



HAL
open science

Pistes d'amélioration de la généralité et de l'efficacité d'un modèle opérationnel de prévision des crues

Paul C. Astagneau

► **To cite this version:**

Paul C. Astagneau. Pistes d'amélioration de la généralité et de l'efficacité d'un modèle opérationnel de prévision des crues. Hydrologie. Sorbonne Université, 2022. Français. NNT : 2022SORUS486 . tel-04028903

HAL Id: tel-04028903

<https://theses.hal.science/tel-04028903>

Submitted on 14 Mar 2023

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.



Sorbonne Université

École doctorale Géosciences, Ressources naturelles et Environnement

INRAE, UR HYCAR, Équipe Hydrologie des bassins versants

Pistes d'amélioration de la généralité et de l'efficacité d'un modèle opérationnel de prévision des crues

Par **Paul Astagneau**

Thèse de doctorat en hydrologie

Dirigée par Vazken Andréassian et François Bourgin

Présentée et soutenue publiquement le 8 décembre 2022

Devant un jury composé de :

Éric GAUME	Université Gustave Eiffel, Nantes	Rapporteur
Hélène ROUX	INP, Toulouse	Rapporteuse
Anne-Catherine FAVRE	Université Grenoble Alpes	Examinatrice
Renaud MARTY	DREAL Centre-Val-de-Loire, Orléans	Examinateur
Ludovic OUDIN	Sorbonne Université, Paris	Examinateur
Pierre RIBSTEIN	Sorbonne Université, Paris	Président du jury
Vazken ANDRÉASSIAN	INRAE, Antony	Directeur de thèse
François BOURGIN	INRAE, Antony	Co-encadrant
Lionel BERTHET	MTE, DGPR/SNRH, Puteaux	Invité

Remerciements

Qui remercier (ou accabler) si ce ne sont les autres ou le hasard ? Je ne vais certainement pas m'auto-congratuler, je n'y suis pour rien dans cette histoire. Vous ne verrez donc, dans les lignes qui vont suivre, aucune mention à un soi-disant mérite qui pourrait m'être attribué. La réussite (ou l'échec) de cette thèse peut finalement être attribuée au hasard, aux constructions sociales qui m'ont toujours placé dans une position favorable dans la société et, surtout, aux autres (seulement en cas de réussite). Les remerciements placés au début d'un manuscrit de thèse ne permettent jamais de témoigner pleinement de la reconnaissance de l'impétrant à l'égard de l'ensemble des personnes qui l'ont, de près ou de loin soutenu. Je vais toutefois entamer une liste plus ou moins exhaustive des personnes sans qui cette thèse n'aurait pu aboutir (ni même démarrer).

Tout d'abord, je souhaiterais remercier mon directeur de thèse, Vazken Andréassian qui m'a appris que la recherche est une affaire d'intuitions à la fois personnelles et collectives (dont la majorité se révèlent être des échecs) avant d'être un travail de synthèse et d'apprentissage. Vazken m'a aussi montré que ce qui paraît beau et compliqué cache souvent un loup. Quelques soient nos aptitudes à formaliser des problèmes en utilisant des outils mathématiques, il vaut mieux ne jamais oublier les hypothèses sur lesquelles reposent un travail de recherche. Je remercie Vazken également pour les nombreuses discussions passionnantes que nous avons eues autour d'un tableau blanc pour essayer de développer nos intuitions respectives, en particulier pendant la pandémie de Covid-19 où les interactions scientifiques réelles manquaient cruellement.

J'aimerais ensuite remercier François Bourgin qui a passé un temps considérable à encadrer mes travaux. Malgré mes débuts balbutiants, mes questions parfois stupides et à côté de la plaque, François est resté patient et a trouvé le juste équilibre entre rectification et place à l'autonomie. Je le remercie particulièrement de m'avoir poussé à placer mes travaux de recherche dans une direction un peu différente des nombreux travaux sur le sujet. Merci aussi de m'avoir incité à rester à jour sur les différentes évolutions de la modélisation hydrologique ainsi que d'avoir proposé l'utilisation de différents outils que je n'aurais jamais imaginé pouvoir manipuler (avec toujours une phrase pour me rassurer sur la facilité d'utilisation de ces outils).

Cette thèse n'aurait jamais vu le jour sans l'encadrement et le temps précieux de Charles Perrin que je souhaiterais remercier infiniment ici. Charles a été d'un soutien inconditionnel tout au long de mon travail, me rassurant régulièrement sur mes multiples interrogations concernant la validité de mes recherches et ma capacité à les poursuivre. Charles est toujours d'un calme olympien qu'on ne saurait apprécier à sa juste valeur tant cela permet de prendre le temps de réfléchir correctement. Malgré ses nombreux commentaires (« de détail ») sur toutes les productions écrites qui ont alimenté ce mémoire, une hiérarchisation des modifications les plus importantes y était toujours associée, ce qui a souvent eu un impact très positif sur mon niveau de stress parfois (souvent) élevé.

Cette thèse n'aurait aussi jamais vu le jour sans le stage qui l'a précédée. Je remercie ainsi mes deux encadrants de stage, Guillaume Thirel et Olivier Delaigue de m'avoir fait confiance en m'initiant à la modélisation hydrologique et de m'avoir guidé et soutenu pour poursuivre en thèse. Je remercie Guillaume pour cette introduction à la rigueur scientifique et rédactionnelle bien illustrée par un article ayant fait l'objet d'interminables relectures par les pairs. Je remercie particulièrement Olivier pour les nombreuses heures passées à m'aider sur de nombreux codes informatiques, et surtout pour tout le travail qu'il investit dans l'équipe, que ce soit pour l'aide immense sur les codes informatiques, les figures, les données, les SIG (par le biais d'une formation de renommée internationale!) ou encore pour l'organisation des cagnottes de départ, des nouveaux arrivants...

J'étends ces remerciements à Jacques de la Rocque, sans qui le stage précédemment mentionné n'aurait jamais été possible puisqu'initié par le stage qui l'a précédé. Je le remercie d'avoir accepté de prendre en stage son neveu paumé et inexpérimenté en le mettant sur le chemin de l'hydrologie.

Je souhaiterais ensuite remercier les sept membres du jury d'avoir accepté d'évaluer cette thèse. Je remercie d'abord la rapportrice de ce travail, Hélène Roux et le rapporteur, Éric Gaume pour leur lecture attentive du manuscrit ainsi que pour les discussions constructives et passionnantes que nous avons eues lors de la soutenance. Je remercie les examinateurs et examinatrices, Anne-Catherine Favre, Renaud Marty et Ludovic Oudin, et Lionel Berthet, invité, d'avoir participé à l'évaluation de mes travaux, en particulier pour s'être positionné·e·s sur des grilles de lecture différentes mais complémentaires des interventions des orateurs et oratrices précédent·e·s. Je remercie le président du jury, Pierre Ribstein d'avoir accepté de mener les questions lors de la soutenance et surtout d'avoir posé des questions courtes et directes permettant de mettre fin à une soutenance de plus de 3 h. J'étends ces remerciements aux membres du comité de suivi, Flora Branger, Pierre-André Garambois, Nicolas Le Moine, Simon Munier et Félicien Zuber, qui ont pris de leur temps pour suivre et aiguiller ce travail depuis ses débuts. Je profite de cette extension des remerciements pour remercier le Service central d'hydrométéorologie et d'appui à la prévision des inondations, financeur de cette thèse.

Merci aux membres de l'équipe Hydro et aux collègues d'INRAE Antony : Alban, Daniela, Helena, Léonard, Paul, François, Gaëlle, Arnaud, Myriam, Cyril, Anne-Lise, Lila, Pierre, Julie, Laure, José, Manon, Morgane, Anthony, Guilherme, Sebastián, Laurent, Ralph, Sarah, Thibault, Taha, Swann, Cédric, Gaïa, Valentin, Benoît, Bruno, Jean-Baptiste, Matthieu, Sylvain, Yannick, Julien, Hocine, Alexis, Timo, Samy, Florent, Marion, Aya, Aliénor, Marie-Line, Valérie, Nathalie, Stéphanie, Hichem, Benjamin, Laura, Saran, Vandoir, Chiara ; pour tous les moments que nous avons passés ensemble aux abords d'un couloir pour discuter de science, de politique ou que sais-je, pendant les nombreuses séances de volley et de frisbee ou encore autour d'une bière bien méritée. Merci de m'avoir supporté, merci pour votre bienveillance, merci pour votre aide, merci pour votre bonne humeur, merci pour vos conseils, merci d'avoir fait partie de ma vie. Cette thèse, je vous la dois entièrement. Je vous ai rejoint en tant qu'étudiant, perdu et sans aspiration. Je pars en tant qu'hydrologue, heureux, reconnaissant et plein d'espérance. J'espère que, quelles que soient les circonstances, nos chemins se recroiseront.

J'adresse une mention spéciale à mes trois camarades du bureau 112 : merci Thibault d'avoir partagé ce chemin commun du stage à la soutenance de thèse, pour notre soutien mutuel dans les moments difficiles et pour notre « bureau » des plaintes ; merci Antoine pour nos discussions interminables (10 % de sciences et 90 % de politique et actualités diverses et variées) ainsi que pour ton grand soutien administratif et mathématique ; merci Shu-Chen pour ta gentillesse tout au long de cette dernière année de thèse rythmée par des agacements réguliers devant des codes informatiques non conciliants. De nombreuses mentions spécifiques pourraient être attribuées à la majorité des personnes mentionnées plus haut, mais cette thèse et ces remerciements sont déjà bien assez longs, et les personnes concernées savent (je l'espère) à quel point elles ont participé à l'aboutissement de ces travaux ainsi qu'à mon épanouissement.

Il est de coutume de mentionner les ami·e·s et la famille comme des bienfaiteur·e·s et des personnes indispensables dans la formation, l'évolution, la construction, la conception, l'alimentation, le bonheur, et le creusement des cernes du thésard. Je ne dérogerai pas à cette règle aujourd'hui. Je remercie avec tout l'amour que je leur porte, ma mère, mon père, mes frères et mes ami·e·s sans qui cette vie n'aurait pas beaucoup de sens.

Je souhaiterais finir ces quelques phrases sur un remerciement général aux personnes qui m'ont montré et me montrent toujours qu'une autre voie que celle tracée par notre propre bulle sociale est possible dans la mesure où l'on apprend à comprendre nos constructions, nos peurs et nos désirs. J'espère qu'ils se reconnaîtront dans ce message.

Résumé

Les inondations font régulièrement des victimes et occasionnent des pertes financières majeures partout dans le monde. Afin d'anticiper au mieux ce phénomène naturel, des systèmes opérationnels de prévision des crues ont été mis en place dans de nombreux pays. Ces systèmes reposent sur des modèles hydrologiques qui transforment des observations et prévisions météorologiques en prévisions de débit dans les rivières. Malgré les nombreuses améliorations apportées aux systèmes de prévision des crues au cours des deux dernières décennies, les prévisions des modèles hydrologiques restent entachées de fortes incertitudes.

En France, le modèle hydrologique GRP est utilisé par la plupart des services opérationnels de prévision des crues du réseau national Vigicrues. Les retours d'expérience ont permis d'identifier plusieurs limitations du modèle, en particulier une sous-estimation fréquente des volumes et pics de crue. Plusieurs aspects de la chaîne de modélisation hydrologique peuvent contribuer à ces problèmes, parmi lesquels la structure du modèle, l'estimation des paramètres ou encore la méthode d'assimilation de données employée. Notre travail de recherche a eu pour objectif de mieux comprendre, par une approche de diagnostic détaillé, les raisons des limites du modèle GRP et d'améliorer la qualité de ses prévisions, en cherchant le degré de complexification adapté au niveau d'information couramment disponible à l'échelle des bassins versants.

Ce travail est basé sur une grande base de données de 229 bassins versants français répartis sur le territoire métropolitain, sur lesquels 10 652 événements de crue ont été sélectionnés. Dans la première partie du manuscrit, nous caractérisons les erreurs en crue du modèle de simulation et cherchons les facteurs pouvant les expliquer. Dans la deuxième partie, nous utilisons l'information donnée par l'intensité de pluie au pas de temps horaire pour modifier le comportement du modèle afin d'améliorer la simulation des crues qui ont lieu à la fin de l'été et au début de l'automne. Dans la dernière partie, nous évaluons si l'amélioration de la structure du modèle améliore la prévision des crues.

Les résultats montrent que le modèle sous-estime fortement le volume des crues qui résulte de fortes intensités de pluie, en particulier en condition de faible humidité antécédente des bassins. Ces sous-estimations sont parfois dues à une mauvaise estimation de la pluie efficace mais plus largement à une dynamique trop lente du routage lorsque la réaction des bassins est très rapide. La modification des flux internes du modèle à partir de l'intensité de pluie horaire permet d'améliorer la simulation des crues estivales sans dégrader la simulation des autres événements. Ces modifications de la structure permettent d'améliorer la qualité des prévisions des crues pour les échéances supérieures à 6 h. Ces améliorations donnent toutefois lieu à une complexification de l'estimation des paramètres et de la méthode d'assimilation. Plusieurs solutions sont proposées pour tenter de dépasser les problèmes d'estimation des paramètres. Plus généralement, les résultats présentés dans ces travaux mettent en évidence les interactions entre le choix de la structure du modèle, l'estimation des paramètres et la méthode d'assimilation employée, et suggèrent donc d'adopter une approche intégrant simultanément ces différentes composantes dans les démarches d'amélioration des modèles de prévision des crues.

La nouvelle version du modèle de prévision proposée à l'issue de ce travail devrait en augmenter l'efficacité globale en conditions opérationnelles, et élargir son applicabilité à des types d'événements et des contextes hydroclimatiques variés.

Abstract

Floods frequently have disastrous human impacts and cause heavy economic losses worldwide. In order to anticipate this natural phenomenon, operational flood forecasting systems have been implemented in many countries. These systems are based on hydrological models that turn meteorological observations and forecasts into streamflow forecasts. Although many improvements have been made over the last two decades, flood forecasts are still associated with a high level of uncertainty.

The GRP hydrological model is used by most of the French operational flood forecasting services (Vigicrues network). Based on feedback from forecasters, some model deficiencies have been identified, in particular systematic underestimation of flood peaks and volumes. Several attributes of the hydrological modelling chain contribute to these issues, among which the model structure, the parameterization technique and the data assimilation procedure. The objective of this work was to better understand the causes of the model failures through a detailed diagnosis framework and to improve the quality of the GRP model's forecasts while searching for the level of complexity compatible with the level of information commonly available at the catchment scale.

This work relies on a large dataset of 229 continental French catchments where 10,652 flood events were selected. In the first part of this manuscript, the model errors in simulating floods are characterized and the explanatory factors are identified. In the second part, hourly rainfall intensities are used to dynamically modify the model's internal fluxes, to improve the simulation of floods occurring in summer and at the beginning of autumn. In the last part, we assess whether this change in model structure leads to better flood forecasts.

Results show that the model highly underestimates flood volumes in summer when high-intensity rainfall events occur, especially under low antecedent soil moisture conditions. These underestimations are sometimes due to poor estimation of effective rainfall but are mainly caused by a too slow runoff routing when the catchment response to rainfall is flashy. Dynamically modifying the internal fluxes of the model with the information provided by hourly rainfall intensities improves flood simulation in summer while maintaining the same level of performance for other flood types. These modifications lead to improvements in flood forecasting for lead times larger than six hours. However, these modifications result in an increase in parameter estimation complexity and in data assimilation complexity. Several options are proposed to overcome this issue and improve parameter estimation. Our results highlight the interactions existing between the choice of model structure, parameter estimation and the data assimilation procedure, which suggests that the improvement of forecasting models should consider all these aspects simultaneously in an integrated approach.

The new flood forecasting model version that was developed in this work should lead to an increase in its overall efficiency in operational conditions and broaden its applicability to varied flood types and hydroclimatic conditions.

Sommaire

Remerciements	iii
Résumé	vii
Introduction générale	1
Contexte	1
Positionnement scientifique et questions de recherche	4
Objectif de la thèse et choix méthodologiques	7
Organisation du mémoire	8
1 Base de données et éléments méthodologiques	11
1.1 Sélection des bassins versants	11
1.2 Modifications de la base de données	13
1.3 Sélection des événements de crues	16
1.4 Caractéristiques des événements	18
1.5 Conclusions sur la base de données	18
1.6 Modèles hydrologiques	20
1.7 Estimation des paramètres et évaluation des performances	27
I Diagnostic approfondi du modèle de simulation : des erreurs aux facteurs explicatifs	29
2 Diagnostic du modèle de simulation GR5H	31
2.1 Résumé	31
2.2 Introduction	32
2.3 Data	34
2.4 Methods	35
2.5 Results	40
2.6 Discussion	43
2.7 Conclusion	48
3 Influence des caractéristiques de la pluie sur les erreurs du modèle	51
3.1 Contexte	51
3.2 Variabilité événementielle par type de bassin	51
3.3 Caractéristiques des événements de pluie et erreurs du modèle	54
3.4 Conclusion	62

4	Analyse des flux internes à l'échelle événementielle	63
4.1	Contexte	63
4.2	Éléments méthodologiques	64
4.3	Résultats et analyses	66
4.4	Conclusion	74
 II Vers une modélisation généraliste plus adaptée à simuler les crues résultant de fortes intensités de pluie		75
5	Modification de la structure du modèle GR5H	77
5.1	Résumé	77
5.2	Introduction	78
5.3	Data and methods	81
5.4	Results	89
5.5	Discussion	95
5.6	Conclusions	99
5.7	Compléments	101
6	Que peut-on attendre d'une modélisation semi-distribuée ?	105
6.1	Contexte	105
6.2	Éléments méthodologiques	107
6.3	Résultats et analyses	108
6.4	Conclusion	117
7	Vers une estimation des paramètres plus adaptée pour simuler les crues	119
7.1	Contexte	119
7.2	Éléments méthodologiques	121
7.3	Résultats et analyses	123
7.4	Conclusion	139
 III Impact de l'amélioration de la structure du modèle hydrologique sur la prévision des crues		141
8	Construction et évaluation d'un modèle de prévision	143
8.1	Contexte	143
8.2	Éléments méthodologiques	145
8.3	Résultats et analyses	152
8.4	Conclusion	171
9	Vers une estimation des paramètres adaptée à la prévision	173
9.1	Contexte	173
9.2	Éléments méthodologiques	174
9.3	Résultats et analyses	175
9.4	Conclusion	183

Conclusion générale	187
Synthèse de la démarche et des résultats	187
Réponses aux questions de recherche	189
Perspectives	191
Bibliographie	193

Liste des figures

1.1	Localisation des 229 bassins versants en France métropolitaine.	13
1.2	Classification des 229 bassins versants selon leur régime hydrologique.	14
1.3	Positionnement des 229 bassins versants sur le graphique adimensionnel de Turc-Budyko.	16
1.4	Exemple de visualisation d'un hydrogramme de crue sélectionné sur le bassin versant la Bruche à Holtzheim.	17
1.5	Répartition spatiale des caractéristiques hydroclimatiques des 10 652 événements de crue sélectionnés (calculées sur la période 2000-2018).	19
1.6	Distribution des caractéristiques hydroclimatiques des 10 652 événements de crue sélectionnés (calculées sur la période 2000-2018).	20
1.7	Schéma du modèle GR5H.	21
1.8	Schéma du modèle GRSD.	25
1.9	Schéma du modèle GRP.	25
2.1	Distribution and localization of the characteristics of 2,990 flood events.	36
2.2	Distribution of catchment seasonal flood volumes compared with seasonal streamflow volume for 2,990 flood events in 229 catchments.	37
2.3	Illustration of the streamflow time windows used to calculate the event bias per catchment and the bias outside periods with events per catchment.	38
2.4	Distribution of GR5H performance in calibration and validation periods over 229 French catchments.	40
2.5	Difference of (a) cumulative rainfall and KGE (b) runoff coefficient and KGE in validation mode between P1 and P2.	41
2.6	Distribution of GR5H bias over 229 catchments and 2,990 events.	41
2.7	Seasonality of GR5H bias over 229 catchments and 2,990 events.	42
2.8	Examples of simulations of summer floods compared with floods that occurred in other seasons.	44
2.9	(a) to (e) present the results of a univariate analysis of the relationships between flood characteristics and event bias β_{ev}^b for 80 catchments in which floods occur in summer. (f) is the linear correlation matrix between the relative flood characteristics.	45
2.10	GR5H summer event bias per catchment plotted against KGE values for catchments where flood events occurred in summer.	46
2.11	(a) Relationship between GR5H simulation of event effective rainfall and event bias (each class contains approximatively 211 events). (b) Relationship between GR5H simulation of event effective rainfall and precipitation intensity relative to event duration.	47
3.1	Localisation des 43 bassins versants méditerranéens.	52
3.2	Répartition des performances globales du modèle GR5H pour deux groupes de bassins versants.	53

3.3	Répartition des performances événementielles du modèle GR5H selon deux groupes de bassins versants et deux types de crues.	54
3.4	Valeurs des indices de variabilité spatiale pour trois hyétogrammes synthétiques. . .	56
3.5	Répartition spatiale et saisonnière de quatre indices de pluie calculés sur 10 652 événements de 229 bassins versants français.	57
3.6	Corrélations linéaires (a) et de rang (b) entre les quatre indices caractérisant les 10 652 événements de pluie.	58
3.7	Répartition spatiale de l'intensité de la pluie moyennée par bassin en été pour quatre gammes de valeurs de la variabilité spatiale des cumuls de pluie.	58
3.8	Analyse univariée du lien entre le biais et le NSE en crue du modèle GR5H et deux caractéristiques des événements de pluie.	59
3.9	Analyse du lien entre la variabilité spatiale de la pluie, l'intensité de la pluie et le biais en crue (a) ou le critère NSE en crue (b) de GR5H.	60
3.10	Analyse du lien entre la variabilité spatiale de la pluie, l'intensité de la pluie, les conditions antécédentes d'humidité du sol (SWI) et les performances en crue de GR5H.	61
4.1	Comparaison entre la pluie efficace simulée par GR5H et le volume de crue observé pour les événements d'été sous-estimés.	66
4.2	Liens entre variabilité spatiale de la pluie, cumul de pluie et deux groupes en fonction de la pluie efficace.	67
4.3	Flux internes de GR5H pour l'hydrogramme de crue du Gardon de Saint-Jean à Corbès en septembre 2014.	68
4.4	Volume d'eau additionnel stocké dans le réservoir de routage par rapport au volume de la pluie efficace en fonction du biais pour les événements estivaux sous-estimés par GR5H.	68
4.5	Distribution du biais événementiel, de la durée des événements, de l'intensité de la pluie et de la variabilité spatiale de la pluie en fonction du retard du débit simulé en été.	69
4.6	Variation du volume d'erreur pour les événements d'été en fonction du volume échangé.	70
4.7	Relations entre les volumes échangés en dehors des événements et le biais en crue en été.	71
4.8	Variation du biais événementiel en fonction du niveau antécédent du réservoir de production et de l'ETR antécédente calculée sur une durée de un mois avant l'événement.	72
4.9	Variations du biais événementiel en fonction de l'ETR antécédente et de l'ETP antécédente pour tous les événements de la base de données.	73
4.10	Liens entre le SWI antécédent (simulation issue du modèle ISBA de Météo-France) et le remplissage antécédent du réservoir de production pour différentes valeurs du ratio ETR sur ETP.	73
5.1	Location of the five hydroclimatic groups of catchments and distributions of four characteristics in these groups.	83
5.2	Diagram of the modifications made to the original GR5H model.	84
5.3	Variation in production rate given a production store level and a rainfall intensity rate.	86
5.4	Variation in production rate when the production store is theoretically empty. . . .	86
5.5	Variation in effective rainfall partitioning with rainfall intensities.	87
5.6	Overall performance (a), performance in simulating 8290 winter events and 2362 summer events (b) and (c), performance outside flood periods (d) for the three modeling hypotheses compared to the reference model.	89

5.7	Distribution of flood performance over 2362 summer floods divided into five groups of catchments.	91
5.8	Variation in event performance against three event characteristics for two groups of catchments.	92
5.9	Simulated and observed hydrographs of six floods that occurred in the Loup River at Tourrettes-sur-Loup and the Estéron River at Sigale.	93
5.10	Variation in additional M3-Pr-B free parameters between calibration periods.	94
5.11	Catchments on which the intensity functions of M3-Pr-B are activated on both calibration periods.	95
5.12	Distributions of the additional M3-Pr-B free parameters for five groups of catchments.	95
5.13	Sensitivity of the additional parameters to the composite criterion E_{comp} over the 229 catchments (a). Distribution of performance in simulating 8290 winter floods and 2362 summer floods for the third modeling hypothesis run with generic i_1 and i_2 values in validation (b and c).	97
5.14	Distribution de six caractéristiques hydroclimatiques pour les cinq groupes de bassins versants. Les distributions sont présentées entre les centiles 5 % et 95 %.	104
6.1	Distribution des performances globales et événementielles en validation croisée pour le modèle de référence, le modèle semi-distribué et le modèle comprenant les fonctions d'intensité.	109
6.2	Comparaison entre les performances événementielles du modèle de référence et le modèle semi-distribué pour cinq groupes de bassins versants (a). Comparaison entre les performances événementielles du modèle avec les fonctions d'intensité et le modèle semi-distribué pour cinq groupes de bassins versants (b).	110
6.3	Distribution des performances événementielles du modèle de référence, du modèle semi-distribué et du modèle avec les fonctions d'intensité en fonction de l'intensité de la pluie (a) et de la variabilité spatiale des cumuls de pluie (b).	111
6.4	Cartes des cumuls de pluie et hydrogrammes observé et simulés de trois événements du bassin versant de la Cèze à Tharoux.	112
6.5	Cartes des cumuls de pluie et hydrogrammes observé et simulés de trois événements du bassin versant du Vistre au Cailar.	113
6.6	Distribution et répartition spatiale des performances événementielles de M0_SD10 et M0_SD50 pour 84 bassins de moins de 250 km ²	116
6.7	Distribution des performances événementielles de M0_SD10 et M0_SD50 pour les 10 % d'événements les plus intenses et les 10 % d'événements les plus variables spatialement pour 84 bassins de moins de 250 km ²	116
7.1	Variation des paramètres du modèle M3 entre les périodes de calage P1 et P2 sur les 229 bassins versants.	123
7.2	Corrélations des paramètres du modèle M3 entre périodes de calage selon deux méthodes d'estimation des paramètres.	124
7.3	Performances globales des modèles M0 et M3 pour deux méthodes d'estimation des paramètres.	124
7.4	Surfaces de réponse des paramètres du modèle M0 en fonction du critère KGE pour le Gardon de Saint-Jean à Corbès.	126
7.5	Surfaces de réponse des paramètres des fonctions d'intensité du modèle M3 en fonction du critère KGE pour le bassin du Gardon de Saint-Jean à Corbès.	127
7.6	Distribution de la variation des paramètres par bassin et par période.	128
7.7	Interactions entre les paramètres du modèle M0 pour le bassin du Doux à Colombier-le-Vieux.	129

7.8	Distribution du taux de remplissage moyen du réservoir de routage du modèle M0 en validation pour les 229 bassins versants (a). Courbe de vidange du réservoir de routage en fin de pas de temps (b).	130
7.9	Distribution de la variation des paramètres de M0 et des paramètres des fonctions d'intensité de M3 par bassin et par période en fonction de cinq critères d'erreur.	131
7.10	Performances événementielles en été en fonction des performances globales pour plusieurs jeux de paramètres des modèles M0 et M3 sur le bassin du Gardon de Saint-Jean à Corbès.	132
7.11	Distribution des performances globales (KGE) du modèle M0 en calage et validation pour plusieurs valeurs du paramètre X_5 fixé pour tous les bassins par période considérée.	133
7.12	Distribution des performances événementielles du modèle M0 (NSEev) en hiver et en été pour plusieurs valeurs du paramètre X_5 fixé pour tous les bassins par période considérée.	133
7.13	Distribution des performances globales (KGE) du modèle M3 en calage et validation pour plusieurs valeurs du paramètre K_L fixé pour tous les bassins par période considérée.	135
7.14	Distribution des performances événementielles (NSEev) du modèle M3 en hiver et en été pour plusieurs valeurs du paramètre K_L fixé pour tous les bassins par période considérée.	136
7.15	Comparaison des performances globales et événementielles des modèles M0 et M3 pour deux méthodes d'estimation des paramètres.	137
7.16	Comparaison des performances événementielles du modèle M3 en été et en validation pour deux méthodes d'estimation des paramètres en fonction de cinq groupes de bassins versants.	138
7.17	Distributions des valeurs des paramètres des fonctions d'intensité du modèle M3 pour deux méthodes d'estimation des paramètres.	139
8.1	Schéma représentant la chaîne de modélisation hydrologique utilisée pour la prévision des crues.	143
8.2	Différences entre la simulation et la prévision des crues.	146
8.3	Schéma représentant la mise à jour du réservoir de routage de GRP par insertion directe du dernier débit observé.	147
8.4	Schéma représentant la mise à jour des réservoirs du modèle M3 par insertion directe du dernier débit observé à l'instant de prévision.	148
8.5	Différence entre les prévisions à instant de prévision fixe et les prévisions à horizon fixe.	149
8.6	Schéma représentant le décalage temporel et l'amplitude de l'erreur des pics de deux prévisions (à horizon fixe) d'un événement de crue du l'Yzeron à Craponne en août 2007.	151
8.7	Distribution des performances globales des modèles GRP, M0 et M3 pour cinq échéances de prévision et <i>l'open loop</i> .	153
8.8	Répartition spatiale du ratio des performances M3/GRP en prévision selon le critère de persistance borné.	155
8.9	Distribution du ratio entre les performances globales (RMSE) des modèles GRP, M0 et M3 pour cinq échéances de prévision et les performances de leur version <i>open loop</i> (OL) respective.	156
8.10	Distribution des performances (RMSE) des modèles GRP, M0 et M3 sur les montées de crue pour cinq échéances de prévision et <i>l'open loop</i> (OL).	157
8.11	Distribution des performances des modèles GRP, M0 et M3 sur les pics de crue pour cinq échéances de prévision et <i>l'open loop</i> .	158

8.12	Distribution des performances événementielles (NSE) des modèles GRP, M0 et M3 pour cinq échéances de prévision et <i>l'open loop</i>	159
8.13	Répartition spatiale du ratio des performances événementielles M3/GRP en prévision selon le critère NSE.	160
8.14	Distribution des critères de dépassement de seuil pour les modèles GRP, M0 et M3 pour cinq échéances de prévision.	162
8.15	Distribution du ratio des RMSE sur les montées de crue entre M3 et GRP en fonction du temps de montée des événements.	163
8.16	Distribution des performances événementielles (NSE) des modèles GRP et M3 en fonction du temps de réaction des bassins versants.	164
8.17	Comparaison des performances événementielles moyennées par bassin et par échéance de prévision (RMSE sur les montées de crue) pour cinq groupes de bassins versants.	165
8.18	Distribution des performances de GRP et M3 sur les pointes de crue en fonction de l'intensité de la pluie moyenne des événements.	166
8.19	Distribution des performances de GRP et M3 sur les pointes de crue en fonction des conditions antécédentes d'humidité du sol (SWI).	167
8.20	Exemples d'hydrogrammes de crue prévus par les modèles GRP et M3 sur quatre bassins versants pour trois échéances de prévision.	168
8.21	Comparaison des performances globales et événementielles des modèles GRP, M0 et M3 utilisés avec et sans méthode de correction des sorties pour quatre échéances de prévision.	170
9.1	Distribution des performances globales et événementielles du modèle M3 selon cinq méthodes d'estimation des paramètres pour quatre échéances de prévision.	176
9.2	Distribution du paramètre K_L selon deux méthodes d'estimation des paramètres du modèle M3 et quatre horizons de prévision.	178
9.3	Exemple d'une crue prévue sur le bassin du Loyre à Voutezac par le modèle M3 selon deux méthodes d'estimation des paramètres : la méthode de référence (a) et la méthode avec les paramètres X_5 et K_L fixés (b).	180
9.4	Distributions des valeurs absolues de la différence entre les paramètres sélectionnés par horizon et leur valeur en version OL en fonction de l'horizon de prévision.	181
9.5	Corrélations des valeurs des paramètres du modèle M3 entre périodes de calage selon quatre horizons de prévision et l'OL.	182
9.6	Dispersion des valeurs des paramètres X_3 et i_2 du modèle M3 entre périodes de calage selon quatre horizons de prévision et l'OL.	183
9.7	Comparaison des performances du modèle M3 entre l'OL et l'échéance 3 h pour plusieurs jeux de paramètres. Exemple sur trois bassins versants. Les points rouges représentent les optima mathématiques.	184

Liste des tableaux

1.1	Distribution de neuf caractéristiques hydro-climatiques et cinq caractéristiques morphologiques des 229 bassins versants.	15
1.2	Liste des indicateurs utilisés pour caractériser les événements sélectionnés.	18
1.3	Liste des variables du modèle GR5H.	23
1.4	Liste des paramètres libres du modèle GR5H.	23
1.5	Liste des paramètres libres du modèle GRP.	26
1.6	Liste des critères de performance utilisés (tous ces critères sont adimensionnels).	28
2.1	Correspondence between bias and bounded bias values.	39
2.2	List of the relative characteristics used for the univariate analysis.	39
4.1	Liste des variables internes et indicateurs.	65
5.1	Event and catchment characteristics used for catchment clustering.	81
5.2	List of the four model versions tested.	88
5.3	Liste des modifications principales non présentées dans ce chapitre. Différentes combinaisons entre les modifications ont été testées.	101
7.1	Fonctions de transformation des paramètres de l'espace transformé vers l'espace réel.	122
7.2	Performances moyennes du modèle M0 pour deux méthodes d'estimation du paramètre X_5	134
7.3	Performances moyennes du modèle M3 pour deux méthodes d'estimation du paramètre K_L	135
7.4	Performances moyennes du modèle M3 avec les paramètres X_5 et K_L fixés à 0.125 et 0.115 respectivement.	135
8.1	Liste des notations des figures 8.3 et 8.4.	148
8.2	Table de contingence caractérisant le franchissement d'un seuil d'alerte.	151
8.3	Performances de GRP et M3 pour les hydrogrammes de la figure 8.20.	169
9.1	Méthodes d'estimation des paramètres testées pour le modèle M3 en prévision.	175

Introduction générale

Contexte

Enjeux de la prévision des crues

Les crues et les inondations qu'elles causent font régulièrement des victimes et occasionnent des destructions majeures partout dans le monde. Plusieurs exemples récents témoignent de la sévérité de ce phénomène naturel. En septembre 2020, la tempête Alex qui s'est abattue dans les Alpes-Maritimes a fait déborder les affluents du Var et de la Roya. Des milliers de personnes ont été affectées, près de 20 personnes ont été tuées, des dizaines d'habitations ont été dévastées et de nombreuses infrastructures ont été détruites (Carrega & Michelot, 2021). En juillet 2021, plusieurs pays européens ont été touchés par des crues dévastatrices, particulièrement en Allemagne où le débordement de l'Ahr, un affluent du Rhin, a causé la mort de 134 personnes (Szönyi *et al.*, 2022). Entre 2001 et 2020, les inondations ont été responsables de 22 % des pertes économiques liées aux catastrophes naturelles, s'élevant à plus de 682 milliards de dollars US (CRED, 2022).

Afin d'anticiper au mieux ce phénomène naturel, des systèmes de prévision opérationnelle des crues existent dans de nombreux pays (Pappenberger *et al.*, 2016). Ces systèmes utilisent des prévisions et observations météorologiques afin de prévoir les débits dans les rivières à des échéances allant de quelques heures à quelques jours (e.g. Pagano *et al.*, 2014 ; Wu *et al.*, 2020). Cette transformation est généralement effectuée à partir de modèles hydrologiques qui sont un élément important de ces systèmes de prévision. Les prévisions en temps réel produites par les modèles hydrologiques sont ensuite expertisées et transmises à la chaîne d'avertissement et d'alerte aux crues.

Les systèmes de prévision des crues existant dans le monde reposent sur différents outils pour prévoir les débits à des échelles spatiales allant du petit bassin versant non jaugé jusqu'à l'échelle régionale, voire continentale. Par exemple, le service australien de météorologie¹ est responsable de la prévision opérationnelle des crues sur tout le territoire australien (ce qui représente 14 fois la superficie de la France). Leur système combine plusieurs modèles hydrologiques de complexités différentes, dont les modèles GR4H (Mathevet, 2005) et Sacramento (Burnash, 1995), pour prévoir les crues des cours d'eau principaux, cela pour des échéances allant de un à sept jours (Hapuarachchi *et al.*, 2022). La prévision des crues soudaines n'est pas assurée directement mais des alertes météorologiques donnant une information sur ce risque de crues sont transmises aux autorités des états. Le système européen d'alerte aux inondations EFAS (Thielen *et al.*, 2009) fournit des prévisions de débit sur les principales

1. bom.gov.au/water/floods/index.shtml

rivières du continent jusqu'à 10 jours à l'avance. Leur système est basé sur le modèle hydrologique LISFLOOD (De Roo, 1999) qui est utilisé pour la prévision opérationnelle depuis 2012.

Malgré les nombreuses améliorations apportées aux systèmes de prévision des crues au cours des deux dernières décennies (e.g. Jain *et al.*, 2018 ; Zanchetta & Coulibaly, 2020), que ce soit au niveau des entrées météorologiques ou en ce qui concerne le développement de modèles plus performants, les prévisions de ces modèles restent entachées de fortes incertitudes (e.g. Berthet *et al.*, 2020 ; Troin *et al.*, 2021 ; Brunner *et al.*, 2021b). Les modèles hydrologiques ont notamment des capacités prédictives plus faibles sur les bassins arides et sur les bassins caractérisés par des conditions sèches à certaines saisons (e.g. McMillan *et al.*, 2016 ; Melsen *et al.*, 2018), en particulier lorsque des crues éclair ont lieu (e.g. Hapuarachchi *et al.*, 2011). Par ailleurs, l'aléa inondation devrait évoluer avec le changement climatique qui intensifie le cycle de l'eau (Douville *et al.*, 2021). L'intensité, la fréquence et la saisonnalité des crues sont modifiées par le changement climatique dans de nombreuses régions du monde (le sens des changements est variable selon les régions concernées). En France, il est établi que le cycle de l'eau s'intensifie et va continuer à s'intensifier en Méditerranée (Tramblay & Somot, 2018), induisant des changements du régime hydrologique de certains bassins versants (Tramblay *et al.*, 2021). Les risques associés à cet aléa sont et seront d'autant plus élevés que de plus en plus de personnes vivent dans des zones inondables dans le monde (Slater & Wilby, 2017). Pour cette raison, l'organisation météorologique mondiale (OMM) a engagé un plan prévoyant d'installer des systèmes de prévision couvrant la majorité des cours d'eau à risque d'inondation, dans tous les pays du monde avant la fin de l'année 2027². L'amélioration de la capacité prédictive des modèles hydrologiques est donc essentielle pour une meilleure anticipation des inondations. Améliorer les modèles de prévision des crues permet, entre autres, d'émettre des alertes plus en avance et ainsi réduire l'impact sur les personnes et les biens.

Prévision opérationnelle des crues en France

En France, l'État organise la gestion des risques naturels à travers les plans de prévention des risques naturels (PPRN), établis par les préfetures³. Le Ministère de la transition écologique a établi une feuille de route sur la prévision des risques naturels à horizon 2025. L'amélioration des outils de prévision des crues fait partie des actions stratégiques prioritaires de cette feuille de route.

L'État organise la surveillance et la prévision des crues à travers le Service central d'hydrométéorologie et d'appui à la prévision des inondations (SCHAPI), créé en 2003 pour faire évoluer le réseau national d'annonce de crues mis en place dès 1984 (Tanguy *et al.*, 2005). Le SCHAPI coordonne les 19 services de prévision des crues (SPC), dont le territoire a été défini en fonction des grands bassins hydrographiques de France métropolitaine, et cinq cellules de veille hydrologique (CVH) dans les départements d'Outre-mer. Le SCHAPI travaille en collaboration avec Météo-France, en charge de la prévision météorologique en temps réel. En 2006, le dispositif Vigicrues⁴ (Berthet *et al.*, 2019 ; Marty *et al.*, 2019 ; Viatgé *et al.*, 2019a) est mis en place pour regrouper les informations à destination de la population sur la

2. public.wmo.int/en/media/press-release/early-warnings-all-initiative-gains-momentum

3. ecologie.gouv.fr/prevention-des-risques-naturels

4. vigicrues.gouv.fr

situation en temps réel des cours d'eau. En 2017, le service national Vigicrues Flash (Javelle *et al.*, 2014, 2019 ; Piotte *et al.*, 2020) est créé pour surveiller plus de 30 000 km de cours d'eau non couverts par Vigicrues. Des outils de modélisation sont utilisés pour émettre des bulletins de vigilance sur le risque d'inondations générées par débordement des cours d'eau⁵. Selon les prévisions faites par les modèles hydrologiques et hydrauliques utilisés, les cours d'eau surveillés par Vigicrues sont classés en vigilance verte (pas de danger particulier), jaune (risque de crue génératrice de débordements), orange (risque de crue génératrice de débordements importants) ou rouge (risque de crue majeure).

Le SCHAPI met à disposition des SPC plusieurs outils de modélisation hydrologique pour la prévision en temps réel. La plateforme PLATHYNES regroupe plusieurs modèles événementiels spatialisés comme le modèle MARINE (Modélisation de l'Anticipation du Ruissellement et des Inondations pour des évènements Extrêmes ; Roux *et al.*, 2011), un modèle distribué fonctionnant au pas de temps variable. Ce modèle est utilisé en particulier pour la prévision des crues éclair. La plateforme de modélisation SMASH (Jay-Allemand *et al.*, 2020) est utilisée au sein du système Vigicrues Flash pour prévoir le débit des petits cours d'eau non jaugés qui sont soumis à des crues rapides. Le modèle de prévision des crues GRP (Viatgé *et al.*, 2019a) est utilisé en France par la plupart des SPC du réseau Vigicrues depuis plus de dix ans. Il découle des travaux de Tangara (2005) et Berthet (2010), et a été appliqué sur de nombreux bassins versants français. Sa structure dérive de celle du modèle GR4J (Perrin *et al.*, 2003) et avait été simplifiée pour permettre une valorisation efficace du débit observé. GRP est un modèle déterministe parcimonieux à trois paramètres libres qui fonctionne du pas de temps infra-horaire au pas de temps journalier. Il intègre une procédure d'assimilation des derniers débits observés en temps réel pour mettre à jour ses états initiaux. Un calcul utilisant l'erreur antérieure est également effectué pour mettre à jour les prévisions futures. Le modèle peut prévoir les débits pour des échéances allant de l'infra-horaire à cinq jours.

Démarches d'amélioration des modèles hydrologiques de prévision

Les retours d'expérience des groupes utilisateurs ont permis d'identifier plusieurs limitations du modèle GRP, en particulier une tendance à la sous-estimation des volumes et pics de crue ainsi qu'un retard à la montée. Par exemple, en juin 2016, la crue exceptionnelle de la Seine, de la Loire et de leurs affluents a été sous-estimée par le modèle GRP (e.g. Peredo, 2021). Plusieurs aspects de la modélisation hydrologique peuvent être à l'origine des difficultés du modèle, parmi lesquels l'estimation des paramètres du modèle (prise en compte des erreurs temporelles, nombre de paramètres, dépendance au pas de temps de fonctionnement), sa structure (estimation des pluies efficaces, prise en compte de l'intensité et de la variabilité spatiale de la pluie), ou encore la méthode d'assimilation de données employée.

Récemment, plusieurs travaux de recherche ont contribué à identifier des pistes d'amélioration. Les travaux de Ficchi (2017) ont notamment permis de développer une structure multi pas de temps capable d'exploiter des données infra-horaires de pluie et de débit, ce qui est particulièrement utile pour prévoir le débit de petits bassins versants très réactifs. Le projet PICS (prévision immédiate des impacts des crues soudaines, 2017-2022, pics.ifsttar.fr)

5. La prévision des inondations par ruissellement urbain, par remontées de nappes ou par submersions marines est assurée par d'autres services de l'État.

a été initié pour améliorer les connaissances et les outils sur la prévision des crues soudaines. Les travaux de Peredo (2021) ont notamment montré l'intérêt d'une modélisation distribuée adaptée pour prévoir les crues du début de l'automne sur le bassin de l'Aude. Les recherches autour du service Vigicrues Flash ont permis de proposer un outil de prévision du débit des petits cours d'eau non jaugés. Des travaux de recherche sont en cours pour améliorer la régionalisation (estimation des paramètres pour des points non jaugés) des paramètres des modèles de la plateforme SMASH (Jay-Allemand *et al.*, 2020 ; Jay-Allemand, 2020).

Positionnement scientifique et questions de recherche

Les travaux de cette thèse s'intéressent à trois grandes questions de recherche que nous resituons par rapport à l'état de l'art.

Modèles efficaces et explicatifs ?

Une des difficultés centrales en modélisation hydrologique est de lier les recherches sur la physique des bassins versants avec les modèles utilisés pour reproduire les débits à l'exutoire de ces derniers (e.g. Hrachowitz & Clark, 2017). Certains modèles empiriques ont des performances très élevées pour simuler les débits, mais ils reposent sur des équations qui sont difficilement interprétables à petite échelle (i.e. à des échelles inférieures à l'échelle du bassin versant). D'autres modèles basés sur des lois d'écoulement de l'eau dans le sol bien décrites à petite échelle ont des performances plus réduites pour reproduire le débit à l'exutoire des bassins versants. Cette problématique est parfois définie comme la capacité de certains modèles à être performants pour de « mauvaises raisons » (e.g. Kirchner, 2006). De nombreux modélisateurs ont donc essayé de trouver une meilleure cohérence entre le fonctionnement des modèles hydrologiques en matière de conceptualisation (i.e. états et variables internes, paramètres) et notre connaissance des processus hydrologiques qui impactent les dynamiques des bassins versants (e.g. Braud *et al.*, 2014 ; Wrede *et al.*, 2015 ; Bouaziz *et al.*, 2021).

Il reste cependant difficile d'établir des liens entre notre connaissance de la physique des versants et la modélisation de la dynamique affectant les débits à l'échelle du bassin versant (e.g. Beven *et al.*, 2019) car ces processus impliquent des échelles spatio-temporelles très différentes (e.g. Blöschl *et al.*, 2019) et difficilement transposables dans les modèles (e.g. Wagener *et al.*, 2010). En d'autres termes, il est souvent difficile d'identifier l'échelle (voire les échelles) à laquelle un processus hydrologique affecte le débit, ce qui signifie qu'il est presque impossible de dériver des lois générales sur les processus hydrologiques affectant le débit en rivière à l'échelle du bassin versant. On parle ainsi de propriétés émergentes : le système à une certaine échelle ne se comporte pas comme une simple addition des propriétés à une échelle inférieure. En ce sens, McDonnell *et al.* (2021) proposent d'unifier la connaissance et la description des processus hydrologiques à toutes les échelles en considérant des échelles d'intérêt (qu'ils appellent « échelle de celui qui regarde »). Un processus hydrologique peut alors être décrit comme un réservoir qui se vide et se remplit à une échelle spatio-temporelle donnée. Chaque processus est imbriqué dans un processus qui agit à une échelle plus large. Par exemple, le ruissellement de subsurface du versant d'une colline peut être décrit comme un réservoir qui se remplit avec l'infiltration de la pluie dans le sol (à l'échelle du versant) jusqu'à être connecté à la rivière (échelle du bassin versant) à partir d'un certain niveau de

saturation qui permet à l'eau de s'écouler à travers les zones saturées jusqu'à la rivière (e.g. van Meerveld & McDonnell, 2006).

Klemeš (2000) soutient qu'un modèle prédictif n'a pas nécessairement vocation à devenir un modèle explicatif car il n'a pas été développé pour tester la validité d'hypothèses théoriques (ou physiques), mais plutôt pour reproduire une variable d'intérêt (voir aussi Shmueli, 2010, pour des explications sur la différence entre explicabilité et prédictibilité des modèles). Les modèles empiriques (qui ont souvent été construits pour être prédictifs) n'ont pas besoin d'être explicatifs pour être utiles dans de nombreux cas pratiques (ici la prévision des crues) mais ils doivent être évalués en dehors de la donnée utilisée pour les construire (i.e. en dehors de la période d'estimation des paramètres). Un modèle explicatif peut quant à lui devenir prédictif à condition qu'il soit capable de reproduire les tendances des données observées (si la donnée que le modèle cherche à reproduire n'est pas utilisée lors de la construction du modèle, il n'a, en théorie, pas besoin d'être évalué sur une période de validation). Nous posons donc la question suivante : ***Est-il possible et souhaitable d'améliorer l'efficacité et la généralité d'un modèle pluie-débit tout en le rendant plus cohérent avec les processus hydrologiques affectant la génération du ruissellement ?*** Nous pourrions ainsi chercher à améliorer l'efficacité du modèle sans en améliorer le pouvoir explicatif, mais nous pourrions aussi tester des hypothèses sur les processus de genèse des crues aux échelles d'intérêt et évaluer leur pouvoir prédictif. Pour vérifier nos hypothèses, nous appliquerons systématiquement un « garde-fou » qui consiste à évaluer la performance de nos hypothèses de modélisation en dehors des périodes d'estimation des paramètres.

Effets indésirables d'une complexification : a-t-on atteint les limites de l'information exploitable à l'échelle du bassin versant ?

La recherche d'une structure générale et efficace pour un grand nombre de bassins versants est le moteur de nombreux développements de modèles conceptuels s'appuyant sur de larges bases de données (Andréassian *et al.*, 2009 ; Gupta *et al.*, 2014), le calage des paramètres jouant le rôle de différenciation entre les bassins versants. Dans certains cas, des jeux de paramètres différents donnent des performances similaires sur un même bassin. Ce problème est souvent défini au moyen de la notion d'équifinalité (Beven, 2006 ; Clark *et al.*, 2016) et est en partie dû au fait que le calage des paramètres compense le manque d'information sur les dynamiques des bassins versants (Andréassian *et al.*, 2012), mais est aussi dû au fait que certains modèles sont intrinsèquement mal formulés mathématiquement (e.g. Kavetski & Clark, 2010). Plus un modèle a un nombre élevé de paramètres libres, plus le risque d'équifinalité est élevé. On peut interpréter l'équifinalité comme la manifestation d'une trop grande liberté accordée à un modèle au regard de l'information disponible. La complexification de la structure des modèles hydrologiques ne serait pas souhaitable sans apport d'information nouvelle.

Le calage multi-objectif permet de formaliser ce problème en explorant les compromis et la sensibilité des paramètres à différents critères de performances ou signatures permettant de décrire plusieurs propriétés des bassins versants. Ces critères peuvent être calculés à partir des chroniques de débit mais aussi à partir de données additionnelles comme les données satellites d'humidité des sols (e.g. Nijzink *et al.*, 2018 ; Li *et al.*, 2018 ; Dembélé *et al.*, 2020) ou encore les données piézométriques (e.g. Pelletier & Andréassian, 2022). L'amélioration du calage des modèles en utilisant uniquement des données de débit peut en effet limiter

la possibilité de réduire l'équifinalité. Une meilleure estimation des paramètres ne permet toutefois pas de compenser les incertitudes liées aux simplifications faites par les modèles.

Parmi les développements réalisés ces dernières années en modélisation hydrologique, des approches fondées sur des méthodes d'apprentissage statistique (ou modèles "boîte noire") ont tenté de dériver des structures à partir de quantités importantes de données, incluant par exemple de nombreux descripteurs de bassins versants (e.g. Kratzert *et al.*, 2018, 2019), dans l'apprentissage des modèles. Une des idées derrière ces approches est qu'il reste de l'information exploitable pour améliorer la simulation des débits en rivière. Cette meilleure exploitation de l'information passerait donc par une adaptation voire une complexification des outils de modélisation, ce qui s'oppose aux affirmations énoncées précédemment. Une suite à ce raisonnement serait de chercher les facteurs de premier ordre qui ont d'une part un impact sur les processus qui affectent la transformation de la pluie en débit à l'exutoire, et d'autre part qui peuvent expliquer les erreurs des modèles (e.g. Gupta *et al.*, 2008). Ici, on revient vers la problématique de l'échelle considérée, mais aussi du contenu en information des données (comme théorisé par Shannon, 1948) pour dériver des liens de cause à effet (pour une démarche explicative) ou des liens prédictifs entre les variables à l'échelle d'intérêt. En résumé, on serait dans une situation où le problème scientifique ne serait pas sous-informé (au sens de Shannon) mais mal informé. Il faudrait ainsi rechercher les facteurs de premier ordre dans la réponse des bassins versants selon le niveau d'information disponible à l'échelle considérée.

La question suivante résume les discussions précédentes pour nos travaux de recherche : ***Pouvons-nous encore améliorer la structure et l'estimation des paramètres d'un modèle de prévision des crues avec l'information disponible en prévision opérationnelle à l'échelle des bassins versants sans augmenter l'équifinalité sous-jacente ?*** Dans le cadre de nos travaux, l'information disponible en prévision opérationnelle à l'échelle des bassins versants correspond aux chroniques de températures, de précipitations et de débits, ainsi qu'aux données hydromorphologiques (topographie, taille...).

Interactions entre les choix de structure, de calage et d'assimilation : quels compromis ?

La plupart des modèles de prévision des crues intègrent des procédures d'assimilation de données dans leur fonctionnement. Ces procédures leur permettent de mieux simuler les crues, en « rattrapant » différentes incertitudes liées au calage des paramètres, à la structure et aux données (e.g. Refsgaard, 1997 ; Jain *et al.*, 2018). L'assimilation de données consiste à utiliser la (ou les) dernière(s) donnée(s) observée(s) pour mettre à jour les états initiaux ou les paramètres du modèle avant l'instant de prévision (i.e. l'instant où on émet une prévision pour une échéance donnée), ou encore pour corriger les débits prévus après cet instant. Les modèles pluie-débit sont ainsi devenus des modèles pluie-débit→débit du fait de l'importance de l'assimilation du dernier débit observé dans la performance des prévisions (Berthet, 2010). Une des difficultés des schémas d'assimilation de données réside dans la quantification des incertitudes des données en temps réel (Liu *et al.*, 2012), afin de ne pas dégrader le signal initial du modèle par du bruit et ainsi ne pas détériorer les prévisions selon l'échéance. Nous avons vu précédemment qu'il existe des interactions fortes entre la complexification d'une structure et l'estimation de ses paramètres. En ajoutant la complexité associée à l'assimi-

lation de données, on ajoute un degré de complexification au modèle. Des choix s'imposent alors sur le compromis optimal entre la complexité des trois piliers du modèle de prévision : la structure, l'estimation des paramètres et l'assimilation de données. L'équilibre choisi peut dépendre de l'échéance de prévision car l'erreur du modèle augmente avec celle-ci du fait de l'atténuation progressive de l'effet de l'assimilation de données (e.g. Li *et al.*, 2009 ; Piazzini *et al.*, 2021). Les choix de modélisation effectués pour développer le modèle GRP ont donné lieu à une simplification de la structure du modèle, et par extension une simplification de l'estimation de ses paramètres. À méthode d'assimilation similaire, Berthet (2010) montre qu'une structure plus complexe n'améliore pas la qualité des prévisions. Par ailleurs, la simplification de la structure du modèle permet une meilleure valorisation du débit observé. Nous nous posons ici la question suivante : ***Quel est le compromis optimal entre la complexification de la structure, de la méthode d'estimation des paramètres et de la procédure d'assimilation de données selon l'échéance de prévision ?***

Objectif de la thèse et choix méthodologiques

La thèse a pour objectif de mieux comprendre, par une approche de diagnostic détaillé, les raisons des limites du modèle GRP et d'améliorer la qualité de ses prévisions, en cherchant le degré de complexification le mieux adapté au niveau d'information disponible à l'échelle des bassins versants.

Pour répondre à cet objectif, nous devons mettre en place un cadre de tests et d'évaluations spécifique. Les questions de recherche que nous avons formulées précédemment amènent à réfléchir au plan d'expérience adéquat pour identifier les limites des modèles pour une prévision générale et efficace des crues. Dans cette thèse, nous nous plaçons dans un cadre d'évaluation qui s'appuie sur un grand échantillon de bassins versants afin d'intégrer des conditions hydroclimatiques variées. Nous utiliserons des modèles continus⁶ (comme le modèle GRP), mais que nous analyserons à partir d'une large sélection d'événements de crue représentant des réponses hydrologiques très différentes. Plusieurs critères d'erreurs seront utilisés ainsi que plusieurs indicateurs sur les caractéristiques hydroclimatiques des crues et des bassins versants. L'évaluation des modèles dans un cadre si général peut poser des problèmes d'identification des liens de cause à effet entre les caractéristiques d'un modèle et ses erreurs car le nombre de filtres d'analyse est élevé. De plus, comme nous l'avons mentionné dans la troisième question de recherche, les modèles de prévision des crues reposent sur trois aspects principaux : la structure, l'estimation des paramètres et l'assimilation de données. Du fait de l'imbrication de ces différents aspects dans l'exercice de prévision, nous devons simplifier le problème pour cibler les améliorations nécessaires et comprendre leur effet sur la qualité des prévisions du modèle. Une méthodologie permettant d'identifier un à un les problèmes associés aux différentes caractéristiques d'un modèle de prévision est développée tout au long de nos travaux.

6. Un modèle continu est un modèle dont les simulations sont effectuées tout au long de l'année, c'est-à-dire sur un maximum de pas de temps avant les simulations d'intérêt, afin d'obtenir une bonne estimation des conditions initiales d'humidité des bassins versants. Il s'oppose aux modèles événementiels dont les simulations sont effectuées seulement pour l'événement de crue modélisé.

Les travaux de recherche antérieurs sur le modèle GRP ont été conduits en considérant les problèmes d'équifinalité, de contenu en information des données, d'échelle d'intérêt et d'efficacité dans un cadre de prévision des crues. Le compromis le plus efficace au regard du cadre d'évaluation employé a été choisi pour développer GRP. Il consiste à mettre un poids très important sur le débit observé à l'instant de prévision. La structure qui en découle est une structure simple qui repose sur trois paramètres libres afin de limiter les problèmes d'équifinalité et augmenter la robustesse du modèle. Les procédures de mise à jour du modèle GRP ont un impact très important sur la qualité de ses prévisions. Dans les travaux de cette thèse, on cherche à explorer d'autres compromis que ceux qui ont été effectués dans GRP. En particulier, nous prenons le parti que l'amélioration des performances d'un modèle hydrologique en simulation conduit à une amélioration de la prévision des crues, notamment lorsque l'assimilation de données ne suffit pas à compenser des limites structurelles.

Organisation du mémoire

L'organisation de ce mémoire dérive de cette dernière hypothèse. Pour améliorer la structure du modèle, nous séparons les travaux effectués avec les procédures de mise à jour des travaux effectués sans ces procédures, du fait de l'importance du dernier débit observé dans les performances du modèle de prévision. Le manuscrit comporte neuf chapitres répartis en trois parties. Les deux premières parties présentent des travaux effectués dans un cadre de simulation des crues (i.e. sans procédure d'assimilation de données). Dans ces parties, nous utilisons le modèle GR5H (Le Moine, 2008 ; Ficchi *et al.*, 2019) qui est la version la plus performante et la plus polyvalente des modèles horaires de la famille GR (e.g. Coron *et al.*, 2017). Nous testons la validité de notre hypothèse initiale dans la dernière partie de ce manuscrit, où les performances de deux versions du modèle GR5H sont comparées avec celles du modèle de référence (GRP) dans un cadre de prévision des crues (i.e. avec des procédures d'assimilation de données).

Le premier chapitre présente la base de données et les éléments méthodologiques généraux de nos travaux. Les travaux de la première partie ont pour but d'étudier les erreurs du modèle de simulation GR5H et d'en proposer des explications. Cette partie est composée de trois chapitres. Le chapitre 2 vise à présenter une analyse détaillée des erreurs du modèle en crue en fonction de plusieurs caractéristiques de ces crues, en particulier la saison d'occurrence du pic de crue. Dans le chapitre 3, nous étudions comment les caractéristiques des événements de pluie associés aux crues peuvent être liées aux erreurs du modèle. Dans le dernier chapitre de cette partie (chapitre 4), nous déterminons si des états et des flux internes du modèle peuvent être à l'origine des déficiences observées.

Plusieurs voies d'amélioration du modèle de simulation sont identifiées à partir du diagnostic de la première partie. Elles constituent les travaux de la deuxième partie. Dans le chapitre 5, plusieurs hypothèses de modélisation sont testées pour améliorer les performances en crue, en particulier lorsque les bassins réagissent à des intensités de pluie élevées. Au vu de l'importance des caractéristiques spatiales de la pluie dans les erreurs du modèle, nous évaluons ensuite l'intérêt d'une distribution spatiale du modèle (chapitre 6). Le chapitre 7 présente un diagnostic de l'estimation des paramètres du modèle GR5H et de la nouvelle

version proposée, cela afin de proposer des améliorations orientées vers la simulation des crues.

Dans la dernière partie de ce mémoire, nous construisons deux modèles de prévision à partir du modèle GR5H et du modèle proposé dans la partie précédente (i.e. une stratégie de mise à jour des états est définie). Dans le chapitre 8, ces modèles sont comparés au modèle GRP dans un cadre d'évaluation spécifique à la prévision des crues, auquel nous ajoutons le cadre d'évaluation développé dans les chapitres précédents. Dans le chapitre 9, plusieurs méthodes d'estimation des paramètres, dont certaines ont été développées dans le chapitre 7, sont testées dans le but d'évaluer les gains potentiels d'une méthode plus complexe.

Pour faciliter la lecture, le contexte bibliographique et les éléments méthodologiques spécifiques à chaque chapitre sont introduits au début des chapitres. Les chapitres 2 et 5 correspondent à des articles publiés dans des revues à comité de lecture.

Chapitre 1

Base de données et éléments méthodologiques

Pour obtenir des résultats les plus généraux possible, plusieurs auteurs, dont Andréassian *et al.* (2009) et Gupta *et al.* (2014), insistent sur l'importance d'utiliser une large base de données contenant des bassins versants avec des conditions hydro-climatiques variées. Ainsi, les résultats obtenus sont moins dépendants des caractéristiques de quelques bassins et de la présence de biais dans les données. Ces conditions sont nécessaires pour identifier les limites d'un modèle hydrologique ainsi que pour développer des hypothèses de modélisation généralisables pour la prévision des crues en France.

1.1 Sélection des bassins versants

La base de donnée de Ficchi (2017), constituée de 240 bassins versants répartis sur la France métropolitaine, sera utilisée pour nos analyses au pas de temps horaire. Cette base de données a été utilisée et développée dans le cadre de travaux sur la structure des modèles GR pour la simulation des crues. De plus, elle contient des séries temporelles du pas de temps 6 minutes au pas de temps journalier, ce qui permettra – dans des travaux futurs – d'étendre les résultats à des pas de temps plus fins. La description qui suit est un résumé de la sélection des bassins faite par Ficchi (2017). Les modifications apportées à cette base de données sont décrites dans la section suivante.

L'échantillon a été constitué en fonction de la variabilité hydro-climatique des bassins, de la disponibilité et de la qualité des données à pas de temps fin, en partant d'une base de données initiale de 3 701 bassins élaborée par l'équipe HYDRO (Delaigne *et al.*, 2020) à partir des données fournies par les services de l'État (Banque Hydro ; Leleu *et al.*, 2014). Les données de pluie et de débit sont disponibles pour des échelles de temps allant du pas de temps six minutes au pas de temps journalier entre 2005 et 2013 (disponibilité des pluviomètres au pas de temps six minutes plus élevée sur cette période). Les chroniques de pluie ont été construites à partir de la réanalyse SAFRAN au pas de temps journalier (Vidal *et al.*, 2010) désagrégée temporellement grâce à une pondération fournie par les polygones de Thiessen des pluviomètres Météo-France au pas de temps six minutes. La réanalyse COMEPHORE (au pas de temps horaire ; Tabary *et al.*, 2012) a été utilisée pour obtenir des données de pluie sur la période d'initialisation avant 2005 et pour l'évaluation de la qualité des données au pas de temps six minutes. Les chroniques de pluie COMEPHORE sont le résultats d'une

fusion entre des données radar et des pluviomètres au sol. Les chroniques de débits aux différents pas de temps ont été constituées à partir de la reconstitution des temps de référence des stations et des interpolations pour les lacunes ne dépassant pas 15 jours d'affilée. Les chroniques d'évapotranspiration potentielle pour les pas de temps infra-journaliers et infra-horaires ont été calculées à partir des données SAFRAN Météo-France disponibles au pas de temps journalier. Les calculs ont d'abord été effectués au pas de temps journalier à partir de la formule d'Oudin *et al.* (2005). Ces chroniques journalières ont ensuite été désagrégées au pas de temps horaire en suivant une distribution parabolique entre 6 h et 19 h (UTC) utilisée par Lobligeois (2014). Deux sous-périodes indépendantes P1 (01/08/2005 – 31/07/2009) et P2 (01/08/2009 – 31/07/2013) ont été définies. Pour obtenir l'échantillon de 240 bassins versants, la sélection a été faite de la manière suivante :

- Des 3701 bassins initiaux, nous sommes passés à 1961 bassins, à partir de critères sur les caractéristiques des bassins : la superficie des bassins est comprise entre 5 km² et 10000 km² ; les bassins sont très peu ou pas influencés (gestion de barrages et autres influences humaines) ; les précipitations neigeuses sont inférieures à 10 % ; le nom des stations ne contient pas de mots tels que « source », « résurgence », « canal » ou encore « étang ».
- 584 bassins ont ensuite été sélectionnés à partir de la qualité, disponibilité et densité des données pluviométriques. Les données des pluviomètres six minutes présentent quelques lacunes. Seules les stations pour lesquelles 90 % des jours de l'année présentent moins de 5 % de lacunes et cinq années de données avec moins de 10 % de lacunes ont été conservées. Plusieurs conditions sont posées concernant la densité des pluviomètres : le nombre de pluviomètres influençant un bassin respecte l'équation de Schaake *et al.* (2000) lorsque le bassin a une taille supérieure à 200 km² (voir équation 2.1 de Ficchi, 2017). Le minimum est de trois stations d'influence pour les bassins de moins de 200 km² et deux pour les bassins de moins de 25 km². Les bassins de plus de 200 km² doivent contenir au moins un pluviomètre en leur sein. Une mesure supplémentaire de densité des pluviomètres d'influence (voir équation 2.3 et seuils empiriques de l'équation 2.4 de Ficchi, 2017) est utilisée pour disqualifier les faibles densités pour les petits bassins dont la plupart présentent déjà très peu de pluviomètres dans leur espace géographique. Ce critère est moins sélectif pour les grands bassins qui ont souvent une densité moins importante de pluviomètres d'influence mais qui contiennent au moins un pluviomètre à l'intérieur de leur espace géographique.
- 261 bassins ont ensuite été écartés car aucune donnée de débits au pas de temps variable n'était disponible. 56 bassins ne présentent pas de données couvrant la période 2005-2013. Seuls les bassins présentant moins de 35 % d'absence de données de débits sur chaque sous-période ont été conservés. Cette sélection sur les données de débits réduit l'échantillon à 223 bassins versants.
- 17 bassins ne respectant pas les critères mentionnés précédemment ont été ajoutés pour assurer une bonne couverture du territoire : trois bassins inférieurs à 5 km², un bassin expérimental de 46 km² localisé en Rhône-Alpes, neuf bassins du sud et du nord de la France (sous-représentés dans l'échantillon) et cinq bassins localisés dans le Sud-Ouest.

L'échantillon final contient des bassins répartis sur toute la France métropolitaine et présente une variabilité hydro-climatique et morphologique importante. La liste des 240 bassins est disponible dans les annexes de Ficchi (2017).

1.2 Modifications de la base de données

Pour les travaux de cette thèse, 11 bassins versants ont été retirés de l'échantillon :

- Huit bassins présentant des données douteuses (inspection visuelle) et/ou des lacunes trop importantes (codes : K2773120, K2783010, K5554580, O3335010, O3395010, O3575510, Q0522530, Q1094010, U1120010).
- Deux bassins ne respectant pas les critères initiaux de sélection (codes : Q7002910 et Q0120060). La fraction de précipitations solides est supérieure à 25 % sur ces deux bassins. Nous n'utiliserons pas de module de prise en compte de l'accumulation et de la fonte de la neige dans nos travaux.

La base de données utilisée dans nos travaux contient 229 bassins versants (fig. 1.1). 219 bassins versants ont un régime hydrologique pluvial (selon la définition de Sauquet *et al.*, 2008), ce qui signifie que les périodes de hautes eaux ont majoritairement lieu en hiver et les périodes d'étiage ont majoritairement lieu en été (fig. 1.2).

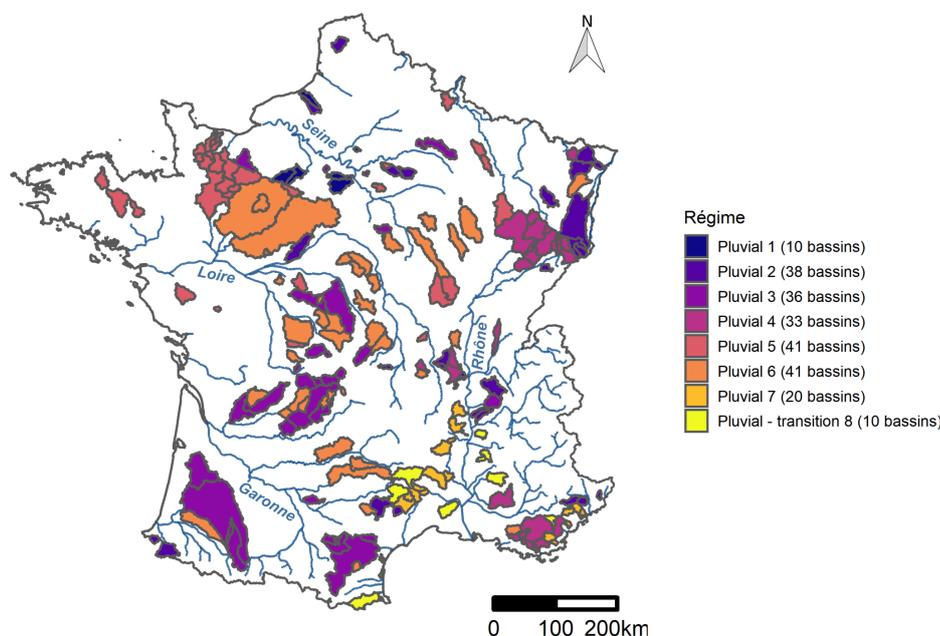


FIGURE 1.1 – Localisation des 229 bassins versants en France métropolitaine. Classification en régimes hydrologiques selon la définition de Sauquet *et al.* (2008).

Dix bassins ont un régime de transition, ce qui signifie que les périodes de hautes eaux ont lieu en hiver et au printemps sous l'influence de la pluie et de la fonte de neige. Les résultats présentés sur les figures 1.1 et 1.2 diffèrent des résultats obtenus par Ficchi (2017) car nous avons refait la classification à partir des modifications apportées à la base de données (voir explications plus loin dans cette section). Les distributions des caractéristiques hydro-climatiques des 229 bassins versants sont présentées dans le tableau 1.1. La figure 1.3 montre qu'une dizaine de bassins ont un bilan hydrique annuel qui témoigne de pertes d'eau (points en bas à droite de la figure). Deux bassins karstiques ont des valeurs de coefficient d'écoulement supérieures à 1 : le bassin de la Laine à Soulaines-Dhuy et le bassin de la Siagne à Callian.

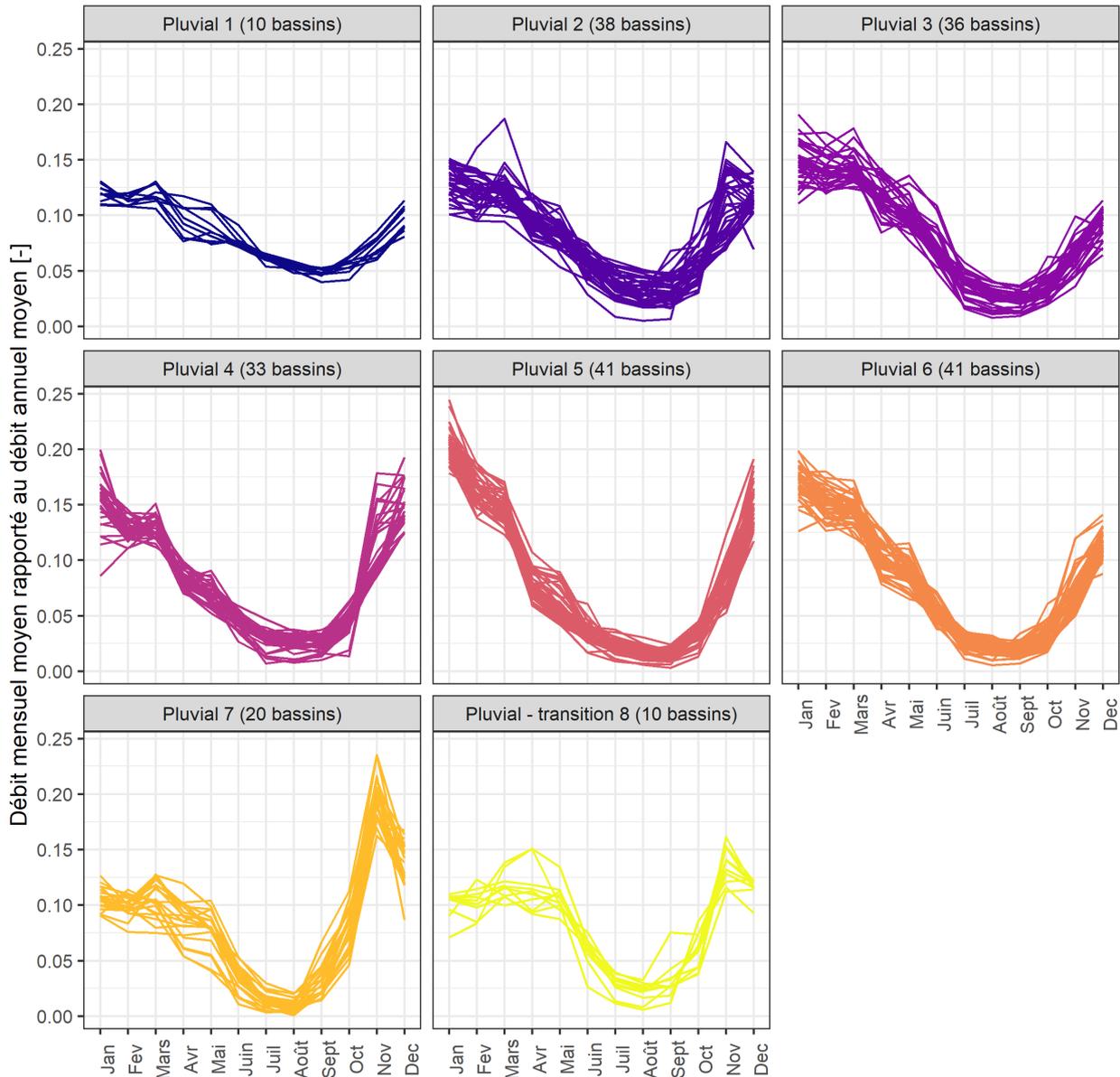


FIGURE 1.2 – Classification des 229 bassins versants selon leur régime hydrologique à partir de la définition de Sauquet *et al.* (2008). Les courbes de régimes ont été établies sur la période 2000-2018.

Tous les travaux présentés dans ce mémoire ont été réalisés avec des données au pas de temps horaire. Les séries temporelles de précipitations que nous utilisons ne proviennent pas de l'interpolation effectuée par Ficchi (2017). Elles ont été extraites et interpolées à l'échelle des bassins versants à partir du produit climatique COMEPHORE de résolution 1 km² (Tabary *et al.*, 2012). À la suite du comité de suivi de thèse de première année, nous avons décidé d'étendre la période temporelle entre le 1^{er} janvier 2000 et le 31 décembre 2018 afin d'inclure plus d'événements de crue dans nos analyses. Les résultats présentés dans le chapitre 2 ont été obtenus sur la période 01/08/2005 - 31/07/2013, initialement définie par d'Andrea Ficchi (extraction SAFRAN : 11/2013 ; extraction Banque Hydro : 12/2013 ; Ficchi, 2017). Les résultats des chapitres suivants ont été obtenus avec la période étendue (extraction SAFRAN : 01/2020 ; extraction Banque Hydro : 05/2020). Nous avons

défini deux sous-périodes indépendantes P1 (01/01/2000 – 30/06/2009) et P2 (01/07/2009 – 31/12/2018) qui incluent les sous-périodes définies par Ficchi (2017). Certains bassins présentent des absences de données de débit. La base de données finale et les sous-périodes définies contiennent au moins autant de données de débit que la base de données de Ficchi (2017).

TABLE 1.1 – Distribution de neuf caractéristiques hydro-climatiques et cinq caractéristiques morphologiques des 229 bassins versants. Les distributions proviennent de Ficchi (2017) et Delaigue *et al.* (2020). P99 correspond au 99^e centile de la pluie.

Caractéristique	Définition/référence	Min	Q25	Med	Moy	Q75	Max
Superficie [km ²]	-	3.54	164.1	354.1	680.3	772.3	7 918
Altitude moyenne [m]	-	70	198	358	383	519	1 060
Pente moyenne [-]	-	0.01	0.03	0.05	0.06	0.08	0.24
Indice topographique [-]	Ducharne (2009)	8.19	12.13	13.26	13.13	14.25	17.33
Densité de drainage [km ²]	Le Moine (2008)	0.06	0.3	0.51	1.19	1.08	19.8
Module (Qm) [mm/an]	-	35	262	349	437	524	1 398
Précipitations annuelles moyennes (Pm) [mm/an]	-	651	818	937	990	1 097	2 108
ETP annuelle moyenne (ETPm) [mm/an]	-	594	664	708	739	767	1 129
Température annuelle moyenne [°C]	-	8.2	9.8	10.4	10.6	11.1	14.3
Autocorrélation des débits à 24 h [-]	-	0.39	0.77	0.84	0.83	0.91	1.00
Coefficient d'écoulement [-]	Qm/Pm	0.04	0.27	0.35	0.38	0.46	1.71
Indice d'aridité [-]	Pm/ETPm	0.66	1.10	1.32	1.37	1.55	3.47
Indice d'intensité des précip. journalières [-]	P99/Pm	7.58	8.57	9.19	10.56	11.53	19.89
Indice de débit de base [-]	Gustard <i>et al.</i> (1992)	0.19	0.46	0.57	0.56	0.65	0.95
Décalage temporel pluie-débit [h]	Ficchi (2017)	1	8	15	21	26	117

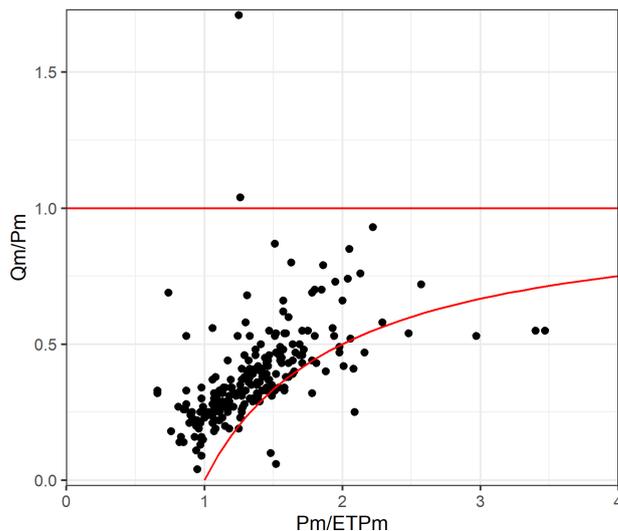


FIGURE 1.3 – Positionnement des 229 bassins versants sur le graphique adimensionnel de Turc-Budyko.

1.3 Sélection des événements de crues

Une grande partie des analyses présentées dans ce mémoire repose sur la sélection d'événements de crues. Cette sélection nous permettra d'évaluer spécifiquement la capacité des modèles à simuler différents types d'événements de crue. Cela nous permettra aussi de lier les hypothèses de modélisation que nous allons formuler à des caractéristiques de crue qui témoignent de réactions très différentes des bassins versants à la pluie. Il existe de nombreuses manières de sélectionner des événements (e.g. Merz *et al.*, 2006 ; Tarasova *et al.*, 2018). Ici, un algorithme similaire à celui employé par Lobligeois (2014) et Ficchi (2017) a été appliqué. Aucune séparation du débit de base n'a été effectuée au préalable. Les étapes de cet algorithme sont les suivantes :

1. Le débit maximal de la période est sélectionné.
2. On vérifie que ce débit maximal n'est pas situé en bordure de chronique (trois jours de part et d'autre).
3. Le début de l'événement est recherché sur une période de 10 jours avant l'instant du débit maximal sélectionné.
4. On vérifie qu'un autre événement n'a pas été sélectionné sur ces 10 jours.
5. Le début de l'événement est l'instant, situé entre 10 jours avant le débit maximal et l'instant du débit maximal, à partir duquel le débit dépasse 20 % du débit maximal sélectionné.
6. La fin de l'événement est recherchée sur une période de 10 jours après l'instant du débit maximal sélectionné.
7. On vérifie qu'un autre événement n'a pas été sélectionné sur ces 10 jours.
8. La fin de l'événement est le moment, situé entre l'instant du débit maximal et 10 jours après, à partir duquel le débit est inférieur à 25 % du débit maximal sélectionné.
9. On vérifie qu'il y a moins de 10 % de données manquantes sur la période sélectionnée.
10. Si le débit maximal de l'événement est inférieur au quantile 95 % du bassin, on arrête la sélection.

11. Si l'événement respecte les conditions précédentes, on le retire de la chronique et on cherche le débit maximal de la nouvelle chronique en répétant les étapes précédentes.

Au maximum, cinquante événements ont été sélectionnés par bassin, cela afin de prendre en compte des événements moins intenses (comme suggéré par Tarasova *et al.*, 2020) tout en gardant un nombre raisonnable pour permettre d'effectuer une inspection visuelle de chaque hydrogramme. Le seuil défini pour trouver la fin de l'événement a été choisi plus haut que celui du début de l'événement pour ne pas sélectionner trop de pas de temps appartenant à la récession de l'hydrogramme par rapport au nombre de points de montée.

L'événement pluviométrique associé à chaque événement de débit a été sélectionné en décalant la période de l'événement de débit d'un temps égal au temps de concentration du bassin en question. Ce temps de concentration a été calculé comme le décalage temporel qui maximise la corrélation croisée entre les chroniques de pluie et de débit, comme formulé par Ficchi (2017). Un ajustement automatique de la fenêtre temporelle de l'événement de pluie a ensuite été effectué :

1. Le début de l'événement de pluie est recherché sur une période qui précède le début préalablement identifié. Le début de l'événement devient l'instant qui suit le dernier instant à partir duquel la pluie est nulle.
2. La fin de l'événement de pluie est recherchée sur une période qui suit la fin préalablement identifiée. La fin de l'événement devient l'instant qui précède l'instant à partir duquel la pluie est nulle.

La sélection automatisée d'événement étant un exercice imparfait, nous avons ensuite inspecté visuellement chaque hydrogramme et chaque hyétogramme et modifié manuellement la sélection quand cela a été jugé nécessaire (voir figure 1.4 pour un exemple de visualisation). Cela concerne une centaine d'événements. L'échantillon final contient 10 652 événements de crues (2 990 pour les travaux présentés dans le chapitre 2).

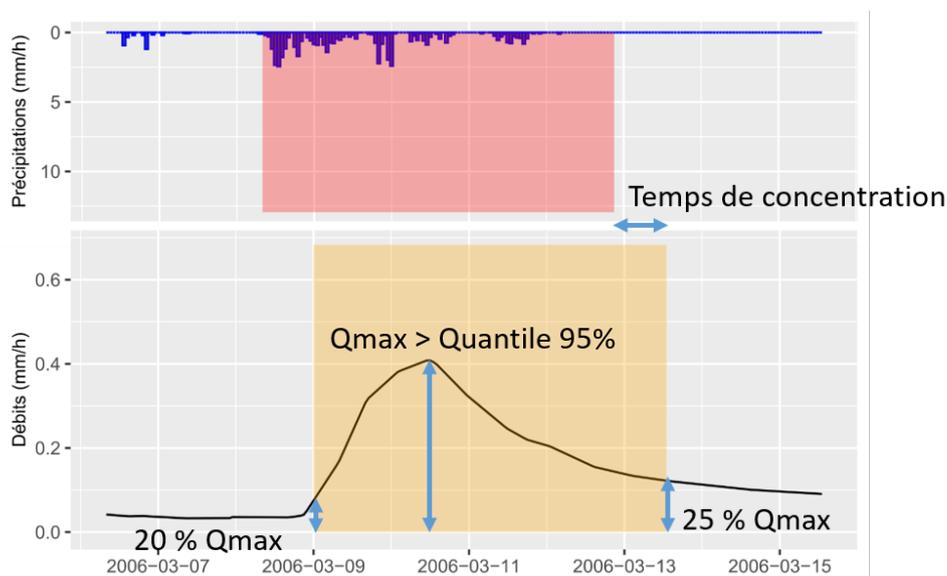


FIGURE 1.4 – Exemple de visualisation d'un hydrogramme de crue sélectionné sur le bassin versant la Bruche à Holtzheim (A2860110). Les rectangles orange et rouge correspondent aux fenêtres temporelles de débit et de pluie respectivement.

1.4 Caractéristiques des événements

Pour caractériser les événements sélectionnés sur notre base de données, nous utilisons plusieurs indicateurs (tableau 1.2). Des indicateurs plus spécifiques sur les événements de pluie seront utilisés dans le chapitre 3.

TABLE 1.2 – Liste des indicateurs utilisés pour caractériser les événements sélectionnés.

Indice	Unité
Durée de l'événement de crue	h
Cumul de débit	mm
Débit maximal	mm/h
Cumul de pluie	mm
Intensité de pluie maximale	mm/h
Rendement défini comme le rapport entre le volume de crue et le cumul de pluie	%

Il existe une variabilité spatiale importante des caractéristiques des événements entre les bassins versants de notre base de données (fig. 1.5 et 1.6). On notera notamment :

- Les événements les plus longs sont localisés sur des affluents de la Loire, de la Seine amont et de la Dordogne. Les bassins correspondants ont une superficie élevée. Les crues des rivières côtières méditerranéennes et de la plupart des affluents du Rhône sont plus courtes en moyenne.
- Les affluents de la Saône et plusieurs affluents en rive droite du Rhône présentent des cumuls de débit élevés.
- Les débits maximaux moyens sont nettement plus élevés dans les Cévennes et pour plusieurs rivières côtières méditerranéennes.
- Les cumuls de pluie suivent la même tendance, avec plusieurs affluents de la Saône et de la Dordogne qui présentent des valeurs élevées.
- Les événements de pluie qui ont lieu dans les Cévennes, sur les bassins méditerranéens et sur plusieurs affluents du Rhône présentent des valeurs élevées d'intensité de pluie maximale par rapport aux autres événements de la base de données.
- Les affluents de la Saône, plusieurs affluents en rive droite de la Loire, quelques affluents de la Dordogne et plusieurs affluents de l'Aisne présentent des rendements événementiels élevés.

1.5 Conclusions sur la base de données

La base de données présentée dans ce chapitre est en adéquation avec les objectifs de nos travaux de recherche car elle est composée d'un large échantillon de bassins versants et d'événements de crue, ce qui nous donnera la possibilité de tirer des conclusions générales sur la performance des modèles hydrologiques que nous allons évaluer. Les événements de crue de cette base de données sont caractérisés par des processus hydrologiques très différents. L'échantillon intègre par exemple des événements longs associés à des cumuls de pluie élevés mais des intensités de pluie faibles. Ces événements ont généralement lieu en hiver sur des bassins où l'humidité antécédente contrôle très largement la réponse hydrologique. Notre

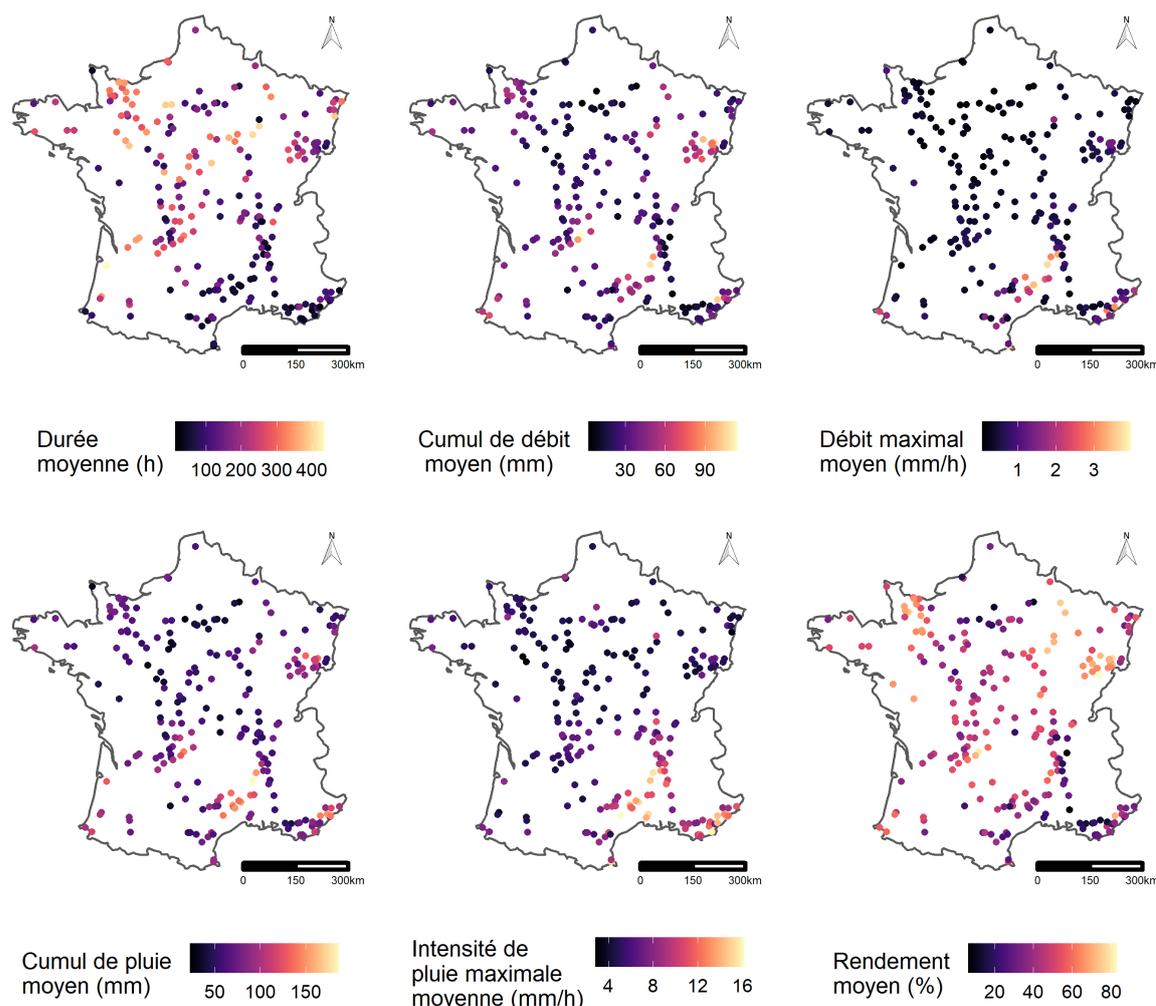


FIGURE 1.5 – Répartition spatiale des caractéristiques hydroclimatiques des 10 652 événements de crue sélectionnés (calculées sur la période 2000-2018).

base de données intègre également des crues associées à des intensités de pluie très élevées qui peuvent induire des réactions très rapides des bassins versants en condition de faible humidité antécédente. La variété des caractéristiques des événements et des bassins nous permettra de bien distinguer les conditions dans lesquelles les modèles ont des performances plus faibles des conditions dans lesquelles les modèles ont des performances élevées. Nous construirons les hypothèses de modélisation à partir de ce diagnostic ciblé.

