

Université d'Aix-Marseille

ÉCOLE DOCTORALE 251 "SCIENCES DE L'ENVIRONNEMENT"

ESTIMATION VARIATIONNELLE DES PARAMÈTRES D'UN MODÈLE HYDROLOGIQUE DISTRIBUÉ.

Thèse pour l'obtention du titre de docteur présentée par
Maxime JAY-ALLEMAND

UMR RECOVER - INRAE *Aix-en-Provence* / UMR G-EAU - INRAE *Montpellier*

Encadrement :

Patrick ARNAUD
Igor GEJADZE
Pierre JAVELLE

Ingénieur de recherche INRAE
Directeur de recherche INRAE
Ingénieur de recherche INRAE

Directeur
Co-directeur
Co-encadrant

Jury :

Hélène ROUX
Arthur VIDARD
Anne-Catherine FAVRE
Roger MOUSSA
François-Xavier LE DIMET

Maître de conférence INP/IMFT
Chargé de recherche INRIA
Professeure IGE
Directeur de recherche LISAH
Professeur émérite LJK

Rapporteur
Rapporteur
Examineur
Examineur
Invité

13 janvier 2021

Résumé

Le calage d'un modèle hydrologique distribué et conceptuel est une étape difficile. En effet, des incertitudes significatives quant à la modélisation des processus hydrologiques existent, le nombre de paramètres à déterminer est très grand et les méthodes de calage classiques ne sont pas adaptées pour converger vers les valeurs des paramètres du modèle. Le but de ces travaux est d'expérimenter l'utilisation d'une méthode d'estimation variationnelle pour calibrer les paramètres d'un modèle hydrologique distribué conçu pour la modélisation des crues éclaircies. Ce modèle est défini sur une grille rectangulaire de 1 km^2 de résolution sur laquelle trois paramètres sont associés à chaque cellule. 23 bassins-versants "aval" et un total 118 sous-bassins emboîtés "amont", essentiellement situés en région méditerranéenne, ont été étudiés. Pour chaque bassin-versant, on dispose des données de débits au niveau de plusieurs stations de jaugeage, des observations de pluies combinant pluviomètres au sol et radars météorologiques et des données d'estimation de l'évapotranspiration de manière continue entre 2007 et 2018 au pas de temps horaire.

Le calage variationnel des paramètres du modèle dépend d'une fonction objectif quadratique qui pénalise l'écart entre les variables observées et simulées. Celle-ci est minimisée sous des contraintes additionnelles a priori. Le gradient de la fonction coût est calculé de manière efficace par le modèle adjoint. Des expériences numériques de calage-validation sont réalisées. Les bénéfices d'un calage distribué par rapport à un calage uniforme des paramètres sont évalués en termes de performances prédictives en validation temporelle, spatiale et spatio-temporelle. Le calage distribué des paramètres sur les bassins "aval" permet au modèle de fournir de meilleures prédictions des débits aux exutoires "amont" utilisés pour la validation. La variabilité spatiale des paramètres optimisés semble permettre de représenter certaines caractéristiques hydrologiques des bassins-versants étudiés. Des expériences numériques complémentaires de calage-validation ont été menées. Parmi ces expériences, certaines intègrent des contraintes spatiales fournies par des descripteurs physiques dans le but d'améliorer les performances prédictives du modèle et la robustesse du calage distribué. D'autres utilisent une fonction objectif dérivant des méthodes d'estimation robuste. Ces dernières expériences explorent des directions de recherches futures. Les résultats sont positifs et les perspectives sont encourageantes notamment pour développer dans le futur des approches permettant de régionaliser les valeurs des paramètres ou bien d'assimiler des observations en temps réel pour faire de la prévision.

Abstract

Calibration of a conceptual hydrological and distributed model is challenging due to a number of reasons, which include fundamental (model adequacy, identifiability) and algorithmic (e.g. local-search versus global-search) issues. The aim of this research work is to investigate the potential of the variational approach for calibrating a simple continuous hydrological model (GRD) involved in several flash-flood modelling applications. This model is defined on a rectangular 1 km^2 resolution grid, with three parameters being associated to each cell. 23 watersheds are chosen as the study benchmark. These watersheds are all located in the metropolitan France and present various hydrological characteristics (shapes, surfaces, bedrock). Twenty-two of them belong to the Mediterranean region and one spreads over the South-East of Paris. For these watersheds, the discharge observations at several gauging stations, gridded rainfall and potential evapotranspiration estimates are continuously available for the 2007-2018 period at an hourly time step.

In the variational approach one looks for the optimal solution by minimizing the standard quadratic cost-function which penalizes the misfit between the observed and predicted values, under some additional a-priori constraints. The cost-function gradient is efficiently computed using the adjoint model. In numerical experiments, the benefits of using the distributed against the uniform calibration are measured in terms of the model predictive performance, in temporal, spatial and spatio-temporal validation. Overall, distributed calibration shows encouraging results, providing better model predictions and relevant spatial distribution of some parameters. Some complementary numerical calibration experiments have been performed. Some of them integrates several environmental descriptors to improve the robustness of the calibrated parameters and the model predictive performances. Others use a new cost function which derive from the concepts of the robust estimation theory. These latter experiments explore future research directions. Future perspectives are identified, such as the development of a regionalisation approach for the model parameters or real-time assimilation of observations in order to produce forecasts.

Remerciements

Les travaux présentés dans cette thèse ont été effectués au sein de projet PICS. Ce travail est le fruit de mes activités de recherche réalisées durant mes trois années de salariat à l'IRSTEA devenu INRAE. Je tiens à apporter mes remerciements les plus chaleureux à l'ensemble des personnes qui m'ont soutenu et ont rendu ce travail possible. En particulier je remercie sincèrement Pierre Javelle, Patrick Arnaud, Igor Gejadze et Pierre-Olivier Malaterre pour m'avoir encouragé et encadré pendant ces trois années. Ils ont toujours été présents et réactifs à mes nombreuses sollicitations. Je remercie vivement l'ensemble de l'équipe d'Hyris Hydrologie, Jean-Alain Fine, Didier Organde et Julie Demargne avec qui j'ai travaillé en étroite collaboration tout au long de cette thèse. C'est notamment grâce à eux que j'ai découvert l'hydrologie lorsqu'ils m'ont accueilli dans leur équipe les années qui ont précédé cette thèse.

Je remercie sincèrement Hélène Roux et Arthur Vidard qui ont accepté d'évaluer mon travail. J'adresse également mes remerciements aux personnes ayant accepté d'être membre du jury le 10 décembre 2020, Roger Moussa, Anne Catherine Favre et Francois-Xavier Le Dimet.

Enfin je souhaite exprimer ma reconnaissance aux nombreuses personnes qui ont contribué de près ou de loin à la réalisation de ces travaux. Je pense en particulier à toutes celles et ceux qui m'ont apporté leurs aides en informatique, qui ont lu et relu ce manuscrit, les personnes qui m'ont soutenus dans l'accomplissement de ce travail et celles qui m'ont permis de m'aérer l'esprit lors de nombreuses et indispensables séances d'escalade.

Pour cette thèse, j'ai obtenu le soutien financier de l'Agence National Française pour la Recherche via le projet PICS (Prévision Immédiate Intégré des Impacts des Crues Soudaines) et de l'équipe RHAX d'IRSTEA/INRAE d'Aix-en-Provence.

Table des matières

1	L'hydrologie et la prévision des crues	17
1.1	De l'hydrologie à la genèse des crues	18
1.1.1	L'hydrologie	18
1.1.2	Les bassins-versants	19
1.1.3	La genèse des crues	19
1.1.4	Caractérisation des crues	20
1.1.5	Les risques engendrés par les crues rapides	20
1.2	Les outils d'anticipation des crues et leurs limites	22
1.2.1	La modélisation pour la prévision des crues	22
1.2.2	Les outils de prévision et d'anticipation des crues rapides	24
1.3	Le développement des modèles distribués pour la prévision des crues rapides . . .	25
1.3.1	Le principe des modèles distribués	25
1.3.2	Les avantages des modèles distribués	27
1.3.3	Difficultés de mise en œuvre des modèles distribués	28
1.4	Contexte, questions scientifiques et objectifs	29
1.4.1	Développer un modèle distribué pour la prévision des débits aux exutoires non-jaugés	30
1.4.2	Décrire les processus hydrologiques locaux en estimant la variabilité spatiale des paramètres des modèles distribués	30
1.4.3	Choisir un algorithme pour estimer spatialement les paramètres d'un modèle distribué	31
1.4.4	Développer une méthodologie pour estimer les paramètres d'un modèle distribué à partir d'un algorithme de calage : méthode de validation et contraintes de calage	33
1.4.5	Contribuer au développement du modèle hydrologique opérationnel du service VigiCrue-Flash	33
2	Préparation des jeux de données	34
2.1	Les sources de données	35
2.1.1	Les données spatiales hydro-météorologiques	35
2.1.2	Les données caractéristiques des bassins-versants	36

2.1.3	Les données spatiales utilisées pour contraindre les paramètres du modèle	37
2.1.4	Les données de débits aux stations hydrométriques	43
2.2	Les bassins-versants étudiés	44
2.2.1	Contexte et zones géographiques étudiés	44
2.2.2	Sélection et critiques des stations hydrométriques	46
2.2.3	Maillage des bassins-versants pour la modélisation distribuée	51
3	Le modèle hydrologique distribué	54
3.1	Présentation du modèle pluie-débit	55
3.1.1	L'opérateur de production	56
3.1.2	L'opérateur de transfert	57
3.2	Développement du modèle de routage : trois variantes proposées	58
3.2.1	Implémentation d'un schéma de routage de maille en maille	59
3.2.2	Modèles conceptuels modélisant un temps de retard	60
3.2.3	Description de la variante 1 : modèle de retard linéaire	67
3.2.4	Description de la variante 2 : modèle de retard non-linéaire	69
3.2.5	Description de la variante 3 : modèle à réservoir linéaire	71
3.2.6	Algorithme général et simplifié du modèle GRD	72
3.3	Étude des réponses hydrauliques élémentaires des trois modèles de routage	73
3.3.1	Protocole	73
3.3.2	Résultats :	75
4	Choix et implémentation de l'algorithme de calage variationnel	77
4.1	Choix de l'algorithme d'estimation variationnel	78
4.2	Principe de l'estimation variationnelle	79
4.2.1	Notations	81
4.2.2	Formulation générique d'un problème d'estimation variationnelle (assimilation de données)	82
4.2.3	Algorithme d'estimation variationnelle	85
4.2.4	La méthode de la L-Curve	85
4.2.5	Régularisation itérative	86
4.2.6	Contraintes d'inégalité et normalisation du vecteur de contrôle	88
4.3	Critères d'estimation utilisés pour le calcul de la fonction coût	91
4.3.1	Fonction quadratique, méthode des moindres carrés	91
4.3.2	Fonction non-gaussienne, méthode d'estimation robuste	92
4.4	Le terme de régularisation	95
4.4.1	Formulation du terme de régularisation de type Tikhonov	95
4.4.2	Formulation de la matrice $B^{1/2}$ pour la méthode de la régularisation itérative	97
4.5	Détails pratiques quant à l'implémentation et l'utilisation de l'algorithme d'estimation variationnelle dans GRD	98
4.5.1	Choix des bornes	98

4.5.2	Détermination automatique du coefficient β pour la méthode de la L-curve	99
4.5.3	Réduction de la taille du vecteur de contrôle	100
4.5.4	Obtention du modèle adjoint	101
4.6	Méthode de calage globale pour l'évaluation de l'ébauche	102
4.7	Conclusions	103
5	Expériences préliminaires de calage et évaluation des modèles de routage	105
5.1	Expériences jumelles de calage variationnel	106
5.1.1	Description des expériences jumelles	106
5.1.2	Résultats de l'expérience jumelle 1	108
5.1.3	Résultats de l'expérience jumelle 2 :	109
5.1.4	Résultats de l'expérience 3 :	110
5.1.5	Conclusion	112
5.2	Comparaison des modèles de routage avec un modèle hydraulique de barré de saint-venant (SIC^2)	112
5.2.1	Création des jeux d'observation pour le calage et la validation à l'aide de SIC^2	112
5.2.2	Calage à l'aide des variantes de GRD	113
5.2.3	Validation temporelle des paramètres calibrés	114
5.2.4	Résultats	115
5.3	Tests de la fonction de régularisation Tikhonov : la méthode de la L-curve et la méthode de régularisation itérative	119
5.3.1	Objectifs	119
5.3.2	Expériences et résultats	120
5.3.3	Conclusions	125
5.4	Conclusions	125
6	Expériences de calage-validation croisées	127
6.1	Objectifs et méthodologie des expériences de calage-validation	128
6.1.1	Objectifs	128
6.1.2	Méthodologie	128
6.1.3	Expériences réalisées	131
6.2	Résultats	135
6.2.1	Comparaison des performances des variantes 1, 2 et 3 du modèle hydrologique	135
6.2.2	Comparaison des performances du modèle utilisant des jeux de paramètres issus des calages uniforme et distribué	138
6.2.3	Critique du calage de l'expérience B pour chaque période, comparaison à l'expérience F	140
6.2.4	Influence de l'ébauche et de la régularisation sur les performances prédictives du modèle	150

6.2.5	Analyse des jeux de paramètres distribués	152
6.2.6	Étude de la robustesse des jeux de paramètres résultant de l'expérience B pour les périodes P1 et P2	157
6.2.7	Analyse de la robustesse des paramètres issus des expériences de calage B, C, D, E et F (variante 2)	163
6.3	Conclusions	165
7	Tentatives d'amélioration de la robustesse du calage des paramètres	167
7.1	Contraire la variabilité spatiale à l'aide de descripteurs	168
7.1.1	Principes et objectifs	168
7.1.2	Expériences réalisées	169
7.1.3	Résultats des expériences	172
7.2	Test d'une fonction objectif basée sur la fonction de distribution Gamma	183
7.2.1	Principe et objectif	183
7.2.2	Expériences réalisées	184
7.2.3	Résultats	184
7.3	Conclusions	189
8	Conclusion générale	190
A	Bassins-versants et positionnement des exutoires sur la grille d'écoulement 1km² pour la modélisation des débits par GRD	196
B	Tableaux des 141 stations hydrométriques sélectionnées dans le cadre de PICS et réparties sur 23 grands bassins-versants	201
C	Masques de réduction du vecteur de contrôle	206
	Bibliographie	213

Table des figures

1.1	Schéma d'un bassin-versant.	19
1.2	Classement simplifié des modèles hydrologiques pluie-débit selon la représentation des processus hydrologiques et de leurs échelles de fonctionnement.	24
1.3	Schéma d'une modélisation distribuée d'un bassin-versant à l'aide d'un maillage rectangulaire. Pour chaque maille, le modèle hydrologique \mathcal{M} transformant la pluie en débit est appliqué. Les débits sont propagés de l'amont vers l'aval à l'aide d'un modèle de routage.	26
2.1	Exemple de la lame d'eau radar Antilope J+1 donnant le cumul de pluie horaire en mm le 2011/11/05 entre 11 et 12 heures (source : Météo-France). De fortes pluies affectent les régions méditerranéennes et la Corse.	36
2.2	Codage des directions d'écoulement associé à la maille M.	37
2.3	Carte des types d'occupation des sols centrée sur le bassin-versant du Gardon (résolution 1 km^2). Les exutoires et les contours des bassins-versants associés sont tracés en noir.	38
2.4	Carte de prédisposition à la production en France Métropolitaine. La légende couleur correspond aux valeurs de l'index normalisé. Carte issue du rapport de stage Mahe (2016) . Projection ETRS89.	39
2.5	Carte de prédisposition à la production centrée sur le bassin-versant du Gardon (résolution 1 km^2). L'indice normalisé $([0, 1])$ est découpé en 5 classes de production croissantes (le code couleur représente le numéro des classes). Les exutoires et les contours des bassins-versants associés sont tracés en noir.	40
2.6	Carte des densités de drainage centrée sur le bassin-versant du Gardon (résolution 1 km^2). L'indice de drainage $([0 : 100])$ est découpé en 5 classes de densités de drainage croissantes (le code couleur représente le numéro des classes). Les exutoires et les contours des bassins-versants associés sont tracés en noir.	41
2.7	Carte des différentes couches géologiques centrée sur le bassin-versant du Gardon (résolution 1 km^2). Chaque couleur représente une couche géologique différente. Les exutoires et les contours des bassins-versants associés sont tracés en noir.	42
2.8	Carte des pentes selon 5 classes croissantes centrée sur le bassin-versant du Gardon (résolution 1 km^2). Les exutoires et les contours des bassins-versants associés sont tracés en noir.	43

2.9	Les secteurs d'étude identifiés lors du premier atelier utilisateurs PICS (source : PICS Newsletter 1).	45
2.10	Dates des événements sélectionnés sur chaque secteur d'étude (source : PICS Newsletter 1).	45
2.11	Données mobilisables par événement et pour chaque bassin-versant concerné (source : PICS Newsletter 1).	46
2.12	Schéma représentant un bassin-versant et son réseau hydrographique éligible aux critères de sélection : surface, nombre de stations de jaugeage et disponibilité des mesures de débit	47
2.13	Localisation des bassins-versants étudiés en France métropolitaine. Les numéros bassins-versants correspondent à la numérotation indiquée dans le tableau 2.2.	49
2.14	Chronique de débits complète mesurée sur l'ardèche à Saint-Martin d'Ardèche	51
2.15	Chronique de débits influencée mesurée sur l'Eyrieux aux Ollières sur Eyrieux.	51
2.16	Chronique de débits incomplète mesurée sur le Vistre au Cailar	51
2.17	Chronique de débits "bruitée" en basses eaux mesurée sur la Siagne à Pégomas	51
2.18	Nombre d'exutoires et de sous-bassins-versants à l'intérieur de chaque méta-bassin-versant.	52
2.19	Bassin-versant du Loing en amont de Chalette sur Loing. Le fond représente les surfaces drainées cumulées et calculées à partir de la grille de direction d'écoulement de résolution 1 km^2 . Les exutoires réels (point bleu) sont positionnés sur le réseau hydrographique alors que les exutoires simulés (points rouge) sont positionnés sur la grille d'écoulement.	53
2.20	Bassin-versant de l'Aude en amont de Carcassonne. Le fond représente les surfaces drainées cumulées et calculées à partir de la grille de direction d'écoulement de résolution 1 km^2 . Les exutoires réels (points bleus) sont positionnés sur le réseau hydrographique alors que les exutoires simulés (points rouges) sont positionnés sur la grille d'écoulement.	53
2.21	Les boxplots montrent la répartition des surfaces des bassins-versants associés aux exutoires situés à l'intérieur des méta-bassins-versants. Les points rouges renseignent la surface de chaque méta-bassin-versant.	53
3.1	Schéma de la nouvelle version du modèle GRD incluant une fonction de routage.	56
3.2	Schéma de routage distribué connectant les mailles $i-1$ et i . Les mailles représentées en vert concernent la partie "hydrologie" du modèle. Les mailles représentées en noire concernent la partie "hydraulique" (routage) du modèle.	59
3.3	Fonction de forme gaussienne $K(t, \mu, \sigma)$ représentée pour différentes valeurs de μ et σ . Courbe violette : $\mu = 0$. et $\sigma = 0.5$. Courbe verte : $\mu = 5$. et $\sigma = 2$	63
3.4	Représentation schématique du noyau Gaussien dans l'espace temporel t' . L'axe du bas représente le compteur l défini à l'équation 3.13.	64

3.5	Fonction de distribution Gamma pour différentes valeurs de son mode μ . Le coefficient d'étalement est fixé à $\sigma = 0.5$. La forme asymétrique de cette fonction se rapproche de celle d'un hydrogramme.	65
3.6	Représentation schématique du noyau Gamma dans l'espace temporel t' . Plus le retard τ augmente et plus cette fonction s'étaie. Les axes du bas représentent les compteurs l et t_l définis par l'équation 3.15.	65
3.7	Schéma représentant les échanges de masse entre les cellules $i - 1$, i et $i + 1$. Les cinq étapes de l'algorithme 2 sont numérotées en rouge.	69
3.8	Schéma représentant les échanges de masse entre les cellules.	70
3.9	Réponse hydraulique élémentaire du modèle de routage de la variante 1. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Le paramètre de vitesse de GRD a été ajusté de telle sorte que la position du pic de crue du débit de pointe le plus fort concorde avec celui de SIC^2 . Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps	74
3.10	Réponse hydraulique élémentaire du modèle de routage de la variante 2. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Les coefficients du modèle de routage de la variante 2 ont été ajustés manuellement pour que les débits de pointe et le retard du pic de crue concordent entre SIC^2 et GRD. Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps	75
3.11	Réponse hydraulique élémentaire du modèle de routage de la variante 3. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Les capacités des réservoirs de routage de GRD ont été ajustées au mieux pour que le débit de pointe du pic de crue le plus fort concorde avec celui de SIC^2 . Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps	75
4.1	Schéma présentant un problème direct, un problème inverse et le processus itératif permettant d'identifier les entrées (paramètres) d'un modèle H à l'aide de données observées.	81
4.2	Processus de convergence : la fonction objectif est minimisée en recherchant de manière itérative la valeur optimale de la variable d'entrée x . Cette recherche s'effectue en se déplaçant d'une distance dx_i depuis la valeur d'origine x_i pour chaque itération i	81
4.3	Exemple d'un graphique type L-curve obtenu après sept cycles d'optimisation pour sept valeurs du coefficient de régularisation β . Le coefficient β évolue de manière croissante sur l'intervalle $[0, 0.01]$. Dans cet exemple, sa valeur optimale est estimée à 10^{-5}	86
4.4	Valeur de la fonction coût Gamma présentée à l'équation 4.28 pour $\mu = 1$ et $\sigma = 1$. 95	

4.5	Exemple d'un maillage 6×3 découpée en trois sous domaines (super-maillages). Les paramètres calés seront uniformes sur chacun des trois sous-domaines.	101
5.1	Schéma représentant le maillage utilisé pour les deux expériences jumelles. Le maillage est constitué de dix pixels numérotés de 1 à 10, de l'amont vers l'aval. Les débits simulés Q_{sim} (output) sont émis depuis le pixel aval numéro 10. Les pluies p_1, p_2, \dots, p_{10} (input) affectent l'ensemble des pixels. Les trois paramètres du modèle sont définis pour chaque pixel.	107
5.2	Résultats de l' expérience 1 pour les variantes 1 (colonne de gauche), 2 (colonne centrale) et 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les paramètres de routage (v, ξ, c_r) sont recherchées.	109
5.3	Résultats de l' expérience 2 pour les variantes 1 (colonne de gauche), 2 (colonne centrale), 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les vitesses de routage v sont recherchées.	110
5.4	Résultats de l' expérience 3 pour les variantes 1 (colonne de gauche), 2 (colonne centrale), 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les vitesses de routage v sont recherchées.	111
5.5	Génération des scénarios de débits pour les périodes de calage et de validation pour une amplitude fixée.	113
5.6	Interfaçage des scénarios de débits entrant, des hydrogrammes simulés par SIC^2 (Qobs) et des hydrogrammes simulés par GRD (Qsim). Le maillage de SIC^2 est constitué d'une succession de 20 biefs de 1km (ligne noire) entrecoupés de 21 noeuds (rond bleu). Le maillage GRD est composé de 21 cellules.	114
5.7	Débits aval simulés par GRD et par SIC après calage des paramètres de vitesse de la variante 1 sur un scénario de débits d'amplitude "medium"	116
5.8	Débits aval simulés par GRD et par SIC après calage des paramètres de "géométrie" de la variante 2 sur un scénario de débits d'amplitude "medium"	116
5.9	Débits aval simulés par GRD et par SIC après calage des capacités des réservoirs de routage de la variante 3 sur un scénario de débits d'amplitude "medium"	117
5.10	Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude " medium ". Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).	117

5.11	Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude " low ". Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).	118
5.12	Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude " high ". Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).	119
5.13	Expérience empirique de calage permettant de tester la méthode d'auto-détermination du coefficient de régularisation (méthode automatique présentée à la section 4.5.2). Le choix du coefficient β est réalisé par la L-curve (courbe noir) et par la méthode empirique automatique (triangles rouge).	122
5.14	Rapports ($C_{tr}^{test2}/C_{tr}^{test1}$) calculés à partir des capacités des réservoirs de transfert optimales trouvées pour les 23 bassins-versants.	124
6.1	Schéma représentant un bassin-versant de calage (utilisation des données de débits à la station de jaugeage aval) et les sous-bassins-versants de validation (utilisation des données de débits aux stations de jaugeage amont). Les contours des bassins-versants sont tracés en noirs. Le réseau hydrographique est représenté en bleu.	131
6.2	Organigramme récapitulant les simulations réalisées sur les 23 groupes de bassins-versants. Les résultats sont ensuite groupés par catégorie de simulations pour les périodes P1 et P2.	133
6.3	Performances obtenues pour les variantes 1, 2 et 3 du modèle GRD après calage distribué des paramètres (expérience B). Le graphique de gauche montre les performances (NSE) obtenues en calage (23 stations de calage, 2 périodes). Le graphique de droite montre les performances (NSE) en validation spatiale (118 stations de validation, 2 périodes).	136
6.4	Performances (NSE) en validation spatio-temporelle (118 stations de validation, 2 périodes) obtenues pour les variantes 1, 2 et 3 du modèle GRD après calage distribué des paramètres (expérience B).	137
6.5	Performances en validation spatiale , 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage uniforme (expérience A , courbe verte) et distribué (expérience B , courbe bleu) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	138

6.6	Performances en validation spatio-temporelle , 118 exutoires de validation sur les périodes de validation P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage uniforme (expérience A , courbe verte) et distribué (expérience B , courbe bleu) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	139
6.7	Relation entre les critères de Nash-Sutcliffe (NSE) et cinq classes du rapport $RS_i = \frac{S_i}{S_{aval}}$ (surface de l'exutoire de validation) / (Surface de l'exutoire de calage). Graphique de gauche : résultat en validation spatiale, expériences A et B . Graphique de droite : résultat en validation spatio-temporelle, expériences A et B	140
6.8	Comparaison des performances obtenues entre les périodes P1 et P2 pour la variante 2 du modèle GRD après calage distribué des paramètres (expérience B). Le graphique de gauche montre les performances (NSE) obtenues en calage (23 stations de calage). Le graphique de droite montre les performances (NSE) en validation spatiale (118 stations de validation).	141
6.9	Débits observés (noir) et simulés pour les expériences B et A (variante 2) à la station de calage "la Siagne à Pégomas" (Y5534030) pour des évènements pluvieux importants entre le 1 et le 30 Novembre 2011. Les pluies moyennes sur le bassin-versant sont représentées par l'histogramme (gris).	142
6.10	Performances (NSE) en calage et en validation spatiales obtenues pour la variante 2 du modèle GRD après calage distribué des paramètres pour les périodes P1 et P2 et pour chacun des 23 bassins-versants (aux exutoires de calage et de validation, expérience B). Les exutoires symbolisés avec le triangle pointant vers le haut correspondent à la période P1. Les exutoires symbolisés avec le triangle pointant vers le bas correspondent à la période P2. Les stations utilisées pour le calage sont marquées avec un contour bleu.	144
6.11	Performances (NSE) obtenues pour la variante 2 du modèle GRD après calage distribué des paramètres pour les périodes P1 et P2 et pour chacun des 23 bassins-versants (calage de référence, expérience F , réalisé à l'aide des données hydrométriques aux 141 exutoires). Les exutoires symbolisés avec le triangle pointant vers le haut correspondent à la période P1. Les exutoires symbolisés avec le triangle pointant vers le bas correspondent à la période P2. Les stations utilisées pour le calage sont marquées avec un contour bleu.	145
6.12	Débits observés (noirs) et simulés (rouges) aux stations V4124010 (l'Eyrieux), Y1415010 (l'Orbiel), Y1002020 (l'Aude) et Y1025010 (l'Aude) sur la période P1. Ces hydrogrammes permettent de visualiser les problèmes de modélisation ou de mesures rencontrées.	149

6.13	Performances en validation spatiale , 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage distribué des paramètres, expériences A, B,C, D et E . Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	151
6.14	Performances en validation spatio-temporelle , 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage distribué des paramètres, expériences A, B,C, D et E . Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	151
6.15	Débits observés (noir) et simulés (variantes 2, expérience A, B et D) à la station de calage "la Siagne à Pégomas" (Y5534030) pour des évènements pluvieux importants entre le 1 et le 30 Novembre 2011. Les pluies moyennes sur le bassin-versant sont représentées par l'histogramme (gris).	152
6.16	Capacités des réservoirs de production c_p obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).	154
6.17	Capacités des réservoirs de transfert c_{tr} obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).	155
6.18	Paramètre de routage ξ obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).	156
6.19	Comparaison des valeurs moyennes des paramètres distribués (c_p, c_{tr}, ξ) obtenues après calage sur les périodes P1 (abscisse) et P2 (ordonnée) pour chacun des 23 bassins-versants. Les valeurs moyennes de l'ébauche (issue d'un calage uniforme des paramètres) sont représentées par les cercles violets. Les valeurs moyenne obtenues après calage distribué sont représentées par des croix rouges. Les moyennes ont été calculées sur l'aire des bassins-versants.	158
6.20	Comparaison des capacités des réservoirs de production c_p obtenues après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.	160
6.21	Comparaison des capacités des réservoirs de transfert c_{tr} obtenues après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.	161

6.22	Comparaison des paramètres de routage ξ obtenus après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes <i>med</i> et les écarts-types σ sont calculés sur l'aire du bassin-versant.	162
6.23	Box-plot des distributions des critères d'adéquation entre les paramètres distribués période P1 et période P2 pour chaque bassin-versant. La stabilité des paramètres obtenus pour les expériences B,C,D,E,F est comparée. Graphique A) coefficient de Pearson. Graphique B) Rapport des valeurs moyennes. Graphique C) Moyenne de la variable θ (équation 6.7). Graphique D) Différence entre les valeurs des paramètres centrées-réduites.	164
7.1	Performances prédictives spatio-temporelles (validation spatio-temporelle) à l'issue des étapes de calage 2 (calage semi-uniforme, colonne de gauche) et 3 (calage distribué, colonne de droite). Les performances sont montrées à l'aide des critères de NSE (graphiques du haut) et de KGE (graphiques du bas). Chaque box-plot est construite à l'aide des résultats obtenus aux 118 exutoires de validation sur les périodes de calage P1 et P2, soit 236 points. Sur chaque graphique, huit box-plots sont tracées : calage uniforme, calage distribué issu de l'expérience B (expérience réalisée au chapitre précédent 6), calages semi-uniformes ou distribués (expériences M1 , M2 , M3 , M4 , M5 , M6)	173
7.2	Box-plots représentant les distributions des critères d'adéquation calculés à partir des paramètres calés sur les périodes P1 et P2 pour chaque bassin-versant. Ces critères ont été calculés à l'issue des résultats des étapes de calage numéro 2 (calage semi-uniforme, colonne de gauche) et numéro 3 (calage distribué, colonne de droite). Les graphiques du haut montre les rapports entre les valeurs moyennes des paramètres. Le graphique du bas (à droite) montre les distributions des coefficients de corrélation de Pearson (lorsque les paramètres sont uniformes où quasi-uniformes le calcul des coefficients de Pearson n'est pas pertinent et le graphique n'est pas affiché).	176
7.3	Performances en validation spatio-temporelle , 118 exutoires de validation sur les périodes de validation P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage uniforme (expérience A , courbe verte), distribué (expérience B , courbe bleu), calage semi-uniforme (expérience M3 , courbe orange) et distribué (expérience M3 , courbe rose) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	177

7.4	Performances prédictives spatio-temporelle (validation spatio-temporelle), 118 exutoires de validation sur les périodes de calage P1 et P2 (236 points), obtenues pour la variantes 2 du modèle GRD après calage. Quatre boxplot sont présentées : performances prédictives spatiales avec calage uniforme, performances prédictives spatiales avec calage distribué EXP B (expérience réalisée au chapitre précédent 6), performances prédictives spatiales avec calage contraint par les descripteurs EXP-M3. Les performances sont décrites à l'aide de trois critères composant le critère de KGE. Le graphique en haut a gauche présente les performances à l'aide du critère de corrélation linéaire (Pearson), le graphique en haut à droite présente les performances à l'aide du critère RMU (mesure du biais) et le graphique en bas présente les performances à l'aide du critère RVAR (mesure de la variabilité).	178
7.5	Relation entre les critères de Nash-Sutcliffe (NSE) et cinq classes du rapport $RS_i = \frac{S_i}{S_{aval}}$ (surface de l'exutoire de validation) / (Surface de l'exutoire de calage). Résultat en validation spatio-temporelle, expériences A , B et M3 (étapes deux et trois).	179
7.6	Capacités des réservoirs de production c_p obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.	180
7.7	Capacités des réservoirs de transfert c_{tr} obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.	181
7.8	Paramètres de routage ξ obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.	182
7.9	Box-plots des distribution des critères d'adéquation entre les paramètres distribués période P1 et période P2 pour chaque bassin-versant. Graphique de gauche : coefficients de Pearson. Graphique de droite : rapports des valeurs moyennes. . .	185
7.10	Performances prédictives spatio-temporelle (validation spatio-temporelle), 118 exutoires de validation sur les périodes de calage P1 et P2 (236 points), obtenues pour la variantes 2 du modèle GRD après calage. Sept box-plots sont présentées : performances prédictives spatio-temporelles de l'expérience EXP B (expérience réalisée au chapitre précédent 6), performances prédictives spatio-temporelles des expériences G1-ITREG , G1 , G2 , G3 , G4 . Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.	186
7.11	Performances du modèle évaluées à l'aide du critère NSE aux 23 exutoires de calage sur les périodes de calage P1 et P2 (46 points par boxplot).	187

7.12 Bassin-versant du Gardon. Comparaison des capacités des réservoirs de production c_p obtenues après calage sur la période P1 entre les expériences G1 et D (graphique de gauche) et les expériences G4 et D (graphique de droite) à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.	188
7.13 Bassin-versant de la Cesse. Comparaison des capacités des réservoirs de production c_p obtenues après calage sur la période P1 entre les expériences G1 et D (graphique de gauche) et les expériences G4 et D (graphique de droite) à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.	188
A.1 Contours des bassins-versants étudiés et positionnement des exutoires.	197
A.2 Contours des bassins-versants étudiés et positionnement des exutoires.	198
A.3 Contours des bassins-versants étudiés et positionnement des exutoires.	199
A.4 Contours des bassins-versants étudiés et positionnement des exutoires.	200
C.1 Masques de réduction du vecteur de contrôle basés sur les données d'occupation du sol pour les 23 bassins-versants et utilisés pour les réservoirs de production.	207
C.2 Masques de réduction du vecteur de contrôle basés sur les données de densité de drainage pour les 23 bassins-versants et utilisés pour les réservoirs de production.	208
C.3 Masques de réduction du vecteur de contrôle basés sur l'indice mixte de production pour les 23 bassins-versants et utilisés pour les réservoirs de production.	209
C.4 Masques de réduction du vecteur de contrôle basés sur les données des couches géologiques de surface pour les 23 bassins-versants et utilisés pour les réservoirs de transfert.	210
C.5 Masques de réduction du vecteur de contrôle basés sur les données des pentes pour les 23 bassins-versants et utilisés pour les réservoirs de transfert.	211
C.6 Masques de réduction du vecteur de contrôle basés sur les surfaces cumulées pour les 23 bassins-versants et utilisés pour les paramètres de routage. La classe 0 correspond aux zones non drainées (zones maritimes); la classe 1 correspond au maille situées sur le réseau hydrographique; la classe 2 correspond aux mailles situées sur les versants.	212

Liste des tableaux

1.1	Exemple d'aléas crues et nombre de décès associés en régions méditerranéennes françaises (Vinet, Freddy et al., 2016)	21
1.2	Domages et décès causés par les crues en région méditerranéenne entre 1990 et 2006 (Llasat Botija et al., 2010)	22
1.3	Inventaire non-exhaustif de modèles hydrologiques distribués existant et exemples d'applications.	27
2.1	Tableau décrivant les types d'occupation des sols décrits par le niveau 1 de la nomenclature de la base de donnée CLC 2012.	38
2.2	Liste des 23 méta-bassins-versants étudiés. Le nom de la rivière, la surface du bassin-versant à l'exutoire aval, le code de la station hydrométrique situé à l'exutoire aval et le nombre de stations internes à chaque bassin-versant sont mentionnés.	48
2.3	Critiques des chroniques de débits aux exutoires aval des 23 groupes de bassins-versants.	51
4.1	Ordre de grandeur des bornes inférieures et supérieures des paramètres du modèle GRD utilisées pour les simulations.	99
5.1	Synthèse des caractéristiques des trois expériences jumelles.	108
5.2	Paramètres de calage choisis pour l'expérience 1	121
5.3	Paramètres de calage choisis pour l'expérience 3	123
5.4	123
6.1	Tableau récapitulatif des 23 méta-bassins-versants étudiés. Le nom de la rivière, la surface du bassin-versant à l'exutoire aval, le code de la station hydrométrique situé à l'exutoire aval et le nombre de stations internes à chaque bassin-versant sont mentionnés.	132
6.2	Tableau récapitulatif des expériences de calage-validation.	135
6.3	Tableau permettant de comparer les performances des expériences B et F par bassin-versant sous forme de score.	148
6.4	Tableau récapitulatif des expériences de calage-validation B, C, D et E.	150

7.1	Liste des six expériences de calage réalisées. Pour chacune des six expériences, les descripteurs utilisés pour contraindre la variabilité spatiale des paramètres de production c_p , de transfert c_{tr} et de routage ξ sont listés.	172
7.2	Expériences de calage réalisées avec la fonction objectif Gamma pour différentes valeurs du coefficient d'étalement noté sp	184
B.1	Liste des 141 exutoires avec stations hydrométriques sélectionnés pour contribuer au projet PICS.	205

Chapitre 1

L'hydrologie et la prévision des crues

Sommaire

1.1	De l'hydrologie à la genèse des crues	18
1.1.1	L'hydrologie	18
1.1.2	Les bassins-versants	19
1.1.3	La genèse des crues	19
1.1.4	Caractérisation des crues	20
1.1.5	Les risques engendrés par les crues rapides	20
1.2	Les outils d'anticipation des crues et leurs limites	22
1.2.1	La modélisation pour la prévision des crues	22
1.2.2	Les outils de prévision et d'anticipation des crues rapides	24
1.3	Le développement des modèles distribués pour la prévision des crues rapides	25
1.3.1	Le principe des modèles distribués	25
1.3.2	Les avantages des modèles distribués	27
	L'utilisation de données spatiales	27
	L'estimation des débits à fines échelles	28
	La variabilité spatiale des caractéristiques et de la dynamique des bassins-versants	28
1.3.3	Difficultés de mise en œuvre des modèles distribués	28
1.4	Contexte, questions scientifiques et objectifs	29
1.4.1	Développer un modèle distribué pour la prévision des débits aux exutoires non-jaugés	30
1.4.2	Décrire les processus hydrologiques locaux en estimant la variabilité spatiale des paramètres des modèles distribués	30
1.4.3	Choisir un algorithme pour estimer spatialement les paramètres d'un modèle distribué	31
1.4.4	Développer une méthodologie pour estimer les paramètres d'un modèle distribué à partir d'un algorithme de calage : méthode de validation et contraintes de calage	33

Ce premier chapitre présente le contexte et les objectifs de ces travaux de thèse. L'hydrologie et les principes généraux responsables de la formation des crues sont décrits dans la première partie. Les crues des rivières sont des phénomènes récurrents et concernent la quasi-totalité des surfaces continentales terrestres. Ces crues impactent notre environnement et notre société (érosion, dégâts matériels, pertes humaines...). Leurs caractéristiques ainsi que les enjeux associés sont exposés dans la deuxième partie. Afin de réduire l'impact des crues dans nos sociétés, celles-ci sont surveillées, modélisées et anticipées. Dans la troisième partie les outils de modélisation et de prévision des crues actuellement disponibles sont présentés. Les outils de prévisions des crues sont souvent basés sur des modèles mathématiques permettant de reproduire les phénomènes naturels de genèse des crues. Ces modèles sont très diversifiés (ils peuvent représenter les phénomènes physiques de plusieurs manières) et opèrent à diverses échelles de modélisation (temporelles et spatiales). Dans la quatrième partie les modèles hydrologiques de types "distribués" sont présentés. Ces types de modèles opèrent à des échelles spatiales affinées et caractérisent spatialement les phénomènes hydrologiques. C'est à l'aide de ce type de modèle que des phénomènes hydrologiques locaux sont étudiés et éventuellement prédits. Cependant ces modèles sont complexes à mettre en œuvre car ils nécessitent beaucoup de données et des ressources informatiques importantes. C'est dans ce contexte que les enjeux et les objectifs de ces travaux de recherches sont détaillés dans la cinquième partie de ce chapitre.

1.1 De l'hydrologie à la genèse des crues

1.1.1 L'hydrologie

L'hydrologie est une discipline scientifique qui étudie les mécanismes responsables des flux d'eau sur terre. C'est l'étude du cycle de l'eau. L'eau est une molécule extrêmement répandue sur la planète. Elle est le seul élément présent naturellement sur terre sous ses trois états : liquide (océans, lacs, rivières, nappes), solide (glaciers, calottes polaires, neige) et gazeux (vapeur d'eau). Son étude s'étend à l'ensemble de ses interactions avec l'atmosphère, la biosphère et la lithosphère. L'hydrologie est donc une discipline très vaste située aux croisements de nombreuses autres disciplines scientifiques telles que la météorologie, la géologie, la glaciologie, la biologie, l'océanographie... Cette discipline fait également appel aux mathématiques et à la physique. Les travaux scientifiques actuels, particulièrement ceux impliqués dans les domaines de la modélisation, font un usage intensif des outils informatiques et demandent des connaissances pour le "calcul numérique". Ces travaux de recherche étudient la formation des crues des rivières et visent à améliorer les outils pour leurs modélisations et leurs prévisions.

Les processus qui régissent la formation des crues sont nombreux. Il s'agit des processus responsables des échanges d'eau entre la surface de la terre et l'atmosphère (la pluie, l'évaporation, l'évapotranspiration...) et des processus responsables des flux d'eau à la surface du sol, dans le sol et dans le sous-sol (les écoulements en rivière, le ruissellement, la percolation, le stockage,

les échanges souterrains...). La crue d'une rivière est générée par des flux d'eau impactant une aire géographique délimitée nommée le bassin-versant.

1.1.2 Les bassins-versants

La notion topographique de bassin-versant est essentielle en hydrologie de surface. C'est une étendue géographique délimitée par une ligne de partage des eaux qui collecte et draine les eaux de surfaces vers un exutoire (figure 1.1).

Un bassin-versant peut-être caractérisé par sa surface, sa pente, sa forme, la nature du sol et du sous-sol, le type d'occupation des sols. Chacune de ces caractéristiques influence la réaction du bassin-versant face à un aléa hydro-climatique (pluie, sécheresse, fonte des neiges...) : on parle de réponse hydrologique. Le temps de réponse hydrologique est la durée séparant le début de l'aléa et la réaction du bassin-versant associée.

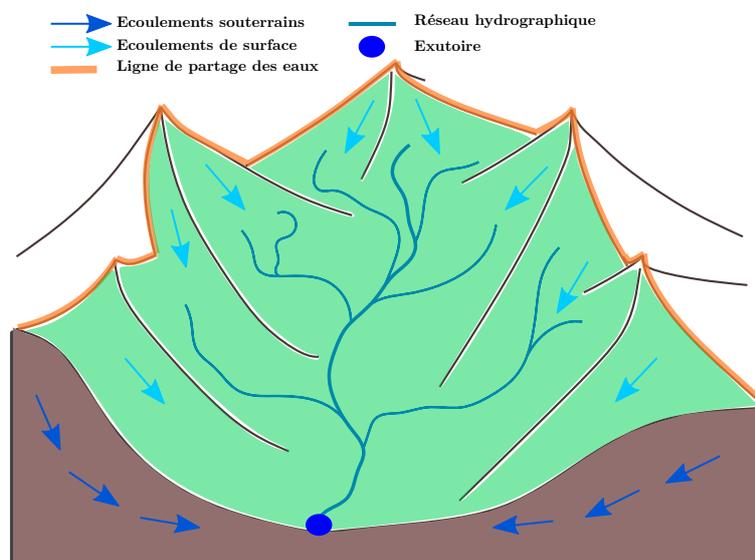


FIGURE 1.1 – Schéma d'un bassin-versant.

1.1.3 La genèse des crues

La crue d'une rivière est une réponse d'un bassin-versant face à un aléa hydro-climatique (forte pluie, fonte des neiges), morphodynamique (la formation d'une onde de submersion provoquée par un glissement de terrain, la vidange d'une poche d'eau sous-glaciaire) ou géodynamique (la rupture d'un barrage après un séisme) (Frédéric et al., 2010). Cet aléa peut-être naturel (orage), induit (rupture d'une digue) ou technologique (vidange d'un barrage). Dans tous les cas, un volume d'eau important est drainé par le bassin-versant jusqu'à son exutoire. Lorsque le débit de la rivière augmente brutalement, on parle de crue. Lorsque le débit maximum est atteint, celui-ci décroît progressivement, on parle de décrue. Une crue est un événement dont la durée peut-être très variable selon le bassin-versant et le caractère de l'aléa (de quelques minutes à plusieurs mois).

1.1.4 Caractérisation des crues

Les crues dépendent des caractéristiques des bassins-versants et du type d'aléa. Il existe des crues dites "rapides" provoquées par exemple par des pluies intenses, souvent localisées et entraînant des phénomènes de ruissellement importants (c'est le cas des épisodes pluvieux intenses observés dans les régions méditerranéennes). Ces crues rapides peuvent être également causées par d'autres aléas tels que la rupture de digues ou de barrages. Lorsque ces crues surviennent, la mise en eau de rivières (parfois "sèches"), de fossés, de talwegs est brutale. Les phénomènes d'érosion associés à ces crues "rapides" sont très puissants (écoulement rapide, charriage de gros matériaux). Les bassins-versants concernés sont de "petites" tailles (de quelques km^2 à plusieurs centaines ou quelques milliers de km^2) et leurs temps de réponse hydrologique sont courts (de quelques minutes à quelques heures) (Gaume et al., 2009). Ces crues sont difficiles à anticiper et sont parfois meurtrières.

A l'inverse, il existe des crues dites "lentes". Ces crues concernent les grands bassins-versants (plusieurs milliers ou centaines de milliers de km^2). Ces dernières peuvent être provoquées par exemple par des pluies continues ou la fonte d'un manteau neigeux sur une grande étendue géographique et pendant plusieurs jours. Ces crues sont plus faciles à anticiper et à prévoir : les observations continues de l'aléa climatique et de la réaction des cours d'eau en amont du bassin-versant renseignent sur l'intensité de la crue à l'aval. Néanmoins, ces crues entraînent des débordements importants des cours d'eau sur de longues périodes de temps. Ces crues sont donc susceptibles d'engendrer des dégâts matériels conséquents et de perturber le fonctionnement d'un écosystème pendant plusieurs jours ou mois.

Une crue peut-être caractérisée par sa période de retour calculée en fonction de sa sévérité (débit de pointe). La période de retour exprime une probabilité. C'est l'inverse d'une probabilité d'apparition annuelle. Par exemple, un débit de pointe centennal correspond à un débit de crue ayant chaque année 1 chance sur 100 d'être vu ou dépassé. L'estimation de la période de retour associée à un débit de pointe est calculée statistiquement à partir de l'historique des crues.

Ces travaux de recherche ciblent l'amélioration des modèles hydrologiques utilisés pour la prévision et l'anticipation des crues rapides.

1.1.5 Les risques engendrés par les crues rapides

Les risques engendrés par les crues sont définis par la combinaison d'aléas, d'enjeux et de vulnérabilités. L'aléa est l'élément susceptible d'impacter un environnement : c'est la crue d'une rivière provoquée le plus souvent par un événement climatique comme une pluie intense. Les enjeux caractérisent les zones directement impactées par la crue (routes, habitations, usines, hôpitaux, champs agricoles...). La vulnérabilité se définit comme la capacité de réaction de ces enjeux face à la crue (mobilité des personnes, valeurs et nature des biens affectés...).

Les enjeux économiques, sociaux, humains, culturels et environnementaux liés aux crues aug-

mentent. La densification des populations dans des zones exposées aux crues est une des raisons de l'augmentation des enjeux (Saint Martin et al., 2016). De plus les changements climatiques actuels affectent les intensités et les périodes de retours des aléas "crues". Par exemple, c'est le cas en Europe de l'Ouest où des événements extrêmes plus fréquents et plus intenses semblent attendus, augmentant ainsi les risques (Dankers and Feyen, 2008).

En France, les crues rapides constituent un risque important. Ces crues sont la cause de nombreuses catastrophes notamment dans les régions méditerranéennes. Ces crues sont provoquées par des événements pluvieux intenses et localisés, appelés épisodes méditerranéens. Ces épisodes surviennent fréquemment en automne. Ces crues sont responsables de dégâts importants : humains, matériels et économiques. Dans la région du Languedoc-Roussillon, en moyenne 10 décès et 150 millions d'euros sont causés chaque année par ces crues (Vinet, 2008). A titre d'exemple, le tableau 1.1 référence quelques événements antérieurs bien documentés et ayant touchés les régions méditerranéennes françaises. Le tableau 1.2 référence les dommages et les décès causés par les crues en région méditerranéenne entre 1990 et 2006.

Face aux enjeux liés aux crues rapides, des outils d'anticipation et d'évaluation des aléas crues ont été développés.

TABLE 1.1 – Exemple d'aléas crues et nombre de décès associés en régions méditerranéennes françaises (Vinet, Freddy et al., 2016)

Date	Lieux	Décès
09/22/1992	Vaison-la-Romaine	49
11/12/1999	Aude	30
09/08/2002	Gard	25
06/15/2010	Var	26
10/03/2015	Alpes-Maritimes	20

TABLE 1.2 – Dommages et décès causés par les crues en région méditerranéenne entre 1990 et 2006 (Llasat Botija et al., 2010)

Pays	Décès	Coût (M€)
Albania	19	15.8
Algeria	1206	198.9
Bulgaria	40	296.0
Croatia	3	-
Egypt	648	91.2
France	177	4042.5
Greece	41	820.1
Israel	32	33.7
Italy	466	19780.3
Libya	-	27.0
Fyrom	2	226.3
Morocco	797	5.8
Portugal	21	8.3
Romania	383	1485.6
Serbia-Montenegro	14	-
Slovenia	-	3.2
Spain	170	757.7
Syria	6	-
Tunisia	49	155.4
Turkey	492	1188.2
Total	4,566	29 136.0

1.2 Les outils d'anticipation des crues et leurs limites

L'objectif des outils d'anticipation des crues est de fournir une information sur l'évolution des débits d'un cours d'eau et de déclencher une alerte en cas de dépassement probable d'un seuil défini. Ces outils permettent de caractériser la crue (débits de pointe, temps de réponse du bassin-versant) et éventuellement de détecter les enjeux locaux susceptibles d'être impactés. Ces outils sont basés sur des observations en temps réel (pluie, débits) et sur des modèles mathématiques permettant de prédire l'évolution des débits.

1.2.1 La modélisation pour la prévision des crues

La modélisation consiste à reproduire le comportement d'un phénomène physique mesurable par le biais d'équations mathématiques. En hydrologie, la modélisation est utile pour étudier le fonctionnement d'un bassin-versant et pour comprendre son fonctionnement hydrologique. Connaissant la quantité de pluie tombée, le modèle hydrologique pluie-débit permet d'estimer le débit à l'exutoire d'un bassin-versant. Un modèle représente les processus hydrologiques à partir d'équations issues de :

- la physique : on parle de modèle à base physique (infiltration par rapport à la porosité du sol, équations de l'hydraulique).
- relations conceptuelles : on parle de modèle conceptuel. C'est le cas des modèles à réservoirs (GR4J).
- relations d'entrées-sorties : on parle de modèle type "boîte noire" qui ne font pas appel à des relations conceptuelles. C'est le cas des modèles basés sur des régressions, des réseaux de neurones ou des analyses fréquentielles (Xu et al., 2017).

Un modèle hydrologique, quel que soit le type de relations mathématiques (physiques, conceptuelles, entrées-sorties), dépend nécessairement de paramètres qui caractérisent les réponses hydrologiques des bassins-versants. Ces paramètres sont soit estimés à partir de la connaissance du bassin-versant (topographie, occupation du sol...) soit calibrés pour représenter au mieux des données de débits éventuellement disponibles à l'exutoire du bassin-versant.

Les modèles hydrologiques utilisés en prévision des crues sont souvent simplifiés et n'intègrent pas l'ensemble des processus physiques de la genèse des crues. En effet, la modélisation de l'ensemble des processus sur des grands espaces géographiques (certaines voire milliers de kilomètres carrés) est difficile et nécessite beaucoup de données : des données météorologiques (température, humidité, évaporation, pluie...), des données géomorphologiques (topographie, caractéristiques géologiques du sol et du sous-sol), des données d'occupation du sol (végétation, urbanisation...) et des données hydrométriques (débits des rivières, humidité du sol, hauteur d'eau des nappes...). Cette liste est bien sûr non exhaustive et ces données ne sont souvent pas ou partiellement disponibles. Les modèles actuels simplifient donc les processus hydrologiques. Ainsi les incertitudes sur les débits simulés peuvent être importantes. Notons que complexifier ces modèles se ferait au détriment des temps de calcul et du nombre de paramètres à estimer.

Les modèles hydrologiques travaillent selon plusieurs échelles spatiales. On distingue les modèles globaux et les modèles distribués :

- Dans le cas d'un modèle global, le bassin-versant tout entier est considéré comme une seule et unique unité spatialement homogène et les débits sont modélisés à l'exutoire du bassin-versant.
- Dans le cas des modèles distribués le bassin-versant est découpé en maille (maillage) et les débits sont modélisés pour chacune de ces mailles puis transférés de l'amont vers l'aval. Ces mailles peuvent être régulières (par exemple des carrés définis sur une grille) ou bien respecter des limites des sous-bassins versants (on parle alors plus souvent de modèles semi-distribués car ils sont composés d'une superposition de plusieurs modèles globaux appliqués sur chaque sous-bassins).

Les phénomènes hydrologiques sont des processus dynamiques avec de l'inertie. Les modèles hydrologiques intègrent donc une dimension temporelle dans la modélisation. La fenêtre temporelle de modélisation (période caractérisant la durée de la simulation) est discrétisée selon un pas de temps. Pour chaque pas de temps, les débits sont calculés par le modèle. Aussi, à chaque

pas de temps, la modélisation dépend du résultat du modèle au pas de temps précédent. Le pas de temps de modélisation est donc déterminant pour correctement reproduire le fonctionnement des bassins-versants selon leurs tailles et leurs formes.

Le schéma 1.2 propose un classement très grossier des modèles pluies-débits selon la représentation des processus hydrologiques (équations basées sur des lois physiques ou équations basées sur des opérateurs conceptuels) et selon leurs échelles de fonctionnement (modèles globaux ou modèles distribués).

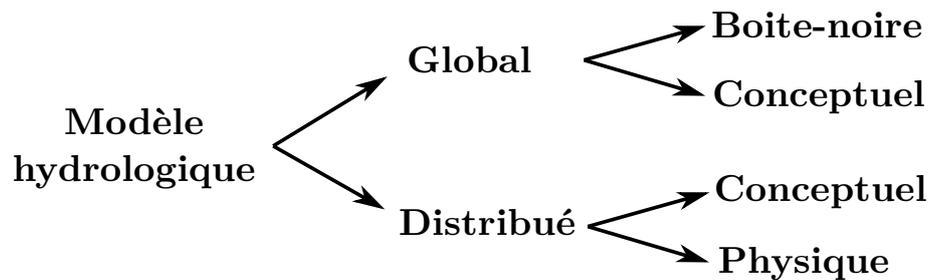


FIGURE 1.2 – Classement simplifié des modèles hydrologiques pluie-débit selon la représentation des processus hydrologiques et de leurs échelles de fonctionnement.

Dans ces travaux de recherche, un modèle hydrologique conceptuel et distribué est utilisé. Ce modèle repose sur des relations mathématiques conceptuelles : plusieurs fonctions caractérisent les phénomènes d'interception, d'infiltration et de ruissellement de la pluie sur le sol et une fonction caractérise le transfert de l'eau de l'amont vers l'aval du bassin-versant. Ce modèle est très simplifié, il fonctionne avec des ressources informatiques limitées et ses opérateurs conceptuels supposent des relations avec des phénomènes hydrologiques. Les modèles à base physique et les modèles de type boîte noire ne sont pas utilisés dans ces travaux. Ces modèles sont soit plus complexes à mettre en œuvre (quantité de données, temps de calcul) soit trop éloignés des concepts hydrologiques (spatialisation des processus hydrologiques difficiles, relations entre paramètres et processus physiques inexistantes).

1.2.2 Les outils de prévision et d'anticipation des crues rapides

Des outils sont déjà opérationnels. Ils couvrent de grandes étendues géographiques (régions, pays, continents) et fournissent des alertes et des prévisions de débits (Alferi et al., 2012). Par exemple, en France métropolitaine le service Vigicrues (<https://www.vigicrues.gouv.fr/>), développé par le SCHAPI (Service central d'hydrométéorologie et d'appui à la prévision des inondations), fournit des prévisions et des cartes de vigilance sur le réseau hydrographique principal. Cependant, cet outil est incomplet et ne concerne pas le réseau secondaire affecté par les crues "rapides". D'autres méthodes spécialisées pour l'anticipation des crues rapides ont donc été développées. Ces méthodes opèrent à plus petites échelles et prennent en compte les observations spatialisées de la pluie fournies par les radars météorologiques.

En France, les méthodes AIGA (Adaptation d'Information Géographique pour l'Alerte en Crue) (Javelle et al., 2014, 2016) et APIC (Carrière et al., 2011) ont été développées spécialement pour l'anticipation des crues rapides. Ces méthodes sont conçues pour anticiper les crues se produisant sur des petits bassins-versants (réseaux hydrographiques non suivis par Vigicrues) et sur lesquels les observations de débit sont indisponibles. La méthode APIC est basée sur la comparaison du cumul des pluies observées par les radars météorologiques avec des quantiles de pluie de référence fournis par la méthode SHYREG (Arnaud et al., 2006). Cette méthode ne tient pas compte des conditions hydrologiques des bassins-versants. La méthode AIGA quant à elle repose sur une modélisation hydrologique distribuée de résolution $1km^2$ et au pas de temps 15 minutes. AIGA produit une estimation des débits en rivière en temps réel à partir des observations de pluies observées par les radars météorologiques. Cette méthode est utilisée par deux web-services : Vigicrues Flash qui couvre le territoire de la France Métropolitaine (de Saint-Aubin et al., 2016) et RHYTMME qui couvre la région Provence-Alpes-Cote-d'Azur (Westrelin et al., 2012).

D'autres systèmes pour l'avertissement des crues, reposant sur une modélisation hydrologique distribuée, ont été développés au niveau européen (European Flood Alert System (Thielen et al., 2008), HYDRATE (Borga et al., 2011)), au Royaume-Uni (Modèle Grid-to-Grid (G2G), (Bell et al., 2007)) et aux États-Unis (FLASH, (Gourley et al., 2017)).

Les développements de chaînes de prévisions plus complètes couplant la prévision de l'aléa avec l'évaluation des enjeux exposés et des vulnérabilités associées sont en cours. Néanmoins leurs développements sont encore à l'état de recherche (projet de recherche ANR PICS (Payrastra, 2018)).

1.3 Le développement des modèles distribués pour la prévision des crues rapides

Dans cette section, les avantages des modèles distribués pour la prévision des crues rapides sont explicités. Les modèles distribués opèrent à des échelles spatiales et temporelles fines. Cela présente un intérêt pour les systèmes d'avertissement des crues rapides. Ces modèles se sont développés grâce à la disponibilité croissante des données spatiales telles que les pluies, les modèles numériques de terrain, l'évapotranspiration, les températures.

1.3.1 Le principe des modèles distribués

Contrairement aux modèles globaux qui considèrent un bassin-versant comme une entité unique (ils décrivent les réponses hydrologiques "moyennes" du bassin-versant), les modèles distribués ont été développés pour utiliser les données spatialisées comme celles mesurées par les radars météorologiques ou par les satellites. Les modèles distribués décrivent les processus hydrologiques locaux. Ils sont discrétisés spatialement sur un maillage (figure 1.3). Sur chaque maille un modèle hydrologique est appliqué. Ce modèle dépend des variables d'entrées (valeurs

des paramètres, pluies, évaporation...) disponibles à la même résolution que celle du maillage. Les variables de sorties du modèle hydrologique (débits, états des sols...) sont calculées pour chacune des mailles. Enfin, les mailles sont connectées entre elles afin de modéliser les phénomènes de transferts/routages (écoulements amont-aval, échanges souterrains...).

Par exemple, les modèles hydrologiques GRD (Javelle et al., 2010) et G2G (Bell et al., 2007), opérationnels pour les systèmes d'avertissement aux crues respectivement en France (Vigicrues Flash) et au Royaume-Uni, utilisent un maillage carré dont la résolution correspond à celle des données d'entrées spatialisées (pluies observées par les radars météorologiques, pluies prévues par les modèles météorologiques...). Les processus hydrologiques sont décrits au niveau de chaque maille par des opérateurs conceptuels ou physiques. En plus des processus hydrologiques, les modèles distribués intègrent un modèle de propagation hydraulique, appelé modèle de routage. Ce modèle permet de transférer les flux d'eau de maille en maille, de l'amont vers l'aval du bassin-versant. La modélisation hydraulique est simplifiée, elle peut être conceptuelle (transfert instantané, modélisation d'un retard pur, hydrogramme unitaire, réservoir linéaire utilisé dans le modèle GRD) ou physique (onde cinématique basée sur les équations de Barré de Saint-Venant pour le modèle G2G).

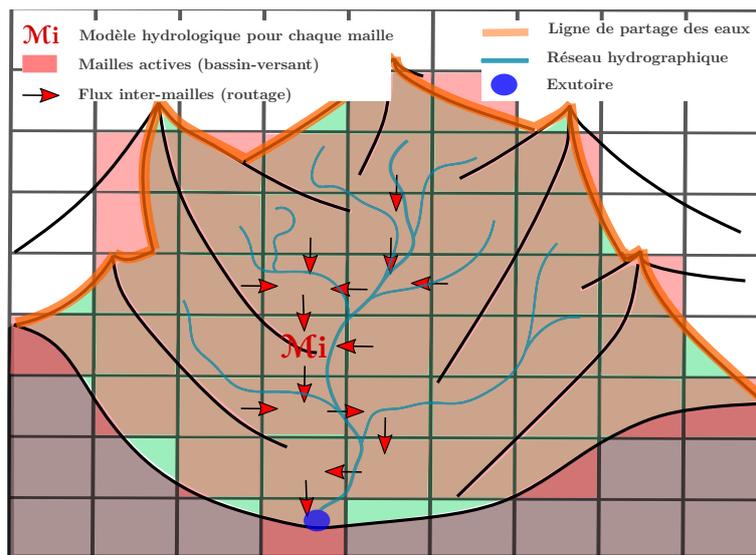


FIGURE 1.3 – Schéma d'une modélisation distribuée d'un bassin-versant à l'aide d'un maillage rectangulaire. Pour chaque maille, le modèle hydrologique \mathcal{M} transformant la pluie en débit est appliqué. Les débits sont propagés de l'amont vers l'aval à l'aide d'un modèle de routage.

Les modèles distribués sont donc plus complexes que les modèles globaux. Ils permettent de modéliser les processus et les spécificités hydrologiques locaux, de connaître les états du modèle spatialement (la précision dépend de la résolution du maillage) et de prendre en compte de nombreuses sources d'observations spatiales. Un inventaire non-exhaustif de modèles distribués, de leurs fonctionnements et applications est présenté dans le tableau 1.3.

TABLE 1.3 – Inventaire non-exhaustif de modèles hydrologiques distribués existant et exemples d’applications.

Nom	Applications	Types de modélisation	références
GRD	Système d’alerte opérationnel Vigicrue-Flash en France	Modèle à réservoirs conceptuel à 2 paramètres	Javelle et al. (2010)
G2G	Système d’alerte opérationnel au Royaume-Uni	Modèle conceptuel à 8 paramètres	Bell et al. (2007)
CREST	Système d’alerte opérationnel aux Etats-Unis	Modèle conceptuel à 10 paramètres	Wang et al. (2011)
LISFLOOD	EFAS (European Flood Alert System)	Modèle conceptuel	Thielen et al. (2008)
MARINE	Gardon d’Anduze	Equations physiques	Roux et al. (2011)

1.3.2 Les avantages des modèles distribués

L’utilisation de données spatiales

Depuis ces dernières décennies, les mesures satellitaires, les radars, les capteurs automatiques ont permis de récolter d’importantes quantités de données disponibles spatialement. Certaines de ces données sont utilisées par les modèles hydrologiques distribués :

- La mesure de la pluie par les radars météorologiques. Cette donnée est disponible en France à la résolution de $1km^2$. Les observations radars de la pluie permettent de prendre en compte leurs variabilités spatiales dans les modèles. En effet, la pluie est la donnée d’entrée principale des modèles hydrologiques adaptés aux crues. La connaissance de sa variabilité spatiale permet de modéliser les processus hydrologiques locaux et d’améliorer la prévision des crues ([Borga et al., 2011](#); [Lobligeois et al., 2014](#)).
- La mesure de la température au sol. Celle-ci est utilisée pour estimer l’évapotranspiration ([Oudin et al., 2005](#)). L’évapotranspiration peut-être intégrée dans les modèles hydrologiques. Sa modélisation est essentielle pour stabiliser temporellement le fonctionnement des modèles hydrologiques. On parle alors de modèle continu. En effet, l’évapotranspiration permet de modéliser les quantités d’eau évaporées depuis le sol. La saturation en eau des sols est un état déterminant qui influence la réponse hydrologique des modèles. Cet état, modélisé par les modèles continus, peut ainsi évoluer dans le temps : passage d’un sol saturé en eau (après une pluie) vers un sol sec (période sèche) et vice-versa.
- Les caractéristiques hydrogéologiques des sols. Ces données sont utilisées pour paramétrer les modèles selon les bassins-versants ([Pokhrel and Gupta, 2010](#)).
- La topographie des bassins-versants. Cette donnée permet de délimiter les bassins-versants et d’orienter l’écoulement de l’eau vers l’exutoire et paramétrer les vitesses de routage.

L'estimation des débits à fines échelles

Les modèles hydrologiques distribués fonctionnent à des échelles spatiales variables. Ils sont discrétisés dans l'espace sur un maillage. La taille des mailles permet d'ajuster le niveau de précision souhaité pour modéliser les processus de transformation pluie-débit. Plus la maille choisie est petite, plus la modélisation permet de représenter les processus hydrologiques à petites échelles (petits bassin-versants, talweg, délimitation précise du contour des bassins-versants...). A l'opposé, le choix d'une grande maille sera adapté pour modéliser de grands bassins-versants sur lesquels les processus locaux ne sont pas prépondérants.

La modélisation des débits opérée par les modèles distribués à fines échelles spatiales est adaptée pour la conception des systèmes d'alerte pour les crues rapides puisque la prévision des débits est disponible spatialement à la résolution du maillage. Cela permet de fournir des prévisions de débits pour de nombreux exutoires (petits bassin-versants sur lesquels aucune observation de débit n'est disponible). Par exemple, le modèle GRD de Vigicrue-Flash ([Javelle et al., 2016](#)), opère sur un maillage carré de résolution $1km^2$ et permet de fournir des prévisions de débits au niveau de 80000 exutoires en France métropolitaine.

La variabilité spatiale des caractéristiques et de la dynamique des bassins-versants

Les modèles distribués permettent de mieux prendre en compte la dynamique des bassins-versants. Ils intègrent la morphologie des bassins-versants tels que les vallées, les versants et les affluents. Ils permettent également de différencier spatialement les processus de transformation pluie-débit en fonction des caractéristiques géomorphologiques (type de sol et de sous-sol, occupation des sols, pentes...). Cependant le choix du paramétrage de ces modèles s'avère délicat. La prise en compte de la variabilité spatiale des caractéristiques des bassins-versants nécessite d'estimer une valeur de chaque paramètre pour chaque maille du modèle.

1.3.3 Difficultés de mise en œuvre des modèles distribués

Les modèles distribués sont plus difficiles à mettre en œuvre que les modèles globaux. Ils nécessitent davantage de données pour fonctionner, ils sont difficiles à calibrer et sont plus "coûteux" en temps de calcul :

- Les données nécessaires pour le fonctionnement de ces modèles sont spatialisées et doivent être interpolées ou extrapolées sur la résolution du maillage choisi pour la modélisation. Ces données sont volumineuses (grilles de pluie pour chaque pas de temps) et requièrent des pré-traitements (calcul de l'évapotranspiration spatiale, calcul des directions d'écoulement à l'aide d'un modèle numérique de terrain).
- Les paramètres doivent être estimés pour chacune des mailles. Pour les modèles à base physique, leurs paramètres sont estimés sur la base de données mesurables (type de sol et sous-sol, humidité des sols, occupation des sols, pentes...). Pour les modèles conceptuels les paramètres ne sont pas mesurables et nécessitent d'être calibrés. Ce calage est difficile.

En effet, un très grand nombre de variables sont à estimer spatialement alors que les observations de débits ne sont disponibles que très localement. Cela pose des problèmes d'identification (algorithmes de calage inadaptés, existence d'une multitude de solutions, instabilité des performances du modèle selon les régimes hydrologiques).

- Les temps et les ressources de calcul sont augmentés. En effet, les quantités de données utilisées par ces modèles nécessitent une capacité de stockage en mémoire importante. De plus, les équations représentant les processus hydrologiques sont calculées pour chaque nœud du modèle. Selon la discrétisation spatiale et temporelle choisie, les temps de calcul deviennent importants.

1.4 Contexte, questions scientifiques et objectifs

Le développement des systèmes de prévision des crues soudaines est un véritable défi en raison du grand nombre de bassins-versants impactés, de leurs tailles variables et de leurs temps de réponse aux pluies pouvant être très courts. Les crues soudaines peuvent s'avérer dangereuses et destructrices. Les systèmes récemment déployés en France et dans le monde, tel que Vigicrues-Flash, sont largement perfectibles. En effet, leurs couvertures géographiques sont partielles et la modélisation des aléas et des enjeux reste encore incomplète.

