



HAL
open science

Diagnostic du colmatage des générateurs de vapeur à l'aide de modèles physiques et statistiques

Sylvain Girard

► **To cite this version:**

Sylvain Girard. Diagnostic du colmatage des générateurs de vapeur à l'aide de modèles physiques et statistiques. Autre. Ecole Nationale Supérieure des Mines de Paris, 2012. Français. NNT : 2012ENMP0051 . pastel-00798355v2

HAL Id: pastel-00798355

<https://pastel.hal.science/pastel-00798355v2>

Submitted on 21 Sep 2015

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.

École doctorale n° 432 : Sciences des métiers de l'ingénieur

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité « Géostatistique »

présentée et soutenue publiquement par

Sylvain GIRARD

le 17 décembre 2012

Diagnostic du colmatage des générateurs de vapeur
à l'aide de modèles physiques et statistiques

Steam generators clogging diagnosis
through physical and statistical modelling

Directeur de thèse : **Hans WACKERNAGEL**

Co-encadrement de la thèse : **Thomas ROMARY**

Pascal STABAT

Jean-Melaine FAVENNEC

Jury

Gérard BIAU, Professeur, École Normale supérieure et Université Pierre et Marie Curie
Clémentine PRIEUR, Professeur, Laboratoire Jean Kuntzmann, Université Joseph Fourier
Marc GRANDOTTO, Habilité à diriger des recherches, DEN/DM2S/STMF/LMEC, CEA
Hans WACKERNAGEL, Professeur, Centre de Géosciences, MINES ParisTech
Thomas ROMARY, Docteur, Centre de Géosciences, MINES ParisTech
Pascal STABAT, Docteur, Centre Énergétique et Procédés, MINES ParisTech
Jean-Melaine FAVENNEC, Ingénieur sénior, STEP, EDF R&D

Président
Rapporteur
Rapporteur
Examineur
Examineur
Examineur
Invité

**T
H
È
S
E**

MINES ParisTech
Centre de Géosciences

Centre de Géosciences, 35 rue Saint Honoré 77305 Fontainebleau, France

Diagnostic du colmatage
des générateurs de vapeur
à l'aide de modèles physiques et statistiques

Version corrigée du 20 septembre 2015

Sylvain Girard

Remerciements

Ces travaux ont été réalisés au sein du groupe P1C du département STEP¹ du service de recherche et développement d'EDF, du centre de géosciences de l'école des Mines à Fontainebleau et du centre « Énergétique et procédés » de l'école des Mines à Paris. Je tiens à remercier les membres de chacune de ces institutions pour l'accueil toujours chaleureux qu'ils m'ont réservé.

Cette thèse a été dirigée par Hans WACKERNAGEL et j'ai bénéficié de l'encadrement de Thomas ROMARY, Pascal STABAT et Jean-Melaine FAVENNEC. Je tiens à les remercier en premier lieu pour leur bienveillance et leur amicalité. J'ai la chance d'avoir accompli ma thèse dans des conditions que l'on pourrait dire idéales. Ces trois années m'ont apporté énormément et je le leur en suis largement tributaire. Ils ont été de bons guides. De ceux, attentionnés, qui guettent le faux pas ou suggèrent telle prise qu'ils savent « crochetante », mais qui cèdent généreusement la tête de la cordée. Je suis honoré de la confiance qui m'a été accordée. L'autonomie dont j'ai disposé m'a été très profitable. Ce sont en effet souvent les détours, les contremarches, les fausses pistes même, qui sont les plus riches d'enseignements.

Je remercie Clémentine PRIEUR, Marc GRANDOTTO et Gérard BIAU d'avoir accepté de s'intéresser à mes travaux et de les évaluer. J'espère que les pages qui suivent pourront leur transmettre une part de mon enthousiasme pour ce sujet.

Les travaux sur le modèle de générateur de vapeur présentés ici sont le fruit d'une collaboration avec Julien NINET et Olivier DENEUX. Je garderais un souvenir sans doute bien différent de cette vaste besogne si je ne l'avais partagée comme ce fût le cas, dans un vrai esprit de camaraderie.

La complicité de Thomas PRUSEK, avec qui j'ai partagé l'expérience du doctorat, m'a aussi été fort appréciable. Le chapitre sur l'échantillonnage des profils de colmatage n'aurait put être écrit sans son aide et ses travaux.

Plusieurs personnes m'ont apporté leur appui, toujours avec beaucoup de diligence et de sagacité. Leurs idées ont plus d'une fois démêlé des situations qui me semblaient inextricables. Je souhaite ainsi remercier Ouardia TOUAZI, Lorenzo AUDIBERT, Franck DAVID, Bertrand IOOSS et Alexandre GIRARD.

Souvent, pendant l'effort des derniers mois, mes pensées se sont tournées vers ceux que j'aime, mes parents, mon frère et Marion. Leur soutien permanent m'est inestimable. Ce sont eux qui me permettent de donner un sens à ces trois années de travail, au-delà des équations.

1. Simulation et traitement de l'information pour l'exploitation des moyens de production

Table des matières

1. Introduction	14
I. Colmatage des générateurs de vapeur	16
2. Générateurs de vapeur des centrales nucléaires	17
2.1. Fonctionnement des générateurs de vapeur	19
2.1.1. Mesures	20
2.2. Dégradation des générateurs de vapeur par dépôt d'oxydes	23
2.2.1. Encrassement	24
2.2.2. Colmatage	24
2.2.3. Opérations de prévention et de maintenance	26
3. Méthodes existantes de diagnostic du colmatage	28
3.1. Examens télévisuels de la plaque supérieure	28
3.1.1. Incertitude	29
3.1.2. Taux de colmatage uniforme équivalent	35
3.2. Méthode par courants de Foucault	38
3.3. Suivi du niveau gamme large en régime stationnaire à pleine puissance . .	39
3.4. Analyse de la réponse dynamique du niveau gamme large	40
3.4.1. Un transitoire bien commode : l'essai périodique R.G.L. 4	40
3.4.2. Méthode à profil choisi	42
3.5. Extension de la méthode à profil choisi par filtrage particulaire	45
3.5.1. Principe du filtrage particulaire	46
3.5.2. Mise à l'épreuve de l'algorithme	50
3.6. Synoptique des méthodes de diagnostic existantes	56
II. Modélisation physique des générateurs de vapeur	58
4. Contexte de la modélisation physique	59
4.1. Présentation de quatre modèles comparables à celui utilisé	60
4.2. Un modèle de référence : THYC	62
4.3. Version initiale du modèle de générateur de vapeur	64
4.3.1. Problèmes inhérents au modèle	65
4.3.2. Méthodologie de calage du modèle	68

5.	Modèle dynamique monodimensionnel de générateur de vapeur	71
5.1.	Cadre général de la modélisation	72
5.1.1.	Résolution du système	73
5.1.2.	Schéma numérique	74
5.1.3.	Structure du modèle mathématique	75
5.2.	Écoulement monophasique monodimensionnel	76
5.2.1.	Équations de conservation pour l'écoulement monophasique	77
5.2.2.	Formulation pratique des équations de conservation	78
5.2.3.	Équations constitutives de l'écoulement monophasique	79
5.3.	Écoulement diphasique monodimensionnel	80
5.3.1.	Opérateurs de moyenne	82
5.3.2.	Taux de vide, titre et rapport de vitesses des phases	83
5.3.3.	Équations de conservation pour l'écoulement diphasique	84
5.3.4.	Équations constitutives de l'écoulement diphasique	87
5.4.	Plaques entretoises	93
6.	Calage et validation du modèle de générateur de vapeur	94
6.1.	Méthodologie de calage du modèle	94
6.2.	Calage en régime stationnaire à puissance nominale	98
6.3.	Calage en régime transitoire	101
6.4.	Validation en régime stationnaire à basse puissance	105
III.	Modèles et méthodes statistiques	111
7.	Analyse de sensibilité globale	112
7.1.	Choix d'une méthode d'analyse de sensibilité	112
7.2.	Indices de Sobol'	114
7.2.1.	Décomposition ANOVA	116
7.3.	Estimation des indices de Sobol'	118
7.3.1.	Estimation des indices d'ordre 1	118
7.3.2.	Estimation des indices totaux	119
7.3.3.	Remarques sur la convergence	120
7.4.	Intervalles de confiance pour les indices de Sobol'	121
7.4.1.	Principe général du <i>bootstrap</i>	121
7.4.2.	Intervalles de confiance <i>bootstrap</i> quantiles	122
8.	Analyse en composantes principales	123
8.1.	Variance des composantes principales	124
8.2.	Composantes principales d'un échantillon	125
8.2.1.	Composantes principales normalisées	126
8.3.	Analyse de sensibilité d'une sortie fonctionnelle	127
8.3.1.	Cas de courbes échantillonnées	127
8.3.2.	Application à l'analyse de sensibilité d'une sortie fonctionnelle	129

9.	Régression inverse par tranche	131
9.1.	Principe général de la régression inverse par tranches	132
9.1.1.	Sous-espace central	132
9.1.2.	Courbe de régression inverse	133
9.2.	Algorithme de la régression inverse par tranches	134
9.2.1.	Détermination de la dimension du sous-espace de projection	135
9.2.2.	Remarque sur l'échantillonnage	137
9.3.	Régression inverse par tranches multidimensionnelle	139
IV.	Diagnostic du colmatage	141
10.	Développement d'une nouvelle méthode de diagnostic du colmatage	142
10.1.	Prétraitement des données	143
10.1.1.	Filtrage de l'effet statique	143
10.1.2.	Amplitude de variation de la mesure de niveau	145
10.2.	Analyse de sensibilité séquentielle	146
10.3.	Réduction de la dimension de sortie	147
10.3.1.	Composantes principales d'un échantillon de simulations	150
10.3.2.	Composantes principales d'un échantillon de réponses mesurées	150
10.3.3.	Retour sur le filtrage de l'effet statique	153
10.4.	Analyse de sensibilité de la sortie réduite	154
10.5.	Réduction de la dimension d'entrée	156
10.5.1.	Régression inverse par tranche monodimensionnelle	157
10.5.2.	Découpage marginal groupé	164
11.	Échantillonnage des profils de colmatage	168
11.1.	Modèle de colmatage par <i>vena contracta</i>	169
11.1.1.	Calage du paramètre non observable	169
11.2.	Analyse de sensibilité du modèle <i>vena contracta</i>	170
11.2.1.	Analyse de sensibilité du paramètre de calage	171
11.2.2.	Analyse de sensibilité de l'allure des profils	174
11.3.	Distribution de profils de colmatage issue du modèle <i>vena contracta</i>	175
11.3.1.	Mélange de particules de diamètres différents	175
11.3.2.	Rétroaction du champ thermohydraulique	177
11.3.3.	Simulation de la distribution de profils	179
12.	Adaptation au transitoire de référence	182
12.1.	Transitoire de référence	182
12.1.1.	Types d'écarts au transitoire de référence	183
12.2.	Méthode de correction des écarts	184
12.2.1.	Déformation uniforme de l'intervalle de temps	184
12.2.2.	Synchronisation d'évènements caractéristiques	188

12.3. Analyse de sensibilité aux écarts de pente et de valeur initiale	189
12.3.1. Plan d'expérience	190
12.3.2. Indices de sensibilité sans déformation temporelle	192
12.3.3. Indices de sensibilité avec déformation temporelle	194
13. Application pratique de la nouvelle méthode de diagnostic	201
13.1. Réduction de la dimension d'entrée avec l'échantillon <i>vena contracta</i> . . .	202
13.2. Définition du nouvel indicateur de colmatage	203
13.3. Application de la nouvelle méthode de diagnostic	205
V. Conclusion	217
14. Perspectives de transposition de la méthodologie à l'étude de l'encrassement	218
14.1. Indicateurs existants	218
14.1.1. Pression de vapeur corrigée	219
14.1.2. Facteur d'encrassement	219
14.1.3. Pression en régime dynamique ?	221
14.2. Une bibliographie disparate mais abondante	222
14.2.1. Nature et morphologie	223
14.2.2. Localisation dans le générateur de vapeur	224
14.2.3. Propriétés thermiques	224
14.3. Pistes d'étude	225
14.3.1. Modélisation physique	225
14.3.2. Amélioration de l'indicateur pression corrigée	226
14.3.3. Amélioration de l'indicateur facteur d'encrassement	227
14.3.4. Vers le diagnostic	227
15. Synthèse et recommandations d'usage	229
15.1. Synthèse des travaux réalisés	229
15.2. Recommandations d'usage et perspectives d'amélioration	231
A. Caractéristiques géométriques des générateurs de vapeur de type 51 B	233
B. Modèles de déposition BOUTHYC et COLMATHYC	234
C. Régression inverse par tranche avec un échantillon uniforme	236
C.1. Régression inverse par tranche monodimensionnelle	236
C.2. Découpage marginal groupé	240
D. Mesures granulométriques	242
E. Régression inverse par tranche avec l'échantillon <i>vena contracta</i>	243
E.1. Régression inverse par tranche monodimensionnelle	243
E.2. Découpage marginal groupé	247

F. Réponses de niveau gamme large mesurées	249
G. Publications	252

Table des figures

2.1.	Schéma d'une tranche nucléaire à eau pressurisée.	18
2.2.	Schéma d'un générateur de vapeur de type 51 B.	21
2.3.	Motif appelé <i>quatre-feuilles</i>	25
2.4.	Schéma du détail d'une de plaque entretoise.	25
2.5.	Photographies de trous foliés sain et colmaté issues d'examens télévisuels.	25
2.6.	Schéma d'un trou folié colmaté.	26
3.1.	Échantillon d'estimations unitaires de taux de colmatage par examen télévisuel.	32
3.2.	Variogramme empirique de données issues d'un examen télévisuel.	33
3.3.	Comportement des résidus.	34
3.4.	Variogramme empirique des résidus.	34
3.5.	Modèle de perte de charge déduit de l'expérience P2C	37
3.6.	Taux de colmatage uniforme équivalent de la tranche n° 1 de la centrale Bravo en mai 2011.	38
3.7.	Faisceau de courbes de réponse de n. g. l. mesurées sur le g. v. n° 3 de la tranche n° 2 de Charlie.	43
3.8.	Quinze configurations de colmatage utilisées par la méthode à profil choisi.	44
3.9.	Base du sous-espace d'initialisation de l'algorithme de filtrage particulaire.	49
3.10.	Diagramme de Tukey des distributions des diagnostics obtenus pour les 5 graines.	52
3.11.	Évolution de l'entropie et de la vraisemblance pour les 5 graines.	53
3.12.	Évolution de l'entropie et de la vraisemblance en fonction de la limite de rééchantillonnage.	54
3.13.	Diagramme de Tukey des diagnostics avec rééchantillonnage systématique pour deux graines.	55
3.14.	Régions explorées par le filtre particulaire.	56
4.1.	Maillage du modèle de générateur de vapeur conçu avec le code THYC.	63
4.2.	Comparaison de courbes de réponse mesurées et simulées avec les version 2 et 7 du modèle de générateur de vapeur.	66
5.1.	Schéma numérique à mailles décalées.	75
5.2.	Schéma d'un sous-canal de l'écoulement secondaire diphasique.	90
6.1.	Illustration par un modèle artificiel du compromis entre l'envergure du domaine de validité et la performance sur un domaine restreint.	95

6.2.	Effet de perturbations du coefficient $M_{\bar{v}r}$ sur la réponse du niveau gamme large lors d'un transitoire.	97
6.3.	Effet d'une perturbation du coefficient $M_{\bar{v}r}$ sur la réponse du niveau gamme large lors d'un transitoire.	98
6.4.	Comparaison en régime nominal des profils de pression.	101
6.5.	Comparaison en régime nominal des profils de titre.	102
6.6.	Comparaison en régime nominal des profils de taux de vide.	102
6.7.	Comparaison en régime nominal des profils d'enthalpie.	103
6.8.	Comparaison en régime nominal des profils de puissance transférée.	103
6.9.	Comparaison en régime nominal des débits transverses.	104
6.10.	Superposition de réponses simulées avec le modèle calé et mesurées.	105
6.11.	Comparaison à basse puissance des profils de pression pour un débit de vapeur de 212 kg s^{-1}	107
6.12.	Comparaison à basse puissance des profils de pression pour un débit de vapeur de 208 kg s^{-1}	107
6.13.	Comparaison à basse puissance des profils de titre.	108
6.14.	Comparaison à basse puissance des profils de taux de vide.	108
6.15.	Comparaison à basse puissance des profils d'enthalpie.	109
6.16.	Comparaison à basse puissance des profils de puissance transférée.	109
6.17.	Comparaison à basse puissance des débits transverses.	110
10.1.	Cinq courbes de réponse simulées, sans prétraitement.	144
10.2.	Cinq courbes de réponse simulées auxquelles sont soustraites leurs valeurs initiales.	144
10.3.	Cinq courbes de réponse simulées auxquelles sont soustraites leurs moyennes temporelles.	144
10.4.	Indices d'ordre 1 séquentiels en jambe chaude.	148
10.5.	Indices d'ordre 1 séquentiels en jambe froide.	148
10.6.	Indices séquentiels totaux en jambe chaude.	149
10.7.	Indices séquentiels totaux en jambe froide.	149
10.8.	Composantes principales (en covariance) d'un échantillon uniforme de courbes de réponses simulées.	151
10.9.	Perturbations de la moyenne par les deux premières directions principales.	151
10.10.	Composantes principales (en corrélation) d'un échantillon uniforme de courbes de réponses simulées.	152
10.11.	Composantes principales (en corrélation) d'un échantillon de courbes de réponses mesurées.	153
10.12.	Composantes principales d'un échantillon de courbes de réponses mesurées avant et après soustraction des moyennes temporelles.	154
10.13.	Indices de sensibilité compacts d'ordre 1 et totaux des scores de la première composante principale.	155
10.14.	Indices de sensibilité compacts d'ordre 1 et totaux des scores de la deuxième composante principale.	156

10.15.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon gaussien et les scores de la première composante principale.	158
10.16.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon gaussien et les scores de la deuxième composante principale.	158
10.17.	Directions de projection pour l'échantillon gaussien et les scores de la première composante principale comme variable de sortie.	160
10.18.	Directions de projection pour l'échantillon gaussien et les scores de la deuxième composante principale comme variable de sortie.	160
10.19.	Nuages des projections sur le sous-espace réduit en fonction des scores de la première composante principale pour l'échantillon gaussien.	162
10.20.	Nuages des projections sur le sous-espace réduit en fonction des scores de la deuxième composante principale pour l'échantillon gaussien.	163
10.21.	Diagrammes de Tukey de \widehat{R}_j pour le découpage marginal groupé de l'échantillon gaussien.	165
10.22.	Directions de projections obtenues par découpage marginal groupé de l'échantillon gaussien.	165
10.23.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la première composante principale, pour l'échantillon gaussien.	166
10.24.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la première composante principale, pour l'échantillon gaussien.	166
10.25.	Schéma du plan de réduction de dimension obtenu par un découpage marginal rassemblant deux courbes de régressions inverses presque rectilignes.	167
11.1.	Nuages de points du paramètre de calage en fonction des cinq variables d'entrée du modèle <i>vena contracta</i>	172
11.2.	Indices de sensibilité des variables d'entrée du modèle <i>vena contracta</i>	173
11.3.	Ajustement de fonctions pour la prédiction du paramètre de calage.	174
11.4.	Profils de colmatages obtenus lors de l'analyse de sensibilité du modèle <i>vena contracta</i>	176
11.5.	Profils de colmatage en fonction du diamètre des particules.	177
11.6.	Base de la distribution de profils de colmatage sans rétroaction.	178
11.7.	Pondération déduite des mesures granulométriques.	179
11.8.	Base de la distribution de profils de colmatage avec rétroaction.	180
11.9.	Échantillon de profils de colmatage issus de la distribution construite avec le modèle <i>vena contracta</i>	181
12.1.	Nuages de points des valeurs initiales des paramètres du transitoire.	185
12.2.	Nuages de points des pentes des paramètres du transitoire.	186
12.3.	Profils de colmatage issus du modèle <i>vena contracta</i>	190
12.4.	Ajustement de la distribution de la température primaire.	191
12.5.	Ajustement de la distribution de la température de l'eau alimentaire.	192
12.6.	Ajustement de la distribution du débit de vapeur.	193

12.7.	Ajustement de la distribution du débit primaire.	193
12.8.	Indices de Sobol' séquentiels sans adaptation au transitoire de référence. . .	195
12.9.	Indices de Sobol' compacts selon la première composante principale sans adaptation au transitoire de référence.	196
12.10.	Indices de Sobol' compacts selon la deuxième composante principale sans adaptation au transitoire de référence.	196
12.11.	Indices de Sobol' séquentiels principale avec adaptation au transitoire de référence en débit de vapeur.	198
12.12.	Indices de Sobol' compacts selon la première composante principale avec adaptation au transitoire de référence en débit de vapeur.	199
12.13.	Indices de Sobol' compacts selon la deuxième composante principale avec adaptation au transitoire de référence en débit de vapeur.	199
12.14.	Indices de Sobol' compacts selon la première composante principale avec adaptation au transitoire de référence en enthalpie alimentaire.	200
12.15.	Indices de Sobol' compacts selon la deuxième composante principale avec adaptation au transitoire de référence en enthalpie alimentaire.	200
13.1.	Direction de projection des profils de colmatage utilisée pour le diagnostic. .	203
13.2.	Nuage de points simulés utilisé pour le diagnostic.	205
13.3.	Diagnostic du colmatage de la tranche n° 2 de la centrale Alpha.	210
13.4.	Diagnostic du colmatage de la tranche n° 3 de la centrale Alpha.	210
13.5.	Diagnostic du colmatage de la tranche n° 4 de la centrale Alpha.	211
13.6.	Diagnostic du colmatage de la tranche n° 1 de la centrale Bravo.	212
13.7.	Diagnostic du colmatage de la tranche n° 2 de la centrale Charlie.	213
13.8.	Diagnostic du colmatage de la tranche n° 3 de la centrale Charlie.	213
13.9.	Diagnostic du colmatage de la tranche n° 4 de la centrale Charlie.	214
13.10.	Diagnostic du colmatage de la tranche n° 1 de la centrale Delta.	215
13.11.	Diagnostic du colmatage de la tranche n° 2 de la centrale Delta.	215
13.12.	Diagnostic du colmatage de la tranche n° 3 de la centrale Delta.	216
13.13.	Diagnostic du colmatage de la tranche n° 4 de la centrale Delta.	216
14.1.	Courbes de réponse de pression de vapeur.	222
C.1.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon uniforme et les scores de la première composante principale.	236
C.2.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon uniforme et les scores de la deuxième composante principale.	237
C.3.	Directions de projection pour l'échantillon uniforme et les scores de la première composante principale comme variable de sortie.	237
C.4.	Directions de projection pour l'échantillon uniforme et les scores de la deuxième composante principale comme variable de sortie.	237
C.5.	Nuages des projections sur le sous-espace réduit en fonction des scores de la première composante principale pour l'échantillon uniforme.	238

C.6.	Nuages des projections sur le sous-espace réduit en fonction des scores de la deuxième composante principale pour l'échantillon uniforme.	239
C.7.	Diagrammes de Tukey de \widehat{R}_j pour le découpage marginal groupé de l'échantillon uniforme.	240
C.8.	Directions de projections obtenues par découpage marginal groupé de l'échantillon uniforme.	240
C.9.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la première composante principale, pour l'échantillon uniforme.	241
C.10.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la deuxième composante principale, pour l'échantillon uniforme.	241
E.1.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon <i>vena contracta</i> et les scores de la première composante principale.	243
E.2.	Diagrammes de Tukey de \widehat{R}_j pour l'échantillon <i>vena contracta</i> et les scores de la deuxième composante principale.	244
E.3.	Directions de projection pour l'échantillon <i>vena contracta</i> et les scores de la première composante principale comme variable de sortie.	244
E.4.	Directions de projection pour l'échantillon <i>vena contracta</i> et les scores de la deuxième composante principale comme variable de sortie.	244
E.5.	Nuages des projections sur le sous-espace réduit en fonction des scores de la première composante principale pour l'échantillon <i>vena contracta</i>	245
E.6.	Nuages des projections sur le sous-espace réduit en fonction des scores de la deuxième composante principale pour l'échantillon <i>vena contracta</i>	246
E.7.	Diagrammes de Tukey de \widehat{R}_j pour le découpage marginal groupé de l'échantillon <i>vena contracta</i>	247
E.8.	Directions de projections obtenues par découpage marginal groupé de l'échantillon <i>vena contracta</i>	247
E.9.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la première composante principale, pour l'échantillon <i>vena contracta</i>	248
E.10.	Nuages des projections sur le sous-espace réduit obtenus par découpage marginal groupé, en fonction des scores de la deuxième composante principale, pour l'échantillon <i>vena contracta</i>	248
F.1.	Faisceaux de courbes de réponse mesurées à la centrale Alpha.	249
F.2.	Faisceaux de courbes de réponse mesurées à la centrale Bravo.	250
F.3.	Faisceaux de courbes de réponse mesurées à la centrale Charlie.	250
F.4.	Faisceaux de courbes de réponse mesurées à la centrale Delta.	251

Liste des tableaux

2.1. Valeurs typiques et incertitudes des principales mesures autour des générateurs de vapeur.	22
3.1. Variance des coefficients du modèle linéaire ajusté sur les données d'examen télévisuel.	35
3.2. Synoptique des méthodes de diagnostic du colmatage.	57
5.1. Comparaison des hypothèses du modèle à deux fluides et du modèle de mélange.	88
6.1. Paramètres de fonctionnement utilisés pour le calage en régime nominal. . .	99
6.2. Paramètres de fonctionnement utilisés pour la validation en régime stationnaire à basse puissance.	106
11.1. Plan d'expérience de l'analyse de sensibilité du modèle <i>vena contracta</i>	171
11.2. Coefficients du modèle de prédiction du paramètre de calage.	174
12.1. Paramètres du transitoire de référence.	183
13.1. Échelle qualitative de gravité du colmatage en fonction de la valeur du nouvel indicateur	206
13.2. Taux de colmatage moyens et uniformes équivalents calculés à partir des résultats d'examens télévisuels.	209
14.1. Valeurs de référence et sensibilités des paramètres affectant la pression. . . .	221
A.1. Caractéristiques du faisceau de tubes.	233
A.2. Caractéristiques des plaques entretoises.	233
D.1. Mesures granulométriques de particules dans l'eau alimentaire.	242

1. Introduction

Les *générateurs de vapeur* (*g. v.*) sont les échangeurs de chaleur qui alimentent les turbines des centrales nucléaires à eau pressurisée. L'eau circulant dans les *g. v.* transporte des particules d'oxydes de fer issues de la corrosion des composants de la centrale. Ces particules s'accumulent dans des trous prévus pour le passage du mélange eau-vapeur et les colmatent progressivement, modifiant ainsi le comportement thermohydraulique des *g. v.* Deux arrêts fortuits des tranches n° 1 et 4 de la centrale de Delta ont eu lieu en 2004 et 2006 pour cause de fissurations de tubes de *g. v.* Le colmatage des *g. v.* a été identifié comme l'une des causes de ces incidents et EDF a alors lancé un programme de recherche pour mettre au point des méthodes de diagnostic de l'état des générateurs de vapeur. De tels diagnostics sont nécessaires pour garantir la sûreté des centrales et optimiser les coûteuses opérations de maintenance permettant de se prémunir du colmatage.

L'une des pistes explorées consiste à analyser le comportement dynamique des *g. v.* à l'aide d'un modèle dynamique monodimensionnel. En effet, la réponse de la différence de pression entre le haut et le bas du *g. v.* lors d'un transitoire de puissance est altérée par le colmatage. Une modélisation adéquate du phénomène permet de reproduire cet effet : pour chaque configuration de colmatage en entrée du modèle, on peut simuler la réponse de la différence de pression. On cherche alors la configuration de colmatage en entrée du modèle permettant d'approcher au mieux les mesures par la simulation. Une première méthode construite à partir de cette idée donnait des résultats encourageants : on pouvait observer l'augmentation progressive du colmatage au cours du temps et sa disparition subite après nettoyage chimique. L'analyse minutieuse des hypothèses formulées implicitement lors du développement de cette première méthode a permis de poser deux questions qui font l'objet de cette thèse :

- « Qu'est-ce-qui, dans la forme des courbes de réponse, est caractéristique de l'état de colmatage ? »
- « Quelle information sur l'état de colmatage peut-on raisonnablement espérer extraire des mesures ? »

L'utilisation de modèles physiques et statistiques a permis de répondre à ces questions et de proposer une nouvelle méthodologie de diagnostic, plus robuste et mieux maîtrisée.

L'exposé est divisé en cinq parties. La première expose la problématique. Le chapitre 2 explique le fonctionnement des *g. v.* et la nature des dégradations qu'ils subissent. Les méthodes de diagnostic du colmatage existantes sont présentées dans le chapitre 3. La seconde partie aborde les travaux effectués sur la modélisation physique des *g. v.* Les résultats principaux de ces travaux sont donnés dans le chapitre 4, qui présente par ailleurs quatre modèles comparables à celui qui a été développé, ainsi qu'un modèle tridimensionnel plus complexe, ayant été choisi comme référence. Le détail des équations physiques du modèle est donné dans le chapitre 5 et la méthodologie mise au point pour le

calibrer et le valider est exposée dans le chapitre 6. La troisième partie fournit les éléments théoriques nécessaires à la compréhension des méthodes statistiques qui ont été employées. Les chapitres 7, 8 et 9 décrivent respectivement la méthode d'analyse de sensibilité globale de Sobol', l'analyse en composante principale utilisée pour réduire la dimension de sortie du modèle et la régression inverse par tranches ou *sliced inverse regression* en anglais (SIR), une technique de réduction de dimension appliquée à l'entrée du modèle. La quatrième partie présente les résultats obtenus à l'aide de ces méthodes. Le chapitre 10 expose comment leur application successive a permis de construire une nouvelle méthode de diagnostic. Les questions du prétraitement des données et de l'échantillonnage des configurations de colmatage sont traitées respectivement dans les chapitres 11 et 12. Le principe de la nouvelle méthode est synthétisé dans le chapitre 13 qui présente ensuite les résultats obtenus par son application à un ensemble de données issues du parc français. La dernière partie se compose des chapitres 14 et 15. Le premier évoque les possibilités d'adaptation des travaux accomplis sur le colmatage à un autre dysfonctionnement des g. v., l'encrassement. Le second, enfin, est une synthèse de l'ensemble des travaux réalisés, suivie de recommandations pour l'usage de la nouvelle méthode proposée.

Première partie

Colmatage des générateurs de vapeur

2. Générateurs de vapeur des centrales nucléaires

Une centrale nucléaire est une installation industrielle de production d'électricité comportant un ou plusieurs réacteurs nucléaires. Ces réacteurs sont le siège de réactions de fission en chaîne produisant de la chaleur qui est utilisée pour vaporiser de l'eau. La vapeur ainsi produite alimente des turbines couplées à un alternateur qui produit l'électricité. On appelle tranche nucléaire une installation comprenant un réacteur et tous les systèmes nécessaires pour produire de l'électricité de façon autonome. Les centrales nucléaires françaises exploitées par EDF sont toutes des centrales à eau pressurisée. Cette dénomination signifie que le fluide caloporteur circulant dans le réacteur pour extraire la chaleur est de l'eau ordinaire maintenue à l'état liquide par une forte pression.

Le circuit contenant l'eau traversant le réacteur est dit *primaire*. La figure 2.1 représente schématiquement une tranche nucléaire à eau pressurisée. Le circuit primaire y est figuré par des flèches orange et bleues (↗ et ↘), respectivement pour les branches dites *chaude* et *froide*. Le premier composant à gauche du schéma, portant le numéro ①, est le réacteur de la tranche. Il se compose d'une cuve métallique étanche dans laquelle est installé le cœur du réacteur, c'est-à-dire l'assemblage des éléments de combustibles (②). De façon simplifiée, il s'agit de cylindres métalliques étanches contenant les matériaux fissiles, en général des oxydes d'uranium enrichi de 3 % à 5 %. Le réacteur comporte aussi des dispositifs de régulation des réactions de fission, notamment des barres de contrôle absorbant les neutrons pouvant être insérées graduellement dans le cœur (③). Plusieurs types de réacteurs sont exploités actuellement en France. Ils se différencient notamment par la puissance qu'ils sont capables de produire. Les centrales nucléaires françaises sont regroupées en quatre « paliers » selon leur puissance électrique nominale : 900 MW, 1300 MW, 1400 MW (palier dit « N4 ») et 1650 MW (réacteur EPR). Les paramètres de fonctionnement qui sont donnés dans cette introduction concernent le palier 900 MW dont les réacteurs produisent en régime nominal une puissance thermique de 2785 MW.

Le composant portant le numéro ④ est un réservoir appelé pressuriseur. Sa fonction est de maintenir l'eau du circuit primaire à l'état liquide. En régime nominal de fonctionnement, l'eau primaire d'un réacteur a une température comprise, suivant la position considérée, entre 285,8 °C et 322,9 °C et une pression de 155 bar.

On peut voir sur le schéma de la figure 2.1 un autre circuit, dit *secondaire*, figuré par des flèches vertes et bleu ciel (↗ et ↘). Le changement de couleur du vert au bleu ciel indique la vaporisation. C'est dans ce dernier que circule l'eau qui alimente les turbines (⑤). Ces dernières sont couplées à un alternateur (⑥) qui produit l'électricité. Les circuits primaire et secondaire sont disjoints : l'eau ayant traversé le réacteur ne se mélange jamais avec l'eau qui est vaporisée. En revanche, l'énergie thermique transportée par le circuit primaire est transférée au circuit secondaire par l'intermédiaire d'échangeurs de chaleur appelés *g. v.* Pour atteindre les turbines, les conduites du circuit secondaire

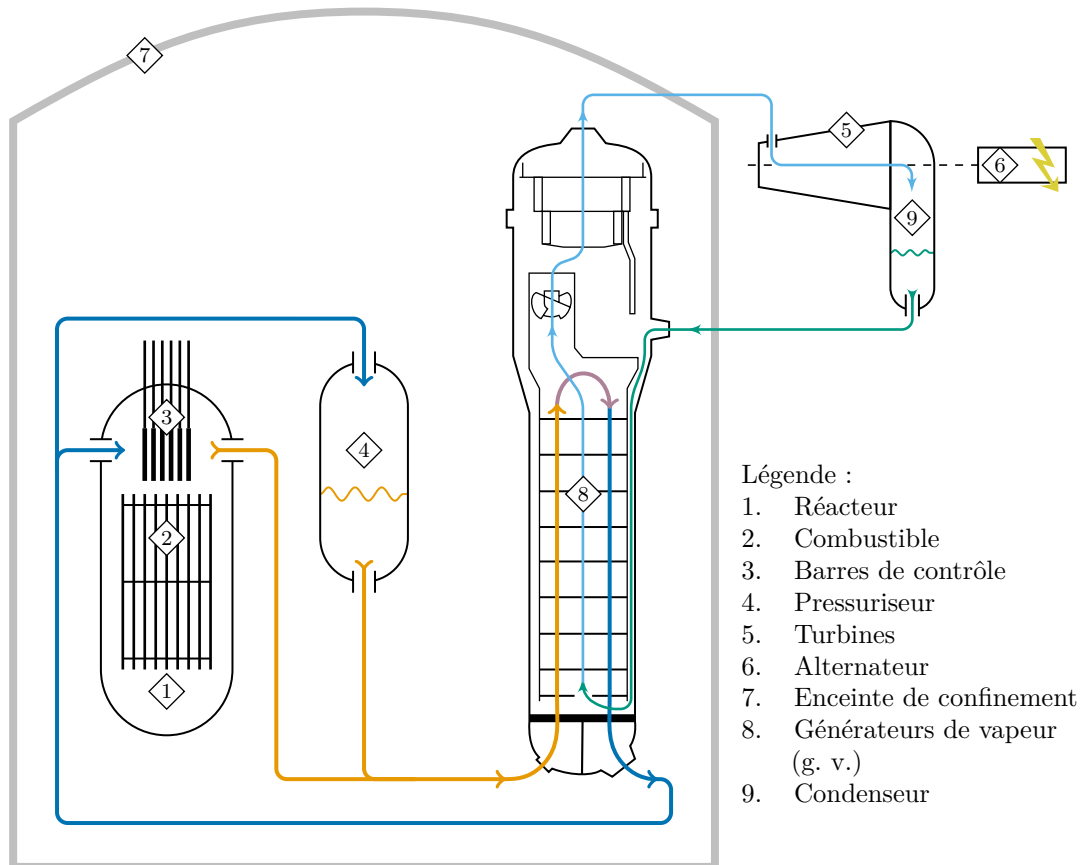


Figure 2.1. : Schéma d'une tranche nucléaire à eau pressurisée.

franchissent une enceinte de béton (7) confinant le réacteur, le pressuriseur et les g. v. Un g. v. unique est représenté dans la figure 2.1 (8). Le circuit primaire des centrales du palier 900 MW comportent en réalité trois boucles disposées autour du réacteur et autant de g. v. Ceux des centrales des palier 1300 MW, N4 et EPR en comportent quatre. Après sa détente dans les turbines, la vapeur est condensée par un échangeur, appelé condenseur (9). Enfin, l'eau sortant du condenseur est réchauffée par d'autres échangeurs, non représentés sur le schéma, et retourne alimenter les g. v.

Le principe de fonctionnement des g. v. est décrit en détail dans la section 2.1 et la section 2.2 présente les dégradations des générateurs de vapeur causées par l'accumulation de dépôts d'oxydes sur leurs composants internes. L'objectif de cette thèse est de proposer une méthode de diagnostic de l'une de ces dégradations appelée *colmatage*.

2.1. Fonctionnement des générateurs de vapeur

Les g. v. sont d'imposants échangeurs de chaleur d'une hauteur d'environ 20 m et d'un diamètre de 3,5 m dans lesquels l'eau du circuit secondaire est vaporisée par la chaleur transportée par l'eau du circuit primaire. Le fonctionnement des g. v. est décrit par Bertin et Pécout (2003) et EDF (1977). La figure 2.2 page 21 représente schématiquement un g. v. de type 51 B et l'annexe A page 233 donne le détail de ses caractéristiques géométriques et de ses paramètres de fonctionnement. C'est ce type de g. v. qui a été étudié dans cette thèse et les paramètres numériques donnés dans la suite de ce chapitre lui sont relatifs. Les g. v. de type 51 A, 51 BI et 51 M sont très similaires au g. v. de type 51 B et on les assimilera à ce dernier. Ces g. v. équipent l'essentiel des tranches du pallier 900 MW.

Un g. v. est composé d'une *boîte à eau* hémisphérique (① sur sur figure 2.2) séparée, par une plaque horizontale dite « à tubes » (②), d'une cuve cylindrique d'une hauteur d'environ 9 m (③), laquelle est elle-même surmontée d'un dôme (④). La boîte à eau est séparée en deux par une plaque verticale (⑤). L'une de ses moitiés reçoit l'eau chaude provenant du réacteur (⑥). L'eau circule ensuite dans un faisceau de 3330 tubes (⑦) traversant la plaque à tubes pour arriver dans la zone d'échange thermique appelée *riser* (⑧). Le faisceau de tube est l'interface entre l'eau des circuits primaire et secondaire. Le transfert thermique s'effectue au travers de la paroi des tubes. Les tubes ont un diamètre intérieur de 22,22 mm et une épaisseur de 1,27 mm. Les tubes sont en forme de *U* et retraversent la plaque à tubes; l'eau refroidie est déversée dans l'autre moitié de la boîte à eau avant de repartir vers le réacteur(⑨). Le débit d'eau primaire est important, de l'ordre de $16\,000\text{ t h}^{-1}$, soit environ 4400 kg s^{-1} .

Passons du côté secondaire. Le g. v. est alimenté en eau à l'état liquide par le haut de la cuve (⑩). La cuve contient une plaque cylindrique, la chemise du faisceau (⑪), séparant la zone de descente de l'eau alimentaire, appelée *downcomer* (⑫), du *riser* qui contient le faisceau de tubes. La chemise du faisceau s'arrête 25 cm au dessus de la plaque à tubes délimitant ainsi la zone d'admission de l'eau alimentaire dans le *riser*. On désigne par *jambe chaude*, la moitié du *riser* contenant les branches des tubes en *U* dans lesquelles l'eau primaire est ascendante. L'eau primaire provient alors du réacteur et est effectivement plus chaude que dans les branches descendantes des tubes en *U* situées dans l'autre moitié du *riser* et appelée *jambe froide*. L'eau circulant dans le *riser* s'échauffe au contact des tubes et son titre en vapeur augmente à mesure qu'elle s'élève. Le g. v. est un échangeur à circulation naturelle, c'est-à-dire que la force motrice qui assure la circulation du fluide est la différence de masse volumique entre la colonne liquide dans le *downcomer* et la colonne de mélange liquide-vapeur dans le *riser* : c'est un thermosiphon. Remarquons que, du fait de la forme en *U* des tubes, l'échangeur est à co-courant dans la *jambe chaude* et à contre-courant dans la *jambe froide*.

En haut du *riser*, juste au dessus de la partie cintrée des tubes appelée *chignon* du faisceau (⑬), le titre du mélange secondaire en régime nominal est de l'ordre de 0,25. Cependant, les turbines des centrales nucléaires nécessitent une vapeur sèche, notamment car des gouttelettes d'eau à grande vitesse endommageraient leurs pales. C'est pourquoi le mélange traverse d'abord des séparateurs appelés cyclones (⑭), au nombre de 3 pour

les g. v. étudiés) dont la forme en hélice permet d'extraire l'eau par centrifugation, puis des cadres sécheurs (⑮), qui sont des enchaînements de chicanes de tôles capturant les gouttelettes d'eau, avant de sortir du g. v. par une ouverture au sommet du dôme (⑯). La vapeur sortant des dispositifs de séparation et de séchage est à saturation et son taux d'humidité résiduelle, appelé taux de primage, est maintenu, en temps normal, inférieur à 0,0025. L'eau liquide recueillie par ces dispositifs est dirigée vers le *downcomer*. Ainsi, l'eau à l'admission du *riser* est presque à saturation ; l'élévation de température nécessaire pour la vaporiser n'est que d'une quinzaine de degrés. On définit le taux de recirculation comme le rapport du débit total traversant le faisceau par le débit de vapeur sortant du g. v. En régime stationnaire, ce dernier est égal au débit d'eau alimentaire. Le taux de recirculation des g. v. 51 B a une valeur en régime nominal d'environ 4,1, égale à l'inverse du titre du mélange à l'admission des cyclones.

Des données issues de 11 tranches équipées de g. v. de type 51 B ou apparentés ont été utilisées pour constituer les exemples présentés dans cette thèse. Elles sont réparties dans quatre centrales auxquelles, pour des raisons de confidentialité, des surnoms ont été attribués : Alpha, Bravo, Charlie et Delta. Les deux premières ont des cinétiques de colmatage plutôt lentes tandis que les deux autres ont atteint des niveaux de colmatage importants ayant nécessité récemment des opérations de nettoyage chimique. Les deux paires diffèrent aussi par leurs conditionnements chimiques. Ce terme sera explicité dans la section 2.2.3.

2.1.1. Mesures

La plupart des mesures effectuées par les capteurs d'exploitation pour la conduite et le contrôle-commande sont enregistrées dans une base de données accessible par une application baptisée « Ouverture du réseau local industriel » (ORLI). La fréquence d'échantillonnage de la plupart des capteurs utilisés dans le cadre de cette thèse est de 2s. Suivant les sites et les capteurs, les données antérieures à 2008 (ou 2006 pour les sites pilotes) peuvent avoir fait l'objet de compression. Celle-ci augmente la bande morte des mesures, de l'ordre de quelques pour-cent, remplaçant ainsi certaines variations de faible amplitude par des plateaux.

Le tableau 2.1 détaille les valeurs typiques en régime nominal des principales mesures autour du g. v. ainsi que les incertitudes associées. Certaines de ces mesures sont effectuées par plusieurs capteurs simultanément. Cette redondance a pour objectif de parer à une défaillance matérielle ou à une dérive de l'un des capteurs. Ces valeurs peuvent varier sensiblement d'une tranche à l'autre, voire d'un g. v. à l'autre, l'instrumentation en place n'étant pas identique. Les données du tableau 2.1 se rapportant au circuit secondaire concernent la centrale Charlie (Douetil, 2008). Précisons qu'il s'agit là des mesures d'exploitation disponibles en continu. D'autres mesures plus précises sont réalisées périodiquement, notamment pour contrôler la puissance échangée par les g. v. Elles n'ont cependant pas été utilisées dans cette thèse.

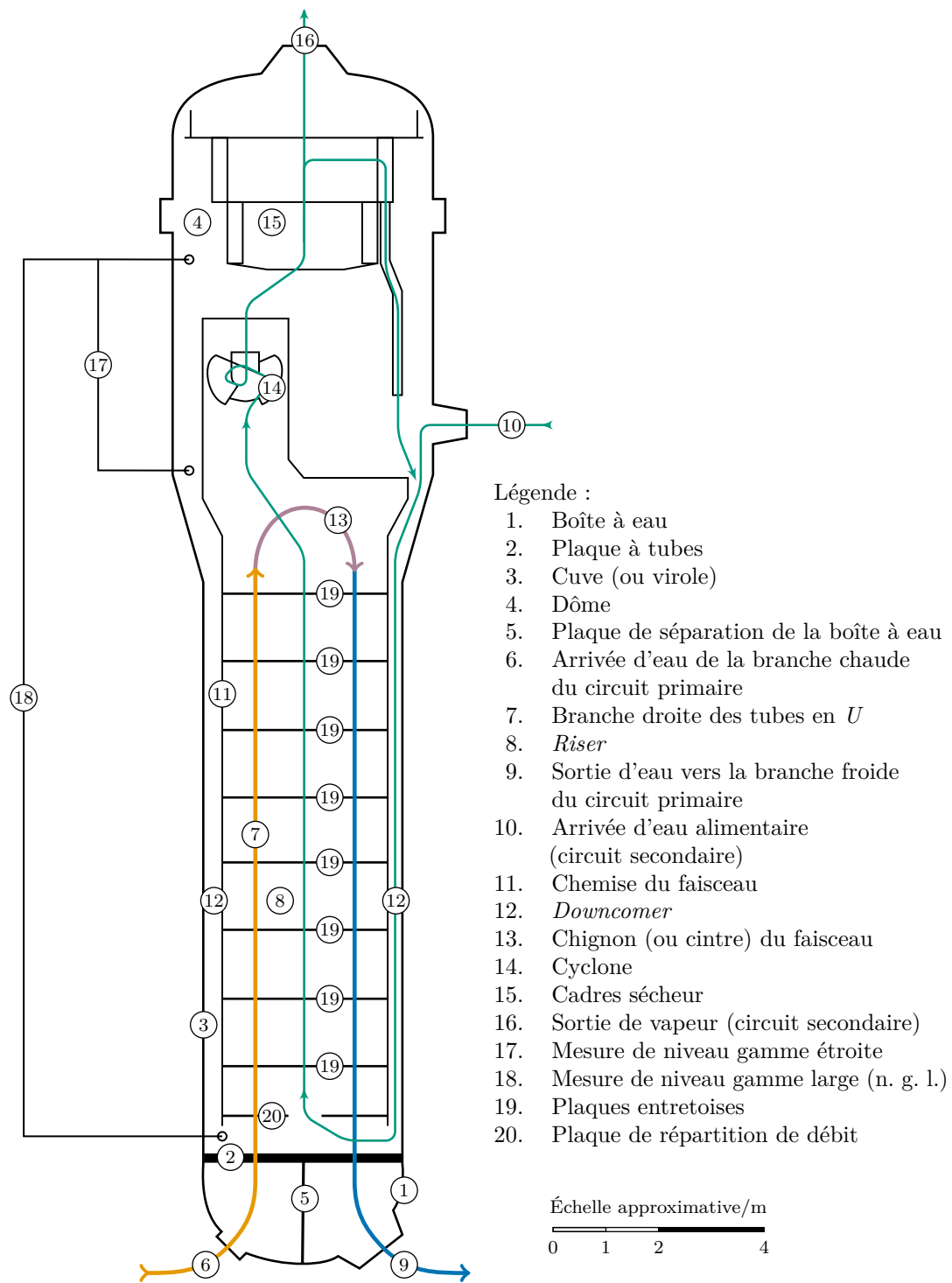


Figure 2.2. : Schéma d'un g. v. de type 51 B en coupe frontale.

Grandeur mesurée	Unité	Valeur typique	Incertitude	Capteurs par g. v.
Pression de la vapeur	bar	56,5	1,6	2
			0,5	1
Débit de vapeur	$t h^{-1}$	1830	45	2
	$kg s^{-1}$	508,3	12,5	
Débit alimentaire	$t h^{-1}$	1813	31,3	2
	$kg s^{-1}$	503,6	12,5	
Température alimentaire	$^{\circ}C$	218	3,6	1
Débit de purge	$t h^{-1}$	50	0,8	1 par tranche
	$kg s^{-1}$	13,8	0,2	
Température de purge	$^{\circ}C$	182	1,5	1 par tranche
Température primaire				
– en entrée de g. v.	$^{\circ}C$	322,9	1	4
– en sortie de g. v.	$^{\circ}C$	285,8	0,6	4
Débit primaire	$t h^{-1}$	15876	612	Essai ponctuel
	$kg s^{-1}$	4410	170	

Tableau 2.1. : Valeurs typiques et incertitudes des principales mesures autour du g. v. accessibles par l'application ORLI (Regaldo et coll., 1984; Douetil, 2008; Deneux et Favennec, 210).

Mesures de niveau dans le générateur de vapeur

Deux capteurs sont destinés à la mesure du niveau d'eau dans le g. v. (Roy, 1985). Le terme de niveau d'eau ne s'applique au sens propre que dans le réservoir, c'est-à-dire l'espace d'arrivée de l'eau alimentaire et de collecte de l'eau extraite par les cyclones, en haut du *downcomer*, à l'extérieur de la chemise du faisceau. En effet, la vapeur dans le *riser* n'est pas séparée de l'eau liquide par une surface libre : le fluide qui y circule est un mélange dont le titre augmente progressivement. La notion de régime d'écoulement décrivant l'agencement des deux phases dans l'écoulement sera abordée dans le chapitre 5. Le niveau dans le réservoir est suivi et régulé par le système de contrôle-commande. Un niveau trop bas ferait courir le risque d'un début de vaporisation dans le cœur car l'extraction de chaleur par les g. v. serait insuffisante. À l'inverse, un niveau excessif entraînerait une humidité importante de la vapeur pouvant être dommageable aux turbines.

Le niveau est estimé à partir de l'écart de pression entre deux piquages situés à des cotes différentes. La mesure est donc affectée, entre autres, par les variations de masse volumique du fluide. Elle est par ailleurs sensible aux pertes de charges parasites et à l'hétérogénéité de l'écoulement dans le *downcomer*. Le *niveau gamme étroite* est utilisé pour contrôler le débit d'eau alimentaire (⑰ sur la figure 2.2). Il est déduit de la différence de pression entre le dôme et le bas du réservoir. Hormis lors de transitoires très rapides pouvant perturber la mesure, par exemple en cas de dépressurisation accidentelle, le niveau gamme étroite est maintenu constant. Le *niveau gamme large* (*n. g. l.*) est quant

à lui déduit de la différence de pression entre le dôme et le bas du *downcomer* (18 sur la figure 2.2). Il est par conséquent beaucoup plus sensible à la température et au débit de l'eau alimentaire, ainsi qu'au taux de recirculation. Il n'est utilisé pour suivre le niveau que lors de régimes lentement variables, notamment comme aide à la conduite manuelle aux basses charges. Lors de transitoires plus rapides, la mesure de n. g. l. n'est plus représentative du niveau car elle est trop affectée par la pression dynamique. C'est cet effet qui est utilisé pour le diagnostic du colmatage par la méthode qui fait l'objet de cette thèse. L'impact du colmatage sur le n. g. l. sera abordé plus en détails dans le chapitre suivant.

2.2. Dégradation des générateurs de vapeur par dépôt d'oxydes

L'eau alimentaire des g. v. transporte des particules et des espèces dissoutes issues de l'oxydation des composants du circuit secondaire. Une partie des produits de corrosion est éliminée par les purges situées en bas des g. v. ; l'essentiel circule cependant dans le *riser* et est susceptible de se déposer sur les parties internes des g. v. Les particules ne sont pas entraînées par la vapeur sortant des g. v. si bien que la matière y entrant s'y accumule.

Des dépôts prélevés sur des tubes extraits de g. v. après plusieurs cycles de fonctionnement ont été analysés (De Vito, 2002; Dijoux, 2003). De nombreuses espèces chimiques interviennent dans leur composition avec une forte prédominance d'un oxyde fer, la magnétite, de formule chimique Fe_3O_4 . Certains atomes de fer de la magnétite sont parfois remplacés par des atomes de zinc ou d'autres espèces non identifiées. Une faible proportion de cuivre a pu être observée pour les centrales dont le circuit secondaire comporte des composants en laiton. L'interface entre les tubes et les dépôts contient des phases riches en calcium et phosphore. Les g. v. subissent des opérations de maintenance visant à éliminer ces dépôts. L'analyse des boues extraites fait elle aussi ressortir la prédominance de la magnétite. Sont aussi détectées des traces de nombreux autres oxydes à base de zinc, phosphore, silicium, chrome, nickel, manganèse, titane, molybdène. . . Voir par exemple à ce sujet les rapports de Lebrun et Petit (2006), Tessier et Petit (2006) et Lebrun et Petit (2007). La concentration en particules varie fortement avec le régime de fonctionnement et le conditionnement chimique du circuit secondaire. La concentration en espèces dissoutes est difficile à mesurer et reste mal connue (Pujet, 2002). La répartition granulométrique des particules de magnétite a fait l'objet de mesures récentes (Couvidou, 2011).

L'accumulation de produits de corrosion dans les g. v. modifie leur fonctionnement. Suivant la localisation des dépôts, on parle d'*encrassement* ou de *colmatage*. Ces deux phénomènes posent des problèmes de performance mais aussi de sûreté. Ils sont décrits dans les deux paragraphes suivants. Le paragraphe 2.2.3 présente quant à lui les différentes techniques mises en œuvre pour y remédier.

2.2.1. Encrassement

On appelle *encrassement* la formation de dépôts à la surface externe des tubes des g. v. Les dépôts d'encrassement augmentent la résistance thermique des parois des tubes des g. v. Cette résistance plus importante diminue la performance de l'échangeur et affecte la pression de vapeur dans le dôme. Outre la dégradation des performances, l'encrassement peut poser des problèmes de sûreté. À mesure que l'encrassement augmente, l'opérateur est contraint d'ouvrir davantage les soupapes de la ligne de vapeur en sortie du g. v., atteignant parfois la limite d'ouverture maximale.

Les travaux cités dans l'introduction de cette section permettent de se faire une idée de la composition des dépôts. On observe qu'elle est variable d'un g. v. à l'autre et probablement aussi suivant la localisation dans le g. v. Des efforts de recherche importants, mesures en situation réelle ou expériences sur maquettes, ont été consentis afin d'essayer de caractériser la structure des dépôts, leur répartition dans le g. v. ou encore le détail de leur impact sur le transfert thermique. La modélisation du phénomène d'encrassement et de son impact à l'échelle d'un g. v. demeure néanmoins un problème ouvert. Cette vaste problématique est abordée dans le chapitre 14 qui ouvre une perspective sur l'adaptation des travaux réalisés pendant cette thèse à l'étude de l'encrassement.

2.2.2. Colmatage

La tenue mécanique du faisceau de tubes est assurée par des plaques percées appelées *plaques entretoises*, disposées à intervalles réguliers dans le g. v. Ces plaques sont indiquées par le numéro ① sur la figure 2.2. Les g. v. de type 51 B et apparentés en comptent 8. La plaque en bas du *riser*, portant le numéro ② sur la figure 2.2, n'est pas une plaque entretoise. On l'appelle *plaque de répartition de débit* car sa fonction est d'homogénéiser les vitesses du fluide dans le *riser*. Elle est percée d'une ouverture d'un diamètre moyen de 55 cm.

Les plaques entretoises sont des plaques en acier de 30 mm d'épaisseur percées de 6660 trous circulaires dans lesquels passent les tubes du faisceau. Chaque tube est entouré de 4 trous, dits *foliés* car disposés autour des tubes à la façon du motif ornemental appelé quatre-feuilles (voir la figure 2.3). Le schéma de la figure 2.4 représente le détail d'une plaque entretoise.

Les plaques entretoises sont sujettes au phénomène de *colmatage* : l'accumulation de produits de corrosion dans les trous foliés entraîne leur obstruction progressive. La figure 2.5 représente deux photographies de trou folié, l'un exempt de dépôt et l'autre fortement colmaté. Ces images ont été obtenues lors d'examen télévisuels de la plaque supérieure de g. v. Cette technique d'inspection est utilisée pour diagnostiquer le colmatage. Elle est décrite dans la section 3.1.

Le colmatage des plaques entretoises réduit l'aire de leur section passante et augmente la perte de charge singulière que subit le fluide secondaire lorsqu'il les traverse. Quatre risques causés par le colmatage ont été identifiés (Adobes, 2011) :

- une perturbation du champ de vitesse peut engendrer des instabilités vibratoires mettant en danger l'intégrité des tubes ;



Figure 2.3. : Motif appelé *quatre-feuilles* sur le fronton de la porte Ouest de l'abbaye de Croyland dans le comté du Lincolnshire, en Angleterre.

Auteur : Janice TOSTEVIN.

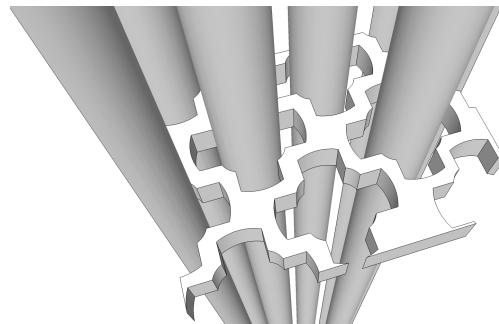


Figure 2.4. : Schéma du détail d'une de plaque entretoise traversée par les tubes du g. v. Les trous foliés sont visibles autour des tubes.

Auteur : Olivier DENEUX.

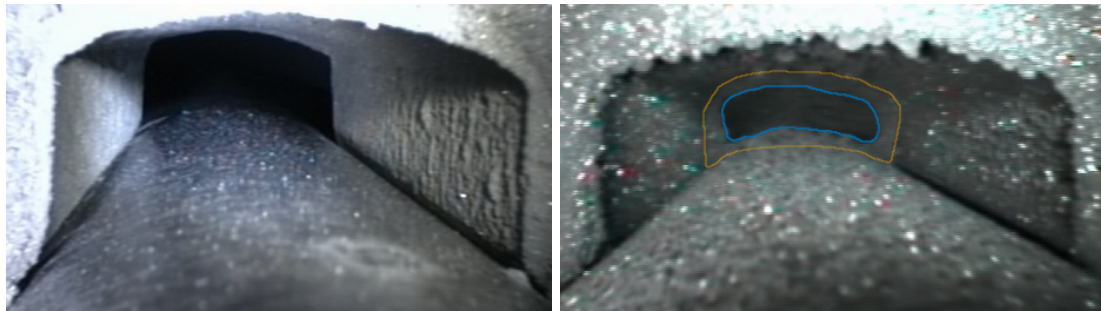


Figure 2.5. : Photographies de trous foliés sain (à gauche) et colmaté (à droite) issues d'examens télévisuels de plaques entretoises supérieures de g. v. Le tracé orange sur la photographie de droite indique la jonction entre le dépôt de colmatage et les bords du trou folié. Le tracé bleu délimite la surface restante pour le passage du fluide. Ces photographie sont prises par le dessus. Le dépôt est donc localisé à la base du trou folié. C'est presque toujours le cas pour les g. v. de type 51 B et aparentés.

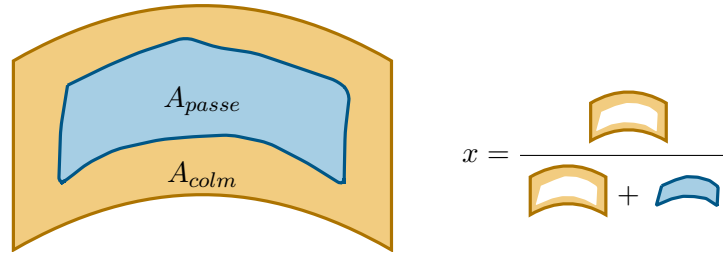


Figure 2.6. : Schéma d'un trou folié colmaté. La surface orange, sur laquelle est inscrit A_{colm} , représente le dépôt de colmatage. La surface bleue, sur laquelle est inscrit A_{passe} , est la surface restante pour le passage du fluide. Le taux de colmatage, v_{colm} , est égal au rapport de l'aire bloquée par l'aire totale.

- lors d'un transitoires très rapides, des oscillations de pression et de température peuvent apparaître survenir et se répercuter sur le cœur ;
- une augmentation localisée du chargement dynamique des plaques entretoises peut rompre les tirants qui les maintiennent en place ;
- une diminution du taux de recirculation, et donc de la masse d'eau dans le g. v., peut compromettre sa capacité à extraire la chaleur résiduelle après une perte d'alimentation en eau.

La réduction de passage d'un trou folié est mesurée par son *taux de colmatage*, égal au rapport de l'aire de la section efficace du dépôt et l'aire du trou folié exempt de colmatage soit, en notant x le taux de colmatage :

$$x = \frac{A_{colm}}{A_{trou}} = \frac{A_{colm}}{A_{colm} + A_{passe}} = 1 - \frac{A_{passe}}{A_{trou}}, \quad (2.1)$$

où A_{colm} , A_{passe} et A_{trou} sont les aires respectivement de la section efficace du colmatage, du passage restant et du trou exempt de colmatage. La figure 2.6 représente schématiquement cette définition. Une définition naturelle du taux de colmatage d'une plaque est son taux de colmatage moyen, c'est-à-dire la moyenne des taux de colmatage de ses trous foliés. On verra cependant que le taux de colmatage uniforme équivalent, défini dans la section 3.1.2, est plus adapté dans certaines situations.

2.2.3. Opérations de prévention et de maintenance

Plusieurs procédés permettent de contenir les phénomènes d'encrassement et de colmatage, ou d'y remédier (Lacroix, 2012). L'augmentation de la basicité de l'eau du circuit secondaire permet de réduire de moitié la quantité d'oxydes qu'elle transporte. Elle est contrôlée par l'ajout d'espèces chimiques à l'eau alimentaire : morpholine, ethanolamine ou ammoniacque (Guivarch et Gay, 2010). Le potentiel hydrogène (pH) de ces différents conditionnements chimiques varient entre 9,2 et 9,8. Cependant, les conditionnements à haut pH déclenchent un phénomène de corrosion ammoniacal en présence de cuivre et sont incompatibles avec les circuits secondaires comportant des éléments en laiton.

Aussi, certaines tranches fonctionnent encore à faible basicité, notamment les tranches n° 2, 3 et 4 de la centrale Charlie et n° 1, 2, 3 et 4 de la centrale Delta (Mercier, 2010). L'ajout de dispersant dans l'eau alimentaire permet de diriger une plus grande proportion d'oxydes vers les purges. Ce procédé, à l'œuvre dans certaines tranches américaines, est testé en 2012 sur l'une des tranches du parc français (Lacroix, 2012).

La plaque à tubes et la zone basse du faisceau sous la plaque de répartition de débit sont nettoyées périodiquement à l'aide de jets d'eau sous pression afin de prévenir la corrosion sous contrainte de la base des tubes. Cette technique de nettoyage par *lançage hydraulique* a été adaptée fin 2006 pour détacher les dépôts de colmatage de la plaque entretoise supérieure de la tranche n° 4 Delta. Un second lançage a été appliqué aux g. v. de la tranche n° 3 Delta à l'été 2007. Les réductions modestes du colmatage obtenues, l'inefficacité du procédé à traiter l'encrassement et sa limitation au nettoyage de la plaque supérieure ont conduit à l'abandonner au profit d'une technique globale, le *nettoyage chimique*.

On distingue, dans la stratégie de maintenance d'EDF, des nettoyages curatifs et préventifs. Il existe plusieurs procédés de nettoyage curatifs. Ils ont en commun leur principe général : la dissolution des dépôts de magnétite par une solution d'acide éthylène diamine tétraacétique (E.D.T.A.) en présence d'une amine et d'hydrazine, à haute température. Les nettoyages préventifs se déroulent dans des conditions chimiques et de température moins agressives et durent moins longtemps. Ils ne permettent d'éliminer qu'une partie des matières déposées et supposent une fréquence d'intervention plus importante. La valeur cible de masse de magnétite ôtée par un nettoyage préventif est d'environ 500 kg par g. v., quand certains nettoyages curatifs éliminent plus de 4 t de matière sèche. Un calcul simple à partir des données géométriques fournies en annexe A permet d'établir que l'essentiel de cette masse est constitué de dépôts d'encrassement et non de colmatage. Cependant, une tranche fortement encrassée a de bonnes chances d'être aussi colmatée. C'est en effet, pour le palier 900 MW, ce que laissent penser les indicateurs de colmatage existants qui sont décrits dans le chapitre suivant. L'essentiel des nettoyages ayant été réalisés pour le parc français sont de type curatif et c'est de ceux-ci dont il sera question par la suite en l'absence de précision.

Les nettoyages chimiques sont des opérations de maintenance lourdes au coût élevé, ne serait-ce que parce qu'elles allongent les durées d'arrêt des centrales entre les cycles. Ils produisent par ailleurs des volumes importants d'effluents qu'il faut traiter. Le remplacement des g. v. est encore plus coûteux (environ trois fois plus) et assujéti à la capacité de production limitée à quelques unités par an. Les enjeux de sûreté liés au colmatage et les fortes contraintes associées aux moyens de s'en prémunir réclament une stratégie de maintenance efficace. Il est nécessaire pour cela de pouvoir diagnostiquer l'état de colmatage des g. v.

3. Méthodes existantes de diagnostic du colmatage

Deux arrêts fortuits des tranches n° 1 et 4 de Delta ont eu lieu en 2004 et 2006, causés par des fuites entre les circuits primaire et secondaire (Juillot Guillard, 2007; Lacroix, 2012). Le colmatage a été identifié comme l'une des causes des fissurations de tubes à l'origine de ces fuites. EDF a alors lancé un programme de recherche pour mettre au point des méthodes de diagnostic du colmatage des g. v.

Le défi principal du diagnostic du colmatage tient à la difficulté d'accès de l'intérieur des g. v. et à l'absence de capteurs internes. Plusieurs méthodes de diagnostic sont actuellement utilisées en routine ou ponctuellement. Des examens télévisuels de la plaque supérieure des g. v. sont réalisés spécifiquement pour diagnostiquer le colmatage lors de certains arrêts de tranche. Cette méthode est présentée dans la section 3.1. D'autres méthodes de diagnostic sont fondées sur l'exploitation de mesures ayant une autre finalité. Le diagnostic par courants de Foucault, décrit dans la section 3.2, tire parti des signaux enregistrés par les sondes à courants de Foucault utilisées lors des arrêts de tranche pour détecter les fissurations de tubes. Deux méthodes utilisent la mesure de n. g. l., une mesure de pression différentielle entre le bas et le haut du g. v. utilisée en exploitation à des fins de régulation et de surveillance. Le suivi du niveau gamme large en régime stationnaire à pleine puissance est décrit dans la section 3.3. Une méthode d'analyse de la réponse dynamique du niveau gamme large est présentée dans la section 3.4. Elle est le point de départ de l'élaboration de la nouvelle méthode de diagnostic proposée dans cette thèse, dont le développement fait l'objet de la partie IV. Les hypothèses de la méthode initiale sont formulées et commentées dans ce chapitre. Parmi elles, l'hypothèse sur la répartition spatiale du colmatage a été jugée la plus contraignante. Un algorithme de filtrage particulière avait été mis au point pour inférer cette répartition à partir de la réponse du n. g. l. Les limites de cet algorithme et les enseignements que l'on peut en tirer sont exposés dans la section 3.5. Un tableau récapitulatif des méthodes exposées est donné en fin de chapitre.

3.1. Examens télévisuels de la plaque supérieure

Les ouvertures en haut des g. v. permettent le passage d'outillages portant une caméra avec laquelle la plaque entretoise supérieure peut être inspectée. Une première méthode d'examen télévisuel a été mise au point en mars 2006. Une valeur de taux de colmatage multiple de 0,25 était attribuée par inspection des clichés des trous foliés. Cette première ébauche a ensuite été progressivement améliorée, notamment grâce à un logiciel calculant

la surface obstruée à partir de contours tracés par un opérateur (Stindel, 2007). La procédure actuelle de traitement est la suivante :

1. Une série de clichés est acquise à l'aide de la caméra robotisée.
2. Pour chaque cliché, l'opérateur ajuste une forme pré-dessinée sur le bord supérieur du trou folié. Ce bord est en général exempt de colmatage. L'ajustement est fait par translation, rotation et homothétie de la forme.
3. Le logiciel calcule l'image de la forme supposée correspondre au bord inférieur du trou par une homothétie de rapport fixé en fonction du matériel utilisé et du milieu (air ou eau). C'est sur le bord inférieur que se dépose la majeure partie du colmatage ; il n'est donc en général pas visible.
4. La forme calculée peut être translatée manuellement afin d'obtenir une meilleure correspondance. Cette étape n'a pas d'impact sur le calcul du taux de colmatage.
5. L'opérateur délimite le contour de la surface libre par des segments ajustés sur le bord interne du dépôt.
6. Le logiciel calcule le rapport des surfaces des contours de la surface libre et du bord inférieur du trou, c'est-à-dire le taux de colmatage.

Lorsque les taux de colmatage semblent très faibles, cette procédure quantitative n'est pas mise en œuvre et les taux sont fixés à une valeur jugée conservatrice de 0,1.

Le robot portant la caméra circule entre les colonnes de tubes, c'est-à-dire dans la direction perpendiculaire à la séparation des jambes chaudes et froides. Seul un trou folié par tube est inspecté et les trous inspectés sont toujours du même côté des tubes. Environ un tiers des colonnes, situées en périphérie de la plaque, ne sont pas accessibles par l'outillage. Le nombre de colonnes inspectées a été fixé à 9. Elles sont espacées aussi régulièrement que possible mais les contraintes d'accès ne permettent pas d'inspecter systématiquement les mêmes colonnes. Le fait que les colonnes périphériques ne puissent être inspectées peut faire craindre un risque d'échantillonnage préférentiel. Il est cependant difficile de quantifier ce risque. Le dépouillement des clichés est fastidieux et seul un sur deux est analysé, avec un objectif de 190 trous inspectés par jambe. Ainsi, les résultats d'examen télévisuels comportent en général entre 150 et 200 estimations unitaires de taux de colmatage. Rappelons que les g. v. de type 51 B comportent 3330 tubes en U , chacun entouré de 4 trous foliés par, soit un total de $3330 \cdot 2 \cdot 4 = 26\,640$ trous foliés sur la plaque supérieur.

3.1.1. Incertitude

Ce sont les moyennes arithmétiques des estimations unitaires des taux de colmatage de chaque demi-plaque qui constituent le diagnostic final utilisé pour définir la stratégie de maintenance. Des tests ont été effectués pour évaluer l'incertitude associée au tracé manuel des contours de colmatage. Les mêmes clichés ont été traités par plusieurs opérateurs mais aucune variabilité important n'a été détecté sur les valeurs moyennes (Stindel, 2007; Monchecourt, 2011). Certains clichés peuvent néanmoins être ambigus à cause d'ombres ou d'un éclairage inadéquat et les contours tracés sont subjectifs. Leur proportion varie

d'un examen à l'autre, suivant l'agencement des tubes et les conditions d'acquisition des images (Audibert, 2012). Un logiciel d'analyse automatique des images est en cours de développement (Audibert et coll., à paraître). Il permettra d'une part, de corriger les écarts de positionnement selon un degré de liberté supplémentaire et d'autre part, de traiter tous les clichés enregistrés, ce qui permettra d'améliorer l'échantillonnage.

L'échantillonnage relativement restreint constitue la source principale d'incertitude sur la moyenne des estimations unitaires. Invoquant le théorème central limite, Stindel (2007) et Monchecourt (2011) proposent pour la moyenne des estimations, m , l'intervalle de confiance à 95 % suivant :

$$\left[m - 1,96 \frac{s}{\sqrt{N}}, m + 1,96 \frac{s}{\sqrt{N}} \right], \quad (3.1)$$

où N est le nombre d'estimations unitaires et s leur écart type. L'incertitude ainsi déterminée est souvent très faible, de l'ordre de 0,002. Il a été décidé d'adopter arbitrairement une valeur conservatrice de 0,05 (Monchecourt, 2011). Les histogrammes de ces échantillons sont en général effectivement d'allure gaussienne. Dans certains cas cependant, la distribution des taux de colmatage est très disparate. L'examen télévisuel du g. v. n° 2 de la tranche n° 1 de la centrale Bravo en mai 2011 (Carradec, 2011) en est un exemple. Environ 85 % des trous inspectés étaient très peu colmatés. Leur taux de colmatage a été estimé globalement à environ 0,15 en jambe chaude et 0,1 en jambe froide. Les 15 % restants des trous inspectés étaient quant à eux fortement colmatés. La forme atypique de ces dépôts compliquant l'estimation, une valeur globale d'environ 0,85 a été avancée. L'impact de la faible proportion de trous fortement colmatés sur la moyenne est limité. On verra dans la section 3.1.2 que leur effet sur la thermohydraulique du g. v. n° 2 n'est cependant pas négligeable. Par ailleurs, l'application du théorème central limite suppose que les valeurs dont on calcule la moyenne sont des réalisations *indépendantes* de variables aléatoires de même loi. Or, les données collectées lors de la plupart des examens télévisuels présentent une structure spatiale apparente qui va à l'encontre de l'hypothèse d'indépendance. Le paragraphe suivant présente une étude de cas illustrant ce point.

Étude de la structure spatiale des estimations par examen télévisuel

La figure 3.1 représente les estimations unitaires de taux de colmatage réalisées lors de l'examen télévisuel de la demi-plaque en jambe chaude du g. v. n° 2 de la tranche n° 3 de la centrale Delta. La couleur évolue du jaune à l'orange pour des taux de colmatage variant de 0,55 à 0,95 et la taille des ronds est proportionnelle au taux de colmatage. Le taux de colmatage dépend manifestement de l'ordonnée qui représente l'éloignement à la bande percée de trous non tubés séparant les deux jambes appelée *rû d'eau*. On peut quantifier cette structuration en calculant le *variogramme empirique* des observations (Rivoirard, 1995, chap. 4). Le variogramme est la demi-variance moyenne de deux points, en fonction de la distance qui les sépare. Soit X la variable aléatoire décrivant le colmatage et $x(i,j)$ le taux de colmatage observé d'un des trous foliés du tube de la $i^{\text{ème}}$ rangée et de la $j^{\text{ème}}$ colonne. Le faisceau de tubes des g. v. 51 B est à pas carré valant $pas = 32,54$ mm. On note \mathbf{r} le vecteur de position d'un trou, que l'on repérera par la position du tube,

sans perte de généralité puisque les trous examinés sont toujours du même côté des tubes. Soit δ_1 et δ_2 les vecteurs unitaires respectivement parallèle et perpendiculaire au rû d'eau. En positionnant l'origine au centre de la plaque, le vecteur de position vérifie $\mathbf{r} = i \cdot pas \delta_1 + (j \cdot pas + \frac{rû}{2}) \delta_2$, où *pas* désigne le pas du faisceau et *rû* la largeur du rû d'eau. On note $\|\mathbf{r}_1 - \mathbf{r}_2\|_2$ la distance séparant deux trous situés en \mathbf{r}_1 et \mathbf{r}_2 . Avec ces notations, le variogramme de X s'écrit :



$$\gamma(h) = \frac{1}{2n(h)} \sum_{\|\mathbf{r}_1 - \mathbf{r}_2\|_2 = h} [x(\mathbf{r}_1) - x(\mathbf{r}_2)]^2, \quad (3.2)$$

où $n(h)$ est le nombre de paires de trous espacés d'une distance h . On peut montrer que la variance de X est la moyenne des valeurs prises par le variogramme pour chacune des distances possibles, pondérées par le nombre de paires espacées de ces distances :

$$\text{var}(X) = \frac{1}{\sum_h n(h)} \sum_h n(h) \gamma(h). \quad (3.3)$$

On parle de variogramme *directionnel* lorsque l'on restreint les distances à une direction donnée.

Le comportement du variogramme à l'origine caractérise la régularité de la variable. Par définition, le variogramme est nul lorsque la distance est nulle. Une discontinuité à l'origine est appelée *effet pépîte* par référence au domaine minier dont le concept de variogramme est issu. L'absence d'un tel décrochement traduit la continuité de la variable étudiée. Plus sa pente à l'origine est faible, plus les variations de la variable associée sont lisses. Notamment, le variogramme est d'allure parabolique à l'origine lorsque la variable est dérivable. Il arrive par ailleurs que le variogramme se stabilise à partir d'une certaine distance. L'existence d'un tel palier signifie que la variable n'est plus structurée spatialement au-delà de cette distance seuil que l'on appelle *portée*.

La figure 3.2 représente les variogrammes des données de la figure 3.1 selon les deux directions indiquées en bas à gauche de la figure 3.1 avec une tolérance angulaire de $\pm 45^\circ$: en bleu () selon δ_1 et en orange () selon r_2 . Les données n'étant pas parfaitement régulièrement réparties, les distances ont été rassemblées par intervalles de $2,2 \cdot pas$ selon δ_1 et $1,8 \cdot pas$ selon δ_2 . La taille des symboles est proportionnelle au nombre de paires comptabilisées, $n(h)$. Il est de l'ordre de 400 à 500 pour les distances médianes. Comme l'observation de la figure 3.1 le laissait supposer, le variogramme selon δ_1 est presque plat. Autrement dit, les données ne sont pas structurées parallèlement au rû d'eau, *au-delà d'une distance d'environ $2 \cdot pas$* . On observerait peut-être une structure à très courte distance si l'on disposait de plus de points séparés d'un pas seulement. Ce type d'information pourrait être exploité pour la construction des modèles de déposition. En effet, le fait que les colmatages de trous proches soient corrélés ou non pourrait permettre d'ajuster les paramètres et l'importance relative des différents mécanismes de déposition proposés. Inspecter des rangs de tubes adjacents serait un gaspillage d'observations ; en revanche l'analyse de tous les clichés collectés au lieu de la moitié permettrait d'en apprendre plus sur la structure du variogramme à l'origine. Cela

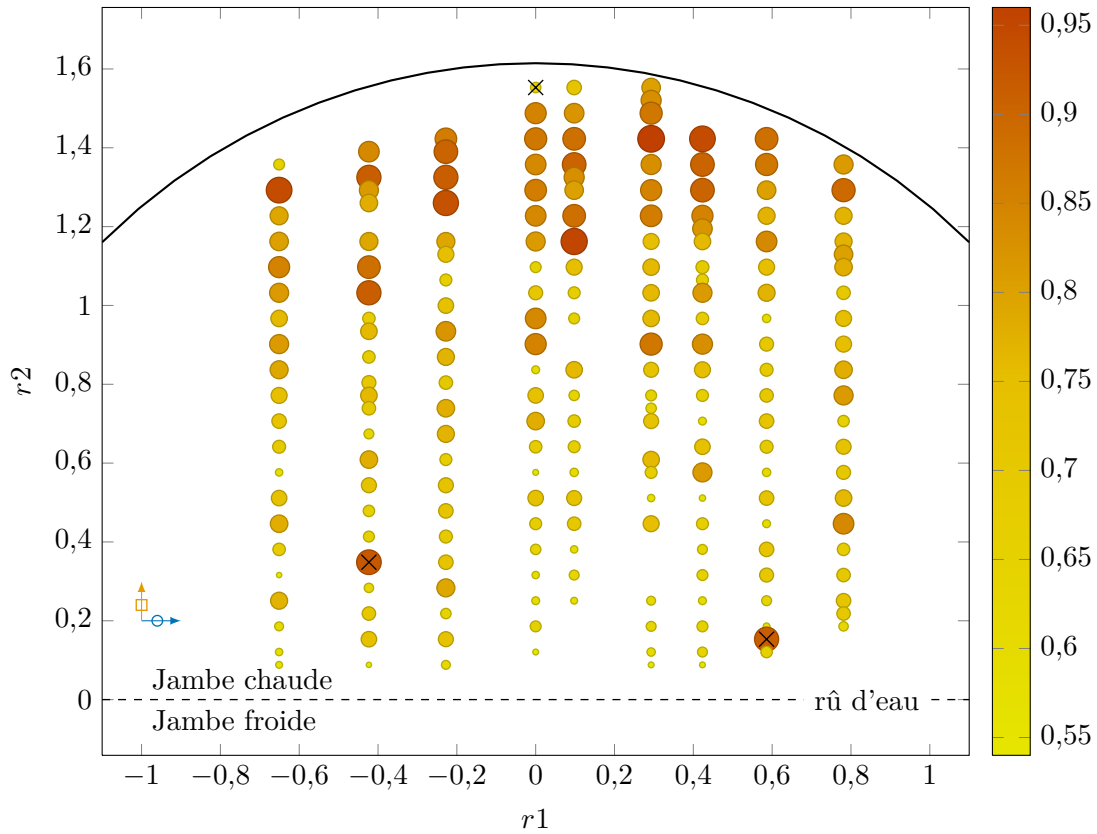


Figure 3.1. : Estimations unitaires de taux de colmatage sur la demi-plaque en jambe chaude du g. v. n° 2 de la tranche n° 3 de la centrale Delta. Les diamètres des cercles sont proportionnels aux taux de colmatage, indiqués par ailleurs par l'échelle de couleur représentée à droite de la figure. Les trois croix noires indiquent des valeurs jugées aberrantes qui ont été écartées de l'échantillon pour ajuster le modèle.

devrait être rendu possible prochainement grâce à l'automatisation du traitement des images (Audibert et coll., à paraître). Le variogramme dans la direction perpendiculaire au rû d'eau comporte, pour des distances de l'ordre de 20 cm, un effet pépite comparable à celui du variogramme selon δ_1 . On peut y voir la conjonction du caractère aléatoire du phénomène de colmatage, des décrochements de dépôts pouvant par exemple survenir de manière chaotique, et l'incertitude de nature aléatoire des estimations unitaires. Au-delà, on observe une structure parabolique qui traduit l'effet des variations des conditions thermohydrauliques de l'écoulement avec l'éloignement au rû d'eau. En effet, les champs transversaux de température, de titre et de vitesse simulés avec « Thermohydraulique des composants » (THYC) présentent eux aussi des gradients marqués dans cette direction. La régularité de cette structure et son origine physique manifeste incite à essayer de la modéliser.

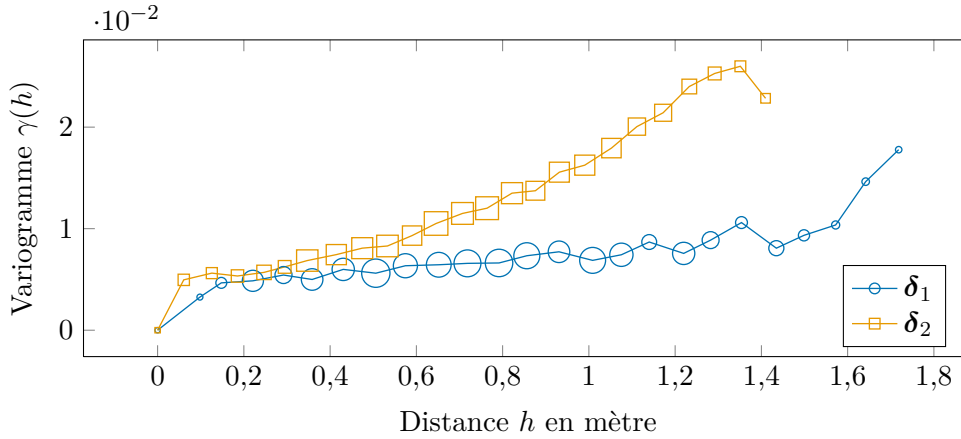


Figure 3.2. : Variogramme empirique des estimations par examen télévisuel représentées par la figure 3.1.

Comme le variogramme s'exprime en fonction du carré des distances, sa structure parabolique suggère un modèle linéaire. La méthode des moindres carrés a été utilisée pour ajuster aux données représentées par la figure 3.1 le modèle suivant :

$$X = a \cdot r_2 + b + \epsilon, \quad (3.4)$$

X est la variable aléatoire représentant le taux de colmatage en un trou folié donné, ϵ un terme d'erreur et r_2 la coordonnée selon δ_2 . Après avoir exclu de l'étude les 3 points, jugés aberrant, marqués d'une croix noire sur la figure 3.1, les estimations suivantes des coefficients du modèle (3.4) ont été obtenues : $\hat{a} = 0,178 \text{ m}^{-1}$ et $\hat{b} = 0,592$. Ce modèle n'est pas très explicatif, son coefficient de détermination valant 0,54. Les estimations des coefficients sont en revanche fortement significatives avec des niveaux de signification¹ très proches de 0. L'histogramme des résidus, représenté à gauche de la figure 3.3, est symétrique et ne comporte pas d'anomalie manifeste. Le terme d'erreur semble par ailleurs homoscédastique à en juger par l'absence de tendance visible dans le nuage de points des résidus en fonction de la variable du modèle, représenté à droite dans la figure 3.3. La figure 3.4 représente les variogrammes des résidus calculés dans les mêmes conditions que ceux de la figure 3.2. Ces deux variogrammes se résument à des effets pépite de variances identiques. Cela signifie que le modèle linéaire proposé saisit l'essentiel de la structure spatiale des données.

La variance, sachant les données, d'une prédiction du modèle en un point de coordonnée r_2 selon δ_2 vaut :

$$\text{var}(\hat{X}(r_2)) = r_2^2 \cdot \text{var}(\hat{A}) + \text{var}(\hat{B}) + 2 \text{cov}(\hat{A}, \hat{B}) + \text{var}(\epsilon). \quad (3.5)$$

1. Lors d'un test statistique, le niveau de signification, appelé *p-value* en anglais, est la probabilité de dépasser la valeur observée de la variable de décision sous l'hypothèse nulle. On teste ici l'hypothèse de nullité des coefficients par un test de Student (Saporta, 2006, chap. 14).

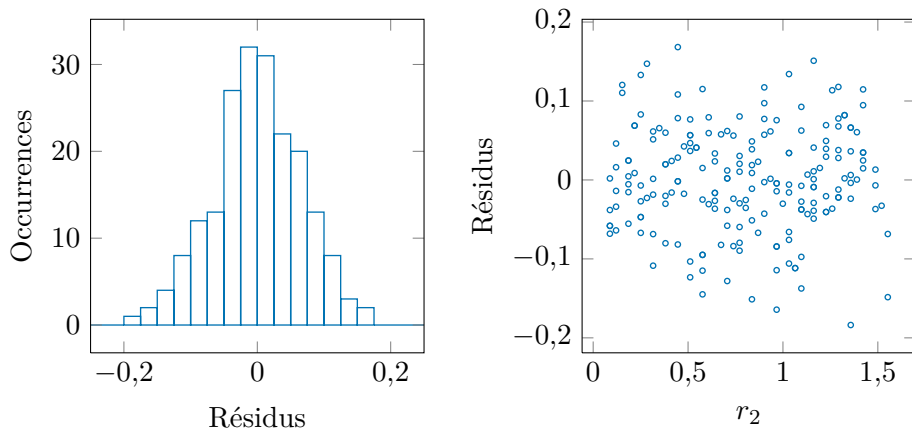


Figure 3.3. : Histogramme des résidus à gauche et nuage de points des résidus en fonction du prédicteur, r_2 , à droite.

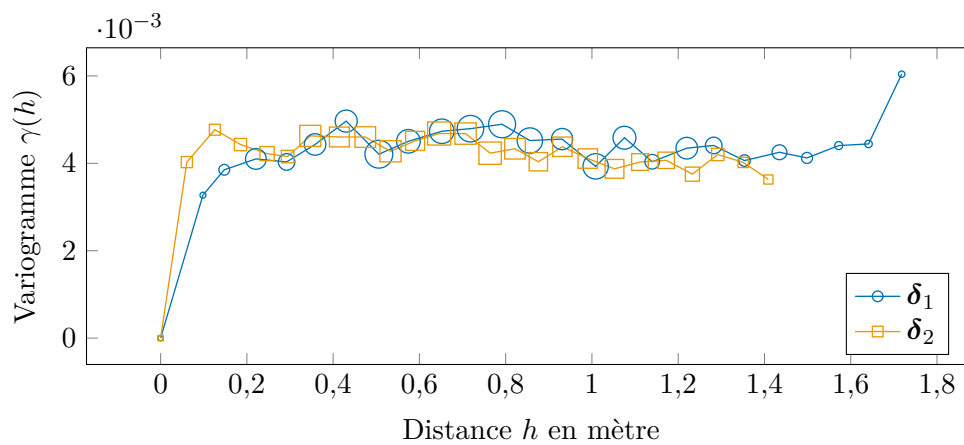


Figure 3.4. : Variogramme empirique des résidus du modèle linéaire ajusté par la méthode des moindres carrés aux données représentées par la figure 3.1.

$\hat{\sigma}_\epsilon^2$	$\text{var}(\hat{A})/\text{m}^{-2}$	$\text{var}(\hat{B})$	$\text{cov}(\hat{A}, \hat{B})/\text{m}^{-1}$
$4,41 \cdot 10^{-3}$	$4,21 \cdot 10^{-7}$	$6,24 \cdot 10^{-10}$	$-1,42 \cdot 10^{-8}$

Tableau 3.1. : Variance des coefficients du modèle linéaire ajusté sur les données d'examen télévisuel.

Sous hypothèse d'homoscédasticité, la variance du terme d'erreur, $\text{var}(\epsilon)$, peut être estimée sans biais par la formule suivante (Saporta, 2006, chap. 16) :

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N-2} \sum_{k=1}^N e_k^2, \quad (3.6)$$

où N désigne le nombre d'observations unitaires et (e_k) les résidus. Soit \mathbf{R} la matrice à N lignes dont la première colonne ne contient que des 1 et la seconde les coordonnées selon δ_2 . La matrice de covariance $\Sigma_{\hat{A}\hat{B}}$ de \hat{A} et \hat{B} peut être estimée par la formule suivante :

$$\Sigma_{\hat{A}\hat{B}} = \hat{\sigma}_\epsilon^2 (\mathbf{R}'\mathbf{R})^{-1}. \quad (3.7)$$

Les équations (3.6) et (3.7) permettent de calculer la variance d'une prédiction ponctuelle du modèle à l'aide de l'équation (3.5). Les variances des coefficients du modèle sont données dans le tableau 3.1. On peut observer que la variance estimée avec la formule (3.5) dépend essentiellement de la variance du terme d'erreur et très peu de la position. En moyenne, pour un indice du tube selon δ_2 variant de 1 à sa valeur maximale possible, 46, l'écart type d'une estimation ponctuelle avec le modèle linéaire est de 0,066. En supposant la distribution de cet estimateur gaussienne, on peut estimer que l'incertitude sur l'estimation du taux de colmatage moyen pour un risque de 5 % est de l'ordre de $\pm 0,13$. Cette valeur est nettement supérieure à la valeur calculée en divisant l'écart type de l'échantillon par la racine carré du nombre d'individus mais aussi à la valeur de 0,05 choisie comme marge de sécurité. On ne peut généraliser ce résultat qu'en réitérant la démarche qui vient d'être exposée pour chaque échantillon. Cette étude de cas souligne néanmoins l'importance d'augmenter la densité d'échantillonnage, ce qui devrait être rendu possible prochainement par l'automatisation de la procédure d'analyse des images (Audibert et coll., à paraître).

3.1.2. Taux de colmatage uniforme équivalent

Afin de comparer plus facilement les diagnostics par examen télévisuel entre eux, les échantillons d'estimations unitaires sont rapportés à deux valeurs en calculant leur moyenne par jambe. Puisque le modèle de g. v. est monodimensionnel, cette opération est aussi nécessaire pour comparer ces diagnostics avec ceux obtenus par analyse de la réponse dynamique du n. g. l. La moyenne des taux de colmatage d'une demi-plaque correspond à la diminution de surface moyenne de cette plaque due au colmatage. Cependant l'effet du colmatage sur la thermohydraulique du g. v. n'est pas linéaire. Ainsi, une plaque pose beaucoup moins de problèmes si tous ses trous ont un taux de colmatage de 0,5 que si

une moitié ont des taux de 0,1 et l'autre de 0,9. Le taux de colmatage moyen n'est pas à même de représenter ces hétérogénéités. Dans cette thèse, une approche différente a été choisie pour rapporter à une dimension des répartitions bidimensionnelles de taux de colmatage. L'idée est de considérer la moyenne des effets plutôt que l'effet de la moyenne. On définit pour cela le *taux de colmatage uniforme équivalent*. Le taux de colmatage uniforme équivalent d'un ensemble de trous foliés est calculé en utilisant un modèle de perte de charge pour le passage du fluide au travers d'une plaque entretoise colmatée. Il correspond au taux de colmatage qu'auraient tous les trous s'ils étaient identiquement colmatés et causeraient la même perte de charge que l'ensemble considéré. Cette mesure permet de prendre en compte la non linéarité de la relation entre colmatage et perte de charge et les hétérogénéités du colmatage des plaques entretoises. Le modèle de perte de charge utilisé est dérivé d'une corrélation empirique modélisant le passage d'un fluide au travers d'une grille à bords arrondis (Idel'cik, 1969, p. 306 et diagramme 8.5). Ses paramètres ont été ajustés grâce à des mesures effectuées sur une maquette de plaque entretoise colmatée baptisée P2C (Pillet et coll., 2010). La figure 3.5 représente les valeurs prises par le modèle et les points de mesure utilisés pour le calibrer pour des taux de colmatage variant de 0 à 0,75. Des mesures ont été effectuées pour des taux plus importants mais le modèle ne permet pas de reproduire fidèlement ces observations. De fait, une plaque entretoise dont les trous sont obstrués à plus de 75 % ne ressemble plus vraiment à une grille. Par ailleurs, il est probable que le passage dans des orifices aussi étroits ait des conséquences sur l'écoulement plus complexes qu'une simple modification des pertes de charge singulières. Le régime d'écoulement, par exemple, est susceptible d'être modifié. Aussi, il a été décidé de ne couvrir avec le modèle que la gamme de taux de colmatage de 0 à 0,7. Cette restriction ne diminue pas l'intérêt de la méthode de diagnostic car la stratégie de maintenance actuelle prescrit que des opérations curatives soient mises en œuvre avant que de tels états de colmatage extrêmes ne soient atteints. Les équations de ce modèle sont détaillées dans la section 5.4. On peut remarquer que la convexité de la relation illustrée par la figure 3.5 implique que le taux de colmatage uniforme équivalent est toujours supérieur ou égale au taux moyen.

Lorsque les taux de colmatage d'une plaque sont très hétérogènes, la forte non linéarité de la relation illustrée par la figure 3.5 entre le taux de colmatage et la perte de charge résultante entraîne une différence significative entre le taux de colmatage moyen et le taux de colmatage uniforme équivalent. Le cas, évoqué dans la section 3.1.1, du g. v. n° 2 de la tranche n° 1 de Bravo examiné en mai 2011 en est un exemple. Le rapport d'examen (Carradec, 2011) conclut à un taux de colmatage moyen en jambe chaude de 0,15. Environ 85 % des trous inspectés étaient très peu colmatés et les clichés correspondant n'ont pas été analysés individuellement en détail. Une valeur de taux de colmatage conservatrice de 0,1 leur a été attribuée. La forme atypique du colmatage des 15 % restant des trous inspectés a empêché d'en faire une estimation précise. Il a été supposé que ce colmatage atypique avait un effet équivalent à un taux de colmatage de 0,45 environ. La valeur exacte n'est pas donnée puisqu'elle influe très peu sur la valeur du taux de colmatage moyen de la demi-plaque. Il en est tout autrement si l'on considère le taux de colmatage uniforme équivalent. La figure 3.6 représente les taux de colmatage uniformes équivalents calculés en supposant un taux de colmatage de 0,1 pour les trous faiblement colmatés

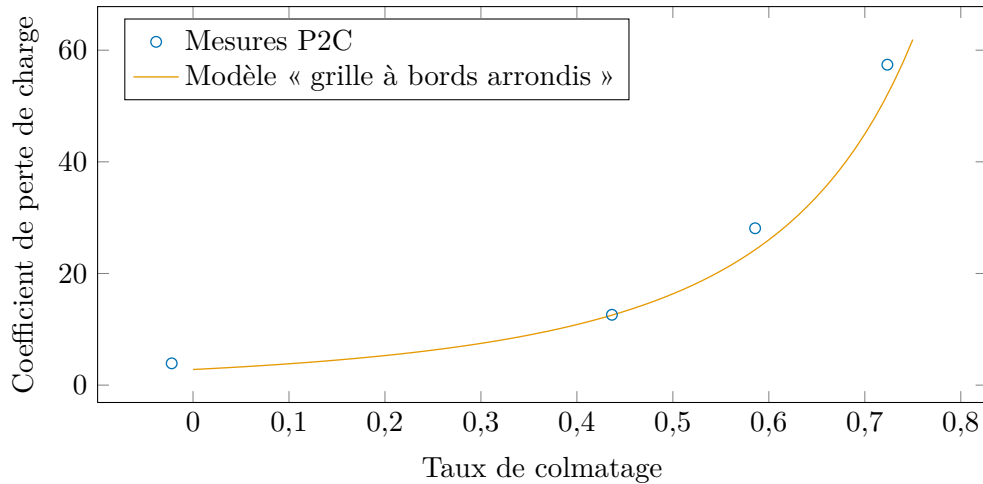


Figure 3.5. : Modèle de perte de charge utilisé et mesures de perte de charge issues des expériences sur la maquette P2C. Le point de donnée en bas à gauche du graphique correspond à une valeur négative de taux de colmatage. Ceci est dû à la précision limitée de la technique d’usinage employée : les trous percés dans cette réplique de plaque entretoise avaient une aire légèrement supérieure à celle de trous foliés réels. Ce n’est cependant pas gênant pour ajuster le modèle.

et en faisant plusieurs hypothèses sur la proportion et le taux de colmatage des trous fortement colmatés. Le taux de colmatage des trous fortement colmatés varie, en abscisse, de 0,4 à 0,65 et chaque courbe correspond à une proportion, de 5 % pour la courbe bleue (—•—) à 30 % pour la courbe vermillon (—▲—). Seule une proportion de 5 % de trous fortement colmatés aboutit à une valeur inférieure ou égale à celle retenue. La valeur des taux de colmatage des trous faiblement colmatés influe très peu sur le taux de colmatage uniforme équivalent. En revanche, la proportion, mal connue du fait de l’échantillonnage restreint, et la valeur du taux fort, difficile à estimer du fait de la forme atypique des dépôts, sont toutes deux déterminantes.

La notion de taux de colmatage uniforme équivalent commence à être utilisée en parallèle du taux de colmatage moyen, comme en témoigne le rapport de stage de Mangin et Goetgheluck (à paraître). Les valeurs de taux moyens et de taux de colmatage uniformes équivalents calculées à partir des examens télévisuels réalisés sur les g. v. étudiés dans cette thèse sont répertoriées dans le tableau 13.2 page 209. Les calculs ont été effectués à partir des mêmes données brutes que celles utilisées par (Mangin et Goetgheluck, à paraître) mais le modèle de perte de charge utilisé est différent et les taux de colmatage n’ont pas été plafonnés.

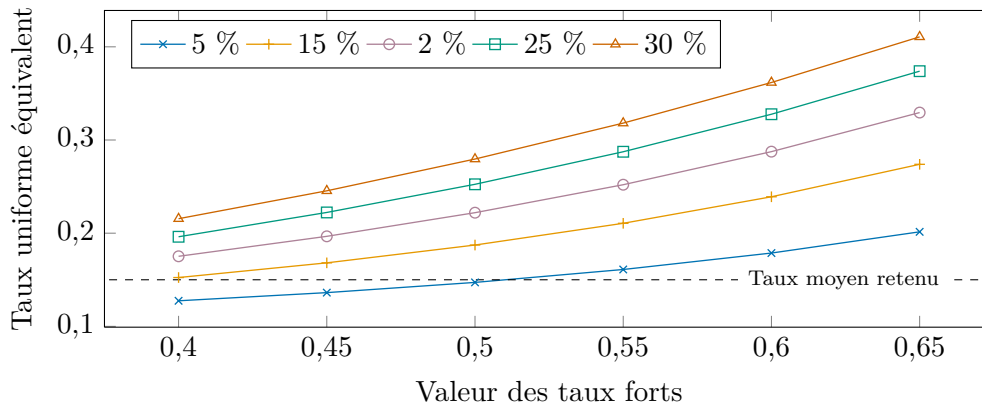


Figure 3.6. : Taux de colmatage uniforme équivalent calculé pour la tranche n° 1 de la centrale Bravo en mai 2011, pour différentes hypothèses sur la proportion de trous fortement colmatés et la valeur de leur taux de colmatage.

3.2. Méthode par courants de Foucault

L'intégrité des tubes des g. v. est contrôlée périodiquement lors des arrêts de tranche à l'aide d'une sonde à courants de Foucault (Pitner et coll., 2011). Ces sondes sont constituées d'une bobine que l'on fait circuler en aller-retour dans les tubes en U . La bobine est alimentée par un courant alternatif et émet un champ magnétique lui aussi alternatif. Lorsqu'un objet conducteur est balayé par le champ magnétique émis par la sonde, il est le siège de courants électriques dits de Foucault (Cheong et coll., 2011). Ces courants génèrent eux-mêmes un champ magnétique qui modifie l'impédance de la bobine. Les signaux d'impédance sont analysés pour détecter des défauts dans les tubes, notamment des fissurations naissantes (Risson et Garbay, 2009).

Les plaques entretoises étant conductrices, leur traversée, à l'intérieur des tubes, produit un signal prononcé. Les passages des bords inférieur et supérieur de la plaque produisent des pics de signes opposés. En l'absence de colmatage, ces deux pics sont d'amplitude comparable. En revanche, lorsque les trous foliés autour du tube sondé sont colmatés, la morphologie du bord inférieur est modifiée et l'amplitude des pics est différente. Les examens télévisuels indiquent en effet qu'en général le colmatage touche la partie basse des trous foliés et qu'il engendre ainsi une asymétrie des deux bords. L'analyse de ces perturbations a conduit à définir un indicateur du colmatage baptisé *ratio sonde axiale (SAX)*. Le ratio SAX est le rapport de la différence d'amplitude des pics et de l'amplitude du plus grand pic (Devinck, 2009). La relation entre le ratio SAX et le taux de colmatage n'a pas été identifiée mais une corrélation positive avec les estimations par examens télévisuels permet de supposer qu'elle est croissante. En extrapolant aux plaques basses la validation par les examens télévisuels des évaluations effectuées en plaque supérieure, la méthode de diagnostic du colmatage par courants de Foucault permet d'estimer la répartition verticale du colmatage. L'incertitude de cette méthode n'a pas fait l'objet d'une quantification précise. L'une des sources principales d'erreur est que la perturbation

du signal dépend de la morphologie du dépôt. Par exemple, des dépôts d'encrassement en amont des trous foliés peuvent fausser l'estimation du colmatage. Une classification des dépôts selon quatre morphologies types a été établie, et des formes de signaux correspondantes ont été identifiées (Barbe et Dumay, 2012). La composition du dépôt ou encore sa densité peuvent aussi influencer fortement le signal. L'effet de ces paramètres a pu être évalué grâce à des mesures effectuées sur des maquettes. Enfin, le signal sature au-delà d'une quantité de dépôt correspondant à un colmatage de l'ordre de 0,5.

Un nouvel indicateur du colmatage fondé sur l'analyse des signaux des sondes à courant de Foucault est en cours de développement (Paul, à paraître). Le lien entre le colmatage et cet indicateur devrait être plus explicite que pour le ratio SAX.

3.3. Suivi du niveau gamme large en régime stationnaire à pleine puissance

Le n. g. l. est une mesure proportionnelle à la différence de pression entre le haut et le bas du *downcomer*. Les deux points de piquage sont indiqués sur la figure 2.2 page 21. La différence de pression entre les deux piquages est due au poids de la colonne d'eau et aux pertes de charge dans le *downcomer* qui varient proportionnellement au carré du débit. En régime stationnaire, tous les débits dans le g. v. sont constants et le débit de vapeur sortant est égal au débit d'eau alimentaire. Le débit (ascendant) dans le *riser* et le débit (descendant) dans le *downcomer* constituent la boucle de recirculation. Le débit d'eau recirculée par les dispositifs de séparation et de séchage correspond à la différence entre le débit sortant (ou entrant) et le débit dans le *riser* (ou le *downcomer*). Le passage au travers des trous foliés des plaques entretoises est responsable de la majeure partie de la résistance hydraulique dans le *riser*. Cette résistance hydraulique augmente à mesure que les trous foliés se colmatent car l'aire de leur section de passage diminue. Cela se traduit par une diminution du débit dans la boucle de recirculation et donc des pertes de charges dans le *downcomer*, soit une augmentation du n. g. l.

Ainsi, la mesure de n. g. l. est impactée par le phénomène de colmatage : on observe en général une augmentation du n. g. l. au fil du temps et une diminution brutale après nettoyage chimique (Crinon, 2009a). Le capteur de n. g. l. est un capteur d'exploitation dont les données peuvent être récupérées à une période de 2 secondes avec l'application ORLI. Un programme d'essais trimestriels a été mis en place par EDF au cours de l'année 2009 pour effectuer un suivi minutieux de cette mesure. Le protocole de cet essai, un calcul d'incertitude de la mesure de n. g. l. et des valeurs seuil sont donnés par Crinon (2009b). Actuellement, cet indicateur est l'un des principaux critères de décision pour la planification des examens télévisuels et des nettoyages de g. v.

Si le n. g. l. en régime stationnaire à pleine puissance permet un suivi tendanciel de l'évolution du colmatage, cette mesure n'est cependant pas suffisamment performante pour permettre un diagnostic absolu. La précision de la mesure de n. g. l. est suffisante pour son usage prévu de régulation, mais un indicateur absolu du colmatage issu de cette mesure comporterait des biais importants. En effet, il arrive d'observer des sauts importants de la mesure de n. g. l. consécutifs, par exemple, à des opérations de maintenance des

capteurs. La détermination des sources de ces biais et la recherche de méthodes pour les prévenir ou les corriger font actuellement l'objet d'un groupe de travail à EDF (Bertrand et Crinon, 2011). Le protocole d'essais (Crinon, 2009a) requiert que les relevés soient faits lorsque « le niveau de la puissance thermique primaire [est] au plus proche de la puissance nominale » et fixe comme limite basse 92 % de la puissance nominale. Cette précaution vaut surtout pour la mesure de pression dans le dôme du g. v. qui est particulièrement sensible au niveau de puissance tandis que le n. g. l. varie assez peu à haute charge. Aussi, du fait de l'incertitude importante mentionnée précédemment, il n'est pas jugé intéressant de corriger la mesure de n. g. l. en fonction du point de fonctionnement (Crinon, 2009a).

3.4. Analyse de la réponse dynamique du niveau gamme large

En modifiant les pertes de charge singulières des plaques entretoises, le colmatage altère non seulement la thermohydraulique des g. v. en régime stationnaire mais aussi leur réponse en régime dynamique. Ce constat est à l'origine du développement, à partir de 2008, d'une nouvelle méthode de diagnostic du colmatage appelée *méthode du n. g. l. dynamique* (Chip et coll., 2008). En analysant des courbes de réponse de n. g. l., on peut espérer pouvoir se débarrasser des biais qui affectent la mesure de n. g. l. en régime stationnaire. La contrepartie de cet avantage est la difficulté plus importante que dans le cas stationnaire de faire le lien entre les données et le colmatage. Pour le suivi en régime stationnaire, la relation entre n. g. l. et colmatage n'a pas été explicitée. Cependant, on peut raisonnablement supposer qu'elle est monotone. Cette hypothèse peut être suffisante pour un suivi tendanciel, d'autant que l'on dispose d'un historique de données très important qui permet de comparer les évolutions des différents g. v. et d'effectuer des déductions à partir des quelques connaissances que l'on a sur leurs états de colmatage passés. Ces derniers sont obtenues par exemple par examen télévisuel ou en comparant les valeurs de n. g. l. avant et après nettoyage. Dès lors que l'on s'intéresse à des courbes de réponse, il devient nécessaire de modéliser le lien entre colmatage et n. g. l. Un modèle monodimensionnel de g. v. permettant de simuler des régimes transitoires a été élaboré spécifiquement pour l'étude du colmatage (Chip et coll., 2008). Ce modèle, ainsi que celui dont il s'inspire et les évolutions ultérieures qui lui ont été apportées sont décrits dans la partie II. Le principe général de la méthode du n. g. l. dynamique est de simuler la réponse du n. g. l. lors d'un transitoire de puissance pour différents états de colmatage et de comparer les courbes obtenues avec des mesures enregistrées par les capteurs d'exploitation lors de transitoires réels. Une fois le modèle disponible, il reste à choisir les transitoires à étudier.

3.4.1. Un transitoire bien commode : l'essai périodique R.G.L. 4

Le choix du type de transitoire utilisé pour le diagnostic dépend essentiellement de quatre critères. Tout d'abord, le transitoire doit pouvoir être modélisé. Il ne doit pas être trop complexe, ni comporter de variations trop brutales qui engendreraient des complications d'ordre numérique. Cependant, la réponse du n. g. l. au transitoire doit être sensible à l'effet du colmatage. Pour que l'approche dynamique ait un intérêt,

les dérivés temporelles des champs thermohydrauliques doivent être importantes. Un transitoire trop « doux » n’apporterait pas d’information de plus que l’étude du régime stationnaire. Par ailleurs, un suivi précis de la cinétique de colmatage n’est possible que si la fréquence de diagnostic est suffisamment élevée, ce qui permet par ailleurs d’en réduire l’incertitude de nature aléatoire. Enfin, le développement de la méthode de diagnostic est grandement facilité si tous les transitoires analysés sont identiques, ou au moins semblables. Cela permet en effet de se focaliser sur l’effet du colmatage lors d’un transitoire particulier plutôt que d’essayer de caractériser de façon générale l’effet du colmatage sur le comportement dynamique du g. v.

La conception d’un transitoire *ad hoc* n’est pas à l’ordre du jour car elle annihilerait l’un des intérêts principaux de l’approche du diagnostic par l’étude du n. g. l., à savoir son coût de mise en œuvre nul. Il se trouve que l’essai *périodique de calibrage des grappes grises (R.G.L. 4)* engendre un transitoire de puissance remplissant remarquablement bien les critères formulés. L’objectif de cet essai est d’actualiser la courbe de calibrage des groupes de grappes de contrôle peu absorbantes, dites « grises », utilisées pour compenser les variations instantanées de réactivité du cœur (Pigeon, 1997). Il s’agit de dispositifs absorbeurs de neutrons pouvant être introduits progressivement dans le cœur du réacteur afin de limiter la puissance de la réaction de fission transmise au fluide primaire. La compensation de puissance correspondant à un niveau d’insertion donné dépend de l’épuisement du cœur. Cette correspondance est représentée par la courbe de calibrage qui est donc mise à jour plusieurs fois par cycle de fonctionnement. Un *cycle de fonctionnement* est la période, d’une douzaine de mois environ, durant laquelle une tranche est exploitée en continu. Une partie du combustible est remplacée lors des *arrêt de tranches*, entre les cycles. Ceux-ci sont par ailleurs mis à profit pour réaliser les opérations de maintenance nécessitant l’arrêt de la tranche, par exemple les nettoyages chimiques des g. v. évoqués dans la section 2.2.3. La prolongation de la durée des arrêts de tranches est l’une des raisons principales du coût très élevé de ces opérations de maintenance. Ainsi, l’essai R.G.L. 4 est réalisé régulièrement, tous les 2 à 3 mois ou 2000 MW j t^{-1} d’épuisement spécifique² soit en moyenne 3 à 4 fois par cycle (Ghattas et Blanc, 1987).

Cet essai présente de multiples avantages pour l’étude du colmatage. Outre sa périodicité adéquate, le transitoire de puissance résultant est rapide et d’amplitude importante, ce qui provoque une réponse dynamique du champ de pression conséquente au regard du bruit de la mesure de n. g. l. Il est néanmoins simple à modéliser car très proche d’une variation linéaire de pente raisonnable. Enfin, les conditions dans lesquelles il est réalisé sont soumises à des règles assez strictes (Pigeon, 1997) si bien que les variations imposées aux conditions aux limites des g. v. sont assez similaires d’un essai à l’autre. L’adaptation de la courbe de calibrage est réalisée sur une diminution de puissance de 100 % à 50 % de la puissance nominale à une vitesse d’environ 3 % par minute (Ghattas et Blanc, 1987). On reviendra sur l’impact des variations des conditions de réalisation des essais R.G.L. 4 dans le chapitre 12.

2. L’épuisement spécifique, aussi appelé *combustion massique*, est l’énergie thermique produite par le cœur, rapportée à la masse de combustible initiale introduite, comptée en ne tenant compte que des noyaux lourds. Il est mesuré usuellement en mégawatt jour par tonne (Reuss, 2003).

L'état de colmatage d'un g. v. ne peut être que partiellement connu à l'aide des méthodes par examen télévisuel ou sonde à courants de Foucault mais on peut raisonnablement supposer que le colmatage augmente monotonement et qu'un g. v. après nettoyage est faiblement colmaté. La figure 3.7 représente les courbes de réponse du n. g. l. enregistrées par les capteurs du g. v. n° 3 de la tranche n° 2 de Charlie lors des essais R.G.L. 4 réalisés de janvier 2002 à janvier 2012. Ce g. v. a subi un nettoyage chimique en 2007. Les examens télévisuels et les estimations par courants de Foucault réalisées avant et après le nettoyage chimique de 2007 indiquent que la tranche est passée d'un taux de colmatage en plaque supérieure de l'ordre de 0,70 à environ 0,15 (de Bouvier, 2007). Les courbes ont été lissées pour en améliorer la lisibilité et leurs moyennes temporelles soustraites afin de faciliter leur comparaison. La méthode de lissage utilisée est le modèle de régression locale polynomial d'ordre 2 de l'algorithme *loess* (MathWorks, 2012) avec une fenêtre de 100s. Les courbes correspondant aux essais les plus anciens sont tracées en vert. Leur début et fin sont indiqués par la lettre *A*. Les courbes suivantes virent progressivement du vert à l'orange jusqu'à la situation juste avant le nettoyage chimique ayant eu lieu en 2007 indiqué par la lettre *B*. La transition marquée par des flèches oranges (\nearrow) de l'état initial (*A*) à l'état avant nettoyage (*B*) se traduit par une rotation des courbes autour d'un point fixe dont l'abscisse est légèrement au-delà de 600s. Cet effet est observé pour tous les g. v. de type 51 B et apparentés et l'on peut conclure qu'il est caractéristique du colmatage. De fait, le nettoyage chimique a un effet inverse très net, indiqué par les flèches bleues (\nwarrow). À partir de l'état après nettoyage (*C*), on observe à nouveau le pivotement caractéristique du colmatage. Les courbes de réponses virent du bleu ciel au mauve et leur évolution, indiquée par des flèches mauves (\searrow), s'oriente vers l'état de colmatage initial. Le fait que le nettoyage chimique soit aussi clairement visible est de bon augure concernant la quantité d'information sur le colmatage que l'on peut extraire de la dynamique du n. g. l.

3.4.2. Méthode à profil choisi

Cette section décrit la méthode de diagnostic par analyse de la réponse dynamique du n. g. l. développée avant le début de cette thèse. Cette méthode sera appelée par la suite la *méthode de diagnostic à profil choisi*. La première version de cette méthode a été mise au point par Chip et coll. (2008) pour les générateurs de vapeur de type 51 B. Elle a été ensuite appliquée aux données disponibles et étendue aux autres types de g. v. par Gay et coll. (2009). Un premier bilan de l'application de cette méthode a été établi par Midou et coll. (2010a) et Midou et coll. (2010b). L'applicabilité de la méthode aux centrales n'effectuant pas de transitoires similaires à l'essai R.G.L. 4 a été étudiée par Ninet et Favennec (2010); Ninet et coll. (2010). La méthode a enfin été adaptée à la nouvelle version du modèle développée pendant la thèse et certains éléments de la nouvelle méthode décrite dans le chapitre 13 lui ont été intégrés (Ninet et coll., 2012). Cette dernière version est actuellement en cours d'industrialisation.

La méthode de diagnostic à profil choisi repose sur le choix d'une famille de configurations de colmatage jugée représentative. La réponse du n. g. l. lors d'un essai R.G.L. 4 est simulée pour chacune de ces configurations de colmatage. On obtient ainsi une famille de

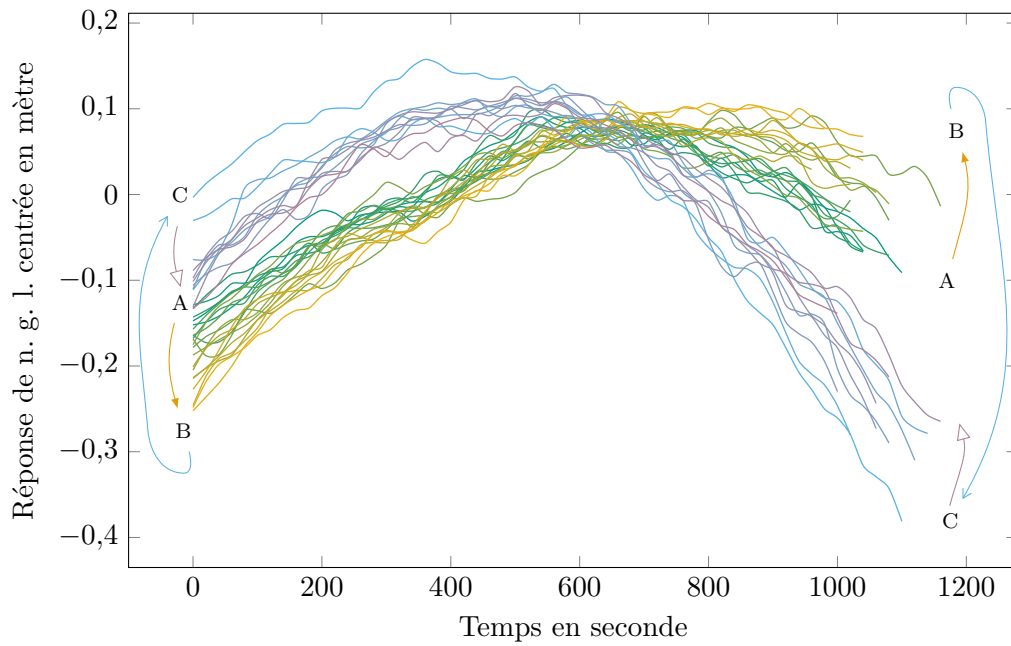


Figure 3.7. : Courbes de réponses de n. g. l. lors d'essais R.G.L. 4 mesurées sur le g. v. n° 3 de la tranche n° 2 de Charlie. Les mesures de pressions des capteurs de n. g. l. des g. v. réels sont converties par une fonction affine en une mesure de niveau en mètre, d'où l'unité indiquée en ordonnée. Ce point est traité plus en détail dans les sections 6.3 et 13.2.

