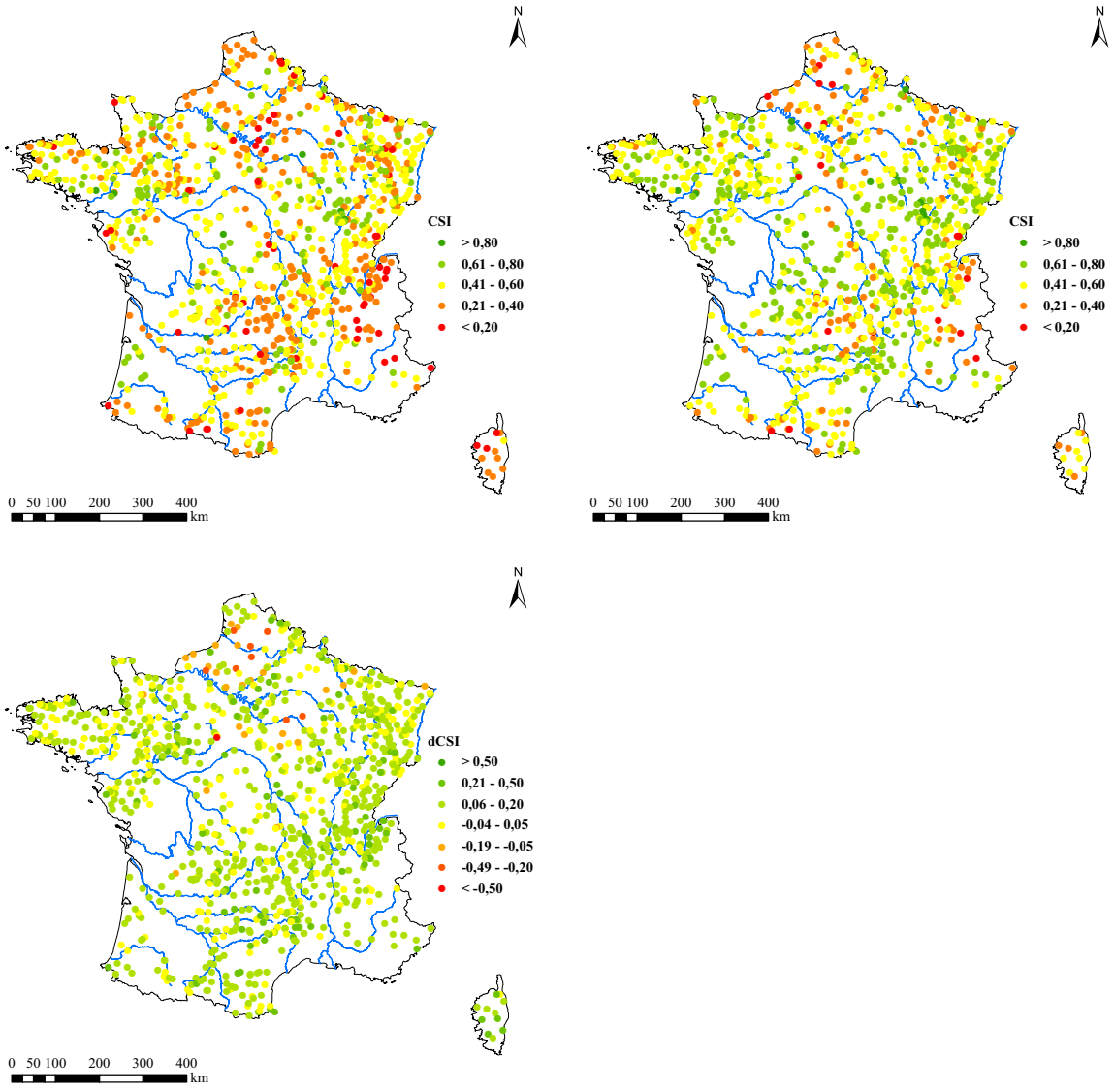


Figure P.20 – Critère : indice de succès critique (CSI) ; Horizon : 24 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