Le projet ANR (Agence National de la Recherche) PICS (Prévisions et Impacts des Crues Soudaines) vise à concevoir et évaluer des chaînes de prévision des crues soudaines offrant jusqu'à 6 heures d'anticipation. Ces prévisions sont constituées d'un chaînage de modèle permettant successivement de :

1. Prévoir les pluies et les débits.
2. Modéliser les zones inondées.
3. Modéliser les impacts.

Ce projet est coordonné par l'Ifsttar et met en interaction des équipes scientifiques et acteurs opérationnels aux compétences variées (météorologues, hydrologues, hydrauliciens, économistes, sociologues, sécurité civile, autorités locales, compagnies d'assurance, gestionnaires d'ouvrages hydroélectriques et de réseaux de transport) provenant d'instituts différents (CCR, Cerema, CNRM, Géosciences Rennes, Irstea, IGE, SCHAPI...).

Une description complète du projet PICS est consultable sur le site internet : <http://pics.ifsttar.fr>. Les travaux de recherche présentés dans cette thèse contribuent au premier maillon du chaînage de modélisation : "Prévoir les pluies et les débits".

Dans la suite de cette partie les principales questions scientifiques et les objectifs de ces travaux de recherche sont présentés.

1.4.1 Développer un modèle distribué pour la prévision des débits aux exutoires non-jaugés

Les débits sont généralement estimés en temps réel à partir des mesures de hauteurs d'eau localisées aux stations de jaugeages. En France métropolitaine, environ 3000 stations de jaugeage automatiques permettent l'estimation des débits en temps réel sur le réseau hydrographique. Ces mesures sont centralisées par le SCHAPI (Service Central d'hydrométéorologie et d'Appui à la Prévision des Inondations) et sont mises à disposition sur le site internet de la banque hydro (<http://www.hydro.eaufrance.fr/>).

Cependant la connaissance des débits des rivières est une information ponctuelle dans l'espace (localisée) et la plupart des stations de jaugeage sont localisées sur le réseau hydrographique principal. Ces mesures ne sont pas disponibles en tous points du réseau et ne concernent pas les petits cours d'eau (affluents et petits bassins-versants). Pourtant, dans le cadre de Vigicrue-Flash (système d'anticipation des crues rapides en France métropolitaine), il est nécessaire d'estimer les débits des rivières au niveau de nombreux exutoires choisis sur le réseau hydrographique secondaire (en amont et en aval de chaque commune) (Javelle et al., 2016, 2014). Dans le cadre de Vigicrue-Flash, ces débits sont donc estimés à l'aide d'un modèle hydrologique distribué sur un maillage carré de résolution $1km^2$. Dans la version opérationnelle du modèle, celui-ci ne possède pas d'opérateur de routage (Javelle et al., 2010). Pour chaque pas de temps, les débits sont calculés aux exutoires des bassins-versants en additionnant les contributions de l'ensemble des débits provenant des mailles amont. Ce choix de conception permet d'accentuer l'anticipation des crues par le modèle. Cependant, ce choix rend le calage des paramètres difficile : les hydrogrammes modélisés et observés ne sont pas synchronisés temporellement ; les observations provenant de plusieurs stations de jaugeage ne peuvent pas être simultanément utilisées pour le calage ; les paramètres calés à l'aide d'une station de jaugeage ne reflètent pas la dynamique des bassins-versants situés à l'amont.

Le premier objectif de ces travaux de recherche est d'améliorer le modèle hydrologique existant. Il s'agit de développer un modèle de routage permettant une continuité de la modélisation le long du réseau hydrographique, de l'amont vers l'aval du bassin-versant. Ce modèle de routage permettra alors d'estimer les débits en chaque point du maillage (et donc aux exutoires non-jaugés) tout en assurant la cohérence des paramètres le long du réseau.

1.4.2 Décrire les processus hydrologiques locaux en estimant la variabilité spatiale des paramètres des modèles distribués

Les modèles distribués ont été développés pour tirer profit des données spatiales. Par exemple, la modélisation hydrologique utilisée dans Vigicrue-Flash intègre les observations de pluie spatiales obtenues par les radars météorologiques. La prise en compte des pluies spatialisées dans les modèles permet d'améliorer l'estimation des débits par les modèles, notamment en région méditerranéenne où les précipitations sont très variables spatialement (Sangati and Borga, 2009; Lobligois et al., 2014). Cependant, le modèle utilisé dans Vigicrue-Flash est "conceptuel" et les processus hydrologiques modélisés sont extrêmement simplifiés. Les paramètres des

opérateurs mathématiques n'ont pas forcément un sens physique. Il est alors difficile d'estimer précisément la valeur de ces paramètres à l'aide de la connaissance des caractéristiques des bassins-versants (occupation des sols, géologie, pente...). Ces paramètres nécessitent donc d'être calibrés. Seulement peu de mesures *in-situ* des débits permettent de valider les sorties du modèle : le nombre de paramètres à calibrer est nettement supérieur au nombre de stations de jaugeage. Les paramètres du modèle utilisés dans Vigicrue-flash sont actuellement calibrés de manière uniforme (c'est-à-dire une unique valeur pour l'ensemble du maillage) à partir des mesures de débits aux stations hydrométriques (Javelle et al., 2010). La modélisation n'intègre pas correctement la variabilité spatiale des caractéristiques hydrologiques du bassin-versant. L'identification de la variabilité spatiale des paramètres semble être une source de progrès intéressante pour espérer simuler correctement les débits aux exutoires non-jaugés (situés à l'intérieur des bassins-versants jaugés). Ceci est l'objectif principal de ces travaux de recherche.

1.4.3 Choisir un algorithme pour estimer spatialement les paramètres d'un modèle distribué

Le calage des paramètres d'un modèle est possible si l'on connaît les variables d'entrées et de sorties de celui-ci. Pour une pluie spatiale observée sur une fenêtre temporelle délimitée, les meilleurs paramètres sont recherchés pour que les débits modélisés correspondent au mieux aux débits observés aux stations de jaugeage. Pour espérer caler l'ensemble des paramètres d'un modèle distribué, il faut que la fenêtre temporelle choisie soit suffisamment longue afin que des événements climatiques affectent à tour de rôle l'ensemble du bassin-versant et entraînent des réactions hydrologiques variées aux stations de jaugeage (crues, étiages).

Néanmoins, un tel calage est difficile. En effet, le modèle est "sur-paramétré" et le problème de calage n'admet pas une solution unique (Beven, 1993). De plus, les algorithmes de calage classique (algorithmes stochastiques probabilistes) ainsi que les autres algorithmes qui ne dépendent pas d'un calcul du gradient ("gradient-free methods") ne sont pas adaptés pour réaliser ce calage :

- Le nombre de variables à calibrer simultanément est trop grand.
- Les débits (sorties du modèle) ne sont pas connus (observés) en chaque maille du maillage. Au contraire, le nombre de stations de jaugeage permettant de mesurer le débit est très inférieur au nombre de mailles utilisées dans le modèle :
- Le choix d'une discrétisation temporelle et spatiale fine est coûteuse en ressources de calcul (temps, mémoire vive).
- Les performances des algorithmes "gradient-free" diminuent lorsque le nombre de variables à identifier augmente.

L'ensemble des méthodes de calage développées pour les modèles distribués met en œuvre des stratégies de régularisation (contraintes appliquées au calage) (De Lavenne et al., 2019). Par exemple, Anderson et al. (2006) tentent d'estimer des valeurs non-uniformes des paramètres d'un modèle distribué à partir des caractéristiques hydrologiques du bassin-versant (géologie, occupations et propriétés des sols). Pokhrel and Gupta (2010) utilise une technique qui consiste

à réduire le nombre de variables à calibrer. Les valeurs (à priori) de paramètres du modèle sont alors estimées à partir de la connaissance des caractéristiques hydrogéologiques des bassins-versants. Ensuite, à la place de caler une valeur de paramètre pour chaque maille, un super-paramètre global est calibré. Celui-ci modifie globalement les valeurs à priori des paramètres selon une loi choisie par le modélisateur (addition ou multiplication par une constante, exposant d'une loi puissance...).

Calibrer l'ensemble des paramètres d'un modèle distribué, sans appliquer de méthodes de réduction du nombre de variables, revient à résoudre un problème inverse de grande dimension. Ce type de problème peut être résolu à l'aide de méthodes d'assimilation de données variationnelles.

Les algorithmes d'assimilation de données sont utilisés depuis des décennies en géo-sciences (météorologie, océanographie, hydraulique et hydrologie). Ces méthodes sont utilisées pour estimer les états et les paramètres de modèles dynamiques décrivant des phénomènes naturels. Une synthèse des applications de ces algorithmes en hydrologie a été écrite par [Liu et al. \(2012\)](#). Le filtre de Kalman et le filtre de Kalman étendu sont les méthodes les plus utilisées ([Sun et al., 2015, 2016](#); [Quesney et al., 2000](#); [Munier et al., 2014](#)). Par exemple, [Munier et al. \(2014\)](#), montre que les performances du filtre de Kalman sur un modèle semi-distribué dépend du degré de spatialisation et du nombre de stations de jaugeage utilisé. [Sun et al. \(2016\)](#) utilise le filtre de Kalman étendu pour ajuster la variation moyenne spatiale des états d'un modèle distribué. Ces applications montrent que le choix de la méthode d'assimilation de données est limité à une approche variationnelle pour calibrer les paramètres d'un modèle distribué.

En estimation variationnelle, le meilleur jeu de paramètres est recherché en minimisant itérativement une fonction objectif. Ces algorithmes recherchent un minimum local autour d'un à priori (valeur initiale de la fonction objectif calculée avec le jeu de paramètre initial). Ces algorithmes requièrent le calcul du gradient de la fonction objectif par rapport aux variables recherchées. Pour des problèmes de petites dimensions, le gradient peut-être calculé par la méthode des différences finies ([Abbaris et al., 2014](#); [Thirel et al., 2010](#)). Pour les problèmes de grandes dimensions, le modèle adjoint permet de calculer le gradient de manière précise et efficace ([Errico, 1997](#)). L'utilisation des méthodes variationnelles impliquant le modèle adjoint est courante en météorologie et en océanographie. Cependant, en hydrologie seulement quelques travaux ont été réalisés dont ceux de [Castaings et al. \(2009\)](#) et [Nguyen et al. \(2016\)](#). Le modèle adjoint a été obtenu à partir de modèles à base "physique" utilisant des équations aux dérivées partielles (Green-Ampt / onde cinématique pour [Castaings et al. \(2009\)](#) et Horton / onde cinématique pour [Nguyen et al. \(2016\)](#)).

L'un des objectifs de ces travaux est d'implémenter et tester un algorithme de calage variationnel pour caler les paramètres distribués d'un modèle hydrologique distribué et conceptuel. Le gradient est calculé par le modèle adjoint. Le modèle adjoint est obtenu par différentiation automatique du code source du modèle par le logiciel Tapenade ([Hascoet and Pascual, 2013](#)).

Les développements associés au modèle hydrologique devront donc respecter les contraintes de différentiabilité.

1.4.4 Développer une méthodologie pour estimer les paramètres d'un modèle distribué à partir d'un algorithme de calage : méthode de validation et contraintes de calage

Le calage de l'ensemble des paramètres d'un modèle distribué est difficile : peu de points d'observation devant le nombre de paramètres, imperfections du modèle et des mesures observées (pluies, débits), sur-paramétrisation du modèle... Ces contraintes rendent le calage sensible aux problèmes "d'équifinalités", lorsque plusieurs jeux de paramètres induisent des performances de modélisation similaires aux exutoires de calage. Les performances du modèle en "validation", c'est-à-dire utilisant des jeux de données différents de ceux utilisés pour le calage, peuvent alors être détériorées. Ces problèmes d'identification ont été largement discutés dans les articles suivants : (Beven, 1993, 2001).

Les conditions de mise en œuvre d'un algorithme variationnel pour caler les paramètres d'un modèle distribué sont donc à choisir rigoureusement. Le jeu de paramètres calé doit permettre d'améliorer les performances du modèle en "validation". Ces travaux ont pour objectif de définir une méthodologie afin d'estimer et d'évaluer les paramètres calibrés. Celle-ci repose sur un processus de double calage/validation spatio-temporel. Les performances du modèle sont évaluées aux exutoires n'ayant pas servi au calage et dans des conditions climatiques différentes de celles choisies lors du calage (période temporelle différente). Le calage est réalisé sous certaines contraintes. Ces contraintes doivent permettre d'améliorer le calage en apportant une source d'information complémentaire aux données de débit servant au calage. Deux types de contraintes seront étudiées et testées : le choix du jeu de paramètre initial (appelé ébauche) et la corrélation spatiale des paramètres (contrainte de variabilité). Le choix de l'ébauche est très important puisqu'il conditionne la solution finale (les algorithmes variationnels recherchent un minimum local). Par exemple, l'ébauche peut être estimée à partir d'un calage global uniforme ou bien à partir des propriétés hydrogéologiques des bassins-versants. Différentes hypothèses seront testées.

1.4.5 Contribuer au développement du modèle hydrologique opérationnel du service VigiCrue-Flash

Enfin ces travaux contribuent à améliorer le modèle utilisé dans le système d'alerte aux crues en France, Vigicrue-Flash. Vigicrue-Flash est donc le point de départ de ces travaux : le modèle utilisé et développé pour ces travaux repose sur le modèle actuellement opérationnel ; les expériences de modélisation sont réalisées avec une discrétisation spatiale de $1km^2$ correspondant à la résolution choisie dans Vigicrue-Flash ; les jeux de données utilisés sont également similaires puisqu'il s'agit des données d'évapotranspiration calculées par la formule de Oudin et al. (2005) et des pluies observées issues des radars météorologiques. Les développements apportés au modèle devront aussi satisfaire les contraintes opérationnelles (temps de calcul réduit).

Chapitre 2

Préparation des jeux de données

Sommaire

2.1	Les sources de données	35
2.1.1	Les données spatiales hydro-météorologiques	35
	Les données de pluies	35
	Les données d'évapotranspiration potentielle	36
2.1.2	Les données caractéristiques des bassins-versants	36
	Le réseau hydrographique	36
	Les contours des bassins-versants	37
	Les directions d'écoulement	37
2.1.3	Les données spatiales utilisées pour contraindre les paramètres du modèle	37
	Les données d'occupation du sol	37
	L'indice de prédisposition à la production	39
	La densité de drainage des bassins-versants	40
	Les données hydrogéologiques	41
	Les pentes des bassins-versants	42
2.1.4	Les données de débits aux stations hydrométriques	43
2.2	Les bassins-versants étudiés	44
2.2.1	Contexte et zones géographiques étudiés	44
2.2.2	Sélection et critiques des stations hydrométriques	46
2.2.3	Maillage des bassins-versants pour la modélisation distribuée	51

Ce chapitre présente les sources de données et les bassins-versants étudiés. Dans une première partie l'ensemble des données hydrologiques, climatiques, météorologiques et géographiques utilisées pour ces travaux sont décrites. Dans une deuxième partie sont présentés les bassins-versants servant aux expériences de calage du modèle hydrologique distribué : les choix des régions étudiées sont brièvement explicités, les données hydrométriques sont critiquées et l'élaboration des maillages des bassins-versants est expliquée.

2.1 Les sources de données

2.1.1 Les données spatiales hydro-météorologiques

Les données de pluies

Les données de pluies proviennent de la réanalyse des images de réflectivité radar fournies par les radars météorologiques. Ces données sont fournies par Météo-France à la résolution spatiale de $1km^2$. L'estimation spatiale de la lame d'eau précipitée est calculée à partir des images de réflectivité radars. Celles-ci sont corrigées (réanalyses) à l'aide, entre autres, des cumuls de pluies enregistrés par les postes pluviométriques au sol. La qualité de l'estimation spatiale de la lame d'eau précipitée varie selon le degré de réanalyse des images radars brutes. Les données utilisées dans ces travaux proviennent des réanalyses disponibles le jour suivant la mesure (J+1). Il s'agit de la lame d'eau nommée "Antilope J+1" produite par Météo-France ([Champeaux, Jean-Louis et al., 2009](#)). Les autres lames d'eau disponibles sont :

- La lame d'eau "Antilope" est disponible 15 minutes après transmission de la mesure de la réflectivité radar. Cette donnée est fournie avec un niveau de réanalyse minimum. Les incertitudes sont fortes. Cette donnée est utilisée pour le suivi en "temps réel" des précipitations. L'historique de ces lames d'eau est disponible sur la période 2006-2018.
- La lame d'eau "Comephore" est disponible plusieurs années après la mesure de la réflectivité radar. Cette donnée intègre une réanalyse complète. L'estimation de la lame d'eau est de meilleure qualité. Cependant, cette donnée est disponible sur la période depuis 1997 jusqu'à 2017 mais n'est pas disponible pour les événements les plus récents.

La lame d'eau "Antilope J+1", est donc un bon compromis entre ces produits. Elle intègre un degré de réanalyse suffisant pour garantir une bonne précision de la lame d'eau estimée. Cette donnée peut-être utilisée pour étudier des événements récents. Par exemple, la figure 2.1 représente la lame d'eau radar "Antilope J+1" le 2011/11/05 entre 11 :00 heure et 12 :00 heure.

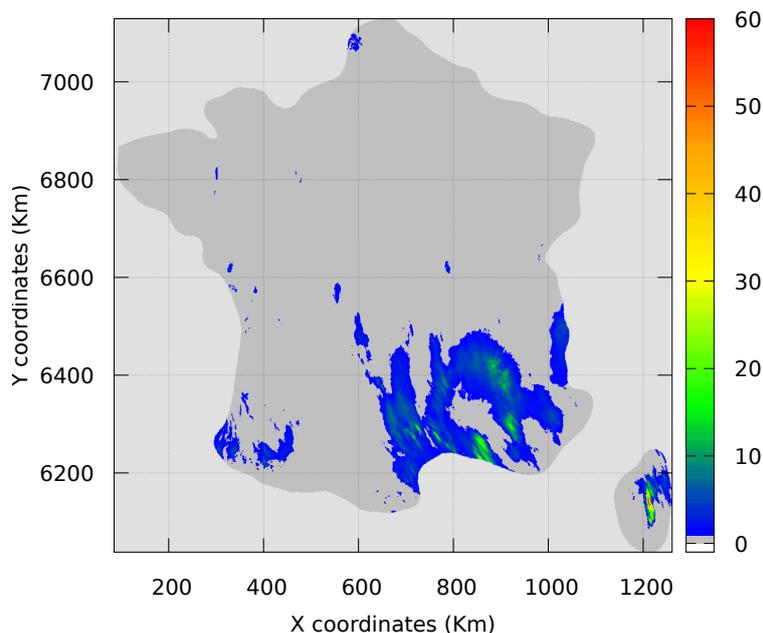


FIGURE 2.1 – Exemple de la lame d’eau radar Antilope J+1 donnant le cumul de pluie horaire en mm le 2011/11/05 entre 11 et 12 heures (source : Météo-France). De fortes pluies affectent les régions méditerranéennes et la Corse.

Les données d’évapotranspiration potentielle

L’évapotranspiration correspond aux phénomènes de transfert d’eau entre le sol et l’atmosphère (Penman and Keen, 1948). Ces phénomènes incluent : les quantités d’eau qui s’évaporent depuis la surface du sol ; les quantités d’eau qui sont interceptées par les végétaux et qui s’évaporent ; les quantités d’eau libérées par la transpiration des végétaux (respiration). L’évapotranspiration réelle est difficilement mesurable. Cette donnée peut être obtenue à partir de modèles mathématiques. On calcule alors l’évapotranspiration potentielle (notée ETP). Dans ces travaux, l’évapotranspiration potentielle est calculée à partir du modèle suggéré par Oudin et al. (2005). Ce modèle dépend de la température journalière moyenne et du rayonnement solaire moyen journalier.

2.1.2 Les données caractéristiques des bassins-versants

Le réseau hydrographique

Le réseau hydrographique présenté sur certains graphiques provient de la base de données CARTHAGE rendue disponible par la SANDRE (<http://www.sandre.eaufrance.fr>).

Les contours des bassins-versants

Les stations hydrométriques sont géolocalisées et définissent des exutoires de bassins-versants. Chaque bassin-versant est associé à un contour délimité par les lignes de crêtes (ligne de partage des eaux) et/ou les zones de drainage (les réseaux d'eau souterrains sont pris en compte pour délimiter la zone drainée par le bassin-versant). Ces données de contour proviennent de la base de données CARTHAGE rendue disponible par la SANDRE (<http://www.sandre.eaufrance.fr>).

Les directions d'écoulement

Le modèle hydrologique distribué utilisé dans ces travaux nécessite de connaître les directions d'écoulement associés au maillage des bassins-versants. Ces directions d'écoulement ont été fournies par Etienne Leblois (INRAE-Lyon) à la résolution spatiale de 1km^2 . Cette grille permet de connaître, pour chaque maille, la direction principale de drainage parmi huit directions. Pour chaque maille M, la direction de l'écoulement est codée suivant le schéma 2.2 :

1	2	3
4	M	5
6	7	8

FIGURE 2.2 – Codage des directions d'écoulement associé à la maille M.

2.1.3 Les données spatiales utilisées pour contraindre les paramètres du modèle

Des descripteurs ont été choisis à priori pour contraindre la variabilité spatiale des paramètres du modèle hydrologique distribué (voir chapitre 7). Ces descripteurs sont choisis à l'échelle de l'ensemble du territoire d'étude. Ils sont spatialisés (résolution 1 km^2) et décrivent localement les caractéristiques de la maille considérée. Ces descripteurs peuvent être de différentes natures : climatiques, géologiques, pédologiques, topographiques ou liés à l'occupation des sols. Les sources de données correspondantes sont présentées ci-dessous.

Les données d'occupation du sol

CORINE Land Cover (CLC) est un inventaire biophysique de l'occupation des sols (<https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>). Cet inventaire est produit par interprétation visuelle d'images prises par satellite. L'échelle de production est le 1/100 000. CLC permet de cartographier des unités homogènes d'occupation des sols d'une surface minimale de 25 ha. Cette base de données a été initiée en 1985. Des mises à jour ont été publiées régulièrement en 1990, 2000, 2006, 2012 et 2018. La publication de la base de données datant de 2012 a été choisie puisqu'elle correspond au milieu de la période climatique étudiée

(2008-2018). Cette base de données permet de connaître les types d'occupation du sol caractérisant les bassins-versants étudiés. Les types d'occupation des sols sont caractérisés par 3 niveaux de nomenclature. La nomenclature niveau 1 a été utilisée. Ce niveau de nomenclature contient 5 classes de type d'occupation des sols recensées dans le tableau 2.1. La figure 2.3 représente les caractéristiques d'occupation du sol par pixel de 1 km^2 centrée sur le bassin-versant du Gardon (Cévennes, France). L'occupation du sol à 1 km^2 est celle associée au type majoritairement présent dans le pixel.

TABLE 2.1 – Tableau décrivant les types d'occupation des sols décrits par le niveau 1 de la nomenclature de la base de donnée CLC 2012.

Type	Description
1	Territoires artificialisés
2	Territoires agricoles
3	Forêts et milieux semi-naturels
4	Zones humides
5	Surfaces en eau

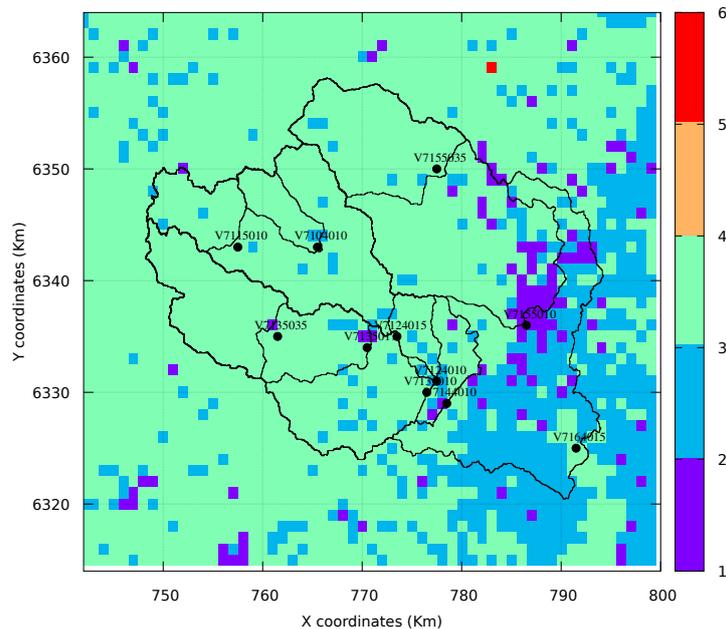


FIGURE 2.3 – Carte des types d'occupation des sols centrée sur le bassin-versant du Gardon (résolution 1 km^2). Les exutoires et les contours des bassins-versants associés sont tracés en noir.

L'indice de prédisposition à la production

L'indice de prédisposition à la production (Mahe, 2016) a été créé à l'aide des descripteurs utilisés par la méthode ANSETR (ANalyse de la SEnsibilité d'un Territoire au Ruissellement) (Dehotin and Breil, 2011). Ces descripteurs sont : la pente, l'occupation du sol, la perméabilité, l'érodabilité et la profondeur du sol. Pour chaque descripteur, la valeur 0 (peu ou pas productif), 1 (moyennement productif) ou 2 (très productif) est affectée selon les caractéristiques du milieu. Ensuite, pour chaque pixel, un score est calculé en additionnant les valeurs associées de l'ensemble des descripteurs utilisés. La figure 2.4 montre la carte de l'indice normalisé (indice variant entre 0 et 1) de prédisposition à la production obtenu pour la France métropolitaine.

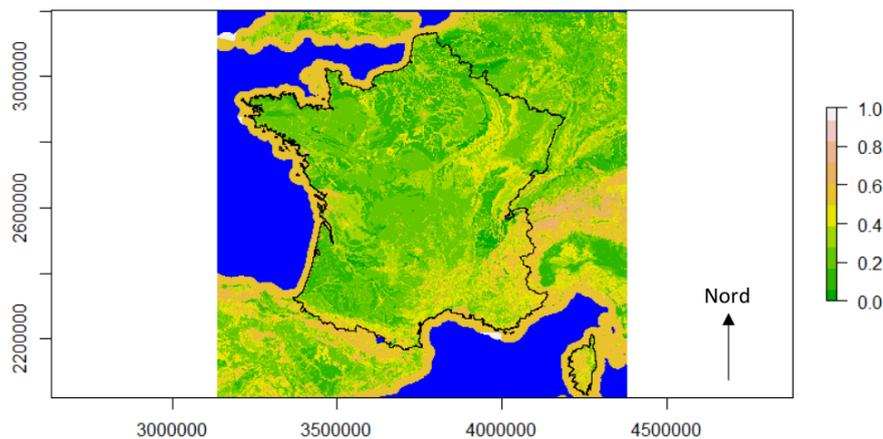


FIGURE 2.4 – Carte de prédisposition à la production en France Métropolitaine. La légende couleur correspond aux valeurs de l'index normalisé. Carte issue du rapport de stage Mahe (2016). Projection ETRS89.

Pour contraindre la variabilité spatiale des paramètres du modèle hydrologique, l'indice de production a été divisé en cinq classes croissantes. Par exemple, la figure 2.5 cartographie la variabilité de l'indice selon les cinq classes de production pour le bassin-versant du Gardon (Cévennes, France).

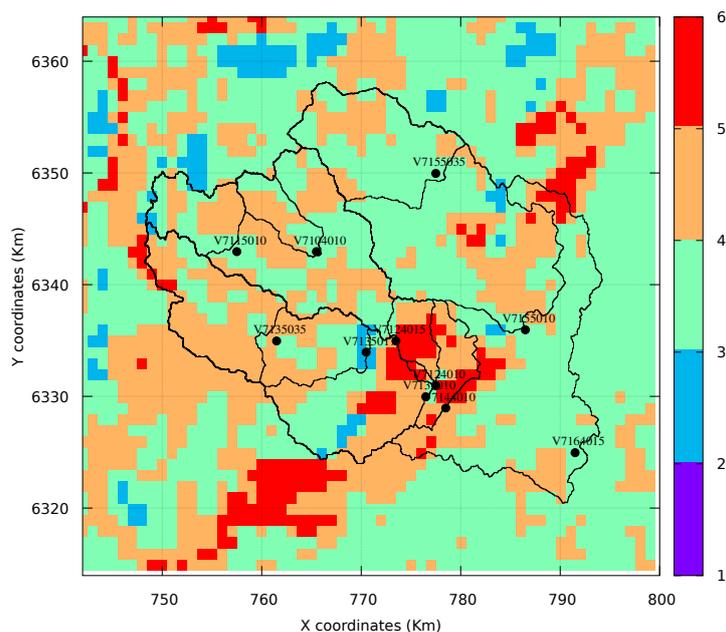


FIGURE 2.5 – Carte de prédisposition à la production centrée sur le bassin-versant du Gardon (résolution 1 km^2). L'indice normalisé ($[0, 1]$) est découpé en 5 classes de production croissantes (le code couleur représente le numéro des classes). Les exutoires et les contours des bassins-versants associés sont tracés en noir.

La densité de drainage des bassins-versants

La densité de réseau est calculé à partir du réseau hydrographique de la base de données Carthage (Organde, 2016). À partir d'un maillage de résolution 100m^2 , chaque maille est indiquée 1 si le réseau hydrographique existe dans la base de données Carthage, 0 sinon. Le calcul de la densité de drainage est ensuite réalisé sur un maillage de résolution 1 km^2 . Pour chaque maille d' 1 km^2 , les valeurs des 100 pixels sont additionnées si bien que la variable varie en théorie entre 0 et 100. La figure 2.6 représente la variabilité de la densité de drainage selon 5 classes de densité croissante (résolution 1 km^2) centrée sur le bassin-versant du Gardon (Cévennes, France).

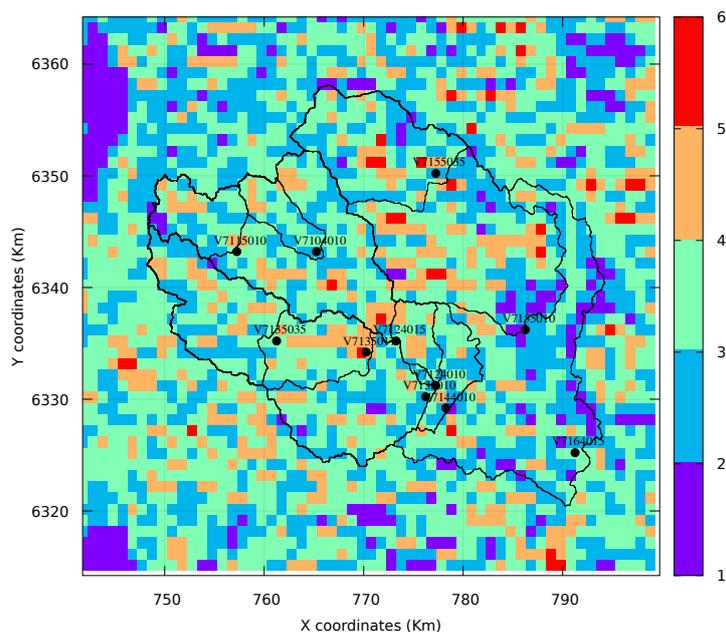


FIGURE 2.6 – Carte des densités de drainage centrée sur le bassin-versant du Gardon (résolution 1 km^2). L'indice de drainage ($[0 : 100]$) est découpé en 5 classes de densités de drainage croissantes (le code couleur représente le numéro des classes). Les exutoires et les contours des bassins-versants associés sont tracés en noir.

Les données hydrogéologiques

La base de données "Million-Géol" est la base de données géo-référencées de la carte géologique à $1/1000000$ (<http://infoterre.brgm.fr/formulaire/telechargement-carte-geologique-metropolitaine-11-000-000>). La carte géologique métropolitaine vectorisée est composée de 6 couches d'informations détaillées (formations géologiques, failles, points remarquables, isobathes, structures du magmatisme et du métamorphisme, grands ensembles structuraux et lithologiques). La couche "formation géologique" a été utilisée pour connaître la variabilité spatiale des formations géologiques qui composent les bassins-versants étudiés. La figure 2.7 représente sous forme d'un indice (résolution 1 km^2) la variabilité des couches géologiques centrée sur le bassin-versant du Gardon (Cévennes, France). Chaque indice représente une couche géologique différente.

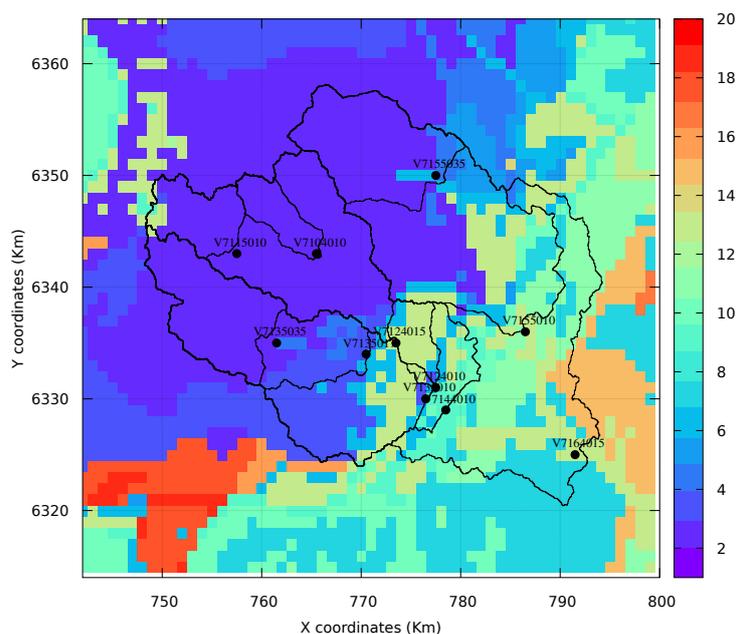


FIGURE 2.7 – Carte des différentes couches géologiques centrée sur le bassin-versant du Gardon (résolution 1 km^2). Chaque couleur représente une couche géologique différente. Les exutoires et les contours des bassins-versants associés sont tracés en noir.

Les pentes des bassins-versants

Les données de pentes proviennent de la base de données Copernicus (<http://land.copernicus.eu/in-situ/eu-dem-derived-products/slope>). Celles-ci sont disponibles à la résolution de 25 m^2 au niveau européen. Ces données ont été agrégées pour obtenir une résolution de 1 km^2 . Les pentes ont été classées selon cinq classes de pentes croissantes pour chaque bassin-versant étudié. La figure 2.8 représente la variabilité des pentes selon 5 classes croissantes (résolution 1 km^2) centrée sur le bassin-versant du Gardon (Cévennes, France).

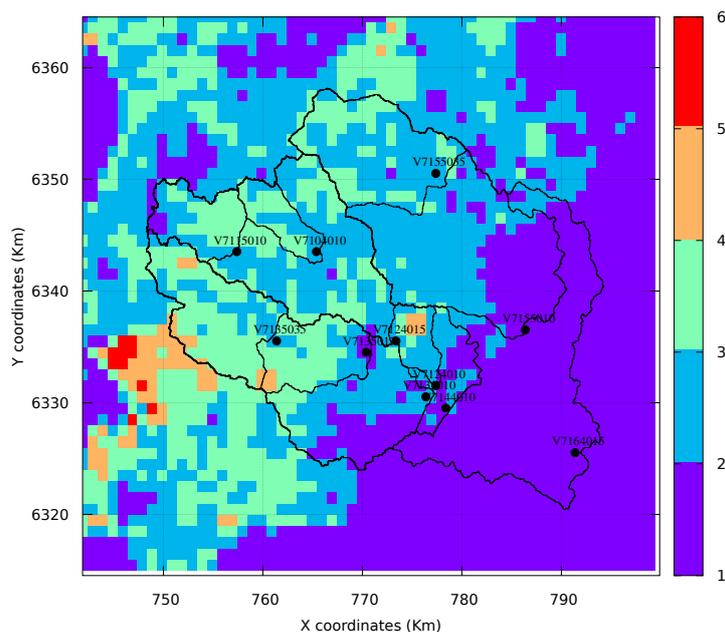


FIGURE 2.8 – Carte des pentes selon 5 classes croissantes centrée sur le bassin-versant du Gardon (résolution 1 km^2). Les exutoires et les contours des bassins-versants associés sont tracés en noir.

2.1.4 Les données de débits aux stations hydrométriques

La mesure des débits est une mesure indirecte qui est obtenue empiriquement par le protocole appelé "jaugeage" : la grandeur mesurée est le niveau de l'eau de la rivière. Cette mesure peut être réalisée en continu à l'aide de capteurs tel que le limnigraphe (limnigraphe à flotteur, à pression ou électronique). L'estimation du débit est obtenue en connaissant la hauteur d'eau et la vitesse d'écoulement en plusieurs points de la section. Les mesures des vitesses d'écoulement en travers de la section d'écoulement peuvent être réalisées sur le terrain à l'aide d'un moulinet ou d'un Doppler (appareil flottant qui mesure les vitesses d'écoulement en fonction de la profondeur le long de la section d'écoulement et permet une estimation directe du débit). Le calcul du débit de la rivière associé à différents niveaux d'eau permet de construire une relation entre le débit et la hauteur d'eau. Cette relation est appelée courbe de tarage. La construction de ces courbes nécessite le plus souvent des outils numériques qui permettent l'extrapolation de la courbe de tarage pour des forts débits (modélisation de l'écoulement de l'eau dans un bief par la formule de Manning-Strickler ou à l'aide de logiciels d'hydraulique résolvant les équations de Barré de Saint-Venant) et l'estimation des incertitudes (logiciel Baratin, (Le Coz et al., 2014)). C'est à partir des courbes de tarage que les débits sont estimés en continu aux stations hydrométriques couramment appelées stations de jaugeage. En France métropolitaine, un réseau de mesures entretenu par les DREAL permet d'estimer en

continu les débits des principaux cours d'eau. Ces mesures sont centralisées par le SCHAPI et accessibles en temps réel via le web service <http://www.hydro.eaufrance.fr/>. Actuellement environ 3000 stations de jaugeage sont opérationnelles et permettent la surveillance des principaux cours d'eau en France. Le pas de temps des mesures est variable. La fréquence des mesures dépend de la variabilité de la grandeur mesurée (ici la hauteur d'eau) : plus la hauteur d'eau varie, plus les mesures sont rapprochées (quelques minutes), inversement moins la hauteur d'eau varie, plus les mesures sont distantes temporellement (de quelques heures à quelques jours). Pour être exploitées, ces données sont souvent interpolées au pas de temps souhaité.

Comme toutes grandeurs mesurées, celles-ci sont parfois entachées d'incertitudes et d'erreurs. La mesure du débit peut-être influencée à différents niveaux : panne du capteur, modification de la section d'écoulement (détarage), branchage ou autres objets pouvant obstruer le champ visuel du capteur. Les incertitudes sur ces mesures sont donc nombreuses et doivent être prises en compte (au moins partiellement) lors de leur exploitation.

2.2 Les bassins-versants étudiés

2.2.1 Contexte et zones géographiques étudiés

Quatre zones géographiques d'étude ont été identifiées dans le cadre du projet PICS (projet ANR présenté au chapitre 1 section 1.4). Ces zones, représentées sur la figure 2.9, sont les suivantes :

1. La rivière du Loing et la rivière de l'Yonne (amont).
2. Le département de l'Aude.
3. Les départements de l'Hérault, du Gard et de l'Ardèche
4. Les départements du Var et des Alpes-Maritimes.

Ces quatre zones ont été sélectionnées à partir d'événements hydrologiques les ayant affectées entre 2010 et 2019 (figure 2.10). Ces événements hydrologiques présentent les caractéristiques suivantes :

- Des réactions d'intensités variées et des dégâts significatifs
- Des données riches quant aux zones inondées et impacts associés
- Un caractère très récent (2018) permettant d'appliquer les dernières méthodes de prévision de pluies.

Pour chacun des événements, une synthèse des données mobilisables a été réalisée (figure 2.11). Il s'agit des données de pluies observées (type de lames d'eau), de prévision de pluie, de débits (stations de jaugeage, mesure in-situ ou post-événement via le programme Hymex), de zones inondées et des impacts (assureurs, décès, interventions et témoignages).

Les modèles hydrologiques utilisent la pluie comme donnée d'entrée et estiment les débits (variable de sortie). Les principaux jeux de données utilisés sont donc les pluies (paragraphe 2.1.1) et les débits observés aux stations de jaugeage (paragraphe 2.1.4). Les données de débits

nécessitant une attention particulière. C'est à l'aide de ces données que les paramètres du modèle hydrologique distribué sont estimés.

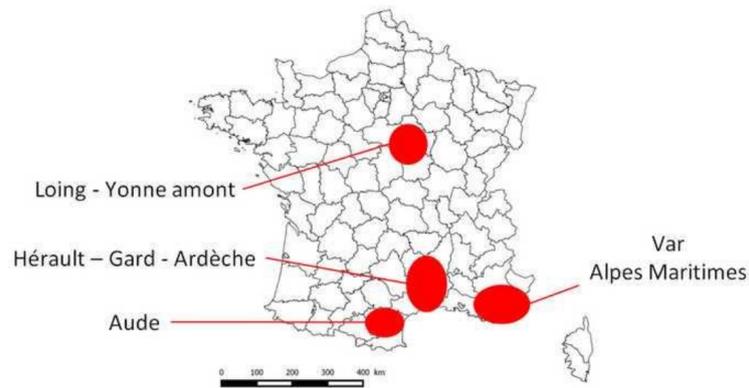


FIGURE 2.9 – Les secteurs d'étude identifiés lors du premier atelier utilisateurs PICS (source : PICS Newsletter 1).

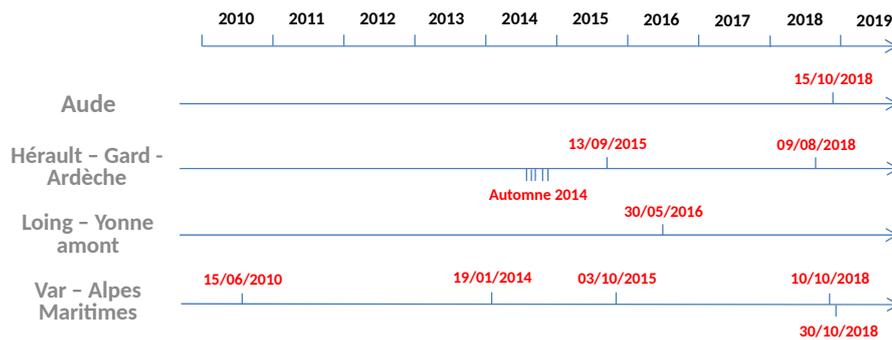


FIGURE 2.10 – Dates des évènements sélectionnés sur chaque secteur d'étude (source : PICS Newsletter 1).

Événement	Bassins concernés	Pluie (observation)		Pluie (prévision)	Débits	Inondation				Impacts						
		panthere 5min	antlope 15 min (Tréf et/ou Tdiff)	antlope 1h (Tréf et/ou J+1)		arome PI	ensembles de prévision immédiate	Stations	Hymex	PHE	enveloppe ZI	Sinistres assureurs	damagis (dégâts)	follow'in (décès)	interventions SDIS	Témoignages (enquêtes en ligne)
15-16/06/2010	Argens	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19/01/2014	Argens et Marevenne/Pansard	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
17-20/09/2014	Orb, Hérault, Vidourle, Gardons, Chassezac	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
29/9 et 6-7/10/2014	Lez - Mosson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10-13/10/2014	Vistre, Gardons, Cèze, Ardèche	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
3-4/11/2014	Ardèche	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
26-28/11/2014	Orb amont	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
12-14/09/2015	Orb et Hérault amont, Cèze, Ardèche	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
03/10/2015	Brague, Frayères, Argentière	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28/05-10/06/2016	Loing, Yonne amont	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
09/08/2018	Ardèche et Cèze aval	0	0	0	0	X	0	0	0	0	0	0	0	0	0	X
10-11/10/2018	Garonne et Preconil	0	0	0	0	X	0	0	0	0	0	0	0	0	0	X
14-15/10/2018	Aude intermédiaire	0	0	0	0	X	0	0	0	0	0	0	0	0	0	X
29/10-01/11/2018	Argens	0	0	0	0	X	0	0	0	0	0	0	0	0	0	X

FIGURE 2.11 – Données mobilisables par évènement et pour chaque bassin-versant concerné (source : PICS Newsletter 1).

2.2.2 Sélection et critiques des stations hydrométriques

Pour chaque zone géographique, des bassins-versants délimités à l'aval par une station de jaugeage (2.1.4) sont sélectionnés. Ces bassins-versants et stations de jaugeages doivent satisfaire les critères d'éligibilité suivants :

- Chaque bassin-versant sélectionné doit inclure au moins une autre station de jaugeage amont (à l'intérieur de son domaine). Les données de débits mesurées aux stations de jaugeage "amont" serviront de "validation" : ces données ne seront pas utilisées pour le calage, mais permettront de vérifier la justesse (précision) des débits prédits par le modèle hydrologique à l'intérieur du bassin-versant.
- Les surfaces des bassins-versants sélectionnés doivent être inférieures à 3000 km^2 : ce seuil permet de sélectionner les bassins-versants suffisamment petits pour être exposés aux crues éclair et suffisamment grands pour inclure une ou plusieurs stations de jaugeages amont. Aussi, ce seuil est suffisamment élevé pour composer un échantillon de bassins-versants de surfaces variées.
- Les stations de jaugeage situées à l'aval de chaque bassin-versant doivent avoir été en service sur l'ensemble de la période 01/01/2008-01/01/2018. Ce critère permet de garantir des chroniques de débits observés les plus complètes possibles sur cette période de 10 ans. De plus, cette période est découpée en deux sous-périodes de cinq ans pour des expériences de calage-validation croisées (voir chapitre 5).
- Les données de débit enregistrées doivent être disponibles et validées au moins 80% du

temps sur la période 01/01/2008-01/01/2018, soit moins de 20% de lacune. En effet, le calage des paramètres du modèle distribué sera meilleur si un grand nombre d'observations est disponible sur la fenêtre temporelle choisie pour le calage (diversité des événements hydrologiques). Le schéma 2.12 représente les critères d'éligibilité des bassins-versants.

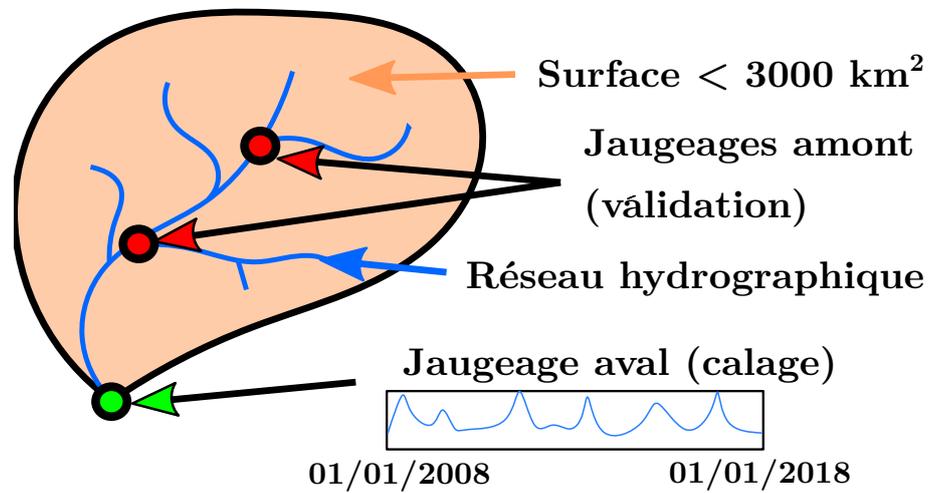


FIGURE 2.12 – Schéma représentant un bassin-versant et son réseau hydrographique éligible aux critères de sélection : surface, nombre de stations de jaugeage et disponibilité des mesures de débit

Vingt-trois méta-bassins-versants sont éligibles aux critères de sélection sur les quatre zones géographiques étudiées. Ces bassins-versants sont listés et numérotés dans le tableau 2.2. La carte de localisation 2.13 permet de les situer en France.

TABLE 2.2 – Liste des 23 méta-bassins-versants étudiés. Le nom de la rivière, la surface du bassin-versant à l'exutoire aval, le code de la station hydrométrique situé à l'exutoire aval et le nombre de stations internes à chaque bassin-versant sont mentionnés.

Numéro	Nom de la rivière	Surface (km^2)	Code	Nombre d'exutoires internes
1	Le Loing	2302	H3201010	9
2	La Cance	381	V3524010	4
3	Le Doux	621	V3744010	3
4	L'Eyrieux	649	V4154010	4
5	L'Ardèche	2264	V5064010	10
6	La Cèze	1112	V5474015	7
7	Le Gardon	1093	V7164015	11
8	L'Aude	1828	Y1232010	12
9	Le Fresquel	935	Y1364010	5
10	L'Orbiel	242	Y1415020	3
11	L'Orbieu	589	Y1564010	4
12	La Cesse	251	Y1605050	2
13	L'Hérault	2208	Y2332015	13
14	L'Orb	1336	Y2584010	12
15	Le Lez	168	Y3204040	4
16	Le Vidourle	503	Y3444020	5
17	Le Vistre	496	Y3534010	2
18	Le Gapeau	535	Y4624010	7
19	L'Argens	2512	Y5312010	11
20	La Giscle	201	Y5444010	3
21	La Siagne	492	Y5534030	6
22	Le Loup	289	Y5615030	2
23	L'Estéron	442	Y6434010	2

118

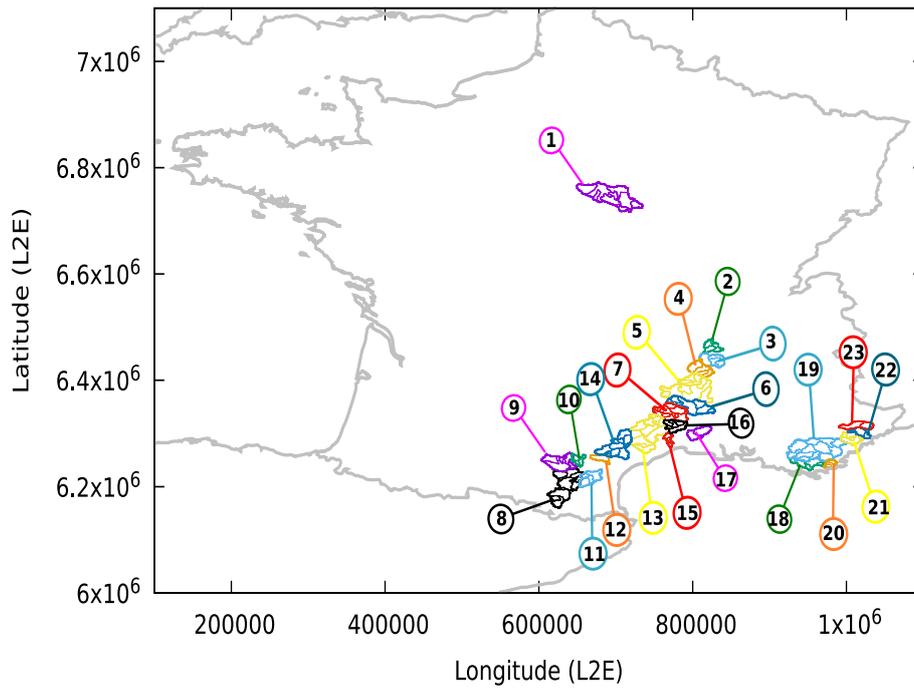


FIGURE 2.13 – Localisation des bassins-versants étudiés en France métropolitaine. Les numéros bassins-versants correspondent à la numérotation indiquée dans le tableau 2.2.

Les chroniques de débits observés aux stations aval de ces 23 méta-bassins-versants sont critiquées à partir d’une étude graphique de celles-ci. Les résultats de ces critiques sont présentés au tableau 2.3. Ces données de débits servent au calage des paramètres du modèle distribué. Leurs qualités influenceront significativement les résultats. Les autres stations de jaugeage localisées à l’intérieur des 23 bassins-versants serviront de validation pour vérifier la prédictibilité des débits à l’intérieur de chaque bassin-versant.

Les graphiques 2.14, 2.15, 2.16, 2.17 illustrent certaines critiques énoncées pour les chroniques de débits des 23 stations de jaugeage retenues.

Stations	Critiques
H3201010 : Le Loing à Challette sur Loing	La chronique est complète.
V3524010 : La Cance à Sarras	La chronique est complète.
V3744010 : Le Doux à Tournon sur Rhône	La chronique est complète.
V4154010 : L’Eyrieux aux Ollières sur Eyrieux	Les débits semblent très influencés en basses eaux. La chronique présente de longues périodes avec des débits constants aux alentours de $10 \text{ m}^3/\text{s}$. Seul les pics de crues sont visibles avec des pointes très raides (crues-décrués).

Stations	Critiques
V5064010 : L'ardèche à Saint-Martin d'Ardèche	La chronique est complète.
V5474015 : La Cèze à Bagnols-sur-Cèze	La chronique est globalement complète. On observe quelques lacunes. On remarque des montées de crues très rapides.
V7164015 : Le Gardon à Ners	La chronique est globalement complète. On observe quelques lacunes.
Y1232010 : L'Aude à Carcassonne pont-neuf	La chronique est complète. Les débits sont très influencés en étiage : ils oscillent rapidement entre 5 et 10 m^3/s .
Y1364010 : Le Fresquel à Carcassonne	La chronique est complète.
Y1415020 : L'Orbiel à Bouilhonnac	La chronique est complète.
Y1564010 : L'Orbieu à Luc sur Orbieu	La chronique est globalement complète. On observe quelques lacunes.
Y1605050 : La Cesse à Mirepeisset	La chronique est globalement complète. On observe quelques lacunes. Les débits semblent un peu influencés en basses eaux. De plus on observe plusieurs régimes hydrologiques en 10 ans.
Y2332015 : L'Hérault à Montagnac	La chronique est globalement complète. On observe quelques lacunes. Les débits sont influencés en basses eaux : ils oscillent rapidement autour de 1 m^3/s .
Y2584010 : L'Orb à Bézier	La chronique est complète.
Y3204040 : Le Lez à Latte	La chronique est complète.
Y3444020 : Le vidourle à Viq le Fesq	Beaucoup de données absentes entre janvier 2008 et octobre 2009. Sur cette période seules les crues sont visibles. On observe aussi un débit constant de 10 m^3/s pendant la période estivale 2009, ce qui nous paraît impossible. A partir d'octobre 2010 les données semblent de bien meilleure qualité.
Y3534010 : Le Vistre au Cairlar	La chronique est incomplète. On observe beaucoup de lacunes avant 2012.
Y4624010 : Le Gapeau à Hyères	La chronique est complète. On observe quelques valeurs de débits anormalement faibles lors de l'été 2016.
Y5312010 : L'Argens à Roquebrune sur Argens	La chronique est complète. On observe des valeurs de débits "bruités" et plus faibles durant l'année 2008 que sur le reste de la période.
Y5444010 : Le Giscle à Cogolin	La chronique est incomplète. Les débits deviennent nuls durant certaines périodes.

Stations	Critiques
Y5534030 : La Siagne à Pégomas	La chronique est complète. Cependant les données sont extrêmement "bruitées" dans la gamme 1-30 m^3/s . Seules les pointes de crues semblent bien représentées.
Y5615030 : Le Loup à Villeneuve-Loubet	La chronique est complète.
T6434010 : L'Estéron au Broc	La chronique est globalement complète. On observe quelques lacunes.

TABLE 2.3 – Critiques des chroniques de débits aux exutoires aval des 23 groupes de bassins-versants.

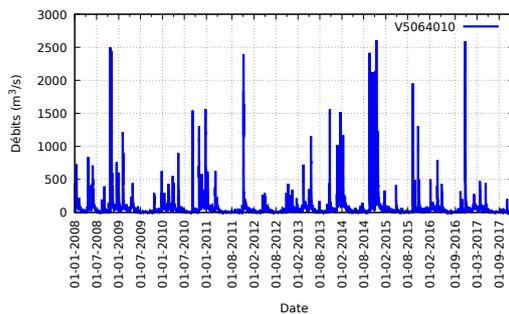


FIGURE 2.14 – Chronique de débits complète mesurée sur l'ardèche à Saint-Martin d'Ardèche

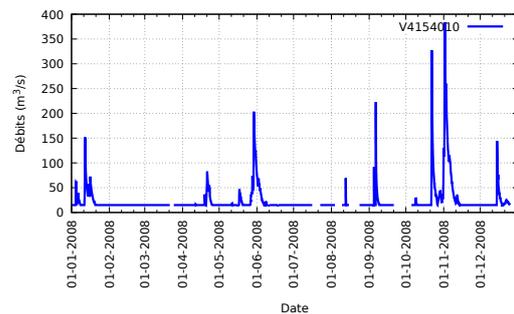


FIGURE 2.15 – Chronique de débits influencée mesurée sur l'Eyrieux aux Ollières sur Eyrieux.

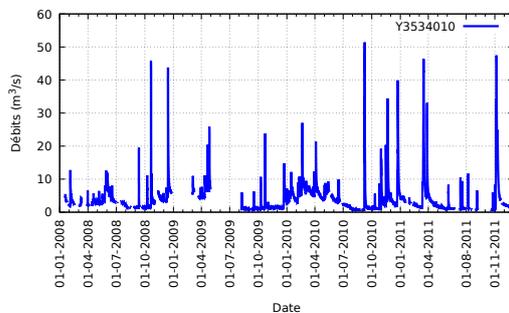


FIGURE 2.16 – Chronique de débits incomplète mesurée sur le Vistre au Caillar

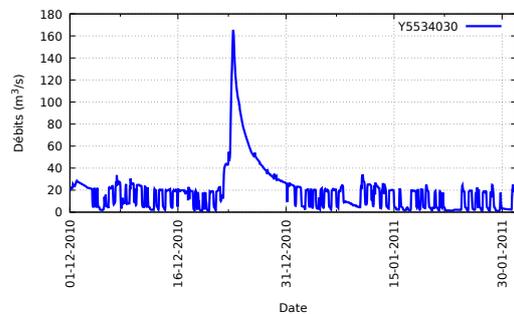


FIGURE 2.17 – Chronique de débits "bruitée" en basses eaux mesurée sur la Siagne à Pégomas

2.2.3 Maillage des bassins-versants pour la modélisation distribuée

Le modèle hydrologique distribué utilisé dans ces travaux fonctionne sur un maillage carré de résolution 1 km^2 . Le maillage utilisé pour la modélisation des 23 méta-bassins-versants est construit à partir de la grille des directions d'écoulement (donnée présentée au paragraphe 2.1.2). A partir de cette donnée, une grille de résolution 1 Km^2 indiquant les surfaces drainées est cal-

culée. Les stations de jaugeage sélectionnées sont géo-référencées (position connue). Le travail consiste à repositionner ces stations sur le maillage utilisé par le modèle (sélection de la maille correspondant à la station de jaugeage). Ce positionnement est réalisé pour chaque station à l'aide d'un "buffer" autour de la géo-localisation exacte des stations de jaugeage. Celui-ci recherche la maille pour laquelle l'écart entre la surface du bassin-versant calculé à partir de la grille d'écoulement 1 km^2 (surface drainé) et sa surface "réelle" (indiquée par la base de donnée CARTHAGE) est minimal. Le positionnement est réalisé avec succès si l'écart entre la surface "modélisée" et "réelle" est inférieur à 10%. En cas d'échec la station de jaugeage n'est pas utilisée.

Les 23 méta-bassins-versants sont délimités à l'aval par les stations de jaugeage présentées dans le tableau 2.3. 118 autres stations de jaugeages sont réparties à l'intérieur de ces 23 méta-bassins-versants. La figure 2.18 indique le nombre de sous-bassins-versants jaugés contenus à l'intérieur de chaque méta-bassin-versant. Au total les données de débit de 141 stations de jaugeages sont utilisées : 23 en calage et 118 en validation (stations situées à l'intérieur des méta-bassins-versants).

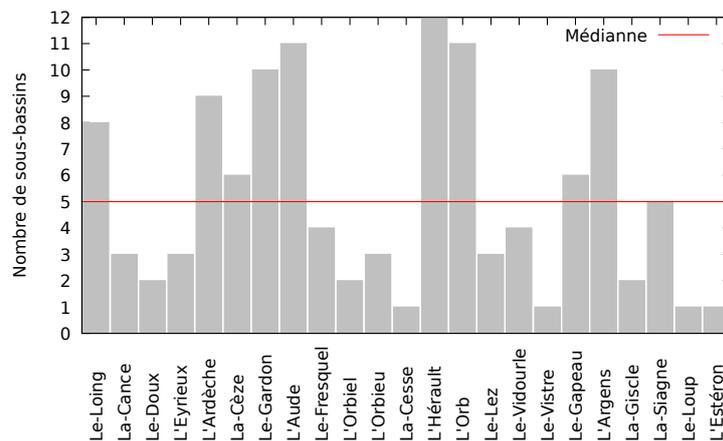


FIGURE 2.18 – Nombre d'exutoires et de sous-bassins-versants à l'intérieur de chaque méta-bassin-versant.

Les cartes (figures 2.19 et 2.20) représentent les bassins-versants du Loing et de l'Aude. Les cartes des 23 méta-bassins-versants et la liste des exutoires sont fournis en annexe : A et B. Les cartes permettent de localiser les stations de jaugeages à l'intérieur des bassins-versants. Chaque exutoire délimite un sous-bassin-versant. Les surfaces des sous-bassins-versants sont très variables. La figure 2.21 montre la répartition de ces surfaces pour chaque méta-bassin-versant.

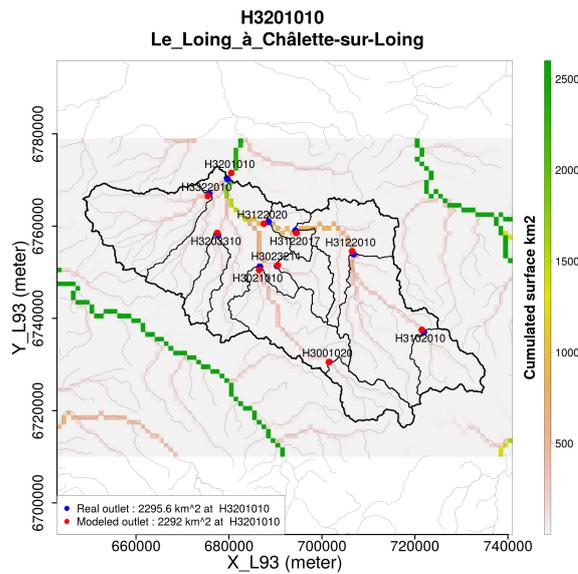


FIGURE 2.19 – Bassin-versant du Loing en amont de Chalette sur Loing. Le fond représente les surfaces drainées cumulées et calculées à partir de la grille de direction d'écoulement de résolution 1 km^2 . Les exutoires réels (point bleu) sont positionnés sur le réseau hydrographique alors que les exutoires simulés (points rouge) sont positionnés sur la grille d'écoulement.

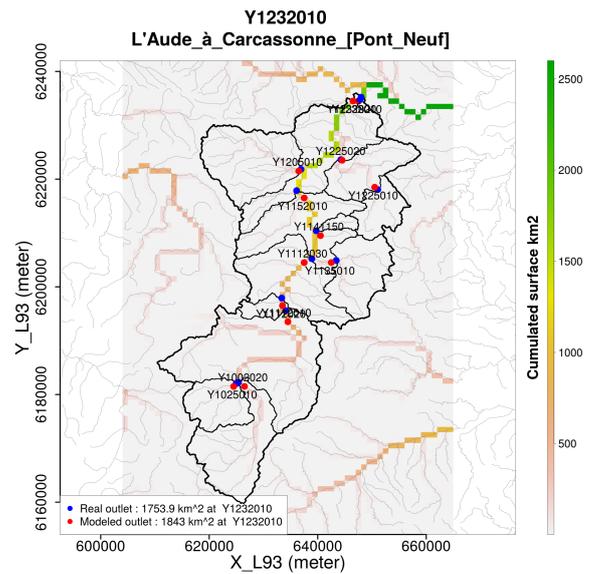


FIGURE 2.20 – Bassin-versant de l'Aude en amont de Carcassonne. Le fond représente les surfaces drainées cumulées et calculées à partir de la grille de direction d'écoulement de résolution 1 km^2 . Les exutoires réels (points bleus) sont positionnés sur le réseau hydrographique alors que les exutoires simulés (points rouges) sont positionnés sur la grille d'écoulement.

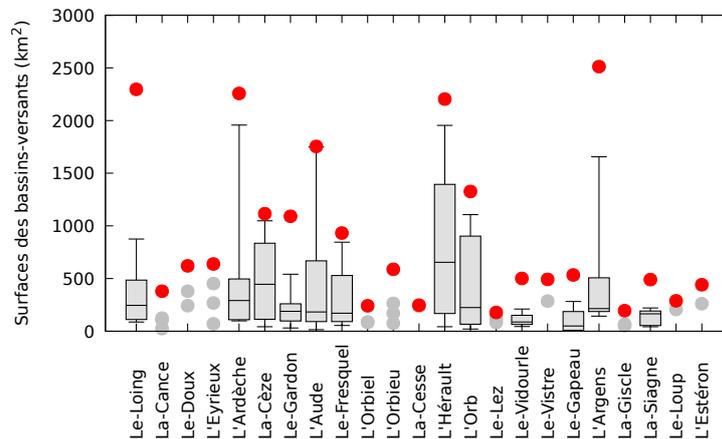


FIGURE 2.21 – Les boxplots montrent la répartition des surfaces des bassins-versants associés aux exutoires situés à l'intérieur des méta-bassins-versants. Les points rouges renseignent la surface de chaque méta-bassin-versant.

Chapitre 3

Le modèle hydrologique distribué

Sommaire

3.1	Présentation du modèle pluie-débit	55
3.1.1	L'opérateur de production	56
3.1.2	L'opérateur de transfert	57
3.2	Développement du modèle de routage : trois variantes proposées	58
3.2.1	Implémentation d'un schéma de routage de maille en maille	59
3.2.2	Modèles conceptuels modélisant un temps de retard	60
	Concepts	60
	Définition du noyau de convolution	61
	Choix du noyau de convolution	62
	Définition d'un modèle linéaire et non-linéaire	66
3.2.3	Description de la variante 1 : modèle de retard linéaire	67
3.2.4	Description de la variante 2 : modèle de retard non-linéaire	69
3.2.5	Description de la variante 3 : modèle à réservoir linéaire	71
3.2.6	Algorithme général et simplifié du modèle GRD	72
3.3	Étude des réponses hydrauliques élémentaires des trois modèles de routage	73
3.3.1	Protocole	73
3.3.2	Résultats :	75

Ce chapitre est consacré à la description du modèle hydrologique utilisé pour ces travaux. Le modèle hydrologique utilisé est un modèle distribué (l'espace est discrétisé à l'aide d'un maillage) et conceptuel (les opérateurs mathématiques utilisés pour décrire les phénomènes hydrologiques ne découlent pas des équations et des lois de la physique). Dans une première partie de ce chapitre, le modèle hydrologique qui simule la "transformation" des données de pluies en débits est détaillé à l'échelle d'une maille. Ce modèle hydrologique découle du modèle développé par [Javelle et al. \(2016, 2014, 2010\)](#) pour l'outil opérationnel Vigicrues Flash. Néanmoins, le modèle originel est dépourvu d'une fonction de routage "réaliste" propageant les débits de l'amont vers l'aval du bassin-versant. Pour ces travaux, trois modèles de routage ont été développés et testés. Ces modèles sont décrits dans la seconde partie de ce chapitre.

3.1 Présentation du modèle pluie-débit

Le modèle hydrologique présenté dans cette partie est appelé GRD. Il s'appuie sur des travaux antérieurs réalisés dans le cadre de la mise en place du système national d'avertissement Vigicrues Flash (Javelle et al., 2016, 2014, 2010). Les modifications que nous avons réalisées portent principalement sur la fonction de routage du modèle et sont détaillées dans la suite.

GRD emprunte les opérateurs de modélisation issus des modèles de la famille GR (Génie-Rural). Les modèles GR sont des modèles à réservoirs globaux ou semi-distribués développés par l'INRAE d'Anthony depuis plus de 30 ans. Ces modèles ont été largement testés et éprouvés. Ils ont démontré leurs performances pour simuler les réponses de nombreux bassins-versants variés sous des conditions climatiques variées et pour différents pas de temps (Perrin et al., 2003; Mouelhi et al., 2006; Lobligeois et al., 2014; Munier et al., 2014; Ficchi et al., 2016; Santos et al., 2018; Riboust et al., 2019; De Lavenne et al., 2019).

Le modèle GRD est défini sur un maillage carré de résolution $1km^2$ (grille) et fonctionne au pas de temps horaire. Pour chaque pas de temps, les entrées du modèle sont les pluies et l'évapotranspiration disponibles à la résolution du maillage. Pour chaque maille (appelée aussi pixel ou cellule), le modèle hydrologique GRD inclut un réservoir de "production" de capacité c_p et un réservoir de "transfert" de capacité c_{tr} . Les débits sont calculés pour chaque maille puis routés vers l'exutoire à l'aide d'un modèle de propagation hydraulique simplifiée dépendant d'un autre paramètre (trois variantes seront proposées dans la suite de ce chapitre). GRD décrit les processus hydrologiques de façon conceptuelle sans relation directe avec les caractéristiques physiques du bassin-versant. La figure 3.1 schématise le fonctionnement du modèle.

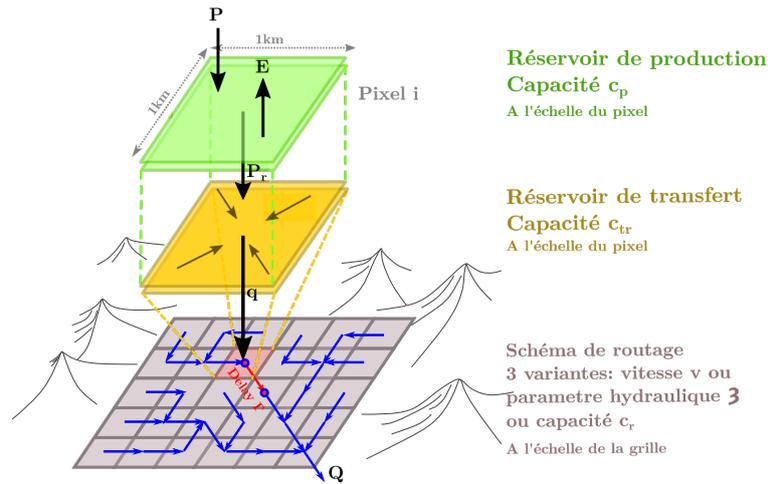


FIGURE 3.1 – Schéma de la nouvelle version du modèle GRD incluant une fonction de routage.