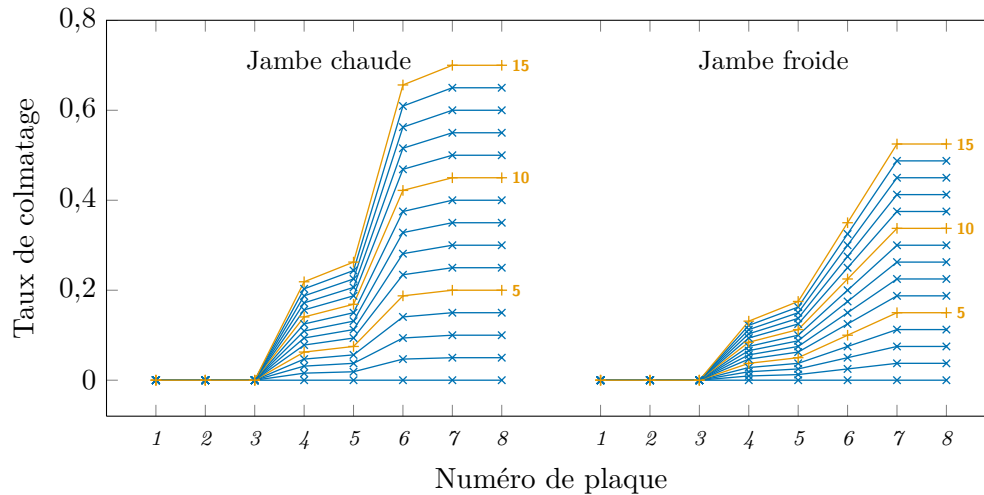


Figure 3.8. : Quinze configurations de colmatage proportionnelles au profil établi par Bernier (2007) utilisées par la méthode de diagnostic à profil choisi. Le taux de colmatage en plaque supérieure côté chaud augmente par pas de 0,05. Le rapport des taux en jambes froide et chaude vaut 0,75.

courbes de réponses simulées. Ces courbes sont ensuite comparées à une courbe enregistrée par les capteurs du g. v. à diagnostiquer. Le diagnostic du colmatage est la configuration testée dont la courbe de réponse simulée est la plus semblable à la courbe de réponse mesurée.

Choix d'un profil de colmatage

Dans la méthode de diagnostic à profil choisi, les configurations de colmatage testées ont toutes été choisies proportionnelles à un profil de colmatage donné. Ceci permet d'ordonner naturellement ces configurations et de les résumer chacune par une valeur unique, par exemple, le taux de colmatage de la demi-plaque supérieure en jambe chaude. La figure 3.8 représente 15 configurations de colmatage utilisées dans la méthode de diagnostic à profil choisi. Le profil dont elles sont issues résulte d'estimation de taux de colmatage par courant de Foucault effectuées pour la tranche n° 4 de Delta (Bernier, 2007). La valeur choisie pour désigner chacune des configurations générées à partir de ce profil est le taux de colmatage en plaque supérieure en jambe chaude. Il s'agit du taux de colmatage le plus important de ce profil et cela permet, si l'on est suffisamment confiant dans le profil choisi, de comparer les résultats avec les examens télévisuels.

Méthode d'analyse des courbes simulées et mesurées

Deux méthodes d'analyse des courbes de réponse avaient été proposées par Chip et coll. (2008). La première consiste à comparer les temps d'atteinte du maximum ; la seconde à minimiser la distance euclidienne des vecteurs représentant les réponses mesurées et

simulées. La deuxième méthode de comparaison a été retenue par la suite car jugée plus robuste.

Dans les deux cas, les données simulées et mesurées subissent un pré-traitement. Les courbes de réponse sont représentées par des vecteurs contenant les valeurs du n. g. l. enregistrées aux instants d'échantillonnage successifs. Les vecteurs simulés et mesurés doivent être de même dimension pour que l'on puisse calculer une distance entre eux. Or, la durée des essais R.G.L. 4 est variable. Ainsi, le premier pré-traitement consiste à adapter la durée du transitoire réel au transitoire simulé par interpolation des données. En effet, on cherche autant que possible à utiliser le même ensemble de simulations pour tous les diagnostics. Ce point sera abordé plus en détail dans le chapitre 12.

Afin de masquer l'effet du colmatage en régime stationnaire sur la valeur du n. g. l. et de se débarrasser des biais associés, la moyenne temporelle des courbes leur est soustraite. On verra dans le chapitre 6 que l'amplitude de variation du n. g. l. est difficile à modéliser de façon totalement fiable. Il a été choisi de contourner cette difficulté en divisant les courbes de réponse centrées par leurs amplitudes respectives. Ce second traitement était motivé par l'hypothèse selon laquelle ce sont les dérivées d'ordre supérieur à 2 du signal qui portent l'information sur le colmatage (Chip et coll., 2008). La division par l'amplitude est en quelque sorte un moyen de filtrer la dérivée d'ordre 1 en divisant par une pente globale. Cette hypothèse sera remise en cause dans la section 10.3.

3.5. Extension de la méthode à profil choisi par filtrage particulière

Le choix d'un profil de colmatage est une hypothèse forte qui entraîne un diagnostic erroné lorsqu'elle n'est pas vérifiée. L'espace de représentation du colmatage est à 16 dimensions, une par demi-plaque, et est donc très vaste. Augmenter simplement la taille de la famille de configurations de colmatage testées n'est pas une solution efficace car même un échantillon de plusieurs dizaines de milliers d'individus couvrira cet espace de façon clairsemée. De plus, on perdrait ainsi l'avantage d'une famille ordonnée de configurations que l'on peut aisément résumer par une valeur unique.

Pour faire face à cette problématique, une transposition du problème dans un cadre stochastique a été proposée par Chip et Midou (2009). L'idée sous-jacente est que le signal dynamique du n. g. l. pourrait permettre d'identifier partiellement la répartition spatiale du colmatage. Plus précisément, on cherche à se faire une idée de la distribution de probabilité des taux de colmatage par demi-plaque conditionnellement aux connaissances que l'on a de l'état du g. v., assimilées ici à la réponse du n. g. l. pendant l'essai R.G.L. 4. Un algorithme de *filtrage particulière* a été développé par Chip et Midou (2009) pour tenter d'échantillonner l'espace de colmatage de manière efficace. Le filtrage particulière est une méthode de Monte-Carlo itérative qui repose sur un *échantillonnage d'importance* adaptatif (Cappé et coll., 2007). Cette méthode relève en général de l'assimilation de données dont l'objet est de corriger les paramètres d'un modèle au fur et à mesure que de nouvelles données sont disponibles. Elle est utilisée ici dans le cadre un peu différent de la résolution d'un problème inverse.

3.5.1. Principe du filtrage particulaire

Notons p_m la densité de probabilité du modèle et p_d celle des observations, \mathbf{X} l'état du modèle, c'est-à-dire la représentation du colmatage, et \mathbf{y} les observations, c'est-à-dire la réponse de n. g. l. mesurée. La formule de Bayes (Saporta, 2006, chap. 1) s'écrit

$$p_m(\mathbf{X}|\mathbf{y}) = \frac{p_d(\mathbf{y}|\mathbf{X}) p_m(\mathbf{X})}{p_d(\mathbf{y})}. \quad (3.8)$$

Dans le domaine de l'assimilation de données, la densité recherchée, $p_m(\mathbf{X}|\mathbf{y})$, est appelée densité de la distribution *a posteriori*. Elle est proportionnelle au produit de la *vraisemblance*, $p_d(\mathbf{y}|\mathbf{X})$, et de la densité de la distribution *a priori*, $p_m(\mathbf{X})$. Dans la suite, les indices d et m seront omis : $p(\cdot)$ désigne la forme fonctionnelle des densités et les arguments permettent d'éviter les ambiguïtés.

La densité de probabilité des observations est inconnue mais l'on peut reformuler le dénominateur de l'équation (3.8) de la façon suivante :

$$p(\mathbf{y}) = \int p(\mathbf{y}, \mathbf{X}) d\mathbf{X} = \int p(\mathbf{y}|\mathbf{X}) p(\mathbf{X}) d\mathbf{X}. \quad (3.9)$$

Cette expression intégrale ne peut en générale pas être calculée par des méthodes d'intégration numériques déterministes, surtout lorsque la dimension est élevée.

Le principe des méthodes de Monte-Carlo est d'approximer une densité de probabilité à l'aide d'un échantillon de réalisations indépendantes de sa distribution. Pour le calcul d'une espérance, par exemple, l'intégrale de la densité sera remplacée par la moyenne de l'échantillon. La variance de cette estimation est inversement proportionnelle au nombre d'individus et ce quelque soit la dimension de l'intégrale, tandis que les méthodes d'intégration déterministes sont inefficaces en grande dimension (Robert et Casella, 2004, p. 21). Plus généralement, on représente une densité $p(\mathbf{X})$ par une moyenne de distributions de Dirac. Les individus de l'échantillon sont appelés *particules* dans le domaine du filtrage particulaire. En notant \mathbf{X}_i la $i^{\text{ème}}$ particule d'un échantillon de taille N , cette approximation s'écrit

$$p(\mathbf{X}) \approx \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{X} - \mathbf{X}_i). \quad (3.10)$$

L'application directe de ce principe à la formule de Bayes donne lieu à ce que l'on appelle l'*échantillonnage d'importance*.

Échantillonnage d'importance

En appliquant l'approximation (3.10) à la formule (3.8) et en choisissant une distribution a priori, $p(\mathbf{X})$ uniforme, le principe de l'échantillonnage d'importance se traduit par l'équation suivante :

$$p(\mathbf{X}|\mathbf{y}) \approx \sum_{i=1}^N w_i \delta(\mathbf{X} - \mathbf{X}_i), \quad (3.11)$$

où les poids w_i ont pour expression :

$$w_i = \frac{p(\mathbf{y}|\mathbf{X}_i)}{\sum_{j=1}^N p(\mathbf{y}|\mathbf{X}_j)}. \quad (3.12)$$

L'utilisation de la formule de Bayes a introduit des poids différents de $\frac{1}{N}$ dans l'approximation (3.11). En choisissant arbitrairement une distribution a priori, l'échantillonnage d'importance permet d'estimer la densité a posteriori à l'aide d'une moyenne pondérée de distributions de Dirac. Dans notre cas, les poids sont calculés à partir de la comparaison des courbes de réponse mesurées et simulées. Soient \mathbf{m} et \mathbf{s}_i les vecteurs correspondant respectivement à l'échantillonnage de la réponse mesurée et simulée. On note par ailleurs d_i la distance euclidienne entre ces deux vecteurs auxquels on a préalablement appliqué le traitement mentionné dans le paragraphe 3.4.2, soit :

$$d_i = \left\| \frac{\mathbf{m} - \bar{\mathbf{m}}}{\max(\mathbf{m}) - \min(\mathbf{m})} - \frac{\mathbf{s}_i - \bar{\mathbf{s}}_i}{\max(\mathbf{s}_i) - \min(\mathbf{s}_i)} \right\|_2, \quad (3.13)$$

où la barre au dessus d'un vecteur indique la moyenne de ses composantes, c'est-à-dire la moyenne temporelle de la réponse. La vraisemblance d'une particule est estimée par la fonction suivante, utilisée pour calibrer heuristiquement la sélectivité du filtre :

$$\hat{v}_i \propto \exp(-5 d_i^2). \quad (3.14)$$

Cette expression peut être interprétée comme une erreur de mesure gaussienne de variance égale à $\frac{1}{10}$. En choisissant comme distribution a priori une distribution uniforme, les poids se déduisent ensuite simplement en divisant la vraisemblance estimée de chaque particule par la somme des vraisemblances des particules de l'ensemble :

$$\hat{w}_i = \frac{\hat{v}_i}{\sum_{j=1}^N \hat{v}_j}. \quad (3.15)$$

Faute de connaissance sur la répartition spatiale du colmatage, la distribution a priori a été choisie uniforme. Ainsi, il est probable que de nombreuses particules auront des poids proches de zéro et modifieront très peu l'approximation. Elles correspondent aux états du modèle produisant des réponses très différentes de la sortie. À l'inverse, les quelques particules dont les poids seront importants détermineront fortement l'approximation. Autrement dit, on n'obtient d'information sur la configuration spatiale réelle du colmatage qu'en testant des configurations qui lui sont proches. L'objectif de l'algorithme de filtrage particulaire est de circonscrire progressivement les régions de l'espace d'état contenant de telles configurations et de les échantillonner plus densément. Il s'agit en quelque sorte d'un moyen d'optimiser l'échantillonnage d'un espace trop vaste pour être exploré exhaustivement en évitant de gaspiller des simulations dans des régions peu vraisemblables.

Échantillonnage d'importance adaptatif

On a vu qu'un échantillon de particules éparpillées dans l'espace d'état permet d'estimer la densité de probabilité du modèle. C'est le principe des méthodes de Monte-Carlo classiques. Elles sont cependant peu efficaces lorsque l'espace d'état est de grande dimension. Une autre possibilité consiste à générer d'abord un premier échantillon de quelques particules et de simuler les états correspondants avec le modèle, puis de déplacer chacune de ces particules selon une règle à définir et de simuler les nouveaux états correspondants. Si l'on peut réaliser un total de N évaluations du modèle et que l'on a généré n particules lors de la première étape, on pourra réaliser $k = N/n$ déplacements. Sous certaines conditions, la règle de déplacement est appelée un *noyau de transition* et cette méthode permet une exploration adéquate de l'espace d'état (Robert et Casella, 2004, chap. 6). L'avantage de cette approche est de permettre d'adapter la distribution a priori, à mesure que l'on obtient de l'information à l'aide des estimations de vraisemblance.

L'une des façons de mettre en œuvre cette idée est d'effectuer périodiquement un rééchantillonnage de l'ensemble des particules. Les particules de poids faible explorent des régions de l'espace d'état peu vraisemblables et n'apportent que peu d'information ; on cherchera donc à les éliminer. À l'inverse, les particules de poids fort seront dupliquées afin d'explorer plus efficacement les régions de forte vraisemblance. En pratique, l'étape de rééchantillonnage consiste en n tirages avec remise dans l'ensemble des n particules affectées chacune d'une probabilité égale à son poids estimé par l'équation (3.15). Le rééchantillonnage n'a d'intérêt que si les poids des particules de l'ensemble sont dissemblables. En effet, si toutes les particules sont de vraisemblances équivalentes, le rééchantillonnage conduira à abandonner l'exploration de certaines régions de l'espace d'état potentiellement intéressantes. À l'inverse, si on laisse l'algorithme évoluer trop longtemps sans rééchantillonner, les particules de l'ensemble risquent de s'éparpiller dans des zones de faible vraisemblance et le rééchantillonnage ne sera plus efficace. L'hétérogénéité de l'ensemble peut être estimée par la fonction suivante appelé *entropie de Shannon* :

$$E(\hat{w}_1, \dots, \hat{w}_n) = - \sum_{i=1}^n \hat{w}_i \text{Log}(\hat{w}_i). \quad (3.16)$$

L'entropie est maximale, égale à $\text{Log}(n)$, lorsque toutes les particules de l'ensemble ont la même vraisemblance ; elle est nulle lorsqu'une seule particule a une vraisemblance non nulle. Un critère de rééchantillonnage peut être défini par une valeur limite d'entropie, E_{lim} , en dessous de laquelle le rééchantillonnage a lieu. Cette limite avait été fixée à $0,3 \text{Log}(n)$ par Chip et Midou (2009).

Algorithme

L'algorithme de filtrage particulaire doit être initialisé avec un ensemble de particules arbitraire. La méthode retenue par Chip et coll. (2008) consiste à générer une famille initiale de configurations de colmatage quadrillant régulièrement un sous-espace choisi de

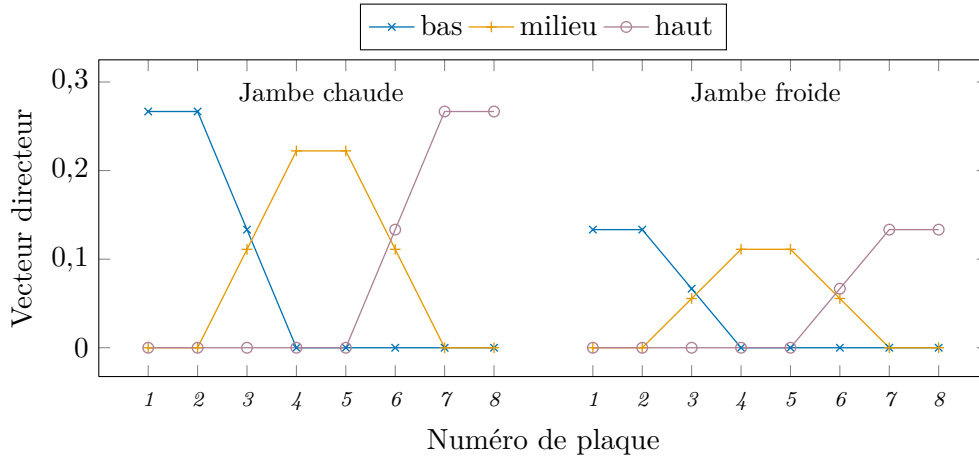


Figure 3.9. : Base du sous-espace d'initialisation de l'algorithme de filtrage particulaire.

l'espace d'état et à utiliser comme ensemble de départ les n configurations de cette famille dont les vraisemblances estimées sont les plus élevées. La figure 3.9 représente la base du sous-espace en question qui est de dimension 3. Les vecteurs directeurs représentés dans la figure 3.9 représentent les parties basses, médianes et haute du g. v. Dans ce sous-espace, le taux de colmatage de la jambe froide est égal à la moitié de celui de la jambe chaude. Ce choix est fondé sur les estimations par examen télévisuel et sonde à courant de Foucault suggérant que le colmatage en jambe froide est en général moins important (Bernier, 2007). On peut remarquer que cette base n'est pas orthogonale. Le sous espace d'initialisation a été échantillonné par pas de 0,05 pour des taux de colmatage variant de 0 à 0,65 en partie haute et médiane et de 0 à 0,3 en partie basse. L'échantillon de profils de colmatage utilisé pour l'initialisation comporte donc $14 \cdot 14 \cdot 7 = 1372$ individus.

Une fois l'algorithme initialisé, les particules sont déplacées dans l'espace d'état à l'aide d'un noyau de transition *isotrope* : à chaque itération le colmatage de chaque demi-plaque augmente de 0,5 avec une probabilité $\frac{1}{3}$, diminue de 0,5 avec une probabilité $\frac{1}{3}$ ou ne varie pas avec une probabilité $\frac{1}{3}$. L'espace d'état est délimité par trois conditions s'exprimant par des inégalités : le taux de colmatage de chaque demi-plaque est positif et inférieur ou égal à 0,65 et le taux de colmatage des demi-plaques en jambe froide est inférieur à celui d'une demi-plaque en jambe chaude correspondante. Le noyau de transition est différent sur les « bords » de l'espace d'état : le taux de colmatage de la demi-plaque satisfaisant le cas d'égalité d'une des conditions délimitant l'espace d'état s'éloigne du bord avec une probabilité $\frac{1}{2}$ et y reste avec une probabilité $\frac{1}{2}$.

Ces deux éléments ayant été précisés, l'algorithme de filtrage particulaire mis en œuvre s'énonce comme suit :

1. Sélectionner parmi l'échantillon d'initialisation les n particules de plus forte vraisemblance ;
2. Pour chaque particule, estimer sa vraisemblance à l'aide de l'équation (3.14) ; en

- déduire les poids avec l'équation (3.15) ;
3. Si l'entropie de l'ensemble, $E(\hat{w}_1, \dots, \hat{w}_n)$, est inférieure à la limite E_{lim} , effectuer un tirage avec remise de n particules dans l'ensemble en affectant chaque particule d'une probabilité égale à son poids w_i ;
 4. Déplacer les particules à l'aide du noyau de transition ;
 5. Simuler les réponses de n. g. l. correspondant aux nouvelles positions des particules ;
 6. Reprendre la procédure à l'étape 2 jusqu'à avoir réalisé k itérations de l'algorithme.

3.5.2. Mise à l'épreuve de l'algorithme

L'algorithme qui vient d'être présenté a été appliqué à l'analyse de 52 courbes de réponses issues de g. v. plus ou moins colmatés de 6 tranches du palier 900 MW. Les résultats de ces exécutions de l'algorithme ont été dépouillés par Ninet (2009) en comparant les profils de colmatage dont les vraisemblances étaient maximales. Pour une même réponse mesurée, les trois profils indiqués comme les plus vraisemblables par l'algorithme sont parfois assez différents, en particulier pour les g. v. fortement colmatés. Par ailleurs, l'évolution des estimations du colmatage d'une demi-plaque au fil des essais R.G.L. 4 n'est souvent pas monotone. Une telle évolution en dents de scie paraît peu plausible. Pour deux des courbes de réponses analysées, l'algorithme a été exécuté à nouveau en augmentant la taille de l'ensemble de 50 à 60 particules et le nombre d'itérations de 20 à 100, faisant passer ainsi le nombre total de simulations, initialisation comprise de 1050 à 6060. Cette augmentation substantielle de la taille des échantillons n'a que marginalement amélioré la régularité des résultats.

Les deux paragraphes suivants décrivent les tests de l'algorithme effectués au cours de cette thèse. Les conclusions de ces tests ont servi de point de départ à la démarche statistique exposée dans le chapitre 10. Le premier avait pour objectif d'évaluer la robustesse de l'algorithme. Il consiste en 5 exécutions indépendantes de celui-ci pour une même courbe de réponse mesurée. Le second test visait à investiguer l'importance de la limite de rééchantillonnage.


Cinq exécutions avec les mêmes données

Cinq exécutions de 20 itérations en plus de l'initialisation ont été réalisées avec un ensemble de 50 particules pour une courbe de réponse mesurée en 2002 sur le g. v. n° 2 de la tranche de n° 2 de Charlie. Les nombres aléatoires utilisés pour décider des déplacements des particules sont obtenus par un *générateur de nombres pseudo-aléatoires* (Robert et Casella, 2004, chap. 2). Cet algorithme génère des séquences de nombres ayant les propriétés de nombres distribués indépendamment selon une loi uniforme. Chaque séquence dépend d'un paramètre d'initialisation que l'on appelle *graine*. Ainsi, les cinq exécutions réalisées correspondent à cinq graines différentes.

La figure 3.10 représente, par des diagrammes de Tukey, les distributions du taux de colmatage pour chaque demi-plaque obtenues avec les cinq graines. Chaque graphique correspond à une graine dont le numéro est indiqué à gauche et chaque diagramme

correspond à une demi-plaque. Ceux de la jambe chaude sont à gauche et ceux de la jambe froide à droite. Les plaques sont numérotées de 1 à 8 de bas en haut. Les bords des boîtes indiquent les premiers et troisièmes quartiles tandis que la barre centrale orange indique la médiane. Les bouts des moustaches sont les valeurs la plus basse et la plus haute à moins de 1,5 fois l'intervalle inter-quartiles respectivement des premiers et troisièmes quartiles. Les distances inter-quartiles sont importantes, souvent supérieures à 0,2. D'une graine à l'autre, les valeurs médianes varient dans une gamme d'environ $\pm 0,075$ en jambe chaude. Les variations en jambe froide sont moins importantes en bonne partie du fait de la proximité des bords de l'espace d'état : le taux de colmatage est minoré par 0 et majoré par le taux de colmatage en jambe chaude. Cette indétermination importante est due au fait que la localisation du colmatage modifie essentiellement la magnitude de l'effet du colmatage sur la forme des courbes de réponse et non sa nature propre. Autrement dit, un taux de colmatage modéré en plaque haute est à peu près équivalent à un colmatage important en plaque basse.

La figure 3.11 représente l'évolution au fil des itérations de l'entropie de l'ensemble et des vraisemblances individuelles des particules. L'entropie, représentée en trait orange, a été normée par l'entropie maximale, $\text{Log}(n)$. La ligne horizontale mauve indique la limite de rééchantillonnage, elle aussi normée, fixée à 0,3. L'entropie de l'ensemble d'initialisation, notée « In. » dans la figure 3.11, est presque maximale. L'ensemble d'initialisation est le même dans les cinq cas puisqu'il est construit de façon déterministe. L'entropie chute d'environ un cinquième dès la première itération, puis décroît en général d'une itération à l'autre ou augmente légèrement. Deux rééchantillonnages ont eu lieu avec les deux premières graines, et un seulement avec les trois autres graines. Les rééchantillonnages ramènent l'entropie environ à la valeur qu'elle avait à la première itération mais reste inférieure à la valeur d'initialisation.

Les barres bleues () indiquent les estimations individuelles de vraisemblances des particules. Elles sont triées, pour chaque itération, par valeurs décroissantes et normées, globalement, par la vraisemblance la plus forte. La vraisemblance la plus forte de chaque itération est indiquée par une croix. La proportion de particules dont la vraisemblance est presque nulle est très importante et ce, dès quelques itérations. Dans certains cas, les rééchantillonnages permettent de diminuer notablement cette proportion mais elle redevient rapidement prépondérante après quelques itérations. La plupart des rééchantillonnages interviennent manifestement trop tard, alors que l'ensemble est déjà dégénéré. Parmi les 1050 simulations effectuées avec chaque graine, un cinquième environ apporte véritablement de l'information et, contrairement au but de ce type d'algorithme, cette proportion a plutôt tendance à décroître au fil des itérations, une bonne part des particules intéressantes étant dans l'ensemble initial.

Variation de la limite d'échantillonnage

On a vu dans le paragraphe précédent que la limite de rééchantillonnage choisie initialement était souvent trop basse pour empêcher la dégénérescence de l'ensemble de particules. La figure 3.12 représente l'évolution de l'entropie et des vraisemblances estimées avec la graine n° 3 pour trois valeurs de la limite de rééchantillonnage : $0,3 \cdot \text{Log}(n)$ comme

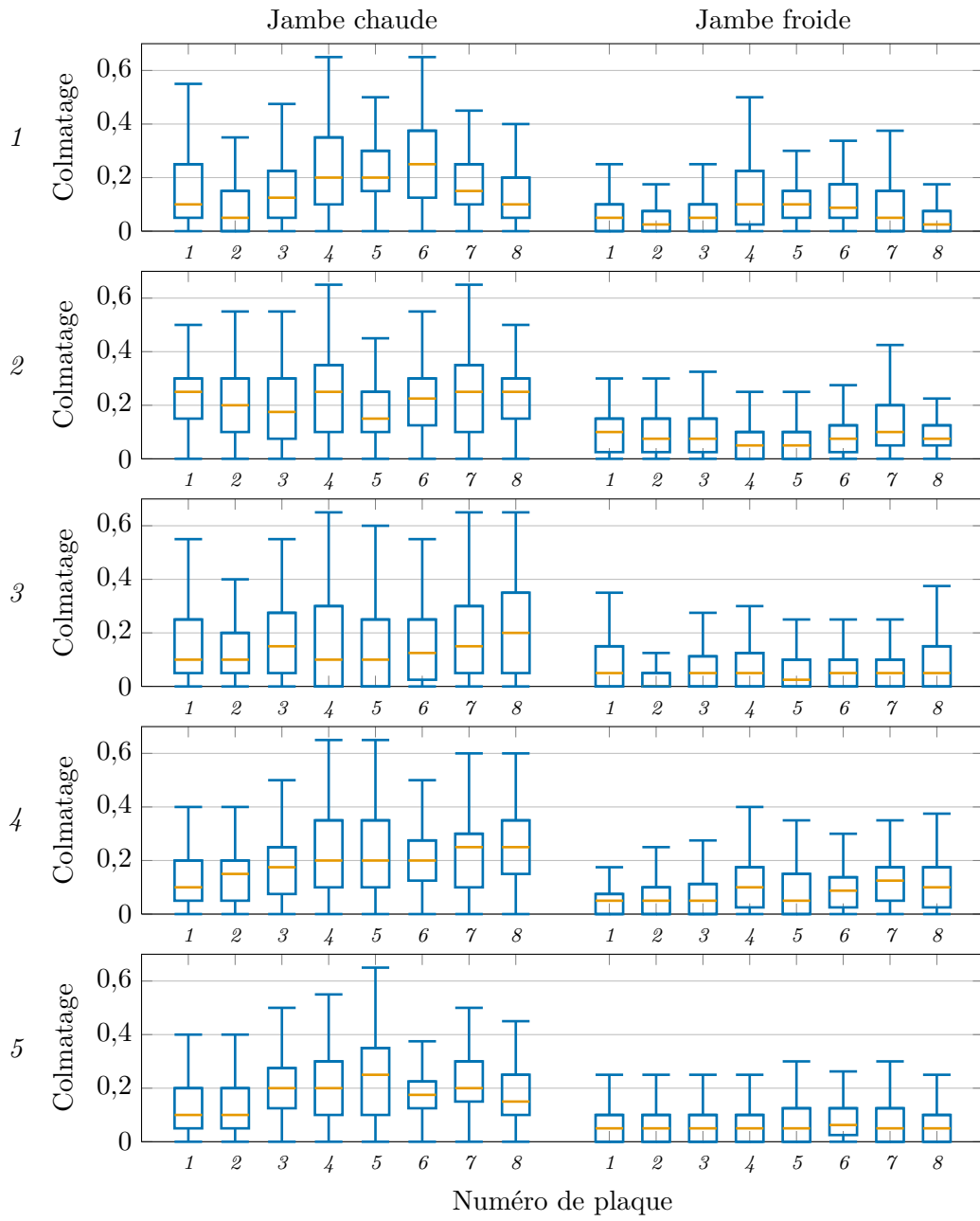


Figure 3.10. : Diagramme de Tukey des distributions des diagnostics obtenus pour les 5 graines.

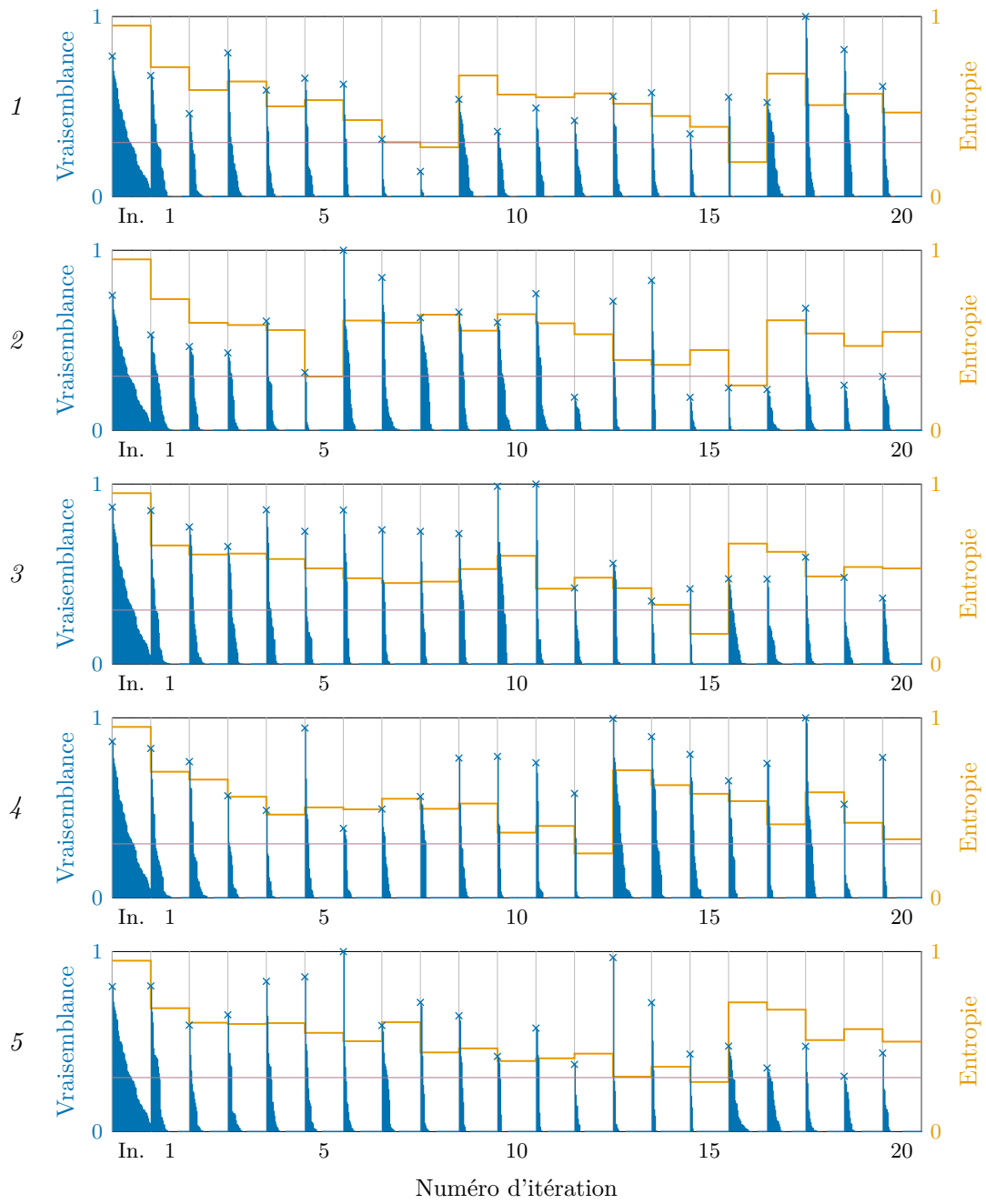


Figure 3.11. : Évolution au fil des itérations de l'entropie de l'ensemble et des vraisemblances individuelles des particules pour les 5 graines.

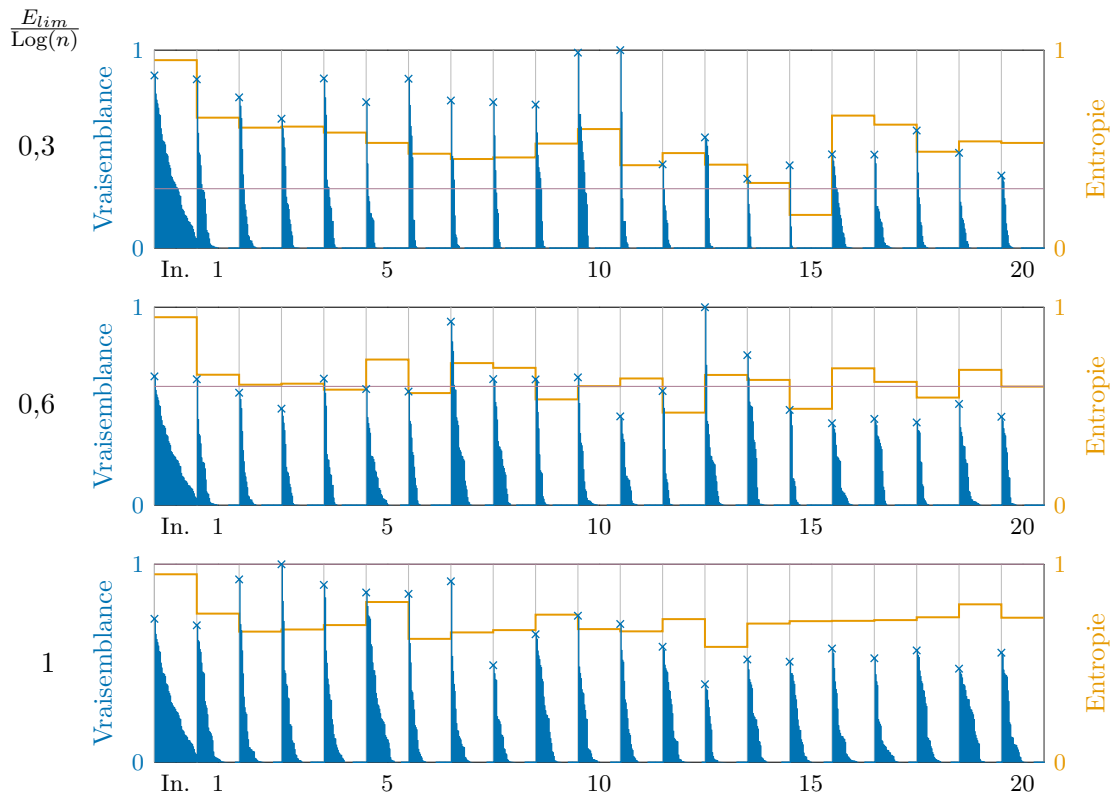


Figure 3.12. : Évolution au fil des itérations de l'entropie de l'ensemble et des vraisemblances individuelles des particules en fonction de la limite d'entropie de rééchantillonnage, E_{lim} .

dans le cas précédent ; $0,6 \cdot \text{Log}(n)$ et $\text{Log}(n)$ soit un rééchantillonnage systématique. La limite de $0,6 \cdot \text{Log}(n)$ conduit à rééchantillonner toutes les trois itérations environ et maintient l'entropie à une valeur proche de la limite de rééchantillonnage. Dans le cas du rééchantillonnage systématique, l'entropie se stabilise autour de $0,7 \cdot \text{Log}(n)$ et près de la moitié des particules contribuent véritablement au diagnostic. Par ailleurs, les vraisemblances les plus fortes de l'ensemble restent conséquentes au fil des itérations.

La figure 3.13 représente les distributions obtenues par rééchantillonnage systématique avec les graines n^{os} 3 et 4. La variance du diagnostic est sensiblement réduite par rapport aux distributions présentées dans la figure 3.10. On constate cependant encore des indéterminations. Celles-ci peuvent être levées partiellement en calculant des moyennes des taux de colmatage de plaques proches, c'est-à-dire en projetant le diagnostic sur un sous-espace de l'espace d'état.

La figure 3.14 représente les projections des profils associés aux particules générées avec les graines n^{os} 3 et 4 sur le sous-espace engendré par les vecteurs *milieu* et *haut* de la

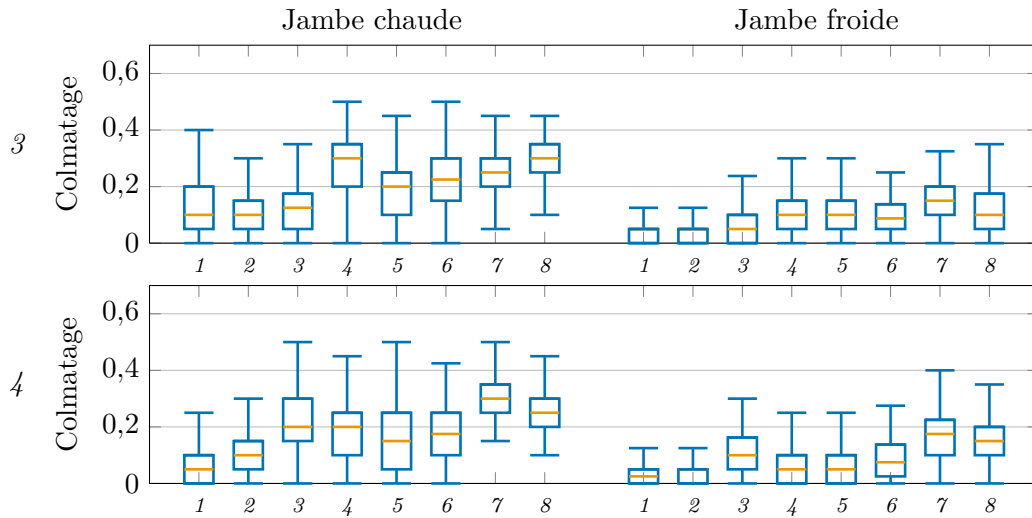


Figure 3.13. : Diagrammes de Tukey des diagnostics obtenus avec les graines n^{os} 3 et 4 en rééchantillonnant à chaque itération.

base représentée par la figure 3.9. Les coordonnées dans ce sous-espace sont des moyennes pondérées des taux de colmatage des plaques n^{os} 3, 4, 5 et 6 selon la direction *milieu* et n^{os} 6, 7 et 8 selon la direction *haut*. Pour chacune des directions de projection, la somme des poids vaut 1 si bien que les moyennes varient dans la même gamme que les taux de colmatage, soit entre 0 et 0,65. La proportion de particules de très faible vraisemblance, représentées en mauve clair, est restreinte pour les deux graines. L’algorithme délimite clairement la zone de forte vraisemblance du sous-espace : entre 0,2 et 0,35 selon le vecteur *haut* et entre 0 et 0,3 selon le vecteur *milieu*. On peut néanmoins observer que les deux graines aboutissent à deux modes différents pour le colmatage *milieu* : proche de 0,05 pour la graine n^o 3 et de 0,15 pour la graine n^o 4. La similitude des effets du colmatage dans différentes localisations est ici explicite : la projection de la région explorée sur le sous-espace *milieu-haut* décrit un quart d’ellipse dont l’excentricité dépend du rapport des sensibilités de la réponse de n. g. l. au colmatage localisé dans les parties médiane et haute.

Le rééchantillonnage systématique augmente de façon appréciable la quantité d’information obtenue pour un nombre de simulations donné. Il ressort cependant des tests réalisés que l’espace à 16 dimensions défini initialement est résolument trop vaste pour être exploré avec le modèle. En effet, la similitude des effets des dépôts de colmatage localisés dans des zones différentes du g. v. est une source d’indétermination. L’importance relative des effets du colmatage suivant sa localisation a été estimée quantitativement par une méthode d’*analyse de sensibilité*, présentée dans le chapitre 7. La détermination des sous-espaces de projection optimaux a quant elle été abordée à l’aide d’une méthode de *réduction de dimension*, présentée dans le chapitre 9. Les résultats de l’application de ces deux méthodes sont présentés dans le chapitre 10.

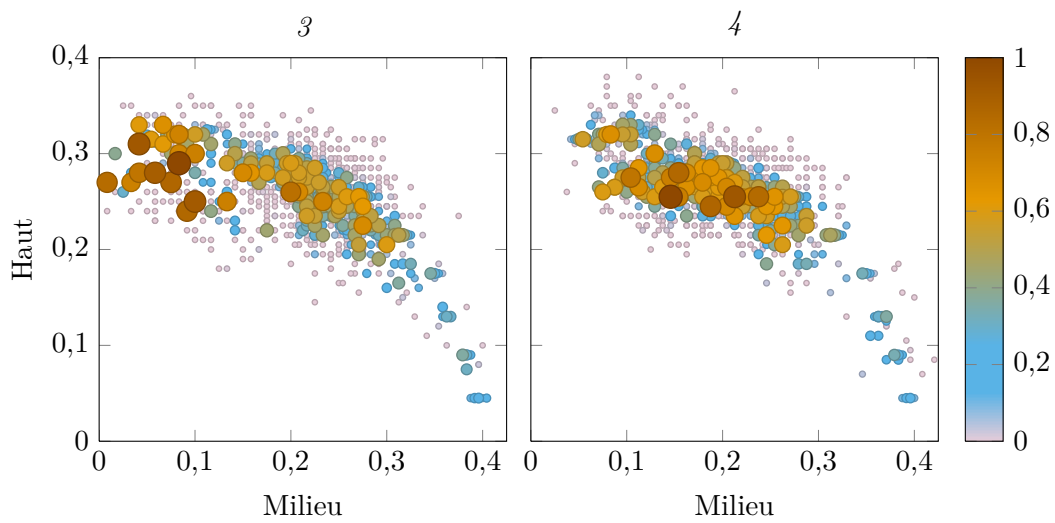


Figure 3.14. : Exploration par l'algorithme du sous-espace *milieu-haut* avec les graines n^{os} 3 et 4. Les vecteurs directeurs *milieu* et *haut* ont été normés de façon à ce que les coordonnées soient homogènes à des taux de colmatage. La vraisemblance normée par la vraisemblance maximale est indiquée par la taille des ronds et par l'échelle de couleurs représentée à droite.

3.6. Synoptique des méthodes de diagnostic existantes

Le tableau 3.2 résume les méthodes de diagnostic qui viennent d'être présentées et indique les pages auxquelles elles sont introduites.

Méthode	Description	Fréquence	Limitations	Page
Examen télévisuel	La plaque supérieure est inspectée par une caméra introduite par le haut des g. v.	Rare et coûteux	Seule la plaque supérieure est visible Échantillonnage restreint	28
Sonde à courants de Foucault	Les sondes utilisées pour contrôler l'intégrité des tubes sont sensibles au colmatage	1 par cycle	Lien implicite entre l'indicateur et le colmatage Sature au-delà de 0,5	38
N. g. l. en régime stationnaire	Le colmatage affecte la perte de charge globale du <i>riser</i> qui est directement liée à la mesure de n. g. l.	3 à 4 par cycle	Biais importants	39
Méthode à profil choisi	Le régime dynamique permet de s'affranchir des biais qui perturbent le diagnostic en régime stationnaire	3 à 4 par cycle	Hypothèses fortes, notamment sur la localisation du colmatage	42
Filtrage particulière	Cette extension stochastique permet d'alléger l'hypothèse faite sur la répartition spatiale du colmatage	3 à 4 par cycle	Les hypothèses précédentes restent en partie nécessaires	45

Tableau 3.2. : Synoptique des méthodes de diagnostic du colmatage. Les périodicités sont données par cycle de fonctionnement car les périodes d'arrêt ont une durée variable selon qu'elles consistent en un simple rechargement (3 à 4 semaines) ou comportent des opérations de maintenance plus complexes (jusqu'à plusieurs mois pour les visites décennales).

Deuxième partie

Modélisation physique
des générateurs de vapeur

4. Contexte de la modélisation physique

Les échangeurs de chaleur sont des systèmes que l'on sait bien modéliser à l'échelle du composant, par exemple en procédant à un bilan thermique et en utilisant la moyenne logarithmique des écarts de températures ou bien la méthode dite du nombre d'unités de transfert (Gicquel, 2009). Cependant, le g. v. est un système dont la géométrie est complexe et l'étude du colmatage nécessite une description plus détaillée des sous-systèmes qui le compose et de leurs interactions. Le découpage en sous-systèmes peut être fait de plusieurs façons et il faut choisir le niveau de détail de la modélisation de chacun d'entre eux en fonction de ce que l'on cherche à observer. Par exemple, la prise en compte du système de contrôle-commande augmente le couplage entre les sous-systèmes mais c'est un élément important du comportement dynamique du g. v.

Pour reproduire l'effet du colmatage de chaque demi-plaque, il est nécessaire de modéliser l'écoulement diphasique dans la partie *riser* du g. v., ce qui rend impossible une solution analytique. La modélisation de l'écoulement diphasique est une difficulté en soi. En effet, les écoulements diphasiques sont des systèmes hors équilibre et peuvent se présenter sous de nombreuses configurations, ou régimes d'écoulement, aux comportements très contrastés. La simulation numérique des écoulements diphasiques aux petites échelles est une discipline en plein essor comme l'atteste l'évolution rapide des codes commerciaux STAR-CD (CD-adapco, 2012), CFX et Fluent (ANSYS UK Ltd., 2012; Chudiak et Warnica, 2012), TransAT (ASCOMP GmbH, 2012) ou encore du code libre Neptune-CFD développé par le Commissariat à l'énergie atomique et aux énergies alternative (CEA), EDF, Areva et l'Institut national de radioprotection et de sûreté nucléaire (I.R.S.N.) (Guelfi et coll., 2005). Cependant son domaine d'application est encore restreint aux géométries simples et les temps de calculs nécessaires sont importants. Une autre approche consiste à manipuler des moyennes temporelles et spatiales des grandeurs thermodynamiques. De cette approche, dite à *milieux interpénétrants*, découlent les modèles à *deux fluides*, dont il existe de nombreuses variantes, suivant les hypothèses simplificatrices retenues. Dans ces modèles, les deux phases sont présentes partout, avec une certaine probabilité. La position de l'interface entre les phases n'est donc plus décrite explicitement. Cependant, complétés par des corrélations empiriques, ces modèles permettent de décrire des géométries plus complexes avec des temps de simulations réduits. Notons dès à présent qu'un modèle plus raffiné ne rendra pas forcément mieux compte de la réalité car il comportera de nombreux paramètres à estimer ce qui nécessite d'avoir une connaissance précise des phénomènes ayant lieu aux petites échelles. Par ailleurs, un modèle plus détaillé est souvent plus difficile à mettre en œuvre et entraîne des temps de calcul importants. En revanche, l'utilisation d'équations simplifiées et de corrélations empiriques a tendance à restreindre le domaine de validité des modèles. Certains paramètres de ces formules sont incertains et seront ajustés à partir de données

de référence : on parle de *calage* du modèle. Le modèle perd en pertinence à mesure que l'on s'éloigne des régimes de fonctionnement utilisés pour le calage. Autrement dit, il n'existe pas de modèle de g. v. « parfait », chacun étant le résultat d'options décidées en fonction de l'usage pour lequel il a été conçu. Cela explique qu'un nombre important de modèles de g. v. aient été développés. Dans sa revue de 1978, Dupont (1978) en dénombre déjà 11 et la liste s'est largement étoffée depuis avec le développement des moyens de calcul. Pour le diagnostic du colmatage, le domaine d'utilisation du modèle est relativement restreint : on cherche à simuler correctement la réponse d'une grandeur donnée lors d'un transitoire lui aussi déterminé. Cela justifie de développer un modèle « sur mesure ». Il doit reproduire le comportement dynamique du g. v. avec suffisamment de précision pour que l'information sur l'état de colmatage apparaisse dans la réponse du n. g. l. mais le temps de calcul nécessaire pour simuler un transitoire de puissance ne doit pas excéder quelques minutes afin de permettre d'explorer de nombreuses configurations de colmatage.

La section 4.1 présente quatre modèles de g. v. utilisés par EDF comparables à celui qui a été utilisé au cours de cette thèse. La section 4.2 décrit le modèle de g. v. développé à partir du code thermohydraulique tridimensionnel THYC. Il s'agit là d'un modèle plus détaillé, ne fonctionnant qu'en régime stationnaire, qui a servi de référence pour le nouveau modèle. Le modèle utilisé pour mettre au point la méthode de diagnostic à profil choisi, ainsi que ceux dont il est inspiré, sont décrits dans la section 4.3. Cette section identifie par ailleurs les défauts de cette version du modèle qui ont conduit au développement d'une nouvelle version pendant la thèse. Le détail des équations physiques du modèle sera donné dans le chapitre 5 présentant cette nouvelle version.

4.1. Présentation de quatre modèles comparables à celui utilisé

Cette section présente quatre modèles dont les domaines d'application s'apparentent à celui du modèle utilisé au cours de cette thèse. Cette revue ne se veut pas exhaustive et d'autres modèles répondant à un cahier des charges similaire ne seront pas présentés. Parmi ceux développés par EDF, on peut citer par exemple le modèle de g. v. issu du code SICLE (Kastelanski, 1983), le modèle LILIGV destiné à l'étude des régulations (Viaud et Augueres, 1991) ou encore le modèle simple monodimensionnel baptisé GV 1D (Gautheron, 2006; Bernier, 2007). Très récemment, un rapport décrivant un modèle analogue à celui utilisé lors de cette thèse a été publié par l'*Electric power research institute* (EPRI) (Tedford, 2012).

Le code PERFGV est destiné à modéliser le fonctionnement en régime permanent des g. v. afin d'en étudier les performances et d'en évaluer certains paramètres thermohydrauliques. Il a été développé par Areva. Trois modèles représentant les familles de g. v. du parc français ont été développés : un pour les g. v. 51 B et apparentés (Pasquale, 2008), un pour les g. v. de remplacement 47/22 (Makarawiez et coll., 2009) et un pour les g. v. 68/19 équipant les tranches du palier 1300 MW (Jurkowski, 2009). Le code PERFGV a été validé pour des niveaux de puissance allant de la puissance nominale à 20 % de la puissance nominale par comparaison de la pression de saturation et du

taux de recirculation avec des mesures obtenues en instrumentant l'un des générateur de vapeur de la tranche n° 1 de la centrale de Paluel (Bremaud, 2010). Les estimations de paramètres thermohydrauliques qu'il produit servent de référence pour de nombreux travaux réalisés par le Service d'études et projets thermiques et nucléaires (SEPTEN), un département d'ingénierie d'EDF (Decout, 2009a; Deotto, 2009; Goetgheluck, 2012). Elles ont été utilisées pour caler la première version du modèle utilisé pour le diagnostic du colmatage, présentée dans la section 4.3 (Chip et coll., 2008).

Le code MANTA, développé par Areva, a pour objectif de simuler le comportement transitoire de l'ensemble d'un réacteur à eau pressurisée, des boucles primaires et des g. v. associés. L'état thermohydraulique des systèmes y est représenté par des équations monodimensionnelles : deux bilans de masse et deux bilans d'énergie ainsi qu'une équation de fermeture décrivant le *glissement*, c'est-à-dire l'écart de vitesse entre les phases, fondé sur le modèle à *vitesse de dérive* appelé *drift flux* en anglais. Ce code a fait l'objet d'une procédure de validation reposant sur des comparaisons à des résultats d'expériences sur maquettes et à des transitoires réels (Brochet, 2000). Un modèle de g. v. dont la partie haute est décrite par deux jambes monodimensionnelles couplées par des liaisons hydrauliques a été développé à l'aide de MANTA et validé récemment à l'aide d'expériences sur maquettes. Cette amélioration des précédents modèles purement monodimensionnels a permis d'améliorer la représentativité du comportement dynamique du modèle (Balleydier, 2010). Ce modèle à deux jambes est utilisé par Areva pour étudier l'impact du colmatage des plaques entretoises sur le comportement dynamique des g. v. (Bittan et coll., 2009).

Le code « Analyse numérique d'un échangeur thermique » (ANETH) a été développé par Areva spécifiquement pour modéliser les générateurs de vapeur. Les modèles construits avec ANETH se composent de *volumes de contrôle* stockant masse et énergie connectés par des *liaisons* modélisant les pertes de charge. Des *capacités thermiques* représentent les solides métalliques. Comme on le verra dans la section 5.1.2, ce schéma général est similaire à celui de la bibliothèque de module ayant servi de point de départ pour la construction du modèle de g. v. avec le logiciel Dymola. La représentation du circuit primaire est monodimensionnelle et l'eau primaire est constamment en phase liquide. Le circuit secondaire est lui aussi monodimensionnel mais ses volumes peuvent comporter un *niveau* défini comme une surface horizontale séparant le volume en deux parties dont le taux de vide est distinct (Nicoli, 2010a). La recherche des niveaux est simplifiée en regroupant les volumes en branches n'admettant qu'une valeur de niveau à un instant donné. À chaque pas de temps, le système d'équations de bilan et de fermeture est résolu, puis les niveaux sont déterminés. Les volumes sans niveaux sont supposés soit monophasiques, soit diphasiques à saturation. Les volumes à niveau comportent deux zones pouvant être diphasiques. Les zones diphasiques sont supposées à saturation tandis que les zones monophasiques peuvent éventuellement être sous-saturées. Chaque zone comporte un bilan de masse et un bilan d'énergie. Le système est fermé par les équations constitutives du fluide et un modèle de séparation par gravité. Plusieurs jeux de corrélations empiriques ont été intégrés à ANETH (Nicoli, 2010b). Le code ANETH a lui aussi été validé par des comparaisons à des résultats d'expériences sur maquettes et à des transitoires réels (Holcblat, 2001; Nicoli, 2009).

Le « Code avancé de thermohydraulique pour les accidents sur les réacteurs à eau »

(CATHARE) est le fruit d'un vaste programme de développement initié en 1979 par quatre partenaires : le CEA, EDF, Areva et l'I.R.S.N. (C.E.A., 2012). CATHARE est développé par une équipe du CEA Sa finalité principale est de simuler les accidents de tranches à eau pressurisée, tel que la perte de réfrigérant primaire par ouverture d'une brèche dans une conduite. D'autres applications ont progressivement vu le jour, certaines en dehors du domaine nucléaire (Geffraye et coll., 2011). Il permet plusieurs niveaux de détail allant de modules non maillés à des modules monodimensionnels et tridimensionnels. Son domaine d'application et sa complexité dépassent nettement ceux des modèles présentés précédemment dans cette partie. Il est comparable au code RELAP5 (Idaho National Laboratory, 2012) développé par l'*Idaho National laboratory*, et utilisé par une partie des exploitants américains. La description des systèmes diphasiques dans CATHARE est un modèle à deux fluides et six équations de bilan permettant de modéliser différents régimes d'écoulement. Les lois de fermeture utilisées ont fait l'objet d'une attention redoublée. Un jeu unique de ces lois est disponible dans chaque révision du code. Leur élaboration repose sur un important ensemble d'expériences (Bestion, 1990). La procédure de validation du code comprend des tests sur boucle complète à échelle réduite.

Ces quatre modèles sont utilisés par le SEPTEN pour l'étude de l'impact du colmatage sur les g. v. Leurs résultats de MANTA, ANETH et CATHARE ont par exemple été comparés pour la prédiction d'instabilités lors d'un transitoire de sur-débit de vapeur (Devidal et Foullon, 2009). Les seuils d'instabilité prédits diffèrent sensiblement mais les résultats obtenus sont globalement concordants. Bien que ces modèles, ou des simulations réalisés avec eux, soient utilisés par EDF, le département de recherche et développement d'EDF ne dispose pas de versions exécutables, et a fortiori modifiables, de ces modèles, pour des raisons de propriété industrielle. Pour certains, l'obtention de documents descriptifs est elle-même délicate.

4.2. Un modèle de référence : THYC

Le logiciel THYC permet de modéliser des écoulements dans un faisceau de crayons chauffants ou dans un faisceau de tubes comme celui des g. v. (Guelfi et Pitot, 2004). Son domaine d'application est comparable à celui du code « *Analysis of the thermal hydraulics of a steam generator/Steam generator analysis package* » (ATHOS/SGAP) développé par l'EPRI (Singhal et Srikantiah, 1991; EPRI, 2008). Une comparaison récente a montré une bonne adéquation de ces deux codes pour les variables principales (David, à paraître). Les principales différences concernent la partie cintrée du faisceau de tubes pour laquelle des options de modélisation différentes ont été prises.

Des modèles de g. v. des centrales françaises ont été construits avec THYC. Dans la suite, notamment dans le chapitre 6, on désignera parfois simplement par « THYC » le modèle de g. v. de type 51 B et apparentés réalisés avec ce code. Le domaine de modélisation est maillé en volumes de contrôle qui peuvent contenir un solide, de l'eau ou bien les deux. Les équations sont moyennées dans chaque maille et lorsqu'un volume de contrôle contient de l'eau et un solide, on définit la porosité volumique du fluide (resp. du solide) dans la maille comme le rapport du volume de fluide (resp. de solide) au

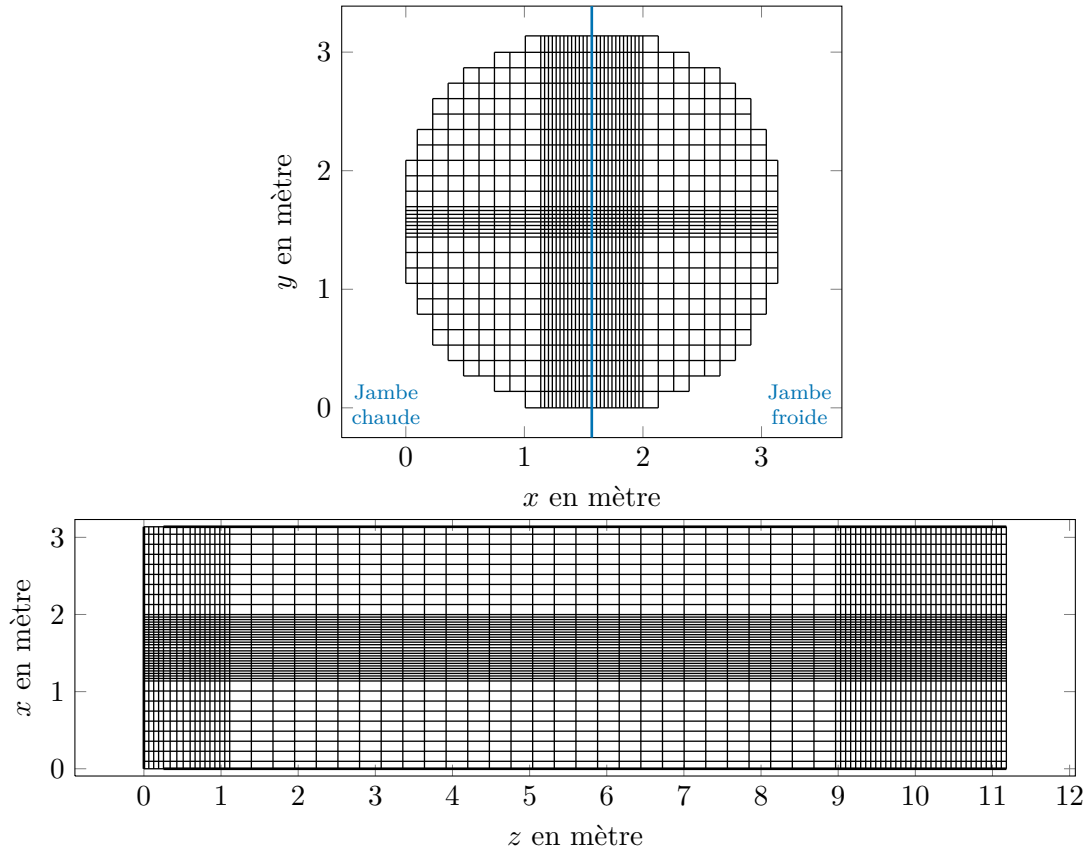


Figure 4.1. : Représentation du maillage du modèle THYC de g. v. 51 B dans un plan horizontal (en haut) et vertical (en bas).

volume de la maille. L'écoulement diphasique est décrit par un modèle à deux fluides. Les équations de conservation sont écrites sous la forme de trois équations de mélange auxquelles s'ajoute une équation d'évolution décrivant la variation de masse de la phase la moins présente, supposée à saturation, et une équation décrivant le déséquilibre des vitesses.

Le maillage de g. v. 51 B de THYC est constitué de $44 \cdot 30 \cdot 78 = 94\,536$ mailles. La figure 4.1 montre le resserrement du maillage en bas et en haut du g. v. ainsi qu'au centre des plans horizontaux.

L'application du code THYC à la modélisation des g. v. a fait l'objet d'une procédure de vérification et de qualification poussée (Tincq et David, 1994; David et Guelfi, 1999). Des expériences sur maquettes et en situations réelles ont permis de valider la simulation du fonctionnement stationnaire des g. v. Ce n'est pas le cas des régimes dynamiques et la régulation du débit de vapeur n'est pas modélisée. La recirculation est prise en compte par l'ajout d'un modèle de retour d'eau, éventuellement maillé. Le débit de recirculation est déduit de la différence de pression entre les mailles d'entrée du retour d'eau, en aval des séparateurs et les mailles d'entrée dans le *riser*, en bas du g. v. En imposant la

puissance de consigne, l'enthalpie et la pression de l'eau alimentaire et le niveau d'eau dans le dôme, une procédure itérative permet de déterminer la pression de la vapeur dans le dôme et d'en déduire celle des mailles d'entrée dans le retour d'eau. Les simulations réalisées avec THYC nécessitent quelques heures pour converger.

Le modèle de g. v. issu de THYC ne permet pas de simuler des transitoires et ses temps de calcul sont trop importants pour l'application des méthodes statistiques envisagées. Néanmoins, la qualité de sa procédure de validation et le fait qu'il soit disponible dans le département de recherche et développement de EDF en font le candidat idéal pour la validation et le calage d'un modèle développé *ad hoc* pour le diagnostic du colmatage. Cette utilisation du modèle THYC fait l'objet du chapitre 6. Par ailleurs, des modèles utilisant THYC pour simuler la formation des dépôts de colmatage et d'encrassement ont été développés (Moleiro et coll., 2010; Pujet, 2001; Prusek, à paraître). L'utilisation de ces modèles pour formuler une hypothèse sur la répartition spatiale du colmatage fait l'objet du chapitre 11.

4.3. Version initiale du modèle de générateur de vapeur

Un modèle dédié au diagnostic du colmatage par analyse de la réponse dynamique du n. g. l. a été développé par Chip et coll. (2008). La version initiale de ce modèle s'inspirait d'un modèle préexistant appelé DRIF-BI (Avenas et coll., 2004), lui-même une traduction en langage Modelica d'un modèle écrit avec le code LEDA d'EDF (Avenas et coll., 2005). Le modèle DRIF-BI écrit en langage Modelica s'est avéré instable numériquement et particulièrement difficile à initialiser, c'est pourquoi il avait été abandonné au profit d'une formulation plus simple (Chip et coll., 2008). Celle-ci a par la suite été améliorée progressivement afin d'en accroître la stabilité et de reproduire plus en détail la dynamique des g. v., réintroduisant ainsi certains éléments de la modélisation physique de DRIF-BI. Dans la suite de l'exposé, en l'absence de précisions, c'est de ce modèle qu'il s'agira.

Parmi les versions successives du modèle, on distinguera les versions numérotées 2 et 7. La version 2 a été utilisée pour mettre au point la méthode de diagnostic à profil choisi décrite dans la section 3.4.2. C'est celle qui était disponible au début de la thèse. Le document de référence la décrivant est le rapport de Midou et coll. (2010a). La version 7 a été utilisée lors du processus d'industrialisation de la méthode de diagnostic à profil choisi (Ninet et coll., 2012). En l'absence de précisions, les résultats présentés ici ont été obtenus avec cette version ou la version précédente, décrite par Deneux et Girard (2011), qui lui était très similaire. C'est à ce jour la version la plus aboutie du modèle. Le détail des équations de la version 7 fait l'objet du chapitre 5 dans lequel sont par ailleurs signalées les différences principales avec la version 2. Les deux prochaines sections exposent les limitations de la version 2 qui ont motivé les développements aboutissant à la version 7. Chacun des paragraphes suivants présente l'un des problèmes identifiés. Les solutions choisies pour y remédier sont énoncées succinctement. Certaines ont nécessité des développements additionnels, d'autres des changements méthodologiques. Elles sont détaillées dans les chapitres 5 et 6.

4.3.1. Problèmes inhérents au modèle

La démarche de simplification ayant abouti à la version 2 avait permis de gagner en flexibilité et en stabilité numérique. Cette version comportait néanmoins certaines lacunes qui sont présentées ci-après.

Instabilités numériques

Lorsque les conditions aux limites du modèle version 2 comportent une rupture de pente dans leurs variations, le n. g. l. entre dans un régime d'oscillations rapidement amorties mais d'amplitude assez importante et croissante avec le taux de colmatage. Le graphique de gauche de la figure 4.2 représente les réponses de n. g. l. lors d'essais R.G.L. 4 simulés avec la version 2 du modèle pour les 15 configurations de colmatage représentées dans la figure 3.8 (chapitre 3, page 44). La teinte des courbes vire du vert (—, débutant en haut à gauche et terminant en bas à droite) à l'orange (—, débutant en bas à gauche et terminant en haut à droite) à mesure que le colmatage augmente depuis un état parfaitement sain à un taux de 0,7 en plaque supérieure en jambe chaude. Les moyennes temporelles des courbes leurs ont été soustraites. Les oscillations sont déclenchées par le début de la chute de puissance. Elles s'amortissent rapidement lorsque le colmatage est faible mais persiste pendant environ 300 s lorsqu'il est élevé.

Le graphique central de la figure 4.2 représente le même faisceau de courbe que la figure 3.7, la couleur des courbes évoluant du vert à l'orange par état de colmatage présumé croissant, en supposant une augmentation monotone du colmatage et une efficacité absolue du nettoyage chimique. Les différences de pression mesurées par les capteurs installés sur les g. v. du pallier 900 MW sont converties par une fonction affine en une mesure de niveau en mètre. Il y a par ailleurs une indétermination de l'amplitude de variation du n. g. l. dans le modèle. Pour ces raisons, les ordonnées du graphique central de la figure 4.2 ont été normalisées par un facteur -4000 . Ce point est important pour la méthode de diagnostic ; il sera abordé plus en détail dans la section 6.3 et dans le chapitre 13. Les oscillations produites par la version 2 ne sont pas observées lors des transitoires réels. Elles sont donc vraisemblablement d'origine purement numérique. Ce phénomène est particulièrement gênant car il provoque une déformation des courbes de réponse dépendante du colmatage et donc susceptible de fausser le diagnostic.

Ces oscillations ont été supprimées dans la version 7 du modèle comme le montre le graphique de droite de la figure 4.2 représentant les courbes de réponse de n. g. l. obtenues avec la version 7 du modèle dans les mêmes conditions de simulations que celles obtenues avec la version 2. Plusieurs erreurs dans le code ont été corrigées au fil des versions, certaines corrélations remplacées et le schéma numérique a été réécrit de façon plus rigoureuse. La cause exacte des oscillations n'a pas pu être déterminée.

Indépendance des deux jambes

L'une des simplifications de la version 2 par rapport à DRIF-BI est l'indépendance des deux jambes. Dans la version 2, chaque jambe du *riser* est représentée par un canal distinct et elles n'interagissent qu'indirectement par le biais de la recirculation : tout

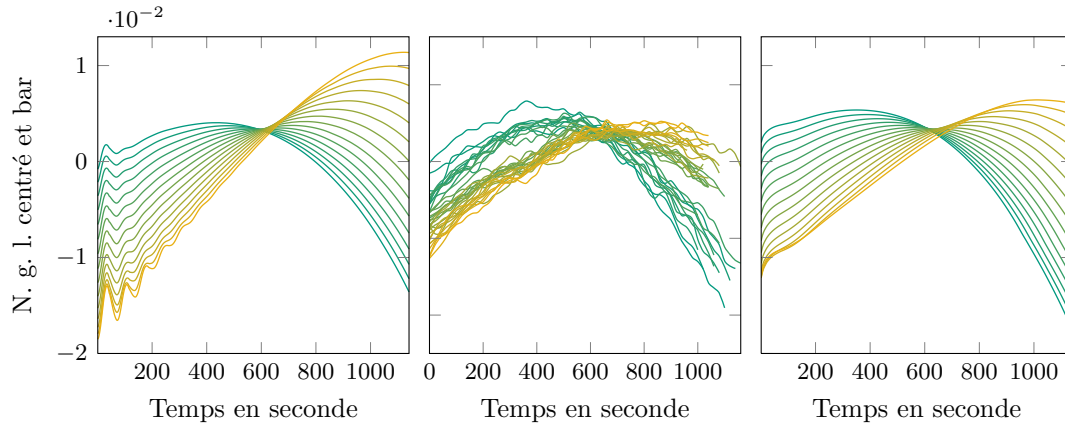


Figure 4.2. : Comparaison de courbes de réponse pour des configurations de colmatage faibles et forte simulées avec la version 2 du modèle, à gauche, mesurées sur le g. v. n° 3 de la tranche n° 3 de la centrale Charlie, au centre, et simulées avec la version 7 du modèle, à droite.

se passe comme si une plaque étanche et calorifugée séparait les deux jambes. Cette approximation impacte fortement les simulations, notamment parce qu'elle autorise des déséquilibres de pression importants entre jambes. Les champs de pression simulés avec THYC sont au contraire très homogènes dans le plant transversal à l'écoulement.

On qualifie de *transverses* les débits traversant perpendiculairement le plan médian séparant les jambes. Les plaques entretoises constituent une restriction importante de la section de passage du *riser*. La résistance à l'écoulement qu'elle engendre dépend de la vitesse et du titre du fluide, lesquels sont en général inégaux entre jambes du fait de l'asymétrie de l'échangeur (la jambe chaude reçoit plus d'énergie du fluide primaire). Aussi on s'attend à observer des débits transverses importants juste en dessous des plaques. Le colmatage réduit l'aire de passage du fluide affectant ainsi les débits transverses. Réciproquement, les débits transverses modifient l'écoulement du fluide dans chaque demi-plaque, et donc sa contribution à la perte de charge totale du *riser*. Les débits transverses introduisent un couplage de l'effet du colmatage des demi-plaques des deux jambes. Notons que les estimations de taux de colmatage par examen télévisuel et sonde à courants de Foucault suggèrent que le colmatage est souvent plus fort en jambe chaude qu'en jambe froide (Romeyer, 2007; Bernier, 2007), ce que soutient le modèle de déposition par *vena contracta* exposé dans la section 11.1 (Prusek, à paraître).

Des liaisons hydrauliques avec perte de charge ont été ajoutées dans le modèle afin de prendre en compte des débits transverses. L'ajustement des coefficients de perte de charge associés est traité dans le chapitre 6. Le transfert thermique par conduction entre jambes a été supposé négligeable devant le transport par convection. Des différences d'effet du colmatage ont été observées par Bittan et coll. (2009) pour trois modèles réalisés avec le code MANTA : à un seul canal, à deux canaux distincts et à deux canaux connectés par des liaisons hydrauliques. L'analyse de sensibilité du modèle présentée dans le chapitre 7

a mis en évidence l'altération de l'effet du colmatage de chaque demi-plaque par les débits transverses.

Superposition des faisceaux de courbes

L'objectif du modèle est de reproduire l'effet du colmatage sur la forme des courbes de réponse du n. g. l. Les échantillons de courbes de réponse représentés par la figure 4.2 comprennent des états de colmatage faibles et forts. L'annexe F page 249 représente par ailleurs l'intégralité des faisceaux de réponse mesurée traités dans le cadre de cette thèse. Les courbes sont triées du vert à l'orange par ordre de colmatage présumé croissant, en faisant l'hypothèse que le colmatage augmente monotonement et qu'il est minimal juste après nettoyage. La plupart des faisceaux de courbes de l'annexe F sont similaires au faisceau central de la figure 4.2. Ceux des g. v. de la tranche n° 1 de la centrale Delta, et dans une moindre mesure de la tranche n° 4, sont sensiblement différents car ils comportent, pour les taux de colmatage importants, des courbes de réponse presque rectilignes dont le point d'arrivée est nettement au dessus du sommet de l'arc formé par les courbes correspondant à des taux plus faibles. Il n'y pas eu d'estimation par examen télévisuel du colmatage de la tranche n° 4 de la centrale Delta, ni de la tranche n° 1 avant son nettoyage chimique en 2007. Plusieurs éléments permettent néanmoins de déduire que ces deux tranches étaient, avant nettoyage, dans un état de colmatage extrême. Ce sont en effet le g. v. n° 2 de la tranche n° 1 et n° 2 de la tranche n° 4 qui ont subi des fissurations de tubes, imputées a posteriori au colmatage (Juillot Guillard, 2007). Cinq examens télévisuels des g. v. de la tranche n° 1 ont été réalisés dans les années qui ont suivies leur nettoyage. En 2009, les trois g. v. de la tranche n° 1 ont été inspectés, puis, en 2008 et 2011, le g. v. n° 2 seulement (Doll, 2011). Lors de chacun de ces examens, une proportion de quelques pour-cent des trous foliés inspectés avait un taux de colmatage supérieur à 0,8, certains trous étant même entièrement obstrués. Ces trous fortement colmatés ont un impact faible sur le taux de colmatage moyen mais sont au contraire déterminants lorsque l'on raisonne avec le taux de colmatage uniforme équivalent. On a vu dans la section 3.1.2, qu'il a été décidé de n'exploiter les résultats de l'expérience P2C pour établir la corrélation donnant la perte de charge en fonction du colmatage que sur la gamme de taux de colmatage de 0 à 0,75. On peut calculer les taux de colmatage uniformes équivalents associés aux examens télévisuels des g. v. de la centrale Delta en utilisant les deux points de donnée mis de côté, correspondant à des taux de colmatage de 0,86 et 0,95, et en extrapolant jusqu'à 1. Compte-tenu du caractère approximatif de ces extrapolations, seuls les ordres de grandeur des taux de colmatage uniformes équivalents sont donnés, à titre indicatif. Les valeurs effectives de taux de colmatage rapporté à une dimension sont probablement situées entre ces valeurs extrapolées et les taux moyens. Les données utilisées pour ces calculs ont été rassemblées par Mangin et Goetgheluck (à paraître). La méthode de calcul des taux de colmatage uniforme équivalent mise en œuvre ici diffère de celle utilisé par Mangin et Goetgheluck (à paraître) notamment par le modèle de perte de charge et le traitement séparé des deux jambes. Le taux de colmatage moyen du g. v. n° 2 valent respectivement 0,15, 0,25 et 0,3 en 2008, 2009 et 2011. Sur cette même période les taux de colmatage uniforme

équivalent valent 0,5, 0,7 et 0,8. Quelque soit la validité de ces calculs, la présence avérée de trous foliés colmatés à plus de 80 % après nettoyage atteste d'un état de colmatage avant nettoyage particulièrement fort.

Sur la figure 4.2, on peut observer que la version 2 du modèle produisait des courbes presque rectilignes, analogues à celle associées aux états de colmatage extrême qui viennent d'être décrits. Cependant, le taux de colmatage le plus fort parmi les profils utilisés pour ces simulations, représentés dans la figure 3.8 page 44, n'est que de 0,7. Il a été choisi, pour la méthode de diagnostic développée dans cette thèse, de se concentrer sur la gamme des colmatages allant de 0 à 0,7 et ce pour deux raisons qui ont été évoquées dans la section 3.1.2. La première est que la corrélation donnant la perte de charge en fonction du colmatage perd en fiabilité à mesure que l'on sort de cette gamme. Il est par ailleurs probable que la modélisation des colmatages forts requiert de prendre en compte d'autres éléments que la perte de charge singulière. On peut par exemple supposer que le régime d'écoulement, c'est-à-dire l'agencement spatial des phases liquide et vapeur, soit fortement modifié par le passage au travers d'une plaque dont les trous sont presque totalement obstrués. Ce type d'effet sort pour le moment du cadre de modélisation choisi, exposé dans le chapitre suivant. La seconde est que les états critiques qu'ont connus les tranches n^{os} 1 et 4 de la centrale Delta, qui ont déclenché le programme de recherche sur le colmatage, ne sont plus susceptibles de se reproduire du fait de la stratégie de surveillance et de prévention du colmatage déployée.

Les degrés de liberté de la version 2 ne permettaient pas d'altérer la forme des courbes de réponse sans attribuer des valeurs peu plausibles aux coefficients et paramètres de calage. Autrement dit, il manque à cette description des g. v. un phénomène physique essentiel pour reproduire correctement la forme de la réponse du n. g. l. Les notions de degré de liberté et de coefficients et paramètres de calage seront abordées plus en détail dans la section 4.3.2. Dans la version 2, l'écoulement diphasique était représenté par le modèle dit *homogène* (Yadigaroglu, 2011a) dans lequel les deux phases ont la même vitesse. Cette situation d'équilibre cinématique simplifie considérablement la description de l'écoulement diphasique car elle permet d'utiliser les mêmes équations que pour décrire un écoulement monophasique. Elle est cependant rarement rencontrée en réalité (Yadigaroglu, 2011b). Un modèle de *glissement*, c'est-à-dire de déséquilibre cinématique entre phases, a été ajouté au modèle. L'un des paramètres de ce modèle permet d'agir sur la forme des courbes de réponses. Le graphique de droite de la figure 4.2 représente les courbes simulées avec la version 7 du modèle pour les mêmes configurations de colmatage que celles utilisées pour le graphique de gauche. Ces courbes simulées sont en meilleure adéquation avec celles représentées dans le graphique central et dans l'annexe F, si l'on met de côté les états avant nettoyage des tranches n^{os} 1 et 4 de la centrale Delta.

4.3.2. Méthodologie de calage du modèle

Certaines corrélations utilisées dans le modèle ont été mises au point pour des géométries et des régimes d'écoulement plus simples que ceux rencontrés dans un g. v. Elles nécessitent par conséquent des ajustements mais, faute de données pour améliorer leur formulation, on se contentera de leur juxtaposer des facteurs multiplicatifs auxquels on s'efforcera

d'attribuer des valeurs raisonnables. Ces facteurs seront appelés par la suite les *coefficients de calage* du modèle. Par ailleurs, les valeurs de certains paramètres sont inconnues et ne peuvent être estimées que très grossièrement à l'aide des corrélations disponibles dans la littérature. Il s'agit principalement des coefficients de perte de charge dans des portions du g. v. dont la géométrie est complexe, tels que les séparateurs cyclones ou bien le *downcomer* et sa jonction avec le *riser*, en bas du g. v. Ces paramètres seront appelés *paramètres de calage*. La méthodologie de calage et de validation du modèle est esquissée dans ce paragraphe. Elle sera décrite en détail dans le chapitre 6 qui lui est consacré. Les coefficients et paramètres de calage constituent les degrés de liberté de la modélisation. Certains coefficients et paramètres de calage ont une importance non négligeable sur l'allure des courbes de réponse de n. g. l. Aussi, la méthodologie permettant de les déterminer est aussi importante que le choix des équations du modèle et leur transcription pour la résolution numérique.

Pour effectuer le calage du modèle, il faut définir des *critères de calage*, c'est-à-dire choisir des grandeurs témoins et des valeurs cibles associées. Les grandeurs témoins sont choisies parmi les grandeurs « importantes » du modèle (ce qui dépend de l'usage envisagé) pour lesquelles il est possible de choisir une valeur cible (mesure ou résultat de simulation d'un autre modèle ayant été validé). Il s'agit ensuite de trouver un compromis entre les différents critères de calage en minimisant simultanément les écarts à 1 des coefficients de calage et fixant les paramètres de calage à des valeurs plausibles. Suivant les cas, le système à optimiser peut être sur-contraint ou sous-contraint. Dans le premier cas, si le modèle n'est pas en parfaite adéquation avec les données dont sont issues les valeurs cibles, certains critères devront être violés. Dans le second cas, plusieurs calages seront possibles et il sera difficile de choisir l'un d'entre eux. Il peut être délicat de constituer le système de calage car les liens entre les grandeurs de calage et les grandeurs témoins sont mal connus.

Calage des pertes de charges par rapport à PERFGV

La puissance totale échangée et le taux de recirculation sont connus pour certains régimes de fonctionnement et constituent des grandeurs témoins évidentes. Néanmoins, les grandeurs témoins mesurables dont on dispose sont en nombre assez restreint et le système de calage de la version 2 était sous-contraint. Afin d'y remédier, il a été choisi de fixer les valeurs des coefficients de perte de charge inconnus de façon à ce que les pertes de charges correspondantes soient égales à celles prédites par le modèle PERFGV. En d'autres termes, les pertes de charge inconnues ont été choisies comme grandeurs témoins et les valeurs cibles sont celles données par PERFGV. Le modèle PERFGV est fiable car il a été soumis à une procédure de qualification, cependant cette dernière ne concernait pas directement les pertes de charges internes au g. v., qui elles restent incertaines. On ne dispose pas de mesures de ces pertes de charge et il est plus prudent d'en faire des paramètres de calage que des grandeurs témoins.

La prise en compte des débits transverses et du déséquilibre cinématique entre phases augmente le nombre de degrés de liberté de la modélisation, c'est-à-dire de paramètres incertains à caler. C'était en quelque sorte l'objectif de ces améliorations, car la version 2

ne permettait pas de reproduire fidèlement certaines observations. Cependant, l'ajout de paramètres de calage nécessite de trouver de nouveaux critères de calage.

Ces considérations ont conduit à mettre au point une nouvelle méthodologie de calage fondée, d'une part, pour le comportement stationnaire, sur la comparaison en régime nominal des profils monodimensionnels de grandeurs thermodynamiques simulées par le modèle de diagnostic du colmatage aux champs thermohydrauliques tridimensionnels simulés avec THYC, et d'autre part, pour le comportement dynamique, sur l'ajustement d'un faisceau de courbes de réponses de n. g. l. La mise en œuvre de cette méthode est présentée dans le chapitre 6.

Validation

Le système de calage admet plusieurs solutions qui peuvent présenter des différences notables. Par ailleurs, certains coefficients de calage, associés aux corrélations empiriques jugées les plus incertaines, ont été fixés à des valeurs résolument différentes de 1. Aussi, il est nécessaire de disposer d'une méthode de validation pour conforter la confiance que l'on a dans les prédictions du modèle. L'approche choisie a été d'effectuer le même type de comparaison avec THYC en régime stationnaire que pour le calage mais pour une puissance basse, de l'ordre de celle atteinte à l'issue d'un essai R.G.L. 4. Les résultats de ce test de la validité de la version 7 sont donnés dans la section 6.4.

5. Modèle dynamique monodimensionnel de générateur de vapeur

Le chapitre précédent a montré l'intérêt d'écrire un modèle dédié au diagnostic du colmatage par l'étude de la réponse dynamique du n. g. l. La nécessité de simuler des régimes transitoires, qui plus est en un temps n'excédant pas quelques minutes, a dès le début orienté la modélisation vers une représentation monodimensionnelle. La réécriture en langage Modelica du modèle de g. v. DRIF-BI (Avenas et coll., 2004) avait démontré la faisabilité de cette approche et ce langage a été choisi pour réaliser le modèle de diagnostic du colmatage. Modelica est un langage orienté objet non causal (Mod, 2005). Un modèle écrit en langage Modelica est constitué d'objets contenant des équations physiques dont l'ordre n'importe pas. Dans la suite, le terme *modèle* désigne suivant le contexte l'ensemble des équations physiques décrivant le fonctionnement du g. v. ou leur écriture en langage Modelica. On appelle *modules*, les sous-ensembles du modèle. Dans l'environnement Modelica, les modules correspondent à des classes « model », au même titre que le modèle lui-même qui en est la classe parente. Les modules représentent des entités physiques du g. v. et sont reliés entre eux par des connecteurs qui transmettent certaines valeurs calculées d'un module à l'autre, en général celles des grandeurs d'état. La section 5.1 détaille le découpage du modèle en modules ainsi que les hypothèses globales de modélisation et le schéma numérique choisi. Par souci de cohésion de l'exposé, la description de chacun des éléments du modèle est précédée de l'essentiel du raisonnement et des calculs mis en œuvre pour le construire. Les principales équations du modèle sont repérables à leur fond gris clair tandis que les hypothèses de modélisation sont signalées par l'abréviation « *Hyp.* » en marge. Les définitions se distinguent des autres équations par le symbole « \equiv » remplaçant « = ».

Les raisons qui ont présidé au choix du langage Modelica sont la bonne lisibilité du code écrit sous forme d'équations physiques facilitant le travail en collaboration et la structure modulaire offrant une flexibilité accrue. De fait, la version 2 était déjà une version revue et corrigée par Julien NINET du modèle initialement écrit par Chip et coll. (2008). Les lacunes identifiées de la version 2 ont conduit à des développements ultérieurs aboutissant à la version 7 du modèle. Celle-ci est le fruit d'un travail de groupe dont une part importante a été accomplie par Julien Ninet et Olivier Deneux. Notamment, la restructuration des modules pour prendre en compte les débits transverses est due à Julien NINET et la réécriture du module représentant le *riser* à Olivier DENEUX. La section 5.1 décrit la structure générale du modèle et son mode de résolution. Les sections 5.2, 5.3 et 5.4 donnent le détail des équations physiques décrivant respectivement l'écoulement monophasique à l'intérieur des tubes, l'écoulement diphasique dans le *riser* et la perte de charge à la traversée des plaques entretoises dans la version 7 du modèle. Les

principales différences avec la version 2 y sont signalées. On se reportera aux rapports de Deneux et Girard (2011); Ninet et coll. (2012) pour la transcription en langage Modelica de ces équations.

5.1. Cadre général de la modélisation

Afin de pouvoir explorer la diversité des configurations spatiales du colmatage, il est nécessaire de réaliser un grand nombre de simulations. Pour cette raison, le temps de calcul pour chaque simulation doit être assez court, de l'ordre d'une dizaine de minutes maximum. Par ailleurs, les données dont on dispose sont issues de mesures toutes externes au g. v. et l'on a très peu de connaissances sur les phénomènes qui s'y déroulent aux petites échelles. La température primaire en branche chaude, par exemple, est mesurée par trois écopés disposées autour de la conduite en amont du g. v. et l'on fait l'hypothèse que la moyenne de ces mesures est représentative de la température moyenne dans la conduite. Il est en revanche impossible de connaître la température du fluide primaire dans chacun des tubes de l'échangeur. Ces considérations ont amené à choisir une représentation monodimensionnelle du g. v. en supposant que l'essentiel des phénomènes qui ont une influence sur le comportement dynamique des grandeurs mesurées ont lieu dans la dimension verticale. On distingue néanmoins la jambe chaude de la jambe froide afin de rendre compte de l'asymétrie du fonctionnement thermohydraulique du g. v. (à co-courant en jambe chaude, à contre-courant en jambe froide) d'une part, et de celle du colmatage d'autre part.

L'utilisation de Modelica incite à décomposer le modèle en parties et sous-parties, appelées modules. Le modèle comprend la cuve du g. v., la partie du circuit primaire circulant dans les tubes en U , la partie haute du g. v. (séparateurs cyclones, cadres sécheurs, dôme et sortie vapeur, jusqu'au capteur de pression de vapeur) et la régulation du débit d'eau alimentaire. Les conditions aux limites du modèle concernent donc l'entrée d'eau alimentaire, la sortie de vapeur ainsi que l'entrée et la sortie d'eau primaire.

La cuve comporte deux parties : la colonne d'arrivée d'eau, appelée *downcomer*, et l'échangeur à proprement parler, appelé *riser*. Le *downcomer* et le *riser* sont chacun dédoublés dans le modèle Modelica afin de représenter les deux jambes. Le *riser* est une alternance de modules « échangeur diphasique » et de modules « plaque entretoise ». Le module « échangeur diphasique » représente le fluide secondaire et l'interface entre ce dernier et les tubes en U . Il est maillé, c'est-à-dire que l'on prend en compte son étendue géométrique selon la direction verticale en le divisant en plusieurs parties appelées *mailles géométriques*. D'autres mailles seront définies plus loin ; elles correspondent aux notions de *volume de contrôle* et d'*élément bipart*. Ces notions sont explicitées dans le paragraphe 5.1.2. Le module « plaque entretoise » est de dimension nulle et modélise la perte de charge singulière au passage des trous foliés.

Les débits transverses entre la jambe chaude et jambe froide du *riser* sont modélisés par un module « tuyau monophasique » dont les connections sont situées en dessous des plaques entretoises. Le module *tuyau monophasique* lie essentiellement le débit à la perte de charge par un coefficient de perte de charge. L'utilisation de ce module ici

est un artifice de modélisation, c’est pourquoi on se contente d’une perte de charge monophasique. En effet, on fixe le coefficient de perte de charge en fonction du débit que l’on veut modéliser afin d’obtenir une relation simple liant la différence de pression entre jambes à la valeur du débit transverse mais on ne cherche pas à modéliser la perte de charge elle-même. Le chapitre 6 traite plus en détail du choix du positionnement des liaisons entre jambes et de l’ajustement de leur coefficient de perte de charge.

Le circuit primaire est modélisé par un module « échangeur monophasique » maillé modélisant le fluide primaire et l’interface de ce dernier avec les tubes en U .

La partie haute du g. v. est modélisée par des modules « tuyau diphasique », représentant la perte de charge due aux séparateurs cyclones, et un module « dôme » modélisant l’ensemble de la séparation eau–vapeur par les cyclones et les cadres sécheurs.

La partie de la régulation prise en compte dans le modèle est l’asservissement des vannes régulant localement le débit alimentaire afin de maintenir le niveau gamme étroite à sa valeur de consigne. Les valeurs du niveau gamme étroite et de débit vapeur utilisées par le système de contrôle-commande sont calculées respectivement dans le module « dôme » et dans un module « tuyau monophasique », à la sortie du dôme.

5.1.1. Résolution du système

Les modèles écrits en langage Modelica ne comportent pour ainsi dire que les équations physiques. Plusieurs logiciels, dont Dymola utilisé par EDF, permettent d’interpréter ce code pour produire un programme compilé intégrant une méthode numérique de résolution du système d’équations. La version de Dymola utilisée est la version 6.0d (Dyn, 2004b,a). Parmi les méthodes numériques qu’elle comporte, *Differential/algebraic system solver* (DASSL) (Petzold, 1982) a été choisie. Cette méthode est appropriée pour la résolution des systèmes implicites d’équations différentielles et algébriques et permet d’initialiser le modèle sans trop de difficulté. Notons cependant que pour une raison non identifiée, le modèle ne peut être initialisé en régime stationnaire. Chaque nouvel état stationnaire est obtenu en simulant l’évolution du système en régime transitoire jusqu’à ce qu’il se stabilise après la perturbation initiale, par exemple à une modification de la configuration de colmatage.

Les fonctions d’état (ou variables différentielles, du point de vue algorithmique) interviennent dans des gradients qui sont discrétisés sur les modules et les mailles des modules maillés. Aussi est-il plus commode d’utiliser des grandeurs définies sur l’ensemble du domaine géométrique. On évite donc les grandeurs n’étant définies que dans l’une des deux phases. Les candidats possibles sont donc la pression p , l’enthalpie massique de mélange \bar{h} , le débit massique \dot{M} , la masse volumique ρ et l’énergie interne e . Le modèle a été construit initialement à partir d’une bibliothèque de modules appelée EDFLib (Bouskela, 2003), devenue par la suite ThermoSysPro (El Hefni et coll., 2011; Bouskela, 2012). La plupart des modules de la version 7 du modèle ont été profondément modifiés ou réécrits entièrement, hormis la transcription en langage Modelica des tables thermodynamiques fondées sur l’*Industrial formulation 1997* (I.F. 97) de l’*International association for the properties of water and steam* (I.A.P.W.S.) (I.A.W.P.S., 2012) qui provient directement de la bibliothèque EDFLib. Ces tables permettent d’accéder faci-

lement à la masse volumique à partir de p et \bar{h} et fournissent également ses dérivées partielles. Compte tenu des hypothèses simplificatrices qui seront vues plus loin, l'énergie interne n'intervient pas en tant que telle dans les équations. Ainsi, les fonctions d'état choisies sont la pression p , l'enthalpie massique de mélange \bar{h} et le débit massique \dot{M} . On ne réalise pas explicitement la discrétisation temporelle, autrement dit, on sait écrire $\frac{\partial p}{\partial t}$ et $\frac{\partial \bar{h}}{\partial t}$ en Modelica. En revanche, la discrétisation spatiale apparaît explicitement dans les équations. Les gradients $\frac{\partial X}{\partial z}$ sont remplacés dans les modules maillés par des différences finies du type $X[n + 1] - X[n]$.

5.1.2. Schéma numérique

Le schéma numérique de type volume fini à mailles décalées de la bibliothèque EDFLib a été adopté (Bouskela, 2003). Le modèle est un assemblage monodimensionnel de volumes de contrôle au centre desquels on calcule les grandeurs intensives. Les flux sont quant à eux calculés aux bords des volumes de contrôle. Lors de l'écriture de l'équation de bilan d'énergie pour le fluide diphasique, on verra dans la section 5.3.3 que l'on fait apparaître, dans la dérivée temporelle, une enthalpie de mélange « pondérée en masse » et, dans le gradient, une enthalpie de mélange « pondérée en volume ». Dans le cas d'un fluide monophasique ou du modèle d'écoulement diphasique dit homogène, ces deux grandeurs sont physiquement équivalentes. Elles sont cependant distinctes lorsque l'on introduit une équation de déséquilibre cinématique et cette distinction intervient dans le schéma numérique. On note donc l'*enthalpie dynamique* pondérée en volume \bar{h} pour la distinguer de l'*enthalpie statique*, h , pondérée en masse. Ainsi, h et p sont calculés au centre des volumes de contrôle tandis que \dot{M} et \bar{h} sont calculés aux bords, le flux sous-jacent à \bar{h} étant $\dot{M} \bar{h}$.

Une des façons de conceptualiser le schéma numérique à mailles décalées est de considérer le modèle mono-dimensionnel comme un enchaînement d'éléments sans dimension. On distingue deux catégories de ces éléments qui doivent être alternés au sein du modèle :

- les volumes de contrôle qui contiennent les équations de bilan de masse et d'énergie ;
- les éléments biports qui contiennent les équations de bilan de quantité de mouvement.

Les éléments biports sont décalés par rapport aux volumes de contrôle de sorte que leurs centres coïncident avec les bords des volumes de contrôle. La figure 5.1 représente les deux *maillages décalés*. Les volumes de contrôle sont représentés en bleu, à gauche et les éléments biports en orange, à droite. Dans le modèle représenté sur la figure 5.1, on effectue un bilan de masse et un bilan d'énergie dans le volume de contrôle central représenté en trait plein et un bilan de quantité de mouvement dans chacun des deux éléments biports.

Le calcul des propriétés du fluide, par exemple de sa masse volumique, au bord d'un volume de contrôle non périphérique nécessite de connaître la pression à cet endroit. On utilisera la convention suivante (représentée en vert, ↗, sur la figure 5.1) : la pression au bord d'un volume de contrôle est obtenue par la moyenne de la pression dans ce volume de contrôle et de la pression dans le volume de contrôle qui partage ce bord.

L'essentiel du transport d'énergie thermique se fait par convection dans le sens du débit. Dans le modèle de g. v., on négligera toujours les transferts thermiques par conduction

Hyp.

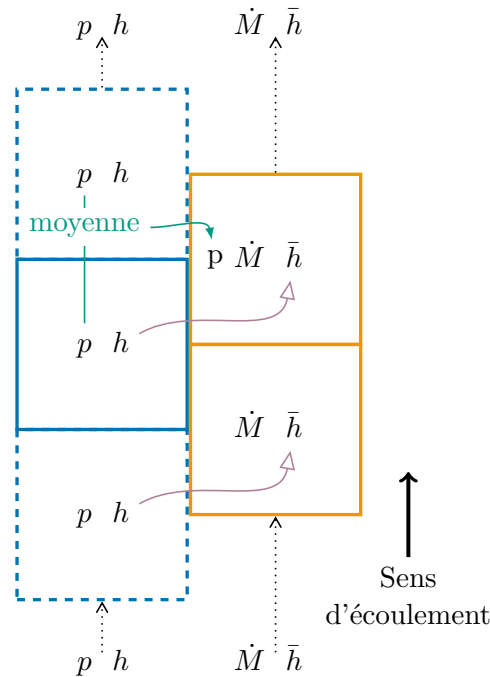


Figure 5.1. : Représentation graphique du schéma numérique à mailles décalées.

dans le sens du débit. Les seuls transferts thermiques par conduction pris en compte sont ceux entre le fluide primaire et le fluide secondaire, à travers la paroi des tubes. Ceux-ci sont représentés par un connecteur de flux entre les volumes de contrôle représentant le fluide primaire et le fluide secondaire, dont on a fait se correspondre les maillages pour éviter d'avoir à faire des interpolations. De ce fait, on peut en général identifier l'enthalpie pondérée en masse h et l'enthalpie pondérée en volume \bar{h} . On utilisera pour ce faire la convention suivante (représentée en mauve, \nearrow , sur la figure 5.1) : l'enthalpie de mélange pondérée en masse h au centre d'un volume de contrôle est égale à l'enthalpie de mélange pondérée en volume \bar{h} dans l'élément biport lui succédant, dans le sens du débit.

Les connecteurs entre les différents modules transmettent par des égalités algébriques les valeurs de p , h , \dot{M} et \bar{h} . La valeur de h est transmise afin de pouvoir gérer d'éventuelles inversions de débit (flèches pointillées, \nearrow , sur la figure 5.1). Le fait de choisir de transmettre \bar{h} au lieu de $\dot{M} \bar{h}$ est une précaution pour éviter les indéterminations en cas de débit nul.

5.1.3. Structure du modèle mathématique

Compte-tenu de la représentation géométrique adoptée pour le modèle de g. v., les principaux éléments à modéliser sont les suivants :

- un écoulement monophasique monodimensionnel (fluide primaire) ;
- un écoulement diphasique monodimensionnel (fluide secondaire) ;
- le transfert thermique entre les deux fluides avec la paroi des tubes comme interface ;

- des pertes de charge, éventuellement diphasiques, dans des géométries complexes (par exemple au passage des plaques entretoises) ;
- les dispositifs de séparation eau-vapeur de la partie haute du g. v. ;
- le contrôle-commande régulant le débit d'eau alimentaire.

Chaque fluide est modélisé par un jeu d'équations de conservation (de masse, d'énergie et de quantité de mouvement) et d'équations constitutives qui décrivent ses propriétés physiques thermodynamiques. On distingue les équations constitutives intrinsèques qui décrivent l'état du fluide et les propriétés thermodynamiques correspondantes des équations constitutives externes qui sont des formules empiriques que l'on appelle *corrélations*. Des corrélations sont utilisées notamment pour calculer les coefficients de transfert thermique aux interfaces des fluides et des parois des tubes ou encore le coefficient multiplicatif à appliquer au coefficient de perte de charge monophasique pour l'adapter au cas d'un fluide diphasique.

La section 5.2 détaille la modélisation des écoulements monophasiques utilisée notamment pour représenter le fluide primaire. Elle permet d'introduire les équations de conservations d'un écoulement monophasique qui serviront de point de départ pour la modélisation du fluide secondaire diphasique qui fait l'objet de la section 5.3. Enfin, le modèle de perte de charge lors de la traversée des plaques entretoises est détaillé dans la section 5.4. L'ajustement des autres coefficients de perte de charge sera abordé dans le chapitre suivant. La modélisation de la régulation n'a pas été modifiée lors de cette thèse et ne sera pas traitée.

5.2. Écoulement monophasique monodimensionnel : modélisation du fluide primaire

Le fluide circulant à l'intérieur des tubes du g. v. est de l'eau maintenue liquide par une pression de 155 bar. La gamme de températures atteintes à l'entrée du g. v., du côté chaud donc, lors d'un essai R.G.L. 4 est de 305 °C à 323 °C.

Le choix d'une représentation monodimensionnelle revient à postuler que la vitesse du fluide est dirigée selon la verticale et ne dépend que de la cote. On néglige la forme cintrée du haut des tubes ; du côté primaire, seule la surface d'échange thermique plus importante dans cette zone est prise en compte. De même, on admet que la pression et l'enthalpie ne dépendent que de la cote, soit :

$$\mathbf{v} = v_z(z,t)\boldsymbol{\delta}_z, \quad p = p(z) \quad \text{et} \quad h = h(z), \quad (5.1)$$

où $\boldsymbol{\delta}_z$ désigne le vecteur unitaire dirigé selon la verticale dans le sens ascendant. L'axe vertical est orienté dans le sens de l'écoulement. Par la suite on omettra les indices z sauf pour désigner la projection du champ de gravité \mathbf{g} selon la direction verticale, afin de rappeler que ce terme change de signe avec le sens de l'écoulement. De la même façon, le symbole τ désigne le seul terme non nul, (τ_{zz}) , du tenseur de contraintes visqueuses $\boldsymbol{\tau}$.

Hyp.

On rappelle la définition du nombre de Reynolds

$$\text{Re} \equiv \frac{\rho v D_c}{\mu}, \quad (5.2)$$

où D_c est une dimension caractéristique (ici le diamètre intérieur des tubes) et μ est la viscosité dynamique du fluide. Le nombre de Reynolds étant important ($\text{Re} \approx 10^6$), l'écoulement est turbulent. Les termes relatifs à la turbulence de l'écoulement dans les équations de conservation d'énergie seront négligés par la suite. L'effet de la turbulence sera pris en compte dans les équations constitutives par le biais d'un coefficient de friction.

Hyp.

5.2.1. Équations de conservation pour l'écoulement monophasique

La démarche rigoureuse pour obtenir les équations de conservation à partir de la formulation intégrale générale d'un bilan est donnée par Ishii et Hibiki (2006, chap. 2). On obtient à partir de là trois équations de conservation pour un écoulement monophasique (Bird et coll., 2006). L'équation dite *de continuité*,

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial z} = 0, \quad (5.3)$$

traduit la conservation de la masse dans le système.

L'équation de *conservation de quantité de mouvement* selon la direction verticale orientée dans le sens de l'écoulement s'écrit :

$$\frac{\partial \rho v}{\partial t} = -\frac{\partial}{\partial z} (\rho v^2 + p + \tau) + \rho g_z. \quad (5.4)$$

Le membre de gauche représente le taux d'accroissement de quantité de mouvement par unité de volume. Le premier terme de la différentielle en z représente l'apport de quantité de mouvement par convection par unité de volume tandis que les deux autres représentent l'apport par transport moléculaire. Le dernier terme représente les forces volumiques agissant sur le fluide, ici la gravité uniquement.

Enfin, l'équation de *conservation d'énergie* s'écrit :

$$\frac{\partial}{\partial t} \left(\frac{\rho v^2}{2} + \rho e \right) = -\frac{\partial}{\partial z} \left[\left(\frac{\rho v^2}{2} + \rho e \right) v - q'' + p v + \tau v \right] + \rho v g_z. \quad (5.5)$$

Le symbole e représente l'énergie interne par unité de masse et q'' le flux de chaleur au travers de la paroi vers le fluide. Le membre de gauche représente le taux d'accroissement d'énergie par unité de volume. Les quatre termes de la différentielle en z représentent dans l'ordre : *i.* l'apport d'énergie par transport convectif par unité de volume ; *ii.* l'apport d'énergie par conduction au travers de la paroi des tubes par unité de volume ; *iii.* le travail des forces de pression par unité de volume ; *iv.* le travail des forces de viscosité par unité de volume. Le dernier terme du membre de droite représente le travail de la gravité.

5.2.2. Formulation pratique des équations de conservation

Ce paragraphe détaille les calculs et hypothèses permettant d'aboutir, à partir des équations qui viennent d'être énoncées, à la formulation utilisée pour écrire le modèle en langage Modelica. En utilisant le débit massique par unité d'aire de la section de passage, $\dot{m} = \frac{M}{A} = \rho v$, l'équation de conservation de la masse s'écrit

$$\frac{\partial \rho}{\partial t} + \frac{\partial \dot{m}}{\partial z} = 0. \quad (5.6)$$

En multipliant l'équation (5.4) par v , on obtient, après réarrangement en utilisant l'équation de continuité, l'équation d'évolution de l'énergie cinétique :

$$\frac{\partial}{\partial t} \left(\frac{\rho v^2}{2} \right) = -\frac{\partial}{\partial z} \left(\frac{\rho v^2}{2} v \right) - \frac{\partial p v}{\partial z} - \left(-\frac{\partial v}{\partial z} \right) p - \frac{\partial \tau v}{\partial z} - \left(-\frac{\partial v}{\partial z} \right) \tau + \rho g_z v. \quad (5.7)$$

On formule alors l'hypothèse suivante, dont on se servira pour les équations constitutives internes : les définitions usuelles des grandeurs thermodynamiques et les équations d'état écrites pour un équilibre thermodynamique sont valables localement hors équilibre en régime dynamique. Hyp.

L'enthalpie par unité de masse h vérifie alors

$$h \equiv e + \frac{p}{\rho}. \quad (5.8)$$

Après soustraction de (5.7) à (5.5), cette équation permet d'obtenir l'équation d'évolution de l'enthalpie :

$$\frac{\partial \rho h}{\partial t} = -\frac{\partial \rho v h}{\partial z} + \frac{\partial q''}{\partial z} - \frac{\partial v}{\partial z} \tau + \frac{Dp}{Dt}. \quad (5.9)$$

On notera que le terme en p est dorénavant une dérivée particulaire : $\frac{Dp}{Dt} = \frac{\partial p}{\partial t} + v \frac{\partial p}{\partial z}$. Cependant, pour les transitoires considérés, le terme de convection $v \frac{\partial p}{\partial z}$ est négligeable devant la variation temporelle $\frac{\partial p}{\partial t}$ et l'on fera l'approximation $\frac{Dp}{Dt} \approx \frac{\partial p}{\partial t}$. Hyp.

Du fait de la dimensionnalité choisie, le tenseur des contraintes visqueuses est dégénéré. Dans l'équation d'évolution de l'enthalpie, on néglige simplement le terme de dissipation par viscosité. Dans l'équation du mouvement, on remplace le terme $\frac{\partial \tau}{\partial z}$ par $\frac{P_m \tau}{A}$ où P_m est le périmètre mouillé de la paroi et A l'aire de la section de passage. De même, on ne dispose pas du profil de température à l'échelle du volume de contrôle ; on remplace $\frac{\partial q''}{\partial z}$ par $\frac{P_c q''}{A}$ où P_c est le périmètre chauffé de la paroi. Hyp.

Concluons ce paragraphe par les formulations pratiques de l'équation de conservation de quantité de mouvement,

$$\frac{\partial \dot{m}}{\partial t} = -\frac{\partial}{\partial z} \left(\frac{\dot{m}^2}{\rho} \right) - \frac{\partial p}{\partial z} - \frac{P_m \tau}{A} + \rho g_z, \quad (5.10)$$

et de l'équation d'évolution de l'enthalpie,

$$\frac{\partial \rho h}{\partial t} = h \frac{\partial \dot{m}}{\partial z} - \frac{\partial \dot{m} h}{\partial z} + \frac{\partial p}{\partial t} + \frac{P_c q''}{A}. \quad (5.11)$$

Pour la transcription en langage Modelica, le premier membre est décomposé à l'aide de l'équation de continuité car la dérivé temporelle du produit ne s'écrit pas naturellement.

Deux termes n'ont pas été explicités dans les équations de conservations écrites ci-dessus : la contrainte imposée par la paroi, τ , dans l'équation du mouvement et la chaleur prélevée au travers de la paroi, q'' , dans l'équation d'évolution de l'enthalpie.

5.2.3. Équations constitutives de l'écoulement monophasique

Les équations d'état utilisées sont celles de l'eau liquide supposées être valides en régime dynamique. En pratique, on utilise les tables thermodynamiques I.F. 97 disponibles dans la bibliothèque EDFLib pour obtenir, à partir des grandeurs d'état, les valeurs et dérivées partielles de la masse volumique ρ , de la température du fluide T_f , de la viscosité dynamique μ , de la conductivité k et de la capacité calorifique C_p :

$$\rho = \rho(p, h), \quad T_f = T_f(p, h), \quad \mu = \mu(p, h), \quad k = k(p, h) \quad \text{et} \quad C_p = C_p(p, h). \quad (5.12)$$

Concernant le terme de viscosité, on cherche à connaître la perte de charge induite par les frictions avec la paroi sans se soucier pas du profil de vitesse compte-tenu de la dimensionnalité adoptée. On appelle *coefficient de friction*, f , le rapport entre la force F exercée par la paroi sur le fluide du fait de son mouvement et le produit d'une surface, A_c , et d'une énergie cinétique, K_c , caractéristiques (Bird et coll., 2006, chap. 6) :

$$f \equiv \frac{F}{A_c K_c}. \quad (5.13)$$

Pour les écoulements dans des tuyaux, le choix usuel de $\frac{\rho v^2}{2}$ comme terme d'énergie cinétique et de la surface mouillée comme surface caractéristique aboutit à l'expression suivante de la contrainte visqueuse :

$$\tau = f \frac{\rho v^2}{2}. \quad (5.14)$$

En général, les coefficients de friction peuvent s'exprimer simplement en fonction du nombre de Reynolds et des caractéristiques géométriques du système. Des séries d'expériences permettent d'établir des diagrammes illustrant la relation du coefficient de friction au nombre de Reynolds de l'écoulement ou bien des *corrélations empiriques*, c'est-à-dire des expressions analytiques ajustées sur les données d'expériences. Dans le cas du fluide primaire, il n'est pas utile d'avoir une description très précise du gradient de pression qui impacte peu le transfert thermique. En pratique, on se contente donc d'un coefficient de friction f constant dont on détermine la valeur par calage de la pression d'entrée sur les mesures des capteurs d'essais. La représentation du champ de pression du fluide secondaire est en revanche cruciale pour l'étude du colmatage et l'on verra dans le

paragraphe 5.3.4 que les lois de fermeture correspondantes ont été choisies avec plus de précautions.

De même que pour la friction, une approche empirique a été adoptée pour décrire le transfert thermique à l'interface du fluide et de la paroi. L'analogie du coefficient de friction est le *coefficient de transfert thermique local*, η , liant le flux de chaleur q'' à la différence de température entre la paroi et le fluide, $T_p - T_f$:

$$q'' = \eta (T_p - T_f) . \quad (5.15)$$

En général, le coefficient de transfert thermique s'exprime par une relation entre le nombre de Nusselt, Nu, et d'autres grandeurs sans dimension comme les nombres de Reynolds, Re, et de Prandtl, Pr. Rappelons l'expression du nombre de Nusselt,

$$\text{Nu} \equiv \eta \frac{D_c}{k} , \quad (5.16)$$

où D_c est une dimension caractéristique et k la conductivité thermique du fluide (Bird et coll., 2006, chap. 14), ainsi que celle du nombre de Prandtl,

$$\text{Pr} \equiv \frac{C_p \mu}{k} , \quad (5.17)$$

où C_p est la capacité calorifique du fluide à pression fixée (Bird et coll., 2006, chap. 9). Dans le cas du fluide primaire, on pose

$$\text{Nu} = \eta \frac{d_{int}}{k} \quad (5.18)$$

où d_{int} est le diamètre interne des tubes. De nombreuses corrélations sont disponibles dans littérature. On choisit ici l'une de celles proposées par V. GNIELINSKI (Rohsenow et coll., 1998, p. 5.26) pour un écoulement liquide ou gazeux dans une conduite circulaire lisse vérifiant $0,5 \leq \text{Pr} \leq 1,5$ et $10^4 \leq \text{Re} \leq 5 \cdot 10^6$

$$\text{Nu} = 0,0214 \left(\text{Re}^{0,8} - 100 \right) \text{Pr}^{0,4} . \quad (5.19)$$

L'utilisation de cette corrélation est légitime pour le fluide primaire pour lequel $\text{Pr} \approx 0,9$ et $\text{Re} \approx 10^6$.

5.3. Écoulement diphasique monodimensionnel : modélisation du fluide secondaire

Le fluide secondaire circulant à l'extérieur des tubes du g. v. est un mélange d'eau et de vapeur dont la proportion varie avec la cote et le point de fonctionnement du réacteur. Les écoulements diphasiques diffèrent des écoulements monophasiques par la présence d'une interface séparant les deux phases. Cette interface est mouvante et peut prendre de nombreuses configurations, parfois très dissemblables, caractérisant différents

régimes d'écoulements. Les écoulements monophasiques sont eux-mêmes décrits par des régimes d'écoulement, par exemple suivant leur degré de turbulence. Ces distinctions restent pertinentes pour l'étude des écoulements diphasiques mais ces derniers sont par ailleurs profondément influencés par la répartition spatiale des deux phases. Collier et Thome (1996, chap. 1) distinguent cinq principaux régimes d'écoulement d'un gaz et d'un liquide à co-courant dans une conduite verticale cylindrique. Les frontières entre ces régimes sont parfois ténues et l'on en trouve dans la littérature de nombreuses variantes et sous-catégories. Dans un écoulement à bulles, par exemple, le gaz est réparti en une multitude de bulles discrètes réparties dans la phase liquide continue. Les bulles peuvent être de très petites tailles et sphériques comme dans une coupe de champagne ou bien plus grosses, en forme de coupoles, comme l'air expiré vers la surface par un plongeur. On conçoit aisément qu'un tel régime aura des propriétés très différentes d'un écoulement annulaire dans lequel l'essentiel du tube est occupé par une phase continue gazeuse tandis qu'un film de liquide recouvre les parois. Le mécanisme de transfert thermique à l'interface du fluide et des parois ne sera par exemple pas le même. Le régime d'écoulement dépend de la géométrie du système, des valeurs locales de paramètres thermodynamiques comme la proportion de gaz et de liquide et d'autres paramètres plus difficilement identifiables comme l'écart à l'équilibre thermodynamique ou la présence de contaminants à l'état de traces.

L'observation directe d'un régime d'écoulement nécessite des parois translucides, ce qui réduit la gamme de flux thermiques possibles. Des techniques de mesures avancées sont mises en œuvre pour tâcher d'observer les régimes d'écoulement lors d'expériences plus proches des situations rencontrées en réalité. L'éventail des techniques actuel est vaste, allant des sondes locales électriques ou optiques aux méthodes de tomographie de rayons X, gamma ou de positrons en passant par des capteurs constitués de grillages d'électrodes (Prasser, 2011). La caractérisation détaillée des régimes d'écoulement dans des systèmes de la complexité et l'envergure d'un g. v. est cependant encore hors de portée.

L'interface entre les phases est dynamique. Les bulles d'un écoulement à bulles apparaissent lors de la vaporisation du liquide, se déplacent rapidement, fusionnent par coalescence ou se résorbent. Le film liquide d'un écoulement annulaire est quant à lui agité de vagues d'amplitude importante qui émettent de fines gouttelettes emportées par le flux gazeux central. De ce fait, les écoulements diphasiques sont en général des systèmes essentiellement hors équilibre, cinématique et thermique. Certains modèles utilisent des méthodes dite de *suivi d'interface* permettant de tenir compte explicitement du comportement de l'interface. Un seul jeu d'équation de conservation décrit alors l'ensemble des phases mais les propriétés des fluides présentent des discontinuités à la traversée de l'interface. Elles sont typiquement mises en œuvre à l'aide des logiciels de simulation numérique de la dynamique des fluides diphasiques évoqués en introduction du chapitre précédent. Cette approche ne permet à l'heure actuelle de traiter que des systèmes de faible étendue et de géométrie simple mais son domaine d'application s'étend progressivement. Ainsi, l'approche dominante demeure celle dite à *milieux interpénétrants*. La forme exacte de l'interface échappant à notre capacité descriptive, on l'interprète en termes statistiques, à l'aide de moyennes temporelles et éventuellement spatiales des grandeurs physiques. Chaque phase est supposée présente en tout point avec une certaine

probabilité, d'où le vocable d'« inter-pénétrant ». Les différents opérateurs de moyenne utilisés font l'objet de la section suivante.

L'application directe du concept de milieux interpénétrants aboutit au modèle à *deux fluides*. Chaque phase est décrite par un jeu d'équations de conservation pour les grandeurs moyennes. Trois lois de fermeture *interfaciales* doivent être ajoutées pour représenter les échanges de masse, de quantité de mouvement et d'énergie entre les phases. Cette formulation est adaptée pour décrire des transitoires brusques (Yadigaroglu, 2011a). Les transitoires plus lents peuvent être abordés à l'aide d'une formulation plus maniable obtenue en sommant les équations de conservation des deux phases. On obtient ainsi un modèle *de mélange* dans lequel les lois de transferts interfaciaux sont remplacées par des lois donnant par exemple le taux de vide en fonction du titre ou bien par des hypothèses sur l'équilibre thermique d'une des phases.

