



**HAL**  
open science

## Comment quantifier l'incertitude prédictive en modélisation hydrologique ?

François Bourgin

► **To cite this version:**

François Bourgin. Comment quantifier l'incertitude prédictive en modélisation hydrologique?. Hydrologie. AgroParisTech, 2014. Français. NNT : 2014AGPT0016 . tel-01130084v1

**HAL Id: tel-01130084**

**<https://pastel.hal.science/tel-01130084v1>**

Submitted on 11 Mar 2015 (v1), last revised 9 Mar 2015 (v2)

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## Doctorat ParisTech

École Doctorale *Géosciences, Ressources Naturelles et Environnement*

### THÈSE

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*

**François BOURGIN**

le 7 Avril 2014

## Comment quantifier l'incertitude prédictive en modélisation hydrologique ?

**Travail exploratoire sur un grand échantillon de bassins versants**

Directeur de thèse : **Vazken ANDRÉASSIAN**

Co-encadrement de la thèse : **Charles PERRIN**

### Jury

**M. Pierre RIBSTEIN**

**Mme Anne-Catherine FAVRE**

**M. Éric GAUME**

**M. Christophe CUDENNEC**

**M. Éric SERVAT**

**M. Massimiliano ZAPPA**

**M. Lionel BERTHET**

METIS - UPMC, Paris

LTHE, Grenoble

IFSTTAR, Nantes

Agrocampus Ouest, Rennes

Hydrosciences, Montpellier

WSL, Birmensdorf

DREAL Auvergne, Clermont-Ferrand

Président

Rapporteur

Rapporteur

Examinateur

Examinateur

Examinateur

Invité

**AgroParisTech**

**Irstea, Unité de recherche Hydrosystèmes et Bioprocédés**

1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony cedex



# Doctorat ParisTech

École Doctorale *Géosciences, Ressources Naturelles et Environnement*

## THÈSE

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*

**François BOURGIN**

le 7 Avril 2014

## **Comment quantifier l'incertitude prédictive en modélisation hydrologique ?**

**Travail exploratoire sur un grand échantillon de bassins versants**

Directeur de thèse : **Vazken ANDRÉASSIAN**

Co-encadrement de la thèse : **Charles PERRIN**

### **Jury**

<b>M. Pierre RIBSTEIN</b>	METIS - UPMC, Paris	Président
<b>Mme Anne-Catherine FAVRE</b>	LTHE, Grenoble	Rapporteur
<b>M. Éric GAUME</b>	IFSTTAR, Nantes	Rapporteur
<b>M. Christophe CUDENNEC</b>	Agrocampus Ouest, Rennes	Examineur
<b>M. Éric SERVAT</b>	Hydrosciences, Montpellier	Examineur
<b>M. Massimiliano ZAPPA</b>	WSL, Birmensdorf	Examineur
<b>M. Lionel BERTHET</b>	DREAL Auvergne, Clermont-Ferrand	Invité



## REMERCIEMENTS

Je tiens à remercier chaleureusement mes encadrants Vazken Andréassian et Charles Perrin, pour m'avoir accompagné dans ce travail de recherche. Merci pour l'enthousiasme, le dynamisme, la bienveillance, les conseils avisés et la vision de l'hydrologie que vous apportez, chacun à votre manière, à l'équipe et à vos étudiants.

J'adresse également mes sincères remerciements aux membres du jury qui m'ont fait l'honneur d'évaluer ces travaux, en premier lieu Pierre Ribstein, en sa qualité de président, Anne-Catherine Favre et Éric Gaume pour leur lecture attentive du manuscrit et leurs suggestions pertinentes, et Lionel Berthet, Christophe Cudennec, Éric Servat et Massimiliano Zappa pour leur participation au jury de thèse et leurs remarques non moins pertinentes.

Je remercie une nouvelle fois Lionel Berthet, ainsi que Nicolas Le Moine, Ludovic Oudin, Eric Parent et Benjamin Renard, pour avoir suivi le déroulement de cette thèse et su prodiguer conseils, critiques et encouragements au cours des réunions du comité de thèse. En tant que membres permanents de l'équipe, Maria-Helena Ramos, Guillaume Thirel et Carina Furusho ont non seulement participé à ces réunions mais également contribué très activement aux différents travaux réalisés au cours de cette thèse ; un grand merci à tous les trois !

Cette thèse a bénéficié du soutien financier du SCHAPI, et je tiens à remercier tout particulièrement André Bachoc et Bruno Janet pour l'intérêt qu'ils ont manifesté pour mes travaux. Les échanges avec plusieurs membres des SPC ont également permis d'assurer à cette thèse une portée opérationnelle, et je les en remercie.

J'adresse par ailleurs mes remerciements au SCHAPI et à Météo-France pour avoir mis à disposition les données hydro-météorologiques utilisées dans cette thèse, et à l'école doctorale GRNE et AgroParisTech pour les formations (y compris lors de stages de terrain) et le suivi académique dont j'ai bénéficié.

Il est difficile de quitter sans une certaine nostalgie l'équipe et le groupement d'Antony. Merci à tous ceux qui ont contribué à faire de cet endroit un lieu où il fait bon aller. Je pense bien sûr aux activités sportives qui nous rassemblent, aux cuvées du nectar d'Antony, et la bienveillance de l'ensemble des personnes qui facilitent les démarches administratives et

l'utilisation des ressources informatiques ; aux discussions prolongées dans les transports, les soirées ici et là, au cours de week-ends classés pas joli-joli où l'hydrologie se fait et se défait au rythme de la pluie.

Je pense plus particulièrement à Florent, mon jumeau de thèse à l'énergie inépuisable, à Pierre au coup droit un peu lézendaire, à Laurent pour sa coronade et à Pierre (non pas le beau, l'autre) qui a apporté de la hauteur de l'extrême à l'équipe. Je pense également aux anciens doctorants désormais docteurs qui nous ont montré la voie de l'hydrologie sociale, aux nouveaux doctorants qui poursuivront, et à tous les étudiants de passage qui j'espère auront pris goût aux études hydrologiques.

Enfin, comment oublier mes amis et les membres de ma famille qui ont su m'accompagner, m'encourager et me supporter (parfois ?) au cours de ces années de thèse. Merci pour votre soutien indispensable ! Et je pense bien sûr à toi, Tafadzwa. Tu sais.

## RÉSUMÉ

La modélisation hydrologique permet de quantifier la transformation pluie-débit au sein d'un bassin versant. Bien que les modèles parviennent généralement à représenter de manière acceptable le fonctionnement des bassins versants, cette représentation, nécessairement simplifiée, reste imparfaite, et une quantification des incertitudes est souhaitable. Cette thèse s'intéresse à la quantification de l'incertitude prédictive en modélisation hydrologique. Le principal objectif de nos travaux est d'explorer différentes méthodes qui permettent d'associer à des simulations ou des prévisions de débits déterministes des distributions probabilistes. Nous distinguons le contexte de simulation du contexte de prévision et adoptons dans ces deux cas une démarche comparative et pragmatique qui permet d'évaluer différentes approches sur un large échantillon de bassins versants français, à l'aide de critères d'évaluation adaptés.

En simulation, nos travaux ont porté sur deux méthodes liées à l'estimation des paramètres des modèles hydrologiques, la méthode GLUE et le calage bayésien, ainsi que sur deux approches plus pragmatiques, l'approche multi-modèles, et le post-traitement statistique. Nos résultats suggèrent que les approches telles que GLUE qui ne s'appuient que sur un ensemble de différents jeux de paramètres ne parviennent pas, en général, à représenter de manière adéquate l'incertitude prédictive totale. L'utilisation d'un modèle d'erreur extérieur au fonctionnement interne du modèle hydrologique est nécessaire. Les méthodes de post-traitement suffisamment flexibles pour caractériser les erreurs résiduelles obtenues en calage parviennent à refléter de manière plus satisfaisante les marges d'erreurs du modèle hydrologique utilisé. Nous proposons également une méthode qui permet d'obtenir une estimation de l'incertitude prédictive pour les bassins non-jaugés, au moyen d'un transfert des marges d'erreurs constatées sur les bassins jaugés. Les résultats indiquent que la méthode est prometteuse et fournit dans la plupart des cas des intervalles de confiance fiables et fins sur les sites non-jaugés.

En prévision, nos travaux ont porté d'une part sur la comparaison de différentes méthodes de post-traitement statistique, et d'autre part sur l'interaction entre l'assimilation de données et le post-traitement au sein d'une chaîne de prévision hydrologique d'ensemble. Les résultats obtenus montrent l'importance de la prise en compte de l'évolution de l'incertitude prédictive en fonction de l'échéance de prévision et mettent en évidence les gains de performance qui peuvent être obtenus quand la quantification de l'incertitude s'appuie sur une meilleure caractérisation de la situation de prévision. Enfin, nos travaux indiquent que l'utilisation conjointe de l'assimilation de données et d'une méthode de post-traitement permet d'améliorer les performances d'une chaîne de prévision hydrologique d'ensemble.





# ABSTRACT

Hydrological modelling aims to quantify the rainfall-runoff relationship at the catchment scale. Although hydrological models are generally able to provide an acceptable representation of the catchment behaviour, this representation is necessarily simplified and imperfect, as a consequence an evaluation of uncertainties is desirable. This thesis focuses on the quantification of predictive uncertainty in hydrological modelling. Our main objective was to explore several methods that can be used to complete the deterministic predictions of a rainfall-runoff model with probabilistic distributions. Two prediction cases were distinguished, namely simulation and forecast. We set up a comparative framework to evaluate different uncertainty quantification methods on a large set of French catchments.

In simulation mode, we focused on two methods related to the parameter estimation issue, the GLUE uncertainty method and the Bayesian framework, as well as two more pragmatic approaches, a multi-model approach and the post-processing method. Our results suggest that the GLUE-like methods based on ensembles of various parameter sets do not provide an adequate representation of the total predictive uncertainty. An external model error is necessary. Post-processing methods that are flexible enough to adequately describe the residual errors of the hydrological model obtained during calibration give more satisfactory estimates of the modelling uncertainty. We also present a method able to transfer model uncertainty estimates from gauged to ungauged catchments. Our results demonstrate that the method holds interesting perspectives, providing in most cases reliable and sharp uncertainty bounds at ungauged locations.

In a forecasting context, we compared several post-processing methods and evaluated the interactions between data assimilation and post-processing in an ensemble forecasting modelling chain. Results show the crucial role of the lead time on the estimates of predictive uncertainty. They also suggest that some improvement can be achieved when the evolution of flows is better taken into account by a post-processing method. Last, we investigated the interactions between data assimilation and post-processing in hydrological ensemble forecasting and showed the benefits of using both in an ensemble forecast system.



# TABLE DES MATIÈRES

<b>Remerciements</b>	<b>v</b>
<b>Résumé</b>	<b>vii</b>
<b>Abstract</b>	<b>ix</b>
<b>Table des figures</b>	<b>xvii</b>
<b>Liste des tableaux</b>	<b>xxi</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Incertitudes en modélisation hydrologique</b>	<b>7</b>
1.1 Introduction . . . . .	9
1.2 Caractérisation et représentation des incertitudes . . . . .	9
1.2.1 Caractérisation des incertitudes . . . . .	9
1.2.2 Représentation des incertitudes . . . . .	11
1.3 Présentation de différentes méthodes d'analyse des incertitudes en hydrologie	12
1.3.1 Tentatives de classification des méthodes d'analyse des incertitudes .	12
1.3.2 Présentation de la méthode GLUE . . . . .	13
1.3.3 Présentation de l'approche bayésienne . . . . .	16
1.3.4 Présentation de l'approche multi-modèles . . . . .	18
1.3.5 Présentation de l'approche de post-traitement . . . . .	19
1.4 Synthèse . . . . .	21
<b>2 Données et modèles pluie-débit</b>	<b>23</b>
2.1 Introduction . . . . .	25
2.2 Données . . . . .	25
2.2.1 Bases de données hydro-climatiques . . . . .	25
2.2.2 Constitution d'un échantillon de bassins versants . . . . .	26
2.3 Modèles pluie-débit . . . . .	31
2.3.1 Simulation et prévision . . . . .	34
2.3.2 Modèle GR4J . . . . .	35
2.3.3 Modèle GRP . . . . .	37
2.4 Synthèse . . . . .	39

<b>3 Méthodes d'évaluation de la qualité des incertitudes prédictives</b>	<b>41</b>
3.1 Introduction . . . . .	43
3.2 Quelques rappels sur les probabilités . . . . .	43
3.3 Critères et outils graphiques d'évaluation . . . . .	45
3.3.1 Évaluer une série d'intervalles prédictifs . . . . .	45
3.3.2 Évaluer une série de quantiles . . . . .	47
3.3.3 Évaluer une série de distributions prédictives . . . . .	48
3.3.4 Évaluer une série de probabilités de dépassement de seuil . . . . .	52
3.4 Choix d'une référence : la climatologie . . . . .	53
3.5 Synthèse . . . . .	54
<b>4 Flow-group conditioning of the PIT diagram</b>	<b>55</b>
Résumé . . . . .	57
Abstract . . . . .	57
4.1 Introduction . . . . .	58
4.1.1 On the specificity of probabilistic forecasts evaluation . . . . .	58
4.1.2 Reliability and the PIT diagram . . . . .	58
4.1.3 Scope of this note . . . . .	59
4.2 Data and methods . . . . .	59
4.2.1 Data Set . . . . .	59
4.2.2 Probabilistic hydrological forecasting system . . . . .	59
4.2.3 PIT Diagram and reliability score . . . . .	60
4.3 Results . . . . .	61
4.3.1 Evaluation of the PIT diagram over the entire flow range . . . . .	61
4.3.2 Evaluation of PIT diagrams for high and low flows . . . . .	61
4.3.3 Looking after hidden compensations when evaluating forecasts with the PIT diagram . . . . .	64
4.4 Conclusions and recommendations . . . . .	66
4.5 Acknowledgements . . . . .	66
<b>5 Présentation des méthodes utilisées en simulation</b>	<b>67</b>
5.1 Introduction . . . . .	69
5.2 Calage et analyse des incertitudes selon la méthode GLUE . . . . .	69
5.2.1 Pourquoi s'intéresser à cette méthode? . . . . .	69
5.2.2 Mise en œuvre de la méthode GLUE . . . . .	70
5.2.3 Analyse des jeux de paramètres . . . . .	71
5.2.4 Synthèse . . . . .	73
5.3 Calage et incertitude prédictive selon l'approche bayésienne . . . . .	74
5.3.1 Mise en œuvre du calage bayésien . . . . .	74
5.3.2 Analyse des jeux de paramètres . . . . .	76
5.3.3 Synthèse . . . . .	77

5.4	Approche multi-modèles . . . . .	78
5.5	Méthodes de post-traitement statistique . . . . .	79
5.6	Synthèse . . . . .	81
<b>6</b>	<b>Comparaison des méthodes utilisées en simulation</b>	<b>85</b>
6.1	Introduction . . . . .	87
6.2	Modalités de la comparaison . . . . .	87
6.2.1	Procédure de calage-contrôle . . . . .	87
6.2.2	Stratégie d'évaluation . . . . .	87
6.3	Évaluation des différentes méthodes en terme de fiabilité et de finesse . . . . .	88
6.3.1	Ensembles de paramètres et incertitude prédictive . . . . .	88
6.3.2	Méthodes de post-traitement . . . . .	91
6.3.3	Approche multi-modèles . . . . .	94
6.4	Analyse des performances globales . . . . .	98
6.5	Synthèse . . . . .	99
<b>7</b>	<b>Transferring model uncertainty estimates from gauged to ungauged catchments</b>	<b>103</b>
	Résumé . . . . .	105
	Abstract . . . . .	105
7.1	Introduction . . . . .	106
7.1.1	Predicting streamflow in ungauged catchments with uncertainty estimates	106
7.1.2	Scope of the paper . . . . .	107
7.2	Data and methods . . . . .	108
7.2.1	Data set . . . . .	108
7.2.2	Rainfall-runoff models . . . . .	109
7.2.3	Regionalisation approach . . . . .	109
7.3	Proposed approach : transfer of relative errors by flow groups . . . . .	110
7.4	Quantitative evaluation of uncertainty bounds . . . . .	111
7.5	Results and discussion . . . . .	115
7.5.1	Reliability, sharpness and overall skill . . . . .	115
7.5.2	Do we need to treat the donor catchments as ungauged? . . . . .	116
7.5.3	Do we need to use groups of relative errors? . . . . .	117
7.5.4	How do the performances of the rainfall-runoff models relate to the characteristics of uncertainty bounds? . . . . .	118
7.6	Conclusions . . . . .	119
7.7	Acknowledgments . . . . .	120
<b>8</b>	<b>Incertitude prédictive en prévision des crues</b>	<b>121</b>
8.1	Introduction . . . . .	123
8.1.1	Contexte de nos travaux en prévision des crues . . . . .	123
8.1.2	Quelques éléments de bibliographie . . . . .	124

8.2	Cadre de comparaison . . . . .	125
8.2.1	Présentation des méthodes testées . . . . .	125
8.2.2	Modalités d'application . . . . .	125
8.3	Présentation des résultats . . . . .	126
8.3.1	Fiabilité, finesse et performance globale . . . . .	126
8.3.2	Capacité à prévoir des dépassements de seuils . . . . .	128
8.4	Vers une meilleure caractérisation de l'incertitude prédictive en prévision . .	130
8.4.1	Motivations et description de l'approche . . . . .	130
8.4.2	Améliorations obtenues . . . . .	130
8.5	Synthèse . . . . .	132
<b>9</b>	<b>Interactions between data assimilation and post-processing</b>	<b>135</b>
	Résumé . . . . .	137
	Abstract . . . . .	137
9.1	Introduction . . . . .	138
9.1.1	Addressing uncertainties in hydrological ensemble forecasting . . . . .	138
9.1.2	Integrating uncertainties in hydrological ensemble forecasting . . . . .	139
9.1.3	Aim and scope of the study . . . . .	140
9.2	Data and methods . . . . .	140
9.2.1	Data set . . . . .	140
9.2.2	PEARP, the Météo-France ensemble forecast . . . . .	142
9.2.3	The GRP rainfall-runoff forecasting model . . . . .	142
9.2.4	Hydrological uncertainty processor . . . . .	143
9.2.5	Ensemble dressing method : an integrator of the meteorological and hydrological uncertainties . . . . .	144
9.2.6	Experiments . . . . .	144
9.2.7	Forecast evaluation methods . . . . .	145
9.3	Results and discussion . . . . .	146
9.3.1	Forecast accuracy . . . . .	146
9.3.2	Reliability . . . . .	148
9.3.3	Sharpness . . . . .	150
9.3.4	Mean CRPSS . . . . .	151
9.4	Summary and conclusions . . . . .	152
9.5	Appendix : Evaluation scores . . . . .	155
9.5.1	Relative bias . . . . .	155
9.5.2	Normalized root-mean-square error . . . . .	156
9.5.3	PIT diagram and alpha score . . . . .	156
9.5.4	Normalized mean interquartile range . . . . .	157
9.5.5	Mean CRPS and mean CRPSS . . . . .	157
9.6	Acknowledgments . . . . .	158

---

<b>Conclusion générale</b>	<b>159</b>
<b>Références bibliographiques</b>	<b>165</b>
<b>A Compléments au chapitre 2</b>	<b>183</b>
A.1 Liste des 202 bassins versants . . . . .	184
A.2 Fonctionnement du modèle GR4J . . . . .	188
A.2.1 Fonction de production . . . . .	189
A.2.2 Fonction de routage . . . . .	190
A.3 Fonctionnement du modèle GRP . . . . .	192
A.3.1 Fonction de production . . . . .	193
A.3.2 Fonction de routage . . . . .	194
A.3.3 Procédures de mises-à-jour . . . . .	194
<b>B Compléments au chapitre 5</b>	<b>197</b>
B.1 Inférence bayésienne . . . . .	198
B.2 Méthodes de post-traitement . . . . .	200
B.2.1 Préalable . . . . .	200
B.2.2 CemaM et CemaO . . . . .	201
B.2.3 MeGa . . . . .	203
B.2.4 MCP . . . . .	204
B.2.5 RQ . . . . .	205





## TABLE DES FIGURES

2.1	Contours des 202 bassins versants sélectionnés. . . . .	29
2.2	Représentation du bilan hydrologique des 202 bassins versants sélectionnés. . . . .	31
2.3	Distribution de quelques caractéristiques des 202 bassins versants sélectionnés. . . . .	32
2.4	Répartition géographique de quelques caractéristiques des 202 bassins versants sélectionnés. . . . .	33
2.5	Illustration des contextes de simulation et de prévision. . . . .	34
2.6	Schéma de la structure du modèle GR4J. . . . .	35
2.7	Schéma de la structure du modèle GRP. . . . .	37
3.1	Illustration d'une fonction de répartition $F_Q(q)$ . . . . .	44
3.2	Illustration de la valeur $p^z$ considérée dans le diagramme de PIT. . . . .	49
3.3	Illustration du diagramme de PIT. . . . .	50
3.4	Illustration du calcul du CRPS. . . . .	51
3.5	Illustration du calcul du score de Brier. . . . .	53
4.1	Common shapes and usual interpretation of the PIT diagram. . . . .	62
4.2	PIT diagrams calculated over the entire flow range, and PIT diagrams calculated for low and high flows. . . . .	63
4.3	Differences between Alpha scores when they are computed over the entire flow range or for either low flows or high flows. . . . .	65
5.1	Distributions des critères de performance NSQ obtenus pour 1000 jeux de paramètres, pour chaque bassin et chaque période. . . . .	72
5.2	Boxplots des écarts interquartiles des distributions marginales des paramètres transformés obtenus avec le critère NSQ. . . . .	73
5.3	Boxplots des valeurs médianes des distributions marginales des paramètres transformés obtenus avec le calage bayésien (MCMC). . . . .	77
5.4	Boxplots des écarts interquartiles des distributions marginales des paramètres transformés obtenus avec le calage bayésien (MCMC). . . . .	78
5.5	Illustration des différentes méthodes testées dans un contexte de simulation. . . . .	83
6.1	Boxplots des indices de fiabilité et de finesse des distributions prédictives obtenues avec la méthode GLUE et le calage bayésien (MCMC). . . . .	89
6.2	Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec la méthode GLUE pour différents seuils d'efficacité. . . . .	90

6.3	Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec le calage bayésien (MCMC). . . . .	92
6.4	Boxplots des indices de fiabilité et de finesse des distributions prédictives obtenues avec neuf méthodes de post-traitement statistique. . . . .	93
6.5	Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec neuf méthodes de post-traitement statistique. . . . .	95
6.6	Boxplots des indices de fiabilité et de finesse des distributions prédictives obtenues avec l'approche multi-modèles. . . . .	96
6.7	Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec l'approche multi-modèles. . . . .	97
6.8	Boxplots des indices de performance globale obtenus par l'ensemble des méthodes testées. . . . .	98
6.9	Indices de performance globale CRPSS. . . . .	100
7.1	Illustration of the proposed approach – Step 1. . . . .	111
7.2	Illustration of the proposed approach – Step 2a. . . . .	112
7.3	Illustration of the proposed approach – Step 2b. . . . .	112
7.4	Illustration of the proposed approach – Steps 3a and b. . . . .	113
7.5	Illustration of the proposed approach – Steps 4a and b. . . . .	113
7.6	Distributions of the three performance criteria. . . . .	115
7.7	Distributions of the three performance criteria, obtained in two cases, i) when the donor catchments are treated as ungauged, and ii) when the donor catchments are treated as gauged. . . . .	117
7.8	Distributions of the three performance criteria, obtained in two cases, i) when 10 groups of relatives errors are used, and ii) when only one group is used. . . . .	118
7.9	Impact of deterministic performance, as quantified by the bounded C2M quadratic criterion, on the three performance criteria for the 907 catchments. . . . .	119
8.1	Boxplots des indices de fiabilité et de finesse des distributions prédictives obtenues avec six méthodes de post-traitement statistique. . . . .	127
8.2	Boxplots des indices de performance globale obtenus avec six méthodes de post-traitement statistique. . . . .	128
8.3	Boxplots des <i>Brier Skill Score</i> obtenus avec six méthodes de post-traitement statistique. . . . .	129
8.4	Fréquences cumulées des gains de fiabilité, finesse et performance globale de la méthode CemaS par rapport à la méthode CemaM. . . . .	131
8.5	Diagrammes en nuage de points des gains de fiabilité et finesse obtenus avec la méthode CemaS par rapport à la méthode CemaM. . . . .	133
8.6	Fréquences cumulées des gains sur le score de Brier obtenus avec la méthode CemaS par rapport à la méthode CemaM. . . . .	134

---

9.1	Locations of the 202 French catchments used in this study. . . . .	141
9.2	Distributions of two deterministic scores, the relative bias (BIAS) and the normalized root-mean-square error (NRMSE). . . . .	147
9.3	PIT diagrams of the 24 h ahead streamflow ensemble forecasts. . . . .	149
9.4	Distributions of the alpha score reliability index. . . . .	150
9.5	Distributions of the normalized mean interquartile range (NMIQR). . . . .	152
9.6	Distributions of the mean CRPSS. . . . .	153
A.1	Schéma de la structure du modèle GR4J. . . . .	188
A.2	Schéma de la structure du modèle GRP. . . . .	192
B.1	Illustration de la transformation NQT. . . . .	201
B.2	Illustration de la méthode CemaM. . . . .	202
B.3	Illustration de la méthode MeGa. . . . .	204
B.4	Illustration de la méthode MCP. . . . .	205
B.5	Illustration de la méthode RQ. . . . .	206
B.6	Illustration des méthodes de référence. . . . .	208



## LISTE DES TABLEAUX

2.1	Description des indices utilisés pour la classification hiérarchique des bassins versants. . . . .	28
2.2	Paramètres du modèle GR4J. . . . .	36
2.3	Paramètres du modèle GRP. . . . .	38
3.1	Critères d'évaluation utilisés en fonction du type de prédiction probabiliste. . .	54
5.1	Liste des dix-sept modèles hydrologiques utilisés pour l'approche multi-modèles.	79
7.1	Characteristics of the 907 catchments. P–precipitation, PE–potential evapo- transpiration, Q–discharge. . . . .	108
8.1	Caractéristiques des méthodes de post-traitement testées en prévision. . . . .	125
9.1	Characteristics of the 202 catchments. P–precipitation, PE–potential evapo- transpiration, Q–discharge. . . . .	142
9.2	Acronyms used for the different experiments used in this study. . . . .	145
A.1	Liste des 202 bassins versants . . . . .	187



# INTRODUCTION GÉNÉRALE





## Contexte de la thèse et problématique scientifique

### Processus de modélisation et analyse des incertitudes

L'hydrologie s'intéresse au fonctionnement de systèmes complexes, naturels ou influencés par l'homme : les bassins versants. L'observation de ces systèmes est limitée, et la compréhension, l'identification des processus physiques qui caractérisent leur dynamique sont imparfaites. La démarche de modélisation de ces systèmes s'accompagne de fait de nombreuses incertitudes et ne peut proposer qu'une approximation, une représentation imparfaite et simplifiée de la réalité.

Malgré ces limites, auxquelles sont confrontées de manière plus générale toutes les sciences de l'environnement, la modélisation est devenue au cours des dernières décennies un outil privilégié pour de nombreuses activités de recherche et applications opérationnelles. En hydrologie, la modélisation est notamment utilisée pour la gestion des ressources en eau, la prévision des crues, le dimensionnement des ouvrages hydrauliques ou encore le test de différents scénarios.

L'utilisation croissante à des fins opérationnelles des résultats de modélisation, intégrés de plus en plus fréquemment dans des systèmes d'aide à la décision, s'est accompagnée d'une réflexion sur la fiabilité des modèles et leurs capacités à répondre de manière adéquate aux problèmes opérationnels. Plusieurs auteurs se sont ainsi penchés sur la question de l'évaluation des modèles (e.g., [Jakeman et al., 2006](#); [Bennett et al., 2013](#)), et différents cadres généraux de modélisation ont été proposés. Par exemple, [Refsgaard et Henriksen \(2004\)](#) et [Refsgaard et al. \(2005\)](#) se sont attachés à formaliser les bonnes pratiques à suivre en s'appuyant sur le concept d'« assurance qualité ».

La reconnaissance des limites des modèles, de leurs imperfections, a progressivement conduit à intégrer l'analyse des incertitudes dans le processus de modélisation ([Refsgaard et al., 2007](#); [Beven, 2009b](#)). Les prévisions d'un modèle ne sont qu'une approximation de la réalité, ce qui se manifeste en pratique par des erreurs, des écarts entre les valeurs prédites et les valeurs observées. Un des objectifs d'une analyse des incertitudes est de comprendre et quantifier cette approximation. L'analyse des incertitudes doit ainsi permettre de renforcer la confiance de l'utilisateur, en évaluant les risques associés à l'application des modèles, et fournir les éléments nécessaires à une prise de décision efficace en contexte incertain.

## Diversité des thématiques liées à l'analyse des incertitudes

Les méthodes d'analyse et de quantification des incertitudes diffèrent par leurs approches, les variables auxquelles elles s'intéressent, et les hypothèses de modélisation sur lesquelles elles reposent. La variété des méthodes proposées est une réponse à la diversité des thématiques que le sujet recouvre. [Montanari \(2011\)](#) évoque par exemple les thématiques suivantes :

- l'analyse et le traitement des données,
- l'estimation des paramètres des modèles hydrologiques,
- l'évaluation et l'identification des modèles,
- la quantification de l'incertitude totale sur les simulations hydrologiques.

Il s'agit donc avant tout de bien distinguer ces différentes questions et les objectifs des différentes méthodes. Certaines méthodes chercheront à quantifier de manière explicite les contributions de différentes sources d'incertitude, tandis que d'autres ne s'intéresseront qu'à la quantification de l'incertitude totale sur les résultats fournis par la modélisation hydrologique.

## Obstacles à la mise en œuvre des méthodes d'analyse d'incertitudes

La prise en compte des incertitudes dans le processus de modélisation est désormais considérée par de nombreux auteurs comme essentielle (voir par exemple [Ramos \*et al.\*, 2010](#)). Cependant, l'analyse et la quantification des incertitudes n'est pas devenue, en pratique, un exercice routinier.

Il existe plusieurs obstacles à la mise en pratique des méthodes d'analyse des incertitudes en modélisation hydrologique. [Pappenberger et Beven \(2006\)](#) ont par exemple présenté plusieurs raisons qui expliqueraient la réticence des modélisateurs et des utilisateurs à s'intéresser au sujet, concluant que la principale difficulté résidait dans la diversité et le manque de maturité de méthodes qui se distinguent par leurs philosophies et leurs approches.

En réponse à une discussion sur la pertinence des méthodes d'analyse des incertitudes initiée par [Beven \(2006b\)](#), [Montanari \(2007\)](#) et [Todini \(2007\)](#) avaient également souligné le manque de clarté dans la définition du problème, et les difficultés de communication liées à l'absence d'une terminologie commune. [Andréassian \*et al.\* \(2007\)](#) et [Hall \*et al.\* \(2007\)](#) avaient quant à eux insisté sur la nécessité d'évaluer de manière plus rigoureuse la capacité des différentes méthodes à restituer de manière fiable l'incertitude qui affectent les résultats des modèles, seul moyen de rendre crédible les estimations communiquées aux utilisateurs.

La diversité des approches proposées dans la littérature ne doit toutefois pas être un obstacle à leur mise en œuvre, et à leur évaluation. Malgré des divergences de vue sur la manière la plus appropriée de traiter les incertitudes en modélisation hydrologique, les discussions se sont le plus souvent concentrées sur des aspects théoriques ou d'ordre philosophique (voir par exemple les échanges récents entre [Beven et al., 2012](#) et [Clark et al., 2012](#)), et l'évaluation pragmatique des méthodes n'a reçu jusqu'ici qu'une attention limitée.

## Objectifs de cette thèse

Le principal objectif de cette thèse est d'explorer les méthodes de quantification de l'incertitude prédictive en modélisation hydrologique, de manière à proposer une approche suffisamment simple et générale pour être applicable à des bassins versants variés, et à divers modèles, dans des contextes de simulation et de prévision.

Pour ce faire, nous nous intéressons à différentes approches capables de compléter les résultats déterministes, en fournissant des distributions prédictives, et cherchons à mettre en évidence, à l'aide d'une approche comparative, leurs forces et leurs faiblesses.

## Organisation du mémoire

Ce mémoire est composée de neuf chapitres.

Le premier chapitre présente quelques rappels concernant les sources d'incertitudes rencontrées en modélisation hydrologique, et dresse un panorama général des principales méthodes d'analyse et de quantification des incertitudes utilisées en hydrologie.

Le deuxième chapitre est consacré à la présentation de l'échantillon de bassins versants retenus pour nos travaux, et aux modèles hydrologiques utilisés.

Le troisième chapitre détaille les critères d'évaluation probabiliste qui nous ont permis de comparer les différentes approches testées au cours de cette thèse.

Le quatrième chapitre correspond à une note technique, consacrée à un outil d'évaluation de la fiabilité des prévisions probabilistes, le diagramme de PIT.

Le cinquième chapitre présente les différentes méthodes qui sont comparées, dans un contexte de simulation, au chapitre six.

Le septième chapitre est un article consacré à la quantification de l'incertitude prédictive pour les bassins non-jaugés. Une méthode qui repose sur un transfert des marges d'erreurs des bassins jaugés aux bassins non-jaugés est proposée.

Le huitième chapitre présente une comparaison de différentes méthodes de post-traitement utilisées pour quantifier l'incertitude prédictive du modèle de prévision des crues GRP.

Le neuvième chapitre est un article qui s'intéresse aux interactions entre l'assimilation de données et le post-traitement d'un modèle de prévision au sein d'une chaîne de prévision hydrologique d'ensemble.

Le dernier chapitre résume les travaux effectués au cours de cette thèse, et évoque quelques perspectives.

# 1

## INCERTITUDES EN MODÉLISATION HYDROLOGIQUE



## 1.1 Introduction

Ce chapitre a pour objectif de dresser un panorama général des méthodes d'analyse et de quantification des incertitudes en modélisation hydrologique. De nombreuses méthodes ont été proposées ces dernières années, et il est difficile d'indiquer dans quelles conditions une méthode peut être utilisée avec confiance. Les avantages et les limites des différentes approches restent largement débattues.

Nous présentons dans un premier temps quelques généralités à propos des différentes sources d'incertitudes rencontrées en modélisation hydrologique. Nous décrivons ensuite un nombre restreint de méthodes qui nous semblent particulièrement intéressantes. Sans être exhaustive, cette présentation a pour objectif de donner un aperçu des principales techniques aujourd'hui disponibles, en discutant de leurs avantages et de leurs limites.

## 1.2 Caractérisation et représentation des incertitudes

### 1.2.1 Caractérisation des incertitudes

La caractérisation des incertitudes en hydrologie a fait l'objet de nombreuses contributions dans la littérature (Beck, 1987; Refsgaard *et al.*, 2007; Montanari *et al.*, 2009; Matott *et al.*, 2009). D'une manière générale, les incertitudes peuvent être caractérisées selon leur type et leurs sources.

En ce qui concerne le type d'incertitude, il est courant de distinguer l'incertitude réductible (ou épistémologique), qui résulte d'un manque de connaissance, de l'incertitude irréductible (ou structurelle, inhérente, stochastique), qui provient d'une variabilité naturelle.

Quant aux sources d'incertitude, les plus fréquemment évoquées sont les incertitudes sur les données d'entrées et les conditions aux limites, l'incertitude liée à l'estimation des paramètres et l'incertitude liée à la structure du modèle. Nous présentons brièvement ces différentes sources d'incertitude.

#### **Incertitude sur les données**

Les incertitudes sur les données issues de l'observation, qui comprennent à la fois les données de forçage comme les précipitations et l'évapotranspiration potentielle, et les données



de réponse comme le débit, résultent à la fois des erreurs sur les mesures mais également du traitement de ces données. Les précipitations sont notamment incertaines en raison de l'échantillonnage spatial du champ de précipitation, puisqu'un nombre limité de pluviomètres est utilisé pour estimer une lame d'eau moyenne sur un bassin versant. Quant aux débits, ils sont estimés en utilisant des hauteurs d'eau mesurées, et une courbe de tarage qui décrit la relation entre hauteur d'eau et débit dans une section. Cette transformation introduit une incertitude sur l'estimation des débits d'un cours d'eau.

Les incertitudes sur les données peuvent être analysées de manière indépendante par rapport au processus de modélisation, que ce soit pour les précipitations (Gotzinger et Bardossy, 2008; Moulin *et al.*, 2009; Renard *et al.*, 2011; McMillan *et al.*, 2011), ou les débits (Jalbert *et al.*, 2011; Le Coz *et al.*, 2014). Au contraire, certaines méthodes tentent de quantifier les incertitudes sur les données lors du processus d'inférence (Ajami *et al.*, 2007; Vrugt *et al.*, 2008; Koskela *et al.*, 2012).

### **Incertitude liée à l'estimation des paramètres**

Les modèles hydrologiques conceptuels utilisent des paramètres qui permettent d'ajuster la réponse du modèle à celle de la transformation pluie-débit d'un bassin versant. Ces paramètres apportent la flexibilité nécessaire pour adapter une structure mathématique fixe à différents bassins versants. Face aux difficultés rencontrées pour spécifier les valeurs des paramètres à partir d'autres informations, l'approche couramment adoptée consiste, lorsque des mesures de débits sont disponibles, à ajuster les paramètres de manière à reproduire au mieux les débits observés à l'exutoire du bassin versant considéré.

L'incertitude sur les paramètres résulte de ce processus d'estimation, et des choix associés à l'approche adoptée. Elle s'exprime le plus souvent par l'obtention d'un ensemble de différents jeux de paramètres, comme nous le verrons par la suite.

### **Incertitude liée au modèle**

L'incertitude liée au modèle traduit la capacité limitée d'un modèle à représenter la dynamique du système modélisé, en raison de sa structure simplifiée par rapport à la complexité du système modélisé, de sa résolution, ou de son implémentation numérique (voir notamment Kavetski et Fenicia, 2011), même en l'absence d'erreurs dans les données.

## Incertitude sur les forçages futurs

Dans un contexte de prévision, les forçages météorologiques futurs ne sont pas connus et constituent une source d'incertitude majeure pour la prévision hydrologique. Afin de tenir compte des incertitudes en modélisation atmosphérique, des prévisions météorologiques à court et moyen terme sont de plus en plus fréquemment proposées sous forme d'ensemble par les centres de prévisions.

Des systèmes de prévisions hydrologiques d'ensemble se sont développés en intégrant ces prévisions météorologiques d'ensemble dans des modèles hydrologiques, comme en témoignent par exemple, en France, les études de [Thirel \*et al.\* \(2008, 2010\)](#), [Randrianasolo \*et al.\* \(2010\)](#) et [Zalachori \*et al.\* \(2012\)](#).

Le développement des systèmes de prévisions hydrologiques d'ensemble constitue un thème de recherche à part entière (voir par exemple la synthèse de [Cloke et Pappenberger, 2009](#)), et de nombreux travaux sont actuellement réalisés au sein de l'initiative internationale HEPEX<sup>1</sup> (e.g., [Schaake \*et al.\*, 2007](#); [Thielen \*et al.\*, 2008](#)).

### 1.2.2 Représentation des incertitudes

La quantification des incertitudes nécessite une représentation mathématique de ce que l'on considère comme incertain. La manière la plus usuelle consiste à adopter la théorie des probabilités, et attribuer à un événement un nombre compris entre 0 et 1. Une probabilité nulle correspond à un événement impossible, alors qu'une probabilité égale à 1 exprime une réalisation certaine. L'usage des probabilités permet de quantifier l'imperfection de nos connaissances quant à la réalisation ou non d'un événement. Lorsque l'on s'intéresse à l'incertitude sur une variable continue, la manière la plus complète de caractériser l'incertitude est alors de l'exprimer à l'aide d'une densité de probabilité (voir le chapitre 3).

Certains auteurs considèrent que l'usage des probabilités n'est pas toujours approprié ([Brown, 2004](#); [Beven, 2009a](#)), et suggèrent d'adopter d'autres formalismes mathématiques. Parmi ces théories, [Montanari \*et al.\* \(2009\)](#) mentionnent par exemple la théorie des ensembles flous, la théorie des possibilités et la théorie de l'évidence. Ces théories non probabilistes n'ont cependant reçu qu'une attention très limitée en modélisation hydrologique.

En revanche, le recours à la notion de scénarios, pour tenir compte d'incertitudes difficilement quantifiables, est très fréquent. Cette approche consiste, de manière générale, à faire

---

1. <http://hepex.irstea.fr>

varier les hypothèses de modélisation. Différents scénarios sont considérés comme autant d'hypothèses plausibles, par exemple sur les variables climatiques (scénarios de changement climatique pour l'étude d'impact à long terme), ou sur les modèles eux-mêmes (de manière à quantifier l'incertitude liée à l'utilisation d'un unique modèle forcément imparfait). Une des difficultés liées à cette approche réside dans la manière d'interpréter et communiquer les résultats obtenus.

## 1.3 Présentation de différentes méthodes d'analyse des incertitudes en hydrologie

### 1.3.1 Tentatives de classification des méthodes d'analyse des incertitudes

Dans un contexte général de modélisation environnementale, [Matott et al. \(2009\)](#) ont proposé de classifier 65 méthodes dans sept catégories :

- *Les méthodes d'analyse des données* : pour évaluer la précision des données issues de l'observation ;
- *Les méthodes d'analyse de l'identifiabilité des modèles* : pour mettre en évidence l'inadéquation des données et suggérer des améliorations de la structure d'un modèle ;
- *Les méthodes d'estimation des paramètres* : pour quantifier l'incertitude sur les paramètres d'un modèle ;
- *Les méthodes d'analyse des incertitudes* : pour quantifier l'incertitude sur les sorties du modèle en propageant les incertitudes associées à plusieurs sources d'incertitude ;
- *Les méthodes d'analyse de sensibilité* : pour examiner dans quelle mesure les variations sur les entrées affectent les sorties d'un modèle ;
- *Les méthodes d'analyse multi-modèles* : pour évaluer l'incertitude d'un modèle ou générer des prévisions d'ensemble à partir de plusieurs modèles ;
- *Les méthodes bayésiennes* : pour quantifier l'incertitude à partir d'une connaissance *a priori* et de la confrontation des résultats aux données d'observation.

Chacune des 65 méthodes répertoriées peut appartenir à plusieurs catégories, en fonction de ses objectifs.

[Montanari \(2011\)](#) se concentre uniquement sur la modélisation hydrologique, et présente un nombre limité de méthodes selon la source d'incertitude à laquelle elles s'intéressent principalement. Il distingue les méthodes qui estiment l'incertitude totale de manière agrégée, et

les méthodes qui considèrent séparément l'incertitude sur les données, l'incertitude sur les paramètres, ou l'incertitude liée à la structure d'un modèle.

L'objectif de la présentation de [Beven \(2009b\)](#) est de fournir une introduction générale au traitement de l'incertitude en modélisation environnementale. [Beven \(2009b\)](#) distingue différentes applications pratiques rencontrées en modélisation hydrologique :

- les situations où les données ne sont pas disponibles, comme pour les bassins non jaugés ou les projections de changement climatique ;
- le cas de la simulation, où des données permettent de réaliser une inférence sur certains paramètres ;
- et celui de la prévision à court et moyen terme, qui se caractérise à la fois par la prévision des données de forçage à court et moyen terme, et par l'utilisation de stratégies d'assimilation de données.

Ces tentatives de classification mettent en évidence l'importance de considérer à la fois les objectifs des méthodes, leur formalisme, et les techniques qu'elles utilisent. Elles révèlent aussi la difficulté à présenter de manière cohérente un ensemble de méthodes aussi variées. Il n'existe pas à notre connaissance de théorie générale du traitement des incertitudes en hydrologie, et nous ne prétendons pas fournir ici un cadre général unificateur.

Loin de prétendre à l'exhaustivité, nous choisissons de présenter les catégories de méthodes suivantes :

- la méthode GLUE ;
- l'approche bayésienne et les différentes méthodes d'inférence associées ;
- l'approche multi-modèles ;
- des méthodes de post-traitement qui cherchent à quantifier l'incertitude prédictive de manière globale.

### 1.3.2 Présentation de la méthode GLUE

Nous présentons la méthode GLUE. Il s'agit vraisemblablement de la méthode la plus communément utilisée dans la littérature hydrologique pour estimer les incertitudes.

#### Motivations et description de l'approche

La méthode GLUE (Generalised Likelihood Uncertainty Estimation) a été introduite par [Beven et Binley \(1992\)](#) comme une méthode générale de calage et d'analyse des incertitudes. La

méthode s'inspire de l'analyse de sensibilité de [Hornberger et Spear \(1981\)](#). Elle a été proposée à la suite d'expériences de modélisation qui révélèrent que différents jeux de paramètres pouvaient représenter de manière acceptable le comportement d'un système complexe. Le parti pris des auteurs est de renoncer à la recherche d'un unique jeu de paramètres, et d'accepter et assumer la multiplicité des jeux de paramètres.

Dans le cadre très général proposé par la méthode GLUE, la performance des différents jeux de paramètres est évaluée selon un ou plusieurs critères choisis pour refléter le degré de crédibilité accordé à chacun des jeux de paramètres ([Smith \*et al.\*, 2008](#)). Une valeur seuil doit également être choisie pour distinguer les jeux de paramètres acceptables des jeux de paramètres non acceptables au regard du critère d'efficacité choisi. Enfin, les débits simulés à partir des différents jeux de paramètres retenus sont utilisés pour calculer les intervalles prédictifs. Il est courant de pondérer les simulations qui constituent l'ensemble final selon le critère de performance, les simulations les plus performantes obtenant plus de poids que les simulations les moins performantes.

L'importance des choix subjectifs associés à l'utilisation de la méthode GLUE a été discutée entre autres par [Freer \*et al.\* \(1996\)](#), [Beven et Freer \(2001\)](#) et [Beven \(2006a\)](#). Le choix de la fonction objectif et le choix de la valeur seuil influencent la sélection des jeux de paramètres considérés comme acceptables et donc les intervalles prédictifs obtenus. La subjectivité de ces choix a été vivement critiquée par certains auteurs, comme nous le verrons plus loin. La littérature abonde néanmoins d'applications de la méthode GLUE en hydrologie et dans de nombreuses autres sciences de l'environnement. La simplicité et la flexibilité de la méthode GLUE sont vraisemblablement les principales raisons de cette large dissémination.

## Évaluations et critiques

Bien que l'utilisation de la méthode GLUE se soit remarquablement répandue dans de nombreux domaines des sciences de l'environnement, l'importance des choix intrinsèques à l'application de GLUE n'a été évaluée que par quelques auteurs.

[Montanari \(2005\)](#), [Xiong \*et al.\* \(2009\)](#) et [Li \*et al.\* \(2010\)](#) ont présenté des résultats qui semblent confirmer l'intuition selon laquelle la variabilité de l'ensemble prédictif augmente lorsque la valeur seuil diminue. Bien que la méthode repose sur une estimation de l'incertitude sur les paramètres plus importante que celle proposée par les méthodes d'inférence statistique, plusieurs travaux, dont ceux de [Montanari \(2005\)](#), [Berthier \(2005\)](#), et [Andréassian \*et al.\* \(2007\)](#), ont souligné que la méthode a tendance à sous-estimer l'incertitude prédictive totale.

D'un point de vue plus théorique, [Mantovan et Todini \(2006\)](#) ont critiqué l'incohérence de la méthode par rapport à l'approche bayésienne, en raison de l'utilisation de critères de performance qui contredisent la cohérence du processus d'assimilation de l'information contenue dans les observations. [Stedinger et al. \(2008\)](#) illustrent quant à eux l'incapacité de la méthode GLUE à restituer les intervalles de confiance dans un cas simple de régression linéaire. Ils mettent en évidence le manque de sensibilité de la méthode à la qualité intrinsèque du modèle utilisé. Comme [Mantovan et al. \(2007\)](#), les auteurs rappellent que la méthode GLUE ignore le modèle d'erreur et ne peut proposer qu'une quantification arbitraire des incertitudes sur la variable simulée. Ils considèrent donc que le cadre théorique de GLUE, notamment promu par le manifeste de [Beven \(2006a\)](#) en faveur du concept d'équifinalité, ne permet pas d'apporter de réponses satisfaisantes au traitement des incertitudes, et recommandent donc la plus grande prudence dans l'utilisation des résultats obtenus par l'application de la méthode GLUE.

### Développements récents et variantes

[Beven \(2006a\)](#) a introduit une extension du concept d'acceptabilité en proposant une nouvelle approche pour identifier les modèles considérés comme acceptables. Il propose dans cette nouvelle version de GLUE de définir les limites d'acceptabilité en spécifiant *a priori* une certaine marge d'erreur associée aux observations. Seuls les paramètres qui permettent d'obtenir des débits simulés se situant à l'intérieur de cette enveloppe définie *a priori* sont considérés comme acceptables. Les pondérations utilisées pour créer les intervalles prédictifs sont définies à l'aide de scores attribués à chaque jeu de paramètres considérés comme acceptables, ces scores pouvant par exemple dépendre de la position de la valeur simulée par rapport à l'intervalle défini autour de la valeur observée. Cette approche a été mise en œuvre dans plusieurs études ([Blazkova et Beven, 2009](#); [Liu et al., 2009](#); [Krueger et al., 2010](#)), mais n'a pas pour l'instant suscité un engouement comparable à la version originale de GLUE, qui reste majoritairement utilisée dans des articles récents.

Par ailleurs, des méthodes qui s'appuient sur des ensembles contraints de jeux de paramètres ont été proposées dans le cadre de la modélisation des bassins non jaugés ([Bardossy, 2007](#); [Yadav et al., 2007](#); [Zhang et al., 2008](#); [Winsemius et al., 2009](#); [Bulygina et al., 2009](#)). Les jeux de paramètres considérés comme acceptables sont ceux qui reproduisent certaines caractéristiques de la réponse hydrologique d'un bassin versant, ces caractéristiques ou « signatures » pouvant être estimées à l'aide de techniques de régionalisation. [Wagner et Montanari \(2011\)](#) ont récemment présenté une synthèse de ces approches qui ont pour objectif de réduire les incertitudes pour les bassins non jaugés.

### 1.3.3 Présentation de l'approche bayésienne

#### Motivations et description de l'approche

Considérer le calage d'un modèle hydrologique comme un problème de régression non-linéaire permet d'utiliser des techniques statistiques bien connues pour l'estimation des paramètres et la quantification des incertitudes.

L'inférence bayésienne s'inscrit dans une démarche qui semble particulièrement bien adaptée au traitement des incertitudes. Selon le paradigme bayésien, toutes les inconnues d'un problème sont décrites par des densités de probabilité. La densité de probabilité jointe des paramètres du modèle statistique considéré s'obtient par la formule de Bayes qui permet une mise à jour d'un état de connaissance, formalisé par des distributions *a priori*, à la lumière de nouvelles informations – les données issues de l'observation. L'estimation des paramètres est conditionnelle aux observations et aux hypothèses faites sur le mécanisme générateur des données. Les densités de probabilités d'autres variables d'intérêt, notamment la densité prédictive des débits, s'obtiennent par propagation des différentes sources d'incertitude considérées.

L'inférence bayésienne a été introduite en modélisation hydrologique comme une alternative aux stratégies de calage traditionnelles. [Kuczera et Parent \(1998\)](#) ont par exemple placé le calage d'un modèle hydrologique dans un cadre statistique bayésien, et quantifié l'incertitude sur les paramètres d'un modèle conceptuel à l'aide de méthodes de simulation de type Monte Carlo. Cette approche a été adoptée dans de nombreux travaux en modélisation hydrologique ([Bates et Campbell, 2001](#); [Vrugt et al., 2003a](#); [Engeland et al., 2005](#); [Yang et al., 2007a](#); [Smith et al., 2008](#); [Laloy et Vrugt, 2012](#), parmi d'autres).

Les travaux les plus récents se sont principalement concentrés sur deux aspects du problème d'inférence :

- la recherche d'un modèle d'erreur global adapté à la modélisation hydrologique ;
- la définition de modèles statistiques d'erreurs pour chacune des sources d'incertitude, de manière à mieux identifier leur contributions à l'incertitude totale.

#### Choix d'une fonction de vraisemblance adaptée aux erreurs résiduelles

La définition d'une fonction de vraisemblance est fondamentale lors du processus d'inférence, puisqu'elle permet de classer et pondérer les différentes hypothèses considérées. Les

premières approches qui ont cherché à spécifier une fonction de vraisemblance adaptée au calage des modèles hydrologiques remontent au début des années 80, lorsque l'utilisation de la méthode du maximum de vraisemblance a été proposée par [Sorooshian et Dracup \(1980\)](#). Il s'agissait alors de mieux respecter les hypothèses d'indépendance et d'homoscédasticité des erreurs résiduelles, grâce à des transformations sur les variables, et à l'utilisation de modèles auto-régressifs.

Différentes fonctions de vraisemblance, dérivées de modèles statistiques d'erreurs de plus en plus complexes, ont depuis été proposées par différents auteurs. [Schaepli et al. \(2007\)](#) ont proposé un modèle de mélange gaussien, [Yang et al. \(2007a,b\)](#) un modèle auto-régressif continu, et [Schoups et Vrugt \(2010\)](#) un modèle d'erreur flexible qui repose sur une composante auto-régressive, et l'utilisation de la loi normale généralisée.

### Séparation des erreurs

Au contraire de l'approche globale qui vient d'être décrite, d'autres recherches se sont concentrées sur la décomposition de l'incertitude totale en différentes sources d'incertitude. La méthode BATEA, *Bayesian Total Error Analysis* ([Kuczera et al., 2006](#); [Kavetski et al., 2006b,a](#); [Thyer et al., 2009](#); [Renard et al., 2010](#)), a par exemple été conçue pour caractériser, quantifier et propager différentes sources d'incertitude (données, paramètres et structure du modèle), de manière à diagnostiquer les principales causes d'erreurs, révéler les faiblesses d'un modèle, et estimer l'incertitude sur les paramètres. Dans cette approche, la fonction de vraisemblance utilisée lors du processus d'inférence ne s'exprime plus par un modèle probabiliste de l'écart entre les valeurs simulées par le modèle hydrologique et les valeurs issues de l'observation, mais par un modèle probabiliste plus complexe obtenu par l'utilisation de probabilités conditionnelles. Des travaux similaires ont été présentés par [Huard et Mailhot \(2006\)](#), [Huard et Mailhot \(2008\)](#), [Ajami et al. \(2007\)](#) et [Vrugt et al. \(2008\)](#).

La séparation des sources d'incertitude peut également être réalisée dans un cadre d'assimilation de données. [Salamon et Feyen \(2010\)](#) utilisent par exemple une méthode de Monte Carlo séquentielle connue sous le nom de filtre à particules pour mettre à jour les états d'un modèle hydrologique distribué. Les méthodes d'assimilation de données permettent également d'estimer de manière jointe les incertitudes sur les états et sur les paramètres d'un modèle hydrologique ([Moradkhani et al., 2005, 2012](#); [Vrugt et al., 2013](#)).



## Difficultés et critiques

Malgré le développement d'algorithmes de plus en plus performants (par exemple les méthodes présentées par [Vrugt et al., 2003a, 2008](#)), l'échantillonnage de la densité de probabilité jointe qui décrit l'incertitude sur les paramètres du modèle probabiliste – à la fois les paramètres du modèle hydrologique et les paramètres qui apparaissent dans les modèles d'erreurs – constitue toujours une difficulté lorsque le problème devient complexe, par exemple lorsque le nombre d'inconnues augmente, ou quand une approche hiérarchique est adoptée ([Kuczera et al., 2010](#)).

Mais au delà des difficultés numériques, les approches qui visent à séparer les contributions des différentes erreurs sont confrontées aux difficultés suivantes :

- Le choix des modèles statistiques d'erreurs, parfois subjectif, influence l'inférence des paramètres, et l'incertitude prédictive (voir parmi d'autres, [Weerts et El Serafy, 2006](#); [Vrugt et al., 2008](#); [Thyer et al., 2009](#); [Salamon et Feyen, 2010](#)). ;
- La spécification d'un modèle d'erreur capable d'exprimer les erreurs dues à l'imperfection du modèle semble extrêmement délicate ([Beven, 2006a](#)), et aucune approche ne semble pour l'instant réellement satisfaisante, malgré les tentatives de [Kuczera et al. \(2006\)](#), [Reichert et Mieleitner \(2009\)](#), ou [Renard et al. \(2010\)](#) ;
- Des compensations, par exemple une correction des précipitations qui compense un problème de structure, peuvent conduire à une inférence incorrecte ou mal posée, comme l'ont montré les travaux de [Renard et al. \(2010\)](#) et [Kuczera et al. \(2010\)](#). [Renard et al. \(2011\)](#) ont montré que seule une meilleure estimation indépendante de l'incertitude sur les données peut permettre de dissocier les contributions de différentes sources d'incertitude.

### 1.3.4 Présentation de l'approche multi-modèles

#### Motivations et description de l'approche

L'approche multi-modèles consiste à utiliser plusieurs modèles hydrologiques considérés comme des hypothèses plausibles du fonctionnement d'un bassin versant. Cette approche est fondée sur la reconnaissance de l'imperfection des modèles pluie-débit, et fait le pari qu'une bonne description de l'incertitude des structures peut être obtenue en utilisant plusieurs modèles. Elle vise deux principaux objectifs : l'amélioration des prévisions par combinaison

des résultats obtenus par plusieurs modèles, et la prise en compte des incertitudes liées à l'imperfection des modèles (e.g., [Georgakakos et al., 2004](#); [Velazquez et al., 2010](#)).

### Présentation de quelques méthodes

Plusieurs méthodes ont été proposées dans la littérature pour combiner les résultats obtenus par un ensemble de modèles. Elles diffèrent dans la manière dont sont pondérés les résultats de chaque modèle (voir par exemple les huit méthodes testées par [Diks et Vrugt, 2010](#)), les poids pouvant être fixes ou variables au cours du temps ([Marshall et al., 2007](#); [Hsu et al., 2009](#)). La plupart partagent le fait d'attribuer plus d'importance aux modèles les plus performants.

La méthode BMA, *Bayesian Model Averaging* ([Hoeting et al., 1999](#)), traite les incertitudes liées aux structures des modèles à l'aide d'une combinaison probabiliste. Une densité de probabilité est construite à l'aide d'un modèle de mélange composé des prévisions de plusieurs modèles. Des applications de la méthode BMA en modélisation hydrologique sont par exemple présentées par [Duan et al. \(2007\)](#) et [Ajami et al. \(2007\)](#).

### Limites de l'approche

Si les approches multi-modèles semblent capables de refléter en partie les incertitudes liées aux imperfections des modèles, la quantification d'une incertitude prédictive totale à partir de ces approches est loin d'être évidente. La variabilité apportée par l'utilisation de différents modèles ne garantit pas une quantification fiable de l'incertitude totale. [Velazquez et al. \(2010\)](#) suggèrent par exemple qu'un ensemble de dix-sept modèles hydrologiques ne semble pas suffisant pour obtenir une quantification fiable des incertitudes pour la plupart des 1061 bassins versants français utilisés dans leur étude. Les auteurs évoquent la nécessité d'un post-traitement de l'ensemble constitué par les simulations des différents modèles pour une utilisation opérationnelle.

## 1.3.5 Présentation de l'approche de post-traitement

### Motivations et description de l'approche

Les méthodes de post-traitement s'appuient sur des simulations déterministes obtenues avec un modèle hydrologique préalablement calé. Elles ne s'intéressent ni à l'incertitude liée à l'estimation des paramètres, ni aux caractéristiques d'autres sources d'incertitude.

Elles adoptent au contraire une approche pragmatique, en se concentrant sur l'analyse des propriétés statistiques des erreurs résiduelles.