3.1.1 L'opérateur de production

Soit $P(t)$ représentant la pluie en fonction du temps et $E(t)$ représentant l'évapotranspiration en fonction du temps. Pour chaque pixel une fonction permet de calculer la quantité de pluie P_r (la pluie efficace) qui participe à l'écoulement. Cette fonction est définie selon les étapes suivantes.

D'abord la pluie nette P_n et l'évapotranspiration nette E_n sont calculées par les équations suivantes :

$$\begin{cases} P_n = P - E, & E_n = 0, & \text{si } P \geq E, \\ P_n = 0, & E_n = E - P, & \text{si } P < E \end{cases} \quad (3.1)$$

Le réservoir de production est rempli par la quantité P_p qui est une fraction de la pluie nette P_n . Ce même réservoir est aussi vidangé par la quantité E_p qui est une fraction de l'évapotranspiration E_n . La variation du niveau du réservoir de production est représentée par l'équation suivante :

$$dh_p = \left[1 - \left(\frac{h_p}{c_p} \right)^2 \right] dP_n - \frac{h_p}{c_p} \left(2 - \frac{h_p}{c_p} \right) dE_n. \quad (3.2)$$

Cette équation est intégrée par rapport au pas de temps δt (les détails de l'intégration de cette équation peuvent être trouvés dans (Edijatno, 1991)) pour obtenir les quantités P_p (quantité de pluie alimentant le réservoir) et E_p (quantité évaporée du réservoir).

$$P_p = c_p \left(1 - \left(\frac{h_p}{c_p} \right)^2 \right) \frac{\tanh\left(\frac{P_n}{c_p}\right)}{1 + \left(\frac{h_p}{c_p}\right) \tanh\left(\frac{P_n}{c_p}\right)}, \quad (3.3)$$

$$E_p = h_p \left(2 - \frac{h_p}{c_p} \right) \frac{\tanh\left(\frac{E_n}{c_p}\right)}{1 + \left(1 - \frac{h_p}{c_p}\right) \tanh\left(\frac{E_n}{c_p}\right)}. \quad (3.4)$$

Notons que h_p est l'état du réservoir au début du pas de temps δt . P_p et E_p sont les volumes d'eau gagnés ou perdus par le réservoir au cours du pas de temps δt . A la fin du pas de temps h_p est mis à jour en ajoutant P_p et en enlevant E_p .

Finalement, la fraction de la pluie P_r qui contribue à l'écoulement mais qui n'entre pas dans le réservoir de production est calculée par :

$$P_r = P_n - P_p \quad (3.5)$$

Le rôle du réservoir de production est de représenter l'humidité du sol. Un réservoir vide ($h_p = 0$) signifie que le sol est complètement sec. La pluie efficace qui participe à l'écoulement est nulle. L'intégralité de la pluie alimente le réservoir de production ($E_p = 0$, $P_p = P_n$, $P_r = 0$). A l'opposé, un réservoir de production plein ($h_p = c_p$) signifie que le sol est complètement saturé : la pluie n'alimente plus le réservoir de production mais contribue totalement à l'écoulement ($E_p = E_n$, $P_p = 0$, $P_r = P_n$).

3.1.2 L'opérateur de transfert

La pluie efficace $P_r(t)$ qui contribue à l'écoulement alimente le réservoir de transfert caractérisé par sa capacité c_{tr} et son état (niveau de remplissage) h_{tr} . Pendant le pas de temps δt , le réservoir de transfert se vidange et émet le débit élémentaire q (débit associé à un pixel). Cette opération est modélisée par un opérateur conservatif. Celui-ci dérive d'une équation différentielle qui décrit l'évolution du niveau de remplissage h_{tr} dans le temps. Cette équation différentielle respecte la condition de conservation de la masse :

$$\frac{dh_t}{dt} + ah_t^b = P_r, \quad (3.6)$$

Il a été montré par (Perrin et al., 2003) que l'équation 3.6 modélise correctement les processus hydrologiques de crue et de sécheresse pour $b = 5$. Cette valeur est issue d'expériences empiriques (tests sur un grand nombre de bassins versants) et n'a pas de justification physique. Pour obtenir une expression du débit élémentaire q , l'équation 3.6 est intégrée sur un pas de temps Δt . Il en résulte l'expression suivante :

$$q = h_t - (h_t^{-4} + c_{tr}^{-4})^{-0.25}, \quad (3.7)$$

où la capacité du réservoir de transfert est exprimée en mm et vaut $c_{tr} = (4a\Delta t)^{-0.25}$.

Davantage de détails concernant les réservoirs de production et de transfert peuvent être trouvés dans [Perrin et al. \(2003\)](#).

3.2 Développement du modèle de routage : trois variantes proposées

La version originelle du modèle intégrait un routage global très simplifié (pixels-exutoires). Les débits à l'exutoire du bassin-versant sont calculés en additionnant les contributions simultanées de chaque pixel. Cette modélisation permet d'anticiper les crues en fonctionnement opérationnel (fonctionnement en temps réel avec des prévisions de pluie). Cependant, ce routage global simplifié pose plusieurs problèmes :

- Les phénomènes hydrauliques responsables de la propagation des crues dans les rivières ne sont pas modélisés (temps de propagation, atténuation...)
- Aux stations de jaugeage, les hydrogrammes observés et modélisés ne sont pas synchronisés temporellement. Cela rend le calage des paramètres du modèle délicat : le problème numérique peut devenir insolvable, les phénomènes de propagation hydraulique peuvent être partiellement modélisés par les paramètres du modèle hydrologique (compensation).
- Incohérence spatiale du modèle : chaque maille est indépendante. Les paramètres du modèle ne peuvent pas être estimés à l'aide de plusieurs stations de mesure du débit le long du réseau. À l'inverse, les paramètres doivent être optimisés spécifiquement pour chaque station de jaugeage.

Ces problèmes sont incompatibles avec les objectifs de ces travaux. En effet, l'un des objectifs est d'identifier les paramètres du modèle de manière spatialisée pour intégrer les spécificités locales des bassins-versants. La description des processus locaux pourrait permettre d'améliorer la prédictibilité des débits aux exutoires non-jaugés. Cet objectif requiert une continuité spatiale du modèle pour décrire les processus hydrologiques et hydrauliques le long du réseau de drainage. L'ajout et le développement d'un modèle de routage pour propager les débits de maille en maille de l'amont vers l'aval du réseau doit pouvoir résoudre ces problèmes. Trois modèles de routage ont été testés et développés. Ces modèles satisfont les caractéristiques suivantes :

- Modéliser conceptuellement le temps de propagation et/ou l'amortissement des crues afin de permettre l'assimilation des débits pour le calage des paramètres du modèle.
- Simplifier la représentation des processus pour limiter le nombre de paramètres à estimer (idéalement 1 seul paramètre par maille)
- Développer un modèle respectant les contraintes de différentiabilité (l'utilisation d'un algorithme de calage variationnel requiert le calcul du gradient, celui-ci est calculé à l'aide du modèle adjoint ([Errico, 1997](#)) obtenu par différentiation automatique du code de calcul à l'aide de Tapenade ([Hascoet and Pascual, 2013](#)))

Le choix du modèle de routage s'oriente vers un modèle simplifié et conceptuel. En effet, l'objectif de GRD n'est pas de modéliser des processus purement hydrauliques. Il semble raisonnable de penser que les processus de propagation des débits à travers le maillage de GRD ne soient pas seulement de nature hydraulique : le maillage est grossier ($1km^2$) et les mailles, localisées sur les versants, pourraient avoir un comportement différent des mailles localisées sur le réseau hydraulique. Pour limiter les processus de compensation des paramètres entre eux, les concepts hydrologiques et hydrauliques représentés par les opérateurs de GRD devront être séparés le plus clairement possible. Aussi nous souhaitons conserver un modèle léger (consommation de mémoire limitée) et rapide (temps de calcul réduit). En effet introduire un modèle de routage physique non-linéaire peut s'avérer très lourd en calcul.

Trois modèles de routage sont présentés, implémentés et testés dans ces travaux. Par souci de simplicité, ces modèles sont présentés en une dimension (1D). On distinguera deux catégories conceptuelles : les modèles modélisant un temps de retard et le modèle associé à la vidange d'un réservoir. Dans la suite ces modèles de routage seront respectivement appelés variante 1, 2 et 3.

3.2.1 Implémentation d'un schéma de routage de maille en maille

Les premiers travaux ont consisté à implémenter un schéma de routage distribué de maille à maille (Organde et al., 2018) tel que présenté à la figure 3.2. Un tel modèle est construit à partir d'un modèle numérique de terrain. Le débit sortant d'une maille amont (notée $i - 1$) est propagé vers la maille aval (notée i). Le débit émissaire de la maille i dépend à la fois du débit provenant du réservoir de transfert du modèle hydrologique (noté q_i) et du débit transitant depuis la maille amont $i - 1$. La fonction \mathcal{R} représente le modèle de routage (équation 3.8) :

$$Q_i = \mathcal{R}(Q_{i-1}) + q_i \quad (3.8)$$

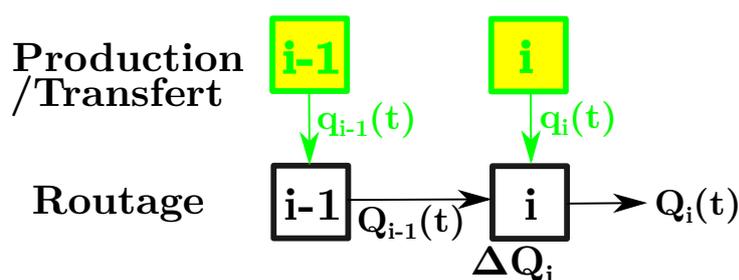


FIGURE 3.2 – Schéma de routage distribué connectant les mailles $i - 1$ et i . Les mailles représentées en vert concernent la partie "hydrologique" du modèle. Les mailles représentées en noire concernent la partie "hydraulique" (routage) du modèle.

Notons qu'en deux dimensions les débits sont propagés de maille en maille de l'amont du bassin-versant vers l'aval suivant huit directions calculées à partir de la grille des directions

d'écoulement présentée au paragraphe 2.1.2. Chaque maille reçoit jusqu'à sept contributions amont et émet un débit vers une seule maille aval. Une maille située à l'amont du bassin-versant ne reçoit pas de contribution. Chaque maille se vidange vers une maille aval selon la direction indiquée sur la grille des directions d'écoulement.

3.2.2 Modèles conceptuels modélisant un temps de retard

Concepts

Ce modèle représente le temps de propagation de l'hydrogramme entre deux mailles. Considérons deux cellules hydrologiques voisines (voir figure 3.2) indexées $i - 1$ et i et définissons les échanges de masse entre elles.

Premièrement, nous supposons que le transfert est réalisé en cascade : il est unidirectionnel de la cellule $i - 1$ à i . Le débit total émissaire de la maille $i - 1$ est propagé vers la maille aval i . Ce débit est retardé par un temps de retard noté τ_{i-1} :

$$\tau_{i-1} = \frac{d_{i-1,i}}{v_{i-1,i}}, \quad (3.9)$$

où $d_{i-1,i}$ et $v_{i-1,i}$ sont la distance et la vitesse de routage entre les deux mailles $i - 1$ et i .

L'implémentation la plus facile de ce modèle est de calculer le débit émissaire de la maille i par la formule 3.10 :

$$Q_i(t) = q_i(t) + Q_{i-1}(t - \tau_{i-1}), \quad (3.10)$$

où t est le pas de temps courant, Q_i est le débit total émis par la maille i , Q_{i-1} est le débit total émis par la maille voisine amont $i - 1$ et q_i est le débit élémentaire émis par le réservoir de transfert à la maille i (équation 3.8).

Notons que le second terme de l'équation 3.10 est, en deux dimensions, la somme de plusieurs contributions provenant des mailles voisines amont (7 contributions au plus) avec leurs propres valeurs de d et v .

L'équation 3.10 n'est pas une formulation explicite du débit $Q(t)$ en fonction du retard τ . $Q(t)$ n'est donc pas différentiable par rapport à τ . Le modèle 3.10 ne respecte pas la condition de différentiabilité énoncée au paragraphe 3.2. Cette formulation n'est pas utilisable avec un algorithme de calage variationnel nécessitant le calcul du gradient (voir paragraphe 1.4.3). Pour contourner ce problème, le second terme de l'équation 3.10 est écrit sous la forme d'un produit de convolution mettant en jeu une fonction K appelée noyau de convolution.

Définition du noyau de convolution

Le membre à droite de l'équation 3.10 est écrit sous la forme de l'intégrale suivante :

$$Q_{i-1}(t - \tau_{i-1}) = \int_{t'=-\infty}^t Q_{i-1}(t') K(t', \tau_{i-1}) dt'. \quad (3.11)$$

Où t est le temps courant, τ_{i-1} est le temps de parcours du débit depuis la maille $i - 1$ vers la maille i et $K(t', \tau_{i-1})$ est un noyau paramétrique dépendant du temps t' et de τ_{i-1} . Le temps t' correspond aux pas de temps antérieurs dans l'intervalle $[-\infty : t]$. Le noyau K doit satisfaire la condition :

$$\int_{t'=-\infty}^t K(t', \tau_{i-1}) dt' = 1, \quad K(t', \tau_{i-1}) > 0 \quad (3.12)$$

La fonction (3.11) dépend explicitement du paramètre τ_{i-1} via K . Le gradient de Q_{i-1} par rapport à τ_i peut être calculé.

En supposant que $Q(t)$ est constant durant le pas de temps Δt , l'équation (3.11) peut être écrite sous la forme discrète suivante :

$$Q_{i-1}(t - \tau_{i-1}) = \sum_{l=1}^L \bar{\beta}_{i-1,l} Q_{i-1}(t - (l-1)\Delta t), \quad (3.13)$$

Où

$$\bar{\beta}_{i-1,l} = \frac{\beta_{i-1,l}}{\sum_{l=1}^L \beta_{i-1,l}},$$

$$\beta_{i-1,l} = K(t - (l-1)\Delta t, \tau_{i-1}), \quad l = 1, \dots, L.$$

A l'opposé de la période temporelle semi-infinie considérée dans l'équation (3.11), L définit une période temporelle finie (en terme de Δt) (équation 3.13). $t - (l-1)\Delta t$, $l = 1, \dots, L$ représente le parcours des pas de temps antérieurs à t appartenant à l'intervalle $] -\infty : t]$. Remarquons que pour une estimation finie des vitesses de routage $v_{i-1,i}$, le temps de retard τ_{i-1} est constant : les coefficients $\beta_{i-1,l}$ ne changent pas dans le temps et peuvent donc être pré-calculés et sauvegardés en mémoire.

Afin d'optimiser le temps de calcul de l'équation 3.13, un coefficient d'élongation temporelle ϵ est introduit. Ce coefficient permet de réduire le nombre de pas de temps parcouru par le compteur l , $l = 1, \dots, L$ en allongeant progressivement la durée du pas de temps. Un nouveau compteur de pas de temps est introduit par la formule suivante :

$$t_0 = t, \quad t_l = t_{l-1} - \epsilon^{l-1} \Delta t, \quad l = 1, \dots, L, \quad (3.14)$$

où Δt est le pas de temps et ϵ est le coefficient d'élongation. Les coefficients $\beta_{i-1,l}$ peuvent alors être calculés à l'aide de la fonction de distribution cumulée \mathcal{K} associée au noyau K suivant

le compteur t_l introduit à l'équation 3.14. L'équation 3.13 devient :

$$Q_{i-1}(t - \tau_{i-1}) = \sum_{l=1}^L \bar{\beta}_{i-1,l} \tilde{Q}_{i-1}(l), \quad (3.15)$$

où

$$\bar{\beta}_{i-1,l} = \frac{\beta_{i-1,l}}{\sum_{l=1}^L \beta_{i-1,l}}, \quad (3.16)$$

$$\beta_{i-1,l} = \int_{t_l}^{t_{l-1}} K(t', \tau_{i-1}) dt = \mathcal{K}(t_l, \tau_{i-1}) - \mathcal{K}(t_{l-1}, \tau_{i-1}), \quad l = 1, \dots, L-1, \quad (3.17)$$

$$\beta_{i-1,L} = \int_{-\infty}^{t_{L-1}} K(t', \tau_{i-1}) dt = 1 - \mathcal{K}(t_{L-1}, \tau_{i-1}), \quad (3.18)$$

$$\tilde{Q}_{i-1}(l) = \bar{Q}_{i-1}(t), \quad t \in [t_{l-1} : t_l]. \quad (3.19)$$

Choix du noyau de convolution

A partir de l'équation 3.11, il est possible d'utiliser différents noyaux $K(t', \tau)$. Un premier choix pourrait être le noyau d'Hayami (Moussa, 1996) qui est dérivé de l'équation de l'onde diffusive. Dans ce cas, on suppose que le transfert de masse entre cellules a une nature hydraulique. Cependant, nous préférons conserver une approche conceptuelle suivant les hypothèses formulées à la section 3.2. Nous proposons d'utiliser et de tester deux noyaux différents : la fonction de densité de probabilité Gaussienne (fonction symétrique) et la fonction de densité de probabilité Gamma (fonction asymétrique).

Utilisation d'une fonction de forme Gaussienne définie par :

$$K(t, \mu, \sigma) = \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right). \quad (3.20)$$

Où μ est l'espérance et σ est la variance. Le mode de la fonction est localisé pour $t = \mu$. Cette fonction est représentée pour deux valeurs de μ et σ sur le graphique 3.3.

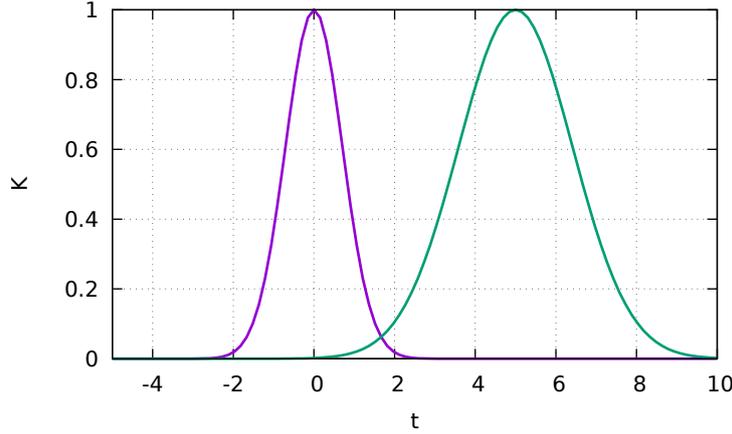


FIGURE 3.3 – Fonction de forme gaussienne $K(t, \mu, \sigma)$ représentée pour différentes valeurs de μ et σ . Courbe violette : $\mu = 0$. et $\sigma = 0.5$. Courbe verte : $\mu = 5$. et $\sigma = 2$.

Le noyau $K(t, \mu, \sigma)$ peut donc s'exprimer en fonction du pas de temps courant t , des pas de temps antérieurs $t' \in] - \infty : t]$, et du temps de retard τ :

$$K(t', t, \tau, \sigma) = \exp\left(-\frac{(t' - (t - \tau))^2}{2\sigma^2}\right). \quad (3.21)$$

$t - \tau$ représente la position du mode de la distribution dans l'espace temporel défini par $t \in] - \infty : t]$. Le noyau 3.21 est utilisé à la place de la fonction $K(t')$ de l'équation 3.11 pour $t' = t - (l - 1)\Delta t$. Le coefficient σ correspond à l'étalement de la fonction. Afin d'éviter des instabilités numériques (variation trop rapide des coefficients $\beta_{i,l}$ de l'équation 3.13), le coefficient d'étalement $\sigma = 0.5$ est utilisé pour les calculs (valeur minimum). Ce noyau est représenté dans l'espace des temps t' sur la figure 3.4.

Le modèle décrit précédemment ressemble au modèle "Lag and Route" décrit dans [Laganier et al. \(2014\)](#) et [Tramblay et al. \(2010\)](#). Cependant, la fonction Gaussienne représente la réponse hydraulique d'une manière plus réaliste. En effet, dans le modèle "Lag and Route" le noyau $K(t')$ est discontinu car il est égal à zéro pour $t' > t - \tau$. Il n'y a pas de discontinuité avec le noyau de forme gaussienne et le débit émit par la cellule $i - 1$ arrive progressivement à la cellule aval i . Ce schéma semble donc plus approprié pour un modèle de routage pixel-pixel. De plus l'absence de discontinuité est une condition nécessaire pour obtenir un modèle différentiable.

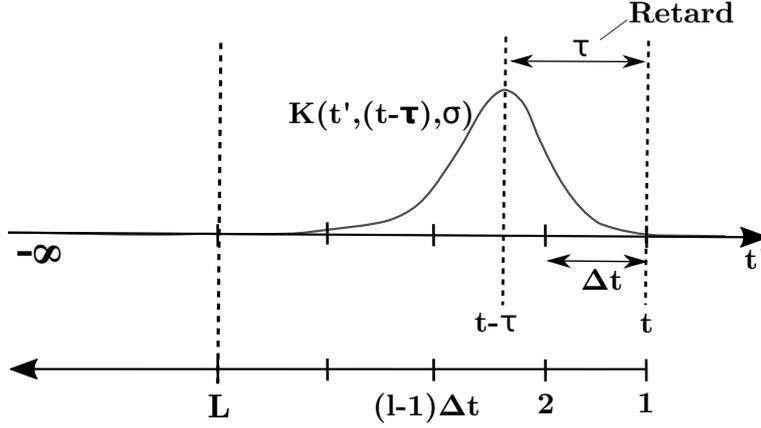


FIGURE 3.4 – Représentation schématique du noyau Gaussien dans l’espace temporel t' . L’axe du bas représente le compteur l défini à l’équation 3.13.

Utilisation de la fonction de densité de probabilité Gamma définie par :

$$f(t; \eta, \theta) = \frac{t^{\eta-1} \exp(-t/\theta)}{\theta^\eta \Gamma(\eta)}, \quad (3.22)$$

Où $\Gamma(\cdot)$ est la fonction Gamma, η est le coefficient de forme et θ est le coefficient d’échelle. C’est une fonction asymétrique, positive, ses arguments sont positifs et l’étalement augmente lorsque son mode se déplace depuis sa valeur minimale 0. Son mode est localisé au point $(\eta - 1)\theta$ et son paramètre de forme peut être exprimé de manière analogue à l’équation 3.20, via μ (paramètre de position dépendant du retard) et σ (paramètre d’étalement) :

$$\eta = 1 + \mu/\sigma,$$

La fonction 3.22 devient

$$f(t; \mu, \sigma) = \frac{t^{\mu/\sigma} \exp(-t/\sigma)}{\sigma^{(1+\mu/\sigma)} \Gamma(1 + \mu/\sigma)}, \quad (3.23)$$

La fonction de densité de probabilité Gamma $f(t; \mu, \sigma)$ est représentée à la figure 3.5 pour différentes valeurs de son mode μ .

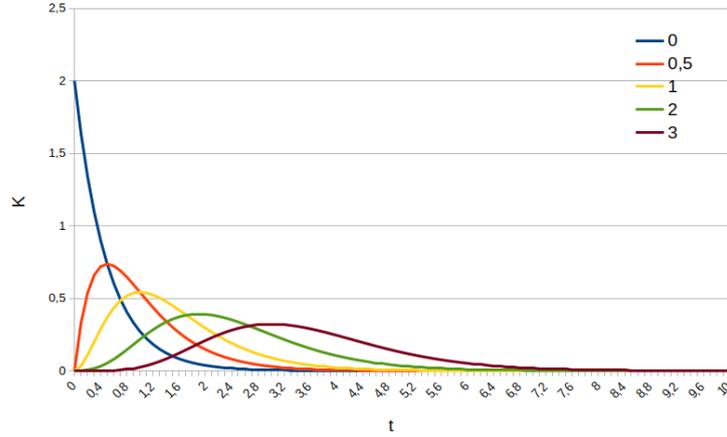


FIGURE 3.5 – Fonction de distribution Gamma pour différentes valeurs de son mode μ . Le coefficient d'étalement est fixé à $\sigma = 0.5$. La forme asymétrique de cette fonction se rapproche de celle d'un hydrogramme.

De la même manière que pour le noyau de forme gaussienne, la fonction 3.23 peut être exprimée en fonction de t , t' et du retard $\mu = \tau$. Nous obtenons alors le noyau K suivant :

$$K(t - t'; \tau, \sigma) = f(t'; \tau, \sigma). \quad (3.24)$$

Le noyau représenté dans l'espace temporel t' est schématisé sur le graphique 3.6.

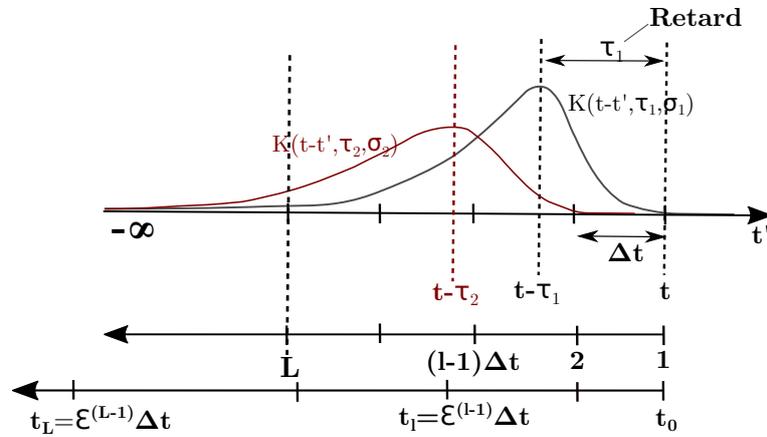


FIGURE 3.6 – Représentation schématique du noyau Gamma dans l'espace temporel t' . Plus le retard τ augmente et plus cette fonction s'étale. Les axes du bas représentent les compteurs l et t_l définis par l'équation 3.15.

Les coefficients $\beta_{i-1,l}$ (équation 3.15) peuvent être calculés à partir de la fonction de distribution cumulée correspondante représentée par la fonction Gamma-Incomplete

suivante :

$$\mathcal{K}(t - t', \tau, \sigma) = \frac{\gamma(1 + \tau/\sigma, t'/\sigma)}{\Gamma(1 + \tau/\sigma)}. \quad (3.25)$$

Cette fonction est calculée par la routine **'gammp'** issue des "Numerical Recipes" Fortran. Si les paramètres τ and σ ne dépendent pas du temps, la fonction $\mathcal{K}(t - t', \tau, \sigma)$ peut être pré-calculée aux points $t - t'$ et tabulée pour simplifier son usage.

Le noyau issu de la fonction de distribution Gamma (équation 3.24) améliore la représentation conceptuelle de la réponse hydraulique par rapport à l'utilisation d'un noyau de forme gaussienne (équation 3.21). En effet, sa forme asymétrique pourrait mieux représenter les phénomènes de propagation des crues. De plus, l'étalement de cette fonction augmente si le retard τ augmente. Cette fonction présente donc des caractéristiques intéressantes pour représenter un transfert hydraulique (voir figure 3.6) dont la vitesse de propagation dépendrait du débit (modèle non-linéaire).

Définition d'un modèle linéaire et non-linéaire

Le modèle linéaire :

Dans le cas d'un modèle linéaire le temps de retard τ est constant quel que soit le débit. La position du mode des fonctions de distribution Gaussienne et Gamma est définie par le temps de retard τ . Celui-ci peut être exprimé via la distance séparant deux cellules hydrologiques $d_{i-1,i}$ et la vitesse de propagation de l'eau $v_{i-1,i}$:

$$\tau_{i-1} = \frac{d_{i-1,i}}{v_{i-1,i}} \quad (3.26)$$

Dans ce cas la vitesse est un paramètre à caler pour chaque cellule. L'inconvénient de cette approche est que $v_{i-1,i}$ ne dépend pas de l'ordre de grandeur du débit. Les temps de réponse sont invariants quelles que soient les amplitudes des hydrogrammes injectés à l'amont du bief. La prédictibilité des débits par GRD dépendra donc de la gamme des débits étudiés. De plus l'amplitude du débit de pointe simulé ne dépend pas du modèle de routage (amortissement très limité). L'amplitude des débits dépendra donc essentiellement des réservoirs de transfert et production dont le fonctionnement au pixel est indépendant du réseau de drainage.

Le modèle non-linéaire conceptuel :

L'idée du routage non linéaire est de considérer que la vitesse $v_{i-1,i}$ évolue en fonction du débit Q_{i-1} transitant entre les cellules $i - 1$ et i . Par exemple, on peut supposer

que $v_{i-1,i}$ est une vitesse locale calculée par la fonction $\Theta(Q)$, qui pour un large canal rectangulaire, est donnée par (Litrico et al., 2010) :

$$\Theta(Q) = \frac{5S_b^{0.3}Q^{0.4}}{3W^{0.4}n^{0.6}},$$

où S_b est la pente locale, W est la largeur et n est le coefficient de frottement de Manning. A partir de cette formule, nous posons

$$v_{i-1,i} = \xi_{i-1}Q_{i-1}^{0.4}, \quad (3.27)$$

où ξ_{i-1} est un paramètre inconnu (appelé paramètre hydraulique). Ce paramètre est distribué spatialement et doit être estimé.

Puisque le délai dépend maintenant de $Q_{i-1}(t)$, le «schéma de routage non linéaire» signifie que le temps de propagation entre les cellules $i - 1$ et i est noté :

$$\tau_{i-1} = \frac{d_{i-1,i}}{\xi_{i-1}Q_{i-1}^{0.4}(t)},$$

3.2.3 Description de la variante 1 : modèle de retard linéaire

Ce modèle, dénoté **variante 1**, utilise :

- Le noyau $K(t', \tau, \sigma)$ Gaussien défini par l'équation 3.21. Pour toutes les simulations le paramètre d'étalement du noyau Gaussien vaut $\sigma = 0.5$.
- La formulation de la vitesse de propagation de l'eau qui ne dépend pas du débit (équation 3.26).
- Ce modèle n'utilise pas de coefficient d'élongation temporelle ϵ

L'algorithme de ce modèle est présenté ci-dessous (algorithme 1). Considérons deux cellules hydrologiques voisines $i - 1$ et i . Introduisons un compteur de pas de temps $l \in [1 : L]$ tel que $l = t - t' + 1$. Ce compteur remonte l'espace temporel depuis le pas de temps courant t vers un pas de temps antérieur $t - L$.

Algorithm 1 schéma de routage linéaire (algorithme simple)

{1 - Calcul des coefficients du noyau Gaussien}
for $l = 1, L$ **do**
 Calcul $\beta_{i-1,l}$ en utilisant les équations 3.17, 3.18, 3.19 pour $\epsilon = 1$
end for
{2 - Calcul du débit transitant depuis la cellule amont $i - 1$ }
 $\tilde{Q}_{i,t} = 0$
for $l = 1, L$ **do**
 $\tilde{Q}_{i,t} = \tilde{Q}_{i,t} + \beta_{i-1,l} \times Q_{i-1,t-(l-1)}$
end for
{3 - Calcul du débit émit par la cellule i }
 $Q_{i,t} = \tilde{Q}_{i,t} + q_{i,t}$

Un second algorithme (algorithme 2), équivalent à l'algorithme 1 est présenté. L'algorithme 2 est plus complexe à mettre en œuvre. Cependant il marque une étape importante vers l'algorithme de routage non-linéaire présenté au paragraphe suivant (algorithme 3). C'est pour cela qu'il est présenté ici. L'algorithme 2 est associé au schéma 3.7. Pour une cellule hydrologique donnée i , nous introduisons une mémoire qui inclut des cellules mémoires de $l = 1, \dots, L$. Chaque cellule mémoire correspond respectivement aux intervalles de temps $t \in [t - L, t]$ (voir figure 3.7). Chaque cellule contient une variable d'état $\tilde{Q}_{i,l}$ et un stock $dQ_{i,l}$ (voir l'algorithme ci-dessous). Soit $Q_i(t)$ le flux émit depuis la cellule i pendant la période $t \in [t - dt, t]$ et $\tilde{Q}_{i-1,l}$ le débit transitant de la cellule $i - 1$ à la cellule i , le transfert entre les cellules hydrologiques $i - 1$ et i s'écrit sous la forme de l'algorithme 2 :

Algorithm 2 schéma de routage linéaire (algorithme complexe)

$Q_{i-1} = \tilde{Q}_{i-1,1}$ {1 - Débit transitant depuis la cellule $i - 1$ }
 $Q_i = Q_{i-1} + q_i$ {2 - Débit émit depuis la cellule i }
{3 - Calcul des coefficients du noyau}
for $l = 1, L$ **do**
 Calcul $\beta_{i-1,l}$ en utilisant les équations 3.17, 3.18, 3.19 pour $\epsilon = 1$
end for
{4 - Étalement en mémoire du débit émit depuis la cellule i }
for $l = 1, L$ **do**
 $\tilde{Q}_{i,l} = \tilde{Q}_{i,l} + \beta_{i,l} Q_i$
end for
{5- Décalage en mémoire du débit transitant depuis la cellule $i - 1$ }
for $l = 1, L - 1$ **do**
 $d\tilde{Q} = \tilde{Q}_{i-1,l+1}$
 $\tilde{Q}_{i-1,l} = d\tilde{Q}$
end for
 $\tilde{Q}_{i-1,L} = 0$

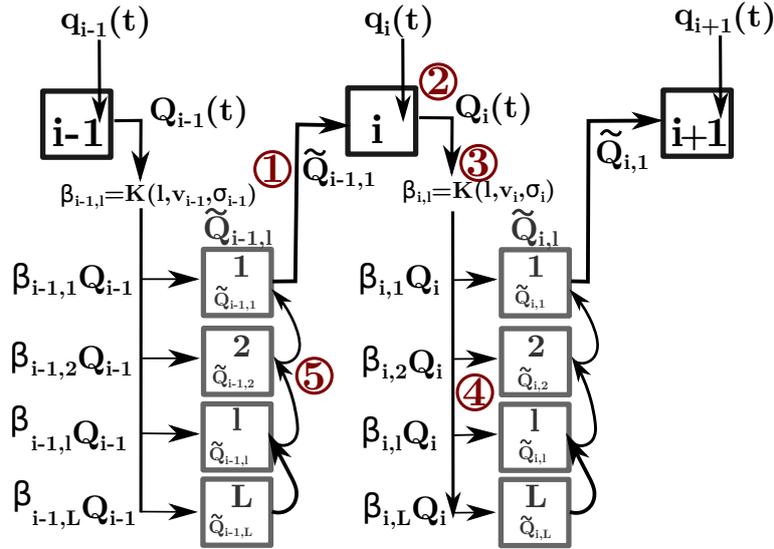


FIGURE 3.7 – Schéma représentant les échanges de masse entre les cellules $i - 1$, i et $i + 1$. Les cinq étapes de l’algorithme 2 sont numérotées en rouge.

Pour un paramètre de vitesse donné, les coefficients $\beta_{i,l}$ calculés à partir du noyau Gaussien n’évoluent pas dans le temps et peuvent être pré-calculés pour l’ensemble des cellules hydrologiques. L’étape 1 de l’algorithme 1 et l’étape 3 de l’algorithme 2 doivent donc être réalisées une seule fois (à l’extérieur de la boucle sur les pas de temps) et les coefficients doivent être stockés dans un tableau.

3.2.4 Description de la variante 2 : modèle de retard non-linéaire

Ce modèle, dénoté **variante 2**, utilise :

- Le noyau $K(t', \tau, \sigma)$ Gamma défini par l’équation 3.24. Pour toutes les simulations le paramètre d’étalement du noyau Gamma vaut $\sigma = 0.5$.
- La formulation de la vitesse de propagation de l’eau dépendant du débit (équation 3.27).
- Ce modèle utilise le coefficient d’élongation temporelle présenté à l’équation 3.14.

Ce modèle généralise la variante 1. Le retard dépend du débit simulé et l’amortissement des crues est représenté par les propriétés de la fonction Gamma (étalement en fonction de la position de son mode). L’algorithme présenté ci-dessous (algorithme 3) découle de l’algorithme 2. Il est associé au schéma 3.8.

Pour une cellule hydrologique donnée i , nous introduisons une mémoire qui inclut des cellules mémoires de $l = 1, \dots, L$. Chaque cellule mémoire correspond respectivement aux intervalles de temps $t \in [t_{l-1}, t_l]$ (voir figure 3.8). Chaque cellule contient une variable d’état $\tilde{Q}_{i,l}$ et un stock $dQ_{i,l}$ (voir l’algorithme ci-dessous). Soit $Q_i(t)$ le flux émit depuis la cellule i pendant la période $t \in [t - dt, t]$ et $\tilde{Q}_{i-1,l}$ le débit transitant de la cellule $i - 1$

un nombre de cellule mémoire L et un coefficient d'élongation β , les instants de temps t_l sont fixés. Ainsi, le retard maximum τ_{max} ne peut pas dépasser la valeur

$$\tau_{max} = t_L = \Delta t \sum_{l=1}^L \beta^{l-1},$$

alors que le délai minimum τ_{min} est limité par la vitesse maximale prise en compte à partir de considérations physiques. Le paramètre τ est discrétisé uniformément avec un pas $d_\tau = (\tau_{max} - \tau_{min})/k_\tau$, ou k_τ est le nombre de discrétisations. Pour un σ donné, nous obtenons une table de la fonction Gamma incomplète $\mathcal{K}(t_l; \tau_k, \sigma)$, $l = 1, \dots, L$, $k = 1, \dots, k_\tau$ avant de lancer le modèle hydrologique. Une interpolation linéaire est utilisée pour calculer $\mathcal{K}(t_l; \tau, \sigma)$.

Pour les calculs, le paramètre d'étalement σ est fixé à la valeur $\sigma = 0.5$. Il doit être suffisamment grand pour éliminer certains artefacts numériques lorsque $Q_i(t)$ varie rapidement.

3.2.5 Description de la variante 3 : modèle à réservoir linéaire

Dans le modèle de Nash (Nash, 1957), le routage est effectué par une série de n réservoirs linéaires de même capacité. Dans le cas d'un modèle distribué sur un maillage fixe, le nombre de réservoirs en cascade est déterminé. Dans ce cas, le décalage et l'amortissement des débits sont paramétrés par les capacités des réservoirs et non par leurs nombres. Dans notre modèle, l'équation de vidange du réservoir de routage découle de l'équation différentielle de conservation de la masse 3.6 écrite à la section 3.1.2. La loi de vidange du réservoir de routage est obtenue par intégration de l'équation 3.6 sur un pas de temps Δt . Cependant, l'intégration diffère par le choix de l'exposant $b = 1$. Il en résulte l'équation de vidange suivante :

$$Q(t) = h_r(t)(1 - \exp(-c_r \Delta t)) \quad (3.28)$$

où $Q(t)$ est le débit émis par le réservoir de routage, c_r la capacité du réservoir de routage, $h_r(t)$ le niveau de remplissage du réservoir de routage et t le pas de temps courant.

Soit Q_{i-1} le débit émis depuis la cellule $i - 1$ et transitant vers la cellule i . Le débit $\tilde{Q}_i(t)$ émis par le réservoir de routage de la cellule i s'écrit alors :

$$\tilde{Q}_i(t) = (h_r^i(t) + \frac{Q_{i-1}(t)}{S_{i-1}})(1 - \exp(-c_r \Delta t)), \quad (3.29)$$

où S_{i-1} est la surface drainée amont.

Le débit total à l'exutoire du pixel i au pas de temps t , est la somme du débit émis

par le réservoir de routage $\tilde{Q}_i(t)$ avec le débit provenant du réservoir de transfert $q_i(t)$ (modèle hydrologique). Ce débit, noté $Q_i(t)$, est donné par la formule :

$$Q_i(t) = \tilde{Q}_i(t) + q_i(t) \quad (3.30)$$

A la fin du pas de temps t , le niveau du réservoir de routage est mis à jour par la formule suivante :

$$h_r^i(t) = h_r^i(t) - \tilde{Q}_i(t) \quad (3.31)$$

Remarquons que le flux amont (noté $Q_{i-1}(t)$) est divisé par la surface drainée amont (notée S_{i-1}). Dans ce cas, la capacité du réservoir réceptionnant le flux amont est indépendante de la surface drainée. L'algorithme du modèle de routage avec réservoir linéaire est présenté ci-dessous (algorithme 4) :

Algorithm 4 schéma de routage du modèle à réservoir

Require: $Q_{i-1}(t)$ {Débit émis par la cellule hydrologique amont}

Require: $q_i(t)$ {Débit émis par le réservoir de transfert}

$h_r^i(t) = h_r^i(t-1) + Q_{i-1}(t)$ {Mise à jour du niveau du réservoir}

$\tilde{Q}_i(t) = (h_r^i(t) + \frac{Q_{i-1}(t)}{S_{i-1}})(1 - \exp(-c_r^i \Delta t))$ {Débit émis par le réservoir de routage}

$h_r^i(t) = h_r^i(t) - \tilde{Q}_i(t)$ {Mise à jour du niveau du réservoir à la fin du pas de temps}

$Q_i(t) = \tilde{Q}_i(t) + q_i(t)$ {Débit total émis par le pixel i }

3.2.6 Algorithme général et simplifié du modèle GRD

L'algorithme général et simplifié du modèle GRD est écrit ci-dessous (algorithme 5). Par commodité, cet algorithme est donné pour un maillage à une seule dimension connectant deux cellules voisines $i-1$ et i . Seules les principales boucles et étapes sont représentées. Les opérateurs de production, de transfert et de routage sont représentés sous forme de fonctions.

Algorithm 5 Algorithme de modélisation pluie-débit distribué

Ensure: $(\xi_i, v_i, c_r^i, \sigma_i, c_{tr}^i, c_p^i) > 0 \forall i$ {Paramètres du modèle}
Ensure: $v_{min} > 0$ {Condition nécessaire sur la vitesse}
Require: $\tau_{max} = dx/v_{min}$ {Calcul du retard maximum possible}
Require: T {T est le nombre de pas de temps de simulation}
Require: n {n est le nombre de cellules à parcourir de l'amont vers l'aval}
 for $t = 1, T$ **do**
 for $i = 1, n$ **do**
 $P_r = Production(Pluie_i, c_p^i, h_t^i)$ {Production}
 $q_i = Transfert(P_r, c_{tr}^i, h_t^i)$ {Transfert}
 if $i == 1$ **then**
 $Q_{i-1} = 0$ {Débit amont nul}
 else
Require: Q_{i-1} {Débit émit à la cellule i-1}
 end if
 $\tilde{Q}_{i-1} = RoutageV1(Q_{i-1}, v_i)$ {Routage variante 1}
 $\tilde{Q}_{i-1} = RoutageV2(Q_{i-1}, \xi_i)$ {Routage variante 2}
 $\tilde{Q}_{i-1} = RoutageV3(Q_{i-1}, c_r^i)$ {Routage variante 3}
 $Q_i = \tilde{Q}_{i-1} + q_i$ {Débit émit à la cellule i}
 end for
 end for
 end for

3.3 Étude des réponses hydrauliques élémentaires des trois modèles de routage

3.3.1 Protocole

Dans cette section, les réponses hydrauliques élémentaires des modèles de routage (variantes 1,2 et 3 présentées aux sections 3.2.3, 3.2.4 et 3.2.5) sont comparées à celles du simulateur *SIC*². *SIC*² est un logiciel de modélisation hydraulique 1-D qui résout les équations de l'hydraulique à surface libre (équations de Barré de Saint-Venant) (Dorchies et al., 2013). On appelle réponse élémentaire, l'étude des signaux en sortie des modèles après une injection "échelon" en entrée. Une injection échelon correspond à une augmentation brutale du débit. Ce débit est maintenu pendant une durée finie puis diminue brutalement jusqu'à sa valeur d'origine.

Les comparaisons entre *SIC*² et GRD sont effectuées sur la base d'une configuration très simple du modèle physique *SIC*² : un routage unidirectionnel dans un bief uniforme de section rectangulaire, de largeur 10 mètres, de longueur 20 kilomètres et de pente 0.5 m pour 1000 m. La condition limite aval est définie à la profondeur normale. Elle se situe dans un second bief, prolongeant le premier, 20 kilomètres plus en aval. Cette condition limite est suffisamment éloignée vers l'aval pour que la ligne d'eau du premier bief ne soit pas influencée (pour la gamme de débits étudiée). La configuration du modèle

GRD (1D) réplique la configuration choisie dans SIC^2 : le maillage est composé d'une série de 21 cellules de $1km^2$. Le routage est effectué entre les centres de chaque cellule, soit une longueur totale de 20 kilomètres. Les réservoirs de production et de transfert sont écartés de la modélisation (capacités de stockage nulles) pour ne s'intéresser qu'au routage et pouvoir comparer les deux modélisations. Pour ces expériences, le pas de temps de simulation est de 15 minutes.

Dans SIC^2 , en entrée du bief amont, un débit de base constant de $1m^3/s$ est injecté. A cela un débit supplémentaire représenté par un signal type "échelon" est injecté pendant 5 heures à la 60^{ème} heure de simulation. Cette injection varie de 1 à $64 m^3/s$ selon les simulations réalisées. Dans GRD, les mêmes hydrogrammes sont injectés dans la cellule amont du modèle. Ces hydrogrammes amont sont préalablement convertis en mm/km^2 afin de concorder avec les spécificités du modèle hydrologique.

Les hydrogrammes aval produits par SIC^2 et le routage de GRD sont comparés pour cette série d'injections variant de 1 à $64 m^3/s$. Ces comparaisons sont visibles sur les graphiques 3.9 3.10 et 3.11. Les coefficients de routage du modèle GRD (vitesses v , paramètres hydrauliques ξ et les capacités des réservoirs de routage c_r) ont été ajustés manuellement. Ces coefficients prennent une valeur uniforme pour l'ensemble des mailles.

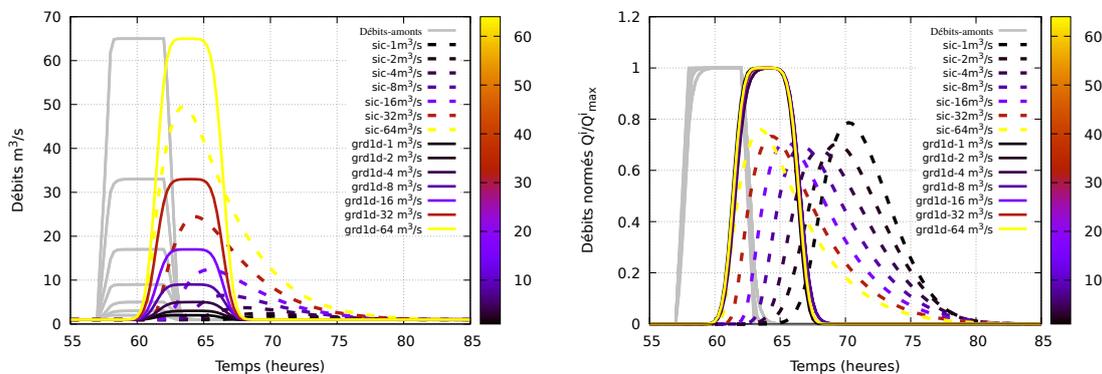


FIGURE 3.9 – Réponse hydraulique élémentaire du modèle de routage de la variante 1. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Le paramètre de vitesse de GRD a été ajusté de telle sorte que la position du pic de crue du débit de pointe le plus fort concorde avec celui de SIC^2 . Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps

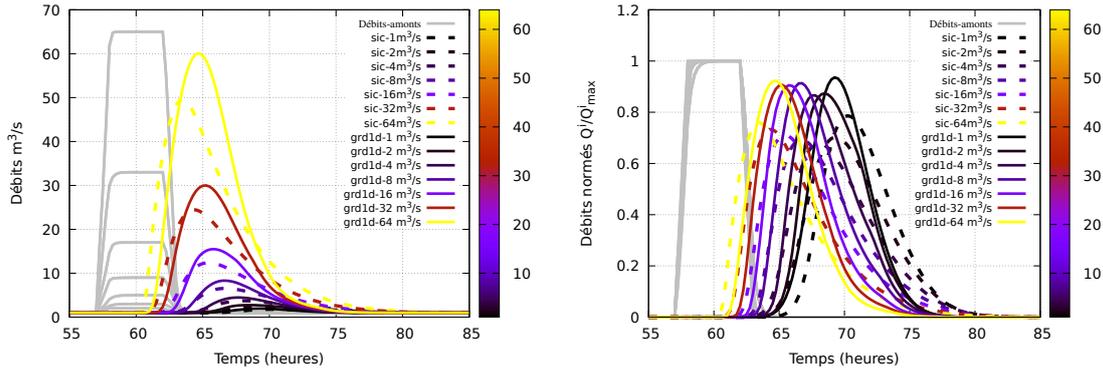


FIGURE 3.10 – Réponse hydraulique élémentaire du modèle de routage de la variante 2. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Les coefficients du modèle de routage de la variante 2 ont été ajustés manuellement pour que les débits de pointe et le retard du pic de crue concordent entre SIC^2 et GRD. Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps

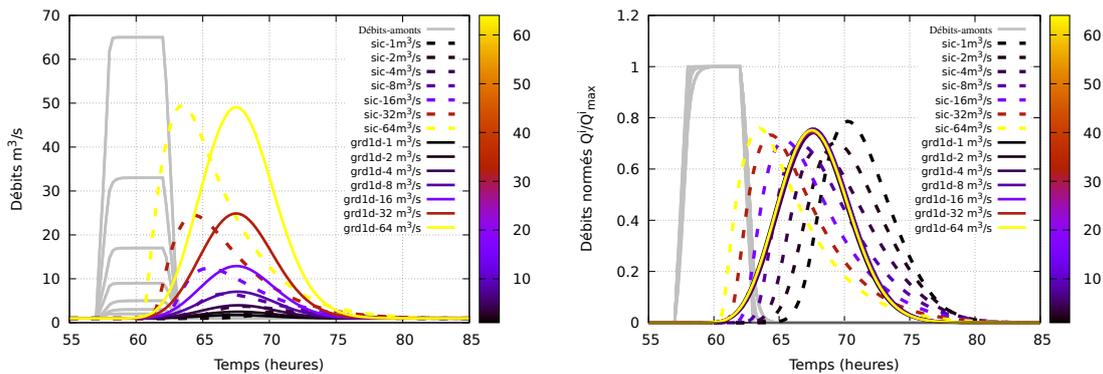


FIGURE 3.11 – Réponse hydraulique élémentaire du modèle de routage de la variante 3. Graphique de gauche : les hydrogrammes obtenus sont en m^3/s . Les capacités des réservoirs de routage de GRD ont été ajustées au mieux pour que le débit de pointe du pic de crue le plus fort concorde avec celui de SIC^2 . Graphique de droite : les hydrogrammes obtenus sont normés (en divisant par le débit maximum) afin de mieux percevoir la position du pic de crue dans le temps

3.3.2 Résultats :

L'analyse des figures 3.9, 3.10 et 3.11 conduit aux observations suivantes :

- 1^{ère} observation : Les variantes de routage 1 et 3 de GRD sont "linéaires". Les temps de propagation ne dépendent pas du débit et sont invariants quelles que soient les amplitudes des hydrogrammes injectés à l'amont du bief. La prédictibilité des débits par GRD dépendra donc de la gamme des débits étudiés.

- 2^{ème} observation : l'amplitude du débit de pointe simulé par la variante 1 ne dépend pas du modèle de routage (amortissement très limité). L'amplitude des débits simulés par GRD dépendra donc essentiellement des réservoirs de transfert et production dont le fonctionnement au pixel est indépendant du réseau de drainage.
- 3^{ème} observation : Le modèle "variante 3", ne permet pas de modéliser un retard indépendamment de l'étalement de la crue. Les hydrogrammes aval sont nécessairement amortis (figure 3.11).
- 4^{ème} observation : Avec le modèle "variante 2", le temps de propagation de la crue dépend du débit injecté en amont. Les hydrogrammes obtenus avec GRD concordent davantage avec les hydrogrammes obtenus par *SIC*².

Les opérateurs "linéaires" (variantes 1 et 3) ne semblent pas suffisants pour modéliser correctement la propagation des crues. Les processus physiques engagés dans ces phénomènes dépendent notamment de l'intensité de la crue et des caractéristiques morphologiques du bassin-versant. Ces processus pourraient devenir prépondérants pour des bassins-versants de grandes tailles (*taille* > 1000km²) ou dotés de caractéristiques morphologiques différentes (faible pente, bassin-versant très allongé...). Dans l'état, la prédictibilité des débits opérée par les variantes 1 et 3, restera très dépendante de la gamme de débits étudiée.

La variante 2 permet de modéliser un retard pur et non-linéaire (c.a.d dépendant du débit). Ce modèle semble mieux reproduire les hydrogrammes propagés par le modèle hydraulique *SIC*². Ce modèle pourrait permettre de représenter les transferts hydrauliques pour une gamme de débits plus large.

Ces trois modèles de routage seront davantage testés et comparés à l'aide d'expériences de calage-validation dont les résultats sont présentés aux chapitres 5 et 6.

Chapitre 4

Choix et implémentation de l'algorithme de calage variationnel

Sommaire

4.1	Choix de l'algorithme d'estimation variationnel	78
4.2	Principe de l'estimation variationnelle	79
4.2.1	Notations	81
4.2.2	Formulation générique d'un problème d'estimation variationnelle (assimilation de données)	82
4.2.3	Algorithme d'estimation variationnelle	85
4.2.4	La méthode de la L-Curve	85
4.2.5	Régularisation itérative	86
4.2.6	Contraintes d'inégalité et normalisation du vecteur de contrôle	88
4.3	Critères d'estimation utilisés pour le calcul de la fonction coût	91
4.3.1	Fonction quadratique, méthode des moindres carrés	91
4.3.2	Fonction non-gaussienne, méthode d'estimation robuste	92
4.4	Le terme de régularisation	95
4.4.1	Formulation du terme de régularisation de type Tikhonov	95
4.4.2	Formulation de la matrice $B^{1/2}$ pour la méthode de la régularisation itérative	97
4.5	Détails pratiques quant à l'implémentation et l'utilisation de l'algorithme d'estimation variationnelle dans GRD	98
4.5.1	Choix des bornes	98
4.5.2	Détermination automatique du coefficient β pour la méthode de la L-curve	99
4.5.3	Réduction de la taille du vecteur de contrôle	100
4.5.4	Obtention du modèle adjoint	101
4.6	Méthode de calage globale pour l'évaluation de l'ébauche	102
	Pré-filtrage	102
	Algorithme de calage "pas à pas"	103
4.7	Conclusions	103

Ce chapitre est consacré à la description de l'algorithme d'estimation variationnel utilisé pour le calage des paramètres du modèle hydrologique GRD. Dans une première partie le choix de l'algorithme d'estimation variationnelle est d'abord présenté. Dans une deuxième partie, les principes de la méthode variationnelle et les algorithmes sont détaillés. Dans les troisième et quatrième parties les critères d'optimisation et la formulation du terme de régularisation sont présentés. Dans une cinquième partie, des détails pratiques quant à l'implémentation et le fonctionnement des algorithmes dans le modèle GRD sont donnés. Dans une sixième et dernière partie, un algorithme de recherche "global", servant à l'estimation du vecteur d'ébauche, est présenté.

4.1 Choix de l'algorithme d'estimation variationnel

Le calage d'un modèle hydrologique distribué est difficile. Tout d'abord, le nombre total des paramètres recherchés peut être très grand (on parle d'un problème de grande dimension). Cette contrainte limite le choix des méthodes de calage. Deuxièmement, l'identification des paramètres n'est pas garantie : ce problème est causé par la "rareté" des observations de débit dans l'espace (peu de stations de jaugeage par rapport aux nombres de mailles), par le contenu informationnel du signal d'entrée (variabilité des précipitations) et éventuellement par la structure du modèle choisi. Néanmoins, ce problème d'identification peut être partiellement contourné en augmentant la longueur et la fréquence des observations pour mieux analyser la dynamique du système.

Pour calibrer les paramètres des modèles distribués, les algorithmes d'estimation variationnelle sont des choix naturels. En effet, de tels algorithmes restent efficaces quelle que soit la dimension du vecteur de contrôle. Le vecteur de contrôle contient l'ensemble des variables recherchées. C'est pourquoi ces méthodes sont couramment utilisées en météorologie et en océanographie pour la prévision opérationnelle et la réanalyse en temps réel (assimilation de données) (Ledimet and Talagrand, 1986; Rabier and Courtier, 1992). Dans ces disciplines, l'algorithme d'estimation variationnel couramment utilisé est connu sous le nom de "4D-Var". Cette méthode converge vers le minimum local d'une fonction objectif (aussi appelée fonction coût) définie sur la fenêtre d'observation (période temporelle d'analyse). L'élément clé de la méthode est le modèle adjoint qui fournit le gradient exact de la fonction coût par rapport à tous les éléments du vecteur de contrôle. Ce gradient est obtenu seulement pour une unique exécution du modèle adjoint (Errico, 1997). Le modèle adjoint permet d'utiliser les méthodes de minimisation basées sur le gradient tel que les algorithmes de type Newton (Newton, quasi-Newton, BFGS). Cependant, le développement du modèle adjoint est très souvent un obstacle pour la mise en œuvre d'une méthode variationnelle. En hydrologie seulement quelques travaux ont été réalisés dont ceux de Castaings et al. (2009) et de Nguyen et al. (2016). Le modèle adjoint a été

obtenu à partir de modèles à base "physique" utilisant des équations aux dérivées partielles (Green-Ampt / onde cinématique pour [Castaings et al. \(2009\)](#) et Horton / onde cinématique pour [Nguyen et al. \(2016\)](#)). [Castaings et al. \(2009\)](#) ont obtenu le modèle adjoint par différentiation automatique alors que [Nguyen et al. \(2016\)](#) l'ont implémenté manuellement. Dans les travaux de ([Castaings et al., 2009](#)), les paramètres du modèle sont estimés pour un seul événement de crue dans un contexte d'expériences jumelles. Dans les travaux de ([Nguyen et al., 2016](#)), des paramètres globaux sont calibrés à l'aide de deux événements de crues. D'autres travaux utilisent des méthodes d'estimation variationnelle pour ajuster les états d'un modèle global ([Seo et al., 2009](#)) ou d'un modèle distribué ([Lee et al., 2012](#)).

Le Filtre de Kalman d'Ensemble (EnKF) est une autre méthode adaptée aux problèmes de grandes dimensions ([Evensen, 1994](#)). Cette méthode est utilisée en météorologie, en océanographie et en hydrologie ([Xie and Zhang, 2010](#); [Clark et al., 2008](#); [Moradkhani et al., 2005](#); [Reichle et al., 2002](#)). Cependant cette méthode est limitée au contexte d'erreur statistique gaussienne et n'est pas pratique pour résoudre des problèmes sous des contraintes additionnelles.

Les méthodes heuristiques et statistiques, tel l'algorithme Nelder-Mead, la chaîne de Markov Monte-Carlo et les algorithmes hybrides conçus pour les applications hydrologiques (SUFI-2, [Abbaspour et al. \(2007\)](#)) ne nécessitent pas le gradient pour évaluer les directions de descente. Cependant, ces méthodes convergent lentement et ne sont pas adaptées pour résoudre des problèmes de grandes dimensions. Leurs performances diminuent lorsque la taille du vecteur de contrôle grandit.

Pour les raisons énoncées ci-dessus, un algorithme de calage variationnel utilisant le modèle adjoint est implémenté et testé. Cet algorithme permet de caler les paramètres du modèle distribués de GRD.

4.2 Principe de l'estimation variationnelle

Un modèle numérique noté H permet de représenter des processus physiques à travers des algorithmes. Ces algorithmes établissent une correspondance entre des variables d'entrées (input) et des variables de sorties (output). Pour un jeu de variables d'entrées connu et défini par le modélisateur (pour une application hydrologique, ces variables sont par exemple la pluie, l'évapotranspiration et les valeurs de paramètres du modèle), le modèle numérique permet de calculer le jeu de variables de sortie correspondant : c'est ce que l'on appelle le "problème direct" (figure 4.1). Cependant, les variables d'entrées des modèles ne sont pas toujours entièrement connues, elles peuvent être difficiles à estimer et le modèle lui-même est imparfait. Le modélisateur dispose souvent d'observations (in-situ) des variables de sorties du modèle. Celles-ci sont nécessaires pour évaluer la justesse

du modèle. Ces observations sont le plus souvent incomplètes (localisées et disponibles sur une fenêtre temporelle réduite) mais elles sont indispensables pour ajuster certains paramètres du modèle (input) afin d'améliorer sa justesse. On parle alors de calage ou d'étalonnage des paramètres du modèle. Lorsque le modélisateur étalonne les paramètres de son modèle, il résout un "problème inverse" (figure 4.1).

Pour évaluer la justesse du modèle, le modélisateur compare les variables de sorties du modèle avec les variables observées. Les variables observées doivent être préalablement "synchronisées" spatialement (géolocalisation / maillage) et temporellement (dates / pas de temps) avec celles du modèle. La mesure de l'écart entre les variables modélisées et observées s'effectue par une fonction (algorithme) appelée "fonction coût" ou "fonction objectif". L'équation 4.1 est un exemple de fonction objectif calculant les résidus entre les grandeurs observées et modélisées.

$$J(X) = \frac{1}{2} \|H(X) - Y^*\|_{\mathcal{Y}}^2, \quad (4.1)$$

où H représente le modèle, X le vecteur de contrôle (les paramètres, les états ou les signaux d'entrées du modèle), $H(X)$ calcul les sorties du modèle (les débits simulés pour une application hydrologique) et $Y^* \in \mathcal{Y}$ représente les observations disponibles (définies dans l'espace des observations \mathcal{Y}).

Les variables d'entrées du modèle sont ajustées de façon à réduire l'écart entre les variables observées et modélisées. Cela revient à minimiser la valeur de la fonction objectif par un processus itératif (figure 4.2) : l'ajustement des paramètres est répété jusqu'à ce que la valeur de la fonction objectif converge vers sa valeur limite minimale 0. C'est un algorithme de calage variationnel (figure 4.1).

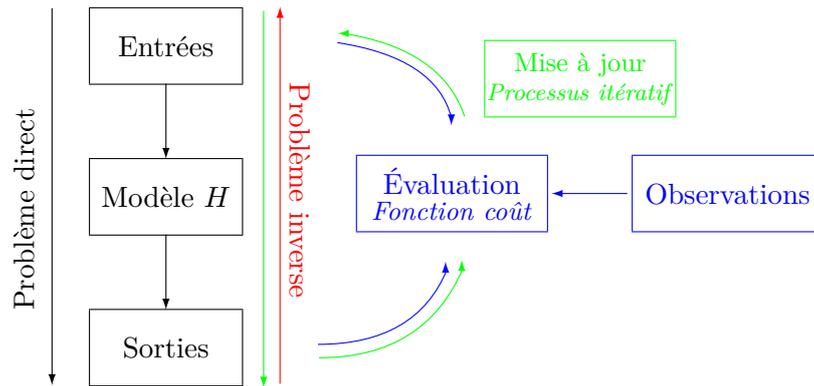


FIGURE 4.1 – Schéma présentant un problème direct, un problème inverse et le processus itératif permettant d’identifier les entrées (paramètres) d’un modèle H à l’aide de données observées.

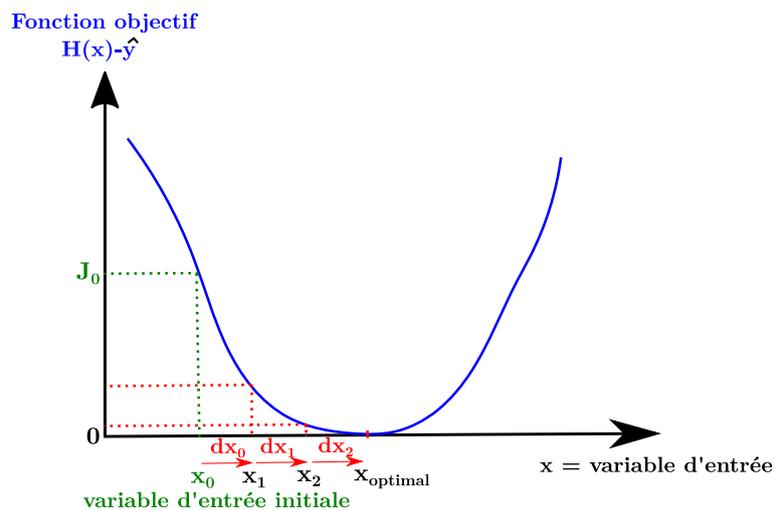


FIGURE 4.2 – Processus de convergence : la fonction objectif est minimisée en recherchant de manière itérative la valeur optimale de la variable d’entrée x . Cette recherche s’effectue en se déplaçant d’une distance dx_i depuis la valeur d’origine x_i pour chaque itération i .

4.2.1 Notations

Les notations suivantes sont couramment utilisées en optimisation variationnelle et seront utilisées dans la suite de ce manuscrit :

- X est le vecteur de contrôle : il s’agit des variables prises en compte dans le modèle tels que les paramètres et les états.
- \bar{X} est le vecteur de contrôle vrai. Il contient les vrais paramètres et les vrais états. Ce vecteur est souvent inconnu.

- X_b est le vecteur de l'ébauche : il s'agit de la connaissance à priori du vecteur de contrôle. Cette connaissance approximative des paramètres et des états du modèle peut provenir par exemple d'observations, d'expertises ou d'autres modélisations.
- X_a est le vecteur analysé. C'est le vecteur recherché par l'algorithme d'assimilation de données. Il contient donc les paramètres et les états trouvés après assimilation.
- Y^* est le vecteur des observations. Il contient les observations disponibles sur le système.
- Y est le vecteur contenant les variables d'état du système modélisé.

4.2.2 Formulation générique d'un problème d'estimation variationnelle (assimilation de données)

Soit X le vecteur de contrôle, vecteur composé des paramètres de contrôle du modèle. Pour le modèle GRD il s'agit des capacités des réservoirs de production et de transfert ainsi que des paramètres du modèle de routage pour chaque pixel du maillage recouvrant un bassin-versant. Les variables d'état du modèle ne sont pas considérées dans ces travaux.

Notons H le modèle mathématiques qui décrit le fonctionnement d'un système naturel à partir des variables d'état du système Y tel que :

$$Y = H(X)$$

Pour modéliser le comportement du système, le vrai vecteur \bar{X} doit être déterminé. Si l'on considère le modèle comme parfait, alors il est possible d'écrire,

$$\bar{Y} = H(\bar{X})$$

En réalité, \bar{X} est incertain et les incertitudes autour de \bar{X} sont notées ϵ_b . On définit alors le vecteur d'ébauche X_b qui est le vecteur de contrôle du système que l'on suppose être le plus proche de la réalité. Celui-ci est évalué de manière à priori :

$$X_b = \bar{X} + \epsilon_b, \quad \epsilon_b \sim N(0, B)$$

où ϵ_b est l'erreur statistique de l'ébauche et B la matrice des covariances d'ébauche.

A cause de la présence d'incertitudes ϵ_b , les prédictions du modèle $Y|X_b = H(X_b)$ (l'évaluation de Y est conditionné par X_b) sont elles aussi incertaines. De plus, les variables observées contiennent des incertitudes (bruits, erreur de mesures...). De ce fait, le vecteur

des observations vaut,

$$Y^* = \bar{Y} + \epsilon_o, \quad \epsilon_o \sim N(0, O)$$

où ϵ_o est l'erreur statistique autour des observations et O est la matrice des covariances d'observation.

Le but de l'assimilation de données est d'obtenir une estimation X_a de X conditionnée par les observations Y^* , c'est à dire $X_a = X|Y^*$. Cette estimation devrait être meilleur que l'ébauche X_b tel que,

$$\|X_a - \bar{X}\| < \|X_b - \bar{X}\|$$

Dans le cadre Bayésien, la densité de probabilité de X conditionné par Y^* est donnée par le formule de Bayes :

$$\mathcal{P}(X|Y^*) = \frac{\mathcal{P}(Y^*|X)\mathcal{P}(X)}{\mathcal{P}(Y^*)} \quad (4.2)$$

En estimation variationnelle, le mode de la fonction $\mathcal{P}(X|Y^*)$ est recherché. Cela revient à maximiser la probabilité $\mathcal{P}(X|Y^*)$. Dans ces travaux nous supposons donc que les erreurs ϵ_b et ϵ_o sont gaussiennes bien que cette hypothèse soit rarement valide en hydrologie. Sous cette hypothèse Gaussienne, le mode de $\mathcal{P}(X|Y^*)$ correspond au minimum de la fonction coût suivante :

$$J(X) = J_0(X) + J_{reg}(X) \quad (4.3)$$

où J_0 correspond au calcul des résidus par la formule :

$$J_0(X) = \frac{1}{2} \|O^{-1/2}(H(X) - Y^*)\|^2,$$

et J_{reg} est le terme de régularisation (pénalité) notée :

$$J_{reg}(X) = \frac{1}{2} \|B^{-1/2}(X - X_b)\|^2$$

L'estimation X_a est obtenue selon la condition optimale suivante :

$$J'(X_a) = 0.$$

Une solution explicite de l'équation ci-dessus peut être trouvée si le modèle H est linéaire (la méthode du BLUE par exemple). Si le modèle H est non-linéaire, la solution X_a est obtenue en résolvant un problème de contrôle optimal à l'aide d'un algorithme de

minimisation itératif :

$$X_a = \arg \min_X J(X) \quad (4.4)$$

La matrice B est très incertaine ou inconnue. Il est alors mieux de considérer la fonction objectif par la formulation de type Tikhonov utilisée pour les problèmes mal posés (Tikhonov, 1963) :

$$J(X) = J_0(X) + \beta J_{reg}(X) \quad (4.5)$$

Où β est le coefficient de régularisation.

Le terme $J_0(X)$ de l'équation 4.3 est le critère d'optimisation. Il correspond aux résidus (somme des écarts) calculés entre les sorties du modèle (débits simulés) et les observations (débits observés).

Le terme $J_{reg}(X)$ de l'équation 4.3 est le critère de régularisation (pénalisation). Il correspond aux résidus calculés entre les paramètres estimés x_a et les paramètres "à priori" x_b . $J_{reg}(X)$ représente une contrainte supplémentaire apportée au problème d'estimation variationnelle : il permet d'ajouter de "l'information" pour mieux contraindre un problème dont les paramètres recherchés sont beaucoup plus nombreux que les observations ; il assure que le minimum recherché de la fonction $J(X)$ soit obtenu pour un vecteur de contrôle analysé X_a proche du vecteur d'ébauche X_b ; il garantit théoriquement l'existence d'une solution unique et peut permettre d'imposer une information "a priori" sur le phénomène. Le coefficient de régularisation β doit être déterminé. Il existe de nombreuses méthodes pour choisir ce coefficient. Dans cette thèse la méthode de la L-curve est utilisée (Hansen and O'Leary, 1993).