Idealement, si les régimes d'écoulement sont connus, on choisira des lois de fermeture leur étant spécialement adaptés. Cette approche qualifiée de *phénoménologique* (Hewitt, 2011) est performante pourvu que l'on puisse déterminer avec certitude le régime d'écoulement. Il faut pour cela disposer de mesures détaillées qui font défaut dans le cas du g. v. Le faisceau de tubes et les plaques entretoises constituent une géométrie particulièrement complexe qui sort du cadre théorique permettant de prédire les régimes d'écoulement. Ainsi, c'est l'approche *empirique*, utilisée pour la plupart des modèles présentés dans le chapitre précédent, qui a été mise en œuvre ici. Les lois de fermeture utilisées sont des corrélations empiriques supposées valides pour tous les régimes d'écoulement. Si les grandeurs intervenant dans une corrélation sont celles qui gouvernent les changements de régime, celle-ci sera à même de les représenter implicitement. Les erreurs commises par ces corrélations sont parfois importantes, en particulier lorsqu'elles sont appliquées à des systèmes éloignés de l'expérience dont elles sont issues. Une procédure de validation du modèle est donc indispensable ; ce sera l'objet du chapitre suivant.

5.3.1. Opérateurs de moyenne

L'interface entre les phases d'un écoulement diphasique se déplace et se déforme rapidement, empêchant ainsi l'écriture simple d'équations de conservation locales comme cela a été fait dans la section 5.2 pour l'écoulement monophasique. Afin de contourner cette difficulté, on suppose que la phase présente en un point donné alterne aléatoirement et l'on opère une moyenne temporelle afin de lisser ces fluctuations. De même que pour un écoulement monophasique, l'opérateur de moyenne temporelle élimine les fluctuations dues à la turbulence. Conjointement, il transforme aussi les deux phases discrètes et mouvantes en phases continues présentes partout dans l'écoulement, avec une certaine probabilité. Étant donné une grandeur locale quelconque λ , sa *moyenne temporelle* est

$$\hat{\lambda} = \frac{1}{\Delta t} \int_{\Delta t} \lambda dt, \quad (5.20)$$

où Δt est le temps d'observation.

Soit ϕ_g l'indicatrice de la phase vapeur, c'est-à-dire la variable valant 1 en présence de vapeur et 0 en présence de liquide. L'application de l'opérateur de moyenne temporelle à

l'indicatrice de la vapeur définit la *fraction moyenne locale de gaz* :

$$\alpha_g \equiv \widehat{\phi}_g = \frac{1}{\Delta t} \int_{\Delta t} \phi_g dt, \quad (5.21)$$

que l'on désignera par la suite simplement par le « taux de vide », noté α . Le complémentaire du taux de vide est la fraction moyenne locale de liquide, α_l . L'opérateur de moyenne temporelle n'interviendra dans la suite qu'implicitement, par le biais du taux de vide.

Un second opérateur de moyenne, spatiale cette fois-ci, permet d'écrire les équations de conservation globalement pour une section du canal. La moyenne transversale d'une grandeur λ sur une section du canal d'aire A est

$$\langle \lambda \rangle = \frac{1}{A} \int_A \lambda dA. \quad (5.22)$$

Ainsi, la moyenne transversale $\langle \alpha \rangle$ du taux de vide est la proportion moyenne pendant la durée d'observation de l'aire de la section transversale occupée par la phase gazeuse. Certaines grandeurs, par exemple la vitesse de la vapeur, ne sont définies que dans l'une des deux phases. Elles seront indicées g pour la phase gazeuse et l pour la phase liquide. On définit, de même qu'auparavant, des moyennes transversales *par phase*. Pour la phase gazeuse par exemple, la moyenne transversale en phase gazeuse de la grandeur λ_g est

$$\langle \lambda_g \rangle_g = \frac{1}{A_g} \int_{A_g} \lambda_g dA, \quad (5.23)$$

où A_g désigne l'aire occupée par la vapeur dans une section transversale du canal.

Un calcul simple permet de montrer que toute grandeur λ_k associée à une phase indicée k vérifie la propriété suivante :

$$\langle \alpha_k \lambda_k \rangle = \langle \alpha_k \rangle \langle \lambda_k \rangle_k. \quad (5.24)$$

5.3.2. Taux de vide, titre et rapport de vitesses des phases

Ce paragraphe introduit les notions et notations nécessaires pour adapter les équations de conservation écrites pour l'écoulement monophasique à l'écoulement diphasique. Il reprend la formulation proposée par Yadigaroglu (2011a) pour une phase liquide et une phase gazeuse (la vapeur) s'écoulant dans une conduite.

On définit le *titre du mélange* comme le rapport du débit massique de vapeur par le débit massique total :

$$\chi \equiv \frac{\dot{M}_g}{\dot{M}}. \quad (5.25)$$

Il s'agit du titre *dynamique*, par opposition au titre *statique* défini comme le rapport des masses. Dans le g. v., il atteint une valeur maximale d'environ 0,5 dans les cintres en jambe chaude, en régime nominal.

Le débit massique est le flux de masse au travers d'une section transversale, soit $\dot{M} = \rho v A = \dot{m} A$. Appliquée à une phase indiquée k , cette formule permet d'exprimer la moyenne transversale de la vitesse de la phase $\langle v_k \rangle_k$ en fonction du débit massique des phases :

$$\langle v_k \rangle_k = \frac{\dot{M}_k}{\rho_k A_k} = \frac{\dot{M}_k}{\rho_k \langle \alpha_k \rangle A}. \quad (5.26)$$

Cette équation peut être reformulée avec le flux de masse de la phase liquide, $\dot{m}_l = \dot{m}(1 - \chi)$, et celui de la phase gazeuse, $\dot{m}_g = \dot{m}\chi$:

$$\langle v_l \rangle_l = \frac{\dot{m}_l}{\rho_l \langle \alpha_l \rangle} = \frac{\dot{m}(1 - \chi)}{\rho_l \langle \alpha_l \rangle} \quad (5.27)$$

et

$$\langle v_g \rangle_g = \frac{\dot{m}_g}{\rho_g \langle \alpha_g \rangle} = \frac{\dot{m}\chi}{\rho_g \langle \alpha_g \rangle}. \quad (5.28)$$

Le flux de masse total, \dot{m} est la somme des flux de masse des phases gazeuse, $\dot{m}_g = \dot{m}\chi$ et liquide, $\dot{m}_l = \dot{m}(1 - \chi)$, soit :

$$\dot{m} = \dot{m}_l + \dot{m}_g \quad (5.29)$$

$$= \rho_g \langle v_g \rangle_g \langle \alpha \rangle + \rho_l \langle v_l \rangle_l \langle 1 - \alpha \rangle. \quad (5.30)$$

Concluons ce paragraphe en introduisant le *rapport des vitesses de phases*, S , aussi appelé *glissement*, qui sera central dans le modèle de déséquilibre cinématique exposé dans la section 5.3.4 :

$$S \equiv \frac{\langle v_g \rangle_g}{\langle v_l \rangle_l}. \quad (5.31)$$

Les équations (5.28) et (5.27) permettent de lier par une équation unique le taux de vide, le titre dynamique et le rapport des vitesses des phases :

$$S = \frac{\rho_l}{\rho_g} \frac{\chi}{1 - \chi} \frac{\langle 1 - \alpha \rangle}{\langle \alpha \rangle}. \quad (5.32)$$

5.3.3. Équations de conservation pour l'écoulement diphasique

L'écriture rigoureuse des équations de conservation d'un écoulement diphasique dans le cas général est complexe et nécessite de longs calculs. On se contente ici du développement proposé par Yadigaroglu (2011a), que l'on trouvera formulé de façon analogue par Collier et Thome (1996, chap. 2) et Ghiaasiaan (2008, chap. 5), dans lequel les deux phases sont supposées s'écouler séparément, comme c'est le cas par exemple du régime annulaire. Néanmoins, Le système d'équations obtenu finalement ne diffère que peu de celui auquel aboutit l'approche rigoureuse exposée par exemple par Ishii et Hibiki (2006, partie II). Le

canal dans lequel s'écoule le fluide est supposé de section constante ce qui est inexact dans la partie cintrée de l'échangeur. Cette approximation est jugée secondaire par rapport au choix de la dimensionnalité qui conduit à négliger les phénomènes transverses qui se déroulent dans cette partie. Les trois équations de bilan suivantes sont obtenues de façon analogue aux équations (5.6), (5.10) et (5.11). Les principales distinctions sont la présence explicite de l'opérateur de moyenne spatiale. Pour chacune des phases, l'équation de continuité s'écrit :

$$\frac{\partial}{\partial t} \langle \rho_k \alpha_k \rangle + \frac{\partial \dot{m}}{\partial z} = \Gamma_k, \quad (5.33)$$

où Γ_k est le débit entrant dans la phase indiquée k par unité de volume. L'interface étant de volume nul, elle ne stocke aucune masse et l'on peut écrire la relation suivante appelée *condition de saut* :

$$\Gamma_l + \Gamma_g = 0. \quad (5.34)$$

L'écriture de l'équation de quantité de mouvement est simplifiée significativement en faisant l'hypothèse que la pression est uniforme dans une section transversale du canal et égale dans les deux phases. Cette hypothèse est difficilement testable. L'observation des résultats de simulation réalisés avec THYC tend à la conforter. Pour la phase indiquée k , la conservation de la quantité de mouvement s'écrit

Hyp.

$$\frac{\partial}{\partial t} \langle \rho_k \alpha_k v_k \rangle = - \frac{\partial}{\partial z} \left(\langle \rho_k \alpha_k v_k^2 \rangle + p \right) + \langle \rho_k \alpha_k \rangle g_z - \frac{P_m \tau}{A} - \frac{P_i \tau_{ik}}{A} + \Gamma_k v_i. \quad (5.35)$$

Les deux derniers termes de cette équation traduisent des transferts au travers de l'interface. Le premier de ces termes est analogue au terme de friction contre la paroi. Il représente l'effet des contraintes visqueuses subies par les phases à leur interface, lorsqu'elles se déplacent l'une par rapport à l'autre et P_i est le périmètre de l'interface. De même que pour l'équation de continuité, on peut écrire la condition de saut suivante :

$$\tau_{ig} + \tau_{il} = 0. \quad (5.36)$$

Le second terme nouveau représente le transfert de quantité de mouvement par convection d'une phase à l'autre. Il est caractérisé par la vitesse v_i de l'interface.

L'équation d'évolution d'enthalpie s'écrit enfin

$$\frac{\partial}{\partial t} \langle \rho_k \alpha_k h_k \rangle = - \frac{\partial}{\partial z} \langle \rho_k \alpha_k h_k v_k \rangle + \langle \alpha_k \rangle \frac{\partial p}{\partial t} + \frac{P_c q''}{A} + \frac{P_i q''_{ik}}{A} + \Gamma_k h_k. \quad (5.37)$$

Celle-ci comporte elle aussi deux termes spécifiques aux écoulements diphasiques. Le premier est analogue au terme de transfert thermique au travers de la paroi. Il est proportionnel au flux thermique de l'interface vers la phase indiquée k , noté q''_{ik} . Le second représente le transfert thermique par convection au travers de l'interface. Comme pour l'écoulement monophasique, on néglige la dissipation d'énergie thermique d'origine visqueuse. Sans cette hypothèse, un terme supplémentaire assorti d'une condition de saut décrirait la répartition de cette énergie entre les deux phases.

Simplification des moyennes transversales de produits

Les équations (5.33), (5.35) et (5.37) contiennent des moyennes transversales de produits de deux, trois ou quatre grandeurs. On ne peut « ouvrir » ces moyennes de produits, c'est-à-dire les remplacer par des produits de moyennes, qu'à condition de connaître la distribution des grandeurs dans une section de canal ce qui sort bien entendu du cadre de la représentation monodimensionnelle. On a vu dans la section 5.3.1 que les produits de deux termes peuvent être traités à l'aide de l'équation (5.24). Les triples produits peuvent être ramenés à des doubles produits, puis ouvert par cette méthode, en sortant de l'opérateur les masses volumiques des phases. Cela revient à supposer que ces dernières sont relativement uniformes dans une section du canal. Cette hypothèse ne pose pas de problème pour le liquide, peu compressible et dont la masse volumique varie très graduellement. La masse volumique de la vapeur est en revanche beaucoup plus sensible à la température. La distinction entre jambe chaude et jambe froide rend cependant l'hypothèse acceptable en séparant en deux jeux d'équations distincts les régions de masse volumique très différente.

Hyp.

Les moyennes de produits de quatre termes, tels $\langle \rho_k \alpha_k v_k^2 \rangle$ ou $\langle \rho_k \alpha_k h_k v_k \rangle$ sont plus problématiques. Le recours usuellement employé est d'ouvrir les moyennes de produits et de leur adjoindre des coefficients de correction multiplicatif. Faute de connaissance sur la répartition transversale de l'enthalpie et des vitesses, on supposera ces coefficients égaux à 1.

Hyp.

Modèle de mélange

Les termes relatifs à l'interface dans les équations (5.33), (5.35) et (5.37) nécessitent de nombreuses lois de fermeture, difficiles à établir sans connaissance du régime d'écoulement. Le *modèle de mélange* est obtenu en sommant les équations écrites pour la phase vapeur et la phase liquide. La *masse volumique de mélange* est une moyenne des masses volumiques des phases pondérée par le taux de vide ,

$$\langle \rho \rangle \equiv \rho_l \langle 1 - \alpha \rangle + \rho_g \langle \alpha \rangle , \quad (5.38)$$

de même que le flux de masse du mélange, \dot{m} , défini par l'équation (5.30) comme une moyenne de flux de masse. La condition de saut permet d'écrire l'équation de continuité dans sa formulation pratique :

$$\frac{\partial \langle \rho \rangle}{\partial t} + \frac{\partial \dot{m}}{\partial z} = 0 . \quad (5.39)$$

Les flux thermiques et les transferts thermiques par convection interfaciaux sont liés par la condition de saut suivante :

$$\Gamma (h_{il} + h_{ig}) + \frac{P_i}{A} q''_{il} + q''_{ig} = 0 . \quad (5.40)$$

Ainsi, la conservation de quantité de mouvement se traduit par

$$\frac{\partial \dot{m}}{\partial t} = -\frac{\partial}{\partial z} \left(\rho_l \langle 1 - \alpha \rangle \langle v_l \rangle_i^2 + \rho_g \langle \alpha \rangle \langle v_g \rangle_g^2 \right) - \frac{\partial p}{\partial z} + \langle \rho \rangle g_z - \frac{P_m \tau}{A}. \quad (5.41)$$

L'équation de conservation d'énergie fait intervenir deux *enthalpies de mélange*. L'une, dans le terme d'évolution temporelle, est pondérée *en masse* :

$$h \equiv (1 - \chi) \langle h_l \rangle_i + \chi \langle h_g \rangle_g. \quad (5.42)$$

L'autre, dans le terme de convection, est pondérée *en volume* :

$$\bar{h} \equiv \frac{\rho_l \langle 1 - \alpha \rangle \langle h_l \rangle_i + \rho_g \langle \alpha \rangle \langle h_g \rangle_g}{\langle \rho \rangle}. \quad (5.43)$$

À l'aide de ces deux définitions, l'équation d'évolution de l'enthalpie de l'écoulement diphasique s'écrit :

$$\frac{\partial}{\partial t} \left(\langle \rho \bar{h} \rangle \right) = \frac{\partial \dot{m} h}{\partial z} + \frac{\partial p}{\partial t} + \frac{P_c q''}{A}. \quad (5.44)$$

5.3.4. Équations constitutives de l'écoulement diphasique

Une partie des équations de fermeture du système qui vient d'être constitué est analogue à celles écrites pour l'écoulement primaire. Les tables I.F. 97 sont à nouveau utilisées et l'on choisit des corrélations décrivant les frictions et le transfert thermique à la paroi. Des équations de fermeture additionnelles sont nécessaires pour pouvoir calculer les termes de transfert à l'interface qui apparaissent dans les équations de conservation du modèle à deux fluides. L'écriture du modèle de mélange permet d'éliminer ces termes mais supprime par la même occasion trois équations de conservation. Le problème de fermeture est en quelque sorte déplacé de la description des échanges interfaciaux à la caractérisation des déséquilibres thermique et cinématique entre phases. Le tableau 5.1 résume ce compromis entre degrés de liberté du modèle et lois de fermeture.

La formulation la plus simple du modèle de mélange, le *modèle homogène*, avait été adoptée dans la version 2 du modèle de g. v. Dans ce modèle, les deux phases ont la même vitesse ce qui permet, en factorisant les termes de vitesse, d'écrire les équations de conservation comme pour un écoulement monophasique. On obtient notamment l'identité $h = \bar{h}$. Les deux phases étaient par ailleurs supposées à saturation : il y avait équilibre cinématique et thermique. Afin d'ajouter un degré de liberté pour tâcher de mieux reproduire la forme des courbes de réponses de n. g. l., l'hypothèse d'équilibre cinématique a été relâchée dans la version 7 à l'aide d'un modèle présenté à la fin de cette section. L'hypothèse d'équilibre thermique a en revanche été conservée.

Hyp.

Friction

L'une des approches couramment utilisée pour estimer le coefficient de friction associé à un écoulement diphasique est de lier le gradient de pression du mélange diphasique, $\left. \frac{\Delta p}{\Delta z} \right|_{lg}$ (*lg* pour *liquide et gaz*), à celui qu'aurait le même écoulement s'il n'était

	Modèle à deux fluides	Modèle de mélange
Équations de conservations	2 × masse	1 × masse
	2 × mouvement	1 × mouvement
	2 × énergie	1 × énergie
Lois de fermeture	Γ	–
	τ_l, τ_g, τ_i	τ
	q_l'', q_g'', q_i''	q''
Contraintes	–	Relation entre les vitesses
	–	Contrainte de l'énergie des deux phases
Variables libres	p, α	p
	$\langle v_l \rangle_l, \langle v_g \rangle_g$	$\langle v_l \rangle_l, \langle v_g \rangle_g$
	$\langle h_l \rangle_l, \langle h_g \rangle_g$	–

Tableau 5.1. : Comparaison entre le modèle à deux fluides et le modèle de mélange utilisé pour représenter le g. v., du point de vue du compromis entre degrés de liberté et lois de fermeture.

constitué que de liquide, $\left. \frac{\Delta p}{\Delta z} \right|_{ls}$ (*ls* pour *liquide seul*). On définit ainsi un multiplicateur diphasique, Φ_{ls}^2 :

$$\Phi_{ls}^2 = \frac{\left. \frac{\Delta p}{\Delta z} \right|_{lg}}{\left. \frac{\Delta p}{\Delta z} \right|_{ls}}. \quad (5.45)$$

Partant de l'équation (5.14), la contrainte visqueuse s'écrit :

$$\tau = \Phi_{ls}^2 f \frac{\rho v^2}{2}. \quad (5.46)$$

Suivant les recommandations formulées par Pierotti (1990), l'ajustement suivant, établi par A. E. RUFFEL à partir de mesures réalisées par J. R. S. THOM, a été utilisée :

$$\Phi_{ls}^2 = 1 + \frac{4200\chi}{(p + 19) \exp\left(\frac{p}{84}\right)}. \quad (5.47)$$

Cette corrélation donne des résultats satisfaisants sur une gamme de pression allant de 17 à 200 bar, encadrant celle du *riser* comprise entre 60 et 64 bar pendant un essai R.G.L. 4. Elle présente par ailleurs l'avantage de sa simplicité qui permet d'éviter certains couplages alourdissant la résolution numérique du système d'équations, facilitant ainsi l'initialisation du modèle.

Pour cette même raison, la corrélation suivante, proposée par FILONENKO et A. D. AL'TSUL (Idel'cik, 1969, p. 58 et diagramme 2.4), a été utilisé pour calculer le coefficient de friction monophasique :

$$f_{ls} = \frac{1}{(1,8 \log(\text{Re}_{ls}) - 1,64)^2}. \quad (5.48)$$

Cette corrélation est valide pour un nombre de Reynolds compris entre 10^4 et $5 \cdot 10^6$, celui de l'écoulement dans le *riser* étant de l'ordre de $3 \cdot 10^5$.

Dans la partie cintrée, on fait l'approximation que l'écoulement est perpendiculaire aux tubes. Les frottements sont représentés par une corrélation décrivant la traversée d'un faisceau de tubes à pas carré due à MOCAN et REVSINA (Idel'cik, 1969, p. 312 et diagramme 8.11) :

$$f_{ls} = 1,52 \left(\frac{pas}{d_{ext}} - 1 \right)^{-0,5} \text{Re}_l^{-0,2} n_{rang}, \quad (5.49)$$

où Re_l est le nombre de Reynolds de la phase liquide (compris environ entre $4 \cdot 10^4$ et $9 \cdot 10^4$ dans la partie cintrée en régime nominal) et n_{rang} le nombre moyen de rangs de tubes du faisceau.

Transfert thermique

Le transfert thermique est traité différemment selon que le fluide secondaire est effectivement diphasique ou à l'état liquide, en bas du g. v. La paire de corrélations adoptée est l'une de celle ayant donné satisfaction dans le modèle THYC (Guelfi et Pitot, 2004).

Le coefficient de transfert thermique dans la partie monophasique est évalué à l'aide d'une corrélation répandue due à F. W. DITTUS et L. M. K. BOELTER (Rohsenow et coll., 1998, p. 5.26) :

$$\eta_{db} = \frac{0,023k}{d_{th}} \text{Re}^{0,8} \text{Pr}^{0,4}, \quad (5.50)$$

où d_{th} est le diamètre thermique du canal valant 24,19 mm. Cette corrélation est valide pour les écoulements dans des conduites dont la longueur est au moins soixante fois supérieure au diamètre hydraulique, dont le nombre de Reynolds est compris entre 2500 et $1,24 \cdot 10^5$ et le nombre de Prandtl entre 0,7 et 120. La plupart des corrélations sont établies pour des écoulements dans des tubes et sont accompagnées de définitions des longueurs caractéristiques associées à cette géométrie. Lorsqu'elles sont utilisées dans des géométries différentes, ces définitions doivent être adaptées ce qui n'est pas nécessairement trivial. Dans sa revue bibliographique, Décossin (2002) signale l'absence de consensus à ce sujet. Ici, les diamètres thermique et hydraulique ont été identifiés et sont tous deux calculés de la façon suivante :

$$d_{th} = d_{hydro} \equiv 4 \frac{pas^2 - \frac{\pi d_{ext}^2}{4}}{4(pas - d_{ext}) + \pi d_{ext}}, \quad (5.51)$$

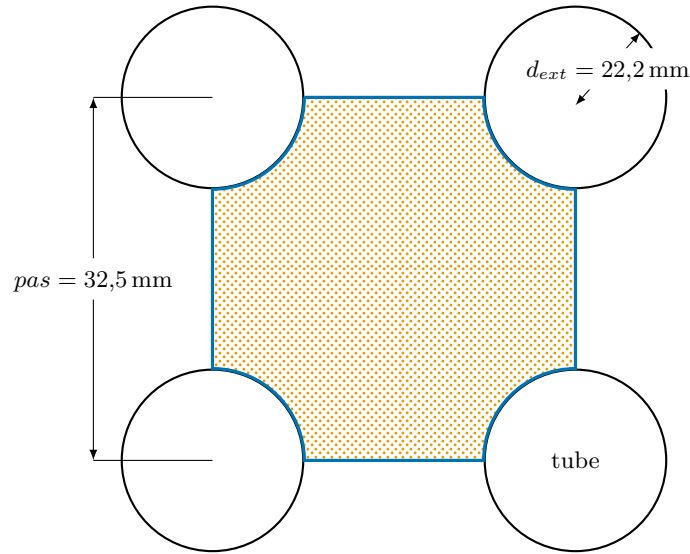


Figure 5.2. : Schéma d'un sous-canal de l'écoulement secondaire diphasique. Le rayon hydraulique est égal au rapport de la surface marquée par des points orange par le périmètre en trait épais bleu. Le diamètre hydraulique vaut 4 fois le rayon hydraulique.

où pas désigne le pas du faisceau de tubes, valant 32,54 mm, et d_{ext} le diamètre externe des tubes valant 22,20 mm. Ces dimensions sont représentées sur la figure 5.2 représentant un sous-canal de l'écoulement secondaire diphasique et sont données dans le tableau A.1 page 233. Le diamètre hydraulique vaut 24,2 mm. Aussi, même en ne considérant comme longueur de canal que la longueur en phase liquide en régime nominal dans la jambe chaude (2,5 m), la première condition de validité est largement vérifiée, le rapport valant environ 100. Toujours en régime nominal considérant la partie monophasique du g. v., le nombre de Reynolds est compris entre $7,82 \cdot 10^4$ et $5,03 \cdot 10^5$ et le nombre de Prandtl entre 0,71 et 0,84. La condition sur le nombre de Reynolds est dépassée d'un facteur 4 environ. Compte-tenu de l'incertitude associé à la géométrie du sous-canal, cela n'a pas été jugé rédhibitoire. Le transfert thermique dans l'écoulement diphasique a néanmoins été jugé l'un des éléments incertains de la modélisation, avec le modèle de déséquilibre cinématique. Ce point sera développé dans le chapitre suivant.

À partir de la cote où apparaissent les premières bulles de vapeur, une corrélation adaptée par Westinghouse de celle proposée par W. H. JENS et P. A. LOTTES (Guelfi et Pitot, 2004, p. 63) a été choisie :

$$\eta_{js} = 57,59 \exp\left(\frac{p}{15,5}\right) (T_p - T_{sat})^3, \quad (5.52)$$

où T_{sat} est la température de saturation du fluide et T_p la température de la paroi. La formule initiale de W. H. JENS et P. A. LOTTES ne s'applique qu'en régime d'ébullition saturée ou sur-saturée. Cette formule adaptée est appliquée au domaine sous-saturé en

conservant la forme du coefficient d'échange. Une variante est possible en conservant la forme du flux thermique (Guelfi et Pitot, 2004). Cette corrélation a été établie spécialement pour les g. v. et a donné satisfaction dans le modèle THYC.

On considère en pratique que l'ébullition sous-saturée accroît nettement le flux thermique vers le fluide secondaire et le coefficient de transfert thermique utilisé dans l'ensemble du *riser* est le maximum de ceux donnés par les deux corrélations précédentes :

$$\eta = \max(\eta_{db}, \eta_{js}). \quad (5.53)$$

Ainsi, la cote de basculement de la corrélation de quidam[Dittus et BOELTER vers la corrélation de JENS et LOTTES est généralement située légèrement en dessous de la cote de vaporisation.

Déséquilibre cinématique : le modèle de flux de dérive

Le modèle de *flux de dérive* (*drift flux*, en anglais) proposé par N. ZUBER et J. A. FINDLAY a été choisi pour décrire le déséquilibre cinématique (Yadigaroglu, 2011a). Les *débits volumiques* de vapeur et de liquide s'expriment en fonction des débits massiques :

$$Q_l \equiv \frac{\dot{M}_l}{\rho_l} \quad \text{et} \quad Q_g \equiv \frac{\dot{M}_g}{\rho_g}. \quad (5.54)$$

Les flux associés,

$$\langle j_l \rangle = \frac{Q_l}{A} \quad \text{et} \quad \langle j_g \rangle = \frac{Q_g}{A}, \quad (5.55)$$

peuvent s'exprimer en fonction du taux de vide, ou bien du titre :

$$\langle j_l \rangle = \langle v_g \rangle_g \langle 1 - \alpha \rangle = \frac{\dot{m}(1 - \chi)}{\rho}, \quad (5.56)$$

$$\langle j_g \rangle = \langle v_g \rangle_g \langle \alpha \rangle = \frac{\dot{m}\chi}{\rho}. \quad (5.57)$$

Ces flux sont homogènes à des vitesses et sont aussi appelés *vitesses superficielles*. De fait, ce sont les vitesses qu'auraient chacune des phases *si elles s'écoulaient seules dans le canal*. Ainsi, on définit les vitesses superficielles *locales* par :

$$j_l \equiv v_l(1 - \alpha) \quad \text{et} \quad j_g \equiv v_g\alpha. \quad (5.58)$$

Le flux volumique total, $\langle j \rangle = \langle j_l \rangle + \langle j_g \rangle$, est la vitesse moyenne à laquelle se déplacent les deux phases. Autrement dit, un plan se déplaçant à la vitesse $\langle j \rangle$ dans le sens de l'écoulement est traversé par des volumes identiques de liquide et de vapeur, en sens opposés. Sa version locale, j , permet de définir la *vitesse locale de dérive du gaz* :

$$v_{gj} \equiv v_g - j. \quad (5.59)$$

En multipliant cette équation par le taux de vide et en intégrant sur une section de canal, on obtient l'identité suivante :

$$\langle \alpha v_{gj} \rangle = \langle \alpha v \rangle - \langle \alpha j \rangle . \quad (5.60)$$

Seul le premier terme du membre de droite peut être ouvert à l'aide de l'équation (5.24) :

$$\langle \alpha v_{gj} \rangle = \langle \alpha \rangle \langle v_g \rangle_g - \langle \alpha j \rangle . \quad (5.61)$$

Le principe du modèle de flux de dérive est de modéliser la distribution du champ de vitesse pour ouvrir les autres termes. Le *coefficient de distribution*,

$$C_0 \equiv \frac{\langle \alpha j \rangle}{\langle \alpha \rangle \langle j \rangle} , \quad (5.62)$$

rend compte des disparités des répartitions transversales de la vitesse et du taux de vide. La *vitesse moyenne de dérive du gaz*,

$$\tilde{v}_{gj} \equiv \frac{\langle \alpha v_{gj} \rangle}{\langle \alpha \rangle} = \langle v_{gj} \rangle_g , \quad (5.63)$$

est quant à elle une moyenne pondérée en taux de vide de la vitesse de dérive.

On tire alors des définitions (5.62) et (5.63) une expression de la vitesse de la phase gazeuse :

$$\langle v_g \rangle_g = C_0 \langle j \rangle + \tilde{v}_{gj} . \quad (5.64)$$

Partant de là, la vitesse de la phase liquide peut être facilement obtenue :

$$\langle v_l \rangle_l = \frac{(1 - C_0 \langle \alpha \rangle) \langle j \rangle - \tilde{v}_{gj} \langle \alpha \rangle}{\langle 1 - \alpha \rangle} . \quad (5.65)$$

On vérifie aisément que l'hypothèse d'équilibre cinématique du modèle homogène implique $C_0 = 1$ et $\tilde{v}_{gj} = 0$ soit, comme attendu :

$$\langle v_l \rangle_l = \langle v_g \rangle_g = \langle j \rangle . \quad (5.66)$$

Quoique le modèle de flux de dérive ait été initialement conçu pour représenter les régimes d'écoulement, tel l'écoulement annulaire, pour lequel il y a effectivement une dérive locale des phases, il constitue un cadre général commode pour corréler titre et taux de vide. Idéalement, les formules donnant C_0 et \tilde{v}_{gj} devraient être choisies en fonction du régime d'écoulement. Il en existe de nombreuses, par exemple celles proposées par Hibiki et Ishii (2003). Celles utilisées dans le modèle de g. v. ne correspondent pas à des régimes d'écoulement spécifiés, ceux-ci étant en majeure partie indéterminés. En revanche, elles

ont été établies à l'issue des expériences « PATRICIA GV2 » (De Crecy, 1985) spécifiques à la géométrie et aux conditions de fonctionnement des g. v. Les formules utilisées sont :

$$C_0 = 0,45 (1 + \sin (\pi [0,5 - \chi])) + 0,1 \quad (5.67)$$

et

$$\tilde{v}_{gj} = C_0 \left(1,382 - 8,155 \sqrt{\frac{\rho_g}{\rho_l}} + 15,7 \frac{\rho_g}{\rho_l} \right). \quad (5.68)$$

La fonction sinus dans la formule donnant C_0 a été ajoutée afin d'obtenir une loi plus lisse, moins contraignante numériquement.

5.4. Plaques entretoises

La version 2 du modèle comportait un modèle de perte de charge au passage des plaques entretoises proposé par E. TOURNIÉ (Bernier, 2007). Le coefficient de perte de charge associé à un trou folié est la somme de quatre coefficients représentant successivement la réduction de section de passage par le bourrelet de colmatage, un premier élargissement, les frottements dans le trou folié et la sortie du trou folié. Depuis, des expériences sur un modèle réduit d'une partie de plaque entretoise ont été réalisées afin de mettre au point une corrélation plus sûre. Ce programme, baptisé P2C, comporte deux séries d'essais, d'abord avec un écoulement monophasique, puis diphasique eau-fréon. Seuls les résultats de la première étape étaient disponibles pendant cette thèse. La maquette utilisée est à l'échelle 1 : 1,72. Elle comporte 25 tubes droits maintenus par quatre plaques entretoises espacées de 651 mm. La troisième plaque est substituée par une plaque plus ou moins colmatée. Huit plaques ont été ouvragées par électro-érosion permettant de tester les taux de colmatage 0, 44, 58, 86 et 95 ainsi que trois morphologies différentes d'arête de colmatage pour un taux de 0,72. À partir de ces mesures, un modèle adapté d'une corrélation conçue pour décrire les écoulements au travers de grilles à bords arrondis (Idel'cik, 1969, p. 306 et diagramme 8.5) a été proposé par Pillet et coll. (2010). La figure 3.5 page 37 représente les mesures et le modèle. La formule utilisée pour déterminer le coefficient de perte de charge, ζ_{pe} , d'une plaque entretoise est

$$\zeta_{pe} = \left(\sqrt{\zeta_0(1-a)} + (1-a) \right)^2 \frac{1}{a}, \quad (5.69)$$

où a est le rapport de l'air de la section passante de la plaque par l'aire de la section de passage en amont de la plaque, et ζ_0 un coefficient tabulé. Dans la corrélation proposée par Idel'cik (1969), ζ_0 dépend de a et du rapport du rayon de l'arrondi par le diamètre hydraulique. Ici, sa valeur a été fixée à 0,3, de façon à reproduire les résultats des expériences P2C. Le coefficient a est une fonction affine du taux de colmatage dont les paramètres dépendent de l'aire de la section du *riser* non occupée par les tubes et des aires passantes des plaques non affectées par le colmatage, par exemple le jeu entre la plaque et la chemise du faisceau ou les trous non tubés.

6. Calage et validation du modèle de générateur de vapeur

Comme l'a illustré le chapitre précédent, la modélisation d'un système complexe requiert un compromis entre degrés de liberté et lois de fermeture du modèle. On voudrait pouvoir « laisser parler les équations » mais les limites de puissance de calcul, voire des théories disponibles, imposent des simplifications et le recours à des corrélations empiriques. Les corrélations disponibles pour modéliser les écoulements diphasiques sont en général issues d'expériences à la géométrie simple, souvent un tube cylindrique court, dans des conditions parfois éloignées de celles des systèmes réels, par exemple sans flux thermique en remplaçant la vapeur par du fréon. Ainsi, les développements théoriques exposés dans le chapitre précédent se doublent d'une approche heuristique d'adaptation des équations aux données disponibles. On a vu dans le chapitre 4 que cette étape délicate n'avait pas été complètement franchie par la version 2. Délicate, car le risque de dégrader le modèle par un calage maladroit n'est pas nul. D'un point de vue plus optimiste, on peut voir là un nouveau compromis. Un calage judicieux restreint le domaine de validité du modèle mais en améliore localement la performance. Ce point de vue est illustré par la figure 6.1 à l'aide d'un modèle factice.

La méthodologie de calage spécifique au modèle de g. v. est présentée dans la section 6.1. Les coefficients et paramètres de calage y sont dénombrés ainsi que les critères de calage et la séquence d'opérations mises en œuvre pour déterminer une solution acceptable de ce système. Les sections 6.2 et 6.3 présentent respectivement les résultats obtenus pour le calage en régime stationnaire et dynamique. Enfin, la validation de cette procédure en régime stationnaire à basse puissance est détaillée dans la section 6.4.

6.1. Méthodologie de calage du modèle

Les valeurs de certains paramètres du modèle sont inconnues et peuvent difficilement être estimées par une modélisation théorique additionnelle. C'est le cas notamment des coefficients de friction dans le circuit primaire ainsi que dans le *riser* et sa partie cintrée, des coefficients de perte de charge singulière des séparateurs cyclones, du *downcomer*¹ et de la jonction entre *downcomer* et *riser* ainsi que de celui régulant le débit des connexions hydrauliques modélisant les débits transverses. Ces valeurs doivent nécessairement être estimées pour compléter le modèle.

1. L'écoulement dans le *downcomer* n'est pas modélisé à proprement parler. Il est simplement représenté par une connexion hydraulique dotée d'un coefficient de perte de charge global représentant la dissipation d'énergie par friction et les pertes de charges singulières dues aux éléments solides encombrant l'écoulement.

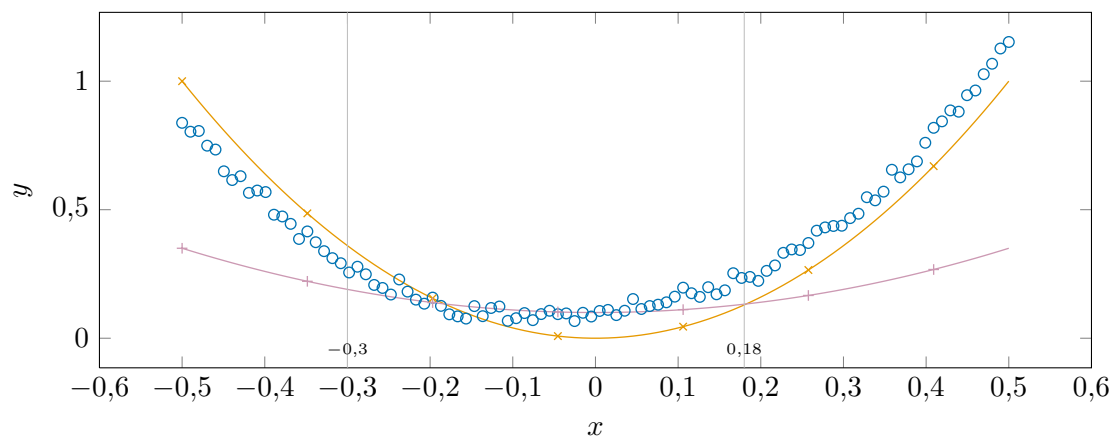


Figure 6.1. : Illustration par un modèle artificiel du compromis entre l'envergure du domaine de validité et la performance sur un domaine restreint. Les points de données (\circ) sont dans l'ensemble bien approchés par le modèle théorique (\times). En revanche, le modèle calé ($+$) est plus performant pour $x \in [-0,3 ; 0,18]$. Les formules utilisées pour générer ces données sont :

$$\circ : y = 4x^2 + 0,3x + \exp\left(\frac{-x^2}{0,3}\right) + \epsilon \text{ avec } \epsilon \sim \mathcal{G}(0, 0,02);$$

$$\times : y = 4x^2;$$

$$+ : y = 4x^2 \cdot 0,25 + 0,1.$$

Les coefficients de calage sont ici 0,25 (multiplicatif) et 0,1 (additif).

À ces paramètres de calage fondamentaux, on ajoute des coefficients de calage modifiant les corrélations jugées les plus incertaines pour déplacer les deux compromis évoqués plus haut : on cherche par là à récupérer des degrés de liberté et améliorer la performance locale du modèle. Les deux éléments jugés les plus incertains de la modélisation sont le transfert thermique à l'interface des tubes et du fluide secondaire et le modèle de flux de dérive. Après une série de tests visant à identifier l'effet des différents coefficients possibles, trois ont été retenus : un coefficient multiplicatif M_{th} modifiant le coefficient de transfert thermique η , et deux coefficients $M_{\tilde{v}_r}$ et $M_{\tilde{v}_c}$ modifiant le paramètre \tilde{v}_{gj} respectivement dans le *riser* (cintre exclu) et dans les séparateurs cyclones.

On dispose pour déterminer des valeurs pertinentes des paramètres et coefficients de calage de données globales en régime nominal, notamment la puissance totale transférée au fluide secondaire ou le taux de recirculation, qui constituent des critères de calage évidents. Afin de contraindre plus le système de calage, il a été choisi de tirer parti indirectement de la procédure de validation expérimentale du modèle THYC en comparant les champs thermohydrauliques tridimensionnels prédits par ce dernier avec les profils monodimensionnels du modèle de diagnostic du colmatage. Cette comparaison nécessite de pratiquer des interpolations qui seront explicitées dans la section suivante. C'est là l'une des principales évolutions méthodologiques par rapport à la version 2. On dispose enfin, grâce à la base de données ORLI, de mesures d'exploitation pour caler le fonctionnement en régime dynamique. Avec le concept illustré par la figure 6.1 en tête, c'est naturellement la réponse du n. g. l. lors du transitoire de puissance qui a été utilisée. Ainsi, les grandeurs témoins sont : la puissance totale transférée et le taux de recirculation, les profils de pression, d'enthalpie, de taux de vide, de titre et de puissance transférée par unité de surface, les débits transverses et l'allure des courbes de réponses de n. g. l. pour différents états de colmatage.

Dans un contexte plus simple, une démarche efficace serait de procéder d'abord à l'analyse des sensibilités des grandeurs témoins aux paramètres et coefficients de calage, puis de construire une méthode numérique d'optimisation. L'interdépendance des variables d'entrée et de sortie de ce problème d'une part, et la difficulté de définir des critères de calages objectifs d'autre part rendent néanmoins cette tâche ardue dans le cas du modèle de g. v. Cette démarche a donc été mise en œuvre « manuellement », c'est-à-dire par une succession d'étapes comportant le choix d'un groupe de paramètres variables et de grandeurs témoins, une analyse sommaire de la sensibilité locale du modèle en appliquant de petites perturbations à ces paramètres et une optimisation manuelle des paramètres choisis. La confrontation de plusieurs tentatives a permis néanmoins de définir une méthode heuristique efficace en deux étapes. Il a été observé une relative indépendance entre d'une part, la pression en bas du *riser* (ou de façon équivalente dans le dôme) et la puissance totale transférée, et d'autre part, les pertes de charges (soit la pente du profil de pression) et le taux de recirculation.

La première étape consiste à s'assurer de l'égalité de la puissance échangée et de la pression d'admission dans le *riser* avec celles issues de la simulation THYC choisie comme référence. La puissance de consigne est calée en déterminant la valeur de la consigne du débit de vapeur sortant, égal au débit alimentaire en régime stationnaire. La puissance échangée est ajustée en modifiant les deux coefficients de calage relatifs au glissement,

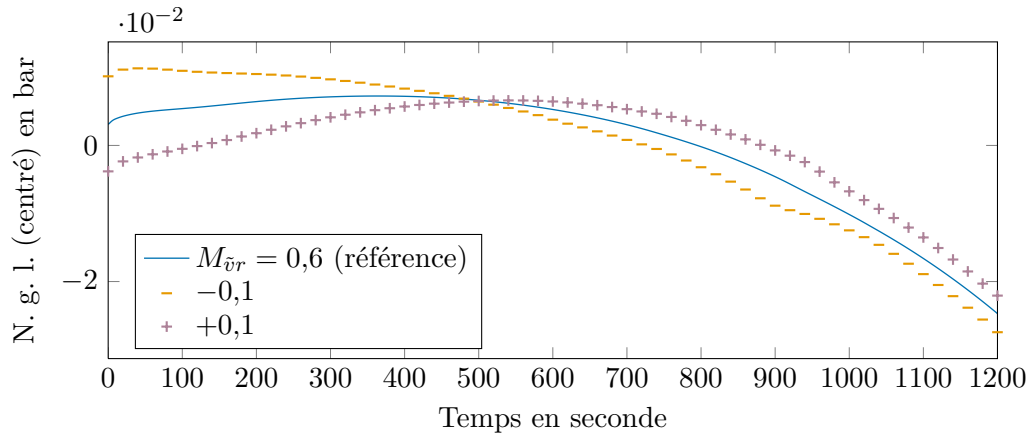


Figure 6.2. : Effet de perturbations du coefficient $M_{\bar{v}r}$ sur la réponse du n. g. l. lors d'un essai R.G.L. 4. Les moyennes temporelles des courbes leur ont été soustraites. Le calage des autres coefficients et paramètres n'a pas été ajusté après les perturbations car elles sont de faible amplitude. L'effet des perturbations est amplifié 5 fois.

$M_{\bar{v}r}$ et $M_{\bar{v}c}$. C'est ici l'écart des deux coefficients qui joue. La valeur « absolue » de $M_{\bar{v}r}$ est quant à elle déterminée par l'observation du comportement dynamique : elle est fixée de façon à donner à la forme des réponses de n. g. l. une évolution en fonction du colmatage conforme à un échantillon de courbes de réponses mesurées. La figure 6.2 illustre l'effet de petites perturbations du coefficient $M_{\bar{v}r}$ autour de la valeur retenue à l'issue du calage. Le calage des autres coefficients et paramètres n'a pas été repris dans son ensemble après les perturbations mais celles-ci sont suffisamment petites pour que la comparaison reste pertinente. L'effet des perturbations a été amplifié par un facteur 5 dans la figure afin d'en améliorer la lisibilité. Le coefficient $M_{\bar{v}r}$ accentue la courbure et diminue la pente globale de la réponse de n. g. l. Enfin, la pression en entrée du *riser* est calée à l'aide du coefficient M_{th} , les deux valeurs variant dans le même sens.

La seconde étape débute par le calage du profil de pression en ajustant les coefficients de friction dans le *riser* et de perte de charge singulière des plaques entretoises. La valeur cible du coefficient de recirculation peut ensuite être atteinte en modifiant les coefficients de perte de charge ζ_{dc} et ζ_{adm} respectivement du *downcomer* et de l'admission dans le *riser*, c'est-à-dire la jonction entre le *downcomer* et le *riser*, en bas du g. v. La répartition des pertes de charge entre le *downcomer* et l'admission n'est pas évidente. Pour beaucoup de modèles de g. v., cette répartition n'a pas une grande importance car le *downcomer* et l'admission sont consécutifs. Elle impacte cependant la forme des courbes de réponse car ces deux éléments sont de part et d'autre du piquage bas de la mesure de n. g. l. La figure 6.3 illustre l'effet de petites perturbations de la répartition finalement retenue. La procédure est la même que pour la figure 6.2, les effets sont amplifiés aussi 5 fois. Diminuer la perte de charge du *downcomer* en faveur de celle de l'admission atténue la courbure de la réponse de n. g. l. Le calage de la répartition se fait donc là encore en

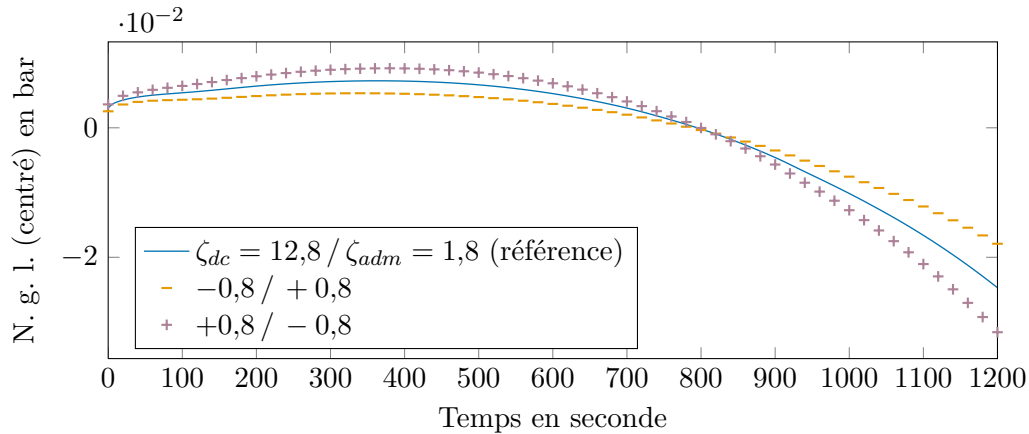


Figure 6.3. : Effet d’une perturbation du coefficient $M_{\tilde{v}r}$ sur la réponse du n. g. l. lors d’un essai R.G.L. 4. Les moyennes temporelles des courbes leur ont été soustraites. Le calage des autres coefficients et paramètres n’a pas été ajusté après les perturbations car elles sont de faible amplitude. L’effet des perturbations est amplifié 5 fois.

régime dynamique à l’aide d’un faisceau de courbes de réponse mesurées.

Les deux étapes ne sont pas complètement indépendantes et il est nécessaire d’itérer ce processus plusieurs fois pour obtenir des valeurs satisfaisantes des grandeurs témoins des deux groupes. S’ajoute à cela l’ajustement du coefficient de perte de charge réglant les débits transverses qu’il faut faire correspondre à ceux observés dans la simulation THYC de référence. Le calcul des débits transverses à partir des champs tridimensionnels simulés avec THYC est exposé dans la section 6.2. Un coefficient unique, ζ_{tr} , est utilisé pour l’ensemble des connexions transverses bien qu’elles ne soient pas dans les mêmes conditions thermohydrauliques. En effet, la précision de la méthode utilisée pour déterminer sa valeur ne semble pas suffisante pour une détermination locale. L’ajustement du coefficient ζ_{tr} est fait au fur et à mesure du processus de calage, sa dépendance aux deux étapes principales n’étant pas trop importante.

6.2. Calage en régime stationnaire à puissance nominale

Les champs thermohydrauliques simulés avec THYC, en premier lieu desquels le champ de pression, ont été utilisés comme critères de calage. Les paramètres de fonctionnement en régime nominal utilisés pour réaliser la simulation THYC de référence sont donnés dans le tableau 6.1.

La plupart des grandeurs intervenant dans les simulations THYC peuvent être extraites des fichiers de résultats. Pour chacun des plans horizontaux, on dispose ainsi d’un tableau de dimension 44×30 , correspondant au maillage illustré par la figure 4.1 page 63, que l’on sépare en deux tableaux de dimension 22×30 correspondant aux deux jambes. Des profils verticaux sont par ailleurs extraits des fichiers de résultats générés par Dymola.

Grandeur	Unité	Valeur
Pression du circuit primaire	bar	155
Enthalpie primaire (branche chaude)	kJ kg^{-1}	1469
.∴ Température primaire	$^{\circ}\text{C}$	322,5
Débit primaire	kg s^{-1}	4592
	t h^{-1}	16531,2
Pression de l'eau alimentaire	bar	65,54
Enthalpie de l'eau alimentaire	kJ kg^{-1}	940,2
.∴ Température de l'eau alimentaire	$^{\circ}\text{C}$	218,98
Débit de vapeur sortant	kg s^{-1}	500,75
Consigne de puissance	MW	923,8
Taux de recirculation		4,13

Tableau 6.1. : Paramètres de fonctionnement utilisés pour le calage en régime nominal par rapport à THYC.

Le nombre de points de ces profils dépend des grandeurs physiques car elles ne sont pas toutes définies sur les mêmes maillages. Le schéma numérique adopté est illustré par la figure 5.1 page 75. Des profils ont été calculés à partir des tableaux issus de THYC en calculant la moyenne dans chaque plan transversal des grandeurs pondérées par les volumes des mailles. Les volumes des mailles dépendent de la géométrie du maillage et de l'encombrement par les tubes et autres pièces solides. Le détail de ces opérations est exposée dans le manuel des programmes développés pour extraire des fichiers de résultat les données simulées avec THYC et Dymola et les interpoler (Girard, 2012a) .

La figure 6.4 compare les profils de pression simulés avec THYC et la version 7 du modèle, à l'issue du calage. L'abscisse représente la cote dans le riser, l'origine ayant été placée à la base du faisceau, juste au-dessus de la plaque à tubes. Les profils représentés couvrent l'ensemble du *riser*, jusqu'en amont des séparateurs cyclones, 11,26 m plus haut. Le cintre des tubes débute juste après la plaque n° 8, à environ 9 m. Les lignes gris clair indiquent les plaques entretoises, numérotées de 1 à 8, ainsi que la plaque de répartition de débit, abrégée p. r. d. Les barres d'erreur sont associées aux profils THYC. Elles représentent les quantiles 0,05 et 0,95 des échantillons utilisés pour calculer les moyennes. Elles sont dans l'ensemble très courtes car le champ de pression est relativement uniforme dans un plan transversal. Les seules hétérogénéités de pression notables sont observées dans la partie cintrée du faisceau de tubes et à l'admission, en bas du g. v.

La pente globale du champ de pression est essentiellement due au poids de la colonne de fluide. La bonne adéquation entre la référence et le modèle calé illustré par la figure 6.4 montre que le champ de masse volumique est correctement reproduit. Les variations plus fines entre les plaques sont le fait des forces de friction. Là encore les résultats sont très satisfaisants, même dans la partie cintrée dont la géométrie tridimensionnelle aurait pu poser problème. Les pertes de charge singulières au passage des plaques diffèrent légèrement mais sont de même ordre de grandeur. Notons que ces pertes de charge

n'étaient pas concernées par la procédure de validation de THYC.

La figure 6.5 montre une très bonne adéquation des profils de titre. Les cotes de vaporisation dans les deux jambes sont bien prédites par le modèle. Les résultats pour le taux de vide, illustrés par la figure 6.6 sont un peu moins bons, mais restent satisfaisants. La taille des barres d'erreur dans ces deux figures indique que les distributions transversales de ces grandeurs sont nettement plus hétérogènes que la pression.

La figure 6.7 représente les profils d'enthalpie de mélange pondérée en masse, c'est-à-dire en titre. Aucune différence importante n'est observée entre les profils de référence et les profils du modèle calé. Le croisement des profils des deux jambes au niveau de la plaque de répartition de débit est bien prédit par le modèle. La figure 6.8 montre elle aussi une bonne adéquation des profils de puissance transférée par unité de surface.

Le traitement des débits transverses est un peu différent. On appelle débits transverses les débits traversant le plan médian séparant les jambes chaude et froide. Dans THYC, chaque maille comporte les valeurs des composantes du débit selon les trois directions de l'espace. À une cote donnée, chaque maille adjacente au plan médian contribue au débit transverse par la composante perpendiculaire au plan médian de son vecteur de débit. Cependant, on peut imaginer des lignes de courant serpentant de part et d'autre du plan médian. Dans une telle configuration il n'y a pas de réel transfert de matière d'une jambe vers l'autre. Or, la représentation monodimensionnelle du g. v. ne comporte que quelques connexions hydrauliques espacées d'environ un mètre. Afin de se prémunir contre cet éventuel artefact, un rang de mailles supplémentaires dans chaque jambe a été utilisé pour calculer les débits transverses. La valeur du débit transverse utilisé pour le calage est la moyenne des sommes dans chaque rang des composantes transversales du débit. La figure 6.9 compare les débits transverses calculés à partir de la simulation THYC et ceux modélisés par des connexions hydrauliques dans la version 7 du modèle. La cote est placée en ordonnée et les plaques sont indiquées de la même façon que dans les graphiques précédents. Les barres fines orange représentent les moyenne des composantes transverses du débit dans chacun des plans transversaux du maillage THYC. C'est en observant ces données que les positions des connexions hydrauliques dans le modèle ont été décidées. On constate en effet qu'en dehors de la zone basse perturbée par l'alimentation d'eau, les débits transverses sont surtout importants juste en dessous des plaques. Les larges barres bleues correspondent à la somme des débits représentés par les fines barres orange sur les régions séparant les plaques. Une interpolation permet de faire correspondre les maillages. Là encore, c'est la moyenne de quatre sommes qui est représentée. Les barres d'erreur noires indiquent les valeurs minimale et maximale de ces quatre valeurs. On remarque que la barre d'erreur pour la zone d'admission, en dessous de la plaque de répartition de débit, s'étend de part et d'autre de l'axe des ordonnées. Ce résultat était prévisible : le flux d'eau alimentaire provenant du *downcomer* est dirigé vers le centre du *riser* et ces courants opposés se rencontrent au niveau du plan médian. Ce phénomène fondamentalement tridimensionnel ne peut être représenté par le modèle monodimensionnel. Aussi, la première connexion hydraulique est placée sous la première plaque entretoise. Les débits passant dans les connexions hydrauliques du modèle sont représentés par les traits noirs terminés d'un rond mauve. En dehors de la partie cintrée, le sens et l'amplitude des débits des débits transverses sont correctement

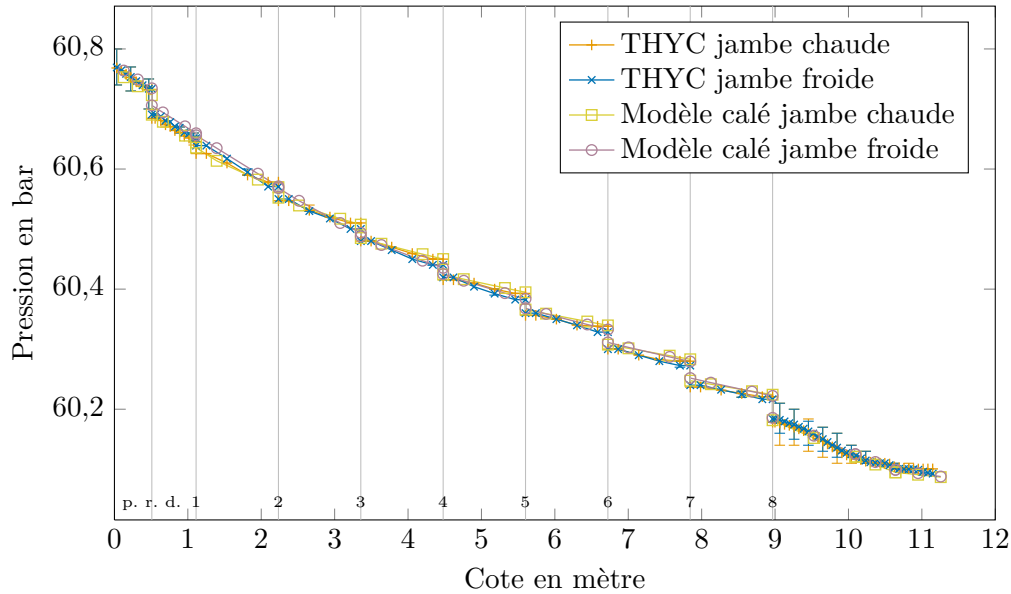


Figure 6.4. : Comparaison en régime nominal des profils de pression simulés avec THYC et la version 7 du modèle calé.

représentés par les connexions hydrauliques de la version 7 du modèle. Notamment, la cote de retournement du sens des débits transverses est bien prédite.

6.3. Calage en régime transitoire

La procédure de calage exposée dans la première section de ce chapitre comporte l'ajustement du coefficient $M_{\tilde{v}r}$ modifiant le modèle de glissement et du rapport $\frac{\zeta_{dc}}{\zeta_{adm}}$ décrivant la répartition des pertes de charges avant et après le piquage bas de la mesure de n. g. l. Ces deux paramètres agissent sur la forme des courbes de réponse de n. g. l., le premier ayant le plus d'influence et ayant été majoritairement utilisé. La capacité du modèle à reproduire fidèlement la forme des courbes de n. g. l. est une condition impérative pour pouvoir l'utiliser pour le diagnostic du colmatage.

Les figures de l'annexe F page 249 représentent les faisceaux de réponses n. g. l. mesurées qui ont été utilisés pour le calage du régime dynamique, à l'exclusion de ceux des tranches n^{os} 1 et 4 de Delta dont les taux de colmatage avant nettoyage excèdent la limite supérieure de la gamme choisie pour le diagnostic, à savoir 0 à 0,7. La figure 6.10 montre la superposition de 15 courbes simulées avec la version 7 du modèle calé (en orange —) et d'un échantillon de 28 courbes mesurées (en bleu —) et de . Les courbes simulées ont été obtenues avec les profils représentés par la figure 3.8 page 44. Les courbes de réponses mesurées proviennent des tranches n^o 4 de la centrale Alpha, n^o 1 de la centrale Bravo, n^o 2 de la centrale Charlie et n^o 2 de la centrale Delta. Elles ont été choisies pour cet exemple afin de représenter autant que possible la diversité des

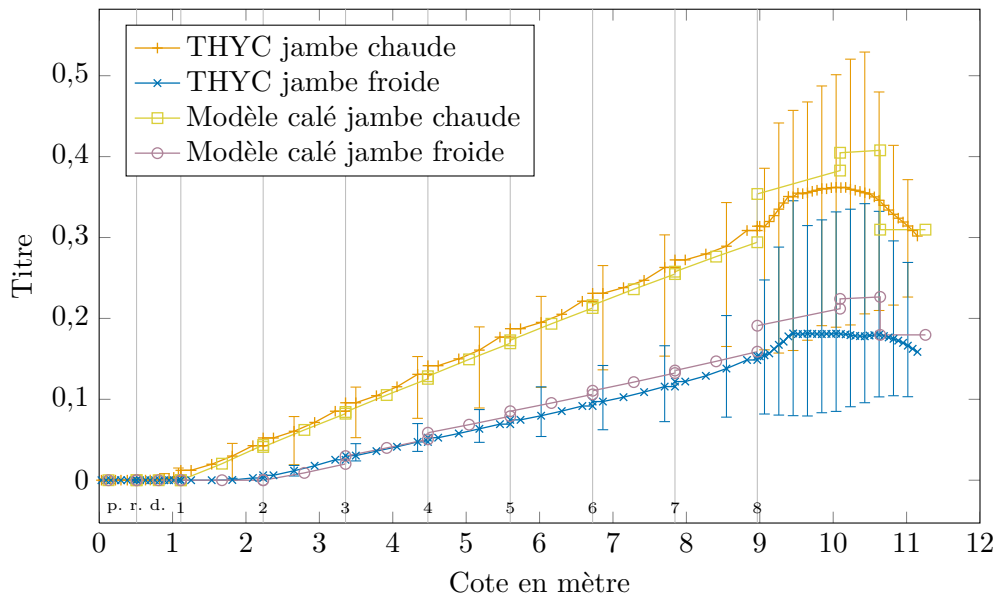


Figure 6.5. : Comparaison en régime nominal des profils de titre simulés avec THYC et la version 7 du modèle calé.

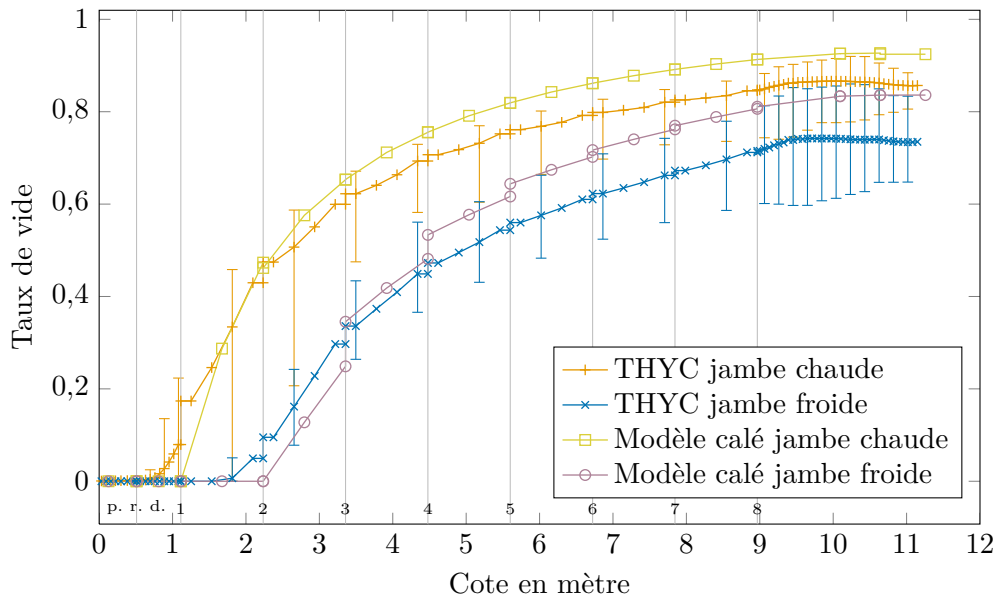


Figure 6.6. : Comparaison en régime nominal des profils de taux de vide simulés avec THYC et la version 7 du modèle calé.

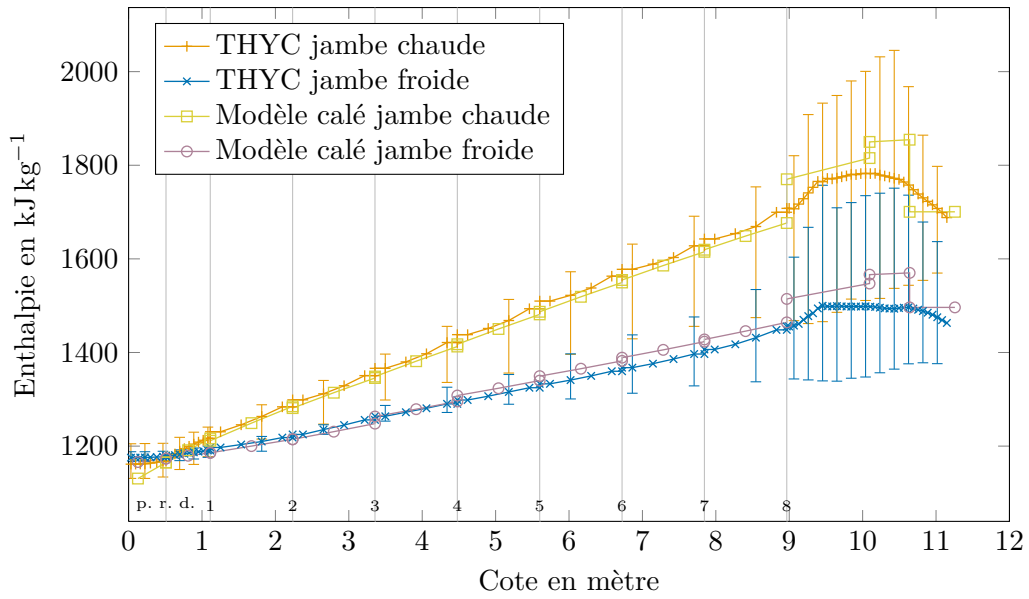


Figure 6.7. : Comparaison en régime nominal des profils d'enthalpie simulés avec THYC et la version 7 du modèle calé.

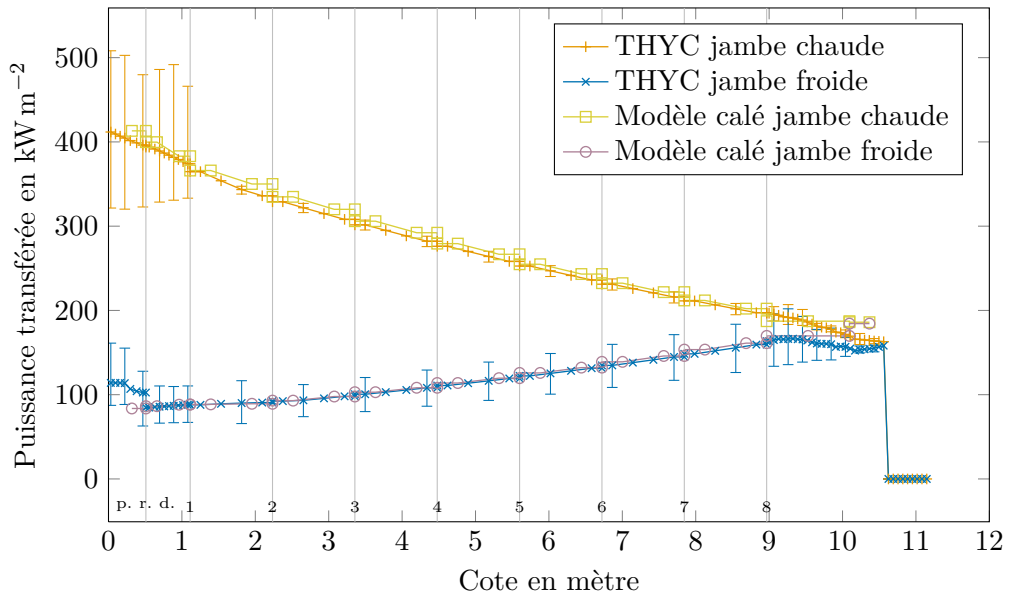


Figure 6.8. : Comparaison en régime nominal des profils de puissance transférée par unité de surface simulés avec THYC et la version 7 du modèle calé.

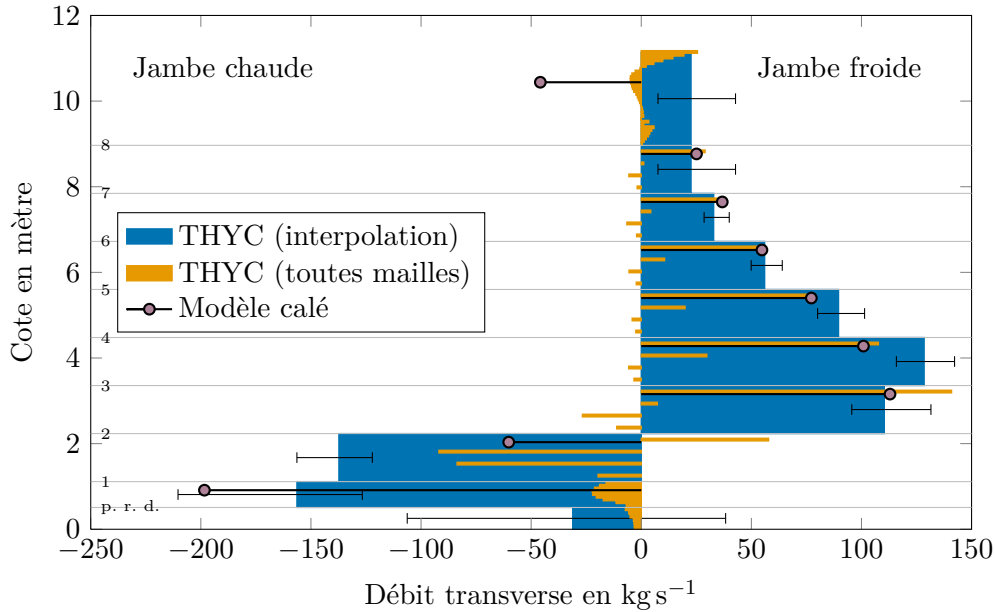


Figure 6.9. : Comparaison en régime nominal des débits transverses simulés avec THYC et la version 7 du modèle calé.

formes que l'on tâche de reproduire avec le modèle. L'adéquation des formes des courbes des deux échantillons est satisfaisante. Les 4 courbes simulées qui débordent du faisceau de courbes mesurées sont celles correspondant aux profils de colmatage très faibles que l'on ne s'attend pas à trouver en réalité. Les nettoyages chimiques ne peuvent en effet éliminer totalement le colmatage et les g. v. utilisés lors des remplacement ne sont pas type 51 B ou apparenté.

Signalons ici un point important pour le développement de la méthode de diagnostic. La valeur du n. g. l. réel est obtenue à partir de la mesure de pression différentielle en lui appliquant une fonction affine afin de correspondre à un niveau d'eau, mesuré en mètres. Les paramètres de cette fonction sont ajustés par l'opérateur pour rendre compte des disparités entre les différents g. v. Par ailleurs, compte tenu de la représentation simplifiée de la géométrie des g. v., il n'est pas possible de reproduire avec exactitude le positionnement des piquages de la mesure de n. g. l. La position du piquage bas est représentée dans le modèle par le rapport $\frac{\zeta_{dc}}{\zeta_{adm}}$ des coefficients de pertes de charges du *downcomer* (en amont du piquage) et de l'*admission* dans le *riser* (en aval). La figure 6.3 illustre donc l'effet du positionnement du piquage bas sur l'amplitude des courbes de réponse. De fait, on peut considérer que le modèle actuel ne permet pas de prédire avec une grande fiabilité cette amplitude. L'expédient à ce problème mis en œuvre dans la première version de la méthode de diagnostic proposée par Chip et coll. (2008) était de diviser chaque courbe de réponse par son amplitude préalablement à la comparaison. Ce prétraitement a été abandonné dans la nouvelle méthode au profit d'un coefficient de proportionnalité entre réponses réelles et simulées unique pour tous les g. v. La

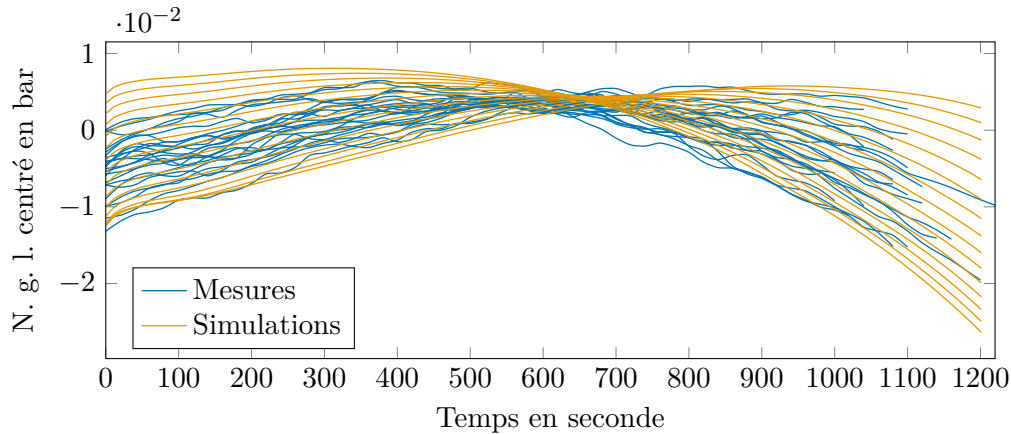


Figure 6.10. : Superposition de réponses simulées avec le modèle calé pour les 15 profils de colmatage représentés dans la figure 3.8 avec des réponses mesurées sur des g. v. des tranches n° 4 de la centrale Alpha, n° 1 de la centrale Bravo, n° 2 de la centrale Charlie et n° 2 de la centrale Delta.

détermination du coefficient de proportionnalité repose sur la comparaison de faisceaux de réponses comme ceux représentés par la figure 6.10. Elle fait partie intégrante de la méthodologie de calage et revient, en quelque sorte, à sélectionner la gamme de taux de colmatage sur laquelle le modèle est défini. La valeur du coefficient utilisée ici est -4000 . Cette même valeur a été utilisée pour produire les diagnostics exposés dans le chapitre 13.

6.4. Validation en régime stationnaire à basse puissance

Malgré les précautions prises lors du calage du modèle, on pourrait craindre d'avoir trop altéré les lois physiques qui le constituent. Afin de vérifier que cela n'est pas le cas, une comparaison du modèle avec THYC a été réalisée en régime stationnaire à basse puissance. Un palier stable à basse puissance ayant eu lieu à la tranche n° 2 de la centrale Charlie a été utilisé pour fixer les valeurs des paramètres de fonctionnement. Elles sont données dans le tableau 6.2.

La figure 6.11 compare les profils de pression obtenus en fixant le débit de vapeur sortant du modèle à 212 kg s^{-1} . Si les pentes des deux profils correspondent, on observe cependant un décalage entre les deux modèles d'environ 0,5 bar, soit un peu moins de 0,8 %. Ce décalage n'est pas rédhibitoire. D'ailleurs, une réduction de 4 kg s^{-1} du débit de vapeur permet de le rattraper comme le montre la figure 6.12. Cette modification de la valeur du débit de vapeur se justifie car la modélisation du retour d'eau est un peu différente dans les deux modèles. Elle n'affecte que très marginalement les profils des autres grandeurs. Ceux-ci sont représentés dans les figures 6.13 à 6.16 pour un débit de vapeur de 208 kg s^{-1} . Comme lors du calage en régime nominal, l'adéquation entre THYC et la version 7 du modèle est excellente. La cote de vaporisation est bien prédite ainsi que le croisement des enthalpies des deux jambes. Seul le taux de vide présente des disparités

Grandeur	Unité	Valeur
Pression du circuit primaire	bar	155
Enthalpie primaire (branche chaude)	kJ kg^{-1}	1354,37
\therefore Température primaire	$^{\circ}\text{C}$	303,04
Débit primaire	kg s^{-1}	4592
	t h^{-1}	16531,2
Pression de l'eau alimentaire	bar	65,54
Enthalpie de l'eau alimentaire	kJ kg^{-1}	788,71
\therefore Température de l'eau alimentaire	$^{\circ}\text{C}$	185,18
Débit de vapeur sortant	kg s^{-1}	212
Consigne de puissance	MW	414,58
Taux de recirculation		9,56

Tableau 6.2. : Paramètres de fonctionnement utilisés pour la validation en régime stationnaire à basse puissance.

qui restent néanmoins acceptables. En dehors de la partie cintrée du faisceau les débits transverses représentés dans la figure 6.17 correspondent bien. La cote de changement de sens des débits transverses se déplace à basse puissance de la plaque n° 2 à la plaque n° 3. Cet effet est bien prédit par le modèle.

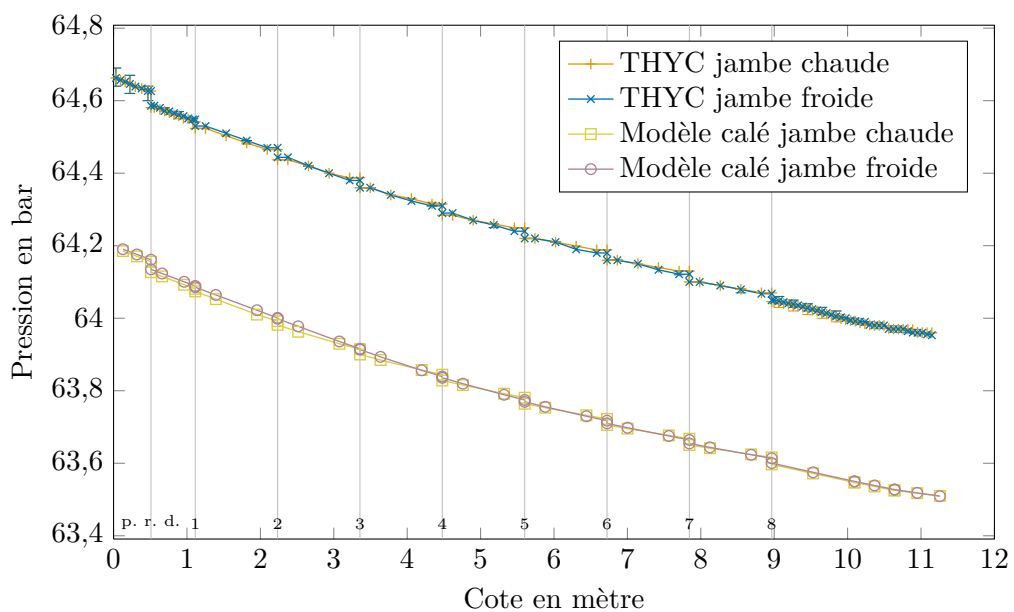


Figure 6.11. : Comparaison à basse puissance des profils de pression simulés avec THYC et la version 7 du modèle calé, pour un débit de vapeur de 212 kg s^{-1} .

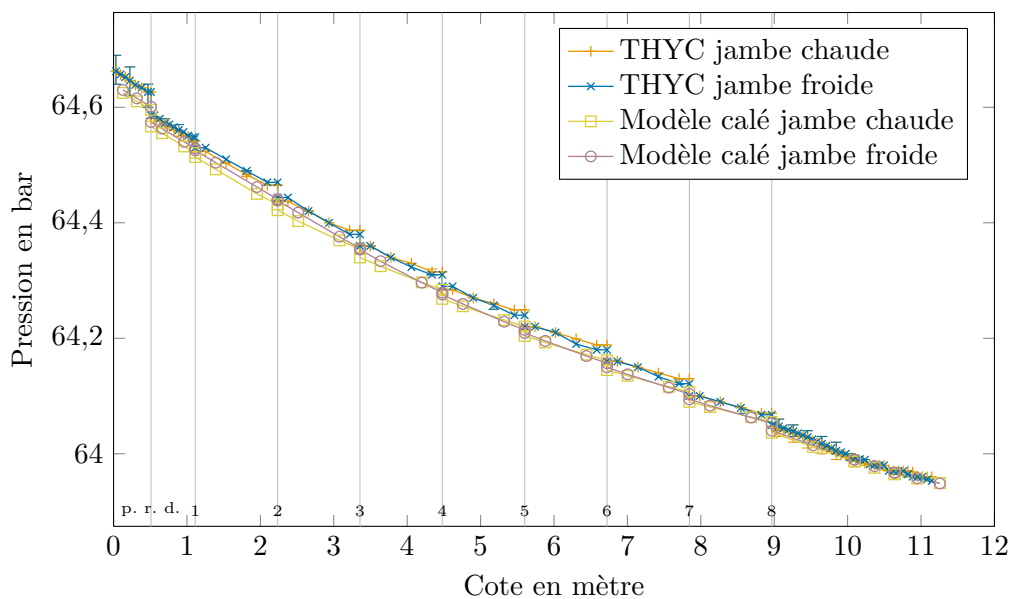


Figure 6.12. : Comparaison à basse puissance des profils de pression simulés avec THYC et la version 7 du modèle calé, pour un débit de vapeur de 208 kg s^{-1} .

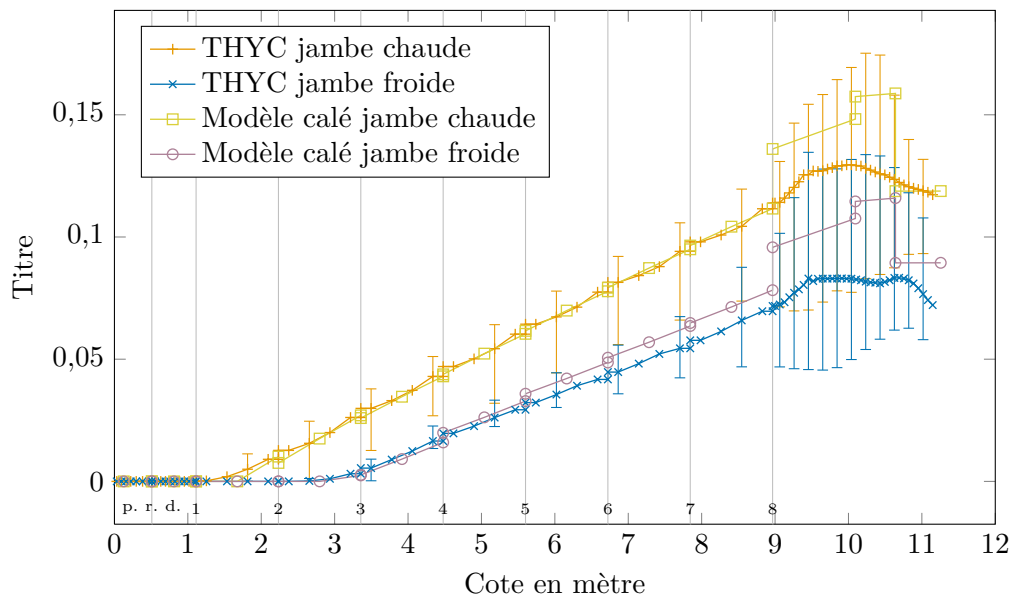


Figure 6.13. : Comparaison à basse puissance des profils de titre simulés avec THYC et la version 7 du modèle calé.

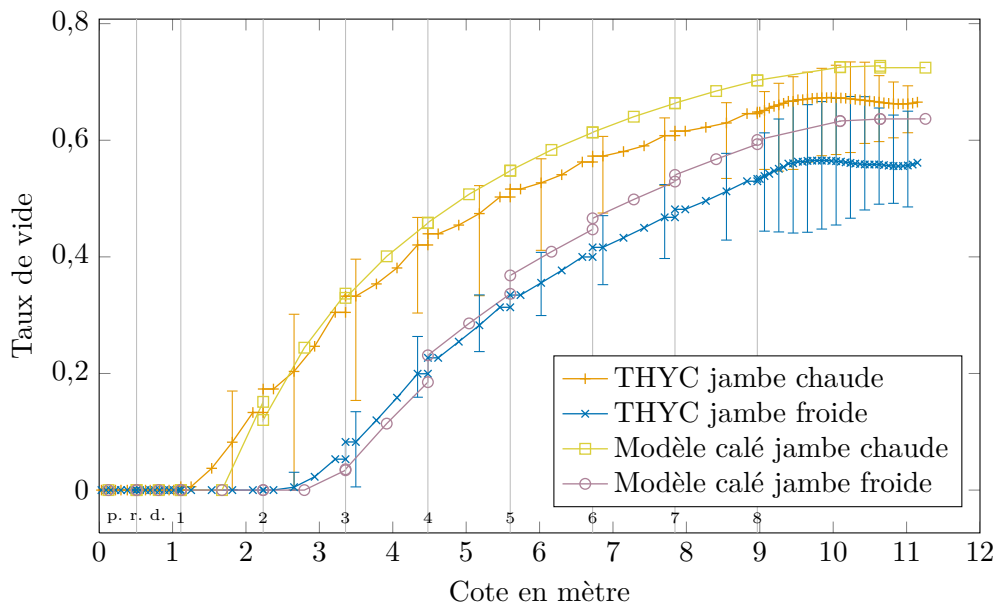


Figure 6.14. : Comparaison à basse puissance des profils de taux de vide simulés avec THYC et la version 7 du modèle calé.

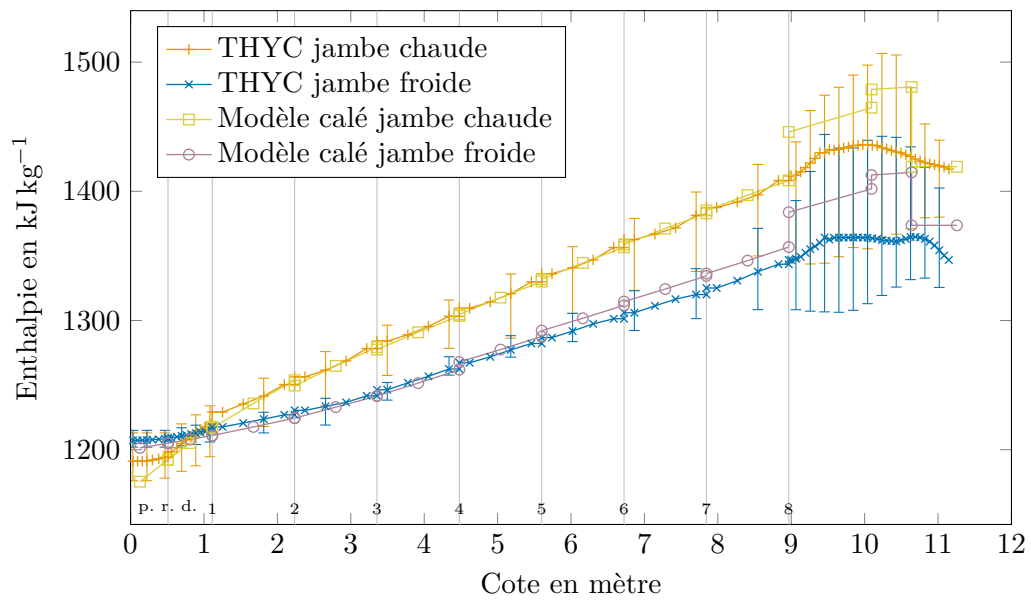


Figure 6.15. : Comparaison à basse puissance des profils d'enthalpie simulés avec THYC et la version 7 du modèle calé.

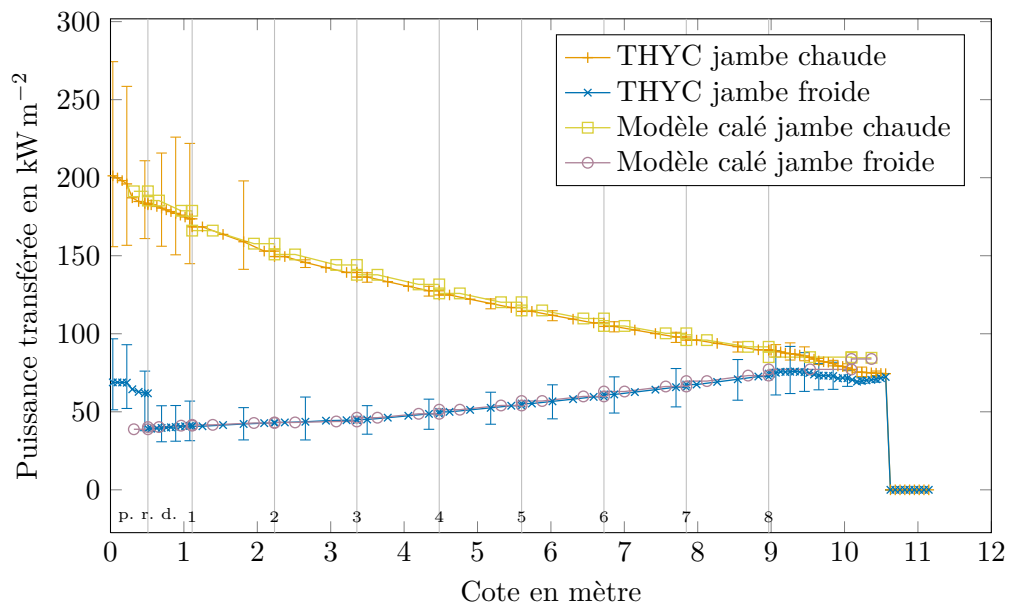


Figure 6.16. : Comparaison à basse puissance des profils de puissance transférée par unité de surface simulés avec THYC et la version 7 du modèle calé.

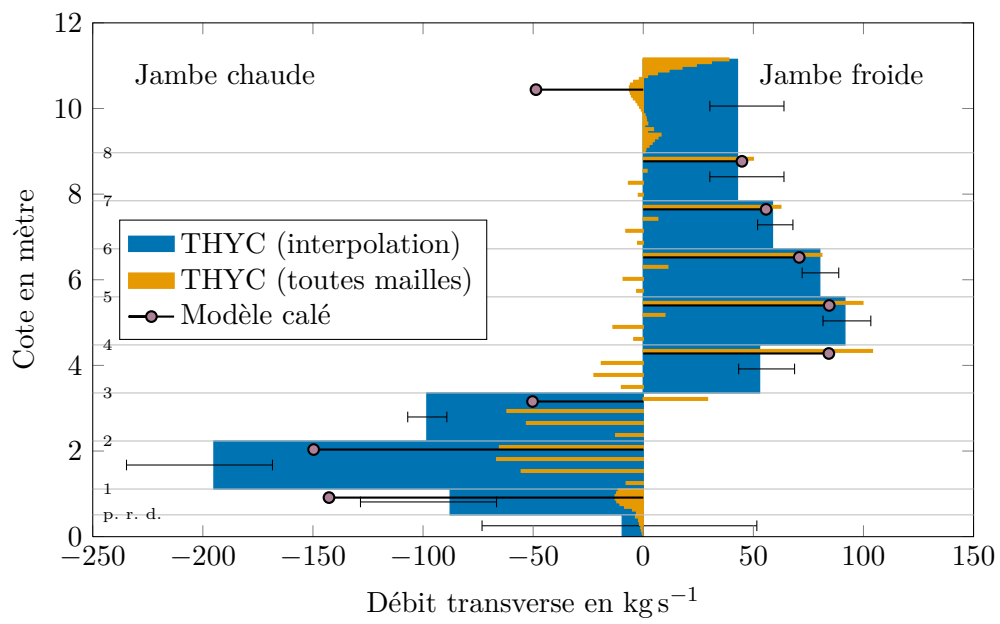


Figure 6.17. : Comparaison à basse puissance des débits transverse simulés avec THYC et la version 7 du modèle calé.

Troisième partie

Modèles et méthodes statistiques

7. Analyse de sensibilité globale

L'analyse de sensibilité d'un modèle est l'étude de l'effet sur ses variables de sortie de variations de ses variables d'entrée. Appliquée au modèle de générateur de vapeur, elle permet d'estimer l'impact d'un dysfonctionnement sur des grandeurs mesurées par les capteurs d'exploitation. Si l'on s'intéresse au comportement du modèle autour d'un point donné de l'espace des paramètres, on procédera à une analyse de sensibilité dite *locale* (Iooss, 2011). Les *indices de sensibilité* sont alors les valeurs au point d'intérêt des dérivées partielles de la fonction représentant le modèle. L'analyse de sensibilité locale revient à estimer le gradient de cette fonction en un point en appliquant aux variables d'entrée de petites perturbations. Ces résultats locaux ne peuvent être étendus à un intervalle étendu que si le modèle est linéaire.

Lorsque la variable d'entrée est un paramètre décrivant un dysfonctionnement, on s'intéresse au contraire à l'ensemble de son domaine de variation. En effet, lors du diagnostic les valeurs de ces variables sont les inconnues du problème. Ce sont donc les techniques de l'analyse de sensibilité *globale* qui ont été utilisées ici. Elles consistent à étudier la sensibilité du modèle aux variations de ses variables d'entrée sur des domaines non ponctuels.

La section 7.1 propose une brève revue des méthodes d'analyse de sensibilité les plus répandues et explique pourquoi la méthode de Sobol' a été choisie ici. Le formalisme de cette méthode est introduit dans la section 7.2 en définissant des indices de sensibilité appelés indices de Sobol'. La section 7.3 présente comment les indices de Sobol' peuvent être estimés à l'aide d'une méthode quasi-Monte-Carlo. Enfin, la construction par *bootstrap* d'intervalles de confiance pour les indices estimés est abordée dans la section 7.4.

7.1. Choix d'une méthode d'analyse de sensibilité

La pratique de l'analyse de sensibilité globale consiste à attribuer aux variables d'entrée des distributions de probabilités et étudier la distribution de la sortie, à l'aide d'un échantillon. La taille de l'échantillon nécessaire dépend de la méthode choisie mais il est souvent d'au moins quelques dizaines d'individus par variable d'entrée et en général plutôt de quelques centaines ou plus. Lorsque le nombre de variables d'entrée est très important ($\gtrsim 20$) il peut être nécessaire d'utiliser au préalable une méthode de *criblage* pour en sélectionner un sous-groupe des plus influentes. On peut utiliser par exemple des plans d'expériences classiques (Saporta, 2006, chap. 21). Dans ce contexte, les variables d'entrée sont souvent appelées *facteurs*. On autorise chaque facteur à prendre un nombre donné de valeurs appelées *niveaux*. Un plan d'expérience consiste en une série de *traitements*, c'est-à-dire d'expérience réalisés pour une combinaison de niveaux des facteurs. Le *plan factoriel complet* explore par exemple l'ensemble des combinaisons

possibles de niveaux des facteurs. Cette exhaustivité permet notamment d'estimer les effets des *interactions* entre variables d'entrée. On parle d'interaction lorsque l'effet de plusieurs variables variant simultanément est plus important que la somme des effets individuels. Il requiert cependant un grand nombre d'évaluations du modèle (2^p où p est le nombre de facteurs) et ne peut être utilisé pour l'étude du colmatage décrit par 16 facteurs. À l'inverse, le plan *un-à-la-fois*, abrégé en anglais O.A.T. pour *one-at-a-time*, ne fait varier qu'un facteur à la fois. Avec deux niveaux par facteur, le nombre d'évaluations nécessaire n'est donc que de $2p + 1$. Cependant, ce plan d'expérience devrait être évité car les effets qu'il mesure sont en général assez localisés et il ne permet pas d'apprécier les effets des interactions entre variables d'entrée (Saltelli et Annoni, 2010). La méthode de Morris (1991) permet de contourner ces écueils. Elle consiste en une succession de plans un-à-la-fois dont le point de départ et les directions de variations sont tirées aléatoirement. La moyenne des effets élémentaires estimés à chaque étape donne des estimations quantitatives dans $[0, +\infty[$ de la sensibilité de chaque variable. Elle permet de distinguer les variables ayant un effet non négligeable et parmi elles, de distinguer celles qui ont un effet linéaire et sans interactions.

Outre une grande dimension d'entrée, l'analyse de sensibilité peut se heurter à l'obstacle d'un modèle nécessitant des temps de calcul très longs. Lorsqu'une évaluation du modèle dure plus de quelques heures, seul un nombre très limité d'entre elles pourra être réalisé. Une façon de traiter ce problème est de construire un *méta-modèle* à partir de ces quelques données et d'utiliser ce dernier pour l'estimation des indices de sensibilité. Un méta-modèle est une approximation du modèle initial disponible en tout point du domaine d'intérêt avec un temps de calcul négligeable. On parle aussi d'approcher la *surface de réponse* du modèle, c'est-à-dire la surface décrite par la sortie dans \mathbb{R}^p lorsque les variables d'entrée varient. On se reportera à la revue de Iooss (2011) pour un aperçu de l'étendue de la panoplie des techniques existantes pour construire une telle approximation.

Une fois que l'on s'est ramené à une situation où l'on peut disposer d'un échantillon conséquent au regard du nombre de variables d'entrée, l'estimation des sensibilités proprement dite, c'est-à-dire la quantification des effets de chaque variable et de leurs interactions, peut débuter. Les méthodes les plus simples d'analyse de sensibilité globale reposent sur la régression linéaire. Elles ne nécessitent pas d'échantillon important mais supposent la linéarité du modèle. La sensibilité des variables est alors mesurée par les coefficients de régression standard ou les coefficients de corrélation partielle (Saporta, 2006, chap. . 3). Au prix du remplacement des variables d'entrée et de sortie par leurs rangs, l'hypothèse de linéarité peut être remplacée par une hypothèse de simple monotonie.

Les méthodes ne nécessitant pas d'hypothèse sur la linéarité ou la monotonie du modèle reposent sur l'analyse de la distribution empirique de la sortie obtenue à l'aide de l'échantillon. Par exemple, les *graphiques de contribution à la moyenne et à la variance de l'échantillon*, éventuellement assortis de tests quantitatifs (Bolado Lavín et coll., 2009), permettent une estimation simple des sensibilités ne nécessitant pas de grands échantillons. À l'inverse, certaines méthodes récentes utilisent l'intégralité des distributions et non seulement ses premiers moments (Borgonovo, 2007). L'approche la plus développée actuellement se focalise sur la variance de la sortie, dont on cherche à attribuer des parts aux variables d'entrée. C'est le cas notamment de la méthode de Sobol' qui est

fondée sur une décomposition fonctionnelle du modèle en une somme de termes dont les variances permettent d'exprimer les effets des variables et leurs interactions. Cette méthode nécessite de réaliser plusieurs centaines d'évaluations du modèle par variable d'entrée. Elle bénéficie en contrepartie d'une bonne robustesse. La seule hypothèse qu'elle nécessite est que la fonction sous-jacente au modèle soit de carré intégrable. C'est par ailleurs une méthode assez mature pour laquelle on dispose d'un certain recul, tant du point de vue théorique que pratique ; en témoignent les nombreux exemples d'application dans la littérature, par exemple : (Saltelli et coll., 2008, p. 169 et chap. 6), (Campolongo et Saltelli, 1997), (Lamboni et coll., 2009).

Trois critères président au choix d'une méthode d'analyse de sensibilité globale parmi les nombreuses disponibles (Iooss, 2011) : les hypothèses que l'on peut faire sur la régularité du modèle, le nombre de variables d'entrée et le nombre d'évaluations du modèle disponibles. Le modèle de g. v. compte 16 variables d'entrée pour décrire le colmatage. Il n'a donc pas été nécessaire de faire appel aux méthodes de criblage. Par ailleurs, une simulation d'un transitoire avec le modèle de g. v. dure en général moins de dix minutes. Le modèle a donc été utilisé directement sans avoir recours à la construction d'un méta-modèle.

Il semble hautement improbable que l'effet du colmatage sur la réponse du niveau gamme large soit linéaire. Il suffit pour s'en convaincre de constater le caractère fortement non linéaire du modèle de perte de charge utilisé, décrit dans les sections 3.1.2 et 5.4. En revanche, l'hypothèse d'intégrabilité prescrite par la méthode de Sobol' ne semble pas poser de problème a priori : le modèle est un système d'équations différentielles faisant intervenir des fonctions plutôt lisses, sans singularités importantes. Le modèle de g. v. ne permet pas d'effectuer de calculs distribués. En revanche, il est possible d'exécuter plusieurs simulations simultanément sur un même ordinateur. La réalisation d'un échantillon de 1000 simulations consécutives par variable d'entrée prendrait environ 111 jours. Il faut donc lancer 15 séries de simulations simultanément pour rapporter cette durée à une semaine. Cela peut être fait en utilisant intensivement 2 machines équivalentes aux stations de travail HP Z600, ou 3 d'entre elles en laissant possible leur utilisation pour d'autres tâches nécessitant moins de puissance de calcul. Ces considérations ont amené à choisir la méthode de Sobol' avec un plan d'expérience de type quasi-Monte-Carlo (Sobol', 1993, 2001). Une autre méthode appelée en anglais *Fourier amplitude sensitivity test* (FAST) a été proposée par R. I. CUKIER, H. B. LEVINE et K. E. SHULER (Saltelli et coll., 2008, chap. 4.7). Elle permet d'estimer les mêmes indices de sensibilité que la méthode de Sobol' et sera présentée dans la section 7.3 une fois que ceux-ci auront été définis.

7.2. Indices de Sobol'

L'analyse de sensibilité s'applique aux modèles pouvant être représentés par des fonctions numériques à plusieurs variables. Une telle fonction n'a pas nécessairement d'expression analytique, mais l'on suppose que l'on peut connaître sa valeur en tout point, par exemple en résolvant numériquement le système d'équations différentielles qu'elle

représente. La fonction qui représente le modèle est une fonction de p variables auxquelles on associe des variables aléatoires appelées *variables d'entrée* du modèle. On associe de même à ces fonctions coordonnées des variables aléatoires appelées *variables de sortie* du modèle. Lorsque les variables d'entrée sont indépendantes, ce que l'on supposera vrai ici, l'ensemble de départ peut être ramené par un codage adéquat au cube unité de dimension p , noté \mathbb{I}^p . On étudie ici le cas d'une variable de sortie unique. Le cas de variables de sortie multiples ou fonctionnelles nécessite des traitements particuliers. Une approche de cette problématique utilisant l'analyse en composantes principales est traitée dans la section 8.3 du chapitre 8.

Notons f la fonction déterministe représentant le modèle étudié :

$$\begin{aligned} f &: \mathbb{I}^p \rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto y = f(\mathbf{x}). \end{aligned} \tag{7.1}$$

On note \mathbf{X} le vecteur aléatoire d'entrée et Y la variable aléatoire de sortie. La variance de la sortie, $V = \text{var}(Y)$, représente la variabilité de la sortie du modèle due aux variations des variables d'entrée.

Lorsque l'on fixe une composante X_i de \mathbf{X} , la variance de Y décroît d'une quantité qui dépend des valeurs choisies. En moyenne, la variance vaut alors $E(\text{var}(Y|X_i))$. Or, la formule de la variance totale (Saporta, 2006, chap. 3) permet, dans les cas où V est finie, d'écrire l'égalité suivante :

$$V = E(\text{var}(Y|X_i)) + \text{var}(E(Y|X_i)). \tag{7.2}$$

Le premier terme est souvent qualifié de *résiduel* ; il représente la variance de la sortie due aux autres variables que X_i . Le second terme mesure la chute de variance. Il est d'autant plus grand que la variance induite par la variable X_i est importante. On peut ainsi définir une mesure, S_i , de l'effet de X_i sur Y , normalisée par la variance totale V :

$$S_i = \frac{\text{var}(E(Y|X_i))}{V}. \tag{7.3}$$

Cette mesure d'importance sera appelée dans la suite *indice de Sobol' d'ordre 1* associé à X_i .

Les S_i représentent l'effet de chaque variable prise séparément. Lorsque l'on applique le même raisonnement à une famille de variables X_i, \dots, X_j , la chute de variance est due aux effets de chacune des variables de la famille ainsi qu'à leurs *interactions*. Pour un modèle donné, il peut arriver que les variables X_i et X_j aient séparément des effets minimes mais que leur impact soit important lorsqu'elles varient simultanément. Cela se traduit par des valeurs faibles de $\text{var}(E(Y|X_i))$ et $\text{var}(E(Y|X_j))$, mais néanmoins une valeur forte de $\text{var}(E(Y|X_i, X_j))$. Des indices de Sobol' mesurant ces effet seront introduits dans la section suivante.

Une autre mesure d'importance consiste à considérer ensemble une variable et toutes ses interactions. Notons \mathbf{X}_{-i} l'ensemble des variables d'entrée privé de X_i . D'après ce

qui précède, si l'on fixe la valeur de \mathbf{X}_{-i} , la variance chute d'une quantité $\text{var}(\text{E}(Y|\mathbf{X}_{-i}))$ qui représente l'effet des variables de \mathbf{X}_{-i} et de leurs interactions entre elles. Son complémentaire à V est donc l'effet de X_i et de ses interactions avec les autres variables, à savoir, d'après la formule de la variance totale, la variance résiduelle $\text{E}(\text{var}(Y|\mathbf{X}_{-i}))$. On appelle *indice de Sobol' total* associé à X_i la quantité

$$S_i^{tot} = \frac{\text{E}(\text{var}(Y|\mathbf{X}_{-i}))}{V} = 1 - \frac{\text{var}(\text{E}(Y|\mathbf{X}_{-i}))}{V}. \quad (7.4)$$

7.2.1. Décomposition ANOVA

Supposons f définie et intégrable sur \mathbb{I}^p et considérons la décomposition suivante,

$$f(\mathbf{X}) = f_0 + \sum_{s=1}^p \sum_{i_1 < \dots < i_s} f_{i_1 \dots i_s}(X_{i_1}, \dots, X_{i_s}), \quad (7.5)$$

où f_0 est une constante et les $f_{i \dots j}$ des fonctions des familles de variables X_i, \dots, X_j . La somme double signifie qu'il y a une fonction $f_{i_1 \dots i_s}(X_{i_1}, \dots, X_{i_s})$ pour chaque sous-famille possible de variables d'entrée : de $f_1(X_1)$ à $f_n(X_n)$ pour les singletons, puis $f_{ij}(X_i, X_j)$ avec $1 \leq \dots < i < j \leq n$ pour les doublets et ainsi de suite jusqu'à $f_{1 \dots n}(X_1 \dots X_n)$ pour la famille complète. Cette décomposition comporte donc 2^n termes.

Définition 1 (Décomposition ANOVA). La décomposition (7.5) est appelée *décomposition ANOVA* de f si chacun de ses termes non constants est d'intégrale nulle selon chacune de ses variables, c'est à dire si

$$\int_0^1 f_{i \dots j}(x_i, \dots, x_j) dx_k = 0 \quad \text{pour } k \in \{i, \dots, j\}. \quad (7.6)$$

Sobol' (1993) a montré que la décomposition ANOVA des fonctions intégrables existe et est unique. Par ailleurs, la condition (7.6) implique que les termes de (7.5) sont orthogonaux. On obtient ainsi, par intégrations successives selon les composantes de \mathbf{x} , les identités suivantes :

$$f_0 = \int_0^1 f(\mathbf{x}) d\mathbf{x} = \text{E}(Y), \quad (7.7)$$

$$f_i = \int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i} = \text{E}(Y|X_i) - f_0, \quad (7.8)$$

$$f_{ij} = \int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i,j} = \text{E}(Y|X_i, X_j) - f_i - f_j - f_0, \quad (7.9)$$

et ainsi de suite pour les termes d'ordre supérieur.

L'acronyme anglais ANOVA fait référence à l'analyse de variance (*analysis of variance*). En effet, si f est de carré intégrable, tous les $f_{i \dots j}$ le sont aussi et on peut élever au carré l'équation (7.5) et l'intégrer sur \mathbb{I}^p :

$$\int_0^1 f^2(\mathbf{x}) d\mathbf{x} - f_0^2 = \sum_{s=1}^p \sum_{i_1 < \dots < i_s} \int_0^1 f_{i_1 \dots i_s}^2 dx_{i_1} \dots dx_{i_s}. \quad (7.10)$$

Dans le cas où les X_i sont indépendants et distribués chacun selon une loi uniforme sur $[0,1]$, le membre de gauche n'est autre que V , la variance de Y , tandis que les termes de la double somme du membre de droite sont les variances des $f_{i\dots j}(X_i, \dots, X_j)$, que l'on notera $V_{i\dots j}$ (Sobol', 1993) :

$$V_{i\dots j} = \text{var}(f_{i\dots j}(X_i, \dots, X_j)) = \int_0^1 f_{i_1\dots i_s}^2 dx_{i_1} \dots dx_{i_s} . \quad (7.11)$$

En fait, seule l'indépendance des variables d'entrée est nécessaire : on peut intégrer dans le modèle des opérations transformant les distributions uniformes en d'autres distributions, par exemple par la méthode de la transformée inverse ou la méthode de rejet (Robert et Casella, 2004, chap. 2). L'équation (7.10) constitue donc une décomposition de la variance de Y selon les effets des variables d'entrée et leurs interactions.

On appelle *indice de Sobol' d'ordre s* associé à X_{i_1}, \dots, X_{i_s} le rapport

$$S_{i_1\dots i_s} = \frac{V_{i_1\dots i_s}}{V} . \quad (7.12)$$

Les indices de Sobol' d'ordre supérieur à 1 représentent les effets d'interactions entre variables. En divisant l'équation (7.10) par la variance totale V , on remarque que la somme des $2^p - 1$ indices de Sobol' vaut 1 :

$$1 = \sum_{s=1}^p \sum_{i_1 < \dots < i_s} S_{i_1\dots i_s} . \quad (7.13)$$

Autrement dit, la variabilité de la sortie est la somme des variabilités induites par chacune des combinaisons des variables d'entrée.

Le nombre d'indices de Sobol' croît géométriquement avec l'ordre, or leur estimation nécessite de nombreuses évaluations du modèle. C'est pourquoi on se contente souvent d'estimer les indices d'ordre 1 et totaux et d'apprécier l'importance des interactions en calculant leur différence. En effet, la somme des $2^{p-1} - 2$ variances des $f_{i\dots j}$ fonctions de X_i ,

$$V_i^{tot} = V_i + \sum_{s=1}^{p-1} \sum_{\substack{i_1 < \dots < i_s \\ i_k \neq i}} V_{ii_1\dots i_s} , \quad (7.14)$$

n'est autre que le numérateur des indices totaux :

$$S_i^{tot} = \frac{V_i^{tot}}{V} . \quad (7.15)$$

La somme des indices totaux est supérieure ou égale à 1. L'égalité est obtenue pour les modèles additifs, c'est-à-dire sans interaction, chaque indice total étant alors égal à l'indice d'ordre 1 correspondant. La différence entre l'indice total et l'indice d'ordre 1 d'une variable est donc une mesure de l'importance des interactions auxquelles elle contribue. L'équation (7.14) pourrait faire penser que l'estimation des indices totaux est coûteuse en temps de calcul. Cependant, on verra dans la section 7.3.2 que celle-ci ne nécessite autant d'évaluation du modèle que pour les indices d'ordre 1. Cela est possible grâce à la formule complémentaire donnée par la seconde égalité de l'équation (7.4)

7.3. Estimation des indices de Sobol'

La méthode *FAST* (Saltelli et coll., 2008, chap. 4.7) (Cukier et coll., 1978) et la méthode de Sobol' (Sobol', 1993, 2001) sont les méthodes d'estimation des indices de Sobol' les plus répandues. La méthode *FAST* est fondée sur une transformée de Fourier multidimensionnelle. Son principal avantage est d'être en général moins coûteuse en temps de calcul. En revanche, elle est plus complexe à mettre en œuvre, peut induire des biais et ne permet pas la construction d'intervalles de confiance, puisque étant déterministe (Iooss, 2011).

Le principe de la méthode de Sobol' est d'estimer les expressions intégrales des indices de Sobol' obtenues à partir de l'équation (7.11) à l'aide d'une méthode de Monte-Carlo. Les méthodes d'intégration numérique classiques sont inefficaces pour estimer ces intégrales car elles sont de dimension trop importante (Robert et Casella, 2004, p. 21). En pratique, les méthodes de Monte-Carlo classiques sont elles-mêmes peu efficaces pour l'estimation des indices de Sobol' dès que le temps nécessaire pour réaliser une évaluation du modèle excède quelques dizaines de secondes. Ainsi, on a recours à une variante de ces méthodes qualifiée de *quasi-Monte-Carlo*. Le terme quasi-Monte-Carlo sera explicité dans la section 7.3.3. Disons simplement pour le moment qu'il s'agit d'une variante des méthodes de Monte-Carlo permettant d'accélérer la convergence d'environ un ordre de grandeur.

Une dizaine de minutes est nécessaires pour simuler un transitoire avec le modèle de g. v. avec une station de travail HP Z600. Ce coût modéré autorise l'utilisation de méthode quasi-Monte-Carlo ; ainsi la méthode de Sobol' a été choisie. Les intervalles de confiances *bootstrap*, décrits dans la section 7.4, permettent de valider la taille des échantillons utilisés.

Compte-tenu de la durée d'une évaluation du modèle, seuls les indices d'ordre 1 et totaux ont été estimés ici. Cela nécessite d'utiliser trois estimateurs \widehat{V} , \widehat{V}_i , et \widehat{V}_i^{tot} des variances V , V_i , et V_i^{tot} . Plusieurs approximations des expressions intégrales des indices de Sobol' ont été proposées (Jansen, 1999; Sobol', 2001; Homma et Saltelli, 1996; Sobol', 1993). Ces différentes approches ont été comparées par Saltelli et coll. (2010) qui proposent un couple de formules permettant d'estimer les indices d'ordre 1 et totaux avec $(p + 2)N$ évaluations du modèle et une convergence optimisée. Ce sont les méthodes décrites dans cet article comme « meilleures pratiques » pour l'estimation conjointe des indices d'ordre 1 et totaux qui ont été utilisées ici.

7.3.1. Estimation des indices d'ordre 1

Étant donné un échantillon des variables d'entrée, représenté par une matrice \mathbf{A} de taille $N \cdot p$, la variance totale V peut être estimée en utilisant la formule usuelle :

$$\widehat{V} = \frac{1}{N} \sum_{k=1}^N f^2(\mathbf{A})_k, \quad (7.16)$$

où \mathbf{A}_k désigne la $k^{\text{ème}}$ ligne de l'échantillon.

En revanche, le numérateur des indices d'ordre 1,

$$V_i = \text{var}(\mathbb{E}(Y|X_i)) = \int_0^1 \left(\int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i} \right)^2 dx_i - f_0^2, \quad (7.17)$$

fait intervenir une somme double qui comporte un nombre bien trop important de termes.

Un expédient à ce problème est de développer le terme carré et de transformer les multiples intégrales à $p - 1$ dimensions en une unique intégrale à $2p - 1$ dimensions en utilisant \mathbf{x}^* , une réalisation de \mathbf{X} indépendante de \mathbf{x} (Homma et Saltelli, 1996) :

$$\left(\int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i} \right)^2 = \int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i} \int_0^1 f(x_i, \mathbf{x}_{-i}^*) d\mathbf{x}_{-i}^*, \quad (7.18)$$

puis

$$\int_0^1 \left(\int_0^1 f(\mathbf{x}) d\mathbf{x}_{-i} \right)^2 dx_i = \int_0^1 f(\mathbf{x}) f(x_i, \mathbf{x}_{-i}^*) d\mathbf{x} d\mathbf{x}_{-i}^* \quad (7.19)$$

Similairement, l'indépendance de \mathbf{x} et \mathbf{x}^* , permet d'exprimer f_0^2 de la façon suivante (Saltelli, 2002) :

$$f_0^2 = \int_0^1 f(\mathbf{x}) f(\mathbf{x}^*) d\mathbf{x}_{-i} d\mathbf{x}_{-i}^*. \quad (7.20)$$

Soit \mathbf{B} un échantillon de \mathbf{X} indépendant de \mathbf{A} et $\mathbf{A}_{\mathbf{B}}^{(i)}$ la matrice obtenue en substituant à la $i^{\text{ème}}$ colonne de \mathbf{A} la $i^{\text{ème}}$ colonne de \mathbf{B} .

La différence des équations (7.19) et (7.20) permet de déduire une estimation de V_i :

$$\widehat{V}_i = \frac{1}{N} \sum_{k=1}^N f(\mathbf{B})_k (f(\mathbf{A}_{\mathbf{B}}^{(i)})_k - f(\mathbf{A})_k). \quad (7.21)$$

7.3.2. Estimation des indices totaux

L'estimateur des indices totaux développé ci-après a été proposé initialement par Jansen (1999) puis généralisé par Sobol' (2001). On cherche une formule approchée de la quantité

$$V_i^{\text{tot}} = V - \text{var}(\mathbb{E}(Y|\mathbf{X}_{-i})). \quad (7.22)$$

La variance V s'exprime simplement :

$$V = \int_0^1 f(\mathbf{x})^2 d\mathbf{x} - f_0^2, \quad (7.23)$$

que l'on peut décomposer en

$$V = \frac{1}{2} \int_0^1 f^2(x_i, \mathbf{x}_{-i}) d\mathbf{x} + \frac{1}{2} \int_0^1 f^2(x_i^*, \mathbf{x}_{-i}) dx_i^* d\mathbf{x}_{-i} - f_0^2. \quad (7.24)$$

On peut par ailleurs, de la même façon que pour l'équation (7.19), écrire le second terme avec une unique intégrale :

$$\text{var}(E(Y|\mathbf{X}_{\rightarrow i})) = \int_0^1 f(\mathbf{x})f(x_i^*, \mathbf{x}) d\mathbf{x} dx_i^* - f_0^2. \quad (7.25)$$

On obtient, de même que précédemment, l'expression recherchée en factorisant la différence des équations (7.24) et (7.25) :

$$V_i^{tot} = \frac{1}{2} \int_0^1 [f(\mathbf{x}) - f(x_i^*, \mathbf{x}_{\rightarrow i})]^2 d\mathbf{x} dx_i^*. \quad (7.26)$$

On déduit finalement l'estimation suivante de V_i^{tot} :

$$\widehat{V}_i^{tot} = \frac{1}{2N} \sum_{k=1}^N [f(\mathbf{A})_k - f(\mathbf{A}_{\mathbf{B}}^{(i)})_k]^2. \quad (7.27)$$

Ainsi, l'estimation des indices d'ordre 1 et totaux par la méthode de Sobol' nécessite le calcul de $f(\mathbf{A})$ et $f(\mathbf{B})$ puis des $f(\mathbf{A}_{\mathbf{B}}^{(i)})$ pour $i \in \{1, \dots, p\}$, soit un total de $(2 + p) \cdot N$ évaluations du modèle.