Les conclusions que nous pourrions tirer des travaux effectués sur cette base de données seront toutefois limitées à des contextes hydroclimatiques similaires à ceux de la France métropolitaine. Une perspective de notre travail serait d'ajouter des bassins versants localisés dans des régions au climat aride, comme certaines parties de l'Australie, ou au contraire beaucoup plus humides. Par ailleurs, les données de débit, de précipitation et de température sont par nature entachées d'incertitudes. Par exemple, les courbes de tarage, qui permettent d'effectuer la transformation hauteur-débit, sont imparfaites du fait de la difficulté à effectuer des jaugeages en crue. L'utilisation d'une grande base de données permet toutefois de limiter les effets globaux des différentes sources d'erreurs sur la modélisation.

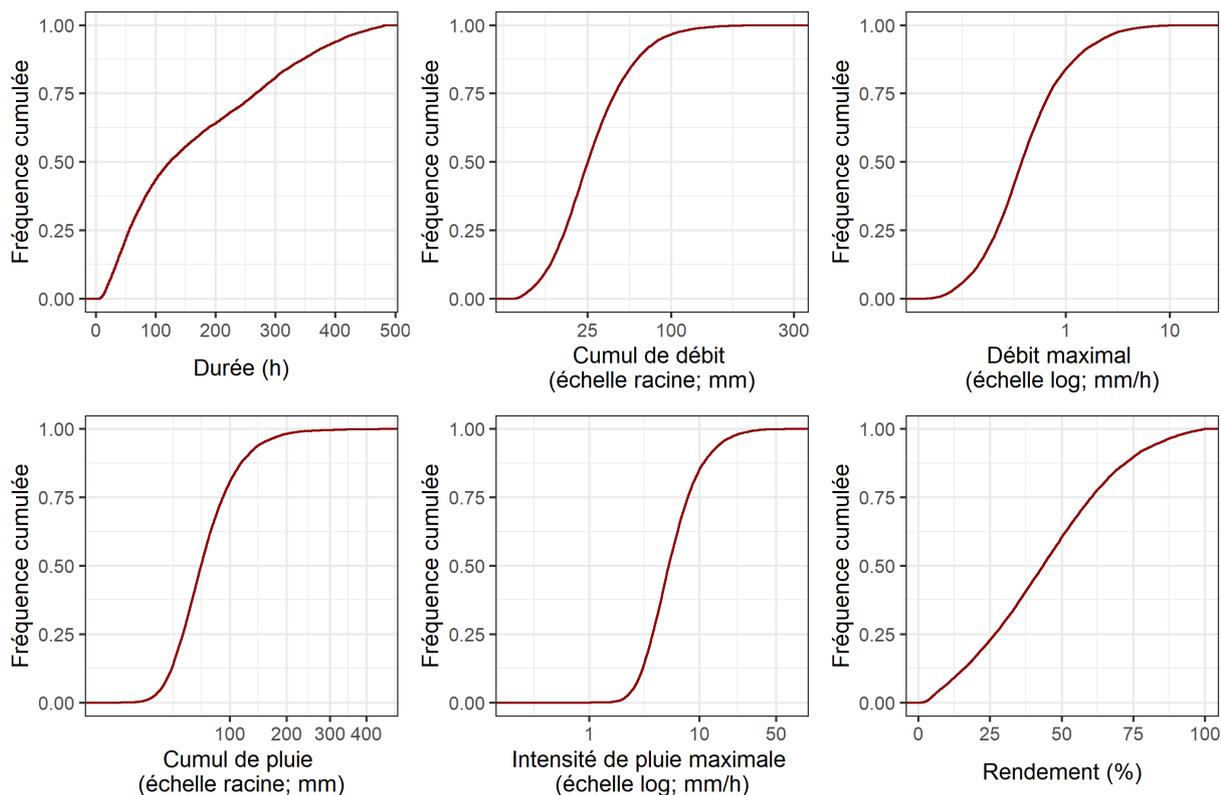


FIGURE 1.6 – Distribution des caractéristiques hydroclimatiques des 10 652 événements de crue sélectionnés (calculées sur la période 2000-2018).

Dans ce chapitre, nous avons présenté une classification des bassins versants en fonction de leur régime hydrologique. Deux autres classifications sont utilisées dans nos travaux : une séparation entre les bassins dits « méditerranéens » et les autres bassins versants de la base de données (chapitre 3) ; une classification orientée vers les caractéristiques des événements de crue et des événements de pluie associés (chapitres 5, 6, 7 et 8).

1.6 Modèles hydrologiques

1.6.1 Le modèle GR5H

Le modèle empirique global GR5H est utilisé dans la première partie de nos travaux (chapitres 2, 3 et 4). Nous utilisons une version qui fonctionne au pas de temps horaire et qui intègre un réservoir d'interception (Ficchi *et al.*, 2019). Les hypothèses de modélisation du chapitre 5 sont construites à partir de cette structure. Nous décrivons ici les principales équations (intégrées sur le pas de temps) de ce modèle. La figure 1.7 représente les états et les flux du modèle. Les définitions des variables et des paramètres sont données dans le tableau 1.3 et le tableau 1.4. On trouvera plus de détails sur le modèle dans les travaux de Le Moine (2008), Lobligois (2014) et Ficchi *et al.* (2019).

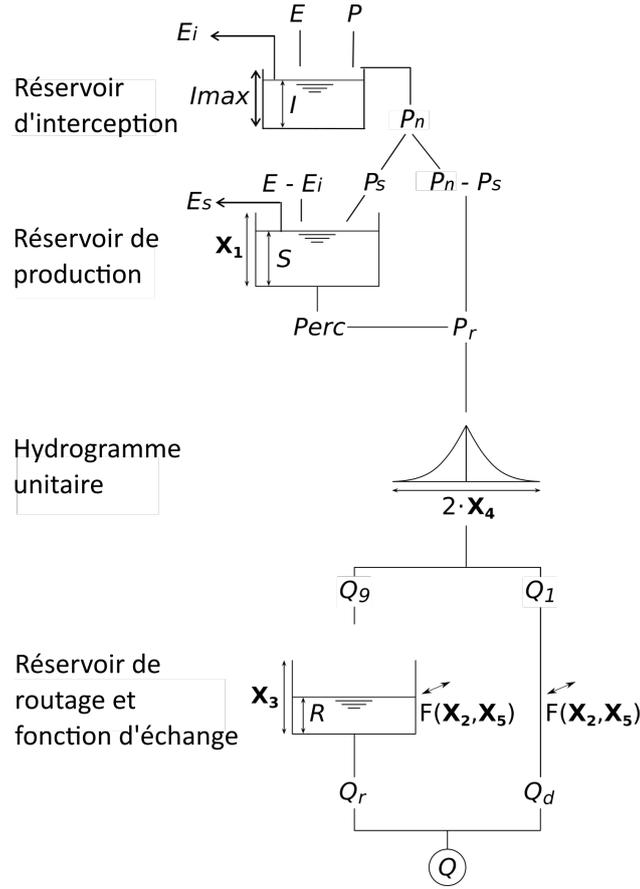


FIGURE 1.7 – Schéma du modèle GR5H (modifié à partir de Le Moine, 2008).

Réservoir d'interception

Au pas de temps courant k , l'évapotranspiration E_i du réservoir d'interception est calculée à partir de l'évapotranspiration potentielle E , des précipitations P et du niveau antécédent du réservoir d'interception I :

$$E_i(k) = \min[E(k), P(k) + I(k-1)] \quad (1.1)$$

La pluie nette est ensuite calculée de la manière suivante :

$$P_n(k) = \max[0, P(k) - (I_{\max} - I(k-1)) - E_i(k)] \quad (1.2)$$

Le niveau du réservoir d'interception devient :

$$I(k) = I(k-1) + P(k) - E_i(k) - P_n(k) \quad (1.3)$$

Réservoir de production

La part de la pluie nette qui remplit le réservoir de production dépend du niveau antécédent de ce réservoir ($S(k-1)$) et de la pluie nette :

$$P_s(k) = \frac{X_1 \left(1 - \left(\frac{S(k-1)}{X_1}\right)^2\right) \tanh\left(\frac{P_n(k)}{X_1}\right)}{1 + \frac{S(k-1)}{X_1} \tanh\left(\frac{P_n(k)}{X_1}\right)} \quad (1.4)$$

Dans le cas où la différence $E(k) - E_i(k)$ est supérieure à zéro, l'évapotranspiration du réservoir de production dépend du niveau antécédent du réservoir de production :

$$E_s(k) = \frac{S(k-1) \cdot \left(2 - \frac{S(k-1)}{X_1}\right) \tanh\left(\frac{E(k)-E_i(k)}{X_1}\right)}{1 + \left(1 - \frac{S(k-1)}{X_1}\right) \tanh\left(\frac{E(k)-E_i(k)}{X_1}\right)} \quad (1.5)$$

Le niveau du réservoir de production devient :

$$S(k) = S(k-1) + P_s(k) - E_s(k) \quad (1.6)$$

La percolation du réservoir de production est calculée comme suit :

$$Perc(k) = S(k) \cdot \left\{ 1 - \left[1 + \left(\frac{4}{21} \cdot \frac{S(k)}{X_1} \right)^4 \right]^{-\frac{1}{4}} \right\} \quad (1.7)$$

La pluie efficace transmise à l'hydrogramme unitaire est calculée comme suit :

$$P_r(k) = P_n(k) - P_s(k) + Perc(k) \quad (1.8)$$

Le flux total d'évapotranspiration est calculé de la manière suivante :

$$ETR(k) = E_s(k) + E_i(k) \quad (1.9)$$

Hydrogramme unitaire (HU)

La courbe en S au temps t , notée $SH(t)$, est définie (indépendamment du pas de temps courant) par :

$$\begin{cases} 0 & t < 0 \\ \frac{1}{2} \cdot \left(\frac{t}{X_4}\right)^{1.25} & 0 < t < X_4 \\ 1 - \frac{1}{2} \cdot \left(2 - \frac{t}{X_4}\right)^{1.25} & X_4 < t < 2 \cdot X_4 \\ 1 & t \geq 2 \cdot X_4 \end{cases} \quad (1.10)$$

Les ordonnées de l'hydrogramme unitaire qui en résultent sont calculées par la formule suivante :

$$HU(j) = SH(j) - SH(j-1) \quad (1.11)$$

où j est un entier entre 1 et le nombre maximum d'ordonnées de l'hydrogramme unitaire n .

Les ordonnées de l'hydrogramme unitaire sont utilisées pour calculer le flux sortant de l'hydrogramme unitaire au pas de temps courant k . Le flux sortant de l'hydrogramme unitaire allant vers la branche indirecte de routage est calculé de la manière suivante :

$$Q_9(k) = \frac{9}{10} \cdot \sum_{j=1}^n HU(j) \cdot P_r(k-j+1) \quad (1.12)$$

Le flux sortant de l'hydrogramme unitaire allant vers la branche pseudo-directe est calculé comme suit :

$$Q_1(k) = \frac{1}{10} \cdot \sum_{j=1}^n HU(j) \cdot P_r(k-j+1) \quad (1.13)$$

TABLE 1.3 – Liste des variables du modèle GR5H.

Notation	Définition	Unité
E	Données d'évapotranspiration potentielle	mm/h
P	Données de précipitations	mm/h
E_n	Évapotranspiration nette	mm/h
E_i	Évapotranspiration du réservoir d'interception	mm/h
I_{max}	Capacité du réservoir d'interception	mm
I	Niveau du réservoir d'interception	mm
P_n	Pluie nette du réservoir d'interception	mm/h
E_s	Évapotranspiration du réservoir de production	mm/h
P_s	Part de la pluie nette qui remplit le réservoir de production	mm/h
S	Niveau du réservoir de production	mm
Perc	Percolation du réservoir de production	mm/h
P_r	Pluie efficace	mm/h
Q_9	Part de la sortie de l'UH qui remplit le réservoir de routage	mm/h
Q_1	Part de la sortie de l'UH qui ne remplit pas le réservoir de routage	mm/h
F	Demi-échange potentiel	mm/h
Q_r	Vidange du réservoir de routage	mm/h
Q_d	Flux venant de la branche directe	mm/h
R	Niveau du réservoir de routage	mm
Q	Débit simulé	mm/h

TABLE 1.4 – Liste des paramètres libres du modèle GR5H.

Notation	Définition	Unité
X_1	Capacité du réservoir de production	mm
X_2	Coefficient d'échange	mm/h
X_3	Capacité du réservoir de routage	mm
X_4	Demi-temps de base de l'hydrogramme unitaire	h
X_5	Seuil d'échange	-

Fonction d'échange

Un flux d'échange entre bassins versants topographiques est ensuite ajouté (ou retiré) à (ou de) chaque flux en fonction du niveau antécédent du réservoir de routage. Le demi-flux d'échange potentiel est calculé à partir de la formule suivante :

$$F(k) = X_2 \left(\frac{R(k-1)}{X_3} - X_5 \right) \quad (1.14)$$

Réservoir de routage

Le niveau du réservoir de routage en début de pas de temps devient :

$$R_0(k) = \max[0, R(k-1) + Q_9(k) + F(k)] \quad (1.15)$$

La vidange de ce réservoir est calculée de la manière suivante :

$$Q_r(k) = R_0(k) \cdot \left\{ 1 - \left[1 + \left(\frac{R_0(k)}{X_3} \right)^4 \right]^{-\frac{1}{4}} \right\} \quad (1.16)$$

Le niveau du réservoir de routage en fin de pas de temps devient :

$$R(k) = R_0(k) - Q_r(k) \quad (1.17)$$

Le flux de la branche pseudo-directe est modifié à partir du flux d'échange :

$$Q_d(k) = \max[0, Q_1(k) + F(k)] \quad (1.18)$$

Le débit simulé final est calculé comme suit :

$$Q(k) = Q_r(k) + Q_d(k) \quad (1.19)$$

1.6.2 Le modèle GRSD

Une version semi-distribuée du modèle GR5H est utilisée dans le chapitre 6. Cette structure, nommée GRSD (fig. 1.8), a été développée par Lobligois (2014) pour prendre en compte explicitement la variabilité spatiale de la pluie. Chaque bassin versant est divisé en sous-bassins versants selon la direction d'écoulement et le choix de la taille des unités spatiales (pour une description de la méthode, voir Le Moine, 2008). Le modèle GR5H est appliqué sur chaque sous-bassin versant avec le module d'interception développé par Fichì *et al.* (2019). Les précipitations et l'ETP sont moyennées indépendamment sur chaque sous-bassin à partir des produits spatialisés bruts (COMPEHORE et SAFRAN). Les débits simulés sur chaque sous-bassin amont sont combinés à partir d'un module de propagation hydraulique linéaire, consistant en une simple translation temporelle basée sur un paramètre de célérité (C en m.s⁻¹) et la distance hydraulique entre exutoires (Lh en m) afin de calculer le débit de l'exutoire jaugé en aval.

1.6.3 Le modèle GRP

Le modèle de prévision GRP (Tangara, 2005 ; Berthet, 2010 ; Viatgé *et al.*, 2019b) est utilisé dans la dernière partie de nos travaux (chapitres 8 et 9). Sa structure dérive de la structure du modèle GR4J (Perrin *et al.*, 2003). Le modèle GRP a trois paramètres libres (tab. 1.5) et intègre une procédure d'assimilation du dernier débit observé pour mettre à jour le niveau du réservoir de routage à l'instant de prévision. Une correction basée sur l'erreur antérieure du modèle peut être appliquée aux sorties du modèle. La figure 1.9 représente les états et les flux du modèle.

Les modules d'interception et de production de GRP sont basés sur les mêmes équations que le modèle GR5H. Contrairement à GR5H, la capacité du réservoir de production est fixée à 275 mm. Le module de routage de GRP est différent de celui de GR5H et se compose d'une seule branche d'écoulement. La pluie efficace estimée par la fonction de production est multipliée par le paramètre d'ajustement *CORR*. La pluie efficace résultant de cet ajustement

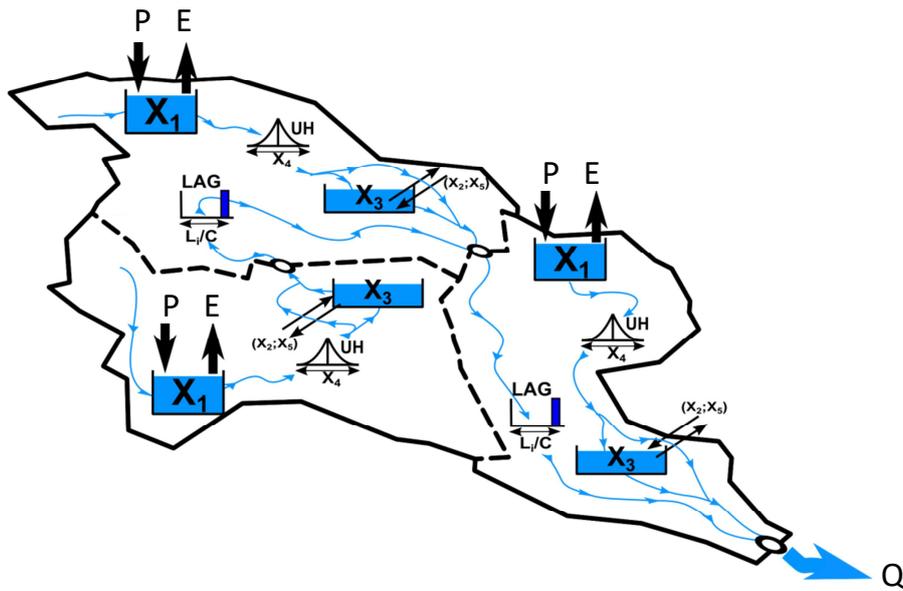


FIGURE 1.8 – Schéma du modèle GRSD. Source : Lobligeois *et al.* (2014). Les entrées de pluie (P) et d'évapotranspiration potentielle (E) sont calculées à l'échelle de chaque sous-bassin.

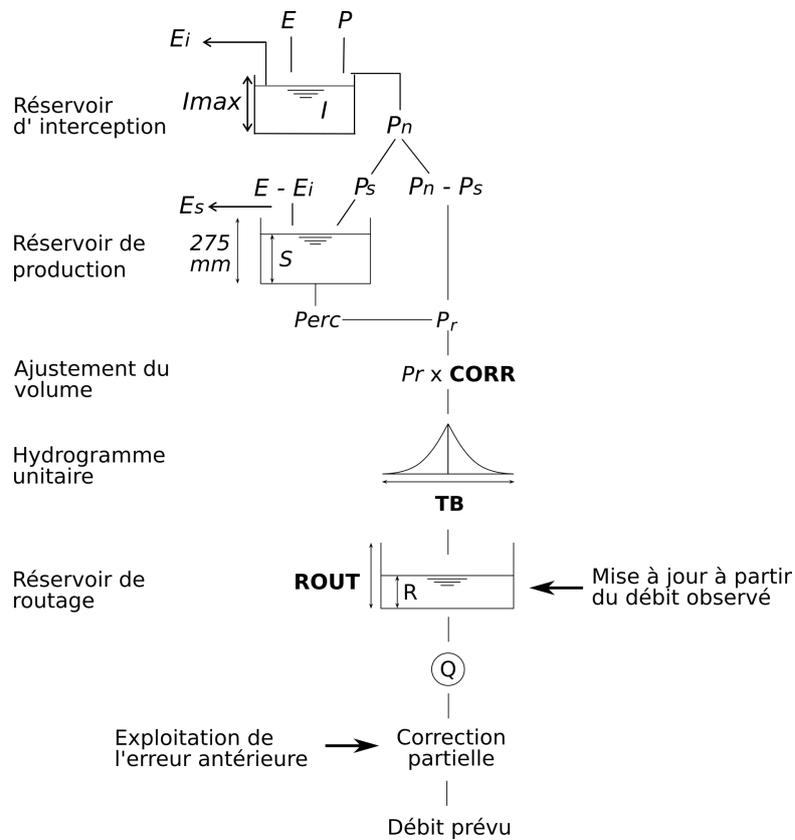


FIGURE 1.9 – Schéma du modèle GRP (modifié à partir de Viatgé *et al.*, 2019b).

TABLE 1.5 – Liste des paramètres libres du modèle GRP.

Notation	Définition	Unité
<i>CORR</i>	Facteur d'ajustement de la pluie efficace	-
<i>TB</i>	Temps de base de l'hydrogramme unitaire	h
<i>ROUT</i>	Capacité du réservoir de routage	mm

alimente un hydrogramme unitaire symétrique caractérisé par un temps de base TB . La courbe en S de cet hydrogramme unitaire est calculée au temps t comme suit :

$$\begin{cases} 0 & t \leq 0 \\ \frac{t^{1.25}}{t^{1.25} + (TB-t)^{1.25}} & 0 \leq t \leq TB \\ 1 & t \geq TB \end{cases} \quad (1.20)$$

Les ordonnées de l'hydrogramme unitaire sont calculées avec l'équation 1.11. Le flux sortant de l'hydrogramme unitaire alimente ensuite le réservoir de routage. Au pas de temps courant k , ce flux est calculé de la manière suivante :

$$QH U(k) = \sum_{j=1}^n H U(j) \cdot P_r(k-j+1) \quad (1.21)$$

avec n le nombre maximum d'ordonnées de l'hydrogramme unitaire.

Le réservoir de routage de GRP suit une loi de vidange quadratique. Son niveau en début de pas de temps est calculé comme suit :

$$R_0(k) = R(k-1) + QH U(k) \quad (1.22)$$

Le débit simulé est égal à la vidange de ce réservoir, qui est calculée comme suit :

$$Q(k) = Q_r(k) = \frac{R_0(k)^2}{R_0(k) + ROUT} \quad (1.23)$$

Le niveau du réservoir de routage en fin de pas de temps devient :

$$R(k) = R_0(k) - Q_r(k) \quad (1.24)$$

Les équations relatives à la mise à jour du réservoir de routage et à la correction des sorties sont présentées dans le chapitre 8 (eq. 8.1 et 8.5).

1.7 Estimation des paramètres et évaluation des performances

Les paramètres des modèles sont estimés à partir d'une procédure de calage effectuée par bassin et sur deux sous-périodes indépendantes (P1 et P2 définies dans la section 1.2). La capacité du réservoir d'interception est estimée au préalable, c'est-à-dire indépendamment de la procédure de calage automatique. Sa valeur a été définie pour chaque bassin versant en minimisant la différence entre les flux d'interception journaliers et les flux d'interception horaires, comme suggéré par Ficchi (2017). Une période de 2 ans précédant le début de chaque sous-période est appliquée pour initialiser les états des modèles (les critères de performance ne sont pas calculés sur ces périodes). L'algorithme de calage est implémenté dans le package `airGR` (Coron *et al.*, 2017, 2020) disponible dans le langage de programmation R (R Core Team, 2021). Cet algorithme procède en deux étapes pour trouver un jeu de paramètres optimal :

- Un préfiltrage de l'espace des paramètres est effectué pour trouver un point de départ (Mathevet, 2005). Ce préfiltrage consiste à tester 3^n jeux de paramètres (avec n le nombre de paramètres libres) et à sélectionner celui ayant la meilleure valeur de la fonction objectif. Cette grille de départ est constituée à partir de trois valeurs par paramètre. Ces valeurs correspondent aux centiles 10 %, 50 % et 90 % estimés sur 900 bassins versants par Perrin *et al.* (2008). Les mêmes paramètres de départ sont utilisés pour tous les bassins.
- Une descente locale de gradient (appelée méthode « pas-à-pas » ; Michel, 1987) est ensuite effectuée à partir du point de départ sélectionné lors du préfiltrage.

Cette procédure de calage est effectuée dans l'espace des paramètres transformés. Ces transformations appliquées aux paramètres sont nécessaires à l'application de la méthode de recherche locale. Celle-ci utilise en effet un pas de recherche fixe quel que soit le paramètre. Les transformations permettent donc d'avoir une sensibilité comparable à ce pas de recherche entre les différents paramètres. Cela permet également d'effectuer la recherche dans une gamme comparable (intervalle $[-10;10]$) pour tous les paramètres transformés. Le critère de performance KGE (Gupta *et al.*, 2009) est utilisé comme fonction objectif dans les chapitres 2, 3, 4, 5, 6 et 7. Le critère de performance NSE (Nash & Sutcliffe, 1970) est utilisé comme fonction objectif dans les travaux sur la prévision (chapitres 8 et 9). Des méthodes alternatives pour estimer les paramètres des modèles sont testées dans les chapitres 7 et 9.

Les performances des modèles sont évaluées sur les sous-périodes de calage ainsi qu'en validation croisée. La validation croisée consiste, pour un bassin donné, à utiliser le jeu de paramètres estimé en calage sur une sous-période donnée pour simuler le débit de l'autre sous-période et vice versa. Plusieurs critères d'erreur sont utilisés dans nos travaux. Nous distinguons les critères calculés sur toute la chronique des débits des critères calculés indépendamment sur chaque événement de crue. Ces critères seront définis dans les sections présentant les éléments méthodologiques spécifiques à chaque chapitre. Le tableau 1.6 présente la liste des critères d'évaluation des performances.

TABLE 1.6 – Liste des critères de performance utilisés (tous ces critères sont adimensionnels).

Notation	Définition	Références	Chapitres
KGE	Critère d'efficacité de Kling et Gupta	Gupta <i>et al.</i> (2009)	2, 3, 5, 6, 7 et 8
NSE	Critère d'efficacité de Nash et Sutcliffe	Nash & Sutcliffe (1970)	8 et 9
β_{ts}	Biais calculé sur toute la chronique de débit d'une sous-période donnée	eq. 2.1	2
β_{ev}	Biais calculé sur un événement de crue	eq. 2.2	2, 3, 4, 5, 6 et 7
β_{cEv}	Biais calculé sur les événements de crue d'une sous-période donnée	eq. 2.3	2 et 4
β_{cWev}	Biais calculé sur la chronique de débit d'une sous-période donnée sans les événements de crue	eq. 2.4	2 et 5
NSE_{ev}	Critère NSE calculé sur un événement de crue	eq. 3.1	3, 5, 6, 7, 8 et 9
ΔQ_{max}	Erreur relative sur les pointes de crue par événement	eq. 6.1	6 et 7
$\Delta_{NSE_{ev}}$	Ratio des valeurs de NSE événementiel entre résolutions spatiales	eq. 6.2	6
CP	Critère de persistance	Kitanidis & Bras (1980)	8 et 9
$\Delta Q_{max}P$	Erreur relative sur les pointes de crue par événement (en fonction de l'échéance)	eq. 6.1 et chapitre 8	8 et 9
POD	Probabilité de détection	eq. 8.6	8
FAR	Taux de fausses alertes	eq. 8.7	8
CSI	Indice de succès critique	eq. 8.8	8
R_{C2MP}	Ratio des valeurs de C2MP	eq. 8.9	8
$R_{NSE_{ev}}$	Ratio des valeurs de NSE événementiel	eq. 8.10	8

Partie I

**Diagnostic approfondi
du modèle de simulation :
des erreurs aux facteurs
explicatifs**

Chapitre 2

Diagnostic du modèle de simulation GR5H

Ce chapitre correspond à un article publié dans la revue *Hydrological Sciences Journal* : Paul C. Astagneau, François Bourgin, Vazken Andréassian and Charles Perrin (2021). When does a parsimonious model fail to simulate floods? Learning from the seasonality of model bias. *Hydrological Sciences Journal*, 66:8, 1288-1305, <https://doi.org/10.1080/02626667.2021.1923720>.

2.1 Résumé

Identifier les situations pour lesquelles un modèle hydrologique présente des performances dégradées est un prérequis pour améliorer sa capacité prédictive. Dans cette étude, nous avons appliqué une méthodologie d'évaluation qui vise à diagnostiquer les déficiences d'un modèle pluie-débit parcimonieux lors de la simulation des crues. Le modèle global GR5H a été évalué sur un large échantillon de 229 bassins versants français où 2 990 événements de crue ont été sélectionnés. Le biais du modèle a été calculé en considérant plusieurs fenêtres temporelles de débit, de la chronique de débit totale aux événements de crue. Les biais ont ensuite été analysés en fonction des saisons et de plusieurs caractéristiques des crues. Les résultats montrent que, malgré ses performances globales élevées, GR5H sous-estime la majorité des événements de crue qui ont lieu en été. En été et en automne, des compensations entre les périodes de crue et les périodes de récession ont été identifiées. Les sous-estimations les plus larges du volume de crue ont lieu pour des événements de pluie très intenses, en particulier en condition de faible humidité antécédente des bassins versants.

2.2 Introduction

2.2.1 On the need to improve hydrological models for flood forecasting

Reliable hydrometeorological predictions are important for mitigating the hazards associated with floods, including loss of lives and livelihoods as well as economic losses (e.g., Carsell *et al.*, 2004; Hallegatte, 2012; Jeuland *et al.*, 2019). Many operational flood forecasting systems have been implemented around the world to produce streamflow forecasts (Pappenberger *et al.*, 2016) and hydrological models are the basis of these flood forecasting systems (Pagano *et al.*, 2014). However, streamflow predictions produced by these models are still subject to large uncertainties (Roundy *et al.*, 2018). Different sources of uncertainty can affect the predictive capability of hydrological models (e.g., Beven, 2016). These uncertainties can be related to the input data, model parameterization or model structural deficiencies for example. Consequently, advances in hydrological modelling are needed to obtain more accurate estimates of flood peak, timing and volume, and therefore issue earlier and better warnings (e.g., Pagano *et al.*, 2014; Jain *et al.*, 2018). Flood forecasting systems have been improved over the years (e.g., Zanchetta & Coulibaly, 2020). For example, the use of continuous models instead of event-based models allowed a reduction of the uncertainty associated with initial conditions (e.g., Berthet *et al.*, 2009; Grimaldi *et al.*, 2021). However, common failures to predict floods are still encountered. For example, the severe floods that occurred in the Seine and Loire rivers and their tributaries (France) in June 2016 were underestimated by the GRP flood forecasting model (e.g., Berthet, 2010; Viatgé *et al.*, 2019a). Hydrological models are also less reliable in arid or dry areas (e.g., McMillan *et al.*, 2016; Melsen *et al.*, 2018), especially when flash floods occur (Hapuarachchi *et al.*, 2011). Improving the predictive capability of hydrological models is therefore essential for improved flood forecasting.

2.2.2 Relevance of large-sample model diagnostics

To be able to improve a model, one must first identify situations where the model fails to yield reliable results, i.e., the first-order factors leading to simulation errors (Gupta *et al.*, 2008). In this regard, various diagnostic methods have been used on hydrological models. Some studies were based on a limited number of catchments (e.g., Butts *et al.*, 2004; Clark *et al.*, 2008; Nicolle *et al.*, 2014). Diagnostics are now increasingly relying on large-sample approaches in order to draw more general conclusions from model performance (e.g., Andréassian *et al.*, 2009; Gupta *et al.*, 2014). Large-sample studies make it possible to establish performance benchmarks, i.e., to determine the current performance of hydrological models across a representative set of catchments (e.g., Seibert, 2001; Seibert *et al.*, 2018). The effects of changes in, for instance, model structure or parameterization can then be assessed in light of the performance of the benchmark. Several recently published country-scale studies investigate different model structures over large sets of catchments representing a variety of hydro-climatic conditions. For example, Lane *et al.* (2019) investigated the predictive capability of four models over 1000 catchments in the UK and the relationships between model performance and catchment attributes, flow regimes and model structures. Knoben

et al. (2020) compared the performance of 38 models over 559 US catchments and highlighted groups of models for which the structural hypotheses were better suited for some catchments rather than others (e.g., seven models had better performance for flashy catchments).

2.2.3 Model evaluation issues

One of the challenges of large-sample hydrology is related to the advantage of exploiting the robust statistical properties of large samples (Mathevet *et al.*, 2006, 2020) while summarizing the results into an understandable outcome but still maintaining a certain degree of precision to find emerging patterns (Gupta *et al.*, 2014). Many large-sample studies used aggregated statistics to assess model performance across comprehensive datasets (e.g., Perrin *et al.*, 2008; Vaze *et al.*, 2010, 2011; Coron *et al.*, 2012; Andréassian *et al.*, 2014). Some comparative studies of rainfall-runoff models found similar levels of performance between model structures when using aggregated metrics (e.g., Perrin *et al.*, 2001; van Esse *et al.*, 2013). It has been widely recognized that the use of single aggregated metrics is not sufficient to assess model performance (e.g., Schaeffli & Gupta, 2007; Euser *et al.*, 2013). For example, Mathevet *et al.* (2020) compared the performance of two conceptual models across a global set of catchments within a multi-objective framework. They found that while the two models yielded similar overall performance, when looking at sub-period statistics, one model performed slightly better in short-term processes. Furthermore, commonly used performance indicators, such as the Nash-Sutcliffe Efficiency criterion (NSE; Nash & Sutcliffe, 1970), are known to be biased towards high flows, although their use in model calibration leads to underestimation of flow variability (Gupta *et al.*, 2009) and a limited ability to reproduce extreme flows (e.g., Oudin *et al.*, 2006; Crochemore *et al.*, 2015). Therefore they must be used cautiously when evaluating the ability of models to simulate floods.

Hydrological signatures have been found useful in improving the identification of patterns between model performance and our understanding of underlying catchment processes (Yilmaz *et al.*, 2008; Hrachowitz *et al.*, 2014; McMillan *et al.*, 2017). They are used as indicators of situations where models perform well and are suitable for further applications. Consequently, a growing number of studies are using sets of hydrological signatures for model evaluation to emphasize particular aspects of the hydrograph (e.g., Shafii & Tolson, 2015; Donnelly *et al.*, 2015; Poncelet *et al.*, 2017; Gnann *et al.*, 2020). Signatures to investigate model performance on high flows can be for example the runoff volume above the 80th percentile of the flow duration curve (Yilmaz *et al.*, 2008). Another way to highlight model errors in flood predictions is to evaluate the simulation of peak flows by calculating specific criteria, such as the time to peak or the peak flow ratio. For example, Mizukami *et al.* (2019) investigated the choice of objective functions to better simulate high flows by calculating the annual peak flow bias.

2.2.4 A focus on events and seasonality

Event-based models are usually evaluated by their ability to reproduce several flood characteristics, such as rising limb, peak flow magnitude or timing (e.g., Borah *et al.*, 2007; Javelle *et al.*, 2010; Stanić *et al.*, 2017; Stephens *et al.*, 2018). To focus on the ability of continuous models to reproduce floods, model performance can also be assessed against a set of events in streamflow time series. By selecting 3620 flood events in 181 catchments, Lobligeois *et al.* (2014) identified where and when accounting explicitly for the spatial vari-

ability of rainfall and potential evapotranspiration inputs (i.e., by using a semi-distributed approach) improved streamflow simulations. Ficchi *et al.* (2016) showed that the peaks and timing of 2400 floods in 240 catchments were better reproduced at sub-daily time steps than at a daily time step. Vergara *et al.* (2016) evaluated the regionalization of the parameters of a hydraulic routing function by calculating peak flow and timing errors in the simulation of 47 563 flood events. de Boer-Euser *et al.* (2017) compared eight hydrological models in one catchment and found similar results in terms of overall performance but clear differences when looking at specific events and metrics. For different purposes, these studies showed that a change of focus in model evaluation, e.g., investigating flood events, can help to identify patterns in the performance of hydrological models.

Streamflow can have strong seasonal variations depending on climate seasonality and catchment characteristics (Berghuijs *et al.*, 2014; Gnann *et al.*, 2020). The seasonal streamflow variations also reflect varying antecedent soil moisture conditions, which can have significant impacts on flood event generation (Blöschl *et al.*, 2013; Berghuijs *et al.*, 2014). In terms of model diagnostics, some studies used signatures to assess the ability of the models to simulate the streamflow regime (e.g., Wang *et al.*, 2008; Massmann, 2019). Other model evaluations investigating the seasonality of model performance relied on the analysis of metrics calculated for each season of the year independently (e.g., Muleta, 2012; Kim & Lee, 2013; Lane *et al.*, 2019).

2.2.5 Scope of the paper

While it is clear that analyzing the performance of hydrological models in simulating flood events can help to characterize model performance, a deeper investigation on the seasonality of model performance considering different streamflow time windows has, to our knowledge, not been conducted to date. In an effort to find patterns of model errors and therefore help target model improvements, we apply a new methodology to assess the simulations of a continuous conceptual rainfall-runoff model over a large sample of flood events at an hourly time step. Bennett *et al.* (2013) suggested that, to evaluate environmental models, we must first look at basic performance criteria and then refine the analyses depending on the problem at hand. Here we follow these recommendations and aim to probe deeper in the analyses by looking at the seasonality of model bias through different disaggregations of the observed hydrograph. The objectives of this paper are (i) to determine whether investigating the seasonality of model bias through different streamflow time windows can provide information on model deficiencies in simulating high flows and (ii) to identify factors causing model weaknesses in flood simulations.

2.3 Data

2.3.1 Catchment set ¹

2.3.2 Selection of flood events

An automated procedure was used to select 2,990 flood events in the catchment dataset: 1681 events in P1 and 1309 events in P2. On average, 13 events were selected per catchment. The number of selected events ranges from 3 to 16 events per catchment. Only events with

1. Les éléments de cette section sont présentés dans le chapitre 1 et ne sont donc pas repris ici.

peak flow higher than the 95th streamflow quantile were selected. Flood event periods were taken from the time when streamflow is higher than 20 % of the event peak flow to the time when streamflow is lower than 30 % of the event peak flow. Each flood event was then visually inspected to avoid overlaps and other errors arising from the automated selection procedure. Figure 2.1 illustrates some of the flood event characteristics per season. In Fig. 2.1, the ratio of flood volume to total water volume was computed as the ratio between the volume of each flood and the total water volume. Winter was considered to extend from the beginning of January to the end of March, spring from April to June, summer from July to September, and autumn from October to December. The highest flow peaks relative to mean catchment flow occurred in summer and autumn, mostly in tributaries of the Rhône River and in the Mediterranean region. Summer events are shorter and of less volume than the other flood events in our dataset. Figure 2.2 shows that the volume of the selected events ranges from 2 % (Spring 5th percentile) to 45 % (Autumn 95th percentile) of the total streamflow volume of the corresponding season. The spring events represent a smaller fraction of the overall spring streamflow volume, whereas autumn events represent a larger proportion of the streamflow volume of the corresponding season.

2.4 Methods

2.4.1 Hydrological model and parameter calibration

The continuous GR5H lumped conceptual rainfall-runoff model (Le Moine, 2008; Lobligeois *et al.*, 2014) was used to simulate hourly streamflow time series at the outlet of each catchment. We used a version that integrates an interception store, as formulated by Ficchi *et al.* (2019). Full mathematical details of the model are given in chapter 1. We performed a continuous split-sample test (Klemeš, 1986) to obtain two parameter sets on P1 and P2, respectively. The model has five free parameters that were estimated for each catchment and each independent sub-period. A warm-up period of 2 years preceding the beginning of each sub-period was applied to initialize the model's states. The hourly precipitation time series of the warm-up period before P1 was built from a uniform disaggregation of the daily time series of the SAFRAN climate reanalysis data. The parameter estimation procedure is based on the algorithm developed by Michel (1987), a local gradient-based optimization procedure preceded by a gross screening of the parameter space (243 parameter sets tested; see the EGD method of Perrin *et al.*, 2008) so as to identify a suitable starting point. We used the same starting parameter sets for each catchment. These parameter sets correspond to the 10th, 50th and 90th percentiles of 900 catchments (see Perrin *et al.*, 2008, for more details). The parameter set that maximized the Kling-Gupta Efficiency criterion (KGE; Gupta *et al.*, 2009) was selected with the automated procedure. The interception store capacity was estimated before applying the parameter estimation algorithm. Its value was defined by minimizing the difference between daily and hourly interception fluxes as suggested by Ficchi (2017). Computations were made in the R environment using the `airGR` package (Coron *et al.*, 2017, 2020).

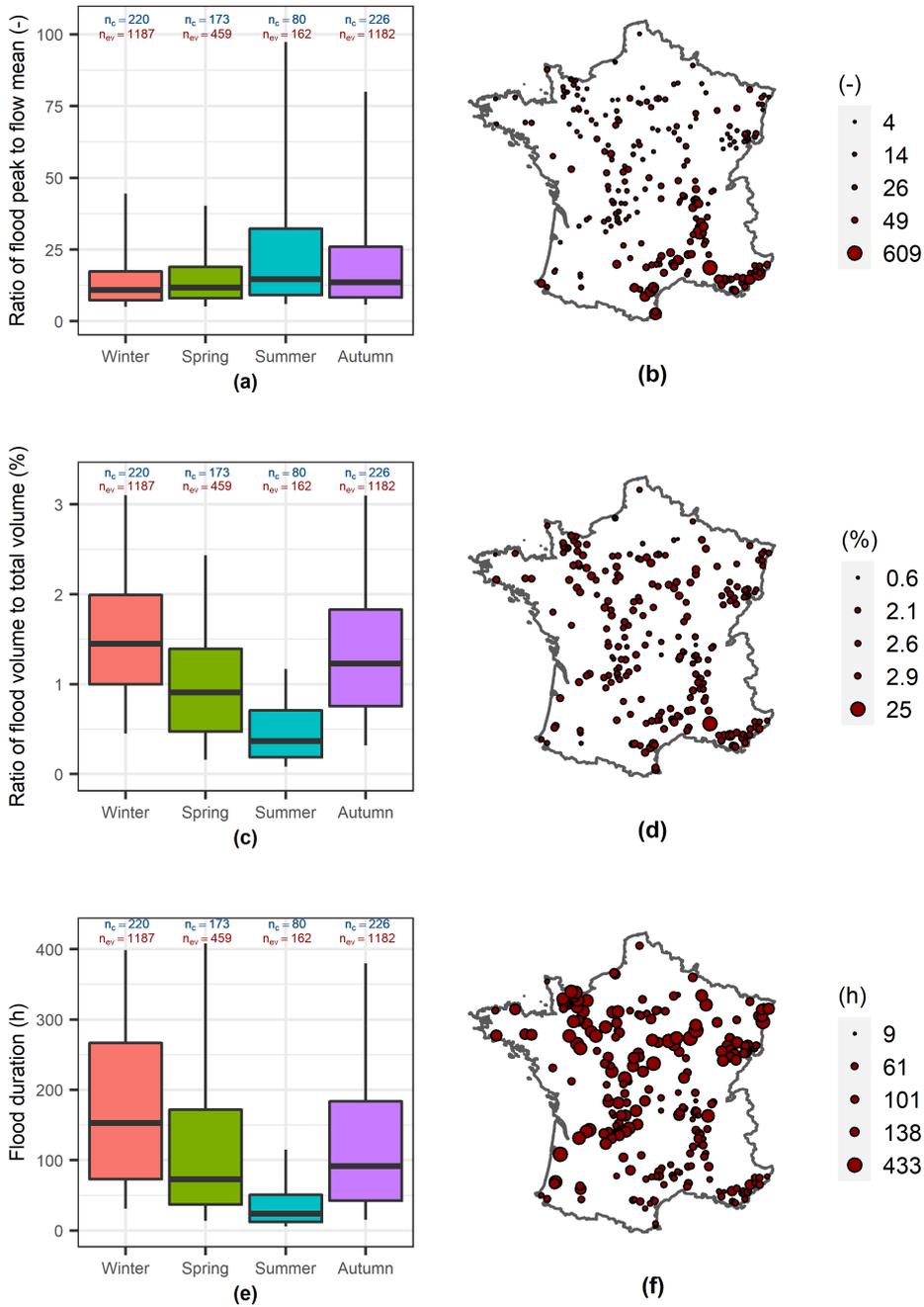


Figure 2.1 – Distribution and localization of the characteristics of 2,990 flood events. The distributions are presented between the 5th and 95th percentiles. Parts (b) and (d) present maximum values and part (f) presents mean values. n_{ev} is the number of flood events of each season and n_c the related number of catchments.

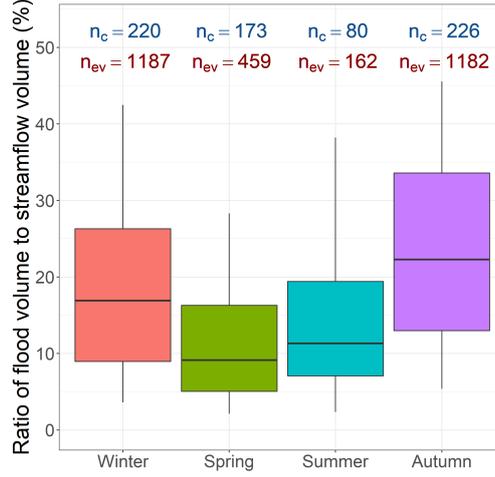


Figure 2.2 – Distribution of catchment seasonal flood volumes compared with seasonal streamflow volume for 2,990 flood events in 229 catchments. n_{ev} is the number of flood events of each season and n_c is the related number of catchments.

2.4.2 Evaluation

GR5H simulations were first assessed by looking at the KGE criterion calculated for the whole time series of each calibration and evaluation period. This gives a general overview of the model performance across our large set of catchments. The simulations were then evaluated by calculating the bias considering four levels of hydrograph disaggregation, i.e., different streamflow periods. Bias is a basic but essential indicator when evaluating how the model performs. Several studies showed the importance of investigating model bias, for example, to evaluate the robustness of rainfall-runoff models (e.g., Coron *et al.*, 2012; Fowler *et al.*, 2016). When investigating specific events, it is an indicator of how well a model is able to simulate flood volumes. These four criteria were calculated after model cross-validation, i.e., model application on P1 and P2 with the parameter sets of P2 and P1, respectively. In other words, the four biases were calculated on P1 with the parameter sets optimized on P2 and on P2 with the parameter sets optimized on P1. The first level of hydrograph disaggregation is the overall bias calculated for the whole time series (including flood flows, but also mean and low flows) of each sub-period independently using the following expression:

$$\beta_{ts} = \frac{\sum_{i=1}^p Q_{sim,i}}{\sum_{i=1}^p Q_{obs,i}} - 1 \in [-1, +\infty) \quad (2.1)$$

with p being the length of the sub-period, and $Q_{sim,i}$ and $Q_{obs,i}$ being the simulated and observed hourly streamflows at time i . β_{ts} is a component of the KGE criterion. The second level of disaggregation is defined as the bias calculated for each event independently, i.e., bias is computed from flows that were observed and simulated during a specific flood event. For a given flood event, the event bias is expressed as follows:

$$\beta_{ev,j} = \frac{\sum_{i=1}^{n_j} Q_{sim,i}}{\sum_{i=1}^{n_j} Q_{obs,i}} - 1 \in [-1, +\infty) \quad (2.2)$$

with n_j the length of the j^{th} event. A single value of β_{ev} was calculated for each of the 2,990 events using cross-validation simulations. The third level of disaggregation is a measure of flood volume errors at the catchment level. It allows us to compare the catchments

with the same weight – as the number of events differs between catchments – and also to compare event bias with overall bias. For each catchment, we calculated the bias for all the events combined on each evaluation period. In other words, the bias was calculated for the streamflow time series during the flood events, i.e., without the times at which there is no flood event. The bias was calculated for each sub-period of each catchment (458 values). For a given catchment and one sub-period, the calculation is expressed as follows:

$$\beta_{cEv} = \frac{\sum_{j=1}^m (\sum_{i=1}^{n_j} Q_{sim,i})}{\sum_{j=1}^m (\sum_{i=1}^{n_j} Q_{obs,i})} - 1 \quad \in [-1, +\infty) \quad (2.3)$$

with m the number of events of the catchment in a given sub-period. The fourth level of bias is calculated in the flow time series without the times at which floods occurred. It is a measure of how the model simulates the water balance outside the selected flood events, i.e., during mean and low flows. It will therefore help to identify compensations between streamflow periods. This bias can be expressed as:

$$\beta_{cWev} = \frac{(\sum_{i=1}^p Q_{sim,i}) - \sum_{j=1}^m (\sum_{i=1}^{n_j} Q_{sim,i})}{(\sum_{i=1}^p Q_{obs,i}) - \sum_{j=1}^m (\sum_{i=1}^{n_j} Q_{obs,i})} - 1 \quad \in [-1, +\infty) \quad (2.4)$$

Figure 2.3 illustrates the streamflow time windows that are considered in the calculation of β_{cEv} and β_{cWev} .

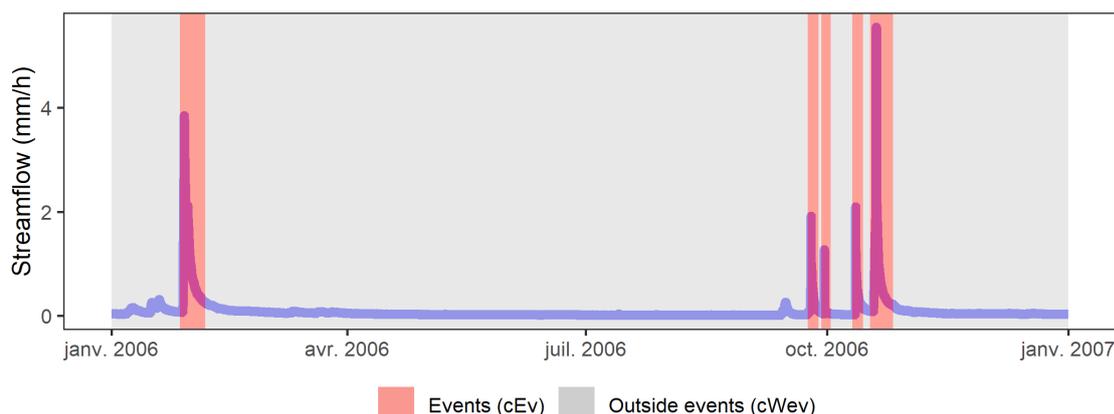


Figure 2.3 – Illustration of the streamflow time windows used to calculate the event bias per catchment and the bias outside periods with events per catchment. The hydrograph presented here corresponds to the Lergue River at Lodève (southern France).

A bounded version of these criteria, β^b , is calculated in order to facilitate comparison of results between catchments (Mathevet *et al.*, 2006), as the bias can tend towards very large values when streamflow is overestimated. The highly skewed distribution introduces difficulties in interpreting mean values and dispersion over a set of criteria values.

$$\beta^b = \frac{\beta}{\beta + 2} \quad \in [-1, 1] \quad (2.5)$$

Table 2.1 presents the corresponding values of β for some values of β^b . Negative values indicate underestimation of observed streamflow, and positive values indicate overestimation of observed streamflow by the model. A value of 0 indicates that there is no bias. We then refined the level of focus by investigating the seasonality of each bias. To calculate the four seasonal biases (β_{ts} , β_{ev} , β_{cEv} and β_{cWev}), we assumed that a flood occurred in a given season if the time of peak was in that season.

Table 2.1 – Correspondence between bias (β) and bounded bias (β^b) values.

β^b	β	β^b	β
-1	-1	0	0
-0.6	-0.75	0.11	0.25
-0.33	-0.5	0.2	0.5
-0.14	-0.25	0.33	1
-0.05	-0.1	0.5	2

2.4.3 Linking flood characteristics with model bias

Considering the catchments on which we identified patterns of seasonal event bias, we performed a univariate analysis between event bias and certain flood characteristics. Table 2.2 lists the characteristics considered in this analysis. Precipitation events were attributed to

Table 2.2 – List of the relative characteristics used for the univariate analysis. All characteristics are expressed as percentages.

Relative maximum hourly precipitation of the event
Relative maximum event streamflow (or peak flow)
Relative mean SWI index of the event (ISBA model; Thirel <i>et al.</i> , 2010a,b; Coustau <i>et al.</i> , 2015)
Relative event runoff coefficient (Q/P)
Relative event duration

each flood event. The time window of a precipitation event was set as the time window of the corresponding flood event negatively shifted by the time of concentration of the catchment (see Table 2.2). The soil wetness index (SWI), calculated by the ISBA (*Interaction Sol-Biosphère-Atmosphère*) surface model (Thirel *et al.*, 2010a,b; Coustau *et al.*, 2015), was used as an indicator of the soil moisture condition of the catchments. It is defined by Barbu *et al.* (2011) as follows:

$$SWI = \frac{w_{tot} - w_{wilt}}{w_{fc} - w_{wilt}} \quad (2.6)$$

with w_{tot} being the root soil moisture (or volumetric water content of the simulated soil column) [-], w_{wilt} the wilting point [-] and w_{fc} the field capacity [-]. We used the daily catchment averaged SWI values. We calculated the mean SWI values of each event. Each characteristic was then divided by its corresponding catchment event average (over both sub-periods) to compare the variability of events within catchments. For example, the relative duration of an event is calculated by dividing its duration by the average event duration of the catchment.

2.5 Results

2.5.1 Overall model performance

Figure 2.4 shows the distributions of KGE values over the 229 catchments and calculated for the calibration and validation periods. The first and third quartiles (Q_1 and Q_3) in

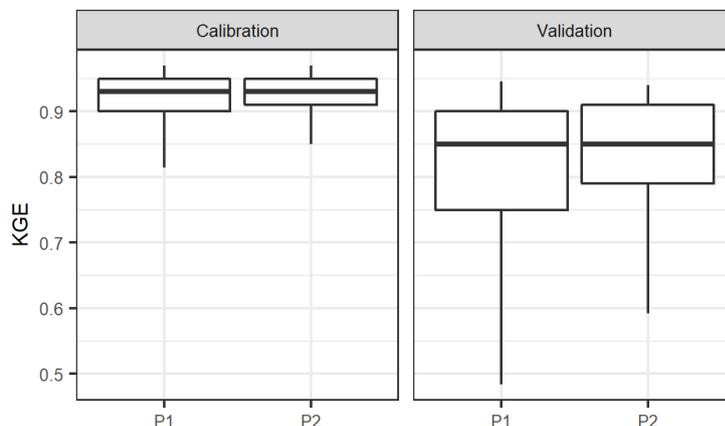


Figure 2.4 – Distribution of GR5H performance in calibration and validation periods over 229 French catchments. The distributions are presented between the 5th and 95th percentiles.

calibration mode are between 0.90 and 0.95 in P1 and between 0.91 and 0.95 in P2. In validation mode, Q_1 and Q_3 are between 0.75 and 0.90 in P1 and between 0.79 and 0.91 in P2. As expected, the distributions show a drop in performance between the calibration and validation periods. The distribution of KGE values is more scattered in P1 than in P2. This can be partly explained by a change in the runoff coefficient and cumulative rainfall between P1 and P2 for some of the catchments in our dataset (see Fig. 2.5). The lowest KGE values in validation mode correspond to a few catchments located in the Mediterranean region. These results are consistent with previous nationwide studies with the GR models (e.g., Lobligeois *et al.*, 2014; Ficchi *et al.*, 2016; Poncelet *et al.*, 2017). Crochemore *et al.* (2015) showed that KGE values between 0.66 and 0.90 are considered to represent good model performance according to expert judgment. We can therefore consider that the GR5H model yields good overall performance in our dataset.

2.5.2 Model bias on four levels of hydrograph disaggregation

We then refined the analysis and investigated whether the model was able to simulate the 2,990 flood events in our catchment dataset. In this regard, the model bias was calculated considering the four levels of hydrograph disaggregation (Fig. 2.6) defined earlier to enable identification of compensations between streamflow periods, i.e., between periods of floods and mean and low flows. The median event bias (β_{ev}^b) of the GR5H model is equal to -0.03, and its distribution is left-skewed towards negative values. The distribution of catchment event bias (β_{cEv}^b) is similar to the distribution of event bias with a median value of -0.02 and a lesser dispersion. When bias is calculated without taking the selected flood events into account (β_{cWev}^b), the median bias has a value of 0.01. The distribution is wider for positive values outside the interquartile range. The model bias calculated for the entire time series (β_{ts}^b) has a median value of 0.001 and a distribution similar to the distribution of β_{cWev}^b .

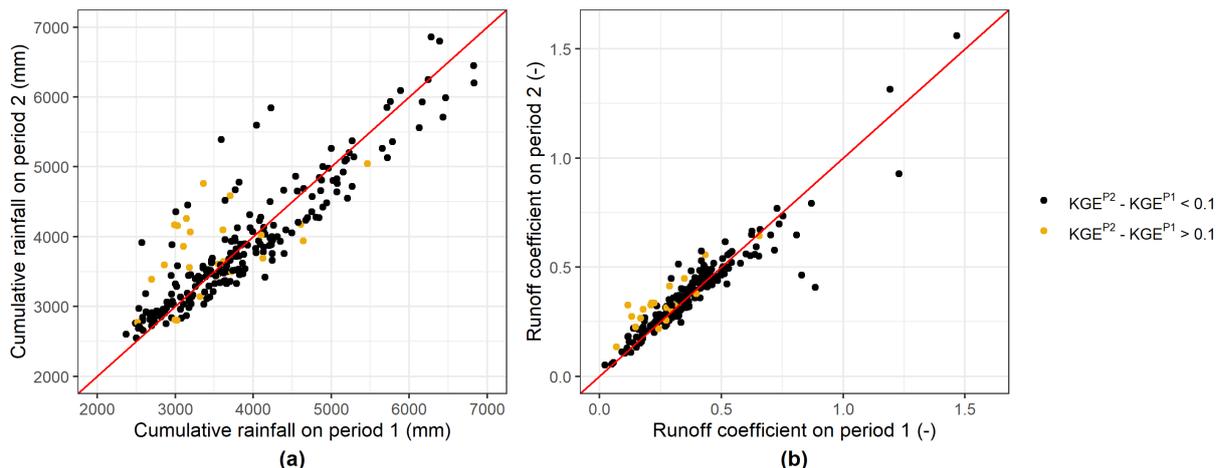


Figure 2.5 – Difference of (a) cumulative rainfall and KGE (b) runoff coefficient and KGE in validation mode between P1 and P2.

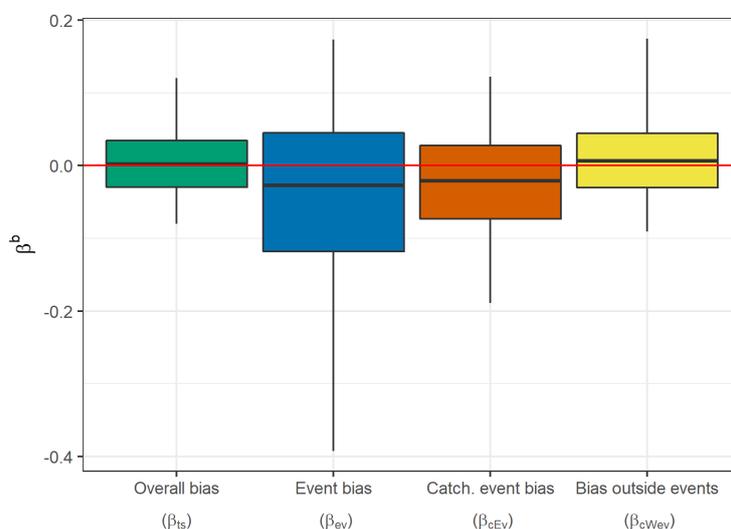


Figure 2.6 – Distribution of GR5H bias over 229 catchments and 2,990 events. The bias (bounded) was calculated for four levels of hydrograph disaggregation. The distributions are presented between the 5th and 95th percentiles. Calculations were made with cross-validation values, i.e., using simulations on P1 and P2 obtained with the parameter sets optimized on P2 and P1 respectively.

The dispersion of the event bias is wider than the distribution of the three other biases. The dispersion of the bias calculated for the whole time series is narrower than the other distributions. Overall, these results show that there are compensations between periods where flood events occur and the rest of the hydrograph. The four bias criteria were also calculated in the calibration period and the results showed similar patterns (not shown here).

2.5.3 Seasonality of model bias

We investigated to what extent certain flood events are better reproduced by the model. As the seasonal variation of streamflow affects antecedent soil moisture conditions, it is a good indicator of the variability in flood characteristics across Metropolitan France. We therefore

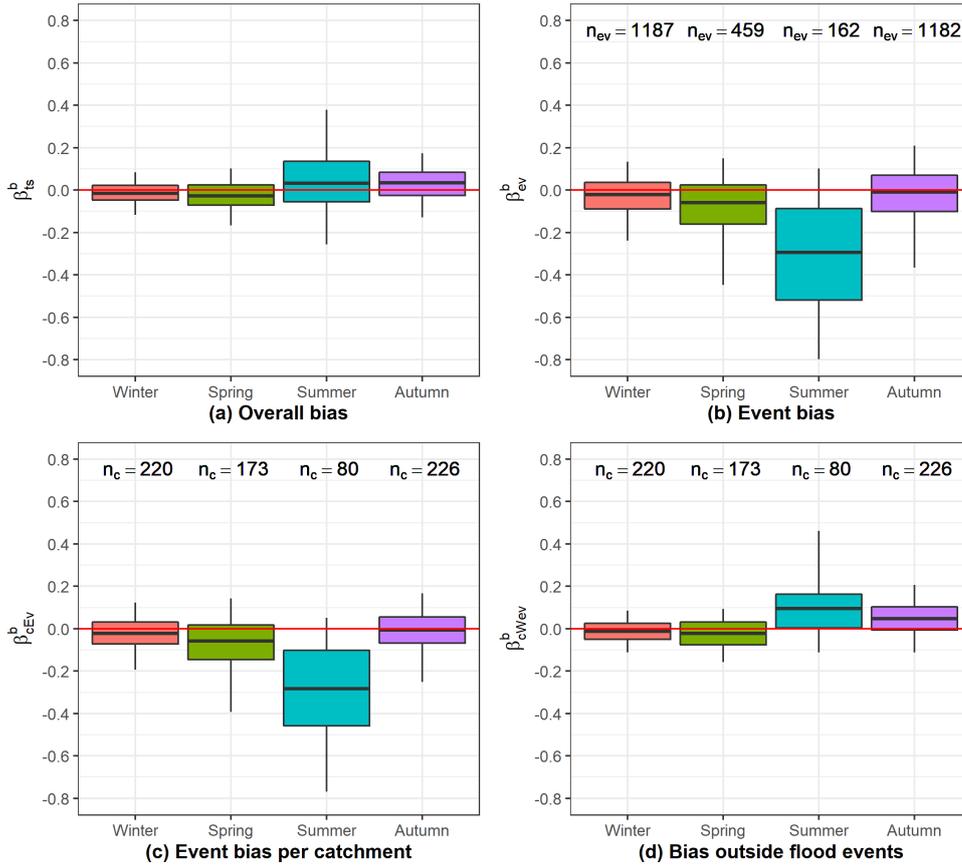


Figure 2.7 – Seasonality of GR5H bias over 229 catchments and 2,990 events. The bias (bounded) was calculated on four levels of hydrograph disaggregation. The distributions are presented between the 5th and 95th percentiles. Calculations were made with cross-validation values. n_c is the related number of catchments and n_{ev} is the related number of flood events. Part (c) shows the seasonal distributions for all catchments.

compared the four bias criteria with the seasonality of floods (Fig. 2.7a to 2.7d). Figure 2.7b shows that more than 75 % of the observed summer floods were underestimated by the GR5H model. The distribution of event bias shows that the underestimation associated with summer events is much larger than with the other flood events of our dataset. Floods that occur in winter and autumn were less underestimated by the model. The distribution of bias across summer events is wider than the bias calculated for the simulations of the other flood events. The largest overestimations of floods were for autumn events. Large underestimations of summer events explain the skewness of the distribution of GR5H event bias presented in Fig. 2.6. Figure 2.7c shows that the seasonal trend of model event bias is also an intra-catchment pattern. The distributions are less dispersed but the tendency remains similar. The seasonal distribution of model bias calculated in the entire hydrograph exhibits a different pattern (Fig. 2.7a). Summer and autumn bias distributions have positive median values. Winter and spring distributions are similar to the distributions in Fig. 2.7b and 2.7c but with a narrower dispersion. When the bias is calculated without flood events (Fig. 2.7d), winter and summer distributions remain similar to the time series bias. Summer and autumn periods are overestimated for 75 % of the corresponding catchments. These

results indicate compensation between periods of flood events and periods where there are no flood events in summer and spring. Event bias seems to affect winter and spring periods less in our dataset.

Figures 2.8a to 2.8h present the GR5H simulations of eight floods that occurred in tributaries of three major French rivers, the Rhône (Doux and Ardèche tributaries), Seine and Garonne rivers. The GR5H model failed to reproduce the volumes of the summer events (Fig. 2.8a, 2.8c, 2.8e, 2.8g), whereas it was able to reproduce the volumes of four floods that occurred in other seasons (Fig. 2.8b, 2.8d, 2.8f, 2.8h). The GR5H model yielded reasonable performance for the rest of the time series, as assessed by the KGE index.

2.5.4 Relationship between event bias and flood characteristics

Based on the 80 catchments where flood events occur in summer, we investigated the relationships between model event bias and various flood characteristics. Figure 2.9 presents the univariate analysis of the links between GR5H bias on floods and flood characteristics. Figure 2.9a shows that GR5H bias is lower for flood events for which the maximum hourly precipitation exceeds 130 % of the mean of maximum event precipitation. Figure 2.9b shows that there is no clear relationship between event peak flow and model bias.

Large underestimations of flood volumes are associated with low SWI index values (Fig. 2.9c) and short-duration events (Fig. 2.9e). Interestingly, low runoff coefficient values are associated with larger underestimations (Fig. 2.9d). As expected, the highest correlation between flood characteristics is between the SWI index and the runoff coefficient (Fig. 2.9f). The runoff coefficient is also positively correlated with flood duration and negatively correlated with maximum precipitation. These results indicate that the largest underestimations of flood volumes by the GR5H model are for short floods occurring in summer under low soil moisture conditions and when high-intensity precipitation events take place.

2.6 Discussion

Our analysis was based on bias calculations considering different periods of the hydrograph to evaluate the capacity of a conceptual rainfall-runoff model to simulate flood events. By refining the time window of analysis and by using the seasonality of streamflow water balance as a proxy of flood variability, we found patterns of seasonal model bias. These patterns of seasonal model bias are linked to soil moisture conditions and specific characteristics associated with the flood events. We now discuss how informative this analysis can be in light of other existing diagnostics.

2.6.1 How informative is the KGE index for summer events?

We have shown that calculating the model bias for different streamflow time windows enabled us to highlight compensations between high-flow periods and periods without floods as well as situations where the model was not able to simulate flood volumes. One may wonder whether these results could be obtained with aggregated statistics such as the KGE criterion. In other words, we wanted to know whether the catchments in which the model yielded low KGE values are the catchments in which the summer floods were underestimated. First, the comparison between overall bias and event bias showed differences in the ability of the GR5H model to simulate the global water balance and to simulate flood volumes. One

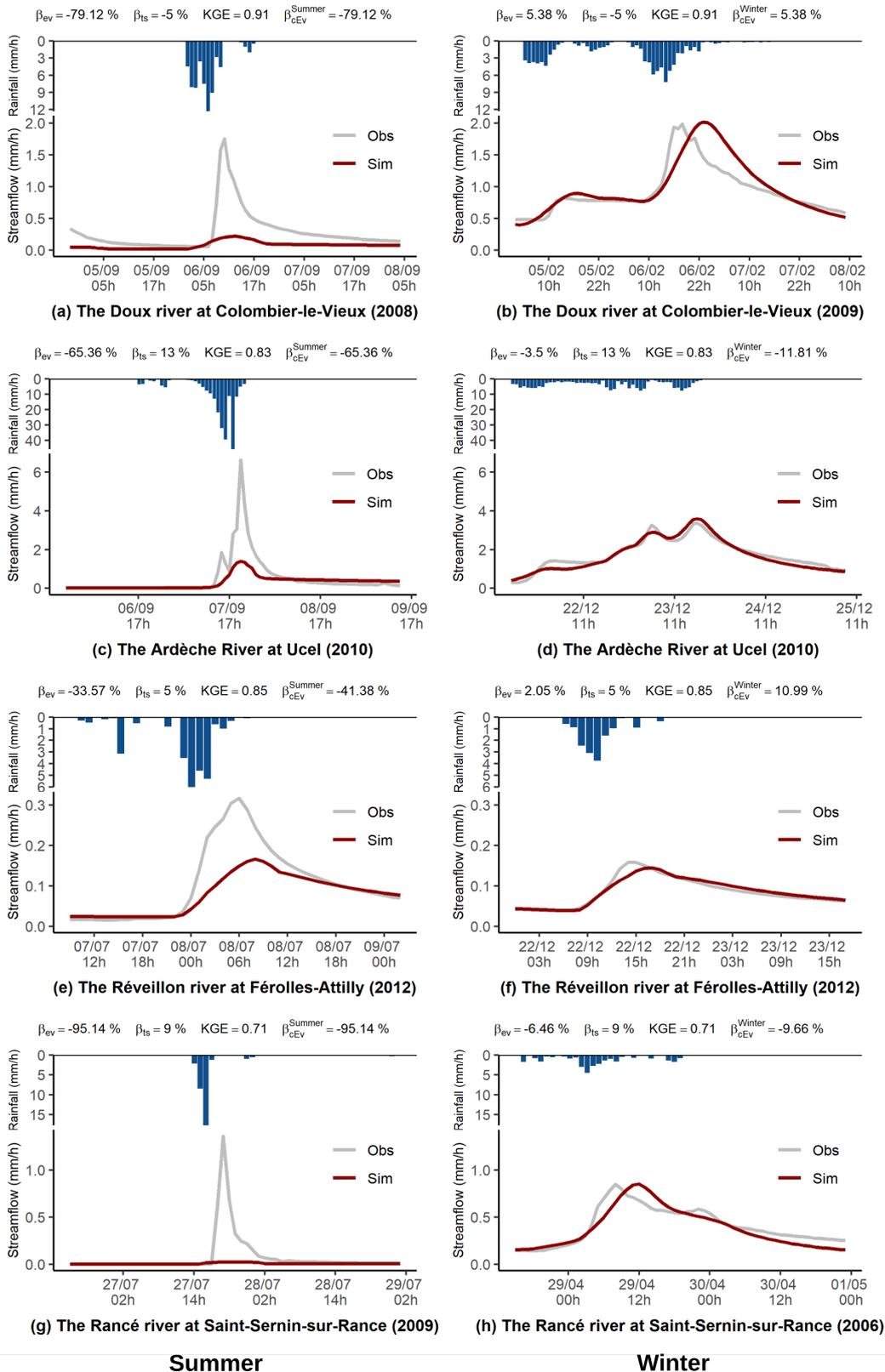


Figure 2.8 – Examples of simulations of summer floods compared with floods that occurred in other seasons. Simulations of GR5H are in validation mode. β_{ev} is the flood event bias. β_{ts} and KGE are the bias and KGE index, respectively, calculated for the entire corresponding validation period.

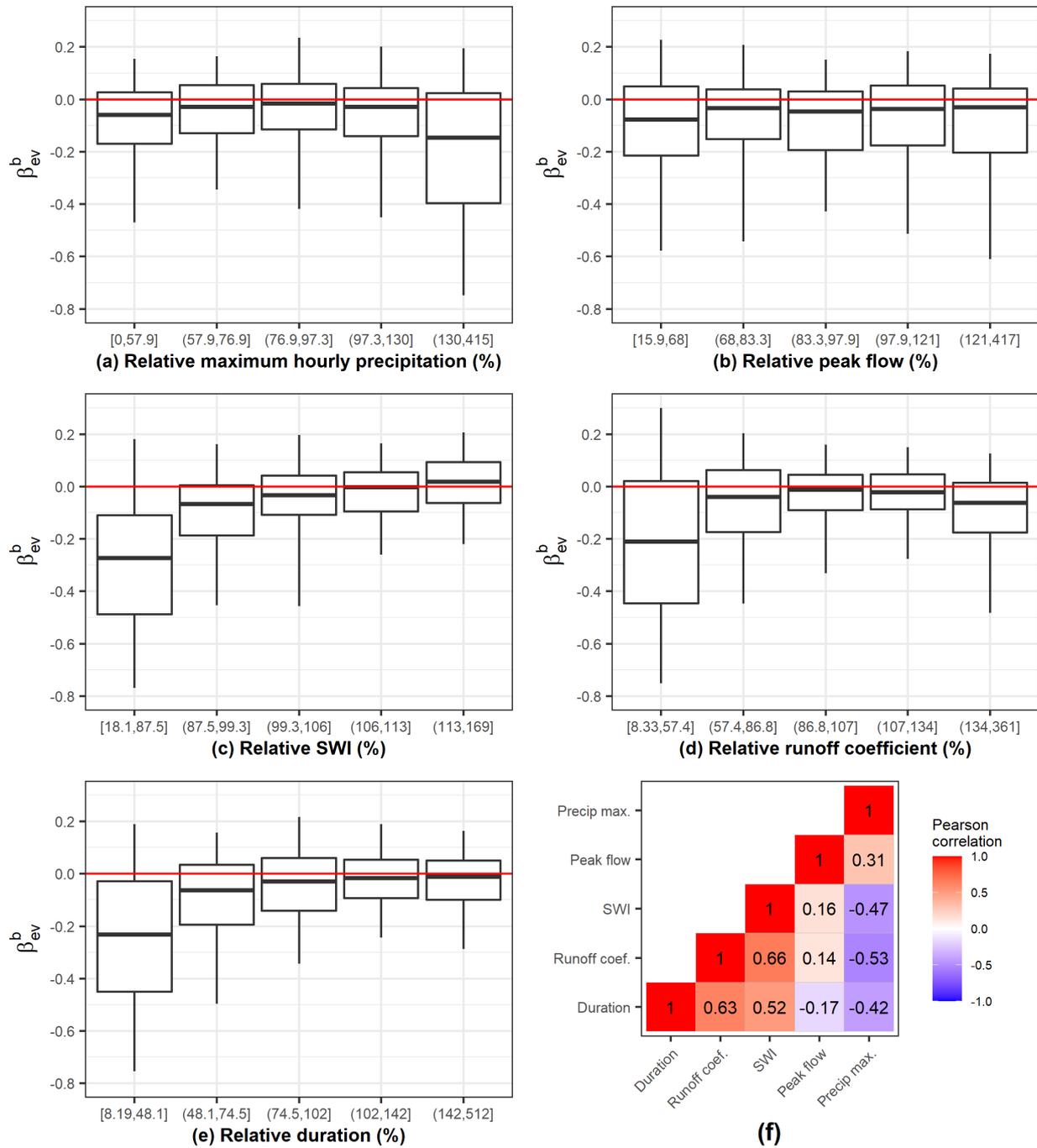


Figure 2.9 – (a) to (e) present the results of a univariate analysis of the relationships between flood characteristics and event bias β_{ev}^b for 80 catchments in which floods occur in summer. Each characteristic class contains approximately 211 events. The flood characteristics presented are relative to the mean flood characteristics of the corresponding catchment. (f) is the linear correlation matrix between the relative flood characteristics.

reason may be the low water volumes associated with summer events compared with the other floods in our dataset (see Fig. 2.1). Mizukami *et al.* (2019) reported similar results when investigating differences in annual peak flow simulations. They showed that improving streamflow bias did not always result in better simulation of flood magnitudes. Furthermore, the results presented in Fig. 2.10 demonstrate that there is no clear relationship between aggregated KGE values and event bias for catchments in which summer floods occur. This is consistent with the findings of Brunner *et al.* (2021a), who showed that while the KGE index integrates a flow bias component, it does not explicitly account for high-flow values. These results are also in line with the study by Mathevet *et al.* (2020), who found that two conceptual models yielded similar results when looking at criteria calculated for the whole time series but found differences in the ability of the model to capture short-term processes when investigating criteria calculated for specific sub-periods. These results show that investigating the seasonality of model bias considering different streamflow time windows can offer new information on the ability of a model to simulate specific flood events.

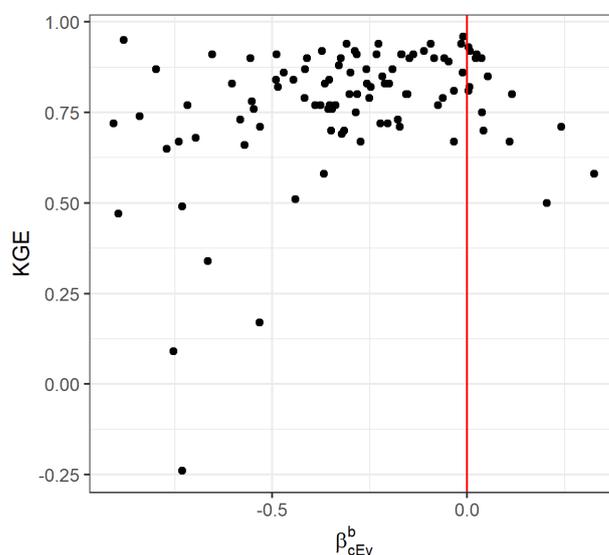


Figure 2.10 – GR5H summer event bias per catchment plotted against KGE values for catchments where flood events occurred in summer.

2.6.2 On the relevance of event and seasonality analyses to improve simulation of floods

To investigate seasonal patterns in model errors, other hydrological signatures could be based on the hydrological regime (or 365-d rolling mean, e.g., Mathevet *et al.*, 2020) or the flow duration curve (e.g., Yilmaz *et al.*, 2008). However, as presented in section 2.5.3, the seasonal model bias did not highlight underestimations by the model of flood events observed in summer. The investigation of specific events along with seasonality clearly enabled the identification of patterns in model errors. Another way to identify these patterns would be to derive the flow duration curve in summer periods and calculate differences above the 80th percentile. But the advantage of flood event selection and analysis lies in the possibility to investigate links to various flood characteristics. Floods can have very different seasonal triggers affecting antecedent soil moisture conditions at different time scales (Blöschl *et al.*, 2013; Berghuijs *et al.*, 2014). Therefore, fast processes are not always considered in overall

performance analyses. The seasonality of floods is a good indicator of flood triggers in metropolitan France. Our results showed that combining event analysis with streamflow seasonality analysis can help to identify where a conceptual model fails to reproduce flood volumes even when the model yields an aggregated performance considered to be reasonable.

2.6.3 First-order factors controlling simulation errors

One of the underlying ideas behind diagnosing a model is to be able to identify where model improvement is needed. In the case of the GR5H model, our results showed a clear seasonal pattern with an unbiased estimation of floods except in summer, where there is a clear underestimation signal. A central question remains regarding the first order factors leading to the identified deficiency. Model deficiency can be related to parameter estimation, either because of the choice of objective function (e.g., Mizukami *et al.*, 2019; Brunner *et al.*, 2021a) or because of data uncertainty (i.e., wrong parameterization because of uncertainty in forcings or streamflows; see e.g., Beven, 2016). However, structural deficiencies may be the underlying cause in the case of our model, as its parsimonious structure could limit its ability to reproduce specific processes occurring in summer, where short-duration processes can take place, such as high-intensity precipitation events. We have seen that this is especially the case under low soil moisture conditions. Furthermore, Fig. 2.11a indicates that the simulation of effective rainfall by the GR5H model partly explains the large underestimations of some of the flood events. When the simulation of effective rainfall is low compared with the other floods of a catchment, the model tends to underestimate the flood volume. Figure 2.11b presents the relationship between a measure of precipitation intensity and flashiness of the catchment response and GR5H simulation of event effective rainfall. Low values of effective rainfall are associated with short-duration events with high precipitation intensity (large y-axis values) and mostly for late spring, summer and early autumn events.

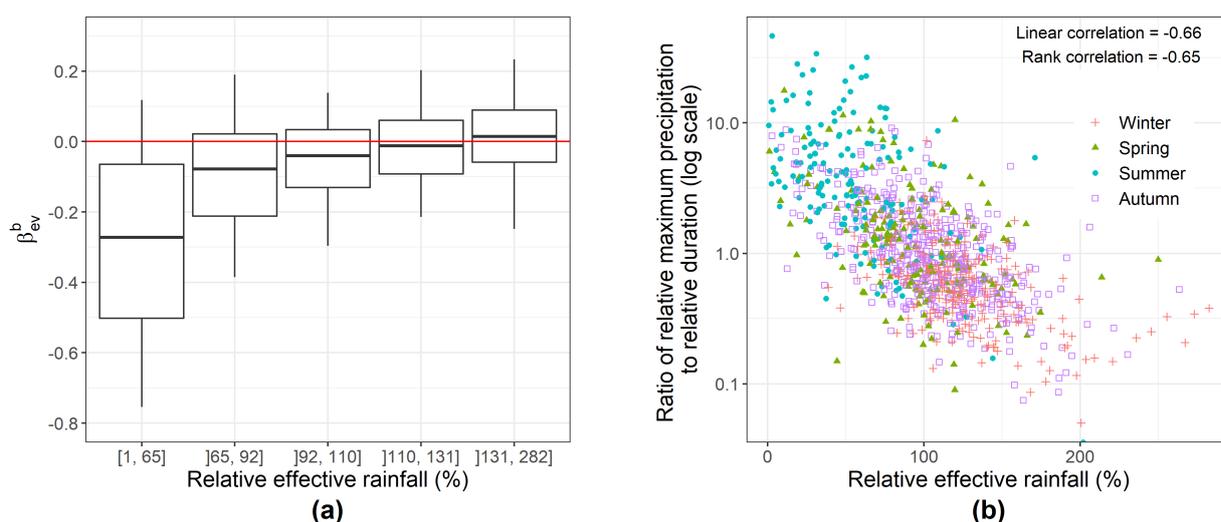


Figure 2.11 – (a) Relationship between GR5H simulation of event effective rainfall and event bias (each class contains approximatively 211 events). (b) Relationship between GR5H simulation of event effective rainfall and precipitation intensity relative to event duration. The results are for 1054 events that took place in 80 catchments in which floods can occur in summer.

Whatever parameter set is chosen, models sometimes do not have the structural ability to reproduce an observed flood. This explanation is consistent with the findings of Mathevet *et al.* (2020), who reported that the shorter the sub-period of evaluation, the greater the difference in performance between the MRX model and the GRX model (similar to GR5H). They concluded that MRX is better able to reproduce short-term processes than GRX because of the more complex structure of the former. The lumped spatial configuration could also be the restraining factor explaining the deficiencies identified, since spatial interactions between precipitations and soil wetness have impacts on flood generation (Tarasova *et al.*, 2020). Short-duration events with high precipitation intensity are sometimes subject to large uncertainties in the input precipitation data (e.g., Zoccatelli *et al.*, 2011; Ruiz-Villanueva *et al.*, 2012; Zhang *et al.*, 2017). If the total amount of rainfall is underestimated during an event, the model cannot simulate the required amount of effective rainfall. This could explain, to some extent, that the model does not perform well on such events. Whatever the first-order factor is, the present methodology, by establishing a finer benchmark for high-flow simulations, helped reveal the difficulties involved in modelling specific flood events. This is particularly relevant to reduce model structure uncertainty and therefore to improve the predictive capability of our model. This work will contribute to future improvements of the GR5H model.

2.7 Conclusion

In this study, we proposed an approach to diagnose the quality of floods simulated by a conceptual rainfall-runoff model over a large set of catchments and flood events. Starting from overall performance criteria calculated for the whole time series, our methodology consisted in computing model bias for selected flood events and looking for seasonal trends and compensations between streamflow periods. From these seasonal patterns, we aimed to link the model bias to certain flood characteristics. We found that while the model yielded reasonable performance for the dataset in terms of aggregated statistics, patterns in model errors were revealed when investigating performance across selected flood events. Using seasonality as an indicator of the variability of flood characteristics enabled us to identify situations where the GR5H model did not manage to reproduce the observed flood volumes. We found that the summer events of our dataset were associated with systematic underestimations by the GR5H model. Short-duration processes, such as high-intensity precipitation events, associated with low soil moisture conditions are not explicitly taken into account in the structure of the GR5H model. For these specific events, this results in a too-low simulation of effective rainfall and therefore underestimation of flood volumes. However, these simulation errors were not detected when considering the KGE index calculated for the entire streamflow time series. By examining both the seasonality of model bias and different streamflow time windows, we have identified compensations between flood events and the rest of the hydrograph in summer and autumn.

Overall, this study confirmed the limitations of using criteria computed on the whole time series to evaluate model performance in simulating high-flow events, even when these criteria are known to be biased towards high flows. The seasonality of streamflow was found to be an indicator of the ability of the GR5H model to simulate specific floods that occurred in some of the catchments of our dataset. It enabled us to refine the analyses and look for links to flood characteristics for these catchments. Future developments of the model could focus on improving the calculation of effective rainfall by accounting for summer flood

generating processes. This could be achieved by better considering the spatial variability of rainfall (e.g., Lobligeois *et al.*, 2014; Loritz *et al.*, 2021) or by taking precipitation intensity into account (e.g., Peredo *et al.*, 2022). Multi-objective calibration could also lead to an improved identification of the parameter sets (e.g., Monteil *et al.*, 2020), using, for example, objective functions that were found to be more suited for flood simulation (e.g., Mizukami *et al.*, 2019).

In this study, we focused on model bias because we considered it a first order property of model errors. Another perspective on our work could be to further analyze model simulations within a multicriteria assessment framework so as to cover more aspects of model performance (Willems, 2009). Also, data uncertainty was not considered for model parameterization or for streamflow result analyses, although it can have an impact on model parametrization and interpretation of model errors (Beven, 2016). Finally, further tests on other catchments, such as dry or arid catchments (e.g., in Australia), could improve the generalizability of our results.

Chapitre 3

Influence des caractéristiques de la pluie sur les erreurs du modèle

3.1 Contexte

Ce chapitre poursuit le diagnostic du modèle de simulation du chapitre précédent. Nous le faisons tout d'abord en caractérisant les différences de performance en crue entre bassins. Nous étudions ensuite les caractéristiques des événements de pluie à l'aide d'indicateurs spécifiques portant sur leur intensité et leur variabilité spatiale. Le but est de séparer les situations où la variabilité spatiale de la pluie explique une partie des erreurs du modèle de celles où c'est l'intensité de la pluie qui semble être un facteur de premier ordre. A minima, on souhaite regarder s'il y a des bassins ou des événements (ou les deux) pour lesquels la variabilité spatiale de la pluie joue un rôle plus important que l'intensité de la pluie, et vice versa. Nous utilisons les mêmes bassins versants que dans le chapitre précédent mais avec une période temporelle élargie (cf. chapitre 1).

3.2 Variabilité événementielle par type de bassin

Notre approche de modélisation est basée sur l'utilisation d'une structure fixe pour tous les bassins de notre base de données. Cette structure repose sur celle des modèles GR qui ont fait l'objet de nombreux travaux scientifiques démontrant leur robustesse mais aussi leurs limites (e.g. Perrin *et al.*, 2003 ; Le Moine *et al.*, 2007 ; Coron *et al.*, 2012 ; Ficchi *et al.*, 2016). Une des difficultés rencontrées lors des diagnostics et améliorations de la structure des modèles GR est d'identifier pour quels types de bassins une structure semble plus adaptée qu'une autre (e.g. Perrin, 2000 ; Pushpalatha *et al.*, 2011 ; van Esse *et al.*, 2013). Jusqu'ici, nous avons présenté des analyses majoritairement basées sur la typologie des événements de crue en laissant de côté la variabilité inter-bassins. L'objectif de l'analyse qui suit est de déterminer si l'erreur événementielle de GR5H dépend du comportement hydrologique des bassins versants. Le but final sera à nouveau de déterminer si les modifications structurelles que nous allons tester par la suite sont bien adaptées et identifiables.

3.2.1 Éléments méthodologiques

Afin de simplifier les analyses, nous distinguons les bassins versants dits « méditerranéens » des autres bassins de notre base de données (fig. 3.1). Cette distinction a été effectuée en fonction de la latitude des exutoires, et diffère de la classification présentée dans le chapitre 1. Ces bassins sont connus pour être fréquemment sujets à des épisodes de pluie très intenses. Nous avons sélectionné des affluents de plusieurs fleuves côtiers méditerranéens et quelques affluents du Rhône.

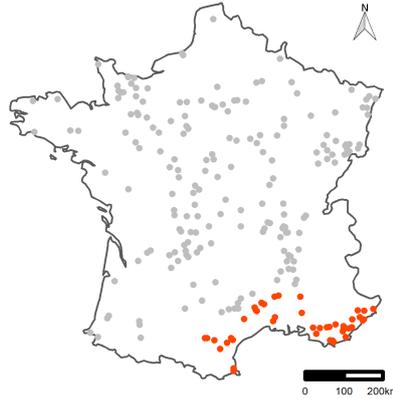


FIGURE 3.1 – Localisation des 43 bassins versants méditerranéens (points orangés).

Deux choix méthodologiques peuvent être faits afin de distinguer la variabilité entre bassins de la variabilité entre crues : attribuer une caractéristique de bassin à chaque événement de crue (e.g. bassin méditerranéen) ou attribuer une caractéristique moyenne des crues à chaque bassin versant (e.g. moyenne des intensités de pluie maximales d'un bassin). La deuxième solution, que nous choisissons ici, permet d'explorer plus simplement les tendances spatiales.

Dans le chapitre précédent (voir section 2.4.2), les performances événementielles du modèle ont été caractérisées au moyen d'un critère de biais (borné). Cela nous a fourni une première indication de la capacité du modèle à reproduire les crues. Cependant, d'autres caractéristiques des hydrogrammes de crue peuvent être explorées. Afin de compléter les analyses, le critère de Nash et Sutcliffe (NSE ; Nash & Sutcliffe, 1970) est calculé pour chaque événement de crue j , avec comme modèle « naïf » la moyenne des débits sur la période considérée (P1 ou P2) plutôt que la moyenne des débits de l'événement.

$$NSE_{ev,j} = 1 - \frac{\sum_{i=1}^{n_j} (Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^{n_j} (\overline{Q_{obs}} - Q_{obs,i})^2} \in] - \infty; 1] \quad (3.1)$$

n_j est la durée de l'événement. $Q_{sim,i}$ est le débit simulé au temps i . $Q_{obs,i}$ est le débit observé au temps i . $\overline{Q_{obs}}$ est la moyenne des débits de la période et du bassin considérés. Pour faciliter les comparaisons, nous utilisons une version bornée entre -1 et 1 de ce critère (C2M ; Mathevet *et al.*, 2006).

$$C2M_{ev,j} = \frac{NSE_{ev,j}}{2 - NSE_{ev,j}} \quad (3.2)$$

L'utilisation de cette version bornée du NSE permet également de faciliter la représentation des résultats car les valeurs de NSE qui sont inférieures à -1 tirent les distributions vers des valeurs très négatives. Pour synthétiser les résultats, les événements de crue ont été séparés

en deux groupes : les événements dont le pic de crue a lieu entre novembre et avril (désigné ici par « hiver ») et les événements dont le pic de crue a lieu entre mai et octobre (désigné ici par « été »).

