Les modèles hydrologiques conceptuelssont-ils robustes face à un climat en évolution ? Diagnostic sur un échantillon de bassins versants français et australiens

Laurent Coron

▶ **To cite this version:**

Laurent Coron. Les modèles hydrologiques conceptuelssont-ils robustes face à un climat en évolution ? Diagnostic sur un échantillon de bassins versants français et australiens. Sciences de la Terre. AgroParisTech, 2013. Français. NNT : 2013AGPT0030 . pastel-00879090

HAL Id: pastel-00879090

https://pastel.hal.science/pastel-00879090

Submitted on 31 Oct 2013

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

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

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

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

(AgroParisTech)

Spécialité : Hydrologie

présentée et soutenue publiquement par

Laurent CORON

le 26 mars 2013

Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution ?

Diagnostic sur un échantillon de bassins versants français et australiens

Jury

M. François ANCTIL *Université Laval (Québec, Canada)*	Rapporteur
M. Thierry LEVIANDIER *ENGEES (Strasbourg, France)*	Rapporteur
Mme Anne-Catherine FAVRE *LTHE (Grenoble, France)*	Examinateur
Mme Christel PRUDHOMME *CEH (Wallingford, Royaume-Uni)*	Examinateur
M. Rémy GARÇON *EDF-DTG (Grenoble, France)*	Examinateur
M. Hubert SAVENIJE *TU-Delft (Delft, Pays-Bas)*	Examinateur
Mme Marie BOURQUI *EDF-R&D (Chatou, France)*	Encadrant
M. Vazken ANDRÉASSIAN *Irstea (Antony, France)*	Directeur de thèse

AgroParisTech
EDF R&D, Laboratoire National d'Hydraulique et d'Environnement
Istea, unité de recherche Hydrosystèmes et Bioprocédés

Thèse de doctorat

Université *AgroParisTech*
École doctorale *Géosciences et Ressources Naturelles*
Encadrement *EDF R&D LNHE, Irstea Antony HBAN*

Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution ?

Diagnostic sur un échantillon de bassins versants français et australiens

Laurent CORON

Thèse soutenue le 26 mars 2013 devant le jury composé de :

M. François ANCTIL	*Université Laval (Québec, Canada)*	Rapporteur
M. Thierry LEVIANDIER	*ENGEES (Strasbourg, France)*	Rapporteur
Mme Anne-Catherine FAVRE	*LTHE (Grenoble, France)*	Examinateur
Mme Christel PRUDHOMME	*CEH (Wallingford, Royaume-Uni)*	Examinateur
M. Rémy GARÇON	*EDF-DTG (Grenoble, France)*	Examinateur
M. Hubert SAVENIJE	*TU-Delft (Delft, Pays-Bas)*	Examinateur
Mme Marie BOURQUI	*EDF-R&D (Chatou, France)*	Encadrant
M. Vazken ANDRÉASSIAN	*Irstea (Antony, France)*	Directeur de thèse

AgroParisTech
EDF R&D Laboratoire National d'Hydraulique et d'Environnement
Irstea, unité de recherche Hydrosystèmes et Bioprocédés

Remerciements

Cette thèse n'existerait pas sans le soutien personnel et scientifique de nombreuses personnes.

Je voudrais en premier lieu remercier Vazken Andréassian, mon directeur de thèse et Marie Bourqui, mon encadrante à EDF, qui m'ont permis de mener ces travaux. Je remercie Vazken, qui m'a recruté, formé et accompagné au fil des années et dont l'enthousiasme scientifique est une réelle source d'inspiration. Je remercie Marie pour son soutien quotidien et l'équilibre qu'avec Vazken elle a su instaurer entre cadrage technique et liberté d'initiative pour me permettre de mener ce projet de thèse à son terme dans d'excellentes conditions.

Je tiens à exprimer ici ma reconnaissance envers les membres de mon jury qui m'ont fait l'honneur d'évaluer mes travaux : François Anctil et Thierry Leviandier, pour avoir accepté le rôle de rapporteurs ainsi qu'Anne-Catherine Favre, Christel Prudhomme, Rémy Garçon et Hubert Savenije pour leurs rôles d'examinateurs. Je les remercie chaleureusement de leur contribution et des discussions très intéressantes que nous avons pu avoir lors de ma soutenance.

J'adresse ensuite ma gratitude aux personnes qui ont permis la réalisation de cette thèse. Je pense à mes supérieurs hiérarchiques et aux personnels administratifs d'EDF ainsi qu'à ces nombreux collègues qui m'ont permis d'utiliser leurs données hydro-météorologiques (EDF R&D, EDF DTG, Irstea, CSIRO, LTHE, Banque Hydro, Météo-France).

Je remercie également tous mes collègues avec qui j'ai pu échanger au cours de ces trois années, sur mes travaux mais aussi sur bien d'autres thématiques. Je leur dois beaucoup et cette thèse ne serait pas ce qu'elle est sans ces discussions dont j'ai beaucoup appris. Je pense particulièrement à mes collègues d'EDF R&D et d'Irstea Antony qui m'ont accompagné au quotidien et plus généralement à tous ceux avec qui j'ai eu le privilège de partager un match de frisbee, une descente en canoë, une expédition spéléologique, ou plus simplement un thé, une pinte, un couscous ou un bon jeu de mots (je laisse la liberté à chacun retrouver le(s) terme(s) lui correspondant). Je remercie tout particulièrement Pierre et Germain, qui ont partagé mon bureau, ainsi que Frédéric Hendrickx et Charles Perrin, qui n'apparaissent pas en tant que tels dans la liste de mes encadrants mais m'ont beaucoup soutenu et conseillé dans la réalisation et la valorisation des travaux restitués ici. Je les remercie notamment pour la qualité des suggestions qu'ils m'ont adressées lors de la rédaction de ce manuscrit. Je remercie aussi Julien Lerat qui a largement contribué au succès de la collaboration avec le CSIRO et à la richesse de mon séjour de six mois à Canberra.

Je remercie enfin mes amis et ma famille pour leur soutien, mes parents qui m'ont transmis cette curiosité des choses et des gens, et surtout je remercie Laura pour tout ce qu'elle m'apporte et que j'ai la chance d'avoir comme compagnon de route.

Résumé

Les hydrologues sont questionnés au sujet de l'évolution à moyen et long terme de la ressource en eau. Pour répondre à ces questions, ils ont recours à des modèles hydrologiques. Une quantification des incertitudes associées à leurs projections leur est souvent demandée et pose la question de la robustesse de ces modèles conceptuels, notamment dans le contexte d'un climat en évolution. En effet, l'utilisation d'un modèle dans des conditions différentes de celles où il a été calé repose sur une hypothèse de transférabilité des paramètres, c'est-à-dire la possibilité d'utiliser les paramètres d'un modèle dans des conditions différentes de celles de sa construction. Nous traitons cette problématique en tentant d'apporter des éléments de réponse aux questions suivantes :

- Quel est le niveau de robustesse des modèles hydrologiques conceptuels face à des conditions climatiques variables ?
- Quelles sont les causes des situations de manque de robustesse et y a-t-il des façons de s'en prémunir ?

Nous y répondons à travers l'étude des performances de modèles conceptuels lors de nombreux tests de transfert temporel de leurs paramètres. Les résultats montrent des corrélations entre les problèmes de robustesse et les écarts de conditions climatiques entre les périodes de calage et de validation des modèles. Des situations de biais systématiques corrélés à des changements de température de l'air ressortent en particulier de ces analyses. Une hétérogénéité existe cependant sur notre échantillon, les variables climatiques ou le type d'erreurs associées aux problèmes détectés pouvant varier d'un bassin à un autre.

L'étude des biais de simulation sur les bassins où les modèles ne sont pas robustes montre des alternances de phases de sous-estimation et surestimation des débits, le biais sur le module décennal pouvant atteindre 20%. Nos travaux révèlent que des comportements très similaires sont obtenus pour diverses périodes et méthodes de calage. Pour les modèles conceptuels utilisés dans ces travaux, les problèmes de robustesse détectés ne résultent pas simplement de quelques "mauvais" calages conduisant à la sélection de paramètres inadaptés pour reproduire le comportement du bassin considéré. Ils semblent plutôt découler de difficultés générales de ces outils de modélisation à reproduire convenablement les bilans moyens de façon simultanée sur différentes périodes.

Ces travaux ouvrent des pistes de réflexion sur la capacité insuffisante de certains modèles hydrologiques conceptuels à reproduire des dynamiques basse fréquence et sur le rôle des erreurs d'estimation des forçages, notamment les variations temporelles de l'évapotranspiration, dans les situations de modélisation incorrecte.

Abstract

Hydrologists are asked to estimate the medium- and long-term evolutions of water resources. To answer these questions, they commonly use conceptual models. In addition, they are often required to provide an estimate of the uncertainties associated with model projections. This raises the question of the robustness of conceptual models, especially in the context of climate evolution. Indeed, using a model in conditions different from those of calibration is based on the hypothesis of parameter transferability, *i.e.* the possibility to use model parameters in conditions different from those used for the model set-up. We focus on this issue with the aim of answering the following questions:

- What is the robustness level of conceptual hydrological models in the context of changing climatic conditions?
- What are the causes for the lack of robustness, and are there ways to prevent it?

We answer these questions by studying the performance of conceptual models through multiple tests of temporal transfer of their parameters. Results show the existence of correlations between the robustness problems and the difference in climate conditions between model calibration and validation periods. The analysis especially points out the situations of systematic bias correlated to differences in air temperature. However, results are heterogeneous in our catchment set, and climate variables or error type associated with the identified problems vary between catchments.

The analysis of simulation biases on catchments where the models are not robust shows alternating phases of flow under- or overestimation, with a possible bias in the mean flow up to 20% over a ten-year period.

Our work reveals that very similar results can be obtained for various periods or calibration methods. The robustness issues faced by conceptual models used in this study do not solely stem from inadequate calibrations leading to the selection of parameters unable reproduce the catchment behavior. They seem to be the consequence of overall difficulties for models to satisfactorily simulate water balances simultaneously on various periods.

This work opens reflections on the limited capacity of some hydrological models to reproduce low-frequency dynamics and raises questions on the role of inputs estimates errors in model failures, especially the temporal variations of evapotranspiration.

Table des matières

Résumé **iii**

Abstract **v**

Table des matières **vii**

Introduction générale **1**

Contexte et problématique scientifique 2
Organisation du mémoire 3

Partie I. Problématique et méthodes d'examination **5**

1. Contexte et complexité de la modélisation hydrologique **7**

1.1. Modélisation de la relation pluie-débit 8
1.1.1. Objet d'étude 8
1.1.2. Philosophies de modélisation 8
1.1.3. Paramétrisation des modèles : nécessité et contraintes 9
1.1.4. Transférabilité des paramètres d'un modèle conceptuel 11
1.2. Modélisation des bassins de montagne 15
1.2.1. Influence des reliefs sur les régimes hydrologiques 15
1.2.2. Difficultés d'estimation des forçages en montagne 17
1.2.3. La neige dans les modèles hydrologiques 18
1.3. Évolutions du climat et incertitudes de projections 20
1.3.1. Changement climatique et études associées 20
1.3.2. Sources d'incertitudes dans les études d'impact 22
1.4. Synthèse du contexte et objectifs de cette thèse 23

2. Comment évaluer la robustesse climatique des modèles pluie-débit ? **25**

2.1. Méthodologie de tests et d'analyse 26
2.1.1. Les outils de mesure de la robustesse des modèles 26
2.1.2. Proposition d'une procédure de test 28
2.1.3. Méthodes d'analyse des résultats 31
2.2. Construction d'une base de données 38
2.2.1. Critères de sélection de bassins versants 38

2.2.2. Données utilisées . . . 39
2.2.3. Caractéristiques des bassins retenus . . . 44
2.3. Modèles hydrologiques considérés . . . 48
2.3.1. Modèle GR4J . . . 48
2.3.2. Modèles Mordor et Mordor6 . . . 49
2.3.3. Modèle SimHyd . . . 51
2.3.4. Module CemaNeige . . . 52
2.3.5. Modèle Cequeau . . . 53
2.4. Mise en place de la procédure . . . 55
2.4.1. Groupes de bassins utilisés pour tester chaque modèle . . . 55
2.4.2. Conditions d'utilisation du GSST . . . 55
2.4.3. Critère de calage et d'évaluation des modèles hydrologiques . . . 57
2.4.4. Variables utilisées pour caractériser les périodes dans la recherche des relations △erreurs=f(△caract.) . . . 58

Partie II. Première phase du diagnostic Étude des symptômes **61**

3. Quantifier la robustesse des modèles hydrologiques face au climat **63**
3.1. Introduction . . . 64
3.1.1. Résumé des choix méthodologiques . . . 64
3.1.2. Niveaux de performances des modèles . . . 64
3.1.3. Précisions quant au calage du module CemaNeige . . . 66
3.2. Conditions de calage, conditions de validation et erreurs des modèles . . . 68
3.2.1. Sélection des tendances significatives . . . 68
3.2.2. Tableau récapitulatif des corrélations obtenues avec GR4J . . . 69
3.3. Zoom sur les cas de problèmes de robustesse en lien avec △température . . . 75
3.3.1. Tableau de corrélation . . . 75
3.3.2. Variations des erreurs sur un sous-groupe choisi . . . 76
3.3.3. Illustration sur deux bassins . . . 79
3.4. Zoom sur les cas de problèmes de robustesse en lien avec △précipitations . . . 84
3.4.1. Tableau de corrélation . . . 84
3.4.2. Variations des erreurs sur un sous-groupe choisi . . . 85
3.4.3. Illustration sur deux bassins . . . 89
3.5. Zoom sur la modélisation des bassins nivaux . . . 93
3.5.1. Introduction . . . 93
3.5.2. Variations des erreurs sur un sous-groupe choisi . . . 93
3.5.3. Illustration sur deux bassins . . . 97
3.6. Synthèse . . . 101

4. Sensibilité des projections à long terme aux choix de modélisation 103
4.1. Introduction . . . 104
4.1.1. Problématique et objectifs . . . 104
4.1.2. Parenthèse terminologique . . . 105
4.1.3. Une valorisation seulement partielle des résultats du GSST . . . 105
4.2. Méthodologie . . . 107
4.2.1. Contexte de l'exercice et origine des données . . . 107
4.2.2. Chaîne de modélisation considérée . . . 108
4.2.3. Détail des sensibilités étudiées . . . 109
4.2.4. Situation de référence . . . 110
4.3. Étude de sensibilité des projections absolues . . . 112
4.3.1. Sensibilité des projections à la modélisation du climat global . . . 112
4.3.2. Sensibilité des projections à la descente d'échelle . . . 112
4.3.3. Sensibilité des projections à la modélisation hydrologique . . . 114
4.3.4. Synthèse des sensibilités constatées sur les projections absolues . . . 116
4.4. Étude de sensibilité des projections relatives . . . 118
4.4.1. L'expression de projections relatives : une pratique courante . . . 118
4.4.2. Méthodologie suivie . . . 119
4.4.3. Niveaux de sensibilité obtenus sur les projections relatives . . . 119
4.4.4. Résultats complémentaires pour d'autres modèles . . . 122
4.4.5. Rappels sur la signification des faisceaux de projections illustrés . . . 123
4.5. Apports de ces études de sensibilité pour nos travaux en temps présent . . . 124
4.5.1. Questions soulevées par nos études de sensibilité des projections . . . 124
4.5.2. Un nouveau regard sur nos tests en temps présent . . . 125

***Partie III. Seconde phase du diagnostic Compréhension et tentatives de remédiation* 131**

5. Recherche des causes du manque de robustesse des modèles 133
5.1. Introduction . . . 134
5.1.1. Quelques rappels des travaux menés jusqu'ici . . . 134
5.1.2. L'éternel dilemme entre modèle et données . . . 134
5.2. Recherche d'explications du côté de la modélisation . . . 136
5.2.1. Sensibilité des résultats au choix du modèle hydrologique . . . 136
5.2.2. Impact de la longueur de la période de calage du modèle . . . 137
5.3. Recherche d'explications du côté de la qualité des données . . . 140
5.3.1. Sensibilité des résultats à la qualité des chroniques de débits . . . 140
5.3.2. Sensibilité des résultats au choix des données climatiques . . . 141

5.4. Analyse des dérives temporelles des erreurs de modélisation 148
5.4.1. Pertinence de cette piste explicative des problèmes de robustesse 148
5.4.2. Analyses des liens entre △temporel et erreurs de modélisation 152
5.4.3. Conclusions 156
5.5. Analyse des parallèles entre biais=f(△caract.) et △param.=f(△caract.) 157
5.5.1. Introduction et mode d'analyse 157
5.5.2. Résultats 158
5.5.3. Fixer les paramètres les plus sensibles est-il une solution ? 167
5.6. Analyse des biais glissants de modélisation 170
5.6.1. Mode de représentation 170
5.6.2. Quelques comparaisons multi-bassins et multi-modèles 172
5.6.3. Interprétations des similarités et liens avec les constatations précédentes 176
5.6.4. Quelle généralisation de ces résultats ? 177
5.7. Conclusions 181

6. Quelles pistes d'amélioration de la robustesse des modèles ? 185
6.1. Introduction 186
6.2. Remplacement du KGE par d'autres critères de calage classiques 187
6.2.1. Introduction 187
6.2.2. Critères de calage considérés 187
6.2.3. Résultats 189
6.2.4. Suffirait-il de retirer la contrainte (trop forte) sur le biais ? 191
6.2.5. Discussion 195
6.3. Méthodes de sélection des paramètres ciblant la robustesse 195
6.3.1. Introduction 195
6.3.2. Concept d'un calage robuste en termes d'extrapolation climatique . . . 196
6.3.3. Une procédure ciblant la réduction des biais moyen terme 198
6.3.4. Conclusions 202
6.4. Quelles pistes du côté de l'évapotranspiration ? 203
6.4.1. Introduction 203
6.4.2. Différentes formulations pour calculer l'ETP 205
6.4.3. Modification de la relation liant ETR et ETP dans GR4J 207
6.4.4. Comparaison des résultats 208
6.4.5. Discussion 212
6.5. Un test de correction des paramètres selon les conditions de transfert 213
6.5.1. Introduction 213
6.5.2. Procédure de test 214
6.5.3. Résultats 216
6.5.4. Discussion 218
6.5.5. Perspectives 220

6.6. L'espace adimensionnel Q/P=f(P/ETP), un outil d'analyse 222
6.6.1. Présentation du mode de représentation 222
6.6.2. Un outil d'analyse pour critiquer les données 223
6.6.3. Un outil d'analyse pour étudier les variations temporelles de bilan . . . 226
6.7. État des lieux 230

Synthèse et conclusion générale **231**

Références **237**

Annexes **247**

A. Listes des figures et tableaux **249**
A.1. Liste des figures 250
A.2. Liste des tableaux 257

B. Pathologies of hydrological models used in changing climatic conditions **259**

C. Compléments sur les données et les modèles **267**
C.1. Compléments sur le contrôle de la qualité des débits 268
C.2. Test de la nécessité de caler le module CemaNeige 271

D. Compléments pour le chapitre 3 **275**
D.1. Pertes performances moyennes liées aux transferts 276
D.2. Tableaux de corrélations entre △erreurs et △caract. 277
D.3. Caractéristiques des sous-groupes d'analyse de tendance 280
D.4. Erreurs de volumes constatées lors des tests de transfert 282

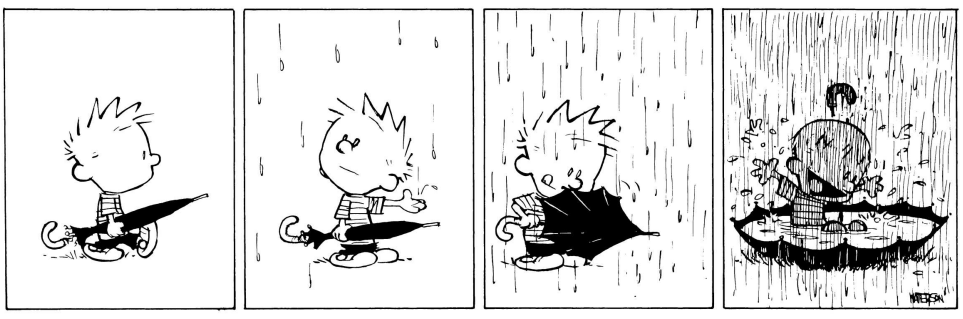
E. Crash testing of hydrological models in contrasted climate conditions... **285**

F. Compléments pour le chapitre 4 **305**

G. Compléments pour le chapitre 5 **311**
G.1. Compléments sur les tests de bootstrap 312
G.2. Compléments sur les analyses des variations des paramètres optimaux 314
G.3. Compléments sur l'exercice de fixation d'un paramètre 324
G.4. Compléments sur les analyses des biais glissants 327

H. Compléments pour le chapitre 6 **331**
H.1. Concept d'un calage robuste en termes d'extrapolation climatique 332
H.2. Compléments sur les tests de correction des paramètres 342

Introduction générale



Calvin and Hobbes, 12/04/1989, Bill Watterson©

Contexte et problématique scientifique

Les hydrologues sont de plus en plus sollicités pour l'estimation des évolutions à moyen et long terme de la ressource en eau face aux changements du climat. Pour répondre à cette problématique, ils alimentent des modèles hydrologiques avec des simulations de climats futurs issues de modèles climatiques. Cependant, cette "projection" des débits futurs n'est pas suffisante et une quantification des incertitudes associées aux projections est souvent indispensable pour en qualifier la vraisemblance.

Dans une étude d'impact du changement climatique conduite il y a près de 20 ans sur 28 bassins versants australiens, *Chiew et al.* (1995) émettaient une nuance terminologique, directement après la présentation de leurs résultats, en ne parlant pas de "prévisions" mais plutôt de "simulations des changements probables tels qu'ils peuvent être estimés actuellement". Cette remarque révélait clairement l'étendue des incertitudes attenantes à ce type d'exercice de projection. Depuis, l'amélioration des capacités de calcul a permis une évolution considérable des modèles de climat qui prennent mieux en compte la diversité des processus atmosphériques. Néanmoins, les inconnues sur les scénarios d'émission de gaz à effet de serre n'ont pas diminué, la reproduction des mécanismes du climat à grande échelle demeure très complexe, et les modèles (climatiques et hydrologiques) restent des modèles et sont donc (largement) imparfaits. Chacune des étapes de la chaîne de modélisation constitue une source possible d'erreurs se propageant jusqu'au résultat final. Comme les autres participants à cette chaîne, l'hydrologue-modélisateur se doit d'estimer la contribution aux erreurs totales induite par "son" maillon, indépendamment de la part relative que celle-ci représente (*Singh et al.*, 2011). Quelle confiance accorderait-il à ses simulations de débits si les conditions climatiques futures étaient parfaitement connues ?

Les modèles hydrologiques sont des outils visant à reproduire la relation entre les forçages climatiques et les écoulements en rivière sur un territoire donné. Dans le cas des modélisations conceptuelles, l'adaptation d'un modèle à un bassin versant se fait via une phase communément appelée "calage", au cours de laquelle les paramètres du modèle sont ajustés pour que les simulations soient au plus proche des phénomènes mesurés (*ex.* dynamique et volumes des écoulements). L'utilisation d'un modèle sur une autre période que celle où il a été calé repose sur une hypothèse de "transférabilité temporelle" des paramètres, c'est-à-dire la possibilité d'utiliser les paramètres d'un modèle dans des conditions différentes de celles de sa construction. La problématique traitée au cours de cette thèse est l'étude de cette transférabilité dans un contexte de climat en perpétuelle évolution. Dans la lignée des recommandations de *Klemeš* (1986), on se posera les questions suivantes : Quels risques prend-on à utiliser un modèle calé dans certaines conditions climatiques pour simuler une période ayant des conditions différentes (*ex.* plus chaude, moins humide) ? Quelles sont les conséquences néfastes sur la qualité des simulations produites et peut-on les limiter ? La figure ci-jointe résume synthétiquement cette problématique.

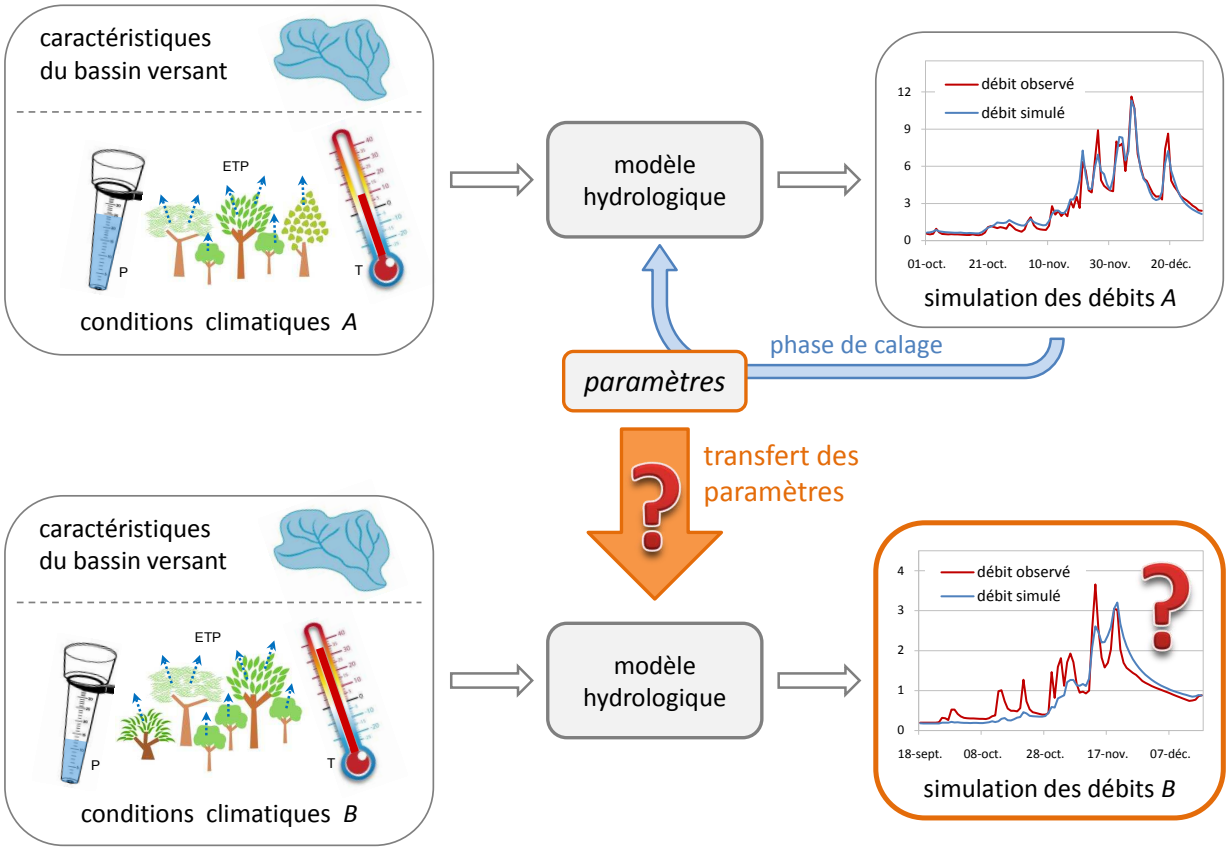


Illustration synthétique de la problématique scientifique

Organisation du mémoire

Ce mémoire s'organise en six chapitres suivis d'un chapitre de conclusion générale.

Le premier chapitre expose le contexte général de nos travaux. Il présente la discipline que constitue la modélisation en hydrologie et discute sa complexité. Il introduit les questionnements à l'origine de cette thèse concernant la robustesse des modèles hydrologiques conceptuels et la nécessité de son évaluation.

Le deuxième chapitre expose la méthodologie suivie. Il rentre dans le vif du sujet en présentant les techniques d'évaluation de la robustesse des modèles hydrologiques conceptuels. Il présente les données et les modèles utilisés pour conduire cette évaluation et détaille la façon dont celle-ci est menée.

Le troisième chapitre marque le début du diagnostic et la transition entre la première partie qui introduit contexte et méthodologie et la deuxième partie où sont initiées les évaluations de la robustesse des modèles étudiés. Il présente les résultats de nombreux tests de transfert temporel

des paramètres de plusieurs modèles hydrologiques sur un large échantillon de bassins. Des relations y sont recherchées entre les erreurs des modèles (en validation) et les changements de conditions climatiques subis lors des transferts de paramètres. Ce troisième chapitre présente notamment des analyses spécifiques sur trois groupes de bassins où des situations différemment problématiques sont constatées.

Le quatrième chapitre poursuit les évaluations des modèles, mais dans le cadre d'une étude d'impact du changement climatique. Il s'intéresse en particulier à la sensibilité des projections émises aux choix de modélisation et notamment au choix de la période de calage du modèle hydrologique. Cette sensibilité est évaluée sur un bassin versant et comparée aux autres éléments de la chaîne de modélisation tels que le choix du modèle de climat global ou de la procédure de descente d'échelle.

Le cinquième chapitre présente un approfondissement des analyses dans le but de comprendre les causes des problèmes de robustesse identifiés dans le troisième chapitre. Il marque la transition entre la deuxième partie portée sur l'évaluation de la robustesse des modèles et la troisième partie où les travaux s'orientent sur la compréhension des situations problématiques et la recherche de solutions. Ce cinquième chapitre présente notamment les analyses des variations temporelles des paramètres optimaux et du lien entre ces variations et les erreurs de simulation des bilans hydrologiques constatées lors des tests de transfert entre des périodes climatiquement contrastées.

Le sixième chapitre expose les explorations de pistes visant à réduire des problèmes de robustesse des modèles conceptuels mis en avant au fil des chapitres précédents. Il présente notamment le test de différents critères de sélection des paramètres des modèles. Une procédure de correction des paramètres est également proposée et évaluée. Ce chapitre discute enfin de la piste des erreurs d'estimation des données comme cause des problèmes constatés sur les modèles et détaille quelques tests sur ce sujet concernant le calcul de l'évapotranspiration.

Ce manuscrit s'achève par un septième chapitre qui résume la problématique et les principales contributions de cette thèse, puis pointe quelques directions pour le prolongement des analyses sur cette thématique de recherche.

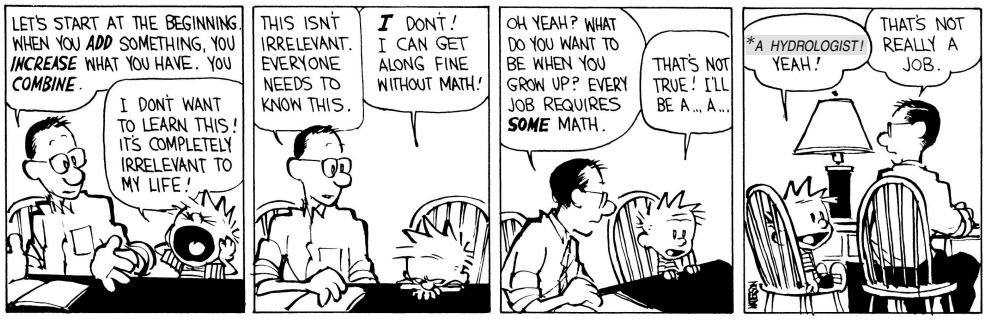
Partie I

Problématique et méthodes d'examination

1

Contexte et complexité de la modélisation hydrologique

L'hydrologie, une discipline à part, entre météorologie et hydraulique, dont la complexité n'est pas toujours reconnue à sa juste valeur...



Calvin and Hobbes, 17/09/1990, Bill Watterson©

** modification qui peut être aperçue sur la porte d'un bureau de doctorants à Irstea-Antony (initialement "a caveman")*

1.1. Modélisation de la relation pluie-débit

1.1.1. Objet d'étude

La modélisation hydrologique vise à représenter les échanges d'eau d'un système complexe : le bassin versant. Cette zone est généralement définie comme le territoire sur lequel toutes les eaux tombées alimentent un même exutoire (cours d'eau, lac, mer, océan). Un bassin versant est délimité par des lignes de partage des eaux que l'on pourra souvent approximer par les lignes de crêtes topographiques. En pratique, l'exercice de modélisation consiste généralement à simuler des débits en un point spécifique d'un cours d'eau. C'est alors la localisation de ce point qui déterminera le bassin versant étudié : le territoire dont ce point constitue l'exutoire. Une illustration de bassin est fournie à la figure 1.1. Il s'agit du bassin de *la Durance à Serre-Ponçon* qui couvre une superficie d'environ 3600 km^2 et où les écoulements ont lieu des sommets au nord et à l'est du bassin vers l'exutoire au sud-ouest.

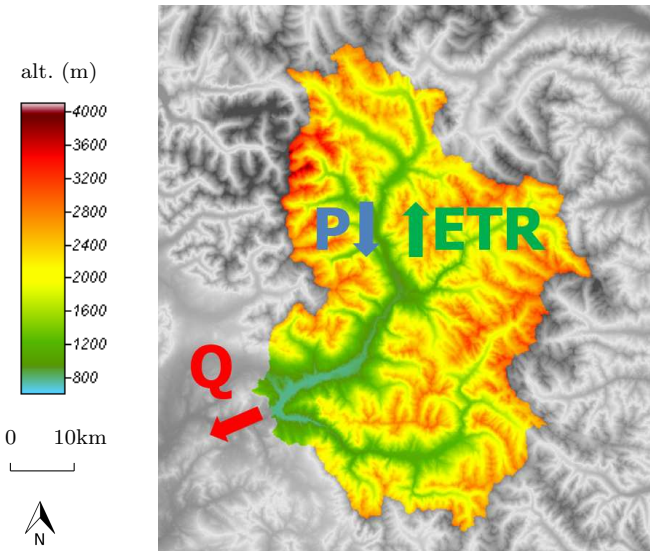


Figure 1.1.: Illustration de l'objet d'étude : le bassin versant

Les flux d'eau symbolisés par les lettres P, ETR et Q correspondent aux précipitations, à l'évapotranspiration et aux écoulements surfaciques (les écoulements souterrains ne sont pas représentés ici mais peuvant être significatifs sur certains bassins).

1.1.2. Philosophies de modélisation

Deux approches de modélisation s'opposent quant au chemin à suivre pour avancer vers une représentation fidèle du comportement d'un système (en hydrologie : le bassin versant). Pour faciliter la description de ces deux visions, reprenons l'analogie utilisée par *Le Moine* (2008) entre l'exercice de modélisation et la résolution d'un puzzle. Une première approche consiste à se concentrer sur chaque pièce du puzzle, une à une, en veillant aux emboîtements de celles-ci selon un schéma qui assure la continuité d'un motif général construit au fur et à mesure. Cette approche porte les dénominations suivantes : *mécaniste*, *réductionniste*, *ascendante* (*upward* ou

bottom-up en anglais). Elle est largement utilisée en science, où les progrès de compréhension locale ont lieu au fil des études et expériences conduites, pour aboutir un jour à une compréhension complète et détaillée du système étudié. La seconde approche, se concentre elle sur la représentation directe du motif final (en hydrologie : la simulation du débit). L'emboîtement précis des pièces du puzzle n'y est plus nécessairement important et certaines pièces peuvent être largement approximées (voire écartées du puzzle) si les efforts associés à leur obtention ou compréhension ne sont pas suffisamment récompensés dans la représentation du motif général. Cette approche porte les dénominations suivantes : *empirique*, *non-réductionniste*, *systémique*, *descendante* (*downward* ou *top-down* en anglais).

Les différents processus du cycle de l'eau (précipitation, solidification, fusion, évaporation, transpiration, infiltration, écoulements surfaciques et souterrains...) sont désormais relativement bien connus et peuvent être modélisés à l'échelle locale. L'approche *ascendante* est donc envisageable en hydrologie et les modèles ainsi construits sont désignés sous le terme *modèles physiques*. Néanmoins, le passage à l'échelle du bassin versant reste extrêmement difficile du fait de la grande hétérogénéité de l'objet d'étude (*ex.* topographie, couverture végétale, composition du sol et du sous-sol...), des difficultés d'estimation des échanges locaux avec l'atmosphère (notamment la forte variabilité spatiale des précipitations) et des difficultés d'adaptation aux possibles évolutions temporelles du système (*c.-à-d.* conditions du bassin non nécessairement stationnaires, en lien par exemple avec une déforestation ou une forte urbanisation). Si tous les verrous entravant cette approche ascendante étaient soulevés, nous disposerions d'un modèle représentatif de tous les processus mécaniques, thermodynamiques et chimiques en action au sein du système bassin versant, à toutes les échelles de temps et d'espace, sans qu'il faille avoir recours à des paramètres nécessitant un calage. Cependant, les verrous sont encore nombreux, en lien notamment avec notre méconnaissance du bassin versant modélisé, et maintiennent la présence de simplifications dans les modèles hydrologiques. Dans une très grande majorité de cas pratiques, l'obtention des mesures physiques permettant l'élaboration *ascendante* de modèles n'est pas possible et contraint les modélisateurs à faire de nombreuses simplifications (*Murphy et al.*, 2006; *Bárdossy*, 2007). Cela conduit à l'émergence de modèles appelés *modèles à base physique* qui sont construits selon les principes de l'approche *ascendante* mais qui comportent certaines zones de conceptualisation (nécessitant paramétrage), mais aussi à l'émergence de modèles purement *conceptuels* issus d'approches *descendantes*.

1.1.3. Paramétrisation des modèles : nécessité et contraintes

La difficulté de mesurer et décrire tous les processus à l'échelle du bassin a favorisé les approches *empiriques* (ou *descendantes*) en hydrologie qui sont aujourd'hui très courantes. Les modèles issus de ce type d'approche sont très nombreux et regroupés sous le terme de *modèles conceptuels*. Deux modèles conceptuels peuvent présenter des structures très différentes (nombre de réservoirs, formulations utilisées, représentation spatiale des processus, etc.), leurs

concepteurs étant libres d'imaginer la *bonne façon* de représenter ou simplifier les processus hydrologiques. Le degré de conceptualisation retenu dépend généralement des objectifs de modélisation, mais aussi de la disponibilité des données nécessaires à la construction du modèle.

Un modèle conceptuel intègre des paramètres permettant d'adapter celui-ci aux caractéristiques du cas d'étude et aux objectifs d'utilisation. Comme évoqué précédemment, l'obtention de mesures pertinentes pour représenter le fonctionnement exhaustif des systèmes hydrologiques reste généralement impossible. Combinée aux simplifications spatiales et temporelles des processus décrits par le modèle, cette situation conduit à des valeurs de paramètres qui ne peuvent pas être directement reliées à des observations. Une phase de calage (manuelle ou automatique) est souvent inévitable pour réduire les biais de simulation. De plus, la conceptualisation peut engendrer une différence entre les valeurs *optimales* des paramètres pour un processus et les valeurs *réelles* représentatives de celui-ci, indépendamment de l'approche utilisée pour les déterminer (*Abebe et al.*, 2010). En plus du *rôle* qui leur est affecté (nous pourrions aussi dire *imaginé*), les paramètres peuvent en effet être partiellement *utilisés* par le modèle pour remplir d'autres fonctions (*ex.* participer à la simulation d'autres processus ou compenser certaines limites de mesure et/ou de conceptualisation). L'utilisation de valeurs physiques pour des paramètres conceptuels et inversement l'interprétation physique de valeurs de paramètres optimisés nécessitent donc une grande précaution. Nous noterons au passage que les modèles *à base physique*, évoqués plus haut, ne diffèrent réellement des modèles conceptuels que par la volonté de leurs auteurs, qui cadre leur structure pour qu'elle soit en accord avec les processus réels. Ces modèles restent en réalité soumis aux mêmes problématiques de validité des paramètres que les modèles purement conceptuels issus de l'approche *descendante*. En effet, la sélection des paramètres ajustables a lieu dans des conditions inévitablement spécifiques et peut donc conduire à divers questionnements sur leur unicité, leur représentativité ou leur généralité.

Pour résumer, les contraintes liées à notre méconnaissance des bassins versants nous conduisent souvent à utiliser des modèles *conceptuels*, qui nécessitent par construction une phase de calage de leurs paramètres. Ces modèles ne décrivent que de façon simplifiée le fonctionnement d'un bassin versant mais peuvent se révéler très performants pour de nombreuses utilisations pratiques. L'usage de ces modèles pose néanmoins certaines questions, car il n'est pas garanti qu'un modèle dont les paramètres sont calés dans des conditions particulières (*ex.* une période, un bassin) puisse reproduire les débits sous d'autres conditions (*ex.* une autre période et/ou un autre bassin) tout en conservant le même niveau de qualité des simulations. Étudier cet aspect revient donc à s'intéresser à la *transférabilité* des paramètres des modèles hydrologiques conceptuels. Nous pouvons alors distinguer la transférabilité temporelle (*c.-à-d.* conditions identiques mais période de simulation différente), la transférabilité climatique (*c.-à-d.* évolution des conditions entre les périodes de calage et de simulation), et la transférabilité spatiale (*c.-à-d.* simulation des débits sur un bassin différent de celui utilisé au calage). Les deux premiers types de transférabilité constituent le sujet central de nos travaux (*cf.* illustration

synthétique de la problématique, page 3). L'étude de la transférabilité spatiale des paramètres correspond aux travaux de régionalisation des paramètres visant à permettre la simulation des débits sur des bassins peu ou pas instrumentés. Ce dernier point ne sera pas traité dans cette thèse.

Avant de poursuivre cette introduction sur la modélisation en hydrologie, marquons un temps d'arrêt pour préciser clairement ce que nous entendrons, dans cette thèse, par les termes *"modèle"* et *"robustesse"*.
Le terme *"modèle"* sera utilisé pour désigner un assemblage de relations mathématiques, réalisé dans le but de remplir un objectif de modélisation (*ex.* reproduire les débits journaliers à l'exutoire d'un bassin versant). Ces relations retranscrivent fidèlement des processus physiques et chimiques dans le cas des modèles physiques et découlent d'un exercice de conceptualisation pour les modèles issus des approches empiriques. Dans nos travaux, le mot *modèle* désigne donc une structure, qui ne se rattache pas à un cas d'étude spécifique. En ce sens, le fait de changer le territoire d'étude ou la période et le critère de calage est sans impact sur cette définition, le modèle restant le même dans toutes ces configurations. Les valeurs des paramètres du modèle, qu'elles soient déterminées par une mesure de terrain ou calage automatique, sont ainsi sémantiquement dissociées du terme *modèle* tel qu'il est utilisé dans ces travaux.
Si l'on se réfère à la définition du dictionnaire Larousse, le terme *"robustesse"* inclut les notions de *force* et *solidité*, mais surtout celle d'une capacité de *résistance* face à des causes d'altération, notamment une exposition à des conditions extrêmes par leur durée ou leur intensité. Dans le cadre de cette thèse, nous désignerons sous l'expression *"robustesse d'un modèle"* sa capacité à conserver un niveau d'erreur équivalent à celui obtenu au calage lorsqu'il est utilisé dans des conditions défavorables, c'est-à-dire qui diffèrent de celles de sa construction. Comme évoqué précédemment, nos travaux portent sur l'étude de la robustesse temporelle et climatique des modèles conceptuels. Elle se traduit par la transférabilité des paramètres de ces modèles dans le temps et sur une large gamme de conditions climatiques.

1.1.4. Transférabilité des paramètres d'un modèle conceptuel

De nombreuses raisons peuvent conduire à l'utilisation d'un modèle hydrologique dans des conditions différentes de celles de sa paramétrisation. C'est par exemple le cas lorsqu'un modèle intervient dans l'estimation de crues extrêmes (*Garavaglia*, 2011), qu'il est utilisé pour émettre des projections des ressources sous changement climatique (*Chiew et al.*, 2009; *Ducharne et al.*, 2010) mais aussi plus simplement lorsqu'un modèle calé sur une période historique figée produit des prévisions saisonnières lors d'années climatiquement extrêmes.

L'identification d'un jeu de paramètres jugé optimal durant le calage d'un modèle n'est pourtant pas une garantie qu'il permette une reproduction fidèle des processus hydrologiques dans toutes les situations (*Andréassian et al.*, 2012). Les conditions rencontrées lors du calage des

paramètres du modèle hydrologique peuvent notamment conduire à la sélection de valeurs peu pertinentes pour d'autres périodes. Un panorama des différentes situations où la transférabilité des paramètres peut être questionnée est présenté en annexe B (*Coron et al.*, 2011). Parmi les plus courantes, nous pouvons lister les deux situations suivantes :

– Problèmes de transférabilité liés à l'homogénéité temporelle des forçages du modèle (*c.-à-d.* évolution de la qualité et/ou disponibilité des données) : Cette situation résulte de la capacité des modèles conceptuels à compenser via leurs paramètres certaines erreurs sur les entrées (*Apaydin et al.*, 2006), notamment les biais dans l'estimation des précipitations ou d'évapotranspiration (*Kokkonen et Jakeman*, 2001; *Oudin et al.*, 2006b). Lors de changements brutaux ou progressifs des réseaux de mesure, ces biais sur les entrées (*ex.* sous-estimation des pluies) peuvent varier. Si ces biais ne sont pas détectés et corrigés, les compensations faites par le modèle au calage sur une période donnée peuvent alors être inadaptées sur d'autres périodes antérieures ou postérieures (*McMillan et al.*, 2011).

– Problèmes de transférabilité liés à une faible identifiabilité des valeurs des paramètres : Selon la structure du modèle considéré, il se peut que certains paramètres n'aient pas un rôle unique et bien déterminé dans le schéma de simulation et puissent indifféremment prendre diverses valeurs au calage (*Beven et Freer*, 2001; *Zin*, 2002). Cela se produit notamment lorsque des paramètres servent à décrire un processus non significatif dans les conditions d'étude (*ex.* accumulation/fonte de neige, apports de nappe lors de sécheresses prononcées), ou lorsque la structure permet des effets de compensation entre paramètres (plusieurs couples de valeurs pouvant conduire au même résultat) (*Wagener et al.*, 2003; *Wriedt et Rode*, 2006; *Abebe et al.*, 2010). Lors du transfert vers d'autres conditions d'utilisation, cette identification incorrecte des valeurs comportementales peut engendrer une diminution de la qualité des simulations (*c.-à-d.* une perte de robustesse).

Une troisième situation, se situant au centre de nos travaux, correspond aux problèmes de transférabilité liés à la dépendance des paramètres aux conditions climatiques rencontrées lors du calage. En effet, le jeu de paramètres optimal sur un bassin peut, dans certains cas, évoluer temporellement avec les variations de climat. Diverses références à ces situations peuvent être trouvées dans la littérature. *Wagener et al.* (2003) ont utilisé un modèle global à cinq paramètres sur un bassin versant anglais et montrent que les périodes estivales et celles concentrant les fortes pluies nécessitent des optimaux différents pour les paramètres contrôlant le transfert rapide des écoulements. *Choi et Beven* (2007) ont échantillonné les chroniques d'un bassin sud-coréen selon des critères de similarités hydrologiques. En calant un modèle sur chaque sous-ensemble ainsi constitué, ils observent comment des paramètres optimaux sur certains ensembles ne conviennent pas à d'autres. *Rosero et al.* (2010) trouvent que les paramètres d'un modèle de surface terrestre sont également fortement influencés par les conditions climatiques, tandis qu'ils sont théoriquement contrôlés par les seules caractéristiques physiques du terrain (*ex.* propriétés du sol, type de végétation, etc.).

L'observation de ce type de résultats n'est pas restreinte à un nombre limité de bassins versants particuliers. Les travaux de *Rosero et al.* (2010) portent sur neuf bassins versants. Des travaux analogues comme ceux de *Vaze et al.* (2010) et *Merz et al.* (2011) portent respectivement sur 61 bassins australiens et 273 bassins autrichiens. Ces deux dernières études aboutissent à des constatations similaires de dépendance des paramètres au climat de calage, entrainant des erreurs de simulation plus grandes lors du transfert des paramètres (*cf.* figure 1.2 pour les résultats de *Vaze et al.*, 2010). Dans leurs travaux, *Merz et al.* (2011) ont pu relier les valeurs des paramètres associés à la simulation de la neige et de l'humidité du sol avec des caractéristiques climatiques moyennes telles que la température de l'air ou l'évapotranspiration (*ETP*) (*cf.* figure 1.3). Nous mentionnerons toutefois l'existence de travaux dans la littérature, faisant état de situations d'indépendance apparente des paramètres au climat (*cf.* travaux de *Niel et al.* (2003) sur 17 bassins africains). Néanmoins, il semble que la stabilité temporelle des paramètres constitue une hypothèse relativement forte, dont les limites risquent d'être atteintes dans des conditions de climat changeant. Dans ce cadre, *de Vos et al.* (2010) font une proposition intéressante pour étudier plus en détail les cas de variations des paramètres optimaux avec le climat : ils suggèrent d'échantillonner la période d'observations disponible selon des critères de similarité climatique, puis d'autoriser le choix de jeux de paramètres différents sur ces sous-ensembles lors de la phase d'optimisation. Plutôt qu'une méthode alternative de calage, ces auteurs présentent celle-ci comme un outil pour analyser le comportement du modèle hydrologique face à différentes conditions et identifier ainsi les points qui pourraient nécessiter une amélioration.

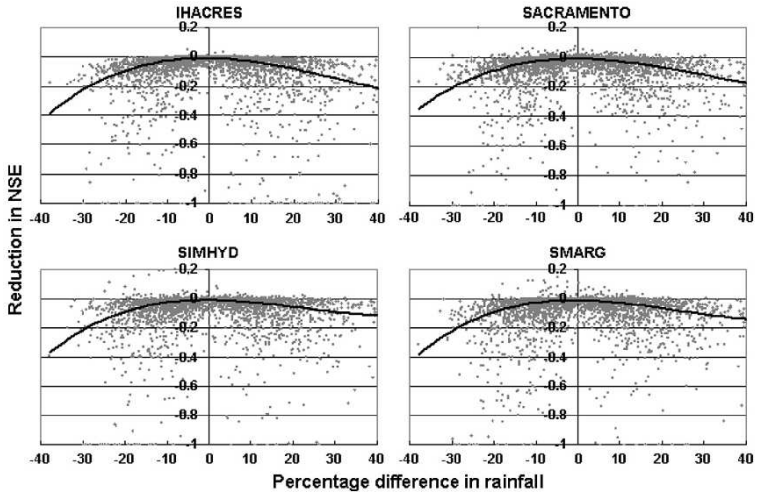


FIGURE 1.2.: Résultats issus de *Vaze et al.* (2010), illustrant le lien entre écarts de précipitations moyennes lors du transfert de paramètres et pertes de performances

Ces diminutions sont étudiées à travers l'étude des différences de critère de Nash et Sutcliffe (1970), pour quatre modèles hydrologiques sur un échantillon de 61 bassins australiens

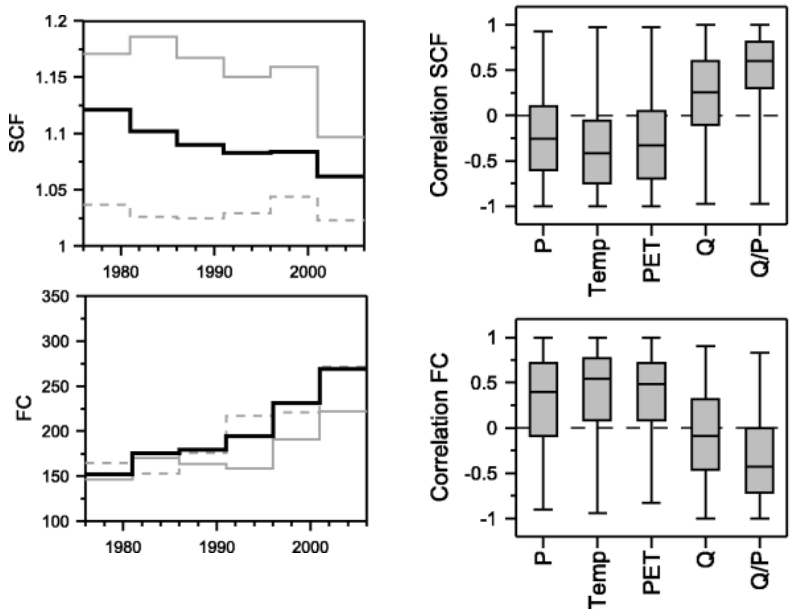


Figure 1.3.: Résultats issus de *Merz et al.* (2011), illustrant la corrélation entre variations climatiques et les valeurs de paramètres d'un modèle hydrologique

Tests menés sur des sous-périodes de 5 ans pour 273 bassins autrichiens. SCF = facteur visant à corriger la sous-captation des précipitations neigeuses, FC = capacité maximale du réservoir de production

Une méthode classique d'évaluation de la dépendance des paramètres d'un modèle au climat et des conséquences y étant associées est le *differential split-sample test* (*DSST*) proposé par *Klemeš* (1986). Il constitue un cas particulier du calage-contrôle simple (*split-sample test*, noté *SST*) dans lequel les périodes de calage et de contrôle sont choisies en fonction de leurs différences climatiques. La dépendance des paramètres au climat de calage est alors étudiée à travers les variations des performances du modèle pour les différentes conditions de test. Une liste non exhaustive d'exemples d'utilisation du *DSST* comprend les travaux suivants : *Refsgaard et Knudsen* (1996); *Donnelly-Makowecki et Moore* (1999); *Xu* (1999); *Seibert* (2003); *Wilby* (2005); *Chiew et al.* (2009); *Vaze et al.* (2010); *Li et al.* (2012); *Seifert et al.* (2012); *Seiller et al.* (2012). La plupart de ces auteurs constatent une diminution des performances des modèles (*c.-à-d.* une augmentation des erreurs sur les débits simulés) lors des transferts de paramètres entre des périodes climatiquement contrastées. Ils concluent que la validité d'un modèle (associé à un jeu optimal de paramètres) pour une utilisation en prévision sous changement climatique dépend du choix plus ou moins judicieux de la période de calage. Ces constatations invitent à une évaluation rigoureuse des incertitudes propres à la structure et la paramétrisation des modèles hydrologiques. Sur cette thématique, nous mentionnerons notamment les travaux de *Butts et al.* (2004); *Reusser et al.* (2008); *Reusser et Zehe* (2011); *Lin et Beck* (2012); *Brigode et al.* (2013). Face à ces problèmes de robustesse, plusieurs équipes travaillèrent sur des méthodes de calage visant à améliorer le degré de transférabilité des paramètres des modèles entres différentes périodes. Nous mentionnerons ainsi les travaux de *Hartmann et Bárdossy* (2005); *Ebtehaj et al.* (2010); *Zhang et al.* (2011); *Gharari et al.* (2013).

Enfin, il est intéressant de noter l'analogie entre le transfert temporel et le transfert spatial des jeux de paramètres d'un modèle hydrologique. Par exemple, *Merz et al.* (2011) notent que les tendances temporelles constatées sur les valeurs des paramètres (du fait d'une évolution graduelle du climat) étaient comparables aux tendances spatiales obtenues entre des régions de climat différent. Un parallèle similaire entre transfert dans le temps et transfert dans l'espace est utilisé par *Singh et al.* (2011). Ils adoptent une approche régionale pour évaluer la *capacité d'extrapolation climatique* de jeux de paramètres via des transferts vers d'autres bassins versants, situés dans une région géographique proche mais plus chaude.

Pour résumer, la question de la robustesse climatique des modèles conceptuels a longtemps été laissée de côté mais les dernières années ont vu un large regain d'attention sur cette thématique. Les études approfondies sur le sujet restent néanmoins peu nombreuses, notamment l'utilisation d'approches systématiques de test mises en place sur un large échantillon de bassins versants.

1.2. Modélisation des bassins de montagne

1.2.1. Influence des reliefs sur les régimes hydrologiques

Les bassins versants situés en zone de montagne présentent des caractéristiques qui leur sont bien spécifiques. Les reliefs forment des barrières au déplacement de masses d'air et forcent celles-ci à s'élever ce qui augmente les précipitations locales. Les bassins montagneux reçoivent ainsi plus d'eau en moyenne que leurs voisins situés en plaine. Leur spécificité principale reste cependant la répartition de leurs écoulements au fil de l'année (on parle aussi de *régime hydrologique*). Les températures relativement basses en altitude augmentent fortement la part des précipitations solides sur ce type de bassins et cette eau reste au sol sous forme de neige sans génération d'écoulement. Elle ne sera libérée que plusieurs jours, semaines ou mois plus tard lorsque les conditions seront réunies pour la fonte du manteau neigeux. Un phénomène de crue de fonte a généralement lieu au printemps (avril-mai-juin) durant lequel les débits augmentent fortement jusqu'à un pic pour redescendre progressivement durant l'été et le début d'automne jusqu'à la reprise des chutes de neige. L'amplitude et le phasage temporel de cette crue de fonte dépendent directement de la quantité de neige reçue pendant l'hiver et du profil de la remontée des températures de l'air au niveau du bassin. Enfin, les étiages sur ce type de bassins ne se produisent pas durant la saison où le rapport précipitations/évapotranspiration est le plus faible mais au cœur de l'hiver lorsque la totalité des précipitations reste stockée sous forme de neige. Ce fonctionnement hydrologique présentant une régularité interannuelle est nommé *régime nival* et illustré par la figure 1.4.

Un bassin de régime nival se différencie nettement d'un bassin de *régime pluvial*, pour lequel la part des précipitations solides reste négligeable et où la répartition des écoulements est plus

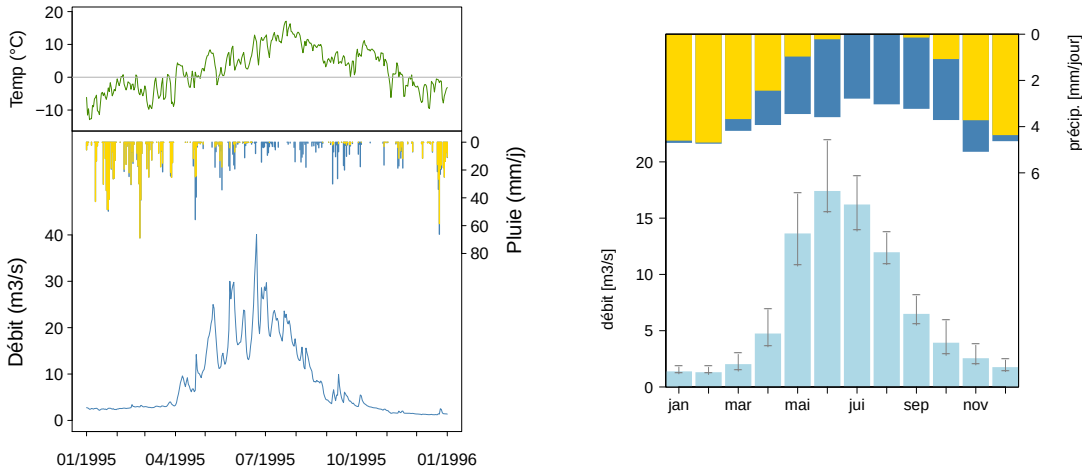


Figure 1.4.: Exemple d'une année hydrologique pour un bassin de régime nival

La Romanche à Mizoën : bassin de 220 km^2 d'altitude médiane 2350 m, où les précipitations solides (orange) réprésentent plus de 50% des apports annuels

corrélée avec celle des précipitations et de l'évapotranspiration potentielle. Si les écoulements sont généralement plus forts en hiver pour les bassins à régime pluvial, ils peuvent présenter de nettes irrégularités interannuelles causées par l'irrégularité des pluies. Deux catégories intitulées *pluvio-nival et nivo-pluvial* regroupent les situations intermédiaires entre régime pluvial et nival. Elles se caractérisent par la présence de deux maximums de débits causés respectivement par la fonte du manteau neigeux (au printemps) et par des précipitations liquides saisonnières plus importantes (généralement en automne-hiver). Pour de tels bassins, deux minimums sont observés : en plein hiver lorsque les précipitations sont majoritairement solides et en plein été lorsque l'évapotranspiration est maximale (parfois combinée à des précipitations plus faibles). Le choix de l'une ou l'autre des dénominations dépend de la composante prépondérante entre pluviale et nivale (*ex.* régime *nivo-pluvial* : crue de fonte majoritaire devant les crues automnales). Enfin, une dernière catégorie dite de *régime nivo-glaciaire* constitue une variante du régime nival dont les caractéristiques sont accentuées par la présence de glaciers : étiages plus prononcés en hiver et retards plus long des écoulements de fonte (maximum des débits en fin d'été). La régularité interannuelle des débits y est également plus marquée que pour les bassins à régime nival. Une illustration de ces différents types de régimes hydrologiques est fournie à la figure 1.5.

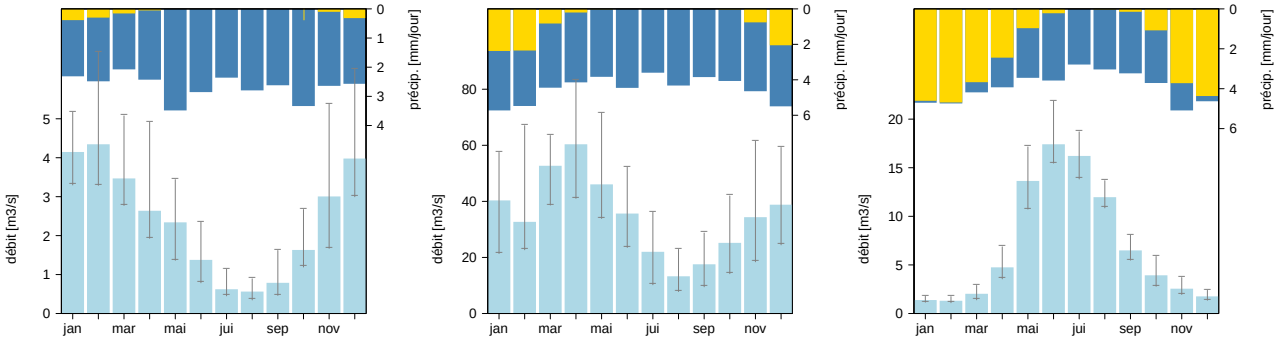


FIGURE 1.5.: Exemples des régimes pluvial, nivo-pluvial et nival

moyennes mensuelles de précipitations liquides (bleu) et solides (orange), moyennes et quantiles 25-75% des débits mensuels (bleu clair) pour l'Aix à Saint-Germain-Laval, le Fier à Vallières, la Romanche à Mizoën

1.2.2. Difficultés d'estimation des forçages en montagne

Le travail de modélisation hydrologique est confronté aux difficultés de mesure des processus déterminants de la relation précipitations - débits, à l'échelle d'un bassin versant. Pour de nombreuses raisons que nous allons évoquer ici, ces difficultés sont amplifiées en zone de montagne, y rendant la modélisation hydrologique d'autant plus complexe.

Une des difficultés majeures de l'hydrologie en zone de montagne réside dans l'estimation des précipitations. Tout d'abord, les précipitations au sein du bassin augmentent en moyenne avec l'altitude tandis que la majorité des appareils de mesure sont situés dans les zones basses (vallées habitées). Cela entraîne une sous-estimation moyenne des volumes d'eau reçus, qui devra être compensée relativement à la topographie du bassin par des gradients correctifs [1]. À cette sous-estimation, liée à l'altitude, s'ajoutent les erreurs de sous-captation principalement causées par les effets aérodynamiques autour de l'appareil de mesure en cas de précipitations solides [2]. Face à ces difficultés, de nombreux efforts sont menés pour estimer au mieux les précipitations solides. C'est le cas au niveau de la mesure : installation de pluviomètres chauffants profilés pour minimiser les effets aérodynamiques, mis en place de systèmes sophistiqués de mesure automatique du manteau neigeux (*ex.* télénivomètres à rayonnement gamma, nivomètres à rayonnement cosmique), campagnes de mesure à la perche des hauteurs de neige en haute montagne (*Gottardi*, 2009). C'est également le cas au niveau de méthodes d'exploitation de ces mesures pour estimer les volumes reçus sur tout un territoire (*Quintana-Seguí et al.*,

1. Les postes pluviométriques et nivométriques situés à haute altitude restent peu nombreux du fait des coûts importants associés à leur installation et maintenance. Les gradients correctifs sont appelés gradients *altitudinaux* et sont appliqués aux données de pluie mais aussi aux données de température de l'air.

2. L'effet aérodynamique est une modification du profil des vents autour de l'appareil de mesure, causée par l'obstacle qu'il représente. D'autres phénomènes tels que l'évaporation des gouttes au contact des parois du pluviomètre et les éclaboussures peuvent aussi engendrer de la sous-captation mais restent minoritaires dans le contexte de la mesure des précipitations en montagne.

2008; *Gottardi et al.*, 2012). Malgré ces efforts, appréhender la répartition spatiale de la neige au sol reste très délicat du fait de sa forte hétérogénéité. Tout d'abord, les effets combinés du vent et du relief engendrent une répartition inégale de la neige lors de sa chute. Une importante redistribution de la neige au sol peut ensuite avoir lieu par l'action du vent et de façon plus réduite par les avalanches. Enfin, une hétérogénéité spatiale est induite par la fonte, certains versants voyant leur couverture neigeuse disparaître bien plus tôt que d'autres du fait de leur plus forte exposition au soleil.

Bien que plus rarement abordée, l'estimation de l'évapotranspiration en montagne reste également une difficulté importante pour la modélisation hydrologique (*Herrnegger et al.*, 2012). Pour réduire, les incertitudes liées au calcul de l'*ETP*, une formulation du type *Penman-Monteith* (*Monteith*, 1965) est généralement recommandée car jugée supérieure aux formulations plus simples (*Allen et al.*, 1998). Néanmoins, certaines composantes de la formule de *Penman-Monteith* telles que le vent ne sont estimées que très approximativement compte tenu du relief. Par défaut, des formulations uniquement dépendantes de la température et de la radiation extra-terrestre restent donc souvent utilisées, par exemple celles de *Thornthwaite* (1948) ou d'*Oudin et al.* (2005).

1.2.3. La neige dans les modèles hydrologiques

Les problèmes d'estimation des forçages sur les bassins versants de montagne viennent renforcer les difficultés inhérentes à la modélisation hydrologique (*cf.* section 1.1.3). Elles conduisent à l'utilisation d'approches conceptuelles (souvent simples) lorsque l'objectif reste la simulation des débits à l'exutoire d'un territoire donné.

La manière la plus simple de modéliser la neige pour l'hydrologie est le modèle conceptuel degré-jour, dont la généralisation à l'échelle d'un bassin versant reste relativement facile. Ce modèle repose sur l'hypothèse que la fonte de la neige peut être considérée proportionnelle à l'écart entre la température de l'air et un seuil de température de fonte. Ce seuil, qui peut être soumis à calage, est généralement situé autour de 0°C lorsque les biais sur l'estimation des champs de températures sont maitrisés. Le calcul de l'ablation du manteau fait ensuite intervenir un paramètre central : le facteur de fonte. Selon l'approche utilisée, celui-ci peut être maintenu constant, quelles que soient les conditions, ou varier temporellement, en fonction de la saison par exemple (*cf.* tests de *Valéry*, 2010). D'autres paramètres peuvent venir compléter cette conceptualisation en ajustant l'inertie de fonte du manteau neigeux. Cette fonte peut ainsi être accélérée selon l'état de murissement de la neige ou ralentie par la modélisation de l'état thermique du manteau limitant l'ablation lors de dépassements courts du seuil de température durant des périodes restant froides en moyenne.

Ce calcul simplifié des processus d'accumulation et d'ablation de neige sur un bassin est parfois mené selon un découpage du territoire en différentes bandes d'altitudes. Cette discrétisation

permet une meilleure représentation de l'accumulation plus importante de neige sur les sommets, puis d'une fonte progressive au printemps et en été des basses altitudes vers les plus hautes en lien avec la remontée progressive de l'isotherme. Nous noterons cependant qu'une forte discrétisation n'entraîne pas nécessairement une amélioration du modèle et pourra au contraire créer des situations non souhaitables (*Valéry*, 2010). C'est le cas lorsqu'un module degré-jour est utilisé sur une zone de trop forte altitude où la température ne dépasse pas suffisamment le seuil de fonte paramétré. En effet, une accumulation continuelle et infinie de neige pourra résulter de cette simplicité des formulations, notamment l'absence d'un module glacière d'écoulement à haute altitude, la non-prise en compte des avalanches, de l'érosion par le vent ou du rôle respectif sur la fonte des différentes composantes du bilan d'énergie. Le niveau adéquat de discrétisation altitudinale d'un module neige de type degré-jour dépend bien entendu de l'amplitude des altitudes présentes sur le bassin étudié. *Valéry* (2010) montra qu'un découpage comprenant entre cinq et dix bandes d'altitudes de surfaces égales semblait, pour ce type de module neige, un bon compromis pour les nombreuses situations testées [3].

Malgré les points de raffinement évoqués ici, il est bon de rappeler que les modules neige conceptuels laissent généralement de nombreux mécanismes de côté : représentation implicite (*c.-à-d.* non explicite) du bilan d'énergie ; pas ou peu de prise en compte réelle des phénomènes de gel, d'écoulement interne d'eau dans le manteau et regel éventuel ; prise en compte non systématique du vieillissement de la neige (modification de sa structure, variation de sa densité, de son contenu en eau et de son albédo), etc. Ces méthodes ont cependant le grand avantage de ne nécessiter que la température de l'air comme donnée d'entrée. De plus, en dépit de leur extrême simplicité, elles restent relativement adéquates lorsque la fonte est dominée par l'apport d'énergie radiative du fait du rayonnement solaire, plutôt que par l'apport de chaleur par advection du fait des mouvements des masses d'air (*Beven*, 2012). Comme souvent en hydrologie, la disponibilité des données représente une contrainte majeure sur les choix de modélisation, pour alimenter les modèles mais aussi pour les valider. En effet, la difficulté d'estimer un stock de neige à l'échelle d'un territoire rend extrêmement complexe l'exercice de validation des modèles de simulation des stocks de neige conçus pour l'hydrologie. Le meilleur réalisme des certains modèles ne sera pas toujours reconnu si les écarts induits sur les débits sortant du modèle hydrologique sont trop peu visibles. Comme les composantes du modèle hydrologique conceptuel, le module neige n'est pas exempt des problèmes d'incertitudes de paramétrisation (*cf. He et al.*, 2011) et une section spécifique de nos travaux traitera des aspects liés à la transférabilité de ces paramètres neige dans un contexte de climat changeant.

Nos travaux portent sur l'estimation de la robustesse des modèles conceptuels. Dans ce cadre, nous avons fait le choix de nous restreindre aux modules neige conceptuels communément utilisés avec les modèles hydrologiques que nous souhaitions tester. Selon les configurations de test, les paramètres de ces modules seront fixés à des valeurs issues d'études préalables

3. Les bandes d'altitudes utilisées par *Valéry* (2010) sont de surfaces égales et non d'amplitude égale en altitude pour limiter les problématiques d'accumulation infinie sur les plus hautes couches évoquées ici.

ou calés. Lorsqu'ils sont calés, leur optimisation est réalisée simultanément à celle des paramètres des modèles hydrologiques par comparaison des débits simulés et observés (ils ne feront donc pas l'objet d'un traitement spécifique). Les configurations testées n'intègrent pas de prise en compte de la dimension glacière, observée à haute altitude, les bassins où ces comportements étaient significatifs ayant volontairement été retirés de notre échantillon. Nous sommes conscients que les modules neige utilisés dans cette thèse ne reflètent pas la diversité des approches existantes sur le sujet. Ce thème de recherche pourrait être largement approfondi par des travaux complémentaires, en comparant divers niveaux de complexité de modélisation ou en s'intéressant à des modules neige *à base physique.*

1.3. Évolutions du climat et incertitudes de projections

1.3.1. Changement climatique et études associées

L'impact des activités anthropiques sur le climat a fait l'objet d'une attention scientifique grandissante au fil des dernières décennies. La mise en commun des compétences scientifiques et techniques sur le sujet à travers le Groupe d'experts Intergouvernemental sur l'Évolution du Climat (*GIEC-IPCC*) a permis d'importants progrès des connaissances et l'établissement d'un consensus international sur cette thématique. L'impact anthropique sur les évolutions de climat passées, principalement sous forme d'émissions croissantes de gaz à effet de serre (*GES*), est aujourd'hui avéré. En dehors de quelques individualités questionnant encore ce point, la phase de démonstration de cet impact anthropique est donc derrière nous et l'attention se porte désormais très largement sur la projection des évolutions futures du climat. Les enjeux associés à ces évolutions sont extrêmement forts, celles-ci conduisant d'ici la fin du siècle à une diminution notable des calottes polaires, une augmentation du niveau moyen des océans et une modification de la disponibilité des ressources en eau à l'échelle régionale et locale (*IPCC*, 2007). Indépendamment des débats sur la perte totale ou partielle de stationnarité du climat (*cf. Milly et al.*, 2008; *Matalas*, 2012), les changements climatiques sont jugés certains mais leur amplitude reste à déterminer. Celle-ci dépend bien entendu des choix politiques de restriction d'émission de *GES* pour le siècle à venir. L'estimation de l'intensité de ces changements comporte également une part d'incertitudes induite par la chaîne de modélisation utilisée pour émettre ces projections.

Évaluer les impacts du changement climatique en hydrologie consiste notamment à produire des projections de l'évolution du débit des rivières à différentes échéances de temps (*ex.* milieu et fin du XXI^{e} siècle). Des études de ce type sont conduites de manière croissante à travers le monde (*ex. Caballero et al.* (2007); *Vicuna et Dracup* (2007); *Steele-Dunne et al.* (2008); *Chiew et al.* (2009); *Görgen et al.* (2010); *Schneider et al.* (2013)). Elles n'en restent pas moins

un exercice délicat du fait de la complexité de la chaîne de modélisation en jeu : (1) scénarios d'émission de gaz à effet de serre, (2) modélisation du climat mondial, (3) procédure de descente d'échelle, (4) modélisation hydrologique. Une illustration de cette chaîne est présentée à la figure 1.6. En quelques mots, elle se décompose de la façon suivante :

- Différents scénarios d'émission de *GES* sont proposés par le *GIEC* pour le siècle à venir, avec pour objectif de représenter diverses évolutions en lien avec des prospectives socio-économiques.
- Ces scénarios sont intégrés aux modèles climatiques de circulation générale (*GCMs*) qui simulent le climat à l'échelle du globe (*ex. CNCM33* du *CNRM*, *MK 3.5* du *CSIRO*, *ECHAM5* du *MaxPlanckInstitute*...).
- Si les maillages des *GCMs* ont été largement affinés au cours des dernières années, des méthodes de descente d'échelle (*MDE*) restent nécessaires pour adapter leurs sorties à un usage hydrologique au niveau des bassins versants. Ces méthodes se distinguent en deux catégories : les méthodes *dynamiques* qui consistent à greffer sur le *GCM* un modèle climatique de plus haute résolution couvrant un territoire plus restreint (modèle de circulation régionale - *RCM*) et les méthodes *statistiques* fondées sur des relations empiriques entre les variables de sortie du *GCM* et le climat local.
- Enfin, la chaîne de modélisation se termine par des modèles hydrologiques qui simulent la transformation entre les forçages climatiques (typiquement précipitations et température) et les débits sur le bassin étudié.

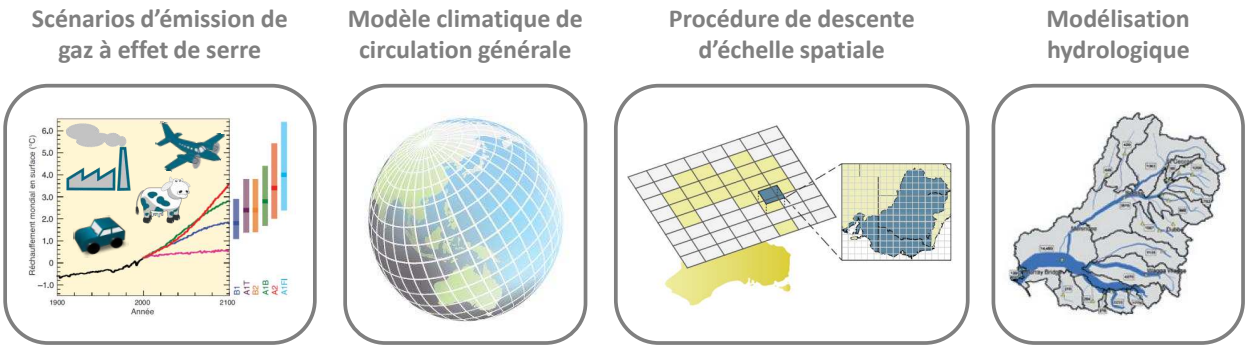


FIGURE 1.6.: Composantes principales de la chaîne de modélisation dans les études d'impact du changement climatique en hydrologie *(figure pour partie issue de CSIRO, 2012)*

1.3.2. Sources d'incertitudes dans les études d'impact

Compte tenu des multiples enjeux liés à l'eau, l'appréciation des incertitudes associées aux projections constitue un point majeur du travail de projection hydrologique sous changement climatique. Chacune des étapes de la chaîne de modélisation est une source d'erreurs qui se propagent en cascade jusqu'au résultat final. Le choix des scénarios d'émission et la modélisation faite au travers des *GCMs* sont généralement considérés comme les sources principales d'incertitudes dans la chaîne de modélisation (*Wilby et Harris*, 2006; *Prudhomme et Davies*, 2009b; *Kay et al.*, 2009; *Arnell*, 2011; *Teng et al.*, 2012). Relativement à celles-ci, les erreurs associées à la modélisation hydrologique ont longtemps été considérées faibles. L'hydrologue a pourtant la lourde tâche d'émettre (et de soutenir) les projections de débits futurs, résultat final de ce type d'études, sans qu'il ne maîtrise pour autant toute la chaîne de modélisation. Il alimente ses modèles avec des forçages climatiques futurs (issus du couple *GCM-MDE*) qui lui sont en quelque sorte imposés, autant d'ailleurs que leurs erreurs associées à ces forçages. De plus, cet exercice de projection des débits futurs fait rarement l'objet d'une validation, ce qui rend impossible la quantification précise des marges d'erreur de modélisation. Seules des méthodes indirectes, telles que les approches multi-modèles ou les études de sensibilité et robustesse des modèles, permettent *une certaine appréciation* des possibles variations sur les projections émises. Ces études doivent être menées indépendamment, au niveau de chaque maillon de la chaîne de modélisation. Quelques exemples sont : *Prudhomme et Davies* (2009a) pour les approches ensemblistes ; *Maurer* (2007) pour les études de sensibilité ; *Vaze et al.* (2010); *Merz et al.* (2011) pour les études de robustesse.

Nos travaux se concentrent sur l'étude de la contribution de la modélisation hydrologique (et notamment du paramétrage) dans l'incertitude des projections de la ressource en eau. Au-delà du choix du modèle, le fait de transférer temporellement des jeux de paramètres entre périodes pose un certain nombre de questions, d'autant plus lorsque les conditions climatiques entre périodes diffèrent sensiblement (*cf.* section 1.1.4). Bien que les choix du scénario d'émission de *GES* ou du *GCM* restent des sources très importantes d'incertitudes, certains travaux suggèrent que la contribution du modèle pluie-débit aux erreurs possibles de projections sous changement climatique ne doit pas pour autant être négligée (*Wilby*, 2005; *Vaze et al.*, 2010). Sans se soucier du classement des contributeurs à incertitude globale de projection, l'hydrologue devrait quoi qu'il en soit évaluer l'élément qui le concerne dans la chaîne de modélisation (*Singh et al.*, 2011).

1.4. Synthèse du contexte et objectifs de cette thèse

Cette thèse est centrée sur la robustesse des modèles conceptuels pluie-débit et de leur paramétrisation dans un contexte de climat changeant, avec un point d'intérêt sur les zones de montagne. Ce chapitre nous a permis de détailler notre objet d'étude et de présenter les éléments de contexte facilitant la compréhension de notre démarche.

Le manque de compréhension et de mesure des processus à l'échelle d'un bassin versant conduit à favoriser des approches conceptuelles de modélisation en hydrologie. Ces approches nécessitent une phase d'ajustement de paramètres (un calage) qui soulève un certain nombre d'interrogations sur les possibilités de transférer ces paramètres pour utiliser le modèle sous d'autres conditions. Les applications pratiques conduisent pourtant les hydrologues à utiliser ce type de modèles dans des conditions différentes de celles où ils ont été calés, comme par exemple dans les études d'impact du changement climatique. Or l'estimation des incertitudes associées aux simulations émises est un point crucial pour l'hydrologue-modélisateur, compte tenu des enjeux liés aux études en hydrologie (dimensionnement d'ouvrages, projections sur la disponibilité de la ressource en eau, ...). Si certaines pistes montrent que la modélisation hydrologique peut constituer une source significative d'incertitudes dans un contexte de climat évolutif, les études approfondies sur le sujet restent très peu nombreuses.

Nos travaux de recherches tentent d'apporter des éléments de réponse aux questions suivantes :

- *Quel est le niveau de robustesse des modèles hydrologiques conceptuels face à des conditions climatiques non stationnaires ?*

  Nous nous interrogerons sur les procédures de test et d'analyse permettant l'obtention de diagnostics robustes sur la *capacité d'extrapolation climatique* des modèles précipitations-débit. Nous étudierons les pertes de performances en lien avec les conditions de transfert des paramètres mais nous chercherons également à mieux caractériser les erreurs de simulation associées (*ex.* volumes, dynamiques, débits extrêmes, simulation de la neige...). Pour limiter les risques de concentrer nos interprétations sur des comportements ou un contexte local trop spécifique, cette approche intégra des comparaisons entre de nombreux bassins et entre plusieurs modèles hydrologiques de structures différentes. Les bassins versants montagneux où les écoulements sont thermosensibles (accumulation/fonte de neige) constitueront un point d'analyse spécifique, compte tenu des fortes augmentations de température attendues dans le cadre du changement climatique.

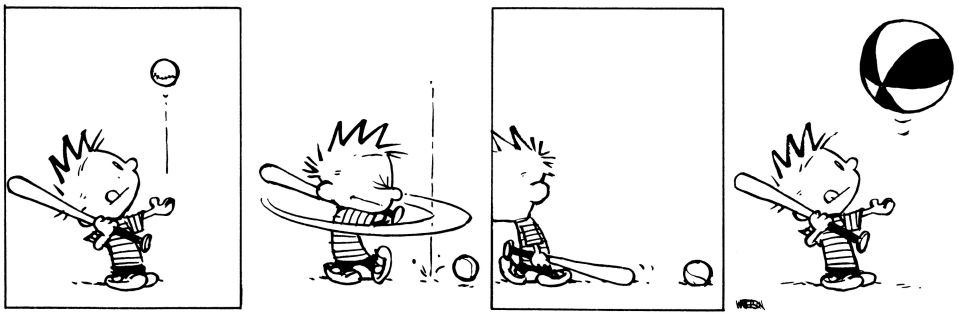
- *Quelles sont les causes de ce manque de robustesse et y a-t-il des façons de s'en prémunir ?*

  Nous tenterons de conduire un diagnostic approfondi pour expliquer les causes des problèmes de transférabilité de paramètres constatés dans les séries de tests. Nous nous lancerons alors à la recherche de diverses pistes pour améliorer cette robustesse et réduire ainsi les erreurs de modélisation dans les situations de climat changeant.

2

Comment évaluer la robustesse climatique des modèles pluie-débit ?

La nécessité de bien choisir ses outils...



Calvin and Hobbes, 20/04/1991, Bill Watterson©

2.1. Méthodologie de tests et d'analyse

Les travaux conduits dans cette thèse portent sur la robustesse des modèles hydrologiques conceptuels dans un contexte de climat changeant. Notre objectif est d'apprécier la *capacité d'extrapolation climatique* des paramètres des modèles pluie-débit. Cette approche se place notamment dans le contexte des études d'impact du changement climatique, où des modèles calés sur une période historique sont utilisés dans des conditions climatiques différentes. Le principe de *validation* des modèles repose sur la comparaison des débits simulés avec des observations. Pour nos travaux, nous sommes donc contraints de travailler sur des données historiques, bien que la variabilité climatique sur les périodes instrumentées ne traduise pas nécessairement l'amplitude des changements de climat attendus.

Dans l'esprit de la discussion menée par *Andréassian et al.* (2009) sur l'évaluation des modèles, nous avons élaboré une méthodologie de *"crash test"*. Nous cherchons effectivement à placer les modèles dans des conditions défavorables de transposition temporelle, avec de fortes différences entre les conditions de calage et d'utilisation pour étudier au mieux leurs limites. Nous nous inspirons donc largement des tests proposés par *Klemeš* (1986) mais en généralisant l'approche pour tenter d'établir des relations entre les erreurs de modélisation et les différences qui caractérisent les périodes de calage et de validation.

2.1.1. Les outils de mesure de la robustesse des modèles

Le *differential split-sample test* (*DSST*) proposé par *Klemeš* (1986), permet une évaluation des modèles dans des conditions climatiques contrastées (exemples d'utilisation discutés dans la section 1.1.4). Son application sur un bassin versant comprend trois étapes :

1. des sous-périodes (souvent en nombre réduit) sont sélectionnées dans la série d'observations disponible en fonction de leurs caractéristiques climatiques ou hydrologiques (*ex.* volumes moyens précipités, température de l'air, etc.) ;
2. un test de calage-contrôle est pratiqué entre les sous-périodes ainsi sélectionnées ;
3. les performances du modèle en contrôle sont comparées afin d'évaluer si elles évoluent notablement avec les écarts climatiques entre les sous-périodes de calage et de contrôle.

Cette procédure, schématisée à la figure 2.1, a néanmoins deux défauts majeurs pour l'obtention de conclusions généralisables. Premièrement, elle nécessite de connaître à l'avance quelle caractéristique climatique risque de jouer un rôle limitant lors du transfert des paramètres du modèle. Si l'influence de différentes caractéristiques climatiques est étudiée, il est généralement difficile de comparer les résultats car les années sélectionnées pour ces évaluations varient avec la caractéristique climatique considérée (*ex.* constitution d'un couple chaud-froid, puis d'un couple humide-sec qui est différent). En reprenant les notations de la figure 2.1, si nous

étudions l'influence d'un écart de précipitation moyenne sur le transfert des paramètres, nous sélectionnerons des sous-périodes A_1, B_1, C_1, D_1, puis si nous recommençons en étudiant cette fois l'influence de la température de l'air nous sélectionnerons des sous-périodes A_2, B_2, C_2, D_2 vraisemblablement différentes des précédentes (la période la plus sèche n'étant pas nécessairement la plus chaude). Le deuxième défaut de l'approche est le nombre de tests de transfert souvent restreint, puisque seules deux ou trois périodes contrastées et indépendantes les unes des autres sont généralement identifiables. Cela limite les possibilités d'établir des conclusions générales et d'identifier de manière robuste le degré de transférabilité des paramètres.

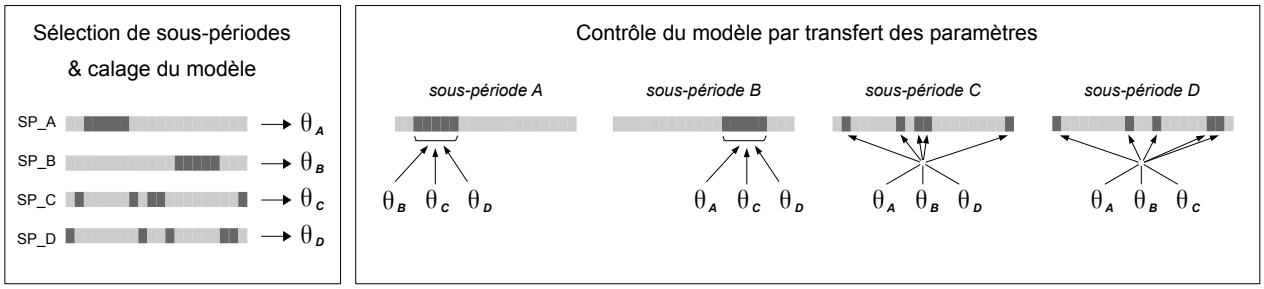


FIGURE 2.1.: Méthodologie de test de calage-contrôle *SST et DSST*

Les sous-périodes utilisées pour caler et valider le modèle peuvent être continues (ex. A et B) ou résulter d'une sélection d'années réparties sur la chronique (ex. C et D).

Selon le choix du modélisateur, les sous-périodes sélectionnées peuvent être continues ou constituer un agrégat d'années hydrologiques n'étant pas nécessairement côte à côte. Cette seconde approche de ré-échantillonnage est communément appelée *bootstrap* et présente plusieurs avantages. D'une part, elle permet la constitution de sous-périodes plus extrêmes en regroupant par exemple toutes les années les plus sèches d'un côté et toutes les années les plus humides de l'autre. Dans le cadre du *DSST*, elle autorise ainsi des tests de transfert de paramètres entre des situations plus contrastées. D'autre part, ce ré-échantillonnage peut permettre de s'affranchir des problèmes liés à des erreurs évolutives sur les entrées telles que la dérive d'une courbe de tarage ou l'enrichissement/appauvrissement d'un réseau de pluviomètres. Néanmoins, elle présente un inconvénient majeur lié au fait que les conditions d'humidité du bassin au début de chaque année ne sont pas identiques. En effet, les réserves d'eau stockées dans le sol (en surface et en profondeur) évoluent en fonction du climat et ne sont pas au même niveau après une année très sèche ou très humide. Cette *mémoire* du bassin versant doit théoriquement être prise en compte lorsque les conditions climatiques d'une période sont caractérisées. En effet, on imagine facilement qu'une année moyenne en termes de précipitations mais précédée d'une période de sècheresse sévère pourra se révéler plus extrême pour le bassin qu'une année sèche précédée d'années humides. Ainsi, la prise en compte de cette mémoire ne peut pas être négligée et devient alors très délicate lorsque la période à caractériser est un agglomérat d'années indépendantes.

2.1.2. Proposition d'une procédure de test

À notre connaissance, les procédures standards de *split-sample* (*SST*) et *differential split-sample test* (*DSST*) sont presque toujours utilisées pour conduire des tests ponctuels, en nombre limité. Pour le *DSST*, ces tests restent souvent focalisés sur l'étude d'impact d'une ou deux variables climatiques sur la transférabilité des paramètres d'un modèle. Afin de surmonter ces limites, nous proposons une généralisation de ces procédures, notre objectif étant de permettre le test de modèles hydrologiques sur un nombre et une variété de situations aussi grands que possible. Nous souhaitons considérer toutes les situations : aussi bien les transferts sous climat équivalents (type *SST*) que très contrastés (type *DSST*).

Principe

La procédure que nous proposons est illustrée à la figure 2.2 et se compose des étapes suivantes :

1. Une fenêtre glissante de longueur choisie est utilisée pour découper la chronique d'observations disponible en sous-périodes. Cette fenêtre est déplacée d'une année hydrologique entre deux sous-périodes, les chevauchements entre périodes étant donc autorisés. Sur la figure 2.2, ces sous-périodes sont indiquées en gris foncé, tandis que les zones en gris clair correspondent au reste de la chronique disponible.
2. Le modèle hydrologique que l'on souhaite tester est calé sur chaque sous-période à l'aide d'une fonction objectif préalablement définie. On retient ainsi un jeu de paramètres θ par sous-période. À cette étape, n'importe quelle fonction objectif ou algorithme de calage peuvent être utilisés. Dans nos tests, nous opterons toutefois pour une procédure automatique de sélection des paramètres assurant une répétabilité de l'exercice.
3. Pour chaque sous-période de calage, le jeu de paramètres optimal est transféré afin de réaliser tous les tests possibles en validation sur les autres sous-périodes. Celles qui chevauchent la sous-période de calage ne sont pas considérées pour maintenir le cadre strict du calage-contrôle entre périodes indépendantes. De plus, une simulation de référence est construite pour chaque sous-période en utilisant le jeu de paramètres optimisé sur celle-ci. Notons que le nombre de tests en contrôle n'est pas constant entre toutes les sous-périodes mais cela n'est pas problématique car tous les résultats seront analysés en même temps.

Chaque sous-période joue alternativement le rôle de période de calage (*donneuse* de paramètres) ou de validation (*receveuse* de paramètres). Nous utiliserons donc par la suite les notations D et R pour distinguer les périodes concernées par le transfert. Par exemple, on notera $\hat{Q}_{R,k}[\theta_D]$ le débit simulé au pas de temps k de la sous-période *receveuse* R en utilisant le jeu de paramètres θ, issu du calage sur la sous-période *donneuse* D. La construction

des sous-périodes de test se fait indépendamment de considérations sur leurs caractéristiques climatiques, celles-ci n'intervenant que plus tard lors des analyses des performances en validation. Les deux étapes qui composent cette démarche de tests (phase de construction des périodes d'intérêt et phase de test du modèle) sont donc réalisées dans un sens opposé à la pratique habituelle puisque nous procédons ainsi : (a) tout d'abord, de nombreux tests en calage-contrôle sont effectués sans distinction sur les conditions de transfert des paramètres, (b) ensuite, les conditions de transferts sont étudiées et mises en face des performances du modèle en validation pour déterminer a posteriori si un lien peut être établi entre les variations de qualité des simulations et un facteur climatique donné. À notre connaissance, une telle généralisation du *split-sample test*, où aucun choix préalable n'est fait sur les sous-périodes testées, n'a pas encore été utilisée sur un grand nombre de cas.

Nous avons intitulé cette approche le *generalized split-sample test* (*GSST*), par analogie aux dénominations de *Klemeš* (1986).

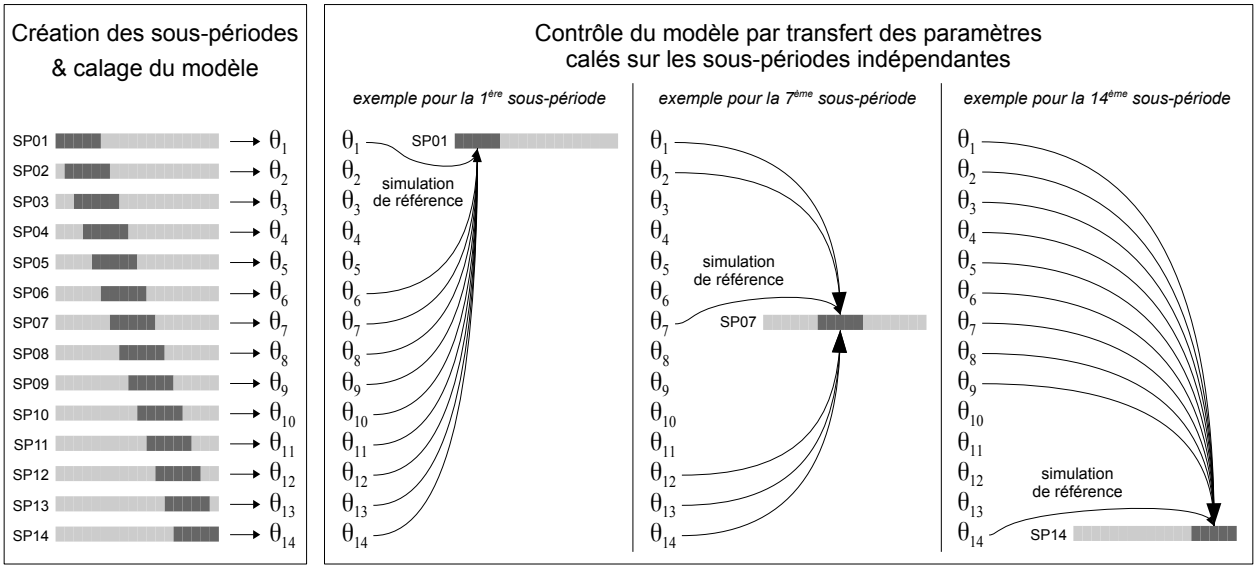


FIGURE 2.2.: Méthodologie de test de calage-contrôle généralisée *GSST*
(dans cet exemple, une chronique totale de 18 ans est découpée en sous périodes de 5 ans)

Nombre de tests réalisables

Le nombre de tests de transfert (N_{tests}) réalisés pour chaque bassin est fonction de la longueur de la série d'observations (L_{chro}) et de la longueur choisie pour les périodes de calage et de contrôle (L_{per}). En vérifiant l'inégalité (assez logique) $L_{chro} \geq 2\, L_{per}$, alors le nombre N_{tests} peut être calculé par la formulation :

$$N_{tests} = (L_{chro} - 2\, L_{per} + 1) \cdot (L_{chro} - 2\, L_{per} + 2) \tag{2.1}$$

Une chronique longue de 18 années découpées en sous-périodes de 5 ans (tel qu'illustré sur la figure 2.2) permettra la réalisation de 90 *split-sample test*s. Dans les tests présentés plus loin, nous disposons plutôt de chroniques de 30 ans que nous découpons en sous-périodes de 10 ans, ce qui permet 132 *split-sample test*s. Des périodes de mise en route sont couramment utilisées pour initialiser les états du modèle, les critères d'erreurs (en calage ou validation) n'étant pas calculés sur celles-ci. Dans le cadre de notre méthodologie, nous ne tenons pas compte de ces périodes d'initialisation dans la constitution des couples de tests, *c.-à-d.* deux périodes sont jugées indépendantes même si l'une d'entre elles chevauche la période de mise en route de l'autre. Cette longueur d'initialisation doit simplement être soustraite pour le calcul de L_{chro} (*ex.* si la série d'observations comporte 30 années et que le modèle est initialisé sur une année, alors on considèrera $L_{chro} = 29$ pour le calcul de N_{tests}).

Avantages

La procédure *GSST* devrait permettre de surmonter les limites abordées précédemment :

1. Elle permet un nombre important de cas d'analyse, qui couvrent de façon continue l'espace des écarts de climat existant dans la série d'observations. Ce grand nombre de tests devrait améliorer la robustesse des interprétations.
2. L'utilisation de périodes continues plutôt que discontinues comme dans les approches de *bootstrap* permet de limiter les effets des années intermédiaires dans l'analyse des résultats (*cf.* discussion section 2.1.1). Les caractéristiques de chaque période (*ex.* pluie, température) pourront être strictement estimées sur les mêmes pas de temps que les critères d'erreur.
3. Elle nous semble moins subjective puisqu'aucun choix de sous-période ou de caractéristique de climat étudié n'est fait *a priori* (*c.-à-d.* avant la réalisation des tests).
4. L'influence de n'importe quelle caractéristique climatique sur la robustesse du modèle peut être étudiée. En effet, toutes les configurations possibles (*c.-à-d.* tous les couples de périodes) sont testées et aboutissent à une liste unique de performances du modèle en validation. Lors de l'analyse de la robustesse des modèles face à un changement de climat, passer d'une variable climatique à une autre (*ex.* pluie $\rightarrow$ température) revient simplement à exprimer cette même liste de performances en changeant la référence choisie pour la visualisation. Notons que cette comparaison n'était pas possible dans l'usage courant du *DSST* puisque le panier de périodes de contrôle n'était pas constant.

Inconvénients

La méthodologie *GSST* est critiquable sur le point suivant : si le nombre de tests est fortement augmenté comparativement aux usages préalables du *DSST*, cela s'accompagne d'une forte redondance de l'information dans les résultats des tests puisque les recouvrements sont autorisés entre les périodes de validation. La procédure *GSST* est toutefois conforme au cadre strict du *split-sample test* original, puisque les sous-périodes composant chaque couple de calage-contrôle sont toujours indépendantes l'une de l'autre. Lors de l'élaboration de cette méthodologie, notre objectif fut de maximiser le nombre de tests en ne se privant d'aucun couple de sous-périodes indépendantes. En effet, ne pas autoriser les recouvrements risquerait dans certains cas d'écarter les couples de sous-périodes les plus contrastés, ou alors l'opération devrait être répétée et conduirait à faire de multiples analyses par bassin au lieu d'une seule dans la configuration proposée ici. Le choix a donc été fait de conserver toutes les périodes de calage disponibles même si elles se chevauchent entre elles [1]. En conséquence, nous devrons garder à l'esprit que les critères de corrélations pouvant être calculés dans nos travaux, le seront entre des points n'étant pas rigoureusement indépendants. Cela ne modifiera toutefois en rien la validité des conclusions, ce qui nous intéresse étant une appréciation générale sur l'existence de liens entre évolutions des performances et changements subis lors de transfert des paramètres.

2.1.3. Méthodes d'analyse des résultats

Cette section présente les techniques numériques et graphiques utilisées dans nos travaux pour évaluer la robustesse des modèles hydrologiques conceptuels face à un climat en évolution.

Critère de mesure de robustesse

Considérons un test de calage-contrôle où un jeu de paramètres θ_D est transféré d'une période D (période *donneuse* où le calage a été effectué) vers une période R (période *receveuse* utilisée pour le contrôle). Avec ces notations, le critère $RMSE$ d'erreur quadratique moyenne, le critère NSE de *Nash et Sutcliffe* (1970) et le critère KGE proposé par *Gupta et al.* (2009) comme alternative au précédent, s'écrivent :

$$RMSE_{\,D\to R} = \sqrt{\frac{1}{n}\cdot\sum_{k=1}^{n}\left(\widehat{Q}_{R,k}\left[\theta_D\right]-Q_{R,k}\right)^2} \tag{2.2}$$

$$NSE_{\,D\to R} = 1-\frac{\sum\limits_{k=1}^{n}\left(\widehat{Q}_{R,k}\left[\theta_D\right]-Q_{R,k}\right)^2}{\sum\limits_{k=1}^{n}\left(\overline{Q_R}-Q_{R,k}\right)^2} \tag{2.3}$$

1. La procédure de découpage étant systématique, une très grande majorité des années auront le même niveau de redondance et il n'y a donc pas de poids plus forts sur certaines années (seules les quelques années aux extrémités de la chronique auront une redondance plus faible).

$$KGE_{\,D\to R} = 1 - \sqrt{(\rho_{kge}-1)^2 + (\alpha_{kge}-1)^2 + (\beta_{kge}-1)^2} \tag{2.4}$$

avec
$$\rho_{kge} = r_{\,pearson\,\left(\widehat{Q}_R[\theta_D]\sim Q_R\right)} = \frac{\sum_{k=1}^{n}\left(\widehat{Q}_{R,k}[\theta_D] - \overline{\widehat{Q}_R[\theta_D]}\right)\cdot\left(Q_{R,k}-\overline{Q_R}\right)}{\sqrt{\sum_{k=1}^{n}\left(\widehat{Q}_{R,k}[\theta_D] - \overline{\widehat{Q}_R[\theta_D]}\right)^2 \cdot \sum_{k=1}^{n}\left(Q_{R,k}-\overline{Q_R}\right)^2}}$$

$$\alpha_{kge} = \frac{\sigma(\widehat{Q}_R[\theta_D])}{\sigma(Q_R)} = \frac{EcartType\left(\widehat{Q}_R[\theta_D]\right)}{EcartType(Q_R)} = \frac{\sqrt{\sum_{k=1}^{n}\left(\widehat{Q}_{R,k}[\theta_D] - \overline{\widehat{Q}_R[\theta_D]}\right)^2}}{\sqrt{\sum_{k=1}^{n}\left(Q_{R,k}-\overline{Q_R}\right)^2}}$$

$$\beta_{kge} = \frac{\mu(\widehat{Q}_R[\theta_D])}{\mu(Q_R)} = \frac{\overline{\widehat{Q}_R[\theta_D]}}{\overline{Q_R}} = \frac{\sum_{k=1}^{n}\widehat{Q}_{R,k}[\theta_D]}{\sum_{k=1}^{n}Q_{R,k}}$$

où $Q_{R,k}$ est le débit observé au pas de temps k de la période R et $\hat{Q}_{R,k}\,[\theta_D]$ est le débit simulé correspondant en utilisant le jeu de paramètres θ_D calé sur la période D.

Analyser les résultats sous forme de critère de biais (*ex.* sur les volumes annuels mais aussi saisonniers ou mensuels) présente l'avantage notable de manipuler des valeurs adimensionnelles parfaitement comparables entre différentes périodes ou bassins versants. Dans leur étude de la stabilité temporelle des paramètres de leur modèle, *Merz et al.* (2011) représentent ainsi des critères de biais en calage et en contrôle sur le même schéma pour montrer que les tendances qu'ils observent sur les erreurs des modèles peuvent être reliées aux conditions de transfert des paramètres du modèle (*cf.* figure 1.3, page 14).

Cependant, lorsque des critères d'erreurs tels que le $RMSE$ ou le NSE sont utilisés, cette comparaison devient plus délicate. Les valeurs $RMSE_{\,D\to R}$ ou $NSE_{\,D\to R}$ obtenues pour différentes périodes de calage D mais une unique période de simulation R peuvent être comparées, puisque les pas de temps où elles sont calculées sont strictement identiques. Néanmoins, le critère d'erreur quadratique est dépendant des volumes en jeu et aura ainsi tendance à être plus élevé pour les périodes (ou bassins) dont les débits sont plus importants. Le critère de *Nash et Sutcliffe* (1970) est quant à lui construit comme un rapport entre le carré des erreurs du modèle et la variance des débits observés. Dans certains cas, on peut considérer que les changements de variance ou volumes entre périodes ont un impact réduit sur la comparaison des résultats de validation du modèle. On se rapproche des limites de validité de cette hypothèse lorsque le panier de périodes de calage/contrôle devient hétérogène en termes de conditions climatiques et de volumes écoulés. Ce problème devient encore plus marqué lorsque les résultats de bassins d'horizons divers sont regroupés pour l'analyse.

En conséquence, nous avons décidé de ne pas construire nos analyses autour de différences de critère NSE pour diagnostiquer l'influence des écarts de climat dans la transférabilité des paramètres de modèles pluie-débit. Nous avons préféré construire le critère de robustesse suivant (noté *MRC* pour *Model Robustness Criterion*) :

$$MRC_{D\to R} = \frac{\varepsilon_{D\to R}}{\varepsilon_{R\to R}} - 1 \tag{2.5}$$

avec ε un critère d'erreur qui sert de fonction objectif lors de la procédure $GSST$.

Ce critère a été construit autour de l'idée d'une évaluation de la pertinence d'un jeu de paramètres comparativement à un jeu de référence optimal (issu d'un calage). $\varepsilon_{D\to R}$ est un critère d'évaluation des erreurs sur la période R du modèle calé sur D (par exemple $\varepsilon_{D\to R} = RMSE_{D\to R}$). $\varepsilon_{R\to R}$ devrait quant à lui correspondre à la plus faible valeur de ε sur R avec le modèle considéré[2]. Comme $\varepsilon_{D\to R}$ et $\varepsilon_{R\to R}$ sont calculés sur les mêmes pas de temps, ils peuvent être comparés pour constituer le critère adimensionnel proposé ici. L'interprétation des valeurs de *MRC* est relativement simple. Ce critère ne prend théoriquement que des valeurs positives. Il vaut 0 lorsque le jeu de paramètres optimisé sur la période D permet une simulation de qualité égale à celle obtenue par calage sur R. Ensuite, plus sa valeur est élevée, moins le jeu de paramètres θ_D est valable pour simuler les débits de la période R. Par exemple, une valeur de *MRC* de 1.0 signifie une augmentation d'erreur relative de 100% (soit un doublement de l'erreur) liée au fait d'utiliser un jeu de paramètres non-optimal (erreur estimée au travers de ε). Par la suite, ces valeurs seront souvent exprimées sous forme de pourcentages. On notera enfin qu'une valeur négative de *MRC* signifierait qu'un jeu de paramètres transféré d'une autre période serait plus pertinent qu'un jeu de paramètres calé sur la période cible. Cela indiquerait une défaillance de la méthode d'optimisation des paramètres sur R, l'optimum global n'y ayant pas été identifié correctement.

Le critère *MRC* proposé ici permet de résoudre une majorité des difficultés mentionnées précédemment concernant le travail de comparaison de performances sur des sous-périodes très variées. La seule condition nécessaire pour permettre la comparaison de *MRCs* issus de conditions diverses est que le rapport $\varepsilon_{D\to R}/\varepsilon_{R\to R}$ soit indépendant des caractéristiques de la période et/ou du bassin concernés (en termes de volumes ou de variabilité). Par exemple, $\varepsilon = NSE$ ne peut pas être utilisé puisque le terme de variance au dénominateur ne disparaît pas dans le calcul du rapport $\varepsilon_{D\to R}/\varepsilon_{R\to R}$. À l'inverse, on pourra utiliser $\varepsilon = RMSE$, $\varepsilon = 1 - NSE$, ou $\varepsilon = 1 - KGE$. On remarquera d'ailleurs que lorsque le critère *MRC* est construit autour d'une erreur quadratique, alors 2-*MRC* s'apparente à un critère de type NSE dans lequel le modèle de référence au dénominateur a été modifié (*Seibert*, 2001; *Lerat et al.*, 2012). Dès lors que le critère ε remplit exigence que nous venons d'évoquer, les *MRCs* peuvent être utilisés pour étudier la transférabilité des jeux de paramètres dans des conditions contrastées.

2. Une vigilance est toutefois requise sur le choix du critère d'erreur ε car s'il conduit à des valeurs trop proches de 0 dans les situations optimales, la division par $\varepsilon_{R\to R}$ sera problématique.

Ils ne devraient pas être trop fortement influencés par le niveau d'erreur du modèle inhérent à une période donnée (niveau causé par les erreurs de données et de structure du modèle) et permettre la mise en valeur des conséquences liées à l'utilisation d'un jeu de paramètres non optimal. Ainsi, un réel diagnostic de la *capacité d'extrapolation climatique* des modèles peut être conduit simultanément sur un large échantillon de cas.

Représentation graphique

Les valeurs du critère *MRC* (*cf.* équation 2.5) nous informent sur la pertinence relative d'un jeu de paramètres pour simuler les débits d'une période différente de celle où il a été calé (transfert $D \rightarrow R$). Dans chaque cas, on peut alors représenter graphiquement la valeur du *MRC* en fonction d'une différence de conditions entre les périodes D et R. Un exemple de cette représentation est fourni à la figure 2.3, où la variable de climat considérée est la température moyenne sur la sous-période. Pour mieux comprendre la façon dont ce graphique est construit et analysé, considérons plusieurs étapes :

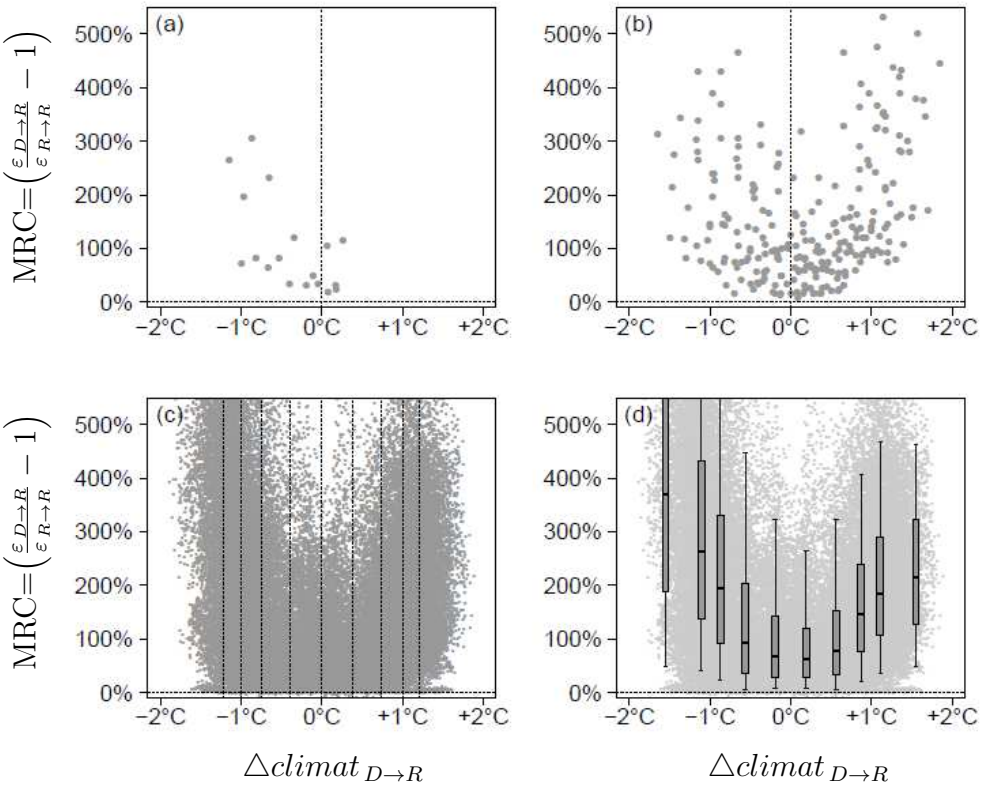


Figure 2.3.: Méthode d'analyse visuelle de la capacité d'extrapolation climatique des paramètres : $\triangle\varepsilon = f(\triangle climat)$

Représentation par nuage de points du type $MRC = f(\triangle climat)$, en considérant ici $\triangle climat = \triangle\overline{T}_{D\rightarrow R}$ et des sous-périodes de test de 10 ans. Les boîtes à moustaches correspondent aux quantiles 5, 25, 50, 75 et 95%.

[figure 2.3.a]

Nous considérons une sous-période *receveuse* R sur un bassin versant donné. Un certain nombre de jeux de paramètres peuvent y être transférés, ceux-ci étant obtenus par calage sur des sous-périodes indépendantes du même bassin (*cf.* section 2.1.2). Nous obtenons ainsi des erreurs en validation ($\varepsilon_{D\to R}$), que nous pouvons comparer au niveau d'erreur de référence obtenu en calage sur la sous-période d'étude ($\varepsilon_{R\to R}$). Pour chaque *split-sample test,* nous disposons donc d'une valeur de *MRC* que nous positionnons en fonction des différences de conditions subies lors du passage de D vers R. Sur cette exemple, la majorité des jeux transférés ont été calés dans des conditions plus froides ($Temp_D < Temp_R$) .

[figure 2.3.b]

Nous pouvons superposer sur le même graphique les points correspondant aux tests sur d'autres périodes receveuses du même bassin. Pour rappel, l'objectif de cette représentation est d'évaluer si les erreurs du modèle peuvent être reliées à un écart de climat lors du transfert du jeu de paramètres. Nous nous intéressons à la forme générale du nuage de points plutôt qu'au positionnement de quelques points spécifiques et nous évaluons si les erreurs de simulation augmentent avec l'écart de climat entre calage et validation.

[figure 2.3.c]

Dans la mesure où toutes les variables représentées sont exprimées relativement, la superposition des résultats peut être poursuivie en ajoutant les points correspondant aux tests sur d'autres bassins. C'est bien là tout l'intérêt de notre procédure de test et d'expression des résultats sous forme de *MRC*. Ce critère numérique et la représentation graphique associée ont été conçus pour pouvoir mener de telles analyses sur de nombreux cas simultanément. Ici encore, c'est la forme générale du nuage de points qui nous intéresse même si elle devient plus difficile à percevoir compte tenu du nombre très important de points. En effet, les valeurs absolues de *MRC* observées en ordonnée présentent peu d'intérêt et c'est bien leurs variations lors d'un déplacement selon l'axe des abscisses qui révèlent si la variable climatique considérée est une source explicative des pertes de performance d'un modèle lors du transfert de ses paramètres.

[figure 2.3.d]

Pour faciliter l'analyse de la forme du nuage, celui-ci est divisé en différentes bandes verticales comportant un même nombre de points (plutôt que de largeur égale) et des quantiles sont calculés sur chacune d'elles (quantiles 5, 25, 50, 75 et 95%). Les boîtes à moustaches ainsi constituées permettent d'apprécier très rapidement la forme générale du nuage et la représentativité d'une tendance éventuellement détectée. La densité de points diminuant pour les écarts de climat les plus marqués, les bandes verticales sont de largeur variable et donc l'espacement entre les boîtes non régulier. Cependant, chaque point du graphique

possède bien "sa" boîte à moustaches (les espaces parfois plus larges situés sur les bords des graphiques restant couverts par la boîte à proximité).

La hauteur des boîtes nous informe sur la gamme d'erreurs obtenue lors des tests de transfert dans les conditions concernées (*ex.* transferts entre des périodes différentes de 1°C $\Leftrightarrow$ *MRC* entre 150 et 400%). Comparer ensuite le positionnement relatif des boîtes, les unes par rapport aux autres nous indique si une différence de climat entre calage et validation conduit ou non à une perte de performance du modèle liée à un transfert inapproprié de ses paramètres. Un positionnement vertical homogène des boîtes, combiné à une large dispersion de chaque boîte, invalide l'hypothèse d'une relation de causalité entre un changement de condition climatique (pour la variable considérée) et les erreurs de modélisation. À l'inverse, une dispersion raisonnable et un positionnement des boîtes plus élevé sur les bords qu'au centre révèlent une possible corrélation entre les conditions climatiques du transfert de paramètres et la qualité des simulations obtenues.

Sur l'exemple illustré, on voit apparaître une forme générale en U, avec des erreurs liées au transfert des paramètres cinq fois plus importantes lorsque la température moyenne diffère de 1.5°C entre les périodes que lorsque cette température est identique.

Précisons enfin que le $\triangle climat$ placé en abscisse sera calculé de façon légèrement différente selon la variable climatique considérée. Deux méthodes sont considérées pour mesurer un écart de conditions climatiques : une différence absolue avec $\triangle climat_{\,D\Rightarrow R} = clim.[D] - clim.[R]$, ou une différence relative avec $\triangle climat_{\,D\Rightarrow R} = (\, clim.[D] - clim.[R] \,) \,/\, clim.[D]$. Pour chaque variable, nous retiendrons la méthode de calcul qui a plus le plus de sens tout en permettant une superposition des nuages issus de bassins très différents. Par exemple, une différence absolue sera utilisée pour les écarts de température ($\triangle \overline{T}_{\,D\Rightarrow R} = \overline{T}_D - \overline{T}_R$) et une différence relative sera utilisée pour les écarts de précipitations ($\triangle \overline{P}_{\,D\Rightarrow R} = (\overline{P}_D - \overline{P}_R)/\overline{P}_R$).

Type de corrélations recherché

Les baisses de performance des modèles lors du transfert de paramètres sont étudiées à travers les variations du critère KGE [3] qui constitue le critère de calage des paramètres dans nos analyses. Plutôt que d'analyser directement les valeurs absolues de KGE, nous préférons son expression sous la forme relative d'un *MRC*, dans lequel $\varepsilon = 1 - KGE$. Comme nous venons de l'évoquer, cette transformation permet la mise en évidence des erreurs liées au transfert des paramètres du modèle. Les trois composantes du KGE peuvent également être utilisées pour caractériser les erreurs commises : ρ_{kge}, α_{kge} et β_{kge} correspondant à une corrélation de *Pearson*, un rapport des écarts-types et un rapport des moyennes entre les débits simulés et observés (*cf.* équation 2.4, page 32).

3. $KGE = 1 - \sqrt{(\rho_{kge} - 1)^2 + (\alpha_{kge} - 1)^2 + (\beta_{kge} - 1)^2}$

Les variations de ces critères d'erreur de simulation sont confrontées aux écarts existants entre les périodes de calage et de validation de nos modèles. Une grande majorité des indicateurs que nous considérerons pour caractériser les périodes traduisent les conditions climatiques. Toutefois, d'autres variables peuvent être utilisées telles que le décalage purement temporel entre les périodes, ou les évolutions de l'impact anthropique sur le bassin (ex. prélèvements agricoles, utilisation intensive d'engrais, imperméabilisation des sols...).

Pour réaliser cette confrontation entre $\triangle erreurs$ et $\triangle caractéristiques$, nous avons opté pour des approches simples basées sur des critères de corrélation et des analyses visuelles tels que ceux présentés sur la figure 2.4. Plutôt que d'utiliser le critère de *Pearson*, qui traduit le degré de corrélation linéaire entre des valeurs, nous utiliserons le critère de *Spearman* (r_s) qui traduit le degré de corrélation linéaire entre leurs rangs. Tout comme le premier, il varie entre -1 et +1, mais il a le grand avantage de ne pas faire d'hypothèse sur le type de relation entre les deux variables (*ex.* linéaire, exponentielle...). Il s'intéresse uniquement à caractériser la monotonie d'une relation. Des valeurs de -1 et +1 indiquent respectivement une anti-corrélation et une corrélation parfaitement monotone entre les variables considérées (*c.-à-d.* $\triangle x \nearrow \Rightarrow \triangle y \searrow$ et $\triangle x \nearrow \Rightarrow \triangle y \nearrow$). Un r_s proche de 0 est le signe d'une absence de relation entre les variables. Sa formulation est donnée par l'équation 2.6, où kx_i est le rang de l'abscisse du point i au sein des abscisses disponibles sur l'échantillon de points et ky_i est le rang de l'ordonnée du point i au sein des ordonnées disponibles sur l'échantillon de points.

$$r_s = \frac{\sum_i \left(kx_i - \overline{kx}\right) \cdot \left(ky_i - \overline{ky}\right)}{\sqrt{\sum_i (kx_i - \overline{kx})^2 \sum_i (ky_i - \overline{ky})^2}} \tag{2.6}$$

Le critère de *Spearman* sera très largement utilisé pour nous guider dans nos analyses des nombreux résultats de tests issus de la procédure $GSST$. Cependant, nous serons amenés à le calculer différemment selon le type d'erreurs étudié, comme illustré sur la figure 2.4. Lorsque nous étudions les variations d'un critère tel que notre *MRC*, nous cherchons une augmentation conjointe des erreurs avec les écarts de conditions entre calage et contrôle, tandis qu'un niveau minimal serait atteint pour des conditions similaires. Dans le cas où une corrélation existerait, nous aurions donc plutôt une forme de nuage de point du type $\vee$ ou $\cup$, telle qu'on peut observer sur la figure 2.3. Pour notre calcul de critère de *Spearman*, nous découpons nos graphiques en deux parties verticales, différenciées par le signe de variation des conditions (*ex.* $P_D > P_R$ d'une part, et $P_D < P_R$ d'autre part), puis nous calculons un critère de *Spearman* sur chacune d'elles. Nous rassemblons ensuite ces valeurs en les moyennant après avoir changé le signe de l'une d'elle. Cette méthode de calcul particulière sera identifiée par la notation r'_s dans le manuscrit. Si la qualité des simulations est caractérisée par un critère de performances (*ex.* KGE, ρ_{kge}), on recherchera des valeurs de r'_s négatives, vérifiant la relation $|\triangle\ caract._{D\rightarrow R}| \nearrow \Rightarrow \triangle perfs \searrow$. Si on étudie par contre un critère d'erreurs (*ex.* $RMSE$, MRC), on recherchera des r'_s négatives,

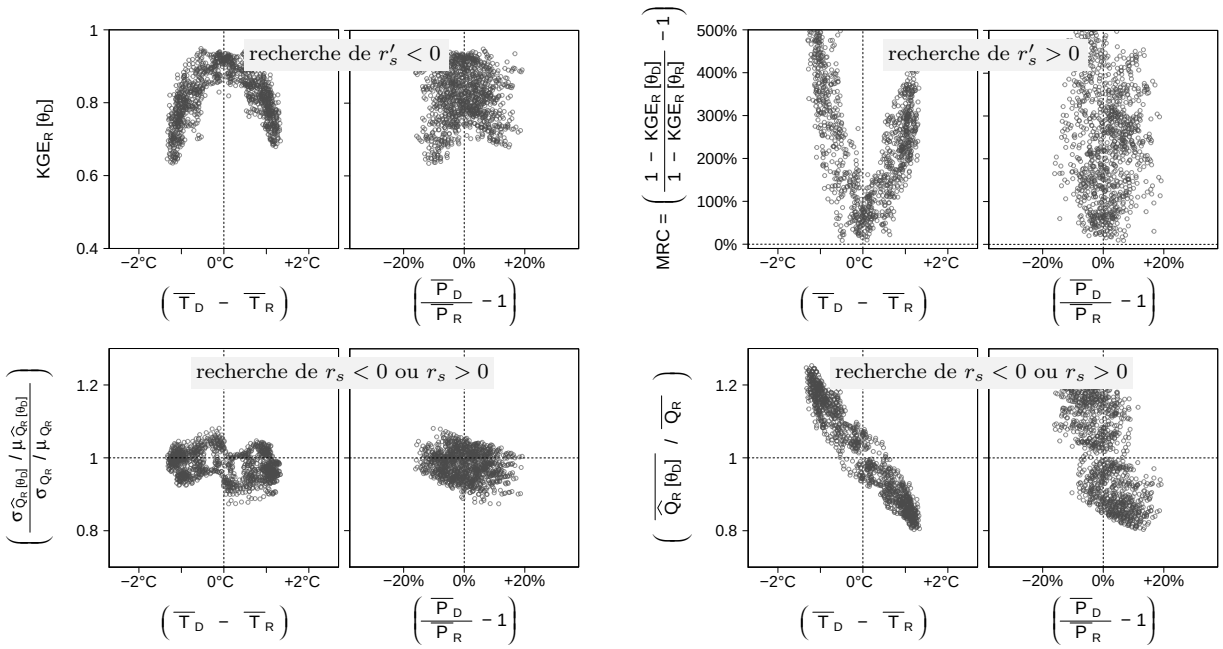


Figure 2.4.: Exemple de nuages de points utilisés pour étudier les liens potentiels entre erreurs de simulation et conditions de transfert des paramètres

vérifiant $|\triangle\ caract._{D \to R}| \nearrow \Rightarrow \triangle erreurs \nearrow$. Pour certains critères d'erreurs, il sera toutefois possible de calculer directement un critère de *Spearman* sur le nuage entier de points, (le critère sera alors noté r_s). C'est notamment ce que nous ferons pour les composantes α_{kge} et β_{kge} qui traduisent les rapports de variance et de biais. Quelques exemples de nuages de points sont fournis à la figure 2.4 et illustrent le type de situations sur lesquelles nous calculerons des critères de *Spearman*.

2.2. Construction d'une base de données

2.2.1. Critères de sélection de bassins versants

Le choix des bassins versants étudiés est une étape importante en modélisation hydrologique. Il peut être motivé par un intérêt particulier pour une rivière ou une zone géographique spécifique (*ex.* crues de la Loire, impacts d'une sécheresse sur une région, etc.). Ce choix peut aussi être uniquement conditionné par une problématique donnée (*ex.* évolution du couvert forestier et conséquences sur le comportement hydrologique, impact de la spatialisation des pluies sur les performances des modèles globaux, voies d'intégration des échanges souterrains dans les modèles pluie-débit, etc.). Dans ce second cas, qui correspond à notre situation, les bassins sont sélectionnés en fonction de l'intérêt qu'ils représentent face à cette problématique et bien évidemment de la disponibilité des données requises pour les tests.

La méthodologie décrite précédemment est construite autour de nombreux tests de calage-validation entre diverses périodes historiques de chaque bassin. Son objectif étant d'évaluer la *capacité d'extrapolation climatique* des paramètres de modèles conceptuels, cette méthodologie nous conduit à rechercher des bassins versants ayant subi des changements prononcés de conditions climatiques et pour lesquels de longues séries sont disponibles. De plus, nos travaux se situent dans le cadre des études menées par *EDF* pour anticiper l'évolution de la disponibilité de la ressource en eau sous l'effet des changements de climat (principalement pour la production hydroélectrique mais aussi pour le refroidissement des centrales thermiques). Les bassins situés en zone de montagne seront donc préférés aux grands ensembles de plaine durant la constitution de notre échantillon. Ces bassins présentent aussi l'avantage d'être moins influencés par les activités humaines ce qui devrait faciliter nos analyses de robustesse des modèles puisque nous pourrons négliger ces impacts.

Deux bases de bassins versants ont été constituées. Un premier échantillon se compose de bassins australiens, où la forte variabilité climatique au cours des dernières décennies nous permet des tests de transfert entre des conditions très marquées. Ces bassins sont répartis le long de la *Great Dividing Range* (Alpes australiennes). La majorité de cet échantillon est située dans les états fédéraux de *New-South-Wales* et *Victoria*, avec un débordement sur les états du *Queensland* (pour les bassins les plus au nord) et de *South-Australia* (pour les bassins les plus à l'ouest). Un second échantillon se compose de bassins français situés sur une large zone couvrant les Alpes, le Massif-Central et les Pyrénées. Leur contexte géographique est présenté à la figure 2.5.

2.2.2. Données utilisées

Échantillon de bassins australiens

Les données des bassins versants australiens ont été obtenues dans le cadre d'une collaboration avec le *CSIRO Land&Water* (Canberra, Australie), au sein duquel j'ai été accueilli de mars à juillet 2011. La variabilité du climat australien permet d'obtenir des situations de forts contrastes climatiques entre différentes périodes sur un même bassin, or le *CSIRO* dispose de séries continues de pluie, température, évapotranspiration potentielle (*ETP*) et débits sur la période pour 1974–2006 pour 228 bassins versants non influencés (parmi lesquels 204 seront retenus, *cf.* section 2.2.2). Cette base est donc extrêmement intéressante au regard des tests de transfert des paramètres que nous souhaitions mener. Elle fut l'une des motivations principales de a collaboration avec le *CSIRO Land&Water*, de pair avec l'intérêt que le *CSIRO* porte à notre problématique comme l'atteste leurs travaux récents (*Vaze et al.*, 2010; *Teng et al.*, 2012).

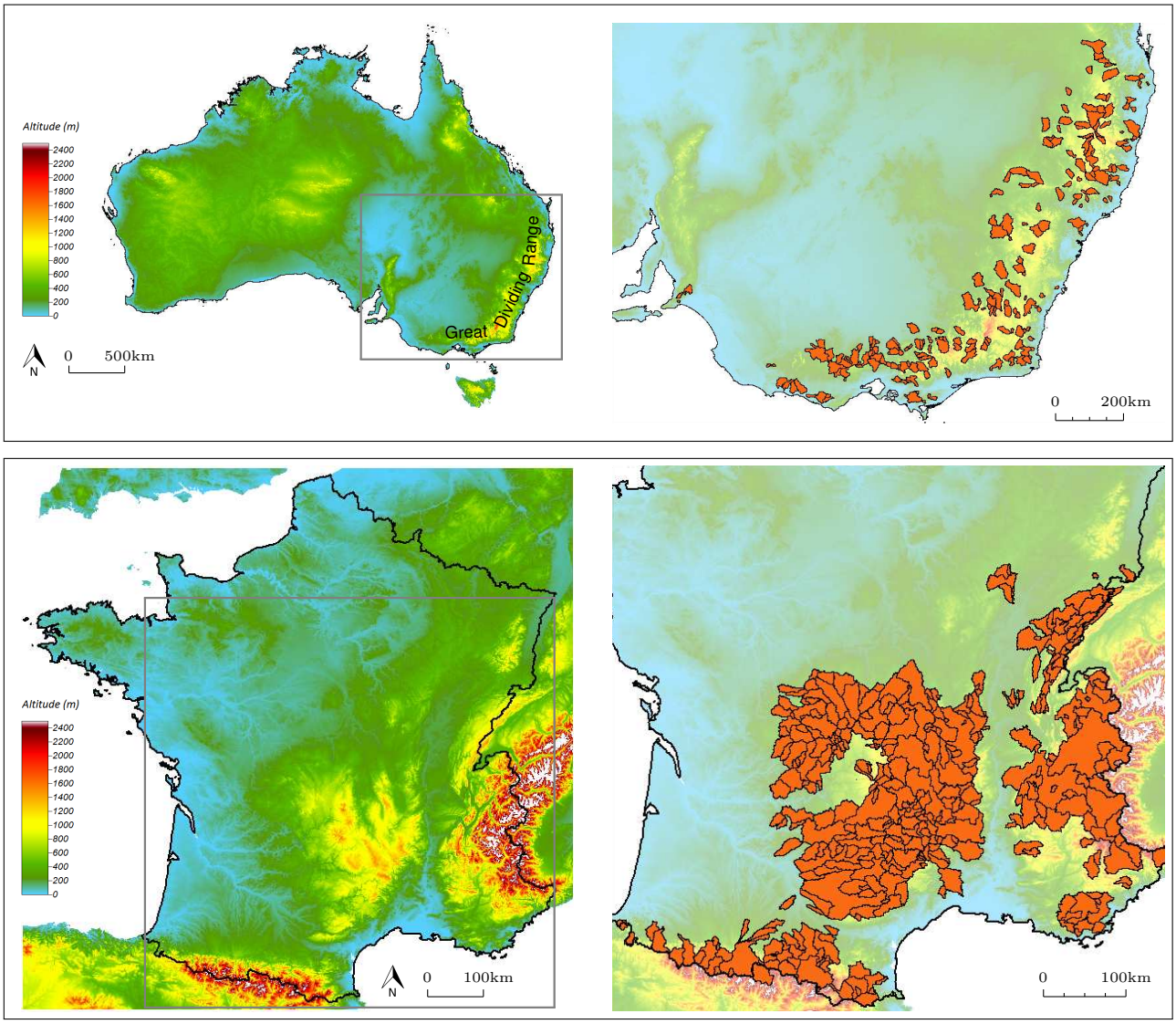


Figure 2.5.: Échantillons de bassins versants français (365) et australiens (204) sélectionnés pour les travaux de thèse

Cet échantillon de bassins australiens est décrit en détail par *Vaze et al.* (2011). Les séries de pluie et d'*ETP* sont issues de la base *SILO Data Drill* (www.longpaddock.qld.gov.au/silo/), qui fournit des données journalières sur une grille ayant une résolution de 0.05° (cellules d'environ 5×5 km) couvrant l'Australie. Ces estimations sont interpolées à partir de mesures ponctuelles de l'*Australian Bureau Of Meteorology*, les valeurs d'*ETP* étant calculées d'après la formulation de *Morton* (1983). Les observations de débits journaliers proviennent des différentes agences étatiques et gouvernementales australiennes, qui assurent également un contrôle de leur qualité. Ces bassins versants ont été sélectionnés pour être non influencés : ils ne comportent pas de retenue notable et ne sont pas sujets aux importants programmes d'irrigation qui peuvent exister en Australie. Enfin, les forçages de la base *SILO Data Drill* sont moyennés pour constituer des séries journalières de pluie et d'*ETP* de bassin. Celles-ci furent notamment utilisées dans divers travaux de recherche australiens sur la ressource en eau et ses évolutions face au changement climatique (*Chiew et al.*, 2008, 2009; *Vaze et Teng*, 2011; *CSIRO*, 2012).

Échantillon de bassins français

L'échantillon de bassins versants français a été constitué par croisement de la base de forçages *SPAZM* et des stations hydrométriques disponibles sur la banque *HYDRO*. La base *SPAZM* est issue du travail mené par *F. Gottardi* durant sa thèse, dont l'objectif était de produire une réanalyse améliorée des mesures de précipitations (liquides et solides) et de températures de l'air en zone de montagne (*Gottardi*, 2009). Cette base repose notamment sur l'établissement de relations linéaires entre les précipitations et l'altitude, qui varient spatialement et temporellement (selon le type de conditions atmosphériques). Ces réanalyses sont conduites sur trois grilles géographiques couvrant les Alpes, le Massif-Central et les Pyrénées. Sur chaque grille, des valeurs de précipitation totale, de température minimale et de température maximale au pas de temps journalier sont disponibles sur des pixels de 1 km². Les bassins versants de la banque *HYDRO* situés dans les zones couvertes par *SPAZM* furent sélectionnés selon des critères de longueur et de qualité des séries de débits (minimum 30 années d'observations, absence d'influences anthropique ou d'erreurs notables sur les débits).