7.3.3. Remarques sur la convergence

Les méthodes de Monte-Carlo classiques consistent à approximer l'espérance d'une variable aléatoire par la moyenne empirique d'un échantillon de réalisations indépendantes de cette variable. Dès lors que l'on peut simuler une distribution uniforme, la distribution de la variable peut être obtenue par exemple par la méthode de la transformée inverse ou la méthode de rejet (Robert et Casella, 2004, chap. 2). Les lois uniformes sont en général simulées à l'aide de *générateurs de nombres pseudo-aléatoires*. Ces algorithmes produisent, à partir d'une valeur initiale appelée graine, une suite de nombres reproduisant fidèlement le comportement statistique de réalisations indépendantes d'une variable distribuée uniformément. Les méthodes de Monte-Carlo reposent sur la loi forte des grands nombres (Saporta, 2006, chap. 12) qui permet d'établir que leur vitesse de convergence est dominée par la racine carré du nombre d'individus de l'échantillon, \sqrt{N} .

La vitesse de convergence des sommes des équations (7.16), (7.21), et (7.27) peut être nettement améliorée en générant les échantillons à l'aide de séquences de Sobol' plutôt qu'avec un générateur de nombres pseudo-aléatoires (Archer et coll., 1997). Les séquences de Sobol' sont des suites de nombres à discrédance faible (Sobol', 1998). On parle alors d'estimation par quasi-Monte-Carlo par opposition à l'estimation par méthode de Monte-Carlo classique. En d'autres termes, les séquences de Sobol' sont des séquences déterministes de nombres ayant l'*allure* de nombres aléatoires uniformément distribués mais qui remplissent l'espace plus efficacement, c'est-à-dire sans laisser de zones vides, ni dans l'espace complet, ni dans les projections sur des sous-espaces de plus petite dimension. Un échantillonnage utilisant de telles séquences permet d'obtenir une vitesse de convergence bornée par $\log(N)^p/N$ (Kucherenko, 2012). Le gain en efficacité est souvent

spectaculaire et l'utilisation de Sobol' pour l'estimation des indices de sensibilité est vivement recommandée (Saltelli et coll., 2010). En pratique, les échantillons \mathbf{A} et \mathbf{B} sont construits en générant une séquence de Sobol' de taille N sur \mathbb{I}^{2p} et en la divisant en deux : les p premières colonnes constituent un échantillon et les p suivantes, l'autre (Sobol', 2001).

Un autre point important est que la précision des estimateurs se dégrade fortement lorsque la moyenne, f_0 , est importante (Sobol', 2001). Il est nécessaire dans ce cas de faire une première estimation de la moyenne et de l'utiliser pour centrer la variable de sortie.

7.4. Intervalles de confiance pour les indices de Sobol'

La convergence des sommes estimées par quasi-Monte-Carlo peut être assez lente. Il est donc important d'apprécier l'incertitude associée aux indices de sensibilité calculés pour connaître la validité de leur classement et savoir si l'échantillon utilisé est assez important. Les estimateurs des indices de sensibilité sont eux-mêmes des variables aléatoires pour lesquelles on peut construire des intervalles de confiance. On verra plus loin que ces intervalles de confiance permettent en outre d'analyser certaines aberrations causées par la convergence partielle des sommes: indices négatifs, somme des indices totaux supérieure à 1 ou indice d'ordre 1 supérieur à l'indice total correspondant.

Les méthodes de *bootstrap* développées par Efron et Tibshirani (1993) permettent, à partir de l'échantillon utilisé pour calculer un estimateur, de construire des intervalles de confiance sans avoir à faire d'hypothèse sur sa distribution. Le principe des méthodes de *bootstrap* est expliqué dans la section 7.4.1 et illustré par leur application à la détermination de l'erreur standard d'un estimateur. La construction d'intervalles de confiance est ensuite abordée dans la section 7.4.2.

7.4.1. Principe général du *bootstrap*

Soit F une distribution de probabilité inconnue et \mathbf{A} un échantillon indépendant de taille N distribué selon F . Une application ϕ permet de calculer à partir de \mathbf{A} un estimateur $\hat{\theta}$ d'une statistique d'intérêt : $\hat{\theta} = \phi(\mathbf{A})$. On cherche à caractériser la qualité de l'estimateur $\hat{\theta}$, par exemple en estimant son erreur type ou en construisant un intervalle de confiance. Le *bootstrap* permet d'atteindre cet objectif de façon automatique et sans information de plus que celle apportée par \mathbf{A} . Il consiste à estimer les propriétés de $\hat{\theta}$ par simulation en utilisant une approximation de la distribution F . Dans sa version dite *non paramétrique*, seule utilisée ici, l'approximation choisie est la fonction de répartition empirique \hat{F} , c'est-à-dire la distribution discrète attribuant à chacun des individus de l'échantillon une probabilité $1/N$.

Un *échantillon bootstrap* est un échantillon de taille N distribué selon \hat{F} . D'après la définition de la distribution empirique, un échantillon *bootstrap* peut être obtenu par tirage avec remise dans \mathbf{A} . On appelle *réplication bootstrap* de $\hat{\theta}$, la grandeur image d'un échantillon *bootstrap* par l'application ϕ . En notant $\mathbf{A}^{(b)}$ un échantillon *bootstrap*, une *réplication bootstrap* s'écrit $\hat{\theta}^{(b)} = \phi(\mathbf{A}^{(b)})$.

Soit $\text{err}_F(\theta)$ l'erreur type de l'estimateur θ . L'estimation par *bootstrap* de $\text{err}_F(\theta)$ consiste à appliquer l'opérateur err à la distribution empirique \hat{F} au lieu de F . On appelle *estimateur bootstrap idéal* de $\text{err}_F(\theta)$ l'erreur type de θ calculée pour un échantillon de taille N distribué selon \hat{F} . En dehors du cas où θ est la moyenne, il n'existe pas de formule permettant de calculer l'estimateur *bootstrap* idéal. Par ailleurs, la taille de la population d'échantillons *bootstrap* vaut $\binom{2N-1}{N}$ et croît donc très vite avec N . Par exemple, la population compte 92 378 individus pour $N = 10$ et est de l'ordre de 10^{58} pour $N = 100$. Il est par conséquent impossible de calculer l'estimateur *bootstrap* idéal en utilisant la population des réplifications *bootstrap*. On peut cependant en obtenir une bonne approximation en calculant l'écart type d'un échantillon de réplifications *bootstrap*. L'intérêt de cette approche est que le calcul d'une réplification *bootstrap* prend en général au plus quelques secondes quand celui de $\hat{\theta}$ peut être beaucoup plus long.

7.4.2. Intervalles de confiance *bootstrap* quantiles

Construire un intervalle de confiance pour un estimateur $\hat{\theta}$ consiste à trouver deux bornes $\hat{\theta}_{inf}$ et $\hat{\theta}_{sup}$ telles que les deux événements $\theta < \hat{\theta}_{inf}$ et $\theta > \hat{\theta}_{sup}$, où θ est la valeur vraie de la statistique estimée par $\hat{\theta}$, aient une probabilité donnée faible. Des intervalles de confiance peuvent être construits en formulant des hypothèses sur la distribution de l'estimateur. Si l'on suppose par exemple que $\hat{\theta}$ suit une loi proche d'une loi gaussienne et que son erreur type vaut $\hat{\sigma}$ alors, un intervalle de confiance simple pour une probabilité seuil α est $[\hat{\theta} - z^{(1-\alpha)} \hat{\sigma}, \hat{\theta} - z^{(\alpha)} \hat{\sigma}]$ où $z^{(\alpha)}$ désigne le quantile α de la distribution gaussienne standard.

Lorsque l'on dispose d'un échantillon conséquent, le *bootstrap* permet de relâcher les hypothèses sur la distribution de $\hat{\theta}$. On définit l'intervalle *bootstrap idéal* de longueur $1 - 2\alpha$ comme étant l'intervalle séparant les quantiles α et $1 - \alpha$ d'une réplification *bootstrap* : $[\theta^{(b)(\alpha)}, \theta^{(b)(1-\alpha)}]$. De même que dans le paragraphe précédent, on se sert alors d'un échantillon de réplifications *bootstrap* pour approximer cet intervalle. Étant données N_b réplifications *bootstrap* de $\hat{\theta}$, l'intervalle *bootstrap* de longueur $1 - 2\alpha$ est borné par les quantiles empiriques α et $1 - \alpha$ de l'échantillon, c'est-à-dire ses $N_b \cdot \alpha^{\text{ème}}$ et $N_b \cdot (1 - \alpha)^{\text{ème}}$ valeurs, en arrondissant si nécessaire.

8. Analyse en composantes principales

Considérons un tableau de données de taille $N \times p$ dont les lignes représentent les individus d'un échantillon de réalisations de p variables aléatoires. Le cadre envisagé ici est général et les p variables peuvent aussi bien représenter les valeurs prises par une mesure à des instants successifs que des grandeurs de natures fondamentalement différentes. Dans le premier cas, les variables sont ordonnées naturellement et l'on pourra les représenter par N courbes faites de p segments. Cependant, ce mode de représentation ne fait pas apparaître clairement les liens qu'il peut y avoir entre des variables consécutives, voire éloignées. Dans le cas général, une autre façon intuitive de représenter ces données est de définir un repère orthonormal de \mathbb{R}^p et d'associer à chaque axe l'une des p variables. Les N individus de l'échantillon sont ainsi représentés par un nuage de N points de \mathbb{R}^p . Il est délicat de représenter graphiquement des espaces de dimension supérieure à 3. Ce nuage de \mathbb{R}^p pourra être représenté par exemple par $p(p-1)/2$ nuages de points à deux dimensions, c'est-à-dire par ses projections orthogonales sur les plans contenant chacune des paires d'axes. Cette représentation, quoique informative sur les corrélations entre variables, peut être malcommode si p est élevé. Certains graphiques seront peu informatifs s'ils font intervenir des variables variant peu selon les individus. À l'inverse, certains traits caractéristiques du nuage de points de \mathbb{R}^p faisant intervenir plus de 2 variables seront décomposés dans plusieurs graphiques et donc difficilement discernables.

L'objectif de l'analyse en composantes principales est de trouver un repère de \mathbb{R}^p permettant une représentation du nuage la plus informative possible au sens où il maximise les variations des projections sur les sous-espaces contenant ses axes (Lebart et coll., 2006, chap. 3). Elle consiste à orienter les axes du repère selon les directions de \mathbb{R}^p dans lesquelles le nuage de points est le plus étendu. Ces directions sont qualifiées de *principales* et les variables aléatoires associées aux projections selon les directions principales sont appelées *composantes principales* du tableau de données. Précisons que les directions principales sont recherchées séquentiellement. La première est le vecteur directeur de la droite sur laquelle la projection des données est la plus étendue. La seconde est ensuite choisie orthogonale à la première de façon à maximiser l'aire de la projection des données sur le plan engendré par les deux premières directions principales. On continue de la sorte jusqu'à obtenir une base de l'espace \mathbb{R}^p . Il y a donc autant de composantes principales que de variables initiales. Remarquons enfin que maximiser l'étendue de la projection selon un sous-espace est équivalent à minimiser la distance des points du nuage à ce sous-espace. Ainsi, le sous-espace engendré par les k premières composantes est celui qui « contient au mieux » les points du nuage.

Une définition plus formelle des composantes principales reposant sur la notion de variance est donnée dans la section 8.1. La section 8.2 détaille comment cette définition se décline pour un échantillon de données et fait le lien avec l'approche géométrique esquissée

en introduction. Enfin, la section 8.3 aborde l'utilisation de l'analyse en composantes principales pour réduire la dimension de la sortie fonctionnelle d'un modèle en vue d'une analyse de sensibilité.

8.1. Variance des composantes principales

Dans l'introduction de ce chapitre, les colonnes du tableau de données ont été associées à des variables aléatoires. Cela permet de donner aux concepts géométriques énoncés ci-dessus une interprétation statistique à l'aide de la notion de variance. En effet, l'étendue de la projection du nuage selon une direction principale est mesurée par la variance de la composante principale associée. Le développement classique qui suit peut être retrouvé en détail dans le livre de Jolliffe (2002, chap. 1).

Soit \mathbf{X} un vecteur aléatoire de dimension p et Σ sa matrice de covariance. Soit \mathbf{c}_1 la première direction principale ; la première composante principale est donc $\mathbf{c}'_1 \mathbf{X}$. Par définition, \mathbf{c}_1 est le vecteur qui maximise la variance $\text{var}(\mathbf{c}'_1 \mathbf{X}) = \mathbf{c}'_1 \Sigma \mathbf{c}_1$. Ajoutons une condition de normalisation, $\mathbf{c}'_1 \mathbf{c}_1 = 1$, et le problème de maximisation peut être résolu par la méthode de Lagrange. Avec λ un coefficient de Lagrange, on cherche à maximiser l'expression :

$$\mathbf{c}'_1 \Sigma \mathbf{c}_1 - \lambda(\mathbf{c}'_1 \mathbf{c}_1 - 1). \quad (8.1)$$

En dérivant selon \mathbf{c}_1 , on obtient l'expression suivante :

$$\Sigma \mathbf{c}_1 = \lambda \mathbf{c}_1, \quad (8.2)$$

d'où l'on déduit que \mathbf{c}_1 est un vecteur propre de Σ associé à la valeur propre λ . Or, la grandeur à maximiser est précisément $\mathbf{c}'_1 \Sigma \mathbf{c}_1 = \lambda$. La première direction principale est donc le vecteur propre de la matrice de covariance de \mathbf{X} associé à sa plus grande valeur propre, notée λ_1 . Cette valeur propre est égale à la variance de la première composante principale.

L'orthogonalité de la seconde direction principale s'écrit $\mathbf{c}'_2 \mathbf{c}_1 = 0$. En multipliant cette expression par λ_1 et en utilisant le fait que \mathbf{c}_1 est un vecteur propre de Σ , on déduit l'expression équivalente suivante :

$$\mathbf{c}'_2 \Sigma \mathbf{c}_1 = 0. \quad (8.3)$$

Cette dernière traduit l'absence de corrélation entre les deux premières composantes principales. Du fait de cette nouvelle contrainte, un second coefficient de Lagrange, μ , est nécessaire pour résoudre le problème de maximisation de la variance de la seconde composante principale $\mathbf{c}'_2 \mathbf{X}$:

$$\mathbf{c}'_2 \Sigma \mathbf{c}_2 - \lambda(\mathbf{c}'_2 \mathbf{c}_2 - 1) - \mu \mathbf{c}'_2 \mathbf{c}_1. \quad (8.4)$$

En dérivant par rapport à \mathbf{c}_2 et en multipliant à gauche par \mathbf{c}'_1 , on obtient :

$$\mathbf{c}'_1 \Sigma \mathbf{c}_2 - \lambda \mathbf{c}'_1 \mathbf{c}_2 - \mu \mathbf{c}'_1 \mathbf{c}_1 = 0. \quad (8.5)$$

Les deux premiers termes de cette équation sont nuls du fait de la condition de non-corrélation des deux premières composantes principales. Par ailleurs, la condition de normalisation implique que $\mathbf{c}'_1 \mathbf{c}_1 = 1$, d'où l'on déduit que μ est nul. Revenant à la formulation initiale du problème de maximisation, on obtient l'expression suivante :

$$\Sigma \mathbf{c}_2 - \mu \mathbf{c}_2 = 0, \quad (8.6)$$

d'où l'on déduit, de même que précédemment, que la deuxième direction principale est donnée par le vecteur propre de Σ associé à sa deuxième valeur propre la plus importante.

Cette procédure peut être répétée pour les p composantes principales. On retiendra en conclusion que les composantes principales de p variables aléatoires sont les combinaisons linéaires de ces variables obtenues par projection selon la base des vecteurs propres de leur matrice de covariance, triés par valeurs propres décroissantes. Les valeurs propres de la matrice de covariance représentent la part de la variance des variables initiales expliquée par chaque composante principale.

Les valeurs propres nulles ou multiples nécessitent certaines précautions. On se reportera au chapitre 2.4 du livre de Jolliffe (2002) pour plus d'explications au sujet de ces cas, rares en pratique, ne remettant pas en cause la définition générale des composantes principales présentée ici.

8.2. Composantes principales d'un échantillon

Le raisonnement exposé dans la section précédente s'applique aux populations de variables aléatoires mais on ne dispose en général que d'un échantillon. Une démarche identique peut-être menée en remplaçant, dans la définition des composantes principales, la variance des variables aléatoires X_1, \dots, X_p par la variance des colonnes d'un échantillon \mathbf{A} . La matrice de covariance de \mathbf{X} est alors naturellement remplacée par la matrice de covariance de l'échantillon,

$$\Sigma_{\mathbf{A}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{A} - \bar{\mathbf{A}})' (\mathbf{A} - \bar{\mathbf{A}}), \quad (8.7)$$

où $\bar{\mathbf{A}}$ est la matrice de terme général $\bar{a}_{ij} = \frac{1}{N} \sum_{k=1}^N a_{kj}$. En désignant dorénavant par \mathbf{c}_i le $i^{\text{ème}}$ vecteur propre de $\Sigma_{\mathbf{A}}$, la $i^{\text{ème}}$ composante principale s'écrit encore $\mathbf{c}'_i \mathbf{X}$.

On appelle *matrice des scores* des composantes principales de \mathbf{A} la matrice

$$\mathbf{S} = \mathbf{X} \mathbf{C}, \quad (8.8)$$

où \mathbf{C} est la matrice contenant en colonnes les directions principales. Le terme général de la matrice des scores, s_{ij} , est le score du $i^{\text{ème}}$ individu de l'échantillon associé à la $j^{\text{ème}}$ composante principale. Les individus de l'échantillon sont des vecteurs de \mathbb{R}^p et leurs scores sont leurs coordonnées dans la base des directions principales. La propriété 1 illustre l'optimalité de la projection sur le sous-espace engendré par les premières directions principales au sens qu'elle préserve au mieux la « forme du nuage ».

Propriété 1. Soit \mathbf{T} une matrice orthonormale de dimension $p \times k$. Les lignes de $\mathbf{T}'\mathbf{A}$ sont les projections des lignes de \mathbf{A} sur un sous-espace de dimension k .

La somme des carrés des distances des lignes de \mathbf{A} à ce sous-espace est minimisée lorsque $\mathbf{T} = \mathbf{C}_k$, où \mathbf{C}_k est la matrice constituée en colonnes des k premières directions principales.

Cette propriété est une formalisation de la description géométrique donnée en introduction : le sous-espace engendré par les k premières directions principales est celui qui « contient au mieux » les données. Elle peut d'ailleurs être choisie comme définition des composantes principales, la définition fondée sur la variance donnée dans la section 8.1 apparaissant alors comme une propriété. Cette approche alternative est proposée dans le livre de Lebart et coll. (2006, chap. 3). Enfin, cette propriété géométrique peut être formulée de façon équivalente en termes statistiques. Sous cette forme, elle s'intègre à la description de la population exposée dans la section 8.1.

Propriété 2. Soit \mathbf{T} une matrice orthonormale de dimension $p \times k$. Soit σ_i^2 la variance résiduelle lors de la prédiction de X_i par une fonction linéaire de $\mathbf{T}'\mathbf{X}$.

Le minimum de $\sum_{i=1}^p \sigma_i^2$ est obtenu pour $\mathbf{T} = \mathbf{C}_k$.

8.2.1. Composantes principales normalisées

Lorsque les variables ont des domaines de variations éloignés, il peut être nécessaire de les normaliser avant de procéder à l'analyse en composantes principales. La variance des variables prenant des valeurs fortes sera prépondérante sur celles des autres variables, et elles accapareront les premières composantes principales, indépendamment de leurs variations relatives. Ce cas se présente notamment lorsque les variables représentent des grandeurs s'exprimant dans des unités différentes.

La normalisation la plus couramment appliquée consiste à diviser les variables centrées par leur écart type, ce qui revient à travailler avec la matrice de corrélation au lieu de la matrice de covariance. Cela ne change pas la façon de calculer les composantes principales mais notons néanmoins qu'il n'y a pas de lien simple entre les composantes principales des variables initiales et celles des variables standardisées.

L'avantage des composantes principales calculées en variance est qu'elles sont directement liées aux variables manipulées, aussi s'interprètent-elles plus aisément. Cependant, l'élimination de l'effet d'échelle grâce la matrice de corrélation est souvent souhaitable.

Une propriété intéressante des composantes principales calculées en corrélation est obtenue en imposant aux vecteurs propres une norme égale à la racine carrée de leur valeur propre associée plutôt qu'une norme unité :

$$\forall k \in \{1, \dots, p\}, \quad \mathbf{c}'_k \mathbf{c}_k = \lambda_k. \quad (8.9)$$

Sous cette condition, la $j^{\text{ème}}$ composante du $k^{\text{ème}}$ vecteur propre n'est autre que le coefficient de corrélation entre la $j^{\text{ème}}$ variable standardisée et la $k^{\text{ème}}$ composante principale.

8.3. Analyse de sensibilité d'une sortie fonctionnelle

Les méthodes d'analyse de sensibilité présentées dans le chapitre 7 s'appliquent au cas d'une sortie scalaire. On peut analyser séparément avec la même procédure chacune des composantes d'une sortie multivariée. Dans le cas d'une sortie fonctionnelle, ou plus précisément, de courbes échantillonnées à des instants fixes, on peut calculer ainsi des indices de sensibilité à chaque pas de temps. Cependant, ces indices de sensibilité séquentiels sont parfois redondants et leur nombre important peut être un obstacle à l'interprétation. Cette approche ne tient pas compte de la corrélation entre les composantes de la sortie.

L'analyse en composantes principales permet de définir de nouvelles variables de sortie en regroupant les variables initiales de façon à faire ressortir les traits caractéristiques de la sortie. Les deux paragraphes suivants décrivent les spécificités de l'analyse en composantes principales appliquée à des courbes échantillonnées et la façon dont elle peut être mise à profit pour l'analyse de sensibilité.

8.3.1. Cas de courbes échantillonnées

Dans le cas de l'étude de la réponse dynamique des g. v., les valeurs des variables sont obtenues en échantillonnant un processus continu et chaque individu est une représentation discrète d'une fonction sous-jacente. Les variables sont donc d'une part, ordonnées naturellement selon l'ordre chronologique et d'autre part, fortement corrélées avec leurs voisines.

Remarquons d'abord que les directions principales peuvent elles-mêmes être vues comme des représentations discrètes de fonctions. La projection sur le sous-espace engendré par les k premières composantes principales est une version discrète d'une décomposition fonctionnelle des réponses du modèle sur une base de k fonctions. Les scores des composantes principales sont les coefficients de cette décomposition. Soit \mathbf{A} un échantillon de réponses du modèle. Les lignes de \mathbf{A} contiennent les valeurs du n. g. l. aux instants d'échantillonnage pour un transitoire et une configuration de colmatage donnés. On note \mathbf{c}_0 le vecteur des moyennes des colonnes de \mathbf{A} . Soit \mathbf{y} une ligne de \mathbf{A} et (s_i) la famille de ses scores, sa projection selon les directions principales s'écrit

$$\mathbf{y} = \mathbf{c}_0 + \sum_{i=1}^p s_i \mathbf{c}_i, \quad (8.10)$$

le nombre de variables p étant ici égal au nombre d'instant d'échantillonnage. On peut considérer que les directions de projection sont des échantillonnages d'une famille de fonctions (ϕ_i) et que le vecteur des moyennes est lui aussi un échantillonnage d'une fonction ϕ_0 . En notant y la fonction représentant la réponse du n. g. l. et \mathbf{t} le vecteur contenant les instants d'échantillonnage, on obtient l'expression suivante :

$$y(\mathbf{t}) = \phi_0(\mathbf{t}) + \sum_{i=1}^p s_i \phi_i(\mathbf{t}). \quad (8.11)$$

Les vecteurs directeurs des directions principales peuvent être représentés par des courbes en reliant simplement les points successifs et en conservant l'ordre chronologique des variables initiales. Cette représentation insiste sur l'interprétation de la projection comme une décomposition fonctionnelle. Une autre interprétation de l'équation (8.10) consiste à considérer chacune des directions de projection comme un « mode de perturbation de la moyenne » (Ramsay et Silverman, 2005, chap. 8.3). On représente sur un même graphique le vecteur des moyennes \mathbf{c}_0 et ce même vecteur auquel on ajoute puis soustrait une petite contribution de l'une des directions de projection. Notons que l'équation (8.10) reste valide si l'on remplace l'échantillon \mathbf{A} par sa version normalisée selon la procédure décrite dans la section 8.2.1. Le passage aux variables normalisées rend cependant caducs les graphiques des perturbations de la moyenne qui deviennent difficilement interprétables. En revanche, la contrainte de normalisation alternative proposée à la fin de la section 8.2.1 permet de comparer l'évolution temporelle de la corrélation des composantes principales avec les variables initiales.

Outre les remarques qui viennent d'être faites, le cadre particulier de variables représentant des données fonctionnelles a donné lieu à des variantes de l'analyse en composantes principales. L'une d'entre elles consiste à considérer la décomposition donnée dans l'équation (8.11) en faisant abstraction des instants d'échantillonnage. L'analyse en composantes principales classique consiste à rechercher les poids de combinaisons linéaires du type $\mathbf{b}'\mathbf{X} = \sum_i b_i X_i$. L'idée de l'analyse en composantes principales *fonctionnelles* (Ramsay et Silverman, 2005, chap. 8) est de remplacer l'indice discret, i , par un indice continu, t . La somme précédente est alors remplacée par une intégrale et le vecteur de poids \mathbf{b} par une fonction de poids $b(t)$:

$$\int b(t) x(t) dt. \tag{8.12}$$

Partant de là, les composantes principales fonctionnelles se définissent similairement aux composantes principales classiques en remplaçant chaque grandeur discrète par son équivalent fonctionnel. La mise en œuvre de l'analyse en composantes principales fonctionnelles est néanmoins plus complexe. En effet, on dispose rarement au départ d'expressions fonctionnelles des données et une première étape de lissage et interpolation ou de projection sur une base de fonctions est nécessaire (Ramsay et Silverman, 2005, chap. 3 et 4). La résolution du problème aux valeurs propres est aussi plus délicate dans le cadre fonctionnel. Les différentes stratégies possibles consistent à reformuler le problème sous une forme matricielle approchée, par exemple en réalisant une discrétisation, en utilisant une nouvelle projection sur une base de fonctions ou encore en ayant recours à des méthodes d'approximation par quadrature (Ramsay et Silverman, 2005, chap 8.4). Cette complexité additionnelle permet d'introduire de façon flexible des hypothèses relatives aux fonctions sous-jacentes et d'éliminer d'éventuels bruits polluant les données tout en préservant le caractère exploratoire de la méthode. L'intérêt de cette opération est de préserver la qualité de l'estimation de la matrice de covariance de la population par la matrice calculée à l'aide de l'échantillon. En effet, il arrive que cette estimation se dégrade fortement avec la dimension lorsque le nombre d'observations est relativement faible par rapport à la dimension, que les instants d'échantillonnage sont espacés et

répartis inégalement ou encore lorsque les données sont fortement bruitées. Viviani et coll. (2005) proposent trois cas d'application dans le domaine médical pour lequel l'analyse en composantes principales fonctionnelles donne des résultats plus satisfaisants que l'analyse en composantes principales classique. Dans les trois cas, la dimension est très élevée par rapport au nombre d'observations.

Dans le cas des réponses de n. g. l. simulées à l'aide du modèle thermohydraulique, le nombre d'individus des échantillons utilisés pour les analyses en composantes principales est au moins 7 fois supérieur au nombre d'instantanés d'échantillonnage. Par ailleurs, les instantanés d'échantillonnage sont proches et régulièrement espacés. Enfin, les courbes simulées ne présentent aucun bruit significatif. Aussi, il ne semble pas nécessaire d'avoir recours à l'analyse en composantes principales fonctionnelles. Un moyen simple de tester la robustesse de l'analyse en composantes principales classique est de vérifier que les composantes principales obtenues (et leurs valeurs propres, dans le cas des composantes normalisées) ne fluctuent pas trop à mesure que l'on augmente la taille de l'échantillon.

La situation est moins favorable pour l'analyse en composantes principales réalisée à partir de mesures de n. g. l. car le ratio entre le nombre d'observations et le nombre de variables est moins important et que les données sont brouillées par le bruit de la mesure de n. g. l. Cependant, on verra dans le chapitre 10 que la similitude importante des directions principales calculées à partir des mesures avec celles calculées à partir des simulations laisse penser que, là encore, l'analyse en composantes principales classique est suffisante.

8.3.2. Application à l'analyse de sensibilité d'une sortie fonctionnelle

Une sortie fonctionnelle d'un modèle comporte souvent des motifs ou au moins une structure globale. Ces traits caractéristiques apparaissent ou disparaissent et sont modifiés dans leur forme et leur amplitude lorsque l'on fait varier les entrées du modèle. Dans le cas où la sortie est une série temporelle, ils représentent en quelque sorte la dynamique du modèle, par opposition à la valeur de la sortie à une date donnée. Une approche complémentaire aux indices séquentiels décrits dans le chapitre 7 a été proposée par Campbell et coll. (2006). Elle consiste à projeter les sorties fonctionnelles sur une base de fonctions judicieusement choisies afin de faire ressortir les traits caractéristiques de la dynamique du modèle. Le nombre d'indices est restreint par la taille de la base, facilitant ainsi grandement l'interprétation, et chaque « effet dynamique » peut être représenté par un jeu d'indices distinct.

Dans certains cas, une base de projection s'impose d'elle-même. En présence de phénomènes périodiques, par exemple, une décomposition de Fourier est tout indiquée. L'analyse en composantes principales permet quant à elle de construire une base de projection de façon adaptative. On sélectionne les premières directions principales de l'échantillon constitué de l'ensemble des simulations réalisées pour l'analyse de sensibilité. Les scores associés sont les nouvelles variables de sortie et l'analyse de sensibilité peut être mise en œuvre comme décrit dans le chapitre 7.

L'analyse en composantes principales s'articule bien avec l'approche choisie pour l'analyse de sensibilité car ces méthodes reposent toutes deux sur le concept de variance

de la sortie. Les valeurs propres constituent une indication du nombre de directions à conserver et permettent de contrôler la part de la variance de la sortie laissée de côté. La propriété 2 assure l'optimalité de la projection au sens qu'elle maximise la part de la variance de la sortie représentée par sa projection selon une base de la taille donnée. Si la décroissance des valeurs propres est rapide, un nombre restreint d'indices de sensibilité suffira à rendre compte de la majeure partie de la variance de sortie. Le procédé de construction des composantes principales permet en outre de se prémunir d'un choix de base inadapté qui distribuerait un effet dynamique dans plusieurs jeux d'indices de sensibilité. Les indices de sensibilité calculés avec les scores des composantes principales seront appelés dans la suite *indices de sensibilité principaux*, par opposition aux indices de sensibilité séquentiels.

L'analyse de sensibilité de sorties fonctionnelles par la méthode de Sobol' et l'analyse en composantes principales a été appliquée par Lamboni et coll. (2009, 2010) à des modèles agronomiques. Ils proposent par ailleurs de mesurer l'impact global des variables d'entrée par des *indices de sensibilité généralisés*. Ces derniers sont définis comme les moyennes des indices de sensibilité de chacune des composantes principales sélectionnées, pondérés par les valeurs propres correspondantes. Les indices de sensibilité généralisés n'ont pas été utilisés pour le diagnostic du colmatage du fait du nombre très restreint de composantes principales ayant des valeurs propres non négligeables.

9. Régression inverse par tranche

L'analyse en composantes principales présentée dans le chapitre précédent peut être utilisée comme méthode de réduction de dimension. Elle permet de trouver un sous-espace sur lequel projeter un échantillon de données en ne perdant qu'une partie contrôlée et restreinte de sa variance. L'analyse en composantes principales utilise un échantillon de données unique dont elle révèle des traits caractéristiques. Dans le domaine de l'apprentissage automatique, on parle de méthode d'apprentissage *non supervisé* (Hastie et coll., 2001, chap. 1). Au contraire, les méthodes d'apprentissage *supervisé* supposent un échantillon d'entrée lié, par une relation inconnue, à un échantillon de sortie. L'échantillon de sortie guide alors l'exploration de l'échantillon d'entrée. Dans le contexte de la réduction de dimension, on cherchera à réduire la dimension des données d'entrée en conservant l'information qu'elles portent à propos de la sortie. Ce type de méthode convient bien pour aborder la problématique du diagnostic : on cherche une expression simple des variables décrivant la dégradation en se laissant guider par ce que l'on peut observer de son effet sur le g. v.

On peut envisager la réduction de dimension de deux façons. La *sélection* de traits caractéristiques consiste à éliminer les variables ne contribuant pas significativement aux effets que l'on cherche à analyser. Les méthodes de criblage évoquées dans le chapitre 7 sur l'analyse de sensibilité relèvent en quelque sorte de cette catégorie. L'*extraction* de traits caractéristiques consiste quant à elle à déduire des variables initiales de nouvelles variables plus adaptées à la description des effets qui nous intéressent. L'analyse en composantes principales fait partie de cette seconde catégorie. C'est cette seconde approche qui a été choisie pour le diagnostic. En effet, les variables décrivant une dégradation des g. v. résultent d'un découpage spatial a priori. Aussi, elles sont toutes de même nature et il paraît naturel de les combiner entre elles. En revanche, rien ne permet de supposer que certaines d'entre elles sont les variables « optimales » recherchées.

La SIR, est une méthode d'apprentissage supervisé d'extraction de traits caractéristiques mise au point par Li (1991); Duan et Li (1991). La « conservation de l'information » s'interprète ici à l'aide de la distribution conditionnelle : conditionner la sortie par les données réduites ou par les données initiales est équivalent. Cette méthode a été choisie pour sa simplicité d'utilisation, sa robustesse et la richesse d'interprétation de ses résultats. En effet, elle ne nécessite pas la mise en œuvre de méthodes d'ajustement ou de lissage, elle est fondée sur une hypothèse peu contraignante sur le « lien inconnu » entre l'entrée et la sortie et les traits caractéristiques extraits sont directement liés aux variables d'entrée.

Le principe général de la méthode est présenté dans la section 9.1. L'algorithme qui en découle est exposé dans la section 9.2. Enfin, la section 9.3 aborde l'adaptation de la méthode au cas où la sortie du modèle est multidimensionnelle.

9.1. Principe général de la régression inverse par tranches

Soit \mathbf{X} un vecteur aléatoire dont les p composantes sont appelées variables d'entrée, et Y une variable de sortie scalaire. Dans le cas du g. v., ces deux variables aléatoires correspondent à l'entrée et à la sortie du modèle thermohydraulique. Ce contexte est analogue à celui mis en place au début du chapitre 7 traitant de l'analyse de sensibilité. L'idée de départ de la SIR est de chercher un sous-espace de \mathbb{I}^p de dimension $K < p$ sur lequel on pourrait projeter \mathbf{X} sans perdre d'information sur le lien entre \mathbf{X} et Y . Plus précisément, on suppose que ce lien peut être décrit par le modèle suivant :

$$Y = g(\boldsymbol{\beta}'_1 \mathbf{X}, \boldsymbol{\beta}'_2 \mathbf{X}, \dots, \boldsymbol{\beta}'_K \mathbf{X}, \epsilon), \quad (9.1)$$

où g est une fonction arbitraire inconnue, ϵ un terme d'erreur de distribution inconnue indépendant de \mathbf{X} et les $\boldsymbol{\beta}_k$ des vecteurs inconnus. Le terme d'erreur ϵ dépend de la fonction g mais on ne cherchera pas à les identifier en pratique. Le modèle (9.1) est équivalent à la formulation suggérée en introduction : la distribution de Y conditionnellement à \mathbf{X} est identique à la distribution de Y conditionnellement aux $\boldsymbol{\beta}'_k \mathbf{X}$, ce que l'on notera

$$F_{Y|\mathbf{X}}(\cdot) = F_{Y|\boldsymbol{\beta}'_1 \mathbf{X}, \boldsymbol{\beta}'_2 \mathbf{X}, \dots, \boldsymbol{\beta}'_K \mathbf{X}}(\cdot). \quad (9.2)$$

Remarquons que ce modèle reste inchangé si l'on remplace les $\boldsymbol{\beta}_k$ par une famille quelconque de vecteurs engendrant le même sous-espace vectoriel. En somme, la réduction de dimension consiste à chercher une base de ce sous-espace, que l'on notera \mathcal{S} . On notera par la suite $\mathbf{P}_{\mathcal{S}}$ le projecteur orthogonal sur \mathcal{S} .

Une troisième façon de formuler le modèle de la SIR est de dire que Y est indépendant de \mathbf{X} conditionnellement à sa projection sur \mathcal{S} (Cook, 1998) :

$$Y \perp \mathbf{X} | \mathbf{P}_{\mathcal{S}} \mathbf{X}, \quad (9.3)$$

ce qui implique que l'on ne perd pas d'information si l'on effectue la régression de Y par les projections de \mathbf{X} sur \mathcal{S} plutôt que par \mathbf{X} .

L'espace engendré par les $\boldsymbol{\beta}_k$ est appelé espace de *réduction efficace de dimension*, ou *effective dimension reduction* (e. d. r.) en anglais, par Li (1991). Toute combinaison linéaire des $\boldsymbol{\beta}_k$ est quant à elle appelée *direction* de réduction efficace de dimension. Compte-tenu de l'indétermination des directions e. d. r., il convient de préciser que l'objectif de la SIR est de trouver l'espace e. d. r. de plus faible dimension. La question de l'existence et de l'unicité éventuelles de cet espace a été élucidée par Cook (1998, chap. 6) à l'aide de la notion de *sous-espace central* qui fait l'objet de la section suivante. Notons que la plupart des modèles de régression supposent $K = 1$ et des conditions additionnelles sur g (Li, 1991). Le modèle de la SIR peut donc être considéré comme peu contraignant. On verra dans la section 10.5.1 qu'un lien peut être établi entre la SIR et l'analyse de sensibilité lorsque K vaut 1.

9.1.1. Sous-espace central

Cook (1998) qualifie les sous-espaces vérifiant (9.3) de *sous-espaces de réduction de dimension pour $Y|\mathbf{X}$* . Remarquons que, d'une part, il n'y a pas unicité des sous-espaces

de réduction de dimension et que, d'autre part, l'espace d'entrée entier est toujours un sous-espace de réduction de dimension puisque les vecteurs de la base canonique vérifient l'identité (9.3). Un sous-espace de réduction de dimension de $Y|\mathbf{X}$ est dit *minimal* si sa dimension est inférieure ou égale à celle de tous les sous-espaces de réduction de dimension de $Y|\mathbf{X}$.

Un bref raisonnement par l'absurde permet de démontrer la propriété suivante (Cook, 1998) :

Propriété 3. Soit $\mathcal{S} = \text{Vect}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K)$ un sous-espace minimal de réduction de dimension pour $Y|\mathbf{X}$ et \mathbf{M} une matrice inversible de taille p .

Le sous-espace $\text{Vect}(\mathbf{M}^{-1}\boldsymbol{\beta}_1, \dots, \mathbf{M}^{-1}\boldsymbol{\beta}_K)$ est un sous-espace minimal de réduction de dimension pour $Y|\mathbf{MX}$.

La non-unicité des sous-espaces de réduction de dimension minimaux pose des problèmes pratiques pour la construction d'une méthode de réduction de dimension. Afin d'y remédier, Cook (1998) propose la notion de *sous-espace central*. Un sous-espace est dit central pour la régression $Y|\mathbf{X}$ si il est un sous-espace de réduction de dimension inclus dans tous les sous-espaces de réduction de dimension. Le sous-espace central existe si et seulement si l'intersection de tous les sous-espaces de réduction de dimension est elle-même un sous-espace de réduction de dimension. Dans ce cas, il est unique et égal à cette intersection. De cette définition découlent les deux propriétés suivantes :

Propriété 4. Le sous-espace central, lorsqu'il existe, est l'unique sous-espace minimal de réduction de dimension.

Le sous-espace minimal de réduction de dimension est une autre appellation du sous-espace e. d. r. qui se trouve être précisément défini de façon unique par le sous-espace central lorsqu'il existe.

Propriété 5. Soit $\mathcal{S} = \text{Vect}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K)$ le sous-espace central pour $Y|\mathbf{X}$ et \mathbf{M} une matrice inversible de taille p .

Le sous-espace $\text{Vect}(\mathbf{M}^{-1}\boldsymbol{\beta}_1, \dots, \mathbf{M}^{-1}\boldsymbol{\beta}_K)$ est le sous-espace central pour $Y|\mathbf{MX}$.

Enfin, Cook (1998) a démontré qu'il suffit que \mathbf{X} ait une densité à support convexe pour que le sous-espace central existe. Ainsi, sous cette condition peu contraignante, le sous-espace e. d. r. défini par Li (1991) existe et est unique, ce que l'on supposera vrai dans la suite de l'exposé.

9.1.2. Courbe de régression inverse

Les méthodes traditionnelles de régression sont mises en échec lorsque le nombre p de variables d'entrée excède la dizaine (Hastie et coll., 2001, chap. 2). La stratégie de la SIR pour contourner ces difficultés est de considérer à la place la régression inverse, $E(\mathbf{X}|Y)$, qui ne dépend que d'une variable. Lorsque y varie, la régression inverse, $E(\mathbf{X}|Y = y)$, décrit une courbe dans \mathbb{R}^p . Son centre, $E(E(\mathbf{X}|Y))$, est l'espérance de \mathbf{X} .

Considérons la condition suivante sur la distribution des variables d'entrée :

Condition 1 (Linéarité). Pour tout vecteur \mathbf{b} dans \mathbb{R}^p , l'espérance conditionnelle $E(\mathbf{b}'\mathbf{X}|\beta'_1\mathbf{X}, \dots, \beta'_K\mathbf{X})$ est linéaire en $\beta'_1\mathbf{X}, \dots, \beta'_K\mathbf{X}$.

Cette condition est difficile à vérifier car elle fait intervenir les β_k qui sont inconnus. Cependant, Li (1991) a montré qu'elle est respectée si \mathbf{X} a une distribution à symétrie elliptique. C'est le cas notamment lorsque \mathbf{X} est un vecteur gaussien. La condition de linéarité permet de contraindre la courbe de régression inverse à être inscrite dans l'espace e. d. r. C'est l'objet du théorème fondamental de la SIR :

Théorème 1. *Si le modèle (9.1) et la condition 1 sont respectés, la courbe de régression inverse centrée $E(\mathbf{X}|Y) - E(\mathbf{X})$ est inscrite dans le sous-espace engendré par les $\Sigma_{\mathbf{X}}\beta_k$, où $\Sigma_{\mathbf{X}}$ est la matrice de covariance de \mathbf{X} .*

D'après la propriété 5, on peut travailler sans perte de généralité avec l'entrée normalisée $\mathbf{Z} = \Sigma_{\mathbf{X}}^{-1/2}[\mathbf{X} - E(\mathbf{X})]$ et les directions de projection $\boldsymbol{\eta}_k = \beta_k \Sigma_{\mathbf{X}}^{1/2}$. Le théorème 1 se reformule ainsi : « si le modèle (9.1) et la condition 1 sont respectés, la courbe de régression inverse $E(\mathbf{Z}|Y)$ est inscrite dans le sous-espace engendré par les $\boldsymbol{\eta}_k$ ». Dès lors, la détermination des directions e. d. r. revient à chercher une base du sous-espace contenant la courbe de régression inverse. Or, on a vu dans le chapitre 8 que l'analyse en composantes principales permet de trouver le sous-espace « contenant le mieux » le nuage de points. Plus formellement, tout vecteur orthogonal à $\text{Vect}(\boldsymbol{\eta}_k)$ est dans le noyau de la matrice de covariance $\text{var}(E(\mathbf{Z}|Y))$. Ainsi, les K vecteurs propres de cette matrice associés à des valeurs propres non nulles sont des vecteurs de $\text{Vect}(\boldsymbol{\eta}_k)$ et la propriété 5 permet d'en déduire une base du sous-espace e. d. r. Ce résultat permet de construire un algorithme pour la recherche d'une base du sous-espace e. d. r. en approchant la courbe de régression inverse à partir d'un échantillon. L'idée générale est d'effectuer une analyse en composantes principales des moyennes par tranches de l'entrée \mathbf{X} . La différence fondamentale avec l'analyse en composantes principales conventionnelle est que les tranches ont été déterminées en utilisant la sortie Y .

9.2. Algorithme de la régression inverse par tranches

Soit \mathbf{A} un échantillon de taille N de \mathbf{X} et $f(\mathbf{A})$ la réalisation de la sortie Y associée. L'algorithme proposé par Li (1991) pour trouver une base du sous-espace e. d. r. est le suivant :

1. Normaliser \mathbf{A} à l'aide de sa matrice de covariance empirique $\widehat{\Sigma}_{\mathbf{A}}$ et obtenir ainsi la matrice \mathbf{C} dont la $i^{\text{ème}}$ ligne a pour expression

$$\mathbf{c}'_i = \widehat{\Sigma}_{\mathbf{A}}^{-1/2}(\mathbf{a}'_i - \bar{\mathbf{a}}'), \quad (9.4)$$

où \mathbf{a}_i est la $i^{\text{ème}}$ ligne de \mathbf{A} et $\bar{\mathbf{a}}$ le vecteur ligne contenant les moyennes des colonnes de \mathbf{A} ;

2. Découper le domaine de variation de $f(\mathbf{A})$ en H tranches, $\mathcal{T}_1, \dots, \mathcal{T}_H$, chacune contenant une proportion p_h des N observations ;

3. Calculer les moyennes par tranches, \mathbf{m}_h :

$$\forall h \in \{1, \dots, H\}, \mathbf{m}_h = \frac{1}{Np_h} \sum_{\{i|f(\mathbf{a}_i) \in \mathcal{T}_h\}} \mathbf{c}_i; \quad (9.5)$$

4. Calculer la matrice de covariance pondérée $\widehat{\Sigma}_{\mathbf{E}(\mathbf{Z}|Y)} = \sum_{h=1}^H p_h \mathbf{m}'_h \mathbf{m}_h$ (les \mathbf{m}_h sont des vecteurs lignes) ;
5. Trouver $(\widehat{\boldsymbol{\eta}}_k)$, la famille de vecteurs propres de $\widehat{\Sigma}_{\mathbf{E}(\mathbf{Z}|Y)}$ classés par valeurs propres décroissantes ;
6. Retourner $\widehat{\boldsymbol{\beta}}_k = \widehat{\Sigma}_{\mathbf{A}}^{-1/2} \widehat{\boldsymbol{\beta}}_k$ pour $k \in \{1, \dots, K\}$.

Les étapes 2 et 3 consistent à approcher la courbe de régression inverse en remplaçant Y par une variable catégorielle à H niveaux. Cette approximation peut sembler grossière à première vue mais comme on ne cherche que les directions de variation de la courbe sans se soucier de sa forme exacte, elle est suffisante. Pour la même raison, le choix du nombre de tranches n'est pas aussi sensible que celui, par exemple, du paramètre de lissage en régression non paramétrique. Le nombre de tranches H sera choisi supérieur à la dimension d'entrée p et inférieur à la moitié du nombre d'individus $N/2$ de façon à pouvoir estimer toutes les éventuelles directions e. d. r. et de s'assurer qu'il y a au moins 2 individus par tranche. En pratique, comme la SIR ne nécessite pas de calculs lourds, il est tout à fait possible de vérifier que le choix de H n'influence pas dramatiquement le résultat en testant plusieurs valeurs. Liquet et Saracco (2012) proposent d'utiliser le *bootstrap* pour sélectionner un nombre optimal de tranches. Ce point sera abordé plus en détails dans la section 9.2.1 concernant la détermination de la dimension de l'espace e. d. r.

La pondération de la matrice de covariance calculée lors de l'étape 4 est surtout nécessaire si l'on utilise des tranches de longueurs identiques plutôt que le critère décrit ci-dessus qui homogénéise la population des tranches en faisant varier leurs tailles.

La dimension de l'espace e. d. r. intervient dans la dernière étape de l'algorithme. Les étapes 4 et 5 ne sont rien d'autre qu'une analyse en composantes principales du nuage de points de \mathbb{R}^p approchant la courbe de régression inverse. Aussi, le résultat sera toujours une famille de p vecteurs propres. Néanmoins, seuls les K premiers sont dans le sous-espace e. d. r. Il est donc nécessaire de déterminer cette dimension.

9.2.1. Détermination de la dimension du sous-espace de projection

Les valeurs propres calculées lors de l'étape 5 de l'algorithme de la SIR permettent de se faire une idée de la dimension du sous-espace e. d. r. En effet, elles représentent la variance du nuage de points approchant la courbe de régression inverse selon la direction de chacun des vecteurs propres. Une valeur propre faible signifie que les points du nuage sont presque contenus dans l'hyperplan orthogonal au vecteur propre correspondant. Il y a donc très peu de chance pour que celui-ci soit véritablement une direction e. d. r. Par exemple, lorsque la dimension e. d. r. est égale à 1, la courbe de régression est une droite de \mathbb{R}^p . Le vecteur propre dont la valeur propre est la plus forte est à peu près

colinéaire à cette droite le long de laquelle sont répartis les points du nuage. La valeur propre correspondante sera donc importante.

Partant de ce constat, on peut, moyennant des hypothèses sur la distribution de l'entrée, comparer les valeurs propres à l'aide de tests statistiques. Li (1991) propose d'utiliser un test du χ^2 pour discerner si une valeur propre est significativement différente de 0. La distribution asymptotique de $\bar{\lambda}_{p-K}$, la moyenne des $p - K$ valeurs propres les plus faibles, est déterminée lorsque \mathbf{X} est gaussien par le théorème suivant :

Théorème 2. *Si \mathbf{X} est distribué selon une loi gaussienne, alors $N(p - K)\bar{\lambda}_{p-K}$ suit asymptotiquement une loi du χ^2 à $(p - K)(H - K - 1)$ degrés de liberté.*

Le théorème 2 permet de tester successivement les hypothèses $K = 0$, $K = 1$, $K = 2$, etc. Chaque fois que la probabilité

$$P_j = \mathbb{P} \left\{ \chi_{(p-j)(H-j-1)}^2 > N(p-j)\bar{\lambda}_{p-j} \right\} \quad (9.6)$$

dépasse un seuil bas, par exemple 0,05, on pourra raisonnablement supposer que le sous-espace e. d. r. compte au moins $j + 1$ dimensions. Une chute brutale de la valeur de P_j entre j et $j + 1$ constitue ainsi une indication que les $j + 1$ premières directions trouvées sont significatives.

Dans sa réponse aux commentaires mettant en doute l'utilité de ce test lorsque la distribution de l'entrée n'est pas gaussienne, Li (1991) évoque l'utilisation du *bootstrap* pour tester les hypothèses $K = j$ contre $K \geq j + 1$. Plus récemment, un algorithme de *bootstrap* plus commode a été proposé par Liquet et Saracco (2008, 2012). Il consiste à calculer pour les différentes valeurs supposées de la dimension e. d. r. un critère de la proximité du sous-espace estimé au sous-espace réel.

Soit \mathbf{M}_j la matrice de taille $p \times j$ dont les colonnes sont les j premiers β_k , c'est-à-dire les j premiers vecteurs propres de la matrice de covariance $\Sigma_{E(Z|Y)}$. On notera \mathcal{S}_j le sous-espace engendré par cette famille de vecteurs et $\mathbf{P}_{\mathcal{S}_j}$ le projecteur $\Sigma_{\mathbf{X}}$ -orthogonal sur ce sous-espace :

$$\mathbf{P}_{\mathcal{S}_j} = \mathbf{M}_j (\mathbf{M}_j' \Sigma_{\mathbf{X}} \mathbf{M}_j)^{-1} \mathbf{M}_j' \Sigma_{\mathbf{X}}. \quad (9.7)$$

Similairement, on définit la matrice $\widehat{\mathbf{M}}_j$ constituée des j premiers $\widehat{\beta}_k$, les vecteurs propres de la matrice $\widehat{\Sigma}_{E(Z|Y)}$ estimée lors de l'étape 4 de l'algorithme de la SIR, $\widehat{\mathcal{S}}_j$ le sous-espace qu'ils engendrent et $\mathbf{P}_{\widehat{\mathcal{S}}_j}$ le projecteur $\widehat{\Sigma}_{\mathbf{X}}$ -orthogonal sur ce sous-espace :

$$\mathbf{P}_{\widehat{\mathcal{S}}_j} = \widehat{\mathbf{M}}_j (\widehat{\mathbf{M}}_j' \widehat{\Sigma}_{\mathbf{X}} \widehat{\mathbf{M}}_j)^{-1} \widehat{\mathbf{M}}_j' \widehat{\Sigma}_{\mathbf{X}}. \quad (9.8)$$

La fonction de risque suivante :

$$R_j = \frac{1}{j} \mathbb{E}[\text{Trace}(\mathbf{P}_{\mathcal{S}_j} \mathbf{P}_{\widehat{\mathcal{S}}_j})], \quad (9.9)$$

exprime la proximité entre les deux sous-espaces \mathcal{S}_j et $\widehat{\mathcal{S}}_j$. Une valeur proche de 1 traduit une similarité importante tandis que la valeur 0 correspond au cas de deux sous-espaces

orthogonaux. Si $j = K$ alors R_j tend vers 1 pour N tendant vers l'infini. Aussi, pour une taille d'échantillon N donnée, une valeur de R_j proche de 1 suggère que la dimension du sous-espace e. d. r. est supérieure ou égale à j . En revanche, si $j > K$, les vecteurs directeurs additionnels ne sont pas des directions e. d. r. Leurs orientations n'ont donc aucune raison de se correspondre et la valeur de R_j restera inférieure à 1 quelque soit N . Remarquons cependant que l'augmentation de j s'accompagne d'un effet antagoniste : à mesure que l'on ajoute des vecteurs directeurs quelconques pour \mathcal{S}_j , l'espace angulaire dans lequel évoluent les vecteurs directeurs de $\widehat{\mathcal{S}}_j$ se ferme. En particulier, pour $j = p$, \mathcal{S}_j et $\widehat{\mathcal{S}}_j$ sont tous deux égaux à \mathbb{R}^p et R_p vaut 1 quelque soit N . Ainsi, il n'y a pas de valeur limite objective de R_j en dessous de laquelle on peut considérer que le $j^{\text{ème}}$ vecteur n'est pas une direction e. d. r.

En pratique, les β_k sont inconnus et l'on ne peut pas calculer les R_j . L'estimateur par *bootstrap* proposé par Liquet et Saracco (2008, 2012) a été utilisé ici. Étant donné un échantillon de taille N utilisé pour chercher les directions e. d. r., on calcule, selon la procédure décrite dans la section 7.4.1 du chapitre 7, N_b répliques *bootstrap* de la famille de vecteurs propres. Pour chaque réplique $\mathbf{P}_{\widehat{\mathcal{S}}_j}^{(b)}$ du projecteur $\mathbf{P}_{\widehat{\mathcal{S}}_j}$ on calcule ensuite la quantité

$$R_j^{(b)} = \frac{1}{j} \text{E}[\text{Trace}(\mathbf{P}_{\widehat{\mathcal{S}}_j} \mathbf{P}_{\widehat{\mathcal{S}}_j}^{(b)})]. \quad (9.10)$$

L'estimateur *bootstrap* de la fonction de risque (9.9) est alors :

$$\widehat{R}_j = \frac{1}{N_b} \sum_{b=1}^{N_b} R_j^{(b)}. \quad (9.11)$$

L'inspection visuelle du graphique des \widehat{R}_j en fonction de j , ou des diagrammes de Tukey des $R_j^{(b)}$, permet d'identifier la dimension de l'espace e. d. r. La valeur du critère stagne ou augmente jusqu'à $j = K$, puis elle chute brusquement avant d'augmenter à nouveau progressivement sous l'effet de la fermeture de l'espace angulaire.

Notons que cette méthode permet, de façon générale, de tester la crédibilité d'un sous-espace e. d. r. estimé et son champ d'application n'est pas restreint à la détermination de sa dimension. Elle a été appliquée par Liquet et Saracco (2008, 2012) pour la détermination d'un nombre optimal de tranches ou pour la détermination de la pondération entre la SIR classique et une de ses variantes fondée sur l'étude de la variance conditionnelle.

9.2.2. Remarque sur l'échantillonnage

Le théorème 1 sur lequel repose la théorie de la SIR nécessite que la distribution de l'entrée vérifie la condition 1 appelée condition de linéarité. On a vu dans la section 9.1.2 que la symétrie elliptique de la distribution d'entrée permet de remplir cette condition. Lorsque l'on dispose d'un modèle numérique, la distribution de l'entrée peut être fixée arbitrairement pourvu que l'on sache la simuler. On peut par exemple choisir une distribution multi-gaussienne : étant à symétrie elliptique, celle-ci vérifie automatiquement

la condition de linéarité. Celle-ci convient bien pour une première analyse du modèle. Cependant, la condition de linéarité devient une restriction gênante lors de l'application de la SIR à la problématique du diagnostic. En effet, le résultat du diagnostic sera conditionné par la distribution des profils de colmatage. Or, on voudrait pouvoir tirer parti de ce conditionnement pour prendre en compte dans le diagnostic les connaissances dont on dispose sur la répartition du colmatage.

Une première approche de ce problème est de forcer la symétrie elliptique de la distribution d'entrée. La façon la plus simple de le faire est de tronquer la distribution initiale en excluant tout les points situés en dehors du plus grand ellipsoïde inscrit dans son support. Cette approche simpliste peut fonctionner lorsque la dimension d'entrée est faible mais sera rapidement mise en échec lorsque la dimension augmente. Prenons l'exemple d'une distribution initiale uniforme sur \mathbb{I}^p . Lorsque $p = 2$, la troncature consiste à éliminer les 4 coins extérieurs au cercle inscrit dans le carré de côté 1, soit déjà environ 21 % de la surface du support initial. L'hypercube de dimension 16 possède quant à lui $2^{16} = 65\,536$ coins dont chacun est d'un volume environ 4,25 fois supérieur au volume de l'hypersphère inscrite. La probabilité qu'au moins un individu d'un échantillon uniforme de taille 10 000 appartienne à l'hypersphère n'excède pas 0,04 !

D. R. BRILLINGER a proposé, dans son commentaire à l'article de Li (1991), une méthode de rééchantillonnage consistant à simuler un échantillon selon une distribution multi-gaussienne puis à sélectionner, pour chaque individu de cet échantillon, l'individu de l'échantillon initial le plus proche. Certains individus de l'échantillon initial seront éliminés par le forçage de symétrie, d'autres apparaîtront plusieurs fois. La moyenne et la variance de la distribution gaussienne peuvent par exemple être choisies égales à celles de l'échantillon initial. Cette idée a été formalisée et développée par Cook et Nachtsheim (1994) qui proposent une méthode pour construire une distribution cible dont le support est l'ellipsoïde de volume minimal contenant une proportion choisie des données initiales. Ces méthodes de rééchantillonnage permettent de traiter les cas de dimension intermédiaire d'une façon plus performante qu'une simple troncature. Elles sont néanmoins tout aussi inefficaces dès que le nombre de variables d'entrée approche la dizaine. On a vu précédemment que la forme du support d'une loi uniforme de grande dimension était trop éloignée d'un ellipsoïde pour permettre un tel rééchantillonnage.

La condition de linéarité n'est cependant pas équivalente à la symétrie elliptique de la distribution d'entrée. C'est une condition moins commode à manipuler mais en définitive bien moins contraignante. La démarche de réduction de dimension présentée jusqu'ici ne prétend pas à l'exactitude : le modèle de la SIR peut par exemple ne pas être parfaitement vérifié, les directions e. d. r. n'en seront pas moins utiles pour représenter les données de façon condensée sans perdre trop d'information. Si l'on s'autorise un biais raisonnable dû à une violation modérée de la condition de linéarité, le spectre des distributions d'entrée utilisables s'élargit significativement. En fait, dès que la dimension d'entrée devient importante, très peu de directions violent la condition de linéarité car la plupart des projections d'un jeu de données de grande dimension sur des sous-espaces de faible dimension sont presque linéaires (Diaconis et Freedman, 1984). Dans leur commentaire à l'article de Li (1991), R. D. COOK et S. WEISBERG exhibent des cas où la SIR fonctionne bien que la distribution d'entrée ne soit pas à symétrie elliptique. La réponse

aux commentaires présente un exemple d'estimation fructueuse des directions e. d. r. par la SIR à l'aide d'un échantillon uniforme en dimension 10. Un résultat asymptotique expliquant cet effet favorable a été développé par Hall et Li (1993) à partir d'un argument bayésien. Une borne supérieure du biais encouru est donnée par Duan et Li (1991) pour le cas $K = 1$. Notons enfin qu'il est possible, a posteriori, de tester si les directions identifiées respectent la condition de linéarité et d'écarter ainsi les cas où l'entorse paraît trop importante.

9.3. Régression inverse par tranches multidimensionnelle

La méthode de la SIR présentée dans la section précédente est applicable lorsque la sortie est scalaire. On considère dans cette section le cas d'une sortie vectorielle, \mathbf{Y} , de dimension q . Plusieurs approches ont été proposées pour adapter la méthode à ce contexte (Li et coll., 2003; Barreda et coll., 2007). Elles sont présentées succinctement dans cette section. L'adaptation la plus directe de la SIR s'appuie sur un *découpage complet* (*complete slicing* en anglais) : des tranches de l'espace de sortie de tailles équivalentes sont construites récursivement en appliquant un découpage selon chacune des q dimensions. Les tranches ne sont plus des segments comme pour la SIR univariée mais des hypervolumes de dimension $q - 1$. Le découpage complet n'a véritablement de sens qu'en dimension 2 ou 3 au-delà de quoi le nombre de tranches et leur volume augmentent drastiquement et il n'est plus possible de les peupler suffisamment pour permettre l'estimation.

En dimension 4 ou plus, le *découpage marginal* (*marginal slicing* en anglais) se pose comme alternative au découpage complet. Il s'agit d'appliquer, au préalable, une transformation à la sortie \mathbf{Y} permettant d'en réduire la dimension. On peut par exemple calculer la médiane des composantes de \mathbf{Y} ou bien effectuer une analyse en composantes principales et effectuer le découpage sur les scores selon les premières composantes. Notons que certaines directions e. d. r. peuvent être éludées lorsque l'on utilise une projection de la sortie sur un sous-espace de plus petite dimension.

Afin de remédier à cet inconvénient, une troisième approche a été proposée. Elle consiste à chercher les vecteurs propres de la matrice Σ_{moy} obtenue par moyenne pondérée des matrices de covariance des régressions inverses selon chaque composante de sortie :

$$\Sigma_{moy} = \sum_{i=1}^q w_i \Sigma_{E(\mathbf{X}|Y_i)}. \quad (9.12)$$

Les w_i sont des poids que l'on choisira tous égaux ou bien proportionnels aux plus grandes valeurs propres obtenues lors de SIR classique effectuées au préalable pour chaque composante de sortie. On parle alors de *découpage marginal groupé* (*pooled marginal slicing* en anglais).

La réduction simultanée de la dimension de l'entrée et de la sortie peut être abordée en exploitant la dualité entre la SIR et une extension non linéaire de l'*analyse canonique* développée par Harold HOTTELING dans les années 30 (Lebart et coll., 2006, chap. 2), (Saporta, 2006, chap. 8). L'objectif de l'analyse canonique est de déterminer les combinaisons linéaires des composantes de \mathbf{Y} les *plus prévisibles* par des combinaisons linéaires

des composantes de \mathbf{X} . Sous une hypothèse gaussienne, les meilleures prédictions au sens des moindres carrés se font à l'aide de régressions linéaires et le problème peut être formulé comme la recherche du maximum de corrélation. On trouve ainsi successivement q couples de combinaisons linéaires canoniques, en imposant la condition qu'elles soient non corrélées.

Cette approche peut être généralisée en recherchant des prédicteurs non linéaires. La recherche des combinaisons linéaires les plus prévisibles revient alors à minimiser le rapport suivant :

$$\frac{E(\text{var}(\boldsymbol{\theta}'\mathbf{Y}|\mathbf{X}))}{\text{var}(\boldsymbol{\theta}'\mathbf{Y})}. \quad (9.13)$$

La formule de la variance totale permet de rapporter ce problème à la recherche du maximum en $\boldsymbol{\theta}$ de

$$\frac{\text{var}(E(\boldsymbol{\theta}'\mathbf{Y}|\mathbf{X}))}{\text{var}(\boldsymbol{\theta}'\mathbf{Y})}. \quad (9.14)$$

Toujours sous la condition de non-corrélation des combinaisons linéaire, un développement analogue à celui mis en œuvre dans la section 8.1 sur du chapitre 8 permet de montrer que ce problème est équivalent à la recherche des valeurs propres de $\Sigma_{E(\mathbf{Y}|\mathbf{X})}$, la matrice de covariance de la régression directe de \mathbf{Y} par \mathbf{X} (Li et coll., 2003). Ainsi, l'algorithme de la SIR peut être mis à profit pour rechercher les combinaisons linéaires les plus prévisibles en échangeant simplement les rôles de l'entrée et de la sortie.

Le découpage en tranches n'est efficace que si la dimension de l'espace à réduire n'est pas trop élevée. Ainsi, la démarche proposée par Li et coll. (2003) est de procéder à une première étape de réduction de dimension du sous-espace dont la dimension est la plus grande, par exemple par analyse canonique (Lebart et coll., 2006, chap. 2) ou analyse en composantes principales. On cherche ensuite le sous-espace e. d. r. de l'espace dont la dimension n'a pas été réduite. On utilisera pour cela l'une des techniques de SIR multivariée mentionnée précédemment. Après avoir déterminé la dimension du sous-espace e. d. r., on échange à nouveau les rôles des deux espaces et l'on procède à une nouvelle réduction de dimension avec l'algorithme de la SIR. Cette étape d'échange puis réduction de dimension est ensuite répétée jusqu'à ce que les fluctuations des résultats deviennent négligeables. Cette procédure, appelée SIR *alternée*, ne peut être conduite entièrement automatiquement. La détermination des dimensions à chaque étape notamment, requiert une interprétation en partie subjective. La visualisation des projections des données selon les différentes directions trouvées joue un rôle important.

Une autre approche a été proposée récemment par Coudret et coll. (à paraître). La SIR monodimensionnelle est d'abord appliquée individuellement à chacune des variables de sortie. Un sous-espace e. d. r. commun est ensuite déterminé en optimisant un critère analogue à celui utilisé pour estimer la dimension e. d. r. Le problème d'optimisation peut être résolu analytiquement. L'article n'ayant été rendu disponible que le 6 juillet 2012, cette méthode n'a pu être mise en œuvre dans cette thèse.

Quatrième partie

Diagnostic du colmatage

10. Développement d'une nouvelle méthode de diagnostic du colmatage

Les parties précédentes ont montré que les réponses de n. g. l. enregistrées lors d'essais R.G.L. 4 sont sensibles à l'influence du colmatage et que le modèle numérique de g. v. permet de simuler ce phénomène de façon satisfaisante. Une première méthode de diagnostic par analyse de courbes de réponse du n. g. l. à l'aide du modèle a été mise au point. Cette méthode, appelée *méthode de diagnostic à profil choisi*, est décrite dans la section 3.4.2. Dans ce même chapitre, on a vu que les autres estimations de colmatage, par examen télévisuel ou par sonde à courants de Foucault, sont rares. De plus, les examens télévisuels ne donnent d'information qu'au sujet de la plaque supérieure et l'incertitude associée aux estimations par courant de Foucault est encore importante. Il est par conséquent délicat d'évaluer la justesse des diagnostics produits simplement en les comparant à des valeurs de référence. Il y aurait là un risque de ne produire qu'une tautologie : la complexité du modèle de g. v. et les grandes dimensions de ses espaces d'entrée et de sortie laissent une marge de manoeuvre importante pour adapter la méthode aux quelques données disponibles. Comment dans ce cas se prémunir d'un surajustement ? Il est certainement plus intéressant de renforcer indépendamment la méthode de diagnostic à profil choisi afin qu'elle apporte des informations complémentaires à celles déjà disponibles.

Chacune des hypothèses sous-jacentes à la méthode de diagnostic à profil choisi a été remise en question. Chaque fois que celles-ci ne pouvaient être étayées par de solides arguments, la méthode a été modifiée, aboutissant finalement à une nouvelle méthode de diagnostic du colmatage que l'on désignera par la suite par *méthode de diagnostic par réduction de dimension*. À l'image du calage du modèle qui était la première étape de cette démarche, il s'agit en quelque sorte d'identifier les degrés de liberté de la méthode et d'éliminer ceux ne pouvant être contraints rationnellement. La méthode de diagnostic à profil choisi comporte plusieurs éléments arbitraires : le choix d'une famille représentative de profils de colmatage, la méthode de prétraitement des courbes de réponse et le choix des variables décrivant le colmatage. De ce constat et de l'analyse des premiers résultats obtenus, notamment avec le filtre particulière, émergent deux questions qui ont orienté les améliorations de la méthode proposée dans cette thèse :

- « Qu'est-ce qui, dans la forme des courbes de réponse de n. g. l., est caractéristique de l'état de colmatage ? »
- « Quelle information sur l'état de colmatage peut-on raisonnablement espérer extraire des mesures ? »

L'ordre dans lequel sont présentées les étapes du travail réalisé pour répondre à ces questions a été choisi afin de limiter les renvois à des paragraphes ultérieurs et l'on

répond ainsi à ces deux questions ensemble, progressivement. La section 10.1 reprend les prétraitements utilisés dans la méthode de diagnostic à profil choisi et justifie le choix de ne retenir que la suppression de la moyenne temporelle. La section 10.2 présente une première analyse de sensibilité séquentielle, c'est-à-dire en considérant la valeur du n. g. l. à chaque pas de temps comme une sortie indépendante. L'analyse en composantes principales est utilisée dans la section 10.3 pour réduire la dimension de sortie du modèle. Une seconde analyse de sensibilité réalisée avec les nouvelles variables de sortie est décrite dans la section 10.4. Enfin, la réduction de la dimension d'entrée est abordée dans la section 10.5.

10.1. Prétraitement des données

10.1.1. Filtrage de l'effet statique

La mesure de n. g. l. en régime stationnaire à puissance nominale est affectée par le colmatage. Cet effet sera qualifié par la suite de *statique* par opposition aux modifications de la réponse *dynamique* que l'on observe lors de transitoires de puissance. La figure 10.1 représente cinq courbes de réponse de n. g. l. simulées pour diverses configurations de colmatage. Elles sont séparées par des écarts importants. Comparativement à ces écarts, leurs formes paraissent assez similaires : il est par exemple difficile de dire quelle est, de la courbe mauve (—+—) et de la courbe verte (—⊖—), celle qui est la plus semblable à la courbe vermillon (—⊞—). Autrement dit, l'amplitude de l'effet statique est grande par rapport à celle de l'effet dynamique. Il est donc nécessaire de filtrer la contribution de l'effet statique pour pouvoir analyser la réponse dynamique du n. g. l. Par ailleurs, la mesure de n. g. l. peut être affectée de biais ou de dérives causés par des phénomènes dont les échelles de temps sont bien supérieures à la durée d'un transitoire. On peut espérer, en filtrant l'effet statique, se débarrasser de ces biais. La méthode de diagnostic à profil choisi actuelle utilise la soustraction de la moyenne pour filtrer cet effet. Cette section propose une validation théorique de ce choix fait arbitrairement dans les premiers temps du développement de la méthode (Chip et coll., 2008).

Les courbes de réponses peuvent être regroupées dans un domaine de variation plus restreint permettant de faire ressortir leurs différences de forme en retranchant à chacune une constante judicieusement choisie. Puisque les transitoires utilisés débutent de points de fonctionnement proches, à savoir un régime stationnaire à puissance nominale, on pourrait être tenté de soustraire à chaque courbe sa valeur initiale. La figure 10.2 montre l'effet de ce filtrage. La variabilité des réponses est nulle à l'instant initial et augmente au cours du transitoire. Il est dorénavant plus facile de comparer la forme des courbes. Par exemple, on remarque que la forme de la courbe vermillon (—⊞—) est plus proche de la courbe verte (—⊖—) que de celle de la courbe mauve (—+—). Cependant, ce choix n'est pas optimal car il annule l'effet du colmatage au début du transitoire, or rien ne permet de savoir a priori si l'information importante pour le diagnostic est bien à la fin du transitoire.

La notion de développement limité permet de formaliser cette remarque. À colmatage donné, la réponse du modèle peut être représentée par une fonction y du temps t dépendant

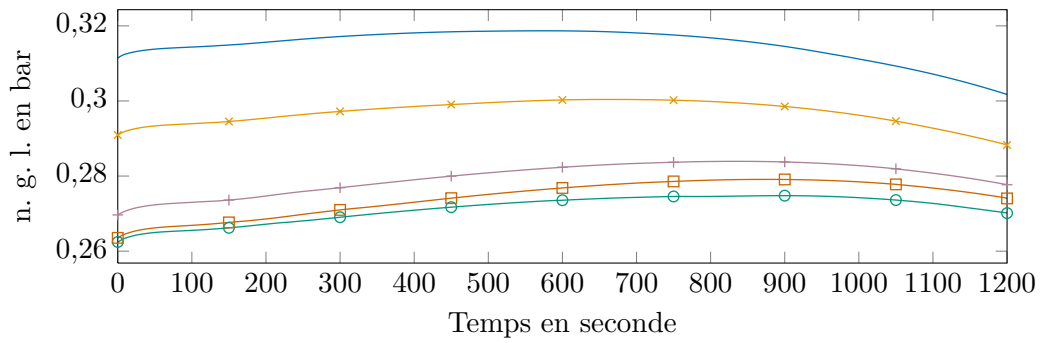


Figure 10.1. : Cinq courbes de réponse simulées, sans prétraitement. Les écarts entre courbes sont importants par rapport aux différences de forme.

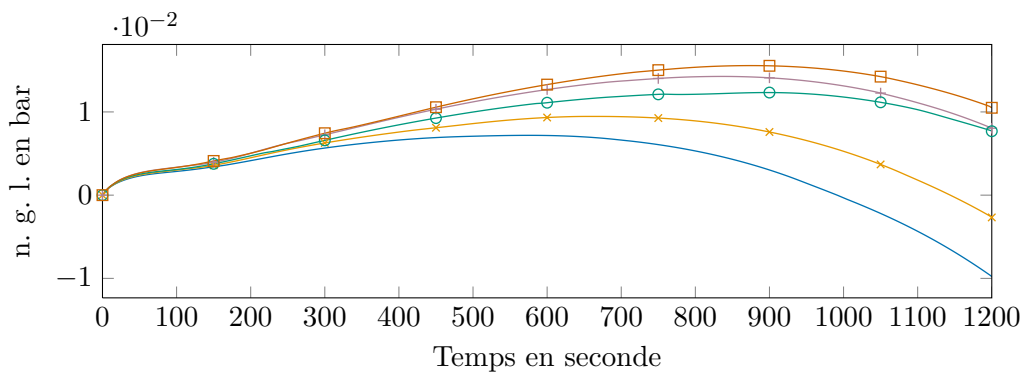


Figure 10.2. : Les cinq courbes de réponse de la figure 10.1 après soustraction de leur valeur initiale. La variance est nulle au début du transitoire et maximale à la fin. On distingue mieux la forme des courbes.

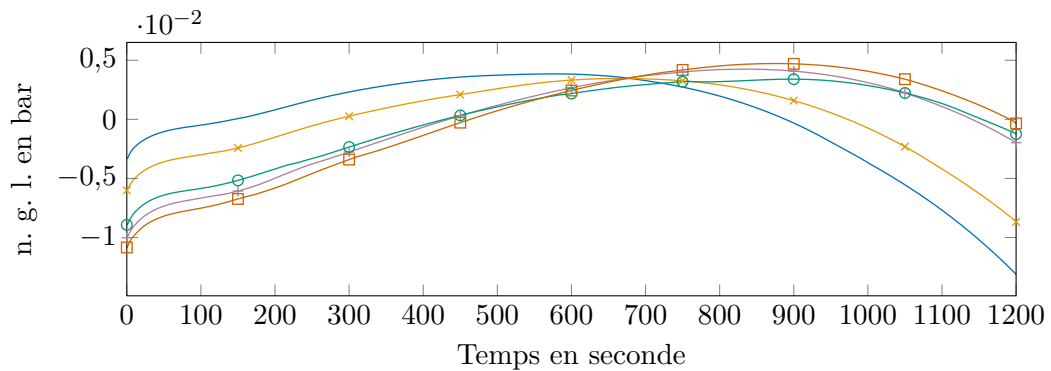


Figure 10.3. : Les cinq courbes de réponse de la figure 10.1 après soustraction de leur moyenne temporelle. La variance est presque nulle vers 650 s ; les différences de forme correspondent à un « pivotement » autour de ce point fixe.

implicitement du transitoire imposé :

$$\begin{aligned} y: \mathbb{R} &\rightarrow \mathbb{R} \\ t &\mapsto y(t). \end{aligned} \tag{10.1}$$

En supposant y plusieurs fois dérivable, le théorème de Taylor permet d'écrire, au voisinage d'une date t_0 , un développement du type :

$$y(t) = \sum_{k=0}^p \frac{y^{(k)}(t_0)}{k!} (t - t_0)^k + r_p(t), \tag{10.2}$$

où $r_p(t)$ désigne un reste négligeable devant $(t - t_0)^p$. Le premier terme de la somme est une constante, $y(t_0)$, et ne dépend donc pas du transitoire. Les suivants, $y'(t_0)(t - t_0)$, $y''(t_0)\frac{(t-t_0)^2}{2}$, $y'''(t_0)\frac{(t-t_0)^3}{6}$, etc. dépendent du temps et font apparaître les dérivées de y . Localement, le premier terme est une bonne approximation de l'effet statique tandis que les termes suivants représentent la partie dynamique du signal que l'on cherche à exploiter. Ainsi, le choix de t_0 détermine la zone du transitoire pour laquelle l'estimation de la contribution statique est la plus juste et retrancher à chaque courbe sa valeur initiale correspond au choix $t_0 = 0$ s. Faute de connaître a priori l'intervalle de temps le plus intéressant, on cherche à minimiser globalement l'erreur commise en évaluant le terme constant. Cela peut être accompli en calculant la moyenne des développements donnés par l'équation (10.2) pour t_0 variant sur tout l'intervalle de temps. Ainsi, l'erreur est répartie de façon homogène sur l'intervalle. Le terme correctif obtenu n'est autre que la moyenne temporelle de chacune des courbes. Le résultat de ce filtrage est illustré par la figure 10.3. Le colmatage a pour effet de faire « pivoter les courbes » autour d'un point fixe à un temps d'environ 650 s. Ce resserrement est un élément important pour la suite car il correspond à un minimum de variance des réponses. Une justification supplémentaire de la validité de cette méthode de filtrage sera apportée dans la section 10.3.3 grâce à l'analyse en composantes principales des courbes de réponse.

10.1.2. Amplitude de variation de la mesure de niveau

Dans le modèle de g. v., le n. g. l. est assimilé à la différence de pression entre l'amont des dispositifs de séparation, au dessus du cintre du faisceau de tubes, et le bas du g. v., au niveau de la jonction entre le *riser* et le *downcomer*. Dans la réalité, cette différence de pression est comparée à celle mesurée sur une colonne d'eau de référence puis convertie par une relation linéaire en une mesure de niveau en mètre. L'amplitude des courbes de réponse simulées et mesurées est donc différente et il est nécessaire de rapporter les deux grandeurs à un même étalon. Il avait été décidé pour la méthode de diagnostic à profil choisi d'adimensionner les deux grandeurs en divisant chaque courbe de réponse par son amplitude. Outre sa simplicité, cette façon de procéder présente l'avantage de compenser d'éventuelles variations de densité de la colonne d'eau de référence. Elle présente néanmoins deux inconvénients.

L'un des intérêts d'étudier un régime dynamique est que la durée du transitoire permet de filtrer partiellement le bruit important de la mesure de n. g. l. Cependant,

l'amplitude est elle-même assez sensible à ce bruit et cette façon de normaliser les mesures peut dégrader la qualité de l'indicateur. L'autre inconvénient plus gênant de la normalisation par l'amplitude est qu'il est délicat de prouver que cette transformation préserve l'information sur le colmatage, comme cela a pu être fait pour le filtrage de l'effet statique par soustraction de la moyenne temporelle. Par ailleurs, l'amplitude des courbes de réponse est affectée par plusieurs coefficients de calage, notamment la répartition des pertes de charge entre le *riser* et le *downcomer* ainsi que les paramètres du modèle de glissement. La procédure de calage décrite dans le chapitre 6 peut aboutir à plusieurs combinaisons de ces paramètres qui induisent une dilatation ou une contraction des faisceaux de courbes de réponse. Ainsi, il a été choisi d'utiliser un unique coefficient de proportionnalité pour rapporter les courbes mesurées et simulées à la même échelle. Cette constante de normalisation peut être déterminée en comparant un faisceau de courbes simulées à des mesures issues d'une tranche dont on sait par ailleurs que l'état de colmatage a fortement varié. Fixer la valeur de la constante multiplicative revient en quelque sorte à choisir les bornes des variations de l'indicateur développé et s'inscrit dans le prolongement de la méthodologie de calage exposée dans le chapitre 6. Cette notion sera reprise plus en détail dans le chapitre 13 lorsque le fonctionnement de la nouvelle méthode de diagnostic aura été exposé.

10.2. Analyse de sensibilité séquentielle

La sortie du modèle de g. v. est la réponse du n. g. l. Elle est de nature fonctionnelle mais on l'échantillonne régulièrement à p pas de temps, ce qui revient à définir p variables de sortie, liées par la dynamique du modèle. La méthode d'analyse de sensibilité décrite dans le chapitre 7 concerne les modèles ayant une variable de sortie unique. Une façon simple de l'utiliser dans ce contexte multidimensionnel est d'ignorer les liens entre les variables de sortie et de les traiter indépendamment. On obtient ainsi pour chaque pas de temps une famille d'indices de Sobol' qui seront dits *séquentiels*. Les indices séquentiels décrivent des courbes qui représentent l'évolution au cours du temps de l'importance relative de chaque demi-plaque. L'importance globale d'une demi-plaque peut être estimée en considérant la durée de l'intervalle de temps pendant lequel ses indices séquentiels sont importants. Les indices séquentiels permettent en outre de savoir si une demi-plaque a de l'effet plutôt sur le début, le milieu ou la fin du transitoire. Ainsi, les formes des trajectoires des indices séquentiels permettent d'identifier différents types d'effets suivant la localisation du colmatage.

Un échantillon de taille 1000 a été généré pour calculer les indices de Sobol' séquentiels associés aux 16 demi-plaques. Les figures 10.4 et 10.5 représentent les indices séquentiels d'ordre 1 respectivement de la jambe chaude et de la jambe froide. Ces indices participent de la même analyse de sensibilité et ne sont présentés dans deux figures que pour en améliorer la lisibilité. Les teintes des courbes indiquent la position dans le g. v. : les plus claires correspondent aux plaques inférieures, les plus sombres aux plaques supérieures. Les numéros des plaques sont par ailleurs indiqués sur les courbes. Les barres d'erreur indiquent les intervalles de confiance *bootstrap* pour un risque de 0,1.

On observe que les indices sont, en général, d'autant plus importants que la plaque est haute dans le g. v. La demi-plaque n° 8 côté froid se détache par son importance. Pour les autres plaques, l'impact de la jambe chaude est plutôt plus important que celui de la jambe froide, quoique cela soit difficile à déterminer pour les plaques inférieures qui affichent une décroissance au début du transitoire. L'essentiel des variations des courbes d'indices séquentiels se concentre dans l'intervalle de 600 s à 900 s. Cet intervalle correspond au point de pivotement des courbes évoqué dans la section 10.1.1. En fonction de la cote de la plaque correspondante, les courbes présentent un pic ou un creux dans l'intervalle central. Cette observation traduit l'existence de deux types d'effets principaux du colmatage sur le n. g. l. : l'un affecte spécialement le milieu du transitoire, l'autre le début et la fin. En partant du bas du g. v., l'amplitude du pic augmente à mesure que l'on s'élève, atteint un maximum pour la plaque n° 2, puis décroît jusqu'à devenir un creux de plus en plus profond. Le pic est beaucoup plus marqué pour la jambe chaude que la jambe froide.

L'absence de différences majeures entre les trajectoires d'indices séquentiels des deux jambes est imputable aux débits transverses qui homogénéisent l'état du fluide. En effet, l'analyse de sensibilité de la version 2 du modèle faisait au contraire ressortir deux comportements très différents. Les résultats de cette analyse sont exposés dans les deux articles regroupés dans l'annexe G page 252. Il est par ailleurs probable que le changement de sens des débits transverses au passage de la plaque n° 2 soit responsable de l'importance du pic en jambe chaude pour cette plaque. Ce phénomène a été décrit dans la section 6.4.

Les indices séquentiels totaux sont représentés dans les figures 10.6 et 10.7. Les indices totaux sont légèrement supérieurs aux indices d'ordre 1 correspondants ce qui traduit l'existence d'interactions, en particulier pour les plaques basses dans l'intervalle central. On remarque par ailleurs que les formes des trajectoires sont très similaires. Ainsi, les interactions ne font qu'amplifier les effets d'ordre 1 mais n'introduisent pas d'effets nouveaux .

10.3. Réduction de la dimension de sortie

Les indices de Sobol' séquentiels ont permis d'étudier de manière essentiellement qualitative l'effet du colmatage sur la réponse dynamique du n. g. l. Cependant, le grand nombre d'indices calculés rend malcommode la quantification précise de l'importance de chaque demi-plaque. Une base de projection des courbes de réponses a été construite par analyse en composantes principales. En plus de permettre de calculer des *indices de sensibilité compacts*, définis dans la section 8.3.2, cette base permet d'analyser l'effet du colmatage sur la forme des courbes de réponse. La méthode de diagnostic à profil choisi compare chaque point de donnée des réponses mesurées et simulées en leur attribuant des poids équivalents. Il est possible d'ajuster ces poids, par exemple en modifiant le filtrage de l'effet statique comme cela est décrit dans la section 10.1.1, mais il faut pour cela connaître la partie du transitoire la plus informative. La base de projection constituée des premières composantes principales est adaptée aux effets du colmatage. Elle permet

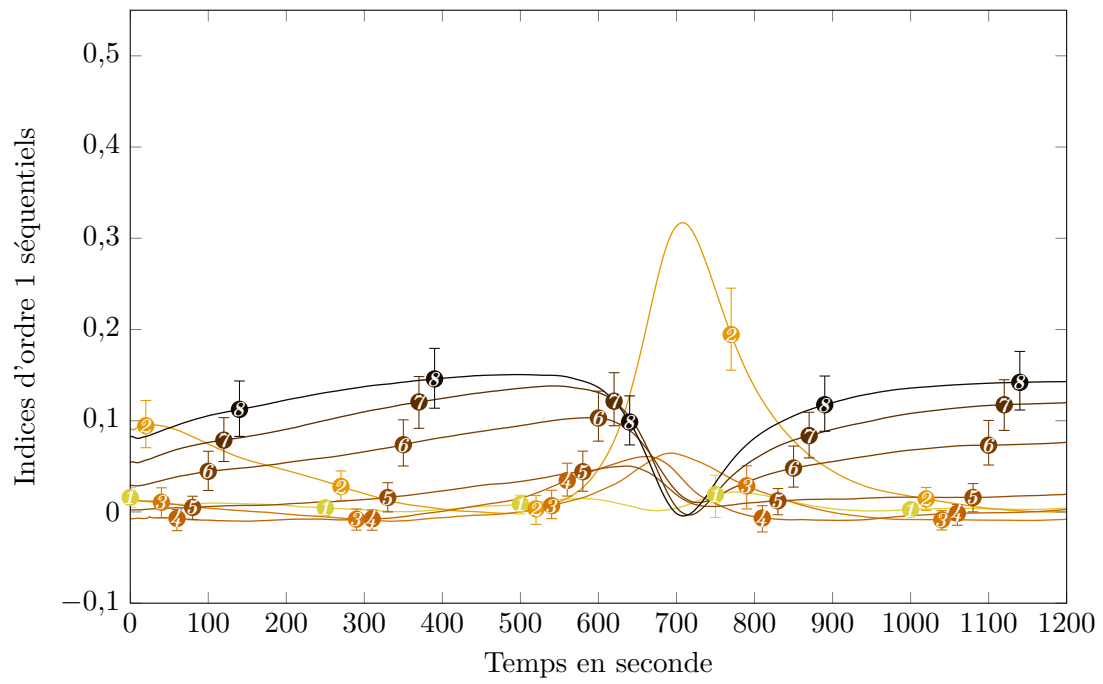


Figure 10.4. : Indices d'ordre 1 séquentiels en jambe chaude.

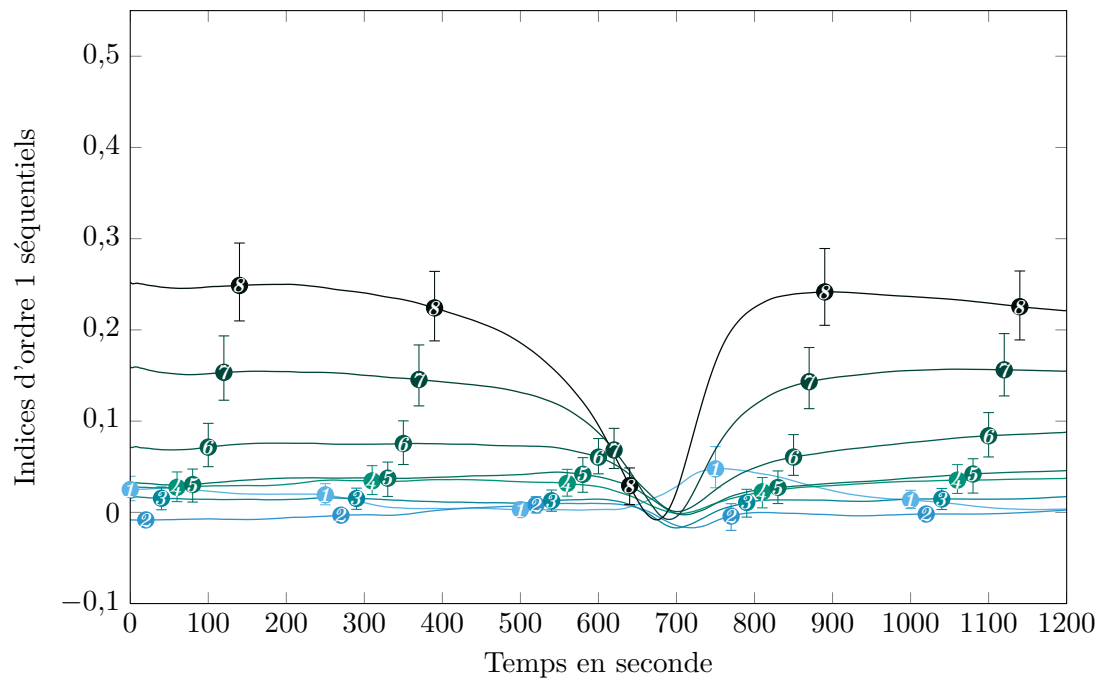


Figure 10.5. : Indices d'ordre 1 séquentiels en jambe froide.

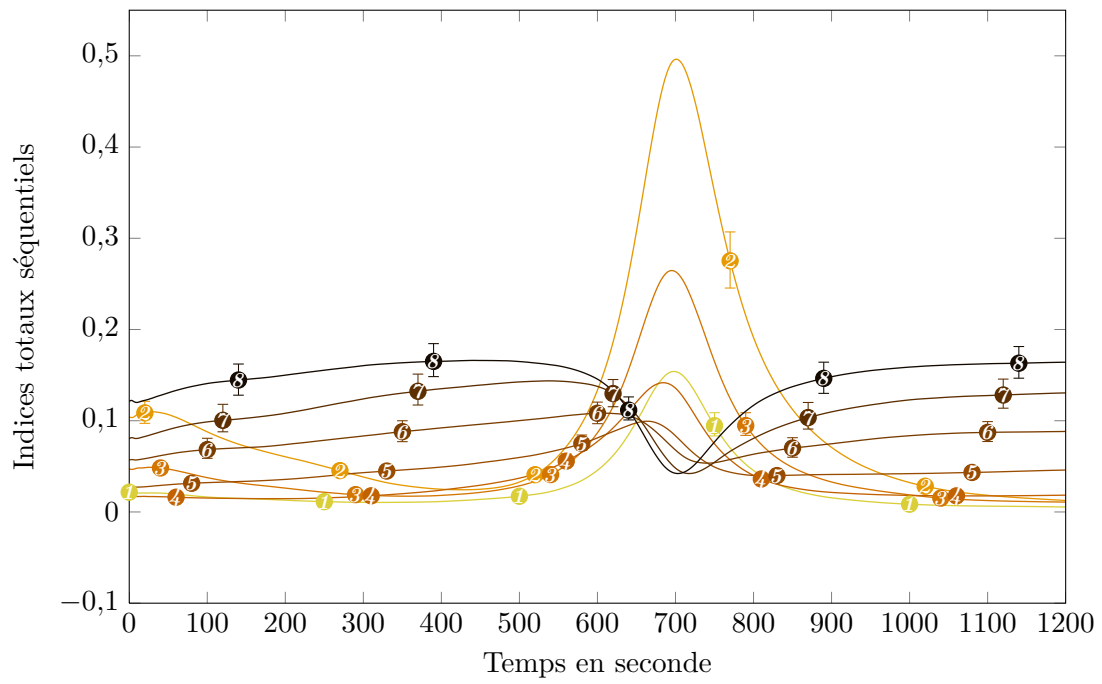


Figure 10.6. : Indices séquentiels totaux en jambe chaude.

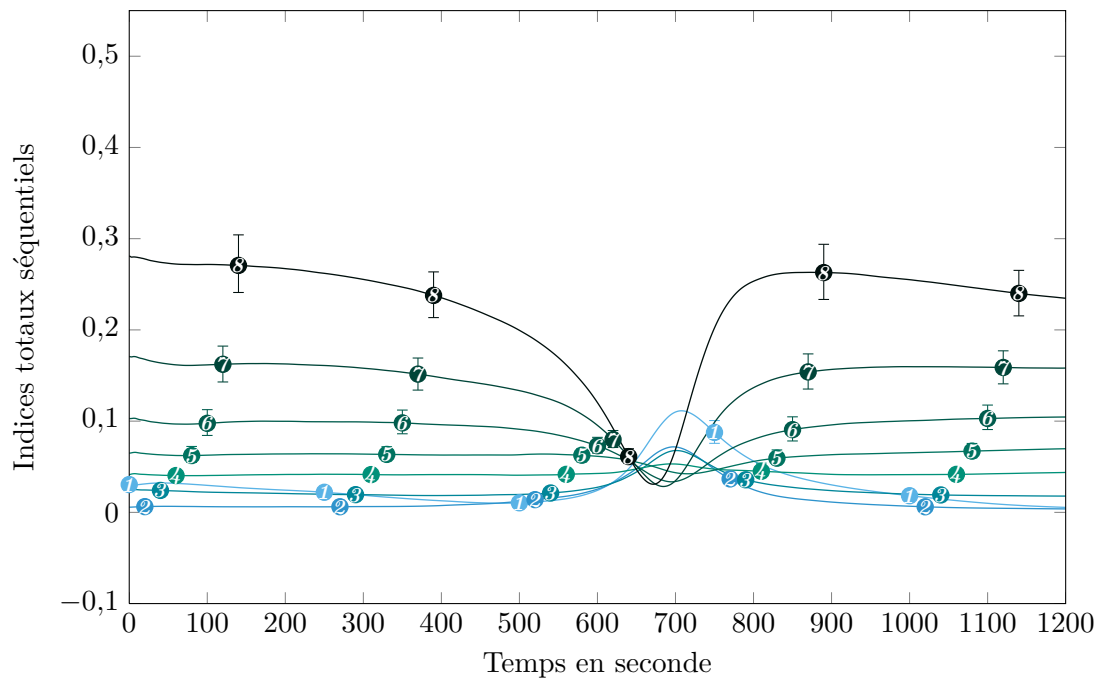


Figure 10.7. : Indices séquentiels totaux en jambe froide.

ainsi de définir une mesure plus performante qu'une simple comparaison point par point.

10.3.1. Composantes principales d'un échantillon de simulations

Un échantillon de 9369 configurations de colmatage avec des taux par demi-plaque distribués uniformément sur $[0 ; 0,65]$ a été généré. La figure 10.8 représente les quatre premières composantes principales calculées avec la matrice de covariance des variables séquentielles. Les valeurs propres des composantes principales décroissent très rapidement : la première rend compte de 96,76 % de la variance et la seconde de 2,91 %. La première composante principale, représentée en bleu (\rightarrow), est presque rectiligne. C'est elle qui est responsable de l'effet de pivotement autour d'un point fixe au milieu du transitoire mentionné dans la section 10.1.1. Ce résultat contredit l'hypothèse qui avait été formulée dans les premiers temps du développement de la méthode selon laquelle ce sont les dérivées du signal d'ordre supérieur ou égal à 2 qui portent l'essentiel de l'information sur l'état de colmatage (Chip et coll., 2008). Le colmatage agit principalement sur la pente globale de la courbe de réponse et donc plutôt sur la dérivée première du signal. La seconde composante principale, représentée en orange (\rightarrow) est d'allure parabolique. Elle affecte principalement la courbure des courbes de réponse. Cet effet, moins marqué que le pivotement autour du point fixe, peut néanmoins être observé sur la figure 10.3. La figure 10.9 représente les perturbations subies par la moyenne lorsqu'on lui ajoute ou qu'on lui soustrait une petite contribution de chacune des deux premières directions principales. Le graphique de gauche représente l'effet de la première direction principale : elle fait « pivoter » la courbe moyenne autour du point fixe d'abscisse environ égale à 650 s. Le graphique de droite représente l'effet de la deuxième direction principale : elle aplatit, ou au contraire dilate la courbe moyenne, les points fixes étant cette fois situés à environ 250 s et 950 s.

La figure 10.10 représente les quatre premières directions principales calculées à partir des variables centrées réduites normées selon leur valeur propre associée. Cette représentation permet de bien discerner les domaines d'influence respectifs des deux premières composantes principales. La première agit essentiellement sur le début et la fin du transitoire. La seconde n'agit que sur un intervalle restreint, au milieu du transitoire, de façon symétrique. On retrouve là les deux familles d'indices séquentiels identifiées dans la section 10.2. La décroissance des valeurs propres est moins rapide que pour les composantes principales « brutes ». De fait, les valeurs prises par le n. g. l. au début et à la fin du transitoire sont environ deux fois plus importantes en valeur absolue qu'au milieu du transitoire, comme on peut le voir sur la figure 10.3. Il y a donc un effet d'échelle entre les variables des extrémités et celles du centre que la normalisation permet de corriger. Ainsi, on privilégiera dans la suite l'analyse en composantes principales fondée sur les corrélations et c'est de celle-ci qu'il s'agira en l'absence de précision.

10.3.2. Composantes principales d'un échantillon de réponses mesurées

L'analyse en composantes principales permet de décomposer l'effet du colmatage sur la forme des courbes de réponse. Il est intéressant d'utiliser cet outil pour apprécier plus

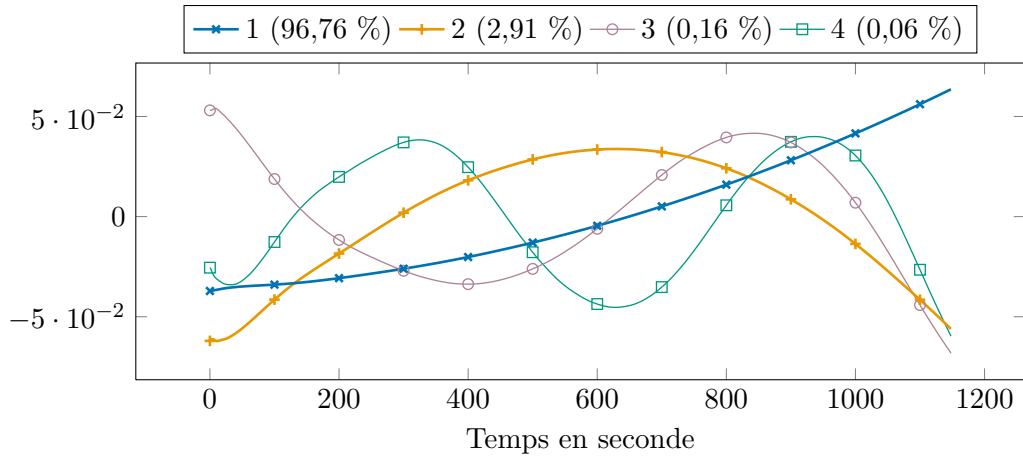


Figure 10.8. : Quatre premières directions principales calculées à partir de la matrice de covariance d'un échantillon uniforme de courbes de réponse simulées. Les deux premières composantes principales représentent presque l'intégralité de la variance de l'échantillon.

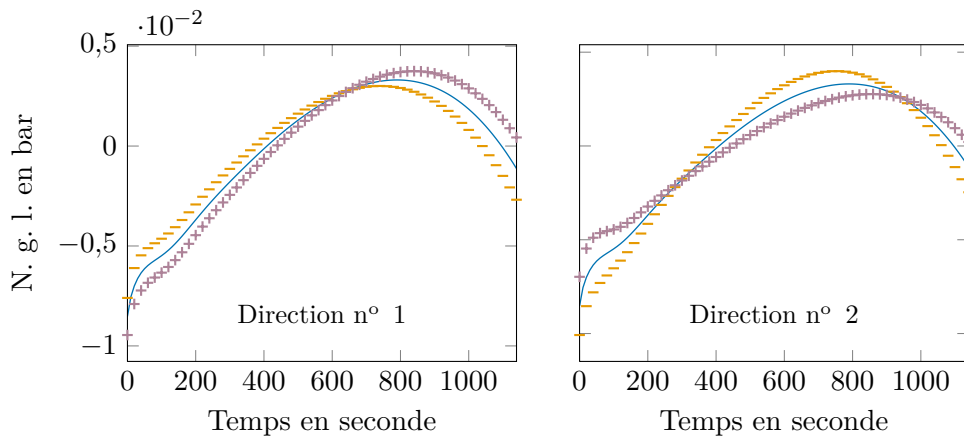


Figure 10.9. : Perturbations négative (-) et positive (+) de la moyenne (—) par la première direction principale, à gauche, et la deuxième, à droite.

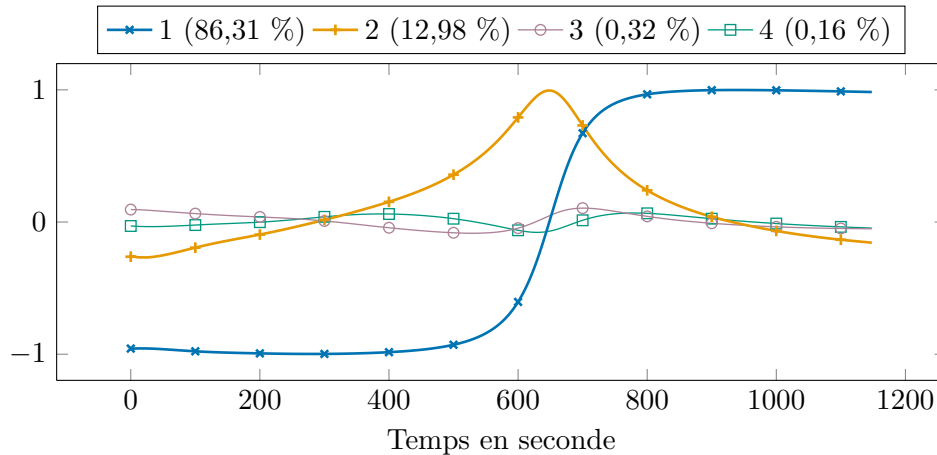


Figure 10.10. : Quatre premières directions principales calculées à partir de la matrice de corrélation d'un échantillon uniforme de courbes de réponse simulées. Chaque direction a une norme égale à la racine carré de la valeur propre associée.

finement la capacité du modèle à reproduire la réponse du n. g. l. La figure 10.11 représente les quatre directions principales obtenues avec un échantillon de $3 \cdot 149 = 447$ courbes de réponse mesurées lors de transitoires réalisés dans les centrales Charlie (tranches n° 2, 3 et 4) et Delta (tranches n° 2 et 3). Ces tranches ont subi des nettoyages chimiques sur la période considérée. L'échantillon contient donc des configurations de colmatage fort et faible. La matrice de corrélation a été utilisée et les directions ont une norme proportionnelle à la racine carré de leur valeur propre. Les réponses ont été tronquées au-delà de la durée du transitoire le plus bref, ce qui explique que l'échelle de temps n'atteint que 800s. Cela a pour effet d'atténuer la forme des directions principales : la branche oblique de la première direction est par exemple moins raide. Elles demeurent néanmoins très semblables à celles obtenues par simulation. Autrement dit, les effets principaux du colmatage sur l'allure des courbes de réponse sont correctement reproduits par le modèle.

Les valeurs propres décroissent moins rapidement que dans le cas simulé : les deux premières composantes principales n'expliquent que 52,94 % de la variance totale. Cela est dû en partie à la brièveté de l'intervalle de temps considéré mais surtout au bruit affectant la mesure de n. g. l. Un lissage par un modèle de régression locale polynomial d'ordre 2 à l'aide de l'algorithme « loess » (MathWorks, 2012) avec une fenêtre de 100 s a été appliqué aux courbes de réponses et une nouvelle analyse en composantes principales a été réalisée avec ces nouvelles données. Les deux premières directions calculées après lissage expliquent 87,31 % de la variance de l'échantillon. Enfin, la figure 10.11 permet d'observer que le rapport signal sur bruit devient assez défavorable au delà de la deuxième composante principale, ce qui incite à ne conserver comme nouvelles variables de sortie que les scores des deux premières.

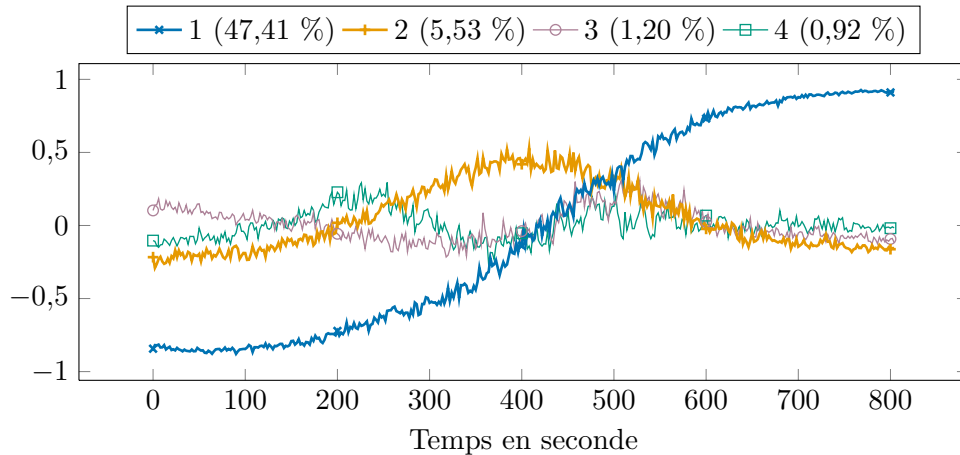


Figure 10.11. : Quatre premières directions principales calculées à partir de la matrice de corrélation d'un échantillon de 447 courbes de réponse mesurées sur les tranches n° 2, 3 et 4 de la centrale Charlie, et n° 2 et 3 de la centrale Delta. Chaque direction a une norme égale à la racine carré de sa valeur propre associée.

10.3.3. Retour sur le filtrage de l'effet statique

L'analyse en composantes principales apporte une justification supplémentaire de la validité de la méthode de filtrage présentée dans la section 10.1.1. La figure 10.12 représente, en traits pleins, les quatre premières directions principales obtenues avec l'échantillon de réponses mesurées utilisé pour la figure 10.11, *sans* soustraction de leur moyenne temporelle. Les courbes de réponses ont été lissées avec l'algorithme *loess* mentionné précédemment. Ce prétraitement n'a ici vocation qu'à améliorer la lisibilité de la figure et ne change pas les conclusions que l'on peut en tirer.

La première direction principale est approximativement constante. Elle correspond à l'effet statique tandis que les composantes principales suivantes représentent l'effet du colmatage sur la réponse dynamique du n. g. l. Les scores de la composante « statique » sont approximativement proportionnelles à la moyenne temporelle des courbes. Or, les composantes principales sont orthogonales ; on peut donc retrancher à chaque courbe sa moyenne temporelle sans affecter les composantes « dynamiques ». Les trois premières directions principales calculées après retranchement des moyennes temporelles sont représentées en traits pointillés sur la figure 10.12. Elles sont très similaires à celles obtenues sans filtrage : la direction « statique » a disparu tandis que la nouvelle première direction se superpose avec la deuxième sans filtrage, la deuxième avec la troisième et ainsi de suite.

Les valeurs propres des composantes calculées sans filtrage de l'effet statique soulignent la différence d'ordre de grandeur avec les effets dynamiques. La première composante principale est en effet largement prédominante avec 99,32 % de variance expliquée. Les valeurs propres calculées après filtrage de l'effet statique illustrent quant à elles l'intérêt

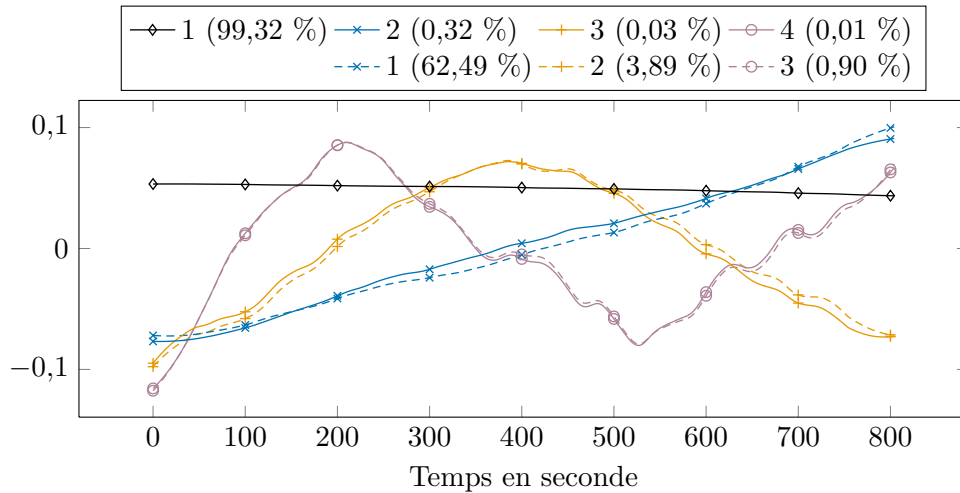


Figure 10.12. : En trait plein, les quatre premières directions principales calculées à partir de la matrice de covariance de l'échantillon de la figure 10.11 *sans* soustraction préalable de la moyenne temporelle. En pointillé, les trois premières directions correspondantes après soustraction de la moyenne temporelle. Les courbes de réponse ont été lissées pour améliorer la lisibilité du graphique.

d'utiliser la matrice de corrélation pour corriger l'effet d'échelle entre les deux premières composantes.

10.4. Analyse de sensibilité de la sortie réduite

La réduction de dimension décrite dans la section précédente a permis de déduire deux nouvelles variables de sortie, à savoir les scores des deux premières composantes principales. On peut calculer, à l'aide des mêmes échantillons que précédemment, des indices de Sobol' compacts, définis dans le chapitre 7. La figure 10.13 représentent les indices de Sobol' compacts d'ordre 1 et totaux des scores de la première composante principale. Chaque paire de barres représente l'indice d'ordre 1 et l'indice total d'une demi-plaque. Les plaques sont disposées de bas en haut comme dans le g. v., celles de la jambe chaude à gauche et celles de la jambe froide à droite. Les barres d'erreur représentent les bornes des intervalles de confiance *bootstrap* pour un risque de 0,1.

La sensibilité de la pente des courbes de réponse augmente avec la cote des plaques. Les sensibilités des deux jambes sont comparables, avec une légère prépondérance, pour la plaque supérieure, de la jambe froide. Les indices d'ordre 1 sont toujours proches des indices totaux correspondants. Autrement dit, il n'y a pas d'interaction notable entre les demi-plaques. La figure 10.14 représente avec les mêmes conventions graphiques les indices des scores de la deuxième composante principale. On observe une asymétrie très prononcée entre jambes : la courbure des réponses de n. g. l. est dominée par les

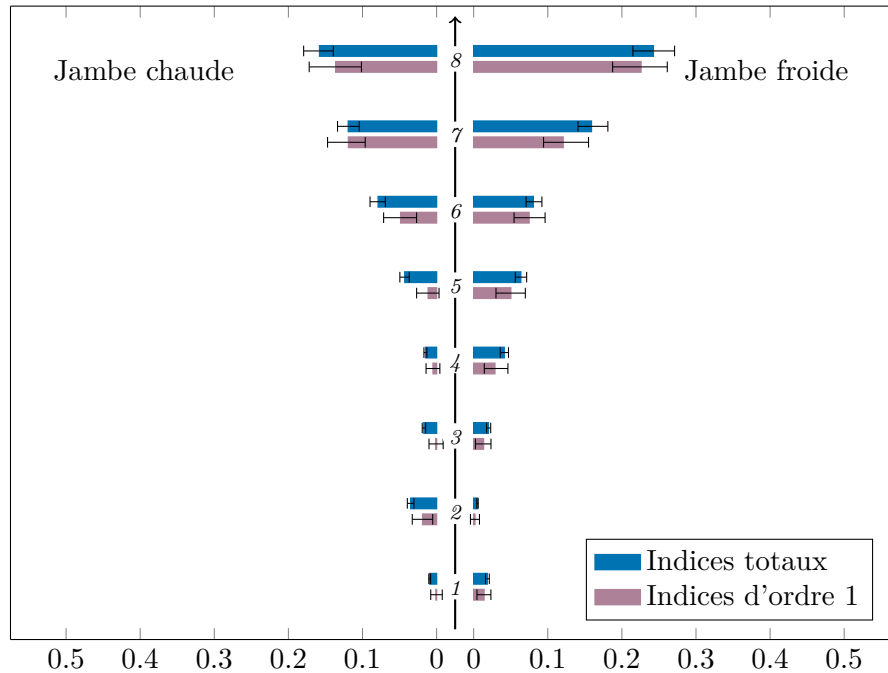


Figure 10.13. : Indices de sensibilité compacts d'ordre 1 et totaux des scores de la première composante principale.

semi-plaques basses en jambe chaude tandis que la jambe froide n'a presque aucun effet. La demi-plaque n° 2 côté chaud totalise près de la moitié de la sensibilité totale des effets d'ordre 1, et les deux suivantes respectivement 15 % et 10 %. Ces demi-plaques interagissent fortement entre elles et avec les autres demi-plaques, notamment avec la demi-plaque n° 1 en jambe chaude qui n'a d'effet que par interaction. La coïncidence de la localisation de cet effet important avec la cote de retournement des débits transverses laisse penser que ces derniers en sont la cause, d'autant plus que cet effet n'était pas présent dans la version 2 du modèle. L'annexe G page 252 regroupe deux articles décrivant les résultats obtenus avec cette version du modèle.

Compte-tenu de l'allure des deux premières composantes principales vues dans la section 10.3.1, il était possible d'anticiper certains des résultats qui viennent d'être exposés. La première direction principale est d'allure rectiligne ; elle agit principalement sur le début et la fin des courbes de réponse. Les plaques dont les indices compacts de la première composante sont importants sont celles dont les indices séquentiels sont importants pendant cet intervalle, autrement dit les plaques présentant un « creux central ». La deuxième composante principale d'allure parabolique agit quant à elle principalement sur le milieu du transitoire en modifiant la courbure des courbes de réponse. Les demi-plaques ayant un impact fort sur les scores de la deuxième composante sont celles qui présentent un « pic central » important avec en tête, la demi-plaque n° 2 en jambe chaude.

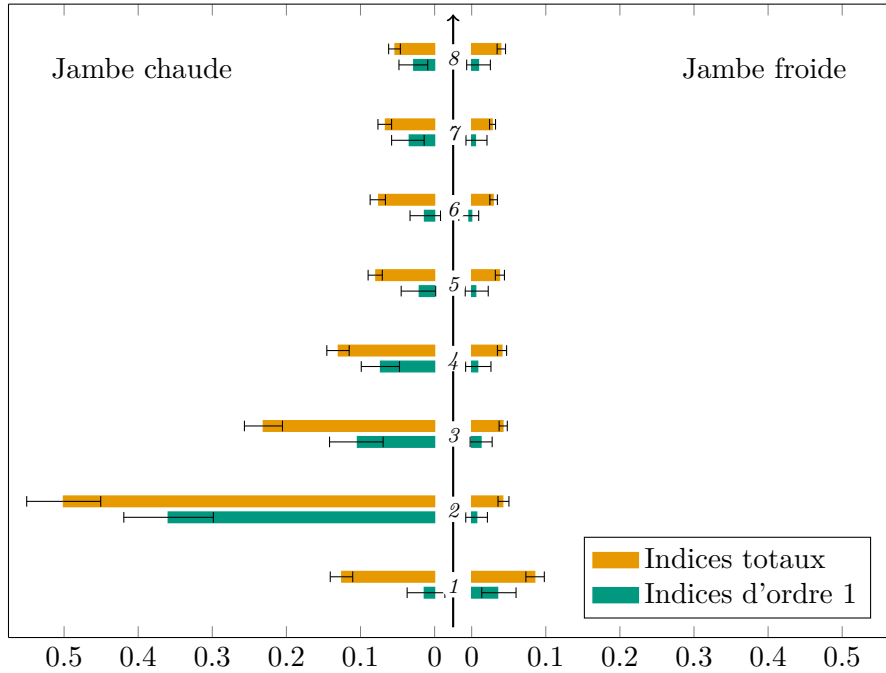


Figure 10.14. : Indices de sensibilité compacts d'ordre 1 et totaux des scores de la deuxième composante principale.

10.5. Réduction de la dimension d'entrée

L'algorithme de filtrage particulière a permis d'observer que le diagnostic du colmatage par analyse des courbes de réponse de n. g. l. est plus précis lorsque l'on considère des projections sur des sous-espaces plus petits. La recherche des sous-espaces optimaux consiste en quelque sorte à regrouper les variables d'entrée selon qu'elles ont des effets analogues sur la réponse de n. g. l. Dans la section 10.2, deux familles de variables d'entrée ont été identifiées : celles ayant une courbe d'indices séquentiels « à pic » ou « à creux ». Cependant, la transition des courbes « à pic » vers les courbes « à creux » est continue et ces familles n'ont pas de définition stricte. La méthode de SIR présentée dans le chapitre 9, permet de trouver des sous-espaces de projections pertinents sans avoir à ajuster de modèle.