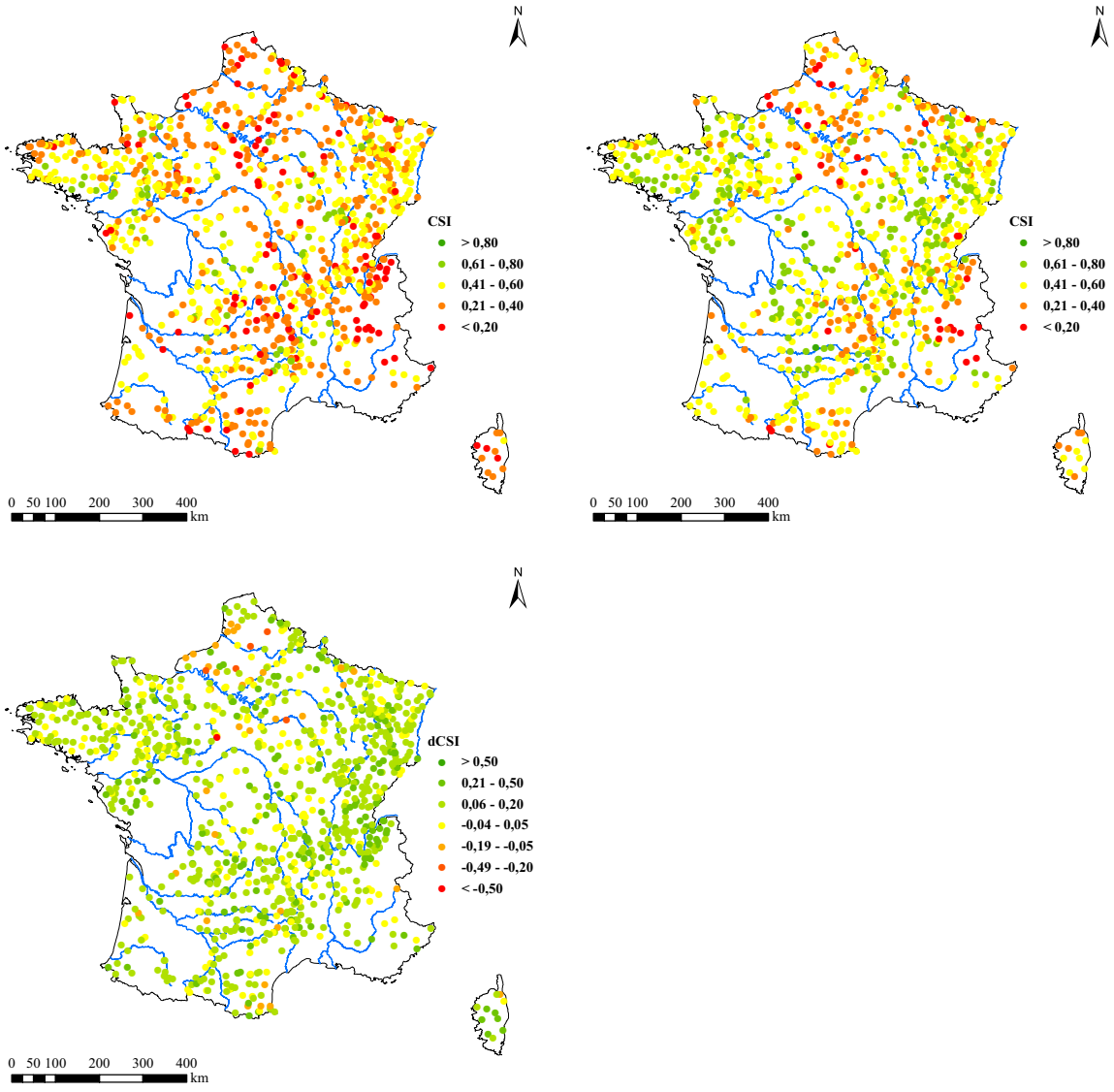


FIGURE P.21 – Critère : indice de succès critique (CSI) ; Horizon : 48 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

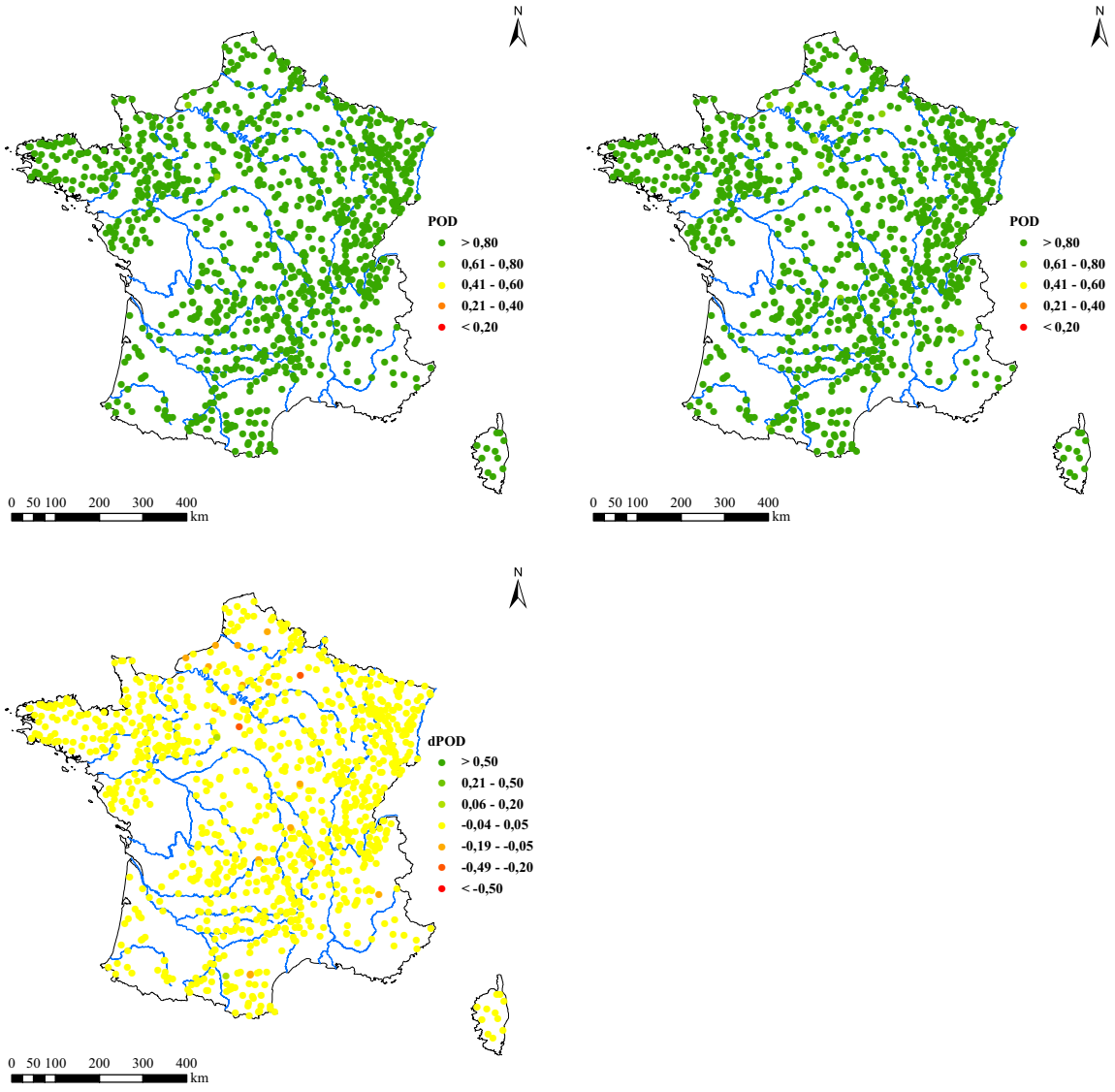


FIGURE P.22 – Critère : probabilité de détection (POD) ; Horizon : 1 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