L'idée principale est que l'incertitude globale se manifeste par des erreurs qui traduisent l'écart entre les valeurs simulées d'un modèle déterministe et les valeurs observées. Ces séries temporelles d'erreurs ont généralement une structure qui diffère de celle d'un signal aléatoire, et contiennent donc une information qui peut être exploitée pour quantifier l'incertitude prédictive de manière agrégée.

Les méthodes proposées dans la littérature se distinguent principalement par les techniques statistiques qu'elles utilisent. Le choix des variables explicatives, l'utilisation de transformations de variables, et le type de régression utilisé sont les principaux éléments qui les caractérisent.

### Présentation de quelques méthodes

Nous présenterons plus en détails les méthodes que nous avons utilisés dans cette thèse au chapitre 5. Nous mentionnons ici d'autres méthodes proposées dans des contextes de simulation et de prévision.

La méthode UNEEC, *Uncertainty Estimation based on local Errors and Clustering*, décrite par [Solomatine et Shrestha \(2009\)](#) dans un contexte de simulation hydrologique, propose d'utiliser les techniques d'apprentissage automatique pour quantifier la distribution des erreurs d'un modèle. La distribution des erreurs empiriques est conditionnée par les données d'entrée et les variables d'état du modèle. Une classification des données est au préalable réalisée pour créer des groupes de données ayant des caractéristiques hydrologiques semblables.

[Krzysztofowicz \(1999\)](#) a proposé une approche bayésienne du traitement des incertitudes prédictives pour la prévision des crues. Le processeur d'incertitude hydrologique présenté par [Krzysztofowicz \(2001\)](#) puis [Maranzano et Krzysztofowicz \(2004\)](#), *Bayesian Processor of Output*, utilise une information *a priori*, construite à partir d'un modèle auto-régressif, et une fonction de vraisemblance. Des applications de cette approche sont présentées par [Reggiani et Weerts \(2008\)](#), [Bogner et Pappenberger \(2011\)](#), et [Biondi et De Luca \(2012\)](#).

### Limites des méthodes de post-traitement

Les méthodes de post-traitement s'appuient sur une analyse et une description des erreurs résiduelles obtenues par un modèle hydrologique sur une période donnée. Elles caracté-

risent de manière globale les erreurs qui résultent d'une combinaison de plusieurs sources d'incertitude, et ne permettent donc pas de séparer la contribution des différentes sources d'incertitude.

## 1.4 Synthèse

La quantification des incertitudes fait désormais partie intégrante de la modélisation hydrologique. Il existe dans la littérature de très nombreuses contributions à ce sujet, ce qui traduit la diversité des questions qui peuvent être associées à cette thématique. S'intéresser aux incertitudes, c'est de manière très générale s'intéresser aux problèmes de la qualité des données utilisées, à la question de l'estimation des paramètres, ou encore aux problématiques liées à l'évaluation et l'identification des modèles hydrologiques.

Les différentes méthodes proposées ces dernières années peuvent être caractérisées selon :

- le contexte de modélisation : simulation ou prévision ;
- le formalisme mathématique : approches probabilistes ou non probabilistes ;
- les différentes sources d'incertitude quantifiées : données, paramètres, modèle ;
- et l'objectif poursuivi : meilleure connaissance du modèle, du système modélisé, ou quantification des marges d'erreurs sur les sorties du modèles.

Il n'existe pas de consensus sur les méthodes les plus appropriées pour aborder la thématique des incertitudes en modélisation hydrologique, et aucune méthode ne peut prétendre répondre de manière adéquate aux différentes attentes exprimées par les utilisateurs ou les gestionnaires.



# 2

## DONNÉES ET MODÈLES PLUIE-DÉBIT



## 2.1 Introduction

Nous présentons dans ce chapitre les données et les modèles hydrologiques utilisés au cours de cette thèse.

La constitution d'un large échantillon de bassins versants s'inscrit dans une démarche scientifique adoptée par l'équipe Hydrologie d'Irstea Antony au sein de laquelle ces travaux ont été réalisés. La multiplication des expériences permet en effet d'évaluer la robustesse et la généralité des méthodes de quantification de l'incertitude prédictive testées, conditions qui nous semblent nécessaires pour des applications opérationnelles, et qui correspondent à nos objectifs de recherche.

Nous avons travaillé au pas de temps journalier en simulation, et au pas de temps horaire en prévision. Cette distinction résulte de considérations avant tout pragmatiques : le choix du pas de temps journalier en simulation s'est imposé pour des raisons de temps de calcul, tandis que le choix du pas de temps horaire reflète la nécessité d'utiliser un pas de temps permettant de bien appréhender la dynamique des crues.

## 2.2 Données

### 2.2.1 Bases de données hydro-climatiques

#### Base de données au pas de temps journalier

La base de données hydro-climatiques au pas de temps journalier utilisée dans cette thèse a été constituée au cours de l'année 2011 au sein de l'équipe Hydrologie d'Irstea Antony, en collaboration avec le SCHAPI et Météo-France. Cette base de données comprend, pour 3701 bassins versants français :

- des données de débits de la banque HYDRO ;
- des données météorologiques issues des réanalyses SAFRAN (résolution de 8 km) ;
- des caractéristiques morphologiques issues d'un modèle numérique de terrain (résolution de 100 m).

La disponibilité des données de débits est inégale suivant les bassins versants, tandis que les données météorologiques SAFRAN couvrent la période 1958-2010 de manière continue (sans lacune).



## Base de données au pas de temps horaire

La base de données utilisée au pas de temps horaire a, quant à elle, été constituée par Florent Lobligois, dans le cadre de sa thèse à Irstea. Nous rappelons ici brièvement les différents traitements de données effectués.

Les données de débits horaires ont été générées à partir des données au pas de temps variable issues de la banque HYDRO. La méthode utilisée suit celle présentée par [Le Moine \(2008\)](#). En particulier, le calcul des débits moyens horaires, calculés par intégration sur chaque heure, s'appuie sur une interpolation linéaire des données au pas de temps variable.

Les données de précipitation proviennent d'une réanalyse des lames d'eau combinant images radar, pluviographes et pluviomètres, récemment proposée par Météo-France. Cette réanalyse, établie sur 10 ans (1997–2006), est disponible au pas de temps horaire et à la résolution kilométrique sur l'ensemble du territoire métropolitain ([Tabary et al., 2012](#)).

Enfin, les données d'évapotranspiration potentielle ont été obtenues par désagrégation des données journalières SAFRAN. Les valeurs horaires suivent une parabole entre 06h00 et 19h00, et sont nulles le reste du temps.

### 2.2.2 Constitution d'un échantillon de bassins versants

La sélection des bassins versants décrite ci-dessous a été effectuée en utilisant la base de données au pas de temps journalier. Nous avons cependant gardé à l'esprit l'objectif de travailler au pas de temps horaire, en nous assurant de la disponibilité des données sur chacune des périodes 1997–2001 et 2002–2006.

#### Pré-sélection

Dans un premier temps, une pré-sélection sur les 3701 stations disponibles a été effectuée :

- Nous avons conservé les bassins dont la surface est supérieure à  $5 \text{ km}^2$  et inférieure à  $50\,000 \text{ km}^2$ .
- Nous avons exclu les bassins versants qui, d'après les méta-données de la banque HYDRO, sont répertoriés comme fortement influencés ou dont la finalité n'est ni l'hydrométrie générale, ni l'alerte de crue ou le suivi d'étiage.
- Nous avons utilisé une estimation des volumes stockés par des barrages situés à l'amont des stations pour exclure les stations potentiellement influencées, en choisissant un

seuil correspondant à un volume stockable de 10 mm de lame d'eau répartie sur le bassin versant.

- Nous avons retenu les stations présentant au moins 20 années de disponibilité sur la période 1980–2009, avec un taux de lacunes annuel inférieur à 10%. Une contrainte supplémentaire sur la disponibilité des débits a également été utilisée : un minimum de 4 ans sur chacune des périodes 1997–2001 et 2002–2006 (avec un taux de lacunes annuel inférieur à 10%) a été requis.
- Enfin, nous avons cherché à éviter les bassins versants fortement influencés par la neige. Pour cela, nous avons exclu les bassins versants pour lesquels le rapport entre précipitations solides et précipitations totales était supérieur à 10%.

A l'issue de cette pré-sélection, 583 bassins versants restaient disponibles.

### Sélection de bassins versants

Nous avons choisi de ne pas conserver l'ensemble de ces bassins versants et de procéder à une sélection d'un peu plus de 200 bassins versants. Il s'agissait avant tout de limiter les temps de calcul, puisque certaines des méthodes utilisées pour la quantification de l'incertitude prédictive requièrent de réaliser de très nombreuses simulations. Il nous a semblé intéressant de profiter de cette nécessaire étape de sélection pour favoriser la diversité hydrologique au sein de l'ensemble de bassins versants retenus, en suivant la démarche adoptée dans la thèse de [Berthet \(2010\)](#).

La répartition géographique de l'ensemble des bassins versants disponibles sur le territoire métropolitain était assez inégale, avec notamment une forte représentation des bassins versants du Nord-Est de la France. La constitution d'un échantillon restreint de bassins versants, sélectionnés suivant certains indices qui caractérisent les régimes hydrologiques et pluvieux, nous a permis de limiter ce déséquilibre.

Nous avons utilisé les mêmes indices que [Berthet \(2010\)](#) pour décrire les régimes hydrologiques et pluvieux. Six indices caractérisent le régime d'écoulement, les étiages, les crues, et le débit de base, tandis que quatre indices sont utilisés pour caractériser les précipitations. Les dix indices sont présentés dans la Table 2.1.

Nous nous sommes appuyés sur ces indices pour réaliser une classification hiérarchique des 583 bassins versants. Les méthodes de classification hiérarchique permettent de constituer des groupes d'éléments homogènes à partir d'une matrice de similarité. Pour chaque paire de bassins versants, la similarité entre les deux bassins a été calculée selon la distance euclidienne entre les dix indices centrés et réduits de chacun des deux bassins. La méthode de

Indice	Description
Écoulement annuel moyen	Lame d'eau, exprimée en mm, qui correspond au débit annuel moyen
Irrégularité des écoulements	Différence entre les moyennes des débits mensuels les plus forts et les plus faibles sélectionnés chaque année, divisée par le débit annuel moyen
Débit spécifique d'étiage	Médiane du dixième débit le plus faible de chaque année, exprimée en $L s^{-1} km^{-2}$
Étiage modulaire	Débit spécifique d'étiage normalisé par un dixième de l'écoulement annuel moyen
Indice d'écoulement de base	Rapport entre le volume de base et le volume d'écoulement total
Coefficient de crue	Médiane des plus forts débits journaliers annuels, divisée par la surface du bassin versant élevée à la puissance 0.8
Précipitation annuelle moyenne	Lame d'eau, exprimée en mm, précipitée en moyenne pendant une année
Irrégularité des pluies	Différence entre les moyennes des précipitations mensuelles les plus fortes et les plus faibles sélectionnés chaque année, divisée par la précipitation annuelle moyenne
Intensité des pluies	Médiane des plus fortes précipitations journalières annuelles
Fraction de jours non pluvieux	

TABLE 2.1 – Description des indices utilisés pour la classification hiérarchique des bassins versants.

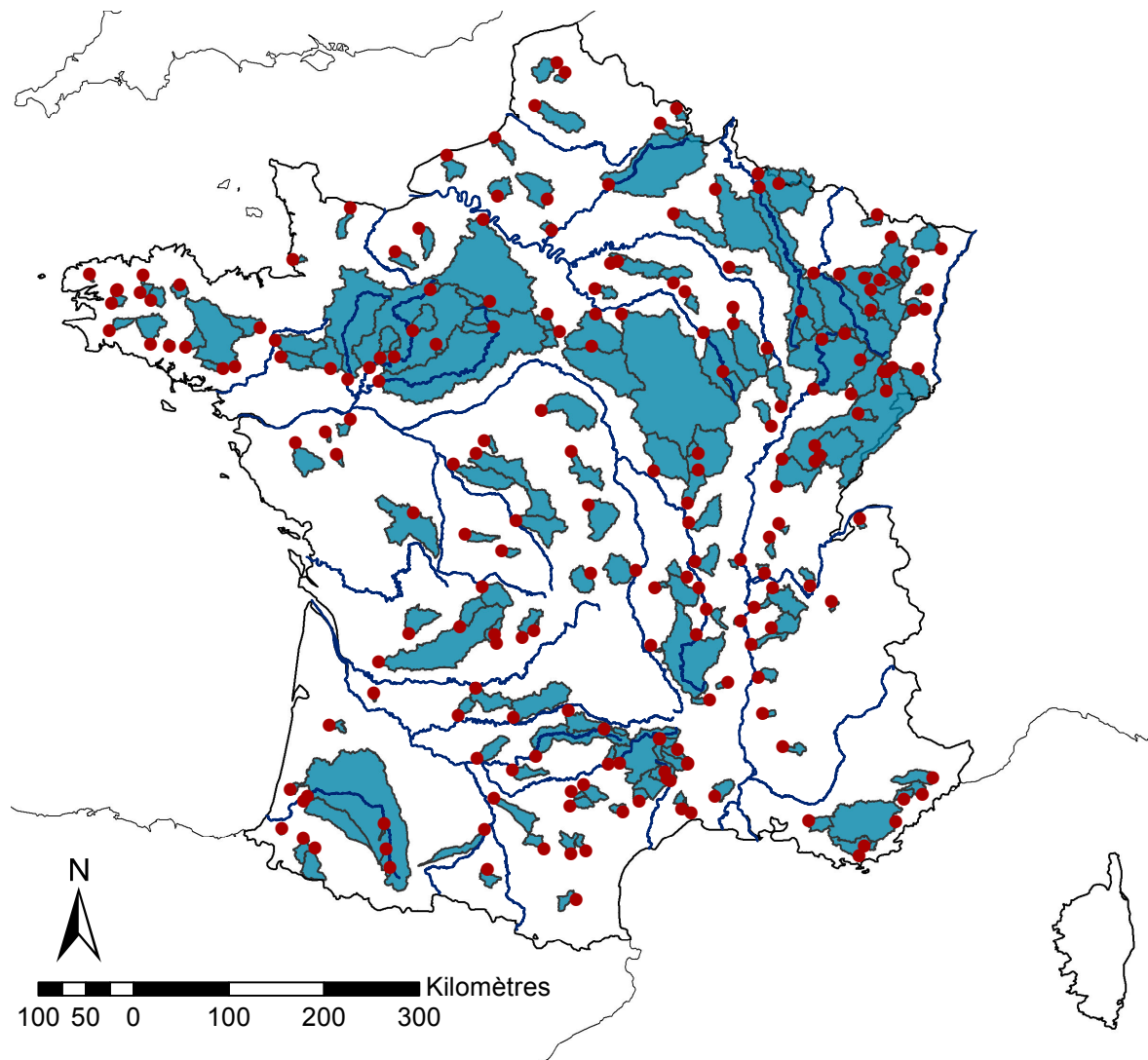


FIGURE 2.1 – Contours des 202 bassins versants sélectionnés, en bleu. Les stations situées à l'exutoire des bassins versants sont représentées par des points rouges.

classification développée par [Bien et Tibshirani \(2011\)](#) nous a semblé intéressante, puisqu'elle permet de désigner un élément représentatif au sein de chaque groupe constitué lors de la classification hiérarchique. Nous avons ainsi sélectionné 100 bassins versants issus du dendrogramme – l'arborescence obtenue par la classification hiérarchique –, auxquels nous avons ajouté les bassins versants les plus éloignés des bassins versants représentatifs.

La classification hiérarchique nous a ainsi permis de sélectionner de manière automatique des bassins versants aux caractéristiques variées. Il nous a toutefois semblé nécessaire de conserver une certaine flexibilité dans la sélection des bassins. Une dernière étape manuelle a permis d'ajouter et d'enlever certains bassins versants, de manière à améliorer la répartition géographique des bassins sélectionnés, et s'assurer que certains bassins pour lesquels nous avons un intérêt particulier soient bien présents dans l'échantillon final. Nous avons par exemple essayé de conserver des bassins versants suivis par les Services de Prévion des

Crues (SPC). A l'issue de cette dernière étape, 202 bassins versants ont été sélectionnés.

Les contours des 202 bassins versants qui constituent l'échantillon final sont tracés sur la Figure 2.1. Cet échantillon comprend des bassins de taille variée, répartis sur l'ensemble du territoire métropolitain, à l'exception des zones montagneuses. Une liste des 202 bassins est présentée dans l'annexe A.1.

### Caractéristiques des bassins versants sélectionnés

La Figure 2.2 présente les 202 bassins versants sélectionnés dans l'espace adimensionnel  $Q/P = f(P/ETP)$ , où Q, P et ETP désignent respectivement les moyennes annuelles des débits, des précipitations et de l'évapotranspiration potentielle; Q/P est le coefficient de ruissellement et P/ETP est l'indice d'aridité. Ce mode de présentation permet de visualiser les principales caractéristiques du bilan hydrologique à l'échelle des bassins versants (Le Moine *et al.*, 2007). Les deux zones grisées correspondent à des bilans qui ne sont pas équilibrés :

- Lorsque Q/P est supérieur à 1, la lame d'eau écoulée à l'exutoire du bassin versant est supérieure à la lame d'eau précipitée. Cette situation suggère que des apports souterrains existent; c'est le cas des bassins versants pour lesquels une influence karstique est connue.
- Lorsque Q/P est inférieur à la limite  $1 - ETP/P$ , la lame d'eau écoulée est inférieure à la différence entre la lame d'eau précipitée et l'évapotranspiration potentielle. Comme l'évapotranspiration potentielle correspond, par définition, à la lame d'eau maximale qui peut être évaporée, cette situation suggère que des pertes souterraines existent.

La Figure 2.2 montre que les bilans hydrologiques des 202 bassins versants sélectionnés sont variés et représentatifs de la diversité rencontrée sur le territoire métropolitain. Seuls quelques bassins sont clairement affectés par des apports ou des pertes souterraines. C'est le cas de La Laine à Soulaines-Dhuys (H1333010), et dans une moindre mesure de la Cuisance à Mesnay (U2655030), pour lesquelles une alimentation karstique est avérée (Le Moine, 2008), tandis que le Rival à Beaufort (V3424310) présente un bilan marqué par d'importantes pertes par infiltration au niveau de la plaine de la Bièvre.

Un résumé de plusieurs caractéristiques hydro-climatiques et morphologiques des bassins constituant l'échantillon final est présenté sur la Figure 2.3, et leur répartition géographique est illustrée sur la Figure 2.4. La densité de drainage a été calculée selon l'approche proposée dans la thèse de Le Moine (2008).

Quelques grandes tendances rencontrées sur le territoire métropolitain se dégagent nettement :

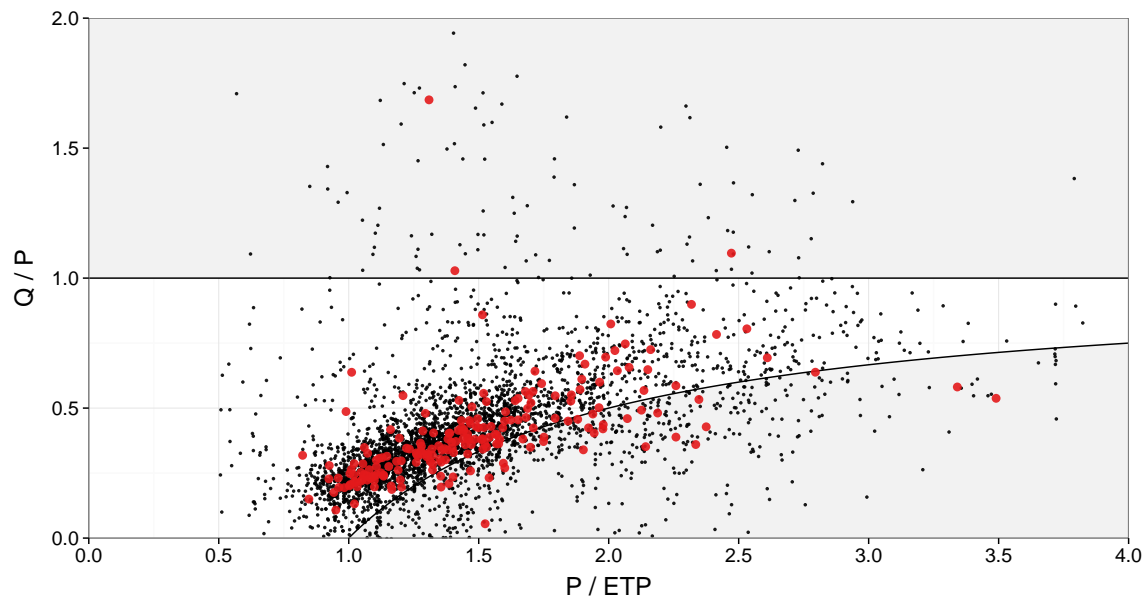


FIGURE 2.2 – Représentation du bilan hydrologique des 202 bassins versants sélectionnés, en rouge, et des 3701 bassins versants de la base de données, en noir.

- Les bassins qui reçoivent le plus de précipitations sont situés sur la façade Atlantique, en Alsace et à l’amont du bassin du Rhône, ainsi que dans la région cévenole.
- La carte des écoulements présente une forte similarité avec la carte des précipitations ; c’est également le cas pour les indices qui caractérisent l’intensité des pluies et l’intensité des crues.
- Un gradient nord-sud caractérise l’évapotranspiration potentielle.
- Les bassins aux plus forts débits de base sont principalement situés au Nord de la France, en Bretagne, en Alsace, et dans le bassin sédimentaire parisien.
- Les bassins les plus élevés sont ceux situés à proximité des principaux reliefs montagneux, des Vosges aux Pyrénées en passant par le Massif Central, sans oublier les Alpes de Haute-Provence.

L’échantillon de bassins versants sélectionnés est varié et représentatif des conditions hydro-météorologiques rencontrées en France métropolitaine. Il nous semble donc particulièrement bien adapté aux objectifs de robustesse et de généralité que nous nous fixons dans cette thèse.

## 2.3 Modèles pluie-débit

Différents modèles pluie-débit ont été utilisés au cours de cette thèse. Les modèles développés au sein de l’équipe d’hydrologie d’Antony ont été privilégiés au regard de l’expertise

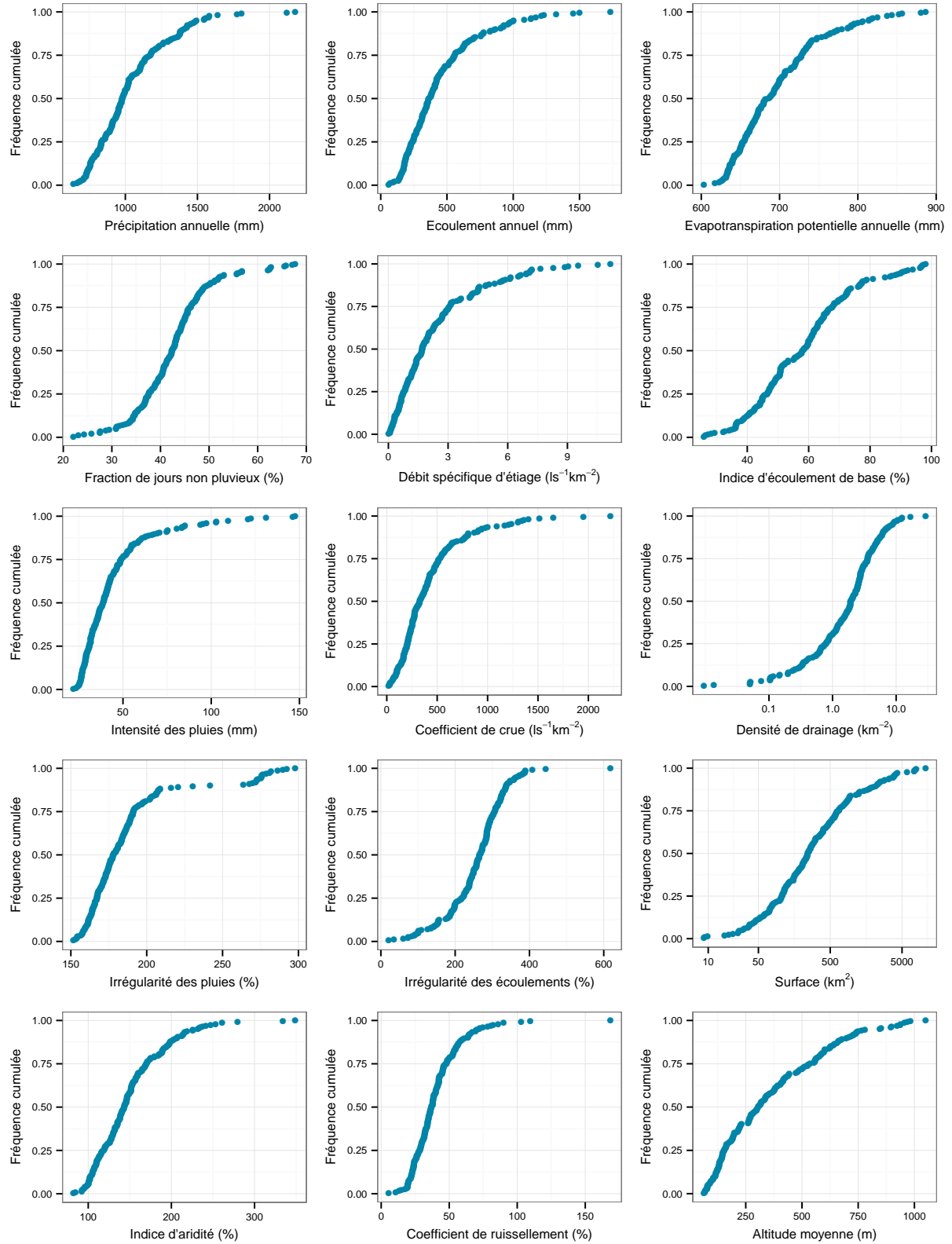


FIGURE 2.3 – Distribution de quelques caractéristiques des 202 bassins versants sélectionnés.



FIGURE 2.4 – Répartition géographique de quelques caractéristiques des 202 bassins versants sélectionnés.



acquise lors de recherches antérieures. Avant de présenter les modèles GR4J et GRP, nous rappelons brièvement ce qui distingue le contexte de simulation de celui de prévision.

### 2.3.1 Simulation et prévision

Les différences entre le fonctionnement d'un modèle hydrologique en simulation et un fonctionnement en prévision sont illustrées sur la Figure 2.5.

Dans un contexte de simulation, les états du modèle évoluent au cours du temps de manière continue, en fonction des forçages disponibles à l'instant  $t$ . En prévision, les derniers débits observés avant l'instant de prévision  $t_0$  peuvent être utilisés par une méthode d'assimilation de données pour ajuster certains états du modèle. A partir de ce point de départ qui définit les conditions initiales, le modèle utilise les forçages disponibles sur une plage de prévision pour produire des débits jusqu'à l'échéance de prévision.

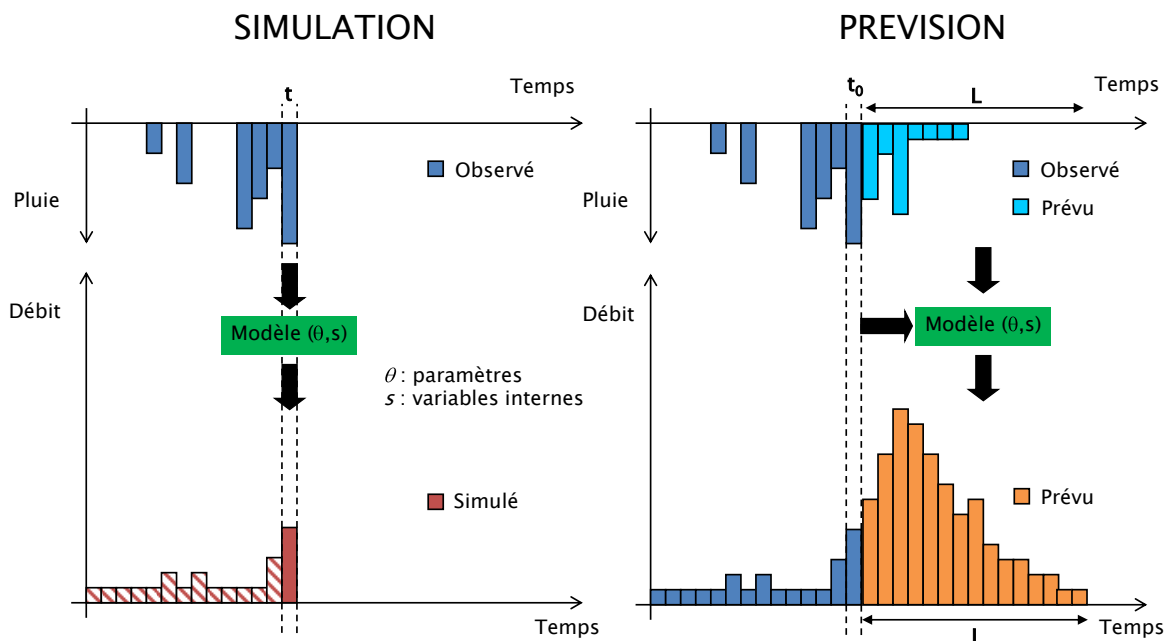


FIGURE 2.5 – Illustration des contextes de simulation et de prévision.

## 2.3.2 Modèle GR4J

### Présentation

Le modèle GR4J (modèle du Génie Rural à 4 paramètres Journalier) est utilisé pour simuler les débits à l'exutoire des bassins versants. Ce modèle pluie-débit global à réservoirs est issu des travaux initiés depuis plus d'une vingtaine d'années à Irstea (anciennement Cemagref). Il a été développé selon une approche empirique et comparative, dans le but d'obtenir un modèle performant, robuste et parcimonieux, pour représenter la transformation pluie-débit. La formulation mathématique de ce modèle qui ne comporte que quatre paramètres à caler (voir Table 2.2) est décrite en détails par Perrin *et al.* (2003) et présentée dans l'annexe A.2.

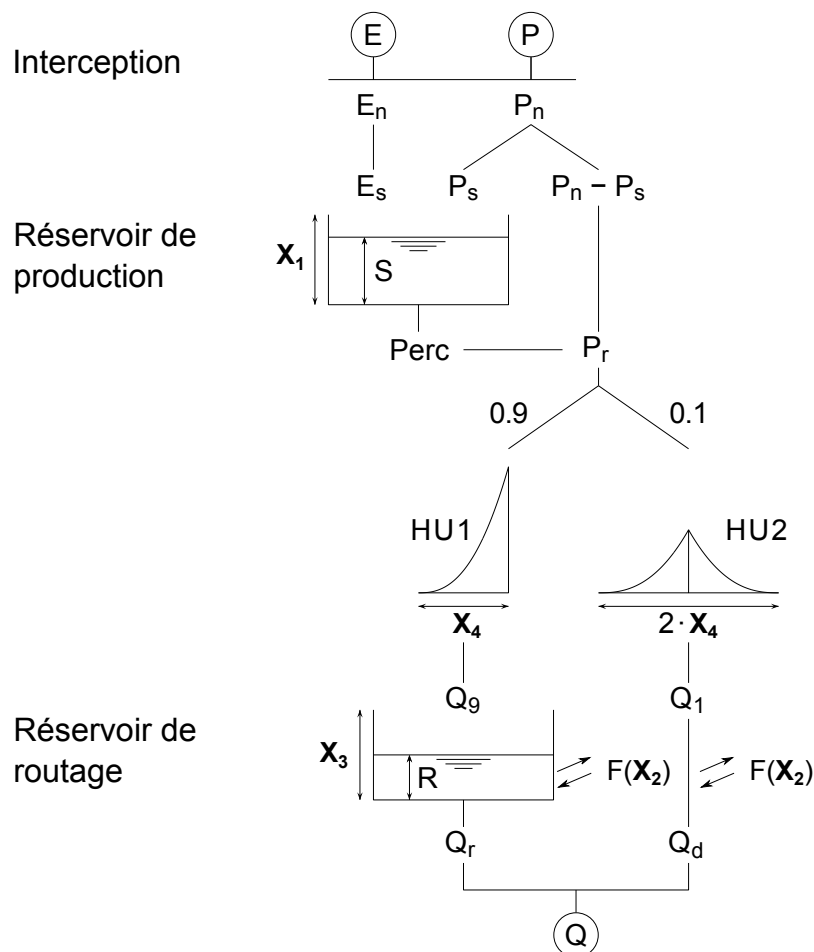


FIGURE 2.6 – Schéma de la structure du modèle GR4J (voir l'annexe A.2 pour la signification des variables intermédiaires).

Un schéma de la structure du modèle GR4J est présenté sur la Figure 2.6. Les réservoirs de production et de routage, associés à deux hydrogrammes unitaires et une fonction d'échange

Paramètre	Signification	Unité
X1	Capacité du réservoir de production	mm
X2	Coefficient d'échanges souterrains	mm
X3	Capacité du réservoir de routage	mm
X4	Temps de base de l'hydrogramme unitaire	j

TABLE 2.2 – Paramètres du modèle GR4J.

avec l'extérieur, permettent de représenter de manière globale la transformation pluie-débit à l'échelle des bassins versants.

### Initialisation et méthode d'optimisation

L'initialisation des états des deux réservoirs du modèle GR4J suit la stratégie présentée par [Le Moine \(2008\)](#), et onze années de mise en route sont utilisées avant la période de simulation.

La procédure d'optimisation utilisée pour estimer les paramètres du modèle GR4J comprend deux étapes. Dans un premier temps, différents jeux de paramètres identifiés lors des travaux de [Perrin et al. \(2008\)](#) sont testés, et le meilleur jeu de paramètres est retenu. Cette étape de pré-filtrage fournit un point de départ à un algorithme de recherche locale, la méthode « pas-à-pas », qui conduit à l'identification d'un jeu de paramètres optimal. Cette procédure s'est révélée efficace et robuste pour différents modèles pluie-débit conceptuels, dont le modèle GR4J ([Mathevet, 2005](#); [Perrin et al., 2008](#)).

L'algorithme de calage utilise des valeurs transformées des paramètres de manière à faciliter la convergence de la méthode de recherche locale. Les correspondances entre les valeurs réelles  $X$  et transformées  $XT$  sont les suivantes :

$$\left\{ \begin{array}{l} XT1 = \log(X1)/1.1 \\ XT2 = \operatorname{argsh}(X2) \\ XT3 = \log(X3) \\ XT4 = 9.99 + 19.98 \cdot (X4 - 20)/19.5 \end{array} \right.$$

### 2.3.3 Modèle GRP

#### Présentation

Le modèle GRP est issu des travaux de thèse de [Tangara \(2005\)](#) et [Berthet \(2010\)](#). Il a été conçu pour la prévision des débits à court terme, et présente la particularité d'intégrer des stratégies d'assimilation des débits passés pour mieux prévoir les crues à venir. Il fonctionne au pas de temps horaire de manière à représenter de manière suffisamment fine la dynamique des débits lors des événements de crues. Sa formulation mathématique est présentée dans l'annexe A.3.

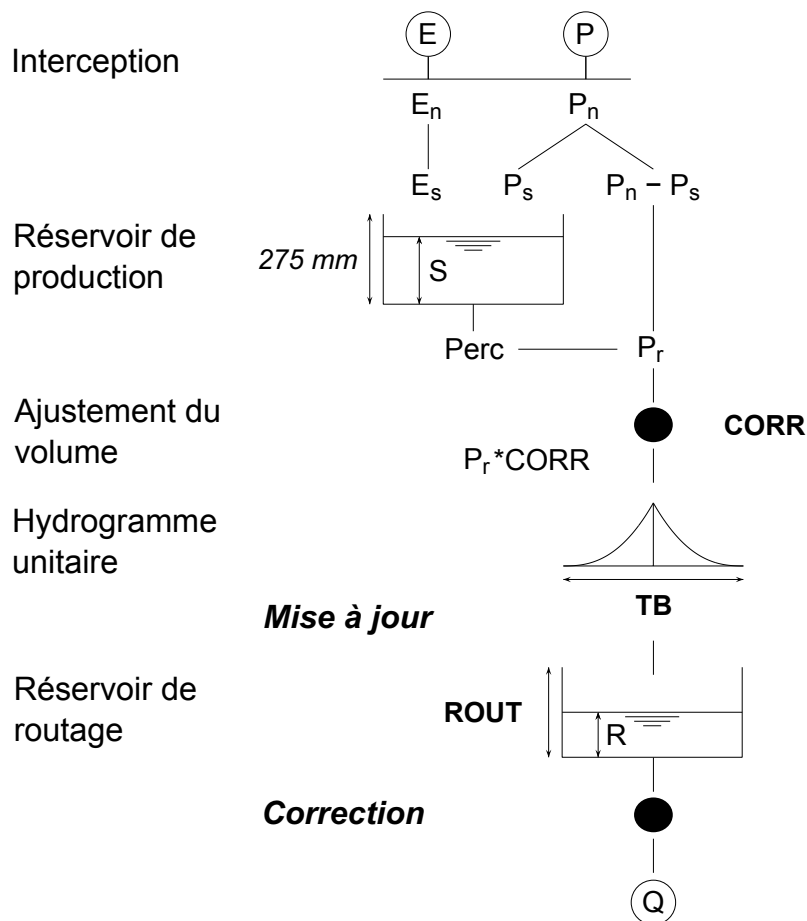


FIGURE 2.7 – Schéma de la structure du modèle GRP (voir l'annexe A.3 pour la signification des variables intermédiaires).

Un schéma de la structure du modèle GRP est présenté sur la Figure 2.7. Ce modèle correspond à une version simplifiée de la structure du modèle GR4J, avec notamment la suppression de la branche directe d'écoulement et le remplacement de la fonction d'échange par un co-

Paramètre	Signification	Unité
CORR	Coefficient d'ajustement de la pluie efficace	-
TB	Temps de base de l'hydrogramme unitaire	h
ROUT	Capacité du réservoir de routage	mm

TABLE 2.3 – Paramètres du modèle GRP.

efficient multiplicatif situé en amont de l'hydrogramme unitaire. Ces modifications ont été comparées à de nombreuses autres alternatives dans le cadre de la thèse de [Berthet \(2010\)](#). Le modèle ne comporte que trois paramètres libres (voir Table 2.3), d'autres paramètres ayant été fixés, comme par exemple la capacité du réservoir de production.

Deux procédures d'assimilation de l'information de débit observé sont utilisées dans le fonctionnement du modèle GRP pour la prévision en temps réel :

- La première est un ajustement de l'état du réservoir de routage. À l'instant de prévision, le niveau dans le réservoir de routage est calculé en fonction du dernier débit observé disponible. Une inversion directe est possible car le débit simulé est une fonction univoque de l'état du réservoir de routage.
- La seconde procédure d'assimilation utilisée est une correction multiplicative du débit prévu. Cette correction dépend de la dernière erreur constatée lors d'une prévision à l'échéance d'une heure.

Les travaux de [Berthet \(2010\)](#) ont montré que ces deux procédures permettent d'améliorer significativement les performances du modèle par rapport à un mode de fonctionnement où seule l'information pluviométrique est disponible. Les mises-à-jour sont parties intégrantes de la structure du modèle qui utilise donc les précipitations et le dernier débit observé comme entrées.

### **Initialisation et méthode d'optimisation**

Au moment de l'initialisation, le taux de remplissage du réservoir de production est fixé à 80%, et celui du réservoir de routage à 50%. Une année de mise en route est considérée suffisante avant le début de la période de prévision.

Les trois paramètres du modèle GRP sont estimés au moyen d'une procédure d'optimisation. Le même algorithme de calage que celui utilisé pour le modèle GR4J est appliqué, mais dans un contexte de prévision, c'est-à-dire que le modèle de prévision GRP fonctionne avec ses mises-à-jour. L'estimation des paramètres est donc réalisée pour un horizon de prévision.

## 2.4 Synthèse

Pour répondre aux objectifs de recherche fixés dans cette thèse, nous avons mis en place un cadre de comparaison qui repose sur un large échantillon de bassins versants pour lesquels des données au pas de temps journalier et horaire sont disponibles sur une période d'au moins dix ans.

Les modèles hydrologiques considérés sont les modèles développés au sein de l'équipe d'hydrologie d'Antony en raison de l'expertise acquise lors de recherches antérieures. Un modèle de simulation, GR4J, et un modèle de prévision, GRP, ont été retenus pour nos travaux.



# 3

## MÉTHODES D'ÉVALUATION DE LA QUALITÉ DES INCERTITUDES PRÉDICTIVES





## 3.1 Introduction

Ce chapitre porte sur l'évaluation des méthodes de quantification de l'incertitude prédictive. Il s'agit d'un point essentiel pour cette thèse qui n'a pourtant reçu qu'une attention limitée en modélisation hydrologique, comme l'ont par exemple souligné [Laio et Tamea \(2007\)](#) et [Franz et Hogue \(2011\)](#). En effet, en dépit du nombre de méthodes proposées pour évaluer l'incertitude en modélisation hydrologique, on observe un manque d'intérêt réel pour les méthodes destinées à évaluer la pertinence des bornes d'incertitude. Des articles ou des ouvrages de référence ne traitent pas de cette question, et il est finalement difficile de savoir ce qui définit le succès d'une méthode de quantification des incertitudes. Nous avons opté pour une évaluation avant tout pragmatique, c'est-à-dire quantitative et adaptée aux multiples expériences que nous avons réalisées.

Après quelques rappels sur les probabilités, nous présentons les différents scores numériques utilisés au cours de cette thèse.

Dans ce chapitre, nous utiliserons le terme générique de prédiction pour désigner les résultats obtenus dans des contextes de simulation ou de prévision.

## 3.2 Quelques rappels sur les probabilités

Dans un cadre probabiliste, l'incertitude sur la valeur d'une variable continue à prédire s'exprime de la manière la plus complète par une fonction de répartition qui caractérise la distribution prédictive. Une illustration du type d'information probabiliste qui peut être communiqué et évalué à partir d'une fonction de répartition est présentée sur la Figure 3.1.

La fonction de répartition  $F_Q(q) = \Pr(Q \leq q)$  de la variable  $Q$ , ici le débit, nous donne accès aux valeurs des quantiles prédictifs. Ainsi, lorsque l'on considère la valeur  $q_\alpha$  du quantile d'ordre  $\alpha$ , nous pouvons exprimer la proposition suivante : *la probabilité que le débit ne dépasse pas la valeur  $q_\alpha$  est égale à  $\alpha$* . Sur l'illustration, deux quantiles d'ordre 0.10 et 0.90 sont mis en évidence en orange, et les flèches indiquent l'information sur la variable prédite que l'on obtient quand on considère ces quantiles.

Les bornes inférieure et supérieure d'un intervalle prédictif centré  $\beta$  sont respectivement définies à partir des quantiles d'ordres  $\frac{1-\beta}{2}$  et  $\frac{1+\beta}{2}$ . L'assertion liée à la communication d'un intervalle prédictif est alors la suivante : *la probabilité que le débit se situe entre les bornes de l'intervalle prédictif est égale à  $\beta$* . Sur l'illustration, les deux quantiles d'ordre 0.10 et 0.90

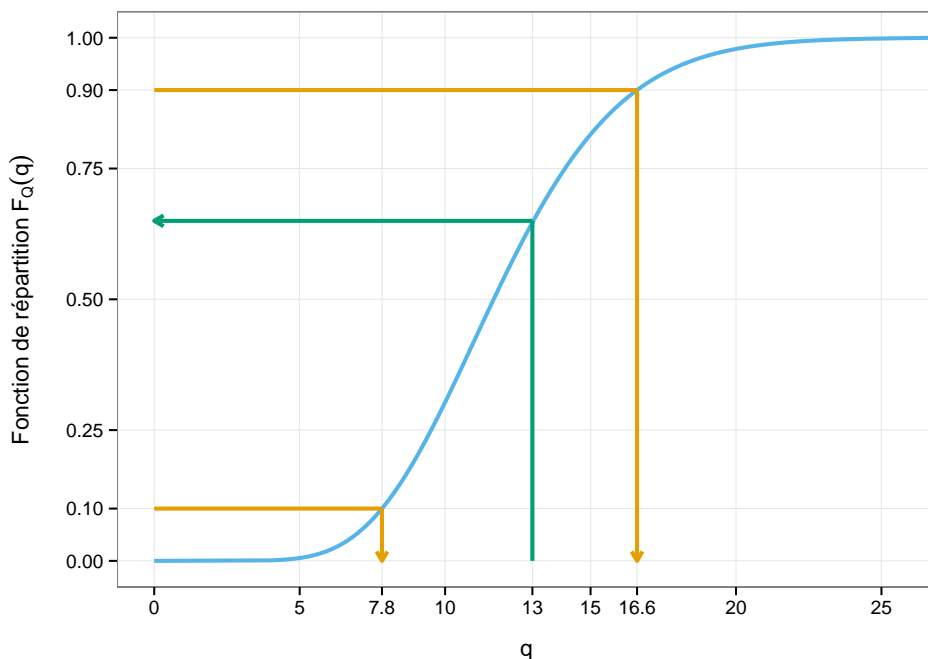


FIGURE 3.1 – Illustration d'une fonction de répartition  $F_Q(q)$ , et du lien entre fréquence au non-dépassement et quantile de débit (flèches orange) ou réciproquement (flèche verte).

permettent de définir les bornes de l'intervalle prédictif à 80% indiquées par les flèches orange.

Par ailleurs, il peut être intéressant, et en particulier dans un contexte de prévision de crues, de communiquer une probabilité de non-dépassement d'un seuil de débit  $q_s$ , par exemple un seuil de vigilance dont la valeur est fixe au cours du temps. Pour accéder à cette information à l'aide de la fonction de répartition, il suffit d'inverser la démarche précédente, et chercher l'ordre du quantile qui correspond à la valeur  $q_s$ . Sur la Figure 3.1, les lignes vertes illustrent ce cheminement et la flèche indique la probabilité de non-dépassement pour une valeur  $q$  donnée. Lorsque plusieurs seuils sont considérés, plusieurs catégories ordonnées de débits sont créées, et la répartition des probabilités attribuées aux différentes catégories s'obtient en calculant la différence des ordres successifs des quantiles qui correspondent aux valeurs des seuils. La somme des probabilités d'appartenance à chaque catégorie doit être égale à 1.

Ces rappels nous semblent nécessaires pour mieux comprendre les critères d'évaluation, car ils utilisent tous l'une ou l'autre de ces informations pour mesurer la qualité d'une série de prédictions probabilistes.

### 3.3 Critères et outils graphiques d'évaluation

Les critères d'évaluation des prédictions probabilistes permettent de mesurer différentes qualités d'une méthode de quantification des incertitudes. Différents critères existent selon le type de prédiction considéré, et chaque critère pénalise à sa manière l'écart entre les informations fournies sous forme de probabilités, et les observations.

Aucun critère numérique ne permet à lui seul de quantifier de manière univoque ce qu'est une « bonne » série de prédictions probabilistes (de même qu'il n'existe pas de critère unique pour évaluer la qualité de prédictions déterministes, voir [Bennett et al., 2013](#) et [Crochemore et al., 2014](#)). Il est donc nécessaire de s'interroger sur ce que l'on attend des informations fournies par des prédictions probabilistes, afin de choisir des critères qui reflètent certaines propriétés attendues. Des outils graphiques d'évaluation peuvent également être utilisés pour mettre en évidence les forces et les faiblesses d'une approche prédictive.

Nous présentons ici plusieurs critères selon les différentes formes de prédiction probabiliste couramment utilisées :

- une série chronologique d'intervalles ;
- une série chronologique de quantiles ;
- une série chronologique de distributions prédictives ;
- une série chronologique de probabilités de dépassement de seuil.

Une présentation plus exhaustive et détaillée des critères d'évaluation est donnée dans les ouvrages de [Wilks \(2011\)](#) et [Jolliffe et Stephenson \(2012\)](#).

#### 3.3.1 Évaluer une série d'intervalles prédictifs

Une stratégie d'évaluation des intervalles prédictifs généralement adoptée en modélisation hydrologique consiste à s'intéresser à deux qualités, la fiabilité et la finesse ([Montanari et Brath, 2004](#); [McIntyre et al., 2005](#); [Yang et al., 2008](#); [Solomatine et Shrestha, 2009](#); [Vrugt et al., 2009](#); [Jin et al., 2010](#)).

##### Fiabilité des intervalles prédictifs

La fiabilité correspond à la validité statistique des intervalles prédictifs associés à un certain niveau de crédibilité. Par exemple, l'estimation des intervalles proposés à un niveau de crédibilité de 90% peut être considérée comme fiable si la fréquence à laquelle les observations se

situent à l'intérieur des intervalles prédictifs est égale à 90%. Le taux de couverture global indique ainsi de manière simple et compréhensible si les intervalles prédictifs sous-estiment ou surestiment les incertitudes. Le taux de couverture peut être calculé pour différents niveaux de crédibilité, et pour différentes situations hydrologiques (par exemple sur les bas débits, les hauts débits ou sur des événements de crues).

### Finesse des intervalles prédictifs

La deuxième caractéristique importante d'un intervalle prédictif est sa finesse. Plus un intervalle est fin, plus l'incertitude sur la variable prédite est faible, et donc plus on a de chance que la valeur réelle soit proche de la valeur prédite. Plusieurs critères évaluant la largeur moyenne des intervalles prédictifs sont proposés dans la littérature, selon la normalisation de la largeur de l'intervalle choisie (voir par exemple [Jin et al., 2010](#)).

### Performance globale des intervalles prédictifs

La fiabilité apparaît comme une condition nécessaire, mais non suffisante pour mesurer la qualité des intervalles prédictifs. Pour être utiles, les intervalles prédictifs doivent à la fois être valides et cohérents par rapport aux données issues de l'observation, mais également suffisamment informatifs pour aider la prise de décisions. Autrement dit, les intervalles prédictifs doivent être fiables et aussi fins que possible.

Pour évaluer la pertinence d'une évaluation des incertitudes, il est tout à fait envisageable d'évaluer séparément fiabilité et finesse. Dans la mesure où ces deux propriétés ne sont pas indépendantes et peuvent se compenser mutuellement, il est aussi possible d'utiliser différents scores combinés.

Le score de Winkler ([Gneiting et Raftery, 2007](#)) permet de tenir compte de ces deux caractéristiques. Pour un intervalle prédictif à  $\beta\%$ , défini par l'intervalle  $[l, u]$ , et un débit observé  $q_{obs}$ , ce score s'écrit :

$$sw_{\beta}(l, u, q_{obs}) = (u - l) + \frac{2}{1 - \beta}(l - q_{obs}) \mathbb{1}\{q < l\} + \frac{2}{1 - \beta}(q_{obs} - u) \mathbb{1}\{q_{obs} > u\} \quad (3.1)$$

où  $\mathbb{1}\{x < y\}$  est la fonction indicatrice qui vaut 1 si  $x < y$ , et 0 sinon.

La pénalité attribuée lorsque le débit se situe en dehors des bornes de l'intervalle est proportionnelle à la distance absolue entre le débit observé et une des bornes de l'intervalle, selon

que le débit se situe au-dessus ou en dessous de l'intervalle. De plus, elle augmente avec le niveau de crédibilité  $\beta$ .

Un score global  $SW_\beta$ , pour un niveau de crédibilité  $\beta$  donné, s'obtient en calculant la moyenne des scores  $sw_\beta$  obtenues pour chaque prédiction.

### 3.3.2 Évaluer une série de quantiles

#### Fiabilité des quantiles

La fiabilité d'une série chronologique de quantiles d'ordre  $\alpha$  est évaluée assez simplement : on s'attend en effet à ce qu'un pourcentage  $\alpha$  des valeurs observées se situent en dessous des quantiles. S'intéresser à la fiabilité des quantiles utilisés pour définir un intervalle prédictif permet de vérifier si les observations ont tendance à se situer au-dessus ou en dessous des bornes de l'intervalle prédictif.

#### Performance globale des quantiles

On ne peut pas parler de finesse pour les quantiles. Cependant, la fiabilité ne suffit pas à caractériser la qualité globale des quantiles, puisqu'elle ne considère que la position relative (au-dessus ou en dessous) d'un quantile par rapport à l'observation. De manière intuitive, on s'attend à ce que la distance entre un quantile et l'observation ait une importance. Un critère, le score de quantile, compare la valeur  $q_\alpha$  du quantile d'ordre  $\alpha$  à un débit observé  $q_{obs}$  de la manière suivante (Gneiting et Raftery, 2007) :

$$s_\alpha(q_\alpha, q_{obs}) = (1 - \alpha)(q_\alpha - q_{obs}) \mathbb{1}\{q_{obs} < q_\alpha\} + \alpha(q_{obs} - q_\alpha) \mathbb{1}\{q_{obs} \geq q_\alpha\} \quad (3.2)$$

Ce score correspond à une fonction de coût asymétrique et linéaire par morceaux. Par exemple, pour un quantile d'ordre 0.90, le score obtenu lorsque l'observation se situe au-dessus du quantile sera égal à la distance absolue entre l'observation et la valeur du quantile, multipliée par l'ordre  $\alpha$  du quantile, tandis que dans le cas contraire, le facteur multiplicatif appliqué à la distance absolue sera égal à  $(1 - \alpha)$ . Un score global  $S_\alpha$  s'obtient en calculant la moyenne des scores obtenus pour chaque prédiction.

Ce critère est particulièrement intéressant en raison de son lien avec le score CRPS (*Continuous Rank Probability Score*). Il est en effet possible de calculer le CRPS à partir d'un nombre fini de quantiles, comme l'ont par exemple montré Bontron (2004) et Laio et Tamea (2007).

Nous pouvons également remarquer que ce score appliqué à une série de quantiles d'ordre 0.5 est égal à la moitié de l'erreur absolue moyenne entre la médiane de la distribution prédictive et l'observation.

### 3.3.3 Évaluer une série de distributions prédictives

#### Fiabilité des distributions prédictives

L'évaluation de la fiabilité d'une série de distributions prédictives est liée à l'évaluation de la fiabilité d'une série de quantiles. En effet, si la série de distributions est correctement spécifiée d'un point de vue statistique, la distribution de l'ordre des quantiles auxquels correspondent les observations doit suivre une loi uniforme.

Pour une prédiction, la valeur considérée est donc l'ordre du quantile qui est pointé par la flèche noire sur la Figure 3.2, notée  $p^z$ . Le diagramme de PIT, *Probability Integral Transform*, permet d'évaluer visuellement la fiabilité d'une série de distributions prédictives, à l'aide d'un diagramme quantile-quantile, ce qui revient à tracer la fonction de répartition des valeurs  $p_n^z$ . Ce type de diagramme présenté sur la Figure 3.3 est fréquemment utilisé en modélisation hydrologique.

Il devient cependant rapidement difficile de visualiser chacune des courbes obtenues pour différents bassins versants ou différentes méthodes prédictives. Ceci a conduit [Renard et al. \(2010\)](#) à proposer un score qui synthétise la fiabilité d'une série de distributions prédictives en quantifiant les déviations par rapport au cas idéal. Ce score  $FIAB_\gamma$  correspond à l'aire entre la diagonale et une courbe PIT. Il est calculé de la manière suivante :

$$FIAB_\gamma = \frac{1}{N} \sum_{n=1}^N \left| p_n^z - p_n^{th} \right| \quad (3.3)$$

où  $N$  est le nombre de prédictions et  $p_n^{th}$  est la valeur de la fonction de répartition des valeurs PIT qui correspond à la valeur  $p_n^z$ .

Le chapitre 4 de cette thèse présente une note technique consacrée à l'interprétation du diagramme de PIT. Nous avons en effet constaté que certaines méthodes de quantification de l'incertitude prédictive obtenaient des scores de fiabilité  $FIAB_\gamma$  plus élevés que ceux obtenus avec d'autres méthodes *a priori* plus satisfaisantes. Il s'est avéré que dans certains cas, des compensations entre différents défauts pouvaient conduire à des diagrammes de PIT d'apparence satisfaisante : c'est-à-dire à des diagrammes qui soient satisfaisants pour de mauvaises raisons.

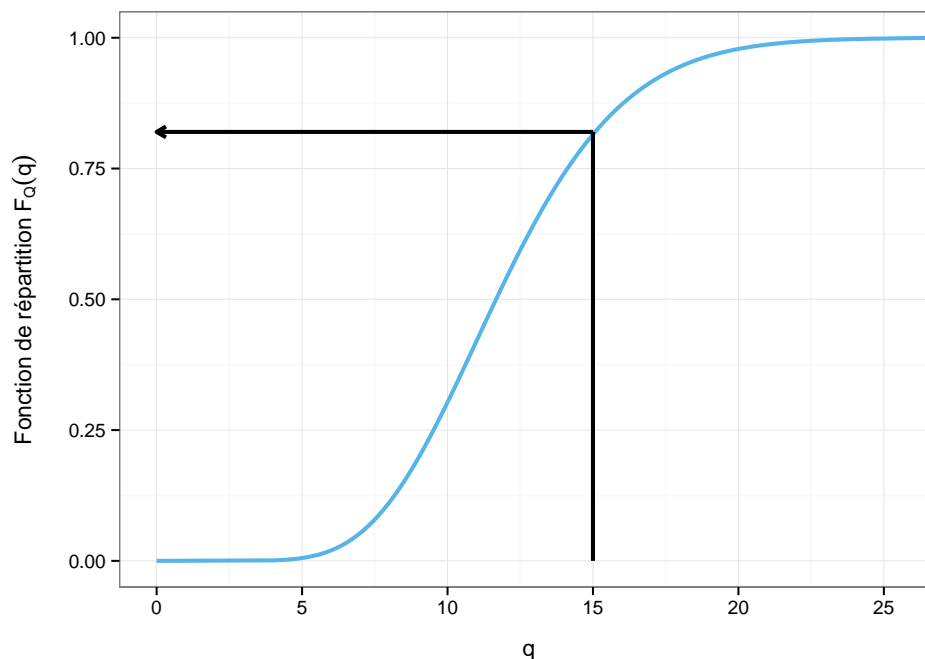


FIGURE 3.2 – Illustration de la valeur  $p^z$  considérée dans le diagramme de PIT.

Ces travaux nous ont amené à utiliser un indice de fiabilité calculé comme la moyenne des scores de fiabilité  $FIAB_\gamma$  obtenus de manière séparée sur les hauts et les bas débits. Cela permet de pénaliser les cas où des compensations ont lieu.

### Finesse des distributions prédictives

La finesse est une caractéristique importante des distributions prédictives. L'incertitude sur une variable prédite est en effet d'autant plus faible que la distribution prédictive est concentrée.

Il existe plusieurs manières de caractériser la finesse – l'étalement, la précision – d'une distribution prédictive. La variance est une mesure classique de la dispersion d'une distribution, et peut à ce titre être utilisée pour définir un indice de précision comme l'a proposé [Renard et al. \(2010\)](#). Cependant d'autres mesures peuvent également être considérées, comme l'écart interquartile ou l'entropie.

Nous avons choisi d'utiliser un critère de finesse proposé dans la thèse de [Bontron \(2004\)](#). Il s'agit de calculer le CRPS, décrit à l'équation 3.4 en remplaçant l'observation par la médiane de la distribution prédictive. Le score obtenu ne dépend que de l'étalement de la distribution.