Les scores de chaque composante principale permettent de définir une nouvelle variable de sortie. Dans la suite, \mathbf{X} désignera le vecteur d'entrée contenant les taux de colmatage des 16 demi-plaques, s_i les scores de la $i^{\text{ème}}$ composante principale, K_i la dimension de l'espace e. d. r. associé et $\beta_{i,j}$ la $j^{\text{ème}}$ direction e. d. r. trouvée par la SIR. Avec ces notations, le modèle de la SIR présenté dans la section 9.1 s'écrit :

$$s_i = g(\beta'_{i,1} \mathbf{X}, \beta'_{i,2} \mathbf{X}, \dots, \beta'_{i,K_i} \mathbf{X}, \epsilon), \quad (10.3)$$

où g est une fonction arbitraire inconnue et ϵ un terme d'erreur indépendant de \mathbf{X} .

La section 10.5.1 expose les résultats obtenus par la SIR « classique », c'est-à-dire en traitant chaque nouvelle variable de sortie indépendamment. Ces résultats obtenus avec un échantillon gaussien seront comparés au fil du texte avec ceux obtenus avec un échantillon uniforme présentés dans l'annexe C p. 236. L'utilisation du découpage marginal groupé, l'une des variantes multidimensionnelles de la SIR, est présentée dans la section 10.5.2.

10.5.1. Régression inverse par tranche monodimensionnelle

Un moyen simple de satisfaire la condition de linéarité énoncée dans la section 9.1.2 est de choisir \mathbf{X} gaussien car la distribution multi-gaussienne est à symétrie elliptique. Un échantillon de 10 031 configurations de colmatage distribuées selon une distribution gaussienne multivariée de moyenne 0,35 et d'écart type $\frac{0,35}{3}$, tronquée en deçà de 0 et au-delà de 0,7 a été généré. Du fait de l'écart type relativement faible choisi, la troncature n'affecte que marginalement la symétrie elliptique de la distribution. Les résultats présentés dans ce chapitre ont été obtenus à partir des scores des composantes principales calculées en corrélation. Ce choix n'affecte pas de manière importante les résultats finaux. Le nombre de tranches n'a pas non plus d'impact fort. Après quelques tests, il a été choisi d'utiliser des tranches d'environ 300 individus.

Dimensions des espaces de réduction de dimension efficace

La dimension de l'espace e. d. r. peut être déterminée à l'aide de l'estimateur *bootstrap* \widehat{R}_j de la fonction de risque définie dans la section 9.2.1. En pratique, on calcule \widehat{R}_j successivement pour chaque valeur de j comprise entre 1 et 16. Lorsque j dépasse la dimension e. d. r. réelle, la valeur de \widehat{R} chute brutalement, après quoi elle augmente à nouveau régulièrement. La dimension e. d. r. est la valeur de j qui précède la chute de \widehat{R} . Ce phénomène est expliqué dans la section 9.2.1. Les figures 10.15 et 10.16 représentent, par des diagrammes de Tukey, les distributions de 1000 répliques *bootstrap* de l'estimateur de la fonction de risque, \widehat{R}_j , en fonction de la valeur de j en choisissant comme variable de sortie respectivement les scores des première et deuxième composantes principales. Les cercles mauves sont les valeurs en dehors de l'intervalle délimité par les bouts des moustaches.

On observe dans les deux cas, une chute brutale pour $j = 4$ ce qui permet de déduire sans ambiguïté que les espaces e. d. r. correspondants aux deux premières composantes principales sont tous deux de dimension 3. La nette augmentation de la variance au passage de $j = 3$ à $j = 4$ conforte cette conclusion. Les dimensions e. d. r. trouvées avec l'échantillon uniforme sont aussi égales à 3 comme l'attestent les diagrammes des figures C.1 et C.2 (annexe C p. 236) qui sont très similaires à ceux des figures 10.15 et 10.16.

Directions de réduction de dimension efficace

Les figures 10.17 et 10.18 représentent les bases des espaces e. d. r. estimées respectivement avec les scores de la première et de la deuxième composante principale. Certaines

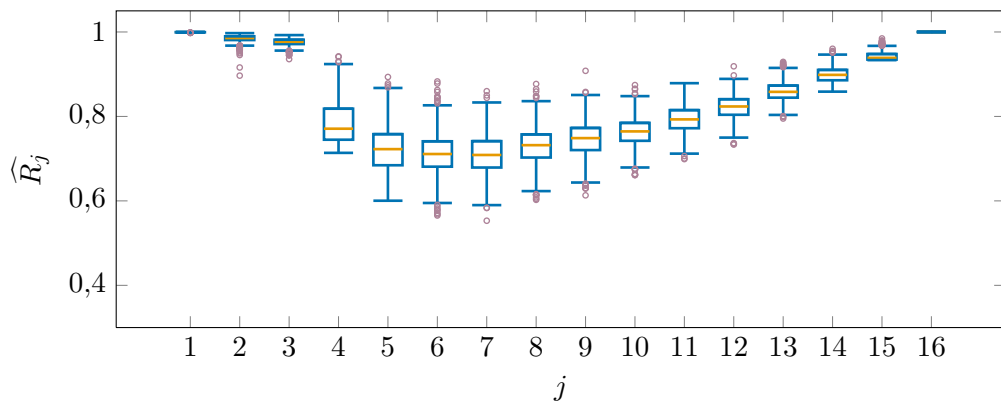


Figure 10.15. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon gaussien et les scores de la première composante principale comme variable de sortie.

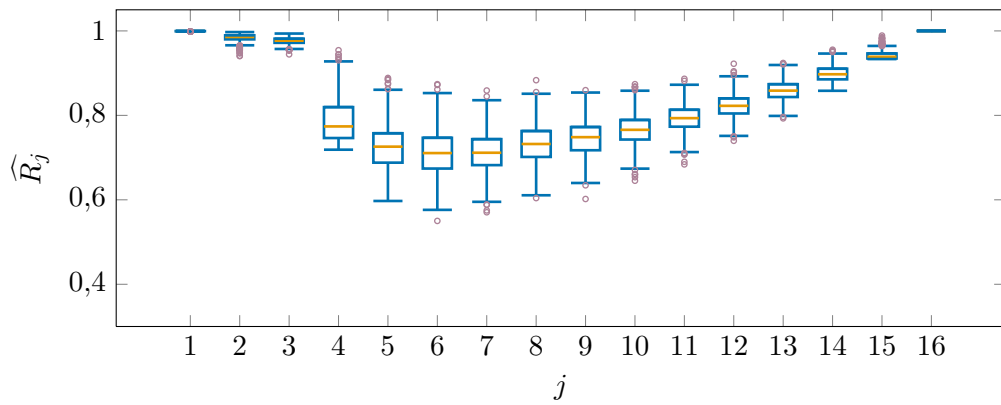


Figure 10.16. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon gaussien et les scores de la deuxième composante principale comme variable de sortie.

directions de projections ne présentent aucune monotonie, par exemple la troisième direction de la première composante principale (—○—, figure 10.17). Seules, elles ne présentent pas d'intérêt pour le diagnostic car des moyennes de taux de colmatage aussi erratiques ne seraient pas interprétables physiquement. On peut cependant chercher à combiner les directions obtenues pour obtenir des moyennes intelligibles. C'est aspect sera abordé dans la section 10.5.2 à l'aide du découpage marginal groupé.

Les premières directions obtenues pour chaque composante principale sont plus régulières. On remarque que leurs composantes évoquent les indices de sensibilités compacts présentés dans les figures 10.13 et 10.14. De fait, on peut montrer par le calcul que si l'espace e. d. r. est une droite et que la fonction de lien inconnue est affine, alors les indices de Sobol' d'ordre 1 sont proportionnels aux carrés des composantes correspondantes du vecteur directeur e. d. r. Plus précisément, en notant ce dernier β , pour chaque variable d'entrée indexée i l'identité suivante est vérifiée :

$$S_i = \frac{(\beta' \delta_i)^2}{p \|\beta\|_2^2}, \quad (10.4)$$

où les (δ_i) sont les vecteurs de la base canonique. Plus la fonction de lien du modèle de la SIR est non linéaire, moins cette identité est valide. Des tests sur des modèles factices ont permis d'observer qu'elle est néanmoins une bonne approximation même pour des modèles éloignés d'une fonction affine. Ainsi, la correspondance observée dans le cas présent entre les carrés des composantes des premières directions e. d. r. de chaque composante principale et les indices compacts correspondants laisse penser que les espaces e. d. r. sont presque des droites. Autrement dit, l'extension du nuage de points dans les directions n^{os} 2 et 3 est probablement très faible. Compte-tenu de l'allure chahutée de ces directions supplémentaires, on peut supposer qu'elles soient des artefacts.

Les directions e. d. r. trouvées avec l'échantillon uniforme sont représentées par les figures C.3 et C.4 de l'annexe C p. 236. Elles sont très similaires à celles trouvées avec l'échantillon gaussien, en particulier les directions $\beta_{1,1}$ et $\beta_{2,1}$ qui sont pour ainsi dire identiques.

Corrélation entre les nouvelles variables d'entrée et de sortie

L'objectif final est de prédire les projections sur l'espace e. d. r. à l'aide des courbes de réponse mesurées, ou plus précisément, à partir de leurs scores selon les deux premières composantes principales. Les scores d'une courbe de réponse mesurée se calculent par projection sur les directions principales. Les courbes de réponse sont échantillonnées à intervalles de temps réguliers mais la durée des transitoires réels est variable. La problématique de l'adaptation des transitoires mesurés à un transitoire choisi comme référence fait l'objet du chapitre 12. Cette paire de scores est tout ce dont on dispose pour prédire les projections selon les directions e. d. r. Ainsi, une direction e. d. r. ne sera intéressante pour le diagnostic, que si elle est fortement corrélée avec au moins une des deux directions principales.

La figure 10.19 représente les nuages de points des coordonnées des individus de l'échantillon gaussien selon les 3 directions e. d. r. en fonction de leurs scores selon la

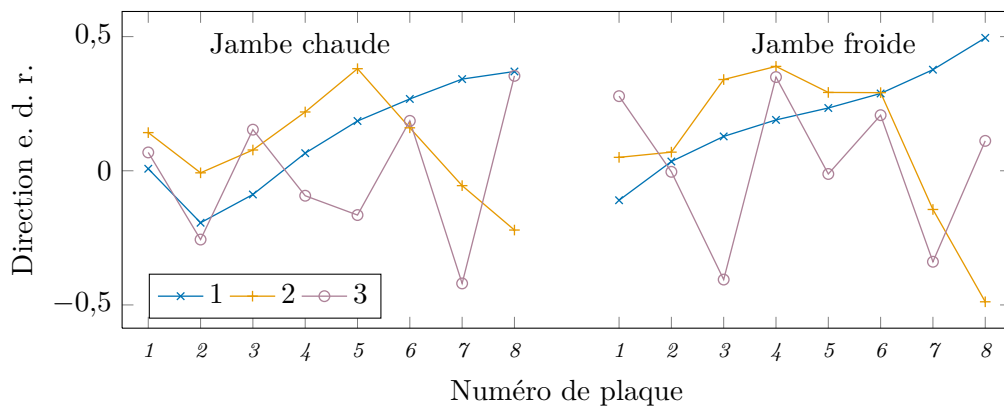


Figure 10.17. : Trois premières directions e. d. r. calculées avec l'échantillon gaussien et les scores de la première composante principale comme variable de sortie.

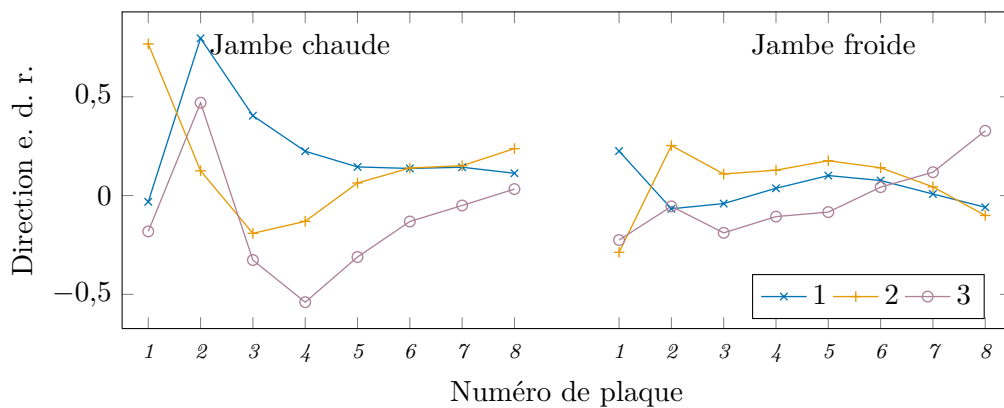


Figure 10.18. : Trois premières directions e. d. r. calculées avec l'échantillon gaussien et les scores de la deuxième composante principale comme variable de sortie.

première composante principale (cercles bleus, \circ). La trajectoire indiquée par des flèches noires (\nearrow) correspond à 15 profils de colmatage utilisés par la méthode de diagnostic à profil choisi, représentés par la figure fig:profil-Bernier page 44. Le sens des flèches indique l'augmentation du colmatage depuis le profil de taux de colmatage nul à celui atteignant 0,7 en plaque supérieure en jambe chaude. Ces profils sont situés dans une région de l'espace des colmatages distincte de celle couverte par l'échantillon gaussien. Ils comportent en effet des valeurs fortes et faibles qui sont très peu probables pour un échantillon gaussien dont les taux de colmatage de chaque demi-plaque sont centré sur 0,35.

Les vecteurs directeurs été normés de façon à ce que cette trajectoire soit contenue dans le cercle unité. Le premier nuage traduit une relation forte et monotone entre s_1 et $\beta'_{1,1}\mathbf{x}$ qui pourrait être utilisée pour le diagnostic. Le coefficient de corrélation correspondant vaut $\rho_{11}^2 = 0,90$. Cette relation semble globale puisque la trajectoire des profils de la méthode de diagnostic à profil choisi traverse le nuage gaussien, en dépit des différences mentionnées précédemment. Autrement dit, il semble que la pente des courbes soit essentiellement déterminée par la moyenne des taux de colmatage avec les poids donnés par $\beta_{1,1}$, sans grande influence de la répartition verticale. Cette hypothèse est étayé par le premier graphique de la figure C.1 de l'annexe C page 236 représentant avec les même convention le nuage obtenu avec un échantillon uniforme. L'échantillon uniforme couvre une région beaucoup plus vaste de l'espace des colmatages que l'échantillon gaussien et le nuage est par conséquent plus épais. La relation entre score et projection e. d. r. est moins linéaire est la corrélation moins forte. Le coefficient de corrélation linéaire vaut $\rho_{11}^2 = 0,87$.

Les deux autres nuages de la figure 10.19 ne font au contraire apparaître aucune relation claire et les coefficients de corrélation linéaire associés sont très faibles, inférieurs à 10^{-4} . Il en est de même des nuages correspondants obtenus avec l'échantillon uniforme, représentés par la figure C.1 de page 236. Cela renforce l'hypothèse selon laquelle ces directions seraient des artefacts. Quoi qu'il en soit elles ne peuvent être exploitées pour le diagnostic.

La figure 10.20 représente les nuages obtenus en utilisant les scores de la deuxième composante principale comme variable de sortie. La relation entre les coordonnées selon la première direction e. d. r. et les scores est moins nette que dans le cas de la première composante principale. La trajectoire des profils de la méthode de diagnostic à profil choisi ne coïncide pas avec le nuage ce qui traduit une localisation de l'effet. Autrement dit, la courbure des courbes de réponse est sensible à la répartition verticale du colmatage. Par ailleurs, cette trajectoire n'est pas monotone en s_2 , si bien que l'application sous-jacente liant le score s_2 à la projection e. d. r. $\beta'_{2,1}\mathbf{x}$ est surjective. Ce dernier point engendre une indétermination rédhibitoire pour le diagnostic. De même que pour la première composante principale, les deux directions e. d. r. suivantes ne sont pas informatives.

Les nuages « croisés », par exemple $\beta'_{1,1}\mathbf{x}$ en fonction de s_2 sont aussi désordonnés. Ils ne sont pas reproduits ici, l'idée d'utiliser simultanément les scores des deux composantes principales étant formalisée par le découpage marginal groupé, exposé dans la section suivante.

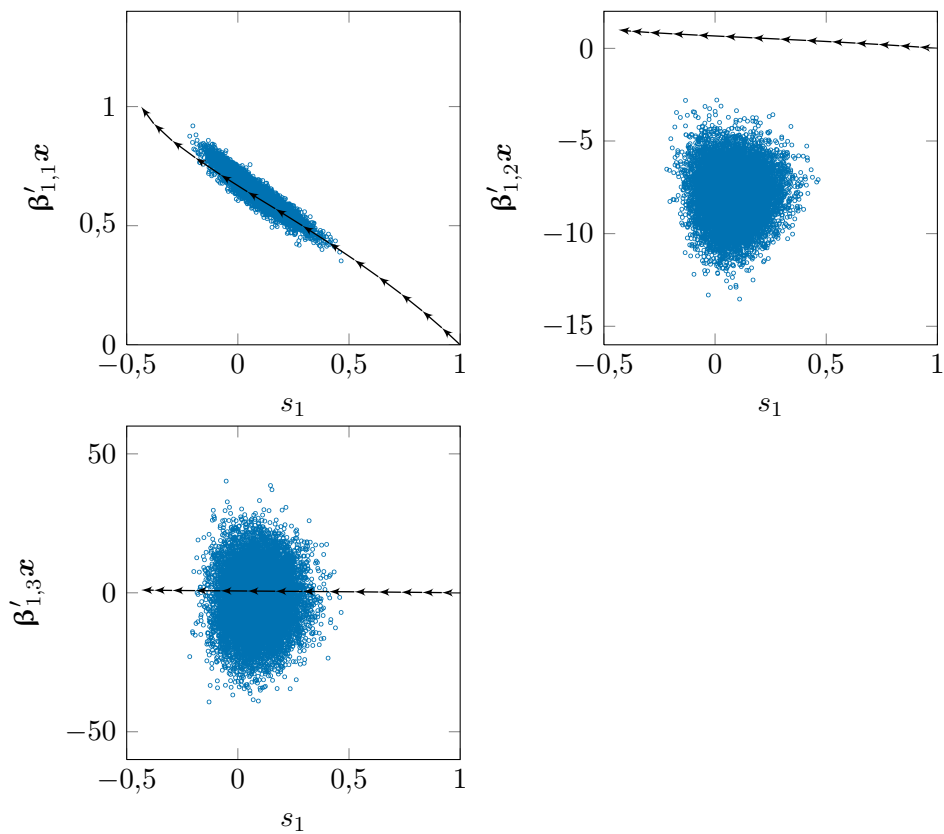


Figure 10.19. : Nuages des projections des individus de l'échantillon gaussien selon les directions e. d. r. obtenues avec les scores de la première composante principale, s_1 , comme variable de sortie, en fonction de s_1 .

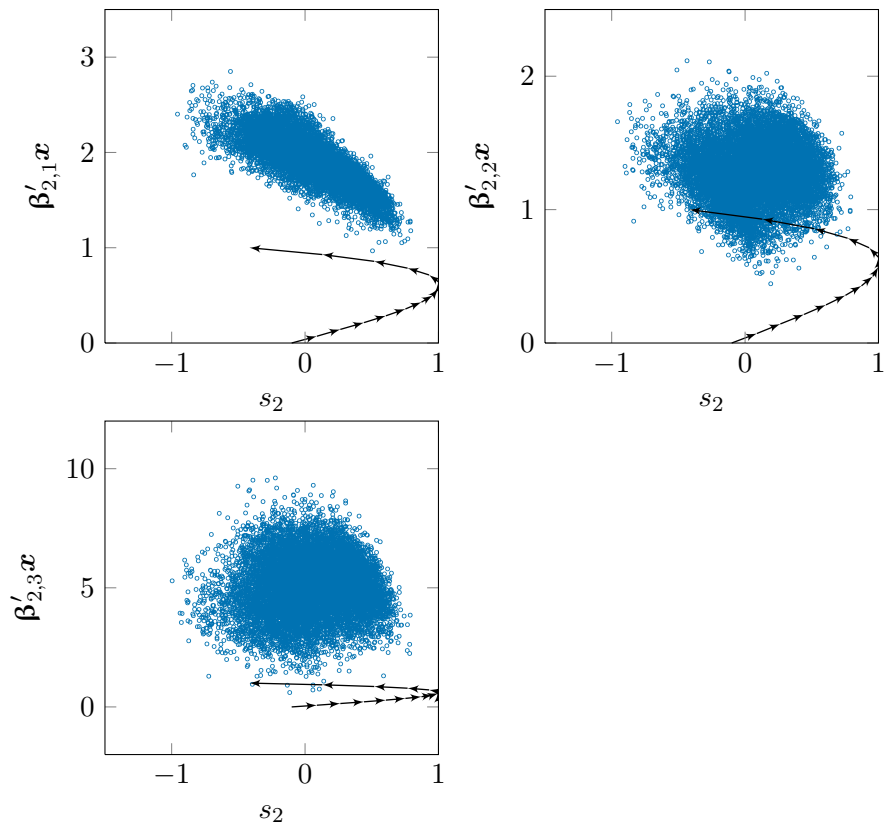


Figure 10.20. : Nuages des projections des individus de l'échantillon gaussien selon les directions e. d. r. obtenues avec les scores de la deuxième composante principale, s_2 , comme variable de sortie, en fonction de s_2 .

10.5.2. Découpage marginal groupé

Parmi les 6 directions e. d. r. identifiées à l'aide de la SIR monodimensionnelle, seule $\beta_{1,1}$, la première direction obtenue avec les scores de la première composante principale, possède les propriétés nécessaires à son utilisation pour le diagnostic :

- une corrélation forte entre les scores et la projection e. d. r. ;
- une relation presque injective entre les scores et la projection e. d. r. ;
- une projection e. d. r. interprétable physiquement, c'est-à-dire pas trop erratique.

Cependant, certaines directions e. d. r. pourraient avoir été perdues lors de la réduction de la dimension de sortie par analyse en composantes principales. En effet, l'analyse en composantes principales est une méthode d'apprentissage non supervisé : les nouvelles variables de sortie ont été obtenues sans tenir compte de l'entrée. Le découpage marginal groupé, décrit dans la section 9.3, peut être utilisé pour essayer de trouver une autre direction e. d. r. dont les projections soient corrélées avec une combinaison des scores des deux composantes principales.

Les diagrammes de Tukey pour la détermination de la dimension obtenue par découpage marginal groupé de l'échantillon gaussien sont représentés dans la figure 10.21. On remarque un décrochement dès $j = 2$. Le minimum de \widehat{R}_j est obtenu pour $j = 3$ et l'on observe ensuite l'augmentation graduelle habituelle. La variance de \widehat{R}_j augmente nettement lors du passage de j de 2 à 3. On peut donc conclure que l'espace e. d. r. est un plan.

La figure 10.22 représente la base de ce plan déterminée par la SIR. On retrouve les directions $\beta_{1,1}$ et $\beta_{1,2}$ obtenues par la SIR monodimensionnelle. La figure 10.25 représente schématiquement la situation en ne considérant que trois variables d'entrée. Les courbes de régression inverse, $E(\mathbf{X}|s_1)$ en mauve (\dashleftarrow) et $E(\mathbf{X}|s_2)$ en vert (\dashrightarrow), sont presque des droites. Les courbes évoluent dans des espaces à trois dimensions et la SIR monodimensionnelle donne pour chacune trois directions e. d. r. Pour chaque courbe, la première direction e. d. r. est dirigée dans l'axe de la courbe et est nettement prépondérante. La première direction de chaque courbe est figurée par une flèche dans la figure 10.25 : en bleu (\nearrow) pour $E(\mathbf{X}|s_1)$ et en orange (\nearrow) pour $E(\mathbf{X}|s_2)$. Le découpage marginal revient en quelque sorte à rechercher le sous-espace contenant les deux courbes de régression inverse. Les variations hors du plan, qui étaient déjà faibles, deviennent véritablement négligeables et l'espace trouvé est le plan représenté par l'aplatissement de couleur bleu ciel entre les directions $\beta_{1,1}$ et $\beta_{1,2}$.

Sans surprise, les nuages des projections selon ces directions en fonction des scores sont similaires à ceux obtenus par la SIR monodimensionnelle. Ces nuages sont représentés par les figures 10.23 et 10.24. Les coordonnées selon $\beta_{1,1}$ sont fortement corrélées avec s_1 et la relation est d'apparence linéaire. Les coordonnées selon $\beta_{1,2}$ sont passablement corrélées avec s_2 mais la trajectoire des profils de la méthode de diagnostic à profil choisi comporte, là encore, un méplat vertical. Les coefficients de corrélation linéaire de ces deux nuages valent respectivement $\rho_1^2 = 0,90$ et $\rho_2^2 = 0,64$. Les nuages « croisés » ne présentent aucune relation identifiable et les coefficients de corrélation correspondants sont inférieurs à 10^{-4} . La méthode de SIR alternée a aussi été testée en prenant comme variables de sortie les scores des quatre premières composantes principales. Elle aboutit finalement

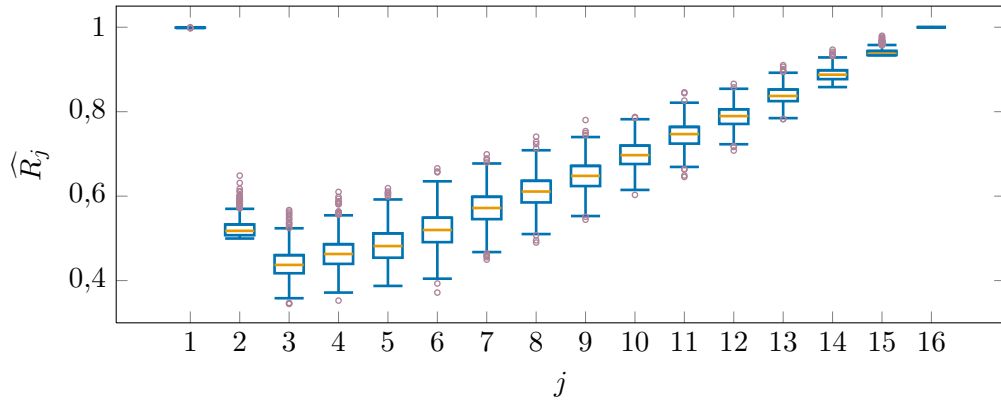


Figure 10.21. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues par découpage marginal groupé de l'échantillon gaussien.

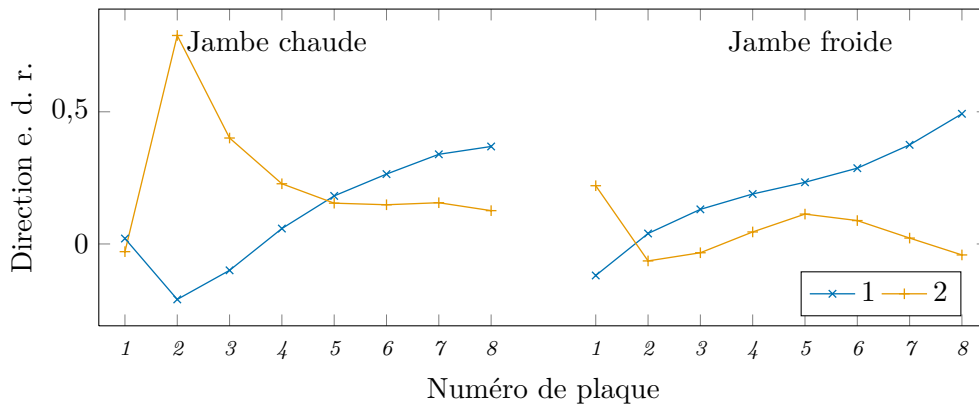


Figure 10.22. : Deux premières directions e. d. r. obtenues par découpage marginal groupé de l'échantillon gaussien.

aux deux même directions identifiées par le découpage marginal groupé.

Les méthodes de SIR multidimensionnelles utilisées ne font pas apparaître de nouvelle direction de projection intéressante pour le diagnostic. La seule véritablement utilisable est donc $\beta_{1,1}$, la première direction e. d. r. obtenue avec les scores de la première composante principale, s_2 , comme variable de sortie. La nouvelle méthode de diagnostic consiste donc à prédire une moyenne des taux de colmatage avec des poids croissant avec l'altitude dans le g. v. par la pente globale des courbes de réponse de n. g. l. La mise en œuvre pratique de cette méthode sera décrite plus en détail dans le chapitre 13.

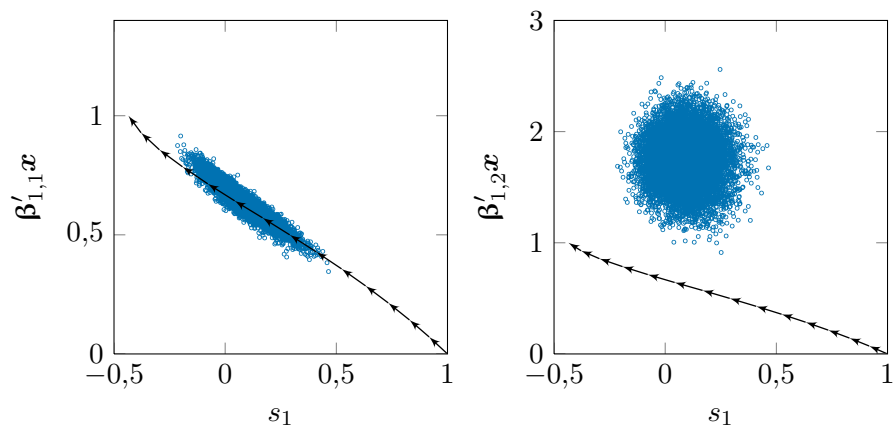


Figure 10.23. : Nuages des projections des individus de l'échantillon gaussien selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la première composante principale, s_1 .

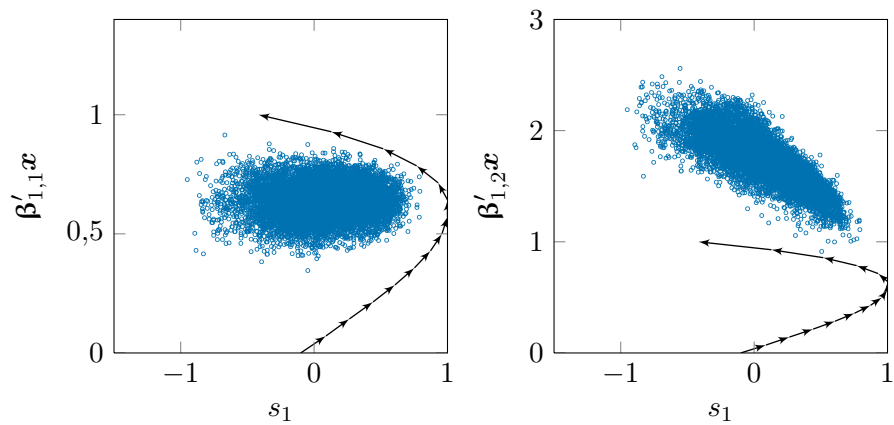


Figure 10.24. : Nuages des projections des individus de l'échantillon gaussien selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la deuxième composante principale, s_2 .

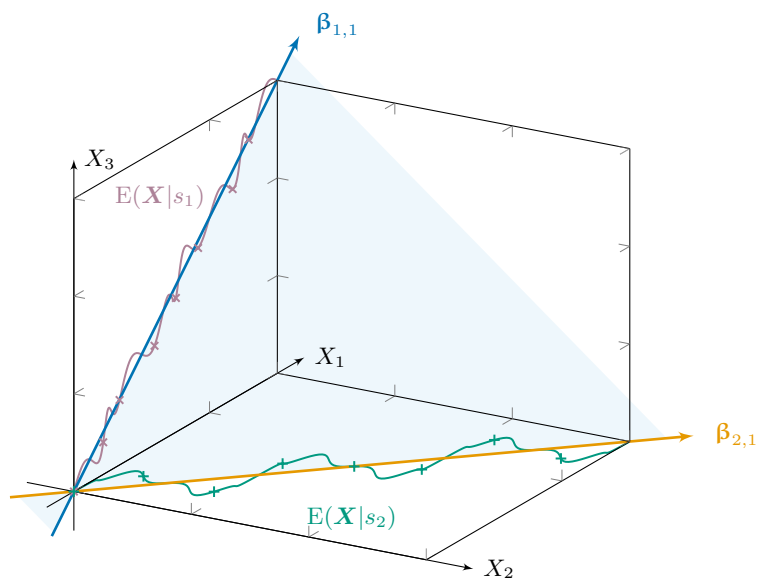


Figure 10.25. : Schéma représentant le plan e. d. r. obtenu par un découpage marginal rassemblant deux courbes de régressions inverses presque rectilignes.

11. Échantillonnage des profils de colmatage

Dans la méthode de diagnostic à profil choisi, la configuration spatiale du colmatage, aussi appelée profil de colmatage, est supposée connue à un facteur multiplicatif près. Cette hypothèse peut conduire à un diagnostic erroné si le profil de colmatage réel est très différent de ceux simulés. On a vu dans la section 3.5.2 page 50 que la réponse du n. g. l. ne permet pas de déterminer le profil de colmatage. En effet, l'algorithme de filtrage particulière est tenu en échec par le nombre trop important de variables et la similarité de leurs effets. Il délimite assez rapidement une zone d'intérêt de l'espace d'état mais ne parvient pas à raffiner cette zone contenant des profils parfois très dissemblables. De fait, il a été montré dans le chapitre 10 que l'information sur le colmatage contenue dans les courbes de réponses peut être résumée à deux variables et que les espaces e. d. r. correspondants sont des droites. Il est donc nécessaire de formuler une hypothèse sur les profils de colmatage.

La réduction de la dimension d'entrée permet de considérer une vaste famille de profils de colmatage possibles sans pour autant rendre le diagnostic inintelligible. Cette hypothèse peut être interprétée comme un conditionnement. En effet, le diagnostic obtenu par la nouvelle méthode est un histogramme de valeurs de la coordonnée e. d. r. choisie comme indicateur du colmatage. Cet histogramme représente la distribution de l'indicateur de colmatage conditionnellement aux données (une réponse de n. g. l. mesurée) et à la distribution de profils de colmatage utilisée pour générer l'échantillon de réponses de n. g. l. simulées. Ainsi, la distribution de profils de colmatage permet de modéliser l'incertitude sur le profil de colmatage réel. Deux distributions ont été utilisées lors du développement de la nouvelle méthode de diagnostic, l'une uniforme sur $[0 ; 0,7]$, l'autre gaussienne de moyenne 0,35 et d'écart type $\frac{0,35}{3}$. Par ailleurs, les régions de l'espace couvertes par des échantillons issus de ces distributions sont très différentes, comme on l'a vu dans la section 9.2.2. Leurs projections selon la direction e. d. r. retenue au chapitre 10 diffèrent elles aussi : la pente du nuage gaussien représenté dans la figure 10.19 est plus faible que celle du nuage uniforme représenté dans la figure C.5 et ce dernier présente une légère courbure pour les colmatage forts. Dans les deux cas, les projections ont une variance faible. En fixant la norme de $\beta_{1,1}$ à 1, la variance de $\beta'_{1,1} \mathbf{X}$ est respectivement de 0,12 pour l'échantillon gaussien et 0,20 pour l'échantillon uniforme. Aussi, les profils de très faible colmatage, correspondant dans la réalité aux g. v. neufs ou après nettoyage, ont une probabilité très faible d'être obtenus en simulant ces distributions. Le nuage obtenu par projection de l'échantillon gaussien selon la direction $\beta_{1,1}$ contient une partie de la trajectoire décrite par les profils, jugés probables, utilisés par la méthode de diagnostic à profil choisi. Le nuage uniforme, au contraire, ne fait que l'effleurer. Ces deux distributions ont été choisies arbitrairement et comportent une proportion importante de profils peu plausibles du point de vue physique. On imagine mal, par exemple, comment deux

demi-plaques adjacentes pourraient avoir des taux de colmatage radicalement différents.

Afin de conditionner le diagnostic par une distribution de profils plus réaliste, un modèle de formation des dépôts de colmatage a été utilisé. Ce modèle exploite les champs thermohydrauliques tridimensionnelles calculés par le modèle de g. v. THYC et repose sur un mécanisme appelé *vena contracta*. Il a été développé par Prusek (à paraître) au cours de sa thèse et est présenté dans la section 11.1. Une analyse de la sensibilité des profils de colmatage aux paramètres de ce modèle est ensuite exposée dans la section 11.2. Enfin, la section 11.3 détaille la façon dont a été construite la distribution de profils de colmatage utilisée ensuite pour le diagnostic.

11.1. Modèle de colmatage par *vena contracta*

Le modèle de colmatage par *vena contracta* utilise des champs thermohydrauliques tridimensionnels simulés avec THYC. Le modèle de g. v. développé à l'aide du code THYC a été décrit dans la section 4.2. Afin de simplifier la mise en œuvre des simulations, on ne simule que le régime nominal. La durée d'un calcul avec THYC est usuellement de 5 à 10 heures sur une station de travail HP Z600.

Le phénomène de déposition est ensuite simulé à l'aide de deux modules appelés BOUTHYC et COLMATHYC. Le module BOUTHYC a été développé par EDF au début des années 2000 (Moleiro et coll., 2010; Pujet, 2001). Il permet de simuler la déposition de particules sur la surface externe des tubes des g. v., autrement dit le phénomène d'encrassement. Le module COLMATHYC est une extension du module BOUTHYC simulant la déposition de particules dans les trous foliés des plaques entretoises (Moleiro et coll., 2010), c'est-à-dire le phénomène de colmatage. Ces deux modules reposent sur les mêmes hypothèses générales. Les particules sont supposées sphériques de diamètre unique d et composées uniquement de magnétite, les espèces dissoutes ne sont quant à elles pas prises en compte. L'eau alimentaire apporte dans le g. v. une fraction massique ϵ_a de particules. Elles sont transportées par le fluide secondaire et peuvent se déposer sur les structures internes du g. v. selon des mécanismes spécifiques à chaque module. Les équations régissant ces mécanismes sont données dans l'annexe B, page 234. Les particules peuvent par ailleurs être éliminées par les purges, en bas de g. v., ou rester en suspension jusqu'aux dispositifs séparateurs et alimenter à nouveau le g. v. en passant par le *downcomer*. Les dépôts sont supposés poreux de porosité unique π_e pour l'encrassement et π_c pour le colmatage.

11.1.1. Calage du paramètre non observable

L'équation (B.5) de l'annexe B page 234 donne la vitesse de déposition du colmatage par le mécanisme de *vena contracta*. Elle fait intervenir un paramètre ω dont la valeur est inconnue. On se propose de la déterminer en fixant la configuration de colmatage de la plaque supérieure à laquelle le modèle doit aboutir à l'issue du phénomène de déposition. Les résultats d'un examen télévisuel sont utilisés comme « valeurs cibles ». Moyennant certaines approximations et en supposant que la valeur de ω est la même dans tout le g. v. et ne varie pas au cours du temps, le modèle COLMATHYC permet d'exprimer les

taux de colmatage unitaires de la plaque supérieure à la fin du processus de déposition, \mathbf{x} , sous la forme d'un système d'équations de la forme suivante (Prusek, à paraître) :

$$\mathbf{x} = \mathbf{a}(\omega) \omega + \mathbf{b}(\omega), \quad (11.1)$$

où $\mathbf{a}(\omega)$ et $\mathbf{b}(\omega)$ sont des vecteurs dépendants des champs calculés avec THYC et de grandeurs calculées par le module COLMATHYC. Ces vecteurs dépendent donc implicitement de ω . En choisissant une configuration de colmatage d'arrivée en plaque supérieure, \mathbf{x}_{cible} , et un temps de déposition total, T_d , la valeur de ω peut être déterminée par la résolution du système non linéaire (11.1). L'algorithme suivant a été utilisé :

1. Choisir une configuration de colmatage d'arrivée en plaque supérieure, \mathbf{x}_{cible} , et fixer le temps de déposition total, T_d , ainsi qu'un critère de convergence numérique, κ .
2. Initialiser le paramètre ω à une valeur arbitraire ω_0 . Fixer $i = 0$.
3. Simuler la déposition pendant toute la durée T_d avec $\omega = \omega_i$. En déduire les vecteurs $\mathbf{a}_i = \mathbf{a}(\omega_i)$ et $\mathbf{b}_i = \mathbf{b}(\omega_i)$.
4. Déduire du système $\mathbf{x}_{cible} = \mathbf{a}_i \omega_{i+1} + \mathbf{b}_i$ une nouvelle valeur ω_{i+1} .
5. Si $|\omega_{i+1} - \omega_i| < \kappa$, incrémenter i d'une unité et retourner à l'étape 3.

Lors de l'étape 4, la nouvelle valeur du paramètre inconnu est déterminée par la méthode des moindres carrés. Notons que le critère de convergence de cet algorithme est d'ordre numérique et ne garantit pas l'adéquation des taux de colmatage simulés en plaque supérieure avec l'examen télévisuel choisi comme référence. Le temps de calcul nécessaire à l'étape 3 est de 10 à 20 minutes et la méthode converge en 10 à 20 itérations lorsque le critère de convergence est fixé à 10^{-4} . Un critère de convergence de 10^{-2} permet de réduire le nombre d'itérations à 5 à 10. Le temps moyen nécessaire à une simulation complète du phénomène de déposition par *vena contracta* est donc de l'ordre de 3 à 4 heures pour une convergence poussée ou de 1 à 2 heures avec un critère de convergence moins strict. Ce temps peut être réduit significativement en choisissant judicieusement la valeur d'initialisation. Une méthode de prédiction de la valeur du paramètre de calage sera exposée dans la section 11.2.1.

11.2. Analyse de sensibilité du modèle *vena contracta*

Le modèle de colmatage par *vena contracta* permet de formuler des hypothèses sur la distribution des profils de colmatage afin de réduire la taille de l'espace à échantillonner. Une distribution de profils de colmatage probables peut être simulée en assignant aux paramètres du modèle des distributions de probabilité. Dans ce but, une analyse de sensibilité a été réalisée afin de déterminer les paramètres du modèle ayant un impact notable sur la forme des profils.

Le paramètre ω étant dépendant des variables d'entrée, on choisit d'en faire une variable interne au modèle, voire une variable de sortie, c'est-à-dire de considérer la méthode de calage comme faisant partie intégrante du modèle dont on analyse la sensibilité. Les variables sélectionnées pour l'analyse de sensibilité sont les suivantes :

Facteur	Unité	Niveaux					
d	μm	0,5	2*	4	6	8	10
ϵ_a	$10^{-9} \text{ kg kg}^{-1}$	8	11*	14			
T_d	jours	4320	6120	7920*			
π_c		0.05*	0,1	0,15			
π_e		0,25	0.30*	0,4			

Tableau 11.1. : Niveaux des facteurs du plan d'expérience pour l'analyse de sensibilité du modèle de colmatage par *vena contracta*. Les valeurs nominales du modèle sont marquées d'un astérisque.

- la fraction massique de particules en entrée du g. v., ϵ_a ;
- le temps de déposition, T_d ;
- la porosité du dépôt d'encrassement, π_e ;
- la porosité du dépôt de colmatage, π_c ;
- le diamètre des particules, d .

Le nombre de variables d'entrée, 5, est restreint mais le temps de calcul supérieur à une heure nécessaire à une évaluation du modèle est en revanche assez contraignant. Un plan factoriel complet à six niveaux pour le diamètre des particules, et à trois pour les autres facteurs a été construit. Les niveaux des facteurs sont donnés dans le tableau 11.1. Six jours ont été nécessaires pour réaliser les 486 simulations avec six processeurs d'une seule machine, le critère de convergence étant fixé à 10^{-2} . A posteriori, la méthode de Morris aurait sans doute abouti aux mêmes conclusions avec un coût moindre en temps de calcul. Le temps de calcul nécessaire au plan factoriel complet est néanmoins raisonnable et il présente l'avantage de sa simplicité de mise en œuvre et d'interprétation.

Le paramètre ω a été calé avec des valeurs de taux de colmatage issues de l'examen télévisuel du g. v. n° 2 de la tranche n° 4 de la centrale Delta en 2006. Il s'agit de l'un des états de colmatage les plus importants ayant été observés. Ce choix permet de simuler le phénomène de déposition dans son intégralité, depuis le g. v. propre jusqu'à un état nécessitant une intervention rapide.

11.2.1. Analyse de sensibilité du paramètre de calage

La valeur du paramètre ω conditionne fortement le phénomène de déposition. Une première analyse de sensibilité a donc été réalisée en le choisissant comme variable de sortie.

Les nuages de points de ω en fonction des cinq facteurs étudiés sont représentés dans la figure 11.1. On peut dès à présent remarquer que les porosités ont très peu d'impact sur ω tandis que le diamètre des particules a un impact fort mais complexe : non linéaire, non monotone, présentant éventuellement une valeur seuil. Pour cette raison, l'importance des quatre premiers facteurs (T_d , π_e , π_c et ϵ_a) est étudiée séparément pour chaque diamètre de particule.

La figure 11.2 représente les indices de sensibilité d'ordre 1 et les interactions d'ordre

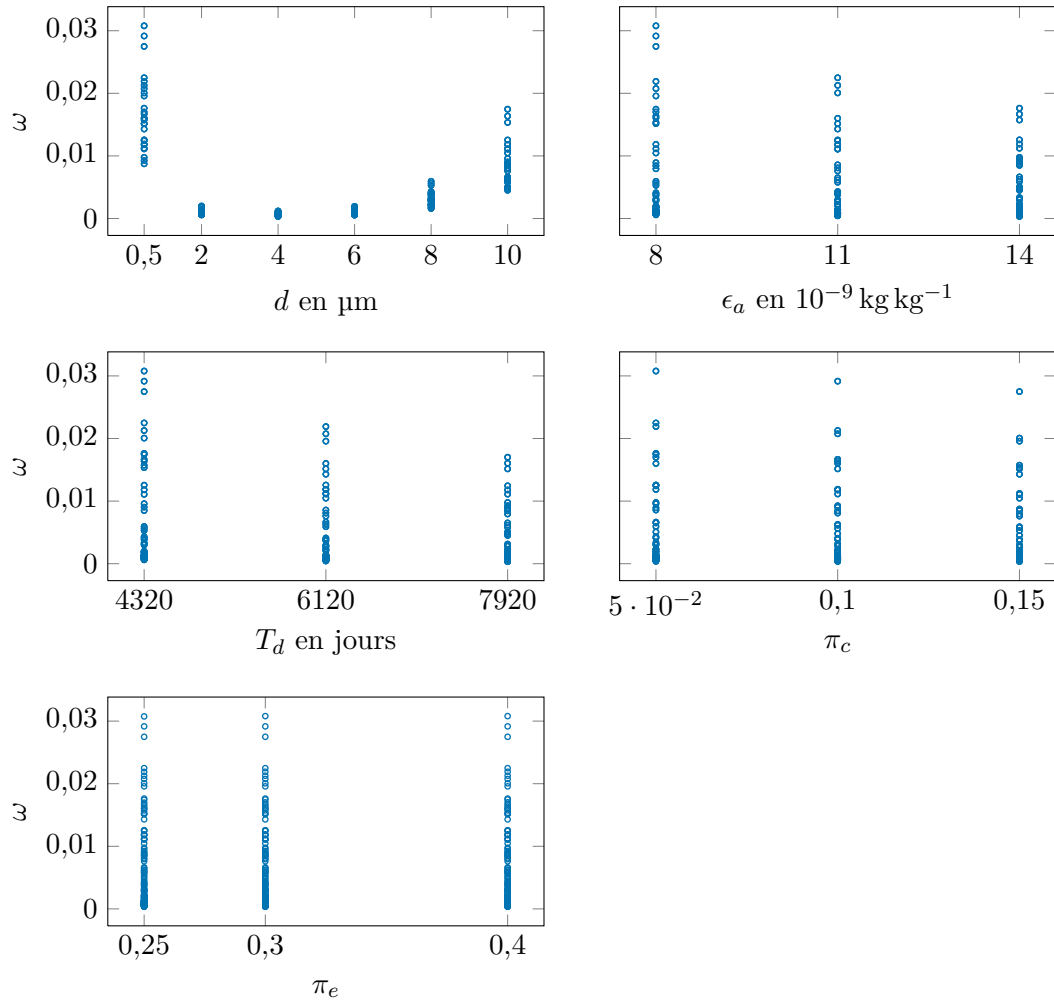


Figure 11.1. : Nuages de points du paramètre de calage ω en fonction des valeurs des cinq facteurs : diamètre des particules (d), fraction massique de particules dans l'eau alimentaire (ϵ_a), temps de déposition total (T_d) et porosités de l'encrassement (π_e) et du colmatage (π_c).

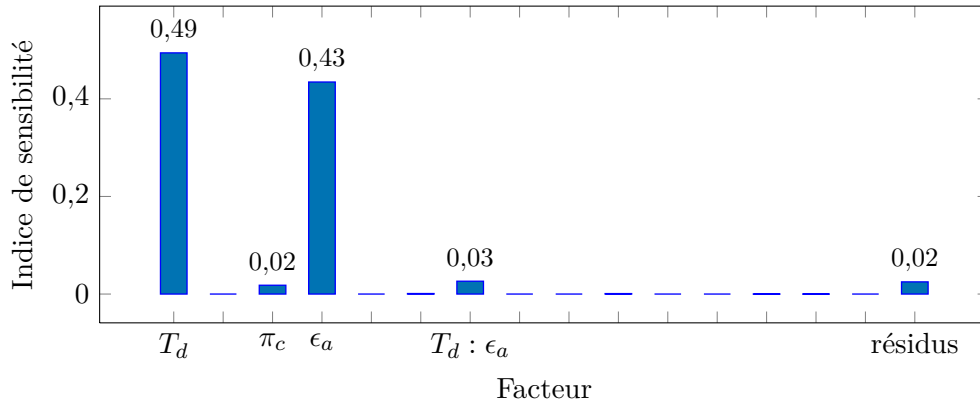


Figure 11.2. : Indices de sensibilité d'ordre 1 du temps de déposition (T_d), de la porosité des dépôts d'encrassement (π_e) et de colmatage (π_c) et de la fraction massique de particules dans l'eau alimentaire (ϵ_a), suivis des interactions entre paires de ces facteurs. Le diamètre des particules vaut $d = 0,5 \mu\text{m}$. Des résultats très similaires ont été obtenus avec les autres valeurs de diamètre. Les interactions à plus de deux facteurs sont négligeables et ne sont pas représentées.

2 obtenues pour la taille de particule $d = 0,5 \mu\text{m}$. Les résultats obtenus pour les autres tailles de particules sont très similaires. La somme des indices d'ordre 1 du temps de déposition et de la fraction massique vaut 0,93. La porosité du colmatage a un effet marginal. La porosité de l'encrassement n'a pour ainsi dire aucun impact, de même que la plupart des interactions, hormis l'interaction à deux facteurs entre le temps de déposition et la porosité du colmatage, qui a un petit effet. Le temps de déposition, la fraction massique, l'interaction entre ces deux facteurs et la porosité du colmatage ont tous des niveaux de signification très proche de 0¹. Les autres facteurs et interactions ne sont pas significatifs. Parmi eux, l'interaction entre le temps de déposition et la porosité du colmatage a le niveau de signification le plus faible, égale à 0,12.

Estimation du paramètre de calage pour accélérer la convergence

Un modèle rudimentaire pour prédire la valeur de ω a été construit à partir des résultats précédents. Ce modèle a pour but d'accélérer la convergence de la méthode de calage en estimant une valeur initiale $\hat{\omega}$ qui soit proche de la valeur finale.

Le temps de déposition et la fraction massique en entrée étant prépondérants, la forme choisie pour le modèle est la suivante :

$$\hat{\omega} = u(d) + v(d) T_d + w(d) \epsilon_a, \quad (11.2)$$

où u , v et w sont trois fonctions du diamètre des particules que l'on peut déterminer

1. On pratique ici un test de Fisher avec pour hypothèse nulle la nullité (individuelle) de chaque indices (Saporta, 2006, chap. 14).

	c_0	c_1	c_2	c_3	c_{-1}
$u (\times 10^2)$	-3,868	1,505	-0,254	$1,666 \cdot 10^{-2}$	4,144
$v (\times 10^6)$	2,127	-0,836	0,142	$-9,429 \cdot 10^{-3}$	-2,262
$w (\times 10^3)$	1,186	-0,463	$7,812 \cdot 10^{-2}$	$-5,125 \cdot 10^{-3}$	-1,269

Tableau 11.2. : Coefficients des fonctions u , v et w du modèle de prédiction du paramètre de calage.

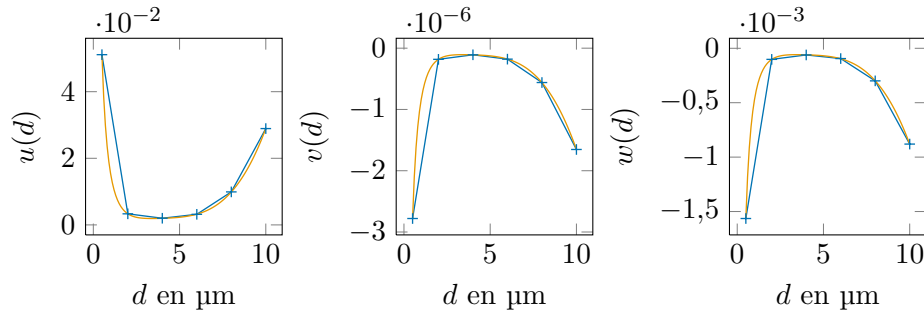


Figure 11.3. : Ajustement (—) de u , v et w à partir des valeurs estimées pour chaque diamètres (+). L'abscisse est le diamètre des particules en μm .

à l'aide des simulations déjà réalisées. On dispose, pour chacune des 6 valeurs de d , de simulations réalisées pour 9 couples de valeurs de T_d et ϵ_a . Des régressions linéaires multiples permettent d'estimer, à partir de ces données, les valeurs prises par les fonctions u , v et w en 6 points. ces valeurs sont ensuite utilisées pour réaliser des ajustements. Les fonctions ajustées sont toutes trois de la forme suivante :

$$x \mapsto c_0 + c_1x + c_2x^2 + c_3x^3 + c_{-1}x^{-1} \quad . \quad (11.3)$$

Les ajustements effectués sont représentés dans la figure 11.3 et les valeurs des coefficients sont données dans le tableau 11.2. L'initialisation de l'algorithme de calage à l'aide de ces fonctions permet une diminution d'environ un tiers du nombre d'itérations (Girard et Prusek, 2012).

11.2.2. Analyse de sensibilité de l'allure des profils

Les graphiques de la figure 11.4 représentent les familles de profils obtenues à la fin de la déposition, par diamètre de particules croissant de gauche à droite et de haut en bas. L'allure des profils de taux de colmatage uniformes équivalents par demi-plaque est contrôlée essentiellement par le diamètre des particules. En effet, pour un diamètre de particule donné, l'allure des profils de colmatage fluctue très peu. Le temps de déposition a une influence importante sur le paramètre de calage ω mais n'influe finalement que sur la cinétique du phénomène de colmatage. L'influence du diamètre n'est sensible qu'à partir de 8 μm .

La figure 11.5 représente des résultats de simulations réalisées avec les valeurs nominales des paramètres du modèle et des diamètres de particule variant de 6 à 10 μm par pas de 0,5 μm . Jusqu'à 8,5 μm , le déplacement n'affecte que la jambe chaude, les taux de colmatage en jambe froide augmentant globalement. On notera que pour les diamètres importants, la configuration en plaque supérieure présente des différences notables avec l'examen télévisuel utilisé comme référence.

D'autres simulations ont permis par ailleurs d'observer l'absence d'effet sur l'allure des profils de colmatage des paramètres de la relation entre le volume déposé et le taux de colmatage, ainsi que de l'examen télévisuel utilisé comme configuration spatiale d'arrivée pour la plaque supérieure. Ces paramètres influencent la cinétique de déposition mais pas la répartition verticale du colmatage moyen (Girard et Prusek, 2012).

11.3. Distribution de profils de colmatage issue du modèle *vena contracta*

L'analyse de sensibilité du modèle de colmatage a révélé que seul le diamètre des particules de magnétite a un impact manifeste sur l'allure des profils de colmatage. Ce paramètre a donc été pris en compte pour la construction de la distribution des profils. D'autres paramètres tels que le temps de déposition total ou la configuration spatiale en plaque supérieure utilisée pour le calage affectent la cinétique de déposition. Cependant, il est plus simple de fixer ces paramètres à leur valeur nominale et de faire varier aléatoirement le temps de déposition partiel pour rendre compte des différentes cinétiques. Ainsi, un profil de colmatage est entièrement déterminé par deux paramètres : un diamètre de particules et un temps de déposition partiel.

11.3.1. Mélange de particules de diamètres différents

Dans la réalité, l'eau alimentaire transporte un mélange de particules de diamètres différents. Le module COLMATHYC ne permet pas de simuler la déposition d'un mélange de particules. Afin de pallier ce manque, une hypothèse a été adoptée : on suppose que le profil de colmatage résultant de la déposition d'un mélange de particules est proche de la moyenne pondérée des profils obtenus pour chaque diamètre de particule. La moyenne est pondérée par des coefficients proportionnels aux masses de chaque ensemble de particules présent dans le mélange.

Des mesures granulométriques ont été réalisées dans la centrale de Charlie (Couvidou, 2011). Le tableau D.1 en annexe (page 242) répertorie les résultats de 22 de ces mesures (Couvidou, 2012). Comme le diamètre des particules influe peu sur la forme des profils en dessous de 8 μm , les données ont été regroupées en 4 classes : de 1 μm à 6 μm , de 6 μm à 8 μm , de 8 μm à 10 μm et de 10 μm à 15 μm . Les particules de diamètre supérieur à 15 μm n'ont pas été prises en compte pour cette application car elles conduisaient à des taux de colmatage supérieurs à 1. Les données du tableau D.1 indiquent que les particules de diamètre très important sont rares. Pour chacune des classes de diamètre de particule, un diamètre représentatif a été choisi et des profils ont été simulés à l'aide du modèle

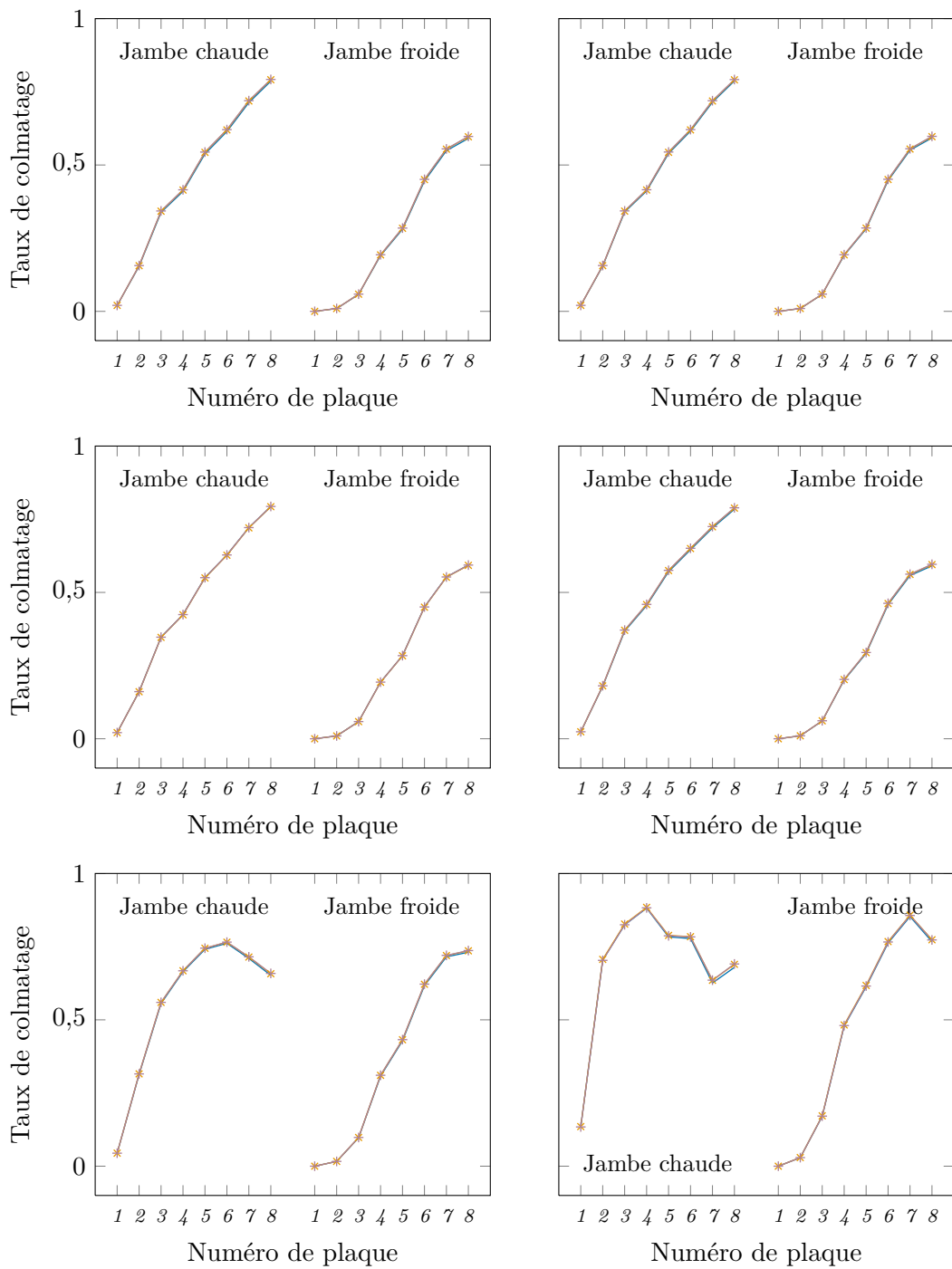


Figure 11.4. : Profils de taux colmatage homogènes équivalents à la fin du processus de dépôt pour chacun des traitements du plan d'expérience.

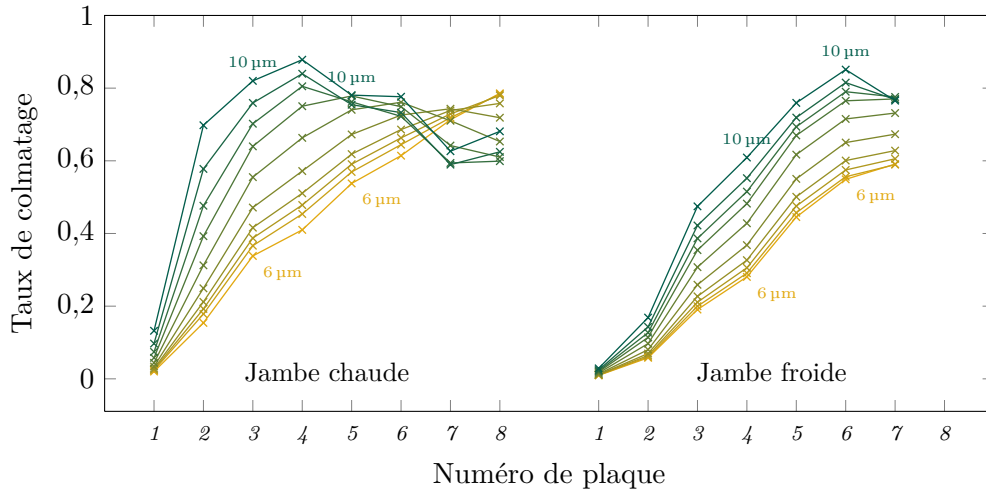


Figure 11.5. : Profils de colmatage homogènes équivalents après 7920 jours de dépositions pour des diamètres de particules variant de 6 à 10 μm par pas de 0,5 μm .

vena contracta pour 22 instants de déposition avec les valeurs nominales des paramètres du modèle. Les profils obtenus ainsi pour des diamètres de particule de 2 μm , 7 μm , 9 μm et 12 μm sont représentés par la figure 11.6. On observe à nouveau qu'à mesure que le diamètre des particules augmente, la position du maximum de colmatage descend dans le g. v. Le choix du diamètre de 2 μm pour la première classe a été fait pour pouvoir réutiliser des simulations déjà réalisées. La figure 11.4 permet de constater que l'allure des profils varie de façon négligeable sur cet intervalle.

Conformément à l'hypothèse formulée au début de ce paragraphe, des pondérations à appliquer aux 4 profils représentatifs ont été calculées à partir de chacun de ces échantillons. Le résultat de ce calcul est représenté graphiquement par la figure 11.7. Le poids appliqué aux profils obtenus pour un diamètre de particule de 2 μm est presque toujours prépondérant. L'observation du tableau D.1 permet de constater que les particules de diamètre supérieur à 6 μm sont rares. En revanche, puisque la masse des particules varie avec le cube du diamètre, les profils associés aux très grosses particules peuvent recevoir des pondérations allant jusqu'à 0,58 pour l'échantillon n° 22.

11.3.2. Rétroaction du champ thermohydraulique

Le colmatage affecte la perte de charge singulière au passage des plaques étroites et modifie donc le champ thermohydraulique. Il est possible de prendre en compte partiellement cet effet en recalculant ponctuellement les champs thermohydrauliques au cours du phénomène de déposition. Chaque actualisation des champs thermohydrauliques nécessite de réaliser une simulation THYC durant plusieurs et elles sont répétées à chaque itération de la méthode de calage du paramètre ω . Pour cette raison, les champs thermohydrauliques n'ont été actualisés que 4 fois pendant la période de déposition.

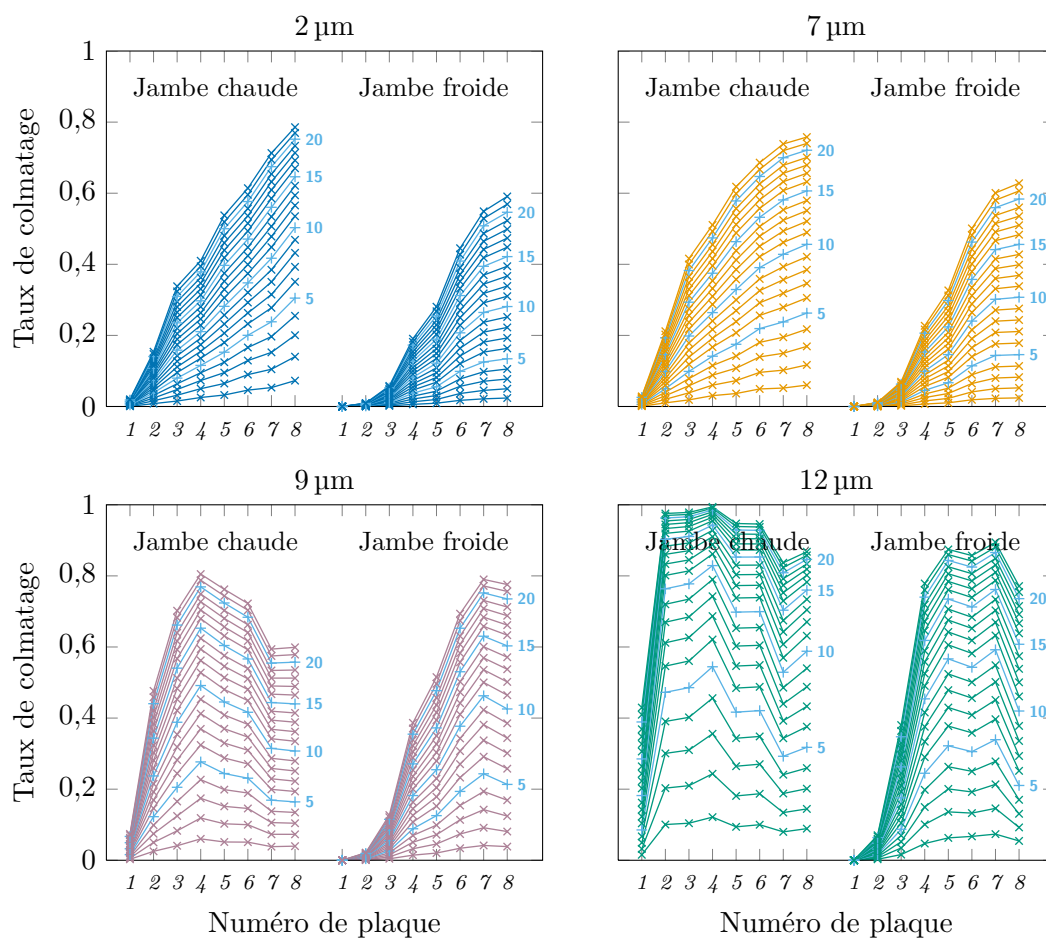


Figure 11.6. : Profils de colmatage obtenus avec le modèle *vena contracta* aux 22 instants de déposition pour les quatre diamètres de particule sélectionnés (2, 7, 9 et 12 μm).

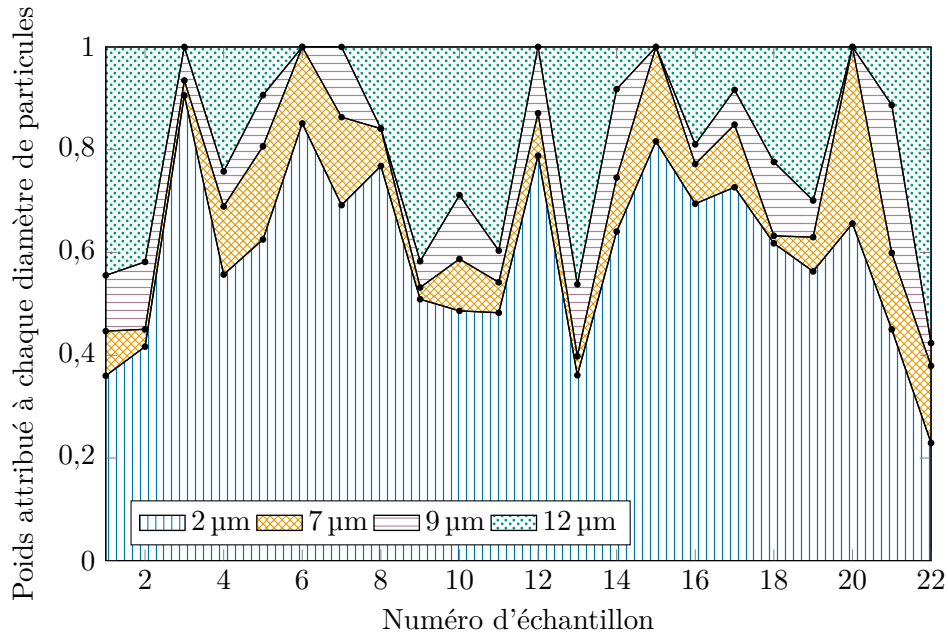


Figure 11.7. : Poids correspondants aux quatre diamètres représentatifs des classes de particules déduits des échantillons granulométriques donnés dans le tableau D.1 de l'annexe D.

Les profils obtenus avec rétroaction pour des tailles de particules de 2, 7 et 9 μm sont représentés par la figure 11.8. La méthode de calage du paramètre ω fonctionne assez mal en conjonction avec la méthode de rétroaction. Les taux de colmatage des plaques supérieures dépassent rapidement la valeur cible et se saturent à 1. Ainsi la rétroaction n'a pas pu être appliquée avec des particules de 12 μm de diamètre.

Des travaux supplémentaires seraient nécessaires pour rendre cette approche plus robuste, notamment en adaptant la méthode de calage du paramètre de modélisation. On peut néanmoins considérer que les profils représentés dans la figure 11.8 sont plausibles et les intégrer à la distribution de profils de colmatage, en plus de ceux obtenus sans rétroaction représentés dans la figure 11.6.

11.3.3. Simulation de la distribution de profils

Considérons d'abord le cas sans rétroaction. On considère d'une part, que l'on ne sait rien de la cinétique de déposition, et d'autre part, que les 22 échantillons granulométriques sont équiprobables. On peut simuler une loi de distribution des profils de colmatage à partir de deux variables aléatoires indépendantes distribuées uniformément. La première détermine un temps de déposition partiel tandis que la seconde sert à sélectionner l'un des échantillons granulométriques. À chaque temps de déposition correspond un jeu de quatre profils de colmatage, un par classe de diamètre de particules. L'échantillon granulométrique sélectionné donne quant à lui les poids à utiliser pour calculer la moyenne

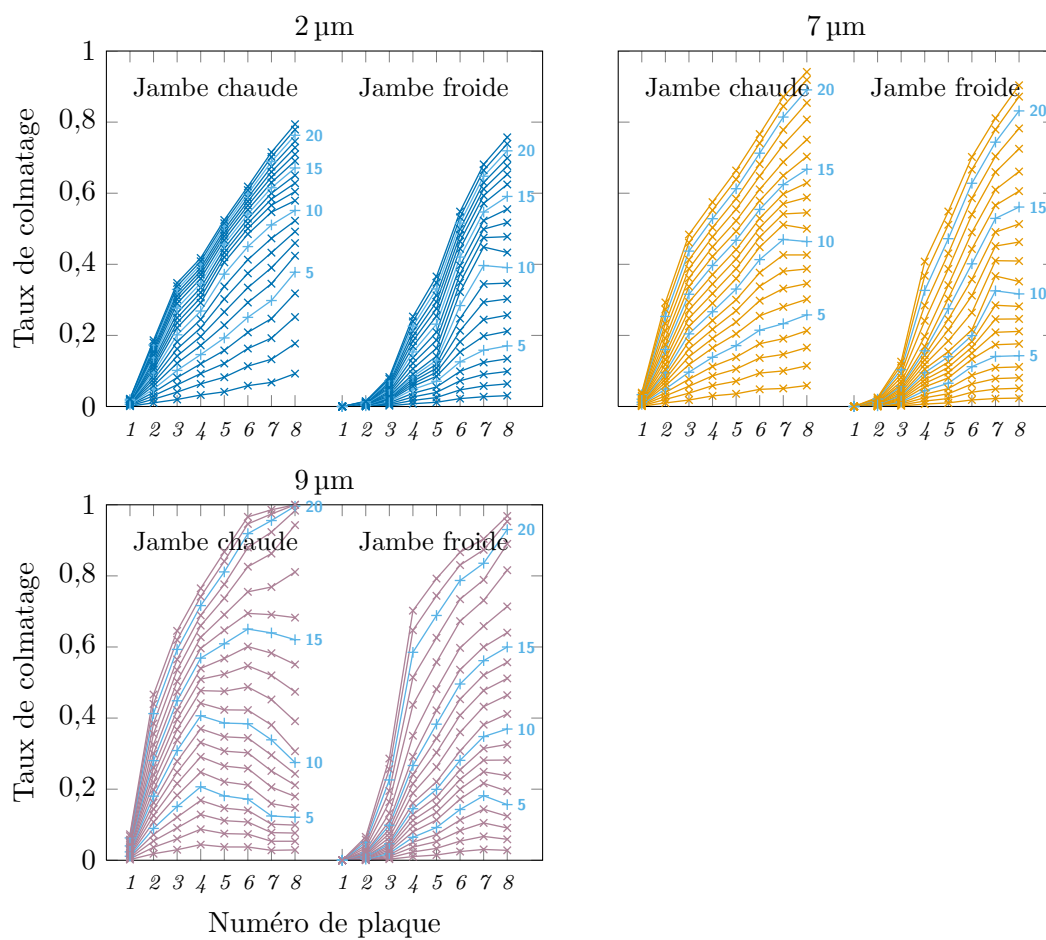


Figure 11.8. : Profils de colmatage obtenus avec le modèle *vena contracta* avec rétroaction thermohydraulique aux 22 instants de déposition pour les trois premiers diamètres de particule sélectionnés (2, 7 et 9 μm).

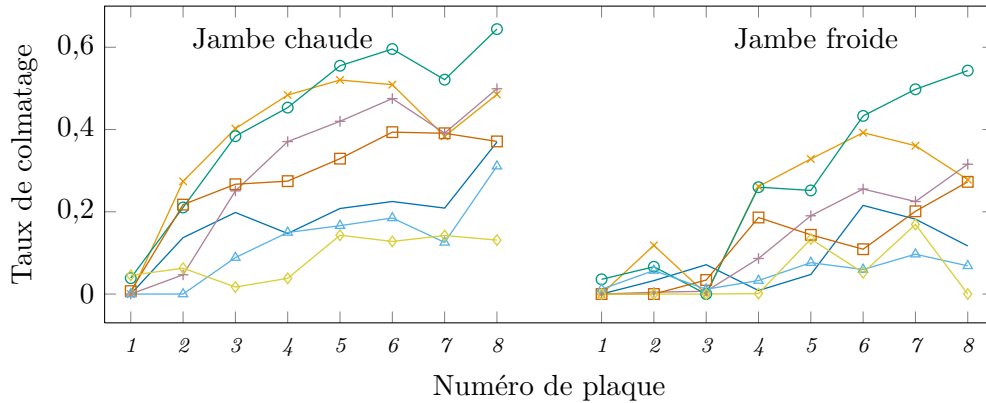


Figure 11.9. : Échantillon aléatoire de 7 profils de colmatage (taux de colmatage uniforme équivalent) obtenus à l'aide de la distribution statistique issue du modèle *vena contracta*.

de ces 4 profils. Un bruit gaussien d'écart type fixé arbitrairement à 0,05 a été ajouté au taux de colmatage de chaque demi-plaque afin de rendre compte sommairement de l'incertitude du modèle.

Le modèle de g. v. a été calé en utilisant des profils dont les taux de colmatage sont compris entre 0 et 0,7. De fait, le modèle de perte de charge singulière des plaques entretoises colmatées décrit dans la section 5.4 n'est véritablement fiable que sur cette plage de taux de colmatage. Tous les profils générés dont certains taux étaient négatifs (à cause du bruit gaussien) ou supérieurs à 0,7 ont été écartés. Ils représentent une proportion d'environ 38 % des couples de répartitions granulométriques et de temps de déposition initialement générés.

L'hypothèse de rétroaction évoquée dans la section précédente peut être intégrée simplement en appliquant la même procédure aux profils obtenus avec rétroaction et en réunissant les deux échantillons obtenus. Un échantillon de 14 233 profils de colmatage dont la moitié ont été obtenus avec rétroaction a été généré. La figure 11.9 représente 7 profils choisis aléatoirement dans cet échantillon.

12. Adaptation au transitoire de référence

Le principe général de la méthode de diagnostic est de comparer les réponses du n. g. l. mesurées lors de transitoires de puissance à des simulations réalisées en faisant varier la configuration de colmatage. Afin de pouvoir comparer plus aisément les résultats de diagnostic, un type de transitoire particulier a été choisi, l'essai R.G.L. 4 décrit dans la section 3.4.1. Cet essai présente de multiples avantages pour l'étude du colmatage. Il est réalisé plusieurs fois par cycle, ce qui permet des diagnostics fréquents. Le transitoire de puissance résultant est rapide et d'amplitude importante, ce qui provoque une réponse dynamique du champ de pression importante au regard du bruit de la mesure de n. g. l. Il est par ailleurs plutôt lisse et peu complexe (presque linéaire) ce qui le rend aisé à modéliser et évite certains problèmes d'ordre numérique causés par des variations trop brutales des conditions aux limites du modèle. Enfin, les conditions dans lesquelles il est réalisé sont soumises à des règles assez strictes (Pigeon, 1997), si bien que les variations imposées aux conditions aux limites des g. v. sont similaires d'un essai à l'autre.

Compte-tenu de ce dernier point, il semble qu'un échantillon unique de simulations pourrait être utilisé pour tous les diagnostics. Cela présente l'avantage de permettre de mutualiser la puissance de calcul, d'obtenir le diagnostic de colmatage immédiatement après l'acquisition des données et de faciliter les comparaisons entre diagnostics. Cela suppose de définir un *transitoire de référence* utilisé pour les simulations qui soit représentatif de l'ensemble des essais R.G.L. 4 exploités. Rappelons cependant que l'objectif de l'essai R.G.L. 4 n'a initialement pas de rapport avec l'étude du colmatage des g. v. : il s'agit d'une manœuvre visant à calibrer le mécanisme de descente des grappes de régulation du cœur des centrales. Aussi, chaque transitoire réel utilisé pour le diagnostic présente des *écarts au transitoire de référence*. Ces écarts n'étaient pas pris en compte dans la méthode de diagnostic à profil choisi : seule la durée des transitoires était considérée pour effectuer l'adaptation des mesures au transitoire de référence. La section 12.1 présente le transitoire de référence choisi et les types de différences que comportent les transitoires réels. Une nouvelle méthode de correction de ces écarts, plus précise que la méthode existante, est proposée dans la section 12.2. Enfin, la section 12.3 présente les résultats d'une analyse de sensibilité réalisée pour choisir judicieusement la grandeur utilisée pour corriger les écarts.

12.1. Transitoire de référence

Dans le modèle thermohydraulique, la puissance transférée par le g. v. est contrôlée par trois conditions aux limites qui varient lors d'un transitoire de puissance : l'enthalpie primaire, l'enthalpie alimentaire et le débit de vapeur sortant du g. v. En régime stationnaire, le débit d'eau alimentaire est égal au débit de vapeur sortant et ces trois grandeurs

	Unité	Valeur initiale	Valeur finale
Enthalpie primaire	kJ kg^{-1}	1463,26	1350,72
∴ Température primaire	$^{\circ}\text{C}$	321,7	302,4
Débit de vapeur	kg s^{-1}	493,85	184,56
Enthalpie Alimentaire	kJ kg^{-1}	941,66	819,34
∴ Température alimentaire	$^{\circ}\text{C}$	219,21	191,97

Tableau 12.1. : Paramètres du transitoire de référence pour les g. v. de type 51 B et apparentés du palier 900 MW. La durée du transitoire est de 1200 s. Le transitoire est simulé par des variations linéaires d'enthalpie mais ce sont des températures qui sont mesurées. Les températures correspondantes sont données à titre indicatif et ont été calculées pour une pression primaire de 155 bar et une pression alimentaire de 66,81 bar

suffisent à calculer la puissance transférée par bilan enthalpique. En régime transitoire, les débits alimentaires et de sortie sont liés par la dynamique du g. v. et contraints par la régulation. La puissance transférée dépend aussi de la dynamique du système. Ainsi, le transitoire de référence est défini par trois fonctions du temps associées à chacune des conditions aux limites. Les variations observées de ces trois grandeurs lors des essais R.G.L. 4 sont approximativement linéaires et sont modélisées par des fonctions affines. Le transitoire de référence est donc déterminé par 7 paramètres : sa durée, ainsi que les valeurs initiales et les pentes des trois fonctions affines. Les valeurs de ces paramètres ont été choisies en calculant les moyennes de données issues d'essais R.G.L. 4 ; elles sont données dans le tableau 12.1.

12.1.1. Types d'écarts au transitoire de référence

Les écarts au transitoire de référence observés sont de plusieurs types : différence de valeur initiale, différence de pente globale, présence de variations non linéaires ou de ruptures de pente et différence de valeur finale. La comparaison des vecteurs de données mesurées et simulées est plus commode lorsqu'ils sont de même dimension. La durée des transitoires étant variable, il faut dans tous les cas procéder à une forme d'adaptation des mesures à la référence choisie. La procédure mise en œuvre dans la méthode de diagnostic à profil choisi dépend uniquement de la durée des transitoires, sans référence aux variations des conditions aux limites. La durée du transitoire mesuré est ajustée à celle du transitoire de référence, puis la réponse du n. g. l. est interpolée régulièrement entre l'instant initial et l'instant final. Ce choix suppose implicitement que la pente des transitoires est constante, égale à la référence.

On peut au contraire décider d'adapter les mesures en tenant compte des écarts au transitoire de référence, ce qui nécessite de choisir une grandeur de référence. Une première approche consisterait à utiliser comme grandeur de référence la valeur de la puissance échangée. Cependant, ce choix comporte deux inconvénients. Le premier est que cette puissance n'est mesurée que de façon indirecte. Par exemple, la puissance électrique à

l'alternateur est une mesure de la puissance de la tranche assez précise mais elle est éloignée du g. v. et peut être impactée par des phénomènes non représentés par le modèle thermohydraulique. De plus, cette mesure est globale pour la tranche et ne représente pas les disparités entre g. v. La température du fluide primaire par exemple, ne varie pas de la même façon dans chaque boucle lors d'un essai R.G.L. 4. Le second inconvénient est qu'un même niveau de puissance peut être atteint par différentes combinaisons de valeurs de ces grandeurs. Les figures 12.1 et 12.2 montrent respectivement les nuages de points des couples de valeurs initiales et de pentes de ces trois grandeurs mesurées lors d'essais R.G.L. 4 des tranches n^{os} 2, 3 et 4 de la centrale Charlie et n^{os} 2 et 4 de la centrale Delta. Ils ne reflètent pas de relation clairement identifiable et les coefficients de corrélation, notés ρ^2 sur chacune des figures, sont faibles. Il y a bien entendu une dépendance entre ces paramètres mais elle semble peu marquée sur la gamme restreinte de régimes de fonctionnement considérée et serait bien délicate à expliciter puisqu'elle implique de nombreux composants de la centrale. Ainsi, on choisira plutôt l'une des trois conditions aux limites utilisées pour simuler les transitoires avec le modèle. Le choix de la grandeur nécessite d'évaluer celle dont les écarts ont le plus d'impact sur le diagnostic. Cette problématique ne sera abordée que dans la section 12.3, à l'aide de l'analyse de sensibilité. Auparavant, on emploiera le terme générique de *grandeur de référence*.

12.2. Méthode de correction des écarts

La correction de la valeur finale de la grandeur de référence peut être effectuée simplement en supprimant la fin de l'une des réponses. Si le transitoire mesuré descend au-delà de la valeur minimale du transitoire de référence, il suffit de tronquer le vecteur des données mesurées. Si le transitoire mesuré descend en-deçà de la valeur minimale du transitoire de référence, c'est le vecteur des données simulées qui sera tronqué.

Une méthode inspirée des techniques d'analyse de données fonctionnelles proposées par Ramsay et Silverman (2005, chap. 7) a été mise au point pour corriger les différences de pente. Le concept de *fonction de déformation temporelle* sur lequel repose cette méthode sera d'abord introduit en l'appliquant à la méthode d'adaptation simple mise en œuvre pour la méthode de diagnostic à profil choisi.

12.2.1. Déformation uniforme de l'intervalle de temps

Les données comparées pour établir le diagnostic du colmatage sont des vecteurs contenant des valeurs de n. g. l. mesurées ou simulées. Ces valeurs correspondent à un échantillonnage à intervalle de temps régulier de fonctions dont l'expression analytique est inconnue. Notons T et T^* les durées respectives du transitoire réel et du transitoire de référence. Soit y^* la fonction du temps à valeur dans $[0, T^*]$ représentant la réponse du n. g. l. simulée avec le transitoire de référence et $\mathbf{t}^* \in [0, T^*]^{p^*}$ le vecteur contenant les p^* instants d'échantillonnage de la fonction simulée. Choisir une période d'échantillonnage constante égale à celle du capteur mesurant le n. g. l., soit 2s, permet de limiter les interpolations à réaliser. On note enfin $\mathbf{y}^* = y^*(\mathbf{t}^*)$ le vecteur de \mathbb{R}^{p^*} contenant les données simulées utilisées pour le diagnostic.

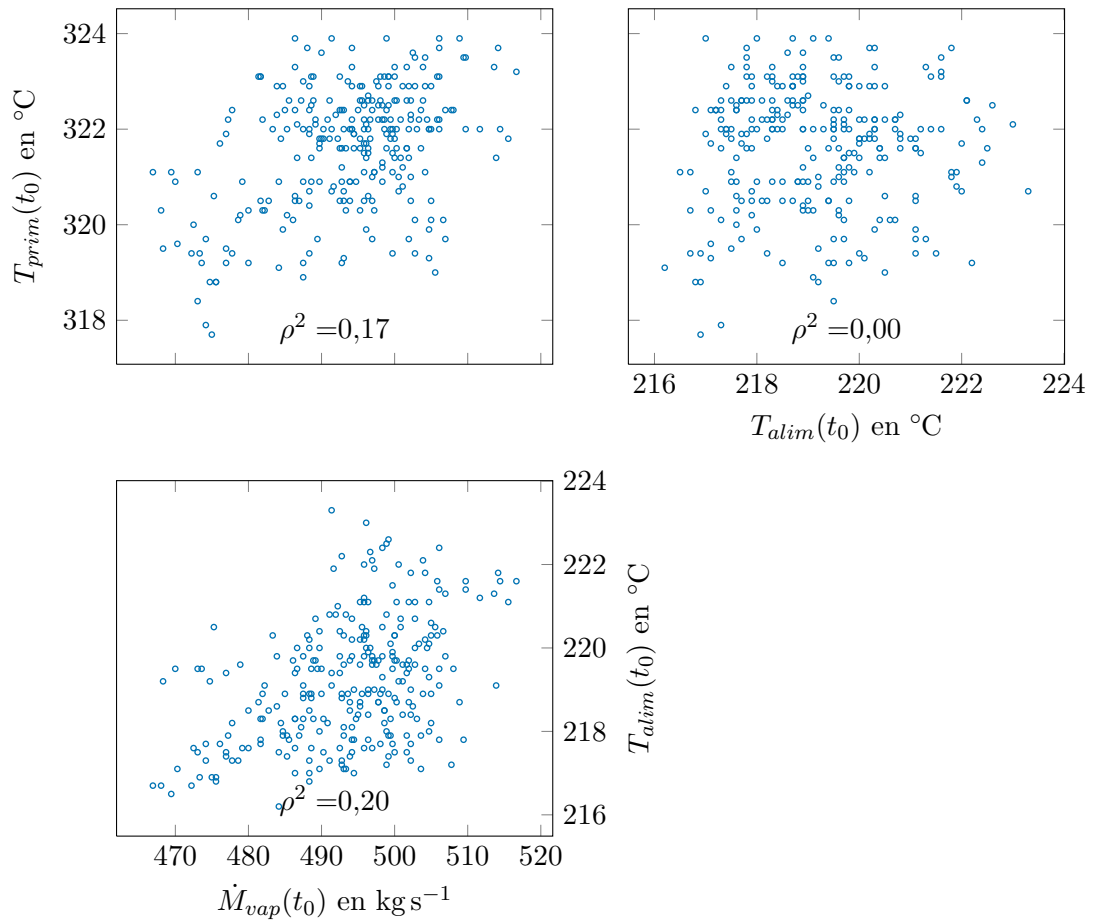


Figure 12.1. : Nuages de points montrant l'absence de corrélation entre les valeurs initiales des trois grandeurs conditions aux limites : la température primaire, $T_{prim}(t_0)$, le débit de vapeur, $\dot{M}_{vap}(t_0)$, et la température alimentaire, $T_{alim}(t_0)$.

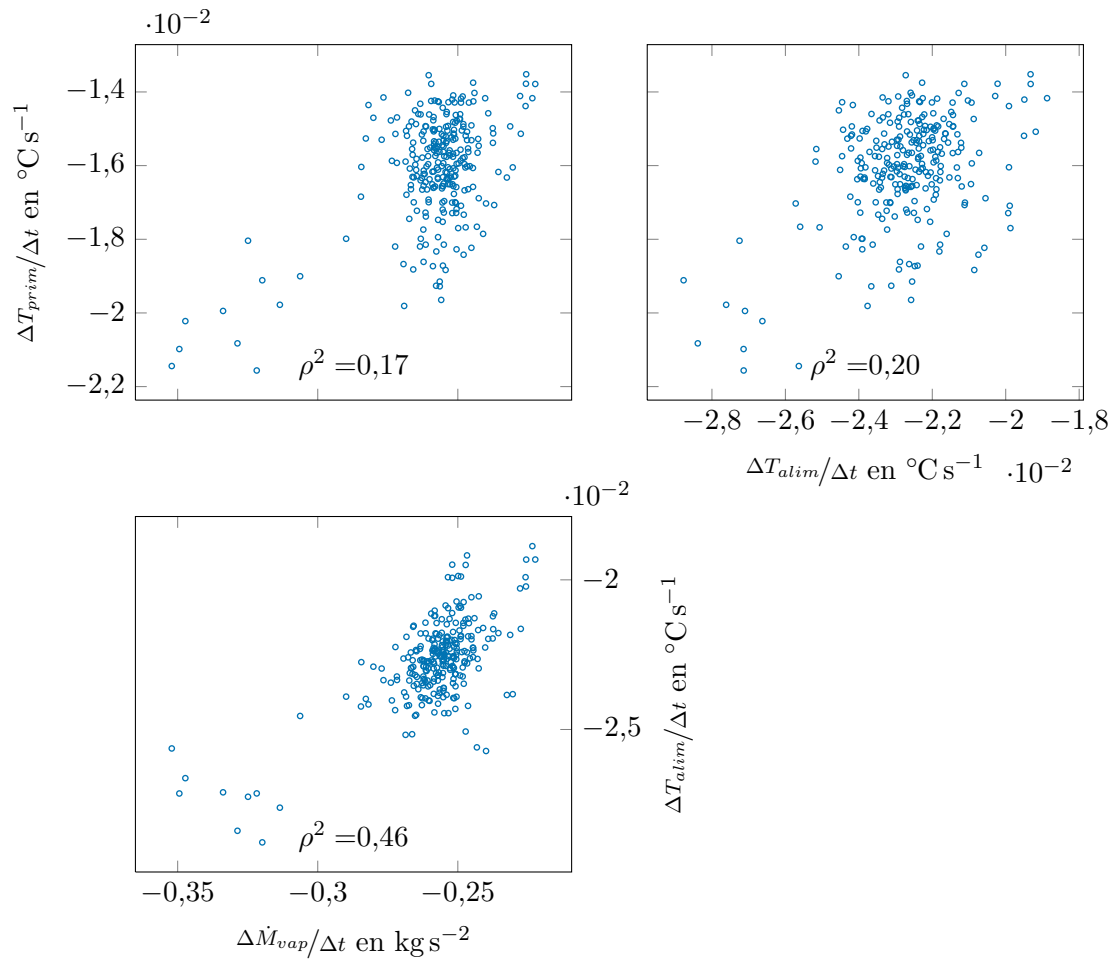


Figure 12.2. : Nuages de points montrant l'absence de corrélation entre les pentes des trois grandeurs conditions aux limites : la température primaire, $\Delta T_{prim}/\Delta t$, le débit de vapeur, $\Delta \dot{M}_{vap}/\Delta t$, et la température alimentaire, $\Delta T_{alim}/\Delta t$.

Soit \mathbf{y} le vecteur de \mathbb{R}^p contenant les données mesurées. Le premier objectif de l'adaptation au transitoire de référence est de faire correspondre la dimension du vecteur de mesures à celle des vecteurs simulés. Autrement dit, on cherche à construire un vecteur $\tilde{\mathbf{y}}$ de \mathbb{R}^{p^*} à partir du vecteur \mathbf{y} de \mathbb{R}^p issu des mesures. En notant y la fonction à valeurs dans $[0, T]$ représentant la réponse du n. g. l. échantillonnée par le capteur de mesure et \mathbf{t} le vecteur de $[0, T]^p$ contenant les instants d'échantillonnage, on obtient l'identité $\mathbf{y} = y(\mathbf{t})$. On cherche à construire une transformation ϕ permettant de passer du « temps simulation » au temps « temps mesuré » :

$$\begin{aligned} \phi : [0, T^*] &\rightarrow [0, T] \\ t &\mapsto \phi(t). \end{aligned} \tag{12.1}$$

On impose aux valeurs initiales et finales de la grandeur de référence du transitoire transformé de coïncider avec celles du transitoire de référence. Par ailleurs, la procédure d'adaptation doit a minima conserver l'ordre de la succession des événements. De cela découlent trois propriétés qui constituent la définition d'une fonction de déformation temporelle :

Définition 2. Une fonction ϕ est appelée fonction de déformation temporelle de l'intervalle de temps $[0, T]$ vers $[0, T^*]$ si elle vérifie

$$\phi(t = 0) = 0, \tag{12.2}$$

$$\phi(t = T^*) = T, \tag{12.3}$$

$$\phi \text{ est strictement croissante.} \tag{12.4}$$

Cette définition ne caractérise pas totalement la transformation à appliquer pour réaliser l'adaptation au transitoire de référence. Pour la méthode de diagnostic à profil choisi, il a été choisi d'imposer qu'un intervalle de temps voie sa durée augmentée ou réduite d'un même facteur quel que soit son instant initial. Cela se traduit par la propriété suivante de la fonction de déformation temporelle ϕ_1 utilisée dans la méthode de diagnostic à profil choisi :

Définition 3 (Condition de déformation uniforme). Une fonction de déformation temporelle ϕ_1 respecte la condition de déformation uniforme si elle est de pente constante.

La fonction ϕ_1 est donc une fonction affine dont on connaît la valeur en deux points. Le vecteur des données adapté, $\tilde{\mathbf{y}}$, vérifie :

$$\tilde{\mathbf{y}} = y(\phi_1(\mathbf{t}^*)). \tag{12.5}$$

Il ne reste qu'à trouver les valeurs prises par y aux instants $\phi_1(\mathbf{t}^*)$. Une façon simple de s'y prendre est d'interpoler linéairement les valeurs du vecteur \mathbf{y} .

12.2.2. Synchronisation d'évènements caractéristiques

La méthode présentée dans la section précédente permet d'obtenir des vecteurs adaptés de même dimension que les vecteurs simulés mais elle ne tient pas compte des écarts au transitoire de référence. Par ailleurs, la propriété 3 est infondée si le transitoire mesuré comporte des ruptures de pente. Une méthode utilisant les mesures de la grandeur de référence pour construire la fonction de déformation temporelle est proposée ci-après. Dans toute cette section, on fait l'hypothèse que la valeur initiale de la grandeur de référence mesurée est celle du transitoire de référence.

Le principe de la nouvelle méthode d'adaptation est inspiré de la méthode de *synchronisation d'évènements caractéristiques* proposée par Ramsay et Silverman (2005, chap. 7). Les évènements peuvent être par exemple l'atteinte d'un extremum, un changement de pente ou encore le passage par une valeur donnée. Ils doivent être présents dans tous les vecteurs de données et facilement repérables. Ces évènements caractéristiques permettent de définir une fonction de déformation temporelle : après déformation à l'aide de la fonction, les évènements du transitoire adapté et du transitoire de référence doivent être simultanés. Cependant, il n'est pas aisé de définir des évènements caractéristiques pour les réponses de n. g. l. Par exemple, l'instant auquel est atteint le maximum de n. g. l. dépend du colmatage. D'autre part, le bruit de la mesure rend délicate l'identification des évènements.

Une solution à ce problème est d'appliquer la synchronisation d'évènements à la grandeur de référence plutôt qu'au n. g. l., en faisant l'hypothèse que la fonction de déformation temporelle ainsi construite pourra être utilisée pour le n. g. l. L'évolution temporelle de la grandeur de référence dépend peu du colmatage et l'on peut choisir cette dernière de façon à disposer d'une mesure comportant peu de bruit. Que l'on choisisse comme grandeur de référence l'enthalpie primaire ou alimentaire, le débit de vapeur ou tout autre image de la puissance échangée, sa variation lors d'un essai R.G.L. 4 sera approximativement linéaire. Aussi, un choix simple d'évènements est le passage par certaines valeurs choisies au préalable. On peut par exemple fixer un nombre d'évènements et les répartir régulièrement sur toute la durée du transitoire de référence.

La construction de la fonction de déformation nécessite de disposer d'une forme fonctionnelle des données qui s'obtient de façon générale par des techniques de lissage ou d'interpolation. Ici, la variation de la grandeur de référence est proche d'une fonction affine par morceaux. On se contentera donc simplement de relier les évènements caractéristiques par des segments de droite. La variation de la grandeur de référence lors de l'essai R.G.L. 4 est monotone hormis de petites fluctuations dont la période est de quelques secondes. Ainsi, la condition de monotonie de la fonction de déformation sera toujours vérifiée si les évènements caractéristiques sont séparés par des intervalles de temps d'une dizaine de secondes ou plus. Un nombre d'évènements compris entre 10 et 100 convient bien pour l'essai R.G.L. 4. L'ajustement de la valeur minimale et l'hypothèse sur la valeur initiale permettent de remplir les deux premières propriétés de la définition des fonctions de déformation temporelle.

La condition de déformation uniforme est quant à elle abandonnée au profit d'une propriété stipulant que les évènements des transitoires mesurés doivent coïncider, après

déformation, avec ceux du transitoire de référence. On définit ainsi une nouvelle contrainte pour la fonction de déformation temporelle, ϕ_2 , utilisée dans la méthode de diagnostic par réduction de dimension :

Définition 4 (Synchronisation d'évènements). Soit ϕ_2 une fonction de déformation temporelle, à laquelle est attachée une famille de K évènements caractéristiques ayant lieu aux instants $(t_i^*)_{1 < i < K}$ et $(t_i)_{1 < i < K}$ respectivement pour la fonction de référence et la fonction mesurée. La fonction ϕ_2 *synchronise* ces évènements si elle vérifie

$$\forall i \in \{1, \dots, K\}, \quad \phi_2(t_i^*) = t_i. \quad (12.6)$$

Les variations de la grandeur de référence entre les évènements ont été approximées par des fonctions affines. Aussi, une fonction de déformation temporelle synchronisant ces évènements est entièrement définie si l'on connaît ses valeurs aux instants des évènements du transitoire de référence. Il suffit en effet de la compléter en reliant les nœuds connus par des segments de droite. La dernière étape est identique à celle de la méthode précédente. Une fois la fonction de déformation temporelle ϕ_2 construite, les données adaptées sont obtenues par interpolation des mesures aux images des instant de mesure.

L'hypothèse selon laquelle la fonction de déformation temporelle construite en utilisant la grandeur de référence peut être appliquée aux données de n. g. l. est implicitement une hypothèse sur la linéarité du lien fonctionnel entre ces deux grandeurs. En effet, si le n. g. l. était une fonction affine de la grandeur de référence, cette méthode d'adaptation fonctionnerait parfaitement. Ce n'est pas le cas, comme en témoigne les exemples de courbes de réponse présentées dans la figure 3.7 page 43. Cependant, puisque l'intervalle de temps est fractionné par les évènements caractéristiques, la méthode fonctionnera aussi si la fonction lien n'est qu'afine par morceaux, les morceaux coïncidant avec les intervalles entre évènements. Les courbes représentées par la figure 3.7 laissent penser que la fonction lien est assez lisse pour pouvoir raisonnablement être approximée par une fonction affine par morceaux. Des tests réalisés sur des exemples artificiels avec la version 2 du modèle ont montré la faisabilité de la méthode (Girard, 2011). Cette méthode d'adaptation a par ailleurs été utilisée avec succès pour appliquer la méthode de diagnostic à profil choisi à des baisses de puissances en fin de cycle, beaucoup plus irrégulières que celles observées lors de l'essai R.G.L. 4 (Ninet et Favennec, 2010; Ninet et coll., 2010). L'analyse de sensibilité présentée dans la section suivante a permis de valider la méthode avec la version 7 du modèle pour l'une des grandeurs de référence envisagées.

12.3. Analyse de sensibilité aux écarts de pente et de valeur initiale

La méthode d'adaptation au transitoire de référence exposée dans la section précédente suppose de choisir une grandeur de référence. Les enthalpies primaire et alimentaire et le débit de vapeur sortant sont les conditions aux limites du modèle utilisées pour simuler un transitoire de puissance. Il est naturel de sélectionner celle de ces trois grandeurs dont l'écart de pente a le plus d'impact sur le diagnostic. Par ailleurs, la méthode d'adaptation

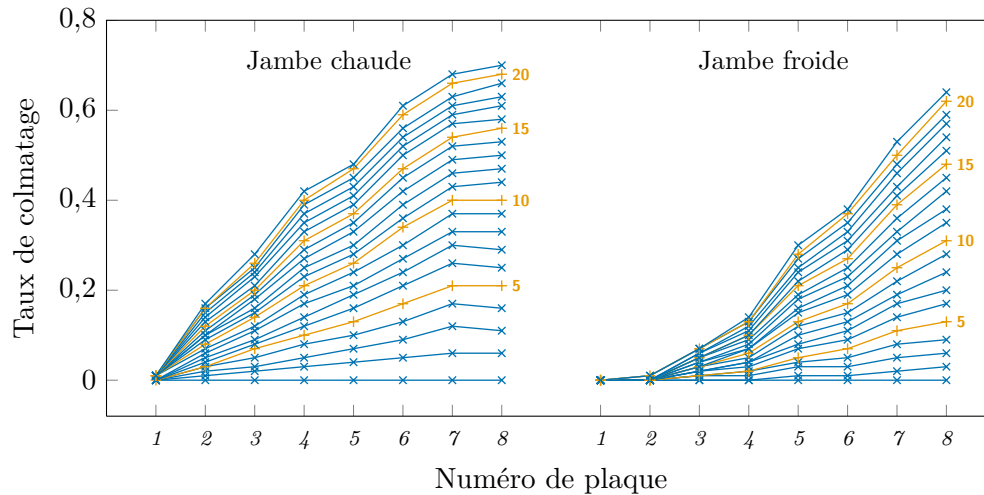


Figure 12.3. : 21 profils de colmatage représentatifs obtenus avec le modèle *vena contracta* pour un diamètre de particule de $2\ \mu\text{m}$.

ne permet pas de corriger les écarts de valeur initiale. Il est donc nécessaire de quantifier l'incertitude causée par ces écarts. Une analyse de sensibilité a été réalisée pour déterminer l'impact des écarts de pente et de valeur initiale de ces trois grandeurs. L'impact d'une valeur différente de débit primaire a été étudié simultanément.

12.3.1. Plan d'expérience

La distribution de configurations de colmatage utilisée est plus simple que celle construite dans le chapitre 11. Les paramètres par défaut du modèle *vena contracta* ont été utilisés avec un diamètre de particules de $2\ \mu\text{m}$ sans réaction pour simuler une famille de 21 configurations de colmatage, représentées par la figure 12.3. Le colmatage a été représenté par une unique variable d'entrée unique distribuée uniformément sur l'intervalle $[0 ; 1]$. À chaque réalisation de la variable aléatoire correspond une configuration de colmatage obtenue par interpolation entre deux configurations de la famille représentée par la figure 12.3. Par exemple, la configuration correspondant à une réalisation de la variable égale à $0,73$ est la moyenne des 15^{ème} et 16^{ème} configurations pondérées par $0,4$ et $0,6$.

Les valeurs initiales et les pentes des trois grandeurs étudiées ont été modélisées par 6 variables aléatoires indépendantes dont les lois ont été ajustées sur des échantillons de mesures. Ces ajustements sont présentés un peu plus loin dans cette section. L'hypothèse simplificatrice d'indépendance des paramètres du transitoire est motivée par l'absence de corrélation illustrée par les nuages de points des figures 12.1 et 12.2. En réalité, ces grandeurs sont bien entendu liées par le mode de fonctionnement de la centrale. Cependant, on ne considère ici qu'une région très restreinte de leur domaine de variation, à savoir celle des conditions dans lesquelles sont réalisés les essais R.G.L. 4. Dans cette région, les liens qui existent entre ces grandeurs apparaissent ténus et l'hypothèse d'indépendance

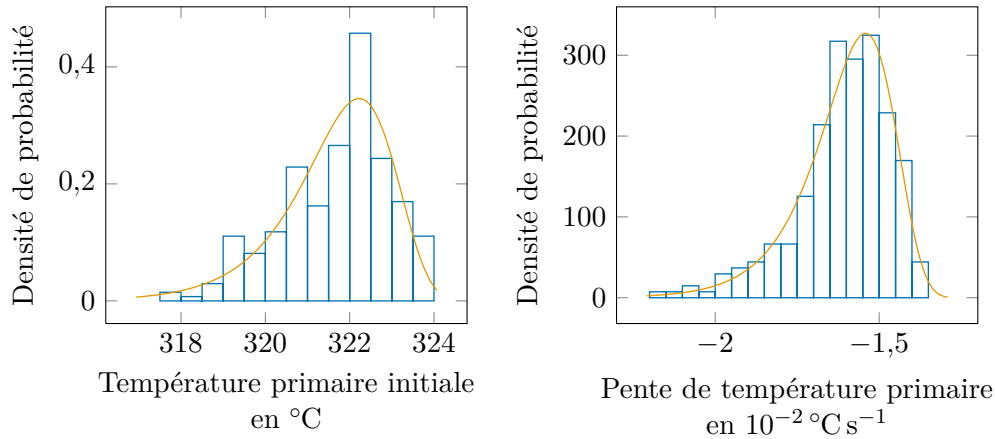


Figure 12.4. : Ajustements des distributions de la valeur initiale (à gauche) et de la pente (à droite) de la température primaire en branche chaude (\square : histogramme des données ; — : densité ajustée).

n'est pas irréaliste.

Les distributions des variables d'entrée ont été ajustées par la méthode du maximum de vraisemblance (Saporta, 2006, chap. 13) à partir de mesures enregistrées lors de 103 essais R.G.L. 4 sur les 3 g. v. des tranches n° 1, 2, et 3 de Charlie et n° 2 et 4 de Delta, entre juillet 2001 et avril 2009. Cette étude de sensibilité a été mise à profit pour vérifier que la valeur du débit primaire n'affecte pas le diagnostic. Le débit primaire n'est mesuré qu'une seule fois par cycle. Des mesures de débit primaire réalisées depuis la mise en service de toutes les tranches équipées de g. v. de type 51 B ont été utilisées pour ajuster la distribution de la variable aléatoire représentant le débit primaire. L'échantillon comprend un total de 628 essais de début de cycle, soit 1884 mesures de débit primaire, les tranches du palier 900 MW comportant trois g. v.

Ajustement des distributions des paramètres de contrôle

Enthalpie primaire L'enthalpie primaire en branche chaude se déduit directement de la température primaire car la pression primaire varie peu autour de sa valeur nominale de 155 bar et n'a que peu d'impact puisque l'eau primaire est liquide et donc très peu compressible. L'histogramme des mesures d'enthalpie primaire initiale est représenté à gauche dans la figure 12.4 et celui des pentes à droite. Les densités ajustées, figurée par des courbes orange, sont des densités de Gumbel de paramètres $\mu_{tp\ max} = 322,222$ et $\sigma_{tp\ max} = 1,063$ pour les valeurs initiales et $\mu_{tp\ pnt} = -1,542 \cdot 10^{-2}$ et $\sigma_{tp\ pnt} = 1,124 \cdot 10^{-3}$ pour les pentes.

Enthalpie alimentaire L'enthalpie alimentaire se déduit elle aussi directement de la température primaire pour la même raison que précédemment : la pression alimentaire varie peu autour de sa valeur nominale de 66,8 bar et n'a que peu d'impact sur l'enthalpie

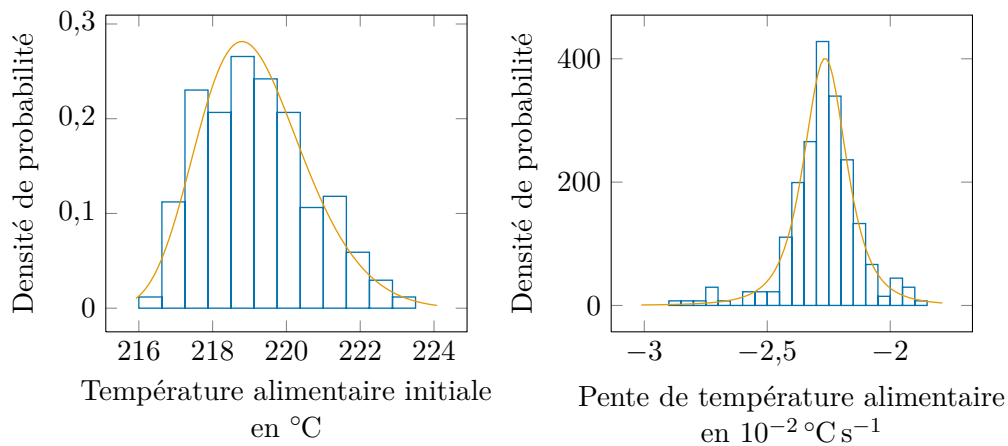


Figure 12.5. : Ajustements des distributions de la valeur initiale (à gauche) et de la pente (à droite) de la température de l'eau alimentaire (: histogramme des données ; — : densité ajustée).

de l'eau alimentaire, elle aussi en phase liquide. L'historgramme des mesures d'enthalpie alimentaire initiale est représenté à gauche dans la figure 12.5 et celui des pentes à droite. L'ajustement des valeurs initiales de température a été réalisé avec une loi des valeurs extrêmes généralisée de paramètres $\mu_{ta\ max} = 218,582$, $\sigma_{ta\ max} = 1,323$ et $\xi_{ta\ max} = -1,513 \cdot 10^{-1}$. L'ajustement des pentes a été réalisé avec une loi de Student non standardisée de paramètres $\mu_{ta\ pnt} = -2,265 \cdot 10^{-2}$, $\sigma_{ta\ pnt} = 9,167 \cdot 10^{-4}$ et $\nu_{ta\ pnt} = 3,003$.

Débit de vapeur L'historgramme des mesures de débit de vapeur initial est représenté à gauche dans la figure 12.6 et celui des pentes à droite. L'ajustement des valeurs initiales de débit vapeur a été réalisé avec une loi de Gumbel de paramètres $\mu_{qv\ max} = 1794,390$ et $\sigma_{qv\ max} = 31,336$. L'ajustement des pentes a été réalisé avec une loi de Student non standardisée de paramètres $\mu_{qv\ pnt} = -9,198 \cdot 10^{-1}$, $\sigma_{qv\ pnt} = 2,411 \cdot 10^{-2}$ et $\nu_{qv\ pnt} = 1,822$.

Débit primaire L'historgramme des mesures de débit primaire est représenté dans la figure 12.7. L'ajustement a été réalisé avec une loi de Student non standardisée de paramètres $\mu_{qprim} = 1,579 \cdot 10^4$, $\sigma_{qprim} = 2,878 \cdot 10^2$ et $\nu_{qprim} = 6,064$.

12.3.2. Indices de sensibilité sans déformation temporelle

Les transitoires d'une durée inférieure à 800 s ont été écartés de l'échantillon car jugés trop éloignés du transitoire de référence. En réalité, de tels transitoires sont rares et l'échantillon filtré comporte encore 930 individus par variables. Une première analyse de

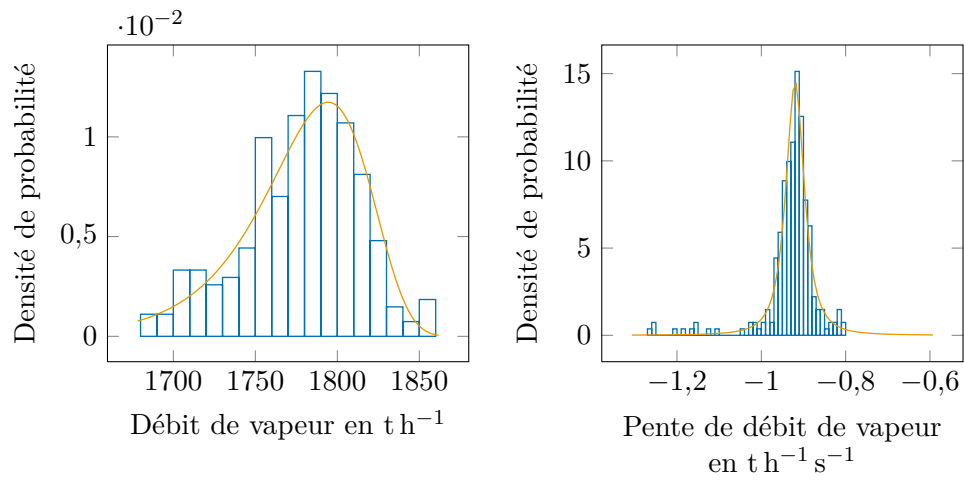


Figure 12.6. : Ajustements des distributions de la valeur initiale (à gauche) et de la pente (à droite) du débit de vapeur sortant (: histogramme des données ; — : densité ajustée).

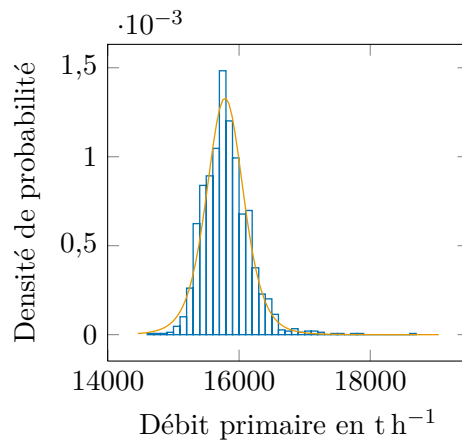


Figure 12.7. : Ajustements des distributions de la valeur initiale (à gauche) et de la pente (à droite) du débit primaire (: histogramme des données ; — : densité ajustée).

sensibilité a été réalisée en tronquant simplement les transitoires de l'échantillon au-delà de 800 s.

La figure 12.8 représente les indices séquentiels d'ordre 1 et totaux correspondants aux six variables d'entrée : colmatage, débit primaire (\dot{M}_{prim}), valeur initiale et pente de l'enthalpie primaire ($h_{prim}(t_0)$ et $\Delta h_{prim}/\Delta t$), valeur initiale et pente de l'enthalpie alimentaire ($h_{alim}(t_0)$ et $\Delta h_{alim}/\Delta t$), valeur initiale et pente du débit de vapeur sortant ($\dot{M}_{vap}(t_0)$ et $\Delta \dot{M}_{vap}/\Delta t$). Comme dans la section 10.2, les barres d'erreur indiquent les bornes des intervalles de confiance *bootstrap*. La courbe bleue (—*) correspond à l'impact du colmatage qui est nettement prédominant. L'enthalpie primaire et le débit de vapeur initiaux ont un impact faible mais non négligeable sur le début et la fin du transitoire. La pente du débit de vapeur a un impact notable sur le milieu du transitoire. Tous les autres paramètres n'ont presque pas d'effet. La comparaison des indices d'ordre 1 et totaux fait apparaître un effet d'interaction entre le colmatage et la pente du débit de vapeur. La courbe d'indices séquentiels du colmatage à la même allure en V que celles des plaques hautes des figures 10.4 à 10.7 débutant page 148. On pouvait s'attendre à ce résultat car ces plaques sont celles qui ont le plus d'impact. La courbe d'indices séquentiels associée à la pente du débit de vapeur a une allure comparable à celle des plaques proches de la zone de début d'ébullition présentant un pic au milieu de l'intervalle de temps.