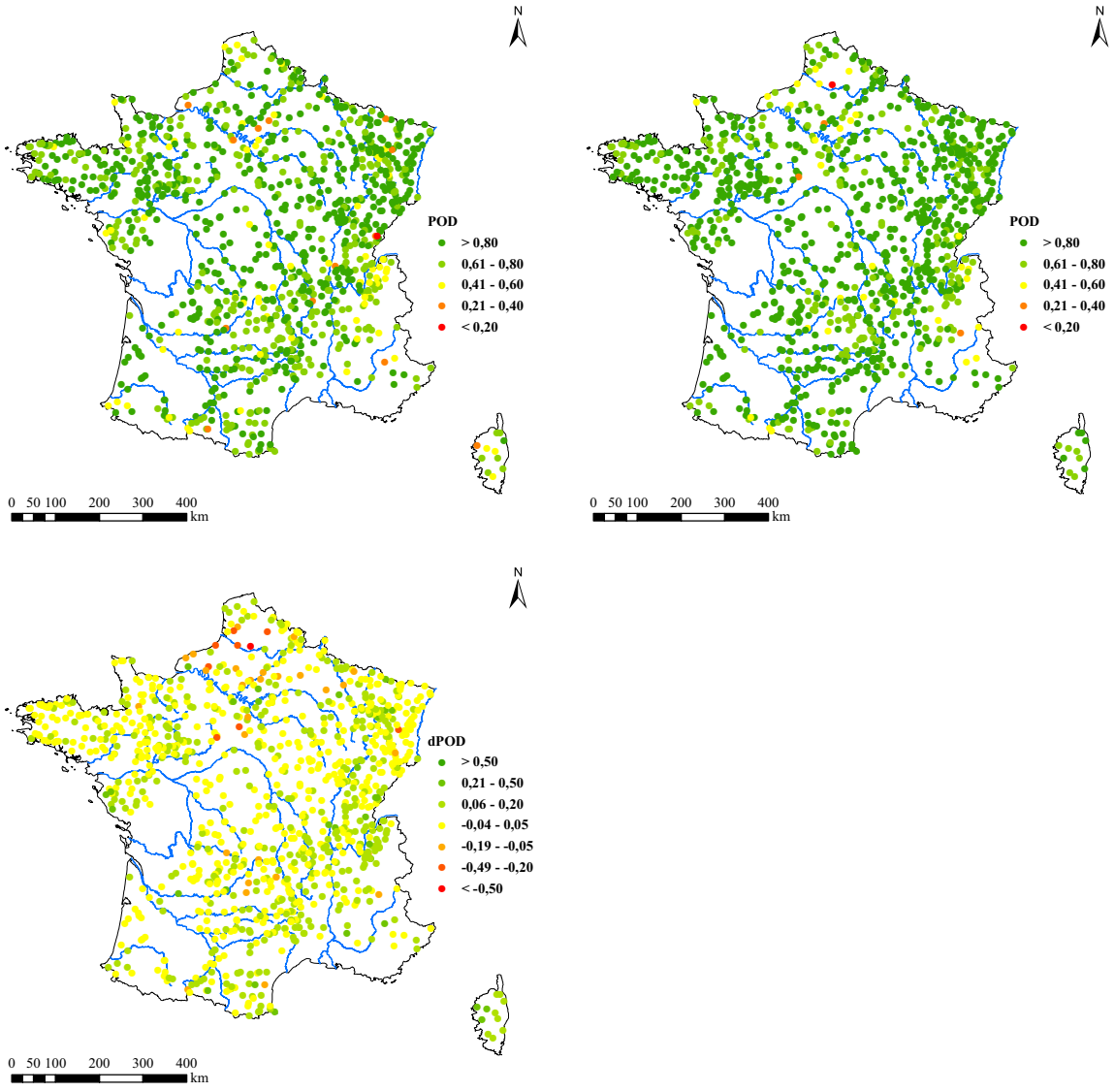


FIGURE P.23 – Critère : probabilité de détection (POD) ; Horizon : 6 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