L'impact des précipitations neigeuses sur les régimes et notre volonté d'utiliser le modèle *Cequeau* nous ont conduit à considérer parfois une approche spatialisée pour les bassins français. Les forçages issus de la base *SPAZM* ont donc été extraits sur des maillages préalablement construits (plutôt que systématiquement moyennés sur tout le bassin comme dans l'échantillon australien). Ces maillages dépendent à la fois des caractéristiques du bassin (*ex.* part solide des précipitations annuelles, amplitude des altitudes) et du modèle hydrologique considéré. Par exemple, *GR4J* est couplé en zone de montagne avec le module *CemaNeige*, selon un maillage en cinq bandes d'altitudes d'égale surface (préconisations de *Valéry*, 2010). Le modèle semi-distribué *Cequeau* est, quant à lui, utilisé sur un maillage dont la résolution dépend de l'hypsométrie du bassin (*c.-à-d.* discrétisation par altitude contrainte par un nombre et

une taille minimum de mailles, *cf.* section 2.3.5). En conséquence, on dispose pour chaque bassin d'une chronique journalière de débits observés et de $3\,N$ chroniques journalières de forçages : précipitation totale, température minimale et température maximale de l'air (N étant le nombre de mailles considéré, *ex.* $N = 5$ pour *GR4J-CemaNeige*). On mentionnera enfin que les calculs de l'évapotranspiration potentielle et la dissociation des précipitations liquides et solides peuvent, selon les cas, être une composante interne au modèle, sujette à optimisation, ou constitue une entrée indépendante dont les paramètres sont préalablement fixés. Ainsi, ces composantes sont intégrées et optimisées pour les modèles *Cequeau* et *Mordor* mais ont une formulation et des paramètres fixes pour les couples *GR4J-CemaNeige Mordor6-CemaNeige* et *SimHyd-CemaNeige.*

Un mot sur la critique des données

Les différentes données utilisées dans nos travaux sont toutes issues de bases ayant subi des analyses de qualité, menées par les différents organismes en charge de la mesure de la donnée ou de son intégration au sein d'une banque de données. Néanmoins, avant d'alimenter des modèles et se lancer dans de multiples tests, il semble nécessaire de jeter un œil aux séries sur lesquelles nous souhaitons construire nos raisonnements.

L'analyse de la qualité des séries de débits est une étape importante, notamment parce que les informations sur les conditions de mesure des hauteurs et la validité des courbes de tarage restent souvent peu disponibles et/ou sont négligées dans les fichiers de chroniques utilisés par le modélisateur. En reprenant toutes les séries de débit des bassins présentant un intérêt pour notre étude, j'ai mené un contrôle de qualité simple, basé sur la visualisation de courbes de doubles-cumuls de débits entre chaque bassin versant et ses plus proches voisins. Cette méthode de détection repose sur le postulat que deux séries voisines de bonne qualité ont des cumuls liés par une relation moyenne de proportionnalité ne présentant pas de cassure. En pratique, on construit une courbe de double-cumuls en représentant les débits cumulés du bassin évalué en fonction des débits cumulés d'un bassin voisin. On analyse ensuite les résidus entre cette courbe et la droite reliant ses deux extrémités (qui sert de référence de la relation de proportionnalité). Si les deux séries comparées sont de qualité adéquate et que l'une ne présente pas de dérive, alors le cumul de ces résidus doit osciller autour de zéro dans des limites *raisonnables*. Si par contre, l'une des séries est erronée, ce cumul de résidus suivra des variations disproportionnées ou des ruptures de pente soudaines et prononcées (*cf.* figure 2.6). Cet exercice est répété ensuite avec d'autres bassins voisins pour s'assurer que la chronique en question est réellement en cause dans les problèmes détectés (*c.-à-d.* que le problème ne provient pas de la série de référence).

Les analyses par doubles-cumuls débit-débit permettent un repérage rapide des anomalies causées par exemple par des modifications de la courbe de tarage, des interpolations linéaires

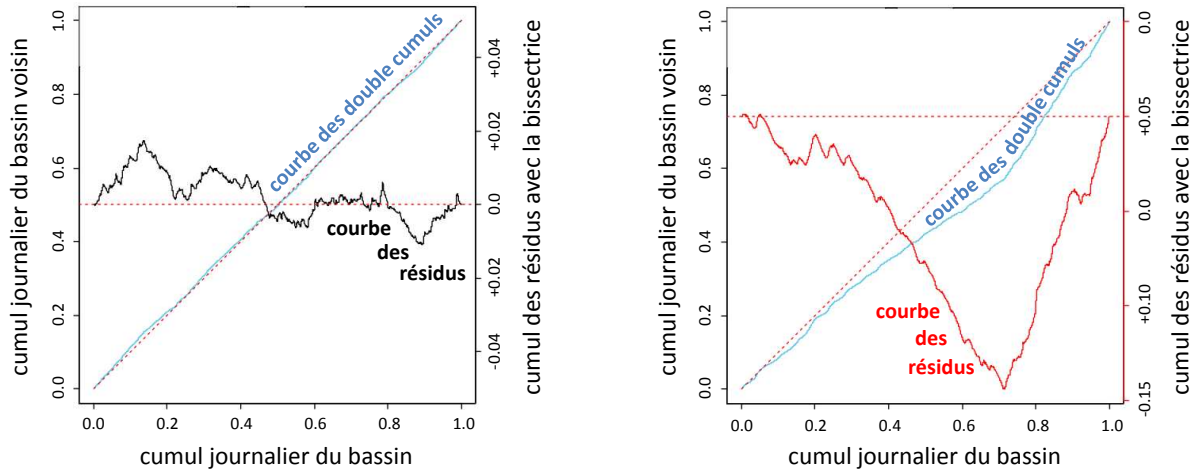


FIGURE 2.6.: Illustration d'analyses de qualité par l'approche doubles-cumuls débit-débit
Série correcte à gauche (*la Grosne à Jalogny*) et érronée à droite (*le Borne à Saint-Jean-de-Sixt*)

notables entre observations de débit distantes dans le temps, ou la présence de données incorrectes. Dans notre cas, nous nous sommes restreints à des détections visuelles des anomalies, sans passer par des approches plus systématiques telles que celle de *Bois* (1987) ou de *Paquet* (2005). Cette procédure a été appliquée à tous les bassins français et australiens *candidats* pour intégrer nos échantillons de test. Dans chaque cas, les débits d'une station sont comparés à ceux des trois stations les plus proches pour assurer l'identification de la série défaillante lorsque des erreurs sont détectées. Toutes les stations jugées douteuses à l'issue de ce test furent simplement retirées du *panier* initial de bassins [4]. Seuls quelques bassins furent retirés de la base (déjà critiquée) du *CSIRO*. En France, où nous avions jusqu'ici pour seul filtre la longueur des séries et la localisation à l'intérieur de la zone *SPAZM*, plus d'un tiers des bassins qui auraient pu être intégrés à notre base furent écartés par cette phase de critique. Précisons toutefois que toutes les séries écartées ne présentaient pas des ruptures aussi nettes que l'exemple présenté sur la figure 2.6. De plus, les bassins *candidats* étant très nombreux, nous pouvions nous permettre d'écarter une chronique tout entière au moindre doute, tandis que le retrait de quelques pas de temps aurait parfois suffi à rétablir la qualité d'une série. Des détails complémentaires sur nos analyses et les types d'erreurs détectés sont présentés en annexe C.1.

En termes de gestion des pas de temps manquants, nous avons fait le choix de fixer un seuil de tolérance à 10% de lacunes par an, au-delà duquel toute l'année hydrologique concernée est retirée des calculs de critères. On précise que les données de débits autant que de climat sont alors retirées pour maintenir une correspondance aussi étroite que possible entre les ca-

4. Dans certains cas, les erreurs identifiées dans les débits auraient pu être déclarées lacunaires et la chronique conservée. Cependant, nos bases de bassins étaient suffisamment conséquentes pour assurer de larges échantillons malgré le retrait des séries d'observations présentant des données erronées.

ractérisations climatiques de nos sous-périodes et les analyses des débits. Ce seuil de tolérance est ensuite appliqué aux sous-périodes constituées, pour ne conserver que celles ayant moins de 10% de données lacunaires au total. Combiné aux analyses par doubles-cumuls débit-débit, ce filtrage des données lacunaires est la cause du retrait de certains bassins des échantillons initiaux (notamment le passage de 228 à 204 bassins australiens [5]).

Les données de forçages (pluie, température) que nous utilisons ne sont pas brutes mais résultent de divers procédés d'interpolation et d'extrapolation puis d'agrégation afin d'obtenir des séries au pas d'espace de nos modèles (moyenne sur une maille, une bande d'altitude ou un bassin). Elles ont subi divers contrôles de qualité par les producteurs de données, puis par les créateurs de ces bases (*Queensland Climate Change Centre of Excellence* pour la base *SILO Data Drill* ; *EDF-DTG* et *Gottardi (2009)* pour la base *SPAZM*). Précisons enfin que la base *SPAZM* représente à ce jour la meilleure source disponible de forçages sur notre zone d'étude en France (bassins de montagne), mais que nous procéderons également à différentes analyses pour apprécier le réalisme de ces données qui alimentent nos modèles et évaluer l'impact du choix de cette base sur nos interprétations (*cf.* sections 5.3.2 et 6.6).

2.2.3. Caractéristiques des bassins retenus

Caractéristiques générales

Les échantillons de bassins versants que nous avons constitués présentent une grande diversité de situations. Cette diversité est illustrée par la figure 2.7 qui fournit les distributions des superficies et altitudes médianes de bassin, des volumes de précipitations et évapotranspirations potentielles annuelles moyennes de bassin et des modules (lames d'eau annuelles moyennes) [6]. On notera que les *ETP* de la base australienne sont estimées par la formule de *Morton* (1983) et que celles de la base française sont obtenues avec la formule d'*Oudin et al.* (2005).

Une large gamme de tailles de bassin est représentée tant pour l'échantillon français qu'australien, avec des bassins allant de quelques dizaines de km² à 2 000 km² pour le plus grand bassin australien (*Wimmera River à Glenorchy, Victoria*) et plus de 10 000 km² pour les plus grands bassins français (*Allier à Moulins* et *Isère à Beaumont-Monteux*). Tous les bassins considérés restent cependant proches des massifs montagneux (*cf.* distribution des altitudes médianes), certains d'entre eux recevant une part de leurs précipitations sous forme de neige. Ils demeurent peu nombreux en Australie et ces apports solides n'influencent pas le régime des écoulements dans une très grande majorité de cas. En France, l'échantillon comporte à

5. Ce filtrage était moins restrictif à l'époque de l'article *Coron et al.* (2012), où nous avions retenu les 216 bassins disposant de moins de 10% de lacunes au total (pas de contrainte annuelle).

6. La fraction solide des précipitations est estimée à l'aide d'une formulation simple de dissociation pluie/neige basée sur un seuil de température journalière de l'air.

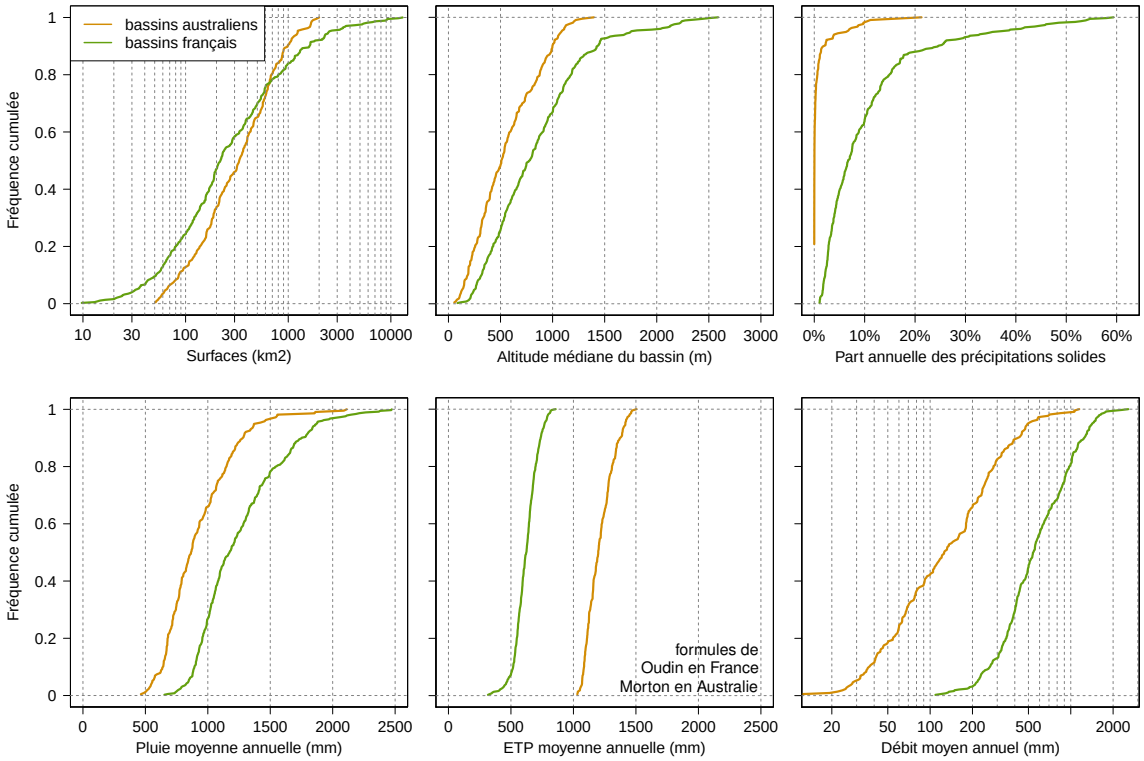


FIGURE 2.7.: Résumé des caractéristiques des bassins étudiés

la fois des bassins versants de basse et moyenne altitude recevant une faible part de neige, et des bassins de haute montagne (dans les Alpes et les Pyrénées) où les précipitations solides conditionnent fortement les écoulements.

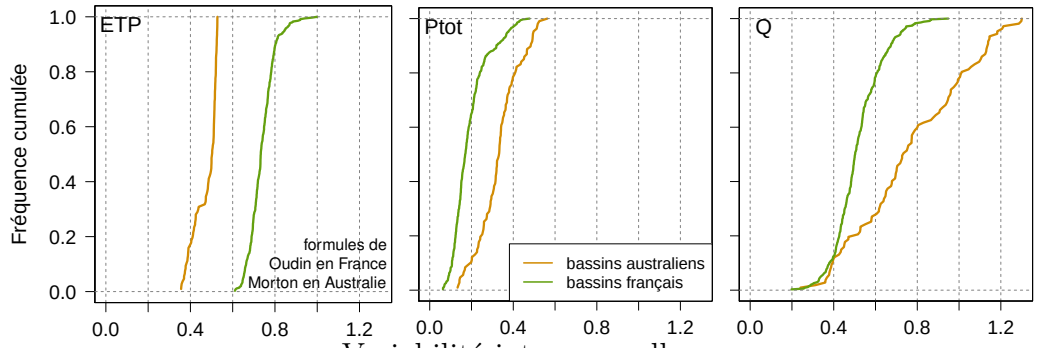
Les distributions de précipitations annuelles ont des formes comparables mais sont décalées d'environ 200 mm. Au-delà d'une différence climatique, ce décalage provient sans doute des effets altitudinaux (*c.-à-d.* pluies en moyenne plus abondantes en altitude, notre échantillon français étant globalement plus haut que l'échantillon australien). Les différences importantes constatées sur les évapotranspirations potentielles reflètent en revanche un écart notable de conditions climatiques entre nos deux échantillons, que l'on retrouve sur les lames d'eau écoulées. Ainsi, notre base de données indique qu'en moyenne environ 50% des précipitations sont converties en débit pour l'échantillon français et seulement 17% sur l'échantillon australien.

Variabilité

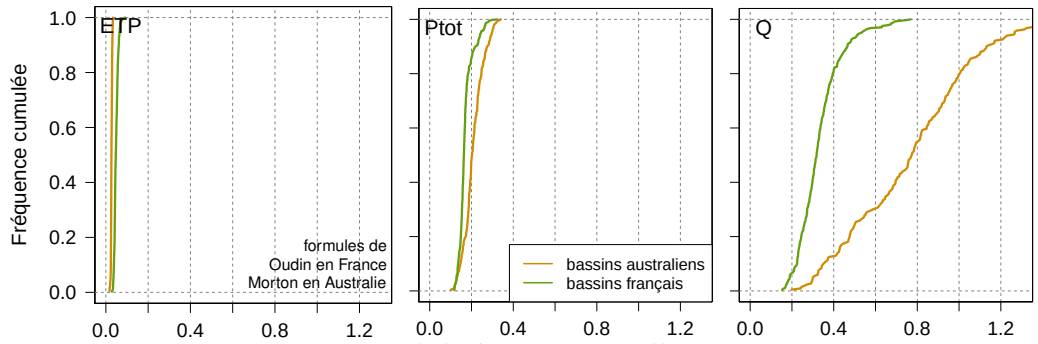
Nous nous intéressons enfin à la variabilité des conditions climatiques sur nos échantillons de bassins, que nous analysons à travers les différents graphiques de la figure 2.8. Celle-ci regroupe les distributions des coefficients de variation (cv = écart-type / moyenne) obtenus sur notre échantillon pour les valeurs mensuelles et annuelles d'ETP, de précipitation totale et de débit ainsi qu'une visualisation des variations observées sur une fenêtre glissante de 10 ans.

Les *ETPs* de notre base australienne suivent globalement un cycle moins prononcé que sur nos bassins français (*c.-à-d.* variabilité intra-annuelle plus faible). Cet écart est sans doute causé par des amplitudes plus grandes sur ces derniers entre les mois d'été et les mois d'hiver, où les températures en montagne diminuent fortement, faisant chuter l'*ETP*. À l'inverse, la variabilité mensuelle des précipitations est plus marquée sur les bassins australiens que français et se retrouve sur les écoulements. La combinaison de ces précipitations temporellement variables et d'*ETP* élevées engendre des situations de très fortes variations intra-annuelles de débits en Australie, certains bassins pouvant aller jusqu'à l'intermittence. Nous noterons également une variabilité assez importante sur certains bassins français, en dépit de précipitations mensuelles plutôt homogènes. Cette situation résulte du cycle saisonnier très marqué des températures qui conditionne les écoulements en haute montagne (via l'*ETP* et surtout l'accumulation et fonte du manteau neigeux*, cf.* section 1.2.1).

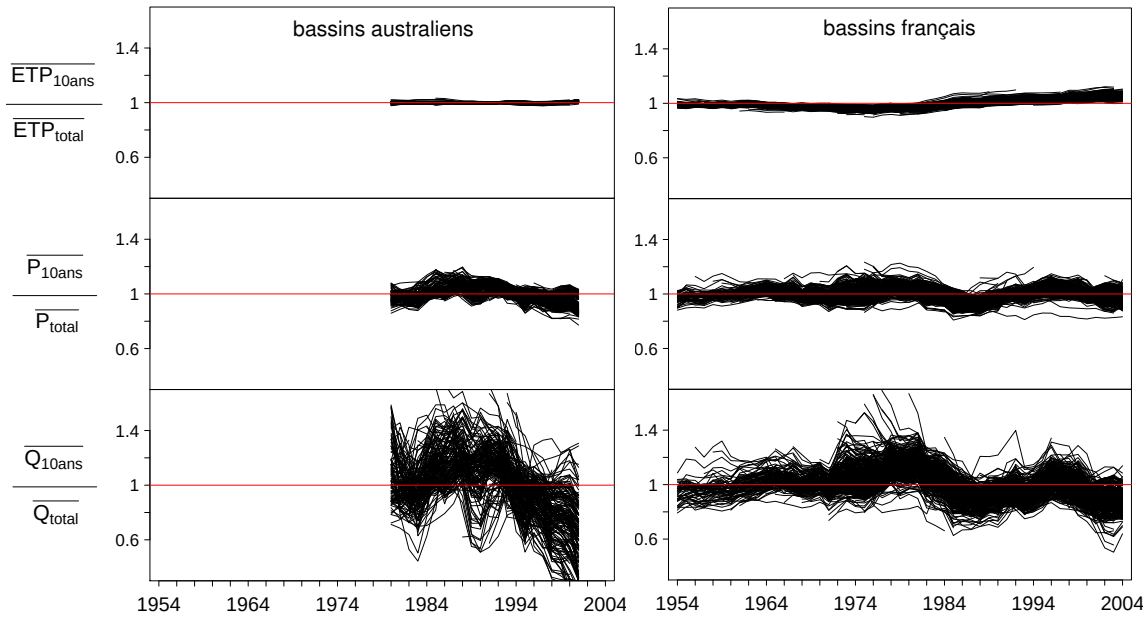
En passant à l'échelle interannuelle, nous retrouvons logiquement une très faible variabilité de l'*ETP* et nous observons une variabilité modérée des précipitations, cependant plus forte sur les bassins australiens que français. Ici encore, cette variabilité des forçages climatiques est amplifiée dans la transformation en écoulement avec des écarts de débits annuels pouvant être conséquents en France et plus encore en Australie. Sur la variabilité long terme enfin, nous constatons que les moyennes sur 10 ans d'*ETP* varient de quelques pourcents, celles de précipitations d'environ 15%, et que cette variabilité s'amplifie sur les débits avec des écarts de 30% en France et 50% en Australie ! Des signaux de basse fréquence émergent d'ailleurs de ces graphiques avec une alternance entre des périodes humides (*ex.* 1975-1985 en France et 1985-1995 en Australie) et des périodes sèches, avec pour les deux échantillons des valeurs inférieures à la moyenne pour les années les plus récentes.



(distribution des coefficients de variation entre les moyennes mensuelles pour chaque bassin)



(distribution des coefficients de variation entre les moyennes annuelles pour chaque bassin)



(Chaque bassin est représenté par une courbe illustrant les moyennes glissantes, normalisées par la moyenne long-terme et affectées à l'année centrale de chaque fenêtre de 10 ans)

FIGURE 2.8.: Aperçu de la variabilité des *ETPs*, précipitations et écoulements sur les bassins étudiés, à différentes échelles temporelles

2.3. Modèles hydrologiques considérés

La comparaison des résultats sur une diversité de cas constitue un point central de notre méthodologie de test. Dans le même esprit, nous souhaitons adopter une approche multi-modèles pour comparer nos résultats pour diverses stratégies de modélisation de la relation précipitations-débit, *c.-à-d.* plusieurs modèles de structures différentes.

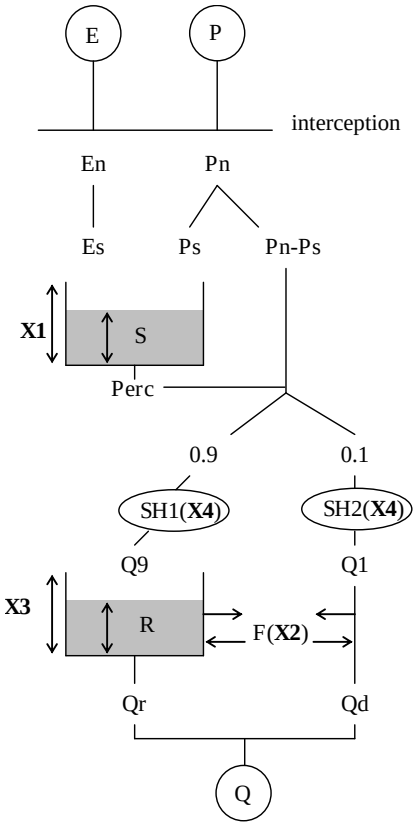
Cinq modèles hydrologiques conceptuels sont utilisés au fil de nos tests de transfert de paramètres dans un contexte de climat changeant : *GR4J* (*Perrin et al.*, 2003), *Mordor6* (*Mathevet*, 2005), *SimHyd* (*Tan et al.*, 2005), *Mordor* (*Garçon*, 1996) et *Cequeau (Charbonneau et al., 1977).* Ce sont tous des modèles conceptuels fonctionnant au pas de temps journalier mais ils diffèrent par leur structure et leur discrétisation spatiale. Les trois premiers sont globaux et comportent peu de paramètres libres. *Mordor* est également global mais comporte un nombre beaucoup plus élevé de paramètres ajustables. *Cequeau* comporte également de nombreux paramètres et est le seul modèle semi-distribué que nous testerons. D'autre part, *Mordor* et *Cequeau* disposent de leur propre module neige tandis que les trois autres modèles nécessitent un couplage avec un module neige dédié lorsque cela s'avère nécessaire. Le module *CemaNeige* de (*Valéry*, 2010) est utilisé pour réaliser ce couplage sur l'échantillon de bassins français.

2.3.1. Modèle *GR4J*

GR4J est un modèle global journalier développé par une approche empirique. Décrit en détail par *Perrin et al.* (2003), il doit être alimenté par des moyennes de bassin de précipitation et d'évapotranspiration potentielle. La figure 2.9 présente sa structure et les équations associées. *GR4J* comporte quatre paramètres à caler (*cf.* tableau 2.1). Leur optimisation est réalisée via un parcours grossier de l'espace des paramètres (préfiltrage) pour identifier la zone probable de convergence, puis une recherche locale. Pour modéliser les écoulements sur des bassins influencés par les précipitations solides, *GR4J* peut être couplé au module *CemaNeige*, qui est décrit dans la section 2.3.4.

Tableau 2.1.: Liste des paramètres du modèle *GR4J*

notation	description	unité
$\theta_{X1(PROD)}$	capacité du réservoir sol	mm
$\theta_{X2(CES)}$	coefficient d'échange souterrain	-
$\theta_{X3(ROUT)}$	capacité du réservoir de routage	mm
$\theta_{X4(THU)}$	constante de temps de l'hydrogramme unitaire	j



$$\text{si } P \geq E,\ Pn = P - E,\quad En = 0,\ \text{si } P < E,\ En = E - P,\quad Pn = 0$$

$$Ps = \frac{Pn.\left(1-\left(\frac{S}{X1}\right)^2\right)}{1+\frac{Pn}{X1}.\left(1+\frac{S}{X1}\right)} \qquad Es = \frac{En.\frac{S}{X1}\left(2-\frac{S}{X1}\right)}{1+\frac{En}{X1}.\left(2-\frac{S}{X1}\right)}$$

$$Perc = S-\left(S^{-4}+\left(\frac{9}{4}.X1\right)^{-4}\right)^{-\frac{1}{4}} \qquad S = S - Perc$$

$$0 \leq j \leq X4,\ SH1(j) = \left(\frac{j}{X4}\right)^{\frac{5}{2}}$$

$$j > X4,\quad SH1(j) = 1$$

$$0 \leq j \leq X4,\ SH2(j) = \frac{1}{2}\left(\frac{j}{X4}\right)^{\frac{5}{2}}$$

$$X4 < j \leq 2.X4, SH2(j) = 1-\frac{1}{2}\left(2-\frac{j}{X4}\right)^{\frac{5}{2}}$$

$$j > 2.X4,\quad SH2(j) = 1$$

$$F = X2.\left(\frac{R}{X3}\right)^{\frac{7}{2}}$$

$$Qd = \max(0, Q1 + F)$$

$$R = \max(\varepsilon, R + Q9 + F)$$

$$Qr = R-\left(R^{-4}+X3^{-4}\right)^{-\frac{1}{4}} \qquad R = R - Qr$$

$$Q = Qr + Qd$$

FIGURE 2.9.: Structure et équations du modèle *GR4J*

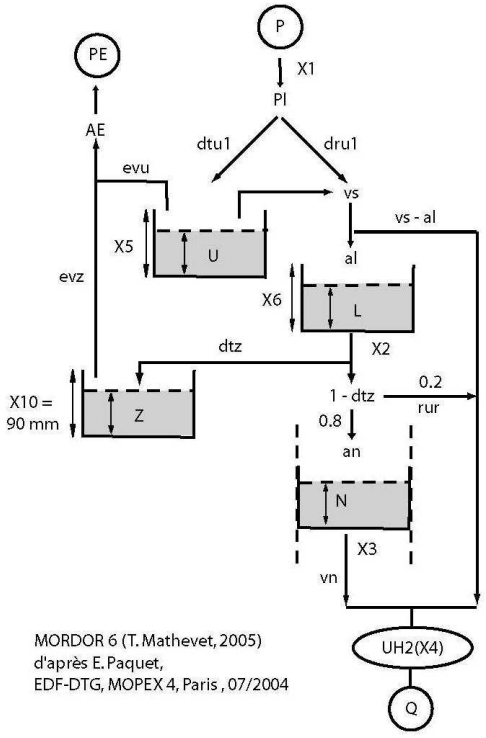
2.3.2. Modèles *Mordor* et *Mordor6*

Mordor est un modèle conceptuel global au pas de temps journalier développé par *Garçon* (1996) qui comporte 23 paramètres ajustables. À la différence de *GR4J*, *Mordor6*, *SimHyd* et *Cequeau*, ce modèle n'apparaîtra que ponctuellement dans nos travaux puisque nous ne disposons de calages et simulations que sur quelques bassins versants. Sa structure n'est pas détaillée ici car elle se rapproche fortement de celle du modèle *Mordor6*, qui sera, quant à lui, largement utilisé pour nos analyses.

Mordor6 est une simplification de *Mordor* réalisée par *Mathevet* (2005). Ce modèle comporte six paramètres à caler (*cf.* tableau 2.2). Le passage de 23 à 6 paramètres a été réalisé en fixant certains paramètres à des valeurs par défaut, en retirant le calcul de l'*ETP* du modèle (celle-ci devenant une entrée, comme pour *GR4J*), ainsi qu'en supprimant le module neige étagé. La figure 2.10 présente la structure et les équations de *Mordor6*. Ce modèle doit être alimenté par des moyennes de bassin de précipitation et d'évapotranspiration potentielle. Si le modèle *Mordor* dispose de son propre module neige (cf. *Paquet*, 2004), ce n'est pas le cas de *Mordor6* qui sera si nécessaire couplé à *CemaNeige* (similairement à ce qui est fait pour *GR4J*).

Tableau 2.2.: Liste des paramètres du modèle *Mordor6*

notation	description	unité
$\theta_{X1(CPL)}$	coefficient multiplicateur des précipitations totales	-
$\theta_{X2(TRL)}$	constante de vidange du réservoir intermédiaire	j
$\theta_{X3(TRN)}$	constante de vidange du réservoir nappe	j
$\theta_{X4(THU)}$	constante de temps de l'hydrogramme unitaire	j
$\theta_{X5(RU)}$	capacité du réservoir de surface	mm
$\theta_{X6(RL)}$	capacité du réservoir intermédiaire	mm



Correction sur la pluie

$Pl = P * X1$

Répartition de la pluie selon U

$dtr1 = Pl.U / X5 \quad dtu1 = Pl - dtr1$

Évolution du réservoir U

$vs = dtr1 + \max(0, U - X5)$

$U = \min(U + dtu1, X5)$

$evu = \min(X5, E.U / X5)$

$U = U - evu$

Évolution du réservoir L

$al = \min(X6 - L, vs.(1 - L / X6))$

$L = L + al \quad vl = L / X2 \quad L = L - vl$

Évolution du réservoir Z

$dtz = vl.(1 - Z / 90) \quad rur = 0.2.vl.(Z / 90)$

$an = 0.8.vl.(1 - Z / 90)$

$Z = Z + dtz$

$evz = \min(Z, (E - evu).Z / 90)$

$Z = \min(90, Z - evz)$

Évolution du réservoir N

$N = N + an$

$vn = \min(N, (N / X3)^3)$

$N = N - vn$

Somme et routage des différentes contributions

$Qt = vs - al + rur + an$

Hydrogramme unitaire de GR4J (Perrin, 2000)

$0 \le j \le X4,$

$SH2(j) = \frac{1}{2}\left(\frac{j}{X4}\right)^{\frac{5}{4}}$

$X4 < j \le 2.X4,$

$SH2(j) = 1 - \frac{1}{2}\left(2 - \frac{j}{X4}\right)^{\frac{5}{4}}$

$j > 2.X4,$

$SH2(j) = 1$

$Q = SH2(X4, Qt)$

Figure 2.10.: Structure et équations du modèle *Mordor6*

2.3.3. Modèle *SimHyd*

SimHyd est un modèle conceptuel global et journalier, initialement développé par *Chiew et al.* (2002) par simplification du modèle *Hydrolog*. Des simplifications complémentaires ont été apportées par *Tan et al.* (2005). Dans la version utilisée ici, les deux paramètres contrôlant l'infiltration sont fixés à des valeurs moyennes et un module de routage (basé sur un réservoir et un hydrogramme unitaire) a été ajouté. Cette version comporte sept paramètres à caler (*cf.* tableau 2.3). La procédure de routage utilisée dans la version de *SimHyd* que nous utiliserons provient des travaux de *Perrin* (2000). La structure et les équations de la fonction de production de *SimHyd* sont présentées sur la figure 2.11. Ce modèle doit être alimenté par des moyennes de bassin de précipitation et d'évapotranspiration potentielle et il peut être couplé à *CemaNeige* pour modéliser les écoulements sur des bassins influencés par les précipitations solides.

TABLEAU 2.3.: Liste des paramètres du modèle *SimHyd*

notation	description	unité
$\theta_{X1(INSC)}$	capacité du réservoir d'interception	mm
$\theta_{X2(SMSC)}$	capacité du réservoir sol	mm
$\theta_{X3(K)}$	constante de vidange du réservoir nappe	j
$\theta_{X4(THU)}$	constante de la convolution	-
$\theta_{X5(ROUT)}$	constante de vidange du réservoir de routage	j
$\theta_{X6(SUB)}$	coefficient *interflow*	mm
$\theta_{X7(CRAK)}$	coefficient de recharge	j

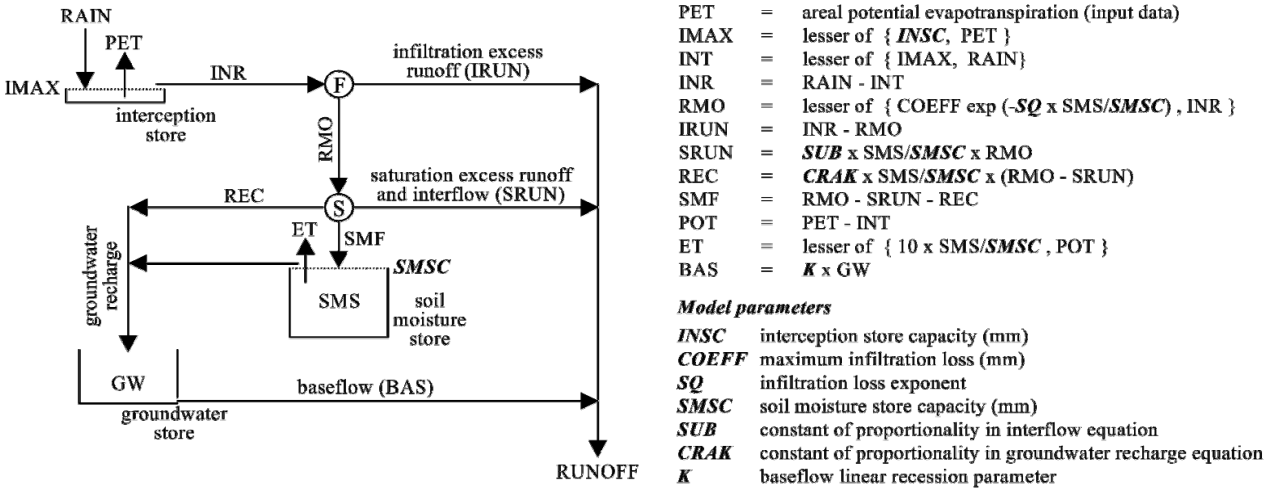
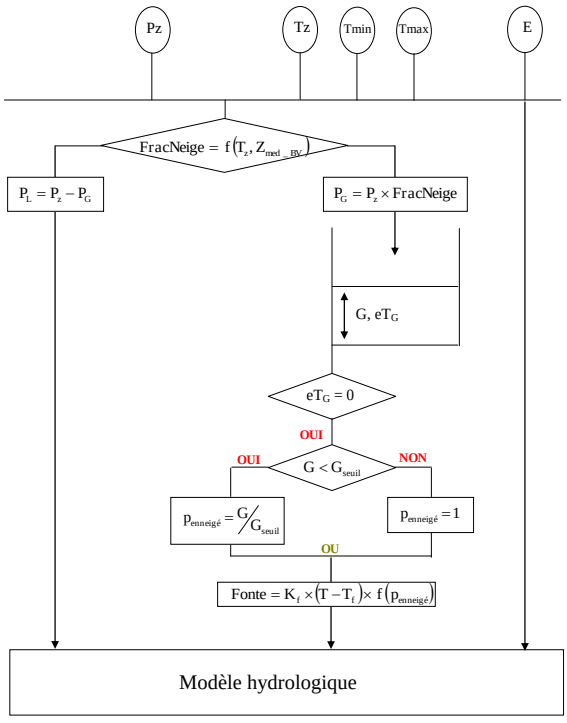


FIGURE 2.11.: Structure et équations de la partie production du modèle *SimHyd*

2.3.4. Module *CemaNeige* (complément pour *GR4J*, *Mordor6* et *SimHyd*)

CemaNeige est un module neige de type degré-jour pouvant *se greffer* en amont de n'importe quel modèle hydrologique. Il a été développé par *Valéry* (2010) qui le testa sur plusieurs centaines de bassins en France, en Suède et au Canada. Il comporte deux paramètres à caler : un facteur de fonte degré-jour (K_f) et un coefficient de pondération de l'état thermique du manteau (C_{TG}). Lorsqu'il est couplé avec un modèle global, il s'utilise sur des *bandes d'altitude* de surface égales. Dans ce cas, les sorties sont ensuite agglomérées pour former une lame d'eau de bassin qui somme les précipitations liquides et la lame de fonte. Cette lame d'eau alimente alors les modèles hydrologiques à la manière d'une pluie de bassin (dans notre cas, il alimentera ainsi *GR4J*, *Mordor6* et *SimHyd*). Compte tenu des résultats de *Valéry* (2010) sur un échantillon de bassins similaire au nôtre, une spatialisation des forçages selon cinq bandes d'altitude de surfaces égales a été utilisée. Dans les cas où la surface du bassin est inférieure à 25 km², une seule bande est considérée pour éviter les situations où très peu ou pas de pixels *SPAZM* étaient associés à l'une des bandes d'altitude.

Le schéma et les équations de fonctionnement commun à pour chaque bande d'altitude sont fournis sur la figure 2.12.



Extrapolation des données d'entrée

$Tz = T + \theta_{altitude} \times (Z_z - Z_{med_BV})$

$Pz = P \times \exp[\beta_{altitude} \times (Z_z - Z_{med_BV})]$

Calcul de la fraction solide

si $Z_{med_BV} < 1500\text{m}$, fonction type *Turcotte et al.* (2007)

si $(Tz\ max) \leq 0°C$, $FracNeige = 1$ si $(Tz\ max) \geq 0°C$, $FracNeige = 0$

sinon, $FracNeige = 1 - (Tz\ max) / ((Tz\ max) - (Tz\ min))$

si $Z_{med_BV} \geq 1500\text{m}$, fonction type *USACE* (1956)

si $(Tz\ moy) < -1°C$, $FracNeige = 1$ si $(Tz\ moy) > +3°C$, $FracNeige = 0$

sinon, $FracNeige = 1 - ((Tz\ moy) - (-1)) / ((+3) - (-1))$

$P_G = Pz \times FracNeige$ et $P_L = Pz - P_G$

Accumulation

$G = G + P_G$

$eT_G = C_{TG} \times eT_G + (1 - C_{TG}) \times Tz$

Calcul de la fonte potentielle

si $eT_G = 0$ et $Tz > T_f$, $F_{pot} = K_f \times (Tz - T_f)$

si $F_{pot} > G$, $F_{pot} = G$

Calcul pourcentage de la zone enneigée

si $G < G_{seuil}$, $p_{enneigé} = G/G_{seuil}$, $= 1$ sinon

Calcul de la fonte effective

$FONTE = [(1 - V_{min}) \times p_{enneigé} + V_{min}] \times F_{pot}$

Actualisation réservoir neige

$G = G - FONTE$

Calcul de la lame d'eau transmise au modèle hydrologique

$Lame_eau = P_L + FONTE$

Figure 2.12.: Structure et équations d'une bande d'altitude du module neige *CemaNeige*

2.3.5. Modèle *Cequeau*

Le modèle hydrologique *Cequeau* est un modèle hydrologique semi-distribué fonctionnant au pas de temps journalier. Il a été initialement développé par l'*INRS-Eau* (*Charbonneau et al.*, 1977; *Morin et al.*, 1981). Son développement a ensuite été poursuivi pour des besoins spécifiques à *EDF-R&D*. La version du modèle utilisée dans cette thèse est la version 5.1.1 d'*EDF R&D*, qui est présentée de façon détaillée dans la note de principe de *Le Moine et Monteil* (2012). Par souci de simplicité, cette version sera simplement appelée *Cequeau* dans ce manuscrit.

La figure 2.13 et le tableau 2.4 fournissent la structure de la partie de production de *Cequeau* (version *EDF R&D*) et la liste des paramètres ajustables. Ce modèle comporte 25 paramètres, éventuellement complétés par 9 paramètres de spatialisation des entrées. Les tests présentés dans cette thèse portent sur une configuration dans laquelle 19 de ces paramètres sont calés.

Parmi les principaux changements apportés par *EDF-R&D* à la version initiale de *Cequeau*, on peut mentionner : la modification du système de maillage pour correspondre à la topographie, la suppression du module *lacs et marécages* et l'ajout d'un module *glacier*, une réécriture des équations de vidange des réservoirs de production (sans changer leur structure), une réécriture complète du module de routage pour la propagation des débits à l'intérieur de la maille, puis jusqu'à l'exutoire du domaine d'étude (propagation désormais effectuée au moyen d'une formulation établie à partir de l'équation 1D de l'onde diffusante, solution de *Hayami (1951)* à célérité C et diffusivité D indépendantes du débit). Les entrées et la fonction de production

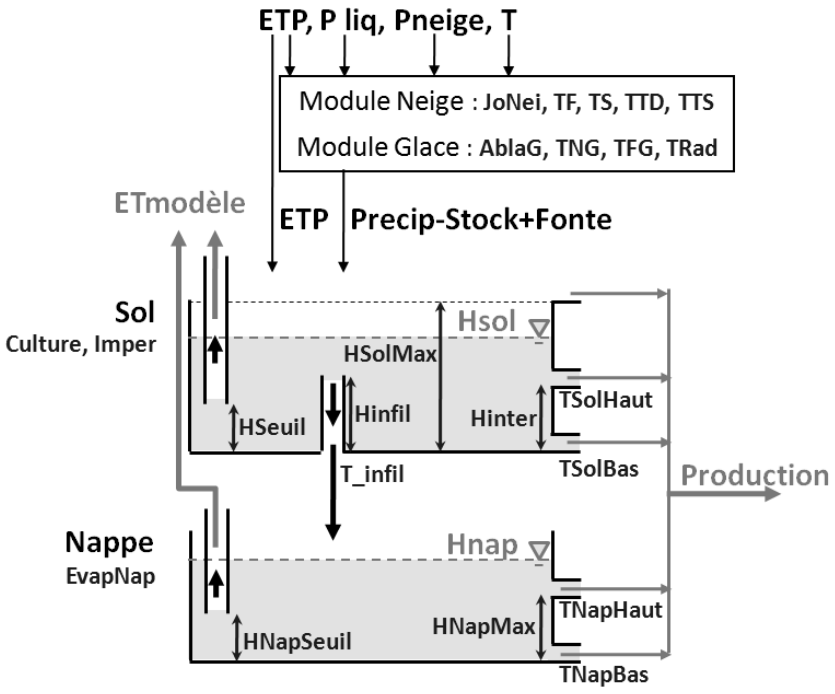


FIGURE 2.13.: Structure de la partie production du modèle *Cequeau*

sont distribuées sur chacune des mailles mais un unique jeu de paramètres est utilisé pour tout le bassin. Le nombre de paramètres à caler varie en fonction des entrées disponibles (*ex.* nécessité éventuelle de caler des paramètres de la méthode d'interpolation des mesures ponctuelles) et de la configuration choisie (*ex.* nombre/type de vidanges de réservoirs, présence du module neige/glace, etc.).

Le choix du nombre de mailles est une question toujours complexe pour les modèles distribués. Un nombre de mailles suffisant doit être utilisé pour tenir compte de l'hétérogénéité spatiale des entrées (notamment à travers les effets orographiques). Dans le même temps, le nombre de mailles a une forte influence sur le temps de calcul du modèle ce qui est particulièrement problématique pour nous puisque l'objectif est de tester le modèle sur plusieurs bassins versants et de faire de multiples calages sur chacun d'entre eux (souvent entre 30 et 40 calages). En conséquence, la procédure suivante a été retenue :

$$Nb_{mailles} = \frac{1}{100} \cdot (Zmax_{BV} - Zmin_{BV})$$

Un bassin comme *la Durance à Serre-Ponçon* dont la station est située à 700 m d'altitude et le point culminant à près de 4000 m sera représenté avec 33 mailles. Quelques verrous sont cependant mis en place pour éviter les situations où le nombre de mailles est trop faible ou celles où la surface moyenne des mailles est trop petite :

$$si\ (Nb_{mailles} < 10)\ \Rightarrow\ Nb_{mailles} = 10$$

$$S_{maille} = {}^{S_{bassin}}/_{Nb_{mailles}} \quad ; \quad si\ (S_{maille} < 5\,km^2) \Rightarrow\ S_{maille} = 5\,km^2$$

Tableau 2.4.: Liste des paramètres du modèle *Cequeau*

Paramètres CEQUEAU						Paramètres fixés pour le calage	
ID	Sous-groupe	Nom	Nom	Unité	ID		
1	Module neige	JoNeig	Jour de l'année au maximum d'ensoleillement pour la fonte de la neige	-	1	calé	
2		TF	Taux potentiel de fonte (facteur degré jour)	mm/°C/j	2	calé	
3		TS	Seuil de température de fonte	°C	3	calé	
4		TTD	Coefficient de déficit calorifique	-	4	calé	
5		TTS	Température du murissement du stock de neige	°C	5	calé	
6	Module glace	AblaG	Taux potentiel de fonte (facteur degré jour)	mm/°C/j	6	non calé	= TF
7		TNG	Temps caractéristique de la transfo neige-glace	j	7	non calé	= 10^4
8		TFG	Seuil de température de fonte	°C	8	non calé	= TS
9		TRad	Surplus de fonte (apporté par le rayonnement, indépendant de la température)	°C	9	non calé	0
10	Réservoir sol	HSolMax	Hauteur maximale du réservoir sol (au delà, ruissellement de surface)	mm	10	calé	
11		Hseuil/HSolMx	Seuil relatif sur Hsol qui détermine si l'évaporation est à taux plein ou réduit	-	11	calé	
12		Hinter/HSolMx	Seuil relatif sur Hsol qui active/désactive la vidange haute du reservoir sol	-	12	calé	
13		Hinfil/HSolMx	Seuil relatif sur Hsol qui active/désactive l'infiltration du reservoir sol vers la nappe	-	13	calé	
14		Culture	Coefficient cultural moyen (sur tout le bassin)		14	non calé	=1
15		Imper	Proportion du bassin imperméable	%	15	non calé	= 0
16		T_SolHaut	Constante de temps de vidange haute du réservoir sol	j	16	calé	
17		T_SolBas	Constante de temps de vidange basse du réservoir sol	j	17	non calé	= 10^6
18		T_Infil	Constante de temps de l'infiltration du réservoir sol vers la nappe	j	18	calé	
19	Réservoir nappe	HNapMax	Hauteur qui active/désactive la vidange haute (rapide) du réservoir nappe	mm	19	calé	
20		T_NapHaut	Constante de temps de vidange haute (rapide) du réservoir nappe	j	20	non calé	= 10^6
21		T_NapBas	Constante de temps de vidange basse (lente) du réservoir nappe	j	21	calé	
22		HNapSeuil/HNapMx	Seuil relatif sur Hnap qui détermine si l'évaporation est à taux plein ou réduit	-	22	non calé	= 0,99
23		HNapBas/HNapMx	Seuil relatif sur Hnap qui active/désactive la vidange basse (lente) du reservoir nappe	-	23	calé	
24	Routage	Célérité	Paramètre de célérité de l'équation 1D de propagation (Hayami)	j	24	calé	
25		Diffusion	Paramètre de diffusion de l'équation 1D de propagation (Hayami)	j	25	calé	
26	Méthode de spatialisation	HP	Hauteur caractéristique de croissance de la pluviométrie	m	26	non calé	x
27		aP	Portée du variogramme des résidus (pour les précipitations)	m	27	non calé	x
28		CT	Gradient altitudinal de température	°C/m	28	non calé	x
29		aT	Portée du variogramme des résidus (pour la température)	m	29	non calé	x
30	Spatialisation	STRNE	Température seuil de formation de la neige	°C	30	calé	
31		Cpsol	Coefficient de correction des précipitations solides	-	31	non calé	= 1
32		Th1	Paramètre 1 de la formule de Thornthwaite	-	32	calé	
33		Th2	Paramètre 2 de la formule de Thornthwaite	-	33	calé	
34		JoEva	Jour de l'année au maximum d'ensoleillement pour l'ETP	-	34	calé	

19 paramètres calés

2.4. Mise en place de la procédure

2.4.1. Groupes de bassins utilisés pour tester chaque modèle

La méthodologie de test *GSST* (*Generalized Split-Sample Test*) est mise en place sur nos deux échantillons de bassins versants (français et australiens). Dans chaque cas, différents modèles et structures sont à notre disposition. Le tableau 2.5 fournit la correspondance entre les configurations de modèles testées et les bassins sur lesquels celles-ci sont utilisées.

Tableau 2.5.: Résumé des modèles et bassins soumis à la procédure de test *GSST*

modèle	échantillon australien	échantillon français
GR4J	204 bassins	×
GR4J + CemaNeige$_{(\theta calés)}$	×	365 bassins
GR4J + CemaNeige$_{(\theta fixes)}$	×	365 bassins
Mordor6	204 bassins	×
Mordor6 + CemaNeige$_{(\theta fixes)}$	×	365 bassins
SimHyd	204 bassins	×
SimHyd + CemaNeige$_{(\theta fixes)}$	×	365 bassins
Mordor	×	6 bassins
Cequeau	×	20 bassins (+345 avec un seul calage)

2.4.2. Conditions d'utilisation du *GSST*

Longueur des sous-périodes de calage-validation

La procédure *GSST* repose sur des séries de calage-contrôle sur un ensemble de sous-périodes obtenues par découpage selon une fenêtre glissante de taille fixe (*cf.* figure 2.2, page 29). Choisir la longueur de cette fenêtre est une tâche délicate puisqu'il s'agit de notre longueur de période de calage (*Yapo et al.*, 1996; *Anctil et al.*, 2004). Celle-ci doit être suffisamment longue pour permettre une identification adéquate des paramètres du modèle. Dans le même temps, l'utilisation de périodes trop longues pourrait se révéler contre-productive vis-à-vis des objectifs de nos travaux. D'une part, elle réduirait les contrastes (notamment climatiques) entre les sous-périodes de transfert des paramètres. D'autre part, des sous-périodes trop longues

entraîneraient une chute importante du nombre de *split-sample tests (SST)* réalisables (*cf.* formule 2.1, page 29), ce qui handicaperait nos analyses. Si 30 années d'observations sont par exemple disponibles, alors un total de 380, 132 et 12 *SST* sont respectivement réalisables pour des longueurs de sous-périodes de 6, 10 et 14 années. Cette longueur de découpage doit donc également être choisie au regard de la longueur totale des séries d'observation disponibles.

Pour nos tests, nous avons choisi une longueur de fenêtre glissante de 10 ans. Cette longueur est commune à tous les modèles, et les années qui la composent sont définies comme des années *hydrologiques* (plutôt que *calendaires*). Cette utilisation d'années hydrologiques vise à éviter de couper nos chroniques françaises au cœur de l'hiver, en pleine période de stockage de neige pour les bassins nivaux. En effet, il est par exemple primordial que la douceur d'un hiver, capturée en partie sur une année calendaire *i*, soit mise en correspondance avec les débits de fonte associés qui auront lieu à l'année calendaire *i+1*. Bien évidemment, les dates de début et fin de ces années hydrologiques diffèrent spatialement (d'un bassin à un autre), et temporellement (d'une année à une autre pour un même bassin). Par souci de simplification, nous avons considéré pour nos tests une plage fixe s'étalant du *1er octobre* au *30 septembre* de l'année suivante. Notons enfin que l'utilisation d'années hydrologiques limite aussi les effets de l'initialisation, les réservoirs des modèles étant plutôt vides en octobre. Pour limiter plus encore les effets néfastes d'une initialisation incorrecte des modèles, nous avons utilisé dans chaque cas une année de mise en route (l'année qui précède la période de 10 ans dans la série d'observations) [7].

Nombre de tests de transfert menés

Pour chaque bassin, la disponibilité des forçages (1976-2006 en Australie et 1948-2010 en France), la présence de débits observés et enfin la longueur de découpage en sous-périodes constituent les trois éléments qui déterminent le nombre de tests de transfert de paramètres pouvant être réalisés (*cf.* formule 2.1, page 29). La figure 2.14 permet de visualiser rapidement le nombre de *split-sample tests (SST)* menés par bassin sur nos deux sous-échantillons (français et australien).

Sur l'échantillon australien, la longueur des séries australiennes permet de réaliser au moins 50 *SST* pour 90% de l'échantillon (183 bassins), un maximum de 156 *SST* étant atteint pour 65% de l'échantillon (134 bassins dont les séries s'étalent sur 31 ans). Les séries françaises sont plus longues et le nombre de *SST* réalisés y est donc nettement plus important, variant de 66 à plus de 1700 tests pour les bassins disposant des séries les plus longues (60 ans).

7. Pour garantir mieux encore des niveaux initiaux réalistes, nous avons dupliqué la période (1+10 ans), ces 22 années étant fournies au modèle et les erreurs n'étant calculées que sur les 10 dernières années.

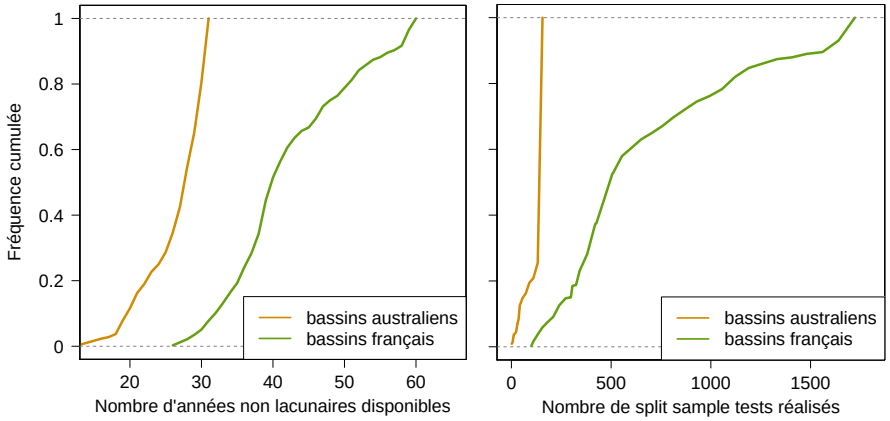


FIGURE 2.14.: Nombre d'années disponibles avec moins de 10% de lacunes et nombre de tests de transfert réalisés entre des sous-périodes indépendantes de 10 ans

(distribution des valeurs par bassin sur nos deux échantillons)

2.4.3. Critère de calage et d'évaluation des modèles hydrologiques

Le choix du critère de calage des modèles hydrologiques est une étape cruciale car ce critère est également celui qui sera utilisé pour l'évaluation de la robustesse des modèles lors de la procédure *GSST* (*cf.* section 2.1.3).

Dans cette thèse, nous avons retenu le critère KGE[8] proposé par *Gupta et al.* (2009). Ce critère présente de nombreux avantages vis-à-vis de notre approche d'analyse :

- Le KGE a été construit sur une décomposition du critère NSE de *Nash et Sutcliffe* (1970), avec pour objectif d'en corriger certains défauts à travers un rééquilibrage des différentes composantes.
- Le KGE est adimensionnel et sa formulation n'est pas liée à un modèle de référence comme c'est le cas pour le NSE qui compare les erreurs du modèle testé à celles d'un modèle naïf de variance nulle (*c.-à-d.* le modèle "moyenne"). Différentes valeurs de KGE peuvent être comparées même lorsqu'elles proviennent de configurations très différentes. Ce point est crucial pour nos travaux puisque nous souhaitons pouvoir rendre compte des comportements moyens des modèles sur de très nombreux bassins et que nous le ferons par le calcul de quantiles sur les valeurs des critères d'évaluation.
- L'expression du KGE est centrée autour de trois variables adimensionnelles qui sont calculées sur toute la période de calage et traduisent divers aspects des performances d'un modèle : le critère de corrélation entre les débits simulés et observés (ρ_{kge}) ; le rapport des variances

8. $KGE = 1 - \sqrt{(\rho_{kge} - 1)^2 + (\alpha_{kge} - 1)^2 + (\beta_{kge} - 1)^2}$, détail de la formule page 32.

des débits simulés et observés (α_{kge}) ; le biais moyen sur les débits modélisés (β_{kge}). Ces trois composantes sont associées de manière equipondérée pour composer le critère KGE et pourront être utilisées indépendamment pour apprécier la source des diminutions de KGE éventuellement constatées dans nos tests.

- Le KGE peut être utilisé sous la forme d'une erreur (ε=1-KGE), plutôt que d'un critère de performance, qui peut être exprimée relativement pour former le critère MRC (*Model Robustness Criterion*) proposé dans la section 2.1.3. [9]

$$MRC_{D\to R} = \frac{\varepsilon_{D\to R} - \varepsilon_{R\to R}}{\varepsilon_{R\to R}} = \frac{KGE_{D\to R} - KGE_{R\to R}}{1 - KGE_{R\to R}} \tag{2.7}$$

Concernant enfin les algorithmes d'optimisations utilisés pour ces calages, nous avons choisi d'utiliser les procédures accompagnant chacun des différents modèles considérés (nous partons du principe qu'elles leurs sont respectivement bien adaptées). Un algorithme de recherche locale, précédé d'une phase de préfiltrage, est utilisé pour l'optimisation des modèles *GR4J-CemaNeige*, *Mordor6-CemaNeige et SimHyd-CemaNeige*. Cette méthode a le grand avantage de permettre une convergence rapide vers un jeu optimal pour un critère d'erreur donné et permet ainsi de réaliser un grand nombre de calages tout en assurant leur répétabilité. Des procédures plus complexes qui s'apparentent pour partie à des algorithmes génétiques sont utilisées pour les modèles *Cequeau* et *Mordor*. Ces calages prennent plus de temps et ont parfois été limitants dans l'utilisation du *GSST*.

2.4.4. Variables utilisées pour caractériser les périodes dans la recherche des relations $\triangle$*erreurs=f($\triangle$caract.)*

L'évaluation de la robustesse climatique des modèles hydrologiques est menée à travers la recherche de liens entre les erreurs de modélisation consécutives à des transferts de paramètres et les différences de caractéristiques entre les périodes impliquées de ces transferts (*c.-à-d.* les périodes de calage et de validation). Nous recherchons ainsi des relations du type $\triangle erreurs = f(\triangle caract.)$, dont la pertinence est évaluée à l'aide d'un critère de corrélation de Spearman (*cf.* section 2.1.3).

La caractérisation des différences entre les périodes de calage et de validation n'est pas une tâche aisée. Nous pouvons en effet considérer toutes sortes d'indicateurs pour capter les particularités d'une période. Dans le cadre de nos travaux, nous portons bien entendu un intérêt particulier aux indicateurs reflétant les conditions climatiques. Nous pourrions également considérer des critères basés sur les débits car ils ont l'avantage de représenter une intégration

9. Si le critère NSE n'est pas directement comparable entre des bassins de dynamiques très différentes (à la différence du KGE), l'utilisation de la forme relative MRC=$(1-NSE_{D\to R})/(1-NSE_{R\to R})$-1 permet son utilisation pour nos analyses, à égalité avec un MRC calculé sur le KGE.

de diverses caractéristiques (précipitations, évapotranspiration, état d'humidité du bassin...). Toutefois, nous choisissons de ne pas concentrer nos analyses sur les débits car il nous semble préférable d'identifier (autant que possible) des variables explicatives sources. De plus, dans une perspective éventuelle de valorisation de nos travaux lors d'études d'impact du changement climatique, nous devons nous restreindre aux variables qui seront à notre disposition pour estimer le risque associé à un transfert de paramètres. Dans de tels cas en effet, seules des évolutions de variables climatiques pouvant être obtenues via des modèles climatiques sont disponibles.

Par conséquent, nos recherches de variables explicatives se font principalement avec les données de températures et de précipitations, qui sont issues de bases de réanalyses, et les estimations d'évapotranspiration, qui sont issues d'un modèle de calcul d'ETP[10]. Un récapitulatif des variables considérées dans nos recherches de corrélation est fourni sur le tableau 2.6.

TABLEAU 2.6.: Variables utilisées pour caractériser les périodes dans la recherche des relations $\triangle erreurs = f(\triangle caract.)$

Variable	Type de calcul réalisé
Température de l'air ($\overline{T}$) Précipitations totales ($\overline{P}$) Précipitations solides ($\overline{P}_{sol}$) Évapotranspiration potentielle ($\overline{ETP}$)	moyennes sur 10 ans moyennes saisonnières interannuelles
Indicateur d'aridité ($\overline{P}/\overline{ETP}$)	moyennes sur 10 ans moy. saisonnières de P sur moy. annuelles d'ETP
Variabilité des températures Variabilité des précipitations totales	écart-type (sur 10 ans) des moyennes annuelles
Anomalie mensuelle des températures Anomalie mensuelle des précipitations	moyenne des anomalies sur 10 ans ex. $\sum_{i=1}^{10} \frac{1}{10} \left(\overline{T}_{i,\,annuel}/max(\overline{T}_{i,\,mensuel})\right)$
Espacement temporel entre périodes (indépendamment de toute considération climatique)	

10. *Rq.* nous avons aussi testé l'utilisation d'une fréquence d'occurrence des types de temps (classification *EDF*) mais elle s'est révélée moins pertinente que les variables climatiques seules et n'a donc pas été retenue.

Partie II

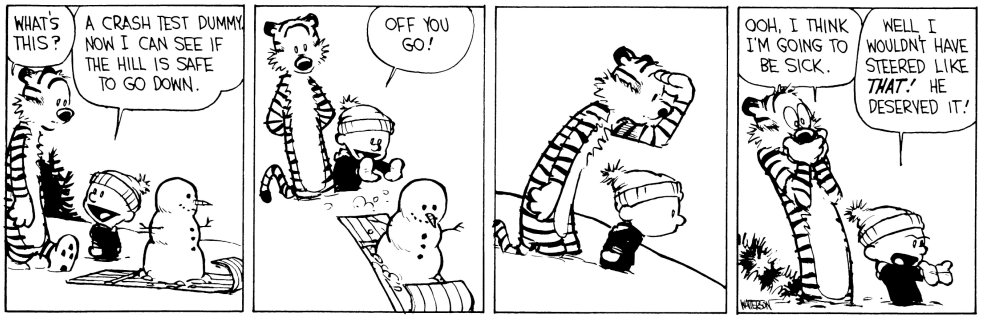
Première phase du diagnostic

Étude des symptômes

3

Quantifier la robustesse des modèles hydrologiques face au climat

Soumettre nos modèles à de véritables crash tests : une manière efficace de savoir ce qu'ils valent...



Calvin and Hobbes, 02/02/1989, Bill Watterson©

3.1. Introduction

3.1.1. Résumé des choix méthodologiques

Notre travail vise à étudier la robustesse temporelle des modèles hydrologiques dans un contexte de climat changeant. Il repose sur des séries de tests de calage-validation systématiques (*cf.* procédure *GSST* présentée à la section 2.1.2). L'échantillon de bassins utilisé se compose de 365 bassins français et 204 bassins australiens (*cf.* section 2.2).

La robustesse des modèles hydrologiques est évaluée à travers les variations des erreurs de simulation (en validation) et leur lien possible avec les différences entre les conditions de calage et de validation. Ce lien est évalué via le critère de corrélation de rang de *Spearman* (*cf.* section 2.1.3). Les erreurs de simulation sont mesurées par le critère KGE et ses trois composantes [1]. Ce critère proposé par *Gupta et al.* (2009) nous sert également de fonction objectif et sera, pour les analyses, exprimé sous une forme d'erreur relative afin de mettre en évidence les erreurs causées par le transfert des paramètres *:* $MRC_{D \to R} = \frac{(1-KGE_{D \to R}) - (1-KGE_{R \to R})}{(1-KGE_{R \to R})}$ (*cf.* sections 2.1.3 et 2.4.3). Enfin, les variables climatiques nous servant à caractériser les conditions de calage et validation sont principalement des moyennes annuelles, saisonnières et des indicateurs de variabilité sur les températures et les précipitations (*cf.* section 2.4.4).

3.1.2. Niveaux de performances des modèles

La procédure *GSST* est mise en place sur tout l'échantillon de bassins selon les conditions rappelées ci-dessus. Nous disposons ainsi d'un très grand nombre de simulations pour nos différents modèles et bassins. Ce chapitre 3 présente les diverses recherches de liens entre les conditions de transfert des paramètres et la qualité des simulations. Avant de nous y plonger, comparons nos simulations aux observations pour dégager un aperçu global des niveaux de performances de modélisation auxquels nous aurons à faire. La figure 3.1 présente les valeurs des critères *KGE, NSE* et du biais sur le volume moyen en calage (en rouge) et en contrôle (en bleu).

Ces performances en calage et validation sont estimées sur chacun des bassins et agglomérées ici sous la forme de boîtes à moustaches qui fournissent les quantiles 5%, 25%, 50%, 75% et 95%. La procédure *GSST* est utilisée sur l'échantillon complet pour les modèles *GR4J*, *Mordor6* et *SimHyd*, et sur seulement une vingtaine de bassins pour le modèle *Cequeau* (compte tenu des temps de calages beaucoup plus importants) et six bassins pour le modèle *Mordor* dont les résultats ne sont pas présentés ici. Nous pouvons compléter ces résultats en représentant les distributions des pertes de performances constatées au niveau de chaque bassin.

1. $KGE = 1 - \sqrt{(\rho_{kge} - 1)^2 + (\alpha_{kge} - 1)^2 + (\beta_{kge} - 1)^2}$, détail de la formule page 32.

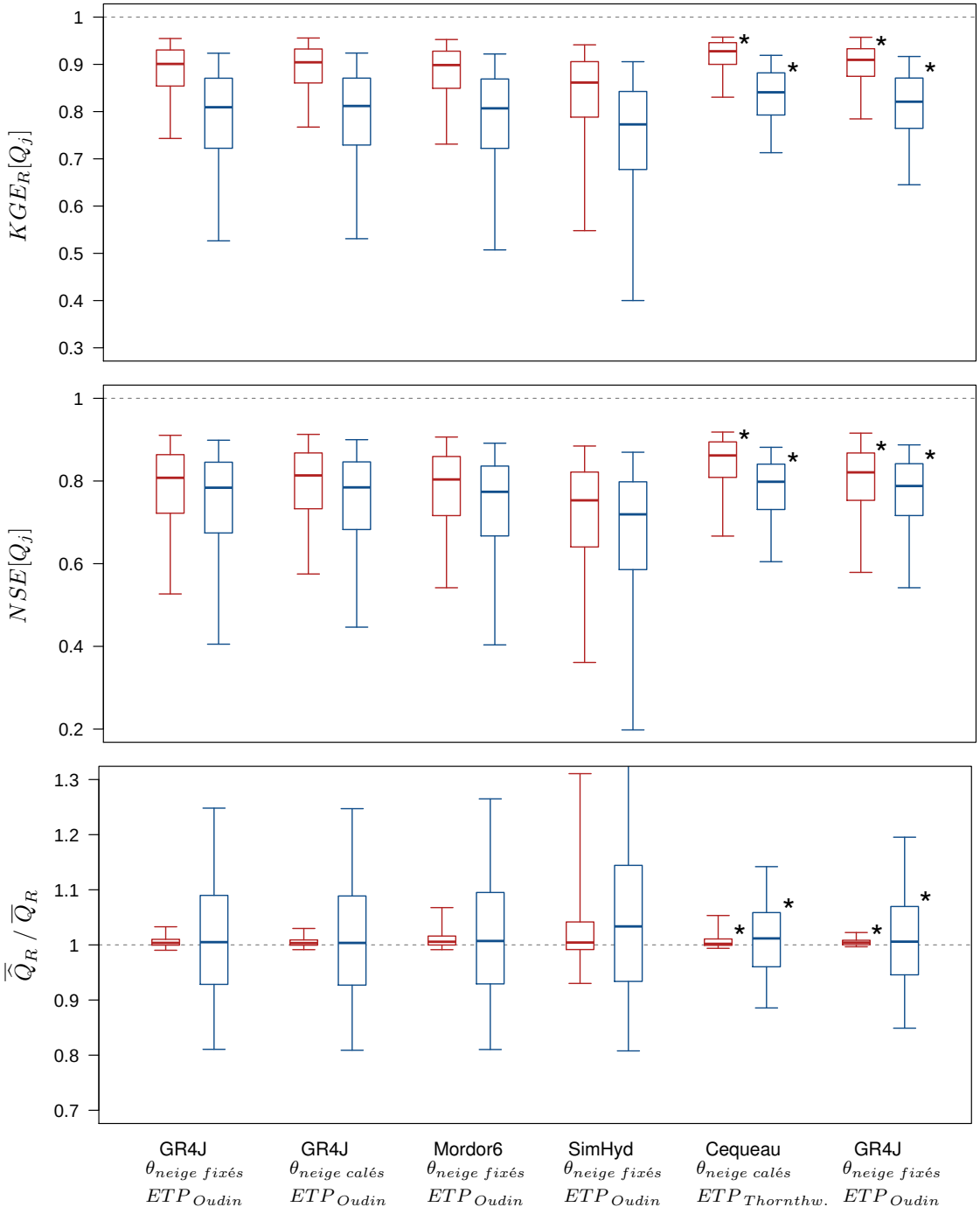


FIGURE 3.1.: Performances moyennes des modèles en calage et en validation sur tout l'échantillon (569 bassins)

trait épais central : valeures médianes, boîtes : quantiles 25 et 75%, moustaches : quantiles 5 et 95%
** résultats obtenus sur un sous-groupe de 20 bassins versants plutôt que sur tout l'échantillon*

Ces distributions sont fournies en annexe D.1 (page 276) pour *GR4J*, *Mordor6* et *SimHyd.* À ce stade, nous nous contenterons néanmoins de la représentation simple proposée ici car elle permet une première appréciation des pertes de performances constatées en moyenne sur nos bassins.

Précisons enfin que le nombre de simulations diffère fortement entre les boîtes concernant respectivement les tests de calage et de validation. En effet, le *GSST* permet un nombre de tests de transfert beaucoup plus important que le nombre de sous-périodes effectivement considérées. Par exemple, une chronique d'observation de 30 ans, découpées en sous-périodes de 10 ans conduira à 21 situations de calage et 132 situations de validation.

Les performances en calage sont tout à fait acceptables pour les modèles considérés avec des *KGE* médians autour de 0.9 (*NSE* autour de 0.8) pour *GR4J*, et *Mordor6* et des valeurs légèrement inférieures pour *SimHyd.* Le passage de calage à validation engendre une augmentation significative des erreurs, notamment visible au niveau des biais moyens (*cf.* figure 3.1). Le modèle *Cequeau*, pour lequel le *GSST* n'a été appliqué que sur une vingtaine de bassins montagneux présente des performances meilleures que *GR4J*, autant en calage qu'en validation. *Cequeau*, qui est semi-distribué, valorise probablement mieux l'information spatiale fournit par la base *SPAZM* que les autres modèles qui sont globaux et dont le module neige étagé en bandes d'altitudes constitue la seule forme de valorisation des entrées spatiales.

3.1.3. Précisions quant au calage du module *CemaNeige*

Aucune modélisation de la neige n'est effectuée sur les bassins australiens où les précipitations solides demeurent faibles et n'impactent pas le régime des écoulements. Sur les bassins français de notre échantillon, la modélisation de la neige est en revanche prise en compte, de façon intégrée pour les modèles *Cequeau* et *Mordor* et par un couplage avec le module *CemaNeige* pour les modèles *GR4J*, *Mordor6*, et *SimHyd.*

Les travaux de *Valéry* (2010), qui développa *CemaNeige*, montrèrent que les dégradations causées par l'utilisation de valeurs fixes plutôt que calées pour les deux paramètres ce module restaient limitées lors d'une analyse moyenne sur un large échantillon de bassins. Pour notre application du *GSST*, nous pouvions donc envisager deux configurations : (a) un *CemaNeige* dont les paramètres sont fixés aux valeurs médianes utilisées par *Valéry* (2010), (b) un *CemaNeige* dont les paramètres sont calés en même que le modèle hydrologique.

À première vue, la deuxième configuration semble de loin la plus adaptée à notre approche générale, puisqu'elle permet de tester la transférabilité de jeux complets de paramètres (incluant le module neige). Nous avons donc mené l'intégralité de la procédure *GSST* pour le couple *GR4J-CemaNeige* (calé) sur notre échantillon français. Dans le même temps, les tests ont été menés à l'identique pour la première configuration (*GR4J-CemaNeige* $\theta_{fixés}$).

Comparaison des performances moyennes en validation

L'influence du calage ou de la fixation des paramètres de *CemaNeige* sur les performances moyennes en calage et en validation du couple *GR4J-CemaNeige* est discutée en annexe C.2. Assez logiquement, les situations de performances équivalentes entre ces configurations sont obtenues sur les bassins où les précipitations solides sont les plus faibles, tandis que la configuration avec *CemaNeige* calé est préférable sur les bassins où ces précipitations sont plutôt importantes. Il est intéressant de constater que la configuration *GR4J-CemaNeige* $\theta_{fixés}$ permet (en moyenne) des erreurs moins importantes en validation sur les bassins où les précipitations solides représentent 5 à 10% des apports annuels. Cette situation est probablement révélatrice de problèmes de robustesse qui sont finalement assez logiques. La variabilité naturelle du climat entraîne, sur de tels bassins, une alternance de périodes où les précipitations solides influencent très peu les écoulements et de périodes où cette influence est plus prononcée. Nous imaginons aisément que des calages sur les sous-périodes faiblement concernées (régime quasi-pluvial) peuvent conduire à la sélection de paramètres irréalistes pour *CemaNeige* dont l'impact ne sera ressenti que lors de transferts sur des périodes où le régime est pluvio-nival. Bien que ces comportements méritent de plus amples investigations, nous ne porterons pas notre attention sur ces situations où les problèmes de transférabilité sont facilement évitables (*ex.* en n'autorisant le calage du module neige que lorsque la part solide des précipitations annuelles dépasse un certain seuil). Nous focaliserons en effet notre attention sur des problèmes de transférabilité plus prononcés et surtout plus délicats à expliquer et éviter, que nous présenterons au fil de ce manuscrit.

Préférence pour la version non calée pour la majorité des tests à venir

Sans pour autant dévoiler la suite de notre exposé, nous sommes contraints de préciser que les deux configurations envisagées pour *CemaNeige* (*c.-à-d.* paramètres calés ou fixés) conduisent à des résultats sensiblement proches sur les critères qui nous paraissent pertinents lors des analyses associées à la procédure *GSST*. Plus précisément concernant ce chapitre 3, le fait de ne pas caler le module neige n'impacte pas réellement les comportements mis en avant dans les sections 3.2.1, 3.3 et 3.4.

Face à cette situation nous avons donc fait le choix d'effectuer la majorité de nos tests avec les modèles *GR4J*, *Mordor6* et *SimHyd* en utilisant la configuration où les paramètres de *CemaNeige* ne sont pas calés. Ce choix qui peut paraître discutable, a pour objectif de minimiser les degrés de liberté des modèles les plus simples et permettre (nous l'espérons) une meilleure compréhension des phénomènes qui conduisent aux manques de robustesse parfois constatés pour nos modèles.

Des résultats faisant intervenir la version calée de *CemaNeige* apparaîtront toutefois à plusieurs reprises dans le manuscrit, en particulier dans la section 3.5 qui traite explicitement de la robustesse des modèles en zone de montagne.

3.2. Conditions de calage, conditions de validation et erreurs des modèles

3.2.1. Sélection des tendances significatives

Une comparaison est effectuée au niveau de chaque bassin entre les variations d'erreurs en validation et les écarts sur diverses caractéristiques (principalement climatiques) entre les périodes de calage et de validation. En pratique, cette comparaison se traduit par le calcul d'une valeur de critère de *Spearman*, notée r_s ou r'_s selon la manière dont elle est calculée (*cf.* section 2.1.3). Donner un aperçu clair des résultats de ces nombreuses comparaisons est une tâche délicate. Le nombre de caractéristiques testées et surtout le nombre de bassins rendent impossible la présentation des différentes corrélations identifiées sur chaque bassin. Nous ne souhaitons pourtant pas renoncer à notre large échantillon qui permet une focalisation sur le comportement des modèles en se détachant des spécificités de quelques cas d'étude.

Une approche pourrait consister à trier les situations selon les valeurs de critère de *Spearman* obtenues pour n'en retenir qu'une seule par bassin (celle donnant la corrélation la plus élevée). Elle présente cependant un défaut majeur puisqu'elle masque les relations de second ordre pouvant exister entre les erreurs du modèle et diverses caractéristiques de période. Nous ne pourrons donc pas apprécier la représentativité réelle d'un lien de corrélation au sein de notre échantillon. Pour illustrer ce défaut, prenons un exemple dans lequel la température moyenne au printemps ($\overline{T}_{MAM}$) et en été ($\overline{T}_{JJA}$) seraient toutes les deux prépondérantes sur la transférabilité des paramètres d'un modèle. Si la première renvoie systématiquement des critères de *Spearman* légèrement plus élevés, l'information sur la représentativité de la seconde disparaîtrait alors de notre analyse. La seule façon de s'en prémunir serait d'augmenter le nombre d'indicateurs utilisés pour caractériser les périodes (*ex.* $\overline{T}_{MAM}$, $\overline{T}_{JJA}$, $\overline{T}_{MAM+JJA}$, mais aussi $\overline{T}_{MAM+SON}$, $\overline{T}_{MAM+JJA+SON}$, ...).

Nous avons donc opté pour une approche dans laquelle la représentativité du lien entre $\triangle erreurs$ et $\triangle caract.$ est estimée de façon indépendante sur chaque caractéristique (*c.-à-d.* de façon non-exclusive vis-à-vis des autres caractéristiques considérées). Pour chaque couple ($\triangle erreurs$,$\triangle caract.$), nous comptabilisons tous les bassins de l'échantillon pour lesquels le critère de *Spearman* dépasse un seuil fixe, marquant la limite de significativité d'une corrélation. Par cette approche, nous tirons pleinement profit du fait d'avoir un large échantillon de bassins qui nous sert à sonder la représentativité des liens que nous étudions ($\triangle erreurs = f(\triangle caract.)$). Nous pourrons par la suite former des groupes de bassins en fonction des liens détectés. Un *zoom* sur ces groupes permettra de mieux analyser la situation, notamment à travers nos analyses visuelles des nuages de points proposées dans la section 2.1.3.

L'approche que nous utilisons (*c.-à-d.* seuil sur le critère de *Spearman* puis analyse visuelle) pourra surprendre car elle ne s'appuie pas sur des tests statistiques visant à rechercher des

corrélations multi-variées pour expliquer les erreurs des modèles, puis tester le degré de signficativilté de celles-ci. En pratique, les nuages de points obtenus à l'aide du GSST sont extrêmement diffus et il nous paraissait délicat de choisir un type de relation entre $\triangle erreurs$ et $\triangle caract.$ Pour ce travail exploratoire, notre objectif était, avant tout, d'apprécier les comportements moyen des modèles face à différentes configurations défavorables. Nous avons donc opté pour des analyses simples et visuelles, qui présentent l'avantage d'être facilement comprises et partagées.

3.2.2. Tableau récapitulatif des corrélations obtenues avec *GR4J*

Mode de représentation

Le tableau 3.1 récapitule les résultats des calculs de corrélation obtenus lorsque *GR4J* est soumis au *GSST* sur tout notre échantillon de bassins. Dans ce tableau, les différentes caractéristiques testées composent les lignes, et les indicateurs d'erreurs composent les colonnes (critères *MRC*, ρ_{kge}, α_{kge}/β_{kge} et β_{kge}). Le contenu des cases indique le taux de bassins concernés par une corrélation de *Spearman* significative ($|r_s| > 0.5$ ou $|r'_s| > 0.5$). Ces taux sont repris à travers un code couleur qui met en évidence les situations les plus courantes. Pour ces analyses, l'échantillon complet est considéré d'un seul tenant, les bassins français et australiens y étant rassemblés. Les tableaux obtenus pour les modèles *Mordor6* et *SimHyd* sont quant à eux disponibles dans l'annexe D.2 (page 277).

Pour rappel, les critères ρ_{kge}, α_{kge} et β_{kge} sont les différentes composantes du KGE (*cf.* équation 2.4, page 32) et correspondent respectivement au critère de corrélation de *Pearson*, au rapport des variances et au rapport des volumes entre les débits simulés et les débits observés. Le *MRC* est, quant à lui, une représentation relative du KGE (*cf.* équation 2.7, page 58).

Notons également que nous avons jugé plus pertinent de considérer le rapport α_{kge}/β_{kge}, au lieu d'étudier les variations de α_{kge}. En notant σ l'écart-type d'une série et μ sa moyenne, le rapport α_{kge}/β_{kge} correspond ainsi à $(\sigma_{\widehat{Q}}/\mu_{\widehat{Q}})/(\sigma_Q/\mu_Q)$, qui n'est autre que le rapport des coefficients de variation entre les simulations et les observations. Cela revient simplement à normaliser les écarts-types avant comparaison et permet de dissocier le biais de la capacité du modèle à reproduire une variance. En effet, une situation où les débits simulés seraient par exemple surestimés de 10% en moyenne, conduirait nécessairement à un écart-type (non normalisé) plus élevé que celui des observations et influencerait notre α_{kge}.

Rappelons enfin que la méthode de calcul d'une corrélation et le signe du critère de *Spearman* dépendent du critère d'erreur considéré (*cf.* section 2.1.3). Une seule colonne retranscrit l'augmentation de l'erreur avec celle de $|\triangle caract.|$ pour les critères *MRC* ($r'_s > 0.5$) et ρ_{kge} ($r'_s < 0.5$). Cette augmentation d'erreur est traduite sur deux colonnes pour les

critères α_{kge}/β_{kge} et β_{kge}, distinguant le test d'une corrélation ($r_s > 0.5$) et d'une anti-corrélation ($r_s < 0.5$). Rappelons enfin que, dans chaque cas, les erreurs et les caractéristiques climatiques sont calculées sur des sous-périodes construites sur une fenêtre glissante de 10 ans. Nous comparerons par exemple le biais sur le débit de la période *receveuse* $\left(\beta_{kge} = \overline{(\widehat{Q}_R)}_{10ans} \,/\, \overline{(Q_R)}_{10ans}\right)$ avec l'écart de température hivernale entre périodes *donneuse* et *receveuse* $\left(\triangle\overline{T}_{DJF} = \overline{(T_{DJF,D})}_{10ans} - \overline{(T_{DJF,R})}_{10ans}\right)$.

Premier niveau d'analyse des résultats

Le tableau 3.1 regroupe de très nombreuses informations. Toutefois, un premier résultat se détache clairement : l'hétérogénéité des situations. Notre tableau comporte en effet beaucoup de jaune et finalement moins de blanc que nous aurions pu l'imaginer, ce qui indique que des situations très diverses sont rencontrées sur l'échantillon tout en étant dans des proportions non négligeables. Parmi les situations qui ressortent, nous mentionnerons les suivantes : Les variations de *MRC* présentent un lien plus fréquent avec la température qu'avec les autres variables. Le critère de biais est quant à lui fréquemment corrélé avec un $\triangle caract.$, celui-ci pouvant aussi bien se rapporter à la température qu'aux précipitations ou à une combinaison des deux. Plus étonnant encore, des taux de représentativité non négligeables sont observés simultanément pour des corrélations et des anti-corrélations entre le biais et les variations d'une même variable climatique. Si les erreurs mesurées à travers de ρ_{kge} ne semblent pas varier en lien avec des changements de conditions climatiques, les corrélations détectées sur les erreurs de variance présentent également des taux importants de représentativité. Tout comme pour les erreurs de volumes, ces analyses révèlent néanmoins des variations de variance tantôt corrélées, tantôt anti-corrélées aux indicateurs climatiques.

Les variations du *MRC* et des autres composantes analysées ici sont intimement liées. Le parallèle entre les corrélations constatées sur les coefficients de variation (σ/μ) et celles sur le *MRC* n'est cependant pas aisé. À ce stade de l'analyse, nous ne connaissons pas l'amplitude des écarts sur σ/μ et leur impact potentiel sur le KGE (et donc sur le *MRC*). La situation est différente pour les corrélations de biais, dont nous avons entrevu les amplitudes à la figure 3.1. Connaissant l'impact très négatif d'un biais sur un critère d'erreur tel que le KGE, nous pouvons d'ores et déjà supposer que les corrélations concernant les *MRCs* se recoupent avec celles concernant les biais (*c.-à-d.* sur le tableau 3.1, il est fort probable que les bassins comptabilisés dans la colonne du *MRC*, à gauche, le soient également dans l'une des deux colonnes du $\overline{\widehat{Q}}/\overline{Q}$, à l'extrémité droite). Par la suite, nous analyserons visuellement les nuages $\triangle erreurs = f(\triangle caract.)$ (*cf.* sections 3.3 et 3.4) ce qui devrait permettre de confronter ces intuitions aux résultats.

TABLEAU 3.1.: Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *GR4J*

Résultats pour GR4J-CemaNeige (θ_{neige} fixés) avec des données SPAZM et une ETP d'Oudin.

Corrélations calculées individuellement sur chacun des 569 bassins puis agglogémérées pour estimer un taux de représentativité au sein de l'échantillon complet (c.-à-d. pourcentage de bassins où $|r_s| > 0.5$).

Le code couleur souligne simplement les taux indiqués dans les cases ($< 5\%$ des bassins concernés en blanc, $\approx 15\%$ en jaune, $\approx 30\%$ et $> 45\%$ en rouge).

	MRC	$r_p\left(\widehat{Q} \sim Q\right)$	$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_S>+0.5$	$r'_S<-0.5$	$r_S<-0.5$	$r_S>+0.5$	$r_S<-0.5$	$r_S>+0.5$
$\overline{T}$	21%	3%	19%	36%	51%	12%
$\overline{T}_{DJF}$	12%	4%	20%	33%	46%	14%
$\overline{T}_{MAM}$	19%	2%	21%	30%	46%	16%
$\overline{T}_{JJA}$	19%	3%	18%	38%	52%	13%
$\overline{T}_{SON}$	8%	2%	16%	31%	42%	9%
$\overline{P}$	8%	3%	28%	9%	20%	22%
$\overline{P}_{DJF}$	5%	2%	29%	12%	12%	31%
$\overline{P}_{MAM}$	6%	3%	31%	12%	11%	37%
$\overline{P}_{JJA}$	7%	3%	25%	18%	20%	25%
$\overline{P}_{SON}$	8%	1%	18%	24%	45%	11%
$\overline{P} / \overline{ETP}$	9%	3%	35%	12%	13%	35%
$\overline{P}_{DJF} / \overline{ETP}$	7%	3%	31%	14%	11%	39%
$\overline{P}_{MAM} / \overline{ETP}$	8%	3%	36%	14%	10%	46%
$\overline{P}_{JJA} / \overline{ETP}$	7%	3%	29%	15%	15%	32%
$\overline{P}_{SON} / \overline{ETP}$	6%	1%	18%	20%	35%	10%
$var(\overline{P}_{an})$	5%	2%	23%	12%	11%	28%
$var(\overline{P}_{an}/\overline{P}_{mens})$	3%	1%	19%	15%	18%	19%
$var(\overline{T}_{an})$	5%	1%	13%	19%	19%	13%
$var(\overline{T}_{an}/\overline{T}_{mens})$	12%	1%	27%	20%	15%	44%
Δ temporel	17%	2%	21%	39%	60%	13%

Second niveau d'analyses des résultats

Les premières constatations majeures étant formulées, intéressons-nous maintenant au détail des résultats visibles sur le tableau 3.1. Pour faciliter sa lecture, raisonnons en lignes et bloc de lignes, ce qui permet de constater la représentativité d'une corrélation entre erreurs et caractéristiques au niveau du critère composite *MRC* et des composantes considérées.

Nous observons des taux de représentativité relativement élevés de corrélation entre les variations du *MRC* et les écarts de température. Que nous considérions des moyennes annuelles ou saisonnières, ces taux sont proches (même si quelques différences sont constatées pour le *MRC*). Cela nous indique que c'est bien une évolution homogène de la température qui est reliée ici à une mauvaise transférabilité des paramètres du modèle, plutôt qu'un changement sur une saison en particulier. Nous constituerons par la suite un sous-groupe de bassins concernés par ces corrélations significatives, pour y analyser plus en détail les comportements (*cf.* section 3.3).

Des taux de représentativité non négligeables sont également constatés entre les erreurs de simulation et les variations de précipitations moyennes. Ces taux demeurent moins importants que ceux observés avec la température, mais feront aussi l'objet d'une analyse spécifique (*cf.* section 3.4).

Les caractéristiques construites sur le rapport entre précipitations et évapotranspiration potentielle conduisent également à des taux de représentativité parfois élevés. Il reste cependant délicat de déterminer si ces taux indiquent une réelle relation ou résultent simplement de l'alternance de liens tantôt avec la température (et donc l'ETP), tantôt avec les précipitations.

Un mot sur la mesure de l'amplitude associée à une corrélation

Avant de continuer, rappelons que le critère de *Spearman* nous informe sur le degré de corrélation entre nos variables, mais ne nous renseigne pas sur l'amplitude des $\triangle erreurs$ en jeu. Les résultats présentés dans le tableau 3.1 ne doivent donc pas être sur-interprétés. Néanmoins, les très forts niveaux de corrélations correspondent nécessairement (par construction) aux plus fortes intensités de pertes de performances (puisque le *panier* de points analysés reste toujours identique, seul le positionnement en abscisse étant modifié, *cf.* section 2.1.3).

Lorsque la tendance est jugée significative, l'amplitude associée à une corrélation peut être estimée par la pente d'une régression linéaire. Ainsi, nous pouvons calculer une pente pour chaque bassin entre le $\triangle erreurs$ étudié (*ex.* le biais en validation) et les variations d'une caractéristique donnée (*ex.* l'écart de températures entre les périodes de calage et validation). Une comparaison de ces pentes permet d'apprécier approximativement l'intensité relative du lien

entre $\triangle erreurs$ et $\triangle caract.$[2] Ce type de comparaisons fonctionne relativement bien lorsqu'une seule caractéristique est considérée mais cette approche devient beaucoup plus problématique lorsque nous mélangeons différentes caractéristiques et souhaitons identifier celle conduisant aux $\triangle erreurs$ les plus intenses. Nous sommes en effet contraints à devoir statuer sur l'importance d'une caractéristique vis-à-vis d'une autre. Prenons un exemple où ces pentes indiquent qu'un $\triangle\overline{T} = +1°C$ et un $\triangle\overline{P} = +100\,mm/an$ conduisent à un même $\triangle erreurs = +15\%$. L'une de ces situations est-elle plus problématique ou sont-elles équivalentes ? Quels critères de jugement, si possible objectifs, devons-nous définir (*ex.* basés sur une variabilité observée ou une probabilité d'occurrence) ?

Face à cette situation, nous avons fait le choix de nous restreindre aux valeurs de critères de *Spearman* lors des comparaisons transversales entre les $\triangle erreurs$ et plusieurs $\triangle caract.$ Les amplitudes d'erreurs en jeu seront ensuite analysées par analyses visuelles des nuages de points et l'utilisation de pentes de régression restreinte aux comparaisons entre bassins, menées sur chaque variable climatique indépendamment des autres.

Méthode de constitution du sous-groupe

Nous construisons nos analyses sur des comparaisons visuelles menées sur des sous-groupes de bassins rassemblant les comportements similaires. Par exemple, nous constituons un sous-groupe comprenant les bassins où des corrélations ont été détectées entre l'augmentation des erreurs lors du transfert des paramètres et les écarts de température moyenne entre les sous-périodes. Dans la continuité du tableau 3.1, cette sélection repose uniquement sur les niveaux des critères de *Spearman* obtenus par bassin lors du traitement des résultats de la procédure $GSST$. En reprenant ainsi le même seuil de significativité, fixé à 0.5, nous avons choisi de baser notre sélection sur les variations du *MRC* pour le modèle *GR4J*. Ces sous-groupes servent pour mener des analyses comparatives entre différents modèles, notamment *Mordor6* et *SimHyd* qui ont aussi été soumis au *GSST* sur la totalité de l'échantillon. Il est vrai que nous aurions pu construire nos sous-groupes de bassins en considérant les variations de biais plutôt que de *MRC* ou en utilisant *Mordor6* ou *SimHyd* plutôt que *GR4J*. Toutefois, ces différences de critères de constitutions des sous-groupes d'analyse n'auraient pas conduità la sélection de bassins très différents. Le tableau 3.2 montre en effet des taux de recouvrement très élevés (supérieurs à 80%) pour les différentes configurations testées. Les figures D.2 et D.3 en annexe (pages 280-281) complètent ces résultats en montrant comment les caractéristiques des bassins sélectionnés diffèrent peu selon la méthode utilisée pour constituer des groupes ($MRC = f(\triangle caract)$ ou $\overline{\widehat{Q}}/\overline{Q} = f(\triangle caract)$).

2. Ce procédé sera par exemple utilisé pour refléter l'intensité des biais de simulation par ajustement de la taille des triangles sur les cartes des figures 3.4 et 3.11.

Tableau 3.2.: Taux de recouvrement entre les sous-groupes utilisés pour les analyses *GSST* en fonction des critères de sélection des bassins

Les critères r et r' indiqués ici correspondent à des corrélations de Spearman calculées sur les nuages de points obtenus par GSST sur chaque bassin. Les précisions sur la méthode de calcul de ces critères sont fournies dans la section 2.1.3.

Rq. sur les bassins français, GR4J, Mordor6 et SimHyd sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).

$\triangle caract.$	sélection critère	sélection modèle	taux de recouvrement
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$	*GR4J* ou *Mordor6*	80%
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$	*GR4J* ou *SimHyd*	88%
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$	*Mordor6* ou *SimHyd*	84%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$	*GR4J* ou *Mordor6*	97%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$	*GR4J* ou *SimHyd*	93%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$	*Mordor6* ou *SimHyd*	95%
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*GR4J*	98%
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*Mordor6*	100%
$\triangle\overline{T}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*SimHyd*	99%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*GR4J*	100%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*Mordor6*	100%
$\triangle\overline{P}$	$r'_s(MRC) > 0.5$ ou $\left\|r_s(\overline{\widehat{Q}}/\overline{Q})\right\| > 0.5$	*SimHyd*	100%

3.3. Zoom sur les cas de problèmes de robustesse en lien avec △température

3.3.1. Tableau de corrélation

Le tableau 3.1 nous a permis de constater l'existence de corrélations entre △*erreurs* et △*température* sur une part non négligeable de notre échantillon de bassins. Nous évoquions précédemment que les cas de mauvaise transférabilité des paramètres semblent liés à une évolution homogène de la température plutôt qu'un changement sur une saison en particulier. Le tableau 3.3 reprend les résultats sur ces variables de température, mais nous y avons détaillé les taux de représentativité sur les échantillons français (365 bassins) et australiens (204 bassins).

TABLEAU 3.3.: Taux de corrélations significatives entre △erreurs et $\triangle\overline{T}$ pour *GR4J*

Sur chaque cellule, résultats sur tout l'échantillon (haut) sur l'échantillon français (bas-gauche) et sur l'échantillon australien (bas-droite). Données et couleurs identiques à celles du tableau 3.1. Rq. en France, GR4J est alimenté par SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ fixés).

	MRC	$r_p\left(\widehat{Q} \sim Q\right)$		$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_S>+0.5$	$r'_S<-0.5$		$r_S<-0.5$	$r_S>+0.5$	$r_S<-0.5$	$r_S>+0.5$
$\overline{T}$	21% 25% 12%	3% 2% 6%		19% 21% 13%	36% 38% 30%	51% 61% 30%	12% 11% 14%
$\overline{T}_{DJF}$	12% 11% 14%	4% 1% 8%		20% 23% 15%	33% 31% 35%	46% 53% 32%	14% 13% 17%
$\overline{T}_{MAM}$	19% 27% 5%	2% 2% 2%		21% 20% 24%	30% 38% 14%	46% 64% 13%	16% 10% 26%
$\overline{T}_{JJA}$	19% 21% 13%	3% 2% 4%		18% 22% 12%	38% 38% 35%	52% 60% 37%	13% 12% 13%
$\overline{T}_{SON}$	8% 8% 9%	2% 2% 2%		16% 19% 12%	31% 29% 31%	42% 46% 32%	9% 10% 8%

Nous remarquons que les corrélations entre les erreurs de simulation et les écarts de température moyenne annuelle ou saisonnière sont plus fréquentes sur l'échantillon français qu'australien. Si nous considérons les corrélations $MRC = f(\triangle\overline{T}_{an})$, 21% de l'échantillon complet sont concernés par un $r'_s > 0.5$ (120 bassins versants), mais cette valeur serait de 25% si elle était calculée sur l'échantillon français seulement et elle tomberait à 12% si elle était calculée sur l'échantillon australiens seulement. Les résultats sont encore plus marqués si nous nous intéressons aux erreurs sur les volumes moyens sur 10 ans avec des corrélations négatives identifiées pour plus de 60% des bassins français. Ces forts taux révèlent la présence régulière du comportement suivant : $\triangle\overline{T}_{an} < 0 \Leftrightarrow$*surestimation du débit moyen* et $\triangle\overline{T}_{an} > 0 \Leftrightarrow$*sous-estimation du débit moyen*. Restons là encore prudents dans nos interprétations en rappelant que ces corrélations indiquent des variations simultanées mais ne constituent pas, à elles seules, un lien avéré de causalité.

3.3.2. Variations des erreurs sur un sous-groupe choisi

Composition du sous-groupe

Nous avons constitué un sous-groupe de bassins pour étudier les problèmes de robustesse des modèles en lien avec les écarts de température entre calage et validation. À partir d'un seuil de corrélation de Spearman choisi à 0.5, nous avons sélectionné les bassins présentant un lien de corrélation entre l'augmentation des erreurs de *GR4J*, estimée via le *MRC*, et les écarts de température moyenne annuelle. Les bassins retenus correspondent aux 21% de l'échantillon complet pour lesquels un lien de corrélation fut détecté entre l'augmentation des erreurs de *GR4J* liées au transfert des paramètres et les écarts de température moyenne annuelle. Cette sélection correspond sur le tableau 3.3, aux bassins comptabilisés dans la cellule en haut à gauche. Elle concerne ainsi 120 bassins, dont 93 français et 27 australiens. Un aperçu des caractéristiques des bassins de ce sous-groupe est disponible en annexe (*cf.* figure D.2, page 280).

Analyse des nuages de points $\triangle erreurs = f(\triangle \overline{T})$

Les nuages de points reliant $\triangle erreurs$ et $\triangle caract.$ sont construits pour chacun des 120 bassins puis rassemblés pour une analyse visuelle commune présentée aux figures 3.2 et 3.3. La procédure de construction de cette représentation est détaillée dans la section 2.1.3.

Ces nuages de points mettent en avant de nettes corrélations entre les augmentations d'erreurs de simulation et les écarts de température annuelle moyenne subis lors du transfert des paramètres [3]. Cette tendance apparaît d'autant plus nettement par contraste avec l'absence de tendance sur les nuages reliant le *MRC* aux écarts de précipitations. Les mêmes tests y étant représentés, nous observons en effet des hauteurs de boîtes beaucoup plus grandes et homogènes, tout comme leur positionnement vertical, signes d'un nuage diffus marquant l'absence complète de corrélation. Il semble important de noter que les trois modèles testés présentent tous des comportements relativement similaires malgré des structures différentes. Si cette représentation ne permet pas d'établir un lien irréfutable de causalité, les différences de niveau d'erreur entre la partie centrale ($\triangle \overline{T}_{an} = 0$°C) et les extrémités sont suffisamment élevées pour susciter notre intérêt et nous pousser à approfondir notre diagnostic.

Nous pensons que la tendance moins prononcée pour *SimHyd* sur $MRC = f(\triangle \overline{T})$ résulte plutôt de problèmes de calage (et donc d'une erreur référente $\varepsilon = 1 - KGE_R[\theta_R]$ plus importante) que d'une réelle supériorité de robustesse par rapport à *GR4J* ou *Mordor6*. Nous noterons enfin une légère dissymétrie, des erreurs plus fortes étant commises lors d'un transfert d'une période froide vers une période chaude que l'inverse (*cf.* figure 3.2). Considérons un écart moyen de 1°C entre les périodes de calage et de validation et raisonnons en moyenne sur

3. Pour rappel, ces erreurs correspondent à des évaluations des simulations par le critère *KGE*, exprimé ici d'une manière relative à une valeur référente : $(1 - KGE_{validation})/(1 - KGE_{calage}) - 1$.

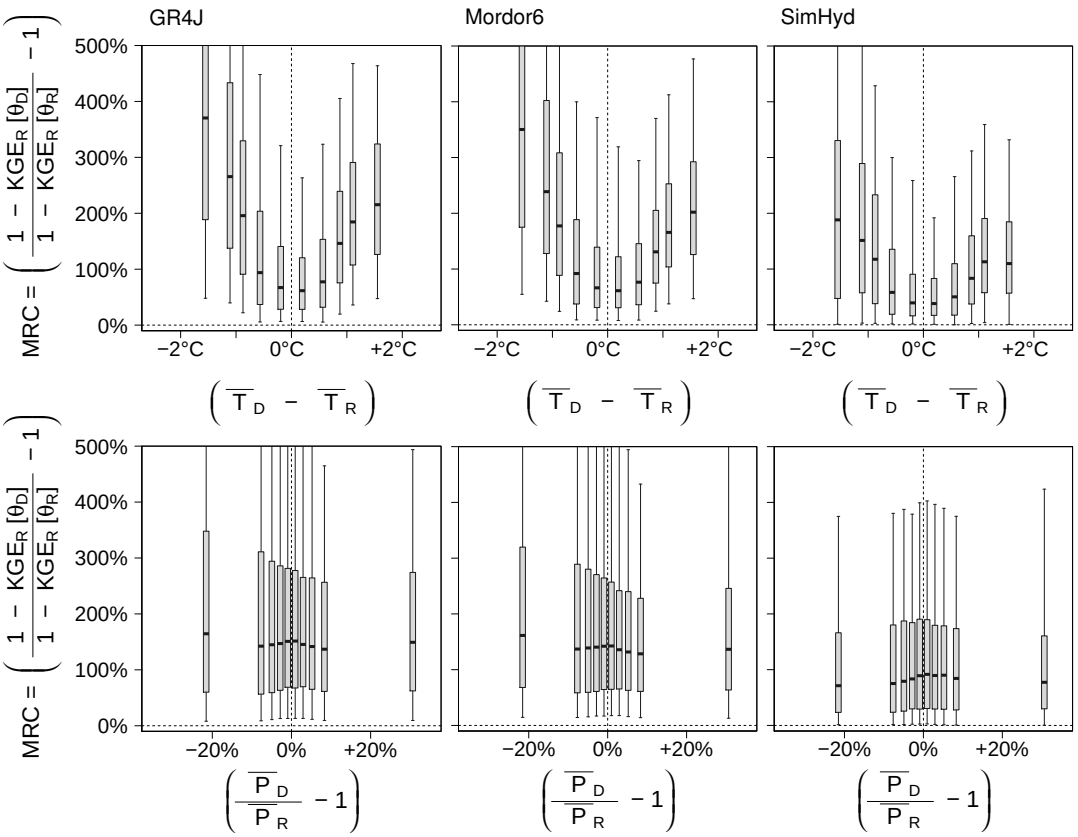


FIGURE 3.2.: Pertes de performances lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures

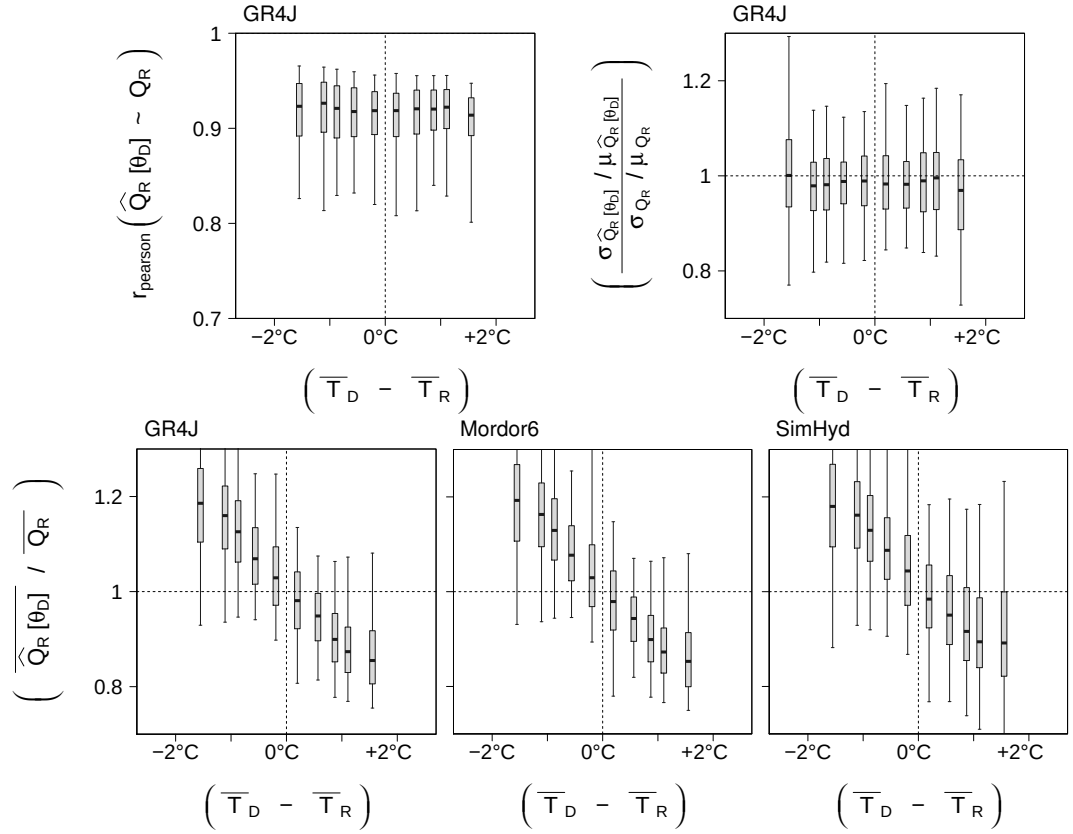


FIGURE 3.3.: Erreurs commises lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures

Rq. sur les bassins français, GR4J, Mordor6 et SimHyd sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

notre sous-groupe. Les erreurs attribuables au transfert des paramètres sont multipliées par 2.4, en moyenne sur le sous-groupe, lorsque la période de calage est plus chaude que la période de simulation $\left(\triangle\overline{T}_{D\to R} = +1°\text{C}\right)$, et multipliées par 3.2 lorsque la période de calage est plus froide que la période de simulation $\left(\triangle\overline{T}_{D\to R} = -1°\text{C}\right)$.

Les formes des nuages de points concernant les sous-composantes du KGE sont également très informatives (*cf.* figure 3.3). Si les corrélations de *Spearman* renvoient un signal sur les rapports de coefficients de variation (*cf.* le tableau 3.3), nous constatons ici que ces variations ne se concrétisent pas en pratique par des différences notables d'erreurs de simulation [4]. Nous retrouvons par contre les liens très nets entre erreurs de volume (biais) et écarts de température entre calage et validation, que laissait entrevoir la forte représentativité de cette corrélation au sein de l'échantillon. Nous pourrions imaginer une augmentation du biais absolu avec $\triangle\overline{T}$, sans que le signe n'importe. Dans un tel cas, nous aurions un nuage en forme de papillon où les points seraient resserrés autour de 1 lorsque $\triangle\overline{T} = 0$ puis se disperseraient à mesure que $\left|\triangle\overline{T}\right|$ augmente. Pourtant, cette situation n'est pas celle observée ici et le signe du biais se révèle fortement corrélé à celui du $\triangle\overline{T}$ sur ce sous-groupe de 120 bassins.

Pour reprendre un résultat déjà entrevu sur le tableau 3.1, nous constatons à nouveau ici qu'une période de calage plus froide que celle de validation (à gauche) est associée à une sur-estimation du débit moyen, tandis qu'une période de calage plus chaude que celle de validation (à droite) est associée à une sous-estimation du débit moyen (comparaisons ici réalisées avec les observations et calculées sur 10 ans). Lorsque cette analyse est répétée sur des quantiles de débits plutôt qu'une moyenne, ces tendances se retrouvent principalement sur les erreurs de volumes concernant les forts débits. Les erreurs demeurent très importantes pour les faibles débits mais ne semblent pas corrélées avec le $\triangle\overline{T}$ comme pour le reste de la gamme.

Cartographie

La figure 3.4 illustre sur une carte les variations de biais sur les 93 bassins français et 27 bassins australiens sélectionnés selon leur valeur de critère $r'_s(MRC\,,\,\triangle\overline{T})$. Le sens des triangles indique le signe des critères de *Spearman* obtenus (triangle vers le haut si $r_s(biais\,,\,\triangle\overline{T}) > +0.5$ et vers le bas si $r_s(biais\,,\,\triangle\overline{T}) < -0.5$) et leur taille reflète l'amplitude des pertes de performance. Toutefois, cette comparaison des amplitudes reste seulement indicative puisqu'elle repose sur une estimation de la pente de la régression linéaire entre biais et écart de température. [5] Les croix grises représentent les bassins de l'échantillon ne faisant pas partie des 120 sélectionnées. Les ronds colorés indiquent quant à eux les quelques rares bassins qui font partie des 120, mais où nous n'avons pas retrouvé de corrélation significative sur le biais (*c.-à-d.* $r'_s(MRC) > 0.5$ mais $|r_s(biais)| < 0.5$).

4. Il est probable que la combinaison de tendances de sens opposés conduise à un nuage diffus.

5. Pour faire ressortir aux mieux les différences, nous fixons une taille minimale et maximale de point et pratiquons une transformation logarithmique sur nos valeurs de pente telle que Taille-Triangle=log(Pente/PenteMax).

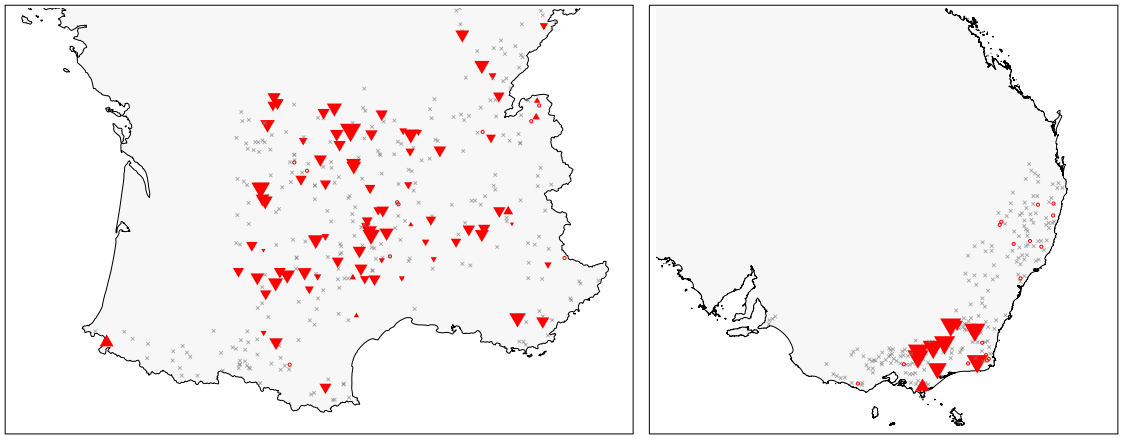


FIGURE 3.4.: Carte des 120 bassins concernés par des corrélations significatives entre erreurs sur les volumes moyens et △températures pour le modèle *GR4J*

Tous les bassins australiens concernés sont situés dans la même zone géographique (État du *New South Wales*), qui ne correspond cependant pas à une situation climatique clairement distincte des zones avoisinantes. En France, les bassins ne suivent pas de répartition spécifique, même si les bassins de haute montagne (Alpes et Pyrénées) semblent moins représentés. Quelques rares cas de corrélations significatives entre biais et $\triangle\overline{T}$ mais de sens opposés au signal moyen sont également visibles.

3.3.3. Illustration sur deux bassins

Afin d'illustrer nos résultats, prenons l'exemple des bassins de *l'Allier à Vieille-Brioude* et de *la Grande Creuse à Fresselines*. Ces bassins ont des superficies respectives de 2260 km^2 et 1230 km^2 et des altitudes médianes de 435 m et 215 m. Ces deux bassins présentent des régimes pluvio-nival avec des modules d'environ 29 m^3/s pour *l'Allier à Vieille-Brioude* et 16 m^3/s pour *la Grande Creuse à Fresselines*. Des séries continues d'observations de débit sont disponibles à leurs exutoires depuis 1919 pour *l'Allier à Vieille-Brioude* (station hydrométrique K2330810) et depuis 1956 pour *la Grande Creuse à Fresselines* (station hydrométrique L4220710). D'après la *Banque Hydro*, les débits de ces deux stations sont de bonne qualité sur toute la période disponible[6]. La longueur de ces séries de débit, combinée avec notre base *SPAZM* (disponible entre 1948 et 2010), nous permet de réaliser un nombre très important de tests de transfert entre périodes indépendantes. Ainsi, les chroniques de ces deux bassins versants ont été respectivement découpées en 52 et 45 sous-périodes de 10 ans et les modèles soumis à près de 1800 et 1250 *split-sample tests* (d'après la procédure *GSST*).

6. Seules les années 1980-1981 pour *l'Allier à Vieille-Brioude* et 1981-1982 pour *la Grande Creuse à Fresselines* présentent des débits validés mais qualifiés de douteux (*cf.* www.hydro.eaufrance.fr).

Les figures 3.5 et 3.6 illustrent la confrontation entre les pertes de performances constatées pour *GR4J* sur ces deux bassins et les écarts de température lors du passage calage-validation. Pour compléter cette comparaison, nous présentons à la figure 3.7 les résultats des tests de *Cequeau* sur *la Grande Creuse à Fresselines*.

Ces figures présentent les nuages d'erreurs relatives que nous pouvons construire par bassin à partir des très nombreux *split-sample tests* réalisés lors du *GSST*. Rappelons en effet que, derrière chaque point, se trouve la simulation d'une sous-période de 10 ans (R) à l'aide d'un jeu de paramètres (θ_D) calé sur une autre sous-période (D). Les deux graphiques inférieurs de ces figures nous le remémorent puisqu'ils fournissent les régimes simulés sur deux périodes *receveuses* de ces bassins. Les différentes courbes représentent les variantes obtenues selon le jeu de paramètres utilisé et sont colorisées selon l'écart de température avec la période simulée. Nous retrouvons ici encore la corrélation entre une sous-estimation des débits et une période de calage plus chaude que la période simulée (orange-rouge), tandis que les cas de surestimations sont rencontrés lorsque la période de calage est plus froide (bleu clair-foncé).

Nous retrouvons clairement une augmentation des erreurs relatives lorsque $\left|\triangle\overline{T}_{D\to R}\right|$ augmente, mais aucun signal en lien avec $\left|\triangle\overline{P}_{D\to R}\right|$ pour ces bassins. La comparaison des figures 3.6 et 3.7 montre des corrélations moins prononcées pour *Cequeau* que pour *GR4J*, mais qui s'expriment plutôt par une plus forte dispersion des performances que par une amélioration globale de celles-ci. Par ailleurs, la figure 3.5 (particulièrement le graphique concernant les erreurs de volume) illustre le fait que des nuages de points ni aléatoires ni clairement organisés peuvent être rencontrés. Ils révèlent que si une majorité des augmentations d'erreurs sont corrélées aux $\triangle\overline{T}$, certaines de leurs variations peuvent être liés à d'autres phénomènes et engendrer des nuages parfois atypiques. Bien que nous ayons choisi des cas *favorables* pour nos illustrations (car ils facilitent la compréhension), gardons toujours à l'esprit que ces formes de nuage où une corrélation globale est visible, sans pour autant expliquer la totalité des variations d'erreurs, constituent plutôt la normalité que l'exception parmi de nos résultats.

À titre indicatif, nous avons également représenté ici les nuages de points obtenus pour une analyse en fonction du seul décalage temporel entre périodes.[7] Trancher sur des liens de causalité entre erreurs et écarts de température, ou erreurs et écarts temporels est souvent une tâche délicate, compte tenu de la tendance continuelle à l'augmentation des températures avec le temps (en particulier sur les bassins français). Cette problématique fera l'objet d'une section dédiée dans le chapitre 5 (*cf.* section 5.4) et nous nous contenterons ici de noter la plus grande netteté de corrélation avec les écarts de température qu'avec les écarts temporels. Pour finir, les formes des nuages concernant les composantes ρ, σ/μ et β illustrent très clairement que ces erreurs sont majoritairement sous forme de biais moyens sur les volumes.

7. L'absence de point dans la partie centrale résulte de l'interdiction d'un chevauchement entre périodes, qui assure l'indépendance entre calage et validation.

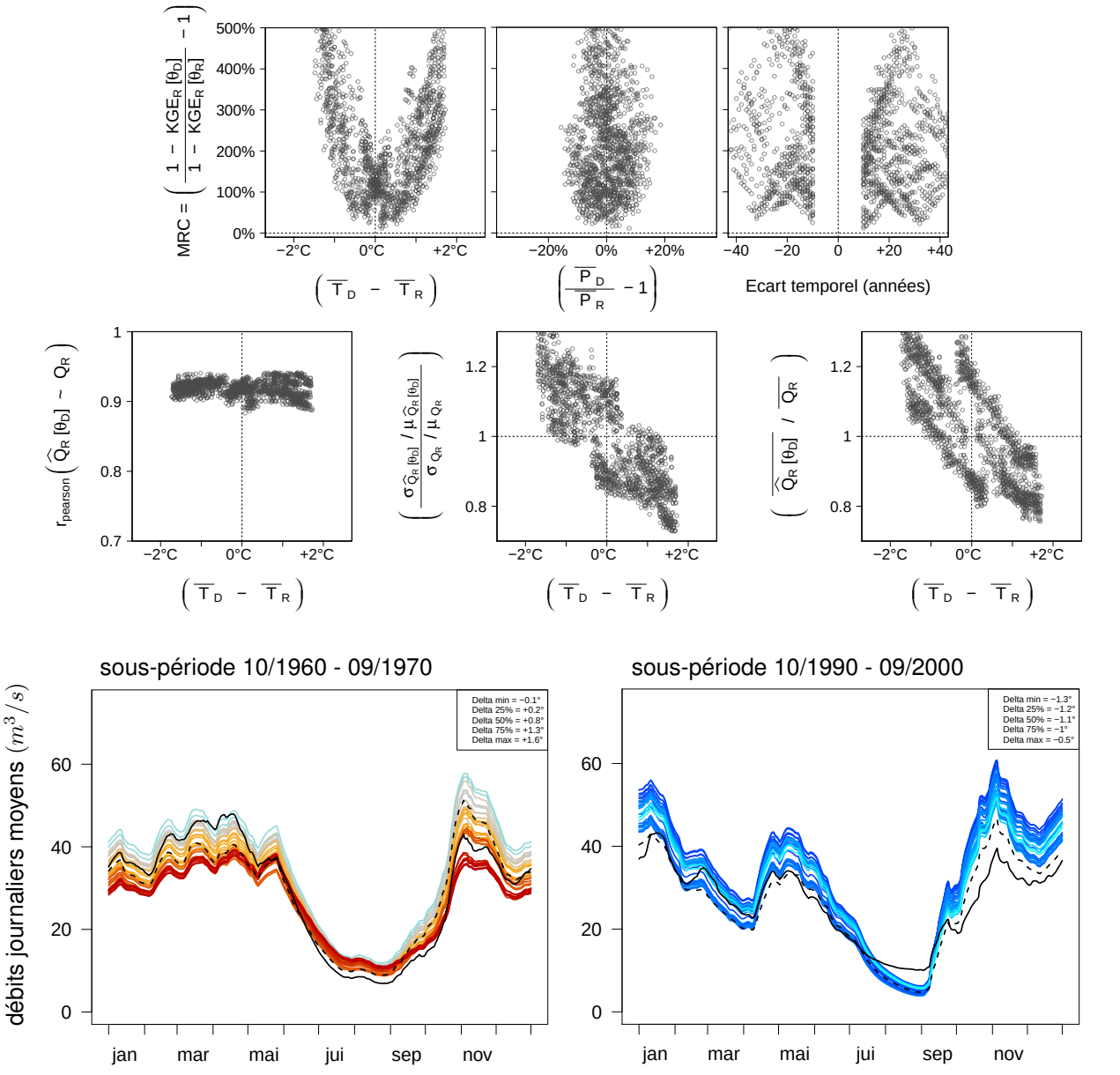


FIGURE 3.5.: Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *l'Allier à Vieille-Brioude*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (bleu foncé -1.5°C, gris 0°C, rouge +1.5°C).

Rq. GR4J est alimenté ici par des données SPAZM et une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} maintenus fixés).

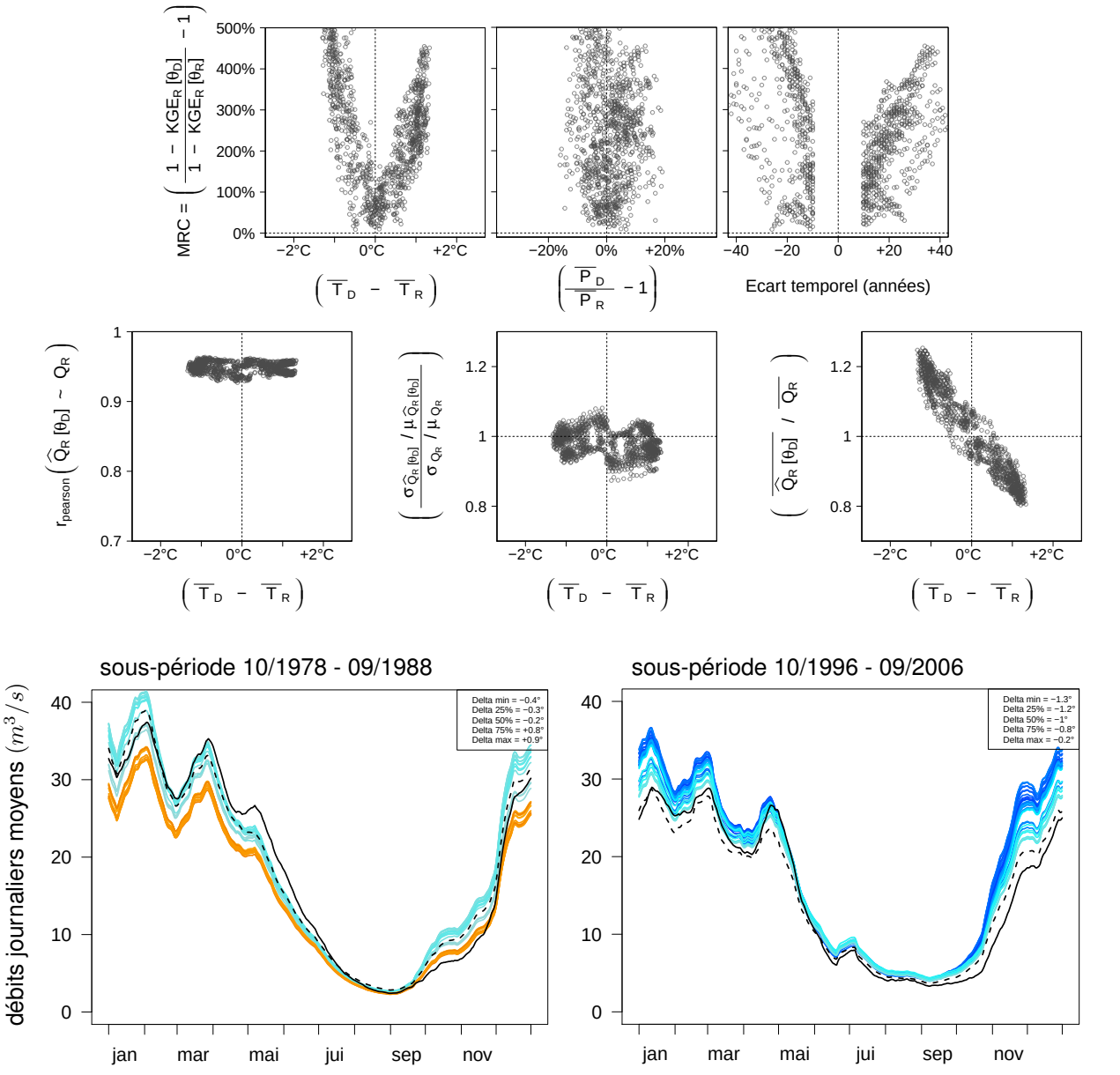


FIGURE 3.6.: Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Grande Creuse à Fresselines*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (bleu foncé -1.5°C, gris 0°C, rouge +1.5°C).

Rq. GR4J est alimenté ici par des données SPAZM et une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} maintenus fixés).

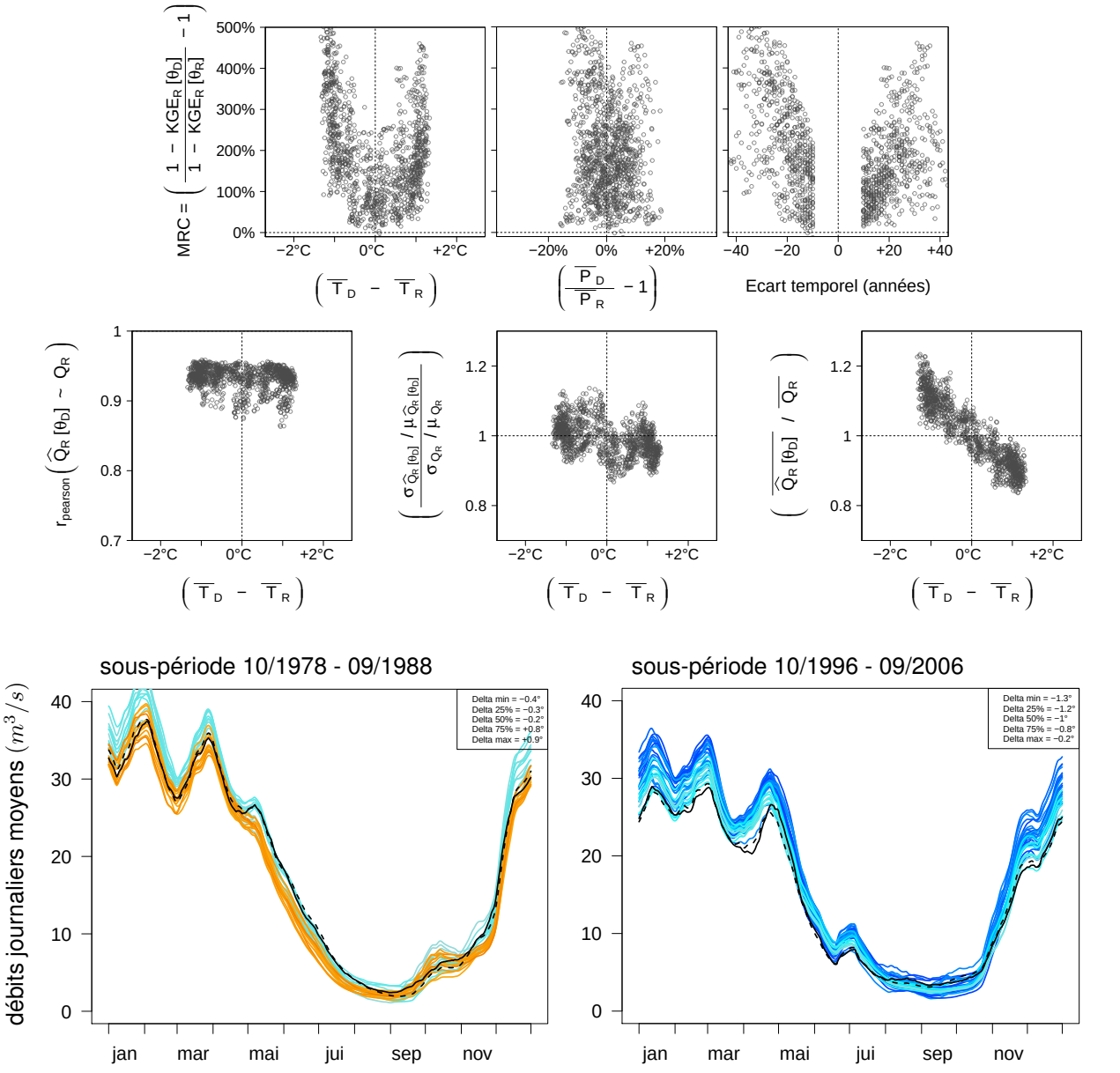


FIGURE 3.7.: Aperçu des erreurs commises par *Cequeau* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Grande Creuse à Fresselines*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (bleu foncé -1.5°C, gris 0°C, rouge +1.5°C).

Rq. Cequeau est alimenté ici par des données SPAZM, il cale un ETP de Thornthwaite et dispose de son propre module neige.

3.4. Zoom sur les cas de problèmes de robustesse en lien avec △précipitations

3.4.1. Tableau de corrélation

Utilisant la même démarche que celle menée pour les △*températures*, nous allons étudier ici les cas de manque de robustesse des modèles en lien avec les △*précipitations*. Le tableau 3.4 reprend les taux de représentativité des corrélations entre erreurs et écarts de précipitations, et précise les taux constatés sur les sous-échantillons français et australiens pris séparément.

Tableau 3.4.: Taux de corrélations significatives entre △erreurs et $\triangle\overline{P}$ pour *GR4J*

Sur chaque cellule, résultats sur tout l'échantillon (haut) sur l'échantillon français (bas-gauche) et sur l'échantillon australien (bas-droite). Données et couleurs identiques à celles du tableau 3.1. Rq. en France, GR4J est alimenté par SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ fixés).

	MRC	$r_p\left(\widehat{Q} \sim Q\right)$		$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_S>+0.5$	$r'_S<-0.5$	$r_S<-0.5$	$r_S>+0.5$	$r_S<-0.5$	$r_S>+0.5$
$\overline{P}$	8% 5% 13%	3% 1% 5%	28% 23% 36%	9% 9% 9%	20% 24% 12%	22% 13% 35%
$\overline{P}_{DJF}$	5% 4% 6%	2% 2% 2%	29% 35% 20%	12% 11% 14%	12% 11% 11%	31% 35% 25%
$\overline{P}_{MAM}$	6% 4% 11%	3% 1% 7%	31% 25% 41%	12% 14% 10%	11% 10% 11%	37% 31% 46%
$\overline{P}_{JJA}$	7% 6% 8%	3% 2% 5%	25% 17% 37%	18% 22% 10%	20% 23% 14%	25% 18% 34%
$\overline{P}_{SON}$	8% 10% 5%	1% 1% 0%	18% 20% 15%	24% 27% 18%	45% 56% 24%	11% 6% 19%

Il existe plusieurs différences notables entre la situation que nous étudions ici ($\triangle erreurs = f(\triangle\overline{P})$) et la précédente concernant les liens entre robustesse et $\triangle\overline{T}$. Tout d'abord, les taux de représentativité rencontrés ici ne sont pas du même ordre que ceux obtenus sur les variables de température. Au total, seuls 8% de l'échantillon complet sont concernés par un $r'_s(MRC) > 0.5$, avec une occurrence plus importante sur l'échantillon australien (13%) que français (5%). La situation est inversée par rapport aux analyses précédentes puisque les corrélations entre erreurs de modélisation et △*températures* sont plus courantes sur l'échantillon français.

Une autre différence majeure réside dans l'hétérogénéité des résultats obtenus ici. Nos analyses des relations entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{T}$ indiquent des critères de *Spearman* très largement de même signe (un signe négatif révélateur d'une anti-corrélation, *cf.* tableau 3.3, page 75). Or ce n'est pas le cas ici. En effet, des taux de représentativité parfois équivalents sont observés pour des corrélations de signes opposés. Cette situation est notamment rencontrée pour la relation entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{P}$. Nous notons d'ailleurs un lien entre le signe de ces corrélations et l'appartenance géographique des bassins, les bassins français présentant plutôt des anti-corrélations entre biais et $\triangle\overline{P}$ et les bassins australiens plutôt des corrélations (*cf.* tableau 3.4). Malgré nos tentatives, nous ne sommes pas en mesure d'expliquer ces différences.

3.4.2. Variations des erreurs sur un sous-groupe choisi

Composition du sous-groupe

Nous avons constitué un sous-groupe de bassins pour étudier les problèmes de robustesse des modèles en lien avec les écarts de précipitations entre calage et validation. La procédure demeure identique : un seuil de significativité choisi à 0.5 et une sélection des bassins présentant un lien de corrélation entre l'augmentation des erreurs de *GR4J*, estimée via le *MRC*, et les écarts de précipitation moyenne annuelle. Cela correspond à 48 bassins (8% de l'échantillon complet) dont 19 français et 29 australiens. Un aperçu des caractéristiques des bassins de ce sous-groupe est disponible en annexe (*cf.* figure D.2, page 280).

Analyse des nuages de points $\triangle erreurs = f(\triangle \overline{P})$

L'analyse visuelle des nuages de points pour ces 48 bassins est menée à travers les figures 3.8, 3.9 et 3.10. La première illustre les variations du critère *MRC* sur ce sous-groupe de bassins face aux écarts de température et précipitation moyennes entre les périodes de calage et validation pour chaque test. Les bassins ayant été *choisis* pour cela, nous retrouvons un lien de corrélation entre erreurs et $\triangle \overline{P}$. En moyenne sur ce sous-groupe, une différence de 10% sur les précipitations subie lors du transfert des paramètres $\left(\triangle \overline{P}_{D \to R} = \pm 10\%\right)$ est associée à une multiplication par un facteur 2-3 des erreurs attribuables à ce transfert.

Nous noterons que, dans le même temps, aucun lien n'est visible avec les $\triangle \overline{T}$ sur ce sous-groupe de bassins. La hauteur plus importante des boîtes au niveau du $\triangle \overline{T} = 0$ résulte des nuages de points des bassins australiens où les données de températures sont parfois très homogènes temporellement (comme constaté sur le bassin exemple à la figure 3.12).

L'analyse des composantes d'erreurs retenues pour nos travaux montre que ces pertes de performances se retrouvent à différents niveaux (*cf.* figure 3.9). Nous distinguons par exemple une légère diminution du critère de corrélation linéaire $\widehat{Q} \sim Q$ lorsque l'écart de précipitations augmente entre les périodes de calage et validation. La variance (σ/μ) semble également affectée (légère tendance négative), mais les analyses que nous avons menées bassin par bassin n'ont pas révélé l'existence d'un réel signal. Nous observons enfin un nuage de forme particulière pour les relations entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle \overline{P}$, avec des boîtes dont la hauteur augmente significativement lorsque l'écart de précipitations devient grand. Cette forme résulte de la présence simultanée, au sein de nos 48 bassins, de situations de corrélation et d'anti-corrélation, déjà visibles sur le tableau 3.4 (différence entre France et Australie). Par conséquent, nous allons scinder notre sous-groupe en deux parties pour mieux visualiser ces situations.