Les figures 12.9 et 12.10 représentent les indices compacts associés aux deux premières composantes principales. Dans tout le reste de ce chapitre, les composantes principales dont il est question sont celles calculées avec la matrice de corrélation. Les mêmes conclusions seraient tirées de l'observation des scores des composantes principales brutes, calculées avec la matrice de covariance. L'impact du colmatage est prédominant sur les scores de la première composante principale. Ceux-ci sont aussi affectés par le débit de vapeur initial et, en moindre mesure, par l'enthalpie primaire initiale. Les scores de la seconde composante principale sont moins dominés par le colmatage et sont assez fortement affectés par la pente de débit vapeur. Seuls la valeur initiale et la pente de l'enthalpie alimentaire ont une influence négligeable. En l'absence de pré-traitement pour compenser l'effet de la pente du débit de vapeur, les scores de la seconde composante principale sont contrôlés pour moitié par des paramètres autres que le colmatage.

On observe, comme dans la section 10.2, que la seconde composante principale est liée au centre du transitoire. Les indices de la figure 12.9 sont dominés par le colmatage et les interactions identifiées à partir des indices séquentiels se retrouvent dans la figure 12.10.

12.3.3. Indices de sensibilité avec déformation temporelle

L'échantillon de courbes de réponse utilisé dans la section précédente a été adapté successivement aux transitoires de référence en débit de vapeur, enthalpie primaire et enthalpie alimentaire.

Déformation selon la pente de débit de vapeur

Parmi les pentes des paramètres de contrôle, celle du débit de vapeur a le plus d'impact. La figure 12.11 représente les indices séquentiels d'ordre 1 et totaux obtenus à partir

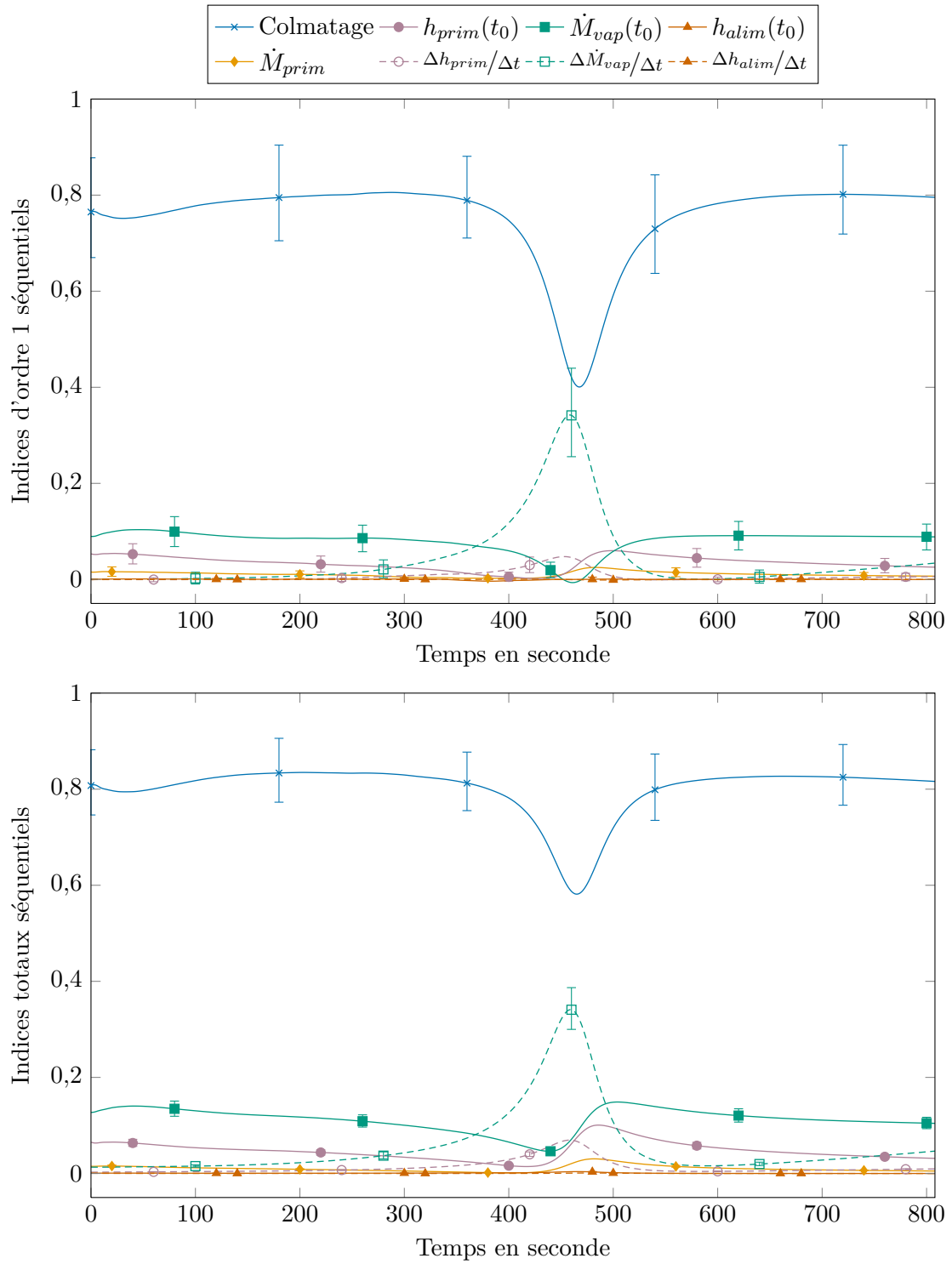


Figure 12.8. : Indices séquentiels d'ordre 1 (en haut) et totaux (en bas) sans adaptation au transitoire de référence (les courbes de réponse sont simplement tronquées au-delà de 800s).

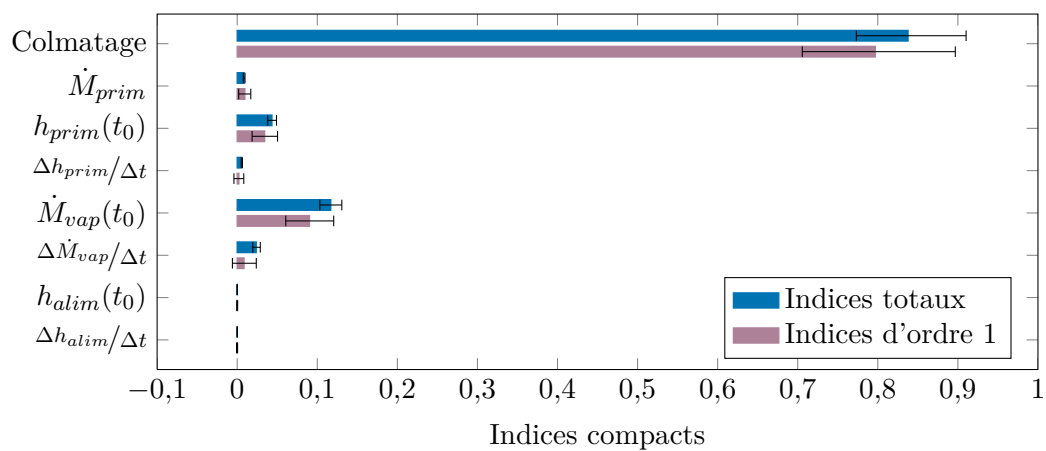


Figure 12.9. : Indices compacts associés à la première composante principale sans adaptation au transitoire de référence.

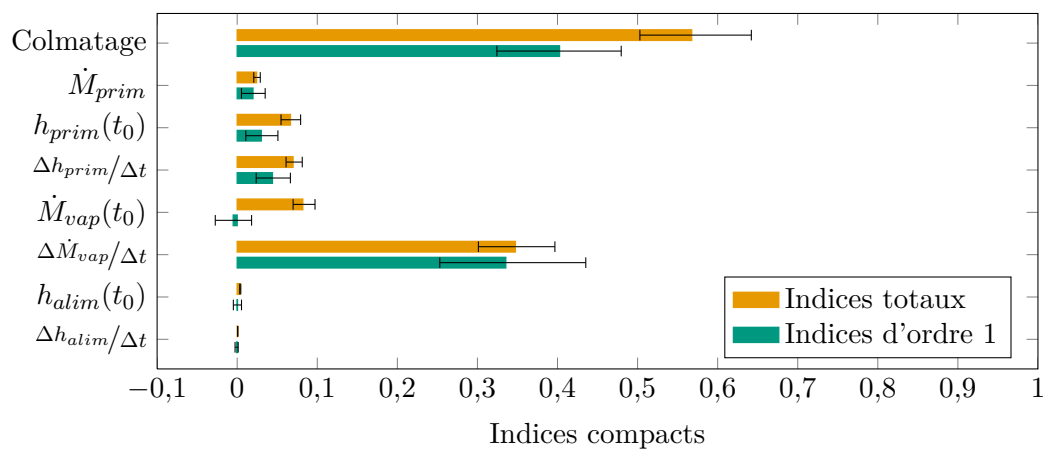



Figure 12.10. : Indices compacts associés à la seconde composante principale sans adaptation au transitoire de référence.

des courbes de réponse d'une durée supérieure à 800 s après adaptation au transitoire de débit de vapeur de référence. Le pic central des indices séquentiels associés à la pente de débit de vapeur est supprimé, ce qui confirme la validité de la méthode de correction. En revanche, l'impact de la valeur initiale du débit de vapeur est augmenté, notamment par une interaction avec le colmatage repérable à la bosse de la courbe verte () du graphique du bas de la figure 12.11.

Les figures 12.12 et 12.13 représentent les indices compacts obtenus avec les courbes de réponse adaptées au transitoire de référence en débit de vapeur. Les indices associés à la première composante principale ne sont pas très différents de ceux de la figure 12.9 obtenus après simple exclusion de la fin des transitoires. La figure 12.13 montre la disparition presque totale de l'effet de la pente du débit de vapeur sur les scores de la deuxième composante principale ainsi que l'apparition d'interaction avec la valeur initiale du débit de vapeur.

L'adaptation des courbes de réponse selon la pente du débit de vapeur renforce légèrement la corrélation entre le colmatage et les scores de la première composante principale en diminuant les interactions. Elle permet par ailleurs de compenser les ruptures de pentes qui n'ont pas été prises en compte lors de cette analyse de sensibilité, ainsi que d'éventuelles erreur de détermination des instant de début et de fin des transitoires.

Déformation selon la pente d'enthalpie primaire ou alimentaire

L'analyse de sensibilité sans déformation temporelle a révélé un impact faible de la pente d'enthalpie primaire. Les figures 12.14 et 12.15 représentent les indices compacts obtenus avec les courbes de réponse adaptées au transitoire de référence en enthalpie primaire. On observe seulement une légère augmentation des interactions pour les scores de la première composante principale. En revanche, la pente du débit d'enthalpie primaire prend une importance prépondérante pour les scores de la seconde composante principale. Autrement dit, lorsque la grandeur de référence choisie n'a que peu d'impact, l'effet de la méthode de correction est plutôt néfaste, ce qui explique l'effet mentionné au début de ce paragraphe. Lors de l'application, avec la version 7 du modèle, de la méthode de diagnostic à profil choisi à l'ensemble des réponses de n. g. l. mesurées disponibles (Ninet et coll., 2012), la puissance électrique a été préférée à la température primaire comme grandeur de référence car cette dernière donnait des résultats plus erratiques.

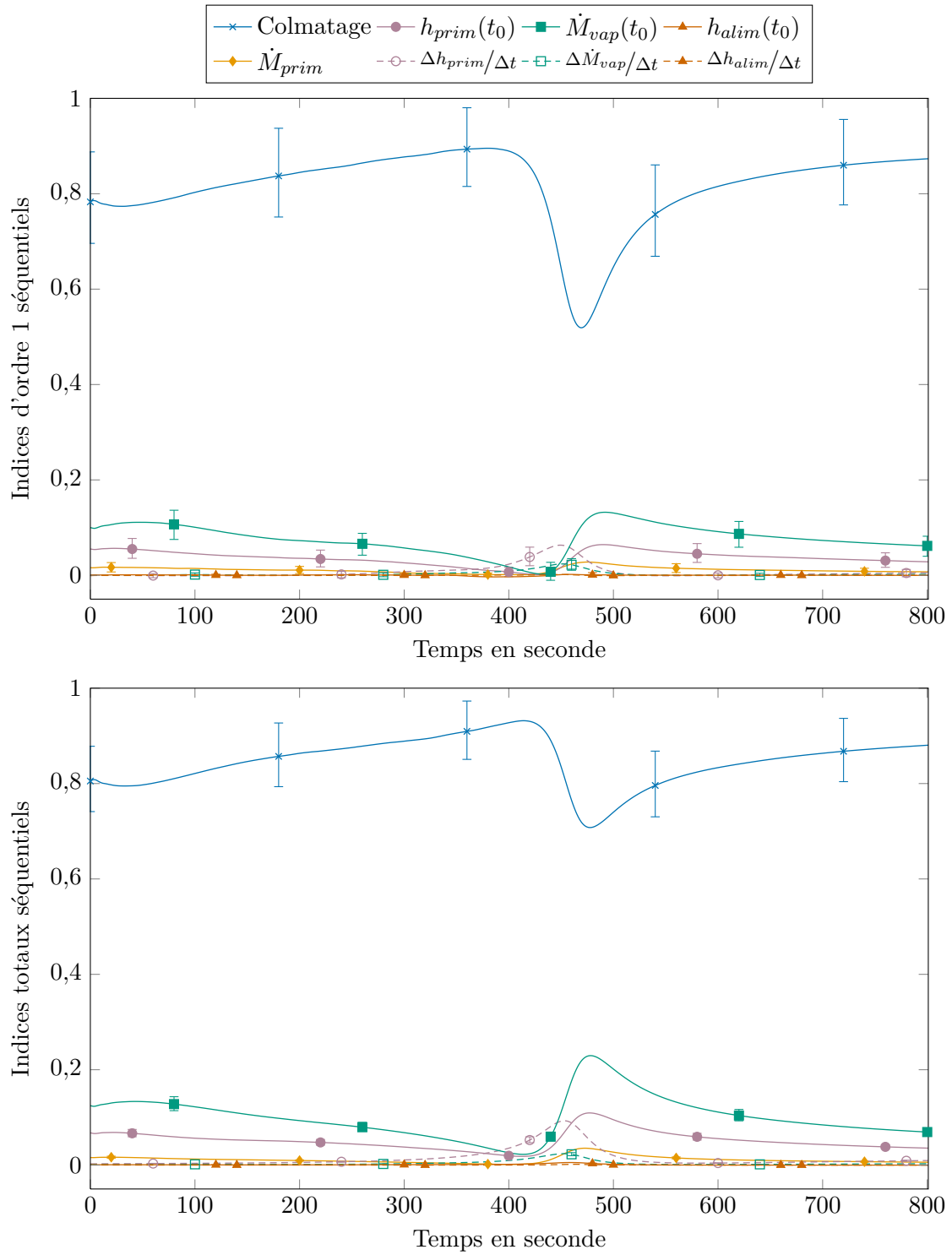


Figure 12.11. : Indices séquentiels d'ordre 1 (en haut) et totaux (en bas) après adaptation au transitoire de référence en débit vapeur.

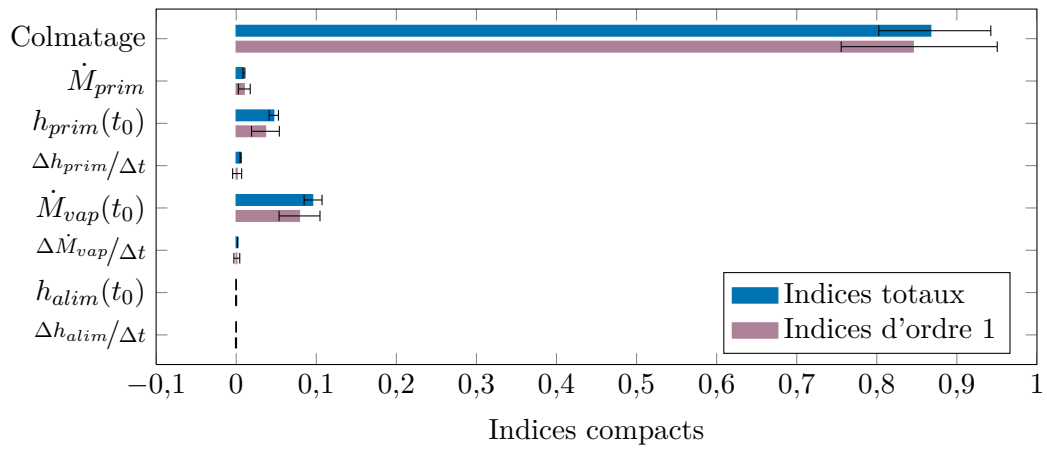


Figure 12.12. : Indices compacts associés à la première composante principale après adaptation au transitoire de référence en débit vapeur.

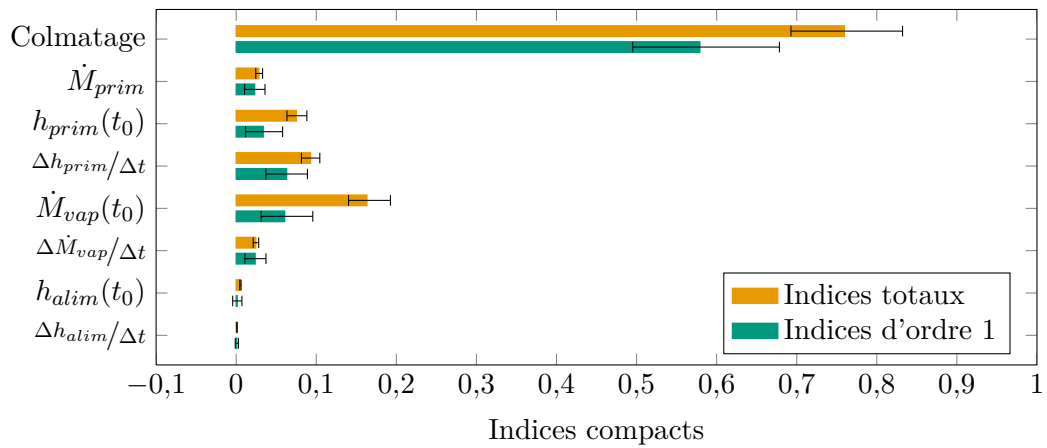


Figure 12.13. : Indices compacts associés à la seconde composante principale après adaptation au transitoire de référence en débit vapeur.

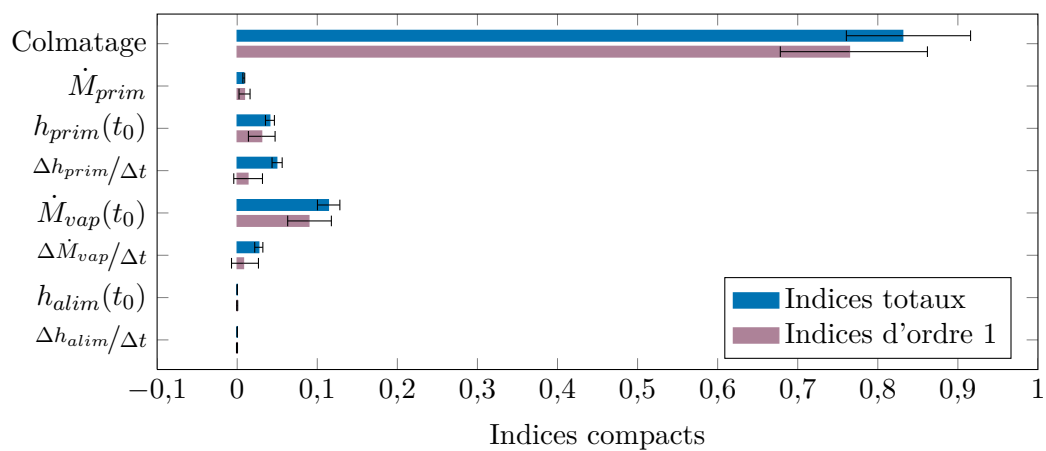


Figure 12.14. : Indices compacts associés à la première composante principale après adaptation au transitoire de référence en enthalpie primaire.

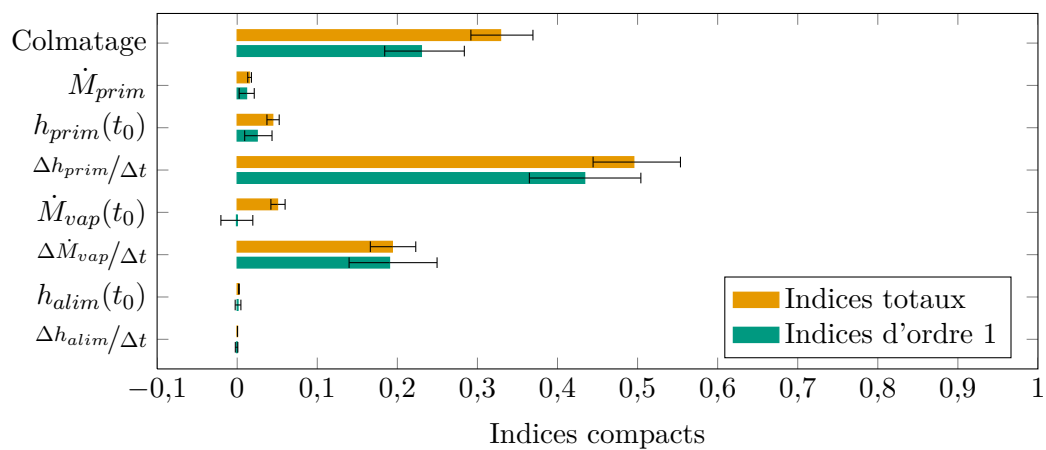


Figure 12.15. : Indices compacts associés à la seconde composante principale après adaptation au transitoire de référence en enthalpie primaire.

13. Application pratique de la nouvelle méthode de diagnostic

Une nouvelle méthode d'analyse des courbes de réponse dynamique du n. g. l. pour diagnostiquer le colmatage a été esquissée à la fin du chapitre 10. L'analyse en composantes principales est utilisée pour rapporter les courbes de réponse à une valeur unique : leur score selon la première composante principale, noté s_1 . Les prétraitements à appliquer pour corriger les écarts du transitoire réel au transitoire choisi comme référence ont été exposés dans le chapitre 12. Une fois ceux-ci appliqués, les scores peuvent être calculés par projection sur la première direction principale. Le principe de la méthode est ensuite d'utiliser un échantillon de simulations pour prédire, à partir des scores s_1 , une moyenne pondérée des taux de colmatage des 16 demi-plaques. Les poids de cette moyenne sont obtenus à l'aide de la SIR : ce sont les composantes de la première direction e. d. r. de vecteur directeur $\beta_{1,1}$.

La distribution de l'échantillon de courbes simulées intervient donc à deux niveaux. Elle influe d'abord sur la direction e. d. r. ; on a vu par exemple que l'application de la SIR a un échantillon gaussien ou uniforme ne donnait pas exactement les mêmes résultats. Néanmoins, cette influence est mineure pour certains effets globaux comme l'action sur la pente des courbes de réponse qui dépend peu de la localisation du colmatage. Le terme *global* s'entend ici dans le sens d'un effet s'appliquant sur l'ensemble de l'espace des taux de colmatage, par opposition à un effet *local* qui n'est observable que dans une restriction de cet espace ou sur un sous-espace. Elle conditionne ensuite le diagnostic : la relation entre le score s_1 et la projection e. d. r. $\beta'_{1,1}\mathbf{x}$ est déduite du nuage de points obtenu à partir des réponses simulées.

Il a été choisi d'utiliser le modèle de déposition par *vena contracta* pour construire une distribution représentative de profils de colmatage. Cette construction a été présentée dans le chapitre 11. Les calculs exposés dans le chapitre 10 pour l'échantillon gaussien, et dans l'annexe C page 236 pour l'échantillon uniforme, ont été repris avec un échantillon de 14 233 individus issus de cette nouvelle distribution. La démarche étant identique à celle mise en œuvre dans le chapitre 10, seuls les grandes lignes sont exposées ci-après, dans la section 13.1. Les graphiques représentant les résultats de chaque étape sont donnés dans l'annexe E page 243. La section 13.2 synthétise le fonctionnement de la méthode de diagnostic proposée et présente une échelle de gravité du colmatage en fonction du nouvel indicateur. L'application de cette méthode à des tranches du parc EDF à partir de données collectées pendant les dix dernières années est enfin exposée dans la section 13.3.

13.1. Réduction de la dimension d'entrée avec l'échantillon *vena contracta*

La SIR monodimensionnelle a d'abord été appliquée en choisissant successivement les scores des deux premières composantes principales comme variables de sortie. Dans les deux cas, la dimension e. d. r. estimée à l'aide de la méthode de *bootstrap* vaut 4 (figures E.3 et E.4 de l'annexe E). Ces dimensions plus grandes que dans les cas gaussien et uniforme s'expliquent par le fait que la région de l'espace de colmatage considérée ici est nettement plus restreinte. En effet, malgré le bruit ajouté au taux de colmatage de chaque demi-plaque, les profils obtenus à l'issue du chapitre 11 sont assez peu diversifiés, comme l'illustre la figure 11.9 page 181. Ainsi, un effet local supplémentaire est observé ici tandis qu'il était filtré avec les échantillons gaussiens et uniformes. De façon imagée, les petites aspérités de la surface de réponse ne sont visibles qu'en l'observant de près.

La direction $\beta_{1,1}$ est identique à celles obtenues avec les échantillons gaussien et uniforme. Cela confirme l'hypothèse formulée au chapitre 10 selon laquelle l'effet correspondant est global. Les autres directions, y compris $\beta_{2,1}$, sont différentes (figure E.3 et E.4 de l'annexe E). On peut en conclure que les effets associés sont locaux, ce qui en limite l'intérêt pour le diagnostic. En effet, le diagnostic sera plus robuste si l'on exploite un effet global car ainsi l'hypothèse sur la distribution des profils de colmatage n'intervient que pour le conditionnement final.

L'espace e. d. r. obtenu par découpage marginal groupé est un plan (figure E.7 l'annexe E), comme pour les échantillons gaussien et uniforme. La médiane de l'estimateur *bootstrap* \hat{R}_j chute de 1 pour $j = 1$ à 0,54 pour $j = 2$ tandis que la valeur pour $j = 3$ est entre les deux, égale à 0,74. En revanche, les variances pour $j = 1$ et $j = 2$ sont presque nulles ce qui signifie que l'espace e. d. r. est bien un plan, du type illustré par la figure 10.25 du chapitre 10. Les deux directions e. d. r. obtenues par découpage marginal groupé (figure E.8 de l'annexe E) sont similaires à $\beta_{1,1}$ et $\beta_{2,1}$ mais intègrent certains éléments des directions mineures de rang supérieur.

L'observation des nuages de points (figures E.5 et E.6) invite à nouveau à ne retenir que $\beta_{1,1}$ pour le diagnostic. Le nuage de $\beta'_{1,1}\mathbf{x}$ en fonction de s_1 est très rectiligne. Il est par ailleurs plus étroit que dans les cas gaussien et uniforme. Cela traduit la petitesse de la région couverte par le nouvel échantillon. Le nuage obtenu après découpage marginal groupé est similaire, quoiqu'un peu plus large. Les coefficients de corrélation linéaire correspondant à ces deux nuages valent respectivement 0,998 et 0,994. C'est la direction de projection obtenue avec la SIR monodimensionnelle qui a été retenue pour le diagnostic car les effets additionnels pris en compte par celle issue du découpage marginal groupé sont locaux et n'apportent pas véritablement d'information supplémentaire.

La direction $\beta_{1,1}$ utilisée pour le diagnostic est représenté par la figure 13.1. Les poids négatifs des demi-plaques n^{os} 1 à 3 en jambe chaude signifie que leur effet est opposé à celui des autres demi-plaques. Ils sont cependant faibles en valeur absolue et il faudrait un colmatage exceptionnellement fort de ces 3 demi-plaques pour fausser notablement le diagnostic. Ce défaut mineur de l'approche par analyse de la réponse du n. g. l. était déjà présent, de façon implicite, dans la méthode de diagnostic à profil choisi. Par ailleurs, la

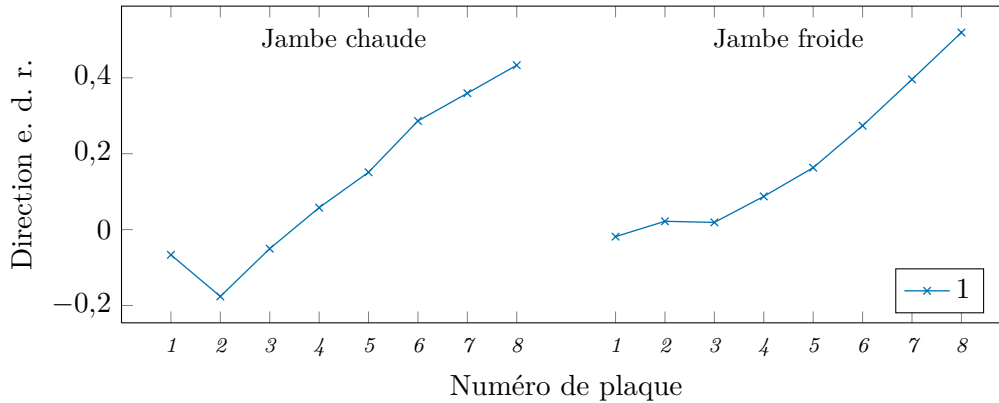


Figure 13.1. : Direction de projection des profils de colmatage utilisée pour le diagnostic.

distribution choisie pour conditionner le diagnostic produit des profils dont les taux de colmatage en bas de g. v. sont faibles, aussi cet effet peut raisonnablement être négligé.

13.2. Définition du nouvel indicateur de colmatage

La figure 13.2 représente le nuage de points utilisé pour le diagnostic. Elle illustre la forte corrélation entre les scores s_1 et les projections $\beta_{1,1}$. Les flèches noires résument la procédure mise en œuvre pour produire un diagnostic :

1. la valeur \hat{s}_1 du score de la courbe de réponse de n. g. l. mesurée est obtenue par projection, après prétraitement, sur la première direction principale ;
2. un intervalle autour de la valeur estimée est choisi, en fonction de la confiance que l'on a dans l'estimation (il est indiqué par les crochets sur l'axe des abscisses dans la figure 13.2) ;
3. les points de simulations dont le score est compris dans cet intervalle sont sélectionnés (en orange sur la figure 13.2) ;
4. L'histogramme des projections $\beta'_{1,1} \mathbf{x}$ des points sélectionnés constitue le diagnostic ; on peut en calculer la valeur moyenne et déduire un intervalle de confiance reflétant l'incertitude due à l'estimation du score de la courbe mesurée, conditionnellement aux hypothèses de modélisation et à la distribution de profils de colmatage choisie.

Une analyse d'incertitude soignée permettrait de déterminer précisément la largeur à attribuer à l'intervalle de sélection, autour de l'estimation \hat{s}_1 . On se contente ici d'une majoration raisonnable établie à partir du modèle suivant :

$$\hat{s}_1(t) = s_1(t) + \epsilon_1(t), \quad (13.1)$$

où $\hat{s}_1(t)$ désigne l'estimation du score à une date t , $s_1(t)$ la valeur vraie du score à cette même date et $\epsilon_1(t)$ un terme d'erreur que l'on suppose indépendant de la valeur vraie.

La différence entre deux estimations successives aux dates t et $t + 1$ s'écrit

$$\begin{aligned}\Delta\widehat{s}_1(t) &= \widehat{s}_1(t+1) - \widehat{s}_1(t) \\ &= s_1(t+1) - s_1(t) + \epsilon_1(t+1) - \epsilon_1(t),\end{aligned}\tag{13.2}$$

En supposant que les termes d'erreur sont indépendants et de même variance σ_ϵ^2 et en notant $\Delta s_1(t)$ la différence entre les valeurs vraies successives, on obtient l'identité suivante :

$$\text{var}[\Delta\widehat{s}_1(t)] = \text{var}[\Delta s_1(t)] + 2\sigma_\epsilon^2,\tag{13.3}$$

d'où l'on déduit la majoration recherchée :

$$\sigma_\epsilon < \sqrt{\frac{1}{2}\text{var}[\Delta\widehat{s}_1(t)]}.\tag{13.4}$$

L'histogramme des différences de scores estimés successivement, sans arrêt de tranche où autre élément perturbateur entre les deux essais R.G.L. 4, est d'allure gaussienne. Aussi, il a été choisi pour le diagnostic de sélectionner les points situés dans l'intervalle

$$\left[\widehat{s}_1 - 1,96\frac{\sigma_{\Delta\widehat{s}}}{\sqrt{2}}, \widehat{s}_1 + 1,96\frac{\sigma_{\Delta\widehat{s}}}{\sqrt{2}}\right],\tag{13.5}$$

où $\sigma_{\Delta\widehat{s}}$ désigne l'écart type des différences de scores estimés successivement, ce qui revient à supposer qu'elles sont distribuées selon une loi gaussienne et à choisir un risque de 0,05.

La linéarité de la fonction liant les scores à la projection e. d. r. contribue à la robustesse du diagnostic. En effet, comme la pente de cette fonction varie très peu, un biais dans l'estimation de s_1 n'aura pas un impact trop important. Pour la même raison, le choix de la valeur du coefficient de normalisation des données mesurées, évoqué dans la section 6.3, aura peut d'impact sur le diagnostic de l'évolution du colmatage des g. v.

La gamme de variation des projections $\beta_{1,1}$ dépend du choix de la normalisation. Le choix de la valeur 4 plutôt que 1 pour l'état le plus colmaté parmi les points de simulation vise à refréner la tentation d'assimiler cet indicateur à un taux de colmatage. La section 3.1.2 du chapitre 3 introduisant la notion de taux de colmatage uniforme équivalent a mis en évidence que la moyenne des taux de colmatage des trous foliés d'une plaque pouvait ne pas être représentative de la moyenne des *effets* du colmatage de ces trous. C'est d'autant plus vrai pour la moyenne des taux de colmatage de plaques distinctes qui sont confrontées à des conditions thermohydrauliques pouvant être très différentes : l'écoulement au travers des plaques basses, par exemple, est monophasique. L'indicateur proposé s'entend, de façon analogue au taux de colmatage uniforme équivalent, comme une moyenne pondérée des taux de colmatage des demi-plaques dont les poids reflètent leurs importances respectives sur le comportement thermohydraulique du g. v. Cet indicateur est pertinent car souvent, c'est précisément les effets du colmatage sur la thermohydraulique que l'on cherche à estimer et à prévenir. Il n'est cependant pas équivalent à un taux de colmatage.

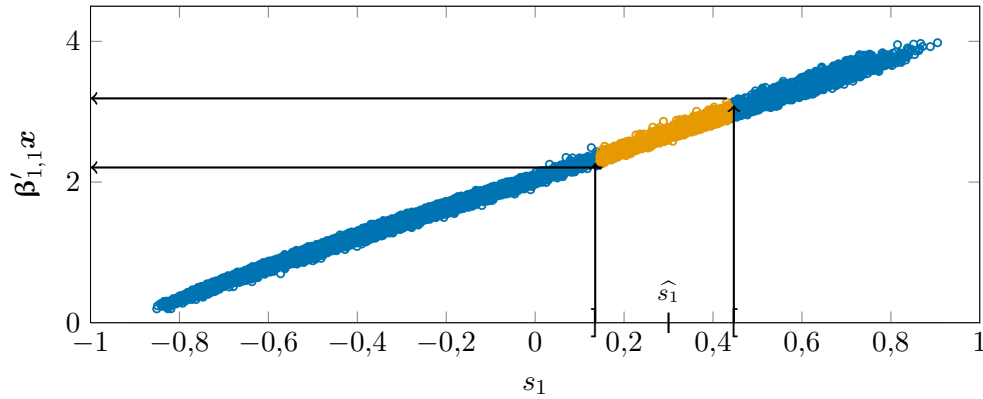


Figure 13.2. : Nuage des coordonnées selon $\beta_{1,1}$ des simulations utilisées pour le diagnostic, en fonction de leurs scores selon la première composante principale. Les flèches noires indiquent le principe de la méthode : un intervalle est délimité autour de l'estimation \hat{s}_1 du score d'une réponse mesurée. Les points simulés à l'intérieur de cet intervalle sont sélectionnés et la distribution de ce sous-échantillon constitue le diagnostic.

L'interprétation de l'échelle de valeurs nouvellement constituée peut se faire à la lumière de l'historique conséquent de mesures disponibles comportant des états de colmatage diversifiés. Bien que ceux-ci ne soient jamais pleinement identifiés, on peut ainsi attribuer à chaque unité de l'échelle proposée un degré de gravité de l'état de colmatage. Les diagnostics obtenus pour les tranches étudiées au cours de cette thèse sont représentés par les figures 13.3 à 13.13, débutant page 210. Ils seront décrits individuellement dans la section 13.3. Cet ensemble de diagnostics a été utilisé pour établir la signification du nouvel indicateur. Le tableau 13.1 donne une description qualitative des états correspondant aux degrés de l'échelle, assortie d'ordre de grandeur du colmatage et du taux de colmatage uniforme équivalent en plaque haute. Ces valeurs sont des ordres de grandeur donnés à titre indicatif. Elles supposent un colmatage faible des plaques inférieures.

13.3. Application de la nouvelle méthode de diagnostic

La méthode de diagnostic dont les détails de mise en œuvre viennent d'être exposés a été appliquée à l'ensemble des mesures de n. g. l. lors d'essais R.G.L. 4 disponibles pour les tranches n^{os} 2, 3 et 4 de la centrale Alpha, n^o 1 de la centrale Bravo, n^{os} 2, 3 et 4 de la centrale Charlie et n^{os} 1, 2, 3 et 4 de la centrale Delta. Les données traitées couvrent la période de 2000 à 2012. Les diagnostics sont représentés par les figures 13.3 à 13.13, chaque figure comprenant les trois g. v. d'une tranche. Les valeurs moyennes de l'indicateur de colmatage pour chaque sous-échantillon correspondant à un essai R.G.L. 4 sont représentées par des cercles bleus (\circ) dont le rayon est proportionnel au nombre d'individus. Les interruptions des traits bleus reliant les cercles correspondent aux arrêts de tranche entre les cycles de fonctionnement. Les traits noirs de part et d'autre des

$\beta'_{1,1}x$	Degré	Description
$1 <$	Colmatage faible	G. v. neuf ou venant d'être nettoyé après avoir atteint un degré de colmatage fort. Les taux de colmatage en plaque supérieure sont inférieurs à 0,2.
$[1 ; 2[$	Colmatage modéré	État actuel de la plupart des g. v. des tranches conditionnées à haut pH ; cet état est aussi celui, après nettoyage, des tranches ayant atteint un degré de colmatage critique. Les taux de colmatage sont de l'ordre de 0,3.
$[2 ; 3[$	Colmatage moyen	État actuel des g. v. des tranches conditionnées à haut pH les plus colmatées ; cet état est celui qu'avaient la plupart des tranches conditionnées à bas pH au début de la période considérée, en 2002. Les taux de colmatage en plaque supérieure sont de l'ordre de 0,45.
$[3 ; 4[$	Colmatage fort	État atteint par la plupart des tranches conditionnées à bas pH avant leur nettoyage en 2007 et 2008. Les taux de colmatage sont de l'ordre de 0,6.
≥ 4	Colmatage critique	Risque tangible de fissuration des tubes. Le taux moyen en plaque haute excède 0,7 et le taux de colmatage uniforme équivalent (calculé en extrapolation) sature à une valeur proche de 1.

Tableau 13.1. : Échelle qualitative de gravité du colmatage en fonction de la valeur du nouvel indicateur

points de diagnostics sont les quantiles 0,05 et 0,95 des sous-échantillons. Les diagnostics réalisés avec la nouvelle méthode sont commentés dans la suite de cette section. Leur interprétation est confrontée à celle que l'on peut faire des taux de colmatage moyens et des taux de colmatage uniformes équivalents calculés à partir des résultats d'examens télévisuels disponibles pour ces tranches. Ces valeurs sont répertoriées dans le tableau 13.2

Les valeurs de l'indicateur pour les trois tranches de la centrale Alpha étudiées sont actuellement toutes proches de 2, ce qui correspond à la transition entre les états de colmatage modéré et moyen sur l'échelle donnée dans le tableau 13.1. Les deux examens télévisuels réalisés en 2007 sur ces tranches avaient conclu à des taux de colmatage faibles en accord avec ce diagnostic. Ces g. v. ont des cinétiques de colmatage plutôt lentes, environ une demie unité du nouvel indicateur en 10 ans, ce que l'on peut attribuer à leur conditionnement à haut pH.

Quoique plus récents, les g. v. de la tranche n° 1 de la centrale Bravo sont dans un état de colmatage sensiblement plus avancé, correspondant à la partie haute du degré moyen de l'échelle proposée. Ce diagnostic est cohérent avec l'examen télévisuel réalisé en 2011 si l'on prend en compte l'hétérogénéité du colmatage. Ce cas a été discuté en détail dans la section 3.1.2. La cinétique de colmatage de cette tranche est par ailleurs légèrement plus rapide que celle des tranches de la centrale Alpha. Elle reste modérée, environ une unité en 10 ans, ce qui est en accord avec le conditionnement à haut pH de cette tranche.

Les g. v. de la tranche n° 2 de Charlie sont passés de 2002 à 2007 d'un colmatage moyen à un colmatage fort. Cette cinétique rapide s'explique par le conditionnement à bas pH des tranches n°s 2, 3 et 4 de Charlie. Le nettoyage chimique de la tranche n° 2 en 2007 a permis de retrouver un état de colmatage faible. La reprise du colmatage s'est faite avec une cinétique comparable à la cinétique avant nettoyage, aboutissant aujourd'hui à un colmatage modéré. Ce diagnostic est cohérent avec les examens télévisuels réalisés après nettoyage qui indiquaient des colmatages faibles. La valeur élevée (0,3) du taux de colmatage uniforme équivalent calculé à partir de l'examen télévisuel du g. v. n° 2 en 2010 confirme la cinétique rapide du colmatage de cette tranche qui augmente d'environ une unité en 5 ans. Le colmatage de la tranche n° 3 a lui aussi une cinétique rapide. L'examen télévisuel du g. v. n° 2 en 2008 indiquait un colmatage de l'ordre de 0,2 ce qui est cohérent avec le nouvel indicateur. Partant d'un état de colmatage moins important, le nettoyage de cette tranche fin 2009 a permis d'atteindre un colmatage faible. La tranche n° 4 de Charlie a eu une cinétique de colmatage très rapide de 2002 à 2008. Ses g. v. n°s 2 et 3 ont atteint la limite haute de l'état de colmatage fort au-delà duquel le nouvel indicateur n'est plus défini. Ce colmatage important est confirmé par l'examen télévisuel du g. v. n° 2 en 2007 qui indiquait un taux moyen en plaque supérieure de l'ordre de 0,54 et un taux de colmatage uniforme équivalent de 0,67. Le nettoyage réalisé en 2008 a ramené l'état de la tranche n° 4 à un colmatage modéré.

Le très fort colmatage de la tranche n° 1 de Delta avant son nettoyage en 2007 a été discuté dans la section 4.3.1. Après nettoyage, les g. v. n°s 1 et 3 étaient faiblement colmatés et le g. v. n° 2 modérément. Ils ont des cinétiques de colmatage comparables à celles des g. v. de Charlie, les tranches de Delta étant elles aussi conditionnées à bas pH. Le colmatage des g. v. n°s 1 et 3 est aujourd'hui modéré et celui du g. v. n° 2 assez fort. On constate que les taux de colmatage uniforme équivalent calculés sont certainement

nettement surestimés, à cause de valeurs de taux de colmatage unitaires au-delà de la gamme de validité du modèle de perte de charge utilisé. En revanche les taux moyens conduisent à un diagnostic probablement trop optimiste. Les tranches n^{os} 2 et 3 de Delta ont des comportements analogues à celles de Charlie, caractérisés par une cinétique rapide. Elles sont passées en 5 ans environ d'un état de colmatage moyen à un état de colmatage fort. Pour ces deux tranches, les nettoyages chimiques ont été particulièrement efficaces permettant de les ramener à un état de colmatage faible, ce que confirme les examens télévisuels. De même que pour la tranche n^o 1, l'indicateur n'est pas défini pour la tranche n^o 4, avant nettoyage. L'état de colmatage après nettoyage est modéré ou moyen, selon les g. v. On constate pour cette tranche un comportement plus erratique du nouvel indicateur de colmatage. Celui-ci peut être imputé au bruit plus important des données, visible sur la figure F.4 de l'annexe F page 249. La cause de ce bruit fort n'a pas été identifiée. Plusieurs exemples de dégradations de l'instrumentation de mesure du n. g. l. ont été évoqués lors d'une réunion récente avec des opérateurs de centrales (Bertrand et Crinon, 2011). On peut par exemple suspecter l'encrassement des sphères de tranquillisation qui protègent la prise de pression haute de mesure n. g. l. d'être à l'origine de la qualité moindre de ces mesures.

Centrale	Tr.	G. v.	Année	Jambe chaude			Jambe froide						
				Moyen	Unif.	Obs.	Moyen	Unif.	Obs.				
Alpha	2	2	2007	0,1	0,1	160	0,09	0,09	145				
	4	2	2007	0,13	0,14	375							
Charlie	2	1	2008	0,06	0,07	210	0,05	0,05	qual.				
			2008	0,09	0,13	203			0,05	0,05	qual.		
			2010	0,11	0,3	366							
		3	2008	0,14	0,28	200	0,05	0,05	qual.				
	3	2	2008	0,21	0,22	191	0,19	0,2	168				
Delta	4	2	2007	0,54	0,67	189	0,49	0,52	134				
			2008	0,3	0,37	337							
	1	2	2008	0,15	0,54	371							
			2009	0,25	0,7	373							
			2011	0,3	0,8	291							
		3	2008	0,08	0,32	373	0,53	0,54	170				
	2	1	2007	0,57	0,59	185							
			2007	0,55	0,63	411				0,53	0,55	39	
			2008	0,56	0,65	412	0,54	0,56	35				
		3	2007	0,56	0,58	188	0,51	0,52	170				
	3	1	2009	2009	0,1	0,1	374	0,52	0,54	181			
				2007	0,74	0,79	201						
				2009	0,1	0,11	372						
				2010	0,1	0,1	385				0,05	0,05	qual.
				2007 ^{av.}	0,68	0,73	75				0,47	0,48	62
	3	2007 ^{ap.}	0,49	0,56	215	0,4	0,43	17					
		2009	0,1	0,1	377								

Tableau 13.2. : Estimations, à partir des résultats bruts d'examen télévisuel, des taux de colmatage moyen et des taux de colmatage uniforme équivalent (colonne « Unif. ») par jambe pour les g. v. des centrales étudiées dans cette thèse. La colonne intitulé « Tr. ». indique le numéro de la tranche et celle intitulé « Obs. » le nombre d'estimations unitaires utilisées pour calculer les taux moyens et uniformes équivalents. La mention « qual. » dans la colonne du nombre d'observation signifie que les clichés n'ont pas été analysés individuellement et qu'une valeur globale a été estimées qualitativement. Les exposants ^{av.} et ^{ap.} distinguent les deux examens télévisuels réalisé sur le g. v. n° 3 de la tranche n° 3 de la centrale Delta ; le premier avant lançage hydraulique et le second après. Les taux de colmatage uniforme équivalent concernant la tranche n° 1 de la centrale Delta ont été calculés avec des valeurs fortes de taux de colmatage unitaires, au-delà de la gamme de validité du modèle de perte de charge.

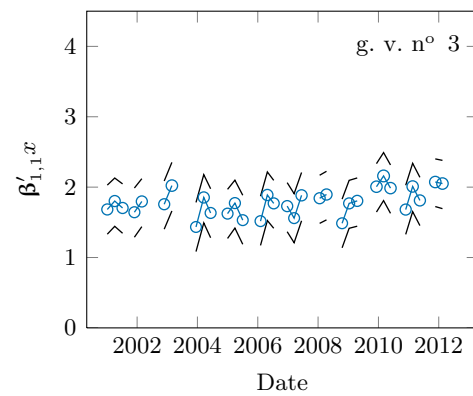
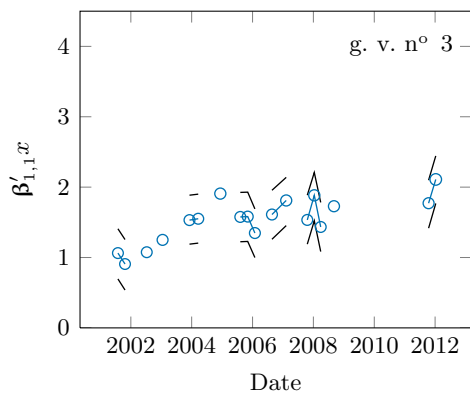
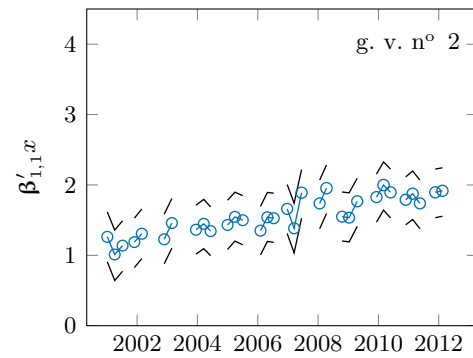
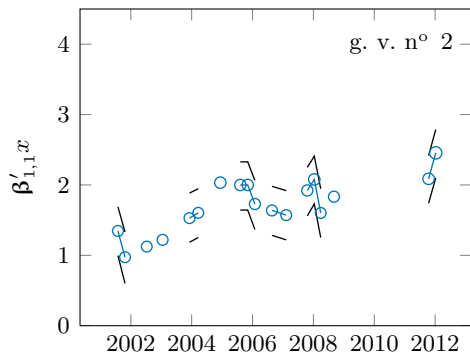
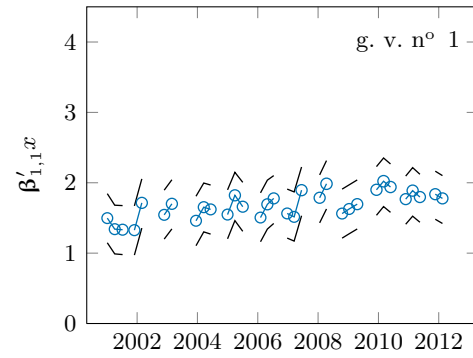
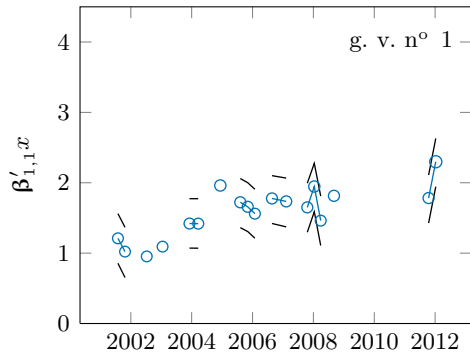


Figure 13.3. : Diagnostic du colmatage
 – centrale Alpha ;
 – tranche n° 2.

Figure 13.4. : Diagnostic du colmatage
 – centrale Alpha ;
 – tranche n° 3.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}x$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◦ 1000 points ; • 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}x$ estimée.

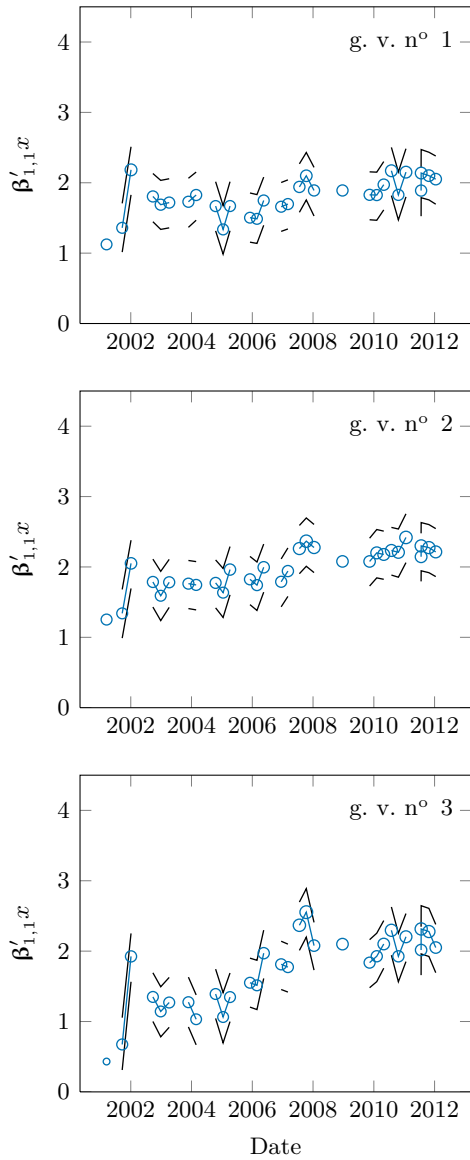


Figure 13.5. : Diagnostic du colmatage
 – centrale Alpha ;
 – tranche n° 4.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◦ 1000 points ; • 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée.

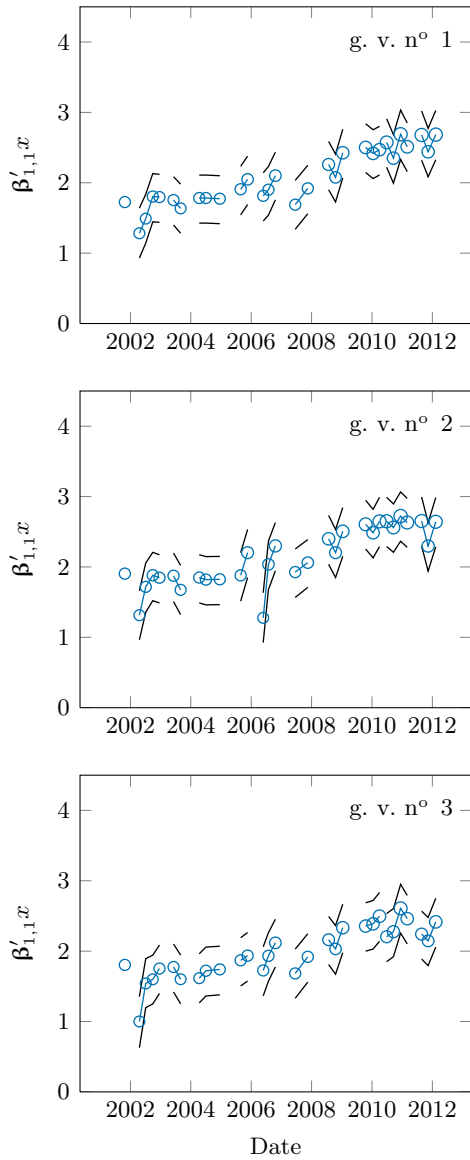


Figure 13.6. : Diagnostic du colmatage
 – centrale Bravo;
 – tranche n° 1.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◦ 1000 points ; • 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée.

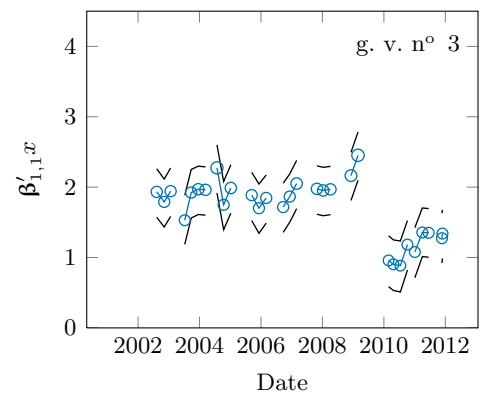
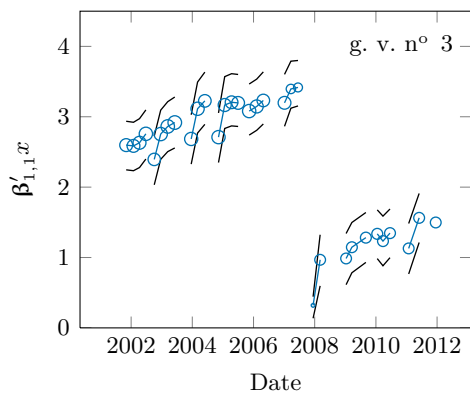
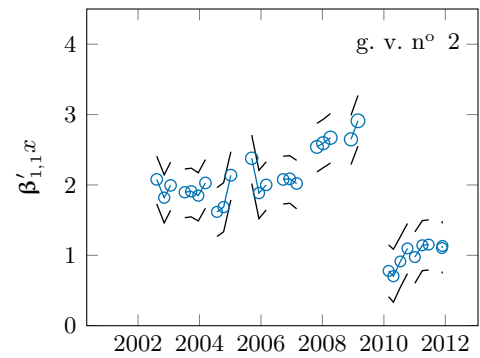
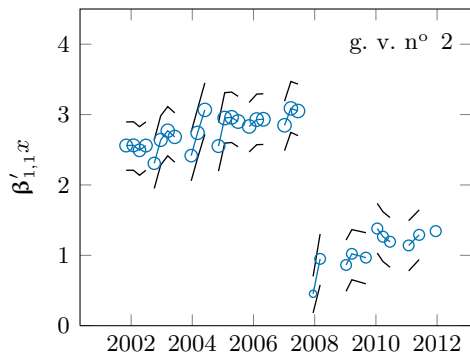
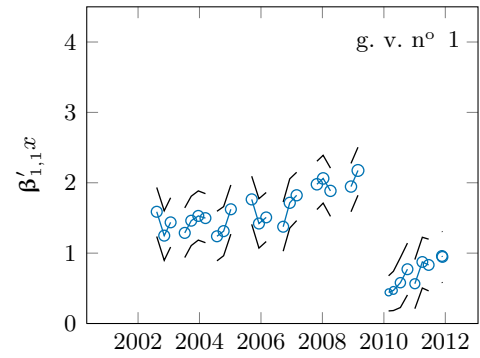
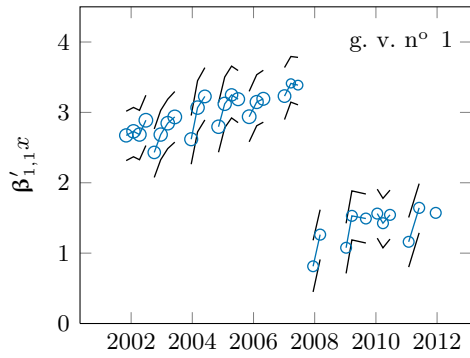


Figure 13.7. : Diagnostic du colmatage
 – centrale Charlie;
 – tranche n° 2.

Figure 13.8. : Diagnostic du colmatage
 – centrale Charlie;
 – tranche n° 3.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}x$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◦ 1000 points ; ◦ 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}x$ estimée.

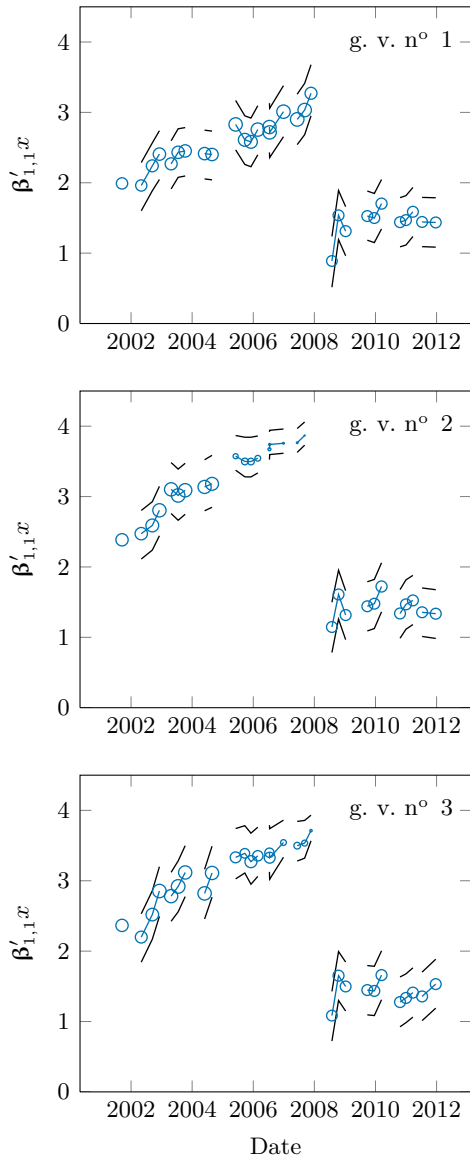


Figure 13.9. : Diagnostic du colmatage
 – centrale Charlie;
 – tranche n° 4.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◦ 1000 points ; • 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}\mathbf{x}$ estimée.

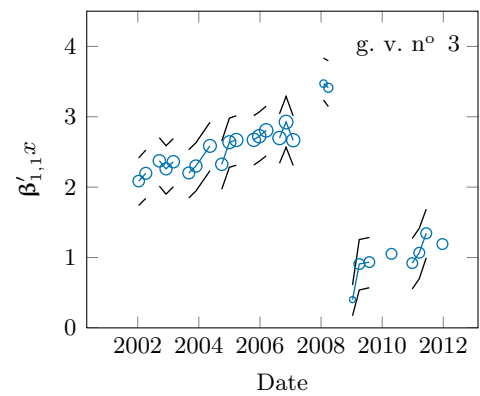
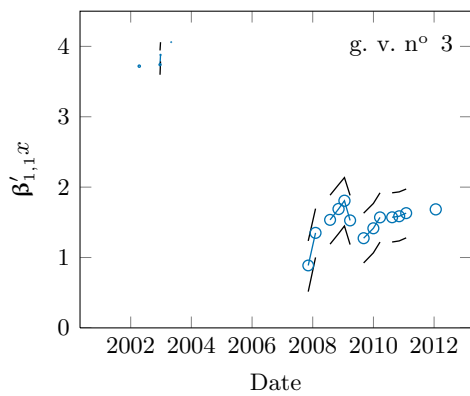
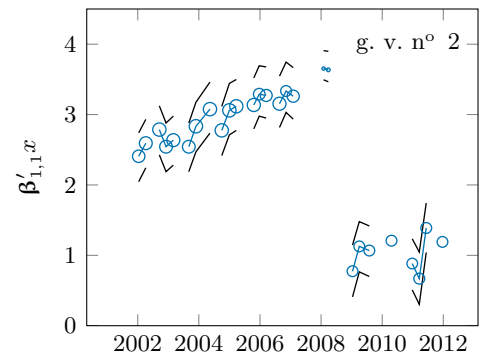
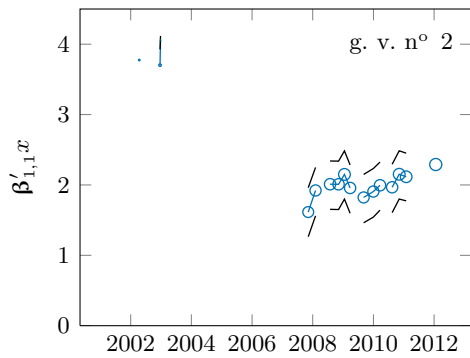
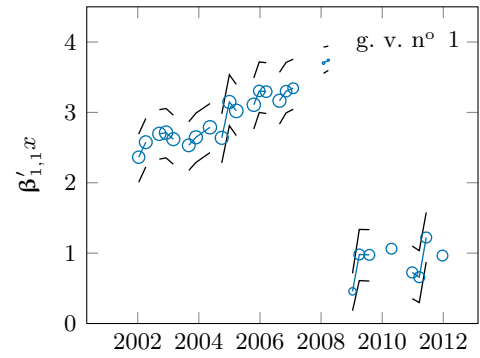
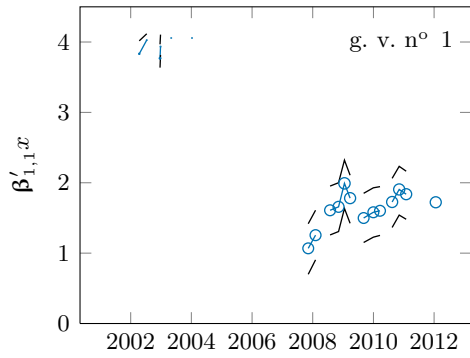


Figure 13.10. : Diagnostic du colmatage
 – centrale Delta;
 – tranche n° 1.

Figure 13.11. : Diagnostic du colmatage
 – centrale Delta;
 – tranche n° 2.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}x$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◐ 1000 points ; ◑ 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}x$ estimée.

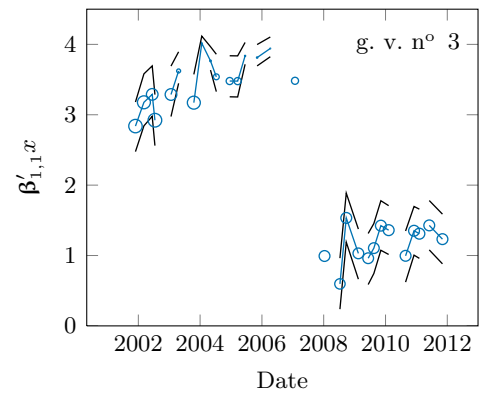
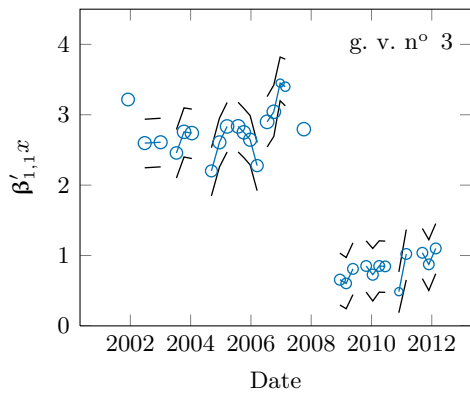
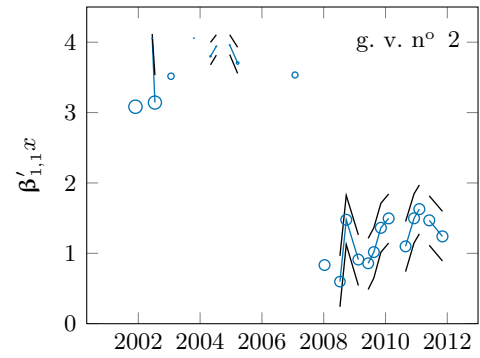
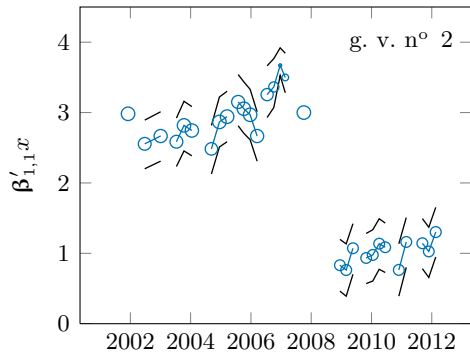
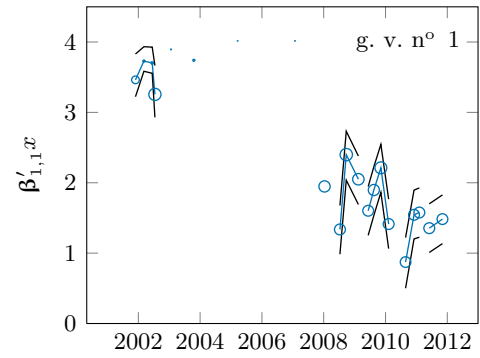
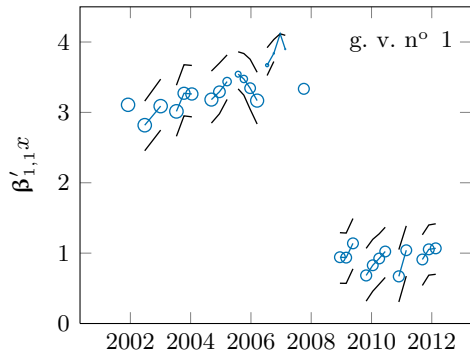


Figure 13.12. : Diagnostic du colmatage
 – centrale Delta;
 – tranche n° 3.

Figure 13.13. : Diagnostic du colmatage
 – centrale Delta;
 – tranche n° 4.

Légende :

- Valeur moyenne de la distribution de $\beta'_{1,1}x$ estimée. La taille est proportionnelle au nombre de points de simulation utilisés : ○ 2000 points ; ◐ 1000 points ; ◑ 500 points.
- Quantiles 0,05 et 0,95 de la distribution de $\beta'_{1,1}x$ estimée.

Cinquième partie

Conclusion

14. Perspectives de transposition de la méthodologie à l'étude de l'encrassement

Le phénomène d'encrassement décrit dans la section 2.2.1 relève de la même cause que le colmatage : l'accumulation de dépôts d'oxydes dans les g. v. Conséquemment, les moyens pour s'en prémunir sont sensiblement les mêmes et certaines méthodes pour diagnostiquer l'un permettent d'obtenir des informations sur l'autre. Lors des examens télévisuels de la plaque entretoise supérieure par exemple, les tubes sont visibles et l'on peut ainsi caractériser qualitativement leur état d'encrassement. Par ailleurs, ces deux phénomènes altèrent simultanément le comportement thermohydraulique des g. v. et l'on peut imaginer qu'ils interagissent, ou au moins que l'un peut affecter le diagnostic de l'autre. Il y a donc une complémentarité naturelle des méthodes de diagnostic du colmatage et de l'encrassement. Ce chapitre expose dans quelle mesure la méthodologie mise en œuvre au cours de cette thèse pourrait être utilisée pour étudier l'encrassement. La section 14.1 propose une revue des indicateurs d'encrassement existants. Un aperçu des travaux relatés dans la littérature est ensuite donné dans la section 14.2. Enfin, des pistes d'étude de l'encrassement par modélisation physique et statistique sont suggérées dans la section 14.3.

14.1. Indicateurs existants

L'une des différences fondamentales entre le colmatage et l'encrassement sont les masses de matière mise en jeu : quelques dizaines de kilogrammes pour le premier contre des centaines, voire plusieurs tonnes, pour le second. Quoique les fractions massiques d'espèces dissoutes et en suspension dans l'eau alimentaire soient très variables, il est possible d'estimer la masse de matière retenue dans un g. v. par un bilan de masse global. L'incertitude associée à un tel bilan sur plusieurs cycles est estimée à $\pm 20\%$ par Mercier (2010). Afin de réduire cette incertitude, la disposition transitoire n° 286 a instauré un programme de mesures régulières des teneurs en fer et en cuivre dans l'eau alimentaire de certaines tranches (Corredera, 2009). Le modèle de déposition BOUTHYC, présenté dans l'annexe B page 234, permet par ailleurs de prédire les masses déposées avec une précision satisfaisante (Gineste, 2009; Moleiro et coll., 2011). Ces estimations de masse peuvent par ailleurs être comparées aux masses extraites lors des nettoyages chimiques qui fournissent des estimations ponctuelles relativement fiables de l'état d'encrassement (passé) des g. v. Ainsi, là où l'on se posait, pour le colmatage, les questions « combien ? » et « où ? » ne subsiste que la deuxième pour l'encrassement. Une part mineure mais non négligeable de matière peut néanmoins se déposer sans constituer d'encrassement, par exemple au fond du g. v., sur la plaque à tubes. Une méthode de *cartographie des dépôts*

d'encrassement (*deposit mapping* en anglais) à l'aide d'une sonde à courant de Foucault est proposée par la société Areva. Elle est actuellement en cours d'expérimentation sur le parc français (Lacroix, 2012) mais il n'existe pour le moment aucune méthode pleinement fonctionnelle permettant de connaître la répartition de la matière déposée dans les g. v.

14.1.1. Pression de vapeur corrigée

L'encrassement a pour principal effet de modifier le transfert thermique au travers des tubes ce qui entraîne, en général, une baisse de la pression de la vapeur sortante. Outre la perte de rendement engendrée, cet effet peut s'avérer problématique lorsque la pression en sortie atteint la limite basse requise par les turbines en dépit d'une ouverture maximale des soupapes de régulation (Decout, 2009b). De fait, la pression de la vapeur à puissance nominale est actuellement l'indicateur principal utilisé pour le suivi de l'encrassement des g. v. (Crinon, 2009b). Cependant, la pression dépend de nombreux facteurs et l'on a vu qu'une puissance donnée pouvait être atteinte par différentes combinaisons de débit et de température des circuits primaire et secondaire. Elle augmente avec la moyenne des températures primaires en amont et aval du g. v. ainsi qu'avec la température de l'eau alimentaire, mais diminue avec la puissance totale transférée et le débit d'eau alimentaire. Ainsi, l'indicateur utilisé est obtenu en corrigeant la pression des écarts à un point de fonctionnement choisi comme référence. Le tableau 14.1 donne les valeurs de référence et les sensibilités actuellement en usage pour le suivi de l'encrassement des g. v. de type 51 B et apparentés (Crinon, 2009b). Le taux de bouchage correspond à la proportion, quelques pour-cent au maximum, des tubes ayant été bouchés volontairement du côté primaire lorsqu'ils étaient suspectés d'être fragilisés. Les sensibilités affichées ont été calculées par analyse de sensibilité locale du modèle PERFGV en faisant varier un paramètre à la fois (Decout, 2009a). Les inconvénients de cette approche ont été abordés dans le chapitre 7. Elle ne permet pas d'estimer les interactions entre variables et est peu précise en extrapolation, c'est-à-dire lorsque l'on considère un régime de fonctionnement notablement différent de la référence, si les effets des variables sont non linéaires.

14.1.2. Facteur d'encrassement

Le *facteur d'encrassement* est un autre indicateur, en usage notamment aux États-Unis (Kreider et coll., 1998; White et coll., 1998; Schwarz, 2001). Il consiste à estimer la résistance thermique globale d'un g. v. par un bilan enthalpique entre les entrées et sorties en l'assimilant à un échangeur à contre-courant simple dans lequel le coefficient de transfert thermique local et les chaleurs spécifiques des deux fluides sont constants dans tout l'échangeur. Plus précisément, le flux thermique local par unité d'aire, q''_{loc} , est lié aux températures des circuits primaire et secondaire, T_1 et T_2 , par le coefficient de transfert thermique local, η_{loc} :

$$q''_{loc} = \eta_{loc} (T_1 - T_2) . \quad (14.1)$$

On définit la *résistance thermique locale*, R''_{loc} , comme l'inverse du coefficient de transfert thermique local :

$$q''_{loc} = \frac{1}{R''_{loc}} (T_1 - T_2) . \quad (14.2)$$

Le symbole seconde, $''$, indique qu'il s'agit d'une résistance par unité d'aire, mesurée en $^{\circ}\text{Cm}^2/\text{W}$. La résistance thermique locale est la somme de la résistance d'un segment de tube propre et de la résistance additionnelle due à l'encrassement. Le facteur d'encrassement local, $\Delta R''_{loc}$, est égal à la différence entre la résistance thermique locale effective et une résistance R''_{loc}^* , choisie comme référence :

$$\Delta R''_{loc} = R''_{loc} - R''_{loc}^* . \quad (14.3)$$

La définition (14.2) peut être étendue à l'ensemble du g. v. :

$$q'' = \frac{\Delta T}{R''} . \quad (14.4)$$

La différence de température ΔT au dénominateur de l'expression précédente peut être exprimée, sous les hypothèses précédemment mentionnées, à l'aide de la *différence de température logarithmique moyenne* :

$$\Delta T = \frac{\Delta T_{entrée} - \Delta T_{sortie}}{\text{Log} \left(\frac{\Delta T_{entrée}}{\Delta T_{sortie}} \right)} . \quad (14.5)$$

Pour un échangeur à contre-courant, les écarts $\Delta T_{entrée}$ et ΔT_{sortie} sont les différences de température d'entrée et de sortie des deux fluides. Pour le g. v., on fixe usuellement $\Delta T_{entrée}$ égal à la différence entre les températures du fluide primaire en branche chaude (en amont du g. v.) et de la vapeur sortante supposée à saturation. L'écart ΔT_{sortie} est quant à lui calculé entre le fluide primaire en branche froide (en aval du g. v.) et l'eau alimentaire. Le facteur d'encrassement global est donc finalement donné par la formule suivante :

$$\Delta R'' = \frac{A_{ech}}{W} \frac{T_{ch} - T_{fr}}{\text{Log} \left(\frac{T_{ch} - T_{vap}}{T_{fr} - T_{vap}} \right)} - R''^* , \quad (14.6)$$

où W est la puissance totale échangée au travers de la surface du faisceaux d'aire A_{ech} , T_{ch} et T_{fr} les températures respectivement des branches chaude et froide du circuit primaire et T_{vap} la température de saturation de la vapeur sortant du g. v. déduite de sa pression qui seule est mesurée. Le choix de la résistance globale de référence, R''^* , importe peu si l'on se concentre sur l'évolution du facteur d'encrassement. La formule (14.6) fait apparaître le facteur d'encrassement comme un autre moyen d'estimer l'encrassement à partir de la pression de vapeur en prenant en compte la diversité des régimes de fonctionnement. Elle peut sembler au premier abord préférable à la méthode de correction de la pression exposée dans le paragraphe précédent car elle ne nécessite pas de choisir

Grandeur	Unité	Référence	Sensibilité en bar/·	
			Inconel 600	Inconel 690
Température branche froide	°C	323,1	0,158	0,164
Température branche chaude	°C	285,8	0,792	0,775
Température alimentaire	°C	219,5	$3,6 \cdot 10^{-2}$	$3,9 \cdot 10^{-2}$
Débit alimentaire	kg s^{-1}	504,2	$-2,9 \cdot 10^{-2}$	$-3,2 \cdot 10^{-2}$
Débit de purge	kg s^{-1}	0	$2,5 \cdot 10^{-2}$	$2,7 \cdot 10^{-2}$
Taux de bouchage	%	0	-0,169	-0,182

Tableau 14.1. : Valeurs de référence et sensibilités des paramètres affectant la pression utilisées actuellement pour le suivi de l'encrassement des g. v. Les sensibilités diffèrent sensiblement suivant l'alliage constituant les tubes (Inconel 600 ou 690).

un régime de fonctionnement de référence, ni de supposer la linéarité des variations des paramètres. En pratique, le facteur d'encrassement est aussi une correction locale. En effet, les hypothèses sur lesquelles la formule (14.6) est fondée ne sont pas valides pour un g. v. et les biais introduits sont d'autant plus grands que l'on considère une gamme de régimes de fonctionnement étendue. Dans les deux cas, il est nécessaire d'une part, de restreindre le domaine d'étude à des régimes de fonctionnement déterminés assez précisément, et d'autre part, de privilégier un suivi tendanciel à des estimations ponctuelles.

14.1.3. Pression en régime dynamique ?

La pression de la vapeur sortant du g. v. est la grandeur la plus impactée par le phénomène d'encrassement. Une première idée pour transposer la méthodologie développée pour l'étude du colmatage à celle de l'encrassement serait d'exploiter la réponse de la pression lors d'un essai R.G.L. 4. La figure 14.1 représente 32 courbes de réponse de la pression de vapeur enregistrées à la sortie des g. v. de la tranche n° 2 de la centrale Charlie, avant et après nettoyage chimique. Ces courbes de réponse de pression ont été enregistrées en même temps que les courbes de réponse de n. g. l. présentées dans la figure 3.7 page 43. De même que précédemment, les moyennes temporelles des courbes leurs ont été soustraites afin de filtrer l'effet statique. Elles ont par ailleurs été lissées avec le même algorithme et les mêmes conventions de couleurs ont été adoptées. Les courbes correspondant aux essais les plus anciens sont tracées en vert. Les courbes suivantes virent progressivement du vert à l'orange jusqu'à la situation juste avant le nettoyage chimique réalisé en 2007. Après nettoyage, elles virent du bleu ciel au mauve. Au contraire de ce qui avait été observé pour le n. g. l. et le colmatage sur la figure 3.7, aucune évolution des formes des courbes n'est détectable au cours du temps. Il n'est donc manifestement pas possible d'appliquer directement à l'encrassement la même méthodologie que pour le colmatage. De par son faible coût en temps de calcul, le modèle physique développé constitue néanmoins un outil adapté pour explorer de nombreux

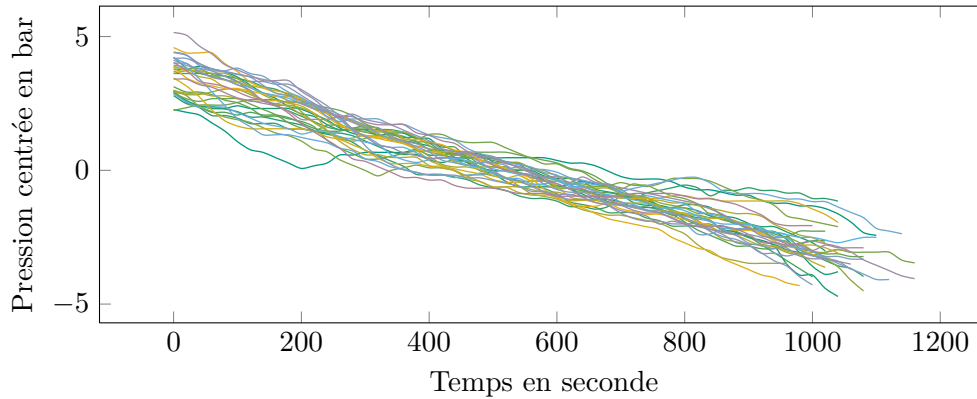


Figure 14.1. : Courbes de réponse de pression de vapeur sortante lors d'essais R.G.L. 4 mesurées sur le g. v. n° 3 de la tranche n° 2 de Charlie.

régimes de fonctionnement et configurations d'encrassement. La section 14.3 expose des pistes d'études utilisant l'analyse de sensibilité et la réduction de dimension.

14.2. Une bibliographie disparate mais abondante

On a vu, en conclusion de la section précédente, que les données disponibles pour l'étude de l'encrassement par la modélisation sont plus pauvres que dans le cas du colmatage car le comportement transitoire n'est pas informatif. La modélisation physique de l'effet de l'encrassement est aussi plus complexe que celle du colmatage. D'une part, le phénomène de transfert thermique est plutôt plus sensible au régime d'écoulement que la perte de charge et les corrélations dont on dispose sont plus approximatives, et d'autre part, la composition et la structure des dépôts jouent probablement un rôle important tandis que la forme du bourrelet colmatant pouvait être ignorée.

De fait, les effets de l'encrassement sur les centrales du parc français sont parfois contraires à l'intuition. La variation de la pression n'est pas toujours monotone, certaines tranches ne regagnent pas de pression après des nettoyages au cours desquels plusieurs tonnes de matière sont extraites et il arrive même que certaines perdent de la pression après nettoyage. Ces observations ne peuvent être expliquées par des modèles de transfert thermique par conduction seule dans des dépôts denses de magnétite pure. Il semble donc qu'un travail d'approfondissement des connaissances sur le phénomène d'encrassement soit indispensable pour progresser sur la voie du diagnostic. Il n'y a actuellement pas de consensus par exemple sur l'épaisseur typique ou la porosité des dépôts d'encrassement. Cependant, le phénomène est étudié depuis longtemps et la littérature recèle une quantité non négligeable d'information. Plusieurs axes d'investigation ont été identifiés pendant la thèse. La morphologie, la nature et la répartition des dépôts d'encrassement d'une part, leur coefficient de transfert thermique, les modèles de transfert thermique locaux (pour un segment de tube) et globaux (pour le g. v.) et les méthodes de diagnostic, d'autre part. Ces éléments sont interdépendants : la nature chimique des dépôts et leur structure

sont par exemple déterminants pour le coefficient de transfert thermique. En effet, les constituants des dépôts affectent leur conductivité thermique et la présence de pores ouverts ou fermés et leur diamètre modifient le mécanisme de transfert thermique.

Concernant le premier axe d'investigation, les techniques de caractérisation mises en œuvre sont coûteuses et chaque expérience ou campagne de mesures n'apporte qu'une quantité modeste d'information. Les mesures et estimations varient fortement d'une source à l'autre. Ceci est dû en partie à la variabilité des modes opératoires mais traduit néanmoins une diversité véritable du phénomène d'encrassement. Il est probable qu'il faille tenir compte de cette diversité pour être à même d'expliquer les comportements contre-intuitifs de certaines centrales. Les paragraphes qui suivent proposent une revue succincte de quelques expériences afin d'illustrer la diversité évoquée.

14.2.1. Nature et morphologie

Des mesures d'épaisseur de dépôts prélevés sur des tubes sont rapportées par Dijoux (2003) qui déplore un échantillonnage restreint et l'endommagement des dépôts par le processus d'extraction, sources d'une incertitude importante. Elles sont réalisées par microscopie optique sur des coupes transversales de tube de g. v. après enrobage et préparation métallographique. Les épaisseurs mesurées varient de quelques μm à près de $150\ \mu\text{m}$. Treize écailles de magnétite présentant une courbure correspondante à celle des tubes ont pu par ailleurs être prélevées dans les boues récupérées après lançage hydraulique (De Vito, 2002). Les épaisseurs de ces écailles sont comprises entre $200\ \mu\text{m}$ et $300\ \mu\text{m}$ environ.

De Vito (2002) propose une morphologie en feuillet de trois couches, du fluide secondaire au tube : magnétite pure à pores ouverts, magnétite et cuivre métallique à pores fermés et magnétite pure à pores fermés. Iglesias et coll. (2003) rapporte, pour des g. v. de centrale à eau lourde de type « Canada deuterium uranium » (CANDU), l'observation d'une structure à deux couches, du côté secondaire, comme du côté primaire. La couche interne du dépôt est dense et entrave la corrosion. La couche externe essentiellement composée de magnétite est moins compacte et contribue significativement à la résistance thermique. Les épaisseurs totales des dépôts mesurées sont respectivement de $0,7\ \mu\text{m}$ du côté primaire et de $25\ \mu\text{m}$ du côté secondaire.

La couche interne des dépôts analysés par Iglesias et coll. (2003), est enrichie en chrome ($\approx 50\ \%$) et appauvrie en nickel tandis que la couche externe comporte $5\ \%$ de chrome, $30\ \%$ de nickel et de fer environ. La présence de cuivre dans le dépôt côté secondaire est supposée responsable d'une moindre résistance thermique que celle du dépôt côté primaire. Turner et coll. (2000) font la même observation : les dépôts côté primaire et secondaire ont des résistances thermiques apparentes similaires bien qu'étant d'épaisseurs très différentes. Ils mettent cependant plutôt en cause les différences morphologiques de ces dépôts.

Une synthèse des observations de tubes extraits des g. v. du parc français de 1997 à 2003 a été constituée par Risson et Garbay (2009). Une synthèse détaillée des méthodes de caractérisation et des mécanismes de déposition et de transfert thermique est proposée par Varrin et coll. (1996). Des données concernant des tranches américaines sont réunies

dans le rapport de Kreider et coll. (1998).

14.2.2. Localisation dans le générateur de vapeur

Très peu de données sont disponibles quant à la localisation des dépôts. Les méthodes de cartographie de dépôt par sondes magnétiques mentionnées plus haut sont en cours de déploiement et les examens télévisuels ne donnent d'information que qualitative et pour une partie restreinte du faisceau de tubes. Ainsi, il semble qu'il faille pour le moment se tourner vers la modélisation. Outre le modèle BOUTHYC (Moleiro et coll., 2010; Pujet, 2001) décrit dans l'annexe B on peut citer un modèle de déposition utilisant ATHOS/SGAP proposé par Srikantiah et Chappidi (2000). La complexité des phénomènes à l'œuvre, notamment le couplage avec la thermohydraulique des g. v., rend très délicate leur modélisation. Pour une application au diagnostic, ces modèles devraient certainement être utilisés conjointement à des méthodes stochastiques, l'analyse de sensibilité entre autres.

14.2.3. Propriétés thermiques

Uhle (1997) citant des travaux de Rassokhin (1973) donne une valeur de $3,69 \text{ W m}^{-1} \text{ K}^{-1}$ pour la conductivité thermique de la magnétite pure. Les mesures de Sridharan et coll. (2002) aboutissent à une valeur comprise entre $1,73 \text{ W m}^{-1} \text{ }^\circ\text{C}^{-1}$ et $3,96 \text{ W m}^{-1} \text{ }^\circ\text{C}^{-1}$ pour une température variant entre $650 \text{ }^\circ\text{C}$ et $21 \text{ }^\circ\text{C}$.

L'effet de différents paramètres comme le flux thermique, le débit ou la porosité des dépôts a été mesuré par Uhle (1997) avec des dépôts de synthèse constitués de magnétite pure. Dans certaines configurations, le dépôt améliore le transfert thermique. Un modèle de transfert thermique au travers de dépôts poreux a été proposé par Uhle (1997) pour expliquer ces observations. Sous certaines conditions sur la distribution des diamètres des pores, la circulation d'eau dans la matrice de magnétite chauffée serait favorable à la vaporisation. Plus récemment, les coefficients de transfert thermique de tronçons de tube neuf, encrassé et nettoyé ont été comparés par Sridharan et coll. (2002). De façon surprenante, à différence de température fixée, le tube encrassé transfère plus de chaleur que le tube nettoyé qui en transfère lui même un peu plus que le tube neuf. Selon l'explication avancée par les auteurs, les dépôts n'étaient pas d'épaisseur suffisante pour générer une isolation significative mais leur surface poreuse augmente le nombre de sites de nucléation par unité d'aire engendrant ainsi un régime d'ébullition sous-saturée, très favorable au transfert thermique. Klimas et coll. (1998) ont mesuré des résistances assez importantes, de l'ordre de $0,05 \cdot 10^{-3} \text{ m}^2 \text{ }^\circ\text{C}^{-1} \text{ W}^{-1}$, dues à des dépôts côté primaire d'épaisseur comprise entre $90 \mu\text{m}$ et $150 \mu\text{m}$. Les résidus après nettoyage d'environ $20 \mu\text{m}$ d'épaisseur n'opposaient en revanche aucune résistance au transfert thermique et des dépôts fins de $2 \mu\text{m}$ augmentaient légèrement le transfert thermique. Ces résultats insistent sur la distinction nécessaire entre les dépôts de porosité importante et les dépôts consolidés, nettement plus denses.

14.3. Pistes d'étude

La pression de vapeur est la grandeur mesurée la plus sensible à l'encrassement mais c'est aussi celle pour laquelle il pose principalement problème. Cette situation est quelque peu différente de celle du colmatage pour lequel les risques posés ne sont pas apparents, le n. g. l. étant un indicateur indirect. La masse de matière dans le g. v. peut être estimée de façon satisfaisante mais les modèles actuels ne permettent pas d'en déduire l'impact sur la pression. Ce constat amène à poser la question suivante : « pour une masse d'encrassement donnée, quels effets peut-on prévoir sur la pression suivant des scénarios probables de composition, de morphologie et de répartition des dépôts ? ». La première étape, déterminante pour la suite, pour répondre à cette question est de réaliser une synthèse de la bibliographie évoquée dans la section précédente. La difficulté est de parvenir à rassembler et comparer des données issues de contextes dissemblables.

14.3.1. Modélisation physique

On peut envisager ensuite une phase de modélisation physique, d'abord locale puis globale. Un premier modèle global de transfert thermique par conduction uniquement au travers de dépôts denses de magnétite pure a été mis au point pendant cette thèse (Girard, 2012b). L'ordre de grandeur, 1 bar à 2 bar, des chutes de pression dans les cas d'encrassement important est correctement prédit par ce modèle. Cependant, la modélisation des effets concurrents à l'isolation thermique est probablement nécessaire pour permettre un diagnostic efficace. La complexité de ces modèles rend difficile leur intégration directe dans un modèle de g. v. permettant des simulations rapides. Aussi, la modélisation pourra, dans un premier temps, n'être que locale. Ce sera l'occasion, d'une part de vérifier que le modèle permet de reproduire les expériences menées sur des tronçons de tubes, celles de Sridharan et coll. (2002) par exemple, et d'autre part de réfléchir à un jeu de paramètres adéquat décrivant les dépôts. En effet, ces modèles plus précis introduisent de nouveaux paramètres, par exemple la distribution des diamètres des pores. Parmi les modèles de transfert thermique dans des dépôts poreux disponibles dans la littérature, on peut citer ceux de Uhle (1997), de C. PAN, B. G. JONES et A. J. MACHIELS (Kreider et coll., 1998, annexe C, p. 171) ou encore de Ul Haq et coll. (2011). Afin de faciliter leur intégration dans le modèle global, on pourra chercher au préalable à modéliser la surface de réponse de ces modèles par des fonctions analytiques sur le domaine correspondant au régime de fonctionnement envisagé pour le diagnostic.

Un raffinement de la modélisation de la partie haute des g. v. pourrait aussi s'avérer nécessaire pour le diagnostic. La description en détail du fonctionnement des dispositifs de séparation et de séchage n'était pas cruciale pour prédire la réponse du n. g. l. mais elle a probablement plus d'importance pour la mesure de pression de vapeur, située en aval. De fait, les g. v. 68/19 et 73/19 équipant respectivement les paliers 1300 MW et N4 ont un comportement face à l'encrassement radicalement différent des g. v. du palier 900 MW. Ce sont surtout ces derniers dont le comportement est incompris actuellement. Or, l'une des différences principales entre ces modèles est le nombre et la taille de leurs séparateurs cyclones : au nombre de 3 et d'un diamètre de 142 cm pour les g. v. de type

51 B, ils sont respectivement 18 et 133 et de diamètres égaux à 50 cm et 20 cm pour les g. v. de type 68/19 et 73/19. Il est donc probable que l'évolution des performances de ces dispositifs ne soit pas identique. Par le terme de *performance* des dispositifs de séparation et de séchage, on entend la limitation de leur perte de charge, de leur taux de débordement, soit la quantité d'eau liquide en sortie appelée *carryover* en anglais et de leur taux de refoulement, soit la quantité de vapeur chaude renvoyée en amont, appelée *carry-under* en anglais. Des dégradations de performance de ces dispositifs pourraient être à l'origine de certaines des pertes de pression observées, ce qui expliquerait l'absence de regain après nettoyage. La modélisation des dispositifs de séparation est traitée en détail par Kolev (2012). La synthèse de Regaldo (1983) contient de nombreuses données qualitatives et quantitatives spécifiques aux équipements du parc français.

Une fois qu'un modèle local satisfaisant aura été obtenu, se posera la question du choix des paramètres décrivant l'encrassement. Le modèle d'encrassement dense mentionné précédemment comporte une épaisseur de dépôt et une conductivité thermique. Cette dernière a été supposée constante et la SIR a été utilisée pour déterminer un découpage spatial adéquat. L'espace e. d. r. trouvé est une droite, ce qui suggère qu'avec ce modèle simple les effets des dépôts localisés différemment sont additifs. La sensibilité de la pression dépend essentiellement de la jambe considérée et de la position par rapport à la limite de vaporisation. Les travaux réalisés avec l'encrassement ont montré qu'un paramétrage parcimonieux est nécessaire pour obtenir un diagnostic robuste.

14.3.2. Amélioration de l'indicateur pression corrigée

Une analyse de sensibilité avec une méthode telle que celle de Sobol' permettrait d'obtenir des résultats plus robustes que l'analyse locale avec un plan d'expérience un-à-la-fois dont les inconvénients ont été évoqués dans le chapitre 7. Son caractère global permettrait de corriger partiellement les erreurs dues à des écarts trop importants au régime de référence et elle permettrait de prendre en compte les interactions entre variables. Il serait par ailleurs intéressant d'intégrer aux variables d'entrée les paramètres de l'encrassement définis lors de l'étape de modélisation physique afin de pouvoir comparer l'impact des écarts à l'état de référence et celui de l'encrassement. Une telle comparaison est indispensable pour déterminer la résolution de la méthode. L'analyse de la sensibilité de la pression aux différents paramètres de l'encrassement permettra en outre de déterminer si le modèle est à même de reproduire les observations faites sur les centrales.

Une première analyse de sensibilité simplifiée du modèle d'encrassement mis au point pendant cette thèse en ne faisant varier que l'épaisseur de dépôt et le débit primaire laisse penser que ce dernier pourrait avoir un impact non seulement sur la pression mais aussi sur les sensibilités relatives des dépôts suivant leur localisation. Cependant, le modèle ne tient pas compte des ajustements faits en dehors du g. v. compensant partiellement l'effet des variations de débit primaire d'un cycle à l'autre. Il sera probablement nécessaire d'ajouter des éléments au modèle pour pouvoir tenir compte de cette variable.

14.3.3. Amélioration de l'indicateur facteur d'encrassement

Le modèle de g. v. permet de comparer le facteur d'encrassement à la résistance thermique réellement imposée et ainsi d'évaluer la pertinence de cet indicateur en fonction du régime de fonctionnement. Ce travail serait plus intéressant si l'on disposait d'un modèle d'encrassement détaillé mais il peut toutefois être entrepris simplement en imposant des résistances thermiques locales. L'inconvénient de cette dernière approche est la difficulté de faire le lien entre les résultats obtenus et la réalité du phénomène d'encrassement. En fonction des résultats, le modèle pourrait ensuite être utilisé pour rechercher une formule macroscopique analogue à celle donnant le facteur d'encrassement mais adaptée à la géométrie et aux régimes de fonctionnement des g. v. Pour mener à bien cette tâche, il est nécessaire de disposer d'un jeu de paramètres décrivant l'encrassement en nombre restreint, ce que l'on pourrait obtenir grâce à des méthodes de réduction de dimension analogues à celles utilisées dans les travaux sur le colmatage.

14.3.4. Vers le diagnostic

Les améliorations des indicateurs existants proposées dans la section précédente pourraient permettre un suivi tendanciel mieux à même de prévoir l'évolution de l'encrassement des g. v. En fonction de leur conditionnement chimique, la masse de matière déposée dans les g. v. peut être estimée et prédite. L'effet sur le g. v. de cette masse de matière dépend de paramètres inconnus comme la répartition des dépôts sur le faisceau de tubes, la masse de matière déposée dans des zones n'impactant pas le transfert thermique (en bas, sur la plaque à tubes par exemple) et la morphologie des dépôts. On peut concevoir plusieurs « scénarios d'encrassement » et les éliminer progressivement, à mesure que les données de pression les infirment. Une telle approche profiterait d'être menée en interaction avec le développement du modèle de déposition BOUTHYC, décrit dans l'annexe B. Ce modèle peut en effet aider à concevoir des scénarios de déposition plausibles. Il profiterait en retour des possibilités de confrontation à la réalité offertes par une meilleure modélisation de l'effet de l'encrassement sur le comportement des g. v.

La démarche qui vient d'être exposée peut par ailleurs être formalisée à l'aide des méthodes d'apprentissage et d'assimilation de données. De fait, l'application de ces méthodes à la détection de l'encrassement des échangeurs de chaleur fait l'objet d'une littérature abondante. Par exemple Riverol et Napolitano (2005) ont utilisé des réseaux de neurones pour détecter l'encrassement d'échangeurs utilisés pour la pasteurisation du lait. Heo et coll. (2000) ont appliqué l'analyse en composantes principales et un modèle d'ondelettes à la détection de l'encrassement de tranches nucléaires coréennes, puis plus récemment, Heo (2008b) a proposé une méthode de surveillance de l'encrassement des échangeurs de centrales nucléaires ainsi qu'une revue des méthodes de surveillance par apprentissage appliquées au domaine nucléaire (Heo, 2008a). Dans le cadre d'un projet récent impliquant les universités de Valenciennes et du Hainaut-Cambrésis, de Poitiers et d'Islande, plusieurs méthodes ont été testées à l'aide d'une boucle d'essais, notamment le filtre de Kalman étendu (Jonsson et coll., 2007), des modèles d'ondelettes (Ingimundardottir et Lalot, 2011) et une méthode par « sous-espaces recursifs » (Lalot

et Mercère, 2008). Notons cependant que ces méthodes s'appliquent pour la plupart à des configurations plus simples que celle du g. v. D'où la nécessité de rechercher une formulation plus simple du problème du diagnostic de l'encrassement.

15. Synthèse et recommandations d'usage

Les méthodes de diagnostic existantes ont été présentées dans le chapitre 3. L'analyse des signaux enregistrés par les sondes à courant de Foucault ne donne qu'une estimation approximative de l'état de colmatage et ne peut différencier les taux supérieurs à 0,5. C'est néanmoins la seule méthode permettant d'estimer la répartition verticale du colmatage dans les g. v. L'interprétation des examens télévisuels est nettement plus directe. Les contraintes de réalisation de ces derniers et leur coût élevé restreignent cependant le nombre de g. v. pouvant être inspectés à une dizaine par an, pour l'ensemble du parc français. Ces contraintes limitent aussi la densité de l'échantillonnage pour chaque examen. Cela peut s'avérer problématique lorsque la répartition du colmatage dans la plaque est très hétérogène, comme c'est le cas par exemple après nettoyage chimique de g. v. très colmatés. Le suivi du n. g. l. en régime stationnaire est réalisé régulièrement mais les biais et dérives des mesures de n. g. l. en limitent fortement la précision. L'approche par analyse de courbes de réponse du n. g. l. est un complément attrayant de ces méthodes : le diagnostic peut être effectué fréquemment, l'utilisation du régime dynamique en améliore la robustesse et le modèle monodimensionnel offre des possibilités d'interprétation détaillée. La méthode de diagnostic à profil choisie proposée initialement reposait cependant sur des hypothèses fortes, notamment sur la répartition verticale du colmatage et sur la façon d'extraire l'information du signal de n. g. l., qu'il fallait examiner.

15.1. Synthèse des travaux réalisés

La première partie du travail, réalisée en collaboration avec Olivier DENEUX et Julien NINET, a consisté en l'amélioration du modèle monodimensionnel de g. v. existant. Elle fait l'objet de la partie II. Les hypothèses associées aux équations du modèle ont été vérifiées soigneusement, conduisant à en réécrire certaines ou à les remplacer par d'autres, plus adaptées. La restructuration du modèle en a amélioré la stabilité numérique et a supprimé des artefacts qui perturbaient le diagnostic. Deux phénomènes supplémentaires ont été pris en compte : les débits perpendiculaires à l'axe de modélisation et le déséquilibre cinématique des phases. Ces ajouts ont augmenté le nombre de degrés de liberté du modèle, permettant ainsi d'améliorer le réalisme des simulations sur la gamme de taux de colmatage considérée, allant de 0 à 0,7. Enfin, une nouvelle méthodologie de calage et de validation du modèle a été mise au point. Elle repose sur la comparaison en régime stationnaire avec le modèle THYC et sur la comparaison en régime dynamique avec les courbes de réponse de n. g. l. enregistrées sur les g. v. du parc français. Les tests de la validité du modèle ont montré des résultats très satisfaisants.

Trois méthodes statistiques ont par la suite été utilisées pour bâtir une nouvelle méthode de diagnostic par analyse de la réponse dynamique du n. g. l. Elles sont présentées, du

point de vue théorique, dans la partie III, et le détail de leur mise en œuvre pour le diagnostic du colmatage est exposé dans la partie IV. L'importance relative du colmatage de chaque demi-plaque a été quantifiée par une analyse de sensibilité globale du modèle avec la méthode de Sobol'. Ces premiers résultats ont en outre mis en évidence l'existence de familles de variables d'entrée ayant des effets de nature similaire ainsi que l'absence d'interactions trop importantes entre elles. L'influence du colmatage sur la forme des courbes de réponse a été explicitée à l'aide de l'analyse en composantes principales. L'effet prépondérant est l'augmentation de la pente globale des courbes : elles « pivotent » autour d'un point fixe situé légèrement au-delà de la moitié du transitoire. Dans une moindre mesure, le colmatage diminue la courbure des courbes de réponse : elles « s'aplatissent » avec deux points fixes situés approximativement aux premier et troisième quarts du transitoire. La décomposition du signal selon les composantes principales a abouti à une nouvelle façon de comparer les courbes mesurées et simulées en ne tenant compte que de l'effet dominant du colmatage. Enfin, la SIR a été employée pour identifier un sous-espace de projection optimal et définir ainsi un nouvel indicateur de colmatage.

La méthode de diagnostic par réduction de dimension proposée implique, par rapport à la méthode à profil choisi, deux transformations fondamentales de l'approche au diagnostic. La première est que l'on ne cherche plus à identifier les taux de colmatage individuels de chacune des demi-plaques. De fait, les tests réalisés avec l'algorithme de filtrage particulaire, et ultérieurement la réduction des dimensions d'entrée et de sortie du modèle, ont montré que la quantité d'information contenue dans les courbes de réponse n'est pas suffisante pour atteindre une telle résolution. L'indicateur proposé tient compte de cette restriction de la résolution du diagnostic et prévient ainsi le risque d'interprétation fallacieuse. Sa forme condensée permet par ailleurs d'étendre l'hypothèse de répartition verticale à un ensemble beaucoup plus vaste de configurations de colmatage sans compromettre l'intelligibilité du diagnostic. Le conditionnement par une distribution de profils de colmatage est ici immédiat. L'algorithme de filtrage particulaire, au contraire, aurait nécessité la construction, plus complexe, d'un nouveau noyau de transition. L'autre changement conceptuel est que les examens télévisuels ne sont plus utilisés pour valider la méthode comme cela avait été envisagé initialement : les deux méthodes sont au contraire utilisées de façon complémentaire. Cette décision repose sur deux arguments. Le premier est que ces deux méthodes ne mesurent pas la même chose. Les examens télévisuels apportent une estimation des taux de colmatage des trous foliés de la plaque supérieure tandis que le diagnostic par analyse de la réponse dynamique du n. g. l. est une évaluation de l'*effet* global du colmatage d'un g. v. Moyennant un modèle de perte de charge, il est possible de faire se correspondre partiellement ces estimations, à l'aide de la notion de taux de colmatage uniforme équivalent. Elles se distinguent néanmoins par leur caractère local ou global. On peut voir dans cette distinction une analogie avec les échelles utilisées par exemple pour caractériser les séismes : seule une correspondance approximative peut être établie entre une intensité mesurée sur l'échelle de Mercalli et une magnitude de moment calculée à partir de sismogrammes. Ainsi, plutôt qu'une comparaison quantitative de grandeurs hétérogènes, on préférera confronter qualitativement les décisions pour la stratégie de maintenance que suggèrent l'un et l'autre de ces diagnostics. Le second argument tient à l'incertitude inhérente à la modélisation

physique des g. v. et du phénomène de colmatage. La répartition de la perte de charge entre le *downcomer* et l'admission dans le *riser*, le coefficient de calage des pertes de charge dans le *riser* et le coefficient multiplicatif appliqué aux courbes de n. g. l. mesurées sont autant de paramètres qui altèrent la définition des taux de colmatage dans le modèle. Ainsi, la variable « taux de colmatage » dans le modèle ne correspond pas exactement à la définition du taux de colmatage comme le rapport de l'aire de la section bloquée par l'air d'un trou exempt de colmatage. Autrement dit, le taux de colmatage n'est en définitive qu'une variable intermédiaire de la méthode de diagnostic par analyse de la réponse du n. g. l.

15.2. Recommandations d'usage et perspectives d'amélioration

L'indicateur proposé n'est pas défini pour des états de colmatage extrêmes comme celui de la tranche n° 1 de la centrale Delta avant son nettoyage en 2007. On pourrait être tenté de procéder à des extrapolations du fait de la linéarité prononcée de la relation entre l'indicateur et la pente globale des courbes de réponse, en particulier pour la distribution de profils de colmatage élaborée dans le chapitre 11. Une approche plus rigoureuse serait d'affiner la modélisation de l'effet local des très forts taux de colmatage. Les résultats d'expériences sur des écoulements diphasiques dans la maquette P2C prochainement disponibles pourraient être utilisés à cette fin.

Les simulations réalisées avec le modèle monodimensionnel indiquent que les trois demi-plaques les plus basses en jambe chaude ont un effet sur la pente globale des courbes de réponse opposé à celui des autres demi-plaques. Cet effet est de faible amplitude et les simulations réalisées avec le modèle de déposition par *vena contracta* suggèrent que les taux de colmatage de ces demi-plaques sont plutôt plus faibles que ceux des plaques hautes. Néanmoins, si pour un g. v. donné les valeurs du nouvel indicateur paraissent anormalement basses, l'hypothèse d'une répartition de colmatage plus importante en bas de g. v. devrait être considérée.

La détermination du facteur multiplicatif appliqué aux réponses mesurées fait partie intégrante de la procédure de calage. Elle doit être effectuée en relation avec l'ajustement de la répartition de la perte de charge entre le *downcomer* et l'admission dans le *riser* d'une part, et du coefficient de calage des pertes de charge dans le *riser* d'autre part. La linéarité de la relation entre l'indicateur de colmatage et la pente globale des réponses limite fortement l'influence de la valeur du coefficient multiplicatif sur le diagnostic. Si cette valeur était modifiée, il serait cependant nécessaire d'établir une nouvelle échelle de gravité du colmatage pour remplacer celle donnée dans le tableau 13.1. Cette échelle de gravité a été proposée comme aide à l'interprétation du nouvel indicateur. Elle ne doit pas occulter le fait que celui-ci est plus performant pour un suivi tendanciel de l'évolution de l'état des g. v. que pour des estimations ponctuelles.

Les comparaisons entre les g. v. d'une même tranche, et a fortiori de tranches différentes, requièrent des précautions. En effet, certaines nuances de conception des modèles de g. v. apparentés au modèle 51 B ont été négligées et surtout, l'instrumentation de mesure du n. g. l. peut varier d'un g. v. à l'autre.

Il est possible de remplacer facilement l'hypothèse faite, à partir du modèle *vena contracta*, sur la répartition verticale : il suffit pour cela de simuler des transitoires avec un nouvel échantillon de configurations de colmatage. Il est nécessaire dans ce cas de recalculer, d'une part, les composantes principales de l'échantillon de courbes simulées et d'autre part, les directions e. d. r. Il est cependant probable que les résultats obtenus à l'issue de ces deux calculs soient similaires à ceux ayant été présentés. En effet, les trois échantillons gaussien, uniforme et déduit du modèle de déposition par *vena contracta*, avaient eux-mêmes conduit à des résultats semblables pour ce qui est des directions principales et de la première direction e. d. r. calculée à partir des scores de la première composante principale.

Certains travaux débutés en fin de thèse n'ont pu être achevés faute de temps. Leur objectif était de corriger les écarts au transitoire de référence des valeurs initiales de la température primaire et du débit de vapeur. L'approche adoptée consiste à simuler, pour un même échantillon de configurations de colmatage, deux transitoires : l'un dans les conditions de référence et l'autre avec des valeurs initiales de température et de débit différentes, déterminées à partir d'une loi de probabilité ajustée sur un échantillon de mesures. Les scores des courbes ainsi simulées peuvent ensuite être déterminés par projection, après avoir ajusté les durées des transitoires. On obtient ainsi pour chaque paire de transitoires, un écart de scores. La modélisation de la surface de réponse des écarts de scores en fonction des écarts de valeurs initiales devrait permettre de déduire une correction simple de ces derniers.

A. Caractéristiques géométriques des générateurs de vapeur de type 51 B

Cette annexe rassemble les principales caractéristiques géométriques des g. v. de type 51 B et de leurs éléments internes. Elles sont identiques, pour la plupart, à celles des g. v. de type 51 A, 51 BI et 51 M. Ces données sont issues de la note de Regaldo et coll. (1984) et de plans de conception des g. v. de type 51 B (Framatome, 1983a,b, 1984).

Élément	Unité	Valeur
Nombre de tubes		3330
Nombre de rangée de tubes		94
Nombre de colonnes de tubes		46
Diamètre extérieur tubes	mm	22,22
Épaisseur des tubes	mm	1,27
Surface d'échange (extérieure)	m ²	4700
Type de pas de perçage		Carré
Valeur du pas	mm	32,54
Masse totale	t	51,5

Tableau A.1. : Caractéristiques du faisceau de tubes des g. v. de type 51 B.

Élément	Unité	Plaques inférieures	Plaque supérieure
Épaisseur	mm	30	
Type de passage des tubes		Broché	
Nombre de trous brochés		6660	6352
Aire les trous brochés	m ²	2,01	1,91
Diamètre extérieur	mm	3118	
Nombre de trous non tubés		208	
Aire des trous non tubés	cm ²	7,91	
Nombre de trous de circulation		–	200
Aire des trous de circulation	cm ²	–	4,54

Tableau A.2. : Caractéristiques des plaques entretoises des g. v. de type 51 B. La plaque supérieure comporte des différences avec les plaques inférieures qui sont indiquées dans la dernière colonne. Les valeurs identiques ne sont pas reprise. Le symbole « – » indique les éléments absents des plaques inférieures.

B. Modèles de déposition BOUTHYC et COLMATHYC

Le logiciel THYC simule les écoulements thermohydrauliques dans des milieux encombrés, de type faisceau d'échangeurs à tubes, qu'il assimile à un milieu poreux de porosité fluide notée π_f (Guelfi et Pitot, 2004). Les modules BOUTHYC et COLMATHYC reposent tous deux sur un couple d'équations principales, l'une modélisant le transport des particules, l'autre la croissance des dépôts. Ces modules sont décrits ici succinctement ; on se reportera aux travaux de Prusek (à paraître) pour une formulation complète des équations. Pour une maille donnée, on note ρ la masse volumique du mélange liquide-vapeur, χ le titre massique du mélange et ε la fraction massique de particules dans l'eau. Cette dernière est ici une grandeur *locale*, à ne pas confondre avec la fraction massique de l'eau alimentaire définie dans la section 11.1. L'équation de transport des particules s'écrit :

$$\frac{\partial}{\partial t} (\pi_f \rho (1 - \chi) \varepsilon) + \text{div } \mathbf{J} = \Psi, \quad (\text{B.1})$$

Le premier terme de cette équation représente la variation temporelle de la masse de particules dans la maille. Le vecteur \mathbf{J} est la somme d'un terme de convection, dépendant de la vitesse \mathbf{u}_l du fluide, et d'un terme de diffusion, dépendant du gradient de fraction massique de particules et d'un coefficient de diffusion turbulente, D :

$$\mathbf{J} = \pi_f \rho (1 - \chi) (\varepsilon \mathbf{u}_l + D \mathbf{grad}(\varepsilon)). \quad (\text{B.2})$$

Le membre de droite, Ψ , représente le transfert volumique de particules vers la paroi.

On note par ailleurs m_d la masse de particules déposées par unité de surface et Φ le flux de particules déposées. L'équation de croissance des dépôts s'écrit :

$$\frac{dm_d}{dt} = \Phi. \quad (\text{B.3})$$

À l'échelle retenue pour la modélisation, le flux Φ est décrit par des corrélations semi-empiriques. Les mécanismes gouvernant l'encrassement dans le module BOUTHYC sont l'ébullition, la diffusion, l'inertie, la sédimentation, et la thermophorèse (Liner et coll., 1992; Turner et coll., 1994). Le mécanisme gouvernant le flux de particules vers les parois des trous foliés et causant le colmatage est appelé *vena contracta* (Rummens, 1999; Rummens et coll., 2004). La « *vena contracta* » est la contraction soudaine que rencontre le fluide à l'entrée des trous foliés. Cette restriction brutale décolle la couche limite et une zone de recirculation se développe, proche du bord. Les faibles vitesses

d'écoulement dans ces zones sont propices à la déposition des particules. Les zones de recirculation sont d'autant plus grandes que la restriction du passage du fluide est importante. Le colmatage par *vena contracta* est donc auto-entretenu. On suppose que toutes les particules transportées vers la paroi par ce mécanisme s'y déposent et ne peuvent être arrachées. Le flux Φ correspondant à ce phénomène est de la forme suivante :

$$\Phi = v_d \rho (1 - \chi) \varepsilon, \quad (\text{B.4})$$

avec v_d une vitesse de déposition dépendante des paramètres thermohydrauliques ayant une influence notable sur le colmatage (Prusek et coll., 2011) : la différence $\Delta\rho$ de masse volumique entre les particules et le fluide, la vitesse u_f du fluide, le titre du fluide ou encore la viscosité μ du fluide. Son expression est la suivante :

$$v_d = \omega k_{geo} \frac{\Delta\rho u_f^2 (1 - \chi) d^2}{\mu}. \quad (\text{B.5})$$

Le coefficient ω est un paramètre non observable dépendant de l'ensemble des paramètres thermohydrauliques. Sa valeur est obtenue à l'aide d'une méthode de calage décrite dans la section 11.1.1 en faisant l'hypothèse qu'elle est identique partout dans le g. v. Le coefficient k_{geo} est une fonction affine du taux de colmatage dont les coefficients dépendent de la géométrie des trous foliés (Prusek et coll., 2011). La vitesse de déposition dépend de la masse volumique des particules ainsi que du carré de leur diamètre. Les particules sont supposées sphériques et composées de magnétite de masse volumique constante valant 5200 kg m^{-3} .

Les équations qui viennent d'être présentées permettent de déterminer la masse de magnétite déposée dans les trous foliés. Le volume du dépôt se déduit directement de sa porosité et de la masse volumique de la magnétite. Il est cependant nécessaire de préciser la morphologie du dépôt, notamment son extension selon la direction verticale, pour pouvoir calculer le taux de colmatage correspondant. Une corrélation liant taux de colmatage et volume a été établie par Morvan et Adobes (2008) à partir de gabarit tridimensionnel réalisés en observant des examens télévisuels Six gabarits ont été réalisés pour les taux de colmatage 0,4, 0,45, 0,5, 0,7, 0,85 et 0,95, puis un modèle exponentiel a été ajusté. En notant x le taux de colmatage et V le volume de dépôt, ce modèle a pour expression

$$x = x_{max} \left(1 - \exp\left(-\frac{V}{V_c}\right) \right). \quad (\text{B.6})$$

Les valeurs de x_{max} et V_c sont respectivement 1,02 et $42,92 \text{ mm}^3$. La valeur de x_{max} légèrement supérieure à 1 peut surprendre mais le modèle exponentiel n'a été choisi que parce qu'il permettait un bon ajustement et non à partir de considérations physiques. L'incertitude sur la forme du dépôt est importante pour les très forts colmatages. Les résultats de simulations comportant des taux de colmatage supérieurs à 0,8 devraient être considérés avec réserve.

C. Régression inverse par tranche avec un échantillon uniforme

Afin de tester la robustesse de la SIR lorsque la distribution d'entrée n'est pas à symétrie elliptique, la démarche présentée dans la section 10.5 a été mise en œuvre en attribuant au taux de colmatage de chaque demi-plaque une distribution uniforme sur l'intervalle $[0 ; 0,65]$.

C.1. Régression inverse par tranche monodimensionnelle

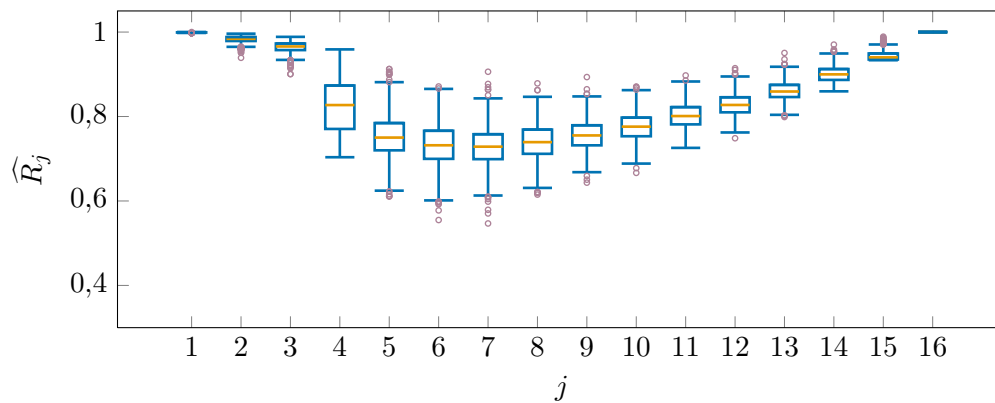


Figure C.1. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon uniforme et les scores de la première composante principale comme variable de sortie.