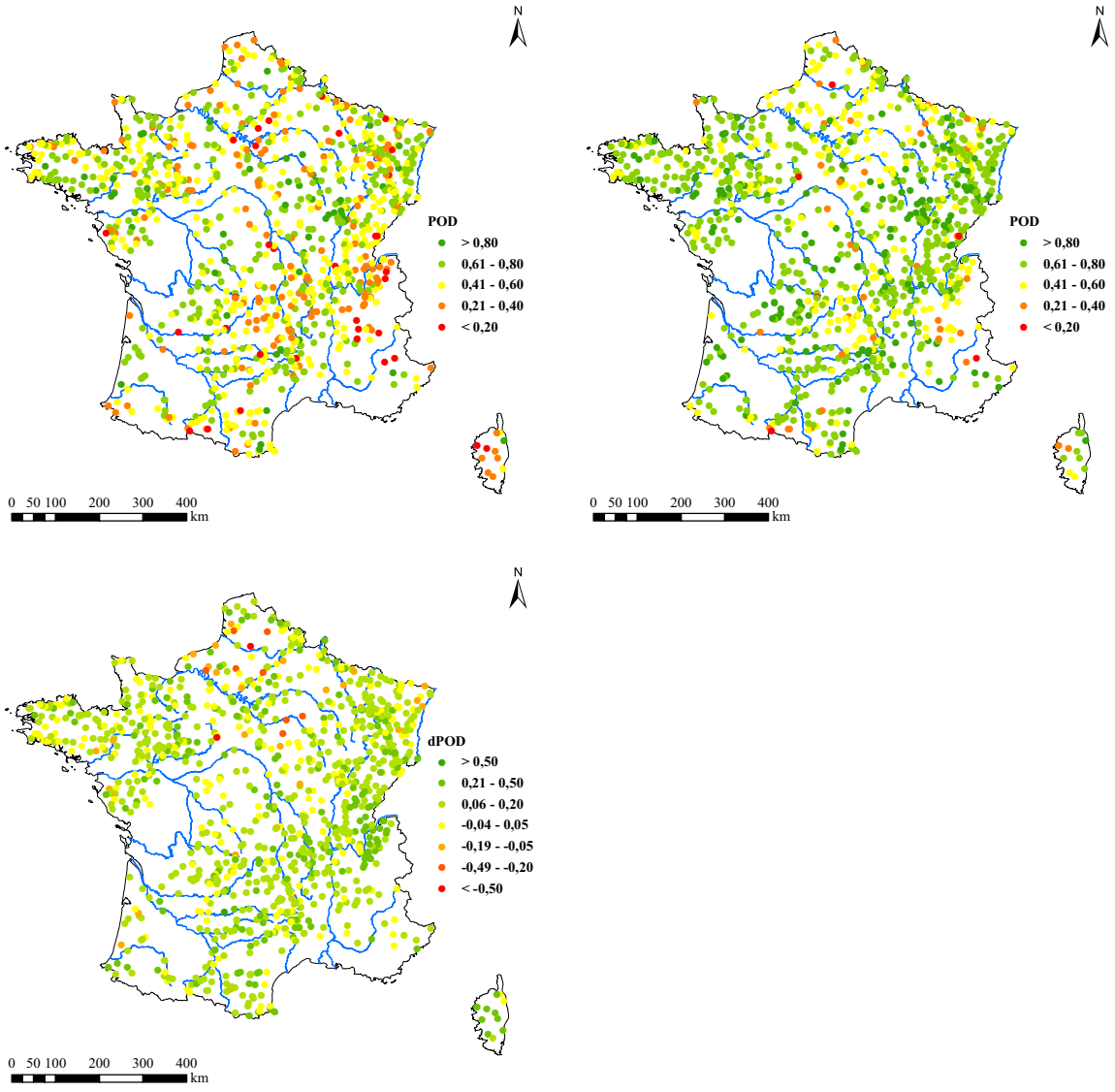


FIGURE P.24 – Critère : probabilité de détection (POD) ; Horizon : 24 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

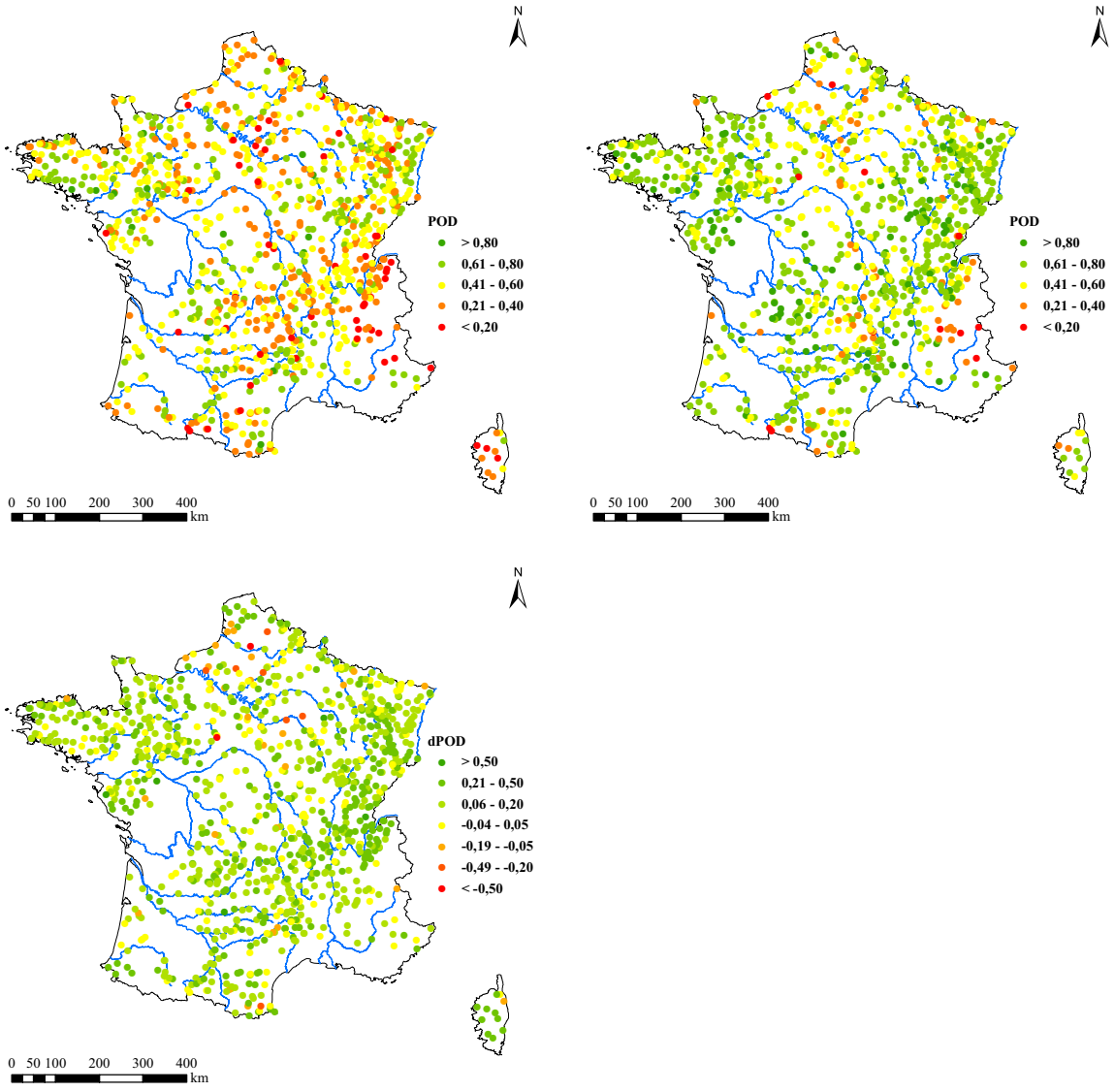


FIGURE P.25 – Critère : probabilité de détection (POD) ; Horizon : 48 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