3.2.2 Résultats

Les résultats présentés en figure 3.2 indiquent que les performances globales de GR5H sont plus faibles pour les bassins méditerranéens que pour les autres bassins de la base de données, en calage et en validation et pour les deux sous-périodes. Par ailleurs, on observe une différence de performances plus marquée entre les sous-périodes de calage pour ces bassins. La répartition spatiale des performances en validation indique que le modèle présente aussi de moins bonnes performances sur certains affluents de la Seine ainsi que sur plusieurs affluents du Rhône. Néanmoins, les performances du modèle ne sont pas dégradées pour tous les bassins méditerranéens.

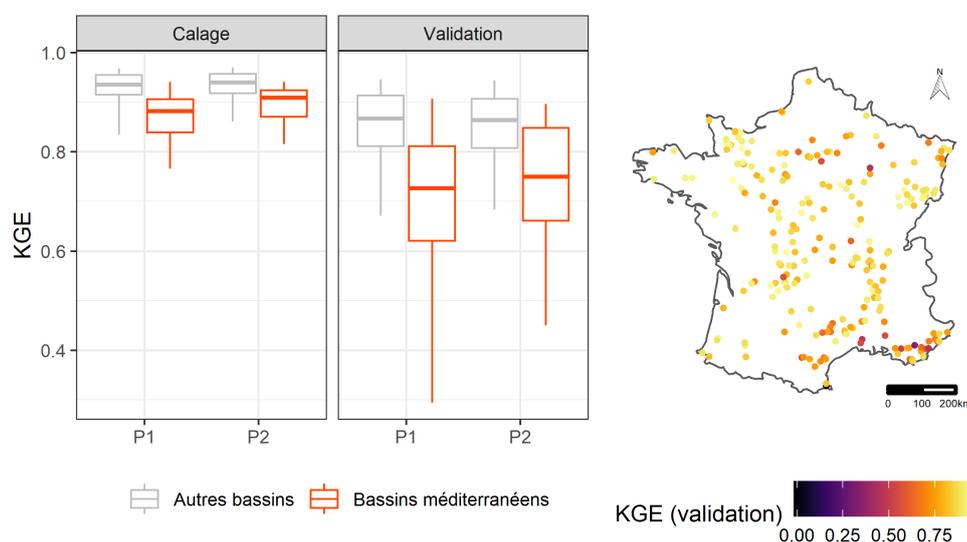


FIGURE 3.2 – Répartition des performances globales du modèle GR5H pour deux groupes de bassins versants. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Les résultats présentés sur la figure 3.3 confirment ceux obtenus au chapitre précédent : les performances événementielles du modèle GR5H sont nettement moins élevées en été qu'en hiver. En particulier, on observe un fort signal de sous-estimation du volume de crue en été. Ce signal est encore plus marqué pour les bassins méditerranéens : les cartes de répartition spatiale des performances événementielles montrent que les sous-estimations les plus fortes sont localisées dans le sud de la France. Certains affluents du bassin parisien présentent toutefois des performances dégradées en été. Plus globalement, les résultats montrent que les crues d'été sont moins bien modélisées pour la majorité des bassins de notre base de données.

L'objectif de cette analyse était de compléter le diagnostic du modèle GR5H en distinguant les tendances observées selon deux groupes de bassins versants. Nous avons vu que les performances du modèle sont moins élevées pour les bassins méditerranéens, avec un déficit du volume en crue plus notable en été. Ces crues d'été semblent être les plus difficiles à

modéliser de notre base de données. De nombreux travaux font état de la difficulté de modéliser ces événements qui sont souvent associés à des épisodes de pluie très intenses et/ou très variables dans l'espace (e.g. Douinot *et al.*, 2016).

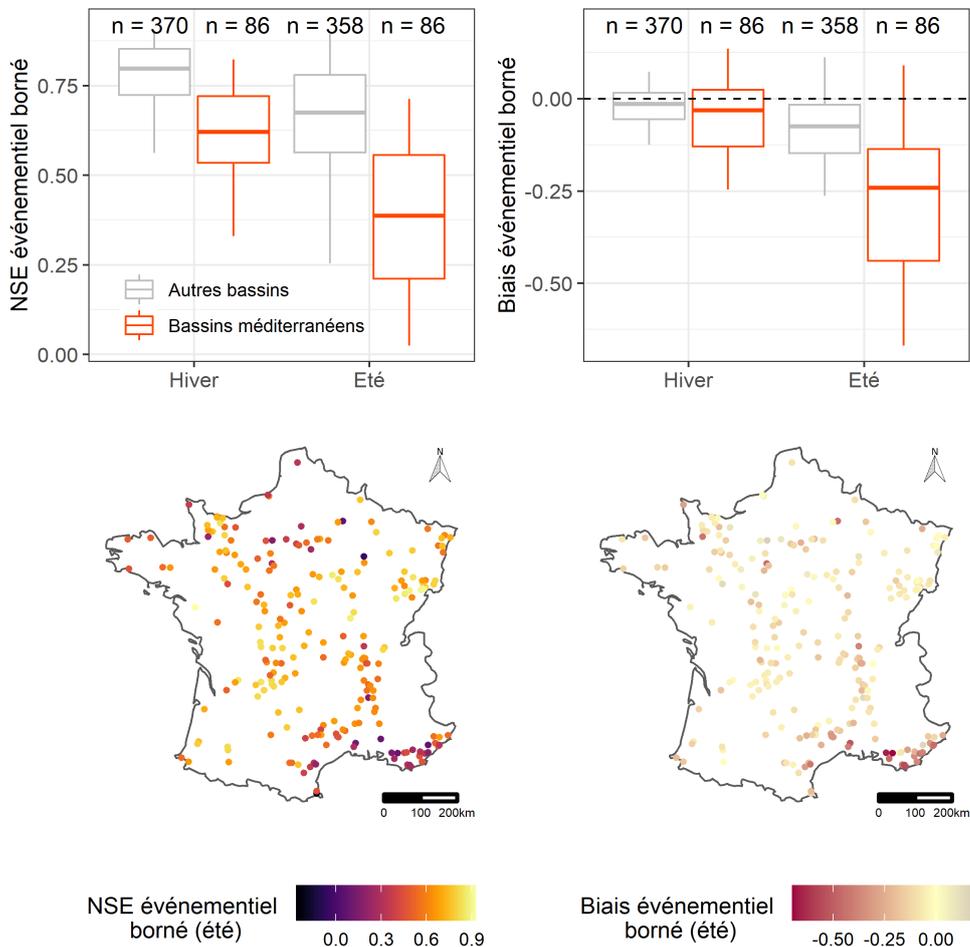


FIGURE 3.3 – Répartition des performances événementielles du modèle GR5H selon deux groupes de bassins versants et deux types de crues. Les résultats sont présentés en validation croisée. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. Les valeurs des critères événementiels présentés sur les cartes correspondent à une moyenne par bassin entre les valeurs des périodes P1 et P2.

3.3 Caractéristiques des événements de pluie et erreurs du modèle

Les propriétés spatio-temporelles de la pluie jouent un rôle dans la réponse hydrologique des bassins versants (e.g. Zocatelli *et al.*, 2015). Si l'événement est localisé à un endroit précis du bassin, s'il s'est déplacé au cours de l'événement ou a arrosé uniformément le bassin, la réaction du bassin sera différente. Ces caractéristiques peuvent avoir un impact sur les performances des modèles hydrologiques (e.g. Lobligeois *et al.*, 2014). Ces liens ont été très largement étudiés dans la littérature (e.g. Lobligeois *et al.*, 2014 ; Loritz *et al.*, 2021), par

exemple en utilisant les *spatial moments of rainfall* qui permettent d'étudier la temporalité de la réponse d'un bassin à des pluies variables dans l'espace (e.g. Zoccatelli *et al.*, 2011 ; Giani *et al.*, 2022). Par ailleurs, les crues rapides provoquées par des événements convectifs sont connues pour être difficiles à modéliser (e.g. Ruiz-Villanueva *et al.*, 2012).

3.3.1 Caractérisation des événements de pluie

Afin de caractériser les propriétés spatiales des événements de pluie qui ont donné lieu aux crues sélectionnées sur notre base de données, nous utilisons plusieurs indices calculés à partir des cumuls horaires spatialisés de résolution 1 km². Ces calculs s'effectuent en utilisant le cumul de pluie horaire sur chaque maille qui recouvre le bassin considéré, et cela pendant l'événement considéré. Les indicateurs utilisés sont les suivants :

- L'indice de Smith (Smith *et al.*, 2004), en mm/h, renseigne sur la variabilité spatiale de la pluie. Lobligeois *et al.* (2014) le calculent de la manière suivante :

$$smithP = \frac{\sum_{t=1}^{n_t} [(\sigma(P_{x,t})_{1 \leq x \leq n_x}) \cdot (\mu(P_{x,t})_{1 \leq x \leq n_x})]}{\sum_{t=1}^{n_t} [\mu(P_{x,t})_{1 \leq x \leq n_x}]} \in [0; +\infty[\quad (3.3)$$

avec $P_{x,t}$ la pluie du pixel x au temps t , n_t la durée de l'événement de pluie (en heures), n_x le nombre de pixels associés au bassin versant, μ la moyenne et σ l'écart-type. Plus la moyenne de la variabilité spatiale à chaque pas de temps est élevée, plus $smithP$ est élevé.

- Un indice de répartition spatiale des cumuls de l'événement. Cet indice permet de déterminer si, sur la durée d'un événement, certaines parties du bassin sont plus arrosées que d'autres et si cela représente une grande ou faible proportion du bassin. Il dérive d'un indice proposé par Tarasova *et al.* (2020) et s'exprime de la manière suivante :

$$spatP = \frac{\sigma(\sum_{t=1}^{n_t} P_{x,t})_{1 \leq x \leq n_x}}{\sum_{t=1}^{n_t} [\mu(P_{x,t})_{1 \leq x \leq n_x}]} \in [0; +\infty[\quad (3.4)$$

Plus $spatP$ est élevé, plus la variabilité spatiale des cumuls d'un événement est élevée. À la différence de $smithP$, $spatP$ ne considère pas la variabilité spatiale à chaque pas de temps.

Deux indicateurs sur l'intensité des pluies à l'échelle du bassin sont utilisés dans cette partie :

- L'intensité maximale de la pluie horaire de l'événement, en mm/h :

$$maxP = \max(\mu(P_{x,t})_{1 \leq x \leq n_x})_{1 \leq t \leq n_t} \in [0; +\infty[\quad (3.5)$$

- L'intensité moyenne de la pluie de l'événement, notée $intP$ (mm/h). Cet indice est défini comme le ratio entre le cumul de pluie de l'événement et sa durée :

$$intP = \frac{1}{n_t} \sum_{t=1}^{n_t} [\mu(P_{x,t})_{1 \leq x \leq n_x}] \in [0; +\infty[\quad (3.6)$$

Afin de faciliter les comparaisons des résultats présentés dans les sections 3.3.2 et 3.3.4, une normalisation « min-max » a été appliquée sur l'espace des 10 652 événements :

$$ind_{norm} = \frac{ind - \min(ind)}{\max(ind) - \min(ind)} \in [0; 1] \quad (3.7)$$

avec *ind* l'indice considéré. Dans la section 3.3.4, l'indice du SWI (défini dans la section 2.4.3) est calculé au pas de temps précédant le début de l'événement au lieu d'être moyenné sur toute sa durée.

Le schéma de la figure 3.4 montre la différence d'information apportée par les deux indices de variabilité spatiale de la pluie (*smithP* et *spatP*). Lorsque l'événement de pluie est réparti sur tout le bassin à chaque pas de temps, les deux indices ont des valeurs proches de 0. Lorsque la pluie est localisée sur la même partie du bassin durant toute la durée de l'événement, les deux indices ont des valeurs élevées. Lorsque la pluie est localisée à différents endroits au cours de l'événement, l'indice *smithP* est élevé alors que l'indice de répartition spatiale des cumuls *spatP* prend une valeur proche de 0.

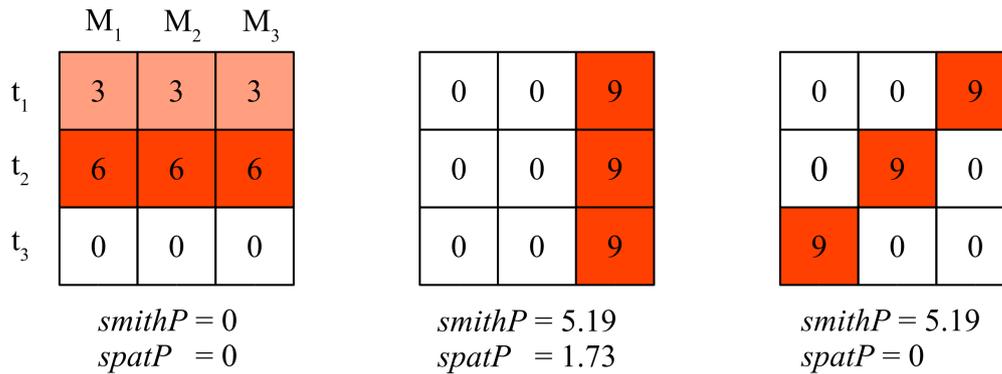


FIGURE 3.4 – Valeurs des indices de variabilité spatiale pour trois hyétogrammes synthétiques. Les valeurs des cumuls de pluie (mm) par pas de temps (t_1 , t_2 et t_3) et par maille (M_1 , M_2 et M_3) sont indiquées dans chaque carré. *smithP* : indice de Smith (mm/h). *spatP* : indice de répartition spatiale des cumuls de pluie.

Cette distinction nous paraît importante pour séparer une information sur la temporalité de la pluie d'une information sur sa répartition spatiale. Le comportement des bassins versants ainsi que les performances du modèle pourraient s'avérer différents selon que l'événement de pluie est localisé à un endroit précis du bassin versant (ou à plusieurs endroits mais le tout ne représentant qu'une faible portion du bassin) ou si l'événement s'est déplacé au cours du temps. Un événement de pluie très localisé donnerait potentiellement lieu à des saturations locales du bassin, engendrant ainsi des comportements plus réactifs pour un cumul de pluie donné alors qu'un événement fortement variable dans le temps et dans l'espace pourrait donner lieu à des réactions moins fortes pour une même intensité de pluie. La plupart des événements de pluie sont probablement caractérisés par cette double variabilité, mais l'indice de Smith ne permet pas cette distinction. Par ailleurs, nous verrons plus loin qu'il recoupe des informations données par les indicateurs d'intensité des pluies. L'indice de variabilité des cumuls de pluie permet de garantir que l'information donnée sur la variabilité spatiale est en partie indépendante de la répartition temporelle de l'événement.

3.3.2 Répartition des indices sur notre base de données

La figure 3.5 présente la répartition spatiale de ces indices sur notre base de données. Les quatre indices ont été moyennés par saison et par bassin. On observe que les événements les plus variables spatialement sont localisés sur des bassins méditerranéens ainsi que sur le bassin du Rhône. Ces événements ont majoritairement lieu en été. Ce résultat était attendu car

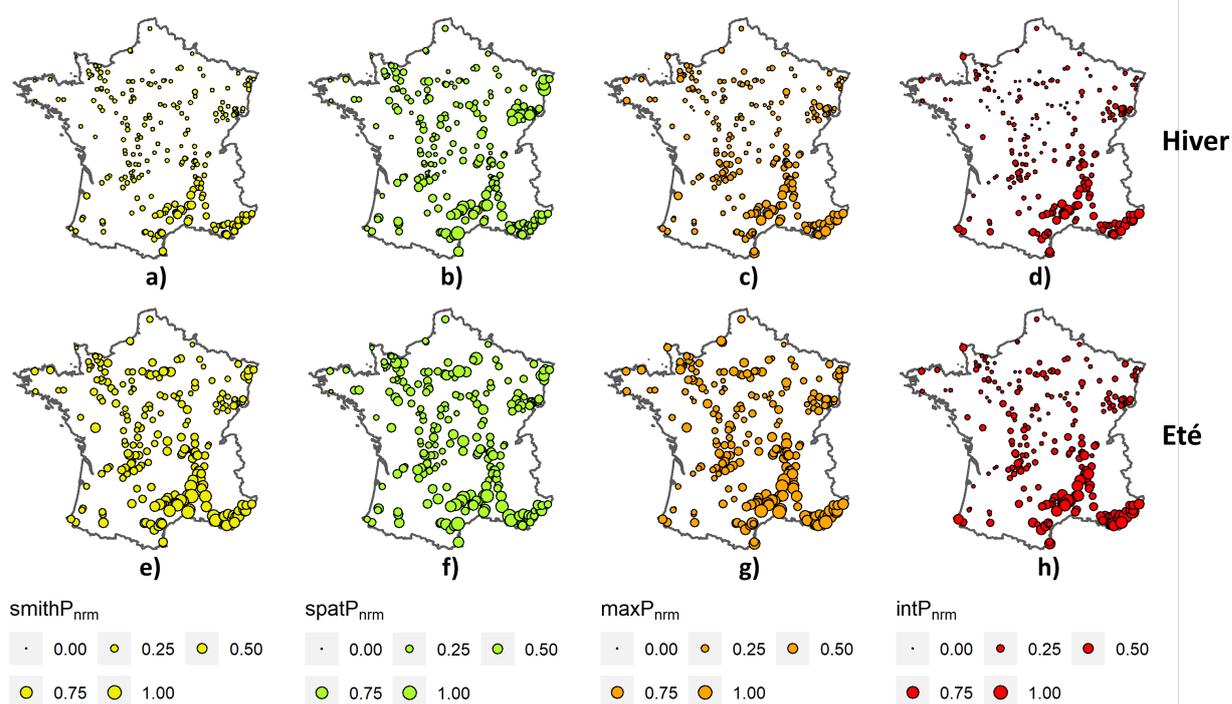


FIGURE 3.5 – Répartition spatiale et saisonnière de quatre indices de pluie calculés sur 10 652 événements de 229 bassins versants français. Les indices ont été moyennés par bassin et par saison puis normalisés. $smithP_{nrm}$: indice de Smith normalisé. $spatP_{nrm}$: indice normalisé de répartition spatiale des cumuls de pluie. $maxP_{nrm}$: intensité de pluie maximale normalisée. $intP_{nrm}$: intensité de pluie normalisée. Les échelles de taille de points sont continues.

la région Cévennes-Vivarais-Méditerranée est connue pour ses épisodes de fortes précipitations (e.g. Gaume *et al.*, 2009 ; Boudevillain *et al.*, 2011 ; Douinot *et al.*, 2016). Cependant, en été, plusieurs bassins répartis sur toute la France présentent des valeurs élevées de $smithP_{nrm}$ et de $spatP_{nrm}$. En hiver, les valeurs de $smithP_{nrm}$ sont faibles pour la majorité des bassins bien que plus élevées pour les bassins méditerranéens. La répartition des valeurs de $spatP_{nrm}$ suit l'évolution saisonnière et spatiale des valeurs de $smithP_{nrm}$. Les disparités entre bassins sont moins importantes que pour $smithP_{nrm}$. Plusieurs bassins de l'est de la France présentent des valeurs élevées de cet indice, en hiver comme en été. L'indice $maxP_{nrm}$ suit la répartition spatiale et saisonnière des valeurs de $smithP_{nrm}$. Les valeurs les plus fortes sont pour des bassins du sud-est de la France et majoritairement en été. L'indice $intP_{nrm}$ suit une répartition proche de $maxP_{nrm}$ à l'exception des bassins non-méditerranéens en été. La durée des événements concernés est plus élevée sur ces bassins malgré la présence de fortes intensités horaires. Comme pour les indices de variabilité spatiale, plusieurs bassins répartis sur toute la France présentent des valeurs élevées d'intensité de pluie en été.

3.3.3 Séparer la variabilité spatiale de l'intensité de la pluie

La figure 3.6 présente les corrélations linéaires et les corrélations de rang des quatre indices de pluie. On observe des valeurs élevées de corrélations entre les indices. En particulier, l'intensité horaire maximale des événements est fortement corrélée avec l'indice de variabilité

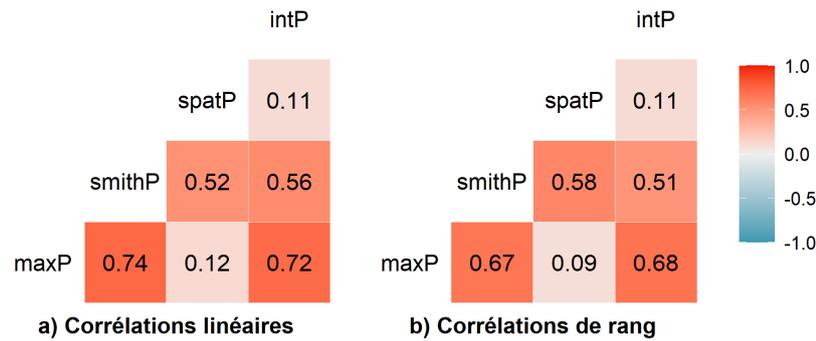


FIGURE 3.6 – Corrélations linéaires (a) et de rang (b) entre les quatre indices caractérisant les 10 652 événements de pluie. *smithP* : indice de Smith. *spatP* : indice de répartition spatiale des cumuls de pluie. *maxP* : intensité de pluie maximale. *intP* : intensité de pluie.

spatiale de Smith. L'indice de variabilité spatiale des cumuls de pluie est faiblement corrélé avec l'intensité de la pluie de l'événement. Cela pourrait indiquer que certains événements de pluie sont caractérisés par de fortes intensités réparties sur tout le bassin lorsque l'on considère toute la durée de l'événement. La différence avec l'indice de Smith est que la variabilité spatiale n'est pas considérée à chaque pas de temps.

On cherche maintenant à séparer les bassins pour lesquels la variabilité spatiale joue un rôle majeur en été, des bassins où c'est plutôt l'intensité de la pluie. Au vu des corrélations de la figure 3.6, nous regardons les bassins où l'indice de répartition spatiale des cumuls donne des valeurs faibles mais où l'indice d'intensité de la pluie de l'événement présente des valeurs élevées et vice versa (fig. 3.7). On observe que les bassins pour lesquels la variabilité spatiale

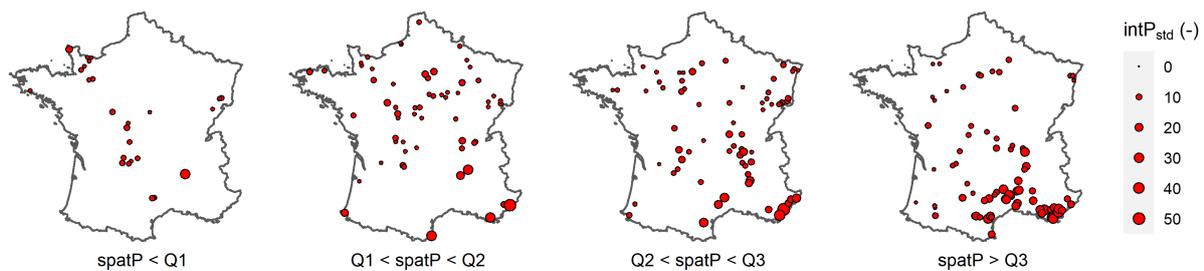


FIGURE 3.7 – Répartition spatiale de l'intensité de la pluie moyennée par bassin en été (mai à octobre) pour quatre gammes de valeurs (à partir des quartiles Q1 à Q3) de la variabilité spatiale des cumuls de pluie *spatP*. $intP_{std}$ est l'intensité de la pluie rapportée au cumul de pluie annuel du bassin. Le cumul de pluie annuel du bassin a été rapporté à une valeur horaire pour obtenir un indice sans dimension.

des cumuls est élevée et l'intensité est forte sont majoritairement localisés dans le sud-est de la France. Par ailleurs, on observe que sur plusieurs bassins non méditerranéens, il y a des événements avec de fortes intensités de pluie, mais répartis sur tout le bassin. Enfin, on observe des valeurs intermédiaires d'intensité de pluie associées à des valeurs intermédiaires de variabilité spatiale de la pluie, cela pour plusieurs bassins versants. Ces observations sont cohérentes avec les faibles valeurs de corrélation obtenues entre *intP* et *spatP* (fig. 3.6).

3.3.4 Liens avec les performances du modèle GR5H en crue

L'analyse qui suit a pour objectif d'explorer les liens apparents entre les caractéristiques des événements de pluie et les erreurs du modèle en crue. Cela permettra d'une part de mieux cibler les modifications à apporter au modèle et d'autre part de mieux identifier les cas où les modifications testées améliorent ou dégradent le modèle.

La figure 3.8 présente une analyse univariée du biais et du NSE en crue en fonction de deux des quatre indices de pluie.

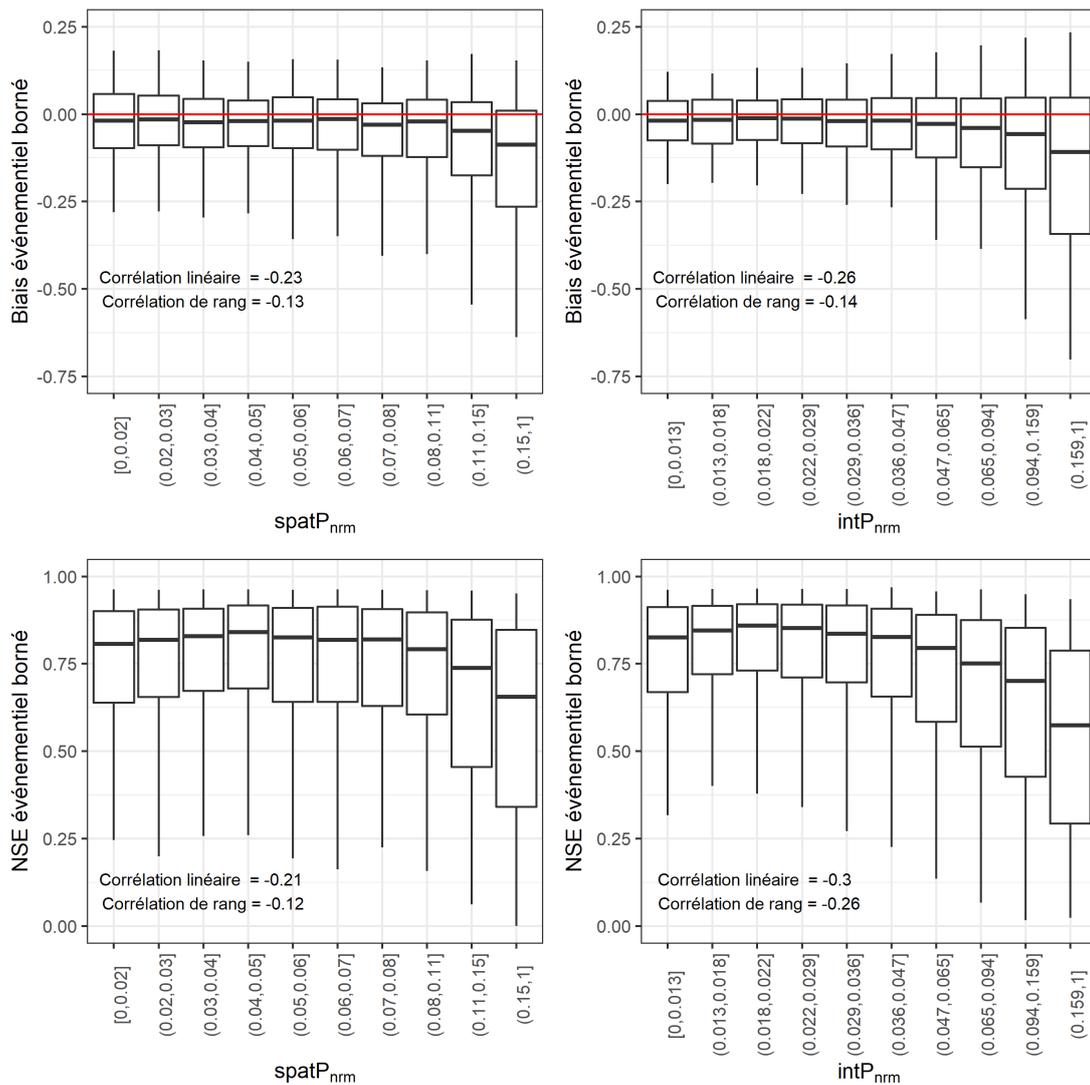


FIGURE 3.8 – Analyse univari\u00e9e du lien entre le biais et le NSE en crue du mod\u00e8le GR5H et deux caract\u00e9ristiques des \u00e9v\u00e9nements de pluie (validation crois\u00e9e). Chaque bo\u00eete \u00e0 moustache contient le m\u00eame nombre de valeurs. Chaque valeur est associ\u00e9e \u00e0 un \u00e9v\u00e9nement de pluie. Les 10 652 \u00e9v\u00e9nements sont consid\u00e9r\u00e9s dans cette analyse. Les bo\u00eetes \u00e0 moustaches sont trac\u00e9es du quantile 5 % au quantile 95 %.

Les valeurs de corr\u00e9lation entre les indices de pluie et les deux crit\u00e8res d'erreur ne sont pas tr\u00e8s \u00e9lev\u00e9es, en particulier les valeurs de corr\u00e9lation de rang. Cependant, pour les deux indices, les valeurs les plus \u00e9lev\u00e9es correspondent aux valeurs les plus fortement n\u00e9gatives du biais \u00e9v\u00e9nementiel et les valeurs les plus faibles du NSE, c'est-\u00e0-dire aux sous-estimations

les plus fortes par le modèle. Ces résultats confirment ceux du chapitre 2 : le modèle sous-estime plus fortement les événements de crue associés à des événements de pluie intenses et présentant une forte variabilité spatiale. Les valeurs du NSE associées aux faibles valeurs de $spatP_{nrm}$ et de $intP_{nrm}$ sont légèrement plus faibles que pour les gammes de valeurs intermédiaires de ces deux indices.

On peut maintenant se poser une question liée à la partie précédente : les sous-estimations du modèle ont-elles lieu lorsque les événements de pluie sont à la fois intenses et variables spatialement ? Ou bien lorsque les événements de pluie sont soit intenses soit variables spatialement ? Un premier élément de réponse peut-être obtenu à partir des résultats présentés dans la section précédente (voir 3.3.3). Il y a des événements et des bassins pour lesquels l'indice de variabilité spatiale des cumuls est faible mais pour lesquels l'intensité de pluie est élevée et vice versa. Il existe aussi des cas où les deux indices présentent des valeurs élevées. On retrouve cette tendance lorsque l'on s'intéresse aux liens entre intensité, variabilité spatiale et erreur du modèle en crue (fig. 3.9a et 3.9b). On observe que les événements de crue sont plus fortement sous-estimés lorsqu'ils sont associés à des événements de pluie intenses ou variables spatialement. Le biais du modèle est encore plus important lorsque les événements sont à la fois intenses et variables dans l'espace. On observe la même tendance pour le critère NSE qui fournit une information plus complète sur l'erreur du modèle que le biais.

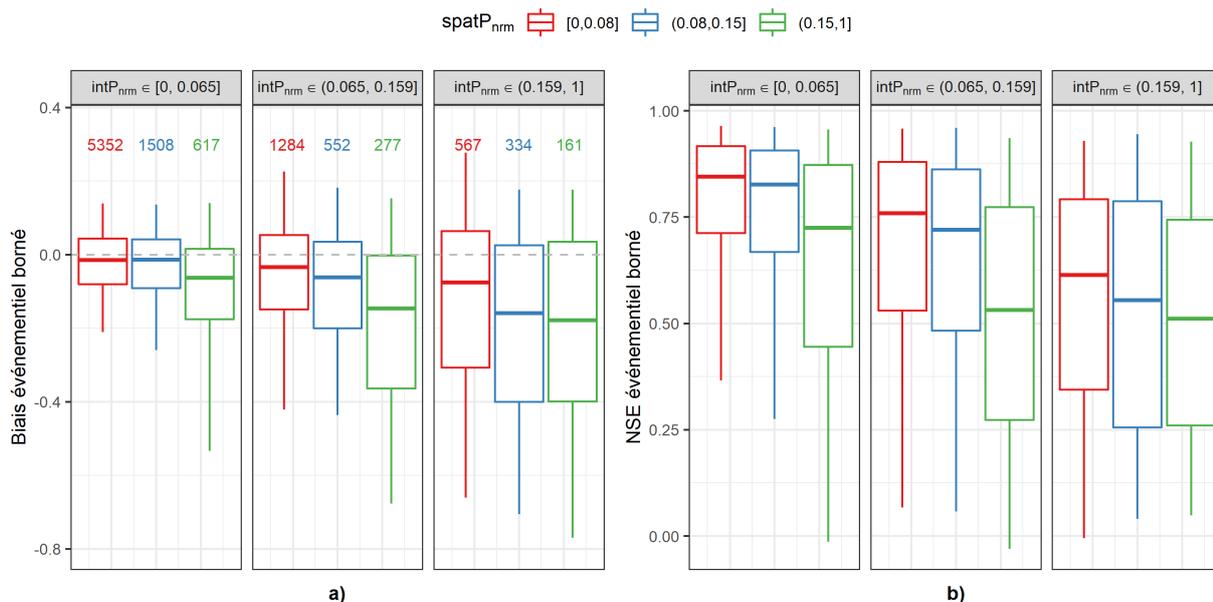


FIGURE 3.9 – Analyse du lien entre la variabilité spatiale de la pluie $spatP_{nrm}$, l'intensité de la pluie $intP_{nrm}$ et le biais en crue (a) ou le critère NSE en crue (b) de GR5H (validation croisée). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 % et contiennent une valeur par événement. Le découpage des classes (centiles 70 % et 90 %) a été choisi de telle sorte que les boîtes à moustaches contiennent un nombre suffisant d'événements mais avec des valeurs d'indices de pluie les plus différentes possibles. Le nombre d'événements par classe est indiqué au dessus de chaque boîte à moustaches.

Les figures 3.10a et 3.10b montrent que la variabilité spatiale et l'intensité de la pluie sont liées aux erreurs du modèle particulièrement lorsque les conditions antécédentes d'humidité sont faibles. En d'autres termes, on remarque que le modèle a moins de difficultés à reproduire

le volume de crue lorsque les bassins sont plus humides, même lorsque les événements de pluie sont intenses et variables dans l'espace. Cependant, les valeurs de critère de NSE dans ce cas (fig. 3.10c et 3.10d) indiquent que, même en conditions humides, les erreurs du modèle sont plus importantes lorsque la pluie est très intense.

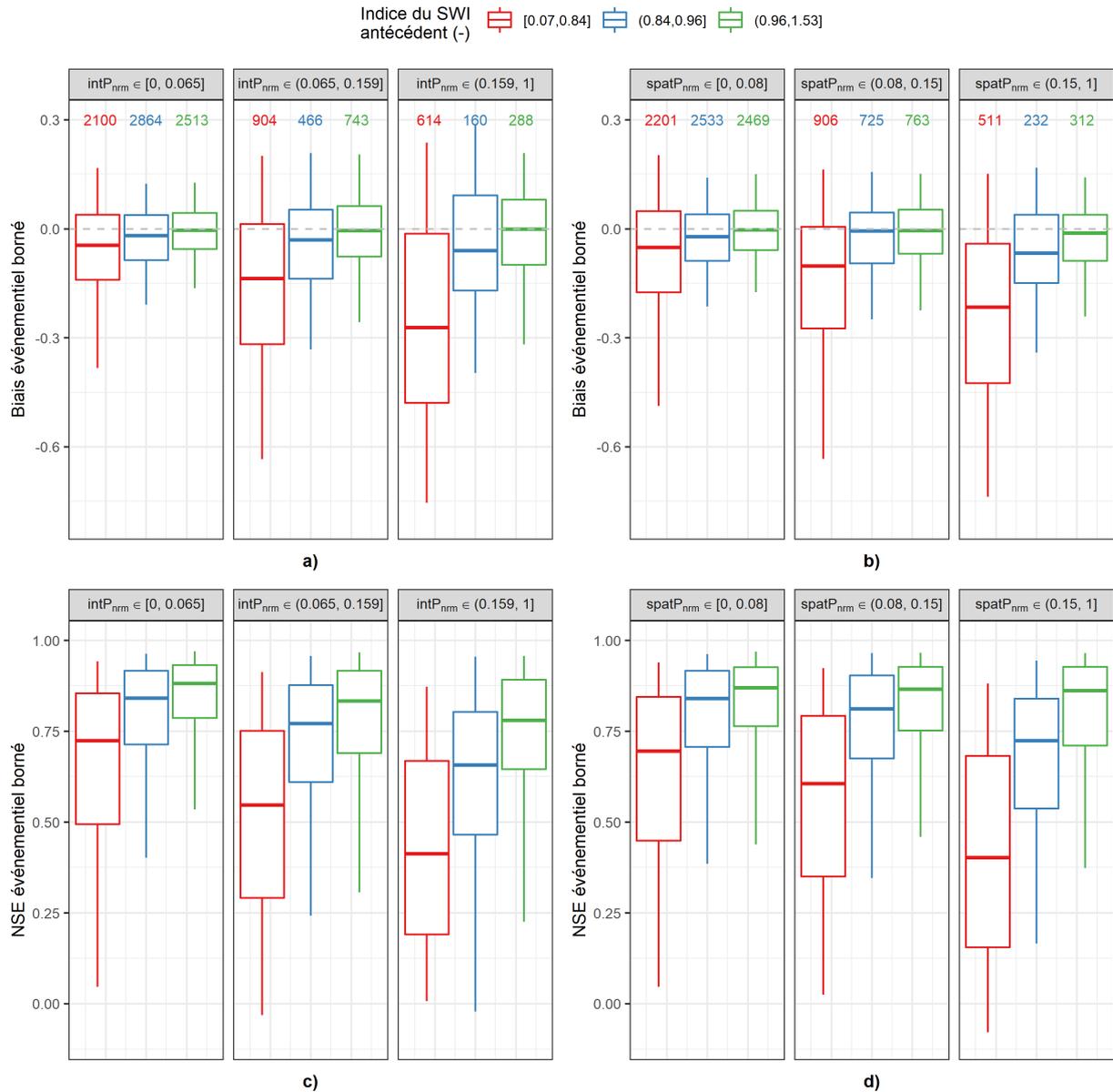


FIGURE 3.10 – Analyse du lien entre la variabilité spatiale de la pluie $spatP_{nrm}$, l'intensité de la pluie $intP_{nrm}$, les conditions antécédentes d'humidité du sol (SWI) et les performances en crue de GR5H (validation croisée). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 % et contiennent une valeur par événement. Le découpage des classes (centiles 70 % et 90 % pour $intP_{nrm}$ et $spatP_{nrm}$ et centiles 33 % et 76 % pour le SWI) a été choisi de telle sorte que les boîtes à moustaches contiennent un nombre suffisant d'événements mais avec des valeurs d'indices de pluie les plus différentes possibles. Le nombre d'événements par classe est indiqué au dessus de chaque boîte à moustaches.

3.4 Conclusion

Les analyses présentées dans ce chapitre ont permis de se rapprocher un peu plus des causes pouvant expliquer les faibles performances du modèle GR5H pour certains événements de crue. Nous avons choisi d'utiliser des indices qui caractérisent deux propriétés des événements de pluie : l'intensité moyenne et la variabilité spatiale des cumuls afin de séparer leur rôle dans la réponse des bassins versants et dans les erreurs du modèle. Les principaux résultats sont les suivants :

- Il existe des différences importantes des erreurs du modèle entre les bassins du sud de la France (pourtour méditerranéen), et le reste des bassins de notre base de données. Les erreurs du modèle pour les événements d'été y sont plus marquées que sur les autres bassins versants.
- Les événements pluvieux qui engendrent les crues observées sur ces bassins sont généralement intenses et fortement variables dans l'espace.
- Les sous-estimations les plus fortes du volume de crue sont associées aux événements de pluie les plus intenses, les plus fortement variables dans l'espace, et qui ont lieu lorsque l'humidité antécédente des bassins est faible.
- Plusieurs événements, soit très intenses et peu variables dans l'espace, soit très variables dans l'espace et peu intenses à l'échelle du bassin versant engendrent des erreurs importantes du modèle.
- Lorsque des événements intenses tombent sur des bassins versants plus humides, les erreurs du modèle peuvent être élevées. Ces erreurs ne sont pas seulement liées aux volumes de crue, comme en atteste les valeurs de biais, mais aussi à la forme des hydrogrammes (cf. valeurs du NSE événementiel).

Ces résultats ont des conséquences fortes sur la recherche d'une structure du modèle qui permettrait de palier les déficiences du modèle. En particulier, on retiendra les éléments suivants :

- La réaction des bassins aux fortes intensités de pluie est probablement très dépendante des conditions antécédentes d'humidité des sols.
- Cette réaction est variable selon le bassin versant considéré, ce qui signifie qu'il pourra être difficile d'obtenir une seule solution non paramétrée pour tous les bassins.
- Les erreurs du modèle n'étant pas uniquement de nature volumétrique, il nous faudra tester des hypothèses à différents niveaux de la structure.
- L'utilisation d'un modèle global risque de limiter les améliorations possibles. Il faudrait pouvoir tester une structure intégrant un degré de discrétisation spatiale de la pluie pour le vérifier.

Chapitre 4

Analyse des flux internes à l'échelle événementielle

4.1 Contexte

Les résultats du diagnostic du modèle, que nous avons présentés dans les chapitres précédents, indiquaient que les sous-estimations du modèle sont très corrélées avec le rendement du réservoir de production. Si on se place dans un cadre d'hypothèse de modification de la structure, ces résultats pointent vers une modification du calcul de la pluie efficace en condition de faible humidité du bassin lors d'événements pluvieux intenses et/ou fortement variables spatialement. Cependant, il reste à déterminer si le modèle sous-estime le débit observé en raison d'une simulation trop faible de la pluie efficace ou lorsque la pluie efficace est faible. En d'autres termes, on cherche à savoir si le calcul de la pluie efficace est un facteur explicatif des erreurs du modèle ou si les erreurs du modèle sont plus importantes au moment où la pluie efficace est faible par rapport au reste de l'année. Cette nuance est importante car elle déterminera la crédibilité des hypothèses structurelles que nous testerons par la suite. Plus généralement, il nous faut identifier les flux et les états qui seraient susceptibles d'influencer les performances du modèle lorsque les intensités de pluie sont élevées.

Il existe, dans la littérature, de nombreuses études qui portent sur les variables internes des modèles hydrologiques. Le but est généralement de lier les flux et les états des modèles à des données complémentaires au débit afin de contraindre l'équifinalité (e.g. Nijzink *et al.*, 2018; Tong *et al.*, 2022), ou tout simplement d'évaluer le réalisme des hypothèses de modélisation (e.g. Bouaziz *et al.*, 2021). L'étude des variables internes des modèles hydrologiques permet aussi de mieux comprendre leur comportement (e.g. Khatami *et al.*, 2019). Par exemple, Ficchi *et al.* (2019) ont montré que les flux d'interception du modèle GR5 ne sont pas cohérents entre pas de temps. Ils évaluent l'intérêt d'utiliser un réservoir d'interception aux pas de temps horaire et infra-horaires afin d'améliorer la cohérence entre les flux internes du modèle.

Dans cette étude, nous explorons les variations de plusieurs variables caractérisant les états et les flux internes du modèle en fonction des performances en crue afin de cibler les facteurs pouvant expliquer certaines déficiences du modèle.

4.2 Éléments méthodologiques

Dans la première partie de cette analyse, nous allons chercher à lier le rendement simulé du modèle (ou pluie efficace) au biais en crue ainsi qu'à des caractéristiques des événements de pluie. Nous définissons la pluie efficace à chaque pas de temps comme suit :

$$Pr = Pn - Ps + Perc \quad (4.1)$$

avec Pn la pluie nette, Ps la part de la pluie nette qui remplit le réservoir de production et $Perc$ la percolation du réservoir de production (les équations et le schéma du modèle sont présentées dans le chapitre 1). Pour chaque événement de crue, un indice de pluie efficace est calculé en prenant en compte les valeurs de Pr calculées sur les pas de temps de l'événement de pluie associé.

$$Pr_{evP} = \sum_{i=1}^{n_t} Pr(i) \quad (4.2)$$

avec n_t la longueur de l'événement de pluie et $i = 1$ le début de l'événement de pluie.

Nous comparons ensuite ce rendement à la différence entre le niveau du réservoir de routage au début et à la fin de l'événement de crue :

$$\Delta R_{evQ} = R(n_j) - R(1) \quad (4.3)$$

avec n_j la durée de l'événement de débit et $R(1)$ le niveau du réservoir de routage au premier pas de temps de l'événement de débit.

Afin de comprendre la dynamique de l'erreur en crue, un indice de dynamique temporelle du biais en crue est calculé pour tous les événements d'été sous-estimés :

$$I_{df} = T_p/n_j \quad (4.4)$$

avec T_p l'instant (exprimé en pas de temps depuis le début de l'événement de débit), entre le pic de débit observé et la fin de la chronique, à partir duquel le ratio Q_{sim}/Q_{obs} devient supérieur à un. Ce ratio est basé sur l'hypothèse que le biais ne change pas de signe plus d'une fois sur la durée de l'événement sous-estimé. Ce n'est bien évidemment pas le cas pour tous les événements. Par conséquent, nous distinguons les événements pour lesquels $I_{df} < 1$ des événements pour lesquels c'est l'inverse. Dans le premier cas, le biais a nécessairement changé de signe au cours de l'événement. Dans le deuxième cas, le débit simulé devient supérieur au débit observé obligatoirement après la fin de l'événement, indiquant ainsi un retard important du modèle.

Pour caractériser l'impact potentiel de la fonction d'échange sur les erreurs du modèle, l'échange total effectué pendant l'événement de crue est calculé (Aex_{evQ}). Il dépend du niveau du réservoir de routage, des deux paramètres d'échange X_2 et X_5 et du flux de la branche pseudo-directe. À chaque pas de temps, il est égal à deux fois le demi-échange potentiel (voir équation 1.14) si le niveau du réservoir de routage et le flux de la branche directe sont suffisamment élevés (en cas de perte d'eau).

$$Aex_{evQ} = \sum_{i=1}^{n_j} Aex(i) \quad (4.5)$$

avec $Aex(i)$ le flux échangé à chaque pas de temps. Dans le même esprit que le biais en dehors des événements (voir équation 2.4), nous calculons le volume échangé en dehors des événements par sous-période de validation et par saison (été ou hiver) :

$$Aex_{cWev} = \sum_{i=1}^p Aex(i) - \sum_{j=1}^m \sum_{i=1}^{n_j} Aex(i) \quad (4.6)$$

p est la longueur de la sous-période considérée et m le nombre d'événements sélectionnés sur cette sous-période.

Enfin, plusieurs indicateurs des conditions antécédentes d'évaporation du bassin (observées et simulées) sont utilisés :

- $ETPa_{1m}$: le ratio entre l'évapotranspiration potentielle (ETP) calculée sur le mois précédant l'événement de pluie et l'ETP annuelle moyenne du bassin.
- $ETRa_{1m}$: le même ratio mais avec l'évapotranspiration réelle (ETR) calculée par le modèle sur le mois précédant l'événement de pluie. À chaque pas de temps, l'ETR est calculée de la manière suivante (voir équations 1.1, 1.5 et 1.9 pour plus de détails) :

$$ETR = E_s + E_i \quad (4.7)$$

avec E_s l'évapotranspiration du réservoir de production et E_i l'évapotranspiration du réservoir d'interception.

- s_a : le taux de remplissage du réservoir de production au pas de temps précédant le début de l'événement de pluie.
- SWI_a : la valeur du SWI (voir équation 2.6) au pas de temps précédant le début de l'événement de pluie.

Le tableau 4.1 résume les notations utilisées dans ce chapitre.

TABLE 4.1 – Liste des variables internes et indicateurs.

Abréviation	Nom	Unité
Q_{evQ}^{obs}	Volume de crue observé	mm
Q_{evQ}^{sim}	Volume de crue simulé	mm
Pr	Pluie efficace	mm/h
Qr	Débit provenant de la branche indirecte	mm/h
Qd	Débit provenant de la branche pseudo-directe	mm/h
Pr_{evP}	Pluie efficace de l'événement	mm
ΔR_{evQ}	Variation du niveau du réservoir de routage	mm
I_{df}	Dynamique du biais en crue	–
Aex_{evQ}	Volume échangé pendant l'événement	mm
Aex_{cWev}	Volume échangé en dehors des événements	mm
$ETPa_{1m}$	Indice d'ETP antécédente (1 mois)	–
$ETRa_{1m}$	Indice d'ETR antécédente (1 mois)	–
s_a	Remplissage antécédent du réservoir de production	–
SWI_a	Indice d'humidité antécédente	–

4.3 Résultats et analyses

4.3.1 Un manque de pluie efficace ou un routage trop lent ?

La figure 4.1a présente les valeurs du volume efficace simulé par événement sous-estimé en été par rapport au volume de crue observé. La figure 4.1b présente la distribution du biais en crue en été pour les événements pour lesquels le volume efficace est inférieur au volume de crue et les événements pour lesquels le volume efficace est supérieur au volume de crue. Les résultats de la figure 4.1a montrent d'une part que, pour la majorité des événements estivaux sous-estimés, la pluie efficace simulée est de l'ordre de grandeur du volume de crue observé (les points sont proches de la bissectrice du graphique), et d'autre part que les plus fortes sous-estimations sont associées à des valeurs faibles de la pluie efficace (points rouges).

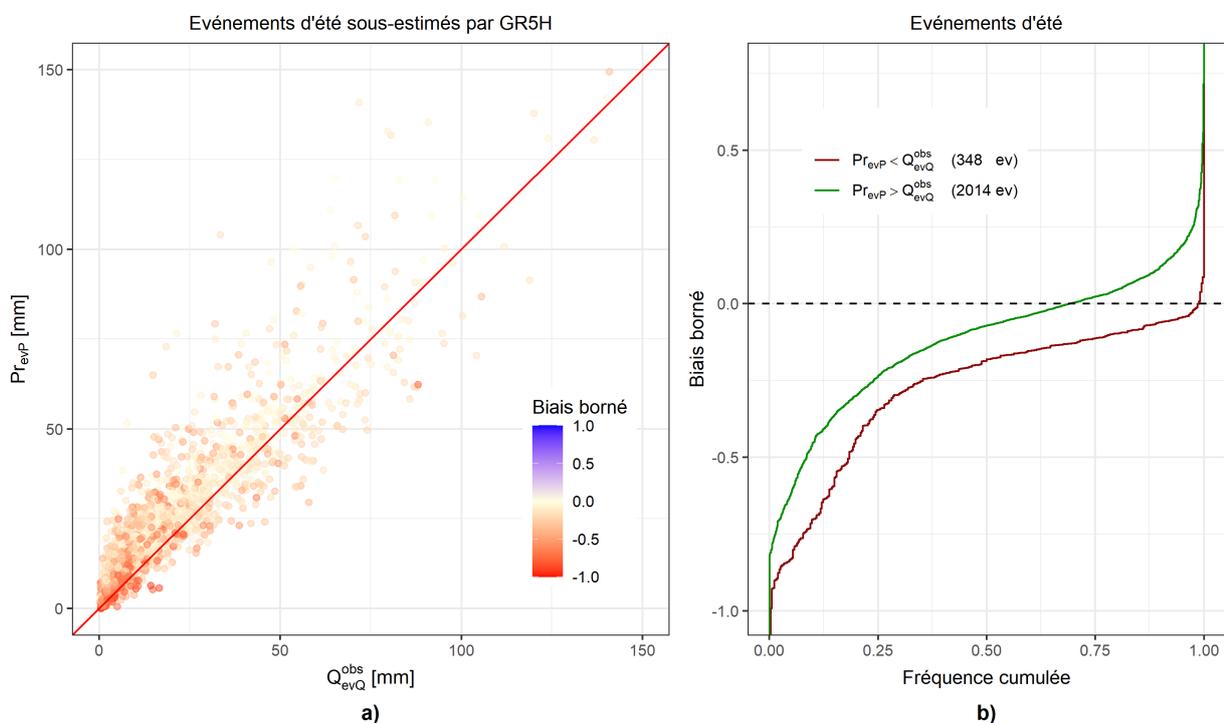


FIGURE 4.1 – Comparaison entre la pluie efficace simulée par GR5H et le volume de crue observé pour les événements d'été sous-estimés (validation croisée).

Par ailleurs, la figure 4.1b montre que les sous-estimations du modèle sont plus fortes lorsque la pluie efficace est inférieure au volume de crue (courbe rouge en dessous de la ligne verticale en pointillés), mais que cette situation ne représente qu'un peu plus de 15 % des événements d'été (348 événements). Ces résultats indiquent que les faibles valeurs de rendement du réservoir de production ne suffisent pas à expliquer les sous-estimations du volume de crue par le modèle. On doit donc chercher une autre composante du modèle où la pluie efficace aurait été stockée sans être transférée et répartie au bon moment. Une autre explication serait que le modèle sous-estime les cinétiques lentes des hydrogrammes (le modèle ne simule pas explicitement le débit de base). Ces cinétiques lentes sont souvent associées à des mécanismes de subsurface et de nombreuses études ont montré qu'une part importante de l'eau des crues provient de ces mécanismes plutôt que de mécanismes de surface (e.g. Hugen-

schmidt *et al.*, 2014 ; Chappell *et al.*, 2017). En l'absence de données permettant de tracer les chemins de l'eau, cette hypothèse reste une explication potentielle des déficiences du modèle.

Les événements de pluie associés aux événements pour lesquels la pluie efficace semble sous-estimée ont un cumul anti-corrélé à leur variabilité spatiale (fig. 4.2). Cela signifie qu'il est possible que la sous-estimation de la pluie efficace soit liée à la variabilité spatiale de la pluie. En d'autres termes, un événement de pluie avec un fort cumul localisé peut ne pas être « détecté » par le modèle car la valeur de pluie à l'échelle du bassin versant est trop faible au vu des conditions initiales du réservoir de production. Les valeurs des coefficients de corrélation n'étant pas très élevées, cette relation reste faible. Une autre explication serait que la pluie est parfois sous-estimée, ce qui engendrerait une sous-estimation de la pluie efficace.

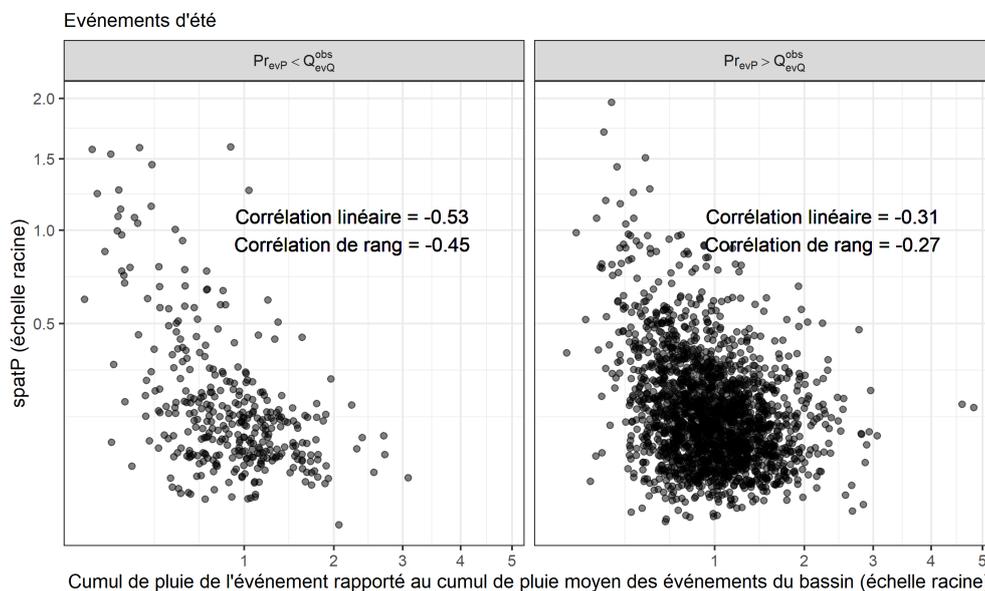


FIGURE 4.2 – Liens entre variabilité spatiale de la pluie, cumul de pluie et deux groupes en fonction de la pluie efficace. 2 362 événements d'été sont considérés ici.

Pour comprendre comment une pluie efficace « bien estimée » peut engendrer une sous-estimation du volume de crue, prenons l'exemple d'une crue importante qui a eu lieu sur le Gardon de Saint-Jean à Corbès en septembre 2014 suite à un événement de pluie intense et fortement variable spatialement (le Gardon est connu pour sa réaction aux fortes intensités de pluie ; e.g. Gaume & Bouvier, 2004). On voit sur la figure 4.3 que le modèle n'a pas réussi à reproduire le volume de cette crue (la courbe rouge est nettement en dessous de la courbe grise). Pourtant, la pluie efficace générée par le modèle est bien du même ordre de grandeur que le débit observé. On remarque aussi que la part du débit qui provient de la branche pseudo-directe fait réagir le modèle au début de l'événement avant que le débit venant de la vidange du réservoir de routage ne prenne le relais. On peut donc faire l'hypothèse que, pour cet événement, une meilleure répartition de la pluie efficace entre branche pseudo-directe (linéaire) et branche indirecte (non linéaire) de l'écoulement aurait permis au modèle de reproduire le volume de crue.

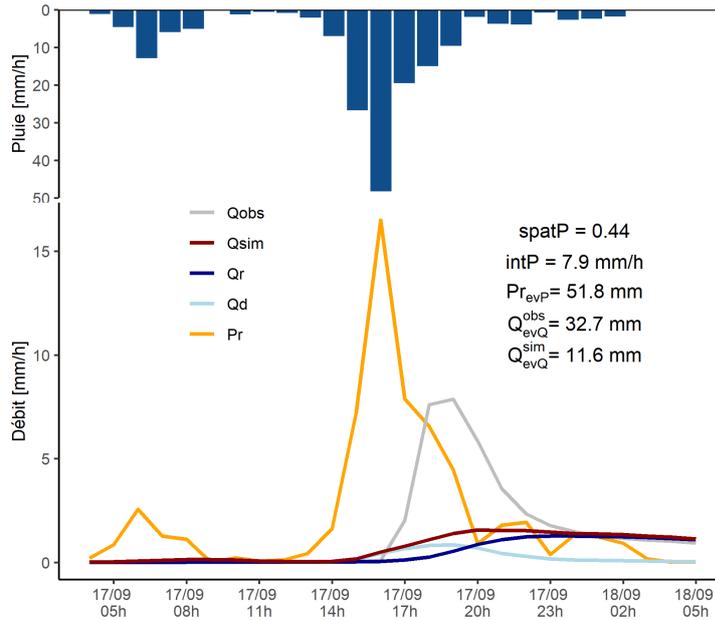


FIGURE 4.3 – Flux internes de GR5H pour l'hydrogramme de crue du Gardon de Saint-Jean à Corbès (262 km²) en septembre 2014 (validation croisée). La pluie est une moyenne spatiale à l'échelle du bassin versant.

En appliquant ce raisonnement à tous les événements estivaux sous-estimés (dont nous avons visualisé les hydrogrammes), on peut regarder si le volume de la pluie efficace est du même ordre de grandeur que la variation du réservoir de routage entre le début et la fin de l'événement de débit (fig. 4.4). Cette figure montre que le biais est le plus important lorsque les points se situent à proximité de la bissectrice, ce qui indique que les sous-estimations lors des événements de crues sont liées à un stockage trop important de la pluie efficace dans le réservoir de routage.

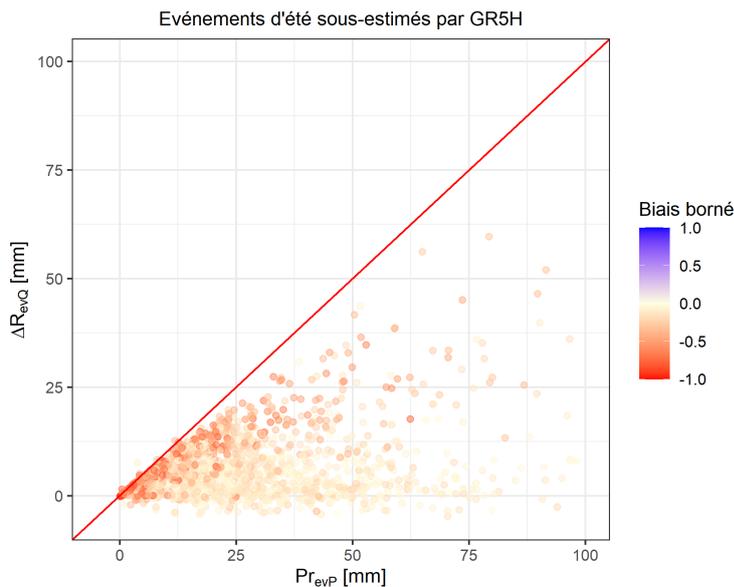


FIGURE 4.4 – Volume d'eau additionnel stocké dans le réservoir de routage par rapport au volume de la pluie efficace en fonction du biais pour les événements estivaux sous-estimés par GR5H (validation croisée).

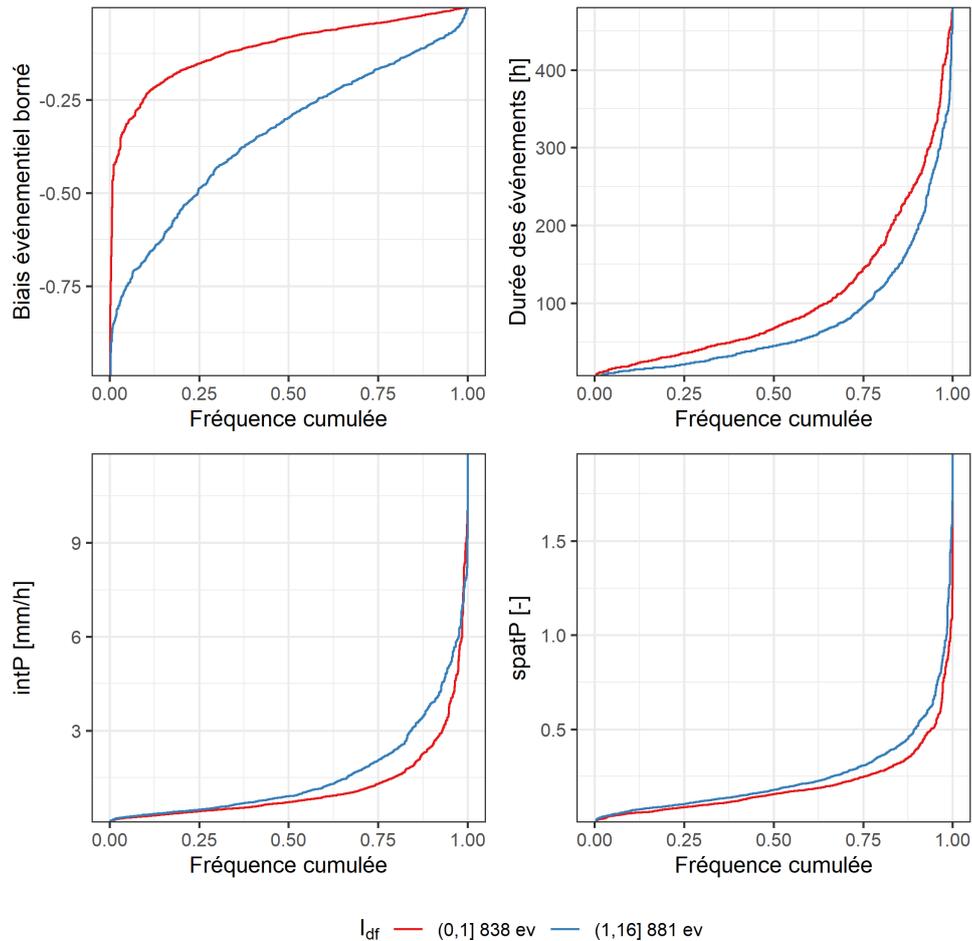


FIGURE 4.5 – Distribution du biais événementiel, de la durée des événements, de l'intensité de la pluie et de la variabilité spatiale de la pluie en fonction du retard du débit simulé en été. 1733 événements d'été sous-estimés par le modèle sont considérés ici. La courbe rouge correspond aux événements pour lesquels le biais du modèle change de signe entre le pic observé et la fin de l'événement. La courbe bleue correspond aux événements pour lesquels le biais du modèle change de signe après la fin de l'événement.

Si on s'intéresse au « réalisme » de la structure de GR5H, on peut se demander s'il est cohérent de chercher à ce que la totalité de la pluie efficace se retrouve dans l'hydrogramme de crue. En effet, la dynamique de la vidange du réservoir de routage observée sur la figure 4.3 indique que ce réservoir semble adapté pour générer une composante plus lente de la cinétique de l'événement. Cette part de l'hydrogramme simulé provient pour partie de l'eau précédemment stockée dans le réservoir. Par ailleurs, nous avons inclus la percolation du réservoir de production dans la variable Pr . À l'échelle d'événements aussi courts et qui ont lieu lorsque le réservoir est faiblement rempli, la percolation représente cependant un volume assez faible (quelques pourcentages de la pluie efficace).

Si la pluie efficace simulée est « suffisamment » élevée mais qu'elle se retrouve stockée dans le réservoir de routage, à quel moment ce volume est-il restitué dans l'hydrogramme simulé ? Cette restitution tardive est-elle liée à des dynamiques spatio-temporelles des événements ? Les résultats présentés en figure 4.5 permettent de répondre en partie à ces questions. On observe que lorsque la pluie est plus intense et/ou plus fortement variable dans l'espace,

le réservoir de routage restitue le volume efficace plus tardivement ($I_{df} > 1$). De plus, les événements pour lesquels le débit simulé dépasse le débit observé après la fin de l'événement sont plus courts que les autres événements d'été sous-estimés. Enfin, par construction, les événements pour lesquels le débit simulé dépasse le débit observé avant la fin de l'événement présentent des biais moins importants.

4.3.2 Impact potentiel de la fonction d'échange sur la simulation des crues

La fonction d'échange du modèle GR5H influe sur le bilan en eau. Elle permet d'ajuster ce bilan à l'échelle annuelle en simulant des échanges entre bassins topographiques. Elle permet aussi de compenser des problèmes d'estimation de l'ETR.

Par conséquent, il pourrait exister un lien entre les échanges simulés à l'échelle des événements et le biais du modèle. On observe sur la figure 4.6 qu'il ne semble pas y avoir de lien évident entre le volume d'erreur et les volumes échangés en crue. Cela signifie qu'il y a des événements pour lesquels le modèle est en sous-estimation et où il perd de l'eau par la fonction d'échange mais qu'il n'existe pas de relation sans équivoque entre les deux. Par ailleurs les volumes sous-estimés ne sont pas souvent du même ordre de grandeur que les volumes échangés. Cependant, il existe des événements pour lesquels le modèle est en sous-estimation et dont les pertes par la fonction d'échange représentent un volume élevé. Pour ces événements, la fonction d'échange exacerbe les sous-estimations du modèle.

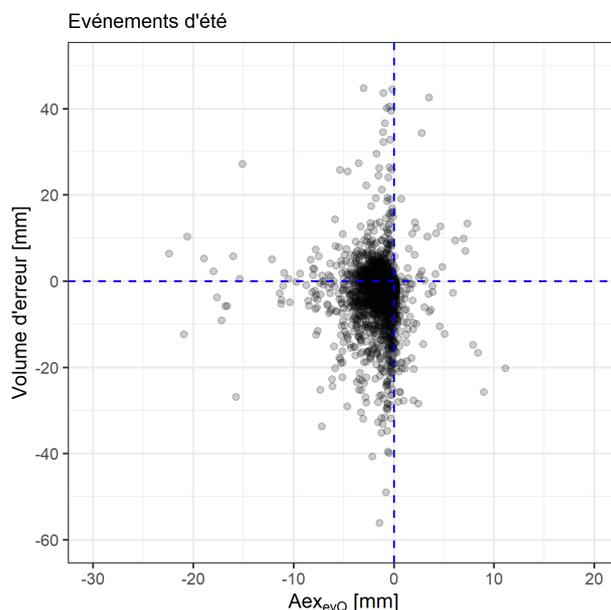


FIGURE 4.6 – Variation du volume d'erreur pour les événements d'été en fonction du volume échangé. 2 362 événements d'été sont considérés ici.

Le fonctionnement des échanges dans GR5H pourrait aussi impacter les conditions initiales du modèle. Par exemple, si un volume important d'eau est perdu en été, le réservoir de routage pourrait avoir un niveau initial trop faible pour pouvoir restituer un volume suffisant au bon moment. Une indication de l'existence d'un tel comportement serait un lien entre les volumes échangés en dehors des événements en été et le biais en crue par bassin.

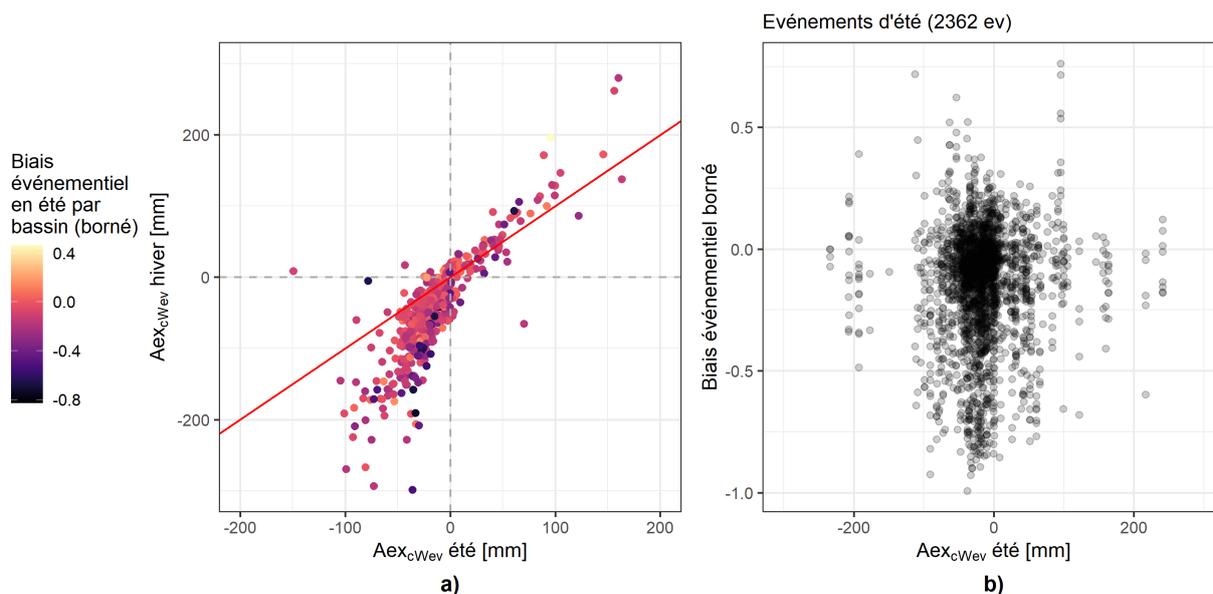


FIGURE 4.7 – Relations entre les volumes échangés en dehors des événements et le biais en crue en été.

La figure 4.7 montre qu'il n'existe pas de lien entre ces deux variables. On observe que les pertes par la fonction d'échange en dehors des événements sont moins importantes en été qu'en hiver (fig. 4.7a). On observe d'autre part que la variabilité la plus importante du biais événementiel en été a lieu lorsque les échanges estivaux en dehors des événements sont les plus faibles (dispersion des points autour de $Aex_{cW_{ev}} = 0$ sur la figure 4.7b). Par ailleurs, on observe des valeurs négatives du biais événementiel quelque soit le signe des échanges en dehors des événements (en été comme en hiver).

Ces résultats sont cohérents avec ceux de Ficchi (2017) qui a analysé les flux d'échanges en crue des modèles GR4 et GR5. Il montre que les flux d'échanges négatifs (pertes) en crue sont plus élevés aux pas de temps horaires et infra-horaires qu'au pas de temps journalier. L'utilisation d'un réservoir d'interception permet une meilleure cohérence de ces flux entre pas de temps. Il montre aussi que l'utilisation de formulations de la fonction d'échange visant à diminuer les pertes en crue n'impacte pas le biais du modèle.

4.3.3 Impact de l'ETP sur les conditions antécédentes en été

Nous allons maintenant nous intéresser à une autre composante du bilan hydrologique. Nous avons vu, dans les chapitres 2 et 3, que les sous-estimations du modèle ont lieu en condition de faible humidité antécédente. On peut donc se demander si ces problèmes de sous-estimation peuvent être liés au calcul de l'ETP et de l'ETR. Le niveau du réservoir de production est-il trop bas pour calculer une pluie efficace suffisante parce que l'ETR antécédente et/ou l'ETP antécédente sont surestimées ?

En premier lieu, la figure 4.8 montre, assez logiquement, que l'ETR antécédente est majoritairement plus élevée en été qu'en hiver. Ensuite, on retrouve les tendances observées dans les chapitres 2 et 3 entre le niveau du réservoir de production et l'erreur du modèle. Le modèle a des performances moins élevées en conditions sèches. Cependant, le dégradé de

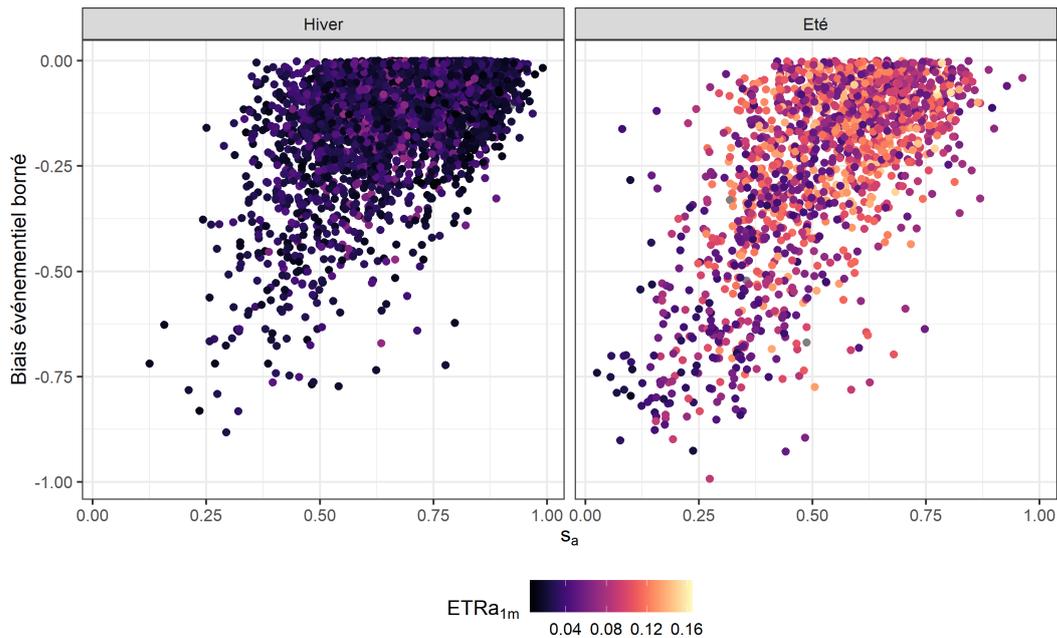


FIGURE 4.8 – Variation du biais événementiel en fonction du niveau antécédent du réservoir de production et de l'ETR antécédente calculée sur une durée de un mois avant l'événement et rapportée à l'ETP annuelle moyenne. Seuls les événements sous-estimés par le modèle sont considérés ici.

couleur représentant l'ETR antécédente indique qu'il n'y a pas de lien direct entre les valeurs fortes d'ETR (fortes pour le bassin considéré) et la sous-estimation des événements lorsque le niveau du réservoir de production est initialement bas. Au contraire, on observe plusieurs valeurs faibles d'ETR lorsque le niveau du réservoir de production est bas en été, ce qui indique que le ratio ETR sur ETP n'est pas très élevé. En résumé, le niveau du réservoir de production est bas en été car l'évapotranspiration y est plus forte qu'en hiver, mais cela ne semble pas expliquer les sous-estimations du modèle. Comme pour la fonction d'échange, il est possible que des valeurs élevées d'ETR avant le début des événements exacerbent les sous-estimations du modèle en été.

Un problème de calcul de l'évapotranspiration pourrait se traduire par une demande en évaporation non satisfaite par l'état du réservoir de production. En d'autres termes, on se demande ici si le ratio ETR/ETP est faible lorsque les événements sont sous-estimés. La figure 4.9 montre que, lorsqu'on se place dans cette configuration, les sous-estimations du volume de crue en été sont plus élevées (points foncés éloignés de la bissectrice). À nouveau, il est possible que ce ne soit pas le facteur explicatif des erreurs du modèle mais simplement un témoin de l'état d'humidité du bassin versant. L'ETP est-elle surestimée ou tout simplement plus élevée lorsque le modèle sous-estime le volume de crue en été ?

En l'absence de mesures d'évaporation ou d'humidité du sol, une manière de savoir si la demande en évaporation est trop forte par rapport à ce qu'elle devrait être est de regarder la différence entre le SWI antécédent et le niveau antécédent du réservoir de production (on compare alors deux simulations, mais issues de modèles très différents). On peut ainsi regarder si les moments où les deux variables sont plus écartées sont les moments où le modèle

a les performances les plus faibles. La figure 4.10 montre l'inverse. Les deux variables sont plus fortement corrélées pour les événements d'été. De plus, les événements pour lesquels l'ETR est éloignée de l'ETP ne correspondent pas aux points les plus dispersés sur le graphique.

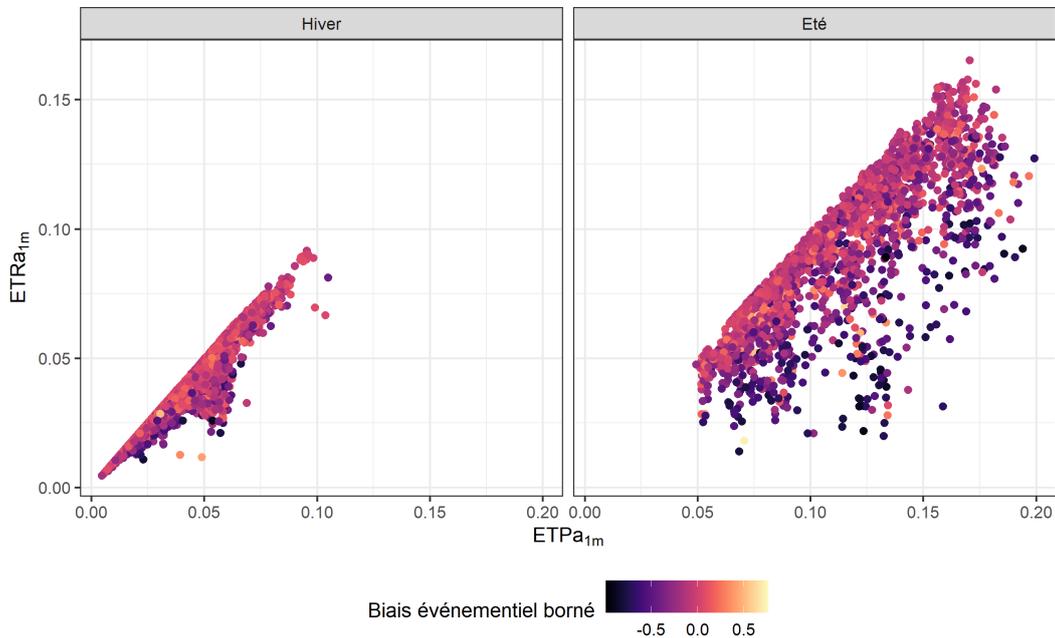


FIGURE 4.9 – Variations du biais événementiel en fonction de l'ETR antécédente et de l'ETP antécédente pour tous les événements de la base de données. L'ETR et l'ETP sont exprimées en ratio par rapport à l'ETP annuelle moyenne.

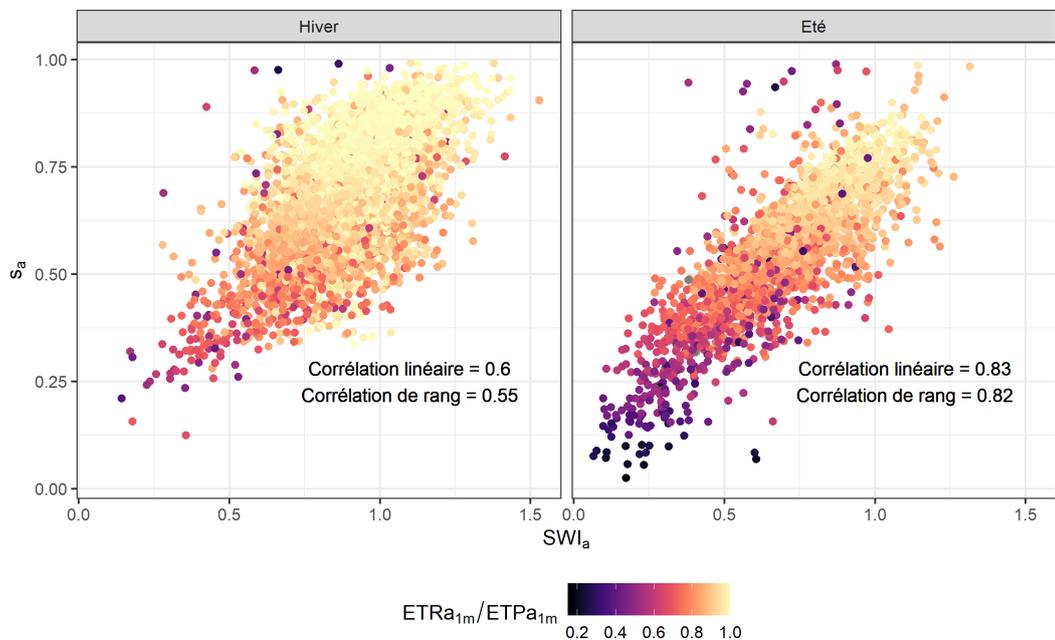


FIGURE 4.10 – Liens entre le SWI antécédent (simulation issue du modèle ISBA de Météo-France) et le remplissage antécédent du réservoir de production pour différentes valeurs du ratio ETR sur ETP.

4.4 Conclusion

Les analyses présentées dans ce chapitre avaient pour but de cibler les éléments de la structure du modèle pouvant expliquer la sous-estimation du volume de crue en été en analysant les flux internes du modèle. Les résultats ont montré que les modifications du modèle qui viseront à améliorer la simulation des crues résultant de fortes intensités de pluie devront s'effectuer au niveau du routage mais pourront aussi s'effectuer au niveau du rendement du réservoir de production. Les résultats ont permis de mettre en évidence les limites du routage de la pluie efficace par la branche indirecte lorsque la pluie est intense et fortement variable dans l'espace et que les événements induits sont de courte durée. Enfin, nous n'avons pas trouvé de liens forts entre la fonction d'échange, l'ETR et l'ETP, et la sous-estimation des crues en été. Cependant, il existe des situations où la pluie efficace est faible et les flux d'échange et d'ETR sont élevés, ce qui donne lieu à une exacerbation des sous-estimations du modèle.

Partie II

**Vers une modélisation généraliste
plus adaptée à simuler
les crues résultant
de fortes intensités de pluie**

Chapitre 5

Modification de la structure du modèle GR5H

Ce chapitre correspond à un article publié dans la revue *Hydrological Processes* : Paul C. Astagneau, François Bourgin, Vazken Andréassian and Charles Perrin (2021). Catchment response to intense rainfall: evaluating modeling hypotheses. *Hydrological Processes*, 36(8), e14676, <https://doi.org/10.1002/hyp.14676>.

5.1 Résumé

Les crues générées par de fortes intensités de pluie sont très difficiles à modéliser. La réponse des bassins versants à ces événements est très hétérogène, ce qui est principalement dû à une association complexe de processus hydrologiques à des échelles spatiales et temporelles fines. L'objectif de cette étude est de trouver une méthode pour informer un modèle hydrologique sur la variabilité de la réponse des bassins versants aux intensités de pluie. Pour cela, nous avons utilisé l'intensité de pluie horaire comme un proxy de l'activation de processus hydrologiques rapides et hétérogènes. Nous avons développé trois hypothèses de modélisation qui augmentent la polyvalence d'un modèle hydrologique en modifiant le calcul du volume efficace ainsi que sa répartition temporelle lors d'événements de pluie intenses. Ces modifications ont été implémentées dans le modèle pluie-débit global GR5H. Les différentes hypothèses de modélisation ont été testées sur 229 bassins versants français sur lesquels 10 652 événements de crues ont été sélectionnés. La performance des différentes versions du modèle a été évaluée en considérant cinq groupes de bassins versants et trois caractéristiques des événements. Les résultats ont montré que l'introduction d'une dépendance dynamique des flux internes du modèle aux intensités de pluie horaires permet d'améliorer la simulation des crues, en particulier pour des bassins méditerranéens. Des valeurs génériques des paramètres des fonctions introduites sont proposées pour limiter l'augmentation de la complexité au calage.

5.2 Introduction

5.2.1 High-intensity rainfall events and heterogeneous processes

Floods caused by high-intensity rainfall events are difficult to predict. In temperate climates, they occur most frequently in the summer or at the beginning of autumn, when soils are relatively dry, causing non-linear streamflow responses (Chappell *et al.*, 2017). It is common to read that infiltration-excess runoff processes (or Hortonian runoff; Horton, 1933) play a major role in the generation of these events (e.g. Lana-Renault *et al.*, 2007; Latron & Gallart, 2008; Manus *et al.*, 2009). However, some studies report saturation mechanisms (e.g. Estrany *et al.*, 2010; Braud *et al.*, 2014) and contributions of subsurface storages to streamflow (e.g. Hugenschmidt *et al.*, 2014; Chappell *et al.*, 2017) to be active even for very intense rainfall events in summer. Overall, there is an interplay of varied flood-generating processes that are scale-dependent (e.g. Blöschl & Sivapalan, 1995; McDonnell *et al.*, 2021), such as local soil saturation (Garambois *et al.*, 2014) and threshold in the connectivity between catchment storages and the stream network through lateral pipe flow processes from the hillslopes (e.g. Uchida *et al.*, 2005; van Meerveld & McDonnell, 2006). Last, these floods are often characterized by short response times to rainfall events making them even harder to predict since fewer data are available for parameter estimation and evaluation of simulations (e.g. Perrin *et al.*, 2007).

5.2.2 On the need to improve model structures

Hydrological models have poorer performance on catchments where these floods occur, especially when there is a strong seasonality of antecedent soil moisture conditions (e.g. McMillan *et al.*, 2016; Melsen *et al.*, 2018). This is mainly due to the non-linearity of the processes involved in the catchment response that hydrological models – which are simplifications in time, space, and processes (e.g. Hrachowitz & Clark, 2017) – struggle to characterize. For example, Astagneau *et al.* (2021a) showed that the GR5H model (Le Moine, 2008; Ficchi *et al.*, 2019) tends to underestimate flood volumes in summer especially when high-intensity rainfall events occur. Recent advances in hydrological modeling have mostly been achieved through the development of better parameterization procedures (e.g. Moussa & Chahinian, 2009; de Lavenne *et al.*, 2019; Mizukami *et al.*, 2019; Pool *et al.*, 2021), the use of additional data to streamflow (e.g. Rakovec *et al.*, 2016; Nijzink *et al.*, 2018; Bouaziz *et al.*, 2021), or the development of small headwater catchment models guided by an improved understanding of underlying hydrological processes, e.g., through isotopic measurements (e.g. Birkel *et al.*, 2014; Kuppel *et al.*, 2018). However, there is still room for model structural improvements and development of other perceptual models at the meso- (100–1000 km²) and regional scales (>10,000 km²; e.g. Fenicia & McDonnell, 2022; Fenicia *et al.*, 2022), since model structure imperfections are one of the main sources of uncertainty in streamflow prediction (e.g. Clark *et al.*, 2008; Knoben *et al.*, 2020). In particular, there is a need for modeling the contribution of highly heterogeneous processes taking place at fine spatiotemporal scales to river flows at larger scales (Blöschl *et al.*, 2019).

5.2.3 Model improvements from better diagnostics

In recent years, model improvement has also been achieved through the use of better evaluation and diagnostic frameworks. In this respect, several studies used hydrological signatures in order to assess model “realism”, adequacy, and performance and therefore go beyond the simple evaluation of models by aggregated statistics (e.g. Yilmaz *et al.*, 2008; Gupta *et al.*, 2008; McMillan *et al.*, 2017). For example, de Boer-Euser *et al.* (2017) compared the performance of 11 models on the Meuse basin and found differences in their ability to reproduce several signatures (e.g. signatures describing flashy dynamics) while yielding similar overall performance. Euser *et al.* (2013) proposed a multi-objective evaluation framework to assess the ability of a hydrological model to simulate different hydrological signatures emphasizing different characteristics of the hydrograph and therefore different processes. Another way to evaluate models is to look for dominant parameters for a given catchment in order to identify simulated dominant processes and then link the results with hydroclimatic conditions (e.g. Herman *et al.*, 2013). These frameworks enable multi-hypothesis testing (e.g. Clark *et al.*, 2011; Blöschl, 2017) and therefore allow us to evaluate the value of making a model more complex for a given catchment (e.g. Hrachowitz *et al.*, 2014). However, identifying situations where the modification of a model component improves streamflow simulation and model robustness remains a challenging task (van Esse *et al.*, 2013).

5.2.4 Fast catchment response in models

When computing runoff production, modelers commonly define several flow components. This can be achieved by setting a partitioning coefficient from a single production store outflow, which can be a fixed parameter as in the GR models (Perrin *et al.*, 2003) or a calibrated parameter as in the IHACRES model (Jakeman & Hornberger, 1993). Flow partitioning can also be applied to multiple runoff production store outflows, such as in the MORDOR model (Garçon, 1996; Garavaglia *et al.*, 2017). In some versions of the HBV model (Bergström, 1976; Parajka *et al.*, 2007; Seibert & Vis, 2012) and in the FLEX-Topo model (Savenije, 2010), a “very-fast” runoff component is calculated from the upper routing store when it reaches a certain threshold.