La figure 3.10 présente à nouveau les liens entre les biais des modèles sur les volumes et $\triangle \overline{P}$ sur nos 48 bassins, mais en distinguant cette fois-ci les cas de corrélation et d'anti-corrélation. Les

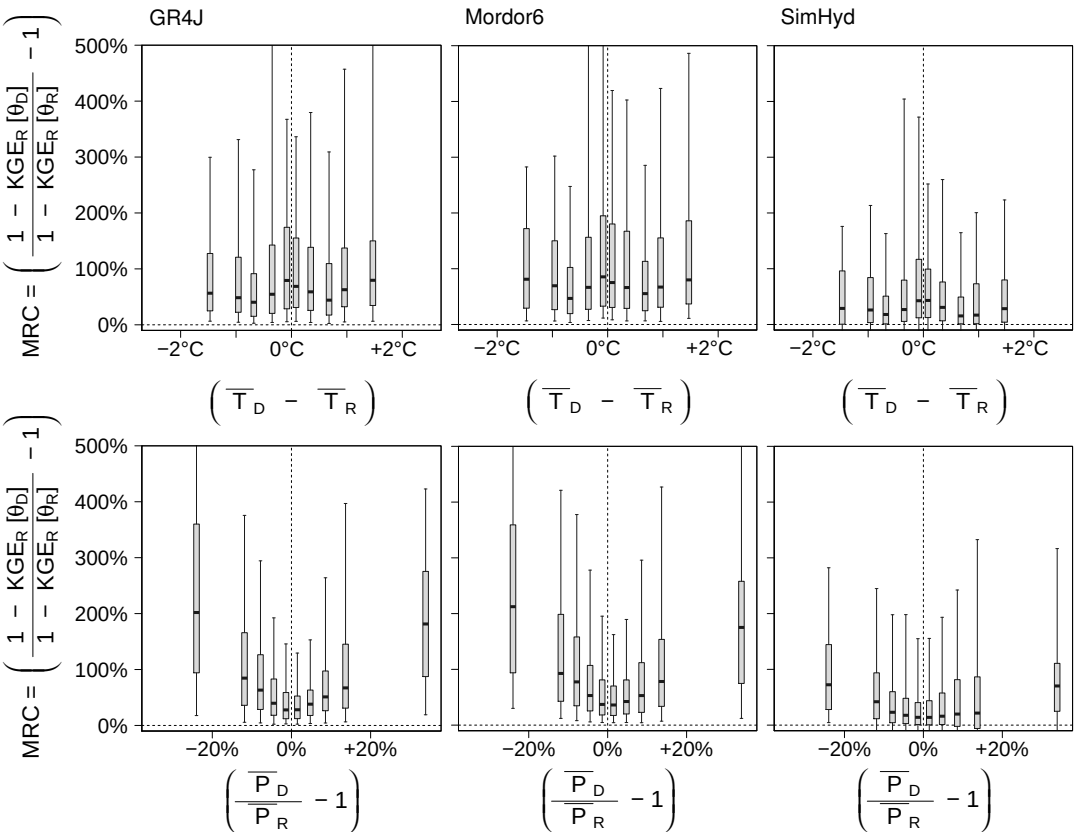


FIGURE 3.8.: Pertes de performances lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations

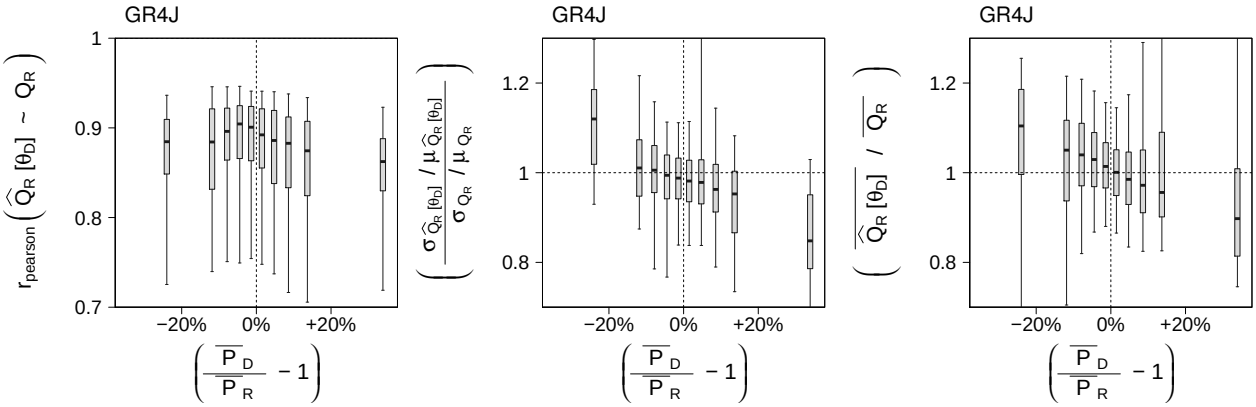


FIGURE 3.9.: Erreurs commises lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations pour GR4J

Rq. sur les bassins français, GR4J, Mordor6 et SimHyd sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

cas de corrélation concernent 32 bassins (graphiques du haut) et les cas d'anti-corrélation en concernent 14 (graphiques du bas), aucun lien significatif n'ayant été trouvé pour deux bassins ($|r_s(biais)| < 0.5$). Cette figure révèle deux situations très contrastées dans lesquelles seuls les signes de tendances sont opposés mais où les formes de nuage restent proches (les pentes sont toutefois plus prononcées dans les cas de corrélation). Ici encore, malgré nos tentatives, nous ne sommes pas en mesure d'expliquer ces différences.

Avant de poursuivre avec d'autres descriptions, notons que deux critères de sélection ont été croisés pour construire la figure 3.10 :

$r'_s(MRC\,,\,\triangle\overline{P}) > 0.5 \quad \& \quad r_s(biais\,,\,\triangle\overline{P}) > +0.5 \quad$ d'une part,
$r'_s(MRC\,,\,\triangle\overline{P}) > 0.5 \quad \& \quad r_s(biais\,,\,\triangle\overline{P}) < -0.5 \quad$ d'autre part.

Les sous-groupes ainsi constitués demeurent relativement réduits. Toutefois, nous aurions pu conduire les mêmes analyses en sélectionnant nos bassins directement sur les corrélations entre biais et écarts de précipitations (*c.-à-d.* $\left|r_s(biais\,,\,\triangle\overline{P})\right| > 0.5$, sans contrainte sur le $r'_s(MRC\,,\,\triangle\overline{P})$). Le tableau 3.4 permet d'ailleurs de constater leur taux élevé de représentativité sur notre échantillon, avec environ 120 bassins dans chaque situation (corrélation ou anti-corrélation). Cet autre exercice d'analyse a été réalisé et a conduit à l'obtention des nuages de points très proches de ceux présentés sur la figure 3.10, aussi bien en termes de pente que de dispersion verticale.

Nous noterons enfin que si les tendances entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{P}$ sont visibles pour *SimHyd* tout autant que pour *GR4J* ou *Mordor6*, ce modèle présente parfois des difficultés à ajuster le bilan hydrologique lors de la phase de calage. Ces difficultés sont à l'origine de la translation vers le haut du nuage de points placés en bas à droite sur la figure 3.10.

Cartographie

La figure 3.11 illustre sur une carte les variations de biais sur les 19 bassins français et 29 bassins australiens qui composent le groupe des 48 bassins sélectionnés selon leur $r'_s(MRC\,,\,\triangle\overline{P})$. Le sens des triangles indique le signe des critères de *Spearman* obtenus (triangle vers le haut si $r_s(biais\,,\,\triangle\overline{P}) > +0.5$ et vers le bas si $r_s(biais\,,\,\triangle\overline{P}) < -0.5$) et leur taille reflète l'amplitude des pertes de performances. [8] Les bassins australiens sont plus nombreux que les bassins français et présentent tous (à une exception près) des liens de corrélations entre erreurs de volume et $\triangle\overline{P}$. Les cas d'anti-corrélation sont quant à eux, majoritaires sur les bassins français inclus dans cette sélection, malgré quelques cas de corrélations significatives.

8. Les détails de cette représentation sont disponible page 79, dans les commentaires de la figure 3.4.

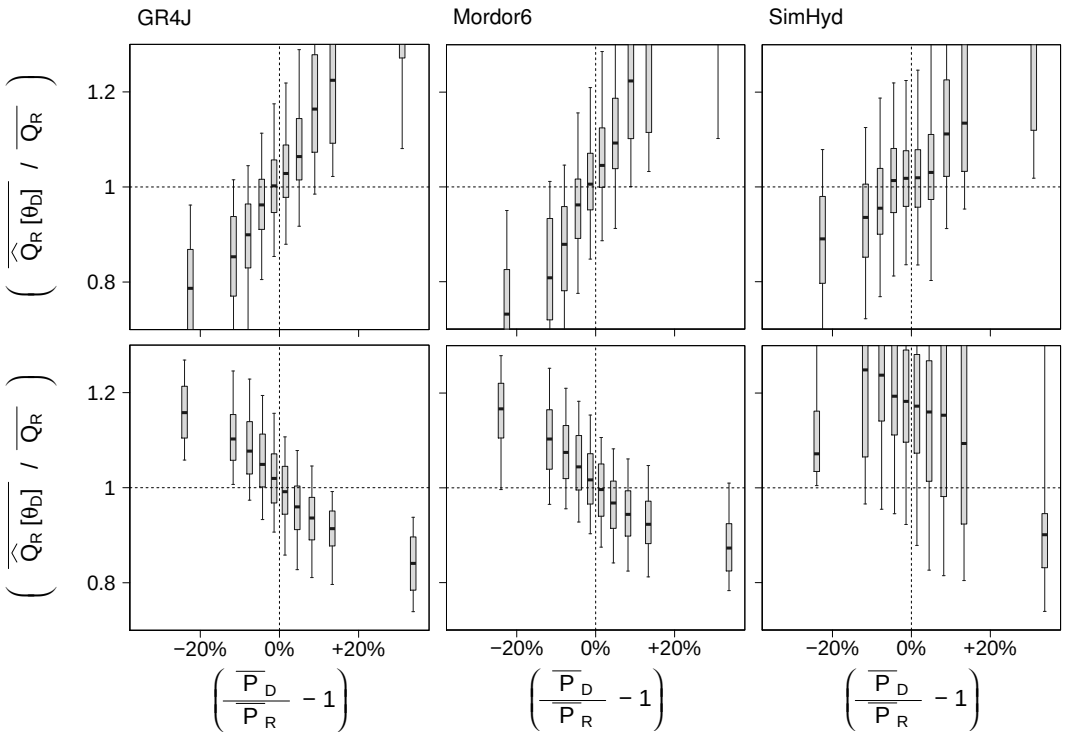


Figure 3.10.: Erreurs commises lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations

32 bassins sont concernés par des corrélations (graphiques du haut)
14 bassins sont concernés par des anti-corrélations (graphiques du bas)

Rq. sur les bassins français, GR4J, Mordor6 et SimHyd sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

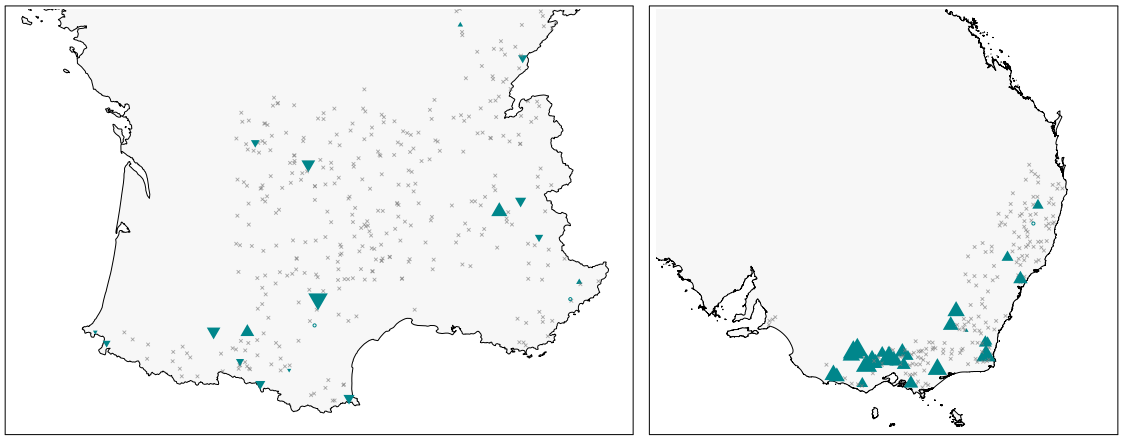


Figure 3.11.: Carte des 48 bassins concernés par des corrélations significatives entre erreurs sur les volumes moyens et △précipitations pour le modèle *GR4J*

3.4.3. Illustration sur deux bassins

Afin d'illustrer nos résultats, prenons l'exemple des bassins de *la Wimmera river à Glenorchy (concrete weir)* et de *la Durance à Embrun (La Clapière).* Mise à part des superficies similaires de 2000 km² et 2300 km², ces bassins diffèrent sur de nombreux points. Le bassin de *la Wimmera river à Glenorchy* (station 415201) est situé en plaine (200 m d'altitude médiane) avec un régime pluvial d'écoulement (les hautes eaux ont lieu lors des périodes de précipitations plus abondantes, lors de l'hiver et du printemps de l'hémisphère sud). À l'inverse, le bassin de *la Durance à Embrun* (station X0310010) est situé en haute montagne (altitude médiane de 2200 m) et les écoulements y sont fortement influencés par la neige (les hautes eaux ont lieu lors de la période de fonte, en début d'été). Cette différence, combinée à des contextes climatiques régionaux très différents, explique la forte différence d'aridité entre ces bassins avec plus de 50% des précipitations totales transformées en écoulements (rapport $\overline{Q}/\overline{P}$) pour *la Durance à Embrun* contre moins de 10% pour *la Wimmera river à Glenorchy.*

Les figures 3.12 et 3.13 illustrent la confrontation entre les pertes de performances obtenues avec *GR4J* sur ces deux bassins et les écarts de précipitations lors du passage calage-validation. Pour compléter cette comparaison, nous présentons à la figure 3.14 les résultats des tests de *Mordor* sur *la Durance à Embrun* (il s'agit bien ici du modèle *Mordor* et non de *Mordor6*, sa version simplifiée que nous utilisons largement par ailleurs) [9]. Précisons enfin que la figure 3.13 correspond à un couple *GR4J-CemaNeige* dont tous les paramètres sont calés (par opposition à l'utilisation générale où nous avons choisi d'utiliser des paramètres fixes pour *CemaNeige*, *cf.* discussion à la section 3.1.3).

Nous retrouvons une augmentation des erreurs relatives lorsque $\left|\triangle\overline{P}_{D\rightarrow R}\right|$ augmente et une organisation des courbes de régime, en lien avec les conditions climatiques impliquées dans les transferts des jeux de paramètres. Néanmoins, les amplitudes et la netteté des corrélations demeurent inférieures à celles observées lors de nos analyses des $\triangle erreurs = f(\triangle\overline{T})$ sur d'autres bassins versants (*cf.* section 3.3).

9. Je remercie d'ailleurs Thibault Mathevet qui m'a fourni les calages de *Mordor* permettant ces analyses et m'a gentiment invité à plusieurs reprises pour que je présente mes travaux à *EDF-DTG* (Grenoble).

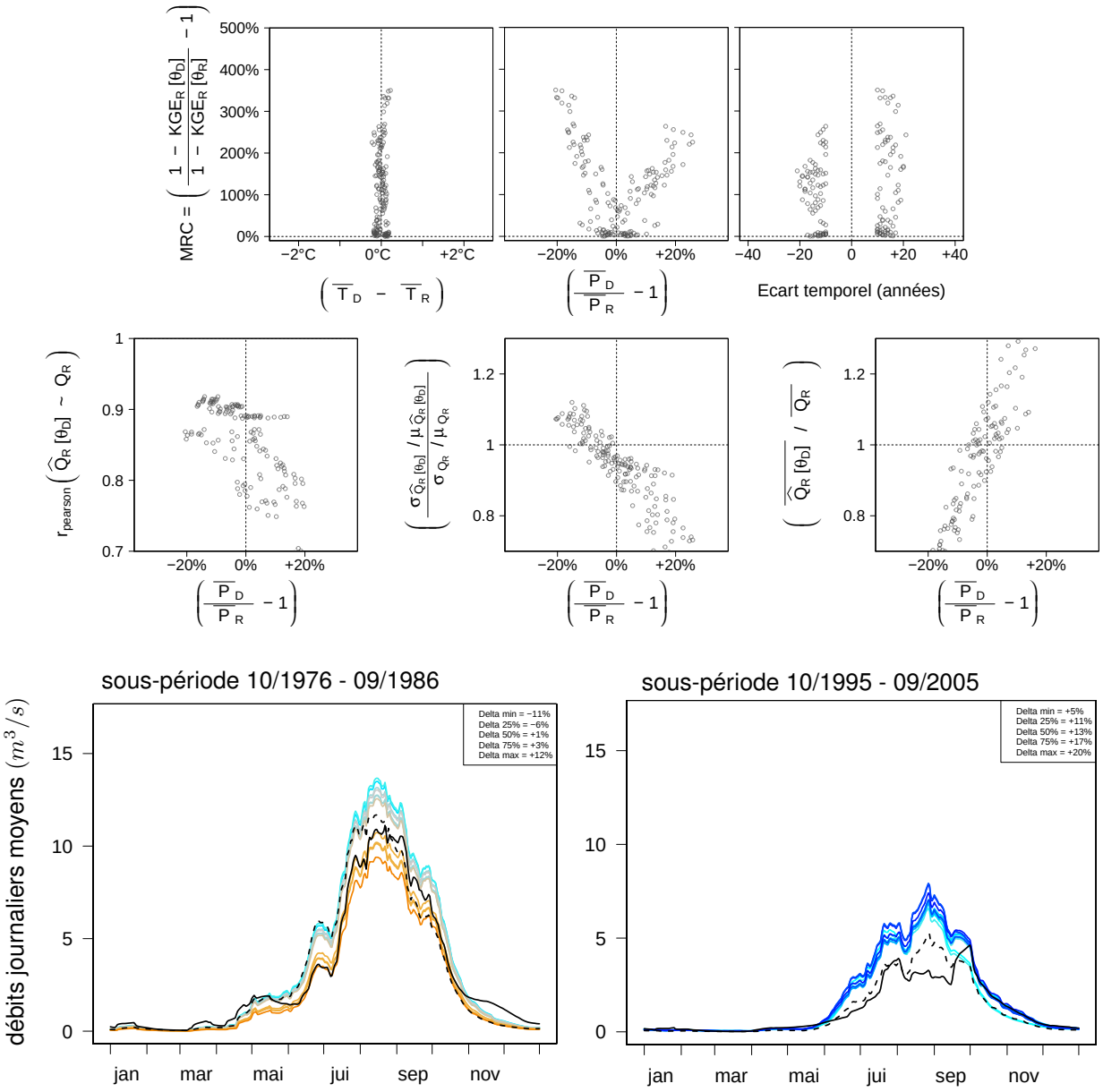


Figure 3.12.: Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Wimmera river à Glenorchy*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (rouge -20%, gris 0%, bleu foncé +20%).

Rq. GR4J est alimenté ici par des données CSIRO.

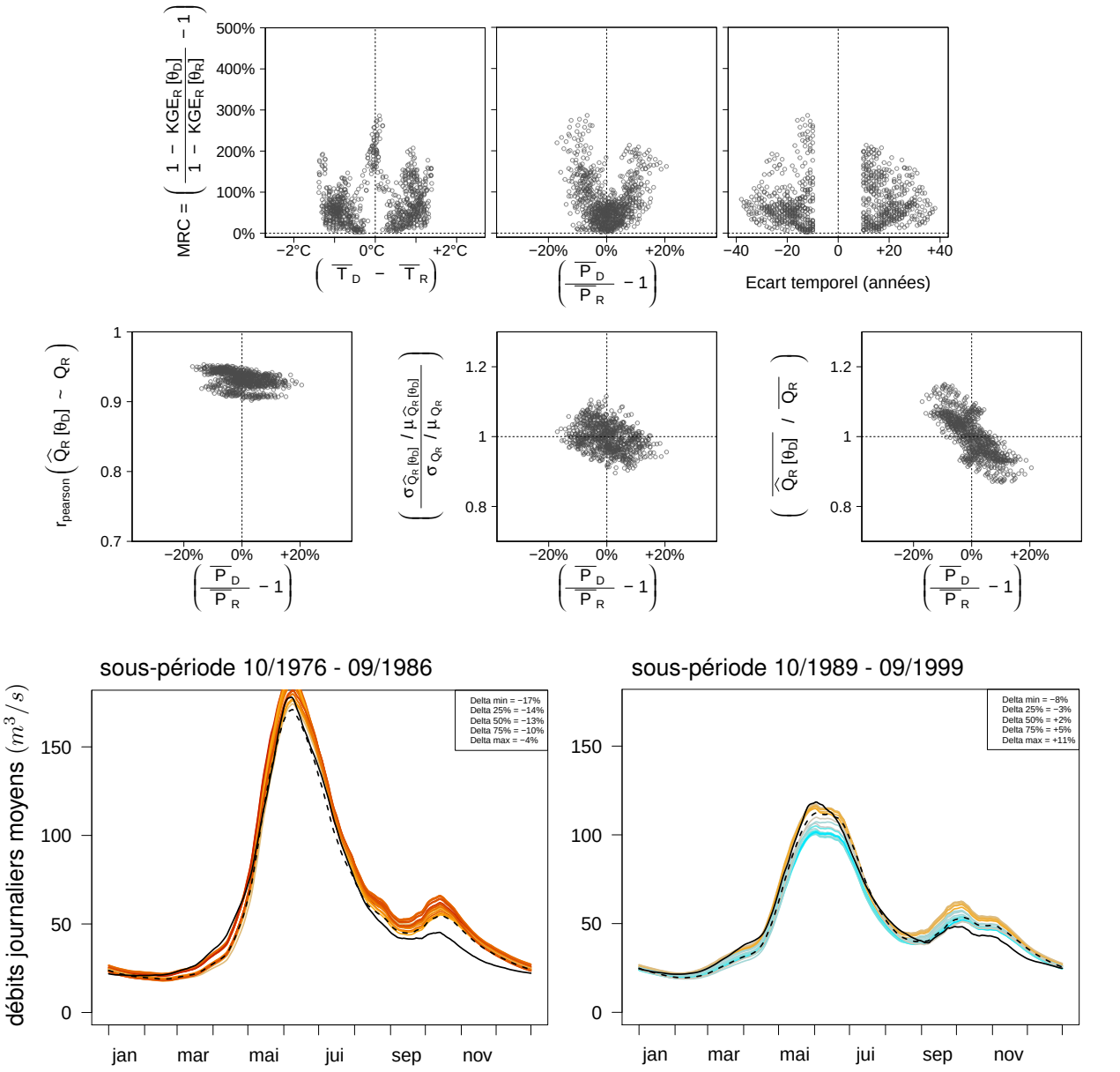


FIGURE 3.13.: Aperçu des erreurs commises par *GR4J* (*CemaNeige* calé) lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Durance à Embrun*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (rouge -20%, gris 0%, bleu foncé +20%).

Rq. GR4J est alimenté ici par des données SPAZM et une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} calés).

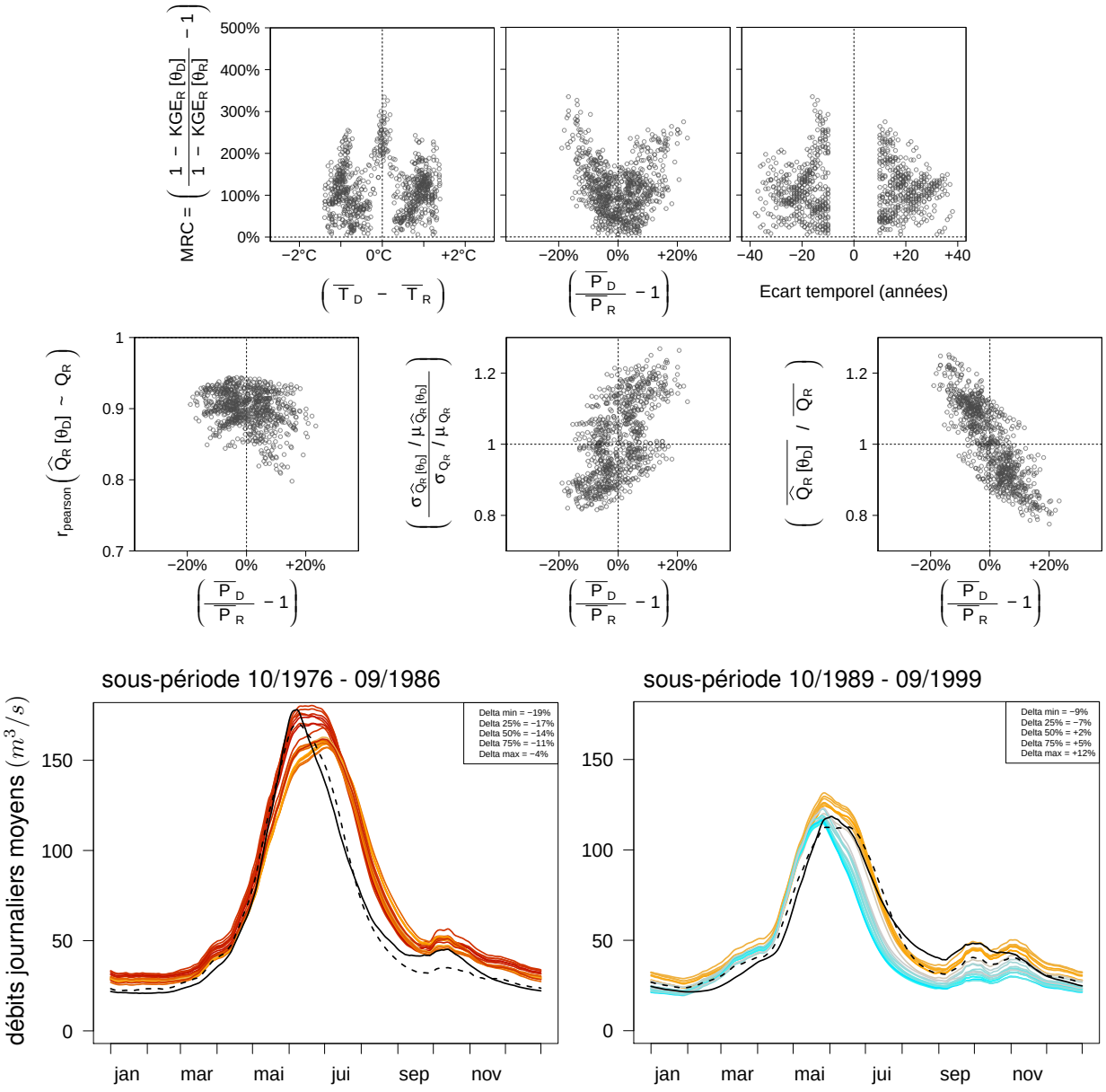


FIGURE 3.14.: Aperçu des erreurs commises par *Mordor* lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Durance à Embrun*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (rouge -20%, gris 0%, bleu foncé +20%)

Rq. Mordor est alimenté ici par des données SPAZM. Il cale son ETP et dispose de son propre module neige.

3.5. Zoom sur la modélisation des bassins nivaux

3.5.1. Introduction

Notre échantillon de bassins français est relativement hétérogène en termes de répartition sur les plaines et reliefs et par conséquent, en termes de conditions climatiques (*cf.* figure 2.7, page 45). Certains bassins présentent notamment des régimes nivaux où la répartition temporelle des écoulements est fortement contrôlée par les processus d'accumulation et d'ablation de neige. L'hydrologie de ces bassins est particulièrement sensible aux variabilités des précipitations et températures qui impactent conjointement les volumes d'eau stockés sous forme solide et le déphasage temporel de leur transformation en écoulements. Pourtant, les nombreux tests que nous avons menés ne font pas ressortir des différences de comportement sur ce type de bassins en particulier. Il est vrai que la plupart des résultats présentés dans les sections 3.2, 3.3 et 3.4 ont été obtenus avec un module neige (*CemaNeige*) soit inactivé, soit utilisé avec des valeurs de paramètres médianes (*c.-à-d.* non calées). Cependant, des tests ont été menés avec *GR4J* en calant systématiquement *CemaNeige* (*cf.* tableau 2.5, page 55). Ils n'ont pas conduit à des interprétations différentes de celles présentées jusqu'ici.

Cette section a pour objectif de conduire une étude spécifique de robustesse temporelle des modèles sur les bassins nivaux, en s'intéressant notamment aux écarts de modélisation de la neige. Nous avons pour cela constitué un groupe de bassins versants, sélectionnés selon la part que représentent les apports solides au sein des précipitations annuelles. Tous les résultats discutés ici reposent sur des tests où *CemaNeige* est calé (à la différence des situations précédentes où des paramètres fixes étaient utilisés pour simplifier les interprétations).

3.5.2. Variations des erreurs sur un sous-groupe choisi

Le sous-groupe de bassins constitué pour réaliser notre zoom comporte 41 bassins, pour lesquels la part annuelle des précipitations solides est estimée supérieure à 20%. Nous insisterons sur le fait qu'il s'agit bien ici d'une estimation, ces 20% d'apports solides n'étant pas mesurés mais issus d'une fonction simplifiée de différentiation pluie/neige (fonction identique à celle présente dans *CemaNeige*, *cf.* section 2.3.4). Ces 41 bassins, généralement situés à haute altitude, sont répartis en France entre les massifs alpin et pyrénéen (*cf.* figure 3.15).

Analyse des nuages de points $\triangle erreurs = f(\triangle caract.)$

De nombreuses simulations sont réalisées à travers notre procédure de test. Les sous-périodes jouent à la fois les rôles de *donneuses* (D) et *receveuses* (R) de paramètres. À l'aide des chroniques de débits pour diverses conditions de calage ($R \Rightarrow R$) et validation ($D \Rightarrow R$), nous

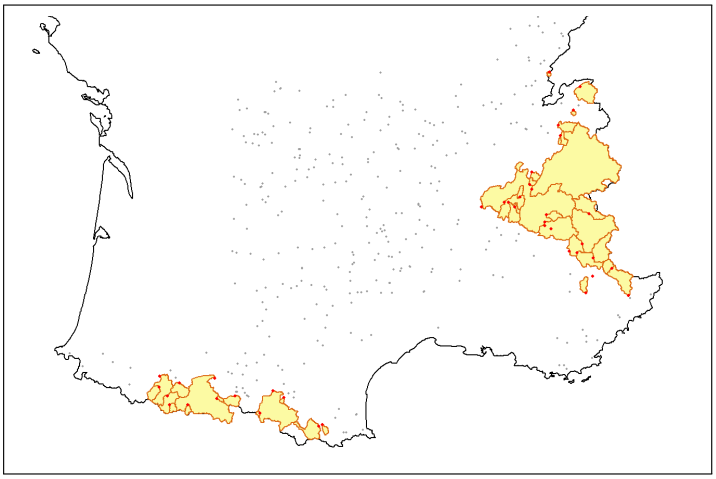


Figure 3.15.: Localisation des 41 bassins où les précipitations solides (estimées) représentent plus de 20% des apports annuels

construisons à nouveau nos graphiques d'analyses $\triangle erreurs = f(\triangle caract.)$. Ceux-ci sont présentés dans la figure 3.16 pour le modèle *GR4J* couplé à *CemaNeige* (dont les paramètres sont maintenant calés).

Les positions et faibles hauteurs des boîtes sur les graphiques concernant le *MRC* nous indiquent des pertes de performances en KGE faibles sur ces 41 bassins versants. Ces pertes sont peu influencées par les conditions de transfert des paramètres du modèle. Les amplitudes observées ici sont en effet, sans comparaison avec celles rencontrées dans les sections 3.3 et 3.4. Des tendances sont toutefois visibles sur les biais sur les volumes moyens ($\overline{\widehat{Q}}/\overline{Q}$) lorsqu'ils sont étudiés au regard des $\triangle\overline{P}$, ainsi que sur la composante de variabilité (σ/μ) qui semble conjointement affectée par des écarts de température et de précipitations moyennes. L'utilisation de variables saisonnières plutôt qu'annuelles dans ces analyses ne met pas en évidence des situations différentes de celles présentées ici, même si les variations de précipitations hivernales et de température printanière renvoient des signaux plus nets que les autres saisons.

Analyse des nuages de points $\triangle neige_{sim.} = f(\triangle\overline{T})$

Parallèlement aux simulations de débits, nous disposons également de simulations de stocks de neige, calculées ici par *CemaNeige* qui modélise les processus d'accumulation et d'ablation de neige de façon indépendante sur cinq bandes d'altitude de surfaces égales. La confrontation avec des grandeurs observées n'est malheureusement pas possible ici, même lorsque des séries de mesure de neige sont disponibles. En effet, les mesures de terrain et les stocks simulés par un module conceptuel restent difficilement comparables. À défaut, nous pouvons tout de même comparer nos séries à une simulation de référence (celle obtenue en calage), comme nous pouvons le faire pour les débits. Dans la lignée des analyses précédentes, ces comparaisons sont donc menées au regard des contrastes de conditions climatiques entre les périodes de calage et de validation.

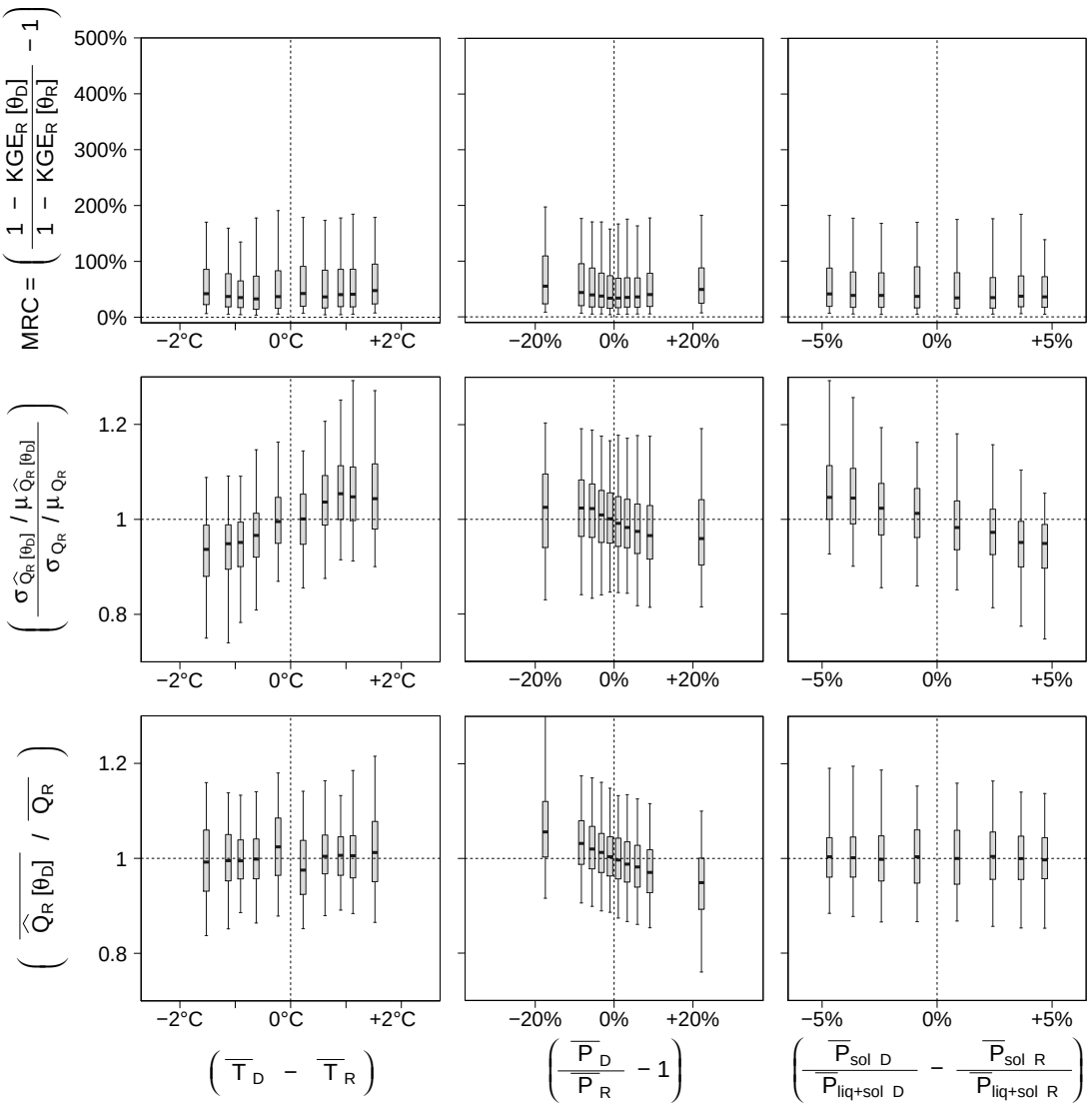


FIGURE 3.16.: Erreurs commises par *GR4J-CemaNeige* lors du *GSST* sur les 41 bassins recevant plus de 20% de précipitations annuelles sous forme solide

GR4J-CemaNeige (θ_{neige} calés) est alimenté ici par des données SPAZM et une ETP d'Oudin. Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Le choix des critères utilisés pour caractériser une série continue de stock de neige n'est pas aisé. Nous avons opté pour deux aspects traduisant de manière relativement simple les écarts volumétriques et temporels entre nos simulations. Après avoir aggloméré les séries obtenues sur les différentes bandes d'altitudes, nous estimons d'une part un stock annuel moyen de neige simulé, pour chaque utilisation du modèle sur une période (moyenne arithmétique de tous les pas de temps) et d'autre part une date de fin de fonte (correspondant au moment où le manteau neigeux simulé a complètement disparu) [10]. Les variations constatées sur ces deux aspects en fonction des conditions de transfert calage-validation sont illustrées à la figure 3.17.

Nous constatons alors la relation suivante : si la période de calage est plus chaude que la période simulée (*c.-à-d.* avec moins de neige), alors un stock de neige plus important sera

10. En pratique, nous évaluons la date moyenne à laquelle le stock de neige du bassin passe sous le seuil des 1 mm dans la période printemps-été.

produit par le module *CemaNeige* ; symétriquement, si la période de calage est plus froide que la période simulée (*c.-à-d.* avec plus de neige), alors un stock de neige moins important sera produit par le module (ce deuxième cas correspondant à l'évolution climatique actuelle).

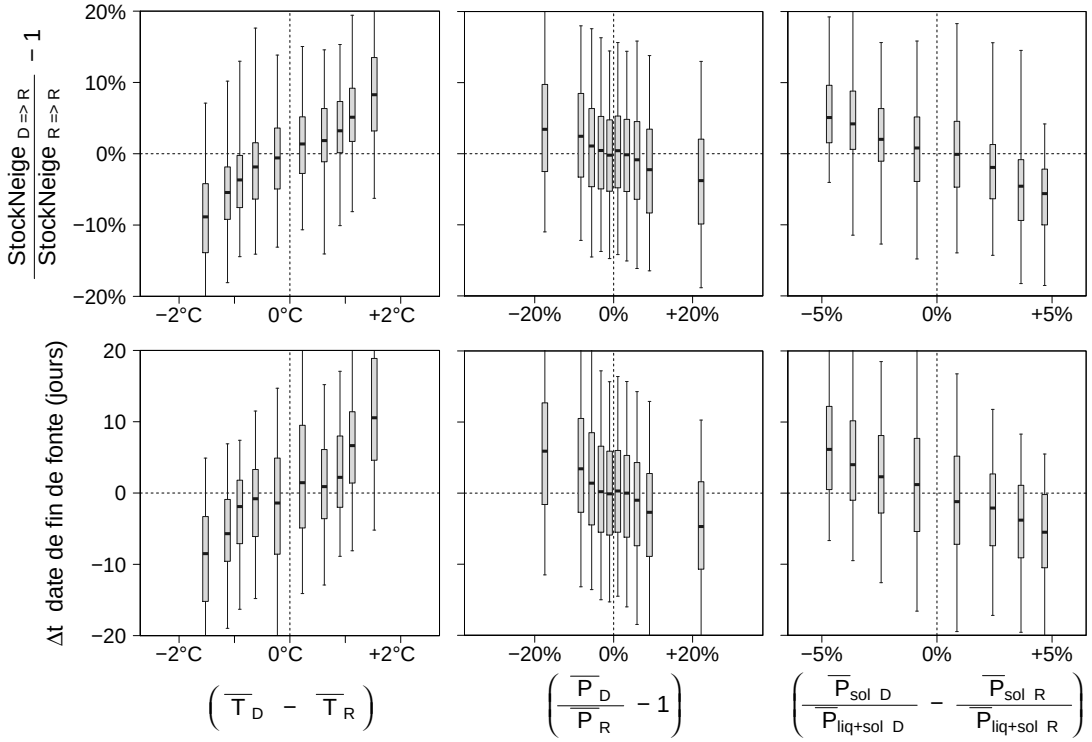


FIGURE 3.17.: Écarts de simulation de neige par *GR4J-CemaNeige* lors du *GSST* sur les 41 bassins recevant plus de 20% de précipitations annuelles sous forme solide

GR4J-CemaNeige (θ_{neige} calés) est alimenté ici par des données SPAZM et une ETP d'Oudin. Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Variations des paramètres du module neige

Nous ne rentrerons pas ici dans l'analyse des évolutions des paramètres du modèle hydrologique car ce point sera traité au sein du chapitre 5. Analysons toutefois brièvement les variations des paramètres du module neige que nous pensons pouvoir juger indépendamment du reste du modèle hydrologique. Pour rappel, *CemaNeige* est basé sur une méthode de fonte degré-jour et comporte deux paramètres libres, soumis à optimisation : le facteur de fonte (K_f) de la méthode degré-jour ainsi qu'un coefficient de pondération de l'état thermique du manteau neigeux (C_{T_G}) qui contrôle les conditions de déclenchement de la fonte (*ex.* de fortes valeurs de C_{T_G} permettent d'augmenter l'inertie du stock et retarder ainsi le démarrage de ma fonte). Ces deux paramètres sont calés en même temps que ceux du modèle hydrologique, par comparaison avec la chronique continue de débits observés. Il n'y a donc pas de confrontation avec des mesures de neige, ni de calage spécifique et/ou indépendant des paramètres de *CemaNeige* sur une période de l'année particulière (*ex.* hiver-printemps). Cette procédure permet pourtant l'identification de jeux de paramètres ($\theta_{modèle} + \theta_{CemaNeige}$) qui assurent une représentation adéquate de la relation précipitations-débit pour les bassins nivaux (*Valéry*, 2010).

Nous nous autorisons à étudier les variations moyennes de K_f et C_{T_G}, sans considérer les variations des paramètres du modèle pluie-débit pour deux raisons majeures. D'abord, les rôles joués par les paramètres de *CemaNeige* nous paraissent relativement indépendants de ceux joués par les paramètres du modèle. Ensuite, nos résultats tendent à montrer qu'il est acceptable de dissocier les écarts de simulation de la neige des erreurs principales de simulation commises sur les débits (nous pensons ici aux erreurs de volume qui créent la majeure partie des variations de *MRC* constatées à travers nos résultats). Pour résumer, il nous semble que les différences de stocks de neige constatées sur la figure 3.17 peuvent être expliquées par les seules variations des paramètres de *CemaNeige* (*c.-à-d.* que les aspects négligés dans ce raisonnement ne suffisent pas à le renverser).

Les différences relatives de valeurs des paramètres de *CemaNeige* obtenues en fonction des conditions de calage sont représentées sur la figure 3.18 [11]. Ces résultats montrent qu'une relation existe entre les valeurs de ces paramètres et la température moyenne (donc également le volume de neige) rencontrée au calage pour les 41 bassins considérés ici.

Nous constatons qu'un calage sur une période plus enneigée aboutit à la sélection d'un facteur K_f plus élevé, le coefficient C_{T_G} ne présentant pas (en moyenne) de tendances. Ces variations du K_f sont probablement la conséquence d'un besoin du modèle de "faire fondre" un stock de neige plus important dans un laps de temps limité, afin de rester en accord avec la chronique de débits. En effet, un phasage incorrect de ces processus engendrerait d'importants écarts entre les débits simulés et observés, en particulier durant la fonte printanière. Ce phasage impacte donc nécessairement la valeur du critère d'erreur utilisé pour l'optimisation et donc les valeurs des paramètres retenues au calage. Un calage sur une période plus enneigée contraint donc une augmentation de la vitesse de fonte. À l'inverse une période de calage plus chaude ne nécessitera pas d'aussi fortes valeurs de K_f (*cf.* figure 3.18).

3.5.3. Illustration sur deux bassins

Afin d'illustrer nos résultats, prenons l'exemple des bassins de *la Dranse à Reyvroz* (500 km²) et de *la Tet à Serdinya* (420 km²). Ces bassins ont respectivement des altitudes médianes 1360 m et 1760 m et la part solide de leur précipitations annuelles est estimée à 26% pour tous les deux (estimation réalisée en utilisant la fonction de dissociation pluie-neige présente dans *CemaNeige*). Leurs régimes sont de type nival, les écoulements y étant fortement influencés par l'accumulation de neige en altitude durant l'hiver, et par l'ablation du stock ainsi constitué créant une crue de fonte au printemps et en été.

11. Nous précisons que les notations D et R ne correspondent plus ici à des situations *donneuse/receveuse*, les comparaisons étant faites entre des valeurs de paramètres obtenues en calage (*c.-à-d.* nous aurions donc pu remplacer ces notations par A et B, ou 1 et 2).

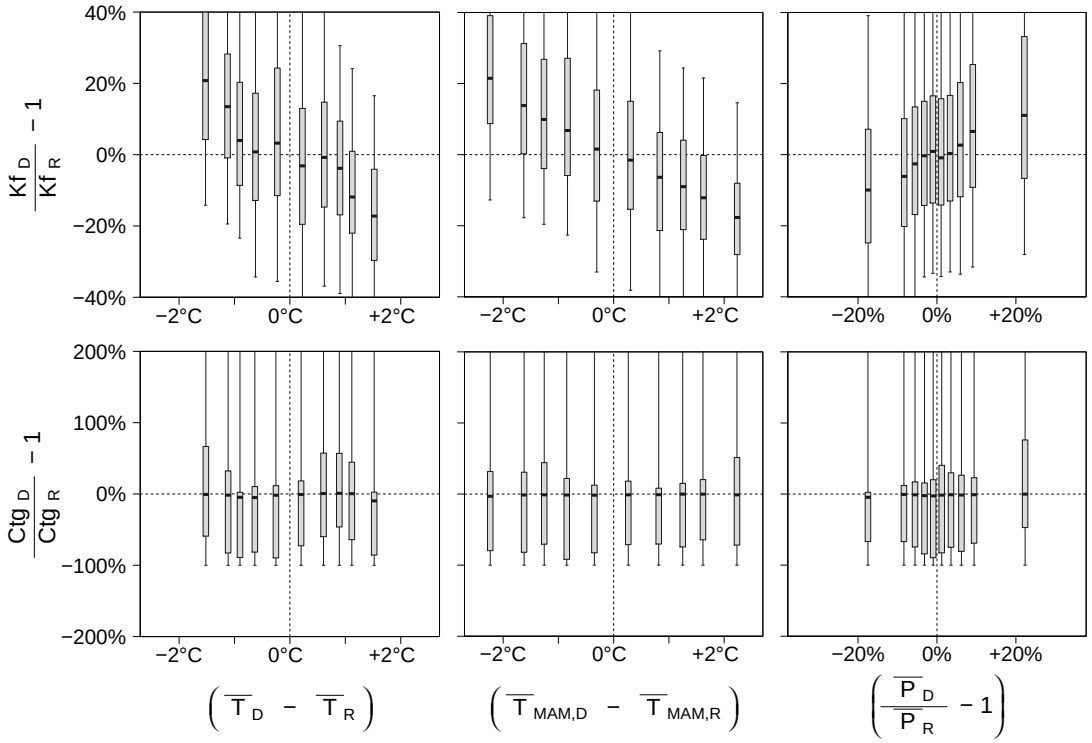


FIGURE 3.18.: Différences des valeurs de paramètres de *CemaNeige* (couplé à *GR4J*) en fonction conditions de calage sur les 41 bassins où $\overline{P}_{solides}/\overline{P}_{totales} > 20\%$

Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Quelques résultats des tests de transfert sont présentés sur la figure 3.19 et 3.20. Nous y retrouvons les nuages de points obtenus lors du *GSST* sur ces bassins, notamment les variations des stocks de neige simulés et des valeurs optimales du paramètre K_f de *CemaNeige* entre les différents calages temporels. Enfin, les débits et les stocks de neige journaliers interannuels que nous avons obtenus en validation avec *GR4J-CemaNeige* sont fournis sur deux périodes de 10 ans. Nous retrouvons des tendances décrites ci-dessus sur les différences de stock de neige simulé en fonction de l'écart de température moyenne subi lors du transfert des paramètres de *CemaNeige*. Ces décalages se retrouvent sur les débits avec un décalage et une amplitude parfois différente de la crue de fonte, qui impactent les critères d'erreurs mesurant la similitude des variances entre les observations et les simulations. Afin de faciliter la compréhension, nous avons volontairement choisi des bassins où la situation est plutôt claire pour illustrer les tendances constatées en moyenne sur le sous-groupe. Précisons néanmoins que tous bassins du sous-groupe considéré ici (41 bassins) n'ont pas des nuages et positionnements de courbes aussi nets. Ajoutons enfin que les liens constatés entre les écarts de stocks de neige simulés et les écarts de température subis lors des transferts semblent propres au module *CemaNeige*. Bien que des variations de stocks simulés soient également obtenues avec *Cequeau* (qui dispose de son propre module neige), elles n'ont pas pu être reliées aux changements de conditions climatiques entre les périodes de calage et de validation.

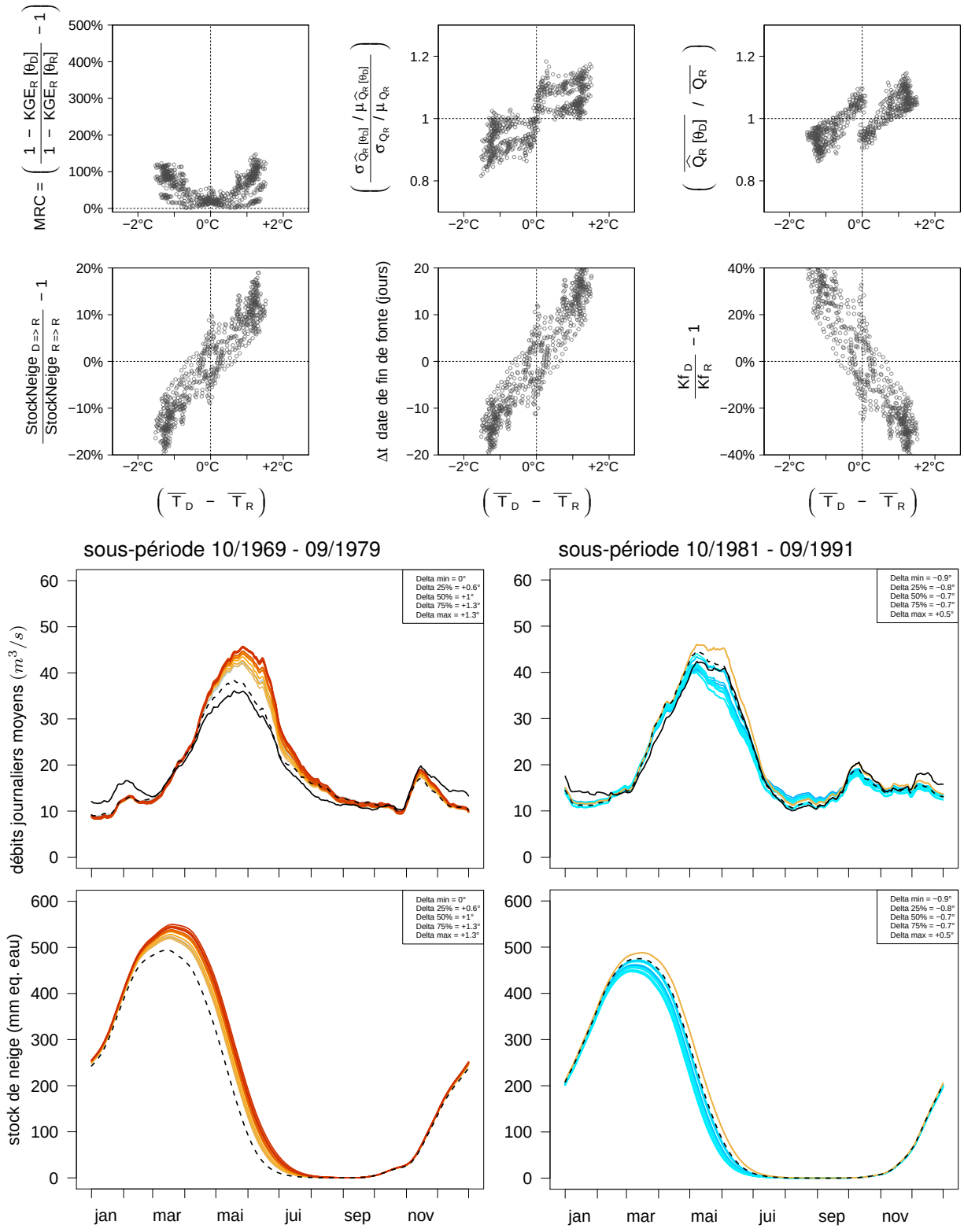


FIGURE 3.19.: Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Drance à Reyvroz*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (bleu foncé -1.5°C, gris 0°C, rouge +1.5°C).

Rq. GR4J est alimenté ici par des données SPAZM et une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} calés).

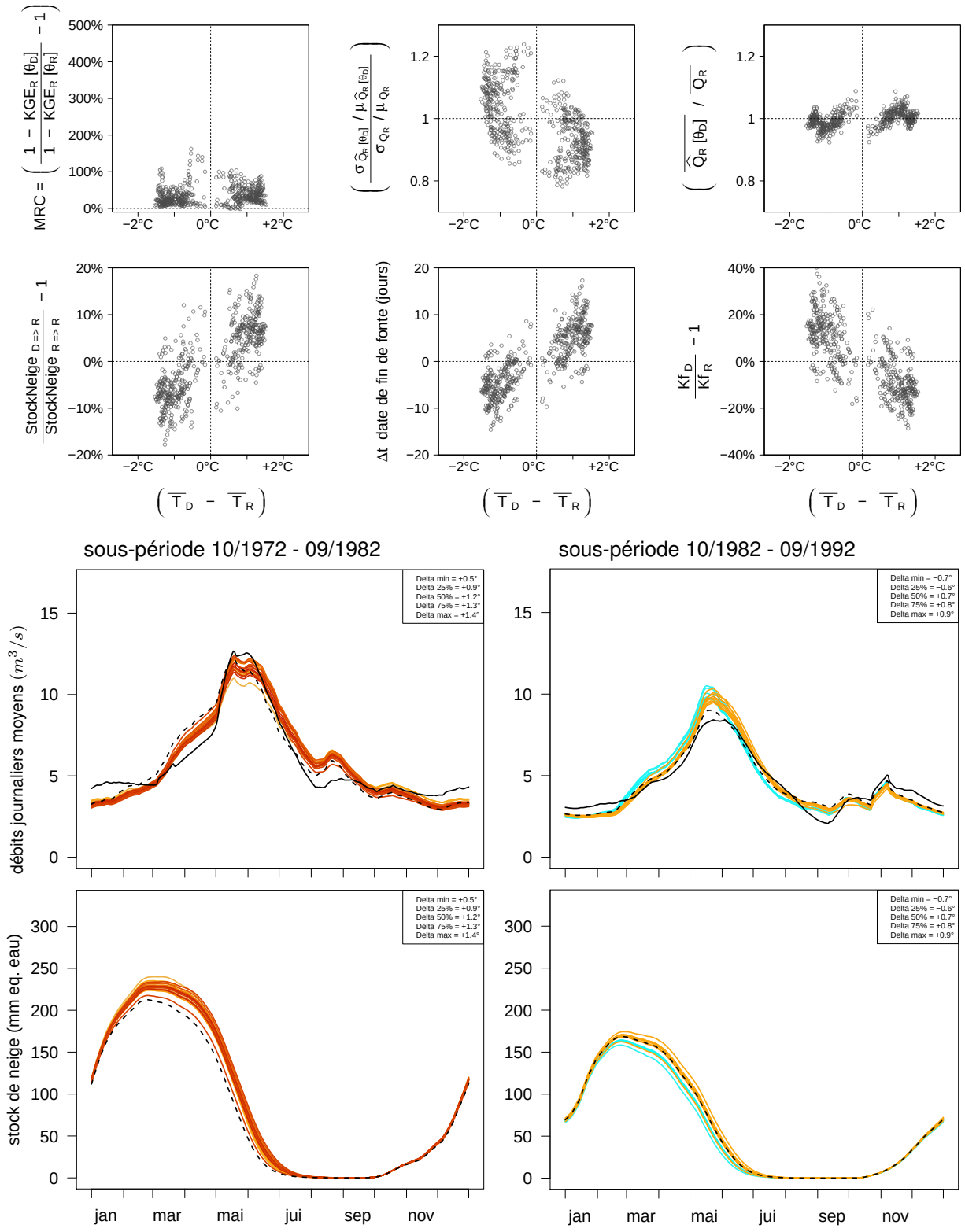


Figure 3.20.: Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Tet à Serdinya*

Débits journaliers lissés sur 31 jours et colorisés selon le code suivant : observations en noir continu, calage en noir pointillé, colorisation des simulations selon le $\triangle T_{D\to R}$ de transfert (bleu foncé -1.5°C, gris 0°C, rouge +1.5°C).

Rq. GR4J est alimenté ici par des données SPAZM et une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} calés).

3.6. Synthèse

Nous avons conduit, à travers la procédure *GSST*, un nombre très important de tests de robustesse de modèles hydrologiques sur un large échantillon de bassins. Ce chapitre a présenté un premier niveau d'analyse des résultats de ces tests, dans lesquels nous avons étudié si les erreurs de simulation pouvaient être reliées aux conditions de transfert des paramètres des modèles (*c.-à-d.* différences entre conditions de calage et de validation).

En résumé, les aspects suivants ont été constatés et discutés :

- Des corrélations entre les erreurs des modèles hydrologiques et les écarts de conditions climatiques subis lors des transferts de paramètres peuvent être détectées. Les erreurs sur les volumes moyens (biais) semblent particulièrement concernées et constituent vraisemblablement le principal contributeur des baisses de performances identifiées à travers les critères classiques de performances, tels que le KGE.

- Une hétérogénéité existe toutefois entre les nombreuses situations rencontrées dans nos tests, sans que nous soyons parvenus à expliquer clairement celle-ci.

  Les problèmes de robustesse semblent corrélés aux écarts de température moyenne sur de nombreux bassins (majoritairement français), notamment lorsque l'on s'intéresse au biais sur le module simulé. Sur d'autres bassins, le comportement des modèles semble impacté par des variations de précipitations moyennes (particulièrement en Australie).

  Le regroupement des bassins où les modèles ont des comportements similaires permet une analyse des amplitudes d'erreurs associées aux transferts de paramètres dans des conditions climatiques plus ou moins contrastées. Ces analyses, menées visuellement, montrent l'importance de l'augmentation des erreurs du modèle en concordance avec les écarts de conditions climatiques. Le biais sur le module décennal peut par exemple atteindre 20% lorsque les transferts sont effectués dans une situation défavorable (*ex.* forts $\left|\triangle\overline{T}\right|$ ou $\left|\triangle\overline{P}\right|$).

- L'existence de problèmes de robustesse des modules neige est également illustrée ici, à travers l'étude de la transférabilité des paramètres de *CemaNeige*. Les résultats montrent comment les paramètres de *CemaNeige* conduisent à des stocks de neige et des crues de fonte différents lorsqu'ils sont transférés entre des périodes contrastées du point de vue de la température de l'air (et donc de la quantité de neige reçue). Ces tests constituent seulement une entrée en matière et nécessiteraient d'être largement complétés notamment par le test de modules plus complexes (et plus physiques). Nous ne poursuivrons cependant pas cet axe de recherche dans cette thèse, car nous préférons concentrer nos analyses sur les biais importants constatés lors de certains transferts, dont la réduction nous paraît un objectif prioritaire. En effet, nous sommes en droit d'attendre d'un modèle qu'il reproduise les bilans hydrologiques moyens.

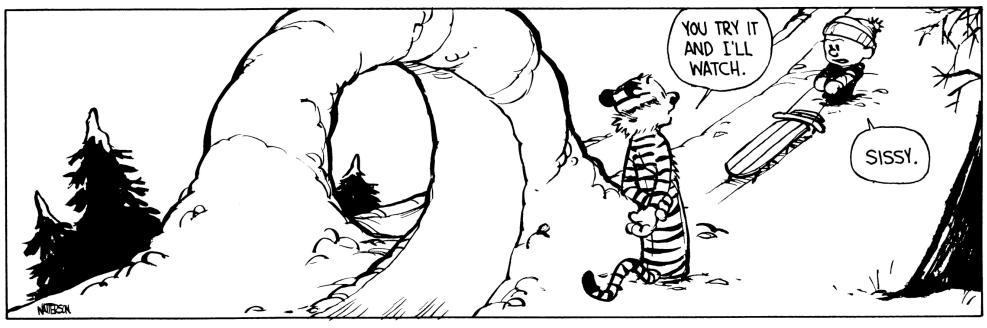
Les échecs mis en avant par le *GSST* montrent que cette capacité n'est pas garantie dans un contexte de climat en évolution et appellent à l'approfondissement de notre diagnostic sur ce point.

- Rappelons pour finir que *corrélation* n'est pas *causalité*. Au vu de nos résultats, notamment l'absence d'explication des différences observées entre bassins, il reste délicat de conclure sur un lien direct de causalité entre les écarts de conditions climatiques subis lors du transfert des paramètres des modèles et les situations où ces modèles sont en difficulté. Cela ne nous empêche pas pour autant de formuler des questions sur l'origine de ces problèmes de robustesse, parmi lesquelles se mélangent des réflexions sur les choix des méthodes de calage des modèles, les choix de structure des modèles ou encore le rôle des données qui les alimentent (*ex.* précipitations totales et solides ou évapotranspiration) et dont l'estimation peut être sujette à des erreurs qui évolueraient temporellement.

4

Sensibilité des projections à long terme aux choix de modélisation

Les études d'impact du changement climatique en hydrologie, un exercice toujours délicat ...



Calvin and Hobbes, 13/01/1989, Bill Watterson©

4.1. Introduction

4.1.1. Problématique et objectifs

Les scénarios d'émission de gaz à effet de serre et les modèles climatiques de circulation globale (*GCMs*) sont souvent considérés comme les sources principales d'incertitude dans les exercices de projection long terme en hydrologie (*cf. Arnell*, 2011; *Teng et al.*, 2012 et discussion de la section 1.3.2). La phase de descente d'échelle spatiale est également cruciale car la méthode utilisée, qu'elle soit statistique (*MDES*) ou dynamique (*MDED*), doit générer des forçages locaux représentatifs des conditions à partir d'un signal de *GCM* produit à des échelles beaucoup plus grandes. Or, les reliefs impactent grandement les conditions climatiques locales et ces effets sont généralement sous-représentés dans les *GCMs* (*Vidal et Wade*, 2008). L'impact du choix de la *MDES*/*MDED* sur les projections émises et leurs marges d'erreur peut donc être également important et sera renforcé en zone de montagne.

Le choix de la période de calage d'un modèle hydrologique influence logiquement les valeurs des paramètres retenus et impacte les simulations réalisées ensuite. Néanmoins, lorsque des périodes suffisamment longues sont utilisées, ces différences de paramètres devraient rester limitées et ne se ressentir que faiblement sur les débits simulés. La procédure *GSST* montre une réalité bien différente. Les erreurs de simulation d'un même modèle varient de façon conséquente selon la période de calage associée. Plus précisément, nos analyses montrent que les périodes de calage ne sont pas seules en cause mais que c'est bien la nature du couple formé par les périodes de calage et de simulation qui conduit ou non à des erreurs significatives de modélisation. Les écarts de conditions climatiques moyennes entre ces périodes semblent une explication possible des variations d'erreurs, bien que notre approche ne permette pas d'affirmer une véritable causalité. Cependant, il n'existe pas de comportement unique et les sensibilités varient d'un bassin à un autre, sans que nous ayons réussi à expliquer ces variations.

Ces résultats soulèvent des questions sur les performances d'un modèle hydrologique conceptuel utilisé pour des projections long terme de débits à des horizons de milieu ou fin de siècle. De fait, la période de simulation est bien distincte de celle de calage pour de telles utilisations. En règle générale, toute la série d'observations disponible est valorisée pour caler le modèle hydrologique dans ce type d'étude. Néanmoins, les résultats obtenus avec le *GSST* nous amènent à questionner la sensibilité des projections aux conditions de calage du modèle :

> *Les débits projetés seraient-ils très différents, si nous n'avions pas 30 années d'observations mais seulement les 10 premières années ou les 10 dernières ?*

Ce chapitre a pour objectif d'évaluer, sur un cas d'étude, cette sensibilité des projections de débits au choix de la période de calage. Celle-ci est comparée à d'autres sensibilités présentes dans la chaîne de modélisation, telles que l'influence des choix du *GCM*, de la *MDES* ou du modèle hydrologique. Nous cherchons ainsi à évaluer l'importance relative de ces choix sur les débits projetés et établir *certains parallèles* avec les classifications des sources incertitudes évoquées dans la section 1.3.2.

4.1.2. Parenthèse terminologique

Les "sensibilités" de projections aux choix de modélisation évoquées dans ce chapitre constituent des sources contributrices à l'incertitude entourant les projections. Nous devons cependant être vigilants sur les termes employés ici. Précisons notamment que la simple caractérisation de ces sensibilités ne suffit pas à l'estimation de marges d'incertitudes sur les débits projetés. Cette estimation nécessite un travail nettement plus complexe qui dépasse largement le champ de nos travaux. Par conséquent, nous veillerons à éviter les termes *"erreurs de projection"* ou *"incertitudes"* car ils englobent plus d'aspects que les seuls impacts des choix de modélisation que nous étudions ici. Nous préférons des termes tels que *"écarts de projections"* ou *"dispersion des débits projetés"* qui sont plus généraux et visent à ne pas tromper le lecteur sur la signification des enveloppes que nous sommes amenés à discuter dans ce chapitre.

4.1.3. Une valorisation seulement partielle des résultats du *GSST*

La procédure *GSST* et son utilisation sur de nombreux bassins ont mis en avant des aspects très intéressants. Tous ces aspects ne seront cependant pas exploités ici.

Le premier apport des analyses menées dans le chapitre 3 est la mise en évidence du rôle majeur du choix de la période de calage sur les erreurs de modélisation commises sur d'autres périodes. Ce point sera largement exploité dans ce chapitre, puisque nous évaluerons comment cette sensibilité à la période de calage se retranscrit sur les projections à long terme de débits. Ces évaluations seront menées sur un cas d'étude pour lequel nous disposons de séries climatiques sur la période 1860-2099.

Les analyses faisant suite au *GSST* permettent toutefois d'aller beaucoup plus loin : elles montrent l'importance du couple formé par les périodes de calage et de simulation. En effet, nous avons vu comment des liens pouvaient exister entre les erreurs d'un modèle hydrologique et les écarts de conditions climatiques subis lors du transfert de ses paramètres. Or, le changement climatique se définit justement par une évolution du climat qui s'amplifiera durant les décennies à venir, avec une accélération probable des changements à mesure que l'on se rapproche de la fin du siècle. Face à ce $\triangle$climat attendu (par rapport aux conditions

historiques du $20^{ème}$ siècle) les analyses présentées au chapitre 3 pourraient servir à estimer la gamme des erreurs commises pour certaines différences climatiques entre calage et simulation (*ex.* amplitude des erreurs de volume que l'on peut attendre pour un $\triangle\overline{T}$ donné). Certaines des évolutions attendues dépassant la variabilité climatique présente dans les séries d'observations, il serait parfois nécessaire d'extrapoler les gammes d'erreurs établies sur les conditions passées. Cet apport notable de nos travaux avec le *GSST* ne sera cependant pas exploité ici, l'étude présentée dans ce chapitre ayant surtout pour but d'illustrer la manière dont les projections à long terme de débits sont impactées par un changement de période de calage du modèle.

Nous estimons dans ce chapitre la sensibilité des projections à long terme aux conditions du calage du modèle hydrologique, à travers le test de multiples jeux de paramètres calés sur différentes sous-périodes temporelles. Nous aurions néanmoins pu utiliser une approche bien différente en ne partant que d'une unique période de calage et en *habillant* la projection déterministe correspondante. Un tel *habillage* serait réalisé en utilisant les estimations obtenues au chapitre 3 quant au manque de capacité d'extrapolation climatique des paramètres sur certains bassins. Pourtant, nous avons choisi de ne pas nous engager dans cette démarche. La raison principale de ce choix est l'hétérogénéité des situations rencontrées sur notre échantillon, c'est-à-dire l'absence d'une sensibilité des modèles aux mêmes variables climatiques sur différents bassins. Les comparaisons entre les nuages $MRC = f\left(\triangle\overline{T}\right)$ et $MRC = f\left(\triangle\overline{P}\right)$ sur les figures 3.2 et 3.8 (pages 77 et 86) illustrent très bien cette hétérogénéité. Nous constatons par exemple que des $\triangle\overline{T}$ similaires entre divers bassins peuvent être associés à des niveaux d'erreur des modèles en transfert totalement différents. En conséquence, la construction d'habillages devrait être menée parallèlement sur divers groupes de bassins, selon leur similarité en termes de robustesse d'un modèle. Le bassin où la projection est émise devrait ensuite subir des tests pour déterminer lequel de ces habillages doit lui être appliqué. Mener cet exercice avec rigueur devient donc rapidement complexe. Par conséquent, nous préférons nous contenter d'étudier la dispersion des débits projetés en considérant de multiples périodes de calage du modèle hydrologique. Bien que cet exercice ne soit pas aussi abouti que celui venant d'être évoqué, cette dispersion traduit tout de même une partie [1] de l'habillage qui pourrait être estimé. L'étude de cette dispersion conserve l'avantage de fournir un aperçu rapide (mais non moins tangible) de la manière dont les projections long terme en hydrologie sont impactées par le choix de la période de calage du modèle.

1. La dispersion ne traduit qu'une partie de l'habillage, tout d'abord parce qu'elle n'est obtenue que sur les données du bassin et n'intègre pas des estimations robustes sur de nombreux cas, et surtout parce qu'elle est limitée par la variabilité climatique observée, sans possibilité d'extrapolation.

4.2. Méthodologie

4.2.1. Contexte de l'exercice et origine des données

L'exercice présenté ici vise à mener une étude (simplifiée) d'impact du changement climatique sur les débits d'un bassin versant, pour apprécier la sensibilité des projections à certains choix dans la chaîne de modélisation. Précisons à nouveau que cet exercice ne prétend pas constituer une étude comparative des différentes sources d'incertitudes. Une telle étude dépasse en effet largement le champ des quelques analyses menées ici (*cf.* parenthèse terminologique précédente). Nous évaluons simplement ici la sensibilité des débits projetés au choix de la période de calage puis comparons celle-ci aux effets causés par un changement de *GCM*, de *MDES* ou de modèle hydrologique.

Pour cela, nous repartons des travaux menés dans le cadre du projet *RIWER-2030* sur la *Durance* amont [2], dans lequel des projections en changement climatique avaient été produites. Je tiens d'ailleurs à remercier Benoit Hingray, chef de ce projet de recherche, qui m'autorise à utiliser les données du projet nécessaire à la réalisation de l'exercice présenté ici. Avant de continuer, précisons toutefois que cette base de modèles et de données date de début 2012 et n'intègre pas les éventuelles modifications ou corrections apportées ensuite lors de la finalisation de *RIWER-2030* ou dans le cadre du projet *R2D2-2050* [3].

La *Durance* amont constitue une zone d'intérêt spécifique dans les études d'impact du changement climatique, du fait de la présence de la retenue de *Serre-Ponçon* (à la confluence de *la Durance* et de *l'Ubaye*) qui marque le point de départ d'une chaîne d'ouvrages hydroélectriques formant la plus importante unité de production de ce type en France. Par ailleurs, cette chaîne d'ouvrages à un rôle prépondérant sur la gestion de l'irrigation dans un très large périmètre provençal, mais ainsi sur l'alimentation en eau de Marseille et l'équilibre des conditions au sein de l'étang de Berre (qui forme l'exutoire de la partie canalisée de la Durance).

Estimer les évolutions de la ressource en eau sur ce territoire est donc un enjeu crucial pour de nombreux acteurs et, en premier lieu, pour *EDF* qui exploite le barrage de *Serre-Ponçon* et la chaîne d'ouvrages associée. Cette zone géographique fait l'objet de plusieurs projets de recherche (*RIWER-2030*, 2011; *AdatpAlp*, 2011; *R2D2-2050*, 2013).

Le bassin considéré pour notre exercice est celui de *la Durance à Embrun (La Clapière)*, d'une superficie d'environ 2300 km². Situé juste à l'amont de *Serre-Ponçon* ce bassin comporte des altitudes variant de 800 à 4000 m et présente un régime d'écoulement de type nival. Les précipitations de novembre à mars sur le bassin sont majoritairement stockées sous forme de

2. RIWER = Regional Climate, Water, Energy Resources and uncertainties from 1960 to 2030. Projet mené par le LTHE et EDF et financé par l'Agence Nationale de la Recherche (www.lthe.fr/RIWER2030/).

3. R2D2 = Risque, ressource en eau et gestion durable de la Durance en 2050 (r2d2- 2050.cemagref.fr).

neige, puis libérées au printemps avec des hautes eaux liées à la fonte atteignant leur maximum au mois de juin. La présence de glaciers sur ce bassin est très limitée et sera négligée par les modèles utilisés ici.

4.2.2. Chaîne de modélisation considérée

La production de projections long terme en hydrologie met en action une véritable chaîne de modélisation. Cette chaîne se compose de différents maillons pour lesquels des variantes sont possibles. Les données du projet *RIWER2030* à notre disposition nous permettent de considérer les configurations suivantes :

- Un scénario d'émission de gaz à effet de serre : *SRA1B*.
  Ce scénario prévoit à une augmentation des émissions jusqu'en 2050 suivie d'une diminution pour la seconde moitié du siècle. Il correspond à un scénario médian de réchauffement mondial, tel qu'il était estimé dans *IPCC* (2007).
- Quatre modèles climatiques globaux : *CNCM33* (1 run disp.), *DMIEH5C* (3 runs disp.), *EGMAM2* (1 run disp.), *MPEH5C* (3 runs disp.).
  Les mécanismes des circulations climatiques à l'échelle globale font intervenir des composantes chaotiques qui sont intégrées dans les *GCM*s. Deux simulations d'un même *GCM* (appelées expériences ou runs) ne seront donc pas équivalentes. Pour apprécier cet aléa, plusieurs runs d'un même *GCM* sont généralement considérés dans les exercices de projection à long terme.
- Trois méthodes de descente d'échelle stochastiques, basées sur une approche du type *analogue* : *ANADTG*, *DDWGEN*, *DSCLIM* avec 100 réalisations disponibles pour chacune
- Cinq modèles hydrologiques : *GR4J*, *Mordor6, SimHyd*, *Cequeau* et *Mordor*.
  Les trois premiers (*GR4J*, *Mordor6* et *SimHyd*) sont utilisés en couplage avec *CemaNeige*, tandis que les deux derniers (*Cequeau* et Mordor) disposent de leur propre module neige. Enfin, 40 jeux de paramètres sont considérés pour chaque modèle, issus de calages sur les observations disponibles (sous-périodes glissantes de 10 ans entre 1959 et 2008).
  (1 calage sur toute la période 1959-2008 et 39 calages sur des sous-périodes glissantes de 10 ans)

L'assemblage des trois premiers maillons nous permet de générer des séries journalières de précipitations et températures sur la période 1860-2099, qui alimentent un modèle hydrologique (le quatrième maillon) pour aboutir à l'émission des projections long terme de débits. De nombreuses combinaisons d'assemblage sont possibles, chacune aboutissant à une série de débits simulés. Parmi ces combinaisons, nous en choisissons une qui servira de référence pour étudier les sensibilités des projections aux choix réalisés au niveau de chaque maillon. Ces sensibilités sont traitées indépendamment les unes des autres, c'est-à-dire que les options d'un seul maillon sont testées à la fois, toutes choses restant égales par ailleurs. Par exemple,

nous étudierons l'impact d'un changement de modèle hydrologique sur les débits projetés en gardant les mêmes *GCM*, *MDES*, mais aussi la même période de calage du modèle.

4.2.3. Détail des sensibilités étudiées

Un seul scénario d'émission de gaz à effet de serre est à notre disposition. Par conséquent, ce maillon à l'amont de tous les autres, pourtant source majeure d'inconnues, ne sera pas soumis à notre étude de sensibilité.

Quatre modèles climatiques globaux sont à notre disposition, plusieurs expériences (ou runs) étant disponibles pour certains. Des différences importantes pouvant avoir lieu entre deux simulations d'un même *GCM*, nous avons fait le choix pour cet exercice simplifié de considérer dans un lot commun les différents *GCMs* et les différents runs d'un *GCM*.[4]

La phase de descente d'échelle est distinguée en deux sous-maillons de chaîne : la méthode utilisée d'une part (parmi les trois structures disponibles) et l'aléa lié à la dimension stochastique des méthodes statistiques de descente d'échelle d'autre part. Cette dimension stochastique se traduit en pratique par l'existence de différentes *réalisations* pour chaque méthode, 100 *réalisations* étant disponibles pour chaque *MDES* dans le cadre du projet *RIWER2030*.

La modélisation hydrologique est également distinguée en deux sous-maillons dans notre analyse de sensibilité : la sélection du modèle d'une part, et l'impact du choix de la période de calage de chaque modèle d'autre part, qui est traité séparément. D'autres distinctions auraient toutefois pu être considérées, telles que le choix de la fonction objectif lors de l'optimisation du modèle ou encore les aléas liés aux questions d'équifinalité[5]. La question du choix de la formule d'évapotranspiration utilisée constitue également un point important. Cette sensibilité est partiellement incluse dans nos travaux au travers des comparaisons multi-modèles : *GR4J*, *Mordor6* et *SimHyd* sont tous les trois alimentés par une ETP d'*Oudin* mais les modèles *Cequeau* et *Mordor* utilisent leurs propres formulations dont les paramètres sont calés avec le reste du modèle. L'impact de ces différences de formulations n'est cependant pas explicitement étudié ici et un travail complémentaire pourrait venir préciser ce point.

4. Il est vrai que ce maillon aurait pu être divisé en deux à l'image de ce que nous faisons pour la descente échelle ou la modélisation hydrologique.

5. Conservation de multiples jeux de paramètres à l'issu du calage, plutôt que de n'en retenir qu'un seul.

4.2.4. Situation de référence

Maillons retenus

L'étude des différentes sensibilités listées ci-dessus est menée autour d'une chaîne de modélisation de référence qui correspond à la situation suivante :

- scénario d'émission de gaz à effet de serre *SRA1B*
- modèle climatique de circulation générale *CNCM33* (unique run disponible)
- méthode de descente d'échelle *DSCLIM*
- réalisation analogue n°90 (de la méthode *DSCLIM*)
- modèle hydrologique *GR4J* couplé au module *CemaNeige* (paramètres calés)
- calage du couple *GR4J-CemaNeige* sur toutes les observations disponibles (période 1961-2008)

Le choix de cette chaîne de modélisation ne relève pas complètement du hasard. Le modèle de climat *CNCM33* produit des simulations médianes comparativement aux deux autres méthodes disponibles. La méthode de descente d'échelle retenue (*DSCLIM*) permet une restitution du régime des débits en temps présent qui est comparable à celle obtenue en utilisant les forçages *SPAZM* (issus de mesures et non d'un modèle de circulation atmosphérique). Les débits étant des intégrateurs du climat, cette similarité entre les débits simulés est recherchée car elle permet, dans une certaine mesure, de contrôler la pertinence de la chaîne *GCM-MDES* d'une façon indépendante à l'erreur du modèle hydrologique (le même modèle étant utilisé dans les deux cas). La réalisation n°90 de *DSCLIM* est une réalisation plutôt médiane parmi les 100 disponibles. Enfin, le modèle hydrologique choisi est *GR4J*, utilisé ici dans une version couplée à *CemaNeige*. Bien qu'il reproduise moins bien les régimes que *Cequeau* ou *SimHyd* en temps présent, il comporte peu de paramètres et une période de calage 10 ans est communément jugée suffisante pour obtenir un jeu de paramètres robuste, ce qui peut être questionné pour les autres modèles.

Projections correspondantes

Les projections de débits mensuels moyens pour cette chaîne sont présentées à la figure 4.1 pour trois périodes de 30 ans correspondant à des situations de temps présent (1979-2008), milieu (2035-2064) et fin (2070-2099) du siècle. Dans ces graphiques, ainsi que dans tous ceux qui suivront dans ce chapitre, les simulations issues de la chaîne de modélisation *GCM-MDES* - modèle hydrologique sont indiquées en bleu. Le régime observé en temps présent est indiqué en rouge pointillé et le régime simulé en temps présent à partir des forçages observés (et non pas de la chaîne *GCM-MDES*) est en rouge continu. À la différence des courbes bleues qui changent donc avec la période et l'assemblage de maillons considéré, ces courbes rouges restent fixes (*rq.* elles sont reproduites sur tous les graphiques).

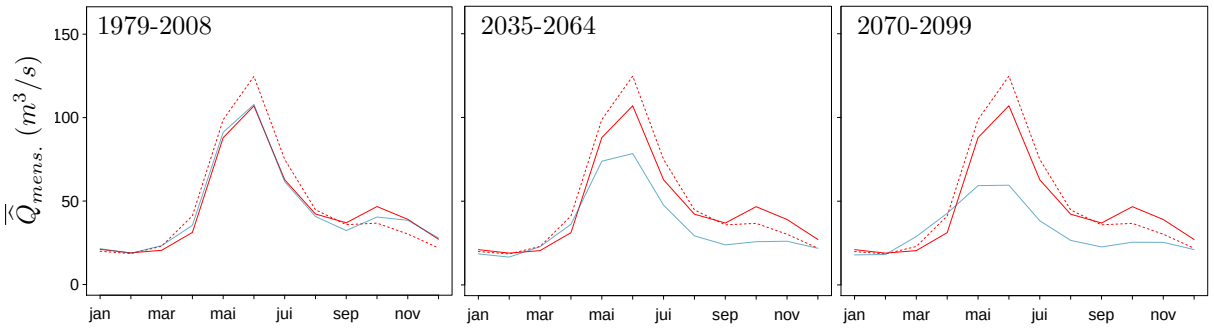


FIGURE 4.1.: Simulation des débits de *la Durance à Embrun* avec la chaîne de référence choisie

bleu : simulation sur la période de 30 ans pour la chaîne de modélisation considérée ; rouge continu : simulation sur 1979-2008 avec forçages observés ; rouge pointillé : observations sur 1979-2008

Mode de représentation et de lecture des résultats

Comme souvent dans ce type d'exercice, nous remarquons que la chaîne de modélisation présente déjà certaines difficultés à reproduire les observations en temps présent. Cela nous invite à lire ces graphiques de manière relative (par comparaison des situations entre les différentes échéances temporelles) plutôt que par analyse des valeurs absolues de débits simulés. Le régime observé est donc volontairement représenté de façon moins visible que la simulation issue des forçages observés qui constitue un meilleur point de repère. Celui-ci permet de constater l'imperfection de la modélisation hydrologique lorsqu'elle est comparée aux observations (rouge continu *vs.* rouge pointillé, sur la période historique). Il permet surtout, pour chaque assemblage des maillons *GCM-MDES*, d'apprécier l'imperfection de la représentation du climat local, indépendamment de l'erreur du modèle hydrologique (rouge continu *vs.* bleu continu, sur la période historique).

En résumé, notre exercice se focalise sur l'estimation des sensibilités des projections aux choix de modélisation et les courbes de projections seront, pour cette raison, analysées relativement les unes aux autres. Les valeurs absolues des projections seront relativement peu commentées, notre attention se portant sur l'épaisseur des faisceaux formés par les tests de différentes options pour un maillon. Cette épaisseur nous servira d'indicateur du degré de sensibilité de la chaîne de modélisation aux différents maillons qui la composent. Enfin, seuls les graphiques concernant les débits sont présentés dans le corps du manuscrit pour assurer une clarté de l'exposé. Certaines figures concernant les forçages (précipitations et températures) sont toutefois disponibles dans l'annexe F, notamment celles illustrant notre situation de référence et l'étude des sensibilités des projections aux choix du *GCM* et de la *MDES* (*cf.* figures F.1 à F.4, pages 307-308).

4.3. Étude de sensibilité des projections absolues à différents maillons de modélisation

4.3.1. Sensibilité des projections à la modélisation du climat global

Les huit configurations disponibles de *GCM* ou runs associés sont successivement testées, les autres maillons de la chaîne restant identiques à la situation de référence choisie. Les dispersions des projections ainsi obtenues sur nos trois horizons sont illustrées à la figure 4.2, les graphiques des forçages correspondants étant renvoyés en annexe (figure F.2, page 307).

La dispersion constatée est déjà relativement importante en temps présent. Elle s'amplifie nettement pour l'échéance milieu de siècle, puis les projections de fin de siècle se resserrent à nouveau pour une majorité de variantes de *GCM*, seulement l'une d'entre-elles se détachant du lot (un run de MPEH5C parmi les trois disponibles).

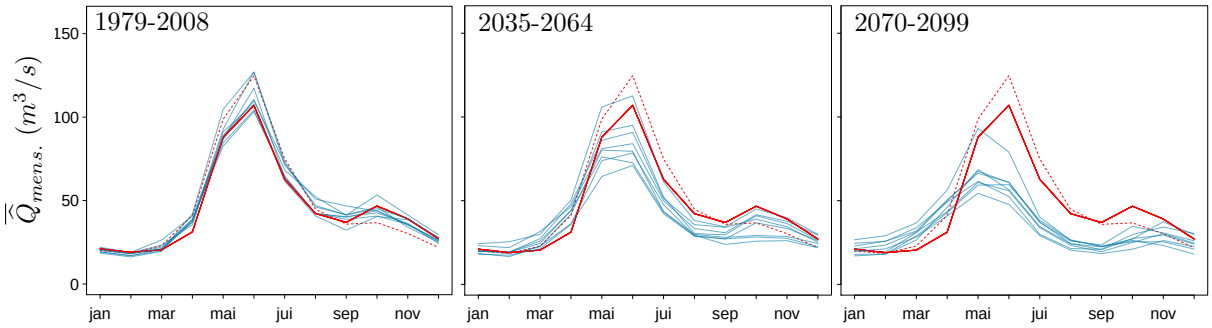


Figure 4.2.: Écarts de simulation des débits de *la Durance à Embrun* selon le *GCM* utilisé (toutes choses égales par ailleurs) *code couleur fig. 4.1*

4.3.2. Sensibilité des projections à la descente d'échelle

Intéressons-nous désormais à la sensibilité des projections face au choix de la méthode de descente d'échelle. Comme précédemment, tous les autres maillons de la chaîne de modélisation sont maintenus dans la situation de référence (*cf.* section 4.2.4).

Nous disposons de trois méthodes de descente d'échelle, toutes statistiques. Dans un premier temps, nous considérons une seule réalisation analogue pour chacune d'elle afin d'apprécier uniquement l'impact d'un changement de méthode sur les débits projetés. Il est vrai que, dans la première phase, le choix de la réalisation analogue retenue pour chaque *MDES* pourrait influencer notre interprétation. Comme nous le verrons ci-dessous, la dispersion des débits liée au choix des réalisations d'une même *MDES* est plus faible que celle causée par un changement

de *MDES*. En conséquence, ce choix d'une seule réalisation pour la première phase d'analyse interfère peu avec notre mesure de sensibilité. En complément, nous avons veillé à choisir pour chaque *MDES* une réalisation plutôt médiane dans le faisceau disponible, qui sera donc représentative de son positionnement.

Sensibilité au choix de la méthode de descente d'échelle

La sensibilité des débits projetés à un changement de méthode de descente d'échelle (entre *DSCLIM*, *ANADTG*, et *DDWGEN*) est illustrée à la figure 4.3 où l'on constate son importance. La dispersion constatée résulte de nettes différences sur la modélisation des volumes précipités entre les méthodes considérées (*cf.* figure F.3, page 308). La méthode *DDWGEN* simule des précipitations hivernales plus fortes que les observations en temps présent, avec une tendance à l'augmentation de celles-ci pour le futur. Cela contraste avec les deux autres méthodes, qui surestiment légèrement les précipitations hivernales en temps présent mais projettent une stagnation (*ANADTG*) ou une diminution (*DSCLIM*) de celles-ci. Les précipitations sur le bassin de *la Durance à Embrun* étant majoritairement stockées en hiver puis générant une crue de fonte lors de la remontée des températures, les différences constatées sur les précipitations hivernales conduisent aux forts écarts visibles ici sur les débits simulés en mai-juin-juillet. [6]

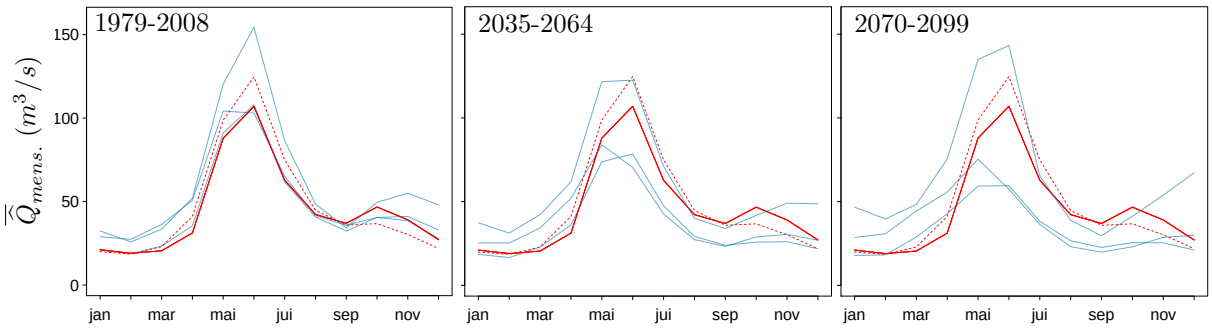


FIGURE 4.3.: Écarts de simulation des débits de *la Durance à Embrun* selon la *MDES* utilisée (toutes choses égales par ailleurs) *code couleur fig. 4.1*

6. À notre connaissance, il semble toutefois exister certaines erreurs dans la façon dont *DDWGEN* (version de janvier 2012) procède à une multiplication des précipitations pour simuler leurs évolutions. L'impact de ces erreurs sur nos analyses globales devrait toutefois resté limité puisque la situation référente n'utilise pas cette MDES et que, de toute façon, nous ne souhaitons apprécier les sensibilités que d'une manière très générale.

Sensibilité au choix de la réalisation analogue

Les simulations obtenues avec les 100 réalisations analogues de *DSCLIM* sont illustrées à la figure 4.4. Les résultats de nos simulations montrent que le faisceau des débits projetés est loin d'avoir une épaisseur négligeable mais cette dispersion reste inférieure à celles liées au choix de la *MDES* ou du *GCM*. Tout comme pour les différentes options de *GCM* ou de *MDES*, ces différences de projections résultent majoritairement d'écarts sur les précipitations modélisées plutôt que sur les températures (*cf.* figure F.4, page 308). Avant de conclure sur ce point, rappelons qu'en pratique, toutes les réalisations analogues disponibles sont utilisées dans une approche ensembliste, la médiane et divers quantiles étant retenus pour former une enveloppe de débits projetés.

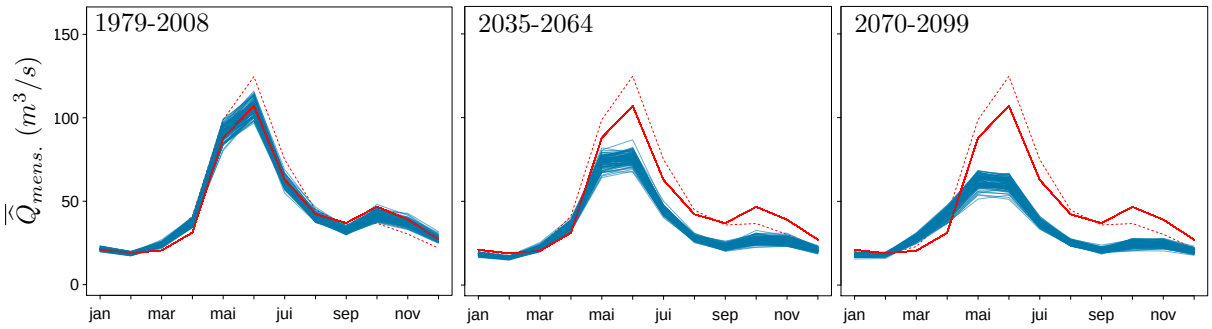


Figure 4.4.: Écarts de simulation des débits de *la Durance à Embrun* selon la réalisation analogue de *MDES* utilisée (toutes choses égales par ailleurs) *code couleur fig. 4.1*

4.3.3. Sensibilité des projections à la modélisation hydrologique

Sensibilité au choix du modèle hydrologique

Le choix du modèle hydrologique et de son niveau de complexité, est souvent source de débats passionnés en hydrologie. Pour cet exercice, nous comparons cinq modèles hydrologiques correspondant à des structures différentes. Classés par nombre croissant de paramètres calés, ces modèles sont les suivants : *GR4J*, *Mordor6*, *SimHyd*, *Cequeau* et *Mordor*. Les trois premiers modèles sont couplés avec le module neige *CemaNeige*, dont les paramètres sont calés avec le modèle. Les différentes projections de débits obtenues pour ces options sont illustrées à la figure 4.5. Comme précédemment, nous n'avons pas pour objectif d'analyser ici la capacité de ces modèles à représenter les observations, mais nous souhaitons juste visualiser la dispersion associée au choix du modèle. Nous noterons que les simulations en temps présent semblent plus proches lorsque les forçages sont issus d'observations (courbes continues rouges) que lorsqu'ils proviennent de la chaîne GCM-*MDES* (*GR4J* donnant des débits plus faibles et *SimHyd* les plus forts). En nous intéressant aux évolutions des débits projetés, les tendances sont logiquement similaires entre les quatre modèles (*GR4J* et *SimHyd* simulant, comme en temps présent, des débits respectivement plus faibles et plus forts que *Cequeau*, *Mordor* et *Mordor6*).

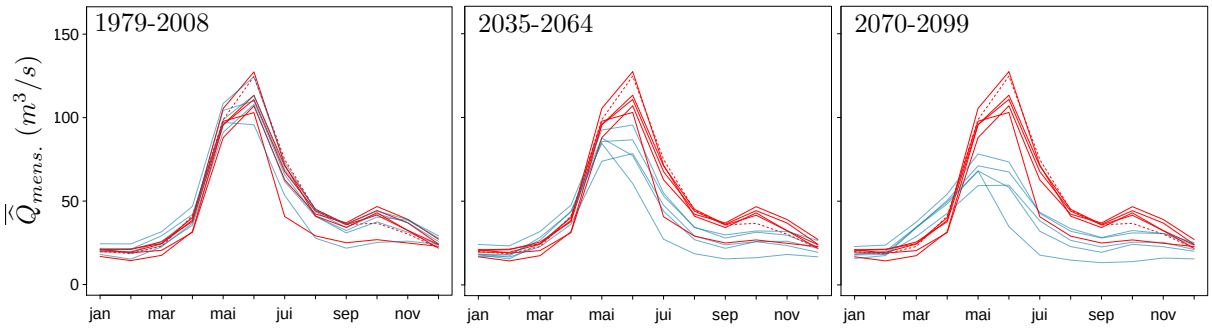


FIGURE 4.5.: Écarts de simulation des débits de *la Durance à Embrun* selon le modèle hydrologique utilisé (toutes choses égales par ailleurs) *code couleur fig. 4.1*

Sensibilité au choix de la période de calage du modèle hydrologique

Le choix de la période de calage peut influer notablement sur les débits simulés sur une autre période, comme nous l'avons constaté au chapitre 3. Pour l'exercice mené ici, nous repartons une fois encore de la chaîne de référence, en faisant seulement varier cette fois-ci le jeu de paramètres utilisé pour *GR4J*. Au total, 39 jeux différents sont testés, issus de calages sur une fenêtre glissante de 10 ans (à la manière de ce qui est fait dans le *GSST*, *cf.* section 2.1.2). Les différents régimes ainsi projetés sont illustrés à la figure 4.6. Nous retrouvons des différences non négligeables mais notons que la répartition des projections au sein du faisceau semble relativement homogène. L'enveloppe constatée n'est en effet pas le fait d'un ou deux jeux de paramètres. Nous détectons donc bien ici une source d'incertitude de simulation liée au choix de la période de calage, qui n'est pas liée à quelques échecs isolés dans la sélection de paramètres adéquats. Rappelons enfin les éléments discutés en introduction, sur le fait que toutes les erreurs associées au choix des conditions de calage ne sont pas retranscrites par ce test de sensibilité. Une part supplémentaire de dispersion provient du manque de capacité d'extrapolation climatique des paramètres mis en avant lors du chapitre 3 et dont les résultats ne sont que partiellement exploités dans ce chapitre (*cf.* section 4.1.3).

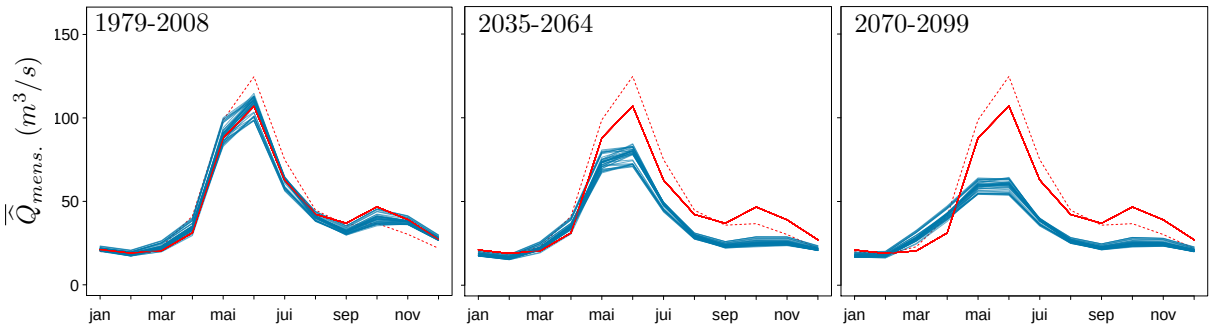


FIGURE 4.6.: Écarts de simulation des débits de *la Durance à Embrun* selon la période de calage du modèle hydrologique (toutes choses égales par ailleurs)

4.3.4. Synthèse des sensibilités constatées sur les projections absolues

L'exercice mené ici nous a permis de visualiser les sensibilités des projections de débits de *la Durance à Embrun* en temps présent et futur, face à différents choix de modélisation parmi les configurations à notre disposition. Ces sensibilités sont étudiées indépendamment les unes des autres, vis-à-vis d'une configuration particulière, choisie comme référence. Afin de faciliter leur comparaison, nous avons regroupé ces résultats sur la figure 4.7, où sont visibles les degrés de dispersion associés aux différents maillons considérés pour les trois périodes considérées dans nos analyses.

D'une manière générale, tous les maillons étudiés ici conduisent à une dispersion non négligeable sur les projections émises. Le choix du *GCM* constitue une source majeure de variabilité des projections, la dispersion étant plus élevée pour l'échéance de milieu de siècle que pour celle de fin de siècle. Pour cette dernière échéance, nous voyons aussi qu'un seul membre du faisceau se détache largement des autres. Concernant l'impact du choix de la *MDES*, nos travaux laissent entrevoir qu'il est également important. Le nombre trop faible d'options comparées et le comportement parfois atypique de la version de *DDWGEN* utilisée ici ne nous permettent cependant pas de tirer de conclusions claires sur ce point. En effet, lorsque *DDWGEN* est laissée de côté, les dispersions obtenues sont presque ramenées à des niveaux équivalents à celles associées aux autres choix de modélisation. Nous devons toutefois reconnaître que le fait d'écarter un comportement atypique est toujours discutable car, sous couvert de filtrage des incohérences, il peut conduire à nier une variabilité des modèles. En ce qui concerne la modélisation hydrologique, l'utilisation d'un modèle plutôt qu'un autre entraîne également des différences importantes sur les projections, mais cette sensibilité reste inférieure à celles liées au choix du *GCM* ou de la *MDES*. Nous noterons que les choix de la réalisation analogue et de la période de calage de *GR4J-CemaNeige* entraînent des niveaux de dispersion équivalents, qui se rapprochent de ceux observés pour un changement de modèle hydrologique, les modèles sélectionnés ayant pourtant des structures bien différentes. Pour finir, précisons que nous avons testé plusieurs versions de la configuration de référence à partir de laquelle les différentes sensibilités sont estimées (notamment en changeant le modèle hydrologique référent) et qu'elles conduisent à des interprétations similaires à celles présentées ici.

Profitons de cette synthèse pour insister sur le fait que les sensibilités discutées jusqu'à présent sont estimées sur des projections *absolues* de débits. Ce qualificatif *absolu* signifie que nous étudions indépendamment les dispersions obtenues sur différentes échéances de simulation (*ex.* 1979-2008 , 2035-2064, 2070-2099). Même si nous sommes amenés à comparer visuellement ces évolutions temporelles, cette comparaison se distingue nettement de l'étude de sensibilité des projections relatives de débits ($\triangle\widehat{Q}_{present \rightarrow futur}$) qui fait l'objet de la section suivante.

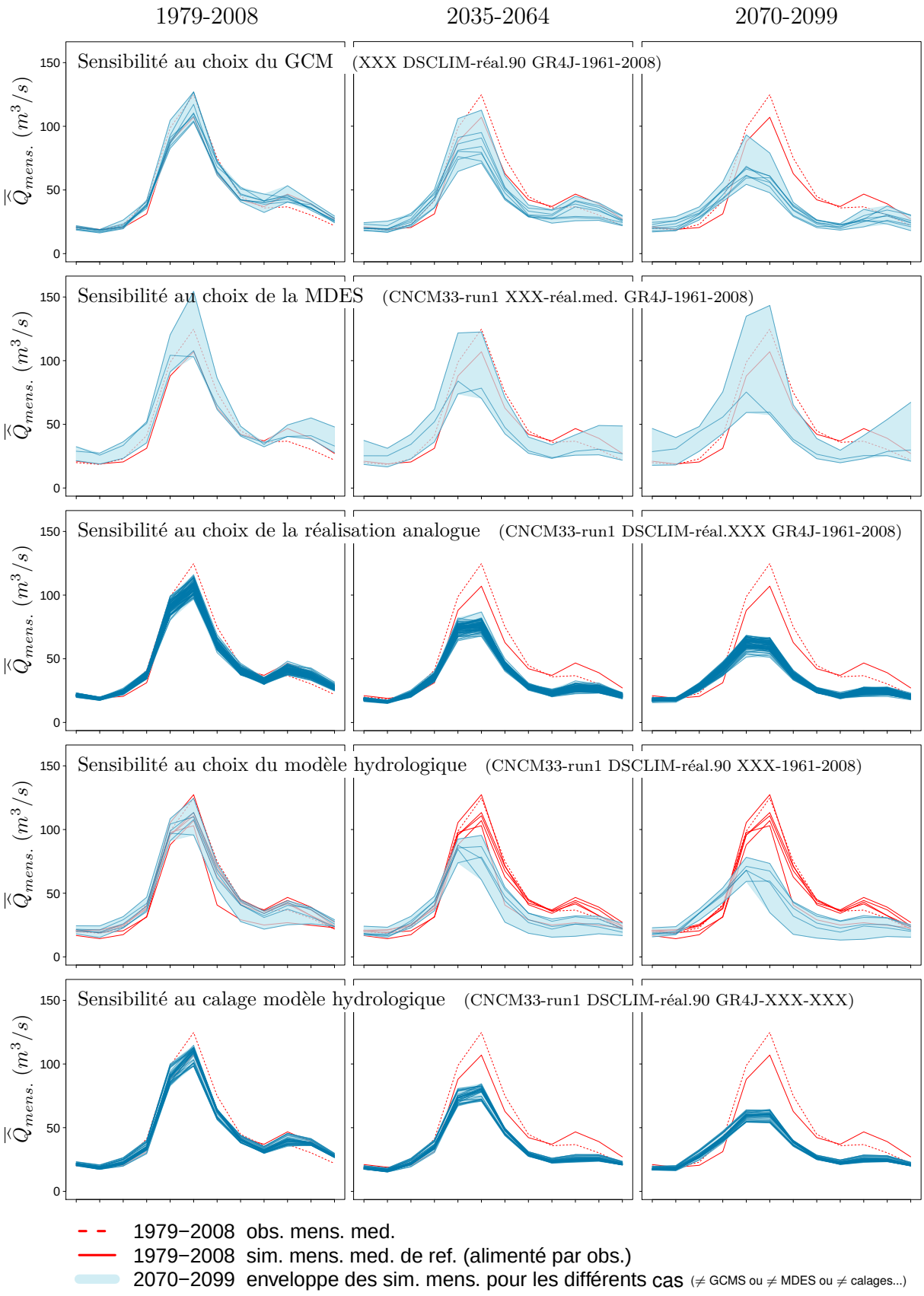


FIGURE 4.7.: Synthèse des sensibilités des débits projetés pour *la Durance à Embrun* au choix effectués sur les différents maillons de la chaîne de modélisation

Écarts liés à chaque maillon analysé indépendamment, les autres étant fixés à une situation choisie (servant de référence) : CNCM33-run1 + DSCLIM-sce90 + GR4J-cal-1961-2008

4.4. Étude de sensibilité des projections relatives à différents maillons de modélisation

4.4.1. L'expression de projections relatives : une pratique courante

D'importants progrès ont été réalisés au cours de la dernière décennie, mais les chaînes de modélisation *GCM-MDES* peinent encore parfois à reproduire les forçages climatiques à la résolution du bassin versant. Cet exercice est notamment difficile en zone de montagne où les reliefs influencent fortement les conditions climatiques locales. À ces imperfections des forçages modélisés viennent s'ajouter celles des modèles hydrologiques. Cet ensemble conduit parfois à des écarts notables entre les simulations issues de la chaîne de modélisation et les observations sur une période historique (*rq.* la situation est favorable sur l'exemple utilisé dans ce chapitre).

Dans les exercices de projection à long terme, il est courant d'exprimer les projections sous forme d'évolutions relatives, par rapport à une situation de référence, plutôt qu'en valeurs absolues. Ces évolutions relatives, aussi désignées sous le terme *d'anomalies* (climatique ou hydrologique) permettent une communication claire sur les changements à prévoir, qu'ils soient à la hausse ou à la baisse. Se pose néanmoins la question de la référence utilisée. Plutôt que de considérer la situation historique observée, une pratique courante consiste à utiliser une simulation de cette période historique avec les mêmes outils de modélisation que ceux utilisés pour émettre les projections. Par exemple, les débits prévus en 2100 par une chaîne *GCM-MDES-modèle-hydrologique* sont comparés avec les débits simulés par cette même chaîne en temps présent. Les différences constatées entre ces simulations (*ex.* -20% sur les débits moyens, +5% sur les débits extrêmes, ...) constituent alors les résultats considérés applicables aux observations. Les défenseurs de ce type d'approche expliquent qu'elle permet de s'affranchir en partie des imperfections des modèles en travaillant, d'une certaine manière, *à erreur constante*. Ce raisonnement repose cependant sur une hypothèse très forte : la stationnarité des erreurs de la chaîne de modélisation. Il suppose en effet que le même type d'erreur sera commis sur les différentes périodes de simulation (*ex.* sous-estimation systématique de 5% des débits).

Sans nous prononcer sur le bien-fondé d'une telle pratique (utilisation d'une référence simulée plutôt qu'observée), nous souhaitons voir à quels résultats elle conduit. Nous reprenons donc nos études de sensibilité précédentes dans ce cadre d'une expression relative des projections.

4.4.2. Méthodologie suivie

Nous repartons des simulations produites précédemment sur *la Durance à Embrun* et exprimons celles-ci sous forme d'évolutions relatives. Pour simplifier cet exercice, nous reprenons les mêmes périodes d'analyses que précédemment, la période 1979-2008 jouant le rôle de référence. Pour chacun des assemblages de maillons testés, nous calculons ainsi une projection relative pour deux horizons (*H1=2035-2064* et *H2*=2070-2099). Ces *anomalies* peuvent être calculées sur diverses caractéristiques des chroniques. Nous pouvons par exemple estimer l'évolution relative du débit moyen annuel, des débits moyens saisonniers, mais aussi de quantiles particuliers sur les faibles et forts débits, de volume de crue de fonte, de caractéristiques d'étiage, etc. Ces projections relatives, notées *PH1* et *PH2*, sont construites sur la base suivante :

$$PH1 = \widehat{Q}_{2035-2064} \,/\, \widehat{Q}_{1979-2008} \; - 1 \qquad PH2 = \widehat{Q}_{2070-2099} \,/\, \widehat{Q}_{1979-2008} \; - 1$$

Dans le même esprit que nos études des projections absolues (*cf.* section 4.3), nous étudions les sensibilités de façon indépendante entre les différents maillons considérés. Par exemple, nous disposons de huit valeurs pour chaque *PH* lors de l'étude de sensibilité liée au choix du *GCM* (puisque nous avons de huit options de *GCM* ou run). Ces valeurs de *PH* sont reportées sur une même figure pour visualiser la dispersion des projections relatives liée à ce choix du *GCM*.

4.4.3. Niveaux de sensibilité obtenus sur les projections relatives

Les études des sensibilités sont menées parallèlement sur les différents maillons de chaîne de modélisation et nous permettent de construire des graphiques synthétiques où les dispersions associées aux projections relatives peuvent être comparées. Le graphique concernant les projections de débits moyens annuels avec *GR4J* est présenté à la figure 4.8 et ceux concernant les différents quantiles de débits et les moyennes saisonnières composent les figures 4.9 et 4.10. Ces nouvelles analyses de sensibilité mettent en avant des éléments très intéressants.

Les choix du *GCMs* et de la *MDES* semblent rester des causes majeures de dispersion des projections, même lorsque celles-ci sont exprimées sous forme relative. Les amplitudes visualisées sur cette représentation signifient que les différences entre les *GCMs* ou entre les *MDES* ne sont pas de simples relations de proportionnalité homogènes dans le temps (*cf.* figure 4.8). En effet, si par exemple deux *GCMs* fournissaient exactement le même signal mais avec un écart de volume de 20% sur tous les horizons, leurs *PH1* et *PH2* seraient par construction confondues, or ce n'est pas le cas ici.

La dispersion des projections relatives associée au choix de la réalisation analogue (pour *DSCLIM*) est également conséquente. Par exemple, la diminution du module de *la Durance à Embrun* projetée pour la fin du siècle oscille entre -21% et -40% selon la réalisation considérée. Ce choix de la réalisation analogue conduisait à une dispersion des régimes simulés plus

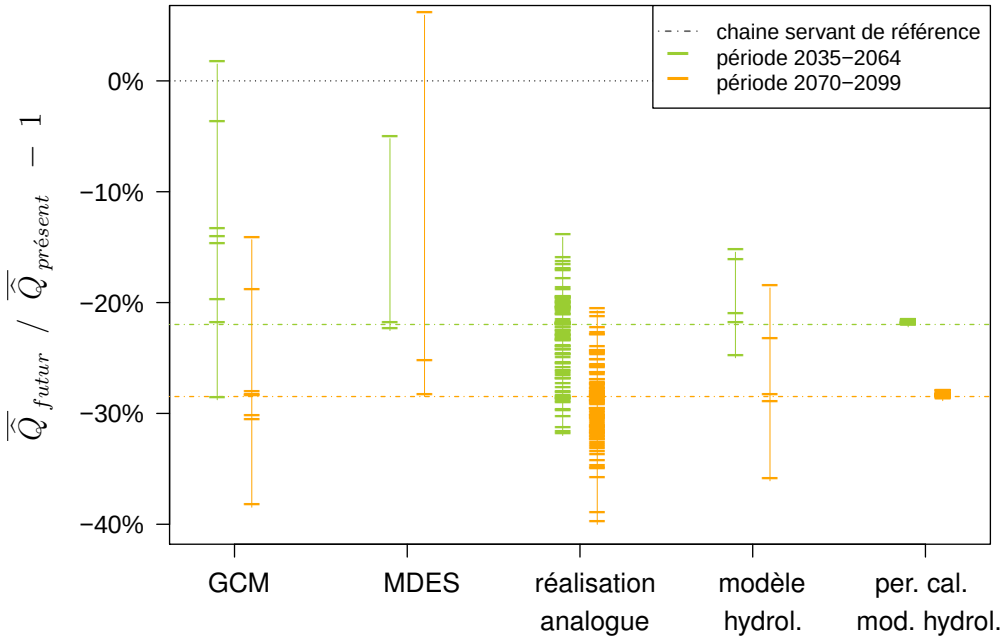


Figure 4.8.: Écarts sur les $\triangle\overline{\widehat{Q}}$ projetés avec *GR4J* sur *la Durance à Embrun* pour les deux horizons considérés, lorsque chaque maillon est considéré indépendamment

faible que le choix du *GCM* ou de la *MDES* lorsque nous considérions les projections absolues (*cf.* figure 4.7). En étudiant les projectives relatives, il apparaît pourtant presque comme un facteur de sensibilité d'un niveau équivalent au choix du *GCM*.

La dispersion associée à un changement de modèle hydrologique reste quant à elle modeste sur les moyennes annuelles, comparativement aux autres maillons de la chaîne (*cf.* figure 4.8). Nous noterons toutefois des différences plus marquées sur les évolutions de débits saisonniers ou de quantiles de débits (*cf.* figures 4.9 et 4.10).

Finissons enfin par la sensibilité des projections relatives au choix de la sous-période de calage du modèle hydrologique, qui constitue pour nous un sujet d'étude privilégié. Il est très intéressant de constater ici comment les options de jeux de paramètres n'entraînent quasiment aucune différence sur les évolutions relatives des débits projetées pour *GR4J*. Alors que nous visualisions à la figure 4.7 un niveau de dispersion des régimes comparable à celui lié au choix de la réalisation analogue, ces différences disparaissent presque entièrement ici ! Il est même surprenant de constater qu'un tel résultat est présent non seulement sur les débits moyens, mais aussi sur différents quantiles de débits plus ou moins extrêmes (*cf.* figure 4.9) et qu'il se retrouve au niveau des différentes saisons (*cf.* figure 4.10).

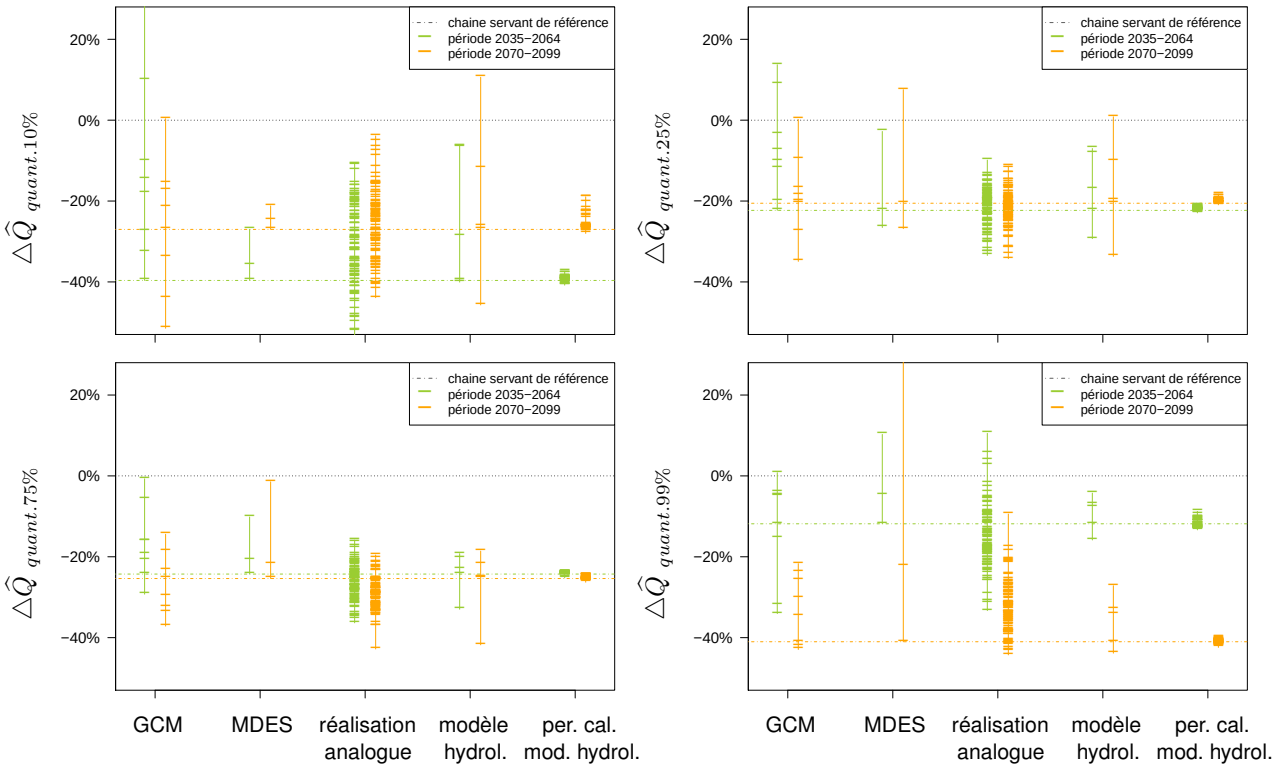


FIGURE 4.9.: Écarts sur les $\triangle\widehat{Q}_{10}$, $\triangle\widehat{Q}_{25}$, $\triangle\widehat{Q}_{75}$ et $\triangle\widehat{Q}_{99}$ projetés avec *GR4J* sur *la Durance à Embrun*, lorsque chaque maillon est considéré indépendamment

Les notations utilisées ici correspondent aux quantiles de non-dépassement.
Par exemple, le Q99 correspond à un fort débit qui n'est dépassé que 1% du temps.

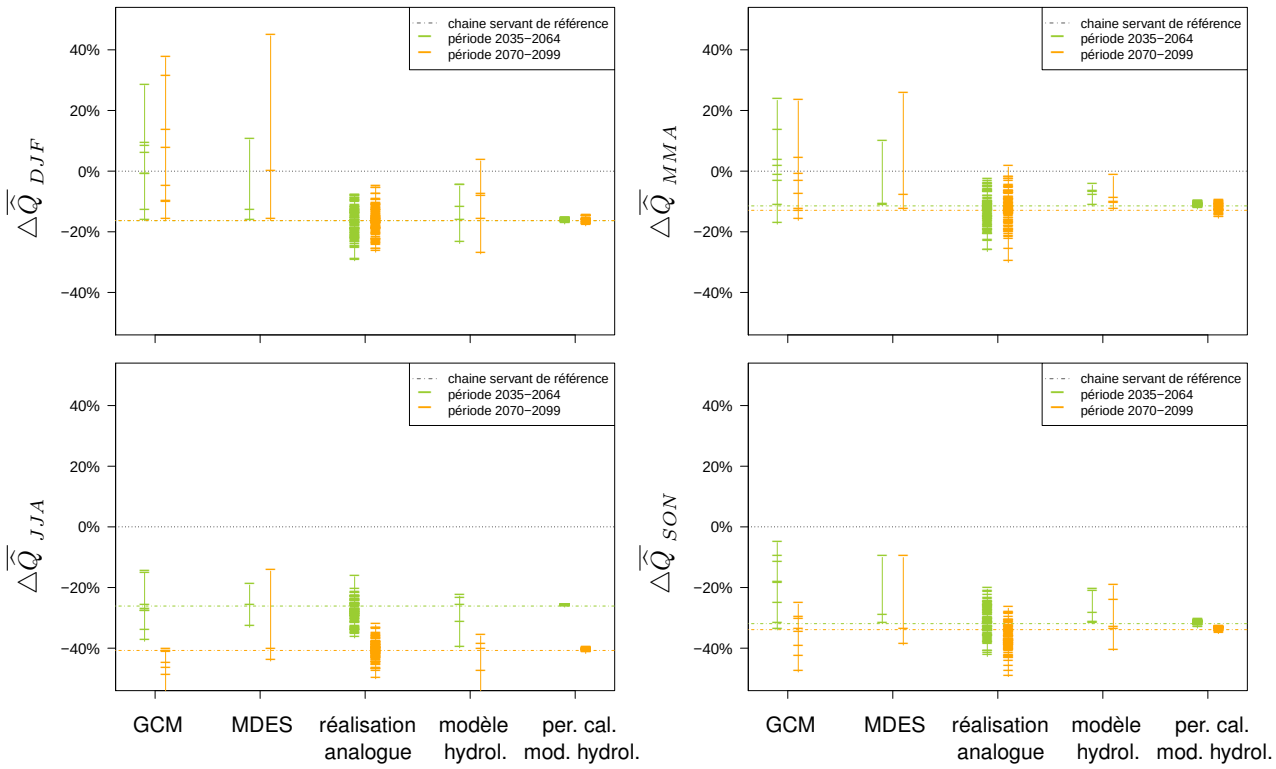


FIGURE 4.10.: Écarts sur les $\overline{\triangle\widehat{Q}}_{saison.}$ projetés avec *GR4J* sur *la Durance à Embrun*, lorsque chaque maillon est considéré indépendamment

4.4.4. Résultats complémentaires pour d'autres modèles

La figure 4.8 semble nous indiquer que la sensibilité des projections émises par *GR4J* au choix de la période de calage est nulle dès lors que ces projections sont exprimées relativement à la situation simulée en temps présent. Face à ce résultat, nous avons recommencé tout notre exercice de projection long terme en changeant le modèle hydrologique utilisé dans la chaîne de référence. Les études de sensibilité sur les projections absolues montrent des résultats similaires à ceux présentés ci-dessous avec la chaîne configuration utilisant *GR4J*. Nous avons placé dans en annexe les figures montrant les dispersions obtenues sur les régimes liées au choix de la réalisation analogue et au choix de la période de calage du modèle hydrologique (*cf.* annexe F, pages 309-310). L'enseignement principal que nous retiendrons de ces figures est un niveau de dispersion équivalent entre ces deux sources. Passons désormais aux études de sensibilité menées sur les projections relatives. Ces analyses sont illustrées par la figure 4.11 qui montre les dispersions des $\triangle\overline{Q}$ pour ces nouvelles chaînes.

Nous constations figure 4.8 une quasi-disparition de la sensibilité au calage de *GR4J* lors du passage de l'étude des projections absolues aux projections relatives. Cette "diminution" de la sensibilité est moins prononcée pour les autres modèles mais reste visible. *Mordor6* présente le

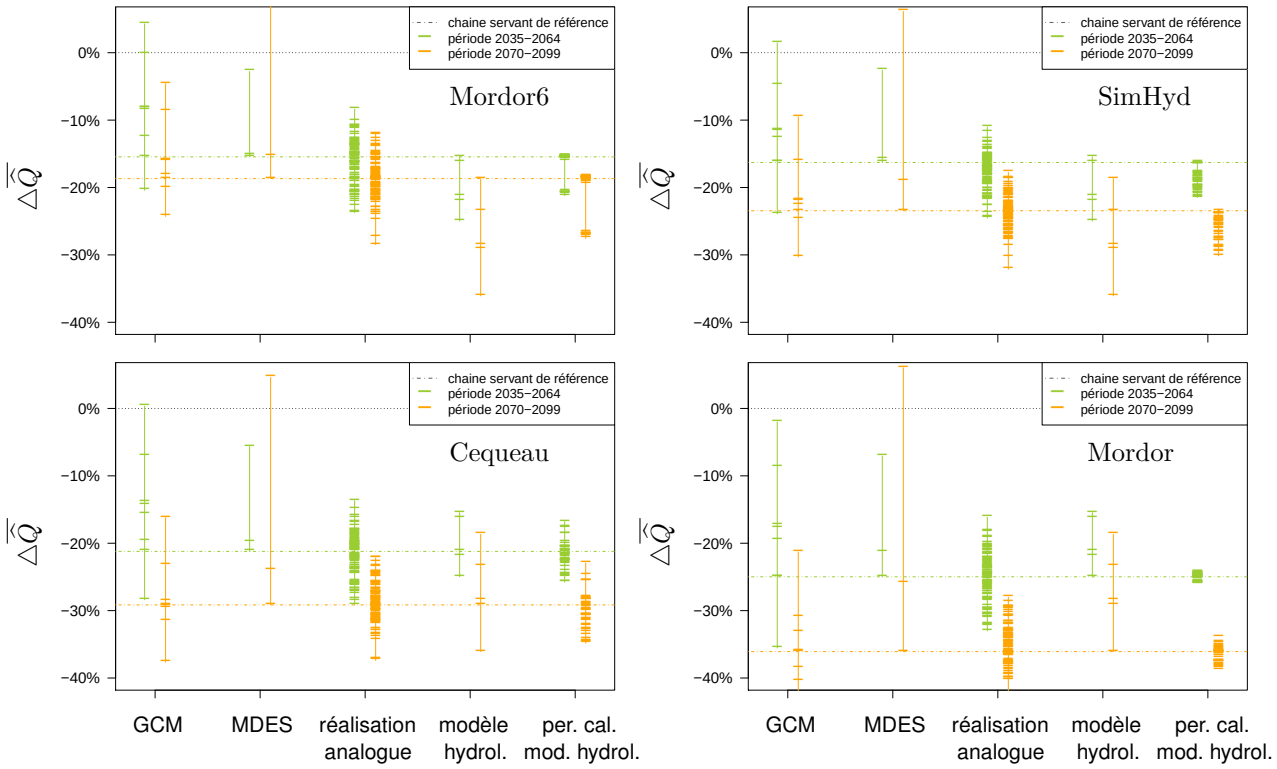


Figure 4.11.: Écarts sur les $\triangle\overline{Q}$ projetés avec *Mordor6*, *SimHyd*, *Cequeau* et *Mordor* sur *la Durance à Embrun* lorsque chaque maillon est considéré indépendamment

Conditions de test identiques à celles utilisées pour construire la figure 4.8, seul le modèle hydrologique étant interverti.

plus de similarités de comportement avec *GR4J*. Bien que quelques jeux de paramètres se détachent du lot, la grande majorité des calages de *Mordor6* conduisent à des $\triangle\overline{Q}$ sensiblement identiques. Une dispersion plus importante est obtenue pour les différents calages des modèles *SimHyd* et *Mordor*. Elle reste toutefois deux fois plus faible que celle liée au choix de la réalisation analogue tandis que des niveaux d'incertitudes similaires étaient perçus lors de l'analyse des projections absolues (*cf.* figures en annexe, pages 309-310). Nous pointerons toutefois un comportement particulier pour *SimHyd* dont la distribution n'est pas centrée autour de la chaîne de référence. Ce décalage révèle que les calages sur différentes sous-périodes conduiront à projeter des débits plus faibles qu'un calage sur toutes les observations (indépendamment des caractéristiques climatiques en jeu). Cette situation résulte vraisemblablement des difficultés de calage de *SimHyd* que nous avons pu également identifier sur quelques bassins lors d'autres analyses. Enfin, le modèle *Cequeau* présente la dispersion la plus importante avec des amplitudes se rapprochant de celles liées au choix de la réalisation analogue.

Pour résumer, la faiblesse de sensibilité des projections relatives au choix de la période de calage observée pour *GR4J* se retrouve en partie sur les autres modèles testés, même si la structure du modèle semble influencer celle-ci.

4.4.5. Rappels sur la signification des faisceaux de projections illustrés

Comme nous l'évoquions en introduction, les analyses menées dans ce chapitre visent à apprécier la sensibilité des projections à divers choix de modélisation. Si les résultats obtenus montrent des enveloppes de projections, celles-ci ne représentent pas des marges incertitudes à proprement parler puisqu'elles ne couvrent que la partie des incertitudes estimée par l'approche multi-modèles (*ex. multi-GCM, multi-MDES, multi-calages du modèle hydrologique...*). C'est donc volontairement (pour limiter les mauvaises interprétations) que nous n'avons pas présenté de résultats combinant toutes les dispersions estimées ici. Notre exercice vise essentiellement à comparer quelques-unes de ces sensibilités pour évaluer l'importance relative de celle qui nous intéresse : le choix de la période de calage du modèle hydrologique. Une estimation plus complète des incertitudes devrait prendre également en compte les erreurs inhérentes à chacun de ces modèles (quand celles-ci peuvent être appréciées). Sur ce point, rappelons d'ailleurs les éléments discutés dans la section 4.1.3, sur le fait que certains résultats issus de nos analyses *GSST* ne sont pas valorisés ici : les constatations d'une augmentation des erreurs de simulation avec l'écart de conditions climatiques subi lors du transfert de paramètres. En effet, nos analyses des problèmes d'extrapolation climatique des paramètres pourraient être utilisées pour ajouter des marges d'erreur supplémentaires autour des dispersions illustrées ici. Pour différentes raisons (*cf.* section 4.1.3) nous avons néanmoins préféré nous contenter d'une étude de la dispersion des débits projetés en lien avec le choix de la période de calage du modèle hydrologique, bien qu'elle ne mette en avant qu'une partie des problèmes de robustesse identifiés au chapitre 3.

4.5. Apports de ces études de sensibilité pour nos travaux en temps présent

4.5.1. Questions soulevées par nos études de sensibilité des projections

Les résultats discutés dans ce chapitre ont été obtenus sur un unique cas d'étude (*la Durance à Embrun*) mais ils n'en restent pas moins troublants. La sensibilité des projections absolues aux conditions de calage du modèle hydrologique est bien visible et se positionne en accord avec les résultats de nos tests de transfert menés précédemment (*cf.* résultats issus du *GSST* présentés dans le chapitre 3). Pourtant, lorsque les simulations futures ne sont plus considérées sous forme absolue mais exprimées relativement à leurs équivalents en temps présent, cette sensibilité tend à disparaître. En simplifiant à l'extrême ce raisonnement et en généralisant des constatations surtout obtenues pour *GR4J-CemaNeige* sur un bassin versant, ces différences notables de sensibilité peuvent conduire à établir les deux points suivants : (a) des problèmes de robustesse des modèles hydrologiques existent en absolu ; (b) ces problèmes résultent en fait d'erreurs de simulation en partie homogènes et stationnaires (sans quoi elles ne disparaîtraient pas aussi nettement lorsqu'elles sont exprimées relativement).

Cette situation soulève un certain nombre de questions quant aux mécanismes à l'origine d'une telle situation. La recherche de réponses à ces questionnements constituera le cœur de notre chapitre 5. Nous y plongeons encore davantage dans l'analyse des simulations mais aussi des valeurs de paramètres des modèles hydrologiques afin de comprendre l'origine de ces comportements. Avant cela, reprenons nos résultats du chapitre 3 pour les analyser sous un nouvel angle de vue, inspiré par les constatations faites dans ce chapitre :

Tandis que les $\triangle\overline{Q}$ projetés semblent indifférents ici au choix de la période de calage, qu'en est-il des $\triangle\overline{Q}$ simulés sur nos sous-périodes historiques ?

Pour y répondre, nous pouvons repartir des tests menés dans le cadre du *GSST* sur notre échantillon de bassins :

1. Nous considérons divers couples de sous-périodes A et B, faisant partie des périodes historiques utilisées au chapitre 3 pour tester nos modèles. Pour simplifier, nous reprendrons nos sous-périodes de 10 ans, construites par décalage d'une fenêtre glissante.
2. Nous considérons divers jeux de paramètres obtenus par calage des modèles sur différentes sous-périodes temporelles. Là encore, nous repartons de la bibliothèque à notre disposition suite à la procédure *GSST* (jeux notés $\theta_{SP[i]}$).
3. Pour chaque couple de périodes A et B et chaque jeu de paramètres, nous pouvons calculer un écart de module simulé $\triangle\overline{\widehat{Q}} = \left(\overline{\widehat{Q}}_B[\theta_{SP[i]}] - \overline{\widehat{Q}}_A[\theta_{SP[i]}] \right) \ / \ \overline{\widehat{Q}}_A[\theta_{SP[i]}]$.
   La comparaison des nombreux $\triangle\overline{\widehat{Q}}$ ainsi obtenus devrait nous dire si le comportement constaté sur les projections long terme est bien réel, *c.-à-d.* si l'expression des simulations sous forme relative révèle une similarité des situations entre différents calages temporels.

4.5.2. Un nouveau regard sur nos tests en temps présent

Reprenons pour commencer le découpage en sous-périodes de 10 ans et les calages des modèles que nous avions réalisés sur *la Durance à Embrun* lors de nos tests avec la procédure *GSST*. Comparons désormais les écarts relatifs de module entre toutes ces sous-périodes, en considérant parallèlement les écarts observés ($\triangle\overline{Q}$) et les écarts simulés ($\triangle\overline{\widehat{Q}}$). Selon les notations ci-dessus, ces écarts s'écrivent $\triangle\overline{Q} = (\overline{Q}_B - \overline{Q}_A) / \overline{Q}_A$ et $\triangle\overline{\widehat{Q}} = (\overline{\widehat{Q}}_B[\theta_{SP[i]}] - \overline{\widehat{Q}}_A[\theta_{SP[i]}]) / \overline{\widehat{Q}}_A[\theta_{SP[i]}]$.

Deux analyses sont alors réalisées :

1. La comparaison entre les écarts simulés et les écarts observés. Pour cette comparaison, nous utilisons le jeu de paramètres calé sur toute la chronique disponible ($\theta_{PT=Pér.Totale}$).
2. La comparaison entre les écarts simulés par les paramètres calés sur nos sous-périodes ($\triangle\overline{\widehat{Q}}[\theta_{SP=Sous-Période}]$) et d'autres écarts simulés à l'aide d'un jeu de référence que nous pendrons égal à θ_{PT}.

La première comparaison ($\triangle\overline{\widehat{Q}}[\theta_{PT}] = f(\triangle\overline{Q})$) nous renseigne sur la capacité générale du modèle à reproduire les débits moyens observés. La seconde comparaison ($\triangle\overline{\widehat{Q}}[\theta_{SP}] = f(\triangle\overline{\widehat{Q}}[\theta_{PT}])$) répond à la question que nous nous posons ici : Des jeux de paramètres qui sont différents et conduisent à des simulations différentes peuvent-ils aboutir à des $\triangle\overline{\widehat{Q}}$ similaires pour un couple de sous-périodes ?

Les résultats de ces comparaisons sur *la Durance à Embrun* sont fournis sur les figures 4.12 et 4.13, où les comportements de différents modèles hydrologiques sont étudiés. Sur les nuages placés à gauche, la dispersion des points met en évidence les difficultés des modèles à reproduire les variations moyennes de débits entre des périodes de 10 ans de ce bassin. Tous les modèles considérés présentent des comportements très similaires, à l'exception de *Mordor* qui se démarque légèrement. Ces différences ne seront toutefois pas discutées ici, notre intérêt se portant principalement sur les nuages placés à droite, où le degré de resserrement des points autour de la bissectrice révèle l'importance des similarités des $\triangle\overline{\widehat{Q}}$. Ces similarités sont extrêmement prononcées pour *GR4J* couplé à *CemaNeige* (que ce dernier soit calé ou non) mais se retrouvent également pour d'autres modèles. Somme toute, nous retrouvons ici les comportements observés lors de l'étude des projections à long terme (*cf.* figures 4.8 et 4.11). De faibles niveaux de dispersion sont obtenus pour *GR4J*, *Mordor* et *Mordor*6, tandis que *SimHyd* et *Cequeau* présentent les niveaux les plus élevés.