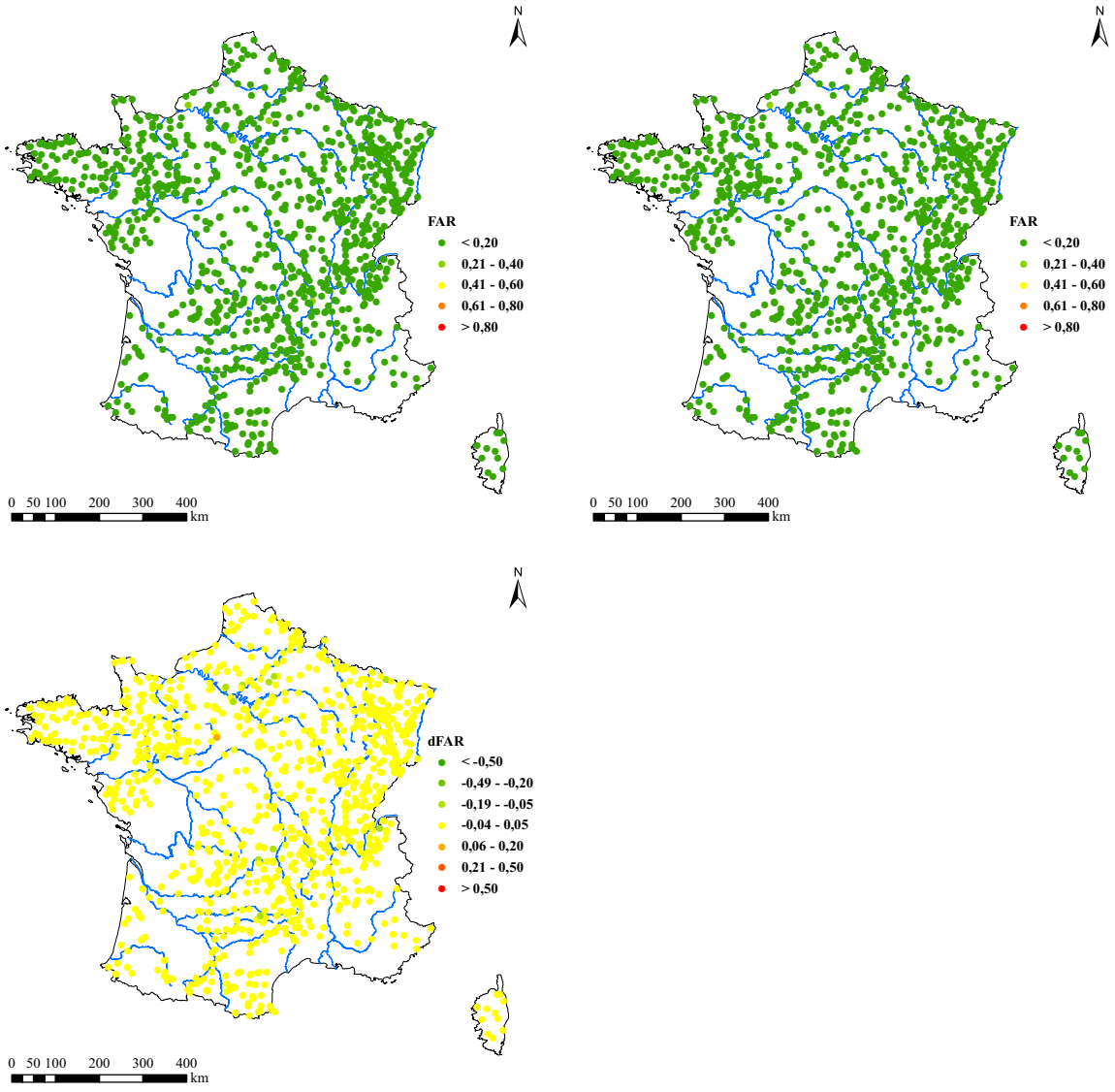


FIGURE P.26 – Critère : taux de fausses alertes (FAR) ; Horizon : 1 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