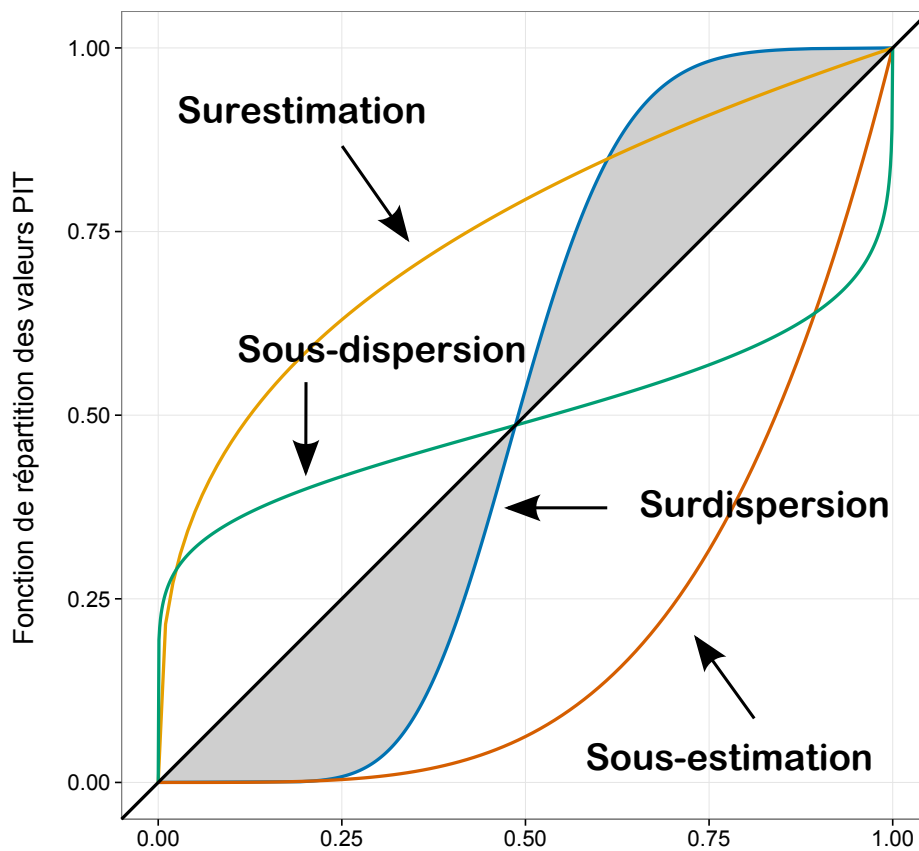


FIGURE 3.3 – Illustration du diagramme de PIT, où chaque courbe correspond à une série de prédictions. La courbe bleue correspond à des prédictions probabilistes surdispersées, puisque l'observation a tendance à se situer trop souvent dans la partie centrale des distributions prédictives. Au contraire, la courbe verte caractérise des prédictions sous-dispersives, qui sous-estiment les incertitudes, les observations se situant alors trop fréquemment au niveau des quantiles supérieurs et inférieurs des distributions. La courbe orange suggère une surestimation systématique des débits observés, tandis que la courbe jaune suggère au contraire une sous-estimation systématique.

### Performance globale des distributions prédictives

On peut considérer qu'un des principaux objectifs d'une quantification des incertitudes est de minimiser l'incertitude prédictive tout en s'assurant de la fiabilité de la méthode prédictive (Gneiting et Raftery, 2007; Renard *et al.*, 2010). Il s'agit donc de proposer des prédictions les plus informatives possible, à condition que ces informations probabilistes soient « justes ».

Le score le plus utilisé pour évaluer la performance globale des prédictions probabilistes est le CRPS moyen (*Continuous Rank Probability Score*). Le CRPS moyen s'obtient à partir des scores individuels CRPS, calculés par l'intégrale suivante :

$$\text{CRPS}(F_X, q_{obs}) = \int_{-\infty}^{\infty} (F_X(x) - \mathbb{1}\{q_{obs} \leq x\})^2 dx \quad (3.4)$$

où  $F_X(x)$  est la fonction de répartition de la distribution prédictive de  $x$ , et  $q_{obs}$  est le débit observé.

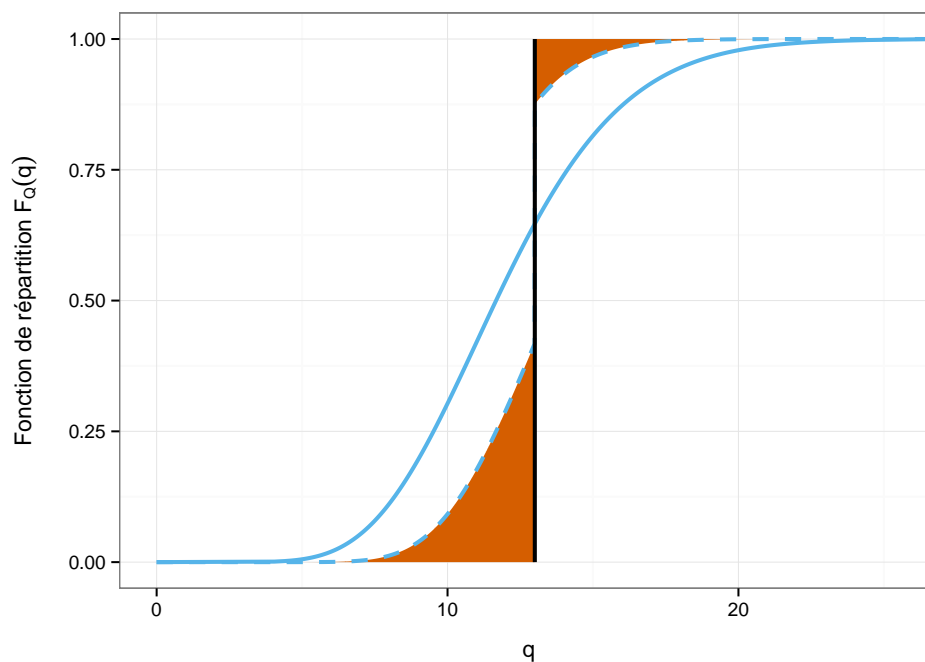


FIGURE 3.4 – Illustration du calcul du CRPS, égal à la surface en orange. La valeur observée est indiquée par la ligne noire verticale ; la distribution prédictive correspond à la courbe bleue.

La Figure 3.4 illustre la manière dont la distribution prédictive est confrontée à une valeur observée. Considérons la fonction de répartition associée à une distribution prédictive et la valeur du débit observé. Le calcul de l'intégrale présentée ci-dessus correspond à la surface orange. Le CRPS quantifie l'écart, la distance, entre la distribution prédictive et une fonction

échelon centrée sur la valeur observée. Une distribution prédictive peu dispersée et centrée, à proximité de la valeur observée, sera moins pénalisée qu'une distribution fortement dispersée, ou qu'une distribution éloignée de la valeur observée. Le CRPS permet de prendre en compte à la fois la position de l'observation dans la distribution prédictive et la finesse de la distribution.

Comme nous l'avons signalé plus haut, le CRPS peut également être calculé à partir du score de quantile  $s_\alpha$  :

$$\text{CRPS}(F, q_{obs}) = 2 \int_0^1 s_\alpha(q_\alpha, q_{obs}) d\alpha \quad (3.5)$$

où par définition  $q_\alpha = F_Q^{-1}(\alpha)$ .

Nous avons utilisé cette équivalence pour obtenir une approximation du CRPS, calculé à partir d'un nombre suffisamment important de quantiles. Pour l'ensemble des travaux présentés dans cette thèse, 99 quantiles ont été utilisés pour caractériser les distributions prédictives.

### 3.3.4 Évaluer une série de probabilités de dépassement de seuil

#### Score de Brier et RPS

Le score de Brier, ou score de probabilité, quantifie la capacité à prédire l'occurrence d'un événement binaire, par exemple un non-dépassement de seuil. Pour chaque prédiction, la probabilité  $p_n$  attribuée à l'événement qui ne s'est pas réalisé, et qui représente donc l'erreur prédictive, est pénalisée de manière quadratique. Le score de Brier, BS, est défini par la moyenne de ces erreurs :

$$\text{BS} = \frac{1}{N} \sum_{n=1}^N p_n^2 \quad (3.6)$$

où N est le nombre de prédictions.

La Figure 3.5 illustre une situation où l'événement, défini par le non-dépassement du seuil, n'est pas réalisé, puisque la valeur observée est supérieure à celle du seuil. Dans ce cas, la probabilité  $p_n$  est égale à  $\alpha$ , l'ordre du quantile correspondant à la valeur du seuil (en vert).

Lorsque plusieurs seuils sont considérés, le RPS (*Rank Probability Score*) peut être utilisé pour évaluer des prédictions probabilistes par catégories. Il s'obtient simplement en calculant la

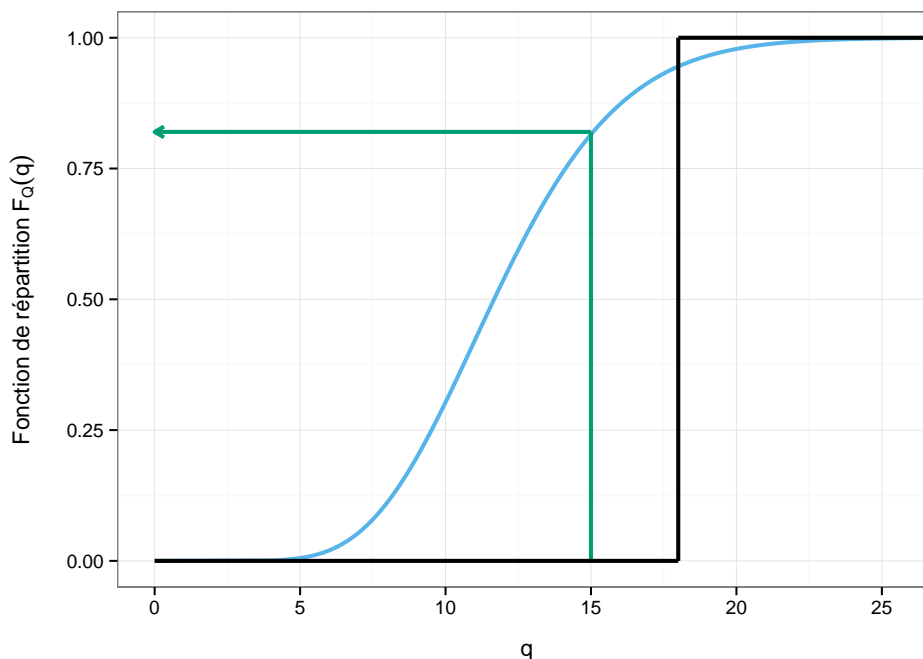


FIGURE 3.5 – Illustration du calcul du score de Brier. La valeur du seuil considérée correspond à l'abscisse de la ligne verticale verte. La probabilité de non-dépassement correspond à l'ordre du quantile pointé par la flèche verte. La valeur observée est indiquée par la ligne noire verticale.

somme des scores de Brier obtenus par chacun des seuils considérés. Au contraire du CRPS, qui s'obtient par intégration des BS, le RPS ne considère qu'une somme des BS. Ceci explique que le BS et le RPS sont adimensionnels, alors que le CRPS a la même unité que la variable prédite.

### 3.4 Choix d'une référence : la climatologie

Les scores probabilistes peuvent être présentés sous forme de gain de performance d'une méthode par rapport à une autre. Ces scores de compétence (*skill score*) sont calculés de la manière suivante :

$$SC = 1 - \frac{\text{Score}^A}{\text{Score}^B} \quad (3.7)$$

où  $\text{Score}^A$  et  $\text{Score}^B$  sont les scores obtenus par les méthodes prédictives A et B respectivement. Nous pouvons remarquer que cette formulation est similaire à celle du critère de performance de [Nash et Sutcliffe \(1970\)](#) couramment utilisé en modélisation hydrologique.

Les scores de compétence présentent l'avantage d'être adimensionnels et sont donc par-

ticulièrement bien adaptés aux études sur de nombreux bassins versants. Nous pourrions également utiliser la transformation proposée par [Mathevet \(2005\)](#),  $f(x) = \frac{x}{2-x}$ , qui permet de limiter la gamme de variation des scores de compétence à l'intervalle  $[-1, 1]$ .

Dans l'ensemble des travaux présentés dans cette thèse, nous avons utilisé la « climatologie » des débits observés comme méthode prédictive de référence. Cette approche possède les caractéristiques suivantes : elle est inconditionnelle, dans le sens où la même prédiction est donnée pour tous les pas de temps ; et fiable, par construction, puisqu'elle est construite à partir des observations. Pour garantir la fiabilité sur chaque période d'évaluation, une climatologie différente est construite pour chaque période.

Les scores liés à la notion de fiabilité ne sont pas directement concernés par cette normalisation. On pourra néanmoins remarquer que la climatologie présente, par construction, un diagramme de PIT idéal.

Type de prédiction	Critère	Qualité évaluée
Intervalle	Taux de couverture	Fiabilité
Intervalle	Largeur moyenne relative	Finesse
Intervalle	Score de Winkler $SW_{\beta}$	Performance globale
Quantile	Pourcentage d'observations en dessous des quantiles	Fiabilité
Quantile	Score de quantile $S_{\alpha}$	Performance globale
Distribution	$FIAB_{\gamma}$	Fiabilité
Distribution	Largeur moyenne relative	Finesse
Distribution	CRPS	Performance globale
Probabilité(s) de dépassement	BS et RPS	Capacité à prévoir un (ou des) dépassement(s) de seuil(s)

TABLE 3.1 – Critères d'évaluation utilisés en fonction du type de prédiction probabiliste.

### 3.5 Synthèse

Nous avons présenté dans ce chapitre plusieurs critères numériques permettant d'évaluer différentes caractéristiques des prédictions probabilistes. La Table 3.1 résume les différents critères utilisés au cours de cette thèse pour évaluer divers types de prédiction, avec la qualité évaluée par ces critères.

# 4

## **FLOW-GROUP CONDITIONING OF THE PIT DIAGRAM: HOW TO AVOID BEING RIGHT FOR THE WRONG REASONS**



Ce chapitre correspond à un article soumis à la revue *Water Resources Research*.

### **Flow-group conditioning of the PIT diagram: how to avoid being right for the wrong reasons**

F. Bourgin<sup>1</sup>, M.H. Ramos<sup>1</sup>, V. Andréassian<sup>1</sup>, C. Perrin<sup>1</sup>, B. Renard<sup>2</sup>

(1) Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, F-92761 Antony Cedex, France

(2) Irstea, UR HHLY, 5 rue de la Doua, CS70077, F-69626 Villeurbanne Cedex, France

## **Résumé**

La fiabilité est un attribut primordial des prévisions probabilistes. Elle correspond à la cohérence statistique de l'information fournie, et est fréquemment évaluée à l'aide du diagramme de PIT. Dans cette note technique, nous examinons les limites de l'utilisation de cet outil d'évaluation en prévision hydrologique. Nous avons appliqué le diagramme de PIT à des prévisions probabilistes de débits, obtenues sur 202 bassins français par un système de prévision composé d'un modèle hydrologique et d'une méthode de post-traitement statistique. Nos résultats montrent qu'il est nécessaire de séparer différents groupes de débits lors de la construction du diagramme de PIT, afin d'éviter que des compensations entre différents groupes de débits ne conduisent à des conclusions erronées. L'application du diagramme de PIT à différents groupes de débits permet une meilleure évaluation des systèmes de prévision hydrologique, et met en lumière des pistes d'amélioration de ces systèmes de prévision.

## **Abstract**

Reliability is a fundamental attribute when evaluating the quality of probabilistic flow predictions and the statistical coherence of the associated uncertainty estimates. It is commonly evaluated using the Probability Integral Transform (PIT) diagram. Here, we investigate a major pitfall of this diagram, when used in the evaluation of streamflow forecasts. An extensive application of the PIT diagram to the probabilistic streamflow forecasts produced by a hydrological model and an uncertainty post-processor is presented for a set of 202 French catchments. Results show the necessity and benefits of conditioning the PIT diagram on at least two specific flow groups of low and high flows (defined using the predicted flows). We show that compensations can occur when all flows are considered together, as usually done in the evaluation of hydrological predictions. The use of conditioning strategies can avoid an



unduly optimistic evaluation of forecast quality and help identifying avenues for improving the probabilistic forecasting system.

## 4.1 Introduction

### 4.1.1 On the specificity of probabilistic forecasts evaluation

The most common way to communicate uncertainty in streamflow predictions for water resources and risk management is through the use of probabilities and prediction intervals (Ramos *et al.*, 2010). While the advantages of probabilistic flow forecasting for decision-making are recognized (Boucher *et al.*, 2012; Ramos *et al.*, 2013), the evaluation of the quality of probabilistic forecasts remains a challenging issue. A specific evaluation framework is required in the probabilistic context and it may depend on how the variable being evaluated is distributed in space and time.

Hydrologists have adopted several verification scores and tools developed in the atmospheric sciences or econometrics (Laio and Tamea, 2007; Franz and Hogue, 2011). Moreover, evaluation strategies in a hydrologic context have been developed to account for the particular space-time variability of flow (e.g., Brown *et al.*, 2010; Zappa *et al.*, 2013). In general, forecast quality evaluation focuses on the agreement between forecasts and observations, on the performance of the forecasts comparatively to a reference system (skill), on the ability of the forecast system to detect critical events, or on the reliability and sharpness of the predictive distributions (Wilks, 2011).

### 4.1.2 Reliability and the PIT diagram

Reliability is related to the consistency of the information conveyed by probabilistic statements. It is often considered that a probabilistic forecast system must produce as sharp as possible predictive distributions while preserving reliability (Gneiting *et al.*, 2007). Assessing reliability is essential to ensure the adequateness of uncertainty quantification in probabilistic forecasting systems to make operational decisions with confidence.

The Probability Integral Transform (PIT) diagram is a graphical tool that has a long tradition in probabilistic forecast verification (see Gneiting *et al.*, 2007, and references therein). Since its introduction in hydrology by Laio and Tamea (2007), the PIT diagram, also known as the predictive QQ plot, has become one of the favorite tools to assess the reliability of predictive

distributions (e.g., [Thyer \*et al.\*, 2009](#); [Engeland \*et al.\*, 2010](#); [Schoups and Vrugt, 2010](#); [Bogner and Pappenberger, 2011](#); [Weerts \*et al.\*, 2011](#); [DeChant and Moradkhani, 2012](#)).

### 4.1.3 Scope of this note

At first glance, the PIT diagram offers a simple and yet powerful way to scrutinize the reliability of predictive distributions. However, its use as a diagnostic tool in hydrological forecasting raises a crucial question: considering the strong variability of catchment behaviors and hydrological signatures, can this diagram adequately inform about the reliability of streamflow forecasts in all situations (i.e., when forecasts are either targeting peak flows or recession periods within hydrological events)? Our goal is to show that there are some pitfalls to be avoided when using the PIT diagram for the evaluation of probabilistic hydrological forecasts and that the use of stratified data can help in revealing hidden biases. This is demonstrated through the application of a hydrological forecasting model combined with an uncertainty post-processor to a large set of catchments.

## 4.2 Data and methods

### 4.2.1 Data Set

A set of 202 unregulated catchments spread over France was used. They represent various hydrological regimes, given the variability in climate, topography and geology in France. Hourly rainfall, potential evapotranspiration (PE) and streamflow data series were available over the 1997–2006 period. PE was estimated using a temperature-based formula ([Oudin \*et al.\*, 2005](#)). Rainfall and temperature data come from a radar-based reanalysis produced by Météo-France ([Tabary \*et al.\*, 2012](#)). Flow data were extracted from the national streamflow HYDRO archive ([www.hydro.eaufrance.fr](http://www.hydro.eaufrance.fr)).

### 4.2.2 Probabilistic hydrological forecasting system

To build a probabilistic forecasting system, we combined a deterministic rainfall-runoff model with an uncertainty post-processor.

The GRP model is a lumped storage-type model designed for flood forecasting and currently used by flood forecasting services in France. It has three free parameters and uses catchment

areal rainfall and PE as inputs. The model also assimilates discharge observations available when issuing a forecast. More details about the model can be found in [Berthet \*et al.\* \(2009\)](#).

The three parameters were calibrated for each catchment by minimizing the root mean square errors (RMSE) during the first five years of available data (1997–2001). The other five years (2002–2006) were used to evaluate the streamflow forecasts with the PIT diagram. Observed rainfalls were used as “perfect” rainfall forecasts over each period. The model was run at the hourly time step and forecasts issued for a 24-hour lead-time were evaluated.

The GRP model is deterministic: a single output is obtained from a single input scenario. One possibility to provide probabilistic forecasts with the model is to use it in combination with a hydrological uncertainty post-processor ([Krzysztofowicz, 1999](#)). In this study, post-processing is performed using a data-based approach: the distribution of relative errors (ratio between observed and predicted flows) obtained during the calibration period is approximated by empirical quantiles. These quantiles are multiplicative coefficients that are then used during the evaluation period to estimate predictive distributions.

### 4.2.3 PIT Diagram and reliability score

The PIT diagram is used to assess the reliability of the probabilistic streamflow forecasts. At each time step, the PIT value is equal to the cumulative distribution  $F_t$  of predicted flows evaluated at the observed flow  $x_t$ , i.e.  $\text{PIT}_t = F_t(x_t)$ . The empirical cumulative distribution of these values over all the time steps of the evaluation period gives the PIT diagram. If the forecasts are reliable, the PIT values follow a uniform distribution on the interval  $[0, 1]$ . Reliability of the probabilistic forecasts implies that the observations should not be preferentially located in specific parts of the predictive distributions, but instead should uniformly span the whole predictive range.

Common shapes of the PIT diagram and their usual interpretation are illustrated in Figure 4.1 (based on [Laio and Tamea, 2007](#)). If the curve systematically lies above the bisector, the observed values are too frequently located in the lower parts of the predictive cumulative distributions, suggesting a systematic bias of over-prediction. Conversely, under-prediction occurs when the observed values are too often located in the upper parts of the predictive distributions. If the PIT values are mostly clustered around the midrange, it means that the probabilistic forecasts are in general too wide and that the observed values do not reach the tails; forecasts are said to be over-dispersive (i.e., predictive uncertainty is over-estimated). Conversely, when the observed values are too often located in the tails, the forecasts are in general too narrow, under-dispersive (i.e., predictive uncertainty is under-estimated).

Renard *et al.* (2010) proposed an index to quantify reliability based on the PIT diagram. The Alpha score is directly related to the area  $A$  between the bisector and the curve (see Figure 4.1) and summarizes the deviation of the PIT diagram from the ideal situation (the bisector). It is defined as  $(1 - 2 \times A)$  and ranges from 0 (poor reliability) to 1 (“perfect” reliability).

## 4.3 Results

### 4.3.1 Evaluation of the PIT diagram over the entire flow range

Figure 4.2 (left) shows the PIT diagrams obtained for the studied catchments during the 5-year evaluation period when all flows are considered. Thin lines represent the results for the catchments, while dotted thick lines represent deviations of  $\pm 0.1$  from the bisector. The 202 catchments were split into four groups (S1 to S4) according to the shape of their PIT diagrams and following the common shapes illustrated in Figure 4.1.

We can see that the forecasting system analyzed here produces, in general, predictive distributions that can be considered fairly reliable: deviations of the curves from the bisector are limited and the PIT diagrams are most often within the bands defined around the bisector. When looking at the different groups, we can see from the shape of the curves that the main deficiencies (although limited) are related to an over-prediction (group S1 with 95 catchments) and an under-prediction (group S2 with 59 catchments) of the forecasting system. The cases of under-dispersive and over-dispersive forecasts (S3 and S4 groups, with 13 and 35 catchments, respectively) occur less often in our data set.

### 4.3.2 Evaluation of PIT diagrams for high and low flows

Two flow groups were defined for each catchment by taking the 30% lowest (low flow group) and the 30% highest (high flow group) predicted flows during the evaluation period. These thresholds were selected in order to obtain two distinct flow subsets, while keeping, at the same time, a large enough number of verification pairs for the statistical evaluation of reliability. Since reliability is an attribute that concerns the conditional distribution of observation given the forecasts, data stratification was done conditionally on the predicted flows rather than the observed flows (see also Mascaro *et al.* (2010)).

Figure 4.2 illustrates the PIT diagrams for each flow group (center and right columns) and catchment group (rows). Several of the catchments where forecasts could be considered

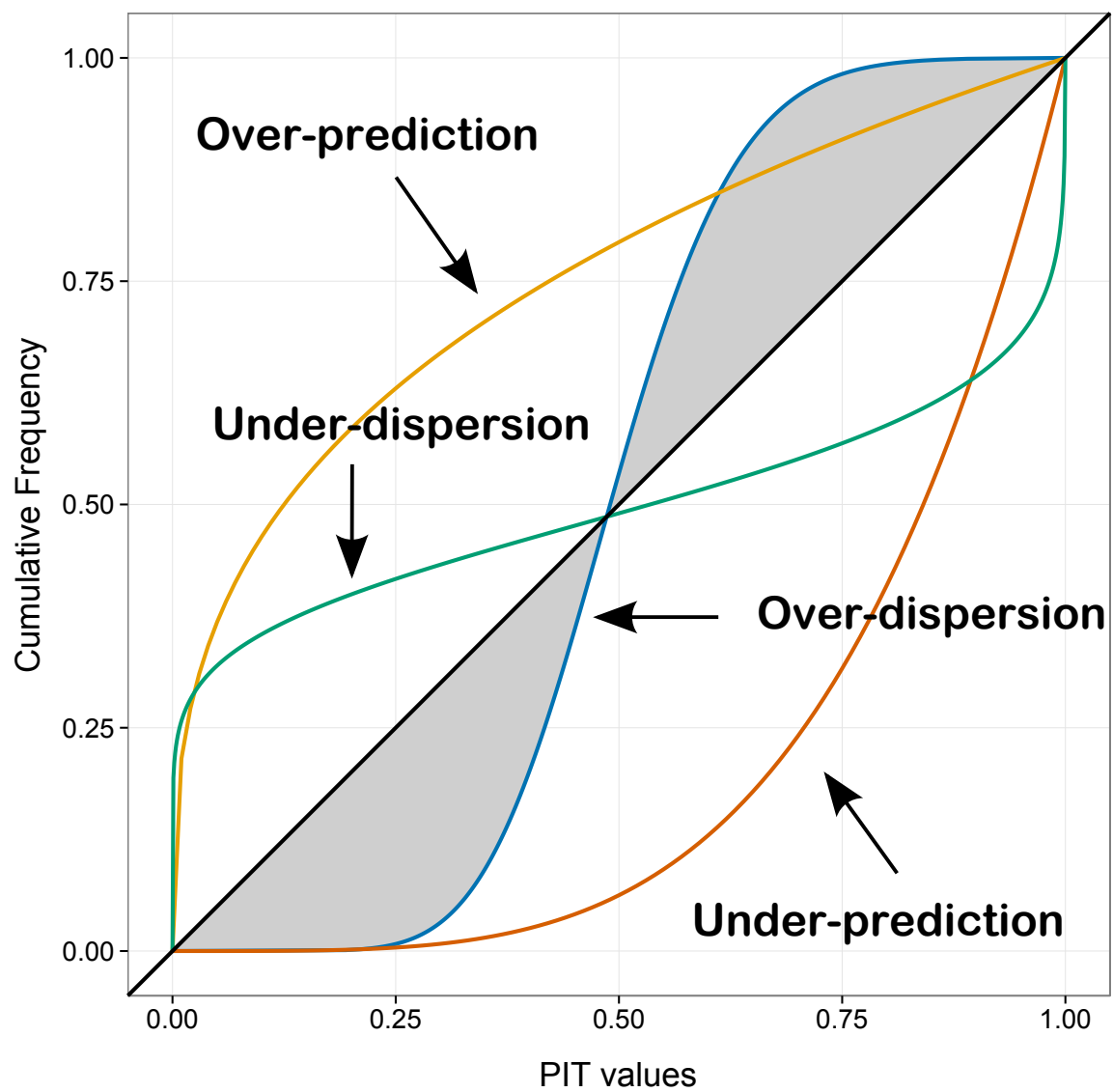


Figure 4.1: Common shapes and usual interpretation of the PIT diagram. The grey area between the blue curve and the bisector can be used to summarize the deviation of the PIT diagram from the ideal situation (cf. the Alpha score in Section 4.2.3).

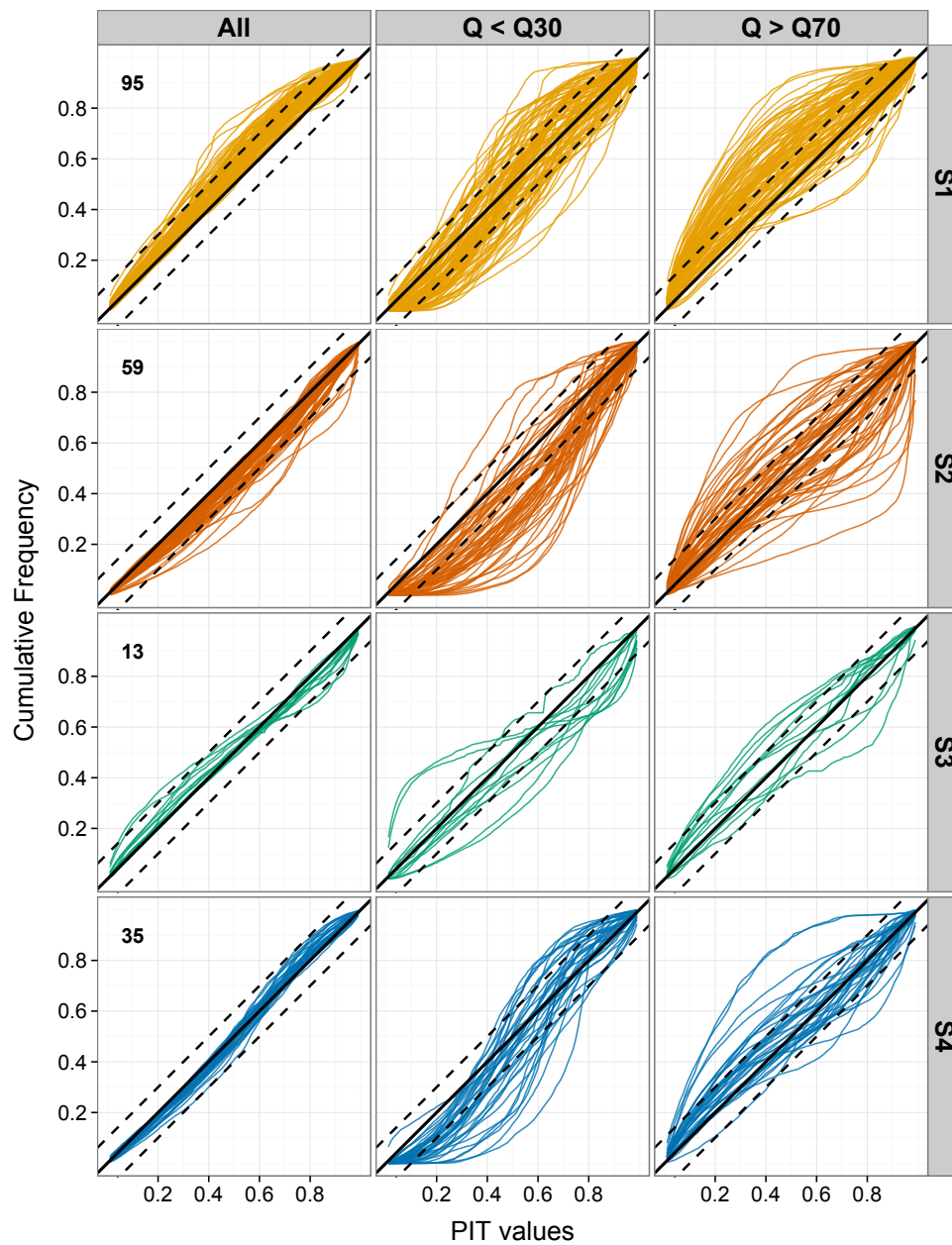


Figure 4.2: PIT diagrams calculated over the entire flow range (left) and PIT diagrams calculated for low (center) and high (right) flows. Each line corresponds to one of the 202 catchments. Four groups of catchments (S1 to S4) are defined based on common shapes of the curves (see Figure 4.1). The number of catchments within each group is given in the upper left corners.

reliable when all flows were analyzed altogether now depict unreliable predictive distributions when high and low flows are analyzed separately. Note that these differences cannot be attributed to the smaller sample sizes used to build the flow-conditional PIT diagrams. Indeed, with a 5-year validation period and hourly data, each high or low flow PIT diagram is built with more than 13,000 data points.

This result shows that the PIT diagrams calculated over the entire flow range can hide some deficiencies of the predictive distributions and, consequently, provide an overly confident evaluation of forecast reliability. Compensation effects may occur: a system with an over-estimation (or over-dispersion) for low flows and an underestimation (or under-dispersion) for high flows may yield a PIT diagram indicating good reliability when flows are considered altogether without data stratification.

### **4.3.3 Looking after hidden compensations when evaluating forecasts with the PIT diagram**

To further illustrate the issues of inconsistency of forecast reliability evaluation across flow groups, we used the Alpha score. Figure 4.3 shows the differences between the Alpha scores computed for each flow group (low and high flows) and the Alpha score computed over the entire flow range. Differences for low flows are plotted in the x-axis and for high flows in the y-axis. If a point is located far away from the position (0,0), it means that inconsistencies exist between the reliability analysis using all flows and the analysis stratified by flow groups. Different behaviors, inconsistencies and compensation effects are highlighted. While consistently reliable predictive distributions are obtained for a few catchments, significantly lower values of the Alpha score, either for low flows (x-axis) or for high flows (y-axis), can be found. Compensations are even more apparent when both scores (for low and high flows) are significantly lower than the one obtained from the PIT diagram computed over the entire flow range. Such a case is illustrated by the catchment represented by a black triangle: while its Alpha score using all flows is 0.94, the scores obtained for low and high flows are only 0.69 and 0.76, respectively. For this particular catchment, the PIT diagrams reveal that uncertainty is under-estimated for high flows while it is overestimated for low flows, illustrating that two non-ideal situations can result in an apparently satisfying evaluation of forecast reliability.

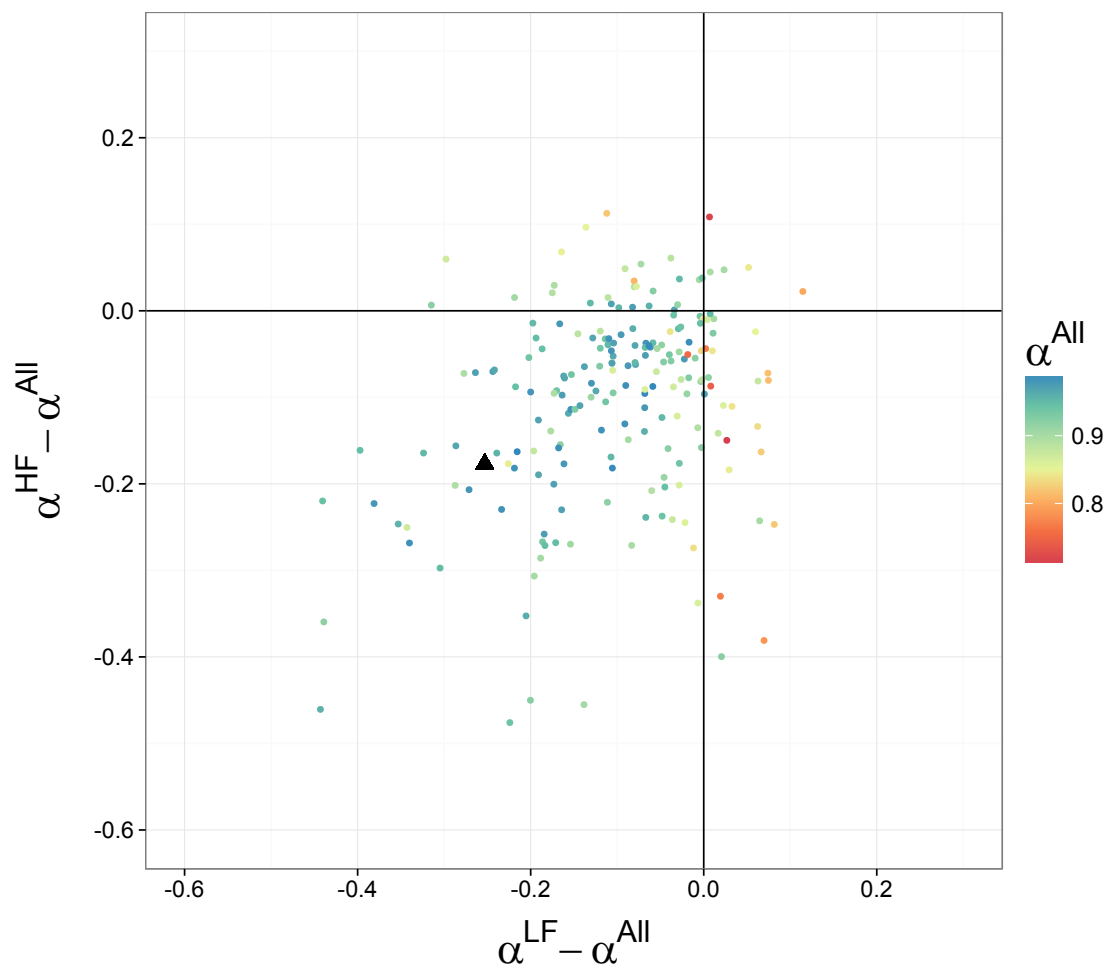


Figure 4.3: Differences between Alpha scores when they are computed over the entire flow range ( $\alpha^{\text{All}}$ ) or for either low flows ( $\alpha^{\text{LF}}$ ) or high flows ( $\alpha^{\text{HF}}$ ). Each dot corresponds to one of the 202 catchments. The dots are coloured according to the Alpha score computed over the entire flow range ( $\alpha^{\text{All}}$ ).



## 4.4 Conclusions and recommendations

This study contributes to understand how tools and scores commonly applied in streamflow forecast verification to detect probability biases can potentially present a risk of misinterpretations. The main issues that may affect the interpretation of the PIT diagram, one of the most common tools to evaluate the reliability of probabilistic prediction systems, were examined on a set of 202 catchments using 24-hour-ahead flow forecasts. Evaluation at each catchment considered either the entire flow range or subsets of high and low predicted flows separately. Results show that compensation effects may occur between flow groups: while the PIT diagram evaluated over the whole flow range may indicate a reliable forecasting system, it may hide a lack of reliability and different probability biases when high or low flows are predicted. A too rapid overall analysis of the quality of a prediction system may thus hide important biases in predictions that are of most interest for forecast users involved in flood forecasting, risk and water resources management.

Other conditioning strategies to better assess forecast attributes may also be helpful depending on the forecast context: focusing on different seasons or on rainfall-driven parts of the predicted hydrographs, or using a measure of spread of the predictive distributions to stratify the data could provide additional insights into potential deficiencies or hidden skills of probabilistic forecasts.

As shown in this study, it is important to pay attention to the necessity and the benefits of a simple conditioning approach that can avoid an unduly optimistic evaluation of forecast quality and help identifying avenues for improving the probabilistic forecasting system. It is also consistent with a more user-oriented evaluation, where high or low predicted flows may be of interest independently.

## 4.5 Acknowledgements

The authors thank Météo-France for providing the meteorological data and Banque HYDRO for the hydrological data. The financial support of SCHAPI to the first author is also gratefully acknowledged.

# 5

## PRÉSENTATION DES MÉTHODES UTILISÉES EN SIMULATION



## 5.1 Introduction

Ce chapitre a pour objectif de présenter les différentes méthodes que nous avons testées dans un contexte de simulation. Nous nous sommes d'abord penchés sur deux approches très communément utilisées en modélisation hydrologique, la méthode GLUE et l'approche bayésienne. Ces deux approches diffèrent dans leur manière de quantifier l'incertitude prédictive, mais partagent le fait d'associer la problématique de l'estimation des paramètres à celle de la quantification de l'incertitude prédictive. Nous présentons dans un deuxième temps d'autres approches de quantification de l'incertitude prédictive, l'approche multi-modèles et l'approche statistique de post-traitement.

## 5.2 Calage et analyse des incertitudes selon la méthode GLUE

### 5.2.1 Pourquoi s'intéresser à cette méthode ?

Comme nous l'avons mentionné au chapitre 1, la méthode GLUE a été utilisée dans de nombreuses études. La plupart des applications de cette méthode font appel à une fonction objectif usuelle pour évaluer les performances obtenues avec différents jeux de paramètres tirés de manière aléatoire, sélectionner certains jeux de paramètres considérés comme acceptables, et pondérer la distribution des résultats obtenus avec l'ensemble de ces jeux de paramètres pour en déduire des intervalles prédictifs exprimant l'incertitude de modélisation.

Certains auteurs sont très critiques envers la méthode GLUE, et considèrent que les choix arbitraires rendent les intervalles prédictifs sans fondements statistiques. Ils ne devraient donc pas être utilisés dans le cadre d'un travail scientifique. Pour les utilisateurs de cette approche, la méthode est une alternative aux autres méthodes plus formelles, mais jugées inadaptées aux cas réels rencontrés en modélisation environnementale.

Les débats théoriques semblent interminables. La méthode continue d'être très utilisée, dans différents contextes de modélisation hydrologique, et il nous a semblé intéressant de tester de manière pragmatique cette approche sur notre large échantillon de bassins versants.

## 5.2.2 Mise en œuvre de la méthode GLUE

### Choix méthodologiques

Trois critères de performance quadratiques, notés NSQ, NSV et NSL ont été utilisés. Ils correspondent respectivement aux critères de Nash-Sutcliffe calculés sur les débits, les racines carrées des débits et les logarithmes des débits. Ces critères sont utilisés dans de nombreuses applications de la méthode GLUE, et sont de manière plus générale représentatifs d'applications s'intéressant respectivement plutôt aux crues, à tous les débits ou aux étiages.

Pour distinguer les paramètres acceptables des paramètres rejetés, nous avons utilisé des seuils de performance relatifs. Les paramètres dont les performances étaient supérieures à 70, 80 et 90% de la valeur maximale obtenue lors de l'étape d'exploration des paramètres (décrite ci-dessous) ont été considérés comme acceptables. Il s'agit d'un choix rendu nécessaire par le fait de travailler sur des bassins versants variés. Il ne nous a en effet pas semblé opportun de fixer un même seuil pour l'ensemble des bassins versants en raison de la dépendance des critères de Nash-Sutcliffe aux caractéristiques des bassins et des périodes de calage.

### Stratégie d'échantillonnage

La méthode d'échantillonnage habituellement utilisée avec la méthode GLUE consiste à générer de manière aléatoire des jeux de paramètres dont les gammes de variations ont été fixées par l'utilisateur. Cette exploration globale des surfaces de réponse est très inefficace, et requiert un nombre très important de simulations. Le fait de travailler sur un grand nombre de bassins nous a conduit à procéder en deux temps.

Nous avons d'abord évalué les performances d'un million de jeux de paramètres pour chaque bassin et chaque période de calage<sup>1</sup>. Ces jeux de paramètres ont été obtenus par un échantillonnage de type latin hypercube, une variante d'un échantillonnage de Monte Carlo, dans l'espace des paramètres transformés. Cette première étape nous a permis d'explorer de manière systématique les surfaces de réponse.

Une deuxième étape a été nécessaire pour obtenir un nombre suffisant de jeux de paramètres. Nous avons donc, pour chaque bassin, chaque période, chaque critère de performance, et chaque seuil de performance, réalisé des tirages aléatoires dans un espace des paramètres réduit, de manière à obtenir un ensemble de mille jeux de paramètres acceptables. Les bornes minimales et maximales de chaque paramètre ont été limitées grâce aux informations

---

1. La procédure de calage-contrôle est présentée au chapitre 6.

obtenues lors de l'exploration systématique des surfaces de réponse à l'aide d'un million de jeux de paramètre. En raison des compensations possibles entre les deux paramètres liés à la capacité des réservoirs, nous avons également estimé les valeurs maximales de la capacité du réservoir de production en fonction de la capacité du réservoir de routage, ce qui a permis de limiter considérablement le nombre de simulations nécessaires pour obtenir mille jeux de paramètres. Les tirages aléatoires ont cette fois été effectués dans l'espace des paramètres non-transformés, afin d'éviter de pondérer les distributions des paramètres en fonction du déterminant de la matrice jacobienne décrivant la transformation des paramètres.

Dans la grande majorité des cas, cette stratégie d'échantillonnage nous a permis d'obtenir facilement des ensembles de mille jeux de paramètres. Nous n'avons rencontré des difficultés qu'avec sept bassins versants sur un total de 202. Certaines surfaces de réponse se sont révélés trop complexes pour qu'une caractérisation satisfaisante soit possible dans le cadre de notre étude. Les cas les plus difficiles sont apparus pour les raisons suivantes :

- des compensations existent entre les différents paramètres du modèle ;
- le modèle est peu sensible aux variations d'un paramètre ;
- le critère de performance est insensible aux différentes simulations obtenues ;
- des fonctionnements différents permettent d'obtenir des performances comparables.

Nous avons renoncé à utiliser les sept bassins concernés. Cela ne représente qu'un faible nombre de cas, et ne remet pas en cause la généralité de nos travaux.

### **Calcul de l'incertitude prédictive**

L'incertitude prédictive s'obtient à partir des simulations obtenues avec les jeux de paramètres considérés comme acceptables. Les quantiles pondérés de la distribution empirique des débits simulés sont calculés à chaque pas de temps. Le poids attribué à chaque jeu de paramètres est proportionnel au critère de performance (voir par exemple [Beven et Freer, 2001](#)).

### **5.2.3 Analyse des jeux de paramètres**

Nous présentons brièvement les caractéristiques des jeux de paramètres retenus. Nous souhaitons mettre en évidence l'impact du critère d'évaluation et l'impact du seuil sur les jeux de paramètres qui constituent l'ensemble prédictif.

### Impact des choix méthodologiques sur les performances

Par construction, les performances individuelles des jeux de paramètres sont meilleures lorsque le seuil augmente. La Figure 5.1 illustre cette évidence pour les expériences NSQ<sub>70</sub>, NSQ<sub>80</sub> et NSQ<sub>90</sub>, qui correspondent à une utilisation du critère de Nash-Sutcliffe avec une valeur seuil de performance respectivement égale à 70, 80 et 90% de sa valeur maximale. Lorsque le seuil augmente, les distributions se déplacent vers de plus fortes valeurs et la variabilité des performances au sein des ensembles de paramètres diminue.

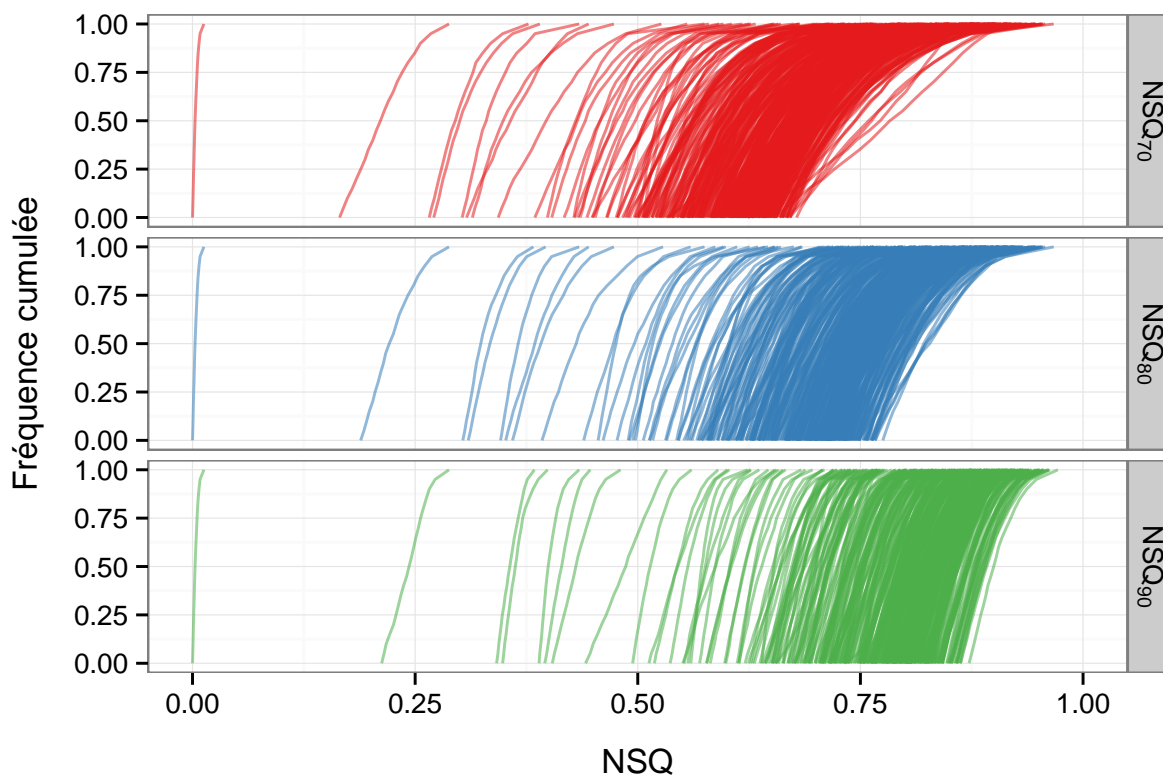


FIGURE 5.1 – Distributions des critères de performance NSQ obtenus pour 1000 jeux de paramètres, pour chaque bassin et chaque période.

### Impact des choix méthodologiques sur les paramètres

Si la surface de réponse est suffisamment bien définie, on s'attend à ce que les paramètres qui obtiennent les meilleures performances se situent dans une zone limitée de l'espace des paramètres, située à proximité d'un optimum global.

La visualisation des jeux de paramètres retenus suggère que c'est en effet le plus souvent le cas. La Figure 5.2 fournit un résumé des distributions marginales des paramètres transformés pour les expériences NSQ<sub>70</sub>, NSQ<sub>80</sub> et NSQ<sub>90</sub>. Il apparaît que la variabilité, mesurée ici par

les écarts interquartiles des distributions marginales des paramètres, diminue lorsque le seuil sur le critère de performance augmente. Les écarts interquartiles restent néanmoins importants au regard des gammes de variation habituelles des paramètres du modèle GR4J. Des résultats similaires sont obtenus avec les deux autres critères NSV et NSL.

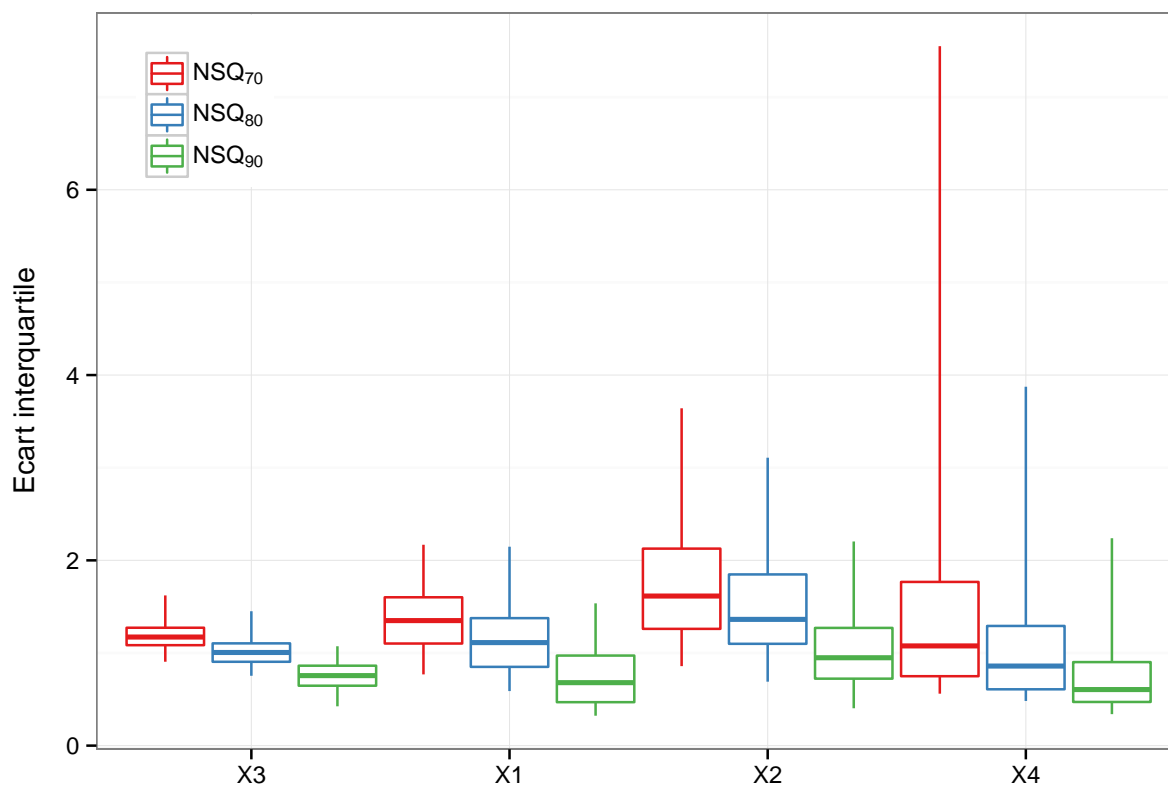


FIGURE 5.2 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des écarts interquartiles des distributions marginales des paramètres transformés obtenus avec le critère NSQ.

#### 5.2.4 Synthèse

La méthode GLUE est une des méthodes d'analyse des incertitudes les plus utilisées en modélisation hydrologique. Nous avons choisi d'appliquer cette approche selon des modalités fréquemment rencontrées dans la littérature. Comme les jeux de paramètres qui composent l'ensemble prédictif varient selon le critère de performance et le seuil utilisé, nous avons réalisé neuf expériences qui reflètent la diversité des résultats qui peuvent être obtenus avec cette méthode.



## 5.3 Calage et incertitude prédictive selon l'approche bayésienne

### 5.3.1 Mise en œuvre du calage bayésien

#### Choix méthodologiques

Nous avons choisi de nous concentrer sur l'application la plus simple possible du calage bayésien.

Lorsque les erreurs résiduelles sont supposées être des réalisations indépendantes d'une loi normale, appelée aussi gaussienne, qu'aucune information *a priori* n'est utilisée, et que la contribution de la variance des résidus est marginalisée, la densité de probabilité *a posteriori* des paramètres du modèle hydrologique, notée  $p(\theta|D)$  où  $D$  désigne de manière générale les données disponibles, est liée aux critères de performance quadratiques, comme l'ont par exemple rappelé [Vrugt et al. \(2003b\)](#), [Kavetski et al. \(2006a\)](#) et [Lu et al. \(2012\)](#) :

$$p(\theta|D) \propto \left[ \sum_{t=1}^n (Q_{o,t} - Q_{s,t}(\theta))^2 \right]^{-n/2} \quad (5.1)$$

où  $n$  est le nombre de pas de temps,  $Q_{o,t}$  et  $Q_{s,t}(\theta)$  sont respectivement le débit observé, et le débit simulé avec les paramètres  $\theta$ , au pas de temps  $t$ .

Une présentation plus complète de cette formulation du calage bayésien est disponible dans l'annexe B.1.

Une stratégie couramment adoptée pour stabiliser la variance des résidus consiste à utiliser une transformation de variable. La transformation de Box-Cox est la plus fréquemment rencontrée ; elle est définie par :

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0, \\ \log(y), & \text{si } \lambda = 0. \end{cases} \quad (5.2)$$

où  $y$  est un vecteur réel et  $\lambda$  est le paramètre de la transformation.

Nous avons testé quatre fonctions de vraisemblance qui correspondent à trois transformations de Box-Cox, avec les paramètres  $\lambda = 0, 0.3$  et  $0.5$ , et à la fonction de vraisemblance classique des moindres carrées, ce qui revient à utiliser  $\lambda = 1$ . Nous noterons ces variantes  $BC_0$ ,  $BC_{0.3}$ ,  $BC_{0.5}$ , et  $BC_1$  respectivement. Ce choix correspond à l'utilisation des transfor-

mations logarithmique et racine carrée que nous utilisons couramment pour le calcul des critères NSL et NSV, ainsi qu'à un choix que l'on retrouve assez souvent dans la littérature,  $\lambda = 0.3$  (e.g., [Vrugt et al., 2009](#)).

L'impact de l'utilisation d'une transformation puissance sur les paramètres et les performances du modèle GR4J a été étudié par [Oudin et al. \(2006\)](#) et [Le Moine \(2008\)](#). Leurs travaux ont montré que le choix de la puissance conduisait à privilégier l'évaluation de certaines gammes de débits. Les faibles valeurs, entre 0 et 0.3, privilégient plutôt les basses eaux tandis que les plus fortes valeurs, supérieures à 1, privilégient plutôt les hautes eaux. Il nous a semblé intéressant de tester l'impact de la transformation de Box-Cox, à la fois sur la distribution des paramètres du modèle hydrologique, mais également sur l'incertitude prédictive qui reflète les hypothèses statistiques du modèle d'erreur.

### Stratégie d'échantillonnage

L'échantillonnage de la distribution jointe des paramètres du modèle hydrologique a été effectué avec un algorithme de Metropolis adaptatif. Il s'agit d'une méthode de Monte Carlo par chaînes de Markov (MCMC) décrite par [Haario et al. \(2001\)](#) et [Haario et al. \(2006\)](#). Le principe des méthodes MCMC est de produire, à l'aide d'une chaîne de Markov, un échantillon de jeux de paramètres dont la distribution empirique représente une approximation de la distribution cible. L'algorithme de Metropolis adaptatif a été utilisé dans plusieurs études en modélisation hydrologique, et a montré de bonnes performances ([Marshall et al., 2007](#); [Smith et Marshall, 2008](#)). Il constitue une alternative à d'autres méthodes adaptatives comme celles proposées par exemple par [Thyer et al. \(2009\)](#), [Laloy et Vrugt \(2012\)](#) ou [Harrison et al. \(2012\)](#).

Nous avons utilisé trois chaînes de Markov en parallèle de manière à évaluer la convergence des trajectoires vers la distribution *a posteriori*. Les trois chaînes ont été initialisées au niveau ou à proximité du maximum de vraisemblance de manière à limiter le nombre de simulations. Pour deux chaînes, le paramètre transformé correspondant à la capacité du réservoir de routage a été perturbé de  $\pm 10\%$ . Chacune des chaînes évolue pendant 5000 itérations, l'adaptation de la matrice de covariance qui permet de guider l'échantillonnage est arrêtée au bout de 4000 itérations, et les 500 derniers jeux de paramètres obtenus pour chaque chaîne sont conservés. Nous avons visualisé les trajectoires des trois chaînes, et calculé le critère de Gelman-Rubin ([Gelman et al., 2004](#)) pour diagnostiquer d'éventuels problèmes de convergence. Ce critère repose sur une comparaison entre la variance entre les chaînes et la variance dans les chaînes.

Nous n'avons rencontré des difficultés que dans un nombre très limité de cas, 11 parmi 1616. Il s'agissait de surfaces de réponse bi-modales pour lesquelles deux des trois chaînes évoluaient à proximité d'un mode différent de celui de la troisième chaîne. L'impact sur les débits simulés s'est révélé négligeable, les deux modes de convergence étant très proches, et nous avons choisi d'utiliser l'ensemble des jeux de paramètres comme s'il s'agissait d'échantillons de la distribution cible.

### **Calcul de l'incertitude prédictive**

Dans le cadre de notre application du calage bayésien, l'incertitude prédictive résulte d'une part de la variabilité apportée par les simulations obtenues avec les différents jeux de paramètres constituant une approximation numérique de la distribution *a posteriori*, et d'autre part de la variabilité apportée par les erreurs résiduelles.

A chaque pas de temps, chacune des valeurs obtenues avec les différents jeux de paramètres est perturbée par une erreur aléatoire qui suit une loi gaussienne de moyenne nulle et de variance constante. Dans notre cas, la variance de ces perturbations est indépendante des jeux de paramètres du modèle hydrologique, et a été estimée à partir des erreurs résiduelles obtenues avec le jeu de paramètres qui correspond au maximum de vraisemblance.

Lorsque la transformation de Box-Cox est utilisée, la transformation inverse est appliquée, à chaque pas de temps, aux distributions des valeurs obtenues après avoir été perturbées dans le domaine transformé.