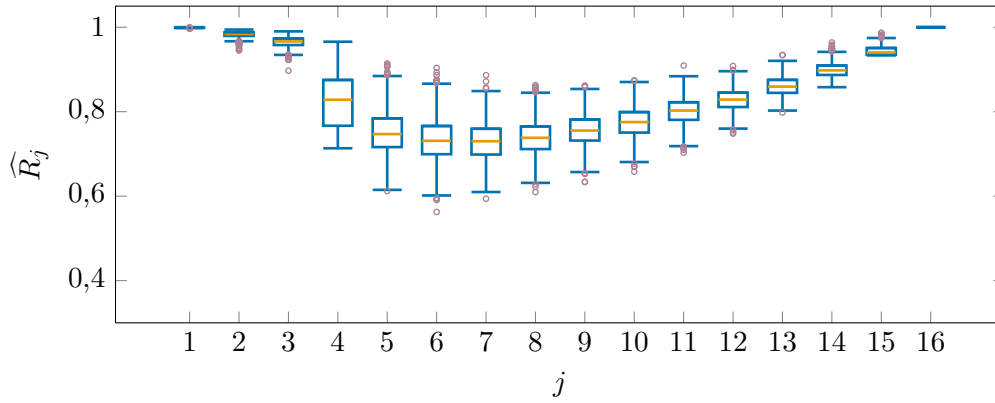


Figure C.2. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon uniforme et les scores de la deuxième composante principale comme variable de sortie.

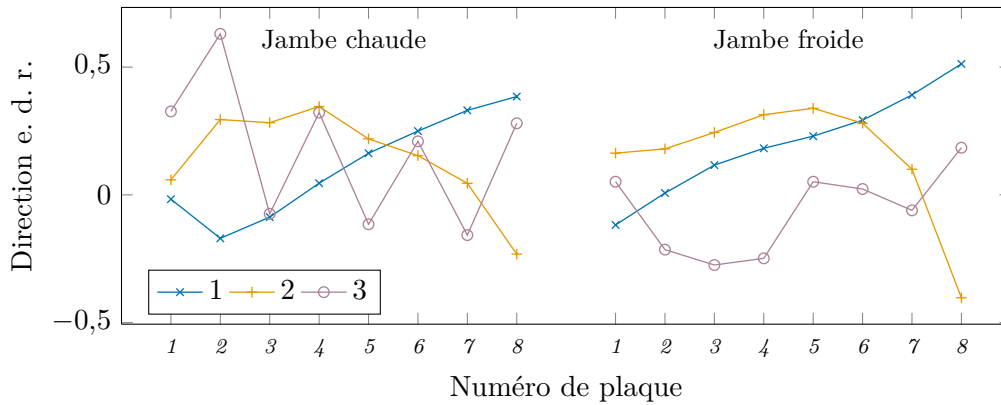


Figure C.3. : Trois premières directions e. d. r. calculées avec l'échantillon uniforme et les scores de la première composante principale comme variable de sortie.

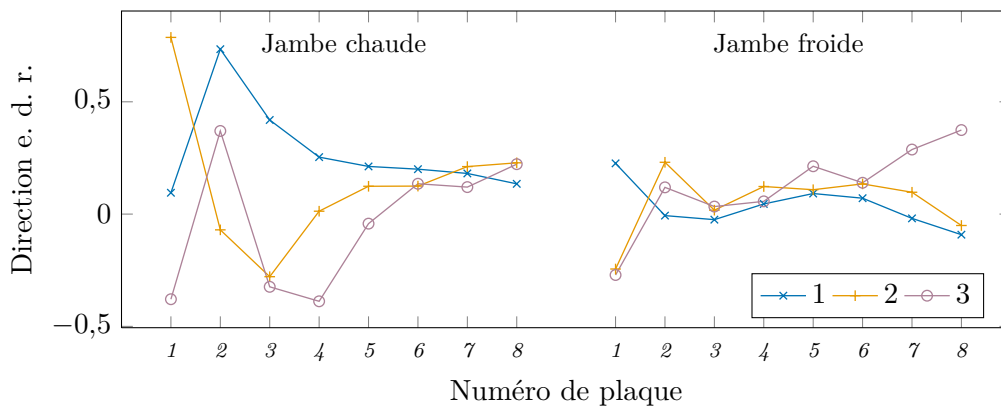


Figure C.4. : Trois premières directions e. d. r. calculées avec l'échantillon uniforme et les scores de la deuxième composante principale comme variable de sortie.

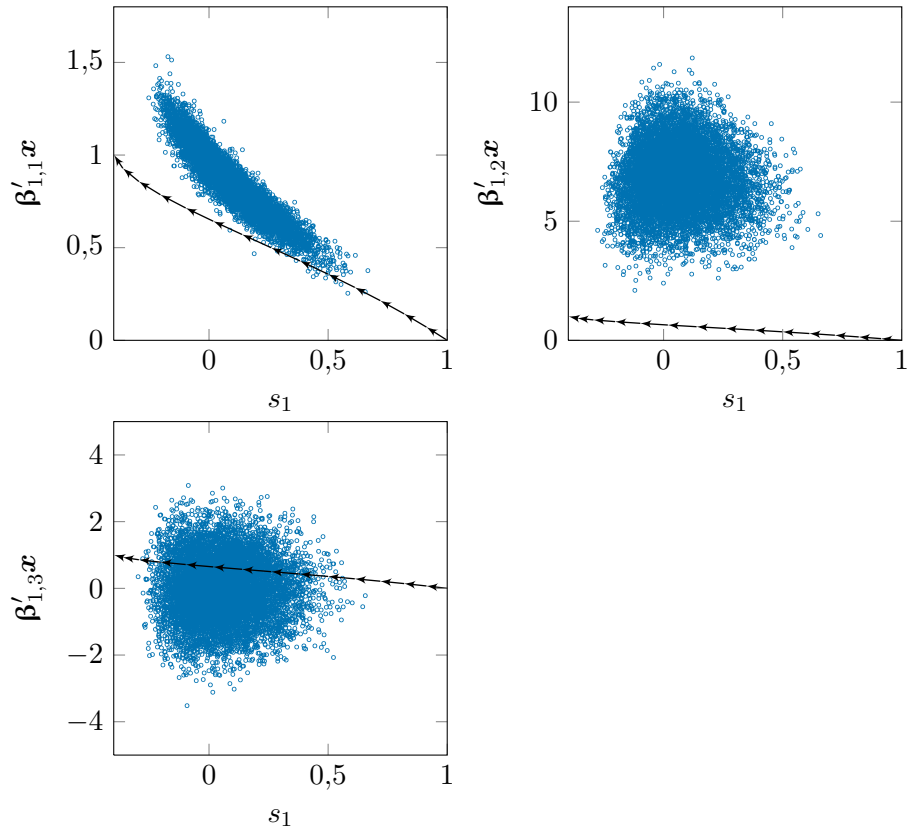


Figure C.5. : Nuages des projections des individus de l'échantillon uniforme selon les directions e. d. r. obtenues avec les scores de la première composante principale, s_1 , comme variable de sortie, en fonction de s_1 .

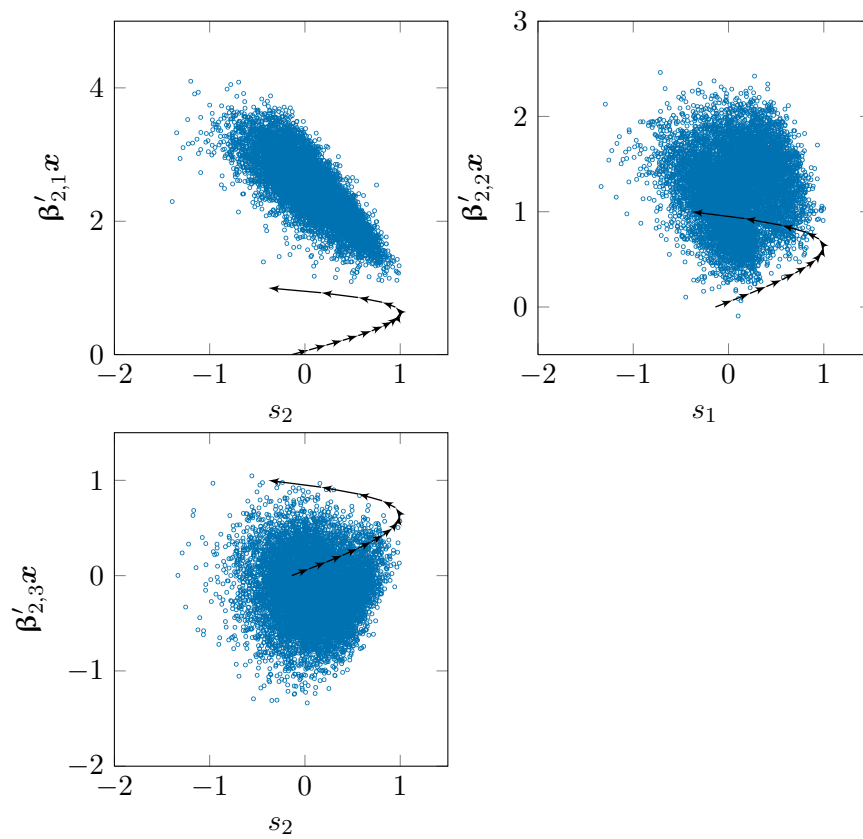


Figure C.6. : Nuages des projections des individus de l'échantillon uniforme selon les directions e. d. r. obtenues avec les scores de la deuxième composante principale, s_2 , comme variable de sortie, en fonction de s_2 .

C.2. Découpage marginal groupé

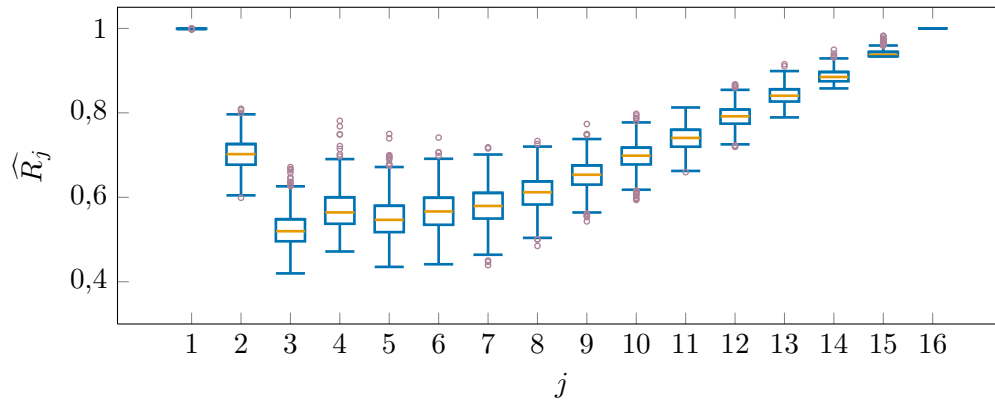


Figure C.7. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues par découpage marginal groupé de l'échantillon uniforme.

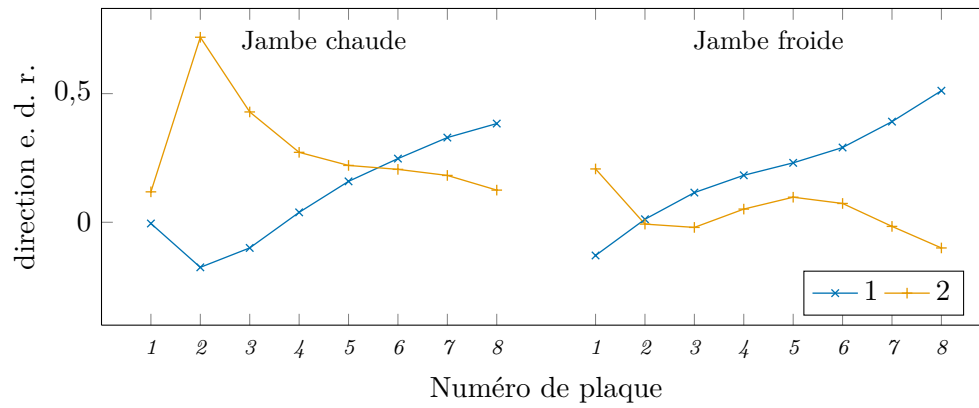


Figure C.8. : Deux premières directions e. d. r. obtenues par découpage marginal groupé de l'échantillon uniforme.

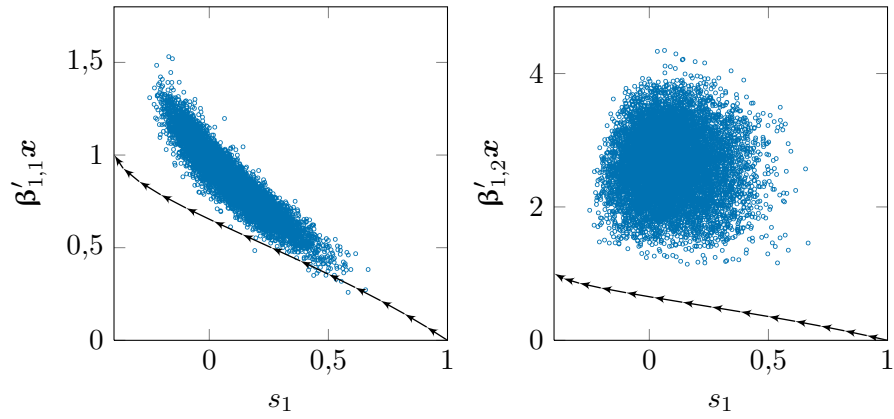


Figure C.9. : Nuages des projections des individus de l'échantillon uniforme selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la première composante principale, s_1 .

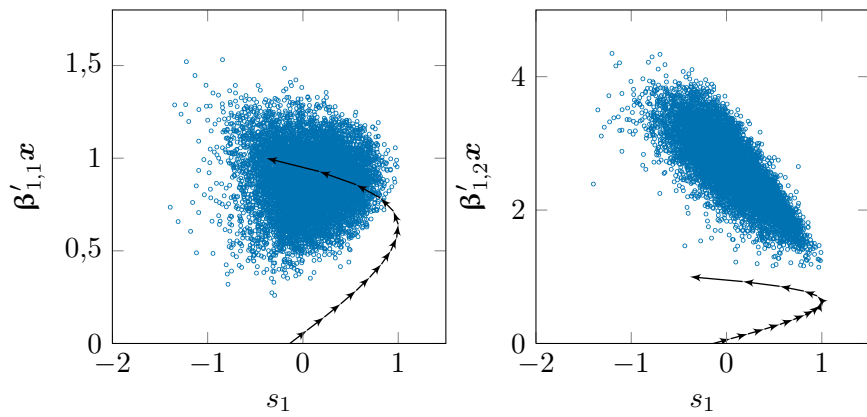


Figure C.10. : Nuages des projections des individus de l'échantillon uniforme selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la deuxième composante principale, s_2 .

D. Mesures granulométriques

Diamètre en μm	Numéro d'échantillon																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	299	228	234	245	229	216	238	243	228	240	233	238	234	491	504	545	541	485	545	91	104	67
2	148	102	118	101	119	90	79	114	101	104	107	80	102	174	199	180	199	190	199	39	38	36
3	67	64	54	48	51	38	46	35	41	40	50	27	28	67	74	86	84	65	70	17	31	19
4	43	29	32	17	19	16	17	32	11	19	14	15	16	36	32	22	35	28	31	19	11	12
5	15	5	4	6	7	2	4	2	6	5	5	4	2	8	1	5	9	10	5	2	8	3
6	8	1	1	3	2	2	1	1	1	2	3	2	2	4	6	3	5	1	5	4	4	3
7	2	1	0	2	4	1	3	1	0	2	0	0	0	2	2	1	2	0	0	2	1	3
8	2	3	1	0	2	0	2	0	1	0	0	0	2	1	0	1	2	3	1	0	2	1
9	3	1	0	1	0	0	0	0	0	2	1	1	1	3	0	0	0	1	1	0	2	0
10	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
11	2	2	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0
12	3	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1
13	1	1	0	0	0	0	0	0	0	1	1	0	2	0	0	0	0	0	0	0	0	2
14	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0
15	8	1	0	0	2	0	3	0	1	3	3	1	0	1	0	0	1	1	0	0	0	0
20	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau D.1. : Effectifs mesurés par tailles de particules pour 22 échantillons prélevés dans les tranches n° 3 et 4 de Charlie. Les diamètres indiqués correspondent à la valeur *inférieure* des classes de diamètre de particule. Par exemple, l'avant-dernière ligne du tableau dénombre les particules détectées dont le diamètre est compris entre 20 et 25 μm . Les particules de diamètre inférieur à 1 μm ne sont pas détectées.

E. Régression inverse par tranche avec l'échantillon *vena contracta*

Les calculs exposés dans le chapitre 10 pour l'échantillon gaussien été effectués avec un échantillon de 14 233 individus issus la distribution de profils de colmatage construite à l'aide du modèle de déposition par *vena contracta*. La construction de cette distribution fait l'objet du chapitre 11.

E.1. Régression inverse par tranche monodimensionnelle

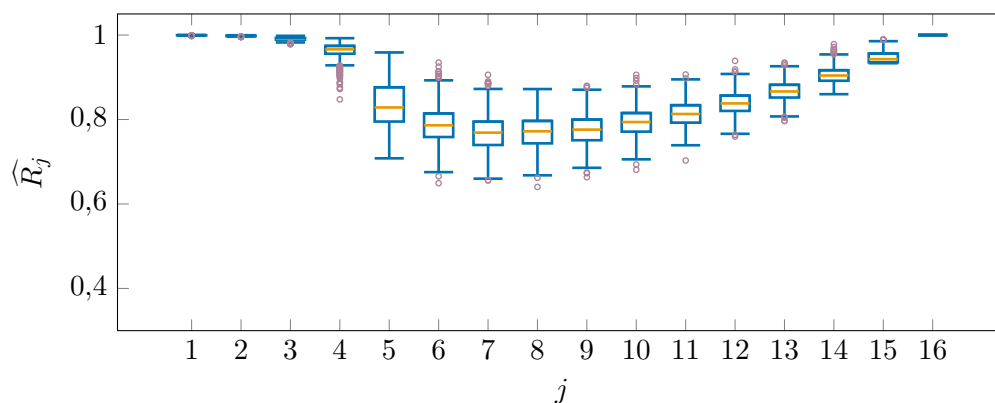


Figure E.1. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon *vena contracta* et les scores de la première composante principale comme variable de sortie.

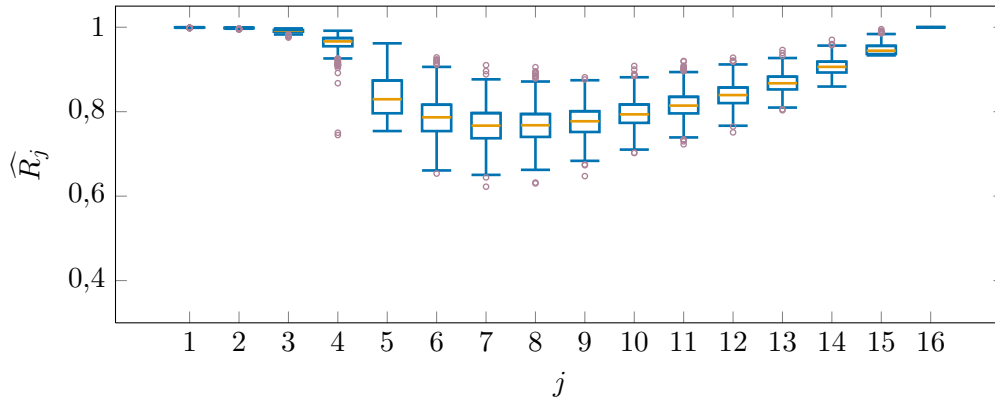


Figure E.2. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues avec l'échantillon *vena contracta* et les scores de la deuxième composante principale comme variable de sortie.

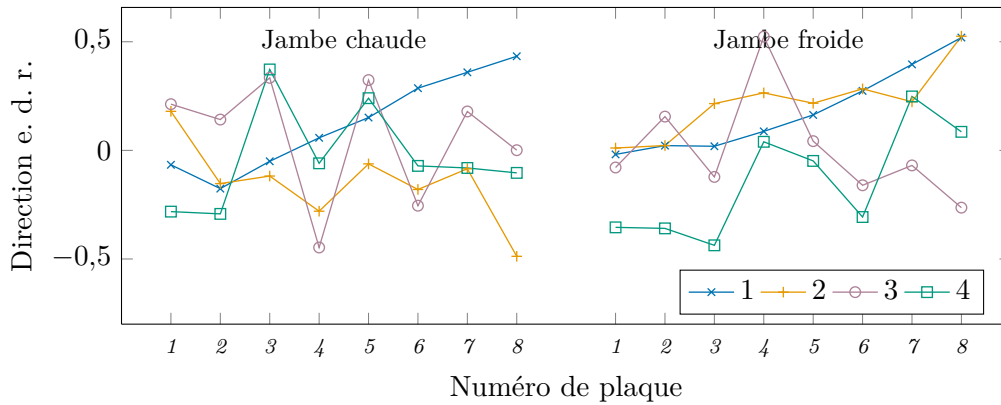


Figure E.3. : Quatre premières directions e. d. r. calculées avec l'échantillon *vena contracta* et les scores de la première composante principale comme variable de sortie.

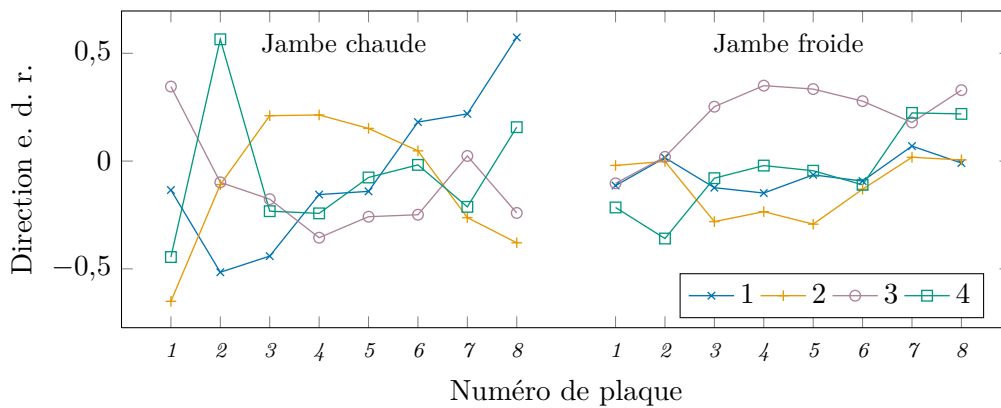


Figure E.4. : Quatre premières directions e. d. r. calculées avec l'échantillon *vena contracta* et les scores de la deuxième composante principale comme variable de sortie.

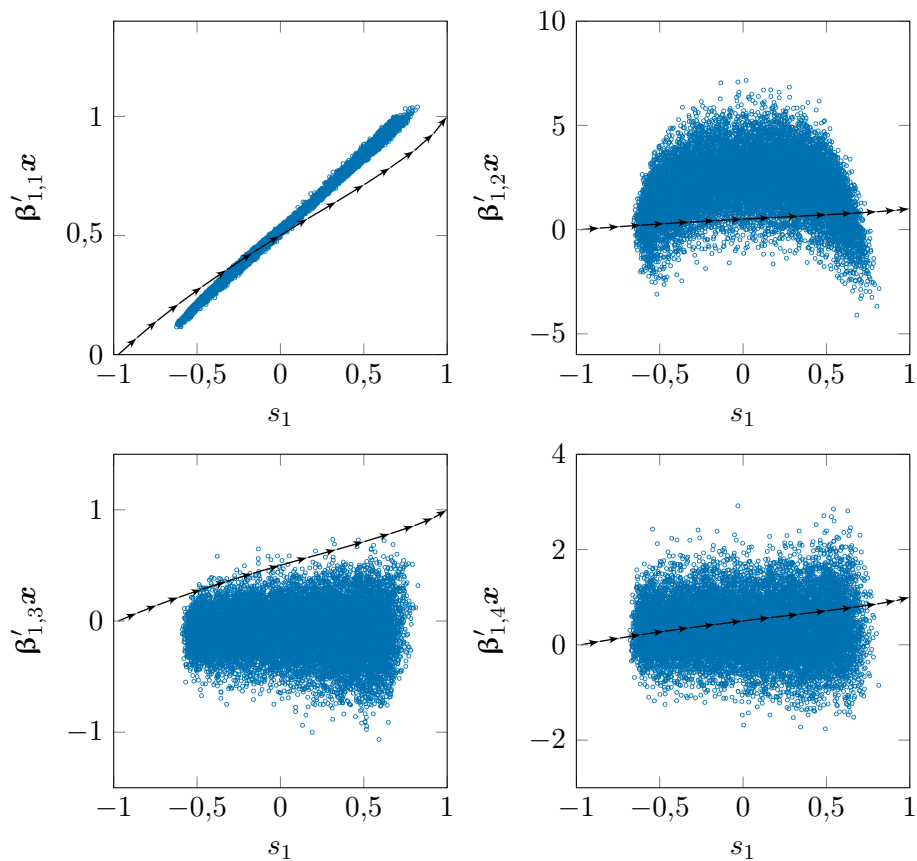


Figure E.5. : Nuages des projections des individus de l'échantillon *vena contracta* selon les directions e. d. r. obtenues avec les scores de la première composante principale, s_1 , comme variable de sortie, en fonction de s_1 .

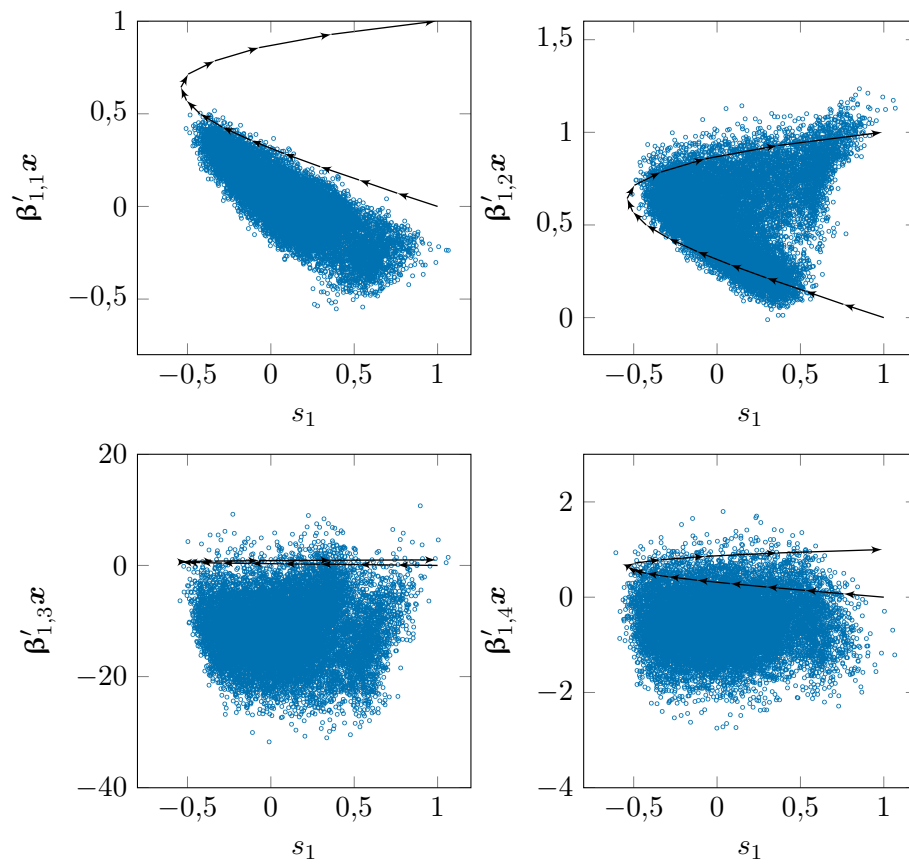


Figure E.6. : Nuages des projections des individus de l'échantillon *vena contracta* selon les directions e. d. r. obtenues avec les scores de la deuxième composante principale, s_2 , comme variable de sortie, en fonction de s_2 .

E.2. Découpage marginal groupé

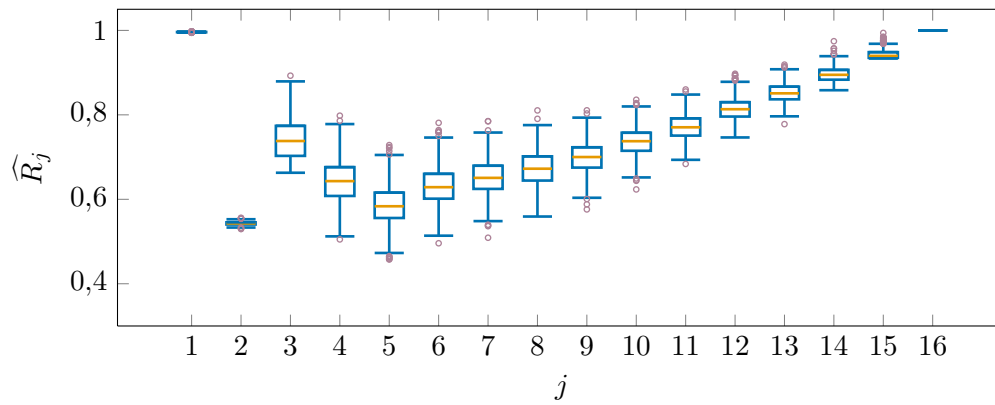


Figure E.7. : Diagrammes de Tukey des distributions des réplifications *bootstrap* de \widehat{R}_j obtenues par découpage marginal groupé de l'échantillon *vena contracta*.

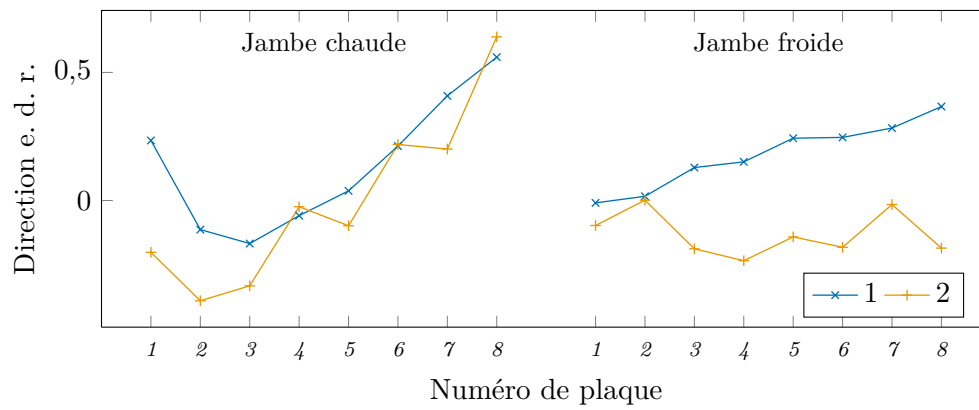


Figure E.8. : Deux premières directions e. d. r. obtenues par découpage marginal groupé de l'échantillon *vena contracta*.

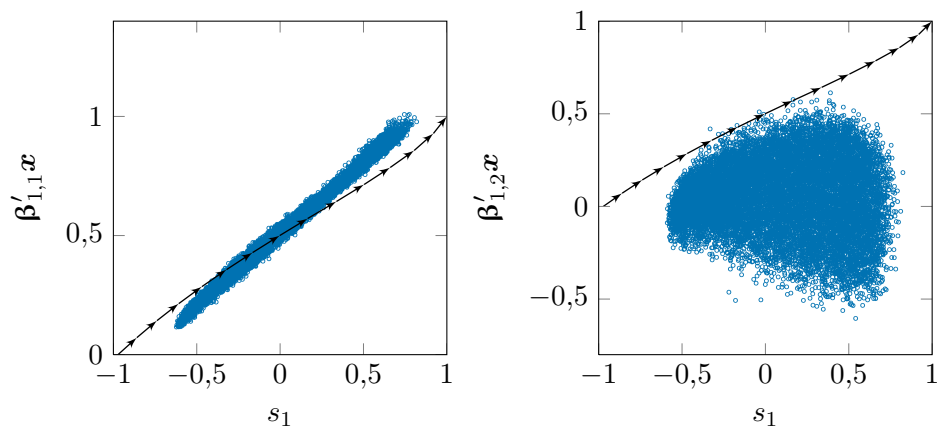


Figure E.9. : Nuages des projections des individus de l'échantillon *vena contracta* selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la première composante principale, s_1 .

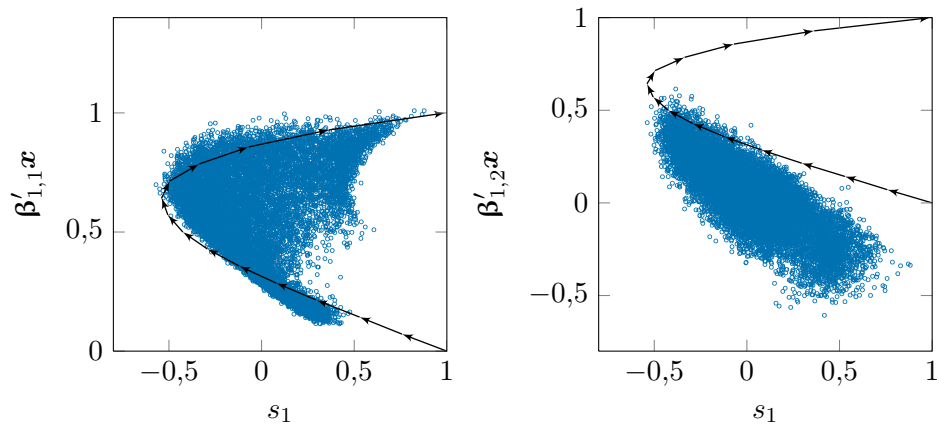


Figure E.10. : Nuages des projections des individus de l'échantillon *vena contracta* selon les directions e. d. r. obtenues par découpage marginal groupé en fonction des scores de la deuxième composante principale, s_2 .

F. Réponses de niveau gamme large mesurées

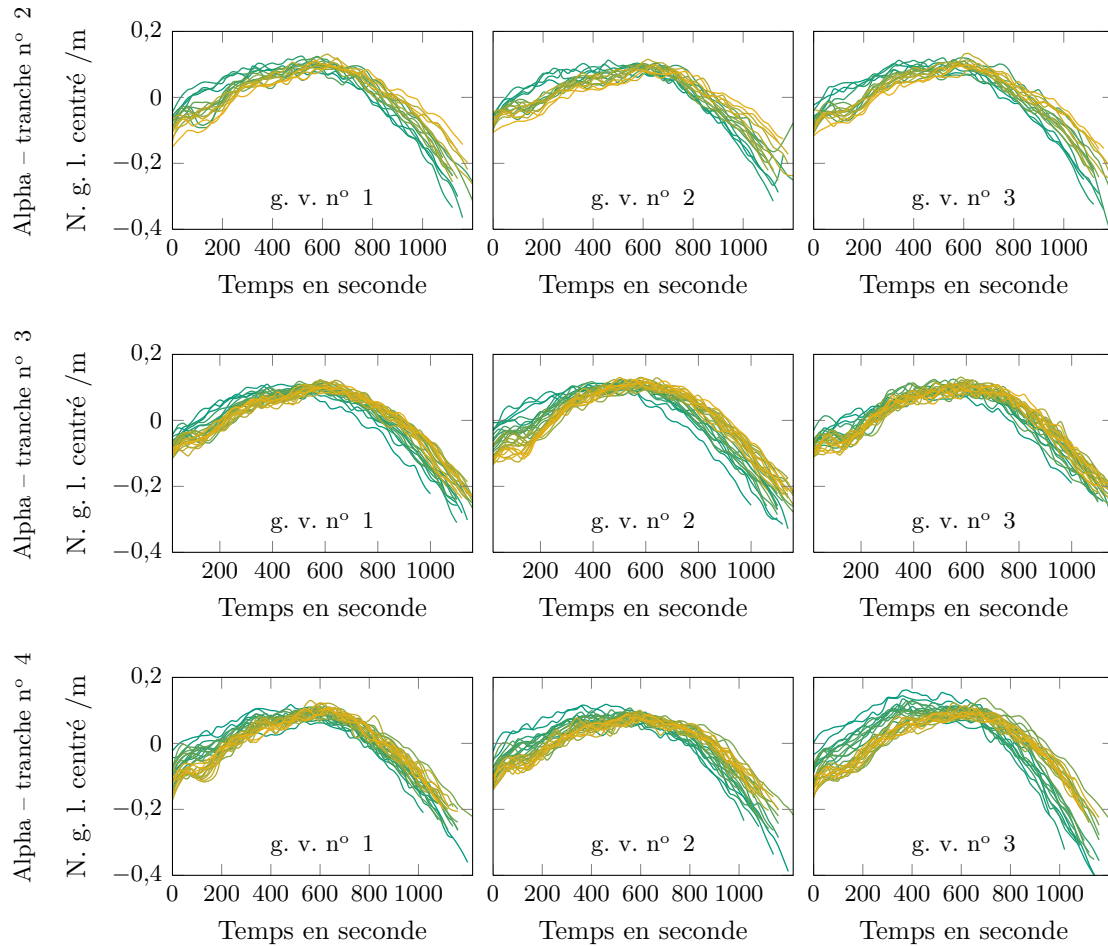


Figure F.1. : Faisceaux de courbes de réponse de n. g. l. mesurées à la centrale Alpha.

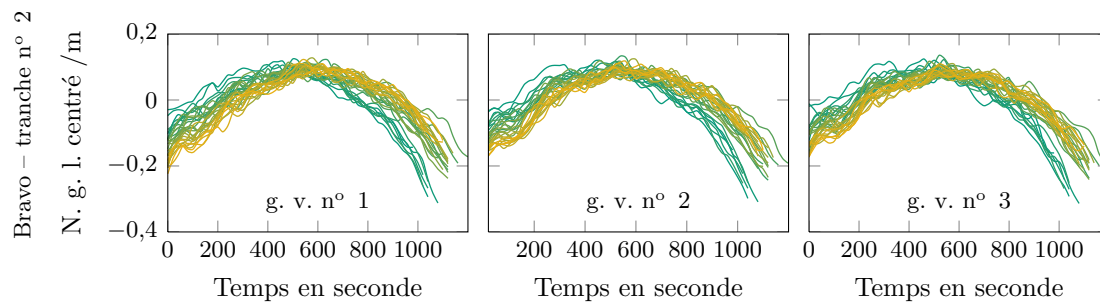


Figure F.2. : Faisceaux de courbes de réponse de n. g. l. mesurées à la centrale Bravo.

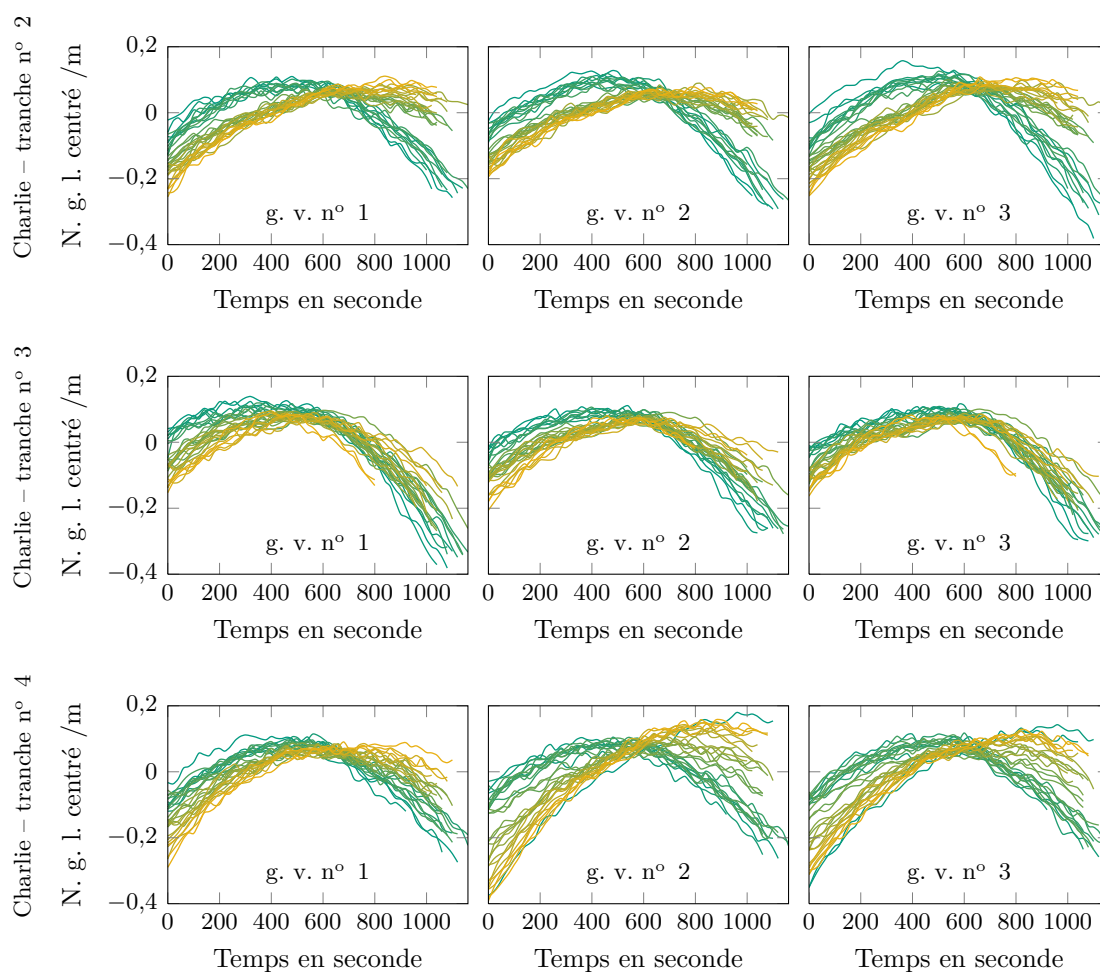


Figure F.3. : Faisceaux de courbes de réponse de n. g. l. mesurées à la centrale Charlie.

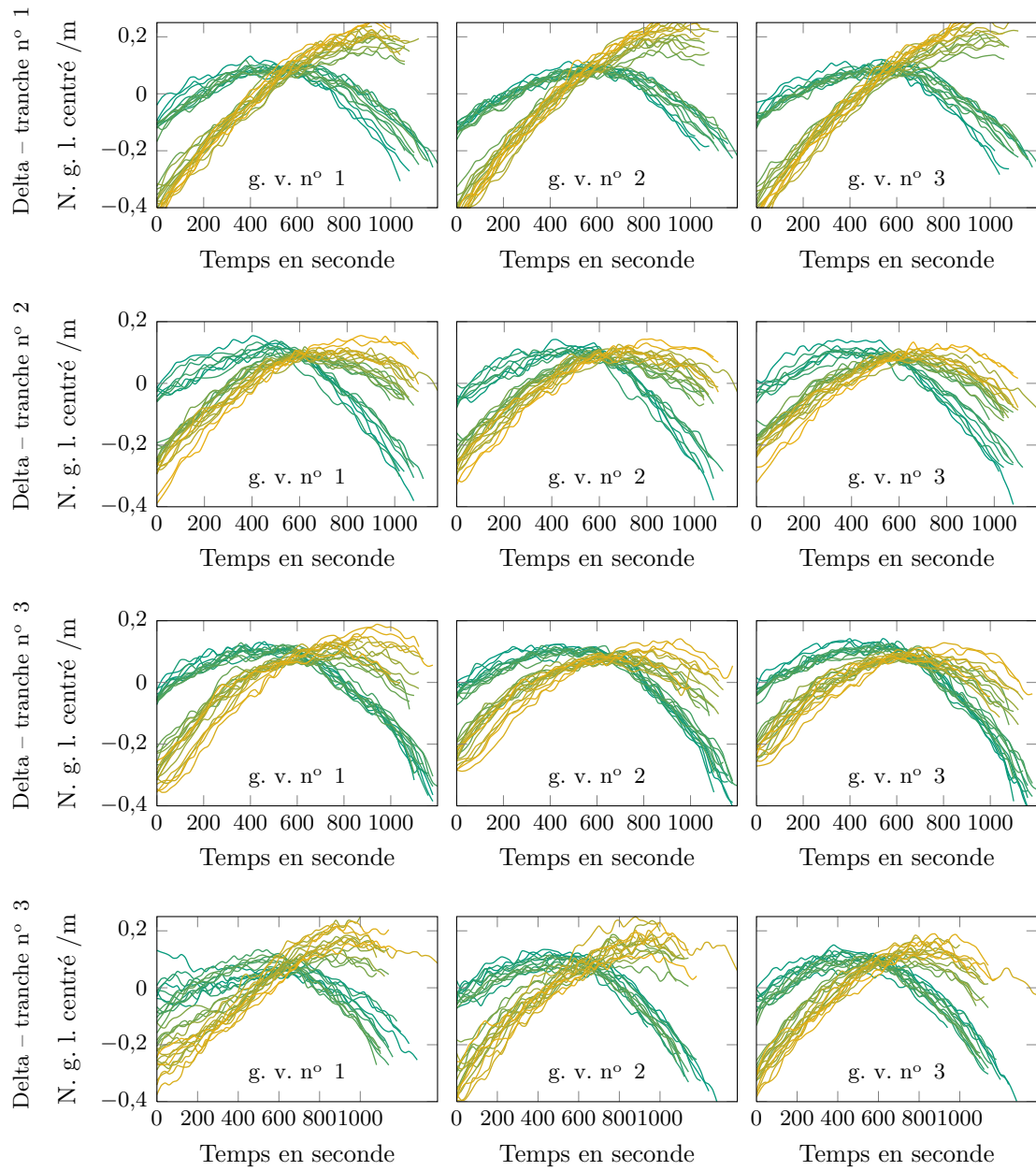


Figure F.4. : Faisceaux de courbes de réponse de n. g. l. mesurées à la centrale Delta.

G. Publications

Cette annexe regroupe deux articles écrits pendant la thèse. Le premier, intitulé *Towards a better understanding of clogged steam generators: a sensitivity analysis of dynamic thermohydraulic model output*¹, a été publié dans les actes de la 19^{ème} Conférence internationale de génie nucléaire (ICONE 19). Cette conférence, initialement programmée du 16 au 26 mai à Chiba, dans les environs de Tokyo, au Japon, a été déplacée les 24 et 25 octobre à Osaka, à cause de la catastrophe de Fukushima. Cet article présente les résultats de l'application de la méthode d'analyse de sensibilité présentée dans le chapitre 7 à la méthode de diagnostic à profil choisi du modèle. Avec ce modèle ne prenant pas en compte les débits transverses, les indices séquentiels indiquaient deux familles de variables d'entrée : l'une regroupant les demi-plaques en jambe chaude et l'autre les demi-plaques en jambe froide. Les indices de sensibilité compacts calculés avec les score de la seconde composante principale sont eux aussi très différents de ceux présentés dans le chapitre 10. On peut observer que l'incertitude sur les indices de Sobol' d'ordre 1 est nettement plus importante que dans le chapitre 10. Cela tient à la formule utilisée, moins performante fondée sur l'équation (7.21) page 119.

Le second article, intitulé *Sensitivity analysis and dimension reduction of a steam generator model for clogging diagnosis*², a été soumis au journal *Reliability engineering and system safety* le 12 septembre 2011 et accepté pour publication le 12 décembre 2012 après révision mineure. Cet article présente l'analyse de sensibilité et la réduction de la dimension d'entrée de la méthode de diagnostic à profil choisi à l'aide de la SIR, décrite dans le chapitre 9. Contrairement à ce qui est montré dans le chapitre 13, les espaces e. d. r. obtenus à partir des scores des deux premières composantes principales sont deux plans presque confondus. L'espace déterminé par découpage marginal groupé est lui aussi un plan. On peut en constituer une base orthogonale dont le premier vecteur ne fait intervenir que les taux de colmatage de la jambe chaude et le deuxième uniquement ceux de la jambe froide. Ces résultats soulignent l'ampleur des modifications apportées au modèle, notamment par la prise en compte des débits transverses.

1. Vers une meilleure compréhension des générateurs de vapeur colmatés : analyse de sensibilité de la sortie dynamique d'un modèle thermohydraulique

2. Analyse de sensibilité et réduction de dimension d'un modèle de générateur de vapeur pour le diagnostic du colmatage

TOWARDS A BETTER UNDERSTANDING OF CLOGGED STEAM GENERATORS: A SENSITIVITY ANALYSIS OF DYNAMIC THERMOHYDRAULIC MODEL OUTPUT

Sylvain Girard

EDF

Phone: +33130878050

sylvain-2.girard@edf.fr

Thomas Romary

Mines ParisTech

Phone: +33164694773

thomas.romary@mines-paritech.fr

Jean-Melaine Favennec

EDF

Phone: +33130878536

jean-melaine.favennec@edf.fr

Pascal Stabat

Mines ParisTech

Phone: +33140519152

pascal.stabat@mines-paristech.fr

Hans Wackernagel

Mines ParisTech

Phone: +33164694760

hans.wackernagel@mines-paristech.fr

Keywords: steam generator, clogging, thermohydraulic model, sensitivity analysis, principal component analysis.

1 INTRODUCTION

Internal parts of Steam Generators (SGs) foul up with iron oxides which causes Tube Support Plate (TSP) clogging which brings about concerns about safety and performance. Means to estimate TSP clogging are needed to optimize the costly maintenance operations. Previous work showed that the dynamic response to a power transient of the wide range level measurement contains informations about the clogging state of steam generators. A diagnosis method based on the comparison of plant WRL response with simulated responses is being developed by EDF. In order to achieve better understanding of the effect of clogging on SGs dynamic behaviour and to assess the potential of a diagnosis method based on the analysis of this behaviour through simulations, it is necessary to determine *i*) what features of the WRL response curves are characteristic of clogging and *ii*) the relative impact of each half TSP on these features.

2 MATERIALS AND METHODS

The objective of the present study is to analyse the sensitivity of the model output (WRL dynamic response) to its input parameters (TSP clogging ratios). A method based on the ANOVA-decomposition and a Monte Carlo computation scheme has been used to compute order 1 and total sensitivity indices for each half-TSP. As the model output is functional, a principal component analysis (PCA) on the samples used to compute sensitivity indices had to be carried out to reduce the dimensionality of the model output and compute 'compact' order 1 and total sensitivity indices for each major principal component. Finally, estimation variability was assessed by construction of BC_a bootstrap confidence intervals.

3 RESULTS

A sensitivity analysis of the dynamic output of a SG model provided a better understanding of the effect of TSP clogging and valuable insights for the development of diagnosis methods: *i*) sequential Sobol indices revealed different behaviours for each SG leg; *ii*) PCA highlighted two main features of the responses curves that characterize clogging; *iii*) Sobol indices on the PCA reduced output allowed to quantify the importance of each half TSP; *iv*) the convergence of the computations was checked with bootstrap confidence intervals.

REFERENCES

- G.E.B. Archer, A. Saltelli, and I.M. Sobol'. *Journal of Statistical Computation and Simulation*, 58:99–120, 1997.
- K. Campbell, M.D. McKay, and B.J. Williams. *SAMO 2004*. 2005.
- J.G. Collier and J.R. Thome. Oxford University Press, 1996.
- B. Efron and R. Tibshirani. Chapman & Hall, 1993.
- T. Homma and A. Saltelli. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.
- I.T. Jolliffe. Springer, 2002.
- Lamboni et al. *Field Crops Research*, 113:312–320, 2009.
- Midou et al. *ICONE18*. 2010.
- I.M. Sobol'. *English Transl.: MMCE*, 1(4), 1993.
- I.M. Sobol'. *Mathematics and Computers in Simulation*, 55:271–280, 2001.
- G. Yadigaroglu. *27th Short Courses - MCMF*. 2010.

TOWARDS A BETTER UNDERSTANDING OF CLOGGED STEAM GENERATORS: A SENSITIVITY ANALYSIS OF DYNAMIC THERMOHYDRAULIC MODEL OUTPUT

Sylvain Girard

EDF

Phone: +33130878050

sylvain-2.girard@edf.fr

Thomas Romary

Mines ParisTech

Phone: +33164694773

thomas.romary@mines-paritech.fr

Jean-Melaine Favennec

EDF

Phone: +33130878536

jean-melaine.favennec@edf.fr

Pascal Stabat

Mines ParisTech

Phone: +33140519152

pascal.stabat@mines-paristech.fr

Hans Wackernagel

Mines ParisTech

Phone: +33164694760

hans.wackernagel@mines-paristech.fr

Keywords: steam generator, clogging, thermohydraulic model, sensitivity analysis, principal component analysis.

ABSTRACT

Tube support plate clogging of steam generators affects their operating and requires frequent maintenance operations. A diagnosis method based on dynamic behaviour analysis is under development at EDF to provide means of optimisation of maintenance strategies. Previous work showed that the dynamic response to a power transient of the wide range level measurement contains informations about the clogging state of steam generators. The diagnosis method consists of comparisons of the measured dynamic response with simulations on a mono-dimensional dynamic steam generator model for various input clogging configurations. In order to assess the potential of this method, a sensitivity analysis has been conducted through a quasi-Monte Carlo scheme to compute sensitivity indices for each half tube support plate's clogging ratio. Sensitivity indices are usually defined for scalar model outputs. Principal component analysis has been used to determine a small subset of variables that condense the information about the shape of the response curves. Finally, estimation variability was assessed by construction of bootstrap confidence intervals. The results showed that half of the preselected input variables have negligible influence and allowed to rank the most important ones. Interactions of input variables have been estimated to exert only a small influence on the output. The effects of clogging on the steam generator dynamics has been characterised qualitatively and quantitatively.

1 INTRODUCTION

Steam Generators (SG) are affected by fouling of their internal elements. This causes clogging of the quatrefoil

holes of the Tube Support Plates (TSP) that can induce safety issues. Means to estimate TSP clogging are needed to optimize maintenance operations.

Clogging reduces the open cross sectional area of the TSPs and thus induces higher pressure drop. This alters the operating point of the SG as well as its dynamic behaviour. Previous studies (Midou et al., 2010) demonstrated that the shape of the Wide Range Level (WRL – the pressure difference measured between the steam dome and the bottom of the downcomer) response curve to a power transient is determined by the clogging state of the TSPs. A diagnosis method based on comparison between measured response curves and simulated ones for varying clogging states is being developed by EDF. A mono-dimensional SG model has been created with Modelica and the Dymola software for that purpose. It is able to simulate the SG behaviour during power transient phases. A 60 % power decrease with a 3 %. min^{-1} rate is performed on French nuclear reactors every three months which allows for frequent diagnosis. The input variables of the model are the clogging ratios of each half TSP, one for the hot and cold legs of the U-tube bundle. Clogging ratios are defined as the ratio of the closed area to the total area of the holes without any clogging. The output of the model consists of the 1200 values of WRL (1 per second) given the clogging ratios of the 16 half TSP.

In order to assess the performance of the method, it is necessary to determine *i*) what features of the WRL response curves are characteristic of clogging and *ii*) the relative impact of each half TSP on these features.

If it happens, for instance, that the clogging ratio of a given half TSP has a negligible influence, it is irrelevant to keep it as an input variable because little or no information about this ratio is contained into the WRL response curve. Discarding irrelevant variables reduces the dimension of the clogging state space that has to be sampled to produce the diagnosis.

The objective of the present study is to analyse the sensitivity of the model output (WRL dynamic response) to its input parameters (TSP clogging ratios). A sensitivity analysis has been conducted through a quasi-Monte Carlo scheme to compute sensitivity indices for each half TSP clogging ratio. As the model output is functional, a Principal Component Analysis (PCA) on the samples used to compute sensitivity indices had to be carried out to reduce the dimensionality of the model output and compute ‘compact’ order 1 and total sensitivity indices for each major principal component. Sequential WRL temporal indices have also proved to be meaningful. Finally, estimation variability was assessed by construction of bootstrap confidence intervals.

2 MATERIALS AND METHODS

2.1 Mono-dimensional steam generator model

The steam generator (SG) model has been developed with the Modelica language using the Dymola software. Modelica is an object-oriented language especially designed for modelling physical systems. It relies on a third party compiler and solver for simulation. These roles are being assumed here by the Dymola software. More specifically, we use a solver named DASSL, provided by Dymola, which is capable of solving differential algebraic equations.

The SG type studied here is the Westinghouse 51. EDF currently operates 48 of these; most of them being about 30 years old. A diagram representing the principal elements of a SG is given on figure 1.

The main elements of the model are :

- primary fluid flowing inside the U-tubes (single-phase flow) ;
- secondary fluid flowing outside the U-tubes (two-phase flow) ;
- thermal transfer between the two fluids and through tube interfaces ;
- two-phase singular pressure drops *e.g.* at the TSP quatrefoil holes ;
- steam-liquid separation devices ;
- feed water flow rate control system.

All these elements are mono-dimensional but the exchanger part is modelled as two channels: one for the *hot leg* (*i.e.* concurrent exchanging side, where primary fluid enters the SG) and one for the *cold leg* (*i.e.* countercurrent exchanging side, where primary fluid exits the SG). The exchanging channels are composed of 20 meshes evenly spaced.

The choice of mono-dimensionality and of the number of meshes is driven by the applications for which the model has been developed. On the one hand, it must be able to simulate the dynamic response of a SG precisely enough so that information about clogging ratios spatial distribution is not lost by averaging processes. On the other hand, computation time for simulation must not exceed five minutes so that it can be used in Monte Carlo methods.

Variations of the boundary conditions of the model allows to simulate a power transient of the plant. The power decrease used in the clogging diagnosis method is modelled by a linear variation of primary and secondary inlet

enthalpies and secondary outlet steam flow rate. The feed water flow rate is being determined by the control system.

Heat transfer coefficients are computed using correlations available in the literature. Pressure drops at the quatrefoil holes are computed using a specific correlation derived from experiments conducted at EDF R&D on a 1:4 scale mock-up of TSP and tubes.

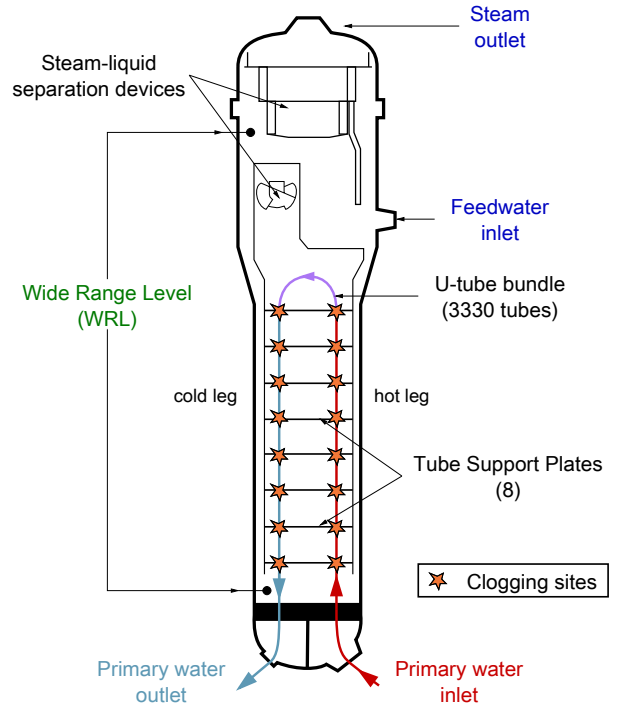


Fig. 1: Westinghouse type 51 SG principal elements

The two phase flow (secondary fluid) in the riser has been modelled using a simple mixture formulation known as the *homogeneous-equilibrium model* (Yadigaroglu, 2010) in which equal phase velocities and thermal equilibrium between the phases is assumed. This choice has been driven by the will to limit the number of unknown parameters to be inferred or calculated by correlations. Correlations available in the literature are usually designed for flows in tubes which can be quite different from a flow in a tube that contains a tube bundle and TSPs. Very little is known about the regime of the two-phase flow in a SG which makes the modelling of velocity slip tricky.

Another argument in favour of the homogeneous formulation is the high level of turbulence that the flow exhibits and the presence of the TSPs. Indeed, going through the quatrefoil holes tend to deconstruct and mix the flow while it slows down the steam more than the liquid. Visual observations made on the experiments mentioned above suggest that we are in presence of a slug-bubbly flow in the upper part of each interval between TSPs (under the TSP) while the flow is bubbly with very small steam structures from the top of the TSP up to approximately the middle of the interval. Flow regimes are described in the book by Collier and Thome (1996, Chap. 1). In the lower part of the intervals, the ratio of the volume of the bubbles to their surface is low and so is the ratio of the ‘floating’ force to frictions that is responsible of the higher velocity of steam. These rough observations were made on a freon-water mixture without heating for a few mass fractions.

When the results of these experiments are available, it will be possible to examine this question in much more details.

Implementing velocity slip in the SG model is being considered for future work. The impact of this amelioration on the results shown here are expected to be negligible.

2.2 Global sensitivity analysis

Sensitivity analysis in the broad sense studies how perturbations of model input variables generate perturbations on the output variables. Local sensitivity analysis is concerned by the effect on the output of small perturbations of the input variables around a given point, whereas global sensitivity analysis looks at the variability of the output on the whole variation domain. Here we focus exclusively on global sensitivity analysis because we want to obtain general information about how TSP clogging affects the WRL response without any particular clogging distribution in mind.

In this study, the components of the input vector are the 16 half TSP clogging ratios so $n = 16$ because Westinghouse 51 SGs have 8 TSP.

Clogging ratios were assumed to range from 0% to 65% as these values cover most real cases.

Let f be a function that represents the model, \mathbf{x} the input variables and u a scalar output variable of the model.

$$\begin{aligned} f &: \mathbb{I}^n \rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto u = f(\mathbf{x}). \end{aligned} \quad (1)$$

For the sake of clarity, in the following general derivation of sensitivity indices the input variables are assumed to take their values in $[0, 1]$ so \mathbb{I}^n denotes the n -dimensional unit hypercube. In (1), the model output consists of a single scalar. We will actually consider several of them and compute a set of sensitivity indices for each one independently.

Quite simple sensitivity indices can be defined when the model is linear or monotonous but this is seldom the case for complex thermohydraulic models. Variance based methods require a large number of simulations but they are model free. Among them is a method introduced by Sobol' (1993) based on the ANOVA-representation of the function f (ANOVA stands for ANalysis Of VAriance) and a Monte Carlo computation scheme. The next section relates the principal theoretical aspects of this method. The following derivation can be found in greater detail in (Sobol', 2001).

2.2.1 ANOVA-representation

Assuming f is an integrable function, let us consider the following decomposition

$$f(\mathbf{x}) = f_0 + \sum_{s=1}^n \sum_{i_1 < \dots < i_s} f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}), \quad (2)$$

where f_0 is a constant and the double sum means that there is a function $f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s})$ for every possible family of input variables: from $f_1(x_1)$ to $f_n(x_n)$ then all the $f_{ij}(x_i, x_j)$ with $1 \leq \dots < i < j \leq n$ and so on up to

$f_{1 \dots n}(x_1 \dots x_n)$. The number of terms in this decomposition is 2^n .

Sobol' (1993) has proved that if we impose to the summands of (2) the following condition

$$\int f_{i_1 \dots j}(x_i, \dots, x_j) dx_k = 0 \quad \text{for } k = i_1, \dots, i_s, \quad (3)$$

the decomposition exists and is unique. It is then called the ANOVA-representation of f . It follows that the summands in (2) are orthogonal and can be expressed as integrals of f .

2.2.2 Sensitivity indices

If we assume f to be square integrable, the $f_{i_1 \dots i_s}$ are also square integrable. Squaring and integrating (2) raises

$$\int f^2(\mathbf{x}) d\mathbf{x} - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} \int f_{i_1 \dots i_s}^2 dx_{i_1} \dots dx_{i_s}. \quad (4)$$

Now if \mathbf{x} is a random vector uniformly distributed in \mathbb{I}^n then $f(\mathbf{x})$ and $f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s})$ are random variables whose variances are respectively

$$D = \int f^2 d\mathbf{x} - f_0^2 \quad (5)$$

and

$$D_{i_1 \dots i_s} = \int f_{i_1 \dots i_s}^2 dx_{i_1} \dots dx_{i_s}, \quad (6)$$

and we have the following equality:

$$D = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} D_{i_1 \dots i_s}. \quad (7)$$

In other words, D measures the variability caused by variations of all the input variables while the $D_{i_1 \dots i_s}$ represent the variability caused by variations of variables in subsets $(x_{i_1}, \dots, x_{i_s})$. Equation (7) states, as expected, that the overall variability is the sum of the variabilities caused by all the possible subsets of input variables.

This leads to define the global sensitivity index of a subset of variables $(x_{i_1}, \dots, x_{i_s})$ by the following ratio

$$S_{i_1 \dots i_s} = \frac{D_{i_1 \dots i_s}}{D} \quad (8)$$

where s is called the order of the indice.

Order 1 sensitivity indices, $S_i = D_i/D$, measure the influence of each half TSP clogging ratio *alone*, that is without considering interactions. Reliable computation of sensitivity indices usually requires numerous model evaluations so it might be off-putting to try to estimate indices of order higher than 1 or 2. In our study we restrict ourselves to order 1 indices due to the high number of input variables. However, as our input variables are highly physically linked, we expect combined effects to play a significant role and completely ignoring them could be misleading.

As a palliative let us consider the variance of the subset of variables constituted by all the variables but one. Indeed, its complementary to the overall variance is the variance of all the subsets that comprise the one left aside

(Homma and Saltelli, 1996). For an input variable x_i , we have the following equality

$$D_i^{tot} = D - D_{\bar{i}} \quad (9)$$

where D_i^{tot} is the variance of all the subsets that comprises x_i and $D_{\bar{i}}$ is the variance of the subset constituted by all variables but x_i . Dividing equation (9) by the overall variance D leads to define *total sensitivity indices* for each input variable as follows

$$S_i^{tot} = \frac{D_i^{tot}}{D}. \quad (10)$$

The difference between total sensitivity indices and order 1 sensitivity indices for a given input variable measures the influence of all the combinations that involve the variable. Thus we can estimate how much of the overall variance is not accounted for by order 1 indices and determine what variables are responsible for it.

2.2.3 Computation scheme

Sobol' (1993) has demonstrated the following theorem

Theorem 1 (Variance of a subset of variables). *Let \mathbf{y} be a subset of variables of the complete family of input variables \mathbf{x} . Let \mathbf{z} be its complementary i.e. $\mathbf{x} = (\mathbf{y}, \mathbf{z})$. The variance $D_{\mathbf{y}}$ on the output of f induced by variations of the variables of \mathbf{y} is equal to*

$$D_{\mathbf{y}} = \int f(\mathbf{x})f(\mathbf{y}, \mathbf{z}') d\mathbf{x}d\mathbf{z}' - f_0^2. \quad (11)$$

The prime symbol indicates here that \mathbf{z}' comes from a different realisation of \mathbf{x} seen as a random variable.

By applying theorem 1 to the case of \mathbf{y} being a single variable x_i and \mathbf{z} being all the input variables but x_i , order 1 and total indices can be computed by estimating integrals : one for f_0 , one for D and two for each input variable.

This can be achieved through a Monte Carlo approach. Let $\boldsymbol{\xi} = (\xi_j)$ and $\boldsymbol{\xi}' = (\xi'_j)$ be two independent samples of size N uniformly distributed in \mathbb{I}^n . The following approximations hold for high enough N

$$\frac{1}{N} \sum_{j=1}^N f(\boldsymbol{\xi}_j) \rightarrow f_0, \quad \frac{1}{N} \sum_{j=1}^N f^2(\boldsymbol{\xi}_j) \rightarrow D + f_0^2, \quad (12)$$

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N f(\boldsymbol{\xi}_j) f(\xi_j^{(1)}, \dots, \xi_j^{(i-1)}, \xi_j^{(i)}, \xi_j^{(i+1)}, \dots, \xi_j^{(n)}) \\ \rightarrow D_i + f_0^2, \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N f(\boldsymbol{\xi}_j) f(\xi_j^{(1)}, \dots, \xi_j^{(i-1)}, \xi_j^{(i)}, \xi_j^{(i+1)}, \dots, \xi_j^{(n)}) \\ \rightarrow (D - D_i)^{tot} + f_0^2. \end{aligned} \quad (14)$$

In this study, random samples of size 1000 have been used. First computations made with samples of size 400 only displayed aberrations that indicated poor convergence.

For each TSP, two sets of 1000 clogging configurations have been built to generate the second members in each

term of the sums in equations (13) and (14). The first set is obtained from the first sample by replacing the clogging ratio of the TSP by the corresponding ratio from the second sample. The second set is obtained from the first sample by replacing all the clogging ratios but the one of the TSP by the corresponding ratio from the second sample. Simulations have also been run with the untouched first sample in order to compute f_0 and D and the first members of the terms of the sums in equations (13) and (14). Eventually, running additional 1000 simulations with the second sample allowed a cross-check by switching the role of the sets of clogging configurations. Thus, a total of 33000 simulations have been run.

LP_{τ} sequences (also known as Sobol sequences) were used to generate the random samples as they are stated to make the sum in equations (12), (13) and (14) converge faster (Archer et al., 1997).

The mean computation time for a simulation with the SG model is 3 minutes on a regular workstation. Hence the total computation time is around 2 month and a half. However, it is possible to launch several parallel threads without any further development so the computations had been carried through in about two weeks on a single workstation.

2.3 Monte Carlo algorithms with Dymola

2.3.1 Initialization of states

A dynamic model is a set of equations that describe the temporal evolution of states. Before running a simulation with Dymola, the states need to be initialized. In the case of this study, the sought after states are those of stable conditions for a given clogging configuration.

For ordinary differential equation models, imposing all the derivatives to equal 0 and setting the values of the boundary conditions and parameters (such as clogging ratios, referred to as *input variables* in section 2.2) is usually sufficient. For more complex models, it is needed to supply values of certain variables called *iteration variables*. When differential algebraic equations are involved, the initialization is likely to fail if the supplied values are too far from 'real' ones, that is values for which the system has a solution. There are about 500 iteration variables for the SG model and most of them need to be estimated with significant precision.

Once a set of appropriate values of the iteration variables has been found for a set of parameters, a convenient means to find values for another set of parameters is to use dynamic simulations. A first set of values for the zero clogging case has been found during the development of the model through a stepwise procedure: going from a cold state with no thermal exchanges and small flow rates and slowly increasing toward the nominal operating point of the SG. This state provides a starting point for the determination of others by dynamic simulation.

For instance, if the stable state corresponding to a homogeneous clogging of 5 % is needed, the procedure would be:

- i) load the zero clogging stable state values of iteration variables ;

- ii) set the clogging ratios to 5 %;
- iii) run a dynamic simulation with Dymola ;
- iv) monitor the simulation for stability (the WRL and circulation ratio are convenient indicators). Go back to step iii) if the states are not stabilized;
- v) save the final states values. They can now be used to initialize the model with all derivatives set to 0.

The new stable state can then be used for dynamic simulation of a power transient.

Changing the clogging configuration amounts to introducing a perturbation into the system. This results in a damped transient state until the new sought for stable state is reached. The more important is the perturbation, the longer the resulting oscillations will take to dampen. If the perturbation is too important, even the dynamic initialization can fail.

2.3.2 Optimization of the stable states research

The computation scheme described in section 2.2.3 implies that numerous stable state researches have to be performed. Therefore this procedure has to be optimized in order to minimize the computation cost of model evaluations and allow for samples of significant size. Optimizing the research of stable states consists in using as a departure point in step i a state that corresponds to a clogging configuration similar to the one sought for: this minimizes the perturbation so that the damping to stable state is as smooth and swift as possible.

Defining the distance between two input configurations of parameters is simple when there are only a few of them and when they affect the model linearly. However, the reasons that motivated the choice of the method presented in section 2.2 precisely make this task non trivial.

The question is to find a function τ that estimates the computation time needed to go from a clogging configuration $\theta^1 = (\theta_{h_1}^1, \dots, \theta_{h_8}^1, \theta_{c_1}^1, \dots, \theta_{c_8}^1)$ to another, θ^2 where the half-TSPs are numbered from bottom to top, the first 8 being on the hot leg and the last 8 on the cold leg.

A heuristic can be derived from simple physical considerations on how clogging affects the SG. Three empirical rules have been used here.

- i) The perturbation, and thus computation time, will increase with differences of individual clogging ratios.
- ii) As the secondary mixture goes through the SG its steam mass fraction increases and singular pressure losses are substantially more important on richer mixtures. Thus, differences of clogging ratios will have a stronger impact on computation time for transition when they are in the hot leg and the higher parts of the SG.
- iii) The TSPs singular pressure drops follow an approximately exponential relation with clogging ratios (this has been observed with the experiments referred at in section 2.1). Thus, computation time will increase, probably at an exponential pace, with the mean values of each couples of clogging ratios $(\theta_{x_i}^1, \theta_{x_i}^2)$.

Let us assume that the effects of all clogging ratio differences add up linearly (which is a crude hypothesis) and

that effects i) and iii) are multiplicative. The sought for function has the following form

$$\tau(\theta^1, \theta^2) = \sum_1^8 w_{hi} \varphi(|\theta_{h_i}^1 - \theta_{h_i}^2|) \gamma\left(\frac{\theta_{h_i}^1 + \theta_{h_i}^2}{2}\right) + w_{ci} \varphi(|\theta_{c_i}^1 - \theta_{c_i}^2|) \gamma\left(\frac{\theta_{c_i}^1 + \theta_{c_i}^2}{2}\right), \quad (15)$$

where the (w_{hi}) and (w_{ci}) are unknown weights that account for rule ii) and φ and γ are two unknown monotonically increasing functions that account respectively for rules i) and iii).

Finding valid values for all the 16 weights seems somewhat illusory considering the roughness of the approach. Using the linear approximation made before, a new set of weights can be defined by averaging groups of consecutive terms in equation (15).

The computation time for 50 transitions from a stable state to another were measured and the weights in (15) were estimated by multiple linear regression using these measurements as second members for various trial φ and γ functions. The validity of these functions was assessed by the maximum error of the computation times for transition estimated with the resulting τ function.

This way, a set of 6 weights could be found with satisfactory goodness of fit: three for the lower, middle and upper part of the SG for both the hot and cold legs. The φ and γ functions eventually selected are

$$\varphi : [0, 1] \rightarrow \mathbb{R} \quad \text{and} \quad \gamma : [0, 1] \rightarrow \mathbb{R} \\ x \mapsto e^{(7.0x)} - 1 \quad \quad \quad x \mapsto e^{(8.4x)} - 1, \quad (16)$$

and the weights values are

$$w_{h \text{ low}} = 0.09 \quad w_{h \text{ mid}} = 0.29 \quad w_{h \text{ up}} = 0.61 \\ w_{c \text{ low}} = 0.05 \quad w_{c \text{ mid}} = 0.08 \quad w_{c \text{ up}} = 0.11. \quad (17)$$

The maximum error is 120 s which is sufficient to ensure computation time of a few minutes. All the weights are positive and respect rule ii). The weights can be interpreted as some crude sensitivity indices for the 6 groups of TSPs.

2.3.3 Implementation of the Monte Carlo computing scheme

The τ function enables to estimate the computation time needed to go from one state to another. An initial set of some 100 stable states has been obtained and was used in the following algorithm:

- i) draw a clogging configuration from the samples constructed for the sensitivity analysis ;
- ii) using the τ function, find the ‘nearest’ stable state available in the stable states database;
- iii) if estimated computation time is above a limit t_{lim1} , go back to i);
- iv) do the transition and dynamic simulation. Store the simulated WRL response curve;
- v) if estimated computation time is above a limit t_{lim2} , add the stable state used for simulation to the database;

vi) go back to i until all the complete sample has been treated.

Multiple threads can be easily run at the same time. The two limits t_{lim1} and t_{lim2} have to be guessed at first and can be adapted as the algorithm goes by. For the algorithm to eventually deal with all the clogging configurations, it has been necessary to add a loosening condition regarding t_{lim1} .

2.4 Preprocessing model output

The aim of this study is to characterize the influence of each half TSP clogging ratios on the shape of the WRL response curve. It is thus necessary to select a set of output variables that correctly represent the shape of the curves and allows to compute sensitivity indices that are interpretable.

2.4.1 Sequential indices

The simplest possible set of output variables is constituted of the values taken by the WRL at each time step. This way $16 \times 1200 = 19200$ order 1 indices and 19200 total indices are obtained. The families of 1200 indices corresponding to a given half TSP will be then referred as the 16 *sequential indices* whose temporal variations during the transient we study.

The starting point of the response curves are quite far from each other because clogging alters the operating point of the SG. Sensors uncertainties and model representativeness issues make it unreliable to use this information in the model based diagnosis method (Midou et al., 2010). Differences in the shape of the curves are of small magnitude compared to distances between individual variation domains: the *static* effect (starting point) hides the *dynamic* effect (shape of the curves). Thus, sequential indices computed without any pre-treatment do not display any appreciable variation because most of the variance of the curves samples is actually due to differences of starting points. The alternative used both for the diagnosis method and the sensitivity analysis presented here is to subtract from each curve its temporal mean. This can be seen as ignoring the first term of the Taylor expansion of the underlying functions and studying the subsequent ones.

Subtracting also from each variable its mean allows to reduce the imprecision due to the fact that the value of f_0 is important compared to the summands of equations (13) and (14). This effect described in (Sobol', 2001) has been quantified when computing confidence intervals as described in section 2.5.

Sequential indices allow to follow the evolution of the relative importance of each input variable throughout the transient. However, considering the 1200 points of the response curves as independent variables is somehow missing the functional nature of the output which is precisely what we are trying to catch.

2.4.2 Principal component analysis

A classical approach to the problem of reducing the dimension of functional data is to use projection on carefully chosen subspaces. In this study, principal component analysis (PCA) has been chosen both because it does not require any inference to be made to construct the subspaces and, as our aim is to attribute shares of overall variances to input variables, using a variance based method for dimension reduction makes good sense. Campbell et al. (2005) and Lamboni et al. (2009) have made similar uses of PCA in a context of sensitivity analysis.

Let $\mathbf{WRL} = (WRL(t))$ be the vector of output variables considered in previous subsection. Principal components (PCs) of \mathbf{WRL} are ordered linear combinations of the $(WRL(t))$ that have maximum variance and are orthogonal to all the preceding PCs. For a set of p variables, up to p PCs can be found, each having less variance than the preceding. Here the number of variables equals the number of time steps so $p = 1200$. Let (\mathbf{pc}_i) be the family of principal components of \mathbf{WRL} . The $pc_i(t)$ are the coefficients in the linear combinations and are named the *loadings* of the i -th PC. Thus, \mathbf{WRL} can be written

$$WRL(t) = \sum_{i=1}^p pc_i(t) \langle \mathbf{WRL}, \mathbf{pc}_i \rangle . \quad (18)$$

By construction, the variance of the PCs are the eigenvalues of the empirical covariance matrix of the sample. They represent the share of the overall variance that is accounted for by each PC. In order to achieve significant dimension reduction, one hopes that the first PCs will account for most of the variance of the variables. In this latter case, projection on the subspace generated by the first few PCs provide a restricted number of variables (the corresponding loadings) that contain most of the information about variance of the whole. The construction of the PCs ensures that such a projection is optimal in terms of sum of variances. The book by Jolliffe (2002) provides PCs derivation, properties and applications in much details and examples.

The remark about subtraction of temporal means made in section 2.4.1 still holds for sensitivity indices computed on PCs loadings: without this preprocessing, the first PC represents more than 99 % of the overall variance because the differences of variation domains are of greater magnitude than variations in scale and shape of the curves.

Once again, subtraction of the means of each variables is helpful to achieve better precision.

A common practice when doing a PCA is to use centered reduced variables. This presents the advantage of normalizing the variables and avoiding the problem of discrepancies in units or range of variation. For instance, if a variable is in the range $[10^3, 10^4]$ while the others vary from 0 to 10, it will cause most of the variance of the set while varying, relatively, as much as the others.

A drawback of working with correlation instead of variances is that the PCs cannot be interpreted as easily. Indeed, when using variances, the PCs can be represented as perturbation of the mean of the sample which allows for simple interpretation of the loadings.

In this study, both approaches will be used and compared.

2.5 Bootstrap confidence intervals

It is important to estimate the accuracy of the computed sensitivity indices. One wants to know for instance, if the ranking of the indices can be trusted as it is or if groups of input variables should be considered. Also, the chosen computation scheme often induces aberrations, such as slightly negative indices or sums of indices that exceed one, due to slow convergence of the sums in equation (13) and (14); confidence intervals allow to decide if these irregularities can be overlooked or if larger sample sizes should be sought.

For each computed sensitivity index \hat{S} , we seek \hat{S}_{lo} and \hat{S}_{up} so that the two events $S < \hat{S}_{lo}$ and $S > \hat{S}_{up}$, where S is the real value of the sensitivity index, have both a probability of α . In this study, α is set to 0.05. As little is known about the distributions involved, bootstrap methods are particularly indicated as they are robust and distribution free.

The general idea behind bootstrap is to draw conclusions about a given estimator by using the empirical distribution upon which the estimator is based. Consider $\hat{S} = \phi(\zeta)$ the estimator of S computed using a sample $\zeta = (\xi, \xi')$ of size $2N$. A *bootstrap sample* ζ^* is obtained by drawing uniformly with replacement from ζ . For each bootstrap sample, a *bootstrap replication* $\hat{S}^* = \phi(\zeta^*)$ is computed in the same way as the estimator was from the original sample. It is now possible to draw inferences on the underlying distribution followed by \hat{S} by analysing the empirical distribution of the bootstrap replications.

Bootstrap *percentile* confidence intervals are constructed by taking the α and $1 - \alpha$ percentiles of the empirical distribution obtained after re-sampling. The *bias-corrected and accelerated* (shortened BC_a) intervals used in this study are derived from the percentiles intervals but include a correction of bias and an *acceleration* that compensates for variation of standard error of S with the value of S . These two corrections occur by shifting the percentiles chosen from the empirical distribution.

Details about bootstrap confidence intervals and their derivation can be found in the book by Efron and Tibshirani (1993). A rule of thumb for the number of necessary bootstrap samples to be used for the construction of confidence intervals is that it should be above 1000. However, it does not require any additional evaluation of the numerical model. Thus, BC_a confidence intervals have been computed for all sequential order 1 and total indices with 2000 bootstrap samples in less than one hour with a regular workstation.

3 RESULTS

3.1 Sequential indices

As an example, 7 WRL response curves of sample ξ after subtraction of their temporal means are displayed in figure 2. The curves are of various shapes and are crossing

each other. This is characteristic of a non-linear behaviour and is auspicious for the resolution of a diagnosis method based on curves comparison.

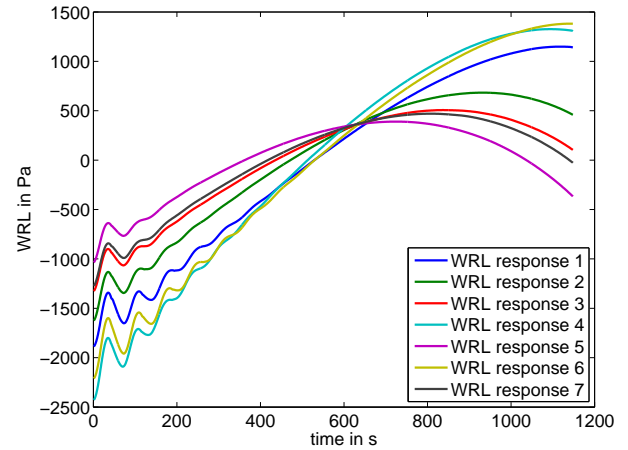


Fig. 2: 7 WRL response curves with their temporal means subtracted

The results obtained for both order 1 and total sensitivity indices are displayed in figures 3 and 4. Red lines correspond to the hot leg and blue ones to the cold leg and the colours darken as the TSPs are higher. The error bars represent the bounds of the BC_a confidence intervals. There are only a few of them for clarity but no discrepancies have been observed on the whole set.

On all presented graphics with a time abscissa, the end time is 1148 s instead of 1200 s. This reflects the fact that actual 60 % power transients performed on plants are usually a little steeper than $3 \%.s^{-1}$; 1148 s is the mean of the observed stopping times in (Midou et al., 2010).

Both sequential order 1 and total sets of indices display two sharp contrasting behaviours for each leg. The hierarchy of the indices is the same in all cases: the higher the TSP is positioned in the SG, the higher are the corresponding sensitivity indices. This is in agreement with physical intuition as stated in section 2.3.2.

As stated in section 2.2.2, the difference between total indices and order 1 indices measures the importance of combined effects.

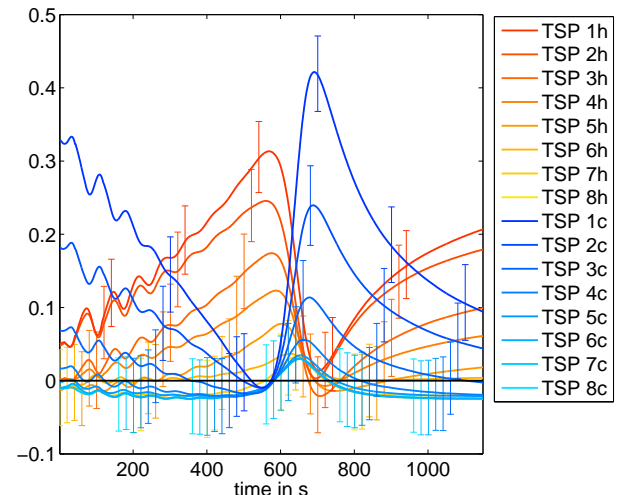


Fig. 3: Sequential order 1 sensitivity indices with BC_a error bars

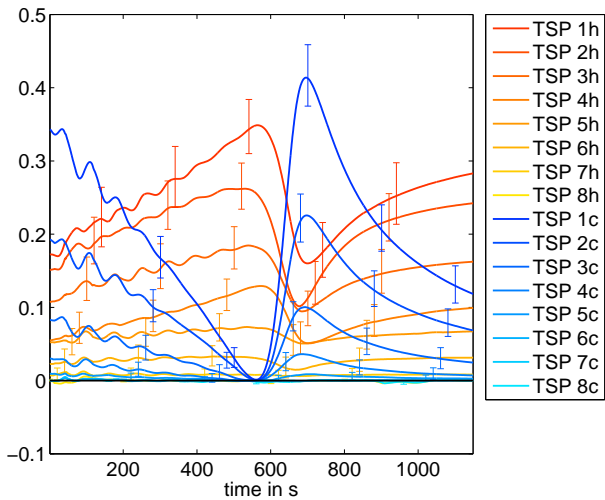


Fig. 4: Sequential total sensitivity indices with BC_a error bars

Four phases are discernible during the transient: *i*) from 0 to 250 s, top hot and cold TSPs prevail with some combined effects for hot leg; *ii*) from 250 s to 650 s, cold TSP have negligible impact and hot combined effects decrease; *iii*) from 650 s to 900 s, the three highest cold TSPs predominate while most of the hot impact is on the form of combined effects; *iv*) from 900 s until end time, hot leg order 1 effects catch up with combined effects while cold leg indices decline again.

There are a few negative order 1 indices but their error bars are roughly centred on the baseline. The corresponding total indices are estimated much more precisely and have near zero values so it is legitimate to consider these indices to be null. On the whole, BC_a intervals are shorter for total indices than for order 1 indices. Their lengths are approximately proportional to the value of the total indices while even very small indices are estimated with substantial uncertainties for order 1 indices.

The different behaviour of the hot and cold legs and the different phases observed during the transient will allow for refinement of the curves comparison methodology. An important result is that only the 4-5 highest TSP of the hot leg and the 3 highest of the cold leg have substantial effect during a significant time interval. From the SG safety analysis point of view, this would imply that lower TSP clogging has little or no effect on SG dynamic behaviour as far as the pressure field is concerned. They might however alter the mass repartition in the SG by modifying the mass fraction of the mixture. For the dynamic behaviour based diagnosis method, this means that no accurate clogging diagnosis can be made about the lower TSP if the WRL is chosen as informative variable.

3.2 Principal Components

A PCA has been conducted on the curves from the ξ sample and the two samples constructed by ‘resampling’ with ξ' all together after subtracting from each curves its temporal mean and from each variable $WRL(t)$ the mean of the WRL values at time t . The first ten PCs obtained with ‘raw’ variables are represented on figure 5 and the first ten PCs obtained after normalizing the variables are represented on 8. The first two PCs account for more than

99.9 % of the overall variance: the first (in dark blue) accounts for 95.54 % and the second (in green) for 4.40 %. Thus the choice of the first two loadings as output variables is straightforward. A resemblance with the order 1 and 2 Legendre polynomials can be noted. Legendre polynomials is a base of function commonly used for projection of functional data (Campbell et al., 2005). Subsequent PCs (in light blue) are rather disorderly and do not look like any general feature of the curves bundle apart from the oscillations in the beginning that have been identified as numerical artefacts. Discarding these PCs is a means to increase the signal-to-noise ratio in addition to radically reduce the dimension of the model output.

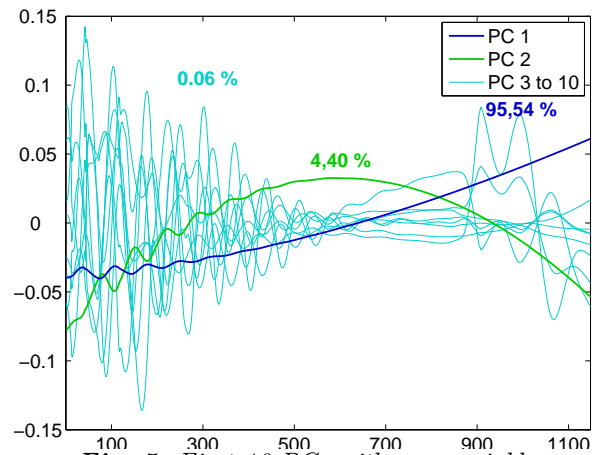


Fig. 5: First 10 PCs with raw variables

A convenient way to interpret the physical meaning of the PCs is to represent them as perturbations of the mean of the curves used for the PCA. On figure 6 the solid black line is the mean of the curves bundle while the ‘-’ and ‘+’ lines are respectively the mean minus or plus k times the first PC ; the value of k is arbitrarily chosen but identical for the two PCs. Figure 7, represent the effect of the second PC on the shape of the response curve with the same conventions. The first PC increases the global slope of the curve by swivelling it round a fixed point around time 650 s. The second PC increases the curvature by ‘bending’ the curve with two fixed points at times 250 s and 900 s. The fixed points of the perturbation of the mean by the first two PCs roughly corresponds to the time limits of the four phases described in section 3.1.

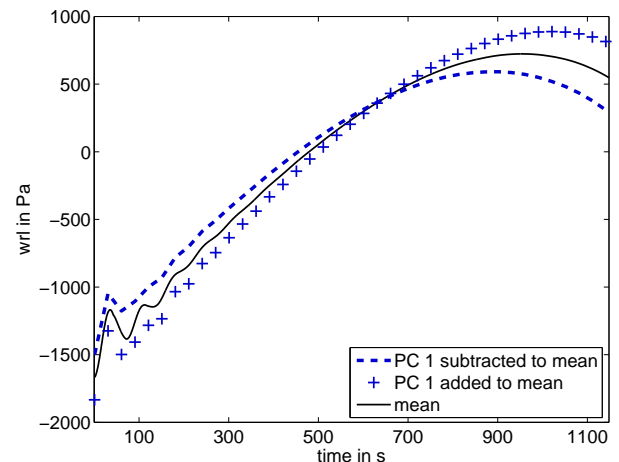


Fig. 6: Effect PC 1 with raw variables represented as perturbations of the mean

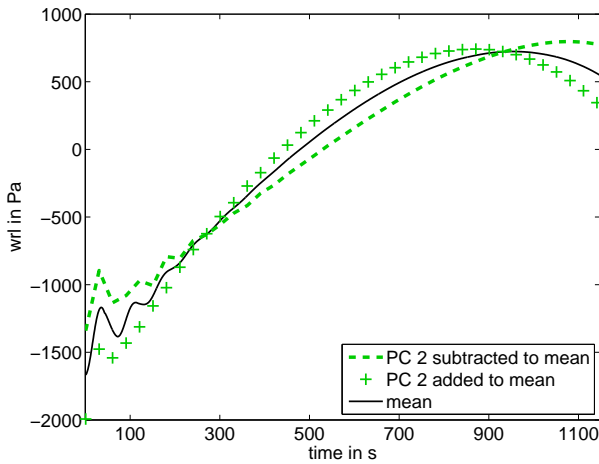


Fig. 7: Effect PC 2 with raw variables represented as perturbations of the mean

The result of the PCA conducted with centred reduced variables is presented on figure 8 with the same colour convention as in figure 5. Once again, the first two loadings are designated output variables as the first PC accounts for 83.83 % and the second for 16.11 %. Using standardized PCA does not enable for simple interpretation of the PC but the first two PCs still have distinct shapes (PC 1 and PC 2 are respectively antisymmetric and symmetric with respect to the axis $t = 650s$) while the subsequent ones are chaotic.

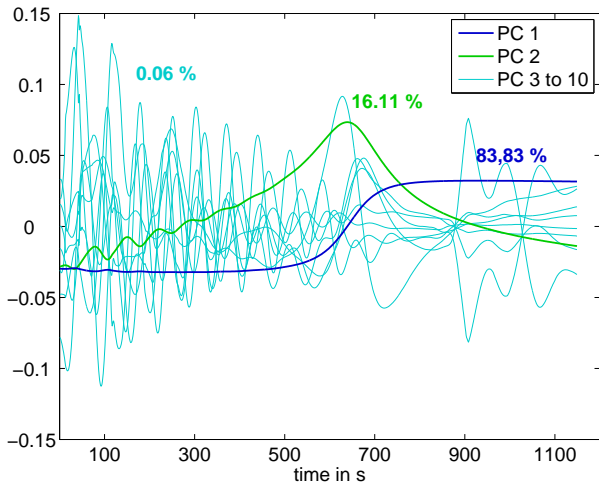


Fig. 8: First 10 PCs with centred reduced variables

For the sensitivity analysis, standardized PCA (figure 8) has been preferred to plain PCA (figure 5 in this study). The first reason for this choice, is that non-normalized PCs are subjected to a scale effect referred to in section 2.4.2. The first PC is the one that tunes the maximum reached in the terminal part of the curves (figure 6). Thus, its important eigenvalue (95.54 %) mostly reflects the variations of amplitude in the response curves which biases the estimation of the relative importance of the PCs. The second reason is that the relatively small eigenvalue of the second variance-PC (4.40 %) means that the corresponding effects to be estimated through sensitivity analysis would be also small and thus uncertainties would be more important: the approximations stated by equations (13) and (14) are erratic when the seconds members of the terms in the sums are too small.

3.3 Sensitivity indices of the reduced dimension model output

The results of the sensitivity analysis conducted on the first two loadings obtained through standardized PCA are displayed on figures 9 and 10 respectively. The TSPs are lined up on the ordinates in ascending order (TSP 1 is the lowest in the SG) with the hot leg TSP at the bottom and the cold leg ones above. For each TSP, two bars represent the corresponding order 1 and total indices: the bar above in lighter hue represents the total index while the other represents the order 1 index. The difference in length between the two bars indicates the combined effects. The error bars indicate the bounds of the BC_a confidence intervals.

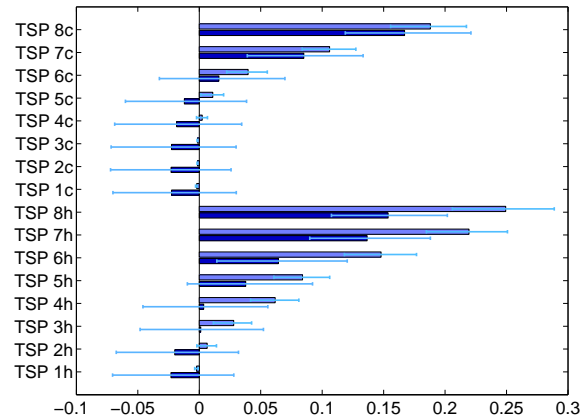


Fig. 9: PC 1 total and order 1 indices with BC_a error bars

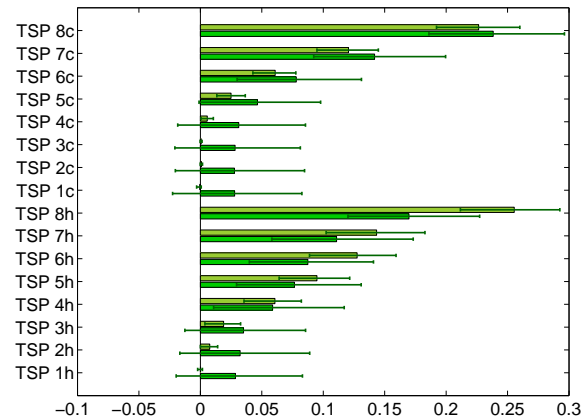


Fig. 10: PC 2 total and order 1 indices with BC_a error bars

As for the sequential indices, total indices have short confidence intervals and their length is proportional to the value of the indices while order 1 indices have longer confidence intervals of constant size. A few order 1 indices for PC 1 are negative. The same reasoning as in section 3.1 leads to consider them as null.

All the PC 2 couples of indices of cold leg TSPs and the first of hot leg TSPs are in wrong order: the order 1 indices exceed the total indices. However, the values of the total indices are always comprised in the confidence intervals of order 1 indices and the confidence intervals of the total indices are small. Thus, it seems legitimate to assume that interactions can be neglected for these variables and that order 1 indices actually equals total indices.

Total sensitivity indices for PC 1 are very low for TSPs 1 to 3 of the hot leg and 1 to 5 of the cold leg. Order 1

indices of the cold leg nearly equals total indices; as confidence intervals are quite large for the order 1 indices, combined effects are considered negligible. There are substantial combined effect TSP 6 to 8 of the hot leg and probably for TSP 4 and 5 but for these last two the important uncertainties on order 1 indices prevent from being positive. As for sequential indices, the higher in the SG a TSP is, the higher are its sensitivity indices.

The ratio of the two PC's eigenvalues is 5, so the most important effects are those related to PC 1 but those related to PC 2 are not negligible. Sensitivity indices of PC 2 display the same features as those of PC 1: same 'important' TSPs, same ranking of the indices and combined effects present only on the hot leg.

Eventually, the result of the sensitivity analysis suggests a reduction of the model input dimension from 16 to 9 because the lower TSPs total indices are very low. With the choice of the WRL measurement as informative variable, clogging ratios of the lower TSPs cannot be estimated by dynamic behaviour analysis. However, clogging mostly affects the SG by increasing the pressure losses and the WRL is an 'integration' of all the pressure losses along the height of the SG. Thus, it can be concluded that the impact safety of clogging of the lower TSP is very limited.

4 CONCLUSION

In order to achieve better understanding of the effect of clogging on SGs dynamic behaviour and to assess the potential of diagnosis method based on analysis of this behaviour through simulations, a sensitivity analysis has been conducted on a mono-dimensional SG model.

A method based on the ANOVA-decomposition and a Monte Carlo computation scheme was used to compute order 1 and total sensitivity indices for each half-TSP. Sequential indices were first computed and exhibited different behaviour for the hot and cold leg as well as four distinct phases during the transient. More compact indices could be computed by conducting a PCA beforehand: the first two PCs contain most of the available information and represent two modes of alteration of the SG behaviour by clogging. The model output could then be reduced to the first two sets of loadings and two sets of sensitivity indices have been computed.

The sensitivity indices indicated that lower TSPs clogging had little or no influence on the WRL response of the SG. This means that no diagnosis can be made about it through the dynamic behaviour analysis method with the WRL as informative variable. However, this also suggests that the influence of the lower TSP clogging on the overall dynamic behaviour of the SG is of second order as the WRL is an accurate indicator of it.

BC_a confidence intervals were used to check the validity of the conclusions drawn from the computed sensitivity indices. They allowed to validate the methodology presented in this study.

There are substantial combined effects for the upper hot leg TSPs only. Thus, although being not exhaustive, the present study can be considered as a thorough analysis of the effect of TSP clogging on SG dynamic.

A new version of the SG model is currently under development at EDF. Indeed, an important hypothesis is being made when modelling the SG exchanging part by two disconnected channels: this amounts to neglecting all transport phenomena between the two legs. In the new version, transport phenomena between these two channels are taken into account by connections whose pressure drop coefficients are being tuned so that corresponding flow rates at nominal power are equal to those computed with the certified EDF 3D CFD code THYC. These transversal connections are expected to impact the sensitivity indices, probably by increasing the amount of combined effects between the hot and cold legs.

REFERENCES

- G.E.B. Archer, A. Saltelli, and I.M. Sobol'. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58:99–120, 1997.
- K. Campbell, M.D. McKay, and B.J. Williams. Sensitivity analysis when model outputs are functions. In K.M. Hanson and F.M. Hemez, editors, *Sensitivity Analysis of Model Output (SAMO 2004)*, pages 81–89. Los Alamos National Laboratory, March 2005.
- J.G. Collier and J.R. Thome. *Convective Boiling and Condensation*. Oxford University Press, 1996.
- B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.
- T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.
- I.T. Jolliffe. *Principal Component Analysis, 2nd edition*. Springer, 2002.
- M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, and H. Monod. Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113:312–320, 2009.
- M. Midou, J. Ninet, A. Girard, and J.-M. Favennec. Estimation of SG TSP blockage: Innovative monitoring through dynamic behavior analysis. In *18th International Conference on Nuclear Engineering ICONE18*, may 2010.
- I.M. Sobol'. Sensitivity estimates for nonlinear mathematical models, in *Matem. Modelirovanie*, 2 (1)(1990) 112 – 118. *English Transl.: MMCE*, 1(4), 1993.
- I.M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55:271–280, 2001.
- G. Yadigaroglu. Basic models for two-phase flows. In *27th Short Courses - Modelling and Computation of Multiphase Flows*. Swiss Federal Institute of Technologie (ETH), February 2010.

Sensitivity Analysis and Dimension Reduction of a Steam Generator Model for Clogging Diagnosis

Sylvain Girard^{a,b,c,*}, Thomas Romary^b, Jean-Melaine Favennec^a,
Pascal Stabat^c, Hans Wackernagel^b

^a*EDF R&D, STEP P1C, 6 Quai Watier, 78400, Chatou, France*

^b*Mines ParisTech, Centre de Géosciences/Équipe de Géostatistique,
35 avenue Saint Honoré, 77305, Fontainebleau, France*

^c*Mines ParisTech, Centre Énergétique et Procédés, 60 Boulevard Saint Michel, 75006,
Paris, France*

Abstract

Nuclear steam generators are subject to clogging of their internal parts which causes safety issues. Diagnosis methodologies are needed to optimize maintenance operations. Clogging alters the dynamic behaviour of steam generators and particularly the response of the wide range level (WRL – a pressure measurement) to power transients. A numerical model of this phenomenon has previously been developed. Its input variables describe the spatial distribution of clogging and its output is a discretization of the WRL dynamic response.

The objective of the present study is to characterize the information about the clogging state of a steam generator that can be inferred from the observation of its WRL response. A methodology based on several statistical techniques is implemented to answer that question. Principal component analysis reveals that clogging alters the WRL response mainly in two distinct ways. Accordingly, the output can be summarized into a vector of dimension 2. A sensitivity analysis is carried out to rank the input variables by magnitude of influence. It has shown that they can be divided into two

*Corresponding author. Tel.: +33130878050.

Email addresses: sylvaingirard@gmail.com (Sylvain Girard),
thomas.romary@mines-paristech.fr (Thomas Romary),
jean-melaine.favennec@edf.fr (Jean-Melaine Favennec),
pascal.stabat@mines-paristech.fr (Pascal Stabat),
hans.wackernagel@mines-paristech.fr (Hans Wackernagel)

groups corresponding to the two sides of the steam generator. Finally, sliced inverse regression is used to reduce the input dimension from 16 to 2. A sampling issue that arises when the input dimension is high is addressed.

The simplification of the original problem yields a diagnosis methodology based on response surface techniques.

Keywords: sensitivity analysis, principal component analysis (PCA), sliced inverse regression (SIR), bootstrap, steam generators, clogging.

1. Introduction

Pressurized light water nuclear power plants mainly consist of two separated water loops that exchange heat. The water from the primary loop goes first through the reactor where it is heated by the nuclear reaction and then through heat exchangers called steam generators (SGs) where it transfers heat to the water of the secondary loop. Steam exits the SGs by their upper opening and then flows through the turbines. A SG consists of a cylindrical tank (approx. 20 m high and 3m wide) that contains the secondary steam-liquid mixture. The primary water enters the SG at its bottom and goes through a bundle of U shaped tubes. Eight circular plates called tube support plates (TSPs) maintain the tube bundle. The tubes fit in circular holes drilled in the TSPs. These holes are surrounded by additional quatrefoil holes to let the secondary steam-liquid mixture flow through. A SG diagram can be found in figure 1.

SGs internal elements foul with iron oxides carried by the secondary feed-water. This causes clogging of the quatrefoil holes that induces safety issues. Means to estimate TSP clogging are needed to optimize maintenance operations. The pressure difference measured between the steam dome and the bottom of the SG is called the wide range level (WRL). Previous studies [1, 2] has shown that the shape of the WRL response curve to a power transient is altered by the clogging state of the TSPs and derived a diagnosis method that utilizing this link. The principle of the method is to compare measured response curves with simulations using with a mono-dimensional SG model. To assess the method's potential and make it reliable, it is necessary to characterise *how much information about the clogging state can be inferred from the WRL response*. This issue breaks down into three closely related questions:

- how does TSP clogging affect the shape of the WRL response?

- Are these effects different in nature and magnitude depending on the location inside the SG?
- What is the simplest formulation of input and output variables that captures these effects?

The methodology presented here to answer these questions relies on computer intensive statistical methods. As the CPU time for a transient simulation with the 1D SG model is around 5 min, large samples of response curves corresponding to different clogging configurations can be generated.

Sensitivity analysis [3] and principal component analysis (PCA) [4] have been carried out to address the first two questions and the simplification of the output. The results suggested the use of a dimension reduction technique called sliced inverse regression (SIR) [5] to simplify the input. Along the process, bootstrap techniques were used to assess the robustness of the results and help with the interpretation. The SG numerical model and the statistical method that have been used are described in section 2. The results are presented and discussed in section 3.

2. Model and methods

2.1. Mono-dimensional steam generator model

The SG type examined here is the Westinghouse 51. EDF currently operates 48 of these, most of them being about 30 years old. A diagram representing the principal elements of a SG is given in figure 1. The SG model has been developed with the Modelica language using the Dymola software. Modelica is an object-oriented language especially designed for modelling physical systems [6, 7]. It relies on a third party compiler and solver for simulation. Here, these roles were assumed by the Dymola software [8]. More specifically, Dymola version 6.0d and a solver named DASSL [9] have been used. It is capable of solving differential algebraic equations. The main elements of the model are:

- primary fluid flow inside the U-tubes (single-phase flow);
- secondary fluid flow outside the U-tubes (two-phase flow);
- thermal transfer between the two fluids and through tube interfaces;
- two-phase singular pressure drops *e.g.* at the TSP quatrefoil holes;

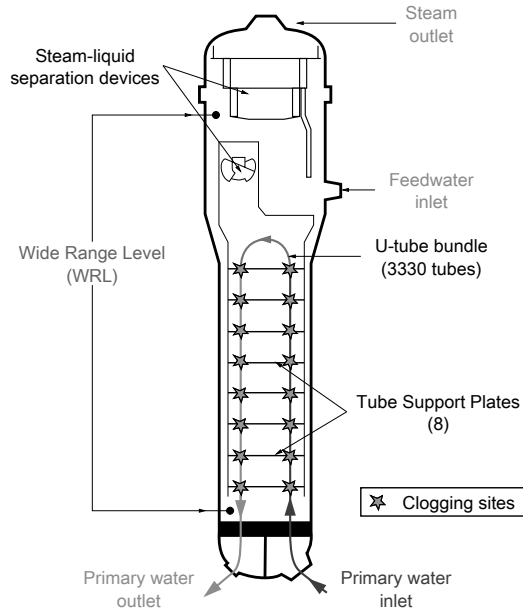


Figure 1: Westinghouse type 51 steam generator.

- steam-liquid separation devices;
- feed water flow rate control system.

All these elements are mono-dimensional but the exchanger part is modelled as two channels: one for the *hot leg* (*i.e.* concurrent exchanging side, where the primary fluid enters the SG) and one for the *cold leg* (*i.e.* counter-current exchanging side, where the primary fluid exits the SG). The exchanging channels are composed of 20 evenly spaced meshes. The choice of mono-dimensionality and of the number of meshes is driven by the applications for which the model has been developed. On the one hand, it must be able to simulate the dynamic response of a SG precisely enough so that information about clogging spatial distribution is not lost by averaging processes. On the other hand, computation time for simulation must be kept low so that computer intensive methods can be used. Typical Monte Carlo methods require hundreds if not thousands of model runs. For instance, the method of Sobol' for sensitivity analysis can often require an order of mag-

nitude of 10^3 simulations per input variable to converge [3]. The number of simulations needed to obtain robust results with SIR is dependent on the model complexity. In the present case, the results were still fluctuating when increasing the sample size from $3 \cdot 10^3$ to $5 \cdot 10^3$ while they were stable for a sample size of 10^4 . In order to be able to use such methods with a reasonable computation time, the computation time had to be under approximately 10 min using regular workstations. Additional details about the model can be found in [2].

2.1.1. Model output definition

A power transient is simulated by varying the model boundary conditions. The transient used in the clogging diagnosis method is a roughly linear power decrease from nominal power to 40% of nominal power in an average time of 1148 s. It is modelled by a linear variation of primary inlet enthalpy and secondary outlet steam flow rate. The feed water flow rate is being determined by the control system. The model output is a vector, \mathbf{w} , of dimension 1148. Its coordinates are the values of the WRL at each 1 s time step.

2.1.2. Model input definition

There are 8 TSPs in the SGs under study and two 1D channels so the vector describing the clogging state, \mathbf{x} , is of dimension 16. Each of its coordinates is a *clogging ratio* associated to a half-TSP. Clogging ratios are defined as the ratio of the blocked area to the total area of the holes without clogging:

$$x_i = \frac{(\text{clogged area of half-TSP})_i}{(\text{total holes area of half-TSP})_i} . \quad (1)$$

Clogging affects the WRL response by increasing the singular pressure drop at TSP crossings. In the model, the corresponding pressure drop coefficients depend on the clogging ratios through a function derived from experiments conducted on a 1:4 scale mock-up of TSPs and tubes [10].

2.1.3. Preliminary analysis

The singular pressure drop at a TSP crossing increases with the clogging ratio and steam fraction and decreases with the pressure of the steam-liquid mixture. The pressure is nearly the same in the two legs and it decreases as the secondary mixture rises inside the SG. The steam fraction equals zero at the bottom of the SG (liquid alone) and increases as the fluid rises and

gets heated by the tubes. Its increase is sharper on the hot leg. From this, clogging is expected to have a greater impact in the hot leg than in the cold leg and in the higher parts of the SG than in the lower.

2.2. Sensitivity analysis of a functional output model

Sensitivity analysis studies how perturbations of the model input variables generate perturbations on its output variables. Here, general information about how TSP clogging affects the WRL response is sought without any particular clogging configuration in mind. Hence, a *global* sensitivity analysis method [3] has been used. A brief review of existing methods is given in section 2.2.1. The selected method consists in estimating sensitivity indices called Sobol' indices through a Monte Carlo computation scheme. These indices are presented in section 2.2.2. A preprocessing issue is addressed in section 2.2.3. Sensitivity analysis is usually applied to univariate or small dimensional output models, hence reduction of the output dimension was needed. Section 2.2.4 details how a convenient projection basis can be constructed using PCA. Section 2.2.5 describes how the validity of the results can be assessed with bootstrap confidence intervals.

2.2.1. Choice of a sensitivity analysis method

Three criteria are to be considered when choosing a global sensitivity analysis method: the number of input variables, the hypotheses that can be made about the regularity of the model (linearity, monotony, integrability) and the computation time for one model evaluation. The sample size needed to obtain robust estimates depends on the model and the chosen method and it usually grows with the number of input variables. When the input variables are too numerous ($\gtrsim 20$), it can be necessary to select beforehand those that are most influential with a screening method such as the Morris method [11].

High computational cost can be another potential hurdle to overcome. Indeed, when the time needed to perform one model evaluation is more than a few hours, it becomes difficult to achieve a substantial sample size. This issue can be addressed through meta-modelling. A meta-model, also known as surrogate model, response surface model or emulator, is a simple function that approximates the initial model response over the domain of interest and has a negligible computation cost. Several approaches have been proposed to build such approximations. Recent techniques include gaussian process

metamodelling [12] and polynomial chaos expansion [13, 14] for the estimation of Sobol' indices.

The objective of screening and metamodelling is to allow to build a sample which size is sufficient for the estimation of sensitivities considering the number of input variables. Once this has been achieved, several estimation techniques are available. The most simple global sensitivity analysis methods are based on linear regression: either standard regression coefficients or partial regression coefficients can provide sensitivity estimates if the model is assumed linear. The linearity hypothesis can be replaced by a less restraining monotony hypothesis by using a rank transformation. The WRL response is unlikely to be linear if only for the highly nonlinear link function between clogging ratio and pressure drop [10].

Many other less restrictive methods rely on the analysis of the empirical distribution of the output based on a sample. For instance, contribution to the sample mean and variance plots [15] allow for graphical estimations of sensitivities without the need for large sample sets. Some recent methods do not rely on moments estimation but instead on global quantitative comparison of histograms [16, 17]. However, the most widespread approach consists in assigning shares of the overall output variance to each input variable. The method of Sobol' [18, 19] and the *Fourier amplitude sensitivity test* (FAST) [20] are two of these so-called *variance based methods*. The former is based on a functional expansion of the model as a sum of terms whose variances can be estimated by a Monte Carlo method. At least a few hundred model evaluations per input variable are generally necessary but the estimates are robust and the only required hypothesis is square integrability. The FAST method is based on a multidimensional Fourier transform. Estimation of sensitivities with the FAST method tends to require smaller samples than with the method of Sobol'. However its application is less straightforward and potential biases require advanced treatment [21].

The number of input variables of the SG model, 16, and its computational cost of a few minutes per model run are small enough for not requiring either screening or metamodelling. The model response is smooth (see figure 2) because it derives from a system of differential equations involving simple functions without singularities. Thus the integrability condition needed for the Sobol' method is likely to be honoured. As will be stated in more detail in the last paragraph of the next section, the method of Sobol' requires $(n + 2) \times N$ model estimates where N is the sample size and n the number of input variables. Let assume that 10 series of simulations can be run

simultaneously, which would typically require only two regular workstations. A conservative upper bound of 10 minutes for the model computational cost leads to a reasonable total computation time of approximately 12 days for a sample size of 10^3 . Thanks to the newer formulae proposed by Saltelli et al. [22], a substantial increase in rate of convergence can be achieved and this sample size is usually sufficient for reliable estimations. These considerations led to choose the method of Sobol' with a quasi-Monte Carlo scheme that will be discussed in the following section.

2.2.2. Sobol' indices

Let us first derive Sobol' indices for a univariate output model. Let f be a function that represents the model, \mathbf{x} the input vector of size n and y the scalar output.

$$\begin{aligned} f &: \mathbb{I}^n \rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto y = f(\mathbf{x}) \end{aligned} \quad (2)$$

The input can be scaled to take values in $[0, 1]$ so \mathbb{I}^n denotes the n -dimensional unit hypercube.

ANOVA-representation. Assuming f is an integrable function, consider the following decomposition,

$$f(\mathbf{x}) = f_0 + \sum_{s=1}^n \sum_{i_1 < \dots < i_s} f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) \quad , \quad (3)$$

where f_0 is a constant and the f_{ij} are functions of subsets of $(x_i)_{i \in \{1, \dots, n\}}$. The double sum means that there is a function $f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s})$ for each possible family of input variables: from $f_1(x_1)$ to $f_n(x_n)$, then all the $f_{ij}(x_i, x_j)$ with $1 \leq \dots < i < j \leq n$ and so on up to $f_{1 \dots n}(x_1 \dots x_n)$. The number of terms in this decomposition is 2^n .

Sobol' [18] has shown that under the following condition on the summands of (3),

$$\int_0^1 f_{i \dots j}(x_i, \dots, x_j) dx_k = 0 \quad \text{for } k = i_1, \dots, i_s \quad , \quad (4)$$

the decomposition exists and is unique. It is called the ANOVA-representation of f . It follows from condition (4) that the summands in (3) are orthogonal and can be expressed as integrals of f . The use of this decomposition has been the subject of several other works. For instance, its treatment by

Rabitz and Aliş [23], who called it *High-dimensional model representation*, is of interest. Archer et al. [24] gave a review of the different concurrent developments that have been made on the subject.

Order 1 Sobol' indices. If f is square integrable, then the $f_{i_1\dots i_s}$ are also square integrable. Squaring and integrating (3) raises

$$\int_0^1 f^2(\mathbf{x}) \, d\mathbf{x} - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} \int_0^1 f_{i_1\dots i_s}^2 \, dx_{i_1} \dots dx_{i_s} \quad . \quad (5)$$

Now if \mathbf{x} is a random vector uniformly distributed in \mathbb{I}^n then $f(\mathbf{x})$ and $f_{i_1\dots i_s}(x_{i_1}, \dots, x_{i_s})$ are random variables whose variances are respectively,

$$D = \int_0^1 f^2 \, d\mathbf{x} - f_0^2 \quad (6)$$

and

$$D_{i_1\dots i_s} = \int_0^1 f_{i_1\dots i_s}^2 \, dx_{i_1} \dots dx_{i_s} \quad , \quad (7)$$

and the following equality holds:

$$D = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} D_{i_1\dots i_s} \quad . \quad (8)$$

In other words, D measures the variability due to variations of all the input variables while $D_{i_1\dots i_s}$ represents the variability caused by variations of the variables from the subset $(x_{i_1}, \dots, x_{i_s})$. Equation (8) states, as expected, that the overall variability is the sum of the variabilities caused by all the possible subsets of input variables.