Les matrices B et O sont respectivement les matrices des covariances d'erreur d'ébauche et des observations. Ces matrices pondèrent les résidus issus des observations et les résidus issus de l'ébauche. Les matrices O et B sont inconnues ou difficiles à évaluer en hydrologie. Pour simplifier leurs déterminations, la matrice O est choisie égale à la matrice identité $O = I$. Le coefficient de régularisation β est alors introduit pour pondérer les deux termes J_0 et J_{reg} de l'équation 4.3. La matrice B doit être définie par le modélisateur. Cette matrice permet de décrire des contraintes supplémentaires nécessaires à la résolution du problème. La représentation de cette matrice sera détaillée au paragraphe 4.4. Elle sera représentée telle que l'ont proposée Gejadze and Malaterre (2017).

4.2.3 Algorithme d'estimation variationnelle

L'algorithme permettant l'estimation variationnelle du vecteur de contrôle X est présenté ci-dessous (algorithme 6). Soit $H(X)$ le modèle direct, $H'_X(X)$ est le modèle tangent linéaire, $(H'_X(X))^T$ le modèle adjoint et N un optimiseur qui calcule le nouveau vecteur de contrôle optimal X^k à chaque itération k .

Algorithm 6 Algorithme d'estimation variationnelle

Require: X_b {Construction du vecteur d'ébauche}

Require: K {Nombre d'itération maximal}

Require: ϵ {Critère d'arrêt}

$X_0 = X_b$ {Initialisation du vecteur de contrôle}

$k = 0$ {Compteur des itérations}

{Boucle avec critères d'arrêt}

while $k \leq K, \|\nabla J(X_k)\| < \epsilon$ **do**

$Y_k = H(X_k), J(X_k)$ {Évaluation du modèle avec le nouveau vecteur X_k }

$\nabla J(X_k) = (H'(X_k))^T O^{-1}(Y_k - Y^*) + \beta(X_k - X_b)$ {Calcul du gradient}

$X_{k+1} = X_k - G_k \nabla J(X_k)$ $\{G_k$ est un opérateur de correspondance entre le gradient et le vecteur de contrôle}

$k = k + 1$ {Nouvelle itération}

end while

$X_a = X_k$

Require: $H(X_a)$ {Évaluation finale du modèle avec le vecteur analysé}

4.2.4 La méthode de la L-Curve

Le choix du paramètre β de l'équation 4.5 est déterminant. Il permet de choisir le bon équilibre entre :

- la minimisation des résidus calculés entre des débits simulés et observés (terme J_0)
- la minimisation de la variance des paramètres imposée par le terme de régularisation J_{reg} .

Autrement dit, si β est très grand, le terme J_0 devient négligeable devant J_{reg} et les paramètres recherchés devront satisfaire les propriétés de J_{reg} . Au contraire, si β est petit, le minimum de J sera un état qui s'approchera le mieux des observations. β est donc un paramètre permettant de réaliser un compromis entre reproduire au mieux les observations disponibles et satisfaire les contraintes imposées par le terme de régularisation. Notons que le choix de l'ébauche est très important. En effet la solution recherchée est contrainte par le vecteur d'ébauche via le terme de régularisation J_{reg} . Cependant, lors de la construction de la L-curve 4.6, la contrainte de régularisation devient efficace seulement lorsque les poids relatifs des terme J_0 et J_{reg} deviennent comparables (Gejadze and Malaterre, 2017).

Le paramètre β peut être déterminé graphiquement par la méthode de la L-curve

(Hansen and O’Leary, 1993). Dans cette méthode, le graphique suivant est construit :

$$\{J_0(X_a(\beta)), J_{reg}(X_a(\beta))\}_{\beta \in (0, \infty)}. \quad (4.6)$$

Les couples de points $(J_0(X_a(\beta)), J_{reg}(X_a(\beta)))$ sont obtenus après convergence de l’algorithme et pour une large gamme de β . Le paramètre β optimal est choisi de façon à ce que les grandeurs J_0 et J_{reg} soient minimisées (figure 4.3). Avec cette valeur du coefficient β , le vecteur de paramètre analysé X_a permet au modèle de reproduire au mieux des observations sans trop s’écarter du vecteur d’ébauche X_b .

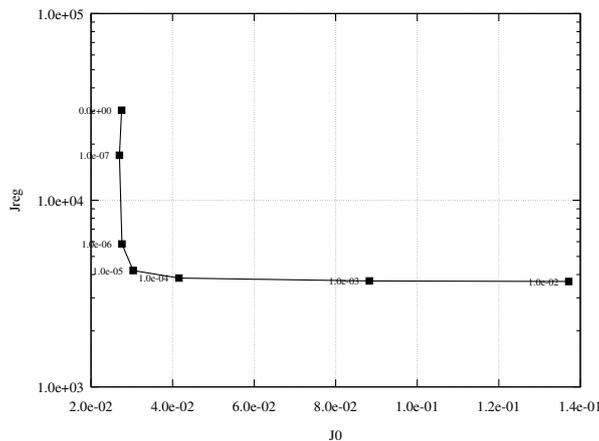


FIGURE 4.3 – Exemple d’un graphique type L-curve obtenu après sept cycles d’optimisation pour sept valeurs du coefficient de régularisation β . Le coefficient β évolue de manière croissante sur l’intervalle $[0, 0.01]$. Dans cet exemple, sa valeur optimale est estimée à 10^{-5} .

Choisir le paramètre de pondération β à l’aide de la L-Curve peut s’avérer très coûteux en temps de calcul et imprécis. En effet, la réalisation de ce graphique nécessite de répéter plusieurs fois le processus de minimisation en affinant le choix du coefficient β .

4.2.5 Régularisation itérative

La méthode de régularisation itérative (Kaltenbacher et al., 20 May. 2008; Alifanov et al., 1995) est plus efficace que la méthode de la L-curve présentée à la section 4.2.4. Dans cette méthode, le numéro de l’itération sert de paramètre de régularisation à la place du coefficient β .

Considérons la fonction coût suivante

$$J(X) = \frac{1}{2} \|O^{-1/2}(H(X) - Y^*)\|^2 + \frac{1}{2} \|B^{-1/2}(X - X_b)\|^2, \quad (4.7)$$

et considérons le changement de variables $X' = B^{-1/2}(X - X_b)$. Alors,

$$X = X_b + B^{1/2}X' \quad (4.8)$$

et la fonction coût 4.7 peut être écrite sous la forme suivante :

$$J(X') = \frac{1}{2} \|O^{-1/2}(H(X_b + B^{1/2}X') - Y^*)\|^2 + \frac{1}{2} \|X'\|^2. \quad (4.9)$$

Dans la méthode de régularisation itérative, la fonction coût 4.9 est écrite sans le terme de régularisation, c'est à dire

$$J(X') \equiv J_0(X') = \frac{1}{2} \|O^{-1/2}(H(X_b + B^{1/2}X') - Y^*)\|^2. \quad (4.10)$$

Pendant le processus de minimisation de l'équation 4.10, la L-curve est construite progressivement :

$$\left\{ J(X'_k, \frac{1}{2} \|X'_k\|) \right\}_{k=0, \dots, K}, \quad (4.11)$$

ou K est le numéro de l'itération courante et X'_k sont les vecteurs de contrôle aux itérations $k \leq K$.

Les solutions de la L-curve "paramétrique" (équation 4.6) et de la L-curve "discrète" 4.11 sont similaires. Cela signifie que les solutions obtenues après minimisation de l'équation 4.5 avec β choisi à partir de l'équation 4.6 et après minimisation de l'équation 4.10 avec la L-curve construite à l'aide de l'équation 4.11 sont à peu près équivalentes. Cependant, la méthode de régularisation itérative est plus efficace et plus robuste ([Gejadze and Malaterre, 2017](#)).

Notons que le changement de variables 4.8 est couramment utilisé dans les algorithmes d'assimilation de données utilisés en météorologie et en océanographie. La raison de ce changement de variable est la suivante : il est plus facile de définir $B^{1/2}$ que B^{-1} ou $B^{-1/2}$ pour les problèmes de grandes dimensions. Les détails quand aux raisons de cette transformation sont précisés dans [Gejadze and Malaterre \(2017\)](#).

Au cours des itérations, la norme $\|X'\|$ est calculée. Les quantités $\|X'_k\|$, $J(X'_k)$ sont sauvées en mémoire pour chaque itération $k = 0, \dots, K$. Le choix de l'itération optimale k^* s'effectue de la manière suivante : l'idée est de limiter la croissance de $\|X'\|$ si la réduction de $J(X'_k)$ devient négligeable. Cette condition est satisfaite en choisissant la plus petite itération k parmi $k \in (1, K)$ lorsque

$$\tan^{-1} \left(\frac{a_k - a_{k-1}}{b_k - b_{k-1}} \right) \times \frac{180}{\pi} > \theta, \quad (4.12)$$

où $a_k = \|X'_k\|/\|X'_K\|$, $b_k = J(X'_k)/J(X'_K)$ et θ est la valeur de l'angle limite en degré (seuil à définir).

L'algorithme d'estimation variationnelle couplé avec une méthode de régularisation itérative est présenté ci-dessous :

Algorithm 7 Variational estimation algorithm with iterative regularization.

Require: X_b {Construction du vecteur d'ébauche}

Require: K {Nombre d'itération maximal}

Require: ϵ {Critère d'arrêt}

$X'_0 = 0$ {Initialisation du vecteur de contrôle}

$k = 0$ {Compteur des itérations}

{Boucle avec critères d'arrêt}

while $k < K$, $\|\nabla J(X_k)\| < \epsilon$ **do**

$X_k = X_b + B^{1/2}X'_k$ {Calcul du vecteur de contrôle X_k (itération k) dans l'espace originel}

$Y_k = H(X_k)$, $J(X_k)$ {Évaluation du modèle avec le nouveau vecteur X_k }

$\nabla J(X_k) = (H'(X_k))^T O^{-1}(Y_k - Y^*)$ {Calcul du gradient}

$X'_{k+1} = X'_k - G_k B^{T/2} \nabla J(X_k)$ { G_k est un opérateur de correspondance entre le gradient et le vecteur de contrôle dans l'espace transformé}

$k = k + 1$ {Nouvelle itération}

end while

$X_a = X_{k^*}$ { k^* est la plus petite valeur de k qui satisfait la condition 4.12}

Require: $H(X_a)$ {Évaluation finale du modèle avec le vecteur analysé}

4.2.6 Contraintes d'inégalité et normalisation du vecteur de contrôle

Dans cette partie, la formulation standard de l'algorithme d'assimilation variationnelle est modifiée afin d'intégrer des contraintes d'inégalités sur le vecteur de contrôle. Cette modification est importante puisque les paramètres du modèle hydrologique sont bornés (par exemple, les capacités des réservoirs de production et de transfert sont toujours positives).

Considérons les contraintes d'inégalités suivantes appliquées au vecteur de contrôle X :

$$X_{min} \leq X \leq X_{max}, \quad (4.13)$$

où X_{min} et X_{max} sont respectivement les bornes minimum et maximum du vecteur de contrôle. Le facteur d'échelle \tilde{X} est introduit tel que

$$X = X_{min} + \tilde{X} \cdot (X_{max} - X_{min}), \quad 0 \leq \tilde{X} \leq 1. \quad (4.14)$$

où \cdot correspond au produit membre à membre (Hadamard). Notons que la normalisation du vecteur de contrôle est importante. Cela permet d'améliorer la convergence.

La matrice B peut être représentée sous la forme

$$B = \sigma_X \cdot I C \sigma_X \cdot I \quad (4.15)$$

où σ_X est le vecteur contenant la variance de X , C est la matrice de corrélation, I la matrice identité.

Le terme de régularisation (pénalité) de l'équation 4.7 peut être écrit sous la forme suivante :

$$\|C^{-1/2}(\tilde{X} - \tilde{X}_b) \cdot \sigma_X^{-1} \cdot (X_{max} - X_{min})\|_{L^2}^2.$$

Si l'on suppose que $\sigma_X^{-1} \cdot (X_{max} - X_{min}) = const = \beta$, la fonction coût 4.7 devient

$$J(\tilde{X}) = \frac{1}{2} \|O^{-1/2}(H(X_{min} + \tilde{X} \cdot (X_{max} - X_{min})) - Y^*)\|^2 + \frac{\beta}{2} \|C^{-1/2}(\tilde{X} - \tilde{X}_b)\|^2, \quad (4.16)$$

et la solution optimale est obtenue en minimisant la fonction J sous les contraintes d'inégalité :

$$\tilde{X}_a = \arg \min_{\tilde{X}} J(\tilde{X}); \quad 0 \leq \tilde{X} \leq 1. \quad (4.17)$$

Ce problème peut être résolu en utilisant le minimiseur LBFGS-B (Limited memory Broyden-Fletcher-Goldfarb-Shanno Bound-constraint) (Zhu et al., 1994). Cet optimiseur utilise une méthode dite Quasi-Newton, optimale pour les problèmes d'optimisation de grande dimension. C'est un algorithme à pas de descente optimal et qui permet de borner le vecteur de contrôle

Cependant, l'implémentation des contraintes d'inégalité (bornes) avec la méthode de régularisation itérative est difficile. Le changement de variable présenté à la section 4.2.5 est alors introduit tel que

$$\tilde{X}' = C^{-1/2}(\tilde{X} - \tilde{X}_b), \quad (4.18)$$

alors,

$$\tilde{X} = \tilde{X}_b + C^{1/2} \tilde{X}', \quad (4.19)$$

et la fonction coût devient

$$J(\tilde{X}') = \frac{1}{2} \|O^{-1/2}(H(X_{min} + (\tilde{X}_b + C^{1/2} \tilde{X}') \cdot (X_{max} - X_{min})) - Y^*)\|^2 + \frac{\beta}{2} \|\tilde{X}'\|^2.$$

De même que précédemment, le terme de régularisation est exclu. Il en résulte la fonction coût suivante :

$$J(\tilde{X}') = \frac{1}{2\beta} \|O^{-1/2}(H(X_{min} + (\tilde{X}_b + C^{1/2} \tilde{X}') \cdot (X_{max} - X_{min})) - Y^*)\|^2. \quad (4.20)$$

Le problème de minimisation est le suivant :

$$\tilde{X}'_a = \arg \min_{\tilde{X}} J(\tilde{X}'); \quad -\tilde{X}_b \leq C^{1/2} \tilde{X}' \leq 1 - \tilde{X}_b. \quad (4.21)$$

Des solveurs spéciaux, tel que le SDP (Sequential Dynamic Programming) ([Gilbert, 2009](#)), doivent être utilisés pour résoudre le problème 4.21. Ces types de solveurs sont très coûteux en temps de calcul.

Comme les contraintes d'inégalité sont linéaires, le problème peut être résolu tel que le propose [Gejadze and Shutyaev \(2020\)](#) :

$$\tilde{X}'_a = \arg \min_{\tilde{X}} J(\tilde{X}'); \quad \frac{u - v}{2} \leq \tilde{X}' \leq \frac{u + v}{2}, \quad (4.22)$$

où u et v satisfont le système suivant

$$|C^{1/2}|v = I, \quad C^{1/2}u = I - 2\tilde{X}_b, \quad I = (1, \dots, 1)^T. \quad (4.23)$$

Le système d'équations 4.23 doit être résolu pour u et v avant minimisation. Remarquons que le vecteur u dépend de \tilde{X}_b et doit être recalculé pour chaque nouveau \tilde{X}_b . Le problème 4.22 peut-être résolu à l'aide du minimiseur LBFGS-B. Lors du processus de minimisation $\frac{\beta}{2} \|\tilde{X}'_k\|$ est utilisé pour sélectionner le numéro approprié de l'itération k^* satisfaisant la condition 4.12.

L'algorithme d'estimation variationnelle intégrant la méthode de régularisation itérative et de contraintes d'inégalité est présenté ci-dessous :

Algorithm 8 Algorithme d'estimation variationnelle avec régularisation itérative et contraintes d'inégalité

Require: X_b {Construction du vecteur d'ébauche}
Require: X_{max}, X_{min} {Bornes maximum et minimum de X }
Require: \tilde{X}_b {Vecteur d'ébauche normé}
Require: K {Nombre d'itération maximal}
Require: ϵ {Critère d'arrêt}
 $\tilde{X}'_0 = 0$ {Initialisation du vecteur de contrôle}
 $k = 0$ {Compteur des itérations}
{Boucle avec critères d'arrêt}
while $k < K, \|\nabla J(X_k)\| < \epsilon$ **do**
 $X_k = X_{min} + (\tilde{X}_b + C^{1/2}\tilde{X}'_k)(X_{max} - X_{min})$ {Calcul du vecteur de contrôle X_k (itération k) dans l'espace originel}
 $Y_k = H(X_k), J(X_k)$ {Évaluation du modèle avec le nouveau vecteur X_k }
 $\nabla J(X_k) = (H'(X_k))^T O^{-1}(Y_k - Y^*)$ {Calcul du gradient}
 $\tilde{X}'_{k+1} = \tilde{X}'_k - (G_k(X_{max} - X_{min})B^{T/2}\nabla J(X_k))$ { G_k est un opérateur de correspondance entre le gradient et le vecteur de contrôle dans l'espace transformé}
 $k = k + 1$ {Nouvelle itération}
end while
 $X_a = X_{k^*}$ { k^* est l'itération k la plus petite qui satisfait la condition 4.12}
Require: $H(X_a)$ {Évaluation finale du modèle avec le vecteur analysé}

4.3 Critères d'estimation utilisés pour le calcul de la fonction coût

Dans cette partie, les critères d'optimisation utilisés sont présentés. En estimation variationnelle, une fonction objectif quadratique est le plus souvent utilisée. Ce type de fonction est présenté à la section 4.3.1. Une fonction coût dérivant des méthodes d'estimation robuste est ensuite présentée à la section 4.3.2.

4.3.1 Fonction quadratique, méthode des moindres carrés

Les méthodes classiques d'estimation utilisent des fonctions coûts quadratiques. Ces fonctions sont composées d'un critère quadratique et d'un terme de pénalité. Elles suivent une approche Bayésienne et assument que la distribution statistique de résidus est Gaussienne. Ce type de fonction admettra en théorie une solution dans un domaine borné. En modélisation hydrologique pluie-débit, la fonction coût est calculée entre les débits observés et simulés aux stations de jaugeage. Soit $P^*(x, t)$ et $E^*(x, t)$ respectivement les données de pluie et d'évapotranspiration définis $\forall x \in \Omega$ et pour chaque pas de temps $t \in (0, T)$. Soit H le modèle hydrologique (présenté au chapitre 3) et $X(x) = (c_p(x), c_{tr}(x), \xi(x))$ le vecteur de paramètres qui inclue les capacités des réservoirs de production et de transferts et les paramètres de routage pour chaque nœud du domaine Ω . Soit N le nombre de station de jaugeage. Les débits simulés aux stations de

jaugeage localisées aux nœuds $x_k \in \Omega$, $k = 1, N$ se calculent par l'équation suivante :

$$Q_k(t) = H(P^*(x, t'), E^*(x, t'), X(x), t), \forall x \in \Omega, t \in (0, T). \quad (4.24)$$

La fonction coût est calculée sur une fenêtre temporelle définie par les pas de temps $t \in (t^*, T)$ et $t' \in (t^*, t)$ où t^* est la période de chauffe (relaxation). La fonction coût s'écrit de la manière suivante :

$$J_0(X) = \sum_{k=1}^N w_k^{-1}(t^*) \int_{t=t^*}^T (H(P^*, E^*, X, t) - Q_k^*(t))^2 dt \quad (4.25)$$

où les coefficient w_k sont des facteurs d'échelle. Si l'on considère

$$w_k(t^*) = \int_{t=t^*}^T (\langle Q_k^* \rangle - Q_k^*(t))^2 dt,$$

où $\langle Q_k^* \rangle$ est la moyenne temporelle de $Q_k^*(t)$ calculée au niveau du nœud k . Dans ce cas la fonction J_0 correspond au calcul

$$J_0 = \sum_{k=1}^N (1 - NSE_k) \quad (4.26)$$

où NSE_k est le critère de Nash-Sutcliffe (Nash-Sutcliffe Efficiency criteria) (Nash and Sutcliffe, 1970) calculé à la $k^{\text{ème}}$ station de jaugeage. Ce critère est très utilisé en hydrologie pour comparer les hydrogrammes simulés et observés. La formulation classique du critère NSE est donnée ci-dessous :

$$NSE_k = 1 - \frac{\sum_i^m (Q_k(t_i) - Q_k^*(t_i))^2}{\sum_i^m (Q_k^*(t_i) - \langle Q_k^* \rangle)^2}, \quad (4.27)$$

où $t_i = (i - 1)\Delta t$ et Δt est le pas de temps. Le critère NSE évolue entre $-\infty$ et 1. Lorsque $NSE = 1$, les débits modélisés et observés sont identiques. Plus le NSE diminue et plus les débits modélisés et observés diffèrent. Notons que la fonction $J_0 = 1 - NSE$ est définie sur l'intervalle $[0, +\infty[$.

4.3.2 Fonction non-gaussienne, méthode d'estimation robuste

Le calage des modèles hydrologiques est difficile. La première raison de cette difficulté est l'existence d'erreurs significatives sur les données d'entrées et de sorties du modèle (mesures de la pluviométrie, mesures de débit). Ces erreurs dépendent de l'intensité de l'événement, de sa localisation et de sa durée. Ainsi, aucune description statistique fiable de ces erreurs est généralement disponible. La deuxième raison est le manque d'adéquation entre les modèles conceptuels et les phénomènes hydrologiques. Les modèles conceptuels reproduisent les caractéristiques globales et moyennes des processus hydrologique mis en

jeu. Par conséquent, la réaction du modèle hydrologique à un signal d'entrée (la pluie) est parfois très différente de la réaction naturelle observée.

A cause des raisons évoquées ci-dessus, les prédictions du modèle ainsi que les données observées contiennent une proportion significative de «valeurs aberrantes» (erreurs d'observation ou erreur de modélisation). Ces valeurs aberrantes peuvent corrompre le fonctionnement des méthodes de calage. Pour contourner ce problème, des méthodes d'estimation dites "robustes" peuvent être considérées pour le calage des paramètres du modèle hydrologique. Une telle méthode sera testée dans le chapitre 7.

La robustesse d'une fonction coût est caractérisé par la proportion maximum de valeurs "extrêmes et/ou inhabituelles" ("outliers") pour laquelle la solution peut-être obtenue. Avec l'approche des moindres carrés, cette proportion est de zéro. Cela signifie qu'une seule et unique valeur "extrême" peut corrompre le processus d'estimation. Pour contourner ce problème, des méthodes d'estimation "robustes" sont disponibles (Rousseeuw and Leroy, 2005). Il s'agit par exemple des méthodes LMedS (Least Median of Squares), LTS (Least Trimmed Squares), M-estimators et Voting (RANSAC) (Fischler and Bolles, 1981). Pour toutes ces méthodes, le calcul des résidus est formulé de la façon suivante :

$$r(t) = \rho(Q(X, t) - Q^*(t))$$

Où ρ est une fonction non-linéaire. Le choix de cette fonction dépend de la méthode utilisée. Cependant, dans les méthodes existantes énoncées ci-dessus, cette fonction n'est pas nécessairement différentiable. Sa différentiabilité est une condition nécessaire à son utilisation dans un contexte d'estimation variationnelle.

Lors de l'utilisation d'une fonction coût classique, tel que présentée à la section 4.3.1 (approche Bayésienne classique), on suppose que les erreurs dominantes proviennent des mesures des débits Q (voir équation 4.24) :

$$Q_k^*(t) = H(\bar{P}(x, t'), \bar{E}(x, t'), h(x, 0), \bar{X}(x), t) + \epsilon.$$

où $\bar{P}(x, t')$, $\bar{E}(x, t')$ et $\bar{X}(x)$ sont respectivement les valeurs "vraies" de la pluie et des paramètres, $\forall x \in \Omega$ et $t' \in (0, t)$. C'est une hypothèse très forte, surtout lorsque l'on utilise un modèle hydrologique conceptuel. En effet, il n'existe pas de valeurs "vrai" de θ . Une nouvelle hypothèse est donc formulée : les débits observés et simulés appartiennent à la même fonction de distribution. Dans ce cas, la distance entre les débits observés et simulés peut être mesurée en comparant leurs probabilités d'occurrence correspondantes. Cette approche suppose de travailler sur une fenêtre temporelle suffisamment longue pour disposer d'une longue séquence d'observation. Après observation des chroniques de débits observés, la fonction de densité de probabilité Gamma a été choisie (voir équation 3.25

chapitre 3). Comme le débit est une grandeur strictement positive, la distance entre les débits observés et simulés est écrite de la façon suivante :

$$\mathcal{P}[Q(t)] - \mathcal{P}[Q^*(t)] = f(Q(t); \mu, \sigma) - f(Q^*(t); \mu, \sigma),$$

Où μ représente la position du mode de la fonction *Gamma* et σ définit son étalement (de manière analogue au paramètre σ de la fonction de distribution Gaussienne). La forme convexe de la fonction coût est obtenue en multipliant cette différence par les résidus normalisés $r(t)$:

$$J_0(X, t) = (f(Q(X, t); \mu, \sigma) - f(Q^*(t); \mu, \sigma))(Q(X, t) - Q^*(t)). \quad (4.28)$$

Comme $f(Q(t); \mu, \sigma) - f(Q^*(t); \mu, \sigma) \leq 1$, les impacts des forts écarts sont proportionnels à $|r(t)|$. À l'opposé, les forts écarts sont proportionnels à $r^2(t)$ avec une approche utilisant les moindres carrés. Avec cette approche, il est important de noter que aussi bien les débits observés $Q^*(t)$ et simulés $Q(t)$ peuvent être considérés comme "extrêmes et/ou inhabituels". Contrairement avec la méthode M-estimators, la fonction suggérée est différentiable.

Finalement, la fonction coût s'écrit de la façon suivante :

$$J_0(X) = \int_0^T J_0(X, t) dt \quad (4.29)$$

La fonction coût Gamma J est représentée à la figure suivante 4.4 :

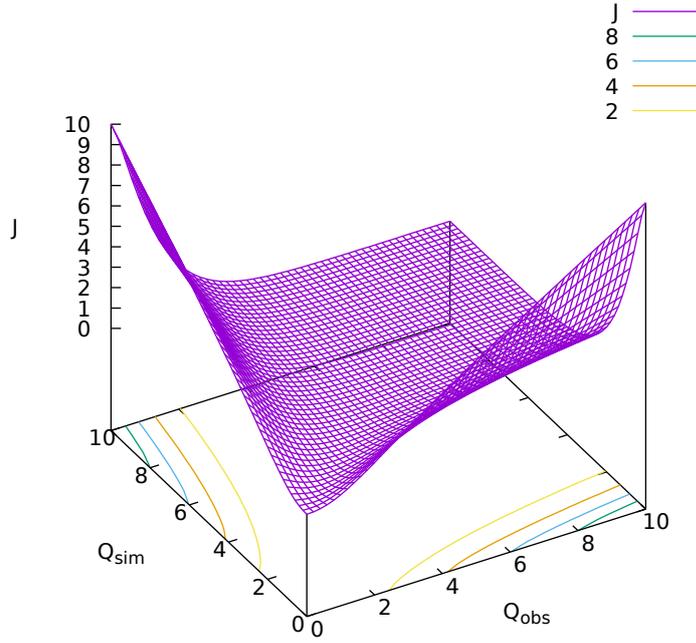


FIGURE 4.4 – Valeur de la fonction coût Gamma présentée à l'équation 4.28 pour $\mu = 1$ et $\sigma = 1$.

Lors de l'application de l'équation 4.28 les paramètres μ et σ doivent être déterminés. Par exemple, ils peuvent être déduits des chroniques d'observation $Q_i^* = Q^*(t_i)$, $i = 1, \dots, N$. Il est alors possible de considérer l'approche suivante :

$$\mu = \bar{Q}^* = \frac{1}{T} \int_0^T Q^*(t) dt,$$

et

$$\sigma = k\bar{Q}^*$$

où k doit être choisi empiriquement.

4.4 Le terme de régularisation

4.4.1 Formulation du terme de régularisation de type Tikhonov

Le terme de régularisation dans les équations 4.3 et 4.5 est formulé de la façon suivante :

$$J_{reg}(X - X_b) = \frac{1}{2} \|B^{-1/2}(X - X_b)\|^2 = \frac{1}{2} (X - X_b)^T B^{-1} (X - X_b).$$

Pour définir ce terme, l'inverse de la matrice des covariances B doit être définie. Au lieu d'inverser la matrice B , celle-ci peut être générée tel que proposé [Gejadze et al. \(2010\)](#).

Cette approche est basée sur l'hypothèse que le vecteur de contrôle X représente une fonction discrétisée d'une variable continue x appartenant à l'espace Sobolev (c'est à dire une fonction qui présente une certaine régularité dans l'espace).

Considérons une fonction en deux dimension $z(x_1, x_2)$ sur un domaine rectangulaire discrétisé sur par grille uniforme de $n \times m$ points de discrétisation. Définissons le vecteur correspondant Z de taille $n \times m$ tel que $Z_{(j-1)m+i} = z_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, m$. En pratique, ce vecteur est un sous-ensemble de X qui représente une des variables distribués du vecteur de contrôle $c_p(x)$, $c_t(x)$, $\xi(x)$. La norme de Sobolev de Z est définie de la façon suivante :

$$J_{reg}(Z) = Z^T B^{-1} Z \equiv \|Z\|_{W_2^2}^2 = J_{reg}^{(1)}(z) + J_{reg}^{(2)}(z),$$

où

$$J_{reg}^{(1)}(z) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m \frac{\alpha_{i,j}}{\sqrt{\omega_{i,j}}} z_{i,j}^2 \quad (4.30)$$

et

$$J_{reg}^{(2)}(z) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m \frac{\alpha_{i,j}}{\sqrt{\omega_{i,j}}} \times \frac{1}{dx} \left[\left(\omega_{i+1/2,j} \frac{z_{i+1,j} - z_{i,j}}{dx} - \omega_{i-1/2,j} \frac{z_{i,j} - z_{i-1,j}}{dx} \right)^2 + \left(\omega_{i,j+1/2} \frac{z_{i,j+1} - z_{i,j}}{dx} - \omega_{i,j-1/2} \frac{z_{i,j} - z_{i,j-1}}{dx} \right)^2 \right] \quad (4.31)$$

où $\omega_{i+1/2,j} = (\omega_{i+1,j} + \omega_{i,j})/2$, etc.

Sur les limites du domaine, lorsque $i = 1$ ou $i = n$ ou $j = 1$ ou $j = m$, ce calcul est effectué en imposant les conditions limites suivantes :

$$\begin{aligned} z(n+1, j) &= z(n, j) \\ z(0, j) &= z(1, j) \\ z(i, m+1) &= z(i, m) \\ z(i, 0) &= z(i, 1) \\ z(n+1, m+1) &= z(n, m) \\ z(0, 0) &= z(1, 1) \\ z(n+1, 0) &= z(n, 1) \\ z(0, m+1) &= z(1, m) \end{aligned}$$

La formulation J_{reg}^1 représente la pénalité causée par l'écart entre le vecteur d'ébauche et le vecteur de contrôle sur le domaine.

La formulation J_{reg}^2 représente la pénalité causée par la variabilité spatiale du vecteur de contrôle sur le domaine.

Les matrices α et ω sont des paramètres à ajuster. ω contrôle le rayon de corrélation spatiale, alors que β contrôle la variance des paramètres. Plus le rayon de corrélation ω augmente et plus les paramètres évolueront graduellement dans l'espace. Le paramètre ω doit être ajusté empiriquement. La matrice α contrôle la variance des paramètres. Si $\alpha = 1$, alors la variance peut être contrôlé par le coefficient de régularisation β (voir équation 4.5). Ce coefficient est déterminé graphiquement par la méthode de la L-curve (méthode décrite à la section 4.2.4). Le choix de ce coefficient est crucial. Il permet de déterminer une solution qui minimise à la fois le critère d'optimisation et les contraintes de régularisation.

4.4.2 Formulation de la matrice $B^{1/2}$ pour la méthode de la régularisation itérative

La matrice $B^{1/2}$ doit être connue explicitement pour la méthode de régularisation itérative. Pour construire cet opérateur, l'approche suivante est choisie. Considérons une fonction $z(x)$ et le vecteur correspondant Z de taille n . Définissons la norme de Sobolev suivante :

$$Z^T B_n^{-1} Z = \|Z\|_{W_2^2}^2 = \frac{1}{2} \sum_{i=1}^n \frac{1}{\sqrt{\omega_i}} z_i^2 + \frac{1}{2} \sum_{i=2}^{n-1} \frac{1}{\sqrt{\omega_i}} \frac{1}{dx} \left(\omega_{i+1/2,j} \frac{z_{i+1,j} - z_{i,j}}{dx} - \omega_{i-1/2,j} \frac{z_{i,j} - z_{i-1,j}}{dx} \right)^2.$$

La matrice B_n^{-1} est une matrice diagonale de longueur n . Sa décomposition de Choleski s'écrit de la manière suivante :

$$B_n^{-1} = B_n^{-1/2} B_n^{-T/2}.$$

Pour un facteur $B_n^{-1/2}$ donné, le produit $X' = B_n^{1/2} X$ peut être défini en résolvant le système

$$B_n^{-1/2} X' = X$$

Considérons maintenant la fonction original discrétisée $z_{i,j}$ et définie sur deux dimensions. Notons z_j la j^{th} colonne de z (dimension n), et z_i la i^{th} ligne de z (dimension m). Premièrement, on applique l'opérateur $B_n^{1/2}$ à chaque colonne en résolvant l'équation $B_n^{-1/2} z'_j = z_j$, $j = 1, \dots, m$. Ensuite, on définit $z = z'$. Enfin, on applique l'opérateur $B_m^{1/2}$ à chaque ligne z_i en résolvant l'équation $B_m^{-1/2} z'_i = z_i$, $i = 1, \dots, n$. Il en résulte la fonction z' qui est une approximation du produit matrice-vecteur impliquant la matrice $B^{1/2}$. Comme B_n^{-1} ne dépend pas de Z , les facteurs $B_n^{-1/2}$, $B_m^{-1/2}$ peuvent être pré-calculés et stockés en mémoire.

4.5 Détails pratiques quant à l'implémentation et l'utilisation de l'algorithme d'estimation variationnelle dans GRD

L'algorithme d'estimation variationnelle est implémenté dans GRD. Quelques détails, contraintes et fonctionnalités supplémentaires sont présentés dans cette partie : les paramètres sont bornés ; les variables du vecteur de contrôle sont normalisées afin d'améliorer la convergence ; la taille du vecteur de contrôle peut être réduite au besoin ; La détermination du coefficient de régularisation de la méthode de la L-curve est simplifiée par un méthode empirique ; le vecteur d'ébauche peut être évalué à l'aide d'algorithmes de calage "global" (stochastique).

4.5.1 Choix des bornes

L'espace des paramètres est borné selon l'équation 4.13. Les opérateurs de GRD sont décrits par des paramètres évoluant dans une gamme de valeurs prédéfinies. Ces bornes sont définies à partir de considérations physiques, numériques ou de connaissances empiriques.

Les contraintes numériques et physiques imposent la borne minimum $X_{min} > 0$ pour l'ensemble des paramètres du modèle GRD. Notons qu'il est important de limiter l'inertie du modèle (retard maximum) afin d'améliorer les performances de calcul (durée de la simulation). Pour cela les bornes minimales des paramètres de routage des variantes 1 et 2 (respectivement v et ξ) sont significativement supérieures à zéro.

Les bornes supérieures X_{max} sont plus délicates à évaluer. Les opérateurs du modèle GRD sont conceptuels et leurs sens physiques "à priori" ne justifient pas des valeurs déterminées pour les bornes supérieures. En revanche, à partir d'expériences numériques, il semble qu'au-delà de certaines valeurs de X , le modèle devient insensible. Par exemple, une vitesse de routage supérieure à $5 m.s^{-1}$, ne modifie plus le temps de retard entre deux mailles pour le modèle GRD fonctionnant au pas de temps horaire. Il est donc préférable d'imposer des limites maximales afin de conserver un modèle dynamique (dont les états varient significativement au cours du temps). De plus ces bornes permettent de mieux contraindre le problème.

Les ordres de grandeurs des bornes inférieures et supérieures sont données dans le tableau 4.1. Les ordres de grandeurs de ces bornes sont généralement utilisés pour les expériences de calage présentées dans les chapitres suivants. Ces bornes sont souvent utilisées pour estimer le vecteur d'ébauche. Elles sont ensuite réajustées (relaxées) en fonction de ce vecteur.

TABLE 4.1 – Ordre de grandeur des bornes inférieures et supérieures des paramètres du modèle GRD utilisées pour les simulations.

Paramètres	Bornes inférieures	Bornes supérieures
c_p	0.1 mm	5000 mm
c_{tr}	0.01 mm	2000 mm
v	0.5 m.s^{-1}	5 m.s^{-1}
ξ	0.5	10
c_r	0.1 min	1000 min

4.5.2 Détermination automatique du coefficient β pour la méthode de la L-curve

Dans ce paragraphe, une méthode empirique permettant de calculer le coefficient de régularisation β est suggérée. Soit la courbe L-curve tracée à partir des couples de points (J_0, J_{reg}) . J_0 est le coût associé au critère de minimisation quadratique et J_{reg} est le coût associé au terme de régularisation. La fonction objectif J est la suivante :

$$J = J_0 + \beta J_{reg},$$

où β est le poids associé au terme de régularisation.

L'objectif est de minimiser à la fois J_0 et J_{reg} . Cela suppose que les grandeurs J_0 et J_{reg} soient du même ordre de grandeur et qu'aucune d'entre elles ne soit négligeables l'une devant l'autre. L'idée est donc de choisir β tel que $J_0 \approx J_{reg}$ au fil des itérations. J_{reg} étant inconnue ou nulle à la première itération du processus d'estimation variationnelle, il convient de l'évaluer. Cette valeur seuil notée J_{reg}^{max} définira la limite supérieure de variabilité du vecteur de contrôle.

La variabilité maximale du vecteur de contrôle peut être déterminée en procédant à un premier cycle d'itération $n \in [1 : N]$ sans contrainte de régularisation, soit $\beta = 0$. Ce premier cycle permet de converger vers un minimum (local) qui donne une estimation de la variabilité maximale du vecteur de contrôle J_{reg}^{max} associée à une valeur minimum du critère noté J_0^{min} . On calcule alors le coefficient de régularisation de la manière suivante :

$$\beta_{opt} = \frac{J_0^{min}}{J_{reg}^{max}} = \frac{J_0^N(\beta = 0)}{J_{reg}^N(\beta = 0)}$$

Le processus itératif est ensuite redémarré en appliquant la méthode de régularisation de type Tikhonov avec le coefficient $\beta = \beta_{opt}$.

Cette méthode empirique de détermination du coefficient de régularisation est testée et comparée avec le graphique de la L-curve. Les expériences et résultats sont présentés au chapitre 5 section 5.3.

4.5.3 Réduction de la taille du vecteur de contrôle

Une méthode permettant de réduire la taille du vecteur de contrôle est implémentée dans l'algorithme de calage variationnel. Le maillage régulier peut être découpé, au besoin, en "super-maille" sur lesquelles les paramètres du modèle prennent une valeur uniforme. Soit un maillage constitué de $1..N$ mailles m d'indices n . Ce maillage est découpé en K sous domaines délimités par les limites j^k . Toutes les mailles m^n prennent les valeurs tel que :

$$\begin{aligned}\forall n \in [1 : j_1], m^n &= 1 \\ \forall n \in [j_k : j_{k+1}], m^n &= k \\ \forall n \in [j_{K-1} : j_K], m^n &= K\end{aligned}$$

où

$$j_K = N \text{ et } j_k \in [1 : N]$$

Le vecteur de contrôle contient alors $K \times p$ variables, où p est le nombre de paramètres par maille. $K \times p$ paramètres sont alors estimés avec l'algorithme variationnel. Ces $K \times p$ paramètres sont uniformisés sur les K super-maillages définies par les domaines $[j_k : j_{k+1}]$ (figures 4.5). En pratique, ces contraintes spatiales sont définies à l'aide d'une matrice présentée à la figure 4.5. Cette matrice est lue par GRD et permet de définir spatialement ces super-maillages.

1	1	2
1	1	2
1	2	2
3	2	2
3	3	3
3	3	3

FIGURE 4.5 – Exemple d’un maillage 6×3 découpée en trois sous domaines (super-maillages). Les paramètres calés seront uniformes sur chacun des trois sous-domaines.

Cette méthode de réduction du vecteur de contrôle est utilisée pour calibrer de manière uniforme les paramètres du modèle sur les mailles ayant des caractéristiques hydrogéologiques similaires (géologie, hydrogéologie et occupation des sols). Ces contraintes seront étudiées dans le chapitre 7. Les données utilisées pour construire ces descripteurs ont été décrites au chapitre 2 section 2.1.3.

4.5.4 Obtention du modèle adjoint

Les méthodes variationnelles requièrent le calcul du gradient ∇J de la fonction objectif J par rapport au vecteur de contrôle X . Ce calcul peut parfois s’avérer complexe. Une solution simple consiste à calculer le taux d’accroissement pour chacune des variables du système et dans chaque direction. Le calcul du gradient va donc demander plusieurs simulations du modèle pour chaque composante du gradient. Dans le cas de problème de grandes dimensions, le nombre de paramètres est très grand et cette méthode n’est pas envisageable. De plus, le gradient ne serait pas exact puisqu’il est calculé de manière approximative.

Le développement du modèle adjoint est alors pertinent car il permet le calcul du gradient par rapport à l’ensemble du vecteur de contrôle. Ce calcul nécessite un seul calcul du modèle direct suivi par un seul calcul du modèle adjoint. Ce calcul est exact car il n’y a pas de passage à la limite.

Tapenade est un outil de différentiation automatique d’un programme informatique (Hascoet and Pascual, 2013). Il génère un code qui calcule les dérivées adjointes de la fonction d’origine. Ce nouveau code permet de calculer très efficacement les gradients d’une fonction.

Le code généré par Tapenade est préalablement validé : le gradient calculé approximativement à partir d’un taux d’accroissement du modèle direct est comparé avec le gradient

calculé par le modèle adjoint. Ce calcul est répété pour chaque paramètre du modèle et pour plusieurs valeurs de δX suffisamment petites (x étant le vecteur de contrôle).

4.6 Méthode de calage globale pour l'évaluation de l'ébauche

Le vecteur d'ébauche X_b doit être estimé avant de débiter l'optimisation variationnelle. Celui-ci peut être estimé à partir d'algorithmes de calage qui implémentent des méthodes de recherche dites "globales". Les paramètres optimaux sont alors recherchés en parcourant l'ensemble de l'espace des paramètres et évitent potentiellement de converger vers des minima locaux. Ces méthodes diffèrent des algorithmes de calage variationnel qui recherchent un minimum local de la fonction objectif autour du vecteur d'ébauche (valeur initiale des paramètres). Cependant, ces algorithmes ne sont pas adaptés pour les problèmes de grandes dimensions. Le vecteur de contrôle est donc réduit au maximum, soit trois paramètres dans le cas de GRD. Considérons un domaine spatial $\Omega(i, j)$ sur lequel les paramètres sont recherchés tels que :

$$c_p(i, j) = \bar{c}_p, \quad c_{tr}(i, j) = \bar{c}_{tr}, \quad v(i, j) = \bar{v}, \quad \forall i, j \in \Omega$$

Dans ce cas, les paramètres prennent une valeur uniforme sur le maillage et seule une valeur moyenne est recherchée pour chacun des paramètres du modèle GRD. Le vecteur de contrôle est constitué de seulement trois éléments. Dans ce cas, le calage est dénommé "calage uniforme".

La méthode de calage présentée ici se décompose en deux étapes successives : le pré-filtrage et l'algorithme "pas à pas".

Pré-filtrage

Soit X^n le n^{eme} paramètre du modèle GRD avec $n \in [1 : N]$ et N le nombre de paramètres décrivant les processus sur chaque maille ($N = 3$, trois paramètres au total pour GRD). Le pré-filtrage consiste à échantillonner avec un pas d'espace régulier d_k , $k \in [1 : K]$, l'espace des paramètres $[X_{min}^n : X_{max}^n]$. L'échantillonnage du paramètre X^n est construit de la manière suivante :

$$X_{k+1}^n = X_k^n + d_k$$

Ensuite l'ensemble des combinaisons des X_k^n , $n = 1..N$, $k = 1..K$ sont réunies dans le vecteur de contrôle \hat{X} tel que :

$$\hat{X} = [(X_1^1, X_1^n, X_1^N), \dots, (X_i^1, X_i^n, X_i^N), \dots, (X_K^1, X_K^n, X_K^N)] \quad \forall i, j, k \in [1 : K] \text{ et } \forall n \in [1 : N]$$

La fonction objectif $J(\hat{X})$ est calculée pour chaque valeur du vecteur \hat{X} . Le jeu de paramètres optimal est celui qui minimise la fonction objectif J .

Le pré-filtrage peut vite devenir très coûteux en temps de calcul. Pour les trois paramètres du modèle GRD, l'espace dans lequel évolue chaque paramètre est échantillonné en cinq valeurs, soit 5^3 combinaisons (vecteur \hat{X}). La fonction objectif est donc évaluée 125 fois ce qui correspond à autant de simulations du modèle direct.

Algorithme de calage "pas à pas"

L'algorithme "pas à pas" est un algorithme de minimisation simple. Cet algorithme converge vers un minimum approximatif de la fonction objectif. Cet algorithme est documenté dans [Michel C. \(1989\)](#) et [Edijatno \(1991\)](#). Il semble suffisant pour caler les paramètres des modèles conceptuels. Dans le cas d'un calage uniforme, cet algorithme est préféré à l'algorithme variationnel (décrit à la section 4.2.3). L'algorithme "pas à pas", couplée au résultat d'un pré-filtrage des paramètres (section 4.6), recherche le minimum "global" de la fonction objectif. En effet, cet algorithme ne converge pas vers un minimum local mais balaye largement la gamme de variation des paramètres. Le calage uniforme pourra déterminer le vecteur d'ébauche, point de départ du calage variationnel.

4.7 Conclusions

Dans ce chapitre, la totalité des méthodes mathématiques et numériques utilisées pour le calage des paramètres du modèle distribué GRD ont été présentée.

Un algorithme d'estimation variationnelle utilisant le modèle adjoint a été décrit et implémenté. Il inclue deux méthodes de régularisation de type Tikhonov (la méthode L-Curve et la méthode itérative). Cet algorithme est utilisé pour caler les paramètres distribués du modèle GRD (expériences de calage présentées aux chapitres 5, 6 et 7). Les simulations réalisées dans le chapitre 5 permettront de tester les principales fonctionnalités de l'algorithme (le fonctionnement du modèle adjoint, les méthodes de régularisation, l'utilisation de l'optimiseur incluant la recherche d'un vecteur de paramètres borné). Les simulations réalisées dans les chapitres 6 et 7 permettront d'évaluer le calage distribué des paramètres du modèle GRD.

Un algorithme de recherche global a été présenté rapidement à la toute fin de ce chapitre (section 4.6). Cet algorithme est implémenté dans GRD et sert à estimer une valeur uniforme pour chaque paramètre du modèle (trois valeurs). Cet algorithme permet de balayer l'espace des paramètres et de rechercher une solution "globale". Il sera utilisé pour évaluer le vecteur d'ébauche dans les expériences de calage aux chapitres 6 et 7.

Deux fonctions objectif pour le calage des paramètres ont été décrites. Ces fonctions

objectifs se distinguent par le choix du critère utilisé pour évaluer l'écart entre les variables observées et modélisées. La fonction objectif quadratique utilisant le critère de NSE (section 4.3.1) sera principalement utilisée pour les expériences de calage réalisées dans les chapitres suivant. La fonction objectif "robuste" (section 4.3.2) sera testée de façon exploratoire dans le chapitre 7.

Enfin, une fonctionnalité du modèle permettant de réduire la taille du vecteur de contrôle en groupant des mailles du domaine a été présentée (section 4.5.3). Cette fonctionnalité sera utilisée dans des expériences de calage réalisées dans le chapitre 7. Pour ces expériences des descripteurs physiques seront utilisés pour contraindre la variabilité spatiale des paramètres. Cela nécessite de grouper les mailles du domaine de modélisation selon les caractéristiques des descripteurs utilisés.

Dans le chapitre suivant (chapitre 5), le modèle GRD et les fonctionnalités de l'algorithme de calage sont testés dans des conditions expérimentales maîtrisées (expériences jumelles et données fictives).

Chapitre 5

Expériences préliminaires de calage et évaluation des modèles de routage

Sommaire

5.1	Expériences jumelles de calage variationnel	106
5.1.1	Description des expériences jumelles	106
5.1.2	Résultats de l'expérience jumelle 1	108
5.1.3	Résultats de l'expérience jumelle 2 :	109
5.1.4	Résultats de l'expérience 3 :	110
5.1.5	Conclusion	112
5.2	Comparaison des modèles de routage avec un modèle hydraulique de barré de saint-venant (SIC^2)	112
5.2.1	Création des jeux d'observation pour le calage et la validation à l'aide de SIC^2	112
5.2.2	Calage à l'aide des variantes de GRD	113
5.2.3	Validation temporelle des paramètres calibrés	114
5.2.4	Résultats	115
5.3	Tests de la fonction de régularisation Tikhonov : la méthode de la L-curve et la méthode de régularisation itérative	119
5.3.1	Objectifs	119
5.3.2	Expériences et résultats	120
	Expérience 1 : Détermination du poids associé à la régularisation par la méthode de la L-curve	121
	Expérience 2 : Comparaison des solutions obtenues avec les méthodes de régularisation L-curve et itérative	122
5.3.3	Conclusions	125
5.4	Conclusions	125

Dans ce chapitre, des expériences numériques sont réalisées afin de tester le modèle hydrologique distribué (présenté au chapitre 3) et l'algorithme d'estimation variationnelle (présenté au chapitre 4).

Dans la première partie nous proposons de tester l'algorithme d'estimation variationnelle pour le calage des paramètres du modèle hydrologique distribué. Pour cela des expériences jumelles sont réalisées. Pour ces expériences les données d'entrées et de sorties du modèle sont connues et artificielles.

Dans la seconde partie, des expériences sont menées afin de tester et comparer les trois modèles de routage conceptuels proposés au chapitre 3. Ces tests sont réalisés dans des conditions expérimentales particulières : les données d'entrées des modèles de routage sont générées artificiellement ; les sorties des modèles de routage sont comparées à celles d'un modèle hydraulique physique (SIC) comme dans le test présenté au chapitre 3 section 3.3.

Dans la troisième partie, les méthodes de régularisation (méthode de la L-curve et méthode de régularisation itérative) décrites aux sections 4.2.4 et 4.2.5 sont testées et comparées. La méthode empirique proposée (chapitre 4, section 4.5.2) est vérifiée. Cette méthode permet déterminer automatiquement le coefficient de régularisation pour la méthode de la L-curve.

5.1 Expériences jumelles de calage variationnel

5.1.1 Description des expériences jumelles

Les objectifs de ces expériences sont de tester le fonctionnement de l'algorithme de calage variationnel et d'évaluer son aptitude à estimer les paramètres distribués du modèle hydrologique.

Une expérience jumelle consiste à résoudre un problème dont les entrées et sorties d'un modèle sont connues et maîtrisées. Soit X le vecteur de contrôle contenant les paramètres du modèle. Ces paramètres sont inconnus et doivent être identifiés. Une expérience jumelle est construite de la manière suivante :

1. Choix d'un jeu de paramètres en tous points du domaine. Ces paramètres sont supposés "vrais" et sont notés \bar{X}
2. Fabrication du vecteur d'observation Y correspondant à \bar{X} tel que $Y = H(\bar{X})$ où H représente le modèle.
3. Choix d'un vecteur d'ébauche X_b différent de l'état vrai \bar{X}
4. Recherche du vecteur analysé X_a qui minimise l'écart entre les sorties du modèle $H(X_a)$ et le vecteur d'observation Y . Cette recherche est effectuée à l'aide de l'algorithme variationnel présenté au chapitre 4.
5. Comparaison des différences entre les vecteurs X_a et \bar{X} .

Les expériences jumelles sont réalisées sur un maillage unidirectionnel constitué d'une série de 10 pixels. Deux expériences impliquant chacune les 3 variantes du modèle GRD (voir section 3.2) sont réalisées. Pour chaque variante de modélisation, trois paramètres par pixel sont recherchés : les capacités des réservoirs de production c_p ; les capacités des réservoirs de transfert c_{tr} ; les paramètres du modèle de routage (les vitesses v , les paramètres hydrauliques ξ ou les capacités du réservoir linéaire de routage c_r selon la variante utilisée). L'algorithme d'estimation variationnelle présenté au chapitre 4 est utilisé pour caler les paramètres du modèle. L'algorithme utilise la fonction objectif quadratique présentée à la section 4.3.1 et la méthode de régularisation itérative présentée à la section 4.2.5.

La durée de la simulation est de 20 heures et le pas de temps est fixé à 15 minutes. En entrée du modèle, une pluie artificielle affectant l'ensemble du domaine (maillage) est construite. Cette pluie artificielle dure deux heures et démarre à la cinquième heure de simulation. Son intensité, comprise entre 0 et 28 millimètres par heure, est non-uniforme spatialement et temporellement (variation sinusoidale). Les débits simulés supposés "vrais" (vecteur d'observation Y) sont construits à partir d'un vecteur pré-déterminé des paramètres du modèle, noté \bar{X} . Ces paramètres sont également non-uniformes spatialement et évoluent de façon sinusoidale du pixel amont au pixel aval du domaine. Pour le calage des paramètres le vecteur d'observation Y est constitué des débits supposés "vrais" émis depuis le pixel aval (le pixel aval du maillage unidirectionnel). La figure 5.1 schématise le fonctionnement des expériences jumelles.

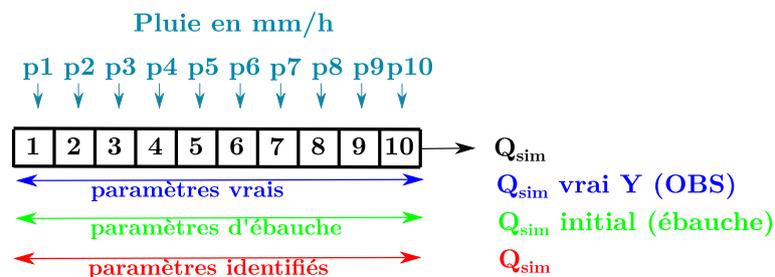


FIGURE 5.1 – Schéma représentant le maillage utilisé pour les deux expériences jumelles. Le maillage est constitué de dix pixels numérotés de 1 à 10, de l'amont vers l'aval. Les débits simulés Q_{sim} (output) sont émis depuis le pixel aval numéro 10. Les pluies p_1, p_2, \dots, p_{10} (input) affectent l'ensemble des pixels. Les trois paramètres du modèle sont définis pour chaque pixel.

Les trois expériences jumelles diffèrent par le choix du vecteur de paramètre initial (le vecteur d'ébauche X_b) et le choix du paramètre de corrélation spatiale de la fonction de régularisation (paramètre ω de l'équation 4.4). Pour la première et troisième expérience, le vecteur d'ébauche X_b est constitué des valeurs moyennes et uniformes des paramètres sur le domaine tel que $X_b^i = \langle \bar{X} \rangle$ pour chaque pixel i du maillage. Pour la seconde expérience, le vecteur d'ébauche X_b est non-uniforme sur le domaine. Il est défini tel que

$X_b^i = \frac{1}{2}(\bar{X}^i + \langle \bar{X} \rangle)$ pour chaque pixel i du maillage. Pour la première et deuxième expérience, le coefficient de corrélation spatiale ω est faible $\omega = 0.1$ (les paramètres seront peu corrélés spatialement). Pour la troisième expérience, ce coefficient de corrélation est augmenté : $\omega = 1$ (les paramètres seront davantage corrélés spatialement). Les caractéristiques de ces trois expériences sont résumées dans le tableau 5.1.

En résumé, les expériences 1 et 2 permettent de mesurer la sensibilité du vecteur d'ébauche sur les paramètres optimaux recherchés. L'expérience 3 permet de montrer l'impact de la fonction de régularisation type Tikhonov sur la solution trouvée.

TABLE 5.1 – Synthèse des caractéristiques des trois expériences jumelles.

	$x_b = \langle \bar{x} \rangle$	$x_b^i = \frac{1}{2}(\bar{x}^i + \langle \bar{x} \rangle)$
$\omega = 0.1$	Expérience 1	Expérience 2
$\omega = 1$	Expérience 3	

Les résultats de ces expériences sont présentés aux sections 5.1.2, 5.1.3 et 5.1.4. Pour chaque expérience, les résultats sont présentés sur une planche de six graphiques. Chaque planche permet de visualiser :

- La répartition spatiale et temporelle de la pluie.
- Les variabilités spatiales des paramètres connus \bar{X} , initialisés X_b et trouvés X_a .
- Les débits "vrais" Y et les débits simulés par le modèle initialement $H(X_b)$ et après calage des paramètres $H(X_a)$.
- L'évolution au cours des itérations du coût associé à la fonction objectif.

5.1.2 Résultats de l'expérience jumelle 1

Les résultats sont présentés sur la figure 5.2.

Pour l'expérience numéro 1, le vecteur de contrôle initial (ébauche) est composé des valeurs de paramètres uniformes et moyennes sur le maillage. Après calage des paramètres à l'aide des débits "vrais" connus au pixel aval, une variabilité spatiale des paramètres est obtenue. Celle-ci ne correspond pas à la variabilité des paramètres connus (\bar{X}). Les capacités des réservoirs de production sont les paramètres les mieux identifiés. En effet leurs variabilités spatiales semblent se rapprocher davantage des paramètres "vrais". Remarquons que la variabilité du paramètre hydraulique ξ de la variante 2 (graphique 5.2, colonne centrale) semble être elle aussi partiellement retrouvée.

La procédure d'optimisation variationnelle fonctionne correctement. La fonction objectif est minimisée. Les débits "vrais" et simulés sont quasiment identiques après calage.

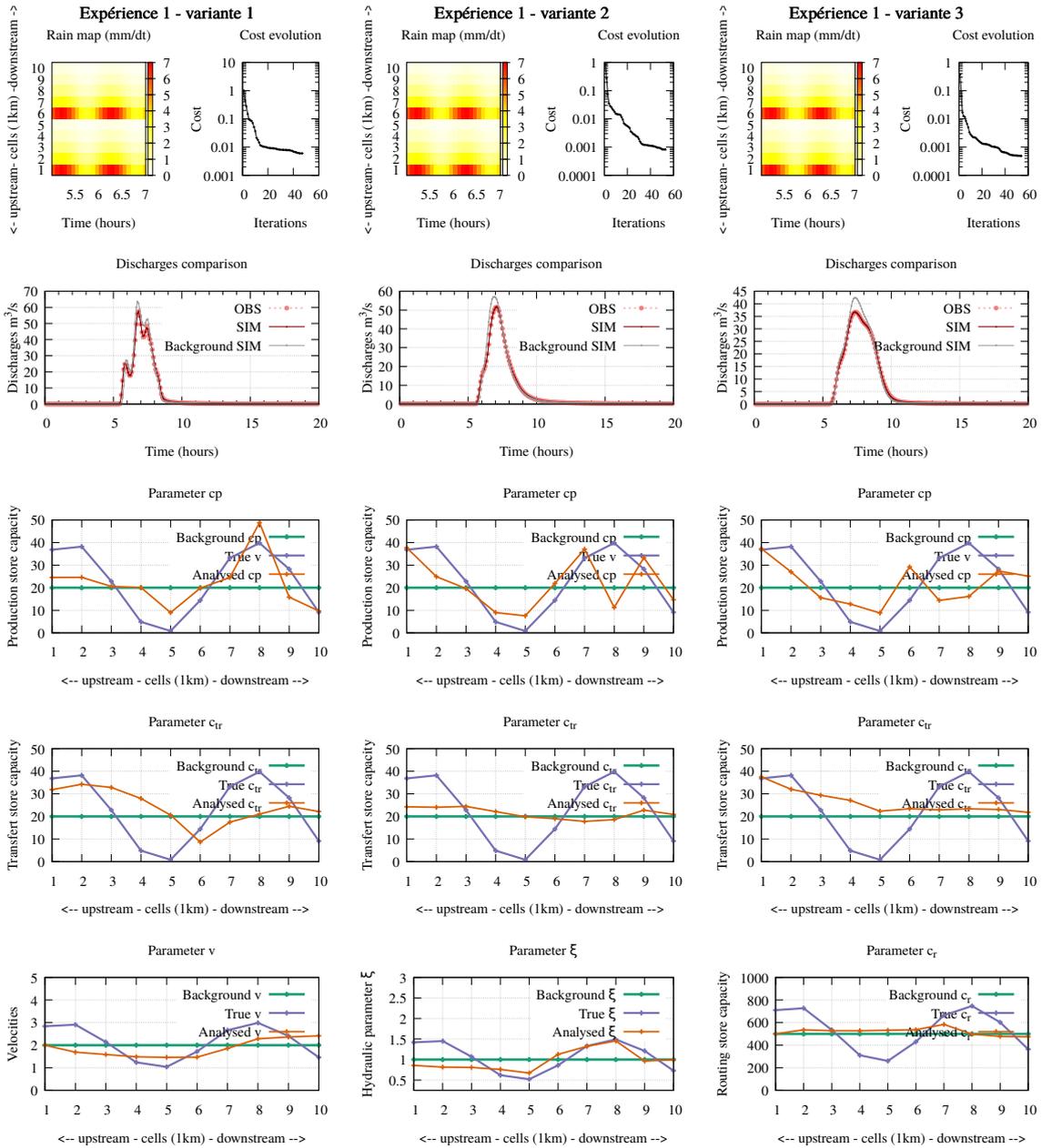


FIGURE 5.2 – Résultats de l'expérience 1 pour les variantes 1 (colonne de gauche), 2 (colonne centrale) et 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les paramètres de routage (v, ξ, c_r) sont recherchées.

5.1.3 Résultats de l'expérience jumelle 2 :

Les résultats sont présentés sur la figure 5.3.

Pour l'expérience numéro 2, le vecteur de contrôle initial (ébauche) est non-uniforme et se rapproche du vecteur contenant les "vrais" paramètres. Après calage, les nouveaux paramètres identifiés restent proches des paramètres d'ébauche mais reflète mieux la varia-

bilité des "vrais" paramètres. Cependant, l'identification des paramètres reste imprécise.

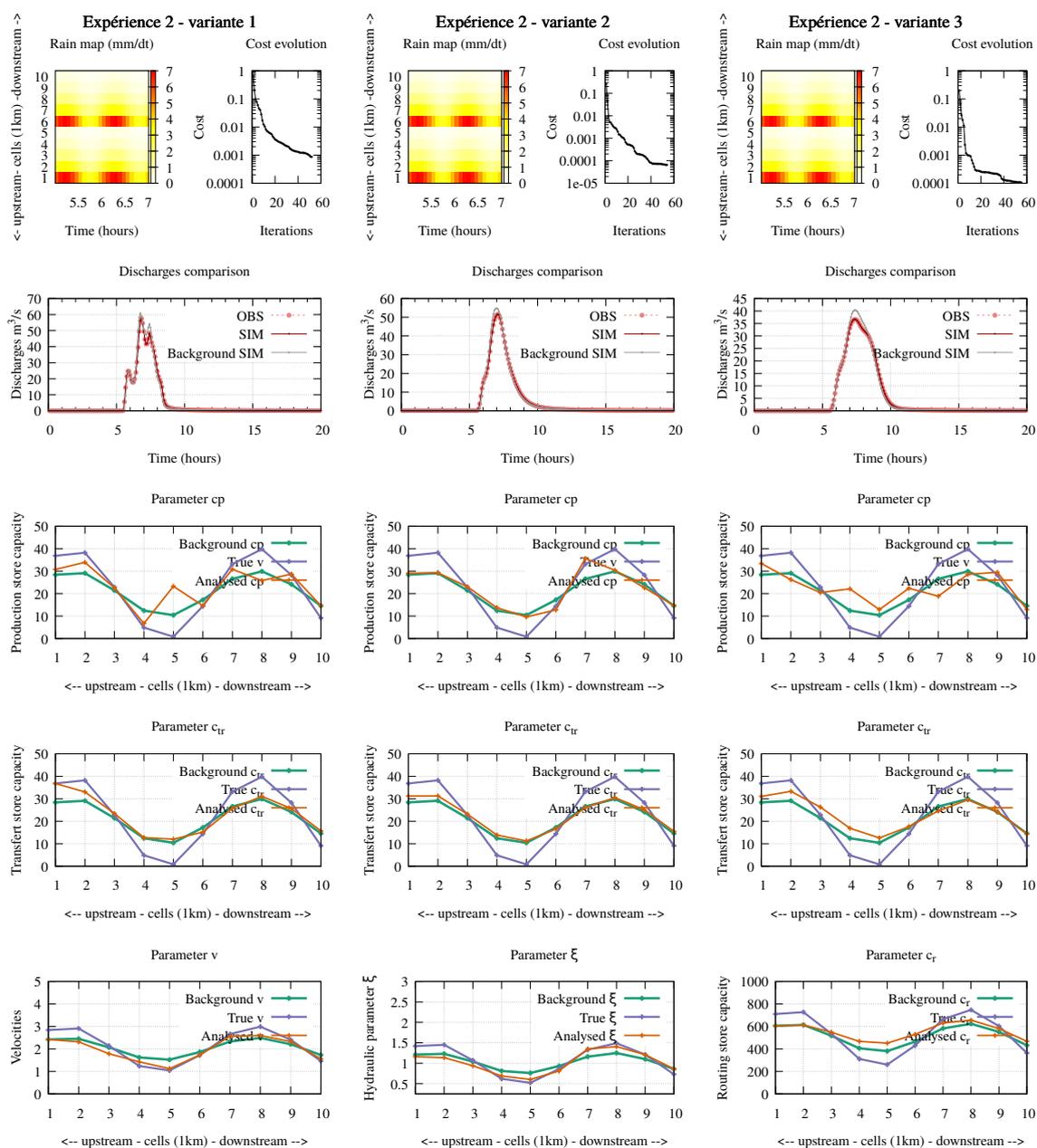


FIGURE 5.3 – Résultats de l'expérience 2 pour les variantes 1 (colonne de gauche), 2 (colonne centrale), 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les vitesses de routage v sont recherchées.

5.1.4 Résultats de l'expérience 3 :

Les résultats sont présentés sur la figure 5.4.

Pour l'expérience numéro 3, le vecteur de contrôle initial (ébauche) est uniforme. Cette expérience diffère de l'expérience numéro 1 par le choix du coefficient de corrélation spa-

tiale. La recherche des paramètres est davantage contrainte par le terme de régularisation J_{reg} de la fonction objectif. Cette contrainte permet d'améliorer sensiblement l'identification des paramètres (surtout pour la variante 2). Notons que cette fonction de régularisation améliore l'identification d'un signal sinusoïdale. Dans la réalité, les variables à identifier n'évolueront pas forcément de façon graduelle dans l'espace.

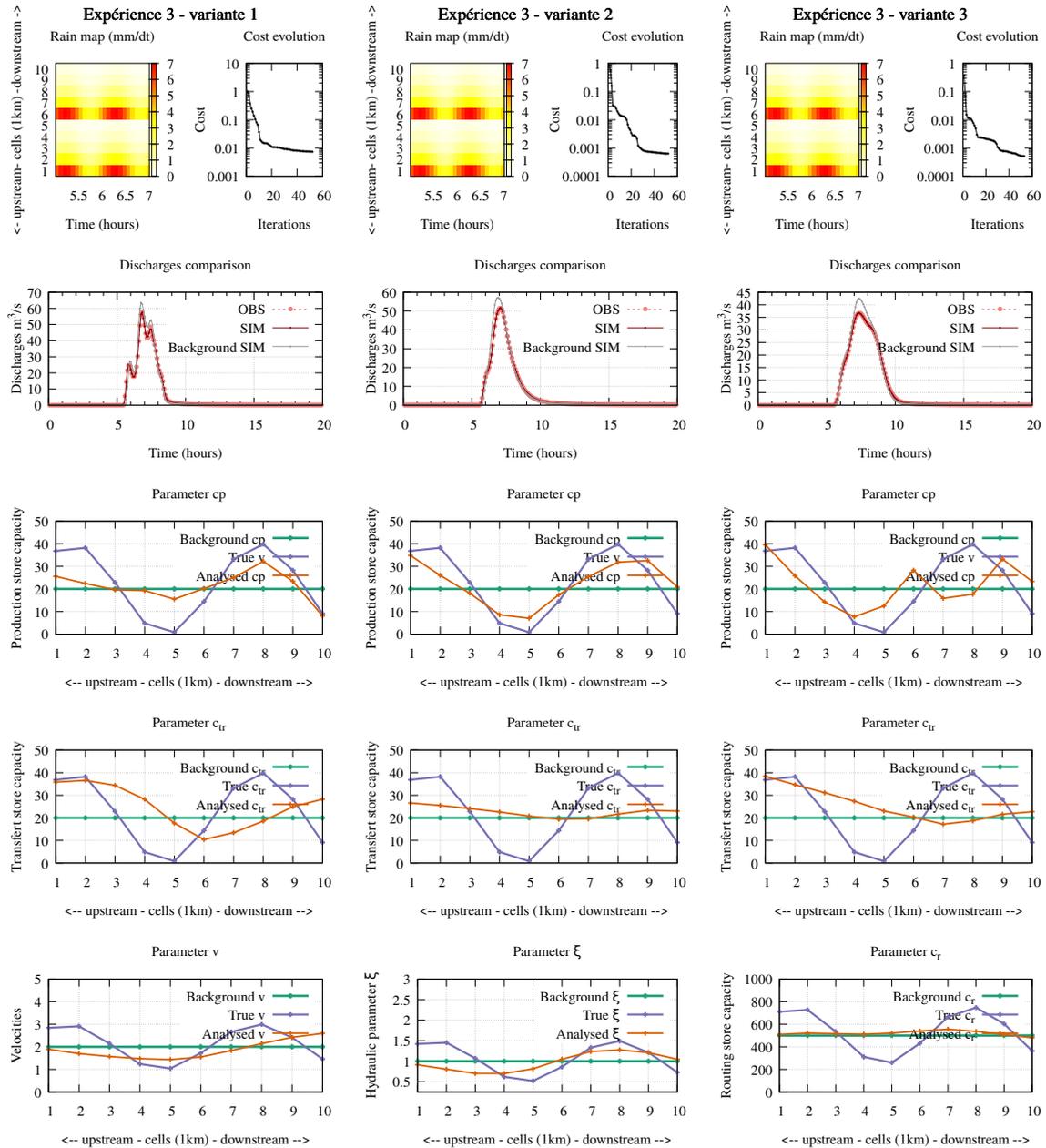


FIGURE 5.4 – Résultats de l'expérience 3 pour les variantes 1 (colonne de gauche), 2 (colonne centrale), 3 (colonne de droite). Les capacités des réservoirs de production c_p , les capacités des réservoirs de transfert c_{tr} et les vitesses de routage v sont recherchées.