### **5.3.2 Analyse des jeux de paramètres**

Nous présentons à présent une brève analyse des jeux de paramètres obtenus avec le calage bayésien. Nous souhaitons mettre en évidence l'impact de la transformation de Box-Cox sur les jeux de paramètres.

#### **Impact des choix méthodologiques sur les paramètres**

La Figure 5.3 fournit un résumé des valeurs médianes des distributions marginales des paramètres transformés pour les quatre expériences effectuées. Nous pouvons constater que les valeurs des réservoirs de routage (X3) et de production (X1) semblent diminuer lorsque le paramètre de la transformation de Box-Cox augmente, au contraire des valeurs du paramètre d'échange (X2) et du temps de base (X4) qui semblent, quant à elles, augmenter.

La Figure 5.4 présente les distributions des écarts interquartiles des distributions marginales des paramètres transformés. Nous constatons que la variabilité, et donc l'incertitude sur les paramètres X3 et X4 a tendance à diminuer lorsque le paramètre de la transformation de Box-Cox augmente. Une possible explication est que les critères qui privilégient les plus hauts débits permettent de mieux discriminer les paramètres liés à la fonction de routage, en raison de l'importance donnée par ces critères à la représentation de la dynamique des crues.

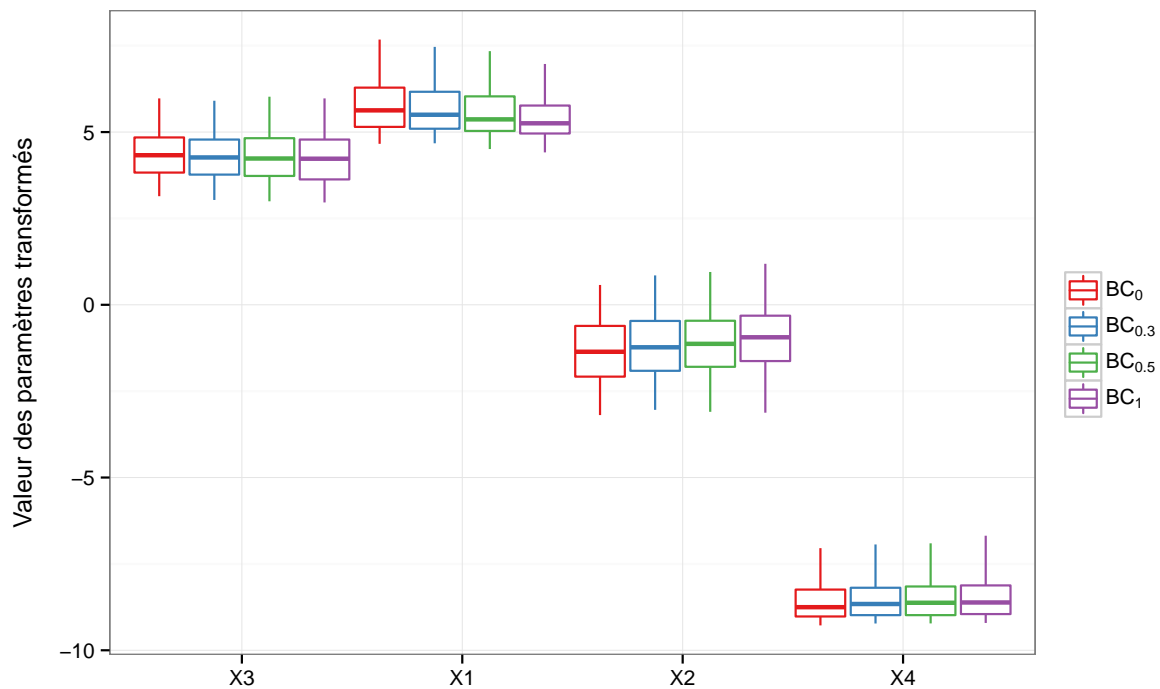


FIGURE 5.3 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des valeurs médianes des distributions marginales des paramètres transformés obtenus avec le calage bayésien (MCMC).

### 5.3.3 Synthèse

Dans le cadre du calage bayésien, le choix d'une fonction de vraisemblance reflète la construction d'un modèle probabiliste capable de représenter la variabilité des débits observés.

Nous avons réalisé quatre expériences de calage selon l'approche bayésienne. Ces expériences dépendent du paramètre de la transformation de Box-Cox qui est utilisé pour limiter l'hétéroscédasticité des résidus. Les jeux de paramètres qui constituent l'échantillon *a posteriori* diffèrent suivant la transformation utilisée. Nous verrons dans quelle mesure cela impacte la quantification de l'incertitude prédictive au prochain chapitre.

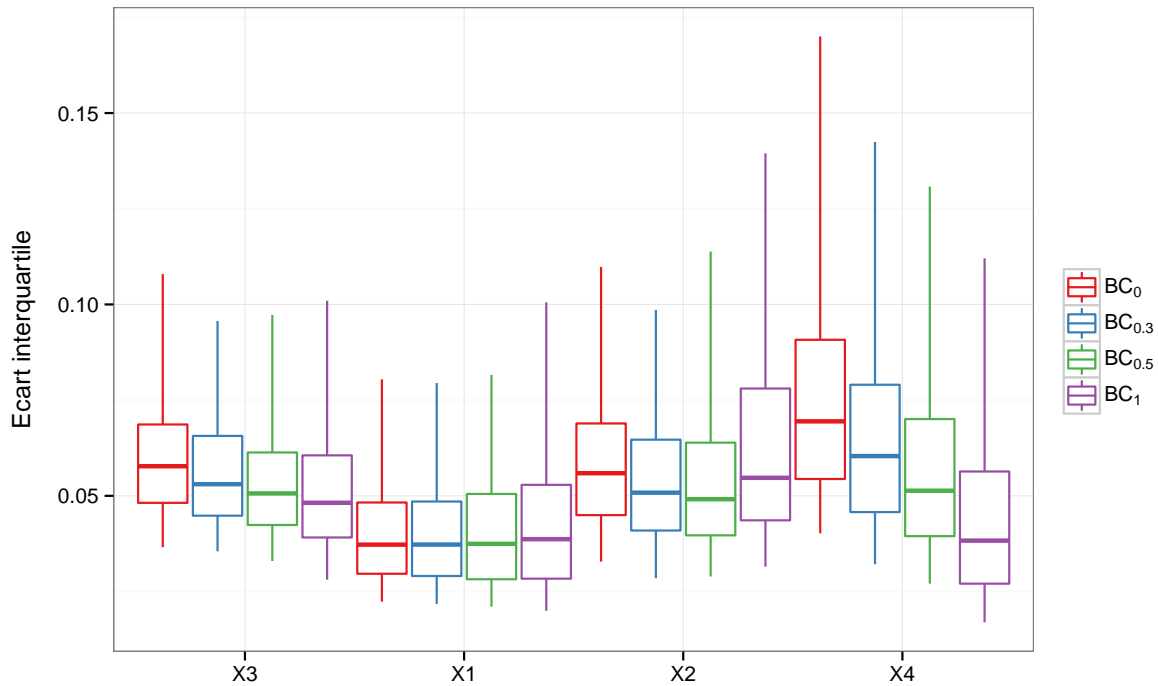


FIGURE 5.4 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des écarts interquartiles des distributions marginales des paramètres transformés obtenus avec le calage bayésien (MCMC).

## 5.4 Approche multi-modèles

La disponibilité des codes de calcul au sein de l'équipe nous a permis d'obtenir des simulations avec dix-sept modèles hydrologiques (voir Table 5.1) précédemment utilisés et décrits par Perrin (2000), Perrin *et al.* (2001) et Mathevet (2005). Nous avons suivi la même démarche que celle décrite dans l'étude de Velazquez *et al.* (2010) :

- Chaque modèle est calé avec le critère de Nash-Sutcliffe calculé sur les racines des débits (NSV).
- Aucune pondération n'est utilisée pour créer l'ensemble prédictif.
- La loi gamma est utilisée pour modéliser la distribution prédictive à chaque pas de temps ; ses deux paramètres sont estimés selon la méthode des moments, c'est-à-dire en calculant la moyenne et l'écart-type de la distribution des débits simulés à chaque pas de temps par l'ensemble des modèles hydrologiques.

N°	Modèle	Version	Nombre de paramètres	Nombre de réservoirs
1	CEQUEAU	CEQU	9	3
2	CREC	CREC	8	3
3	GARDENIA	GARD	7	3
4	GR4J	GR4J	4	2
5	HBV	HBV0	9	3
6	HYMOD	HYM0	6	5
7	IHACRES	IHAC	6	3
8	MOHYSE	MOHY	7	2
9	MORDOR	MORD	6	4
10	NAM	NAM0	10	7
11	PDM	PDM0	8	4
12	SACRAMENTO	SACR	13	6
13	SIMHYD	SIMH	8	3
14	SMAR	SMAR	9	3
15	TANK	TANK	10	5
16	TOPMODEL	TOPM	8	3
17	WAGENINGEN	WAGE	8	4

TABLE 5.1 – Liste des dix-sept modèles hydrologiques utilisés pour l’approche multi-modèles. Voir [Perrin \(2000\)](#) pour plus de détails concernant les versions de ces modèles.

## 5.5 Méthodes de post-traitement statistique

5

Les méthodes de post-traitement statistique utilisées sont décrites ici succinctement. Une présentation plus détaillée est disponible dans l’annexe B.2.

### Une méthode empirique de quantification de l’incertitude (CemaM)

Avant que cette thèse ne débute, une méthode empirique de quantification de l’incertitude prédictive a été proposée au sein de l’équipe d’hydrologie d’Antony par Claude Michel, et testée dans le cadre d’un stage de DEA ([Berthier, 2005](#)). Cette méthode consiste à décrire la distribution empirique des erreurs relatives en fonction des valeurs du débit simulé par un modèle hydrologique ; des quantiles de la distribution empirique des erreurs relatives sont estimés par tranches de débits. Ces quantiles sont des coefficients multiplicatifs qui peuvent être utilisés pour définir les intervalles prédictifs en fonction du débit simulé par le modèle. Pour un intervalle prédictif à 90% par exemple, les bornes de l’intervalle seront définies à partir des quantiles d’ordre 0.05 et 0.95.

Comme ces quantiles dépendent de la tranche du débit simulé sur lesquels ils ont été estimés, il est nécessaire de définir une manière de tenir compte des cas d’extrapolation ; nous

avons choisi d'utiliser une extrapolation à valeur constante. Par ailleurs, nous avons fixé à 10 le nombre de tranches de débits sur lesquelles les quantiles sont estimés. Il s'agit d'un compromis entre le fait que le nombre de valeurs disponibles sur une période de cinq ans est limité, et la nécessité de décrire de manière suffisamment fine les variations des distributions d'erreurs en fonction du débit simulé.

Afin de mettre en évidence l'intérêt d'utiliser différentes tranches de débits, nous avons testé une version simplifiée de la méthode CemaM, la méthode CemaO, qui ne considère quant à elle qu'une seule tranche de débits.

### **Approche méta-gaussienne (MeGa)**

La méthode proposée par [Montanari et Brath \(2004\)](#) utilise un modèle statistique appelé méta-gaussien, introduit en hydrologie par [Kelly et Krzysztofowicz \(1997\)](#), pour décrire la relation statistique entre un débit simulé et l'erreur prédite. Le modèle méta-gaussien est construit autour de la transformation en quantiles normaux, une transformation empirique et non-paramétrique qui permet d'obtenir des densités marginales gaussiennes. [Montanari et Brath \(2004\)](#) proposent de caractériser la dépendance entre les deux variables par une régression linéaire dans le domaine normal, et d'en déduire par transformation inverse les intervalles prédictifs du débit.

[Montanari et Grossi \(2008\)](#) ont présenté une modification de la méthode, que nous avons utilisée. Le modèle méta-gaussien est appliqué séparément aux erreurs positives et négatives, et la densité prédictive finale est construite comme une densité mélange. Une régression linéaire est utilisée dans le domaine normal pour modéliser séparément ces erreurs, de manière à obtenir des distributions prédictives des erreurs qui évoluent en fonction de la valeur du débit simulé.

### **Traitement de l'incertitude prédictive conditionnelle (MCP)**

Dans la méthode proposée par [Coccia et Todini \(2011\)](#), *Model Conditional Processor*, la variable aléatoire prédite par le modèle statistique utilisé n'est pas l'erreur mais le débit. L'approche consiste donc à estimer la distribution prédictive du débit observé en fonction du débit simulé. Les auteurs proposent d'utiliser le modèle méta-gaussien pour obtenir une distribution jointe multi-variée qu'ils supposent multi-normale, ce qui revient à utiliser une régression linéaire entre les débits simulés et les débits observés dans le domaine normal.

### Régression quantile (RQ)

L'objectif de l'étude présentée par [Weerts \*et al.\* \(2011\)](#) était d'estimer l'incertitude globale d'un système de prévision des crues à partir d'une analyse des propriétés statistiques des séries temporelles des erreurs. La régression quantile est utilisée pour caractériser la relation entre les deux variables préalablement transformées dans le domaine normal. La régression quantile, introduite à l'origine en économie, permet de s'affranchir de l'hypothèse de normalité des résidus habituellement formulée pour les modèles de régression linéaire, et d'étendre les techniques de régression aux quantiles. Elle permet de décrire la variation des quantiles d'une variable d'intérêt – en l'occurrence l'erreur résiduelle – en fonction d'une variable explicative – ici la hauteur d'eau simulée –, et donc d'estimer des marges d'erreurs.

### Transformations de variables et normalité des résidus

Nous avons également appliqué différentes variantes d'une méthode que l'on peut considérer comme une méthode de référence. Il s'agit de faire l'hypothèse que les erreurs additives suivent une distribution gaussienne, de moyenne nulle et dont la variance est égale à la variance empirique des erreurs estimée sur la période de calage. Les variantes de cette méthode utilisent des transformations de Box-Cox, de manière similaire à ce qui est réalisé dans le cadre du calage bayésien. Nous les noterons  $BC_0$ ,  $BC_{0.3}$  et  $BC_{0.5}$ . Pour conserver une cohérence dans les acronymes utilisés, nous désignerons la méthode de référence, sans transformation de Box-Cox, par le terme  $BC_1$ .

## 5.6 Synthèse

Ce chapitre avait pour objectif de présenter les différentes méthodes que nous avons testées dans un contexte de simulation. Nous avons d'abord décrit la manière dont les deux méthodes liées à l'estimation des paramètres, la méthode GLUE et l'approche bayésienne, ont été appliquées. Nous avons ensuite présenté deux autres approches, l'approche multi-modèles et l'approche de post-traitement, pour laquelle différentes méthodes ont été mises en œuvre. Ces quatre approches sont résumées à l'aide d'une illustration sur la Figure 5.5.

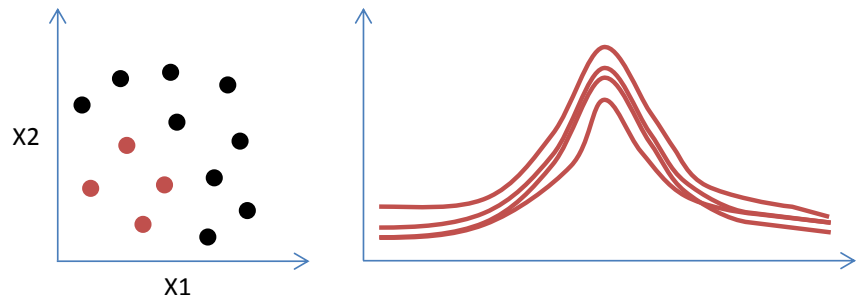
Bien sûr, aucune comparaison ne peut prétendre à l'exhaustivité, et nous avons dû faire des choix quant aux modalités d'application de ces méthodes. Ces choix influencent inévitablement les résultats que nous présentons au prochain chapitre. Nous avons néanmoins tenté de rendre compte de l'impact de ces choix en réalisant plusieurs expériences pour les



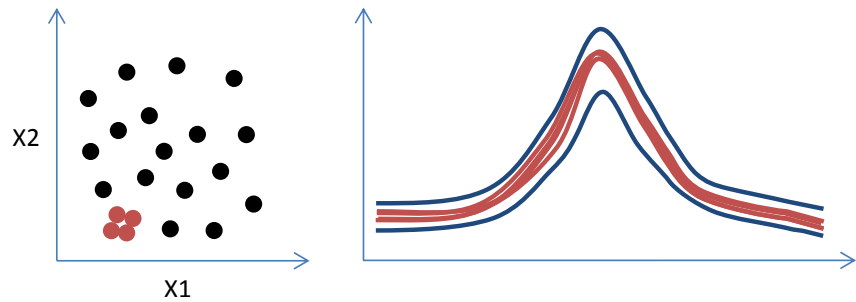
méthodes liées à l'estimation des paramètres du modèle hydrologique, et en considérant différentes méthodes de post-traitement.

**GLUE**

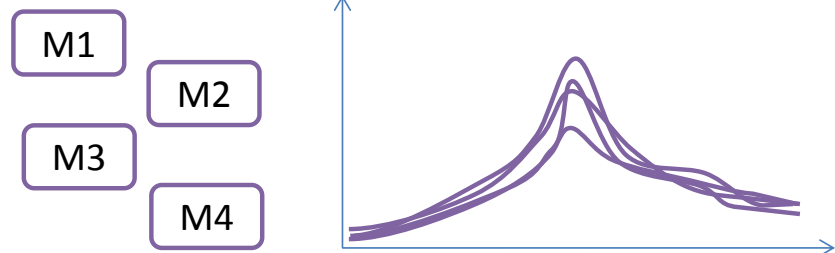
Ensemble de jeux de paramètres

**Approche bayésienne**

Ensemble de jeux de paramètres + modèle d'erreurs résiduelles

**Approche multi-modèles**

Ensemble de modèles

**Post-traitement**

Modèle d'erreurs résiduelles

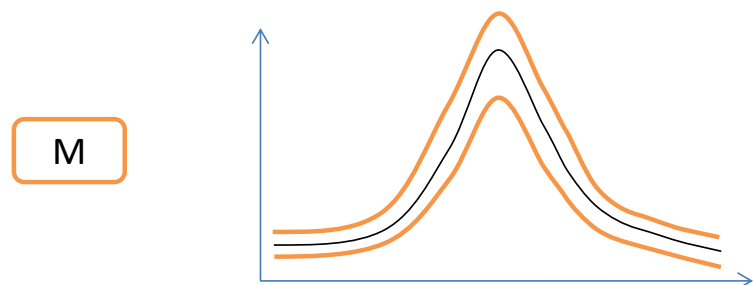


FIGURE 5.5 – Illustration des différentes méthodes testées dans un contexte de simulation. Pour la méthode GLUE et l'approche bayésienne, des paramètres (en rouge) sont sélectionnés dans l'espace des paramètres, ici à deux dimensions ( $X_1$ ,  $X_2$ ). L'approche bayésienne utilise également un modèle d'erreurs résiduelles. L'approche multi-modèles s'appuie sur un ensemble de simulations obtenues avec plusieurs modèles (M1–M4). Les méthodes de post-traitement utilisent un modèle d'erreurs résiduelles.



# 6

## COMPARAISON DES MÉTHODES UTILISÉES EN SIMULATION



## 6.1 Introduction

Nous présentons dans ce chapitre une comparaison des résultats obtenus avec les méthodes présentées au chapitre précédent. Quatre types d'approches ont été retenus et différentes variantes ont été testées.

Nous rappelons dans un premier temps les modalités de la comparaison. Nous présentons ensuite les résultats suivant la stratégie d'évaluation présentée au chapitre 3.

## 6.2 Modalités de la comparaison

### 6.2.1 Procédure de calage-contrôle

Pour cette comparaison, toutes les méthodes ont été appliquées suivant une procédure de test classique, le calage-contrôle :

- La période totale (1997-2006) est divisée en deux sous-périodes de cinq ans.
- Une période est utilisée pour estimer les paramètres, l'autre pour évaluer les performances sur de nouvelles données.
- Le rôle des deux périodes est ensuite inversé.

Cette procédure de calage-contrôle permet ainsi d'obtenir des résultats en contrôle sur l'intégralité de la période. Elle s'entend ici au sens large, puisqu'elle comprend à la fois l'estimation des paramètres d'un ou plusieurs modèles hydrologiques, mais également celle des paramètres des modèles statistiques qui décrivent l'incertitude prédictive.

### 6.2.2 Stratégie d'évaluation

La stratégie d'évaluation se concentre sur les caractéristiques des distributions prédictives suivantes :

- la fiabilité, qui correspond à la cohérence statistique des probabilités ;
- la finesse, qui correspond à la capacité à concentrer les probabilités ;
- la performance globale.

Les indices de fiabilité et de finesse présentés au chapitre 3 sont dans un premier temps utilisés ; le CRPS est ensuite appliqué en tant que score global de performance.

## 6.3 Évaluation des différentes méthodes en terme de fiabilité et de finesse

### 6.3.1 Ensembles de paramètres et incertitude prédictive

Nous nous concentrons dans un premier temps sur les résultats obtenus avec les deux approches qui s'appuient sur des ensembles de paramètres pour quantifier l'incertitude prédictive, l'approche GLUE et l'approche du calage bayésien. Dans le cas de la méthode GLUE, les ensembles de jeux de paramètres utilisés pour quantifier l'incertitude prédictive varient suivant le critère de performance et la valeur seuil, tandis que dans le cas du calage bayésien, les ensembles varient suivant la fonction de vraisemblance utilisée.

La Figure 6.1 synthétise à l'aide de boxplots les distributions des indices de fiabilité et de finesse obtenus sur les 195 bassins considérés, pour les deux périodes de calage (haut) et de contrôle (bas). Ces indices ne varient pas indépendamment l'un de l'autre, et les Figures 6.2 et 6.3 permettent de visualiser ce lien. Pour faciliter la lecture de ces graphiques, une version bornée sur [-100%, 100%] de l'indice de finesse est utilisée en abscisse ; la fonction de transformation est la suivante :  $f(x) = \frac{x}{2-x}$ .

En ce qui concerne la méthode GLUE, il apparaît que pour les trois critères de performance, les indices de finesse augmentent lorsque la valeur seuil augmente, ce qui correspond au fait que l'augmentation du seuil conduit à une réduction de la variabilité des paramètres. De plus, nous pouvons remarquer pour une même valeur seuil, les distributions les plus fines sont obtenues avec le critère NSL, tandis que les distributions les plus larges sont obtenues avec le critère NSQ. Ceci s'explique par le fait que le critère NSL cible en priorité les bas débits, alors que le critère NSQ cible avant tout les hauts débits. Les paramètres qui obtiennent de bonnes performances avec le critère NSL n'apportent donc pas nécessairement la variabilité suffisante pour simuler les plus hauts débits, ce qui explique que les distributions prédictives soient, en moyenne, plus fines que celles obtenues avec les paramètres les plus performants au sens du critère NSQ. Par ailleurs, l'augmentation de la valeur seuil ne semble pas avoir un impact systématiquement positif sur les indices de fiabilité, et une grande variabilité est constatée pour toutes les variantes testées.

Nous pouvons également constater que la variabilité apportée par les ensembles de jeux de paramètres obtenus lors du calage bayésien (expériences MCMC\_BC<sub>0</sub>–MCMC\_BC<sub>1</sub>) est faible : les indices de finesse sont élevés alors que les indices de fiabilité sont faibles. Ces résultats sont à comparer avec ceux obtenus lorsque la variabilité apportée par le modèle d'er-

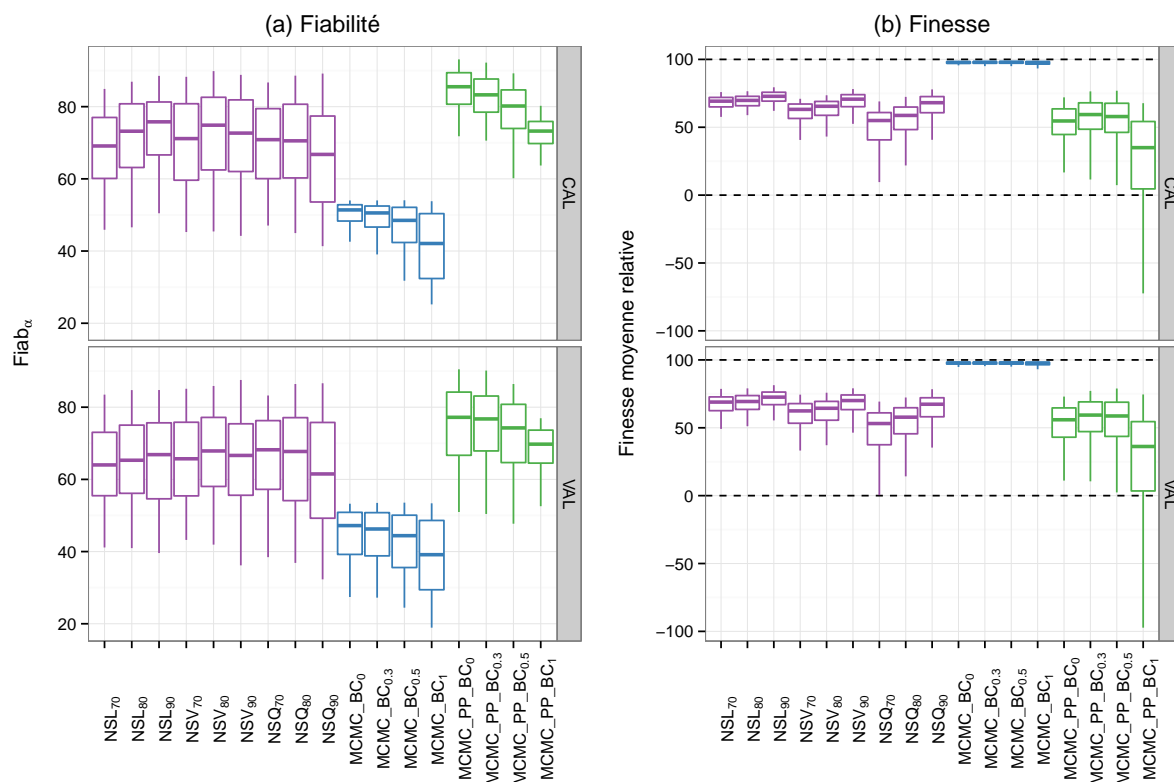


FIGURE 6.1 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices (a) de fiabilité et (b) de finesse des distributions prédictives obtenues avec la méthode GLUE (neuf premiers boxplots) et le calage bayésien (MCMC), pour deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans sur 195 bassins.



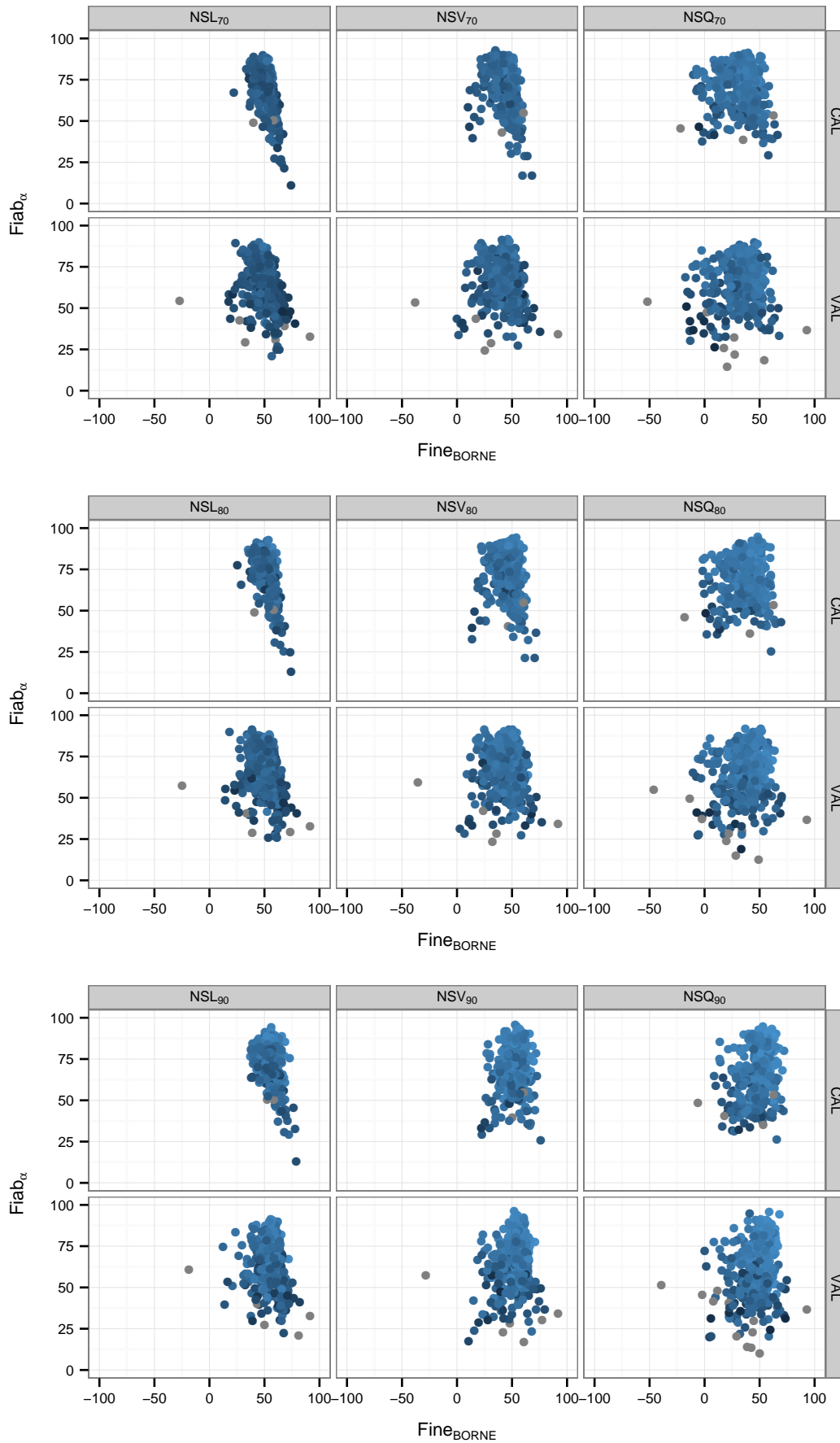


FIGURE 6.2 – Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec la méthode GLUE pour différents seuils d’efficacité (de haut en bas, correspondant à 0,7, 0,8 et 0,9 fois la valeur maximale), pour deux périodes de calage (haut), et deux périodes d’évaluation (bas) de cinq ans.

reur s'ajoute à l'incertitude liée à l'estimation des paramètres (expériences MCMC\_PP\_BC<sub>0</sub>–MCMC\_PP\_BC<sub>1</sub>) Les indices de fiabilité sont alors nettement plus élevés, supérieurs à ceux obtenus avec la méthode GLUE, tandis que les indices de finesse indiquent que la largeur moyenne des distributions prédictives a en général fortement augmenté.

Les résultats obtenus avec le calage bayésien varient suivant la transformation de données utilisée. Cette sensibilité reflète d'une part la sensibilité du calage des paramètres hydrologiques au critère de performance, et d'autre part la sensibilité du modèle d'erreur à la transformation des données. L'estimateur des moindres carrés présente la particularité d'obtenir les distributions prédictives les plus fines lorsque seule l'incertitude liée à l'estimation des paramètres est propagée, mais les plus larges lorsque l'incertitude totale est considérée. Cela semble confirmer l'intérêt d'utiliser une transformation de données pour tenir compte de l'hétéroscédasticité des erreurs.

### 6.3.2 Méthodes de post-traitement

Nous examinons à présent les résultats obtenus avec les approches qui utilisent la comparaison entre les débits simulés par le modèle GR4J et les débits observés. Le modèle GR4J a été calé avec le critère de performance NSV.

La Figure 6.4 présente les indices de fiabilité et de finesse obtenus avec les neuf méthodes de post-traitement présentées au chapitre précédent, et la Figure 6.5 illustre les covariations de ces deux indices.

Les quatre premières méthodes, BC<sub>0</sub>, BC<sub>0.3</sub>, BC<sub>0.5</sub> et BC<sub>1</sub>, utilisent les mêmes modèles d'erreur que les quatre variantes du calage bayésien. Nous constatons que des résultats similaires sont obtenus pour les méthodes BC<sub>0.5</sub> et BC<sub>1</sub>, mais que des différences significatives existent pour les méthodes BC<sub>0</sub>, BC<sub>0.3</sub>. Ces différences s'expliquent par :

- le fait que les paramètres optimaux du modèle hydrologique ne sont pas identiques – sauf pour la variante BC<sub>0.5</sub> qui correspond à un critère NSV –, puisque le calage bayésien ne dissocie pas critère de calage et modèle d'erreur ;
- la contribution de l'incertitude liée à l'estimation des paramètres – même limitée – à l'incertitude prédictive.

Les quatre méthodes MeGa, MCP, RQ et CemaM présentent des indices de fiabilité élevés en calage, mais bien plus faibles et variables sur les périodes de contrôle. Ces diminutions ne concernent pas uniquement ces quatre méthodes – nous pouvions déjà noter une diminution générale des indices de fiabilité pour la méthode GLUE et le calage bayésien –, mais

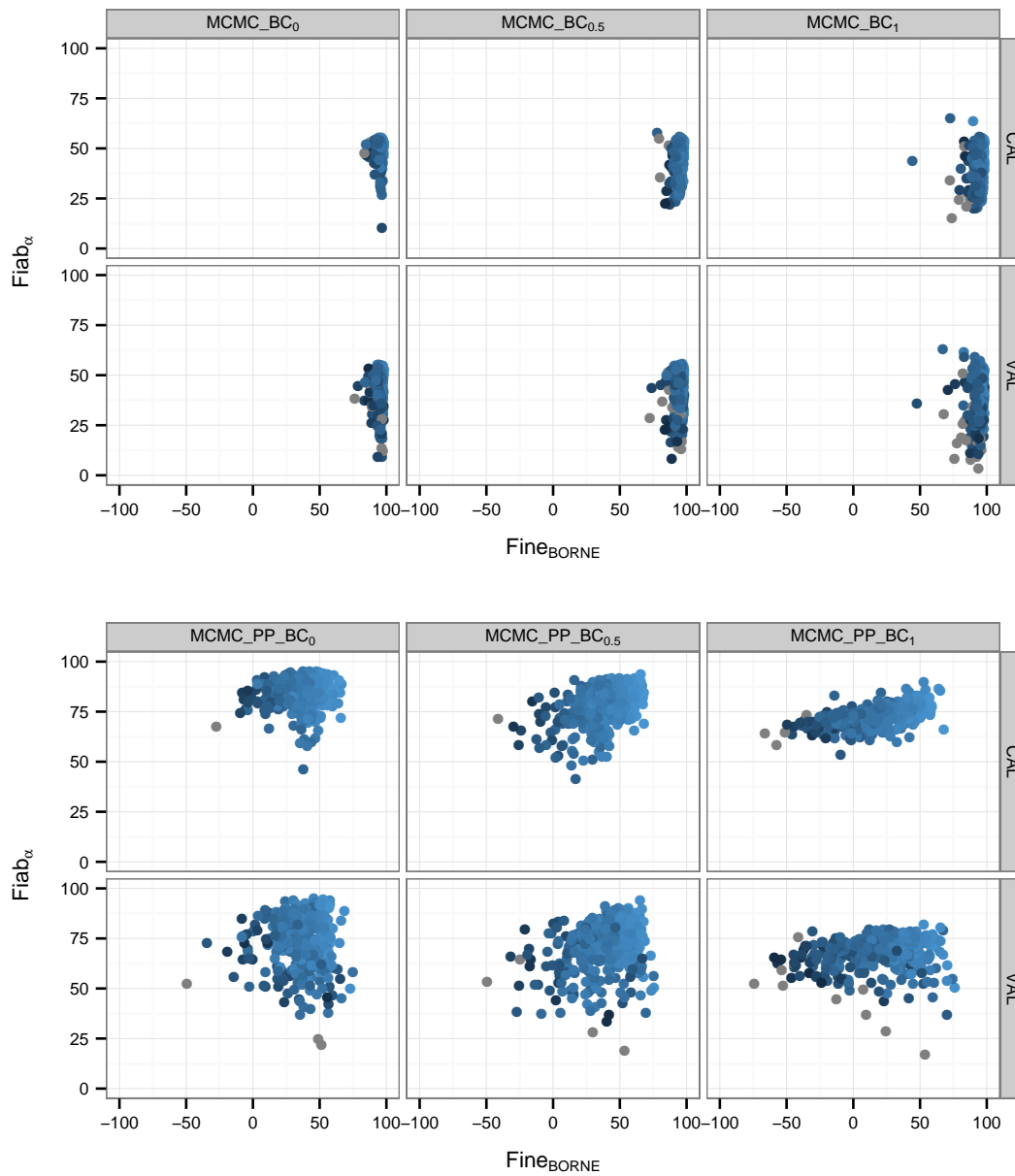


FIGURE 6.3 – Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec le calage bayésien (MCMC), pour deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans. Les indices obtenus avec la transformation BC<sub>0.3</sub> sont similaires à ceux obtenus avec la transformation BC<sub>0.5</sub>, et ne sont pas présentés pour faciliter la visualisation.

apparaissent plus nettement en raison des valeurs élevées obtenues en calage. Cela indique que les caractéristiques des erreurs ne sont pas totalement (robustement) extrapolables entre la période de calage et la période de validation (contrôle).

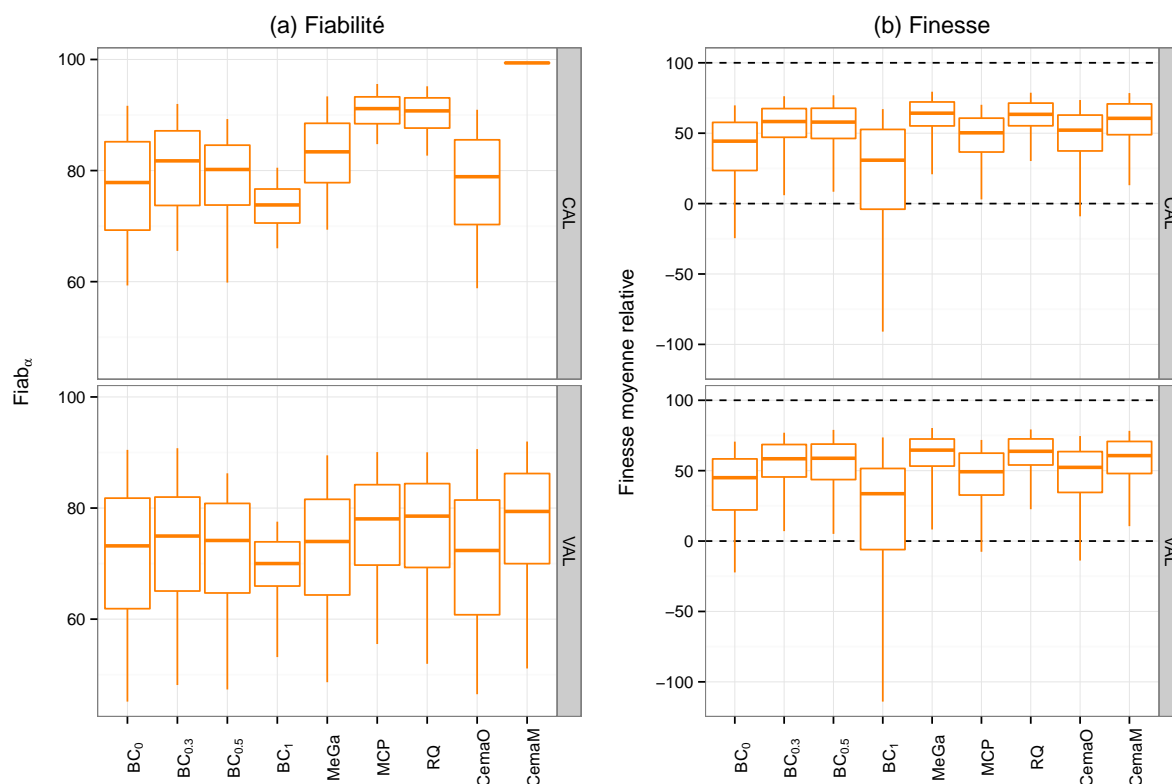


FIGURE 6.4 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices (a) de fiabilité et (b) de finesse des distributions prédictives obtenues avec neuf méthodes de post-traitement statistique, pour deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans sur 195 bassins.

Le cas de la méthode CemaM est particulièrement intéressant. Le caractère empirique de la méthode permet par construction d'obtenir des indices de fiabilité égaux à 1 pendant la période de calage. Les autres méthodes sont, quant à elles, limitées par leur formalisme statistique, et parviennent avec plus ou moins de succès à refléter les caractéristiques des erreurs. A cet égard, nous pouvons remarquer que l'indice de fiabilité utilisé ici – la moyenne des scores mesurant les déviations du diagramme de PIT calculés séparément pour les bas et hauts débits simulés –, permet de pénaliser à juste titre la méthode CemaO, qui ne tient pas compte de la variabilité des distributions des erreurs relatives selon le débit simulé, mais qui obtiendrait néanmoins des indices de fiabilité égaux à 1 si le diagramme de PIT était construit avec l'ensemble des données.

Ainsi, le net avantage apporté par la flexibilité de la méthode CemaM semble disparaître sur les périodes indépendantes de contrôle, et des indices de fiabilité semblables sont alors

obtenus avec les méthodes MCP, RQ et CemaM. Il n'en reste pas moins que la méthode CemaM reste à de nombreux égards plus simple.

En terme de finesse, les résultats indiquent que les méthodes  $BC_{0.3}$ ,  $BC_{0.5}$  MeGa, RQ et CemaM produisent en général les distributions les plus fines, tandis que la méthode  $BC_1$  se distingue en présentant des indices de finesse très variables et négatifs dans plus de 25% des cas.

### 6.3.3 Approche multi-modèles

Nous nous intéressons à présent aux résultats obtenus avec l'approche multi-modèles. Pour rappel, les dix-sept modèles utilisés (voir [Velazquez et al., 2010](#)) ont été calés avec le critère de performance NSV, et une loi gamma a été estimée à chaque pas de temps par la méthode des moments.

La Figure 6.6 présente les indices de fiabilité et de finesse obtenus, et la Figure 6.7 illustre les covariations de ces deux indices.

Nous constatons que les indices de fiabilité et de finesse sont assez semblables à ceux obtenus avec les méthodes de post-traitement les plus satisfaisantes, à savoir la méthode RQ et la méthode CemaM. Les indices de fiabilité semblent légèrement plus faibles, et les indices de finesse légèrement plus forts.

Il est particulièrement intéressant de noter une diminution des indices de fiabilité pour les périodes de contrôle par rapport aux périodes de calage. Ce résultat peut sembler contre-intuitif dans la mesure où la variabilité des distributions prédictives ne dépend que de la variabilité des simulations obtenues avec les différents modèles, et ne fait pas l'objet d'un apprentissage statistique sur une période donnée. On remarque d'ailleurs que les indices de finesse diminuent sur les périodes de contrôle, ce qui indique que la variabilité augmente ; c'est la conséquence du passage d'une période de calage, où les simulations sont contraintes, à une période indépendante de contrôle, où les simulations ne sont plus contraintes. Ces résultats suggèrent que la diminution des indices de fiabilité est principalement attribuable aux pertes de performance des modèles pluie-débit utilisés pour créer l'ensemble prédictif.

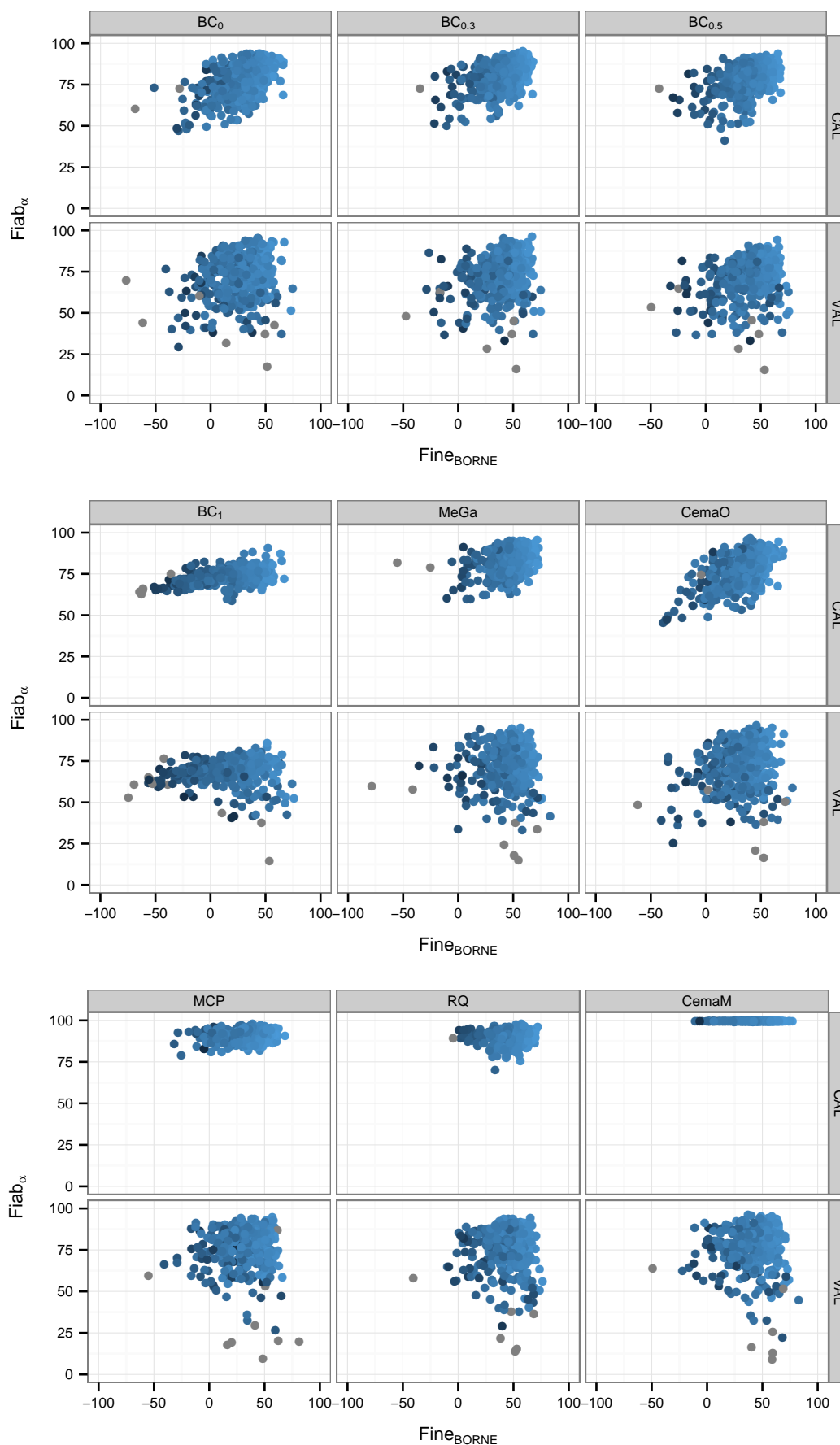


FIGURE 6.5 – Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec neuf méthodes de post-traitement statistique, pour deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans.

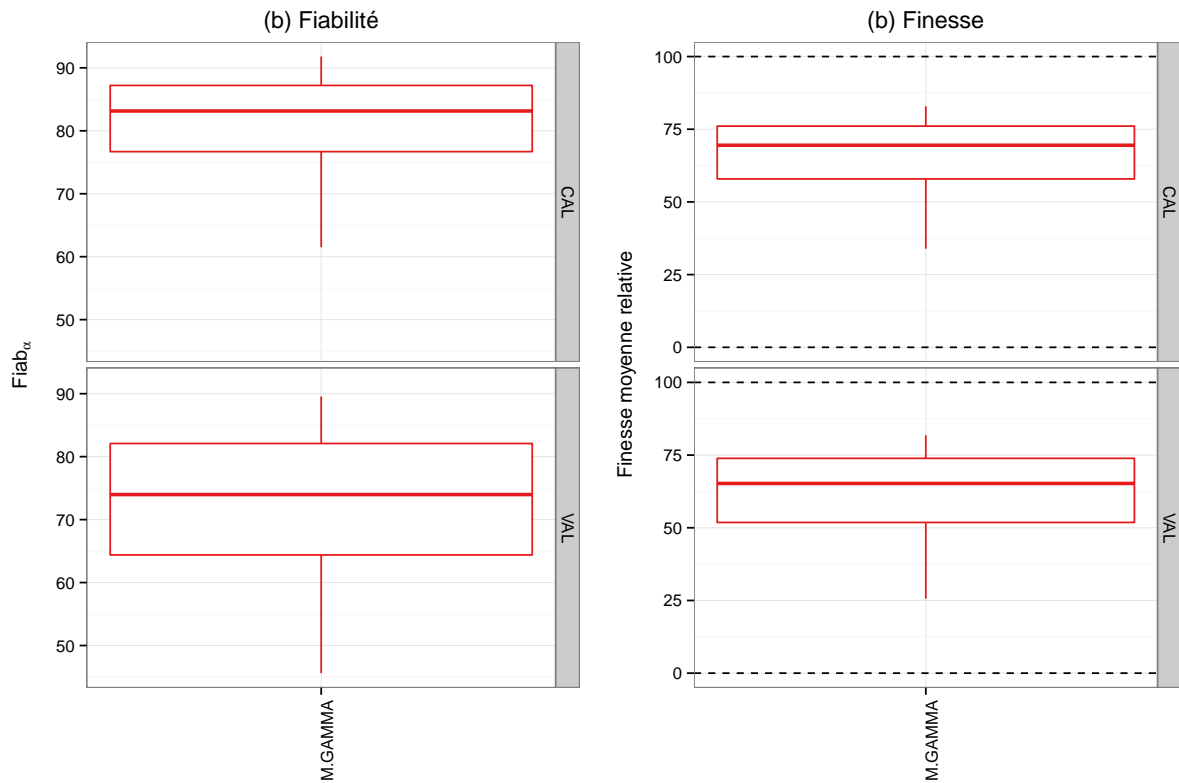


FIGURE 6.6 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices (a) de fiabilité et (b) de finesse des distributions prédictives obtenues avec l’approche multi-modèles, pour deux périodes de calage (haut), et deux périodes d’évaluation (bas) de cinq ans sur 195 bassins.

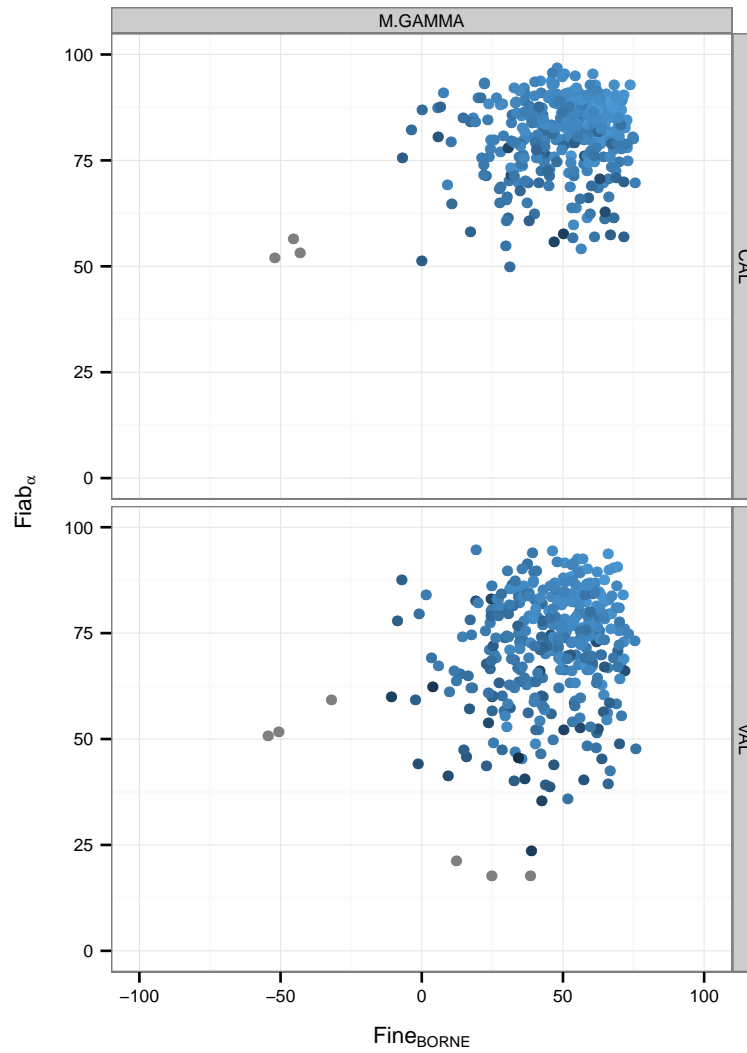


FIGURE 6.7 – Diagrammes en nuage de points entre les indices de fiabilité et de finesse obtenus avec l’approche multi-modèles pour deux périodes de calage (haut), et deux périodes d’évaluation (bas) de cinq ans.



## 6.4 Analyse des performances globales

Nous abordons à présent l'analyse des performances globales des méthodes de quantification de l'incertitude prédictive. La performance globale est quantifiée par le score de compétence CRPSS. Nous cherchons à mettre en évidence les méthodes les plus performantes.

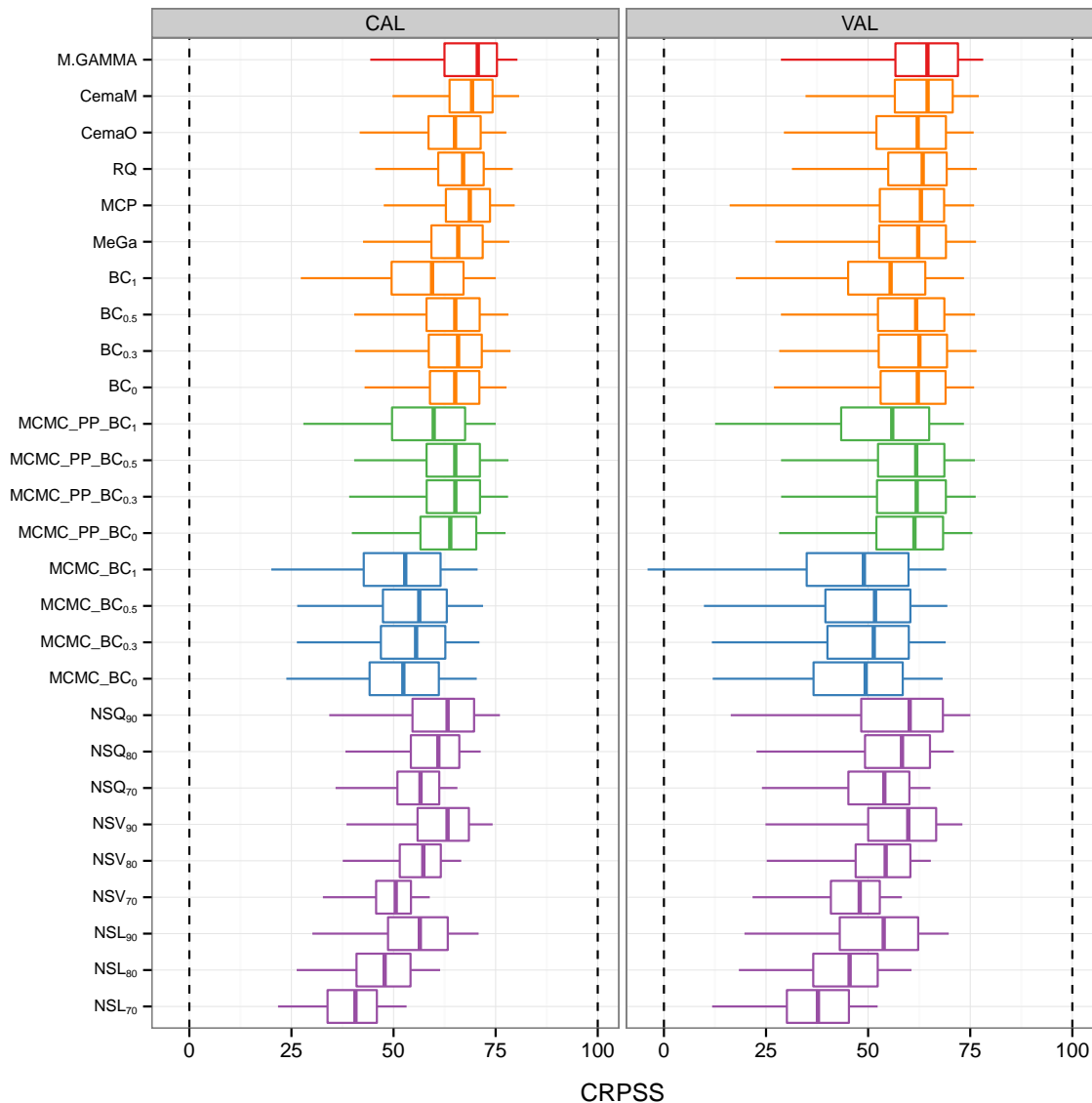


FIGURE 6.8 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices de performance globale obtenus par l'ensemble des méthodes testées, pour deux périodes de calage (gauche), et deux périodes d'évaluation (droite) de cinq ans sur 195 bassins.

La Figure 6.8 présente les boxplots des indices de performance obtenus avec les vingt-sept méthodes testées, sur les périodes de calage et de contrôle. Nous pouvons remarquer que :

- Les méthodes les moins performantes sont celles liées au formalisme GLUE ; seules les

variantes  $NSQ_{70}$ ,  $NSQ_{80}$  et  $NSQ_{90}$  obtiennent des performances comparables à celles obtenues par un calage bayésien.

- L'utilisation d'une transformation de variable permet d'améliorer significativement les performances des méthodes liées au calage bayésien, et celles des méthodes de référence de post-traitement ( $BC_0$ ,  $BC_{0.3}$ ,  $BC_{0.5}$ ) ; la prise en compte de l'hétéroscédasticité des erreurs est nécessaire, et les exposants 0.3 et 0.5 semblent produire les meilleurs résultats.
- Les performances des méthodes bayésiennes ne sont généralement pas dues à l'incertitude liée à l'estimation des paramètres, puisque l'on observe que peu de différence entre ces méthodes et les méthodes de référence de post-traitement ( $BC_0$ ,  $BC_{0.3}$ ,  $BC_{0.5}$ ).
- Parmi les méthodes de post-traitement, les méthodes CemaM et RQ semblent se distinguer sur les périodes de contrôle. Ces deux méthodes sont les plus flexibles, puisqu'elles estiment directement des quantiles prédictifs. La méthode RQ suit néanmoins un certain formalisme statistique (les quantiles évoluent de manière linéaire dans un espace transformé), ce qui peut expliquer que ses performances soient légèrement moins bonnes que celles obtenues avec la méthode CemaM.
- L'approche multi-modèles obtient des performances comparables à celles de la méthode CemaM. L'approche gagne en finesse ce qu'elle perd en fiabilité.

La Figure 6.9 présente une comparaison des indices de performance obtenus par les différentes méthodes. Pour faciliter la visualisation, une version bornée du CRPSS est utilisée, et deux méthodes ( $NSQ_{80}$ ,  $NSL_{80}$ ) ne sont pas présentées. La méthode CemaM a été choisie comme approche de référence. Les résultats suggèrent que pour certains bassins, aucune méthode n'est capable de produire des résultats satisfaisants, et qu'il existe une forte corrélation entre les performances obtenues par les différentes méthodes.