This leads to define the Sobol' index of a subset of variables $(x_{i_1}, \dots, x_{i_s})$ by the following ratio,

$$S_{i_1\dots i_s} = \frac{D_{i_1\dots i_s}}{D} \quad , \quad (9)$$

where s is called the order of the index. Order 1 Sobol' indices, $S_i = D_i/D$, measure the influence of each half TSP clogging ratio *alone* while higher order indices measure the interactions. With 16 input variables there are as many as 120 order 2 indices. Estimating Sobol' indices requires numerous model evaluation so higher order indices were not computed.

Total Sobol' indices. the input variables are strongly physically linked so completely ignoring interactions could be misleading. As a palliative, consider the sum D_i^{tot} of the variances caused by all subsets that include a given variable x_i . Dividing this quantity by D , the overall variance, one defines total Sobol' indices, S_i^{tot} :

$$S_i^{tot} = \frac{D_i^{tot}}{D} \quad . \quad (10)$$

The difference between the total index and the order 1 index of a given variable quantifies its interactions with other variables.

A simple calculation [25] shows that D_i^{tot} and the variance D_{-i} caused by the subset of all variables except x_i , sum up to D :

$$D_i^{tot} = D - D_{-i} \quad . \quad (11)$$

Hence, each total Sobol' index can be deduced from the estimation of the variance of one subset of variables.

Computation scheme. Sobol' [19] has demonstrated that the variances corresponding to subsets of variables can be expressed as multidimensional integrals. Classical numerical methods are ineffective because of the high dimension of these integrals. However, Monte Carlo estimates can be derived and some are provided by Sobol' [18, 19]. Saltelli et al. [22] proposed a review of the available formulae for the computation of first order and total Sobol' indices and singled out a pair of them which have been used here as 'best practice' formulae.

Let \mathbf{A} and \mathbf{B} be two samples of N clogging configurations, that is to say two matrices of size $N \times n$. We denote by $\mathbf{A}_{\mathbf{B}}^{(i)}$ the matrix obtained from \mathbf{A} by replacing its i -th column by the i -th column from \mathbf{B} and by $f(\mathbf{A})_k$ (resp. $f(\mathbf{B})_k$ and $(f(\mathbf{A}_{\mathbf{B}}^{(i)}))_k$) the response of the model corresponding to the k -th individuals of sample \mathbf{A} (resp. \mathbf{B} and $\mathbf{A}_{\mathbf{B}}^{(i)}$). The formula used here to compute the denominator of first order indices is the following [25, 26]:

$$\widehat{D}_i = \frac{1}{N} \sum_{k=1}^N f(\mathbf{B})_k (f(\mathbf{A}_{\mathbf{B}}^{(i)})_k - f(\mathbf{A})_k) \quad . \quad (12)$$

The one used for total indices is [27]:

$$\widehat{D}_i^{tot} = \frac{1}{2N} \sum_{k=1}^N [f(\mathbf{A})_k - f(\mathbf{A}_{\mathbf{B}}^{(i)})_k]^2 \quad . \quad (13)$$

Estimating order 1 and total indices of n input variables with a Monte Carlo sample size of N requires $(n + 2) \times N$ model evaluations [22]: N for \mathbf{A} , N for \mathbf{B} and $n \times N$ for the $\mathbf{A}_{\mathbf{B}}^{(i)}$ with i varying from 1 to n .

In this context, standard Monte Carlo relying on pseudo-random numbers is only moderately effective. This is due to the tendency of pseudo-random sequences to aggregate into clusters which is detrimental especially in high dimension. Substantial improvement is achieved by using a quasi-Monte Carlo procedure based on low-discrepancy uniform sequences such as Sobol' sequences [24]. Here a Sobol' sequence has been used to generate the samples, following the procedure prescribed by Sobol' [19]. The input vector coordinates vary from 0 to 0.65 which covers most of practical clogging cases.

2.2.3. Preprocessing of the output

The increased pressure drop due to clogging alters both the full power 'static' values of the WRL and its dynamic behaviour. The 'static' value is presently used for cursory diagnosis of clogging. Examining the dynamic response is meant to retrieve more detailed information and to sidestep the issue of sensor bias. As the range of variation of the WRL 'static' value over the years of plant operation is large compared to the dynamic variations of the WRL during a power transient, it has been necessary to pre-process the data by removing the 'static' value trend. Indeed, Sobol' indices computed on unprocessed output reflect only the variance due to differences in WRL initial value.

A straightforward corrective action would be to subtract from each curve its initial value. However, this would arbitrarily eliminate the variance of the first sequential variables. Subtracting a constant is a crude correction and choosing this constant to be equal to the value taken by the initial variable concentrates all available accuracy on the beginning of the curves.

Let $w : t \mapsto w(t)$ be the WRL response function. For a given time t_0 , one can write a Taylor expansion of w of the form

$$w(t) = w(t_0) + w'(t_0)(t - t_0) + w''(t_0)\frac{(t - t_0)^2}{2} + \dots \quad (14)$$

Averaging (14) for t from 0 to 1148 makes the temporal mean appear as the first term of the 'average' expansion. Subtracting the temporal mean instead of the initial value is a means to distribute the error along the time interval. In this way no assumption is made *a priori* about the most informative part of the response curves.

Five sample WRL response curves are displayed in the left panel of figure 2. The difference in ‘static’ value can be appreciated by the difference in initial value. However, differences in the shape of the curves are difficult to assess. In the right panel of figure 2, the same curves are presented with their temporal mean subtracted. Their shapes appear more contrasted. For instance, the circle-marked and x-marked curves are approximately equidistant from the square-marked one in the left graph but the right graph shows that the circle-marked curve has a much more similar shape.

The validity of the subtraction of a constant has been investigated using the PCA results in section 3.1.2.

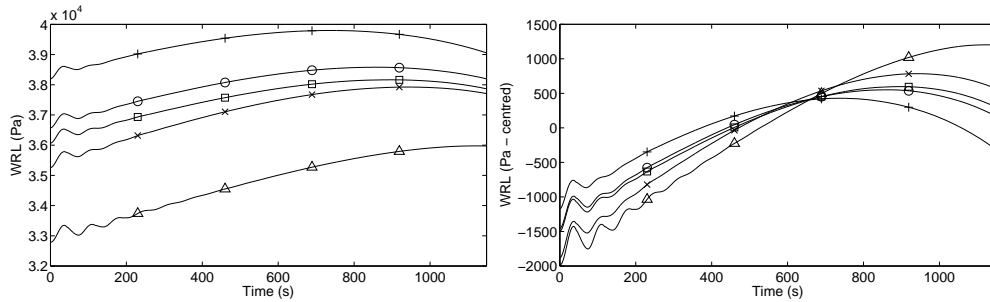


Figure 2: Simulated WRL response curves before and after subtraction of temporal means.

2.2.4. Reduction of output dimension

The most straightforward implementation of sensitivity analysis consists in considering the value of the WRL at each time step as distinct output variables. However this yields a large number of indices which makes the ranking of the input variables and the analysis of their interactions cumbersome. Moreover, this approach does not take into account the functional nature of the output because it is blind to the high correlation of the sequential variables. One way to tackle this issue is to expand the time series onto an appropriate orthogonal basis [28–30].

PCA is a simple method to obtain such a basis directly from the data. The optimization criterion used in PCA is the maximization of variance along the directions of the basis. As our aim is to attribute shares of overall variance to input variables, using a variance based method for dimension reduction makes sense.

Principal components. The principle of PCA is to progressively build an orthogonal projection basis by adding directions so that the spanned space fits the data the most adequately. Considering the output vector, $\mathbf{w} = (w(t))$, as a random vector leads to define principal components (PCs) as ordered linear combinations of the original variables, $(w(t))$, that are orthogonal and have maximum variance [4]. For a set of p variables, up to p PCs can be found. Coordinates along each PC are called *scores*; they constitute a new set of variables, each representing a smaller share of total variance than the previous one. The sample used for the sensitivity analysis can be used to estimate the PCs and their scores: the eigenvectors of its empirical covariance matrix are the PCs and their variances are the corresponding eigenvalues. The scores are then easily obtained by projection.

A common practice in PCA is to use standardised variables to avoid scale problems. For instance, if a variable lies in the interval $[10^3, 10^4]$ while the others vary from 0 to 10, the former would most of the variance of the dataset while varying, relatively, as much as the others. Both types of PCA (with untransformed variables and standardised variables) have been used in this study, leading to different types of interpretation.

Sensitivity analysis on scores. Eigenvalues usually drop quite quickly in magnitude. PCs of low eigenvalue describe small fluctuations in the dataset and one can neglect them without losing much information. Keeping only the r most prominent PCs allows to sum up the effect of clogging on the shape of the WRL response with a manageable number of variables.

The WRL response curves lie in a p -dimensional space. Excluding the $p - r$ PCs of lowest variances comes down to selecting the r -dimensional subspace that most nearly encloses the data, the curves being very ‘flat’ along the directions left aside. Then, Sobol’ indices can be computed on PC scores in exactly the same way as sequential indices. In addition of being less numerous, these new Sobol’ indices present the advantage of being linked to PCs whose shapes can be interpreted. A more sophisticated approach using the notion of generalized sensitivity indices has been proposed by [30]. Here, the small number of PCs with a substantial eigenvalue made it unnecessary.

2.2.5. Assessing indices validity

It is important to estimate the accuracy of the computed sensitivity indices. One wants to know for instance, if the ranking of the indices can be trusted as it is or if groups of input variables should be considered. In addi-

tion, the chosen computation scheme sometimes induces aberrations, such as slightly negative indices or sums of indices that exceed one, due to slow convergence of the Monte Carlo estimates; confidence intervals allow to decide if these irregularities can be overlooked or if larger samples should be used.

For each sensitivity index S , an estimator \hat{S} has been computed. Building a confidence interval consists in finding \hat{S}_{lo} and \hat{S}_{up} so that the two events $S < \hat{S}_{lo}$ and $S > \hat{S}_{up}$ have both a given small probability. As little is known about the distributions involved, bootstrap methods are particularly indicated as they are robust and distribution free.

Bootstrap confidence intervals. The general idea behind bootstrap is to draw conclusions about a given estimator by using the empirical distribution upon which the estimator is based. The estimator \hat{S} is linked to the sample used to compute it, ζ , by a function ϕ : $\hat{S} = \phi(\zeta)$. A *bootstrap sample* $\zeta^{(b)}$ is obtained by drawing uniformly with replacement from ζ . For each bootstrap sample, a *bootstrap replication* $\hat{S}^{(b)} = \phi(\zeta^{(b)})$ is computed in the same way as the estimator. It is possible to draw inferences on the underlying distribution followed by \hat{S} by analysing the empirical distribution of the bootstrap replications.

Bootstrap *percentile* confidence intervals are constructed by taking the α and $1 - \alpha$ percentiles of the empirical distribution obtained after re-sampling. The *bias-corrected and accelerated* (shortened BC_a) intervals used in this study are derived from the percentile intervals but include a correction of bias and an *acceleration* that compensates for variation of the standard error of S with the value of S . These two corrections consist in shifts of the percentiles finally chosen from the empirical distribution.

Details about bootstrap confidence intervals and their derivation can be found in the book by Efron and Tibshirani [31].

2.3. Sliced inverse regression

The dimensionality of the model input was chosen on a physical basis: it is the most detailed description of clogging that can be reasonably described by a 1D model. However, as the results of the sensitivity analysis will show, variations in shape of the response can be satisfactorily accounted by a smaller number of variables. Discarding irrelevant variables is necessary to ensure that the diagnosis method is only used in its applicability domain. It also reduces the size of the space to be sampled and allows for the construction of meaningful graphical representations.

Among dimension reduction techniques, sliced inverse regression (SIR) [5] bears several practical advantages. It is quite robust, easy to implement and based on a very generic model. A generic formulation of the SIR method is given in section 2.3.1. For SIR to be really useful, it is necessary to determine the dimension the input space can be reduced down to without loss of information. A bootstrap technique intended to address this issue is presented in section 2.3.2. Finally, section 2.3.3 details what adjustments are needed in the multivariate output case.

2.3.1. SIR principle

The basic idea behind SIR is to find a limited number of linear combinations of the predictors that are sufficient to retrieve the information from the regression.

The following model is assumed,

$$y = g(\beta'_1 \mathbf{x}, \dots, \beta'_q \mathbf{x}, \epsilon) \quad , \quad (15)$$

where (β_k) is a family of unknown vectors, g is an unknown function taking value in \mathbb{R}^{q+1} and ϵ is independent from \mathbf{x} .

The space $\text{Span}[(\beta_k)]$ is called the *efficient dimension reduction* (e.d.r.) subspace and its elements e.d.r. *directions*. This terminology emphasizes the fact that g is arbitrary and that the β_k themselves are not identifiable.

The inverse regression curve $\mathbb{E}(\mathbf{x}|y)$ lies in \mathbb{R}^n . If model (15) holds, it stays always close to a q -dimensional subspace. Appropriate conditions on the distribution of \mathbf{x} will ensure that it falls into the e.d.r. subspace.

Confining the inverse regression curve to the e.d.r. subspace. Consider the following condition,

Condition 1 (Linearity). *For any \mathbf{b} in \mathbb{R}^n , the conditional expectation $\mathbb{E}(\mathbf{b}'\mathbf{x}|\beta'_1 \mathbf{x}, \dots, \beta'_q \mathbf{x})$ is linear in $\beta'_1 \mathbf{x}, \dots, \beta'_q \mathbf{x}$.*

Such a condition is difficult to check because it involves the unknown (β_k) . It is however satisfied if \mathbf{x} has an elliptically symmetric distribution, such as the Gaussian distribution [5].

Theorem 1. *Under model (15) and condition 1, the centred inverse regression curve $\mathbb{E}(\mathbf{x}|y) - \mathbb{E}(\mathbf{x})$ lies in the subspace spanned by $(\Sigma\beta_{\mathbf{k}})$, where Σ is the covariance matrix of \mathbf{x} .*

Hence, substituting \mathbf{x} by its standardised version implies under condition 1 that the inverse regression curve is contained in the e.d.r. subspace.

SIR algorithm. The model can produce a sample of WRL responses for an arbitrary distribution of \mathbf{x} . This sample can then be used to estimate, first the inverse regression curve and then, in the same manner as in section 2.2.4, the q -dimensional subspace that most adequately contains it.

The following algorithm given by Li [5] has been used:

1. Standardise \mathbf{x} using its empirical covariance matrix $\widehat{\Sigma}$:
 $\tilde{\mathbf{x}}_i = \widehat{\Sigma}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$.
2. Divide the range of variation of y into H slices, I_1, \dots, I_H , each containing a proportion p_h of the N observations.
3. Compute the slice averages, $(\hat{\mathbf{m}}_h)$, of the input individuals:
 $\forall h \in \{1, \dots, H\}, \hat{\mathbf{m}}_h = \frac{1}{p_h N} \sum_{\{i|y \in I_h\}} \tilde{\mathbf{x}}_i$.
4. Compute $\widehat{V} = \sum_{h=1}^H p_h \hat{\mathbf{m}}_h \hat{\mathbf{m}}_h'$.
5. Find $(\hat{\boldsymbol{\eta}}_k)$, the family of eigenvectors of \widehat{V} sorted by decreasing eigenvalues.
6. Output $(\hat{\boldsymbol{\beta}}_k) = (\widehat{\Sigma}^{-1/2} \hat{\boldsymbol{\eta}}_k)_{k \in \{1, \dots, q\}}$.

In the last step of the algorithm, the $n - q$ eigenvectors with the smallest eigenvalues are left aside. It is necessary to determine the dimension of the e.d.r. subspace in order to avoid missing information or including spurious directions.

2.3.2. e.d.r. subspace dimension determination

Li [5] proposed a statistical test to determine the dimension q . Unfortunately, it relies on an assumption of Gaussian distribution for \mathbf{x} . As a non Gaussian distribution has been investigated here, the bootstrap approach devised by Lique and Saracco [32] has been preferred.

Let B_k and \widehat{B}_k be the matrices whose columns are respectively the vectors $(\boldsymbol{\beta}_k)$ and their estimators $(\hat{\boldsymbol{\beta}}_k)$ with k in $\{1, \dots, k\}$. Let P_k and \widehat{P}_k be the Σ -orthogonal and $\widehat{\Sigma}$ -orthogonal projectors onto the spaces spanned by these same vectors,

$$P_k = B_k(B_k' \Sigma B_k)^{-1} B_k' \Sigma \quad ; \quad \widehat{P}_k = \widehat{B}_k(\widehat{B}_k' \widehat{\Sigma} \widehat{B}_k)^{-1} \widehat{B}_k' \widehat{\Sigma} \quad . \quad (16)$$

The following risk function,

$$R_k = \frac{1}{k} \mathbb{E}[\text{Trace}(P_k \widehat{P}_k)] \quad , \quad (17)$$

expresses the closeness of the two subspaces. If $k = q$, the criterion tends to 1 when the sample size tends to infinity. Thus, for a given sample size N , a value of R_k close to 1 suggests that the e.d.r. dimension is greater than or equal to k . Contrarily, if $k > q$, additional vectors are not e.d.r. directions. Accordingly, they point in directions at an angle with the e.d.r. subspace and the value of R_k will stay below 1 whatever the sample size. Let us note however that the increase in k has also the opposite effect as each additional direction vector reduces the remaining angular space in which the non-e.d.r. direction points. In particular, for $k = n$ both projectors are the identity transformation and R_k is equal to 1 whatever the sample size. Hence, there is no value of R_k under which one can objectively assert that the k -th vector is not an e.d.r. direction. However, the variation of the risk function as k varies from 1 to n usually allows to deduce the e.d.r. dimension unequivocally.

A bootstrap estimate \hat{R}_k of R_k can be formed as follows [31]: for a given bootstrap replication of the sample used to conduct the SIR, the plug-in estimator of R_k is

$$\hat{R}_k^{(b)} = \frac{1}{k} \mathbb{E} \left[\text{Trace} \left(\hat{P}_k \hat{P}_k^{(b)} \right) \right] . \quad (18)$$

For \mathcal{B} bootstrap replications, the bootstrap estimate is

$$\hat{R}_k = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \hat{R}_k^{(b)} . \quad (19)$$

Then, the e.d.r. dimension is determined by computing the bootstrap estimate, \hat{R}_k , of the risk function for each value of k from 1 to n and inspecting the graph of \hat{R}_k against k . The value of the criterion stagnates or increases until $k = q$. It then drops abruptly before rising again as the remaining angular space closes. The variance of the bootstrap estimate is also greater for non-e.d.r. direction and can be used to come to a decision in ambiguous cases.

2.3.3. Multivariate output SIR

Several approaches have been proposed to adapt SIR to a multivariate context [33]. In this paper, the following adaptation of model (15) is adopted:

$$\mathbf{y} = g(\beta'_1 \mathbf{x}, \dots, \beta'_q \mathbf{x}, \epsilon) , \quad (20)$$

where \mathbf{y} stands for the multivariate output.

Building on what has been done for the sensitivity analysis in section 2.2.4, SIR has been carried out with the scores of the r selected PCs as output variables. Then a method called Pooled Marginal Slicing (PMS) has been applied to the r -dimensional output [33].

Pooled Marginal Slicing principle. Applying the SIR algorithm up to step 4 to each of the r components of the multivariate output yields a set of weighted covariance matrices $(\widehat{V}_i)_{i \in \{1, \dots, r\}}$. A convex combination with a set of weights (w_i) can be formed,

$$\widehat{V}_{pool} = \sum_{i=1}^r w_i \widehat{V}_i \quad . \quad (21)$$

The e.d.r. directions are finally estimated by executing the second half of the SIR algorithm with \widehat{V}_{pool} .

3. Results and discussion

3.1. Sensitivity analysis of the SG model

Sequential indices are presented in section 3.1.1 and the projection basis obtained by PCA is presented in section 3.1.2. This is compared to PCs obtained with data from industrial sites. Finally, section 3.1.3 details the ‘compact’ sensitivity indices computed with the reduced dimension output.

3.1.1. Sequential Sobol’ indices

The size of the Monte Carlo samples has been fixed to 1000 so a total of 18000 transient simulations have been run for the sensitivity analysis. Using 10 concurrent threads on two workstations, the total computation time was approximately one week.

Sequential order 1 and total indices are represented in figures 3 and 4. Indices are grouped in graphics by hot and cold leg variables. The shade of the curves corresponds to the height of the TSPs: light curves are associated to the lower TSP and dark ones to the higher. The error bars represent the bounds of the BC_a confidence intervals. Only a few of them are presented for readability but no discrepancies have been observed on the whole set. Both sequential order 1 and total sets of indices display two sharp contrasting behaviours for each leg. The ranking of the indices is the same in all cases:

the higher the TSP is positioned in the SG, the higher are the corresponding sensitivity indices. This is in agreement with the preliminary analysis conducted in section 2.1.3.

As stated in section 2.2.2, the difference between total indices and order 1 indices measures the amount of interaction. Comparison of figure 3 and 4 shows that interactions are globally weak and that they involve only the highest TSPs, except for the interval between 650 s and 750 s where hot leg TSPs interact strongly. It can be noted that the corresponding order 1 indices (left panel of figure 3) briefly become negative. The very low variance of the output in this interval probably causes this lack of convergence of the Monte Carlo estimates. Indeed, the time-centred response curves cross at approximately 700 s, as can be seen on the right panel of figure 2. It seems legitimate to consider these indices as null because their error bars are roughly centred on the baseline.

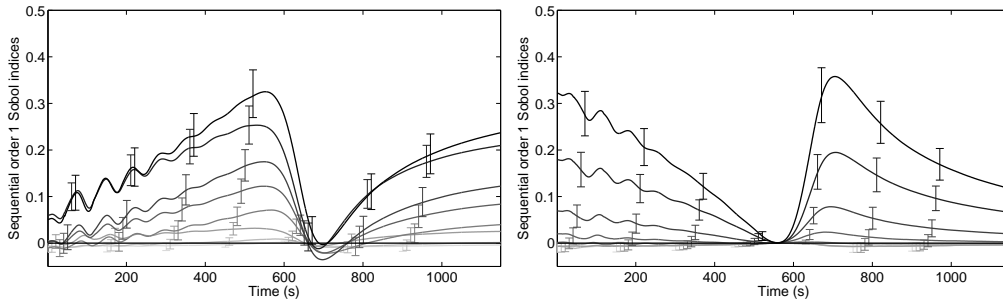


Figure 3: Sequential order 1 Sobol' indices. Both plots pertain to the same sensitivity analysis; indices corresponding to the hot leg are represented on the left and those corresponding to the cold leg on the right in order to ease readability.

3.1.2. Dimension reduction of the model output

Sequential indices revealed that the impact of clogging changes qualitatively with the SG leg and quantitatively with the level of the TSPs. However, the large number of sequential indices makes it difficult to estimate precisely the impact of each TSP and the interactions. In order to reduce the output dimension, a 'raw' and a normalized PCA have been carried out. The resulting PCs have been compared to those obtained with plant data.

The first 10 PCs obtained with a uniform sample with 'raw' and normalized variables are displayed in figure 5. The normalized PCs in the right panel of figure 5 have been multiplied by the square root of their eigenvalue.

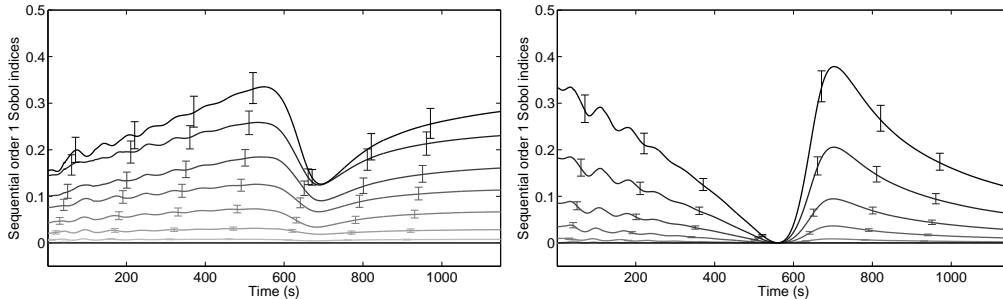


Figure 4: Sequential total Sobol' indices. Both plots pertain to the same sensitivity analysis; indices corresponding to the hot leg are represented on the left and those corresponding to the cold leg on the right in order to ease readability.

Hence, it is actually the sequential correlation coefficients between the time step variables and the PCs that are represented. In both cases, the first 2 PCs account for more than 99.9 % of the overall variance. It shows in the normalized variables graphic: the correlation coefficients of the next PCs almost do not depart from the baseline meaning that these PCs are only marginally correlated with the original variables. The low variance PCs can be seen in more details in the left panel because they all have an \mathcal{L}_2 norm equal to one. They are rather disorderly and do not look like any general feature of the curves except from the oscillations in the beginning that have been identified as numerical artefacts.

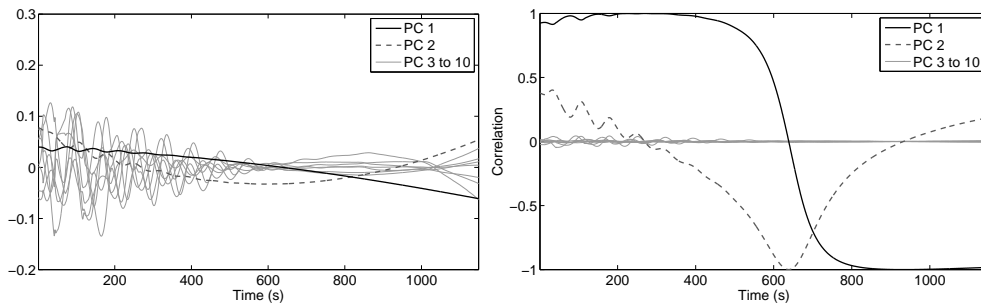


Figure 5: First 10 PCs obtained with the uniform sample (left: 'raw' variables; right: normalized variables)

The first 2 'raw' PCs are polynomials of degree 1 and 2. The first PC increases the global slope of the curves by spinning it round a fixed point around time 650 s. The second PC increases the curvature by 'bending' the

curves with two fixed points at times 250 s and 900 s. The PC 1 correlation coefficients curve is S-shaped with 2 plateaux at $+1$ and -1 from approximately 0 s to 400 s and 800 s to 1148 s. In between there are two sharp inflexions. This means that the original variables of the beginning of the time interval are highly correlated with PC 1 while those at the end are highly anti-correlated. The PC 1 correlation coefficients curve is V-shaped and points towards 1 around 650 s. Only the time steps variables of the middle of the interval are substantially correlated with the PC 2.

PCA on measured data. A PCA has been carried out on 291 measured response curves from 5 EDF units. The 97 processed transients (there are 3 SGs per unit) spread over a period of 10 years and each unit has undergone a chemical cleaning at some points. Hence, the data include a wide array of clogging configurations, from very low clogging just after the chemical cleaning, to very high clogging just before.

The first 3 PCs obtained without preprocessing and the first 2 PCs obtained with the preprocessing described in section 2.2.3 are displayed in figure 6. On the left panel, the first PC is nearly a constant and its scores are proportional to the temporal mean of the curves. The PCs are orthogonal by construction so PC 2 and 3 are very similar to PC 1 and 2 from the right panel. This validates the chosen preprocessing. The PCs obtained with measurements are similar to those found with the simulations. This shows that the main effects of clogging on WRL are correctly represented by the model and that PCA is an appropriate tool to represent them.

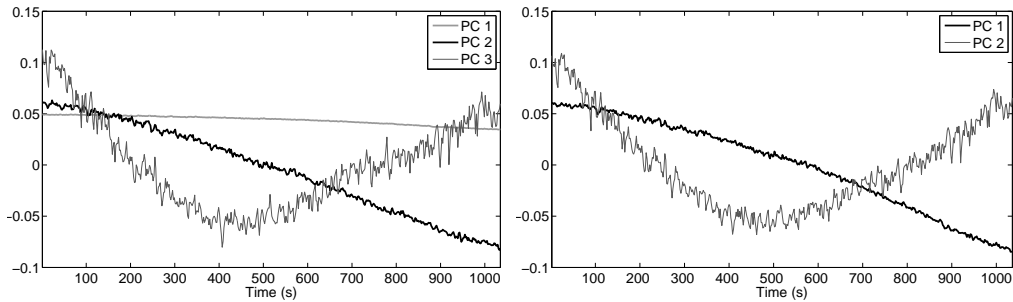


Figure 6: First PCs obtained on 291 measured response curves (left: before subtraction of the temporal mean; right: after subtraction of the temporal mean).

3.1.3. Sobol' indices of reduced dimension output

The first 2 PCs obtained in the previous section account for almost all of the variance of the sample. Comparison with plant data showed that they satisfactorily represent the effects of clogging on the WRL response. It is straightforward to select those 2 PCs to build a projection basis for the reduced dimension output sensitivity analysis. Sobol' indices computed with 'raw' and normalized PC scores did not differ fundamentally and only the former are presented here.

The results of the sensitivity analysis conducted on the first 2 sets of standardised PC scores are displayed in figure 7. Each couple of bars corresponds to a TSP. They are lined up from bottom to top in ascending order, hot leg first. The light bars represent total indices and the dark bars represent order 1 indices. The length difference of the two bars of a couple represents the interaction in which the input variable is involved. The error bars indicate the bounds of the BC_a confidence intervals.

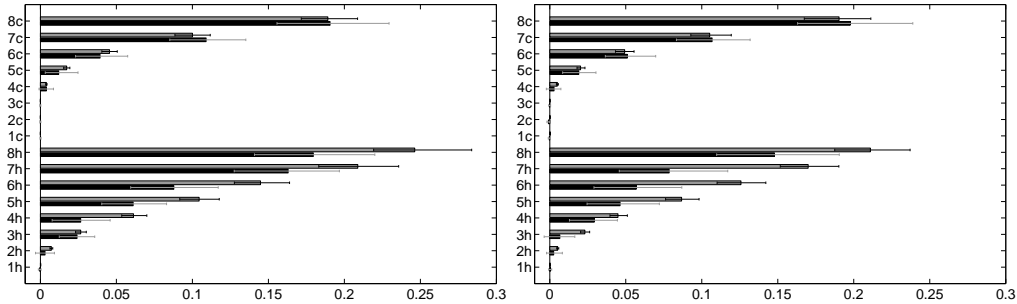


Figure 7: Order 1 and total Sobol' indices computed with PC scores (left: PC 1; right: PC 2).

As for the sequential indices, total indices have shorter confidence intervals and their length is proportional to the value of the indices while order 1 indices have longer confidence intervals of constant size. A few total indices are slightly smaller than their first order counterpart. However, these differences are always small estimates for total indices are always within the confidence intervals of order 1 indices. Thus, it seems sensible to assume that these order 1 indices actually equal the total indices and that no interaction is involved here. The ranking of the indices is again in agreement with the preliminary analysis. On the whole, interactions are rather limited. Taking the confidence intervals into account, only TSP 4 to 8 on the hot leg and to a

lesser extent TSP 7 and 8 on the cold leg seem to be involved in substantial interactions.

3.2. Dimension reduction of the SG model input

The sequential Sobol' index curves in figures 3 and 4 are almost proportional. This suggests that clogging of TSPs of a same leg affect the WRL response in a similar manner. In addition, sensitivity analysis on the PC scores showed that there are little input variable interactions. These observations give credibility to model (15) so SIR is well indicated to simplify the model input. The high dimension of the input raises a sampling issue. It is addressed in section 3.2.1. Then, results of PC-wise univariate SIR and multivariate SIR are detailed in section 3.2.2 and 3.2.3.

3.2.1. Note on sampling scheme

Gaussian sampling. A simple means to satisfy condition 1 is to choose a Gaussian distribution for \mathbf{x} . A sample of 10^4 response curves for clogging ratios following a multivariate Gaussian distribution of mean $0.65/2$ and standard error $0.65/6$ has first been simulated. A few individuals with negative or very high clogging ratios have been trimmed without affecting too much the elliptic symmetry of the distribution. The first two PCs obtained with this sample are similar in shape to those found with a uniform sample displayed in figure 5. However, the shares of explained variance are different: the ratio of the first eigenvalue to the second is much higher in the Gaussian sample case. The standard deviation of the sequential variables is also globally lower in the Gaussian case, especially in the middle of the time interval. This is due to the fact that the Gaussian sample covers a volume much smaller than the uniform sample. At best, the Gaussian sample can efficiently cover the hypersphere inscribed into the hypercube $[0, 0.65]^{16}$. This would not cause much trouble in low dimension, but here the hypercube looks more like a sea urchin than a cube: it has $2^{16} = 65536$ 'corners' having each a volume approximately 4.25 times higher than the volume of the inscribed hypersphere. The previous observations tend to show that the Gaussian sample is unable to capture what happens inside the 'corners' of the hypercube. Yet, sampling extensively the hypercube while preserving an elliptic contour for \mathbf{x} is rendered difficult by its shape. Indeed, trimming and re-weighting a uniform sample, following for instance the guidelines of Cook and Nachtsheim [34], is unlikely to succeed because the probability that at least one individual out of a 10^4 size sample falls into the inscribed hypersphere is lower than 0.04!

Flexibility of the linearity condition. Condition 1 is actually weaker than elliptic symmetry and SIR can yield sensible results in cases that does not exactly comply with it. It has been shown by Diaconis and Freedman [35] that most low-dimensional projections from a high-dimensional data set are approximately Gaussian. Hall and Li [36] extended this result showing that low-dimensional projections of high-dimensional data are almost linear. As an illustration, a simulation example of e.d.r. directions correctly identified by SIR with a uniform sample in dimension 10 is given in the rejoinder of [5]. Here the dimension is higher and the data are relatively smooth because they are produced by a model so it can be expected that SIR would work in spite of a violation of condition 1. The bootstrap dimension determination method has been successfully tested with a strongly non elliptically distributed input [32]. A 10^4 size uniform sample of WRL response curves has been simulated in order to investigate the model's behaviour inside the 'corners'. The results obtained with this uniform sample are presented below.

3.2.2. Marginal slicing

Marginal slicing has been applied to the first two PCs of the data set. Using 'raw' or normalized PCA made but little difference so only the results with normalized PCA are presented here. The number of slices had also a very limited influence. Here, 33 slices of cardinal 303 have been used.

Bootstrap estimates of the risk function for the e.d.r. space dimension have been computed using 500 bootstrap replications. Corresponding box plots for the uniform samples are given in figure 8. In both plots, the mean of \widehat{R}_k is first close to 1, then decreases steeply down to around 0.7 and eventually climbs up until it reaches 1 for $k = 16$. The variance of \widehat{R}_k is close to 0 on the initial plateau, then it soars at the beginning of the drop in mean and eventually decreases regularly down to 0 as k increases up to 16. The increase in mean in the third part of the plots is a consequence of the growth of the basis. Additional directions progressively restrict the angular domain where the directions found with the bootstrap replications may differ from those found with the original sample. This shows through the progressive reduction of the variance of \widehat{R}_k as k increases. When k equals 16, \widehat{P}_k and the $\widehat{P}_k^{(b)}$ are proportional to the identity.

The dimension of the e.d.r. space is given by the highest value of k for which the mean of \widehat{R}_k is nearly 1 and its variance nearly 0 [32]. Here, it is equal to 2 for both sets of PC scores. The first two directions found with the uniform sample are displayed in figure 9. The plot on the left and on the

right display the two e.d.r. directions obtained respectively with the scores of the first and second PC as output variable. The black circle-marked and grey square-marked lines represent respectively the coordinates of the first and second e.d.r. vectors in both plots. Each mark correspond to a half-TSP. Once again, they are numbered in ascending order from the bottom of the SG to its top, hot leg first. The lines are broken at the limit between the hot and cold leg; however, each pair of colour and mark correspond to a unique e.d.r. direction in each plot. The Gaussian sample yielded similar results but the directions were a little less monotonous which goes against physical reasoning. It was not able to retrieve the second e.d.r. direction with the PC 1 scores.

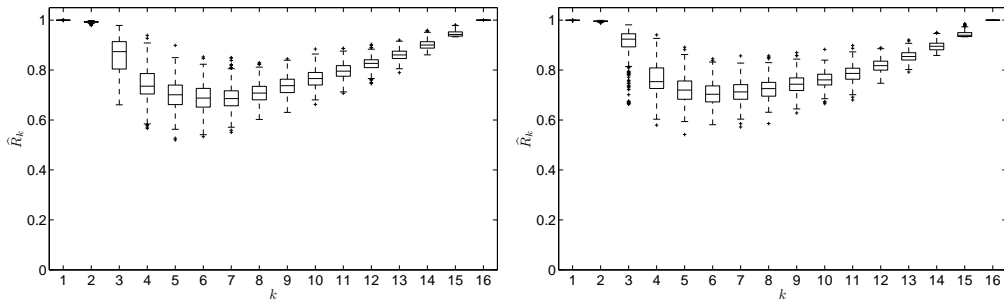


Figure 8: Box plots of \widehat{R}_k values with uniform sample for the PC 1 (left:) and PC 2 (right:) scores.

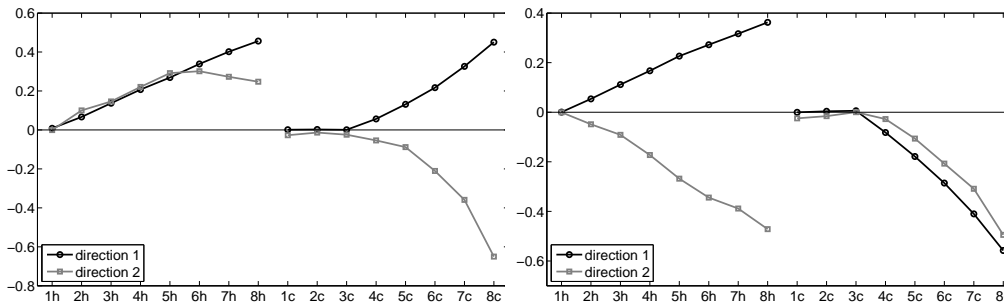


Figure 9: e.d.r. direction obtained with uniform sample for the PC 1 (left:) and PC 2 (right:) scores.

3.2.3. Pooled Marginal Slicing

The dimension of the e.d.r. space yielded by PMS (see section 2.3.3) with the two sets of PC scores is equal to 2 as can be seen on the left panel of figure 10. The right panel of figure 10 displays an orthogonal basis of the plane spanned by the first 2 directions found with the uniform sample. The vectors have been combined so that hot and cold legs are as separated as possible between. They are normalized to have a \mathcal{L}_1 -norm equal to 1 so that the coordinates vary in the same range as clogging ratios. Using PMS made SIR more robust to changes in the sampling scheme. Indeed, the basis obtained with the Gaussian sample was nearly the same as the one displayed in figure 10. The \widehat{R}_k values also indicated that the e.d.r. space is a plane but in a less obvious manner.

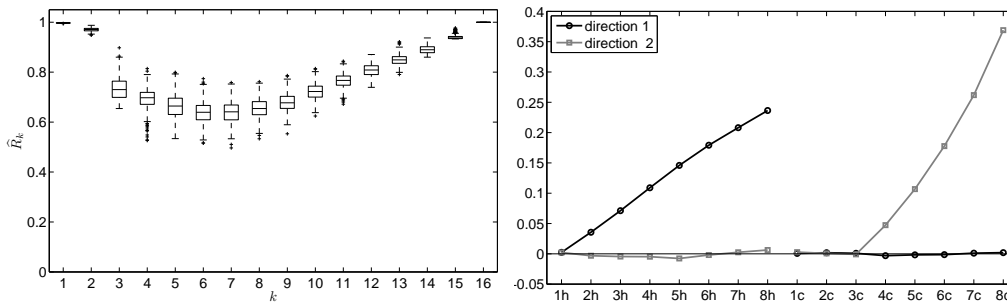


Figure 10: Box plots of \widehat{R}_k values (left:) and e.d.r. space basis (right:) obtained by PMS with the uniform sample.

The two e.d.r. directions found correspond to weighted averages of the clogging ratios of each leg. This means that a clogging diagnosis based on WRL response curve analysis will consist of hot and cold average clogging ratios. The results obtained here illustrate the fact that SIR can provide interesting projection direction even when the linearity condition is not fully satisfied. When the input dimension is large, elliptically contoured distributions are unlikely to be able to efficiently cover the domain of interest. In such cases, uniform sampling is a straightforward alternative to Gaussian sampling and the bootstrap method proposed by [32] can be used to determine the e.d.r. subspace dimension.

4. Conclusion

A methodology combining several statistical techniques has been carried out with a 1D SG model. It allowed to characterize the information about the clogging state of a SG that can be inferred from its WRL response to a power transient. The study has shown that :

- clogging affects the WRL response in two distinct ways. It alters its global slope and its curvature.
- These effects depend on the leg of the SG and the elevation of the clogging sites. Clogging of the hot leg and cold leg have a different impact and the former is predominant. The higher is the clogging site in the SG, the greater is the magnitude of the alteration.
- The WRL response curves can be summarised by vectors of size 2, each coordinate describing respectively the global slope and the curvature of the curves. The clogging state of individual half-TSPs cannot be identified by analysing the WRL response. The diagnosis actually consists in average clogging ratios of each leg.

The low dimensions of the simplified input and output provide a convenient framework for future development of a diagnosis methodology. Two response surfaces, one for each direction of the e.d.r. subspace basis, can be built by swapping the input and output. Then, any measured WRL response can be projected on the PC basis yielding two coordinates. The average clogging of each leg is then indicated by the heights of the two response surfaces associated to the couple of coordinates.

The methodology can be easily adapted to other diagnosis contexts. Here are three remarks to serve that purpose. The derivation of the diagnosis method is mainly based on the dimension reduction achieved with PCA and SIR. However, the basis of the e.d.r. subspace that SIR outputs may not be the most pertinent for the diagnosis. The sensitivity analysis provides valuable insights on the role of each input variable and suggests meaningful combinations of the e.d.r. directions found with SIR.

When the input dimension is high and it is suspected that important features may appear only for extreme values of the input variables, the SIR should be carried out with both a Gaussian sample and a uniform sample. Possible inconsistencies in the results can be caused by a too strong violation of the linearity hypothesis in the uniform sample case. In such a situation,

the sensitivity analysis can be used to remove the least influential input variables prior to the SIR. Keeping only the 3 to 5 most prominent variables allows to build a Gaussian sample that covers a more reasonable portion of the hypercube domain.

Finally, in situations where the output is more complex, that is, when there are more PCs with non-null eigenvalue, the reduction of the output dimension can be carried out in a more sophisticated way. One drawback of using the PC scores as the new output variables is that this choice is independent of the input. Using Hotelling's theory of most predictable variates, Li et al. [37] have proposed an extension of SIR that relies on the data to find the output projection basis.

References

- [1] M. Midou, J. Ninet, A. Girard, J.-M. Favennec, Estimation of SG TSP blockage: innovative monitoring through dynamic behavior analysis, in: Proc. 18th International Conference on Nuclear Engineering (ICONE18), 2010.
- [2] J. Ninet, J.-M. Favennec, Determination of applicability of EDF steam generator monitoring algorithm to pressurized water reactors worldwide, Tech. Rep. 1021079, EPRI, 2010.
- [3] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, Wiley Online Library, 2008.
- [4] I. T. Jolliffe, Principal Component Analysis, 2nd edition, Springer, 2002.
- [5] K. C. Li, Sliced inverse regression for dimension reduction (with discussion), Journal of the American Statistical Association 86 (414) (1991) 316–327.
- [6] Modelica association, Official webpage of the Modelica language, <http://www.dymola.com>, 2012.
- [7] Modelica language specification – version 2.2, Modelica association, 2005.
- [8] Dassault System, Official webpage of the Dymola (Dynamic Modeling Laboratory) software, <http://www.dymola.com>, 2012.

- [9] L. R. Petzold, Description of DASSL: a differential/algebraic system solver, Tech. Rep. SAND-82-8637; CONF-820810-21, Sandia National Laboratories, Livermore, CA (USA), 1982.
- [10] J. Pillet, N. Nimambeg, C. Pinto, Résultats des essais monophasiques sur la maquette P2C pour la détermination des pertes de charge induites par le colmatage, Tech. Rep. H-I84-2009-02838, EDF, 2010.
- [11] M. D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics* (1991) 161–174.
- [12] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of sobol indices for the gaussian process metamodel, *Reliability Engineering & System Safety* 94 (3) (2009) 742–751.
- [13] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering & System Safety* 93 (7) (2008) 964–979.
- [14] G. Blatman, B. Sudret, Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, *Reliability Engineering & System Safety* 95 (11) (2010) 1216 – 1229.
- [15] R. Bolado Lavín, W. Castaings, S. Tarantola, Contribution to the sample mean plot for graphical and numerical sensitivity analysis, *Reliability Engineering & System Safety* 94 (6) (2009) 1041–1049.
- [16] E. Borgonovo, Measuring uncertainty importance: investigation and comparison of alternative approaches, *Risk Analysis* 26 (5) (2006) 1349–1361.
- [17] E. Borgonovo, A new uncertainty importance measure, *Reliability Engineering & System Safety* 92 (6) (2007) 771–784.
- [18] I. M. Sobol', Sensitivity Estimates for Nonlinear Mathematical Models, in *Matem. Modelirovanie*, 2 (1)(1990) 112-118, English Transl.: *MMCE* 1 (4).
- [19] I. M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* 55 (2001) 271–280.

- [20] R. Cukier, H. Levine, K. Shuler, Nonlinear sensitivity analysis of multiparameter model systems, *Journal of computational physics* 26 (1) (1978) 1–42.
- [21] J.-Y. Tissot, C. Prieur, A bias correction method for the estimation of sensitivity indices based on random balance designs, *Reliability Engineering & System Safety* 107 (0) (2012) 205–213.
- [22] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications* 181 (2) (2010) 259–270.
- [23] H. Rabitz, O. F. Aliş, General foundations of high-dimensional model representations, *Journal of Mathematical Chemistry* 25 (3) (1999) 197–233.
- [24] G. E. B. Archer, A. Saltelli, I. M. Sobol’, Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *Journal of Statistical Computation and Simulation* 58 (1997) 99–120.
- [25] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety* 52 (1) (1996) 1–17.
- [26] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145 (2) (2002) 280–297.
- [27] M. J. Jansen, Analysis of variance designs for model output, *Computer Physics Communications* 117 (1-2) (1999) 35–43.
- [28] K. Campbell, M. McKay, B. Williams, Sensitivity analysis when model outputs are functions, *Reliability Engineering & System Safety* 91 (10-11) (2006) 1468–1472.
- [29] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod, Multivariate global sensitivity analysis for dynamic crop models, *Field Crops Research* 113 (2009) 312–320.
- [30] M. Lamboni, H. Monod, D. Makowski, Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models, *Reliability Engineering & System Safety* 96 (2010) 450–459.

- [31] B. Efron, R. Tibshirani, An introduction to the bootstrap, Chapman & Hall, 1993.
- [32] B. Liqueet, J. Saracco, Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR α method, Communications in Statistics Simulation and Computation 3 (6) (2008) 1198–1218.
- [33] L. Barreda, A. Gannoun, J. Saracco, Some extensions of multivariate sliced inverse regression, Journal of Statistical Computation and Simulation 77 (1) (2007) 1–17.
- [34] R. D. Cook, C. J. Nachtsheim, Reweighting to achieve elliptically contoured covariates in regression, Journal of the American Statistical Association 89 (426) (1994) 592–599.
- [35] P. Diaconis, D. Freedman, Asymptotics of graphical projection pursuit, The Annals of Statistics 12 (3) (1984) 793–815.
- [36] P. Hall, K. C. Li, On almost linearity of low dimensional projections from high dimensional data, The Annals of Statistics 21 (2) (1993) 867–889.
- [37] K. C. Li, Y. Aragon, K. Shedden, C. Thomas Agnan, Dimension reduction for multivariate response data, Journal of the American Statistical Association 98 (461) (2003) 99–109.

Bibliographie

- ADOBES André : Bilan du projet GV-SCOPE, qui a fédéré les actions menées par EDF-R&D de 2007 à 2009, en réponse au colmatage des plaques entretoises de générateurs de vapeur du parc nucléaire français. Rapport technique H-I84-2009-03284-FR, EDF, novembre 2011.
- ANSYS UK LTD. : Site officiel du code ANSYS CFD, 2012. <http://www.ansys.com/>.
- ARCHER G. E. B., SALTELLI Andrea et SOBOL' Il'ya Meerovich : Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58:99–120, 1997.
- ASCOMP GMBH : Site officiel du code TransAT, 2012. <http://www.ascomp.ch/transat/>.
- AUDIBERT Lorenzo : Conversation personnelle, 2012.
- AUDIBERT Lorenzo, FILIOT Pierre-Louis et BOUTBOUL Dustin : Méthodes et algorithmes pour l'exploitation automatique des examens télévisuels du colmatage des générateurs de vapeurs. Rapport technique H-P1B-2012-01989-FR, EDF, à paraître.
- AVENAS Cécile, ALVAREZ Dominique, SCHUCK Ugo, SOUYRI Anick et WALTER O. : Code LEDA – Note descriptive et note d'utilisation du modèle de générateur de vapeur DRIFBI – Application au GV 47/22. Rapport technique HI-81/05/038/A, EDF, septembre 2005.
- AVENAS Cécile, DESVIGNES Nicolas et SCHUCK Ugo : Note de conception du modèle DRIF-BI en Modelica – Formulation physique et utilisation. Rapport technique HI-81/03/29/A, EDF, avril 2004.
- BALLEYDIER Sylvain : Qualification of manta 2 channels sg model on MEGEVE and on ASME experiments. Rapport technique NEPD-F DC 10307, Areva, novembre 2010.
- BARBE Vicent et DUMAY David : évaluation du colmatage des plaques entretoises – bilan des connaissances sur les procédés par sonde interne. Rapport technique EDEECE110317/A, EDF, juin 2012.
- BARREDA Laure, GANNOUN Ali et SARACCO Jérôme : Some extensions of multivariate sliced inverse regression. *Journal of Statistical Computation and Simulation*, 77(1):1–17, 2007.

- BERNIER Valérie : Impact du colmatage sur les paramètres de fonctionnement des GV 51B sur la base du modèle monodimensionnel des générateurs de vapeur de remplacement du palier 900 MWe. Rapport technique ANREEC070136/A, EDF, juillet 2007.
- BERTIN Jean-Claude et PÉCOUT Jean-Michel : Présentation générale de l'évaporateur des générateurs de vapeur des tranches rep 900, 1300 et 1450 mw. Rapport technique NTD 00/0059, EDF, mars 2003.
- BERTRAND Pierre et CRINON Renaud : 1^{ère} réunion du groupe de travail « rendre plus robuste l'indicateur NGL utilisé pour le suivi du colmatage » (26 mai 2011). Compte-rendu de réunion, octobre 2011. D4550.01-11/2929.
- BESTION Dominique : The physical closure laws in the CATHARE code. *Nuclear Engineering and Design*, 124(3):229–245, 1990.
- EDF : Générateur de vapeur. Notice Technique B326, EDF, 1977.
- BIRD R. Byron, STEWART Warren E. et LIGHTFOOT Edwin N. : *Transport Phenomena*. Wiley, 2^{ème} édition, mars 2006.
- BITTAN Joseph Jeremy, DENIS Loleh et GOURJAULT Cristophe : 1300 MW - GV 68/19 – plaques entretoises supérieures colmatées – modélisation bi-canal de la zone chauffante avec MANTA. Rapport technique NEPR-F DC 10176, Areva, mars 2009.
- BOLADO LAVÍN Ricardo, CASTAINGS William et TARANTOLA Stefano : Contribution to the sample mean plot for graphical and numerical sensitivity analysis. *Reliability Engineering & System Safety*, 94(6):1041–1049, 2009.
- BORGONOVO Emanuele : A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.
- BOUSKELA Daniel : Bibliothèque Modelica thermohydraulique. Structure et recommandations de conception. Rapport technique H-P12-2003-02281, EDF, novembre 2003.
- BOUSKELA Daniel : Site officiel de la bibliothèque ThermoSysPro, 2012. <http://www.modelica.org/libraries/thermosyspro>.
- BREMAUD Éric : Note de synthèse sur la qualification du code PERFGV version 2. Rapport technique EEG DC 2119, Areva, mars 2010.
- BROCHET L. : Note de synthèse de qualification du code MANTA. Rapport technique EPTA DC 1505, Framatome, septembre 2000.
- CAMPBELL Katherine, MCKAY Michael D. et WILLIAMS Brian J. : Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, 91(10-11):1468–1472, 2006.

- CAMPOLONGO Francesca et SALTELLI Andrea : Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety*, 57(1):49–69, 1997.
- CAPPÉ Olivier, GODSILL Simon J. et MOULINES Éric : An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings*, 95(5):899–924, 2007.
- CARRADEC Ronan : CNPE de Bravo– tranche 1; Expertises télévisuelles réalisées sur la plaque entretoise 9 (PE8) du GV 2 (n°208) – VD3 2011. Rapport technique EDIAT110460/A, EDF, juin 2011.
- CD-ADAPCO : Site officiel du code STAR-CD, 2012. <http://www.cd-adapco.com/index.html>.
- C.E.A. : Site officiel du code CATHARE, 2012. <http://www-cathare.cea.fr/>.
- CHEONG Yong-Moo, CHAUDARY M. Sabir, EDWARDS Paul, GROSSER Paul, RODDA John et KHAN A. Asghar : *Eddy Current Testing at Level 2: manual for the Syllabi Contained in IAEA-TECDOC-628.Rev. 2 "Training Guidelines for Non Destructive Testing Techniques"*. IAEA, 2011.
- CHIP Vincent et MIDOU Mickael : Approche stochastique pour l'estimation du colmatage des générateurs de vapeur d'une centrale nucléaire. Rapport technique H-P1B-2009-02146, EDF, août 2009.
- CHIP Vincent, MIDOU Mickael, FAVENNEC Jean-Melaine et SIROS Frédérique : étude d'estimation du colmatage par traitement des données de process sur un modèle physique de GV. Rapport technique H-P1B-2009-03213, EDF, octobre 2008.
- CHUDIAK Mike et WARNICA Dave : Steam generator management program: thermal-hydraulic analysis of a recirculating steam generator using commercial computational fluid dynamics software. Rapport technique 1024838, EPRI, février 2012.
- COLLIER John Gordon et THOME John .R. : *Convective Boiling and Condensation*. Oxford University Press, 1996.
- COOK Robert Dennis : *Regression Graphics: ideas for Studying Regressions Through Graphics*. Wiley Online Library, 1998.
- COOK Robert Dennis et NACHTSHEIM Christopher J. : Reweighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association*, 89 (426):592–599, 1994.
- CORREDERA Géraldine : Renforcement de la surveillance de la chimie du circuit secondaire et suivi de la propreté des générateurs de vapeur. Rapport technique EDEECH080312, EDF, février 2009.

- COUDRET Raphaël, GIRARD Stéphane et SARACCO Jérôme : A new sliced inverse regression method for multivariate response regression. <http://hal.inria.fr/hal-00714981>, à paraître.
- COUVIDOU Philippe : Granulométrie des produits de corrosion du circuit secondaire. Rapport technique EDLCHM100404/A, EDF, juin 2011.
- COUVIDOU Philippe : Conversation personnelle, décembre 2012.
- CRINON Renaud : évaluation de l'efficacité des NCGV 2007-2008 sur la performance des GV. Rapport technique D4550.31-09/2003, EDF, octobre 2009a.
- CRINON Renaud : Surveillance de la performance des générateurs de vapeur. Rapport technique D4550.31-08/2832, EDF, mars 2009b.
- CUKIER R.I., LEVINE H.B. et SHULER K.E. : Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics*, 26(1):1–42, 1978.
- DAVID Franck : Benchmark comparison of steam generator thermal-hydraulic and flow induced vibration industry codes. Rapport technique, EPRI, à paraître.
- DAVID Franck et GUELFY Antoine : THYC-ECHANGEUR version 3, application aux générateurs de vapeur, fiches de qualification – partie 2. Rapport technique HT-33/95/017/B, EDF, décembre 1999.
- de BOUVIER P., O. a. M. : Avis technique CEIDRE : évaluation du colmatage résiduel des plaques entretoises des GV de Charlie2 après nettoyage chimique (octobre 2007). Rapport technique EDEDFM070308/A, EDF, juin 2007.
- DE CRECY F. : étude du comportement du secondaire des générateurs de vapeur – Résultats des expériences Patricia GV2. Rapport technique TT/SETRE/173, C.E.A., 1985.
- DE VITO Sophie : Caractérisation physico-chimique de dépôts côté secondaire de tubes de générateurs de vapeur issus de la centrale de Charlie 1. Rapport technique D.5710/ECH/2002/005093/00, EDF, juillet 2002.
- DÉCOSSIN Étienne : Coefficient d'échange thermique monophasique en assemblage combustible REP – synthèse bibliographique. Rapport technique HI-84/02/01 1/A, EDF, avril 2002.
- DECOUT Antoine : Conditions de référence et coefficients de sensibilité pour le suivi de la pression fournie par les GV. Rapport technique ENITSM080250, EDF, février 2009a.
- DECOUT Antoine : Effet prévisionnel de l'encrassement des GV sur les performances des tranches 1300 MW avec/sans augmentation de puissance et besoin en lessivages chimiques. Rapport technique ENITSM080171, EDF, mars 2009b.

- DENEUX Olivier et FAVENNEC Jean-Melaine : Projet EPO – bilan des possibilités d’amélioration des mesures primaires des tranches nucléaires. Rapport technique H-P1C-2010-02032-FR, EDF, novembre 210.
- DENEUX Olivier et GIRARD Sylvain : Description du modèle dynamique de GV 51B sous Dymola/Modelica pour l’évaluation du colmatage. Rapport technique H-P1C-2011-00910-FR, EDF, juin 2011.
- DEOTTO Geoffrey : Détermination du taux de circulation à partir des mesures de niveau gamme large. état d’avancement de l’application d’un modèle de transposition NGL/CR à l’ensemble des GV du parc. Rapport technique ENREEC090630, EDF, décembre 2009.
- DEVIDAL Éric et FOULLON Clémentine : 900 MWe – GV 51B – Plaques entretoises supérieures colmatés – Comparaison CATHARE/MANTA/ANETH. Rapport technique NEPR-F DC 10201, Areva, mars 2009.
- DEVINCK Jean-Michel : Mode opératoire appliqué pour la détermination de l’indicateur SAX lié au colmatage sur les plaques entretoises quadrifoliées des GV 51B, 51 BI et 68/19. Rapport technique EDEECE080127/B, EDF, janvier 2009.
- DIACONIS Persi et FREEDMAN David : Asymptotics of graphical projection pursuit. *The Annals of Statistics*, 12(3):793–815, 1984.
- DIJOUX Michel : Synthèse de l’évaluation et de la caractérisation de l’encrassement secondaire des GV de Charlie 1 avant nettoyage chimique. Rapport technique D.5710/IRCE/2003/006672/00, EDF, novembre 2003.
- DOLL Mathieu : CNPE de Delta– tranche 1, expertises télévisuelles réalisées sur la plaque entretoise n° 8 (PE8) du GV2 (GV n° 101) – VP-2011. Rapport technique EDIAT110588/A, EDF, juillet 2011.
- DOUETIL Thierry : Projet DEREK : détermination des incertitudes des mesures utilisées par le modèle VALI du circuit secondaire de Charlie tranche 3. Rapport technique H-P1C-2008-01482-FR, EDF, décembre 2008.
- DUAN Naihua et LI Ker Chau : Slicing regression: a link-free regression method. *The Annals of Statistics*, 19(2):505–530, 1991.
- DUPONT Jean-François : *Influence du Degré de Simplification du Modèle de l’Hydrodynamique Diphasique sur le Comportement Dynamique Simulé d’un Générateur de Vapeur*. Thèse de doctorat, École Polytechnique Fédérale de Lausanne, 1978. Thèse n° 311.
- Dymola : dynamic modeling laboratory user’s manual – Dymola 6 addition (version 6.1)*. Dynasim AB, 2004a.
- Dymola : dynamic modeling laboratory user’s manual (version 5.3a)*. Dynasim AB, 2004b.

- EFRON Bradley et TIBSHIRANI Robert : *An Introduction to the Bootstrap*. Chapman & Hall, 1993.
- EL HEFNI Baligh, BOUSKELA Daniel et LEBRETON Gregory : Dynamic modelling of a combined cycle power plant with ThermoSysPro. *Dans 8th Modelica conference proceedings*, pages 365–375, Dresden, Allemagne, mars 2011. Modelica association et Fraunhofer IIS EAS.
- EPRI : ATHOS/SGAP, Version 3.1 Analysis of the thermal hydraulics of a steam generator/steam generator analysis package, version 3.1. Rapport technique 1016564, EPRI, novembre 2008.
- FRAMATOME : Générateur de vapeur série 51 – Plaque entretoise brochée. Plan, juillet 1983a. 430 CPY 1614 C.
- FRAMATOME : Générateur de vapeur série 51 – Plaque entretoise supérieure brochée. Plan, novembre 1983b. 430 CPY 1615 F.
- FRAMATOME : Générateur de vapeur de rechange modèle 51B – Ensemble vue intérieure. Plan, octobre 1984. FRA EE-GP-DB-0049 C.
- GAUTHERON G. : Principe et résultats d'une modélisation thermo-hydraulique de générateur de vapeur. Rapport technique ENREEC020546, EDF, juin 2006.
- GAY Aurélien, MIDOU Mickael, FAVENNEC Jean-Melaine, et NINET Julien : Rapport de stage d'Aurélien Gay sur l'étude d'estimation du colmatage par traitement des données process sur un modèle physique de GV. Rapport technique H-P1B-2009-02094, EDF, septembre 2009.
- GEFFRAYE Geneviève, ANTONI O., FARVACQUE M., KADRI D., LAVIALLE G., RAMEAU B. et RUBY A. : CATHARE 2 V2.5_2: a single version for various applications. *Nuclear Engineering and Design*, 241(11):4456–4463, 2011. 13th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-13).
- GHATTAS S. et BLANC Jean-Michel : EP RGL 4 – Essai périodiques de calibrage des grappes grises. Rapport technique EP/NE/DC.0476, EDF/Framatome, 1987.
- GHIAASIAAN Seyed Mostafa : *Two-phase flow, boiling and condensation in conventional and miniature systems*. Cambridge University Press, 2008.
- GICQUEL Renaud : *Systèmes énergétiques*, volume 1. Transvalor – Presse des Mines, février 2009.
- GINESTE Pierre : Encrassement des tubes des GV : le code BOUTHYC à l'épreuve des mesures sur les tranches 900 MW et 1300 MW du parc français. Rapport technique H-I84-2008-04544-FR, EDF, avril 2009.

- GIRARD Sylvain : Ajustement de courbes de réponse à un transitoire de puissance de référence pour l'estimation du colmatage. Rapport technique H-P1C-2011-01438-FR, EDF, septembre 2011.
- GIRARD Sylvain : Extraire des champs et profils de résultats de simulations avec THYC et Dymola (manuel pratique). Rapport technique xxx, EDF, 2012a.
- GIRARD Sylvain : Réunion de thèse – bilan de la troisième année. Présentation orale, juillet 2012b.
- GIRARD Sylvain et PRUSEK Thomas : Analyse de sensibilité du modèle de colmatage par vena contracta à l'échelle du générateur de vapeur. Rapport technique H-P1C-2012-01742-FR, EDF, 2012.
- GOETGHELUCK Jean-François : Cahier des charges : comportement thermohydraulique des GV colmatés – coefficients de correction I'PV et qualification de l'écoulement en haut de GV. Rapport technique ENREEC100256/B, EDF, janvier 2012.
- GUELFY Antoine, BOUCKER Marc, HÉRARD Jean-Michel, PÉTURAUD Pierre, BESTION Dominique, BOUDIER P., FILLION P., GRANDOTTO Marc et HERVIEU E. : A new multi-scale platform for advanced nuclear thermal-hydraulics status and prospects of the NEPTUNE project. *Dans 11th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH)*, Avignon, France, octobre 2005.
- GUELFY Antoine et PITOT Samuel : Note de principe THYC version 4.1, partie 1 : modélisation. Rapport technique H-I84/03/020/A, EDF, avril 2004.
- GUIVARCH Michel et GAY Nadine : Doctrine de maintenance générateurs de vapeur REP – propreté du secondaire. Rapport technique D4550.01-10/5517, EDF, décembre 2010.
- HALL Peter et LI Ker Chau : On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, 21(2):867–889, 1993.
- HASTIE Trevor, TIBSHIRANI Robert et FRIEDMAN Jerome : *The elements of statistical learning*, volume 1. Springer Series in Statistics, 2001.
- HEO Gyunyoung : Condition monitoring using empirical models: technical review and prospects for nuclear applications. *Nuclear Engineering and Technology*, 40(1):49–68, 2008a.
- HEO Gyunyoung : Development of a fouling monitor for safety-related heat exchangers. *Nuclear Engineering and Design*, 238(6):1295–1303, 2008b.
- HEO Gyunyoung, CHOI Seong Soo et CHANG Soon Heung : Thermal power estimation by fouling phenomena compensation using wavelet and principal component analysis. *Nuclear engineering and design*, 199(1):31–40, 2000.

- HEWITT Geoffrey F. : Empirical and phenomenological models for multiphase flows – I. Vertical flows. *Dans 28th Short Courses - Modelling and Computation of Multiphase Flows* Swi (2011).
- HIBIKI Takeshi et ISHII Mamoru : One-dimensional drift-flux model and constitutive equations for relative motion between phases in various two-phase flow regimes. *International Journal of Heat and Mass Transfer*, 46(25):4935–4948, 2003.
- HOLCBLAT A. : Générateur de vapeur modèle 55/19 – note de synthèse sur la qualification du code ANETH version4. Rapport technique EEG DC 2865, Framatome, mars 2001.
- HOMMA Toshimitsu et SALTELLI Andrea : Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.
- I.A.W.P.S. : Site officiel de l’I.A.W.P.S., 2012. <http://www.iapws.org/>.
- IDAHO NATIONAL LABORATORY : Site officiel du code RELAP5, 2012. <http://www.inl.gov/relap5/relap5-3.htm>.
- IDEL’CIK I. E. : *Memento des Pertes de Charge - Coefficients de Pertes de Charge Singulières et de Pertes de Charge par Frottement*. Collection de la Direction des Études d’Électricité de France. Édition Eyrolles, 1969. Traduit du russe.
- IGLESIAS A.M., CALDERON R. et OTHERS : Thermal resistance contributions of oxides growth on Incoloy 800 steam generator tubes. *Nuclear engineering and design*, 219(1):1–10, 2003.
- INGIMUNDARDOTTIR H. et LALOT Sylvain : Detection of fouling in a cross-flow heat exchanger using wavelets. *Heat Transfer Engineering*, 32(3-4):349–357, 2011.
- IOOSS Bertrand : Revue sur l’analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1):3–25, 2011.
- ISHII Mamoru et HIBIKI Takeshi : *Thermo-Fluid Dynamics of Two-Phase Flow*. Springer Science, 2006.
- JANSEN Michiel J.W. : Analysis of variance designs for model output. *Computer Physics Communications*, 117(1-2):35–43, 1999.
- JOLLIFFE Ian T. : *Principal Component Analysis*. Springer, 2^{ème} édition, 2002.
- JONSSON G.R., LALOT Sylvain, PALSSON O.P. et DESMET B. : Use of extended kalman filtering in detecting fouling in heat exchangers. *International journal of heat and mass transfer*, 50(13-14):2643–2655, 2007.
- JUILLOT GUILLARD Delphine : Arrêts fortuits de Delta 1 et 4, synthèse des études de compréhension. Rapport technique NEEG-F DC 10017, AREVA, septembre 2007.

- JURKOWSKI Romuald : Paramètres de fonctionnement du générateur de vapeur 68/19 en cas d'abaissement de la température primaire calculés avec le code PERFGV faisceau alliage 600. Rapport technique NEEG-F DC 10176, Areva, mars 2009.
- KASTELANSKI Pierre : Modélisation d'un générateur de vapeur PWR à l' aide du code SICLE. Rapport technique HI/4572-02, EDF, août 1983.
- KLIMAS S. J., G. Miller D. et W. Turner Carl : The effect of the removal of steam generator tube ID deposits on heat transfer. Rapport technique 11985, AECL, mai 1998.
- KOLEV Nikolay Ivanov : *Multiphase flow dynamics: nuclear thermal hydraulics*, volume 5. Springer Verlag, 2^{ème} édition, 2012.
- KREIDER M. A., WHITE G. A. et VARRIN R. D., J. : Steam generator thermal performance degradation case studies. Rapport technique TR-110018, EPRI, juillet 1998.
- KUCHERENKO Sergei : Monte Carlo and quasi-Monte Carlo methods. *Dans SAMO summer school 2012*. J.R.C., juillet 2012.
- LACROIX Régis : Générateurs de vapeur REP AP06/09 – colmatage et encrassement du secondaire des GV – stratégie de maintenance. Rapport technique D4550.01-11/3257 ind. 0, EDF, janvier 2012.
- LALOT Sylvain et MERCÈRE G. : Detection of fouling in a heat exchanger using a recursive subspace identification algorithm. *Dans Proceedings of the 19th International Symposium on Transport Phenomena*, pages 17–21, 2008.
- LAMBONI Matieyendou, MAKOWSKI David, LEHUGER Simon, GABRIELLE Benoit et MONOD Hervé : Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113:312–320, 2009.
- LAMBONI Matieyendou, MONOD Hervé et MAKOWSKI David : Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96:450–459, 2010.
- LEBART Ludovic, PIRON Marie et MORINEAU Alain : *Statistique exploratoire multidimensionnelle*. Sciences Sup. Dunod, 4^{ème} édition, 2006.
- LEBRUN Julien et PETIT D. : Rapport d'expertise : analyse physico-chimique de boues GV provenant du CNPE de Delta – tranche 3. Rapport technique EDLCHM060375, EDF, août 2006.
- LEBRUN Julien et PETIT D. : Rapport d'expertise : analyse physico-chimique de boues GV provenant du CNPE de Charlie – tranche 2. Rapport technique EDLCHM070186, EDF, avril 2007.
- LI Ker Chau : Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 86(414):316–327, 1991.

- LI Ker Chau, ARAGON Yves, SHEDDEN Kerby et THOMAS-AGNAN Christine : Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, 98(461):99–109, 2003.
- LINER Y., CARVER M. B., TURNER Carl W. et CAMPAGNA A. O. : Simulation of magnetite particulate fouling in nuclear steam generators. *Dans ASME meeting on Steam Generator Sludge Deposition in Recirculating and Once Through Steam Generator Upper Tube Bundle and Support Plates*, volume 8, 1992.
- LIQUET Benoît et SARACCO Jérôme : Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR α method. *Communications in Statistics Simulation and Computation*, 3(6):1198–1218, 2008.
- LIQUET Benoît et SARACCO Jérôme : A graphical tool for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. *Computational Statistics*, 27(1):103–125, 2012.
- MAKARAWIEZ Gregory, BOYARD Guillaume et JURKOWSKI Romuald : Palier R.E.P. 900 MW – CP0, CP1 et CP2 générateur de vapeur de remplacement – modèle GVRA 47/22, paramètres de fonctionnement calculés avec le code PERFGV. Rapport technique NEEG-F DC 10177, Areva, avril 2009.
- MANGIN Matthieu et GOETGHELUCK Jean-François : Détermination d’ajustements statistiques reliant le taux de colmatage en PE supérieure aux mesures de niveau gamme large des GV 51B, 47/22 et 68/19 – Mise en application pour le parc en exploitation. Rapport technique ENREEC120325A, EDF, à paraître.
- MATHWORKS : Fonction smooth.m, septembre 2012. <http://www.mathworks.fr/fr/help/curvefit/smooth.html>.
- MERCIER Stéphane : Estimation d’encrassement de la partie secondaire des générateurs de vapeur de l’ensemble des tranches du parc. Rapport technique EDEECH080172, EDF, janvier 2010.
- MIDOU Mickael, NINET Julien, FAVENNEC Jean-Melaine, GAY Aurélien et CHIP Vincent : Estimation du colmatage des générateurs de vapeur de type 51b et 47/22 : méthode du NGL dynamique. Rapport technique H-P1B-2009-03213, EDF, janvier 2010a.
- MIDOU Mickael, NINET Julien, GIRARD Alexandre et FAVENNEC Jean-Melaine : Estimation of SG TSP blockage: innovative monitoring through dynamic behavior analysis. *Dans 18th International Conference on Nuclear Engineering (ICONE18)*, mai 2010b.
- Modelica Language Specification - Version 2.2*. Modelica Association, février 2005.
- MOLEIRO Edgar, OUKACINE Fadila et ADOBES André : Secondary fouling and TSP flow blockage: a model based on transport equations. *Dans Steam Generator Secondary Side Management Conference*, San Antonio, mars 2010. EPRI SGMP.

- MOLEIRO Edgar, OUKACINE Fadila, PRUSEK Thomas et GIRARD Sylvain : Réunion de travail sur le transfert thermique au travers de dépôts d'encrassement. Compte-rendu non publié, mai 2011.
- MONCHECOURT Dorian : Méthodologie d'évaluation du taux de réduction de section des passages foliés en PE 9 des GV de type 68/19. Rapport technique EDIAT110416/A, EDF, juillet 2011.
- MORRIS Max D. : Factorial sampling plans for preliminary computational experiments. *Technometrics*, pages 161–174, 1991.
- MORVAN Jacques et ADOBES André : Proposition d'une loi reliant pourcentage et volume de colmatage dans le cas d'un colmatage de bord d'attaque pour un passage folié de GV de type 51B. Rapport technique Compte-rendu I84-08-41, EDF, octobre 2008.
- NICOLI Alexandre : Palier REP 900 MWe – CP1 / CP2 – modèle GVRP 55 / 19 compléments à la note de synthèse de qualification du code ANETH. Rapport technique PEEG-F DC 10204, Areva, avril 2009.
- NICOLI Alexandre : Projet RCE – description de la modélisation physique du code ANETH version 4.07. Rapport technique PEEG-F DC 10097, Areva, septembre 2010a.
- NICOLI Alexandre : Projet RCE – description des corrélations du code ANETH version 4.07. Rapport technique PEEG-F DC 10099, Areva, octobre 2010b.
- NINET Julien : Analyse des premiers résultats fournis par l'algorithme stochastique d'évaluation du colmatage des GV. Non publié, 2009.
- NINET Julien, DENEUX Olivier et GIRARD Sylvain : évaluation du colmatage par la méthode du NGL dynamique : validation et application au parc de GV 51B, 51BI et 47/22. Rapport technique H-P1C-2012-00688-FR, EDF, avril 2012.
- NINET Julien et FAVENNEC Jean-Melaine : Determination of applicability of EDF steam generator monitoring algorithm to pressurized water reactors worldwide. Rapport technique 1021079, EPRI, décembre 2010.
- NINET Julien, FAVENNEC Jean-Melaine et LARRIGNON Guillaume : Estimation du taux de colmatage des générateurs de vapeur par analyse du comportement dynamique du NGL : étude de faisabilité concernant l'applicabilité de la méthode aux tranches REP hors EDF. (étude réalisée pour l'EPRI). Rapport technique H-P1C-2010-03382-EN, EDF, 2010.
- PASQUALE Lydia : Palier R.E.P. 900 MWe, générateur de vapeur modèle g.v. 51 b, paramètres de fonctionnement calculés avec le code PERFGV. Rapport technique NEEG-F DC 10191, Areva, novembre 2008.
- PAUL Nicolas : Estimation du colmatage des générateurs de vapeurs par analyse des signaux issus des sondes SAX. Rapport technique H-P1B-2012-01666-FR, EDF, à paraître.

- PETZOLD Linda R. : Description of DASSL: a differential/algebraic system solver. Rapport technique SAND-82-8637; CONF-820810-21, Sandia National Laboratories, Livermore, CA (USA), septembre 1982.
- PIEROTTI G. : Perte de charge et transfert thermique dans les faisceaux de tubes. Rapport technique HT-37/90.16A, EDF, décembre 1990.
- PIGEON L. : Dossier de spécifications de l'intégration de l'EP RGL4 dans SAPEC : calibrage des grappes grises. Rapport technique HP-13/97/019/A, EDF, novembre 1997.
- PILLET Joël, NIMAMBEG Nasser et PINTO Christophe : Résultats des essais monophasiques sur la maquette P2C pour la détermination des pertes de charge induites par le colmatage. Rapport technique H-I84-2009-02838, EDF, mai 2010.
- PITNER Patrice, GUIVARCH Michel et LACROIX Régis : Doctrine de maintenance – le faisceau tubulaire du générateur de vapeur. Rapport technique D4550.01-10/5646/B, EDF, décembre 2011.
- PRASSER Horst-Michael : Advanced two-phase flow instrumentation. *Dans 28th Short Courses - Modelling and Computation of Multiphase Flows Swi* (2011).
- PRUSEK Thomas : *Modélisation et simulation numérique du colmatage à l'échelle du sous-canal dans les générateurs de vapeur*. Thèse de doctorat, Université d'Aix et Marseille (AMU), à paraître.
- PRUSEK Thomas, MOLEIRO Edgar, OUKACINE Fadila, OUARDIA Ouardia et ADOBES André : Tube fouling and tube support plate blockage in steam generators: inverse method for the estimation of unobservable parameters in a deposit model. *Dans 19th International Conference on Nuclear Engineering (ICONE19, Tokyo, mai 2011*.
- PUJET Stéphane : Encrassement secondaire des GV : modélisation des dépôts de particules et calculs de démonstration. Rapport technique HI-84/01/010/A, EDF, 2001.
- PUJET Stéphane : Analyse du REX et des études de laboratoire sur le transport et le dépôt et produits de corrosion. Rapport technique HI-84/02/008/A, EDF, mai 2002.
- RAMSAY James O. et SILVERMAN Bernard Walter : *Functional Data Analysis*. Springer, 2^{ème} édition, 2005.
- RASSOKHIN N. G. : Thermal conduction of deposits of iron oxides. *Teploenergetika*, 20 (9):12–15, 1973.
- REGALDO C. : Travaux de recherche et développement effectués sur les dispositifs de séparation et de séchage des générateurs de vapeur des type 51, 68/19 et N4. Rapport technique E SE EC 83-24 A, Cre/CJ, EDF, novembre 1983.
- REGALDO C., de SURGY J. et SOGORB B. : Générateurs de vapeur des tranches à eau légère (conception – évolution). Rapport technique DR001673, EDF, avril 1984.

- REUSS Paul : *Précis de neutronique*. EDP Sciences, janvier 2003.
- RISSON Élodie et GARBAY Éric : Synthèse des expertises de tubes extraits de GV de 1997 à 2009. Rapport technique EDLMEM090373, EDF, juillet 2009.
- RIVEROL C. et NAPOLITANO V. : Estimation of fouling in a plate heat exchanger through the application of neural networks. *Journal of Chemical Technology and Biotechnology*, 80(5):594–600, 2005.
- RIVOIRARD Jacques : *Concepts et méthodes de la géostatistique*. Numéro C-158 dans Documentation du centre de géostatistique. École des mines de Paris, octobre 1995.
- ROBERT Christian P. et CASELLA George : *Monte Carlo statistical methods*. Springer Verlag, 2004.
- ROHSENOW Warren M., HARTNETT James P. et CHO Young I., éditeurs. *Handbook of Heat Transfer*. McGraw-Hill, third édition, 1998.
- ROMEYER Bertrand : Hypothèse d'étude de colmatage enveloppe des examens effectués à St Alban 1. Rapport technique ENREEC070385, EDF, décembre 2007.
- ROY : Fonctionnement des générateur de vapeur. Rapport technique E-SE/TH 84-84 A, EDF, octobre 1985.
- RUMMENS Helena Elisabeth Cornelia : *The thermalhydraulics of tube-support fouling in nuclear steam generators*. Thèse de doctorat, Carleton university, Canada, septembre 1999.
- RUMMENS Helena Elisabeth Cornelia, ROGERS J. T. et TURNER Carl W. : The thermal hydraulics of tube support fouling in nuclear steam generators. *Nuclear Technology*, 148(3):268–286, 2004.
- SALTELLI Andrea : Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.
- SALTELLI Andrea et ANNONI Paola : How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517, 2010.
- SALTELLI Andrea, ANNONI Paola, AZZINI Ivano, CAMPOLONGO Francesca, RATTO Marco et TARANTOLA Stefano : Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.
- SALTELLI Andrea, RATTO Marco, ANDRES Terry, CAMPOLONGO Francesca, CARIBONI Jessica, GATELLI Debora, SAISANA Michaela et TARANTOLA Stefano : *Global sensitivity analysis: the primer*. Wiley Online Library, 2008.
- SAPORTA Gilbert : *Probabilités, analyse des données et statistique*. Editions Technip, 2006.

- SCHWARZ Thomas : Heat transfer and fouling behaviour of siemens PWR steam generators – long-term operating experience. *Experimental thermal and fluid science*, 25(5):319–327, 2001.
- SINGHAL A.K. et SRIKANTIAH G. : A review of thermal hydraulic analysis methodology for PWR steam generators and ATHOS3 code applications. *Progress in Nuclear Energy*, 25(1):7–70, 1991.
- SOBOL' Il'ya Meerovich : Sensitivity estimates for nonlinear mathematical models, in *Matem. Modelirovanie*, 2 (1)(1990) 112-118. *English Transl.: MMCE*, 1(4), 1993.
- SOBOL' Il'ya Meerovich : On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*, 47:103–112, 1998.
- SOBOL' Il'ya Meerovich : Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55:271–280, 2001.
- SRIDHARAN A., HOCHREITER L.E., CHEUNG F.B. et WEBB RL : Effect of chemical cleaning on steam generator tube performance. *Heat Transfer Engineering*, 23(1):38–47, 2002.
- SRIKANTIAH G. et CHAPPIDI P. R. : Particle deposition and fouling in PWR steam generators. *Nuclear engineering and design*, 200(1):285–294, 2000.
- STINDEL Marc : Méthode d'évaluation par examen télévisuel du taux de colmatage de la plaque entretoise supérieure des GV 900 MW : évolutions, incertitudes et retour d'expérience. Rapport technique EDIAT070539, EDF, juillet 2007.
- 28th Short Courses - Modelling and Computation of Multiphase Flows*, février 2011. Swiss Federal Institute of Technologie (ETH).
- TEDFORD Nathan : Steam generator management program: dynamic analysis of a steam generator – Part 1: method development, steady-state verification, and steady-state validation. Rapport technique 1025134, EPRI, février 2012.
- TESSIER Jean-François et PETIT D. : Rapport d'expertise : analyse physico-chimique de boues GV provenant du CNPE de Delta – tranche 4. Rapport technique EDLCHM060441, EDF, septembre 2006.
- TINCQ Didier et DAVID Franck : THYC-ECHANGEUR version 3, application aux générateurs de vapeur, fiches de qualification – partie 1. Rapport technique HT-33/94/029/A, EDF, décembre 1994.
- TURNER C. W., KLIMAS S. J. et BRIDEAU M. G. : Thermal resistance of steam-generator tube deposits under single-phase forced convection and flow-boiling heat transfer. *The Canadian Journal of Chemical Engineering*, 78(1):53–60, 2000.

- TURNER Carl W., LINER Y. et CARVER M.B. : Modelling magnetite particle deposition in nuclear steam generators and comparisons with plant data. *Dans 2nd Steam Generator and Heat Exchanger Conference*, pages 51–64. Société Nucléaire Canadienne, 1994.
- UHLE Jennifer Lee : *Boiling Heat Transfer Characteristics of Steam Generator U-Tube Fouling Deposits*. Thèse de doctorat, Massachusetts Institute of Technology, 1997.
- UL HAQ Inam, CINOSI Nicolas, BLUCK Michael, HEWITT Geoffrey et WALKER Simon : Modelling heat transfer and dissolved species concentrations within PWR crud. *Nuclear Engineering and Design*, 241(1):155–162, 2011.
- VARRIN R. D., J., SILVA E., BARSHAY S., LURIE S., GAY J., ELTER J., ROTHAM M., ESPOSITO J., BARKATT A. et LABUDA E. : Characterization of PWR steam generator deposits. Rapport technique TR-106048, EPRI, février 1996.
- VIAUD C. et AUGUERES M.J. : Le modele de representation de générateur de vapeur et poste d'eau destine a l'étude de la régulation des générateurs de vapeur LILLIGV 1 .0 – Note de principe. Rapport technique HP-35/91.10, EDF, juin 1991.
- VIVIANI Roberto, GRÖN Georg et SPITZER Manfred : Functional principal component analysis of fMRI data. *Human Brain Mapping*, 24(2):109–129, 2005.
- WHITE G. A., KREIDER M. A. et VARRIN R. D., J. : A global fouling factor methodology for analyzing steam generator thermal performance degradation. Rapport technique, Knolls Atomic Power Lab., Schenectady, NY (United States), 1998.
- YADIGAROGLU Georges : Basic models for two-phase flows. *Dans 28th Short Courses - Modelling and Computation of Multiphase Flows Swi* (2011).
- YADIGAROGLU Georges : Basic models for two-phase flows. Exposé, février 2011b. 28th Short Courses - Modelling and Computation of Multiphase Flows.

Glossaire

47/22

Type de générateur de vapeur équipant des tranches du palier 900 MW. Il n'a pas été étudié dans cette thèse. 60

51 B

Type de générateur de vapeur équipant des tranches du palier 900 MW étudié dans cette thèse. D'autres modèles analogues (51 A, 51 BI et 51 M) lui ont été assimilés. 19–21, 24, 25, 29, 30, 42, 60, 62, 63, 104, 183, 191, 219, 226, 231, 233

68/19

Modèle de générateur de vapeur équipant des tranches du palier 1300 MW. Il n'a pas été étudié dans cette thèse. 60, 225, 226

73/19

Type de générateur de vapeur équipant des tranches du palier N 4 (voir ce terme). Il n'a pas été étudié dans cette thèse. 225, 226

Alpha

Nom de code donné à l'une des centrales étudiées dans cette thèse. Elle est conditionnée à haut pH (voir Bravo, Charlie et Delta). 20, 101, 105, 205, 207, 209–211, 249

ANETH

« Analyse numérique d'un échangeur thermique ». Modèle de générateur de vapeur (voir chap. 4). 61, 62

ANOVA

Acronyme anglais de *analysis of variance* signifiant « analyse de variance ». Caractérise notamment la décomposition présentée dans la section 7.2.1. 116

ATHOS/SGAP

« *Analysis of the thermal hydraulics of a steam generator/Steam generator analysis package* », soit en français : « Analyse de la thermohydraulique des générateurs de vapeur/Paquet pour l'analyse des générateurs de vapeur ». Modèle de générateur de vapeur (voir chap. 4). 62, 224

bootstrap

Utilisé ici comme substantif et comme adjectif, ce terme anglais désigne des méthodes d'inférence fondées sur un rééchantillonnage de l'échantillon utilisé pour l'estimation. Le *bootstrap* a été utilisé dans cette thèse pour établir des intervalles de confiance pour les indices de sensibilité (voir chap. 7), et pour estimer les dimensions des espaces de réduction de dimension efficace (voir chap. 9). Le mot *bootstrap* signifie

en français « tirants de botte ». Ce nom fait référence à l'expression *to pull oneself up by one's bootstraps*, signifiant « se hisser en se tractant par ses tirants de botte ». Elle serait une allusion à un épisode des *Aventures du baron de Münchhausen* écrit par Rudolf Erich RASPE, au cours duquel le baron se sort d'un marais en se tirant par sa queue-de-cheval. 112, 118, 121, 122, 135–137, 146, 154, 157, 158, 165, 194, 202, 236, 237, 240, 243, 244, 247

BOUTHYC

Modèle de formation des dépôts d'encrassement utilisant THYC (voir l'annexe B). 169, 218, 224, 227, 234

Bravo

Nom de code donné à l'une des centrales étudiées dans cette thèse. Elle est conditionnée à haut pH (voir Alpha, Charlie et Delta). 20, 30, 36, 38, 101, 105, 205, 207, 212, 250

CANDU

Acronyme de « Canada deuterium uranium ». Type de réacteur nucléaire à uranium naturel à l'eau lourde pressurisée, développé par le groupe *Atomic energy of Canada limited*. 223

CATHARE

« Code avancé de thermohydraulique pour les accidents sur les réacteurs à eau ». Code thermohydraulique dont est issu un modèle de générateur de vapeur (voir chap. 4). 61, 62

CEA

Commissariat à l'énergie atomique et aux énergies alternative (<http://www.cea.fr/>, anciennement C.E.A., sigle de « Commissariat à l'énergie atomique »). Institut public français de recherche sur les énergies bas carbone (nucléaire et renouvelables), la défense, les technologies pour l'information et les technologies pour la santé. 59, 62

Charlie

Nom de code donné à l'une des centrales étudiées dans cette thèse. Elle est conditionnée à bas pH (voir Alpha, Bravo et Delta). 20, 27, 42, 43, 50, 66, 101, 105, 152, 153, 175, 184, 191, 205, 207, 209, 213, 214, 221, 222, 242, 250

COLMATHYC

Modèle de formation des dépôts de colmatage utilisant THYC (voir l'annexe B). 169, 170, 175, 234

DASSL

Acronyme anglais de *Differential/algebraic system solver* signifiant « méthode numérique de résolution des systèmes algébriques d'équations différentielles ». 73

Delta

Nom de code donné à l'une des centrales étudiées dans cette thèse. Elle est conditionnée à bas pH (voir Alpha, Bravo et Charlie). 14, 20, 27, 28, 30, 32, 44, 67, 68, 101, 105, 152, 153, 171, 184, 191, 205, 207, 209, 215, 216, 231, 251

downcomer

Mot anglais signifiant ici « canal de descente » et désignant l'espace entre la virole externe des générateurs de vapeur et la chemise du faisceau, par lequel l'eau alimentaire et l'eau liquide retenue par les dispositifs de séparation et de séchage sont amenées en bas du générateur de vapeur (voir *riser*). 19, 20, 22, 23, 39, 72, 94, 97, 100, 104, 145, 146, 169, 231

DRIF-BI

Modèle de générateur de vapeur (voir chap. 4). 64, 65, 71

Dymola

Logiciel utilisé pour interpréter le code du modèle de générateur de vapeur écrit en langage Modelica. 61, 73, 98, 99

EDFLib

Bibliothèque de modules écrits en langage Modelica pour la modélisation des centrales de production d'électricité. Cette bibliothèque a servi de point de départ pour le développement de la version 2 du modèle de générateur de vapeur. Une autre bibliothèque baptisée ThermoSysPro est désormais utilisée à EDF. 73, 74, 79

e. d. r.

Abréviation anglaise de *effective dimension reduction* soit « [de] réduction de dimension efficace » en français. Ce terme qualifie les directions de projection obtenues avec la SIR (voir ce terme), ainsi que le sous-espace qu'elles engendrent et la dimension de ce sous-espace. 132–140, 156, 157, 159–168, 201, 202, 204, 226, 232, 237–241, 244–248, 252

EPR

Marque désignant type de réacteur récent conçu par le groupe Areva (<http://www.areva.com>), d'une puissance nominale électrique de 1650 MW. Cette marque est dérivée du sigle E.P.R., signifiant initialement *European pressurized reactor* (Réacteur européen à eau pressurisée) puis *Evolutionary power reactor* (Réacteur de puissance de nouvelle génération). 17, 18

EPRI

« *Electric power research institute* », soit en français « Institut de recherche sur l'énergie électrique » (<http://http://my.epri.com>. Consortium international à but non lucratif, de recherche sur la production d'électricité, dont EDF fait partie). 60, 62

examen télévisuel

Technique d'inspection de la plaque entretoise supérieure à l'aide d'une caméra introduite par le haut des générateurs de vapeur, afin d'en estimer l'état de colmatage. 24, 25, 28–30, 33, 35, 37–40, 42, 44, 49, 66, 67, 142, 169–171, 175, 205, 207–209, 218, 224, 229, 230, 235

FAST

Acronyme anglais de *Fourier amplitude sensitivity test* ce que l'on pourrait traduire

par « Analyse de sensibilité par les amplitudes de Fourier » en français. Cette méthode d'analyse de sensibilité est évoquée dans le chapitre 7. 114, 118

g. v.

Générateur de vapeur (voir chap. 2). 14, 15, 17–30, 32, 35–45, 49, 50, 55, 57, 59–69, 71–76, 80–83, 87–94, 96, 97, 99, 100, 104, 105, 114, 118, 127, 131, 132, 142, 145–147, 154, 165, 168, 169, 171, 177, 181–184, 191, 202, 204–209, 218–231, 233, 235

I.A.P.W.S.

« *International association for the properties of water and steam* » soit en français l'« Association internationale pour les propriétés de l'eau et de la vapeur » (<http://www.iapws.org/>). Association internationale à but non lucratif dédiée à l'étude des propriétés de l'eau et de la vapeur, notamment thermodynamiques et dans un contexte industriel. 73

I.F. 97

« *Industrial formulation 1997* » soit en français « Formulation industrielle 1997 ». Tables des propriétés thermodynamiques de l'eau proposée par l'I.A.P.W.S. 73, 79, 87

I.R.S.N.

« Institut national de radioprotection et de sûreté nucléaire » (<http://www.irsn.fr>). Institut public français de recherche et d'expertise sur les risques nucléaires et radiologiques. 59, 62

LEDA

Code thermohydraulique dont est issu un modèle de générateur de vapeur (voir chap. 4). 64

MANTA

Nom donné en avril 1997 au « Nouveau code de fonctionnement d'ensemble », abrégé N.F.C.E. Code thermohydraulique dont est issu un modèle de générateur de vapeur (voir chap. 4). 61, 62, 66

méthode de diagnostic à profil choisi

Méthode de diagnostic du colmatage par analyse du comportement dynamique des générateurs de vapeur existante au début de cette thèse. 42, 44, 60, 64, 142, 143, 145, 147, 161, 164, 168, 182–184, 187, 189, 197, 202, 229, 252

méthode de diagnostic par réduction de dimension

Nouvelle méthode de diagnostic du colmatage par analyse du comportement dynamique des générateurs de vapeur proposée dans thèse. 142, 189, 230

Modelica

Langage de programmation utilisé pour écrire le modèle de générateur de vapeur. 64, 71–74, 78, 79

Monte-Carlo (de)

Qualifie des méthodes d'estimation numérique par une approche stochastique. Le

terme, dû à Nicholas METROPOLIS, fait référence aux jeux de hasard des casinos de Monte-Carlo. 45, 46, 48, 118, 120

N4

Nom donné au palier des tranches nucléaires française d'une puissance nominale électrique de 1400 MW. 17, 18, 225

n. g. l.

Niveau gamme large. Mesure de pression différentielle entre le bas et le haut du générateur de vapeur. Elle est fortement impactée par le colmatage. 22, 23, 28, 35, 39–43, 45, 46, 50, 55, 57, 60, 64, 65, 67–71, 87, 96–98, 101, 104, 127, 129, 142, 143, 145–147, 150, 152–154, 156, 165, 168, 182–184, 187–189, 197, 201–203, 205, 208, 221, 225, 229–231, 249–251

ORLI

« Ouverture du réseau local industriel ». Application permettant d'accéder à l'historique des mesures d'exploitation des tranches nucléaires françaises. 20, 22, 39, 96

P2C

Expérience sur une maquette de plaque entretoise à échelle réduite visant à déterminer la perte de charge singulière engendrée par différents taux de colmatage. 36, 37, 67, 93, 231

PERFGV

Modèle de générateur de vapeur (voir chap. 4). 60, 69, 219

pH

Mesure de l'activité chimique des ions hydrogène en solution. Elle caractérise l'acidité de la solution. Une solution dite acide, si son pH est inférieur à 7, neutre si son pH est égal à 7 et basique si son pH est supérieur à 7. 26, 206, 207

quasi-Monte-Carlo (de)

Qualifie une méthode de Monte-Carlo (voir ce terme) lorsque celle-ci est appliquée en utilisant des nombres quasi-aléatoires, issus par exemple d'une séquence de Sobol' (voir section 7.3.3). 112, 114, 118, 120, 121

R.G.L. 4 (essai)

Essai réalisé périodiquement pour mettre à jour la courbe de calibration des grappes de contrôle faiblement absorbantes, dites grises. Le transitoire de puissance résultant est exploité pour le diagnostic du colmatage. 41–43, 45, 50, 65, 70, 76, 88, 97, 98, 142, 182–184, 188–191, 204, 205, 221, 222

riser

Mot anglais signifiant ici « canal de montée » et désignant l'espace cylindrique dans lequel l'eau du circuit secondaire s'échauffe au contact des tubes dans lesquels circule l'eau du circuit primaire. Le *riser* est alimenté en eau par le bas, par le *downcomer* (voir ce terme). 19, 20, 22–24, 39, 57, 59, 63, 65, 66, 71, 72, 88, 89, 91, 93, 94, 96, 97, 99, 100, 104, 145, 146, 231

SEPTEN

« Service d'études et projets thermiques et nucléaires ». Département d'ingénierie d'EDF. 61, 62

SIR

Acronyme anglais de *sliced inverse regression*, soit « régression inverse par tranches » en français. Cette méthode de réduction de dimension est présentée dans le chapitre 9. 15, 131–140, 156, 157, 159, 164, 165, 201, 202, 226, 230, 236, 252

taux de colmatage uniforme équivalent

Étant donné un ensemble de trous foliés colmatés à des taux variables, leur taux de colmatage uniforme équivalent est le taux de colmatage qu'ils auraient s'ils étaient tous colmatés identiquement en engendrant la même perte de charge. 26, 36–38, 67, 174, 181, 205–207, 209, 230

ThermoSysPro

Bibliothèque de modules écrits en langage Modelica pour la modélisation des centrales de production d'électricité (<http://www.modelica.org/libraries/thermosyspro>). 73

THYC

« Thermohydraulique des composants ». Code thermohydraulique dont est issu un modèle de générateur de vapeur (voir chap. 4). 32, 60, 62–64, 66, 70, 85, 89, 91, 96, 98–105, 107–110, 169, 170, 177, 229, 234

vena contracta

Littéralement « veine contractée » en latin. Point dans un écoulement où l'aire de la section de passage est minimale. Le terme serait dû à Evangelista TORRICELLI. Il se réfère ici au mécanisme de formation des dépôts de colmatage dominé par le décollement de couche limite en amont des trous foliés. Un modèle décrivant ce phénomène est décrit dans l'annexe B. 66, 169–171, 177, 178, 180, 181, 190, 201, 231, 232, 234, 235, 243–248

version 2

Version initiale du modèle de générateur de vapeur utilisé dans cette thèse. Les améliorations réalisées au cours de la thèse ont abouti à la version 7. 64–66, 68, 69, 71, 72, 87, 93, 94, 96, 147, 155, 189

version 7

Version finale du modèle de générateur de vapeur utilisé dans cette thèse, développée à partir de la version 2. 64–66, 68, 70, 71, 73, 87, 99–105, 107–110, 189, 197

Westinghouse

Entreprise américaine spécialisée dans la conception et la fabrication d'assemblages de combustibles nucléaires, les services spécialisés à l'industrie et l'ingénierie associée, et la conception et la construction de nouvelles centrales (<http://www.westinghousenuclear.com/>). 90

Index des notations

La nomenclature qui suit est divisée en 8 paragraphes :

- grandeurs physiques représentées par des caractères latins, p. 316 ;
- grandeurs physiques représentées par des caractères grecs, p. 318 ;
- indices et exposants des grandeurs physiques, p. 319 ;
- opérations sur des grandeurs physiques, p. 320 ;
- grandeurs statistiques représentées par des caractères latins, p. 320 ;
- grandeurs statistiques représentées par des caractères grecs, p. 322 ;
- indices et exposants des grandeurs statistiques, p. 322 ;
- opérations sur des grandeurs statistiques, p. 323.

On trouvera pour chaque symbole une description succincte, l'unité associée et la page où la notation est utilisée pour la première fois.

Symbole	Signification	Unité	Page
Physique, caractères latin			
A	Aire d'une section de canal	m^2	78
A_{colm}	Aire de la section efficace d'un dépôt de colmatage	m^2	26
A_{ech}	Aire de la surface d'échange externe du faisceau de tubes	m^2	221
$\mathbf{a}(\omega)$	Vecteur dépendant des champs calculés avec THYC et de grandeurs calculées par le module COLMATHYC		170
A_{passe}	Aire de la section de passage restante d'un trou folié colmaté	m^2	26
A_{trou}	Aire d'un trou folié sain	m^2	26
$\mathbf{b}(\omega)$	Vecteur dépendant des champs calculés avec THYC et de grandeurs calculées par le module COLMATHYC		170
C_0	Coefficient de distribution du modèle de flux de dérive		92
C_p	Capacité calorifique à pression constante	$J\ ^\circ C^{-1}\ kg^{-1}$	79
D	Coefficient de diffusion des particules	$m^2\ s^{-1}$	234
d	Diamètre des particules de magnétite	m	169
D_c	Distance caractéristique	m	77
d_{hydro}	Diamètre hydraulique	m	90
d_{int}	Diamètre interne des tubes	m	80
d_{ext}	Diamètre externe des tubes	m	90
d_{th}	Diamètre thermique	m	90
e	Énergie interne par unité de masse	$J\ kg^{-1}$	73

f	Coefficient de friction		79
F	Force de friction exercée par la paroi sur le fluide du fait de son mouvement	N	79
g_z	Composante verticale du champ de gravitation (négative)	m s^{-2}	77
h	Enthalpie (pondérée en masse) par unité de masse	J kg^{-1}	76
\bar{h}	Enthalpie de mélange (pondérée en volume) par unité de masse	J kg^{-1}	73
j	Vitesse superficielle locale (flux de débit volumique)	m s^{-1}	91
\mathbf{J}	Terme de convection et de diffusion des particules	$\text{kg m}^2 \text{s}^{-1}$	234
k	Conductivité thermique	$\text{W m}^{-1} \text{°C}^{-1}$	79
K	Nombre d'évènements caractéristiques		189
K_c	Énergie caractéristique	J	79
k_{geo}	Constante géométrique, fonction affine du taux de colmatage		235
m_d	Masse de particules déposée par unité de surface	kg m^{-2}	234
\dot{M}	Débit massique	kg s^{-1}	73
\dot{m}	Flux de masse	$\text{kg s}^{-1} \text{m}^{-2}$	78
M_{th}	Coefficient de calage du transfert thermique dans le <i>riser</i>		96
$\dot{M}_{vap}(t_0)$	Débit de vapeur initial	kg s^{-1}	184
$M_{\bar{v}c}$	Coefficient de calage de la vitesse moyenne de dérive du gaz dans les cyclones		96
$M_{\bar{v}r}$	Coefficient de calage de la vitesse moyenne de dérive du gaz dans le <i>riser</i> (cintre exclu)		96
Nu	Nombre de Nusselt		80
p	Pression	bar	73
p	Nombre de pas de temps d'échantillonnage		184
pas	Pas du faisceau de tubes	m	31
P_c	Périmètre chauffé	m	78
P_i	Périmètre de l'interface entre les phases	m	85
P_m	Périmètre mouillé	m	78
Pr	Nombre de Prandtl		80
Q	Débit volumique	kg m^{-3}	91
q''	Flux thermique au travers de la paroi vers le fluide considéré	W m^{-2}	77
q''_{ik}	Flux thermique au travers de l'interface vers la phase indicée k	W m^{-2}	85
R''	Résistance thermique	$\text{°C m}^2 \text{W}^{-1}$	220
Re	Nombre de Reynolds		77
$r\hat{u}$	Largeur du rû d'eau séparant les deux jambes.	m	31
S	Rapport des vitesses des phases		84
T	Durée d'un transitoire	s	184

t	temps	s	76
T_1	Température du fluide primaire	°C	220
T_2	Température du fluide primaire	°C	220
$T_{alim}(t_0)$	Température initiale de l'eau alimentaire	°C	184
T_{ch}	Température de l'eau primaire, en branche chaude		221
T_d	Temps de déposition total	j	170
T_f	Température du fluide	°C	79
T_{fr}	Température de l'eau primaire, en branche froide		221
T_p	Température de la paroi	°C	80
$T_{prim}(t_0)$	Température initiale de l'eau du circuit primaire (en branche chaude)	°C	184
T_{sat}	Température de saturation		91
T_{vap}	Température de la vapeur sortante, supposée à saturation		221
\mathbf{t}	Vecteur contenant les instants d'échantillonnage	s	184
\tilde{v}_{gj}	Vitesse moyenne de dérive du gaz	m s^{-1}	92
v	Vitesse	m s^{-1}	76
V	Volume du dépôt de colmatage	m^3	235
V_c	Volume caractéristique d'un dépôt de colmatage	m^3	235
v_d	Vitesse de déposition des particules	m s^{-1}	235
v_{gj}	Vitesse locale de dérive du gaz	m s^{-1}	92
W	Puissance totale transférée du circuit primaire au circuit secondaire		221
x	Taux de colmatage		26
x_{max}	Taux de colmatage asymptotique de la corrélation entre taux de colmatage et volume de dépôt		235
y	Fonction du temps représentant la valeur du niveau gamme large	bar	184
\mathbf{y}	Vecteur contenant les valeurs échantillonnées du niveau gamme large	bar	184
z	Position selon la direction verticale (cote)	m	76
Physique, caractères grecs			
α	Taux de vide		83
Γ_k	Débit entrant dans la phase indiquée k , par unité de volume	$\text{kg s}^{-1} \text{m}^{-3}$	85
$\Delta R''$	Facteur d'encrassement global	°C m ² W	221
Δt	Temps d'observation	s	82
$\Delta \dot{M}_{vap}/\Delta t$	Pente du débit de vapeur	kg s^{-2}	184
$\Delta T_{alim}/\Delta t$	Pente de la température de l'eau alimentaire	°C s ⁻¹	184
$\Delta T_{prim}/\Delta t$	Pente de la température de l'eau du circuit primaire (en branche chaude)	°C s ⁻¹	184
ϵ_a	Fraction massique de particule dans l'eau alimentaire		169

ζ_{adm}	Coefficient de perte de charge à l'admission dans le <i>riser</i>		98
ζ_{dc}	Coefficient de perte de charge dans le <i>downcomer</i>		98
η	Coefficient de transfert thermique	$\text{W m}^{-2} \text{ } ^\circ\text{C}^{-1}$	80
κ	Critère de convergence numérique de la méthode d'estimation du paramètre ω		170
λ	Grandeur quelconque		82
μ	Viscosité dynamique	Pa s	77
π_c	Porosité des dépôts de colmatage		169
π_e	Porosité des dépôts d'encrassement		169
π_f	Porosité fluide dans le modèle THYC		234
ρ	Masse volumique du fluide	kg m^{-3}	73
τ	Composante verticale du tenseur de contraintes visqueuses	N m^{-2}	76
ϕ	Fonction de déformation temporelle		187
Φ	Flux de particules déposées	$\text{kg m}^{-2} \text{ s}^{-1}$	234
ϕ_g	Fonction indicatrice de la phase gazeuse		83
Φ_{ls}^2	Multiplicateur diphasique du coefficient de perte de charge		88
χ	Titre (en vapeur) du mélange		83
Ψ	Transfert volumique de particules vers la paroi	$\text{kg m}^{-3} \text{ s}^{-1}$	234
ω	Paramètre non observable du modèle de déposition par <i>vena contracta</i>		170
ε	Fraction massique locale de particules		234
ζ_{pe}	Coefficient de perte de charge d'une plaque entretoise		93

Physique, indices et exposants

\cdot_c	Grandeur caractéristique		77
\cdot^*	Grandeur de référence		184
\cdot_f	Grandeur relative au fluide		79
\cdot_g	Grandeur relative à la phase gazeuse		83
\cdot_l	Grandeur relative à la phase liquide		83
\cdot_{lg}	Grandeur relative au mélange de liquide et de gaz		88
loc	Grandeur locale		220
\cdot_{ls}	Grandeur relative au liquide seul		88
\cdot_p	Grandeur relative à la paroi		80
\cdot''	Grandeur définie par unité de surface		77
\cdot_{sat}	Grandeur à saturation		91
\cdot_z	Composante selon la direction verticale, orientée vers le haut		76

Physique, opérateurs

$\hat{\cdot}$	Opérateur de moyenne temporelle	82
$\langle \cdot \rangle$	Opérateur de moyenne transversale	83
$\langle \cdot \rangle_k$	Opérateur de moyenne transversale sur la surface occupée par la phase indiquée k	83
$\frac{D \cdot}{Dt}$	Dérivée particulière	78
$\frac{\partial \cdot}{\partial t}$	Dérivée partielle par rapport au temps	74
$\frac{\partial \cdot}{\partial z}$	Dérivée partielle par rapport à la cote	74
\cdot_i	Grandeur relative à l'interface entre les phases	85
$\tilde{\cdot}$	Grandeur adaptée à l'aide d'une fonction de déformation temporelle	187

Statistique, caractères latin

\mathbf{A}	Échantillon des variables d'entrée (individus en lignes, variables en colonnes)	118
$\mathbf{A}^{(b)}$	Échantillon <i>bootstrap</i> obtenu à partir de \mathbf{A}	121
$\mathbf{A}_B^{(i)}$	Matrice obtenue en substituant à la $i^{\text{ème}}$ colonne de l'échantillon \mathbf{A} , la $i^{\text{ème}}$ colonne de l'échantillon \mathbf{B}	119
\mathbf{A}_k	$k^{\text{ème}}$ ligne (individu) de l'échantillon \mathbf{A}	118
\mathbf{B}	Échantillon des variables d'entrée (individus en lignes, variables en colonnes)	119
c_i	$i^{\text{ème}}$ direction principale	124
\mathbf{C}	Matrice contenant en colonnes les directions principales	125
\mathbf{C}_k	Matrice constituée en colonnes des k premières directions principales	126
d_i	Distance entre la réponse mesurée et la réponse simulée pour la particule indiquée i	47
$E(\hat{w}_1, \dots, \hat{w}_n)$	Entropie de Shannon d'un ensemble de particules	48
E_{lim}	Entropie limite de rééchantillonnage	48
F	Distribution de probabilité	121
f	fonction mathématique représentant le modèle	115
f_0	Moyenne de la sortie du modèle (premier terme de la décomposition ANOVA)	116
$f_{i\dots j}$	Terme de la décomposition ANOVA, fonction des variables d'indices i, \dots, j	116
g	Fonction de lien arbitraire du modèle de la SIR	132
H	Nombre de tranches (pour la SIR)	135
K	Dimension e. d. r.	132
k	Nombre d'itérations du filtre particulière	48

\mathbf{m}	Vecteur représentant une réponse mesurée		47
\mathbf{M}_j	Matrice constituée en colonne des j premières directions e. d. r.		136
\mathbf{M}	Matrice inversible de dimension égale au nombre de variables d'entrée		133
\mathbf{m}_h	Moyenne des variables d'entrée sur un échantillon pour la tranche indiquée h		135
N	Nombre d'estimation unitaires lors d'un examen télévisuel		30
n	Nombre de particules de l'ensemble		48
N_b	Nombre de réplifications <i>bootstrap</i>		122
$n(h)$	Nombre de paires de points espacés d'une distance h		31
p	nombre de variables d'entrée		113
$prj \cdot$	Projecteur orthogonal (sur le sous-espace en indice)		132
$p(\cdot)$	Fonction représentant une densité de probabilité		46
p_d	Densité de probabilité des observations pour le filtre particulière		46
p_m	Densité de probabilité du modèle pour le filtre particulière		46
\mathbf{r}	Vecteur position d'un trou folié	m	31
R_j	Fonction de risque pour l'estimation de la dimension e. d. r.		137
$R_j^{(b)}$	Estimateur <i>bootstrap</i> de la fonction de risque pour la détermination de la dimension e. d. r.		137
\mathcal{S}	Sous-espace de réduction de dimension		132
\mathbf{s}_i	Vecteur représentant la réponse simulée pour la particule indiquée i		47
S_i	Indice de Sobol' d'ordre 1 de la variable indiquée i		115
$S_{i_1 \dots i_s}$	Indice de Sobol' d'ordre s des variables indiquées i_1, \dots, i_s		117
s_{ij}	Score du $i^{\text{ème}}$ individu d'un échantillon, associé à sa $j^{\text{ème}}$ composante principale		125
\mathbf{S}	Matrice des scores		125
S_i^{tot}	Indice de Sobol' total de la variable indiquée i		116
\mathcal{T}	Tranche (intervalle)		135
\mathbf{T}	Matrice représentant une transformation orthonormale		126
V	Variance totale de la sortie du modèle		115
\hat{v}_i	Estimation de la vraisemblance de la particule indiquée i		47
w_i	Poids attribué à la particule indiquée i pour l'échantillonnage d'importance		47

\widehat{w}_i	Poids de la particule indiquée i déduit des estimations de vraisemblance de l'ensemble	47
X	Variable aléatoire représentant un taux de colmatage.	31
\mathbf{X}	Vecteur aléatoire d'entrée	115
Y	Variable aléatoire de sortie	115
\mathbf{Z}	Vecteur aléatoire des variables d'entrées normalisées	134

Statistique, caractères grecs

α	Risque des intervalles de confiance	122
β_k	$k^{\text{ème}}$ direction e. d. r.	132
$\Delta\widehat{s}_1(t)$	Différence entre deux estimations successives du score selon la première composante principale, aux dates t et $t + 1$	204
$\Delta s_1(t)$	Différence entre deux valeurs vraies du score selon la première composante principale, aux dates t et $t + 1$	204
η_k	$k^{\text{ème}}$ direction e. d. r. normalisée	134
$\widehat{\theta}^{(b)}$	Réplication <i>bootstrap</i> de l'estimation $\widehat{\theta}$	121
$hestim$	Statistique estimée	121
λ	Coefficient de Lagrange	124
λ_i	$i^{\text{ème}}$ valeur propre	124
μ	Coefficient de Lagrange	125
ρ^2	Coefficient de corrélation linéaire	161
Σ	Matrice de covariance	124
$\sigma_{\Delta\widehat{s}}$	Écart type des différences de scores selon la première composante principale estimés successivement	204
ϕ	Application produisant une estimation	121
$\delta(\cdot)$	Distribution de Dirac	46
δ_i	Vecteur directeur unitaire de la $i^{\text{ème}}$ direction de l'espace	31
$\gamma(h)$	Variogramme en fonction de la distance h	m ² 31
β_{ij}	$j^{\text{ème}}$ direction principale, la variable de sortie étant le score selon la $i^{\text{ème}}$ composante principale	156

Statistique, indices et exposants

\cdot^*	Dans un intégrande, réalisation d'une variable aléatoire indépendante de sa première occurrence	119
$\cdot^{(b)}$	Se rapportant au <i>bootstrap</i>	121
\cdot_{-i}	Relatif à l'ensemble des variables d'entrée privé de celle indiquée i	116

Statistique, opérateurs

$\bar{\cdot}$	Moyenne des composantes d'un vecteur	47
$\hat{\cdot}$	Estimation	118
err·	Erreur type	122
$E(\cdot)$	Espérance d'une variable aléatoire	115
$E(\cdot \cdot)$	Espérance conditionnelle d'une variable aléatoire sachant une autre	115
$\ \cdot\ _2$	Norme euclidienne d'un vecteur	31
$\text{var}(\cdot)$	Variance d'une variable aléatoire	115
$\text{var}(\cdot \cdot)$	Variance conditionnelle d'une variable aléatoire sachant une autre	115
$\text{Vect}(\cdot)$	Sous-espace vectoriel engendré par la famille de vecteurs en argument	133
$z(\cdot)$	Quantile de la distribution gaussienne standard	122

Diagnostic du colmatage des générateurs de vapeur à l'aide de modèles physiques et statistiques

Résumé Les *générateurs de vapeur* sont d'imposants échangeurs de chaleur qui alimentent les turbines des centrales nucléaires à eau pressurisée. Au cours de leur exploitation, des dépôts d'oxydes s'y accumulent et obstruent progressivement des trous prévus pour le passage du fluide. Ce phénomène, appelé *colmatage*, pose des problèmes de sûreté. Une méthode de diagnostic est nécessaire pour optimiser la stratégie de maintenance permettant de s'en prémunir. La piste explorée dans cette thèse est l'analyse de la réponse dynamique des générateurs de vapeur lors de transitoire de puissance, à l'aide d'un modèle physique monodimensionnel.

Deux améliorations ont été apportées au modèle existant au cours de la thèse : la prise en compte des débits perpendiculaires à l'axe du générateur de vapeur et la modélisation du déséquilibre cinématique entre la phase liquide et la phase vapeur. Ces éléments ont ajouté des degrés de liberté permettant de mieux reproduire le comportement réel des générateurs de vapeur. Une nouvelle méthodologie de calage et de validation a alors été proposée afin de garantir la robustesse du modèle.

Le problème inverse initial était mal posé car plusieurs configurations spatiales de colmatage peuvent donner des réponses identiques. La magnitude relative de l'effet des dépôts suivant leur localisation a été évaluée par analyse de sensibilité avec la *méthode de Sobol'*. La dimension de la sortie fonctionnelle du modèle a au préalable été réduite par une *analyse en composantes principales*. Enfin, une méthode de réduction de dimension appelée *régression inverse par tranches* a été mise en œuvre pour déterminer des sous-espaces de projection optimaux pour le diagnostic. Une méthode de diagnostic plus robuste et mieux maîtrisée que celle existante a pu être proposée grâce à cette nouvelle formulation.

Mots clefs Générateur de vapeur, colmatage, diagnostic, modélisation physique, analyse de sensibilité, réduction de dimension.

Steam generators clogging diagnosis through physical and statistical modelling

Abstract *Steam generators* are massive heat exchangers feeding the turbines of pressurised water nuclear power plants. Internal parts of steam generators foul up with iron oxides which gradually close some holes aimed for the passing of the fluid. This phenomenon called *clogging* causes safety issues and means to assess it are needed to optimise the maintenance strategy. The approach investigated in this thesis is the analysis of steam generators dynamic behaviour during power transients with a monodimensional physical model.

Two improvements to the model have been implemented. One was taking into account flows orthogonal to the modelling axis, the other was introducing a *slip* between phases accounting for velocity difference between liquid water and steam. These two elements increased the model's degrees of freedom and improved the adequacy of the simulation to plant data. A new calibration and validation methodology has been proposed to assess the robustness of the model.

The initial inverse problem was ill posed: different clogging spatial configurations can produce identical responses. The relative importance of clogging, depending on its localisation, has been estimated by sensitivity analysis with the *Sobol' method*. The dimension of the model functional output had been previously reduced by *principal components analysis*. Finally, the input dimension has been reduced by a technique called *sliced inverse regression*. Based on this new framework, a new diagnosis methodology, more robust and better understood than the existing one, has been proposed.

Keywords Steam generators, clogging, diagnosis, physical modelling, sensitivity analysis, dimension reduction.