Some hydrological models explicitly simulate infiltration-excess overland flow. Their formulation is often based on the Green–Ampt infiltration equation (Green & Ampt, 1911) or one of its extensions (Beven, 2021) to estimate the soil infiltration capacity rate. Then, if the intensity rate is higher than the infiltration capacity rate, ponding starts and surface runoff is generated. This is the case for some versions of TOPMODEL (Beven & Kirby, 1979; Beven *et al.*, 2021) where soil infiltration is based on an extension of the Green–Ampt equation (Beven, 1984). Some hydrological models, like the SWAT model (Arnold *et al.*, 2012), include the SCS-CN method (SCS, 1956) to estimate surface runoff. Several studies have focused on modifying the SCS-CN formulation to improve streamflow simulation (e.g. Mishra *et al.*, 2004; Verma *et al.*, 2020, 2021). For example, Pang *et al.* (2020) modified the SCS-CN to take slopes and precipitation intensity into account. They found that their new formulation performed better than their reference model in simulating floods of the Chao River watershed. In the JULES land-surface model (Best *et al.*, 2011), a local surface runoff component is calculated when the throughfall rate becomes higher than the rate of maximum infiltration. Llargeron *et al.* (2018) showed that calculating a variable maximum infiltration rate was not sufficient to overcome model deficiencies in the case of flooding from intense rainfall. Beven (2021) argues that while we can make local predictions of infiltration-

excess runoff, it is not possible to estimate its contribution to the flood hydrograph at larger scales. One way to deal with this issue is to increase model temporal and spatial resolutions with “exhaustive” process description, but these models suffer from overparameterization problems (e.g. Beven, 1993; Andréassian *et al.*, 2012), especially when no other data than streamflow are available for model validation.

Several conceptual rainfall–runoff models use dedicated functions to simulate more runoff when high-intensity rainfall events occur, even when the soil moisture content is low. For example, the MODSUR model (Ledoux *et al.*, 1989) has a parameter controlling the threshold at which the production function generates runoff for low antecedent soil moisture conditions. This parameter depends on the soil type. Another attempt at simulating infiltration-excess runoff in a conceptual way was made by Willems (2014). The conceptual infiltration equation was changed to simulate more runoff in the case of high rainfall intensities and wetness conditions without, however, making the surface runoff coefficient dependent on rainfall intensities. Recently, Peredo *et al.* (2022) introduced a function in the semi-distributed version of the GR5H model (Lobligeois *et al.*, 2014) to generate more effective rainfall in the case of high-intensity rainfall events under low antecedent soil moisture conditions, and obtained some improvements in the simulation of autumn floods of the Aude River without degrading the simulation of winter floods.

5.2.5 Scope of the paper

We need to improve the ability of hydrological models to characterize the catchment response to intense rainfall, especially under low antecedent soil moisture conditions and at scales where field work experiments are not always available for model validation. We have seen that several issues arise with the current perceptual models applied at the catchment scale when only streamflow time series are available.

The objective of this study is to evaluate the benefit of making a better use of rainfall intensity to inform the structure of a hydrological model on catchment response to intense rainfall events. Three questions arise from our objective:

1. Which fluxes of a lumped rainfall–runoff model should vary with rainfall intensity at the event and catchment scale?
2. Can we identify catchments and hydro-meteorological conditions for which rainfall intensity rate functions significantly improve flood simulation?
3. Is there enough information in rainfall and streamflow time series to derive stable-in-time parameter values associated with the new modeling functions?

To address these questions, we developed a large number of modeling hypotheses aimed at dynamically modifying the storages and fluxes of a lumped rainfall–runoff model using rainfall intensity rates. Three of these hypothesis are discussed in detail in this article. We first present the data used to evaluate our hypotheses. Then, we introduce the methodological basis behind our hypotheses and the framework to evaluate performance and adequacy in regard to our objectives. The main results are then presented, analyzed, and discussed considering seasonality and catchment hydroclimatic conditions before summarizing the main conclusions of our research.

5.3 Data and methods

5.3.1 Catchment set and hydroclimatic data ¹

5.3.2 Catchment clustering

Aiming to investigate the links between hydroclimatic properties and flood-generating processes, we divided our catchment set into five groups by performing a clustering procedure. We used four indicators based on the selected rainfall and streamflow events and five indicators based on catchment hydroclimatic properties (see Tab. 5.1).

Table 5.1 – Event and catchment characteristics used for catchment clustering.

Characteristic	Type	Definition/reference	Abbrev.
Variability in event rainfall intensity [-]	Event-based	Eq. 5.1	CV _{intP}
Variability in event cumulative rainfall spatial variability [-]	Event-based	Eq. 5.2	CV _{spatP}
Variability in event antecedent soil wetness condition [-]	Event-based	Coustau <i>et al.</i> (2015)	CV _{antSWI}
Variability in event duration [-]	Event-based	–	CV _{duration}
Rainfall–runoff lag time [h]	Hydroclimatic	Ficchi (2017)	TL
Wetness index [-]	Hydroclimatic	Pm/PEm	Wi
Runoff coefficient [-]	Hydroclimatic	Qm/Pm	Rc
Magnitude of high flows [-]	Hydroclimatic	Q99/Qm	Q _{mag}
Flow autocorrelation at 24 h [-]	Hydroclimatic	–	R _{Q_24h}
Magnitude of intense rainfall [-]	Hydroclimatic	P99/Pm; Ficchi (2017)	P _{mag}

Note: Q99 is the 99th streamflow percentile. P99 is the 99th rainfall percentile. Qm is the mean streamflow. Pm is the mean rainfall. PEm is the mean potential evapotranspiration.

Two indices were calculated to characterize rainfall events: the mean rainfall intensity, *intP*, and the spatial variability in cumulative rainfall, *spatP* (or “spatial coefficient of variation in precipitation event volume”; Tarasova *et al.*, 2020). They were calculated as follows:

$$intP = \frac{1}{n_t} \sum_{t=1}^{n_t} \left[\mu (P_{x,t})_{1 \leq x \leq n_x} \right] \quad (5.1)$$

$$spatP = \frac{\sigma (\sum_{t=1}^{n_t} P_{x,t})_{1 \leq x \leq n_x}}{\sum_{t=1}^{n_t} \left[\mu (P_{x,t})_{1 \leq x \leq n_x} \right]} \quad (5.2)$$

in which $P_{x,t}$ is the rainfall of grid cell x at time t , n_t is the duration of the rainfall event (in time steps), n_x is the number of grid cells of the catchment (based on a 100-m resolution flow direction grid), μ is the mean, and σ is the standard deviation. Large values of *intP* indicate that there was a large amount of rainfall in a short period. Large values of *spatP*

1. Les éléments de cette section sont présentés dans le chapitre 1 et ne sont donc pas repris ici.

indicate that a large proportion of the rainfall event occurred in a small portion of the catchment. Antecedent soil moisture conditions can have a significant influence on flood-generating processes (Blöschl *et al.*, 2013; Berghuijs *et al.*, 2014). We used the soil wetness index (SWI) calculated by the ISBA surface model (Thirel *et al.*, 2010a,b; Coustau *et al.*, 2015). The SWI was aggregated at the catchment scale and retrieved at the start of rainfall events to attribute one value to each flood event.

Coefficients of variability CV_c (i.e. the ratio of standard deviation to mean value of a given event characteristic c) were calculated to associate event characteristics with catchment properties. In other words, for each flood characteristic, a value of its variability between events was attributed to each catchment.

We used unsupervised random forest (RF; Breiman, 2001) and hierarchical clustering algorithms available in the R environment (Liaw & Wiener, 2002; R Core Team, 2021) to classify the catchments according to the aforementioned characteristics (Tab. 5.1). RF is a powerful tool for detecting non-linear dependencies between variables and has been used in many hydrological studies (e.g. Saadi *et al.*, 2019; Stein *et al.*, 2021). More details about the use of random forest algorithms for water science applications can be found in Tyrallis *et al.* (2019).

The clustering procedure led to five groups of catchments (Figs. 5.1 and 5.14). We give here a few comments on their main characteristics:

- Group 1: they have lower values of rainfall and flow magnitude and lower values of variability in flood antecedent wetness conditions than the catchments of groups 4 and 5 but higher values than the catchments of groups 1 and 2.
- Group 2: they have the slowest flow dynamics with the highest rainfall–runoff lag times and the highest values of flow autocorrelation at 24 h.
- Group 3: they have slow flow dynamics but higher rainfall–runoff coefficients and higher values of wetness index than the catchments of group 2.
- Group 4: they are characterized by very different hydroclimatic conditions than the catchments of groups 1 to 3. They have faster flow response to rainfall, higher variability in event rainfall intensity, and higher variability in event antecedent wetness conditions, which indicates a strong flow seasonality. They are mostly located in the Rhône basin and the Mediterranean area. Seven catchments are tributaries of the Seine River. They share most of the characteristics of the other catchments of group 4 except for the magnitude of rainfall, which falls in the lower part of the distribution.
- Group 5: they have lower variability in cumulative rainfall spatial variability. It means that the rainfall fields associated with their flood events can be spatially highly variable, but, if so, this variability does not change much between events. They have higher values of streamflow magnitude than the catchments of group 4. They are all located in the Rhône basin and the Mediterranean area².

5.3.3 Hydrological modeling

We made three modeling hypotheses on the possible dependence of the storages and fluxes of the GR5H lumped hydrological model on rainfall intensity rates (Fig. 5.2), starting with a “reference” model version. All versions include an explicit interception store as proposed by Ficchi *et al.* (2019). The GR5H model has been used in several hydrological modeling studies and usually yields high streamflow performance over large sets of catchments (e.g.

2. Ce groupe diffère en partie du groupe de bassins méditerranéens défini au chapitre 3.

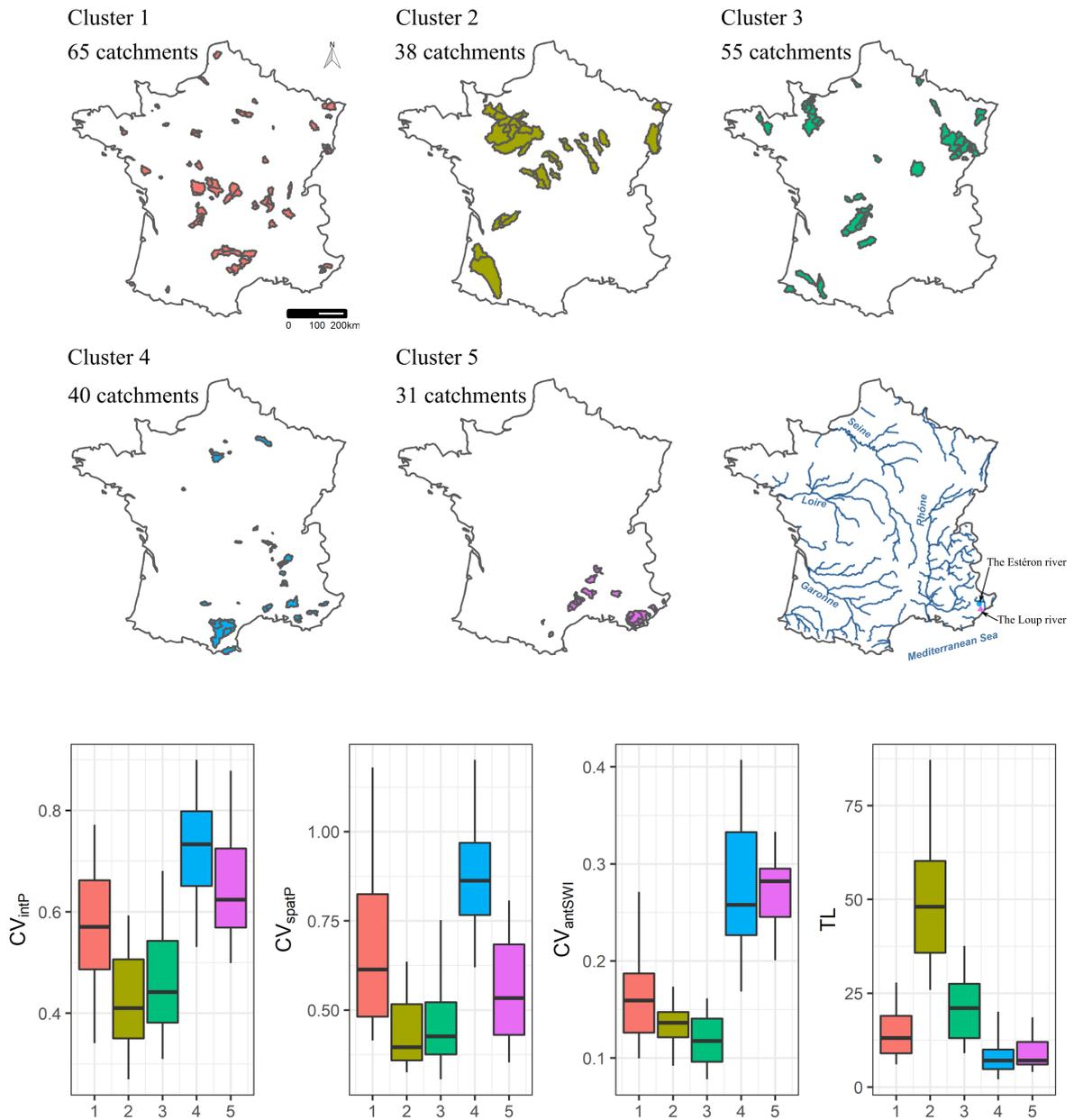


Figure 5.1 – Location of the five hydroclimatic groups of catchments and distributions of four characteristics in these groups. The distributions are presented between the 5th and 95th percentiles.

Ficchi *et al.*, 2016, 2019; Astagneau *et al.*, 2021a). Its structure may, however, experience difficulties to simulate streamflow response to specific hydrological events, in particular the high-intensity rainfall events.

We first introduce the relevant specificities of the reference model version and then the rationales behind the three modeling hypotheses.

Reference model

The GR5H rainfall–runoff model used in this study consists of an interception store, a production store, two unit hydrographs (UHs), and a routing store. A net rainfall flux (P_n) is calculated from the interception store and either fills the production store or becomes effective rainfall (P_r). The instantaneous production rate (or the fraction of net rainfall that becomes effective rainfall) is calculated as follows:

$$\eta = \left(\frac{S}{X_1}\right)^2 = s^2 = 1 - \frac{P_s}{P_n} = \frac{P_r}{P_n} \quad (5.3)$$

with S the level of the production store (mm), X_1 the production store capacity (mm), s the production store filling rate, and P_s (mm/h) the rainfall infiltrating (and remaining in) the production store (see Astagneau *et al.*, 2021a, for the complete set of equations of GR5H).

A percolation flux calculated from the production store is then added to P_r . A total of 10% (fixed partitioning coefficient $B = 0.1$) of P_r goes to a direct branch where it is routed to the catchment outlet by a symmetric UH. The remaining 90% ($1 - B = 0.9$) of P_r goes to a second branch where it is routed to the catchment outlet by an asymmetric UH and a routing store. The flow exiting the routing store (Q_r) is added to the flow exiting the symmetric UH (Q_d) to form the final simulated flow.

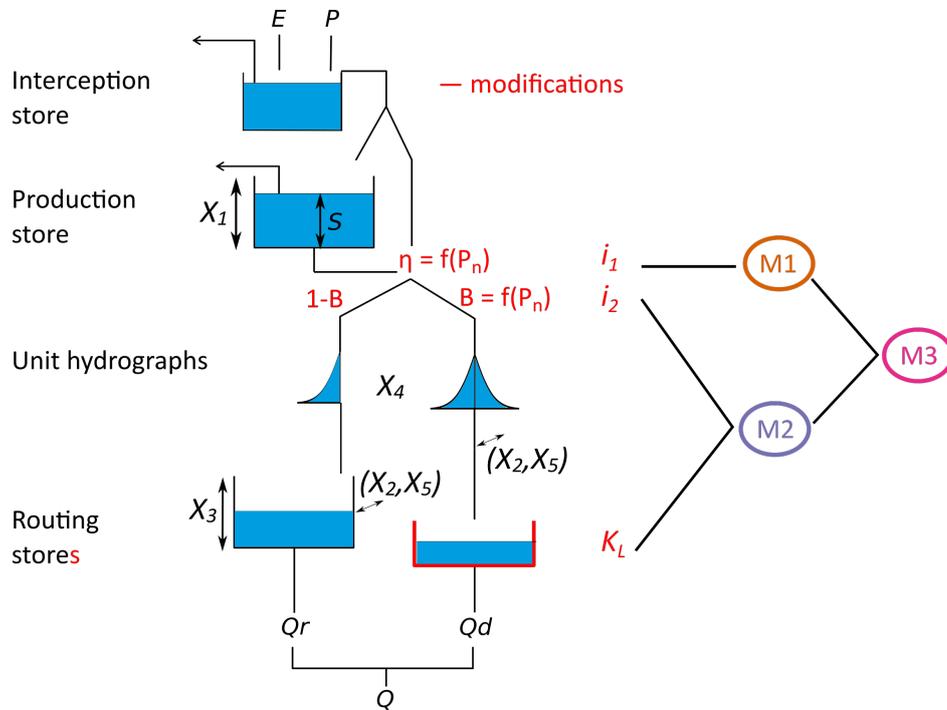


Figure 5.2 – Diagram of the modifications (in red) made to the original GR5H model (in black).

Modeling hypotheses

- *Volume hypothesis*: The first hypothesis is based on the assumption that, at the scale of a rainfall event, when intensities are high and soil moisture is low, the reference model underestimates the effective rainfall volume (part of net rainfall that does not infiltrate the production store).
- *Temporal distribution hypothesis*: The second hypothesis is based on the assumption that there is enough simulated effective rainfall, but that the temporal distribution of effective rainfall should be modified when rainfall intensities are high to allow more of the effective rainfall to reach the outlet during the event.
- The third hypothesis is a combination of the first and second hypotheses.

The concept of "enough effective rainfall" can be seen as fundamentally conceptual: It is only used here to explicate our modeling hypotheses. Also, the effective rainfall volume and its temporal distribution are closely related concepts and their separation depends to some extent on the chosen time window.

Our starting points to construct the first modeling hypothesis are the formulations proposed by Ficchi (2017) and Peredo *et al.* (2022). They tested a modification of the effective rainfall rate calculation within the production store of the GR5H model. Equation 5.3 implies that P_r depends on P_n . However, the instantaneous production rate does not: It only depends on the reservoir filling rate. Therefore, when the production store level is low, the production rate becomes small even for large values of rainfall intensity that will mostly fill the production store. The time required for the production store to reach a level high enough to produce more effective rainfall might be too long in the case of fast flood events under low antecedent soil moisture conditions. To cope with this problem, Ficchi (2017) proposed making the instantaneous production rate dependent on the net rainfall intensity using the following expression (also used but in a different context by Saadi, 2020):

$$\eta' = (1 - \gamma) \cdot s^2 + \gamma \quad (5.4)$$

$$\gamma = 1 - \exp(-P_n \cdot i_1) \quad (5.5)$$

with i_1 an additional parameter in h/mm and γ a rainfall intensity coefficient. This formulation enables the production store to produce more effective rainfall when its level is low but when rainfall intensities are high (Fig. 5.3 and 5.4). When the production store level is high or when rainfall intensities are small, the production rate tends toward the reference GR5H calculation (Eq. 5.3). The function is disabled when $i_1 = 0$.

For the second modeling hypothesis, we made the effective rainfall partitioning between the two branches of the routing function vary with rainfall intensities: The assumption is that for larger rainfall intensities, faster processes occur in the catchment. Therefore, the model needs to route the effective rainfall faster than usual by diverting a higher fraction of P_r to the direct branch (see Andrieu *et al.*, 2021, for an application with geomorphological instantaneous unit hydrographs). Note that this dynamic effective rainfall partitioning into two branches (direct and indirect) is conceptual, which means that it aims to reproduce the catchment behavior rather than any specific local physical process.

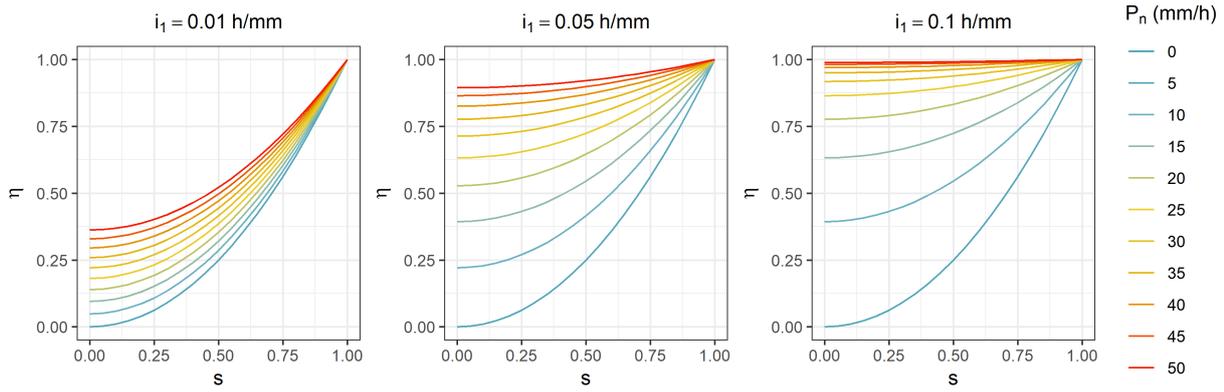


Figure 5.3 – Variation in production rate given a production store level and a rainfall intensity rate (first modeling hypothesis).

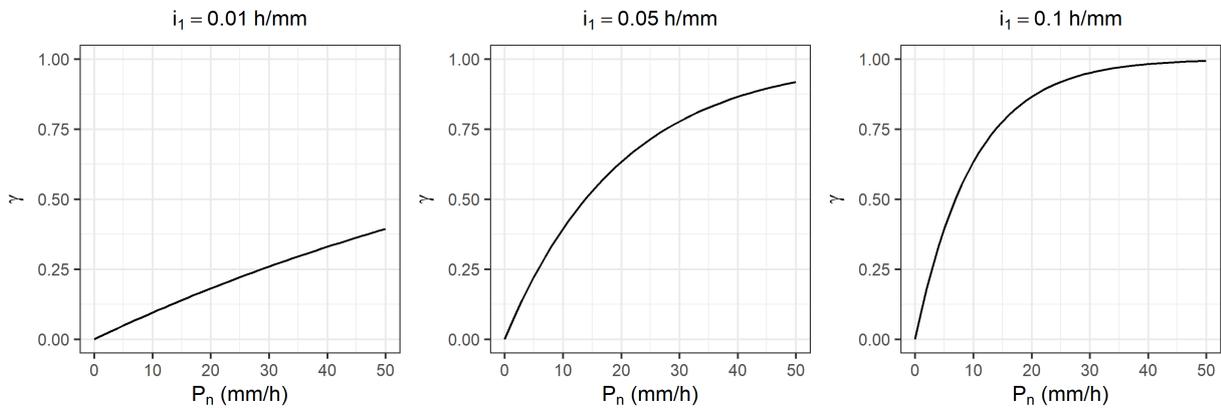


Figure 5.4 – Variation in production rate (first modeling hypothesis) when the production store is theoretically empty ($\eta'(s = 0) = \gamma$).

The partitioning coefficient B is calculated as follows:

$$B = 0.9 \cdot \tanh \left(\left(\sum_{j=t_1}^t P_n(j) \cdot i_2 \right)^2 \right) + 0.1 \quad (5.6)$$

$$t_1 = t - 2 \cdot X_4 + 1$$

with X_4 the UH half-time constant and i_2 an additional parameter in h/mm. When i_2 equals 0, B equals 0.1. This is a way to disable the function when rainfall intensities are low or when there is no rainfall at all. The shape of the intensity function differs from the first modeling hypothesis because B only depends on P_n here. We made the assumption that when intensities are low, B should not vary from its original value. B should also gradually increase with rainfall rates (Fig. 5.5). A rain smoothing function was introduced to place less emphasis on 1-hour rainfall values that can be subject to large uncertainties especially in the case of large rainfall intensities. Therefore B varies with the net rainfall intensity of the current time step and also of the previous time steps.

The shape of the unit hydrograph used in the direct branch UH was modified (exponent changed from 1.25 to 2.5) because of the resulting hydrograph shape when B increases (not shown here; for more details, see Fig. 9.11 of Le Moine, 2008). This change was applied for

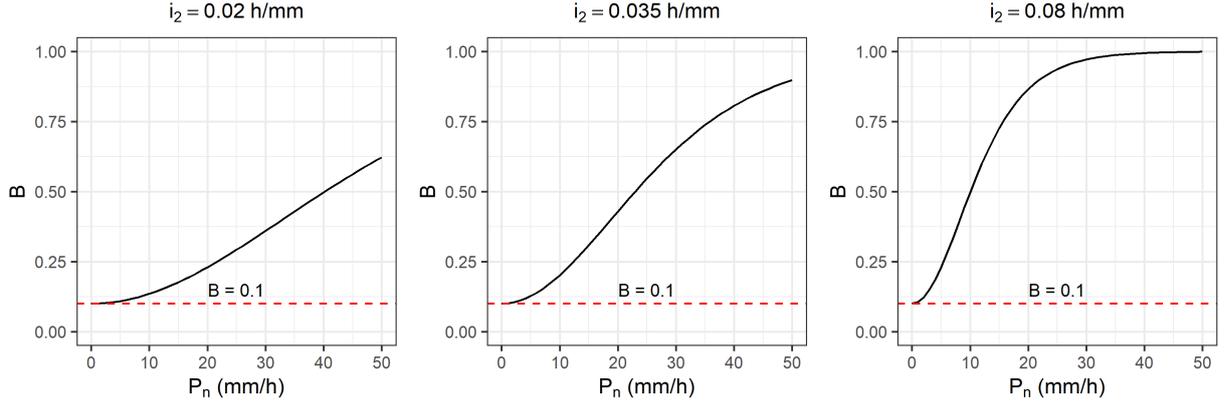


Figure 5.5 – Variation in effective rainfall partitioning with rainfall intensities (second modeling hypothesis).

all three modeling hypotheses. A linear store was added after the UH of the direct branch to improve the transfer of effective rainfall. The outflow of the linear store is calculated as follows (discrete formulation; see Appendix C of Le Moine, 2008):

$$Q_{d,\Delta t} = K_L \cdot L_t \quad (5.7)$$

with $Q_{d,\Delta t}$ the specific volume (mm) exiting the linear store between time $t - 1$ and time t , $K_L \in]0; 1]$ the dimensionless linear store coefficient, and L_t the updated store level.

$$L_t = L_{t-1} + Q_{dHU,\Delta t} \quad (5.8)$$

with $Q_{dHU,\Delta t}$ the specific volume (mm) exiting the direct UH between time $t - 1$ and time t and entering the linear store.

Table 5.2 summarizes the four model versions used in this study. Overall, the proposed modifications increasingly modify model functioning as rainfall intensity increases, and are neutral in the case of low rainfall intensity (and in the case of high soil moisture for the volume hypothesis). In total, approximately 100 versions were tested but are not detailed here for the sake of brevity. They mostly consisted in testing different shapes of the intensity functions, adding a dependency of the partitioning coefficient to the production store filling rate, routing the supplementary effective rainfall by the direct-branch, and changing the shapes of the UHs.

5.3.4 Calibration and evaluation

We estimated the free parameters of the four model versions (see Tab. 5.2) following the method used in Astagneau *et al.* (2021a): The models were calibrated for each catchment on two independent sub-periods P1 and P2 (see Sect. 5.3.1) and the EGD (exhaustive gridding discretization) calibration algorithm (Perrin *et al.*, 2008) was applied considering a warm-up period of 2 years to initialize the states of the models. The EGD method proceeds in two steps: First, a screening of the parameter space is performed to find a starting point from 3^n parameter sets (with n the number of parameters); second, a local-gradient-based optimization is run to find the parameter set that gives the highest performance. The Kling-Gupta efficiency criterion (KGE; Gupta *et al.*, 2008) was used as objective function. The GR5H model was run and modified in the R environment (R Core Team, 2021; Astagneau *et al.*, 2021b) with the `airGR` package (Coron *et al.*, 2017, 2020).

Table 5.2 – List of the four model versions tested.

Modification	Equation(s)	Abbreviation	No. of free param.
None	5.3	M0 (<i>Benchmark GR5H</i>)	5
Production store	5.4 and 5.5	M1-Pr	6
Partitioning coefficient and direct branch	5.6, 5.7 and 5.8	M2-B	7
Production store, partitioning coefficient and direct branch	5.4, 5.5, 5.6, 5.7 and 5.8	M3-Pr-B	8

Note: Pr is the effective rainfall. B is the partitioning coefficient.

Model performance was first evaluated following the evaluation framework of Astagneau *et al.* (2021a): The overall performance in simulating the entire hydrograph is first assessed using the KGE index; then, the performance of the models in simulating floods is evaluated by calculating two performance metrics on each flood event independently – a volume error criterion (or event relative bias, β) and the Nash–Sutcliffe efficiency criterion (NSE; Nash & Sutcliffe, 1970). The catchment mean flow was used as the benchmark for the calculation of NSE on events. The NSE was calculated to cover more properties of the flood hydrograph than the volume efficiency criterion alone. We calculated a bounded version of these criteria to enable a better comparison between catchments and between events (Mathevet *et al.*, 2006). Both bounded criteria take values between -1 and 1. Negative values of β indicate underestimation of the flood volume, positive values indicate overestimation. Negative values of the bounded NSE criterion indicate that the catchment mean flow is a better predictor than the model in simulating a flood event. When the NSE value tends toward 1, the observed and simulated flood hydrographs tend to match perfectly. This bounded criterion enables a better graphical representation of the model results because values of the original NSE criterion that are lower than -1 draw the distribution towards large negative values. This transformation equals zero when NSE is zero. β was also calculated on the streamflow time series without the observed and simulated flood events to investigate possible compensations between high-flow and low-flow periods. We considered two seasons for the evaluation of flood performance. Flood peaks occurring between May and October were labeled as summer floods. Flood peaks occurring between November and April were labeled as winter floods. Flood simulations were then analyzed according to the five groups of catchments defined in Section 5.3.2 and three event characteristics: mean rainfall intensity, spatial variability in cumulative rainfall, and antecedent soil moisture conditions (SWI). All the results are presented using cross-validation simulations, which means that the criteria were calculated on P1 while model simulations were generated with the parameter sets estimated on P2 and vice versa. As the GR5H model was used in a continuous mode, the simulations of the flood events that we have evaluated were not run separately. They were extracted from the two continuous time series (P1 and P2) of model simulations (cross-validation values). We performed the Wilcoxon rank test (Wilcoxon, 1945) to detect statistically significant changes in performance between the three modeling hypotheses and the reference model (at significance level 0.05). The resulting p-values are reported in the text.

5.4 Results

5.4.1 Overview of model performance

First, Figure 5.6 shows the distribution of model performance over the 229 catchments and 10,652 flood events. In calibration mode, for both sub-periods, all three model versions show an improvement over the reference version M0 (p-values of 0.39, 1.8e-3 and 6.0e-4 in the order of model numbering): they have higher median values and a smaller dispersion of the KGE index than the reference model (5.6a). M3-Pr-B exhibits the highest KGE values and M2-B the second highest (see Tab. 5.2 for naming). In validation mode, all three models have similar KGE values to M0. M2-B and M3-Pr-B have higher median KGE values than M0 for both sub-periods (p-values of 0.19 and 0.18 respectively).

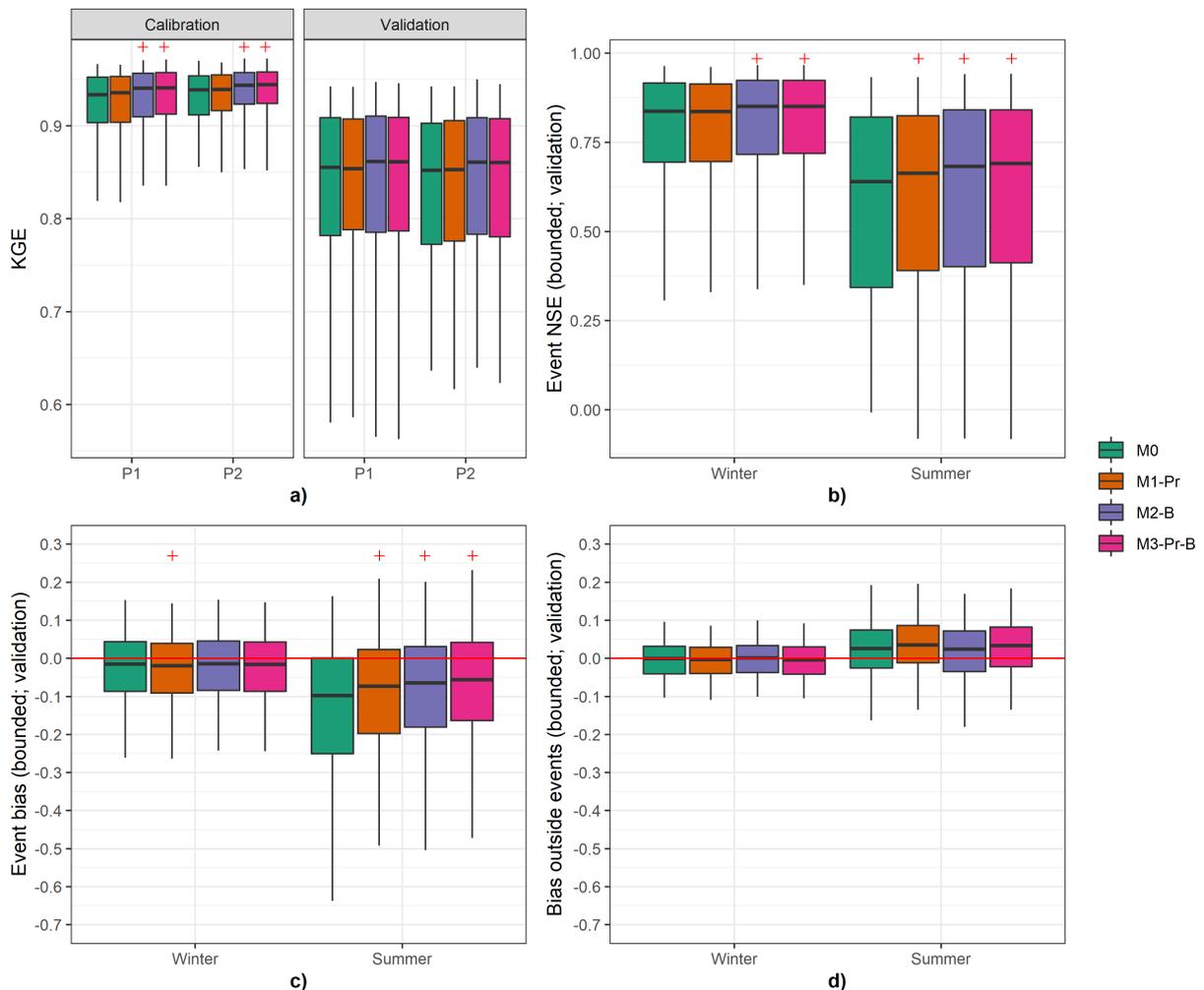


Figure 5.6 – Overall performance (one value for each catchment) (a), performance in simulating 8290 winter events and 2362 summer events (b) and (c), performance outside flood periods (d) for the three modeling hypotheses compared to the reference model (cross-validation values). The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles.

Event performance as assessed by the NSE index increases in summer for all model versions (p-values of 0.03, 7.0e-5 and 7.2e-6 in the order of model numbering) with M3-Pr-B having the highest median NSE values and M2-B the second highest (5.6b). In winter, M1-Pr performs equally to M0 (p-value of 0.50). The median NSE values of M2-B and M3-Pr-B are higher than the median NSE of M0 (p-values of 1.3e-7 and 7.1e-9 respectively). In terms of volume error performance in summer, M3-Pr-B has the median value closest to zero and the smallest dispersion (5.6c; p-value < 2.2e-16). Both M1-Pr and M2-B have median values closer to zero and smaller dispersion of volume error than M0 (p-values of 1.0e-9 and 1.2e-14 respectively), with M2-B having the smallest bias of the two. In winter, the volume error of M2-B and M3-Pr-B is close to the reference model (p-values of 0.80 and 0.32 respectively). There is a slight degradation of the volume error of M1-Pr toward negative values compared to M0 (p-value of 0.01). The distributions of the volume error values calculated outside the periods of flood events show that M1-Pr and M3-Pr-B tend to produce larger overestimations of streamflow in summer (p-values of 0.08 and 0.36 respectively) than M0 and M2-B (5.6d). In winter, the median values of this criterion are close to the reference model with a slight degradation toward negative values for M1-Pr and M3-Pr-B (p-values of 0.40 and 0.50 respectively). M2-B exhibits a similar distribution to M0 (p-value of 0.63). This overview of model performance showed that the proposed model versions have similar overall performance but that simulation of summer floods is improved both in terms of NSE and volume error. Simulation of winter floods is not degraded and is even improved for M2-B and M3-Pr-B. The volume hypothesis leads to an overestimation of volume outside periods of flood events. Overall, the largest improvements in terms of flood performance compared to the reference model are achieved by the combination of the two modeling hypotheses (M3-Pr-B).

5.4.2 Variation in performance within hydroclimatic groups

We refine the analysis by looking at model performance in simulating summer floods considering each group of catchments identified in Sect. 5.3.2. Figure 5.7 shows that summer flood performance increases for all groups of catchments. The smallest increase in summer flood performance is for catchments of group 2 (for the NSE distributions, p-values of 0.92, 0.52 and 0.46 in the order of model numbering). The largest increases in summer flood performance are found for groups 4 and 5 where the three modeling hypotheses perform better than the reference model, especially in terms of event bias. For these two groups, the reference model greatly underestimates flood volumes in summer and yields the lowest flood performance compared to the other three groups. The M3-Pr-B model has the highest flood performance, especially for catchments of group 5 (for the NSE distributions, p-value of 7.4e-4). The volume hypothesis (M1-Pr) performs better in terms of event bias than the temporal distribution hypothesis (M2-B) for catchments of group 5 (p-values of 3.5e-9 and 9.0e-7 respectively). For the other four groups, M2-B has higher flood performance than M1-Pr. We will only focus on M3-Pr-B and catchments of groups 4 and 5 in the analyses that follow.

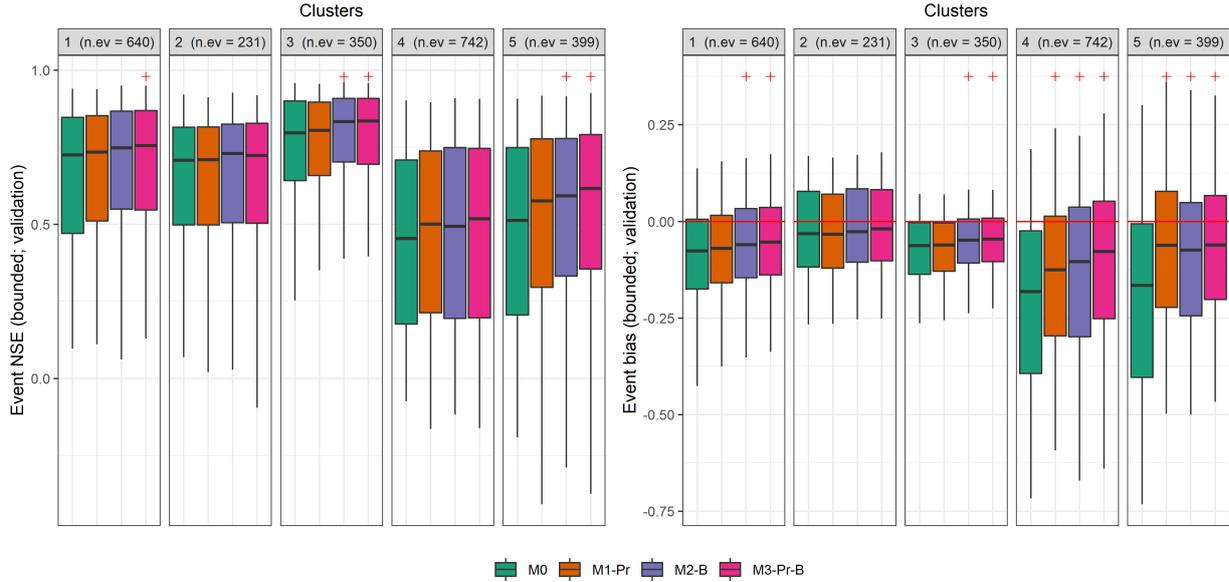


Figure 5.7 – Distribution of flood performance (cross-validation values) over 2362 summer floods divided into five groups of catchments. The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles.

5.4.3 Performance and flood characteristics

Considering all floods occurring in the catchments of groups 4 and 5, we investigate the relationships between bias in simulating floods and three event characteristics (defined in Sect. 5.3.2) for the reference model and the third modeling hypothesis (Fig. 5.8). Figure 5.8a and 5.8b show that the reference model underestimates flood volumes when there are large rainfall intensities. For these events, the event bias values of M3-Pr-B are less dispersed and the median is closer to 0.

For catchments of group 4, the event bias values of the reference model tend to be lower with increasing spatial variability in cumulative rainfall (Fig. 5.8c). The event bias obtained with the new model version follows the same trend but with less dispersed values and a median value closer to 0. The decrease in performance with increasing rainfall spatial variability is less pronounced for catchments of group 5 (Fig. 5.8d). The reference model clearly underestimates flood volumes for both groups of catchments under low antecedent soil moisture conditions. Under these conditions, M3-Pr-B performs better, especially for floods of group 5.

Figure 5.9 shows the simulation of six flood hydrographs for two different Mediterranean catchments with typical streamflow responses to intense rainfall. The flood hydrographs presented on Fig. 5.9 were retrieved from two different sub-periods (validation periods P1 and P2), which means that the parameter values are not necessarily the same between the events (two different calibration periods were used). The Loup River at Tourrettes-sur-Loup (206 km², group 5) and the Estéron River at Sigale (262 km², group 4) are located in southern France (see Fig. 5.1). The Estéron River is a tributary of the Var River, and the Loup River flows directly into the Mediterranean Sea. These catchments are characterized by high seasonal variations in streamflow and steep slopes. Severe floods can occur, usually in winter or in the beginning of autumn and sometimes due to intense convective rainfall events.

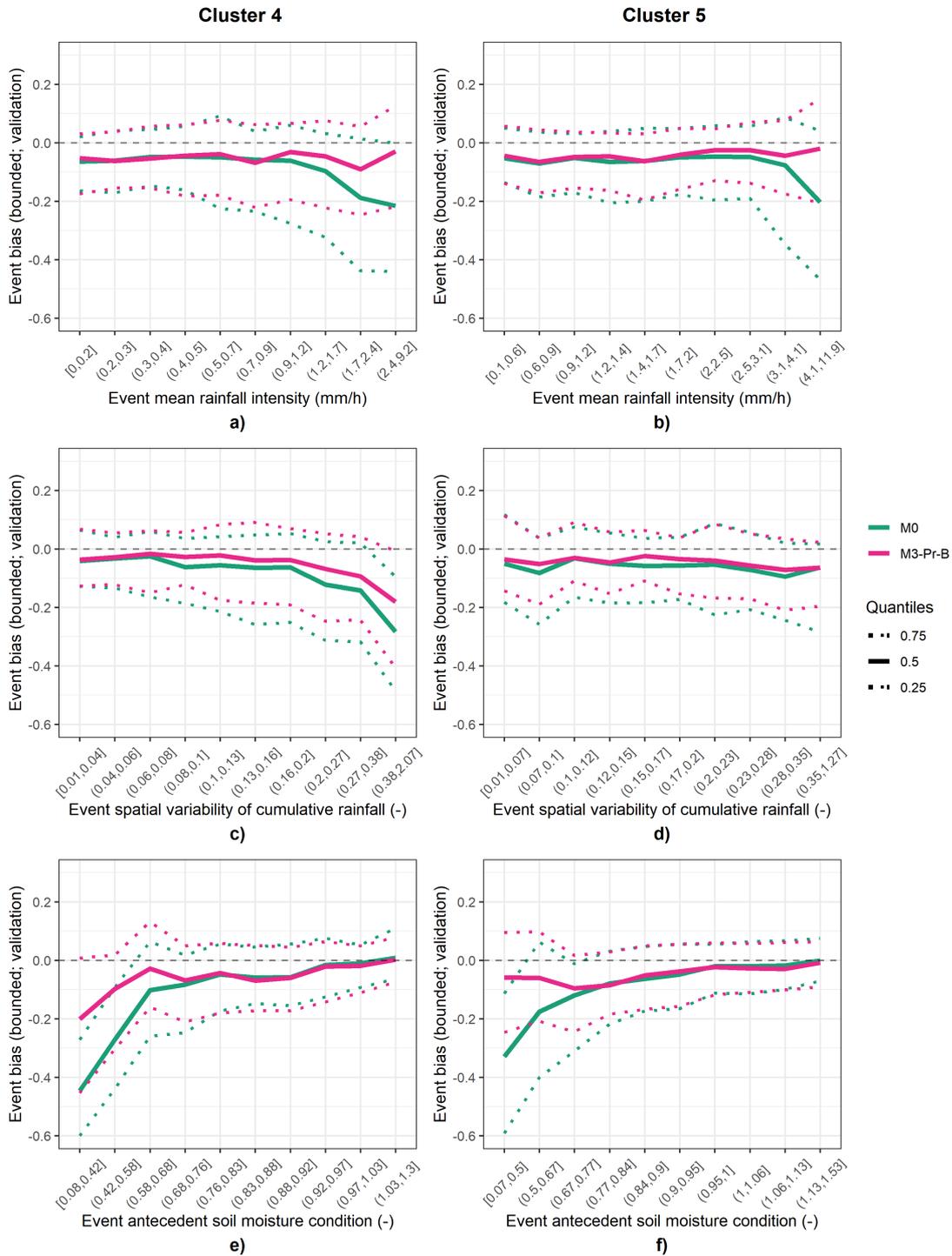


Figure 5.8 – Variation in event performance (cross-validation values) against three event characteristics for two groups of catchments. Floods were divided into 10 quantile classes per characteristic and cluster group. A total of 1932 events for cluster 4 and 1508 events for cluster 5 are considered here.

These examples were selected to illustrate our results and complement the presentation of the scores with boxplots. They cannot be considered as representative of the 10,652 events of our dataset in the sense of exploiting the outputs of a clustering procedure.

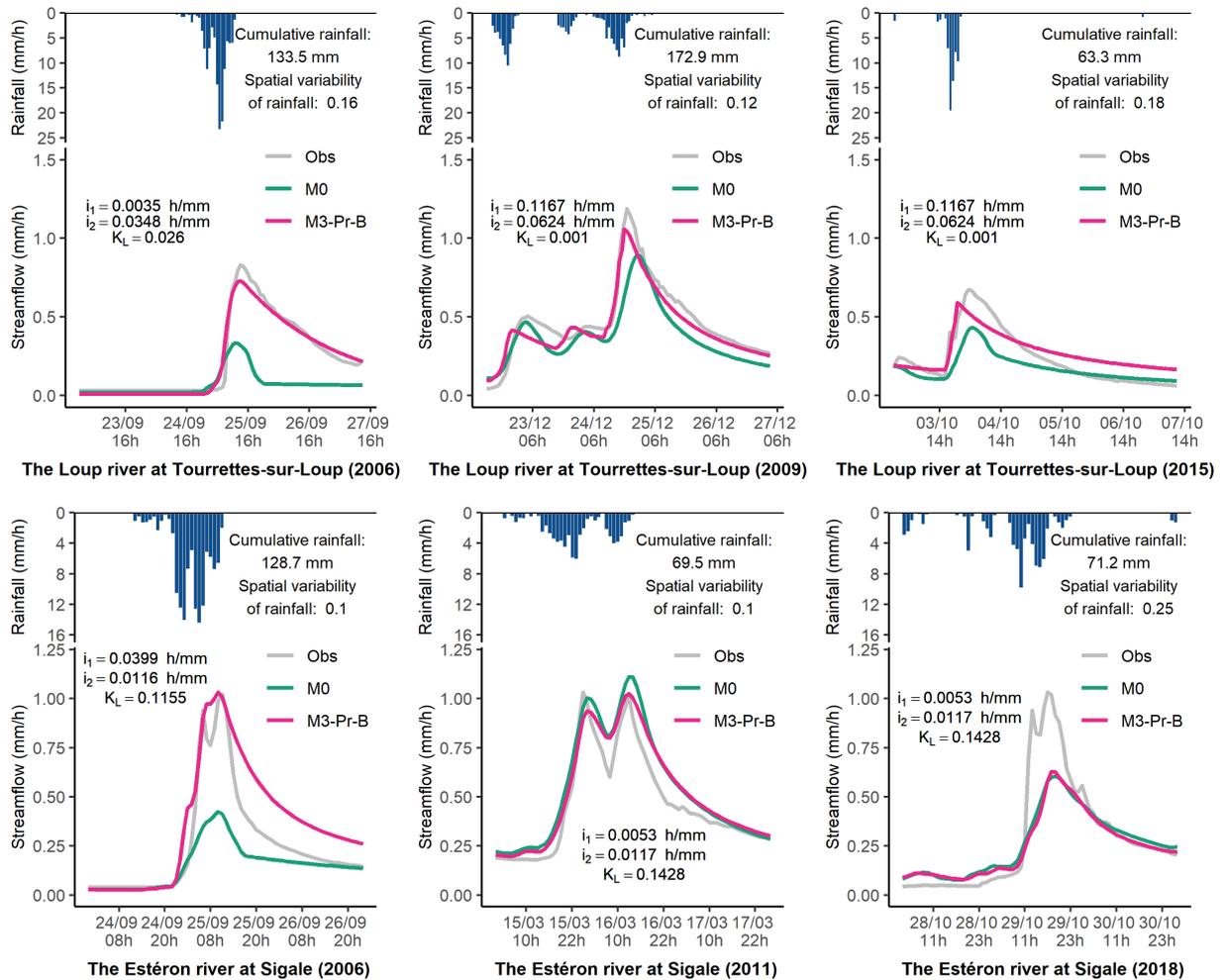


Figure 5.9 – Simulated (M0 and M3-Pr-B) and observed hydrographs of six floods that occurred in the Loup River at Tourrettes-sur-Loup and the Estéron River at Sigale (cross-validation values).

The summer flood of September 2006 that occurred in the Loup River followed an intense rainfall event with intensities of up to 23 mm/h but evenly distributed over the entire catchment area. For this event, the reference model clearly underestimated the peaks and the overall flood volume. The M3-Pr-B on the other hand managed to reproduce the flood volume and the flood peak better, although it is still underestimated. The flood of December 2009 is associated with a larger amount of, but less intense, rainfall. Both models are able to reproduce the hydrograph shape. In the beginning of October 2015, a smaller flood event (in terms of peak and volume) was followed by a short and intense rainfall event. The spatial variability in this rainfall event is slightly larger than the other two events and the total amount of rainfall is almost three times smaller. For this event, the reference model failed to simulate the catchment response, whereas the M3-Pr-B model was able to reproduce the flood peak, volume, and timing better, with a slight overestimation of the recession volume.

The flood of September 2006 that occurred in the Estéron River followed a large and intense rainfall event with intensities of up to 14 mm/h mainly distributed over the entire catchment area. The reference model response to this event was clearly limited compared to the catchment response. The new model version, on the other hand, managed to reproduce the peak and timing of the flood but overestimated the recession volume. The March 2011 flood was associated with a smaller amount of rainfall and smaller intensities but was reproduced well by both models. In October 2018, a rainfall event with cumulative rainfall in the same order of magnitude as the winter flood of 2011, but highly variable in space, led to a fast and intense catchment response. Both models greatly underestimated the flood volume and peak of this event.

5.4.4 Additional model parameters

The most complex model that we have introduced in this study is M3-Pr-B, which has one additional reservoir and three additional free parameters compared to M0. These parameters control the modification of fluxes and storages of the model with varying rainfall intensities. The additional functions we proposed were also made to deactivate the functions in the case of low intensities and, for certain values of i_1 , i_2 and K_L , are equivalent to M0. Figure 5.10 shows that these parameters vary greatly between calibration periods, especially i_1 and to some extent i_2 . For some catchments, i_1 or i_2 is null in one sub-period and above zero in the other, indicating that the intensity function is disabled for one period but not for the other (see Eq. 5.4, 5.5 and 5.6 and the related explanations). The same behaviour is observed for K_L , with several values equal to 1 but with smaller variations between periods and for more catchments.

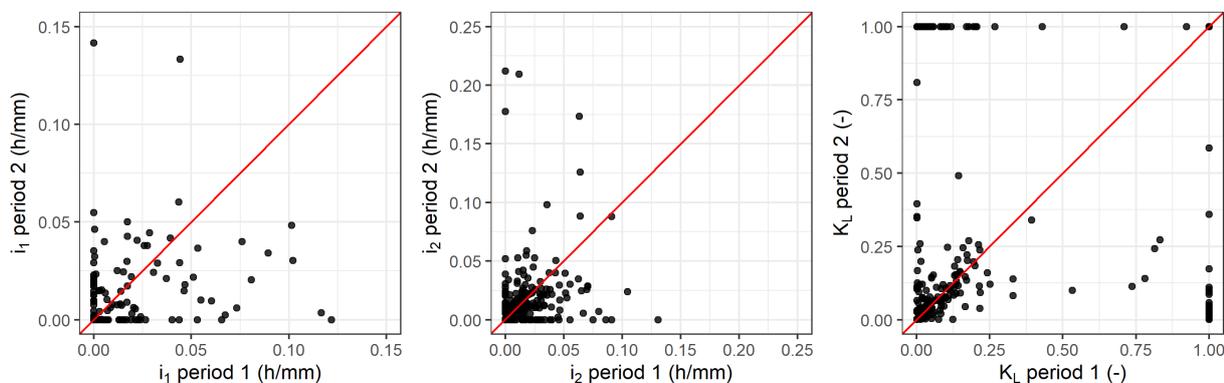


Figure 5.10 – Variation in additional M3-Pr-B free parameters between calibration periods.

As the additional parameters of the third modeling hypothesis vary between calibration periods, the intensity functions are not activated in all periods. As a consequence, on some catchments both the reference model and the new modeling hypotheses are selected as the best performing models. Figure 5.11 shows that the same modeling hypotheses are activated in both calibration periods for 118 catchments of our dataset.

We investigated whether there are differences of parameter distributions between the five hydroclimatic groups of catchments. Results indicate that i_1 follows different distributions for groups 4 and 5 (Fig. 5.12) with more values above zero, especially for catchments of group 5 (85 % of pairs (catchment, sub-period) have i_1 values greater than zero). This means that the first modeling hypothesis is more often activated for catchments of group 5

than for the other catchments. The distributions of i_2 values are very similar between the five groups. The distributions of K_L are different between the five groups. Group 5 has the lowest values of K_L and group 2 has the highest number of values equal to 1 (no linear store on the direct branch).

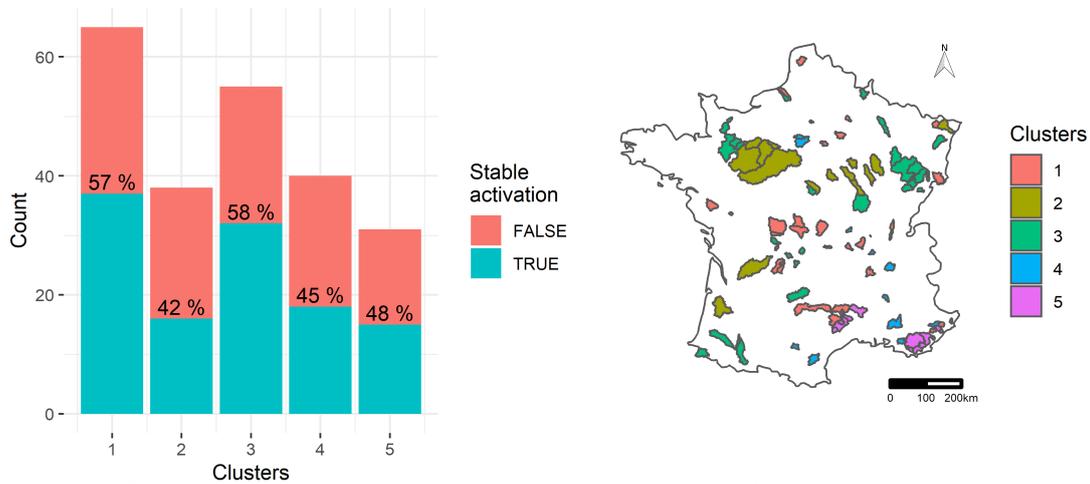


Figure 5.11 – Catchments on which the intensity functions of M3-Pr-B are activated on both calibration periods.

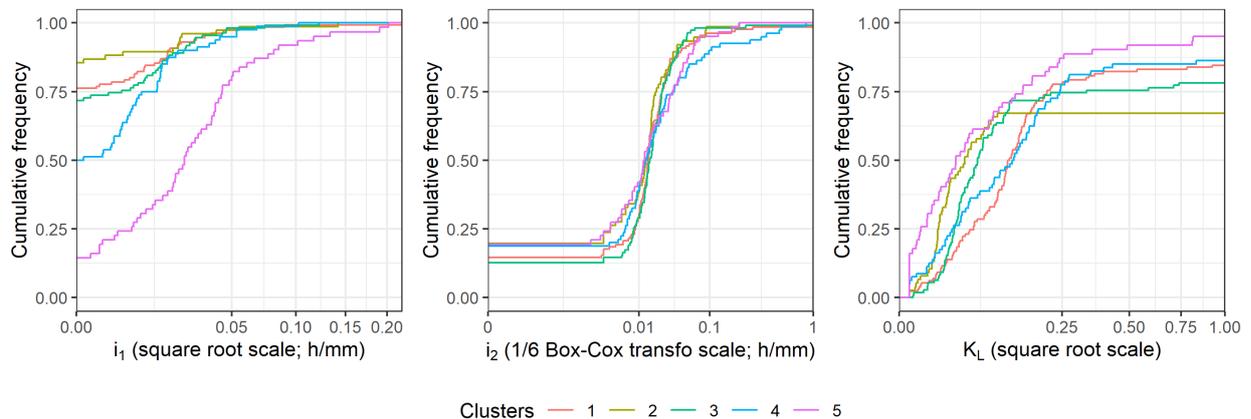


Figure 5.12 – Distributions of the additional M3-Pr-B free parameters for five groups of catchments.

5.5 Discussion

5.5.1 On the available information content to identify additional parameter values

Activation of intensity-rate functions between periods

The previous results indicate that the additional parameters of the proposed intensity functions vary between the two calibration periods, which implies some variations in the intensity function activation. Also, the parameter distributions (Fig. 5.12) and the performance of

the different modeling hypotheses (Fig. 5.7) demonstrate that the additional parameters and their sensitivity to the chosen metrics differ between the groups of catchments. Parameter identifiability can therefore be considered as catchment-dependent. We refer to “parameter identifiability” for issues related to the dependence of parameters to groups of catchments and to the variation of parameter values between calibration periods.

This issue was also raised by van Esse *et al.* (2013) who found that a more flexible model often yields equifinality of possible structures (also demonstrated here as an equifinality of parameters) for a given catchment. One reason can be that on some catchments, most of the flood events that we have selected occur in winter, which suggests that the intensity functions are activated for a very small number of data points. Peredo *et al.* (2022) found that a function similar to the first modeling hypothesis affects the effective rainfall volume simulated on the Aude catchment only for a very few time steps and mostly between July and December. Therefore, the estimation of the added parameters can become very uncertain, especially since there are more parameters to be estimated compared to the initial model structure. Furthermore, as suggested by Astagneau *et al.* (2021a), the KGE index calculated on the whole streamflow time series might not be sufficiently informative to evaluate model performance in simulating summer floods, especially since summer floods are often shorter than winter floods and therefore less data points are available for calculating a performance criterion. Although we used a multi-objective framework for model evaluation, only the KGE index was used for parameter estimation in this study. This parameter identifiability issue could also be related to a possible temporal variability in hydroclimatic conditions between the calibration periods. Finally, there might just not be enough information in the rainfall time series at the hourly time step to derive optimal parameter values at the event scale.

Can we estimate generic parameter values?

To limit the parameter identifiability issue, one solution would be to either fix the values of the additional parameters, or prescribe them using climatic or physical descriptors. Here, we investigated the use of generic values of i_1 and i_2 for the third modeling hypothesis (M3-Pr-B). We selected the combination (i_1, i_2) – for which X_1 to X_5 and K_L were calibrated on P1 and P2 with the KGE objective function – that yielded the maximum mean value of a composite criterion in calibration over the 229 catchments of our dataset. This criterion is calculated as follows for each catchment and calibration period:

$$\begin{aligned}
 E_{comp} &= \frac{1}{2} \left(KGE^b + \overline{NSE}_{ev}^b \right) \\
 KGE^b &= \frac{KGE}{2 - KGE} \\
 \overline{NSE}_{ev}^b &= \frac{1}{n_{ev}} \cdot \sum_{i=1}^{n_{ev}} \frac{NSE_i}{2 - NSE_i}
 \end{aligned} \tag{5.9}$$

KGE^b is a bounded version of the KGE index³. \overline{NSE}_{ev}^b is the average bounded NSE calculated on the events of one catchment and one calibration period. n_{ev} is the number of selected events on one catchment and one calibration period.

$i_1 = 0.002$ and $i_2 = 0.018$ were selected as the best generic parameter values for the intensity functions of M3-Pr-B. Figure 5.13a shows that i_2 (temporal distribution hypothesis) is well identified among the different parameter sets that were tested. Interestingly, an

3. Cette formulation correspond à la formulation du C2M présentée au chapitre 3 (eq. 3.2).

increase in i_1 (volume hypothesis) when i_2 (temporal distribution hypothesis) tends toward its optimal value degrades the composite criterion, whereas an increase in i_2 when i_1 tends toward its optimal value improves the composite criterion. This means that the temporal distribution hypothesis is better adapted to make the GR5H model simulate catchment response to different flood-generating processes across different hydroclimatic conditions and therefore improves its versatility.

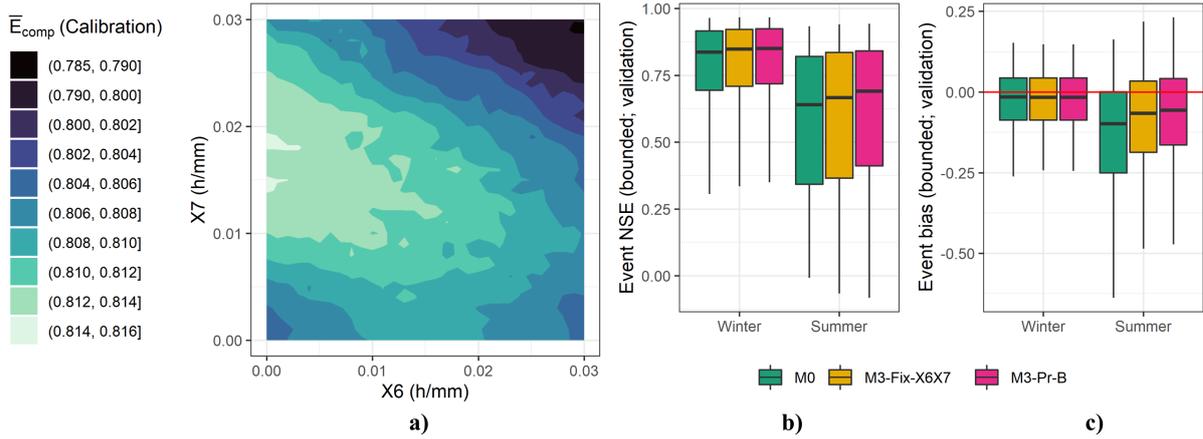


Figure 5.13 – Sensitivity of the additional parameters to the composite criterion E_{comp} over the 229 catchments (a). Distribution of performance in simulating 8290 winter floods and 2362 summer floods for the third modeling hypothesis run with generic i_1 and i_2 values in validation (b and c). M3-Fix- i_1i_2 is the model version with fixed i_1 and i_2 values for all catchments. The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles.

The third hypothesis run with the best generic i_1 and i_2 parameters (M3-Fix- i_1i_2) improves the performance of the reference model M0 for winter and summer events (validation), especially in terms of bias in summer (p-value of $5.3e-15$). However, the performance in simulating floods in validation periods is lower than the fully calibrated version. Looking back at the performance of the fully calibrated volume hypothesis (M1-Pr; Fig. 5.6), we can see that M3-Fix- i_1i_2 achieves the same level of performance as M1-Pr in simulating summer floods but M3-Fix- i_1i_2 seems to be more suited to simulating winter floods. Although the volume hypothesis is almost not activated ($i_1 = 0.002$) when looking for generic values of the additional parameters, the results of Sect. 5.4.2 indicate that this hypothesis is needed to simulate the summer floods of several catchments in cluster 5. However, it might be too costly to activate this function for the other catchments of our dataset where summer flood-generating processes may have different properties at the catchment scale.

Overall, the generic parameters that we have identified can be seen as starting points to improve the ability of the GR5H model to better simulate streamflow in the case of flooding from intense rainfall. Their use enables a reduction of the parameter space during the calibration process, which can be especially useful when not enough data are available or when parameter identifiability is low.

5.5.2 Rainfall intensity and hypothesis testing at the catchment scale

A better temporal distribution of increased runoff production

Our results show that the second modeling hypothesis (temporal distribution hypothesis) performs better than the first hypothesis (volume hypothesis) in terms of overall performance and flood performance over 229 catchments and 10,652 flood events. However, for several catchments of the Rhône River and the Mediterranean arc, the first modeling hypothesis seems to be needed to reproduce flood volume in the case of large intensities, as indicated by the event bias values for summer events in cluster 5 (Fig. 5.7). The more frequent positive values of the associated additional parameter (i_1 ; Fig. 5.12) for the third hypothesis (combination of the first and the second hypotheses) confirm that the volume hypothesis is needed to simulate summer floods in cluster 5. For these catchments, this could mean that under low antecedent wetness conditions, the runoff coefficient is higher in the case of intense rainfall events. Therefore, improving the temporal distribution of effective rainfall is not sufficient to simulate the flood hydrograph. However, we expect the model to be able to simulate more volume at the event scale with the first hypothesis, as the water balance is affected at the event scale in that case whereas this is not the case for the second hypothesis. This increased volume in the case of large intensities results in larger flow overestimation outside flood periods (Fig. 5.6d), which indicates that the temporal distribution of the additional volume simulated at the event scale should be improved as well.

Here, we made the assumption that the effective rainfall volume produced by the model can be separated from its temporal distribution. This is certainly not always true, as a more delayed distribution of effective rainfall would result in a lower simulated flood volume and therefore in a lower simulated runoff coefficient at the event scale. The performance of the third modeling hypothesis (e.g. Figs. 5.6b and 5.6c) shows that the model needs both an increase in runoff production (first hypothesis) and a faster transfer at the event scale (second hypothesis). The third modeling hypothesis adapts to more heterogeneous flood-generating processes in summer while keeping its ability to simulate floods under wetter conditions.

Overall, the results indicate that more complexity is needed to simulate floods of catchments in groups 4 and 5. Conversely, the intensity functions do not seem to be needed for catchments in cluster 2. The catchments in cluster 2 are associated with the most constant hydroclimatic conditions between events (low variability in rainfall intensities, low variability in rainfall spatial variability, low variability in antecedent wetness conditions, low variability in flood duration) and with the slowest catchment response to rainfall (high time lag and high flow autocorrelation at 24 h) indicating very uniform flood-generating processes that the reference model seemed to capture well. This confirms that a robust parsimonious model can perform well under wetter and more uniform hydroclimatic conditions without unnecessary complexity (e.g. Jakeman & Hornberger, 1993), but some adaptation of its structure might be needed to simulate streamflow response to more heterogeneous processes (e.g. Knoben *et al.*, 2020), especially at the hourly time step (Ficchi *et al.*, 2019).

Spatial variability in rainfall intensities

The analysis of performance in simulating floods in clusters 4 and 5 depending on three causative processes further indicates the value of our modeling hypotheses in a Mediterranean context. Figures 5.8a and 5.8b indicate that the floods following the 10 % most intense

rainfall events are less underestimated by the new model version. The upper part of the error distribution also indicates that there might be a small bias towards overestimation of 25 % of these events. Extreme rainfall events, especially in summer, are known to be subject to large uncertainties (e.g. Ruiz-Villanueva *et al.*, 2012; Zhang *et al.*, 2017) in their estimation, making the interpretation of the upper part of the error distribution difficult to establish.

Overall, the results show that bias in simulating floods associated with high-intensity rainfall events under low antecedent soil moisture conditions is reduced for many catchments located in the Mediterranean area. However, we noticed that for catchments in cluster 4, and for events with large values of the spatial variability index, the new model still largely underestimates flood volumes, while this is not the case for the catchments in cluster 5. This result can first appear counter-intuitive since we used a spatially lumped model in our study. But looking at the cluster's hydroclimatic properties (Fig. 5.1 and 5.14), we can see that there is a higher variability in the rainfall spatial variability coefficient between events for catchments in cluster 4 than for catchments in cluster 5. This could mean that the lumped GR5H model is better able to reproduce flood volumes when rainfall events usually affect the same area of the catchment. This explanation is consistent with the findings of Lobligeois *et al.* (2014), who showed that the impact of spatial resolution of rainfall on model performance is catchment-dependent. Another explanation could be that the flood events associated with high spatial variability for catchments in cluster 4 are not associated with very intense rainfall at the catchment scale for the whole duration of the events, whereas the intensities associated with the flood events of catchments in cluster 5 in the case of large values of spatial variability are sufficient to make the new model simulate more volume. These low intensity values could also be the result of large uncertainties that are sometimes associated with short-duration rainfall events (e.g. Ruiz-Villanueva *et al.*, 2012; Zhang *et al.*, 2017). Overall, separating rainfall spatial variability from rainfall intensity allowed us to highlight the benefit of making better use of rainfall intensity information for flood simulation. However, many rainfall events occurring in summer are intense and highly variable in space, and the benefits of using a finer spatial resolution remains to be investigated in order to generalize the findings of some recent studies (e.g. Loritz *et al.*, 2021; Peredo *et al.*, 2022).

5.6 Conclusions

The objective of this study was to use rainfall intensity as a proxy for the activation of fast hydrological processes in order to improve the simulation of summer floods by a conceptual lumped rainfall–runoff model. We worked at the catchment scale, with a large catchment set, and made three modeling hypotheses on the dependence of the fluxes of our model to hourly rainfall intensity rates:

- The first hypothesis consists in assuming that large rainfall intensities increase the volume of effective rainfall.
- The second hypothesis consists in assuming that large rainfall intensities induce a faster routing of effective rainfall to the catchment outlet.
- The third hypothesis combines the first two hypotheses, modifying both the volume and the routing of effective rainfall.

Our results showed that the three hypotheses increased the ability of the hydrological model to simulate summer floods, especially in terms of volume error, and particularly for tributaries of the Rhône River and catchments located in the Mediterranean area. The third hypothesis shows the highest performance in capturing the streamflow response to more heterogeneous storms in summer, while maintaining good performance in simulating winter floods. We name this new model GR5H-RI. Our results indicate that there is a clear benefit in introducing a dependency of the storages and fluxes of a lumped conceptual model on rainfall intensities (at least at the hourly time step). Our work obviously has some limitations; the lumped configuration tested here may be restrictive for some catchments, especially where the spatial organization of rainfall varies between events. Furthermore, the values of the additional parameters are difficult to estimate, because the intensity-dependent functions are only activated on very few time steps. To overcome this issue, we proposed default values for two of the additional parameters, which cause a decrease in model performance in comparison to the fully calibrated model, but still enable a better simulation of floods in comparison to the reference model. Regionalizing the parameters of the intensity-dependent function (with relationships between these parameters and catchment meteorological properties) would be another solution.

Our hope is that a better simulation of floods from intense rainfall will lead to improvements in flood forecasting, providing that our functions are compatible with the assimilation of real-time rainfall and streamflow data. There is a potential for using the proposed modeling hypotheses in a Mediterranean context, especially since an intensification of extreme rainfall events is to be expected with climate change in some parts of this region (e.g. Trambly & Somot, 2018). Further tests will be conducted in future studies to evaluate the value of these functions to simulate and forecast extremes, for example by using post-estimations of peak discharges of extreme events that occurred in Mediterranean catchments (e.g. the 2nd October 2020 Alex storm in south-eastern France; Payrastra *et al.*, 2022). Finally, even if this paper only discusses the application of the three above modelling hypotheses to the GR5H structure, we believe that they can be applied to other models. Depending on the specificities of each model, the first, second or third hypothesis may prove to be the most suitable.

5.7 Compléments

5.7.1 Autres modifications testées

Plusieurs versions du modèle GR5H (dont celles de Ficchi, 2017; Saadi, 2020; Peredo, 2021) ont été testées pour aboutir aux hypothèses de modélisation présentées dans ce chapitre. Le tableau 5.3 présente les modifications principales que nous avons effectuées dans la structure du modèle.

TABLE 5.3 – Liste des modifications principales non présentées dans ce chapitre. Différentes combinaisons entre les modifications ont été testées.

Modification	Niveau de la modif.
Modification du rendement à partir de l'équation 2 de Peredo <i>et al.</i> (2022)	Production
Modification du rendement à partir de l'équation 2 de Saadi <i>et al.</i> (2021)	Production
Modification du rendement à partir d'une fonction sigmoïde à un paramètre	Production
Modification du rendement à partir d'une fonction sigmoïde à deux paramètres	Production
Routage de la surproduction par la branche pseudo-directe	Prod. et rout.
Utilisation d'une structure à deux hydrogrammes unitaires (cf. Ficchi, 2017)	Routage
Modification dynamique du partage à partir d'une fonction sigmoïde à deux paramètres	Routage
Optimisation du paramètre d'asymétrie de l'hydrogramme unitaire de la branche pseudo-directe	Routage
Modification de l'exposant de l'hydrogramme unitaire de la branche pseudo-directe	Routage
Modification dynamique du partage en fonction de la pluie efficace	Routage
Modification dynamique du partage en fonction du taux de remplissage antécédent du réservoir de production	Routage
Modification dynamique du partage en fonction du taux de remplissage antécédent du réservoir de production et de l'intensité de la pluie	Routage
Ajout d'une troisième branche de routage avec un réservoir exponentielle (cf. Pushpalatha <i>et al.</i> , 2011)	Routage

Les résultats obtenus à partir de ces modifications ont montré que la modification de certains flux et certains stocks du modèle à partir de l'intensité de pluie horaire a un impact de premier ordre sur la simulation des crues. Le choix du niveau de la modification (e.g. routage ou production) a un impact de deuxième ordre. La forme des fonctions associées a souvent un impact de troisième ordre. Nous introduisons quatre de ces modifications dans les paragraphes qui suivent. Les raisonnements qui sont à la base de ces quatre modifications ont eu un impact significatif sur la formulation des fonctions présentées en 5.3.3.

Modification du rendement à partir d'une fonction sigmoïde à un paramètre

Les équations 5.4 et 5.5 ont été retenues pour modifier le rendement du réservoir de production. Elles sont illustrées par les figures 5.3 et 5.4. Ces formulations impliquent que l'augmentation du rendement en fonction de l'intensité de la pluie est activée même pour de faibles intensités de pluie. Par ailleurs, la forme de la fonction exponentielle associée implique que plus l'intensité de pluie augmente, moins l'augmentation du rendement est élevée (cela ne signifie pas que le rendement est moins élevé pour des intensités plus fortes, seulement que, pour un niveau du réservoir donné, la différence de rendement entre 1 mm/h et 2 mm/h d'intensité de pluie est plus élevée que la différence de rendement entre 2 mm/h et 3 mm/h d'intensité de pluie). Ce constat nous a incité à tester plusieurs formulations de l'équation 5.5. Une de ces fonctions alternatives est similaire à l'équation 5.6 (sans la fonction de lissage de la pluie). La réécriture de l'équation 5.5 à partir de la fonction sigmoïde ainsi utilisée s'écrit de la manière suivante :

$$\gamma = \tanh(P_n \cdot i_1)^2 \quad (5.10)$$

Le principe de cette fonction à un paramètre est que la modification du rendement de la pluie n'est pas activée pour les faibles intensités de pluie. Par ailleurs, l'augmentation du rendement avec l'intensité de pluie s'effectue de manière plus progressive.

L'utilisation de cette fonction (à la place de l'équation 5.5) donne lieu à une baisse des performances du modèle M1-Pr sur l'échantillon de bassins versants. Cette baisse de performance s'explique dans une certaine mesure par une activation de la fonction d'intensité sur moins de pas de temps (à cause du seuil de « démarrage » de la fonction). De plus, la fonction s'active moins rapidement et de manière plus brutale sur les fortes intensités horaires, induisant ainsi un comportement plus erratique du rendement.

Routage de la surproduction par la branche pseudo-directe

Les résultats des chapitres 3 et 4 ont montré que le routage du modèle M0 ne permet pas de reproduire le volume de certains événements de crue. Avant de développer le modèle M2-B, nous avons testé une version alternative à M1-Pr qui consiste à effectuer le routage de la pluie efficace « additionnelle » par la branche pseudo-directe du modèle. Cela permet théoriquement de réduire le temps de transfert de cette pluie efficace additionnelle. À chaque pas de temps, on peut calculer : la pluie efficace de référence P_{r0} calculée à partir de l'intégration de l'équation 5.3; la pluie efficace P_r à partir de l'intégration de l'équation 5.4. Router la pluie efficace « additionnelle » par la branche pseudo-directe revient à ce que le partage B dépende du ratio P_{r0}/P_r . En gardant le principe selon lequel B revient à son état initial (i.e. $B = 0$) lorsque la fonction d'intensité n'est pas activée, on peut écrire :

$$B = 1 - 0.9 \times \frac{P_{r0}}{P_r} \quad (5.11)$$

Les performances obtenues avec cette fonction sont inférieures aux performances de M1-Pr. La fonction d'intensité est activée pour moins de bassins versants. Les valeurs de i_1 sont généralement plus faibles que celles obtenues avec la version initiale de M1-Pr.

Asymétrie de l'hydrogramme unitaire de la branche pseudo-directe

Le modèle M2 présenté dans ce chapitre contient un réservoir linéaire au niveau de la branche pseudo-directe, après l'hydrogramme unitaire (HU) symétrique. L'ajout de ce réservoir est le résultats de plusieurs essais sur le routage de la pluie efficace par cette branche. Lorsque la part de la pluie efficace qui passe par la branche pseudo-directe augmente, l'hydrogramme unitaire seul n'arrive pas à effectuer le routage correctement. Avant d'ajouter un réservoir linéaire en aval de l'HU, nous avons cherché à adapter l'HU. Un des tests principaux a été d'asymétriser (au sens de la section 6.3.1 et de la figure 6.1 de Berthet, 2010) la forme de l'HU en introduisant un paramètre supplémentaire dans le calcul de la courbe en S au temps t (notée $SH(t)$; voir eq. 1.10) :

$$\begin{cases} 0 & t < 0 \\ \frac{1}{2} \times \left(\frac{t}{X_4 \times (1 + X_{asym})} \right)^{2.5} & 0 < t < X_4 \times (1 + X_{asym}) \\ 1 - \left(\frac{1}{2 \times (1 - X_{asym})^{2.5}} \right) \times \left(2 - \frac{t}{X_4} \right)^{2.5} & X_4 \times (1 + X_{asym}) < t < 2 \times X_4 \\ 1 & t \geq 2 \times X_4 \end{cases} \quad (5.12)$$

X_{asym} est le paramètre d'asymétrie à caler. Comme mentionné dans la section 5.3.3, l'exposant de cet hydrogramme unitaire est égal à 2.5 (contrairement au modèle de référence; voir eq. 1.10). Les performances obtenues avec cette fonction sont inférieures aux performances de M2-B bien que supérieures aux performances de M2-B sans réservoir linéaire. Cette baisse de performance s'explique dans une certaine mesure par des formes erratiques de certains hydrogrammes de crue.

Modification du partage en fonction du réservoir de production et de l'intensité de la pluie

Nous avons montré dans les chapitres 2, 3 et 4 que le modèle M0 n'arrive pas à reproduire le volume des crues qui ont lieu en condition de faible humidité antécédente des bassins et lorsque les intensités de pluie sont élevées. Ce constat a motivé les choix de modification de la structure présentées dans ce chapitre. Cependant, le coefficient de partage dynamique que nous avons proposé dépend uniquement de l'intensité de la pluie nette. Nous avons testé une version des modèles M2-B et M3-Pr-B (sans fonction de lissage de la pluie) dans laquelle le coefficient B dépend à la fois de l'intensité de la pluie et du niveau antécédent du réservoir de production. Dans cette version, la modification de B est « désactivée » lorsque le niveau du réservoir de production est élevé. L'équation 5.6 devient ainsi :

$$B = 0.9 \cdot \tanh(P_n \cdot i_2)^2 \times (1 - s)^\alpha + 0.1 \quad (5.13)$$

s est le taux de remplissage du réservoir de production et α une constante. Plusieurs valeurs de α ont été testées. Les performances de M2-B et M3-B-Pr pour les crues d'été ne sont pas améliorées avec cette version plus complexe. Les performances sont dégradées en hiver mais restent similaires à celles de M0. Par ailleurs, cette version du modèle induit des sur-réactions en étiage pour plusieurs bassins.

5.7.2 Caractéristiques de la classification de bassins

La figure 5.14 présente les distributions des caractéristiques hydroclimatiques – non présentées en figure 5.1 – qui sont utilisées pour le clustering de bassin (cf. tab 5.1).

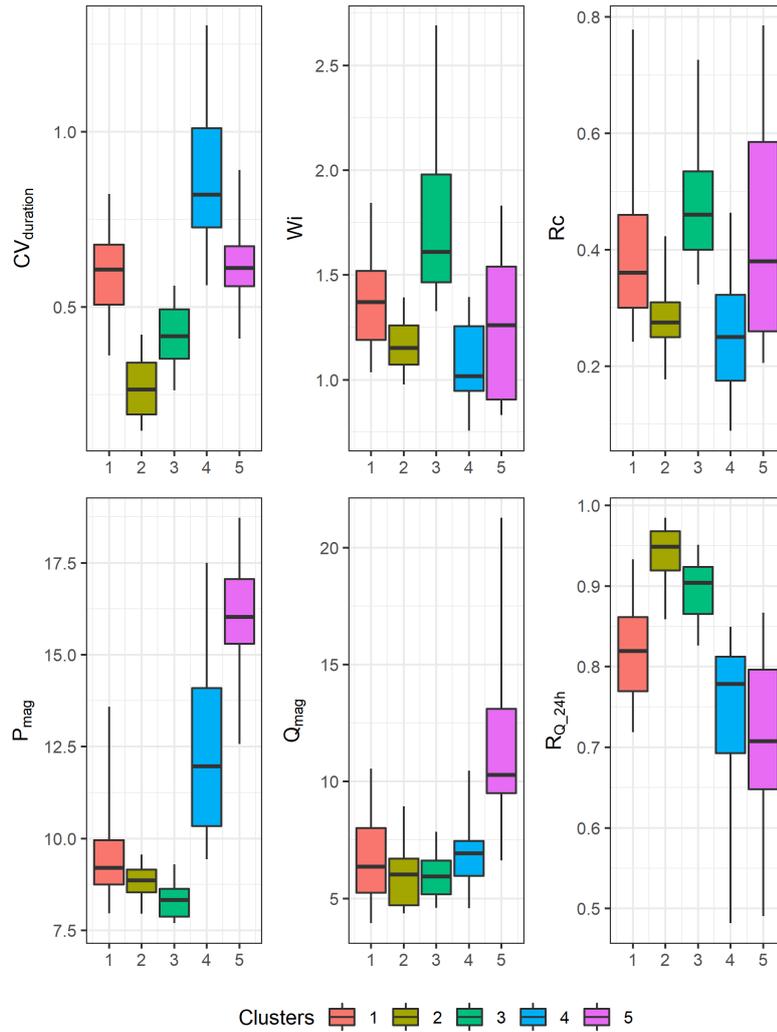


FIGURE 5.14 – Distribution de six caractéristiques hydroclimatiques pour les cinq groupes de bassins versants. Les distributions sont présentées entre les centiles 5 % et 95 %.

Chapitre 6

Que peut-on attendre d'une modélisation semi-distribuée ?

6.1 Contexte

Le diagnostic des difficultés du modèle GR5H présenté dans la partie I (cf. chapitres 2 et 3) a montré que les caractéristiques spatio-temporelles des événements de pluie ont un impact sur les performances en crue. En particulier, les événements qui ont lieu en été et au début de l'automne sur certains bassins sont majoritairement sous-estimés par le modèle. Nous avons vu que certains des événements de pluie associés à ces crues sont intenses et peu variables dans l'espace, que d'autres présentent une forte variabilité spatiale mais des intensités plus faibles, mais que beaucoup de ces événements sont à la fois intenses et variables dans l'espace. Dans le chapitre précédent, nous avons tenté de palier ces déficiences en exploitant l'information donnée par l'intensité de pluie horaire à l'échelle du bassin versant. Assez logiquement, les gains de performance les plus significatifs ont été obtenus pour des événements de crue résultant de pluies très intenses à cette échelle. Toutefois, de nombreux événements de crue restent sous-estimés par le modèle. Une part importante de ces événements sont associés à des valeurs élevées de variabilité spatiale des cumuls de pluie mais aussi à des intensités de pluie élevées au regard du bassin considéré.

Une manière de prendre en compte l'impact de la variabilité spatiale de la pluie dans la réponse d'un bassin versant consiste à discrétiser celui-ci en sous-unités spatiales homogènes. Ces sous-unités peuvent être des sous-bassins versants comme dans le modèle semi-distribué GRSD (Lobligeois *et al.*, 2014 ; de Lavenne *et al.*, 2019), des unités hydrologiques présentant des caractéristiques communes (HRU pour *Hydrological Response Units* en anglais) comme dans le *Dynamic TOPMODEL* (Beven & Freer, 2001a), ou encore des mailles régulières comme dans les modèles *Grid-to-Grid* (G2G) (Bell *et al.*, 2007) et SMASH (Jay-Allemand *et al.*, 2020). Ces sous-unités sont connectées entre elles, par exemple par une fonction de délai qui peut dépendre de la distance entre les exutoires des sous-bassins versants ou par des fonctions de transfert plus complexes entre les différents réservoirs associés à chaque maille ou HRU. Quelle que soit la complexité de la méthode de discrétisation spatiale utilisée, une des hypothèses derrière la modélisation spatialisée est que la variabilité de la réponse hydrologique au sein d'un bassin versant joue un rôle de premier ordre sur la variabilité des débits à l'exutoire principal de celui-ci. Certaines crues rapides en Méditerranée sont par exemple régies par des processus de saturation très localisés (e.g. Douinot *et al.*, 2018). Plus généra-

lement, les interactions spatiales entre la pluie et les conditions antécédentes d'humidité des sols ont un impact sur les processus de genèse des crues (e.g. Tarasova *et al.*, 2020).

Il existe de nombreux travaux qui portent sur l'impact de la discrétisation spatiale de la pluie sur les performances des modèles pluie-débit. Lobligeois *et al.* (2014) expliquent que les conclusions sont contrastées dans la littérature car beaucoup d'études ont porté sur des bassins et des événements en particulier. À partir d'un grand échantillon de bassins versants français, ils montrent que l'impact de la discrétisation spatiale sur les performances en crue du modèle GRSD dépend du bassin et de l'événement considéré. Les différences de performance avec la version globale du modèle restent généralement minimales. Cependant, les performances en crue augmentent pour des bassins méditerranéens et des bassins situés dans les Cévennes, en particulier lorsque la variabilité spatiale de la pluie est élevée.

Loritz *et al.* (2021) expliquent que l'absence de supériorité (parfois même les performances dégradées) des modèles distribués sur les modèles globaux pour simuler les débits à l'exutoire des bassins versants peut être expliquée par la difficulté à paramétrer les modèles distribués. En effet, lorsque les paramètres varient entre sous-unités spatiales, les modèles distribués souffrent de problèmes de sur-paramétrisation (ou équifinalité Beven, 1993 ; Andréassian *et al.*, 2012). Le nombre de paramètres à estimer devient trop grand pour trouver un jeu de paramètres unique au calage (conduisant à des surfaces de réponse très plates et un grand nombre de jeux de paramètres équivalents). Ils expliquent que les bassins versants sont des filtres passe-bas qui lissent le signal porté par la pluie. Pour certains événements, la multiplication de la structure par le nombre de mailles spatiales devient ainsi redondante pour simuler le débit. Cependant, lorsque les processus hydrologiques qui interagissent à des échelles spatiales fines dominent la réponse du bassin, par exemple en été lors d'événements convectifs, la discrétisation spatiale des modèles devient nécessaire pour modéliser le débit. Cette problématique s'inscrit plus largement dans la difficulté qu'ont les hydrologues à décrire les processus hydrologiques en fonction des échelles spatiales considérées (e.g. Blöschl & Sivapalan, 1995 ; McDonnell *et al.*, 2021). En d'autres termes, on peut se demander si la structure d'un modèle global est adaptée à la sous-unité spatiale considérée. Par exemple, la courbe de rendement du modèle GR5H n'est peut-être pas adaptée pour une sous-unité spatiale correspondant au bassin intermédiaire. De même, l'hydrogramme unitaire est potentiellement redondant avec la fonction de routage hydraulique dans le modèle GRSD (voir par exemple Lerat *et al.*, 2012 ; Lobligeois, 2014, pour des travaux sur ce sujet). Les fonctions d'intensité que nous avons développées dans le chapitre précédent pourraient se révéler être une manière indirecte de « capter » la variabilité sous-maille (temporelle et/ou spatiale) de la réponse du bassin lors d'événements intenses et fortement variables dans l'espace.

L'objectif de ce chapitre est de déterminer si la prise en compte explicite de la variabilité spatiale de la pluie permet d'améliorer les performances en crue du modèle GR5H au regard des améliorations déjà apportées par les fonctions d'intensité développées dans le chapitre précédent. On cherchera en particulier à évaluer si les améliorations apportées par les fonctions d'intensité sont redondantes par rapport à la distribution spatiale de la pluie. Les travaux de Peredo *et al.* (2022) avaient notamment mis en évidence l'intérêt de combiner les deux approches sur le bassin versant de l'Aude.

6.2 Éléments méthodologiques

Nous utilisons le modèle semi-distribué GRSD développé par Lobligeois (2014) pour prendre en compte explicitement la variabilité spatiale de la pluie (fig. 1.8). Ce modèle est basé sur la structure du modèle GR5H utilisé dans la première partie de notre travail (structure avec un seul hydrogramme unitaire). Chaque bassin versant est divisé en sous-bassins versants selon la direction d'écoulement et le choix de la taille des unités spatiales (pour une description de la méthode, voir Le Moine, 2008). Le modèle GR5H est appliqué sur chaque sous-bassin versant avec le module d'interception développé par Ficchi *et al.* (2019). Les débits simulés sur chaque sous-bassin amont sont combinés à partir d'un module de propagation hydraulique linéaire, consistant en une simple translation temporelle basée sur un paramètre de célérité (C en m.s^{-1}) et la distance hydraulique entre exutoires (L_h en m) afin de calculer le débit de l'exutoire jaugé en aval. Les précipitations et l'ETP sont moyennées indépendamment sur chaque sous-bassin à partir des produits spatialisés bruts (COMPEHORE et SAFRAN).