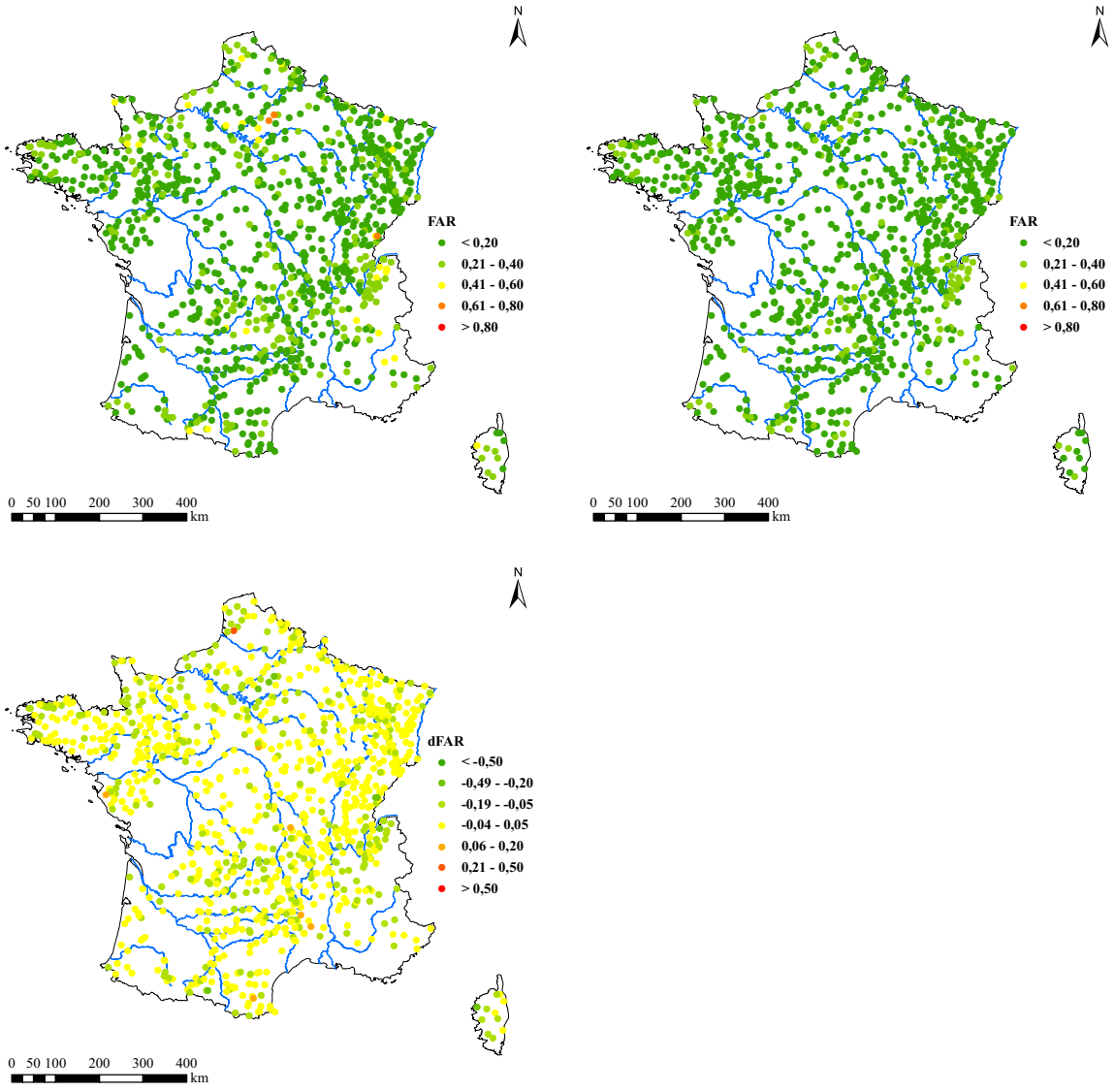


FIGURE P.27 – Critère : taux de fausses alertes (FAR) ; Horizon : 6 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

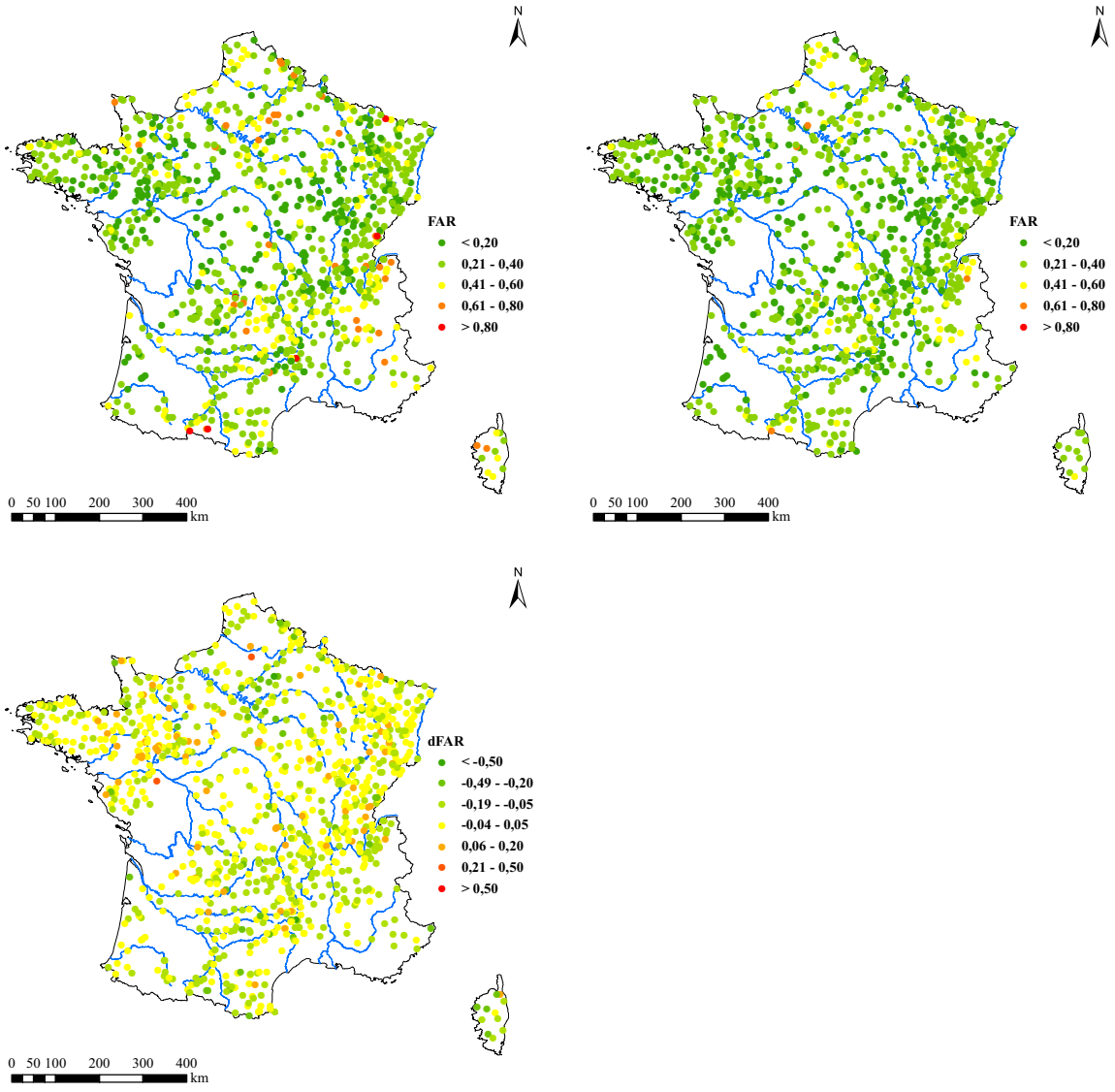


FIGURE P.28 – Critère : taux de fausses alertes (FAR) ; Horizon : 24 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