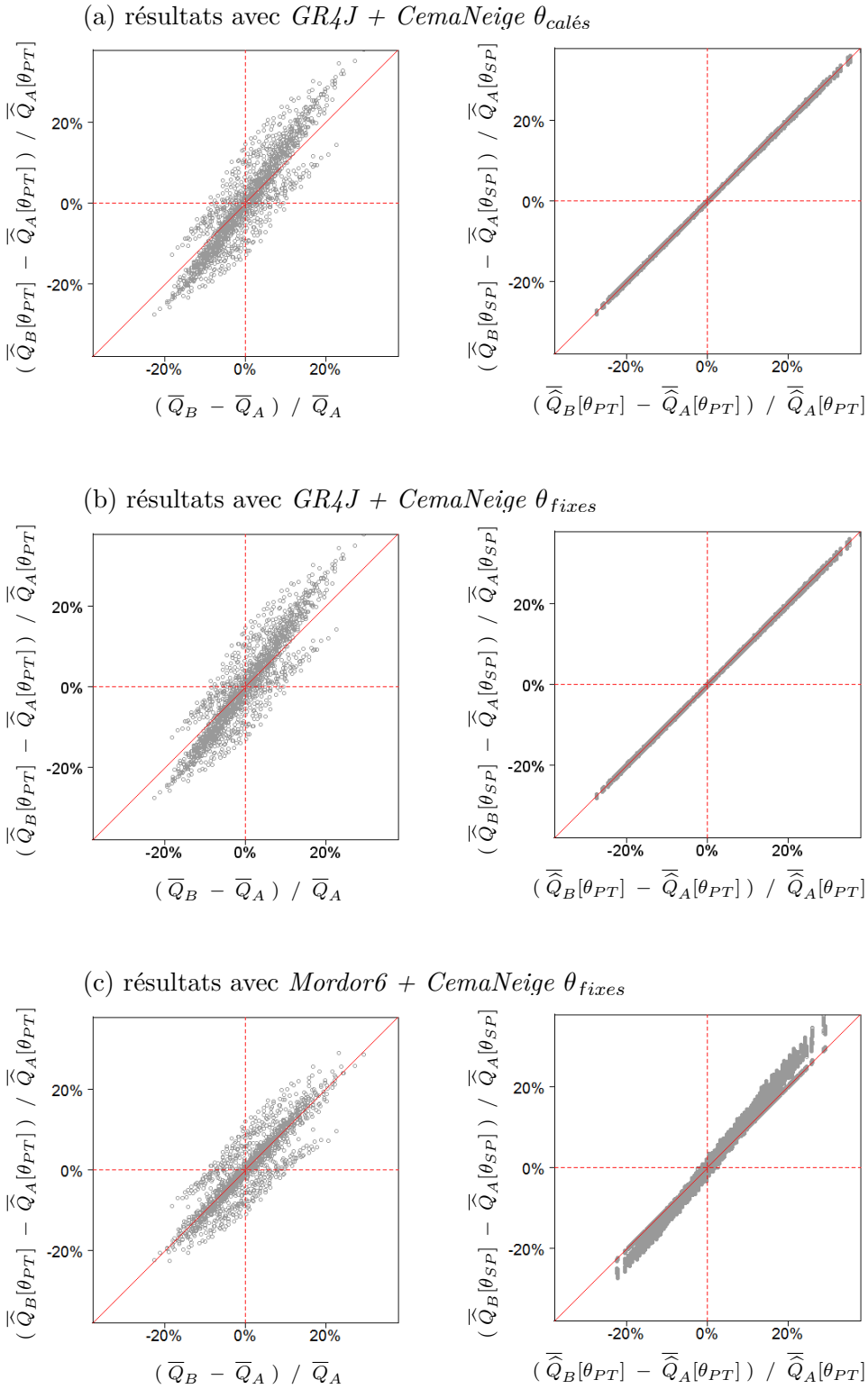


Figure 4.12.: Comparaison des écarts de modules observés et simulés par différents modèles sur les périodes historiques de *la Durance à Embrun* (modèles 1 à 3)

Rq. GR4J et Mordor6 sont alimentés ici par des données SPAZM et une ETP d'Oudin et ils sont couplés à CemaNeige (θ_{neige} maintenus fixes ou calés selon le cas).

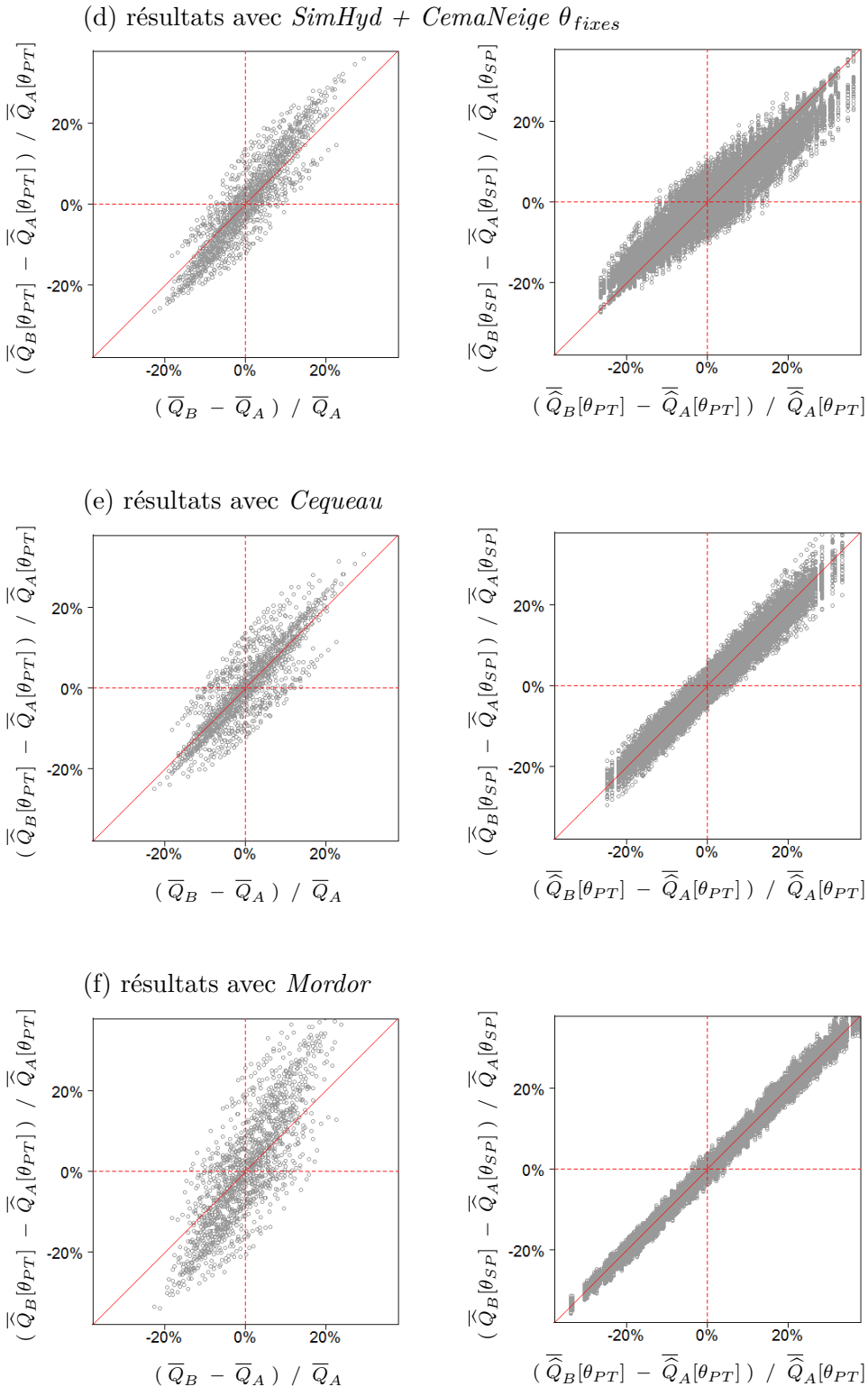


FIGURE 4.13.: Comparaison des écarts de modules observés et simulés par différents modèles sur les périodes historiques de *la Durance à Embrun* (modèles 4 à 6)

Rq. SimHyd est alimentés ici par des données SPAZM et une ETP d'Oudin. Il est couplé à CemaNeige (θ_{neige} maintenus fixes). Cequeau et Mordor sont aussi alimentés par des données SPAZM mais calent chacun leur propre ETP et disposent de leur propre module neige.

Lorsque ces représentations sont répétées sur tous les bassins de notre échantillon, nous constatons que cette similitude des variations simulées de débits entre différents jeux de paramètres n'est pas spécifique au bassin de *la Durance à Embrun*.

La figure 4.14 montre les résultats agrégés d'une comparaison sur un groupe de 20 bassins pour lesquels les modèles *GR4J* et *Cequeau* ont été utilisés suivant la procédure *GSST*. Les nuages placés à gauche révèlent la tendance de ces modèles à simuler des variations de débit moyen plus faibles que celles observées, avec des erreurs qui sont en moyenne plus faibles pour *Cequeau* sur ces bassins. Les nuages placés à droite sur la figure 4.14 permettent la comparaison des $\triangle\overline{\widehat{Q}}$ et illustrent la similitude des variations relatives simulées entre les différents jeux de paramètres. Compte tenu du nombre très important de situations évaluées, ces nuages sont nécessairement plus diffus que lorsqu'ils sont construits sur un bassin, mais l'utilisation des boîtes à moustaches fait ressortir un resserrement particulièrement net des points autour de la bissectrice pour les deux modèles. Pour *GR4J*, une situation identique est obtenue lorsque tous les bassins de notre échantillon sont considérés (*cf.* figure 4.15).

Ces constatations éveillent bien évidemment notre attention et nous encouragent à approfondir notre diagnostic pour comprendre les mécanismes pouvant conduire à une telle similitude des $\triangle\overline{\widehat{Q}}$, tandis que les simulations absolues diffèrent.

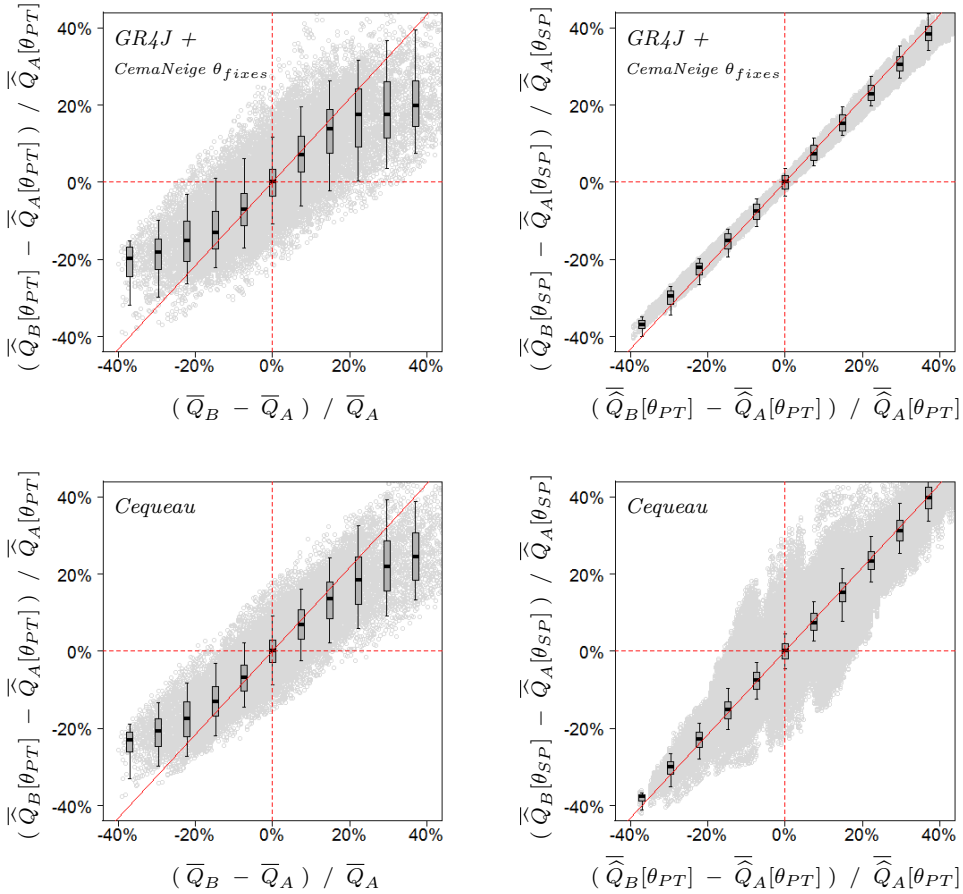


FIGURE 4.14.: Comparaisons aggrégées des écarts de modules observés et simulés par *GR4J+ CemaNeige* θ_{fixes} et *Cequeau* sur les périodes historiques d'un groupe de 20 bassins

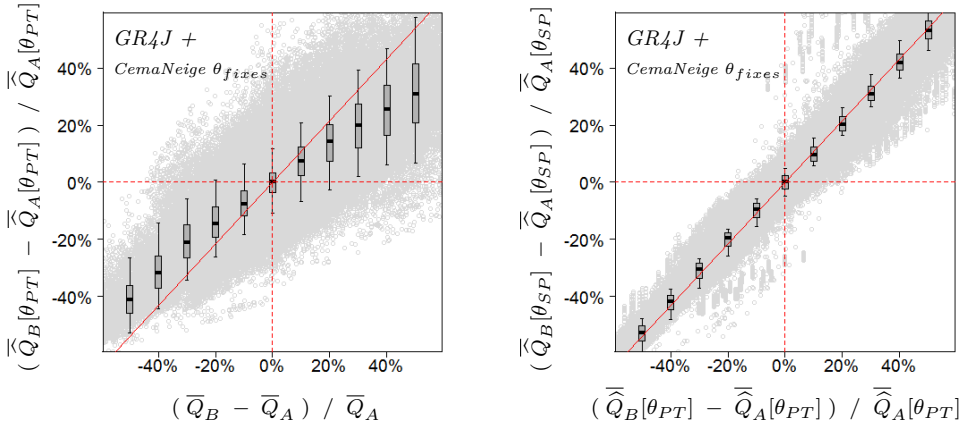


FIGURE 4.15.: Comparaisons aggrégées des écarts de modules observés et simulés par *GR4J+ CemaNeige* θ_{fixes} sur les périodes historiques de tout l'échantillon (569 bassins)

Rq. GR4J-CemaNeige (θ_{neige} fixes) est alimenté par des données SPAZM, une ETP d'Oudin. Cequeau est aussi alimenté par des données SPAZM mais cale sa propre ETP et son module neige. Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Partie III

Seconde phase du diagnostic

Compréhension et tentatives de remédiation

5

Recherche des causes du manque de robustesse des modèles hydrologiques

L'enthousiasme sans faille du chercheur...

Calvin and Hobbes, 03/06/1995, Bill Watterson©

5.1. Introduction

5.1.1. Quelques rappels des travaux menés jusqu'ici

La robustesse des modèles hydrologiques est évaluée à travers des séries de tests de transfert entre périodes indépendantes : les *split-sample test*s, préconisés par *Klemeš* (1986). Nous retrouvons assez logiquement une augmentation des erreurs de simulation lors du passage d'une période de calage à une période de validation (*c.-à-d.* l'utilisation d'un jeu de paramètres non optimal). L'importance de cette augmentation d'erreur varie selon le bassin et selon les périodes considérées pour le test de transfert. Dans un certain nombre de cas, nous avons détecté des liens de corrélation entre les amplitudes d'augmentation d'erreurs et les conditions de transfert des paramètres du modèle, plus précisément les différences de caractéristiques climatiques entre les périodes de calage et de validation. Des situations de corrélations en lien avec les écarts de températures et de précipitations moyennes ont notamment été discutées (*cf.* sections 3.3 et 3.4).

Au cours de ces travaux, nous n'avons pas réussi à expliquer l'hétérogénéité des situations rencontrées (*ex.* absence de cohérence géographique ou climatique dans les résultats). Néanmoins, nous avons pu mettre en avant que les situations de manque de robustesse les plus notables étaient liées à d'importants biais de simulation sur les volumes moyens. Nous tentons dans ce chapitre de comprendre les mécanismes conduisant à ces erreurs quasi systématiques.

5.1.2. L'éternel dilemme entre modèle et données

Face à des écarts significatifs entre observations et simulations, l'hydrologue-modélisateur se demande généralement à qui attribuer la responsabilité de ces erreurs. Est-ce la qualité des données, la qualité du modèle ou un mélange des deux qui est en cause ?

Problèmes liés aux modèles

Dès qu'une difficulté de simulation est rencontrée par un modèle, il est tentant pour le modélisateur de rejeter la faute sur les données. Cela est d'autant plus vrai que les métiers de collecte et constitution de base d'observations sont souvent dissociés du travail de modélisation hydrologique. Pourtant, les défaillances de modélisation peuvent aussi être perçues de manière positive. Elles constituent potentiellement une source d'information pour la compréhension des processus hydrologiques et sur notre manière de les appréhender via un modèle (*cf.* travaux d'*Andréassian et al.* (2010) et leur séminaire sur l'étude de monstres hydrologiques).

Les erreurs de modélisation traduisent une incapacité du modèle à reproduire la (les) variable(s) cible(s) que constituent les débits en un ou plusieurs points d'un territoire d'étude. Ces erreurs peuvent avoir lieu en dépit d'une bonne qualité des données utilisées pour caler puis alimenter un modèle lors de son utilisation. Elles peuvent résulter d'une inadaptation de la procédure d'optimisation des paramètres ou d'une inadaptation du modèle lui-même, à travers sa structure (ou encore d'une inadaptation de l'hydrologue mais il s'agit là d'un autre débat...). Les choix conceptuels de structure sont multiples et complexes à trancher (*ex.* représentation explicite ou non d'une zone de sol saturée, d'une nappe souterraine, des variations de couverture de sol ou d'un réseau hydrographique). Une structure *légère* sera facile à contraindre mais risque d'être mise en défaut (en calage) sur des cas spécifiques. Une structure *riche* a plus de souplesse pour traiter tous les cas, y compris les situations très particulières, mais il sera sans doute plus difficile de la contraindre d'une façon à assurer sa robustesse (en validation).

Problèmes liés aux données

Les performances des modèles hydrologiques sont impactées par la qualité des données qui les alimentent. Rappelons d'ailleurs que ces entrées des modèles sont elles-mêmes issues de modèles d'estimation (*ex.* estimation de la conversion hauteur-débit à partir quelques jaugeages, interpolation d'un champ de précipitation à partir de mesures localisées, etc.). Ces erreurs peuvent alors être ponctuelles (*ex.* estimation erronée d'une crue en extrapolation ou de l'intensité d'un orage), constantes (*ex.* biais systématique d'estimation de la lame d'eau précipitée sur le bassin) ou tendancielles (*ex.* dérive de la relation hauteur-débit).

Lors de la phase d'utilisation du modèle, des erreurs sur l'estimation des débits (lors d'une forte crue par exemple) entraîneront une diminution des performances du modèle, et ce même si les débits modélisés sont proches des débits *réels* (et loin des débits observés erronés). Par ailleurs, des erreurs sur les forçages se propageront au sein du modèle et, malgré un lissage partiel, se traduiront en erreurs sur les débits simulés. Le type d'erreurs commises sur les données lors de cette phase d'utilisation (*ex.* estimation erronée sur un évènement ou biais systématique sur une période) déterminera l'ampleur des variations sur les critères de performance du modèle (*ex.* diminutions localisées ou étendues).

Les problèmes de qualité des données lors de la phase de calage d'un modèle peuvent impacter les performances de celui-ci de diverses façons. Certaines erreurs n'auront pas de conséquences visibles sur la qualité des simulations réalisées, notamment si elles demeurent localisées ou concernent des pas de temps ayant peu d'impact sur les variations du critère de calage utilisé, ou encore si elles sont systématiques mais stationnaires (*Oudin et al.*, 2006b). Sur ce dernier point, en effet, les modèles conceptuels intègrent souvent un paramètre qui pourra servir à compenser un biais systématique sur l'estimation du champ de précipitations par exemple. Certaines erreurs ponctuelles ou les erreurs tendancielles influenceront par contre la sélection

des paramètres avec des conséquences potentiellement plus néfastes. Qu'elles se situent au niveau des forçages ou des débits, ces erreurs peuvent en effet conduire à l'identification de valeurs inappropriées des paramètres du modèle. Des valeurs permettant une adéquation entre les simulations et les observations erronées sont alors choisies au détriment d'autres valeurs permettant une meilleure représentation de la relation entre les forçages *réels* et les débits *réel*s.

Mentionnons enfin une dernière situation que nous ne classerons pas avec les problèmes de qualité des données mais qui y est apparentée : le manque de données informatives sur certains processus pourtant importants. Nous évoquons ici ces situations où le modèle aurait la capacité de représenter certains processus, mais où les données qui l'alimentent ne permettent pas la sélection des paramètres adéquats lors du calage. Cette situation est notamment rencontrée lorsque les pas de temps ou les pas d'espace des données d'entrée ne permettent pas de restituer certains processus. Nous citerons par exemple le cas des bassins très réactifs concernés pas des orages localisés qui nécessitent une forte résolution temporelle et spatiale des données, sans laquelle le signal entrant dans le modèle est trop lissé pour permettre une représentation des phénomènes réels. À l'inverse, nous mentionnerons les bassins influencés par les vidanges et recharges lentes d'une nappe phréatique qui nécessitent, pour être captées par un modèle, une période de calage relativement longue et, selon les cas, une connaissance des processus sur une zone plus large que le seul territoire du bassin versant.

5.2. Recherche d'explications du côté de la modélisation

5.2.1. Sensibilité des résultats au choix du modèle hydrologique

Face à l'existence possible d'erreurs de modélisation, la stratégie adoptée a été celle du recours à une approche multi-modèles. En réalisant les mêmes tests sur des modèles de structures et de niveaux de complexité différents, nous pouvons ainsi discerner dans nos résultats si le choix du modèle modifie sensiblement les conclusions et, si oui, quelles structures permettent une meilleure robustesse temporelle face à un climat changeant.

Les résultats présentés tout au long de ce rapport montrent des différences relativement faibles entre modèles vis-à-vis des comportements étudiés. Une analyse plus détaillée des chroniques mettrait, bien entendu, en avant des différences dans les variables simulées. Néanmoins, nos travaux portaient initialement sur l'étude des baisses de performances liées au transfert des paramètres, que nous avons pu relier à des erreurs sur le bilan hydrologique annuel dans une majorité de cas. Nous nous sommes ensuite concentrés sur la compréhension des mécanismes conduisant à ces biais, et n'avons, par conséquent, pas affiné l'étude des différences entre modèles sur d'autres aspects que ces erreurs sur le volume moyen. Si ce choix peut paraître

relativement fort, précisions toutefois que la capacité d'un modèle à reproduire le bilan hydrologique annuel semble un critère minimal que nous sommes en droit d'attendre d'un modèle hydrologique. Face aux amplitudes des erreurs constatées sur les volumes moyens lors de nos tests, l'étude d'autres critères plus raffinés nous paraît secondaire (dans un premier temps tout du moins). Rappelons enfin que les travaux de cette thèse visent aussi à alimenter les réflexions concernant les erreurs de modélisation dans les études d'impact du changement climatique, dans lesquelles les estimations du module à différents horizons constituent l'un des objectifs majeurs.

Plus spécifiquement, la similitude des comportements entre les différents modèles testés a été (et sera) étudiée à plusieurs occasions durant nos travaux. Les comparaisons suivantes peuvent notamment être citées :

- mise en avant des similitudes de comportement à travers le tableau 3.2 (page 74) qui montre le recouvrement des sous-groupes de bassins concernés par des corrélations entre △erreurs en validation et △conditions entre les périodes de calage et de validation
- étude des similitudes de comportement via les analyses de nuages de points (*cf.* figures 3.2 et 3.8, pages 77 et 86) et dans les visualisations associées sur quelques exemples illustratifs (*cf.* figures pages 82-83 et pages 91-92)
- étude des similitudes de comportement (présentée plus loin) à travers les analyses des courbes de biais glissants (*cf.* figure 5.19)

5.2.2. Impact de la longueur de la période de calage du modèle

Introduction

Lorsque les travaux présentés dans cette thèse ont été initiés, nous nous sommes largement interrogés sur la longueur des sous-périodes à choisir dans la procédure *GSST* (*cf.* section 2.4.2). Cette longueur de calage devait être suffisamment longue pour permettre une identification adéquate des paramètres du modèle et suffisamment courte pour permettre un nombre important de tests entre des périodes indépendantes et contrastées (notamment climatiquement).

Afin d'apprécier l'impact qu'aurait ce changement de longueur de sous-périodes de calage, nous avons souhaité reproduire les mêmes tests de transfert pour une sélection de bassins, en considérant des sous-périodes de 20 ans (contre 10 ans précédemment). Pour que le nombre de tests de transfert demeure suffisant, nous avons identifié les bassins de notre échantillon disposant de plus de 50 années d'observations, sur lesquels nous avons reproduit la procédure de tests multiples. Pour faciliter les analyses, nous avons choisi de nous concentrer sur les bassins qui faisaient partie du sous-groupe où nous avions identifié des liens entre robustesse et △températures lors des transferts des paramètres de *GR4J* (*cf.* section 3.3). Cette double

sélection conduit à la formation d'un groupe de 31 bassins sur lesquels nous avons observé des corrélations entre $\triangle\varepsilon_R$ et $\triangle\overline{T}_{D\to R}$ lors du *GSST* sur 10 ans de calage et où la longueur des séries permet un nombre élevé de tests pour un nouveau découpage en sous-périodes de 20 ans. Une simple comparaison des nuages de points $MRC_{D\to R} = f(\triangle\overline{T}_{D\to R})$ ou $\overline{\widehat{Q}_R}[\theta_D]/\overline{Q_R} = f(\triangle T_{D\to R})$ obtenus pour chaque configuration (calages-validations sur 10 ou 20 ans) permet rapidement de statuer sur l'impact d'une plus grande longueur de calage sur nos résultats.

Comparaisons des résultats

Les nuages de points issus du *GSST* sur notre sélection de 31 bassins sont présentés sur les figures 5.1 et 5.2 pour *GR4J* et *Mordor6*. Le processus de construction des boîtes à moustaches est inchangé, le nuage étant découpé en 10 bandes verticales d'un nombre égal de points (*cf.* section 2.1.3). Le fait d'augmenter la durée des sous-périodes de calage et validation réduit nécessairement les contrastes entre celles-ci, mais ne perturbe pas réellement le travail de comparaison.

Ces nouveaux tests montrent qu'une augmentation de la longueur de la période de calage ne réduit pas les problèmes de transférabilité temporelle des paramètres sur les 31 bassins étudiés ici, qu'ils soient mesurés par le critère *MRC* ou à travers le biais (*cf.* figures 5.1 et 5.2). La responsabilité des problèmes mis en avant par la procédure *GSST* ne peut donc pas être simplement rejetée sur le fait de caler le modèle sur une période trop courte pour pouvoir capter certaines dynamiques lentes du bassin. Il est vrai que la structure parcimonieuse de *GR4J* ne semble pas permettre, à première vue, la représentation de telles dynamiques lentes. Les tests menés avec *Mordor6* donnent cependant les mêmes résultats, tandis que ce dernier dispose de plusieurs réservoirs qui devraient lui permettre de mieux s'en sortir si ces dynamiques lentes étaient en causes (*cf.* schéma de structure, figure 2.10, page 50). À première vue, cet argument ne semble donc pas valide même si des tests avec d'autres structures de modèles pourraient être menés pour conforter ce résultat.

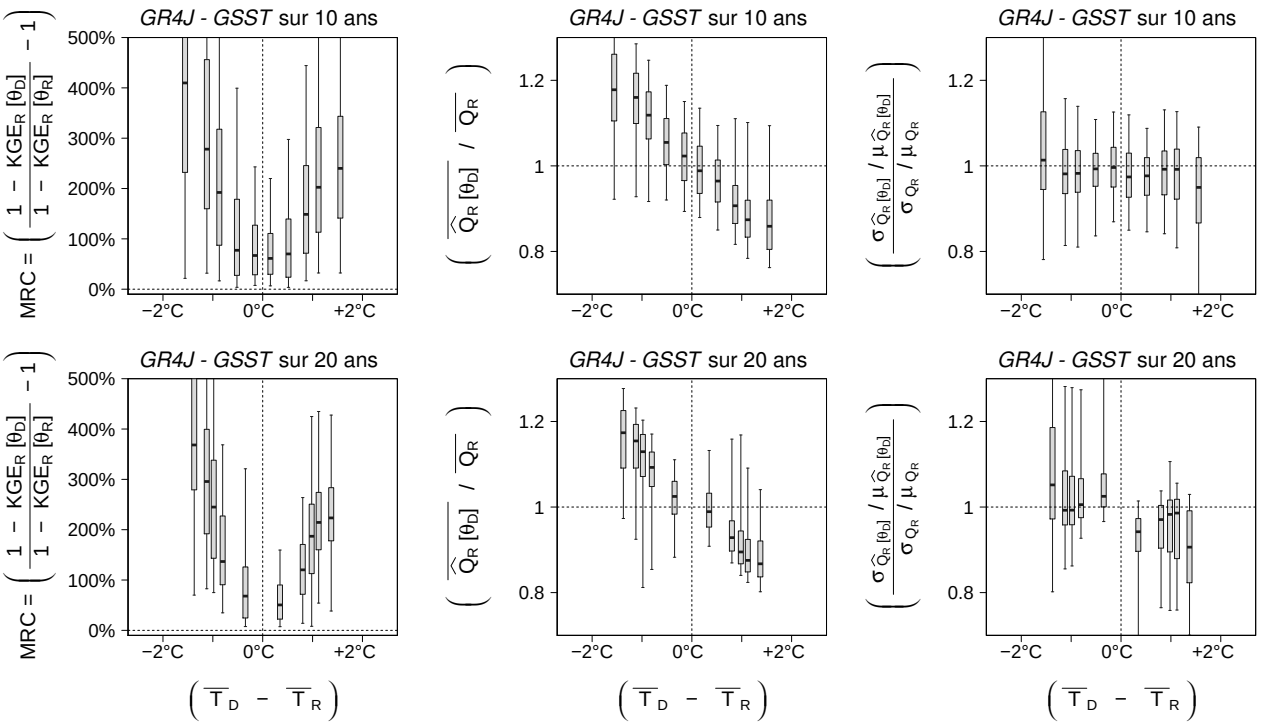


FIGURE 5.1.: Comparaison des erreurs observées lors des transferts de *GR4J* avec des périodes de 10 ans (en haut) et 20 ans (en bas)

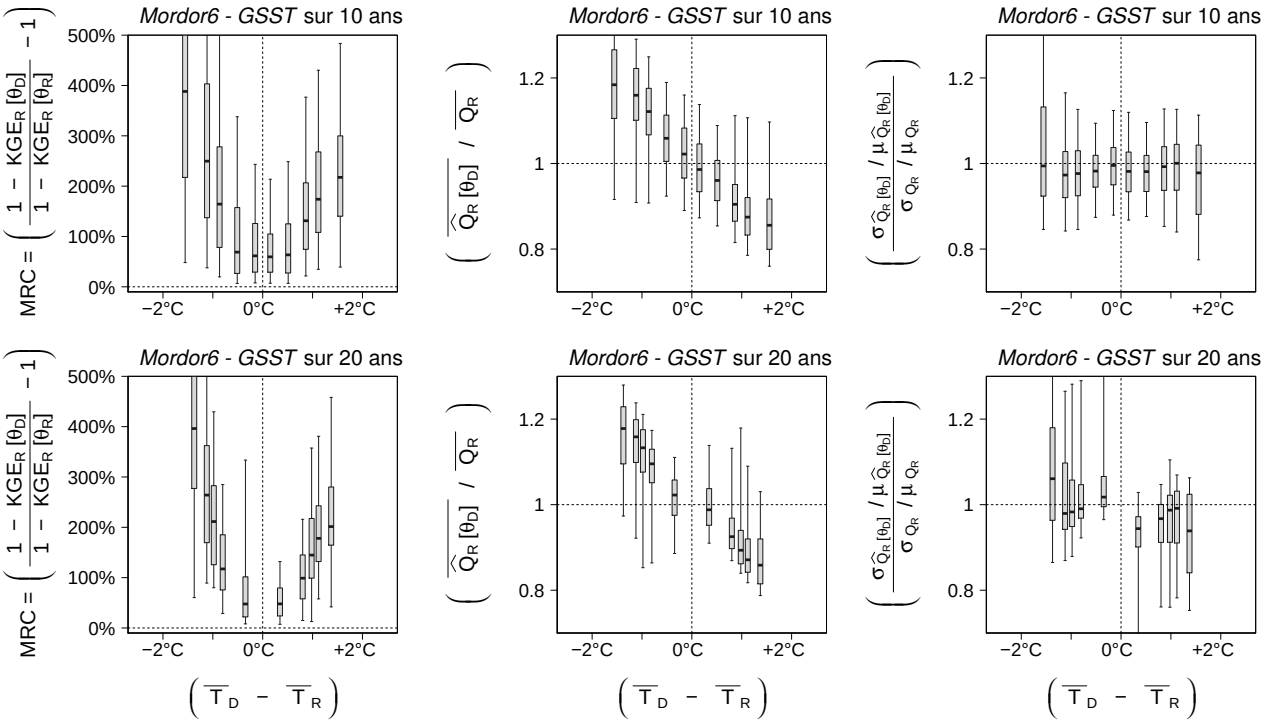


FIGURE 5.2.: Comparaison des erreurs observées lors des transferts de *Mordor6* avec des périodes de 10 ans (en haut) et 20 ans (en bas)

Rq. sur les bassins français, GR4J et Mordor6 sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

5.3. Recherche d'explications du côté de la qualité des données

5.3.1. Sensibilité des résultats à la qualité des chroniques de débits

Malgré des efforts continuels de contrôle de la qualité des séries de débits mises à disposition des utilisateurs, les erreurs y restent potentiellement nombreuses. Notre travail se concentre sur la robustesse des modèles hydrologique dans un climat en évolution et nous avons choisi (dans la mesure du possible) d'écarter en amont des travaux les situations problématiques liées à la qualité des débits. Ainsi, lors de la constitution de notre échantillon de bassins, toutes les séries de débits potentiellement intéressantes furent soumises à un contrôle de qualité simple, seules les séries où aucun problème notable n'avait été identifié étant retenues. Les détails de cette procédure sont fournis dans la section 2.2.2 et l'annexe C.1 et nous n'en rappellerons que les principes ici. Ce contrôle qualité est basé sur la visualisation de courbes de doubles-cumuls de débits entre chaque station hydrométrique et ses plus proches voisines. Cette méthode de détection des erreurs repose sur le postulat que deux séries voisines de bonne qualité ont des cumuls liés par une relation moyenne de proportionnalité ne présentant pas de cassure. L'existence de tronçons de pente moyenne différente ou la présence de ruptures soudaines de cette relation (zones localisées de pente presque nulle ou infinie) sont autant d'indicateurs d'erreurs sur l'une des séries comparées. Malgré sa relative simplicité, nous pensons que cette méthode de contrôle nous a permis d'écarter les cas d'erreurs flagrantes sur les débits, notamment les situations causées par un changement de la courbe de tarage où la présence de données erronées mais non déclarées lacunaires. Quelques exemples de situations écartées de l'échantillon sont disponibles dans l'annexe C.1.

Une évaluation du rôle potentiel des erreurs de débits sur nos interprétations a toutefois été réalisée *a posteriori*. En effet, nous disposons des travaux de *Renard* (2008) ainsi que des retours faits par les DREALs auprès d'Irstea Aix-en-Provence sur la qualité des séries. Nous avons donc pu établir une liste de bassins français sur lesquels ce retour avait été fait et les débits qualifiés de bonne qualité. Une comparaison rapide a été menée pour apprécier la sensibilité des résultats à ce critère de qualité des séries de débits, en considérant d'une part, les résultats obtenus sur le groupe de 120 bassins où $\triangle\varepsilon = f(\triangle\overline{T})$, et d'autre part, les résultats obtenus lorsque nous ne retenons, parmi ces 120, que les 43 bassins présents dans la liste évoquée ci-dessus. Les résultats correspondants sont présentés sur la figure 5.3 et ne mettent en évidence aucune différence globale qui aurait influencé les interprétations faites dans ces travaux.

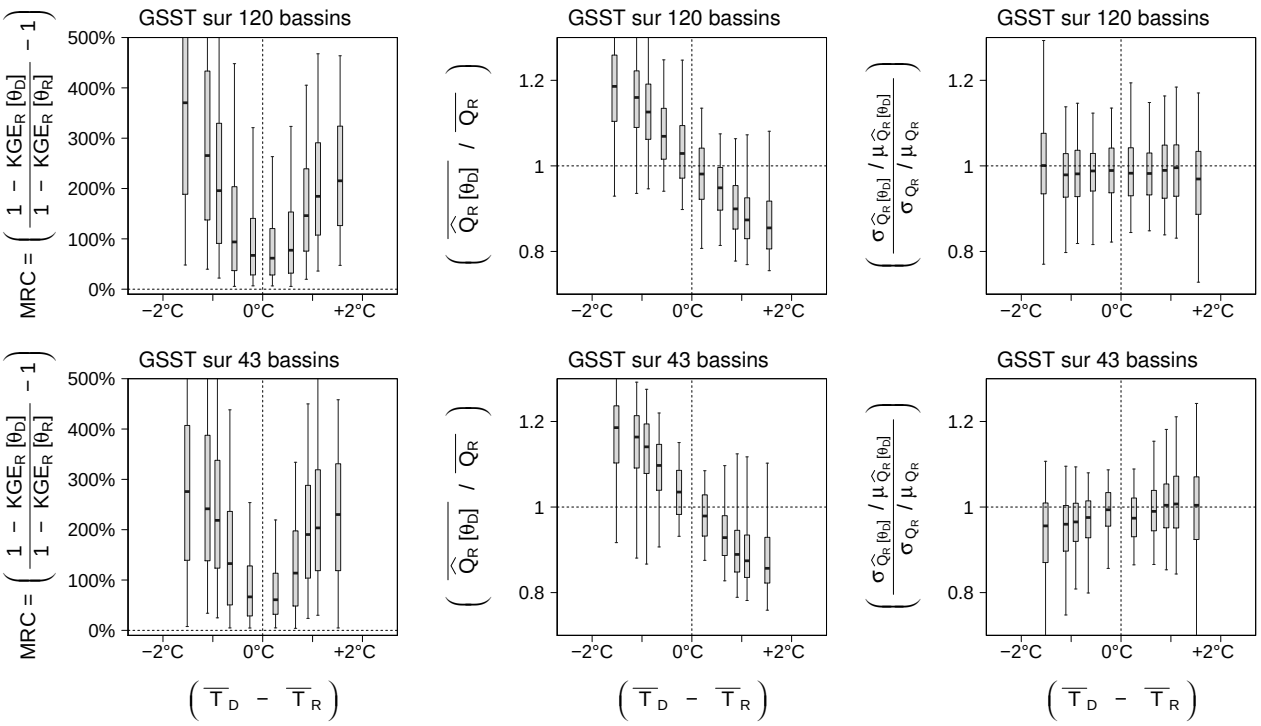


FIGURE 5.3.: Sensibilité des résultats du *GSST* à la qualité des chroniques de débits observées

Sur les bassins français, GR4J est alimenté par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés). Les tests de GSST sont conduits sur des sous-périodes de 10 ans. Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

5.3.2. Sensibilité des résultats au choix des données climatiques

Les erreurs d'estimation des données d'entrées des modèles hydrologiques (*c.-à-d.* précipitations, températures et/ou évapotranspiration potentielle) peuvent avoir des conséquences très néfastes sur la capacité d'un modèle hydrologique conceptuel à simuler des débits en accord avec les observations (*cf.* section 5.1.2).

Ces erreurs peuvent provenir directement de problèmes de mesure ou d'estimation (*ex.* sous-captation par un poste pluviométrique ou formulation inadaptée pour le calcul d'une *ETP*). Elles peuvent également provenir des méthodes d'interpolation et d'extrapolation utilisées pour obtenir des champs de forçages climatiques sur tout le territoire du bassin versant à partir d'observations ponctuelles (*ex.* polygones de Thiessen, krigeage, réanalyses).

Une étude ciblant les problèmes de qualité des mesures pourrait être menée (sur quelques cas) où la stationnarité des erreurs sur chaque appareil serait scrupuleusement suivie. Une telle approche n'a cependant pas été menée car elle ne rentrait pas dans le cadre général de nos travaux dans lesquels de très nombreux cas sont étudiés simultanément. Nous nous reposons donc sur les contrôles de qualité (déjà nombreux) menés par les producteurs de ces données.

À la place, nous nous concentrons directement sur les données qui alimentent nos modèles : les séries issues des modèles de réanalyses climatiques (*ex. SAFRAN* et *SPAZM* en France ou *SILO Data Drill* en Australie). Nous optons pour cela une approche multi-sources (similaire à celle utilisée pour étudier le rôle des erreurs de modélisation) puisque nous avons la chance de disposer de deux bases de forçages distinctes sur le territoire français : *SAFRAN* et *SP*AZM.

Présentation des bases disponibles et retour sur les choix précédents

Depuis le début de nos travaux sur les bassins français, deux bases de données climatiques sont à notre disposition :

- La base *SAFRAN*, créée par *Météo-France*, combine des observations de surface et des analyses de modèles météorologiques pour produire des données horaires de température, humidité, vent, précipitations et rayonnement, interpolées sur une grille de 8 km x 8 km couvrant la France métropolitaine (*Quintana-Seguí et al.*, 2008). Dans le cadre de nos travaux, nous utilisons la version journalière de *SAFRAN* pour laquelle nous disposons également d'un calcul de l'évapotranspiration potentielle de *Penman-Monteith*.
- La base *SPAZM*, créée par *EDF,* est une réanalyse des observations journalières de température et précipitations issue des travaux de *Gottardi* (2009). En repartant des mesures des réseaux de *Météo-France* et d'*EDF*, cette base a été construite dans le but d'améliorer l'estimation des champs de température et précipitations en zone de montagne. Elle fournit ces champs au pas de temps journalier sur une grille de 1 km x 1 km couvrant une large moitié sud de la France (*cf.* section 2.2.2 pour plus de détail).

Dès l'initiation des travaux de cette thèse, nous souhaitions travailler sur un large échantillon de bassins mais avec la volonté particulière d'y inclure des bassins montagneux. Ceux-ci présentent plusieurs avantages. D'une part, les signaux du changement climatique global y semblent particulièrement marqués, ce qui permet des contrastes plus importants entre différentes périodes temporelles, or nous recherchons justement ces contrastes pour notre analyse de capacité d'extrapolation climatique des modèles. D'autre part, le fonctionnement hydrologique de ces bassins montagneux est thermosensible du fait des processus de stockage saisonniers d'eau sous forme de neige, puis relargage lors d'une crue de fonte printanière. Étudier l'impact d'un changement de température sur la capacité des modèles hydrologiques à suivre les modifications de régime d'écoulement est donc très intéressant, notamment pour *EDF* qui utilise de tels modèles pour gérer sa production d'hydroélectricité. Dans ce contexte, notre choix s'est orienté vers la base *SPAZM* qui semble mieux adaptée à fournir des données de précipitations en altitude. Elle nous permet également de disposer de forçages sur des zones de plaines plus faiblement influencées par les reliefs.

Nous souhaitons revenir ici sur ce choix en nous interrogeant sur l'impact de l'origine des données de forçages dans nos interprétations. Bien que les réseaux de mesure utilisés dans la

construction des bases *SPAZM* et *SAFRAN* soient partiellement communs, la méthode de réanalyse reste très spécifique à chacune d'elle. Au-delà de la question des écarts entre les données de ces bases, du point de vue de l'alimentation des modèles, nous souhaitions surtout déterminer si nous aurions obtenu des résultats différents en utilisant la base *SAFRAN* plutôt que la base *SPAZM*.

Configurations testées

De nouveaux tests de transfert sont menés pour évaluer la sensibilité de nos analyses au choix de la base de forçages. Ceux-ci sont conduits avec le modèle *GR4J*, couplé avec le module *CemaNeige*. Dans la continuité des tests menés précédemment, les paramètres de *CemaNeige* ne sont pas calés ici (*cf.* section 3.1.3). *GR4J* est alimenté avec des données de précipitation et d'*ETP* de bassin, cette dernière étant couramment calculée dans nos tests par la formule d'*Oudin et al.* (2005), principalement à partir des séries de température de l'air. Pour ces nouveaux tests, nous comparons également l'influence du changement de la formulation d'*ETP*, puisqu'un calcul de l'évapotranspiration potentielle de *Penman-Monteith* est disponible avec la base *SAFRAN* au pas de temps journalier (*Monteith*, 1965). Précisons enfin que le même découpage en sous-périodes glissantes de 10 ans est utilisé pour ces nouvelles utilisations du *GSST*.

Les trois configurations testées sont les suivantes :

1. *Base de forçages* $SPAZM + ETP_{Oudin}$

   Entrées de *CemaNeige* : précipitations et températures journalières *SPAZM* à 1 km moyennées sur cinq bandes d'altitudes de surfaces égales en entrée de *CemaNeige*.
   Entrées de *GR4J* : lames d'eau journalières de bassin issues de *CemaNeige* et *ETPs* journalières de bassin calculées sur les températures *SPAZM* avec la formule d'Oudin.

2. *Base de forçages* $SPAZM + ETP_{Penman-Monteith}$

   Entrées de *CemaNeige* : précipitations et températures journalières *SPAZM* à 1 km moyennées sur cinq bandes d'altitudes de surfaces égales en entrée de *CemaNeige*.
   Entrées de *GR4J* : lames d'eau journalières de bassin issues de *CemaNeige* et *ETPs* journalières de bassin de *Penman-Monteith* issues de la base *SAFRAN*.

3. *Base de forçages* $SAFRAN + ETP_{Oudin}$

   Entrées de *CemaNeige* : précipitations et températures journalières *SAFRAN* à 8 km moyennées par bassin puis redécomposées sur cinq bandes d'altitudes de surfaces égales en utilisant les gradients altitudinaux moyens issus de *Valéry* (2010).
   Entrées de *GR4J* : lames d'eau journalières de bassin issues de *CemaNeige* et *ETPs* journalières de bassin calculées sur les températures *SAFRAN* avec la formule d'Oudin.

Ces travaux ont été initiés dans l'optique de comparer les données de précipitation et température issues des bases *SPAZM* et *SAFRAN*. Ils pourraient être repris pour cibler la question du choix de la formulation d'*ETP* en ajoutant par exemple une quatrième configuration dans laquelle nous considérerions les précipitations et températures de la base SAFRAN et l'*ETP* de *Penman-Monteith* (également calculée sur des données *SAFRAN*).

La figure 5.4 donne un aperçu des différences constatées entre les données issues de la base *SAFRAN* et celles issues de la base *SPAZM*. Nous y retrouvons notamment des différences quant à l'estimation des précipitations sur les bassins les plus arrosés, les volumes de précipitation estimés étant plus élevés pour *SPAZM* que pour *SAFRAN*. Ces différences ne sont pas surprenantes au regard des travaux *Gottardi* (2009), concepteur de la base *SPAZM*, qui travailla justement à améliorer l'estimation des précipitations en altitude et corriger les sous-estimations constatées sur les reliefs pour la réanalyse *SAFRAN*.

Afin de faciliter l'analyse des résultats du *GSST* pour les trois configurations considérées, nous nous concentrerons ici encore sur une sélection composée de bassins sur lesquels nous avions détecté une corrélation entre les erreurs de modélisation et l'écart de température entre les périodes de calage et validation (*cf.* section 3.3.2). Ces bassins ont l'avantage de constituer des nuages de points marqués par les tendances et ils facilitent ainsi les comparaisons rapides. Un sous-groupe de 93 bassins français où les trois configurations de forçages sont disponibles est donc constitué et utilisé pour mener notre étude de sensibilité au choix de la base de forçages.

Comparaison des résultats

Les résultats obtenus lors de ces tests sont illustrés à la figure 5.5. La situation utilisée dans tous les tests présentés jusqu'ici a été placée au centre (données de précipitations et températures issues de *SPAZM* et formulation d'*Oudin* pour le calcul de l'*ETP*). Les deux autres configurations nouvellement testées sont placées de part et d'autre.

Les graphiques composant les colonnes de gauche et du centre ont en commun l'utilisation des champs de précipitations et température de la base *SPAZM* et diffèrent sur l'*ETP* fournie au modèle. Malgré des différences (parfois grandes) entre les de séries d'*ETP* considérées ici, la forte similitude entre les nuages obtenus révèle que le remplacement de la formulation d'*ETP* ne semble pas impacter les problèmes de robustesse auxquels sont confrontés les modèles sur ces bassins. Les graphiques composant les colonnes de droite et du centre ont seulement en commun la formule utilisée pour l'*ETP* (celle d'*Oudin*), et diffèrent quant aux champs de précipitations et température (*SPAZM* ou *SAFRAN*) qui servent à alimenter le modèle et à calculer l'*ETP*. En ce qui concerne les variations du critère *MRC*, les différences sont plus marquées que lors de la comparaison précédente. Bien qu'une tendance similaire reste visible, la nette augmentation de la hauteur des boîtes indique un nuage plus diffus avec des situations

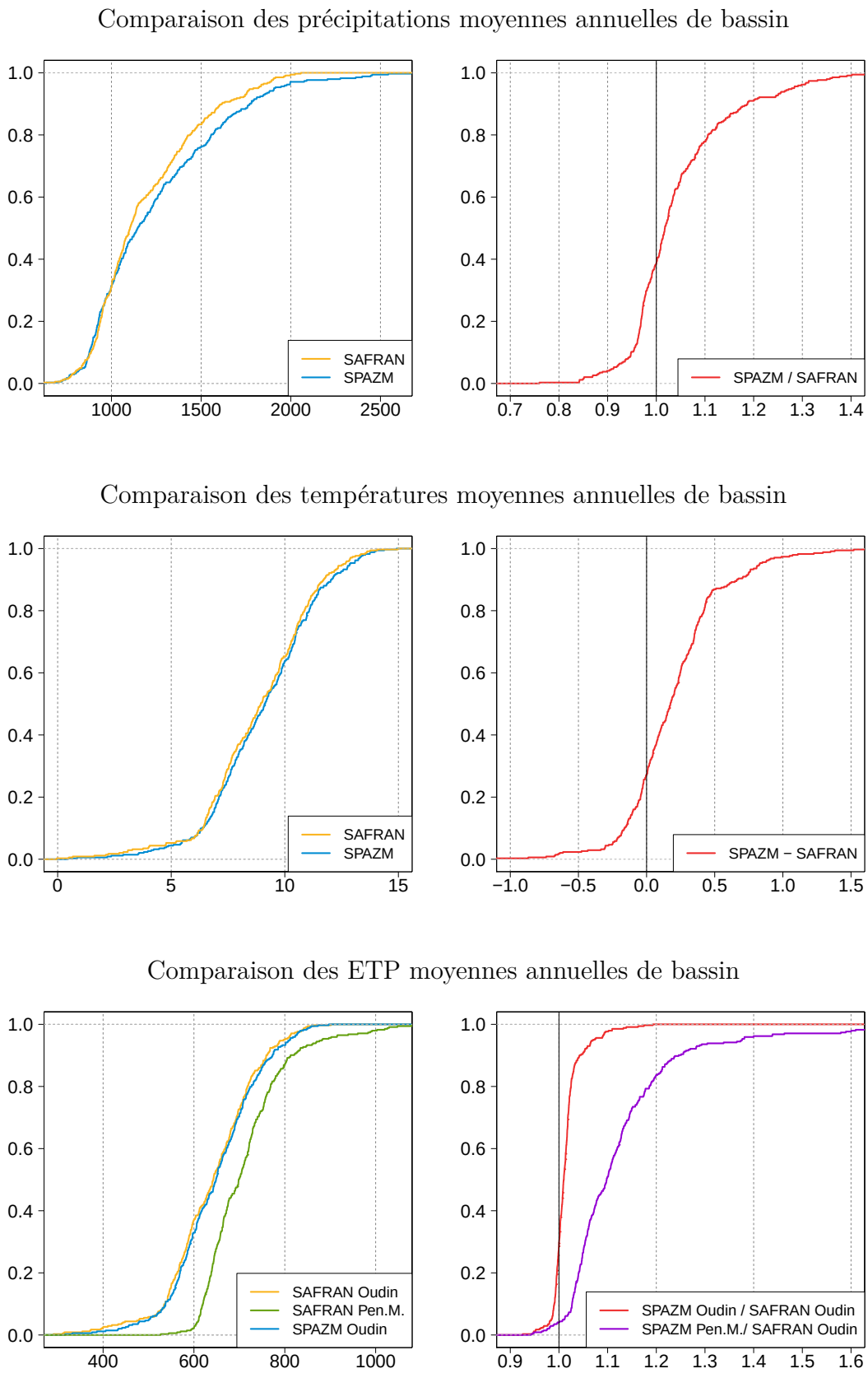


FIGURE 5.4.: Aperçu des différences de données entre les bases *SAFRAN* et *SPAZM* sur les 365 bassins français de l'échantillon

de pertes de performance pouvant être élevées même avec des $\triangle\overline{T}_{SAFRAN}$ proches de 0. Enfin, une augmentation de hauteur des boîtes est visible sur les graphiques concernant les erreurs de volume, mais y est nettement moins prononcée, tandis que la relation entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{T}$ reste d'ampleur similaire.

Discussion

En dépit de données de mesures partiellement communes, nous avons remarqué des différences non négligeables entre les deux bases de forçages françaises à notre disposition (*SPAZM* et *SAFRAN*). Pourtant, le test de changement de base de données mené sur 93 bassins versants n'a pas mis en évidence de différence radicale vis-à-vis des erreurs sur les volumes moyens constatées lors du transfert des paramètres du modèle *GR4J*.

Nous pensions que la méthode de calcul de l'*ETP* pouvait constituer un point explicatif des erreurs de modélisation liée à des transferts entre des périodes de températures moyennes différentes. Parmi nos *a priori*, nous jugions que la formule d'*Oudin*, qui dépend uniquement de la température et du rayonnement extraterrestre, pouvait manquer une partie du signal qu'aurait capté une formule plus complexe telle que celle de *Penman-Monteith* (*Oudin et al.*, 2005; *Monteith*, 1965). Nos comparaisons laissent penser que ce n'est pas le cas mais des travaux complémentaires sur cette question de l'évapotranspiration paraissent maintenant essentiels.

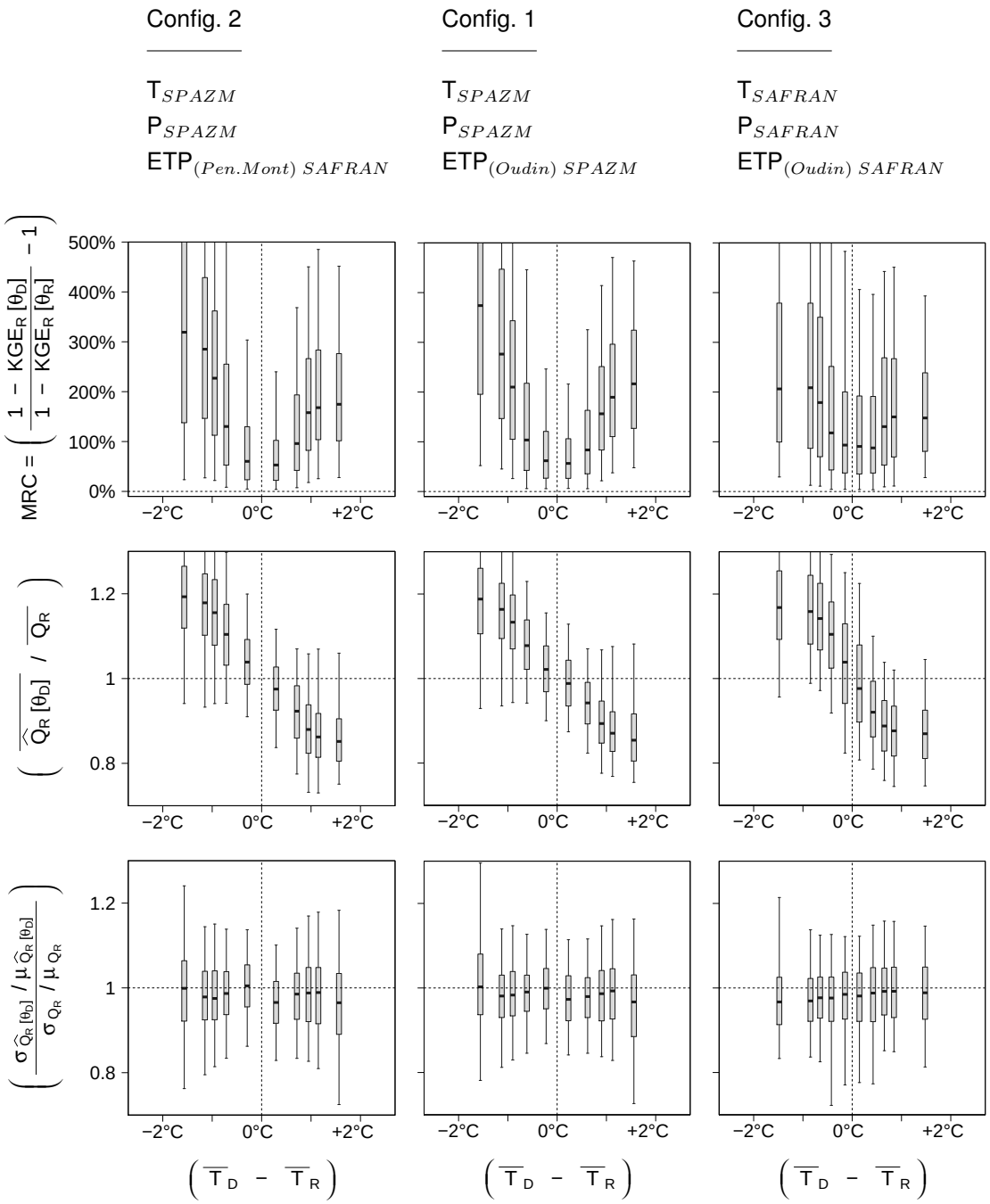


FIGURE 5.5.: Comparaison des pertes de performances lors des transferts entre périodes selon la base de forcages utilisée (SPAZM ou SAFRAN)

Résultats obtenus pour GR4J-CemaNeige (θ_{neige} fixés) sur une sélection de 93 bassins français. Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

5.4. Analyse des dérives temporelles des erreurs de modélisation

5.4.1. Pertinence de cette piste explicative des problèmes de robustesse

La procédure *GSST* fait intervenir de nombreux tests de transfert temporel des paramètres. Or, les analyses des variations d'erreurs de *GR4J* en fonction des conditions de transfert montrent une augmentation moyenne de celles-ci avec l'écart temporel entre les périodes de calage et de validation (*cf.* tableau 3.1, page 71). Dans ce chapitre, nous recherchons les causes de ce manque de robustesse des modèles. Dans ce cadre, nous souhaitons déterminer la nature de ce lien entre les erreurs de modélisation et le décalage temporel subi lors du transfert. Si une réelle corrélation était établie, elle pourrait en effet être le signe que les difficultés rencontrées par les modèles que nous évaluons ici ne sont finalement pas liées aux changements de conditions climatiques d'utilisation de ces modèles. Nous pourrions alors envisager diverses causes explicatives à cette situation, par exemple les dérives dans l'estimation des données alimentant les modèles ou les évolutions lentes de la couverture végétale du bassin.

Quelques mots sur les dérives temporelles dans l'estimation des données

Des situations d'évolution continue des erreurs de mesure des débits peuvent être rencontrées sur des stations hydrométriques sujettes à un détarage progressif. Elles peuvent être causées par des évolutions lentes du lit de la rivière au niveau de la station de jaugeage (érosion ou dépôt sédimentaire, développement de la végétation, etc.) qui entraînent une dérive de validation de la relation hauteur-débit utilisée.

Les évolutions temporelles des erreurs de données peuvent aussi se retrouver au niveau des précipitations. Une évolution du réseau d'appareils de mesure (ajout ou retrait de postes, renouvellement du matériel de mesure) pourra engendrer des variations sur les erreurs d'estimation. En effet, le type de postes pluviométriques, leur environnement direct (exposition au vent, proximité d'arbres...) et la couverture géographique du territoire (vallées, sommets et versants différemment arrosés selon les flux atmosphériques) sont autant d'éléments qui influencent la justesse des estimations des précipitations à l'échelle d'un bassin versant. Comme l'a montré *Oudin et al.* (2006b), des erreurs systématiques peuvent être compensées par le modèle et ne seront pas problématiques si elles demeurent stationnaires (*c.-à-d.* la compensation de celles-ci par le modèle restera valide). Cependant, si ces biais dans les estimations évoluent temporellement du fait d'une évolution du réseau, alors ils deviendront problématiques vis-à-vis de la transférabilité temporelle des paramètres d'un modèle hydrologique.

Les estimations de la température de l'air sont beaucoup moins sujettes à des évolutions temporelles que celles des précipitations. Ce champ est plus lisse spatialement (autant horizontalement que verticalement) et donc plus facile à approximer (notamment l'estimation du gradient de température avec l'altitude comparativement au gradient de précipitation).

Enfin, l'estimation de l'évapotranspiration potentielle du bassin peut être sujette à des erreurs évoluant temporellement. Il est vrai que les calculs d'ETP font généralement intervenir la variable température que nous pouvons juger comme faiblement concernée par des erreurs tendancielles. Rappelons néanmoins que l'ETP est une grandeur théorique d'évapotranspiration maximale d'un gazon (fétuque manade). Calculée de manière externe ou interne au modèle hydrologique, elle interviendra dans l'estimation de l'évapotranspiration réelle (ETR). Les éventuelles erreurs d'estimation des grandeurs intervenant dans le calcul de l'ETP (température, vent, radiation) se traduiront donc en erreurs sur le champ d'ETP estimé. De plus, il existe une grandeur théorique intermédiaire qui adapte l'évapotranspiration potentielle à la couverture végétale des sols pour déterminer une évapotranspiration maximale ($ETM = Kc \cdot ETP$, où Kc est appelé coefficient cultural) [1]. Or, la couverture des sols, et donc ce coefficient de pondération de l'ETP, peut évoluer temporellement. Si elles existent, ces évolutions ressortiront particulièrement lorsque de longues périodes sont considérées (comme c'est le cas dans nos tests). Ces évolutions peuvent être d'origine *anthropique directe* (*ex.* imperméabilisation des sols, mise en place de réseaux de drainage, déforestation, changement de type de culture) et se dérouler sur des périodes relativement courtes (de quelques années à quelques décennies). Ces évolutions peuvent également être d'origine plus *naturelle*, telles que des cycles de destruction et reprise de la végétation avec, s'ils existent, des impacts anthropiques indirects (*ex.* incendies, températures extrêmes, sécheresse) ou bien des situations d'évolution lente du type de végétation, en lien avec la colonisation d'une espèce et/ou avec les évolutions climatiques. Tout comme pour les erreurs concernant les précipitations, des erreurs d'estimation de l'ETP (ou plutôt de l'ETM) peuvent être compensées par le modèle si elles sont systématiques et stationnaires (*Andréassian et al.*, 2004; *Oudin et al.*, 2006b) mais deviennent problématiques lorsqu'elles varient temporellement.

Problématique d'analyse

Les données de température utilisées sur notre échantillon français mettent en avant une tendance temporelle à l'augmentation (autour de 0.9 °C en moyenne tous les 20 ans). Les sous-périodes continues que nous utilisons dans la procédure $GSST$ sont donc marquées par cette augmentation (*cf.* figure 5.6). Ainsi, le fait de transférer des paramètres pour un $\triangle T$ de ± 1°C signifie également un transfert temporel de ± 20 ans (en moyenne sur l'échantillon).

1. Ce calcul intermédiaire ($ETM = Kc \cdot ETP$) n'est pas présent dans les modèles hydrologiques les plus simples mais peut être rencontré pour des modèles plus complexes tels que *Cequeau*.

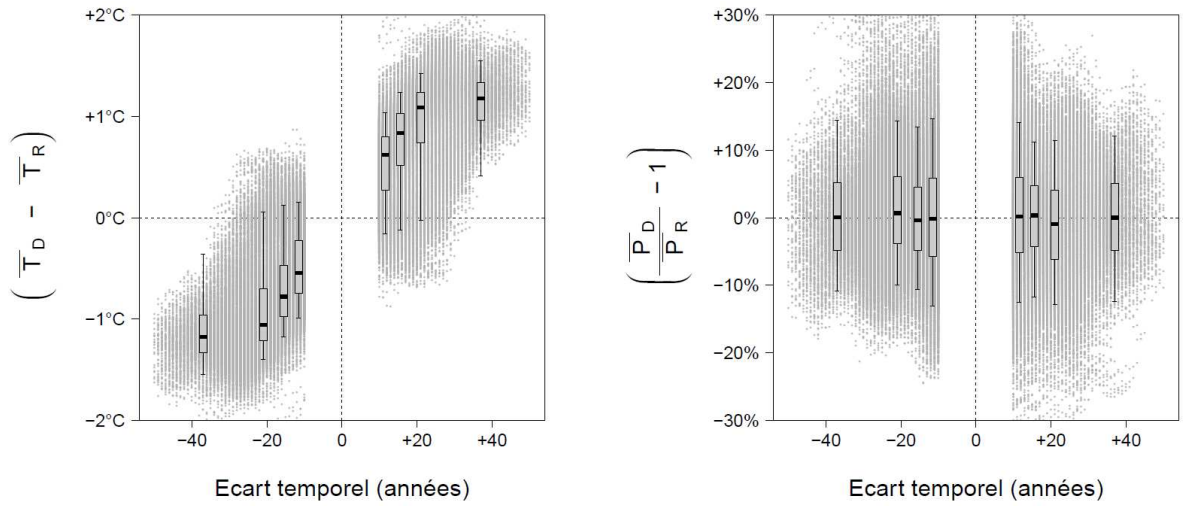


Figure 5.6.: Comparaison entre écarts climatiques et écarts temporels pour toutes les situations de transfert testées

Rq. vide central causé par le non-chevauchement entre périodes de calage et contrôle.
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Cette situation est problématique lorsque nous constatons des corrélations entre les variations d'erreurs de simulation et les variations de température moyenne entre les périodes de calage et de validation. En effet, il peut alors être difficile de déterminer si nous détectons (a) le rôle des variations de température entre les périodes de calage et de validation, ou (b) le rôle de l'éloignement temporel entre les périodes de calage et de validation. Or, les implications associées à ces deux points sont très différentes du point de vue de nos évaluations de la robustesse temporelle et climatique des modèles hydrologiques :

- Le premier point ($\triangle\varepsilon = f(\triangle\overline{T})$), analysé du point de vue des défaillances éventuelles de modélisation, soulève la question de la capacité d'extrapolation climatique des modèles.
- Ce même premier point ($\triangle\varepsilon = f(\triangle\overline{T})$), analysé du point de vue des problèmes de mesure et d'estimation des données, ouvre les réflexions sur d'éventuels liens entre les conditions climatiques et les erreurs d'estimation des données qui alimentent les modèles (bien évidemment, nous pensons à l'estimation de l'ETP ou à la sous-captation de la neige mais n'oublions pas des impacts indirects tels que la présence de végétation dans la rivière, en corrélation avec la température, qui perturbe la mesure des faibles débits).
- Le second point ($\triangle\varepsilon = f(\triangle temps)$) nous conduit par contre directement vers la question des erreurs de mesure et de leurs évolutions temporelles discutée précédemment (*ex.* dérive des courbes de tarages, évolution des réseaux de mesures ou de la couverture végétale).

Remarque : Nous noterons la présence d'une bande centrale vide sur la figure 5.6. Ce vide est causé par la contrainte de non-chevauchement entre les périodes de calage et contrôle. Considérer deux périodes de 10 ans continues et indépendantes entraîne nécessairement un écart temporel d'au moins 10 ans entre les barycentres de celles-ci.

Commentaire préalable (antérieur aux tests)

La prochaine section vise à départager les rôles des écarts de température et des écarts temporels sur les erreurs de modélisation. Avant de nous y plonger, autorisons-nous à exprimer un *a priori* sur ce que nous pensions trouver :

Nous avons constitué un groupe de 120 bassins versants pour lesquels $\triangle\varepsilon = f(\triangle\overline{T})$ et nous souhaitons déterminer si la véritable corrélation ne se situe pas, en réalité, avec le $\triangle temps$ (impliquant les questions de dérive de la qualité des données et de la relation pluie-débit). Pourtant, l'étude des variations des erreurs relativement aux $\triangle\overline{T}$ et $\triangle temps$ nous conduit à douter de cette seconde possibilité :

1. Les nuages représentant $MRC = f(\triangle\overline{T})$ et $MRC = f(\triangle temps)$ (en haut sur la figure 5.7) laissent penser que $\triangle temps$ serait, tout comme $\triangle\overline{T}$, une variable explicative des variations d'erreurs de modélisation. Néanmoins, les corrélations sont moins nettes, comme le révèlent les hauteurs relatives des boîtes et leur positionnement vertical (dispersion plus importante et médianes de MRC moins élevées). Nous avons donc le sentiment que les $\triangle\overline{T}$ constituent une variable explicative plus forte que $\triangle temps$.

2. Les nuages illustrant les biais en fonction des écarts de températures (en bas sur la figure 5.7) mettent en avant une nette anti-corrélation entre $(\overline{\widehat{Q}}_R[\theta_D]/\overline{Q}_R)$ et $(\overline{T}_D - \overline{T}_R)$. Cette anti-corrélation est moins nette lorsque nous considérons la relation $biais = f(\triangle temps)$. Plus encore, il serait extrêmement surprenant que des dérives tendancielles dans l'estimation des données causent à elles seules une telle situation. En effet, il est peu probable que des dérives de mesures sur des bassins très différents conduisent, en moyenne, à des biais de simulation de même signe (*ex.* surestimation lors d'un calage sur une période antérieure). Une dérive de la courbe de tarage ou une évolution du réseau de pluviomètres devrait entraîner indifféremment une tendance négative ou positive.

Pour résumer, la comparaison des niveaux de dispersion des nuages de points combinée au systématisme des erreurs de volumes nous conduisent à considérer peu probable l'hypothèse selon laquelle les dérives d'estimations des données seraient la cause première des problèmes de robustesse constatés (les relations $\triangle erreurs = f(\triangle\overline{T})$ n'étant qu'un signal induit, artéfact de la relation $\triangle\overline{T} = f(\triangle temps)$). Précisons toutefois que nous ne pouvons pas totalement écarter le fait que les dérives d'estimations des entrées des modèles puissent elles-mêmes être liées aux évolutions de température (*ex.* erreurs d'estimation de l'ETP).

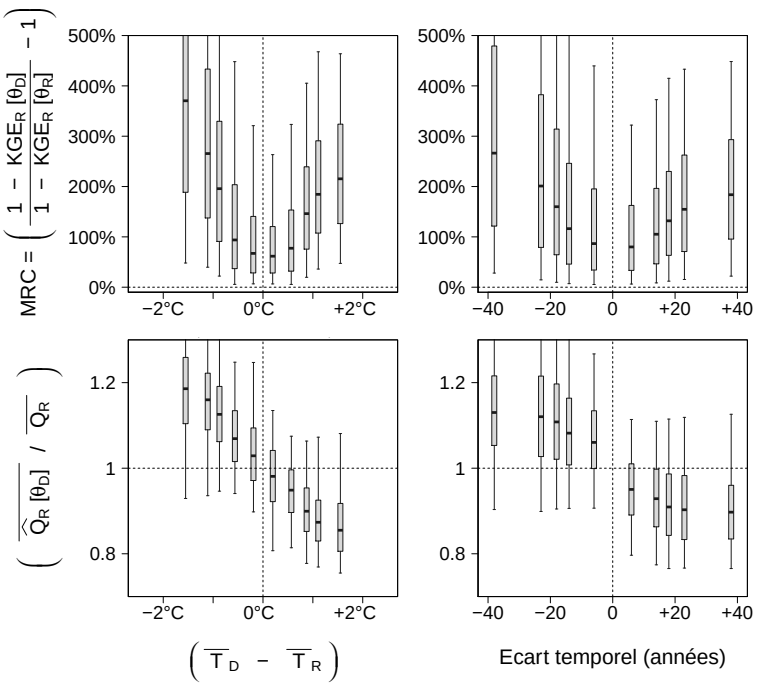


Figure 5.7.: Analyse comparative des pertes de performances selon les △températures et les écarts temporels subis lors des transferts

Résultats obtenus pour GR4J sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures. Sur les bassins français, GR4J est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

5.4.2. Analyses des liens entre △*temporel* et erreurs de modélisation

Tests de rééchantillonnage des années (approche bootstrap)

Lors de la mise en place de notre méthodologie de test *GSST*, nous avons évoqué l'alternative constituée par une approche de type *bootstrap* (*cf.* discussion à la section 2.1.1). Les sous-périodes utilisées pour caler et valider les modèles hydrologiques n'y sont plus nécessairement continues et peuvent constituer un agrégat d'années hydrologiques temporellement espacées. Cependant, toutes nos analyses sont sujettes à la difficulté de considérer les phénomènes de *mémoire* des bassins [2]. Cette prise en compte étant beaucoup plus délicate dans le cadre du *bootstrap* que du *GSST*, nous avons préféré ce dernier pour réaliser le cœur de nos travaux (*cf.* section 2.1.1). Le *bootstrap* présente pourtant un avantage notable sur le *GSST*, lorsque l'on souhaite dissocier le rôle des variations climatiques du rôle des dérives temporelles sur des écarts de performances constatés lors de transferts de paramètres. Cet avantage est lié à l'absence de contraintes sur la contigüité des années qui composent les périodes de calage

2. Ce terme *mémoire* désigne le fait que la réponse d'un bassin versant aux forçages climatiques peut être influencée par les conditions antérieures à la période d'étude, de plusieurs mois voire de quelques années.

et validation dans les tests. Avec le *bootstrap*, nous pouvons volontairement construire des sous-périodes composées d'années réparties tout au long de la série disponible. Les erreurs systématiques sur les entrées du modèle liées à des dérives temporelles sont alors effacées (au moins en partie), puisque toutes les périodes testées contiennent à la fois des années situées au début, au milieu et à la fin des chroniques de forçages et de débit.

Quelques tests ponctuels de *bootstrap* ont été menés pour apprécier si les résultats s'en trouvaient modifiés. La procédure de constitution des sous-périodes de test par *bootstrap* est détaillée dans l'annexe G.1. Les biais en validation obtenus pour trois bassins et deux modèles hydrologiques sont donnés à la figure 5.8. Les ronds gris correspondent aux nombreux tests issus de la procédure *GSST,* déjà aperçus précédemment, tandis que les ronds rouges illustrent les résultats issus des nouveaux tests par *bootstrap.*

Pour des $\triangle\overline{T}$ équivalents entre calage et validation, les résultats obtenus sur trois bassins avec l'approche *bootstrap* indiquent des erreurs sur les volumes moyens moins fortes en validation que celles obtenues avec la procédure *GSST*. Ces biais ne disparaissent cependant pas et nous retrouvons des tendances similaires de surestimation des débits lors des transferts froid→chaud et de sous-estimation lors des transferts chaud→froid. Ces nouveaux tests nous permettent de confirmer notre *a priori* sur le rôle plus important des variations de température sur les erreurs de simulation, que du simple éloignement entre périodes.

Plusieurs hypothèses peuvent être émises pour expliquer le fait que les points rouges ne sont pas localisés précisément au centre du nuage de points gris. Nous pouvons notamment évoquer l'existence possible de phénomènes de *mémoire* sur notre analyse, qui impacteront la façon de caractériser la température moyenne d'un agrégat d'années indépendantes. En effet, les $\triangle\overline{T}$ estimés pour des couples de périodes continues et discontinues ne sont pas réellement comparables, ce qui peut conduire à une incertitude sur le positionnement latéral des points rouges. De plus, le nombre de tests réalisés ici demeure restreint, or le processus de calage peut conduire ponctuellement à une mauvaise identification du jeu de paramètres optimal.

Précisons cependant que les difficultés d'utilisation du *bootstrap* et d'interprétation des résultats associés évoqués ici sont liées à notre contexte d'analyse et ne sont pas spécifiques au *bootstrap* en lui-même. Dans un contexte de calage seul, cette approche reste tout à fait valide et peut même permettre la sélection de jeux plus robustes qu'un calage sur une période continue de même longueur.

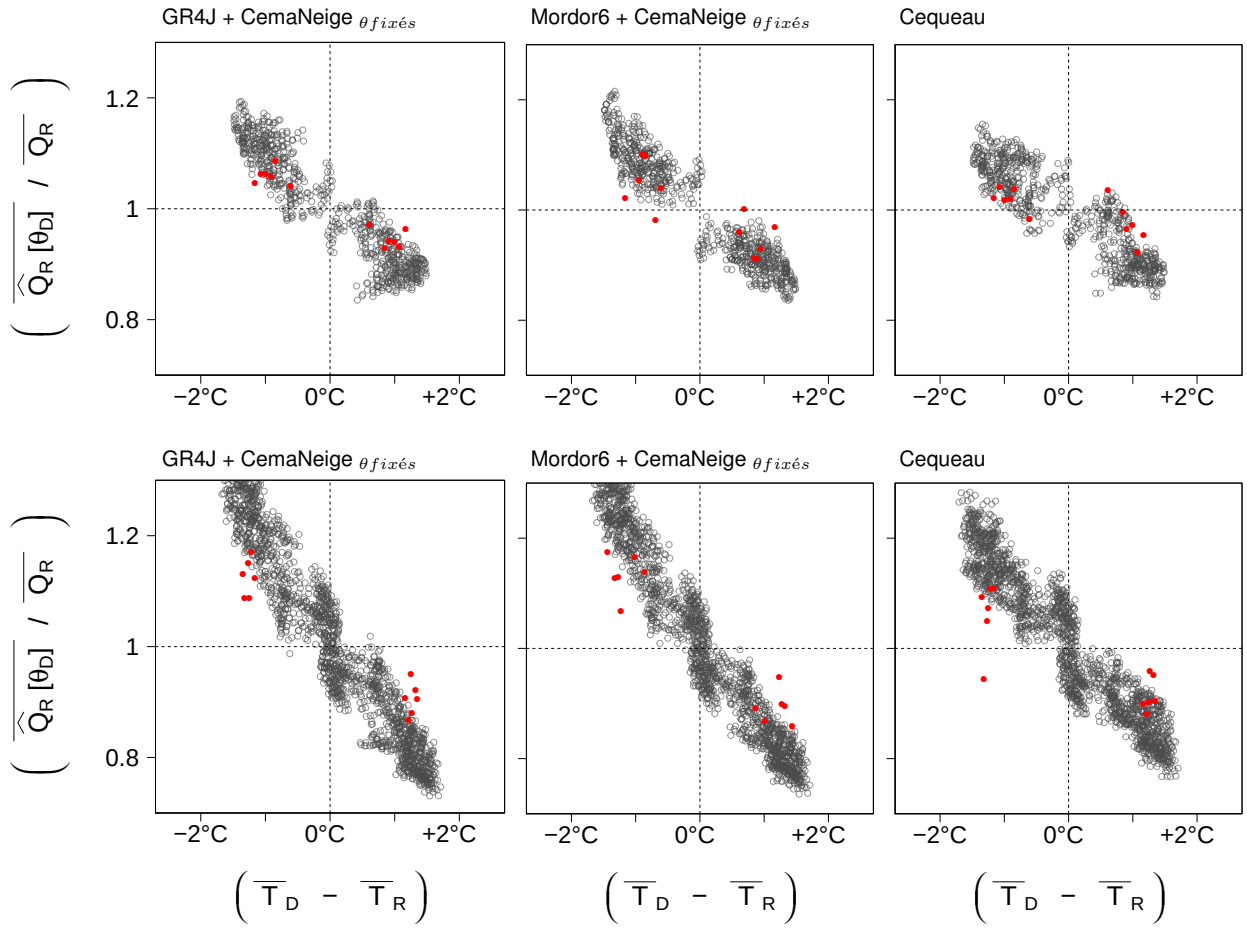


Figure 5.8.: Quelques tests de transfert *bootstrap* avec *GR4J* sur les bassins de *la Dourbie à Nant* et *du Lot à Banassac*

Cercles gris : résultats obtenus par GSST. Points rouges : résutats obtenus par bootstrap, sur des périodes sélectionnées pour maximiser conjointement l'écart de température moyenne entre l'agglomérat d'années et la dispersion temporelle de ces années.

GR4J, Mordor6 et Cequeau sont ici alimentés par des données SPAZM, les deux premiers utilisant une ETP d'Oudin tandis que Cequeau cale une ETP de Thornthwaite.

Comparaison directe des effets respectifs de △climat et △temps sur les erreurs de modélisation

Sans passer nécessairement par une multiplication des tests par l'approche *bootstrap*, nous disposons d'un autre moyen pour départager $\triangle\overline{T}$ et $\triangle temps$, dans leurs influences respectives sur les interprétations menées dans ces travaux. Cette approche repose simplement sur l'introduction d'une nouvelle forme de représentation des résultats. Jusqu'ici, les variations des erreurs constituaient les ordonnées de nos graphiques et les variations de chaque caractéristique, prise séparément en constituaient l'abscisse (*ex.* $\triangle\varepsilon = f(\triangle\overline{T})$ ou $\triangle\varepsilon = f(\triangle\overline{P})$ ou $\triangle\varepsilon = f(\triangle\overline{P}/\overline{E})$). Nous pouvons ainsi imaginer une représentation où deux caractéristiques de périodes sont considérées et constituent les abscisses et ordonnées du graphique. L'espace ainsi

constitué est alors découpé selon un pas fixe, propre à chaque caractéristique, pour former une grille de pixels sur laquelle nous plaçons les résultats de nos tests de transfert (*ex.* erreurs constatées pour les situations où $\triangle\overline{T}\epsilon\,[-1.2°\text{C}; -1.0°\text{C}]$ et $\triangle\overline{P}\epsilon\,[+10\%; +15\%]$). Enfin, nous analysons sur chaque pixel les résultats des tests associés pour en faire ressortir un niveau d'erreur exprimé à travers un code couleur [3]. Cette représentation, visible sur la figure 5.9, sera désignée ici sous les termes de représentation *en damier*.

Comparativement à nos analyses précédentes, sur la forme des nuages de points $\triangle erreurs = f(\triangle caract.)$, la représentation en damier présente deux inconvénients : (1) Les erreurs étant indiquées par un code couleur, seul un quantile peut être illustré par graphique *en damier* (*ex.* valeur médiane), tandis que la représentation $\triangle\varepsilon = f\,(\triangle caract.)$ permet d'avoir, sur un même graphique, un aperçu de la forme complète du nuage (via les boîtes à moustaches). (2) Cette méthode contraint à un découpage régulier de l'espace x-y sous forme de grille qui a pour conséquence un nombre différent de points pour chaque pixel [4]. Afin de limiter les erreurs d'interprétation liées à ce deuxième aspect (*c.-à-d.* sous/sur-représentation de certains cas) nous réduisons la taille des pixels lorsque trop peu de points sont disponible (taille proportionnelle au nombre de points sous le seuil des 50 valeurs par pixel et fixe au-dessus).

La figure 5.9 présente les résultats obtenus pour GR4J sur les 120 bassins de notre sous-groupe témoin (où nous avions détecté des corrélations entre *MRC* et $\triangle\overline{T}$). Avant d'analyser les variations de couleurs, notons les particularités du damier représenté ici. Tout d'abord, nous retrouvons la bande centrale vide, causée par la contrainte de non-chevauchement entre les périodes de calage et contrôle. Ensuite, nous remarquons l'absence de pixels d'analyse dans certaines parties de l'espace (*ex.* $\triangle t < 0$ & $\triangle\overline{T} > 0$ ou encore $\triangle t > 0$ & $\triangle\overline{T} < 0$) qui résulte de l'évolution temporelle des températures (*cf.* figure 5.6).

L'organisation des couleurs sur les différents damiers de la figure 5.9 confirme la prédominance de la corrélation entre les erreurs de modélisation et les écarts de température, comparativement à celle entre $\triangle erreurs$ et $\triangle temps$. Ce résultat est particulièrement visible sur le graphique *en damier* concernant le critère *MRC* (à gauche), où nous distinguons clairement qu'un déplacement horizontal n'entraîne pas de changement de couleur, à la différence d'un déplacement vertical. Par exemple, si la différence de température est nulle $\left(\triangle\overline{T} \sim 0\right)$, alors un écart temporel (même de ± 30 ans) ne s'accompagne pas d'augmentation de *MRC*, tandis qu'à écart temporel constant, notre critère *MRC* augmente avec $\left|\triangle\overline{T}\right|$. Même s'il est un peu moins marqué, ce résultat se retrouve sur le damier de droite qui concerne les erreurs de volume constatées en validation lors de nos tests.

3. Ce niveau peut être une moyenne mais aussi des quantiles spécifiques et peut porter sur différents critères d'erreur (*ex. MRC*, biais, variance...).

4. Ce qui n'est pas le cas pour les analyses par nuage de points dans lesquelles chaque boîte à moustaches concerne un nombre identique de points.

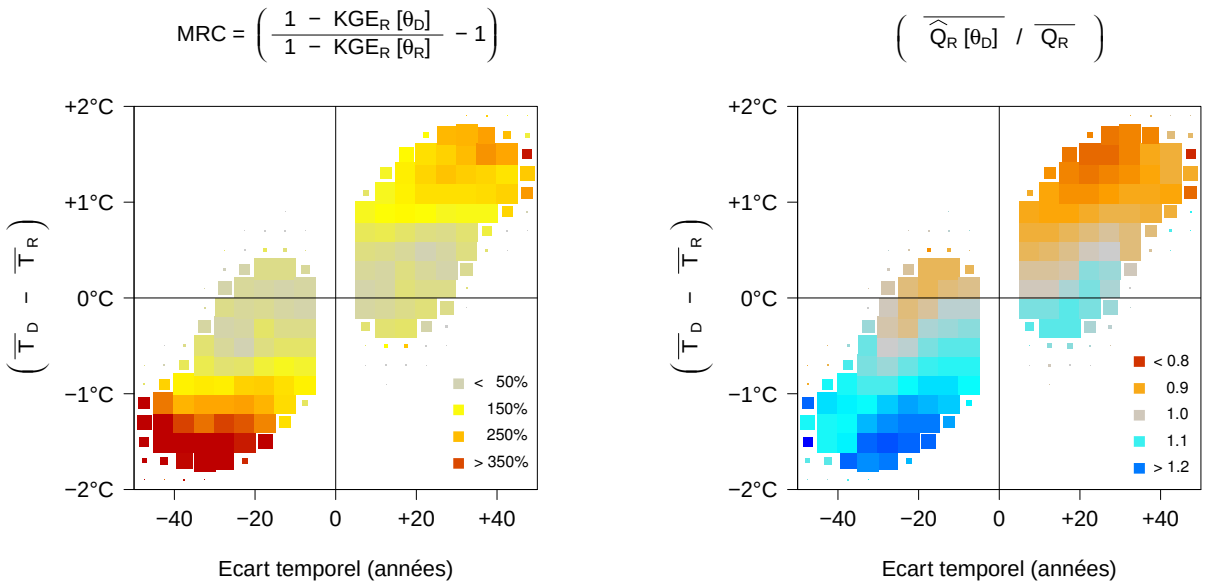


Figure 5.9.: Visualisation conjointe des liens entre pertes de performances, △températures et écarts temporels subis lors des transferts

Résultats obtenus pour GR4J sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures. Sur les bassins français, GR4J est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés).

5.4.3. Conclusions

Cette section avait pour objectif de déterminer si les évolutions temporelles de qualité des données et/ou de la relation pluie-débit constituaient une variable explicative de certaines situations de manque de robustesse des modèles. Plus précisément, nous souhaitions dissocier les corrélations entre erreurs de modélisation et $\triangle\overline{T}$, d'une part, et erreurs de modélisation et $\triangle temps$, d'autre part. Deux pistes ont été explorées pour cela : une approche basée sur des tests de *bootstrap*, menés sur trois bassins versants et deux modèles hydrologiques, et une nouvelle méthode de visualisation (en damier), permettant une comparaison des rôles respectifs de $\triangle\overline{T}$ et $\triangle temps$ sur les performances de *GR4J* pour un sous-groupe de 120 bassins.

Ces résultats tendent à monter que : (a) les erreurs de simulation et le $\triangle\overline{T}$ subi entre calage et validation constituent bien un signal principal, et (b) les écarts temporels ne sont pas, à eux seuls, une variable explicative des pertes de performances moyennes des modèles, mais sont induites par l'augmentation moyenne de température avec le temps ($\triangle\overline{T} = f(\triangle temps)$).

5.5. Analyse des parallèles entre *biais=f(△caract.)* et *△param.=f(△caract.)*

5.5.1. Introduction et mode d'analyse

Les nombreux tests que nous avons menés jusqu'ici nous ont permis de constater des cas d'erreurs systématiques de simulation qui semblent liés aux conditions de transfert des paramètres (*cf.* chapitre 3). Les biais observés sur les volumes moyens lors du *GSST* constituent le résultat le plus marquant et expliquent vraisemblablement une grande partie des variations de critères d'erreurs tels que le *KGE* ou le *NSE* lors du passage de conditions de calage à celles de validation. Nous souhaitons désormais plonger au cœur des modèles pour comprendre si une trace de ces corrélations se retrouve dans les valeurs optimales des paramètres. Lorsque nous observons, par exemple, qu'un calage sur une période plus froide ou plus humide est associé à une surestimation quasi systématique des débits sur de nombreux bassins, nous nous interrogeons sur les évolutions des paramètres sélectionnés qui conduisent à ce comportement. Nos travaux se rapprochent ainsi des démarches de *Le Lay et al.* (2007); *Merz et al.* (2011) ou encore *Reusser et Zehe* (2011).

Une fois de plus, nous nous sommes interrogés sur le choix des bassins versants sur lesquels conduire ces nouvelles analyses. Nous pouvions rechercher directement une relation entre les variations temporelles des conditions climatiques sur un bassin et les variations temporelles des valeurs de paramètres jugées optimales pour un modèle donné. Nous avons cependant opté pour une approche se plaçant dans la lignée des tests précédents où nous avions formé des groupes de bassins selon les comportements des modèles face au *GSST*. Nous reprenons ici deux groupes respectivement composés de 120 et 32 bassins sur lesquels nous avions relié robustesse des modèles et $\triangle$*températures* ou $\triangle$*précipitations* lors de nos *split-sample tests* (*cf.* sections 3.3 et 3.4).

Le mode de représentation et d'analyse des variations des paramètres en lien avec les conditions de transfert est très similaire à ce qui fut utilisé pour l'étude des variations de *KGE*, exprimé sous la forme du *MRC* (*cf.* section 2.1.3). Pour chaque situation de transfert ($D \rightarrow R$), la valeur du paramètre utilisé (θ_D) est exprimée relativement à celle obtenue par calage sur la période simulée (θ_R). Prenons un exemple où une surestimation des débits est constatée ($\overline{\widehat{Q}}_R[\theta_D]/\overline{Q}_R > 0$) en correspondance avec un calage sur une période plus froide ($\overline{T}_D - \overline{T}_R < 0$) ou plus humide ($(\overline{P}_D - \overline{P}_R)/\overline{P}_R > 0$). Nous voulons savoir quelles différences causent cette situation au niveau des paramètres et nous étudions pour cela le rapport $(\theta_D - \theta_R)/\left|\theta_R\right|$. L'avantage de cette représentation est de permettre un parallèle avec l'étude des nuages d'erreurs et l'expression des paramètres sous la forme relative permet de s'affranchir des différences de valeurs absolues de paramètres liées à chaque bassin versant. Nous restons toutefois conscients

que pour certains paramètres, une variation relative de 20% n'aura pas le même sens si elle s'opère sur des valeurs très faibles ou très fortes. Toutefois, cette expression relative permet de rassembler, dans un même nuage de points, les résultats issus de bassins très différents qui pourront ainsi être analysés ensemble.

Remarque : Les valeurs obtenues pour les rapports $(\theta_D - \theta_R)/|\theta_R|$ ont peu d'importance et nous n'y prêtons pas réellement d'attention lors des comparaisons entre les nuages de points concernant des paramètres différents, compte tenu du degré inégal d'impact de différents paramètres sur les simulations. Nous nous concentrons sur l'étude de la forme des nuages. Cette étude reste qualitative puisque c'est visuellement que nous apprécions les tendances à travers le positionnement vertical des boîtes à moustaches, pondéré par la hauteur des boîtes (*ex.* de fortes hauteurs de boîtes est le signe de l'absence de corrélation, même si une légère tendance est visible).

5.5.2. Résultats

Les résultats complets de ces analyses ont été placés dans l'annexe G.2 afin de ne pas alourdir le contenu du mémoire. Trois paramètres par modèle sont représentés ici, parmi lesquels ceux dont les variations nous paraissent apporter un éclairage sur les erreurs systématiques de volume mises en avant par le *GSST*. Ces analyses sont nécessairement propres à chaque modèle, compte tenu des différences de structure et de complexité.

À chaque fois, deux analyses parallèles sont présentées selon le sous-groupe de bassins :

1. le sous-groupe de 120 bassins où nous avions détecté des corrélations entre les erreurs de *GR4J* en validation et les écarts de température subis lors des transferts temporels [5]. Les résultats correspondants seront placés en haut dans les figures qui suivent.
2. le sous-groupe de 32 bassins où nous avions détecté des corrélations entre les erreurs de *GR4J* en validation et les écarts de précipitations subis lors des transferts temporels [6]. Les résultats correspondants seront placés en bas dans les figures qui suivent.

Comportements observés pour *GR4J*

GR4J est un modèle conceptuel alimenté par des moyennes journalières de bassin de précipitations et d'*ETP*. Il comporte quatre paramètres à caler : capacité du réservoir sol ($\theta_{X1(PROD)}$), coefficient d'échange souterrain ($\theta_{X2(CES)}$), capacité du réservoir de routage ($\theta_{X3(ROUT)}$) et constante de temps de l'hydrogramme unitaire ($\theta_{X4(THU)}$). Le détail de sa structure est fourni

5. Sur ces 120 bassins, nous avons constaté une surestimation des débits lorsque la période calage est plus froide que celle de validation et inversement.

6. Sur ces 32 bassins, nous avons constaté une surestimation des débits lorsque la période calage est plus humide que celle de validation et inversement.

dans la section 2.3.1. Les variations relatives des trois premiers paramètres, relativement aux conditions de transferts, sont présentées sur la figure 5.10 pour nos deux cas d'études, tous les nuages complémentaires étant reportés en annexe (*cf.* figures G.3 à G.5).

Les variations de $(\theta_D - \theta_R)/|\theta_R|$ mettent en avant des tendances sur le paramètre $\theta_{X2(CES)}$ en lien avec les conditions de transfert, nettement visibles sur les deux groupes de bassins étudiés et particulièrement marquées sur le groupe de 120 bassins où $\triangle\varepsilon = f(\triangle\overline{T})$ (graphiques du haut). Une tendance est également visible sur $\theta_{X1(PROD)}$ pour le groupe de 32 bassins où $\triangle\varepsilon = f(\triangle\overline{P})$. Dans les autres cas, même si des tendances sur les médianes des boîtes semblent émerger, les hauteurs des boîtes sont trop importantes pour que nous nous risquions à interpréter ces variations en lien avec le delta climat.

Le paramètre $\theta_{X2(CES)}$ contrôle un processus qui s'apparente à des échanges souterrains avec les bassins versants voisins. Il s'agit d'une variable d'ajustement pouvant ajouter ou retirer de l'eau du système *bassin versant* conceptualisé. En moyenne sur l'échantillon, $\theta_{X2(CES)}$ prend des valeurs négatives, ce qui signifie que le modèle fait sortir de l'eau du système. Nous constatons ici des valeurs de $\theta_{X2(CES)}$ plus faibles (*c.-à-d.* plus négatives) que les valeurs référentes pour les situations de sous-estimations des débits et, à l'inverse, des valeurs de $\theta_{X2(CES)}$ plus élevées (*c.-à-d.* moins négatives) que les valeurs référentes pour les situations de surestimations des

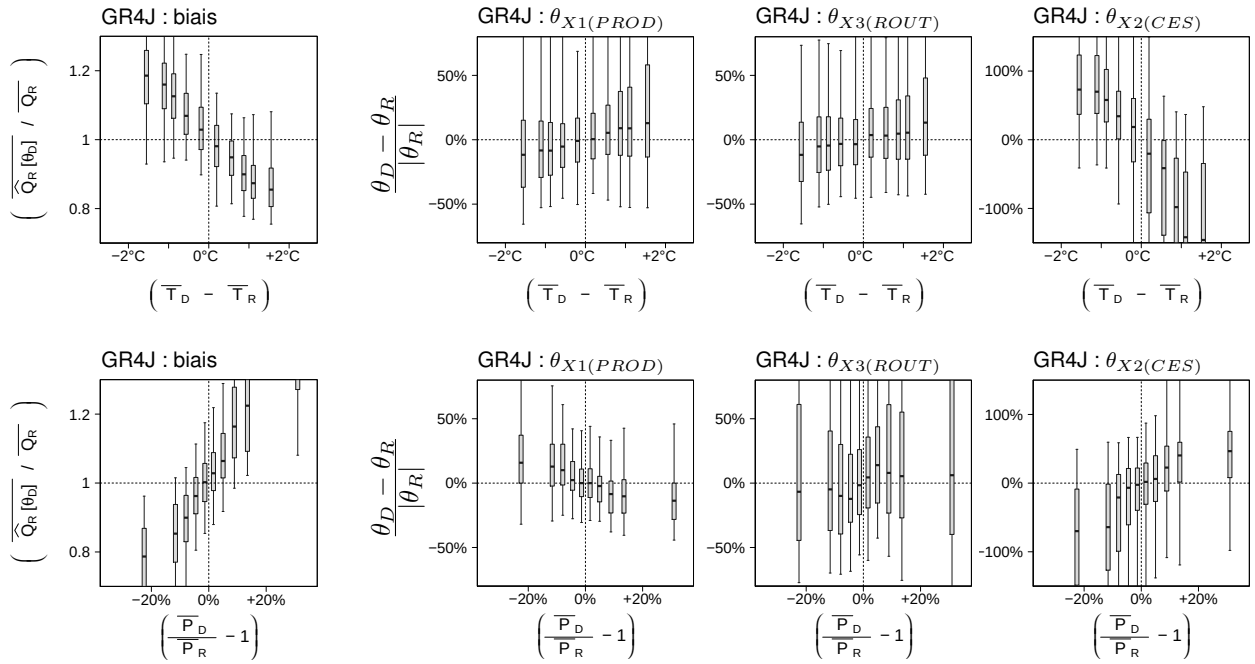


FIGURE 5.10.: Erreurs de volume et $\triangle\theta$ observés pour *GR4J* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$, en haut, et △erreurs et $\triangle\overline{P}$, en bas

Rq. sur les bassins français, GR4J est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés).

débits. Cette situation est logique. Un $\theta_{X2(CES)}$ trop faible conduit effectivement à faire sortir trop d'eau du bassin versant et donc à sous-estimer les écoulements de surface. À l'inverse, un $\theta_{X2(CES)}$ trop fort (*c.-à-d.* pas assez négatif) ne fera pas sortir assez d'eau et entraînera les surestimations de débits constatées.

Comportements observés pour *Mordor6*

Mordor6 est un modèle global à réservoir, alimenté par des moyennes journalières de bassin de précipitations et d'*ETP*. Il comporte une structure assez différente de *GR4J* avec ses six paramètres à caler : coefficient multiplicateur des pluies ($\theta_{X1(CPL)}$), capacité du réservoir de surface ($\theta_{X5(RU)}$), capacité du réservoir intermédiaire ($\theta_{X6(RL)}$), constante de vidange du réservoir intermédiaire ($\theta_{X2(TRL)}$), constante de vidange du réservoir nappe ($\theta_{X3(TRN)}$), constante de temps de l'hydrogramme unitaire ($\theta_{X4(THU)}$). Le détail de sa structure est fourni dans la section 2.3.2.

Les variations relatives des paramètres $\theta_{X1(CPL)}$, $\theta_{X3(TRN)}$ et $\theta_{X5(RU)}$, relativement aux conditions de transfert, sont présentées sur la figure 5.11 pour nos deux groupes d'étude, les nuages obtenus pour tous les paramètres étant reportés en annexe (*cf.* figures G.6 à G.8). Les trois paramètres retenus ici permettent (plus ou moins directement) d'ajuster l'équilibre entre précipitations (P), évapotranspiration (ETR) et écoulement de surface (Q). Le paramètre $\theta_{X1(CPL)}$ est celui permettant d'ajuster le plus facilement le bilan du bassin en multipliant les précipitations. Il peut être inférieur ou supérieur à 1, la confrontation avec les débits observés au calage déterminant si un ajout ou un retrait d'eau du système est nécessaire. Sur notre échantillon, il a généralement une valeur inférieure à 1, correspondant à une sortie d'eau du système, en cohérence avec les valeurs négatives du $\theta_{X2(CES)}$ de *GR4J*. Le paramètre $\theta_{X5(RU)}$ et, dans une moindre mesure, le paramètre $\theta_{X2(TRL)}$ permettent d'ajuster la mise à disposition d'eau dans des réservoirs conceptuels sujets à évaporation. Ils peuvent donc aussi influencer la répartition de l'eau entre P, ETR et Q mais de façon moins directe que $\theta_{X1(CPL)}$.

Comme nous pouvions le supposer au regard des constatations faites pour *GR4J*, le paramètre $\theta_{X1(CPL)}$ varie (relativement) en lien avec les conditions de transfert sur les deux groupes de bassins étudiés. L'utilisation d'un $\theta_{X1(CPL)}$ trop faible (comparativement à la référence calée) entraînera une sous-estimation des débits ; à l'inverse, l'utilisation d'un $\theta_{X1(CPL)}$ trop fort (*c.-à-d.* trop proche de 1) "ne fera pas sortir assez d'eau du système" et conduira à une surestimation de débits. Nous noterons enfin la présence de tendances sur $\theta_{X3(TRL)}$ mais nous ne pensons pas qu'elles soient directement liées au processus d'équilibrage du bilan qui influence les variations de $\theta_{X1(CPL)}$. Une explication probable de ces $\triangle\theta_{X3(TRL)}$est la suivante : le passage des périodes de calage à celles de validation implique des variations absolues des grandeurs P, ETR, Q. Or, des variations de volume de précipitations non compensées par l'évapotranspiration induiront des variations sur les volumes d'eau qui transitent au sein de la

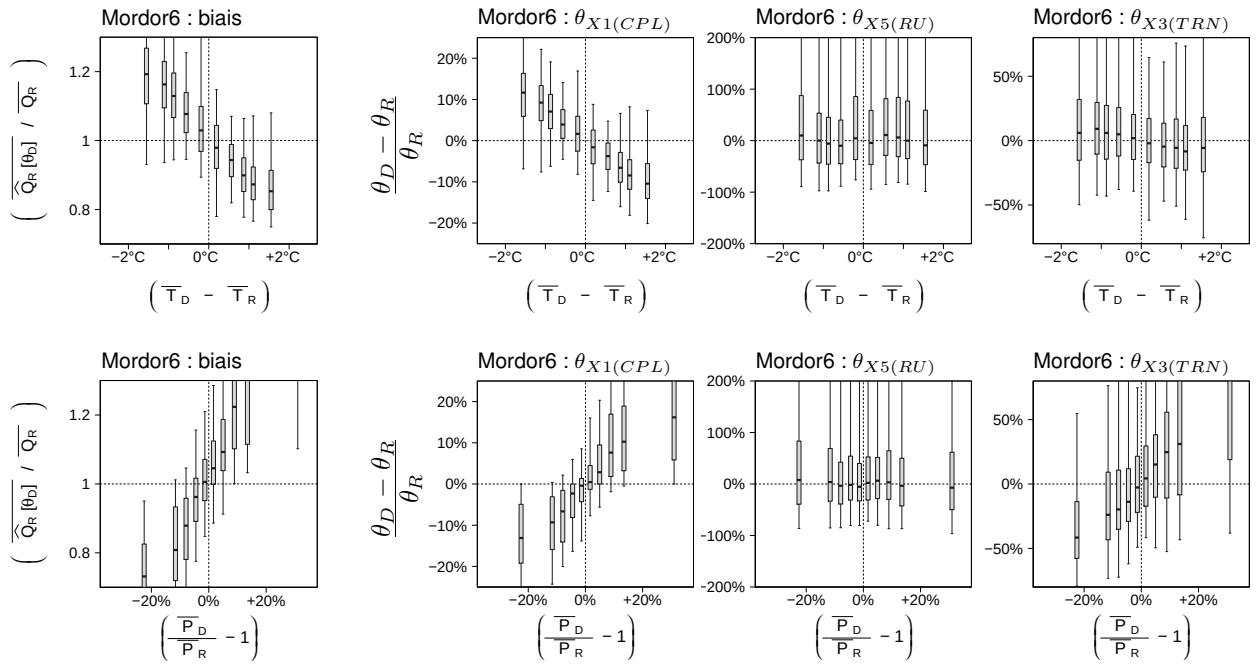


FIGURE 5.11.: Erreurs de volume et $\triangle\theta$ observés pour *Mordor6* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$, en haut, et △erreurs et $\triangle\overline{P}$, en bas

Rq. sur les bassins français, Mordor6 est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés).

structure du modèle, notamment par les réservoirs conceptuels. Le temps caractéristique de réaction du bassin à une pluie devant rester inchangé, cette situation pourra donc contraindre le modèle à augmenter ou diminuer la vitesse de vidange de ses réservoirs. Le fait que des tendances sur $\theta_{X2(TRL)}$ et $\theta_{X3(TRN)}$ soient particulièrement visibles sur les graphiques obtenus pour le groupe de 32 bassins où $\triangle\varepsilon = f(\triangle\overline{P})$ va dans le sens de ce raisonnement (*c.-à-d.* plus d'eau s'écoulant dans les réservoirs → de plus forts ajustements de $\theta_{X2(TRL)}$ et $\theta_{X3(TRN)}$, *cf.* figures G.6 à G.8 en annexe).

Comportements observés pour *SimHyd*

Tout comme *GR4J* et *Mordor6*, *SimHyd* est un modèle global à réservoir, alimenté par des séries journalières de précipitations et d'*ETP* de bassin. Dans la version utilisée dans cette thèse, *SimHyd* comporte sept paramètres à caler : capacité du réservoir d'interception $\theta_{X1(INSC)}$, capacité du réservoir sol $\theta_{X2(SMSC)}$, coefficient *interflow* $\theta_{X6(SUB)}$, coefficient de recharge $\theta_{X7(CRAK)}$, constante de vidange du réservoir nappe $\theta_{X3(K)}$, constante de vidange du réservoir de routage $\theta_{X5(ROUT)}$, constante de la convolution $\theta_{X4(THU)}$. Le détail de sa structure est fourni dans la section 2.3.3.

Les variations relatives des paramètres $\theta_{X1(INSC)}$, $\theta_{X2(SMSC)}$ et $\theta_{X7(CRAK)}$, en fonction des conditions climatiques de transfert, sont présentées sur la figure 5.12 pour nos deux sous-groupes de bassins, les nuages obtenus pour tous les paramètres étant reportés en annexe (*cf.* figures G.9 à G.11). À la différence de *GR4J* ou *Mordor6*, le seul moyen pour le modèle *SimHyd* d'ajuster l'équilibre entre P, ETR et Q se fait via la mise à disposition d'eau dans des réservoirs conceptuels sujets à évaporation. Ceci est principalement fait à travers le réglage de la capacité du réservoir sol $\theta_{X2(SMSC)}$, mais aussi par la capacité du réservoir d'interception $\theta_{X1(INSC)}$ ou le coefficient $\theta_{X7(CRAK)}$ qui contrôle la part de l'eau rechargeant la nappe et donc le volume restant dans le réservoir sol.

Des variations sont visibles sur ces trois paramètres pour le sous-groupe de 120 bassins où les erreurs de modélisation sont corrélées aux écarts de température en transfert (graphiques sur haut). Elles sont particulièrement nettes sur le paramètre $\theta_{X2(SMSC)}$ (*cf.* faible hauteur des boîtes relativement à la pente globale). Les variations de $\theta_{X1(INSC)}$ et $\theta_{X2(SMSC)}$ sont en cohérence avec les biais constatés sur les débits moyens : des capacités de réservoir plus grandes que nécessaire conduisent à l'évaporation de volumes plus importants d'eau et donc

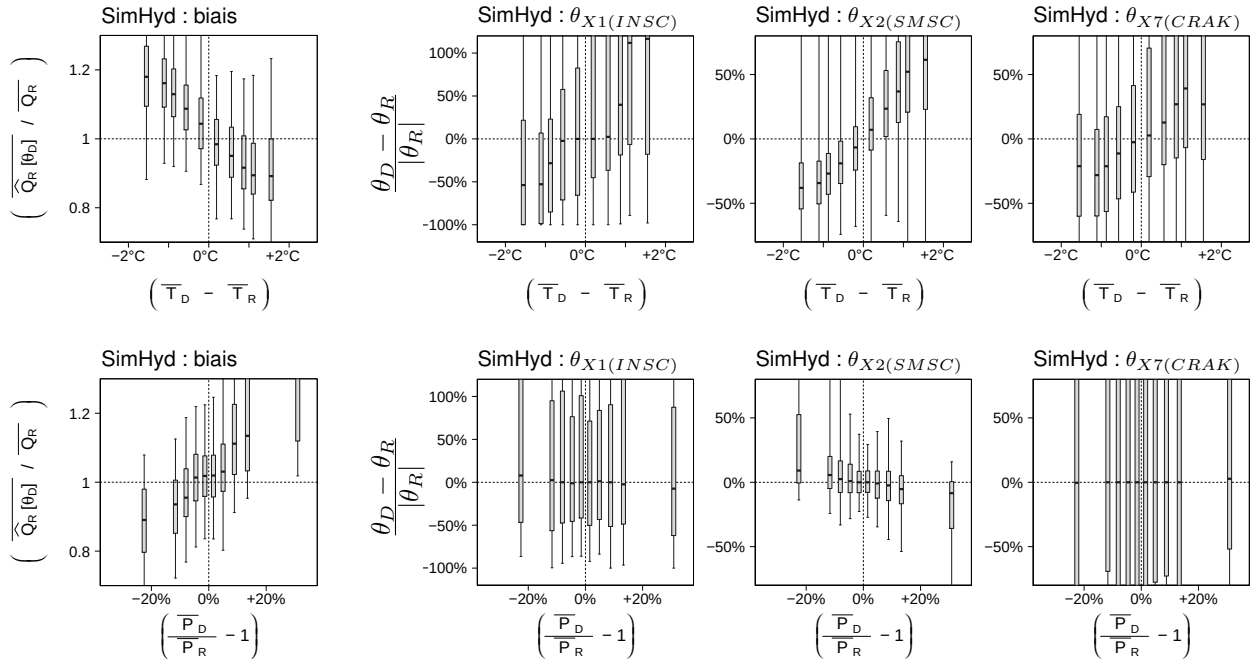


Figure 5.12.: Erreurs de volume et $\triangle\theta$ observés pour *SimHyd* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$, en haut, et $\triangle$erreurs et $\triangle\overline{P}$, en bas

Rq. sur les bassins français, SimHyd est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95%.

à une sous-estimation des débits. Concernant le paramètre $\theta_{X7(CRAK)}$ nous pourrions nous attendre à ce que les situations de sous-estimation correspondent à des valeurs plus faibles que nécessaire (*c.-à-d.* moins d'eau partant vers le réservoir de recharge, et donc plus d'eau restant dans le réservoir sol, sujet à évaporation). Or nous constatons ici des variations inversées. Cette inversion s'explique probablement par un lien entre les valeurs de $\theta_{X7(CRAK)}$ et de $\theta_{X2(SMSC)}$. En effet, les formulations au sein du modèle indiquent : (Recharge du réservoir nappe) = $\theta_{X7(CRAK)}$*(Taux de remplissage du réservoir sol)*(Volume d'eau entrant pour infiltration). À recharge vers la nappe équivalente, l'augmentation de $\theta_{X2(SMSC)}$ entraîne une diminution du taux de remplissage et donc une augmentation du $\theta_{X7(CRAK)}$. Or, le paramètre $\theta_{X2(SMSC)}$ semble la façon la plus simple d'ajuster les bilans. Celui-ci varie donc prioritairement, ce qui induit les variations constatées de $\theta_{X7(CRAK)}$.

Les résultats obtenus sur le second groupe de bassins sont beaucoup moins probants. La tendance sur $\theta_{X2(SMSC)}$ est à peine visible et nous constatons une grande variabilité des valeurs des paramètres, sans relation apparente avec le $\triangle\overline{P}$ (*cf.* figure G.10). Cette situation révèle probablement les limites de ce travail visant à interpréter des variations de paramètres (*cf.* discussion menée plus loin). Néanmoins, les résultats obtenus sur le groupe de 120 bassins où nous avions détecté un lien entre erreurs et $\triangle\overline{T}$ (*c.-à-d.* explication des biais par les variations de $\theta_{X2(SMSC)}$), nous semblent suffisamment nets pour qu'ils ne soient pas remis en cause.

Comportements observés pour *Cequeau*

Cequeau est un modèle conceptuel semi-distribué, utilisé ici dans une version où 19 paramètres sont calés et dont la structure est décrite dans la section 2.3.5. Ce modèle a seulement été utilisé sur quelques bassins de notre échantillon, parmi lesquels des bassins où nous avons retrouvé des corrélations entre les erreurs de simulation et les conditions de transfert des paramètres (*cf.* exemples fournis dans les sections 3.3.3 et 3.4.3). Nous avons tenté de mener des analyses visuelles sur les variations de paramètres, similairement à celles menées pour *GR4J*, *Mordor6* et *SimHyd*, à la seule différence que ce travail a été conduit par bassin plutôt que sur des nuages rassemblant plusieurs bassins. Cependant, ces analyses n'ont pas permis de retrouver, au niveau des paramètres, les signaux expliquant les pertes de performances constatées (notamment les biais systématiques sur les débits).

En guise d'exemple, nous présentons ici des nuages représentant les variations des paramètres *a priori* intéressantes pour nos analyses, sur deux bassins versants (*cf.* figures 5.13 et 5.14). Il s'agit des deux bassins modélisés par *Cequeau* pour lesquels les corrélations étaient les plus nettes entre $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{T}_{D\rightarrow R}$. D'une manière générale, les nuages de points y sont très diffus et ne font pas ressortir de corrélations telles que nous avions pu en rencontrer pour *GR4J*, *Mordor6* ou *SimHyd*. Il est cependant vrai que ces nuages ne sont représentés ici que sur un bassin à chaque fois et que l'utilisation des boîtes à moustaches pour les autres modèles

permet de mieux apprécier le niveau de densité des points (pas toujours visible sur les nuages obtenus avec *Cequeau*) et lisser les situations atypiques ponctuelles. Le fait que des tendances ne se détachent pas clairement pour ce modèle résulte probablement du nombre important de paramètres libres et des possibles interactions pouvant exister entre ceux-ci. Il est vrai que nous ne conduirons pas ici une identification précise des paramètres de *Cequeau* concernés par de telles interactions, néanmoins, la présence de 19 paramètres libres nous laisse supposer l'existence de multiples combinaisons possibles pour ajuster les équilibres entre P, ETR et Q.

Limites de l'approche

Le travail consistant à interpréter des variations de paramètres, plutôt que de se maintenir au niveau des variables simulées, présente un certain nombre de limites. Ces limites sont directement liées à la nécessité de conceptualiser la relation précipitations-débits au sein de modèles dont les paramètres doivent être calés (*cf.* discussion à la section 1.1.3).

Cette conceptualisation peut en effet induire des interactions entre les valeurs de certains paramètres. Ainsi, lorsque plusieurs paramètres ont un impact sur un flux d'eau au sein du modèle, leurs valeurs ne seront pas totalement indépendantes même si la relation associée est indirecte. C'est par exemple le cas de $\theta_{X2(CES)}$ et $\theta_{X3(ROUT)}$ pour *GR4J* qui déterminent le volume d'eau échangé avec des bassins voisins, ou encore des paramètres $\theta_{X2(SMSC)}$ et $\theta_{X7(CRAK)}$ pour *SimHyd* qui déterminent le volume d'eau transféré au réservoir conceptuel de nappe. Parallèlement, lorsque le nombre de paramètres libres augmente, les possibilités d'équifinalité des jeux de paramètres apparaissent (*c.-à-d.* le fait que différentes combinaisons de paramètres conduisent à la même simulation). Il devient alors plus difficile d'interpréter directement les valeurs prises par les paramètres lors de la phase de calage, et les nuages de points correspondant aux variations des paramètres deviennent très diffus. Cette situation est observée pour le modèle *Cequeau*, et probablement en partie pour le modèle *SimHyd*.

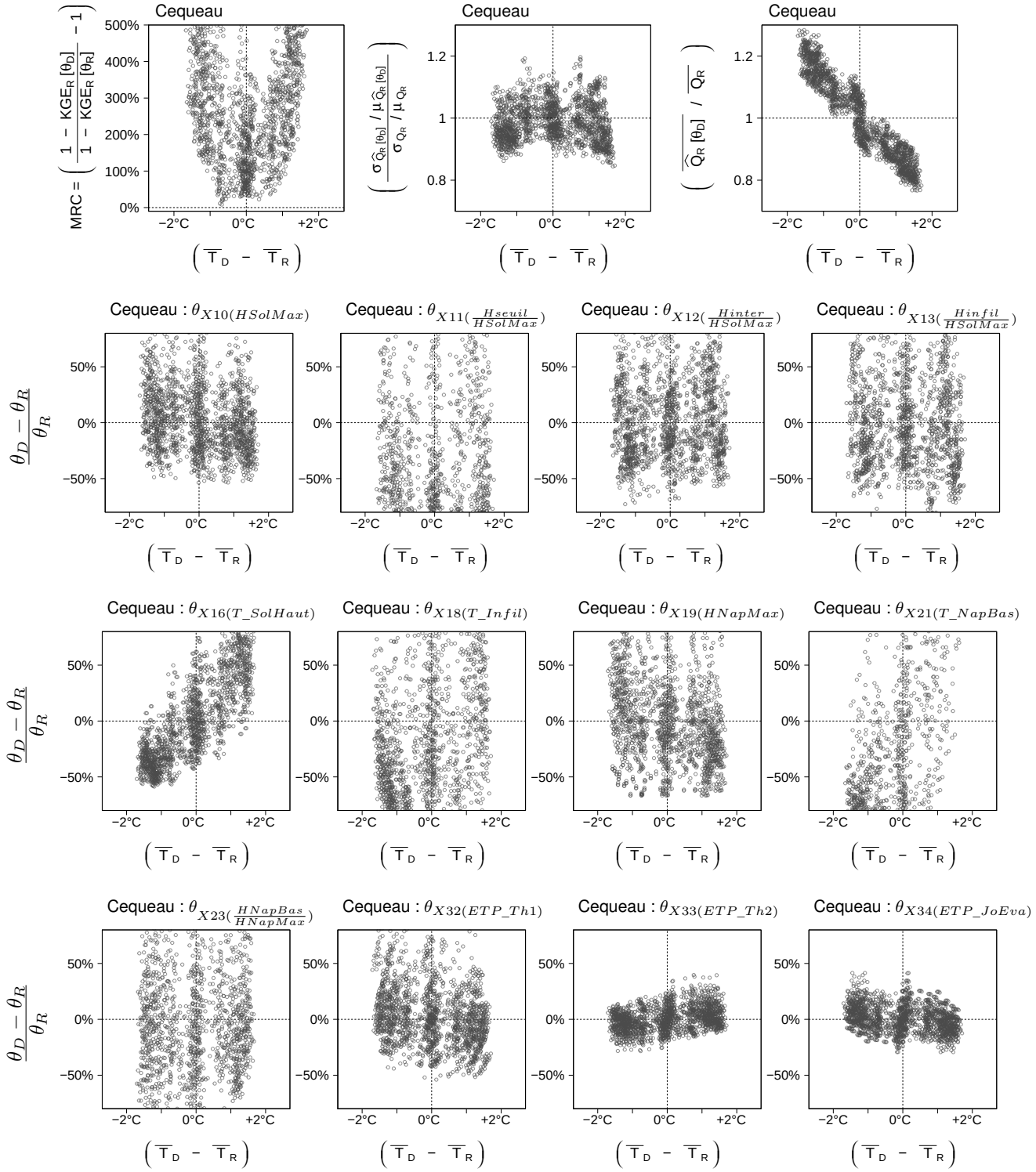


FIGURE 5.13.: Erreurs de volume et $\triangle\theta$ observés pour *Cequeau* lors des transferts entre périodes sur le bassin *du Lot à Banassac*

Rq. Cequeau est alimenté ici par des données SPAZM.
Il cale une ETP de Thornthwaite et dispose de son propre module neige.

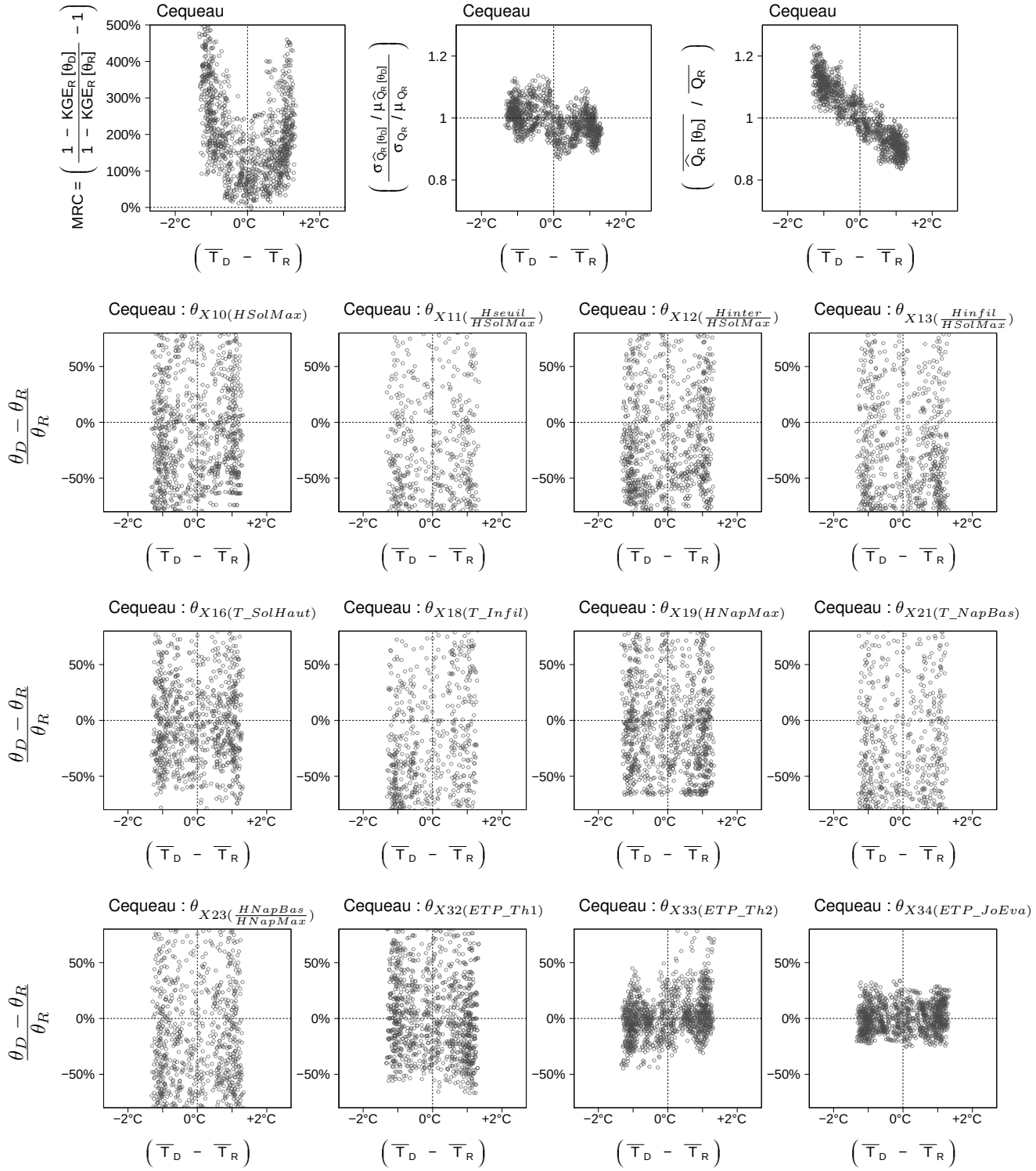


Figure 5.14.: Erreurs de volume et $\triangle\theta$ observés pour *Cequeau* lors des transferts entre périodes sur le bassin de *la Grande Creuse à Fresselines*

Rq. Cequeau est alimenté ici par des données SPAZM.
Il cale une ETP de Thornthwaite et dispose de son propre module neige.

5.5.3. Fixer les paramètres les plus sensibles est-il une solution ?

Point de départ

Pour les modèles les plus simples, il est possible de remonter le chemin des erreurs de modélisation. À partir des variations d'un critère composite tel que le *KGE*, nous pouvons ainsi remonter vers le type d'erreurs causant ces variations (principalement des biais) puis poursuivre jusqu'à la source que constituent les valeurs de paramètres sélectionnées au calage. Nous pouvons ainsi montrer comment les corrélations entre les erreurs de simulation et nos $\triangle caract.$ se retrouvent sur les paramètres qui permettent l'ajustement du bilan en eau du bassin. Selon la structure du modèle, il peut s'agir d'un coefficient correctif des précipitations, d'une variable permettant une augmentation/diminution de l'évapotranspiration ou encore d'un coefficient d'échange souterrain permettant l'export/import d'eau depuis/vers des bassins voisins.

Partant de ces constatations, nous avons souhaité conduire un dernier test avant de clôturer cette analyse des valeurs des paramètres. En menant un raisonnement raccourci et (peut-être) naïf, nous posons la question suivante : "*Il est possible d'identifier le(s) paramètre(s) trop dépendant(s) des conditions climatiques de calage pour certains modèles. Pourrions-nous simplement fixer ce(s) paramètre(s) pour résoudre nos problèmes de robustesse ?*"

Principe de l'exercice

Pour répondre à cette question, nous avons mené l'exercice suivant pour *GR4J* et *Mordor6* :

1. Nous repartons d'un groupe de bassins pour lequel nous avons pu identifier une corrélation entre les erreurs de modélisation et les conditions de transfert des paramètres. Pour cela, nous avons choisi d'utiliser le sous-groupe de 120 bassins où $\triangle\varepsilon = f(\triangle\overline{T})$ car il présente l'avantage de comporter un nombre élevé de bassins et de conduire à des nuages de points marqués par de nettes tendances *:* $MRC = f(\triangle\overline{T})$ ou $\triangle\theta = f(\triangle\overline{T})$.
2. À l'aide des figures 5.10 et 5.11, nous identifions pour chaque modèle le(s) paramètre(s) dont les variations au calage semblent le plus expliquer les problèmes de robustesse en transfert [7]. Pour cet exercice, nous choisissons de ne retenir qu'un paramètre par modèle : celui dont la tendance avec $\triangle\overline{T}$ est la plus nette, en moyenne sur nos 120 bassins ($\theta_{X2(CES)}$ pour *GR4J* et $\theta_{X1(CPL)}$ pour *Mordor6*).
3. Pour chaque bassin versant, nous fixons ce paramètre θ_{Xk}, à la valeur obtenue lors du calage sur toute la période disponible (celle-ci étant donc "réaliste" pour le bassin).
4. Nous recommençons ensuite notre procédure *GSST* puis les analyses graphiques associées. Le modèle est donc soumis aux mêmes calages et validations sur les sous-périodes de 10 ans, à la seule différence que l'un de ses paramètres est à présent maintenu fixe.

7. Maintenant que nous avons conclu qu'il s'agissait des paramètres permettant l'ajustement du bilan, un simple regard à la structure d'un modèle devrait nous permettre de deviner les paramètres à identifier ici.

5. Nous comparons enfin ces nouveaux résultats avec les anciens pour déterminer si les importants biais causés par les transferts de paramètres sont résorbés et/ou si les valeurs de paramètres retenues au calage ont été modifiées.

Résultats

Les résultats détaillés de cet exercice sont fournis dans l'annexe G.3. Les informations principales de ces résultats sont reprises sur les figures 5.15 et 5.16. Comparativement à la situation initiale (où tous les paramètres étaient calés), nous retrouvons quasiment les mêmes tendances sur les biais en validation. Pour chacun des modèles, le fait de fixer le paramètre dont les variations étaient reliées aux conditions de transfert n'a donc pas de réel impact en termes de réduction des problèmes de robustesse (mais impacte tout de même les performances globales du modèle). L'analyse des valeurs des paramètres nous montre que les tendances précédemment observées sur les paramètres $\theta_{X2(CES)}$ pour *GR4J* et $\theta_{X1(CPL)}$ pour *Mordor6* sont *transférées* vers d'autres paramètres. Les paramètres concernés sont ceux qui permettent de réaliser, en second recours, un ajustement du bilan maintenant que les paramètres d'ajustement les plus directs sont maintenus fixes.

Discussion

Le petit exercice mené ici sur deux modèles hydrologiques (*GR4J* et *Mordor6*) permet de confirmer un résultat déjà entrevu à plusieurs reprises au fil de nos analyses : l'ajustement du bilan au calage et les erreurs associées en validation (sur les volumes moyens) constituent un point central dans l'explication des problèmes de robustesse temporelle des modèles hydrologiques mis en évidence par le *GSST*.

Par son apparente simplicité, cet exercice de fixation de certains paramètres, incite à une prise de distance sur la façon dont nous pouvons concevoir ou appréhender le fonctionnement d'un modèle conceptuel et le rôle respectif de ses paramètres. Bien que la formulation suivante soit (peut-être) trop brutale, cet exercice nous rappelle qu'un modèle reste souvent *un agglomérat de blocs partiellement ou intégralement conceptuels*, qui vise à reproduire des observations, dans un cadre où coexistent contraintes de structure et degrés de liberté (via les paramètres ajustables). Le fait de fixer l'un des paramètres d'un modèle ne bouleversera pas nécessairement sa capacité à simuler des débits, mais doit nous amener à considérer celui-ci comme un modèle différent du précédent. Ce nouveau modèle dispose alors de contraintes de structure identiques mais de degrés de liberté différents. En conséquence, la phase de calage pourra potentiellement conduire à une redistribution des rôles des paramètres, qui procurent ces degrés de liberté. Par exemple, un paramètre influant sur les volumes évaporés pourra servir à représenter la sous-captation des pluies ou l'export d'eau souterraine vers d'autres bassins. Cette situation est observée ici pour *GR4J* et *Mordor6* où le fait de fixer les paramètres qui ajustent le bilan en eau du système entraîne un transfert de ce rôle vers d'autres paramètres, sans que les différences sur les performances en calage et en validation ne soient très marquées.

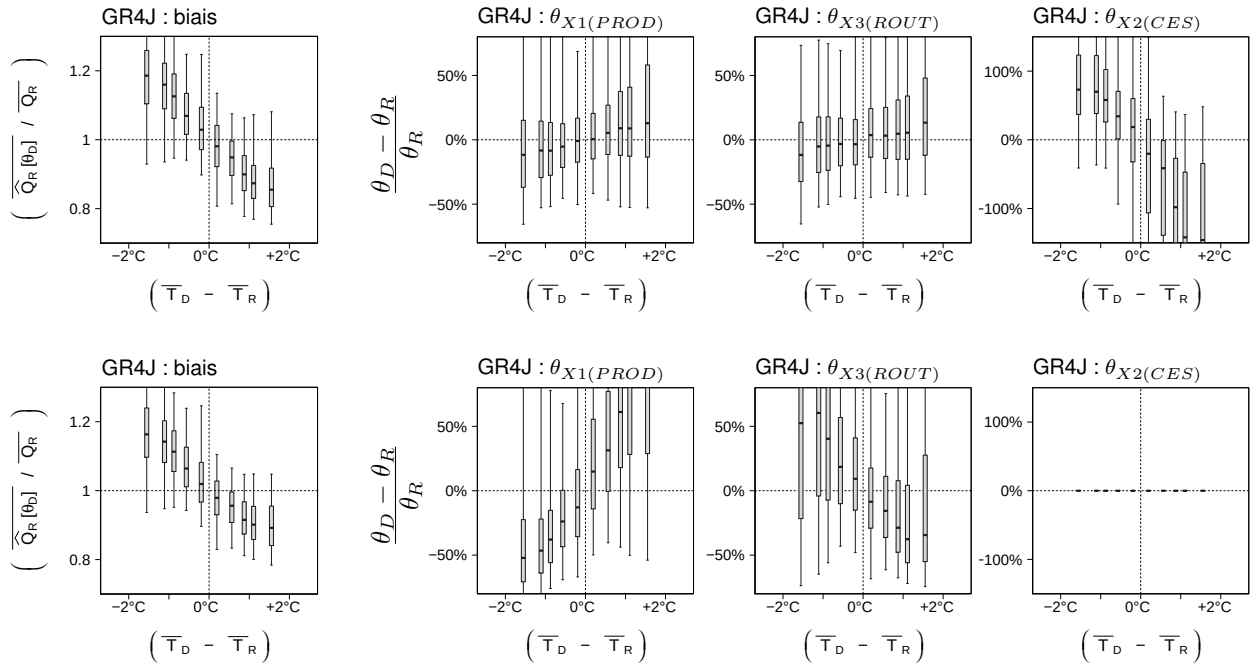


FIGURE 5.15.: Biais et $\triangle\theta$ observés pour *GR4J* lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$

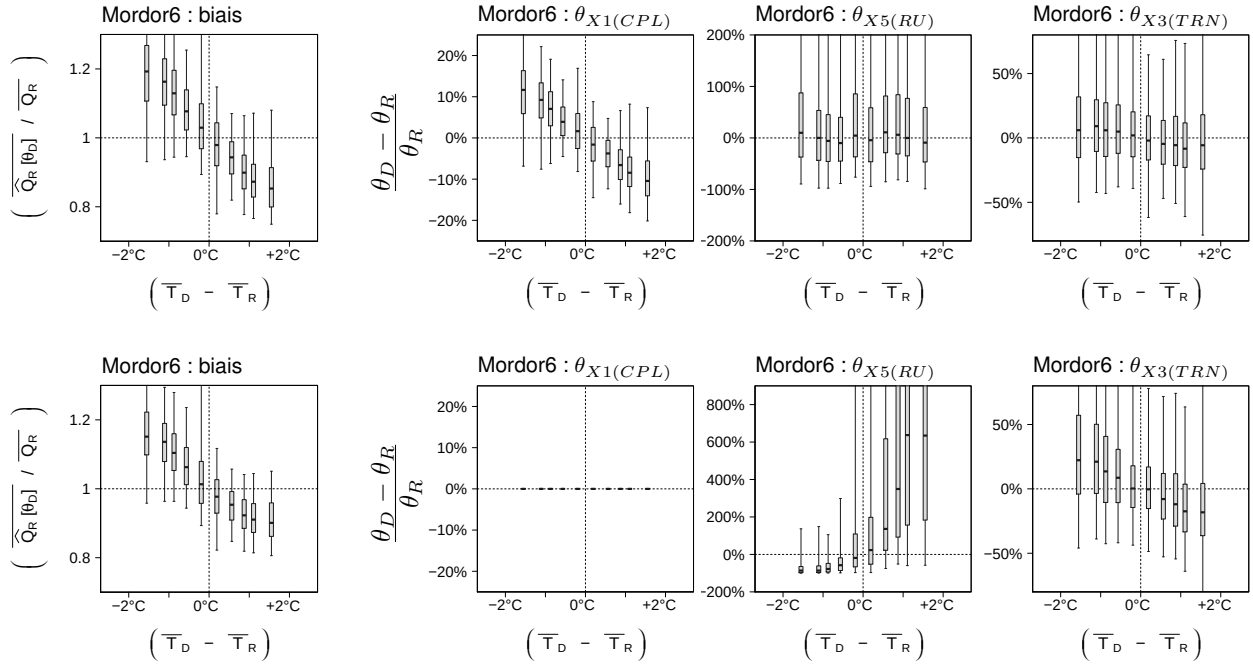


FIGURE 5.16.: Biais et $\triangle\theta$ observés pour *Mordor6* lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$

GR4J et Mordor6 sont alimentés par des données SPAZM, une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

5.6. Analyse des biais glissants de modélisation

Les travaux d'évaluation de la robustesse temporelle des modèles hydrologiques nous ont progressivement conduits à étudier en détail la question de l'équilibrage des bilans et son rôle explicatif des variations de performances constatées lors des transferts des jeux de paramètres. Les symétries constatées sur les nuages représentant $\overline{\widehat{Q}}/\widehat{Q} = f(\triangle caract.)$ sont particulièrement étonnantes. Par exemple, si un transfert effectué d'une période donneuse (A) vers une période receveuse (B) conduit à une surestimation moyenne de 10% des débits, alors une inversion des rôles (*c.-à-d.* un transfert de B vers A) conduit à une sous-estimation des débits de 10%. À travers une analyse des biais glissants, nous souhaitons avancer ici dans notre compréhension des causes des cas de manque de robustesse temporelle des modèles, et de cette symétrie quelque peu surprenante sur les erreurs de volume commises en validation.