Comme Lobligeois *et al.* (2014), nous effectuons un calage global, c'est-à-dire que tous les sous-bassins versants ont le même jeu de paramètres. L'hypothèse que nous faisons ici est que la variabilité spatiale des caractéristiques physiques des bassins a un rôle moins important que celui de la variabilité spatiale des pluies (Andréassian *et al.*, 2004). Cela permet de s'affranchir des problèmes de calage des modèles semi-distribués et de se concentrer sur l'impact de la distribution spatiale de la pluie sur les performances du modèle. Le modèle a ainsi six paramètres libres : les cinq paramètres de GR5H et le paramètre de célérité. Dans un premier temps, nous appliquons une discrétisation spatiale de résolution 50 km^2 pour tous les bassins versants. À la fin de cette étude, nous explorons l'impact d'une discrétisation de résolution 10 km^2 pour tous les bassins qui ont une superficie inférieure à 250 km^2 .

Trois modèles sont comparés dans la première partie de cette analyse :

- GR5H avec deux hydrogrammes unitaires (M0 dans le chapitre 5) ;
- GRSD à une résolution de 50 km^2 (M0_SD50) ;
- GR5H comprenant les deux fonctions d'intensité et le réservoir linéaire sur la branche directe (M3 ; voir chapitre 5).

Nous appliquons le même cadre d'évaluation des performances que dans le chapitre précédent, des performances globales des modèles aux performances événementielles en fonction de caractéristiques des événements de pluie. Pour compléter les analyses, nous utilisons un critère simple d'erreur sur les pointes de crue calculé pour chaque événement j de la manière suivante :

$$\Delta Q_{max,j} = \frac{\max[Q_{sim,j}]}{\max[Q_{obs,j}]} - 1 \quad (6.1)$$

avec $Q_{sim,j}$ la chronique des débits simulés de l'événement j et Q_{obs} la chronique des débits observés de l'événement j . Comme pour le critère de biais (voir eq. 2.2), nous utilisons une version bornée de ce critère. Il est supérieur à zéro lorsque la valeur du pic simulé est supérieure à la valeur du pic observé et inférieure à zéro quand c'est le contraire.

Dans la deuxième partie de l'analyse, nous comparons GRSD avec une résolution spatiale de 10 km² (M0_SD10) à M0_SD50. Pour comparer les performances des deux versions, nous calculons le critère suivant pour chaque bassin et chaque saison :

$$\Delta_{NSE_{ev}} = \frac{\overline{NSE}_{ev}^{M0_SD10} + 1}{\overline{NSE}_{ev}^{M0_SD50} + 1} \quad (6.2)$$

avec, pour le bassin et la saison considérés, $\overline{NSE}_{ev}^{M0_SD10}$ la moyenne des valeurs de NSE événementiel borné de M0_SD10 et $\overline{NSE}_{ev}^{M0_SD50}$ la moyenne des valeurs de NSE événementiel borné de M0_SD50. Lorsque $\Delta_{NSE_{ev}}$ est supérieur à 1, la performance en crue de M0_SD10 est supérieure à la performance de M0_SD50.

6.3 Résultats et analyses

6.3.1 Distribution spatiale à 50 km²

Les distributions des valeurs de KGE en calage sont similaires pour M0 et M0_SD50 (fig. 6.1). Comme présenté dans le chapitre précédent, les valeurs des quantiles de la distribution des KGE de M3 en calage sont supérieures aux valeurs associées à M0. En validation, les valeurs des quantiles des distributions des KGE de M0_SD50 et M3 sont légèrement plus élevées que celles des quantiles de la distribution des KGE de M0. Les valeurs des quantiles de la distribution des KGE de M3 sont plus élevées que celles des quantiles de la distribution des KGE de M0_SD50. Ces résultats sont cohérents avec ceux de Lobligeois *et al.* (2014) qui montrent que les gains de performance obtenus par une distribution spatiale de la pluie existent mais sont peu significatifs en moyenne sur un large échantillon de bassins.

Les distributions des valeurs de biais en crue en hiver sont similaires pour les trois modèles. En été, la distribution des biais de M0_SD50 en crue est plus proche de 0 (non biaisé) que la distribution des biais de M0. La distribution des biais de M3 en crue est la plus proche de 0 et la moins dispersée des trois distributions.

Les distributions des valeurs de NSE événementiel de M0 et M0_SD50 sont similaires en hiver. Les valeurs des quantiles de la distribution des NSE de M3 sont plus élevées que celles des distributions associées aux deux autres modèles. Les valeurs des quantiles de la distribution des NSE de M0_SD50 en été sont plus élevées que celles des distributions associées à M0 mais moins élevées que les valeurs associées à M3.

Les distributions de performance sur les pics de crue en hiver indiquent que le modèle M0_SD50 sous-estime plus largement les pics que les deux autres modèles. En été, le critère sur les pointes de crue est amélioré pour M0_SD50 par rapport à M0. M3 a les meilleures performances pour simuler les pics en été.

Ces premiers résultats montrent que la distribution spatiale de la pluie à une résolution de 50 km² permet d'améliorer les performances pour les crues d'été sur notre base de données. Les gains de performance obtenus sont moins importants que ceux obtenus avec le modèle M3 en été et en hiver.

La figure 6.2a présente les performances événementielles de M0 et M0_SD50 selon le type de bassins (voir classification effectuée dans le chapitre 5) et la saison considérée. Les gains de performance liés à la distribution spatiale de la pluie semblent être plus importants pour les bassins du groupe 2 en été. Ce groupe de bassins contient les plus grands bassins de notre

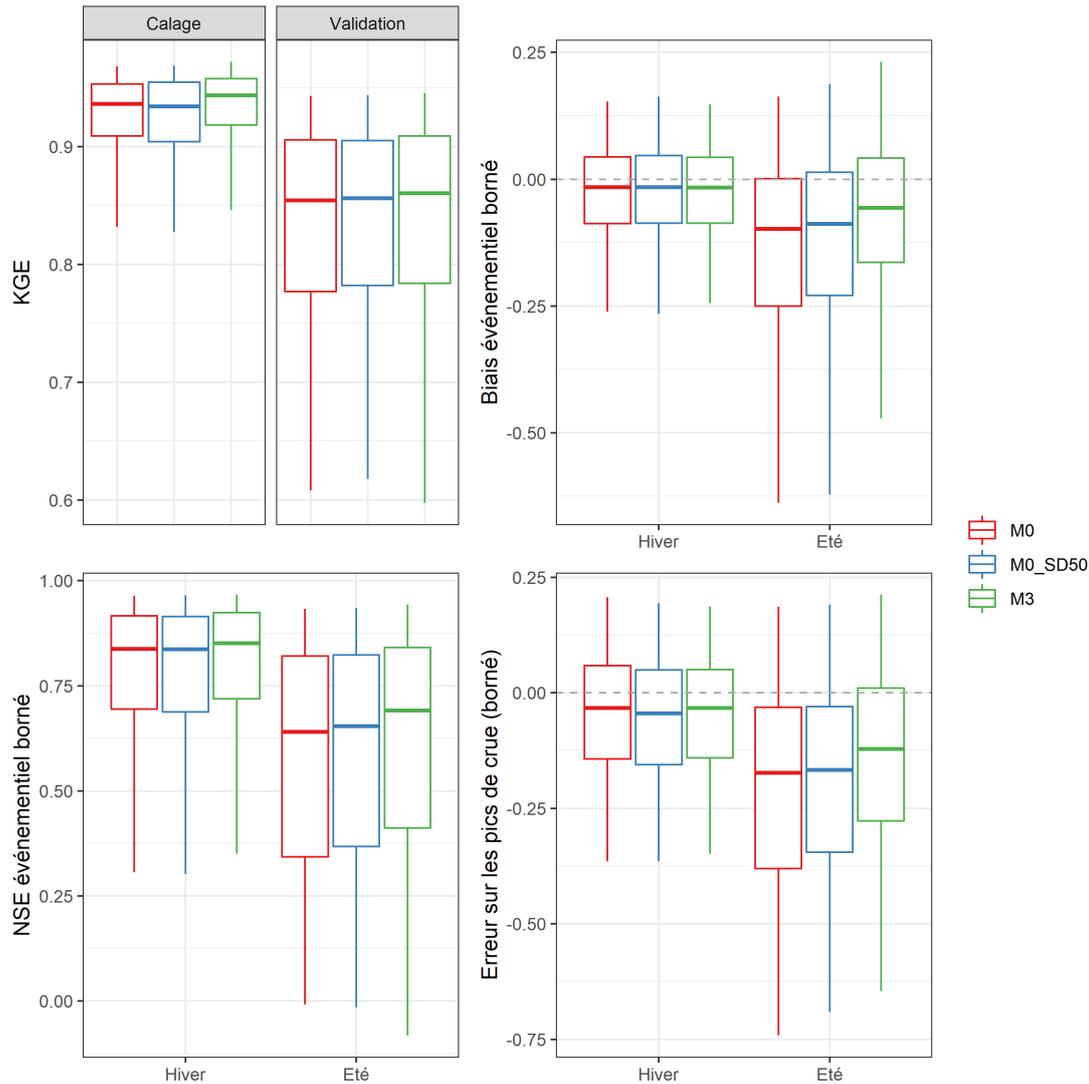


FIGURE 6.1 – Distribution des performances globales et événementielles en validation croisée pour le modèle de référence (M0), le modèle semi-distribué (M0_SD50) et le modèle comprenant les fonctions d'intensité (M3). 8 290 événements d'hiver et 2 362 événements d'été sont considérés ici. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

base de données. On observe des gains de performance en été pour les bassins du groupe 4, bien que la dispersion des points soit élevée. Les performances en crue en hiver sont fortement dégradées pour quelques bassins du groupe 5.

La figure 6.2b montre que les performances de M3 sont plus élevées que les performances de M0_SD50 pour les crues d'hiver pour plusieurs bassins. En été, la dispersion des points est très forte, en particulier pour les bassins des groupes 4 et 5. Ces résultats indiquent peut-être que les deux structures ont un rôle différent en été sur ces bassins : sur certains bassins, la variabilité spatiale de la pluie a un impact plus important sur le débit que l'intensité de la pluie et vice versa.

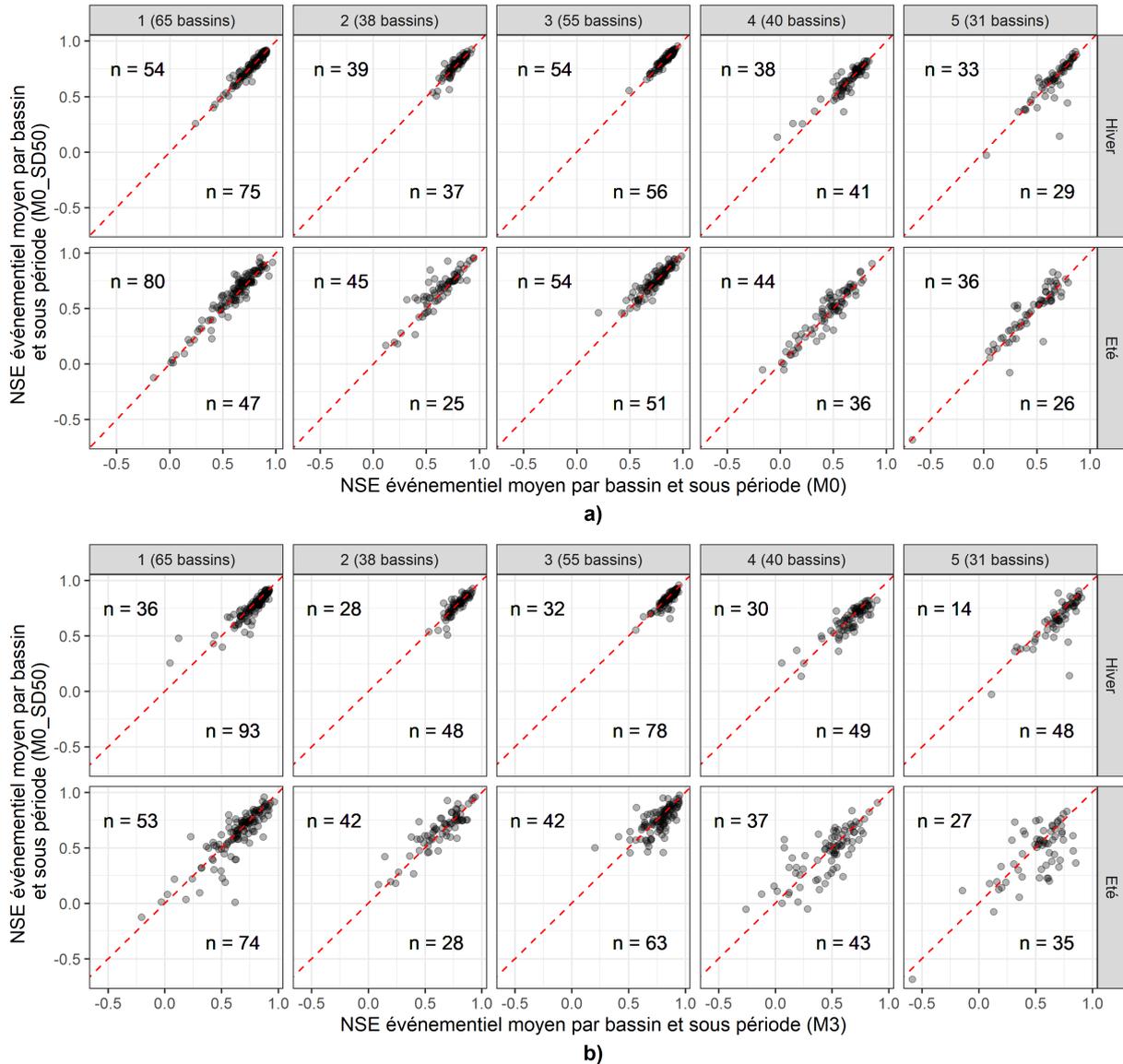


FIGURE 6.2 – Comparaison entre les performances événementielles du modèle de référence (M0) et le modèle semi-distribué (M0_SD50) pour cinq groupes de bassins versants (a) (voir classification effectuée dans le chapitre 5). Comparaison entre les performances événementielles du modèle avec les fonctions d'intensité (M3) et le modèle semi-distribué (M0_SD50) pour cinq groupes de bassins versants (b). Chaque point correspond à la moyenne du NSE événementiel pour une sous-période, une saison et un bassin donné.

Pour poursuivre cette hypothèse, nous reprenons les deux indices caractérisant les événements de pluie (voir chapitres 3 et 5). Les figures 6.3a et 6.3b présentent les performances événementielles des trois modèles en fonction de ces indices.

La figure 6.3a montre que les trois modèles ont des performances similaires sur les 50 % d'événements ayant les intensités de pluie de bassin les plus faibles. Plus l'intensité de pluie de l'événement est élevée, plus la différence de performances entre M3 et les deux autres modèles est élevée. Les performances de M0_SD50 restent proches des performances de M0 pour les événements de crue associés aux pluies les plus intenses. La figure 6.3b montre que M3 a des performances supérieures aux deux autres modèles pour les deux premières classes

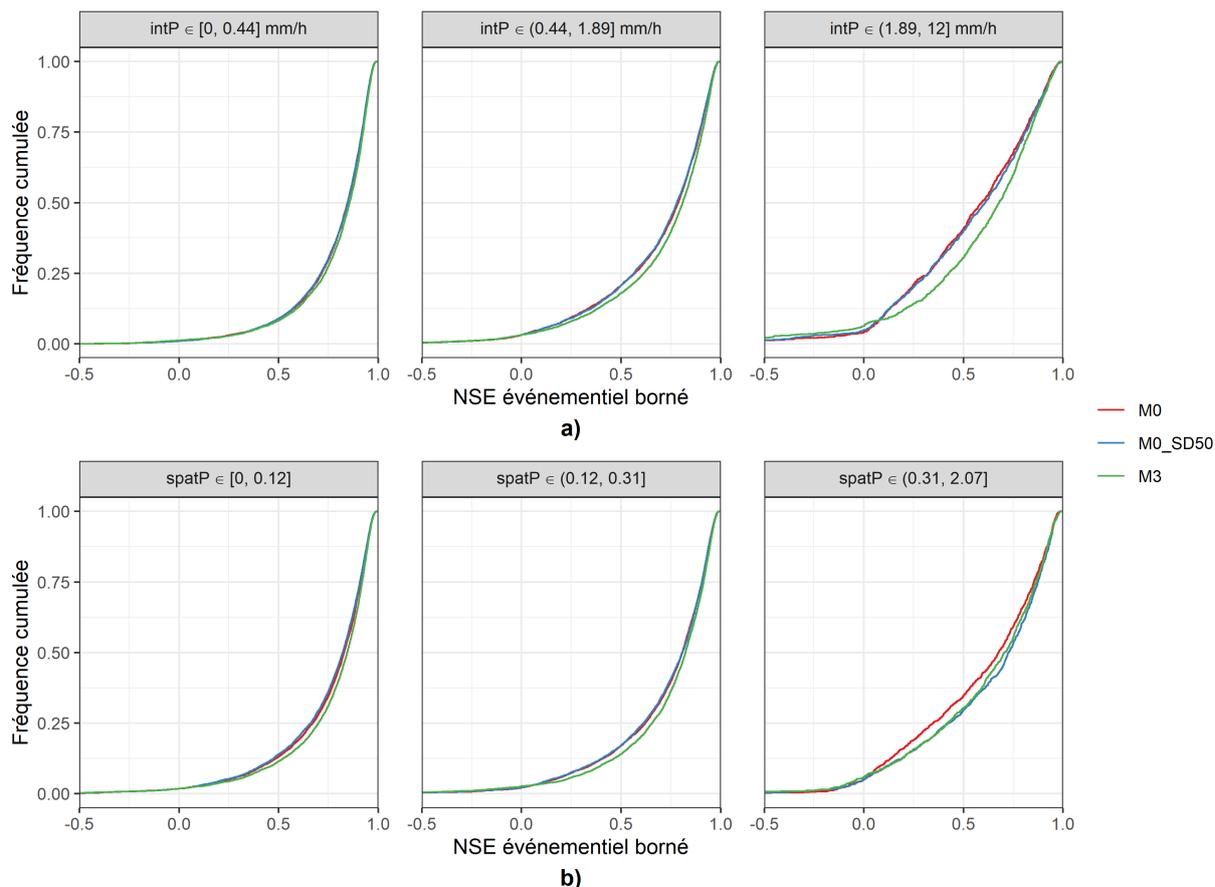


FIGURE 6.3 – Distribution des performances événementielles du modèle de référence (M0), du modèle semi-distribué (M0_SD50) et du modèle avec les fonctions d'intensité (M3) en fonction de l'indice d'intensité moyenne de la pluie $intP$ (a) et de la variabilité spatiale des cumuls de pluie $spatP$ (b). Les résultats sont présentés en validation croisée. Les 10 652 événements sont considérés dans cette analyse. Les bornes des classes d'intensité de pluie correspondent aux quantiles 50 % et 90 %.

de variabilité spatiale de la pluie, qui contiennent 90 % des événements (cela ne signifie pas que M3 a de meilleures performances pour chacun de ces événements). Pour les 10 % d'événements de crue associés aux pluies les plus fortement variables dans l'espace, M3 a de meilleures performances que M0, et M0_SD50 a les meilleures performances des trois.

En résumé, M3 a de meilleures performances que M0_SD50 en crue, en particulier pour les événements associés à des pluies très intenses. M0_SD50 améliore les performances de M0 lorsque les événements de pluie sont très variables dans l'espace. Dans ce dernier cas, M3 a des performances proches de celles de M0_SD50 mais cela ne correspond probablement pas toujours aux mêmes événements/bassins.

Les figures 6.4 et 6.5 illustrent les observations précédentes pour deux bassins versants : la Cèze à Tharoux, un affluent en aval du Rhône en rive droite et le Vistre au Cailar, un fleuve côtier méditerranéen. Nous avons fait le choix d'utiliser des échelles de temps, de pluie et de débit différentes entre les événements pour faciliter la visualisation car les événements de pluie ont des caractéristiques très différentes.

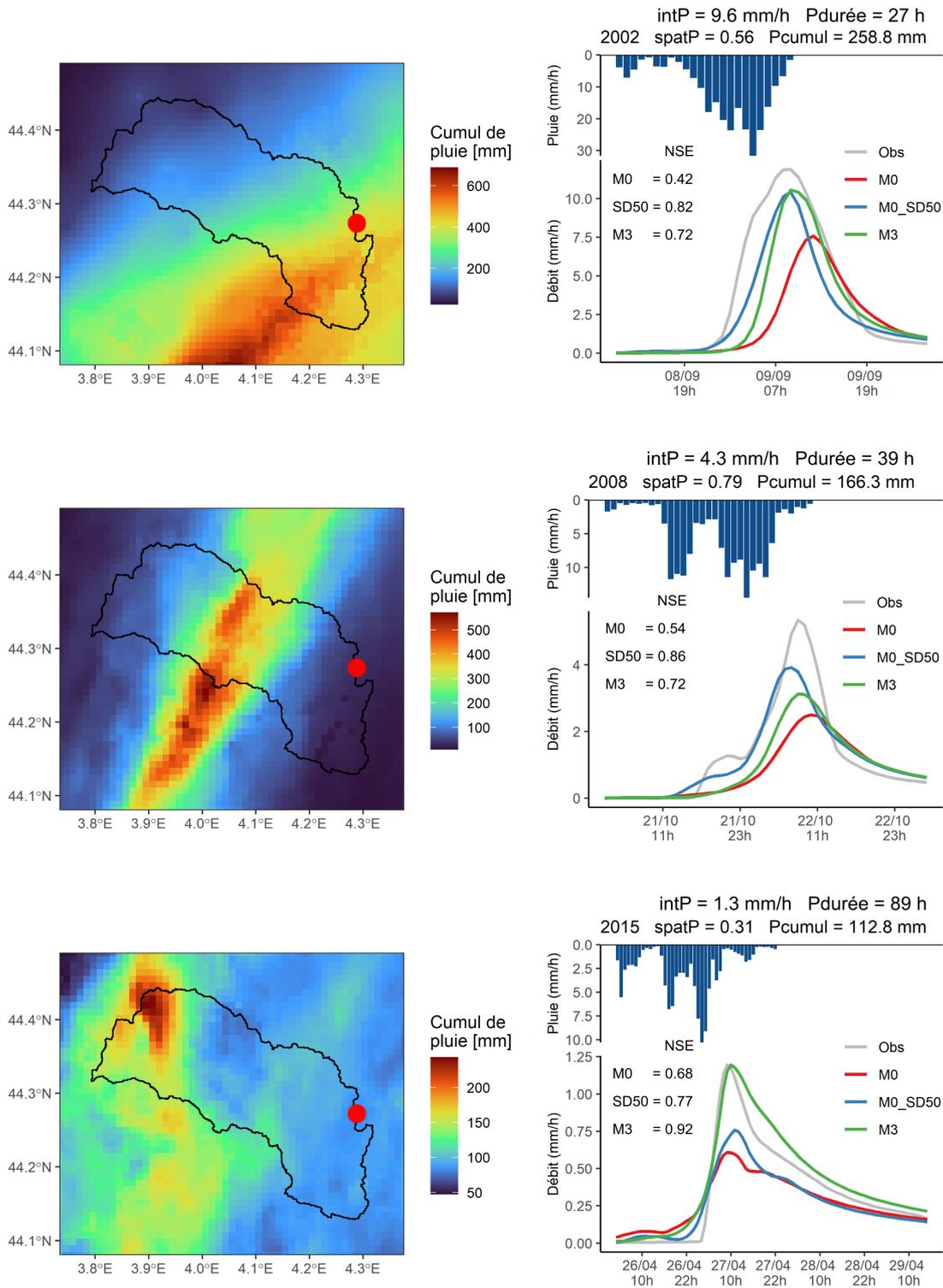


FIGURE 6.4 – Cartes des cumuls de pluie et hydrogrammes observé et simulés (validation croisée) de trois événements du bassin versant de la Cèze à Tharoux (665 km²). Les points rouges correspondent à la position de l'exutoire jaugé du bassin versant. M0 est le modèle de référence. M0_SD50 est le modèle semi-distribué. M3 est le modèle qui comprend les fonctions d'intensité. spatP est l'indice de variabilité spatiale des cumuls de pluie. Pcumul est le cumul de pluie de l'événement. Pdurée est la durée de l'événement de pluie. intP est l'indice d'intensité moyenne de la pluie. La performance des simulations est présentée avec le critère NSE borné.

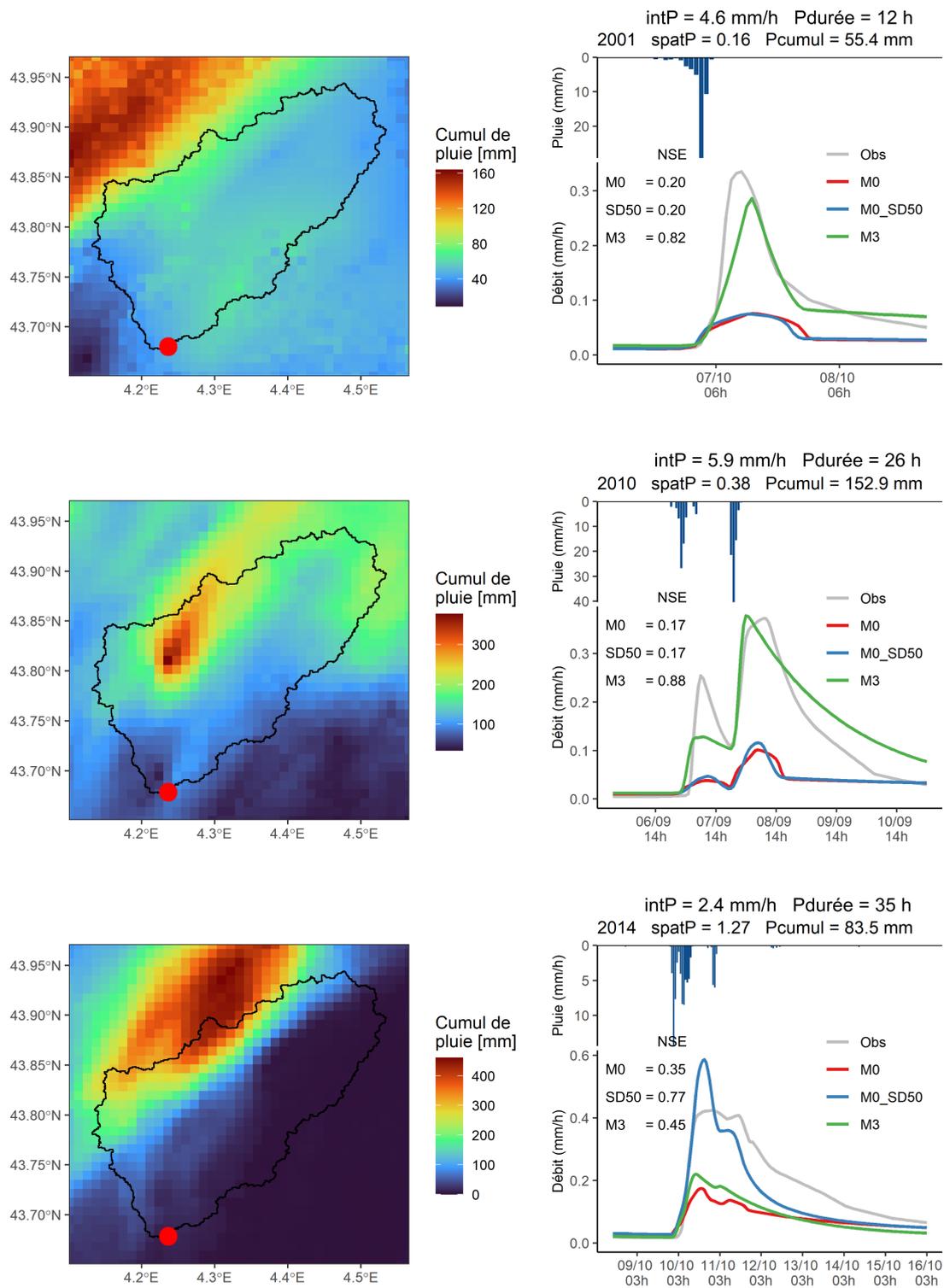


FIGURE 6.5 – Cartes des cumuls de pluie et hydrogrammes observé et simulés (validation croisée) de trois événements du bassin versant du Vistre au Cailar (496 km²). Les points rouges correspondent à la position de l'exutoire jaugé du bassin versant. M0 est le modèle de référence. M0_SD50 est le modèle semi-distribué. M3 est le modèle qui comprend les fonctions d'intensité. spatP est l'indice de variabilité spatiale des cumuls de pluie. Pcumul est le cumul de pluie de l'événement. Pdurée est la durée de l'événement de pluie. intP est l'indice d'intensité moyenne de la pluie. La performance des simulations est présentée avec le critère NSE borné.

L'événement qui a eu lieu en septembre 2002 sur la Cèze est caractérisé par un fort cumul de pluie à l'échelle du bassin versant, des intensités horaires de bassin allant jusqu'à plus de 30 mm/h, et une forte variabilité spatiale des cumuls de pluie. La carte de pluie montre que la partie sud-est du bassin (proche de l'exutoire jaugé) a reçu des cumuls de pluie allant jusqu'à 600 mm. Le modèle de référence a du mal à reproduire le volume de crue et présente un fort retard. Les modèles M0_SD50 et M3 reproduisent mieux le volume et le pic de crue bien qu'ils soient toujours en sous-estimation. M0_SD50 a un meilleur timing pour cet événement. Le modèle de référence n'arrive pas à reproduire le timing de l'événement probablement parce que celui-ci est majoritairement tombé proche de l'exutoire. Le modèle M3 réagit mieux que M0 car les intensités horaires de bassin permettent aux fonctions d'intensité de réagir.

L'événement d'octobre 2008 présente une forte variabilité spatiale de la pluie avec des cumuls localisés dans la partie centrale du bassin atteignant jusqu'à trois fois le cumul de pluie du bassin. Les intensités de bassin sont plus faibles que pour l'événement précédent. Le modèle de référence n'arrive pas à reproduire cet événement. Le modèle M3 reproduit un peu mieux la dynamique de l'événement mais sous-estime encore fortement le volume et le pic. Le modèle M0_SD50 reproduit mieux le volume de la crue et la montée, et sous-estime beaucoup moins la pointe.

En avril 2015, un événement caractérisé par un cumul et des intensités de bassin proches de ceux d'octobre 2008 mais une pluie plus homogène a fait réagir le bassin versant moins fortement. Bien qu'une petite partie du nord-ouest du bassin (en altitude) ait reçu des cumuls de pluie plus importants que les cumuls de bassin, une grande partie du bassin a été arrosée avec des cumuls du même ordre de grandeur. M0 et SD_50 sous-estiment la crue ainsi générée. M3 reproduit mieux la dynamique de l'événement.

L'événement d'octobre 2001 qui a eu lieu sur le Vistre fait suite à un événement de pluie dont les cumuls sont répartis de manière homogène sur tout le bassin. L'événement de pluie a été beaucoup plus intense en dehors du bassin. La moitié de l'événement de pluie (28 mm) est tombée en une heure, ce qui a induit une réaction très rapide du bassin versant. M0 et M0_SD50 sous-estiment une large partie du volume de crue. M3 arrive à reproduire la dynamique de la crue mais reste en sous-estimation.

En septembre 2009, des cumuls plus importants sont tombés sur le bassin sous la forme de deux événements de pluie courts et intenses et dont les intensités horaires de bassin atteignent 40 mm/h. La variabilité spatiale des cumuls de pluie est plus élevée que pour l'événement précédent car une petite partie du bassin a reçu des cumuls atteignant deux fois le cumul de pluie de bassin. Seul M3 simule correctement la dynamique de l'événement ainsi induit, en particulier le deuxième pic de crue.

L'événement d'octobre 2014 présente des caractéristiques très différentes. L'événement de pluie a eu lieu en bordure du bassin avec des cumuls allant jusqu'à 400 mm mais une grande partie du bassin n'a reçu que très peu d'eau voire pas d'eau du tout. Le cumul de pluie à l'échelle du bassin est ainsi très faible par rapport au cumul que certaines parties du bassin ont reçu. Les intensités horaires de bassin sont plus faibles que les intensités des deux autres événements présentés ici. Seul M0_SD50 simule correctement le volume de crue avec une légère surestimation du pic.

Les observations précédentes mettent en évidence plusieurs comportements :

- Lorsque la variabilité spatiale des cumuls de pluie et les intensités de pluie de bassin sont élevées, M3 et M0_SD50 captent mieux la variabilité spatio-temporelle sous-jacente. La variabilité spatiale n'est pas toujours séparable de la variabilité temporelle (e.g. un événement qui se déplace).
- Lorsque les intensités horaires de bassin sont élevées et que c'est ce signal qui engendre le plus de variabilité de la réponse du bassin, c'est-à-dire que la variabilité spatiale de la pluie peut-être élevée mais qu'elle a un rôle moins important (e.g. l'événement d'avril 2015 sur la Cèze et l'événement de septembre 2010 sur le Vistre), M0_SD50 n'a pas de meilleures performances que M0, et M3 permet de reproduire une partie de la dynamique des crues ainsi générées.
- Lorsque la variabilité spatiale de la pluie est très élevée mais que les intensités et le cumul de bassin sont faibles par rapport aux conditions « habituelles » de génération des crues du bassin, la distribution spatiale du modèle permet de capturer la variabilité de la réponse du bassin versant. Les modèles M0 et M3 n'ont pas les capacités intrinsèques de reproduire la dynamique de ce type d'événements.

6.3.2 Distribution spatiale à 10 km²

La figure 6.6 montre que les performances événementielles des deux versions de GRSD sont très proches avec une faible amélioration des performances en hiver pour M0_SD10 par rapport à M0_SD50. La répartition spatiale des performances montre que la discrétisation spatiale à 10 km² améliore les performances en hiver et en été pour plusieurs bassins du sud et du sud-est. On n'observe pas de tendances spatiales entre les bassins pour lesquels cette discrétisation spatiale dégrade les performances du modèle. Le bassin pour lequel M0_SD10 dégrade le plus les performances en hiver est un bassin avec un rendement Q/P supérieur à 1. Les deux bassins pour lesquels M0_SD10 améliore le plus les performances sont des bassins qui ont une superficie de 42 km² et 63 km². La discrétisation spatiale à 50 km² ne permet pas de diviser ces bassins en sous-bassins versants.

Les résultats de la figure 6.7 indiquent que la discrétisation spatiale plus fine du modèle n'améliore pas les performances pour les événements les plus intenses ni pour les événements les plus variables dans l'espace. Les performances dégradées de M0_SD10 pour certains bassins peuvent s'expliquer par une non-adaptation de la structure du modèle à une résolution de 10 km². En effet, le bilan en eau est plus difficile à simuler pour des sous-bassins plus petits car il est plus compliqué d'établir les conditions aux limites.

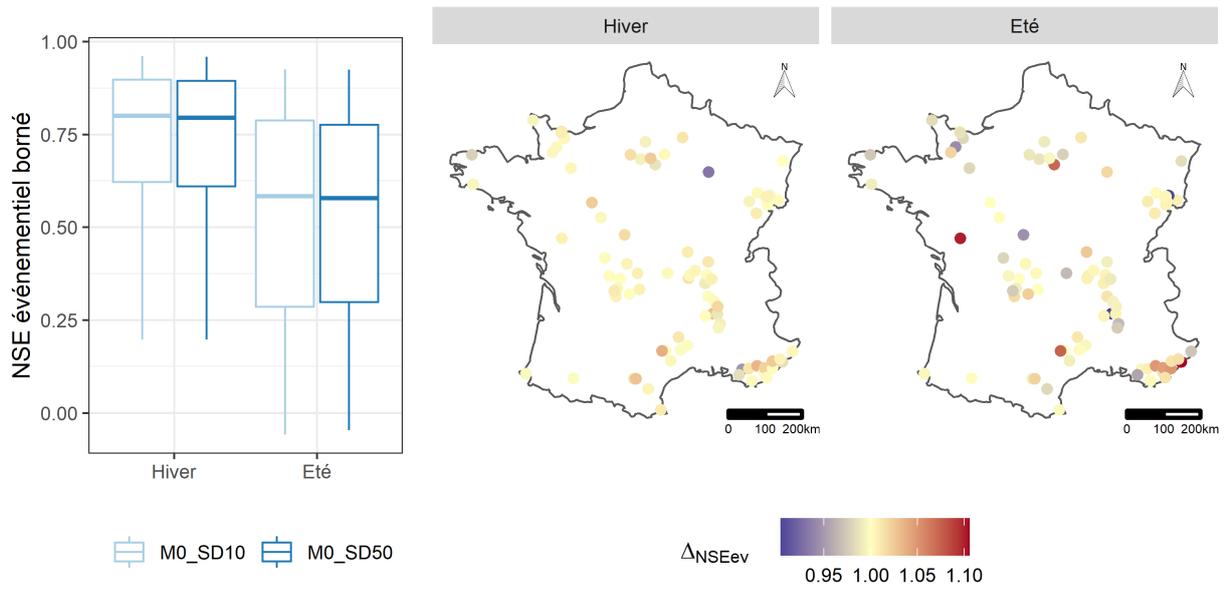


FIGURE 6.6 – Distribution et répartition spatiale des performances événementielles de M0_SD10 et M0_SD50 pour 84 bassins de moins de 250 km². 2838 événements d’hiver et 1037 événements d’été sont considérés ici. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. $\Delta_{NSEev} > 1$ signifie que M0_SD10 a de meilleures performances événementielles que M0_SD50.

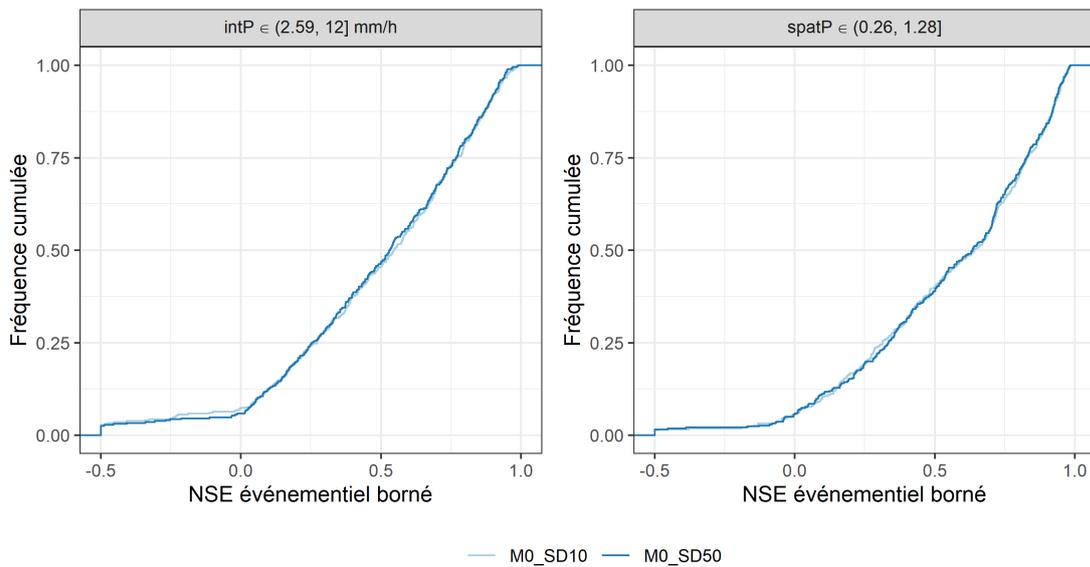


FIGURE 6.7 – Distribution des performances événementielles (validation) de M0_SD10 et M0_SD50 pour les 10 % d’événements les plus intenses et les 10 % d’événements les plus variables spatialement pour 84 bassins de moins de 250 km².

6.4 Conclusion

Les résultats présentés dans ce chapitre ont montré que la prise en compte de la distribution spatiale de la pluie dans un schéma de modélisation semi-distribuée permet d'améliorer les performances du modèle pour certains événements d'été, en particulier lorsque la variabilité spatiale de la pluie est très élevée. Ces analyses confirment celles de Lobligeois *et al.* (2014). Nous avons vu que l'amélioration des performances par une distribution simple de la pluie n'est pas équivalente à l'amélioration des performances apportée par la structure M3. Nous retiendrions qu'il existe un rôle différencié de ces deux structures lorsque la réaction du bassin est majoritairement portée par la variabilité spatiale ou lorsque c'est la variabilité temporelle qui joue un rôle plus important. Cependant, lorsque ces deux caractéristiques affectent la dynamique des événements de crue, les deux structures peuvent apparaître soit redondantes soit complémentaires. Une perspective de ce travail serait de distribuer spatialement les fonctions d'intensité pour déterminer s'il existe une complémentarité des deux approches (pour des exemples, voir Goullet, 2016 ; Peredo *et al.*, 2022). Cela soulèverait des questions d'adaptation des formulations à l'échelle considérée et de redondance des fonctions de routage.

Le travail présenté ici comporte des limites. En particulier, l'application d'un calage global limite la portée de nos résultats. Dans cette analyse, nous avons présenté des exemples d'hydrogrammes qui illustrent les différences entre les modèles. Il existe toutefois une diversité très importante de cas observés qu'il est difficile de résumer. La généralité de nos résultats est démontrée par les distributions statistiques sur l'ensemble de l'échantillon de bassins et d'événements.

Enfin, les performances plus faibles de la version distribuée par rapport aux performances de la version qui comprend les fonctions d'intensité ne signifient pas qu'il n'est pas possible d'améliorer la simulation des crues rapides avec une meilleure discrétisation spatiale. Les résultats indiquent plutôt qu'une distribution spatiale et un calage simples du modèle GR5H ne permettent pas de résoudre l'intégralité des déficiences du modèle.

Chapitre 7

Vers une estimation des paramètres plus adaptée pour simuler les crues

7.1 Contexte

L'estimation des paramètres des modèles hydrologiques a fait l'objet de nombreux travaux, en particulier avec l'augmentation de la puissance de calcul des ordinateurs ces dernières décennies qui permet d'explorer l'espace des paramètres à partir de millions de simulations. Dans le cas des modèles pluie-débit, il est fréquent de recourir à des procédures de calage pour estimer la valeur des paramètres qui ne sont pas nécessairement liés à des caractéristiques facilement mesurables des bassins versants (e.g. de Lavenne *et al.*, 2019). Dans nos travaux, les paramètres des différents modèles ont été estimés à partir d'une telle procédure, qui permet d'obtenir un jeu de paramètres dit « optimal » sur chaque sous-période de calage au regard d'une fonction coût mesurant la différence entre la chronique des débits observés et la chronique des débits simulés. Cette méthode d'estimation des paramètres est très efficace mais elle présente des limites qui peuvent nuire à la robustesse et à la capacité prédictive des modèles hydrologiques.

Une des difficultés du calage des modèles hydrologiques est le « surparamétrage ». Ce terme désigne le fait qu'un modèle peut contenir un nombre trop important de paramètres libres au regard de l'information disponible pour contraindre leur estimation. Le surparamétrage donne souvent lieu à des problèmes d'équifinalité (e.g. Beven, 2006 ; Clark *et al.*, 2016), i.e. des jeux de paramètres différents donnent des performances similaires au regard de la fonction coût choisie. Ces difficultés ont poussé certains auteurs à proposer un nombre maximal de paramètres qu'il est possible d'estimer à partir des chroniques de débits (e.g. Jakeman & Hornberger, 1993 ; Perrin *et al.*, 2001). Ce nombre maximal dépend du modèle utilisé, notamment sa formulation mathématique, car les paramètres interagissent entre eux selon le degré d'activation d'une fonction donnée, ce qui peut changer la forme des surfaces de réponse. La complexité maximale et nécessaire d'un modèle dépend aussi de la résolution temporelle des données (Ficchi *et al.*, 2019) et donc du niveau d'information disponible. Le surparamétrage peut aussi se traduire par le « surcalage », c'est-à-dire le fait que l'optimum mathématique n'est pas nécessairement l'optimum hydrologique du bassin considéré (Andréassian *et al.*, 2012). En d'autres termes, l'algorithme de calage peut déterminer un optimum mathématique sur une période donnée mais les performances baissent lorsque le jeu de paramètres correspondant est utilisé sur une autre période (e.g. Fowler *et al.*, 2016).

Cette notion est liée à la robustesse des modèles hydrologiques (pour plus de détails sur la notion de robustesse, voir Coron, 2013 ; Thirel *et al.*, 2015 ; Royer-Gaspard, 2021). Lorsque les différences de performance entre la période de calage et la période de contrôle sont élevées, il y a généralement une variabilité importante des paramètres entre périodes, ce qui n'est pas souhaitable lorsqu'on se place en contexte de stationnarité comme nous le faisons en simulation et prévision des crues. Les paramètres deviennent ainsi dépendants de la période climatique utilisée.

Plus généralement, les difficultés d'identification des paramètres des modèles hydrologiques produisent une incertitude paramétrique. Cette incertitude paramétrique est liée à l'incertitude des données (e.g. les paramètres s'adaptent aux « mauvaises » données) et de l'incertitude liée à la structure des modèles (e.g. les paramètres compensent la mauvaise adéquation d'un modèle au fonctionnement du bassin considéré). Plusieurs auteurs suggèrent qu'il est important de prendre cette incertitude en compte dans le développement des modèles hydrologiques (e.g. Clark *et al.*, 2011 ; Beven, 2018). Cela peut se faire dans un cadre d'inférence statistique en effectuant des calages bayésiens (e.g. Laloy & Vrugt, 2012). D'autres auteurs considèrent que l'incertitude paramétrique est une manifestation de l'incertitude totale de la chaîne de modélisation. L'incertitude paramétrique découle ainsi de ce que certains auteurs nomment l'incertitude épistémique (pour l'incertitude qui provient de notre manque de connaissance ; Beven, 2016). Ils suggèrent de considérer un ensemble de jeux de paramètres « acceptables » pour un bassin, une période et un critère donnés (e.g. Beven & Freer, 2001b). Cette méthode demande cependant de définir un seuil d'acceptabilité sur les performances du modèle, avec un choix souvent arbitraire.

Afin d'améliorer l'identification des paramètres des modèles hydrologiques, il est possible d'utiliser des fonctions objectif qui sont plus en adéquation avec le problème donné. Par exemple, une fonction objectif peut être orientée vers certaines caractéristiques des hydrogrammes (e.g. utiliser une transformation des débits pour mieux simuler les bas débits Pushpalatha *et al.*, 2012 ; Santos *et al.*, 2018). Pour mieux simuler les hauts débits, Mizukami *et al.* (2019) proposent une fonction objectif qui met plus de poids sur les pics annuels. Ils obtiennent une amélioration de la simulation des pics annuels mais une dégradation des performances pour d'autres critères caractérisant les hauts débits. Une autre manière d'améliorer l'identification des paramètres est de se placer dans un cadre multi-objectif (pour une revue sur le sujet, voir Efstratiadis & Koutsoyiannis, 2010). On peut ainsi évaluer les compromis qui existent entre différents objectifs de modélisation en définissant plusieurs jeux de paramètres optimaux (e.g. Monteil *et al.*, 2020). Cela permet aussi de diminuer les possibilités de surajustement (*overfitting* en anglais) des paramètres à des caractéristiques spécifiques des hydrogrammes (Pool *et al.*, 2018). Par exemple, Moussa & Chahinian (2009) testent plusieurs fonctions objectif pour caler les paramètres d'un modèle conceptuel sur le bassin versant du Gardon d'Anduze. Leurs résultats montrent que les paramètres varient très largement entre fonctions objectif. Ils proposent un critère multi-objectif à partir d'un front de Pareto pour améliorer la simulation des crues sur ce bassin.

L'identification des paramètres des modèles hydrologiques s'inscrit dans la recherche du contenu en information des données (comme théorisé par Shannon, 1948). Cette notion de contenu en information est très liée aux questions de recherche de nos travaux (cf. introduction générale) dont on peut décliner des questions spécifiques à ce chapitre. Les données à

notre disposition nous permettent-elles d'augmenter le nombre de paramètres du modèle ? Un objectif de simulation des crues permet-il de contraindre une partie de l'équifinalité inhérente à notre exercice de modélisation ? Reste-t-il de l'information pour complexifier l'estimation des paramètres et la structure du modèle que nous utilisons ? Les objectifs de ce chapitre sont de caractériser les problèmes d'identification des paramètres des modèles utilisés dans nos travaux et proposer des méthodes d'estimation des paramètres pour pallier ces défauts et améliorer la simulation des crues. Pour répondre à ces objectifs, nous caractérisons le niveau d'identifiabilité des paramètres des modèles M0 et M3 (cf. fig. 5.2). Nous le faisons en regardant la variation des paramètres entre périodes de calage, en analysant les surfaces de réponse du modèle, en explorant les interactions entre certains paramètres et en analysant les variations des paramètres au regard de plusieurs fonctions objectif. Dans un second temps, nous proposons une réduction de la complexité paramétrique des modèles en fixant deux paramètres. Un paramétrage multi-objectif est introduit pour améliorer l'identification des paramètres.

7.2 Éléments méthodologiques

Pour caractériser le niveau d'identifiabilité des paramètres, nous avons choisi d'explorer l'espace des paramètres le plus largement possible. L'idée de ce travail est d'explorer les surfaces de réponse du modèle de manière empirique afin d'obtenir un premier aperçu des difficultés d'identification des paramètres. Une autre méthode aurait été de mener une analyse de sensibilité, par exemple avec les indices de Sobol (Sobol, 2001). Cependant, ce type de méthode ne permet pas d'identifier précisément les équifinalités entre paramètres.

Dans un premier temps, trois millions de jeux de paramètres du modèle M0 (X_1 à X_5) ont été testés pour chaque bassin et chaque sous-période de calage (ce qui revient à tester environ 20 valeurs par paramètre). Les paramètres ont été générés à partir d'un échantillonnage par hypercube latin (McKay *et al.*, 1979), une méthode qui dérive du tirage de Monte-Carlo. L'échantillonnage a été effectué dans l'espace des paramètres transformés (tableau 7.1).

Dans un deuxième temps, pour chaque bassin et chaque sous-période, les intervalles de recherche des paramètres X_1 , X_3 et X_4 ont été réduits à partir des jeux de paramètres pour lesquels le critère NSE global est supérieur à 95 % de la valeur maximale de ce critère. Les intervalles de recherche des paramètres X_2 et X_5 sont difficiles à réduire sur une grande partie des bassins (voir les figures 7.4 et 7.6). Afin de réduire ces intervalles, nous avons échantillonné ces paramètres autour de leur valeur calée pour chaque bassin et chaque sous-période :

- Le paramètre X_2 est échantillonné dans l'espace transformé en ajoutant 1 et en soustrayant 1 à sa valeur calée.
- L'intervalle de X_5 est déterminé dans l'espace non-transformé en ajoutant 0.15 et en soustrayant 0.15 à sa valeur calée. X_5 est ensuite échantillonné dans l'espace transformé en utilisant les nouvelles bornes.

Les choix méthodologiques précédents ont été déterminés à partir d'une inspection visuelle des surfaces de réponse des paramètres sur chaque bassin versant. Les valeurs seuil utilisées sont arbitraires et tiennent compte, d'une part de la convergence des paramètres, et d'autre part de la taille des tableaux de données ainsi générés. Plusieurs valeurs ont été testées avant de choisir les valeurs utilisées dans ce chapitre.

Dans un troisième temps, un million de jeux de paramètres du modèle M3 ont été testés en utilisant les intervalles précédemment contraints pour les paramètres X_1 à X_5 , cela afin de réduire les temps de calcul associés.

TABLE 7.1 – Fonctions de transformation des paramètres de l’espace transformé vers l’espace réel. L’intervalle des paramètres dans l’espace transformé est $[-9.99; 9.99]$. X^T désigne le paramètre transformé. X^R désigne le paramètre réel.

Par.	Signification	Unité	Fonction de transformation	Intervalle
X_1	Capacité du réservoir de production	mm	$X^R = 140 \times \exp(0.5 \times X^T) - 0.9$	$[0.05; 20673]$
X_2	Coefficient d’échange	mm/h	$X^R = X^T/24$	$[-0.42; 0.42]$
X_3	Capacité du réservoir de routage	mm	$X^R = 80 \times \exp(0.5 \times X^T) - 0.5$	$[0.05; 11813]$
X_4	Demi-temps de base des hydrogrammes unitaires	h	$X^R = \exp(0.3 \times X^T) \times 10 + 0.001$	$[0.5; 200]$
X_5	Seuil d’échange	–	$X^R = (X^T + 10) / 20$	$]0; 1[$
i_1	Coefficient d’intensité de la pluie (production)	h/mm	$X^R = \exp(X^T \times 0.31 - 3.80) - 0.001$	$[0; 0.5]$
i_2	Coefficient d’intensité de la pluie (partage)	h/mm	$X^R = \exp(X^T \times 0.28 - 3.45) - 0.002$	$[0; 0.5]$
K_L	Coefficient de vidange du réservoir linéaire	–	$X^R = \exp(X^T \times 0.23 - 2.25) - 0.01$	$[0.001; 1]$

Les critères événementiels présentés dans ce chapitre ont été calculés par bassin et par sous-période pour alléger les sauvegardes des tableaux contenant les résultats des quatre millions de jeux de paramètres testés (voir eq. 2.3 pour un exemple avec le biais).

Les résultats de ce chapitre sont séparés en deux parties. La première partie présente différents problèmes liés à l’estimation des paramètres des modèles M0 et M3. Les étapes de test sont les suivantes :

- Présentation de la variation des paramètres entre périodes de calage.
- Analyse de l’identifiabilité des paramètres au regard de la fonction objectif.
- Exploration des interactions entre paramètres.
- Analyse de la variation des paramètres au regard d’autres critères d’évaluation.

La seconde partie présente des solutions que nous avons testées pour pallier certains des défauts identifiés dans la première partie :

- Plusieurs valeurs génériques du paramètre de seuil d’échange (X_5) sont testées pour le modèle M0. Les paramètres X_1 à X_4 sont calés par bassin et par sous-période en utilisant la même valeur de X_5 pour tous les bassins.
- La même méthode est appliquée pour le coefficient de vidange du réservoir linéaire (K_L) du modèle M3.
- Les jeux de paramètres des modèles M0 et M3 sont sélectionnés à partir des quatre millions de jeux de paramètres testés en utilisant la fonction objectif suivante :

$$Fobj = \frac{1}{2} \times (KGE + NSE_{cEv}) \quad (7.1)$$

avec NSE_{cEv} le critère NSE calculé pour une sous-période et un bassin donnés sur tous les pas de temps des événements sélectionnés.

7.3 Résultats et analyses

7.3.1 Identifiabilité des paramètres

Variation des paramètres entre périodes de calage

Une des manifestations de l'incertitude paramétrique est la variation des paramètres optimaux entre périodes de calage. Dans le chapitre 5, nous avons mis en évidence la variabilité des paramètres des fonctions d'intensité entre périodes de calage (fig. 5.10). La figure 7.1 montre que cette variabilité caractérise aussi les paramètres de la fonction d'échange X_2 et X_5 . La dispersion des valeurs des paramètres X_1 , X_3 et X_4 entre sous-périodes est plus faible.

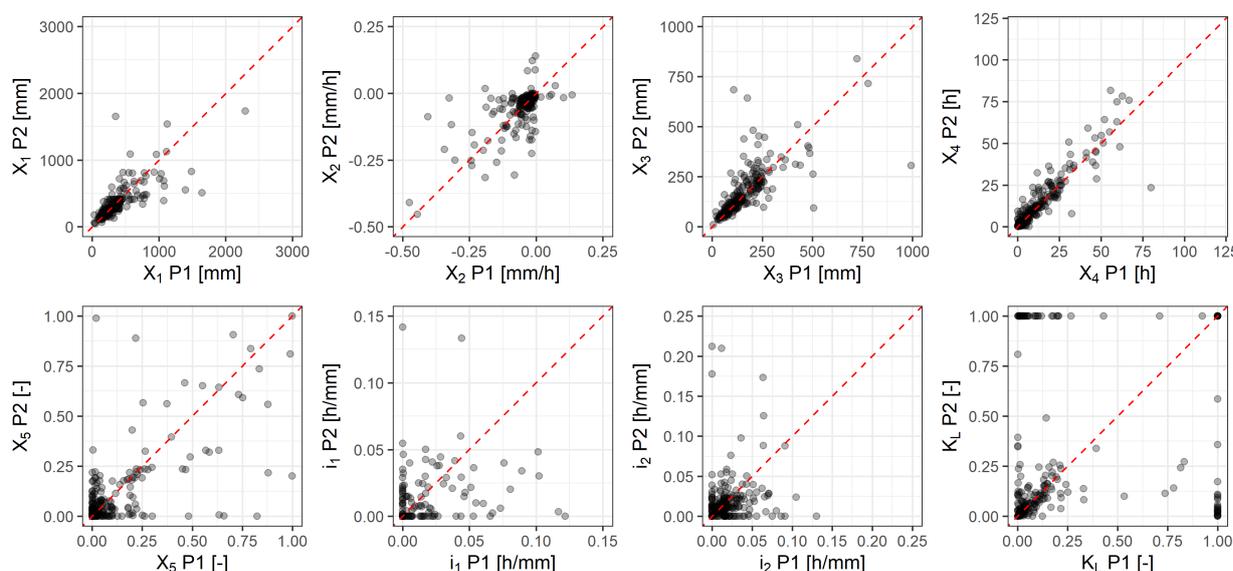


FIGURE 7.1 – Variation des paramètres du modèle M3 entre les périodes de calage P1 et P2 sur les 229 bassins versants.

Une explication de cette instabilité des paramètres pourrait être que l'algorithme de calage est limité pour certains bassins, en ne permettant pas une exploration assez large des paramètres ou en identifiant des optima mathématiques très éloignés des optima hydrologiques. Pour le vérifier, on regarde la différence de stabilité des paramètres pour deux méthodes d'estimation des paramètres : le calage utilisé dans les travaux précédents et les jeux de paramètres optimaux sélectionnés à partir de l'exploration large de l'espace des paramètres. Les valeurs de corrélation entre ces deux méthodes ne sont pas très différentes et ne permettent pas de conclure sur les limites de l'algorithme de calage (fig. 7.2). On notera cependant que les corrélations sont plus élevées pour les paramètres X_4 , i_2 et K_L lorsqu'on utilise une exploration plus large de l'espace des paramètres. À l'inverse, les valeurs de corrélation sont plus faibles pour i_1 dans ce cas. Les corrélations de rang et les corrélations linéaires indiquent deux comportements différents pour le paramètre X_2 . Cela est probablement dû à la présence de quelques valeurs très dispersées. Les valeurs de corrélation linéaire peuvent être très sensibles dans ce cas.

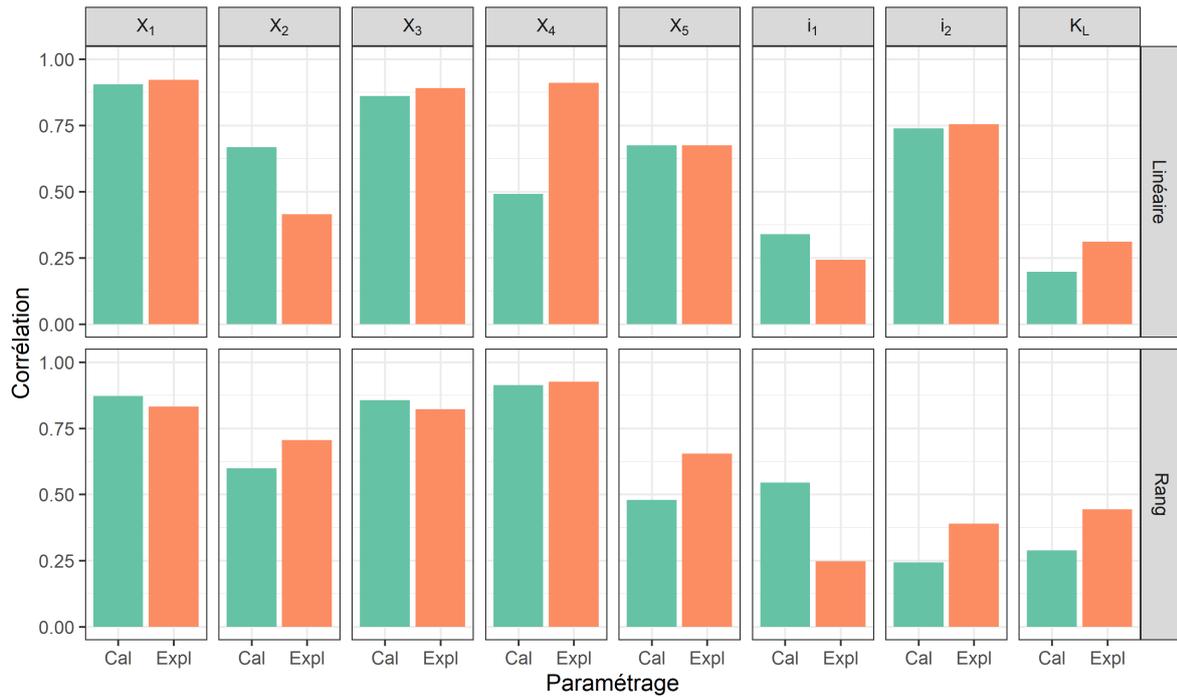


FIGURE 7.2 – Corrélations (linéaire et de rang) des paramètres du modèle M3 entre périodes de calage selon deux méthodes d’estimation des paramètres. Cal : calage des paramètres à partir de l’algorithme utilisé dans les chapitres précédents. Expl : sélection du meilleur jeu de paramètres par période et par bassin à partir d’une exploration d’un grand ensemble de jeux de paramètres. La même fonction objectif KGE est utilisée pour les deux méthodes d’estimation des paramètres.

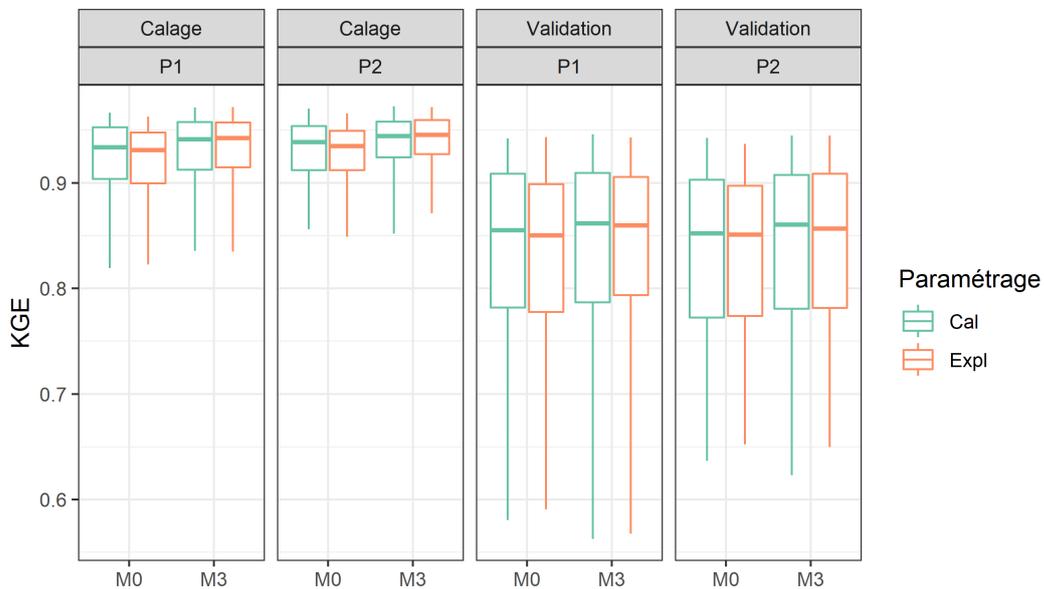


FIGURE 7.3 – Performances globales des modèles M0 et M3 pour deux méthodes d’estimation des paramètres. Cal : calage des paramètres à partir de l’algorithme utilisé dans les chapitres précédents. Expl : sélection du meilleur jeu de paramètres par période et par bassin à partir d’une exploration d’un grand ensemble de jeux de paramètres. La fonction objectif KGE est utilisée pour les deux méthodes d’estimation des paramètres. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Des optimums locaux ou une limitation de l'algorithme de calage lorsque le nombre de paramètres est plus élevé (e.g. huit paramètres libres pour le modèle M3) pourraient engendrer des baisses de performance. Les performances en calage du modèle M0 sont plus faibles sur les deux sous-périodes lorsque les jeux de paramètres sont sélectionnés à partir de l'exploration large de l'espace des paramètres plutôt que par l'algorithme de calage (fig. 7.3). À l'inverse, les valeurs de KGE de M3 sont plus élevées en calage pour les deux sous-périodes lorsque les jeux de paramètres sont sélectionnés à partir du million de jeux de paramètres. Ces résultats indiquent que l'algorithme de calage est moins adapté pour certains bassins lorsque le modèle a huit paramètres libres que lorsqu'il en a cinq. En validation, les performances sont similaires pour les deux méthodes d'estimation des paramètres, cela pour les deux modèles. Ces résultats sont cohérents avec ceux de Mathevet (2005), qui a montré que l'utilisation de méthodes globales plutôt que des méthodes locales pour le calage du modèle GR4H peut donner lieu à des performances légèrement plus élevées en calage sans pour autant améliorer les performances en validation.

Identifiabilité des paramètres au regard du critère KGE

Pour comprendre les problèmes d'identification des paramètres sur certains bassins, nous avons visualisé les surfaces de réponse des paramètres de M0 et M3 pour tous les bassins de notre base de données. Les figures 7.4 et 7.5 présentent ces surfaces de réponse pour le bassin versant du Gardon de Saint-Jean à Corbès (situé dans les Cévennes) qui correspond à une configuration fréquemment observée.

On voit que les paramètres X_1 (capacité du réservoir de production), X_3 (capacité du réservoir de routage) et X_4 (demi-temps de base des HU) sont bien identifiables, c'est-à-dire que les points pour lesquels les valeurs de KGE sont les plus élevées correspondent à un intervalle restreint des valeurs de ces paramètres. On observe que ce n'est pas le cas pour les paramètres de la fonction d'échange X_2 et X_5 , et pour le paramètre K_L (coefficient de vidange du réservoir linéaire). Les surfaces de réponse sont plus plates et ont une forme différente entre les deux périodes. Cela signifie que des valeurs très différentes de ces paramètres donnent des performances équivalentes. Sur ce bassin, les paramètres i_1 et i_2 semblent bien identifiés bien que les pentes des surfaces de réponse ne soient pas aussi importantes que celles des paramètres X_1 , X_3 et X_4 . On notera toutefois que les pentes observées dépendent du choix des fonctions de transformation, des intervalles de recherche ainsi que du seuil d'acceptabilité. Ces choix ont un impact sur l'échelle des graphiques présentés.

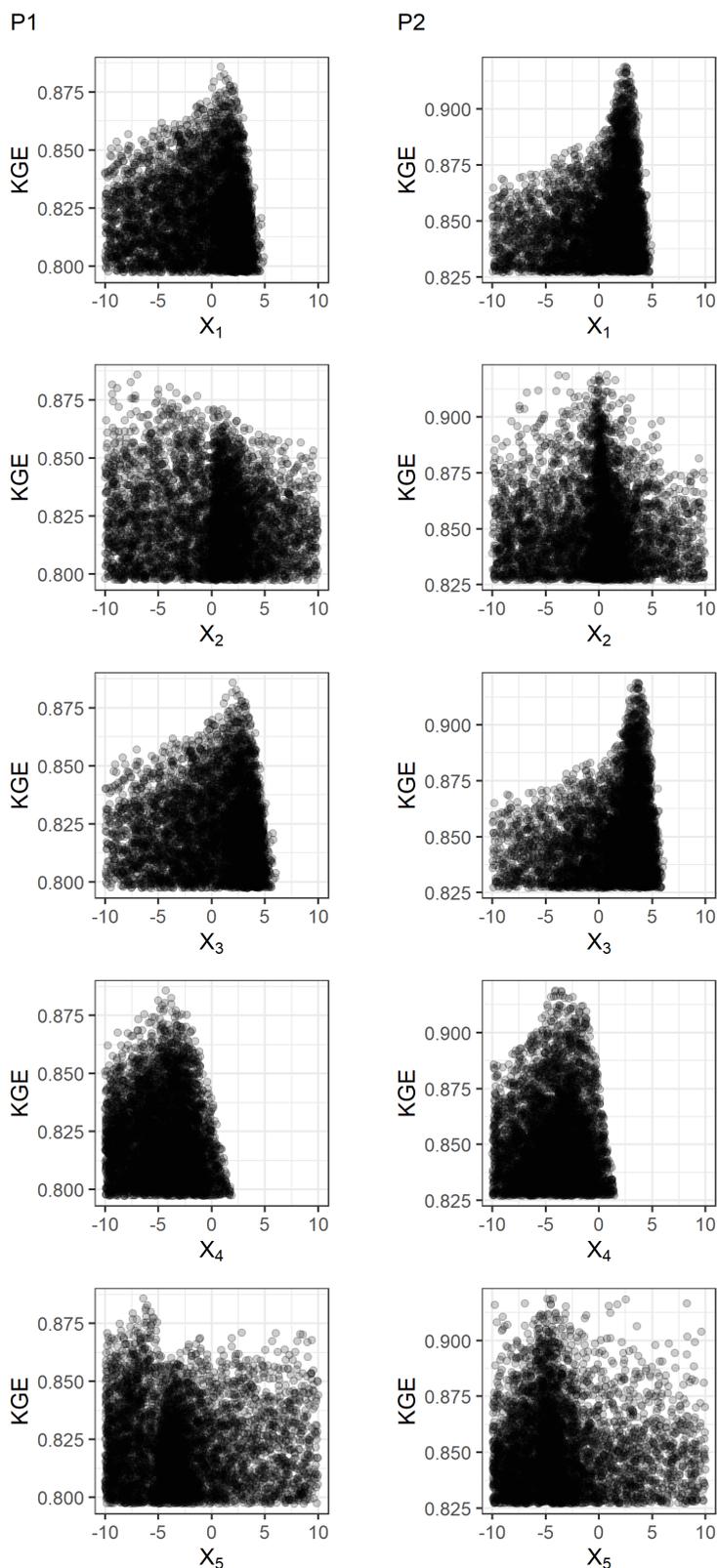


FIGURE 7.4 – Surfaces de réponse des paramètres du modèle M0 en fonction du critère KGE pour le Gardon de Saint-Jean à Corbès. Les deux lignes correspondent aux deux sous-périodes de calage P1 et P2. Seuls les jeux de paramètres pour lesquels les valeurs de KGE sont supérieures à 90 % de la valeur maximale du KGE pour le bassin et la période considérés ont été retenus ici.

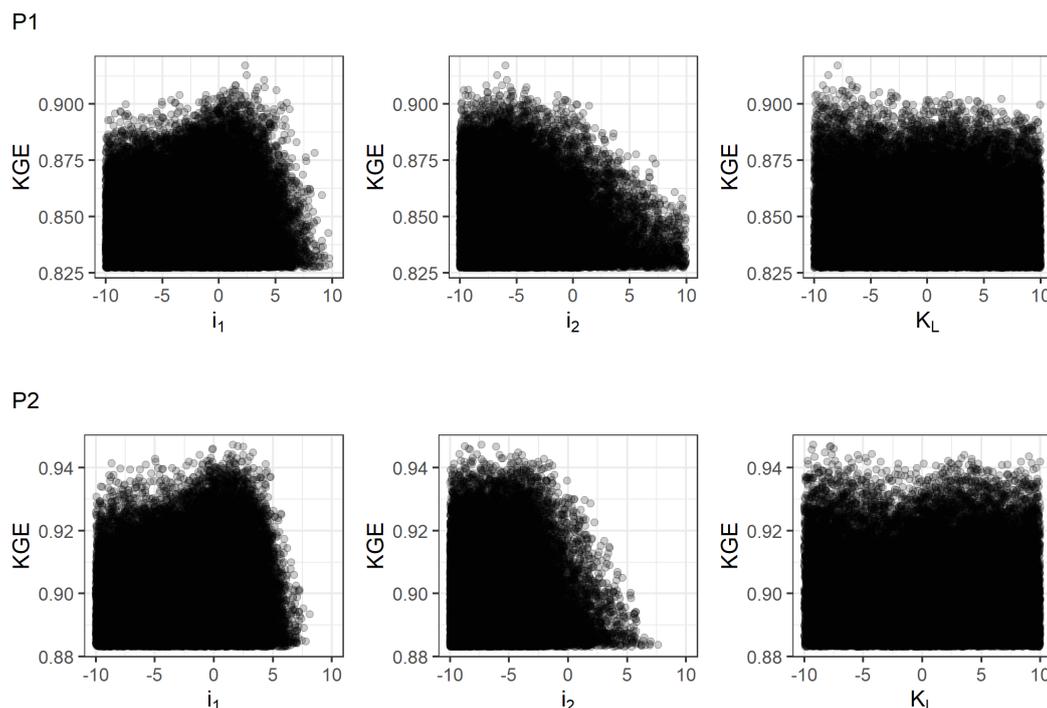


FIGURE 7.5 – Surfaces de réponse des paramètres des fonctions d’intensité du modèle M3 en fonction du critère KGE pour le bassin du Gardon de Saint-Jean à Corbès. Les deux lignes correspondent aux deux sous-périodes de calage P1 et P2. Seuls les jeux de paramètres pour lesquels les valeurs de KGE sont supérieures à 90 % de la valeur maximale du KGE pour le bassin et la période considérés ont été retenus ici. Les valeurs des paramètres sont présentées dans l’espace transformé.

Les surfaces de réponse sur le bassin du Gardon sont représentatives d’une grande partie de notre base de données. Pour quantifier ces observations, nous calculons la variation min.–max. des paramètres par bassin et par sous-période selon un seuil d’acceptabilité du critère KGE (fig. 7.6). Plus la variation min.–max. est élevée, plus l’intervalle des paramètres contraints est large, ce qui indique une équi-finalité élevée au regard du critère KGE. On observe tout d’abord que le paramètre de seuil d’échange X_5 et le coefficient de vidange du réservoir linéaire K_L sont les plus difficiles à contraindre. La majorité des valeurs de variation des paramètres est proche de 20, c’est-à-dire la gamme totale de l’intervalle de recherche. Ensuite, les paramètres X_1 , X_3 et X_4 présentent les valeurs les plus faibles de variation min.–max., confirmant ainsi les observations faites sur le Gardon. Les paramètres i_1 et i_2 semblent plus faciles à contraindre que les paramètres X_5 et K_L . Les distributions de variation min.–max. sont toutefois éloignées des distributions de X_1 , X_3 et X_4 . La distribution de X_2 se situe entre les deux. On observe très peu de différences entre les sous-périodes P1 et P2. Les difficultés d’identification des paramètres ne semblent donc pas spécifiques à la période.

Interactions entre les paramètres

Les difficultés d’identification des paramètres peuvent aussi révéler des interactions entre certains paramètres. Par exemple, les paramètres de capacité des deux réservoirs (X_1 et X_3) du modèle M0 interagissent sur plusieurs bassins. La figure 7.7 illustre cette interaction sur le bassin du Doux pour les jeux de paramètres associés à des valeurs de KGE supérieures à

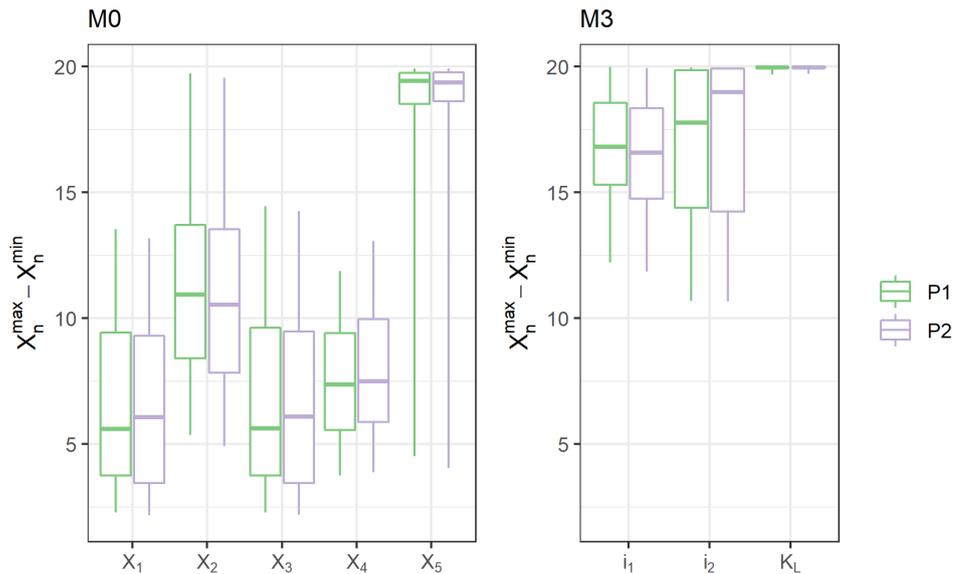


FIGURE 7.6 – Distribution de la variation des paramètres par bassin et par période. Seuls les jeux de paramètres pour lesquels les valeurs de KGE sont supérieures à 95 % de la valeur maximale du KGE pour le bassin et la période considérés ont été retenus ici. Les valeurs de variation sont exprimées dans l'espace des paramètres transformés. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. Les valeurs des paramètres sont présentées dans l'espace transformé.

90 % de la valeur maximale du KGE. On observe sur la figure qu'il y a une anti-corrélation très forte entre les deux paramètres sur ce bassin. Cela signifie que le modèle a des performances équivalentes pour un réservoir de production grand et un réservoir de routage petit et lorsque le réservoir de production est petit et le réservoir de routage est grand. Étant donné que les deux paramètres sont bien identifiés au calage pour la majorité des bassins, et que les variations entre périodes de calage sont faibles par rapport aux autres paramètres, l'algorithme de calage est suffisant pour trouver un optimum pour le bassin malgré cette interaction. En effet, on observe que les valeurs les plus élevées du KGE ne sont pas situées aux extrémités des deux nuages de points et que les valeurs correspondantes des paramètres sont proches entre les deux périodes.

L'interaction entre les paramètres de la fonction d'échange (X_2 et X_5) et la capacité du réservoir de routage (X_3) semble causer davantage de problèmes d'identification. On observe deux modes très distincts sur une grande majorité des bassins. Ces deux modes sont représentés par les asymptotes sur la figure 7.7 :

- Lorsque X_2 est positif, X_5 varie fortement lorsque X_2 tend vers 0^+ . X_5 tend vers une valeur seuil ($\simeq 0.25$) lorsque X_2 tend vers des valeurs élevées. Plus X_3 augmente, plus les points sont éloignés de $X_2 = 0$ et $X_5 = 0.25$.
- Lorsque X_2 est négatif, X_5 varie fortement lorsque X_2 tend vers 0^- . X_5 tend vers une valeur seuil ($\simeq 0.25$) lorsque X_2 tend vers des valeurs négatives importantes. Plus X_3 augmente, plus les points sont éloignés de $X_2 = 0$ et $X_5 = 0.25$.

La présence de ces deux modes signifie tout d'abord que le modèle a des performances équivalentes pour deux méthodes d'estimation des paramètres très différentes. La fonction d'échange du modèle GR5H a un fonctionnement complexe et difficile à interpréter. D'après l'équation 1.14 (cf. chapitre 1), lorsque X_2 est négatif, si le taux de remplissage du réservoir

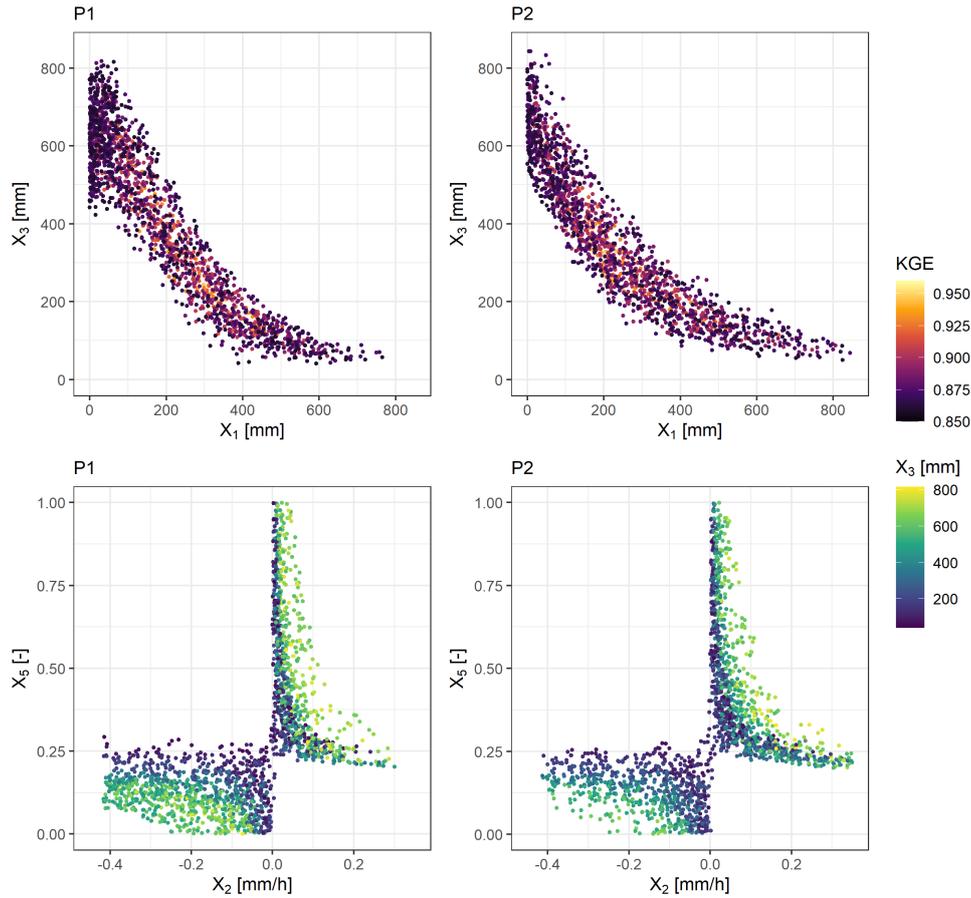


FIGURE 7.7 – Interactions entre les paramètres du modèle M0 pour le bassin du Doux à Colombier-Le-Vieux. Seuls les jeux de paramètres pour lesquels les valeurs de KGE sont supérieures à 90 % de la valeur maximale du KGE pour le bassin et la période considérée ont été retenus ici.

de routage est supérieur à X_5 , le flux est négatif et lorsque c'est l'inverse, le flux est positif. Cela signifie grossièrement que le modèle perd de l'eau en crue par la fonction d'échange et en gagne en étiage. De la même manière, lorsque X_2 est positif, le modèle perd de l'eau en étiage et en gagne en crue. L'intensité du flux est déterminée par X_2 mais aussi par la différence entre le taux de remplissage du réservoir de routage et le paramètre X_5 . Les deux paramètres ont donc un rôle initialement différencié, l'un contrôlant les pertes moyennes du bassin (X_2) et l'autre la saisonnalité de ces pertes (X_5). In fine, leur fonctionnement interagit car les deux paramètres contrôlent le flux total.

Lorsque X_2 a des valeurs proches de zéro, les flux calculés sont plus faibles, ce qui explique l'insensibilité de X_5 car la fonction d'échange a moins d'impact sur le débit. Le paramètre X_3 compense cette diminution du flux en diminuant, augmentant ainsi le taux de remplissage moyen du réservoir et donc la différence potentielle avec le paramètre X_5 .

X_5 semble tendre vers une valeur seuil, ici $X_5 = 0.25$. Cette valeur seuil pourrait être liée au taux de remplissage du réservoir de routage qui contrôle le signe de l'échange. En effet, si ce taux de remplissage est toujours supérieur à X_5 et que X_2 est négatif, l'échange sur le bassin sera toujours négatif. Cela pourrait expliquer les deux modes observés. Les courbes de la figure 7.8a montrent d'ailleurs que le taux de remplissage moyen du réservoir de routage (moyenne sur tous les pas de temps) est majoritairement situé entre 0.2 et 0.3.

De plus, visuellement, cette valeur seuil correspond à l'endroit où la courbe de vidange du réservoir de routage augmente rapidement (fig 7.8b). Cela indiquerait que la valeur seuil de X_5 correspond à la transition entre le comportement en étiage et le comportement en crue simulés par le modèle.

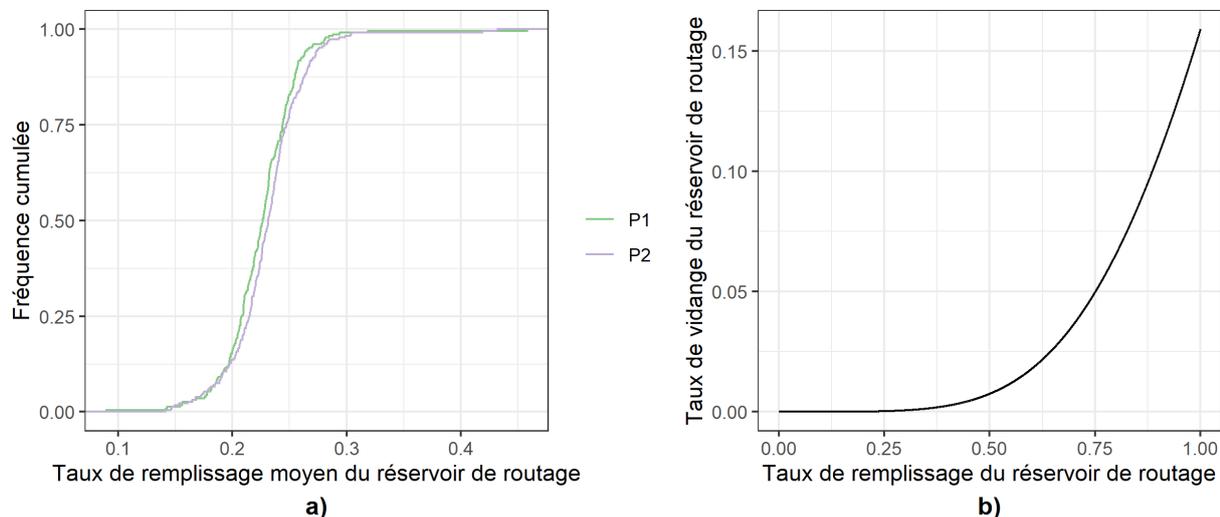


FIGURE 7.8 – Distribution du taux de remplissage moyen du réservoir de routage (moyenne sur tous les pas de temps) du modèle M0 en validation pour les 229 bassins versants (a). Courbe de vidange du réservoir de routage en fin de pas de temps (b).

Ces interactions entre paramètres montrent que l'information portée par le débit et l'utilisation d'une fonction objectif globale (dont une des composantes est le biais) ne permet peut-être pas de paramétrer la saisonnalité de l'échange mais simplement le bilan en eau du bassin considéré. Ce comportement est d'ailleurs celui qui est simulé par la version à quatre paramètres du modèle (voir Perrin *et al.*, 2003). Dans notre configuration, on fait face à un problème sous-informé qui contient trop de degrés de liberté. Au calage, l'optimum mathématique est caractérisé par un des deux modes de fonctionnement sur une période et vice versa, engendrant ainsi une incertitude paramétrique très élevée.

Identifiabilité des paramètres au regard d'autres critères d'erreur

Comme évoqué au début de ce chapitre, une manière de contraindre l'équifinalité est d'explorer d'autres critères d'erreur. En particulier, on peut se demander s'il est adéquat d'utiliser une fonction objectif calculée sur toute la chronique pour estimer les paramètres des fonctions d'intensité qui ont un rôle très événementiel. En outre, nous avons trouvé dans le chapitre 2 que le KGE global n'est pas corrélé avec les biais en crue en été (voir fig. 2.10). La figure 7.9 présente les variations min.-max. pour chaque paramètre et différentes fonctions objectif.

On observe que les différences de contrainte entre fonctions objectif ne sont pas très élevées pour les paramètres X_1 (capacité du réservoir de production), X_2 (coefficient d'échange) et X_3 (capacité du réservoir de routage). Les variations min.-max. du paramètres X_4 (demi-temps de base des HU) sont beaucoup plus faibles lorsque les critères événementiels sont utilisés, ce qui indique que le paramètre a un rôle très important pour simuler les événements. Étonnement, les variations min.-max. de X_5 (seuil d'échange) sont plus faibles lorsque les critères

événementiels sont utilisés, en particulier le NSE sur les événements d'été. Les paramètres des fonctions d'intensités (i_1 , i_2 et K_L) sont plus largement contraints lorsque les critères événementiels sont utilisés pour la sélection des jeux de paramètres, en particulier lorsque le NSE sur les événements d'été est utilisé. Ces résultats indiquent que les paramètres du modèle ayant un rôle événementiel pourraient être mieux identifiés en utilisant une fonction objectif qui donne plus de poids aux événements de crue.