5.1.5 Conclusion

L'algorithme de calage variationnel fonctionne correctement et permet de minimiser la fonction objectif. Les trois paramètres des trois variantes du modèle GRD sont identifiés pour chacune des mailles du maillage (3×10 paramètres). Les solutions trouvées correspondent à des minima de la fonction objectif. Cependant, ces minima correspondent à des minima "locaux". En effet il existe plusieurs combinaisons de paramètres permettant de réduire le coût associé à la fonction objectif. Ces solutions ne correspondent pas nécessairement aux "vrais" jeux de paramètres comme le montrent ces trois expériences jumelles. Le choix du vecteur d'ébauche permet de contraindre le problème et d'améliorer l'identification des paramètres. Le terme de régularisation ajoute une contrainte supplémentaire (corrélation spatiale des paramètres) et permet lui aussi d'améliorer l'estimation spatiale des paramètres. Ce dernier résultat est sans doute vrai dans la mesure où la contrainte spatiale ajoutée (régularisation de type Tikhonov dont la propriété est de lisser les paramètres spatialement) correspond à la forme de la solution recherchée (signal sinusoïdale).

5.2 Comparaison des modèles de routage avec un modèle hydraulique de barré de saint-venant (SIC^2)

Ces expériences ont pour but de comparer et d'évaluer les trois modèles de routage proposés au chapitre 3. Comme pour les expériences jumelles réalisées précédemment (section 5.1), ces expériences sont menées sur un maillage unidimensionnel. Les variantes de routage du modèle GRD (variante 1, 2 et 3) sont comparées avec le modèle hydraulique physique SIC^2 (Dorchies et al., 2013). Ces expériences ne sont pas des expériences jumelles puisque les données de débits sont générées avec le modèle SIC^2 . Ces données de débits servent de référence et sont considérés comme "vrais".

Les variantes de routage du modèle GRD (variante 1, 2 et 3) vont alors être utilisées pour modéliser les crues générées par le modèle hydraulique SIC^2 . L'objectif est de comparer les modèles de routage conceptuels utilisés dans GRD avec un modèle de routage hydraulique à base physique. Dans ce cas précis, les fonctions de production et de transfert du modèle hydrologique sont écartées de la modélisation.

5.2.1 Création des jeux d'observation pour le calage et la validation à l'aide de SIC^2

SIC^2 (Simulation Intégrée des Canaux et de leur Contrôle) est un logiciel de simulation hydraulique développé par IRSTEA (Dorchies et al., 2013). À l'aide de SIC^2 , nous simulons la propagation de crues dans un tronçon de rivière artificielle. La longueur du tronçon de rivière fait 20 km. Ce tronçon est constitué d'une succession de 20 biefs de 1

km de longueur, de géométrie identique (canal rectangulaire de largeur 20 mètres) et de pente constante (0.05%)(voir figure 5.6).

À l'amont de chaque bief, des débits sont générés de manière aléatoire et sont injectés à chaque pas de temps. Ces chroniques de débits sont créés par un générateur aléatoire à partir d'une loi "log-normale" tout en respectant une corrélation spatiale et temporelle. Le mode μ de la loi "log-Normal" permet d'ajuster l'amplitude médiane du signal généré. Un débit de base constant égale à $1 \text{ m}^3/\text{s}$ est injecté à l'amont du réseau.

Le pas de temps des simulations est fixé à 15 minutes. Sous ces conditions, la propagation des crues est simulée dans le réseau sur une première période de 40 jours (soit 960 heures)(voir figure 5.5). Les hydrogrammes obtenus au nœud aval (bief numéro 20, nœud numéro 21) seront utilisés pour caler les paramètres de routage du modèle hydrologique GRD.

Sur une période de 10 jours qui suit la première période de 40 jours (soit 240 heures), 100 nouveaux hydrogrammes sont générés aléatoirement sur le même principe (voir figure 5.5). Ces hydrogrammes seront utilisés pour effectuer des validations temporelles du calage du modèle GRD.

Trois amplitudes de débits sont testées : des amplitudes "medium", "low" et "high" respectivement de mode μ , $\mu/2$, $2 \times \mu$ sont caractérisées par des débits d'amplitudes variables (amplitude moyenne, amplitude deux fois plus petite et amplitude deux fois plus grande).

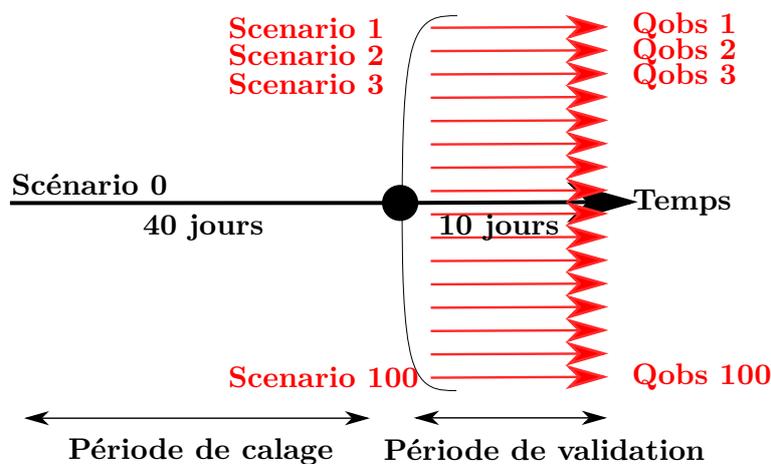


FIGURE 5.5 – Génération des scénarios de débits pour les périodes de calage et de validation pour une amplitude fixée.

5.2.2 Calage à l'aide des variantes de GRD

Le modèle hydrologique est constitué de 21 pixels successifs de 1 km^2 . Le routage est effectué entre les centres de chaque pixel, soit 20 tronçons de 1 km , représentant chacun un bief construit dans SIC^2 (voir figure 5.6). Le pas de temps de simulation est fixé à 15 minutes. Les débits qui ont été générés de manière aléatoire dans SIC^2 à l'entrée de

chaque bief sont convertis en lame d'eau précipitée sur le pixel correspondant. Le débit de base ($1 \text{ m}^3/\text{s}$) est assuré en injectant un flux constant à l'entrée du pixel amont.

Les paramètres initiaux (ébauche) du modèle de routage hydrologique sont fixés à des valeurs uniformes sur l'ensemble du maillage. Elles ont été déterminées arbitrairement, connaissant l'ordre de grandeur des intensités pluvieuses et des débits à simuler. Les vitesses initiales v sont égales à 1 m/s (variante 1), le coefficient "hydraulique" ξ est égale à 0.5 (variante 2) et les capacités initiales des réservoirs de routage c_r sont égales à 100 s^{-1} (variante 3). Les réservoirs de production et de transfert sont supprimés de la modélisation afin d'étudier uniquement le schéma de routage. Le coefficient d'étalement de la loi Gamma a été évalué à 0.5 . Ce coefficient est uniforme sur le maillage et son estimation provient de la comparaison entre la forme de la fonction de distribution Gamma avec un hydrogramme simulé par *SIC*² dans un canal rectangulaire.

Pour les variantes de GRD, un calage spatial des paramètres de routage du modèle (la vitesse pour la variante 1, le coefficient "hydraulique" ξ pour la variante 2 et la capacité du réservoir de routage c_r pour la variante 3) est réalisé à l'aide des hydrogrammes générés par *SIC*² aux biefs n°20 (1 point de contrôle) et sur une période de 40 jours (appelée fenêtre d'assimilation). Ce calage est réalisé pour chacune des trois amplitudes de débits (medium, low, high). Les critères de Nash-Sutcliffe (critère NSE présenté à l'équation 4.27) sont calculés entre les hydrogrammes simulés par GRD et ceux simulés par *SIC*².

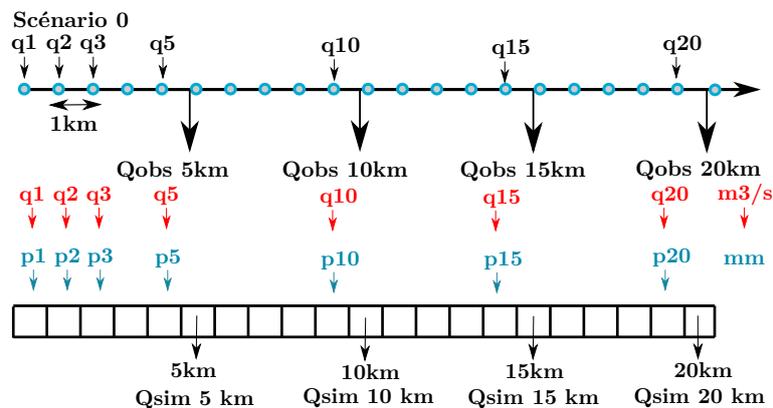


FIGURE 5.6 – Interfaçage des scénarios de débits entrant, des hydrogrammes simulés par *SIC*² (*Qobs*) et des hydrogrammes simulés par GRD (*Qsim*). Le maillage de *SIC*² est constitué d'une succession de 20 biefs de 1km (ligne noir) entrecoupés de 21 noeuds (rond bleu). Le maillage GRD est composé de 21 cellules.

5.2.3 Validation temporelle des paramètres calibrés

Pour chaque variante, le jeu de paramètres calibrés est utilisé pour modéliser les 100 hydrogrammes qui ont été préalablement générés pour la validation. Le modèle est d'abord lancé sur la première période de 40 jours et les états finaux sont sauvegardés. Le modèle est ensuite relancé 100 fois pour simuler la propagation des 100 scénarios générés

pour la validation. Les expériences sont les suivantes :

- 1) calage sur un scénario d'amplitude "medium" : validation sur les scénarios d'amplitudes "medium", "high" et "low".
- 2) calage sur un scénario d'amplitude "low" : validation sur les scénarios d'amplitudes "medium", "high" et "low".
- 3) calage sur un scénario d'amplitude "high" : validation sur les scénarios d'amplitudes "medium", "high" et "low".

5.2.4 Résultats

En calage : les graphiques 5.7, 5.8 et 5.9 représentent les débits observés (SIC) et simulés (GRD) après calage sur un scénario de débits amont d'amplitude "medium". Chaque variante (1, 2 et 3) permet de très bien modéliser la propagation des débits après calage (figures 5.7, 5.8 et 5.9). Néanmoins, les débits avalés simulés de la variante 1 sont davantage "bruités" et les pointes de crues sont légèrement sur-estimées. En effet les phénomènes d'amortissement/diffusion sont très négligés.

En validation : Pour chaque expérience 1), 2), 3), les critères de Nash-Sutcliffe (NSE) sont calculés entre les hydrogrammes simulés (GRD) et observés (SIC^2) à l'aval des biefs numéros 5, 10, 15 et 20 (pixels 6, 11, 16 et 21 dans GRD) pour chaque scénario de validation. Les résultats sont analysés en calculant des statistiques (médiane, quantile 90 et écart-type) sur les 100 valeurs de ce critère. Les graphiques 5.10, 5.11 et 5.12 représentent ces statistiques sous la forme de "box-plot". Les performances de la variante 2 dépassent ou égalisent celles de la variante 3. Ces écarts sont davantage perceptibles lorsque les calages et validations sont opérés sur des scénarios d'amplitudes antagonistes. Les performances de la variante 1 sont nettement inférieures. Ce résultat montre clairement la déficience du modèle de routage utilisé dans la variante 1. En revanche la cascade de réservoirs linéaires, utilisé dans la variante 3, semble suffisante pour modéliser la propagation des crues dans un bief. Ce modèle s'apparente d'ailleurs à la cascade de Nash ([Nash, 1957](#)).

Conclusion : L'algorithme d'estimation variationnelle a permis d'estimer les paramètres de routage pour chaque maille du réseau à partir des débits "observés" (générés par SIC^2) à l'exutoire aval du réseau. Le modèle de routage "variante 2" est bon et permet de représenter correctement les phénomènes de propagations hydrauliques des crues : en validation temporelle et spatio-temporelle, ses performances dépassent celles des variantes 1 et 3. Les performances de ce modèle restent bonnes en validation pour des scénarios de pluies d'amplitudes différentes de ceux utilisés en calage. Les performances de ce modèle sont stables le long du réseau (validation spatio-temporelle). En effet, les critères de NSE calcu-

lés à l'amont et à l'aval du réseau (cellules 6, 11, 16 et 21) sont équivalents. Au contraire, les performances des variantes 1 et 3 se dégradent en validation temporelle et spatio-temporelle pour des scénarios de pluie d'amplitudes différentes de ceux utilisés en calage. De plus, pour la variante 1, les performances varient entre l'aval et l'amont du réseau : elles deviennent meilleures à l'amont du réseau lorsque les phénomènes hydrauliques sont moins prépondérants.

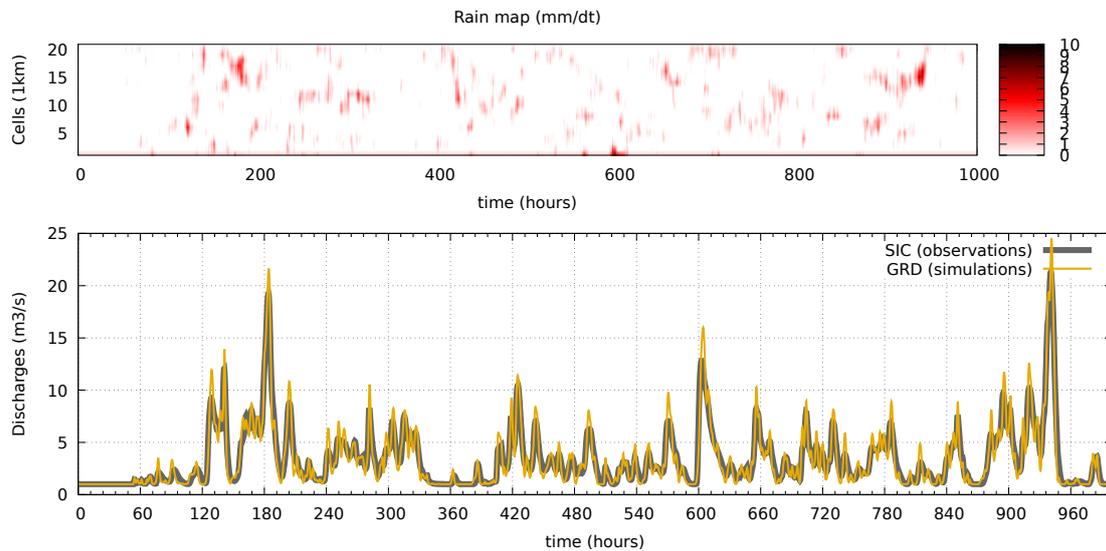


FIGURE 5.7 – Débits aval simulés par GRD et par SIC après calage des paramètres de vitesse de la variante 1 sur un scénario de débits d'amplitude "medium"

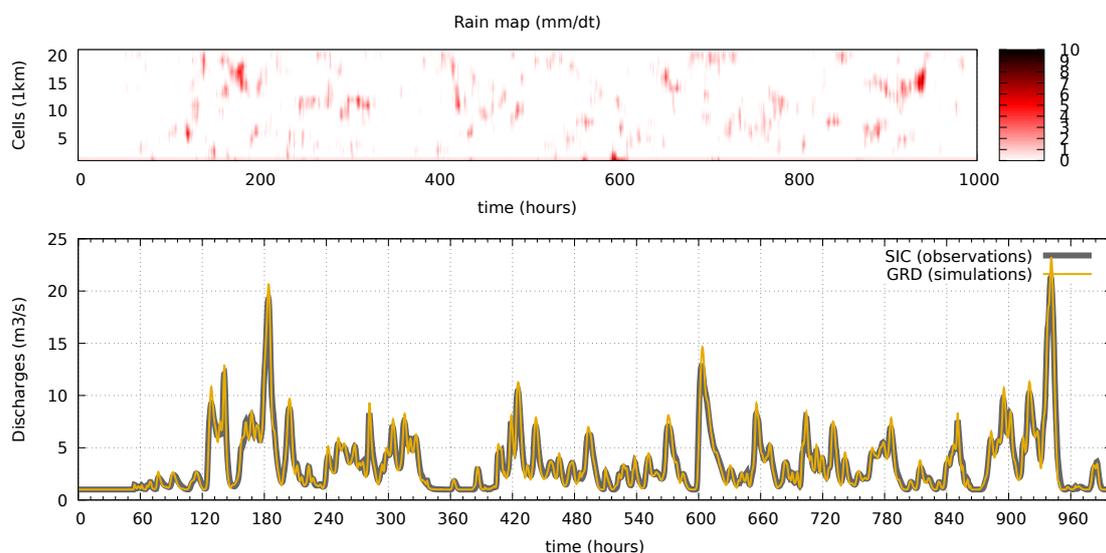


FIGURE 5.8 – Débits aval simulés par GRD et par SIC après calage des paramètres de "géométrie" de la variante 2 sur un scénario de débits d'amplitude "medium"

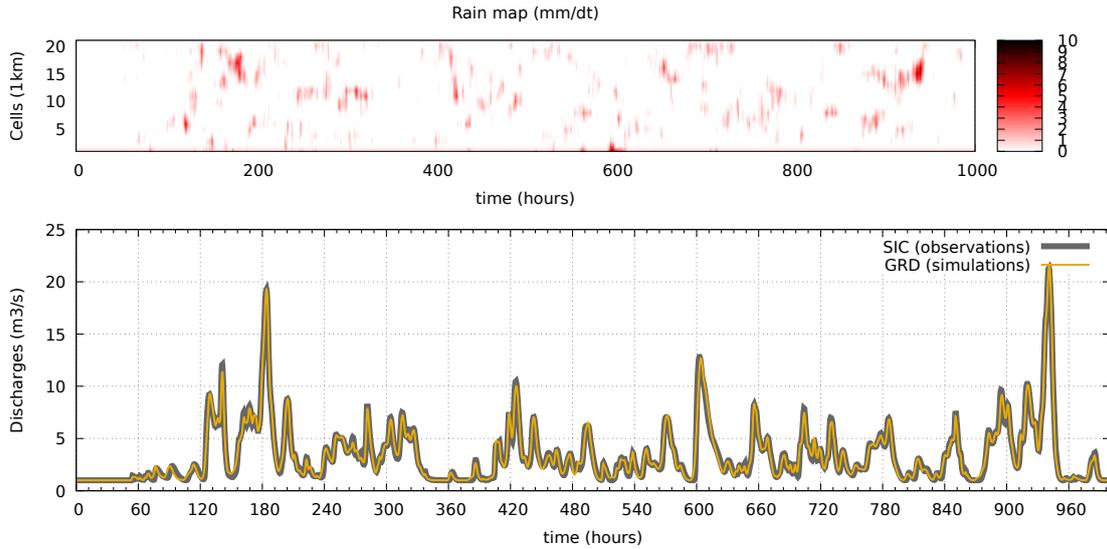


FIGURE 5.9 – Débits aval simulés par GRD et par SIC après calage des capacités des réservoirs de routage de la variante 3 sur un scénario de débits d'amplitude "medium"

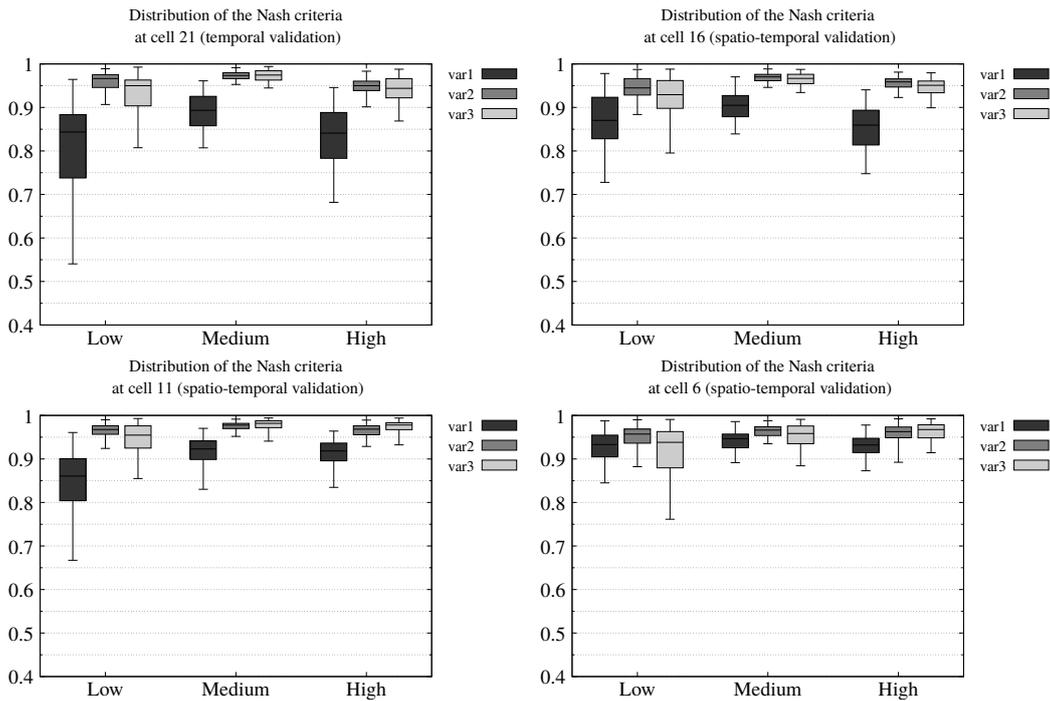


FIGURE 5.10 – Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude "medium". Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).

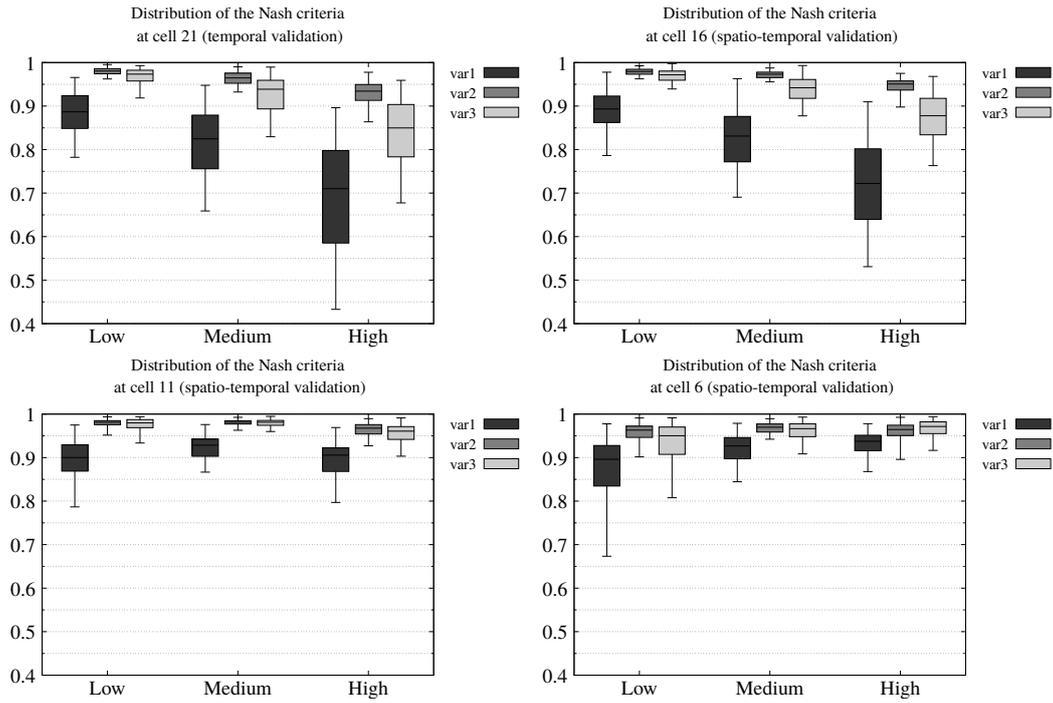


FIGURE 5.11 – Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude "low". Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).

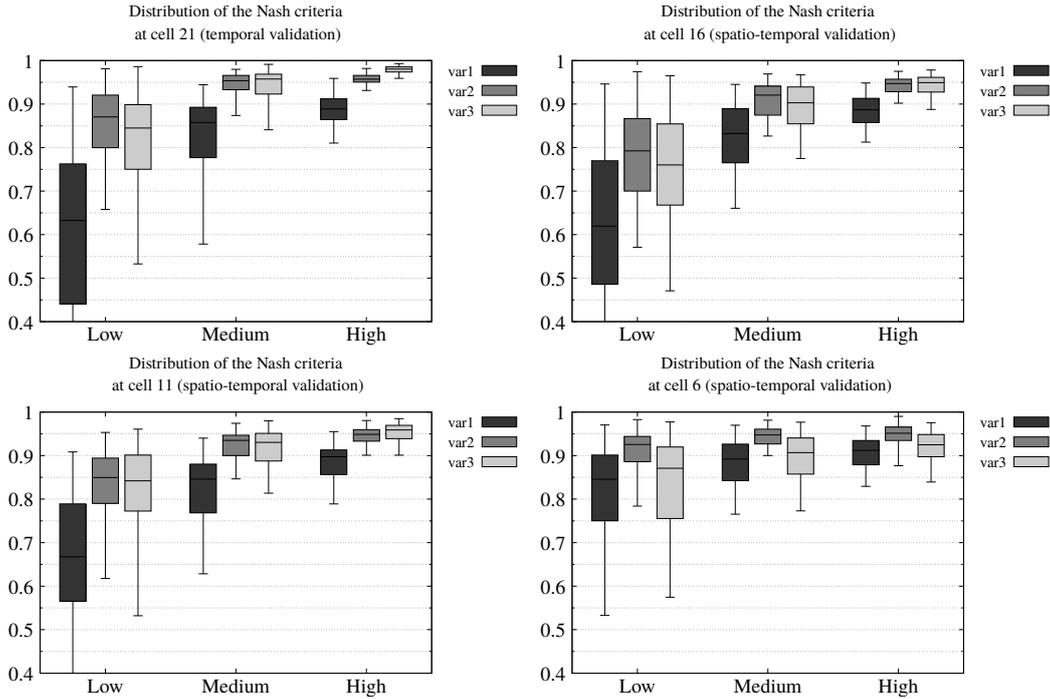


FIGURE 5.12 – Évaluation de la prédictibilité des débits par des variantes 1,2 et 3 à partir des paramètres estimés sur un scénario d'amplitude **"high"**. Distribution des performances obtenues en validation sur 100 scénarios, pour chacune des amplitudes de débits "medium", "low", "high" et à l'aval des biefs 5,10,15 et 20 (pixels 6,11,16 et 21 dans GRD).

5.3 Tests de la fonction de régularisation Tikhonov : la méthode de la L-curve et la méthode de régularisation itérative

5.3.1 Objectifs

Dans cette partie, des expériences de calage sont réalisées. Ces expériences permettent de tester les techniques de régularisation présentées au chapitre 4. Le terme de régularisation associé à la fonction objectif (section 4.4) est utilisé pour contraindre le calage des paramètres du modèle. Les deux méthodes, celles de la L-Curve (section 4.2.4) et de la régularisation itérative (section 4.2.5) sont testées pour paramétrer la régularisation.

En estimation variationnelle, la régularisation permet de contraindre la solution recherchée. Elle est apportée par une fonction, notée J_{reg} , qui est ajoutée à la fonction objectif J . La fonction J_{reg} dépend de paramètres. Un paramètre α permet d'estimer la déviation (variance) par rapport à l'ébauche ; un paramètre ω permet d'estimer la corrélation spatiale des paramètres dans l'espace (lissage) ; un paramètre β permet d'ajuster l'ordre de grandeur du terme J_{reg} par rapport J_0 . α et ω sont évalués empiriquement. La fonction de régularisation peut être introduite de deux manières différentes :

1. La régularisation de type Tikhonov avec construction de la L-curve : la fonction objectif J présentée à l'équation 4.3 est minimisée. Le terme de régularisation devient "actif" lorsque les termes J_0 et J_{reg} deviennent comparables. Le coefficient β doit être choisi précisément pour rendre la fonction objectif J sensible aux termes J_0 et J_{reg} . Le coefficient β est évalué à l'aide d'une courbe L-Curve (voir section 4.2.4). Remarquons que si l'ébauche est uniforme la fonction objectif J devient sensible à J_{reg} seulement après plusieurs itérations (lorsque le vecteur de contrôle aura suffisamment dévié de l'ébauche). Lors des premières itérations la fonction objectif n'est pas contrainte (Gejadze and Malaterre, 2017).
2. La régularisation de type Tikhonov itérative : un changement de variable est introduit et le numéro de l'itération sert de paramètre de régularisation. Le "degré" de régularisation est conditionné par le nombre d'itérations réalisés. Cette méthode est décrite à la section 4.2.5. Notons qu'avec cette méthode, le problème est contraint dès la première itération.

L'élaboration de la L-curve est très coûteuse en ressource de calcul. C'est pour cela qu'une méthode empirique est envisagée pour déterminer plus efficacement ce paramètre. Cette méthode a été présentée à la section 4.5.2. Elle est testée dans cette partie lors de l'expérience 1.

La méthode itérative est plus efficace puisque le "degré" de régularisation correspond au numéro de l'itération. Celui-ci est évalué au fur et à mesure des itérations. Cette méthode est cependant plus complexe à mettre en œuvre et présente quelques limitations (par exemple, l'implémentation de cette méthode ne permet pas d'utiliser une ébauche et des bornes non-uniformes spatialement). Les solutions trouvées pour la méthode de la L-curve et pour la méthode itérative sont comparées dans l'expérience 2.

5.3.2 Expériences et résultats

Deux expériences de calage des paramètres du modèle GRD sont réalisées à partir du jeu de données présenté au chapitre 2 (section 2.2). Le modèle GRD présente les caractéristiques suivantes : modèle de routage "variante 2" (section 3.2.4), maillage carré de résolution 1 km^2 , pas de temps de simulation horaire. Trois paramètres par maille sont recherchés pour chaque bassin-versant (les capacités des réservoirs de production, les capacités des réservoirs de transfert et le paramètre hydraulique de routage). La fonction objectif est calculée entre les débits observés et simulés aux stations de jaugeage situées à l'aval des bassins-versants (stations sélectionnées pour le calage à la section 2.2.2). Le calage est réalisé sur une période temporelle d'un an (du 01/01/2011 au 01/01/2012) succédant à une période de "chauffe" d'un an (du 01/01/2010 au 01/01/2011) du modèle. L'objectif de ces tests n'est pas d'évaluer le calage du modèle, mais de tester les méthodes de régularisation. C'est pourquoi une période de calage "courte" (1 an) est choisie. Deux expériences sont réalisées.

Expérience 1 : Détermination du poids associé à la régularisation par la méthode de la L-curve

Pour ces tests la fonction de régularisation J_{reg} présentée à la section 4.4.1 est utilisée. Les paramètres de simulation sont présentés au tableau suivant :

TABLE 5.2 – Paramètres de calage choisis pour l’expérience 1

Paramètres	Valeurs choisies
Coefficients de déviation $\alpha_{cp}, \alpha_{ct}, \alpha_{\xi}$	$(\alpha_{cp} = 1, \alpha_{ct} = 1, \alpha_{\xi} = 1)$
Coefficients de corrélation $(\omega_{cp}, \omega_{ct}, \omega_{\xi})$	$(\omega_{cp} = 10, \omega_{ct} = 10, \omega_{\xi} = 5)$
Vecteur d’ébauche $x_b = (\bar{c}_p, \bar{c}_{tr}, \bar{\xi})$	$x_b = (200, 100, 1)$
Coefficients de régularisation β	$\beta = [10^{-7} : 10^{-1}]$
Borne min x_{min}	$x_{min} = \frac{1}{2}x_b$
Borne max x_{max}	$x_{max} = 2x_b$

Deux méthodes sont mises en œuvre pour déterminer le coefficient de régularisation optimal :

1. Les calages sont réalisés pour une large gamme du coefficient de régularisation β . La courbe L-curve est tracée pour chaque couple de points (J_0, J_{reg}) .
2. Les calages sont réalisés avec un coefficient de régularisation calculé à l’aide de la méthode automatique empirique présenté à la section 4.5.2. A la fin de la procédure de minimisation, le point (J_0, J_{reg}) correspondant est ajouté sur le graphique de la L-curve

Les graphiques type L-curve obtenus après calage des paramètres pour chaque bassin-versant sont tracés sur la figure 5.13.

Resultats : Les graphiques type L-curve obtenus après calage des paramètres pour chaque bassin-versant sont tracés sur la figure 5.13.

Le graphique en L-curve permet d’estimer le coefficient de régularisation β . En effet, les courbes présentent un angle marqué lorsque le critère J_{reg} augmente fortement alors que le critère J_0 n’est pas davantage réduit.

La méthode automatique (présenté à la section 4.5.2) permet d’estimer efficacement le coefficient de régularisation β en deux cycles d’optimisation (au lieu des sept cycles ayant été réalisés pour construire les courbes L-curve). Les points (J_0, J_{reg}) correspondant aux coefficients β optimaux se situent aux points de courbure maximale des courbes L-curve.

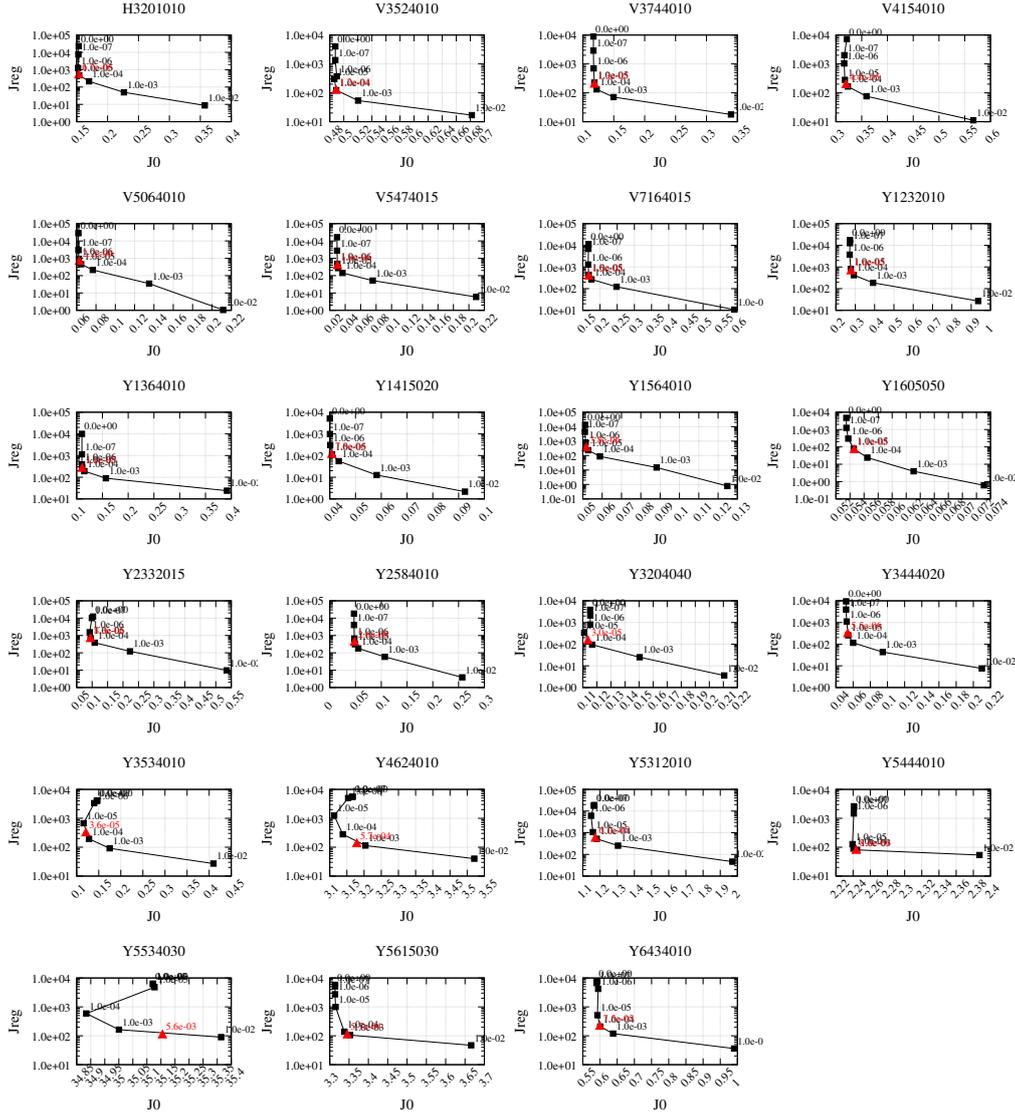


FIGURE 5.13 – Expérience empirique de calage permettant de tester la méthode d’auto-détermination du coefficient de régularisation (méthode automatique présentée à la section 4.5.2). Le choix du coefficient β est réalisé par la L-curve (courbe noir) et par la méthode empirique automatique (triangles rouge).

Expérience 2 : Comparaison des solutions obtenues avec les méthodes de régularisation L-curve et itérative

Deux méthodes de régularisation type Tikhonov sont testées : la méthode de la L-curve (section 4.2.4) et la régularisation itérative décrite (section 4.2.5). Ces expériences consistent à comparer les solutions trouvées par l’algorithme variationnel à l’aide de ces deux méthodes dans les mêmes conditions expérimentales :

- **Test 1** : utilisation de la fonction de régularisation 4.4.1, le degré de régularisation optimale est choisie à l’aide de la L-curve par la méthode proposée à la section

4.5.2

- **Test 2** : utilisation de la fonction de régularisation itérative décrite à la section 4.2.5

Les paramétrages des calages sont définis dans le tableau 5.3 :

TABLE 5.3 – Paramètres de calage choisis pour l’expérience 3

Paramètres	Valeurs choisies
Vecteur d’ébauche $x_b = (\bar{c}_p, \bar{c}_{tr}, \bar{\xi})$	$x_b = (200, 100, 1)$
Borne min x_{min}	$x_{min} = \frac{1}{2}x_b$
Borne max x_{max}	$x_{max} = 2x_b$

Le paramètre de corrélation ω est définis empiriquement (tableau 5.4). Notons que les valeurs des paramètres de corrélation ω ne sont pas comparable entre les deux méthodes. Néanmoins, au regard de la variabilité des paramètres calés, leurs valeurs ($\omega = 10$ pour la L-curve et $\omega = 1$ pour la méthode itérative) lissent les paramètres de manière équivalente. Les paramètres de déviation α (équations 4.30 et 4.31) sont fixés à 1. Pour la méthode de la L-curve, le paramètre de régularisation β est recherché à la place de α .

TABLE 5.4

Test	α	ω
1	1	10
2	1	1

Résultats : Les capacités optimales des réservoirs de transfert, obtenues après calage et pour les deux méthodes de régularisation, sont comparées spatialement pour chaque bassin-versant. Les rapports ($C_{tr}^{test2}/C_{tr}^{test1}$) sont tracés sur la figure 5.14.

Les solutions trouvées sont sensiblement différentes. Premièrement, les différences situées à l’intérieur des bassins-versants s’expliquent par la méthode de régularisation qui, dans le cas de la méthode itérative, impose que l’optimisation soit contrainte dès la première itération. Les différences localisées à l’extérieur des bassins-versants s’expliquent par un choix des paramètres de corrélation différent entre les deux méthodes. Notons que la sensibilité des paramètres situés sur les pixels à l’extérieur des bassins-versants est uniquement liée à l’utilisation de la fonction de régularisation (corrélation spatiale).

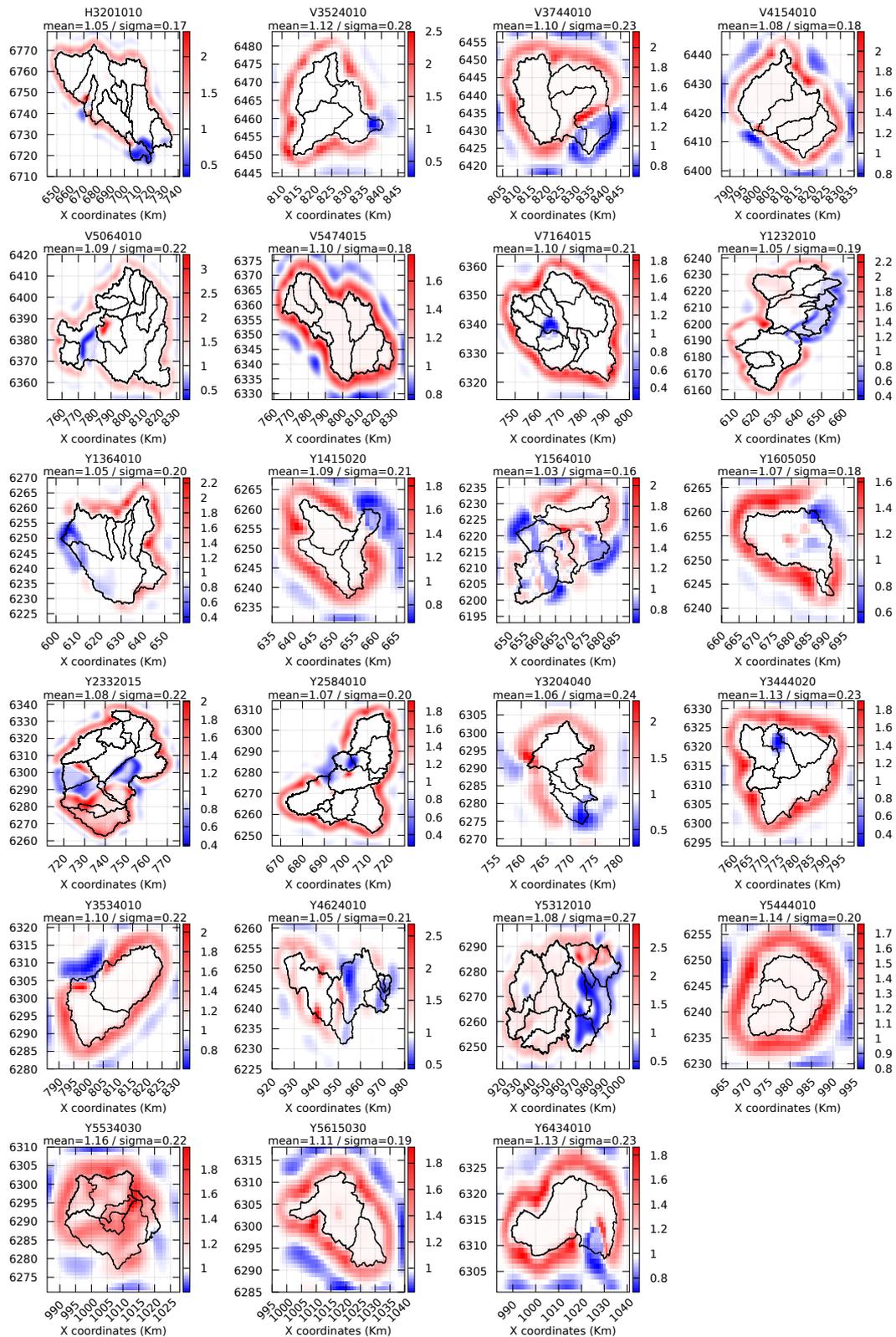


FIGURE 5.14 – Rapports ($C_{tr}^{test2} / C_{tr}^{test1}$) calculés à partir des capacités des réservoirs de transfert optimaux trouvées pour les 23 bassins-versants.

5.3.3 Conclusions

Deux façons d'utiliser la régularisation de type Tikhonov ont été testées. Ces deux méthodes mènent à des résultats sensiblement différents selon les bassins-versants pour lesquels les paramètres ont été calés. Bien que la régularisation itérative est une méthode plus robuste (Gejadze and Malaterre, 2017), le choix de la méthode dépendra de la configuration des expériences de calage :

- Les paramètres d'ébauche sont uniformes sur le maillage : la régularisation itérative sera préférée puisque cette méthode est plus efficace et permet une régularisation de la solution dès les premières itérations. De plus, des différences parfois sensibles ont été observées sur les paramètres calés.
- Les paramètres d'ébauche sont non-uniformes sur le maillage : la méthode de régularisation avec L-curve est choisie puisque l'implémentation actuelle de la méthode de régularisation itérative ne permet pas d'utiliser un vecteur d'ébauche et des bornes non-uniformes. La méthode automatique présentée à la section 4.5.2 permet de déterminer le coefficient de pondération de la régularisation avec une bonne précision. Cette méthode sera donc préférée par rapport à la construction du graphique L-curve. En effet, celle-ci permet d'obtenir une bonne estimation du coefficient de régularisation β et nécessite seulement deux cycles d'optimisation.

5.4 Conclusions

Dans ce chapitre, l'algorithme d'estimation variationnelle et les modèles de routage ont été testés et évalués.

Les expériences jumelles réalisées à la section 5.1 ont permis de s'assurer du bon fonctionnement de l'algorithme d'estimation variationnelle et du modèle adjoint. Ces expériences ont permis d'évaluer les capacités de l'algorithme à identifier un jeu de paramètres distribués. L'importance quant au choix de l'ébauche a été soulignée. La fonction de régularisation type Tikhonov, dont les propriétés imposent de lisser les paramètres spatialement, est adaptée si la forme de la solution est sinusoidale.

Les trois variantes de routage proposées au chapitre 3 (sections 3.2.3, 3.2.4 et 3.2.5) ont été testées et évaluées. D'un point de vue purement hydraulique, la variante 2 est le meilleur modèle. Cette variante permet de représenter correctement les phénomènes de propagation des crues issues d'événements de pluie de nature différente de ceux utilisés en calage.

La fonction de régularisation de type Tikhonov doit aider à identifier les paramètres distribués. Elle peut-être utilisée de deux façons différentes (la méthode itérative et la méthode de la L-curve). Ces deux méthodes ont été testées. Elles fonctionnent toutes les deux correctement mais mènent à des solutions parfois sensiblement différentes selon les bassins-versants. La méthode itérative sera préférée et utilisée lorsque les paramètres

d'ébauche sont uniformes. Au contraire, si l'ébauche est spatialement non-uniforme la méthode de la L-curve sera utilisée. Dans ce dernier cas, la méthode présentée au chapitre 4 section 4.5.2 permettra d'estimer efficacement le coefficient de régularisation.

Dans le chapitre suivant, des expériences de calage-validation sont réalisées avec de vrais données de pluies et de débits et pour le jeu de bassins-versants présentés au chapitre 2.

Chapitre 6

Expériences de calage-validation croisées

Sommaire

6.1	Objectifs et méthodologie des expériences de calage-validation . . .	128
6.1.1	Objectifs	128
6.1.2	Méthodologie	128
6.1.3	Expériences réalisées	131
6.2	Résultats	135
6.2.1	Comparaison des performances des variantes 1, 2 et 3 du modèle hydrologique	135
6.2.2	Comparaison des performances du modèle utilisant des jeux de paramètres issus des calages uniforme et distribué	138
6.2.3	Critique du calage de l'expérience B pour chaque période, comparaison à l'expérience F	140
6.2.4	Influence de l'ébauche et de la régularisation sur les performances prédictives du modèle	150
6.2.5	Analyse des jeux de paramètres distribués	152
6.2.6	Étude de la robustesse des jeux de paramètres résultant de l'expérience B pour les périodes P1 et P2	157
6.2.7	Analyse de la robustesse des paramètres issus des expériences de calage B, C, D, E et F (variante 2)	163
6.3	Conclusions	165

Ce chapitre est consacré aux expériences de calage-validation réalisées sur les 23 groupes de bassins-versants présentés au chapitre 2 (section 2.2). Les paramètres du modèle hydrologique distribué GRD sont calibrés pour chacun des 23 groupes de bassins-versants. Ce calage est réalisé à l'aide des observations de pluie issues des radars météorologiques (voir section 2.1.1), des données spatiales de l'évapotranspiration potentielle et des données de débits disponibles aux stations hydrométriques délimitant les exutoires aval des bassins-versants. L'objectif est d'estimer pour chaque bassin-versant, un jeu de

paramètres distribués et représentatif de la variabilité spatiale de ses caractéristiques hydrologiques. On tentera de démontrer que ces jeux de paramètres distribués améliorent les performances prédictives du modèle. Dans une première partie de ce chapitre, la méthodologie de ces expériences est détaillée. Dans une seconde partie, les résultats de ces expériences sont présentés.

6.1 Objectifs et méthodologie des expériences de calage-validation

6.1.1 Objectifs

Les expériences de calage-validation croisées présentées dans ce chapitre ont pour objectif de répondre aux questions suivantes : est-ce avantageux de caler et d'utiliser des paramètres spatialement variables par rapport à des paramètres spatialement uniformes ? Les jeux de paramètres distribués (variables spatialement) sont-ils représentatifs des caractéristiques hydrologiques des bassins-versants étudiés ? Améliorent-ils la modélisation des débits à l'intérieur des bassins-versants étudiés, aux exutoires "non-jaugés" ? Quelles sont les principales contraintes et difficultés associées à cette approche et quelles sont les pistes d'amélioration suggérées ?

6.1.2 Méthodologie

Les simulations hydrologiques sont effectuées à partir des données de pluies et débits disponibles sur la période du 01/01/2008 au 01/01/2018. Cette période est divisée en deux sous-périodes de 5 ans chacune : la période P1, du 01/01/2008 au 01/01/2013, et la période P2, du 01/01/2013 au 01/01/2018. Les expériences de calage-validation sont réalisées pour ces deux périodes distinctes. Avant chacune des deux périodes, une année antérieure pleine est utilisée pour stabiliser les niveaux des réservoirs. Cette période est appelée période de "chauffe" du modèle. On suppose qu'une durée d'un an est suffisante (Perrin et al., 2003).

Le modèle hydrologique et ses trois variantes de routage sont décrits au chapitre 3 section 3.2. Les paramètres du modèle GRD sont : les capacités des réservoirs de production notées c_p ; les capacités des réservoirs de transfert notées c_{tr} ; les paramètres du modèle de routage notés v (variante 1), ξ (variante 2) et c_r (variante 3). Ces paramètres sont distribués sur un maillage carré de résolution 1 km^2 . En calage, tous les paramètres sont calés simultanément.

La méthode d'estimation variationnelle présentée au chapitre 4, nécessite un jeu de paramètres initial pour démarrer. Ce jeu de paramètres est appelé "ébauche". Les données observées (les pluies spatialisées et les débits aux exutoires en aval des bassins-versants)

et les incertitudes de modélisation ne permettent pas d'identifier un jeu de paramètres distribués unique et représentatif des vraies caractéristiques hydrologiques spatiales du bassin-versant (non-unicité de la solution, problème montré lors des expériences jumelles présentées au chapitre 5). De ce fait, le choix de l'ébauche est important puisqu'elle permet de contraindre le calage distribué des paramètres (recherche d'un optimum autour de l'ébauche). Le calage est donc réalisé en deux étapes successives.

Lors de la première étape, un jeu de paramètres optimum et spatialement uniforme est recherché pour chaque domaine : $c_p(i, j) = \bar{c}_p$, $c_{tr}(i, j) = \bar{c}_{tr}$, $v(i, j) = \bar{v}$, $\xi(i, j) = \bar{\xi}$, $c_r(i, j) = \bar{c}_r$, où i et j représentent les indices des coordonnées des cellules du maillage. Lors de cette étape, le vecteur de contrôle (le vecteur contenant les paramètres à identifier) ne contient que trois éléments ($\bar{X}_{variante1}(i, j) = (\bar{c}_p, \bar{c}_{tr}, \bar{v})$, $\bar{X}_{variante2}(i, j) = (\bar{c}_p, \bar{c}_{tr}, \bar{\xi})$ et $\bar{X}_{variante3}(i, j) = (\bar{c}_p, \bar{c}_{tr}, \bar{c}_r)$). Ce problème consiste à estimer l'optimum global correspondant au vecteur de contrôle analysé \bar{X}_a . C'est un problème de petite dimension (seulement trois variables sont recherchées) qui est résolu par l'algorithme de recherche global présenté à la section 4.6. Cette première étape est dénommée **calage uniforme**.

Lors de la seconde étape, un jeu de paramètres distribués et variables spatialement est estimé pour chaque bassin-versant étudié. Ce calage utilise comme point de départ (ébauche) le résultat du calage uniforme $X_b(i, j) = \bar{X}_a$. Lors de cette étape, trois paramètres par maille sont calibrés à l'aide de l'algorithme d'estimation variationnelle présenté au chapitre 4. Cette seconde étape est dénommée **calage distribué**.

La méthode de calage autorise de borner la valeur des paramètres (voir section 4.5.1). Ces bornes sont définies pour chacune des deux étapes du calage. Des considérations physiques et numériques imposent les bornes inférieures telles que $(c_p, c_{tr}, v, \xi, c_r) > 0$. Pour le calage uniforme, les bornes supérieures sont choisies afin de préserver la dynamique du modèle. Les valeurs de ces bornes sont présentées au chapitre 4 section 4.5.1. Pour le calage distribué, les bornes sont recalculées automatiquement pour chaque bassin-versant en fonction de la valeur de l'ébauche tel que $X_{max} = b \times X_b$ et $X_{min} = X_b/b$, où $b = 4$ est utilisé pour ces expériences. Ces bornes encadrent l'ébauche très largement. Le problème est finalement peu contraint par ces bornes.

L'étape de validation du modèle consiste à vérifier les performances prédictives du modèle à l'aide de données qui n'ont pas été utilisées pour le calage. Les performances du modèle sont évaluées à l'aide des critères de NSE (qui a servi au calage, voir section 4.3.1) et KGE. Le critère de KGE est un compromis parmi trois critères (la corrélation, le biais et la variance). Ce critère est formulé ci-dessous à l'équation 6.1 :

$$KGE = 1 - \sqrt{(P - 1) + (Rmu - 1) + (Rvar - 1)} \quad (6.1)$$

où Rmu mesure le biais entre les débits moyens simulés $\overline{Q_{sim}}$ et observés $\overline{Q_{obs}}$,

$$Rmu = \frac{\overline{Q_{sim}}}{\overline{Q_{obs}}}, \quad (6.2)$$

$Rvar$ mesure le rapport entre les variances des débits simulés σ_{sim} et observés σ_{obs} ,

$$Rvar = \frac{\sigma_{sim}}{\sigma_{obs}}, \quad (6.3)$$

et P est le coefficient de corrélation linéaire de Pearson calculé entre les débits simulés et observés pour chaque pas de temps $t \in [1, T]$,

$$P = \frac{\sum_t^T (Q_{sim}^t - \overline{Q_{sim}})(Q_{obs}^t - \overline{Q_{obs}})}{\sqrt{\sum_t^T (Q_{sim}^t - \overline{Q_{sim}})^2} \sqrt{\sum_t^T (Q_{obs}^t - \overline{Q_{obs}})^2}}. \quad (6.4)$$

La validation temporelle consiste à vérifier les performances prédictives du modèle sur la période qui n'aura pas servi au calage. Le jeu d'observations (voir sections 2.1.1 et 2.2.2) est divisé en deux sous-jeux de données : l'un est utilisé pour le calage et l'autre est utilisé pour la validation. Chaque période (P1 et P2) peut donc être utilisée à tour de rôle pour le calage ou pour la validation.

La validation spatiale consiste à vérifier les performances prédictives du modèle aux stations de jaugeage qui n'ont pas servi au calage. Les données de débit disponibles aux stations de jaugeage sont utilisées soit pour le calage (les stations hydrométriques situées à l'aval de chaque bassin-versant sont utilisées pour le calage) soit en validation (les stations hydrométriques situées à l'intérieur des bassins-versants sont utilisées pour la validation). Si les données de débits de la station hydrométrique sont utilisées en calage, le bassin-versant correspondant est dénoté "bassin de calage". Inversement, si les données de débits de la station hydrométrique sont utilisées en validation, le bassin-versant correspondant est dénoté "bassin de validation". La figure 6.1 schématise un bassin-versant de calage associé à des sous-bassins-versants de validation.

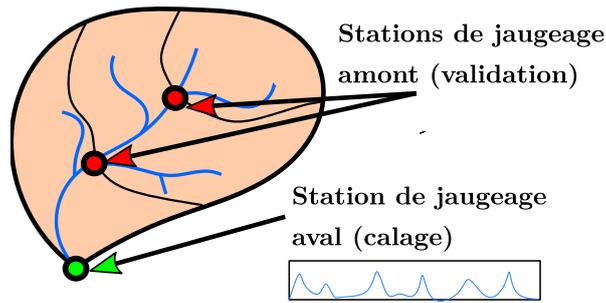


FIGURE 6.1 – Schéma représentant un bassin-versant de calage (utilisation des données de débits à la station de jaugeage aval) et les sous-bassins-versants de validation (utilisation des données de débits aux stations de jaugeage amont). Les contours des bassins-versants sont tracés en noirs. Le réseau hydrographique est représenté en bleu.

Par la suite, on distingue deux niveaux de validation :

- La validation spatiale : les performances prédictives du modèle sont calculées sur les bassins de validation mais sur la période de calage.
- La validation spatio-temporelle : les performances prédictives du modèle sont calculées sur les bassins de validation et sur la période de validation.

6.1.3 Expériences réalisées

À partir de la méthodologie de calage-validation présentée à la section précédente, des simulations permettant le calage des paramètres et l'analyse des performances du modèle en validation sont réalisées. Ces simulations sont réalisées pour le jeu de données présentées au chapitre 2. Les bassins-versants modélisés sont listés dans le tableau 6.1 et sont cartographiés en annexe A. Les stations de jaugeages pour lesquelles les données de débits sont utilisées en calage et en validation sont listés en annexe B.

TABLE 6.1 – Tableau récapitulatif des 23 méta-bassins-versants étudiés. Le nom de la rivière, la surface du bassin-versant à l'exutoire aval, le code de la station hydrométrique situé à l'exutoire aval et le nombre de stations internes à chaque bassin-versant sont mentionnés.

Numéro	Nom de la rivière	Surface (km^2)	Code	Nombre d'exutoires internes
1	Le Loing	2302	H3201010	9
2	La Cance	381	V3524010	4
3	Le Doux	621	V3744010	3
4	L'Eyrieux	649	V4154010	4
5	L'Ardèche	2264	V5064010	10
6	La Cèze	1112	V5474015	7
7	Le Gardon	1093	V7164015	11
8	L'Aude	1828	Y1232010	12
9	Le Fresquel	935	Y1364010	5
10	L'Orbiel	242	Y1415020	3
11	L'Orbieu	589	Y1564010	4
12	La Cesse	251	Y1605050	2
13	L'Hérault	2208	Y2332015	13
14	L'Orb	1336	Y2584010	12
15	Le Lez	168	Y3204040	4
16	Le Vidourle	503	Y3444020	5
17	Le Vistre	496	Y3534010	2
18	Le Gapeau	535	Y4624010	7
19	L'Argens	2512	Y5312010	11
20	La Giscle	201	Y5444010	3
21	La Siagne	492	Y5534030	6
22	Le Loup	289	Y5615030	2
23	L'Estéron	442	Y6434010	2

118

Les simulations sont réalisées dans l'ordre suivant :

1. Simulations pour le calage des paramètres : calages **uniformes** suivis des calages **distribués** en utilisant les données de débits disponibles aux exutoires des bassins-versants pour les périodes temporelles P1 et P2.
2. Simulations pour les validations spatiales : Les calages uniforme et distribué sont suivis de **validations spatiales** en utilisant les données de débits aux stations de validation (stations de jaugeage situées à l'intérieur des bassins-versants et non utilisées pour le calage) sur les périodes de calage P1 et P2
3. Simulation pour les validations spatio-temporelles : Les calages uniformes et distribués sont suivis de validations **spatio-temporelles** à l'aide des données de débits aux stations de validation sur les périodes de validation P2 et P1.

L'organigramme suivant, figure 6.2 schématise la structure méthodologique des simulations permettant le calage des paramètres et la validation des résultats.

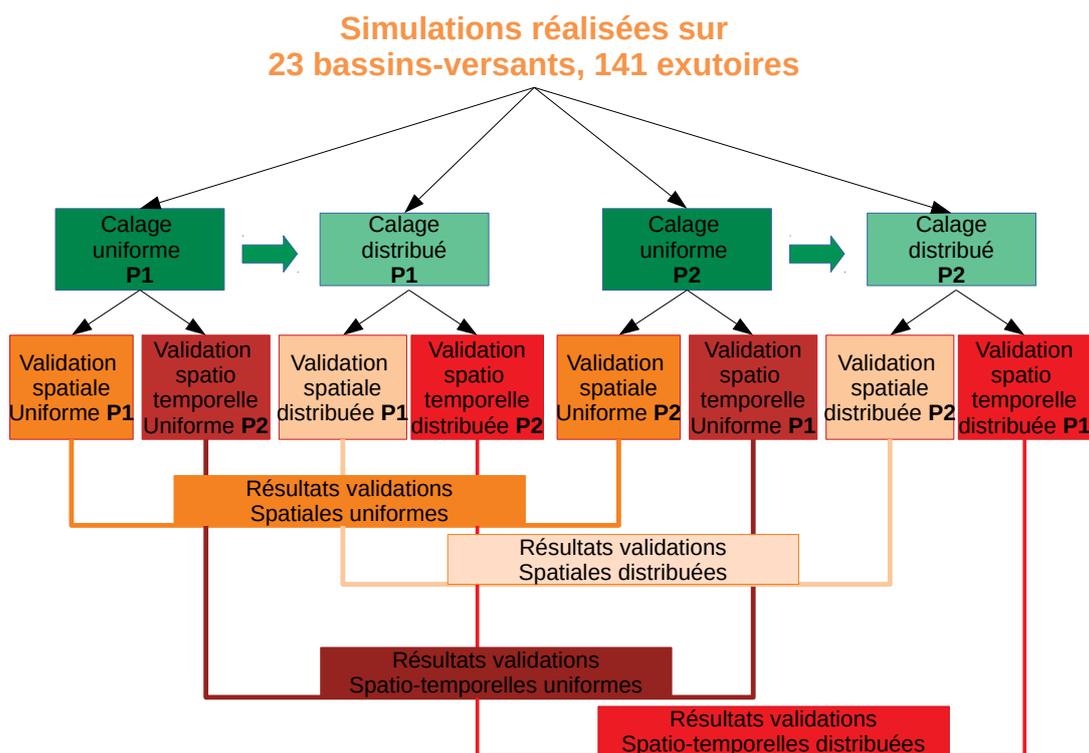


FIGURE 6.2 – Organigramme récapitulant les simulations réalisées sur les 23 groupes de bassins-versants. Les résultats sont ensuite groupés par catégorie de simulations pour les périodes P1 et P2.

Les simulations permettant le calage et la validation des résultats se déclinent en six expériences selon la paramétrisation de l’algorithme de calage et la configuration du modèle. Les caractéristiques de ces six expériences sont présentées ci-dessous et sont synthétisées dans le tableau 6.2 :

1. **Expérience A** : expérience qui consiste au calage uniforme des paramètres sur les domaines à l’aide d’un algorithme de recherche stochastique (voir section 4.6). Le calage s’effectue à l’aide des données aux exutoires aval des 23 bassins-versants (voir tableau 6.1). Pour cette expérience les variantes 1, 2 et 3 du modèle sont testées.
2. **Expérience B** : expérience qui consiste au calage distribué des paramètres avec régularisation itérative. L’ébauche est identifiée par le calage uniforme (expérience A) comme présentée dans la section méthodologie 6.1.2). Le calage s’effectue à l’aide des données aux exutoires situés en aval des 23 bassins-versants (voir tableau 6.1). Pour cette expérience les variantes 1, 2 et 3 du modèle sont testées.
3. **Expérience C** : expérience qui consiste au calage distribué des paramètres sans

régularisation et sans se soucier du point de départ. L'ébauche est alors définie arbitrairement de telle sorte que ($c_p = 200 \text{ mm}$, $c_{tr} = 100 \text{ mm}$, $\xi = 1$) indépendamment des bassins-versants modélisés. Le calage s'effectue à l'aide des données aux exutoires situés en aval des 23 bassins-versants (voir tableau 6.1). Pour cette expérience, seule la variante 2 du modèle est utilisée.

4. **Expérience D** : expérience qui consiste au calage distribué des paramètres sans régularisation mais dont l'ébauche est identifiée par un calage préalable uniforme (expérience A) comme présentée dans la section méthodologie 6.1.2). Le calage s'effectue à l'aide des données aux exutoires situés en aval des 23 bassins-versants (voir tableau 6.1). Pour cette expérience, seule la variante 2 du modèle est utilisée.
5. **Expérience E** : expérience qui consiste au calage distribué des paramètres avec régularisation. L'ébauche est identifiée par un calage uniforme (expérience A) comme présentée dans la section méthodologie 6.1.2). L'identification du coefficient optimal de régularisation β est réalisée à partir de la méthode simplifiée de la L-Curve présentée à la section 4.2.4. Le calage s'effectue à l'aide des données aux exutoires situés en aval des 23 bassins-versants (voir tableau 6.1). Pour cette expérience, seule la variante 2 du modèle est utilisée.
6. **Expérience F** : expérience qui consiste au calage distribué des paramètres à l'aide des données de débits disponibles à l'ensemble des 141 exutoires (voir tableau B). La régularisation n'est pas utilisée mais l'ébauche est identifiée par un calage uniforme (expérience A). Pour cette expérience, seule la variante 2 du modèle est utilisée.

Le tableau suivant (tableau 6.2) synthétise la paramétrisation des six expériences de **calage uniforme / calage distribué** :

TABLE 6.2 – Tableau récapitulatif des expériences de calage-validation.

Nom	Type de calage		Régularisation de type Tikhonov		Ébauche calculée depuis un calage uniforme	Nombre de station de calage	Variantes du modèle
	Uniforme	Distribué	Méthode itérative	Méthode L-curve			
A	×					23	1,2,3
B		×	×		×	23	1,2,3
C		×				23	2
D		×			×	23	2
E		×		×	×	23	2
F		×			×	118 + 23 = 141	2

Remarque : les données de débits et de pluies sont entachés d’incertitude et les choix quand aux stations de jaugeage utilisées pour le calage et pour la validation sont critiquables. Cependant les simulations réalisées sont toutes impactées de la même manière et les résultats des expériences sont comparées entre-elles. De ce fait les problèmes liés aux données ne devraient pas fausser les conclusions.

6.2 Résultats

6.2.1 Comparaison des performances des variantes 1, 2 et 3 du modèle hydrologique

L’expérience de calage **B** a été réalisée pour les trois variantes du modèle hydrologique. Ces variantes se distinguent par la modélisation du routage des hydrogrammes entre chaque pixel du maillage. Dans cette partie les performances des trois variantes du modèle, obtenues aux stations de calage et aux stations de validation à l’issue de l’expérience de calage **B**, sont comparées. Ces expériences se focalisent sur le fonctionnement hydrologique du modèle et sont réalisées avec des données hydrologiques réelles. Elles se distinguent donc des expériences réalisées au chapitre précédent (chapitre 5) qui visaient à évaluer uniquement le fonctionnement des modèles de routage avec des données fictives.

Les critères de performances NSE sont calculés aux exutoires de calage (23 exutoires) et de validation (118 exutoires) pour chaque période P1 et P2. Les valeurs du critère

NSE (23×2 et 118×2) sont triées par ordre croissant et sont présentées sous forme de distribution statistique sur les graphiques 6.3 et 6.4.

Aux stations de calage (figure 6.3, graphique de gauche), les performances des trois variantes sont comparables. Les performances sont excellentes pour 90% des stations ($NSE > 0.8$). Les performances sont mauvaises pour une seule station correspondant à la période de calage P1 ($NSE < 0.4$). Il s'agit du bassin-versant de la Siagne à la station Y5534030 "la Siagne à Pégomas" (la figure 6.10 permet de l'identifier). La cause de ces mauvaises performances en calage sur ce bassin-versant est vraisemblablement lié aux données de pluie et de débits. Ce problème sera discuté dans la prochaine partie.

Les performances en validation spatiale (figure 6.3, graphique de droite) sont identiques entre les trois variantes. Les performances des modèles sont bonnes pour environ 60% des stations de validation ($NSE > 0.7$). Elles se dégradent pour les 40% restantes. Pour 15% d'entre elles, les critères NSE calculés sont négatifs.

Les performances en validation spatio-temporelle (figure 6.4) sont aussi très comparables entre les trois variantes. On remarquera que les performances de la variante 3 sont légèrement inférieures entre les fréquences 0.3 et 0.4 et que les performances de la variante 1 sont légèrement inférieures pour les fréquences supérieures à 0.4.

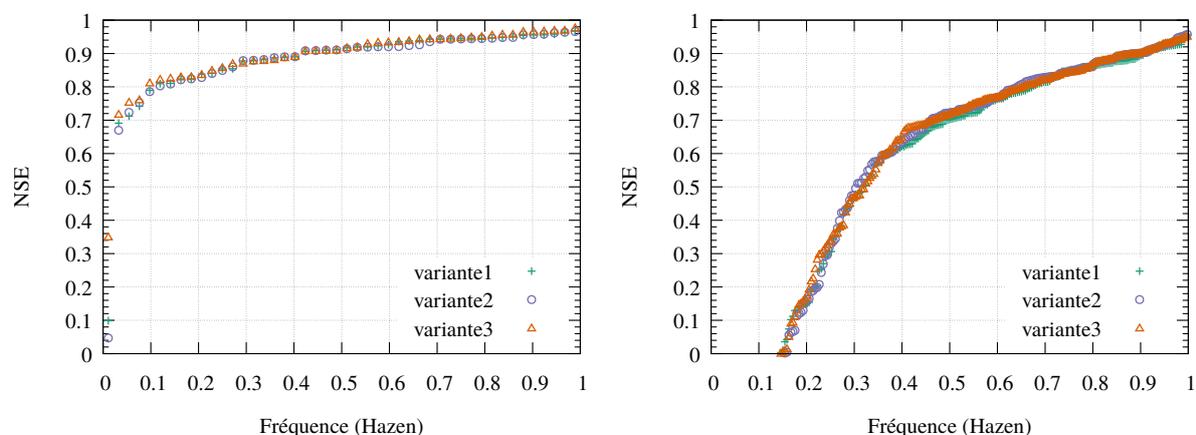


FIGURE 6.3 – Performances obtenues pour les variantes 1, 2 et 3 du modèle GRD après calage **distribué** des paramètres (**expérience B**). Le graphique de gauche montre les performances (NSE) obtenues en calage (23 stations de calage, 2 périodes). Le graphique de droite montre les performances (NSE) en validation spatiale (118 stations de validation, 2 périodes).