5.6.1. Mode de représentation

Pour comprendre ces phénomènes, nous avons étudié les variations temporelles du biais pour chaque bassin en procédant de la manière suivante :

1. pour chaque bassin, nous reprenons la bibliothèque de jeux de paramètres obtenue par calage successifs sur les sous-périodes de 10 ans (selon la méthode de découpage illustrée à la figure 2.2, page 29) ;
2. chacun de ces jeux est utilisé pour simuler la période complète disponible du bassin considéré puis l'erreur sur le volume moyen est calculée sur une fenêtre glissante de 10 ans également [8] ;
3. ces courbes de biais glissants sont analysées à travers une représentation visuelle dans laquelle les différentes sous-périodes sont placées sur l'axe des abscisses et les biais de modélisation correspondants ($\overline{\widehat{Q}}/\overline{Q}$) sur l'axe des ordonnées ;

Pour faciliter la compréhension de ces représentations, décomposons leur construction :

. **Figure 5.17.a & 5.17.b**
Commençons par ne considérer qu'un seul jeu de paramètres θ_{D_1}, issu d'un calage sur l'une des sous-périodes (D_1). Cette sous-période de calage est représentée par un point de plus grande taille afin de la distinguer des autres sous-périodes sur lesquelles le jeu θ_{D_1} n'est plus optimal. Compte tenu du critère de calage utilisé, l'erreur de volume obtenue sur cette période de calage est nécessairement presque inexistante ($\overline{\widehat{Q}}/\overline{Q} \backsimeq 1$, *cf.* graphique a) [9].
Nous étudions les biais obtenus sur une fenêtre glissante de 10 ans qui correspondent aux

8. Cette façon de procéder est équivalente au fait de rependre les biais obtenus pour tous les transferts effectués via le *GSST*, en y rajoutant les transferts entre périodes non indépendantes jusque-là écartées.

9. L'importance du biais dans la fonction objectif n'est pas spécifique à nos conditions de tests (utilisation du critère *KGE*) mais se vérifie pour une grande majorité des critères de calage couramment utilisés.

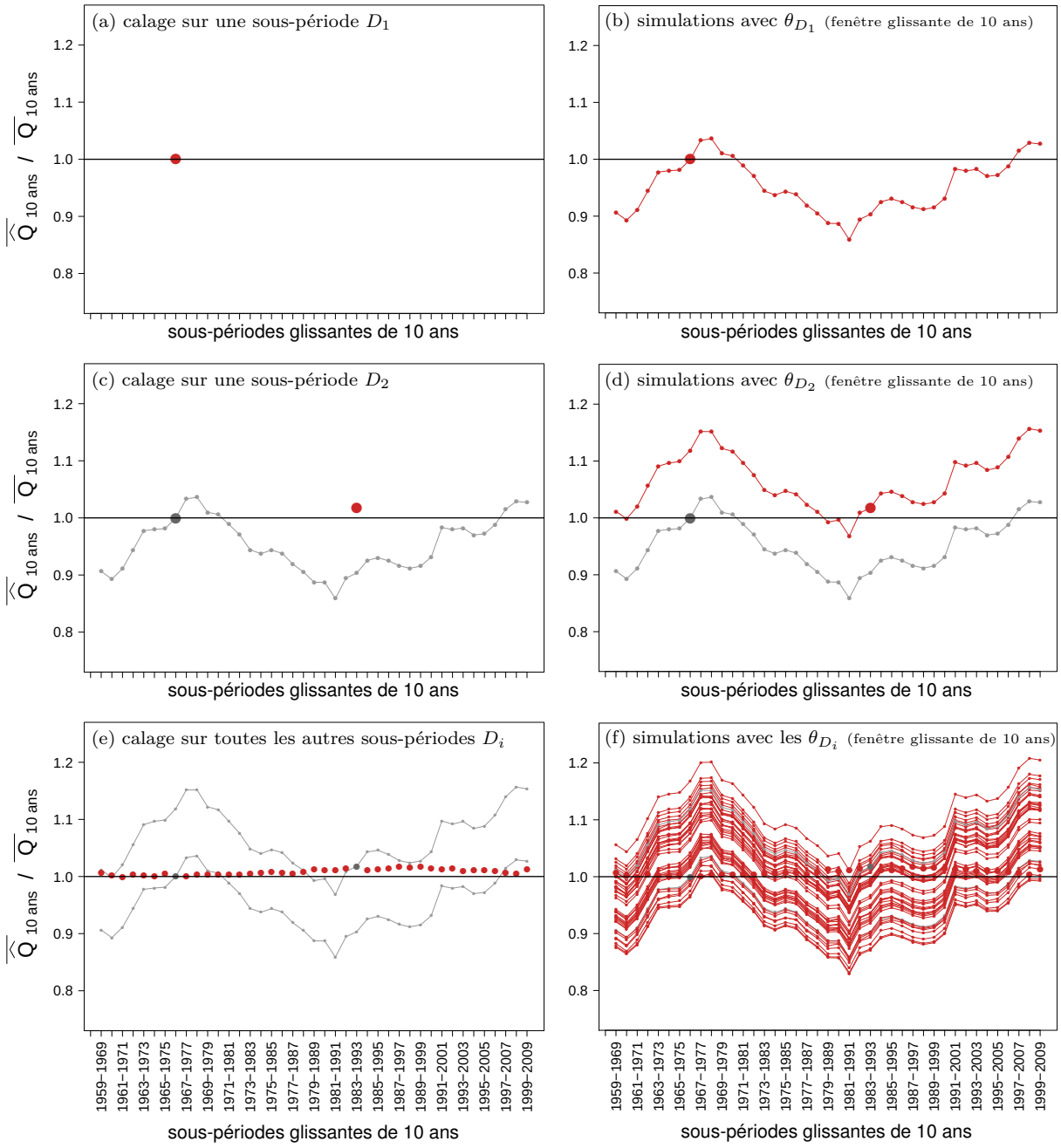


FIGURE 5.17.: Décomposition de l'analyse visuelle des biais glissants

Courbes des biais sur 10 ans glissants obtenus avec GR4J sur le bassin de l'Ubaye à Barcelonnette. Chaque courbe comporte un point de plus forte taille indiquant la sous période de calage. GR4J-CemaNeige (θ_{neige} fixés) est alimenté ici par des données SPAZM et une ETP d'Oudin.

rapports $\overline{\widehat{Q}}_{R_i}[\theta_{D_1}]/\overline{Q}_{R_i}$ pour toutes les sous-périodes R_i (*cf.* graphique b). La courbe ω_{D_1} ainsi construite montre l'existence de situations de sous-estimation et de surestimation. Précisons que le fait d'avoir identifié ou non, une caractéristique climatique qui varie de façon corrélée avec le biais glissant n'a pas d'importance dans la démonstration menée ici.

- **Figure 5.17.c & 5.17.d**
  Un second jeu de paramètres (θ_{D_2}) est considéré et le même exercice de représentation est reproduit. Ici encore, l'erreur de volume moyen sur la période D_2 est presque inexistante (*cf.* graphique c) et nous constatons des alternances de périodes de sous-estimation et de surestimation (*cf.* graphique d).
  Nous notons alors que les deux courbes obtenues ($\omega_{D_1} = \overline{\widehat{Q}}_{R_i}[\theta_{D_1}]/\overline{Q}_{R_i}$ & $\omega_{D_2} = \overline{\widehat{Q}}_{R_i}[\theta_{D_2}]/\overline{Q}_{R_i}$) ont des formes très proches. Un point les différencie néanmoins : un décalage vertical homogène. Nous pourrions presque croire qu'une translation verticale a été utilisée pour passer d'une des courbes à l'autre, mais elles résultent bien de deux calages indépendants sur des sous-périodes différentes.

- **Figure 5.17.e & 5.17.f**
  Nous recommençons cet exercice pour tous les jeux de paramètres à disposition (jeux obtenus par calage sur la fenêtre glissante de 10 ans).
  Nous vérifions que les erreurs de volume restent faibles en calage (*cf.* graphique e), puis nous étudions les courbes de biais glissant et retrouvons le même *parallélisme* surprenant entre celles-ci (*cf.* graphique f).

La figure G.16 (page 329 en annexe) donne un exemple d'utilisation de ce mode de représentation, où nous avons placé côte à côte des exemples de nuages de points tels qu'ils furent utilisés jusqu'ici et les courbes de biais glissants proposées ici.

5.6.2. Quelques comparaisons multi-bassins et multi-modèles

Les résultats que nous venons de présenter sur la figure 5.17 sont assez surprenants. Avant de les discuter, prolongeons encore un peu les analyses visuelles en répondant aux trois points suivants :

- Nous avons constaté une similitude des comportements entre nos calages sur différentes sous-périodes de 10 ans mais comment se comporte un calage obtenu sur toute la période disponible ? Est-il possible qu'un calage sur une période trop courte ne permette pas au modèle pas de capter une dynamique lente qu'il arriverait à représenter s'il était calé sur une période étendue ?
- Nous avons constaté une situation particulière avec le modèle *GR4J-CemaNeige*, mais est-elle spécifique à ce modèle ? Une structure de modélisation et un niveau de complexité différents conduiraient-ils à d'autres résultats ?

- Nous avons constaté une situation sur un bassin versant (*l'Ubaye à Barcelonnette*), mais est-elle spécifique à ce bassin versant ? La similitude des comportements entre différents calages est-elle constatée sur d'autres bassins versants ?

Comparaison avec des calages sur toute la série disponible

Le graphique 5.18.a constitue un prolongement de la décomposition menée sur la figure 5.17, puisque nous avons seulement ajouté (en rouge) la courbe de biais glissant obtenue pour un calage sur la période complète (1959-2009). Elle montre très clairement que la mise à disposition de tous les pas de temps au calage ne permet pas une réduction des variations de biais sur 10 ans glissants, par la captation d'un signal lent indiscernable lors d'un calage sur une sous-période (résultat retrouvé sur les autres modèles et bassins).

Comparaison avec d'autres bassins

La figure 5.18 nous montre que ce parallélisme des courbes du biais glissant n'est pas propre à un cas d'étude. En réalité, nous l'avons retrouvé sur tous les bassins de notre échantillon, même si quelques fois, les calages (peut-être incorrects) sur une ou deux sous-périodes altèrent partiellement le degré de parallélisme. Mentionnons toutefois les situations où les biais glissants sont presque stables temporellement pour un jeu de paramètres et où nous observons alors des courbes des biais (ω_{D_i}) non seulement plates pour tous les jeux de paramètres (calés sur des sous-périodes) mais aussi quasiment superposées (*cf.* figure 5.18.d).

Comparaison avec d'autres modèles

Les différents graphiques de la figure 5.19 montrent le peu d'influence du choix du modèle sur ces variations de biais. Des résultats semblables sont constatés pour des modèles journaliers peu paramétrés mais de structures différentes, tels que *GR4J*, *Mordor6*, *SimHyd* (*cf.* graphiques 5.19.a-b-c) [10]. Ces résultats sont également proches pour des modèles plus complexes tels que *Cequeau* ou *Mordor* (*cf.* graphiques 5.19.d-e). Nous aurions pu penser que les calages sur toute la période de ces modèles beaucoup plus paramétrés permettraient une meilleure captation de certains processus que les modèles plus simples. Nous constatons ici que ce n'est pas le cas, un comportement similaire étant même obtenu sur un modèle beaucoup plus simple : le modèle annuel de *Mouelhi et al.* (2006) (*cf.* graphique 5.19.f).

10. *GR4J*, *Mordor6* et *SimHyd* sont couplés ici avec *CemaNeige* (θ_{neige} fixés). Les mêmes analyses menées avec *CemaNeige* calé donnent toutefois des résultats identiques à ceux présentés ici, *cf.* figure 5.22.

Pour rappel, les structures des différents modèles conceptuels sont présentées de façon détaillée dans la section 2.3 et le modèle annuel de *Mouelhi,* qui ne comporte qu'un seul paramètre (α), peut être résumé par la formule suivante :

$$Q_{an(i)} = P_{an(i)} \left(1 - 1/ \left[1 + \left(\frac{0.7 \cdot P_{an(i)} + 0.3 \cdot P_{an(i-1)}}{\alpha \cdot ETP_{an(i)}} \right) \right]^{0.5} \right) \tag{5.1}$$

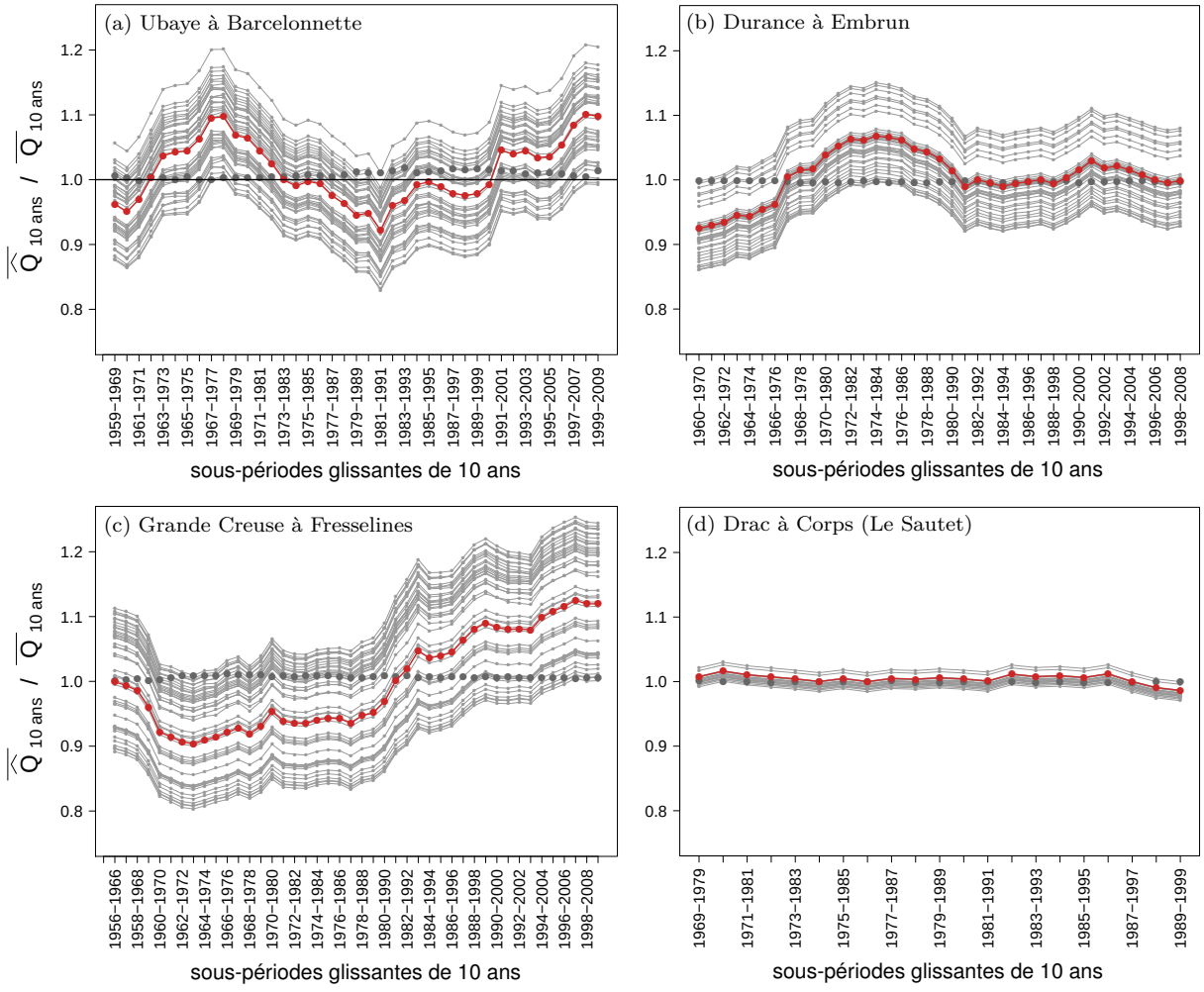


Figure 5.18.: Analyse visuelle multi-bassins des biais glissants sur 10 ans

Tests réalisés avec GR4J + CemaNeige $_{\theta fixés}$, alimenté par SPAZM et une ETP d'Oudin. Courbes liées aux calages sur les sous-périodes de 10 ans en gris et sur la période complète en rouge (chaque courbe grise comporte un point de plus forte taille indiquant la sous période de calage).

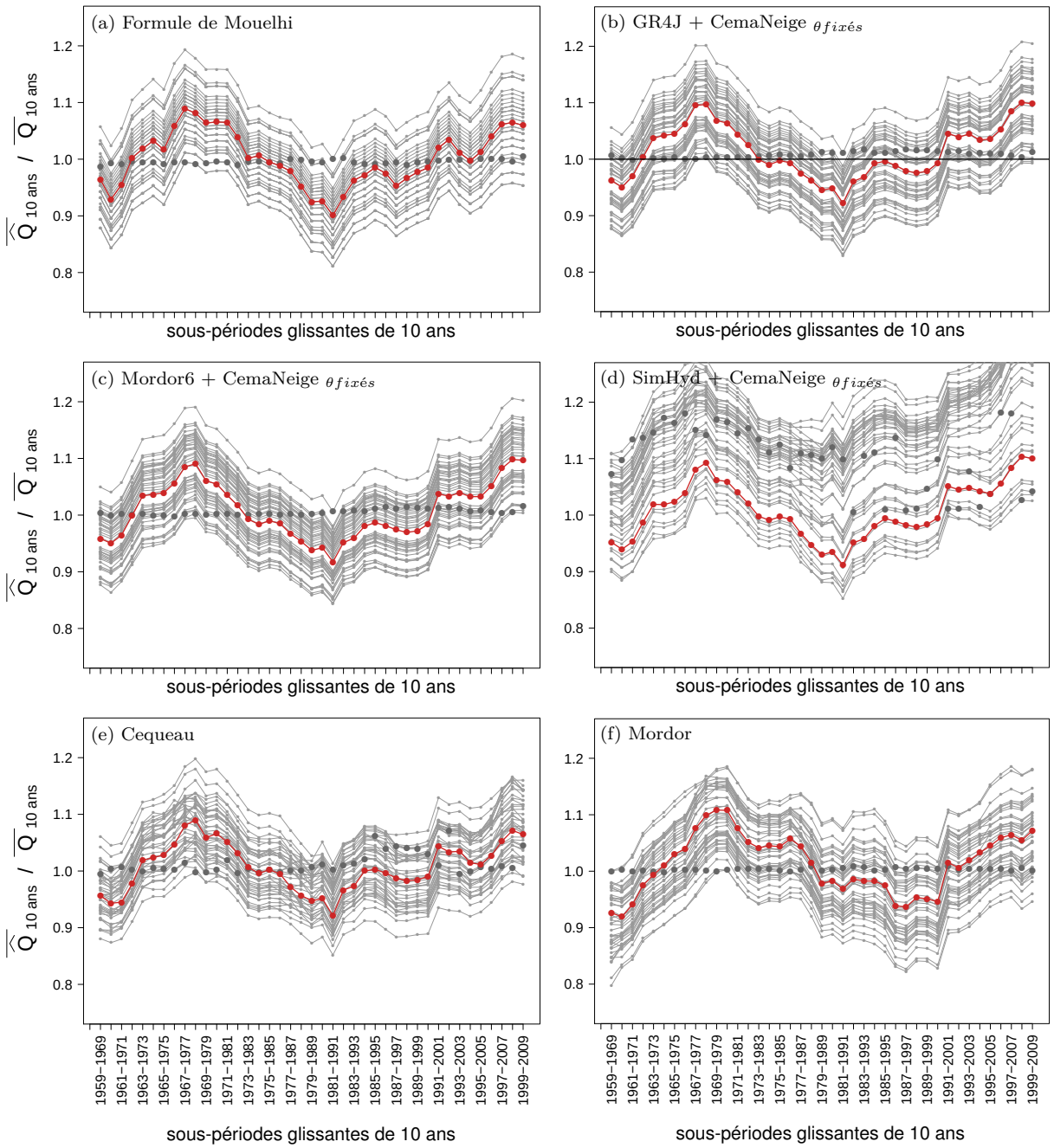


FIGURE 5.19.: Analyse visuelle multi-modèles des biais glissants sur 10 ans

Simulations réalisées par différents modèles sur le bassin de l'Ubaye à Barcelonnette. Courbes liées aux calages sur les sous-périodes de 10 ans en gris et sur la période complète en rouge (chaque courbe grise comporte un point de plus forte taille indiquant la sous période de calage).

Tous les modèles sont ici alimentés par des données SPAZM. GR4J, Mordor6 et SimHyd sont couplés à CemaNeige et utilisent une ETP d'Oudin. Mouelhi utilise aussi une ETP d'Oudin tandis que Cequeau et Mordor calent leur propre ETP et disposent d'un module neige interne.

5.6.3. Interprétations des similarités et liens avec les constatations précédentes

Soyons honnêtes, ce parallélisme entre les courbes du biais glissant, obtenu pour les différents calages sur notre découpage temporel nous surprend. Il permet cependant de comprendre la plupart des problèmes de robustesse discutés jusqu'ici.

Prenons une situation où le calage sur la période complète induit des phases de sous-estimation et surestimation des débits moyens formant ainsi un *signal* de variations relatives du biais glissant *(ω)*. Nos analyses montrent que : (1) tous les jeux de paramètres conduisent au même *signal* (c'est le parallélisme que nous évoquions, *ex.* $\omega_{D_1} \| \omega_{D_2}$) ; (2) un calage sur une sous-période D conduit (nécessairement) à un faible biais sur celle-ci (*cf.* ronds gris foncé de plus grande taille sur la figure 5.18). La coexistence de ces deux aspects aboutit à une situation pouvant être résumée de la façon suivante : *pour chaque jeu de paramètres (θ_D), la courbe de "signal du biais" (ω_D) se positionne verticalement de telle sorte que l'erreur moyenne de volume en calage soit faible (c.-à-d.* $\overline{\widehat{Q}_D}[\theta_D]/\overline{Q_D} \backsimeq 1$*).*

Une conséquence majeure de ce comportement est que le positionnement de la sous-période D au sein la courbe ω_D déterminera les biais des simulations associées ($\overline{\widehat{Q}}_{R_i}[\theta_D]/\overline{Q}_{R_i}$). Si D constitue le point le plus haut du *signal* ω, alors tous les transferts de paramètres conduiront à des sous-estimations des débits moyens (et inversement). Le fait que ce *signal* ω soit quasiment identique pour tous les jeux explique la forte symétrie observée sur nos nuages de points $\overline{\widehat{Q}}/\overline{Q} = f(\triangle caract.)$.

Ainsi, les erreurs de volume que nous observons lors des transferts des paramètres ne proviennent pas de mauvais calages ou de calages trop spécifiques à une période. Elles résultent d'une incapacité générale du modèle à reproduire les *bilans moyen terme* (sur 10 ans) observés sur certains bassins. Les transferts de paramètres conduisant aux plus forts biais en validation sont simplement ceux effectués entre les sous-périodes "antagonistes" du point de vue de cette incapacité (*c.-à-d.* les périodes formant des points hauts et des points bas sur la courbe ω obtenue par un seul calage sur la période complète). Lorsque le modèle représente correctement les bilans temporels (*c.-à-d.* courbe ω plate), les calages sur des périodes plus courtes n'entraînent pas d'erreur sur les volumes moyens lors des transferts de paramètres vers d'autres périodes (*cf.* figure 5.18.d).

5.6.4. Quelle généralisation de ces résultats ?

Critères de caractérisation de l'homogénéité des comportements

L'étude par bassin des courbes ω de signal biais obtenues pour différentes sous-périodes de calage et différents modèles hydrologiques met en avant les mêmes parallélismes et similarités de forme. Les résultats présentés précédemment sur quelques exemples se retrouvent donc sur tout notre échantillon et nous cherchons ici à caractériser cette homogénéité des comportements à travers la construction de deux critères numériques (détail en annexe G.4) :

$$\rho_i = \sigma\left[\omega_{SP[i]}^{M_X} - \omega_{PT}^{M_X}\right] \,/\, \sigma\left[\omega_{PT}^{M_X}\right] \qquad et \qquad \rho'_{M_1\,,\,M_2} = \sigma\left[\omega_{PT}^{M_2} - \omega_{PT}^{M_1}\right] \,/\, \sigma\left[\omega_{PT}^{M_1}\right] \tag{5.2}$$

où σ représente l'écart-type et où $\omega_{SP[i]}^{M_X}$ et $\omega_{PT}^{M_X}$ sont respectivement
les biais glissants obtenus avec les jeux de paramètres du modèle M_X
calé sur la sous-période $SP[i]$ et la période totale PT

Le premier critère (ρ_i) mesure le degré de parallélisme des comportements entre différents calages d'un même modèle sur chaque bassin. Le second critère ($\rho'_{M_1\,,\,M_2}$) mesure par bassin le degré de parallélisme des comportements entre différents modèles hydrologiques (seul le calage sur la période totale étant considéré à chaque fois). Dans les deux cas, des valeurs proches de 0 correspondent à un parallélisme maximal. Avant de passer à l'analyse des valeurs de ρ_i et $\rho'_{M_1\,,\,M_2}$, jetons un œil aux variations temporelles de biais constatées sur notre échantillon de bassins. La figure 5.20 nous donne les distributions des valeurs de $\sigma\left[\omega_{PT}^{M_X}\right]$, c'est-à-dire des écarts-types des biais glissants obtenus pour un calage sur toute la période disponible de chaque bassin. En moyenne sur notre échantillon complet, les biais glissants sur 10 ans présentent un écart-type supérieur à 5%, avec des valeurs nettement plus élevées pour certains bassins. Ces valeurs d'écart-type sont élevées et révèlent l'existence de sous-périodes de 10 ans sur lesquelles les erreurs sur le volume moyen atteignent probablement 10 voire 20%.

Similarités entre différents calages temporels

Pour chaque bassin, les courbes $\omega_{SP[i]}$ (obtenues par calage d'un modèle sur différentes sous-périodes) sont comparées à la courbe ω_{PT} (obtenue par calage du même modèle sur toute la période du bassin). Elles sont représentées en gris sur les figures 5.18 et 5.19, les courbes ω_{PT} étant indiquées en rouge. La formulation adimensionnelle du critère ρ_i permet sa comparaison entre différents calages et différents bassins. En rassemblant les valeurs de ρ_i obtenues sur tout notre échantillon puis en exprimant quelques quantiles, il devient possible d'avoir un aperçu rapide et complet du parallélisme illustré jusqu'ici sur quelques exemples seulement.

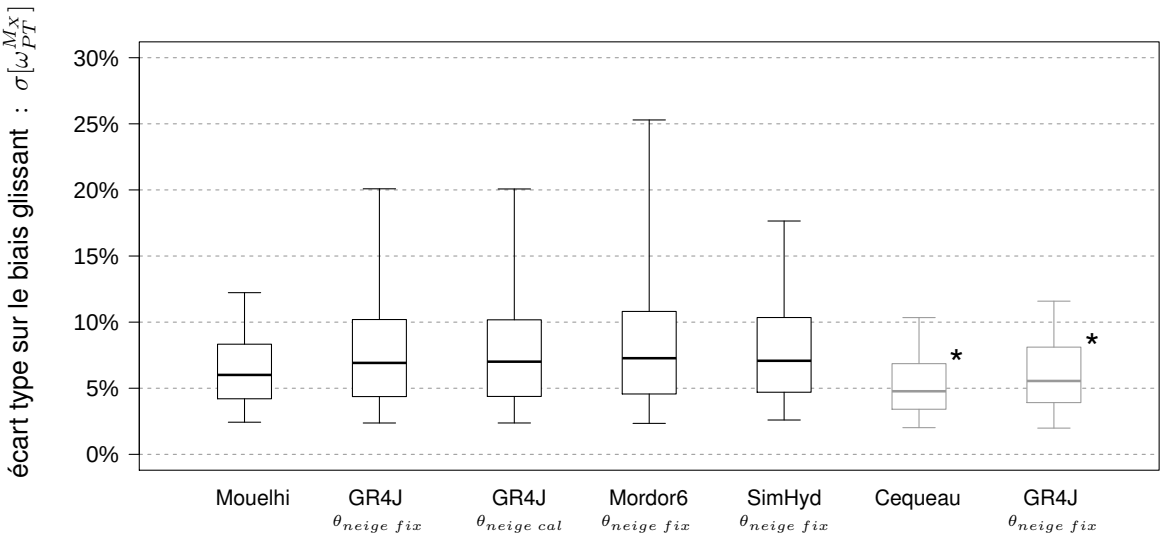


Figure 5.20.: Distributions de l'écart-type des biais glissants sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$, pour tout l'échantillon (569 bassins) et pour l'échantillon français seulement (365 bassins)

** Écarts-types calculés seulement sur les 365 bassins français.*
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% des distributions.

Tous les modèles sont ici alimentés par des données SPAZM. GR4J, Mordor6 et SimHyd sont couplés à CemaNeige et utilisent une ETP d'Oudin. Mouelhi utilise aussi une ETP d'Oudin tandis que Cequeau cale une ETP de Thornthwaite et dispose d'un module neige interne.

Ces résultats sont présentés sur la figure 5.21, pour chacun des modèles testés. Les valeurs médianes inférieures à 0.1 révèlent qu'en moyenne, les "défauts" de parallélisme entre $\omega_{SP[i]}$ et ω_{PT} (*c.-à-d.* entre une courbe grise et une courbe rouge) sont 10 fois plus faibles que les variations temporelles des biais constatées sur les bassins.

Il est intéressant de noter que le parallélisme diminue avec l'augmentation du degré de paramétrisation des modèles. D'une manière générale, nous retrouvons toutefois que le fait de caler un modèle sur une période temporelle réduite (10 ans) entraîne relativement peu de différence du point de vue du signal ω. Ce comportement, très prononcé pour *GR4J* et pour *Mordor6*, révèle donc que l'origine des forts biais observés lors de la procédure *GSST* n'est pas une mauvaise sélection de paramètres et que ceux-ci ne peuvent pas être solutionnés par un calage sur une série plus longue. Le problème identifié semble être tout autre : les modèles n'arrivent pas toujours à reproduire simultanément tous les bilans hydrologiques moyen terme (10 ans).

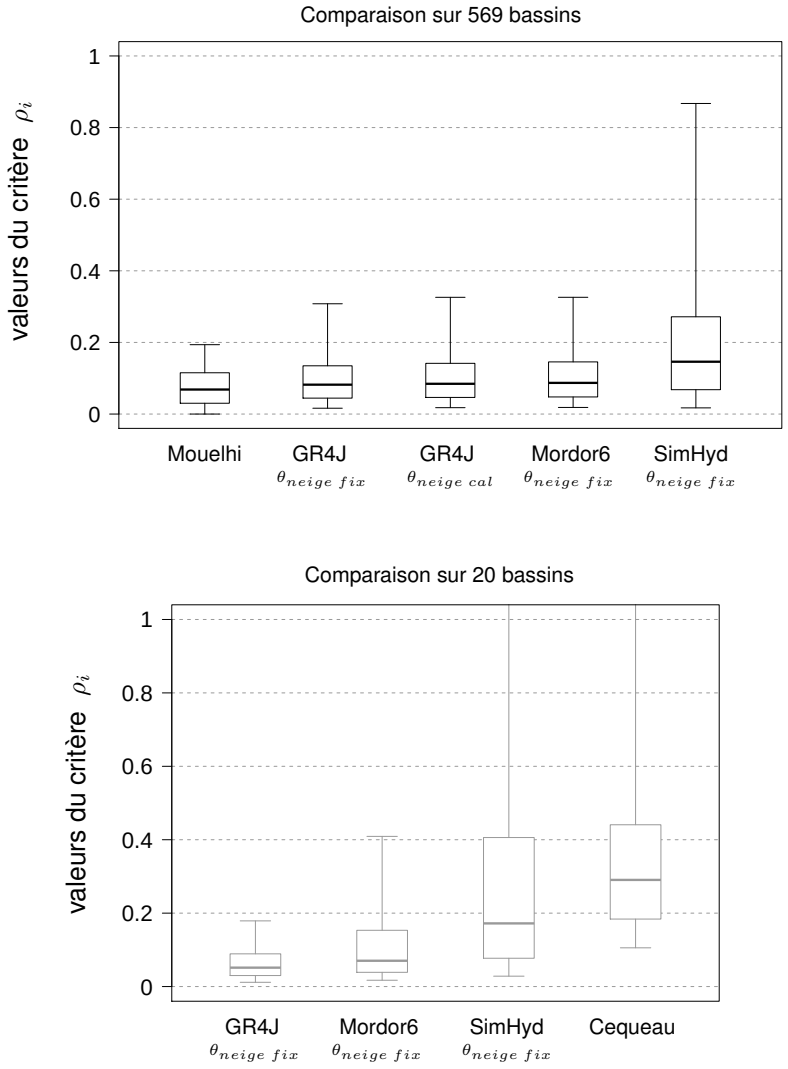


FIGURE 5.21.: Distributions du critère ρ_i mesurant la similitude des biais glissants entre différents calages temporels des modèles

** Écarts-types calculés seulement sur les 365 bassins français.*
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% des distributions.

Tous les modèles sont ici alimentés par des données SPAZM. GR4J, Mordor6 et SimHyd sont couplés à CemaNeige et utilisent une ETP d'Oudin. Mouelhi utilise aussi une ETP d'Oudin tandis que Cequeau cale une ETP de Thornthwaite et dispose d'un module neige interne.

Similarité entre différents modèles hydrologiques

Pour chaque bassin, nous comparons désormais les courbes de biais glissants (ω_{PT}), leurs similarités étant mesurées à l'aide du critère $\rho'_{M_1\,,\,M_2}$. En guise de courbe de référence, nous utilisons ici le signal obtenu pour *GR4J-CemaNeige* ($\theta_{CemaNeige}$ fixés). Cette analyse comparative, est conduite sur la totalité de notre échantillon (569 bassins), sauf pour *Cequeau* qui n'est comparé à *GR4J-CemaNeige* que sur les 365 bassins français. Les résultats correspondants sont présentés sur la figure 5.22. Ici encore, des valeurs proches de zéro indiquent un parallélisme maximum.

Les similitudes inter-modèles sont moins marquées que les similitudes inter-périodes (distributions de ρ_i *vs.* distributions de $\rho'_{M_1\,,\,M_2}$). Les valeurs de $\rho'_{M_1\,,\,M_2}$ indiquent des différences de degré de similitude selon les modèles considérés. Le modèle *GR4J* couplé à *CemaNeige* calé est logiquement très proche de *GR4J* couplé à *CemaNeige* non calé. Ce degré de similitude augmente lorsque l'on considère d'autres modèles tels que *Mordor6*, *SimHyd*, *Cequeau* et le modèle annuel de *Mouelhi*. Nous ne nous risquerons cependant pas ici à interpréter ces variations, mais nous contenterons de constater que les $\rho'_{M_1\,,\,M_2}$ demeurent toujours inférieurs à 1 et souvent inférieurs à 0.5, avec des médianes entre 0.2 et 0.4 pour les différentes comparaisons. Ces valeurs confirment donc que des modèles de structures pourtant différentes ont des comportements relativement proches en termes d'erreur sur les débits moyens. Autrement dit, tous les modèles testés ici rencontrent des difficultés semblables à reproduire les bilans hydrologiques observés sur différentes plages des séries disponibles.

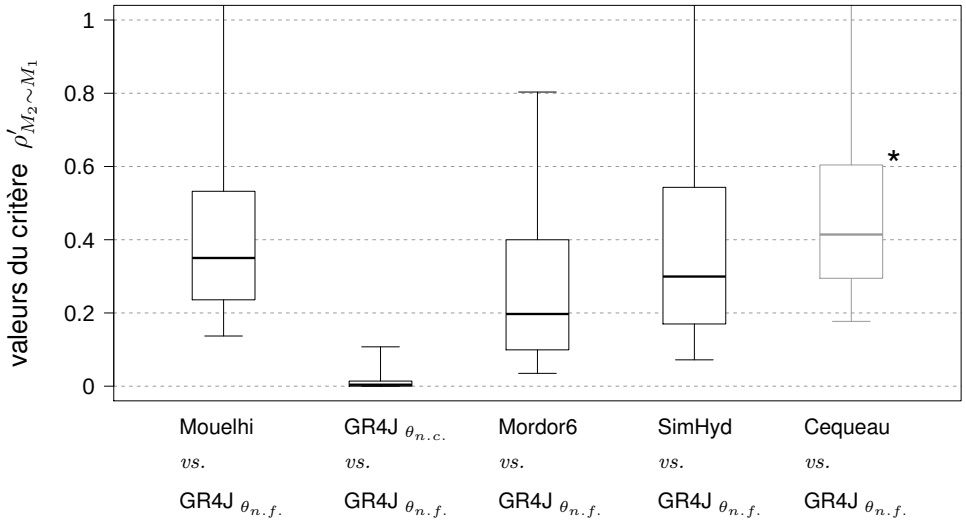


Figure 5.22.: Distributions du critère $\rho'_{M_2\sim M_1}$ mesurant la similitude des biais glissants entre entre différents modèles hydrologiques

*Tests sur tout l'échantillon (569 bassins), * ou sur les 365 bassins français seulement.*
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% des distributions.
Tous les modèles sont ici alimentés par des données SPAZM. GR4J, Mordor6 et SimHyd sont couplés à CemaNeige et utilisent une ETP d'Oudin. Mouelhi utilise aussi une ETP d'Oudin tandis que Cequeau cale une ETP de Thornthwaite et dispose d'un module neige interne.

5.7. Conclusions

Qu'avons-nous appris ?

Ce chapitre nous a permis d'en apprendre plus sur les causes du manque de robustesse des modèles mis en valeur par nos nombreux tests de transfert temporel des paramètres. Nous retiendrons particulièrement les points suivants :

• Les travaux de critique des données nous laissent penser que les problèmes de qualité des débits ne constituent pas une explication majeure des difficultés de transférabilité des paramètres des modèles mis en avant dans cette thèse. En ce qui concerne les données de forçages climatiques, bien qu'elles soient partiellement construites à partir des mêmes mesures, les différentes bases disponibles sur notre échantillon français ne conduisent pas à des séries identiques (avec des écarts parfois significatifs). Néanmoins, les modèles présentent des problèmes de robustesse similaires entre ces différentes configurations, en particulier au niveau des erreurs systématiques sur les volumes auxquelles nous portons une attention particulière. Si le rôle explicatif de ces données de forçage sur les échecs des modèles conceptuels n'a pas pu être établi, les différences constatées entre leurs sources alternatives laissent ouvertes certaines questions sur ce point.

• En essayant de comprendre les erreurs systématiques commises par nos modèles lors de certains transferts, nous sommes remontés jusqu'aux variations de paramètres optimaux entre les périodes impliquées dans ces situations d'échec. Ces analyses montrent comment les paramètres permettant l'ajustement du bilan lors du calage présentent des valeurs optimales qui varient temporellement en relation avec l'évolution des conditions climatiques. Sur ce point, précisons que les paramètres concernés ne doivent pas être perçus comme la source des problèmes mais comme des degrés de liberté à disposition du modèle pour tenter de reproduire des observations lors du calage. Le fait de figer les paramètres sensibles aux variations climatiques ne résout d'ailleurs pas les problèmes de robustesse constatés pour *GR4J* et *Mordor6*, cette sensibilité étant simplement transférée vers les autres paramètres d'ajustement du bilan.

• La procédure *GSST* et les analyses d'erreurs associées ont mis en avant l'importance des erreurs systématiques de volumes pour expliquer le manque de robustesse constaté pour les modèles que nous testons. Un calage sur une sous-période conduit les modèles à ajuster leurs paramètres pour *fermer* le bilan hydrologique mais cet ajustement n'est pas toujours transférable à d'autres périodes. Ce comportement se retrouve quelle que soit la période de calage considérée et nous comprenons désormais que les erreurs révélées par le *GSST* ne sont qu'une résultante des difficultés générales des modèles à reproduire les volumes moyens écoulés sur toutes les sous-périodes. Cela explique d'ailleurs l'équivalence des résultats obtenus lors de l'allongement de la sous-période de calage utilisée pour le *GSST*.

• La similitude des variations de biais moyen terme ($\overline{\widehat{Q}_{10\,ans}}/\overline{Q_{10\,ans}}$) entre des jeux de paramètres issus de différents calages corrobore les résultats obtenus dans le chapitre 4. En effet, la mise en évidence du parallélisme des courbes de biais glissants équivaut à montrer que des simulations absolues différentes deviennent équivalentes lorsqu'elles sont exprimées sous forme relative (*c.-à-d.* $\overline{\widehat{Q}}[\theta_A] \neq \overline{\widehat{Q}}[\theta_B]$ mais $\triangle\overline{\widehat{Q}}_{X\to Y}[\theta_A] \simeq \triangle\overline{\widehat{Q}}_{X\to Y}[\theta_B]$). Ainsi, nous aurions très bien pu calculer un degré de corrélation entre $\triangle\overline{\widehat{Q}}[\theta_{SP}]$ et $\triangle\overline{\widehat{Q}}[\theta_{PT}]$ à la place du critère ρ_i qui mesure la similarité des courbes de biais glissants sur 10 ans.

L'approche *GSST* était-elle inutile ?

La première phase de notre diagnostic consistait à rechercher des corrélations entre les conditions de transfert des paramètres et les pertes de performances (*cf.* chapitre 3). Nous avons vu dans cette seconde phase qu'une source explicative majeure de ces pertes était la difficulté des modèles à reproduire simultanément les volumes moyens écoulés sur différentes périodes. Dans ce cadre, le *signal de biais moyen terme* constitue un outil d'analyse simple mais très informatif. Un seul calage des modèles par bassin (*ex.* sur toute la série disponible) suffit à nous renseigner sur les biais commis lors des transferts de paramètres entre différentes sous-périodes.

Face à cette constatation, nos nombreux efforts pour mettre en place la procédure *GSST* et en analyser les résultats semblent presque se révéler inutiles. Notre travail avec le *GSST* ne perd toutefois pas son sens, puisque sans les nombreuses analyses présentées précédemment, nous n'aurions pas pu montrer les points suivants :

1. les erreurs de volume sur les débits moyens constituent vraisemblablement le contributeur principal des pertes de performances constatées lors des transferts de paramètres,
2. la signature de ces erreurs de bilan peut se retrouver au niveau des variations de valeurs optimales des paramètres qui assurent la fermeture du bilan hydrologique lors du calage des modèles hydrologiques,
3. les calages obtenus sur différentes sous-périodes sont égaux face aux difficultés de représentation des volumes moyens sur différentes parties de la chronique (*cf.* similarité des courbes ω traduisant les variations relatives de biais).

De plus, nous avons certes pu démontrer sur un grand nombre de cas une similitude des comportements entre différents calages pour des modèles comme *GR4J* ou *Mordor*6. La réalisation de multiples calages temporels reste néanmoins incontournable pour vérifier cette similitude avec d'autres modèles. Enfin, rappelons que cette similitude a seulement été établie vis-à-vis des erreurs moyennes de volumes écoulés et l'analyse détaillée de toutes les erreurs induites lors du transfert d'un jeu de paramètres nécessite la procédure *GSST* (*ex.* études des erreurs lors des crues ou des étiages, de la capacité à reproduire une crue de fonte, etc.) .

En résumé, la réalisation de tests de transfert demeure incontournable pour étudier les problèmes de robustesse des modèles de façon exhaustive et le *GSST* représente, dans ce cadre, un puissant outil. Pour la suite de nos travaux, nous pourrons toutefois nous en passer pour certaines situations, compte tenu des connaissances acquises ici. Nous pourrons par exemple nous contenter d'un unique calage (sur toute la série disponible) pour étudier la capacité des modèles *GR4J* et *Mordor6* à reproduire des bilans hydrologiques sur différentes périodes temporelles. La situation plus floue observée sur les autres modèles nous contraint à la vigilance pour ces derniers, où des investigations complémentaires sont nécessaires avant d'effectuer de tels raccourcis.

Quelles suites ?

À l'issue de ce chapitre, plusieurs aspects restent en suspens. Nous avons identifié les difficultés des modèles à reproduire les bilans moyen terme de façon simultanée sur différentes sous-périodes temporelles, et si des corrélations peuvent parfois être établies entre les biais des modèles et les écarts de conditions climatiques subis lors des transferts de paramètres. Néanmoins, nous n'avons pas déterminé la cause réelle de cette situation. Nous pourrions prolonger encore nos tentatives de compréhension mais préférons tenter notre chance autrement en passant à la recherche de mesures qui permettraient d'éviter les comportements problématiques identifiés. Sans pour autant espérer la mise au point de procédures améliorant la robustesse des modèles, nous comptons sur cette phase de recherche de solutions pour apporter d'autres éléments de diagnostic et dresser quelques pistes de poursuite de nos travaux.

6

Quelles pistes d'amélioration de la robustesse des modèles ?

Ces nombreuses questions auxquelles
il n'est pas toujours facile de répondre...

Calvin and Hobbes, 21/07/1987, Bill Watterson©

6.1. Introduction

Résumé des travaux précédents :

La procédure *GSST* a permis une évaluation systématique de la robustesse temporelle des modèles hydrologiques conceptuels (*cf.* chapitre 3). L'étude approfondie des cas problématiques a mis en avant les erreurs de volumes comme cause majeure des pertes de performances constatées lors des transferts entre des périodes climatiquement contrastées. Ces problèmes de robustesse semblent liés à l'incapacité des modèles testés à reproduire les variations temporelles observées des équilibres entre précipitations, évapotranspiration et écoulement. Cette incapacité conduit à la situation suivante : un calage des modèles sur une sous-période donnée permet un équilibrage du bilan hydrologique moyen sur celle-ci par un ajustement des paramètres, mais cet ajustement n'est pas transférable vers d'autres périodes présentant des caractéristiques climatiques contrastées (*cf.* chapitre 5).

Les travaux menés jusqu'ici peuvent également être résumés par l'analogie suivante : nous avons réalisé un *diagnostic* de la robustesse temporelle de modèles conceptuels, en partant des *symptômes* que constituent les pertes de performances constatées lors de certains transferts temporels. L'analyse de ces *symptômes* nous a conduits à remonter la chaîne qu'ils composent (*c.-à-d.* pertes de performances ← erreurs de volumes ← équilibrages de bilans non transférables) jusqu'à identifier la *pathologie* à leur origine (*c.-à-d.* l'incapacité des modèles testés à reproduire les variations observées des équilibres entre P, ETR et Q). La cause de cette *pathologie* n'a néanmoins pas pu être identifiée jusqu'ici, malgré les comparaisons menées entre plusieurs configurations de tests et plusieurs modèles hydrologiques.

Objectifs du chapitre :

La phase de diagnostic n'a pas abouti à l'identification d'une origine claire des difficultés rencontrées, mais nous ne sommes pas dépourvus d'idées quant aux éléments qui pourraient y conduire. À défaut de pouvoir proposer une méthodologie générale pour augmenter de la robustesse des modèles, ce chapitre présente les différentes pistes que nous avons explorées (à des degrés plus ou moins avancés).

6.2. Remplacement du *KGE* par d'autres critères de calage classiques

6.2.1. Introduction

Les problèmes de robustesse temporelle de modélisation hydrologique semblent communs aux différents modèles testés (*cf.* sections 3.3.2, 3.4.2 et 5.6.4). Tous les résultats présentés jusqu'ici ont été obtenus pour des calages de ces modèles par le critère KGE, proposé par *Gupta et al.* (2009).

Nous souhaitons déterminer ici si d'autres critères de calage classiques conduisent à une situation équivalente, c'est-à-dire si le manque de transférabilité climatique des paramètres est imputable au critère KGE et si l'utilisation d'un critère plus standard, tel que le NSE n'aurait pas engendré de telles erreurs. Nous avons donc reproduit une partie de nos tests en changeant la fonction objectif considérée au calage. Ces tests ont été menés sur un modèle hydrologique (*GR4J-CemaNeige*, alimenté par une ETP d'Oudin), tout l'échantillon de bassins ou seulement une sous-partie étant utilisés selon les cas.

Précisons enfin que cette section se focalise sur le remplacement du KGE par d'autres fonctions objectif classiques et que la recherche d'autres critères ou procédures de calage ciblant spécifiquement la robustesse des modèles fera l'objet d'une discussion séparée (*cf.* section 6.3).

6.2.2. Critères de calage considérés

Le critère KGE a été construit sur une décomposition du critère de *Nash et Sutcliffe* (1970), avec pour objectif d'en corriger certains défauts à travers un rééquilibrage des différentes composantes. Nous reprenons ici nos tests de calage-validation en remplaçant ce critère par d'autres fonctions classiques en modélisation hydrologique. Comme pour l'étude de sensibilité des résultats à la longueur des sous-périodes dans le *GSST* (*cf.* section 5.2.2), nous avons choisi de nous concentrer sur les bassins où des liens avaient été détectés entre robustesse et écarts de température lors des transferts des paramètres (*cf.* section 3.3). Sur cette sélection de 120 bassins, le nuage de points liant $\overline{\widehat{Q}}/\overline{Q}$ et $\triangle\overline{T}_{D\to R}$ est marqué par une nette tendance ($\overline{T}_D < \overline{T}_R \Leftrightarrow \overline{\widehat{Q}}_R[\theta_D] > \overline{Q}_R$ et inversement), lorsqu'un critère KGE est utilisé au calage. Une comparaison des nuages de points correspondants permet donc facilement d'apprécier l'impact du choix de la fonction objectif sur nos résultats.

Une comparaison est réalisée entre les trois fonctions objectif suivantes pour le calage de *GR4J* (alimenté par une *ETP* d'Oudin et couplé à *CemaNeige* dont les paramètres sont fixés) :

1. le critère $KGE[Q_j]$, sur les débits journaliers (critère utilisé jusqu'ici) ;
2. le critère $NSE[\sqrt{Q_j}]$, qui correspond à un NSE sur la racine des débits journaliers ;
3. le critère de $RMSE[\sqrt{Q_j}] \cdot (\,1 + |\,B\,|\,)$, qui combine un critère d'erreur quadratique sur la racine des débits journaliers et un critère de biais sur le débit moyen de la période considérée ($B = \overline{\widehat{Q_j}}\ /\ \overline{Q_j}$).

Les critères $NSE[\sqrt{Q_j}]$ et $RMSE[\sqrt{Q_j}] \cdot (1 + |\,B\,|)$ sont basés sur un calcul d'erreur quadratique et s'apparentent aux fonctions classiquement utilisées en modélisation hydrologique. Leurs formulations sont identiques à celles fournies dans les équations 2.2 et 2.3 (page 31), les débits subissant simplement une transformation racine carrée. Cette transformation réduit l'influence des très forts débits et tend ainsi à rééquilibrer le poids des différents pas de temps sur la sélection des paramètres. Les critères alternatifs, ainsi constitués, permettent l'identification de jeux de paramètres mieux adaptés pour simuler toute la gamme des débits *(Oudin et al., 2006a)*. L'introduction du terme multiplicatif $(1 + |B|)$ au niveau du second critère a pour effet de renforcer le poids des erreurs de volume dans la méthode de sélection. Comparativement à l'utilisation d'une erreur $RMSE[\sqrt{Q_j}]$ seule, cet ajout améliore fortement les erreurs de biais en calage sur notre échantillon de bassins australiens (*cf.* figure 6 dans *Coron et al.* (2012), en annexe E).

Remarque préliminaire :
Contrairement aux comparaisons précédemment réalisées, nous indiquons ici les valeurs absolues de KGE plutôt que leur expression relative sous la forme MRC_{KGE}. La formulation relative met en valeur les pertes de performances liées au transfert mais n'a de sens que lorsqu'elle concerne le critère utilisé pour l'optimisation. En effet, elle revient à *normaliser* l'erreur de modélisation par l'erreur de référence obtenue pour le modèle sur la période simulée ($MRC = \varepsilon_R[\theta_D]/\varepsilon_R[\theta_R] - 1$). Pour des calages sur des critères différents que le KGE, les valeurs de MRC_{KGE} peuvent donc prendre des valeurs négatives car θ_R ne conduit pas nécessairement au plus fort $KGE_R[\theta_i]$ atteignable. Pour nos trois fonctions objectif alternatives, nous devrions calculer les trois critères *MRCs* suivants :

- $MRC_{KGE} = (1 - KGE_R[\theta_D])\ /\ (1 - KGE_R[\theta_R])\ - 1$
- $MRC_{NSE} = (1 - NSE_R[\theta_D])\ /\ (1 - NSE_R[\theta_R])\ - 1$
- $MRC_{RMSE\cdot(1+|B|)} = (RMSE_R[\theta_D] \cdot (1 + |B_R[\theta_D]|))\ /\ (RMSE_R[\theta_R] \cdot (1 + |B_R[\theta_R]|))\ - 1$

Ces trois *MRCs* seraient difficilement comparables entre eux, ce qui explique notre utilisation des valeurs de KGE (non normalisées) qui ont l'avantage d'être adimensionnelles et comparables entre des bassins très divers (ce qui n'est pas le cas pour NSE et $RMSE \cdot (1 + |B|)$, *cf.* discussion section 2.1.3).

6.2.3. Résultats

La figure 6.1 fournit les nuages de points $\triangle\varepsilon = f(\triangle\overline{T})$ issus du *GSST* pour ces trois alternatives de calage de *GR4J* sur notre sélection de 120 bassins. Les valeurs absolues de KGE en validation ainsi que les trois composantes d'erreur couramment considérées dans nos travaux y sont représentées.

Les cloches de *KGE* obtenues sont globalement similaires, même si les calages sur les deux critères nouvellement testés conduisent à des diminutions de performances légèrement plus faibles pour les plus grands écarts de température subis en transfert (cette différence est particulièrement visible pour les calages sur $NSE[\sqrt{Q_j}]$ et les transferts tels que $\overline{T}_D < \overline{T}_R$). Les nuages concernant les biais (sur lesquels nous concentrons nos analyses) ne présentent, par contre, pas de différence significative de forme entre les trois critères de calage testés. Enfin, l'absence de tendance et le degré de dispersion des nuages obtenus sur les deux autres mesures d'erreurs (corrélation $\widehat{Q_j} \sim Q_j$ et rapport des coefficients de variations $(\sigma_{\widehat{Q_j}}/\mu_{\widehat{Q_j}})/(\sigma_{Q_j}/\mu_{Q_j})$) sont également proches à une différence près : la variabilité est sous-estimée en moyenne pour les calages sur $NSE[\sqrt{Q_j}]$ et $RMSE[\sqrt{Q_j}] \cdot (1 + |B|)$, quelles que soient les conditions de transfert. Ce comportement, montré par *Gupta et al.* (2009) lors de leur décomposition du critère NSE, est donc vérifié ici sur un grand nombre de tests.

Les fonctions objectif testées ici sont toutes trois basées sur la comparaison des débits simulés et observés au pas de temps journaliers. Selon le critère, cette comparaison peut être directe (*ex.* $RMSE$), relative à un modèle de variance nulle (*ex.* NSE), ou peut combiner plusieurs composantes (*ex.* KGE). Malgré ces différences, ces trois critères conduisent à des situations proches en termes d'erreurs commises lors des transferts temporels. Si les résultats présentés ici ne concernent qu'un seul modèle et un sous-groupe de 120 bassins, les tests menés sur notre échantillon australiens (*cf. Coron et al.*, 2012) laissent penser que ces conclusions sont généralisables à un plus grand nombre de bassins et à d'autres modèles (*Mordor6* et *SimHyd* à minima).

La figure 6.2 vient compléter ces résultats en présentant les écarts-types des courbes de biais glissants obtenues sur tout notre échantillon (569 bassins). Calculés pour chaque bassin sur la courbe ω correspondant à un calage sur toute la période, ces écarts-types $\left(\sigma\left[\omega_{PT}^{M_X}\right]\right)$ révèlent les difficultés du modèle à reproduire les bilans moyen terme de façon simultanée sur toutes les périodes. Si des valeurs légèrement plus faibles sont constatées lorsqu'un critère *NSE* est utilisé, les écarts-types restent du même ordre de grandeur pour les trois critères de calage considéré ici. Nos recherches doivent donc être poursuivies puisque la réduction de ces erreurs systématiques de volume nous paraît cruciale pour améliorer la transférabilité temporelle des paramètres des modèles (*cf.* section 5.6).

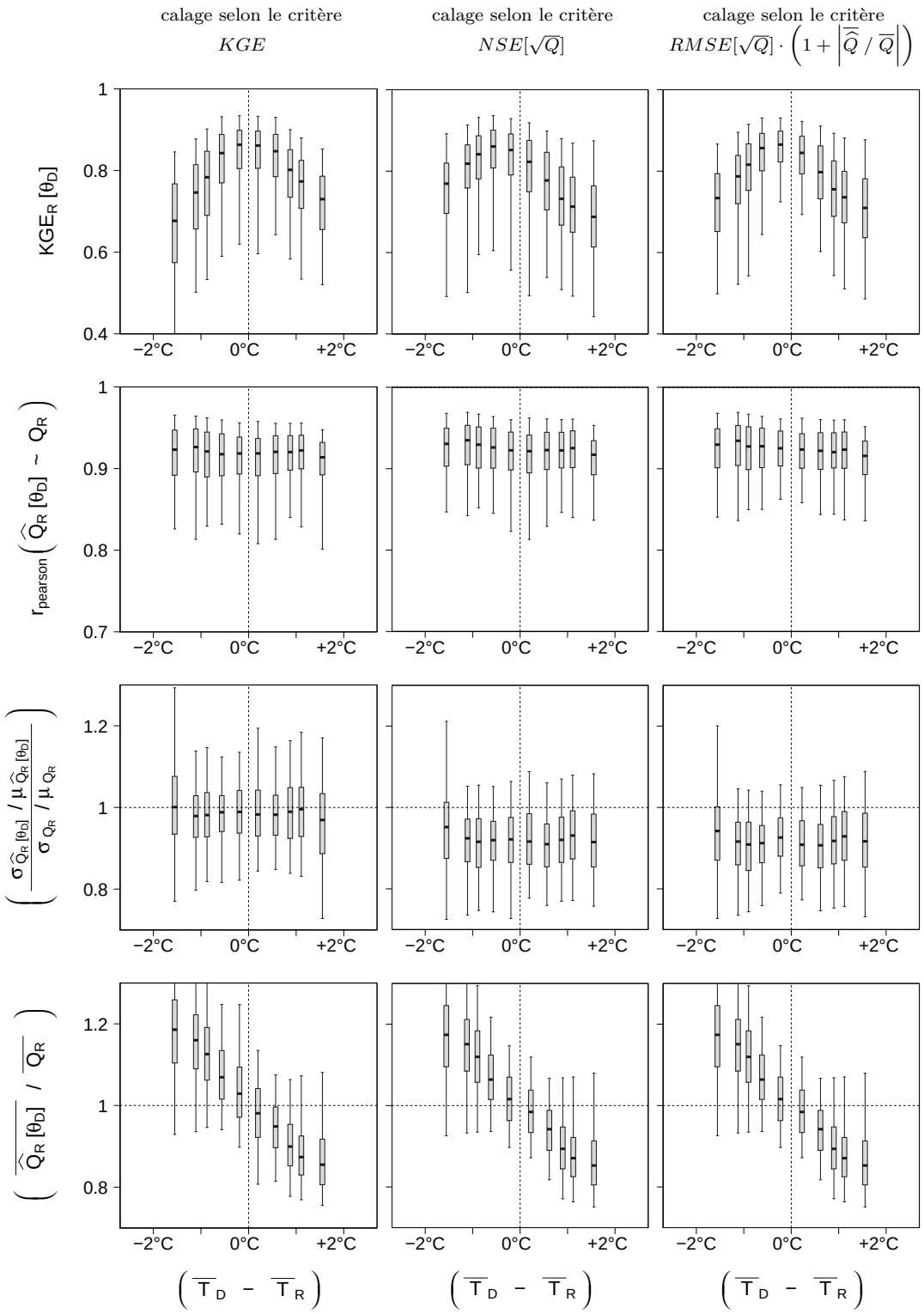


FIGURE 6.1.: Pertes de performance lors des transferts entre périodes sur les 120 bassins concernés par des corrléations significatives entre △erreurs et △température

Rq. sur les bassins français, GR4J est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés).
Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

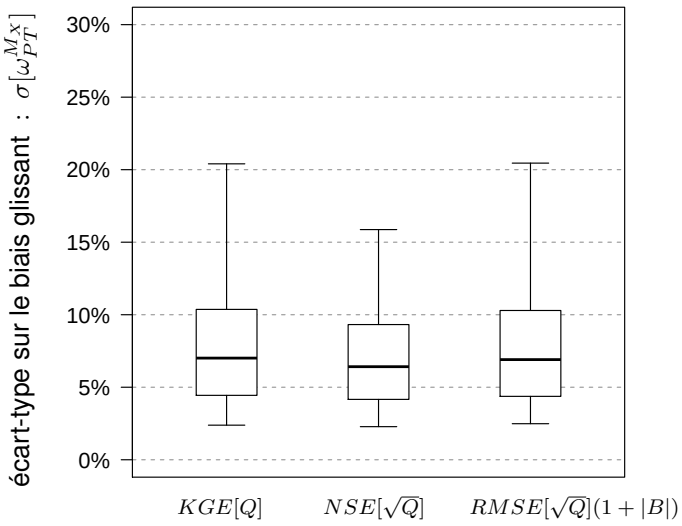


FIGURE 6.2.: Distributions de l'écart-type du biais glissant sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$, pour tout l'échantillon, selon la fonction objectif utilisée pour caler *GR4J*

Résultats sur 569 bassins pour GR4J-CemaNeige$_{\theta fixés}$, alimenté par une ETP d'Oudin. Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95%.

6.2.4. Suffirait-il de retirer la contrainte (trop forte) sur le biais ?

Les nombreuses analyses que nous avons menées montrent comment les équilibrages des volumes moyens lors de la phase de calage expliquent les erreurs systématiques de volume constatées en validation. Ces résultats furent explicitement montrés lors de notre étude des biais glissants de modélisation, à la section 5.6. Notons aussi que les différentes fonctions objectifs testées jusqu'ici (*KGE*, *NSE*, *RMSE*) accordent toutes une importance élevée à ces erreurs de volumes et contraignent fortement les modèles à minimiser le biais durant la phase de calage.

Partant de ces observations, nous nous sommes interrogés sur la possibilité d'une contrainte trop importante sur $\overline{\widehat{Q}}/\overline{Q}$ lors du calage, qui serait respectée au détriment d'une meilleure représentation des dynamiques d'écoulement. La réalité n'est évidemment pas aussi simple. Une variation de débit simulée par un modèle est généralement conditionnée par la gamme de débits associée. La capacité du modèle à reproduire les variations observées sera donc affectée par les erreurs éventuelles de volumes. Nous avons toutefois souhaité conduire un petit test, sur quelques exemples, pour étudier l'effet d'un retrait de cette contrainte de biais.

Nous avons pour cela repris le critère de *Gupta et al.* (2009), en considérant cette fois-ci la formulation KGE_S qui comporte les coefficients s_ρ, s_α et s_β, permettant la pondération des différentes composantes (*cf.* formulation ci-dessous).

$$KGE_{S\ D\to R} = 1 - \sqrt{\left[s_\rho \cdot (\rho_{kge} - 1)\right]^2 + \left[s_\alpha \cdot (\alpha_{kge} - 1)\right]^2 + \left[s_\beta \cdot (\beta_{kge} - 1)\right]^2} \qquad (6.1)$$

Pour rappel, les composantes ρ_{kge}, α_{kge} et β_{kge} correspondent respectivement au critère de corrélation de *Pearson*, au rapport des variances et au rapport des volumes entre les débits simulés et les débits observés (*cf.* équation 2.4, page 32).

Ce KGE_S nous permet de conduire très facilement notre test sur le retrait de la contrainte de biais, qui est clairement identifiée dans ce critère (par opposition aux *NSE* et *RMSE*). Il nous suffit de comparer nos résultats obtenus avec le KGE classique (*c.-à-d.* s_ρ=s_α=s_β=1) issus d'un KGE_S pour lequel s_ρ=1, s_α=1 et s_β=0.

Sans vouloir reproduire notre procédure *GSST* sur la totalité de notre échantillon, nous avons évalué la pertinence de cette piste sur quelques cas ponctuels. Les figures 6.3 et 6.4 présentent les résultats comparatifs obtenus pour deux bassins versants précédemment rencontrés dans nos exemples : *l'Allier à Vieille Brioude* et *la Grande Creuse à Fresselines*. Nous y avions détecté une corrélation moyenne entre les variations de température et les problèmes de robustesse des modèles hydrologiques. Ces figures regroupent quelques-uns des nuages de points intéressants ainsi que les biais glissants obtenus pour les différents calages (période complète en rouge et sous-périodes de 10 ans en gris).

Ces comparaisons montrent que le fait de relâcher la contrainte de biais lors du calage de *GR4J*, ne permet pas d'améliorer sa capacité à représenter simultanément les différents bilans moyen terme. Les analyses de biais glissants ($\overline{\widehat{Q}_{10\,ans}}/\overline{Q_{10\,ans}} = f(temps)$) permettent de voir particulièrement bien comment l'utilisation d'un coefficient s_β nul entraîne une translation verticale des courbes (car il n'y a plus de contrainte sur le biais en calage) sans que leurs formes moyennes ne changent. Ce que nous désignions sous le terme de *"signal du biais"* au chapitre précédent (*cf.* section 5.6) ne semble donc pas modifié par un retrait du critère de biais au calage. Sur ces deux exemples, il est d'ailleurs intéressant de noter qu'il n'y a ni une amélioration ni une détérioration de ce *signal*. Ajoutons enfin que les quelques tests que nous avons menés sur d'autres bassins versants conduisent à des conclusions identiques.

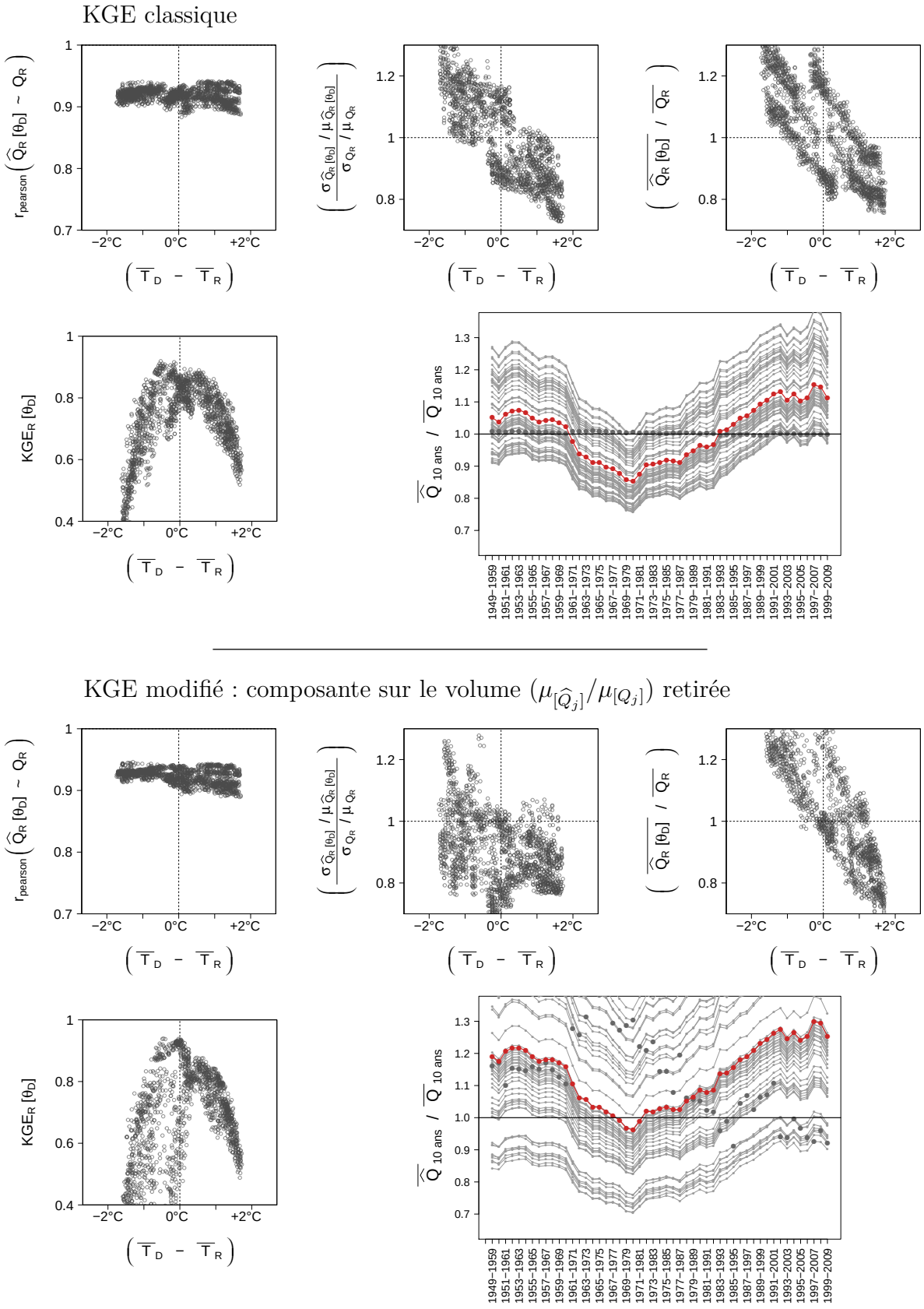


FIGURE 6.3.: Impact du retrait de la composante biais du *KGE* sur la robustesse de *GR4J* pour le bassin de *l'Allier à Vieille Brioude*

Résultats du GSST sur 10 ans pour GR4J-CemaNeige$\theta_{fixés}$ (données SPAZM et ETP d'Oudin). Courbes liées aux calages sur les sous-périodes de 10 ans en gris et sur la période complète en rouge (chaque courbe grise comporte un point de plus forte taille indiquant la sous période de calage).

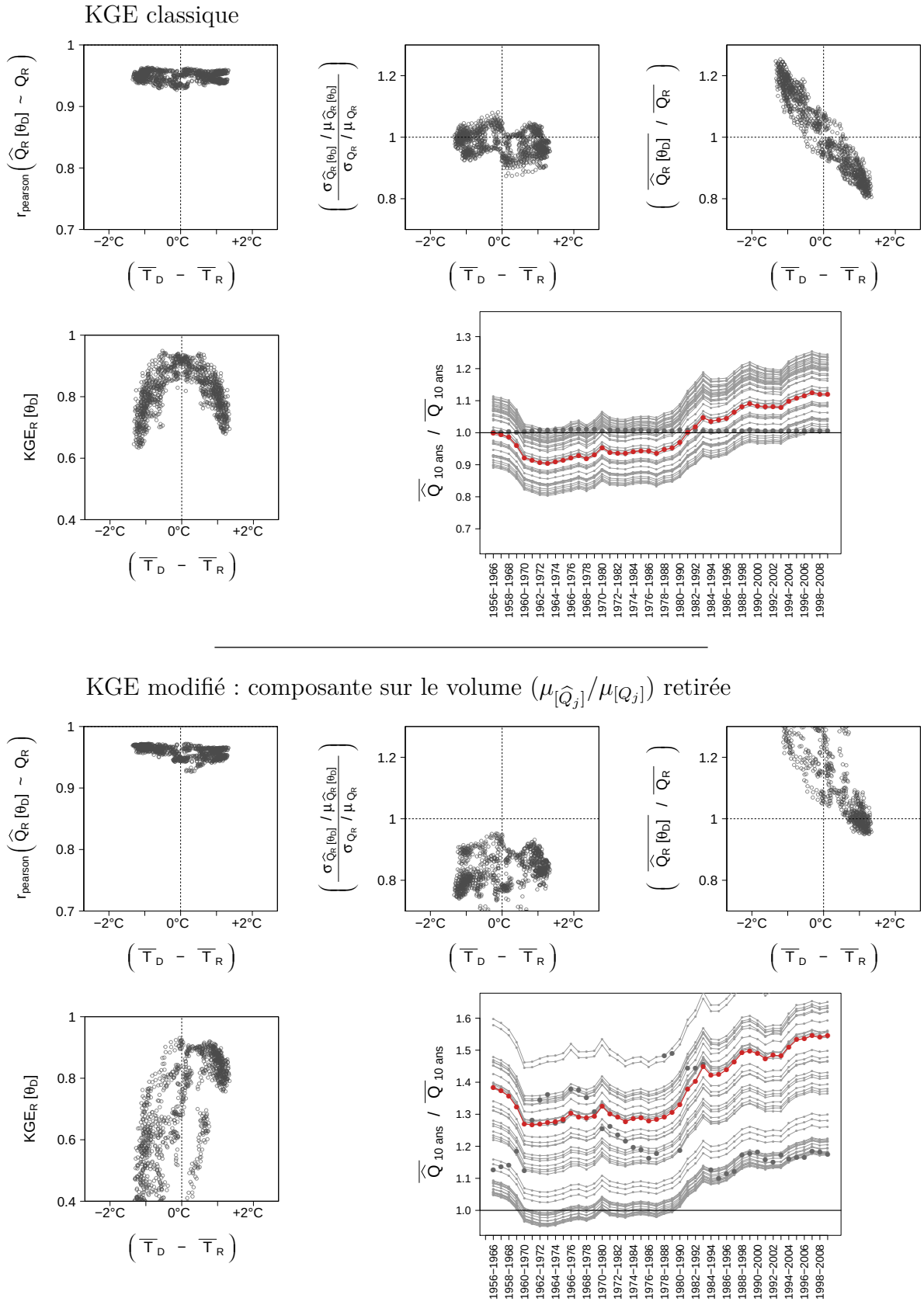


FIGURE 6.4.: Impact du retrait de la composante biais du *KGE* sur la robustesse de *GR4J* pour le bassin de *la Grande Creuse à Fresselines*

Résultats du GSST sur 10 ans pour GR4J-CemaNeige$\theta_{fixés}$ (données SPAZM et ETP d'Oudin). Courbes liées aux calages sur les sous-périodes de 10 ans en gris et sur la période complète en rouge (chaque courbe grise comporte un point de plus forte taille indiquant la sous période de calage).

6.2.5. Discussion

L'utilisation des critères *NSE* et *RMSE* plutôt que *KGE* ne semble pas permettre une amélioration du manque de robustesse de *GR4J* entre des périodes climatiquement contrastées. Le retrait de la contrainte sur le biais au calage, testé sur quelques cas, ne semble pas non plus réduire les difficultés du modèle à représenter les variations des bilans moyen terme. Bien que ces différents tests ne concernent que *GR4J*, nos résultats précédents nous conduisent fortement à penser que des conclusions équivalentes seraient obtenues avec les autres modèles de notre panel.

La similitude des résultats entre ces différents critères de calage soulève une interrogation : *Les modèles pourraient-ils subir un antagonisme entre une représentation convenable des variations de débits au pas de temps journaliers (c.-à-d. signal haute fréquence) et une représentation convenable des variations des équilibres du bilan en eau sur des échelles de temps beaucoup plus grandes (c.-à-d. signal basse fréquence) ?* Une fois n'est pas coutume, l'échec d'une piste de solution à nos problèmes de robustesse nous conduit à imaginer de nouveaux tests.

6.3. Méthodes de sélection des paramètres ciblant la robustesse temporelle et climatique

6.3.1. Introduction

Le remplacement du critère *KGE* par d'autres critères classiques pour caler nos modèles hydrologiques ne semble pas permettre la diminution des erreurs commises lors des transferts des jeux de paramètres, notamment au niveau des erreurs de volumes.

L'idée selon laquelle nous pourrions trouver une façon de caler nos modèles qui résoudrait (ou au moins réduirait) leur manque de robustesse temporelle et climatique était pourtant séduisante. Ne souhaitant pas abandonner si tôt cette piste, reprenons nos réflexions en y intégrant les constatations que nous avons pu faire au fil de nos nombreux tests. Deux options émergent quant à la direction à prendre pour résoudre notre problématique via une meilleure procédure de sélection des paramètres :

- *Option 1*
  Se focaliser sur la robustesse climatique des jeux, *c.-à-d.* rechercher une procédure permettant la sélection de jeux de paramètres capables de représenter convenablement toute la gamme de conditions climatiques et aplatir ainsi les nuages de *MRC* étudiés dans le chapitre 3.

. *Option 2*
Se focaliser sur la réduction des biais moyen terme, *c.-à-d.* rechercher une procédure permettant la sélection de jeux de paramètres capables de reproduire les bilans moyen terme observés et aplatir ainsi les courbes étudiées à la section 5.6.

6.3.2. Concept d'un calage robuste en termes d'extrapolation climatique

Lorsque les travaux de cette thèse furent initiés, la phase de diagnostic de la robustesse des modèles hydrologiques face aux changements de climat ne nous semblait qu'une première étape. La seconde partie de nos travaux devait être la recherche de voies d'amélioration de cette robustesse (notamment par une meilleure méthode de sélection des jeux de paramètres). À mesure que nous progressions dans nos premiers diagnostics (présentés au chapitre 3), nous gardions à l'esprit le concept d'une procédure de calage qui serait robuste vis-à-vis des variations de conditions climatiques. Sur ce thème nous pouvons mentionner les travaux de *Bárdossy et Singh* (2008) et *Gharari et al.* (2013) qui convoitent également des procédures d'identification de paramètres robustes. Dans notre cas, il s'agissait de trouver la méthode de sélection des jeux de paramètres qui maximiserait la capacité d'extrapolation climatique des modèles hydrologiques, autrement dit, orienterait cette sélection vers des jeux capables de représenter convenablement la relation pluie-débit sur toute la gamme de conditions climatiques.

Ces idées furent influencées par nos méthodes d'analyse via les nuages de points reliant l'augmentation des erreurs de simulation avec l'écart de climat subi lors du transfert temporel des paramètres. Nous pensons particulièrement aux graphiques $MRC = f(\triangle climat)$ qui se caractérisent par des formes en $\cup$ ou $\vee$ lorsqu'une corrélation entre pertes de performances et conditions de transfert est détectée (*cf.* section 2.1.3). L'objectif du *calage climatiquement robuste* serait d'aplatir ces formes, *c.-à-d.* avoir une faible augmentation des erreurs lorsque les périodes de calage et de validation deviennent contrastées. Il est cependant possible que cette meilleure extrapolation climatique des paramètres se fasse au détriment des performances lors de transferts sous climats similaires. Cet objectif et l'éventuel compromis associé sont schématisés sur la figure 6.5.

Nous avons ainsi imaginé et testé une procédure de sélection des paramètres dont l'objectif était d'identifier les jeux permettant un niveau de performances équivalent sur toute la gamme de conditions climatiques. Dans cette procédure, détaillée dans l'annexe H.1, chaque jeu de paramètres est évalué selon deux critères de qualité : la variabilité (σ) et la moyenne (μ) des performances associées sur une gamme climatique élargie. Ces critères σ et μ s'apparentent à des évaluations du degré d'aplatissement et du positionnement vertical de la courbe représentative du jeu dans un graphique tel que celui de la figure 6.5.

Cette procédure présente cependant plusieurs limites importantes qui rendent son utilisation difficile en pratique. L'une des limites majeures est liée à notre incapacité à prédire si le

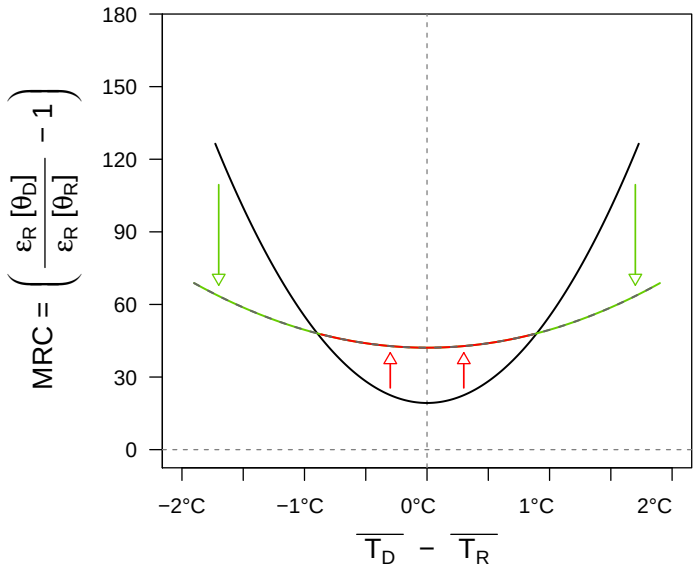


FIGURE 6.5.: Schématisation du principe d'un calage climatiquement robuste pour des bassins où les erreurs en transfert sont corrélées avec les écarts de températures

Transferts des paramètres entre des périodes donneuses (D) et receveuses (R) flèches descendantes vertes : objectif de diminution des erreurs lors des transferts entre périodes constrastées ; flèches ascendantes rouges : effet collatéral éventuel d'augmentation des erreurs lors des transferts entre périodes similaires

manque de robustesse d'un modèle sur le bassin étudié sera lié à des écarts de température, de précipitations, ou d'autres phénomènes non détectés. Cette incapacité fut mise en avant par la forte hétérogénéité obtenue lors de la recherche de corrélation entre erreurs de modélisation et changements de conditions climatiques (*cf.* section 3.2). Or, sans cette capacité à prédire quelles conditions de transfert sont problématiques, le calage *climatiquement robuste* devient difficile puisque, par construction, il est mené pour une variable climatique particulière. Plus généralement, cette procédure de *calage* ne paraît plus aussi pertinente que lors de sa conception, compte tenu des comportements interprétés depuis. Nous pensons en particulier ici aux analyses des courbes de biais glissants $\overline{\widehat{Q}}/\overline{Q} = f(t)$) qui expliquèrent (en partie) l'origine des erreurs systématiques de volume constatées lors de certains transferts de paramètres. Ces différents points expliquent le placement en annexe de la présentation détaillée de cette procédure ainsi que des résultats illustratifs associés et des éléments de discussion la concernant (annexe H.1, pages 332-341).

6.3.3. Une procédure ciblant la réduction des biais moyen terme

Les analyses des biais glissants ont révélé les difficultés des modèles à reproduire les bilans hydrologiques moyens de manière simultanée sur différentes sous-périodes temporelles. Il fut très frappant de constater les fortes similarités des comportements entre les jeux de paramètres issus de calages sur diverses sous-périodes ou entre des modèles de structures pourtant différentes (*cf.* section 5.6). Plus étonnant encore, ces biais de simulation sur les bilans moyen terme restent quasiment identiques pour des calages sur la totalité de la série disponible.

Ces résultats nous interrogent sur la capacité des modèles à reproduire des dynamiques basse fréquence. Cette incapacité induit des ajustements de bilan lors des calages sur des sous-périodes qui ne sont pas transférables temporellement et conduisent aux erreurs observées lors du *GSST*. Ainsi, nous pensons avoir trouvé là une source importante des situations de manque de robustesse des modèles lors de transferts temporels et climatiques de leurs paramètres. Concentrer nos efforts sur la réduction de ces biais moyen terme nous semble donc une piste pertinente[1], dans notre quête d'une méthode de sélection des paramètres qui résoudrait (ou au moins réduirait) les problèmes de robustesse des modèles.

L'une des caractéristiques notables de cette piste est qu'elle peut ne cibler que la réduction des erreurs de volume et être ainsi menée indépendamment de toute considération des variations climatiques observées sur la chronique. En effet, son exploration ne nécessite pas la recherche préalable d'un lien entre robustesse et évolutions des conditions climatiques (qui n'aboutit d'ailleurs pas toujours). Cela constitue une différence notable avec des approches telles que le *calage climatiquement robuste* et représente un grand avantage pour nos travaux compte tenu de notre difficulté à identifier clairement un indicateur climatique explicatif des variations temporelles des erreurs pour de nombreux bassins (*cf.* section 3.2). En effet, nous pouvons à nouveau bénéficier pleinement de tout l'échantillon de bassins, *c.-à-d.* pouvoir intégrer toute la diversité des situations dans nos analyses sans devoir nous cantonner à des sous-groupes de bassins ayant des comportements similaires (tel que celui des 120 bassins où nous constations un lien entre MRC et $\triangle\overline{T}$).

Mise en place des tests

Nos travaux ont mis en avant des erreurs de volume moyen terme parfois importantes, même lorsque l'intégralité de la chronique est utilisée pour le calage. Cette découverte troublante nous conduit à la question suivante : *Voyons-nous ici le signe d'une réelle incapacité des modèles testés à reproduire les dynamiques lentes ? Ou est-il possible que ces modèles puissent reproduire celles-ci mais que les exigences qui leur sont portées par ailleurs (pour simuler les*

1. Cette idée nous semble en tout cas plus pertinente que celle du calage robuste que nous venons d'évoquer.

dynamiques journalières) les en empêchent ? En d'autres termes, *les contraintes de reproduction des variations hautes fréquences, imposées via une fonction objectif telle que les KGE ou NSE, sont-elles antagonistes avec une reproduction des variations d'équilibre moyen terme du bilan des bassins ?*

Pour répondre à cette question, nous considérons un critère de calage ciblant spécifiquement et uniquement la réduction de variance sur les erreurs de volume moyen terme. Ce critère est un écart-type des biais glissants sur 10 ans, dont la formulation est précisée ci-dessous et qui correspond au $\sigma\left[\omega_{PT}^{M_X}\right]$ rencontré dans nos analyses précédentes (*cf.* section 5.6.4). Toute la chronique disponible est utilisée, l'objectif de calage recherché étant schématisé sur la figure 6.6.

$$RMSE\left[\,biais_{\,10\,ans}\,\right] = \sigma\left[\omega_{PT}^{M_X}\right] = \sqrt{\frac{1}{(n-10)} \cdot \sum_{k=1}^{n-10} \left(\sum_{a=k}^{k+10} \widehat{Q}_a \;/\; \sum_{a=k}^{k+10} Q_a \;-\; 1 \right)^2} \qquad (6.2)$$

où $\widehat{Q}_a$ et Q_a sont respectivement les débits moyens annuels simulés et observés, n étant le nombre total d'années des séries disponibles.

Ce critère de calage retire toute contrainte sur la reproduction des débits au pas de temps journalier. Ce choix est particulièrement brutal et pose un certain nombre de questions quant à la qualité des simulations journalières issues de calages sur de tels critères. Précisons néanmoins que nous ne sommes plus réellement ici à la recherche du critère améliorant la robustesse du modèle, mais que nous souhaitons en premier lieu tester si notre modèle est capable, dans l'absolu, de reproduire les variations basse fréquence présentes dans les observations.

Résultats

Nous avons (naïvement) entrepris de tester ce critère avec *GR4J* sur tout notre échantillon de bassins. Cependant, la comparaison des valeurs de $RMSE\left[\,biais_{\,10\,ans}\,\right]$ obtenus en calage avec celles issues de jeux optimisés selon les critères *KGE* ou *NSE* nous a rapidement permis de constater que les optimums globaux n'avaient pas été atteints dans un grand nombre de situations. Le remplacement de l'algorithme de recherche locale (classique pour *GR4J*) par l'algorithme utilisé pour *Cequeau* n'a pas conduit à des résultats plus concluants, malgré la présence de composantes aléatoires lors de la recherche. Nous nous sommes donc rabattus sur un calage par tirage Monte-Carlo d'un million de jeux de paramètres qui ne nous permet cependant pas de généraliser les tests de notre critère à de nombreux bassins. Les résultats obtenus sur deux bassins versants sont illustrés à la figure 6.7, où nous représentons le compromis entre une réduction de l'erreur quadratique sur les débits journaliers et une réduction

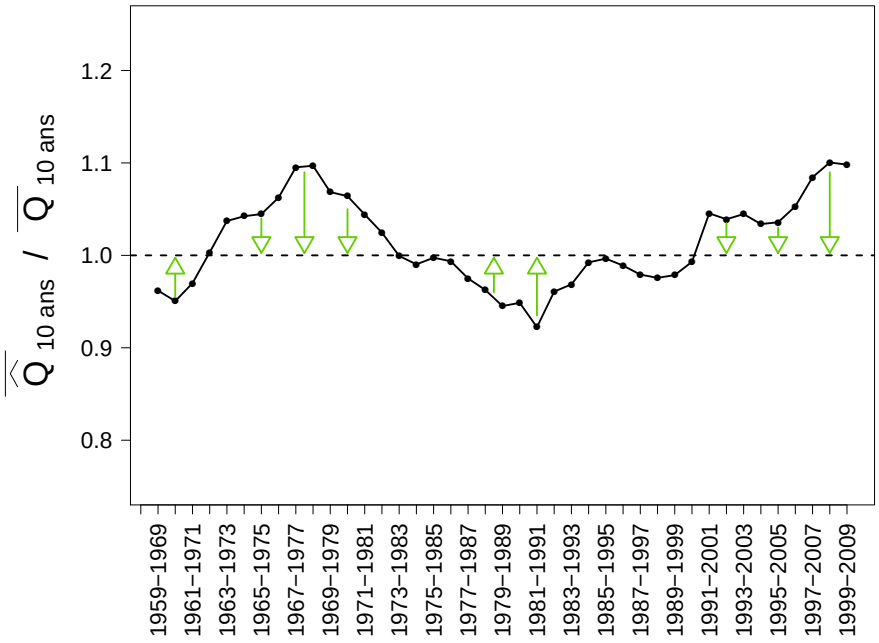


Figure 6.6.: Schématisation du principe d'un calage visant à réduire les biais moyen terme

Courbe noire : biais glissant sur 10 ans obtenu lors d'un calage sur toute la période avec un critère classique (ex. KGE) ; flèches vertes : effet recherché (réduction des biais moyen terme).

de l'écart-type des biais glissants sur 10 ans pour chacun des points testés ainsi que les courbes de biais glissants obtenues pour deux points notables.

Les résultats sont clairs : le critère proposé ici ne permet qu'une réduction marginale de l'écart-type des biais glissants, bien qu'il cible précisément cette réduction. Les nombreux jeux de paramètres testés ne conduisent d'ailleurs qu'à de faibles variations des écarts-types. Ces exemples ne sont pas atypiques, puisque le degré de variation constaté est conforme aux niveaux obtenus en moyenne sur notre échantillon (entre 5 et 10%, *cf.* figure 6.2). Nous noterons enfin que le calage sur l'erreur quadratique des débits journaliers assure une faible erreur de volume en moyenne (*c.-à-d.* $\overline{\widehat{Q}}/\overline{Q} \simeq 1$), qui n'est plus garanti par le critère $RMSE\,[\,biais_{\,10\,ans}\,]$, où seules les variations de ces erreurs sont ciblées mais pas leur moyenne.

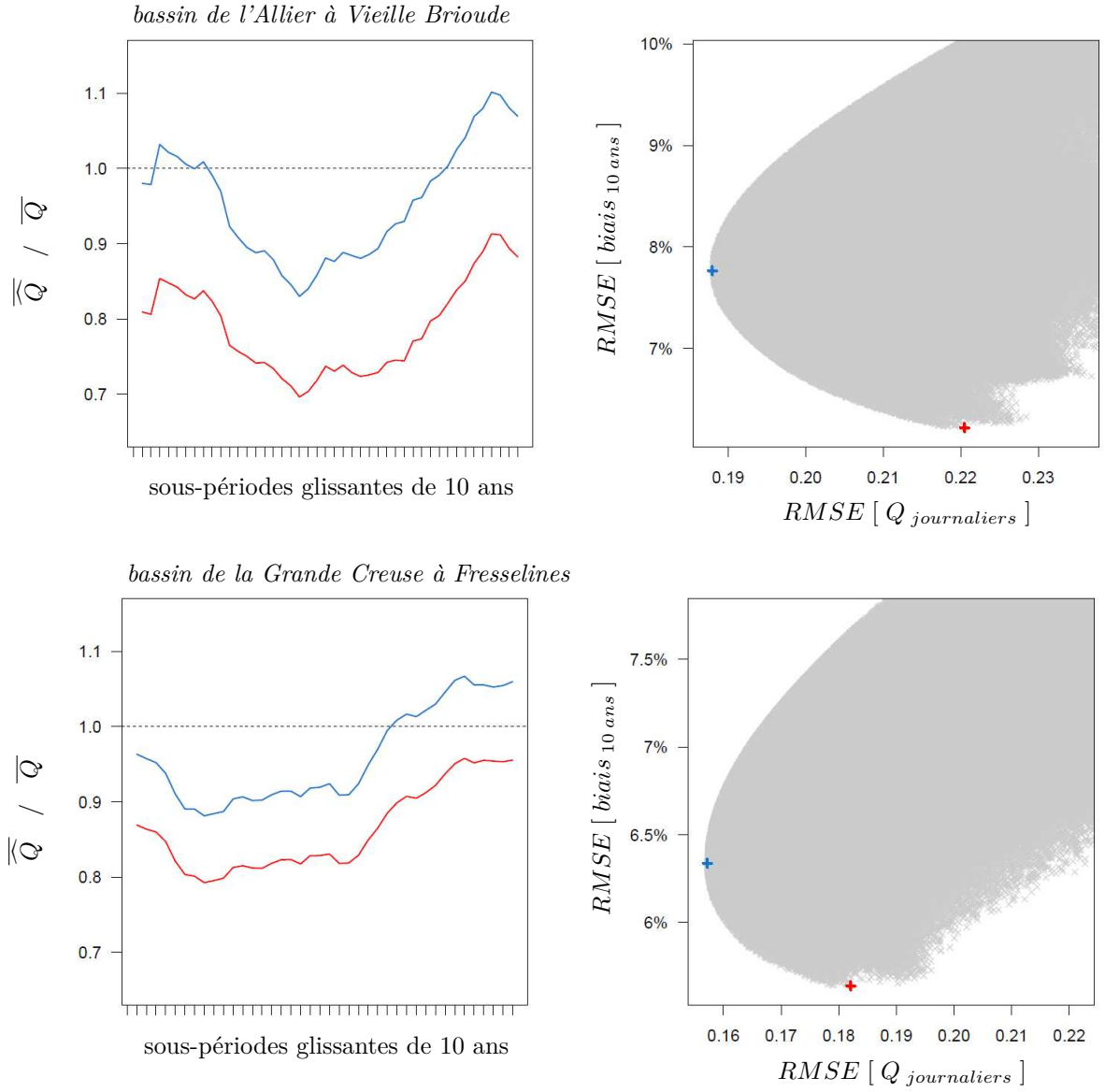


FIGURE 6.7.: Illustration du compromis entre un calage classique ($RMSE[Q]$) et un calage sur l'écart-type des biais moyen terme ($RMSE[biais_{10\ ans}]$)

Résultats obtenus par tirage aléatoire d'un million de paramètres pour GR4J-CemaNeige$_{\theta fixés}$, alimenté par des données SPAZM et une ETP d'Oudin.

6.3.4. Conclusions

Diverses tentatives ont été menées dans cette section ainsi que dans la précédente (section 6.2) pour déterminer si le critère de calage constituait une cause (et donc aussi une solution) des difficultés de reproduction des bilans moyen terme par nos modèles. Le remplacement du *KGE* par des critères tels que le *NSE* n'a pas permis une amélioration de la robustesse des modèles *GR4J, Mordor6*, *SimHyd*. Ces critères alternatifs restent toutefois très apparentés au *KGE* et nous prenons ici le contre-pied en proposant d'autres méthodes de sélection des paramètres, conçues dans le but de cibler directement les problèmes de transférabilité mis en avant par le *GSST*.

Menées sur le modèle *GR4J* (alimenté par une *ETP* d'Oudin et couplé à *CemaNeige*), ces tentatives n'ont pas permis de progresser vers une amélioration de la robustesse des modèles. De nombreuses questions restent cependant ouvertes quant à l'existence de structures qui ne seraient pas sujettes à ces problèmes de robustesse lorsque la *bonne* procédure de sélection des paramètres leur est appliquée. Cette piste n'est donc aucunement fermée par les résultats présentés ici, mais s'élargit aux réflexions sur le test de modèles plus complexes et d'éventuelles modifications de *GR4J* qui permettraient un gain de robustesse temporelle et climatique. La capacité des modèles à reproduire les bilans moyen terme semble déterminante pour expliquer les problèmes de transférabilité des paramètres et les efforts futurs devront continuer d'être concentrés sur ce point, notamment dans le choix de la procédure de calage des modèles.

6.4. Quelles pistes du côté de l'évapotranspiration ?

6.4.1. Introduction

Les cas les plus notables de mauvaise transférabilité rencontrés dans nos travaux semblent majoritairement liés à l'incapacité du modèle à reproduire les équilibres entre précipitations, évapotranspiration et écoulements sur différentes plages temporelles simultanément. Malgré notre mauvaise compréhension de l'origine de ces problèmes, nous explorons ici la piste d'une estimation incorrecte de l'évapotranspiration réelle (ETR). Cette piste semble en effet vraisemblable compte tenu des corrélations constatées entre les problèmes de transfert des paramètres et les variations de température ou de précipitations entre les périodes impliquées dans ces transferts.

Les erreurs d'estimation de l'ETR (si elles sont en cause) peuvent avoir deux origines :

1. Erreurs dans l'estimation de l'évapotranspiration potentielle (qui impactera ensuite l'estimation de l'évapotranspiration réelle).

2. Erreurs dans l'estimation de l'évapotranspiration réelle, à partir de la connaissance de l'évapotranspiration potentielle et de la disponibilité en eau (estimée via le niveau de remplissage de réservoirs dans les modèles conceptuels).

Nous avons ainsi imaginé deux pistes, visant à rechercher l'explication des problèmes de robustesse du côté de l'évapotranspiration. Ces pistes sont schématisées sur la figure 6.8.

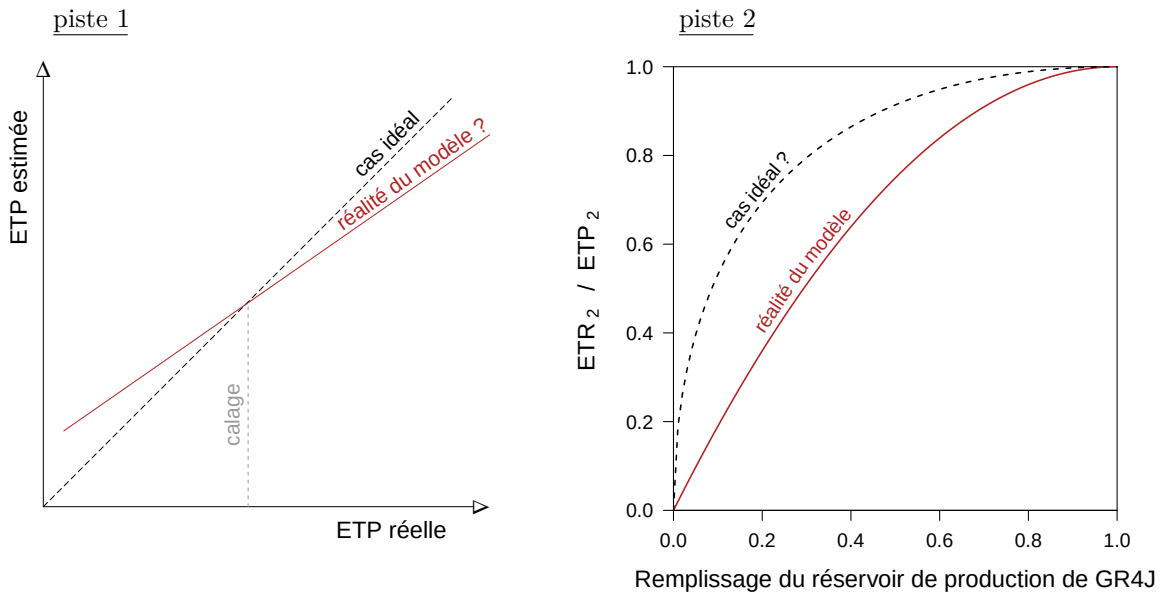


FIGURE 6.8.: Réflexions sur les causes possibles d'erreur d'estimation de l'évapotranspiration

Discussion autour de la piste 1

Les erreurs d'estimation de l'ETP constituent notre première piste. Pour bien comprendre la façon dont ce type d'erreurs impacterait la robustesse des modèles, simplifions le raisonnement au maximum. Considérons la situation théorique de données de forçages (P et T) exemptes de toute erreur, d'un bassin versant où les échanges souterrains sont connus et d'un modèle hydrologique reproduisant strictement les débits observés et n'étant sujet à aucun problème de robustesse lorsqu'il est alimenté par des données de forçages fidèles à la réalité. La formulation d'ETP utilisée permettrait alors le calcul des volumes d'eau évaporés tout au long de la série de données. Si cette formule est calée avec le modèle hydrologique, alors les paramètres d'ETP seront ajustés pour équilibrer le bilan du bassin et assurer l'égalité $\widehat{ETR} = ETR$. Une formulation incorrecte entraînerait une courbe $\widehat{ETP} = f(ETP)$ différente de la bissectrice à 45° (*cf.* figure 6.8). Cette courbe serait en dessous de la bissectrice en cas d'hyposensibilité et au-dessus de la bissectrice en cas d'hypersensibilité.

Si cette erreur était bien à l'origine de nos problèmes de robustesse, les erreurs de volume constatées sur un certain nombre de bassins ($\overline{T}_D > \overline{T}_R \Leftrightarrow \overline{\widehat{Q}}_R < \overline{Q}_R$ et inversement, *cf.* section 3.3) laissent penser à une hyposensibilité de l'estimation d'ETP. En effet, une hyposensibilité conduirait à une sous-estimation de l'ETP (et donc de l'ETR) lors de transferts des périodes froides vers les périodes chaudes et expliquerait la surestimation des débits constatée dans de tels cas. Rappelons qu'à ce stade, ces réflexions restent théoriques et que le graphique de gauche de la figure 6.8 est un schéma illustratif, rien ne nous laissant d'ailleurs penser que la courbe rouge soit effectivement une droite.

Discussion autour de la piste 2

La seconde piste concerne la façon de calculer une évapotranspiration réelle (ETR) à partir d'un niveau potentiel (ETP) supposé correct. Cette ETR dépend à la fois de l'ETP et de la disponibilité en eau pour répondre à cette demande évaporatoire. Conceptuellement, l'ETR est déterminée au sein du modèle hydrologique par une (ou plusieurs) relation(s) entre l'ETP et les niveaux des réservoirs évaporants. Des relations inadaptées pourraient par exemple induire une diminution trop importante de l'ETR lorsque les taux de remplissage des réservoirs évaporants diminuent.

Prenons pour exemple le modèle *GR4J*, pour lequel l'évapotranspiration réelle est calculée de la façon suivante : Si $P>ETP$ alors nous avons directement $ETR=ETP$. Si $ETP>P$ alors l'ETR est la somme de deux composantes : $ETR_1=P$ et ETR_2 qui dépend de l'ETP restante ($ETP_2=ETP-P$) et du niveau du réservoir de production par la relation illustrée en rouge sur la figure 6.8. Dans une situation où cette courbe rouge ne serait pas idéale, nous pourrions avoir une ETR_2 s'éloignant trop lentement ou trop rapidement de l'ETP_2 lorsque l'*humidité du bassin* varie.

En supposant que ce type d'erreurs est bien à l'origine de nos problèmes de robustesse, certaines situations constatées (*ex.* $\overline{T}_D < \overline{T}_R \Leftrightarrow \overline{\widehat{Q}}_R > \overline{Q}_R$ sur 120 bassins ou $\overline{P}_D > \overline{P}_R \Leftrightarrow \overline{\widehat{Q}}_R > \overline{Q}_R$ sur 32 bassins, *cf.* sections 3.3 et 3.4) laissent penser à une ETR_2 s'éloignant trop rapidement du niveau potentiel (*cf.* courbe pointillée sur la figure 6.8). En effet, un réchauffement ou un assèchement subi lors du transfert des paramètres s'accompagne d'une diminution moyenne du taux de remplissage du réservoir de production de *GR4J*, entrainant elle-même une baisse de l'*ETR*, qui si elle est trop importante, engendre une surestimation des débits. Rappelons encore que nous ne formulons ici que des hypothèses, imaginées pour expliquer les biais parfois très importants obtenus lors des transferts temporels de paramètres de nos modèles hydrologiques. Ces pistes vont toutefois dans le même sens que les résultats de *Lobligeois* (2010), qui a initié une étude de sensibilité de plusieurs modèles d'évapotranspiration au changement climatique.

L'évaluation de ces deux pistes, comme source d'explication des problèmes de robustesse constatés, repose sur les analyses suivantes :

1. La comparaison des comportements constatés lors des transferts pour plusieurs formules d'*ETP* (approche multi-modèles telle que celle menée sur la modélisation hydrologique).
2. Un test spécifique sur *GR4J* dans lequel nous tentons de modifier la formulation du calcul de l'*ETR* en fonction du taux de remplissage du réservoir de production.

6.4.2. Différentes formulations pour calculer l'*ETP*

Notre première piste consiste à considérer les erreurs d'estimation de l'évapotranspiration potentielle (*ETP*) comme source d'explication des problèmes de robustesse temporelle constatés via la procédure *GSST*. Plusieurs éléments de réponse à cette piste ont déjà été trouvés au fil des tests présentés dans les chapitres précédents. Nous n'allons donc pas présenter ici de nouveaux tests mais nous contenter de rappeler les différentes formules de calcul d'*ETP* utilisées et d'en commenter les résultats.

Au total, trois formules d'*ETP* ont été rencontrées dans nos tests :

1. La formule d'*Oudin et al.* (2005) est de loin la plus utilisée dans nos travaux puisqu'elle nous a servi pour estimer l'*ETP* alimentant les modèles *GR4J*, *Mordor6* et *SimHyd* dans la majorité des tests de transfert de paramètres que nous avons menés.
2. La formule de *Penman-Monteith* (*Monteith*, 1965) fut croisée lors de notre étude de sensibilité des résultats au choix de la base de forçages (*cf.* section 5.3.2). Ces séries d'*ETP* de *Penman-Monteith* ont été mises à notre disposition avec la base *SAFRAN*.
3. La formule de *Thornthwaite* (1948) a été croisée plusieurs fois, sans toujours être explicitement nommée, puisque c'est la formule présente au sein du modèle *Cequeau*. Elle comporte deux paramètres qui sont calés, dans nos tests, en même temps que ceux du modèle hydrologique (*Cequeau*).

Pour rappel, des comparaisons de résultats faisant intervenir différentes formules furent présentées dans le chapitre précédent (figure 5.5, page 147), lors des analyses de la sensibilité des résultats aux choix de la base de forçages. Nous y rencontrions la formule d'*Oudin* utilisée avec températures issues des bases *SPAZM* ou *SAFRAN* ainsi que la formule de *Penman-Monteith* à notre disposition avec les données *SAFRAN*. Par ailleurs, quelques parallèles furent établis ponctuellement entre des résultats faisant intervenir les formules d'*Oudin* (associée à *GR4J*, *Mordor6* et *SimHyd*) et de *Thornthwaite* (associée au modèle *Cequeau*), *cf.* exemples illustratifs des liens entre $\triangle erreurs$ et $\triangle \overline{T}_{D \rightarrow R}$ sur *la Grande Creuse à Fresselines*, (pages 3.6-3.7) ou comparaison des biais glissants sur *l'Ubaye à Barcelonnette* (page 175). D'une manière générale, ces différents tests ne nous ont pas permis d'identifier distinctement une amélioration de la robustesse temporelle des modèles avec le changement de formule d'*ETP*. Par exemple, les formulations d'*Oudin* et de *Penman-Monteith* diffèrent par leur complexité. Elles conduisent à des valeurs relativement différentes (*cf.* figure 5.4), sans pour autant que des niveaux différents de robustesse moyenne ressortent sur les bassins testés.

Nous devons toutefois pointer les limites des comparaisons menées. En effet, les effets d'un changement de la seule formulation servant à calculer l'*ETP* n'ont pas été étudiés, d'autres éléments variant entre les configurations testées. Par exemple, les comparaisons sur 120 bassins présentées à la section 5.3.2 concernent un *GR4J-CemaNeige,* alimenté par des données P et T issues de *SPAZM* et une *ETP* d'*Oudin,* et un *GR4J-CemaNeige,* alimenté par des données P et T issues de *SPAZM* et l'*ETP* de *Penman-Monteith* à notre disposition. Or, cette *ETP* de *Penman-Monteith* a été calculée à partir de données issues de la base *SAFRAN*. Les données de températures utilisées pour calculer les *ETP* de bassin sont donc différentes. Dans cette section, nous comparions également les performances obtenues entre ce *GR4J-CemaNeige,* alimenté par des données P et T issues de *SPAZM* et l'*ETP* de *Penman-Monteith,* et un *GR4J-CemaNeige,* alimenté par des données P et T issues de *SAFRAN* et l'*ETP* d'*Oudin.* Si dans ce cas les données de températures utilisées par les deux formules d'*ETP* sont bien identiques, la comparaison fait intervenir un changement sur les précipitations et températures alimentant *CemaNeige* (puis *GR4J*). À l'époque de ces tests, nous souhaitions déterminer l'impact d'un changement de données de forçages, en considérant notamment mais pas exclusivement l'usage de différentes formulations pour le calcul de l'*ETP*. Afin de traiter plus exhaustivement cette question de l'*ETP*, nous pourrions compléter ces comparaisons par un test de *GR4J-CemaNeige,* alimenté par des données P et T issues de *SAFRAN* et l'*ETP* de *Penman-Monteith.* Enfin, les tests faisant intervenir l'*ETP* de *Thornthwaite* ne concernent que le modèle *Cequeau*, et les comparaisons des résultats associés impliquent donc un changement de modèle hydrologique en plus d'un changement de formule d'*ETP*. Pour résumer, les travaux initiés ici ébauchent quelques conclusions mais nécessitent d'être repris et approfondis pour que cette question du rôle des estimations d'*ETP* sur la robustesse des modèles soit traitée de façon exhaustive.

Rq. Le modèle *Mordor* dispose de sa propre formulation d'*ETP* et représente donc une configuration intéressante pour étudier le rôle attribuable à la formule d'*ETP* dans les problèmes de robustesse mis en avant par le *GSST*. Cependant, son utilisation fut trop anecdotique pour que nous l'évoquions ici. Toutefois, au moment de finaliser ce manuscrit, des tests complémentaires avec ce modèle menés par *EDF-DTG* semblent indiquer une meilleure robustesse moyenne (même si toutes les erreurs de volumes ne sont pas pour autant résorbées).

6.4.3. Modification de la relation liant *ETR* et *ETP* dans *GR4J*

Notre seconde piste concerne la façon de calculer une évapotranspiration réelle (*ETR*) à partir d'une *ETP*. Tout comme la piste précédente, elle ne sera pas traitée en détail, les tests présentés ici constituant seulement un début d'exploration. Nous souhaitons en quelque sorte estimer la potentialité de cette piste, comme source d'explication des problèmes de robustesse constatés dans nos tests. Ils sont menés avec le modèle *GR4J* et consistent en une modification de la relation liant la part d'*ETR* issue du réservoir de production aux niveaux d'*ETP* et de remplissage de ce réservoir (*cf.* section 6.4.1). La formulation originale dans *GR4J* ainsi que plusieurs options alternatives sont représentées sur la figure 6.9 (le paramètre α contrôlant la forme de cette relation).

La contrainte de forme de cette courbe a été relâchée et nous avons reproduit nos tests de calage-contrôle en considérant α comme un paramètre libre du modèle, optimisé conjointement avec les quatre autres paramètres de *GR4J*. La figure 6.10 présente les écarts de paramètres

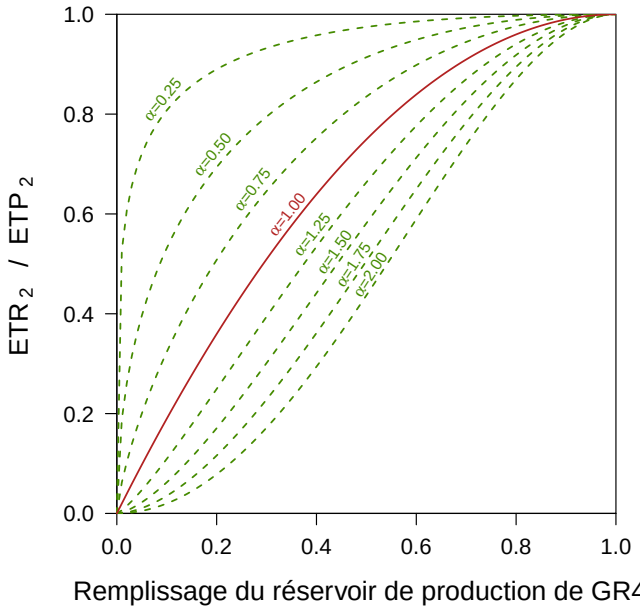


FIGURE 6.9.: Rendement évaporatoire du réservoir sol de *GR4J* : situation originale (continu rouge) et alternatives proposées (pointillés verts)

engendrés par l'ajout de ce nouveau degré de liberté. Elle met en avant des résultats proches pour tous les paramètres de *GR4J*, sauf $\theta_{X2(CES)}$. Ce résultat n'est pas réellement surprenant puisque nous avions vu que ce paramètre subissait les variations les plus importantes pour ajuster le bilan lors de la phase de calage. Or le paramètre α constitue un autre moyen pour le modèle d'ajuster son bilan. Par conséquent, les valeurs prises par de $\theta_{X2(CES)}$ dans cette version de *GR4J* à cinq paramètres ne sont pas indépendantes de celles prises par α. Notons enfin que le paramètre α varie sur notre échantillon selon une gamme relativement large. La médiane de cette distribution se situe autour de 0.9. Cette différence avec la configuration initiale de *GR4J* (où nous avions α=1), induit un maintien de l'ETR_2 à des niveaux plus proches de l'ETP_2 pour des taux de remplissage équivalents du réservoir de production. Si en moyenne le modèle fait ainsi sortir plus d'eau via l'*ETR*, les bilans n'en sont pas nécessairement affectés puisque nous notons dans le même temps une légère diminution des valeurs de $\theta_{X3(ROUT)}$ qui aura pour effet de réduire l'export d'eau par échange souterrain.

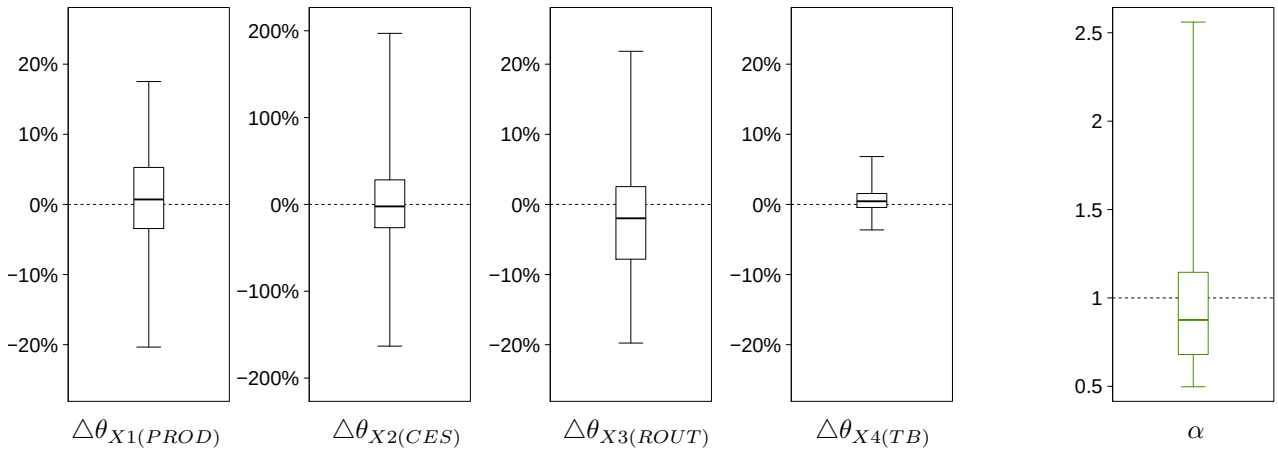


Figure 6.10.: Variations des paramètres calés de *GR4J* engendrée par l'ajout du paramètre α modifiant la relation ETR∼ETP *(quantiles 5, 25, 50, 75 et 95%)*

$\triangle\theta = \theta\,[\,GR4J_{initial}\,]\;/\;\theta\,[\,GR4J_{modif}\,]\;-\;1$ *échelle différente pour* $\theta_{X2(CES)}$
α *= paramètre de forme de la relation liant ETR/ETP au taux de remplissage du réservoir sol*

Résultats obtenus sur 410 bassins pour des calages sur toute la période disponible.
Alimentation de GR4J par SPAZM, une ETP d'Oudin et couplage avec CemaNeige ($\theta_{fixés}$).
Bassins issus des 569 dont le taux de lacunes sur la période totale est inférieur à 10%.

6.4.4. Comparaison des résultats

Les résultats de ces tests sont évalués sous l'angle des variations de biais de simulation $(\overline{\widehat{Q}}/\overline{Q})$entre les différentes sous-périodes temporelles, la réduction de ces variations apparaissant comme une condition nécessaire à l'amélioration de la robustesse du modèle. Pour cela, nous étudions les écarts-types des biais glissants sur 10 ans ($\sigma[\omega_{PT}]$, *cf.* annexe G.4) obtenus

pour les alternatives faisant intervenir différentes formules d'*ETP*, ainsi que la modification de *GR4J* présentée ci-dessus. En raison de plages temporelles légèrement différentes entre les différentes sources de données, la comparaison rigoureuse des biais glissants n'a été menée que sur 211 bassins français (parmi les 365 initiaux). Les résultats obtenus sont synthétisés sur la figure 6.11.

Nous retrouvons ici des niveaux similaires d'erreurs sur les biais glissants moyen terme entre les différentes configurations testées (numérotées de A à E), avec des écarts-types atteignant 7% en moyenne. Conformément aux discussions précédentes, le changement de la formule utilisée pour calculer l'*ETP* et des données de forçages ne conduit pas à une amélioration nette des performances. Si nous notons une légère supériorité des performances pour *Cequeau*, il nous est délicat de conclure sur le rôle de la formulation d'*ETP*. En effet, la structure et le degré de paramétrisation de *Cequeau* sont particulièrement éloignés de ceux de *GR4J*, *Mordor6* ou *SimHyd*. Enfin, le fait de libérer la contrainte sur la relation entre *ETP* et *ETR* dans *GR4J* ne permet aucunement d'améliorer les erreurs de reproduction par le modèle des bilans moyen terme observés (*cf.* configuration E).

La figure 6.12 vient compléter ces résultats en présentant les degrés de similitude entre les courbes de biais obtenues par bassin entre ces configurations. Cette méthode de mesure, détaillée en annexe G.4, fut notamment utilisée dans le chapitre 5 pour comparer les comportements entre différents calages temporels ou différents modèles hydrologiques. Les valeurs les plus basses indiquent une similarité de forme maximale, une valeur de 0.2 signifiant que le coefficient de variation (écart-type/moyenne) des biais glissants est cinq fois supérieur aux dissemblances entre les alternatives comparées. Pour rappel, nous obtenions des valeurs inférieures à 0.1 lorsque nous comparions les courbes de biais issues de différents calages temporels (*cf.* section 5.6.4).

Analysées de pair avec la figure 6.11, ces distributions révèlent une situation qui n'était pas visible dans les analyses précédentes : tandis que les courbes de biais ont des écarts-types équivalents, leurs formes présentent des différences. Le passage d'une *ETP* d'*Oudin* calculée sur des températures *SPAZM* à une *ETP* de *Penman-Monteith* calculée sur des données *SAFRAN* entraîne des dissemblances sur les variations temporelles de biais moyen terme. Ces dissemblances augmentent lorsque la configuration C est utilisée, correspondant à un changement de formule d'*ETP* (*Oudin*→*Thornthwaite*) mais aussi de modèle hydrologique (*GR4J*→*Cequeau*). Les écarts les plus marqués sont obtenus entre les configurations D et A, correspondant à un changement complet de la base de forçages (*SPAZM*→*SAFRAN*) mais en conservant la même formule d'*ETP* (*Oudin*) et le même modèle hydrologique (GR4*J*). Le test de modification du calcul de l'*ETR* dans *GR4J* (config. E) fait exception avec des courbes dont les variations restent très proches de la situation référente (config. A).

La figure 6.13 fournit une illustration des différences de courbes rencontrées sur deux bassins : *la Grande Creuse à Fresselines* et *la Durance à Embrun*. Nous y retrouvons les similitudes entre

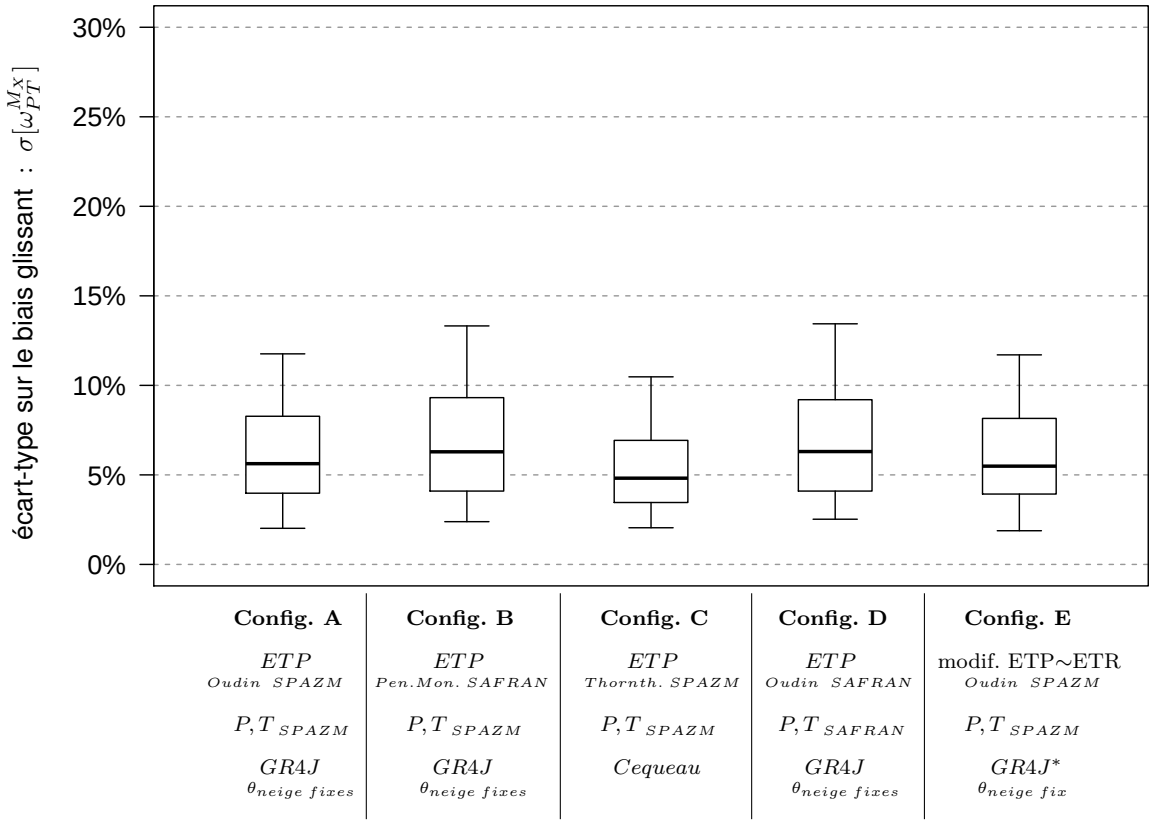


FIGURE 6.11.: Écarts-types du biais glissant sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$ pour différentes configurations sur un groupe de 211 bassins français *(quant. 5, 25, 50, 75 et 95%)*

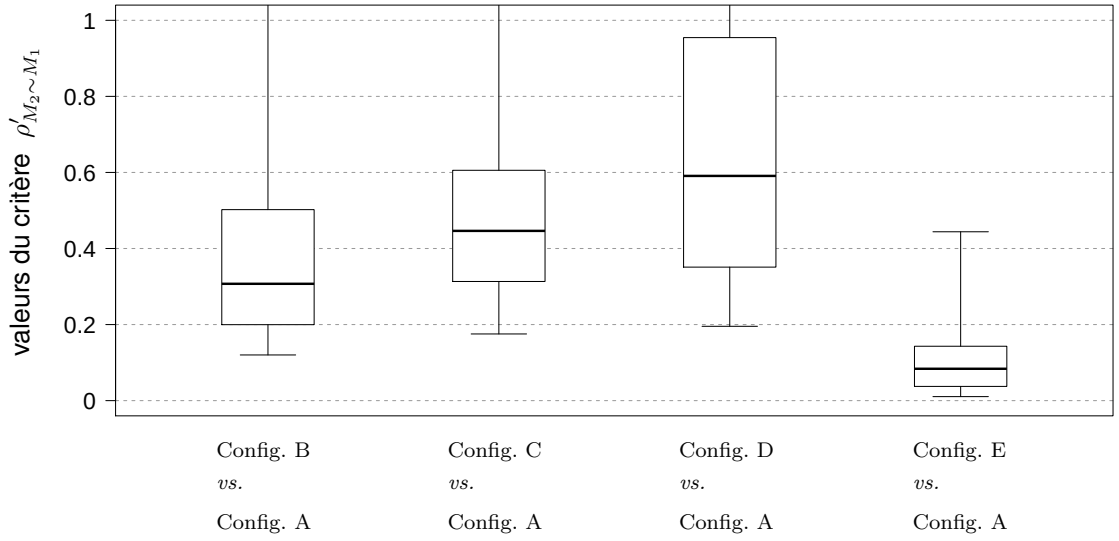


FIGURE 6.12.: Mesures de la similitude des biais glissants obtenus sur 211 bassins français pour différentes configurations de calcul d'évapotranspiration

les configurations A et E. Le remplacement d'une *ETP* d'*Oudin* par les valeurs de *Penman-Monteith* issues de *SAFRAN* (config. B) provoque des nettes différences de comportement de *GR4J* sur *la Durance à Embrun*. Nous retrouvons également un comportement différent pour la configuration C, mais nous rappelons que celle-ci implique un changement de modèle hydrologique (or *GR4J* et *Cequeau* ont des structures bien différentes). Enfin, les différences non négligeables observées entre les configurations A (en rouge) et D (en bleu) posent quant à elles des questions sur l'estimation des précipitations, seule l'origine des données de précipitations et température changeant entre ces configurations. Rappelons sur ce point que *la Durance à Embrun*, où les différences sont les plus marquées, comporte un régime nival et des altitudes élevées, ce qui soulève des réflexions quant à l'extrapolation altitudinale des précipitations et l'estimation des volumes de neige stockés sur les sommets (ces volumes étant d'ailleurs libérés durant le printemps et l'été, une période où l'*ETP* est maximale).

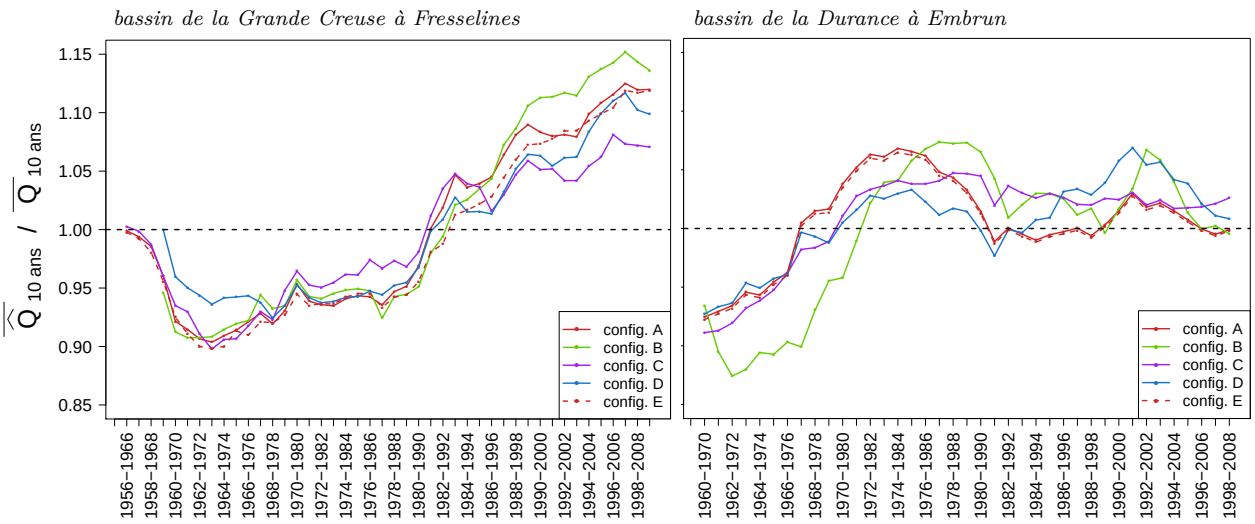


FIGURE 6.13.: Exemples des courbes de biais glissants pour différentes configurations de forçages sur *la Grande Creuse à Fresselines* et *la Durance à Embrun*

config A. $ETP_{Oudin\ SPAZM}$ $(P,T)_{SPAZM}$ $GR4J_{\theta_{neige\ fixes}}$
config B. $ETP_{Pen.Mont.\ SAFRAN}$ $(P,T)_{SPAZM}$ $GR4J_{\theta_{neige\ fixes}}$
config C. $ETP_{Thornth.\ SPAZM}$ $(P,T)_{SPAZM}$ $Cequeau$
config D. $ETP_{Oudin\ SAFRAN}$ $(P,T)_{SAFRAN}$ $GR4J_{\theta_{neige\ fixes}}$
config E. $modif.\ ETP \sim ETR_{Oudin\ SPAZM}$ $(P,T)_{SPAZM}$ $GR4J^{*}_{\theta_{neige\ fixes}}$

6.4.5. Discussion

Plusieurs indices nous laissent penser que la modélisation incorrecte des variations des forçages et notamment de l'évapotranspiration réelle constitue un point majeur pouvant expliquer une majorité des cas de manque de robustesse constatés dans nos travaux. Parmi ces indices, mentionnons le fait que les pertes de performances semblent liées à des erreurs conséquentes sur les volumes écoulés, dont les variations temporelles sont incorrectement reproduites (*cf.* courbes des biais glissants). Or, ces erreurs de reproduction des bilans semblent corrélées avec les écarts de température ou de précipitations moyennes entre les périodes de calage et de validation, pour un certain nombre de cas. Dans le même temps, nous savons que l'évapotranspiration réelle est une grandeur extrêmement complexe à estimer à l'échelle d'un bassin versant. Celle-ci dépend de multiples variables liées à certaines conditions atmosphériques (température, vent, exposition aux radiations), aux conditions d'humidité du sol du bassin versant (en lien avec les volumes de précipitations, la présence de rivières et lacs ou encore l'affleurement de nappes) et au type de couverture du sol (type de végétation). Pour complexifier le tout, cette couverture du sol est elle-même fortement liée aux deux premiers points (*ex.* variations saisonnières de la végétation mais aussi évolution progressive sur de nombreuses années).

L'estimation de l'*ETR* est donc complexe. Les différences constatées entre l'*ETP* d'*Oudin* calculée sur les températures de la base *SAFRAN* et l'*ETP* de *Penman-Monteith* fournie avec cette base, sont symptomatiques des difficultés d'estimation de ces grandeurs (*cf.* section 5.3.2). La pondération de cette *ETP* par la disponibilité en eau, pour passer ainsi de l'*ETP* à l'*ETR,* rajoute ensuite une couche supplémentaire de difficulté d'estimation.

Par conséquent, tout pourrait laisser croire que nous tenons la cause de nos problèmes de robustesse des modèles hydrologiques conceptuels. Les différentes pistes que nous avons explorées dans cette direction n'ont pas permis d'aboutir à des résultats encourageants. Nous n'avons pas réussi à améliorer les estimations pour obtenir de meilleures performances en extrapolation climatique, mais les différences de forme sur les courbes de biais glissants constatées lors des changements de base de forçages montrent l'importance que ces données ont sur la qualité des simulations. Traiter rigoureusement cette question dépassait largement le cadre de cette thèse, qui constitue (seulement) une première plongée dans la problématique de robustesse temporelle et climatique des modèles hydrologiques. Nos différentes analyses confirment toutefois que la question des données de forçages, et notamment l'évapotranspiration, constitue un axe de recherche crucial pour la poursuite des travaux sur ce thème.

6.5. Un test de correction des paramètres selon les conditions de transfert

6.5.1. Introduction

Dans un certain nombre de cas, les erreurs de simulation sont corrélées aux variations de conditions climatiques subies lors du transfert temporel des paramètres du modèle. Or nous avons vu que pour certains modèles peu paramétrés, il était possible de retrouver ces corrélations au niveau des valeurs des paramètres permettant l'ajustement du bilan lors du calage (*cf.* section 5.5). Rappelons que ces corrélations ne concernent pas les valeurs absolues des paramètres, mais les variations relatives entre les périodes impliquées dans le transfert. Par exemple, nous constatons sur certains bassins que le paramètre $\theta_{X1(CPL)}$ de *Mordor6* sera en moyenne plus élevé que la valeur optimale lorsqu'il est calé sur une période où la température moyenne de l'air est inférieure (*cf.* figure 5.11, page 161).

Puisque nous disposons de sous-groupes de plusieurs dizaines de bassins sur lesquels une tendance moyenne peut être établie, l'idée suivante nous est venue :

Si nous disposons d'un jeu de paramètres et que nous connaissons le $\triangle$climat lors du passage du calage à la simulation, pourrions-nous déterminer la correction souhaitable des paramètres pour améliorer la robustesse du modèle ?

Bien entendu, déterminer de telles relations et plus encore pouvoir les utiliser sur un bassin spécifique devient complexe dès que nous sortons du cadre théorique de nos tests systématiques. En effet, il est préalablement nécessaire d'avoir clairement identifié, sur le bassin étudié, les relations liant (a) les variations de paramètres spécifiques, (b) les erreurs du modèle et (c) les changements de conditions entre calage et validation. Nous sommes conscients que ce type de configuration favorable n'est certainement pas la normalité en pratique, ni au niveau des structures de modèles (des interactions entre paramètres étant souvent présentes), ni au niveau des séries d'observations (*cf.* section 3.2, où nous montrons que les corrélations entre climat et erreurs sont très hétérogènes entre les bassins).

Par conséquent, nous pouvons d'ores et déjà affirmer que cette piste ne constitue pas une voie généralisable d'amélioration de la robustesse des modèles. Elle s'apparente plutôt à une astuce pour corriger des paramètres qui devraient théoriquement varier peu temporellement (et climatiquement). Conscients des nombreuses limites de cette piste, la curiosité nous pousse tout de même à l'explorer. Nous sommes curieux de déterminer si le simple ajustement *post-calage* des paramètres, selon une relation préalablement établie, améliorerait effectivement la robustesse moyenne d'un modèle utilisé dans des conditions lui étant défavorables.

6.5.2. Procédure de test

La procédure mise au point pour explorer cette piste est liée au comportement de *GR4J* et *Mordor6* sur les 120 bassins où des liens ont été détectés entre robustesse et écarts de température lors des transferts des paramètres (*cf.* section 3.3). Utilisé à plusieurs reprises lors des explorations menées dans ce chapitre, ce sous-groupe a l'avantage de présenter des tendances relativement claires entre les erreurs de volume et le $\triangle\overline{T}_{D\to R}$ (*ex.* $\overline{T}_D < \overline{T}_R \Leftrightarrow \widehat{Q}_R > Q_R$ et inversement). De plus, des tendances associées se retrouvent sur les variations de certains paramètres de *GR4J* et *Mordor6*, notamment ceux qui permettent d'ajuster le bilan au calage (*cf.* sections 5.5.2).

La procédure que nous avons imaginée se décompose de la façon suivante :

1. En reprenant les résultats issus du *GSST*, nous identifions les paramètres de *GR4J* et *Mordor6* qui semblent varier (en moyenne) avec les variations temporelles de température sur nos 120 bassins (*cf.* section 5.5).
   Il s'agit principalement du paramètre $\theta_{X2(CES)}$ pour *GR4J* et du paramètre $\theta_{X1(CPL)}$ pour *Mordor6*. Nous considérons également les paramètres $\theta_{X3(ROUT)}$ et $\theta_{X1(PROD)}$ pour *GR4J*, ainsi que le paramètre $\theta_{X3(TRN)}$ pour *Mordor6*, dont les variations sont moins prononcées mais ne peuvent pas pour autant être écartées. Tous ces paramètres permettent directement ou indirectement de participer à l'ajustement du bilan au calage, certaines interactions pouvant exister entre eux (*ex.* $\theta_{X2(CES)}$ et $\theta_{X3(ROUT)}$ pour *GR4J*).