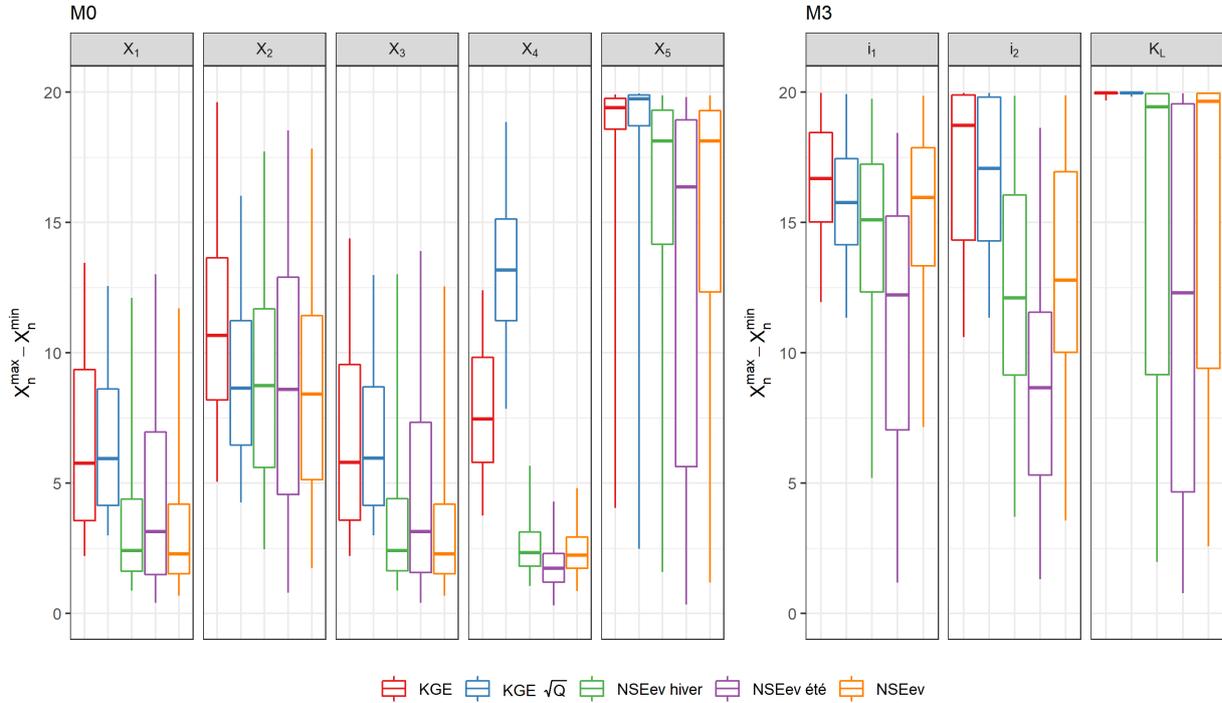


FIGURE 7.9 – Distribution de la variation des paramètres de M0 et des paramètres des fonctions d'intensité de M3 par bassin et par période en fonction de cinq critères d'erreur. Seuls les jeux de paramètres pour lesquels les valeurs d'un critère donné sont supérieures à 95 % de la valeur maximale de ce critère pour le bassin et la période considérés ont été retenus ici. Les valeurs de variation sont exprimées dans l'espace des paramètres transformés. $KGE\sqrt{Q}$: critère KGE calculé sur la racine carrée des débits. NSEev hiver : critère NSE calculé sur les événements d'hiver (une valeur par période et par bassin considérés ; voir eq. 2.3 pour un calcul similaire avec le biais). NSEev été : critère NSE calculé sur les événements d'été. NSEev : critère NSE calculé sur tous les événements. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Les faibles valeurs de variation min.–max. n'indiquent pas nécessairement que les paramètres convergent vers des paramètres optimaux pour le bassin. En effet, l'utilisation de critères sur les événements peut donner lieu à une dégradation des performances en contrôle si le nombre de points utilisés au calage n'est pas suffisant. Des problèmes d'insensibilité ou, à l'inverse, de surcalage peuvent ainsi avoir lieu. De plus, il est possible que l'utilisation de critères événementiels dégrade l'identification des paramètres qui jouent un rôle sur le bilan en eau et donc sur les conditions antécédentes. Il faut donc explorer les compromis entre la fonction objectif globale et une fonction objectif qui met plus de poids sur les hauts débits. La figure 7.10 présente un exemple de compromis entre un critère événementiel et un critère global sur le bassin du Gardon. Pour le modèle M0, on observe que, pour augmenter la valeur

du critère événementiel, il faut nécessairement diminuer la valeur du critère global. De plus, il existe plusieurs jeux de paramètres avec des valeurs très différentes des deux critères. Pour le modèle M3, on observe que pour des valeurs très proches de NSE global, il existe des valeurs différentes de NSE sur les événements d'été. Cela signifie qu'il serait possible de trouver un compromis qui améliorerait la performance pour les crues d'été mais qui ne dégraderait pas trop fortement les performances globales. Le calage mono-objectif utilisé dans nos travaux ne permet pas de trouver un tel compromis.

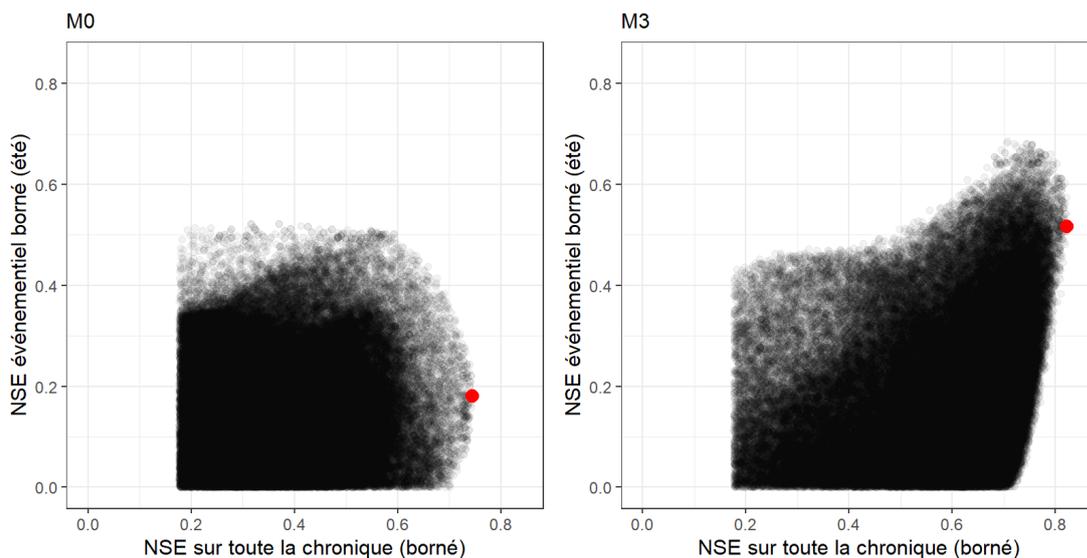


FIGURE 7.10 – Performances événementielles en été en fonction des performances globales pour plusieurs jeux de paramètres des modèles M0 et M3 sur le bassin du Gardon de Saint-Jean à Corbès (période P2). Le NSE borné est utilisé pour caractériser les performances globales afin d'avoir la même nature de critère d'erreur en abscisse et en ordonnée. Les points rouges désignent les valeurs maximales du NSE global.

7.3.2 Voies d'amélioration

Peut-on fixer le seuil d'échange (paramètre X_5) ?

Pour réduire l'incertitude paramétrique associée à la fonction d'échange, nous testons une configuration où X_5 a la même valeur pour tous les bassins. On observe une baisse des performances globales en calage pour toutes les valeurs de X_5 testées par rapport à la version où le paramètre est libre (fig. 7.11). En validation, pour certaines valeurs du paramètre, les performances sont similaires aux performances de la version où le paramètre est libre. Les valeurs optimales de X_5 ne sont pas les mêmes entre périodes. Cependant, si on prend la valeur optimale sur la période P1 au calage, et qu'on regarde les performances associées en validation sur la période P2, on remarque que la distribution des performances est très similaire à la valeur optimale sur P2. Cette observation est valable pour l'expérience inverse.

Les performances événementielles baissent en hiver et augmentent en été lorsque X_5 est fixé (fig. 7.12), ce qui indique que l'estimation des paramètres de la fonction d'échange a un impact sur la simulation des crues en été. Ces résultats contrastent avec les résultats trouvés au chapitre 4, où nous avons vu que les erreurs du modèle en crue ne sont pas corrélées avec

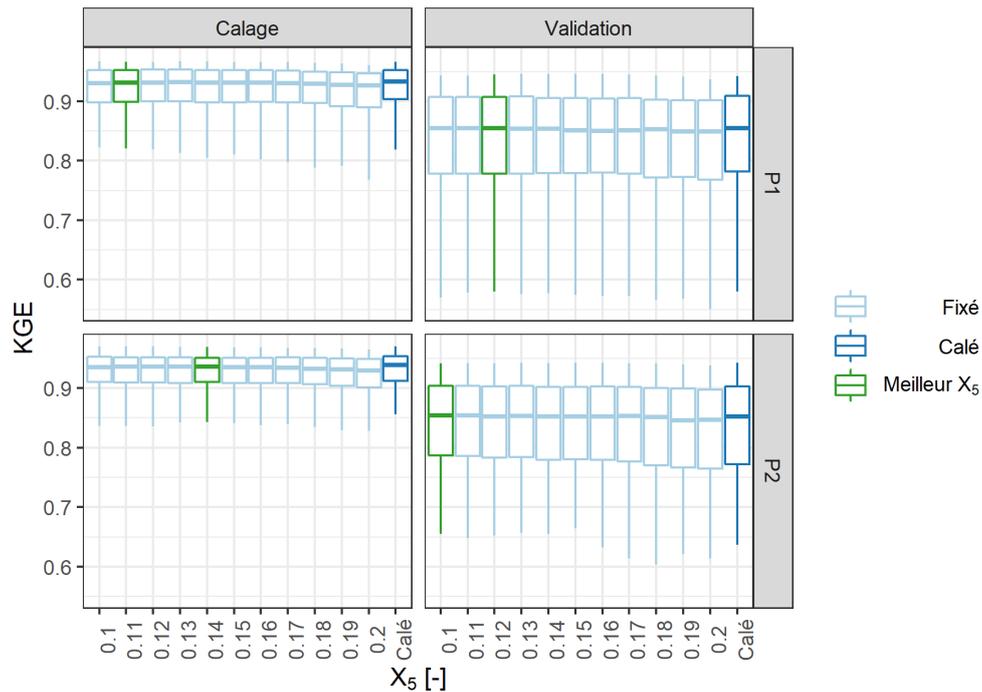


FIGURE 7.11 – Distribution des performances globales (KGE) du modèle M0 en calage et validation pour plusieurs valeurs du paramètre X_5 fixé pour tous les bassins par période considérée. Les boîtes à moustaches vertes représentent les distributions de KGE dont la moyenne est la plus élevée (sans compter la version où X_5 est libre). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

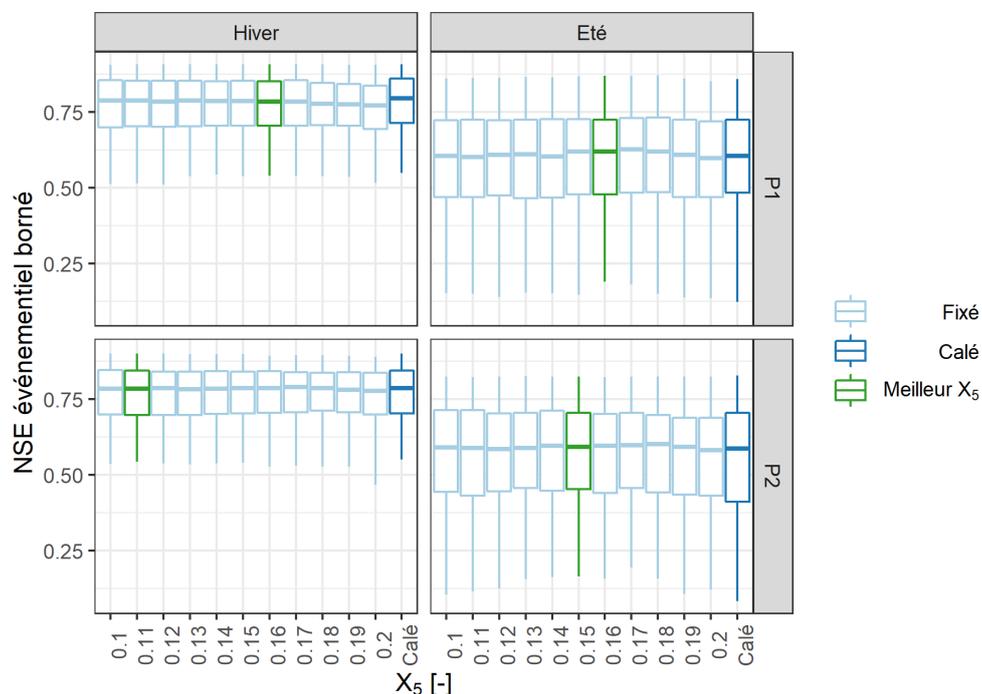


FIGURE 7.12 – Distribution des performances événementielles du modèle M0 (NSEEv) en hiver et en été pour plusieurs valeurs du paramètre X_5 fixé pour tous les bassins par période considérée. Les boîtes à moustaches vertes représentent les distributions de NSE événementiel dont la moyenne est la plus élevée (sans compter la version où X_5 est libre). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

les flux d'échange. Les différences de performance sont toutefois faibles entre les différentes valeurs de X_5 testées. Comme pour le KGE global, on observe une différence entre les deux sous-périodes. Les valeurs optimales de X_5 pour le NSE événementiel sont différentes des valeurs optimales pour le KGE global.

Le tableau 7.2 présente les performances moyennes du modèle M0 pour les paramètres X_5 génériques optimaux (c'est-à-dire fixés à une même valeur pour tous les bassins). On observe que les performances moyennes de M0 changent très peu lorsqu'on utilise une valeur unique de X_5 pour tous les bassins.

TABLE 7.2 – Performances moyennes du modèle M0 pour deux méthodes d'estimation du paramètre X_5 . Les valeurs des critères sont des moyennes des valeurs calculées sur chaque bassin.

Période	Configuration	X_5 [-]	X_5 fixé		X_5 calé	
			KGE	NSEev	KGE	NSEev
P1	Calage	0.11	0.916	0.738	0.919	0.742
P2	Calage	0.14	0.924	0.743	0.927	0.750
P1	Validation	0.14	0.819	0.707	0.819	0.710
P2	Validation	0.11	0.827	0.714	0.825	0.714

Peut-on fixer le coefficient du réservoir linéaire (paramètre K_L) ?

Nous avons vu précédemment que le paramètre K_L est difficile à contraindre, en particulier lorsqu'on utilise un critère global. Pour réduire la complexité paramétrique du modèle, nous testons une configuration où K_L a la même valeur pour tous les bassins (le paramètre X_5 est libre ici). On observe une baisse des performances globales en calage (fig. 7.13) entre la version avec le meilleur K_L fixé et la version avec K_L libre, cela pour les deux sous-périodes de calage. Les performances sont équivalentes en validation sur P1 et dégradées sur P2. Les différences de performance globale moyenne sont cependant très faibles (quelques centièmes de KGE, voir tableau 7.3). Les valeurs optimales de K_L sont très proches entre périodes de calage. Les valeurs optimales en validation sont plus faibles que les valeurs optimales en calage. Les performances varient peu entre les différentes valeurs de K_L .

Les résultats de la figure 7.14 indiquent que les performances événementielles en hiver (en validation) de la version avec le paramètre K_L optimal sont proches des performances de la version avec le paramètre K_L libre, cela pour les deux sous-périodes. En été, les performances baissent pour les deux sous-périodes. Les valeurs optimales de K_L en validation selon le NSE événementiel en hiver sont proches des valeurs optimales obtenues en calage selon le KGE global. En été, les valeurs optimales sont très différentes entre les deux périodes et éloignées des valeurs optimales obtenues en calage à partir des valeurs de KGE global.

Le tableau 7.3 présente les performances moyennes du modèle M3 pour les K_L génériques optimaux. On observe que les performances moyennes de M3 changent très peu lorsqu'on utilise une valeur unique de K_L pour tous les bassins.

Comme les résultats précédents montrent qu'il est possible de fixer les paramètres X_5 et K_L sans impacter trop fortement les performances moyennes du modèle, nous proposons de fixer ces deux paramètres pour le modèle M3. Nous choisissons les valeurs $X_5 = 0.125$ et $K_L = 0.115$. Le tableau 7.4 présente les performances du modèle M3 avec ces valeurs fixées.

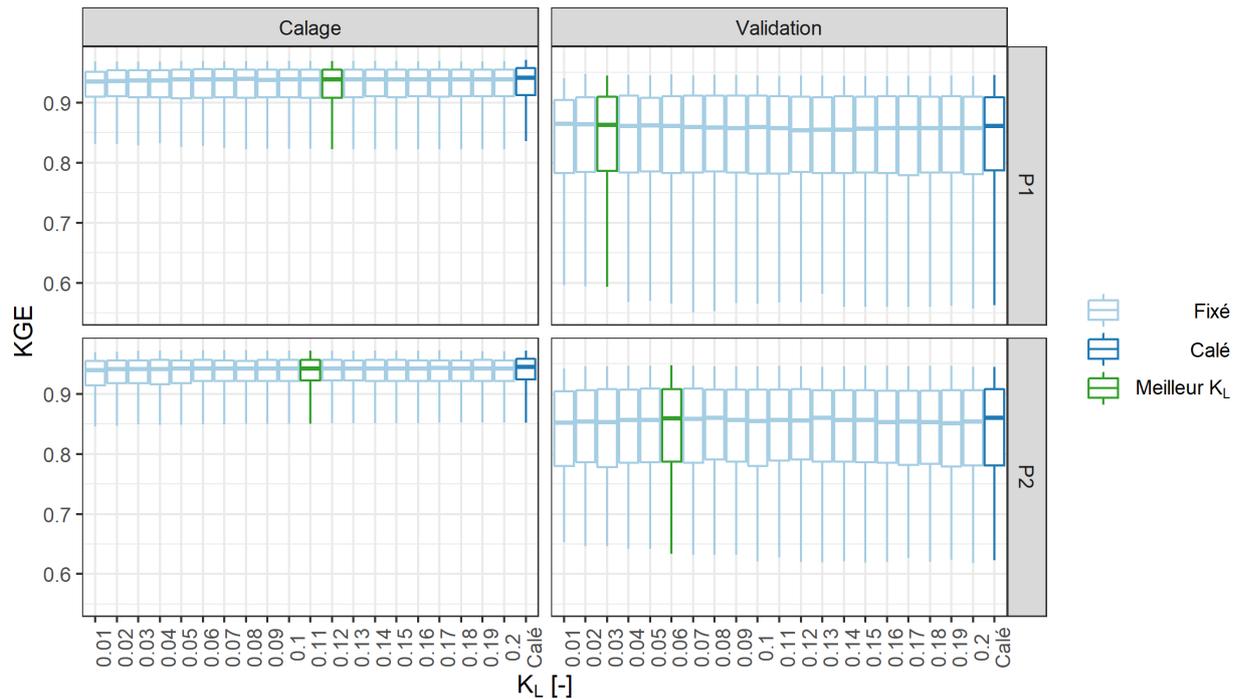


FIGURE 7.13 – Distribution des performances globales (KGE) du modèle M3 en calage et validation pour plusieurs valeurs du paramètre K_L fixé pour tous les bassins par période considérée. Les boîtes à moustaches vertes représentent les distributions de KGE dont la moyenne est la plus élevée (sans compter la version où K_L est libre). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

TABLE 7.3 – Performances moyennes du modèle M3 pour deux méthodes d'estimation du paramètre K_L . Les valeurs des critères sont des moyennes des valeurs calculées sur chaque bassin.

Période	Configuration	K_L fixé			K_L calé	
		K_L [-]	KGE	NSEev	KGE	NSEev
P1	Calage	0.12	0.924	0.764	0.927	0.768
P2	Calage	0.11	0.931	0.767	0.933	0.772
P1	Validation	0.11	0.823	0.726	0.823	0.731
P2	Validation	0.12	0.828	0.729	0.831	0.731

TABLE 7.4 – Performances moyennes du modèle M3 avec les paramètres X_5 et K_L fixés à 0.125 et 0.115 respectivement. Les valeurs des critères sont des moyennes des valeurs calculées sur chaque bassin et chaque sous-période.

Période	Configuration	X_5 et K_L fixés		X_5 et K_L calés	
		KGE	NSEev	KGE	NSEev
P1	Calage	0.922	0.761	0.927	0.768
P1	Validation	0.823	0.722	0.823	0.731
P2	Calage	0.928	0.761	0.933	0.772
P2	Validation	0.829	0.726	0.831	0.731

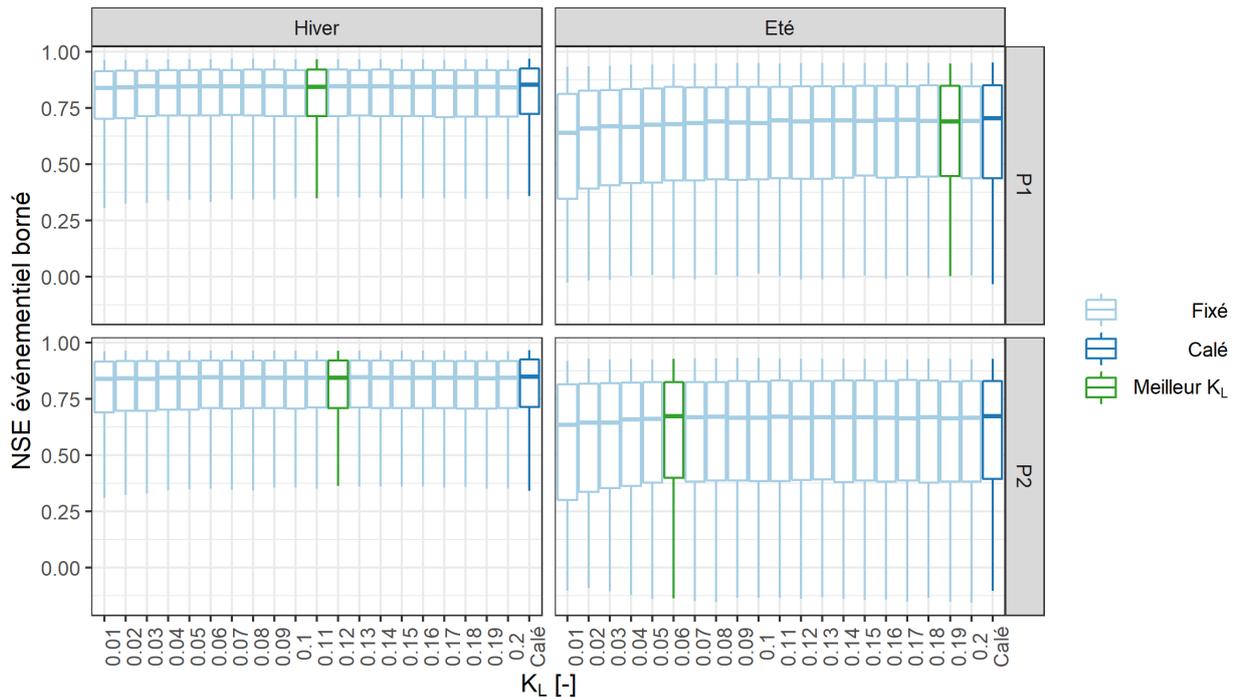


FIGURE 7.14 – Distribution des performances événementielles (NSEEv) du modèle M3 en hiver et en été pour plusieurs valeurs du paramètre K_L fixé pour tous les bassins par période considérée. Les boîtes à moustaches vertes représentent les distributions de NSE événementiel dont la moyenne est la plus élevée (sans compter la version où K_L est libre). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Paramétrage multi-objectif

On se demande maintenant s'il est possible d'améliorer la simulation des crues à partir d'une méthode d'estimation des paramètres qui met plus de poids sur les hauts débits. Les résultats précédents ont montré que les paramètres qui ont une dynamique plus événementielle semblent être mieux contraints par les critères calculés sur les événements de crue. La figure 7.15 présente les performances globales et événementielles lorsque les modèles sont paramétrés en utilisant une fonction multi-objectif (voir section 7.2 et eq. 7.1). En calage et en validation, et pour les deux sous-périodes, les valeurs de KGE global sont moins élevées pour la version avec le critère multi-objectif que les valeurs de KGE global de la version avec le KGE comme fonction objectif. C'est un résultat attendu puisque le critère multi-objectif met moins de poids sur le KGE. En revanche, le NSE événementiel augmente en calage et en validation pour les deux modèles, cela pour les deux sous-périodes. Le critère événementiel augmente pour les événements d'hiver et les événements d'été. Bien que les critères de biais et de pics en été n'aient pas été utilisés au calage, ils sont tous deux améliorés pour les deux modèles en calage et en validation. Les différences de performance entre les deux méthodes d'estimation des paramètres sont du même ordre de grandeur pour les deux modèles. L'ordre des performances entre les deux modèles reste systématiquement le même. Contrairement aux conclusions de Mizukami *et al.* (2019), l'inclusion d'un objectif de simulation des crues au calage ne donne pas lieu à une baisse des performances pour les critères portant sur les hauts débits. Plusieurs raisons peuvent expliquer ces différences. Tout d'abord, la fonction objectif que nous avons utilisée n'intègre que partiellement un critère calculé sur les événe-

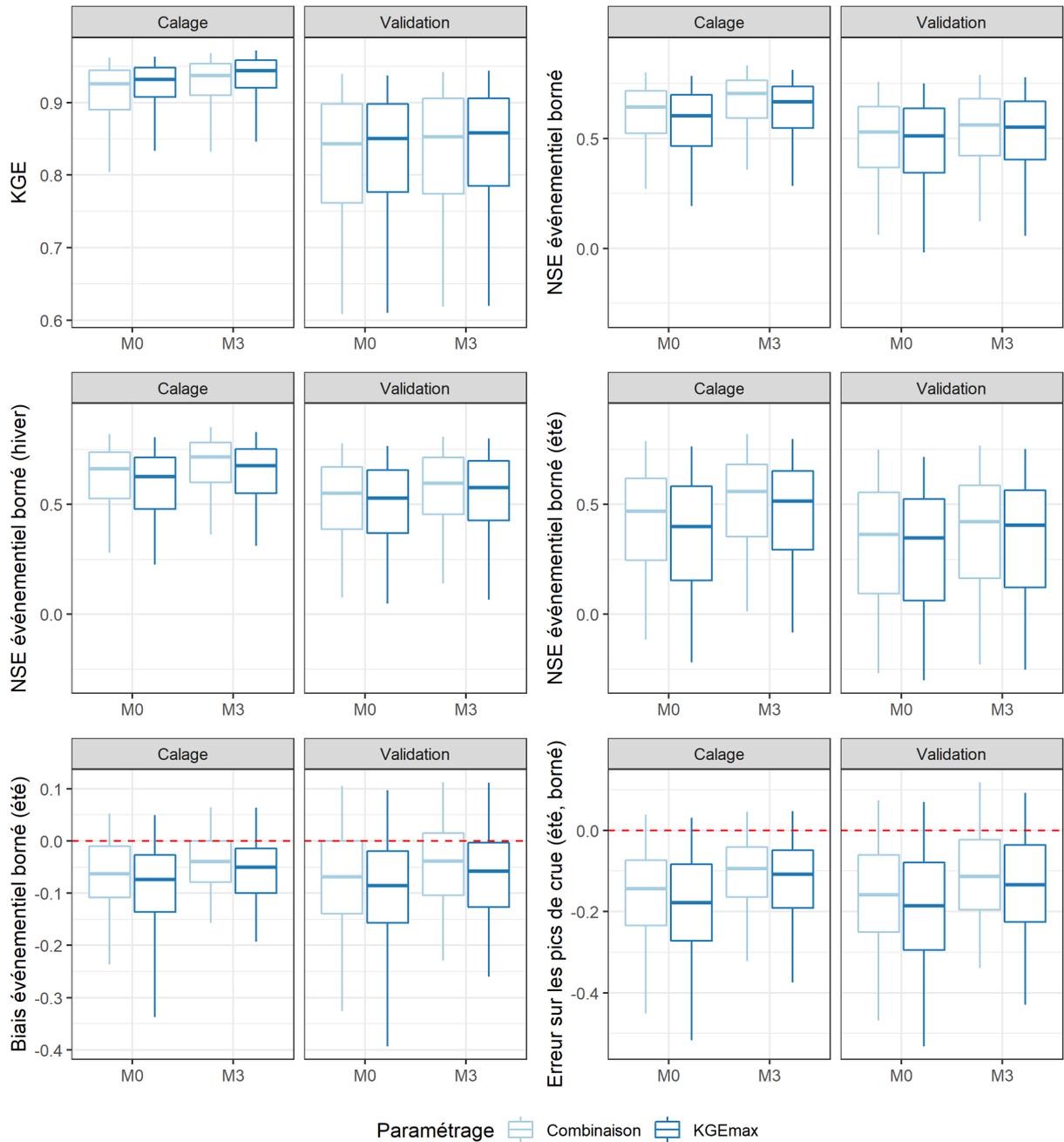


FIGURE 7.15 – Comparaison des performances globales et événementielles des modèles M0 et M3 pour deux méthodes d’estimation des paramètres. Combinaison : sélection du jeu de paramètres optimal à partir d’un critère composite incluant un critère événementiel (eq. 7.1). KGEmax : sélection du jeu de paramètres optimal à partir de la valeur maximale du critère KGE. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

ments, ce qui n'est pas le cas du critère APFB (*annual peak flow bias*) utilisé par Mizukami *et al.* (2019). Ce critère est défini comme le ratio entre la moyenne des pics annuels simulés et la moyenne des pics annuels observés. Ensuite, nous n'avons pas calculé les mêmes critères pour caractériser les hauts débits. Enfin, leur critère se base sur les pics annuels, ce qui signifie que très peu de points portent l'erreur au calage comparé aux critères qui sont calculés sur tous les événements.

La figure 7.16 présente les différences de performances en été pour les deux méthodes d'estimation des paramètres selon le bassin considéré.

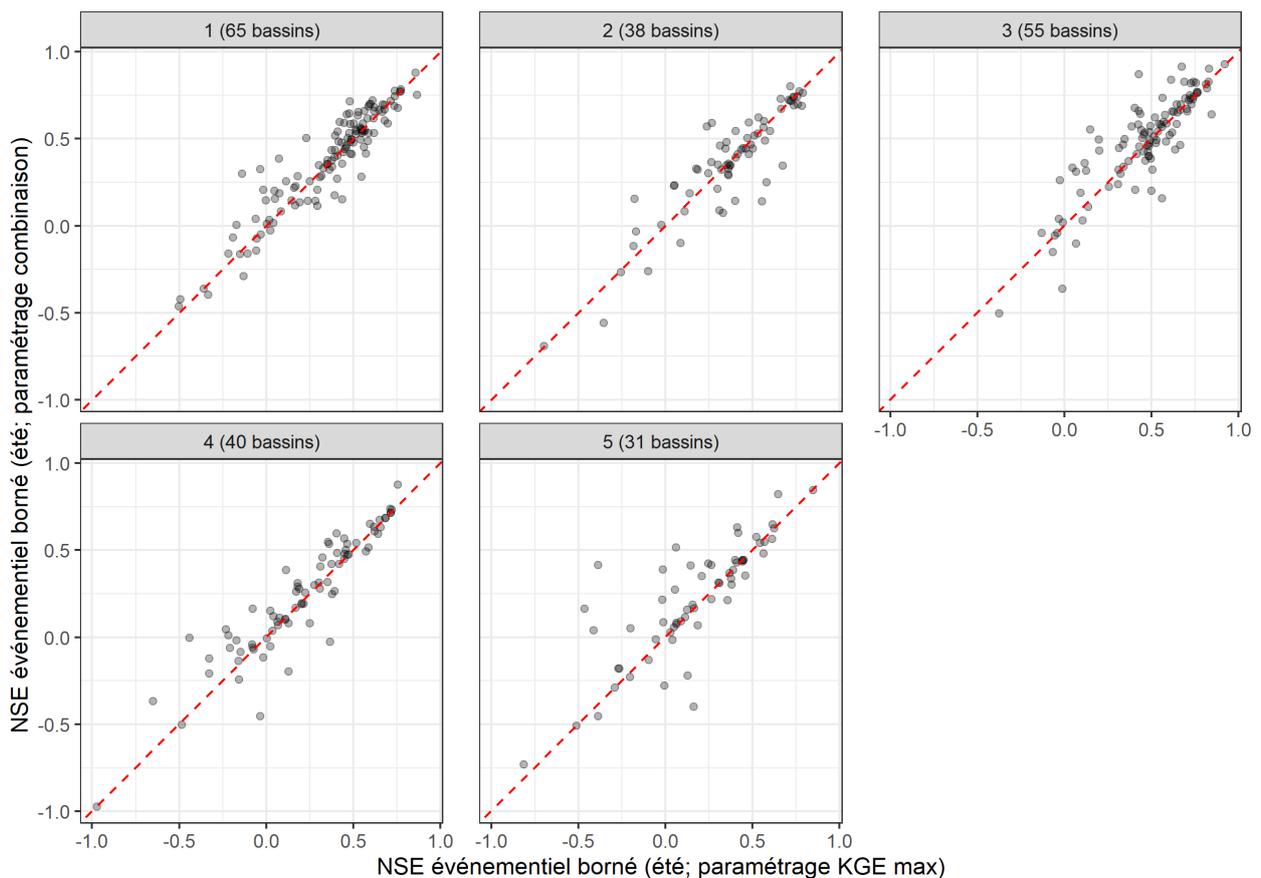


FIGURE 7.16 – Comparaison des performances événementielles du modèle M3 en été et en validation pour deux méthodes d'estimation des paramètres en fonction de cinq groupes de bassins versants. Combinaison : sélection du jeu de paramètres optimal à partir d'un critère composite incluant un critère événementiel (eq. 7.1). KGE_{max} : sélection du jeu de paramètres optimal à partir de la valeur maximale du critère KGE.

On observe quelques améliorations pour les bassins des groupes 1, 3, 4 et 5. Le paramétrage multi-objectif ne semble pas avoir d'intérêt pour les bassins du groupe 2. Le modèle M3 ne présentait pas d'améliorations significatives par rapport à M0 sur ces bassins (voir chapitre 5). On observe la présence de bassins des groupes 4 et 5 pour lesquels le paramétrage multi-objectif dégrade fortement les performances en été. La Vis à Saint-Laurent-le-Minier (affluent de l'Hérault) fait partie de ces bassins. Le modèle M3 surestime certains pics d'événements résultant de fortes intensités de pluie sur ce bassin. Cependant, la forme des hydrogrammes

de ces événements est caractéristique d'un débordement de la rivière. Le paramétrage multi-objectif exacerbe la surestimation de ces événements.

Les valeurs des paramètres des fonctions d'intensité (i_1 et i_2) sont en effet plus élevées sur notre base de données lorsque la méthode d'estimation des paramètres avec la fonction multi-objectif est utilisée (fig. 7.17). Le paramètre K_L prend des valeurs plus faibles.

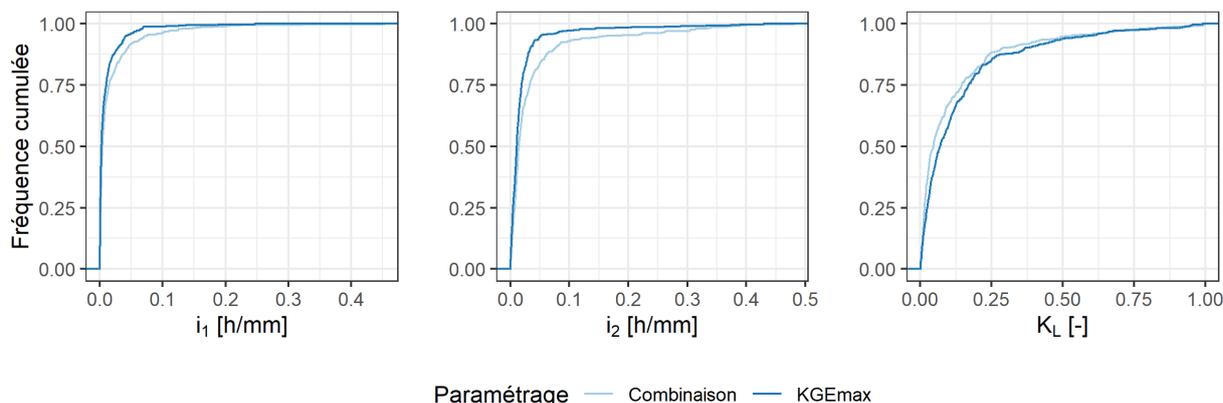


FIGURE 7.17 – Distributions des valeurs des paramètres des fonctions d'intensité du modèle M3 pour deux méthodes d'estimation des paramètres. Combinaison : sélection du jeu de paramètres optimal à partir d'un critère composite incluant un critère événementiel (eq. 7.1). KGE_{max} : sélection du jeu de paramètres optimal à partir de la valeur maximale du critère KGE.

7.4 Conclusion

L'objectif de ce chapitre était de caractériser les problèmes d'identification des paramètres des modèles utilisés dans nos travaux, cela afin de proposer des solutions pour réduire la complexité paramétrique et améliorer la simulation des crues. Pour répondre à cet objectif, nous avons généré un grand ensemble de jeux de paramètres des modèles M0 et M3 afin d'explorer l'identifiabilité des différents paramètres au regard de plusieurs critères d'erreur. Les résultats ont permis de mettre en évidence les difficultés d'identification des paramètres de la fonction d'échange et des fonctions introduites dans les travaux sur la structure du modèle. Ces difficultés se traduisent par des variations importantes des paramètres entre périodes de calage ainsi qu'une équifinalité élevée sur la plupart des bassins. Les paramètres de la fonction d'échange interagissent avec le réservoir de routage selon deux modes de fonctionnement caractérisés par des valeurs très distinctes des paramètres mais des performances équivalentes. L'information portée par un critère d'erreur calculé sur les débits ne suffit pas à déterminer le mode de fonctionnement qui est le plus adéquat pour le bassin considéré. Pour réduire l'incertitude paramétrique, nous avons cherché à fixer les paramètres X_5 et K_L . Les résultats ont montré qu'il est possible de fixer ces paramètres sans impacter trop fortement les performances moyennes du modèle. Cette analyse contraste avec les résultats trouvés au chapitre 5, où la recherche de généricité pour les paramètres i_1 et i_2 avait donné lieu à une baisse des performances en crue. L'utilisation de critères événementiels permet de réduire la gamme de variation des paramètres qui ont un rôle important à l'échelle événementielle, en

particulier le demi-temps de base de l'hydrogramme unitaire et les paramètres des fonctions d'intensité. Pour cette raison, nous avons testé une méthode d'estimation des paramètres basée sur une fonction objectif qui intègre une composante événementielle. Cette configuration permet d'améliorer les performances en crue des modèles M0 et M3, en particulier la simulation des volumes et pics de crue en été.

Les modifications de la structure qui visent à améliorer la simulation des crues, combinées à l'utilisation d'une fonction d'échange à deux paramètres, impliquent une incertitude paramétrique élevée. Les résultats présentés dans ce chapitre indiquent que l'information portée par un critère d'erreur comme le KGE ne suffit pas pour garder autant de degrés de liberté. Le problème est donc sous-informé au vu de l'objectif que nous utilisons. Nous avons montré qu'il est possible de réduire cette incertitude paramétrique en fixant certains paramètres et qu'il est possible d'améliorer le calage à partir d'objectifs plus adaptés pour la simulation des crues. Il y a donc un compromis à trouver entre la complexification de la structure et l'incertitude paramétrique. Ce compromis aura un rôle important dans le développement du modèle de prévision. Une perspective de ce travail serait d'utiliser des données additionnelles au débit pour continuer à réduire l'incertitude paramétrique (e.g. des données satellite d'humidité des sols, cf. Nijzink *et al.*, 2018).

Les résultats de ce chapitre sont basés sur des choix méthodologiques qui peuvent en partie sembler arbitraires. En particulier, l'utilisation de seuils de critères pour analyser la variation des paramètres associés aux performances les plus élevées est discutable. Nous avons testé plusieurs valeurs de seuils de sélection des jeux de paramètres considérés comme « acceptables ». Les conclusions de nos analyses ne semblent pas dépendre de ces seuils (non montré ici). Par ailleurs, les fonctions de transformation des paramètres que nous avons utilisées dérivent des travaux précédents sur les modèles GR. Nous avons adapté ces fonctions à la nouvelle structure que nous utilisons ainsi qu'à l'exercice d'exploration de l'espace des paramètres qui s'est avéré coûteux en temps de calculs. Des travaux complémentaires pourraient porter sur l'influence des fonctions de transformation des paramètres sur la convergence de la méthode d'estimation des paramètres.

Partie III

**Impact de l'amélioration
de la structure du modèle
hydrologique
sur la prévision des crues**

Chapitre 8

Construction et évaluation d'un modèle de prévision

8.1 Contexte

L'hypothèse de départ de nos travaux est que l'amélioration des performances d'un modèle hydrologique en simulation devrait conduire à une amélioration des performances en prévision des crues, en particulier lorsque les bassins versants réagissent à des intensités de pluie élevées. Pour cette raison, nous avons décidé de mener les travaux sur l'amélioration du modèle indépendamment de méthodes visant à exploiter l'information sur le débit observé par des techniques d'assimilation (e.g. mise à jour des états ou paramètres du modèle) ou de post-traitements (e.g. basés sur des modèles d'erreur). Nous avons ainsi effectué nos travaux dans un cadre de simulation et non de prévision des débits, cela en utilisant le « meilleur » modèle de simulation à notre disposition, ici le modèle GR5H. L'objectif des travaux de ce chapitre est de vérifier la validité de notre postulat de départ, en comparant les performances du modèle que nous avons développé avec les performances du modèle GRP – utilisé par de nombreux services opérationnels en France – dans un cadre de prévision déterministe. L'infirmité de ce postulat signifierait que l'assimilation de données et/ou le post-traitement sont suffisants pour compenser les limites structurelles du modèle que nous avons cherché à traiter.

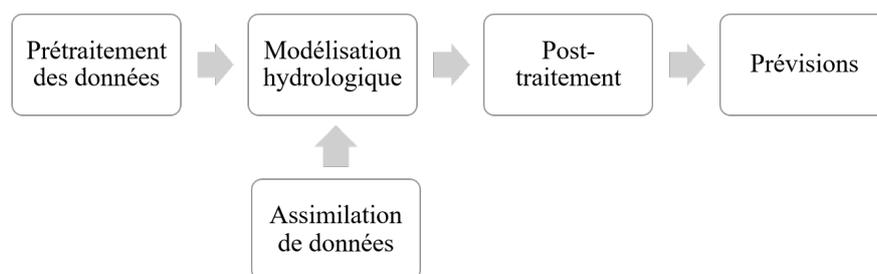


FIGURE 8.1 – Schéma représentant la chaîne de modélisation hydrologique utilisée pour la prévision des crues.

Les modèles de prévision des crues utilisés par les services opérationnels (en France et dans de nombreux pays) sont généralement couplés à des méthodes de mise à jour des états ainsi que des méthodes de correction des sorties (e.g. Pagano *et al.*, 2014; Pappenberger *et al.*, 2016). L'erreur à (ou jusqu'à) l'instant de prévision est utilisée pour modifier les états initiaux (e.g. Piazzi *et al.*, 2021), les sorties du modèle (e.g. Li *et al.*, 2015) et/ou les paramètres (e.g. Yang & Michel, 2000). Des données additionnelles au débit peuvent également être assimilées pour modifier les états des modèles, e.g. des données satellite d'humidité des sols (e.g. Yan & Moradkhani, 2016) ou encore des données piézométriques (e.g. Pelletier, 2021). Les données de débit restent cependant majoritairement utilisées pour la prévision en temps réel (Piazzi *et al.*, 2021), du fait qu'elles représentent la principale variable cible de la modélisation hydrologique.

L'assimilation séquentielle des débits – c'est-à-dire l'utilisation de l'erreur à chaque pas de temps pour mettre à jour les états d'un modèle – est très utilisée pour améliorer la prévision des modèles pluie-débit (Liu *et al.*, 2012). Le filtre de Kalman (Kalman, 1960) ainsi que de nombreuses variantes de ce filtre comme le filtre de Kalman étendu (Miller *et al.*, 1994) sont fréquemment appliqués pour mettre à jour les états d'un modèle à partir des débits disponibles à l'instant de prévision. Ces filtres permettent d'attribuer un degré de confiance relatif au débit assimilé en temps réel en réalisant des mises à jour partielles des états (Berthet, 2010). Cela permet de tenir compte du fait que ces données sont entachées d'incertitudes liées aux techniques d'observation (e.g. Horner *et al.*, 2018). L'assimilation de données employée dans les modèles de prévision des crues se base de plus en plus sur des méthodes ensemblistes comme le filtre Kalman d'ensemble (Evensen, 1994) ou le filtre particulaire (Arulampalam *et al.*, 2002) dont le formalisme permet une quantification des incertitudes prédictives dans un cadre probabiliste.

Les travaux de Berthet (2010) sur l'assimilation de données dans le modèle GRP ont montré qu'une mise à jour totale du réservoir de routage (ou insertion directe, i.e. basée sur une relation univoque entre le débit simulé par le modèle et le niveau de ce réservoir ; voir fig. 8.3) couplée à une correction des sorties (exploitant le caractère autorégressif de l'erreur du modèle) donnent les meilleures performances sur un large échantillon de bassins versants français. Les performances du modèle pluie-débit GRP – dont la structure est simplifiée par rapport au modèle de simulation GR5H pour faciliter l'exploitation et valoriser au mieux l'information sur le débit observé – sont très dépendantes de l'utilisation de l'erreur à l'instant de prévision. Les choix d'assimilation dans le modèle GRP impliquent que les corrections du modèle sont appliquées de manière déterministe. Pour l'utilisation opérationnelle, la quantification des incertitudes prédictives est effectuée a posteriori dans un cadre de post-traitement statistique avec la méthode QUOIQUE (Bourgin, 2014; Bourgin *et al.*, 2014; Viatgé *et al.*, 2019a). Dans un cadre probabiliste, les prévisions ont plusieurs attributs qui caractérisent leur qualité (Bradley *et al.*, 2016). Généralement, trois de ces attributs sont évalués : la précision, la fiabilité et la finesse. Peredo (2021) les définit de la manière suivante :

- la précision mesure l'erreur entre la prévision et l'observation ;
- la fiabilité mesure la capacité de l'ensemble à reproduire la fréquence d'un événement donné ;
- la finesse correspond à la dispersion de l'ensemble.

Dans un cadre déterministe, on évalue seulement la précision des prévisions.

Afin de comparer équitablement les performances en prévision du modèle développé dans cette thèse avec le modèle de référence GRP, nous avons besoin d'utiliser la même méthode de mise à jour. Contrairement aux méthodes ensemblistes, les méthodes de mise à jour par insertion directe dépendent du modèle utilisé. Par exemple, un modèle constitué de deux branches de routage et de deux réservoirs de routage ne peut pas être mis à jour de la même manière qu'un modèle constitué d'une seule branche de routage, car à chaque instant de prévision, il existera plusieurs niveaux initiaux des réservoirs permettant de reproduire le débit observé. Il nous faudra donc développer une méthode proche de celle de GRP mais adaptée aux modèles utilisés dans les chapitres précédents. Cette fois-ci, on se retrouve à explorer les compromis entre complexification de la structure du modèle et choix de la méthode d'assimilation. Le compromis optimal pourra dépendre de l'échéance de prévision car l'erreur du modèle augmente avec celle-ci, du fait de l'atténuation progressive de l'effet bénéfique de l'assimilation de données (Li *et al.*, 2009 ; Piazzini *et al.*, 2021).

Dans ce chapitre, nous présentons le modèle GRP et nous proposons une méthode de mise à jour par insertion directe pour les modèles M0 (modèle GR5H initial) et M3 (modèle GR5H modifié). Nous comparons ensuite les performances des trois modèles (sans correction des sorties) à plusieurs échéances en utilisant des critères d'erreur globaux et événementiels. Les performances des modèles GRP et M3 sont ensuite analysées en fonction de plusieurs caractéristiques des bassins et des événements. Enfin, nous présentons l'effet de la correction des sorties sur ces comparaisons.

8.2 Éléments méthodologiques

La terminologie et les méthodes employées en prévision des crues diffèrent du cadre de la simulation des crues. Les modèles testés ici sont toujours utilisés de manière continue (i.e. sur tous les pas de temps). Pour tester les modèles sur les séries passées, on se place successivement à chaque pas de temps de la série et on lance le modèle en prévision comme s'il était utilisé en temps réel. À chaque instant de prévision t , certains états des modèles sont mis à jour à partir du débit observé. Une fois les états initiaux mis à jour, les prévisions sont lancées entre $t + 1$ et $t + H$, H étant l'horizon de prévision (ou échéance de prévision)¹. C'est une manière de se replacer dans un contexte de prévision proche du temps réel. Ainsi, on utilise toute l'information disponible jusqu'à l'instant de prévision pour émettre une prévision dans le futur. En temps réel, on utilise des prévisions de pluie entre $t + 1$ et $t + H$. Pour les travaux de ce chapitre, nous utilisons les scénarios de pluies dites « parfaites », i.e. les chroniques de pluie de bassin observées que nous avons utilisées dans les chapitres précédents. Cela nous permet de comparer les modèles de prévision indépendamment des incertitudes provenant des produits de pluie, ou du moins avec une incertitude moins importante qu'en temps réel.

Les prévisions des modèles GRP, M0 et M3 sont analysées dans ce chapitre. Le calage de ces trois modèles est effectué en simulation. Cela signifie que nous prenons les paramètres calés sur chaque bassin et chaque sous-période indépendamment des procédures de mise à jour des états initiaux. Bien que Berthet (2010) ait montré que le modèle GRP a de

1. Ces deux termes seront utilisés de manière interchangeable dans la suite du chapitre

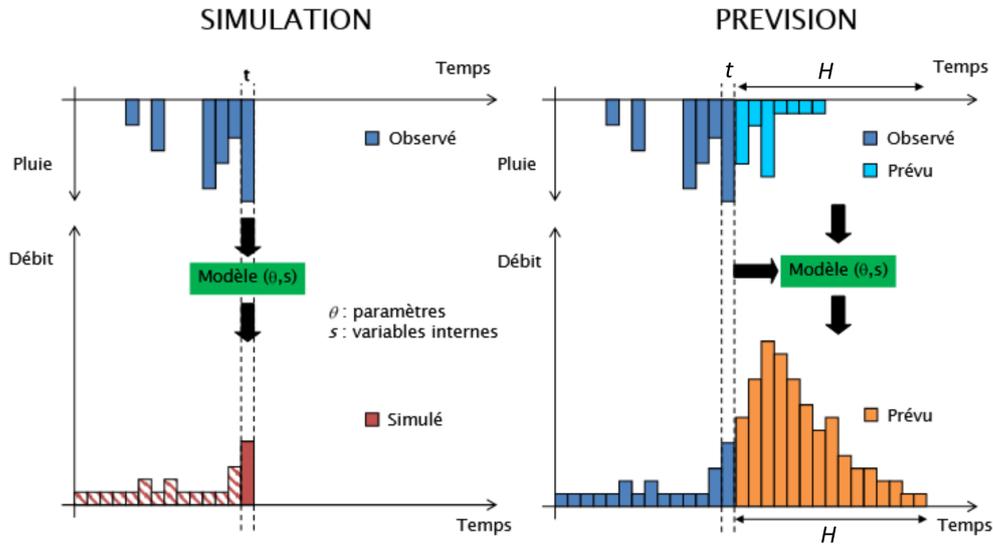


FIGURE 8.2 – Différences entre la simulation et la prévision des crues. H est l'horizon de prévision. t est l'instant de prévision. Schéma modifié à partir de Bourgin (2014).

meilleures performances (en particulier à courte échéance) lorsque qu'il est calé par horizon de prévision, nous souhaitons analyser les performances indépendamment des compensations paramétriques induites par une telle procédure. Nous explorons l'impact d'une estimation des paramètres par horizon sur les performances du modèle M3 dans le chapitre suivant. Les travaux de Gérard (2010) ont montré que l'utilisation du critère KGE pour caler le modèle GRP donne lieu à une dégradation des prévisions. Cette dégradation concerne principalement le critère de persistance, le rapport des variances et la variabilité du biais aux horizons 1 h et 24 h. Pour cette raison et contrairement aux simulations précédentes, les trois modèles ont été calés en utilisant le critère NSE global comme fonction objectif.

8.2.1 Mise à jour des réservoirs et correction des sorties

La mise à jour des conditions initiales du modèle GRP à partir du dernier débit observé s'effectue au niveau du réservoir de routage. Les travaux de Berthet (2010) ont montré qu'il n'est pas nécessaire de faire remonter la mise à jour à l'hydrogramme unitaire ni au réservoir de production. À chaque instant de prévision, le niveau du réservoir de routage est le niveau permettant de reproduire exactement le débit observé (fig. 8.3; voir section 1.6.3 pour les équations complètes du modèle et fig. 1.9 pour le schéma complet). Le niveau mis à jour R_t^* se calcule de la manière suivante :

$$R_t^* = \frac{\sqrt{Q_{obs,t}^2 + 4 \times ROU \times Q_{obs,t}} - Q_{obs,t}}{2} \quad (8.1)$$

avec $Q_{obs,t}$ le débit observé à l'instant de prévision et ROU [mm] le paramètre de capacité à un pas de temps du réservoir de routage.

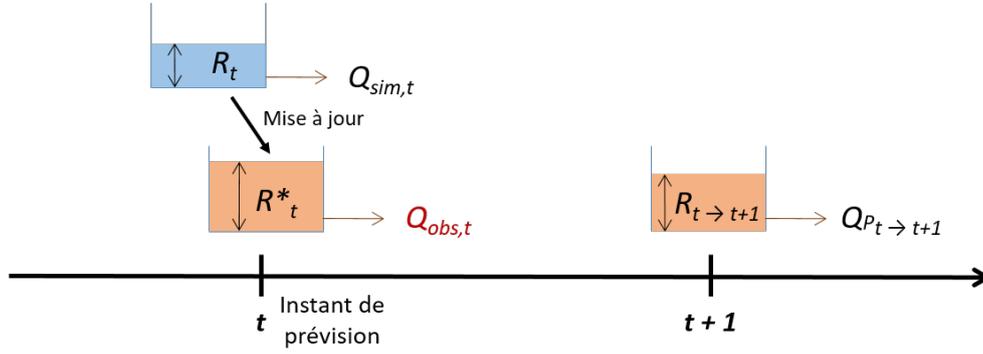


FIGURE 8.3 – Schéma représentant la mise à jour du réservoir de routage de GRP par insertion directe du dernier débit observé. Les notations sont reportées dans le tableau 8.1. Schéma modifié à partir de Viatgé & Perrin (2019).

Les modèles M0 et M3 ont deux branches de routage. Cela signifie que la mise à jour par insertion directe que nous allons proposer est nécessairement différente de la mise à jour de GRP. Lors de ses tests sur la structure du modèle GRP, Berthet (2010) évalue une version à deux branches de routage. Pour mettre à jour le réservoir de routage, il propose de trouver le niveau dont la vidange $Q^*_{R,t}$ est égale à $Q_{obs,t} - Q_{D,t}$, avec $Q_{D,t}$ le flux provenant de la branche pseudo-directe. Cependant, cette règle implique que lorsque $Q_{D,t} \geq Q_{obs,t}$, le réservoir de routage se vide presque totalement pour s'ajuster au débit observé. Pour limiter cet effet, nous proposons de déterminer une vidange cible $Q^*_{R,t}$ calculée comme suit :

$$Q^*_{R,t} = \frac{Q_{R,t}}{Q_{D,t} + Q_{R,t}} \times Q_{obs,t} = \alpha \times Q_{obs,t} \quad (8.2)$$

où α est la fraction du débit simulé qui provient de la branche indirecte à l'instant de prévision (avant mise à jour). Cette méthode implique que le modèle M0 ne pourra jamais totalement s'ajuster au débit à l'instant t , ce qui pourrait avoir un impact sur les performances à très courte échéance.

Comme le modèle M3 intègre un réservoir linéaire sur la branche pseudo-directe, on peut aussi calculer une vidange cible $Q^*_{D,t}$:

$$Q^*_{D,t} = \frac{Q_{D,t}}{Q_{D,t} + Q_{R,t}} \times Q_{obs,t} = (1 - \alpha) \times Q_{obs,t} = Q_{obs,t} - Q^*_{R,t} \quad (8.3)$$

L'hypothèse que nous faisons ici est que l'information portée par l'erreur du modèle à un instant t ne permet pas d'estimer un partage « observé » des écoulements. On préfère donc reprendre le partage initialement simulé par le modèle. Cette méthode de mise à jour est illustrée par la figure 8.4

La formulation de la vidange du réservoir de routage de GR5H (fonction instantanée en puissance 5 du niveau) n'est pas inversible analytiquement. Il n'y a donc pas de solution analytique donnant le niveau du réservoir dont la vidange est égale à $Q^*_{R,t}$. Pour trouver un niveau approché à chaque instant t , 1000 valeurs entre $X_3/1000$ et X_3 sont testées (voir Pelletier (2021) pour une procédure similaire appliquée au modèle GR6J). Le réservoir linéaire de la branche pseudo-directe de M3 est lui inversible. Le niveau du réservoir linéaire mis à jour L^*_t à l'instant t est calculé comme suit :

$$L^*_t = Q^*_{D,t} \times \frac{1 - K_L}{K_L} \quad (8.4)$$

où K_L [-] est le coefficient de vidange du réservoir linéaire.

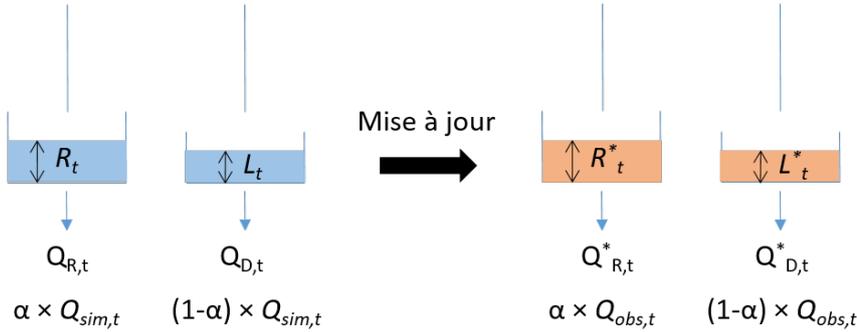


FIGURE 8.4 – Schéma représentant la mise à jour des réservoirs du modèle M3 par insertion directe du dernier débit observé à l’instant de prévision. Les notations sont reportées dans le tableau 8.1.

Contrairement au modèle GRP, les modèles M0 et M3 intègrent une fonction d’échange qui agit au niveau des fonctions de routage. Pour simplifier la méthode de mise à jour et pour de ne pas ajouter de degrés de liberté supplémentaires dans le calcul de R_t^* , les échanges sont calculés à partir du niveau du réservoir de routage estimé par le modèle de simulation. Lorsque la mise à jour s’applique aux deux réservoirs de routage, cela revient à ne pas avoir d’échange à l’instant de prévision. Lorsque la mise à jour s’applique seulement au réservoir de routage de la branche indirecte, cela revient à avoir un échange seulement sur la branche pseudo-directe à l’instant de prévision.

TABLE 8.1 – Liste des notations des figures 8.3 et 8.4.

Notation	Définition	Unité
t	Instant de prévision	h
$Q_{obs,t}$	Débit observé à t	mm/h
$Q_{sim,t}$	Débit simulé à t	mm/h
$Q_{P_t \rightarrow t+1}$	Débit prévu à t pour $t + 1$	mm/h
R_t	Niveau du réservoir de routage à t avant la mise à jour	mm
R_t^*	Niveau du réservoir de routage mis à jour à t	mm
$R_{t \rightarrow t+1}$	Niveau du réservoir de routage prévu à t pour $t + 1$	mm
L_t	Niveau du réservoir linéaire à t avant la mise à jour	mm
L_t^*	Niveau du réservoir linéaire mis à jour à t	mm
$Q_{R,t}$	Flux simulé qui vient de la branche indirecte à t avant la mise à jour	mm/h
$Q_{D,t}$	Flux simulé qui vient de la branche pseudo-directe à t avant la mise à jour	mm/h
$Q_{R,t}^*$	Vidange cible du réservoir de routage de la branche indirecte à t	mm/h
$Q_{D,t}^*$	Vidange cible du réservoir linéaire de la branche pseudo-directe à t	mm/h
α	Fraction du débit simulé qui vient de la branche indirecte à t avant la mise à jour	–

Dans la dernière partie des résultats de ce chapitre, les trois modèles sont couplés à une méthode de correction des sorties. La méthode employée ici est une méthode de type autorégressif développée par Tangara (2005). Elle est employée dans le modèle GRP en fonctionnement opérationnel. Une autre méthode développée par Berthet (2010) est également utilisée en fonctionnement opérationnel mais elle n'est pas utilisée dans nos travaux.

Le débit prévu $Q_{P_{t \rightarrow t+H}}$ à l'instant t pour un horizon H donné est corrigé à partir du calcul suivant :

$$\hat{Q}_{P_{t \rightarrow t+H}} = Q_{P_{t \rightarrow t+H}} \times \left(\frac{Q_{obs,t}}{Q_{P_{t-1 \rightarrow t}}} \right)^{0.45} \quad (8.5)$$

avec $\hat{Q}_{P_{t \rightarrow t+H}}$ le débit corrigé à l'instant t pour un horizon H donné et $Q_{P_{t-1 \rightarrow t}}$ le débit prévu à l'instant $t - 1$ pour l'instant t . Le principe de cette correction est qu'elle s'efface lorsque l'on s'éloigne de l'instant de prévision car l'erreur du modèle entre $t - 1$ et t est très faible devant l'erreur à $t + H$ pour H suffisamment grand.

8.2.2 Critères d'erreur

Tous les critères sont calculés à horizon fixe. En d'autres termes, toutes les prévisions réalisées successivement sur chaque pas de temps de la série de données pour un horizon donné sont mises à la suite pour créer une série de débits prévus à cet horizon (fig. 8.5). Cette série temporelle est ensuite comparée aux observations selon le critère choisi. Pour simplifier la présentation des résultats, les valeurs des critères d'erreur sont présentées en validation croisée sans distinguer les sous-périodes temporelles. La plupart des résultats sont présentés pour les horizons 3 h, 6 h, 12 h, 24 h et 48 h. Lorsque le critère utilisé le permet, les performances du modèle de simulation (i.e. sans mise à jour des états initiaux) sont présentées. On se référera aux modèles de simulation par l'abréviation OL (pour *open loop* en anglais) dans la suite du chapitre. Pour calculer les critères événementiels, les prévisions à horizon fixe retenues sont celles qui tombent dans la fenêtre de l'événement observé. Cela permet de comparer les performances aux différents horizons et en OL sur les mêmes pas de temps.

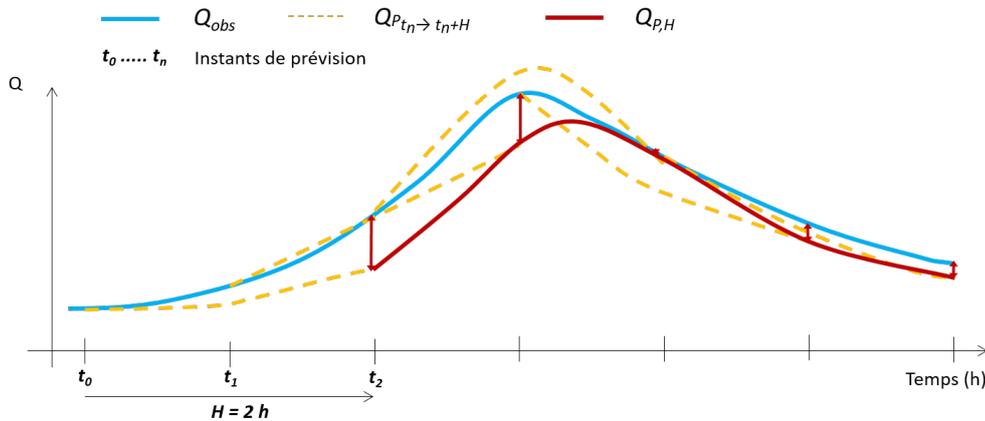


FIGURE 8.5 – Différence entre les prévisions à instant fixe et les prévisions à horizon fixe. Schéma modifié à partir de Viatgé & Perrin (2019). H est l'horizon de prévision. Q_{obs} est la chronique des débit observés. $Q_{P_{t_n \rightarrow t_n+H}}$ est le débit prévu à l'instant t_n pour l'instant $t_n + H$. $Q_{P,H}$ est le débit prévu à horizon fixe.

Il existe de nombreux critères pour évaluer la performance des prévisions hydrologiques déterministes. Comme dans les travaux précédents, nous distinguons deux familles de critères : les critères calculés sur tous les points de la chronique des débits (critères globaux) et les critères calculés sur les événements préalablement sélectionnés (critères événementiels). En complément, nous calculons des critères liés à la table de contingence (prévision de dépassement de seuil).

Le critère de persistance (CP ; Kitanidis & Bras, 1980) est souvent utilisé pour évaluer la qualité des prévisions des modèles hydrologiques. Le critère de persistance est un critère quadratique dont la formulation est proche du critère NSE mais dont le modèle « naïf » est la persistance des débits (le débit prévu est constant et égal au débit observé à l'instant de prévision). Berthet (2010) explique qu'en prévision (et en particulier à très courte échéance), la moyenne des débits observés (utilisée comme référence dans le critère NSE) est un modèle « naïf » très facile à battre du fait de la connaissance du débit observé à l'instant de la prévision. L'utilisation du critère CP pose toutefois des problèmes d'interprétation des performances en prévision. En effet, il n'est pas comparable entre horizons de prévision, la qualité du modèle naïf n'étant pas constante avec l'horizon de prévision. De plus, lorsqu'on compare plusieurs modèles, le modèle naïf a moins d'importance puisqu'il est le même pour tous les modèles. On cherche ainsi à savoir si les différences entre les modèles donnent lieu à une amélioration ou à une dégradation des performances.

Nous utilisons une version bornée (avec une formulation similaire au critère C2M présenté au chapitre 3) de ce critère (C2MP calculé sur toute la chronique des débits ; Berthet, 2010) car il évite de donner trop de poids aux valeurs fortement négatives du critère CP et a souvent été employé dans les travaux portant sur le modèle GRP. Nous présentons aussi les performances globales (i.e. les performances calculées sur toute la chronique) avec les critères KGE et NSE pour les raisons évoquées précédemment. Nous ajoutons l'erreur quadratique moyenne (RMSE) à nos analyses car elle est souvent utilisée en prévision des crues et son interprétation sur un bassin donné est simple, avec généralement une augmentation progressive avec l'horizon de prévision. Nous rapportons la RMSE au débit moyen du bassin considéré pour faciliter la comparaison des performances entre bassins.

Pour analyser les performances événementielles, nous utilisons le critère NSE (comme dans les chapitres précédents), la RMSE calculée sur les montées de crue (définies comme la fenêtre temporelle allant du début de l'événement observé jusqu'au débit maximal observé) et un critère d'erreur sur les pics de crue que nous avons adapté pour la prévision. Lorsqu'on trace des hydrogrammes prévus à horizon fixe, on peut observer des hydrogrammes qui semblent très bien reproduire les observations mais qui présentent un retard (fig. 8.6). Lorsque ce retard est supérieur à l'échéance de prévision, la conclusion que l'on tire de l'observation peut être biaisée car on aura tendance à penser que la prévision est de bonne qualité mais qu'elle est simplement en retard de quelques heures. On se retrouve en fait dans un cas où le pic prévu à échéance fixe fait partie de la récession des hydrogrammes prévus à instant de prévision fixe. Par exemple, pour un horizon de 3 h, si on se place à l'instant du pic observé t_p , la prévision qui a été émise à l'instant de prévision $t - 3$ pour l'instant t_p peut sous-estimer très fortement le pic observé. La prévision faite à l'instant de prévision t_p pour l'instant $t + 3$ peut elle reproduire l'amplitude du pic observé grâce à l'information sur l'erreur du modèle assimilée à l'instant de prévision t_p . Cela pose un problème pour calculer un critère d'erreur sur les pics lorsque la fenêtre temporelle pour sélectionner le pic prévu est la fenêtre

totale de l'événement considéré. Pour pallier ce défaut, nous proposons de réduire la fenêtre de recherche du pic prévu en fonction de l'horizon de prévision (H). Le critère que nous utilisons dans ce chapitre se base sur une fenêtre temporelle de recherche du pic prévu qui démarre à $t_p - H/2$ et qui se termine à $t_p + H/2$. $H/2$ est arrondi à l'unité supérieure pour les horizons impaires. Le critère $\Delta Q_{max}P$ est ensuite calculé selon l'équation 6.1.

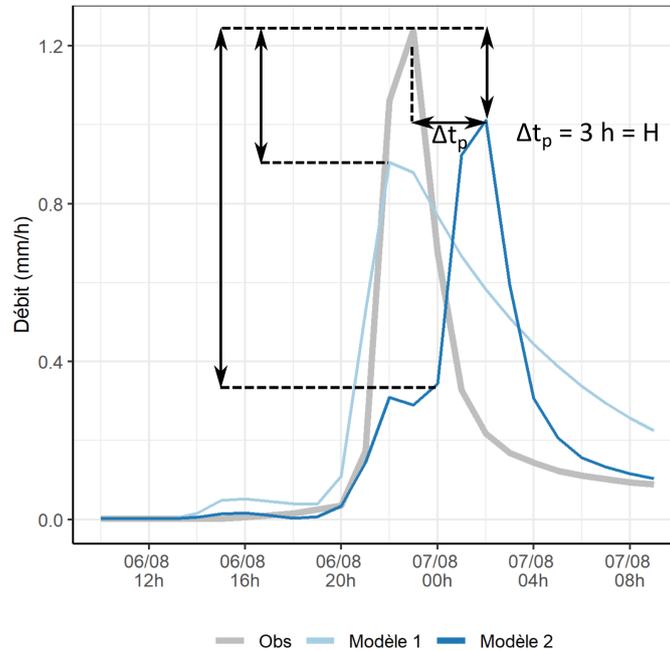


FIGURE 8.6 – Schéma représentant le décalage temporel et l'amplitude de l'erreur des pics de deux prévisions d'un événement de crue de l'Yzeron à Craponne en août 2007. Δt_p est le décalage temporel entre le pic observé et le débit maximal prévu par le modèle 2 à l'horizon 3 h.

Pour juger de la qualité d'une prévision déterministe, il est possible de définir un seuil d'alerte dépassé ou non par les prévisions et les observations à chaque instant t . La table de contingence dénombre les quatre cas possibles (tab. 8.2).

TABLE 8.2 – Table de contingence caractérisant le franchissement d'un seuil d'alerte.

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

Trois critères dérivant de la table de contingence sont calculés :

— La probabilité de détection (valeur idéale : 1) :

$$POD = \frac{AV}{AV + AM} \quad (8.6)$$

— Le taux de fausses alertes (valeur idéale : 0) :

$$FAR = \frac{FA}{AV + FA} \quad (8.7)$$

- L'indice de succès critique (valeur idéale : 1) :

$$CSI = \frac{AV}{AV + AM + FA} \quad (8.8)$$

Nous définissons le seuil de dépassement comme le quantile 95 % des débits.

Deux ratios sont calculés par bassin et sous-période de validation pour analyser la répartition spatiale des performances :

- Le ratio entre les valeurs de C2MP de M3 et les valeurs de C2MP de GRP.

$$R_{C2MP} = \frac{C2MP_{M3} + 1}{C2MP_{GRP} + 1} \in [0; +\infty[\quad (8.9)$$

- Le ratio entre les valeurs de NSE événementiel borné de M3 et les valeurs de NSE événementiel borné de GRP.

$$R_{NSEev} = \frac{\mu(NSEev_{M3}) + 1}{\mu(NSEev_{GRP}) + 1} \in [0; +\infty[\quad (8.10)$$

avec $\mu(NSEev_{M3})$ et $\mu(NSEev_{GRP})$ les moyennes par bassin et par sous-période des NSE événementiels bornés de M3 et GRP. Lorsque la valeur de ces ratios est supérieure à 1, le modèle M3 a de meilleures performances que le modèle GRP.

8.3 Résultats et analyses

8.3.1 Comparaison des performances de GRP, M0 et M3

La figure 8.7 présente les performances globales des trois modèles pour cinq horizons de prévision et l'OL.

- À l'horizon 3 h, les valeurs de C2MP de GRP sont plus élevées que celles de M3 et M0. À 6 h, les distributions des valeurs de C2MP de GRP et de M3 sont proches. La médiane et les quantiles 75 % et 95 % de C2MP de M3 sont plus élevés que ceux de GRP, c'est l'inverse pour les quantiles 5 % et 25 %. Les valeurs de C2MP de M3 deviennent plus élevées que celles de GRP à partir de l'horizon 12 h. M0 a les valeurs de C2MP les moins élevées des trois modèles pour les horizons 3 h, 6 h, 12 h et 24 h. M0 a des valeurs de C2MP plus élevées que GRP à 48 h.
- Les valeurs de NSE de GRP sont plus élevées que celles de M3 et M0 à l'horizon 3 h. Les valeurs de NSE de M3 deviennent plus élevées que celles de GRP et M0 à partir de l'horizon 6 h et pour l'OL. Les valeurs de NSE de M0 sont plus élevées que les valeurs de NSE de GRP à 48 h et pour l'OL. Pour les autres horizons, les valeurs des quantiles des distributions de NSE de GRP sont plus élevées.
- Les distributions des valeurs de KGE suivent les mêmes tendances. On observe cependant des valeurs de quantile 5 % nettement plus faibles pour M0 et M3 par rapport à GRP pour tous les horizons (OL non compris).
- Les distributions de RMSE de GRP et M3 sont proches à 3 h et à 6 h. Pour les autres horizons, les valeurs de RMSE de M3 sont plus faibles que les valeurs de RMSE de GRP. M0 a les valeurs les plus élevées de RMSE à 3 h, 6 h et 12 h. À 24 h et 48 h et pour l'OL, les valeurs de RMSE de M0 sont plus faibles que celles de GRP mais plus élevées que celles de M3.

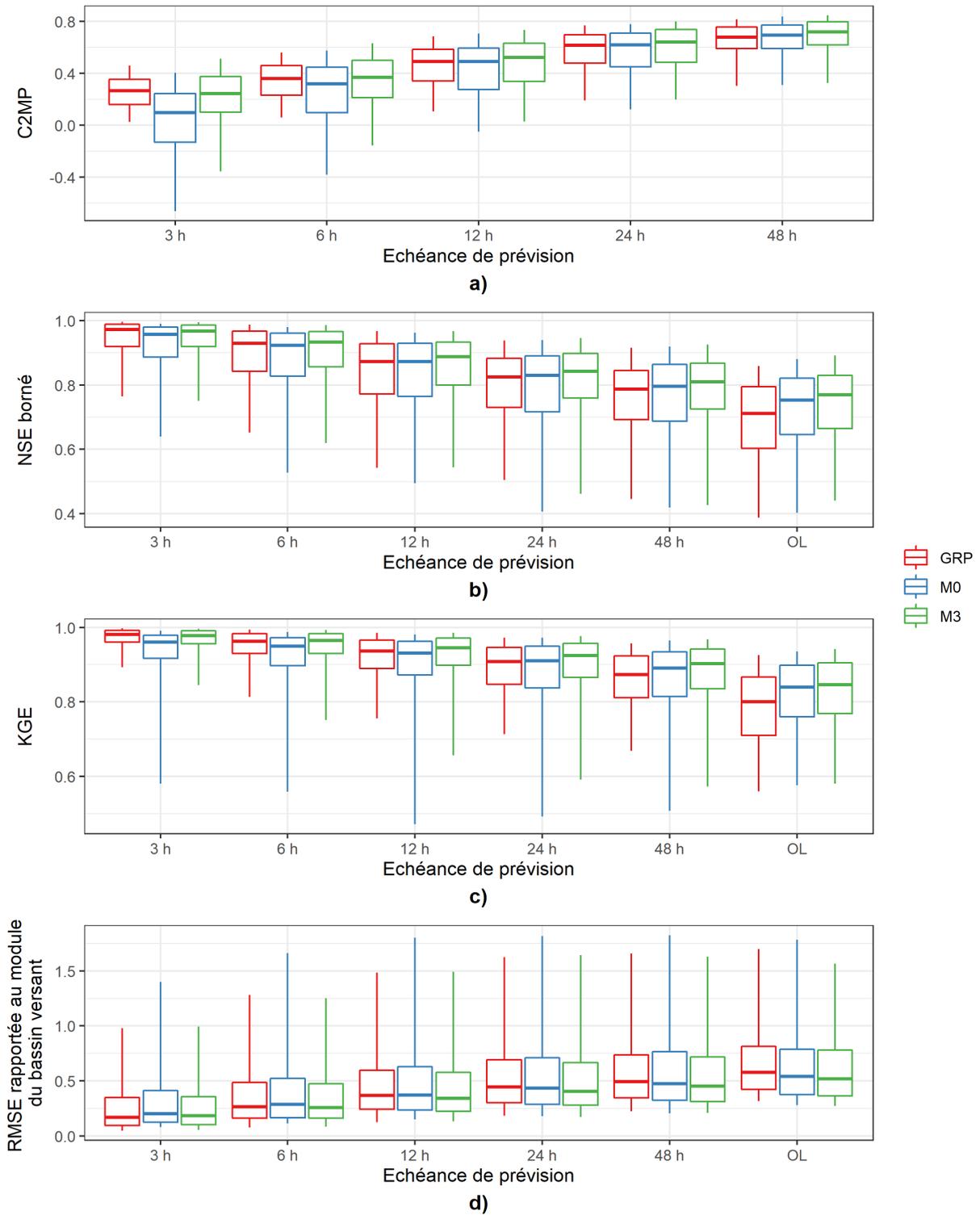


FIGURE 8.7 – Distribution des performances globales (validation croisée) des modèles GRP, M0 et M3 pour cinq échéances de prévision et *l'open loop* (OL). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Ces premiers résultats montrent que le modèle M3 a de meilleures performances globales que le modèle GRP sauf à très courte échéance, en particulier au regard du C2MP. Le modèle M0 présente les moins bonnes performances à part en OL et à 48 h. Les résultats indiquent aussi qu'il semble y avoir des bassins pour lesquels les modèles M0 et M3 dégradent fortement les performances par rapport à GRP.

La figure 8.8 présente la répartition spatiale du ratio des performances globales entre M3 et GRP selon le C2MP. À l'horizon 3 h, plusieurs bassins présentent une dégradation des valeurs de C2MP. La majorité de ces bassins est située dans la moitié nord de la France. Quelques bassins méditerranéens présentent des valeurs dégradées du C2MP. On observe une amélioration des performances pour plusieurs affluents du Rhône et en amont de la Loire. On observe les mêmes tendances à 6 h avec cependant moins de valeurs dégradées du C2MP. Aux horizons 12 h et 24 h, la plupart des bassins présentent une amélioration des performances ou très peu de différences entre GRP et M3. Les valeurs de C2MP sont fortement dégradées pour un affluent du Rhône, le Rival à Beaufort, pour les quatre horizons. Ce bassin a un rendement Q/P très faible (0.05) même au regard de son indice d'humidité P/ETP (1.71). Il est situé sur le plateau karstique du Vercors ainsi que sur le massif karstique du Devoluy². Si des pertes et résurgences karstiques influencent le comportement hydrologique de ce bassin, il est possible que la fonction de correction de la pluie efficace de GRP permette un meilleur bouclage du bilan en eau que la fonction d'échange de GR5H.

La figure 8.9 présente les RMSE de GRP, M0 et M3 rapportées aux RMSE de leur version OL respective. Une valeur du ratio de RMSE égale à 1 indique que les performances de l'OL sont égales aux performances des prévisions. Lorsque le ratio est inférieur à 1, cela signifie que le modèle a une RMSE inférieure à celle de sa version OL, et donc qu'il est meilleur. Le but de ce graphique est de comparer l'impact des techniques de mise à jour sur les performances des modèles de prévision. On observe la même tendance pour les cinq horizons de prévision : les valeurs du ratio des RMSE de M0 sont plus élevées que celles de M3 et GRP ; les valeurs de ce ratio pour GRP sont les plus faibles. Ces résultats indiquent que la procédure de mise à jour des états initiaux de GRP permet des gains de performance plus importants par rapport à la version OL que les procédures de M0 et M3. La procédure de mise à jour de M3 semble plus performante que celle de M0, en particulier pour l'horizon 3 h. Cela est peut-être dû à la capacité de M3 de simuler le débit observé à l'instant de prévision (après mise à jour) grâce au réservoir linéaire de la branche de routage pseudo-directe.

Nous nous intéressons maintenant à la performance des prévisions selon les 10 652 événements de crue sélectionnés sur notre base de données. La figure 8.10 présente la distribution des RMSE sur les montées de crue pour les trois modèles. Les distributions des valeurs de RMSE sont plus faibles pour le modèle M3 que pour les modèles GRP et M0 pour les événements d'hiver et d'été, cela pour tous les horizons de prévision sauf en hiver à l'horizon 3 h où la distribution de M3 est similaire à celle de GRP. Les différences les plus marquées entre GRP et M3 sont observées en été. M0 a des distributions de RMSE similaires à GRP en hiver à part à 3 h où les valeurs de RMSE de M0 sont plus élevées. En été, les distributions de RMSE de M0 sont plus faibles que celles de GRP à part à l'horizon 3 h où les distributions de RMSE sont similaires.

2. www.isere.fr/Documents/environnement/ODE/Contexte_hydrographique.pdf

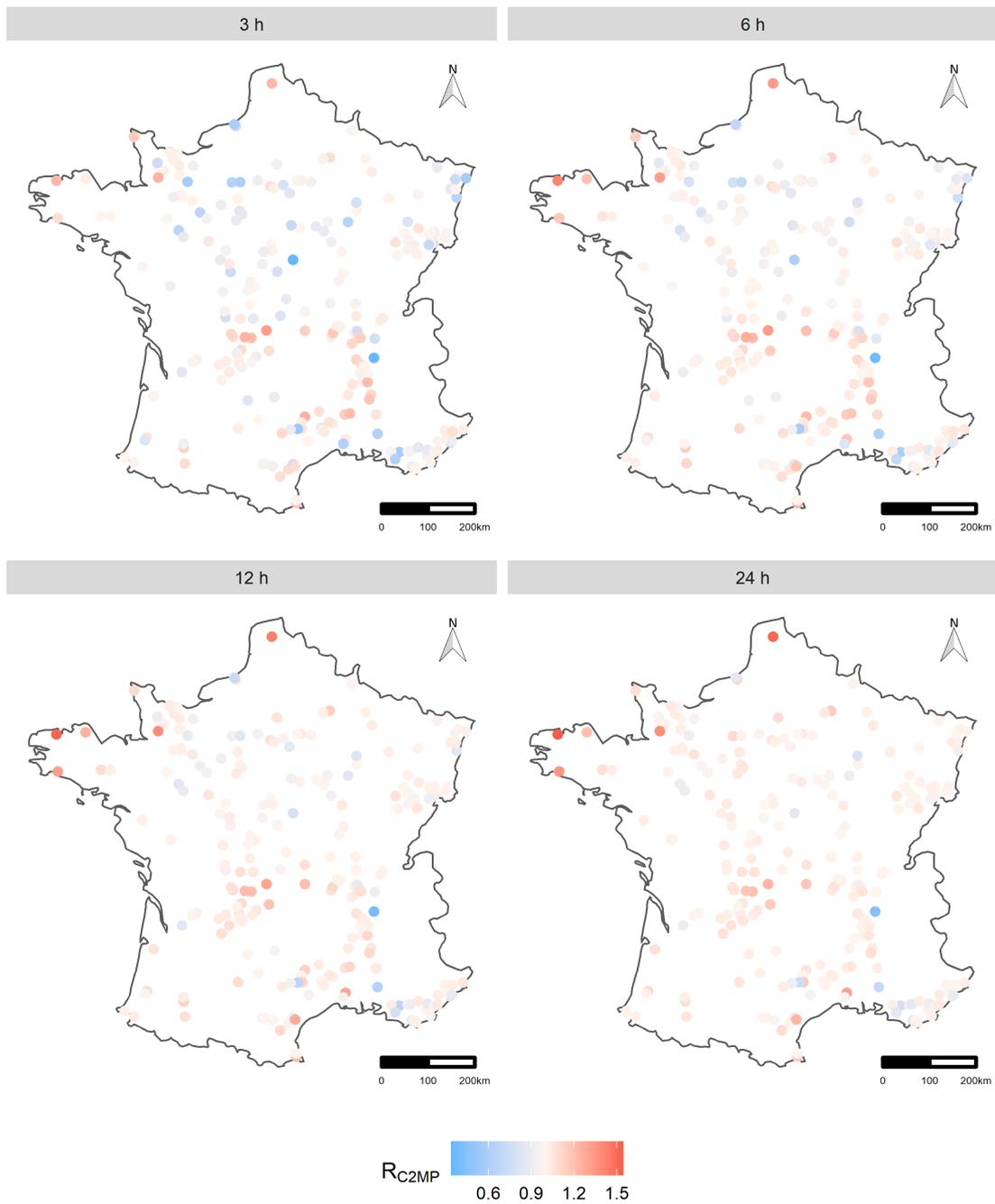


FIGURE 8.8 – Répartition spatiale du ratio des performances M3/GRP en prévision selon le critère de persistance borné (C2MP). Les performances sont présentées pour quatre échéances de prévision en moyennant le ratio sur les deux sous-périodes de validation. $R_{C2MP} > 1$ signifie que M3 a des performances plus élevées que GRP.

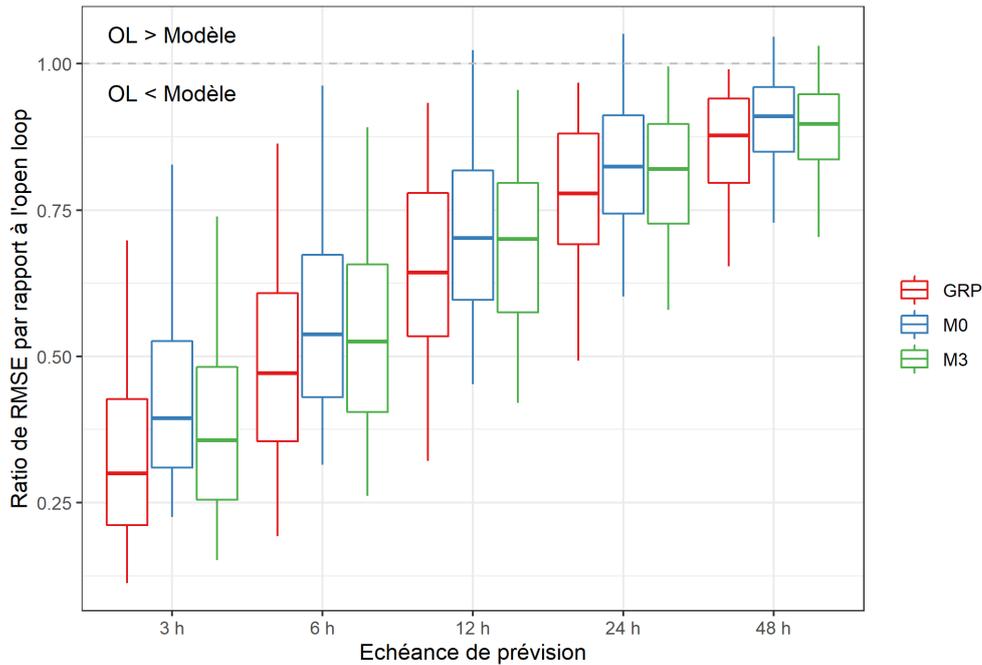


FIGURE 8.9 – Distribution du ratio entre les performances globales (RMSE; validation croisée) des modèles GRP, M0 et M3 pour cinq échéances de prévision et les performances de leur *open loop* (OL) respective. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

La figure 8.11 montre que le modèle M3 a de meilleures performances sur les pics de crue que GRP et M0 en été et en hiver pour tous les horizons de prévision excepté l'horizon 3 h pour lequel les performances sont proches. Les différences de performances sont plus marquées en été. On observe que le modèle M0 a de meilleures performances sur les pics de crue que GRP à part à l'horizon 3 h en hiver et aux horizons 3 h et 6 h en été où les performances sont dégradées par rapport à GRP.

On observe sur la figure 8.12 que le modèle M3 présente des valeurs plus élevées de NSE que les modèles GRP et M0 pour les événements d'été et les événements d'hiver, cela pour tous les horizons de prévision. Le modèle M0 a des performances plus élevées que le modèle GRP pour les événements d'hiver et les événements d'été pour tous les horizons excepté à 3 h où GRP a de meilleures performances.

Cette vue d'ensemble des performances événementielles montre que le modèle M3 a les meilleures performances des trois modèles, en particulier à partir de l'horizon 6 h. À 3 h les performances sont soit équivalentes à celles de GRP soit plus élevées (montées de crue en été). Le modèle M0 a des performances plus élevées que le modèle GRP à partir de l'horizon 12 h et en particulier sur les pics de crue. Nous obtenons des conclusions similaires pour d'autres critères de performance (e.g. biais) calculés sur les événements de crue (non montré ici par souci de concision).

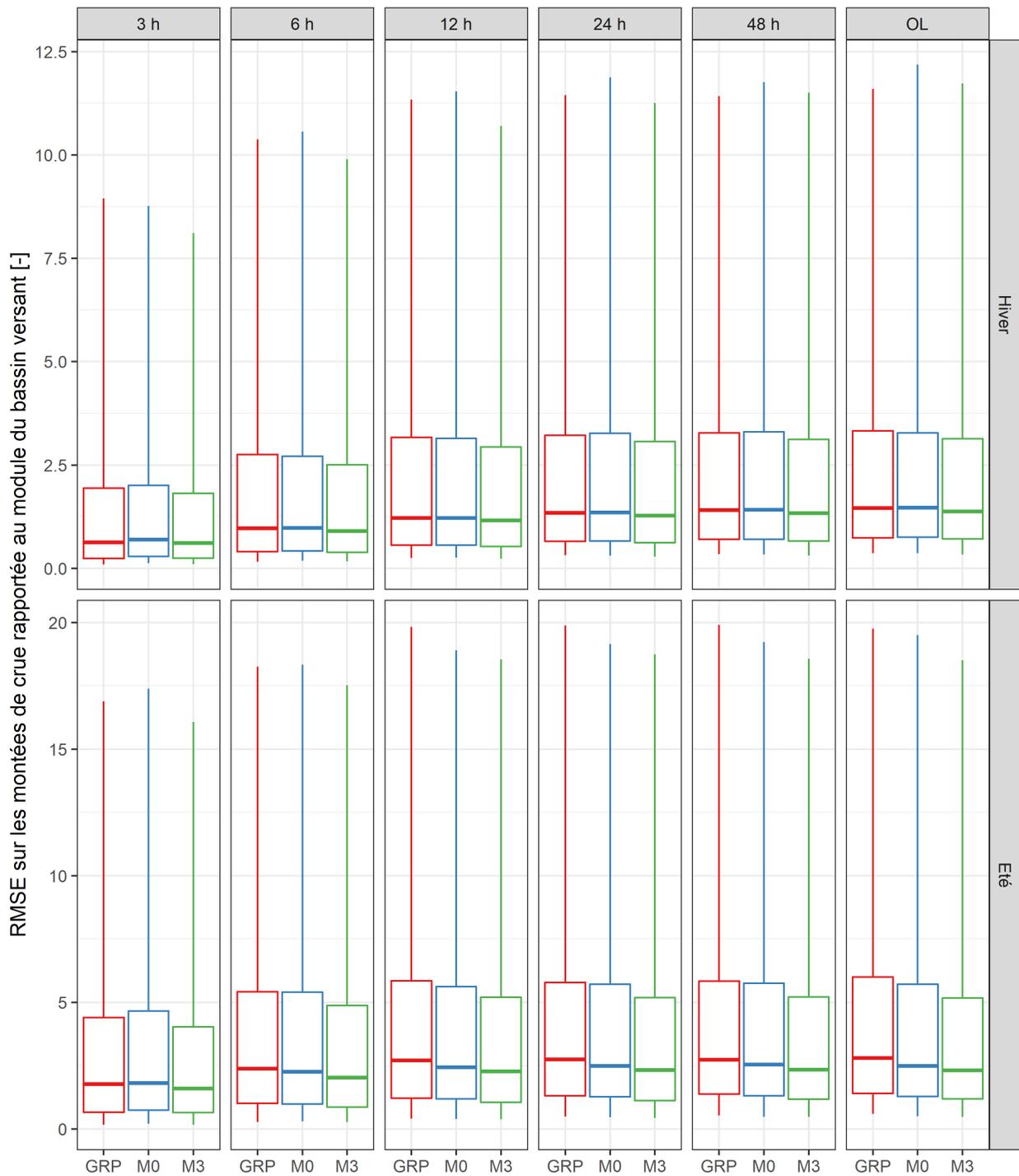


FIGURE 8.10 – Distribution des performances (RMSE; validation croisée; une valeur de critère par événement) des modèles GRP, M0 et M3 sur les montées de crue pour cinq échéances de prévision et *l'open loop* (OL). 8 290 événements d'hiver et 2 362 événements d'été sont considérés ici. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. L'échelle des valeurs du critère pour les événements d'hiver est différente de l'échelle des valeurs du critère pour les événements d'été, cela afin de faciliter la visualisation à très courte échéance.