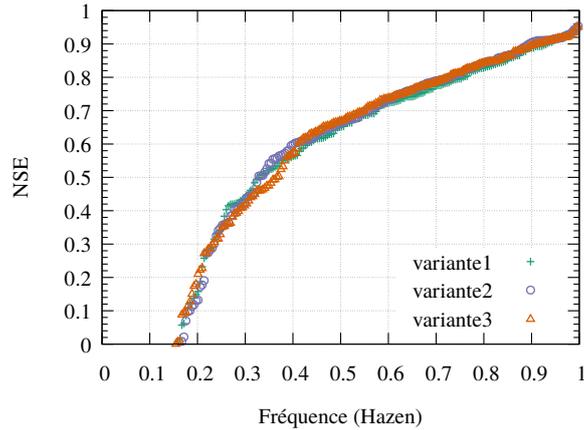


FIGURE 6.4 – Performances (NSE) en validation spatio-temporelle (118 stations de validation, 2 périodes) obtenues pour les variantes 1, 2 et 3 du modèle GRD après calage **distribué** des paramètres (**expérience B**).

Conclusion : Premièrement, ces résultats indiquent que l’algorithme de calage fonctionne correctement. Le modèle GRD permet de simuler des réponses hydrologiques variées. Les performances du modèle utilisant des paramètres variables spatialement sont en général très bonnes aux exutoires de calage.

Deuxièmement, ces résultats montrent que les trois modèles de routage n’influencent pas significativement les performances prédictives du modèle hydrologique GRD fonctionnant au pas de temps horaire avec une résolution spatiale de $1km^2$. Bien que d’un point de vue conceptuel la variante 2 semble être un meilleur modèle de propagation des crues (cf chapitre 5 section 5.2), le modèle de routage ne semble pas être l’opérateur prépondérant permettant d’améliorer les performances du modèle hydrologique. Cela peut s’expliquer par un problème de compensation des paramètres de transfert et de routage et par des incertitudes trop importantes sur les observations ne permettant pas de bien distinguer ces deux opérateurs assez proches.

Pour la suite du manuscrit, nous avons choisi de montrer uniquement les résultats obtenus pour la variante 2. La raison est que les trois variantes donnent des résultats similaires et que leurs présentations détaillées ne présentent pas d’intérêt particulier. Le choix de la variante 2 repose donc sur les résultats obtenus au chapitre 5. Un travail supplémentaire doit encore être mené pour départager ou améliorer ces variantes, mais cet objectif sort des cadres fixés dans ces travaux de thèse.

6.2.2 Comparaison des performances du modèle utilisant des jeux de paramètres issus des calages uniforme et distribué

Les performances du modèle hydrologique utilisant des jeux de paramètres calés de manière uniforme et distribué sont comparées. Ces résultats sont issus des expériences de calage uniforme **A** et distribué **B** utilisant la variante 2 du modèle hydrologique (voir tableau 6.2).

Les performances prédictives du modèle sont évaluées à l'aide des critères NSE (équation 4.27) et KGE (équation 6.1) aux exutoires de validation sur les périodes de calage (validation spatiale) et de validation (validation spatio-temporelle). Les performances aux exutoires de validation période P1 et P2 (118×2 points) sont représentées sous forme de distributions statistiques (fréquences) aux graphiques 6.5 et 6.6.

Le calage distribué des paramètres améliore significativement les performances prédictives du modèle. Les gains sont plus conséquents en validation spatio-temporelle. Ce résultat est important puisqu'il s'agit de l'étape d'évaluation la plus difficile pour le modèle. Les gains sont compris entre 5% et 15% en termes de NSE et 2% à 10% en termes de KGE. Les gains sont cependant plus limités si les performances du modèle sont évaluées à l'aide du critère KGE qui n'est pas utilisé pour le calage.

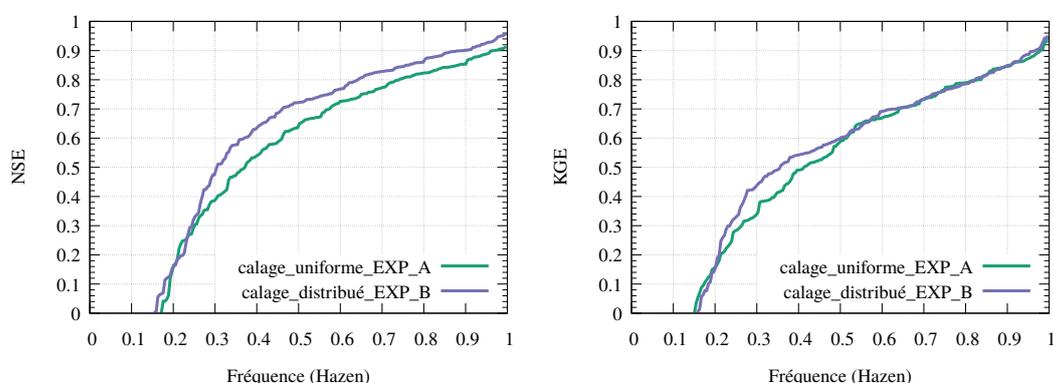


FIGURE 6.5 – Performances en **validation spatiale**, 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage **uniforme** (expérience **A**, courbe verte) et **distribué** (expérience **B**, courbe bleu) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

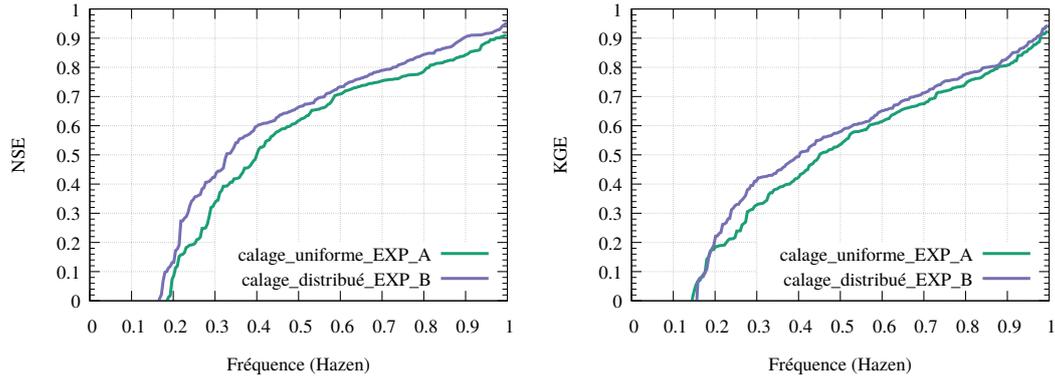


FIGURE 6.6 – Performances en **validation spatio-temporelle**, 118 exutoires de validation sur les périodes de validation P1 et P2, obtenues pour la variante 2 du modèle GRD après calage **uniforme** (expérience **A**, courbe verte) et **distribué** (expérience **B**, courbe bleu) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

Les performances prédictives du modèle sont ensuite étudiées en fonction de l'éloignement des bassins-versants de validation par rapport à l'exutoire de calage aval. Les exutoires sont classés selon le critère RS suivant :

$$RS_i = \frac{S_i}{S_{aval}},$$

où i est l'indice de l'exutoire de validation, S_i la surface du bassin-versant associée à l'exutoire i , S_{aval} la surface associée à l'exutoire de calage.

La figure 6.7 permet de visualiser la répartition des critères de Nash-Sutcliffe (NSE) en fonction de cinq classes du rapport RS pour les exutoires de validation respectivement en validation spatiale et en validation spatio-temporelle. Les gains apportés par le calage distribué des paramètres sont compris entre 5% et 15% en termes de NSE quelle que soit la valeur du rapport RS . De plus, les performances prédictives obtenues avec un calage distribué des paramètres sont moins hétérogènes et donc plus "stables" (box-plot réduite autour de la valeur médiane). Les performances prédictives du modèle deviennent meilleures lorsque le rapport RS augmente. Ces performances diminuent significativement pour un ratio $RS \leq 0.2$. Cela signifie que la prédictibilité des débits est plus difficile pour les exutoires situés dans les zones en amont des 23 méta-bassins-versants. Ce résultat est cohérent : plus on s'éloigne des exutoires situés en aval (exutoires ayant servi au calage) et plus la sensibilité des événements hydrologiques affectant l'amont des bassins-versants diminue. Ce résultat est obtenu à la fois pour les calages uniforme et distribué des paramètres.

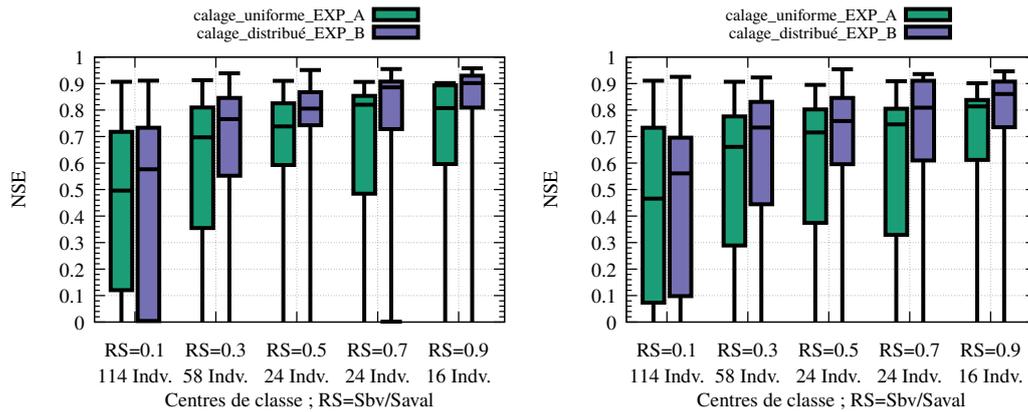


FIGURE 6.7 – Relation entre les critères de Nash-Sutcliffe (NSE) et cinq classes du rapport $RS_i = \frac{S_i}{S_{aval}}$ (surface de l'exutoire de validation) / (Surface de l'exutoire de calage). Graphique de gauche : résultat en validation spatiale, expériences **A** et **B**. Graphique de droite : résultat en validation spatio-temporelle, expériences **A** et **B**

Conclusions : Ces résultats montrent que le calage distribué des paramètres améliore les performances prédictives du modèle hydrologique. L'utilisation d'un jeu de paramètres distribués par rapport à un jeu de paramètres uniformes est préférable. Par exemple un score de performance NSE supérieur à 0.6 est obtenu pour 60% des exutoires de validation avec un jeu de paramètres distribués contre 50% avec un jeu de paramètres uniformes. L'algorithme d'estimation variationnelle basé sur le modèle adjoint est efficace pour calibrer les paramètres d'un modèle hydrologique distribué. Techniquement, ce calage peut être réalisé sur des grandes étendues géographiques et à partir de données observées sur de longues périodes temporelles. Néanmoins, l'hydrologie des bassins-versants situés en amont (éloignés de l'exutoire ayant servi au calage) reste difficile à modéliser malgré des paramètres variables spatialement. L'application réalisée est cependant extrême car les expériences sont privées d'un grand nombre d'observations. Par exemple, l'information des quelques bassins "amonts" aurait pu être utilisée, de même que l'information de l'ensemble des deux périodes d'observation.

6.2.3 Critique du calage de l'expérience **B** pour chaque période, comparaison à l'expérience **F**

Les performances du modèle hydrologique distribué sont analysées à l'aide des données de débits observés aux exutoires de calage et de validation sur les périodes temporelles P1 et P2. Les résultats des expériences **B** et **F** sont discutés. Seuls les résultats de modélisation issus du modèle variante 2 sont présentés. Les objectifs de cette analyse sont d'abord de critiquer les débits simulés par rapport aux débits observés et ensuite de me-

sur les performances du modèle par rapport aux conditions expérimentales suggérées (calage distribué réalisé à l'aide des données observées d'une seule station de jaugeage située à l'aval de chaque bassin-versant pour l'expérience **B**, versus un calage utilisant toutes les jauges disponibles expérience pour l'expérience **F**).

Le graphique 6.8 présente les performances du modèle hydrologique aux exutoires de calage et de validation (expérience **B**) en distinguant les périodes temporelles de calage P1 et P2. Ces performances sont présentées sous forme de distribution statistique des critères de performance NSE sur les périodes de calage.

Les performances du modèle GRD aux exutoires de calage et de validation sont globalement moins bonnes pour la période P1. Il semble possible que ces différences soient causées par la qualité des données disponibles sur la période P1 (données de pluies et débits plus anciennes et de qualité moindre).

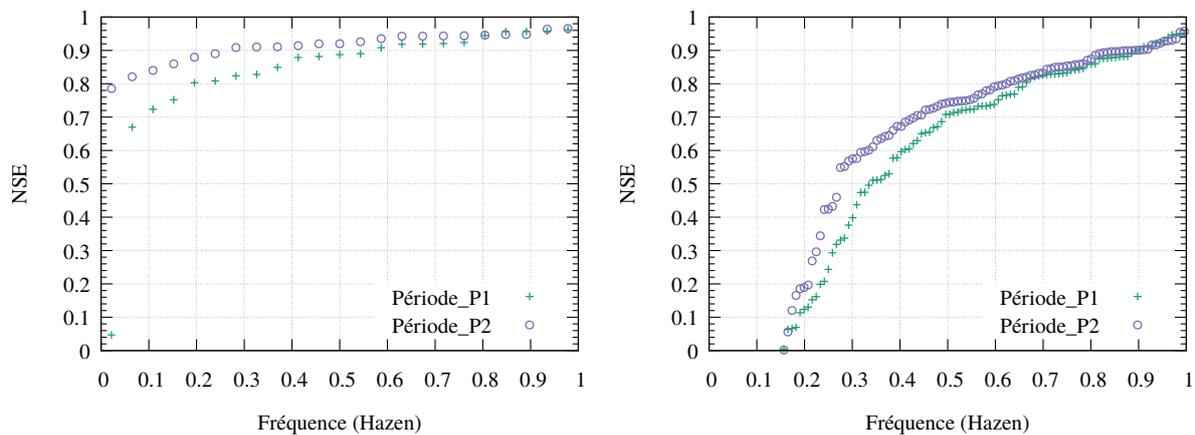


FIGURE 6.8 – Comparaison des performances obtenues entre les périodes P1 et P2 pour la variante 2 du modèle GRD après calage **distribué** des paramètres (expérience **B**). Le graphique de gauche montre les performances (NSE) obtenues en calage (23 stations de calage). Le graphique de droite montre les performances (NSE) en validation spatiale (118 stations de validation).

En calage, les performances sont mauvaises ($NSE < 0.1$) pour le seul bassin-versant de la Siagne (station Y5534030) sur la période P1 (voir figure 6.10). Les causes de ces mauvaises performances en calage sur ce bassin-versant sont recherchées. Pour cette station et contrairement à la période P1, le NSE calculé sur la période P2 est supérieur à 0.8. L'analyse des chroniques des débits observés pour cette station montre que les débits faibles sont très variables (bruits sur les mesures ou influences en basses eaux). Cette variabilité est vraisemblablement artificielle mais ne semble pas être à l'origine des défaillances du modèle sur la période P1 (les performances sont bonnes sur la période P2). L'observation approfondie des chroniques de débits observés et simulés en période P1,

met en évidence une succession de deux évènements pluvieux majeurs en novembre 2011 (voir figure 6.9). Le premier évènement entraîne une forte crue de la Siagne. Le deuxième évènement n'entraîne pas de réaction du cours d'eau alors que la pluie observée moyenne sur le bassin-versant semble largement supérieure à celle du premier évènement. Il semble très probable que ce deuxième évènement soit responsable de la défaillance du calage. Les données observées de pluies et de débits semblent incohérentes pour cette date. Un évènement pluvieux de cette intensité devrait entraîner une réaction marquée du bassin-versant, or aucune réaction n'est mesurée à la station de jaugeage aval. Les résultats de la simulation avec des paramètres calés de manière uniforme (expérience **A**) montrent que le bassin-versant réagit fortement à l'évènement pluvieux du 21/11/2011. Remarquons que cette réaction est marquée bien que la paramétrisation du modèle permet d'absorber intégralement la crue précédente datant du 07/11/2011. Le calage distribué (expérience **B**) permet d'atténuer drastiquement la crue du 21/11/2011, tout en générant une légère réaction du bassin-versant lors de l'évènement précédent (07/11/2011). En ce sens, la simulation est améliorée par le calage distribué. En revanche les performances sur la période temporelle P1 restent très mauvaises ($NSE < 0.1$). Cet exemple montre que le calage peut être très sensible aux incertitudes sur les données.

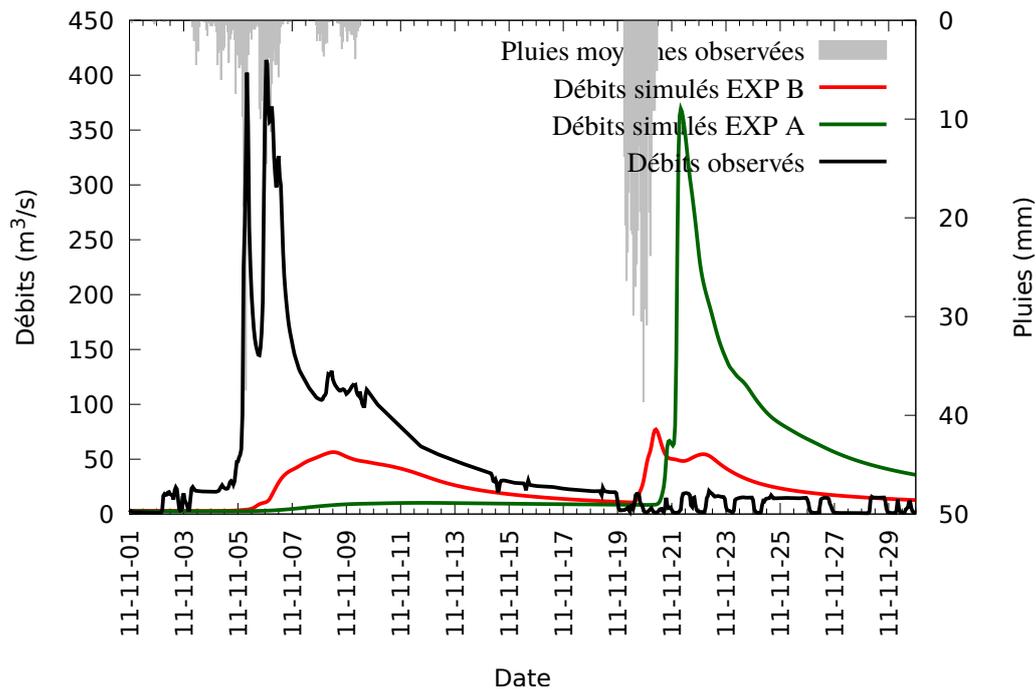


FIGURE 6.9 – Débits observés (noir) et simulés pour les expériences **B** et **A** (variante 2) à la station de calage "la Siagne à Pégomas" (Y5534030) pour des évènements pluvieux importants entre le 1 et le 30 Novembre 2011. Les pluies moyennes sur le bassin-versant sont représentées par l'histogramme (gris).

Les critères de NSE obtenus aux stations de calage et de validation pour les périodes de calage P1 et P2 sont cartographiés sur la figure 6.10. Pour les bassins-versants de l'Aude, l'Orb, l'Argens, le Gapeau, le Vidourle et la Siagne, les performances aux stations de validation sont médiocres. Pour la Siagne, un problème de données a été identifié en période P1 et explique en partie ces mauvaises performances (problème de calage). Pour l'Argens, le Vidourle et le Gapeau, les stations pour lesquelles les performances du modèle en validation sont mauvaises sont principalement situées en amont. Les causes de ces mauvaises performances sont variées : les données de débits peuvent être incomplète, très incertaines ou influencées ; le calage distribué réalisé dans les conditions expérimentales proposés peut rendre la modélisation très incertaine (variabilité spatiale des paramètres non représentative des caractéristiques hydrologiques des bassins-versants) ; le modèle lui-même est imparfait, simplifié et ne permet pas, par exemple, de simuler les réponses hydrologiques naturelles.

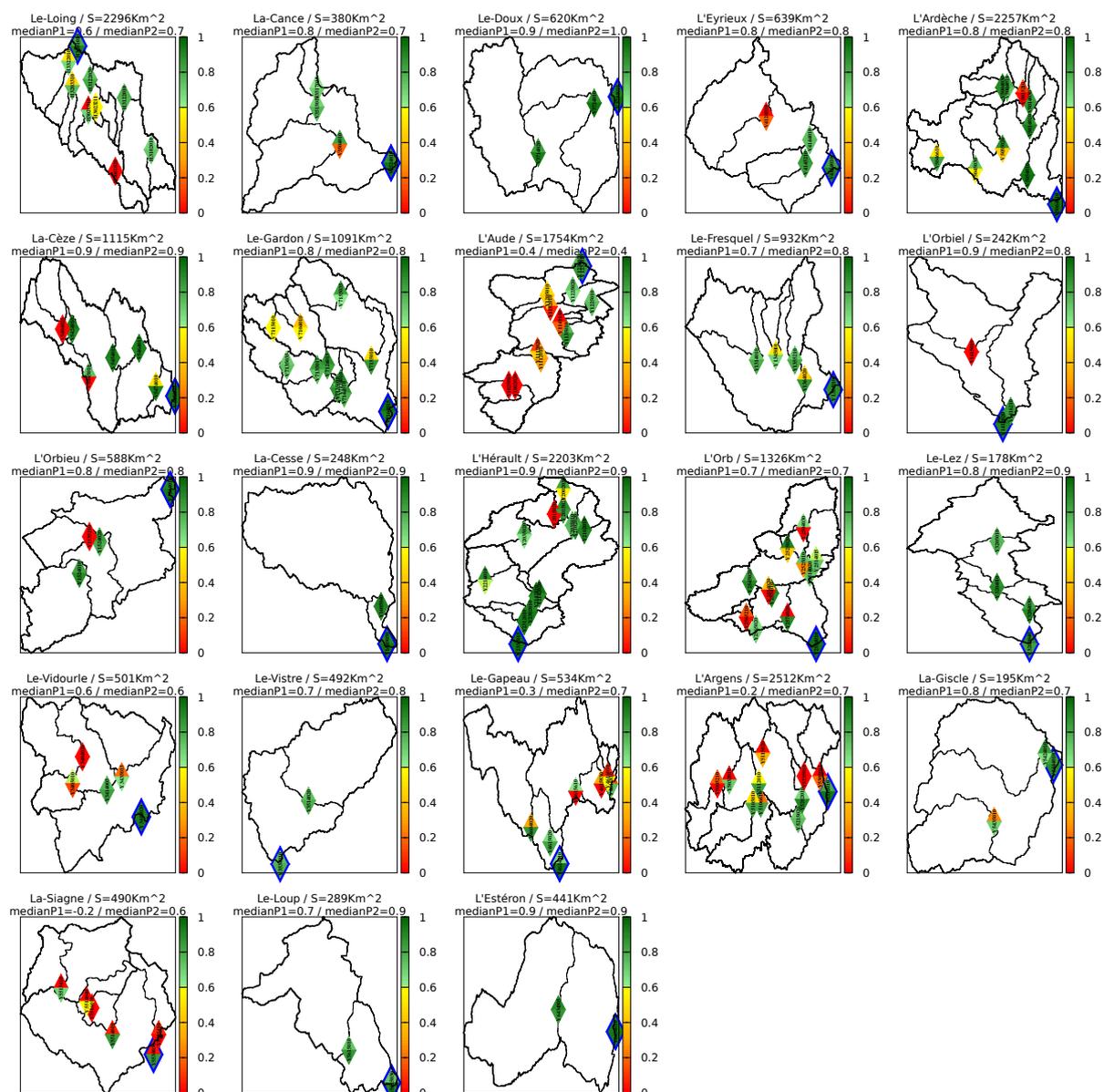


FIGURE 6.10 – Performances (NSE) en calage et en validation spatiales obtenues pour la variante 2 du modèle GRD après calage **distribué** des paramètres pour les périodes P1 et P2 et pour chacun des 23 bassins-versants (aux exutoires de calage et de validation, expérience **B**). Les exutoires symbolisés avec le triangle pointant vers le haut correspondent à la période P1. Les exutoires symbolisés avec le triangle pointant vers le bas correspondent à la période P2. Les stations utilisées pour le calage sont marquées avec un contour bleu.

L'expérience de calage **F** permet d'identifier le meilleur jeu de paramètre possible puisque l'ensemble des données de débits sont utilisées pour le calage. La comparaison des résultats de l'expérience **B** et **F** doit permettre ainsi de distinguer :

- les stations de jaugeage pour lesquelles les données et les défaillances du modèle sont responsables des mauvaises performances ;

- les stations de jaugeages pour lesquelles le calage distribué des paramètres ne permet pas de représenter correctement les processus hydrologiques.

Les performances du modèle pour l'expérience **F** sont présentées sur la figure 6.11 à l'aide du critère de NSE tout en distinguant les périodes temporelles P1 et P2.

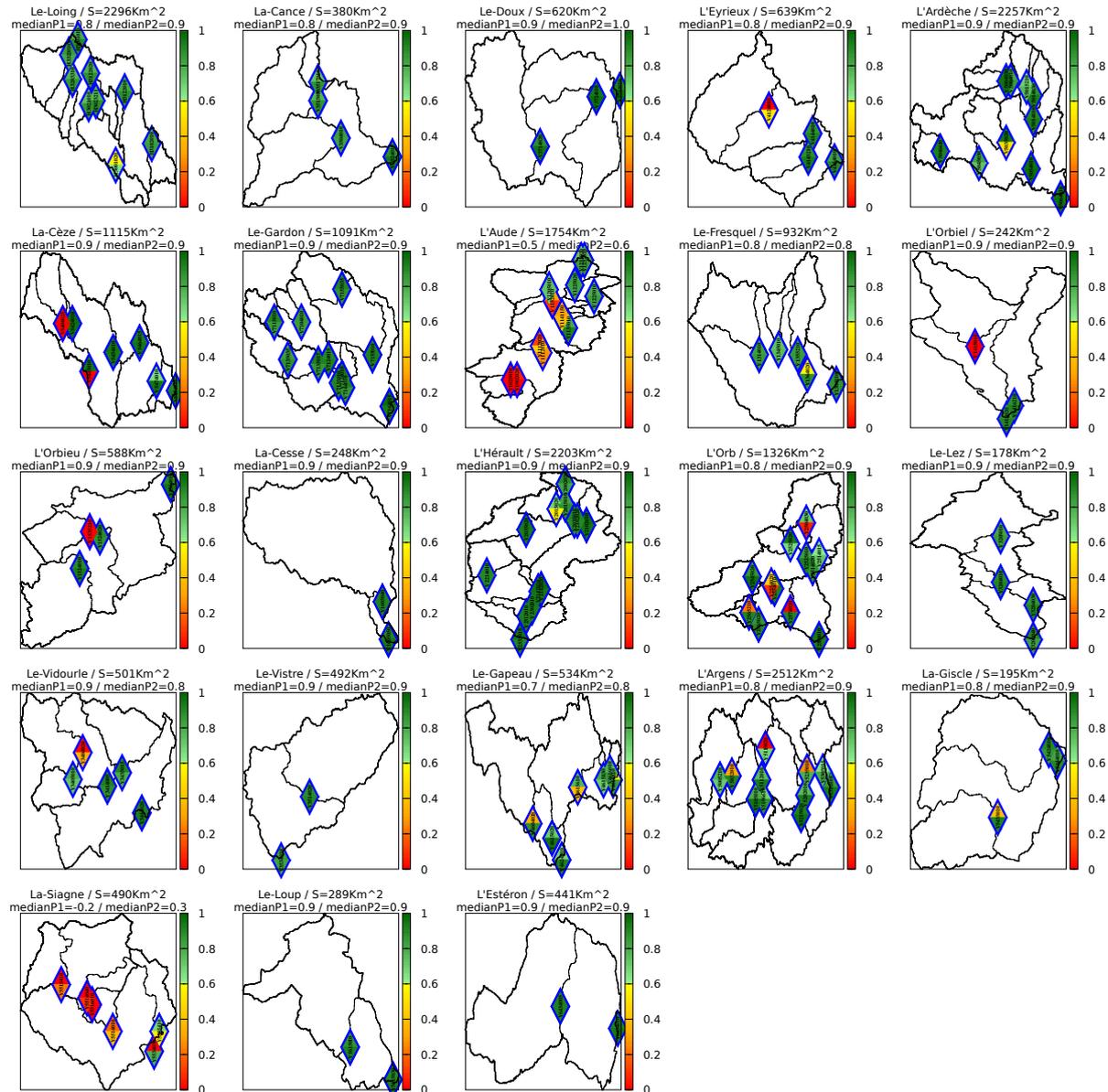


FIGURE 6.11 – Performances (NSE) obtenues pour la variante 2 du modèle GRD après calage **distribué** des paramètres pour les périodes P1 et P2 et pour chacun des 23 bassins-versants (calage de référence, expérience **F**, réalisé à l'aide des données hydrométriques aux 141 exutoires). Les exutoires symbolisés avec le triangle pointant vers le haut correspondent à la période P1. Les exutoires symbolisés avec le triangle pointant vers le bas correspondent à la période P2. Les stations utilisées pour le calage sont marquées avec un contour bleu.

La comparaison des performances à l'issue des expériences **B** et **F** est réalisée à l'aide des critères suivants. Les performances sont classées en deux catégories : mauvaises lorsque le $NSE \leq 0.6$ (échecs) et bonnes lorsque $NSE > 0.6$ (succès). Pour chaque bassin-versant et chacune des deux expériences (**B** et **F**), les échecs et succès sont comptabilisés (tableau 6.3). Deux scores sont ensuite calculés pour chaque bassin-versant.

Le potentiel modélisable du bassin-versant se calcule selon la formule suivante :

$$PM = \frac{Nb \text{ succès en calage}}{Population} \quad (6.5)$$

Plus le score PM est proche de 1 et plus le bassin-versant est "modélisable" par le modèle GRD. Inversement, un score PM proche de 0 indique que le modèle n'arrive pas à simuler correctement les hydrogrammes même lorsque ceux-ci sont utilisés pour le calage.

Le score de calage se calcule selon la formule suivante :

$$SC = \frac{Nb \text{ succès en validation}}{Nb \text{ succès en calage}} \quad (6.6)$$

Plus le score SC est proche de 1 et plus le calage réalisé par l'expérience **B** est satisfaisant. Inversement, plus le SC est petit et plus le calage réalisé par l'expérience **B** est défectueux aux stations de validation.

Les nombres de succès et d'échecs en calage sont comptabilisés dans le tableau 6.3. L'expérience **F** montre que au total 48 chroniques de données hydrologiques (stations et périodes comprises) sont indescriptibles par le modèle. Cela représente 47 chroniques utilisées en validation pour l'expérience **B**. Ce nombre représente environ 20% de la population statistique des stations de validation étudiées (118 exutoires, deux périodes). La cause de ces échecs de modélisation peut être attribuée en partie à la qualité des données de débits et de pluies. En effet, les débits observés sont parfois fortement influencés (barrage, prélèvements, rejets) et les données de pluies et débits peuvent être entachées de fortes incertitudes. Certaines chroniques de débits douteuses sont tracées sur la figure 6.12. Remarquons que, pour certaines stations dont les débits sont fortement influencés en basses-eaux, les pics de crues peuvent être bien représentés (figure 6.12, stations V4124010 sur l'Eyrieux).

Les résultats des scores PM et SC sont présentés dans le tableau 6.3. Le score $PM > 0.8$ pour 16 bassins-versants. De plus, pour 14 d'entre eux, les nombres d'échecs en calage pour l'expérience **F** ne dépassent pas 1. Cela signifie que globalement, les chroniques de débits observés appartenant à ces bassins-versants sont modélisables et ne semblent pas être affectées par de fortes incertitudes sur les données dans la mesure où

le calage de l'expérience **F** a utilisé toute l'information disponible.

Inversement, les bassins-versants de la Siagne et de l'Aude sont peu "modélisables" ($PM = 0.25$ et $PM = 0.50$ respectivement). Cela signifie que les performances du modèle aux stations hydrométriques utilisées en calage sont globalement mauvaises. Pour ces bassins-versants, les mauvaises performances en calage peuvent être attribuées à la qualité des données.

Pour les six bassins-versants du Gapeau, du Vidourle, de l'Argens, du Loing, de l'Ar-dèche et du Gardon, les scores SC sont faibles ($0.53 < SC < 0.79$) alors que les scores PM sont élevés ($0.79 < PM < 1$). Cela signifie que les jeux de paramètres issu du calage par l'expérience **B** ne permettent pas de simuler correctement les débits aux exutoires de validation. Pour ces bassins-versants, les calages distribués des paramètres (dans les conditions de l'expérience **B**) ne semblent pas représentatifs des variabilités spatiales des caractéristiques hydrologiques. Pour cinq autres bassins-versants (la Cèze, l'Orb, le Fresquel, l'Hérault et la Cance) les scores SC sont corrects mais pourraient être améliorés ($0.8 < SC < 1$). Au total, les chroniques de débits observés pour 41 stations hydrométriques ne sont pas correctement simulées par le modèle à l'issue de l'expérience **B** alors qu'elles le sont pour l'expérience **F**.

Pour dix bassins-versants, les scores SC valent 1. Cela signifie que le calage réalisé avec les données hydrométriques des exutoires en aval (expérience **B**) permet de simuler correctement les hydrogrammes aux stations de validation lorsque cela est possible (pour cinq bassins-versants le score PM est inférieur à 1).

TABLE 6.3 – Tableau permettant de comparer les performances des expériences B et F par bassin-versant sous forme de score.

Nom	Surface (km^2)	Nombre d'échec		Nombre de succès		Pop.	PM	SC
		B	F	B	F			
La Siagne	492	10	9	2	3	12	0.25	0.67
L'Aude	1828	14	12	10	12	24	0.50	0.83
L'Orbiel	242	2	2	4	4	6	0.67	1.00
L'Orbieu	589	2	2	6	6	8	0.75	1.00
Le Gapeau	535	7	3	7	11	14	0.79	0.64
La Cèze	1112	4	3	10	11	14	0.79	0.91
L'Orb	1336	8	5	16	19	24	0.79	0.84
Le Vidourle	503	4	2	6	8	10	0.80	0.75
Le Doux	621	1	1	5	5	6	0.83	1.00
La Giscle	201	1	1	5	5	6	0.83	1.00
L'Argens	2512	12	3	10	19	22	0.86	0.53
L'Eyrieux	649	1	1	7	7	8	0.88	1.00
Le Fresquel	935	2	1	8	9	10	0.90	0.89
Le Loing	2302	7	1	11	17	18	0.94	0.65
L'Ardèche	2264	5	1	15	19	20	0.95	0.79
L'Hérault	2208	3	1	23	25	26	0.96	0.92
Le Gardon	1093	5	0	17	22	22	1.00	0.77
La Cance	381	1	0	7	8	8	1.00	0.88
La Cesse	251	0	0	4	4	4	1.00	1.00
Le Vistre	496	0	0	4	4	4	1.00	1.00
Le Loup	289	0	0	4	4	4	1.00	1.00
L'Estéron	442	0	0	4	4	4	1.00	1.00
Le Lez	168	0	0	8	8	8	1.00	1.00
TOTAL		89	48	193	234	282		

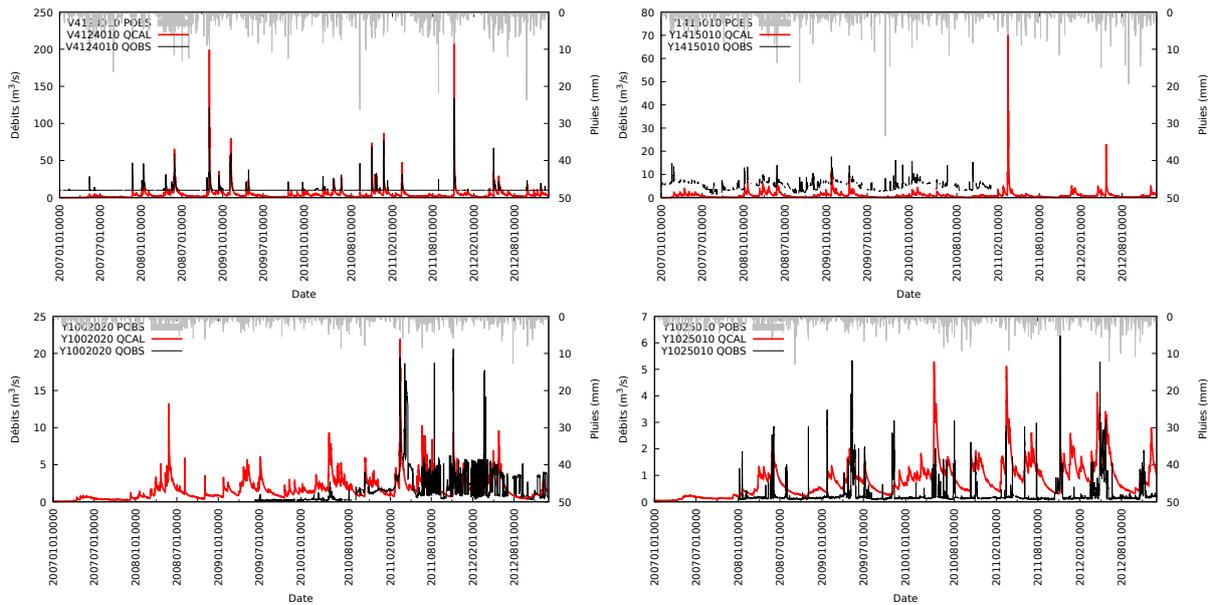


FIGURE 6.12 – Débits observés (noirs) et simulés (rouges) aux stations V4124010 (l’Eyrieux), Y1415010 (l’Orbiel), Y1002020 (l’Aude) et Y1025010 (l’Aude) sur la période P1. Ces hydrogrammes permettent de visualiser les problèmes de modélisation ou de mesures rencontrés.

Conclusions : Ces expériences ont permis d’identifier la proportion d’exutoires dont les données de débits ne sont pas modélisables (48 chroniques, ce qui représente environ 20% des chroniques observées utilisées en validation) et pour lesquels le calage distribué des paramètres ne permet pas de représenter correctement les processus hydrologiques (41 chroniques qui représentent à leurs tours environ 17% des chroniques utilisées en validation). L’amélioration de la technique de calage distribué des paramètres (apport de contraintes supplémentaires, modification du critère d’optimisation) pourrait permettre d’améliorer significativement la modélisation de ces chroniques. Les bassins-versants dont les paramètres semblent le moins représentatif des caractéristiques hydrologiques sont le Gapeau, le Vidourle, l’Argens, le Loing, l’Ardèche et le Gardon. Pour les autres bassins-versants, le modèle permet de représenter correctement le fonctionnement hydrologique aux exutoires où les données le permettent ($0.8 < SC \leq 1$). Les incertitudes sur les données peuvent fortement influencer le calage. Les simulations réalisées sont toutes impactées de la même manière. De ce fait les problèmes liés aux données ne devraient pas fausser les conclusions.

6.2.4 Influence de l'ébauche et de la régularisation sur les performances prédictives du modèle

La qualité du calage des paramètres est influencée par le choix du vecteur d'ébauche et des méthodes de régularisation mises en œuvre. Dans cette partie, des expériences de calage (impliquant la variante 2 du modèle) ont été réalisées dans le but de montrer leurs effets. Il s'agit des expériences **B**, **C**, **D** et **E** dont les caractéristiques sont rappelées dans le tableau 6.4.

TABLE 6.4 – Tableau récapitulatif des expériences de calage-validation B, C, D et E.

Nom	Type de régularisation		Ébauche calculée depuis un calage uniforme
	Tickonov-itérative	Tickonov-Lcurve	
B	×		×
C			
D			×
E		×	×

Les graphiques 6.13 et 6.14 représentent les performances prédictives spatiales et spatio-temporelle des expériences de calage **A** (référence), **B**, **C**, **D** et **E** sous forme de distribution statistique. Chaque courbe de distribution contient 118×2 points (exutoires de validation, période P1 et P2). Les performances sont analysées à l'aide du critère de NSE (utilisé en calage) et du critère KGE.

La comparaison des résultats de l'expérience **C** avec les expériences **B**, **D** et **E** montre que l'ébauche estimée à partir du calage uniforme (expérience **A**) permet d'améliorer significativement les performances obtenues avec un calage distribué.

La comparaison des résultats des expériences **D** avec les expériences **B** et **E** indique que les méthodes de régularisation mises en œuvre ne permettent pas d'améliorer les performances prédictives du modèle. Ces formes de régularisation (type Tikhonov) supposent que les paramètres soient lissés spatialement. Cette contrainte ne semble pas appropriée aux paramètres hydrologiques du modèle. Par exemple, l'absence de fonction de régularisation (expérience **D**) améliore nettement les prédictions des débits à l'exutoire aval du bassin-versant de la Siagne (voir figure 6.15) bien que les données hydrométriques semblent douteuses pour l'événement considéré.

La comparaison des expériences **B** et **E** montrent que la méthode de régularisation itéra-

tive est meilleure que la méthode de la L-Curve. Ce meilleur résultat peut être attribué à l'implémentation de la méthode qui permet de prendre en compte ces contraintes dès les premières itérations.

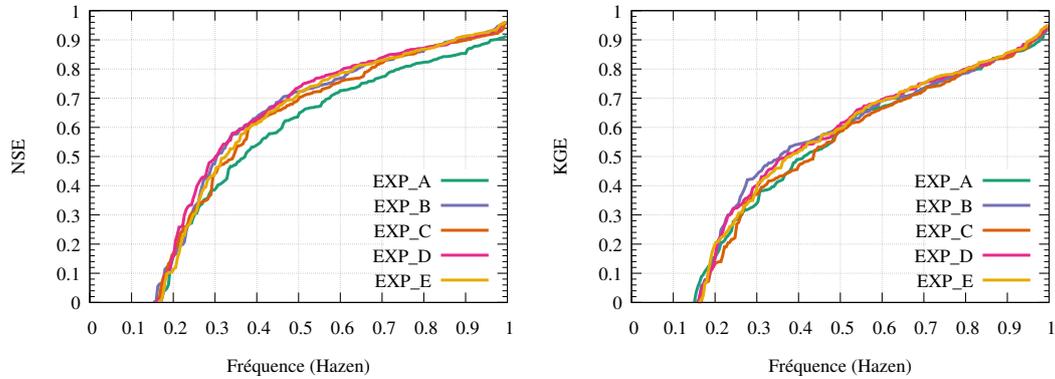


FIGURE 6.13 – Performances en **validation spatiale**, 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage **distribué** des paramètres, expériences **A**, **B**, **C**, **D** et **E**. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

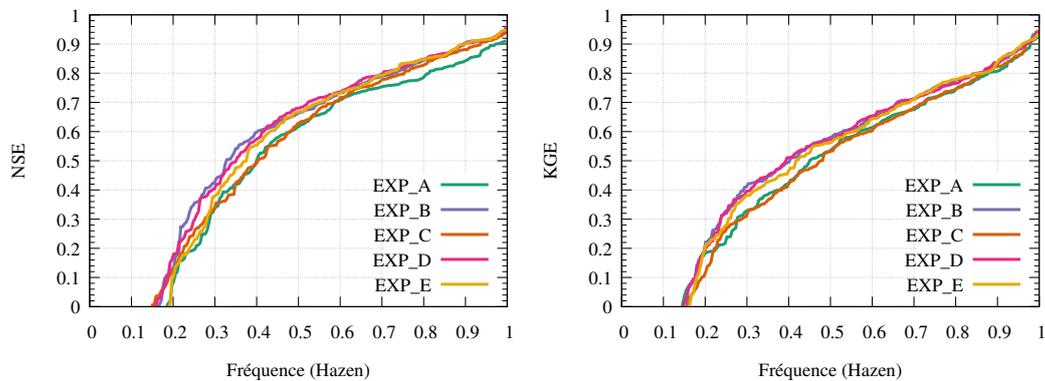


FIGURE 6.14 – Performances en **validation spatio-temporelle**, 118 exutoires de validation sur les périodes de calage P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage **distribué** des paramètres, expériences **A**, **B**, **C**, **D** et **E**. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

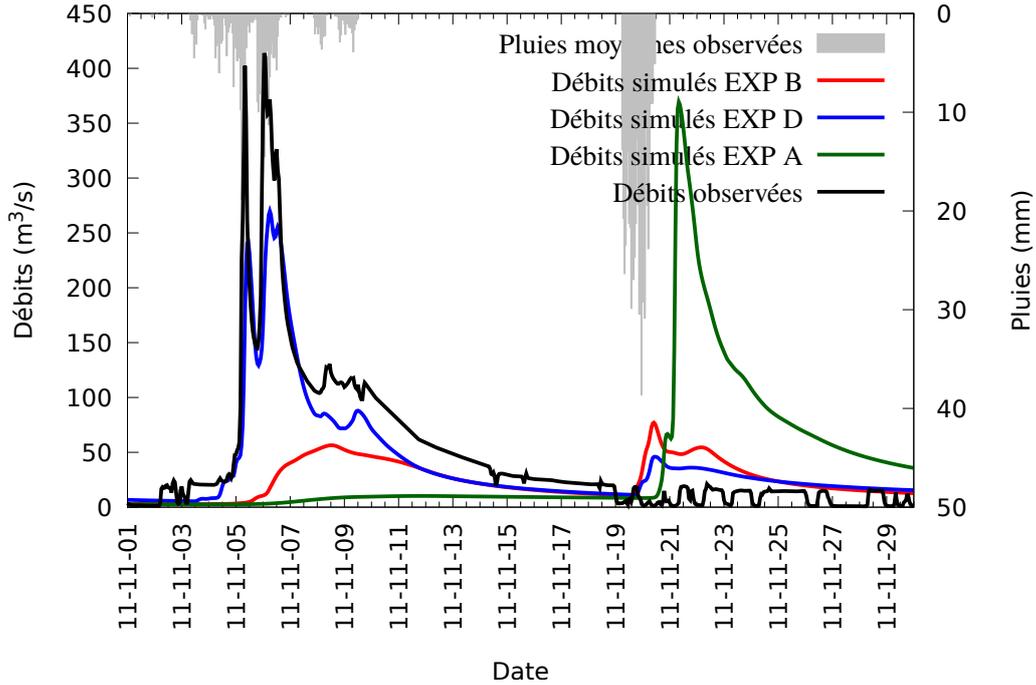


FIGURE 6.15 – Débits observés (noir) et simulés (variantes 2, expérience A, B et D) à la station de calage "la Siagne à Pégomas" (Y5534030) pour des évènements pluvieux importants entre le 1 et le 30 Novembre 2011. Les pluies moyennes sur le bassin-versant sont représentées par l'histogramme (gris).

Conclusions : La détermination du vecteur d'ébauche à partir d'un calage uniforme par un algorithme de recherche global conditionne le succès du calage variationnel et les bonnes performances prédictives du modèle utilisant un jeu de paramètres distribués. L'introduction de fonctions de régularisation de type Tikhonov ne permet pas d'améliorer significativement les performances prédictives du modèle. La structure de cette contrainte de régularisation ne semble pas appropriée à l'identification des paramètres du modèle hydrologique.

6.2.5 Analyse des jeux de paramètres distribués

Dans cette partie, les jeux de paramètres obtenus sur les périodes P1 et P2 et résultant de l'expérience de calage **B** sont analysés. Les jeux de paramètres sont cartographiés pour chaque période et chaque bassin-versant. La variabilité spatiale de ces jeux de paramètres est interprétée.

Les graphiques 6.16, 6.17 et 6.18 présentent les cartes des paramètres (respectivement c_p, c_{tr} et ξ) obtenues après calage distribué pour les périodes P1 et P2 et pour les 23 bassins-versants, expérience **B**.

Ces cartes mettent en évidence la forte variabilité spatiale des paramètres. Les paramètres varient significativement autour de la valeur moyenne. L'écart-type peut dépasser 50% de la valeur moyenne des paramètres ($\frac{\sigma}{\mu}$). Cette variabilité est obtenue quelle que soit la période de calage, P1 ou P2. On observe aussi que la variabilité des paramètres est aussi forte en amont qu'en aval des bassins-versants. Ce résultat indique que la fonction objectif, dépendante des données de débits aux exutoires situés en aval des bassins-versants, est sensible aux paramètres des mailles situées à l'amont des bassins-versants (quelles que soient leurs surfaces et formes). Bien que la variabilité spatiale des paramètres améliore globalement la prédiction des débits aux exutoires de validation, celle-ci ne garantit pas des performances prédictives satisfaisantes pour l'ensemble des exutoires de validation (on a vu précédemment que les simulations des débits au niveau de 41 exutoires utilisés en validation pouvaient être significativement améliorées).

Les cartes du paramètre de routage (figure 6.18) mettent en évidence une sensibilité marquée de la fonction objectif par rapport aux mailles situées sur le réseau hydrographique (regroupant les débits accumulés les plus forts). En effet, lors du calage les valeurs du paramètre de routage sont exclusivement modifiées sur les mailles situées sur le réseau. A l'inverse, la valeur de l'ébauche est généralement conservée sur les mailles "versant". Pour la plupart des bassins-versants, la valeur moyenne du paramètre de routage est aux alentours de $\xi = 10$. Ce coefficient correspond à des vitesses de drainage globalement rapides (inférieures à 0.6 m.s^{-1} en étiage, 1.5 m.s^{-1} pour les débits médians, supérieures à 10 m.s^{-1} en crue). Ces valeurs semblent cohérentes puisque la plupart des bassins-versants étudiés se situent dans des zones montagneuses (excepté le bassin-versant du Loing dont le coefficient moyen de routage est plus faible et vaut $\xi \approx 2$). Cependant, des valeurs de coefficient de drainage plus faibles sont parfois observées sur le réseau hydrographique. Ces valeurs peuvent paraître étonnantes. Les causes sont nombreuses : le modèle de routage est conceptuel ; les processus de routage modélisés ne correspondent pas exclusivement à de l'écoulement en rivière (écoulement souterrain, de ruissellement de surface, piézométrique des nappes) ; le calage entraîne des phénomènes de compensation des temps de propagation entre les différents drains des bassins-versants ; les comportements hydrauliques des rivières sont variables (largeurs, pentes, méandres, seuils , barrages...).

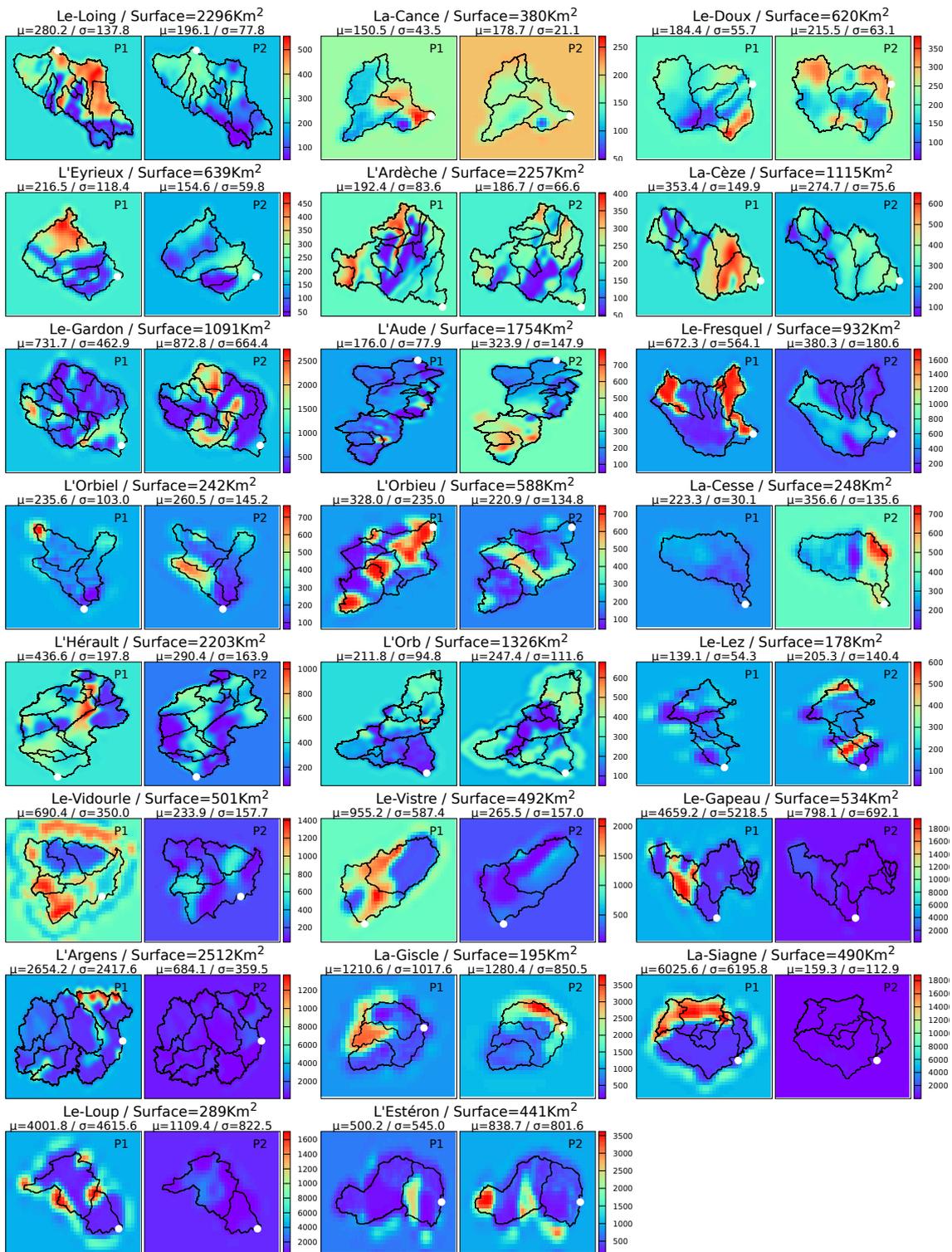


FIGURE 6.16 – Capacités des réservoirs de production c_p obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).

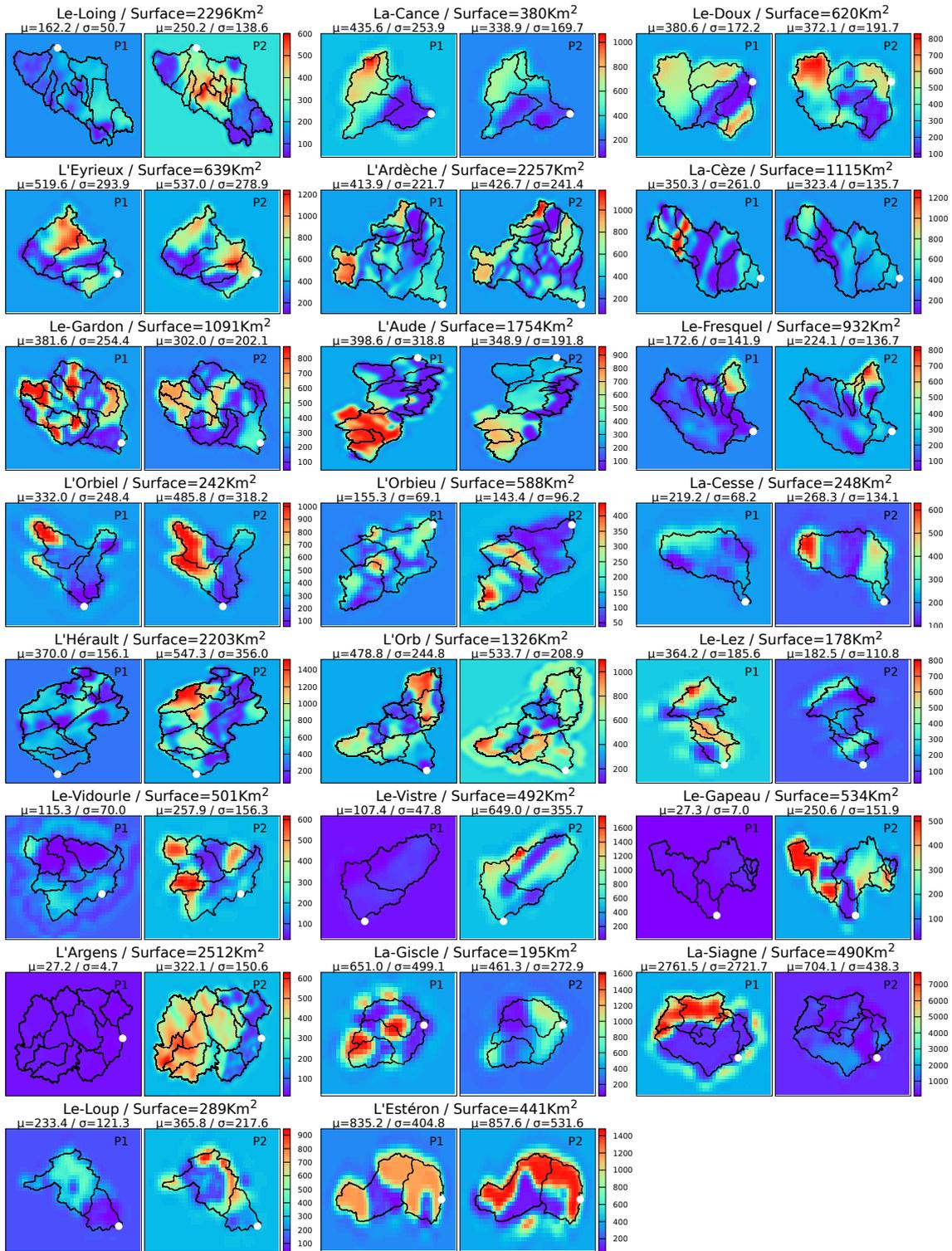


FIGURE 6.17 – Capacités des réservoirs de transfert c_{tr} obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).

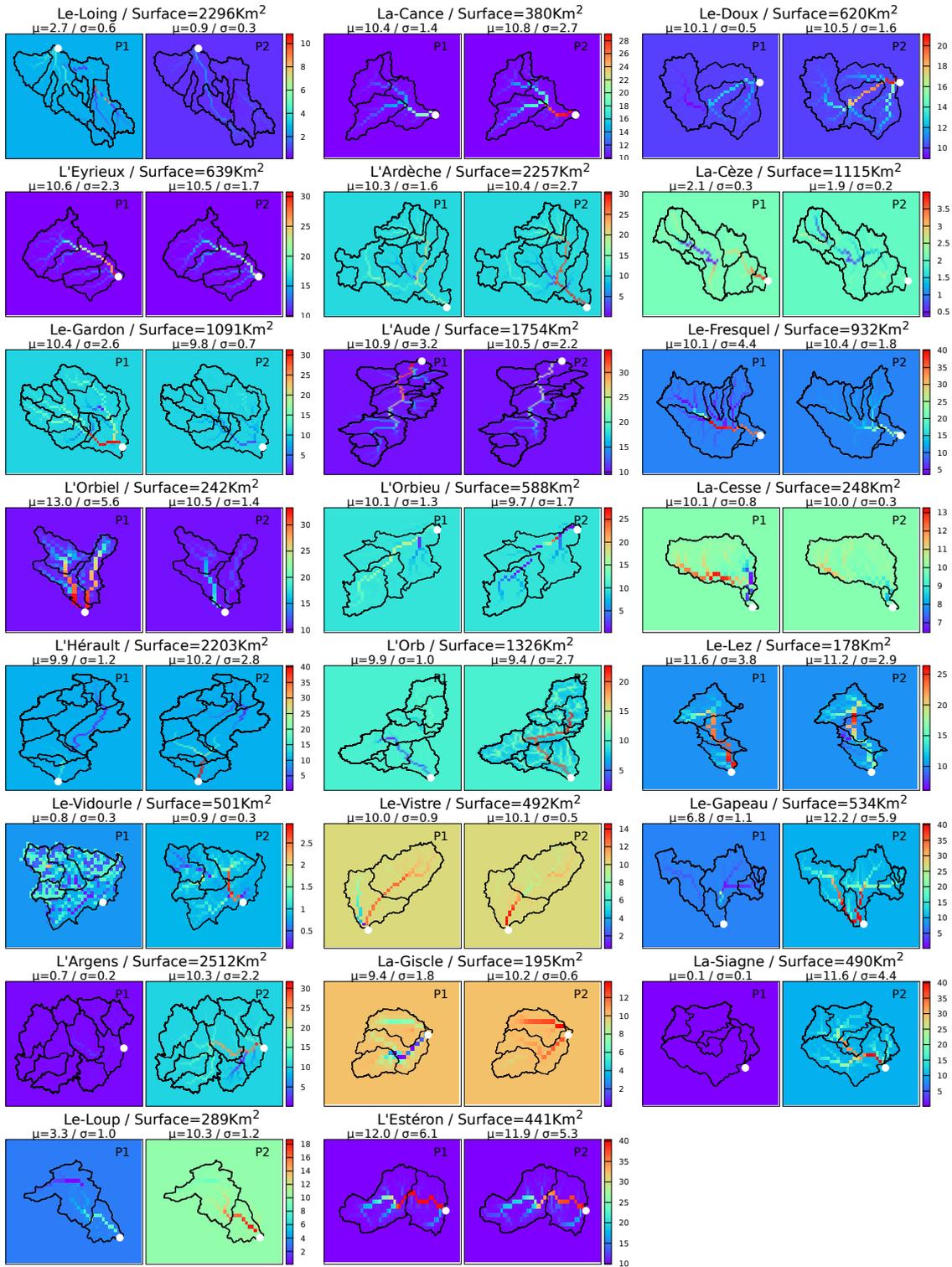


FIGURE 6.18 – Paramètre de routage ξ obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire des bassins-versants (contours des bassins-versants délimités en noir).

Conclusions : Les données de débits à l’aval des bassins-versants modélisés suffisent pour caler et créer une variabilité spatiale des paramètres du modèle. Pour tous ces bassins-versants les paramètres calés sont très variables spatialement. Cependant, cette variabilité ne garantit pas des performances satisfaisantes pour l’ensemble des exutoires de validation situés à l’intérieur des bassins-versants modélisés. Les jeux de paramètres calés sont très différents entre les deux périodes analysées. Il est difficile d’établir des liens entre les paramètres de production c_p et de transfert c_{tr} et des descripteurs environnementaux. Quant aux paramètres de routage ξ , ils semblent être corrélés au réseau de drainage. Les valeurs de ce paramètre indiquent que les vitesses d’écoulement modélisées sont du même ordre de grandeur que celles observées en rivière.

6.2.6 Étude de la robustesse des jeux de paramètres résultant de l’expérience **B** pour les périodes **P1** et **P2**

Dans cette partie les différences et similitudes des jeux de paramètres obtenus sur les périodes de calage **P1** et **P2** sont comparées. Il s’agit ici d’évaluer la robustesse (stabilité) de ces jeux de paramètres. Celle-ci est discutée au regard des résultats de l’expérience de calage **B**.

Les paramètres distribués obtenus après calage sur les périodes **P1** et **P2** (expérience **B**) présentent des écarts importants (voir figures 6.16, 6.17 et 6.18). Les causes de ces écarts peuvent être attribuées à : 1) l’utilisation de données hydrologiques issues de climatologies différentes (deux périodes temporelles courtes et distinctes) ; 2) des problèmes d’identification où d’équifinalité mathématiques (Beven, 1993, 2001). L’équifinalité d’une solution provient du manque de données, de la dépendance au vecteur d’ébauche, des incertitudes de modélisation et du choix de la fonction objectif.

Le graphique 6.19 permet de comparer les paramètres obtenus après calage sur les périodes **P1** et **P2**. Pour chacune des deux périodes et chaque bassin-versant, les moyennes spatiales des paramètres sont calculées. Les paramètres moyens issus de l’expérience **A** (utilisés comme valeur d’ébauche) et les paramètres moyens issus de l’expérience **B** sont comparés.

Pour les deux tiers des bassins-versants modélisés, les paramètres uniformes (issus du calage de l’expérience **A**) sont globalement stables entre les deux périodes. Au contraire, ces paramètres sont instables pour les bassins-versants de la Siagne, de l’Argens, du Gapeau, du Loup, du Vidourle et du Vistre. L’écart entre les cercles violets et les croix rouges montrent l’effet du calage distribué sur la stabilité des valeurs moyennes des paramètres.

Le calage distribué des paramètres ne modifie pas beaucoup les valeurs moyennes (les cercles violets et croix rouges sont proches). Les jeux de paramètres, issus de l'expérience de calage **B** pour les périodes P1 et P2, diffèrent donc en termes de variabilité spatiale. Au regard de ce résultat, le choix des paramètres uniformes d'ébauche (résultat du calage de l'expérience **A**) n'explique pas les différences constatés en terme de variabilité spatiale des paramètres distribués (résultat de l'expérience **B**).

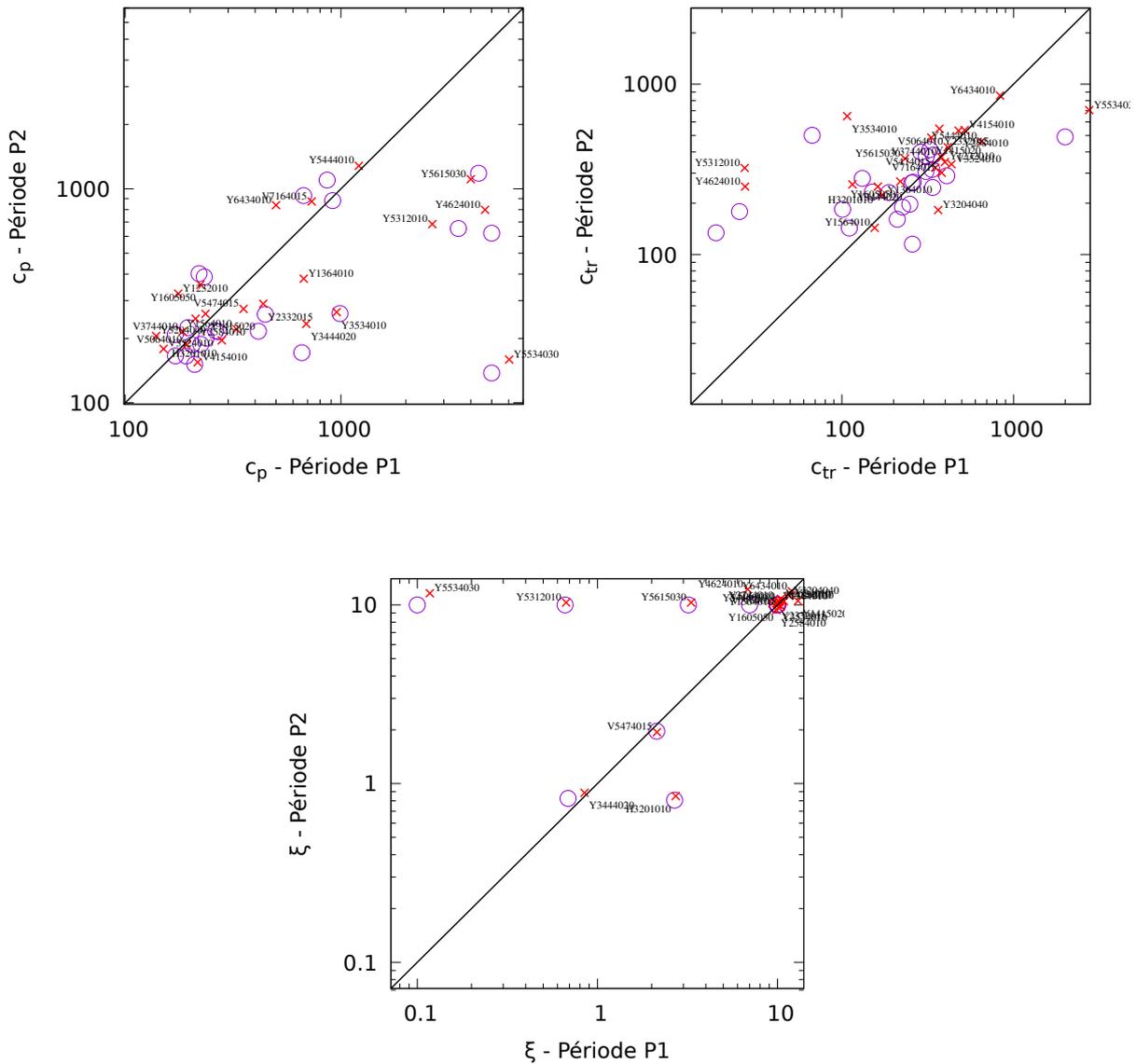


FIGURE 6.19 – Comparaison des valeurs moyennes des paramètres distribués (c_p , c_{tr} , ξ) obtenues après calage sur les périodes P1 (abscisse) et P2 (ordonnée) pour chacun des 23 bassins-versants. Les valeurs moyennes de l'ébauche (issu d'un calage uniforme des paramètres) sont représentées par les cercles violets. Les valeurs moyennes obtenues après calage distribué sont représentées par des croix rouges. Les moyennes ont été calculées sur l'aire des bassins-versants.

Le critère présenté à l'équation 6.7 permet de comparer et de cartographier les différences entre les jeux de paramètres calés sur les périodes P1 et P2. Soit X le vecteur de contrôle, le critère θ est calculé par :

$$\theta = abs \left(\frac{(X^{P1} - \mu_X^{P1}) - (X^{P2} - \mu_X^{P2})}{\mu_X^{P1,P2}} \right) \times sgn \left(\frac{(X^{P1} - \mu_X^{P1})}{(X^{P2} - \mu_X^{P2})} \right) \quad (6.7)$$

Ce critère vaut 0 si les paramètres sont identiques entre les deux périodes. Plus ce critère s'éloigne de sa valeur optimale $\theta = 0$, plus les écarts entre les paramètres calés sur les périodes P1 et P2 sont importants. Pour chaque bassin-versant, ce critère est normé par la valeur moyenne des paramètres toutes périodes confondues $\mu_X^{P1,P2}$. Ses valeurs sont donc comparables entre les bassins-versants modélisés.