2. Le groupe de 120 bassins est divisé en deux sous-groupes, que nous noterons A et B. Ce découpage est aléatoire, mais nous vérifions que les caractéristiques de ces groupes demeurent semblables (*ex.* surface et altitudes des bassins mais aussi volumes de précipitations ou d'écoulements estimés, *cf.* figure H.5 en annexe, page 343).

3. Sur chacun des paramètres retenus, nous estimons une relation liant l'écart séparant la valeur transférée de la valeur calée et l'écart de température moyenne subi lors du transfert. Cette relation entre $\left(\theta_{Xi\,[R]} - \theta_{Xi\,[D]}\right) \;/\; |\,\theta_{Xi\,[D]}\,|$ et $\triangle\overline{T}_{D\to R}$ est estimée indépendamment sur les sous-groupes de bassins A et B.
   Notre objectif étant de déterminer une relation corrective des paramètres, il nous faut trouver la modification qu'il *faudrait faire* pour *ramener* le paramètre donneur $\theta_{Xi\,[D]}$ vers la valeur $\theta_{Xi\,[R]}$ qui constitue notre cible [2]. Nous exprimons pour cela les variations des $\theta_{Xi\,[R]}$ en fonction des $\theta_{Xi\,[D]}$ et inversons ainsi les rôles sont par rapport à nos représentations habituelles (*c.-à-d.* estimation $(\theta_R - \theta_D) \,/\, |\theta_D|$ ici contre $(\theta_D - \theta_R) \,/\, |\theta_R|$ précédemment).
   Nous avons choisi d'approximer ces relations par des régressions linéaires, qui sont illustrées à la figure 6.14. Par souci de simplification, seules les pentes de ces régressions sont retenues

2. Pour rappel, θ_R est le jeu obtenu en calage sur la période receveuse, *c.-à-d.* la période que nous souhaitons simuler en transférant le jeu θ_D.

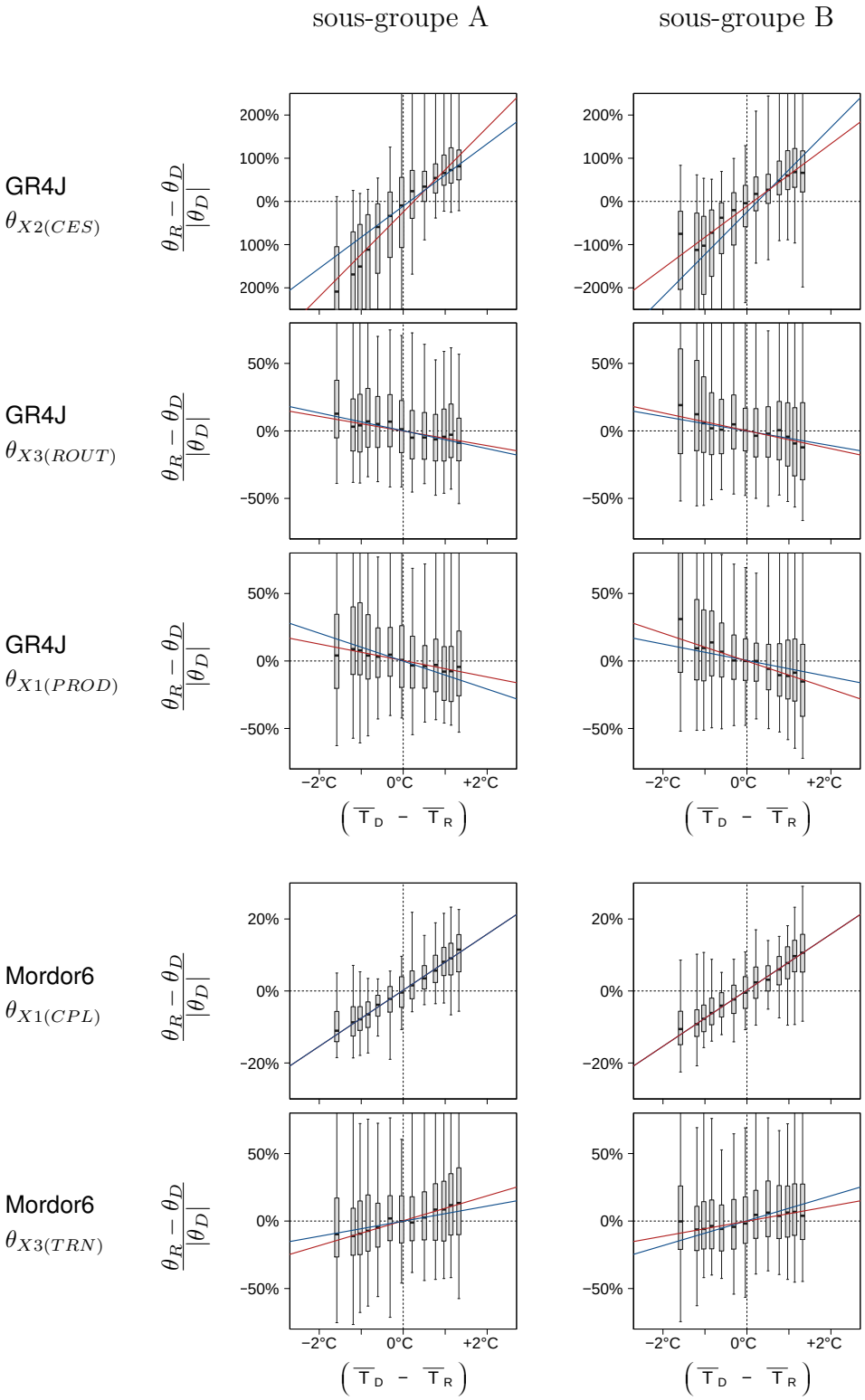


FIGURE 6.14.: Estimations de régressions linéaires entre $\triangle\theta$ et $\triangle\overline{T}$ pour certains paramètres de *GR4J* et *Mordor6*

En rouge : régression linéaire construite à partir des données du sous-groupe (calage) ; en bleu : régression linéaire issue de l'autre sous-groupe (transferts B→A et A→B) (validation).

Estimations réalisées sur les 120 bassins concernés par des corrléations entre △erreurs et △températures, divisés en deux sous-groupes. Sur les bassins français, modèles alimentés par des données SPAZM, une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).

pour la suite de la procédure, les ordonnées à l'origine étant considérées nulles. Pour chaque paramètre θ_{Xi}, nous estimons ainsi le couple de pentes $p_{[Xi\,,A]}$ et $p_{[Xi\,,B]}$, qui constituent les facteurs correctifs de ce paramètre en fonction du $\triangle\overline{T}_{D\to R}$, respectivement obtenus sur les sous-groupes indépendants A et B.

4. Les facteurs correctifs étant déterminés, nous testons s'ils permettent d'améliorer la robustesse des modèles sur le sous-groupe non utilisé pour leur estimation (*c.-à-d.* $p_{[Xi\,,A]}$ est utilisé pour corriger les bassins du groupe B et inversement).
   Pour évaluer cette procédure corrective, nous reprenons tous les jeux obtenus par calage sur les différentes sous-périodes et nous les transférons temporellement en effectuant la modification suivante :
   $$\theta'_{Xi[D\to R]} = \theta_{Xi[D]} \; + \; p_{[Xi]} \cdot \triangle\overline{T}_{D\to R} \cdot |\theta_{Xi[D]}| \tag{6.3}$$
   Ces corrections sont appliquées aux paramètres $\theta_{X1(PROD)}$, $\theta_{X2(CES)}$ et $\theta_{X3(ROUT)}$ pour *GR4J*, et aux paramètres $\theta_{X1(CPL)}$ et $\theta_{X3(TRN)}$ pour *Mordor6*.
   Un même jeu θ_D sera transformé autant de fois qu'il sera transféré vers d'autres périodes (tous les ${\theta'}_{D\to R}$ étant différents car fonction du $\triangle\overline{T}_{D\to R}$ subi lors du transfert). Afin de rester dans un cadre de validation, nous veillons enfin à ce que le facteur $p_{[Xi]}$ utilisé provienne du sous-groupe auquel le bassin considéré n'appartient pas.

5. Les résultats de ces simulations sont analysés selon notre approche habituelle par nuage de points. La comparaison des formes de nuages ainsi obtenues avec la situation initiale nous renseigne sur les progrès de robustesse éventuellement réalisés.

6.5.3. Résultats

Les résultats obtenus avec cette procédure pour *GR4J* et *Mordor6* sont présentés aux figures 6.15 et 6.16, où ils sont comparés à la situation initiale (*GSST* simple avec transfert des paramètres sans procédure corrective). Les illustrations des valeurs corrigées de paramètres ont, quant à elles, été placées dans l'annexe H.2 (pages 344-345).

La procédure de correction des paramètres $\theta_{X2(CES)}$, $\theta_{X3(ROUT)}$ et $\theta_{X1(PROD)}$ de *GR4J* en fonction des conditions de transfert permet certaines améliorations de la robustesse sur les bassins testés ici, dans des proportions toutefois variées. Une augmentation de la robustesse est constatée sur les deux sous-groupes de 60 bassins (*cf.* augmentations du *MRC* moins prononcées et pentes du biais moins fortes sur la figure 6.15, particulièrement au niveau des médianes de distributions). Cette diminution est nettement plus marquée sur le sous-groupe A (coefficients correctifs estimés sur le sous-groupe B) que sur le sous-groupe B (coefficients correctifs estimés sur le sous-groupe A).

Les coefficients correctifs estimés sur le sous-groupe A par régression linéaire ($\triangle\theta_{Xi} = p_{[Xi,\,A]} \cdot \triangle\overline{T}$) ne permettent donc pas une amélioration de la robustesse sur le sous-groupe B aussi

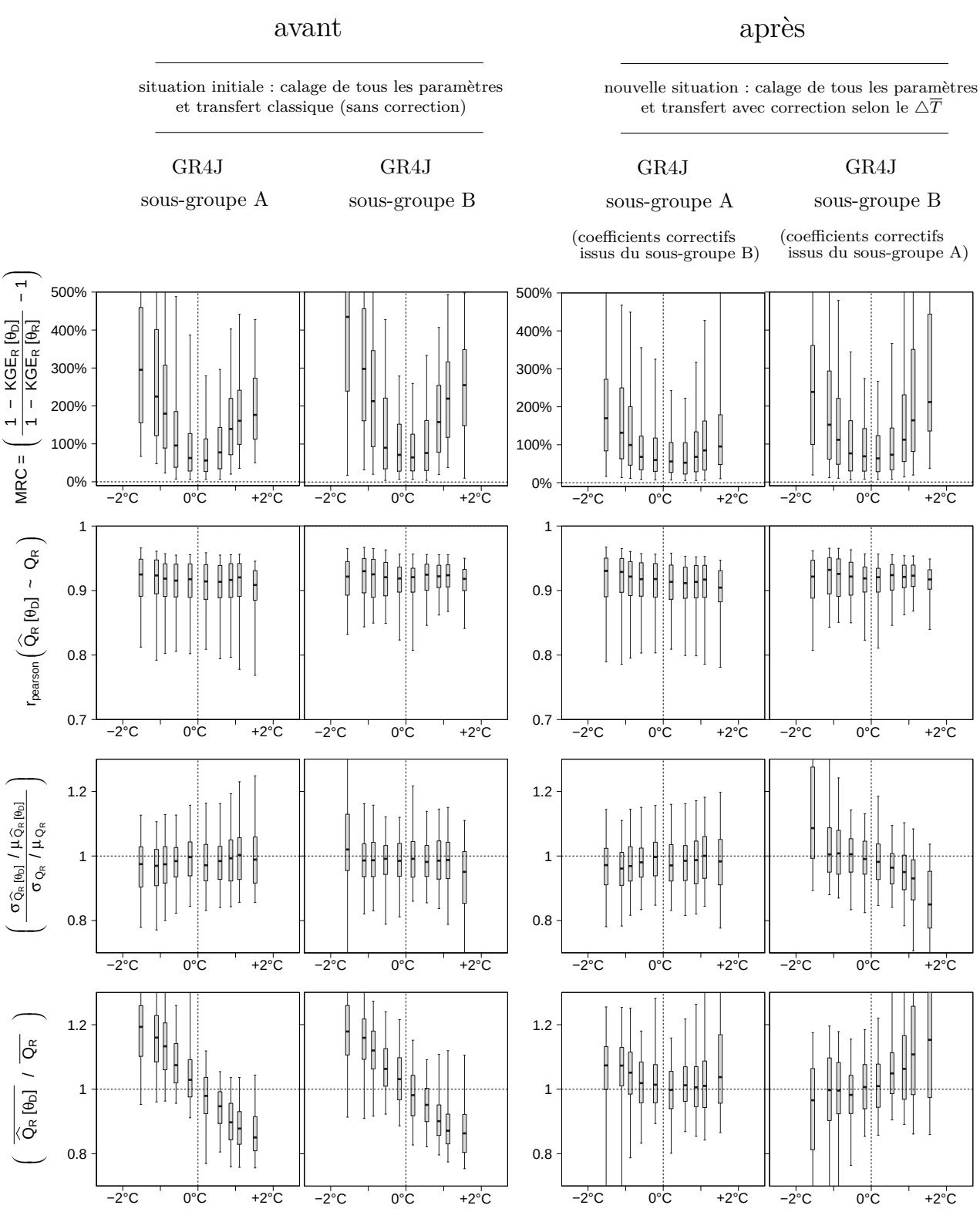


FIGURE 6.15.: Niveaux de robustesse obtenus pour *GR4J* avec une correction des paramètres selon les conditions de transfert

Tests sur les 120 bassins concernés par des corrléations significatives entre $\triangle\varepsilon$ et $\triangle\overline{T}$, et divisés en deux sous-groupes (A et B). Coefficients correctifs des paramètres déterminés par régression linéaire sur l'autre sous-groupe de bassins (cf. figure 6.14). Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Rq. sur les bassins français, GR4J est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés).

efficace que lorsque les rôles de A et B sont inversés. La correction des paramètres semble même trop importante puisqu'elle va jusqu'à inverser la tendance constatée sur le biais (*cf.* passage de $\overline{T}_D - \overline{T}_R \nearrow \Leftrightarrow \overline{\widehat{Q}} / \overline{Q} \searrow$ à $\overline{T}_D - \overline{T}_R \nearrow \Leftrightarrow \overline{\widehat{Q}} / \overline{Q} \nearrow$). Cette inversion se retrouve d'ailleurs sur les valeurs du paramètre $\theta'_{X2(CES)}$ obtenues après la procédure de correction (*cf.* figure H.6, page 344 en annexe). Enfin, nous constatons des variations de paramètres plus dispersées avec la procédure corrective que dans la situation initiale (dispersion détectable à travers la hauteur plus importante des boîtes, *cf.* figure H.6).

La procédure de correction des paramètres $\theta_{X1(CPL)}$ et $\theta_{X3(TRN)}$ de *Mordor6* selon les conditions de transfert a des effets bénéfiques nettement plus prononcés. L'augmentation des erreurs de simulation avec le contraste thermique subi lors du transfert semble quasiment supprimée lorsque nous considérons les résultats moyens sur nos deux sous-groupes de 60 bassins (*cf.* figure 6.15). En effet, les portions des nuages de *MRC* et de biais qui concernent des transferts où $\triangle\overline{T}_{D\to R} \neq 0$ sont recentrées sur des valeurs similaires à celles obtenues avec $\triangle\overline{T}_{D\to R} \simeq 0$. Les variations relatives du paramètre $\theta'_{X1(CPL)}$ selon le $\triangle\overline{T}_{D\to R}$ confirment simultanément le succès de cette procédure de compensation et le fort lien entre $\triangle\theta_{X1(CPL)}$ et $\triangle erreurs_{D\to R}$ sur les 120 bassins utilisés ici (*cf.* figure H.7, page 345 en annexe).

6.5.4. Discussion

Nous avons testé sur une partie de l'échantillon une procédure de correction de paramètres de *GR4J* et *Mordor6* en fonction des conditions climatiques dans lesquelles ceux-ci étaient transférés temporellement. Par construction, cette procédure n'a pu être menée que sur les modèles et les bassins où nous avions réussi à isoler (en partie) une relation entre les variations de quelques paramètres et les variations de certaines caractéristiques climatiques (*cf.* section 5.5). Ce test exploratoire est donc soumis à une double contrainte particulièrement limitante, quant au choix du modèle, d'une part, et au choix des bassins, d'autre part.

Nous constatons une différence d'efficacité de cette procédure lorsqu'elle est utilisée pour *GR4J* ou *Mordor6*. L'une des causes probables de cette différence réside dans la façon dont l'ajustement du biais au calage est réalisé pour chacun de ces modèles. La manière la plus directe de modifier les équilibres entre $\overline{P}$, $\overline{ETP}$ et $\overline{Q}$ pour *Mordor6* est une multiplication des précipitations (via $\theta_{X1(CPL)}$), qui a lieu à l'amont du modèle hydrologique et donc indépendamment de ce qu'il s'y passe. Il existe ainsi une relation assez directe entre une augmentation ou une diminution de $\theta_{X1(CPL)}$ et les conséquences en termes de biais de simulation. Dans ce cadre, l'utilisation d'une régression linéaire pour déterminer la relation corrective semble bien adaptée. Cette relation est moins directe pour *GR4J*, où l'ajustement du bilan est majoritairement fait via un échange souterrain. Les volumes importés ou exportés dépendent de la valeur du coefficient d'échange ($\theta_{X2(CES)}$) mais également de la capacité du réservoir de routage ($\theta_{X1(ROUT)}$) et de la quantité d'eau qui y transite. La capacité du réservoir de

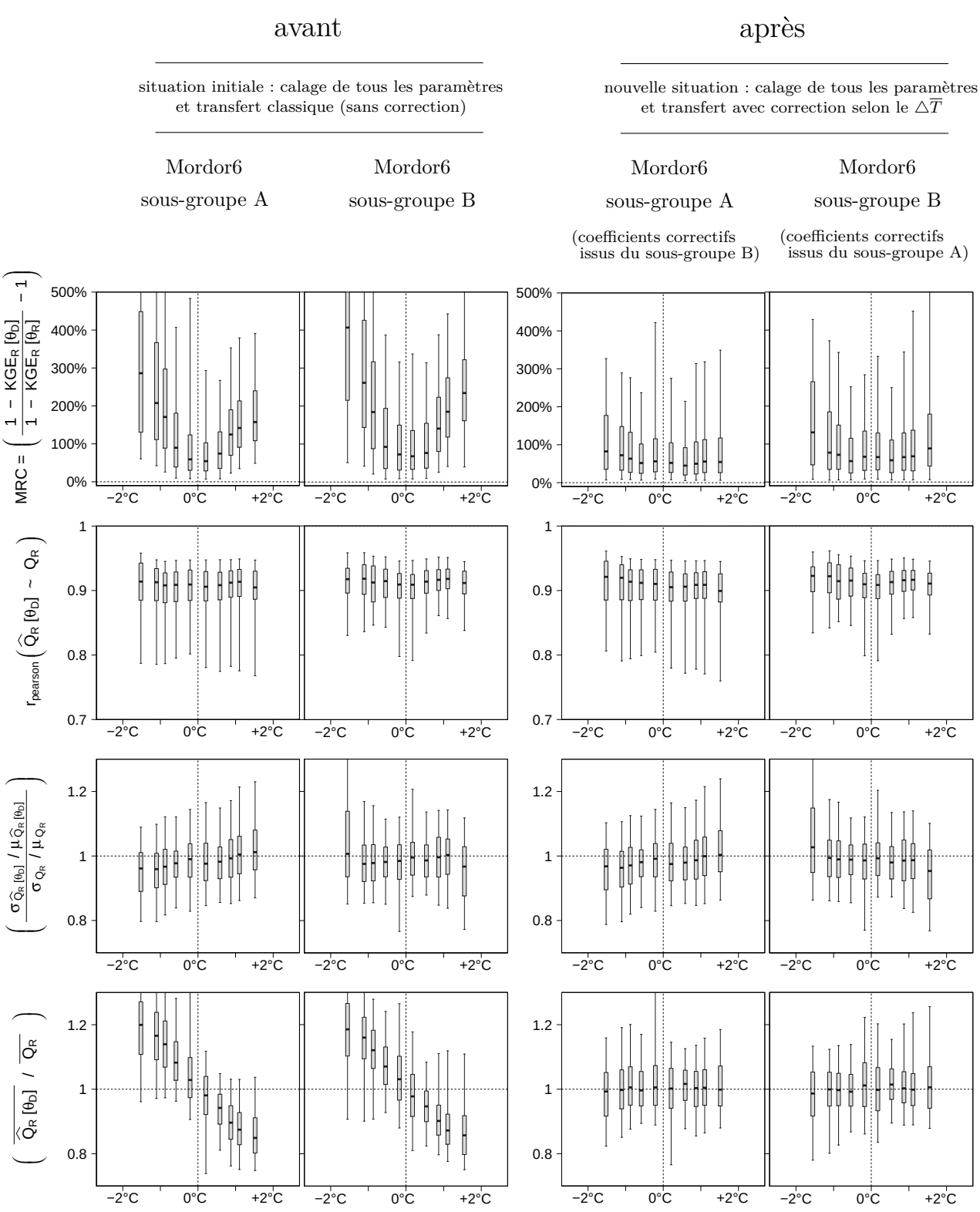


FIGURE 6.16.: Niveaux de robustesse obtenus pour *Mordor6* avec une correction des paramètres selon les conditions de transfert

Tests sur les 120 bassins concernés par des corrléations significatives entre $\triangle\varepsilon$ et $\triangle\overline{T}$, et divisés en deux sous-groupes (A et B). Coefficients correctifs des paramètres déterminés par régression linéaire sur l'autre sous-groupe de bassins (cf. figure 6.14).
Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% sur chaque bande veticale.

Rq. sur les bassins français, Mordor6 est alimenté par des données SPAZM, une ETP d'Oudin et il est couplé à CemaNeige (θ_{neige} fixés).

production ($\theta_{X1(PROD)}$) est un autre moyen pour le modèle *d'évacuer* plus ou moins d'eau du système par évapotranspiration. La situation est donc moins triviale que pour *Mordor6* et nous voyons également que le choix d'une régression linéaire n'est pas réellement adapté pour caractériser la relation $\triangle\theta_{X2(CES)} = f(\triangle\overline{T})$ (*cf.* figure 6.14). Ces différents points contribuent vraisemblablement à l'efficacité amoindrie de cette procédure de correction sur *GR4J*. Ce travail pourrait être affiné et nous pourrions aussi évaluer cette procédure corrective des paramètres pour diverses variations de la structure des modèles (*ex.* en ajoutant ou retirant certains degrés de liberté). Une idée, indéniablement extravagante, serait de rechercher une structure délibérément problématique en termes de robustesse climatique mais plus facilement corrigible. Malgré ces différentes idées, plusieurs caractéristiques intrinsèques à une telle procédure corrective limitent les perspectives de cette piste.

6.5.5. Perspectives

Des résultats plus que satisfaisants ont été obtenus pour le modèle *Mordor6* sur nos deux sous-groupes de 60 bassins, qui jouent alternativement le rôle d'échantillon de calage et de validation dans la procédure corrective testée. Devons-nous pour autant nous satisfaire de celle-ci et voir ici la solution (tant espérée) aux problèmes de robustesse temporelle et climatique de nos modèles ? Comme nous le laissions entrevoir en introduction, la réponse est malheureusement négative.

Plusieurs raisons rendent l'utilisation de cette piste très difficile en pratique (par opposition à notre cadre théorique des tests, qui est favorable sur de nombreux aspects).
Tout d'abord, cette procédure corrective ne peut être utilisée que sur des modèles dont la structure conduit à l'existence de corrélations entre (a) les variations de quelques paramètres, (b) les erreurs de simulation et (c) les variations climatiques du bassin étudié. Ce premier point constitue, déjà en lui-même, une barrière importante à l'utilisation généralisée de cette approche corrective. Elle nécessite, d'une part, d'une conceptualisation du modèle qui conduit à l'émergence de corrélations claires sur les paramètres et, d'autre part, de chroniques d'observation suffisamment longues pour établir ces corrélations (via le *GSST* par exemple).
Rappelons ensuite que les relations correctives ne sont estimées ici que sur une seule variable climatique (celle jugée la plus influente). Il est toutefois possible que d'autres variations climatiques influencent la qualité des transferts et que l'efficacité de la procédure corrective soit ainsi conditionnée par l'importance de ces influences de deuxième ordre.
Estimer des relations correctives simultanément *multi-paramètres* et *multi-variables-climatiques*, tout en souhaitant que ces relations soient robustes, semble difficilement réalisable en pratique. Comme nous l'avons vu, les liens entre robustesse des modèles et évolution des conditions climatiques ne sont pas homogènes sur notre échantillon. Les variables explicatives des problèmes de transférabilité des paramètres peuvent en effet varier entre les bassins. Pour certains bas-

sins, nous détectons une sensibilité des erreurs de modélisation à $\triangle\overline{T}$ mais pas à $\triangle\overline{P}$ (*cf.* section 3.3.2). Cette sensibilité est inversée pour d'autres bassins (*c.-à-d.* lien avec $\triangle\overline{P}$ mais pas $\triangle\overline{T}$, *cf.* section 3.4.2), tandis que pour d'autres encore, nous n'avons pas réussi à identifier des liens du type $\triangle erreurs = f(\triangle climat)$, malgré des problèmes de transférabilité du même ordre de grandeur. Il nous faudrait donc raisonner au cas par cas, les relations correctives étant alors estimées sur chaque bassin. Si des projections sous changement climatique doivent être émises, nous pourrions ainsi imaginer une étude du type *GSST* sur toute la chronique disponible et l'estimation de coefficients correctifs sur les paramètres et les variables climatiques qui paraissent pertinents. Ces paramètres seraient alors ajustés en fonction des changements de climat prévus (*c.-à-d.* en fonction des scénarios utilisés mais aussi de l'horizon de projection considéré). De nombreux travaux complémentaires sont toutefois nécessaires pour évaluer la robustesse d'une telle approche.

En complément de ces limites et d'une manière plus générale, nous ne pouvons nous satisfaire qu'à moitié de ce type d'approche qui n'élimine en rien la cause du manque de robustesse climatique des modèles hydrologiques testés. Elle s'apparente à une méthode *curative* visant à corriger, dans la mesure du possible, une défaillance connue dans notre système de modélisation. Idéalement, nous souhaiterions plutôt trouver des méthodes *préventives* qui permettraient notamment une représentation adéquate des bilans par les modèles et feraient disparaître ainsi une source majeure des pertes de performances révélées par le *GSST*. Parmi ces méthodes *préventives*, nous pensons bien sûr à d'éventuelles modifications de la structure des modèles en difficulté, pour leur permettre de mieux représenter des dynamiques basse fréquence. Émerge ainsi de ces réflexions l'idée d'approches comparatives entre différentes structures de modèles, où une attention particulière serait portée sur les variations temporelles des biais sur les débits moyens simulés.

6.6. L'espace adimensionnel Q/P=f(P/ETP), un outil d'analyse

6.6.1. Présentation du mode de représentation

En calculant les moyennes des séries temporelles de précipitations ($\overline{P}$), d'évapotranspiration potentielle ($\overline{ETP}$) et de débits ($\overline{Q}$), nous pouvons positionner chaque bassin au sein d'une représentation adimensionnelle où $x = \overline{P}/\overline{ETP}$ et $y = \overline{Q}/\overline{P}$ (*cf.* figure 6.17). Pour les bassins où les écoulements souterrains sont négligeables sur le bilan long terme, le placement sur cet espace est contraint par deux limites *physiques* : la contrainte d'infériorité des écoulements face aux précipitations en moyenne (*c.-à-d.* limite $\overline{Q} = \overline{P}$ soit $y = 1$) et la contrainte d'infériorité du déficit d'écoulement ($\overline{P} - \overline{Q}$) face à l'évapotranspiration potentielle en moyenne (*c.-à-d.* limite $\overline{P} - \overline{Q} = \overline{ETP}$ soit $y = 1 - 1/x$). Les zones pour lesquelles ces contraintes ne sont pas respectées sont représentées en gris sur la figure 6.17. En pratique, les bilans *réalistes* se positionnent sur une zone plus restreinte que l'espace laissé en blanc ici, les deux frontières n'étant jamais atteintes.

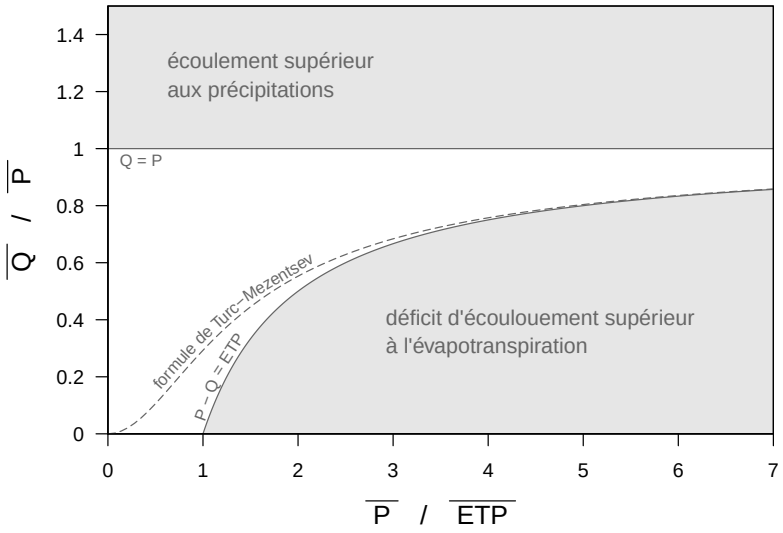


Figure 6.17.: Présentation de la représentation adimensionnelle $\overline{Q} \,/\, \overline{P} = f\left(\overline{P} \,/\, \overline{ETP}\right)$

Les zones grisées délimitent le positionnement théorique des bassins où les écoulements souterrains sont négligeables sur le bilan long terme.

Cette représentation adimensionnelle des bassins versants peut constituer un outil d'analyse à part entière pour mieux comprendre le fonctionnement de leurs cycles hydrologiques ou détecter d'éventuelles anomalies dans les données. En remplaçant les débits observés par des grandeurs simulées, *Le Moine et al.* (2007) ont utilisé cette représentation pour étudier la capacité de différentes structures de modèles à représenter des flux souterrains entre bassins. Dans le cadre de nos travaux, cet outil a toutefois été relativement peu exploité mais constitue une piste autour de laquelle nous pourrions baser des travaux futurs.

La relation de *Turc-Mezentsev* (1954; 1955) constitue un modèle certes très simplifié mais qui permet de nombreuses analyses des bilans hydrologiques (*cf. Mouelhi et al.*, 2006; *Andréassian et Perrin*, 2012). Cette relation peut être traduite au sein de cet espace adimensionnel sous la forme d'une courbe (indiquée en pointillé sur la figure 6.17). *Le Moine et al.* (2007) proposent de généraliser cette relation et de définir ainsi une famille de courbes définies par l'équation ci-dessous, la relation de *Turc-Mezentsev* classique étant obtenue pour α=1 et γ=2.

$$y = 1 - \frac{1}{\left(1 + \left(\frac{x}{\alpha}\right)^{\gamma}\right)^{1/\gamma}} \tag{6.4}$$

6.6.2. Un outil d'analyse pour critiquer les données

Au cours de nos travaux, nous avons abordé la question de la sensibilité de nos analyses au choix de la base de forçages utilisée pour alimenter les modèles (*cf.* sections 5.1.2 et 5.3.2). La représentation adimensionnelle $\overline{Q} \,/\, \overline{P} = f\left(\overline{P} \,/\, \overline{ETP}\right)$ présente dans ce cadre un outil mettant en évidence les différences entre ces bases. Si nous disposons par ailleurs de connaissances sur la zone *réelle* de positionnement des bassins dans cet espace, cette représentation peut également permettre de critiquer le réalisme des données issues d'une base comparativement à une autre.

La figure 6.18 présente la position des bassins de notre échantillon au sein de cet espace adimensionnel. Les grandeurs $\overline{P}$, $\overline{ETP}$ et $\overline{Q}$ utilisées ici sont calculées à partir des séries temporelles utilisées dans la plupart de nos tests. Les données utilisées pour les bassins australiens sont celles mises à notre disposition par le *CSIRO*. Les données utilisées pour les bassins français correspondent aux précipitations issues de *SPAZM*, aux *ETP* d'*Oudin* calculée à partir des températures issues de la base *SPAZM* et aux débits issus de la banque *HYDRO*. Quelques relations de *Turc-Mezentsev* généralisées sont représentées en pointillés verts (pour α=0.9 et γ=[0.5 ;1 ;2.5]), la relation classique de *Turc-Mezentsev* étant en pointillés larges rouges.

Nous constatons la présence de certains bassins dans les zones grisées, les bassins les plus montagneux au sein de notre échantillon étant particulièrement concernés. Cette présence peut être une indication de deux phénomènes.

- Une première explication est l'existence d'échanges souterrains significatifs entre les bassins versants concernés et leur voisinage (sous forme de gain pour la zone grisée supérieure ou de pertes pour la zone grisée inférieure droite). Bien que cette explication soit probable ponctuellement, nos critères de constitution de l'échantillon nous ont conduits à retenir majoritairement des bassins de montagne (situés dans les Alpes, le Jura ou les Pyrénées) où les échanges avec les eaux souterraines ne devraient (*a priori*) pas représenter un processus déterminant dans les cycles hydrologiques. Ce point mériterait néanmoins quelques vérifications plus approfondies.

- Une seconde explication est l'existence d'erreurs d'estimation des grandeurs qui composent ces bilans hydrologiques ($\overline{P}$, $\overline{ETP}$ et $\overline{Q}$). En effet, le positionnement des bassins de moyenne et haute montagne dans la zone inférieure droite pourrait être causé par une surestimation des précipitations et/ou une sous-estimation de l'évapotranspiration potentielle.

Nous voyons ainsi un exemple d'utilisation de cette représentation adimensionnelle comme un outil pour critiquer les données que nous utilisons. Reprenons avec ce nouvel outil les comparaisons menées précédemment entre différentes bases de forçages (*cf.* section 5.3.2). Présentés à la figure 6.19, ces résultats mettent en avant quelques points intéressants sur les 365 bassins où nous disposons des trois configurations. Nous constatons par exemple une réduction notable du nombre de points (69 *vs.* 106) sortant de la zone théorique lorsque l'*ETP* d'*Oudin*, calculée sur les températures *SPAZM*, est remplacée par celle de *Penman-Monteith*, calculée sur les données *SAFRAN* (*cf.* config. 2 *vs.* config. 1). Le remplacement de la base *SPAZM* par la base *SAFRAN* (pour une même formule d'*ETP*) entraîne également une diminution des points dans la zone grisée inférieure droite (93 *vs.* 105), mais au détriment du passage de quelques points dans la zone grisée supérieure (*cf.* config. 3 *vs.* config. 1).

Restons toutefois prudents face à ces résultats car il nous est difficile de dire quelle configuration est la plus proche de la réalité. La représentation adimensionnelle des bilans ouvre néanmoins des champs de réflexions. La possibilité d'une correction trop importante des précipitations *SPAZM* en altitude en fait partie. Rappelons toutefois que la base *SAFRAN* souffre

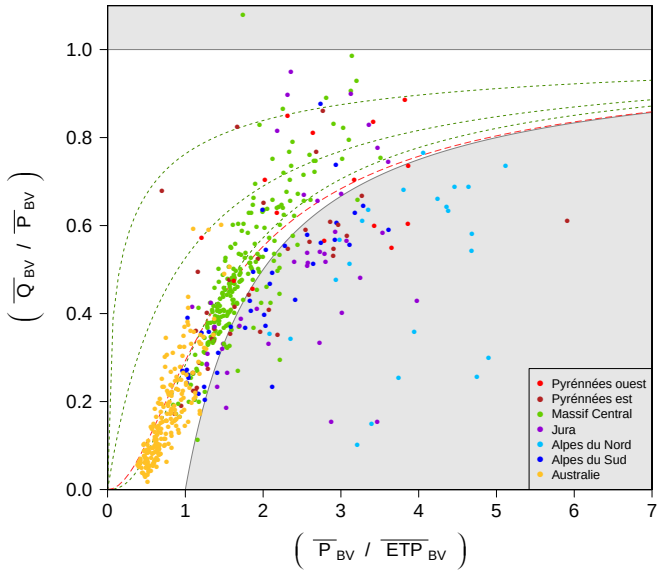


Figure 6.18.: Représentation des bilans observés pour les bassins de notre échantillon au sein de l'espace adimensionnel $\overline{Q} / \overline{P} = f\left(\overline{P} / \overline{ETP}\right)$

Pour chaque bassin, valeurs moyennées sur toute la série disponible des débits observés, des précipitations et ETP moyennes de basssin estimées (base SPAZM et formule d'Oudin). Le code couleur indique la localisation des bassins représentés, avec une distiction selon le massif montagneux d'appartenance pour les bassins français. Les lignes pointillées sont quelques courbes de la relation de Turc-Mezentsev généralisée.

d'une mauvaise prise en compte de l'impact orographique sur les volumes de précipitations (*Gottardi*, 2009) et que le développement de la base *SPAZM* repose sur la volonté d'améliorer cette représentation. Il n'est donc pas possible de trancher aussi nettement sur un réalisme inférieur des précipitations issues de *SPAZM* par rapport à celles issues de *SAFRAN*. Un second champ également très important est la possibilité d'une sous-estimation globale de l'évapotranspiration en zone de montagne. Les températures y étant nettement plus faibles qu'en plaine, il est possible que les formulations usuelles d'estimation de l'*ETP* ne conduisent pas à des volumes représentatifs des processus réels (*Herrnegger et al.*, 2012).

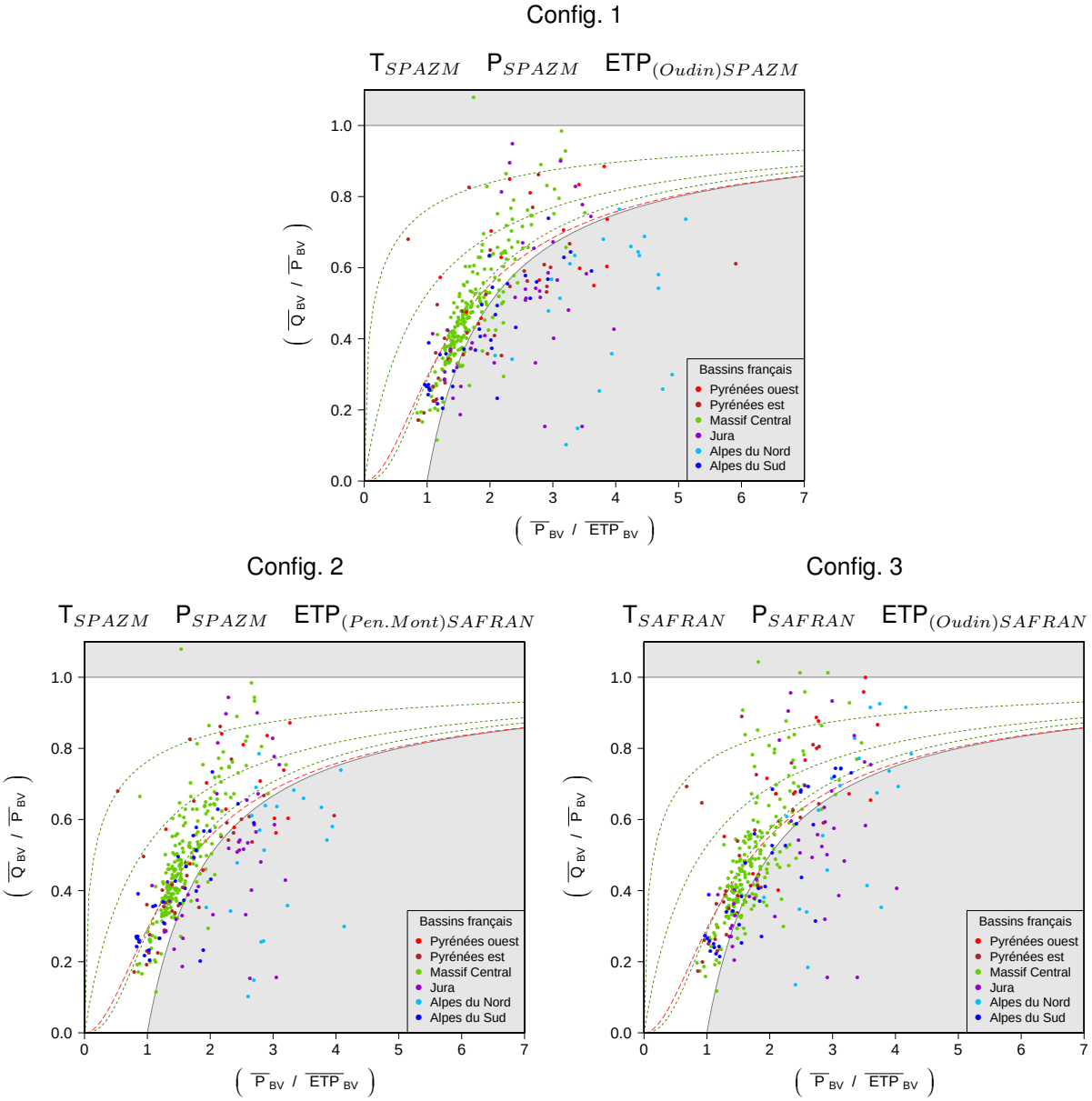


FIGURE 6.19.: Comparaison entre différentes bases de forçages au sein de l'espace adimensionnel $\overline{Q}/\overline{P} = f\left(\overline{P}/\overline{ETP}\right)$ pour 365 bassins français

Valeurs moyennées sur toute la série disponible des débits observés et des précipitations et ETP estimées. Le code couleur indique la localisation des bassins représentés et les lignes pointillées sont quelques courbes de la relation de Turc-Mezentsev généralisée.

6.6.3. Un outil d'analyse pour étudier les variations temporelles de bilan

La représentation adimensionnelle des bilans peut également être utilisée pour étudier leurs variations temporelles, lorsque des sous-périodes suffisamment longues sont considérées. Nous utiliserons ici les sous-périodes de 10 ans construites pour la procédure *GSST* (découpage selon une fenêtre glissante). Chaque bassin est donc représenté par plusieurs points dans l'espace adimensionnel, selon les grandeurs $\overline{P_{SP[i]}}$, $\overline{ETP_{SP[i]}}$ et $\overline{Q_{SP[i]}}$ obtenues sur les différentes sous-périodes $SP[\,i\,]$.

Plusieurs situations peuvent alors être rencontrées :

- Si les variables $\overline{P}$, $\overline{ETP}$ et $\overline{Q}$ ne varient pas temporellement pour un bassin, alors tous les points correspondant à l'analyse des bilans moyen terme seront superposés. Cette situation est toutefois extrêmement rare et présente peu d'intérêt pour nous puisqu'elle ne permet aucune analyse.
- Si les variables $\overline{P}$, $\overline{ETP}$ et $\overline{Q}$ varient temporellement, nous pouvons étudier la forme du nuage de points ainsi constitué et tenter des parallèles avec les comportements des modèles sur le bassin (*ex.* sa capacité à assurer un faible biais sur les différentes périodes temporelles pour un même calage).

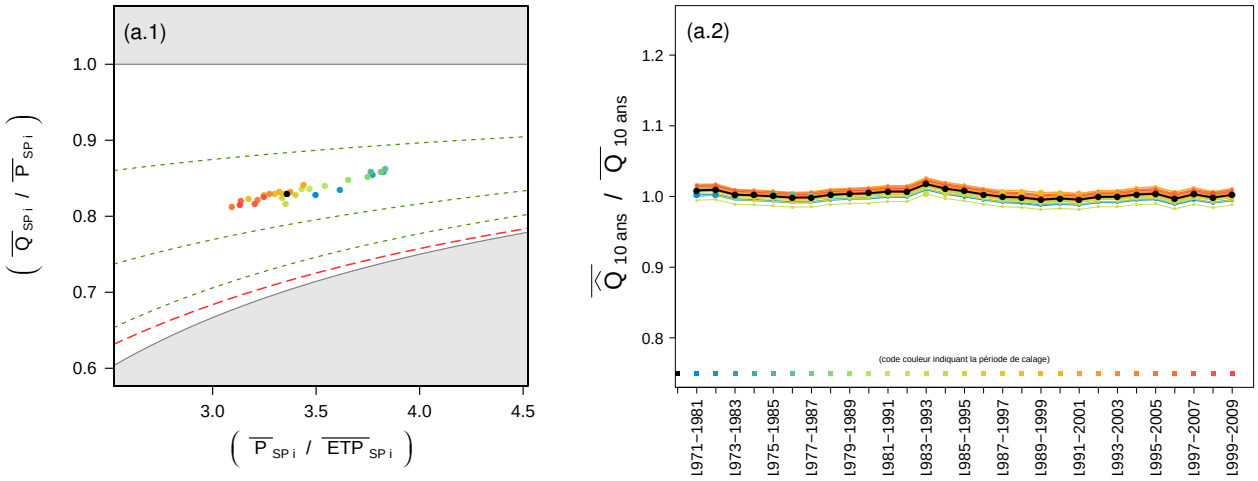
Quelques exemples de représentations adimensionnelles des bilans moyen terme sont fournis aux figures 6.20 et 6.21. Nous représentons chaque sous-période par un point de couleur différente (le point noir correspondant au bilan calculé sur toute la série disponible). Afin d'établir le parallèle avec la robustesse des modèles hydrologiques sur ces bassins, nous avons reporté à droite de chaque représentation les graphiques issus de l'analyse des biais glissants de modélisation pour *GR4J* (*cf.* section 5.6). Avant de passer à l'analyse des résultats, notons deux conventions de représentation pour ces figures. Tout d'abord, le code couleur utilisé ici traduit la progression temporelle entre les sous-périodes et ne reflète pas spécifiquement une variation des conditions climatiques entre celles-ci (même s'il est vrai qu'une tendance à l'augmentation des températures existe). Ensuite, les représentations adimensionnelles des bilans ont été réalisées selon des échelles mobiles, assurant un zoom sur la partie du graphique où se situe le bassin considéré. Nous avons toutefois veillé à ce que les amplitudes restent identiques sur les deux axes ($x = \overline{P}/\overline{ETP}$ et $y = \overline{Q}/\overline{P}$) pour faciliter les comparaisons entre bassins.

Ces figures permettent une constatation très intéressante :

Les résultats obtenus sur *la Bienne à Jeurre* et sur *la Venelle à Selongey* illustrent des situations où le modèle reproduit convenablement les débits sur toutes les sous-périodes, c'est-à-dire où nous obtenons des courbes de biais glissants plates et donc superposées (*cf.* graphiques a.2, b.2). Or, dans ces situations, les bilans de ces sous-périodes forment dans l'espace adimensionnel un alignement de points plutôt en accord avec l'une des relations de *Turc-Mezentsev* généralisées évoquées précédemment (*cf.* graphiques a.1, b.1).

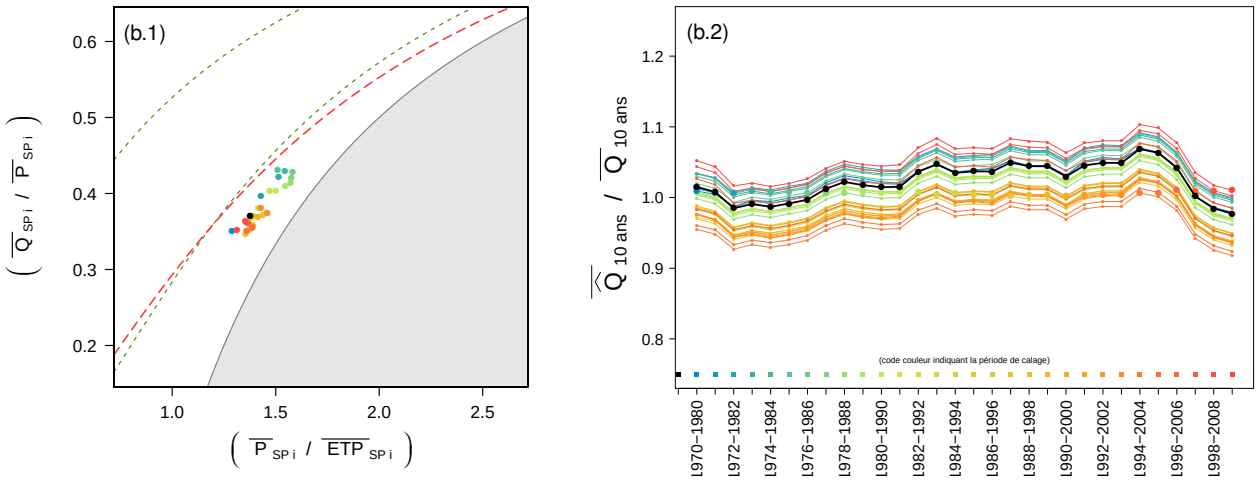
(a) bassin de *la Bienne à Jeurre*

S=590km^2, Z_{med}=940m, $\overline{P}$=1880mm, $\overline{ETP}_{oudin}$=560mm, $\overline{Q}$=1590mm



(b) bassin de *la Venelle à Selongey*

S=56km^2, Z_{med}=430m, $\overline{P}$=910mm, $\overline{ETP}_{oudin}$=640mm, $\overline{Q}$=340mm



(c) bassin de *l'Azergues à Chatillon*

S=340km^2, Z_{med}=530m, $\overline{P}$=990mm, $\overline{ETP}_{oudin}$=670mm, $\overline{Q}$=400mm

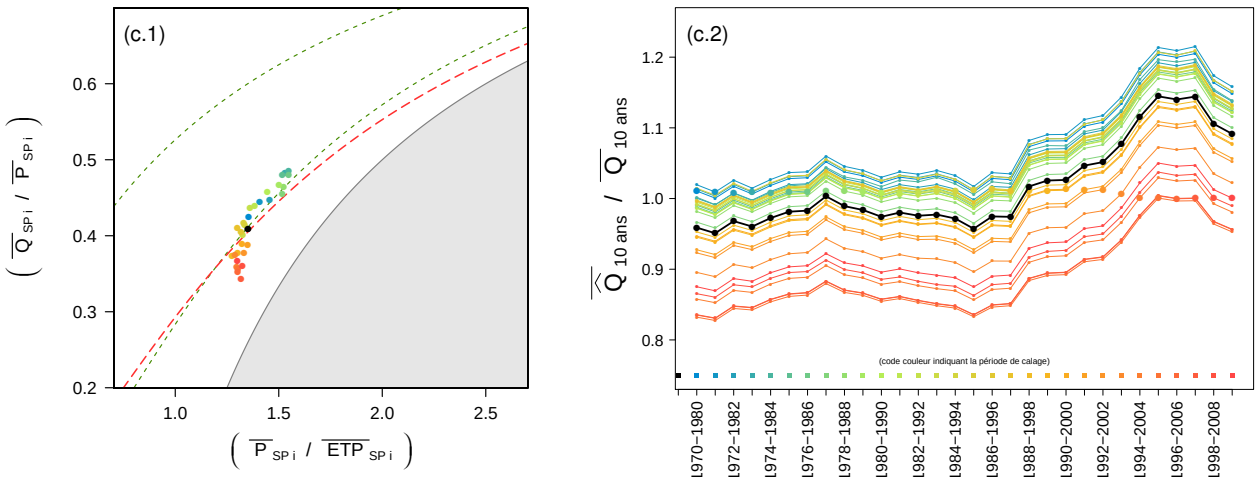


FIGURE 6.20.: Exemples d'analyse des variations temporelles de bilan dans l'espace adimensionnel Q/P=f(P/ETP) et parallèle avec les biais glissants de *GR4J* (ex. 1-3)

Rq. GR4J est alimenté ici par des SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Le code couleur n'a pas d'autre signification que la différentiation temporelle.

Les résultats obtenus sur le bassin de *l'Azergues à Chatillon* illustrent une situation intermédiaire intéressante (*cf.* graphiques c.1 et c.2). Les erreurs de volume sont stables sur plus de la moitié de la chronique (1970 à 1998) et les points correspondants sont alignés dans l'espace *Q/P=f(P/ETP)*. Les biais diffèrent sur la dernière partie de la chronique (sous-périodes représentées en rouge), et les points correspondants sortent de l'alignement jusqu'ici respecté.

Les trois autres exemples illustrent enfin des cas où le modèle (*GR4J*) a de grandes difficultés à reproduire les bilans moyens termes, *c.-à-d.* signal de biais glissant marqué par de fortes variations (*cf.* graphiques d.2, e.2 et f.2). Nous remarquons que les bilans des sous-périodes correspondantes, exprimés dans l'espace adimensionnel, forment soit un alignement dont la pente moyenne diffère de celle d'une relation de *Turc-Mezentsev* généralisée (*cf.* graphique d.1), soit des nuages de points relativement éclatés (*cf.* graphiques e.1 et f.1). Les résultats obtenus sur le bassin de *l'Ubaye à Barcelonnette* sont particuliers puisque les grandeurs $\overline{P}$, $\overline{ETP}$ et $\overline{Q}$ estimées se situent dans une zone grisée de la représentation adimensionnelle. Ce positionnement nous ramène à nos discussions précédentes sur les problèmes d'erreurs systématiques dans l'estimation de ces grandeurs (*cf.* section 6.6.2).

Nous pouvons ainsi mettre en évidence une concordance entre les situations de réussite ou d'échec des modèles et la présence ou l'absence d'un alignement des bilans moyen terme sur une courbe de type *Turc-Mezentsev* généralisée, au sein de l'espace $\overline{Q}/\overline{P} = f(\,\overline{P}/\overline{ETP}\,)$. Cette concordance ne doit cependant pas être sur-interprétée. Elle nous permet de reconnaître une équivalence dans la capacité d'extrapolation climatique de modèles conceptuels (plus ou moins complexes) et d'approches beaucoup plus simples, telles que la formule de *Mouelhi et al.* (2006) croisée dans la section 5.6.4 ou la relation de *Turc-Mezentsev* utilisée ici. Néanmoins, cette concordance ne constitue pas, à elle seule, une preuve que les situations problématiques de biais des modèles résultent uniquement d'erreurs dans l'estimation des composantes du bilan réel (débit moyen et surtout précipitations et évapotranspiration moyennes de bassin).

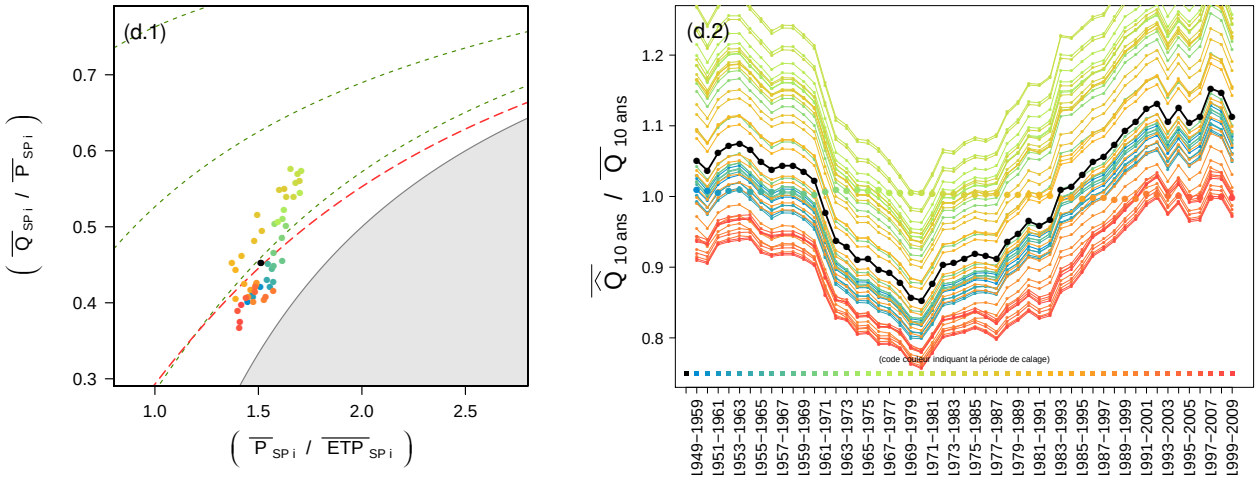
Au final, deux portes restent donc entrouvertes sur les causes des problèmes de robustesse rencontrés dans nos travaux :

(a) une capacité insuffisante des modèles testés à reproduire certaines dynamiques lentes d'équilibre hydrologique des bassins,

(b) des biais non constants dans l'estimation des composantes du bilan, notamment les précipitations et évapotranspiration moyennes de bassin, qui évoluent en fonction des conditions climatiques moyennes.

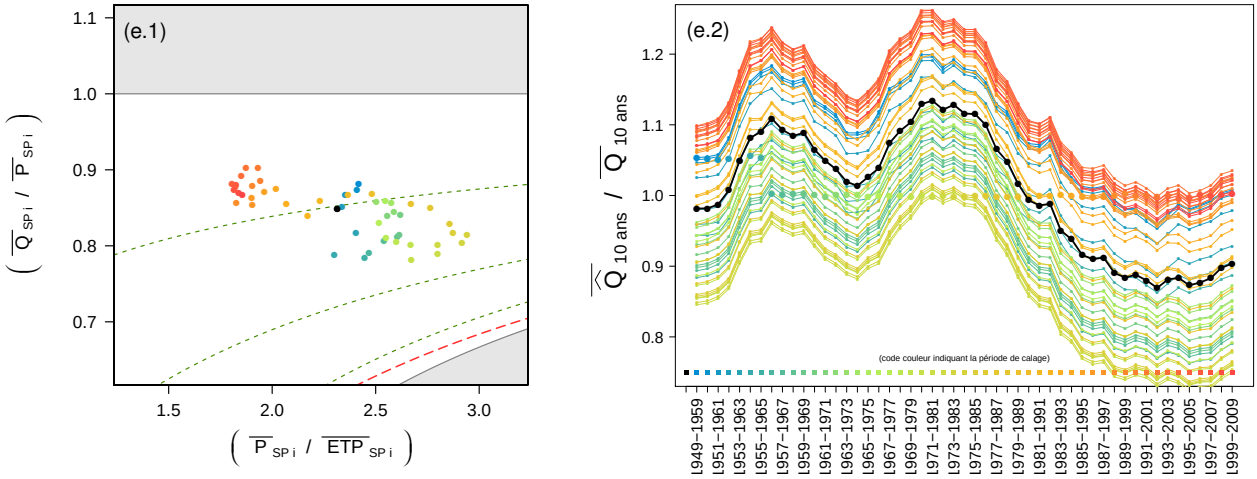
(d) bassin de *l'Allier à Vieille-Brioude*

S=2260km^2, Z_{med}=1070m, $\overline{P}$=880mm, $\overline{ETP}_{oudin}$=570mm, $\overline{Q}$=410mm



(e) bassin de *la Nive des Aldudes à St-Etienne-de-Baigorry*

S=160km^2, Z_{med}=650m, $\overline{P}$=1680mm, $\overline{ETP}_{oudin}$=700mm, $\overline{Q}$=1420mm



(f) bassin de *l'Ubaye a Barcelonnette*

S=540km^2, Z_{med}=2270m, $\overline{P}$=1210mm, $\overline{ETP}_{oudin}$=410mm, $\overline{Q}$=600mm

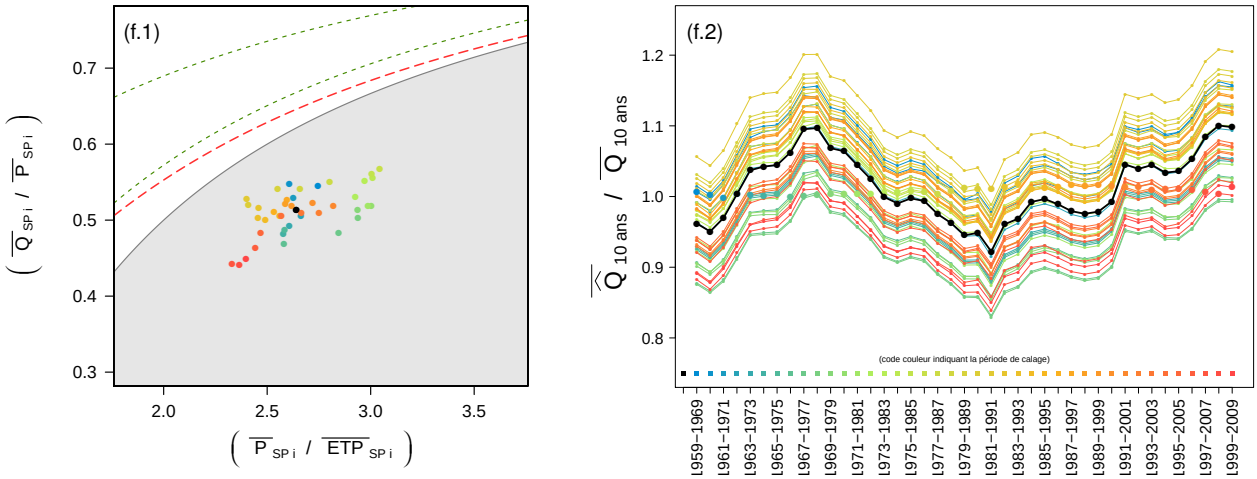


FIGURE 6.21.: Exemples d'analyse des variations temporelles de bilan dans l'espace adimensionnel Q/P=f(P/ETP) et parallèle avec les biais glissants de *GR4J* (ex. 4-6)

Rq. GR4J est alimenté ici par des SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Le code couleur n'a pas d'autre signification que la différentiation temporelle.

6.7. État des lieux

Ce chapitre présente les différentes pistes que nous avons envisagées pour permettre une amélioration de la robustesse des modèles hydrologiques conceptuels. Elles sont explorées ici à des degrés divers et pourront largement faire l'objet d'approfondissements lors de travaux futurs.

Bien qu'une procédure de correction post-calage des paramètres ait pu montrer son efficacité sur les modèles *GR4J* et *Mordor6,* aucune des pistes testées n'a permis de résoudre, à la source, le manque de transférabilité des paramètres mis en avant par le *GSST*. Au-delà des limites de la méthode corrective proposée ici, il peut sembler légitime d'attendre d'un modèle qu'il reproduise un bilan hydrologique moyen, or les biais significatifs de modélisation constatés au chapitre 5, montrent que cette reproduction n'est pas garantie pour les modèles considérés dans nos travaux (même en situation de calage sur toute la série disponible). Par conséquent, la recherche de solutions *préventives* plutôt que *curatives* reste l'objectif ciblé.

Concernant les pistes de solutions *préventives*, les tests de changements de fonction objectif n'ont pas permis un gain de robustesse pour *GR4J* (alimenté par une *ETP* d'Oudin et couplé à *CemaNeige*), même lorsque les problèmes de biais sur les bilans moyen terme sont explicitement ciblés. Les travaux portant sur le rôle potentiel des erreurs d'estimation de l'évapotranspiration (potentielle et réelle) n'ont pas non plus permis de réelles avancées mais cette porte demeure largement ouverte pour de futurs travaux qui devront traiter cette question plus exhaustivement qu'elle n'est abordée ici. Enfin, nous avons vu qu'un retour à l'étude des données observées pouvait également constituer un outil pour mieux comprendre les situations d'échec des modèles hydrologiques.

Pour résumer, si nous ne sommes pas en mesure de fournir ici une liste de préconisations permettant d'augmenter la robustesse des modèles, nous pouvons tout de même dresser un bilan de nos compréhensions et des points qu'il semble intéressant d'analyser à la suite de ces travaux. La difficulté des modèles testés à reproduire les bilans hydrologiques moyens de façon simultanée sur différentes périodes constitue l'une des causes majeures des pertes de performance constatées lors des transferts de paramètres. L'identification des causes de cette situation devra constituer un point central dans les travaux qui feront suite à cette thèse. Les liens parfois détectés entre les biais des modèles et les variations climatiques (notamment de température moyenne de l'air) portent nos réflexions vers deux pistes explicatives qui devront faire l'objet d'approfondissements : (a) une structure des modèles étudiés ici qui serait inadaptée pour représenter certaines dynamiques basse fréquence ; (b) des erreurs dans l'estimation des moyennes de précipitation et surtout d'évapotranspiration (potentielle et réelle) qui seraient non-stationnaires.

Synthèse et conclusion générale

L'exercice du bilan : un retour sur les accomplissements mais aussi une confrontation aux espérances initiales...

Calvin and Hobbes, 16/08/1986, Bill Watterson©

Retour sur la problématique de cette thèse

La modélisation hydrologique, telle que nous la connaissons aujourd'hui, n'a pas été initialement pensée dans un objectif de robustesse. À juste titre, les pionniers de cette discipline ont fait porter leurs efforts sur la reproduction des débits observés, sur un nombre très limité de cas d'études, en conceptualisant puis traduisant mathématiquement les processus attenants à la relation précipitation-débit. Ce travail était déjà suffisamment délicat compte tenu de toutes les difficultés existantes pour alimenter les modèles en données et les caler (souvent manuellement). Les premiers modèles hydrologiques constituaient surtout des outils de compréhension ou de reconstitution de mesures manquantes et, dans de nombreux cas, les périodes de calage et d'utilisation n'étaient pas réellement distinctes. Comme *de Marsily et al.* (1992) l'ont très bien dit, les hydrologues essayaient juste de faire "de leur mieux". Les besoins opérationnels qui pouvaient nécessiter une sortie du champ strict des conditions observées en calage (*ex.* dimensionnement d'ouvrages de protection ou de stockage de l'eau) étaient remplis par des approches statistiques d'analyse des données.

Au fil du temps, l'usage des modèles hydrologiques s'est généralisé et ceux-ci sont aujourd'hui utilisés dans des contextes de plus en plus variés, notamment en situation d'extrapolation, où les conditions de simulation diffèrent sensiblement de celles rencontrées lors du calage du modèle. C'est par exemple le cas pour déterminer des phénomènes extrêmes (crues ou sécheresses), pour simuler les débits de bassins non jaugés ou pour émettre des projections à long terme dans un climat en évolution. Ces extrapolations peuvent ainsi être temporelles, spatiales et climatiques. Ce nouveau contexte d'utilisation place la question de leur robustesse comme une problématique centrale. L'estimation de cette robustesse devient d'autant plus nécessaire aujourd'hui que les hydrologues sont sollicités pour fournir des marges d'incertitudes en complément des valeurs estimées par une approche déterministe.

Certains hydrologues ont pointé très tôt cette nécessité de soumettre les modèles pluie-débit à des tests rigoureux et systématiques de robustesse. L'une des publications les plus notables sur ce sujet est bien sûr celle de *Klemeš* (1986), qui propose une liste de tests à réaliser selon le contexte d'utilisation du modèle. Ces tests – *split-sample test*, *proxy-basin test*, *differential split-sample test* et *proxy-basin differential split-sample* test – visent à évaluer la robustesse temporelle, spatiale et climatique d'un modèle. Les travaux menés dans le cadre de cette thèse traitent essentiellement des questions de robustesse temporelle et climatique des modèles hydrologiques conceptuels.

Résumé des travaux accomplis et principales conclusions

Une généralisation de la procédure de split-sample test

Dans un premier temps, nous avons travaillé sur une généralisation de la procédure de Klemeš par une systématisation des tests, dont l'objectif était de permettre une évaluation exhaustive de la robustesse temporelle et climatique de modèles hydrologiques conceptuels. Intitulée *Generalized Split-Sample Test* (*GSST*), cette procédure consiste à réaliser un grand nombre de tests de transfert temporel des paramètres des modèles à l'aide d'un découpage des séries d'observation. Les variations des erreurs de modélisation ainsi obtenues sont analysées au regard des conditions climatiques associées aux transferts. Le *GSST* a été utilisé pour différents modèles hydrologiques sur plusieurs centaines de bassins français et australiens.

Un outil de diagnostic puissant... mais déroutant

Un premier niveau d'analyse de ces résultats a été présenté au chapitre 3. Il montre l'existence de corrélations entre les problèmes de robustesse temporelle des modèles et les écarts de conditions climatiques subis lors des transferts. Ces erreurs prennent particulièrement la forme d'erreurs systématiques sur les volumes moyens, mais une forte hétérogénéité existe en ce qui concerne les variables climatiques associées aux problèmes détectés. Si les erreurs de modélisation semblent corrélées aux écarts de températures moyennes sur de nombreux bassins, des situations bien différentes sont observées sur d'autres bassins (*ex.* corrélations aux écarts de précipitations) sans que nous parvenions à expliquer ces disparités de comportement.

Un nouveau regard sur l'importance relative des incertitudes dans une étude d'impact du changement climatique

Le chapitre 4 a illustré cette sensibilité des simulations aux conditions de calage du modèle hydrologique dans le contexte d'une étude d'impact du changement climatique. Celle-ci a été comparée aux autres sensibilités inhérentes à un tel exercice de projection (*ex.* choix du modèle climatique global (*GCM*), de la méthode de descente d'échelle (*MDES*), etc.). Cet exercice ne constitue pas une évaluation exhaustive des incertitudes et tous les résultats mis en avant par le *GSST* n'y sont pas intégrés. Il permet tout de même une appréciation de l'impact de certains choix de modélisation sur les projections à long terme de débit. Sur le bassin étudié (*la Durance à Embrun*), ces analyses montrent que le choix de la période de calage du modèle hydrologique induit une dispersion sur les projections absolues comparable à celle causée par le faisceau de réalisations analogues d'une *MDES* et se rapprochant de celle constatée lors d'un changement de modèle hydrologique. Pourtant, les différences de simulation causées par un changement de période de calage sont nettement plus faibles lorsque ces analyses portent sur les projections relatives (c'est-à-dire les $\triangle\widehat{Q}$ par rapport à une situation de référence simulée). Ce comportement est constaté sur les projections à long terme, mais aussi sur les simulations en temps présent (issues du *GSST*). Il se retrouve pour tous les modèles testés mais est, de

loin, le plus marqué pour le modèle *GR4J* couplé à *CemaNeige* (c'est-à-dire que *GR4J* semble fiable s'il est utilisé pour évaluer des changements relatifs).

Un impact significatif des conditions de calage

Ces résultats nous ont conduits à approfondir notre diagnostic dans le chapitre 5. Des tests complémentaires y sont menés pour étudier la sensibilité de nos interprétations au contexte d'étude (*ex.* base de données utilisée, choix de découpage lors du *GSST*). Les analyses sont ensuite centrées sur les variations temporelles des paramètres optimaux et révèlent l'importance de la question de l'équilibrage des bilans dans les problèmes de robustesse constatés. En effet, les paramètres des modèles optimisés sur certaines périodes pour minimiser le biais ne sont pas toujours transférables vers d'autres périodes du même bassin. En accord avec les résultats obtenus au chapitre 3, les variations temporelles des paramètres peuvent être reliées aux variations de conditions climatiques pour un certain nombre de bassins de notre échantillon.

Une difficulté des modèles à représenter simultanément un ensemble de sous-périodes

Les analyses des biais glissants ($\overline{\widehat{Q}_{10\,ans}}/\overline{Q_{10\,ans}}$) réalisées pour divers calages et présentées dans le chapitre 5 font émerger l'un des résultats les plus marquants de nos travaux (déjà pressenti au chapitre 4) : si les calages sur différentes sous-périodes conduisent à des simulations bien distinctes, les variations relatives des modules simulés sont très similaires pour les modèles testés. Ainsi, les courbes de biais correspondant aux jeux de paramètres calés sur différentes sous-périodes ou même sur la chronique complète sont toutes parallèles. Dès lors qu'un modèle considéré présente des difficultés à reproduire simultanément les débits moyens observés sur différentes périodes, on peut prévoir des erreurs systématiques lors du transfert de ses paramètres. Ainsi, les situations de pertes de performances mises en avant par le *GSST* ne semblent pas résulter de la sélection de paramètres non robustes car inadaptés, elles découlent plutôt d'une incapacité des modèles testés à reproduire convenablement les bilans hydrologiques moyens sur toutes les sous-périodes considérées.

Des solutions peuvent être imaginées...
mais les pistes concrètes restent peu satisfaisantes

Le sixième chapitre présente une compilation de pistes envisagées pour réduire les problèmes de robustesse des modèles mis en évidence par le *GSST*. Explorées à des degrés divers, ces pistes incluent des tests d'autres méthodes de sélection des paramètres et des réflexions sur la façon dont l'évapotranspiration est modélisée. Ce chapitre rend également compte du test d'une méthode de correction des paramètres de *GR4J* et *Mordor6*, appliquée après le calage en fonction des écarts climatiques subis lors du transfert à venir. Bien que les résultats soient encourageants, nous ne sommes pas entièrement satisfaits de ce type d'approche qui s'apparente à une méthode "curative" visant à corriger une défaillance connue dans notre système de modélisation. Nous lui aurions préféré des méthodes préventives permettant de faibles biais

de simulation sur toutes les sous-périodes temporelles, faisant ainsi disparaître une source majeure des pertes de performances révélées par le *GSST*. Le chapitre s'achève par la suggestion d'un retour à l'étude des données observées dans le but de mieux comprendre les variations de comportements des bassins et expliquer peut être ainsi les situations d'échec des modèles hydrologiques. De nombreuses questions restent notamment en suspens concernant l'estimation de l'évapotranspiration réelle.

Perspectives

Les résultats de cette thèse ont été obtenus grâce à l'utilisation de plusieurs modèles hydrologiques et d'un grand nombre de bassins versants (569 pour l'échantillon au complet). Ce type d'approche présentait l'avantage de s'affranchir des risques de tomber sur un cas très (trop) spécifique, pour se concentrer sur les comportements moyens et assurer une généralité plus grande à l'interprétation des résultats. Nos travaux ont cependant montré la difficulté d'interprétation des cas particuliers.

La poursuite des analyses initiées ici semble donc passer par un recoupement rigoureux des causes pouvant conduire aux difficultés rencontrées par les modèles, et particulièrement les erreurs de reproduction des bilans hydrologiques sur différentes périodes temporelles. Deux axes majeurs émergent des travaux menés ici : (a) une amélioration de la capacité des modèles testés à simuler une dynamique d'écoulement basse fréquence ; (b) une explication des limites de transférabilité par des erreurs non stationnaires d'estimation des composantes du bilan hydrologique, notamment l'évapotranspiration moyenne.

Nous suggérons la mise en place d'un échantillon (peut être plus restreint) de bassins versants bien instrumentés où les séries d'observation disponibles seraient longues et bien renseignées pour permettre une estimation aussi précise que possible des flux moyens entrants et sortants des bassins versants. La difficulté du travail de critique des séries ne doit pas être sous-estimée, les interprétations de données ponctuelles ou tendances atypiques étant souvent délicates en pratique. Des modèles pourraient ensuite être utilisés pour expérimenter différentes formules d'évapotranspiration potentielle et diverses méthodes de calcul de l'évapotranspiration réelle au sein du modèle hydrologique. D'autres modifications de structures des modèles hydrologiques pourraient être testées dans l'optique d'augmenter leur capacité à reproduire les bilans hydrologiques moyen terme observés. Le problème de la robustesse des modules neige conceptuels, seulement entrevu dans cette thèse, pourrait également être approfondi dans ce cadre, mais il nous semble moins prioritaire que les travaux portant sur l'évapotranspiration. Enfin, la comparaison des comportements entre des modèles conceptuels et des modèles à base physique apporterait certainement des éclairages intéressants sur les processus conduisant aux problèmes de robustesse mis en avant dans cette thèse, en particulier concernant l'estimation des flux d'eau entrants et sortants des bassins versants.

235 pages en quatre mots

Transférabilité

C'est l'objectif de nos travaux : l'étude de la transférabilité temporelle des paramètres de modèles hydrologiques conceptuels dans un climat en évolution. Cette transférabilité, indicateur de la robustesse des modèles, est étudiée à l'aide de la procédure *GSST (Generalized Split-Sample Test).*

Hétérogénéité

C'est ce qui caractérise les problèmes de robustesse des modèles sur notre échantillon de bassins. Bien que les liens entre ces problèmes et les variations de température de l'air semblent récurrents, les erreurs de modélisation ne peuvent pas toujours être reliées à des variables climatiques où sont parfois corrélées à une variable différente de celle constatée sur le bassin voisin sans que nous réussissions à en expliquer la cause.

Similarité

C'est certainement notre résultat le plus marquant : nous constatons une similarité des comportements relatifs entre des jeux calés sur différentes sous-périodes temporelles, mais aussi entre différents modèles hydrologiques. Ces similarités sont particulièrement visibles dans l'étude des $\triangle\overline{\widehat{Q}}$ au chapitre 4 et des courbes de biais glissants au chapitre 5.

Causalité

C'est le grand champ qui reste à explorer. Si nos analyses ont permis un diagnostic relativement complet de la robustesse temporelle et climatique des modèles hydrologiques conceptuels, elles n'ont pas permis d'expliquer leurs causes réelles. Les difficultés de reproduction temporelle des bilans hydrologiques, même en situation de calage sur toute la chronique, demeurent partiellement incomprises. Qu'une capacité insuffisante des modèles à simuler des dynamiques lentes ou que les erreurs d'estimation de grandeurs telles que l'évapotranspiration soient en cause, l'établissement d'une causalité est essentielle pour espérer apporter des solutions aux problèmes de robustesse mis en avant par le *GSST* dans nos travaux.

Références

Calvin and Hobbes, 16/06/1986, Bill Watterson©

Abebe, N. A., F. L. Ogden, et N. R. Pradhan (2010), Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model : Implications for parameter estimation, *Journal of Hydrology*, 389(3-4), 301–310, doi :10.1016/j.jhydrol.2010.06.007.

AdatpAlp (2011), Climate adaptation and natural hazard management in the alpine space - final report, *Alpine Space Programme & Bavarian State Ministry of the Environment and Public Health, www.adaptalp.org*, *Rapport technique*, pp. 96, Munich, Allemagne.

Allen, R. G., L. S. Pereira, D. Raes, et M. Smith (1998), Crop evapotranspiration - guidelines for computing crop water requirements - FAO irrigation and drainage paper 56, *Food and Agriculture Organization of the United Nations (FAO) - Organisation des Nations Unies pour l'Alimentation et l'Agriculture (ONUAA)*, *Rapport technique*, pp. 15, Rome, Italie.

Anctil, F., C. Perrin, et V. Andréassian (2004), Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models, *Environmental Modelling & Software*, 19(4), 357–368, doi :10.1016/S1364-8152(03)00135-X.

Andréassian, V., et C. Perrin (2012), On the ambiguous interpretation of the turc-budyko nondimensional graph, *Water Resources Research*, 48, W10601, doi :10.1029/2012WR012532.

Andréassian, V., C. Perrin, et C. Michel (2004), Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, *Journal of Hydrology*, 286(1-4), 19–35, doi :10.1016/j.jhydrol.2003.09.030.

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M.-H. Ramos, et A. Valéry (2009), Crash tests for a standardized evaluation of hydrological models, *Hydrology and Earth System Sciences*, 13(10), 1757–1764, doi :10.5194/hess-13-1757-2009.

Andréassian, V., C. Perrin, E. Parent, et A. Bárdossy (2010), The court of miracles of hydrology : can failure stories contribute to hydrological science ?, *Hydrological Sciences Journal*, 55(6), 849–856, doi :10.1080/02626667.2010.506050.

Andréassian, V., N. Le Moine, C. Perrin, M.-H. Ramos, L. Oudin, T. Mathevet, J. Lerat, et L. Berthet (2012), All that glitters is not gold : the case of calibrating hydrological models, *Hydrological Processes*, 26(14), 2206–2210, doi :10.1002/hyp.9264.

Apaydin, H., A. S. Anli, et A. Ozturk (2006), The temporal transferability of calibrated parameters of a hydrological model, *Ecological Modelling*, 195(3-4), 307–317, doi :10.1016/j.ecolmodel.2005.11.032.

Arnell, N. W. (2011), Uncertainty in the relationship between climate forcing and hydrological response in UK catchments, *Hydrology and Earth System Sciences*, 15(3), 897–912, doi : 10.5194/hess-15-897-2011.

Beven, K., et J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249(1–4), 11–29, doi :10.1016/S0022-1694(01)00421-8.

Beven, K. J. (2012), *Rainfall-Runoff Modelling : The Primer. Second edition*, Wiley-Blackwell, pp. 488, ISBN : 047071459X.

Bois, P. (1987), Contrôle de séries chronologiques corrélées par étude du cumul des résidus de la corrélation, *Deuxièmes journées hydrologiques de l'ORSTOM*, Montpellier, France.

Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrology and Earth System Sciences*, 11(2), 703–710, doi :10.5194/hess-11-703-2007.

Bárdossy, A., et S. K. Singh (2008), Robust estimation of hydrological model parameters, *Hydrology and Earth System Sciences*, 12(6), 1273–1283, doi :10.5194/hess-12-1273-2008.

Brigode, P., L. Oudin, et C. Perrin (2013), Hydrological model parameter instability : A source of additional uncertainty in estimating the hydrological impacts of climate change ?, *Journal of Hydrology*, 476, 410–425, doi :10.1016/j.jhydrol.2012.11.012.

Butts, M. B., J. T. Payne, M. Kristensen, et H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *Journal of Hydrology*, 298(1-4), 242–266, doi :10.1016/j.jhydrol.2004.03.042.

Caballero, Y., S. Voirin-Morel, F. Habets, J. Noilhan, P. LeMoigne, A. Lehenaff, et A. Boone (2007), Hydrological sensitivity of the adour-garonne river basin to climate change, *Water Resources Research*, 43, W07448, doi :200710.1029/2005WR004192.

Charbonneau, R., J. Fortin, et G. Morin (1977), The CEQUEAU model : description and examples of its use in problems related to water resource management, *Hydrological Sciences Bulletin*, 22(1), 93–202.

Chiew, F. H. S., P. H. Whetton, T. A. McMahon, et A. B. Pittock (1995), Simulation of the impacts of climate change on runoff and soil moisture in australian catchments, *Journal of Hydrology*, 167(1-4), 121–147, doi :10.1016/0022-1694(94)02649-V.

Chiew, F. H. S., M. C. Peel, et A. W. Western (2002), Application and testing of the simple rainfall-runoff model SIMHYD, *Mathematical Models of Small Watershed Hydrology and Applications*, pp. 335–367, Littleton, Colorado : VP Singh & DK Frevert, Water Resources Publication edition.

Chiew, F. H. S., J. Vaze, N. R. Viney, P. Jordan, J.-M. Perraud, L. Zang, J. Teng, J. Pena Arancibia, R. Morden, A. Freebairn, J. Austin, P. Hill, C. Wiesemfeld, et R. Murphy (2008), Rainfall-runoff modelling across the murray-darling basin. a report to the australian government from the CSIRO murray-darling basin sustainable yields project., *CSIRO*, *Rapport technique*, pp. 70, Canberra, Australie.

Chiew, F. H. S., J. Teng, J. Vaze, D. A. Post, J. M. Perraud, D. G. C. Kirono, et N. R. Viney (2009), Estimating climate change impact on runoff across southeast australia : Method, results, and implications of the modeling method, *Water Resources Research*, 45, W10414, doi :200910.1029/2008WR007338.

Choi, H. T., et K. Beven (2007), Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *Journal of Hydrology*, 332(3-4), 316–336, doi :10.1016/j.jhydrol.2006.07.012.

Coron, L., V. Andréassian, M. Bourqui, C. Perrin, et F. Hendrickx (2011), Pathologies of hydrological models used in changing climatic conditions : a review, *Hydro-climatology : Variability and change (IUGG2011)*, 344, 39–44, IAHS Publications, Melbourne, Australie.

Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, et F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions : An experiment on 216 australian catchments, *Water Resources Research*, 48, W05552, doi :10.1029/2011WR011721.

CSIRO (2012), Climate and water availability in south-eastern australia : A synthesis of findings from phase 2 of the south eastern australian climate initiative (SEACI), *CSIRO*, *Rapport technique*, pp. 41, Canberra, Australie.

de Marsily, G., P. Combes, et P. Goblet (1992), Comment on "Ground-water models cannot be validated", by L.F. konikow and J.D. bredehoeft, *Advances in Water Resources*, 15, 367–369.

de Vos, N. J., T. H. M. Rientjes, et H. V. Gupta (2010), Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering, *Hydrological Processes*, 24, 2840–2850, doi :10.1002/hyp.7698.

Donnelly-Makowecki, L., et R. Moore (1999), Hierarchical testing of three rainfall-runoff models in small forested catchments, *Journal of Hydrology*, 219(3-4), 136–152, doi :10.1016/S0022-1694(99)00056-6.

Ducharne, A., F. Habets, C. Pagé, E. Sauquet, P. Viennot, M. Déqué, S. Gascoin, A. Hachour, E. Martin, L. Oudin, L. Terray, et D. Thiéry (2010), Climate change impacts on water resources and hydrological extremes in northern france, *XVIII Conference on Computational Methods in Water Resources*, pp. 8, Barcelone, Espagne.

Ebtehaj, M., H. Moradkhani, et H. V. Gupta (2010), Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling, *Water Resources Research*, 46, W07515, doi :10.1029/2009WR007981.

Garavaglia, F. (2011), Méthode SCHADEX de prédétermination des crues extrêmes, *Université de Grenoble*, *Thèse de doctorat*, pp. 279, Grenoble, France.

Garçon, R. (1996), Prévision opérationnelle des apports de la durance à serre-ponçon à l'aide du modèle MORDOR. bilan de l'année 1994-1995, *La Houille Blanche*, (5), 71–76, doi :10.1051/lhb/1996056.

Gharari, S., M. Hrachowitz, F. Fenicia, et H. H. G. Savenije (2013), An approach to identify time consistent model parameters : sub-period calibration, *Hydrology and Earth System Sciences*, 17(1), 149–161, doi :10.5194/hess-17-149-2013.

Gottardi, F. (2009), Estimation statistique et réanalyse des précipitations en montagne utilisation d'ébauches par types de temps et assimilation de données d'enneigement application aux grands massifs montagneux français, *INPG*, *Thèse de doctorat*, pp. 284, Grenoble, France.

Gottardi, F., C. Obled, J. Gailhard, et E. Paquet (2012), Statistical reanalysis of precipitation fields based on ground network data and weather patterns : Application over french mountains, *Journal of Hydrology*, 432–433, 154–167, doi :10.1016/j.jhydrol.2012.02.014.

Görgen, K., J. Beersma, H. Buiteveld, G. Brahmer, M. Carambia, O. De Keizer, P. Krahe, E. Nilson, R. Lammersen, C. Perrin, et D. Volken (2010), Assessment of climate change impacts on discharge in the rhine river basin : Results of the RheinBlick2050 project, *KHR-CHR*, *Rapport technique*, pp. 229, Lelystad, Netherlands.

Gupta, H. V., H. Kling, K. K. Yilmaz, et G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria : Implications for improving hydrological modelling, *Journal of Hydrology*, 377(1-2), 80–91, doi :10.1016/j.jhydrol.2009.08.003.

Hartmann, G., et A. Bárdossy (2005), Investigation of the transferability of hydrological models and a method to improve model calibration, *Advances in Geosciences*, 5, 83–87.

Hayami, S. (1951), On the propagation of flood waves, *Bulletins - Disaster Prevention Research Institute, Kyoto University*, 1, 1–16.

He, M., T. S. Hogue, K. J. Franz, S. A. Margulis, et J. A. Vrugt (2011), Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, *Advances in Water Resources*, 34(1), 114–127, doi :10.1016/j.advwatres.2010.10.002.

Herrnegger, M., H.-P. Nachtnebel, et T. Haiden (2012), Evapotranspiration in high alpine catchments – an important part of the water balance !, *Hydrology Research*, 43(4), 460–475, doi :10.2166/nh.2012.132.

IPCC (2007), Climate change 2007 : Synthesis report. contribution of working groups i, II and III to the fourth assessment report of the intergovernmental panel on climate change, *Intergovernmental Panel on Climate Change (IPCC) - Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC)*, *Rapport technique*, pp. 104, Genève, Suisse.

Kay, A., H. Davies, V. Bell, et R. Jones (2009), Comparison of uncertainty sources for climate change impacts : flood frequency in england, *Climatic Change*, 92(1-2), 41–63, doi : 10.1007/s10584-008-9471-4.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, 31(1), 13–24, doi :10.1080/02626668609491024.

Kokkonen, T. S., et A. J. Jakeman (2001), A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications, *Water Resources Research*, 37(9), 2345–2352, doi : 10.1029/2001WR000299.

Le Lay, M., S. Galle, G. M. Saulnier, et I. Braud (2007), Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability : A case study in west africa, *Water Resources Research*, 43, W07420, doi :10.1029/2006WR005257.

Le Moine, N. (2008), Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme desmodèles pluie-débit ?, *UPMC*, *Thèse de doctorat*, pp. 264, Paris, France.

Le Moine, N., et C. Monteil (2012), CEQUEAU - EDF R&D version 5.1.1, note de principe, *EDF R&D*, *Rapport technique*, pp. 33, Chatou, France.

Le Moine, N., V. Andréassian, C. Perrin, et C. Michel (2007), How can rainfall-runoff models handle intercatchment groundwater flows ? theoretical study based on 1040 french catchments, *Water Resources Research*, 43, W06428, doi :200710.1029/2006WR005608.

Lerat, J., V. Andréassian, C. Perrin, J. Vaze, J. M. Perraud, P. Ribstein, et C. Loumagne (2012), Do internal flow measurements improve the calibration of rainfall-runoff models ?, *Water Resources Research*, 48, W02511, doi :10.1029/2010WR010179.

Li, C. Z., L. Zhang, H. Wang, Y. Q. Zhang, F. L. Yu, et D. H. Yan (2012), The transferability of hydrological models under nonstationary climatic conditions, *Hydrology and Earth System Sciences*, 16(4), 1239–1254, doi :10.5194/hess-16-1239-2012.

Lin, Z., et M. B. Beck (2012), Accounting for structural error and uncertainty in a model : An approach based on model parameters as stochastic processes, *Environmental Modelling & Software*, 27–28, 97–111, doi :10.1016/j.envsoft.2011.08.015.

Lobligeois, F. (2010), Etude comparative des modèles d'ETP sur des évènements climatiques extrêmes & analyse des stratégies de prise en compte de l'irrigation dans le modèle hydrologique CEQUEAU, *UMPC Sisyphe - EDF R&D*, *Rapport technique*, pp. 74, Paris, France.

Matalas, N. (2012), Comment on the announced death of stationarity, *Journal of Water Resources Planning and Management*, 138(4), 311–312, doi :10.1061/(ASCE)WR.1943-5452.0000215.

Mathevet, T. (2005), Quels modèles pluie-débit globaux au pas de temps horaire ? développements empiriques et comparaison de modèle sur un large échantillon de bassins versants, *ENGREF*, *Thèse de doctorat*, pp. 354, Paris, France.

Maurer, E. P. (2007), Uncertainty in hydrologic impacts of climate change in the sierra nevada, california, under two emissions scenarios, *Climatic change*, 82(3-4), 309–325, doi :10.1371/journal.pone.0009932.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, et R. Woods (2011), Rainfall uncertainty in hydrological modelling : An evaluation of multiplicative error models, *Journal of Hydrology*, 400, 83–94, doi :10.1016/j.jhydrol.2011.01.026.

Merz, R., J. Parajka, et G. Blöschl (2011), Time stability of catchment model parameters - implications for climate impact analyses, *Water Resources Research*, 47, W02531, doi :10.1029/2010WR009505.

Mezentsev, V. (1955), Du nouveau sur le calcul de l'évaporation totale (yechio raz o rastchetie srednevo summarnovo ispareniia), *Meteorologiya i Gidrologiya (Russian Meteorology and Hydrology)*, 5, 24–26.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, et R. J. Stouffer (2008), Stationarity is dead : Whither water management ?, *Science*, 319(5863), 573–574, doi :10.1126/science.1151915.

Monteith, J. (1965), Evaporation and environment, *Symposia of the Society for Experimental Biology (in The State and Movement of Water in Living Organisms)*, 19, 205–234, Cambridge University Press, Swansea, Royaume-Uni.

Morin, G., J. Fortin, J. Lardeau, et W. Sochanska (1981), Modèle CEQUEAU : Manuel d'utilisation, *Université du Québec, INRS-Eau*, *Rapport technique*, pp. 362, Québec, Canada.

Morton, F. (1983), Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, *Journal of Hydrology*, 66, 1–76, doi :10.1016/0022-1694(83)90177-4.

Mouelhi, S., C. Michel, C. Perrin, et V. Andréassian (2006), Linking stream flow to rainfall at the annual time step : The manabe bucket model revisited, *Journal of Hydrology*, 328(1-2), 283–296, doi :16/j.jhydrol.2005.12.022.

Murphy, C., R. Fealy, R. Charlton, et J. Sweeney (2006), The reliability of an "off-the-shelf" conceptual rainfall runoff model for use in climate impact assessment : uncertainty quantification using latin hypercube sampling, *Area*, 38(1), 65–78, doi :10.1111/j.1475-4762.2006.00656.x.

Nash, J., et J. Sutcliffe (1970), River flow forecasting through conceptual models part i. a discussion of principles, *Journal of Hydrology*, 10(3), 282–290, doi :10.1016/0022-1694(70)90255-6.

Niel, H., J. E. Paturel, et E. Servat (2003), Study of parameter stability of a lumped hydrologic model in a context of climatic variability, *Journal of Hydrology*, 278(1-4), 213–230, doi :10.1016/S0022-1694(03)00158-6.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, et C. Loumagne (2005), Which potential evapotranspiration input for a lumped rainfall–runoff model ? : Part 2-towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, *Journal of Hydrology*, 303(1–4), 290–306, doi :10.1016/j.jhydrol.2004.08.026.

Oudin, L., V. Andréassian, T. Mathevet, C. Perrin, et C. Michel (2006a), Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, *Water Resources Research*, 42, W07410, doi :10.1029/2005WR004636.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, et C. Michel (2006b), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *Journal of Hydrology*, 320(1-2), 62–83, doi :10.1016/j.jhydrol.2005.07.016.

Paquet, E. (2004), Évolution du modèle hydrologique MORDOR : modélisation du stock nival à différentes altitudes, *La Houille Blanche*, (2), 75–82, doi :10.1051/lhb:200402008.

Paquet, E. (2005), Critique et homogénéisation de séries climatologiques. outil KRITIKOR, notice méthodologique, *EDF-DTG*, *Rapport technique*, pp. 26, Grenoble, France.

Perrin, C. (2000), Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative, *INPG*, *Thèse de doctorat*, pp. 291, Paris, France.

Perrin, C., C. Michel, et V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, 279(1-4), 275–289, doi :10.1016/S0022-1694(03)00225-7.

Prudhomme, C., et H. Davies (2009a), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. part 1 : baseline climate, *Climatic Change*, 93(1), 177–195, doi :10.1007/s10584-008-9464-3.

Prudhomme, C., et H. Davies (2009b), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. part 2 : future climate, *Climatic Change*, 93(1), 197–222, doi :10.1007/s10584-008-9461-6.

Quintana-Seguí, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, et S. Morel (2008), Analysis of near-surface atmospheric variables : Validation of the SAFRAN analysis over france, *Journal of Applied Meteorology and Climatology*, 47(1), 97–107, doi :10.1175/2007JAMC1636.1.

R2D2-2050 (2013), Risque, ressource en eau et gestion durable de la durance en 2050, *Irstea, EDF, UPMC, SCP, ACTEON, https ://r2d2-2050.cemagref.fr*, *Projet de recherche*, Lyon, France.

Refsgaard, J. C., et J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resources Research*, 32(7), 2189–2202.

Renard, B. (2008), Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en france, *INPG*, *Thèse de doctorat*, pp. 268, Grenoble, France.

Reusser, D. E., et E. Zehe (2011), Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resources Research*, 47, W07550, doi : 201110.1029/2010WR009946.

Reusser, D. E., T. Blume, B. Schaefli, et E. Zehe (2008), Analysing the temporal dynamics of model performance for hydrological models, *Hydrology and Earth System Sciences*, 5(6), 3169–3211, doi : 10.5194/hessd-5-3169-2008.

RIWER-2030 (2011), Regional climate, water, energy resources and uncertainties from 1960 to 2030, *LTHE, EDF, ANR, www.lthe.fr/RIWER2030*, *Projet de recherche*, Grenoble, France.

Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, et G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition zones during the warm season, *Journal of Geophysical Research*, 115, 21 pp., doi :201010.1029/2009JD012035.

Schneider, C., C. L. R. Laizé, M. C. Acreman, et M. Flörke (2013), How will climate change modify river flow regimes in europe ?, *Hydrology and Earth System Sciences*, 17(1), 325–339, doi :10.5194/hess-17-325-2013.

Seibert, J. (2001), On the need for benchmarks in hydrological modelling, *Hydrological Processes*, 15(6), 1063–1064, doi :10.1002/hyp.446.

Seibert, J. (2003), Reliability of model predictions outside calibration conditions, *Nordic Hydrology*, 34(5), 477–492, doi :10.2166/nh.2003.028.

Seifert, D., T. O. Sonnenborg, J. C. Refsgaard, A. L. Højberg, et L. Troldborg (2012), Assessment of hydrological model predictive ability given multiple conceptual geological models, *Water Resources Research*, 48, W06503, doi :201210.1029/2011WR011149.

Seiller, G., F. Anctil, et C. Perrin (2012), Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, *Hydrology and Earth System Sciences*, 16(4), 1171–1189, doi :10.5194/hess-16-1171-2012.

Singh, R., T. Wagener, K. van Werkhoven, M. E. Mann, et R. Crane (2011), A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate – accounting for changing watershed behavior, *Hydrology and Earth System Sciences*, 15(11), 3591–3603, doi : 10.5194/hess-15-3591-2011.

Steele-Dunne, S., P. Lynch, R. McGrath, T. Semmler, S. Wang, J. Hanafin, et P. Nolan (2008), The impacts of climate change on hydrology in ireland, *Journal of Hydrology*, 356(1-2), 28–45, doi :10.1016/j.jhydrol.2008.03.025.

Tan, K. S., F. H. S. Chiew, R. B. Grayson, P. J. Scanlon, et L. Siriwardena (2005), Calibration of a daily rainfall-runoff model to estimate high daily flows, *MODSIM 2005 International Congress on Modelling and Simulation*, pp. 2960–2966, The Modelling and Simulation Society of Australia and New Zealand, Melbourne, Australie.

Teng, J., J. Vaze, F. H. S. Chiew, B. Wang, et J.-M. Perraud (2012), Estimating the relative uncertainties sourced from GCMs and hydrological models in modelling climate change impact on runoff, *Journal of Hydrometeorology*, 13, 122–139, doi :10.1175/JHM-D-11-058.1.

Thornthwaite, C. W. (1948), An approach toward a rational classification of climate, *Geographical Review*, 38(1), 55–94.

Turc, L. (1954), Le bilan d'eau des sols : relation entre les précipitations, l'évapotranspiration et l'écoulement, *Annales agronomiques*, Série A(5), 491–595.

Turcotte, R., L.-G. Fortin, V. Fortin, J.-P. Fortin, et J.-P. Villeneuve (2007), Operational analysis of the spatial distribution and the temporal evolution of the snowpack water equivalent in southern québec, canada, *Nordic Hydrology*, 38(3), 211, doi :10.2166/nh.2007.009.

USACE (1956), Snow hydrology, *U.S. Army Coprs of Engineers (USACE) North Pacific Division*, *Rapport technique*, pp. 437, Portland, Oregon, USA.

Valéry, A. (2010), Modélisation précipitations-débit sous influence nivale. Élaboration d'un module neige et évaluation sur 380 bassins versants, *AgroParisTech*, *Thèse de doctorat*, pp. 303, Paris, France.

Vaze, J., et J. Teng (2011), Future climate and runoff projections across new south wales, australia : results and practical applications, *Hydrological Processes*, 25(1), 18–35, doi :10.1002/hyp.7812.

Vaze, J., D. A. Post, F. H. S. Chiew, J.-M. Perraud, N. R. Viney, et J. Teng (2010), Climate nonstationarity - validity of calibrated rainfall-runoff models for use in climatic changes studies, *Journal of Hydrology*, 394(3-4), 447–457, doi :10.1016/j.jhydrol.2010.09.018.

Vaze, J., F. H. S. Chiew, J.-M. Perraud, N. R. Viney, D. A. Post, J. Teng, B. Wang, J. Lerat, et M. Goswami (2011), Rainfall-runoff modelling across southeast australia : Datasets, models and results, *Australian Journal of Water Resources*, 14(2), 101–116.

Vicuna, S., et J. A. Dracup (2007), The evolution of climate change impact studies on hydrology and water resources in california, *Climatic Change*, 82, 327–350, doi :10.1007/s10584-006-9207-2.

Vidal, J.-P., et S. Wade (2008), A framework for developing high-resolution multi-model climate projections : 21st century scenarios for the UK, *International journal of climatology*, 28(7), 843–858, doi :10.1002/joc.1593.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, et H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling : dynamic identifiability analysis, *Hydrological Processes*, 17(2), 455–476, doi :10.1002/hyp.1135.

Wilby, R. L. (2005), Uncertainty in water resource model parameters used for climate change impact assessment, *Hydrological Processes*, 19(16), 3201–3219, doi :10.1002/hyp.5819.

Wilby, R. L., et I. Harris (2006), A framework for assessing uncertainties in climate change impacts : Low-flow scenarios for the river thames, UK, *Water Resources Research*, 42, W02419, doi :200610.1029/2005WR004065.

Wriedt, G., et M. Rode (2006), Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH, *Advances in Geosciences*, 9, 145–150.

Xu, C. (1999), Climate change and hydrologic models : A review of existing gaps and recent research developments, *Water Resources Management*, 13(5), 369–382, doi :10.1023/A:1008190900459.

Yapo, P. O., H. V. Gupta, et S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models : sensitivity to calibration data, *Journal of Hydrology*, 181(1-4), 23–48, doi :10.1016/0022-1694(95)02918-4.

Zhang, H., G. H. Huang, D. Wang, et X. Zhang (2011), Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering, *Advances in Water Resources*, 34(10), 1292–1303, doi :10.1016/j.advwatres.2011.06.005.

Zin, I. (2002), Incertitudes et ambiguité dans la modélisation hydrologique : discussion, développements méthodologiques et application à l'hydrologie de crue en ardèche, *INPG*, *Thèse de doctorat*, pp. 200, Grenoble, France.

Annexes

A

Listes des figures et tableaux

A.1. Liste des figures

1.1. Illustration de l'objet d'étude : le bassin versant 8
1.2. Résultats issus de *Vaze et al.* (2010), illustrant le lien entre écarts de précipitations moyennes lors du transfert de paramètres et pertes de performances . 13
1.3. Résultats issus de *Merz et al.* (2011), illustrant la corrélation entre variations climatiques et les valeurs de paramètres d'un modèle hydrologique 14
1.4. Exemple d'une année hydrologique pour un bassin de régime nival 16
1.5. Exemples des régimes pluvial, nivo-pluvial et nival 17
1.6. Composantes principales de la chaîne de modélisation dans les études d'impact du changement climatique en hydrologie 21

2.1. Méthodologie de test de calage-contrôle *SST et DSST* 27
2.2. Méthodologie de test de calage-contrôle généralisée *GSST* 29
2.3. Méthode d'analyse visuelle de la capacité d'extrapolation climatique des paramètres : $\triangle\varepsilon = f(\triangle climat)$ 34
2.4. Exemple de nuages de points utilisés pour étudier les liens potentiels entre erreurs de simulation et conditions de transfert des paramètres 38
2.5. Échantillons de bassins versants français (365) et australiens (204) sélectionnés pour les travaux de thèse 40
2.6. Illustration d'analyses de qualité par l'approche doubles-cumuls débit-débit . 43
2.7. Résumé des caractéristiques des bassins étudiés 45
2.8. Aperçu de la variabilité des *ETPs*, précipitations et écoulements sur les bassins étudiés, à différentes échelles temporelles 47
2.9. Structure et équations du modèle *GR4J* 49
2.10. Structure et équations du modèle *Mordor6* 50
2.11. Structure et équations de la partie production du modèle *SimHyd* 51
2.12. Structure et équations d'une bande d'altitude du module neige *CemaNeige* . 52
2.13. Structure de la partie production du modèle *Cequeau* 53
2.14. Nombre d'années disponibles avec moins de 10% de lacunes et nombre de tests de transfert réalisés entre des sous-périodes indépendantes de 10 ans 57

3.1. Performances moyennes des modèles en calage et en validation sur tout l'échantillon (569 bassins) 65
3.2. Pertes de performances lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle$températures 77
3.3. Erreurs commises lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle$températures 77
3.4. Carte des 120 bassins concernés par des corrélations significatives entre erreurs sur les volumes moyens et $\triangle$températures pour le modèle *GR4J* 79

3.5. Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *l'Allier à Vieille-Brioude* 81
3.6. Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Grande Creuse à Fresselines* 82
3.7. Aperçu des erreurs commises par *Cequeau* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Grande Creuse à Fresselines* 83
3.8. Pertes de performances lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations 86
3.9. Erreurs commises lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations pour GR4J 86
3.10. Erreurs commises lors des transferts entre périodes sur les 48 bassins concernés par des corrélations significatives entre △erreurs et △précipitations 88
3.11. Carte des 48 bassins concernés par des corrélations significatives entre erreurs sur les volumes moyens et △précipitations pour le modèle *GR4J* 88
3.12. Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Wimmera river à Glenorchy* 90
3.13. Aperçu des erreurs commises par *GR4J* (*CemaNeige* calé) lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Durance à Embrun* . . . 91
3.14. Aperçu des erreurs commises par *Mordor* lors des transferts entre périodes en lien avec $\triangle\overline{P}$ pour le bassin de *la Durance à Embrun* 92
3.15. Localisation des 41 bassins où les précipitations solides (estimées) représentent plus de 20% des apports annuels 94
3.16. Erreurs commises par *GR4J-CemaNeige* lors du *GSST* sur les 41 bassins recevant plus de 20% de précipitations annuelles sous forme solide 95
3.17. Écarts de simulation de neige par *GR4J-CemaNeige* lors du *GSST* sur les 41 bassins recevant plus de 20% de précipitations annuelles sous forme solide 96
3.18. Différences des valeurs de paramètres de *CemaNeige* (couplé à *GR4J*) en fonction conditions de calage sur les 41 bassins où $\overline{P}_{solides}/\overline{P}_{totales} > 20\%$ 98
3.19. Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Drance à Reyvroz* 99
3.20. Aperçu des erreurs commises par *GR4J* lors des transferts entre périodes en lien avec $\triangle\overline{T}$ pour le bassin de *la Tet à Serdinya* 100

4.1. Simulation des débits de *la Durance à Embrun* avec la chaîne de référence choisie 111
4.2. Écarts de simulation des débits de *la Durance à Embrun* selon le *GCM* utilisé (toutes choses égales par ailleurs) 112
4.3. Écarts de simulation des débits de *la Durance à Embrun* selon la *MDES* utilisée (toutes choses égales par ailleurs) 113
4.4. Écarts de simulation des débits de *la Durance à Embrun* selon la réalisation analogue de *MDES* utilisée (toutes choses égales par ailleurs) 114

4.5. Écarts de simulation des débits de *la Durance à Embrun* selon le modèle hydrologique utilisé (toutes choses égales par ailleurs) 115
4.6. Écarts de simulation des débits de *la Durance à Embrun* selon la période de calage du modèle hydrologique (toutes choses égales par ailleurs) 115
4.7. Synthèse des sensibilités des débits projetés pour *la Durance à Embrun* au choix effectués sur les différents maillons de la chaîne de modélisation 117
4.8. Écarts sur les $\triangle\overline{\widehat{Q}}$ projetés avec *GR4J* sur *la Durance à Embrun* pour les deux horizons considérés, lorsque chaque maillon est considéré indépendamment . . 120
4.9. Écarts sur les $\triangle\widehat{Q}_{10}$, $\triangle\widehat{Q}_{25}$, $\triangle\widehat{Q}_{75}$ et $\triangle\widehat{Q}_{99}$ projetés avec *GR4J* sur *la Durance à Embrun*, lorsque chaque maillon est considéré indépendamment 121
4.10. Écarts sur les $\triangle\overline{\widehat{Q}}_{saison.}$ projetés avec *GR4J* sur *la Durance à Embrun*, lorsque chaque maillon est considéré indépendamment 121
4.11. Écarts sur les $\triangle\overline{\widehat{Q}}$ projetés avec *Mordor6*, *SimHyd*, *Cequeau* et *Mordor* sur *la Durance à Embrun* lorsque chaque maillon est considéré indépendamment . . 122
4.12. Comparaison des écarts de modules observés et simulés par différents modèles sur les périodes historiques de *la Durance à Embrun* (modèles 1 à 3) 126
4.13. Comparaison des écarts de modules observés et simulés par différents modèles sur les périodes historiques de *la Durance à Embrun* (modèles 4 à 6) 127
4.14. Comparaisons aggrégées des écarts de modules observés et simulés par *GR4J+ CemaNeige* θ_{fixes} et *Cequeau* sur les périodes historiques d'un groupe de 20 bassins 129
4.15. Comparaisons aggrégées des écarts de modules observés et simulés par *GR4J+ CemaNeige* θ_{fixes} sur les périodes historiques de tout l'échantillon (569 bassins) 129

5.1. Comparaison des erreurs observées lors des transferts de *GR4J* avec des périodes de 10 ans (en haut) et 20 ans (en bas) 139
5.2. Comparaison des erreurs observées lors des transferts de *Mordor6* avec des périodes de 10 ans (en haut) et 20 ans (en bas) 139
5.3. Sensibilité des résultats du *GSST* à la qualité des chroniques de débits observées 141
5.4. Aperçu des différences de données entre les bases *SAFRAN* et *SPAZM* sur les 365 bassins français de l'échantillon 145
5.5. Comparaison des pertes de performances lors des transferts entre périodes selon la base de forcages utilisée (SPAZM ou SAFRAN) 147
5.6. Comparaison entre écarts climatiques et écarts temporels pour toutes les situations de transfert testées 150
5.7. Analyse comparative des pertes de performances selon les $\triangle$températures et les écarts temporels subis lors des transferts 152
5.8. Quelques tests de transfert *bootstrap* avec *GR4J* sur les bassins de *la Dourbie à Nant* et *du Lot à Banassac* 154
5.9. Visualisation conjointe des liens entre pertes de performances, $\triangle$températures et écarts temporels subis lors des transferts 156

5.10. Erreurs de volume et $\triangle\theta$ observés pour *GR4J* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$, en haut, et $\triangle$erreurs et $\triangle\overline{P}$, en bas . . . 159
5.11. Erreurs de volume et $\triangle\theta$ observés pour *Mordor6* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$, en haut, et $\triangle$erreurs et $\triangle\overline{P}$, en bas . . . 161
5.12. Erreurs de volume et $\triangle\theta$ observés pour *SimHyd* lors des transferts entre périodes sur 120 bassins et 32 bassins, respectivement concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$, en haut, et $\triangle$erreurs et $\triangle\overline{P}$, en bas . . . 162
5.13. Erreurs de volume et $\triangle\theta$ observés pour *Cequeau* lors des transferts entre périodes sur le bassin *du Lot à Banassac* . . . 165
5.14. Erreurs de volume et $\triangle\theta$ observés pour *Cequeau* lors des transferts entre périodes sur le bassin de *la Grande Creuse à Fresselines* . . . 166
5.15. Biais et $\triangle\theta$ observés pour *GR4J* lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$ . . 169
5.16. Biais et $\triangle\theta$ observés pour *Mordor6* lors des transferts entre périodes sur les 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$ 169
5.17. Décomposition de l'analyse visuelle des biais glissants . . . 171
5.18. Analyse visuelle multi-bassins des biais glissants sur 10 ans . . . 174
5.19. Analyse visuelle multi-modèles des biais glissants sur 10 ans . . . 175
5.20. Distributions de l'écart-type des biais glissants sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$, pour tout l'échantillon (569 bassins) et pour l'échantillon français seulement (365 bassins) 178
5.21. Distributions du critère ρ_i mesurant la similitude des biais glissants entre différents calages temporels des modèles . . . 179
5.22. Distributions du critère $\rho'_{M_2 \sim M_1}$ mesurant la similitude des biais glissants entre entre différents modèles hydrologiques . . . 180

6.1. Pertes de performance lors des transferts entre périodes sur les 120 bassins concernés par des corrléations significatives entre $\triangle$erreurs et $\triangle$température . 190
6.2. Distributions de l'écart-type du biais glissant sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$, pour tout l'échantillon, selon la fonction objectif utilisée pour caler *GR4J* . . . 191
6.3. Impact du retrait de la composante biais du *KGE* sur la robustesse de *GR4J* pour le bassin de *l'Allier à Vieille Brioude* . . . 193
6.4. Impact du retrait de la composante biais du *KGE* sur la robustesse de *GR4J* pour le bassin de *la Grande Creuse à Fresselines* . . . 194
6.5. Schématisation du principe d'un calage climatiquement robuste pour des bassins où les erreurs en transfert sont corrélées avec les écarts de températures . 197
6.6. Schématisation du principe d'un calage visant à réduire les biais moyen terme 200

6.7. Illustration du compromis entre un calage classique ($RMSE[Q]$) et un calage sur l'écart-type des biais moyen terme ($RMSE[biais_{10\,ans}]$) 201
6.8. Réflexions sur les causes possibles d'erreur d'estimation de l'évapotranspiration 203
6.9. Rendement évaporatoire du réservoir sol de *GR4J* : situation originale (continu rouge) et alternatives proposées (pointillés verts) 207
6.10. Variations des paramètres calés de *GR4J* engendrée par l'ajout du paramètre α modifiant la relation ETR∼ETP 208
6.11. Écarts-types du biais glissant sur 10 ans $\sigma\left[\omega_{PT}^{M_X}\right]$ pour différentes configurations sur un groupe de 211 bassins français 210
6.12. Mesures de la similitude des biais glissants obtenus sur 211 bassins français pour différentes configurations de calcul d'évapotranspiration 210
6.13. Exemples des courbes de biais glissants pour différentes configurations de forçages sur *la Grande Creuse à Fresselines* et *la Durance à Embrun* 211
6.14. Estimations de régressions linéaires entre $\triangle\theta$ et $\triangle\overline{T}$ pour certains paramètres de *GR4J* et *Mordor6* 215
6.15. Niveaux de robustesse obtenus pour *GR4J* avec une correction des paramètres selon les conditions de transfert 217
6.16. Niveaux de robustesse obtenus pour *Mordor6* avec une correction des paramètres selon les conditions de transfert 219
6.17. Présentation de la représentation adimensionnelle $\overline{Q}\,/\,\overline{P} = f\left(\overline{P}\,/\,\overline{ETP}\right)$ . . . 222
6.18. Représentation des bilans observés pour les bassins de notre échantillon au sein de l'espace adimensionnel $\overline{Q}\,/\,\overline{P} = f\left(\overline{P}\,/\,\overline{ETP}\right)$ 224
6.19. Comparaison entre différentes bases de forçages au sein de l'espace adimensionnel $\overline{Q}\,/\,\overline{P} = f\left(\overline{P}\,/\,\overline{ETP}\right)$ pour 365 bassins français 225
6.20. Exemples d'analyse des variations temporelles de bilan dans l'espace adimensionnel Q/P=f(P/ETP) et parallèle avec les biais glissants de *GR4J* (ex. 1-3) 227
6.21. Exemples d'analyse des variations temporelles de bilan dans l'espace adimensionnel Q/P=f(P/ETP) et parallèle avec les biais glissants de *GR4J* (ex. 4-6) 229

C.1. Exemple d'une fiche d'analyse de qualité par doubles-cumuls 269
C.2. Illustration d'erreurs détectées par doubles-cumuls : dérive de courbe de tarage 270
C.3. Illustration d'erreurs détectées par doubles-cumuls : données incorrectes . . . 270
C.4. Performances de *GR4J-CemaNeige* en calage et en validation sur l'échantillon français (365 bassins), selon si les paramètres de *CemaNeige* sont calés ou fixés 272
C.5. Part solide des précipitations et altitudes et des groupes de bassins où le calage de *CemaNeige* a une effet négatif/nul/positif sur la robustesse temporelle . . 273

D.1. Résumé des pertes de performances relatives liées au transfert des paramètres 276
D.2. Résumé des caractéristiques de deux sous-groupes de bassins : 120 bassins où $MRC = f(\triangle\overline{T})$ et 48 bassins où $MRC = f(\triangle\overline{P})$ 280

D.3. Résumé des caractéristiques de deux sous-groupes de bassins : 360 bassins où $\triangle(\widehat{\overline{Q}}/\overline{Q}) = f(\triangle\overline{T})$ et 244 bassins où $\triangle(\widehat{\overline{Q}}/\overline{Q}) = f(\triangle\overline{P})$ 281
D.4. Erreurs de volumes obtenues avec *GR4J* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures 282
D.5. Erreurs de volumes obtenues avec *Mordor6* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures 283
D.6. Erreurs de volumes obtenues avec *SimHyd* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures 283

F.1. Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* avec la chaîne de référence choisie 307
F.2. Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon le *GCM* utilisé (toutes choses égales par ailleurs) 307
F.3. Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon la *MDES* utilisée (toutes choses égales par ailleurs) 308
F.4. Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon la réalisation analogue utilisée (toutes choses égales par ailleurs) 308
F.5. Écarts de simulation des débits de *la Durance à Embrun* selon la réalisation analogue utilisée pour la *MDES* (toutes choses égales par ailleurs) 309
F.6. Écarts de simulation des débits de *la Durance à Embrun* selon la période de calage de chaque modèle hydrologique (toutes choses égales par ailleurs) . . . 310

G.1. Méthode de sélection des années pour l'approche *bootstrap* 313
G.2. Illustration d'un test de bootstrap sur le bassin de *la Dourbie à Nant* 314
G.3. Biais de *GR4J* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{T}$ 315
G.4. Biais de *GR4J* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 316
G.5. Biais de *GR4J* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 317
G.6. Biais de *Mordor6* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{T}$ 318
G.7. Biais de *Mordor6* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 319
G.8. Biais de *Mordor6* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 320
G.9. Biais de *SimHyd* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{T}$ 321
G.10. Biais de *SimHyd* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 322

G.11. Biais de *SimHyd* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ 323
G.12. Erreurs commises par *GR4J* lors des transferts entre périodes sur 120 bassins concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$ 324
G.13. $\triangle\theta$ obtenus pour *GR4J* lors du test de fixation d'un paramètre pour un *GSST* sur les 120 bassins où △erreurs=f($\triangle\overline{T}$) 325
G.14. Erreurs commises par *Mordor6* lors des transferts entre périodes sur 120 bassins concernés par des corrélations significatives entre △erreurs et $\triangle\overline{T}$ 325
G.15. $\triangle\theta$ obtenus pour *Mordor6* lors du test de fixation d'un paramètre pour un *GSST* sur les 120 bassins où △erreurs=f($\triangle\overline{T}$) 326
G.16. Exemples de représentation côte à côte des analyses par nuage de points et des courbes de biais glissants 329

H.1. Schématisation du principe d'un calage climatiquement robuste pour des bassins où les erreurs en transfert sont corrélées avec les écarts de températures . 333
H.2. Illustration du compromis entre qualité des simulations (μ) et robustesse (σ) sur une gamme élargie de conditions climatiques 336
H.3. Exemples de résultats de calages robustes de *GR4J* sur le bassin de *l'Allier à Vieille-Brioude* 339
H.4. Exemples de résultats de calages robustes de *GR4J* sur le bassin de *la Grande Creuse à Fresselines* 340
H.5. Résumé des caractéristiques des deux sous-groupes de 60 bassins utilisés dans le test de correction des paramètres de *GR4J* et *Mordor6* 343
H.6. Pertes de performance et $\triangle\theta$ obtenus pour *GR4J* lors du test de correction des paramètres selon les conditions de transfert 344
H.7. Pertes de performance et $\triangle\theta$ obtenus pour *Mordor6* lors du test de correction des paramètres selon les conditions de transfert 345

A.2. Liste des tableaux

2.1. Liste des paramètres du modèle *GR4J* 48
2.2. Liste des paramètres du modèle *Mordor6* 50
2.3. Liste des paramètres du modèle *SimHyd* 51
2.4. Liste des paramètres du modèle *Cequeau* 54
2.5. Résumé des modèles et bassins soumis à la procédure de test *GSST* 55
2.6. Variables utilisées pour caractériser les périodes dans la recherche des relations $\triangle erreurs = f(\triangle caract.)$ 59

3.1. Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *GR4J* 71
3.2. Taux de recouvrement entre les sous-groupes utilisés pour les analyses *GSST* en fonction des critères de sélection des bassins 74
3.3. Taux de corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$ pour *GR4J* 75
3.4. Taux de corrélations significatives entre $\triangle$erreurs et $\triangle\overline{P}$ pour *GR4J* 84

D.1. Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *GR4J* 277
D.2. Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *Mordor6* 278
D.3. Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *SimHyd* 279

B

Pathologies of hydrological models used in changing climatic conditions: a review

Éléments de contexte liés à cette publication

Article court publié en lien avec mon exposé oral lors de la conférence de l'*International Union of Geodesy and Geophysics (IUGG)*, en juillet 2011 à Melbourne (Australie). Il s'agit d'un article de discussion autour des problèmes pouvant apparaître lors de l'utilisation de modèles hydrologiques dans des conditions différentes du calage.

Référence complète

Coron, L., V. Andréassian, M. Bourqui, C. Perrin, et F. Hendrickx (2011), Pathologies of hydrological models used in changing climatic conditions : a review, dans *Hydro-climatology : Variability and change (IUGG 2011)*, vol. 344, 39–44, IAHS Publ., Melbourne, Australia.

Hydro-climatology: Variability and change (Proceedings of symposium J-H02 held during IUGG2011 in Melbourne, Australia, July 2011) (IAHS Publ. 344, 2011).

Pathologies of hydrological models used in changing climatic conditions: a review

L. CORON[1,2], V. ANDRÉASSIAN[2], M. BOURQUI[1], C. PERRIN[2], F. HENDRICKX[1]

1 *EDR R&D, Chatou, France*
laurent.coron@edf.fr

2 *Cemagref, Antony, France*

Abstract Characterizing the impact of climate change on hydrology is not as simple as feeding a previously calibrated hydrological model with future climate scenarios. Nevertheless, hydrological modelling is often considered as a small contributor to the overall uncertainty in climate change impact studies. Running a model under conditions that can be significantly different from those used for calibration raises questions relative to the actual extrapolation capacity of the model. As hydrological models (as complex as they may be) are always a simplification of reality, they can never fully integrate all aspects of the rainfall–runoff relationship. Consequently, we prefer to consider them as patients that can certainly be in good health in average conditions, but may also be affected with pathologies when exposed to unusual conditions (namely conditions they have not been properly trained or structured for). Focusing on the robustness issues linked with non-stationary climatic conditions, this paper reviews some of the typical pathologies rainfall–runoff models can suffer from when asked to predict discharges under climate conditions different from the calibration ones.

Key words hydrological model; climate change; model calibration; parameter stability

INTRODUCTION

Characterizing the impact of climate change on hydrology is an increasing field of study throughout the world. However, this task is sensitive to the numerous uncertainties induced by the complexity of the modelling chain (emissions scenarios → global circulation models → downscaling techniques → rainfall–runoff models). Several authors agree that the major sources of uncertainty are the first steps of the modelling chain (mainly emissions scenarios and climate models) and that hydrological modelling is a lower contributor to total uncertainty (Wilby, 2005; Wilby & Harris, 2006; Kay *et al*., 2009; Prudhomme & Davies, 2009). Whether or not progress is made by climate modellers to improve the estimation of future climatic forcings, it is our responsibility as hydrologists to provide trustworthy simulations when a rainfall–runoff (RR) model is run under conditions that may be significantly different to those used for calibration (e.g. future *vs*. current conditions). Indeed, many unknowns still remain concerning the actual climate extrapolation capacity of hydrological models. In this paper, we review the main robustness issues of RR models which are of importance when models are used in a climate change context.

The ideal case: the healthy model

We will not discuss in detail here how an ideal model should look like (number of modules, stores, spatial representation of inputs, etc.). Let us just mention that such a model should be able to integrate correctly the various aspects of the rainfall–runoff relationship, and that it should then be able to simulate correct discharges when provided with correct inputs. Within the context of this article, one should insist on this latter aspect by adding that the model performances should remain good at the application stage, however climatically-different the simulation period may be.

The hard reality: models are full of simplifications that make them dependent on the calibration period

Due to the lack of knowledge on the true functioning of the hydrological system, the lack of data for model construction and the complexity of measuring each process within the hydrological sphere, a RR model incorporates large simplifications (Murphy *et al*., 2006). All hydrological models remain at some point conceptual and empirical. Thus, their parameters require manual or

Copyright © 2011 IAHS Press

automatic calibration. One would ideally like to measure these parameters in the field, but due to the model conceptualisation, direct use of field parameter estimations are rarely possible in practice and calibration remains necessary to reduce bias (Merz *et al.*, 2011). Moreover, whichever approach is used, the estimated parameters may not be the "true" representative parameters of the processes in the catchment, leading to uncertainties in the prediction of the model variable (Abebe *et al*., 2010). As will be discussed below, model parameters may indeed be affected to various extents by various elements during calibration, which may divert them from their correct values, i.e. those values that would be optimal over the long term.

A REVIEW OF COMMON PATHOLOGIES

Dependency of model parameters on the inputs quality and availability

Quite logically, the quality of a RR model is highly dependent on the inputs it was fed with during calibration. Therefore, incorrect estimation of the input or a too short calibration period may affect the model parameter determination and thus bias simulations.

Several studies have assessed the impact of input quality on the optimal parameter set obtained from calibration. Among them, one can mention the work of Andréassian *et al*. (2001, 2004) and Oudin *et al*. (2006). They studied the impact that random and systematic errors in rainfall and potential evapotranspiration (PE) estimates can have on model performance and parameter values for simple RR models. They showed how the models can sometimes use their free parameters to smooth input errors, without decreasing the simulation's quality. Such conclusions are in agreement with the findings of Kokkonen & Jakeman (2001) who worked on the impact of PE estimates on model performance.

Apart from errors in input rainfall depths, several works were made on the effect of incorrect estimation of spatial variability (e.g. poor coverage of the basin with a gauging network), as reviewed by Brath *et al*. (2004). These authors also carried out some tests on the influence of reducing the rainfall gauging network for a distributed model. Interestingly, they found that model performance did not noticeably decrease when fed with a spatially uniform rainfall, as long as the network size remained sufficient to measure the correct amount of overall rainfall. These results contrast with the previous findings of Chaubey *et al*. (1999) who observed a decrease in performance of the Agricultural Non-Point Source Pollution model (AGNPS) and a higher variability in the estimated parameters when the spatial variability of rainfall was deteriorated.

The dependency of model parameters on input quality may also exist for inputs other than climate forcings, as illustrated by Apaydin *et al*. (2006). These authors studied the transferability of the parameters of the SLURP distributed model between two temporally spaced periods. They observed a greater performance loss when transferring model parameters from the oldest to the most recent period than the contrary. They attributed this finding to the wider range of precipitation, but also to the better distribution of land cover data in the recent period. This yielded a more accurate estimation of the model parameters.

Finally, the length of the calibration period may significantly affect model calibration if it is too short. A large number of studies were carried out on this topic and were reviewed by Perrin *et al.* (2007). The general rule about calibration length is to have climatic and flow conditions sufficiently diverse to give a representative picture of their natural variability, thus allowing for an exhaustive activation of the hydrological processes at work in the basin. Beyond this statement, very few recommendations exist on the minimum length required for calibration. Indeed, depending on the studies and models used, this length may vary from two to ten years (Yapo *et al*., 1996; Anctil *et al*., 2004; Brath *et al*., 2004; Perrin *et al.*, 2007).

Dependency of model parameters on the climate of the calibration period

Once the free parameters have been calibrated, the model structure should ideally provide a good simulation of the rainfall–runoff relationship. Therefore, it should be able to simulate discharges

without significant performance loss, when fed with other climate inputs. In reality, model parameters may sometimes be dependent on the climate they have experienced during the calibration period. However, one must say there is currently no consensus in the literature on this dependency or on the impact it may have on the model simulations.

Merz *et al*. (2011) calibrated the semi-distributed HBV model on a series of 5-year periods between 1976 and 2006 for 273 Austrian catchments. They showed how the parameters representing snow and soil moisture processes could vary depending on the calibration period. They established a link between these variations and climatic characteristics such as air temperature and potential evapotranspiration. Similarly, Wagener *et al*. (2003) found in their test using the DYNIA approach (see next section) that the model parameter that controls rapid groundwater recharge or rapid surface runoff was dependent on the calibration climate. Furthermore, they illustrated how this parameter had two optima both with relatively high identifiability: small values were required during summer periods whereas high values were needed during storm events.

With regard to these findings, questions can be raised on the validity of using a model to simulate discharges under climatic conditions that are different from the calibration ones. Several authors investigated the topic carrying out various tests derived from the differential split-sample test proposed by Klemeš (1986). One can mention the work of Wilby (2005) on uncertainties in climate change impact studies. He found that the projections' uncertainty due to the choice of the calibration period could be of the same order as the uncertainty due to greenhouse gas emission scenarios. He concluded that the transferability of model parameters was dependent on the representativeness of the calibration period. Also working on this issue of climatic parameter transferability, Vaze *et al*. (2010) calibrated four commonly-used models (SIMHYD, Sacramento, MARG, IHACRES) on 61 catchments in southeast Australia over the wettest and driest 10, 20, 30 and 40 years and simulated the other periods. They illustrated how the performance decrease and bias increase between calibration and validation could be related to the difference in annual rainfall. They concluded that models calibrated on average or wet conditions had more difficulty in simulating dry periods than the other way round. Finally, one can mention the study by Choi & Beven (2007) who used the Generalized Likelihood Uncertainty Estimator (GLUE) approach to evaluate TOPMODEL over one basin in South Korea. After sampling the time series into several clusters according to their hydrological similarities, they found that parameters that were optimal for some clusters were not convenient for use on others (particularly the ones obtained on the dry clusters, which contrasts with the findings of Vaze *et al*. (2010)).

However, some authors have drawn conclusions regarding the ability of hydrological models calibrated in specific climatic conditions to be used in others. This is the case for Chiew *et al*. (2009) who calibrated the model SIMHYD over a long period and used it to simulate discharges on particularly wet and dry periods. The authors concluded on the suitability of RR models for climate change impact studies when they are calibrated against a sufficiently long period. Clear decreases of model performance were however observed in some cases between calibration and simulations and the authors considered the subject as worthy of further investigation. One can also mention the case of Niel *et al*. (2003) who studied 17 African catchments and noted time stable parameters for about two thirds of the basins and found no obvious link between parameter values and climate for the last third, although rainfall and runoff significantly decreased over the years.

Although no general consensus has been reached yet and despite the differences in model sensitivity, it seems that parameters cannot directly be transferred from one period to another without paying attention to the climatic differences. Obviously this applies to all climate change impact studies.

Low identifiability of parameter values

The identifiability level of a parameter expresses how well the parameter is defined within a model structure (Abebe *et al*., 2010). This level is high if changes in the parameter value have significant effects on the simulated discharges and thus on the calibration criteria, allowing an accurate

estimation of the parameter correct value. Conversely, this level can be low if it has only little effect on the discharges. Potentially, a random value could then be considered as optimal. Methods to analyse parameters identifiability are usually based on a Monte-Carlo approach. Wilby (2005) carried out such an analysis on the CATCHMOD model by testing random parameter sets and looking at the evolution of the simulation's quality relative to individual parameter values. Another method proposed to investigate this question is the Dynamic Identifiability Analysis (DYNIA) of Wagener *et al*. (2003) that aims to identify the most informative parts of the hydrograph for the calibration of each model parameter. Applying this method on a five-parameter RR model, the authors showed that some parts of the hydrograph (recession limbs, rainy periods, etc.) contain more information for the identification of some parameters.

Low identifiability may happen in the case of structural problems in the model, when a parameter does not have a precise role in the model functioning (in spite of the modeller's intention) or when several parameters interact and compensate each other (Wilby, 2005; Abebe *et al*., 2010). However, low identifiability issues are generally not independent from the two "pathologies" mentioned above. Indeed, insufficient input quality or climate variability (linked with calibration length) will induce problems on the determination of the model parameters.

The low identifiability of a parameter may or may not have an impact on model performance: it will have no impact if the process represented by the parameter is insignificant in the case study (e.g. the parameters of a snow module on a basin without snow influence). But if the process in question is significant in some periods and not in others, then it may have a serious impact, because the choice of the calibration period will influence the correctness of the parameter estimate. Wilby (2005) found that many parameters of the CATCHMOD model had a low identifiability on the Thames basin. However, some of them showed better identifiability when calibrated on wet periods. Abebe *et al*. (2010) applied DYNIA on the HBV model. They also found that identifiability was higher on wet periods than on dry ones. Using the same methodology on the WaSiM-ETH model, Wriedt & Rode (2006) illustrated how the snow melt runoff parameter was only identifiable during winter runoff and how low flow conditions were not suitable to calibrate parameters controlling fast runoff processes.

One could argue that the above results are trivial and that nothing will be problematic as long as the calibration period is rich enough in terms of climate variability. However, this is not completely true: hydrologists may face cases where a process that was insignificant during the calibration period (in spite of its length), becomes relevant in the period of interest for simulation. This is particularly true in climate change impact studies. For example, if groundwater exchanges between catchments have a visible impact on discharges only during extended periods of drought, they will not be properly accounted for in a model that was calibrated on a period where droughts were not severe enough. Similar examples could be made with glaciers melt processes in mountainous areas or regarding the role of evapotranspiration in the water balance.

DISCUSSION

The various pathologies listed in this paper cause real problems to hydrologists but should not be seen as fatal. Some preventive and/or curative measures may exist, or they could be found, provided that the problem source is clearly identified first. Following Andréassian *et al*. (2001), we propose to classify model parameters into four categories (considering that calibrations are always made with sufficient input data, i.e. several years):

– **Category A** The parameter remains stable over time (it is not affected by changes in the climate of the calibration period or variance in the quality of inputs estimates). An example of such a parameter can be found in the simplest models, where one unique parameter controls the time shift between rainfall and flow (see Oudin *et al*., 2006).

– **Category B** The parameter is affected by wrong estimates of the inputs from the calibration period. However, it converges towards a single value when the input quality is improved.

Examples of such parameters are the evapotranspiration module parameters, the rainfall multiplication factor or the exchange with groundwater if it is the only way to adjust the water balance (see Andréassian *et al.*, 2001)

– **Category C** The parameter is affected by the climatic properties of the calibration period. This may happen either because the level of identifiability varies depending on the climate, or because the model parameters depend on the climate of the calibration period due to model conceptualization deficiency (Wriedt & Rode, 2006; Abebe *et al*., 2010; Merz *et al*., 2011).

– **Category D** The parameter has an erratic behaviour with no apparent link to the inputs. This reveals the poor adaptation of the model for the case study. It may happen if the process represented by the parameter is not significant, or if the model shows conceptual problems inducing equifinality through parameters compensation (Mo *et al*., 2006; Abebe *et al*., 2010). Parameters of this category are clearly undesirable in a hydrological model.

Parameters from category A are the ideal ones for modellers since they are easy to calibrate and can be globally trusted whatever the future conditions of model use.

Parameters from category B may be the source of problems if they are transferred between periods having different inputs quality/availability levels (e.g. the measurements network has evolved and provide more accurate inputs). However, the problem can be addressed by working on the input quality and choosing cautiously the calibration period to ensure that the model parameters do not compensate for incorrect inputs estimates. This may lead to the use of only the most recent years, as suggested by Apaydin *et al*. (2006).

Parameters from categories C and D are the most problematic, particularly in the case of climate change impact studies where the model using conditions are known to be different from the calibration ones. However, some sub-distinctions can be made:

1. The parameter falls in category C or D because the process it represents is not significant for the available time series, although the model structure is parsimonious enough to ensure the need for the parameter. If possible, one should first try to increase the length of the calibration period (and thus the climate variability), which may add periods where the process is significant. If all available data are already used, then the problem can be solved either by setting the parameter to an *a priori* coherent value or simply by inactivating the related part of the model (if the process in question is believed to remain insignificant).
2. The parameter falls in category C because it has a clear dependence on the climatic conditions of the calibration period, although its identifiability is high. One solution is to extract from the historical data the sub-periods that are climatically closest to the test period, and calibrate the model on those only. However, this remains problematic if the test conditions are different from all the available data, as is often the case in climate change impact studies. In this case, a second solution could be to use the established relation between parameter values and climate to extrapolate the parameter values corresponding to the future climate. However, this option is qualified by Merz *et al*. (2011) as *inelegant* since it does not respect the usual philosophy of modelling which is to have time invariant parameters that are able to work correctly under time variant conditions (such as the rainfall or the air temperature).
3) For all the other situations, the impossibility of determining the parameter's true value is closely linked with the inappropriate conceptualization of the model. Therefore, no simple solution exists because modellers may be reluctant to change the structure of a model they have been using for years or decades.