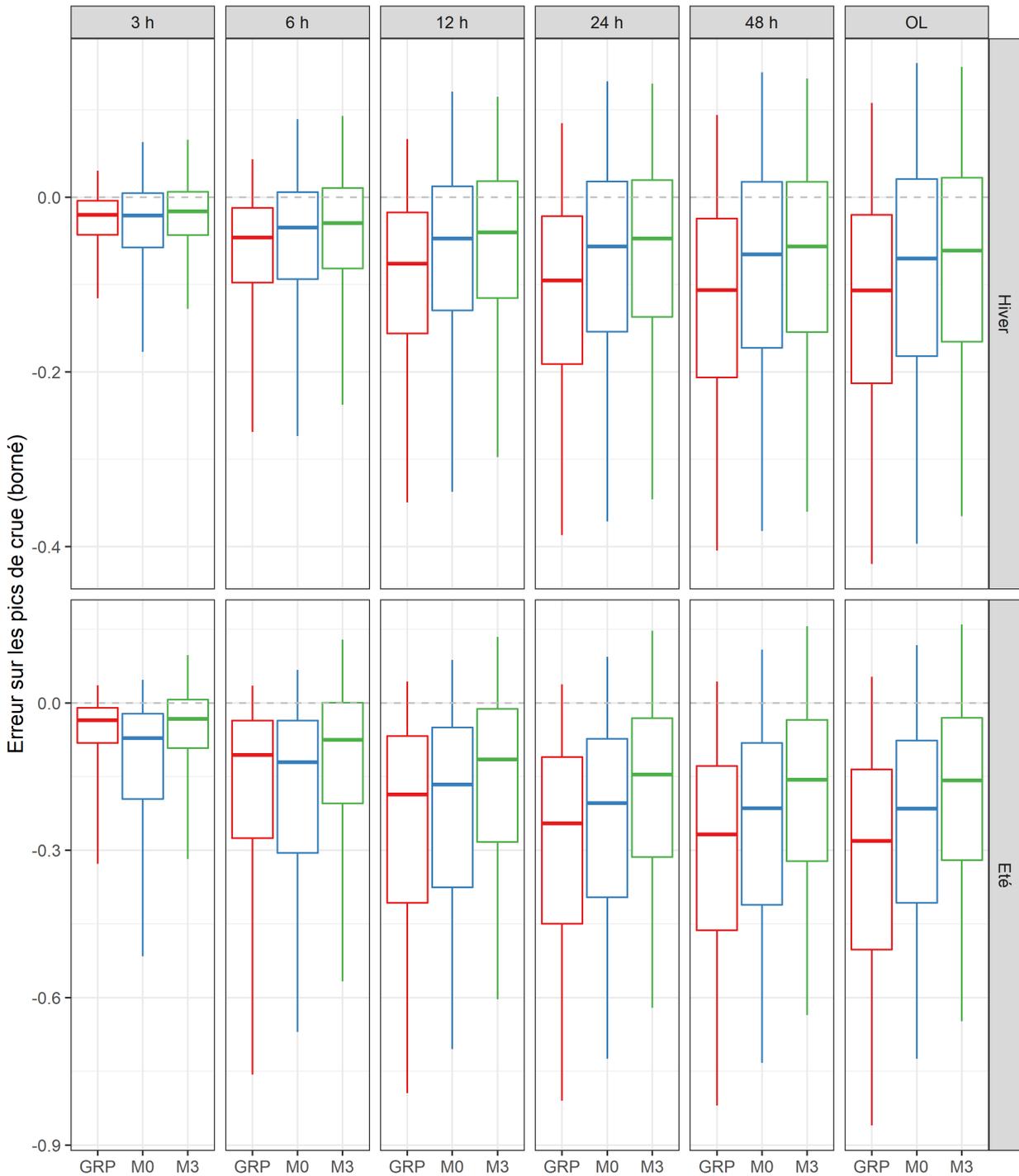


FIGURE 8.11 – Distribution des performances (validation croisée; une valeur de critère par événement) des modèles GRP, M0 et M3 sur les pics de crue pour cinq échéances de prévision et l'open loop (OL). 8 290 événements d'hiver et 2 362 événements d'été sont considérés ici. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. L'échelle des valeurs du critère pour les événements d'hiver est différente de l'échelle des valeurs du critère pour les événements d'été, cela afin de faciliter la visualisation à très courte échéance.

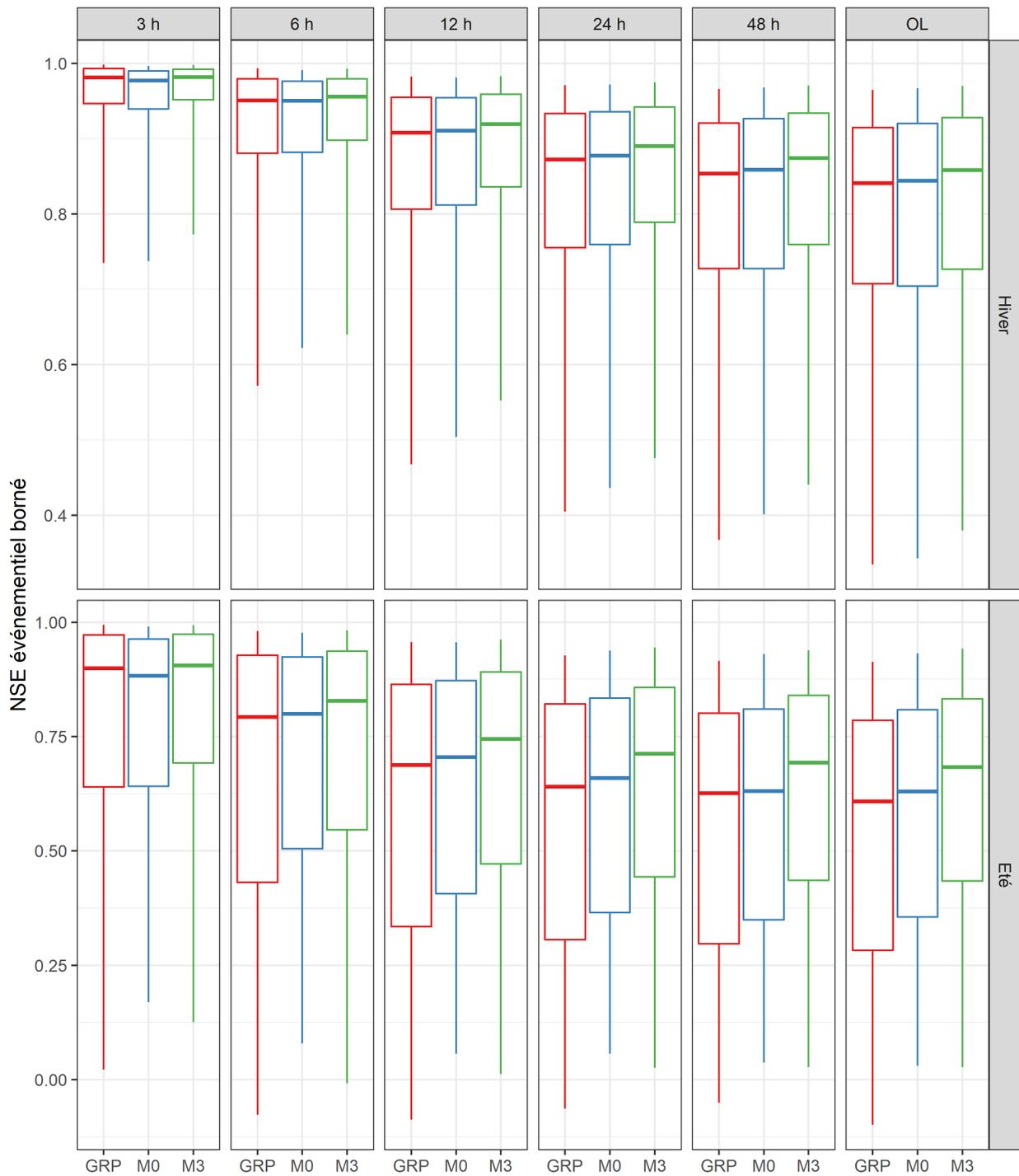


FIGURE 8.12 – Distribution des performances événementielles (NSE; validation croisée; une valeur de critère par événement) des modèles GRP, M0 et M3 pour cinq échéances de prévision et *l'open loop* (OL). 8290 événements d'hiver et 2362 événements d'été sont considérés ici. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %. L'échelle des valeurs du critère pour les événements d'hiver est différente de l'échelle des valeurs du critère pour les événements d'été, cela afin de faciliter la visualisation à très courte échéance.

La figure 8.13 présente la répartition spatiale des performances événementielles comparées entre M3 et GRP.

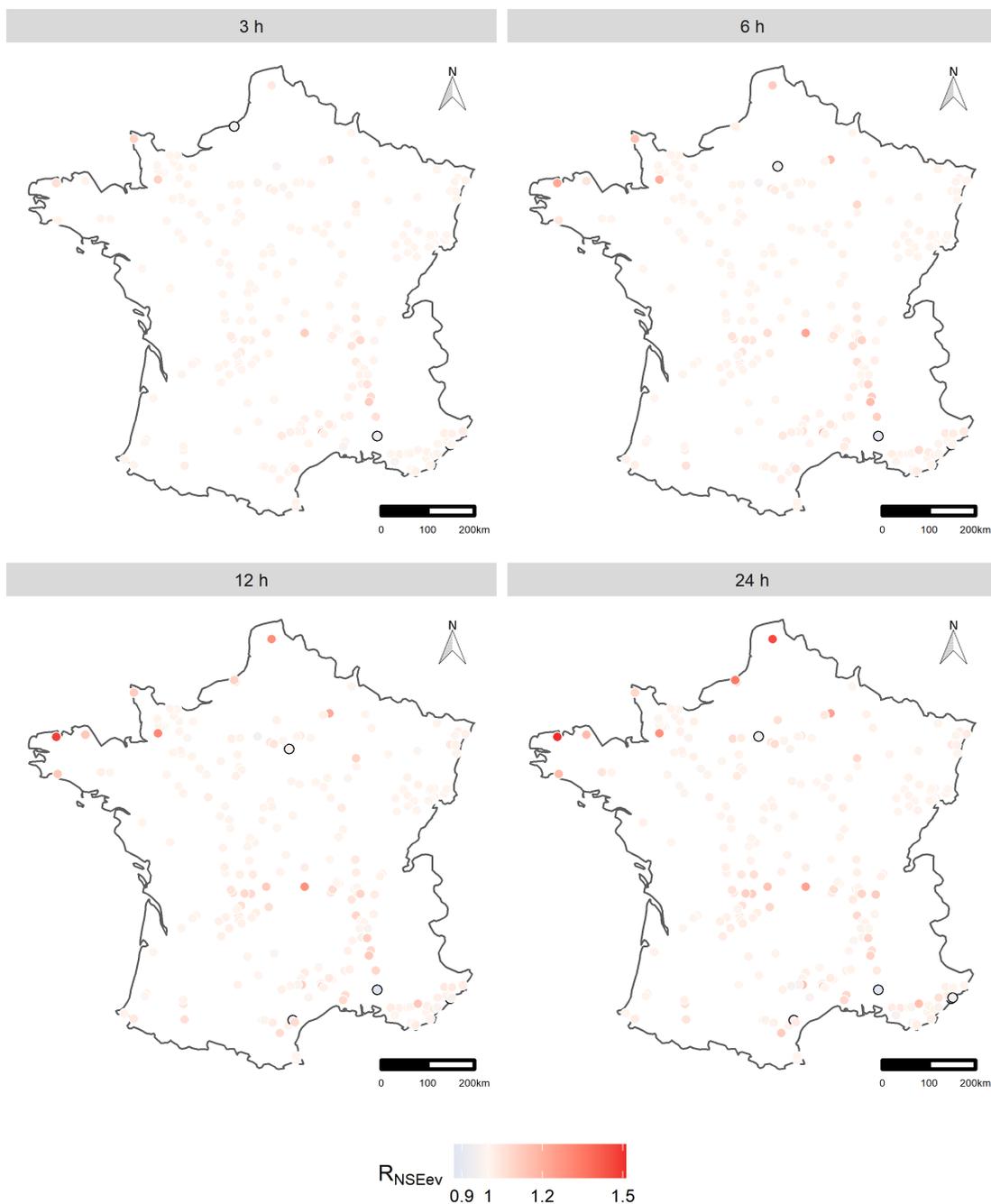


FIGURE 8.13 – Répartition spatiale du ratio des performances événementielles M3/GRP en prévision selon le critère NSE. Les performances sont présentées pour quatre échéances de prévision en moyennant le ratio par bassin et sur les deux sous-périodes de validation. $R_{NSE_{ev}} > 1$ signifie que M3 a des performances plus élevées que GRP.

On observe que pour tous les horizons, la majorité des bassins présente une amélioration des performances événementielles (points rouges ; valeurs du ratio supérieures à 1) ou des performances similaires entre M3 et GRP (ratio égal à 1). À 3 h, les améliorations sont majoritairement localisées sur des affluents du Rhône. Pour les trois autres horizons, les

améliorations sont localisées sur des affluents du Rhône (dont les Cévennes), sur des bassins du sud-est méditerranéen et sur des affluents de la Loire sur la partie amont la plus au sud. Plusieurs bassins bretons et du nord de la France présentent des améliorations importantes aux horizons 12 h et 24 h. Une explication de ces améliorations de performance peut être le fait que M3 a un réservoir de production dont la capacité est calée, ce qui n'est pas le cas du modèle GRP. Le modèle M3 a des performances nettement inférieures à GRP sur un affluent très en aval du Rhône, le Coulon à Oppède. Ce bassin a un rendement Q/P très faible par rapport aux autres bassins de notre base de données (égal à 0.03). Il s'agit d'un bassin dont le fonctionnement est lié à un système karstique³. La fonction de correction de la pluie efficace de GRP agit peut-être comme un coefficient de correction surfacique (de la même manière qu'un des paramètres du modèle Gardénia; voir Le Mesnil, 2021, pour une modélisation conceptuelle de bassins karstiques avec les modèles GR5H et Gardénia), permettant ainsi à GRP de mieux prévoir les crues sur ce bassin.

La figure 8.14 présente les distributions des critères liés à la table de contingence. M3 présente les valeurs les plus élevées du CSI à tous les horizons sauf à l'horizon 3 h pour lequel les distributions de CSI de GRP et M3 sont similaires. Les valeurs de CSI de GRP sont plus élevées que celles de M0 à part aux horizons 24 h et 48 h pour lesquels M0 a des valeurs plus élevées de ce critère. Le modèle M3 a des valeurs de POD similaires aux valeurs de GRP sauf à 12 h où M3 a de meilleures performances. M0 a des valeurs de POD plus faibles que les valeurs de GRP pour tous les horizons. GRP a des valeurs plus élevées de FAR pour tous les horizons sauf à 3 h où la distribution de FAR de GRP est similaire à celle de M3. Les valeurs de FAR de M3 sont plus faibles que les valeurs de FAR de M0 pour tous les horizons. En résumé, le modèle M3 a de meilleures performances pour ces critères d'alerte pour tous les horizons de prévision, sauf à l'horizon 3 h pour lequel les performances de GRP et M3 sont similaires. Les différences les plus notables sont pour le taux de fausses alertes. Le modèle M0 a les performances les plus faibles pour ces critères.

Les résultats précédents ont montré que le modèle M3 a de meilleures performances que les modèles GRP et M0 pour presque toutes les échéances. Les différences de performance les plus notables concernent les événements d'été. Les pics de crue en été sont notamment largement mieux simulés par M3. De nombreux bassins pour lesquels le modèle M3 a de meilleures performances que le modèle GRP sont des affluents du Rhône. Les améliorations pour les bassins méditerranéens semblent moins marquées que les améliorations obtenues pour M3 par rapport à M0 en simulation. Les résultats montrent qu'il semble difficile d'obtenir de meilleures performances que celles du modèle GRP à très courte échéance. Plusieurs raisons peuvent expliquer ces observations. Les prévisions à 3 h sont beaucoup plus dépendantes de l'erreur à l'instant de prévision qu'à plus longue échéance. La structure du modèle a donc moins d'importance à courte échéance, contrairement aux procédures de mise à jour. Les résultats de la figure 8.9 appuient ce raisonnement. La procédure de mise à jour de GRP semble plus efficace que les techniques que nous avons développées pour les modèles M0 et M3. Par ailleurs, les erreurs sont probablement plus autocorrélées à courte échéance qu'à plus longue échéance pour la majorité des événements de crue. Il est toutefois possible que ce ne soit pas le cas de certains événements qui ont une dynamique de montée très rapide. Nous allons explorer cette hypothèse dans les sections qui suivent.

3. www.sircc.fr/le-bassin-versant

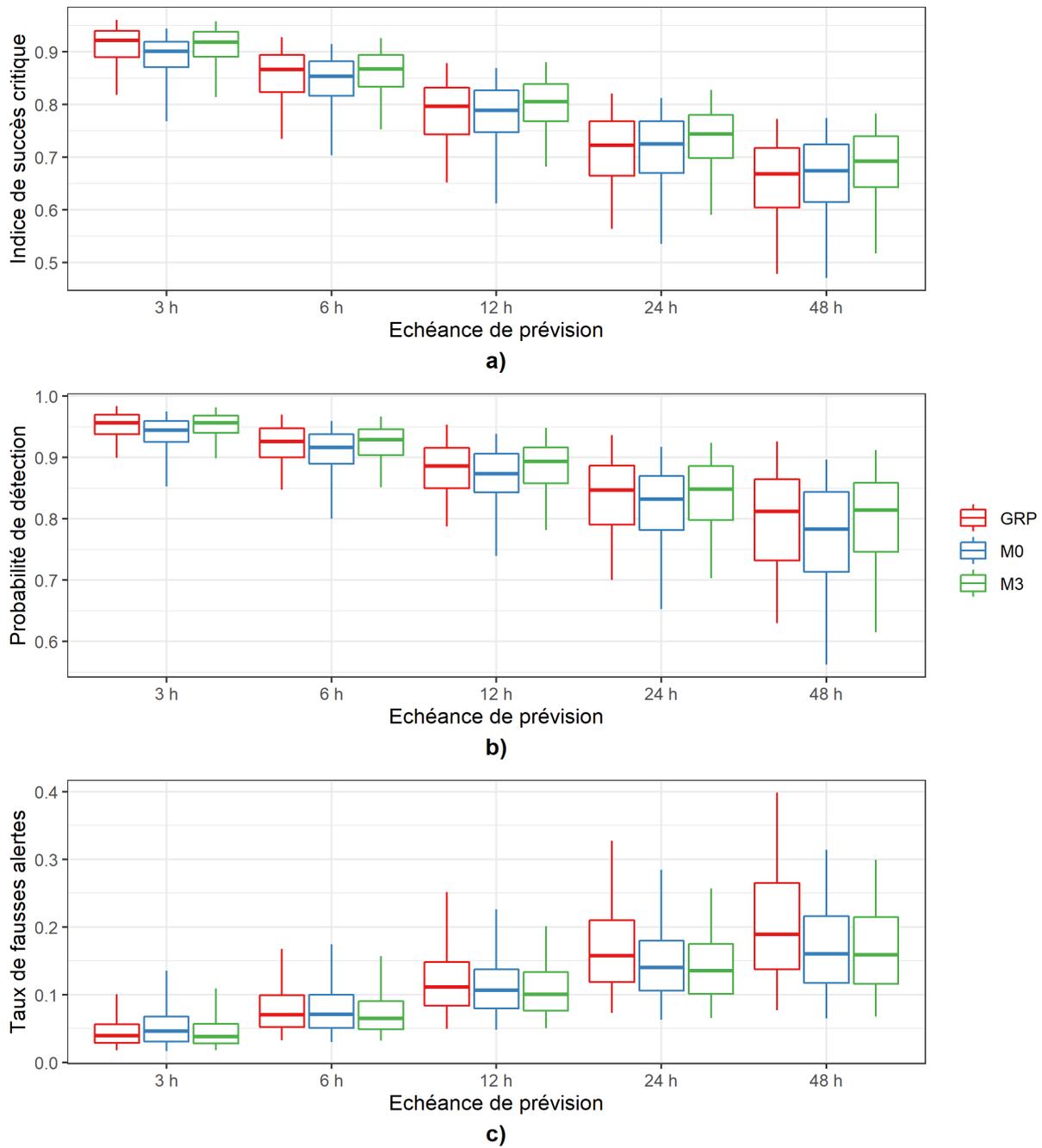


FIGURE 8.14 – Distribution des critères de dépassement de seuil (a : CSI ; b : POD ; c : FAR) pour les modèles GRP, M0 et M3 pour cinq échéances de prévision (validation croisée). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

8.3.2 Performances en fonction des caractéristiques des bassins et des événements

En comparant les performances sur les montées de crue aux temps de montée (fig. 8.15), on observe que plus le temps de montée est court, meilleures sont les performances de M3 par rapport à GRP, cela pour tous les horizons de prévision. Ces différences augmentent plus l'horizon est petit. Entre 65 % et 70 % des événements dont le temps de montée est inférieur à 12 h sont mieux prévus par M3 que par GRP aux horizons 3 h et 6 h. Le quantile 50 % du ratio des performances est égal à 1 voire supérieur à 1 pour les événements dont le temps de montée est supérieur à 24 h pour les quatre horizons de prévision. Ces résultats indiquent que la dynamique des crues a un fort impact sur les différences de performance entre M3 et GRP, même à très courte échéance. L'erreur à l'instant de prévision contient moins d'information lorsque de fortes variations du débit ont lieu en très peu de temps. Dans ce cas, seule une amélioration de la structure du modèle permet d'améliorer les prévisions.

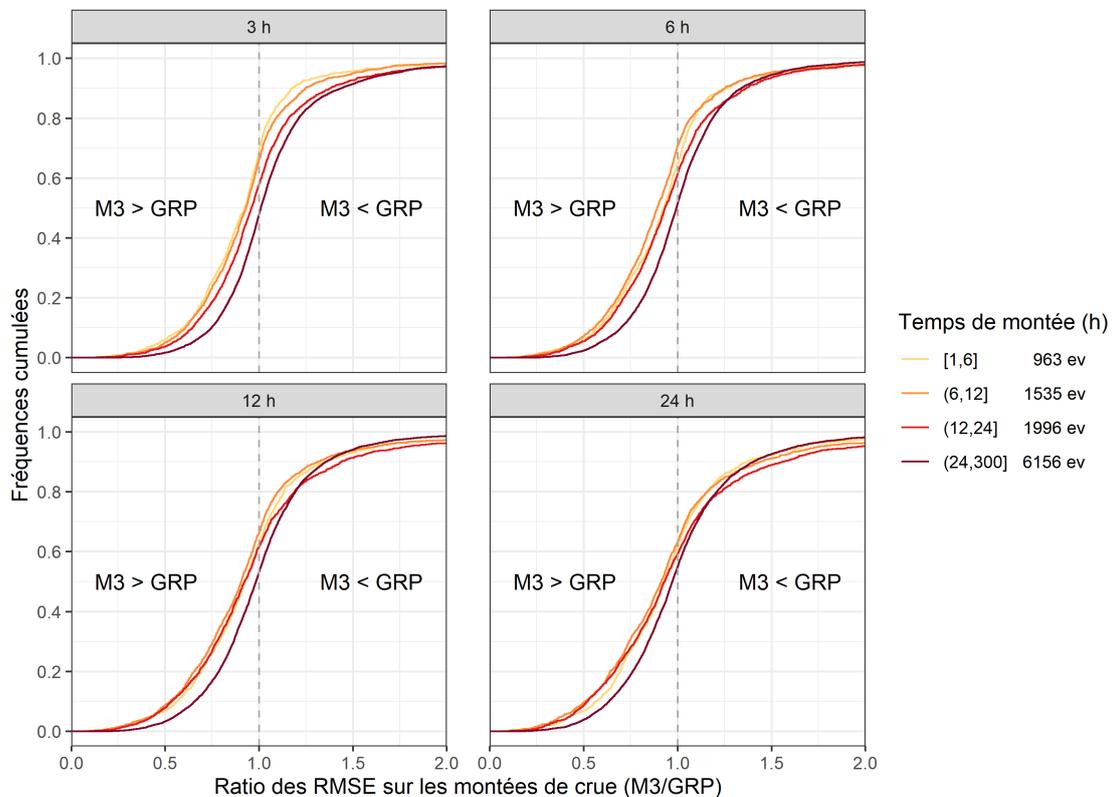


FIGURE 8.15 – Distribution du ratio des RMSE sur les montées de crue entre M3 et GRP en fonction du temps de montée des événements. Les résultats sont présentés en validation croisée pour cinq échéances de prévision. Le découpage des temps de montée a été choisi en fonction des échéances de prévision. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Les résultats présentés sur la figure 8.16 confortent cette hypothèse. On observe que les événements de crue des bassins dont le temps de concentration est inférieur à 21 h sont mieux prévus par M3 que par GRP, cela pour les quatre horizons de prévision. Pour les bassins dont le temps de concentration est supérieur à 21 h, GRP a de meilleures performances à 3 h et à 6 h. Cela signifie que lorsque le bassin a un comportement lent, GRP est davantage

capable d'exploiter l'erreur à l'instant de prévision que M3 à courte échéance, même si la structure de M3 simule aussi bien ces événements que la structure de GRP. À nouveau, les erreurs à courte échéance sont plus autocorrélées lorsque la dynamique du bassin est lente. La structure et l'assimilation de données plus efficace du modèle GRP permettent de mieux exploiter ces caractéristiques des erreurs.

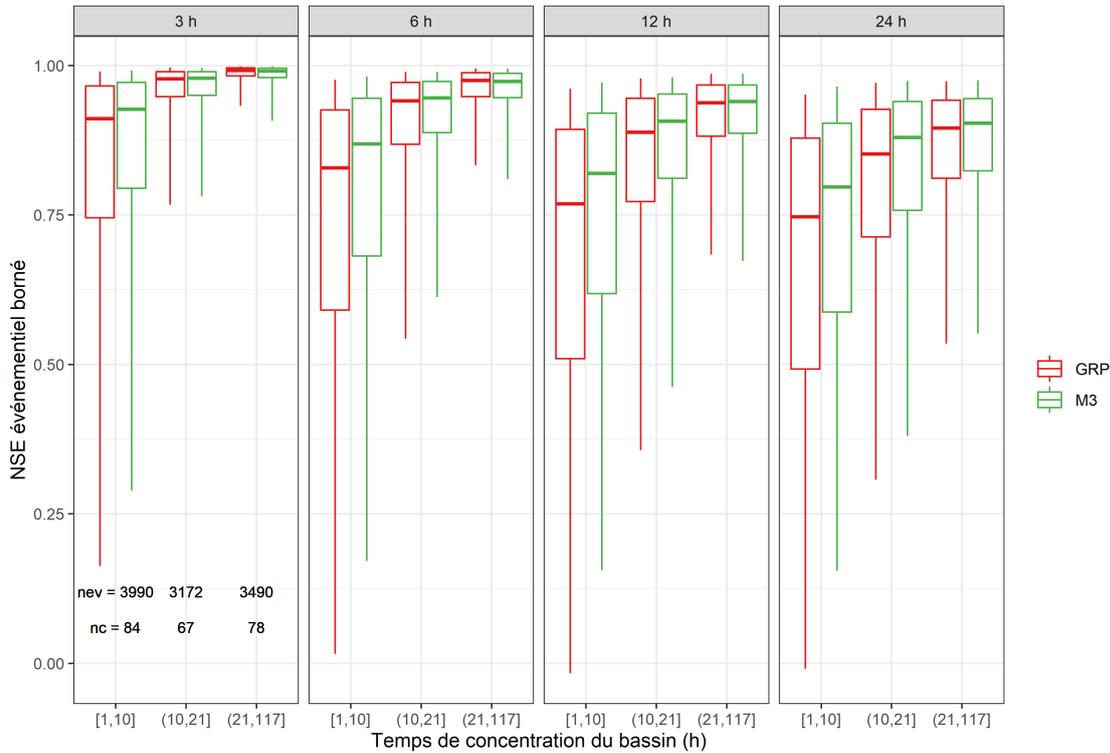


FIGURE 8.16 – Distribution des performances événementielles (NSE) des modèles GRP et M3 en fonction du temps de réaction des bassins versants. Les résultats sont présentés en validation croisée pour quatre échéances de prévision. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

Nous avons utilisé une classification de bassins selon plusieurs descripteurs dans le chapitre 5. La figure 8.17 présente les performances sur les montées de crue selon cette classification. On observe plusieurs améliorations des performances par le modèle M3 pour les bassins des groupes 1 et 3 pour les quatre horizons. Une majorité des bassins des groupes 4 et 5 ont de meilleures performances sur les montées de crue avec le modèle M3. Pour ces bassins, on observe une variabilité importante des performances entre horizons de prévision. GRP a de meilleures performances pour plusieurs bassins du groupe 2. Ce groupe contient les bassins qui ont les dynamiques les plus lentes de notre base de données.

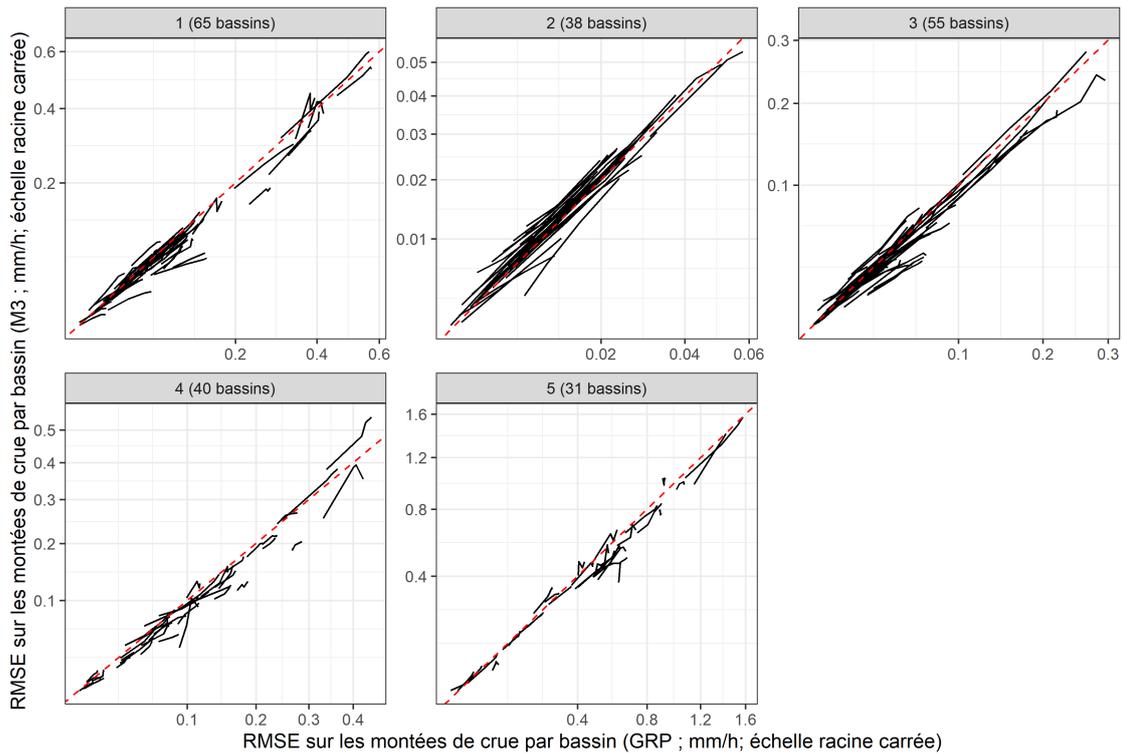


FIGURE 8.17 – Comparaison des performances événementielles moyennées par bassin et par échéance de prévision (RMSE sur les montées de crue ; validation croisée) pour cinq groupes de bassins versants. Chaque « fil » relie les performances d'un bassin pour quatre échéances de prévision (3 h, 6 h, 12 h et 24 h). Les échelles de RMSE sont différentes entre les groupes de bassins afin de faciliter la visualisation.

Le modèle M3 a été développé pour mieux prendre en compte l'intensité des pluies par rapport à M0, en particulier en condition de faible humidité antécédente. Par ailleurs, le modèle GRP est connu pour sous-estimer fréquemment les pics de crue. Nous regardons maintenant l'impact du changement de structure sur la capacité des prévisions de M3 et GRP à simuler les pics de crue lorsque les intensités de pluie sont élevées (fig. 8.18) et lorsque les conditions antécédentes d'humidité des sols sont faibles (fig. 8.19). L'erreur des deux modèles augmente avec l'accroissement de l'intensité de pluie moyenne, cela pour les quatre horizons de prévision. À 3 h, on observe très peu de différences entre les deux modèles. À 6 h, 12 h et 24 h, l'erreur sur les pics de crue est moins élevée pour M3 que pour GRP, en particulier pour les intensités de pluie les plus élevées. Les sous-estimations des pics de crue sont plus fortes pour les deux modèles en condition de faible humidité antécédente à part à l'horizon 3 h. M3 sous-estime moins fortement les pics de crue que GRP aux horizons 6 h, 12 h et 24 h. Ces différences sont plus marquées lorsque l'humidité antécédente est faible.

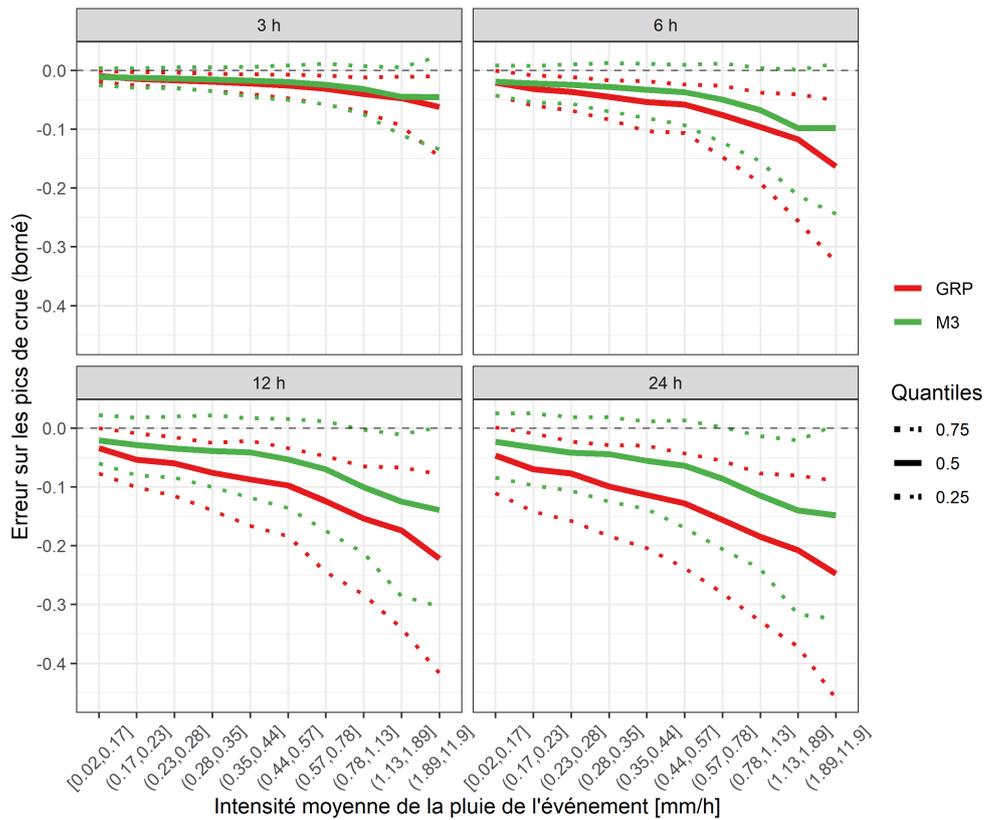


FIGURE 8.18 – Distribution des performances de GRP et M3 sur les pointes de crue en fonction de l'intensité de la pluie moyenne des événements (validation croisée). Les résultats sont présentés pour quatre échéances de prévision. Les événements ont été répartis en dix classes de quantile d'intensité moyenne de la pluie.

La figure 8.20 présente quatre hydrogrammes prévus par les modèles GRP et M3 aux horizons 3 h, 12 h et 24 h sur quatre bassins versants de notre base de données. Les performances événementielles de GRP et M3 sur ces hydrogrammes sont présentées dans le tableau 8.3. À nouveau, les hydrogrammes présentés ici ne sont que des illustrations des analyses précédentes et ne constituent pas un échantillon représentatif des résultats.

L'Ardèche à Ucel (478 km²) est un affluent en aval du Rhône en rive droite. De nombreux événements de crue ont lieu au début de l'automne sur ce bassin. L'événement qui a eu lieu en novembre 2014 est caractérisé par des intensités de pluie allant jusqu'à 20 mm/h. Le modèle M3 prévoit mieux le pic de crue que GRP pour les trois horizons de prévision. À 3 h, le pic prévu par GRP à échéance fixe est décalé de 3 h, ce qui signifie qu'il a été prévu à l'instant du pic observé. Bien que ce pic soit du même ordre de grandeur que le pic prévu par M3, la prévision est de moins bonne qualité au regard de l'horizon considéré.

L'Ill à Osthuse (3 296 km²) est un affluent du Rhin en rive gauche. Il est caractérisé par une dynamique lente et des périodes de hautes eaux qui ont majoritairement lieu en hiver. La crue de décembre 2010 est un exemple caractéristique de ce bassin. L'événement de pluie associé à cet événement est réparti sur cinq jours. Les prévisions des deux modèles sont satisfaisantes pour les trois horizons mais le pic de crue est sous-estimé à 12 h et à 24 h. Le modèle M3 surestime une partie de la montée de crue et sous-estime plus fortement le pic que GRP.

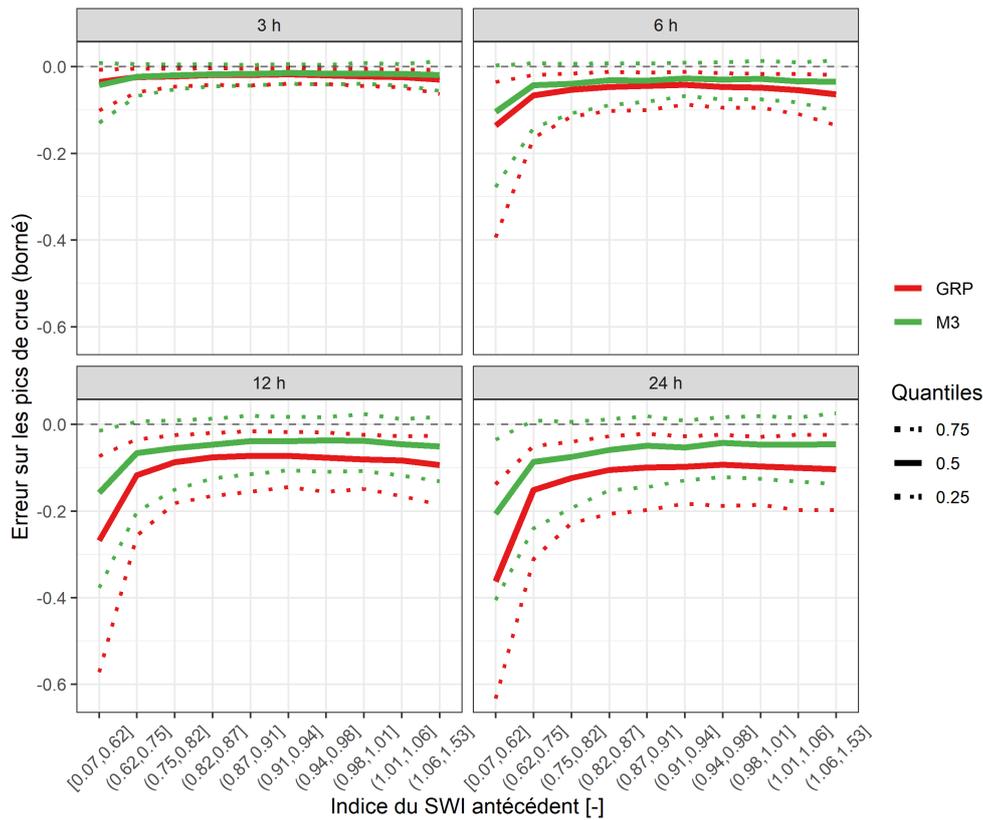


FIGURE 8.19 – Distribution des performances de GRP et M3 sur les pointes de crue en fonction des conditions antécédentes d'humidité du sol (SWI; validation croisée). Les résultats sont présentés pour quatre échéances de prévision. Les événements ont été répartis en dix classes de quantile d'humidité antécédente.

L'Estéron au Broc (443 km²) est un bassin côtier méditerranéen dont les hautes eaux ont lieu en hiver et au début de l'automne. On observe que les deux modèles sous-estiment le volume et le pic d'une crue qui a eu lieu en novembre 2014, en particulier à 12 h et à 24 h. Le modèle M3 prévoit mieux le pic et la montée de crue que GRP pour ces horizons. La différence est plus marquée à 24 h. À 3 h, les prévisions des deux modèles sont proches mais le pic et la montée de crue sont un peu mieux simulés par M3. Le décalage du pic à l'horizon 12 h est dû à l'assimilation de la surestimation du débit juste avant la montée de crue (12 h avant). On ne retrouve pas ce décalage à 24 h car les prévisions du débit 24 h avant le pic surestiment moins le débit observé.

La Cèze à Tharoux (665 km²) est un affluent en aval du Rhône en rive droite. Les périodes de hautes eaux ont lieu en automne sur ce bassin. La crue de septembre 2015 est caractérisée par des intensités de pluie allant jusqu'à 28 mm/h et une montée de crue d'une durée de 5 h. M3 prévoit mieux la montée, le volume et le pic de crue que GRP pour les trois horizons. Pour cet événement, la sous-estimation importante du volume de crue par GRP n'est pas rattrapée par son assimilation de données, même à courte échéance.

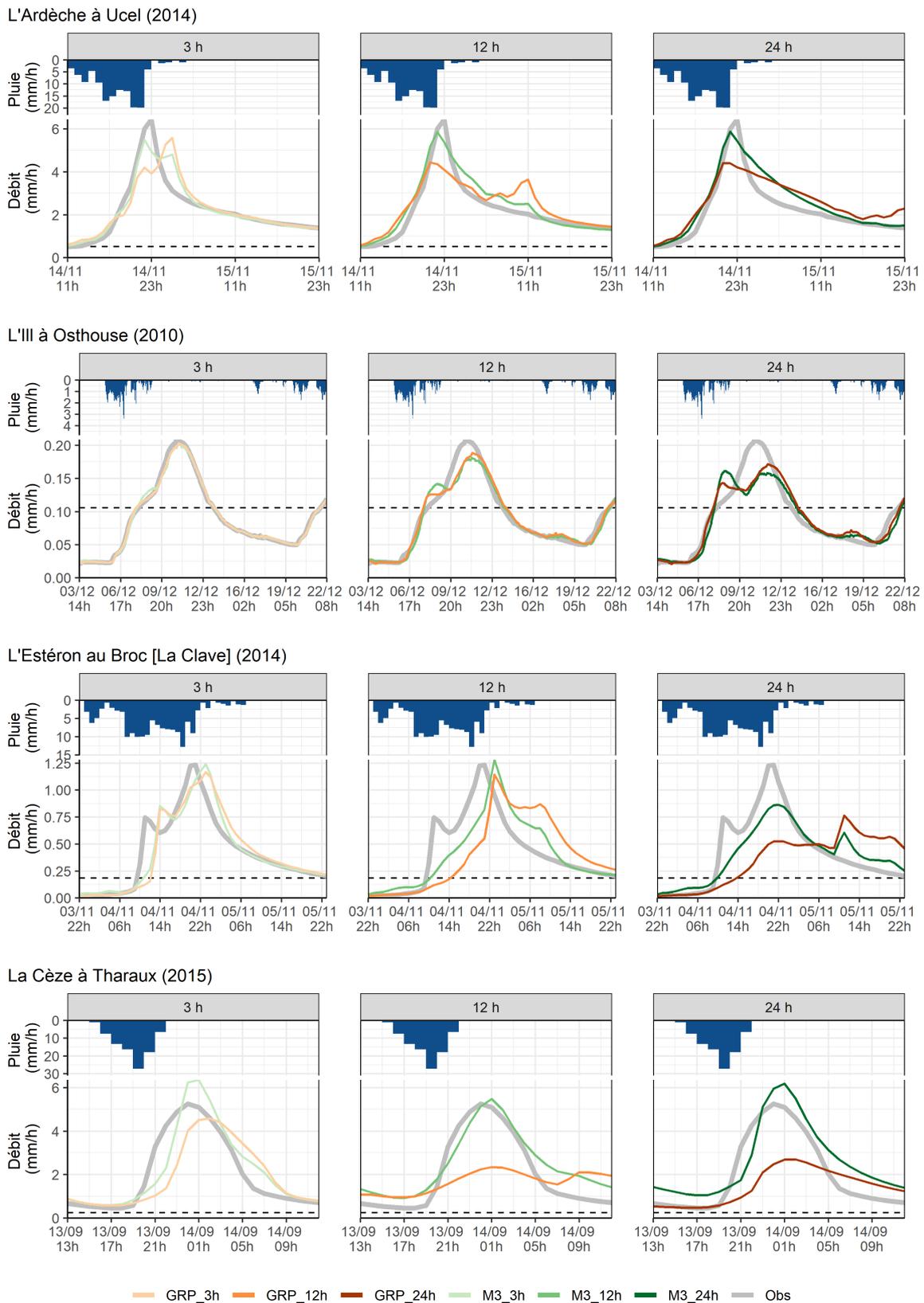


FIGURE 8.20 – Exemples d'hydrogrammes de crue (à horizon fixe) prévus par les modèles GRP et M3 sur quatre bassins versants pour trois échéances de prévision (validation). La ligne en pointillés noirs est le quantile 95 % des débits.

TABLE 8.3 – Performances de GRP et M3 pour les hydrogrammes de la figure 8.20. RMSEm [-] est la RMSE sur les montées de crue rapportée au module du bassin versant. ΔQ_{\max} [-] est l'erreur sur les pics de crue (borné). NSEev est le NSE événementiel borné.

Horizon	Critère	L'Ardèche à Ucel (2014)		L'Ill à Osthouse (2010)		L'Estéron au Broc (2014)		La Cèze à Tharoux (2015)	
		GRP	M3	GRP	M3	GRP	M3	GRP	M3
3 h	RMSEm	6.92	3.54	0.13	0.20	4.69	4.14	20.41	12.78
	ΔQ_{\max}	-0.18	-0.15	-0.03	-0.04	-0.05	0.01	-0.13	0.21
	NSEev	0.81	0.93	0.99	0.99	0.80	0.84	0.69	0.84
12 h	RMSEm	6.31	3.36	0.47	0.57	8.76	5.61	22.87	7.25
	ΔQ_{\max}	-0.31	-0.09	-0.12	-0.14	-0.07	0.04	-0.55	0.04
	NSEev	0.84	0.94	0.96	0.94	0.41	0.71	0.56	0.91
24 h	RMSEm	5.99	3.02	0.77	0.93	8.72	4.71	24.38	10.62
	ΔQ_{\max}	-0.32	-0.09	-0.22	-0.25	-0.57	-0.30	-0.49	0.18
	NSEev	0.86	0.91	0.87	0.83	0.44	0.82	0.56	0.84

8.3.3 Correction des sorties

Berthet (2010) montre que la mise à jour des sorties du modèle par la méthode de Tangara (2005) a une influence sur les prévisions de GRP jusqu'à l'horizon 6 h. Nous avons vu que la comparaison des performances des modèles dépend de l'horizon de prévision, GRP ayant des performances à l'horizon 3 h plus élevées voire équivalente à M3 selon certains critères. Nous regardons maintenant si l'ajout d'une mise à jour des sorties (cf. eq. 8.5) change l'ordre des performances. La figure 8.21 présente les performances globales et événementielles de GRP, M0 et M3 en appliquant ou non la méthode de Tangara (2005). Comme Berthet (2010), on observe que la mise à jour des sorties a une influence importante sur les performances aux horizons 3 h et 6 h et très peu d'influence aux horizons 12 h et 24 h. On observe une légère dégradation des performances à l'horizon 24 h. L'ordre des performances des modèles reste inchangé pour tous les horizons. On observe une influence de la mise à jour sur les performances événementielles à l'horizon 3 h. À 12 h et à 24 h, la mise à jour dégrade les performances événementielles des trois modèles. L'ordre des performances des modèles reste inchangé pour tous les horizons.

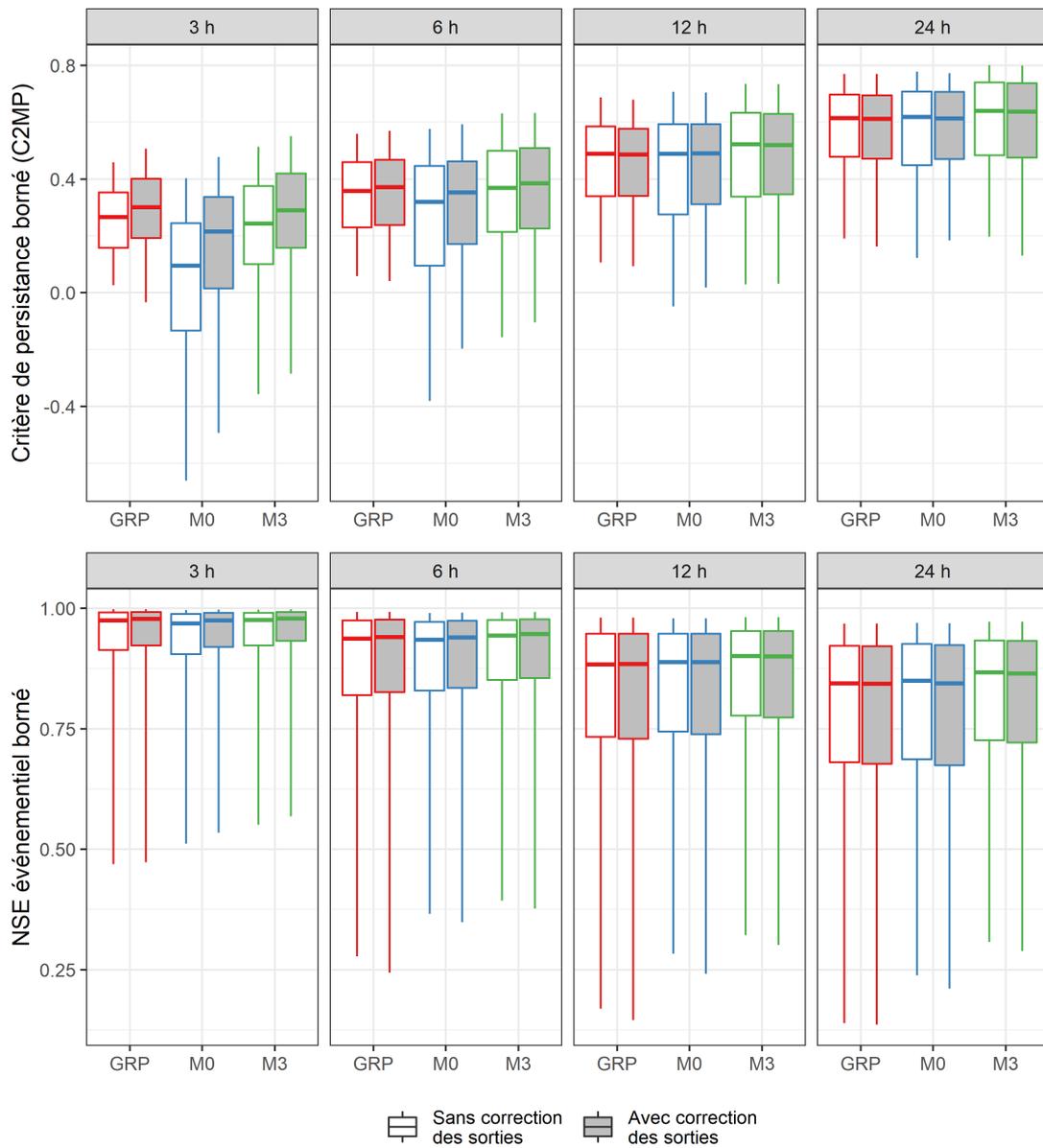


FIGURE 8.21 – Comparaison des performances globales (C2MP) et événementielles (NSEEv) des modèles GRP, M0 et M3 utilisés avec et sans méthode de correction des sorties pour quatre échéances de prévision (validation croisée). Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

8.4 Conclusion

L'objectif des travaux de ce chapitre était d'évaluer l'influence d'un changement de structure sur la qualité des prévisions déterministes et vérifier notre hypothèse selon laquelle l'amélioration de la structure du modèle de simulation permet des gains en prévision. Pour répondre à cet objectif, nous avons construit une méthode de mise à jour des états initiaux des modèles M0 et M3 s'inspirant de la méthode de mise à jour du modèle GRP. Les prévisions des trois modèles ont ensuite été évaluées à plusieurs échéances à l'aide de nombreux critères événementiels et globaux et en fonction de plusieurs indicateurs de crue et caractéristiques des bassins.

Les résultats ont montré que le modèle M3 a de meilleures performances globales que le modèle GRP sauf à très courte échéance selon le critère C2MP. Le modèle M0 a de moins bonnes performances globales que M3 et GRP sauf à l'horizon 48 h. La répartition spatiale des performances globales a permis d'identifier que GRP est plus performant pour les premiers horizons de prévision sur des bassins situés sur la moitié nord de la France. On observe quelques dégradations des performances avec le modèle M3 pour des bassins du sud de la France. Les meilleures performances globales de GRP à courte échéance s'expliquent en partie par une meilleure assimilation de l'erreur à l'instant de prévision. Les moins bonnes performances du modèle M0 par rapport à GRP et M3 à très courte échéance sont liées à la méthode de mise à jour employée qui ne permet pas de reproduire exactement le débit observé à l'instant de prévision.

Les prévisions événementielles de M3 sont plus performantes que les prévisions de GRP à partir de l'horizon 6 h. Le modèle M3 simule beaucoup mieux les pointes de crue que le modèle GRP et en particulier pour les événements d'été. À 3 h, les performances événementielles de M3 et GRP sont presque équivalentes. M3 prévoit toutefois mieux les montées de crue pour cet horizon. M0 a de meilleures performances événementielles que GRP à partir de l'horizon 12 h mais reste moins performant que GRP à très courte échéance et moins performant que M3 pour tous les horizons. L'utilisation du modèle M3 semble particulièrement utile pour prévoir les crues de nombreux affluents du Rhône et plusieurs affluents en amont de la Loire. On observe aussi des améliorations pour plusieurs bassins côtiers méditerranéens, mais ces améliorations ne sont pas aussi importantes que ce qui aurait pu être attendu avec l'utilisation des fonctions d'intensité en prévision. Les performances événementielles du modèle sont très dépendantes de la dynamique des crues et du temps de réaction des bassins. L'utilisation de M3 améliore la prévision des montées de crue sur une majorité d'événements dont les temps de montée sont inférieurs à 12 h. Par ailleurs, la prévision des pics de crue est particulièrement améliorée lorsque les événements de crue sont associés à des fortes intensités de pluie et/ou lorsque l'humidité antécédente des sols est faible. Enfin, l'utilisation d'une méthode simple de mise à jour des sorties des modèles améliore les prévisions à courte échéance mais ne change pas les conclusions de nos analyses.

Les conclusions de ce chapitre sont dépendantes de la méthode de mise à jour que nous avons employée ainsi que de la manière dont nous avons paramétré les modèles. Le choix de paramétrer le modèle en simulation pourrait notamment avoir une influence sur les performances à très courte échéance. Par ailleurs, nous avons vu dans le chapitre précédent que le paramétrage de la fonction d'échange de GR5H donne lieu à une forte incertitude para-

métrique. La pertinence de l'utilisation de cette fonction d'échange en contexte de prévision reste à démontrer. Il est en effet possible que les différences de performances entre M3 et GRP pour les bassins méditerranéens soient liées à cette fonction d'échange, en particulier pour les bassins à faible rendement annuel. Plus généralement, les résultats présentés dans ce chapitre impliquent de réfléchir à nouveau aux compromis entre la complexification de la structure, les gains de performance et l'incertitude paramétrique. L'amélioration des prévisions déterministes pour une partie des événements liés à des dynamiques rapides des bassins versants justifie-t-elle une complexification aussi importante de la structure, de l'estimation des paramètres et de la mise à jour ?

Chapitre 9

Vers une estimation des paramètres adaptée à la prévision

9.1 Contexte

Dans le chapitre précédent, les modèles de prévision ont été calés en simulation en utilisant une fonction mono-objectif. Tous les paramètres libres des deux modèles ont été ajustés par bassin et sous-période.

Les résultats du chapitre 7 ont montré que certains paramètres du modèle M3 sont difficilement identifiables, en particulier les paramètres X_5 (fonction d'échange) et K_L (réservoir linéaire). Ces difficultés d'identification induisent une forte variabilité des valeurs de ces paramètres entre périodes de calage. L'utilisation de valeurs fixées pour tous les bassins n'engendre pas de dégradations importantes des performances en validation. Les procédures de mise à jour des états initiaux ayant un impact sur le fonctionnement du modèle, il est possible que ces résultats changent en prévision. L'incertitude paramétrique a peut-être un impact plus important sur les performances en prévision. À l'inverse, l'utilisation de l'erreur à chaque instant de prévision permet peut-être de compenser une partie de cette incertitude.

Considérer que les jeux de paramètres du modèle de prévision peuvent différer de ceux estimés en simulation peut sembler contre-intuitif. En effet, lorsqu'on se place dans un contexte stationnaire (ce que l'on fait en prévision des crues), les paramètres sont par définition des invariants temporels (e.g. Moradkhani *et al.*, 2005) caractérisant de façon unique le comportement hydrologique des bassins versants. Cependant, comme nous l'avons vu dans le chapitre 7, les jeux de paramètres estimés au calage sont nécessairement associés à un certain degré d'incertitude car les modèles hydrologiques sont loin de reproduire parfaitement la relation pluie-débit (Berthet, 2010), cela quels que soient les modèles utilisés.

Si les paramètres du modèle de prévision peuvent être différents de ceux du modèle de simulation, il faut déterminer le degré de liberté qu'il est nécessaire d'accorder pour obtenir les meilleures performances en prévision. On trouve plusieurs options dans la littérature. Une manière de faire varier les paramètres d'un modèle de prévision est de les mettre à jour à partir de l'assimilation du débit sur plusieurs pas de temps avant l'instant de prévision (e.g. Yang & Michel, 2000; Ercolani & Castelli, 2017). On peut aussi choisir de caler les paramètres par horizon de prévision, ce qui est plus inhabituelle dans la littérature. Berthet

(2010) a étudié l'influence des mises à jour des états et des corrections des sorties sur les paramètres du modèle GRP. Il trouve que la mise à jour du réservoir de routage a une forte influence sur les paramètres du modèle. Il montre que le calage des paramètres en même temps que la mise à jour du modèle améliore les performances en validation à toutes les échéances. La variation des paramètres entre horizons de calage n'est pas très importante et dépend principalement de la procédure de correction des sorties. Il suggère de caler le modèle par horizon pour obtenir les meilleures performances possibles en validation. Si un seul horizon de calage est utilisé, il est préférable de caler à un horizon de un à plusieurs jours plutôt qu'à très courte échéance afin d'obtenir de bonnes performances pour tous les horizons.

Des travaux plus récents sur le modèle GRP (Pinna, 2017) ont montré que le calage de certains paramètres du modèle en simulation plutôt qu'en prévision améliore la performance des prévisions. En particulier, le calage du paramètre de correction de la pluie efficace en simulation permet de réduire les sous-estimations des volumes de crue. Ces résultats montrent que le calage des paramètres lorsque les procédures d'assimilation sont activées peut avoir un impact négatif sur le comportement du modèle. Li *et al.* (2015) montrent par exemple que l'estimation des paramètres du modèle GR4J combinée à une procédure de correction des sorties peut accentuer les effets indésirables d'une telle procédure et proposent une méthode de correction qui permet d'obtenir des paramètres plus proches de ceux du modèle de simulation. Bellier (2018) cale le modèle GRP en simulation afin d'éviter ces compensations paramétriques.

L'objectif de ce chapitre est d'explorer différentes méthodes d'estimation des paramètres du modèle M3 afin de proposer une stratégie adaptée à la prévision opérationnelle. Pour cela, nous évaluons dans un premier temps si l'utilisation de valeurs fixes pour les paramètres X_5 et K_L ne dégrade pas les performances en prévision. Nous étudions ensuite l'impact de la sélection d'un jeu de paramètres en simulation à partir d'une fonction objectif intégrant une composante événementielle sur les prévisions du modèle (l'utilisation de cette méthode d'estimation des paramètres avait donné lieu à une amélioration des performances événementielles en simulation ; cf. chapitre 7). Enfin, nous analysons les performances d'une estimation des paramètres effectuée par horizon de prévision afin de vérifier que l'ajout d'un degré de liberté supplémentaire dans l'estimation des paramètres améliore bien les performances et n'impacte pas la robustesse du modèle. En ce sens, nous explorons la dépendance des paramètres aux horizons de prévision et l'impact d'une telle méthode sur la variabilité des paramètres entre périodes de calage. Nous étudions ainsi les potentielles interactions entre le choix de la méthode d'estimation des paramètres et la mise à jour du modèle.

9.2 Éléments méthodologiques

Cinq méthodes sont appliquées pour estimer les paramètres du modèle M3 (tab. 9.1).

L'exploration de l'espace des paramètres présentée dans le chapitre 7 est utilisée pour les méthodes « Expl », « Comb » et « Hori ». La fonction objectif utilisée pour la méthode « Comb » est la suivante :

$$Fobj = \frac{1}{2} \times (NSE + NSE_{cEv}) \quad (9.1)$$

TABLE 9.1 – Méthodes d’estimation des paramètres testées pour le modèle M3 en prévision.

Méthode	Signification
Ref	Calage de référence en simulation (à partir de l’algorithme EGD ; voir 2.4.1)
X5KL fixés	Calage en simulation en utilisant les mêmes valeurs de X_5 et K_L pour tous les bassins : $X_5 = 0.125$ et $K_L = 0.115$. (voir tab. 7.4)
Expl	Sélection du meilleur jeu de paramètres en simulation selon le critère NSE à partir d’une large exploration de l’espace des paramètres
Comb	Sélection du meilleur jeu de paramètres en simulation selon un critère composite intégrant une composante événementielle (eq. 9.1)
Hori	Sélection du meilleur jeu de paramètres par horizon de prévision selon le critère NSE

avec NSE_{cEv} le critère NSE calculé pour une sous-période et un bassin donnés sur tous les pas de temps des événements sélectionnés et NSE le critère calculé sur toute la chronique pour une sous-période et un bassin donnés.

La sélection du jeu de paramètres optimal par horizon de prévision est effectuée à partir des 2 500 meilleurs jeux de paramètres (selon le critère NSE) sélectionnés par bassin et par sous-période en simulation. Les jeux de paramètres optimaux par horizon ne se trouvent pas nécessairement parmi ces 2 500 jeux de paramètres. Ce choix est motivé par les temps de calcul associés au modèle de prévision. Par ailleurs, les paramètres ainsi sélectionnés ne sont pas trop éloignés des paramètres optimaux du modèle de simulation, ce qui devrait en principe permettre d’éviter une sur-adaptation des paramètres à l’horizon de calage. Nous reprenons trois des critères utilisés dans le chapitre 8 pour évaluer les performances des différentes méthodes d’estimation des paramètres : le C2MP, le NSE événementiel borné et le critère d’erreur sur les pics de crue (adapté à la prévision à horizon fixe).

9.3 Résultats et analyses

La figure 9.1 présente les performances des différentes méthodes d’estimation des paramètres de M3 par rapport à la méthode de référence. À 3 h, les performances de « X5KL fixés » sont plus élevées que les performances de « Ref » selon le C2MP et le NSE événementiel. Les performances sont équivalentes à 6 h et un peu inférieures à 12 h et 24 h selon ces deux critères. On n’observe pas de différences pour la simulation des pics de crue. Ces résultats indiquent qu’il est possible de réduire la complexité paramétrique du modèle en passant de huit à six paramètres libres sans dégrader trop fortement les performances en prévision. Ces résultats indiquent aussi que la complexité nécessaire peut dépendre de l’horizon de prévision. À court horizon, l’information portée par le débit a plus d’importance qu’à longue échéance, diminuant ainsi la valeur de l’information portée par la structure du modèle. Il est ainsi préférable de diminuer le degré de liberté du modèle à courte échéance.

Les valeurs de C2MP de « Expl » sont plus élevées que les valeurs de C2MP de « Ref » pour les quatre horizons de prévision et en particulier à 3 h et à 6 h. Les performances de « Expl » selon le NSE événementiel sont plus élevées que celles de « Ref » à 3 h et équivalentes

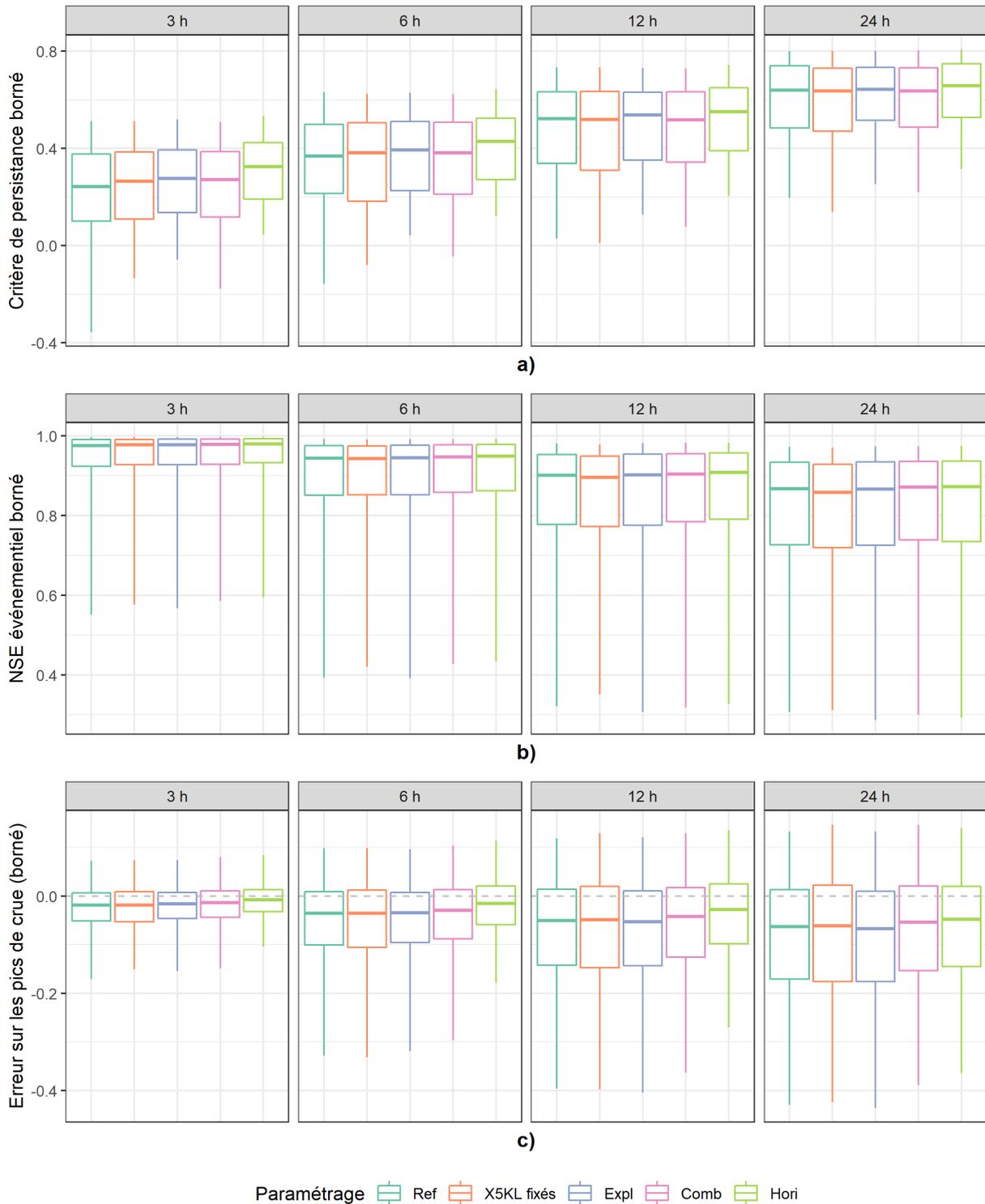


FIGURE 9.1 – Distribution des performances globales et événementielles du modèle M3 selon cinq méthodes d’estimation des paramètres (validation croisée) pour quatre horizons de prévision. Les boîtes à moustaches sont tracées du quantile 5 % au quantile 95 %.

aux autres horizons. La configuration « Expl » améliore les performances sur les pics de crue à 3 h et à 6 h et les dégrade à 24 h. Les performances à 12 h sont équivalentes à « Ref » selon ce critère. Ces résultats n'étaient pas attendus puisque nous obtenions des performances similaires entre ces deux méthodes pour le modèle de simulation. Nous y revenons plus loin dans ce chapitre.

Les performances globales de « Comb » sont moins élevées que les performances de « Expl » mais restent plus élevées que les performances de « Ref », cela pour tous les horizons de prévision. Les performances événementielles de cette méthode sont plus élevées que celles de « Expl » pour tous les horizons et selon les deux critères. Ces résultats sont cohérents avec ceux obtenus dans le chapitre 7 : l'utilisation d'une méthode d'estimation des paramètres qui intègre une composante événementielle dégrade les performances globales mais améliore les performances événementielles.

Les performances globales de « Hori » sont nettement supérieures aux performances des quatre autres méthodes, cela pour tous les horizons de prévision. Les performances événementielles de « Hori » selon le NSE sont supérieures aux performances des autres méthodes à 3 h, 6 h et 12 h et équivalentes à « Comb » à l'horizon 24 h. Les performances de « Hori » sur les pics de crue sont aussi nettement plus élevées que les performances des autres méthodes, en particulier à 3 h et à 6 h.

Ces résultats montrent qu'il est possible d'améliorer les performances du modèle M3 en prévision en utilisant une méthode d'estimation des paramètres plus adaptée à la prévision.

L'amélioration du C2MP à 3 h par les méthodes « X5KL fixés » et « Expl » n'était pas attendue. Une explication possible se trouve au niveau du réservoir linéaire de M3. La figure 9.2 présente les distributions du paramètre (K_L) de vidange du réservoir linéaire pour les méthodes « Ref » et « Expl ». On observe que les distributions du paramètre K_L sont très différentes lorsque « Expl » a de meilleures performances en C2MP que « Ref », cela pour les quatre horizons de prévision. K_L a des valeurs plus faibles pour la méthode « Expl » dans ce cas. Par ailleurs, plus de 25 % des valeurs de K_L sont égales à 1 lorsque « Expl » a de meilleures performances en C2MP que « Ref ». Cela signifie que le réservoir linéaire n'est pas « activé » pour ces bassins et sous-périodes (le réservoir est « transparent » : tout ce qui sort du réservoir sur le pas de temps est égal à ce qui y entre). La méthode initiale d'estimation des paramètres de M3 a été construite de manière à ce que le modèle M0 soit testé lors de la phase de préfiltrage (voir 2.4.1 et 5.3.4), cela afin d'activer les fonctions d'intensité et le réservoir linéaire seulement si la différence au calage par rapport à M0 est suffisamment élevée. Cette méthode induit une difficulté en prévision car, sans réservoir linéaire sur la branche directe, le modèle M3 a une mise à jour proche de celle de M0. En effet, dans ce cas, seul le réservoir de routage sur la branche indirecte est mis à jour à l'instant de prévision. Nous avons vu dans le chapitre précédent que cette mise à jour est moins efficace à courte échéance que la mise à jour de M3 et de GRP, en particulier à courte échéance.

La figure 9.3 illustre la difficulté de mettre à jour le modèle M3 sans réservoir linéaire, en particulier lorsque la branche directe porte une partie importante du débit. La Loyre est un sous-affluent en rive droite de la Dordogne par la Vézère. La crue d'octobre 2000 n'est pas caractérisée par des intensités de pluie très élevées. Cependant, le modèle M3 simule cet événement avec une part importante du débit venant de la branche directe (courbe grise des figures 9.3c et 9.3d), ce qui signifie que les fonctions d'intensité ont été activées. On observe sur les figures 9.3a et 9.3b que la simulation de l'événement par l'OL est similaire pour les

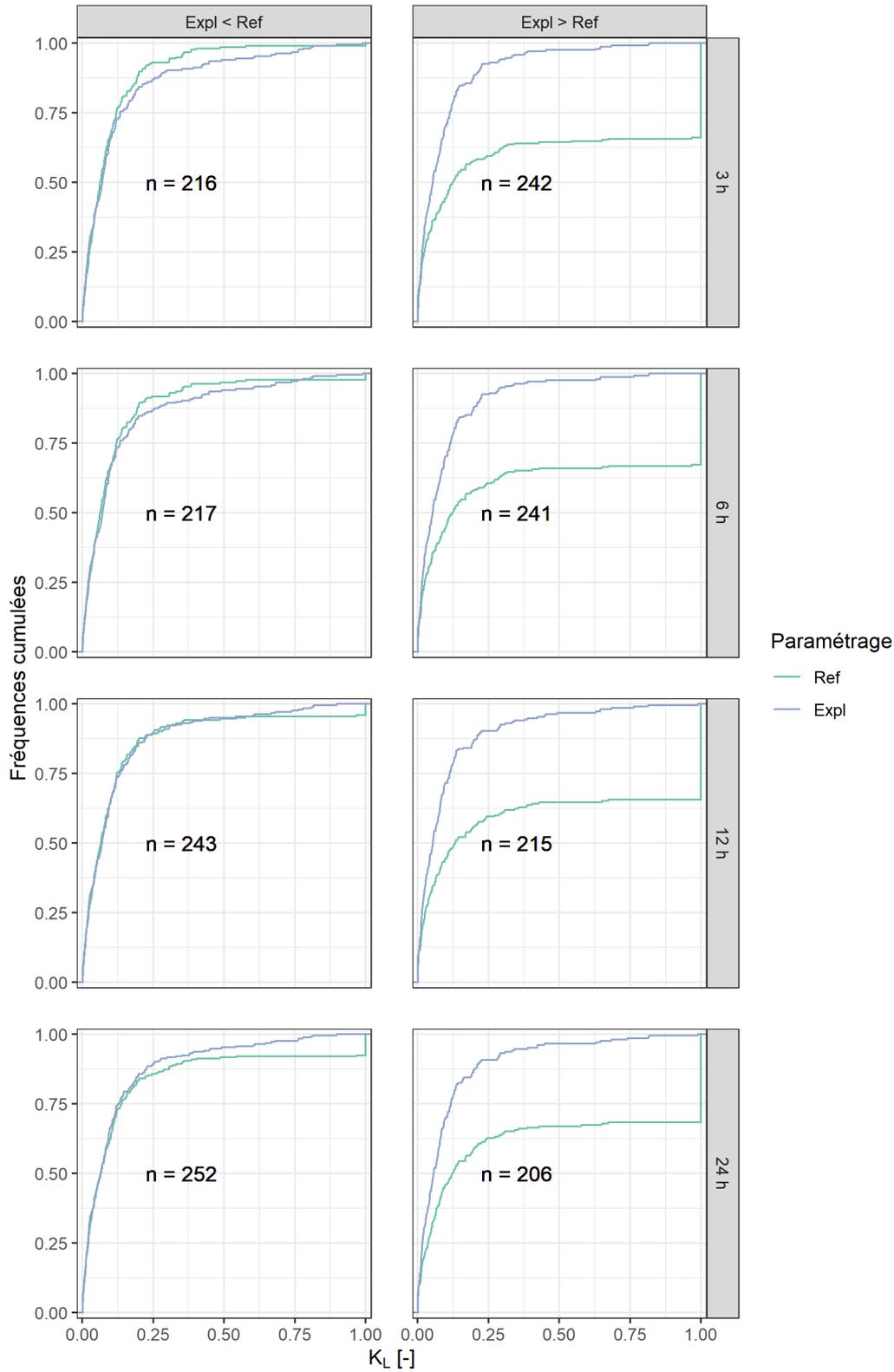


FIGURE 9.2 – Distribution du paramètre K_L selon deux méthodes d'estimation des paramètres du modèle M3 et quatre horizons de prévision. K_L ne varie pas en fonction de l'horizon. Les distributions sont séparées en deux groupes : les bassins et périodes pour lesquels la méthode « Expl » a de meilleures performances selon le C2MP (validation croisée) que la méthode de référence « Ref », et les bassins et périodes pour lesquels c'est l'inverse.

deux configurations (« X5KL fixés » et « Ref »). En revanche, « X5KL fixés » prévoit mieux la crue à 3 h que « Ref ». On observe sur les figures 9.3c et 9.3d que le réservoir de routage de la branche indirecte est mis à jour de la même manière pour les deux configurations. Comme le flux de la branche indirecte Q_r est moins élevé que celui de la branche pseudo-directe Q_d , la différence entre le niveau de l'OL (R) et le niveau mis à jour (R^*) n'est pas très élevée. La figure 9.3d montre que le réservoir linéaire est mis à jour lors de la montée de crue sous-estimée par l'OL, ce qui permet au modèle de mieux prévoir la crue que lorsque le réservoir linéaire n'est pas activé.

Nous avons vu sur la figure 9.1 que l'estimation des paramètres par horizon de prévision améliore très largement les performances du modèle M3. Nous cherchons maintenant à étudier l'influence de cette méthode sur la variation des paramètres du modèle. La figure 9.4 présente les écarts des valeurs des paramètres par horizon aux valeurs obtenues pour la version OL. Les variations des paramètres X_1 , X_2 , X_5 et i_1 sont similaires entre horizons de prévision. Pour les paramètres X_3 , X_4 , i_2 et K_L , les distances euclidiennes augmentent plus l'horizon de prévision diminue. Les variations des paramètres X_2 , X_5 , i_1 , i_2 et K_L sont plus importantes que les variations des paramètres X_1 , X_3 et X_4 . Ces résultats indiquent d'une part que les paramètres de la fonction d'échange et des fonctions d'intensité sont les plus sensibles à une estimation des paramètres en prévision par rapport à une estimation des paramètres en simulation et d'autre part que les paramètres X_3 , X_4 , i_2 et K_L ont une dépendance à l'horizon de prévision choisi.

Une des manifestations des difficultés d'identification des paramètres est la variation des valeurs de ces paramètres entre périodes de calage (voir chapitre 7). La figure 9.5 présente la corrélation des paramètres (valeurs réelles, i.e. non transformées) du modèle M3 entre périodes de calage selon l'horizon de prévision. On observe très peu de changements pour le paramètre X_1 à part à 3 h où la corrélation linéaire est plus faible. La corrélation maximale obtenue pour le paramètre X_2 est pour l'échéance 12 h. On observe des corrélations plus faibles en prévision pour le paramètre X_3 par rapport à l'OL. Pour le paramètre X_4 , les corrélations à courte échéance sont un peu plus faibles que la corrélation en OL. Ces résultats sont cohérents avec les résultats de Pinna (2017) qui ont montré qu'un calage du temps de base de l'hydrogramme unitaire et de la capacité du réservoir de routage de GRP en simulation améliore les performances en comparaison avec un calage de ces paramètres en prévision. On observe quelques variations de la corrélation entre horizons de prévision pour le paramètre X_5 . Les corrélations obtenues pour le paramètre i_1 sont plus faibles en prévision qu'en simulation. Pour le paramètre i_2 , la corrélation baisse lorsque l'horizon de prévision diminue. La corrélation diminue d'environ 0.5 entre l'OL et l'horizon 3 h. À l'inverse, la corrélation de K_L entre périodes de calage augmente plus l'horizon diminue. Cela est probablement dû à l'activation plus systématique du réservoir linéaire en prévision (voir fig. 9.2). En résumé, le calage par horizon de prévision proposé ici diminue les corrélations entre périodes de calage pour la plupart des paramètres. Ces diminutions semblent plus marquées pour les deux paramètres des fonctions d'intensité. À l'inverse, la stabilité du paramètre K_L est améliorée avec cette méthode, en particulier à courte échéance.

Ces résultats sont à nuancer avec l'utilisation de la corrélation linéaire comme indicateur de variation des paramètres entre périodes de calage. Les valeurs de corrélation linéaire peuvent baisser très largement lorsque qu'il y a des points plus dispersés que la moyenne. La figure 9.6 montre par exemple que les corrélations du paramètre X_3 entre périodes selon l'ho-

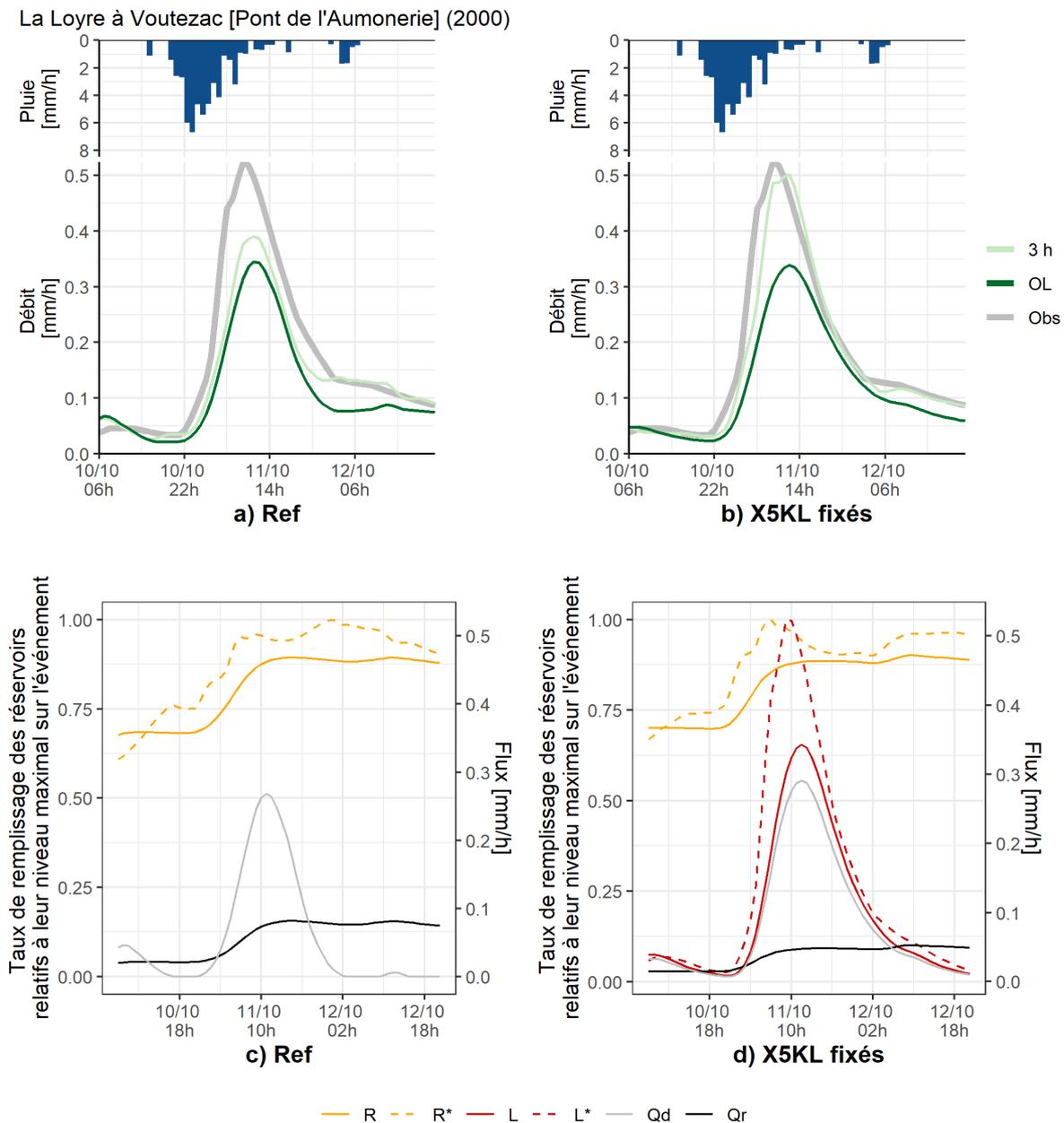


FIGURE 9.3 – Exemple d’une crue prévue sur le bassin du Loyre à Vutezac (sous-affluent de la Dordogne par la Vézère) par le modèle M3 selon deux méthodes d’estimation des paramètres : la méthode de référence (a) et la méthode avec les paramètres X_5 et K_L fixés (b). Les figures (c) et (d) présentent les variations des réservoirs des deux branches de routage ainsi que les flux venant de ces deux branches. R et R^* sont les niveaux du réservoir de la branche indirecte avant et après mise à jour. L et L^* sont les niveaux du réservoir de la branche pseudo-directe avant et après mise à jour. Q_r et Q_d sont les flux des branches indirecte et pseudo-directe respectivement (calculés par l’OL).

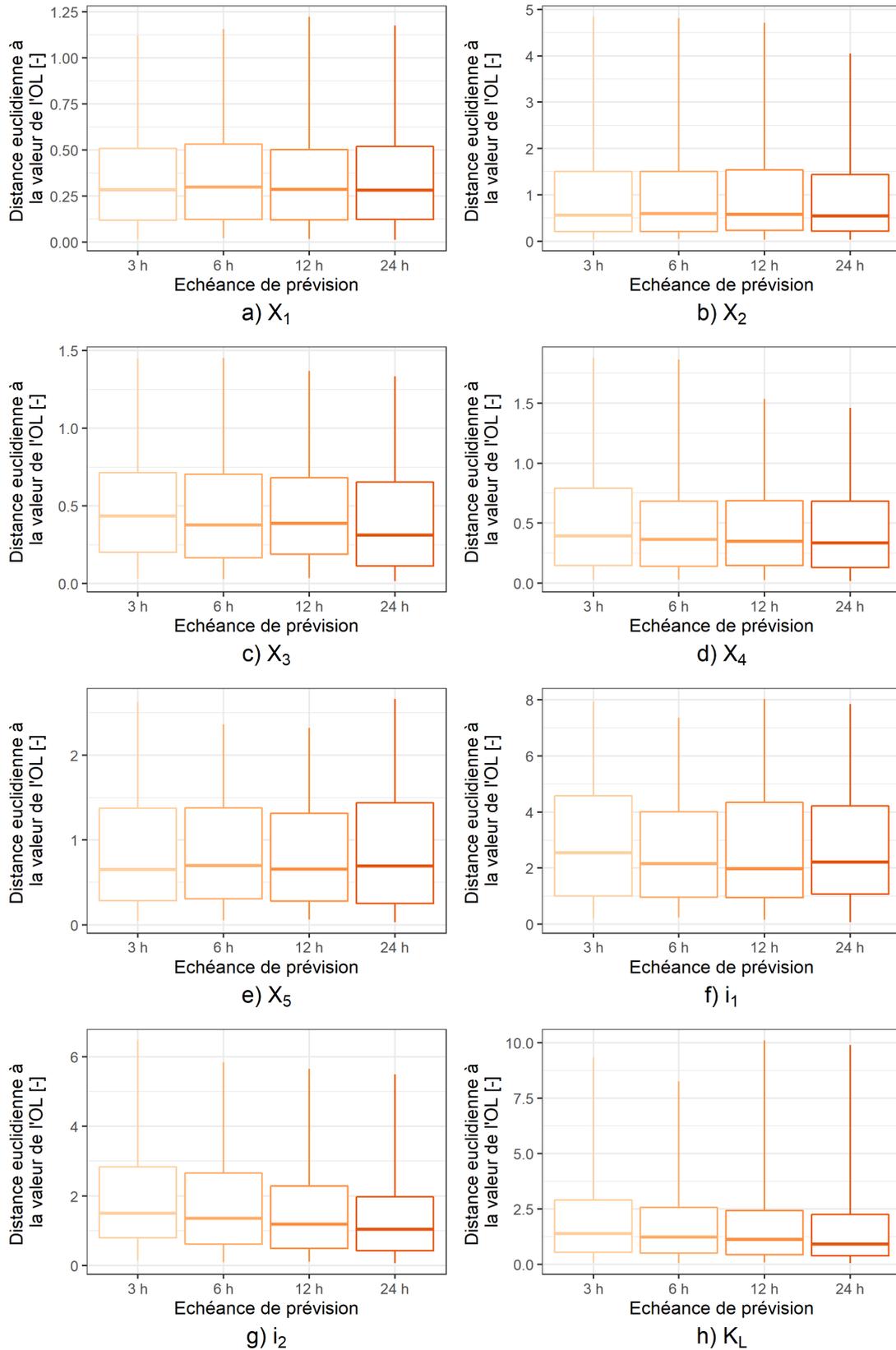


FIGURE 9.4 – Distributions des valeurs absolues de la différence entre les paramètres sélectionnés par horizon (« Hori ») et leur valeur en version OL (« Expl ») en fonction de l’horizon de prévision. Les distances ont été calculées dans l’espace des paramètres transformés.