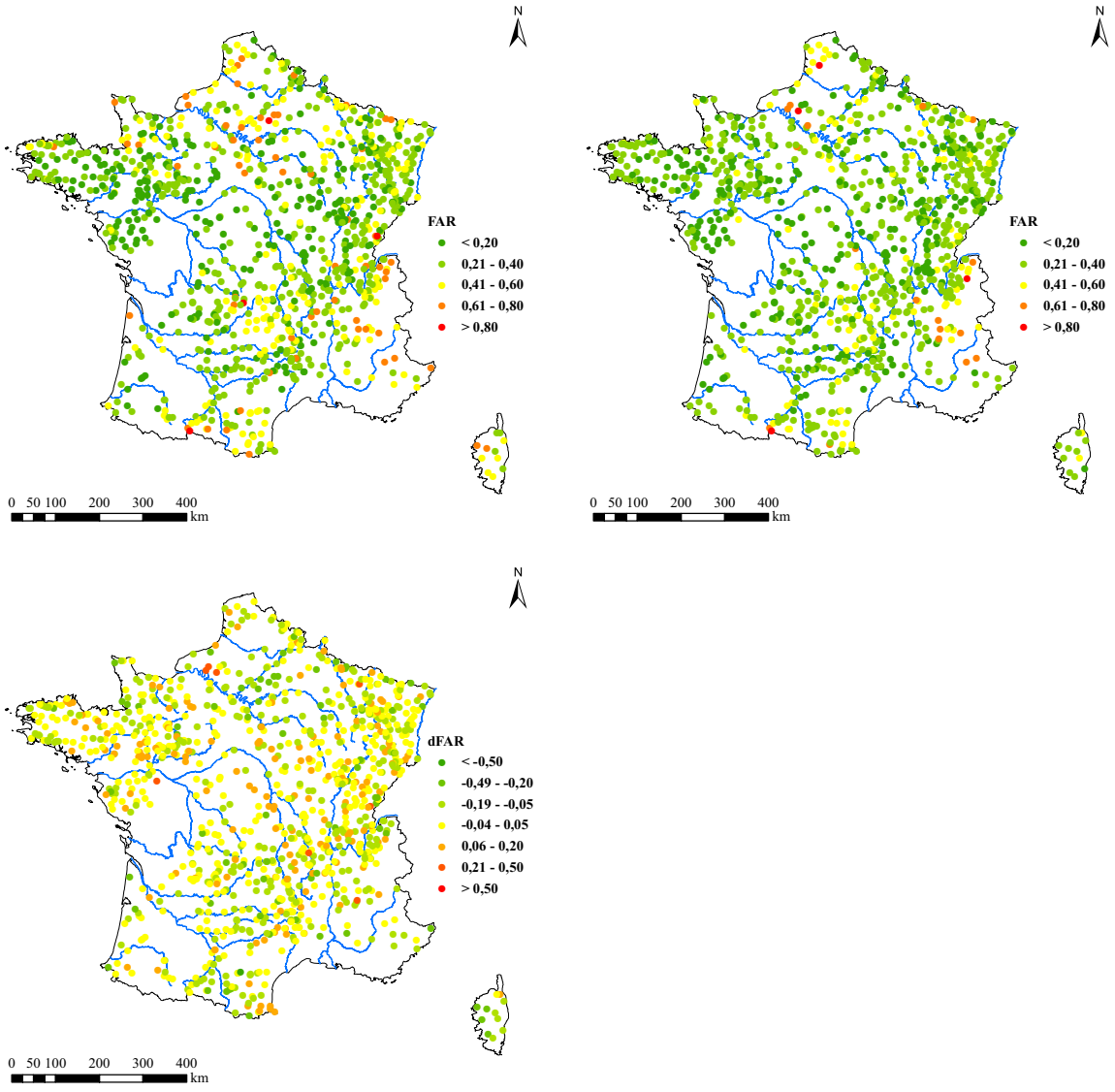


FIGURE P.29 – Critère : taux de fausses alertes (FAR) ; Horizon : 48 h.
En haut, à gauche, le modèle GRP-ARIMA ; à droite, le modèle GRP-RNA. En bas, différence des valeurs de ces deux modèles.

Q

Annexe

Méthode des plus proches voisins

La méthode des plus proches voisins est une approche très simple et totalement empirique. Elle consiste à rechercher dans une bibliothèque d'événements hydrométéorologiques ceux qui ressemblent le plus à l'événement en cours. Nous l'employons pour déterminer à travers son calage les données les plus pertinentes à apporter à un modèle de prévision Pluie – Débit, car son principe ne repose sur aucune conception hydrologique.

Q.1 Les mêmes causes entraînent les mêmes effets

Ce modèle suppose seulement que les processus en jeux sont stationnaires et que nous disposons d'une série de données contenant à peu près tous les types d'événements de crue possible (ce qui implique notamment dans notre cas qu'il est suffisant de connaître la pluie et les débits pour caractériser les événements de crue). Le modèle recherche les événements présentant les données hydrométérologiques les plus proches des conditions présentes dans la chronique passée, pour en déduire les débits futurs.