Nous avons calculé les coefficients de corrélation des rangs de Spearman entre les couples de performances obtenues avec la méthode CemaM et les autres méthodes ; ils sont supérieurs à 0.8 dans 22 cas sur 26 – seules les méthodes  $NSL_{70}$ ,  $NSL_{80}$ ,  $NSL_{90}$  et  $NSV_{70}$  présentent de plus faibles coefficients. Ces résultats suggèrent qu'il est difficile d'identifier certaines méthodes comme plus adaptées que d'autres pour certains bassins versants. Il semble au contraire que les mêmes difficultés sont partagées par les différentes méthodes.

## 6.5 Synthèse

Différentes approches existent pour quantifier l'incertitude prédictive d'un modèle hydrologique. Nous avons souhaité dans cette thèse réaliser une comparaison de plusieurs méthodes.

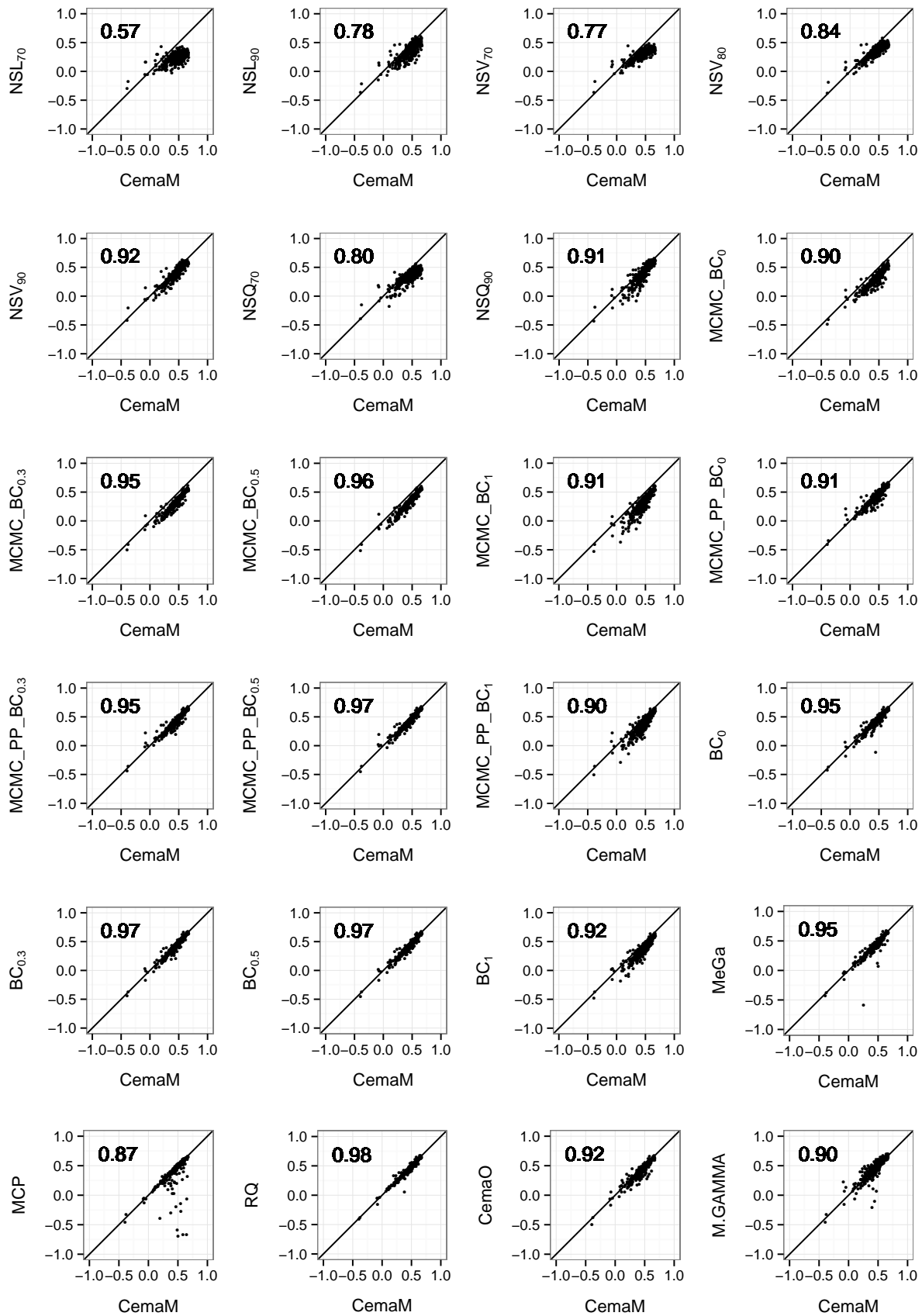


FIGURE 6.9 – Indices de performance globale CRPSS, pour 195 bassins et deux périodes d'évaluation de cinq ans. Une version bornée est utilisée pour faciliter la visualisation.

La comparaison de vingt-sept méthodes a été réalisée sur 195 bassins versants selon une procédure de calage-contrôle. Les caractéristiques de fiabilité, de finesse et la performance globale des distributions prédictives ont été examinées.

Les principales conclusions de cette comparaison sont les suivantes :

- L'incertitude liée à l'estimation des paramètres ne semble pas suffisante pour caractériser l'incertitude prédictive totale ; l'hypothèse selon laquelle les ensembles de paramètres pourraient refléter de manière implicite les erreurs résiduelles ne semble pas vérifiée sur notre échantillon de bassins versants. Les méthodes liées au formalisme GLUE sont globalement les moins satisfaisantes.
- Les résultats obtenus avec le calage bayésien montrent la faible contribution de l'incertitude liée à l'estimation des paramètres à l'incertitude totale. L'incertitude totale est en grande partie contrôlée par le modèle d'erreur qui est associé à la fonction de vraisemblance. L'utilisation d'une transformation de variable permet d'améliorer significativement les performances en raison d'une meilleure prise en compte de l'hétéroscédasticité des erreurs.
- Les méthodes de post-traitement que nous avons appliquées au cours de cette thèse ont pour objectif de caractériser les distributions prédictives de manière conditionnelle au débit simulé par un modèle hydrologique. Elles utilisent différents formalismes statistiques pour caractériser ces distributions, ce qui conduit à des résultats qui diffèrent en terme d'indices de fiabilité, de finesse, et de performance globale. Il apparaît que les méthodes les plus flexibles obtiennent des indices de fiabilité plus élevés que les méthodes qui reposent sur de fortes hypothèses statistiques, malgré des diminutions importantes lors du passage des périodes de calage aux périodes de contrôle, et de meilleures performances globales.
- Enfin, l'approche multi-modèles obtient des performances comparables à celles obtenues avec les méthodes de post-traitement les plus satisfaisantes ; les distributions prédictives sont néanmoins, en général, moins fiables.



# 7

## **TRANSFERRING MODEL UNCERTAINTY ESTIMATES FROM GAUGED TO UNGAUGED CATCHMENTS**



Ce chapitre correspond à un article soumis à la revue *Hydrology and Earth System Sciences*.

### **Transferring model uncertainty estimates from gauged to ungauged catchments**

F. Bourgin<sup>1</sup>, V. Andréassian<sup>1</sup>, C. Perrin<sup>1</sup>, L. Oudin<sup>2</sup>

(1) Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, F-92761 Antony Cedex, France

(2) UPMC Université Paris 06, UMR 7619 Metis, Case 105, 4 place Jussieu, F-75005 Paris, France

## **Résumé**

La simulation des débits pour les bassins non-jaugés représente un véritable défi, et la quantification de l'incertitude prédictive dans ce contexte est un tâche délicate. Dans cet article, nous présentons une méthode qui permet de transférer une estimation de l'incertitude prédictive des bassins jaugés aux bassins non jaugés. Nous avons testé cette approche sur un échantillon de 907 bassins versants français, en évaluant trois qualités de l'incertitude prédictive: la fiabilité, la finesse et la performance globale. Les résultats obtenus montrent que la méthode est capable de produire, dans la plupart des cas, des intervalles prédictifs à la fois fiables et fins à l'exutoire des bassins versants non-jaugés.

## **Abstract**

Predicting streamflow hydrographs in ungauged catchments is a challenging issue, and accompanying the estimates with realistic uncertainty bounds is an even more complex task. In this paper, we present a method to transfer model uncertainty estimates from gauged to ungauged catchments and we test it over a set of 907 catchments located in France. We evaluate the quality of the uncertainty estimates based on three expected qualities: reliability, sharpness, and overall skill. Our results show that the method holds interesting perspectives, providing in most cases reliable and sharp uncertainty bounds at ungauged locations.



## 7.1 Introduction

### 7.1.1 Predicting streamflow in ungauged catchments with uncertainty estimates

Predicting the entire runoff hydrograph in ungauged catchments is a challenging issue that has attracted much attention during the last decade. In this context, the use of suitable conceptual rainfall-runoff models has proved to be useful, and because traditional calibration approaches based on observed discharge data cannot be applied in ungauged catchments, other approaches are required. Various methods have been proposed for the estimation of rainfall-runoff model parameters in ungauged catchments, as reported by the recent synthesis of the Prediction in Ungauged Basins (PUB) decade (Blöschl *et al.*, 2013; Hrachowitz *et al.*, 2013; Parajka *et al.*, 2013).

The estimation of predictive uncertainty is deemed good practice in any environmental modelling activity (Refsgaard *et al.*, 2007). In hydrological modelling, the topic has been widely discussed for years, and there is still no general agreement about how to adequately quantify uncertainty. In practice, various methodologies are currently used.

For gauged catchments, the methodologies include Bayesian calibration and prediction approaches (see e.g., the review of Liu and Gupta, 2007), informal methods related to the GLUE framework (Beven and Freer, 2001), multi-model approaches (Duan *et al.*, 2007; Velazquez *et al.*, 2010) and other total uncertainty quantification methods (Montanari and Brath, 2004; Solomatine and Shrestha, 2009; Weerts *et al.*, 2011; Ewen and O'Donnell, 2012). A comprehensive review of the topic can be found in Matott *et al.* (2009) and Montanari (2011).

While many methods have been proposed for gauged catchments, only a few have been proposed for the estimation of predictive uncertainty on ungauged catchments. McIntyre *et al.* (2005) presented a GLUE-type approach consisting of transferring ensembles of parameter sets obtained on donor (gauged) catchments to target (ungauged) catchments. More recently, a framework based on constrained parameter sets was applied in several studies (Yadav *et al.*, 2007; Zhang *et al.*, 2008; Winsemius *et al.*, 2009; Bulygina *et al.*, 2011, 2012; Kapangaziwiri *et al.*, 2012). It is a two-step procedure. The first step consists in estimating with uncertainty various summary metrics of the hydrographs, also called “signatures” of the catchments, or gathering other “soft” or “hard” information at the target ungauged catchment. The second step is the selection of an ensemble of model parameter sets. “Acceptable” or “behavioural” parameter sets are those that yield sufficiently close simulated summary

metrics compared to regionalized metrics. The reader can refer to [Wagener and Montanari \(2011\)](#) for a comprehensive description of this framework.

One difficulty of the above mentioned approaches lies in the interpretation of the uncertainty bounds obtained from the parameter ensemble predictions. As noted by [McIntyre et al. \(2005\)](#) and [Winsemius et al. \(2009\)](#), the uncertainty bounds cannot easily be interpreted as confidence intervals, and thus it remains difficult to use them in practice. In addition, solely relying on an ensemble of model parameter sets to quantify total predictive uncertainty is often not sufficient when imperfect rainfall-runoff models are used.

A pragmatic alternative consists in addressing separately the parameter estimation and the uncertainty estimation issues. It has been argued by several authors ([Montanari and Brath, 2004](#); [Andréassian et al., 2007](#); [Ewen and O'Donnell, 2012](#)) that a posteriori characterization of modelling errors of a “best” or “optimal” simulation can yield valid uncertainty bounds at gauged locations. As stated by [Solomatine and Shrestha \(2009\)](#),

The historical model residuals (errors) between the model prediction  $\hat{y}$  and the observed data  $y$  are the best available quantitative indicators of the discrepancy between the model and the real-world system or process, and they provide valuable information that can be used to assess the predictive uncertainty.

Similarly, one could argue that the model residuals between the model prediction and the observed data at *neighbouring gauged locations* are, perhaps, the best available indicators of the discrepancy between the model and the real-world system at *the target ungauged location*.

The only attempt we are aware of to apply a total uncertainty estimation approach at ungauged location is the one presented by [Roscoe et al. \(2012\)](#). They quantified uncertainty for river stage prediction at ungauged locations by first interpolating the residual errors at ungauged locations, and then applying quantile regression to these errors.

### 7.1.2 Scope of the paper

The aim of this paper is to provide an estimation of the total uncertainty affecting runoff prediction at ungauged locations when a rainfall-runoff model and a regionalisation scheme is used.

To our knowledge, a framework based on residual errors and total uncertainty quantification has not yet been extensively tested in the context of prediction in ungauged catchments.

This paper contributes to the search for methods able to provide uncertainty estimates when runoff predictions in ungauged catchments are sought.

## 7.2 Data and methods

Our objective is not to develop a new parameter regionalisation approach. Therefore, we purposely chose to use ready-to-use materials and methods and only focus on the uncertainty quantification issue. This study can be considered as a follow-up of the work made by [Oudin \*et al.\* \(2008\)](#) on the comparison of regionalisation approaches. We only provide here an overview of the data set, the rainfall-runoff models and the parameter calibration and regionalisation approach, since the details can be found in [Oudin \*et al.\* \(2008\)](#).

### 7.2.1 Data set

A database of 907 French catchments was used. They represent various hydrological conditions, given the variability in climate, topography, and geology in France. This set includes fast responding Mediterranean catchments with intense precipitation as well as larger, groundwater-dominated catchments. Some characteristics of the data set are given in Table 7.1. Catchments were selected to have limited snow influence, since no snowmelt module was used in the hydrological modelling. Daily rainfall, runoff, and potential evapotranspiration (PE) data series over the 1995–2005 period were available. Meteorological inputs originate from Météo-France SAFRAN reanalysis ([Vidal \*et al.\*, 2010](#)). PE was estimated using the temperature-based formula proposed by [Oudin \*et al.\* \(2005\)](#). Hydrological data were extracted from the HYDRO national archive ([www.hydro.eaufrance.fr](http://www.hydro.eaufrance.fr)).

Table 7.1: Characteristics of the 907 catchments. P–precipitation, PE–potential evapotranspiration, Q–discharge.

	Percentiles				
	0.05	0.25	0.50	0.75	0.95
Catchment area (km <sup>2</sup> )	27	73	149	356	1788
Mean annual precipitation (mm/y)	753	853	978	1176	1665
Mean annual potential evapotranspiration (mm/y)	549	631	659	700	772
Mean annual runoff (mm/y)	133	233	344	526	1041
Q/P ratio	0.17	0.27	0.34	0.45	0.68
P/PE ratio	1.06	1.25	1.47	1.83	2.9
Median elevation (m)	76	149	314	645	1183

## 7.2.2 Rainfall-runoff models

Two daily, continuous lumped rainfall-runoff models were used:

- The GR4J rainfall-runoff model, an efficient and parsimonious daily lumped continuous rainfall-runoff model described by Perrin *et al.* (2003).
- The TOPMO rainfall-runoff model, inspired by TOPMODEL (Beven and Kirkby, 1979). This version was tested on large data sets and showed performance comparable to that of the GR4J model, while being quite different (Michel *et al.*, 2003; Oudin *et al.*, 2008, 2010).

Using these two models rather than a single one makes it possible to draw more general conclusions.

The GR4J and TOPMO models have four and six free parameters respectively. On gauged catchments, parameter estimation is performed using a local gradient search procedure, applied in combination with a pre-screening of the parameter space (Mathevet, 2005; Perrin *et al.*, 2008). This optimization procedure has proved to be efficient in past applications for the conceptual models used here. As objective function, we used the Nash and Sutcliffe (1970) criterion computed on root square transformed flows. This criterion was shown to yield a good compromise between different objectives (Oudin *et al.*, 2006).

## 7.2.3 Regionalisation approach

By definition, no discharge data is available for calibrating parameter sets at ungauged location. Thus, other strategies are needed to estimate the parameters of hydrological models for prediction in ungauged catchments.

Oudin *et al.* (2008) assessed the relative performance of three classical regionalisation schemes over a set of French catchments: spatial proximity, physical similarity and regression. Several options within each regionalisation approach were tested and compared. Based on their results, the following choices were made here for the regionalisation approach, as they offered the best regionalisation solution:

- Poorly modelled catchments were discarded as potential donors: only catchments with a performance criterion in calibration above 0.7 were used as possible donors.
- The spatial proximity approach was used. It consists of transferring parameter sets from neighbouring catchments to the target ungauged catchment. Proximity of the ungauged catchments to the gauged ones was quantified by the distances between catchments centroids.

- The output averaging option was chosen. It consists of computing the mean of the streamflow simulations obtained on the ungauged catchment with the set of parameters of the donor catchments.
- The number of neighbours was set to 4 and 7 catchments for GR4J and TOPMO respectively.

### 7.3 Proposed approach: transfer of relative errors by flow groups

Transferring calibrated model parameters from gauged catchments to ungauged catchment is a well established approach when parameters cannot be inferred from available data. The method presented here extends the parameter transfer approach to the domain of uncertainty estimation.

The main idea underlying the proposed approach is (i) to treat each donor as if it was ungauged (simulating flow through the above described regionalisation approach), (ii) characterize the empirical distribution of relative errors for each of these donors, and (iii) transfer model uncertainty estimates to the ungauged catchment.

The methodology used to transfer model uncertainty estimates can be described by the following steps, illustrated by Figs. 7.1 to 7.5:

#### 1. Selection of catchments

Here we consider a target catchment as ungauged, called TUC. This catchment has  $n$  neighbouring gauged catchments, called  $NGC_1, NGC_2, \dots, NGC_n$ . If the  $NGC_i$  catchment was now considered ungauged, one could also consider  $n$  neighbouring catchments, called  $NGC_i^1, NGC_i^2, \dots, NGC_i^n$ . Obviously, the TUC catchment would be excluded from this set of second order donor catchments.

#### 2. Application of the parameter regionalisation scheme to the donor catchments $NGC_i$

- a. Apply the parameter regionalisation scheme to obtain a simulated discharge time series for each  $NGC_i$  using neighbours  $NGC_i^j$ .
- b. Compute the relative errors of streamflow reconstitution, and create 10 groups of relative errors according to the magnitude of the simulated discharge. The groups are based on the quantiles of the simulated discharges, so that each group is equally populated. The subdivision into flow groups allows accounting for the heteroscedasticity of model errors.

#### 3. Computation of the multiplicative coefficients

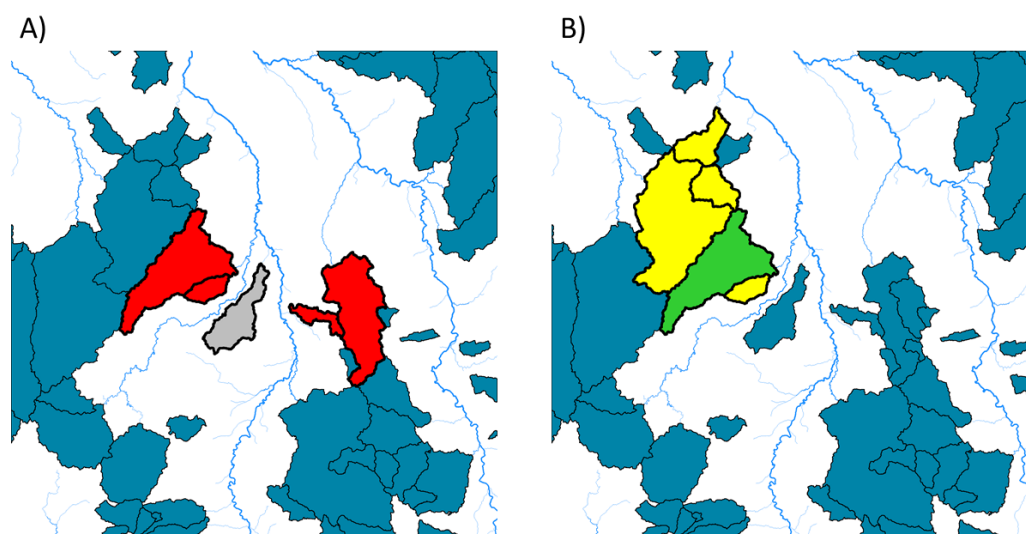


Figure 7.1: Illustration of the proposed approach – Step 1: in (A), a target catchment (grey) is considered as ungauged; this catchment has  $n$  neighbouring gauged catchments (red). In (B), if one of the neighbouring catchment is now considered ungauged (green), we also consider  $n$  neighbouring catchments (yellow). Note that the target catchment is excluded from this set of second order donor catchments.

- a. Put together the relative errors from the donors according to the group they belong to.
  - b. Compute the empirical quantiles of the relative errors distribution within each group. Each quantile of relative error can be considered a multiplicative coefficient. These multiplicative coefficients will be used to obtain the prediction bounds.
4. Computation of the uncertainty bounds for the target catchment TUC
- a. Apply the parameter regionalisation scheme to obtain a simulated discharge time series for TUC using the parameter sets of the neighbouring catchments  $NGC_i$ .
  - b. Multiply the simulated discharge by the set of multiplicative coefficients obtained at Step 3b to obtain the uncertainty bounds.

Some of the methodological choices made here will be further discussed in Sect. 7.5.

## 7.4 Quantitative evaluation of uncertainty bounds

We assessed the relevance of the 90% uncertainty bounds by focusing on three characteristics: reliability, sharpness and overall skill.

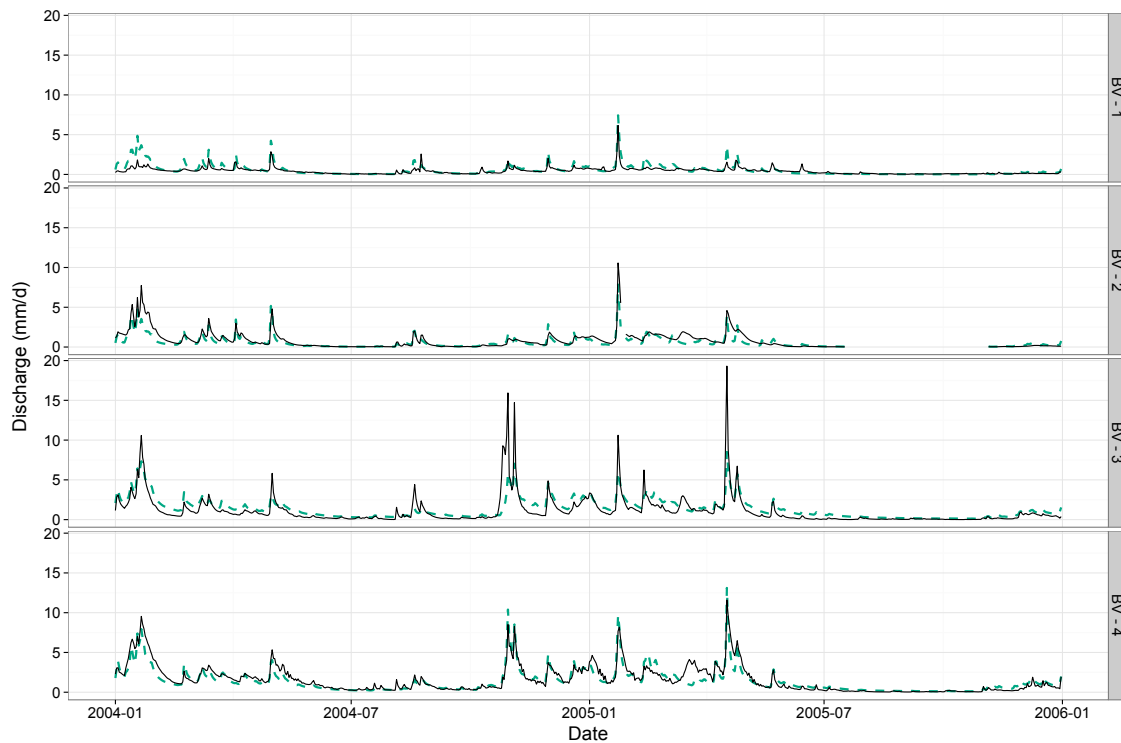


Figure 7.2: Illustration of the proposed approach – Step 2a: simulated (green) and observed (black) discharge time series for four donor catchments treated as ungauged, i.e., in which model parameters must be estimated from a regionalisation approach.

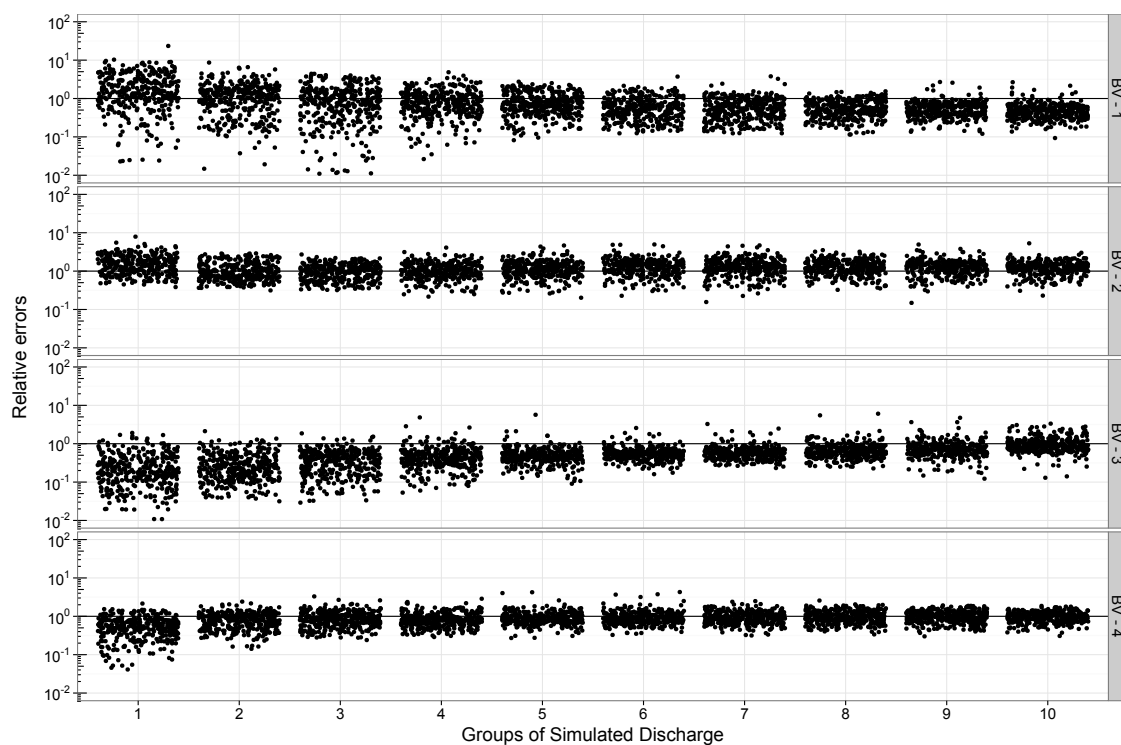


Figure 7.3: Illustration of the proposed approach – Step 2b: relative errors by flow groups; groups of relative errors are defined according to the magnitude of the simulated discharge.

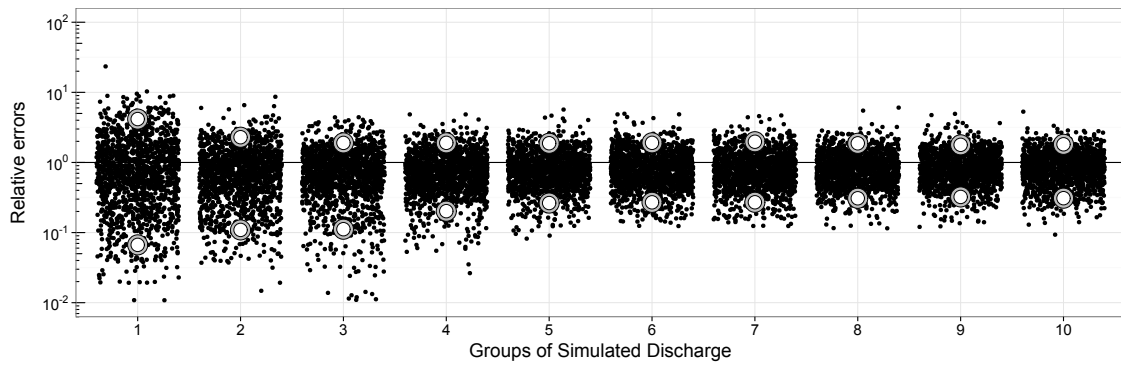


Figure 7.4: Illustration of the proposed approach – Steps 3a and b: aggregating the relative errors observed at the donors catchments; red dots correspond to the empirical quantiles (5 and 95%) of the relative errors distribution within each group.

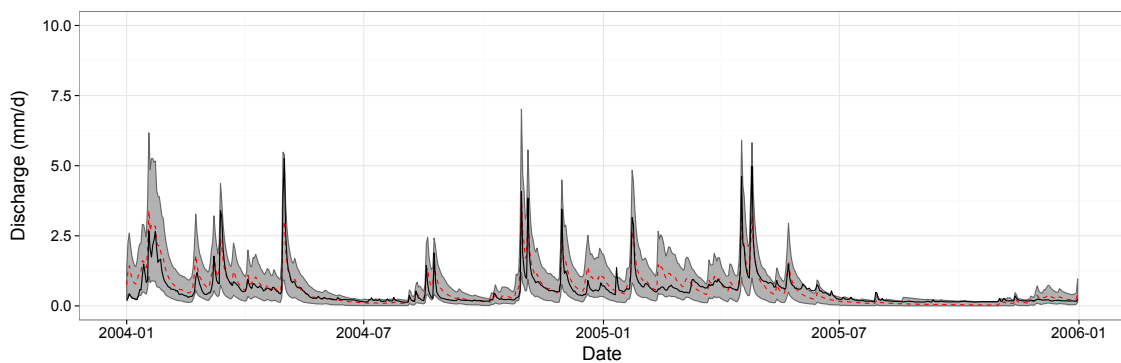


Figure 7.5: Illustration of the proposed approach – Steps 4a and b: simulated (red) and observed (black) discharge time series for the ungauged catchments; 90% uncertainty bounds in grey.



Reliability refers to the statistical consistency of the uncertainty estimation with the observation, i.e., a 90% prediction interval is expected to contain approximately 90% of the observations if prediction errors are adequately characterized by the uncertainty estimation. To estimate the reliability, we used the coverage ratio (CR) index, computed as the percentage of observations contained in the prediction intervals.

Sharpness refers to the concentration of predictive uncertainty. We used a quantitative index based on the average width of the uncertainty bounds. To ease comparison between catchments, we used the width of the 90% intervals of historical flows [Q5, Q95], where Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve, as a benchmark. The ratio (R) between these two values provides information about the reduction of uncertainty obtained by the application of the rainfall-runoff and the methodology presented here, compared to the climatology. The value  $1 - R$  indicates the percentage of reduction of the average width. We call this criterion the average width index (AWI). It is positive if the average width is reduced, and negative otherwise.

Uncertainty bounds that are as sharp as possible and reasonably reliable are sought: indeed sharp intervals that would consistently miss the target would be misleading, while overly large intervals that would successfully cover the observations at the expense of sharpness would be of limited value for decision making.

To complete the assessment of the prediction bounds, we used the interval score (Gneiting and Raftery, 2007). The interval score (IS) accounts for both reliability and sharpness and provides an overall assessment of the quality of the prediction bounds. The scoring rule of the interval score is defined as

$$S = (u - l) + \frac{2}{1 - \beta}(l - q)\mathbf{1}\{q < l\} + \frac{2}{1 - \beta}(q - u)\mathbf{1}\{q > u\} \quad (7.1)$$

where  $[l, u]$  is the prediction interval and  $q$  is the observed value;  $\mathbf{1}\{x < y\}$  is the indicator function, equal to 1 if  $x < y$  and 0 otherwise, and  $\beta$  is equal to 90% since a 90% interval is sought here. IS is the average value of  $S$  over the time series.

To ease comparison between catchments and evaluate the skill of the prediction bounds, we used the unconditional climatology as a benchmark and computed the interval skill score

$$ISS = 1 - \frac{IS}{IS^{\text{clim}}} \quad (7.2)$$

where  $IS^{\text{clim}}$  is the interval score obtained with the 90% climatological interval [Q5, Q95] (Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve).

The skill score is positive when the prediction bounds are more skillful than the climatological interval, and negative otherwise.

## 7.5 Results and discussion

### 7.5.1 Reliability, sharpness and overall skill

Figure 7.6 shows the distributions of the three criteria used to evaluate the uncertainty bounds on the 907 catchments. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) are used to synthesize the variety of scores over the 907 catchments of the data set.

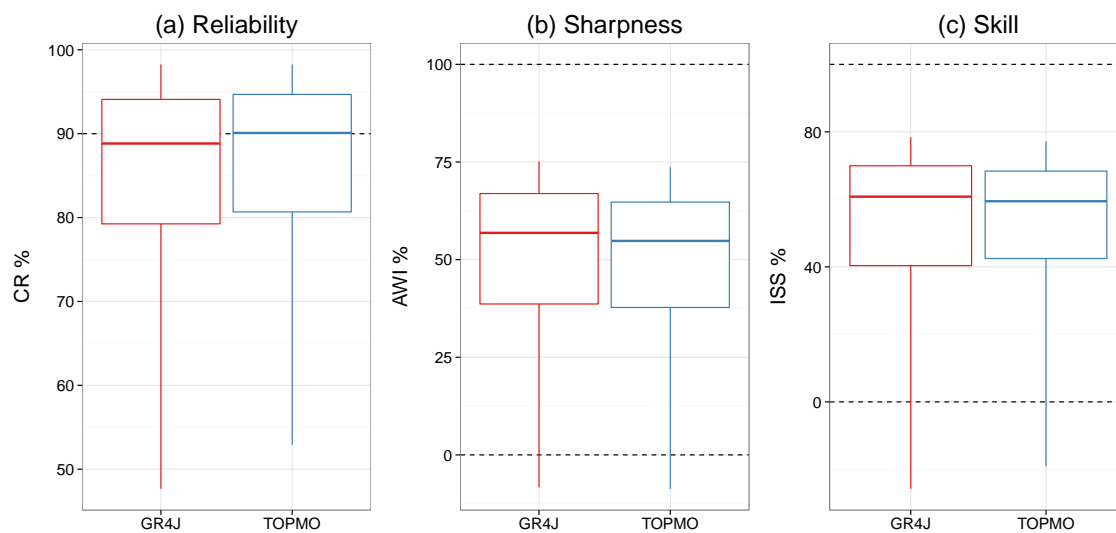


Figure 7.6: Distributions of the three performance criteria. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

#### Reliability

For both models, half of the catchments (from the lower quartile to the upper quartile) have CR values between 80 and 95%. The median values are equal to 89 and 90% for GR4J and TOPMO respectively. Since a value of 90% is expected for 90% prediction bounds, these results suggest that the prediction bounds are, in most cases, able to reflect the magnitude of errors when predicting runoff hydrographs in ungauged catchments. The CR values fall below 70% for around 14% of the catchments with GR4J, and 13% with TOPMO, which indicates cases where the proposed approach yields predictive bounds that might be too narrow or biased (i.e., not well centered on the observations).

## Sharpness

Regarding sharpness, it can be seen that for GR4J, half of the catchments (from the lower quartile to the upper quartile) have AWI values between 39 and 67%, while for TOPMO corresponding values are equal to 38 and 65%. The median values are equal to 57 and 55% for GR4J and TOPMO respectively. The higher the AWI values, the lower the predictive uncertainty is. Since it would be utopic to expect that no errors will be made when predicting runoff hydrographs for ungauged catchments, we considered here uncertainty reduction values between 30 and 80% as quite satisfactory. Note that negative values are seen for 7% of the catchments with both GR4J and TOPMO, which indicates cases where the approach yield prediction intervals whose average width is larger than the width of the historical [Q5, Q95] interval (Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve).

## Overall skill

Finally, Fig. 7.6(c) shows that the predictive skill is positive for most catchments (around 92%) for both models. For both models, half of the catchments (from the lower quartile to the upper quartile) have ISS values between 40 and 70%. The median values are equal to 61 and 59% for GR4J and TOPMO respectively. While it might be argued that the unconditional climatology is not a very challenging benchmark, we consider that it is still a positive and reassuring result.

### 7.5.2 Do we need to treat the donor catchments as ungauged?

A critical step of the proposed approach is to apply the regionalisation scheme to obtain a simulated discharge time series for each donor catchment (Step 2). This is done because we expect that predictive uncertainty at ungauged locations is larger than predictive uncertainty at gauged location, i.e., when the rainfall-runoff is calibrated with observed discharge data. To assess the impact of this methodological choice, we applied the methodology described earlier to transfer uncertainty estimates, but this time the donor catchments are treated as gauged.

Similarly to Fig. 7.6, Fig. 7.7 shows the distributions of the three criteria obtained in the two cases: whether or not the donor catchments are treated as ungauged. We can observe for both models a drop in reliability, whereas sharpness increases. This is because the relative errors are smaller when the donor catchments are treated as gauged, yielding narrower but

less reliable prediction bounds for the target catchment. Interestingly, this results in skill scores that are quite similar: improvements in terms of sharpness compensate decreases in terms of reliability.

Note that reliability is generally considered as a prevailing characteristic over sharpness. Therefore, the benefit of treating the donor catchments as ungauged clearly appears in Fig. 7.7a.

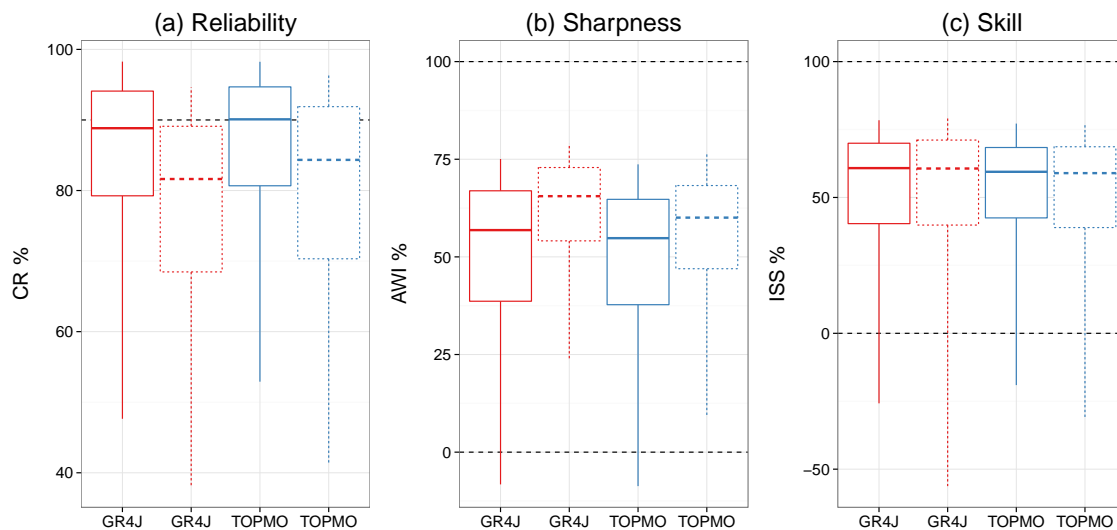


Figure 7.7: Distributions of the three performance criteria, obtained in two cases, i) when the donor catchments are treated as ungauged (continuous lines) and ii) when the donor catchments are treated as gauged (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

### 7.5.3 Do we need to use groups of relative errors?

Another critical step of the proposed approach is to use 10 groups of relative errors. The groups are defined according to the magnitude of the simulated discharge (Step 2). This was done to take into account the fact that the characteristics of errors usually change according to the magnitude of the simulated discharge. To assess the impact of this methodological choice, we again applied the methodology described earlier to transfer model uncertainty estimates, but this time only one group is used instead of 10.

Figure 7.8 shows the distributions of the three criteria obtained in the following two cases: whether 10 groups or only one group of relative errors are used. For both models, reliability slightly increase, whereas both sharpness and skill decrease. It appears that improvements in terms of reliability are not sufficient to compensate decreases in terms of sharpness when overall skill is considered.

While it could be argued that using only one group is the preferable option because of the slight improvement in terms of reliability, in our opinion, the improvement is not sufficiently important to compensate the decrease in terms of uncertainty reduction and skill. We definitely prefer to maintain different flow groups.

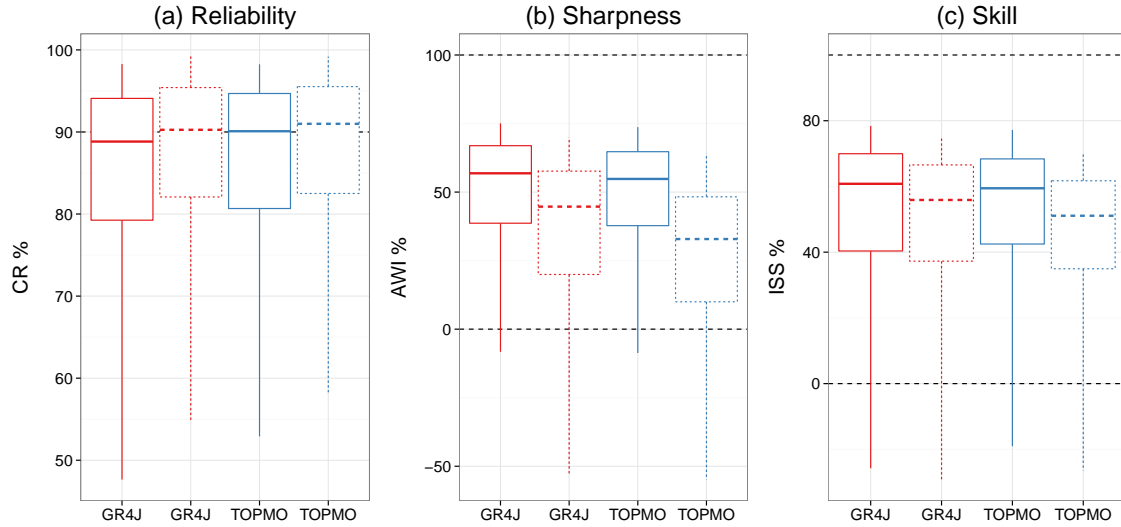


Figure 7.8: Distributions of the three performance criteria, obtained in two cases, i) when 10 groups of relative errors are used (continous lines) and ii) when only one group is used (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

#### 7.5.4 How do the performances of the rainfall-runoff models relate to the characteristics of uncertainty bounds?

To gain insights into the possible relationships between the performance of the deterministic rainfall-runoff simulations and characteristics of the uncertainty bounds at ungauged locations, the three criteria used to characterize the uncertainty bounds are plotted in Fig. 7.9 as function of a quadratic efficiency criteria for the 907 catchments. The quadratic efficiency criterion is the C2M (Mathevet *et al.*, 2006), a bounded version of the Nash and Sutcliffe (1970) efficiency (NSE) criterion. The equations are:

$$\text{C2M} = \frac{\text{NSE}}{2 - \text{NSE}} \quad (7.3)$$

$$\text{NSE} = 1 - \frac{\sum_{t=1}^n (Q_{o,t} - Q_{s,t})^2}{\sum_{t=1}^n (Q_{o,t} - \mu_o)^2} \quad (7.4)$$

where  $n$  is the total number of time-steps,  $Q_{s,t}$  is the simulated discharge at time-step  $t$ ,  $Q_{o,t}$  is the observed discharge at time-step  $t$ , and  $\mu_o$  is the mean of the observed discharges. The

advantage of this bounded version is avoiding large negative values difficult to plot.

A trend appears between deterministic performance and characteristics of the prediction bounds at ungauged locations, for the two rainfall-runoff models. The reliability index exhibits larger variability compared to the sharpness index, and the stronger link is seen for the skill score. Reliability is relatively less affected by the poor deterministic performance of the simulation at ungauged location because there are cases where poor performance at neighbour locations leads (through the transfer of relative errors) to wide prediction bounds that are able to cover the observed values. We can also observe that skill scores and C2M scores are strongly related, which indicates that when the transfer of model parameters performs well, the transfer of model uncertainty estimates performs well too.

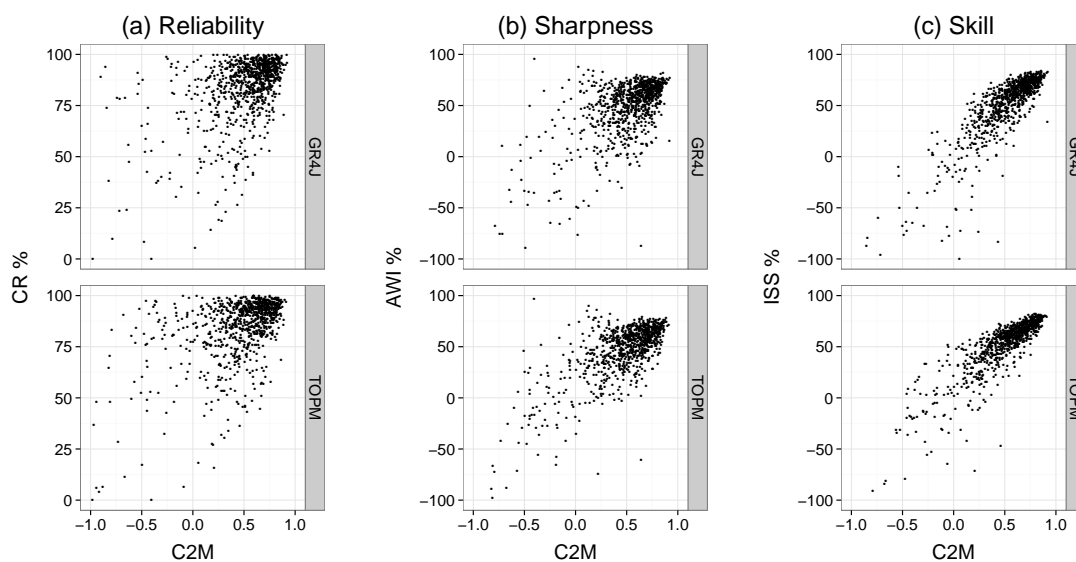


Figure 7.9: Impact of deterministic performance, as quantified by the bounded C2M quadratic criterion, on the three performance criteria for the 907 catchments. Note that for easing visualisation, the lower limits of AWI (b) and ISS (c) values are set to -100% but lower values of AWI are obtained in 7 cases for both models, and lower values are obtained in 18 and 22 cases for GR4J and TOPMO respectively.

## 7.6 Conclusions

Runoff hydrograph prediction in ungauged catchments is notoriously difficult, and attempting to estimate the predictive uncertainty in that context is a further challenge. We proposed a method allowing the transfer of model uncertainty estimates from gauged to ungauged catchments. The method extends the parameter transfer approach to the domain of uncertainty estimation.

We evaluated the approach over a large set of 907 catchments by assessing three expected qualities of uncertainty estimates, reliability, sharpness and overall skill. Our results demonstrate that the method is generally able to reflect model errors at ungauged locations and provide reasonable reliability. We applied two different rainfall- runoff models (GR4J and TOPMO) to ensure that the presented results are not model-specific.

Although we used a transfer based on spatial proximity, the approach is independent of the regionalisation scheme used to obtain deterministic prediction at ungauged locations, and any other similarity measure could be a basis for transferring residual errors.

Last, we would like to stress that the results presented in this study are expressed in terms of dimensionless measures. As such, they can provide a basis for future comparisons when prediction in ungauged catchments with uncertainty estimates is performed.

## **7.7 Acknowledgments**

The authors thank Météo-France for providing the meteorological data and Banque HYDRO for the hydrological data. The financial support of SCHAPI to the first author is also gratefully acknowledged.

# 8

## INCERTITUDE PRÉDICTIVE EN PRÉVISION DES CRUES





## 8.1 Introduction

### 8.1.1 Contexte de nos travaux en prévision des crues

Les travaux présentés dans ce chapitre concernent la quantification de l'incertitude en prévision hydrologique. Ils s'appuient sur l'expérience acquise au sein de l'équipe en matière de modélisation et assimilation de données pour la prévision des crues.

La prévision des crues représente un contexte de modélisation hydrologique particulier. Par rapport au contexte de simulation, la prévision introduit les notions d'instant de prévision et d'échéance, et les techniques d'assimilation des données permettent d'apporter des informations supplémentaires au modèle utilisé sur l'état du bassin à l'instant de la prévision (typiquement le débit observé), afin d'améliorer les prévisions du modèle.

Les travaux réalisés pendant la thèse de [Berthet \(2010\)](#) ont montré que la manière la plus efficace de profiter de la connaissance des débits observés les plus récents était, pour le modèle de prévision GRP, d'ajuster le niveau du réservoir de routage – l'état le plus à l'aval du modèle. Cette approche de mise-à-jour directe s'est révélée plus performante que d'autres méthodes d'assimilation probabilistes comme le filtre de Kalman ou le filtre à particules, et est désormais considérée comme une composante à part entière du modèle de prévision. Elle présente néanmoins un désavantage par rapport aux méthodes d'assimilation probabilistes : elle est déterministe, et ne fournit pas d'estimation de l'incertitude prédictive.

Dans un contexte de prévision des crues, les forçages météorologiques utilisés par un modèle hydrologique sont des prévisions forcément incertaines de l'évolution future de l'atmosphère. La propagation de l'incertitude des prévisions météorologiques ne suffit cependant pas à caractériser l'incertitude totale en prévision hydrologique. Même si on disposait d'une connaissance parfaite des forçages météorologiques, les prévisions obtenues avec un modèle hydrologique seraient imparfaites, puisqu'un modèle n'est qu'une représentation simplifiée du fonctionnement complexe d'un bassin versant. Les deux principales sources d'incertitude en prévision des crues sont la prévision des précipitations et le modèle hydrologique utilisé ([Krzysztofowicz, 1999](#)).

Dans cette thèse, nous cherchons à quantifier l'incertitude du modèle de prévision GRP lorsque les forçages météorologiques sont connus – étant entendu que cette connaissance reste imparfaite en raison des limitations des systèmes et réseaux d'observations. Nous nous sommes intéressés à des méthodes capables de produire des prévisions probabilistes, à partir d'un modèle déterministe. Notre attention s'est portée sur les méthodes de post-traitement

statistique. Dans un contexte de prévision des crues, les méthodes de post-traitement offrent un compromis intéressant entre les approches qui ont l'ambition de séparer l'ensemble des sources d'incertitude, et les approches plus globales de post-traitement des erreurs totales de la chaîne de prévision (e.g., [Fortin et al., 2006](#); [Boucher et al., 2012](#); [Brown et Seo, 2013](#)).

### 8.1.2 Quelques éléments de bibliographie

Notre objectif est d'obtenir une série chronologique de distributions prédictives pour quantifier l'incertitude de prévision. La distribution prédictive doit varier au cours du temps, et en fonction de l'horizon de prévision, afin de refléter la manière dont varient les marges d'erreurs selon les situations de prévision. La formulation explicite de ce problème peut être attribuée à [Krzysztofowicz \(1999\)](#) qui propose de construire un modèle d'erreur du modèle hydrologique utilisé, sous l'hypothèse que les pluies futures sont connues. Plusieurs développements récents s'inspirent de cette approche.

Certaines approches sont paramétriques ou semi-paramétriques; elles supposent que les erreurs suivent une distribution statistique de loi connue, en utilisant le plus souvent une transformation de variables. Il s'agit alors d'estimer les paramètres qui caractérisent ces distributions ([Krzysztofowicz et Maranzano, 2004](#); [Montanari et Grossi, 2008](#); [Coccia et Todini, 2011](#); [Smith et al., 2012](#); [Leedal et al., 2013](#)). Certaines méthodes utilisent des paramètres qui peuvent varier suivant certaines variables explicatives ([Morawietz et al., 2011](#); [Pianosi et Raso, 2012](#); [Yan et al., 2012](#); [Regonda et al., 2013](#)).

D'autres méthodes considèrent le problème comme un problème de régression de quantiles. Il s'agit alors d'estimer, suivant certaines variables explicatives, les séries chronologiques de plusieurs quantiles. Des méthodes de régression paramétriques ([Weerts et al., 2011](#)) ou non-paramétriques ([Solomatine et Shrestha, 2009](#); [Van Steenbergen et al., 2012](#)) peuvent être utilisées pour construire un modèle pour chacun des quantiles considérés. Enfin, [Hopson et Webster \(2010\)](#) et [Ewen et O'Donnell \(2012\)](#) ont proposé des approches qui reposent sur la notion de voisinage.

## 8.2 Cadre de comparaison

### 8.2.1 Présentation des méthodes testées

Les méthodes de post-traitement statistique testées ici ont déjà été présentées dans le chapitre 5. Par rapport à la comparaison réalisée dans un contexte de simulation, seule la méthode de référence  $BC_{0.3}$  a été conservée. Par conséquent, six méthodes, dont les principales caractéristiques sont rappelées dans la Table 8.1, sont comparées dans un contexte de prévision.

Méthode	Caractéristiques
$BC_{0.3}$	transformation Box-Cox (0.3), les erreurs suivent une distribution normale
MeGa	transformation NQT, régression linéaire entre les erreurs et les débits simulés
MCP	transformation NQT, régression linéaire entre les débits observés et simulés
RQ	transformation NQT, régression quantile entre les erreurs et les débits simulés
CemaO	distribution empirique des erreurs relatives
CemaM	distribution empirique des erreurs relatives par tranches de débits

TABLE 8.1 – Caractéristiques des méthodes de post-traitement testées en prévision. La transformation NQT désigne la transformation en quantiles normaux.

### 8.2.2 Modalités d'application

Nous avons appliqué les six méthodes de post-traitement aux chroniques de débits simulés par le modèle GRP pour les 202 bassins versants français présentés au chapitre 2. Nous avons utilisé la version opérationnelle du modèle GRP, qui comprend notamment un utilitaire de calage automatique (voir chapitre 2). Les trois paramètres du modèle sont calés de manière à minimiser les erreurs quadratiques pour un horizon de prévision que nous avons choisi, pour cette étude, égal à 24 heures. Les travaux de [Berthet \(2010\)](#) ont montré l'intérêt de caler les paramètres du modèle lors d'un fonctionnement en prévision (c'est-à-dire avec assimilation des données) ; le choix d'un horizon de prévision de 24 heures permet d'obtenir des performances équilibrées pour différents horizons de prévision.

Les échéances de 1h, 3h, 6h, 12h, 24h, et 48h ont été considérées pour l'application du post-traitement. Il s'agit d'un compromis entre temps de calcul et représentativité de l'évolution des résultats en fonction de l'horizon de prévision. La procédure classique de calage-contrôle a été suivie sur la période 1996–2007, pour laquelle nous disposons des données au pas de temps horaire.

En prévision, les erreurs dépendent fortement de l'horizon, car l'effet des corrections associées aux mises-à-jour du modèle disparaît progressivement. Nous en avons tenu compte en appliquant les méthodes de post-traitement pour chaque horizon de prévision.

## 8.3 Présentation des résultats

### 8.3.1 Fiabilité, finesse et performance globale

La Figure 8.1 résume les indices de fiabilité et de finesse obtenus avec les six méthodes testées, sur les périodes de calage et de contrôle.

Pour toutes les méthodes, les indices de finesse se dégradent avec l'horizon de prévision. Cela traduit logiquement le fait que les erreurs de prévision sont plus importantes à mesure que les effets de l'assimilation de données se dissipent. Les méthodes MeGa, RQ, et CemaO présentent les indices de finesse les plus élevés, tandis que les méthodes BC<sub>0,3</sub> et MCP se distinguent par de plus faibles indices. La méthode CemaM se situe entre ces différentes méthodes.

L'évolution des indices de fiabilité selon les horizons de prévision diffère selon les méthodes. Pour trois méthodes, MeGa, RQ, CemaO, les indices de fiabilité diminuent en calage avec l'horizon de prévision, ce qui indique que les hypothèses sur lesquelles reposent les méthodes sont moins bien adaptées aux erreurs plus complexes rencontrées aux horizons plus lointains. On constate également une tendance à la baisse des indices de fiabilité pour la méthode CemaM en contrôle ; le transfert temporel des marges d'erreurs est moins robuste aux horizons les plus lointains, ce que nous attribuons à une plus grande variabilité des erreurs rencontrées dans ces situations. Par ailleurs, les indices de fiabilité augmentent puis diminuent pour la méthode BC<sub>0,3</sub>, et présentent une nette tendance à la hausse pour la méthode MCP. Ces situations traduisent également les évolutions complexes des distributions d'erreurs selon l'horizon de prévision.

La Figure 8.2 présente les indices de performance globale obtenus avec les six méthodes testées, sur les périodes de calage et de contrôle. Nous constatons que les indices de performance globale suivent des tendances similaires aux indices de finesse : les performances en terme de CRPSS se dégradent nettement, quelque soit la méthode, pour les horizons les plus lointains. La valeur médiane du CRPSS passe par exemple, pour la méthode CemaM en contrôle, de 0.89 pour un horizon de prévision de 6h, à 0.71 pour un horizon de 48h. Les performances globales des différentes méthodes sont assez similaires ; seuls des gains de

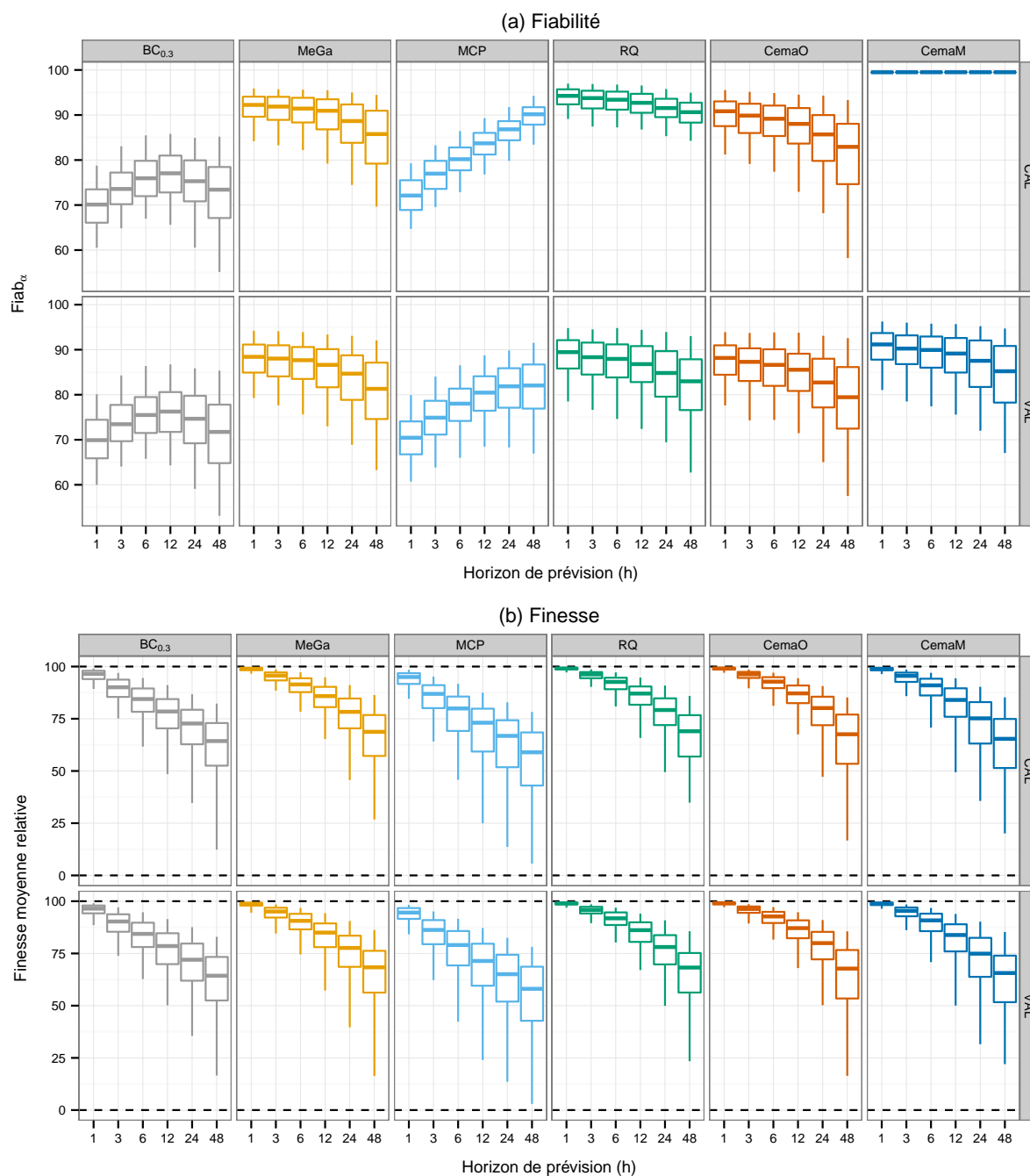


FIGURE 8.1 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices (a) de fiabilité et (b) de finesse des distributions prédictives obtenues avec six méthodes de post-traitement statistique, pour six horizons de prévision, deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans sur 202 bassins.

performance sont à noter pour la méthode CemaM aux plus larges horizons.