CONCLUSIONS

All hydrological models remain simplifications of the real world and incorporate parameters that need to be calibrated to some extent. However, these parameters may sometimes be dependent on

the calibration conditions and therefore not be appropriate for use under other conditions. Making a parallel with the medical domain, we reviewed here some of the "pathologies" a model may suffer from due to the parameters' dependency on the calibration conditions (e.g. input quality, average climate, diversity of hydrological processes, etc.). Although these findings are highly dependent on the model and given case study, they raise questions regarding the validity of transferring model parameters from one period to another temporarily spaced and/or climatically different. This is of particular concern for climate change studies where the model application conditions are known to be different from the calibration ones.

REFERENCES

Abebe, N. A., Ogden, F. L. & Pradhan, N. R. (2010) Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *J. Hydrol.* **389**(3-4), 301–310.

Anctil, F., Perrin, C. & Andréassian, V. (2004) Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall–runoff forecasting models. *Environ. Model. Softw.* **19**(4), 357–368.

Andréassian, V., Perrin, C. & Michel, C. (2004) Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *J. Hydrol.* **286**(1-4), 19–35.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. (2001) Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250**(1-4), 206–223.

Apaydin, H., Anli, A. S. & Ozturk, A. (2006) The temporal transferability of calibrated parameters of a hydrological model. *Ecological Modelling* **195**(3-4), 307–317.

Brath, A., Montanari, A. & Toth, E. (2004) Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *J. Hydrol.* **291**(3-4), 232–253.

Chaubey, I., Haan, C. T., Grunwald, S. & Salisbury, J. M. (1999) Uncertainty in the model parameters due to spatial variability of rainfall. *J. Hydrol.* **220**(1-2), 48–61.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C. & Viney, N. R. (2009) Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method. *Water Resour. Res.* **45**, W10414.

Choi, H. T. & Beven, K. (2007) Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *J. Hydrol.* **332**(3-4), 316–336.

Kay, A., Davies, H., Bell, V. & Jones, R. (2009) Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change* **92**(1-2), 41–63.

Klemeš, V. (1986) Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31**(1), 13–24.

Kokkonen, T. S. & Jakeman, A. J. (2001) A comparison of metric and conceptual approaches in rainfall–runoff modeling and its implications. *Water Resour. Res.* **37**(9), 2345–2352.

Merz, R., Parajka, J. & Blöschl, G. (2011) Time stability of catchment model parameters – implications for climate impact analyses. *Water Resour. Res.* **47**, W02531.

Mo, X., Pappenberger, F., Beven, K. J., Liu, S., De Roo, A. & Lin, Z. (2006) Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model. *Hydrol. Sci. J.* **51**(1), 45–65.

Murphy, C., Fealy, R., Charlton, R. & Sweeney, J. (2006) The reliability of an 'off-the-shelf' conceptual rainfall runoff model for use in climate impact assessment: uncertainty quantification using Latin hypercube sampling. *Area* **38**(1), 65–78.

Niel, H., Paturel, J. E. & Servat, E. (2003) Study of parameter stability of a lumped hydrologic model in a context of climatic variability. *J. Hydrol.* **278**(1-4), 213–230.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. & Michel, C. (2006) Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *J. Hydrol.* **320**(1-2), 62–83.

Perrin, C., Oudin, L., Andréassian, V., Rojas-Serna, C., Michel, C. & Mathevet, T. (2007) Impact of limited streamflow data on the efficiency and the parameters of rainfall–runoff models. *Hydrol. Sci. J.* **52**(1), 131.

Prudhomme, C. & Davies, H. (2009) Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. *Climatic Change* **93**(1), 177–195.

Vaze, J., Post, D. A., Chiew, F., Perraud, J. M., Viney, N. & Teng, J. (2010) Climate nonstationarity – Validity of calibrated rainfall–runoff models for use in climatic changes studies. *J. Hydrol.* **394**(3-4), 447–457.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. & Gupta, H. V. (2003) Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. *Hydrol. Processes* **17**(2), 455–476.

Wilby, R. L. (2005) Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrol. Processes* **19**(16), 3201–3219.

Wilby, R. L. & Harris, I. (2006) A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK. *Water Resour. Res.* **42**, W02419.

Wriedt, G. & Rode, M. (2006) Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. *Adv. Geosci.* **9**, 145–150.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall–runoff models: sensitivity to calibration data. *J. Hydrol.* **181**(1-4), 23–48.

C

Compléments sur les données et les modèles utilisés

C.1. Compléments sur le contrôle de la qualité des débits

Cette section présente quelques illustrations complémentaires sur la critique des données de débit par une approche doubles-cumuls (cf. section 2.2.2, page 42). Cette approche s'apparente très largement des travaux de *Bois* (1987) ou de *Paquet* (2005) sur la détection par analyse des résidus d'erreurs cumulés.

Sa mise en place se fait de la manière suivante pour chaque bassin :

- identification des trois plus proches voisins (distance selon x et y seulement),
- détermination de la plage de données commune entre les trois bassins,
- construction de doubles-cumuls entre le bassin étudié et chacun des voisins,
  - nous construisons des cumuls débit-débit et pluie-pluie sur les valeurs journalières,
  - afin d'assurer la comparaison, nous ne considérons pas les pas de temps pour lesquels des données sont lacunaires pour un bassin (ou plus),
  - nous calculons alors les résidus entre la courbe de double cumul et la droite reliant le premier et dernier point,
- analyse visuelle pour détecter les différents cas de figure, parmi lesquelles nous pouvons peut citer :
  - la présence de données incorrectes qui devraient être déclarées lacunaires,
  - la présence d'interpolations linéaires importantes (perturbant les bilans de volumes dans les doubles cumuls),
  - l'existence d'une modification de la courbe de tarage entraînant des différences de volumes estimés,
  - la présence d'un barrage sur tout ou partie de la série qui assure un débit minimal ou écrête les crues et modifie le phasage des écoulements,
  - la présence d'une nappe importante ou d'un barrage qui gouverne fortement le débit de base.

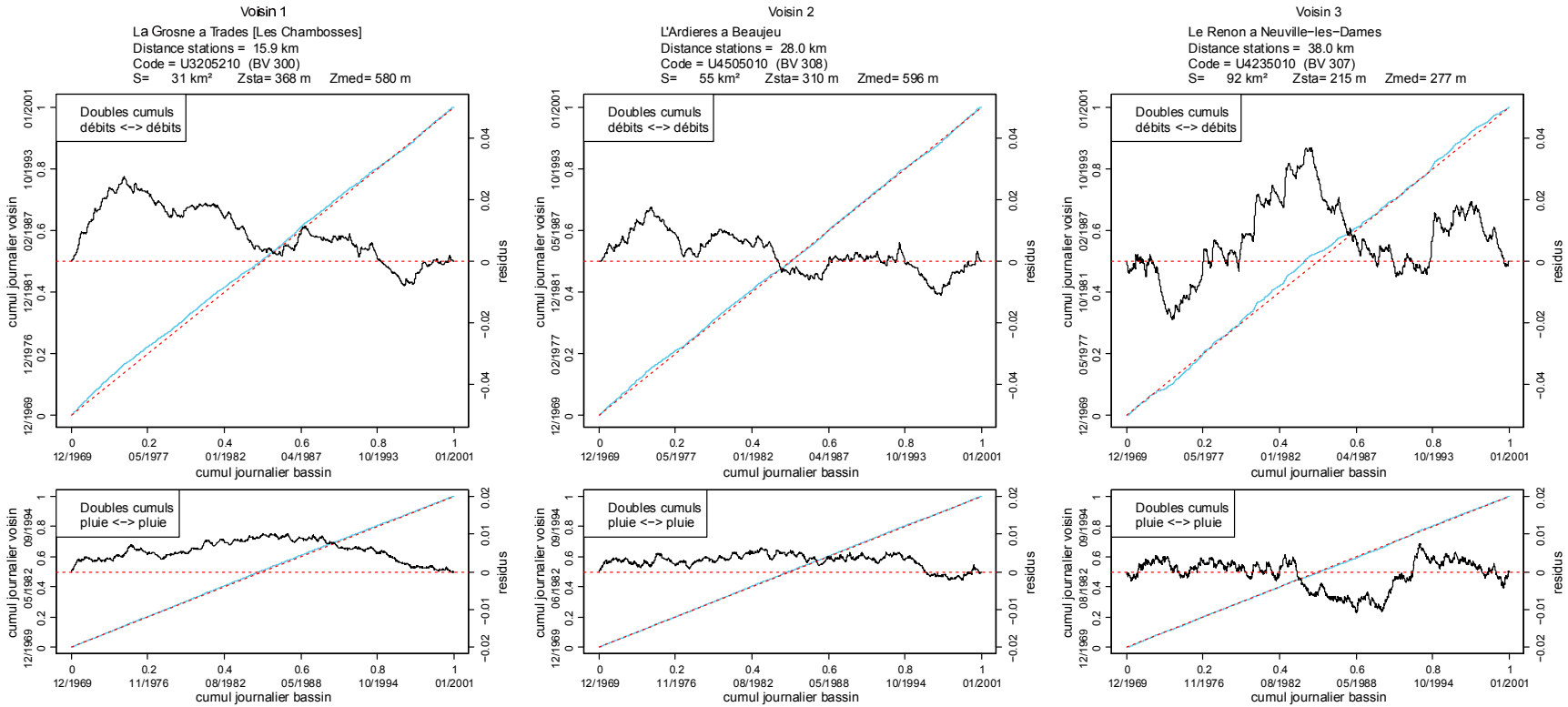


FIGURE C.1.: Exemple d'une fiche d'analyse de qualité par doubles-cumuls
(aucune erreur flagrante détectée ici)

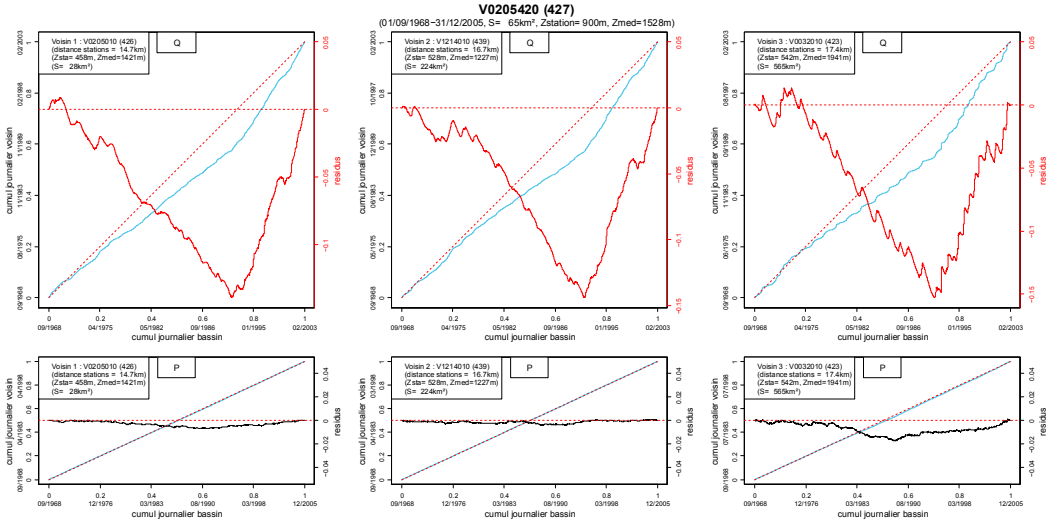


FIGURE C.2.: Illustration d'erreurs détectées par doubles-cumuls : dérive de courbe de tarage (la droite de doubles-cumuls présente deux phases distinctes de pentes différentes)

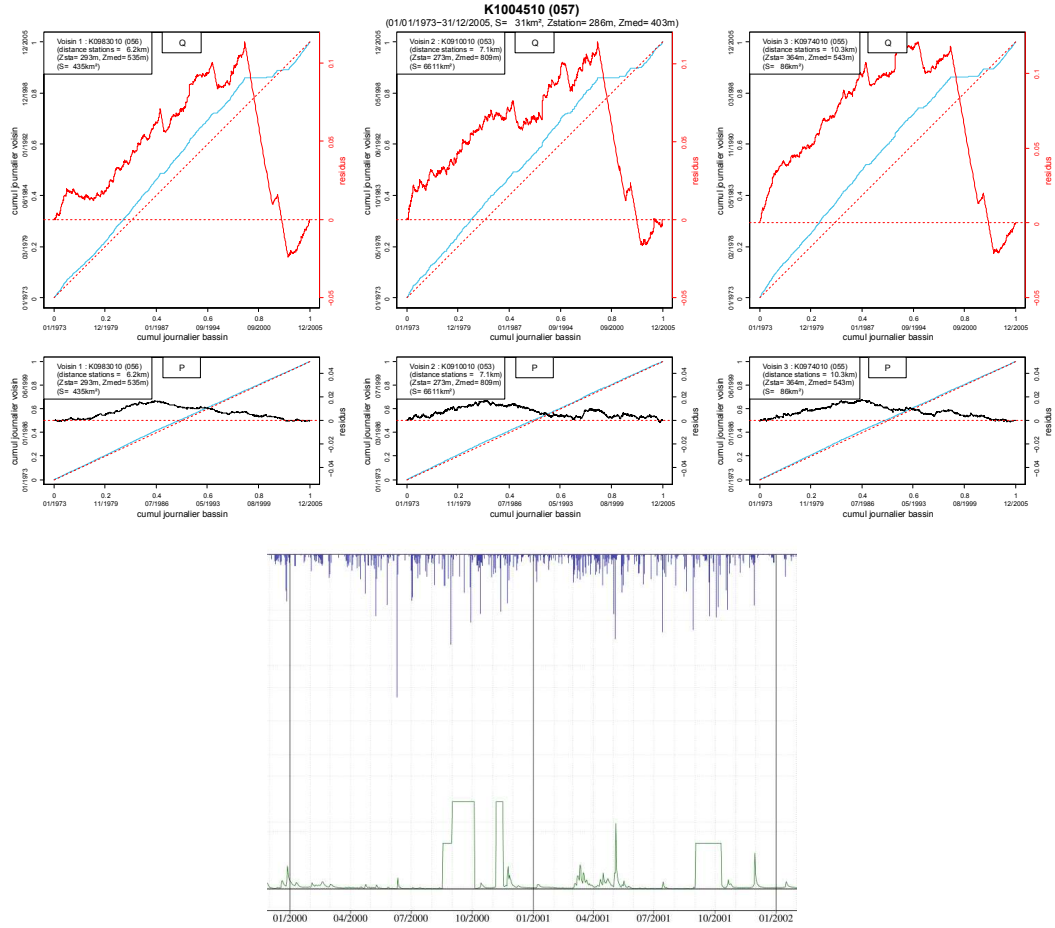


FIGURE C.3.: Illustration d'erreurs détectées par doubles-cumuls : données incorrectes (des section horizontales ou verticales sont visibles sur la droite de doubles-cumuls, l'étude de la chronique permet de retrouver les plages de données incorrectes correspondantes)

C.2. Test de la nécessité de caler le module *CemaNeige*

Les éléments présentés ici apportent des compléments sur le choix de ne pas utiliser le module *CemaNeige* dans sa version calée lorsque celui-ci est couplé aux modèles *GR4J*, Mordor6 et *SimHyd*, pour conduire la procédure *GSST* sur notre échantillon de bassins français.

Contexte

CemaNeige est un module neige conceptuel développé par *Valéry* (2010), pouvant se greffer au-dessus de n'importe quel modèle hydrologique. Lorsqu'il est couplé avec un modèle global, il s'utilise sur des bandes d'altitude de surfaces égales. Dans ce cas, les sorties sont ensuite agglomérées pour former une lame d'eau de bassin qui somme les précipitations liquides et la lame de fonte. Il comporte deux paramètres à caler : (1) un facteur de fonte degré-jour, (2) un coefficient de pondération de l'état thermique du manteau. Les détails du fonctionnement de ce module sont présentés dans la section 2.3.4, page 52.

Nos travaux sont construits autour de tests de transférabilité temporelle et climatique de paramètres (calage dans certaines conditions et utilisation dans d'autres conditions). Les paramètres du module neige sont donc concernés par cette étude, tout autant que ceux du modèle hydrologique. Or, *Valéry* (2010) montra qu'il était possible de fixer les deux paramètres de *CemaNeige*, tout en conservant des performances acceptables en moyenne sur son échantillon de tests. Nous avons donc la possibilité d'étudier deux configurations dans la cadre du *GSST* : une première où tous les paramètres seraient calés (modèle hydrologique et module neige) et une seconde où les paramètres du module neige seraient maintenus à des valeurs fixes. Étudier la perturbation causée par le transfert des paramètres neige entre des périodes contrastées est très intéressant. Cela ne constitue néanmoins qu'une partie de nos travaux car nous souhaitons aussi pouvoir évaluer la transférabilité des paramètres du modèle hydrologique seulement. Une approche type *GSST* dans laquelle les paramètres de *CemaNeige* sont maintenus fixes a l'avantage de permettre cette évaluation.

Performances du couple *GR4J-CemaNeige* pour les deux configurations

Les performances associées à ces deux configurations (paramètres du module neige calés ou fixés) sont comparées sur l'ensemble des situations rencontrées sur notre échantillon de 365 bassins français. Celles-ci sont illustrées sur la figure C.4 qui présente les performances en calage et validation avec un zoom sur la différence entre ces configurations en validation. Notre échantillon comprend des bassins à régime nival mais également de nombreux bassins de plus basse altitude.

Nous constatons que le fait de fixer ou non les paramètres du module *CemaNeige* a une influence limitée sur les performances obtenues pour une majorité de bassins de notre échantillon français. En validation, nous constatons qu'il existe des cas où le calage des paramètres de *CemaNeige* est préférable (partie de droite du graphique c) et des cas où leur fixation est préférable (partie de gauche du graphique c). Ces derniers cas, moins nombreux, correspondent ainsi à des situations où le calage des paramètres de *CemaNeige* entraîne une diminution de la robustesse du modèle dans son ensemble (couple *GR4J-CemaNeige*). Notons cependant que les différences de simulation concernées, captées à travers le critère KGE, restent modérées.

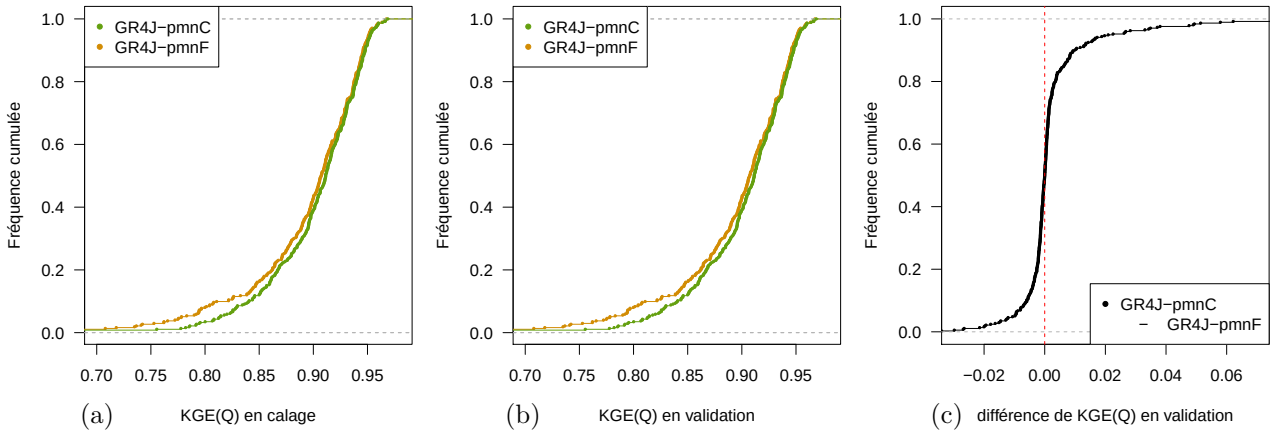


FIGURE C.4.: Performances de *GR4J-CemaNeige* en calage et en validation sur l'échantillon français (365 bassins), selon si les paramètres de *CemaNeige* sont calés ou fixés

en vert : paramètres du module neige calés (pmnC),
en orange : paramètres du module neige fixés à leurs valeurs médianes (pmnF)

Caractéristiques des bassins sensibles à l'une des configurations

Comme nous venons de le voir, le fait de caler ou fixer les paramètres de *CemaNeige* n'influence pas les performances de la même façon pour tous les bassins de notre échantillon :

- le calage de *CemaNeige* semble préférable pour certains bassins ;
- la fixation de paramètres de *CemaNeige* semble préférable pour d'autres bassins ;
- le fait de fixer ou caler *CemaNeige* est indifférent pour un troisième groupe de bassins.

Nous souhaitons nous intéresser ici aux caractéristiques des bassins composant ces trois groupes pour mieux comprendre cette situation. Pour cela, commençons par fixer un seuil de différence de KGE (choisi à 0.005) en dessous duquel nous considérons équivalant de caler ou fixer les paramètres de *CemaNeige*, vis-à-vis des performances globales de simulation. Nous constituons alors trois sous-groupes de bassins français, selon la situation : indifférence entre calage et fixation, calage préférable, fixation préférable.

Quelques caractéristiques intéressantes de ces sous-groupes sont présentées à la figure C.5. Bien que les distributions ne soient pas disjointes, nous pouvons noter les tendances suivantes :

1. Les bassins pour lesquels il est indifférent de fixer ou caler les paramètres de *CemaNeige* sont ceux où il y a peu de neige. Ce résultat est peu surprenant.
2. Assez logiquement, nous retrouvons qu'il est plutôt préférable de caler le module neige sur les bassins les plus élevés qui reçoivent une part plus importante de précipitations solides.
3. Les bassins pour lesquels une utilisation de paramètres fixés semble préférable constituent un groupe intermédiaire.

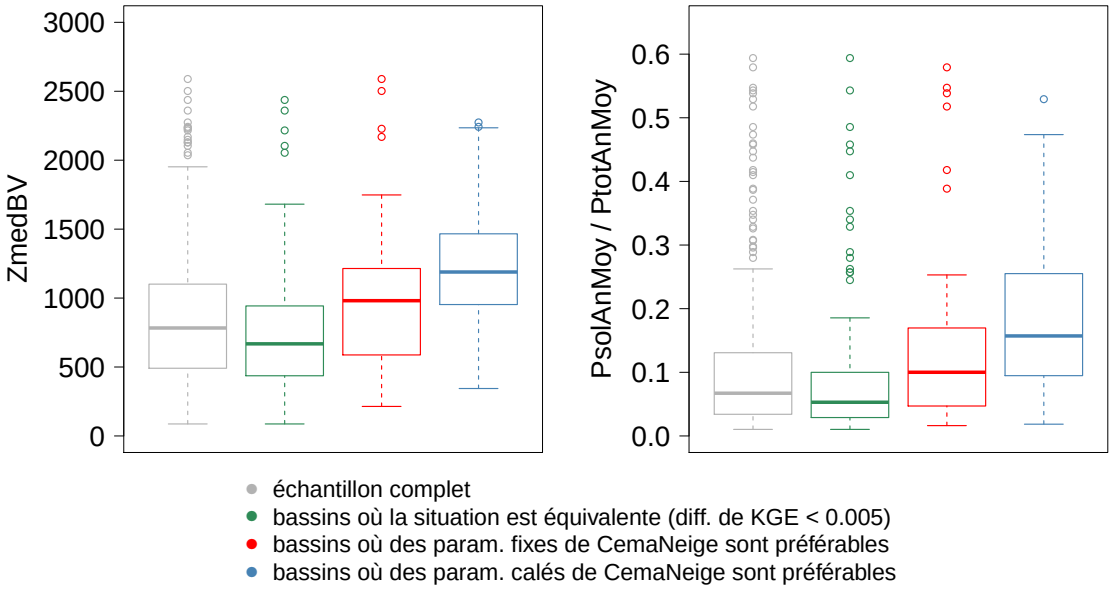


FIGURE C.5.: Part solide des précipitations et altitudes et des groupes de bassins où le calage de *CemaNeige* a une effet négatif/nul/positif sur la robustesse temporelle

Robustesse temporelle estimée (via le critère KGE) sur les débits simulés par le couple GR4J-CemaNeige, sur tous l'échantillon français (365 bassins)
trait épais central : valeures médianes, boîtes : quantiles 25 et 75%, moustaches : quantiles 5 et 95%

Conclusion

La question de la nécessité de caler ou non *CemaNeige* ne trouve pas de réponse unique sur notre échantillon de bassins français. Il est préférable de caler celui-ci sur certains bassins et de maintenir ses paramètres fixes sur d'autres. La position des bassins dans l'un ou l'autre de ces groupes s'explique assez logiquement par des critères géoclimatiques. Les diminutions de performance liées à l'utilisation de paramètres fixes (plutôt que calés) demeurent néanmoins raisonnables et nous nous autorisons donc à utiliser une telle configuration dans nos travaux pour étudier la robustesse des modèles hydrologiques seuls.

D

Complément pour le chapitre 3

D.1. Pertes performances moyennes liées aux transferts

Illustration des pertes de performances constatées entre une situation de calage et une situation de validation lors des très nombreux tests de transfert réalisés. Sur chaque sous-période simulée et pour chaque modèle, nous comparons donc les erreurs obtenues avec des paramètres issus d'un calage sur une autre sous-période aux erreurs obtenues en calage (*c.-à-d.* avec le jeu optimal sur cette sous-période simulée). Ces résultats viennent compléter ceux présentés dans la section 3.1.2 (notamment la figure 3.1, page 65).

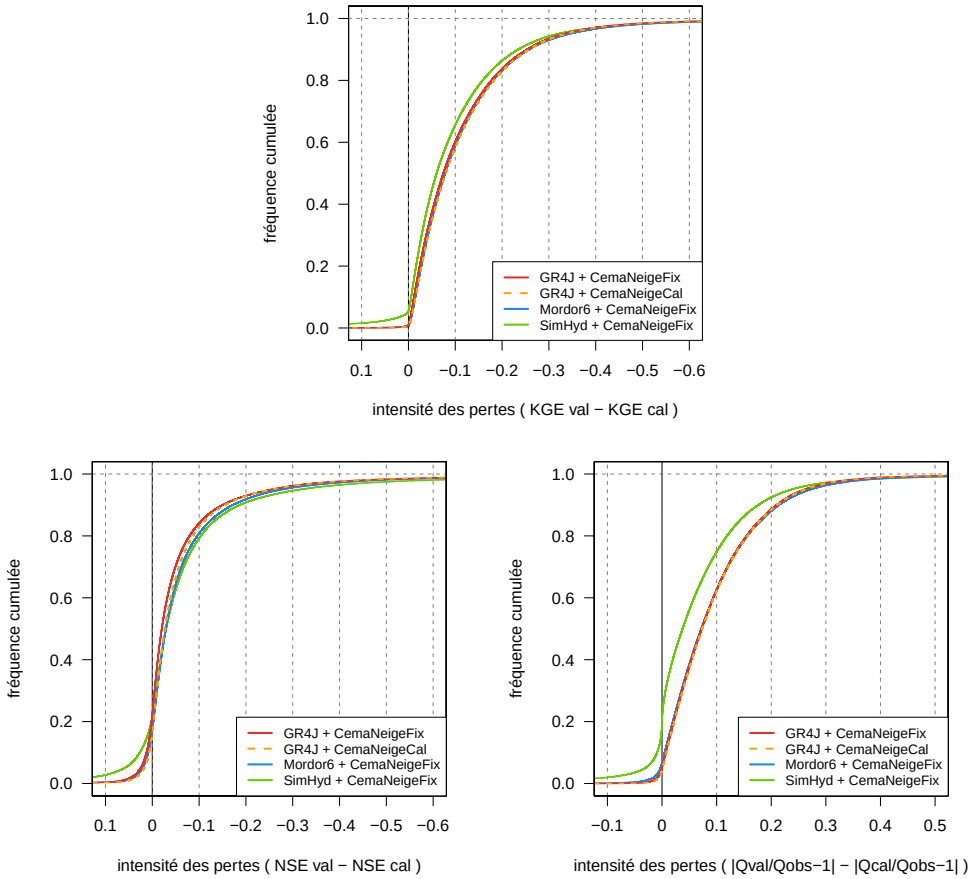


FIGURE D.1.: Résumé des pertes de performances relatives liées au transfert des paramètres

Distributions obtenues après agglomération de tous les résultats issus du GSST sur l'échantillon complet (569 bassins) avec des sous-périodes de 10 ans glissants. Modèles alimentés par une ETP d'Oudin dans toutes les configurations présentées ici.

Exemples de résultats :
Dans 40% des tests réalisés, la diminution de KGE entre une configuation optimale (calage) et une situation de transfert (validation) est inférieure à 0.05 (*c-à-d.* elle est plus grande dans 60% des cas). KGE étant notre critère de calage, dans l'idéal aucune configuation où $KGE_{cal} < KGE_{val}$ ne doit être rencontrée. Nous détectons donc ici quelques situations problématiques pour SimHyd.

D.2. Tableaux de corrélations entre △erreurs et △caract.

Tableaux présentant le résultats des analyses de corrélations entre les erreurs de modélisation et les conditions de transfert des paramètres des modèles lors du *GSST*. Les explications permettant la lecture de ces tableaux sont présentées dans la section 3.2.2.

Tableau de corrélation pour *GR4J*

TABLEAU D.1.: Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *GR4J*

Résultats pour GR4J-CemaNeige (θ_{neige} fixés) avec des données SPAZM et une ETP d'Oudin.

Corrélations calculées individuellement sur chacun des 569 bassins puis agglogémérées pour estimer un taux de représentativité au sein de l'échantillon complet (c.-à-d. pourcentage de bassins où $|r_s| > 0.5$).

Le code couleur souligne simplement les taux indiqués dans les cases ($< 5\%$ en blanc, $\approx 15\%$ en jaune, $\approx 30\%$ et $> 45\%$ en rouge).

	MRC	$r_p\left(\widehat{Q} \sim Q\right)$	$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_s>+0.5$	$r'_s<-0.5$	$r_s<-0.5$	$r_s>+0.5$	$r_s<-0.5$	$r_s>+0.5$
$\overline{T}$	21%	3%	19%	36%	51%	12%
$\overline{T}_{DJF}$	12%	4%	20%	33%	46%	14%
$\overline{T}_{MAM}$	19%	2%	21%	30%	46%	16%
$\overline{T}_{JJA}$	19%	3%	18%	38%	52%	13%
$\overline{T}_{SON}$	8%	2%	16%	31%	42%	9%
$\overline{P}$	8%	3%	28%	9%	20%	22%
$\overline{P}_{DJF}$	5%	2%	29%	12%	12%	31%
$\overline{P}_{MAM}$	6%	3%	31%	12%	11%	37%
$\overline{P}_{JJA}$	7%	3%	25%	18%	20%	25%
$\overline{P}_{SON}$	8%	1%	18%	24%	45%	11%
$\overline{P} / \overline{ETP}$	9%	3%	35%	12%	13%	35%
$\overline{P}_{DJF} / \overline{ETP}$	7%	3%	31%	14%	11%	39%
$\overline{P}_{MAM} / \overline{ETP}$	8%	3%	36%	14%	10%	46%
$\overline{P}_{JJA} / \overline{ETP}$	7%	3%	29%	15%	15%	32%
$\overline{P}_{SON} / \overline{ETP}$	6%	1%	18%	20%	35%	10%
$var(\overline{P}_{an})$	5%	2%	23%	12%	11%	28%
$var(\overline{P}_{an}/\overline{P}_{mens})$	3%	1%	19%	15%	18%	19%
$var(\overline{T}_{an})$	5%	1%	13%	19%	19%	13%
$var(\overline{T}_{an}/\overline{T}_{mens})$	12%	1%	27%	20%	15%	44%
Δ temporel	17%	2%	21%	39%	60%	13%

Tableau de corrélation pour *Mordor6*

TABLEAU D.2.: Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *Mordor6*

Résultats pour Mordor6-CemaNeige (θ_{neige} fixés) avec des données SPAZM et une ETP d'Oudin.

Corrélations calculées individuellement sur chacun des 569 bassins puis agglogémérées pour estimer un taux de représentativité au sein de l'échantillon complet (c.-à-d. pourcentage de bassins où $|r_s| > 0.5$).

Le code couleur souligne simplement les taux indiqués dans les cases ($< 5\%$ en blanc, $\approx 15\%$ en jaune, $\approx 30\%$ et $> 45\%$ en rouge).

	MRC	$r_p(\widehat{Q} \sim Q)$	$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_s>+0.5$	$r'_s<-0.5$	$r_s<-0.5$	$r_s>+0.5$	$r_s<-0.5$	$r_s>+0.5$
$\overline{T}$	21%	3%	17%	38%	54%	10%
$\overline{T}_{DJF}$	11%	3%	18%	34%	47%	12%
$\overline{T}_{MAM}$	20%	2%	21%	33%	49%	16%
$\overline{T}_{JJA}$	21%	3%	16%	40%	54%	10%
$\overline{T}_{SON}$	10%	3%	14%	32%	46%	8%
$\overline{P}$	9%	3%	27%	10%	15%	26%
$\overline{P}_{DJF}$	5%	2%	31%	11%	8%	36%
$\overline{P}_{MAM}$	9%	3%	33%	12%	8%	39%
$\overline{P}_{JJA}$	7%	3%	26%	17%	17%	29%
$\overline{P}_{SON}$	6%	1%	16%	25%	43%	11%
$\overline{P} / \overline{ETP}$	11%	3%	35%	10%	10%	43%
$\overline{P}_{DJF} / \overline{ETP}$	8%	2%	34%	12%	8%	43%
$\overline{P}_{MAM} / \overline{ETP}$	11%	3%	39%	13%	9%	48%
$\overline{P}_{JJA} / \overline{ETP}$	7%	3%	31%	14%	12%	37%
$\overline{P}_{SON} / \overline{ETP}$	4%	1%	16%	21%	32%	11%
$var(\overline{P}_{an})$	5%	1%	21%	13%	10%	28%
$var(\overline{P}_{an}/\overline{P}_{mens})$	4%	1%	19%	14%	17%	18%
$var(\overline{T}_{an})$	4%	1%	13%	21%	22%	13%
$var(\overline{T}_{an}/\overline{T}_{mens})$	11%	1%	30%	17%	14%	43%
Δ temporel	16%	3%	19%	42%	62%	11%

Tableau de corrélation pour *SimHyd*

TABLEAU D.3.: Taux de corrélations significatives entre les erreurs de simulation et diverses conditions de transfert des paramètres pour *SimHyd*

Résultats pour SimHyd-CemaNeige (θ_{neige} fixés) avec des données SPAZM et une ETP d'Oudin.

Corrélations calculées individuellement sur chacun des 569 bassins puis agglogémérées pour estimer un taux de représentativité au sein de l'échantillon complet (c.-à-d. pourcentage de bassins où $|r_s| > 0.5$).

Le code couleur souligne simplement les taux indiqués dans les cases ($< 5\%$ en blanc, $\approx 15\%$ en jaune, $\approx 30\%$ et $> 45\%$ en rouge).

	MRC	$r_p\left(\widehat{Q} \sim Q\right)$	$\left(\frac{\sigma_{\widehat{Q}} / \mu_{\widehat{Q}}}{\sigma_Q / \mu_Q}\right)$		$\left(\overline{\widehat{Q}} / \overline{Q}\right)$	
	$r'_S>+0.5$	$r'_S<-0.5$	$r_S<-0.5$	$r_S>+0.5$	$r_S<-0.5$	$r_S>+0.5$
$\overline{T}$	16%	3%	22%	27%	49%	9%
$\overline{T}_{DJF}$	8%	4%	22%	24%	43%	13%
$\overline{T}_{MAM}$	14%	2%	24%	23%	44%	14%
$\overline{T}_{JJA}$	13%	3%	22%	30%	49%	10%
$\overline{T}_{SON}$	9%	3%	19%	22%	39%	8%
$\overline{P}$	4%	4%	17%	13%	21%	19%
$\overline{P}_{DJF}$	4%	2%	16%	17%	9%	29%
$\overline{P}_{MAM}$	3%	3%	22%	16%	10%	31%
$\overline{P}_{JJA}$	5%	3%	21%	18%	19%	21%
$\overline{P}_{SON}$	6%	2%	18%	20%	41%	11%
$\overline{P} / \overline{ETP}$	5%	3%	22%	15%	11%	32%
$\overline{P}_{DJF} / \overline{ETP}$	6%	2%	21%	19%	9%	36%
$\overline{P}_{MAM} / \overline{ETP}$	5%	3%	26%	17%	10%	40%
$\overline{P}_{JJA} / \overline{ETP}$	5%	3%	23%	16%	13%	29%
$\overline{P}_{SON} / \overline{ETP}$	4%	2%	17%	19%	32%	11%
$var(\overline{P}_{an})$	4%	2%	16%	14%	10%	24%
$var(\overline{P}_{an}/\overline{P}_{mens})$	2%	1%	14%	17%	16%	15%
$var(\overline{T}_{an})$	5%	2%	14%	17%	17%	13%
$var(\overline{T}_{an}/\overline{T}_{mens})$	9%	1%	22%	22%	13%	41%
Δ temporel	12%	3%	25%	31%	54%	12%

D.3. Caractéristiques des sous-groupes d'analyse de tendance

La figure D.2 présente les caractéristiques des deux sous-groupes de bassins utilisés lors des analyses menées au chapitre 3, plus particulièrement dans les sections 3.3 et 3.4.

Cette figure permet de constater qu'il n'existe pas de différence majeure entre les caractéristiques des bassins où les erreurs de modélisation semblent soit liées à des écarts de température (distributions en orange) soit à des écarts de précipitations (distributions en vert). Les bassins pour lesquels $MRC = f(\triangle\overline{T})$ sont majoritairement français et ceux pour lesquels $MRC = f(\triangle\overline{P})$ sont majoritairement australiens, ce qui explique les différences de distributions constatées sur les $ETPs$ et les débits.

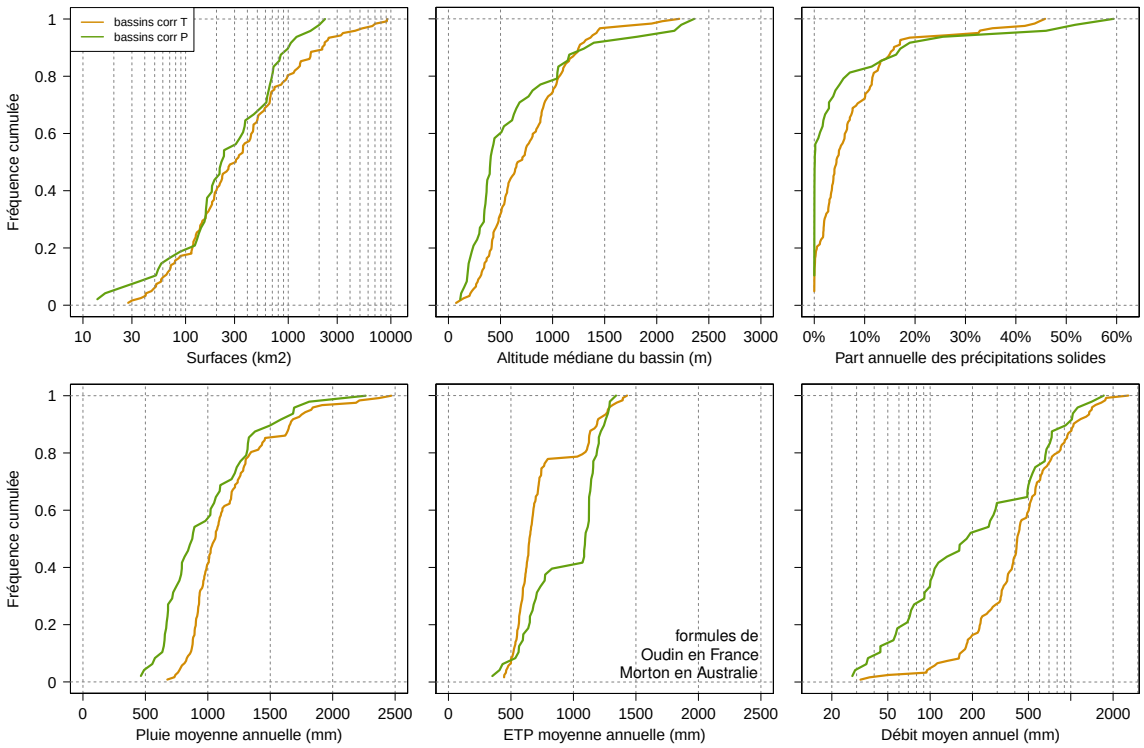


FIGURE D.2.: Résumé des caractéristiques de deux sous-groupes de bassins : 120 bassins où $MRC = f(\triangle\overline{T})$ et 48 bassins où $MRC = f(\triangle\overline{P})$

Ces groupes de bassins sont ceux utilisés dans les analyses du chapitre 3 (sections et 3.3 et 3.4).

La figure D.3 présente les caractéristiques de deux autres sous-groupes de bassins qui ne sont pas utilisés dans nos analyses mais correspondent aux bassins pour lesquels des corrélations significatives ($|r_s| > 0.5$) sont détectés sur les variations d'erreurs de volume moyen. En analysant les résultats du *GSST* sur 10 ans avec *GR4J-CemaNeige*, il est ainsi possible d'identifier 360 bassins pour lesquels $\triangle(\overline{\widehat{Q}}/\overline{Q}) = f(\triangle\overline{T})$ et 244 bassins pour lesquels $\triangle(\overline{\widehat{Q}}/\overline{Q}) = f(\triangle\overline{P})$. À la différence des sous-groupes constitués sur $MRC = f(\triangle caract.)$, il y a un léger recouvrement ici, certains bassins présentant les deux types de corrélations.

Cette figure complémentaire montre comment les caractéristiques physiques, géographiques et climatiques des bassins n'expliquent pas, à elles seules, l'hétérogénéité des comportements des modèles constatée lors de l'utilisation du *Generalized Split-Sample Test (GSST)* sur tout notre échantillon.

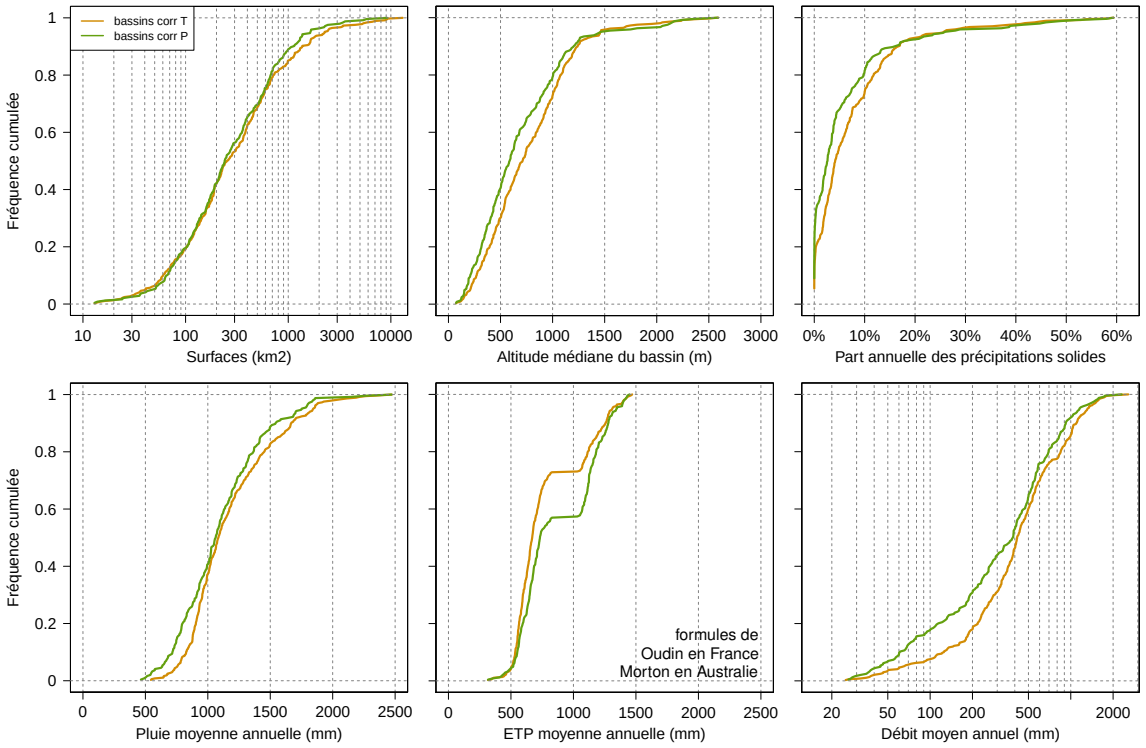


FIGURE D.3.: Résumé des caractéristiques de deux sous-groupes de bassins : 360 bassins où $\triangle(\overline{\widehat{Q}}/\overline{Q}) = f(\triangle\overline{T})$ et 244 bassins où $\triangle(\overline{\widehat{Q}}/\overline{Q}) = f(\triangle\overline{P})$

D.4. Erreurs de volumes constatées lors des tests de transfert

Illustration des erreurs de volumes constatées en validation lors des nombreux tests de transferts réalisés sur un sous-groupe de 120 bassins où un lien de corrélation est détecté entre les variations de *MRC* et de température moyenne de l'air. Ces résultats viennent compléter ceux présentés dans la section 3.3.2 (notamment la figure 3.3, page 77).

Nous noterons que les erreurs de volume qui ressortent lors d'un calcul de moyenne proviennent plutôt d'erreurs sur les forts débits tandis que les nuages deviennent très diffus pour les débits les plus faibles, où les erreurs demeurent très importantes mais ne sont pas corrélées avec le $\triangle\overline{T}$ comme pour le reste de la gamme.

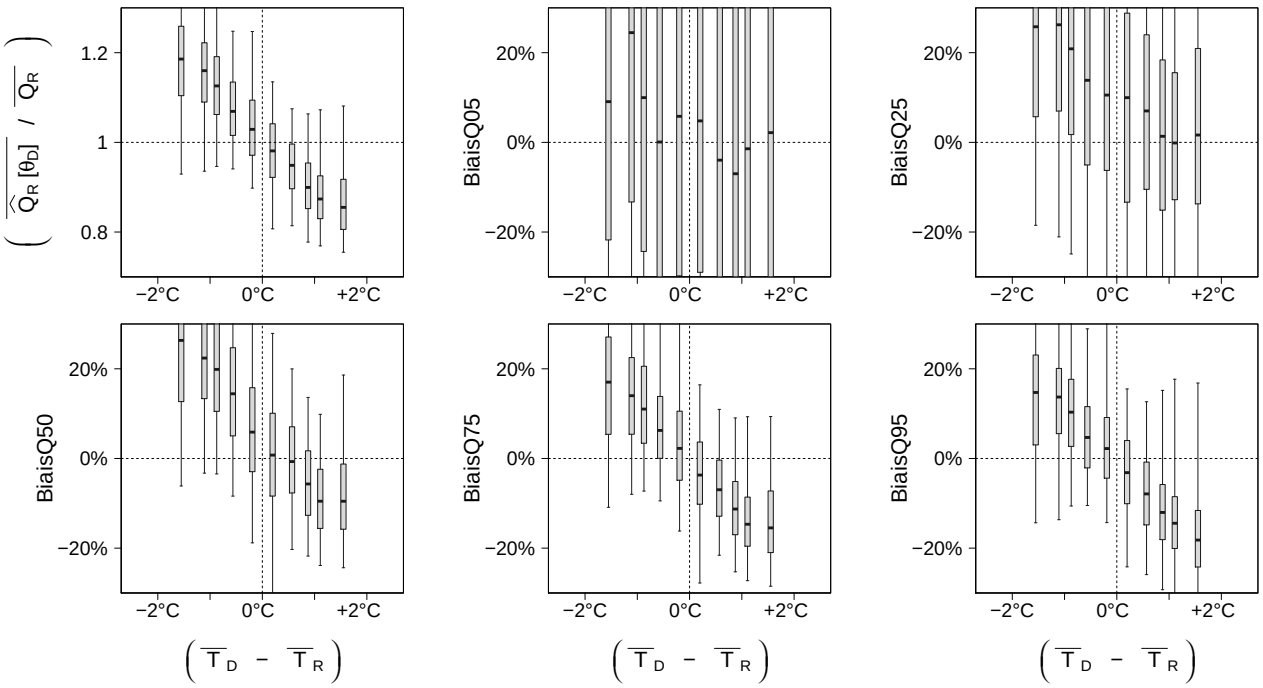


FIGURE D.4.: Erreurs de volumes obtenues avec *GR4J* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures

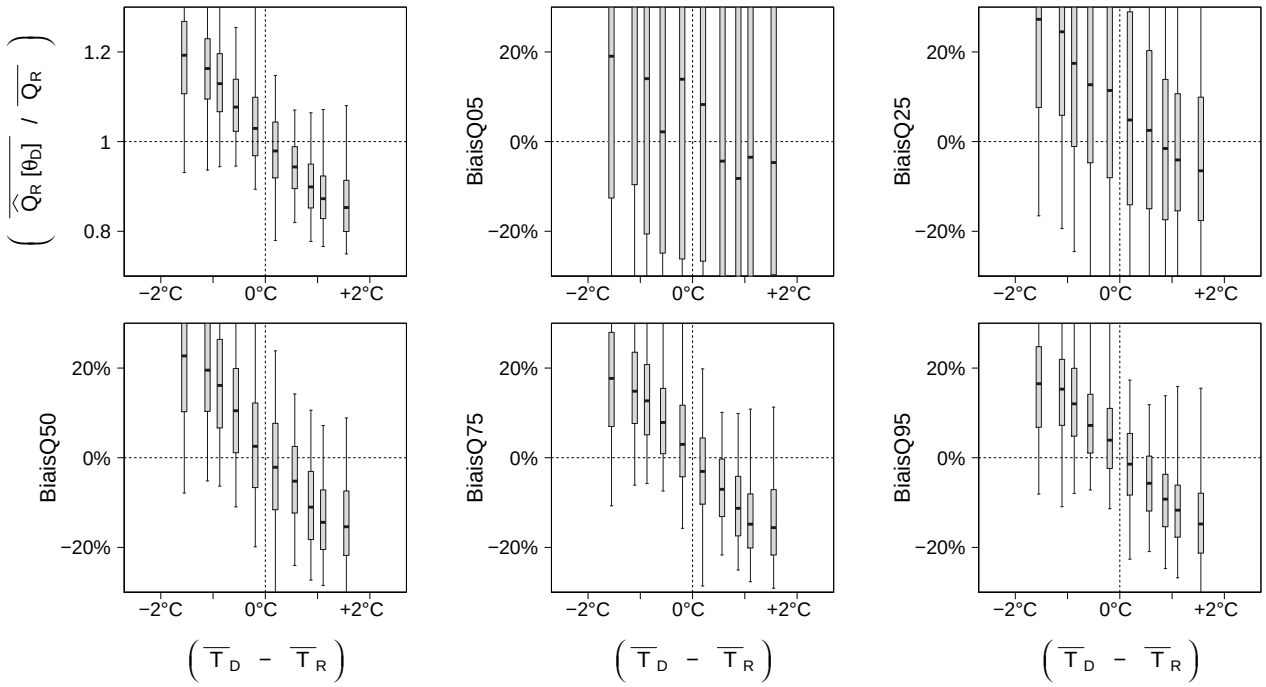


FIGURE D.5.: Erreurs de volumes obtenues avec *Mordor6* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures

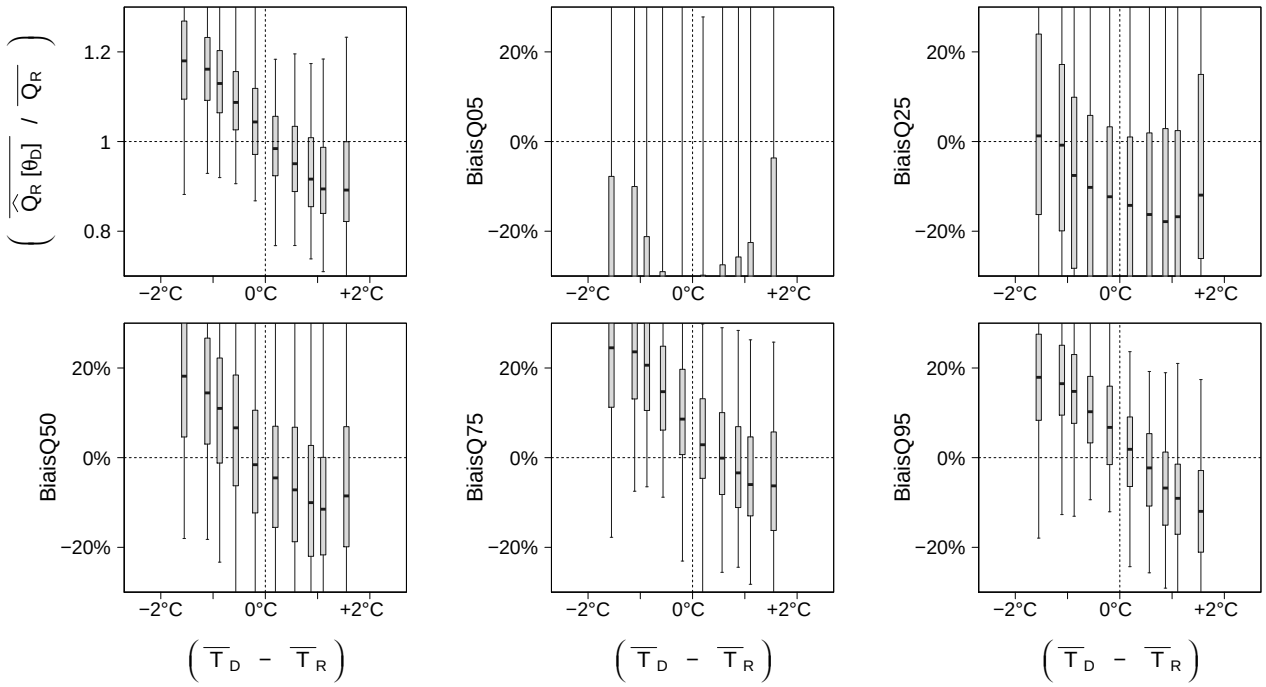


FIGURE D.6.: Erreurs de volumes obtenues avec *SimHyd* lors du *GSST* sur les 120 bassins concernés par des corrélations significatives entre △erreurs et △températures

Les notations utilisées ici correspondent aux quantiles de non-dépassement.
Par exemple, le Q95 correspond à un fort débit qui n'est dépassé que 5% du temps.

Rq. sur les bassins français, GR4J, Mordor6 et SimHyd sont alimentés par des données SPAZM et une ETP d'Oudin et couplés à CemaNeige (θ_{neige} fixés).

E

Crash testing of hydrological models in contrasted climate conditions...

Éléments de contexte liés à cette publication

Une partie des travaux présentés dans cette thèse ont été réalisés dans le cadre d'une collaboration avec le *CSIRO Land&Water* (Australie). Celle-ci s'est notamment traduite par mon séjour durant 5 mois (mars-juillet 2011) au sein de l'équipe de Jai Vaze et Francis Chiew à Canberra. Ce séjour m'a permis notamment l'utilisation de leur base de bassins versants qui présente une richesse vis-à-vis de notre thématique de recherche. La forte variabilité des conditions climatiques en Australie permet en effet d'obtenir des situations de forts contrastes dans les chroniques d'observation entre des périodes différentes d'un même bassin versant. Autrement dit, ces données constituent un terrain de jeu très intéressant pour les tests de transfert que nous souhaitions réaliser. Au-delà du simple échange de données, ce fut bien entendu une collaboration scientifique plus large autour de mon sujet d'étude, puisque mes travaux sur la robustesse climatique se placent notamment dans la continuité des tests menés par le *CSIRO Land&Water* (*cf. Vaze et al.*, 2010).

Les fruits de cette collaboration ont notamment été valorisés lors de l'*IUGG2011*[1], où j'ai présenté ces travaux lors d'un exposé oral, puis dans l'article scientifique fourni ici, qui fut publié en avril 2012 dans *Water Resources Research*.

Je tiens à remercier ici encore toutes les personnes qui ont permis ce projet de collaboration, notamment : Vazken Andréassian (Irstea) qui donna l'impulsion initiale, mes collègues hydrologues de *Chatou* pour leur soutien au projet et plus largement les départements *LNHE* et *MFEE* d'*EDF-R&D* qui prirent en charge la majorité des frais liés à mon séjour. J'adresse des remerciements particuliers à Marie Bourqui (*EDF-R&D*) et Julien Lerat (*CSIRO*) pour leur persévérance dans l'affrontement des nombreuses étapes administratives, qui furent nécessaires pour permettre mon détachement d'*EDF* vers le *CSIRO*. Merci également à Francis Chiew et Jai Vaze pour leur accueil au *CSIRO* et leur disponibilité. Enfin, je remercie à nouveau Julien Lerat qui fut un collègue et ami durant ces cinq mois australiens, ainsi qu'une source de nombreux conseils pertinents qui m'aidèrent à valoriser au maximum mon séjour, tant du point de vue professionnel que personnel.

Référence complète de l'article

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions : An experiment on 216 australian catchments. *Water Resources Research*, 48(5), W05552.

1. Conférence de l'*International Union of Geodesy and Geophysics*, 28 juin - 07 juillet 2011, Melbourne.

WATER RESOURCES RESEARCH, VOL. 48, W05552, doi:10.1029/2011WR011721, 2012

Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments

L. Coron,[1,2] V. Andréassian,[2] C. Perrin,[2] J. Lerat,[3] J. Vaze,[3] M. Bourqui,[1] and F. Hendrickx[1]

Received 7 December 2011; revised 26 March 2012; accepted 15 April 2012; published 26 May 2012.

[1] This paper investigates the actual extrapolation capacity of three hydrological models in differing climate conditions. We propose a general testing framework, in which we perform series of split-sample tests, testing all possible combinations of calibration-validation periods using a 10 year sliding window. This methodology, which we have called the generalized split-sample test (GSST), provides insights into the model's transposability over time under various climatic conditions. The three conceptual rainfall-runoff models yielded similar results over a set of 216 catchments in southeast Australia. First, we assessed the model's efficiency in validation using a criterion combining the root-mean-square error and bias. A relation was found between this efficiency and the changes in mean rainfall (P) but not with changes in mean potential evapotranspiration (PE) or air temperature (T). Second, we focused on average runoff volumes and found that simulation biases are greatly affected by changes in P. Calibration over a wetter (drier) climate than the validation climate leads to an overestimation (underestimation) of the mean simulated runoff. We observed different magnitudes of these models deficiencies depending on the catchment considered. Results indicate that the transfer of model parameters in time may introduce a significant level of errors in simulations, meaning increased uncertainty in the various practical applications of these models (flow simulation, forecasting, design, reservoir management, climate change impact assessments, etc.). Testing model robustness with respect to this issue should help better quantify these uncertainties.

Citation: Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, *48*, W05552, doi:10.1029/2011WR011721.

1. Introduction

1.1. Challenges in Predicting Hydrological Response Under Variable Climatic Conditions

[2] Quantifying the impacts of climate change on streamflow has been an increasing concern in the past few years and has been the focus of many studies throughout the world [*Caballero et al.*, 2007; *Vicuna and Dracup*, 2007; *Steele-Dunne et al.*, 2008; *Chiew et al.*, 2009; *Görgen et al.*, 2010]. The modeling steps associated with this task include (1) selecting emission scenarios, (2) running global circulation models (GCMs), (3) downscaling the GCM's output to a scale that can be used for hydrology, and (4) running hydrological models that simulate the rainfall-runoff (RR) transformation at the catchment scale. Step 4 is often considered to contribute less than the other steps to the overall uncertainty [*Wilby and Harris*, 2006; *Prudhomme and Davies*, 2009; *Kay et al.*, 2009; *Arnell*, 2011; *Teng et al.*, 2011]. However, the uncertainty associated with the estimation of parameters of hydrological models cannot be neglected [*Wilby*, 2005; *Vaze et al.*, 2010b; *Merz et al.*, 2011]. This is even truer when RR models are run under climatic conditions significantly different from calibration conditions (e.g., projections of future conditions versus current conditions). Indeed, many unknowns remain concerning the actual transposability over time of model parameters under contrasted conditions. Although this transposability is a critical issue in the context of climate change impact studies where nonstationary conditions are explicitly considered, it also has implications in other more operational model applications (like forecasting, design, etc.), in which model robustness is essential to provide reliable results. Besides, fluctuations in climate also exist in historical time series (such as Hurst-Kolmogorov behaviors [see *Koutsoyiannis*, 2011]), which may question parameters transferability whenever a model is used to simulate flows on a period whose climatic conditions are different from those in model calibration.

1.2. Parameter Dependency on Calibration Period Climate

[3] Because of the lack of knowledge and data on the true functioning of the hydrological system, all hydrological models remain to some extent conceptual and empirical [*Murphy et al.*, 2006]. As a result, deriving physically

[1]LNHE, EDF R&D, Chatou, France.
[2]Hydrosystems and Bioprocesses Research Unit, Irstea, Antony, France.
[3]Black Mountain Laboratories, CSIRO Land and Water, Acton, ACT, Australia.

Corresponding author: L. Coron, LNHE, EDF R&D, 6 quai Watier, F-78401 Chatou CEDEX, France. (laurent.coron@edf.fr)

©2012. American Geophysical Union. All Rights Reserved.
0043-1397/12/2011WR011721

meaningful values for the model's parameters via calibration remains a challenging task. In their discussion on model "pathologies," *Coron et al.* [2011] reviewed various situations where parameter estimation can be hampered and induce values that transfer poorly to other periods. Common examples are cases where input quality and/or availability evolve as well as issues related to low levels of parameter identifiability. Problematic situations may also emerge from the climatic dependency of model parameters.

[4] Existing hydrological models have been developed using either top-down (conceptual models) or bottom-up (physically based) approaches. However, both types of models suffer from the same problems when it comes to the calibration of their parameters. At calibration stage, the optimal set can vary over time in accordance with seasonal and/or long-term climatic variations. *Wagener et al.* [2003] applied a five-parameter lumped model to an English catchment and showed that summer and rain storm periods require different optima for the parameters controlling rapid water transfer. *Choi and Beven* [2007] sampled the times series of a South Korean catchment according to a hydrological similarity measure. Calibrating TOPMODEL parameters on each cluster, they found that optimal sets on some clusters were not convenient for use on others. *Rosero et al.* [2010] found that parameters from the Noah land surface model, which should in principle be controlled only by physical site characteristics (e.g., soil and vegetation type), were also strongly influenced by climatic conditions. These findings are not restricted to a limited number of particular catchments. Although the work from *Rosero et al.* [2010] was based on only nine catchments, *Vaze et al.* [2010b] and *Merz et al.* [2011] conducted studies over 61 Australian and 273 Austrian catchments, respectively, and observed similar dependencies. *Merz et al.* [2011] established a link between the HBV parameters representing snow and soil moisture processes with climatic characteristics such as air temperature and PE. Cases of apparent independence may also be observed (see the work by *Niel et al.* [2003] on 17 African catchments). Most of these past results indicate that the assumption of parameter stability over time is strong. Parameter values can vary seasonally because of differences in dominant hydrological processes controlling runoff generation in different seasons but may also change on longer time scales in relation to climate variability (e.g., modifications of annual groundwater balance, vegetation change, etc.). Recently, *de Vos et al.* [2010] made an interesting proposition to further investigate the reasons for this disturbing dependency: they suggested clustering time series according to climatic similarities and allowing parameters to vary over these clusters during calibration. Rather than an alternative optimization method, they presented it as a tool for investigating model functioning and thus identifying the possible needs for improvements.

[5] A parallel can be made between transferring parameters over time and space. For example, *Merz et al.* [2011] observed temporal trends on parameter values (due to climate evolution) that were comparable to variations over space when moving between regions with different climates. A similar trade of space for time was made by *Singh et al.* [2011], who used a regional approach to evaluate extrapolation skills of parameter sets by transferring them to other catchments in warmer climatic zones.

[6] Note that, to ease reading, the term climate is abusively used hereafter to designate the main characteristic of precipitation, temperature, etc., over a time-limited period (often 10 years long here), whereas climate usually refers to characteristics over an extended period of time (typically several decades).

[7] A typical testing procedure to investigate parameter dependency on climate and related consequences on model efficiency is the differential split-sample test (DSST) proposed by *Klemeš* [1986]. This is a specific case of the split-sample test (SST), where calibration and validation periods are chosen according to their climatic differences. The parameter dependency on the calibration periods is analyzed through the evolution of the model's performance on this test. Examples of applications of DSST include the studies by *Refsgaard and Knudsen* [1996], *Donnelly-Makowecki and Moore* [1999], *Xu* [1999], *Seibert* [2003], *Wilby* [2005], *Chiew et al.* [2009], *Vaze et al.* [2010b], and *Bastola et al.* [2011]. Most of these authors observed decreases in model performance (i.e., larger model errors) after transferring parameter sets between climatically contrasted periods. They concluded that a model's suitability for climate change impact studies depends on the judicious choice of the calibration period. In that context, some recommend the use of extended periods to ensure sufficiently diverse climatic and flow conditions during calibration to give a representative picture of their natural variability. Others suggest the use of calibration periods whose climatic conditions are closer to future ones, often corresponding to the recent records. However, this second option prevents from benefiting from the full information available to quantify modeling uncertainty.

1.3. A Need for Further Investigation

[8] Increasingly, hydrologists use RR models over wide ranges of climatic conditions not necessarily encountered during the calibration stage, and need to have an estimate of the uncertainties associated with their simulations. Several studies have emphasized the limitations in the transfer of parameters between climatically contrasted periods, some of which were mentioned above. However, very few investigations have been conducted using a methodology that would provide general conclusions on this issue, i.e., on the basis of a large number of catchments and using different models. The two main studies by *Vaze et al.* [2010b] and *Merz et al.* [2011] showed that differences in climate between calibration and validation could significantly affect model performances. For instance, *Vaze et al.* [2010b] found that transferring parameters to a drier climate was particularly problematic and concluded that such transfer should not be made for changes in mean rainfall greater than 15%.

[9] The above mentioned studies require complementary work (1) to develop more general testing procedures able to provide comparable results under various conditions and over a wide range of parameter transfer conditions, thus resulting in more robust conclusions on parameter transferability, (2) to apply such procedures on a variety of cases for a better quantification of model robustness under a changing climate, and (3) to identify the situations where parameters are not transferable and, if possible, explain why.

[10] This study extends the work of *Vaze et al.* [2010b] by enlarging the catchment set and proposing a new

generalized methodology to evaluate the validity of RR models for use under nonstationary climatic conditions. Some of the conclusions of these authors are confirmed, but new insights are also provided (e.g., on catchment-specific behaviors).

[11] This paper is organized as follows. The catchment set and hydrological models used are presented in section 2. The methods used to evaluate the parameter transferability are described in section 3. The results are presented in section 4, starting from the analysis of the entire data set and then distinguishing different behaviors. A discussion around the methodology and the results is provided in section 5. Conclusions are summarized in section 6.

2. Catchment Set and Models

2.1. Study Area

[12] The choice of the study area was greatly influenced by the scope of this work. We needed sufficiently variable climatic conditions to make it possible to select contrasted periods for model testing in the most extreme conditions. In this perspective, we used a set of 216 catchments in southeast Australia, where climate variability is notoriously greater than in many other places in the world. The initial set was composed of 228 catchments, but 12 catchments were not used in our testing procedure because of insufficient data availability (see section 3.4). This catchment set has been described in detail by *Vaze et al.* [2010a]. The catchments are located on a large zone from south Queensland to west Victoria along the Great Dividing Range [Figure 1]. The range of climate and physical characteristics of the data set is summarized in Table 1. Rainfall is the most important driver of runoff in Australia and is much more variable both temporally and spatially than the other climate variables [*Vaze et al.*, 2010a]. Only 15% of the rainfall becomes runoff on average for the catchment set. Significant variations in rainfall and considerable variations in streamflow can occur between years, as shown by the interannual variability coefficients (see Table 1). Variability also exists between longer periods, as shown in Figure 2, which plots the series of relative mean PE, rainfall and streamflow values over a 10 year sliding window. The three graphs illustrate how the mean climate over a decade can differ from the climate over the entire record (32 years in most cases). While ranges of ±10% in mean rainfall and ±3% in mean PE between 10 year subperiods can be observed for most catchments, this results in relative variations on streamflow that can be much larger (up to 50% in absolute value). Note that on most catchments, the 1980s were wetter than average, while the end of the period was much drier. This is a quite interesting contrast for the objectives of our study. Across the data set, there is also a spatial variation of average conditions in terms of rainfall, PE and runoff. Catchments located east and south of the Great Dividing Range are wetter than the catchments located inland. Moreover, catchments are summer dominated in the

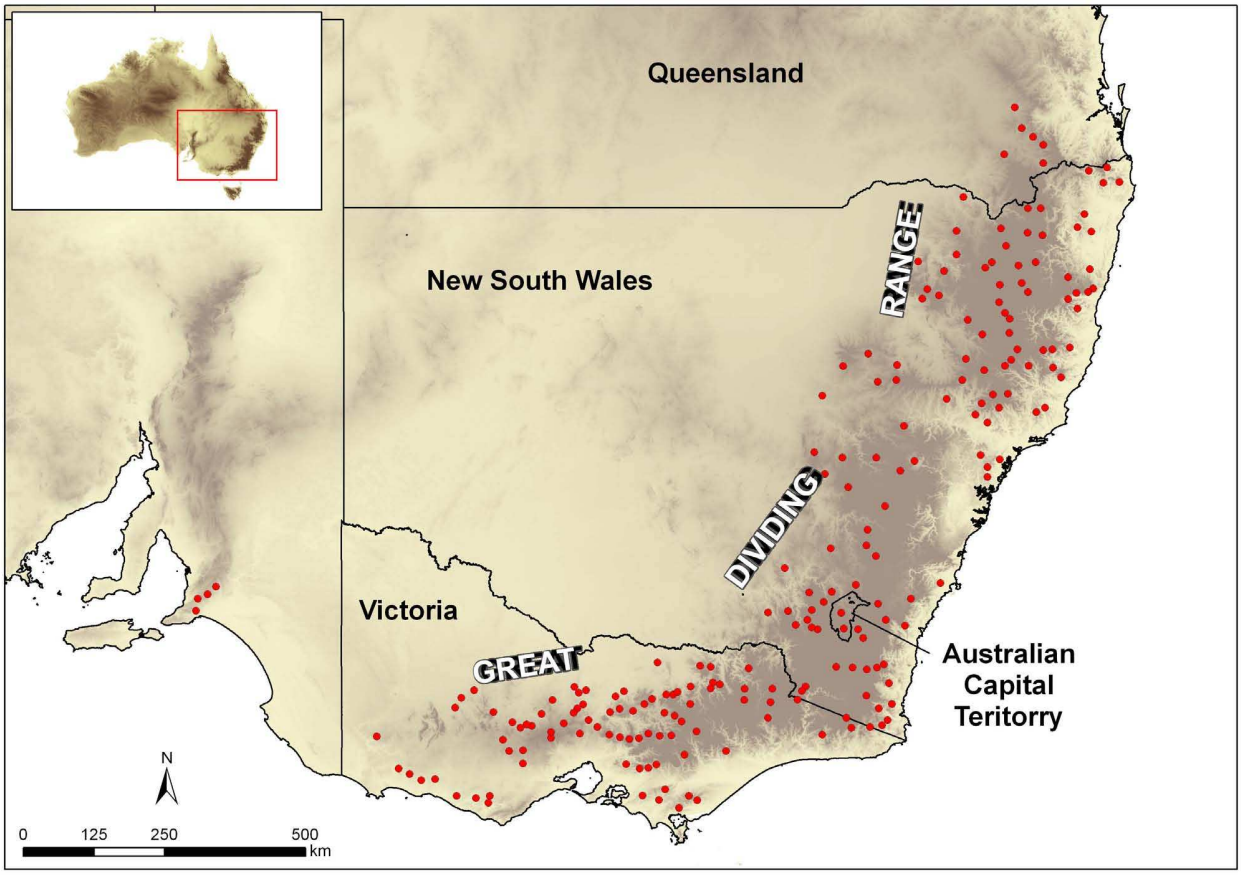


Figure 1. Locations of the 216 catchments.

Table 1. Percentiles of the Distributions of a Few Catchment Characteristics on the Entire Set of 216 Catchments[a]

	Statistics Over the Entire Catchment Set				
	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
Catchment surface (km^2)	70	160	330	630	1240
Mean annual potential evaporation PE_{ma} (mm)	1070	1120	1200	1290	1410
Mean annual rainfall P_{ma} (mm)	570	720	860	1100	1400
Mean annual runoff Q_{ma} (mm)	30	60	120	250	500
Aridity index P_{ma}/PE_{ma} (%)	45	59	73	92	121
Rainfall-runoff yield Q_{ma}/P_{ma} (%)	4	9	15	23	36
Interannual variability of PE (%)	2.2	2.5	2.7	2.8	3.1
Interannual variability of P (%)	13	17	20	23	29
Interannual variability of Q (%)	29	46	70	91	114

[a]Values of interannual variability correspond to coefficients of variation calculated on 10 year periods.

north (i.e., most rainfall and runoff occur in summer) and winter dominated in the south, while interannual variability is greater in the north than in the south.

[13] Daily data of rainfall and PE were derived from the SILO Data Drill, which provides daily data for 0.05° surface grids (∼5 × 5 km) across Australia (http://www.longpaddock.qld.gov.au/silo/). These estimates are interpolated from point measurements made by the Australian Bureau of Meteorology. Daily PE is computed using Morton's wet environment algorithms [*Morton*, 1983]. Daily streamflow data for the 216 catchments were obtained from relevant state government agencies and were checked for errors. For a majority of catchments, continuous records of rainfall, PE and runoff were available for the 1974–2006 period. Most of the catchments range in size between 100 and 1000 km^2, with a median value around 330 km^2. They are mostly unregulated with no major storage or irrigation schemes. These data were partly used in recent Australian projects such as the Murray-Darling Sustainable Yields project [*Chiew et al.*, 2008], the South-Eastern Australian Climate

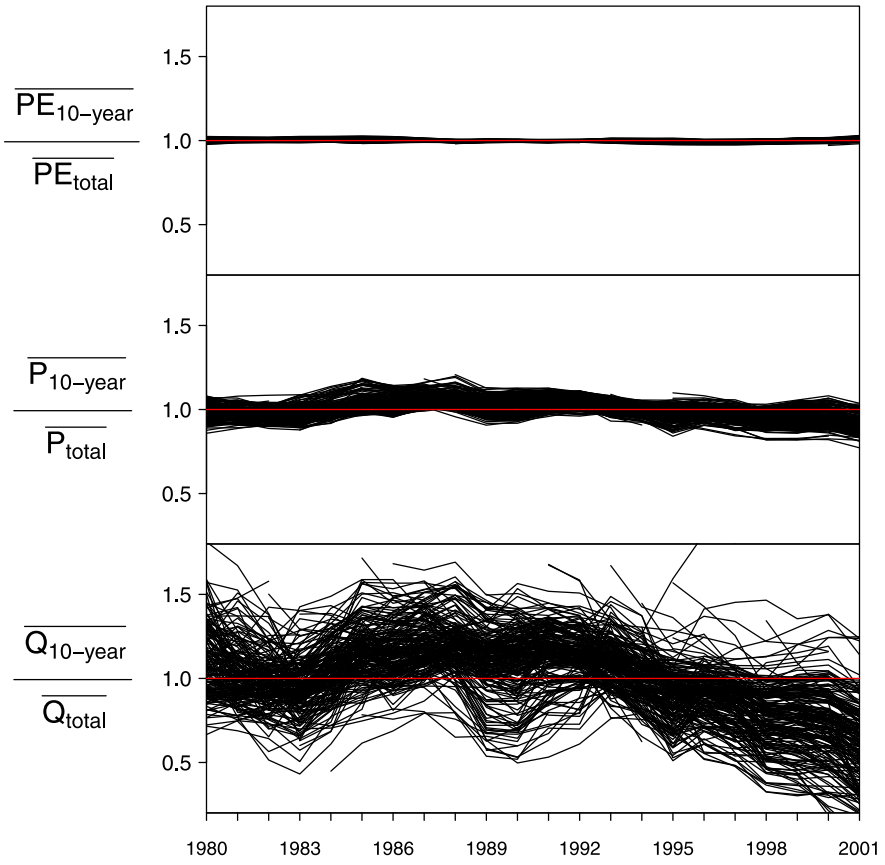


Figure 2. Relative long-term climate variability of potential evapotranspiration (PE), rainfall (P), and streamflow (Q) over the catchment set. For each catchment, a line corresponds to the series of mean values over a 10 year sliding window. Values are expressed relative to the average value on the total record (usually 1974–2006). Each value is plotted at the central year of the 10 year window.

Initiative (http://www.seaci.org/) and the study for the New South Wales Office of Water on climate impact on runoff [*Vaze and Teng*, 2011].

2.2. Hydrological Models

[14] Three daily lumped continuous reservoir-type RR models were used in this study: GR4J, MORDOR6 and SIMHYD. These models had already been applied to large data sets in previous studies [*Chahinian et al.*, 2006; *Oudin et al.*, 2008; *Perrin et al.*, 2008; *Chiew et al.*, 2009]. Table 2 gives an overview of the characteristics of these models as well as the references where detailed descriptions are provided. In spite of their parsimony (only a few free parameters), they showed a good level of efficiency in past applications and correspond to different representations of the RR transformation. GR4J is based on two stores and four parameters to calibrate, while MORDOR6 and SIMHYD both have four stores and six and seven free parameters, respectively.

3. Crash Test Methodology and Analysis Method

[15] Our general objective was to study the transfer of parameter sets between climatically contrasted periods. In the vein of the discussion of *Andréassian et al.* [2009] on model evaluation, we implemented a "crash test" methodology for models to be used in changing climatic conditions (typically such as climate change impact studies), i.e., a testing methodology putting models in extremely demanding application conditions, in order to explore their application limits. Loss of robustness caused by inappropriate parameter transfers are analyzed through the variations in model performance.

3.1. Generalized Split-Sample Test

[16] The differential split-sample test (DSST) discussed in section 1.2 allows us to evaluate models in contrasted climatic conditions. The method usually follows three steps: (1) A small number of subperiods are selected according to one climatic characteristic (e.g., mean rainfall or temperature for the catchment). (2) The calibration-validation test is applied on these periods. (3) The validation performances are compared to evaluate whether they vary significantly when climatic characteristics differ between calibration and validation periods.

[17] This procedure has two limitations for obtaining robust and generalizable conclusions. First, it requires knowing in advance which climatic characteristics most likely play a key role in limiting the parameter set transfer. If the influence of different characteristics is tested, it is often difficult to compare the results because the subperiods used are different. Indeed they are selected according to the climatic property studied: for example, the driest period may differ from the warmest one. Second, the number of transfer tests is usually small, as often only two or three contrasted periods can be identified. This limits the possibility of drawing general conclusions and discovering the main drivers of parameter transferability from the results themselves. Indeed, it might be hard to distinguish the effect of the climate difference from other aspects potentially influencing parameter transfer.

[18] To overcome these limitations, we propose a generalization of the standard SST and DSST schemes. The objective is to test the model in as many and as varied climatic configurations as possible, including similar and contrasted conditions between calibration and validation. The problem was approached the other way round compared to what is usually done: numerous tests of parameter transfer were carried out and the results were analyzed to determine afterward whether the variations in the transfer quality were related to climatic aspects. This approach will be called the generalized split-sample test (GSST) hereafter.

[19] The GSST procedure simply consists of a series of calibration-validation tests on subperiods of equal length, considering all possible configurations. This procedure is based on the following steps (see Figure 3).

[20] 1. A sliding window of the chosen length (5 years on the graph) is used to define subperiods. Between two periods, the window is moved by 1 year (i.e., one hydrological cycle), thus allowing the subperiods to overlap. In Figure 3, these subperiods are the dark gray bars, while light gray represents the remaining part of the time series.

[21] 2. The hydrological model(s) are calibrated on each subperiod using a previously selected function. This provides one parameter set θ per period. At this step, any objective function or calibration algorithm can be used.

[22] 3. For each calibration subperiod, the optimized parameter set is used to perform all the possible validation tests on independent subperiods. Validation subperiods overlapping with the calibration one are not considered to ensure strict independence of calibration and validation conditions (see Figure 3). Moreover, a reference flow series is simulated for the calibration period using the parameter set obtained after calibrating the model on that period. Note that the number of validation tests will not be the same for all calibration periods. But this is not a problem as all results will be analyzed together.

Table 2. Overview of the Characteristics for the Three Models Tested[a]

	GR4J	MORDOR6	SIMHYD Plus Routing
Number of free parameters	4	6	7
Structure overview: production	A rainfall interception by PE, a nonlinear SMA store, an intercatchment GW exchange function	A rainfall correction factor, a nonlinear SMA store, a lower evaporative store	An interception store, a SMA nonlinear store
Structure overview: transfer	Two unit hydrographs, a nonlinear routing store	A direct flow component, an infiltration store, a linear routing store, a unit hydrograph	A groundwater store, a nonlinear routing store, a convolution delay
Sources and first publications	*Perrin et al.* [2003]	*Mathevet* [2005] (original MORDOR version from *Garçon* [1996])	*Chiew et al.* [2002] (with simplifications from *Tan et al.* [2005])

[a]PE, potential evapotranspiration; SMA, soil and moisture capacity; GW, groundwater.

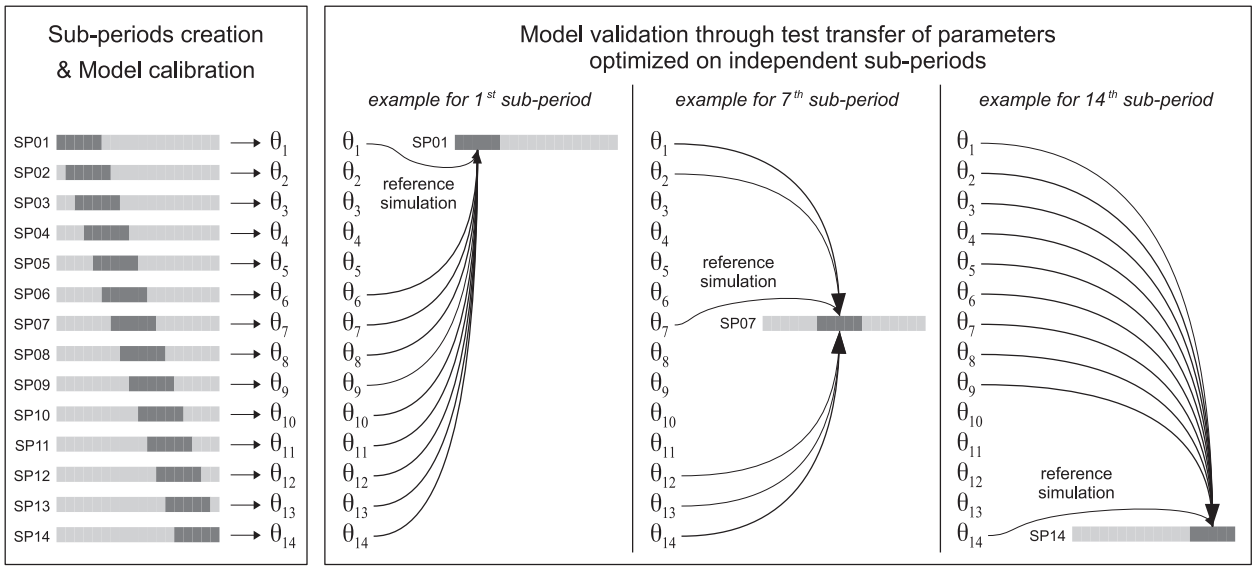


Figure 3. Illustration of the proposed generalized split-sample test (GSST) procedure (example with 18 years available and 5 year subperiods).

[23] To our knowledge, the SST, in which periods are not a priori selected but all combinations are tested, had not been used in this way before. The GSST procedure is meant to overcome the limitations previously mentioned: (1) It leads to a large number of test cases for the analysis, providing a continuum of the possible range of climatic differences existing in the observed series; for instance, with an 18 year long time series, sampling into 5 year periods and 3 year periods generates 90 and 182 SSTs, respectively. (2) It is less subjective because no choice is made before the tests. (3) The influence of changes in any climate characteristics on model robustness can be studied (ranging from mean interannual variables to indices characterizing the mean seasonal contrasts within the year). Indeed, all possible configurations are tested and lead to a unique list of validation performances. Switching between rainfall and temperature in the analysis simply means expressing these performances relative to one or the other climatic characteristic (the common practice where periods are selected with respect to one characteristic or the other leads to results which are not directly comparable).

[24] It could be argued that if the number of tests is increased there will be considerable redundancy in the tests carried out since the subperiods are not independent. Actually, our intention was to multiply the number of SSTs to study the entire range of climatic differences available between periods, even if each of them does not radically differ from all the others. The important point is that the calibration and validation periods are actually independent, which is in agreement with the original SST scheme.

3.2. Which Criteria Can Quantify the Extrapolation Capacity of a Given Parameter Set?

[25] When a model is used to simulate discharges, errors will arise: (1) for reasons which were already noticeable during calibration (data and model structure errors, identifiability issues, etc.) and (2) by the move from the calibration period to another period leading to the use of less than optimal parameters for this application period [*Merz et al.*, 2011]. In this study, we investigate this second aspect under a wide range of conditions. Separating these two sources of error is essential to achieving an informative evaluation of the extrapolation capacity of hydrological models: a model may work well in calibration but show poor transposability over time.

[26] Let us consider a split sample test where a parameter set θ is transferred from a period D ("donor", i.e., calibration) to a period R ("receiver", i.e., validation). With these notations, the root-mean-square error (RMSE) the Nash-Sutcliffe efficiency (NSE) [*Nash and Sutcliffe*, 1970] and the bias on total volumes can be written as

$$\mathrm{RMSE}_{D\to R} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\hat{Q}_{R,k}[\theta_D] - Q_{R,k}\right)^2} \tag{1}$$

$$\mathrm{NSE}_{D\to R} = 1 - \frac{\sum_{k=1}^{n}\left(\hat{Q}_{R,k}[\theta_D] - Q_{R,k}\right)^2}{\sum_{k=1}^{n}\left(\overline{Q}_R - Q_{R,k}\right)^2} \tag{2}$$

$$\mathrm{BIAS}_{D\to R} = \frac{\sum_{k=1}^{n}\hat{Q}_{R,k}[\theta_D] - \sum_{k=1}^{n}Q_{R,k}}{\sum_{k=1}^{n}Q_{R,k}} \tag{3}$$

in which $Q_{R,k}$ is the observed discharge at time step k on period R, $\hat{Q}_{R,k}[\theta_D]$ the simulated discharge at time step k on period R using the parameter set θ optimized on D, and n is the total number of time steps in period R.

[27] The advantage of the relative formulation of bias is to provide values that are comparable between periods and catchments. In their study on the time stability of parameters, *Merz et al.* [2011] plotted model bias in validation and

in calibration on the same figure to show that their trend on model error was indeed caused by the parameter transfer.

[28] Another way to emphasize the performance losses caused by the parameter transfer is to study the evolving performance from calibration to validation. $RMSE_{D\to R}$ or $NSE_{D\to R}$ values for different D periods but a single R period can be directly compared since all errors are calculated on the same time steps. Differences or ratios can be computed to highlight the quality of a given parameter set compared to another. However, the RMSE is dependent on the mean volume and will tend to be greater for periods (or catchments) showing larger discharges. NSE is built around a ratio between the squared model error and the variance of observed flows. Under certain conditions, one can assume that changes in variance or volumes between periods have a limited impact on the comparison results. Limitations appear when the periods compared show contrasted climatic properties and hence contrasted flow levels. It becomes even more complicated when results from different catchments are analyzed together. Contrary to what is often done in the literature, we decided not to use differences in NSE to conclude on the influence of changes in climate on parameter transferability. Instead, we defined the following model robustness criteria (MRC):

$$MRC_{D\to R} = \frac{\varepsilon_{D\to R}}{\varepsilon_{R\to R}} - 1 \tag{4}$$

where ε is the objective function to be minimized during calibration. The main idea is that the quality of a given parameter set is assessed relative to a reference set, obtained through calibration. $\varepsilon_{D\to R}$ is one estimate of the model error on period R using the parameters calibrated on period D (e.g., $\varepsilon_{D\to R} = RMSE_{D\to R}$). It varies depending on the ability of the parameter set optimized on period D to simulate discharges on period R. $\varepsilon_{R\to R}$ should be the smallest value of ε achievable on period R with the model. $\varepsilon_{D\to R}$ and $\varepsilon_{R\to R}$ are comparable since they are computed on the same "receiver" period. MRC should theoretically be positive. Its interpretation is straightforward. It takes a value of 0 if the parameter set optimized on D gives the same fit it would have if it was calibrated on R. The higher the value, the less suitable the parameter set for the receiving period R. For example, a MRC value of 0.2 means that there has been a 20% error increase due to the use of a transferred parameter instead of the optimal one. Note that a negative value would mean that the parameter set optimized on period D performs better on period R than the reference set optimized over period R. This would be the indication of a problem in parameter optimization on period R, where the global optimum had not been identified properly. Here this happens in a very limited number of calibration runs (less than 1%), indicating that this has only a marginal influence on our results.

[29] The formulation of MRC overcomes most of the difficulties mentioned previously in comparing performances. The only requirement to allow comparing MRC values obtained under various conditions is that the ratio $\varepsilon_{D\to R}/\varepsilon_{R\to R}$ must be independent from the period or catchment characteristics (in terms of volumes or dynamic). For example, $\varepsilon = RMSE$ can be used, whereas $\varepsilon = -NSE$ cannot since a variance term would remain in MRC and results from various catchments could therefore not be mixed. When ε is the mean square error, MRC is a modified version of –NSE in which the benchmark model at the denominator has been changed [see *Lerat et al.*, 2012]. In the NSE formulation, this benchmark is the mean observed flow value over the R period while in MRC, it is the flow simulated by the tested model (e.g., GR4J) using the parameter set optimized on R. Provided ε shows the expected properties mentioned above, MRC is fully comparable over various conditions in terms of climate, catchment scale or dynamic. MRC values should not be significantly affected by the imperfect fit of the model to a specific period (caused by data and model structure errors). Conversely, the influence of using transferred rather than optimized parameters is highlighted. Therefore the climatic extrapolation capacity of a parameter set can be quantified and the results analyzed together on a large number of case studies.

3.3. Methodology for Analyzing the Results

[30] Variations in MRC values were analyzed relative to the differences in climate between the calibration and validation periods, aiming to investigate the potential link between the quality of parameter transfers and the variation in climate from calibration to validation. Changes in climate were expressed as ratios (e.g., 10% less rainfall) or differences (e.g., +1°C). We built graphs where each MRC value was plotted against the corresponding change of the selected climate variable. An example is shown in Figure 4 for the MORDOR6 model and rainfall variations. Figure 4a shows several parameter transfer tests carried out on a single receiving period. Because all values on the x and y axes are relative, the results for all the other receiving periods can be plotted on the same graph. The plotting procedure was then repeated for all the catchments: this provides a large cloud of points, as shown in Figure 4b. To extract the information contained in the graph, the cloud was divided into vertical slices with the same number of points (instead of slices of equal width, which would be less robust). In each slice, the distribution of MRC values is summarized by a box plot (showing the 0.05, 0.25, 0.5, 0.75, and 0.95 percentiles) (see Figure 4c). Because of the lesser density on the left and right sides of the graph, the slices have different widths and the box plots are therefore not evenly spaced. Nevertheless, every point has a corresponding box plot and the voids appearing on the graph are always covered by the neighboring box plot. The vertical spread for each box plot indicates the range of performance loss obtained for the corresponding test conditions. Comparing the relative position of the box plots indicates whether a change in climate from calibration to validation causes loss of robustness due to inappropriate parameter transfers.

[31] To better understand the interpretation of this graph, let us take the example of the box plots shown in Figure 4c. The climate variable used here is mean rainfall. The box plot obtained for about +30% in rainfall (right-hand side of the graph) represents the cases of donor periods that are wetter than the receiver period (i.e., parameter transfer from wet to dry). This box plot is compared to the box plot for 0% change in rainfall. The latter indicates the "usual" performance loss when parameters are transferred under similar climate conditions (median of about 20%). The higher values shown by the box plot on the right indicate

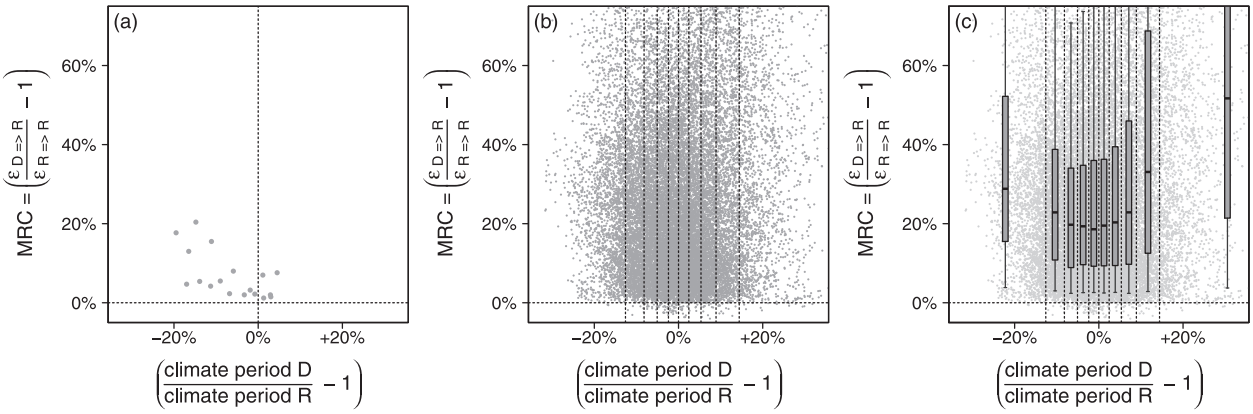


Figure 4. Procedure followed to illustrate the results. The relative loss of performance is plotted against the relative evolution of climate conditions. (a) Dotty plot for a single period, (b) dotty plot for all periods of all catchments, and (c) summary of dotty plots as box plots.

that using a parameter set calibrated on a period 30% wetter than the validation climate is likely to reduce the quality of the model's simulations (here the box plot median increases from 20% to 50%). In a way, this representation puts into perspective the loss of efficiency due to a common transfer and the loss stemming from the climatic extrapolation conditions.

3.4. Crash Test Conditions

[32] The models' parameters were calibrated to minimize the following objective function:

$$\varepsilon = \mathrm{RMSE}[\sqrt{Q}](1 + |\mathrm{Bias}|) \quad (5)$$

[33] The combination of RMSE$[\sqrt{Q}]$ and bias gives weight to dynamic representation as well as water balance. The models' ability to simulate mean runoff is of particular importance in the context of climate change impact studies. Using square-root-transformed flows to compute the RMSE reduces the influence of high flows during calibration and was found to give a good compromise between alternative criteria [*Oudin et al.*, 2006a]. Given the small number of free parameters in the tested models, we used a prior systematic inspection of the parameter space followed by a simple steepest descent local search procedure to determine the most likely zone of convergence. This approach proved efficient for such parsimonious models compared to more complex search algorithms [see *Edijatno et al.*, 1999; *Mathevet*, 2005].

[34] Choosing the subperiod length used in the sampling methodology (see section 3.1) is a difficult task (see, e.g., the discussions by *Yapo et al.* [1996] and *Anctil et al.* [2004]). The calibration period should be long enough to allow for correct parameter determination. At the same time, using overly long periods may play against the study's objectives, as it would reduce the contrast between periods. Also, the number of independent test periods per catchment decreases when the subperiod length increases. We repeated the work with 10 year long and 5 year long periods and will present here the results obtained with the 10 year long calibration periods. Because of the length of available records (more than 30 years) and the high variability of the Australian climate (compared to other parts of the world), using 10 year periods still provides significant differences in mean climate (see section 2.1 and Figure 2) and therefore does not change the conclusions.

[35] The number of tests that could be made on the catchments following the GSST procedure depended on data availability. Subperiods with more than 10% missing values were excluded from the tests. As a result, 12 catchments were not considered from the initial set because of insufficient record length or excessive gaps in data. Therefore, a total of 216 catchments were used in the tests. Fifty tests or more were made for 183 catchments (85% of the set). A maximum of 156 split-sample tests was reached for 134 catchments (62% of the set), corresponding to permutations of 10 year periods over the 1974–2006 period.

3.5. Climate Variables Investigated

[36] Various climate variables can be used in the analysis. Examples include mean annual or seasonal rainfall, PE or temperature, the number of extreme rain or drought events, aridity index, etc. Here we based the analysis on the common variables P, PE, and T with averages computed over the test period. For each split sample test, we determined the changes in mean P, PE and T between calibration and validation periods and plotted these ΔP, ΔPE, and ΔT against each other. Scatter plots were summarized in box plots using the representation introduced in section 4.3. Figure 5 shows the existence of correlations between variables: the link between ΔT and ΔPE is not surprising; their anticorrelation with ΔP indicates that an increase in temperature (or PE) on these catchments generally coincides with a decrease in precipitation.

4. Results

4.1. Calibration Results

[37] We first provide an overview of calibration performance to evaluate the quality of the reference parameter sets. The box plots in Figure 6 show the distribution of calibration performance in terms of NSE calculated on root-square discharges (Figure 6, left) and bias on total volume (Figure 6, right). The calibration results for our tests are shown in black, corresponding to the objective function ε defined in

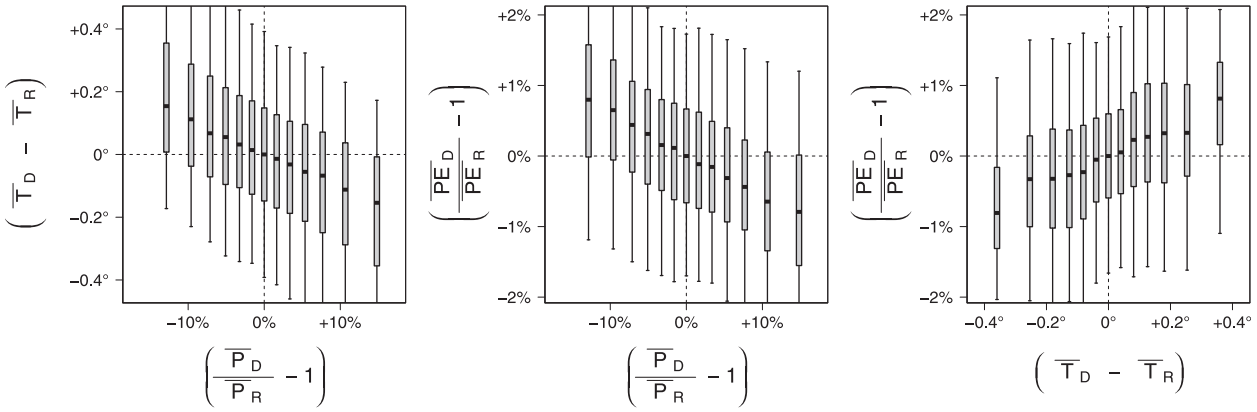


Figure 5. Correlation between variations in P, PE, and T for all the split-sample tests over the entire catchment set.

equation (5). As a source of comparison, we plotted the results obtained for $\varepsilon = \mathrm{RMSE}[\sqrt{Q}]$ in gray. We discussed above how the comparison of NSE values between periods or catchments relies on strong assumptions, not always valid in our context. These results only aim at checking that the models perform reasonably well in calibration. We note the benefit from adding the bias constraint in the objective function and the limited consequences on NSE$[\sqrt{Q}]$ values, which is in agreement with the results found by *Viney et al.* [2009] and the theoretical comments made by *Gupta and Kling* [2011].

4.2. Which Climate Property Causes Problems for Parameter Transfer?

[38] We applied the GSST procedure over 10 year periods for 216 catchments. As in Figure 4, the variations in model robustness criteria (MRC defined in equation (4)) are plotted against the differences in climate between calibration and validation for all catchments. These graphs are provided in Figure 7. To facilitate the analysis of Figure 7, the individual points are not shown. The nine graphs correspond to the three models and three climate variables considered (mean P, PE, and T). The shape of each scatterplot is then analyzed to determine which climatic characteristics influence parameter transferability and to what extent this affects model performance.

[39] First, we note the relatively wide vertical range of the box plots. This indicates that performance losses can be low or high depending on the period or catchment, independently from the mean climate evolution. The center part of each graph gives an overview of the level of error obtained when parameters are transferred under similar climate conditions. We observe a median (thick black line within the box plot) at about 18–20% for GR4J, MORDOR6 and SIMHYD. These values mean that transferring parameters to another period with a similar mean climate leads to an 18%–20% increase in model error ($\varepsilon = \mathrm{RMSE}[\sqrt{Q}](1 + |\mathrm{Bias}|)$) on average compared to calibration. This performance loss is a combination of two aspects. First, there is an incompressible loss of performance when going from calibration to validation due to the model's approximations (inputs, parameters and structure). Second, average conditions such as the total rainfall can remain stable between the two periods, whereas meaningful differences exist (e.g., daily variability in rainfall and runoff).

[40] We can now evaluate whether an additional loss occurs when the climate conditions between the "donor" and "receiver" periods differ significantly. In spite of the vertical spread, trends are visible on the three left graphs in Figure 7 (case of mean rainfall). Considering the large number of points in the scatterplot, these trends on the box

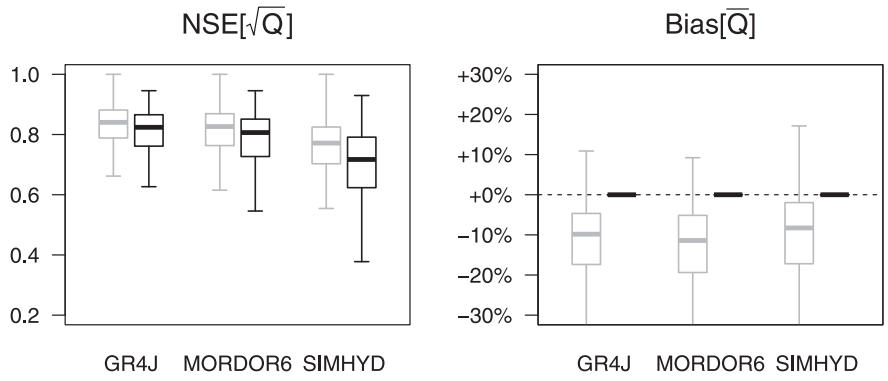


Figure 6. Calibration performance on the entire catchment set with two objective functions: $\varepsilon = \mathrm{RMSE}[\sqrt{Q}]$ (in gray) and $\varepsilon = \mathrm{RMSE}[\sqrt{Q}](1 + |\mathrm{Bias}|)$ (in black).

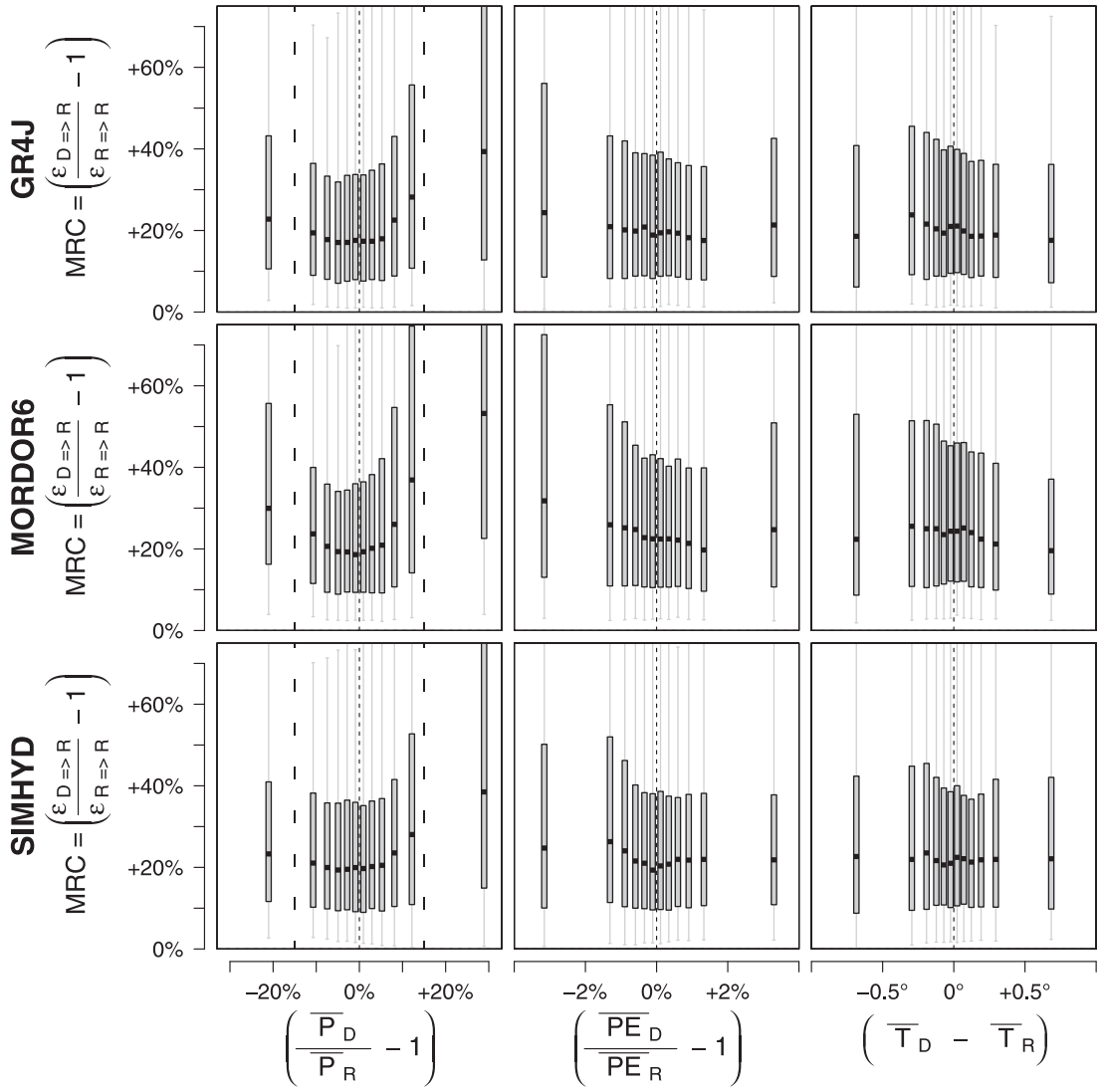


Figure 7. Study of models deficiencies: Performance loss due to the parameter transfer plotted against changes in mean P, PE, and T for the 216 catchments, with $\varepsilon = \mathrm{RMSE}[\sqrt{Q}](1 + |\mathrm{Bias}|)$.

plot medians, but also on the 25th and 75th percentiles, are definitely not the result of isolated cases and should be considered significant, although they are small. This indicates that a link can be established between model performance (estimated through the objective function) and the difference in mean rainfall between calibration and simulation periods. The performance loss is not symmetric for positive and negative changes in mean rainfall. We can observe a greater loss when donor periods are wetter than receiver periods. In addition, the trend seems to be the strongest for MORDOR6 and the weakest for SIMHYD. For example, a +15% difference in mean rainfall leads to an average increase in *MRC* from about 18 to 30% for GR4J, 18 to 40% for MORDOR6 and 20 to 30% for SIMHYD. Interestingly, similar trends are not visible for PE and T, in spite of the correlation observed between ΔP, ΔPE and ΔT (although we notice slightly larger performance losses when $\Delta\mathrm{PE} < 0$). These differences likely result from the fact that the environment in southeast Australia is water limiting (as opposed to energy limiting), which may cause a higher sensitivity of flows to rainfall compared to PE and temperature (sensitivity measured through the objective function).

[41] The results obtained here are in general agreement with the findings reported by *Vaze et al.* [2010b], although the methodology has been modified to provide better robustness on conclusions (more tests and a different analysis procedure). Amplitudes cannot be compared directly but the general shapes are similar: a change in mean rainfall reduces parameter transferability and this transferability seems better from dry to wet than vice versa.

4.3. Consequences of These Deficiencies on Volume Predictions

[42] In this section, we attempt to specifically quantify the ability of models to correctly predict mean runoff over a period in the context of varying climatic conditions. Mean runoff is a basic but nonetheless crucial indicator for

water resources management under current conditions as well as potential future ones.

[43] Bias values have the advantage of being directly comparable. Therefore, we plotted the bias in validation against the climate differences between calibration and validation (results not presented here). We found trends for all three climatic characteristics (P, PE, and T), although they were stronger for changes in rainfall. All models showed a tendency to overestimate flows in the validation period when the calibration period was wetter and cooler and to underestimate flows when the calibration period was dryer and warmer. Nevertheless, no clear conclusion could be drawn on the individual role played by each climate variable because of the dependency observed between ΔP, ΔPE, and ΔT (see Figure 5). In spite of its many advantages, the representation method used to build Figure 7 has one drawback: only one climate characteristic can be analyzed at a time.

[44] To overcome this limitation and to make sure that the results can be compared with studies carried out in alpine regions (i.e., energy-limited conditions), we used a representation where two climate variables could be displayed at a time. We could then determine if only one or both variables influence the validation bias when they vary. Figure 8 shows the median bias on simulated flows plotted for all possible combinations of two climate variables between ΔP, ΔPE, and ΔT. Note that this representation illustrates the models deficiencies in relation with calibration

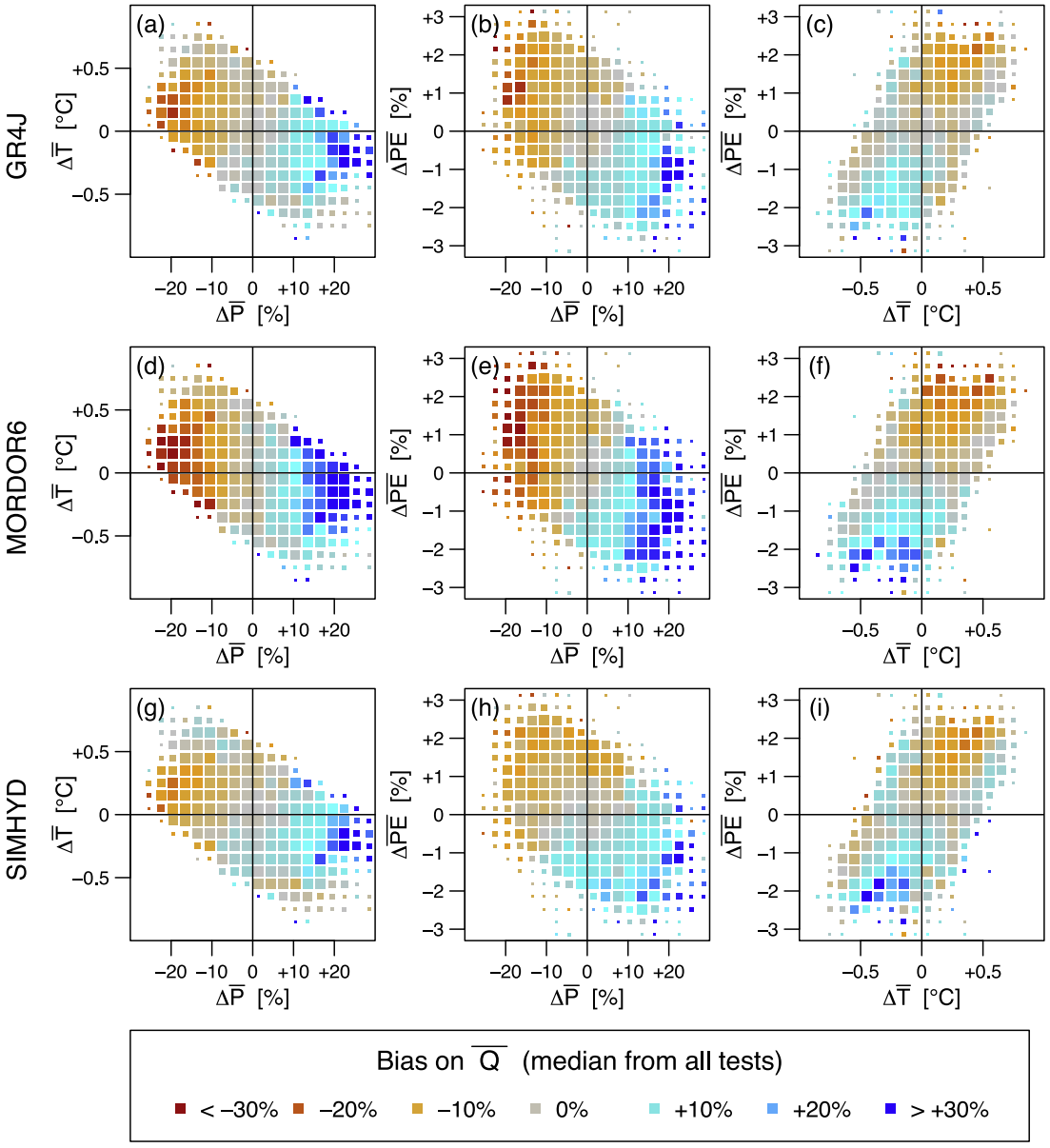


Figure 8. Study of models deficiencies: bias on simulated discharges as a function of changes in P, PE, and T during parameter transfer (median bias from an aggregation of results on 216 catchments; the changes in climate are computed with the climate of the receiver period as reference).

conditions. It should not be interpreted as a plot of the actual flow sensitivity to variations in climate. Compared to the graphs previously presented, this representation also has limitations: (1) only one percentile of the box plots can be shown at a time; (2) the cloud gridding does not ensure an equal number of points in each cell but equal distribution along the *x* and *y* axes. To limit the impact on interpretations of this second limitation, the cell sizes are modified to reflect the number of points available. The size is proportional to the number of points under 50 values and is fixed above 50 values (e.g., few points are available for $\Delta T > 0$ and $\Delta PE < 0$ or for $\Delta T > 0$ and $\Delta P > 0$).

[45] The median bias values obtained for all three models over the 216 catchments show the following variations. If we consider absolute values of bias, we note a symmetrical pattern along the *y* axis for Figures 8a, 8d, and 8g. Indeed, for constant T, the absolute bias increases with changes in P, while the contrary is not true (i.e., no symmetry along the *x* axis). As a result, we can conclude that on average for the data set, changes in P influence the bias more than changes in T. This is to be expected as all the 216 catchments used in this study are in a water-limited area where relative changes in P have larger impacts on runoff than changes in T or PE. The situation is not as clear on Figures 8b, 8e, and 8h (ΔPE and ΔP). Changes in P and PE have a combined effect on the validation bias, although the greatest changes in P ($\pm 15\%$) seem to have more impact than the greatest changes in PE ($\pm 2\%$). Finally, ΔPE and ΔT also have some effect on the validation bias. None of them seems to be more influential than the others (with respect to the points available). A parallel can be made with the results from *Oudin et al.* [2006b], who showed that biased rainfall inputs have a larger impact on models efficiency than biased PE inputs.

[46] The comparison of results between GR4J, MORDOR6 and SIMHYD indicates that even though the intensity varies depending on the model's structure, the general behaviors are similar. This constitutes a difference with the findings of *Vaze et al.* [2010b], who obtained contrary results depending on the model used. Here we observed the smallest and strongest absolute bias for the SIMHYD and MORDOR6 models, respectively, when P changes between calibration and validation. This relation between the change in climate from calibration to validation and the simulation bias was also observed by *Merz et al.* [2011], although the study area was completely different (Australia versus Austria). Further investigations are needed to determine whether these results are related to the model and objective function used in model calibration. On average, we observed that a 20% absolute bias was introduced when mean rainfall differed by 10%–20% and PE differed by 1%–2% between calibration and validation.

4.4. Are Models Deficiencies Similar on All Catchments?

[47] A significant link between climate difference during parameters transfer and bias on simulated discharges was found when the results from the entire data set were analyzed together. However, we do not have information on the homogeneity of this link between different catchments. Interpretations made on a single catchment are always difficult to generalize. However, the use of GSST provides an average of 125 SSTs per catchment and calibrations were made on 10 year periods. Therefore, conditions can be considered sufficient to make a rough analysis at the catchment scale of the consequences that parameter time transfer may have on model efficiency.

[48] For each catchment, a library of validation tests is available and can be classified according to differences in climate between calibration and validation. We can estimate a median value of simulation bias for a specific change in climate (e.g., $+10\%$ in rainfall, $-0.5^{\circ}C$, etc.). A series of maps can then be built for a quick overview of the spatial variability of parameter transferability issues (one map per climate difference). Here we present the results for changes in mean rainfall as they have the greatest impact on bias for our catchment set. Because the results are quite similar for the three models tested, only the results for GR4J are shown here. The maps in Figure 9 are for $\Delta P = -10\%$, 0%, $+10\%$, where each catchment is represented by one symbol. The triangle direction (upward/downward) shows the sign, while the color indicates the intensity of the bias obtained during simulations. Black crosses correspond to cases where this level of rainfall difference was not available for the catchment during the GSST procedure.

[49] In accordance with the previous findings, for a great majority of catchments, we found that simulation bias was close to zero when the mean rainfall was similar between calibration and validation. Mean flows were overestimated when rainfall was greater during calibration than validation and vice versa. Some catchments were exceptions to these generic results with two possible situations: (1) both ΔP and ΔPE affect the simulation bias (see in Figure 8). For example, the combination of $\Delta P = 0$ and $\Delta PE < 0$ can lead to a positive bias or the combination of $\Delta P < 0$ and $\Delta PE > 0$ to a negative bias; (2) other climatic characteristics may differ and affect the results, although these changes are not detected when using mean P, T, and PE; (3) other issues which are catchment specific and unrelated to climate variations may induce robustness loss due to parameter transfer. Despite data verification, cases of changes in input quality and availability or changes in rating curves between periods are never completely avoidable.

[50] Generally, we found that the magnitude of models deficiencies differs between catchments. For some catchments, transferred parameters were suitable for other periods, even with contrasted climate. For other catchments, a clear relation was established between simulation quality (particularly bias on total volume) and the differences in climate between calibration and validation. In spite of various attempts, we could not relate the intensity of these changes with catchment or climate characteristics. However, we noted that the largest models deficiencies occurred on catchments with relatively low annual runoff yield. Besides, as mentioned in section 1.2, temporal changes of nonclimatic characteristics (e.g., land use, quality of input data, etc.) may lead to performance losses, with impacts possibly increasing with the time elapsed between calibration and validation periods. This may contribute to explain the cases of strongest models deficiency, potentially caused by the combined negative impact of differences in climate conditions and in other aspects between periods.

[51] Two examples of the possible situations are presented in Figure 10. These catchments should not be

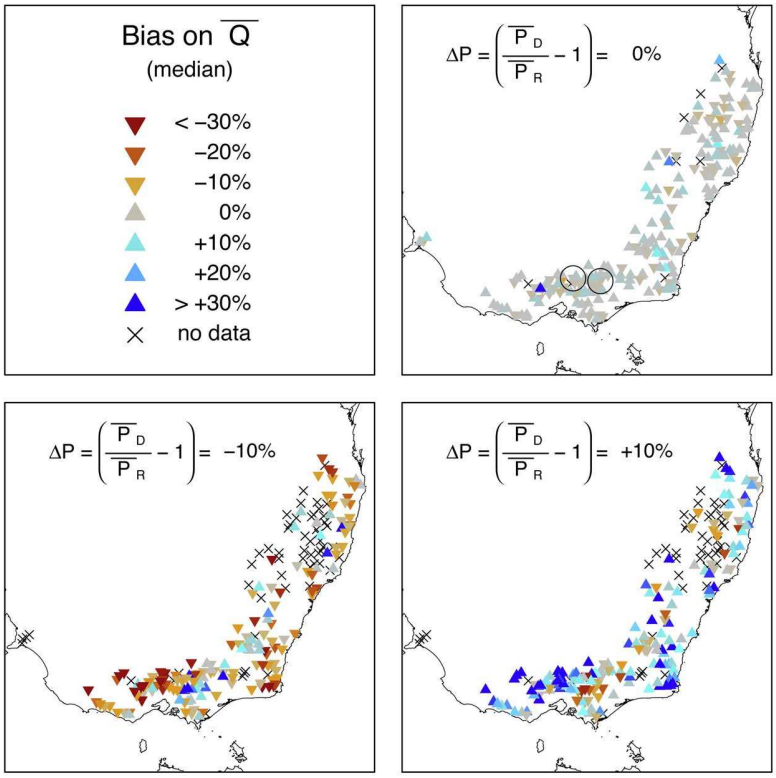


Figure 9. Study of models deficiencies: maps of median bias on discharges simulated by GR4J for $\Delta P = -10\%, 0\%, +10\%$

considered as representing average conditions obtained over two subgroups. Indeed, we obtained a large variety of results depending on the catchment analyzed. In some cases, the simulation quality varied independently from any climatic characteristic, while distinct correlations were found on others. The results for the Rose River at Matong North (179 km^2) and the Pranjip Creek River at Moorilim (818 km^2) catchments are used as an illustration. The location of these two catchments is shown by black circles on the first map in Figure 9. They were chosen because they illustrate some of the contrasted situations we encountered, with similar climate variations. Figure 10 (left) shows the variations of *MRC* with respect to changes in rainfall between calibration and validation. The results for all the calibration-validation combinations are plotted on the graph. The performance losses are clearly stable for the first catchment, while they are greatly influenced by changes in rainfall for the second one. For both catchments, the solid black dots are the results of parameter transfers to the receiver period 1978–1987. The corresponding simulations are illustrated in Figure 10 (right) as mean monthly flows over this 10 year period. The solid and dashed lines correspond to the observation and the reference simulation obtained from calibration, respectively. The envelope with horizontal shading (vertical shading) shows the range of simulated values when the calibration period was drier (wetter) than the validation period. All simulations are relatively similar in the case of the Rose River (the two envelopes even overlap sometimes). Contrary to this, the range of simulated values is very large for the Pranjip River and the curves of the different simulations are positioned in accordance with the climate difference. When classified from most overestimating to most underestimating, the curve order is indeed almost identical to the classification of climate differences between calibration and validation.

5. Discussion

5.1. Methodological Choices

[52] The objective function ε used for these results was RMSE$[\sqrt{Q}](1 + |\text{Bias}|)$. Other criteria such as RMSE[Q] $(1 + |\text{Bias}|)$ or directly RMSE$[\sqrt{Q}]$ and RMSE[Q] were tested. The results are not presented here but the overall shapes obtained in Figures 7 and 8 were similar. While here we focus more on the diagnostic part, studies investigating the reduction of robustness loss during parameter transfer may be the topic of future investigations. Among possible sources are the use of more complex objective functions considering, for example, error heteroscedasticity as suggested by *Thyer et al.* [2009] or *Schoups and Vrugt* [2010].

[53] In addition to keeping the approach simple, we aimed at obtaining an overall and robust view of the topic. As a result, we compared three conceptual models with different structures (although all were lumped) and used a relatively large catchment set. Considering the potential implications of these results, it is indeed important to determine whether the findings for one model over one catchment are isolated or similar in a number of other cases. With the same idea of maximizing the number of points for

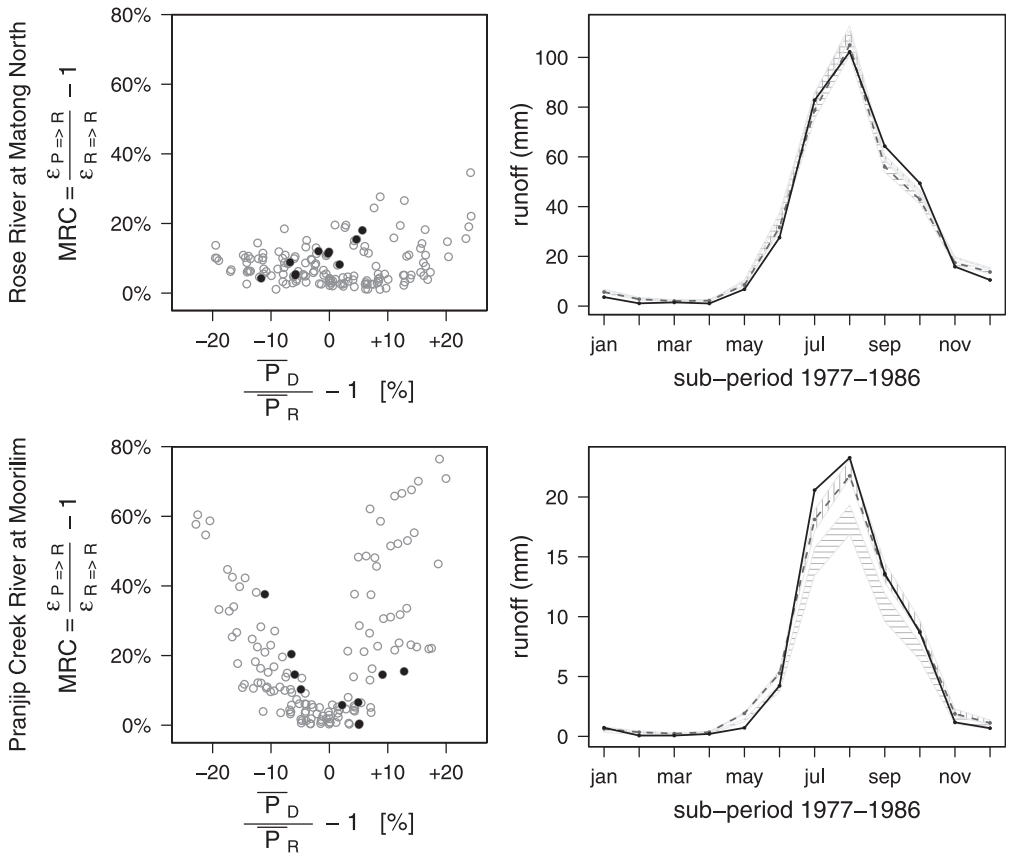


Figure 10. Examples of test results for two catchments: Rose River at Matong North (up) and Pranjip Creek River at Moorilim (bottom). (left) Performance losses during parameter transfer expressed relative to changes in mean rainfall. (right) Range of monthly flows simulated over the 1978–1987 period for various parameter transfers (horizontal and vertical shading indicates the simulation envelope for calibration on wetter or drier climate than validation).

analysis, we authorized periods to overlap in our sampling methodology. As a result, each point will not be entirely independent of the others with which it shares one calibration year or more. Because are more interested in the general picture than computing statistical relation, this does not affect the interpretation of results. Besides, forbidding overlap would require selecting a period, which would reduce the range of climatic contrasts tested for each catchment.

5.2. Results

[54] Using a parameter set under climatic conditions differing from the calibration period can lead to decreased simulation qualities. For instance, we observed performance losses when rainfall changed during parameter transfer and these losses were greater for wet to dry transfers than the other way round. These results corroborate the findings of *Vaze et al.* [2010b] and have direct implications on the use of hydrological models under contrasted climatic conditions (among which, climate change impact studies). Moreover, we observed models deficiencies in the form of an average tendency to overestimate (underestimate) discharges when parameters are optimized in a wetter (drier) climate. This is consistent with the results from *Merz et al.* [2011], although the characteristics of the two catchment sets used differ significantly. High biases were obtained for ranges of climate differences between optimization and application stages similar to possible future climate evolutions in the next decades (as projected by climate models). Since mean flow estimates are crucial indicators in water management plans for current and future conditions, these results should therefore be a particular source of concern. At the same time, we found that the magnitudes of these models deficiencies were not homogeneous but somewhat catchment dependent. This might explain why there is currently no consensus in the literature on this question, but it surely makes the situation more complex for hydrologists.

[55] Further work is needed to understand the hydrological mechanisms behind the robustness issues observed in southeast Australia because of inappropriate parameter transfer. The analysis of the link between parameter values and calibration conditions should contribute to that (this is out of the scope of this paper but will be reported in due course). The possible causes for inappropriate parameter transfer are various (see section 2). Among these, we believe that incorrect simulation of the water budget might be a major one. When transferring parameters, we indeed assume that (1) the adjustments made during parameter calibration provide a satisfactory representation of the water budget and (2) they remain valid over time (i.e., the adjustments made on one period are suitable for another). However, this might not always be true

and could cause increased bias when parameters are transferred. For instance, intercatchment groundwater flows (IGF) can play an important role in the catchment water budget. Even when they are explicitly represented in the models, the IGF formulations are extremely simplified and the associated parameter value remains difficult to identify [*Le Moine et al.*, 2007]. Similarly, the water balance can be strongly affected by errors in inputs that models tend to compensate during calibration to close the water budget. However, these errors might change over time, as the volumes in action vary, which will impact the parameters' transferability if this is not taken into account [*McMillan et al.*, 2011]. Last, the sensitivity of runoff to climate is not straightforward in southeast Australia [*Potter and Chiew*, 2011]. It seems that the RR relationship can sometimes show a form of elasticity, i.e., the prevailing hydrological processes may change between periods of different regimes [*Harman et al.*, 2011]. If this is the case, parameters calibrated under a specific climate would not be representative of the processes that are active under other conditions.

[56] Numerous unknowns remain around the cases of robustness loss sometimes observed when a model is used in a changing climate. The study of other climatic characteristics than mean conditions and the comparison between problematic cases with nonproblematic cases may provide new leads for a better understanding of the phenomena involved. Further research is obviously needed to build models actually able to cope with nonstationary conditions in all catchments. Before this can be achieved, preliminary testing seems preferable to obtain a rough estimation of the parameters' transferability on a given catchment. The procedures proposed in this paper are one way to achieve such testing.

6. Conclusions

[57] Rainfall-runoff models are essential tools for the prediction of river flows. Once calibrated under historical climatic conditions, they are sometimes fed with forcings with different climatic characteristics. This raises questions on the validity of such parameter transfer or what could be called the climatic extrapolation capacity of hydrological models. Following the path opened by *Klemeš* [1986], recent research illustrates how significant it can be, with direct implications on the prediction quality. However, there is no consensus in the literature, as most research was carried out on isolated cases and the results are not always comparable. With this in mind, we proposed a generalized testing procedure (GSST) and the associated analysis methods, with the objective of obtaining robust interpretations on this topic.

[58] When applied to 216 catchments in southeast Australia using three conceptual RR models, this methodology led to the following results. Wide ranges of performance losses were observed between calibration and validation periods. Using the same error criteria as for parameter calibration, we found a tendency toward increased simulation error, with a greater difference in mean rainfall during the parameter transfer, but nothing of similar magnitude for changes in mean PE or temperature. Bias in total volumes was affected by changes in both mean rainfall and PE. We also observed a tendency to overestimate mean runoff when the calibration period was wetter (wet to dry parameter transfer) and to underestimate mean runoff when the calibration period was drier (dry to wet parameter transfer). Even though the tendencies were observed for a majority of the catchments in the data set, we found that their intensity could greatly vary between catchments.

[59] The results obtained here corroborate previous findings obtained by others on large catchment sets to highlight the possible lack of robustness when models are used under a changing climate. They have important implications when using hydrological models as decision-making tools in a wide range of applications (flood risk management, water availability, hydropower, climate change impact studies, etc.). Therefore, we believe that research should be pursued on the improvement of methods to diagnose parameter transferability under a changing climate. Further research is needed to apply similar testing procedures with other models and on different catchment sets (e.g., catchments which are energy limited instead of water limited like those studied herein) with the same objective: determining how relevant the errors in parameter transfers due to climate differences are compared to usual transfer errors in similar conditions. The differences may provide new insights into the behavior of models and hydrological systems in changing conditions. Further research is also needed to analyze what is happening in problematic cases, i.e., what makes the parameters unsuitable on a different period and look for solutions to reduce these robustness losses by means of model structure adjustment, the choice of objective function and/or constraints during parameter optimization. All these constitute exciting challenges for the coming decade on prediction under change that will be launched by the International Association of Hydrological Sciences.

[60] **Acknowledgments.** This study was part of a collaboration between research teams from LNHE, EDF R&D (France), CSIRO Land and Water (Australia), and Irstea HBAN (France). The authors would like to thank EDF R&D for supporting this study and CSIRO Land and Water for hosting the first author of the study for a 5 month period and for making available the large data sets used. The constructive review comments made by A. Efstratiadis, J. C. Refsgaard, and an anonymous reviewer on the first version of the manuscript helped improve the text and are gratefully acknowledged.

References

Anctil, F., C. Perrin, and V. Andréassian (2004), Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models, *Environ. Modell. Software*, *19*(4), 357–368, doi:10.1016/S1364-8152(03)00135-X.

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M.-H. Ramos, and A. Valéry (2009), Crash tests for a standardized evaluation of hydrological models, *Hydrol. Earth Syst. Sci.*, *13*(10), 1757–1764, doi:10.5194/hess-13-1757-2009.

Arnell, N. W. (2011), Uncertainty in the relationship between climate forcing and hydrological response in UK catchments, *Hydrol. Earth Syst. Sci.*, *15*(3), 897–912, doi:10.5194/hess-15-897-2011.

Bastola, S., C. Murphy, and J. Sweeney (2011), Evaluation of the transferability of hydrological model parameters for simulations under changed climatic conditions, *Hydrol. Earth Syst. Sci. Discuss.*, *8*(3), 5891–5915, doi:10.5194/hessd-8-5891-2011.

Caballero, Y., S. Voirin-Morel, F. Habets, J. Noilhan, P. LeMoigne, A. Lehenaff, and A. Boone (2007), Hydrological sensitivity of the Adour-Garonne river basin to climate change, *Water Resour. Res.*, *43*, W07448, doi:10.1029/2005WR004192.

Chahinian, N., et al. (2006), Compilation of the MOPEX 2004 results, in *Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment-MOPEX*, edited by V. Andréassian et al., *IAHS Publ., 301*, 313–338.

Chiew, F. H. S., M. C. Peel, and A. W. Western (2002), Application and testing of the simple rainfall-runoff model SIMHYD, in *Mathematical Models of Small Watershed Hydrology and Applications*, edited by V. P. Singh and D. K. Frevert, pp. 335–367, Water Resour. Publ., Littleton, Colo.

Chiew, F. H. S., et al. (2008), Rainfall-runoff modelling across the Murray-Darling Basin: A report to the Australian government from the CSIRO Murray-Darling Basin Sustainable Yields Project, CSIRO, Canberra.

Chiew, F. H. S., J. Teng, J. Vaze, D. A. Post, J. M. Perraud, D. G. C. Kirono, and N. R. Viney (2009), Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, *Water Resour. Res.*, *45*, W10414, doi:10.1029/2008WR007338.

Choi, H. T., and K. Beven (2007), Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *J. Hydrol.*, *332*(3–4), 316–336, doi:10.1016/j.jhydrol.2006.07.012.

Coron, L., V. Andréassian, M. Bourqui, C. Perrin, and F. Hendrickx (2011), Pathologies of hydrological models used in changing climatic conditions: A review, in *IUGG 2011—Hydro-climatology: Variability and Change, IAHS Publ.*, *344*, 39–44.

de Vos, N. J., T. H. M. Rientjes, and H. V. Gupta (2010), Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering, *Hydrol. Processes*, *24*, 2840–2850, doi:10.1002/hyp.7698.

Donnelly-Makowecki, L., and R. Moore (1999), Hierarchical testing of three rainfall-runoff models in small forested catchments, *J. Hydrol.*, *219*(3–4), 136–152, doi:10.1016/S0022-1694(99)00056-6.

Edijatno, N. D. O. Nascimento, X. Yang, Z. Makhlouf, and C. Michel (1999), GR3J: A daily watershed model with three free parameters, *Hydrol. Sci. J.*, *44*, 263–277, doi:10.1080/02626669909492221.