Q.1.1 Définir une distance entre l'événement présent et les événements passés

Il existe une multitude d'approches permettant de définir la distance entre deux événements hydrométéorologiques. Ce mémoire n'ayant pas pour objet d'optimiser cette approche, nous en testons une et suivons la méthodologie de Karlsson et Yakowitz (1987). Chaque pas de temps est caractérisé par un vecteur contenant les dernières observations de débits et pluies. La proximité entre les conditions hydrométéorologiques actuelles et celles d'un événement passé est traduite par une distance entre le vecteur du pas de temps actuel t et un pas de temps passé t'. Cette distance se décompose en une distance sur les débits d_Q et une distance sur les pluies d_P :

$$d^2 = d_Q^2 + \beta \cdot d_P^2 \quad \text{avec}$$

$$d_P^2 = \sum_{i=-H}^{N} \alpha_P^{i+\Delta_P} \cdot (P_{t-i} - P_{t'-i})^2 \quad \text{et} \quad d_Q^2 = \sum_{i=0}^{N} \alpha_Q^i \cdot (Q_{t-i} - Q_{t'-i})^2$$

où H est l'horizon de prévision. Ces deux distances sont construites sur des poids affectés aux observations de pluie et de débit diminuant avec leur ancienneté. Le modèle compte quatre paramètres. Les deux premiers α_Q et α_P, compris entre 0 et 1, permettent d'évaluer le nombre de données de débit (respectivement de pluie) réellement prises en compte par le modèle. Le paramètre β permet d'évaluer l'importance relative des informations apportées par les données de pluie et de débit. Enfin, un quatrième paramètre, le décalage temporel Δ_P, nous informe sur les données de pluie les plus importantes.
Le nombre de pas de temps N est choisi égal à 96 (comme chez Karlsson et Yakowitz, 1987).

Q.1.2 Sélection de 10 événements proches à la situation actuelle

Pour chaque pas de temps (de la période de contrôle), les dix événements passés les plus semblables sont sélectionnés. Chaque événement est caractérisé par sa distance à l'événement présent. L'estimation du débit à l'horizon H est alors une moyenne des débits passés à l'horizon H de la date de référence, pondérée par l'inverse du carré de la distance à l'événement présent :

$$\widehat{Q}_{t+H|t} = \left(\sum_{i=1}^{10} \frac{1}{d_{t_i}^2} \right)^{-1} \cdot \sum_{i=1}^{10} \frac{1}{d_{t_i}^2} \cdot Q_{t_i+H}$$

Q.1.3 Calage et contrôle

L'évaluation du modèle des plus proches voisins nécessite de séparer les chroniques de données disponibles en trois sous-parties pour être conforme aux préconisations de Klemeš (1986) :

- une période « bibliothèque » où le modèle recherche des événements passés similaires.
- une période de calage,
- une période de contrôle.

Nous disposons de chroniques de 10 ans. Pour que la « bibliothèque » soit suffisamment riche, la chronique est scindée en une première partie de 5 ans servant de bibliothèque, la seconde partie est utilisé pour 3 ans de calage et 2 ans de contrôle. Les périodes de 5 ans sont ensuite inversées.

Q.2 Résultats

Q.2.1 Enseignements sur l'importance des données

Les valeurs calées du paramètre α_Q, sont distribuées sur l'intervalle $[0, 1]$: le modèle peut n'utiliser que le débit antérieur sur quelques bassins (mais très logiquement le critère de persistance est alors proche de 0 au contrôle). Au contraire, d'autres bassins présentent une valeur de α_Q proche de 1 : ils utilisent quasiment tous les débits des $N = 96$ derniers pas de temps. La médiane du paramètre α est située autour de 0.4 (pour les horizons de 1 et 6 heures) et 0.5 (pour l'horizon de 24 heures), ce qui implique que le plus souvent un très petit nombre d'observations de débit sont réellement prises en compte par cette approche.

Le paramètre analogue pour les pluies, α_P, est compris entre 0,4 et 1 : un schéma des données de pluies ne se réduisant pas à la seule donnée du pas de temps courant est donc nécessaire, ce qui traduit probablement l'importance des pluies futures. La médiane et la

moyenne de ce paramètre se situe autour de 0,7 : les données au-delà de 10 heures (avant et après $t + \Delta_P$) ne sont pas nécessaires dans plus de la moitié des cas.

Le paramètre Δ_P est centré sur la date de prévision (avec un léger décalage d'une heure). Ses valeurs sont resserrées autour de quelques heures.

Le dernier paramètre, β, mesure l'importance relative de l'information de pluies par rapport à l'information contenue dans le débit antérieur. Les valeurs de β sont faibles : en moyenne à 0,1 et pour neuf bassins sur dix, inférieures à 0,25. On constate cependant que la distribution s'étale doucement au-delà de 1. Cette observation peut signifier :
- que la pluie apporte le plus souvent moins d'information jugée utile par la méthode que le débit,
- que le pas de temps n'est pas adéquat, peut-être faudrait-il prendre des pas de temps plus grands pour la pluie (des cumuls sur plusieurs heures).

Q.2.2 Performances de cette approche en prévision

La méthode des plus proches voisins a été appliquée sur nos 178 bassins pour des prévisions à des horizons de 1, 6 et 24 heures. En contrôle, les performances médianes et moyennes se situent au niveau de celles du modèle de persistance avec un nombre non négligeable de bassins sur lesquels cette méthode obtient des résultats très nettement inférieurs.

À la vue de ces résultats, il ne nous paraît pas opportun d'explorer les propositions d'amélioration de Shamseldin et O'Connor (1996), qui permettent notamment d'extrapoler dans une certaine mesure au-delà des plus forts débits de la chronique de référence. D'une part, les performances de la méthode initiale sont très faibles. D'autre part, l'hypothèse de linéarité sur les perturbations ne paraît pas fondée et les performances présentées dans l'article ne sont pas tellement plus élevées que les nôtres.

Q.3 Conclusions

La méthode des plus proches voisins ne permet pas ici d'atteindre des niveaux de performance comparables à ceux du modèle plus conceptuel GRP. L'absence totale de conceptualisation peut expliquer ce constat au pas de temps horaire.

Quelle valeur peut-on accorder aux informations sur l'importance des données que nous venons de tirer de la méthode des plus proches voisins ? Ces informations confirment bien nos connaissances *a priori*. Mais d'un autre côté, l'approche « dégrade » l'information puisqu'en partant de plus de données que le simple débit actuel, les performances sont inférieures à celles du modèle de persistance qui ne s'appuie que sur le débit au pas de temps courant.

F'Murrr (1983). Tonnerre et mille sabots. *Le Génie des Alpages*, **7 :** 40 – 41