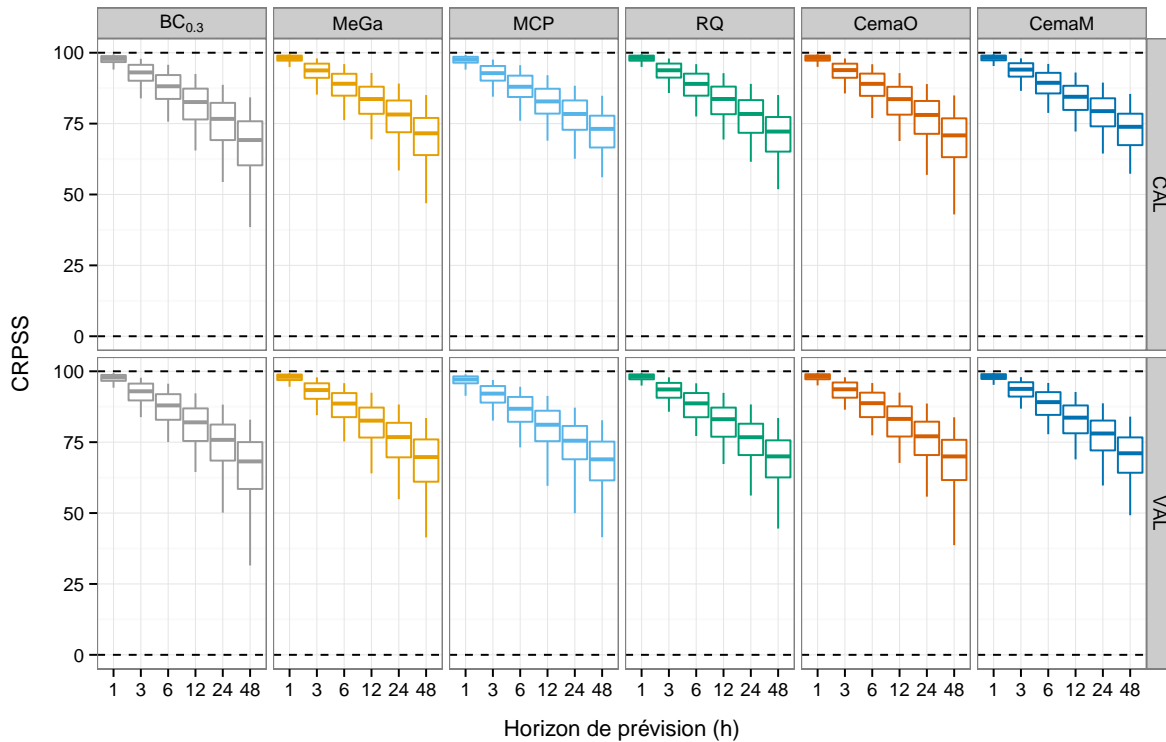


FIGURE 8.2 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des indices de performance globale obtenus avec six méthodes de post-traitement statistique, pour six horizons de prévision, deux périodes de calage (haut), et deux périodes d'évaluation (bas) de cinq ans sur 202 bassins versants.

### 8.3.2 Capacité à prévoir des dépassements de seuils

Dans un contexte de prévision des crues, il est intéressant d'évaluer la capacité des approches prédictives à prévoir des dépassements de seuils.

Le score de Brier a été calculé pour deux seuils qui correspondent, pour chaque bassin et chaque période d'évaluation, aux quantiles 0.7, et 0.9 des débits observés sur la période d'évaluation. Un seuil défini à partir du  $q_{70}$  peut paraître assez faible ; il permet néanmoins d'évaluer au sens large les périodes de hautes eaux. Le BSS (*Brier Skill Score*) présenté sur la Figure 8.3 utilise la climatologie des débits comme référence. Des scores proches de 100 indiquent de bonnes performances, alors que des scores négatifs indiquent que la méthode est moins performante qu'une approche où la même probabilité de dépassement, basée sur la climatologie des débits, serait donnée à chaque instant de prévision.

La Figure 8.3 confirme que les performances se dégradent, pour toutes les méthodes et pour

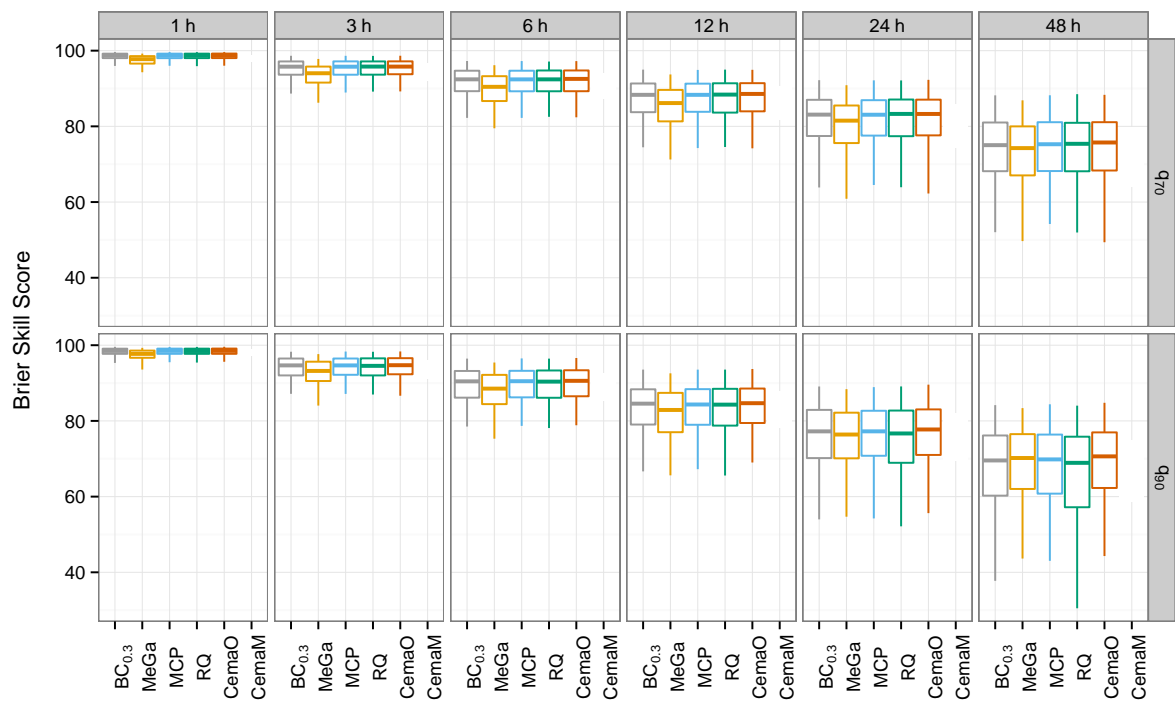


FIGURE 8.3 – Boxplots (quantiles 5, 25, 50, 75 et 95%) des *Brier Skill Score* obtenus avec six méthodes de post-traitement statistique, pour six horizons de prévision et deux seuils de débits sur 202 bassins.



les deux seuils considérés, lorsque l'échéance de prévision augmente. La comparaison des résultats obtenus avec les différentes méthodes indiquent que les méthodes MeGa, RQ et CemaM sont globalement les plus satisfaisantes, mais les différences apparaissent faibles au regard de la variabilité des performances que l'on peut constater suivant les bassins et les périodes d'évaluation.

## 8.4 Vers une meilleure caractérisation de l'incertitude prédictive en prévision

### 8.4.1 Motivations et description de l'approche

La méthode CemaM reflète la variabilité des marges d'erreurs suivant les valeurs du débit prévu par le modèle. Faire dépendre l'incertitude prédictive du débit prévu relève du bon sens hydrologique, et les autres méthodes de quantification de l'incertitude que nous avons testées s'appuient également sur la valeur du débit prévu pour faire évoluer dans le temps les distributions prédictives.

Il est néanmoins souhaitable d'aller vers une meilleure caractérisation de l'incertitude prédictive. Comme le suggère par exemple [Garçon \*et al.\* \(2009\)](#), on peut s'attendre à ce que l'incertitude dépende également du contexte hydrologique, et il semble dès lors souhaitable de tenir compte de cette réalité hydrologique.

Nous décrivons à présent une extension de la méthode CemaM qui vise à mieux caractériser les situations hydrologiques. Deux situations de prévision sont distinguées : la situation de montée correspond au cas où le débit simulé pour l'horizon de prévision considéré est plus important que le débit observé à l'instant de prévision, tandis que la situation de descente correspond au cas contraire. La méthode CemaM est ensuite appliquée séparément à ces deux échantillons d'erreurs. Nous appelons cette variante CemaS.

### 8.4.2 Améliorations obtenues

Pour quantifier les améliorations apportées par une meilleure caractérisation de l'incertitude prédictive, nous avons calculé le rapport des indices de fiabilité, de finesse et de performance globale obtenus avec les deux méthodes CemaM et CemaS. Un résumé des gains pour chaque critère est présenté sur la Figure 8.4.

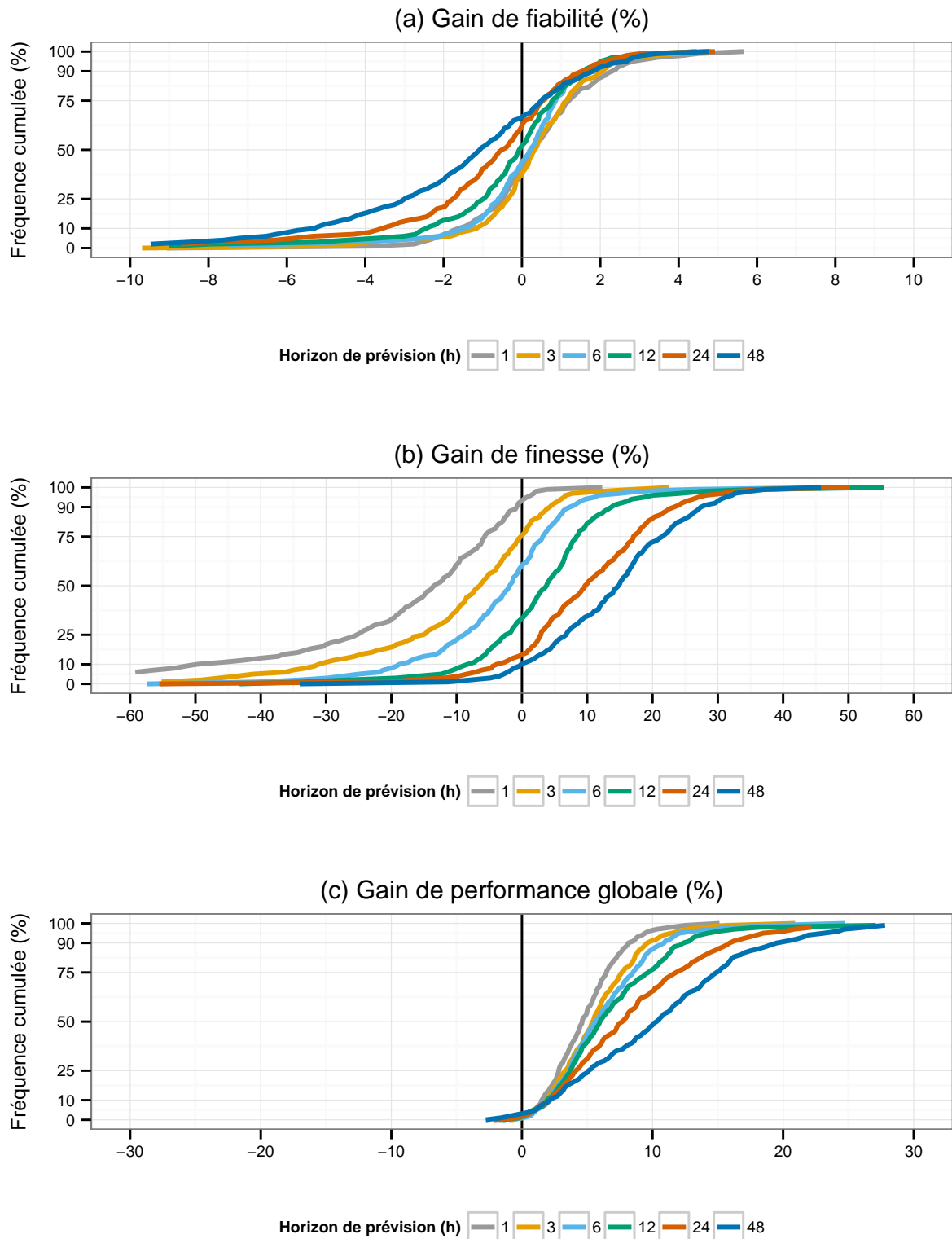


FIGURE 8.4 – Fréquences cumulées sur 202 bassins et deux périodes d'évaluation de cinq ans, des gains de (a) fiabilité, (b) finesse et (c) performance globale de la méthode CemaS par rapport à la méthode CemaM, pour six horizons de prévision.

On constate que la situation est contrastée en terme de gain de fiabilité, puisque des gains positifs et négatifs sont obtenus aux horizons les plus courts, tandis que des pertes de fiabilité sont plus fréquentes pour les horizons de 24 et 48h. On note toutefois que les gains, positifs ou négatifs sont le plus souvent très limités, de l'ordre de quelques pourcents. Il n'y a *a priori* pas de raisons évidentes qui pourraient expliquer d'importantes variations des indices de fiabilité, puisque les deux méthodes utilisent la même manière de caractériser les distributions d'erreurs relatives. Il semble néanmoins que le fait de considérer séparément deux échantillons différents conduise à une légère perte de robustesse aux plus longs horizons.

Les variations en terme de finesse sont nettement plus importantes. On note un déplacement des indices vers les valeurs positives lorsque l'horizon de prévision augmente ; les distributions prédictives obtenues avec la méthode CemaS sont, dans la plupart des cas, plus fines en moyenne que celles de la méthode CemaM à partir d'un horizon de 12h.

Enfin, nous constatons que les gains de performance – le score de performance étant le CRPS moyen – sont, pour tous les horizons, positifs dans plus de 90% des cas, avec des gains médians compris entre 5 et 10%. Les gains les plus importants apparaissent pour les plus longs horizons.

La Figure 8.5 illustre les covariations des gains de fiabilité et finesse obtenus sur notre échantillon de 202 bassins. Nous pouvons remarquer le déplacement des gains de finesse quand les horizons augmentent, et la situation contrastée en terme de gains de fiabilité. Il est également intéressant de constater que des gains de performance peuvent être obtenus lorsque des gains de fiabilité et de finesse sont négatifs. Ces exemples sont rares mais nous rappellent la difficulté d'évaluer toute la richesse de l'information apportée par les prévisions probabilistes.

Enfin, la Figure 8.6 présente les distributions des gains obtenus pour le score de Brier, pour deux seuils de débits ( $q_{70}$  et  $q_{90}$ ), deux périodes, et six horizons de prévision. Les gains sont positifs dans 90% des cas, quel que soit l'horizon et pour les deux seuils considérés, et augmentent aux plus longues échéances. Les gains médians varient de 6 à 18% selon les horizons pour le seuil  $q_{70}$ , et de 5 à 9% pour le seuil  $q_{90}$ .

## 8.5 Synthèse

Dans ce chapitre, différentes méthodes de quantification de l'incertitude prédictive ont été comparées pour le modèle de prévision GRP. Nous nous sommes placés dans une situation idéale de prévision où les pluies futures sont supposées connues, de manière à séparer l'incertitude liée au modèle de prévision de l'incertitude totale.

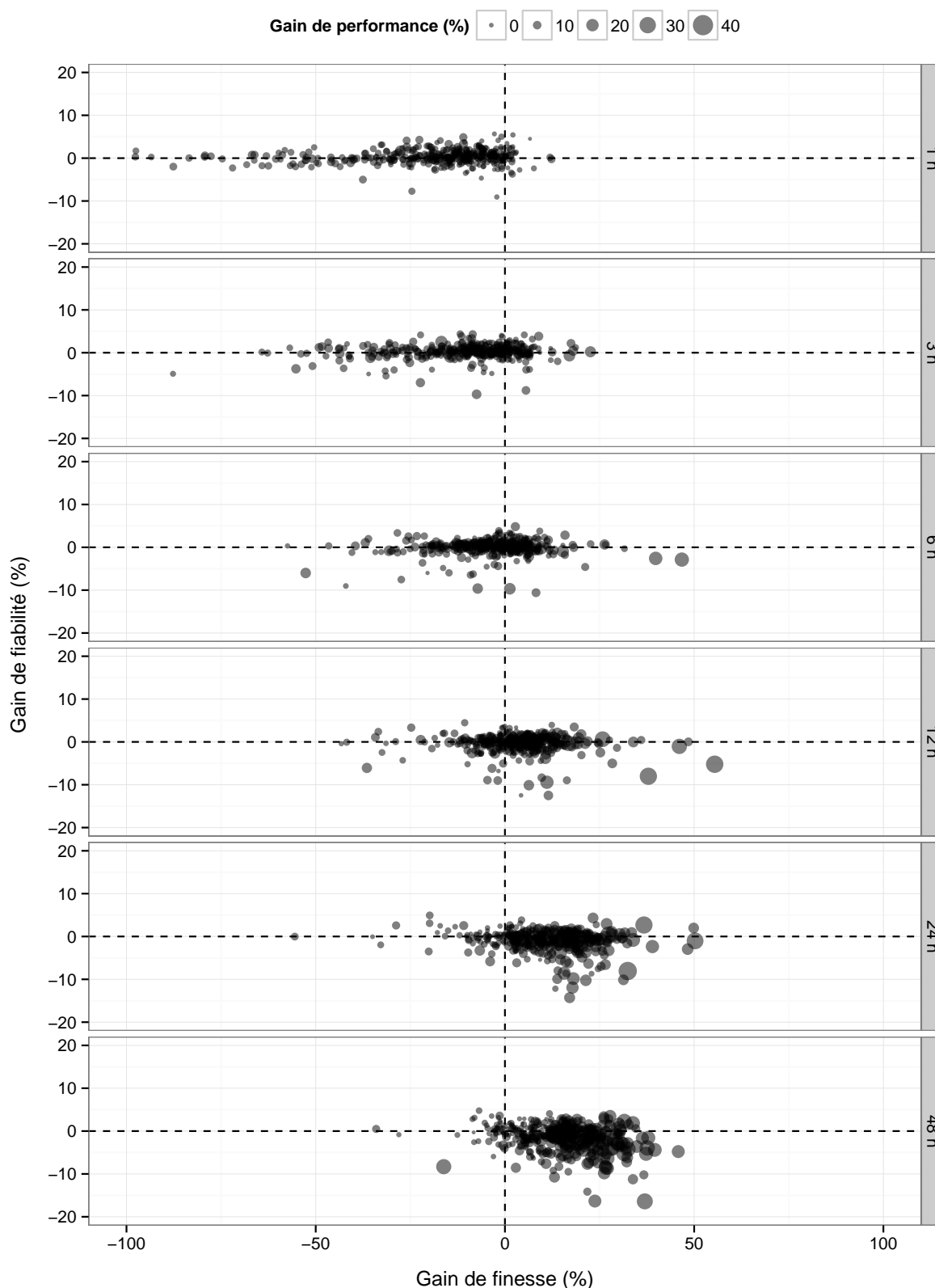


FIGURE 8.5 – Diagrammes en nuage de points des gains de fiabilité et finesse obtenus, sur 202 bassins et deux périodes d'évaluation de cinq ans, avec la méthode CemaS par rapport à la méthode CemaM, pour six horizons de prévision (en ligne). La taille des points dépend du gain de performance globale. Pour faciliter la visualisation, cinq indices de finesse inférieurs à -100% ne sont pas présentés pour la première ligne (1h), et un indice de fiabilité égal à -30% n'est pas présenté pour les deux dernières lignes (24 et 48h).

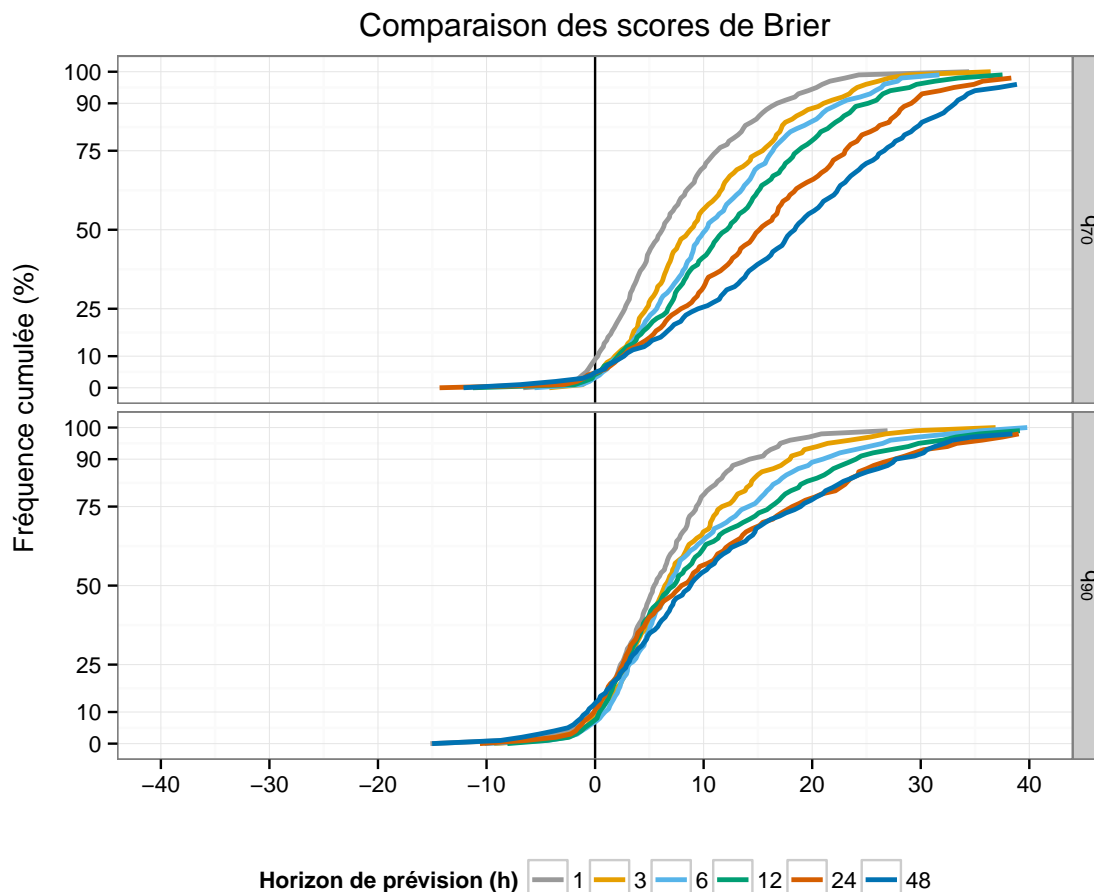


FIGURE 8.6 – Fréquences cumulées sur 202 bassins et deux périodes d'évaluation de cinq ans, des gains sur le score de Brier obtenus avec la méthode CemaS par rapport à la méthode CemaM, pour deux seuils de débits (ligne) et six horizons de prévision.

Les résultats obtenus mettent en évidence le fait que certaines méthodes ont plus de difficultés à refléter la diversité des situations rencontrées en prévision des crues que d'autres méthodes plus flexibles, ce qui se traduit par de plus faibles indices de fiabilité. En particulier, nos résultats suggèrent que l'attention doit être portée sur une estimation suffisamment flexible des quantiles des distributions prédictives.

Nous avons également montré qu'il est possible d'améliorer la performance des prévisions probabilistes obtenues avec une méthode de post-traitement statistique, en caractérisant de manière plus réaliste les situations hydrologiques.

Il nous semble à ce stade essentiel de poursuivre à la fois les efforts de développements, et les travaux de comparaison des méthodes de post-traitement en prévision hydrologique. Ces travaux pourraient notamment s'inscrire dans le cadre de l'inter-comparaison des méthodes de post-traitement actuellement menée au sein de l'initiative HEPEx ([Van Andel et al., 2013](#)).

# 9

## **INVESTIGATING THE INTERACTIONS BETWEEN DATA ASSIMILATION AND POST-PROCESSING IN HYDROLOGICAL ENSEMBLE FORECASTING**



Ce chapitre correspond à un article soumis à la revue *Journal of Hydrology*.

## **Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting**

F. Bourgin<sup>1</sup>, M.H. Ramos<sup>1</sup>, G. Thirel<sup>1</sup>, V. Andréassian<sup>1</sup>

(1) Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, F-92761 Antony Cedex, France

## **Résumé**

Nous étudions dans cet article dans quelle mesure l'assimilation de données et le post-traitement statistique contribuent à la performance d'un système de prévisions hydrologiques d'ensemble. Nous comparons, sur un large échantillon de bassins versants, quatre variantes du système de prévision: sans assimilation de données et sans post-traitement, sans assimilation de données mais avec post-traitement, avec assimilation de données mais sans post-traitement, et enfin avec assimilation de données et post-traitement. Les résultats obtenus indiquent que les deux stratégies agissent de manière complémentaire. L'assimilation de données a un effet très positif, mais qui s'estompe rapidement, sur la justesse des prévisions, tandis que le post-traitement a un effet bénéfique, et durable, sur la fiabilité des prévisions. L'utilisation de ces deux méthodes est par conséquent recommandée pour la prévision hydrologique d'ensemble.

## **Abstract**

We investigate how data assimilation and post-processing contribute, either separately or together, to the skill of a hydrological ensemble forecasting system. Based on a large catchment set, we compare four forecasting options: without data assimilation and post-processing, without data assimilation but with post-processing, with data assimilation but without post-processing, and with both data assimilation and post-processing. Our results clearly indicate that both strategies have complementary effects. Data assimilation has mainly a very positive effect on forecast accuracy. Its impact however decreases with increasing lead time. Post-processing, by accounting specifically for hydrological uncertainty, has a very positive and longer lasting effect on forecast reliability. As a consequence, the use of both techniques is recommended in hydrological ensemble forecasting.



## 9.1 Introduction

### 9.1.1 Addressing uncertainties in hydrological ensemble forecasting

Developing and improving operational hydrological ensemble forecasting systems is a critical step toward better decision-making and risk management. The skill of operational hydrological ensemble forecasting systems is limited by two main sources of uncertainty (Krzysztofowicz, 1999): meteorological uncertainty and hydrological uncertainty. From a pragmatic point of view, the need to properly account for these two main sources of uncertainty arises because (i) a hydrological forecaster has no choice but to rely on uncertain meteorological forecasts and (ii) even with accurate inputs, hydrological forecasts will remain uncertain due to our limited knowledge of initial conditions and the inherent limitations of the forecast model used.

Meteorological uncertainty is commonly addressed by propagating an ensemble (or multi-scenario) input of weather forecasts. For instance, several operational and pre-operational flood forecasting systems across the globe have been set up to be forced by ensemble numerical weather predictions (see Cloke and Pappenberger, 2009, for a review). Addressing the hydrological uncertainty issue is less common, although a general framework of probabilistic forecasting that includes a hydrological post-processing method has been introduced fifteen years ago by Krzysztofowicz (1999). Since then, a number of other hydrological uncertainty processors have been proposed (Montanari and Brath, 2004; Montanari and Grossi, 2008; Solomatine and Shrestha, 2009; Coccia and Todini, 2011; Morawietz *et al.*, 2011; Weerts *et al.*, 2011; Ewen and O'Donnell, 2012; Pianosi and Raso, 2012; Smith *et al.*, 2012; Van Steenbergen *et al.*, 2012; Yan *et al.*, 2012), but their use is not widespread for operational ensemble forecasting.

Although generally dealt with separately, statistical post-processing and data assimilation (also called real-time model updating in the engineering community) can be intrinsically related in the hydrological forecasting framework. Both represent techniques that may be used in a forecasting system to improve the quality of the forecasts (i.e., to provide more accurate and reliable forecasts) and to, ultimately, enhance the usefulness of the forecasts in decision-making. Since forecasting deals with an uncertain future, these techniques aim to bring additional information to the forecast procedure and take into account the various uncertainty sources (or at least the major uncertainty sources) affecting the forecasting chain. This is usually achieved by merging information from model and observations.

While data assimilation and post-processing share a general goal, the techniques applied may differ in the practice of hydrological forecasting. These differences usually draw the separation between what is defined as data assimilation and what is defined as post-processing in a modelling framework. The definitions used in this study are the following: we use the term “post-processing” when using the hydrological uncertainty processor (Section 9.2.4), whose primary purpose is to dress deterministic forecasts with uncertainty based on distributions of past model errors and, this way, build probabilistic forecasts. “Data assimilation” refers to techniques applied to perform the updating of the system before it issues a deterministic forecast. Here it concerns the state updating of the hydrological model and a model error correction applied to its output (Section 9.2.3).

The fact that data assimilation has the potential to improve real-time streamflow forecasting is widely accepted (see [Liu \*et al.\*, 2012](#), for a review). In contrast to probabilistic and ensemble-based data assimilation methods (e.g., [Weerts and El Serafy, 2006](#); [Salamon and Feyen, 2010](#); [Moradkhani \*et al.\*, 2012](#); [Vrugt \*et al.\*, 2013](#)), deterministic updating schemes are designed to improve forecasts without producing probabilistic outputs. They may be easier to implement, mainly operationally, but at the price of leaving the uncertainty quantification issue unanswered. In these cases, the use of statistical post-processing methods together with data assimilation procedures provides a way to reduce and quantify the predictive uncertainty in the hydrological forecasts.

### 9.1.2 Integrating uncertainties in hydrological ensemble forecasting

“Ensemble dressing” is an intuitive and operationally-appealing method that allows integration of uncertainties from hydrological modelling and meteorological (ensemble) forcing. The main difference with other ensemble-based post-processors (e.g., [Wang and Bishop, 2005](#); [Fortin \*et al.\*, 2006](#); [Brown and Seo, 2010](#); [Boucher \*et al.\*, 2012](#); [Brown and Seo, 2013](#)) is that, for ensemble dressing, hydrological modelling errors are assessed separately, and later combined with ensemble forecasts. Distributions of modelling errors are obtained from long time series of simulated and observed data (i.e., learning from the past), and then applied to ensemble forecasts to obtain the total predictive distribution.

In recent studies, the use of ensemble dressing has been implemented and tested to improve the skill of hydrological ensemble forecasting systems. For instance, [Reggiani \*et al.\* \(2009\)](#) present a Bayesian ensemble uncertainty processor for medium-range ensemble flow forecasts in the Rhine river basin. [Hopson and Webster \(2010\)](#) use an uncertainty processor based on the k-nearest neighbors (k-NN) resampling method to dress probabilistic medium-

range forecasts for two large basins in Bangladesh. [Zalachori \*et al.\* \(2012\)](#) compare different strategies based on pre- and post-processing methods to remove biases in a streamflow ensemble prediction system developed for reservoir inflow management in French catchments, while [Pagano \*et al.\* \(2013\)](#) present a hydrological application of ensemble dressing for 128 catchments in Australia.

The studies mentioned above are similar in that they focus on post-processors for operational applications and on the overall evaluation of the quality of post-processed forecasts. Like in the studies that develop and test data assimilation techniques, most of the forecast assessment is on the benefits (in terms of quality) that post-processors or data assimilation may bring to forecast quality (accuracy, reliability, sharpness, etc.) at fixed forecast lead times. Little is known about the interactions between these two components of a forecasting system and the impacts of implementing both post-processing and data assimilation on the performance of the forecasts along the forecast lead times.

### 9.1.3 Aim and scope of the study

This study aims to shed light on the interactions between data assimilation and post-processing in hydrological ensemble forecasting. We address the following questions:

1. How does data assimilation impact hydrological ensemble forecasts?
2. How does post-processing impact hydrological ensemble forecasts?
3. How does data assimilation interact with post-processing to improve the quality and skill of hydrological ensemble forecasts over the forecast lead times?

We address these questions with the help of a large set of catchments, making it possible to draw more general and robust conclusions.

## 9.2 Data and methods

### 9.2.1 Data set

A set of 202 unregulated catchments spread over France was used (Figure 9.1). The catchments represent various hydrological conditions, given the variability in climate, topography, and geology in France. This set includes fast responding Mediterranean basins with intense precipitation as well as larger, groundwater-dominated basins. Some characteristics of the

data set are given in Table 9.1. Catchments were selected to have limited snow influence, since no snowmelt module was used in the hydrological modelling (Section 9.2.3).

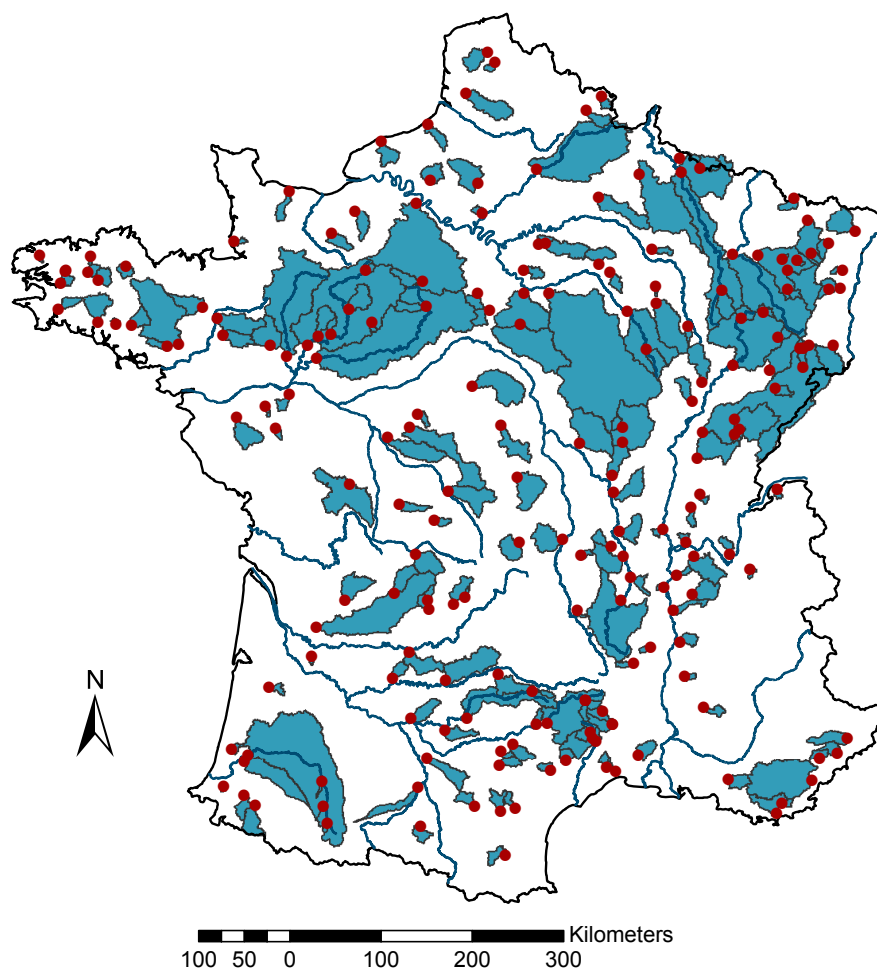


Figure 9.1: Locations of the 202 French catchments used in this study (dots correspond to the gauging stations, and blue colour is catchment areas).

Potential evapotranspiration (PE), precipitation, and discharge data were available at hourly time steps over the 1997–2006 period. Temperature inputs originate from the SAFRAN reanalysis (Vidal *et al.*, 2010). PE was estimated using a temperature-based formula (Oudin *et al.*, 2005). Precipitation data come from a reanalysis dataset recently produced by Météo-France based on weather radar and rain gauge network (Tabary *et al.*, 2012). River discharge data were extracted from the HYDRO national archive ([www.hydro.eaufrance.fr](http://www.hydro.eaufrance.fr)).

	Percentiles				
	0.05	0.25	0.50	0.75	0.95
Catchment area (km <sup>2</sup> )	31	108	245	653	3761
Mean annual precipitation (mm/y)	725	848	957	1158	1465
Mean annual potential evapotranspiration (mm/y)	645	668	701	745	828
Mean annual runoff (mm/y)	143	232	344	513	964
Q/P ratio	0.18	0.27	0.35	0.47	0.68
P/PE ratio	0.93	1.14	1.36	1.66	2.14
Mean elevation (m)	86	155	306	535	843
Discharge autocorrelation at 48 h	0.28	0.5	0.66	0.81	0.94

Table 9.1: Characteristics of the 202 catchments. P–precipitation, PE–potential evapotranspiration, Q–discharge.

## 9.2.2 PEARP, the Météo-France ensemble forecast

A short-range meteorological ensemble prediction system, the Météo-France PEARP EPS (Nicoulaou, 2002), was used to produce hydrological ensemble forecasts. The PEARP EPS runs once a day at 18:00 UTC; it has 11 members, a 60 h forecast range, and a 0.25° (ca. 25 km in France) grid resolution. A spatial disaggregation to an 8 km × 8 km grid, which includes bias correction, was applied to the PEARP forecasts. Bias correction was applied to precipitation forecasts using a multiplying factor obtained from a comparison between the mean of the PEARP ensemble and the Météo-France SAFRAN reanalysis over a complete year (March 2005–March 2006). Details can be found in Thirel *et al.* (2008). PEARP forecasts were available over the 2005–2009 period, but only the period matching the observed data could be used here, i.e. from August 2005 to December 2006.

PEARP forecasts were already used at the daily time step in recent hydrological studies (Thirel *et al.*, 2008; Randrianasolo *et al.*, 2010). Overall, they showed good quality over France at this time step. The quality for short-term forecasting at hourly time steps (with either raw and post-processed forecasts) is first assessed here.

## 9.2.3 The GRP rainfall-runoff forecasting model

The GRP model is a continuous, lumped storage-type model designed for flood forecasting. Its structure was derived from the GR4J model (Perrin *et al.*, 2003) and is composed of a production function and a routing function. The production function consists of a non-linear soil moisture accounting (SMA) reservoir and a volume adjustment coefficient. The routing function includes a unit hydrograph and a non-linear routing store. The GRP model uses

catchment areal rainfall and PE as inputs; it is parsimonious with three parameters to be calibrated against observed data: one in the production function (the volume adjustment coefficient) and two for the routing function (the base time of the unit hydrograph and the total capacity of the routing store). In this study, the three free parameters were calibrated for each catchment by minimizing the root mean square errors (RMSE) during the first five years of available data (1997–2001).

Importantly, the hourly version of the GRP model uses together two data assimilation procedures for flood forecasting. The first exploits the last available observed discharge to directly update the routing store state, and the second exploits the last relative error to correct the model output with a multiplicative coefficient. More details about the forecasting model GRP and the two assimilation procedures can be found in [Berthet \*et al.\* \(2009\)](#).

#### 9.2.4 Hydrological uncertainty processor

We used a hydrological uncertainty processor (HUP) to evaluate the conditional errors of the hydrological model. Only hydrological uncertainty is considered by the HUP here since the model is run with observed weather data. The meteorological uncertainty is subsequently considered through the joint use of the HUP with the PEARP forecasts, as described in Section 9.2.5. The HUP used here is a data-based and non-parametric method that was applied by [Andréassian \*et al.\* \(2007\)](#) to assess model simulation uncertainties and compute empirical uncertainty bounds to flow simulations. Here it is applied to produce probabilistic flow forecasts. The basic idea is to estimate empirical quantiles of relative errors stratified by different flow groups. The HUP is trained during the period used for calibrating the parameters of the hydrological model (1997–2001). Note that it is possible that this approach yield optimistic uncertainty estimates, since errors are usually larger during an independent period than during the calibration period. Since forecast error characteristics vary with forecast range when data assimilation is used, the HUP is trained at several lead times separately.

For each catchment, the HUP is trained as described below:

1. The hydrological model is run with observed weather data as input and the time series of relative errors is evaluated:  $Q_{fct}/Q_{obs}$ , where  $(Q_{fct}, Q_{obs})$  are the pairs of discharge forecasts and observations.
2. The time series is stratified into 20 groups according to the magnitude of the  $Q_{fct}$ . The limits of each group are fixed so that each group contains the same number of values.

3. Within each group, an empirical distribution of relative errors is defined and 99 quantiles are estimated (corresponding to the percentiles 1%, 2%, ..., 98%, 99%).

Application of the HUP for another forecast period is described by the last step:

4. Once defined during the training period, the empirical quantiles of relative errors can be applied to any forecast discharge at a certain lead time. The limits of each group are the same as those obtained during the training period. Note that when data assimilation is not used, the empirical quantiles of relative errors are the same whatever the forecast lead time is. Given a discharge forecast  $Q_{fct}$ , we first determine the flow group  $Q_{fct}$  belongs to; then  $Q_{fct}$  is multiplied by the 99 quantiles of relative errors; the 99 values obtained describe the predictive distribution at the considered time step and for a given forecast horizon. In cases of extrapolation (i.e., when the forecast discharge is out of the range of the flow groups defined during the training phase of the HUP), values of relative errors from the nearest flow groups (i.e., the lowest or the highest flow groups) are used.

Preliminary studies carried out to compare this approach to other similar post-processing approaches suggest that it can yield similar results in terms of forecast performance, while being simpler in its application.

### 9.2.5 Ensemble dressing method: an integrator of the meteorological and hydrological uncertainties

The ensemble dressing method is used as an integrator of the meteorological and hydrological uncertainties. It consists in two steps. Firstly, each time an ensemble PEARP forecast is available, the hydrological model is run with the ensemble forecast and the HUP is applied, according to Step 4 of Section 9.2.4, to each of the 11 members of the ensemble for each lead time considered. Secondly, the  $11 \times 99$  values obtained at each lead time are pooled together and an empirical cumulative distribution is estimated. From this distribution, 99 quantiles are retained as the members of the dressed ensemble.

Application and evaluation of the ensemble dressing method for the ensemble forecasts is done over an independent period, the 17-month period from August 2005 to December 2006.

### 9.2.6 Experiments

The hydrological ensemble forecast system combines meteorological and streamflow data from observation networks, the Météo-France PEARP ensemble forecast, the GRP rainfall-

runoff model with its two data assimilation functions, the hydrological uncertainty processor (HUP) and the ensemble dressing method. Hereafter we will use the term “post-processing” to describe the joint use of the HUP and the ensemble dressing method, while the term “data assimilation” will refer to the two updating techniques used in the GRP model.

In order to assess the benefits of data assimilation and post-processing, considered together or separately, different configurations of the forecasting chain were analyzed. Our experiments comprise a chain without data assimilation and post-processing (NoDA-NoPP), without data assimilation but with post-processing (NoDA-PP), with data assimilation but without post-processing (DA-NoPP), and with both data assimilation and post-processing (DA-PP). The characteristics of the experiments and the acronyms used are given in Table 9.2.

In particular, the NoDA-NoPP experiment corresponds to the situation where the hydrological model is run in simulation mode, i.e., without using recent streamflow observations for data assimilation, and is then driven by the PEARP ensemble forecast when the forecast is issued. When data assimilation is used, the state of the routing reservoir of the hydrological model is first updated based on the last observed discharge, and the second procedure is then applied separately at each streamflow ensemble member. This structured analysis allows us to identify the influence of data assimilation and post-processing separately to assess the benefits of both components when used together in the forecasting chain.

	Without data assimilation	With data assimilation
Without post-processing	NoDA-NoPP	DA-NoPP
With post-processing	NoDA-PP	DA-PP

Table 9.2: Acronyms used for the different experiments used in this study.

### 9.2.7 Forecast evaluation methods

The evaluation of the performance of probabilistic forecasts should reflect the different facets of probabilistic forecasts. In this study, the forecasts obtained from the four experiments set up (Table 9.2) were evaluated with both deterministic and probabilistic scores. We aimed to assess the influence of data assimilation and post-processing on the following characteristics of ensemble forecasts: accuracy of the ensemble mean, overall sharpness and reliability of the whole ensemble, and overall forecast quality of the ensemble.

More specifically, we evaluated the accuracy of the ensemble mean values with the relative bias (BIAS) and the normalized root-mean-square error (NRMSE). To assess the overall reliability of the forecasts, we used the Probability Integral Transform (PIT) diagram (see e.g.,



Laio and Tamea, 2007; Thyer *et al.*, 2009) and an index that quantifies deviation from the ideal case, the alpha score (Renard *et al.*, 2010). The overall sharpness of the forecasts was measured with an index based on the interquartile range that we called normalized mean interquartile range (NMIQR). Finally, we assessed the overall forecast quality of the whole ensemble with the mean Continuous Rank Probability Skill Score (mean CRPSS). The mean CRPSS is computed with the unconditional streamflow climatology as the reference. These scores are presented in more details in Appendix 9.5.

## 9.3 Results and discussion

### 9.3.1 Forecast accuracy

Figure 9.2 shows the distributions of the two deterministic scores used to assess forecast accuracy: the relative bias (BIAS) and the normalized root-mean-square error (NRMSE). Each score is computed for lead times 6 h, 12 h, 24 h and 48 h and for all 202 catchments. The distribution of the 202 values is summarized with boxplots.

We note that forecast accuracy decreases with increasing lead time for the four experiments. For NoDA experiments (NoDA-NoPP and NoDA-PP), the loss of performance is quite limited: it is only related to the decreasing performance of the PEARP ensemble precipitation forecasts. For DA experiments (DA-NoPP and Da-PP), the decrease is stronger and the performances converge toward those of NoDA experiments: the effects of the two DA procedures used in the GRP forecasting model vanish with larger horizons; the decrease in performance of the hydrological model is then added to the losses in performance of the PEARP ensemble precipitation forecasts. Figure 9.2 also reveals that post-processing does not significantly impact forecast accuracy, whether or not DA is used. DA has a much stronger impact on the ensemble mean values than post-processing, especially for shorter lead times and, to a lower extent, for larger lead times. The two DA procedures used in the GRP forecasting model have been designed to improve the performance of deterministic forecasts and, as can be seen, they clearly help improving the mean of the ensemble forecasts. Post-processing on the other hand primarily aims to account for hydrological uncertainty. Its capability to reduce overall bias and squared errors in the mean of the ensemble forecasts is limited here. Nonetheless, for all lead times, forecast accuracy is best when DA and PP are used together, which indicates the benefits of the combined use of data assimilation and post-processing.

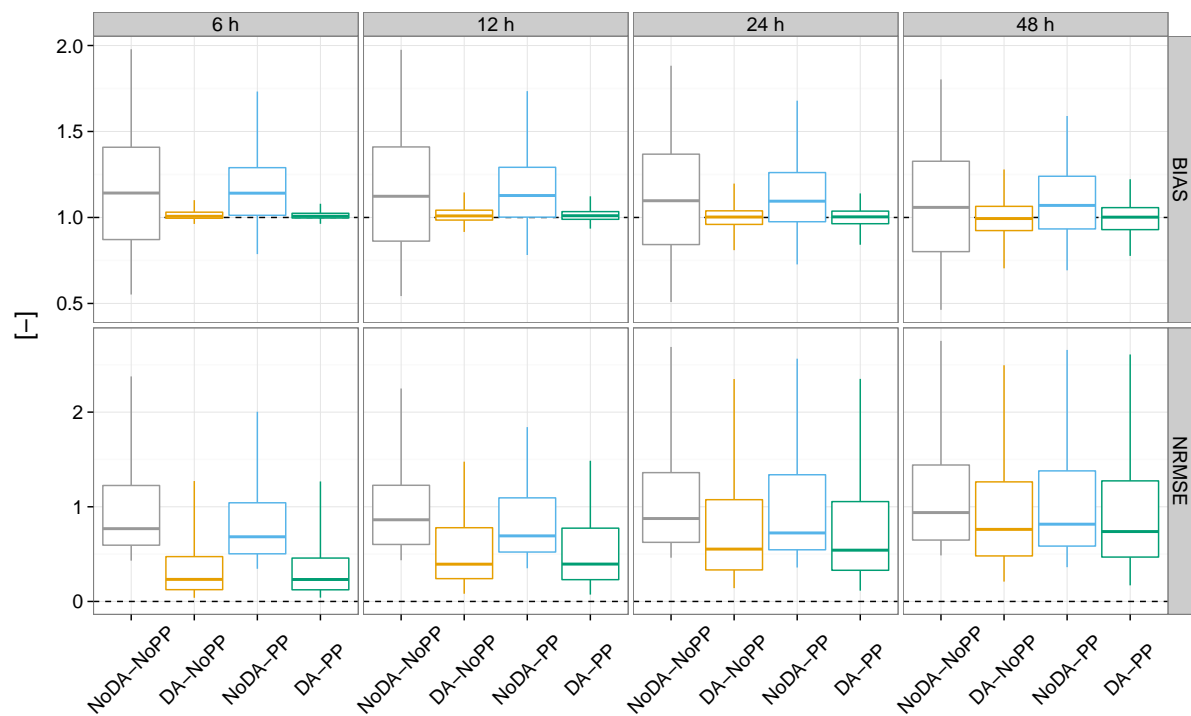


Figure 9.2: Distributions of two deterministic scores, the relative bias (BIAS) and the normalized root-mean-square error (NRMSE), for ensemble streamflow forecasts from the four experiments (see Table 9.2) and lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set.

### 9.3.2 Reliability

Figure 9.3 presents the PIT diagrams obtained for each of the 202 catchments, when considering 24 h ahead ensemble forecasts. Similar figures were obtained for the other lead times (not shown).

From Figures 9.3a and 9.3b, it can be seen that most of the curves are almost horizontal straight lines, while they would follow the bisector (black lines in the graphs) in the ideal case of reliable ensemble predictions. Figures 9.3a and 9.3b clearly reveal that the raw ensembles are lacking reliability for all of the catchments. The impact of post-processing on reliability is apparent when looking at the results in Figures 9.3c and 9.3d: the curves of the ensemble streamflow forecasts with post-processing follow the ideal situation much more closely than the curves shown in Figures 9.3a and 9.3b (ensemble streamflow forecasts without post-processing). It means that the overall reliability of the ensembles is clearly improved with post-processing and this for both cases, with and without DA. A comparison of solely Figures 9.3c and 9.3d confirms also the positive impact of data assimilation on the reliability of the ensembles: the PIT curves of the dressed ensembles are substantially closer to the diagonal (perfect reliability) when DA is applied.

The PIT diagrams convey a visual evaluation of the overall reliability of probabilistic forecasts. To quantify it, we used the alpha score, a reliability index that measures the deviation of the PIT curves from the ideal situation. Figure 9.4 presents the distributions of the alpha scores obtained for each experiment over the 202 catchments. Results in Figure 9.4 confirm the visual evaluation obtained with the PIT diagrams: the two experiments that do not account for hydrological uncertainty (NoDA-NoPP and DA-NoPP) lack reliability. Their alpha values are almost always below 0.5, while the alpha values obtained when hydrological uncertainty is taken into account (NoDA-PP and DA-PP) are almost always higher than 0.5. The benefits of DA is also apparent when comparing, on one hand NoDA-NoPP and DA-NoPP, and on the other hand NoDA-PP and DA-PP, although it can be also seen that DA alone (comparing NoDA-NoPP to DA-NoPP) cannot correct under dispersion of the ensemble forecasts. Post-processing is then a necessary step to achieve reliable forecasts in the forecasting chain analyzed.

These results suggest that for the 202 catchments studied the spread obtained by propagating solely the precipitation ensembles into the hydrological model is too small to properly reflect the range of forecast errors. The deterministic data assimilation strategy used here is effective in improving the reliability of the ensemble forecasts, but it is not sufficient to correct the under dispersion of the streamflow ensemble forecasts as revealed by the PIT diagrams in

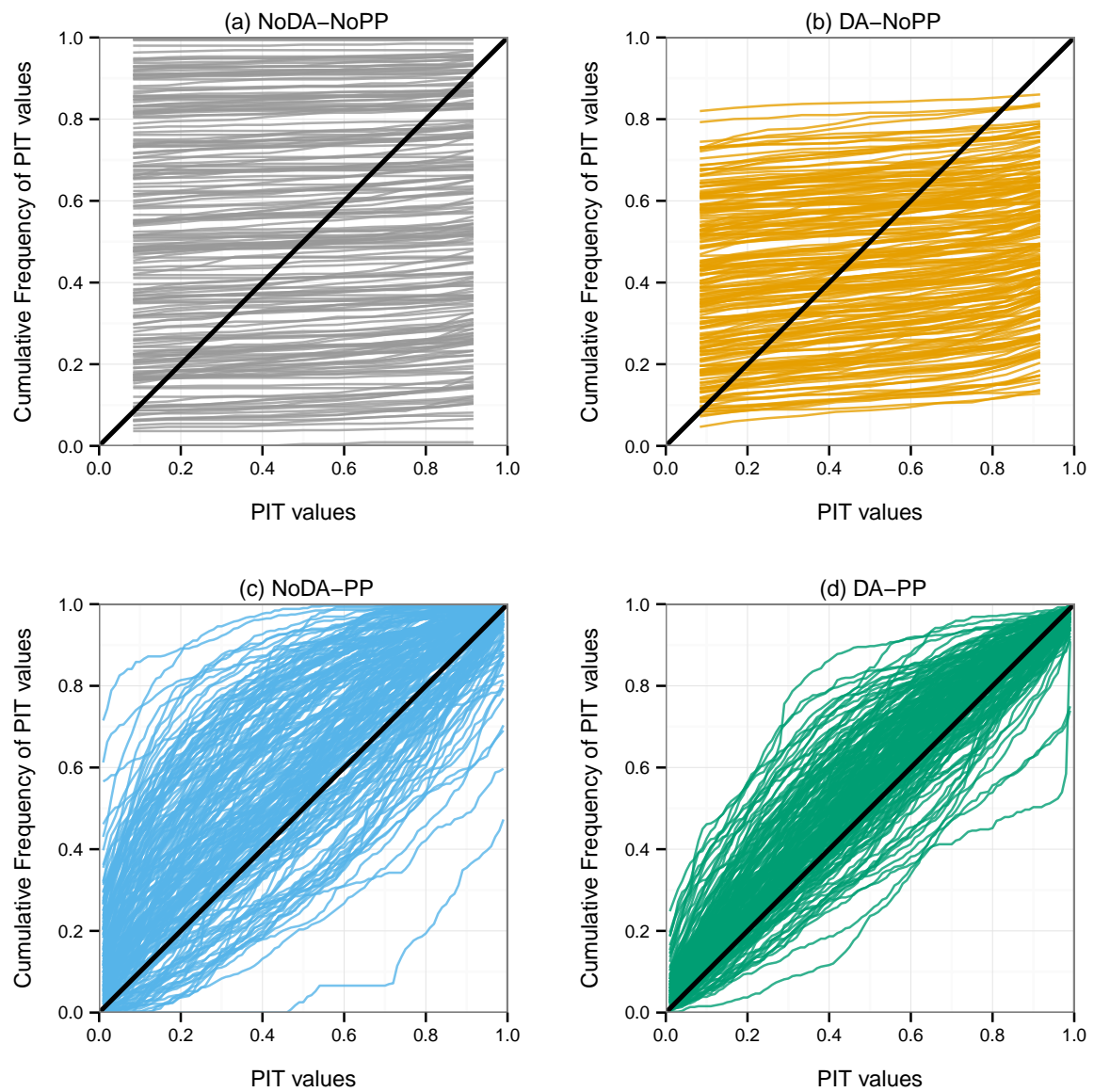


Figure 9.3: PIT diagrams of the 24 h ahead streamflow ensemble forecasts from the four experiments (see Table 9.2). Each line represents one of the 202 catchments of the data set.

Figure 9.3 and the alpha scores in Figure 9.4. This is a strong indication that the hydrological uncertainty issue should be specifically addressed in order to improve the overall reliability of hydrological ensemble forecasts.

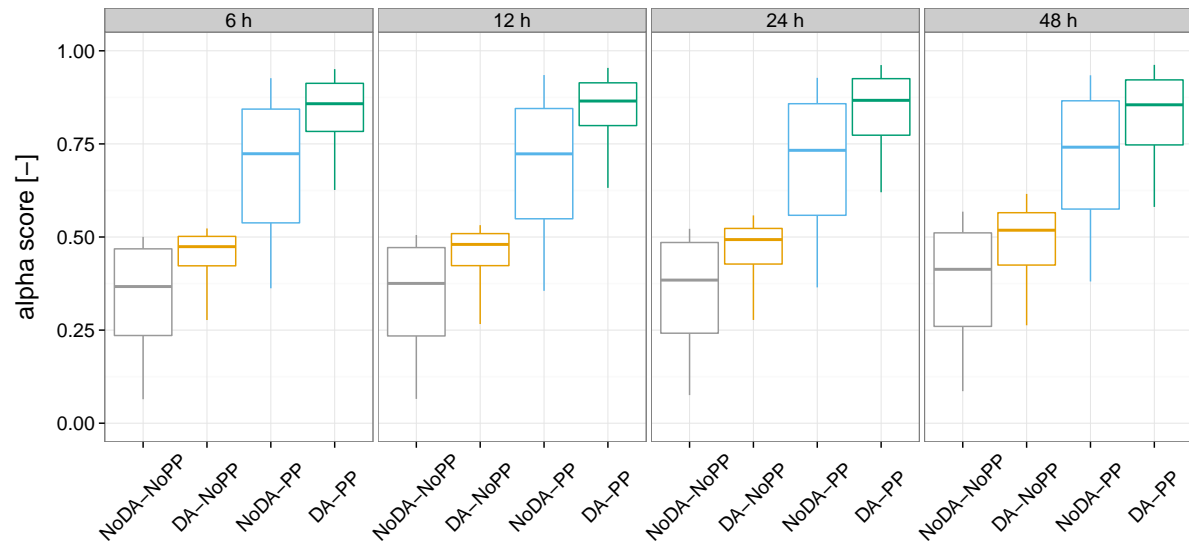


Figure 9.4: Distributions of the alpha score reliability index for streamflow ensemble forecasts from the four experiments (see Table 9.2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score is 1.0.

### 9.3.3 Sharpness

Sharpness is a desirable characteristic of any probabilistic forecast. The sharper the forecast, the less uncertain it is, and thus the more information is conveyed. The four experiments we used made it possible to investigate how meteorological and hydrological uncertainties interact and affect sharpness. Figure 9.5 shows the distributions of a sharpness index, the normalized mean interquartile range (NMIQR), over 202 catchments.