Ce critère est positif si les variations autour des moyennes des paramètres μ_X^{P1} et μ_X^{P2} concordent entre les périodes P1 et P2 : c'est à dire ($X^{P1} > \mu_X^{P1}, X^{P2} > \mu_X^{P2}$) ou ($X^{P1} < \mu_X^{P1}, X^{P2} < \mu_X^{P2}$).

Ce critère est négatif si les variations autour de la moyenne des paramètres μ_X^{P1} et μ_X^{P2} sont antagonistes entre les deux périodes : c'est à dire si ($X^{P1} > \mu_X^{P1}, X^{P2} < \mu_X^{P2}$) ou ($X^{P1} < \mu_X^{P1}, X^{P2} > \mu_X^{P2}$).

Ce critère est calculé et cartographié pour chacun des bassins-versants sur les graphiques 6.20, 6.21 et 6.22. Ces cartes mettent en évidence les différences significatives existantes entre les jeux de paramètres calés sur les périodes P1 et P2. En effet, pour de nombreuses régions, les paramètres du modèle présentent des dynamiques hydrologiques contradictoires entre les deux périodes P1 et P2 ($\theta < 0$). D'une part, le choix de l'ébauche n'explique pas ces différences. De plus les caractéristiques physiques des bassins-versants n'ont pas pu changer drastiquement sur les deux périodes successives de 5 ans. Ces différences en termes de variabilité spatiales s'expliquent probablement par le choix des données hydrologiques spatialisées qui sont utilisées pour le calage (pluies).

Les valeurs négatives de θ (couleur rouge) sur les graphiques 6.20, 6.21 et 6.22 indiquent donc les zones où les paramètres du modèle sont très instables entre les deux périodes. Pour ces zones les paramètres sont sensibles aux jeux de données utilisés en calage (événements hydrologiques différents). Les valeurs nulles et positives de θ (couleurs verte et bleue) indiquent les zones où les paramètres sont plus stables vis-à-vis de la période de calage P1 ou P2. Dans ce cas les écarts constatés semblent davantage liés à des différences entre la climatologie moyenne des deux périodes.

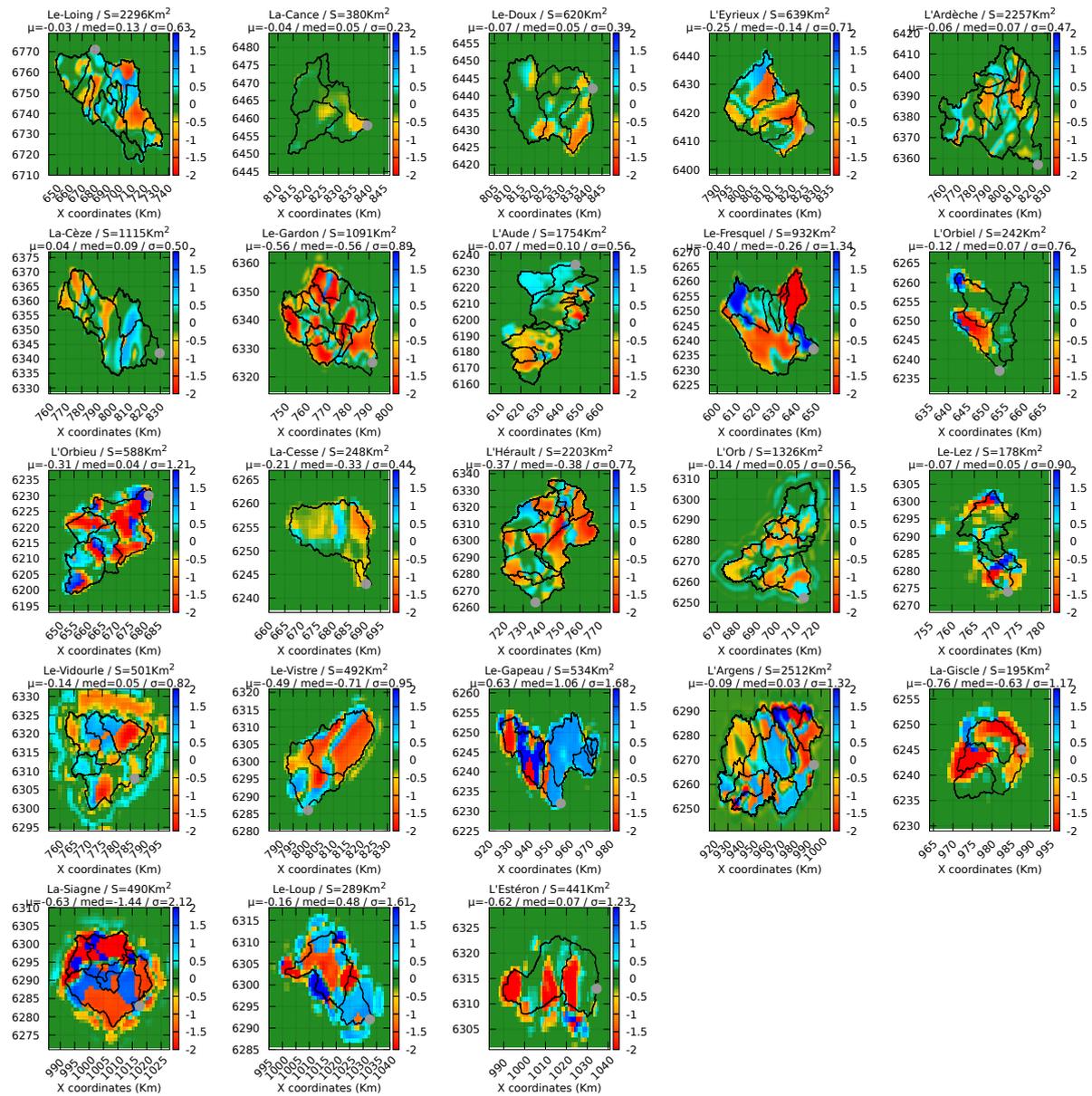


FIGURE 6.20 – Comparaison des capacités des réservoirs de production c_p obtenues après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.

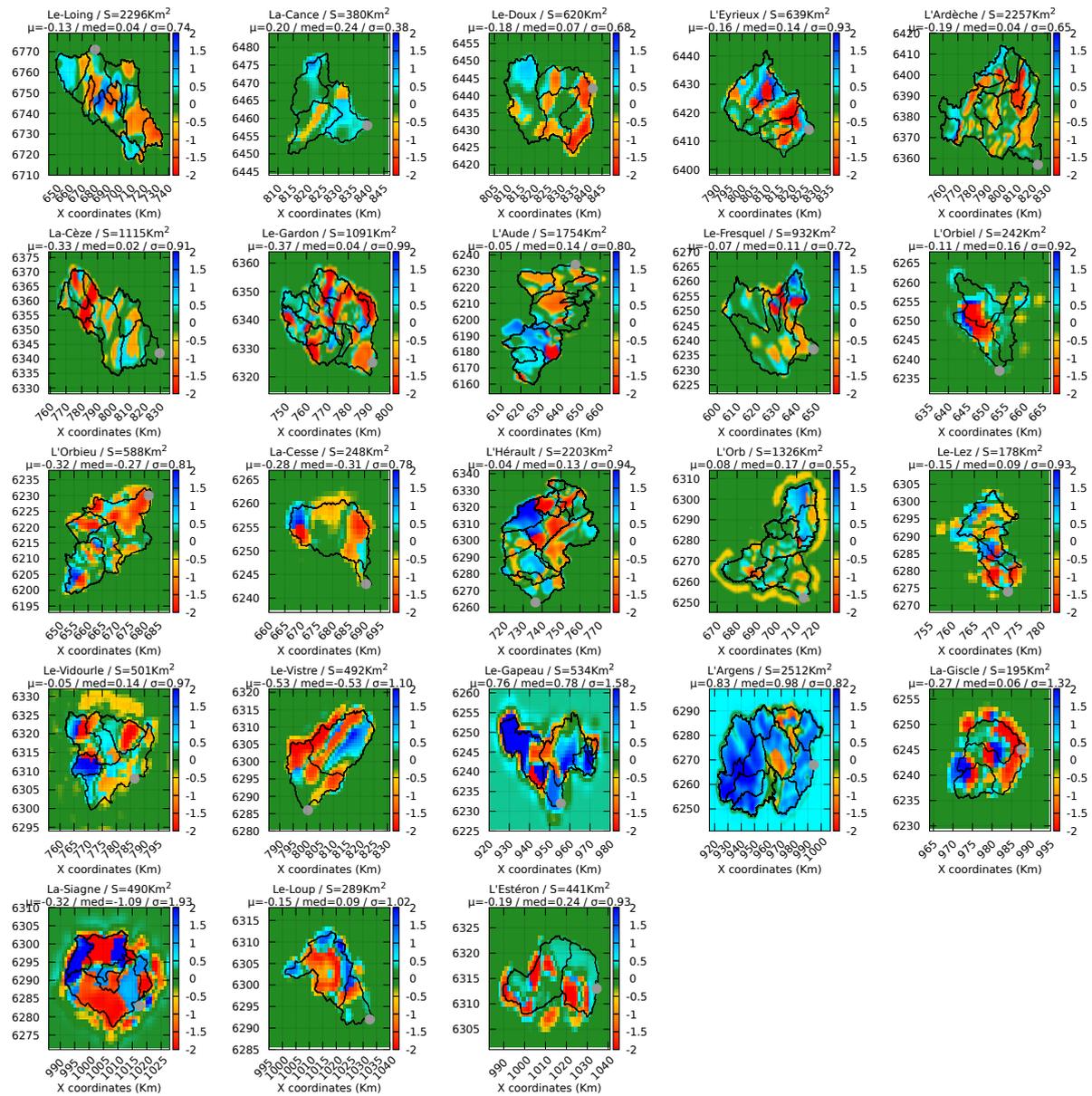


FIGURE 6.21 – Comparaison des capacités des réservoirs de transfert c_{tr} obtenues après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.

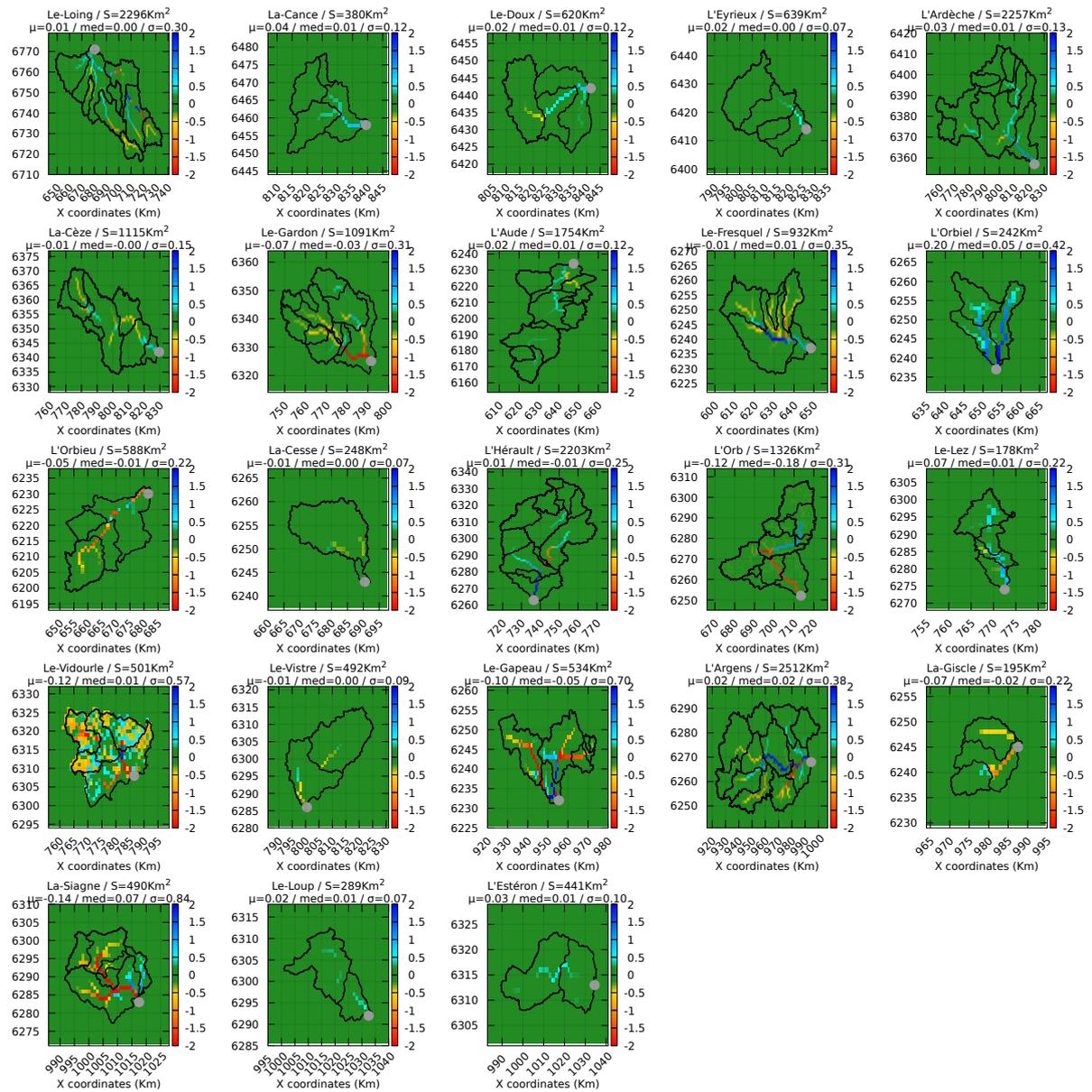


FIGURE 6.22 – Comparaison des paramètres de routage ξ obtenus après calage pour les périodes P1 et P2 à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.

Conclusions : Bien que des écarts importants soit constatés entre les deux périodes d'analyse, les paramètres de certains bassins-versants restent relativement cohérents selon les périodes ($\bar{\theta} \approx 0, \bar{\theta} \geq 0$). C'est le cas des bassins-versants de la Cance, du Doux, de l'Ardèche, de la Cèze, de l'Orbiel et du Gapeau. Ce résultat est encourageant et semble montrer que la variabilité spatiale des paramètres du modèle GRD pourrait être mieux contraintes afin de limiter sa sensibilité aux périodes d'analyses. Pour cela, une période d'analyse plus longue pourrait être considérée et des descripteurs environnementaux pourrait permettre de contraindre la variabilité spatiale des paramètres.

En dépit des différences constatées sur les jeux de paramètres, nous remarquons que les performances prédictives spatio-temporelles du modèle restent bonnes (graphiques 6.6) .

6.2.7 Analyse de la robustesse des paramètres issus des expériences de calage **B, C, D, E et F (variante 2)**

Dans cette partie l'étude de la robustesse des jeux de paramètre obtenus pour les périodes P1 et P2 est élargie aux jeux de paramètres issues des expériences de calage **B, C, D, E et F**.

La robustesse des paramètres du modèle est analysée en fonction des méthodologies de calage mises en œuvre pour les expériences **B, C, D, E et F**. Le graphique 6.23 permet de comparer la stabilité des paramètres calés entre les deux périodes temporelles P1 et P2 obtenues sous plusieurs configurations : utilisation des fonctions de régularisation et identification préalable ou non du vecteur d'ébauche par un calage uniforme (voir la description des expériences dans le tableau 6.2). Quatre critères sont utilisés et sont représentés sous forme de distribution :

- La distribution des coefficients de corrélation Pearson permet de mesurer l'existence d'une relation entre les paramètres des deux périodes.
- La distribution des rapports calculés entre les valeurs moyennes des paramètres obtenues sur les deux périodes permet de quantifier le biais.
- La distribution des moyennes du coefficient $\bar{\theta}$ (critère présenté à l'équation 6.7) permet de quantifier et d'indiquer la direction des écarts.
- La distribution des moyennes des différences ϕ (différences calculées à partir des paramètres centrés-réduits sur les périodes P1 et P2) permet de mesurer l'écart absolu.

La stabilité des paramètres entre les périodes P1 et P2 pour les expériences **B, D et E**

est comparable. On remarque que l'expérience **C** permet d'obtenir un jeu de paramètres plus stable. Le choix des paramètres d'ébauche issus des résultats du calage uniforme de l'expérience **A** pour chaque période affecte négativement la stabilité des paramètres. Ce résultat peut paraître surprenant puisque la prédétermination de l'ébauche par un calage uniforme permet d'améliorer significativement les performances prédictives spatio-temporelles du modèle. L'expérience **F** permet aussi d'obtenir des paramètres plus stables. Ce résultat était attendu puisque davantage de données de débits sont utilisés pour le calage. Remarquons que les paramètres restent relativement peu robustes et instables (coefficient de Pearson inférieur à 0.6) malgré l'utilisation des données de débits aux 141 exutoires.

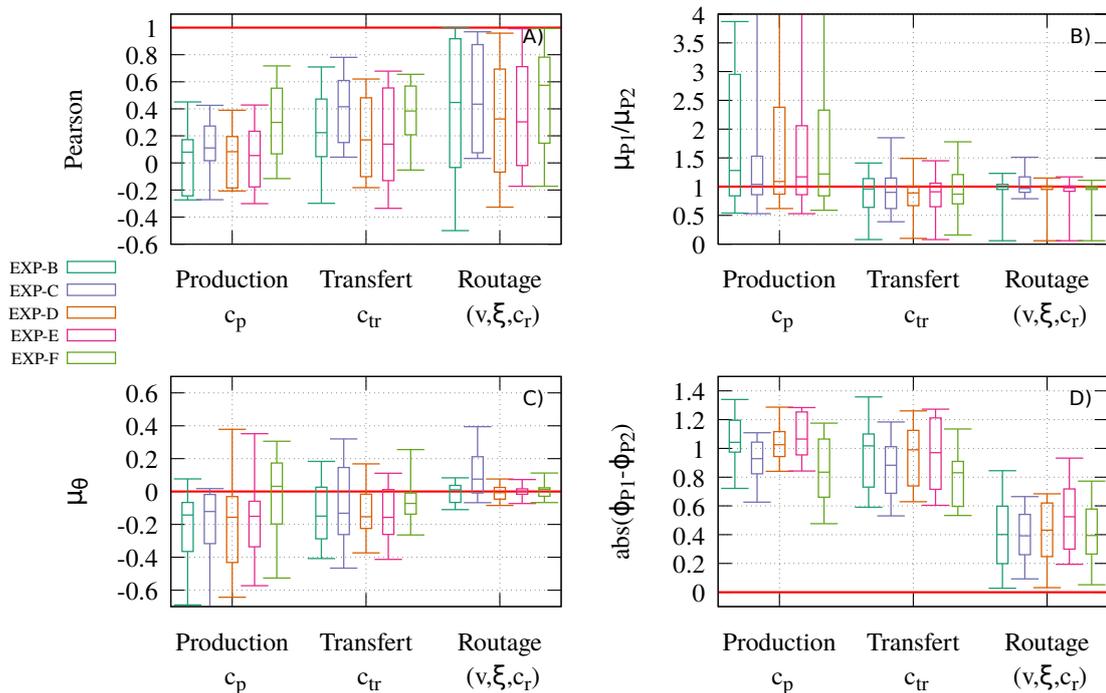


FIGURE 6.23 – Box-plot des distributions des critères d'adéquation entre les paramètres distribués période P1 et période P2 pour chaque bassin-versant. La stabilité des paramètres obtenus pour les expériences B,C,D,E,F est comparée. Graphique A) coefficient de Pearson. Graphique B) Rapport des valeurs moyennes. Graphique C) Moyenne de la variable θ (équation 6.7). Graphique D) Différence entre les valeurs des paramètres centrées-réduites.

Conclusions : Les jeux de paramètres distribués calés sur des périodes d'analyse différentes sont relativement instables. Des écarts importants sont constatés en termes de variabilité spatiale. Ces écarts semblent en partie provoqués par le choix des données utilisées pour le calage. Les calages sont réalisés sur seulement 5 ans d'observation, ce qui rend la robustesse des paramètres fortement tributaire de l'échantillonnage. Ces problèmes de robustesse se retrouvent aussi dans les approches de modélisation globales (Brigode et al., 2013; Dakhlaoui et al., 2017) et ne semblent pas liés à la procédure de calage distribué (voir les résultats du calage uniforme sur la figure 6.19). Le paramètre de production est le paramètre le plus instable. L'utilisation de fonction de régularisation a peu d'effet sur la stabilité des jeux de paramètres.

6.3 Conclusions

Dans ce chapitre des expériences de calage-validation croisées ont été réalisées sur un échantillon de 23 groupes de bassins-versants. Les performances prédictives du modèle hydrologique distribué GRD ont été évaluées en validation spatiale et spatio-temporelle sur 118 exutoires amont.

Les résultats des simulations effectuées avec les trois variantes du modèle n'ont pas montré de différences significatives en termes de performances. L'opérateur hydraulique ne semble pas être le composant essentiel permettant d'améliorer la prédiction des débits. Néanmoins, ce résultat devrait être vérifié. Un travail supplémentaire doit encore être mené pour départager ou améliorer ces variantes, mais cet objectif sort des cadres fixés dans ces travaux de thèse.

Ces expériences ont permis de démontrer que l'utilisation de jeux de paramètres distribués améliore significativement les performances prédictives du modèle hydrologique aux exutoires de validation ainsi que sur la période temporelle de validation. Il est donc possible et bénéfique d'estimer la variabilité spatiale des paramètres d'un modèle hydrologique conceptuel distribué à l'aide d'un algorithme d'estimation variationnelle utilisant le modèle adjoint.

Une marge de progression importante existe afin d'améliorer la représentation spatiale des caractéristiques hydrologiques des bassins-versants. Les paramètres distribués ont été calés à partir des données de débits disponibles aux exutoires avals des bassins-versants sur deux périodes temporelles distinctes. Ce calage a permis d'introduire une variabilité spatiale de chaque paramètre à l'intérieur des bassins-versants. Cependant, on a constaté que cette variabilité est très différente selon la période d'analyse. Elle dépend du choix

du jeu de données et de la longueur de la période de calage. Cette variabilité ne garantit pas toujours une bonne prédiction des débits aux exutoires de validation. En effet, les débits observés pour environ 17% des exutoires de validation ne sont pas bien simulés par le modèle utilisant des jeux de paramètres distribués.

La mise en œuvre de la méthode de calage variationnelle nécessite de pré-déterminer une valeur moyenne uniforme des paramètres pour chaque bassin-versant (ébauche). Celle-ci doit être identifiée à l'aide d'un algorithme de recherche global. Cette ébauche garantit le succès de la méthode variationnelle. Une méthodologie de calage en deux étapes est donc nécessaire pour réaliser un calage distribué des paramètres à savoir un calage uniforme à l'aide d'un algorithme de recherche global suivi d'un calage distribué à l'aide d'un algorithme d'estimation variationnelle.

Les méthodes de régularisation testées (Tikhonov - L-curve, Tikhonov - itérative) ne permettent pas d'améliorer la corrélation des jeux de paramètres trouvés à partir de données de périodes temporelles différentes. Ces méthodes ne permettent pas non plus d'améliorer significativement les performances prédictives du modèle. Pourtant, nous avons montré l'intérêt de la régularisation sur des cas tests au chapitre 5. C'est donc que la structure des fonctions de régularisation testées ne semble pas appropriée à l'identification des paramètres du modèle hydrologique.

Par ailleurs, il faut noter le caractère "extrême" de ces expériences. En effet, le calage des paramètres a été réalisé sur des fenêtres temporelles réduites (deux fois cinq ans) et en utilisant seulement les données de débits aux stations de jaugeage aval. Le calage du modèle nécessitera sûrement de prendre en compte les observations de débits au niveau de plusieurs stations de jaugeage par bassin-versant et de rallonger la fenêtre temporelle d'analyse.

Afin d'améliorer le calage distribué des paramètres (adéquation des jeux de paramètres trouvés indépendamment de la période d'analyse et améliorations des performances prédictives spatio-temporelles), plusieurs pistes sont suggérées : l'estimation d'une ébauche non-uniforme basée sur des descripteurs physiques caractérisant les bassins-versants ; l'utilisation d'une fonction objectif limitant le poids des erreurs sur les débits extrêmes. Ces deux pistes sont explorées dans le chapitre suivant.

Chapitre 7

Tentatives d'amélioration de la robustesse du calage des paramètres

Le calage distribué des paramètres est perfectible. En effet, les résultats des précédentes expériences (chapitre 6) ont montré qu'environ 17% des chroniques de débits observées ne sont pas correctement simulées par le modèle en validation spatiale et spatio-temporelle. Les précédentes expériences ont aussi montré que les jeux de paramètres distribués sont très dépendants des données observées ayant servi pour leurs calages (différences et incohérences entre les jeux de paramètres calés sur les périodes temporelles P1 et P2). Ces problèmes de modélisation sont causés par des jeux de paramètres distribués peu robustes et peu représentatifs des caractéristiques hydrologiques des bassins-versants concernés. Ce manque de robustesse en calage des modèles conceptuels est un problème connu. Par exemple, [Dakhlaoui et al. \(2017\)](#) se sont aperçus que certains paramètres des modèles conceptuels sont très dépendants des conditions climatiques. [Brigode et al. \(2013\)](#) ont montré que le manque de robustesse du calage des paramètres des modèles conceptuels rend leurs utilisations délicates pour prévoir les conditions hydrologiques sous différents scénarios climatiques. Ces derniers recommandent par ailleurs d'accentuer les recherches visant à améliorer la structure des modèles et des méthodes de calage.

Dans ce chapitre des expériences sont mises en œuvre pour tenter d'améliorer la méthodologie de calage permettant d'estimer les paramètres distribués du modèle GRD. Les améliorations recherchées sont : d'une part accroître la stabilité et la robustesse des jeux de paramètres calés sur les deux périodes d'analyse ; d'autre part l'amélioration des performances prédictives spatiales et spatio-temporelles du modèle aux exutoires pour lesquels les performances sont médiocres. Pour cela deux stratégies sont mises en œuvre :

1. Contraindre la variabilité spatiale des paramètres selon des descripteurs choisis à priori. Ces descripteurs sont choisis à l'échelle de l'ensemble du territoire d'étude. Ils décrivent localement les caractéristiques de la maille considérée. Ces descrip-

teurs peuvent être de différentes natures : climatiques, géologiques, pédologiques, topographiques ou liés à l'occupation des sols. Les sources de données correspondantes ont été présentées au chapitre 2. Plusieurs combinaisons de descripteurs sont testées pour contraindre l'identification des réservoirs de production, les réservoirs de transfert et les paramètres hydrauliques du modèle GRD.

2. Utiliser une fonction objectif plus robuste. Cette fonction objectif a été présentée en détail au chapitre 4. Celle-ci devrait permettre de réduire la sensibilité du calage aux événements extrêmes ainsi qu'aux variables observées ou modélisées aberrantes. Ainsi, l'objectif est d'accroître la robustesse des jeux de paramètres estimés à partir de données provenant de périodes différentes.

Ce chapitre se découpe en deux parties, chacune abordant l'un des points présentés ci-dessus.

7.1 Contraindre la variabilité spatiale à l'aide de descripteurs

7.1.1 Principes et objectifs

Les expériences de calage précédentes (chapitre 6) ont permis d'estimer des paramètres de production et de transfert très variables spatialement. On s'est aperçu que la variabilité spatiale de ces paramètres est différente selon les jeux de données utilisés pour le calage. Ces jeux de paramètres ne semblent donc pas optimaux pour représenter des caractéristiques hydrologiques des bassins-versants. En effet, il semble que l'identification de ces paramètres soit en grande partie influencée par le choix de la période temporelle contenant les données utilisées en calage ([Jay-Allemand et al., 2019](#)).

Le modèle hydrologique GRD est conceptuel mais ses opérateurs permettent de représenter les phénomènes hydrologiques. Des relations entre les paramètres du modèle et des descripteurs sont donc susceptibles d'exister. L'utilisation de telles relations pourrait permettre d'améliorer le calage distribué des paramètres (gain de performances prédictives spatio-temporelles, augmentation de la robustesse des paramètres par rapport aux données permettant leurs identifications). Néanmoins, la structure du modèle est très simplifiée et les choix quant aux relations descripteurs-paramètres sont délicats. Par exemple des relations empiriques entre les paramètres des modèles hydrologiques globaux GR5H et GR6J et des descripteurs morphologiques et hydro-climatiques des bassins-versants ont été recherchés dans les travaux de thèse de [Lobligeois \(2014\)](#) et [Poncelet \(2016\)](#). Dans ces travaux, les paramètres des modèles sont calés sur des échantillons de bassins-versants jaugés. Des relations sont ensuite recherchées entre les valeurs des paramètres et des descripteurs. Globalement, aucune relation forte n'a été établie, mise à part pour le paramètre temporel de l'hydrogramme unitaire.

Dans ce chapitre, une nouvelle méthodologie est testée. Les paramètres du modèle distribué sont calés en imposant leurs variabilités spatiales par des descripteurs. Les descripteurs utilisés sont de natures climatiques, géologiques, pédologiques, topographiques ou liés à l'occupation des sols. Le principe des expériences est le suivant : contraindre la variabilité spatiale des paramètres selon la variabilité spatiale de chaque descripteur. Dans ce cas, l'algorithme d'estimation variationnelle permet d'identifier des valeurs du paramètre à caler selon des zones géographiques ayant des caractéristiques semblables. Cette approche impose de minimiser la fonction objectif sous des contraintes spatiales fortes. Aussi dans cette approche, les valeurs des paramètres ne sont pas déduites à priori à partir des propriétés des descripteurs, mais estimées directement par l'algorithme d'estimation variationnelle. Cette approche diffère donc des méthodes de calage des modèles distribués mise en œuvre par Pokhrel and Gupta (2010), Anderson et al. (2006) et Samaniego et al. (2010). Dans ces dernières, quelques "super-paramètres" (additions de constantes, facteurs multiplicateurs, exposants) sont calés et servent à relaxer une estimation non-uniforme et à priori de la valeur des paramètres (cette estimation étant réalisée à partir des caractéristiques spatiales des bassins-versants).

Notons que les contraintes spatiales imposées par ces descripteurs permettent de régulariser (pénaliser) la solution. Cependant la façon de prendre en compte ces contraintes est structurellement très différente de la fonction de régularisation de type Tikhonov utilisée au chapitre précédent. La technique utilisée ici permet de réduire le nombre de paramètre à caler (réduction de la taille du vecteur de contrôle) en agissant sur la structure du maillage. Cette procédure a été décrite au chapitre 4 section 4.5.3).

Dans les expériences suivantes, le calage distribué des paramètres, dont leurs variabilités spatiales sont issues des descripteurs, est qualifié de calage **semi-uniforme**.

Dans cette partie, nous tentons de répondre aux questions suivantes : contraindre la variabilité spatiale des paramètres du modèle à l'aide de descripteurs permet-elle d'améliorer le calage ? Cette nouvelle variabilité reflète-t-elle davantage les caractéristiques hydrologiques des bassins-versants et permet-elle d'améliorer les performances prédictives du modèle ? Permet-elle d'obtenir des jeux de paramètres plus robustes et donc moins sensibles aux données utilisées en calage ? Les expériences suivantes tenteront d'apporter des réponses à ces questions.

7.1.2 Expériences réalisées

Une nouvelle étape de calage est introduite. Le calage distribué des paramètres du modèle GRD est alors réalisé en trois étapes successives (au lieu de deux étapes lors des expériences précédentes) :

1. **Calage uniforme** : calage des paramètres spatialement uniformes sur chaque domaine. Trois valeurs de paramètres sont recherchées. Cette étape de calage est

résolue par l'algorithme de recherche global présenté à la section 4.6.

2. **Calage semi-uniforme** : Ce calage est réalisé en prenant comme ébauche le calage uniforme. Lors de cette étape, les paramètres sont contraints spatialement par des descripteurs. Ces descripteurs imposent la variabilité spatiale des paramètres : l'algorithme d'estimation variationnelle identifie les valeurs des paramètres selon des zones géographiques ayant des caractéristiques semblables. La fonction régularisation de type Tikhonov n'est pas prise en compte ($\beta = 0$). Les bornes autorisent les paramètres à dévier de +/- 300% de la valeur uniforme d'ébauche.
3. **Calage distribué** : Ce calage est réalisé en prenant comme ébauche le calage semi-uniforme. Pour cette étape l'algorithme de calage variationnel est utilisé. La fonction régularisation de type Tikhonov n'est pas prise en compte ($\beta = 0$). Pour cette étape les bornes minimales et maximales sont variables spatialement et décrivent une enveloppe autour de l'ébauche. Les paramètres sont autorisés à dévier de +/- 50% maximum de la valeur locale de l'ébauche. Les bornes sont davantage contraignantes que celles utilisées à l'étape précédente "**calage semi-uniforme**". L'idée est de relaxer le calage semi-uniforme tout en conservant la variabilité imposée par les descripteurs. Cette contrainte pourrait être imposée par l'utilisation de la fonction de régularisation de type Tikhonov. Cependant, les précédentes expériences ont montré que la structure de cette fonction ne semblait pas appropriée pour décrire les variabilités des paramètres du modèle hydrologique. Il n'est donc pas souhaitable d'utiliser cette fonction pour ces expériences (en tout cas pas de cette manière).

La même méthodologie que celle utilisée pour les expériences de calage-validation croisées décrites au chapitre 6 est appliquée. Pour rappel, les expériences de calage-validation croisées sont réalisées à l'aide des données de pluies et de débits disponibles sur les périodes temporelles P1 (du 01/01/2008 au 01/01/2013) et P2 (du 01/01/2013 au 01/01/2018). Les paramètres du modèle sont calibrés à l'aide des observations de débits disponibles aux stations avals de chaque bassin-versant. Les performances prédictives spatio-temporelles du modèle sont évaluées (performance aux exutoires n'ayant pas servi au calage et sur une période différente de celle utilisée pour le calage).

Six expériences de calage-validation croisées sont réalisées. Chaque expérience utilise un jeu de descripteurs différent. Les relations à priori entre les descripteurs et les paramètres du modèle sont décrites ci-dessous :

Les différents descripteurs testés pour contraindre la variabilité spatiale des capacités des réservoirs de production c_p :

- Les données d’occupation du sol issues de la base de donnée Corine Land Cover. Ces données permettent de distinguer les types d’occupation du sol (forêt, culture, ville...). Les types d’occupation du sol influent sur les quantités d’eau stockées en surface et relarguées par évapotranspiration. Les types d’occupation du sol ont été regroupés en cinq classes selon la nomenclature du premier niveau proposée dans cette base de données. Les données utilisées ont été présentées au chapitre 2 à la section 2.1.3. Ce descripteur est noté **CLC**. Ce descripteur est cartographié pour tous les bassins-versants dans l’annexe (C.1).
- L’Indice Mixte de Production : cet indice caractérise les pixels en fonction de leurs prédispositions à la production (Mahe, 2016). Un pixel peu productif stockera davantage les précipitations qu’un pixel fortement productif qui à l’opposé contribuera à l’écoulement. Les données utilisées ont été présentées au chapitre 2 à la section 2.1.3. Ce descripteur est noté **IMP**. Ce descripteur est cartographié pour tous les bassins-versants dans l’annexe (C.3).
- La densité de drainage : plus la densité de drainage est forte et plus les précipitations sont susceptibles de participer à l’écoulement et donc de rejoindre le réservoir de transfert. Les données utilisées ont été présentées au chapitre 2 à la section 2.1.3. Ce descripteur est noté **DD**. Ce descripteur est cartographié pour tous les bassins-versants dans l’annexe (C.2).

Les descripteurs utilisés pour contraindre la variabilité spatiale des capacités des réservoirs de transfert c_{tr} :

- Les données des couches géologiques affleurantes en surface. Ces données proviennent de la base de donnée GEOL du BRGM présentées au chapitre 2 à la section 2.1.3. Les propriétés du sous-sol influent sur le transfert de l’eau de l’amont vers l’aval (vitesses d’infiltration, volumes stockés, vitesses de drainage). Les réservoirs de transfert pourraient donc être corrélés avec les propriétés géologiques du sol. Ce descripteur est noté **GEOL**. Ce descripteur est cartographié pour tous les bassins-versants dans l’annexe (C.4).
- La pente moyenne local : plus la pente est forte et plus l’écoulement vers le réseau de drainage est favorisé et rapide. Les données utilisées ont été présentées au chapitre 2 à la section 2.1.3. Ce descripteur est noté **PENTE**. Ce descripteur est cartographié pour tous les bassins-versants dans l’annexe (C.5).

Le descripteur utilisé pour contraindre la variabilité spatiale des paramètres de routage ξ :

Les résultats des expériences de calage menées au chapitre 6 ont montrés que les paramètres de routage du modèle hydrologique sont corrélés au réseau de drainage. La

variabilité de ce paramètre est contraint par un descripteur permettant de séparer les mailles situées sur le réseau hydrographique et les mailles situées sur les versants. Cette distinction s'opère à l'aide d'un seuil arbitraire sur les surfaces cumulées. Lorsque la surface amont cumulée est supérieure à 10 km^2 , les mailles sont considérées comme des mailles appartenant au réseau de hydrographique principale. À l'inverse, lorsque la surface amont cumulée est inférieure à 10 km^2 , les mailles appartiennent aux versants. Ce descripteur est noté **SURF**. Ce descripteur est cartographié pour tous les bassins-versants dans l'annexe (C.6).

Une combinaison de descripteurs propre à chaque expérience est utilisée. Les couples "expériences / descripteurs" sont listés dans le tableau 7.1.

TABLE 7.1 – Liste des six expériences de calage réalisées. Pour chacune des six expériences, les descripteurs utilisés pour contraindre la variabilité spatiale des paramètres de production c_p , de transfert c_{tr} et de routage ξ sont listés.

Nom	c_p	c_{tr}	ξ
M1	CLC	GEOL	SURF
M2	CLC	PENTE	SURF
M3	DD	PENTE	SURF
M4	DD	GEOL	SURF
M5	IMP	GEOL	SURF
M6	IMP	PENTE	SURF

7.1.3 Résultats des expériences

Analyse des performances prédictives spatio-temporelles : La figure 7.1 présente les résultats des critères de performance NSE et KGE obtenus aux stations de validation et sur la période de validation sous forme de box-plot. Ces résultats correspondent aux performances prédictives spatio-temporelles du modèle à l'issue de la deuxième et troisième étape de calage (respectivement le calage **semi-uniforme** et **distribué**). Les performances prédictives spatio-temporelles de chaque expérience de calage (tableau 7.1) sont comparées avec les performances prédictives obtenues pour les expériences de référence **A** (calage uniforme) et **B** (calage distribué) décrites au chapitre précédent (voir chapitre 6 section 6.1.3).

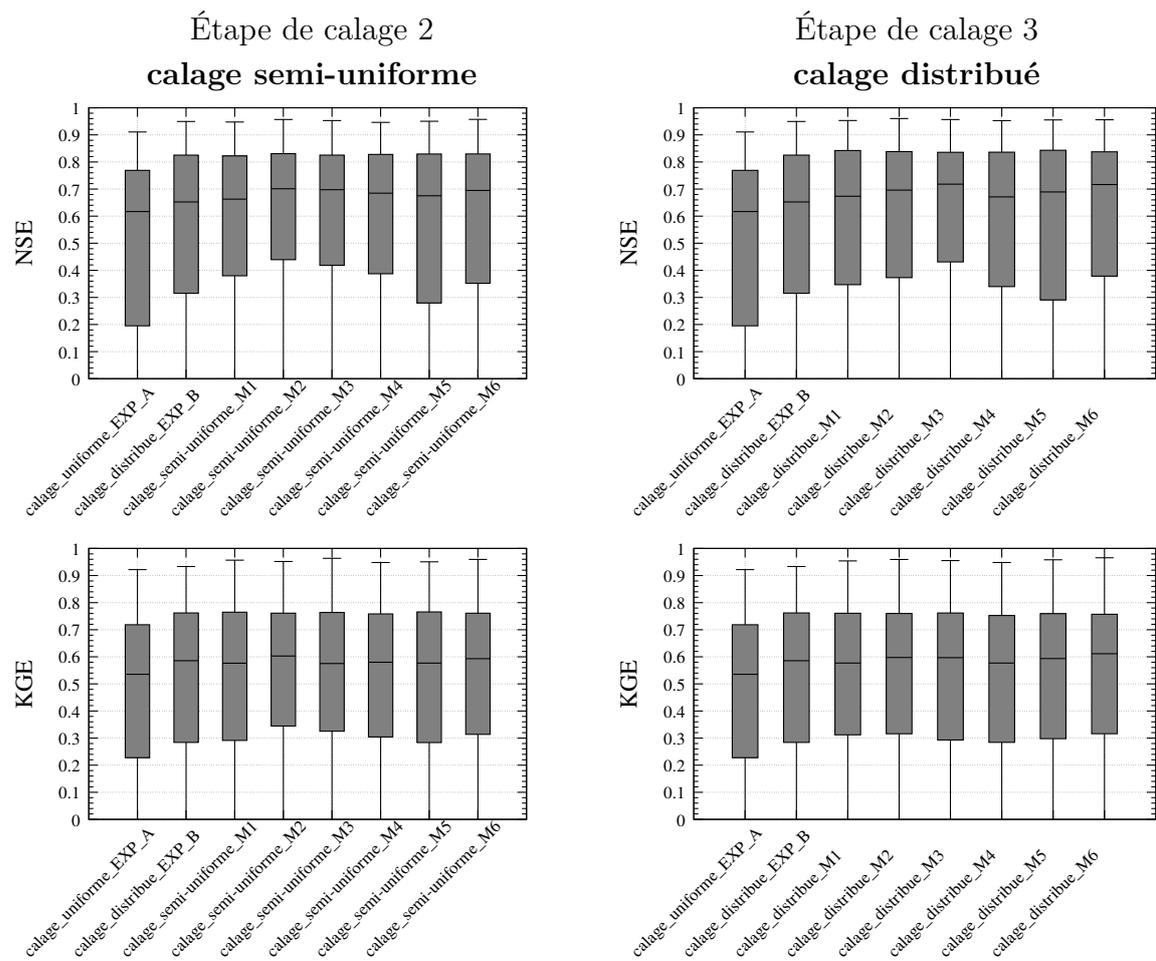


FIGURE 7.1 – Performances prédictives spatio-temporelles (**validation spatio-temporelle**) à l'issue des étapes de calage 2 (calage semi-uniforme, colonne de gauche) et 3 (calage distribué, colonne de droite). Les performances sont montrées à l'aide des critères de NSE (graphiques du haut) et de KGE (graphiques du bas). Chaque box-plot est construite à l'aide des résultats obtenus aux 118 exutoires de validation sur les périodes de calage P1 et P2, soit 236 points. Sur chaque graphique, huit box-plots sont tracés : calage uniforme, calage distribué issu de l'expérience **B** (expérience réalisée au chapitre précédent 6), calages semi-uniformes ou distribués (expériences **M1**, **M2**, **M3**, **M4**, **M5**, **M6**)

À l'issue de la deuxième étape du calage, environ 10 à 20 valeurs de paramètres ont été calibrées selon les bassins-versants et les descripteurs choisis. Alors qu'un nombre réduit de paramètres ont été calibrés, les performances prédictives spatio-temporelles mesurées à l'aide du critère de NSE égalisent ou dépassent celles obtenues avec l'expérience **B** (calage distribué de référence). Les meilleures performances (médiane et variance) sont obtenues pour les expériences **M2**, **M3** et **M6**. Ces trois expériences imposent que la variabilité des capacités des réservoirs de transfert soit contraintes par les pentes. À l'issue de la troisième étape de calage, l'ensemble des paramètres distribués sont calibrés. Si l'on regarde la valeur médiane des distributions, les performances prédictives

spatio-temporelles sont légèrement améliorées. En revanche l'étalement des distributions augmente sensiblement. Les meilleurs résultats sont obtenus pour les expériences **M3** et **M6**. Le gain de performances à l'issue de la troisième étape est faible par rapport au nombre de paramètres calibrés. Cependant les bornes supérieures et inférieures des paramètres sont davantage contraignantes que celles utilisées pour l'expérience **B**. Cela pourrait expliquer ce gain de performances limité.

Les gains performances prédictives spatio-temporelles mesurées à l'aide du critère de KGE sont moins visibles. Au regard de ce critère l'expérience **M2** semble légèrement meilleurs. Au total, les gains de performances en termes de NSE médian atteignent 10% ($NSE = 0.62$ pour l'expérience **A** et $NSE = 0.72$ pour l'expérience **M3**) et la distribution statistique est davantage resserrée (les quantiles de NSE 25% vaut $NSE_{0.25} \approx 0.2$ pour l'expérience **A**, $NSE_{0.25} \approx 0.32$ pour l'expérience **B**, $NSE_{0.25} \approx 0.44$ pour l'expérience **M3**).

Analyse de la stabilité des paramètres : La figure 7.2 permet d'évaluer l'adéquation des jeux de paramètres calés entre les deux périodes temporelles P1 et P2. On évalue la robustesse des paramètres calés à l'aide des deux mêmes critères qu'au chapitre 6 : le rapport des valeurs moyennes calculées sur chaque période $\mu(X)_{P1}/\mu(X)_{P2}$ et le coefficient de corrélation de Pearson. Ces deux critères sont calculés pour chacun des 23 bassins-versants et les résultats sont tracés sous forme de box-plots. Le critère de corrélation de Pearson est calculé seulement pour les jeux de paramètres issus de l'étape de calage distribué (étape 3). Les valeurs des critères de robustesse calculés pour les expériences **M1-M6** sont comparés celles des expériences **A** et **B** (expériences réalisées au chapitre 6).

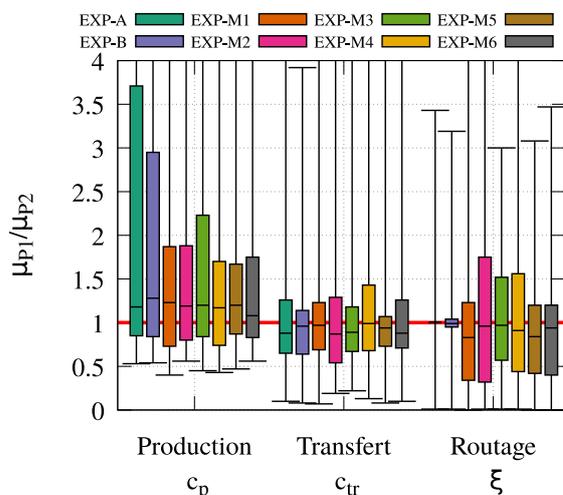
L'analyse des rapports des valeurs moyennes des paramètres $\mu(X)_{P1}/\mu(X)_{P2}$ montrent que les calages distribués et semi-uniformes améliorent la stabilité du paramètre de production. Cependant l'utilisation de descripteurs ne permet pas d'améliorer davantage ce critère. On remarque aussi que l'étape numéro trois de calage a tendance à légèrement étaler les distributions du critère μ_{P1}/μ_{P2} . Notons que la fonction de régularisation n'est pas utilisée pour les expériences **M1-M6** alors qu'elle l'est pour l'expérience **B**.

- L'analyse des distributions des coefficients de corrélation de Pearson montrent que :
- L'expérience **M2** permet d'améliorer la robustesse des capacités des réservoirs de production. En revanche, la robustesse des capacités des réservoirs de transfert diminue.
 - L'expérience **M3** permet d'améliorer la robustesse des capacités des réservoirs de transfert. En revanche, la robustesse des capacités des réservoirs de production diminue.

- Les expériences **M2**, **M3** et **M5** permettent d'améliorer significativement la robustesse des paramètres de routage, alors que l'expérience **M6** la dégrade.
- La robustesse des paramètres de l'expérience **M1** est toujours inférieure aux autres expériences.

Au regard des résultats précédents, les descripteurs testés dans l'expérience **M3** ("densité de drainage" pour les réservoirs de production et "pentes" pour les réservoirs de transfert) sont retenus. Néanmoins ce choix est discutable : les performances prédictives obtenues pour expériences **M2**, **M3** et **M6** sont proches ; la classification utilisée pour chaque descripteur doit être approfondie (celle-ci influence directement le nombre de valeurs de paramètres à caler et donc la variabilité spatiale obtenue après l'étape de calage semi-uniforme) ; la robustesse des paramètres obtenus après calage sous la contrainte de descripteurs n'est pas significativement améliorée. Des investigations supplémentaires sont donc nécessaires afin d'affiner ces résultats.

Étape de calage 2
calage semi-uniforme



Étape de calage 3
calage distribué

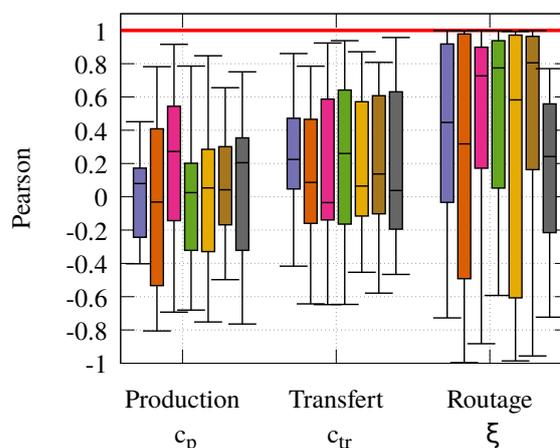
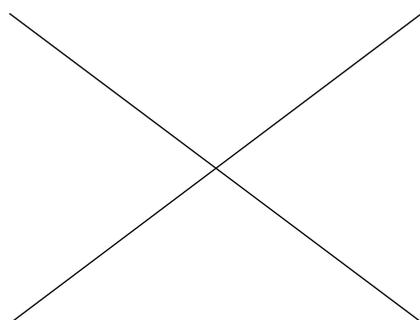
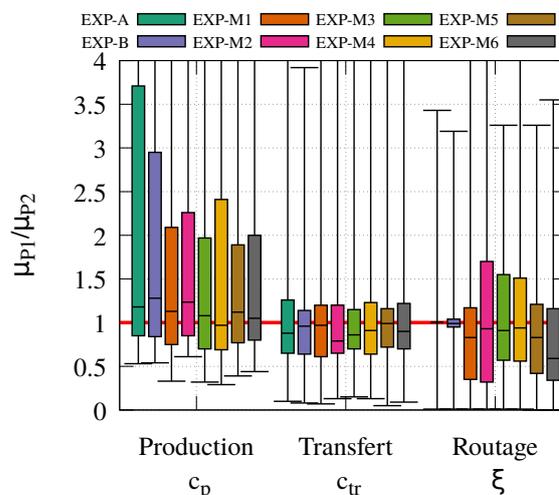


FIGURE 7.2 – Box-plots représentant les distributions des critères d'adéquation calculés à partir des paramètres calés sur les périodes P1 et P2 pour chaque bassin-versant. Ces critères ont été calculés à l'issu des résultats des étapes de calage numéro 2 (calage semi-uniforme, colonne de gauche) et numéro 3 (calage distribué, colonne de droite). Les graphiques du haut montre les rapports entre les valeurs moyennes des paramètres. Le graphique du bas (à droite) montre les distributions des coefficients de corrélation de Pearson (lorsque les paramètres sont uniformes où quasi-uniformes le calcul des coefficients de Pearson n'est pas pertinent et le graphique n'est pas affiché).

Évaluation des performances de l'expérience M3 : Les performances prédictives spatio-temporelles à l'issu des résultats de l'expérience **M3** sont comparées aux performances prédictives issus des expériences **A** et **B** (expériences décrites au chapitre 6) sur le graphique 7.3. Les gains sont significatifs par rapport au critère NSE, négligeables ou nuls par rapport au critère KGE.

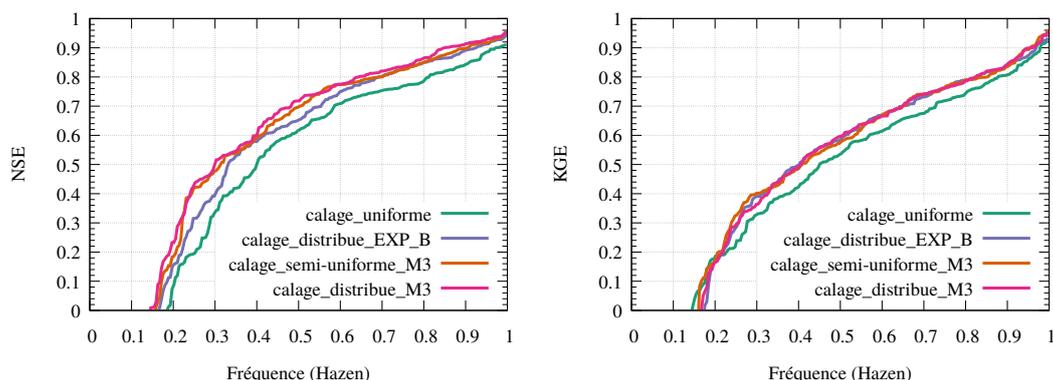


FIGURE 7.3 – Performances en **validation spatio-temporelle**, 118 exutoires de validation sur les périodes de validation P1 et P2, obtenues pour la variantes 2 du modèle GRD après calage **uniforme** (expérience **A**, courbe verte), **distribué** (expérience **B**, courbe bleu), calage semi-uniforme (expérience **M3**, courbe orange) et distribué (expérience **M3**, courbe rose) des paramètres. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

Une analyse approfondie des performances est réalisée à l'aide d'une décomposition du critère de KGE. Les coefficients de Pearson, RVAR et RMU, présentés au chapitre 6 section 6.1.2, sont calculés séparément. Ces résultats complémentaires sont présentés à la figure 7.4. Notons que la fonction objectif de notre calage n'était par le KGE, mais le NSE.

Il est intéressant de remarquer que le coefficient de corrélation linéaire de Pearson est amélioré pour l'expérience **M3** alors que les autres composantes du KGE sont légèrement dégradées. Plus globalement le calage distribué, par rapport au calage uniforme, apporte une hausse significative des critères de corrélation de Pearson et des critères RVAR. Le critère de biais (RMU), calculé à partir des débits moyens, n'est pas amélioré.

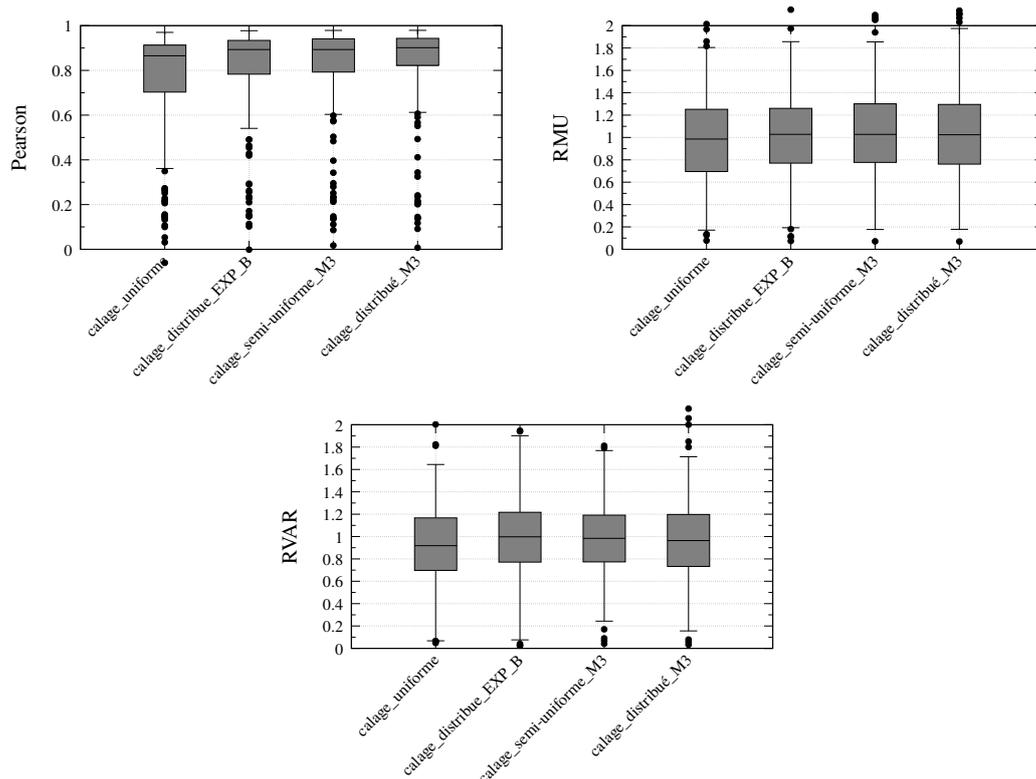


FIGURE 7.4 – Performances prédictives spatio-temporelle (**validation spatio-temporelle**), 118 exutoires de validation sur les périodes de calage P1 et P2 (236 points), obtenues pour la variantes 2 du modèle GRD après calage. Quatre boxplot sont présentées : performances prédictives spatiales avec calage uniforme, performances prédictives spatiales avec calage distribué **EXP B** (expérience réalisée au chapitre précédent 6), performances prédictives spatiales avec calage contraint par les descripteurs EXP-M3. Les performances sont décrites à l’aide de trois critères composant le critère de KGE. Le graphique en haut à gauche présente les performances à l’aide du critère de corrélation linéaire (Pearson), le graphique en haut à droite présente les performances à l’aide du critère RMU (mesure du biais) et le graphique en bas présente les performances à l’aide du critère RVAR (mesure de la variabilité).

Sur le graphique 7.5, les performances prédictives du modèle à l’issue de l’expérience **M3** sont analysées en fonction de l’éloignement des exutoires de validation par rapport à l’exutoire de calage. Les performances à l’issue des trois étapes du calage sont affichées et comparées aux performances du calage distribué expérience **B**. A chaque étape de calage les performances prédictives du modèle sont améliorées. Les gains sont nettement plus importants à l’issue de l’étape numéro 2 du calage (passage d’un calage uniforme à un calage semi-uniforme). Les gains apportés par l’étape numéro 3 du calage sont faibles. À l’issue des trois étapes de calage, les performances prédictives du modèle sont globalement meilleures que celles de l’expérience **B** (médianes supérieures et/ou boxplots davantage resserrées). Cependant pour les bassins-versants les plus éloignés de l’exutoire de calage (classe $RS = [0 : 0.2]$), les performances prédictives évoluent peu.

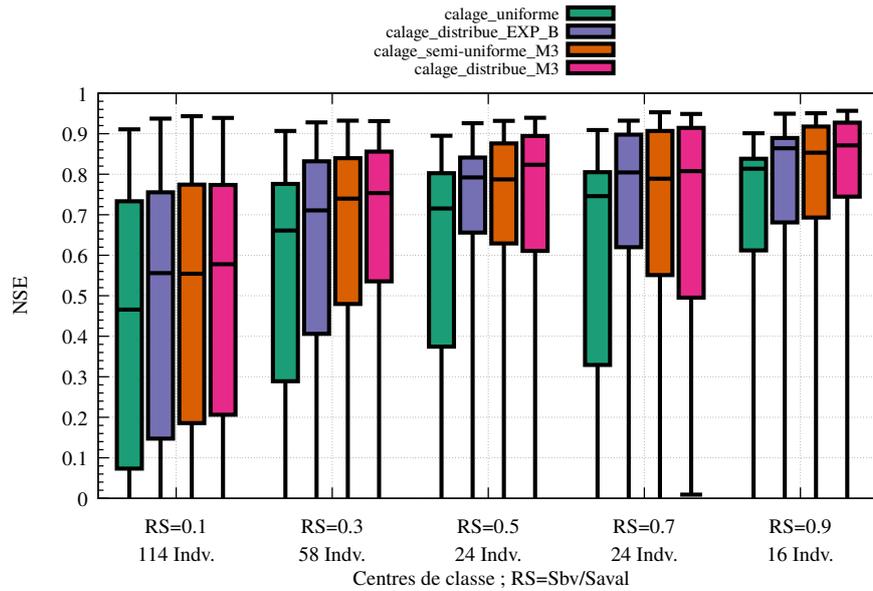


FIGURE 7.5 – Relation entre les critères de Nash-Sutcliffe (NSE) et cinq classes du rapport $RS_i = \frac{S_i}{S_{aval}}$ (surface de l'exutoire de validation) / (Surface de l'exutoire de calage). Résultat en validation spatio-temporelle, expériences **A**, **B** et **M3** (étapes deux et trois).

Analyse des jeux de paramètres distribués : Les jeux de paramètres obtenus à l'issue de l'expérience **M3** sont cartographiés par bassin-versant et pour chaque période temporelle P1 et P2 sur les figures 7.6, 7.7 et 7.8. A l'issue de la 3^{eme} étape de calage, les paramètres sont très variables spatialement. Bien que des contraintes spatiales ont été apportées par des descripteurs, les paramètres calés sur les périodes temporelles P1 et P2 restent très différents (le Loing, l'Hérault, le vistre, le Vidourle, l'Argens). Ce résultat confirme l'analyse précédente concernant les critères de robustesse. Ce problème n'est pas forcément lié à la procédure de calage distribuée puisqu'il existe aussi pour le calage uniforme des paramètres (voir les résultats des critères de robustesse sur la figure 7.2). Ce problème de robustesse est en partie lié aux modèles hydrologiques conceptuels (Dakhlaoui et al., 2017; Brigode et al., 2013). Le calage n'a peut-être pas été assez contraint entre les étapes 2 et 3, ce qui pourrait accentuer les différences entre les jeux de paramètres. Notons aussi que les conditions expérimentales de calage sont très pénalisantes puisque seules les observations de débits aux exutoires aval des bassins-versants sont utilisées.

Ce résultat montre que les descripteurs utilisés pour contraindre la variabilité spatiale ne suffisent à expliquer le fonctionnement des opérateurs du modèle hydrologique GRD. Cependant, leurs utilisations permettent de régionaliser naturellement les paramètres à l'extérieur des bassins-versants. Les performances prédictives du modèle doivent alors être vérifiées sur des bassins-versants voisins. Ces résultats ouvrent donc des perspectives intéressantes pour la régionalisation des paramètres du modèle.

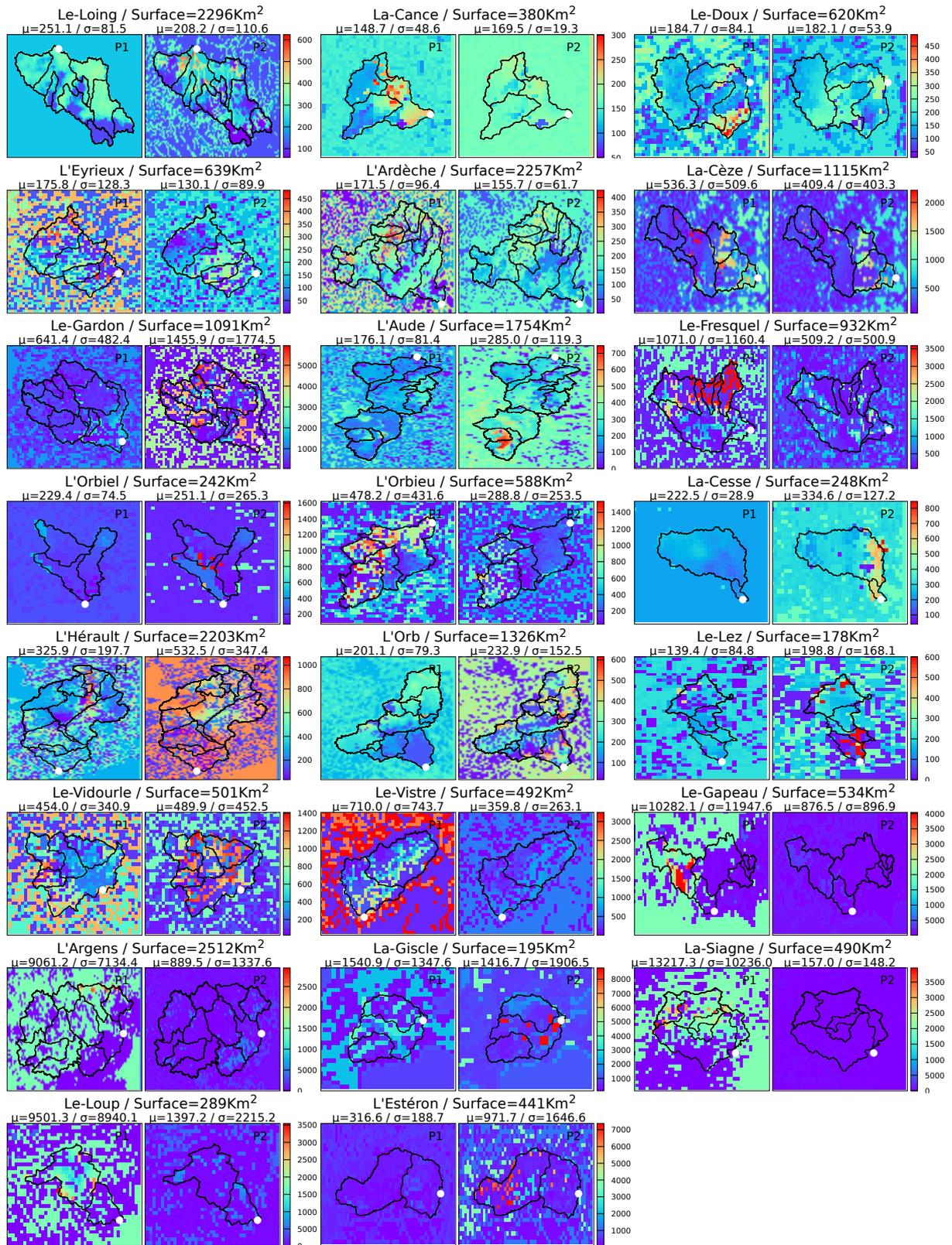


FIGURE 7.6 – Capacités des réservoirs de production c_p obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.

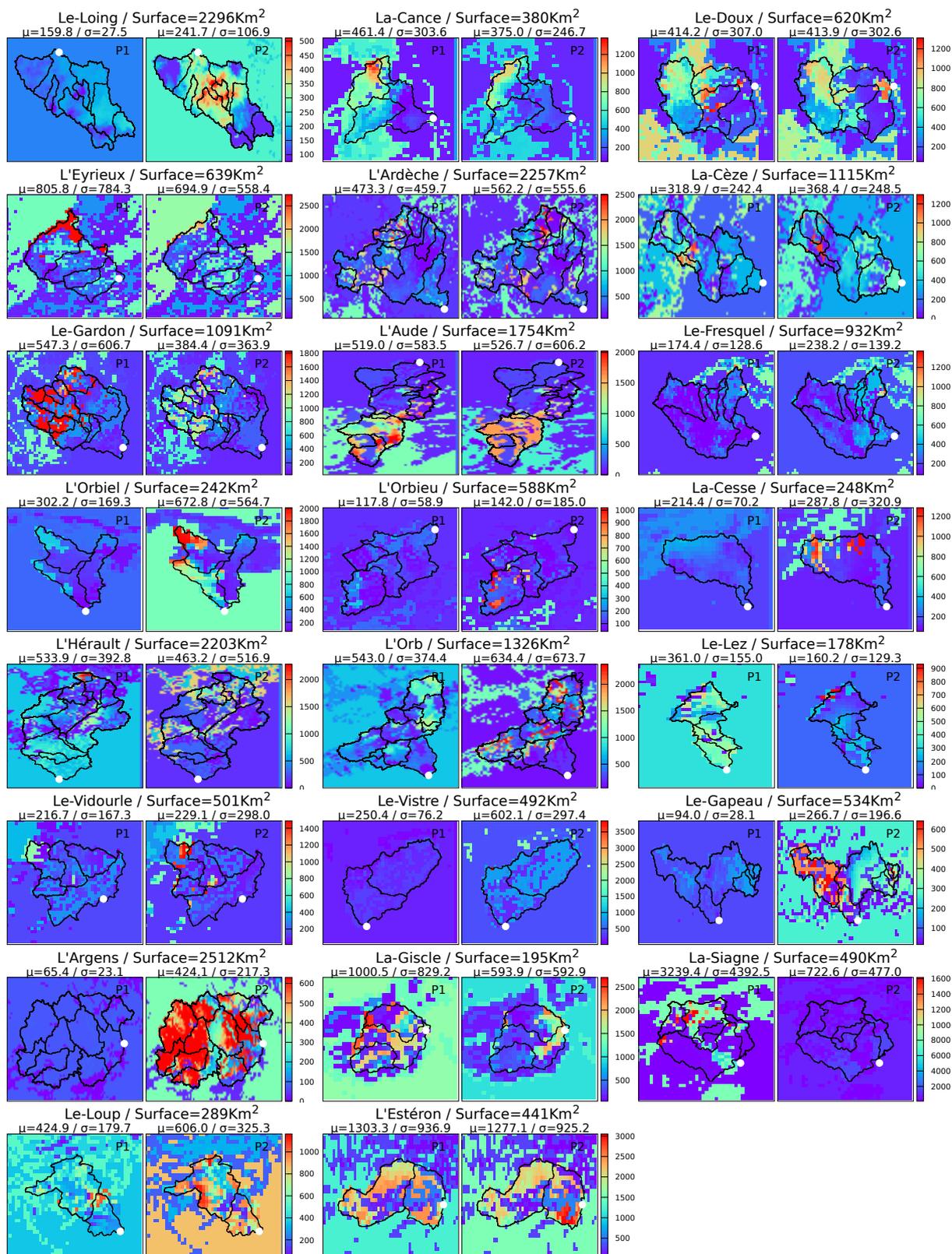


FIGURE 7.7 – Capacités des réservoirs de transfert c_{tr} obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.