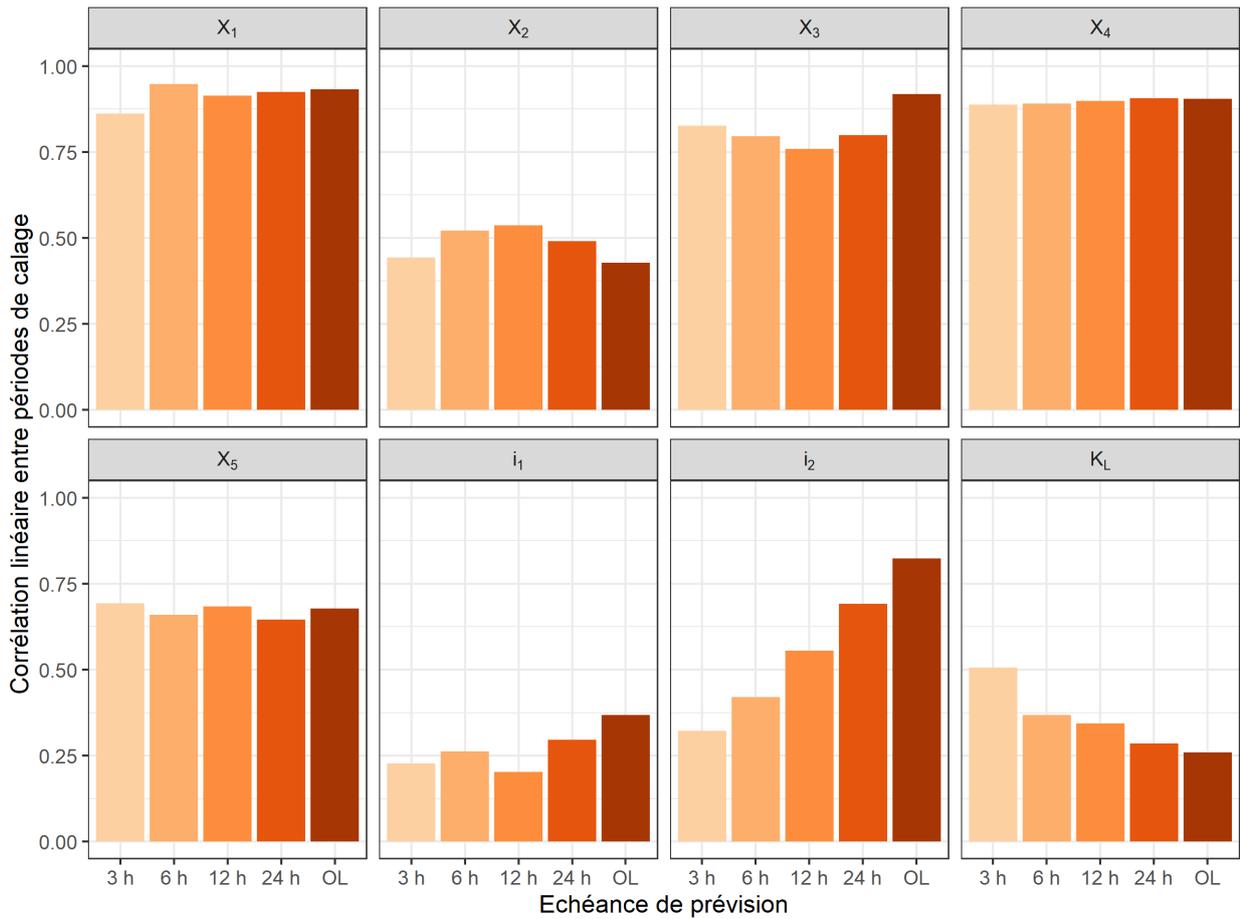


FIGURE 9.5 – Corrélations des valeurs (réelles, i.e. non transformées) des paramètres du modèle M3 entre périodes de calage selon quatre horizons de prévision et l’OL.

Le rizon de prévision dépend de la position de quelques points très dispersés. En revanche, pour le paramètre i_2 , on observe que la dispersion des points diminue avec l’augmentation de l’horizon de prévision.

Comme les performances obtenues avec la méthode « Hori » sont supérieures aux performances obtenues avec les autres méthodes, les jeux de paramètres optimaux au calage peuvent être différents entre l’OL et les différents horizons de prévision. Pour chaque bassin et chaque sous-période, nous avons tracé les performances obtenues par horizon en fonction des performances obtenues en OL pour les 2 500 jeux de paramètres testés. Pour l’horizon 3 h où les différences avec l’OL sont plus marquées que pour les autres horizons, nous observons trois configurations sur notre base de données (fig. 9.7) :

- Le jeu de paramètres optimal pour l’OL est le même que le jeu de paramètres optimal pour l’horizon de prévision considéré. Ce cas est illustré par le graphique représentant cet optimum pour le bassin de l’Aille à Vidauban, un affluent de l’Argens qui est un bassin côtier méditerranéen.

- Le jeu de paramètres optimal pour l'OL est différent du jeu de paramètres optimal pour l'horizon de prévision considéré. Plusieurs valeurs très éloignées du critère de performance en OL correspondent à des valeurs très proches du critère en prévision. Cette configuration est illustrée par l'exemple sur le Tarn à Mostuéjols, un des principaux affluents de la Garonne.
- Le jeu de paramètres optimal pour l'OL est différent du jeu de paramètres optimal pour l'horizon de prévision considéré. L'optimum mathématique de l'OL est très différent de l'optimum mathématique de l'horizon de prévision considéré. En d'autres termes, il n'est pas possible de trouver un compromis permettant des performances élevées à la fois pour l'OL et pour l'échéance de prévision visée. Ce cas est illustré par l'exemple sur l'Orbieu à Saint-Martin-des-Puits, un affluent de l'Aude.

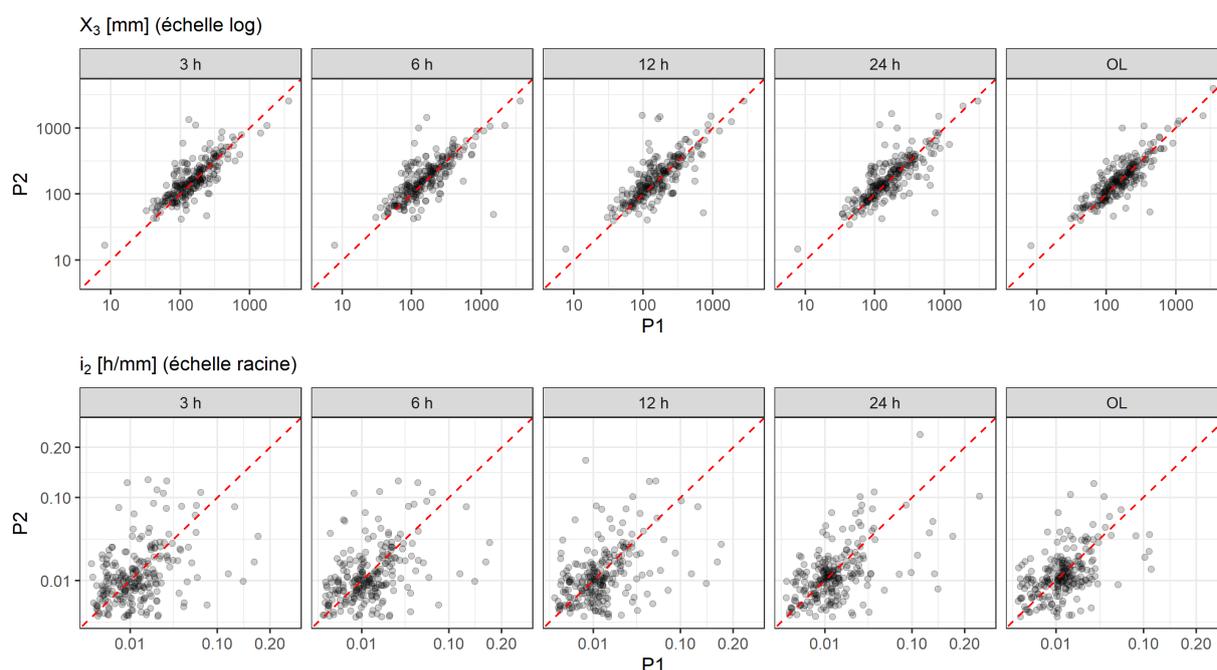


FIGURE 9.6 – Dispersion des valeurs des paramètres X_3 et i_2 du modèle M3 entre périodes de calage selon quatre horizons de prévision et l'OL.

La forme du front de Pareto¹ du troisième exemple indique que le meilleur jeu de paramètres pour l'échéance visée ne se trouve pas nécessairement parmi les 2 500 meilleurs jeux de paramètres en OL.

9.4 Conclusion

Les performances en prévision de plusieurs méthodes d'estimation des paramètres du modèle M3 ont été présentées dans ce chapitre. Les résultats ont montré que :

- L'utilisation de valeurs fixes des paramètres X_5 et K_L pour tous les bassins ne donne pas lieu à une dégradation des performances en prévision à part pour le critère C2MP à longue échéance.

1. Lors d'une optimisation multi-objectif, il existe plusieurs points qui correspondent à des compromis entre les objectifs. Un optimum de Pareto est un point représentant une partie de la limite où il n'est pas possible d'améliorer un objectif sans en dégrader un autre. Le front de Pareto est l'ensemble des optima de Pareto.

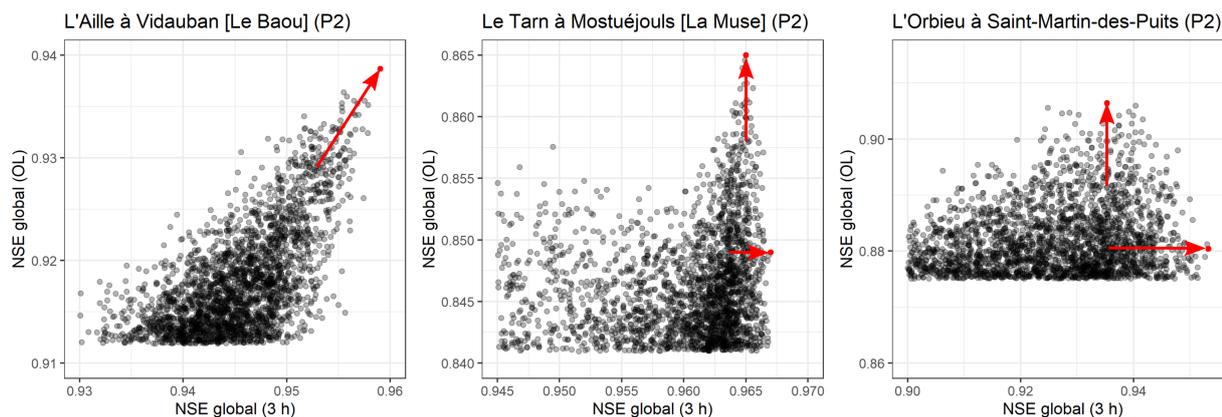


FIGURE 9.7 – Comparaison des performances du modèle M3 entre l'OL et l'échéance 3 h pour plusieurs jeux de paramètres. Exemple sur trois bassins versants. Les points rouges représentent les optima mathématiques.

- La sélection du jeu de paramètres en simulation à partir d'une exploration étendue de l'espace des paramètres permet une amélioration des performances à courte échéance. Cela est majoritairement dû à une activation systématique du réservoir linéaire qui permet une mise à jour plus efficace du modèle à l'instant de prévision.
- La sélection des jeux de paramètres en simulation à partir d'une fonction objectif qui intègre une composante événementielle améliore les performances événementielles.
- La méthode qui consiste à sélectionner un jeu de paramètres par échéance conduit à une amélioration très importante des performances globales et événementielles pour tous les horizons de prévision.
- L'estimation des paramètres par échéance induit des variations importantes des valeurs des paramètres de la fonction d'échange et des paramètres des fonctions d'intensité par rapport aux paramètres de la version OL.
- L'estimation des paramètres par échéance induit une dégradation des corrélations des paramètres entre périodes de calage. Cet effet est plus marqué pour les deux paramètres des fonctions d'intensité. À l'inverse, les corrélations associées au paramètre de vidange du réservoir linéaire augmentent plus l'horizon diminue.
- Au calage, il existe des compromis différents entre les performances de l'OL et les performances en prévision.

L'intérêt du réservoir linéaire pour la mise à jour des conditions initiales de la branche directe pose question. Si le modèle de simulation sous-estime le volume de crue comme dans l'exemple présenté sur la figure 9.3, la non conservation du volume peut avoir un intérêt. Si le modèle de simulation est simplement en retard, induisant ainsi une sous-estimation de la montée, on peut se demander si la mise à jour ne devrait pas s'appliquer au niveau de l'hydrogramme unitaire de la branche directe. Ces résultats montrent que la construction d'un modèle de prévision implique des interactions entre le choix de la structure du modèle et le choix de la méthode de mise à jour. Ces résultats ont été mis en évidence par le choix de la méthode d'estimation des paramètres qui ajoute ainsi un degré supplémentaire d'interaction.

L'utilisation d'un calage multi-objectif entre horizons de prévision et d'un calage différencié selon les paramètres seraient des perspectives intéressantes de ce travail (dans la continuité des travaux de Pinna, 2017). La dépendance de certains paramètres aux horizons de prévision renforce l'intérêt de caler certains paramètres en fonction de l'horizon. Les variations importantes entre les valeurs des paramètres en prévision et les valeurs des paramètres de la version OL pour plusieurs paramètres indiquent qu'il y aurait un intérêt à adopter un mode de calage hybride entre une estimation des paramètres en simulation et une estimation des paramètres en prévision. Certains paramètres comme la capacité du réservoir de production (X_1) seraient estimés en simulation car ils dépendent très fortement du type de bassins. Les paramètres qui ont un rôle plus important sur les dynamiques de crue comme les paramètres des fonctions d'intensité seraient calés par horizon pour que le modèle s'adapte mieux à l'information portée par le débit observé à l'instant de prévision.

L'influence des différentes méthodes d'estimation des paramètres sur les prévisions des modèles GRP et M0 n'a pas été explorée dans cette étude. L'impact d'une méthode d'estimation des paramètres par échéance sur les performances de ces modèles devra être étudié pour confirmer ou infirmer les résultats du chapitre précédent et généraliser les résultats de ce chapitre. La plupart des méthodes d'estimation des paramètres que nous avons proposées dans ce chapitre ne sont pas applicables en condition opérationnelle car les temps de calculs nécessaires pour tester quatre millions de jeux de paramètres sont très élevés. Elles constituent toutefois des pistes intéressantes pour proposer une méthode d'estimation des paramètres adaptée aux conditions opérationnelles.

Conclusion générale

Synthèse de la démarche et des résultats

L'objectif de ces travaux était d'améliorer un modèle opérationnel de prévision des crues à l'aide d'un diagnostic détaillé, en cherchant à identifier un degré de complexification adapté au niveau d'information disponible. Pour répondre à cet objectif, nous avons utilisé une large base de données de 229 bassins versants et 10 652 événements de crue (19 ans de données au pas de temps horaire). Un diagnostic des erreurs du modèle de simulation GR5H en fonction de plusieurs caractéristiques des crues a d'abord été effectué. Nous avons ensuite évalué si les caractéristiques des événements de pluie associés aux événements de crue pouvaient être liées aux erreurs du modèle. Une analyse a été menée pour déterminer si le calcul de certains flux internes du modèle pouvaient être à l'origine des déficiences identifiées. Ce diagnostic détaillé nous a permis de proposer des voies d'amélioration de la simulation des crues par le modèle GR5H. Plusieurs hypothèses sur la réaction des bassins versants à des intensités de pluie élevées ont ainsi été formulées puis évaluées. Nous confrontons ensuite ces hypothèses à l'intérêt d'une discrétisation spatiale du modèle pour améliorer la simulation des crues. Les hypothèses de modélisation proposées nous ont amené à questionner la complexité associée aux paramètres du modèle et à la manière de les estimer. En ce sens, un diagnostic a été proposé et plusieurs méthodes visant à améliorer l'estimation des paramètres dans un contexte de simulation des crues ont été testées. Nous avons ensuite construit deux modèles de prévision à partir de GR5H et de l'hypothèse de modélisation la plus performante (i.e. une stratégie de mise à jour des états a été définie), et nous les avons comparés au modèle de référence (GRP) dans un cadre de prévision des crues. Une méthodologie d'évaluation des performances des prévisions déterministes est proposée pour évaluer en détail les différences entre les modèles. Enfin, différentes méthodes d'estimation des paramètres du modèle de prévision développé dans ces travaux ont été testées afin de proposer un nouveau modèle de prévision des crues plus général et plus efficace.

Les résultats ont montré que le modèle GR5H a des performances élevées pour de nombreux bassins versants et de nombreux événements de crue mais qu'il sous-estime la majorité des crues qui ont lieu en été et au début de l'automne (chapitre 2). Ces sous-estimations sont particulièrement marquées lorsque les intensités de pluie sont élevées et lorsque l'humidité antécédente est faible (chapitres 2 et 3). La première partie de notre diagnostic a aussi mis en évidence les limites de l'utilisation de critères calculés sur la chronique complète des débits pour évaluer la capacité du modèle à simuler les crues (chapitre 2). **Les erreurs du modèle en crue sont très liées à la variabilité spatiale des cumuls de pluie, à l'intensité de la pluie à l'échelle du bassin versant et aux conditions antécédentes d'humidité des sols** (chapitre 3). Certains bassins versants, en particulier dans le sud de la France, ont

des dynamiques très rapides et des variations saisonnières importantes. Le modèle GR5H a des difficultés à réagir dans ces conditions. Cela est partiellement dû à une mauvaise estimation de la pluie efficace mais principalement à un routage trop lent de celle-ci (chapitre 4).

Ce diagnostic détaillé du modèle de simulation nous a conduit à formuler des hypothèses de modélisation pour améliorer les performances du modèle lorsque des événements de pluie intenses ont lieu en saison sèche (chapitre 5). La modification du rendement instantané des pluies en fonction de l'intensité horaire, combinée à un partage dynamique des écoulements en fonction de l'intensité de la pluie, améliore la simulation des crues en été et en hiver, en particulier pour les bassins méditerranéens et plusieurs affluents du Rhône. La modification de la dynamique des flux du modèle avec l'intensité de pluie horaire a un impact plus important sur les performances qu'une distribution simple de la pluie (chapitre 6). Celle-ci permet une amélioration de la simulation des crues associées à des pluies fortement variables dans l'espace. La complexification de la structure donne toutefois lieu à des problèmes d'identification des paramètres (chapitres 5 et 7). Le diagnostic de l'estimation des paramètres de GR5H et du nouveau modèle proposé dans cette thèse a mis en évidence des limitations souvent récurrentes dans les études sur les paramètres des modèles hydrologiques (chapitre 7). Nous avons identifié des interactions indésirables entre les paramètres de la fonction d'échange dans le modèle GR5H ainsi qu'une difficulté à contraindre ces paramètres et les paramètres des fonctions d'intensité. L'utilisation d'une valeur unique du paramètre seuil d'inversion des échanges (X_5) et d'une valeur unique du coefficient de vidange du réservoir linéaire (K_L) permet de réduire la complexité paramétrique sans dégrader les performances du modèle.

Trois modèles de prévision ont ensuite été comparés afin de déterminer si l'amélioration de la structure du modèle en simulation conduit à une amélioration de la prévision des crues (chapitre 8). La construction du modèle de prévision a mis en exergue la difficulté d'assimiler le dernier débit observé pour mettre à jour les états initiaux d'un modèle à deux branches d'écoulement. Nous avons proposé une méthode basée sur le partage simulé des flux venant des deux branches de routage afin de mettre à jour les états du modèle GR5H et du modèle développé dans les travaux de cette thèse. Cette méthode est moins efficace que la méthode de mise à jour du modèle GRP. Cependant, **les résultats montrent que le modèle GR5H qui inclut les fonctions d'intensité a des performances plus élevées que GRP à partir de l'horizon 6 h selon la plupart des critères de performance**. Les différences sont très marquées pour les événements et les bassins qui ont des dynamiques très rapides (événements avec des temps de montée très courts et bassins avec des faibles temps de concentration). À très courte échéance, le modèle GRP a de meilleures performances sur les bassins de la moitié nord de la France. Cela s'explique par une meilleure assimilation du dernier débit observé à l'instant de prévision, qui a un impact plus fort sur les prévisions à très courte échéance. Le modèle GR5H a quant à lui des performances plus faibles que le modèle GRP, à part pour les horizons supérieurs à 24 h. Les différentes méthodes d'estimation des paramètres du nouveau modèle ont montré que l'utilisation d'un réservoir linéaire sur la branche directe permet une meilleure assimilation du dernier débit observé (chapitre 9). Par ailleurs, l'estimation des paramètres par échéance de prévision améliore les performances, particulièrement pour les premiers horizons car les valeurs de plusieurs paramètres varient très largement en prévision par rapport à leur valeur estimée en simulation.

Réponses aux questions de recherche

Les résultats obtenus dans cette thèse permettent d'apporter des éléments de réponse aux questions de recherche que nous avons formulées dans l'introduction générale.

Est-il possible et souhaitable d'améliorer l'efficacité et la généralité d'un modèle pluie-débit tout en le rendant plus cohérent avec les processus hydrologiques affectant la génération du ruissellement ?

Les hypothèses de modélisation que nous avons testées ont permis d'améliorer d'une part l'efficacité du modèle en crue et d'autre part sa généralité, c'est-à-dire sa capacité à simuler différents types de crue et de comportements des bassins versants. Avons-nous rendu le modèle plus cohérent avec les processus affectant la génération du ruissellement ? Le modèle que nous utilisons est un modèle empirique, c'est-à-dire qu'il n'a pas été développé à partir de formulations théoriques sur les écoulements de l'eau dans les différentes couches du sol. C'est un modèle à finalité prédictive (et non explicative, cf. Shmueli, 2010) qui a été construit en fonction de sa capacité à simuler le débit, notamment en extrapolation temporelle. Les hypothèses de modélisation que nous avons développées avaient pour but d'améliorer les simulations du modèle lorsque les bassins versants réagissent à de fortes intensités de pluie. Nous avons testé une hypothèse sur le calcul de la pluie efficace et une hypothèse sur la répartition temporelle de cette pluie efficace. Ces hypothèses sont exprimées sous forme d'équations empiriques. Les performances des modèles intégrant ces équations ont été évaluées en utilisant de nombreux indicateurs sur les caractéristiques des crues, des bassins et de la pluie. Ces hypothèses correspondent à des lois de comportement à l'échelle des bassins versants mais ne doivent toutefois pas être considérées comme des lois décrivant la génération du ruissellement à petite échelle lors de fortes intensités de pluie. Dans le jargon de la météorologie, on les qualifierait plutôt de « paramétrisation » de la variabilité sous-maille, temporelle et spatiale, de la réaction des bassins versants. Nous soutenons qu'à l'échelle du bassin versant, il n'est pas possible de proposer des lois théoriques sur la contribution de processus hydrologiques appartenant à des échelles spatio-temporelles fines au débit à l'exutoire des bassins versants. Ces lois ne seraient validables qu'en utilisant des données permettant de retracer les chemins de l'eau dans le bassin versant. Nous avons choisi d'adopter une voie différente pour l'amélioration des modèles de prévision des crues. Cette voie est un compromis entre le développement de modèles purement prédictifs, mais dont l'interprétation peut s'avérer complexe, et le développement de modèles théoriques qui ne nous semblent pas adaptés aux échelles spatio-temporelles auxquelles nous nous intéressons.

Peut-on encore améliorer la structure et l'estimation des paramètres d'un modèle de prévision des crues avec l'information disponible en prévision opérationnelle à l'échelle des bassins versants sans augmenter l'équifinalité sous-jacente ?

Nos travaux d'amélioration de la généralité du modèle pluie-débit GR5H ont nécessité une complexification de sa structure. Cette complexification a induit une complexification de l'estimation des paramètres. De plus, les paramètres des fonctions d'intensité sont difficiles à contraindre lorsqu'un calage mono-objectif est effectué à partir de la chronique de débit. Par ailleurs, le paramètre supplémentaire de la fonction d'échange – introduit dans les travaux de Le Moine (2008) pour permettre une variation saisonnière des échanges – induit lui aussi une complexité paramétrique importante car il interagit avec les autres paramètres

de la fonction de routage. Ces difficultés à contraindre certains paramètres peuvent être vues comme une manifestation de la limite de l'information disponible à l'échelle du bassin versant pour simuler le débit. Si la simulation du débit à partir de la pluie et de la température est un problème sous-informé ou mal-informé, il est nécessaire de réduire le nombre de degrés de liberté ou de contraindre le problème à partir d'objectifs différents. C'est ce qui nous a poussé à réduire le nombre de paramètres libres du modèle. Nos tests n'ayant pas donné lieu à une dégradation des performances, nous en déduisons qu'il est possible d'améliorer la généralité d'un modèle pluie-débit tout en limitant les problèmes d'identification des paramètres. De plus, les tests réalisés pour paramétrer le modèle à partir d'une fonction objectif intégrant une composante événementielle semblent prometteurs pour contraindre certains paramètres du modèle. Bien que cela n'ait pas été testé dans nos travaux, l'utilisation d'une méthode d'estimation des paramètres très orientée vers la simulation des crues risque d'induire une dégradation des simulations en basses eaux. Par conséquent, il n'est peut-être pas possible d'augmenter la généralité du modèle jusqu'à permettre une simulation performante pour des objectifs de modélisation très différents sans apport d'information nouvelle dans le calage des paramètres. La nature de l'information nécessaire pour combiner un objectif de simulation des crues avec un objectif de simulation des étiages peut toutefois être variée et intégrée dans un cadre multi-objectif, par exemple en utilisant diverses fonctions du débit à des échelles spatio-temporelles différentes (e.g. multi- pas de temps, échelle régionale...) ou encore en intégrant des données de piézométrie (e.g. Pelletier & Andréassian, 2022).

Quel est le compromis optimal entre la complexification de la structure, de la méthode d'estimation des paramètres et de la procédure d'assimilation de données selon l'échéance de prévision ?

Les résultats que nous avons présentés dans la deuxième partie de ce mémoire ont montré que la complexification de la structure mène à une complexification de l'estimation des paramètres. Les résultats de la troisième partie ont montré qu'il est nécessaire d'adapter le schéma d'assimilation du dernier débit observé lorsque la structure devient plus complexe (la complexité additionnelle est ici caractérisée par les deux branches de routage de GR5H par rapport à l'unique branche de routage de GRP). Enfin, les résultats du dernier chapitre ont montré qu'une adaptation de l'estimation des paramètres du modèle de simulation induit des comportements différents en prévision et qu'une méthode d'estimation des paramètres dépendant de l'horizon améliore les performances. En résumé, nous avons mis en évidence les liens étroits entre les trois piliers du modèle de prévision. Changer un des trois piliers nécessite l'adaptation des deux autres. La comparaison entre les différentes versions d'un modèle peut s'avérer complexe pour cette raison. Comment dépasser cette complexité dans le développement et l'amélioration d'un modèle de prévision ? Dans cette thèse, nous avons commencé par ce qui nous semblait le plus évident : l'amélioration de la structure du modèle. Nous avons ensuite ajouté l'analyse des paramètres puis la construction d'une méthode de mise à jour adaptée à la structure du modèle. Le principe que nous soutenons est que l'assimilation du dernier débit observé pour mettre à jour les états initiaux du modèle est une « béquille » qui permet de compenser les différents types d'erreurs associés à la prévision des crues. En réduisant une partie des erreurs du modèle, la nécessité d'utiliser cette « béquille » devrait diminuer. Cependant, les résultats de nos tests ont montré que l'assimilation de données reste très importante pour la qualité des prévisions à très courte échéance. Ce poids de l'assimilation dépend du bassin et des événements considérés car les améliorations de la structure restent utiles à très courte échéance lorsque les bassins ont des temps de réaction

très faibles. En résumé, le compromis optimal entre la complexification de la structure, de l'estimation des paramètres et de l'assimilation de données dépend non seulement de l'échéance mais aussi de la complexité du bassin modélisé.

Perspectives

Les travaux de recherche présentés dans ce mémoire se sont focalisés sur l'amélioration d'un seul modèle hydrologique. Une perspective intéressante de notre travail serait de tester les hypothèses de modélisation sur la réaction des bassins versants aux fortes intensités de pluie en utilisant d'autres modèles hydrologiques. Cela permettrait de statuer sur la généralité de nos hypothèses pour modéliser la variabilité sous maille, temporelle et spatiale, du comportement des bassins versants. En particulier, la pertinence de ces hypothèses en fonction de la discrétisation spatio-temporelle des modèles pourrait être explorée car les comportements que nous avons cherché à modéliser dépendent de l'échelle considérée. Les compromis entre le choix de la structure et l'estimation des paramètres pourraient aussi dépendre de cette échelle. L'étude de la dépendance spatiale de nos hypothèses de modélisation pourrait aussi nous amener à réfléchir à la régionalisation des paramètres du modèle. La régionalisation permet notamment d'effectuer des estimations du débit pour des points non jaugés. Ces estimations représentent un enjeu très important pour la prévision des crues en France. Des travaux sur le transfert de l'information des stations intermédiaires jaugées dans un contexte de prévision spatialisée sont en cours au sein de notre équipe de recherche.

Les résultats de nos recherches pourront avoir des implications sur la prévision opérationnelle des crues en France avec le modèle GRP. Le modèle développé dans cette thèse pourrait être intégré dans le logiciel utilisé par les SPC. Son intégration soulève toutefois plusieurs questions qu'il sera nécessaire d'explorer. Tout d'abord, nous avons vu que le nouveau modèle a des performances dégradées par rapport à GRP sur certains bassins versants. Nous pouvons donc imaginer que les deux modèles soient proposés aux services opérationnels, et que le plus performant soit utilisé en temps réel selon le bassin d'intérêt. Une perspective intéressante serait de combiner les deux modèles dans une approche multi-modèles en fonction de l'échéance de prévision. Ensuite, nos travaux pourraient contribuer à l'amélioration de l'anticipation des crues soudaines (cf. la démarche initiée dans le projet PICS sur la prévision immédiate des crues soudaines). Pour le vérifier, il faudrait tester la nouvelle approche de prévision au pas de temps infra-horaire et au sein d'une structure semi-distribuée. Enfin, nous avons exploré plusieurs méthodes d'estimation des paramètres du modèle, mais il reste à déterminer laquelle sera la mieux adaptée à la prévision opérationnelle (nombre de paramètres, calage en fonction de l'échéance, fonction objectif...).

L'évaluation de la qualité du modèle de prévision développé dans cette thèse a été faite dans un cadre déterministe. L'objectif principal de cette évaluation était de déterminer si les changements effectués dans la structure du modèle permettent une amélioration des prévisions par rapport au modèle de référence qui est utilisé dans les services de prévision des crues. Une question reste cependant en suspens : l'amélioration de la structure d'un modèle hydrologique permet-elle une diminution de l'incertitude prédictive? Autrement dit, nous n'avons pas déterminé si l'incertitude liée au modèle hydrologique a été diminuée au sein de l'incertitude prédictive totale. Pour répondre à cette question, il serait nécessaire d'utiliser une méthode de quantification des incertitudes prédictives dans notre cadre d'évaluation (la

méthode de post-traitement statistique QUOIQUE est généralement utilisée par les SPC ; Bourgin, 2014) ainsi que plusieurs produits de pluie utilisés en temps réel. Ce cadre d'évaluation soulève des questions scientifiques et techniques qui dépassent le cadre de cette thèse, comme les interactions entre l'assimilation de données et les méthodes de post-traitement statistique (e.g. Bourgin *et al.*, 2014). Par ailleurs, l'incertitude liée aux prévisions météorologiques est probablement trop élevée pour que l'amélioration de la structure du modèle puisse réduire de manière conséquente l'incertitude prédictive totale. La complexification de la structure proposée dans nos travaux n'est peut-être pas justifiée pour certains bassins et à certaines échéances au vu des gains de performance obtenus.

Nous avons discuté à plusieurs reprises de la recherche d'un compromis de modélisation adapté au niveau d'information disponible. Cette notion est très utilisée pour construire les modèles qui reposent sur des méthodes d'apprentissage statistique. La recherche sur l'application de ces modèles à l'hydrologie opérationnelle s'est fortement développée ces dernières années (e.g. Nevo *et al.*, 2022). Bien que leur comportement soit difficilement interprétable, il semblerait que leur utilisation permette des avancées significatives dans de nombreux cas d'application comme la régionalisation des paramètres des modèles hydrologiques. La valeur ajoutée des gains de performance obtenus dans nos travaux pourrait diminuer si ces modèles s'avéraient avoir une capacité prédictive plus élevée que nos modèles empiriques pour la prévision des crues. En attendant une telle conclusion, des perspectives s'ouvrent sur le couplage de ces méthodes avec des modèles hydrologiques classiques (e.g. Slater *et al.*, 2022), cela afin de combiner les bénéfices des deux approches pour améliorer le niveau prédictif des modèles de prévision en hydrologie.

Bibliographie

- Andréassian, V., Bourgin, F., Oudin, L., Mathevet, T., Perrin, C., Lerat, J., Coron, L. & Berthet, L. (2014). Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency. *Water Resources Research* **50**, 8356–8366, <https://doi.org/10.1002/2013wr014761>.
- Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J. & Berthet, L. (2012). All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes* **26**, 2206–2210, <https://doi.org/10.1002/hyp.9264>.
- Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C. & Loumagne, C. (2004). Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds. *Water Resources Research* **40**, W05209, <https://doi.org/10.1029/2003wr002854>.
- Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.H. & Valéry, A. (2009). HESS Opinions "Crash tests for a standardized evaluation of hydrological models". *Hydrology and Earth System Sciences* **13**, 1757–1764, <https://doi.org/10.5194/hess-13-1757-2009>.
- Andrieu, H., Moussa, R. & Kirstetter, P.E. (2021). The event-specific geomorphological instantaneous unit hydrograph (e-GIUH): The basin hydrological response characteristic of a flood event. *Journal of Hydrology* **603**, 127158, <https://doi.org/10.1016/j.jhydrol.2021.127158>.
- Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Liew, M.W.V., Kannan, N. & Jha, M.K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE* **55**, 1491–1508, <https://doi.org/10.13031/2013.42256>.
- Arulampalam, M., Maskell, S., Gordon, N. & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing* **50**, 174–188, <https://doi.org/10.1109/78.978374>.
- Astagneau, P.C., Bourgin, F., Andréassian, V. & Perrin, C. (2021a). When does a parsimonious model fail to simulate floods? Learning from the seasonality of model bias. *Hydrological Sciences Journal* **66**, 1288–1305, <https://doi.org/10.1080/02626667.2021.1923720>.
- Astagneau, P.C., Thirel, G., Delaigue, O., Guillaume, J.H.A., Parajka, J., Brauer, C.C., Viglione, A., Buytaert, W. & Beven, K.J. (2021b). Technical note: Hydrology modelling R packages – a unified analysis of models and practicalities from a user perspective. *Hydrology and Earth System Sciences* **25**, 3937–3973, <https://doi.org/10.5194/hess-25-3937-2021>.

- Barbu, A.L., Calvet, J.C., Mahfouf, J.F., Albergel, C. & Lafont, S. (2011). Assimilation of soil wetness index and leaf area index into the ISBA-a-gs land surface model: grassland case study. *Biogeosciences* **8**, 1971–1986, <https://doi.org/10.5194/bg-8-1971-2011>.
- Bell, V.A., Kay, A.L., Jones, R.G. & Moore, R.J. (2007). Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrology and Earth System Sciences* **11**, 532–549, <https://doi.org/10.5194/hess-11-532-2007>.
- Bellier, J. (2018). *Prévisions hydrologiques probabilistes dans un cadre multivarié : quels outils pour assurer fiabilité et cohérence spatio-temporelle ?* Thèse de doctorat, Université Grenoble Alpes, <https://tel.archives-ouvertes.fr/tel-01950725>.
- Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D. & Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling & Software* **40**, 1–20, <https://doi.org/10.1016/j.envsoft.2012.09.011>.
- Berghuijs, W.R., Sivapalan, M., Woods, R.A. & Savenije, H.H.G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research* **50**, 5638–5661, <https://doi.org/10.1002/2014WR015692>.
- Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments. Rapport technique Rep. RHO 7, SMHI, <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1456191&dsid=3856>.
- Berthet, L., Andréassian, V., Perrin, C. & Javelle, P. (2009). How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. *Hydrology and Earth System Sciences* **13**, 819–831, <https://doi.org/10.5194/hess-13-819-2009>.
- Berthet, L. (2010). *Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique*. Thèse de doctorat, Agro-ParisTech, Paris, <https://tel.archives-ouvertes.fr/pastel-00529652>.
- Berthet, L., Bourgin, F., Perrin, C., Viatgé, J., Marty, R. & Piotte, O. (2020). A crash-testing framework for predictive uncertainty assessment when forecasting high flows in an extrapolation context. *Hydrology and Earth System Sciences* **24**, 2017–2041, <https://doi.org/10.5194/hess-24-2017-2020>.
- Berthet, L., Valéry, A., Garçon, R., Marty, R., Moulin, L., Puygrenier, D., Piotte, O., Lay, M.L., Janet, B. & Duquesne, F. (2019). Cohérence des prévisions et place de l'expertise : les nouveaux défis pour la prévision des crues. *La Houille Blanche* **105**, 5–12, <https://doi.org/10.1051/lhb/2019001>.
- Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B. & Harding, R.J. (2011). The joint UK land environment simulator (JULES), model description – part 1: Energy and water fluxes. *Geoscientific Model Development* **4**, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>.

- Beven, K. & Kirby, M.J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* **24**, 43–69, <https://doi.org/10.1080/02626667909491834>.
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology* **320**, 18–36.
- Beven, K. (2016). Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal* **61**, 1652–1665, <https://doi.org/10.1080/02626667.2015.1031761>.
- Beven, K. (2021). The era of infiltration. *Hydrology and Earth System Sciences* **25**, 851–866, <https://doi.org/10.5194/hess-25-851-2021>.
- Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., Dadson, S., Everard, N., Fowler, H.J., Freer, J., Hannah, D.M., Heppell, K., Holden, J., Lamb, R., Lewis, H., Morgan, G., Parry, L. & Wagener, T. (2019). Developing observational methods to drive future hydrological science: Can we make a start as a community? *Hydrological Processes* **34**, 868–873, <https://doi.org/10.1002/hyp.13622>.
- Beven, K. & Freer, J. (2001a). A dynamic TOPMODEL. *Hydrological Processes* **15**, 1993–2011, <https://doi.org/10.1002/hyp.252>.
- Beven, K. & Freer, J. (2001b). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**, 11–29, [https://doi.org/10.1016/s0022-1694\(01\)00421-8](https://doi.org/10.1016/s0022-1694(01)00421-8).
- Beven, K.J. (1984). Infiltration into a class of vertically non-uniform soils. *Hydrological Sciences Journal* **29**, 425–434, <https://doi.org/10.1080/02626668409490960>.
- Beven, K.J. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* **16**, 41–51, [https://doi.org/10.1016/0309-1708\(93\)90028-E](https://doi.org/10.1016/0309-1708(93)90028-E).
- Beven, K.J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *WIREs Water* **5**, <https://doi.org/10.1002/wat2.1278>.
- Beven, K.J., Kirkby, M.J., Freer, J.E. & Lamb, R. (2021). A history of TOPMODEL. *Hydrology and Earth System Sciences* **25**, 527–549, <https://doi.org/10.5194/hess-25-527-2021>.
- Birkel, C., Soulsby, C. & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research* **50**, 3481–3501, <https://doi.org/10.1002/2013WR014925>.
- Blöschl, G. & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes* **9**, 251–290, <https://doi.org/10.1002/hyp.3360090305>.
- Blöschl, G. (2017). Debates-hypothesis testing in hydrology: Introduction. *Water Resources Research* **53**, 1767–1769, <https://doi.org/10.1002/2017WR020584>.
- Blöschl, G., Bierkens, M.F., Chambel, A., Cudenneq, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B. & Viglione, Alberto, e. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal* **64**, 1141–1158, <https://doi.org/10.1080/02626667.2019.1620507>.

- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. & Savenije, H. (eds.) (2013). *Runoff Prediction in Ungauged Basins*. Cambridge University Press, <https://doi.org/10.1017/CBO9781139235761>.
- Borah, D.K., Arnold, J.G., Bera, M., Krug, E.C. & Liang, X.Z. (2007). Storm event and continuous hydrologic modeling for comprehensive and efficient watershed simulations. *Journal of Hydrologic Engineering* **12**, 605–616, [https://doi.org/10.1061/\(asce\)1084-0699\(2007\)12:6\(605\)](https://doi.org/10.1061/(asce)1084-0699(2007)12:6(605)).
- Bouaziz, L.J.E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C.C., Niel, J.D., Dewals, B.J., Drogue, G., Grelier, B., Melsen, L.A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A.H., Willems, P., Savenije, H.H.G. & Hrachowitz, M. (2021). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences* **25**, 1069–1095, <https://doi.org/10.5194/hess-25-1069-2021>.
- Boudevillain, B., Delrieu, G., Galabertier, B., Bonnifait, L., Bouilloud, L., Kirstetter, P.E. & Mosini, M.L. (2011). The Cévennes-Vivarais Mediterranean hydrometeorological observatory database. *Water Resources Research* **47**, W07701, <https://doi.org/10.1029/2010wr010353>.
- Bourgin, F., Ramos, M., Thirel, G. & Andréassian, V. (2014). Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. *Journal of Hydrology* **519**, 2775–2784, <https://doi.org/10.1016/j.jhydrol.2014.07.054>.
- Bourgin, F. (2014). *Comment quantifier l'incertitude prédictive en modélisation hydrologique ? Travail exploratoire sur un grand échantillon de bassins versants*. Thèse de doctorat, Irstea Antony, AgroParisTech, <https://pastel.archives-ouvertes.fr/tel-01130084>.
- Bradley, A.A., Demargne, J. & Franz, K.J. (2016). Attributes of forecast quality. *Handbook of Hydrometeorological Ensemble Forecasting*, p. 1–44, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-40457-3_2-1.
- Braud, I., Ayrat, P.A., Bouvier, C., Branger, F., Delrieu, G., Coz, J.L., Nord, G., Vandervaere, J.P., Anquetin, S., Adamovic, M., Andrieu, J., Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., Didon-Lescot, J.F., Domergue, J.M., Douvinet, J., Dramais, G., Freydier, R., Gérard, S., Huza, J., Leblois, E., Bourgeois, O.L., Boursicaud, R.L., Marchand, P., Martin, P., Nottale, L., Patris, N., Renard, B., Seidel, J.L., Taupin, J.D., Vannier, O., Vincendon, B. & Wijbrans, A. (2014). Multi-scale hydrometeorological observation and modelling for flash flood understanding. *Hydrology and Earth System Sciences* **18**, 3733–3761, <https://doi.org/10.5194/hess-18-3733-2014>.
- Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Brunner, M.I., Melsen, L.A., Wood, A.W., Rakovec, O., Mizukami, N., Knoben, W.J.M. & Clark, M.P. (2021a). Flood spatial coherence, triggers, and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated models. *Hydrology and Earth System Sciences* **25**, 105–119, <https://doi.org/10.5194/hess-25-105-2021>.

- Brunner, M.I., Slater, L., Tallaksen, L.M. & Clark, M. (2021b). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water* **8**, <https://doi.org/10.1002/wat2.1520>.
- Burnash, R.J.C. (1995). The NWS River Forecast System – Catchment Modeling. *Computer models of watershed hydrology* (ed. V.P. Singh), p. 311–366, Water Resources Publications, Colorado, <https://www.wrpllc.com/books/cmwhn.html>.
- Butts, M.B., Payne, J.T., Kristensen, M. & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* **298**, 242–266, <https://doi.org/10.1016/j.jhydrol.2004.03.042>.
- Carrega, P. & Michelot, N. (2021). Une catastrophe hors norme d'origine météorologique le 2 octobre 2020 dans les montagnes des Alpes-Maritimes. *Physio-Géo* p. 1–70, <https://doi.org/10.4000/physio-geo.12370>.
- Carsell, K.M., Pingel, N.D. & Ford, D.T. (2004). Quantifying the benefit of a flood warning system. *Natural Hazards Review* **5**, 131–140, [https://doi.org/10.1061/\(asce\)1527-6988\(2004\)5:3\(131\)](https://doi.org/10.1061/(asce)1527-6988(2004)5:3(131)).
- Chappell, N.A., Jones, T.D., Tych, W. & Krishnaswamy, J. (2017). Role of rainstorm intensity underestimated by data-derived flood models: Emerging global evidence from subsurface-dominated watersheds. *Environmental Modelling & Software* **88**, 1–9, <https://doi.org/10.1016/j.envsoft.2016.10.009>.
- Clark, M.P., Kavetski, D. & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research* **47**, W09301, <https://doi.org/10.1029/2010WR009827>.
- Clark, M.P., Schaefli, B., Schymanski, S.J., Samaniego, L., Luce, C.H., Jackson, B.M., Freer, J.E., Arnold, J.R., Moore, R.D., Istanbuloglu, E. & Ceola, S. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research* **52**, 2350–2365, <https://doi.org/10.1002/2015wr017910>.
- Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T. & Hay, L.E. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* **44**, W00B02, <https://doi.org/10.1029/2007WR006735>.
- 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**, W05552, <https://doi.org/10.1029/2011WR011721>.
- Coron, L., Thirel, G., Delaigue, O., Perrin, C. & Andréassian, V. (2017). The Suite of Lumped GR Hydrological Models in an R package. *Environmental Modelling and Software* **94**, 166–171, <https://doi.org/10.1016/j.envsoft.2017.05.002>.
- Coron, L. (2013). *Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution?* Thèse de doctorat, Irstea Antony, AgroParisTech, <https://tel.archives-ouvertes.fr/pastel-00879090/>.

- Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C. & Michel, C. (2020). *airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling*. R package version 1.6.9.27. <https://cran.r-project.org/web/packages/airGR/index.html>.
- Coustau, M., Rousset-Regimbeau, F., Thirel, G., Habets, F., Janet, B., Martin, E., de Saint-Aubin, C. & Soubeyroux, J.M. (2015). Impact of improved meteorological forcing, profile of soil hydraulic conductivity and data assimilation on an operational hydrological ensemble forecast system over france. *Journal of Hydrology* **525**, 781–792, <https://doi.org/10.1016/j.jhydrol.2015.04.022>.
- CRED (2022). Disasters in numbers. Rapport technique, Center for research on the Epidemiology of Disasters, CRED, Brussels, <https://www.emdat.be/publications>.
- Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S.P., Grimaldi, S., Gupta, H. & Paturel, J.E. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal* **60**, 402–423, <https://doi.org/10.1080/02626667.2014.903331>.
- de Boer-Euser, T., Bouaziz, L., Niel, J.D., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G. & Willems, P. (2017). Looking beyond general metrics for model comparison – lessons from an international model intercomparison study. *Hydrology and Earth System Sciences* **21**, 423–440, <https://doi.org/10.5194/hess-21-423-2017>.
- de Lavenne, A., Andréassian, V., Thirel, G., Ramos, M.H. & Perrin, C. (2019). A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research* **55**, 8821–8839, <https://doi.org/10.1029/2018WR024266>.
- De Roo, A. (1999). LISFLOOD: a rainfall-runoff model for large riverbasins to assess the influence of land use changes on flood risk. *Ribamod: river basin modelling, management and flood mitigation* (ed. P. Balabanis), European Commission.
- Delaigue, O., Génot, B., Lebecherel, L., Brigode, P. & Bourgin, P. (2020). Database of watershed-scale hydroclimatic observations in France. <https://webgr.inrae.fr/base-de-donnees>.
- Dembélé, M., Hrachowitz, M., Savenije, H.H.G., Mariéthoz, G. & Schaeffli, B. (2020). Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets. *Water Resources Research* **56**, e2019WR026085, <https://doi.org/10.1029/2019wr026085>.
- Donnelly, C., Andersson, J.C. & Arheimer, B. (2015). Using flow signatures and catchment similarities to evaluate the e-HYPE multi-basin model across europe. *Hydrological Sciences Journal* **61**, 255–273, <https://doi.org/10.1080/02626667.2015.1027710>.
- Douinot, A., Roux, H., Garambois, P.A. & Dartus, D. (2018). Using a multi-hypothesis framework to improve the understanding of flow dynamics during flash floods. *Hydrology and Earth System Sciences* **22**, 5317–5340, <https://doi.org/10.5194/hess-22-5317-2018>.
- Douinot, A., Roux, H., Garambois, P.A., Larnier, K., Labat, D. & Dartus, D. (2016). Accounting for rainfall systematic spatial variability in flash flood forecasting. *Journal of Hydrology* **541**, 359–370, <https://doi.org/10.1016/j.jhydrol.2015.08.024>.

- Douville, H., Raghavan, K., Renwick, J., Allan, R.P., Arias, P.A., Barlow, M., Cerezo-Mota, R., Cherchi, A., Gan, T.Y., Gergis, J., Jiang, D., Khan, A., Mba, W.P., Rosenfeld, D., Tierney, J. & Zolina, O. (2021). Water cycle changes. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, p. 1055–1210, Cambridge University Press, <https://www.ipcc.ch/report/ar6/wg1>.
- Ducharne, A. (2009). Reducing scale dependence in TOPMODEL using a dimensionless topographic index. *Hydrology and Earth System Sciences* **13**, 2399–2412, <https://doi.org/10.5194/hess-13-2399-2009>.
- Efstratiadis, A. & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal* **55**, 58–78, <https://doi.org/10.1080/02626660903526292>.
- Ercolani, G. & Castelli, F. (2017). Variational assimilation of streamflow data in distributed flood forecasting. *Water Resources Research* **53**, 158–183, <https://doi.org/10.1002/2016WR019208>.
- Estrany, J., Garcia, C. & Batalla, R.J. (2010). Hydrological response of a small Mediterranean agricultural catchment. *Journal of Hydrology* **380**, 180–190, <https://doi.org/10.1016/j.jhydrol.2009.10.035>.
- Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. & Savenije, H.H.G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences* **17**, 1893–1912, <https://doi.org/10.5194/hess-17-1893-2013>.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99**, 10143, <https://doi.org/10.1029/94jc00572>.
- Fenicia, F. & McDonnell, J.J. (2022). Modeling streamflow variability at the regional scale: (1) perceptual model development through signature analysis. *Journal of Hydrology* **605**, 127287, <https://doi.org/10.1016/j.jhydrol.2021.127287>.
- Fenicia, F., Meißner, D. & McDonnell, J. (2022). Modeling streamflow variability at the regional scale: (2) development of a bespoke distributed conceptual model. *Journal of Hydrology* **605**, 127286, <https://doi.org/10.1016/j.jhydrol.2021.127286>.
- Ficchi, A. (2017). *An adaptive hydrological model for multiple time-steps: Diagnostics and improvements based on fluxes consistency*. Thèse de doctorat, UPMC, Irstea Antony, <https://hal.archives-ouvertes.fr/tel-01619102v1>.
- Ficchi, A., Perrin, C. & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *Journal of Hydrology* **538**, 454–470, <https://doi.org/10.1016/j.jhydrol.2016.04.016>.
- Ficchi, A., Perrin, C. & Andréassian, V. (2019). Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching. *Journal of Hydrology* **575**, 1308–1327, <https://doi.org/10.1016/j.jhydrol.2019.05.084>.

- Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L. & Peterson, T.J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research* **52**, 1820–1846, <https://doi.org/10.1002/2015wr018068>.
- Garambois, P., Larnier, K., Roux, H., Labat, D. & Dartus, D. (2014). Analysis of flash flood-triggering rainfall for a process-oriented hydrological model. *Atmospheric Research* **137**, 14–24, <https://doi.org/10.1016/j.atmosres.2013.09.016>.
- Garavaglia, F., Lay, M.L., Gottardi, F., Garçon, R., Gailhard, J., Paquet, E. & Mathevet, T. (2017). Impact of model structure on flow simulation and hydrological realism: from a lumped to a semi-distributed approach. *Hydrology and Earth System Sciences* **21**, 3937–3952, <https://doi.org/10.5194/hess-21-3937-2017>.
- 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* **82**, 71–76, <https://doi.org/10.1051/lhb/1996056>.
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D. & Viglione, A. (2009). A compilation of data on European flash floods. *Journal of Hydrology* **367**, 70–78, <https://doi.org/10.1016/j.jhydrol.2008.12.028>.
- Gaume, E. & Bouvier, C. (2004). Analyse hydro-pluviométrique des crues du Gard et du Vidourle des 8 et 9 septembre 2002. *La Houille Blanche* **90**, 99–106, <https://doi.org/10.1051/lhb:200406014>.
- Gérard, L. (2010). Sensibilité des performances d'un modèle de prévision des crues au critère de calage. Rapport de stage, Cemagref Antony, INP Toulouse, <https://webgr.inrae.fr/wp-content/uploads/2012/07/2010-MASTER-GERARD.pdf>.
- Giani, G., Rico-Ramirez, M. & Woods, R.A. (2022). Are moments of rainfall spatial variability useful for runoff modelling in operational hydrology? *Hydrological Sciences Journal* <https://doi.org/10.1080/02626667.2022.2092405>.
- Gnann, S.J., Howden, N.J.K. & Woods, R.A. (2020). Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. *Hydrology and Earth System Sciences* **24**, 561–580, <https://doi.org/10.5194/hess-24-561-2020>.
- Goullet, J. (2016). *A la recherche des résolutions spatiales et temporelles caractéristiques du comportement hydrologique d'un bassin versant*. Mémoire de master, UPMC, Irstea, Antony, <https://webgr.inrae.fr/wp-content/uploads/2017/06/an2016-pub00053045.pdf>.
- Green, W.H. & Ampt, G.A. (1911). Studies on soil physics. *The Journal of Agricultural Science* **4**, 1–24, <https://doi.org/10.1017/S0021859600001441>.
- Grimaldi, S., Nardi, F., Piscopia, R., Petroselli, A. & Apollonio, C. (2021). Continuous hydrologic modelling for design simulation in small and ungauged basins: A step forward and some tests for its practical use. *Journal of Hydrology* **595**, 125664, <https://doi.org/10.1016/j.jhydrol.2020.125664>.

- Gupta, H.V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. & Andréassian, V. (2014). Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences* **18**, 463–477, <https://doi.org/10.5194/hess-18-463-2014>.
- Gupta, H.V., Kling, H., Yilmaz, K.K. & Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Gupta, H.V., Wagener, T. & Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22**, 3802–3813, <https://doi.org/10.1002/hyp.6989>.
- Gustard, A., Bullock, A. & Dixon, J. (1992). *Low flow estimation in the United Kingdom*. Institute of Hydrology, <https://nora.nerc.ac.uk/id/eprint/6050>.
- Hallegatte, S. (2012). *A Cost Effective Solution to Reduce Disaster Losses in Developing Countries: Hydro-Meteorological Services, Early Warning, and Evacuation*. The World Bank, <https://openknowledge.worldbank.org/handle/10986/9359>.
- Hapuarachchi, H.A.P., Wang, Q.J. & Pagano, T.C. (2011). A review of advances in flash flood forecasting. *Hydrological Processes* **25**, 2771–2784, <https://doi.org/10.1002/hyp.8040>.
- Hapuarachchi, H.A.P., Bari, M.A., Kabir, A., Hasan, M.M., Woldemeskel, F.M., Gamage, N., Sunter, P.D., Zhang, X.S., Robertson, D.E., Bennett, J.C. & Feikema, P.M. (2022). Development of a national 7-day ensemble streamflow forecasting service for Australia. *Hydrology and Earth System Sciences* **26**, 4801–4821, <https://doi.org/10.5194/hess-26-4801-2022>.
- Herman, J.D., Reed, P.M. & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research* **49**, 1400–1414, <https://doi.org/10.1002/wrcr.20124>.
- Horner, I., Renard, B., Coz, J.L., Branger, F., McMillan, H.K. & Pierrefeu, G. (2018). Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research* **54**, 1952–1976, <https://doi.org/10.1002/2017wr022039>.
- Horton, R.E. (1933). The Role of infiltration in the hydrologic cycle. *Transactions, American Geophysical Union* **14**, 446, <https://doi.org/10.1029/TR014i001p00446>.
- Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H.H.G. & Gascuel-Oudou, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research* **50**, 7445–7469, <https://doi.org/10.1002/2014wr015484>.
- Hrachowitz, M. & Clark, M.P. (2017). HESS opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences* **21**, 3953–3973, <https://doi.org/10.5194/hess-21-3953-2017>.
- Hugenschmidt, C., Ingwersen, J., Sangchan, W., Sukvanachaikul, Y., Duffner, A., Uhlenbrook, S. & Streck, T. (2014). A three-component hydrograph separation based on geochemical tracers in a tropical mountainous headwater catchment in northern Thailand. *Hydrology and Earth System Sciences* **18**, 525–537, <https://doi.org/10.5194/hess-18-525-2014>.

- Jain, S.K., Mani, P., Jain, S.K., Prakash, P., Singh, V.P., Tullos, D., Kumar, S., Agarwal, S.P. & Dimri, A.P. (2018). A brief review of flood forecasting techniques and their applications. *International Journal of River Basin Management* **16**, 329–344, <https://doi.org/10.1080/15715124.2017.1411920>.
- Jakeman, A.J. & Hornberger, G.M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* **29**, 2637–2649, <https://doi.org/10.1029/93WR00877>.
- Javelle, P., Demargne, J., Defrance, D., Pansu, J. & Arnaud, P. (2014). Evaluating flash-flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system. *Hydrological Sciences Journal* **59**, 1390–1402, <https://doi.org/10.1080/02626667.2014.923970>.
- Javelle, P., Fouchier, C., Arnaud, P. & Lavabre, J. (2010). Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *Journal of Hydrology* **394**, 267–274, <https://doi.org/10.1016/j.jhydrol.2010.03.032>.
- Javelle, P., Vinet, C.S.M.F. & Payrastre, O. (2019). Towards real time assessment of flood risk damage : an application of the AIGA method in the south of France. *Geophysical Research Abstracts* (ed. EGU), vol. 21, <https://meetingorganizer.copernicus.org/EGU2019/EGU2019-15204.pdf>.
- Jay-Allemand, M. (2020). *Estimation variationnelle des paramètres d'un modèle hydrologique distribué*. Thèse de doctorat, Université d'Aix-Marseille, <http://www.theses.fr/2020AIXM0400>.
- Jay-Allemand, M., Javelle, P., Gejadze, I., Arnaud, P., Malaterre, P.O., Fine, J.A. & Orlande, D. (2020). On the potential of variational calibration for a fully distributed hydrological model: application on a mediterranean catchment. *Hydrology and Earth System Sciences* **24**, 5519–5538, <https://doi.org/10.5194/hess-24-5519-2020>.
- Jeuland, M., Hansen, K., Doherty, H., Eastman, L.B. & Tchamkina, M. (2019). The economic impacts of water information systems: A systematic review. *Water Resources and Economics* **26**, 100128, <https://doi.org/10.1016/j.wre.2018.09.001>.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82**, 35–45, <https://doi.org/10.1115/1.3662552>.
- Kavetski, D. & Clark, M.P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. impact of time stepping schemes on model analysis and prediction. *Water Resources Research* **46**, W10511, <https://doi.org/10.1029/2009wr008896>.
- Khatami, S., Peel, M.C., Peterson, T.J. & Western, A.W. (2019). Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resources Research* **55**, 8922–8941, <https://doi.org/10.1029/2018WR023750>.
- Kim, H. & Lee, S. (2013). Assessment of a seasonal calibration technique using multiple objectives in rainfall-runoff analysis. *Hydrological Processes* **28**, 2159–2173, <https://doi.org/10.1002/hyp.9785>.

- Kirchner, J.W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* **42**, W03S04, <https://doi.org/10.1029/2005WR004362>.
- Kitanidis, P.K. & Bras, R.L. (1980). Real-time forecasting with a conceptual hydrologic model: 2. applications and results. *Water Resources Research* **16**, 1034–1044, <https://doi.org/10.1029/wr016i006p01034>.
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal* **31**, 13–24, <https://doi.org/10.1080/02626668609491024>.
- Klemeš, V. (2000). Common sense and other heresies: Selected papers on hydrology and water resources engineering. *Canadian Water Resources Journal* **25**, 108–108, <https://doi.org/10.4296/cwrj2501108>.
- Knoben, W.J.M., Freer, J.E., Peel, M.C., Fowler, K.J.A. & Woods, R.A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research* **56**, e2019WR025975, <https://doi.org/10.1029/2019wr025975>.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences* **22**, 6005–6022, <https://doi.org/10.5194/hess-22-6005-2018>.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S. & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences* **23**, 5089–5110, <https://doi.org/10.5194/hess-23-5089-2019>.
- Kuppel, S., Tetzlaff, D., Maneta, M.P. & Soulsby, C. (2018). EcH2O-iso 1.0: water isotopes and age tracking in a process-based, distributed ecohydrological model. *Geoscientific Model Development* **11**, 3045–3069, <https://doi.org/10.5194/gmd-11-3045-2018>.
- Laloy, E. & Vrugt, J.A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAMZS and high-performance computing. *Water Resources Research* **48**, W01526, <https://doi.org/10.1029/2011wr010608>.
- Lana-Renault, N., Latron, J. & Regüés, D. (2007). Streamflow response and water-table dynamics in a sub-Mediterranean research catchment (Central Pyrenees). *Journal of Hydrology* **347**, 497–507, <https://doi.org/10.1016/j.jhydrol.2007.09.037>.
- Lane, R.A., Coxon, G., Freer, J.E., Wagener, T., Johnes, P.J., Bloomfield, J.P., Greene, S., Macleod, C.J.A. & Reaney, S.M. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences* **23**, 4011–4032, <https://doi.org/10.5194/hess-23-4011-2019>.
- Largerón, C., Cloke, H., Verhoef, A., Martínez-de-la Torre, A. & Mueller-Quintino, A. (2018). Impact of the representation of the infiltration on the river flow during intense rainfall events in Jules. Rapport technique, ECMWF, <https://www.ecmwf.int/en/elibrary/18245-impact-representation-infiltration-river-flow-during-intense-rainfall-events-jules>.

- Latron, J. & Gallart, F. (2008). Runoff generation processes in a small Mediterranean research catchment (Vallcebre, Eastern Pyrenees). *Journal of Hydrology* **358**, 206–220, <https://doi.org/10.1016/j.jhydrol.2008.06.014>.
- Le Mesnil, M. (2021). *Signatures Hydrologiques des Bassins Karstiques*. Thèse de doctorat, Montpellier SupAgro, <https://tel.archives-ouvertes.fr/tel-03578569>.
- Le Moine, N. (2008). *Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?* Thèse de doctorat, UPMC, Cemagref, <https://hal.inrae.fr/tel-02591478>.
- Le Moine, N., Andréassian, V., Perrin, C. & Michel, C. (2007). How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. *Water Resources Research* **43**, W06428, <https://doi.org/10.1029/2006WR005608>.
- Ledoux, E., Girard, G., Marsily, G., Villeneuve, J.P. & Deschenes, J. (1989). Spatially distributed modeling: Conceptual approach, coupling surface water and groundwater. *Unsaturated Flow in Hydrologic Modeling*, p. 435–454, Springer Netherlands.
- Leleu, I., Tonnelier, I., Puechberty, R., Gouin, P., Viquendi, I., Cobos, L., Foray, A., Baillon, M. & Ndima, P.O. (2014). La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques. *La Houille Blanche* **100**, 25–32, <https://doi.org/10.1051/lhb/2014004>.
- Lerat, J., Perrin, C., Andréassian, V., Loumagne, C. & Ribstein, P. (2012). Towards robust methods to couple lumped rainfall-runoff models and hydraulic models: A sensitivity analysis on the Illinois River. *Journal of Hydrology* **418-419**, 123–135, <https://doi.org/10.1016/j.jhydrol.2009.09.019>.
- Li, H., Luo, L., Wood, E.F. & Schaake, J. (2009). The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research* **114**, <https://doi.org/10.1029/2008jd010969>.
- Li, M., Wang, Q.J., Bennett, J.C. & Robertson, D.E. (2015). A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts. *Hydrology and Earth System Sciences* **19**, 1–15, <https://doi.org/10.5194/hess-19-1-2015>.
- Li, Y., Grimaldi, S., Pauwels, V.R. & Walker, J.P. (2018). Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *Journal of Hydrology* **557**, 897–909, <https://doi.org/10.1016/j.jhydrol.2018.01.013>.
- Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R News* **2**, 18-22, https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf.
- Liu, Y., Weerts, A.H., Clark, M., Franssen, H.J.H., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O. & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences* **16**, 3863–3887, <https://doi.org/10.5194/hess-16-3863-2012>.

- Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P. & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences* **18**, 575–594, <https://doi.org/10.5194/hess-18-575-2014>.
- Lobligeois, F. (2014). *Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues ?* Thèse de doctorat, UPMC, AgroParisTech, Paris, <https://hal.archives-ouvertes.fr/tel-01134990>.
- Loritz, R., Hrachowitz, M., Neuper, M. & Zehe, E. (2021). The role and value of distributed precipitation data in hydrological models. *Hydrology and Earth System Sciences* **25**, 147–167, <https://doi.org/10.5194/hess-25-147-2021>.
- Manus, C., Anquetin, S., Braud, I., Vandervaere, J.P., Creutin, J.D., Viallet, P. & Gaume, E. (2009). A modeling approach to assess the hydrological response of small Mediterranean catchments to the variability of soil characteristics in a context of extreme events. *Hydrology and Earth System Sciences* **13**, 79–97, <https://doi.org/10.5194/hess-13-79-2009>.
- Marty, R., Faucard, Y., Hans, P.A., Reinbold, D. & Berthet, L. (2019). Se préparer pour une meilleure anticipation des crues, premier bilan des actions entreprises par le service de prévision des crues Loire-Cher-Indre. *La Houille Blanche* **105**, 26–32, <https://doi.org/10.1051/lhb/2019004>.
- Massmann, C. (2019). Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments. *Hydrological Processes* **34**, 4–20, <https://doi.org/10.1002/hyp.13566>.
- Mathevet, T. (2005). *Quels modèles pluie-débit globaux pour le pas de temps horaire ? Développement empirique et comparaison de modèles sur un large échantillon de bassins versants*. Thèse de doctorat, CEMAGREF (Antony), ENGREF (Paris), France, <https://hal.inrae.fr/tel-02587642>.
- Mathevet, T., Michel, C., Andréassian, V. & Perrin, C. (2006). A bounded version of the nash-sutcliffe criterion for better model assessment on large sets of basins. *IAHS Publication* **307**, 211–219, <https://hal.inrae.fr/hal-02588691>.
- Mathevet, T., Gupta, H., Perrin, C., Andréassian, V. & Le Moine, N. (2020). Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds. *Journal of Hydrology* **585**, 124698, <https://doi.org/10.1016/j.jhydrol.2020.124698>.
- McDonnell, J.J., Spence, C., Karran, D.J., van Meerveld, H.J.I. & Harman, C.J. (2021). Fill-and-spill: A process description of runoff generation at the scale of the beholder. *Water Resources Research* **57**, e2020WR027514, <https://doi.org/10.1029/2020wr027514>.
- McKay, M.D., Beckman, R.J. & Conover, W.J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245, <https://doi.org/10.1080/00401706.1979.10489755>.
- McMillan, H., Westerberg, I. & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes* **31**, 4757–4761, <https://doi.org/10.1002/hyp.11300>.

- McMillan, H., Booker, D. & Cattoën, C. (2016). Validation of a national hydrological model. *Journal of Hydrology* **541**, 800–815, <https://doi.org/10.1016/j.jhydrol.2016.07.043>.
- Melsen, L.A., Addor, N., Mizukami, N., Newman, A.J., Torfs, P.J.J.F., Clark, M.P., Uijlenhoet, R. & Teuling, A.J. (2018). Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences* **22**, 1775–1791, <https://doi.org/10.5194/hess-22-1775-2018>.
- Merz, R., Blöschl, G. & Parajka, J. (2006). Spatio-temporal variability of event runoff coefficients. *Journal of Hydrology* **331**, 591–604, <https://doi.org/10.1016/j.jhydrol.2006.06.008>.
- Michel, C. (1987). *Hydrologie appliquée aux petits bassins ruraux*. Cemagref, Antony, France, https://belinra.inrae.fr/index.php?lvl=notice_display&id=225112.
- Miller, R.N., Ghil, M. & Gauthiez, F. (1994). Advanced data assimilation in strongly non-linear dynamical systems. *Journal of the Atmospheric Sciences* **51**, 1037–1056, [https://doi.org/10.1175/1520-0469\(1994\)051<1037:adain>2.0.co;2](https://doi.org/10.1175/1520-0469(1994)051<1037:adain>2.0.co;2).
- Mishra, S.K., Jain, M.K. & Singh, V.P. (2004). Evaluation of the SCS-CN-based model incorporating antecedent moisture. *Water Resources Management* **18**, 567–589, <https://doi.org/10.1007/s11269-004-8765-1>.
- Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H.V. & Kumar, R. (2019). On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrology and Earth System Sciences* **23**, 2601–2614, <https://doi.org/10.5194/hess-23-2601-2019>.
- Monteil, C., Zaoui, F., Le Moine, N. & Hendrickx, F. (2020). Multi-objective calibration by combination of stochastic and gradient-like parameter generation rules – the caRamel algorithm. *Hydrology and Earth System Sciences* **24**, 3189–3209, <https://doi.org/10.5194/hess-24-3189-2020>.
- Moradkhani, H., Sorooshian, S., Gupta, H.V. & Houser, P.R. (2005). Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources* **28**, 135–147, <https://doi.org/10.1016/j.advwatres.2004.09.002>.
- Moussa, R. & Chahinian, N. (2009). Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. *Hydrology and Earth System Sciences* **13**, 519–535, <https://doi.org/10.5194/hess-13-519-2009>.
- Muleta, M.K. (2012). Improving model performance using season-based evaluation. *Journal of Hydrologic Engineering* **17**, 191–200, [https://doi.org/10.1061/\(asce\)he.1943-5584.0000421](https://doi.org/10.1061/(asce)he.1943-5584.0000421).
- Nash, J.E. & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Nevo, S., Morin, E., Rosenthal, A.G., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Levi, N.P., Reich, O., Gilon, O., Maor, R.,

- Timnat, S., Shechter, T., Anisimov, V., Gigi, Y., Levin, Y., Moshe, Z., Ben-Haim, Z., Hassidim, A. & Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences* **26**, 4013–4032, <https://doi.org/10.5194/hess-26-4013-2022>.
- Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Lay, M.L., Besson, F., Soubeyroux, J.M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augereau, B. & Morice, E. (2014). Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. *Hydrology and Earth System Sciences* **18**, 2829–2857, <https://doi.org/10.5194/hess-18-2829-2014>.
- Nijzink, R.C., Almeida, S., Pechlivanidis, I.G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Nooijen, R.R.P., Savenije, H.H.G. & Hrachowitz, M. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research* **54**, 8332–8362, <https://doi.org/10.1029/2017wr021895>.
- Oudin, L., Andréassian, V., Mathevet, T., Perrin, C. & Michel, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research* **42**, W07410, <https://doi.org/10.1029/2005wr004636>.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. & Loumagne, C. (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**, 290–306, <https://doi.org/10.1016/j.jhydrol.2004.08.026>.
- Pagano, T.C., Wood, A.W., Ramos, M.H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S. & Verkade, J.S. (2014). Challenges of Operational River Forecasting. *Journal of Hydrometeorology* **15**, 1692–1707, <https://doi.org/10.1175/JHM-D-13-0188.1>.
- Pang, S., Wang, X., Melching, C.S. & Feger, K.H. (2020). Development and testing of a modified SWAT model based on slope condition and precipitation intensity. *Journal of Hydrology* **588**, 125098, <https://doi.org/10.1016/j.jhydrol.2020.125098>.
- Pappenberger, F., Pagano, T.C., Brown, J.D., Alfieri, L. & Lavers, D. A., e.a. (2016). Hydrological Ensemble Prediction Systems Around the Globe. *Handbook of Hydrometeorological Ensemble Forecasting* (eds. Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H.L. Cloke & J.C. Schaake), p. 1–35, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-40457-3_47-1.
- Parajka, J., Merz, R. & Blöschl, G. (2007). Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. *Hydrological Processes* **21**, 435–446, <https://doi.org/10.1002/hyp.6253>.
- Payrastre, O., Nicolle, P., Bonnifait, L., Brigode, P., Astagneau, P.C., Baise, A., Belleville, A., Bouamara, N., Bourgin, F., Breil, P., Brunet, P., Cerbelaud, A., Courapied, F., Devreux, L., Dreyfus, R., Gaume, E., Nomis, S., Poggio, J., Pons, F., Rabab, Y. & Sevrez, D. (2022). Tempête Alex du 2 octobre 2020 dans les Alpes-Maritimes : une contribution de la communauté scientifique à l'estimation des débits de pointe des crues. *LHB* <https://doi.org/10.1080/27678490.2022.2082891>.

- Pelletier, A. (2021). *Nappes et rivières : la piézométrie peut-elle améliorer la prévision des étiages des cours d'eau ?* Thèse de doctorat, Inrae Antony, Ecole des Ponts, Sorbonne Université, <https://tel.archives-ouvertes.fr/tel-03783485>.
- Pelletier, A. & Andréassian, V. (2022). On constraining a lumped hydrological model with both piezometry and streamflow: results of a large sample evaluation. *Hydrology and Earth System Sciences* **26**, 2733–2758, <https://doi.org/10.5194/hess-26-2733-2022>.
- Peredo, D. (2021). *Quels gains d'une modélisation hydrologique adaptée et d'une approche d'ensemble pour la prévision des crues rapides ?* Thèse de doctorat, Inrae Antony, Sorbonne Université, <https://hal.sorbonne-universite.fr/tel-03469456v1>.
- Peredo, D., Ramos, M.H., Andréassian, V. & Oudin, L. (2022). Investigating hydrological model versatility to simulate extreme flood events. *Hydrological Sciences Journal* <https://doi.org/10.1080/02626667.2022.2030864>.
- Perrin, C., Michel, C. & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology* **242**, 275–301, [https://doi.org/10.1016/S0022-1694\(00\)00393-0](https://doi.org/10.1016/S0022-1694(00)00393-0).
- Perrin, C. (2000). *Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative*. Thèse de doctorat, Institut national polytechnique de Grenoble, Cemagref, Antony, <https://tel.archives-ouvertes.fr/tel-00006216>.
- Perrin, C., Andréassian, V., Serna, C.R., Mathevet, T. & Le Moine, N. (2008). Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research* **44**, W08447, <https://doi.org/10.1029/2007WR006579>.
- Perrin, C., Michel, C. & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* **279**, 275 – 289, [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7).
- 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. *Hydrological Sciences Journal* **52**, 131–151, <https://doi.org/10.1623/hysj.52.1.131>.
- Piazzì, G., Thirel, G., Perrin, C. & Delaigle, O. (2021). Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. *Water Resources Research* **57**, e2020WR028390, <https://doi.org/10.1029/2020WR028390>.
- Pinna, T. (2017). *Peut-on améliorer les performances du modèle GRP pour la prévision des crues ?* Rapport de Master, Paris, <https://webgr.inrae.fr/publications/memoires-de-master>.
- Piotte, O., Montmerle, T., Fouchier, C., Belleudy, A., Garandeau, L., Janet, B., Jauffret, C., Demargne, J. & Organde, D. (2020). Les évolutions du service d'avertissement sur les pluies intenses et les crues soudaines en France. *La Houille Blanche* **106**, 75–84, <https://doi.org/10.1051/lhb/2020055>.

- Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V. & Perrin, C. (2017). Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. *Water Resources Research* **53**, 7247–7268, <https://doi.org/10.1002/2016wr019991>.
- Pool, S., Vis, M. & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal* **63**, 1941–1953, <https://doi.org/10.1080/02626667.2018.1552002>.
- Pool, S., Vis, M. & Seibert, J. (2021). Regionalization for ungauged catchments — lessons learned from a comparative large-sample study. *Water Resources Research* **57**, e2021WR030437, <https://doi.org/10.1029/2021wr030437>.
- Pushpalatha, R., Perrin, C., Le Moine, N. & Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology* **420-421**, 171–182, <https://doi.org/10.1016/j.jhydrol.2011.11.055>.
- Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T. & Andréassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology* **411**, 66–76, <https://doi.org/10.1016/j.jhydrol.2011.09.034>.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Rakovec, O., Kumar, R., Attinger, S. & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research* **52**, 7779–7792, <https://doi.org/10.1002/2016wr019430>.
- Refsgaard, J.C. (1997). Validation and intercomparison of different updating procedures for real-time forecasting. *Hydrology Research* **28**, 65–84, <https://doi.org/10.2166/nh.1997.0005>.
- Roundy, J.K., Duan, Q. & Schaake, J. (2018). Hydrological predictability, scales, and uncertainty issues. *Handbook of Hydrometeorological Ensemble Forecasting*, p. 1–29, Springer Berlin Heidelberg.
- Roux, H., Labat, D., Garambois, P.A., Maubourguet, M.M., Chorda, J. & Dartus, D. (2011). A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments. *Natural Hazards and Earth System Sciences* **11**, 2567–2582, <https://doi.org/10.5194/nhess-11-2567-2011>.
- Royer-Gaspard, P. (2021). *De la robustesse des modèles hydrologiques face à des conditions climatiques variables*. Thèse de doctorat, INRAE Antony, Sorbonne Université.
- Ruiz-Villanueva, V., Borga, M., Zocatelli, D., Marchi, L., Gaume, E. & Ehret, U. (2012). Extreme flood response to short-duration convective rainfall in South-West Germany. *Hydrology and Earth System Sciences* **16**, 1543–1559, <https://doi.org/10.5194/hess-16-1543-2012>.
- Saadi, M. (2020). *Représentation de l'urbanisation dans la modélisation hydrologique à l'échelle du bassin versant*. Thèse de doctorat, Sorbonne Université, <https://tel.archives-ouvertes.fr/tel-03250292>.

- Saadi, M., Oudin, L. & Ribstein, P. (2019). Random forest ability in regionalizing hourly hydrological model parameters. *Water* **11**, 1540, <https://doi.org/10.3390/w11081540>.
- Saadi, M., Oudin, L. & Ribstein, P. (2021). Physically consistent conceptual rainfall–runoff model for urbanized catchments. *Journal of Hydrology* **599**, 126394, <https://doi.org/10.1016/j.jhydrol.2021.126394>.
- Santos, L., Thirel, G. & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences* **22**, 4583–4591, <https://doi.org/10.5194/hess-22-4583-2018>.
- Sauquet, E., Gottschalk, L. & Krasovskaia, I. (2008). Estimating mean monthly runoff at ungauged locations: an application to France. *Hydrology Research* **39**, 403–423, <https://doi.org/10.2166/nh.2008.331>.
- Savenije, H.H.G. (2010). HESS opinions: “Topography driven conceptual modelling (FLEX-topo)”. *Hydrology and Earth System Sciences* **14**, 2681–2692, <https://doi.org/10.5194/hess-14-2681-2010>.
- Schaake, J., Duan, Q., Smith, M. & Koren, V. (2000). Criteria to select basins for hydrologic model development and testing. *Preprints, 15th Conference on Hydrology* p. 10–14.
- Schaefli, B. & Gupta, H.V. (2007). Do Nash values have value? *Hydrological Processes* **21**, 2075–2080, <https://doi.org/10.1002/hyp.6825>.
- SCS (1956). *National Engineering Handbook*. Soil Conservation Service USDA.
- Seibert, J. & Vis, M.J.P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences* **16**, 3315–3325, <https://doi.org/10.5194/hess-16-3315-2012>.
- Seibert, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes* **15**, 1063–1064, <https://doi.org/10.1002/hyp.446>.
- Seibert, J., Vis, M.J.P., Lewis, E. & van Meerveld, H. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes* **32**, 1120–1125, <https://doi.org/10.1002/hyp.11476>.
- Shafii, M. & Tolson, B.A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research* **51**, 3796–3814, <https://doi.org/10.1002/2014wr016520>.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell system technical journal* **27**, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science* **25**, <https://doi.org/10.1214/10-sts330>.
- Slater, L., Arnal, L., Boucher, M.A., Chang, A.Y.Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R.L., Wood, A. & Zappa, M. (2022). Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models. *Hydrology and Earth System Sciences Discussions (in review)*, <https://doi.org/10.5194/hess-2022-334>.

- Slater, L.J. & Wilby, R.L. (2017). Measuring the changing pulse of rivers. *Science* **357**, 552–552, <https://doi.org/10.1126/science.aao2441>.
- Smith, M.B., Koren, V.I., Zhang, Z., Reed, S.M., Pan, J.J. & Moreda, F. (2004). Runoff response to spatial variability in precipitation: an analysis of observed data. *Journal of Hydrology* **298**, 267–286, <https://doi.org/10.1016/j.jhydrol.2004.03.039>.
- Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation* **55**, 271–280, [https://doi.org/10.1016/s0378-4754\(00\)00270-6](https://doi.org/10.1016/s0378-4754(00)00270-6).
- Stanić, M., Todorović, A., Vasilić, Ž. & Plavšić, J. (2017). Extreme flood reconstruction by using the 3DNet platform for hydrological modelling. *Journal of Hydroinformatics* **20**, 766–783, <https://doi.org/10.2166/hydro.2017.050>.
- Stein, L., Clark, M.P., Knoben, W.J.M., Pianosi, F. & Woods, R.A. (2021). How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. *Water Resources Research* **57**, e2020WR028300, <https://doi.org/10.1029/2020wr028300>.
- Stephens, C., Johnson, F. & Marshall, L. (2018). Implications of future climate change for event-based hydrologic models. *Advances in Water Resources* **119**, 95–110, <https://doi.org/10.1016/j.advwatres.2018.07.004>.
- Szönyi, M., Roezer, V., Deubelli, T., Ulrich, J., MacClune, K., Laurien, F. & Norton, R. (2022). PERC floods following “Bernd”. Rapport technique, Zurich Insurance Company, [https://pure.iiasa.ac.at/id/eprint/18144/1/Zurich_PERC_Analysis_Bernd_ENG\(2\).pdf](https://pure.iiasa.ac.at/id/eprint/18144/1/Zurich_PERC_Analysis_Bernd_ENG(2).pdf).
- Tabary, P., Dupuy, P., L’Henaff, G., Gueguen, C., Moulin, L., Laurantin, O., Merlier, C. & Soubeyroux, J.M. (2012). A 10-year (1997–2006) reanalysis of Quantitative Precipitation Estimation over France: methodology and first results. *IAHS Publication* **351**, 255–260, http://hydrologie.org/redbooks/a351/abs_351_0255.pdf.
- Tangara, M. (2005). *Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global*. Thèse de doctorat, Ecole pratique des hautes études, Paris, <https://webgr.inrae.fr/publications/theses>.
- Tanguy, J.M., Carriere, J.M., le Trionnaire, Y. & Schoen, R. (2005). Réorganisation de l’annonce des crues en France. *La Houille Blanche* **91**, 44–48, <https://doi.org/10.1051/lhb:200502005>.
- Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R. & Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research* **56**, e2019WR026951, <https://doi.org/10.1029/2019wr026951>.
- Tarasova, L., Basso, S., Zink, M. & Merz, R. (2018). Exploring controls on rainfall-runoff events: 1. time series-based event separation and temporal dynamics of event runoff response in Germany. *Water Resources Research* **54**, 7711–7732, <https://doi.org/10.1029/2018wr022587>.

- Thielen, J., Bartholmes, J., Ramos, M.H. & de Roo, A. (2009). The European flood alert system – part 1: Concept and development. *Hydrology and Earth System Sciences* **13**, 125–140, <https://doi.org/10.5194/hess-13-125-2009>.
- Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S. & Habets, F. (2010a). A past discharges assimilation system for ensemble streamflow forecasts over France – part 1: Description and validation of the assimilation system. *Hydrology and Earth System Sciences* **14**, 1623–1637, <https://doi.org/10.5194/hess-14-1623-2010>.
- Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S., Regimbeau, F. & Habets, F. (2010b). A past discharge assimilation system for ensemble streamflow forecasts over France – part 2: Impact on the ensemble streamflow forecasts. *Hydrology and Earth System Sciences* **14**, 1639–1653, <https://doi.org/10.5194/hess-14-1639-2010>.
- Thirel, G., Andréassian, V. & Perrin, C. (2015). On the need to test hydrological models under changing conditions. *Hydrological Sciences Journal* **60**, 1165–1173, <https://doi.org/10.1080/02626667.2015.1050027>.
- Tong, R., Parajka, J., Széles, B., Greimeister-Pfeil, I., Vreugdenhil, M., Komma, J., Valent, P. & Blöschl, G. (2022). The value of satellite soil moisture and snow cover data for the transfer of hydrological model parameters to ungauged sites. *Hydrology and Earth System Sciences* **26**, 1779–1799, <https://doi.org/10.5194/hess-26-1779-2022>.
- Tramblay, Y., Ribes, A., Somot, S., Neppel, L., Lucas-Picher, P., Vinet, F. & Sauquet, E. (2021). Impacts du changement climatique sur les pluies intenses et les crues en Méditerranée. *LHB* **107**, 1–5, <https://doi.org/10.1080/00186368.2021.1912971>.
- Tramblay, Y. & Somot, S. (2018). Future evolution of extreme precipitation in the mediterranean. *Climatic Change* **151**, 289–302, <https://doi.org/10.1007/s10584-018-2300-5>.
- Troin, M., Arsenault, R., Wood, A.W., Brissette, F. & Martel, J.L. (2021). Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. *Water Resources Research* **57**, e2020WR028392, <https://doi.org/10.1029/2020wr028392>.
- Tyralis, H., Papacharalampous, G. & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **11**, 910, <https://doi.org/10.3390/w11050910>.
- Uchida, T., van Meerveld, I.T. & McDonnell, J.J. (2005). The role of lateral pipe flow in hillslope runoff response: an intercomparison of non-linear hillslope response. *Journal of Hydrology* **311**, 117–133, <https://doi.org/10.1016/j.jhydrol.2005.01.012>.
- van Esse, W.R., Perrin, C., Booij, M.J., Augustijn, D.C.M., Fenicia, F., Kavetski, D. & Lobligois, F. (2013). The influence of conceptual model structure on model performance: a comparative study for 237 French catchments. *Hydrology and Earth System Sciences* **17**, 4227–4239, <https://doi.org/10.5194/hess-17-4227-2013>.
- van Meerveld, H.J.T. & McDonnell, J.J. (2006). Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the panola hillslope. *Water Resources Research* **42**, W02410, <https://doi.org/10.1029/2004WR003778>.

- Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Teng, J. & Viney, N.R. (2011). Conceptual rainfall–runoff model performance with different spatial rainfall inputs. *Journal of Hydrometeorology* **12**, 1100–1112, <https://doi.org/10.1175/2011jhm1340.1>.
- Vaze, J., Post, D., Chiew, F., Perraud, J.M., Viney, N. & Teng, J. (2010). Climate non-stationarity – validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology* **394**, 447–457, <https://doi.org/10.1016/j.jhydrol.2010.09.018>.
- Vergara, H., Kirstetter, P.E., Gourley, J.J., Flamig, Z.L., Hong, Y., Arthur, A. & Kolar, R. (2016). Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States. *Journal of Hydrology* **541**, 421–433, <https://doi.org/10.1016/j.jhydrol.2016.06.011>.
- Verma, R.K., Verma, S., Mishra, S.K. & Pandey, A. (2021). SCS-CN-based improved models for direct surface runoff estimation from large rainfall events. *Water Resources Management* **35**, 2149–2175, <https://doi.org/10.1007/s11269-021-02831-5>.
- Verma, S., Singh, P., Mishra, S., Singh, V., Singh, V. & Singh, A. (2020). Activation soil moisture accounting (ASMA) for runoff estimation using soil conservation service curve number (SCS-CN) method. *Journal of Hydrology* **589**, 125114, <https://doi.org/10.1016/j.jhydrol.2020.125114>.
- Viatgé, J. & Perrin, C. (2019). Formation sur le modèle GRP : construction et principe de fonctionnement. Support de formation, Irstea, Antony.
- Viatgé, J., Berthet, L., Marty, R., Bourgin, F., Piotte, O., Ramos, M.H. & Perrin, C. (2019a). Vers une production en temps réel d'intervalles prédictifs associés aux prévisions de crue dans vigicrues en France. *La Houille Blanche* p. 63–71, <https://doi.org/10.1051/lhb/2019016>.
- Viatgé, J., Pinna, T., Ficchi, A., Perrin, C., Dorchies, D., Garandeau, L. & Tilmant, F. (2019b). Vers une plus grande flexibilité temporelle du modèle opérationnel de prévision des crues GRP. *La Houille Blanche* p. 72–80, <https://doi.org/10.1051/lhb/2019017>.
- Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M. & Soubeyroux, J.M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology* **30**, 1627–1644, <https://doi.org/https://doi.org/10.1002/joc.2003>.
- Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao, P.S.C., Basu, N.B. & Wilson, J.S. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research* **46**, W05301, <https://doi.org/10.1029/2009wr008906>.
- Wang, A., Li, K.Y. & Lettenmaier, D.P. (2008). Integration of the variable infiltration capacity model soil hydrology scheme into the community land model. *Journal of Geophysical Research* **113**, <https://doi.org/10.1029/2007jd009246>.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* **1**, 80, <https://doi.org/10.2307/3001968>.

- Willems, P. (2009). A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling & Software* **24**, 311–321, <https://doi.org/10.1016/j.envsoft.2008.09.005>.
- Willems, P. (2014). Parsimonious rainfall–runoff model construction supported by time series processing and validation of hydrological extremes – part 1: Step-wise model-structure identification and calibration approach. *Journal of Hydrology* **510**, 578–590, <https://doi.org/10.1016/j.jhydrol.2014.01.017>.
- Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H.H.G., Uhlenbrook, S., Kavetski, D. & Pfister, L. (2015). Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments. *Hydrological Processes* **29**, 2731–2750, <https://doi.org/10.1002/hyp.10393>.
- Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F. & Robertson, D.E. (2020). Ensemble flood forecasting: Current status and future opportunities. *WIREs Water* **7**, <https://doi.org/10.1002/wat2.1432>.
- Yan, H. & Moradkhani, H. (2016). Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. *Advances in Water Resources* **94**, 364–378, <https://doi.org/10.1016/j.advwatres.2016.06.002>.
- Yang, X. & Michel, C. (2000). Flood forecasting with a watershed model: a new method of parameter updating. *Hydrological Sciences Journal* **45**, 537–546, <https://doi.org/10.1080/02626660009492355>.
- Yilmaz, K.K., Gupta, H.V. & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research* **44**, W09417, <https://doi.org/10.1029/2007wr006716>.
- Zanchetta, A. & Coulibaly, P. (2020). Recent advances in real-time pluvial flash flood forecasting. *Water* **12**, 570, <https://doi.org/10.3390/w12020570>.
- Zhang, X., Zwiers, F.W., Li, G., Wan, H. & Cannon, A.J. (2017). Complexity in estimating past and future extreme short-duration rainfall. *Nature Geoscience* **10**, 255–259, <https://doi.org/10.1038/ngeo2911>.
- Zocatelli, D., Borga, M., Viglione, A., Chirico, G.B. & Blöschl, G. (2011). Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response. *Hydrology and Earth System Sciences* **15**, 3767–3783, <https://doi.org/10.5194/hess-15-3767-2011>.
- Zocatelli, D., Borga, M., Chirico, G.B. & Nikolopoulos, E. (2015). The relative role of hillslope and river network routing in the hydrologic response to spatially variable rainfall fields. *Journal of Hydrology* **531**, 349–359, <https://doi.org/10.1016/j.jhydrol.2015.08.014>.