It can be seen that the ensemble spreads of three experiments, NoDA-NoPP, DA-NoPP and DA-PP, increase significantly with increasing lead time, while it is more stable over lead times for the experiment NoDA-PP. For NoDa-NoPP and DA-NoPP, the median value of NMIQR over the 202 catchments raises in a very close behavior for both experiments, from around 0.05 for 6 h ahead forecasts to 0.13 for 48 h ahead forecasts. For the experiment DA-PP, the increase in the median values is much more important: from 0.07 at 6 h to 0.32 at 48 h. These results indicate that forecast uncertainty increases with increasing lead time as the result of increasing meteorological uncertainty alone (NoDA-NoPP and DA-NoPP) or as the result of increasing meteorological and hydrological uncertainties considered together and

with DA (DA-PP). Comparing DA-NoPP and DA-PP reveals the impact of post-processing: taking into account hydrological uncertainty leads to more spread and less sharpness in ensemble forecasts. Comparing NoDA-NoPP and DA-NoPP shows that the propagation of meteorological uncertainty has a rather similar impact on ensemble sharpness whether or not DA is used to update the states of the forecasting model. Remarkably, the ensemble spreads obtained without DA but with post-processing (NoDA-PP) is stable across the lead times with a median value over the 202 catchments around 0.52. This is because statistical post-processing reflects the large errors obtained when the forecasting model does not use DA (see Figure 9.2). In this case, the spread obtained when taking hydrological uncertainty into account is so large that the increasing spread of the PEARP ensemble forecasts with increasing lead time has no visible impact on the spread of the post-processed ensemble: hydrological uncertainty dominates meteorological uncertainty.

Not surprisingly, sharper forecasts are obtained when only meteorological uncertainty is taken into account (NoPP experiments). This is to the detriment of reliability: ensemble forecasts with only meteorological uncertainty are sharper but not reliable, reflecting the presence of under dispersion (as shown in Section 9.3.2). The use of post-processing (PP experiments) leads to ensembles that are more spread out because they attempt to handle hydrological uncertainty and reflect hydrological forecast errors. Ensembles are thus less sharp but, on the other hand, achieve reliability. At this point, it should be remembered that sharp but unreliable forecasts should be considered with caution. Unreliable forecasts can convey a wrong impression of certainty that results from having neglected one or several important sources of uncertainty.

#### 9.3.4 Mean CRPSS

The analysis of the impacts of data assimilation and post-processing on two important characteristics of probabilistic forecasts, reliability and sharpness, showed that post-processing was necessary to improve reliability, but at the cost of lower sharpness, i.e., greater ensemble spread and uncertainty, even if sharpness could be improved with the application of a data assimilation procedure. We now turn our attention to the mean CRPSS, a probabilistic score that provides an assessment of the overall quality of ensemble forecasts.

Figure 9.6 shows the distributions of the mean CRPSS over 202 catchments. We note that performance decreases with increasing lead time for the two experiments with data assimilation: median values of the CRPSS are equal to 0.84 (DA-NoPP) and 0.87 (DA-PP) for 6 h range forecasts, and equal to 0.45 (DA-NoPP) and 0.57 (DA-PP) for 48 h range forecasts. Mean

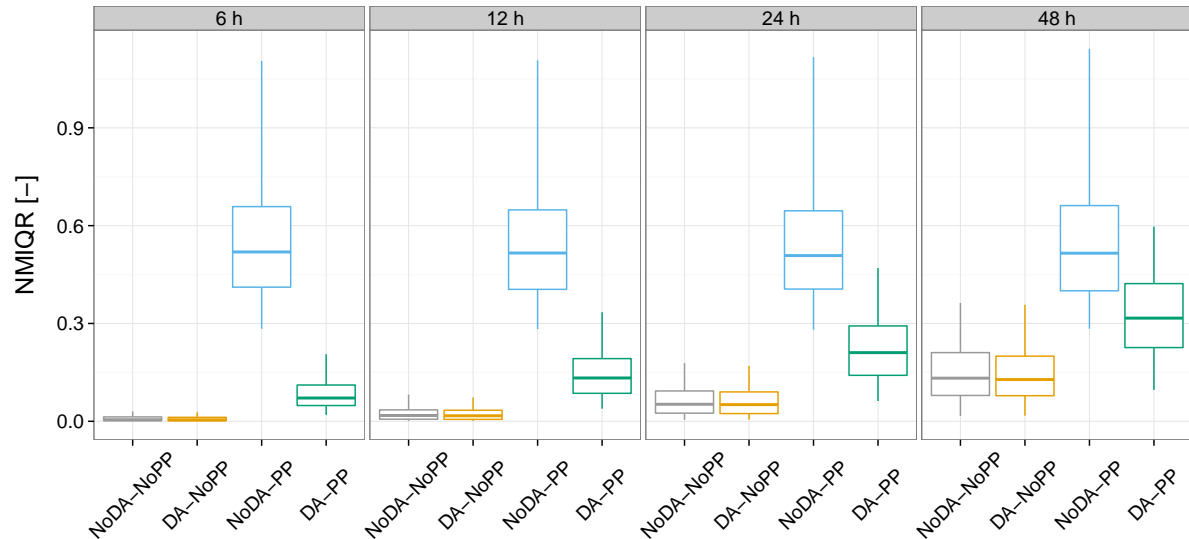


Figure 9.5: Distributions of the normalized mean interquartile range (NMIQR) for streamflow ensemble forecasts from the four experiments (see Table 9.2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score is 0.

CRPSS values of the two experiments without data assimilation decrease only slightly but are much lower than values obtained with data assimilation (median values around 0.10 for NoDA-NoPP and around 0.45 for NoDA-PP). This is especially true for shorter lead times and, to a lower extent, for larger lead times. Furthermore, the comparison with the reference climatology shows that data assimilation alone is sufficient to generate skillful forecasts for more than 95% of the catchments for lead times up to 24 h, but post-processing (DA-PP) is necessary to achieve forecasts that have better overall performance than climatology at 48 h.

These results show the general added value of data assimilation and post-processing to the overall quality of ensemble forecasts. When evaluating the overall quality of ensemble forecasts with the CRPSS, the benefits in terms of reliability overcome the loss of sharpness that results from accounting for hydrological uncertainty. The streamflow ensemble forecasts that explicitly account for both sources of uncertainty, meteorological and hydrological uncertainties, through post-processing, while reducing as much as possible hydrological uncertainty, here through data assimilation, are the most skillful forecasts.

## 9.4 Summary and conclusions

We investigated the relative contributions of data assimilation and post-processing to the skill of hydrological ensemble forecasts. The study assessed the benefits of data assimilation and

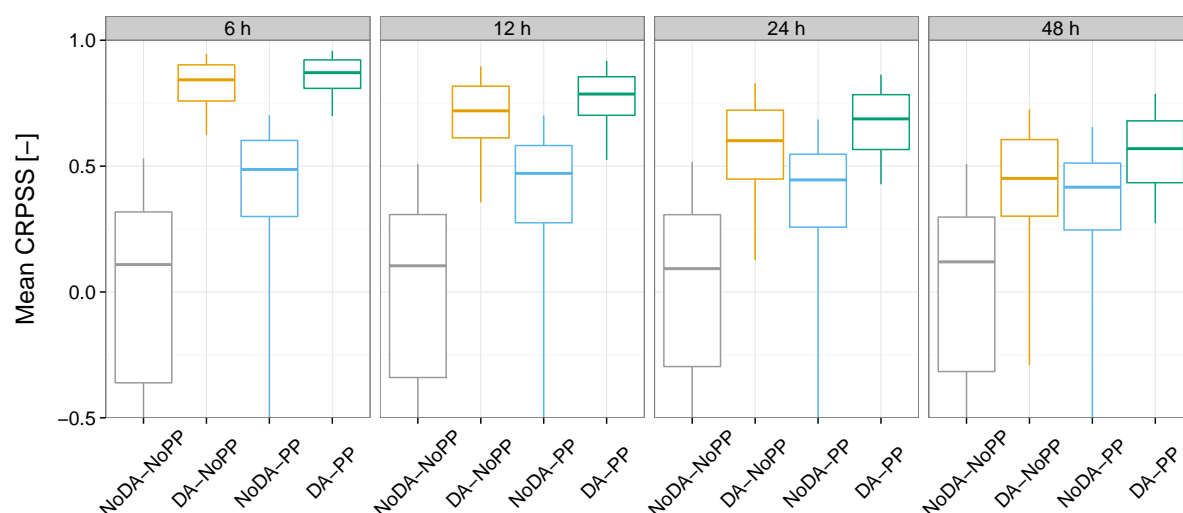


Figure 9.6: Distributions of the mean CRPSS for streamflow ensemble forecasts from the four experiments (see Table 9.2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score is 1.0.

post-processing with the help of four configurations of a short-range hydrological ensemble forecasting system: without data assimilation and post-processing (NoDA-NoPP), without data assimilation but with post-processing (NoDA-PP), with data assimilation but without post-processing (DA-NoPP), and with both data assimilation and post-processing (DA-PP).

We applied deterministic and probabilistic scores to streamflow forecasts of a large catchment set which brought into light the main general conclusions listed below:

- We verify the well-known fact that short-range hydrological forecasts benefit from data assimilation. Data assimilation has a strong impact on improving the quality of the ensemble mean, and a much lesser effect on the variability of the ensemble members (i.e., their spread).
- The benefits of a simple yet efficient hydrological uncertainty processor to improve the reliability and the overall quality of the short-range hydrological ensemble forecasts were demonstrated. Post-processing has a strong impact on forecast reliability.
- The benefits of the combined use of data assimilation and post-processing were demonstrated: both contribute to achieve reliable and sharp forecasts, with impacts acting differently according to the target lead time. The stronger impact on forecast reliability comes from the use of post-processing. Adding data assimilation to the system helps in improving sharpness and reliability at all lead times, with higher gains in performance at shorter lead times.



We acknowledge some limitations. It was only possible to evaluate the forecasting chain over a 17-month period of ensemble forecasts, since this was the common period between observations and forecasts we had available. Furthermore, PEARP ensembles are ran only once a day, which limits the number of hourly evaluation pairs. For these reasons, it was not possible to evaluate flows over specific flooding thresholds. However, with increasing data archives, we expect that such an issue will be treated in future work.

Our study considered only one data assimilation technique (state updating with error output correction) and one post-processing method (ensemble dressing with hydrological errors) together with one rainfall-runoff model forecasting (GRP model). There are several other techniques and models in the literature that could also be tested using the methodology presented here. For instance, a comparison between different configurations of the method used, or different hydrological uncertainty processors, including methods that take into account the autocorrelation of errors (e.g., [Schoups and Vrugt, 2010](#)) could be investigated. Besides, while a bias correction was applied to the PEARP forecasts, a more sophisticated pre-processor (see e.g., [Verkade et al., 2013](#)) could be used to further investigate how meteorological and hydrological biases interact and contribute to the quality of the final hydrological ensemble.

Also, the effectiveness of a data assimilation technique or a post-processing method (and hence the choice of the procedures to operate in a forecasting system) is affected by different sources of uncertainties present in a flow forecasting system, including the forcing data, initial conditions, parameter uncertainty and model structural uncertainty. In our study, we followed the works of [Krzysztofowicz \(1999\)](#) and focused on a decomposition of the total uncertainty into meteorological and hydrological uncertainty. Observational or parameter uncertainties were thus not explicitly considered. Additional sources of uncertainty may however affect the performance of data assimilation techniques and post-processors, as well as the way they interact in the forecasting system. Further investigations would be necessary to better assess the extent to which this may affect forecast quality.

Although our findings may be related to the configuration used, they are based on common techniques and on the study of a large set of catchments, which helps in giving robustness and generality to the results obtained. The study also shows that, for a given system configuration, it is interesting to analyze how data assimilation and/or post-processing techniques set up to improve forecast quality affect the attributes of the forecasts and interact to provide overall good forecasts. The aim of a forecaster may then be to achieve a good combination of hydrological model, data assimilation and post-processing procedures that results in an overall good quality of his/her operational system (eventually over specific space and time

scales of interest), rather than to search for the best data assimilation technique or post-processor available, without taking into account how they will interact between them and with the probabilistic forecasting system as a whole.

Despite those limitations, our results strongly suggest that data assimilation and post-processing techniques based on hydrological uncertainty processors should be more widely tested to foster their implementation in pre-operational and operational hydrological ensemble forecasting systems and their use in real-time probabilistic forecasting. The use of both strategies is highly recommended since they have complementary effects: data assimilation has a very positive effect on forecast accuracy, and thus helps reduce hydrological uncertainty, but its impact diminishes with lead time, while post-processing, by accounting for hydrological uncertainty, has a very positive and longer lasting efficient effect on forecast reliability.

## 9.5 Appendix: Evaluation scores

The evaluation scores used in this article are defined and briefly described below. For more details, the reader may refer to [Wilks \(2011\)](#).

### 9.5.1 Relative bias

The relative bias (BIAS) is defined as the ratio between the mean of deterministic forecasts and the mean of observations,

$$\text{BIAS} = \frac{\sum_{k=1}^N Q_{fct}(k)}{\sum_{k=1}^N Q_{obs}(k)} \quad (9.1)$$

where  $(Q_{fct}(k), Q_{obs}(k))$  is the  $k$ th of  $N$  pairs of deterministic forecasts and observations.

Values higher (lower) than 1 indicate an overall overestimation (underestimation) of the observed values.

### 9.5.2 Normalized root-mean-square error

The root-mean-square error (RMSE) is a widely used measure of accuracy for point forecasts,

$$\text{RMSE} = \left[ \frac{1}{N} \sum_{k=1}^N (Q_{fct}(k) - Q_{obs}(k))^2 \right]^{1/2} \quad (9.2)$$

where  $(Q_{fct}(k), Q_{obs}(k))$  is the  $k$ th of  $N$  pairs of forecasts and observations.

The lower the RMSE, the better. For a perfect deterministic forecast,  $\text{RMSE} = 0$ .

The normalized root-mean-square error (NRMSE) is obtained by dividing the RMSE by the mean runoff. The use of a non-dimensional score facilitates the comparison of the results obtained over different catchments.

### 9.5.3 PIT diagram and alpha score

The Probability Integral Transform (PIT) diagram is a graphical tool used to assess the reliability of probabilistic forecasts (Gneiting *et al.*, 2007; Laio and Tamea, 2007). The PIT diagram corresponds to the empirical cumulative distribution of the PIT values, which are defined for each pair of forecasts and observations as the value that the cumulative predictive distribution  $F$  reaches at the observation,  $p^{obs} = F(Q_{obs})$ . It is analogous to a cumulated version of the rank histogram. If the forecasts are reliable, the PIT values follow a uniform distribution on the interval  $[0, 1]$  and the PIT curve is close to the 1:1 line. Reliability of the probabilistic forecasts implies that the observations should not be preferentially located in specific parts of the predictive distributions, but instead should uniformly span the whole predictive range.

The alpha score is an index proposed by Renard *et al.* (2010) to reflect the overall reliability of probabilistic forecasts. The alpha score is directly related to the PIT diagram. It is defined as  $1 - 2A$ , where  $A$  is the area between the bisector and the PIT curve,

$$A = \frac{1}{N} \sum_{k=1}^N \left| p^{obs}(k) - p^{th}(k) \right| \quad (9.3)$$

and where  $(p^{obs}(k), p^{th}(k))$  is the  $k$ th of  $N$  pairs of observed and theoretical PIT values.

The alpha score ranges from 0 to 1. 0 indicates poor reliability while values close to 1 indicate perfect reliability.

### 9.5.4 Normalized mean interquartile range

To assess the sharpness of probabilistic forecasts, we defined the mean interquartile range (MIQR) as the mean of the interquartile range of forecasts over the evaluation data. The interquartile range, defined as the range between the upper quartile (75th percentile) and the lower quartile (25th percentile) of a distribution, is a robust measure of the spread of a distribution. MIQR is computed as

$$\text{MIQR} = \frac{1}{N} \sum_{k=1}^N (Q_{fct}^{75}(k) - Q_{fct}^{25}(k)) \quad (9.4)$$

where  $(Q_{fct}^{25}(k), Q_{fct}^{75}(k))$  is the  $k$ th of  $N$  pairs of quartiles of the forecasts.

Similarly to the NRMSE, we divided the MIQR by the mean runoff to obtain an a non-dimensional score.

### 9.5.5 Mean CRPS and mean CRPSS

For a forecast–observation evaluation pair, the Continuous Rank Probability Score (CRPS) (e.g., [Matheson and Winkler, 1976](#); [Gneiting et al., 2007](#)) measures the quadratic distance between two cumulative distribution functions, the cumulative predictive distribution  $F(x)$  and a Heaviside function based on the observed value  $\mathbb{1}\{Q_{obs} \leq x\}$ :

$$\text{CRPS}(F, Q_{obs}) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}\{Q_{obs} \leq x\})^2 dx \quad (9.5)$$

The mean CRPS,  $\overline{\text{CRPS}}$ , is the average value of the CRPS over the  $N$  pairs of evaluation data:

$$\overline{\text{CRPS}} = \frac{1}{N} \sum_{k=1}^N \text{CRPS}(k) \quad (9.6)$$

The mean Continuous Rank Probability Skill Score (CRPSS) is a skill score based on the CRPS. Skill scores (SS) are used to assess the relative quality of two forecasting systems. They are generally defined as:

$$\text{SS} = 1 - \frac{\text{Score}^A}{\text{Score}^B} \quad (9.7)$$

where  $\text{Score}^A$  and  $\text{Score}^B$  are the scores of the forecasting system A and B respectively. The forecasting system B is usually termed the reference forecast.

Climatology is commonly used as a reference. To compute the mean CRPSS with the unconditional climatology as the reference, an unconditional streamflow ensemble forecast is first obtained from the empirical distribution of all observed discharges over the evaluation period, and then used for all forecast occasions.

## **9.6 Acknowledgments**

The authors thank Météo-France for providing the meteorological data and Banque HYDRO for the hydrological data. The financial support of SCHAPI to the first author is also gratefully acknowledged.

# CONCLUSION GÉNÉRALE



La quantification de l'incertitude prédictive en modélisation hydrologique a fait l'objet de nombreuses contributions. Il n'existe pas de consensus sur ce sujet, et diverses méthodes sont couramment utilisées.

Dans cette thèse, nous nous sommes intéressés à différentes méthodes de quantification de l'incertitude prédictive. Nous avons cherché à mettre en évidence, à l'aide d'une approche comparative, leurs forces et leurs faiblesses. L'objectif de nos travaux était de proposer, au final, une approche suffisamment simple et générale pour être applicable à des bassins versants variés, et à divers modèles, dans des contextes de simulation et de prévision.

## Résultats obtenus

Un premier aspect central de ce travail a été la mise en place de méthodes d'évaluation et le choix de métriques pour estimer les performances des méthodes de quantification des incertitudes testées. Cette question a été récurrente tout au long de ce travail et a mis en évidence la difficulté de sélectionner des approches d'évaluation rigoureuses et généralistes, tant les points de vue sur la question peuvent être variés.

Nous nous sommes en particulier interrogés sur l'interprétation du diagramme de PIT qui est couramment utilisé pour évaluer un attribut essentiel des prévisions probabilistes, la fiabilité. Nous avons mis en évidence que certaines limitations du diagramme de PIT peuvent mener à une évaluation trop optimiste de certains systèmes de prévision, et qu'une analyse plus détaillée est nécessaire pour éviter ces écueils. Nous nous sommes appuyés sur ces travaux pour définir un indice de fiabilité utilisé lors de la comparaison des différentes méthodes testées.

Dans un contexte de simulation hydrologique, nos travaux ont montré que les approches qui ne s'appuient que sur un ensemble de différents jeux de paramètres ne parviennent pas, en général, à représenter de manière adéquate l'incertitude prédictive totale. A cet égard, l'approche bayésienne est plus satisfaisante que la méthode GLUE, car elle complète l'incertitude liée à l'estimation des paramètres par une incertitude globale portée par un modèle d'erreur. Les méthodes de post-traitement s'appuient uniquement sur un modèle d'erreur extérieur au fonctionnement interne du modèle hydrologique. Lorsqu'elles sont suffisamment flexibles, ces méthodes parviennent à caractériser les erreurs résiduelles obtenues lors du calage du modèle hydrologique utilisé, ce qui permet de refléter les marges d'erreurs effectivement constatées.

Nous nous sommes appuyés sur ces résultats pour proposer une méthode de quantification



de l'incertitude prédictive pour les bassins non-jaugés. Cette nouvelle approche repose sur un transfert géographique des erreurs relatives constatées pour des stations jaugés aux stations non-jaugés. En cela, la méthode étend et complète les approches de régionalisation utilisées pour obtenir des simulations pour les bassins non-jaugés. Cette approche contraste notamment avec les méthodes proposées jusqu'ici, qui s'appuient uniquement sur une quantification de l'incertitude liée à l'estimation des paramètres et qui sont donc sujettes aux mêmes limitations que les méthodes similaires appliquées pour les bassins jaugés.

Dans un contexte de prévision, nous avons montré l'importance de la prise en compte de l'évolution de l'incertitude prédictive en fonction de l'échéance de prévision. Nous avons également mis en évidence les gains de performance qui peuvent être obtenus lorsque la quantification de l'incertitude s'appuie sur une meilleure caractérisation de la situation de prévision.

Enfin, nous avons étudié les interactions entre l'assimilation de données et le post-traitement d'un modèle de prévision au sein d'une chaîne de prévision hydrologique d'ensemble. Les résultats obtenus indiquent que les deux stratégies agissent de manière complémentaire ; leur utilisation est donc recommandée dans une chaîne de prévision hydrologique d'ensemble.

## **Quelques perspectives**

Nous évoquons à présent quelques perspectives aux travaux menés pendant cette thèse.

En ce qui concerne la quantification de l'incertitude prédictive en simulation hydrologique, certains points pourraient faire l'objet de travaux complémentaires.

D'abord, une combinaison de l'approche multi-modèles avec celle du post-traitement mériterait d'être étudiée. On pourrait par exemple envisager d'appliquer une méthode de post-traitement à chacun des modèles hydrologiques utilisés, et considérer une distribution prédictive totale composée des distributions obtenues pour chaque modèle. Malgré de bonnes performances globales, l'approche multi-modèles souffre en effet d'un manque de fiabilité qui pourrait être corrigé grâce à l'utilisation d'une méthode de post-traitement.

Par ailleurs, il nous semble que la méthode proposée pour quantifier l'incertitude prédictive des bassins non-jaugés mériterait d'être comparée aux autres méthodes proposées dans la littérature, afin de mieux mettre en évidence les avantages et les limitations des différentes approches. Il serait intéressant d'étudier dans quelle mesure la méthode proposée permet d'améliorer la fiabilité de l'incertitude prédictive par rapport aux méthodes qui ne s'appuient que sur des ensembles de paramètres.

En ce qui concerne la quantification de l'incertitude prédictive en prévision, les travaux de comparaison des méthodes de post-traitement pourraient être poursuivis dans le cadre de l'inter-comparaison actuellement menée au sein de l'initiative HEPEX (Van Andel *et al.*, 2013). D'autre part, la poursuite du développement d'une méthode qui s'appuie sur une meilleure caractérisation du contexte hydrologique de prévision nous semble essentielle et prioritaire. Ces développements pourraient s'appuyer sur une méthode d'identification des états internes du modèle les plus informatifs pour distinguer des distributions d'erreurs aux caractéristiques différentes les unes des autres, ainsi que sur des méthodes statistiques non-paramétriques pour caractériser ces distributions.

Enfin, la mise en œuvre opérationnelle des méthodes de post-traitement que nous avons testées dans cette thèse pourrait contribuer à accompagner les Services de Prévision des Crues (SPC) et le Service Central Hydrométéorologique d'Appui à la Prévision des Inondations (SCHAPI) dans leur évolution vers un affichage et une communication des prévisions des crues et des incertitudes associées.



## RÉFÉRENCES BIBLIOGRAPHIQUES

- AJAMI, N. K., Q. Y. DUAN et S. SOROOSHIAN : An integrated hydrologic Bayesian multimodel combination framework : Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1):W01403, 2007. Cité pp. 10, 17 et 19.
- ANDRÉASSIAN, V., J. LERAT, C. LOUMAGNE, T. MATHEVET, C. MICHEL, L. OUDIN et C. PERRIN : What is really undermining hydrologic science today? *Hydrological Processes*, 21:2819–2822, 2007. Cité pp. 4, 14, 107 et 143.
- BARDOSSY, A. : Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, 11(2):703–710, 2007. Cité p. 15.
- BATES, B. C. et E. P. CAMPBELL : A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, 37(4):937–947, 2001. Cité p. 16.
- BECK, M. B. : Water quality modeling : A review of the analysis of uncertainty. *Water Resources Research*, 23(8):1393–1442, 1987. Cité p. 9.
- BENNETT, N. D., B. F. W. CROKE, G. GUARISO, J. H. A. GUILLAUME, S. H. HAMILTON, A. J. JAKEMAN, S. MARSILI-LIBELLI, L. T. H. NEWHAM, J. P. NORTON, C. PERRIN, S. A. PIERCE, B. ROBSON, R. SEPPELT, A. A. VOINOV, B. D. FATH et V. ANDRÉASSIAN : Characterising performance of environmental models. *Environmental Modelling & Software*, 40:1–20, 2013. Cité pp. 3 et 45.
- BERTHET, L. : *Prévision des crues à pas de temps fin : comment mieux assimiler l'information de débit dans un modèle Pluie - Débit?* Thèse de doctorat, AgroParisTech/ENGREF, Paris, 2010, 603 pp. Cité pp. 27, 37, 38, 123, 125 et 192.
- BERTHET, L., V. ANDRÉASSIAN, C. PERRIN et P. JAVELLE : 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, 2009. Cité pp. 60 et 143.

- BERTHIER, C.-H. : Quantification des incertitudes des débits calculés par un modèle pluie-débit empirique. Mémoire de D.E.A., Université Paris-Sud 11, 2005, 55 pp. Cité pp. 14, 79 et 201.
- BEVEN, K. : A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2):18–36, 2006a. Cité pp. 14, 15 et 18.
- BEVEN, K. : On undermining the science? *Hydrological Processes*, 20(14):3141–3146, 2006b. Cité p. 4.
- BEVEN, K. et A. BINLEY : The future of distributed models : model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298, 1992. Cité p. 13.
- BEVEN, K. et J. FREER : 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, 2001. Cité pp. 14, 71 et 106.
- BEVEN, K. J. : Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J. F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson. *Stochastic Environmental Research and Risk Assessment*, 23(7):1059–1060, 2009a. Cité p. 11.
- BEVEN, K. J. : *Environmental modelling : An uncertain future ?* Routledge, London, 2009b. ISBN 9780415457590. Cité pp. 3 et 13.
- BEVEN, K. J. et M. J. KIRKBY : A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69, 1979. Cité p. 109.
- BEVEN, K., P. SMITH, I. WESTERBERG et J. FREER : Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al. *Water Resources Research*, 48:W11801, 2012. Cité p. 5.
- BIEN, J. et R. TIBSHIRANI : Hierarchical Clustering With Prototypes via Minimax Linkage. *Journal of the American Statistical Association*, 106(495):1075–1084, 2011. Cité p. 29.
- BIONDI, D. et D. L. DE LUCA : A Bayesian approach for real-time flood forecasting. *Physics and Chemistry of the Earth*, 42-44:91–97, 2012. Cité p. 20.
- BLAZKOVA, S. et K. BEVEN : A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation : Skalka catchment, Czech Republic. *Water Resources Research*, 45:W00B16, 2009. Cité p. 15.

- BLÖSCHL, G., M. SIVAPALAN, T. WAGENER, A. VIGLIONE et H. SAVENIJE : *Runoff Prediction in Ungauged Basins : Synthesis Across Processes, Places and Scales*. Cambridge University Press, 2013. ISBN 1107028183. Cité p. 106.
- BOGNER, K. et F. PAPPENBERGER : Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research*, 47:W07524, 2011. Cité pp. 20 et 59.
- BONTRON, G. : *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, Grenoble, 2004, 276 pp. Cité pp. 47 et 49.
- BOUCHER, M. A., D. TREMBLAY, L. DELORME, L. PERREAULT et F. ANCTIL : Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology*, 416:133–144, 2012. Cité pp. 58, 124 et 139.
- BROWN, J. D. : Knowledge, uncertainty and physical geography : towards the development of methodologies for questioning belief. *Transactions of the Institute of British Geographers*, 29(3):367–381, 2004. Cité p. 11.
- BROWN, J. D. et D. J. SEO : Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrological Processes*, 27(1):83–105, 2013. Cité pp. 124 et 139.
- BROWN, J. D., J. DEMARGNE, D.-J. SEO et Y. LIU : The Ensemble Verification System (EVS) : A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling & Software*, 25(7):854–872, 2010. Cité p. 58.
- BROWN, J. D. et D.-J. SEO : A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts. *Journal of Hydrometeorology*, 11(3):642–665, JUN 2010. ISSN 1525-755X. Cité p. 139.
- BULYGINA, N., C. BALLARD, N. MCINTYRE, G. O'DONNELL et H. WHEATER : Integrating different types of information into hydrological model parameter estimation : Application to ungauged catchments and land use scenario analysis. *Water Resources Research*, 48:W06519, 2012. Cité p. 106.
- BULYGINA, N., N. MCINTYRE et H. WHEATER : Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. *Hydrology and Earth System Sciences*, 13(6):893–904, 2009. Cité p. 15.

- BULYGINA, N., N. MCINTYRE et H. WHEATER : Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. *Water Resources Research*, 47:W02503, 2011. Cité p. 106.
- CLARK, M. P., D. KAVETSKI et F. FENICIA : Reply to comment by K. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling". *Water Resources Research*, 48:W11802, 2012. Cité p. 5.
- CLOKE, H. L. et F. PAPPENBERGER : Ensemble flood forecasting : A review. *Journal of Hydrology*, 375(3-4):613–626, 2009. Cité pp. 11 et 138.
- COCCIA, G. et E. TODINI : Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences*, 15 (10):3253–3274, 2011. Cité pp. 80, 124, 138, 204 et 205.
- CROCHEMORE, L., C. PERRIN, V. ANDRÉASSIAN, U. EHRET, S. SEIBERT, S. GRIMALDI, H. GUPTA et J.-E. PATUREL : Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal*, 2014. Cité p. 45.
- DECHANT, C. M. et H. MORADKHANI : Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48:W04518, 2012. Cité p. 59.
- DIKS, C. G. H. et J. A. VRUGT : Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment*, 24(6):809–820, 2010. Cité p. 19.
- DUAN, Q., N. K. AJAMI, X. GAO et S. SOROOSHIAN : Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5):1371–1386, 2007. Cité pp. 19 et 106.
- ENGELAND, K., B. RENARD, I. STEINSLAND et S. KOLBERG : Evaluation of statistical models for forecast errors from the HBV model. *Journal of Hydrology*, 384(1-2):142–155, 2010. Cité p. 59.
- ENGELAND, K., C. Y. XU et L. GOTTSCHALK : Assessing uncertainties in a conceptual water balance model using Bayesian methodology. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 50(1):45–63, 2005. Cité p. 16.

- EWEN, J. et G. O'DONNELL : Prediction intervals for rainfall-runoff models : raw error method and split-sample validation. *Hydrology Research*, 43(5):637–648, 2012. Cité pp. 106, 107, 124 et 138.
- FORTIN, V., A.-C. FAVRE et M. SAID : Probabilistic forecasting from ensemble prediction systems : Improving upon the best-member method by using a different weight and dressing kernel for each member. *Quarterly Journal of the Royal Meteorological Society*, 132(617):1349–1369, 2006. Cité pp. 124 et 139.
- FRANZ, K. J. et T. S. HOGUE : Evaluating uncertainty estimates in hydrologic models : borrowing measures from the forecast verification community. *Hydrology and Earth System Sciences*, 15(11):3367–3382, 2011. Cité pp. 43 et 58.
- FREER, J., K. BEVEN et B. AMBROISE : Bayesian estimation of uncertainty in runoff prediction and the value of data : An application of the GLUE approach. *Water Resources Research*, 32(7):2161–2173, 1996. Cité p. 14.
- GARÇON, R., B. HOUDANT, F. GARAVAGLIA, T. MATHEVET, E. PAQUET et J. GAILHARD : Human assessment of hydrometeorological forecasts and communication of their uncertainties in a decision making context. *Houille Blanche-Revue Internationale De L Eau*, 5(5):71–80, 2009. Cité p. 130.
- GELMAN, A., J. B. CARLIN, H. S. STERN et D. B. RUBIN : *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd édn, 2004. ISBN 1-584-88388-X (rel). Cité p. 75.
- GEORGAKAKOS, K. P., D. J. SEO, H. GUPTA, J. SCHAAKE et M. B. BUTTS : Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298(1-4):222–241, 2004. Cité p. 19.
- GNEITING, T., F. BALABDAOUI et A. E. RAFTERY : Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 69:243–268, 2007. Cité pp. 58, 156 et 157.
- GNEITING, T. et A. E. RAFTERY : Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. Cité pp. 46, 47, 51 et 114.
- GOTZINGER, J. et A. BARDOSSY : Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resources Research*, 44:W00B07, 2008. Cité p. 10.



- HAARIO, H., E. SAKSMAN et J. TAMMINEN : An adaptive Metropolis algorithm. *Bernoulli*, 7 (2):223–242, 2001. Cité p. 75.
- HAARIO, H., M. LAINE, A. MIRA et E. SAKSMAN : DRAM : Efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006. Cité p. 75.
- HALL, J., E. O'CONNELL et J. EWEN : On not undermining the science : coherence, validation and expertise. Discussion of Invited Commentary by Keith Beven Hydrological Processes, 20, 3141-3146 (2006). *Hydrological Processes*, 21(7):985–988, 2007. Cité p. 4.
- HARRISON, K. W., S. V. KUMAR, C. D. PETERS-LIDARD et J. A. SANTANELLO : Quantifying the change in soil moisture modeling uncertainty from remote sensing observations using Bayesian inference techniques. *Water Resources Research*, 48:W11514, 2012. Cité p. 75.
- HOETING, J. A., D. MADIGAN, A. E. RAFTERY et C. T. VOLINSKY : Bayesian model averaging : A tutorial. *Statistical Science*, 14(4):382–401, 1999. Cité p. 19.
- HOPSON, T. M. et P. J. WEBSTER : A 1-10-Day Ensemble Forecasting Scheme for the Major River Basins of Bangladesh : Forecasting Severe Floods of 2003-07. *Journal of Hydrometeorology*, 11(3):618–641, 2010. Cité pp. 124 et 139.
- HORNBERGER, G. M. et R. C. SPEAR : An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 12(1):7–18, 1981. Cité p. 14.
- HRACHOWITZ, M., H. H. G. SAVENIJE, G. BLÖSCHL, J. J. MCDONNELL, M. SIVAPALAN, J. W. POMEROY, B. ARHEIMER, T. BLUME, M. P. CLARK, U. EHRET, F. FENICIA, J. E. FREER, A. GELFAN, H. V. GUPTA, D. A. HUGHES, R. W. HUT, A. MONTANARI, S. PANDE, D. TETZLAFF, P. A. TROCH, S. UHLENBROOK, T. WAGENER, H. C. WINSEMIUS, R. A. WOODS, E. ZEHE et C. CUDENNEC : A decade of Predictions in Ungauged Basins (PUB)a review. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 58(6):1198–1255, 2013. Cité p. 106.
- HSU, K.-I., H. MORADKHANI et S. SOROOSHIAN : A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, 45:W00B12, 2009. Cité p. 19.
- HUARD, D. et A. MAILHOT : A Bayesian perspective on input uncertainty in model calibration : Application to hydrological model "abc". *Water Resources Research*, 42(7):W07416, 2006. Cité p. 17.
- HUARD, D. et A. MAILHOT : Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Research*, 44(2):W02424, 2008. Cité p. 17.

- JAKEMAN, A. J., R. A. LETCHER et J. P. NORTON : Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21(5):602–614, 2006. Cité p. 3.
- JALBERT, J., T. MATHEVET et A. C. FAVRE : Temporal uncertainty estimation of discharges from rating curves using a variographic analysis. *Journal of Hydrology*, 397(1-2):83–92, 2011. Cité p. 10.
- JIN, X. L., C. Y. XU, Q. ZHANG et V. P. SINGH : Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, 383(3-4):147–155, 2010. Cité pp. 45 et 46.
- JOLLIFFE, I. T. et D. B. STEPHENSON : *Forecast verification : a practitioner's guide in atmospheric science*. Wiley-Blackwell, Oxford, 2nd édn, 2012. Cité p. 45.
- KAPANGAZIWIRI, E., D. A. HUGHES et T. WAGENER : Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 57(5):1000–1019, 2012. Cité p. 106.
- KAVETSKI, D., G. KUCZERA et S. W. FRANKS : Bayesian analysis of input uncertainty in hydrological modeling : 1. Theory. *Water Resources Research*, 42(3):W03407, 2006a. Cité pp. 17, 74 et 198.
- KAVETSKI, D., G. KUCZERA et S. W. FRANKS : Bayesian analysis of input uncertainty in hydrological modeling : 2. Application. *Water Resources Research*, 42(3):W03408, 2006b. Cité p. 17.
- KAVETSKI, D. et F. FENICIA : Elements of a flexible approach for conceptual hydrological modeling : 2. Application and experimental insights. *Water Resources Research*, 47:W11511, 2011. Cité p. 10.
- KELLY, K. S. et R. KRZYSZTOFOWICZ : A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics*, 11(1):17–31, 1997. Cité pp. 80, 200 et 203.
- KOSKELA, J. J., B. W. F. CROKE, H. KOIVUSALO, A. J. JAKEMAN et T. KOKKONEN : Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment. *Water Resources Research*, 48:W11513, 2012. Cité p. 10.
- KRUEGER, T., J. FREER, J. N. QUINTON, C. J. A. MACLEOD, G. S. BILOTTA, R. E. BRAZIER, P. BUTLER et P. M. HAYGARTH : Ensemble evaluation of hydrological model hypotheses. *Water Resources Research*, 46:W07516, 2010. Cité p. 15.

- KRZYSZTOFOWICZ, R. : Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9):2739–2750, 1999. Cité pp. 20, 60, 123, 124, 138 et 154.
- KRZYSZTOFOWICZ, R. : The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1-4):2–9, 2001. Cité p. 20.
- KRZYSZTOFOWICZ, R. et C. J. MARANZANO : Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology*, 293(1-4):57–73, 2004. Cité p. 124.
- KUCZERA, G., D. KAVETSKI, S. FRANKS et M. THYER : Towards a Bayesian total error analysis of conceptual rainfall-runoff models : Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331(1-2):161–177, 2006. Cité pp. 17 et 18.
- KUCZERA, G. et E. PARENT : Monte Carlo assessment of parameter uncertainty in conceptual catchment models : the Metropolis algorithm. *Journal of Hydrology*, 211(1-4):69–85, 1998. Cité p. 16.
- KUCZERA, G., D. KAVETSKI, B. RENARD et M. THYER : A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. *Water Resources Research*, 46:W07602, 2010. Cité p. 18.
- LAIO, F. et S. TAMEA : Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, 11(4):1267–1277, 2007. Cité pp. 43, 47, 58, 60, 146 et 156.
- LALOY, E. et J. A. VRUGT : High-dimensional posterior exploration of hydrologic models using multiple-try DREAM((ZS)) and high-performance computing. *Water Resources Research*, 48:W01526, 2012. Cité pp. 16 et 75.
- LE COZ, J., B. RENARD, L. BONNIFAIT, F. BRANGER et R. LE BOURSICAUD : Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves : A Bayesian approach. *Journal of Hydrology*, 509:573–587, 2014. Cité p. 10.
- 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 ?* Thèse de doctorat, Université Pierre et Marie Curie, Paris, 2008, 324 pp. Cité pp. 26, 30, 36 et 75.
- LE MOINE, N., V. ANDRÉASSIAN, C. PERRIN et C. MICHEL : How can rainfall-runoff models handle intercatchment groundwater flows ? Theoretical study based on 1040 French catchments. *Water Resources Research*, 43(6):W06428, 2007. Cité p. 30.

- LEEDAL, D., A. H. WEERTS, P. J. SMITH et K. J. BEVEN : Application of data-based mechanistic modelling for flood forecasting at multiple locations in the Eden catchment in the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences*, 17 (1):177–185, 2013. Cité p. 124.
- LI, Z., Q. SHAO, Z. XU et X. CAI : Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method : A case study of SWAT model applied to Yingluoxia watershed in northwest China. *Journal of Hydrology*, 385(1-4):76–83, 2010. Cité p. 14.
- LIU, Y., A. H. WEERTS, M. CLARK, H. J. H. FRANSSSEN, S. KUMAR, H. MORADKHANI, D. J. SEO, D. SCHWANENBERG, P. SMITH, A. van DIJK, N. van VELZEN, M. HE, H. LEE, S. J. NOH, O. RAKOVEC et P. RESTREPO : Advancing data assimilation in operational hydrologic forecasting : progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10):3863–3887, 2012. Cité p. 139.
- LIU, Y. L., J. FREER, K. BEVEN et P. MATGEN : Towards a limits of acceptability approach to the calibration of hydrological models : Extending observation error. *Journal of Hydrology*, 367 (1-2):93–103, 2009. Cité p. 15.
- LIU, Y. Q. et H. V. GUPTA : Uncertainty in hydrologic modeling : Toward an integrated data assimilation framework. *Water Resources Research*, 43(7):W07401, 2007. Cité p. 106.
- LU, D., M. YE et M. C. HILL : Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification. *Water Resources Research*, 48:20, 2012. Cité p. 74.
- MANTOVAN, P. et E. TODINI : Hydrological forecasting uncertainty assessment : Incoherence of the GLUE methodology. *Journal of Hydrology*, 330(1-2):368–381, 2006. Cité p. 15.
- MANTOVAN, P., E. TODINI et M. L. V. MARTINA : Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment : Incoherence of the GLUE methodology". *Journal of Hydrology*, 338(3-4):319–324, 2007. Cité p. 15.
- MARANZANO, C. J. et R. KRZYSZTOFOWICZ : Identification of likelihood and prior dependence structures for hydrologic uncertainty processor. *Journal of Hydrology*, 290(1-2):1–21, 2004. Cité p. 20.
- MARSHALL, L., D. NOTT et A. SHARMA : Towards dynamic catchment modelling : a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes*, 21(7):847–861, 2007. Cité pp. 19 et 75.

- MASCARO, G., E. R. VIVONI et R. DEIDDA : Implications of Ensemble Quantitative Precipitation Forecast Errors on Distributed Streamflow Forecasting. *Journal of Hydrometeorology*, 11 (1):69–86, 2010. Cité p. 61.
- MATHESON, J. E. et R. L. WINKLER : Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976. Cité p. 157.
- MATHEVET, T. : *Quels modèles pluie-débit globaux au pas de temps horaire ? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants*. Thèse de doctorat, ENGREF, Paris, 2005, 463 pp. Cité pp. 36, 54, 78 et 109.
- MATHEVET, T., C. MICHEL, V. ANDRÉASSIAN et C. PERRIN : A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS-AISH Publication*, (307):211–219, 2006. Cité p. 118.
- MATOTT, L. S., J. E. BABENDREIER et S. T. PURUCKER : Evaluating uncertainty in integrated environmental models : A review of concepts and tools. *Water Resources Research*, 45: W06421, 2009. Cité pp. 9, 12 et 106.
- MCINTYRE, N., H. LEE, H. WHEATER, A. YOUNG et T. WAGENER : Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, 41(12):W12434, 2005. Cité pp. 45, 106 et 107.
- MCMILLAN, H. K., M. P. CLARK, W. B. BOWDEN, M. DUNCAN et R. A. WOODS : Hydrological field data from a modeller's perspective : Part 1. Diagnostic tests for model structure. *Hydrological Processes*, 25(4), 2011. Cité p. 10.
- MICHEL, C., C. PERRIN et V. ANDRÉASSIAN : The exponential store : a correct formulation for rainfall-runoff modelling. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 48(1):109–124, 2003. Cité p. 109.
- MONTANARI, A. : Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 41(8):W08406, 2005. Cité p. 14.
- MONTANARI, A. : What do we mean by 'uncertainty' ? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes*, 21(6):841–845, 2007. Cité p. 4.
- MONTANARI, A. : Uncertainty of Hydrological Predictions. In PETER, W., éd. : *Treatise on Water Science*, p. 459–478. Elsevier, Oxford, 2011. ISBN 978-0-444-53199-5. Cité pp. 4, 12 et 106.

- MONTANARI, A. et A. BRATH : A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 40(1):W01106, 2004. Cité pp. 45, 80, 106, 107 et 138.
- MONTANARI, A. et G. GROSSI : Estimating the uncertainty of hydrological forecasts : A statistical approach. *Water Resources Research*, 44:W00B08, 2008. Cité pp. 80, 124, 138 et 203.
- MONTANARI, A., C. A. SHOEMAKER et N. van de GIESEN : Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology : An overview of issues and challenges. *Water Resources Research*, 45:W00B00, 2009. Cité pp. 9 et 11.
- MORADKHANI, H., C. M. DECHANT et S. SOROOSHIAN : Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, 48:W12520, 2012. Cité pp. 17 et 139.
- MORADKHANI, H., K. L. HSU, H. GUPTA et S. SOROOSHIAN : Uncertainty assessment of hydrologic model states and parameters : Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5):W05012, 2005. Cité p. 17.
- MORAWIETZ, M., C.-Y. XU, L. GOTTSCHALK et L. M. TALLAKSEN : Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. *Journal of Hydrology*, 407(1-4):58–72, 2011. Cité pp. 124 et 138.
- MOULIN, L., E. GAUME et C. OBLÉD : Uncertainties on mean areal precipitation : assessment and impact on streamflow simulations. *Hydrology and Earth System Sciences*, 13(2):99–114, 2009. Cité p. 10.
- NASH, J. E. et J. V. SUTCLIFFE : River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970. Cité pp. 53, 109 et 118.
- NICOLAU, J. : Short-range ensemble forecasting. In *Proceedings WMO/CBS Technical Conferences On Data Processing and Forecasting Systems, Cairns, Australia*, p. 2–3, 2002. Cité p. 142.
- OUDIN, L., V. ANDRÉASSIAN, C. PERRIN, C. MICHEL et N. LE MOINE : Spatial proximity, physical similarity, regression and ungaged catchments : A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44(3):W03413, 2008. Cité pp. 108 et 109.



- LOUDIN, L., F. HERVIEU, C. MICHEL, C. PERRIN, V. ANDRÉASSIAN, F. ANCTIL et C. LOUMAGNE : 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(1-4):290–306, 2005. Cité pp. 59, 108 et 141.
- LOUDIN, L., V. ANDRÉASSIAN, T. MATHEVET, C. PERRIN et C. MICHEL : Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42(7):W07410, 2006. Cité pp. 75 et 109.
- LOUDIN, L., A. KAY, V. ANDRÉASSIAN et C. PERRIN : Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, 46:W11558, 2010. Cité p. 109.
- PAGANO, T. C., D. L. SHRESTHA, Q. J. WANG, D. ROBERTSON et P. HAPUARACHCHI : Ensemble dressing for hydrological applications. *Hydrological Processes*, 27(1):106–116, 2013. Cité p. 140.
- PAPPENBERGER, F. et K. J. BEVEN : Ignorance is bliss : Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(5):W05302, 2006. Cité p. 4.
- PARAJKA, J., A. VIGLIONE, M. ROGGER, J. L. SALINAS, M. SIVAPALAN et G. BLÖSCHL : Comparative assessment of predictions in ungauged basins - Part 1 : Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, 17(5):1783–1795, 2013. Cité p. 106.
- PERRIN, C. : *Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative*. Thèse de doctorat, INPG, Grenoble, 2000, 530 pp. Cité pp. 78 et 79.
- PERRIN, C., C. MICHEL et V. ANDRÉASSIAN : Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242(3-4):275–301, 2001. Cité p. 78.
- PERRIN, C., C. MICHEL et V. ANDRÉASSIAN : Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289, 2003. Cité pp. 35, 109, 142 et 188.
- PERRIN, C., V. ANDRÉASSIAN, C. R. SERNA, T. MATHEVET et N. LE MOINE : Discrete parameterization of hydrological models : Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research*, 44(8):W08447, 2008. Cité pp. 36 et 109.
- PIANOSI, F. et L. RASO : Dynamic modeling of predictive uncertainty by regression on absolute errors. *Water Resources Research*, 48:W03516, 2012. Cité pp. 124 et 138.

- RAMOS, M. H., S. J. van ANDEL et F. PAPPENBERGER : Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences*, 17(6):2219–2232, 2013. Cité p. 58.
- RAMOS, M.-H., T. MATHEVET, J. THIELEN et F. PAPPENBERGER : Communicating uncertainty in hydro-meteorological forecasts : mission impossible ? *Meteorological Applications*, 17(2):223–235, 2010. Cité pp. 4 et 58.
- RANDRIANASOLO, A., M. H. RAMOS, G. THIREL, V. ANDRÉASSIAN et E. MARTIN : Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmospheric Science Letters*, 11(2):100–107, 2010. Cité pp. 11 et 142.
- REFSGAARD, J. C. et H. J. HENRIKSEN : Modelling guidelines - terminology and guiding principles. *Advances in Water Resources*, 27(1):71–82, 2004. Cité p. 3.
- REFSGAARD, J. C., H. J. HENRIKSEN, W. G. HARRAR, H. SCHOLTEN et A. KASSAHUN : Quality assurance in model based water management - review of existing practice and outline of new approaches. *Environmental Modelling & Software*, 20(10):1201–1215, 2005. Cité p. 3.
- REFSGAARD, J. C., J. P. van der SLUIJS, A. L. HOJBERG et P. A. VANROLLEGHEM : Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software*, 22(11):1543–1556, 2007. Cité pp. 3, 9 et 106.
- REGGIANI, P., M. RENNER, A. H. WEERTS et P. van GELDER : Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resources Research*, 45:W02428, 2009. Cité p. 139.
- REGGIANI, P. et A. H. WEERTS : A Bayesian approach to decision-making under uncertainty : An application to real-time forecasting in the river Rhine. *Journal of Hydrology*, 356(1-2):56–69, 2008. Cité p. 20.
- REGONDA, S. K., D.-J. SEO, B. LAWRENCE, J. D. BROWN et J. DEMARGNE : Short-term ensemble streamflow forecasting using operationally-produced single-valued streamflow forecasts - A Hydrologic Model Output Statistics (HMOS) approach. *Journal of Hydrology*, 497:80–96, 2013. Cité p. 124.
- REICHERT, P. et J. MIELEITNER : Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resources Research*, 45:W10402, 2009. Cité p. 18.



- RENARD, B., D. KAVETSKI, G. KUCZERA, M. THYER et S. W. FRANKS : Understanding predictive uncertainty in hydrologic modeling : The challenge of identifying input and structural errors. *Water Resources Research*, 46:W05521, 2010. Cité pp. 17, 18, 48, 49, 51, 60, 146 et 156.
- RENARD, B., D. KAVETSKI, E. LEBLOIS, M. THYER, G. KUCZERA et S. W. FRANKS : Toward a reliable decomposition of predictive uncertainty in hydrological modeling : Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47:W11516, 2011. Cité pp. 10 et 18.
- ROSCOE, K. L., A. H. WEERTS et M. SCHROEVERS : Estimation of the uncertainty in water level forecasts at ungauged river locations using quantile regression. *International Journal of River Basin Management*, 10(4):383–394, 2012. Cité p. 107.
- SALAMON, P. et L. FEYEN : Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. *Water Resources Research*, 46:W12501, 2010. Cité pp. 17, 18 et 139.
- SCHAAKE, J. C., T. M. HAMILL, R. BUIZZA et M. CLARK : The hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society*, 88(10):1541–1547, 2007. Cité p. 11.
- SCHAEFLI, B., D. B. TALAMBA et A. MUSY : Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology*, 332(3-4):303–315, 2007. Cité p. 17.
- SCHOUPS, G. et J. A. VRUGT : A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46:W10531, 2010. Cité pp. 17, 59 et 154.
- SMITH, P., K. J. BEVEN et J. A. TAWN : Informal likelihood measures in model assessment : Theoretic development and investigation. *Advances in Water Resources*, 31(8):1087–1100, 2008. Cité pp. 14 et 16.
- SMITH, P. J., K. J. BEVEN, A. H. WEERTS et D. LEEDAL : Adaptive correction of deterministic models to produce probabilistic forecasts. *Hydrology and Earth System Sciences*, 16(8):2783–2799, 2012. Cité pp. 124 et 138.
- SMITH, T. J. et L. A. MARSHALL : Bayesian methods in hydrologic modeling : A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resources Research*, 44:W00B05, 2008. Cité p. 75.

- SOLOMATINE, D. P. et D. L. SHRESTHA : A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, 45:W00B11, 2009. Cité pp. 20, 45, 106, 107, 124 et 138.
- SOROOSHIAN, S. et J. A. DRACUP : Stochastic parameter estimation procedures for hydrologic rainfall-runoff models : Correlated and heteroscedastic error cases. *Water Resources Research*, 16(2):430–442, 1980. Cité p. 17.
- STEDINGER, J. R., R. M. VOGEL, S. U. LEE et R. BATCHELDER : Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, 44:W00B06, 2008. Cité p. 15.
- TABARY, P., P. DUPUY, G. L'HENAFF, C. GUEGUEN, L. MOULIN, O. LAURANTIN, C. MERLIER et J.-M. SOUBEYROUX : A 10-year (1997-2006) reanalysis of Quantitative Precipitation Estimation over France : methodology and first results. In *Weather Radar and Hydrology*, vol. 351, p. 255–260. IAHS, 2012. ISBN 0144-7815. Cité pp. 26, 59 et 141.
- TANGARA, M. : *Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global*. Thèse de doctorat, EPHE, Paris, 2005, 374 pp. Cité p. 37.
- THIELEN, J., J. SCHAAKE, R. HARTMAN et R. BUIZZA : Aims, challenges and progress of the Hydrological Ensemble Prediction Experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. *Atmospheric Science Letters*, 9(2):29–35, 2008. Cité p. 11.
- THIREL, G., F. REGIMBEAU, E. MARTIN, J. NOILHAN et F. HABETS : Short- and medium-range hydrological ensemble forecasts over France. *Atmospheric Science Letters*, 11(2):72–77, 2010. Cité p. 11.
- THIREL, G., F. ROUSSET-REGIMBEAU, E. MARTIN et F. HABETS : On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Predictions. *Journal of Hydrometeorology*, 9(6):1301–1317, 2008. Cité pp. 11 et 142.
- THYER, M., B. RENARD, D. KAVETSKI, G. KUCZERA, S. W. FRANKS et S. SRIKANTHAN : Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling : A case study using Bayesian total error analysis. *Water Resources Research*, 45:W00B14, 2009. Cité pp. 17, 18, 59, 75 et 146.
- TODINI, E. : Hydrological catchment modelling : past, present and future. *Hydrology and Earth System Sciences*, 11(1):468–482, 2007. Cité p. 4.

- VAN ANDEL, S. J., A. WEERTS, J. SCHAAKE et K. BOGNER : Post-processing hydrological ensemble predictions intercomparison experiment. *Hydrological Processes*, 27(1):158–161, 2013. Cité pp. 134 et 163.
- VAN STEENBERGEN, N., J. RONSYN et P. WILLEMS : A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software*, 33:92–105, 2012. Cité pp. 124 et 138.
- VELAZQUEZ, J. A., F. ANCTIL et C. PERRIN : Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments. *Hydrology and Earth System Sciences*, 14(11):2303–2317, 2010. Cité pp. 19, 78, 94 et 106.
- VERKADE, J. S., J. D. BROWN, P. REGGIANI et A. H. WEERTS : Post-processing ecmwf precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology*, 501:73–91, 2013. Cité p. 154.
- VIDAL, J.-P., E. MARTIN, L. FRANCHISTEGUY, M. BAILLON et J.-M. SOUBEYROUX : A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30(11):1627–1644, 2010. Cité pp. 108 et 141.
- VRUGT, J. A., H. V. GUPTA, L. A. BASTIDAS, W. BOUTEN et S. SOROOSHIAN : Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, 39(8):1214, 2003a. Cité pp. 16 et 18.
- VRUGT, J. A., H. V. GUPTA, W. BOUTEN et S. SOROOSHIAN : A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8):1201, 2003b. Cité p. 74.
- VRUGT, J. A., C. J. F. ter BRAAK, M. P. CLARK, J. M. HYMAN et B. A. ROBINSON : Treatment of input uncertainty in hydrologic modeling : Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44:W00B09, 2008. Cité pp. 10, 17 et 18.
- VRUGT, J. A., C. J. F. ter BRAAK, C. G. H. DIKS et G. SCHOUPS : Hydrologic data assimilation using particle Markov chain Monte Carlo simulation : Theory, concepts and applications. *Advances in Water Resources*, 51:457–478, 2013. Cité pp. 17 et 139.
- VRUGT, J. A., C. J. F. ter BRAAK, H. V. GUPTA et B. A. ROBINSON : Response to comment by Keith Beven on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?". *Stochastic Environmental Research and Risk Assessment*, 23(7):1061–1062, 2009. Cité pp. 45 et 75.

- WAGENER, T. et A. MONTANARI : Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, 47:W06301, 2011. Cité pp. 15 et 107.
- WANG, X. et C. BISHOP : Improvement of ensemble reliability with a new dressing kernel. *Quarterly Journal of the Royal Meteorological Society*, 131(607, A):965–986, APR 2005. ISSN 0035-9009. Cité p. 139.
- WEERTS, A. H., H. C. WINSEMIUS et J. S. VERKADE : Estimation of predictive hydrological uncertainty using quantile regression : examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences*, 15(1):255–265, 2011. Cité pp. 59, 81, 106, 124, 138 et 206.
- WEERTS, A. H. et G. Y. H. EL SERAFY : Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, 42(9):W09403, 2006. Cité pp. 18 et 139.
- WILKS, D. S. : *Statistical methods in the atmospheric sciences*. Academic, Oxford, 3rd éd., 2011. Cité pp. 45, 58 et 155.
- WINSEMIUS, H. C., B. SCHAEFLI, A. MONTANARI et H. H. G. SAVENIJE : On the calibration of hydrological models in ungauged basins : A framework for integrating hard and soft hydrological information. *Water Resources Research*, 45:W12422, 2009. Cité pp. 15, 106 et 107.
- XIONG, L. H., M. WAN, X. J. WEI et K. M. O’CONNOR : Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 54(5):852–871, 2009. Cité p. 14.
- YADAV, M., T. WAGENER et H. GUPTA : Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8):1756–1774, 2007. Cité pp. 15 et 106.
- YAN, J., G. Y. LIAO, M. GEBREMICHAEL, R. SHEDD et D. R. VALLEE : Characterizing the uncertainty in river stage forecasts conditional on point forecast values. *Water Resources Research*, 48:W12509, 2012. Cité pp. 124 et 138.
- YANG, J., P. REICHERT, K. C. ABBASPOUR, J. XIA et H. YANG : Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology*, 358(1-2):1–23, 2008. Cité p. 45.

- YANG, J., P. REICHERT, K. C. ABBASPOUR et H. YANG : Hydrological modelling of the Chaohe Basin in China : Statistical model formulation and Bayesian inference. *Journal of Hydrology*, 340(3-4):167–182, 2007a. Cité pp. 16 et 17.
- YANG, J., P. REICHERT et K. C. ABBASPOUR : Bayesian uncertainty analysis in distributed hydrologic modeling : a case study in the Thur River basin (Switzerland). *Water Resources Research*, 43(10):W10401, 2007b. Cité p. 17.
- ZALACHORI, I., M. H. RAMOS, R. GARÇON, T. MATHEVET et J. GAILHARD : Statistical processing of forecasts for hydrological ensemble prediction : a comparative study of different bias correction strategies. *Adv. Sci. Res.*, 8:135–141, 2012. Cité pp. 11 et 140.
- ZAPPA, M., F. FUNDEL et S. JAUN : A ‘Peak-Box’ approach for supporting interpretation and verification of operational ensemble peak-flow forecasts. *Hydrological Processes*, 27(1):117–131, 2013. Cité p. 58.
- ZHANG, Z., T. WAGENER, P. REED et R. BHUSHAN : Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. *Water Resources Research*, 44:W00B04, 2008. Cité pp. 15 et 106.