Garçon, R. (1996), Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. Bilan de l'année 1994–1995, *La Houille Blanche*, *5*, 71–76, doi:10.1051/lhb/1996056.

Görgen, K., et al. (2010), Assessment of Climate Change Impacts on Discharge in the Rhine River Basin: Results of the RheinBlick2050 Project, *CHR Rep. I-23*, Int. Comm. for the Hydrol. of the Rhine Basin, Lelystad, Netherlands.

Gupta, H. V., and H. Kling (2011), On typical range, sensitivity, and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics, *Water Resour. Res.*, *47*, W10601, doi:10.1029/2011WR010962.

Harman, C. J., P. A. Troch, and M. Sivapalan (2011), Functional model of water balance variability at the catchment scale: 2. Elasticity of fast and slow runoff components to precipitation change in the continental United States, *Water Resour. Res.*, *47*, W02523, doi:10.1029/2010WR009656.

Kay, A., H. Davies, V. Bell, and R. Jones (2009), Comparison of uncertainty sources for climate change impacts: Flood frequency in England, *Clim. Change*, *92*(1–2), 41–63, doi:10.1007/s10584-008-9471-4.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*(1), 13–24, doi:10.1080/02626668609491024.

Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics and uncertainty, *J. Am. Water Resour. Assoc.*, *47*(3), 481–495, doi:10.1111/j.1752-1688.2011.00543.x.

Le Moine, N., V. Andréassian, C. Perrin, and C. Michel (2007), How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, *Water Resour. Res.*, *43*, W06428, doi:10.1029/2006WR005608.

Lerat, J., V. Andréassian, C. Perrin, J. Vaze, J.-M. Perraud, P. Ribstein, and C. Loumagne (2012), Do internal flow measurements improve the calibration of semidistributed rainfall-runoff models? *Water Resour. Res.*, *48*, W02511, doi:10.1029/2010WR010179.

Mathevet, T. (2005), Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparaison de modèle sur un large échantillon de bassins versants, PhD thesis, L'École Natl. du Génie Rural, des Eaux et des For., Paris.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, *400*, 83–94, doi:10.1016/j.jhydrol.2011.01.026.

Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*, W02531, doi:10.1029/2010WR009505.

Morton, F. (1983), Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, *J. Hydrol.*, *66*, 1–76, doi:10.1016/0022-1694(83)90177-4.

Murphy, C., R. Fealy, R. Charlton, and J. Sweeney (2006), The reliability of an "off-the-shelf" conceptual rainfall runoff model for use in climate impact assessment: Uncertainty quantification using Latin hypercube sampling, *Area*, *38*(1), 65–78, doi:10.1111/j.1475-4762.2006.00656.x.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I. A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

Niel, H., J. E. Paturel, and E. Servat (2003), Study of parameter stability of a lumped hydrologic model in a context of climatic variability, *J. Hydrol.*, *278*(1–4), 213–230, doi:10.1016/S0022-1694(03)00158-6.

Oudin, L., V. Andréassian, T. Mathevet, C. Perrin, and C. Michel (2006a), Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, *Water Resour. Res.*, *42*, W07410, doi:10.1029/2005WR004636.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006b), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *J. Hydrol.*, *320*(1–2), 62–83, doi:10.1016/j.jhydrol.2005.07.016.

Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, *44*, W03413, doi:10.1029/2007WR006240.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*(1–4), 275–289, doi:10.1016/S0022-1694(03)00225-7.

Perrin, C., V. Andréassian, C. Rojas-Serna, T. Mathevet, and N. Le Moine (2008), Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, *Water Resour. Res.*, *44*, W08447, doi:10.1029/2007WR006579.

Potter, N. J., and F. H. S. Chiew (2011), An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models, *Water Resour. Res.*, *47*, W00G10, doi:10.1029/2010WR010333.

Prudhomme, C., and H. Davies (2009), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: Baseline climate, *Clim. Change*, *93*(1), 177–195, doi:10.1007/s10584-008-9464-3.

Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*(7), 2189–2202.

Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, and G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106, doi:10.1029/2009JD012035.

Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, *46*, W10531, doi:10.1029/2009WR008933.

Seibert, J. (2003), Reliability of model predictions outside calibration conditions, *Nord. Hydrol.*, *34*(5), 477–492.

Singh, R., T. Wagener, K. van Werkhoven, M. E. Mann, and R. Crane (2011), A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate—Accounting for changing watershed behavior, *Hydrol. Earth Syst. Sci.*, *15*(11), 3591–3603, doi:10.5194/hess-15-3591-2011.

Steele-Dunne, S., P. Lynch, R. McGrath, T. Semmler, S. Wang, J. Hanafin, and P. Nolan (2008), The impacts of climate change on hydrology in Ireland, *J. Hydrol.*, *356*(1–2), 28–45, doi:10.1016/j.jhydrol.2008.03.025.

Tan, K. S., F. H. S. Chiew, R. B. Grayson, P. J. Scanlon, and L. Siriwardena (2005), Calibration of a daily rainfall-runoff model to estimate high daily flows, in *MODSIM 2005 Int. Congr. on Modelling and Simulation*, edited by A. Zerger and R. M. Argent, pp. 2960–2966, Modell. Simul. Soc. of Australia and New Zealand, Melbourne, Australia.

Teng, J., J. Vaze, F. H. S. Chiew, B. Wang, and J.-M. Perraud (2011), Estimating the relative uncertainties sourced from GCMs and hydrological models in modelling climate change impact on runoff, *J. Hydrometeorol.*, *13*, 122–139, doi:10.1175/JHM-D-11-058.1.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.

Vaze, J., and J. Teng (2011), Future climate and runoff projections across New South Wales, Australia: Results and practical applications, *Hydrol. Processes*, *25*(1), 18–35, doi:10.1002/hyp.7812.

Vaze, J., F. H. S. Chiew, J.-M. Perraud, N. R. Viney, D. A. Post, J. Teng, B. Wang, J. Lerat, and M. Goswami (2010a), Rainfall-runoff modelling across southeast Australia: Datasets, models and results, *Aust. J. Water Resour.*, *14*(2), 101–116.

Vaze, J., D. A. Post, F. H. S. Chiew, J. M. Perraud, N. R. Viney, and J. Teng (2010b), Climate nonstationarity—Validity of calibrated rainfall-runoff models for use in climatic changes studies, *J. Hydrol.*, *394*(3–4), 447–457, doi:10.1016/j.jhydrol.2010.09.018.

Vicuna, S., and J. A. Dracup (2007), The evolution of climate change impact studies on hydrology and water resources in California, *Clim. Change*, *82*, 327–350, doi:10.1007/s10584-006-9207-2.

Viney, N. R., J.-M. Perraud, J. Vaze, F. H. S. Chiew, D. A. Post, and A. Yang (2009), The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, paper presented at *18th World IMACS/MODSIM Congr.*, Modell. Simul. Soc. of Australia and New Zealand, Cairns, Australia.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Wilby, R. L. (2005), Uncertainty in water resource model parameters used for climate change impact assessment, *Hydrol. Processes*, *19*(16), 3201–3219, doi:10.1002/hyp.5819.

Wilby, R. L., and I. Harris (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, *42*, W02419, doi:10.1029/2005WR004065.

Xu, C. (1999), Operational testing of a water balance model for predicting climate change impacts, *Agric. For. Meteorol.*, *98-99*, 295–304, doi:10.1016/S0168-1923(99)00106-9.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, *181*(1–4), 23–48, doi:10.1016/0022-1694(95)02918-4.

F

Compléments pour le chapitre 4

Présentation des illustrations complémentaires proposées ici

Les figures F.1 à F.4 constituent des compléments à la section 4.3. Elles représentent les températures et précipitations journalières interannuelles pour les trois périodes d'études et la dispersion causée sur celles-ci par un changement de GCM, de méthode de descente d'échelle statistique (*MDES*) ou de réalisation analogue pour la MDES considérée.

Les figures F.5 et F.6 constituent quant à elles des compléments à la section 4.4, concernant un exercice de remplacement de *GR4J-CemaNeige* par d'autres modèles hydrologiques dans la chaîne servant de référence pour nos analyses. Plus précisément, ces figures illustrent la sensibilité des projections absolues, faites par différents modèles hydrologiques, au choix de la réalisation analogue et au choix de la période de calage.

Compléments sur la sensibilité des forçages climatiques au choix du GCM

FIGURE F.1.: Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* avec la chaîne de référence choisie

bleu : simulation sur la période de 30 ans pour la chaîne de modélisation considérée, rouge pointillé : 1979-2008 observations

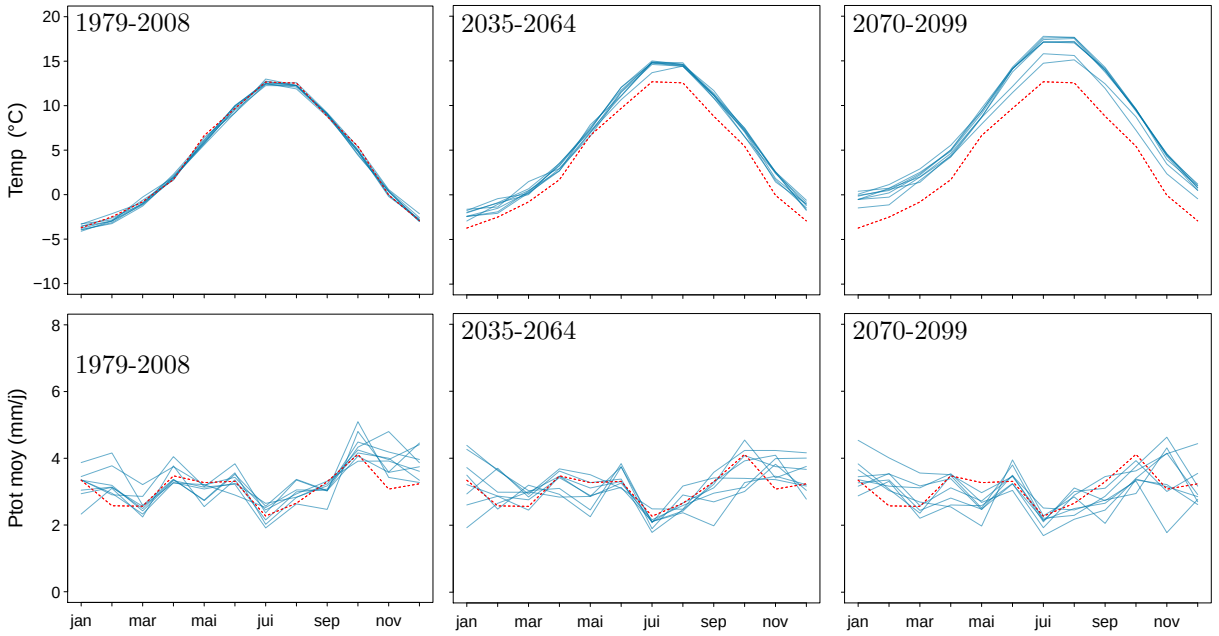


FIGURE F.2.: Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon le *GCM* utilisé (toutes choses égales par ailleurs) *code couleur fig. F.1*

Compléments sur la sensibilité des forçages climatiques au choix de la MDES

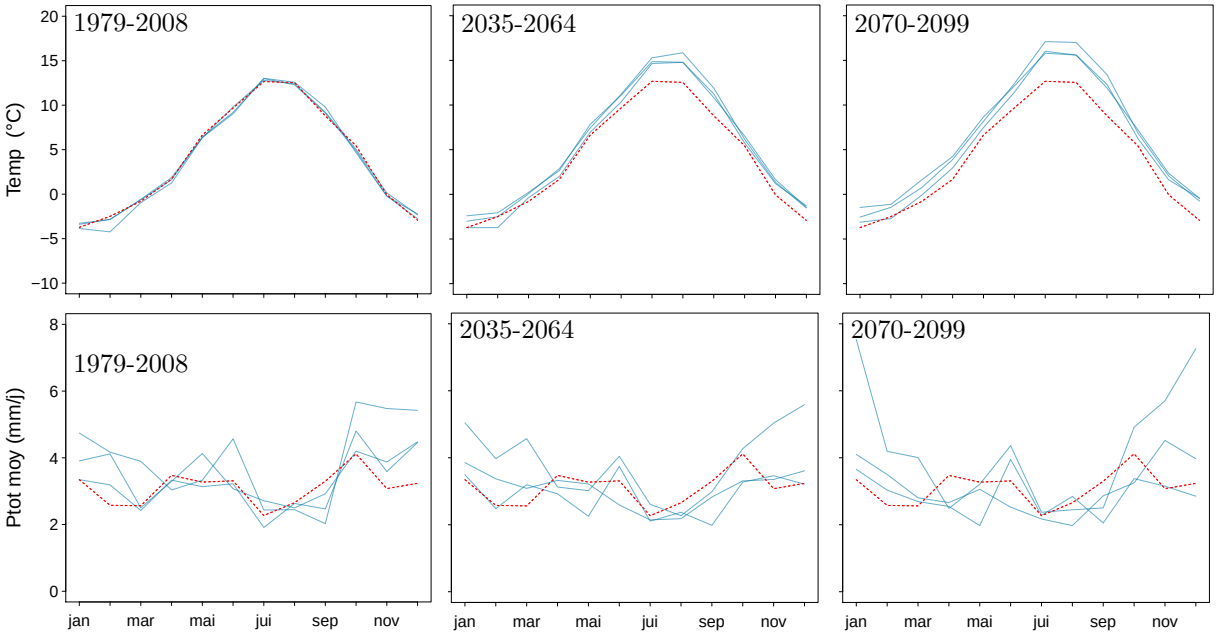


FIGURE F.3.: Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon la *MDES* utilisée (toutes choses égales par ailleurs) *code couleur fig. F.1*

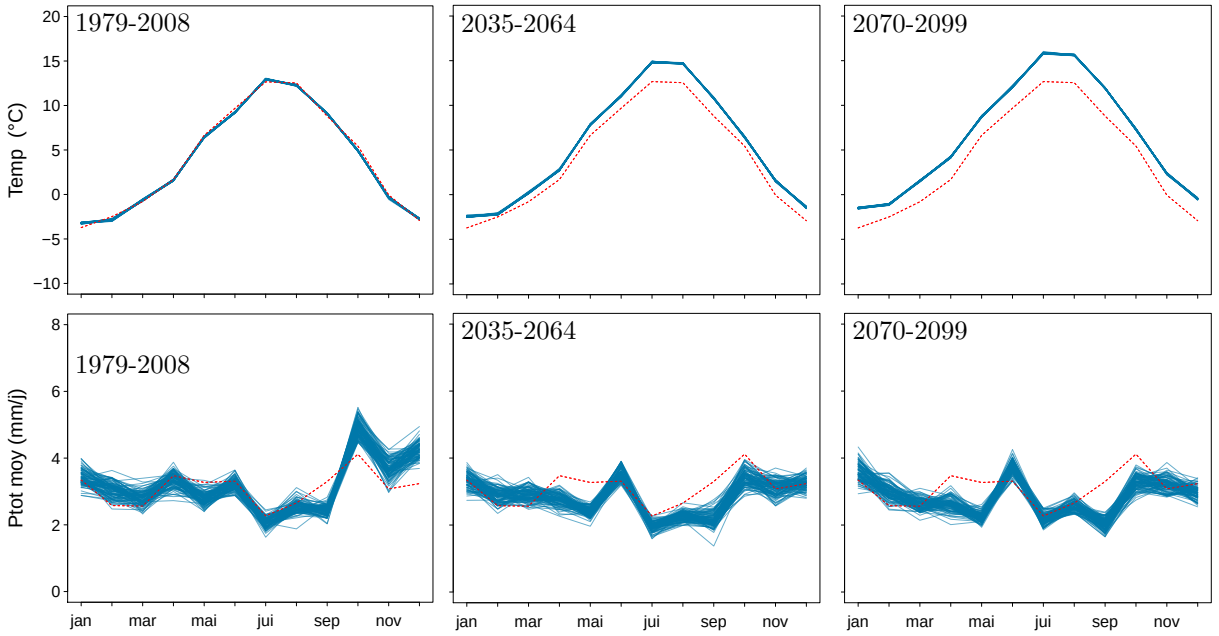


FIGURE F.4.: Écarts de simulation des $\overline{T}_{mens.}$ et $\overline{P}_{mens.}$ de *la Durance à Embrun* selon la réalisation analogue utilisée (toutes choses égales par ailleurs)

Compléments sur la sensibilité des projections absolues au choix de la réalisation analogue pour les différents modèles hydrologiques

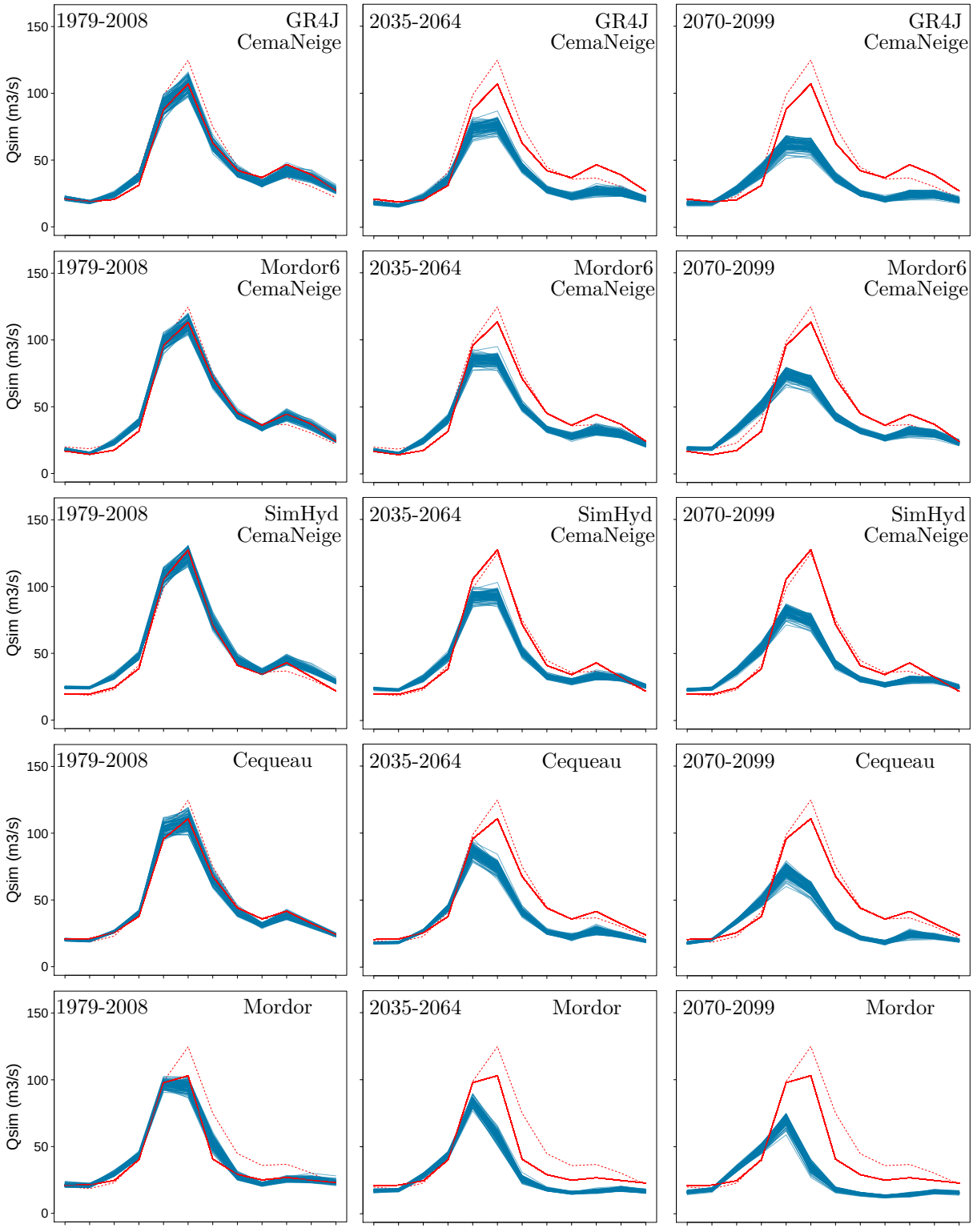


FIGURE F.5.: Écarts de simulation des débits de *la Durance à Embrun* selon la réalisation analogue utilisée pour la *MDES* (toutes choses égales par ailleurs)

Compléments sur la sensibilité des projections absolues au choix de la période de calage pour les différents modèles hydrologiques

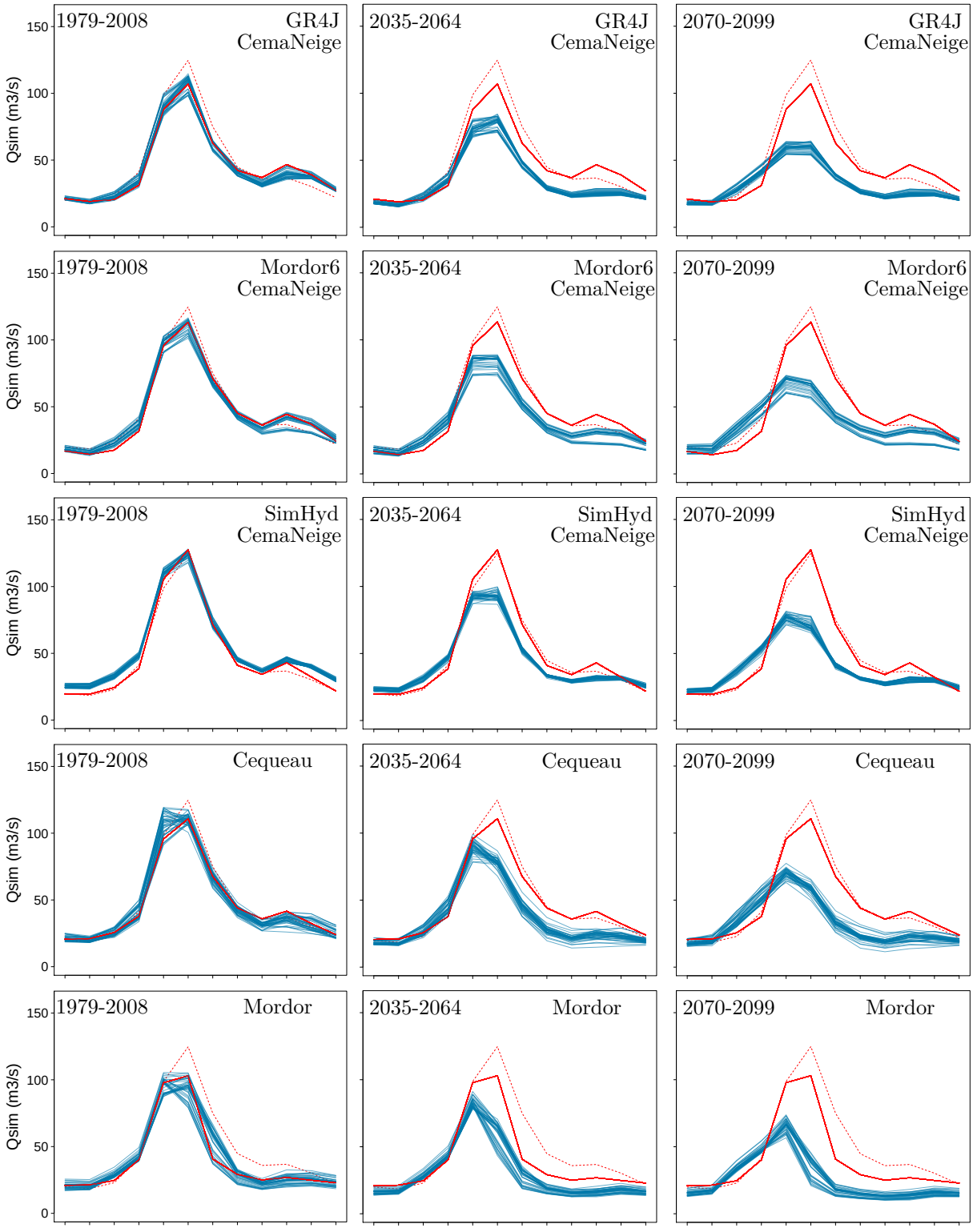


FIGURE F.6.: Écarts de simulation des débits de *la Durance à Embrun* selon la période de calage de chaque modèle hydrologique (toutes choses égales par ailleurs)

G

Compléments pour le chapitre 5

G.1. Compléments sur les tests de bootstrap

Les éléments ci-dessous viennent compléter la section 5.4.2 dans laquelle une approche de type *bootstrap* est utilisée pour conduire une analyse des liens entre les erreurs de modélisation et les espacements temporels entre les périodes de calage et de validation.

Procédure d'échantillonnage des années pour nos tests

En considérant des chroniques de N années disponibles sur un bassin donné et une longueur de périodes tests de 10 années, la procédure d'échantillonnage utilisée est la suivante :

1. Nous construisons le vecteur d'indice des années [1, 2, 3, ..., N-2, N-1, N]
2. Nous effectuons un tirage aléatoire de 10 valeurs d'indice parmi ce vecteur (par exemple $k = 1$ et $candidat\,[k] = [3, 28, 13, 51, 38, 46, 8, 22, 45, 30]$).
3. Nous calculons alors :
   3.a. La caractéristique climatique étudiée sur ces 10 années ($CaraClim(k)$)
   Dans cet exemple, nous considérons $\overline{T}$: $CaraClim(k) = moy\left(\overline{T}_{an.}[candidat[k]]\right)$
   3.b. Un indicateur de l'espacement entre ces années ($IndEsp(k)$)
   $IndEsp(k) = ecart_type\,(candidat(k)–min\,(candidat\,[k]))$
4. Nous répétons les étapes 1 à 3 un très grand nombre de fois (ex : 500 000) et nous construisons un nuage de points illustrant pour chaque *candidat* les valeurs correspondantes de $CaraClim$ (ici sur l'axe x) et $IndEsp$ (ici sur l'axe y). Un exemple de ce type de nuage est fourni sur la figure G.1. Les points situés en bas représentent les *candidats* pour lesquels les années sont très proches les unes des autres et ceux situés en haut, à l'inverse les cas où les années sont temporellement réparties le long de la chronique totale. Nous notons que le nuage a ici tendance à se resserrer en haut ce qui signifie qu'il est difficile d'avoir un *candidat* regroupant des années à la fois très froides et réparties temporellement. Ce comportement n'est pas systématique et on imagine aisément la situation inverse, dans laquelle la sélection d'années spécifiques permet de constituer un agglomérat dont le climat moyen est plus extrême.
5. Nous normalisons les vecteurs $CaraClim$ et $IndEsp$ pour qu'ils varient sur un intervalle de largeur 1, centré sur 0 (variations sur $[-0.5; +0.5]$).
   Puis nous construisons les deux classements suivants :
   $classement\,[-\alpha \cdot CaraClimNorm + (1-\alpha) \cdot IndEspNorm]$ // période froide
   $classement\,[+\alpha \cdot CaraClimNorm + (1-\alpha) \cdot IndEspNorm]$ // période chaude
   Ils ordonnent les *candidats* selon le compromis qu'ils représentent entre nos deux critères (climat extrême et répartition temporelle des années). Le critère alpha permet d'ajuster le poids relatif donné à l'un des critères comparativement à l'autre. Pour notre exemple, avec alpha=0.5, les 100 meilleurs *candidats* de chaque cas sont représentés en vert dans la figure G.1.

6. Enfin, nous identifions parmi ces meilleurs *candidats* les couples qui sont indépendants (pas d'année en commun) et nous sélectionnons l'un d'entre eux. Cette sélection est visible sur la figure G.1, où les *candidats* "froids" et "chauds" sont représentés en bleu et rouge sur le nuage de points, tandis que le graphique de droite permet de visualiser la répartition temporelle des années qui les composent.

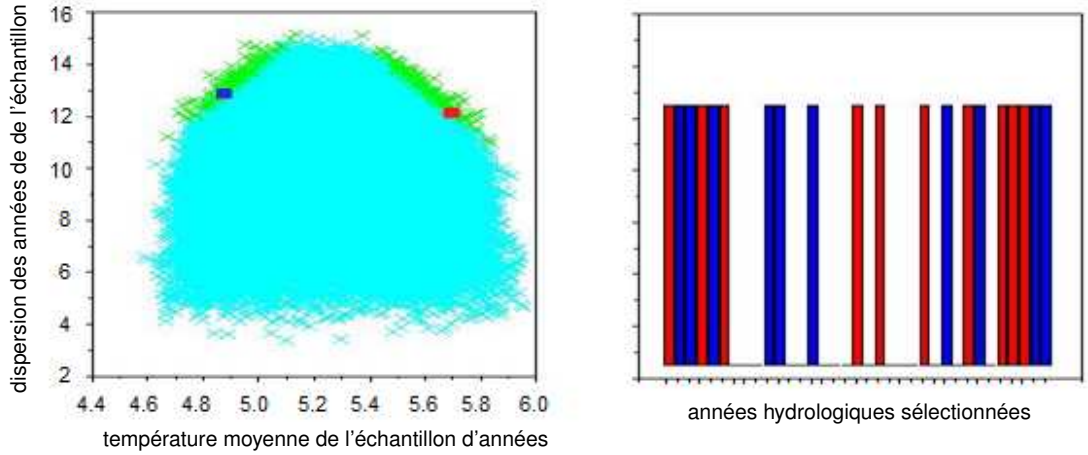


FIGURE G.1.: Méthode de sélection des années pour l'approche *bootstrap*
Totalité des groupes d'années sélectionnés en bleu clair, rééchantillonages intéressants en vert, couple retenu en bleu foncé (bloc froid) et en rouge (bloc chaud)

Exemple sur un bassin

Des tests de bootstrap sont menés sur le bassin de *la Dourbie à Nant* où d'importants biais sont constatés lors des transferts de paramètres, en lien avec les changements de température moyenne entre les périodes de calage et de validation. Divers couples de périodes sont sélectionnés selon la procédure de bootstrap décrite ci-dessus et soumis à l'exercice de transfert de paramètres. Les résultats ainsi obtenus sont illustrés sur la figure G.2. Les points rouges sur le graphique de droite correspondent aux nouveaux tests par *bootstrap*, tandis que le nuage de points gris correspond aux résultats précédents issus du *GSST*. Les deux nuages ainsi constitués étant en accord, on peut affirmer que la situation constatée via *GSST* (simulations biaisées à la suite des transferts de paramètres) n'est pas le résultat du $\Delta temporel$ mais bien celui du $\triangle\overline{T}$. Cette affirmation est bien entendu valable pour ce bassin uniquement et des comparaisons similaires doivent être réalisées pour en déduire une conclusion générale. Ces résultats vont toutefois dans le même sens que notre intuition selon laquelle les causes des problèmes de robustesse des modèles mis en avant par le *GSST* ne se résument pas une simple dérive temporelles des conditions de test.

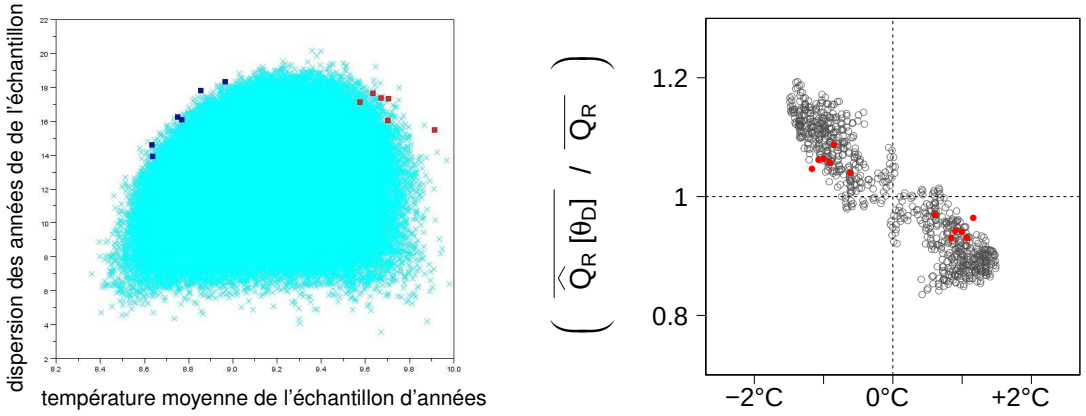


FIGURE G.2.: Illustration d'un test de bootstrap sur le bassin de *la Dourbie à Nant*

Résultats obtenus avec GR4J-CemaNeige (θ_{neige} fixés) avec des données SPAZM et une ETP d'Oudin. Couples sélectionnés en rouge et bleu foncé à gauche et résultats des points obtenus en rouge à droite.

G.2. Compléments sur les analyses des variations des paramètres optimaux

Les éléments fournis ici viennent compléter la section 5.5 où sont conduites des analyses des variations temporelles des paramètres optimaux des modèles. Ces variations sont analysées au regard des variations de conditions climatiques. Plus spécifiquement, nous tentons ici d'établir des parallèles entre $\overline{\widehat{Q}}/\overline{Q} = f(\triangle caract.)$ et $\triangle param. = f(\triangle caract.)$

Modèle *GR4J*

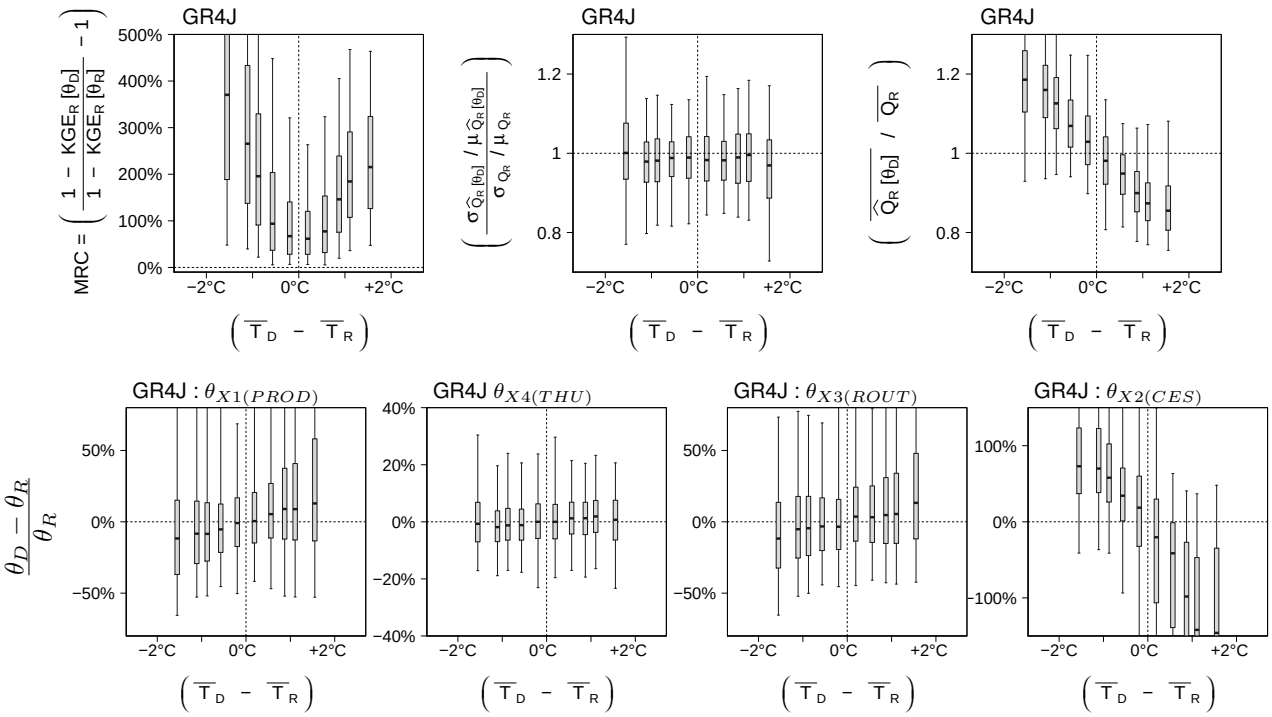


FIGURE G.3.: Biais de *GR4J* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre $\triangle$erreurs et $\triangle\overline{T}$ $(r'_s[MRC] > 0.5)$

GR4J est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

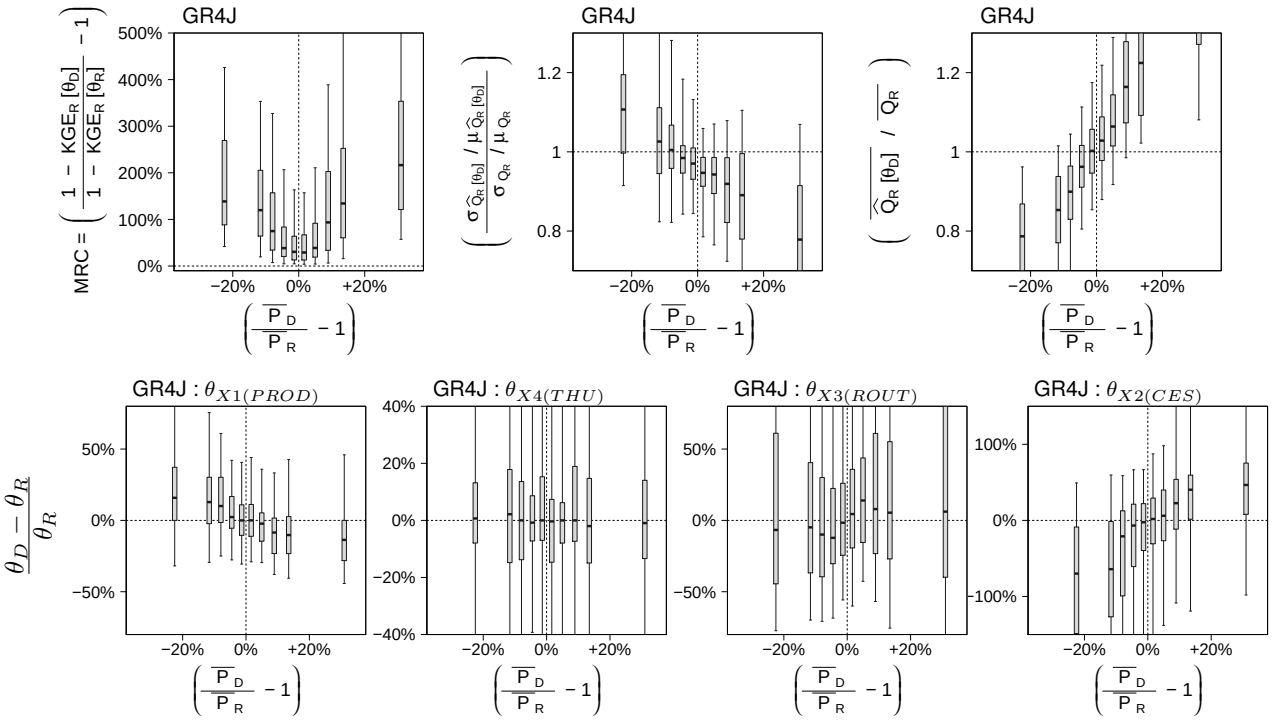


FIGURE G.4.: Biais de *GR4J* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5 \;\&\; r_s[biais] > 0.5)$

GR4J est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

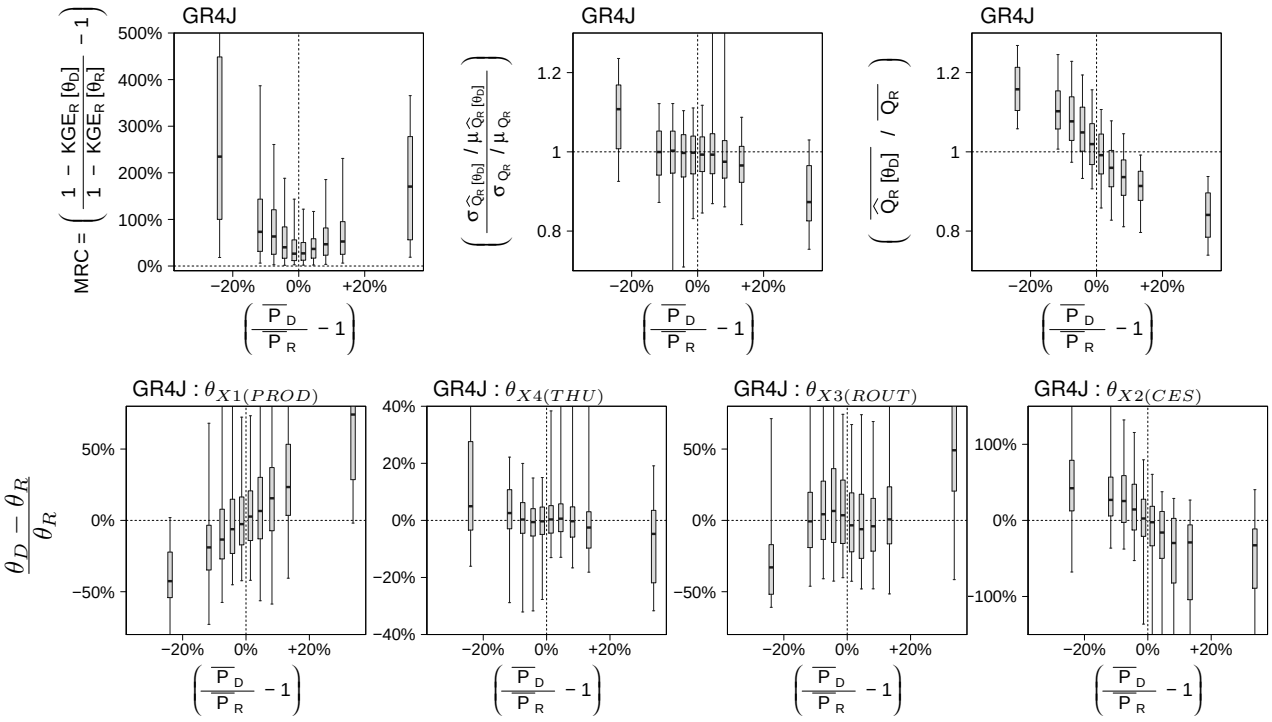


FIGURE G.5.: Biais de *GR4J* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre $\triangle$erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5 \;\&\; r_s[biais] < -0.5)$

GR4J est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

Modèle *Mordor6*

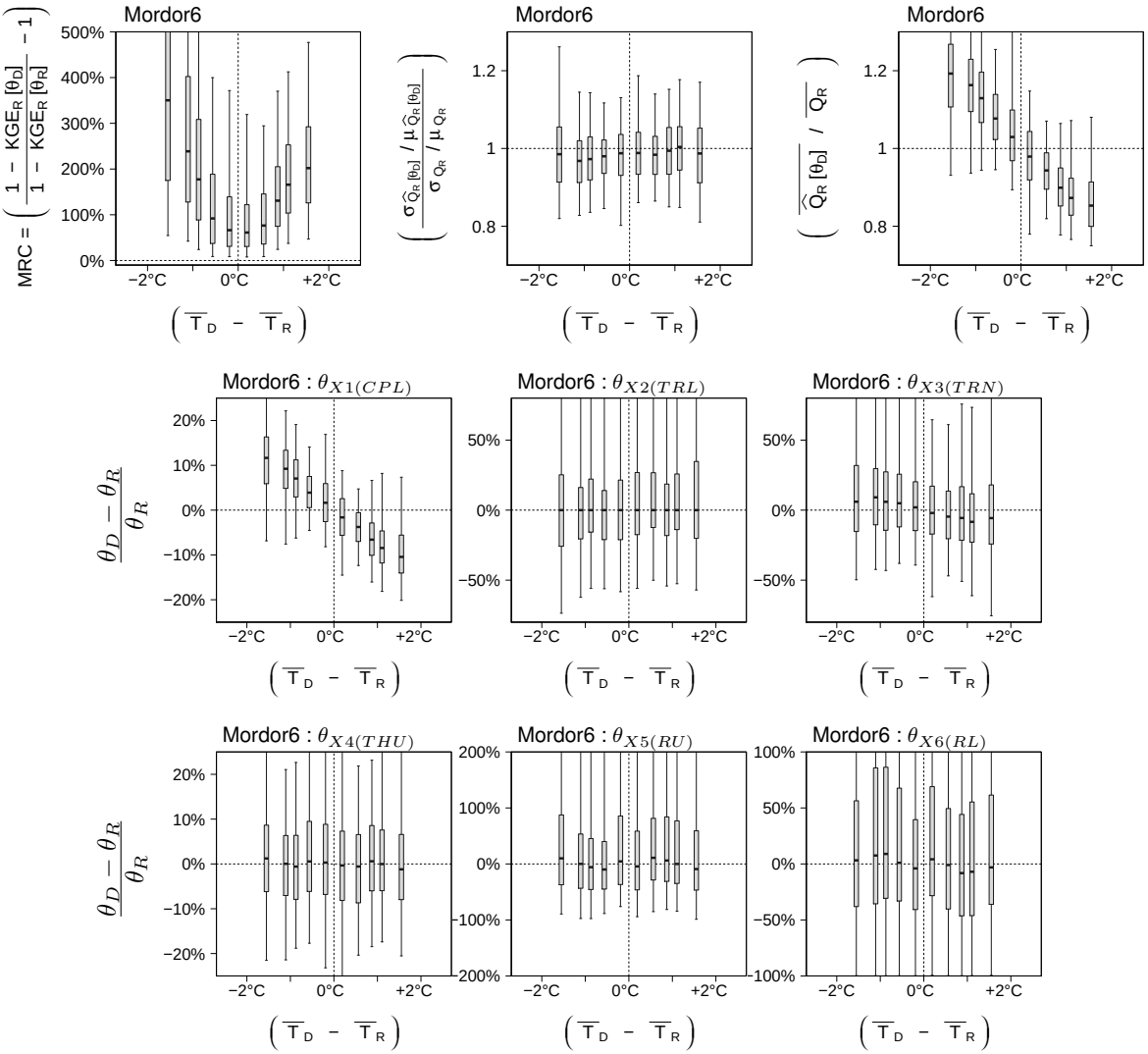


FIGURE G.6.: Biais de *Mordor6* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{T}$ $(r'_s[MRC] > 0.5)$

Mordor6 est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

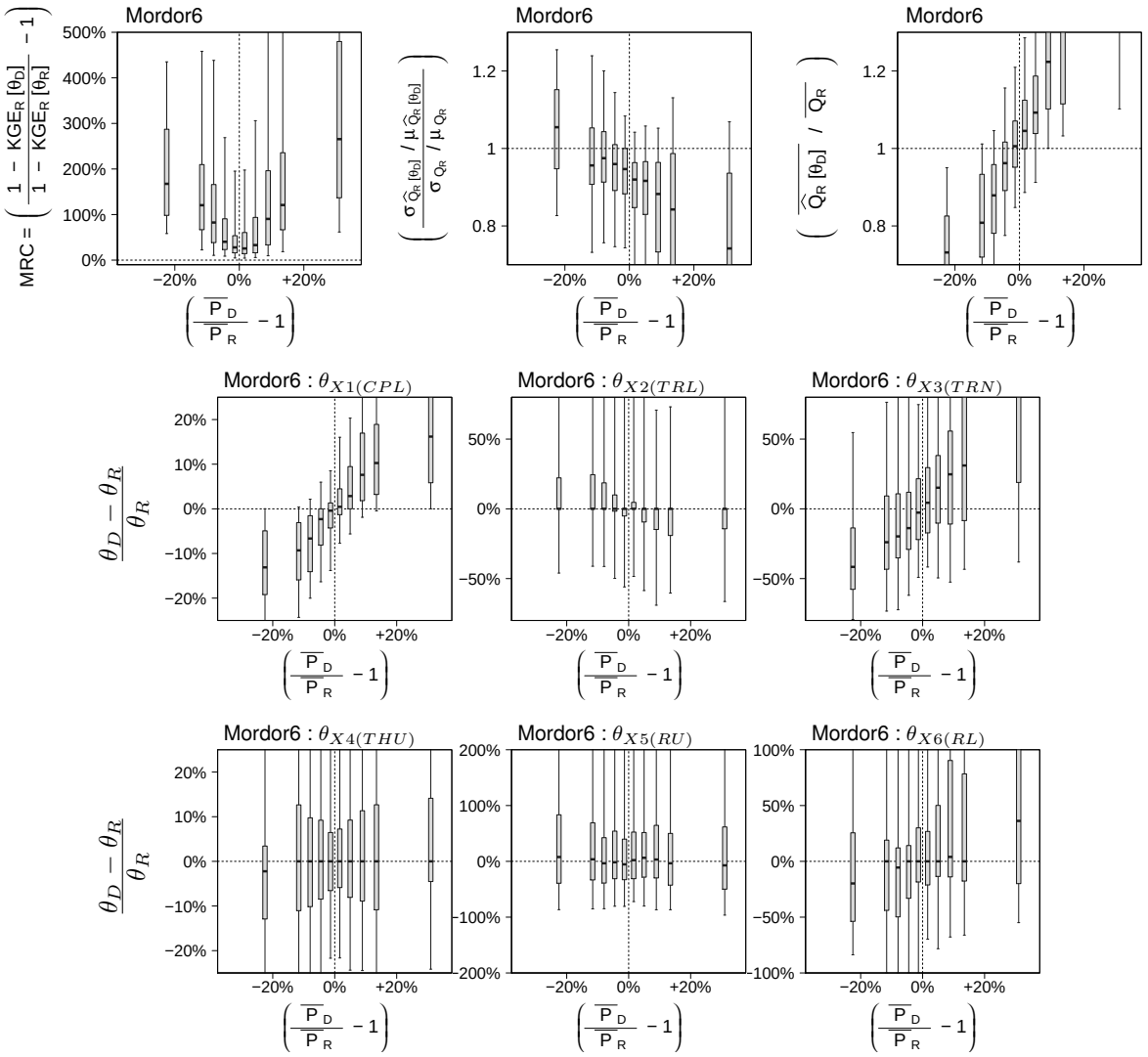


FIGURE G.7.: Biais de *Mordor6* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre $\triangle$erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5 \;\&\; r_s[biais] > 0.5)$

Mordor6 est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

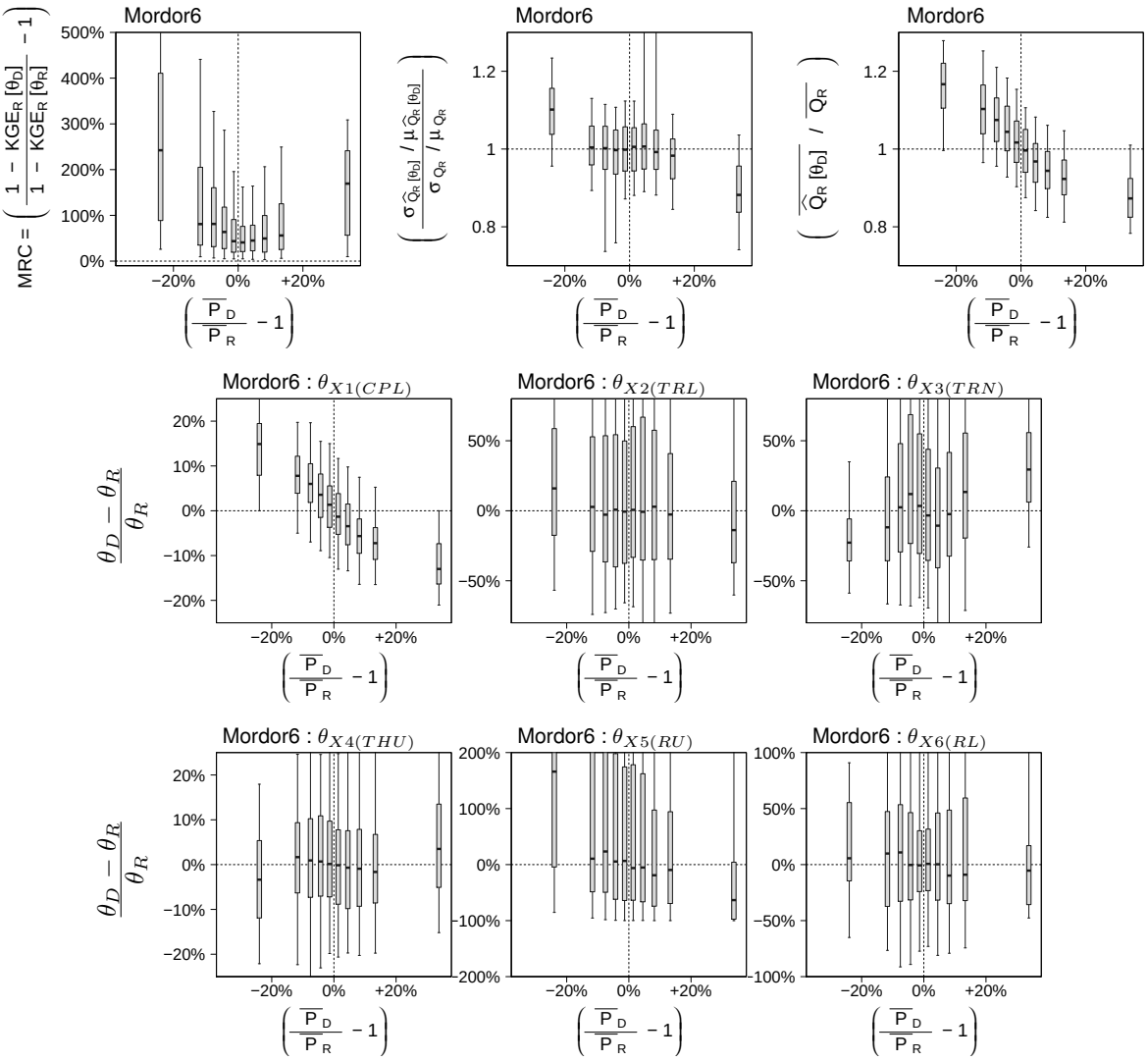


FIGURE G.8.: Biais de *Mordor6* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre $\triangle$erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5 \;\&\; r_s[biais] < -0.5)$

Mordor6 est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

Modèle *SimHyd*

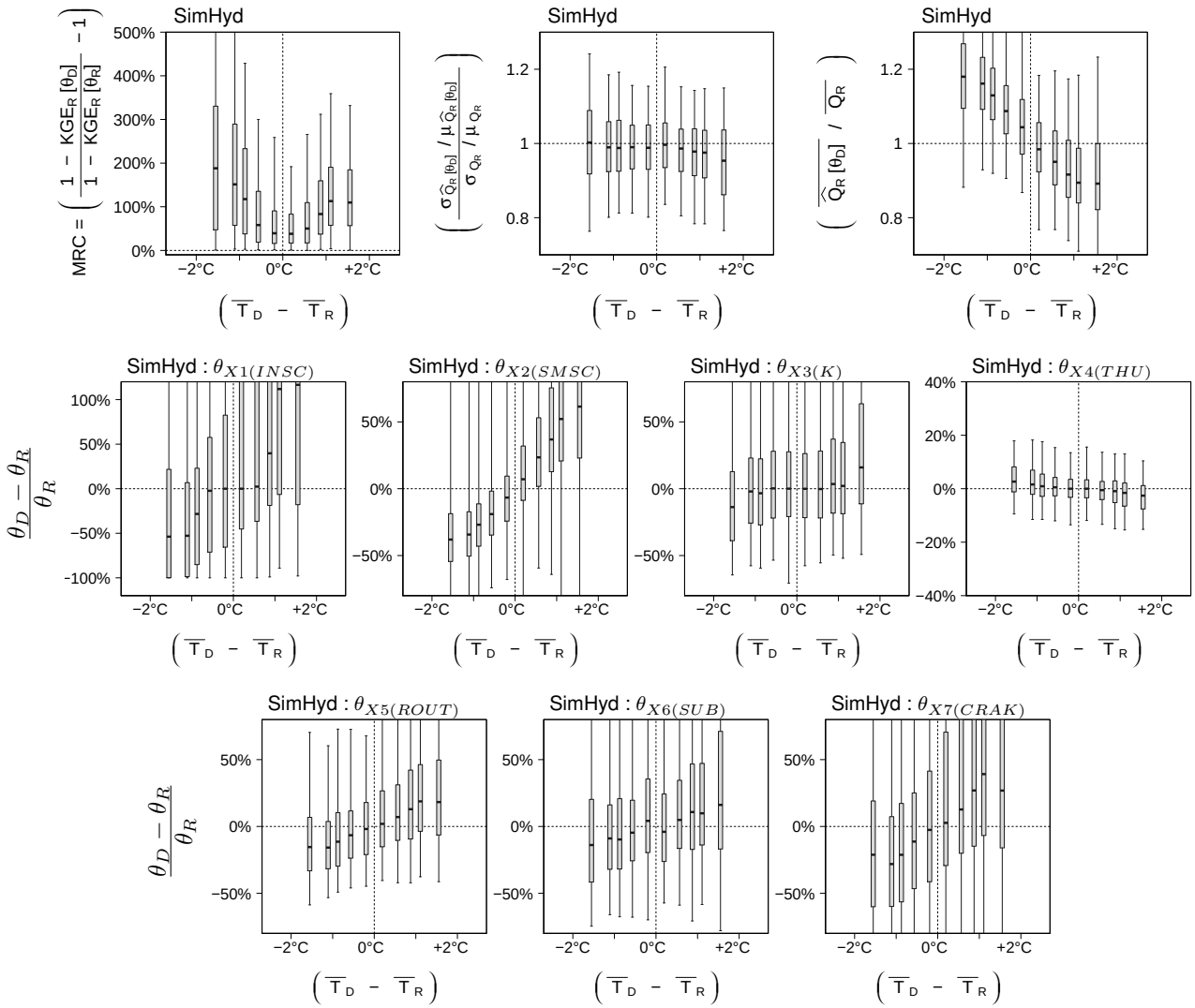


FIGURE G.9.: Biais de *SimHyd* lors des transferts entre périodes sur 120 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{T}$ $(r'_s[MRC] > 0.5)$

SimHyd est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

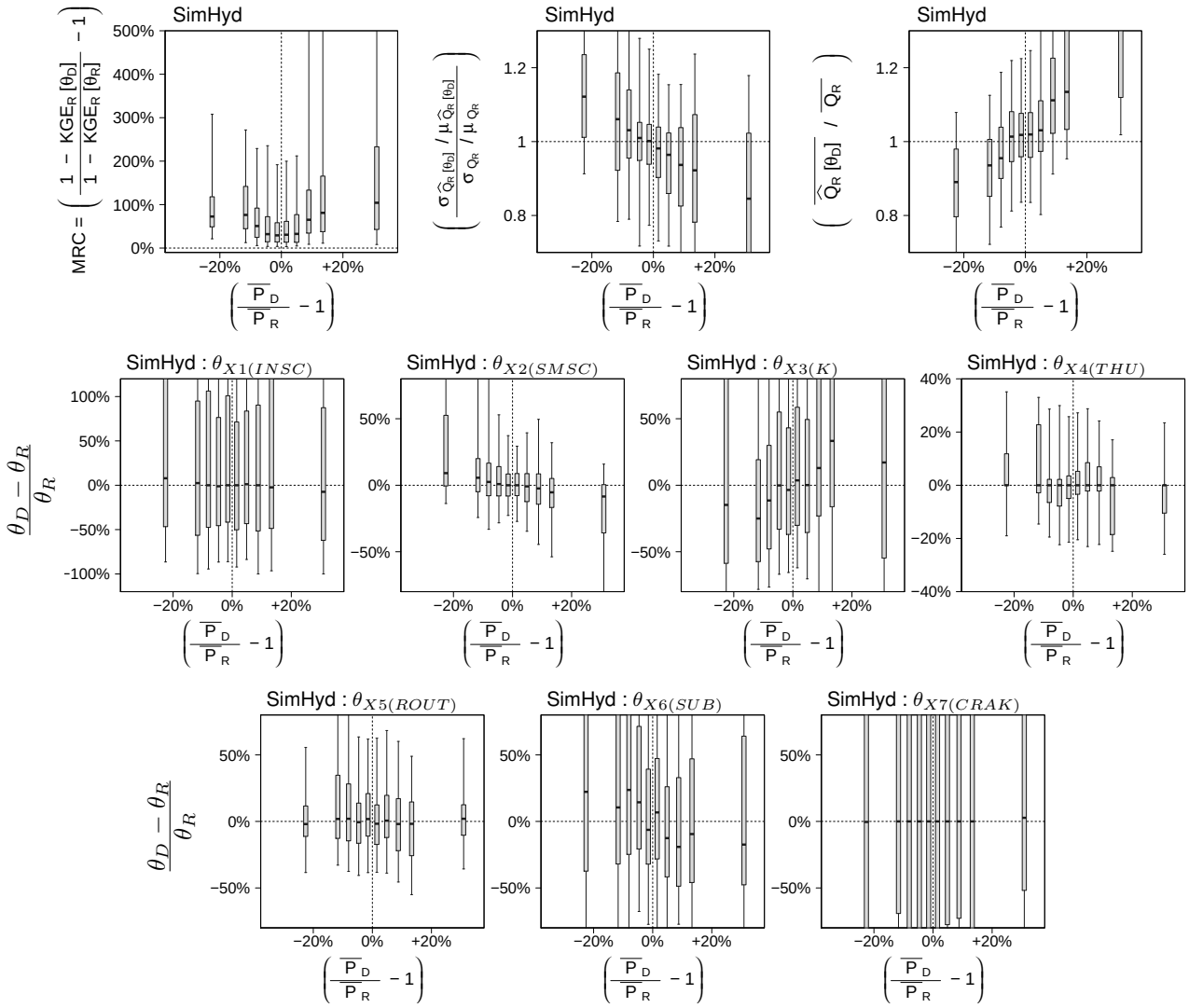


FIGURE G.10.: Biais de *SimHyd* lors des transferts entre périodes sur 32 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5$ & $r_s[biais] > 0.5)$

SimHyd est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

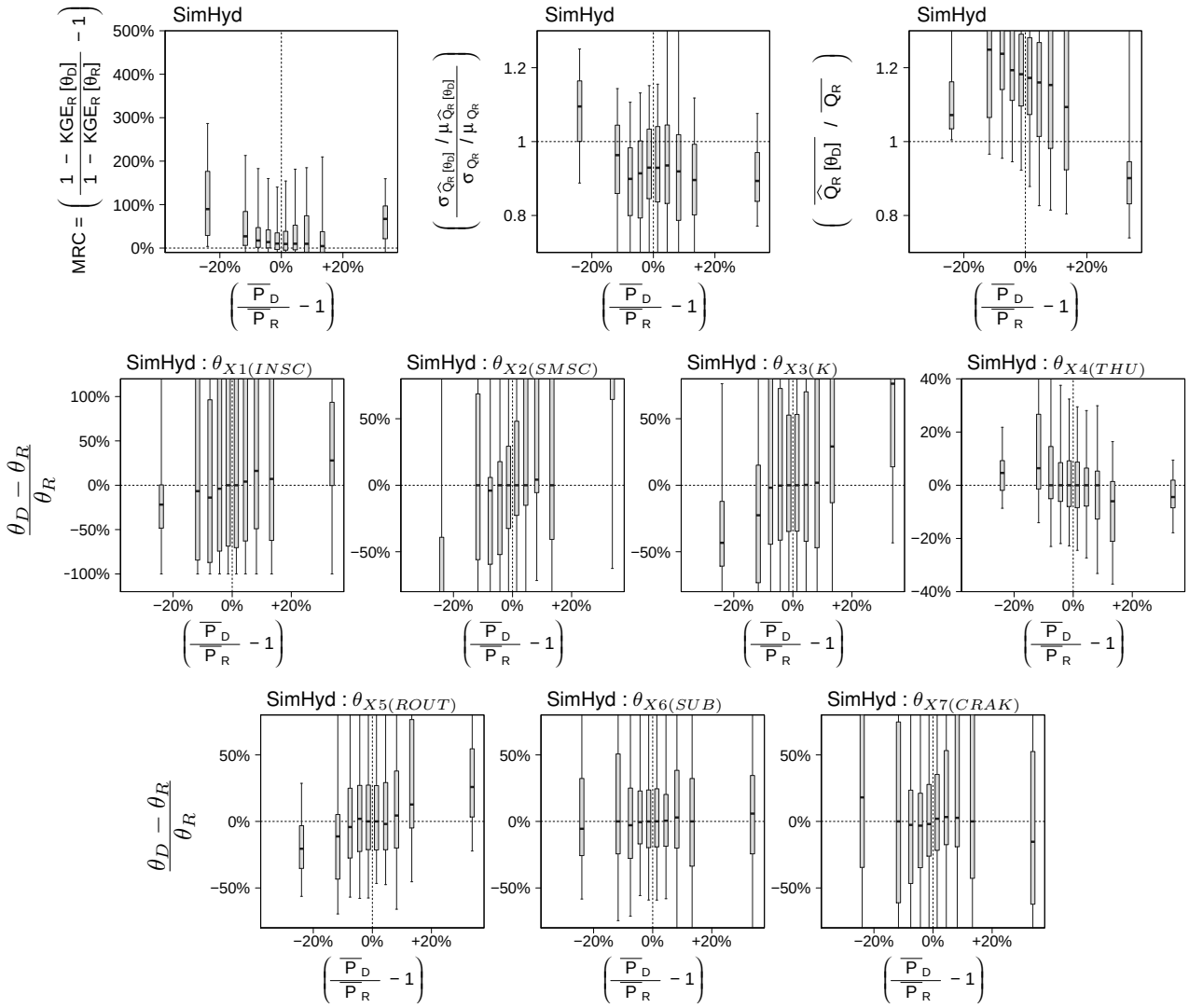


FIGURE G.11.: Biais de *SimHyd* lors des transferts entre périodes sur 14 bassins concernés par des corrélations entre △erreurs et $\triangle\overline{P}$ $(r'_s[MRC] > 0.5 \;\&\; r_s[biais] < -0.5)$

SimHyd est alimenté par des données SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Les boîtes à moustaches indiquent les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

G.3. Compléments sur l'exercice de fixation d'un paramètre

Les paramètres optimaux des modèles les plus parcimonieux varient parfois temporellement de façon nettement corrélée aux changements de conditions climatiques. Nous nous donc poser la question suivante : "*Il est possible d'identifier le(s) paramètre(s) trop dépendant(s) des conditions climatiques de calage pour certains modèles. Ne pourrions-nous pas simplement fixer ce(s) paramètre(s) pour résoudre nos problèmes de robustesse ?*"

Les figures G.12 à G.15 viennent compléter les illustrations placées dans le corps du manuscrit (section 5.5.3) en présentant les variations d'erreurs de simulation et de valeurs des paramètres avant et après avoir fixé le paramètre le plus sensible à $\triangle\overline{T}$ sur *GR4J* et *Mordor6* (tous deux couplés à *CemaNeige* dont les paramètres sont maintenus fixes).

Résultats pour *GR4J*

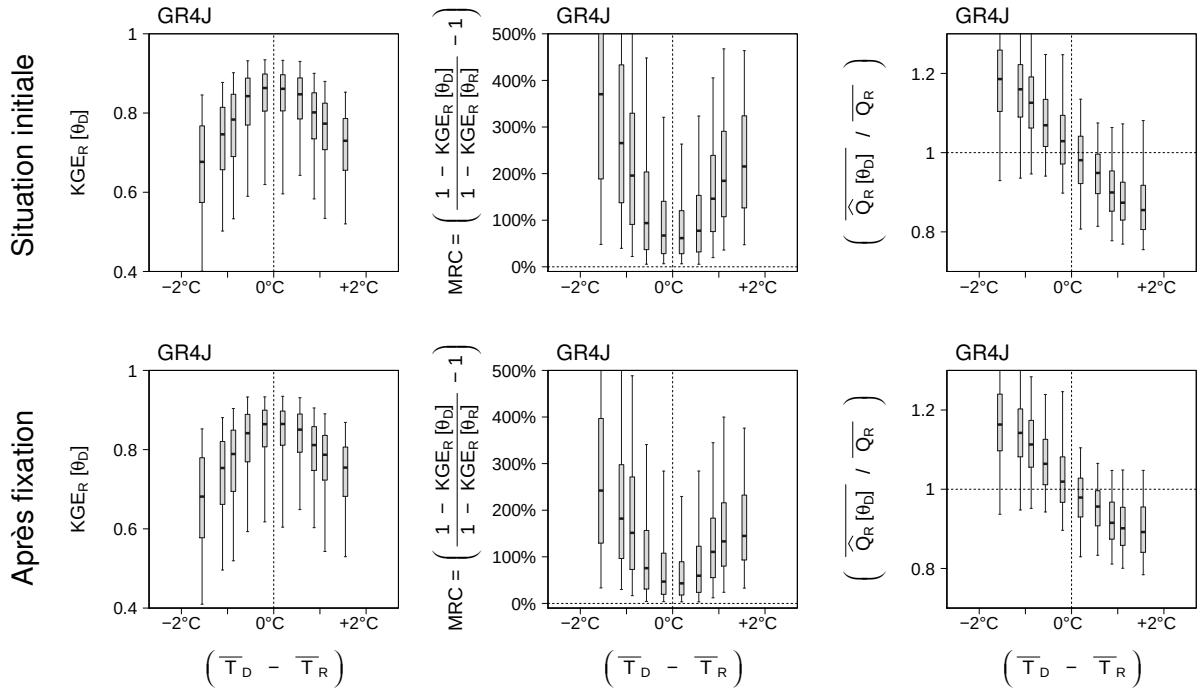


FIGURE G.12.: Erreurs commises par *GR4J* lors des transferts entre périodes sur 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$

Sur les bassins français : données SPAZM, ETP d'Oudin, couplé à CemaNeige (θ_{neige} fixés). Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

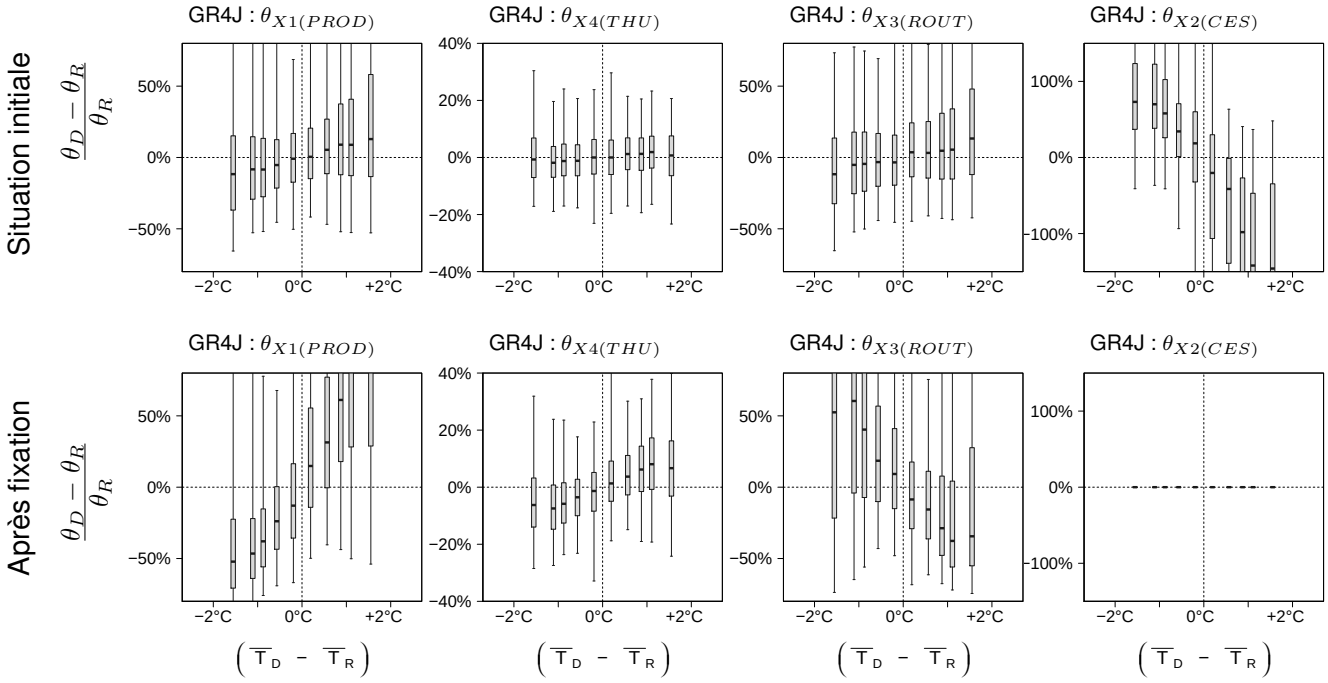


FIGURE G.13.: $\triangle\theta$ obtenus pour *GR4J* lors du test de fixation d'un paramètre pour un *GSST* sur les 120 bassins où $\triangle$erreurs=f($\triangle\overline{T}$)

(sur les bassins français : données SPAZM, ETP d'Oudin, couplé à CemaNeige (θ_{neige} fixés))

Résultats pour *Mordor6*

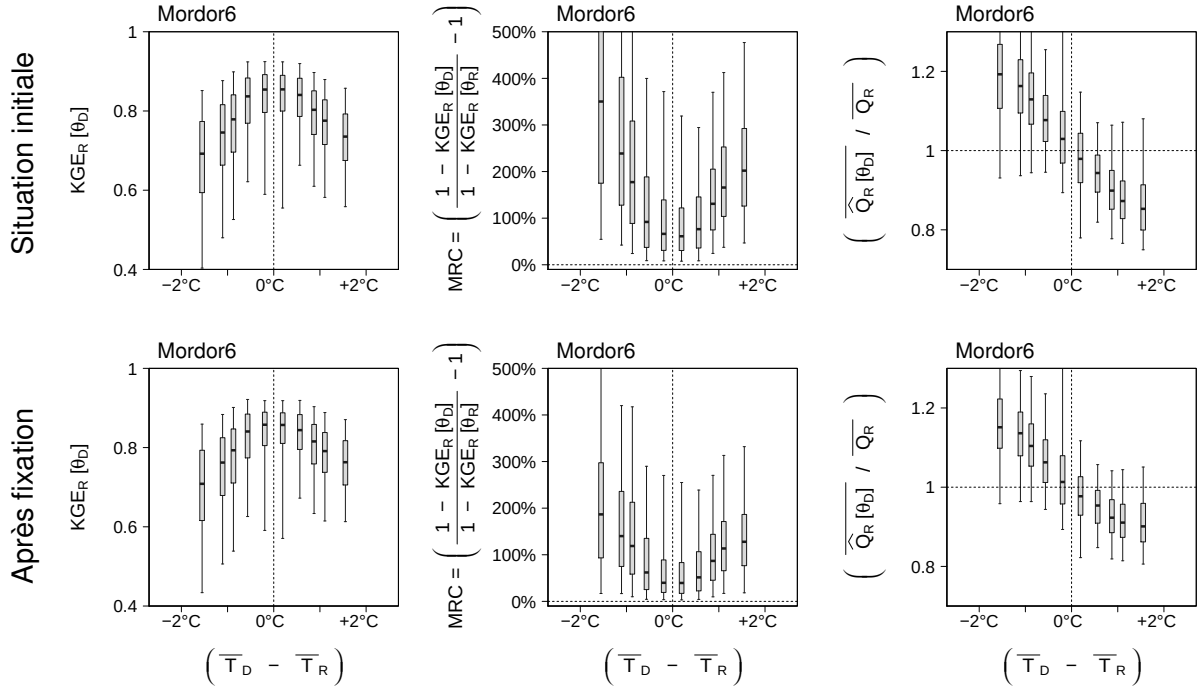


FIGURE G.14.: Erreurs commises par *Mordor6* lors des transferts entre périodes sur 120 bassins concernés par des corrélations significatives entre $\triangle$erreurs et $\triangle\overline{T}$

(sur les bassins français : données SPAZM, ETP d'Oudin, couplé à CemaNeige (θ_{neige} fixés))

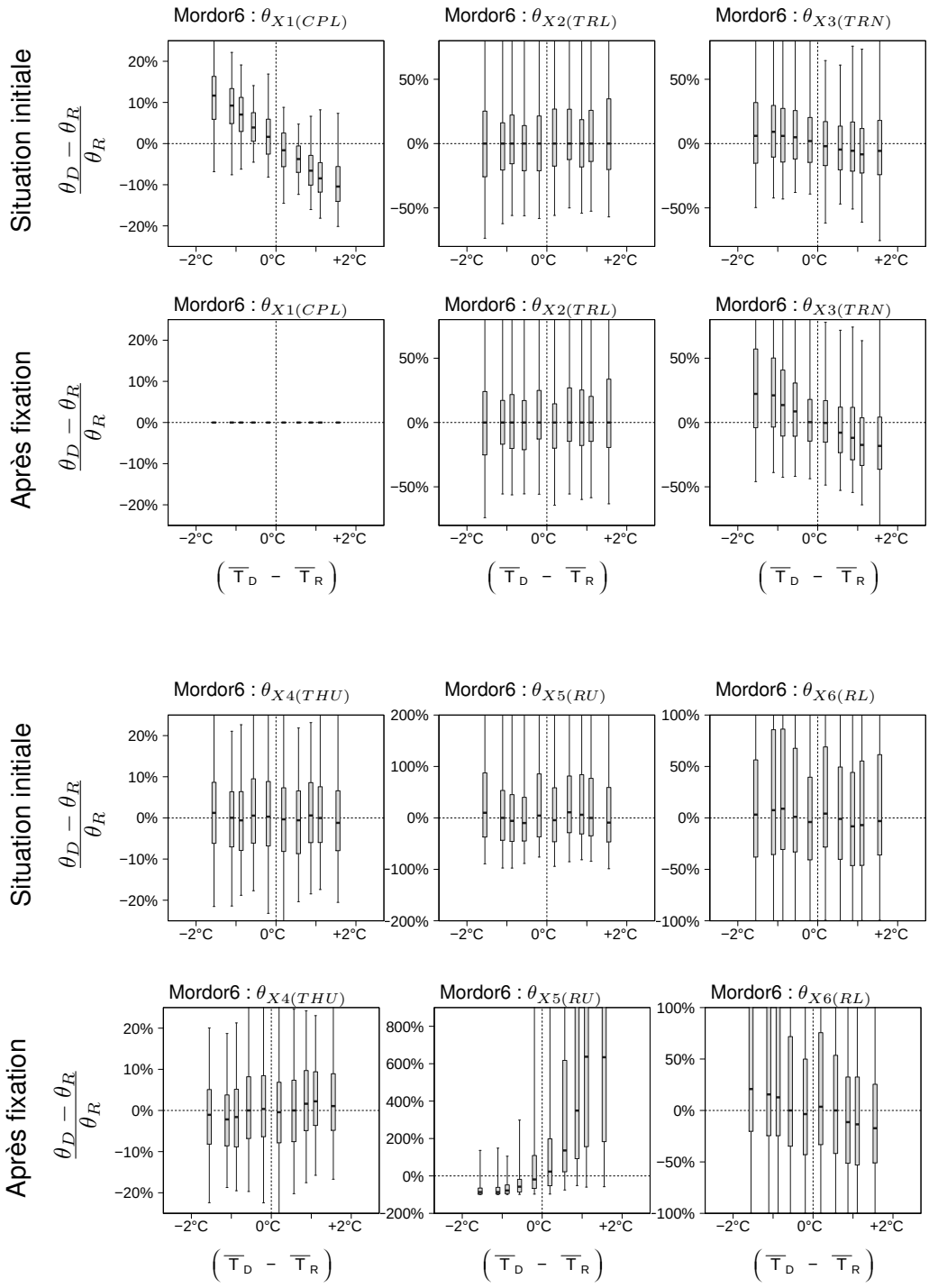


FIGURE G.15.: $\triangle\theta$ obtenus pour *Mordor6* lors du test de fixation d'un paramètre pour un *GSST* sur les 120 bassins où $\triangle$erreurs=f($\triangle\overline{T}$)

Sur les bassins français : données SPAZM, ETP d'Oudin, couplé à CemaNeige (θ_{neige} fixés). Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

G.4. Compléments sur les analyses des biais glissants

La section 5.6 présente une analyse des biais glissants de modélisation ($\overline{\widehat{Q}_{10\,ans}}/\overline{Q_{10\,ans}}$) obtenus pour différentes sous-périodes de calage et différents modèles. Nous apportons ici quelques compléments sur les critères utilisés pour la mesure de similarité entre les courbes de ces biais glissants. La figure G.16 fournit quant à elle une représentation côte à côte des analyses par nuages de points menées dans le chapitre 3 et des analyses par courbes de biais glissants menées dans le chapitre 5.

Mesure de similarité des comportements entre différents calages temporels

Pour chaque bassin et chaque modèle, nous considérons des biais glissants sur 10 ans obtenus en simulation en utilisant les jeux de paramètres calés sur nos différentes sous-périodes et du jeu calé sur toute la période disponible. Ces biais glissants seront respectivement notés $\omega_{SP[i]}$ et ω_{PT} et correspondent aux courbes grises et à la courbe rouge sur les graphiques de la figure 5.18. Le biais glissant $\omega_{SP[i]}$ est donc composé des différentes erreurs de volume moyen sur la fenêtre glissante de simulation et s'écrit :

$$\omega_{SP[i]} = \left[\frac{\overline{\widehat{Q}}_{SP[1]}\left[\theta_{SP[i]}\right]}{\overline{Q}_{SP[1]}}, \frac{\overline{\widehat{Q}}_{SP[2]}\left[\theta_{SP[i]}\right]}{\overline{Q}_{SP[2]}}, ..., \frac{\overline{\widehat{Q}}_{SP[p-1]}\left[\theta_{SP[i]}\right]}{\overline{Q}_{SP[p-1]}}, \frac{\overline{\widehat{Q}}_{SP[p]}\left[\theta_{SP[i]}\right]}{\overline{Q}_{SP[p]}} \right]$$

Dans notre cas, la taille de la fenêtre sur laquelle nous estimons ce biais glissant est identique à celle ayant servi au découpage des périodes de calage. Les n sous-périodes de calage sont donc les mêmes que les p périodes de calcul du biais glissant.

Pour chaque sous-période de calage $SP[i]$, nous souhaitons évaluer le degré de parallélisme entre la courbe $\omega_{SP[i]}$ (en gris sur la figure G.16) et la courbe ω_{PT} est issue du calage sur tout la période et nous sert de référence (en rouge sur la figure G.16). Nous calculons ainsi l'écart-type de la différence entre $\omega_{SP[i]}$ et ω_{PT} :

$$\sigma\left[\omega_{SP[i]} - \omega_{PT}\right] = \frac{1}{p} \cdot \sqrt{\sum_{k=1}^{p} \left(\frac{\overline{\widehat{Q}}_{SP[k]}\left[\theta_{SP[i]}\right] - \overline{\widehat{Q}}_{SP[k]}\left[\theta_{PT}\right]}{\overline{Q}_{SP[k]}} \right)^2}$$

Plus les valeurs de $\sigma\left[\omega_{SP[i]} - \omega_{PT}\right]$ sont faibles, plus les courbes de $\omega_{SP[i]}$ et ω_{PT} sont de forme similaire sans impliquer leur degré de superposition qui n'est pas mesuré ici (*c.-à-d.* les variations relatives comptent mais pas la *translation verticale*). Plutôt que d'exprimer de valeurs absolues de ces $\sigma\left[\omega_{SP[i]} - \omega_{PT}\right]$, nous avons préféré les diviser par l'écart-type de notre référence ω_{PT}. Cette division permet d'évaluer rapidement l'ordre de grandeur des imperfections du parallélisme entre $\omega_{SP[i]}$ et ω_{PT} vis-à-vis des variations propres du biais glissant. Il s'agit en quelque sorte du ratio bruit~signal, que nous pouvons noter :

$$\rho_i = \sigma\left[\omega_{SP[i]} - \omega_{PT}\right] \;/\; \sigma\left[\omega_{PT}\right]$$

Quelques quantiles sur ce ρ_i nous permettent d'estimer rapidement, sur chaque bassin et pour un modèle donné, le degré de parallélisme de nos courbes $\omega_{SP[i]}$. En regroupant ensuite les ρ_i obtenus sur tous les bassins, nous obtenons un aperçu global du degré de parallélisme des comportements sur notre échantillon (*cf.* figure 5.21, page 179).

Mesure de similarité des comportements entre différents modèles

Parallèlement, nous souhaitons caractériser les différences pouvant exister entre les modèles que nous avons testés. Le fait que tous les calages sur des sous-périodes conduisent aux mêmes variations de biais qu'un calage sur l'intégralité des observations est un premier résultat très intéressant. Le fait que différents modèles ayant des structures et paramétrisations différentes entraîne également des variations de biais identiques est un second résultat, pour partie indépendant du premier.

Pour estimer cette similitude des comportements entre modèles, nous avons construit un second critère sur le même schéma que ρ_i, à la différence près que nous comparons ici des biais glissants entre modèles. Pour simplifier (et compte tenu des valeurs constatées de ρ_i), nous n'utilisons qu'une courbe de biais glissant par modèle et par bassin : celle issue du calage sur la période totale, notée ω_{PT}^{M}, la notation M indiquant le modèle considéré. Le critère vise à comparer les signaux ω_{PT}^{M} issus de deux modèles (M_1 et M_2), l'un des deux servant de référence. Il est calculé sur chaque bassin et s'écrit :

$$\rho'_{M_1\,,\,M_2} = \sigma\left[\omega_{PT}^{M_2} - \omega_{PT}^{M_1}\right] \;/\; \sigma\left[\omega_{PT}^{M_1}\right]$$

Quelques quantiles sur ce $\rho'_{M_1\,,\,M_2}$ permettent d'estimer rapidement le degré de similitude des comportements entre les différents modèles testés (*c.-à-d.* similitude de nos courbes ω_{PT}, représentées en rouge sur nos exemples). Les résultats correspondants sont fournis dans le corps du manuscrit à la figure 5.22, page 180).

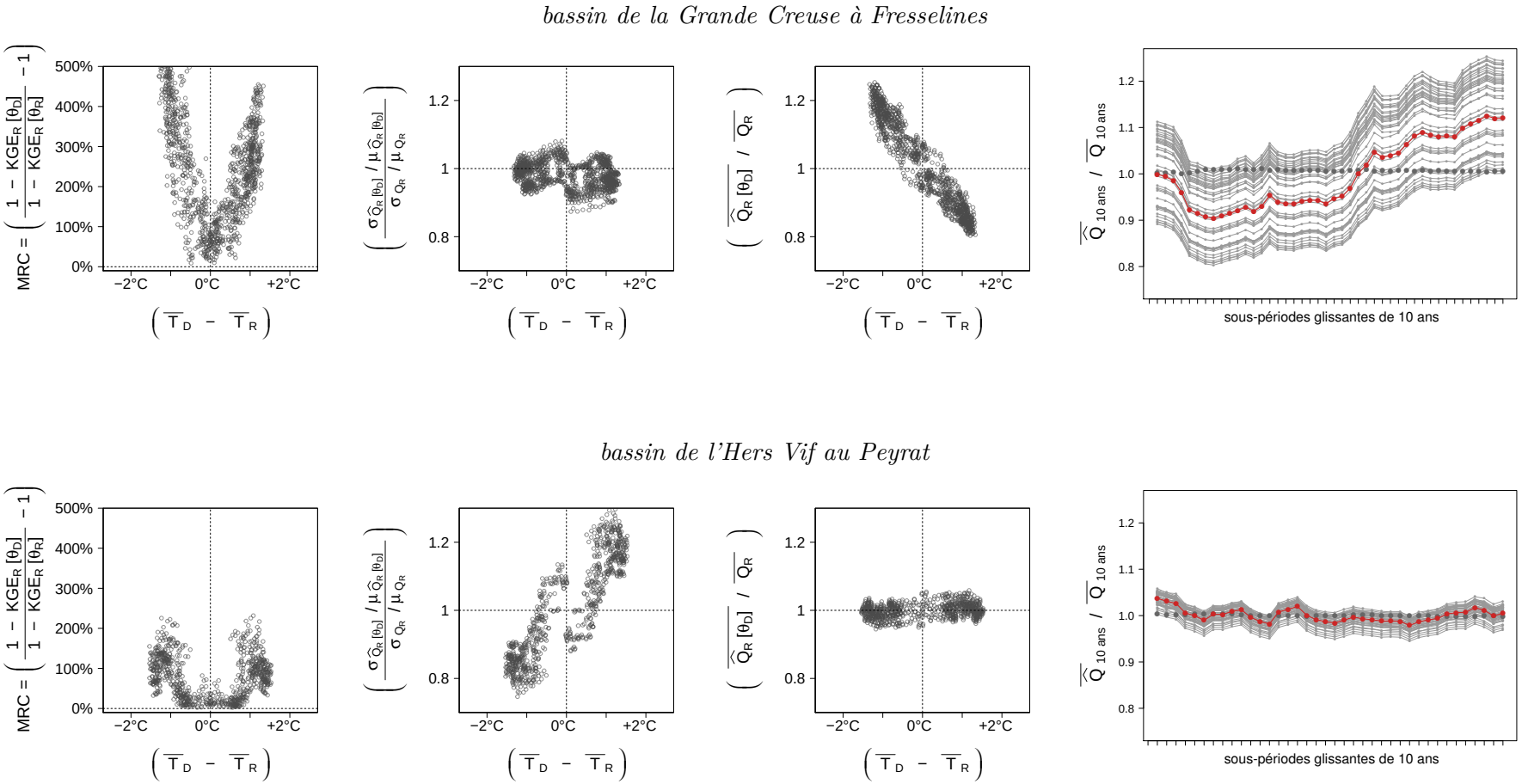


FIGURE G.16.: Exemples de représentation côte à côte des analyses par nuage de points et des courbes de biais glissants

H

Compléments pour le chapitre 6

H.1. Concept d'un calage robuste en termes d'extrapolation climatique

Présentation du concept

La phase de diagnostic de la robustesse des modèles hydrologiques face aux changements de climat ne nous semblait qu'une première étape lorsque nous avons initié ces travaux. La seconde partie de nos travaux devait être la recherche de voies d'amélioration de cette robustesse (notamment par une meilleure méthode de sélection des jeux de paramètres). Initialement ambitionnée comme le cœur de nos travaux, cette seconde partie nous a semblé de moins en moins atteignable au fil de notre avancée sur la problématique, c'est-à-dire au fil de notre compréhension ou absence de compréhension des phénomènes conduisant aux pertes de performances constatées lors des tests de transfert temporel des paramètres des modèles.

Lors de cette progression, nous gardions cependant à l'esprit le concept d'une procédure de calage qui serait *robuste* vis-à-vis des variations de conditions climatiques. Il s'agissait de trouver la méthode de sélection des jeux de paramètres qui maximiserait la capacité d'extrapolation climatique des modèles hydrologiques, autrement dit, orienterait cette sélection vers des jeux capables de représenter convenablement toute la gamme de conditions climatiques, quitte à perdre en efficacité localement [1].

À défaut d'avoir pu construire une telle solution aux problèmes mis en évidence lors du *GSST*, nous avons tout de même souhaité exposer nos réflexions ici. Cette section vise donc à présenter un concept de *calage climatiquement robuste*, puis de l'illustrer par quelques cas d'application menés lors de ces réflexions pour conclure sur les limites de cette approche.

Principe

Les outils à notre disposition pour estimer la robustesse des modèles ont constitué une source majeure d'inspiration pour nos réflexions autour du concept d'un *calage climatiquement robuste*, tout particulièrement notre méthode de représentation des résultats issus du *GSST* sous la forme de nuages de points dont nous analysons l'allure (*cf.* section 2.1.3). Cette représentation se base sur l'expression des erreurs relatives de modélisation ($\varepsilon_R[\theta_D]/\varepsilon_R[\theta_R]-1$) en fonction de l'écart de conditions climatiques subi lors du transfert (*ex.* $\triangle\overline{T}_{D\rightarrow R}$). L'augmentation des erreurs avec l'écart de climat se traduit ainsi par une forme en $\cup$ ou $\vee$. L'objectif du *calage climatiquement robuste* est d'aplatir cette forme, c'est-à-dire avoir une faible augmentation des erreurs lorsque les périodes de calage et de validation deviennent contrastées. Cependant,

1. Par *perdre localement*, nous signifions ici *perdre sur des conditions climatiques spécifiques*, pour gagner en moyenne sur toute la gamme des conditions.

il est possible que cette meilleure extrapolation climatique des paramètres sélectionnés se fasse au détriment des performances lors de transfert sous climat similaire. Cet objectif et l'éventuel compromis associé sont schématisés sur la figure H.1.

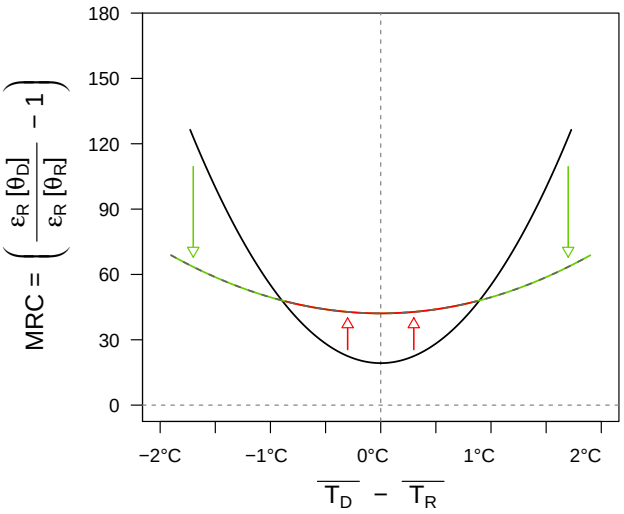


FIGURE H.1.: Schématisation du principe d'un calage climatiquement robuste pour des bassins où les erreurs en transfert sont corrélées avec les écarts de températures

transferts des paramètres entre des périodes donneuses (D) et receveuses (R)
flèches descendantes vertes : objectif de diminution des erreurs lors des transferts entre périodes constrastées ; flèches ascendantes rouges : effet collatéral éventuel d'augmentation des erreurs lors des transferts entre périodes similaires

Description de la procédure

Lors des analyses faisant suite au *GSST*, les comparaisons étaient effectuées sur des résultats de transfert du type $\varepsilon_{D\to R}/\varepsilon_{R\to R} = f(\triangle climat_{D\to R})$. La procédure de *calage climatiquement robuste* que nous avons imaginé est quelque peu différente. Elle repose sur le tirage aléatoire d'un grand nombre de jeux de paramètres qui sont évalués sur leur capacité à reproduire convenablement une gamme élargie de conditions climatiques. Nous ne considérons plus de transferts (*c.-à-d.* plus de $\triangle climat_{D\to R}$) mais nous estimons directement les performances sur différentes conditions climatiques (*ex.* erreurs sur des périodes de température faible, moyenne ou forte).

Cette procédure se décompose selon les étapes suivantes :

1. Mise en place de groupes de sous-périodes selon leurs caractéristiques climatiques.

   a. Constitution de sous-périodes à partir de la série historique disponible, qui seront utilisées pour évaluer les performances du modèle. Pour simplifier, nous reprenons les mêmes sous-périodes que dans nos tests précédents (découpage selon une fenêtre glissante de 10 ans).

   b. Caractérisation des conditions climatiques sur chaque sous-période (*ex.* moyennes annuelles/saisonnières ou variabilité interannuelle/infra-annuelle des précipitations et températures, indice d'aridité...).

   c. Regroupement des sous-périodes selon leurs similarités climatiques. Cette étape s'apparente au découpage vertical des nuages de points dans nos précédentes analyses (*cf.* section 2.1.3). Elle consiste à analyser ensemble les périodes climatiquement semblables et à distinguer les périodes contrastées.

2. Tirage aléatoire de jeux de paramètres et simulation des chroniques.

   a. Tirage aléatoire d'un très grand nombre de jeux de paramètres entre des bornes réalistes pour le bassin versant considéré [2]. Pour les résultats présentés ici, seul le modèle *GR4J* est considéré (alimenté par une *ETP* d'Oudin et couplé à *CemaNeige* dont les paramètres sont fixés).

   b. Simulation de la chronique complète avec chaque jeu de paramètres et extraction des erreurs sur chaque sous-période ($\varepsilon_{10\,ans}$). Pour ces tests, nous avons considéré le critère d'erreur $RMSE[\sqrt{Q}]$, qui constitue également notre fonction objectif. Il est vrai que ce critère donnera mécaniquement des valeurs plus fortes pour les périodes (et les bassins) plus humides. Pour évacuer (au moins partiellement) cet effet et faciliter les comparaisons graphiques, ces erreurs sont normalisées et nous construisons $\varepsilon'_{10\,ans} = \varepsilon_{10\,ans}/min(\varepsilon_{10\,ans})$. L'erreur obtenue pour chaque jeu est donc normalisée par l'erreur minimale obtenue sur la même sous période (*c.-à-d.* avec le meilleur jeu). Cette division par l'erreur obtenue en calage se rapproche donc fortement de notre formulation $MRC = \varepsilon_{D\to R} \,/\, \varepsilon_{R\to R} - 1$. Elle permet de mettre en avant les erreurs explicitement liées au fait de transférer temporellement (et climatiquement) les paramètres.

   c. Regroupement des erreurs normalisées ($\varepsilon'_{10\,ans}$) pour déterminer les niveaux sur chacun des groupes climatiques ($\varepsilon'_{regroupement\,climatique}$). Ainsi, nous disposons par exemple du vecteur $\varepsilon'_{\overline{T}_{regroupements}}$ qui fournit les erreurs moyennes en validation sur les différentes plages de température constituées en regroupant nos sous-périodes de test.

2. Les valeurs obtenues par calage sur diverses sous-périodes temporelles sont une source particulièrement utile pour nous aider à définir les bornes entre lesquelles sont tirés aléatoirement un million de jeux de paramètres.

3. Estimation de la capacité d'extrapolation climatique des jeux

   a. Pour chaque jeu de paramètres, estimation d'un critère de moyenne et de variance des erreurs entre les différents groupes climatiques ($\mu[\varepsilon'_{regrpmt\,clim.}]$ et $\sigma[\varepsilon'_{regrpmt\,clim.}]$). Ces critères nous informent sur la capacité du jeu à fournir des performances simultanément bonnes (critère μ) et stables (critère σ) sur toute la gamme de climat disponible.

   b. Construction d'un nuage de points en utilisant les critères $\mu[\varepsilon'_{regrpmt\,clim.}]$ et $\sigma[\varepsilon'_{regrpmt\,clim.}]$ obtenus pour chaque jeu comme des coordonnées. Chaque jeu est ainsi positionné en fonction du compromis qu'il représente entre la qualité des simulations qu'il permet, d'une part, et la robustesse des performances sur toute une gamme de conditions climatiques, d'autre part. Un exemple de nuage ainsi obtenu est fourni (en gris) sur la figure H.2.

   c. Identification des jeux de paramètres composant le front de Pareto entre performances et capacité d'extrapolation climatique. Ces jeux sont représentés en noir sur la figure H.2, celui indiqué par un carré rouge étant un exemple de jeu pouvant être retenu à l'issue de ce *calage climatiquement robuste*.

   d. Identification du jeu de paramètres optimal sur toute la période (*c.-à-d.* jeu fournissant un niveau minimal de $RMSE[\sqrt{Q}]$ moyenne sur toutes les années). Ce jeu est représenté par un + bleu foncé sur la figure H.2.

   e. Identification des jeux de paramètres optimaux sur les différentes sous-périodes de tests (*c.-à-d.* pour chaque sous-période de 10 ans, nous identifions le jeu fournissant la $RMSE[\sqrt{Q}]$ minimale sur celle-ci, c'est-à-dire celui fournissant l'erreur utilisée à l'étape 2.b pour la normalisation). Ces jeux sont représentés par des x bleus clair sur la figure H.2.

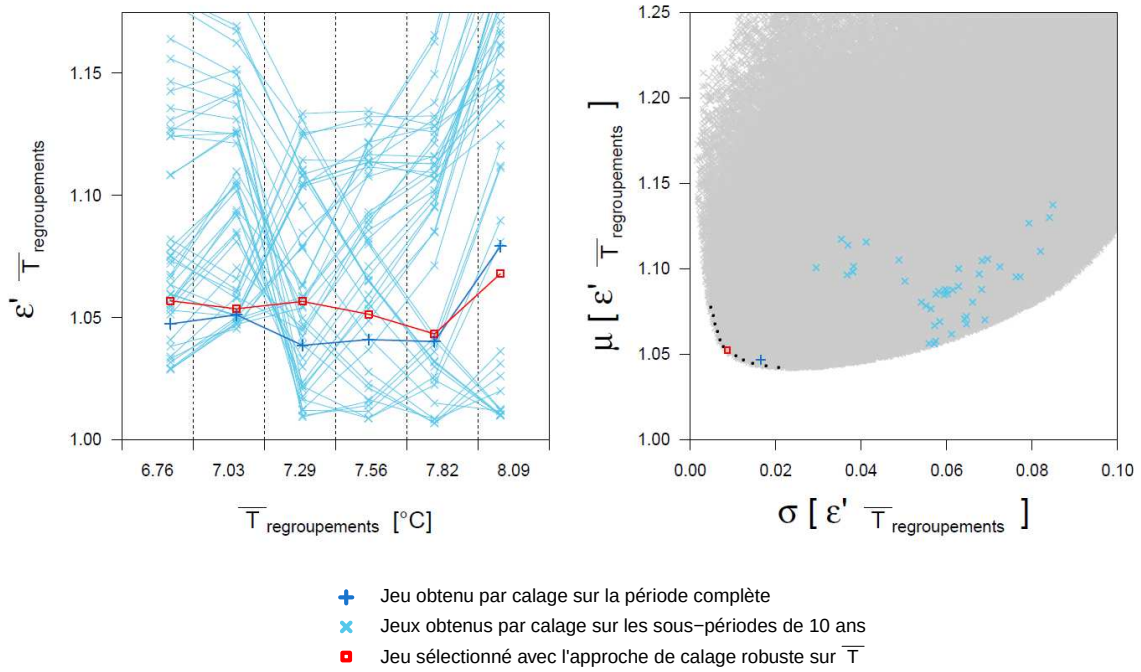


FIGURE H.2.: Illustration du compromis entre qualité des simulations (μ) et robustesse (σ) sur une gamme élargie de conditions climatiques

Résultats obtenus avec GR4J-CemaNeige (θ_{neige} fixés) alimenté par SPAZM et une ETP d'Oudin.

À gauche : variations des erreurs relatives en fonction des températures en validation (un jeu de paramètres = une courbe) ; à droite : compromis entre qualité des simulations (μ) et robustesse des performances (σ) (un jeu de paramètres = un point).

En gris : ensemble des jeux de paramètres tirés aléatoirement ; en noir et rouge : front de Pareto des jeux fournissant le meilleur compromis entre performances et capacité d'extrapolation climatique ; en bleu foncé et bleu clair : jeux respectivement obtenus par calage sur toute la période et sur les différentes sous-périodes de 10 ans.

Exemples de résultats

En guise d'exemples, nous avons considéré les bassins de *l'Allier à Vieille-Brioude* et de *la Grande Creuse à Fresselines*, où nous avions identifié des liens entre robustesse et $\triangle\overline{T}$ lors des transferts des paramètres de *GR4J-CemaNeige*. Ces deux bassins ont déjà joué le rôle d'exemples illustratifs dans la section 3.3.3, où nous fournissions d'ailleurs quelques-unes de leurs caractéristiques.

Les résultats obtenus sur ces bassins avec notre procédure de *calage climatiquement robuste* sont fournis sur les figures H.3 et H.4. Les résultats correspondant à une configuration de calage (*c.-à-d.* lorsque toute la chronique disponible est considérée) constituent les deux graphiques supérieurs. Les quatre graphiques inférieurs montrent quant à eux les résultats obtenus sur les deux moitiés de cette chronique et permettent d'évaluer, en validation, la pertinence de notre procédure de sélection *robuste* par rapport à un calage classique.

Comme sur l'illustration précédente, les graphiques placés à gauche sur ces figures montrent les niveaux d'erreur obtenus, pour les jeux d'intérêt, sur les différents groupes climatiques : $\varepsilon'_{\overline{T}_{regroupements}} = f(\overline{T}_{regroupements})$. Chaque jeu y est représenté par une série de points (un par bande climatique) indiquant les performances du jeu sur les différents regroupements climatiques. Pour faciliter la lecture, ces points sont reliés entre eux, ce qui permet une visualisation rapide de la capacité d'extrapolation climatique associée à chaque jeu. Cette représentation est apparentée à celle que nous avons rappelée schématiquement sur la figure H.1 et qui fut largement utilisée dans nos précédentes analyses. La seule différence est que les conditions climatiques sont exprimées directement ici, plutôt que sous forme relative (*c.-à-d.* $\overline{T}$ plutôt que $\triangle\overline{T}$). La moyenne (μ) et la variance (σ) de chacune de ces courbes constituent les coordonnées des points associés sur les nuages placés à droite. Ces nuages illustrent le compromis entre performance moyenne et robustesse climatique des jeux sur la gamme de température considérée.

Ces différentes représentations des performances des jeux de paramètres sur les groupes climatiques sont très informatives.

Nous retrouvons en bleu clair le fait que les jeux calés sur des sous-périodes de 10 ans sont peu transférables vers des périodes plus chaudes ou plus froides (pour ces bassins et ce modèle). Chacune de ces courbes présente un point bas indiquant la gamme climatique dans laquelle se situe plutôt la sous-période de 10 ans correspondante puis les erreurs augmentent avec l'écart de température (*ex.* graphique H.3.a). Ainsi, un jeu fournissant de faibles erreurs relatives sur les périodes froides verra celles-ci augmenter avec la température de la période simulée et inversement. Les très faibles erreurs obtenues sur les conditions les plus proches de celles de calage peuvent compenser les fortes erreurs commises sur les conditions les plus contrastées et aboutir à un critère μ relativement faible. Le critère de variance σ, indicatif de la robustesse climatique du jeu, prendra par contre des valeurs élevées (*ex.* graphique H.4.b).

Les jeux issus de la procédure de *calage climatiquement robuste* (en rouge) sont, par construction, positionnés au plus près du cas idéal de variance nulle et erreur normalisée nulle. Néanmoins, il est frappant de constater que ces jeux ne battent que faiblement les calages classiques (en bleu foncé) réalisés sur toute la chronique (ou demi-chronique). En effet, ces calages classiques constituent déjà un compromis très intéressant entre performance moyenne (μ) et transférabilité sur des périodes contrastées (σ). Les tests de validation par découpage de la chronique en deux moitiés continues montrent une nette diminution des performances relatives. Si cette diminution est compréhensible, il est plus troublant d'obtenir des performances similaires à celles obtenues avec le jeu calé classiquement sur l'autre demi-chronique (*ex.* graphiques H.3.e-f et H.4.c-d). Plus troublant encore, nous constatons certains cas où le transfert du calage classique permet de meilleures simulations (tant sur μ que sur σ) que le jeu sélectionné pour sa *robustesse climatique* sur l'autre demi-chronique et également transféré (*ex.* graphique H.4.e-f). Combinés à des considérations plus générales, ces résultats alimentent la discussion menée ci-dessous quant à la véritable pertinence de cette piste de *calage robuste*.

Discussion

La méthodologie de calage proposée ici permet théoriquement la sélection de jeux de paramètres plus robustes en termes de transférabilité climatique. Cependant, elle présente plusieurs limites importantes qui rendent son utilisation difficile en pratique. Les résultats présentés ici, pourtant obtenus dans des conditions très favorables, confirment les limites de cette piste de *calage climatiquement robuste*, comme solution aux problèmes de robustesse identifiés dans nos travaux.

Une première contrainte liée de cette procédure est la nécessité de réaliser un très grand nombre de simulations. Si elle ne constitue pas une véritable limite, elle peut toutefois s'avérer contraignante selon la complexité du modèle et le pas de temps de simulation. Une alternative pourrait être de pondérer les années lors du calage selon le caractère extrême qu'elles représentent du point de vue climatique. L'idée serait de réduire ainsi l'influence des années moyennes, si elles sont en plus grand nombre, pour mieux contraindre le modèle à représenter les situations très contrastées. La construction du critère de pondération adéquat restera toutefois une tâche délicate. Cette procédure alternative aurait néanmoins l'avantage d'être identique au calage standard et serait ainsi nettement plus simple à mettre en place et beaucoup moins gourmande en temps de calcul pour la plupart des modèles.

Une seconde contrainte, qui constitue cette fois une limite majeure de cette procédure, est liée par notre difficulté à prédire si le manque de robustesse d'un modèle sur le bassin étudié sera lié à des écarts de température, de précipitations, ou d'autres phénomènes non détectés. Cette difficulté fut mise en avant par l'hétérogénéité obtenue lors de la recherche de corrélation entre erreurs de modélisation et changements de conditions climatiques (*cf.* section 3.2). En effet, les résultats peuvent différer d'un bassin à un autre, sans que nous parvenions à expliquer cette répartition géographique. Or, sans cette capacité à prédire quelles conditions de transfert sont problématiques, nous ne pouvons pas réaliser un calage *climatiquement robuste*, puisque celui-ci est mené pour une variable climatique particulière. Il devient alors nécessaire de conduire au préalable une procédure du type *GSST*, qui requiert de longues séries d'observations, ce dernier point étant problématique pour de nombreux bassins. Cette limite à la procédure est donc particulièrement handicapante.

D'une manière plus générale, si nous sommes conscients de problèmes de transferts climatiques que nous ne pouvons résoudre, nous devons nous interroger sur notre objectif. Est-il préférable d'avoir un modèle reproduisant mieux toute la gamme des conditions mais avec des performances altérées ou souhaitons-nous un modèle performant sur le type de conditions que nous souhaitons simuler ? De la même manière : si nous savons qu'un modèle n'arrive pas à représenter fidèlement les forts débits et les faibles débits avec un même jeu de paramètres, préférons-nous une unique configuration permettant un compromis acceptable ou plusieurs configurations (*ex.* plusieurs jeux de paramètres voire même structures de modèle) spécifiques

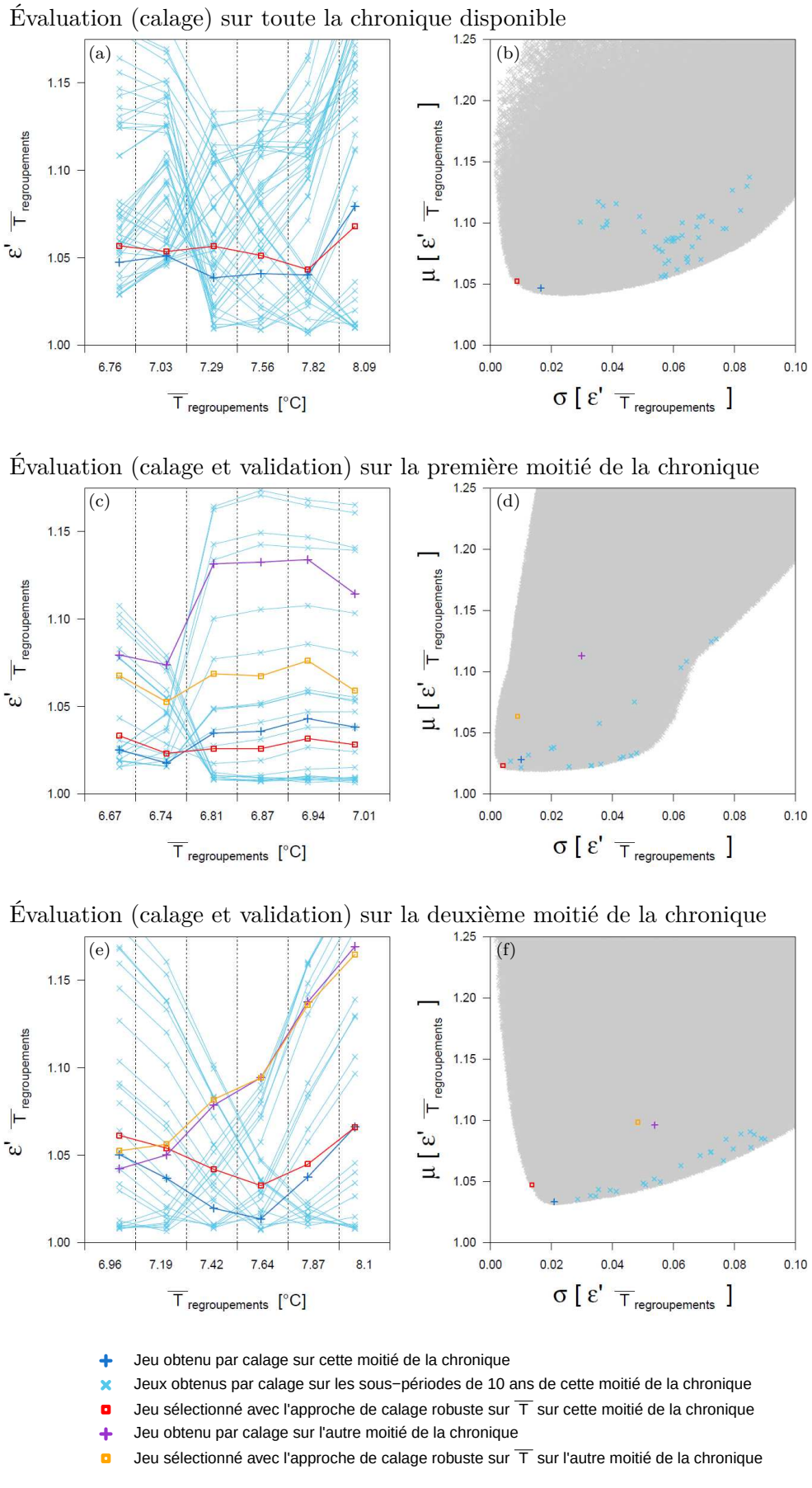


FIGURE H.3.: Exemples de résultats de calages robustes de *GR4J* sur le bassin de *l'Allier à Vieille-Brioude*

Résultats pour GR4J-CemaNeige (θ_{neige} fixés) alimenté par SPAZM et une ETP d'Oudin.

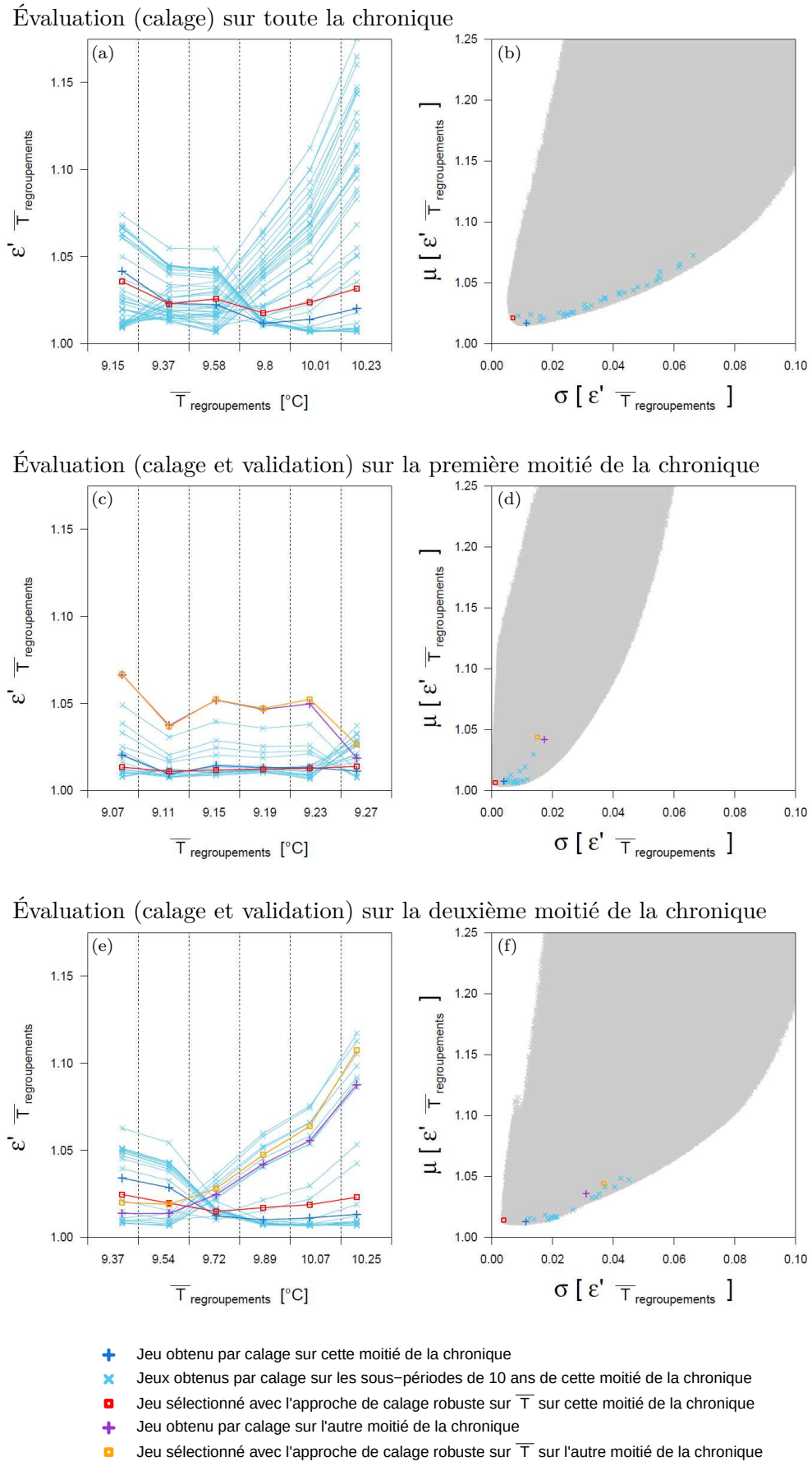


FIGURE H.4.: Exemples de résultats de calages robustes de *GR4J* sur le bassin de *la Grande Creuse à Fresselines*

Résultats pour GR4J-CemaNeige (θ_{neige} fixés) alimenté par SPAZM et une ETP d'Oudin.

pour chacune de ces tâches (simulation des crues ou simulation des étiages) ? Nous comprenons ainsi que la procédure de calage *climatiquement robuste* peut se révéler plus pénalisante que bénéfique selon la situation (*cf.* graphique H.4.e-f). En effet, à défaut de pouvoir résoudre le manque de robustesse des modèles, peut-être préfèrerions-nous caler notre modèle sur la période qui se rapproche le plus climatiquement de celle de simulation. Par exemple, on pourra préférer caler un modèle sur les années chaudes pour mener des études d'impact du changement climatique sur des bassins où des corrélations entre manque de robustesse et $\triangle\overline{T}$ entre périodes sont constatés. Dans ce cadre, une sélection des années par *bootstrap* constitue une approche pertinente.

Pour résumer, le concept d'un calage assurant une meilleure extrapolation climatique, que nous imaginions initialement comme une solution ultime, perd sa pertinence. Les difficultés des modèles à reproduire les variations d'équilibre entre précipitations, évapotranspiration et écoulements semblent être une cause majeure des pertes de performance constatées lors des transferts temporels de paramètres (*cf.* section 5.6, où nous étudions ces biais moyen terme). Or nos travaux montrent que le fait de caler sur l'intégralité de la chronique plutôt que sur des sous-périodes, de tester différents modèles ou même de choisir un critère de calage spécifiquement focalisé sur l'équilibrage des bilans moyen terme ne permettait pas réellement d'améliorer cette défaillance constatée. Au vu de ces résultats, il semble peu réaliste qu'une approche telle que celle proposée ici (sans réflexion sur les entrées ou la structure des modèles) permette de résoudre les problèmes de robustesse rencontrés. En conclusion, il ne nous paraît pas judicieux de poursuivre cette piste de *calage climatiquement robuste*.

H.2. Compléments sur les tests de correction des paramètres

Les éléments fournis ici viennent compléter la section 6.5, où est présentée une procédure de correction des paramètres *post-calage* dans le but de se prémunir des pertes de performance des modèles subies lors des transferts entre des périodes climatiquement contrastées. Cette correction est appliquée à l'issue du calage et son amplitude dépend de l'écart de température entre les conditions de calage et d'utilisation du modèle hydrologique. Elle est testée sur deux sous-groupes de 60 bassins pour les modèles *GR4J et Mordor6* (tous deux alimentés par une ETP d'Oudin et couplés à *CemaNeige* dont les paramètres sont maintenus fixes pour cet exercice). Ces deux groupes jouent alternativement les rôles d'échantillons de calage et de contrôle de la procédure corrective.

La figure H.5 fournit quelques caractéristiques de ces deux sous-groupes de 60 bassins.

Les figures H.6 et H.7 viennent quant à elles compléter les figures 6.15 et 6.16 (pages 217 et 219) en fournissant les variations des paramètres optimaux avant la procédure de correction, et les valeurs prises après celle-ci.

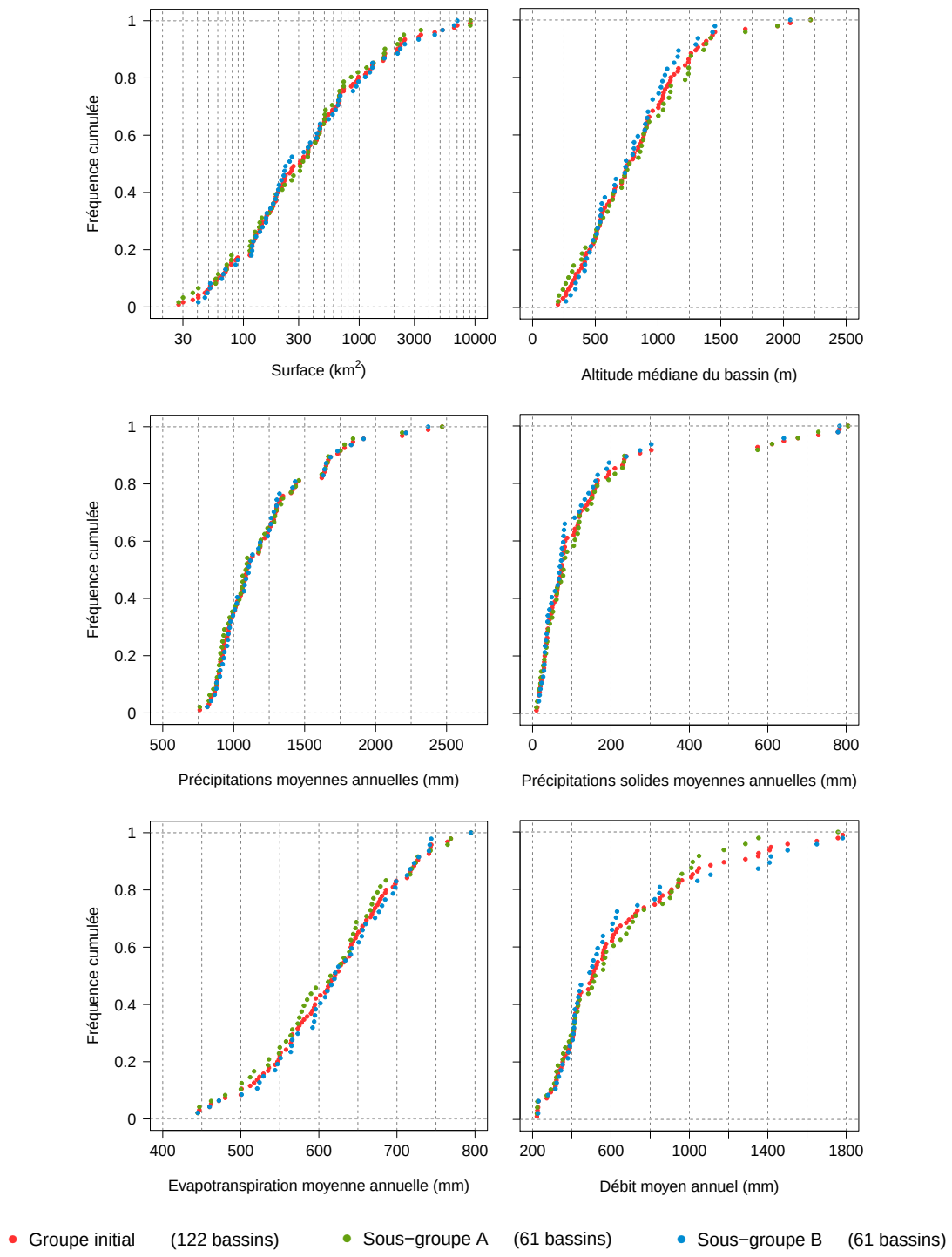


FIGURE H.5.: Résumé des caractéristiques des deux sous-groupes de 60 bassins utilisés dans le test de correction des paramètres de *GR4J* et *Mordor6*

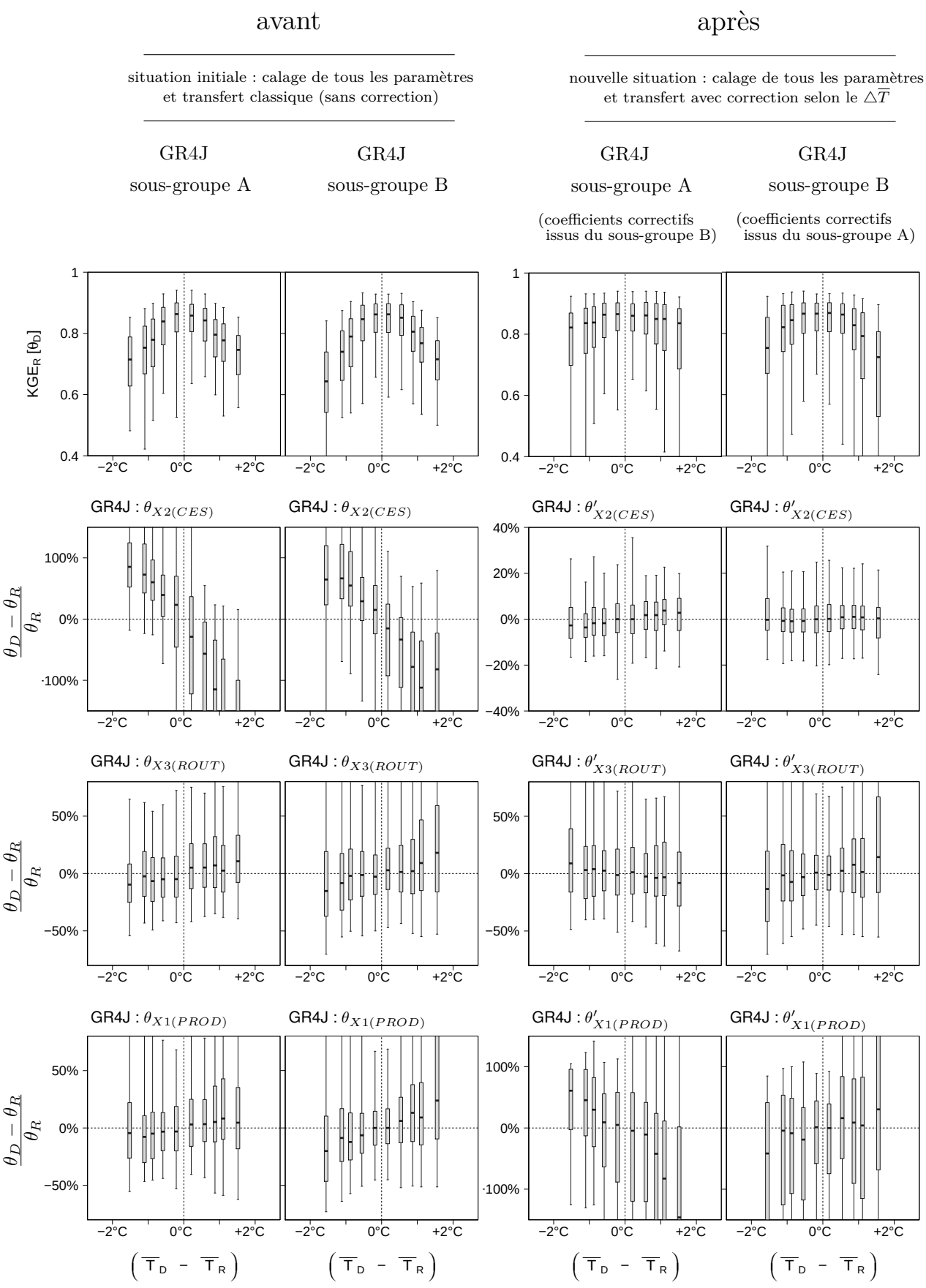


FIGURE H.6.: Pertes de performance et $\triangle\theta$ obtenus pour *GR4J* lors du test de correction des paramètres selon les conditions de transfert

Tests sur 120 bassins où $\triangle$erreurs=f($\triangle\overline{T}$), divisés en deux sous-groupes (A et B).
En France, GR4J alimenté par SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés).
Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

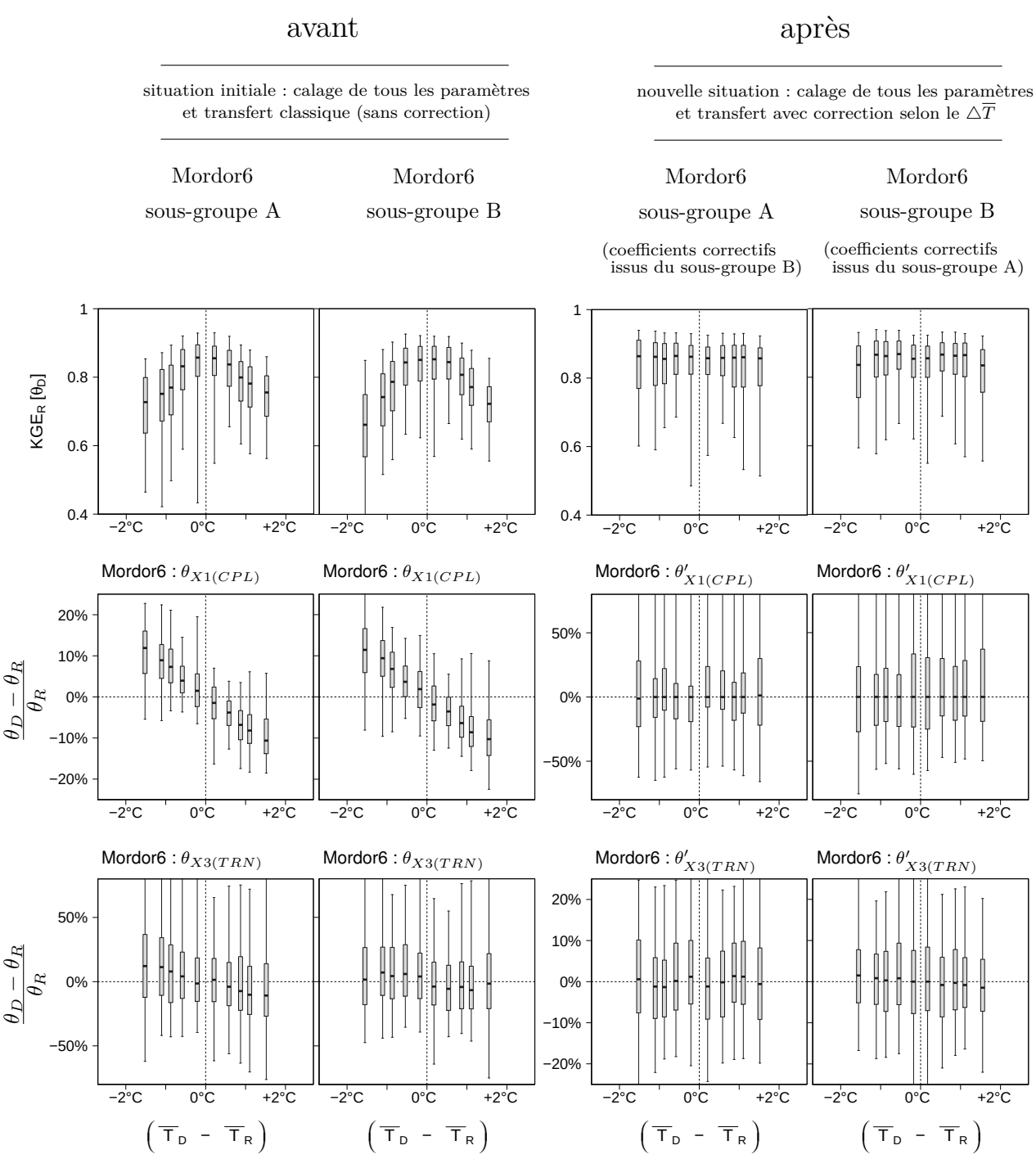
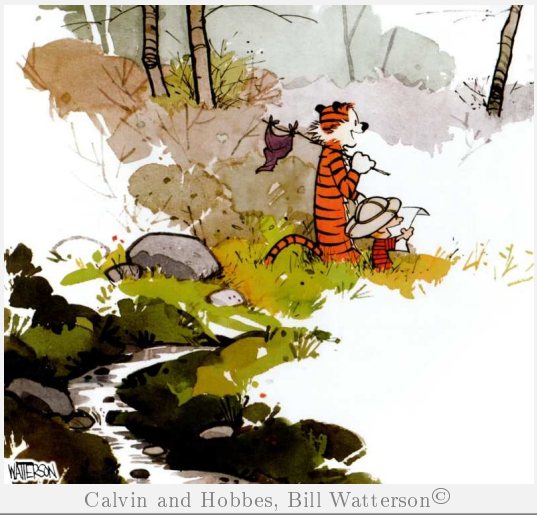


FIGURE H.7.: Pertes de performance et $\triangle\theta$ obtenus pour *Mordor6* lors du test de correction des paramètres selon les conditions de transfert

Tests sur 120 bassins où $\triangle$erreurs=f($\triangle\overline{T}$), divisés en deux sous-groupes (A et B). En France, Mordor6 alimenté par SPAZM, une ETP d'Oudin et couplé à CemaNeige (θ_{neige} fixés). Boîtes à moustaches indiquant les quantiles 5, 25, 50, 75 et 95% pour chaque bande verticale.

Et maintenant, où allons-nous ?



Calvin and Hobbes, Bill Watterson©

Thèse de doctorat

Laurent Coron

Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution ?

Résumé

Les hydrologues sont questionnés au sujet de l'évolution à moyen et long terme de la ressource en eau. Pour répondre à ces questions, ils ont recours à des modèles hydrologiques. Une quantification des incertitudes associées à leurs projections leur est souvent demandée et pose la question de la robustesse de ces modèles conceptuels, notamment dans le contexte d'un climat en évolution. En effet, l'utilisation d'un modèle dans des conditions différentes de celles où il a été calé repose sur une hypothèse de transférabilité des paramètres, c'est-à-dire la possibilité d'utiliser les paramètres d'un modèle dans des conditions différentes de celles de sa construction.

Nous traitons cette problématique en tentant d'apporter des éléments de réponse aux questions suivantes :

- Quel est le niveau de robustesse des modèles hydrologiques conceptuels face à un climat en évolution ?
- Quelles sont les causes des situations de manque de robustesse et y a-t-il des façons de s'en prémunir ?

Nous y répondons à travers l'étude des performances de modèles conceptuels lors de nombreux tests de transfert temporel de leurs paramètres. Les résultats montrent des corrélations entre les problèmes de robustesse et les écarts de conditions climatiques entre les périodes de calage et de validation des modèles. Des situations de biais systématiques corrélés à des changements de température de l'air ressortent en particulier de ces analyses. Une hétérogénéité existe cependant sur notre échantillon, les variables climatiques ou le type d'erreurs associées aux problèmes détectés pouvant varier d'un bassin à un autre.

L'étude des biais de simulation sur les bassins où les modèles ne sont pas robustes montre des alternances de phases de sous-estimation et surestimation des débits, le biais sur le module décennal pouvant atteindre 20%. Nos travaux révèlent que des comportements très similaires sont obtenus pour diverses périodes et méthodes de calage. Pour les modèles conceptuels utilisés dans ces travaux, les problèmes de robustesse détectés ne résultent pas simplement de quelques "mauvais" calages conduisant à la sélection de paramètres inadaptés pour reproduire le comportement du bassin considéré. Ils semblent plutôt découler de difficultés générales de ces outils de modélisation à reproduire convenablement les bilans moyens de façon simultanée sur différentes périodes.

Ces travaux ouvrent des pistes de réflexion sur la capacité insuffisante de certains modèles hydrologiques conceptuels à reproduire des dynamiques basse fréquence et sur le rôle des erreurs d'estimation des forçages, notamment les variations temporelles de l'évapotranspiration, dans les situations de modélisation incorrecte.

Mots-clés

modélisation hydrologique ; robustesse ; évolutions climatiques ;
paramètres ; transférabilité temporelle ; calage ; bassins versants.