# A

## COMPLÉMENTS AU CHAPITRE 2

## A.1 Liste des 202 bassins versants

Nom de la station	Code	S [km <sup>2</sup> ]	Z [m]	P [mm]	E[mm]	Q[mm]
L'Ill à Didenheim	A1080330	668	390	977	670	316
La Lièpvrette à Lièpvre	A2332110	107	602	1407	623	546
Le Giessen à Sélestat	A2352020	277	501	1206	634	409
L'Ehn à Niedernai	A2612010	57	443	1047	657	302
La Moder à Schweighouse-sur-Moder	A3301010	622	267	859	672	287
La Zorn à Saverne	A3422010	184	424	1102	649	385
La Moselle à Toul	A5730610	3346	442	1180	636	622
La Meurthe à Azerailles	A6221010	964	530	1488	626	637
Le Vacon à Barbas	A6443010	36	303	1082	648	412
Le ruisseau des Amis à Marainviller	A6543010	38	264	821	666	281
La Mortagne à Sainte-Hélène	A6621210	98	484	1494	640	537
La Meurthe à Laneuveville-devant-Nancy	A6921010	2787	382	1142	653	424
La Sarre Rouge à Vasperviller	A9013050	90	503	1210	630	513
La Sarre à Keskastel	A9091050	881	318	973	661	350
La Rosselle à Forbach	A9532010	168	273	862	650	349
Le Mouzon à Circourt-sur-Mouzon	B1092010	401	378	1004	640	363
La Meuse à Stenay	B3150020	3915	319	980	645	417
Le Ton à Écouvies	B4224310	310	298	968	633	508
La Chiers à Carignan	B4631010	1978	282	937	640	424
L'Helpe Mineure à Maroilles	D0137020	267	193	903	634	478
La Solre à Ferrière-la-Grande	D0206010	115	190	907	633	410
La Laquette à Witternesse	E3518510	79	115	966	627	224
L'Aa à Wizernes	E4035710	392	135	1020	618	464
L'Authie à Dompierre-sur-Authie	E5505720	792	115	895	634	333
L'Yères à Touffreville-sur-Eu	G1003010	255	152	963	632	348
Le Durdent à Vittefleury	G6003010	354	125	998	652	350
La Seine à Nod-sur-Seine	H0100010	373	404	955	662	398
La Seine à Courtenot	H0400020	2392	318	908	671	216
L'Aube à Bar-sur-Aube	H1201010	1298	323	952	670	395
La Laine à Soullaines-Dhuys	H1333010	22	200	890	680	1500
La Barbuise à Pouan-les-Vallées	H1513210	171	145	743	680	174
La Superbe à Saint-Saturnin	H1603010	366	124	739	673	145
L'Yonne à Courlon-sur-Yonne	H2721010	10687	263	870	688	275
Le Loing à Châlette-sur-Loing	H3201010	2302	170	753	695	174
Le Loing à Épisy	H3621010	3917	148	733	697	165
Le ru d'Ancoeur à Blandy	H3923010	188	117	705	690	93
L'Oeuf à Bondaroy	H4002010	241	125	665	700	72
La Juine à Méréville	H4033010	210	129	640	693	146
La Suize à Villiers-sur-Suize	H5033310	83	405	988	652	377
Le Bruxenelle à Brusson	H5173110	131	155	867	688	252
L'Orgeval à Boissy-le-Châtel	H5723011	105	148	761	677	191
Le Grand Morin à Pommeuse	H5732010	769	164	748	676	224
L'Aisne à Givry	H6221010	2888	195	944	656	360
La Vesle à Châlons-sur-Vesle	H6412020	828	131	711	678	156
L'Oise à Sempigny	H7401010	4320	137	844	649	264
Le Thérain à Beauvais	H7742010	755	157	842	636	240
Le Sausseron à Nesles-la-Vallée	H7853010	102	107	716	668	171
L'Andelle à Vascoeuil	H8212010	377	159	973	642	333
L'Eure à Saint-Luperce	H9021010	315	224	730	664	162
L'Eure à Louviers	H9501010	5891	162	672	668	138
L'Orbiquet à Beuvillers	I1103010	334	179	891	649	268

Nom de la station	Code	S [km <sup>2</sup> ]	Z [m]	P [mm]	E[mm]	Q[mm]
La Dives à Beaumais	I2021010	283	142	747	659	181
La Drôme à Sully	I5352010	241	115	907	663	312
Le Thar à Jullouville	I7913610	73	93	1034	668	444
Le Trieux à Saint-Péver	J1711710	184	218	1106	647	469
Le Douroun au Ponthou	J2404010	26	193	1234	656	564
L'Aber-Benoit à Plabennec	J3213020	27	80	1132	669	622
L'Élorn à Commana	J3403020	10	261	1377	637	998
Le ruisseau de Mougau Bihan à Commana	J3404110	9	232	1312	636	980
L'Aulne à Scignac	J3601810	117	199	1260	642	632
L'Hyères à Trébrivan	J3713010	258	197	1105	655	551
La Douffine à Saint-Ségal	J3834010	140	163	1295	659	777
L'Odet à Ergué-Gabéric	J4211910	203	145	1263	666	772
L'Éllé à Arzano	J4742010	576	165	1187	661	547
L'Evel à Guénin	J5613010	316	106	942	682	359
Le Coët-Organ à Quistinic	J5704810	46	132	1157	680	602
Le Meu à Montfort-sur-Meu	J7353010	477	96	787	682	217
La Seiche à Bruz	J7483010	809	70	758	703	199
Le Semnon à Bain-de-Bretagne	J7633010	406	81	803	700	253
L'Oust à Saint-Gravé	J8502310	2454	121	910	676	323
L'Arz à Molac	J8813010	161	91	1017	693	448
La Loire à Bas-en-Basset	K0550010	3260	969	951	603	345
Le Furan à Andrézieux-Bouthéon	K0614010	175	662	880	680	422
Le Lignon de Chalmazel à Poncins	K0773220	662	743	936	650	382
L'Aix à Saint-Germain-Laval	K0813020	197	750	1011	632	455
Le Rhins à Saint-Cyr-de-Favières	K0983010	435	553	953	686	385
L'Arconce à Montceaux-l'Étoile	K1173210	593	358	920	696	293
La Selle à la Celle-en-Morvan	K1284810	135	558	1228	650	702
L'Arroux à Étang-sur-Arroux	K1321810	1792	431	999	673	379
L'Arroux à Rigny-sur-Arroux	K1341810	2271	407	984	673	373
L'Aron à Verneuil	K1773010	1468	299	977	700	361
Le Lidenne à Couteuges	K2365510	48	697	758	651	142
La Morge à Maringues	K2783010	668	512	835	697	187
Le Couzon à Courpière	K2944010	72	721	1044	632	561
Le Sioulet à Miremont	K3273010	470	740	954	640	409
L'Aumance à Hérisson	K5383010	929	342	832	700	218
L'Auron à Bourges	K5653010	575	192	811	715	193
La Sauldre à Salbris	K6402510	1240	219	832	692	247
L'Indre à Saint-Cyran-du-Jambot	K7312610	1707	201	805	720	247
La Tourmente à Villeloin-Coulangé	K7414010	109	150	736	722	178
La Briance à Condat-sur-Vienne	L0563010	605	386	1136	719	431
Le Clain à Vivonne	L2321610	1824	146	828	733	256
La Petite Creuse à Fresselines	L4411710	853	393	916	694	314
L'Ardour à Folles	L5034010	129	437	1098	702	440
La Brame à Oradour-Saint-Genest	L5323010	232	267	951	729	304
La Claise au Grand-Pressigny	L6202030	886	117	762	737	151
L'Hoëne à la Mesnière	M0014110	75	220	830	659	272
L'Orne Saosnoise à Montbizot	M0243010	502	103	756	706	169
La Tortue à Saint-Michel-de-Chavaignes	M0416010	45	140	730	715	209
La Gée à Fercé-sur-Sarthe	M0535010	113	86	725	722	180
La Vègre à Asnières-sur-Vègre	M0583020	401	122	758	706	247
La Sarthe à Saint-Denis-d'Anjou	M0680610	7523	135	768	694	211
L'Ozanne à Trizay-lès-Bonneval	M1034020	267	187	713	673	174
Le Loir à Durtal	M1531610	7918	130	686	699	133
La Mayenne à Chambellay	M3630910	4166	154	863	676	315
L'Oudon à Châtellais	M3771810	732	81	772	704	185



Nom de la station	Code	S [km <sup>2</sup> ]	Z [m]	P [mm]	E[mm]	Q[mm]
L'Hyrome à Saint-Lambert-du-Lattay	M5214020	153	99	737	730	194
Le Beuvron à Andrezé	M6014010	38	106	771	728	269
L'Ouin à Mauléon	M7044010	60	174	837	721	349
La Maine à Remouillé	M7453010	595	80	894	731	309
L'Arize au Mas-d'Azil	O0744030	220	600	1101	747	556
La Louge à Muret	O0984010	481	307	779	787	379
Le Girou à Cépet	O2344010	524	215	768	814	135
Le Tarnon à Florac	O3064010	132	982	964	636	828
La Mimente à Florac	O3084320	126	937	1457	645	854
Le Tarn à Millau	O3401010	2143	896	1135	652	675
La Muze à Montjoux	O3454310	112	780	991	680	397
Le Rancé à Saint-Sernin-sur-Rance	O3754010	289	625	1104	727	615
Le Gijou à Vabre	O4194310	207	704	1187	707	669
Le Dadou à Paulinet	O4704030	71	678	1179	719	623
Le Serre à Coussergues	O5055010	108	765	1103	689	297
L'Aveyron à Laguépie	O5292510	1604	597	993	719	337
La Vère à Bruniquel	O5754020	310	267	833	795	210
Le Lemboulas à Lafrançaise	O5964020	403	186	802	800	161
Le Dourdou à Conques	O7874010	545	507	1070	734	411
Le Célé à Orniac	O8133520	1246	417	1112	742	472
La Lémance à Cuzorn	O8394310	220	228	901	774	181
Le Céou à Saint-Cybranet	P2484010	573	272	930	773	182
La Loyre à Voutezac	P3234010	104	371	1175	724	533
La Loyre à Saint-Viance	P3274010	254	292	1118	740	424
La Corrèze à Tulle	P3502510	354	601	1381	664	907
La Montane à Eyrein	P3614010	42	627	1431	665	927
L'Engranne à Baigneaux	P5715010	30	72	826	805	177
L'Isle à Corgnac-sur-l'Isle	P6081510	448	331	1127	731	436
L'Isle à Abzac	P7261510	3757	214	981	764	317
La Lizonne à Saint-Séverin	P8284010	624	143	913	771	269
L'Oussouet à Trébons	Q0115710	32	938	1396	702	972
L'Échez à Bordères-sur-l'Échez	Q0224020	154	440	1222	762	593
La Laysa à Villefranque	Q0435010	38	262	961	795	527
L'Adour à Saint-Vincent-de-Paul	Q3120010	7707	228	1013	791	358
Le Luy à Saint-Pandelon	Q3464010	1144	133	1133	817	457
Le Saison à Mauléon-Licharre	Q7322510	498	729	1806	713	1454
La Bidouze à Aïcirits-Camou-Suhast	Q8032510	246	367	1558	788	654
Le Mendialçu à Hasparren	Q8345910	17	132	1416	825	909
Le Grand Arriou à Moustey	S2224610	113	69	1046	791	274
Le Magescq à Magescq	S4214010	77	64	1168	806	419
La Saône à Monthureux-sur-Saône	U0020010	231	352	1024	652	406
Le Coney à Xertigny	U0104010	64	418	1191	642	656
Le Breuchin à la Proiselière-et-Langle	U0415010	123	561	1640	628	1138
La Saône à Ray-sur-Saône	U0610010	3761	334	1113	662	517
La Vingeanne à Oisilly	U0924020	609	314	942	674	326
L'Ognon à Chassey-lès-Montbozon	U1044010	852	410	1380	666	634
La Norges à Genlis	U1235020	271	273	834	698	321
Le Saint-Nicolas à Rougemont-le-Château	U2305210	9	713	1316	655	1084
La Savoureuse à Vieux-Charmont	U2345040	235	490	1547	659	825
Le Rhome à Lachapelle-sous-Chaux	U2345410	19	571	2119	634	1232
La Rosemontoise à Chaux	U2345820	27	641	2179	624	1171
Le Cusancin à Cusance	U2425250	154	577	1329	657	959
La Loue à Champagne-sur-Loue	U2634010	1264	620	1473	635	1324
La Furieuse à Salins-les-Bains	U2635010	41	589	1548	641	1213
La Cuisance à Mesnay	U2655030	61	587	1582	640	1734

Nom de la station	Code	S [km <sup>2</sup> ]	Z [m]	P [mm]	E[mm]	Q[mm]
Le Doubs à Neublans-Abergement	U2722010	7346	580	1368	641	777
La Seille à Saint-Usuge	U3424010	938	306	1238	707	481
Le Sevron à Bény	U3445020	60	287	1147	721	487
Le Vieux Jonc à Buellas	U4216010	121	272	1020	728	240
Le Morgon à Villefranche-sur-Saône	U4525210	68	337	807	765	231
Le Foron à Sciez	V0355010	51	656	1296	655	566
L'Albane à Chambéry	V1318210	53	557	1480	691	520
Le Furans à Arbignieu	V1464310	159	525	1427	671	702
La Bourbre à Tignieu-Jamezieu	V1774010	696	375	1037	739	341
La Sereine à Montluel	V3005610	78	283	1022	737	213
La Véga à Pont-Évêque	V3225420	86	300	1025	759	312
La Valencize à Chavanay	V3315010	37	646	908	698	295
Le Rival à Beaufort	V3424310	467	472	1089	714	60
La Galaure à Saint-Uze	V3614010	228	415	995	734	291
La Véore à Beaumont-lès-Valence	V4034020	195	379	1014	748	199
La Glueyre à Gluiras	V4145210	72	852	1345	662	866
Le Jabron à Souspierre	V4455010	79	525	915	725	294
L'Ardèche à Meyras	V5004030	99	899	1775	635	1132
Le Toulourenc à Malaucène	V6035010	157	847	987	672	255
Le Gardon de Saint-Germain à Saint-Germain-de-Calberte	V7105210	31	674	1579	722	760
Le Gardon de Mialet à Généralgues	V7124010	244	563	1449	747	692
Le Gardon de Saint-Jean à Corbès	V7135010	262	537	1380	769	756
L'Agly à Saint-Paul-de-Fenouillet	Y0624020	218	730	946	736	390
Le Fresquel à Villepinte	Y1314010	211	222	788	819	154
L'Orbiel à Bouilhonnac	Y1415020	242	480	1024	754	346
L'Argent Double à la Redorte	Y1435410	108	425	1024	780	302
L'Arre au Vigan	Y2015010	155	664	1389	728	929
La Vis à Saint-Laurent-le-Minier	Y2035010	306	640	1379	731	968
L'Hérault à Laroque	Y2102010	916	626	1363	739	613
La Lergue à Lodève	Y2214010	181	565	1266	741	714
La Mare au Pradal	Y2525010	116	584	1428	734	576
Le Lez à Montferrier-sur-Lez	Y3204010	116	146	866	856	552
Le Salaison à Mauguio	Y3315080	53	77	815	881	227
Le Vistre à Bernis	Y3514020	280	80	729	887	232
L'Arc à Meyreuil	Y4022010	297	342	689	814	103
Le Réal Collobrier à Pierrefeu-du-Var	Y4615610	71	338	935	843	228
Le Gapeau à Hyères	Y4624010	536	316	871	852	190
L'Argens à Roquebrune-sur-Argens	Y5312010	2514	350	804	838	186
La Siagne à Callian	Y5514040	166	956	975	692	1002
Le Loup à Tournettes-sur-Loup	Y5615010	206	1050	1058	707	485
L'Estéron au Broc	Y6434010	443	923	1017	709	432

TABLE A.1: Liste des 202 bassins versants. S désigne la surface, Z l'altitude moyenne et P, E et Q désignent respectivement les moyennes annuelles des précipitations, de l'évapo-transpiration potentielle et des débits.

## A.2 Fonctionnement du modèle GR4J

Cette présentation du fonctionnement du modèle GR4J est issue de Perrin *et al.* (2003) (voir également le site de l'équipe : <http://webgr.irstea.fr>).

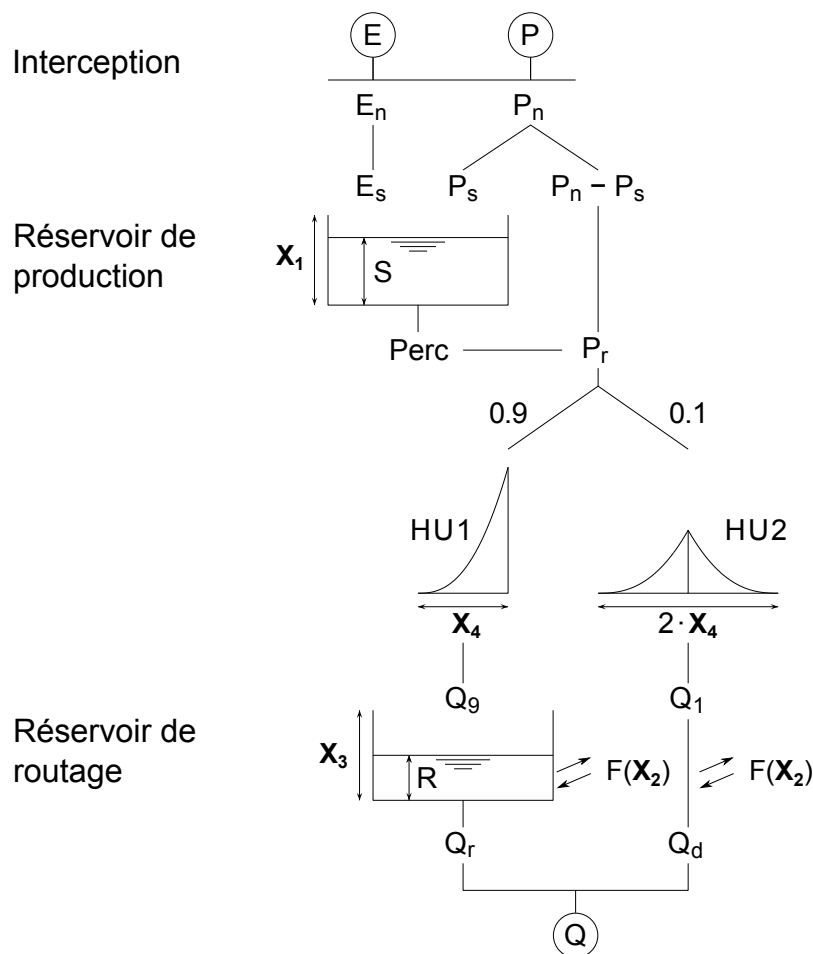


FIGURE A.1 – Schéma de la structure du modèle GR4J.

On désigne par  $P$ (mm) la hauteur de pluie et par  $E$ (mm) l'évapotranspiration potentielle (ETP).  $P$  est une estimation de la pluie de bassin et  $E$  peut être issue d'une courbe d'ETP moyenne interannuelle.

Les équations suivantes correspondent aux équations intégrées sur un pas de temps.

### A.2.1 Fonction de production

La première opération est la neutralisation de P par E pour déterminer une pluie nette  $P_n$  et une évapotranspiration nette  $E_n$ , calculées par :

$$\text{Si } P > E \text{ alors } P_n = P - E \text{ et } E_n = 0$$

$$\text{Si } P < E \text{ alors } P_n = 0 \text{ et } E_n = E - P$$

Dans le cas où  $P_n$  est différente de zéro, une partie  $P_s$  de  $P_n$  alimente le réservoir de production et est calculée par :

$$P_s = X1 \cdot \frac{\left(1 - \left(\frac{S}{X1}\right)^2\right) \cdot \tanh\left(\frac{P_n}{X1}\right)}{1 + \frac{S}{X1} \cdot \tanh\left(\frac{P_n}{X1}\right)} \quad (\text{A.1})$$

où X1 et S sont respectivement la capacité maximum et le niveau du réservoir de production.

Dans le cas contraire, lorsque  $E_n$  est différent de zéro, une quantité d'évaporation  $E_s$  est retirée du réservoir de production. Elle est donnée par :

$$E_s = S \cdot \frac{\left(2 - \left(\frac{S}{X1}\right)^2\right) \cdot \tanh\left(\frac{E_n}{X1}\right)}{1 + \left(1 - \frac{S}{X1}\right) \cdot \tanh\left(\frac{E_n}{X1}\right)} \quad (\text{A.2})$$

Le contenu du réservoir est mis à jour par :

$$S = S - E_s + P_s \quad (\text{A.3})$$

Une percolation Perc issue du réservoir de production est alors calculée :

$$\text{Perc} = S \cdot \left\{ 1 - \left[ 1 + \left( \frac{4}{9} \frac{S}{X1} \right)^4 \right]^{-\frac{1}{4}} \right\} \quad (\text{A.4})$$

Le contenu du réservoir est de nouveau mis à jour par :

$$S = S - \text{Perc} \quad (\text{A.5})$$

La quantité d'eau  $P_r$  qui atteint finalement la partie routage du modèle est :

$$P_r = \text{Perc} + (P_n - P_s) \quad (\text{A.6})$$

## A.2.2 Fonction de routage

$P_r$  est divisée en deux composantes d'écoulement, 90% étant routés par un hydrogramme unitaire HU1 et un réservoir de routage et 10% par un unique hydrogramme unitaire HU2.

HU1 et HU2 dépendent du même paramètre  $X_4$ , temps de base de HU1 exprimé en jours. Les ordonnées des hydrogrammes sont calculées à partir des courbes en S (cumul de la proportion de la pluie unitaire traitée par l'hydrogramme en fonction du temps), notées respectivement SH1 et SH2 :

$$\text{SH1}(t) = \begin{cases} 0 & : t \leq 0 \\ \left(\frac{t}{X_4}\right)^{\frac{5}{2}} & : 0 < t < X_4 \\ 1 & : t \geq X_4 \end{cases} \quad (\text{A.7})$$

$$\text{SH2}(t) = \begin{cases} 0 & : t \leq 0 \\ \frac{1}{2} \left(\frac{t}{X_4}\right)^{\frac{5}{2}} & : 0 < t \leq X_4 \\ 1 - \frac{1}{2} \left(2 - \frac{t}{X_4}\right)^{\frac{5}{2}} & : X_4 < t < 2X_4 \\ 1 & : t \geq 2X_4 \end{cases} \quad (\text{A.8})$$

Les ordonnées de HU1 et HU2 sont alors calculées par :

$$\text{UH1}(j) = \text{SH1}(j) - \text{SH1}(j-1)$$

$$\text{UH2}(j) = \text{SH2}(j) - \text{SH2}(j-1)$$

où  $j$  est un entier.

Un échange souterrain en eau (perte ou gain) est calculé par :

$$F = X_2 \cdot \left(\frac{R}{X_3}\right)^{\frac{7}{2}} \quad (\text{A.9})$$

où  $R$  est le niveau dans le réservoir,  $X_3$  la capacité à un jour du réservoir et  $X_2$  le coefficient

d'échange en eau qui peut être positif dans le cas d'apports, négatif dans le cas de pertes vers des nappes profondes ou nul.

Le niveau dans le réservoir de routage est mis à jour en ajoutant la sortie  $Q_9$  de l'hydrogramme HU1 et F :

$$R = \max(0; R + Q_9 + F) \quad (\text{A.10})$$

Il se vidange ensuite en une sortie  $Q_r$  donnée par :

$$Q_r = R \cdot \left\{ 1 - \left[ 1 + \left( \frac{R}{X3} \right)^4 \right]^{-\frac{1}{4}} \right\} \quad (\text{A.11})$$

Le niveau dans le réservoir devient :

$$R = R - Q_r \quad (\text{A.12})$$

La sortie  $Q_1$  de l'hydrogramme HU2 est soumise au même échange pour donner la composante d'écoulement  $Q_d$  :

$$Q_d = \max(0; Q_1 + F) \quad (\text{A.13})$$

Le débit total  $Q$  est alors donnée par :

$$Q = Q_r + Q_d \quad (\text{A.14})$$

### A.3 Fonctionnement du modèle GRP

Cette présentation du fonctionnement au pas de temps horaire du modèle GRP s'inspire de la description faite par Berthet (2010) (voir également le site de l'équipe : <http://webgr.irstea.fr>).

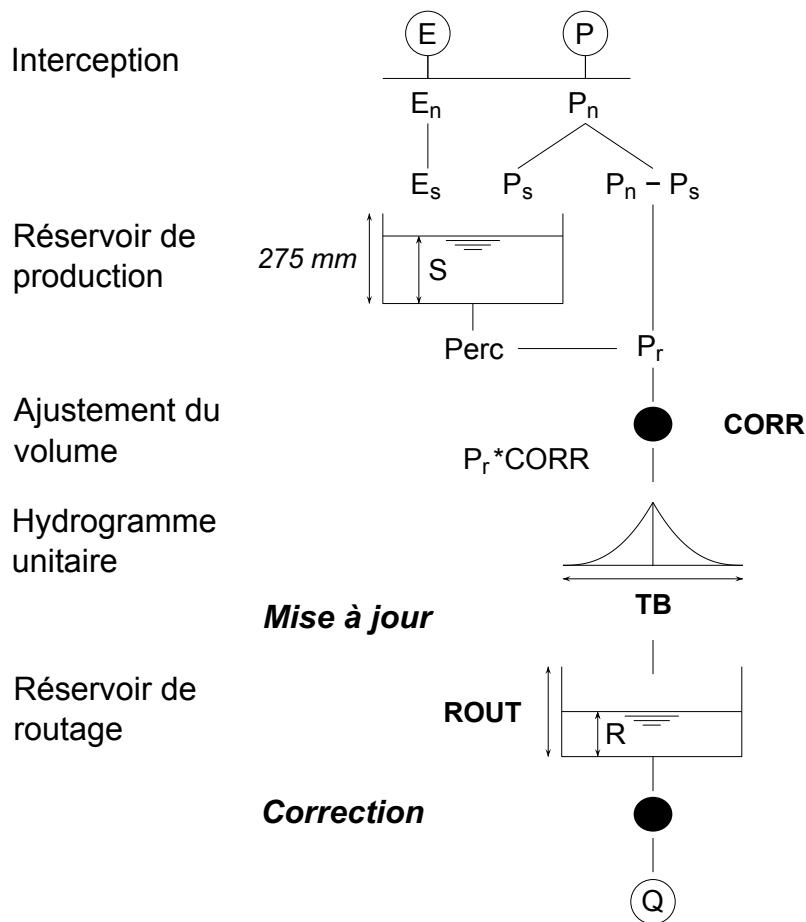


FIGURE A.2 – Schéma de la structure du modèle GRP.

On désigne par  $P(\text{mm})$  la hauteur de pluie et par  $E(\text{mm})$  l'évapotranspiration potentielle (ETP).  $P$  est une estimation de la pluie de bassin et  $E$  peut être issue d'une courbe d'ETP moyenne interannuelle.

Les équations suivantes correspondent aux équations intégrées sur un pas de temps.

### A.3.1 Fonction de production

La première opération est la neutralisation de P par E pour déterminer une pluie nette  $P_n$  et une évapotranspiration nette  $E_n$ , calculées par :

$$\text{Si } P > E \text{ alors } P_n = P - E \text{ et } E_n = 0$$

$$\text{Si } P < E \text{ alors } P_n = 0 \text{ et } E_n = E - P$$

Dans le cas où  $P_n$  est différente de zéro, une partie  $P_s$  de  $P_n$  alimente le réservoir de production et est calculée par :

$$P_s = A \cdot \frac{\left(1 - \left(\frac{S}{A}\right)^2\right) \cdot \tanh\left(\frac{P_n}{A}\right)}{1 + \frac{S}{A} \cdot \tanh\left(\frac{P_n}{A}\right)} \quad (\text{A.15})$$

où A et S sont respectivement la capacité maximum (fixée à 275 mm) et le niveau du réservoir de production.

Dans le cas contraire, lorsque  $E_n$  est différent de zéro, une quantité d'évaporation  $E_s$  est retirée du réservoir de production. Elle est donnée par :

$$E_s = S \cdot \frac{\left(2 - \left(\frac{S}{A}\right)^2\right) \cdot \tanh\left(\frac{E_n}{A}\right)}{1 + \left(1 - \frac{S}{A}\right) \cdot \tanh\left(\frac{E_n}{A}\right)} \quad (\text{A.16})$$

Le contenu du réservoir est mis à jour par :

$$S = S - E_s + P_s \quad (\text{A.17})$$

Une percolation Perc issue du réservoir de production est alors calculée :

$$\text{Perc} = S \cdot \left\{ 1 - \left[ 1 + \left(\frac{S}{K}\right)^4 \right]^{-\frac{1}{4}} \right\} \quad (\text{A.18})$$

où  $K = A/5.25$

Le contenu du réservoir est de nouveau mis à jour par :

$$S = S - \text{Perc} \quad (\text{A.19})$$



La quantité d'eau qui atteint finalement la partie routage du modèle est :

$$P_r^{\text{CORR}} = \text{CORR} \cdot (\text{Perc} + (P_n - P_s)) \quad (\text{A.20})$$

où CORR est le coefficient d'ajustement du volume.

### A.3.2 Fonction de routage

L'hydrogramme unitaire HU dépend du paramètre TB, temps de base exprimé en heures. Les ordonnées des hydrogrammes sont calculées à partir d'une courbe en S (cumul de la proportion de la pluie unitaire traitée par l'hydrogramme en fonction du temps) :

$$\text{SH}(t) = \begin{cases} 0 & : t \leq 0 \\ \frac{t^\alpha}{t^\alpha + (\text{TB} - t)^\alpha} & : 0 < t < \text{TB} \\ 1 & : t \geq \text{TB} \end{cases} \quad (\text{A.21})$$

où  $\alpha = 1.25$  est un paramètre fixe.

Le niveau dans le réservoir de routage est mis à jour en ajoutant la sortie de l'hydrogramme HU et se vidange ensuite en une sortie  $Q_{prev}$  donnée par :

$$Q_{prev} = \frac{R^2}{R + \text{ROUT}} \quad (\text{A.22})$$

où ROUT désigne la capacité du réservoir de routage.

### A.3.3 Procédures de mises-à-jour

Deux procédures de mises-à-jour sont utilisées dans le fonctionnement du modèle GRP pour la prévision en temps réel.

La mise-à-jour du réservoir de routage est directe, c'est-à-dire qu'elle vise à ajuster le niveau du réservoir de routage en fonction du dernier débit observé  $Q_{obs}$ . Le réservoir de routage étant quadratique, on obtient :

$$R_{maj} = \frac{2 \cdot \text{ROUT}}{1 + \sqrt{1 + 4 \cdot \frac{\text{ROUT}}{Q_{obs}}}} \quad (\text{A.23})$$

Cet ajustement est réalisé chaque fois qu'un débit observé est disponible à l'instant de prévision.

La seconde procédure d'ajustement est donnée par l'équation suivante :

$$Q_{maj} = Q_{prev} \cdot \epsilon^\gamma \quad (\text{A.24})$$

où  $\gamma = 0.45$  est un paramètre fixe, et le rapport  $\epsilon = \left(\frac{Q_{prev}}{Q_{obs}}\right)_{t-1}$  désigne la dernière erreur multiplicative disponible à l'instant de prévision lorsque l'on considère une échéance de prévision d'une heure.



# B

## COMPLÉMENTS AU CHAPITRE 5

## B.1 Inférence bayésienne

Selon le paradigme bayésien, l'incertitude sur des paramètres inconnus, traités comme des variables aléatoires, s'exprime par une densité de probabilité. La densité de probabilité *a posteriori* est obtenue lors du processus d'inférence à partir d'une loi *a priori* et de l'expression d'une fonction de vraisemblance  $f(\mathbf{x})$  qui correspond à la densité de la loi de probabilité des observations  $\mathbf{x}$  utilisées pour l'inférence.

L'inférence bayésienne permet d'estimer les paramètres inconnus  $\theta$  à l'aide de la formule de Bayes :

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta} \quad (\text{B.1})$$

où  $\pi(\theta)$  et  $\pi(\theta|\mathbf{x})$  désignent respectivement la densité de la loi de probabilité *a priori* et la densité de la loi de probabilité *a posteriori* des paramètres  $\theta$ .

Dans un contexte de modélisation hydrologique, il est courant de considérer la formulation suivante :

$$Q_{o,t} = Q_{s,t}(\theta) + \epsilon_t(\theta), t = 1, \dots, n \quad (\text{B.2})$$

où  $n$  est le nombre de pas de temps,  $Q_{o,t}$  et  $Q_{s,t}(\theta)$  sont respectivement le débit observé et le débit simulé avec les paramètres  $\theta$ , et  $\epsilon_t(\theta)$  désigne l'erreur additive résiduelle, au pas de temps  $t$ .

Dans un tel cadre de modélisation déterministe, ce sont les hypothèses faites sur la distribution des erreurs résiduelles qui déterminent l'expression de la fonction de vraisemblance.

Lorsque les erreurs résiduelles sont supposées être des réalisations indépendantes d'une loi normale, qu'aucune information sur la distribution *a priori* des paramètres n'est utilisée, et que la contribution de la variance des résidus est marginalisée, la densité de probabilité *a posteriori* des paramètres libres du modèle est liée aux critères de performance quadratiques. Le développement ci-dessous suit la présentation de [Kavetski et al. \(2006a\)](#).

Suivant les hypothèses énoncées ci-dessus, les erreurs  $\epsilon_t$  suivent une loi normale de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (\text{B.3})$$

En utilisant la loi *a priori* suivante pour  $\sigma^2$ ,  $p(\sigma^2) \propto 1/\sigma^2$ , connue sous le terme de prior non

informatif de Jeffrey, et en considérant que les erreurs sont indépendantes, la densité jointe des paramètres  $p(\theta, \sigma^2 | D)$  s'exprime par :

$$p(\theta, \sigma^2 | D) \propto \prod_t f(\epsilon_t) \cdot \frac{1}{\sigma^2} \quad (\text{B.4})$$

On obtient alors :

$$\begin{aligned} p(\theta, \sigma^2 | D) &\propto \prod_t \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma^2} \\ &\propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_t \epsilon_t^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma^2} \\ &\propto \sigma^{-n-2} \exp\left(-\frac{\sum_t \epsilon_t^2}{2\sigma^2}\right) \\ &\propto \sigma^{-n-2} \exp\left(-\frac{SS}{2\sigma^2}\right) \end{aligned} \quad (\text{B.5})$$

où  $SS = \sum_t \epsilon_t^2$  est la somme des erreurs quadratiques.

La densité des paramètres  $p(\theta | D)$  s'obtient en marginalisant la contribution du paramètre  $\sigma^2$ . Une transformation de variable permet de résoudre le calcul de manière analytique. En posant  $\sigma^2 = SS/2s$ , on obtient  $d\sigma^2 = (-SS/2s^2) ds$ , d'où

$$\begin{aligned} p(\theta | D) &= \int p(\theta, \sigma^2 | D) d\sigma^2 \\ &\propto \int \sigma^{-n-2} \exp\left(-\frac{SS}{2\sigma^2}\right) d\sigma^2 \\ &\propto \int (SS/2s)^{-(n-2)/2} \exp(-s) (-SS/2s^2) ds \\ &\propto SS^{-n/2} \int (1/2s)^{-(n-2)/2} \exp(-s) (-2s^2) ds \\ &\propto SS^{-n/2} \end{aligned} \quad (\text{B.6})$$

La densité de probabilité jointe *a posteriori* des paramètres du modèle déterministe considéré est liée à la somme des erreurs quadratiques et au nombre de pas de temps utilisés pour l'inférence.

## B.2 Méthodes de post-traitement

Nous décrivons ici plus en détails les méthodes de post-traitement présentées succinctement au chapitre 5.

### B.2.1 Préalable

#### Notations

Nous utilisons les notations suivantes :  $n$  est le nombre de pas de temps,  $Q_{o,t}$  et  $Q_{s,t}$  sont respectivement le débit observé et le débit simulé par un modèle, au pas de temps  $t$ .

En utilisant ces notations, les erreurs relatives  $e_t^r$  sont définies par :

$$e_t^r = \frac{Q_{o,t}}{Q_{s,t}} \quad (\text{B.7})$$

et les erreurs additives par :

$$e_t = Q_{o,t} - Q_{s,t} \quad (\text{B.8})$$

Les méthodes de post-traitement utilisées ici permettent d'obtenir des quantiles prédictifs  $q_{\alpha,t}$  d'ordre  $\alpha$  ( $q_{\alpha} = F_Q^{-1}(\alpha)$ , où  $F$  est la fonction de répartition du débit) pour différentes valeurs de  $\alpha$ , de manière à fournir une approximation discrète des distributions prédictives.

#### Transformation en quantiles normaux

La transformation en quantiles normaux est utilisée par les méthodes MeGa, MCP et RQ qui s'inspirent toutes les trois des travaux de [Kelly et Krzysztofowicz \(1997\)](#).

Il s'agit d'une transformation non-paramétrique qui permet d'obtenir des densités marginales gaussiennes. Elle est définie par :

$$\text{NQT}(x) = \Phi^{-1}(F_X(x)) \quad (\text{B.9})$$

où  $F_X(x)$  est la fonction de répartition d'une variable  $X$  et  $\Phi^{-1}$  désigne la réciproque de la fonction de répartition de la loi normale centrée réduite  $\mathcal{N}(0, 1)$ .

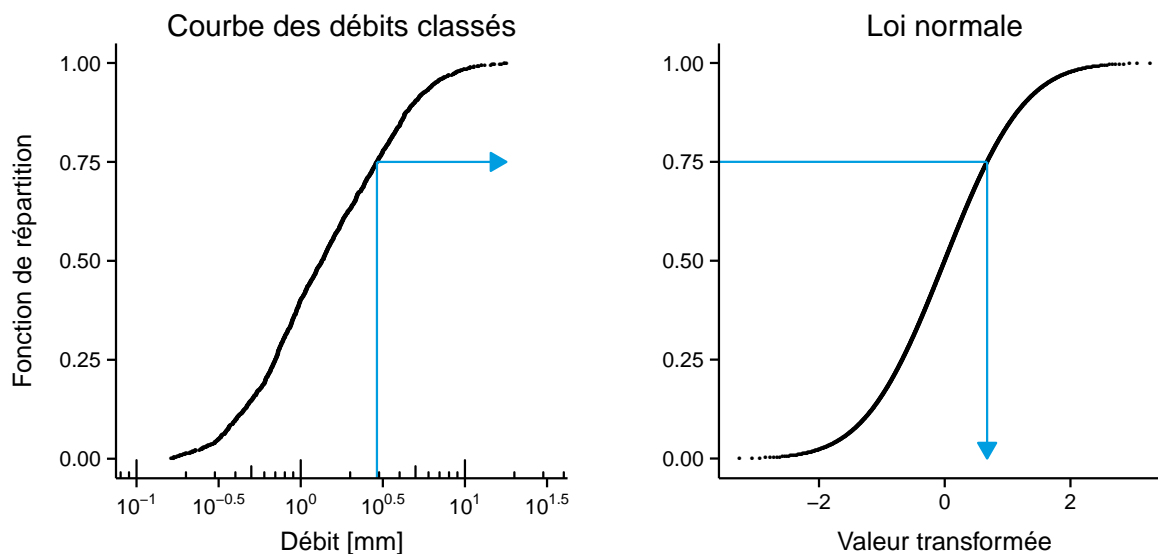


FIGURE B.1 – Illustration de la transformation NQT. Pour une valeur de la courbe des débits classés, la réciproque de la fonction de répartition de la loi normale centrée réduite  $\mathcal{N}(0, 1)$  est utilisée pour calculer la valeur transformée.

Le principe de la transformation est illustrée sur la Figure B.1.

### B.2.2 CemaM et CemaO

Les méthodes CemaM et CemaO s'appuient sur une analyse des erreurs relatives  $e_t^r$ .

Le fait de considérer des erreurs relatives plutôt que des erreurs additives permet d'obtenir des valeurs positives lorsque les marges d'erreurs viennent compléter la valeur simulée déterministe. Le seul inconvénient est que les situations pour lesquelles le débit simulé est nul doivent faire l'objet d'un traitement particulier. Nous n'avons cependant pas rencontré cette situation pour les simulations de débits sur nos bassins versants français, et nous n'avons donc considéré que les cas où les débits simulés sont strictement positifs.

La méthode empirique de quantification de l'incertitude prédictive proposée au sein de l'équipe d'hydrologie d'Antony par Claude Michel et testée dans le cadre d'un stage de DEA (Berthier, 2005) consiste à calculer des quantiles de la distribution empiriques des erreurs relatives pour différentes tranches de débits simulés. Plus précisément, la méthode peut être décrite par les étapes suivantes :

1. Calcul des erreurs relatives à partir d'une chronique de débits observés et simulés.
2. Création de N tranches de débits simulés. Les tranches de débits sont définies de manière à contenir un même nombre de valeurs.



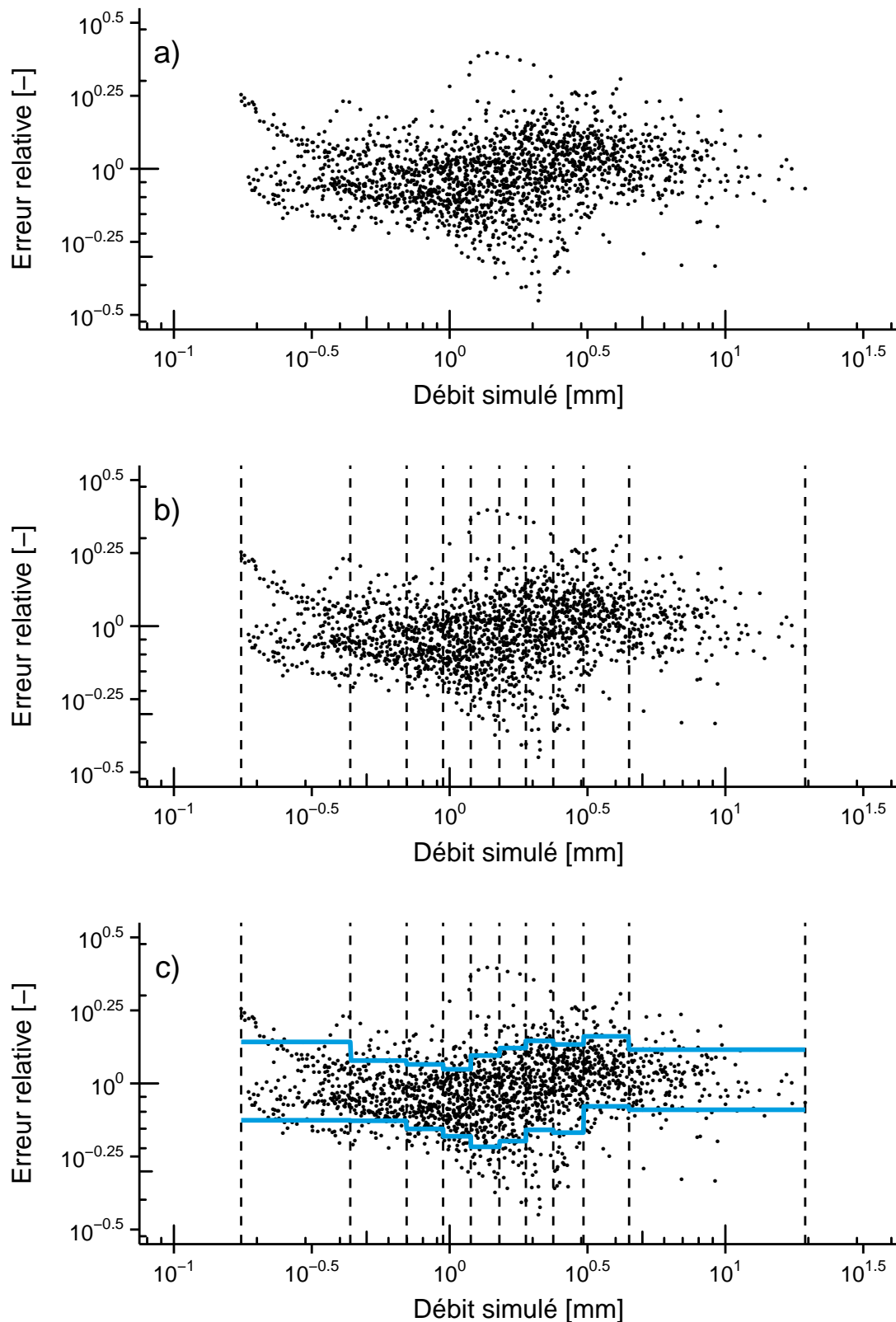


FIGURE B.2 – Illustration de la méthode CemaM. a) Nuage d’erreurs relatives en fonction du débit simulé. b) Création de  $N = 10$  tranches de débits simulés contenant le même nombre de points. c) Calcul des quantiles d’ordre 10% et 90% pour chaque tranche de débits.

3. Calcul de  $K$  quantiles d'ordre  $\alpha^k$  de la distribution empirique des erreurs relatives, pour chacune des tranches de débits.

Le fonctionnement de la méthode est illustré sur la Figure B.2.

Ces trois étapes permettent d'obtenir  $K \times N$  coefficients multiplicatifs notés  $M_k^n$ . Ces coefficients multiplicatifs sont utilisés pour obtenir les quantiles de la distribution prédictive :

$$q_{\alpha^k} = Q_s^i \times M_k^i \quad (\text{B.10})$$

où  $Q_s^i$  est un débit simulé appartenant à la  $i$ ème tranche de débits, et  $q_{\alpha^k}$  est le quantile prédictif d'ordre  $\alpha^k$ .

La méthode CemaO est une version simplifiée de la méthode CemaM. Elle ne considère qu'une seule tranche de débits,  $N = 1$ . Les coefficients multiplicatifs sont donc identiques pour toutes les valeurs de débits simulés.

### B.2.3 MeGa

La méthode MeGa proposée par [Montanari et Grossi \(2008\)](#) s'inspire des travaux de [Kelly et Krzysztofowicz \(1997\)](#) et s'appuie sur un modèle statistique appelé méta-gaussien pour décrire la relation entre les débits simulés  $Q_{s,t}$  et les erreurs additives  $e_t$ .

Une relation linéaire est utilisée dans le domaine transformé :

$$\text{NQT}(e_t) = a\text{NQT}(Q_{s,t}) + \epsilon_t \quad (\text{B.11})$$

où  $a$  est le coefficient de la relation linéaire et  $\epsilon_t$  suit une loi normale de moyenne nulle et de variance  $\sigma^2$ .

La méthode peut être décrite par les étapes suivantes :

1. Calcul des erreurs additives.
2. Séparation des erreurs positives et négatives. Les étapes suivantes s'appliquent à chaque population.
3. Calcul des valeurs transformées pour les débits simulés et les erreurs.
4. Caractérisation de la relation linéaire entre ces deux variables.
5. Calcul des quantiles dans le domaine transformé.
6. Calcul des quantiles dans le domaine réel.

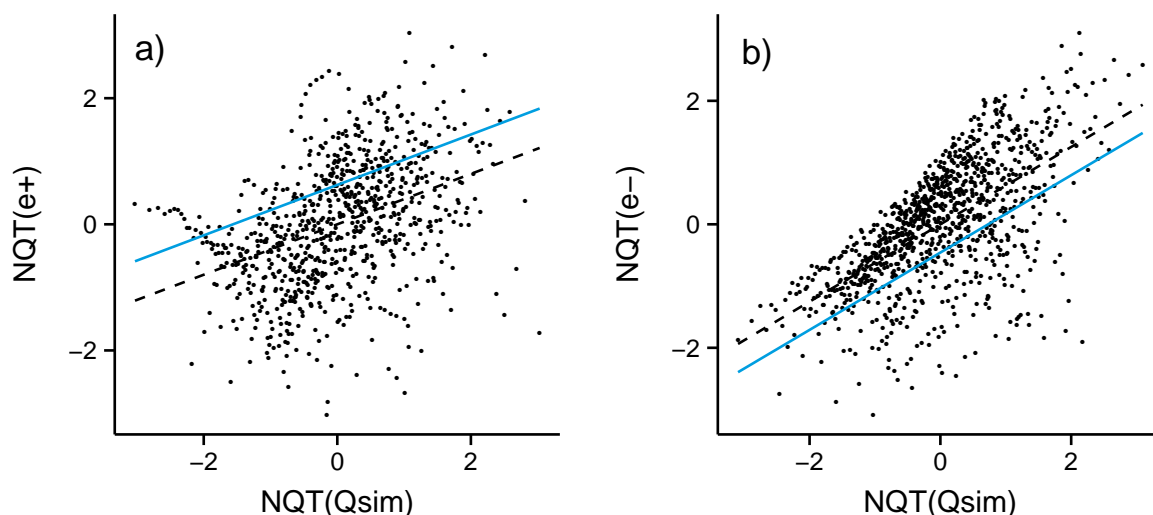


FIGURE B.3 – Illustration de la méthode MeGa. Les erreurs (a) positives et (b) négatives font l'objet d'un traitement séparé. Pour chaque nuage de points, la droite en pointillé désigne la droite de régression linéaire et la droite en bleue représente la courbe des quantiles d'ordre (a) 90% et (b) 10%.

La Figure B.3 présente une illustration du fonctionnement de la méthode MeGa.

## B.2.4 MCP

Le formalisme de la méthode MCP proposée par [Coccia et Todini \(2011\)](#) est proche de celui de la méthode MeGa. La différence est que la variable aléatoire prédite par le modèle statistique méta-gaussien utilisé n'est pas l'erreur mais le débit.

Le modèle repose sur l'hypothèse d'une relation linéaire entre les débits simulés  $Q_{s,t}$  et les débits observés  $Q_{o,t}$  dans le domaine normal :

$$\text{NQT}(Q_{o,t}) = a\text{NQT}(Q_{s,t}) + \epsilon_t \quad (\text{B.12})$$

où  $a$  est le coefficient de la relation linéaire et  $\epsilon_t$  suit une loi normale de moyenne nulle et de variance  $\sigma^2$ .

La méthode peut être décrite par les étapes suivantes :

1. Calcul des valeurs transformées pour les débits simulés et les débits observés.
2. Séparation de deux populations en fonction du débit simulé. Les étapes suivantes s'appliquent à chaque population.
3. Caractérisation de la relation linéaire entre les deux variables.

4. Calcul des quantiles dans le domaine transformé.
5. Calcul des quantiles dans le domaine réel.

Coccia et Todini (2011) propose de séparer le domaine en deux sous-domaines et d'utiliser une relation linéaire pour chaque sous-domaine. Nous avons choisi de séparer les données qui contenaient les 10% des débits simulés les plus forts du reste des données.

La Figure B.4 illustre le fonctionnement de la méthode.

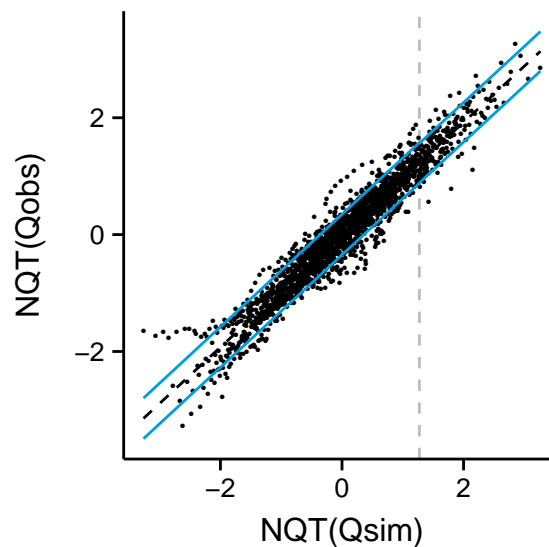


FIGURE B.4 – Illustration de la méthode MCP. La droite de régression linéaire est en pointillé et les droites en bleue représentent les quantiles d'ordre 10% et 90%. La droite verticale en pointillé marque la séparation entre deux populations. Sur cet exemple, la différence entre les caractéristiques de la relation linéaire pour chaque population est à peine visible.

### B.2.5 RQ

La régression quantile est une méthode de régression introduite à l'origine en économie. Elle permet de s'affranchir de l'hypothèse de normalité des résidus habituellement formulée pour les modèles de régression linéaire et d'étendre les techniques de régression aux quantiles.

On suppose que le quantile  $q_\alpha$  d'une variable aléatoire  $Y$  peut s'exprimer sous la forme d'une fonction  $f$  qui dépend d'une variable explicative  $X$  et des paramètres  $\beta_\alpha$ .

La régression quantile consiste à estimer, pour une valeur de  $\alpha$ , les paramètres  $\beta_\alpha$  en minimi-

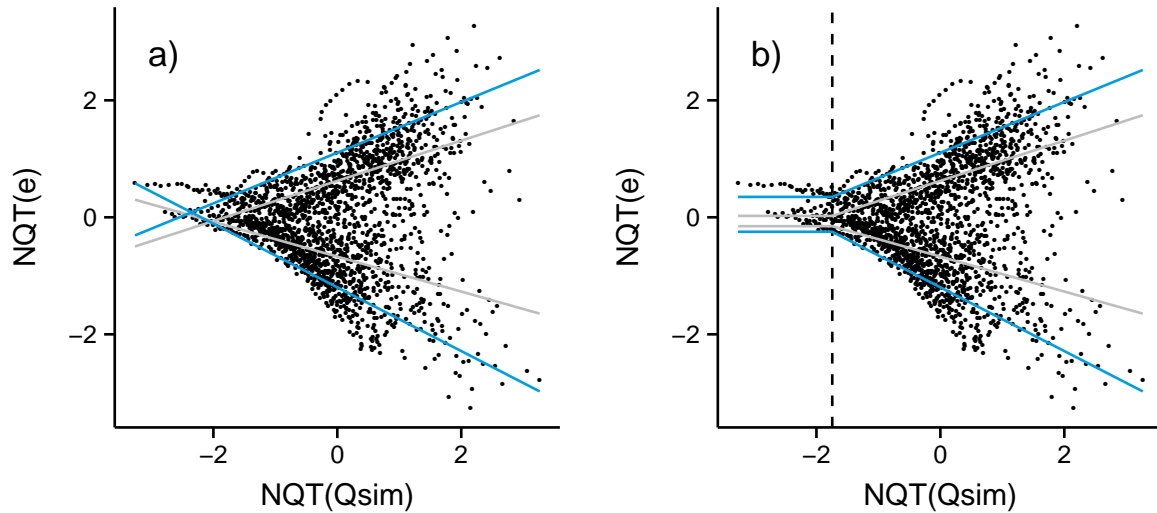


FIGURE B.5 – Illustration de la méthode RQ. a) Les droites bleues correspondent aux valeurs des quantiles conditionnels d'ordre 10% et 90%, et les droites grises aux quantiles d'ordre 30% et 70%. On remarque que les droites se croisent pour les bas débits. b) Correction du croisement des quantiles suivant l'approche présentée par [Weerts \*et al.\* \(2011\)](#).

sant la quantité :

$$\sum_i [\rho_\alpha(y_i - f(x_i, \beta_\alpha))] \quad (\text{B.13})$$

où  $(x_i, y_i)$  sont les réalisations des variables X et Y et  $\rho_\alpha(x)$  est une fonction définie par :

$$\rho_\alpha(x) = \begin{cases} x \cdot (\alpha - 1) & \text{si } x < 0 \\ x \cdot \alpha & \text{sinon} \end{cases} \quad (\text{B.14})$$

[Weerts \*et al.\* \(2011\)](#) proposent d'appliquer cette méthode de régression dans le domaine transformé entre les variables  $\text{NQT}(e_t)$  et  $\text{NQT}(Q_{s,t})$ . Le modèle est décrit par les relations linéaires suivantes :

$$\text{NQT}(e_t)_\alpha = a_\alpha \text{NQT}(Q_{s,t}) + b_\alpha \quad (\text{B.15})$$

avec les paramètres  $(a, b)_\alpha$  définis pour chaque valeur de  $\alpha$ .

Dans le domaine réel, le quantile  $q_{\alpha,t}$  d'ordre  $\alpha$  s'obtient par :

$$q_{\alpha,t} = Q_{s,t} + \text{NQT}^{-1}[\text{NQT}(e_t)_\alpha] \quad (\text{B.16})$$

où  $\text{NQT}^{-1}$  désigne la réciproque de la transformation NQT.

La méthode peut être décrite par les étapes suivantes :

1. Calcul des erreurs additives.
2. Calcul des valeurs transformées pour les débits simulés et les erreurs.
3. Calcul des paramètres  $(a, b)_\alpha$  de la régression quantile pour différentes valeurs de  $\alpha$ .
4. Calcul des quantiles d'ordre  $\alpha$  dans le domaine transformé.
5. Correction des quantiles lorsque ceux-ci se croisent.
6. Calcul des quantiles d'ordre  $\alpha$  dans le domaine réel.

La Figure B.5 illustre le fonctionnement de la méthode.

### Transformations de variables et normalité des résidus

Ces méthodes reposent sur l'hypothèse que les erreurs additives calculées après l'utilisation d'une transformation T de variable pour les débits simulés et observés suivent une loi normale de moyenne nulle et de variance  $\sigma^2$ .

Le formalisme mathématique est simplement :

$$T(Q_{o,t}) \sim \mathcal{N}(T(Q_{s,t}), \sigma^2) \quad (\text{B.17})$$

Nous avons utilisé des transformations de Box-Cox avec des paramètres  $\lambda = 0, 0.3, 0.5$  et  $1$ .

L'approche peut être décrite par les étapes suivantes :

1. Calcul des valeurs transformées pour les débits simulés et les débits observés.
2. Calcul des erreurs additives.
3. Calcul de la variance des erreurs.
4. Calcul des quantiles dans le domaine transformé.
5. Calcul des quantiles dans le domaine réel.

La Figure B.6 présente le fonctionnement de cette approche pour deux transformations.

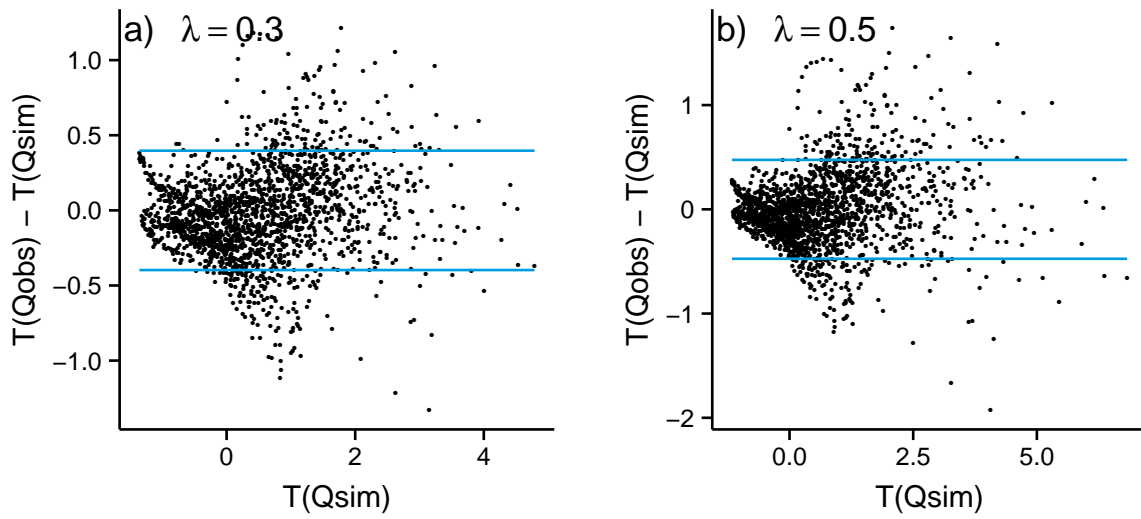


FIGURE B.6 – Illustration des méthodes de référence. L'impact de la transformation  $T$  sur le nuage de points est notable. Les droites bleues désignent les quantiles d'ordre 10% et 90%.