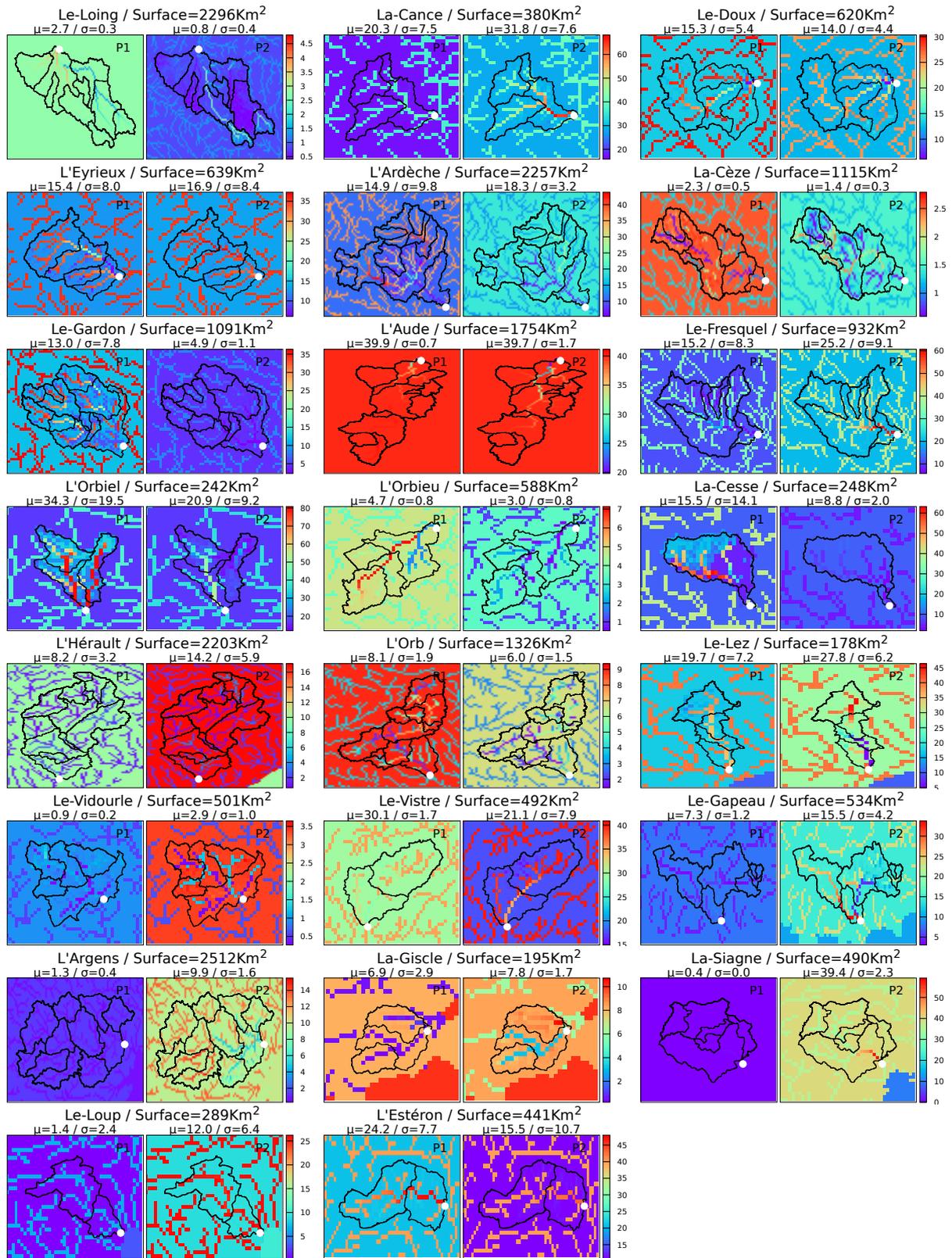


FIGURE 7.8 – Paramètres de routage ξ obtenues après calage sur les périodes temporelles P1 (gauche) et P2 (droite) pour chaque bassin-versant. Les moyennes μ et écarts-types σ sont calculés sur l'aire du bassin-versant.

Conclusion : Les résultats sont encourageants. Les contraintes spatiales apportées par des descripteurs choisis "à priori" permettent d'égaliser voir d'améliorer les performances prédictives du modèle tout en réduisant drastiquement le nombre de paramètres à estimer. En ce sens le calage distribué est amélioré et les descripteurs "densité de drainage" et "pente" améliorent le mieux la représentation des processus hydrologiques simulés par le modèle conceptuel GRD. De plus, ce type de calage offre l'opportunité de régionaliser les paramètres du modèle à l'extérieur des bassins-versants jaugés. Cependant la robustesse des paramètres n'est pas améliorée et ces résultats sont exploratoires. Le choix et l'utilisation des descripteurs devront être affinés (tests de nouveaux descripteurs, nombres de classes par descripteur) et davantage de tests doivent être effectués pour évaluer l'influence des bornes et du poids de l'ébauche lors de la minimisation de la fonction coût. Enfin, l'approche de régionalisation devra être expérimentée en considérant plusieurs bassins-versants de calage et plusieurs bassins-versants voisins pour la validation.

7.2 Test d'une fonction objectif basée sur la fonction de distribution Gamma

Les précédentes expériences de calage-validation ont montré que les paramètres distribués pouvaient être très différents entre les deux périodes. Le calage manque de robustesse même avec la méthode proposée qui vise à améliorer la cohérence des paramètres à l'aide de descripteurs. Les causes de ce manque de robustesse sont :

- Une climatologie différente des deux périodes temporelles rendant le calage sensible aux événements hydrologiques extrêmes. Ce problème est connu et concerne les modèles conceptuels (Dakhlaoui et al., 2017; Brigode et al., 2013).
- La sensibilité du calage aux incertitudes de modélisation et aux incertitudes sur les données observées. La réaction du modèle hydrologique à un signal d'entrée (la pluie) est parfois très différente de la réaction naturelle observée. Ces différences peuvent corrompre le fonctionnement des méthodes de calage.

Pour tenter de contourner ces problèmes, une méthode d'estimation dite "robuste" est suggérée pour caler les paramètres du modèle hydrologique.

7.2.1 Principe et objectif

Pour tenter d'améliorer la robustesse des paramètres calés, de nouvelles expériences de calage-validation sont réalisées avec une méthode d'estimation dite "robuste". Il s'agit d'utiliser la fonction objectif décrite au chapitre 4 section 4.3.2. Cette fonction objectif

doit permettre de limiter l'influence des valeurs incertaines et extrêmes qu'elles soient simulées ou observées. Elle est basée sur la fonction cumulée de densité de probabilité Gamma. L'objectif est d'obtenir des paramètres plus stables entre les deux périodes temporelles et moins dépendants des jeux de données.

Le protocole de calage-validation est le même que celui utilisé lors des précédentes expériences. Les paramètres du modèle sont calés à l'aide des débits observés aux exutoires aval de chaque méta-bassins-versants et pour chaque période temporelle P1 et P2. L'adéquation entre les jeux de paramètres estimés pour chacune des deux périodes temporelles P1 et P2 est évaluée. Les performances prédictives spatio-temporelles du modèle sont vérifiées.

7.2.2 Expériences réalisées

Cinq expériences de calage-validation croisées sont réalisées. La fonction objectif utilisée dépend du paramètre d'étalement sp de la fonction de distribution de la loi Gamma. Quatre valeurs de ce paramètre sont testées. En modifiant le paramètre d'étalement de la loi Gamma, la fonction objectif donne des poids différents aux différentes classes de débits. Plus le paramètre sp est grand, plus un poids fort est donné aux valeurs extrêmes. La fonction objectif Gamma combinée à la méthode de régularisation itérative est testée dans une sixième expérience. Ces expériences sont récapitulées dans le tableau 7.2.

TABLE 7.2 – Expériences de calage réalisées avec la fonction objectif Gamma pour différentes valeurs du coefficient d'étalement noté sp .

Nom	sp	Régularisation
G1-ITREG	10	Tikhonov-itérative
G1	10	-
G2	50	-
G3	100	-
G4	500	-

7.2.3 Résultats

Les résultats sont tout d'abord analysés en termes de robustesse. La stabilité des paramètres obtenues à l'issue des calages pour les périodes temporelles P1 et P2 est présentée à la figure 7.9. Deux critères sont calculés : le coefficient de corrélation de Pearson et le rapport entre les valeurs moyennes des paramètres obtenus sur les périodes temporelles P1 et P2. Les résultats sont comparés avec ceux issus de l'expérience **B** (expérience réalisée au chapitre 6). Pour rappel, les paramètres utilisés dans l'expérience B sont les suivants : ébauche uniforme ; fonction coût quadratique calculant le critère de NSE aux exutoires aval ; régularisation itérative type Tikhonov.

Les expériences G1 et G1-ITREG permettent d'obtenir un gain sensible de robustesse. Ce gain est notable en ce qui concerne les capacités des réservoirs de transfert et les paramètres de routage. Il est moindre pour les capacités des réservoirs de production. La robustesse des paramètres a tendance à diminuer pour les expériences G2, G3 et G4. La fonction objectif Gamma combiné avec une méthode de régularisation itérative semble apporter un effet positif sur la robustesse des paramètres estimés.

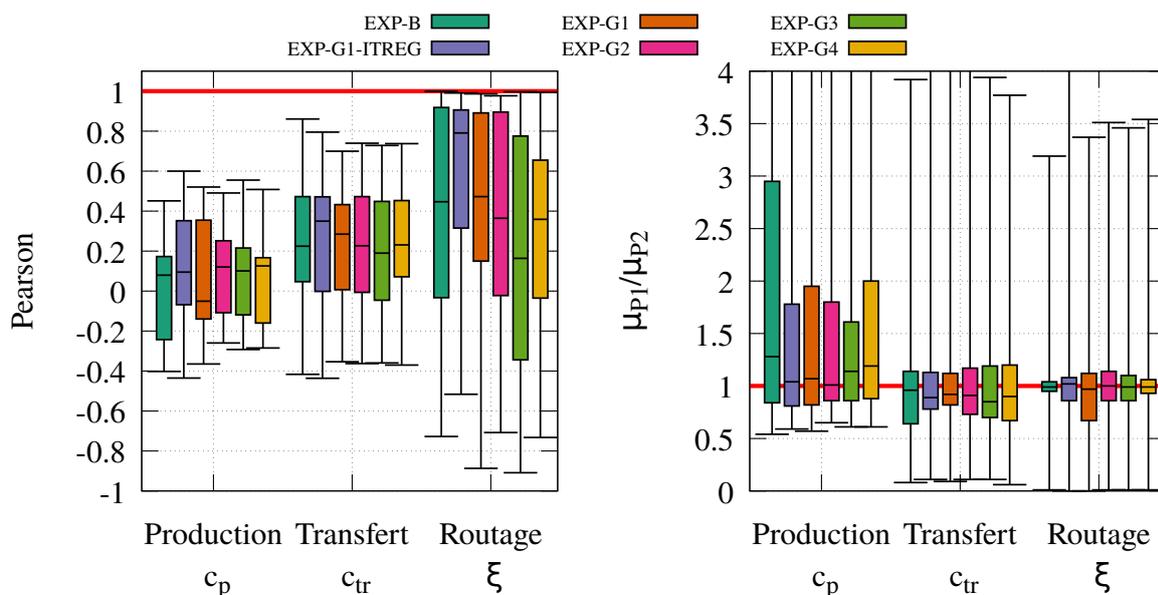


FIGURE 7.9 – Box-plots des distribution des critères d'adéquation entre les paramètres distribués période P1 et période P2 pour chaque bassin-versant. Graphique de gauche : coefficients de Pearson. Graphique de droite : rapports des valeurs moyennes.

Les résultats des performances prédictives spatio-temporelles sont analysés à l'aide des critères de performance NSE et KGE sur la figure 7.10. Les performances prédictives du modèle des expériences **G1-ITREG**, **G1**, **G2**, **G3**, **G4** sont comparées avec celles de l'expérience **B**. Notons que l'analyse des performances avec le critère NSE est pénalisante pour les expériences **G1-ITREG**, **G1**, **G2**, **G3**, **G4** puisque, contrairement à l'expérience **B**, le critère de NSE n'a pas été utilisé pour le calage des paramètres.

Lorsque le paramètre d'étalement sp de la loi Gamma diminue (de l'expérience **G4** à l'expérience **G1**), la fonction coût devient moins sensible aux valeurs extrêmes de débits. Plus ce coefficient diminue et plus les performances prédictives s'améliorent légèrement (augmentation de la médiane et réduction de la distribution). Pour l'expérience **G1**, les performances prédictives spatio-temporelles deviennent légèrement meilleures que celles obtenues pour l'expérience **B**.

Pour l'expérience **G4**, la fonction objectif Gamma est davantage sensible aux valeurs

extrêmes de débits. En ce sens, les résultats devraient se rapprocher de ceux de l'expérience **B**. Par rapport à cette dernière, les performances prédictives du modèle sont légèrement dégradées. Cette dégradation des performances provient probablement des choix des critères : pour l'expérience **B**, le NSE est utilisé en calage et pour l'analyse des performances ; pour l'expérience **G4** le NSE est utilisé uniquement pour l'analyse des performances.

L'estimation des paramètres lors de l'expérience **G1-ITREG** combine la fonction objectif Gamma avec la méthode de régularisation itérative type Tikhonov. Pour cette expérience, la distribution des critères de performance NSE se rétrécit autour de la médiane. L'utilisation combinée de la fonction objectif Gamma et de la régularisation semble avoir un effet positif sur les performances prédictives du modèle.

Ces résultats sont néanmoins à tempérer, car ces progrès sont perceptibles seulement avec le critère de NSE. En effet, les performances analysées avec le critère KGE sont globalement stables ou légèrement dégradées.

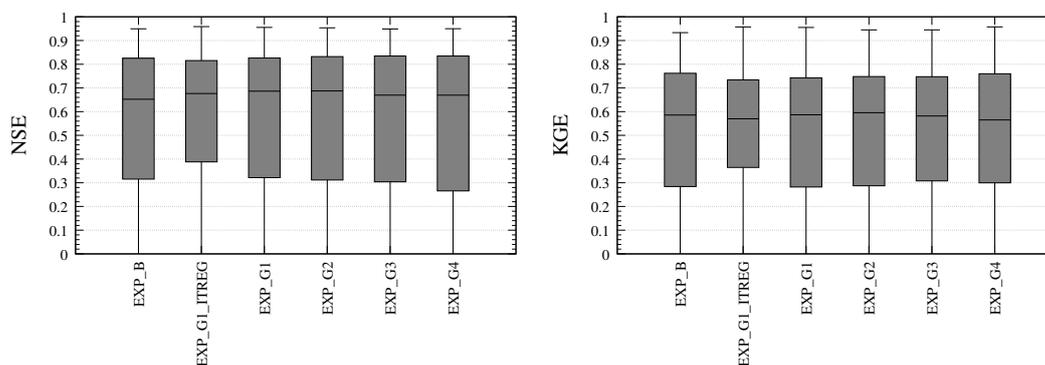


FIGURE 7.10 – Performances prédictives spatio-temporelle (**validation spatio-temporelle**), 118 exutoires de validation sur les périodes de calage P1 et P2 (236 points), obtenues pour la variantes 2 du modèle GRD après calage. Sept box-plots sont présentées : performances prédictives spatio-temporelles de l'expérience **EXP B** (expérience réalisée au chapitre précédent 6), performances prédictives spatio-temporelles des expériences **G1-ITREG**, **G1**, **G2**, **G3**, **G4**. Le graphique de gauche montre les performances en terme de NSE. Le graphique de droite montre les performances en terme de KGE.

Les jeux de paramètres estimés changent en fonction de la fonction coût utilisée et du choix de la valeur du paramètre d'étalement sp . Aux stations de calage, les performances varient peut. Seule l'utilisation de la régularisation dégradent les performances. Ce résultat est visible sur la figure 7.11 et illustre le fait que plusieurs jeux de paramètres donnent des performances équivalentes aux stations de calage. Seulement, ces jeux de paramètres sont différents. Les analyses des indicateurs de robustesses et des performances spatio-temporelles les montrent (voire figures 7.9 et 7.10).

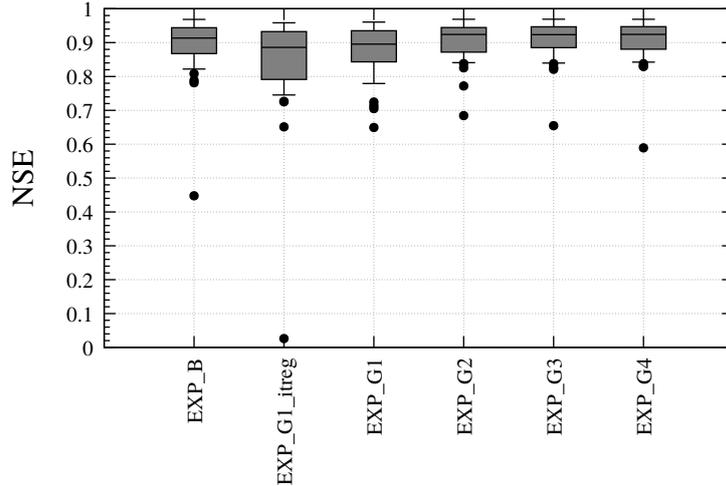


FIGURE 7.11 – Performances du modèle évaluées à l’aide du critère NSE aux 23 exutoires de calage sur les périodes de calage P1 et P2 (46 points par boxplot).

Afin d’illustrer les différences entre les jeux de paramètres estimés à l’aide de la fonction coût Gamma pour différentes valeurs du coefficient sp , les variabilités spatiales des capacités des réservoirs de production résultant des expériences **G1**, **G4** et **D** sont comparées pour deux bassins-versants (le Gardon et la Cesse). Pour rappel, l’expérience **D** utilise la fonction objectif quadratique (NSE) pour l’estimation des paramètres. Contrairement à l’expérience **B**, l’expérience **D** n’utilise pas de régularisation ($\beta = 0$). Il est donc plus juste de comparer les paramètres des expériences **G1** et **G4** avec ceux résultant de l’expérience **D**.

Les figures 7.12 et 7.13 permettent d’illustrer les différences entre les jeux de paramètres obtenus sur les bassins-versants du Gardon et de la Cesse. Les jeux de paramètre résultant des expériences **G4** et **G1** sont comparés avec les jeux de paramètres résultant de l’expérience **D**. Le critère θ est utilisé pour comparer les jeux de paramètres. Ce critère est présenté au chapitre 6, partie 6.2.6 (équation 6.7).

Lorsque le paramètre d’étalement sp est grand, la minimisation de la fonction objectif Gamma mène à des jeux de paramètre proche de la fonction objectif quadratique calculant le critère de NSE. En effet les jeux de paramètres résultant des expériences **G4** et **D** sont proches ($\theta \approx 0$ sur les bassins-versants).

Lorsque le paramètre d’étalement sp est petit, les valeurs fortes de débits ont des poids faibles. La minimisation de la fonction objectif Gamma mène à un jeu de paramètres radicalement différent (sans pour autant altérer les performances à la station de calage). Ces jeux de paramètre diffèrent de ceux trouvés par l’expérience **D** ($\theta \neq 0$) et donc de ceux trouvés par l’expérience **G4**.

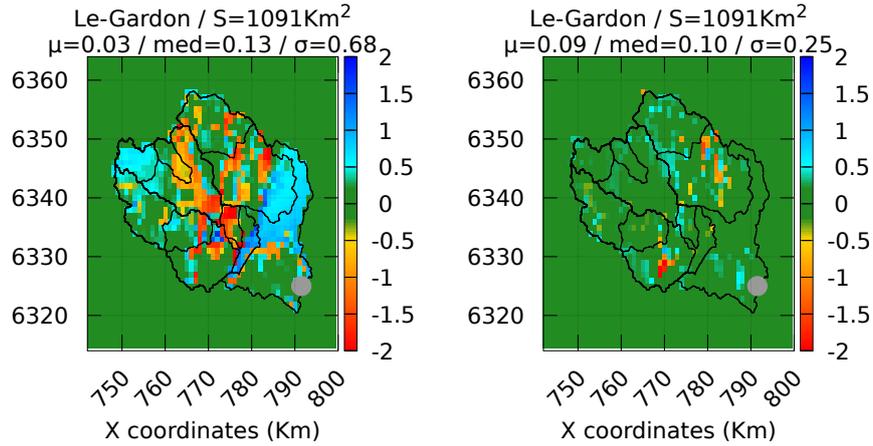


FIGURE 7.12 – Bassin-versant du Gardon. Comparaison des capacités des réservoirs de production c_p obtenues après calage sur la période P1 entre les expériences G1 et D (graphique de gauche) et les expériences G4 et D (graphique de droite) à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.

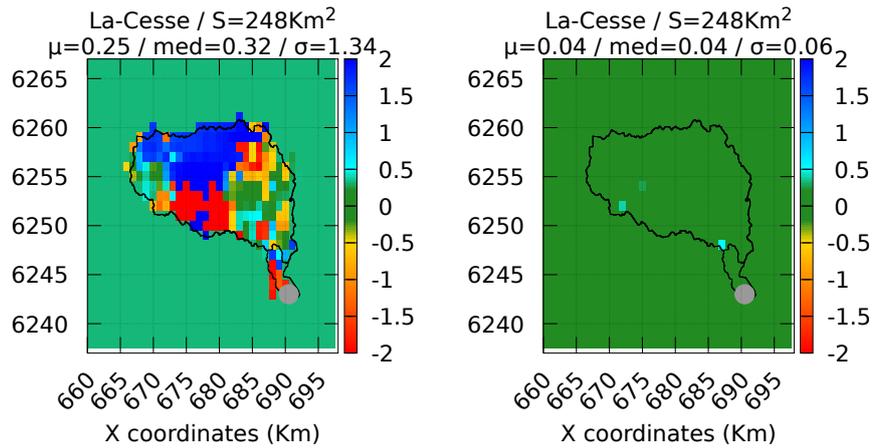


FIGURE 7.13 – Bassin-versant de la Cesse. Comparaison des capacités des réservoirs de production c_p obtenues après calage sur la période P1 entre les expériences G1 et D (graphique de gauche) et les expériences G4 et D (graphique de droite) à l'aide du critère θ présenté à l'équation 6.7. Les moyennes μ , les médianes med et les écarts-types σ sont calculés sur l'aire du bassin-versant.

Conclusion : L'utilisation de la fonction objectif Gamma pour le calage des paramètres permet d'améliorer la robustesse des paramètres (notamment les capacités des réservoirs de transfert et les paramètres de routage). Les performances prédictives spatio-temporelles du modèle sont également légèrement améliorées. Ces résultats sont exploratoires mais encourageants. En effet, cette fonction coût est paramétrable et permet de filtrer les débits observés ou simulés en minimisant leurs influences lorsque leurs valeurs sont trop extrêmes. L'amélioration de cette fonction objectif doit être considérée : celle-ci pourrait servir à améliorer la robustesse du calage et exclure automatiquement les données observées aberrantes susceptibles de corrompre le calage. Une telle fonction pourrait permettre de s'affranchir d'une critique minutieuse des données (l'identification de données aberrantes est difficile, fastidieuse et conduit parfois à exclure la totalité des données au lieu des seules valeurs douteuses). Des tests complémentaires doivent être réalisés. Il serait notamment intéressant de combiner l'utilisation de la fonction coût Gamma avec les contraintes spatiales apportées par des descripteurs.

7.3 Conclusions

Dans ce dernier chapitre, deux techniques ont été expérimentées pour tenter d'améliorer la robustesse du calage distribué des paramètres du modèle GRD.

D'abord l'utilisation de descripteurs (de nature climatique, géologique, pédologique, topographique ou lié à l'occupation des sols) permet de contraindre la variabilité spatiale des paramètres du modèle tout en réduisant drastiquement la taille du vecteur de contrôle. L'utilisation des descripteurs "densité de drainage" et "pente" ont permis d'améliorer la représentation des processus hydrologiques simulés par le modèle. Bien que cette approche n'a pas permis d'améliorer la robustesse des paramètres trouvés, les performances prédictives du modèle le sont. De plus cette méthode ouvre des perspectives intéressantes pour espérer régionaliser les paramètres à l'extérieur des bassins-versants jaugés.

La deuxième technique se base sur une fonction coût paramétrable construite pour limiter l'influence des valeurs extrêmes observées et simulées sur le calage. Bien que la conception de cette fonction doit être affinée et que son utilisation dans ce contexte est exploratoire, la robustesse du calage est sensiblement améliorée.

Des tests complémentaires doivent être réalisés. Ces tests doivent permettre d'améliorer le choix des descripteurs. Des expériences combinant l'utilisation de la fonction objectif Gamma avec les meilleurs descripteurs pourraient ensuite être menées.

Chapitre 8

Conclusion générale

Les évolutions du modèle hydrologique distribué, trois variantes de routage proposées :

Le modèle hydrologique distribué GRD a été amélioré pour rendre la modélisation totalement distribuée. En effet dans les précédentes versions du modèle, les processus de routage vers les exutoires des bassins-versants étaient représentés de manière globale (c'est à dire un routage direct de la maille vers l'exutoire). Désormais, le routage est effectué de l'amont vers l'aval et de maille en maille (Jay-Allemand et al., 2019; Organde et al., 2018). Trois variantes de routage représentant les processus de transfert hydraulique entre chaque maille ont été développés. Ces modèles sont conceptuels et se distinguent par leurs complexités (linéaire et non-linéaire) et par les concepts introduits (modélisation d'un retard pur ou vidange d'un réservoir).

Tout d'abord les trois modèles de routage ont été testés et évalués sur un maillage unidimensionnel. Leurs aptitudes à représenter les phénomènes de propagation hydraulique ont été comparées avec le simulateur SIC résolvant les équations de l'hydraulique à surface libre (équations de Barré de Saint-Venant) dans un bief large et de géométrie trapézoïdale. Ces tests ont montré que le modèle simulant un retard pur et non-linéaire en fonction du débit (variante 2) permettait de bien représenter les phénomènes de propagation des crues.

Lors de modélisations hydrologiques à l'aide de données réelles, les trois variantes de routage suggérées mènent à des résultats très similaires. En effet couplé au modèle hydrologique, aucun de ces trois modèles n'apporte un gain de performance quant à la prédiction des débits aux exutoires des bassins-versants. Premièrement, les processus de routage au sein des bassins-versants ne sont pas de nature purement hydraulique. Deuxièmement, des compensations avec les autres composantes du modèle hydrologique (notamment le module de transfert) pourraient induire des problèmes d'équifinalité. Des expériences complémentaires pourraient être réalisées afin de départager ces trois variantes. Les résultats devraient être analysés en fonction des caractéristiques des bassins-versants modélisés (surfaces, formes et pentes par exemple).

L'intérêt de l'analyse variationnelle pour le calage des modèles hydrologiques :

Un algorithme de calage variationnel a été implémenté afin d'estimer la valeur des paramètres du modèle hydrologique distribué. Cet algorithme nécessite le calcul du gradient d'une fonction objectif par rapport à l'ensemble des paramètres du modèle. Ce gradient est calculé à l'aide du modèle adjoint. Le modèle adjoint a été généré avec l'outil de différentiation automatique Taped à partir du code source de l'algorithme du modèle hydrologique. Ce code source a été optimisé et pensé pour permettre sa différentiation automatique. Par exemple, les trois modèles de routage ont été conçus de façon à respecter les conditions nécessaires de différentiabilité.

Dans ces travaux de recherche cet algorithme d'estimation variationnelle a été testé pour identifier les paramètres du modèle hydrologique distribué. Les paramètres distribués doivent permettre de représenter la variabilité spatiale des caractéristiques hydrologiques des bassins-versants. Les bénéfices d'un tel calage pour la prédiction des débits à l'intérieur des bassins-versants jaugés ont été étudiés. Une méthodologie de calage-validation croisée a été appliquée afin d'évaluer les performances prédictives du modèle. L'exercice de calage est réalisé dans les conditions les plus difficiles : les paramètres sont estimés à l'aide des données de débits observés aux exutoires situés en aval de 23 groupes de bassins-versants. Dans ces groupes de bassins, seul le bassin le plus à l'aval a servi au calage du modèle. Les performances prédictives du modèle sont alors évaluées à l'intérieur des 23 groupes de bassins-versants, aux exutoires des sous-bassins de validation dont les données de débit n'ont pas servi au calage des paramètres (performances prédictives spatiales). Ces performances sont également évaluées sur des périodes temporelles différentes de celles choisies pour le calage des paramètres (performances prédictives temporelles).

La méthodologie de calage implique deux étapes distinctes. La première étape consiste à caler une valeur uniforme de chaque paramètre sur l'ensemble des bassins-versants. Lors de cette première étape l'utilisation d'un algorithme de recherche global est recommandée. Cette phase est déterminante. Elle conditionne le succès de la procédure de calage car elle permet d'évaluer la valeur d'ébauche qui est le point de départ du calage variationnel. La deuxième étape consiste à caler les paramètres de façon distribuée à l'aide de l'algorithme variationnel : pour chaque maille (1 km^2) les valeurs des paramètres du modèle (trois paramètres par maille) sont identifiées.

Le calage distribué à l'aide de l'algorithme variationnel a permis de créer une variabilité spatiale des paramètres à l'intérieur de chaque bassin-versant. Les performances prédictives du modèle utilisant un jeu de paramètres variables spatialement (deuxième

étape du calage) ont été comparées aux performances prédictives du modèle utilisant un jeu de paramètres uniformes spatialement (première étape du calage). L'utilisation d'un jeu de paramètres variables spatialement apporte un gain significatif en termes de performances prédictives spatio-temporelles. Par exemple un score de performance NSE supérieur à 0.6 est obtenu pour 60% des exutoires de validation spatio-temporelle contre 50% lorsqu'un jeu de paramètres uniformes est utilisé. Ces résultats montrent qu'il est bénéfique de réaliser un calage distribué des paramètres par rapport à un calage uniforme (Jay-Allemand et al., 2019). La variabilité spatiale des paramètres à l'intérieur des bassins-versants semble donc représentative de certaines caractéristiques hydrologiques. Les performances ne sont pas identiques selon la localisation des exutoires de validation au sein des bassins-versants. Les débits aux exutoires de validation les plus éloignés des exutoires de calage restent plus difficiles à modéliser. En ce sens, le calage distribué pourrait être significativement amélioré. Il reste que pour 20% des exutoires de validation, les déficiences du modèle ou la qualité médiocre des données de débits ne permettent pas de modéliser ou de valider correctement les débits.

Des solutions proposées pour mieux contraindre le calage mais des progrès encore à accomplir :

Le calage distribué est largement perfectible. Les jeux de paramètres trouvés sont très variables spatialement et très différents selon la période d'analyse choisie. Ces problèmes d'identification sont provoqués par les jeux de données (erreurs de mesure de la pluie et des hauteurs d'eau, courbes de tarage imprécises), par l'absence de relation précise entre les paramètres du modèle et les variables hydrologiques (les paramètres du modèle ont un sens physique très limité) et par les imprécisions du modèle.

Afin d'améliorer le calage distribué des paramètres plusieurs pistes ont été étudiées : dans un premier temps l'utilisation d'une fonction de régularisation de type Tikhonov a été testée pour contraindre le calage et améliorer le conditionnement ; puis des descripteurs (de nature climatique, géologique, pédologique...) ont été utilisés pour contraindre la variabilité spatiale des paramètres ; enfin l'utilisation d'une fonction objectif dérivant des méthodes d'estimation "robustes" a été expérimentée.

Les contraintes de régularisation type Tikhonov imposent que les valeurs des paramètres varient graduellement dans l'espace. Ces contraintes ne permettent ni d'améliorer les performances prédictives du modèle ni de stabiliser les jeux de paramètres trouvés entre les deux périodes d'analyse choisies. Ce type de contraintes ne semble donc pas approprié au problème puisqu'il ne décrit pas la variabilité spatiale des propriétés hydrologiques (type de sol, pente...).

Des contraintes spatiales ont été apportées à partir de relations entre des descripteurs

physiques et les paramètres du modèle. Le choix des descripteurs est difficile. En effet, le modèle est conceptuel et les paramètres ne représentent pas nécessairement un processus physique précis. Pour cela, plusieurs combinaisons de descripteurs ont été testées. Globalement les résultats sont positifs. Les performances prédictives spatio-temporelles du modèle utilisant les jeux de paramètres distribués contraints par ces descripteurs sont améliorées. Cela signifie que ces jeux de paramètres sont davantage représentatifs du fonctionnement des bassins-versants. En plus d'améliorer les performances, cette méthode permettra peut-être de mieux régionaliser les paramètres à l'extérieur des bassins-versants jaugés. Cependant, les jeux de paramètres calés restent encore très différents selon la période d'analyse choisie et manquent de robustesse.

L'existence d'erreurs significatives sur les données observées et sur les débits simulés peuvent rendre le calage difficile et potentiellement instable. Les méthodes d'estimation robustes sont basées sur une fonction objectif qui permet d'exclure une proportion de valeurs extrêmes et erronées. Une telle méthode a été testée. Les jeux de paramètres trouvés sont sensiblement plus robustes et ont engendré un léger gain de performances prédictives spatio-temporelles du modèle. Ces résultats sont exploratoires et cette fonction objectif pourrait être davantage testée et améliorée.

Le problème du manque de robustesse des jeux de paramètre n'est pas complètement lié au calage distribué. Il est aussi présent pour des modélisations globales et provient du fait que les jeux de paramètres identifiés sont très dépendants du choix des données et de la période d'analyse. Des travaux complémentaires doivent être menés pour résoudre ce problème.

Par exemple, l'utilisation de descripteurs doit être approfondie. D'autres descripteurs ou combinaisons de descripteurs pourraient être testées et leurs degrés de précision devraient être étudiés.

L'utilisation d'une fonction objectif robuste et paramétrable est une piste intéressante. Son développement spécifique pour le calage des paramètres d'un modèle hydrologique doit être poursuivie. Celle-ci devrait être comparée à d'autres fonctions objectifs telle que le KGE.

La structure du modèle hydrologique pourrait être améliorée afin d'y intégrer la modélisation de nouveaux processus et limiter les incertitudes de modélisation.

La procédure de calage-validation utilisée pour les expériences numériques manque de signification statistique. En effet celle-ci est basée sur seulement deux périodes temporelles de cinq ans contenant des données différentes. Premièrement, une fenêtre d'analyse plus longue pourrait être considérée (augmentation du nombre de données). Cependant, celle-ci resterait limitée par la disponibilité des observations (pluies radar) et par les ressources de calcul nécessaires (temps et mémoire vive). Deuxièmement, la procédure de calage-validation pourrait être enrichie en considérant de multiples combinaisons aléatoires de

données utilisées soit pour le calage soit pour la validation. Cette nouvelle procédure donnerait aux résultats davantage de signification statistique et permettrait sûrement d'affiner les conclusions concernant la mise en œuvre du calage distribué, les choix des descripteurs et les choix de la fonction objectif à utiliser.

Des perspectives de recherches intéressantes pour la régionalisation et la prévision des débits :

L'identification ou le calage des paramètres à l'extérieur des bassins-versants jaugés (sans observation de débits) est aussi un challenge. Une telle estimation serait extrêmement bénéfique aux modèles de prévision des crues voulant fonctionner en sites non jaugés. L'utilisation de descripteurs permet de contraindre spatialement le calage. De ce fait, les paramètres du modèle sont automatiquement estimés à l'extérieur des bassins-versants jaugés. Bien que les relations entre les descripteurs et les paramètres du modèle hydrologique doivent être établies plus précisément, cette méthode offre des perspectives intéressantes. Il serait intéressant de construire des expériences de calage en étudiant les performances prédictives du modèle en validation sur des bassins-versants voisins (et non des sous-bassins). Avec une telle approche, les données disponibles aux exutoires de plusieurs bassins-versants jaugés distants géographiquement pourraient servir à identifier les paramètres d'une région. Cela pourrait contribuer à stabiliser les jeux de paramètres (augmentation du nombre de données) et peut-être d'affiner le choix des descripteurs à utiliser.

Les jeux de paramètres distribués devraient être testés dans le cadre du service Vigicrues-Flash afin d'évaluer les gains de performances potentiels en fonctionnement opérationnel. Pour cela, les paramètres distribués du modèle doivent être fixés (choix d'un jeu de paramètres distribués au détriment d'un autre). Sinon il serait possible de considérer plusieurs jeux de paramètres issus de calages différents. Dans ce cas, plusieurs scénarios de prévision pourraient être prédits (ensemble de simulations). Le choix du scénario le plus probable pourrait être affiné au cours du temps soit par le gestionnaire soit par des méthodes de détection automatique tel que le "Tree-Based Model Predictive Control".

Pour la suite, des travaux de recherche pourraient cibler le fonctionnement en temps réel du modèle. Dans ce cas, les variables d'états du modèle (niveaux initiaux des réservoirs de production et de transfert) peuvent être réajustées en temps réel dès qu'un nouvel ensemble d'observations est disponible. Il s'agit alors d'appliquer un processus classique d'assimilation de données. Le même algorithme variationnel doit être utilisé. En revanche la fenêtre temporelle d'analyse sera plus courte. L'assimilation en temps-réel des états du modèle offre des perspectives intéressantes pour fournir et améliorer les prévisions des

débits à courtes échéances. Cette structure d'assimilation est d'ailleurs déjà fonctionnelle dans le modèle GRD.

Annexe A

Bassins-versants et positionnement des exutoires sur la grille d'écoulement $1km^2$ pour la modélisation des débits par GRD

Les figures suivantes représentent les contours des bassins-versants étudiés dans le cadre de PICS. Les positionnements des exutoires (réels et modélisés) sont tracés sur les cartes. Le fond de carte inclue, les surfaces drainées cumulées ainsi que le réseau hydrographique.

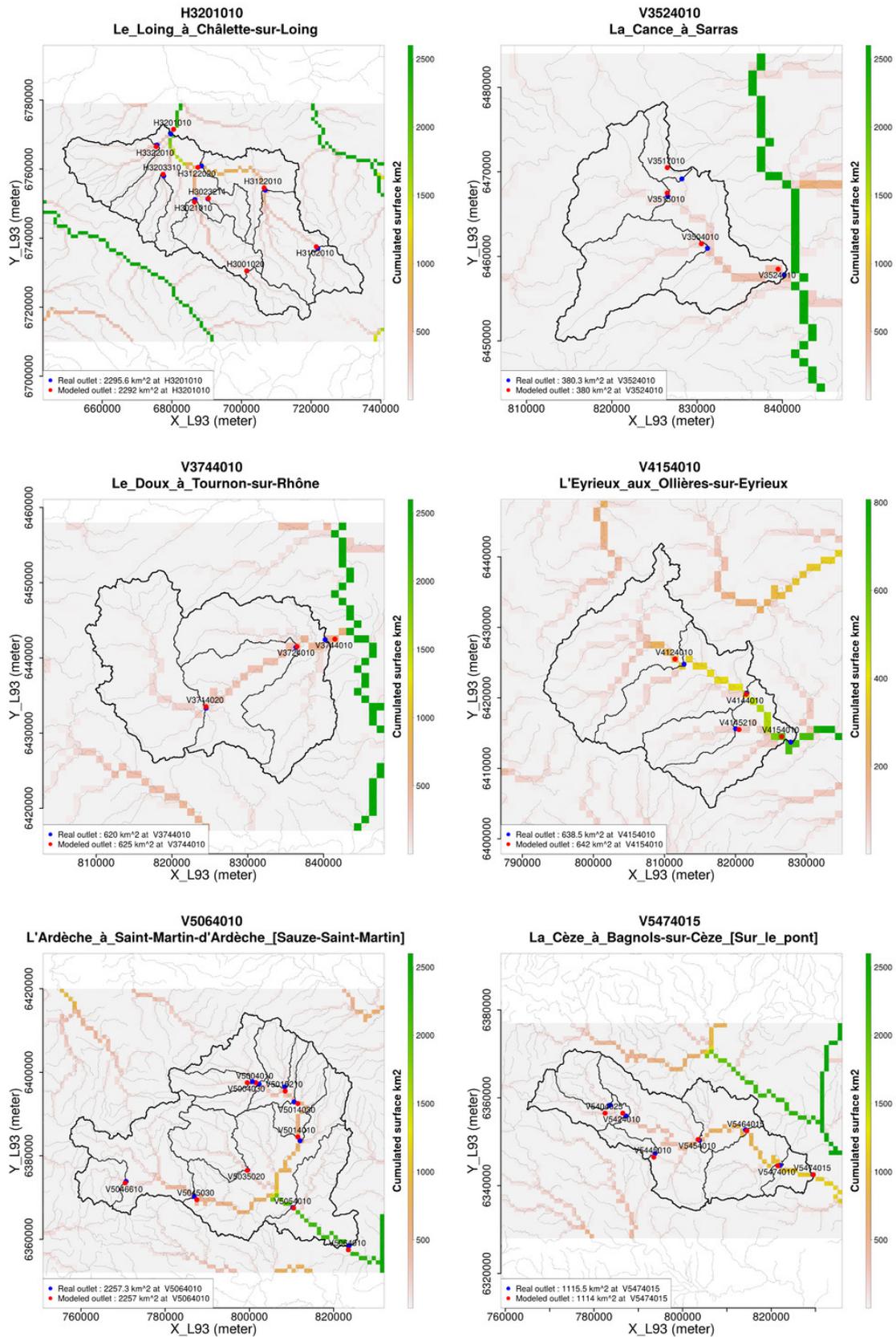


FIGURE A.1 – Contours des bassins-versants étudiés et positionnement des exutoires.

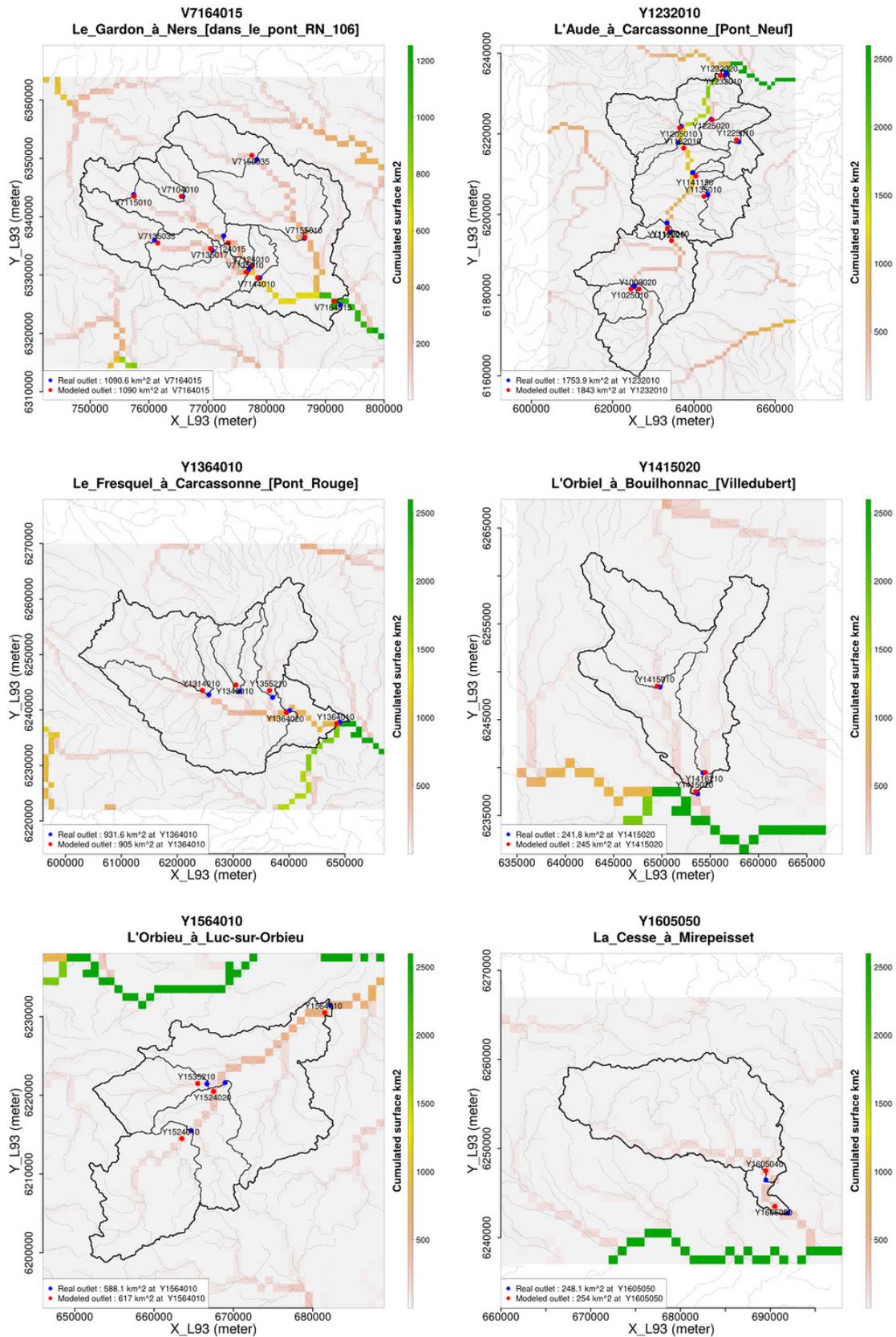


FIGURE A.2 – Contours des bassins-versants étudiés et positionnement des exutoires.

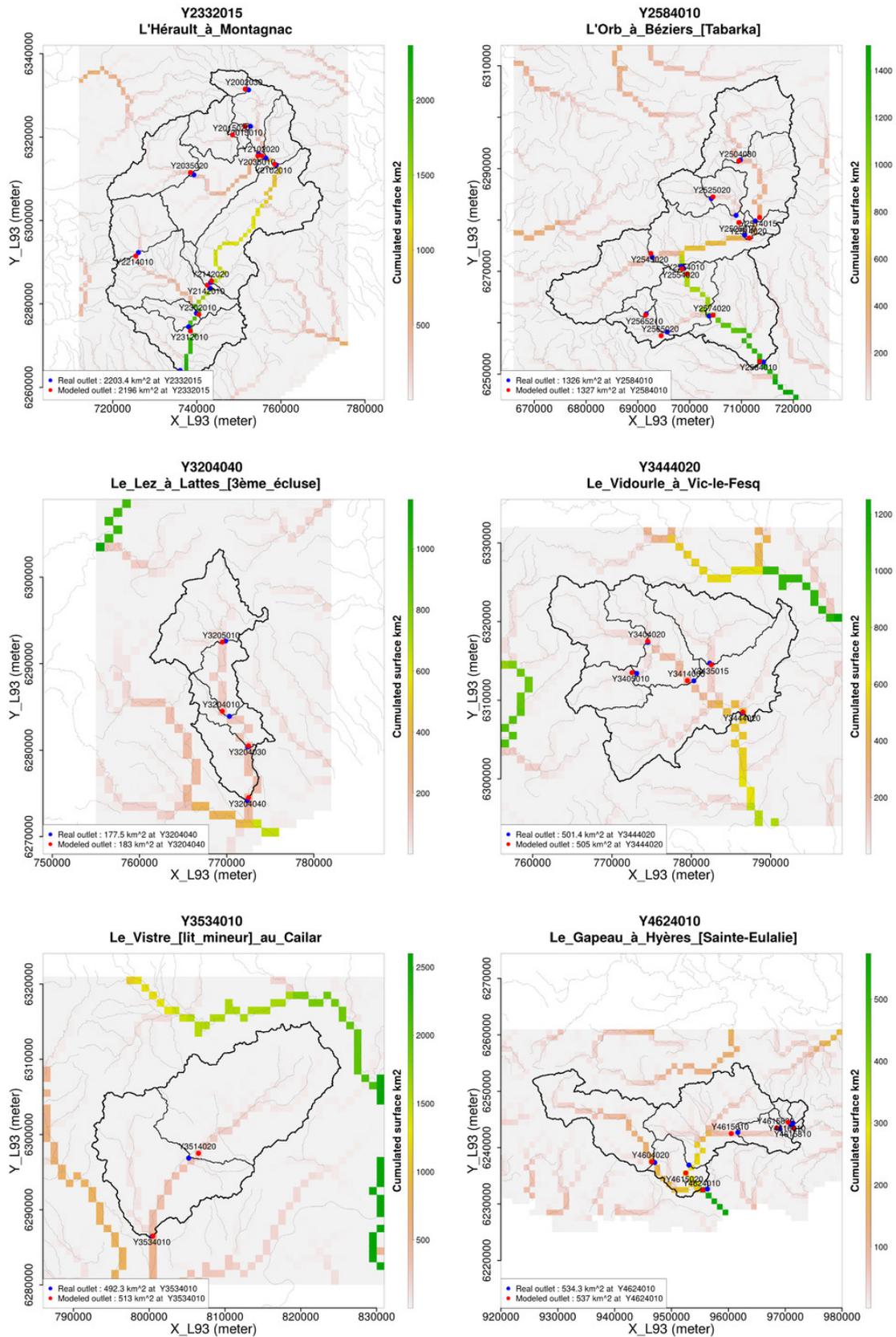


FIGURE A.3 – Contours des bassins-versants étudiés et positionnement des exutoires.

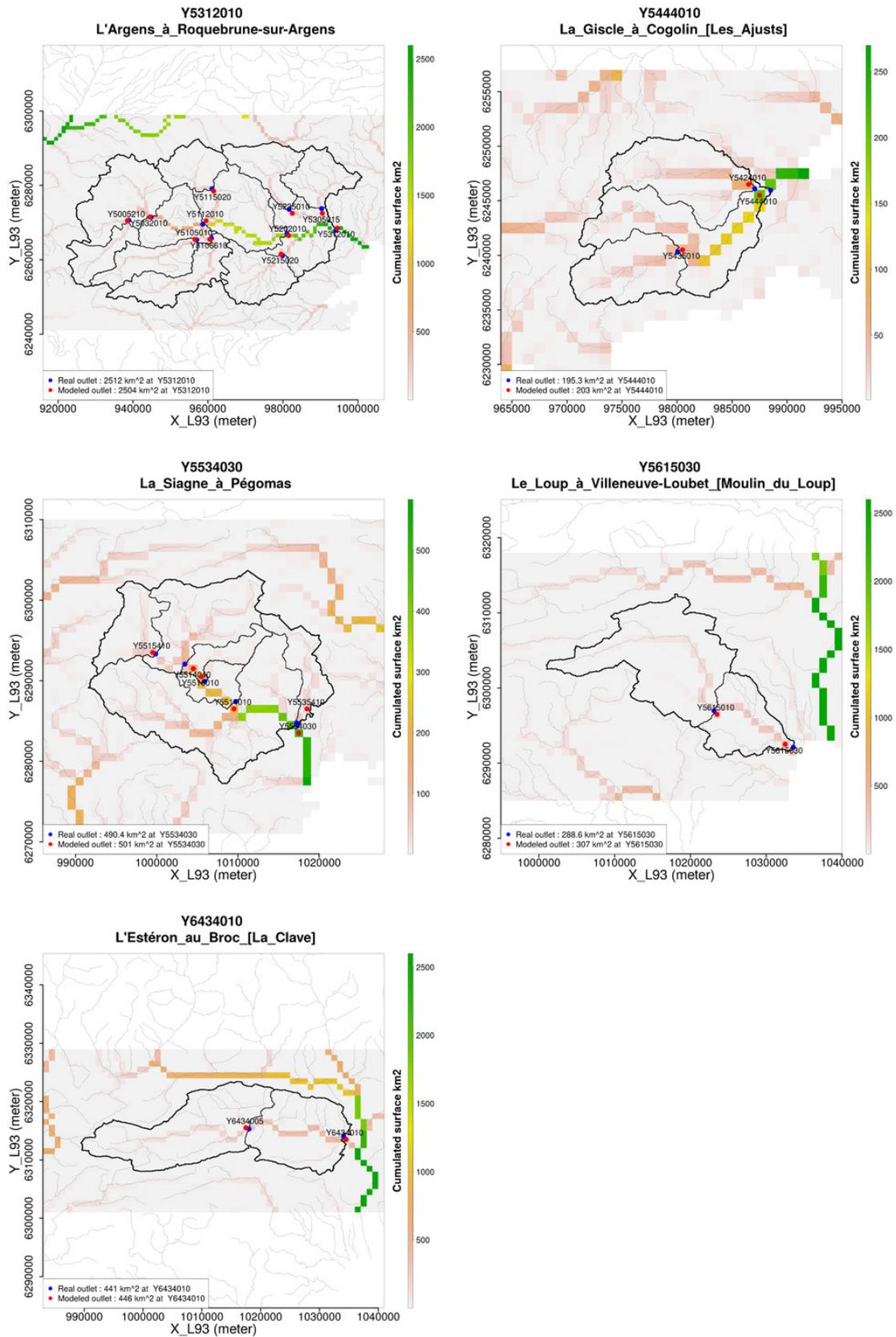


FIGURE A.4 – Contours des bassins-versants étudiés et positionnement des exutoires.

Annexe B

Tableaux des 141 stations hydrométriques sélectionnées dans le cadre de PICS et réparties sur 23 grands bassins-versants

Code	Nom station	Surface km^2
code	name	surfirstea
H3001020	Le Loing à Saint-Martin-des-Champs [Le Moulin de la Prairie]	126.09
H3021010	Le Loing à Montbouy	409.42
H3023211	L'Aveyron à la Chapelle-sur-Aveyron	97.3875
H3102010	L'Ouanne à Toucy	153.645
H3122010	L'Ouanne à Charny	561.53
H3122020	L'Ouanne à Gy-les-Nonains	875.105
H3201010	Le Loing à Châlette-sur-Loing	2295.5775
H3203310	Le Puiseaux à Saint-Hilaire-sur-Puiseaux	87.0875
H3322010	La Bezonde à Pannes	338.275
V3504010	La Cance à Annonay	124.565
V3515010	La Deume à Saint-Julien-Molin-Molette [La Garinière]	106.3725
V3517010	Le Ternay à Savas [Ternay]	25.1975
V3524010	La Cance à Sarras	380.26
V3714020	Le Doux à Lamastre	242.835
V3724010	Le Doux à Colombier-le-Vieux	379.5525
V3744010	Le Doux à Tournon-sur-Rhône	620.015
V4124010	L'Eyrieux au Cheylard	268.7675

Code	Nom de la station	Surface km^2
V4144010	L'Eyrieux à Beauvène [Pont de Chervil]	452.0825
V4145210	La Glueyre à Gluiras [Tisonèche]	70.7175
V4154010	L'Eyrieux aux Ollières-sur-Eyrieux	638.5325
V5004010	L'Ardèche à Pont-de-Labeaume	291.6725
V5004030	L'Ardèche à Meyras [Pont Barutel]	98.9125
V5014010	L'Ardèche à Vogüé	621.4225
V5014030	L'Ardèche à Ucel	476.2625
V5015210	La Volane à Vals-les-Bains	108.995
V5035020	La Beaume à Rosières	199.8675
V5045030	Le Chassezac à Gravières	495.8475
V5046610	L'Altier à Altier [La Goulette]	103.3625
V5054010	L'Ardèche à Vallon-Pont-d'Arc	1957.255
V5064010	L'Ardèche à Saint-Martin-d'Ardèche [Sauze-Saint-Martin]	2257.3375
V5404025	La Cèze à Gaujac [Barrage de Sénéchas]	113.35
V5424010	La Cèze à Bessèges	228.0925
V5445010	L'Auzonnet aux Mages	42.625
V5454010	La Cèze à Tharoux	663.5625
V5464015	La Cèze à Montclus	835.02
V5474010	La Cèze à la Roque-sur-Cèze	1048.49
V5474015	La Cèze à Bagnols-sur-Cèze [Sur le pont]	1115.4875
V7104010	Le Gardon de Saint-Martin à Saint-Étienne-Vallée-Française [Roq.]	30.4425
V7115010	Le Gardon de Sainte-Croix à Gabriac [Pont Ravagers]	48.2675
V7124010	Le Gardon de Mialet à Générargues [Roucan]	244.085
V7124015	Le Gardon de Mialet à Mialet [en amont du camping]	219.7475
V7135010	Le Gardon de Saint-Jean à Corbès [Roc Courbe]	260.5
V7135017	Le Gardon de Saint-Jean à Saint-Jean-du-Gard [Bâtiment communal]	157.715
V7135035	Le Gardon à Saumane	97.515
V7144010	Le Gardon d'Anduze à Anduze [sous le pont du train]	540.79
V7155010	Le Gardon d'Alès à Alès [Pont Vieux]	315.2175
V7155035	Le Gardon à Sainte-Cécile-d'Andorge [Barrage 2 Ste Cécile d'Andorge]	115.385
V7164015	Le Gardon à Ners [dans le pont RN 106]	1090.5675
Y1002020	L'Aude à Escouloubre [aval]	182.615
Y1025010	La Bruyante à Escouloubre [Usson-les-Bains]	91.345
Y1112010	L'Aude à Belvianes-et-Cavirac	641.1
Y1112020	L'Aude à Quillan [OA RD 118]	668.5675
Y1135010	La Salz à Cassaignes	131.035

Code	Nom de la station	Surface km^2
Y1141150	Le Ruisseau de Lavalette à Alet-les-Bains [Alet-les-Bains]	15.25
Y1152010	L'Aude à Limoux [Pont de Fer]	1164.3175
Y1205010	Le Sou à Saint-Martin-de-Villereglan	196.8575
Y1225010	Le Lauquet à Greffeil	68.34
Y1225020	Le Lauquet à Saint-Hilaire [Le Pech]	176.64
Y1232010	L'Aude à Carcassonne [Pont Neuf]	1753.935
Y1232020	L'Aude à Carcassonne [Pont Vieux]	1749.5175
Y1314010	Le Fresquel à Villepinte	212.95
Y1345010	Le Lampy à Raissac-sur-Lampy	54.8175
Y1355210	La Rougeanne à Moussoulens	128.215
Y1364010	Le Fresquel à Carcassonne [Pont Rouge]	931.5625
Y1364020	Le Fresquel à Pezens	845.425
Y1415010	L'Orbiel à Lastours	87.2025
Y1415020	L'Orbiel à Bouilhonnac [Villedubert]	241.7525
Y1416210	La Clamoux à Malves-en-Minervois	84.04
Y1524010	L'Orbieu à Saint-Martin-des-Puits	171.05
Y1524020	L'Orbieu à Lagrasse [OA RD 3]	264.2625
Y1535210	Le ruisseau le sou à Lagrasse	75.3325
Y1564010	L'Orbieu à Luc-sur-Orbieu	588.15
Y1605040	La Cesse à Bize-Minervois	235.9875
Y1605050	La Cesse à Mirepeisset	248.0925
Y2002030	L'Hérault à Valleraugue	42.4075
Y2015010	L'Arre au Vigan [La Terrisse]	154.29
Y2015020	L'Arre au Vigan	133.145
Y2035010	La Vis à Saint-Laurent-le-Minier	499.79
Y2035020	La Vis à Blandas [Lafoux de la Vis]	222.985
Y2102010	L'Hérault à Laroque	911.83
Y2102020	L'Hérault à Ganges	808.23
Y2142010	L'Hérault à Gignac [aval]	1408.715
Y2142020	L'Hérault à Gignac	1380.3075
Y2214010	La Lergue à Lodève	181.08
Y2302010	L'Hérault à Canet	1895.05
Y2312010	L'Hérault à Aspiran	1953.61
Y2332015	L'Hérault à Montagnac	2203.4075
Y2504030	L'Orb à Cazilhac [Truscas]	193.27
Y2514015	L'Orb à Bédarieux	344.7925

Code	Nom de la station	Surface km^2
Y2514020	L'Orb à Hérépian	373.435
Y2525010	La Mare au Pradal	114.93
Y2525020	La Mare à Saint-Gervais-sur-Mare	61.2475
Y2545020	Le Jaur à Olargues [amont]	225.1075
Y2554010	L'Orb à Vieussan [Vieussan aval]	904.44
Y2554020	L'Orb à Vieussan [Vieussan amont]	903.3425
Y2565020	Le Vernazobre à Saint-Chinian	66.5975
Y2565210	L'Illouvre [source] à Babeau-Bouldoux [Poussarou]	20.4225
Y2574020	L'Orb à Cessenon-sur-Orb	1107.145
Y2584010	L'Orb à Béziers [Tabarka]	1326.0475
Y3204010	Le Lez à Montferrier-sur-Lez [Lavalette]	114.495
Y3204030	Le Lez à Montpellier [Pont Garigliano]	141.04
Y3204040	Le Lez à Lattes [3ème écluse]	177.525
Y3205010	Le Lez au Triadou [Pont du Lien]	83.3975
Y3404020	Le Vidourle à Conqueyrac [Barrage de Conqueyrac]	84.0675
Y3405010	Le Vidourle [Rieu Massel] à Pompignan [Barrage de Ceyrac]	44.6875
Y3414005	Le Vidourle à Quissac	210.3875
Y3435015	Le Criulon à la Rouvière [Barrage de La Rouvière]	91.805
Y3444020	Le Vidourle à Vic-le-Fesq	501.4225
Y3514020	Le Vistre à Bernis	286.4575
Y3534010	Le Vistre [lit mineur] au Cailar	492.3225
Y4604020	Le Gapeau à Solliès-Pont	186.8625
Y4615020	Le Réal Martin à la Crau [Decapris]	282.0925
Y4615610	Real Collobrier Pond de Fer	70.67
Y4615810	Real Collobrier Valescure	9.2575
Y4615830	Real Collobrier Collobriere	29.355
Y4616010	Real Collobrier Mauret	8.405
Y4624010	Le Gapeau à Hyères [Sainte-Eulalie]	534.3175
Y5005210	Le Cauron à Bras [Pont de l'Avocade]	143.5975
Y5032010	L'Argens à Châteauvert	506.775
Y5105010	Le Caramy à Vins-sur-Caramy [Les Marcounious]	202.34
Y5106610	L'Issole à Cabasse [Pont des Fées]	235.2975
Y5112010	L'Argens à Carcès [aval]	1143.6325
Y5115020	La Bresque à Salernes [Les Vingalières]	167.52
Y5202010	L'Argens aux Arcs	1655.7125
Y5215020	L'Aille à Vidauban [Le Baou]	227.8175

Code	Nom de la station	Surface km^2
Y5235010	La Nartuby à Trans-en-Provence	195.2425
Y5305015	L'Endre au Muy	187.1275
Y5312010	L'Argens à Roquebrune-sur-Argens	2512.045
Y5424010	La Giscle à Cogolin	66.06
Y5435010	La Môle au Lavandou [Destel]	50.64
Y5444010	La Giscle à Cogolin [Les Ajusts]	195.275
Y5514010	La Siagne à Montauroux [Veyans]	220.845
Y5514040	La Siagne à Callian [Ajustadoux]	164.5675
Y5515410	La Siagnole à Mons [Le Moulin]	53.93
Y5516010	Les Tuves [source] à Montauroux	191.835
Y5534030	La Siagne à Pégomas	490.4225
Y5535410	La Mourachonne à Pégomas	42.935
Y5615010	Le Loup à Tourrettes-sur-Loup [Les Vallettes]	206.4925
Y5615030	Le Loup à Villeneuve-Loubet [Moulin du Loup]	288.6225
Y6434005	L'Estéron à Sigale [Pont du Coude]	261.8675
Y6434010	L'Estéron au Broc [La Clave]	441.0025

TABLE B.1 – Liste des 141 exutoires avec stations hydrométriques sélectionnés pour contribuer au projet PICS.

Annexe C

Masques de réduction du vecteur de contrôle

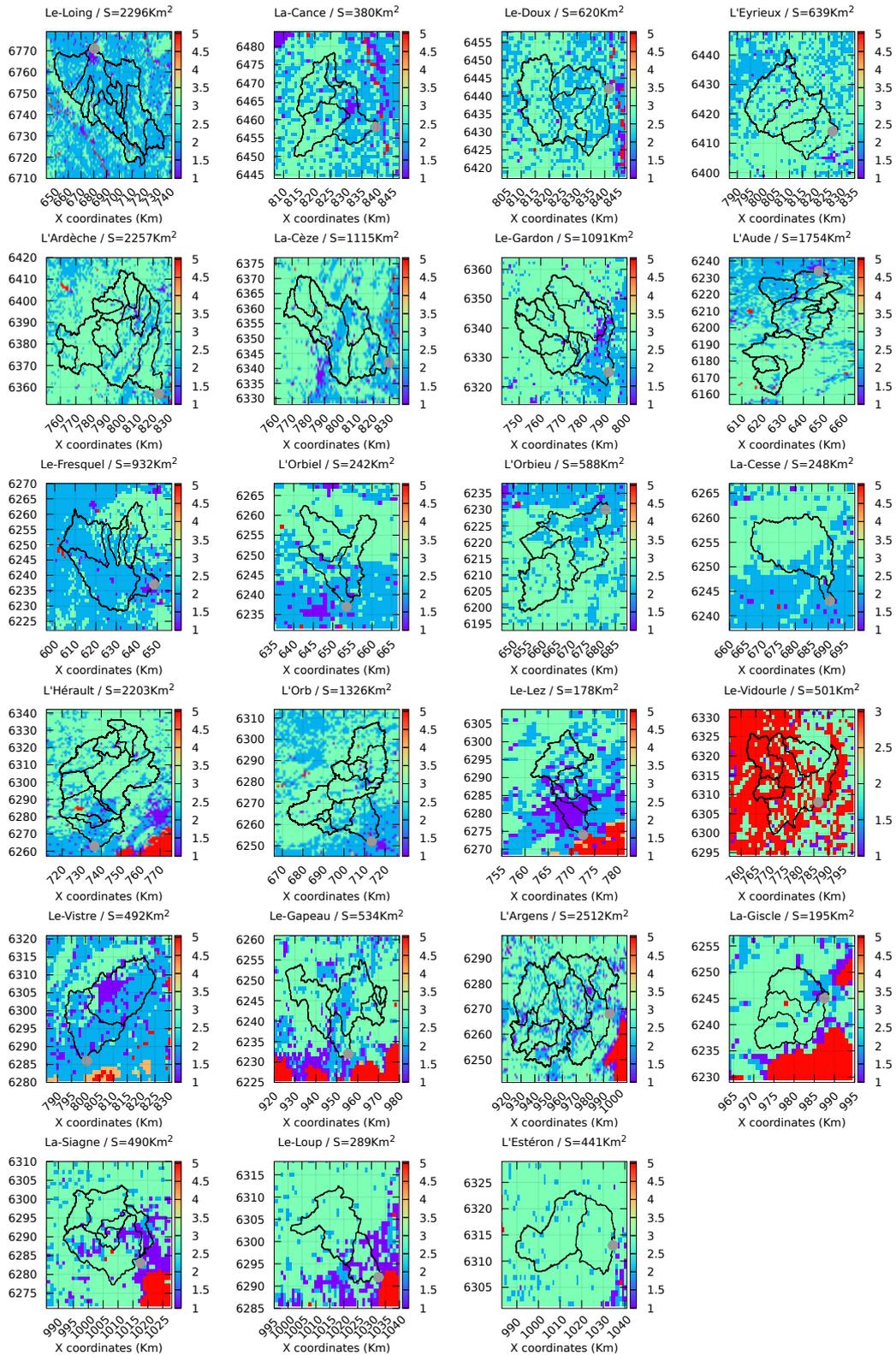


FIGURE C.1 – Masques de réduction du vecteur de contrôle basés sur les données d'occupation du sol pour les 23 bassins-versants et utilisés pour les réservoirs de production.

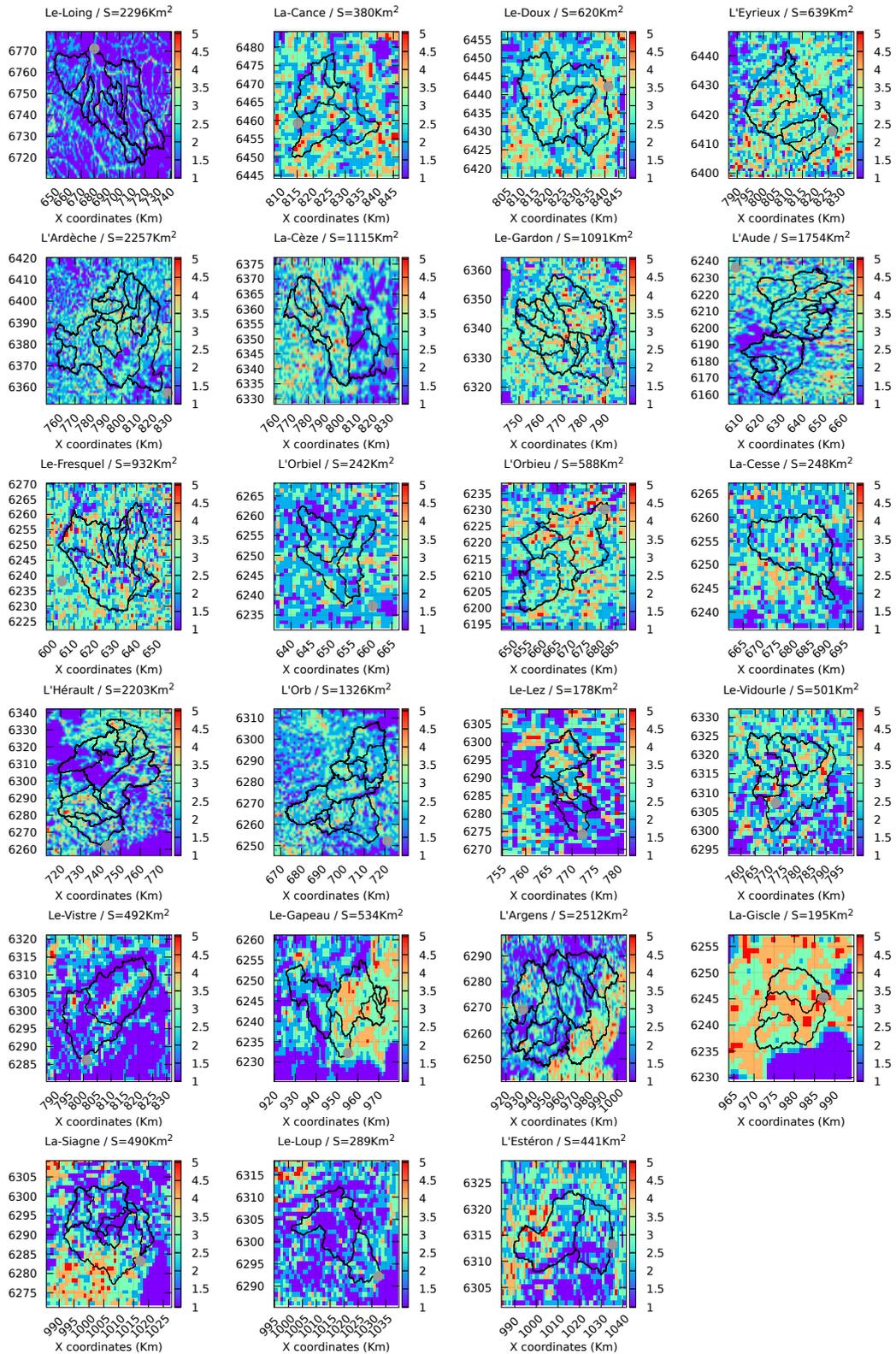


FIGURE C.2 – Masques de réduction du vecteur de contrôle basés sur les données de densité de drainage pour les 23 bassins-versants et utilisés pour les réservoirs de production.

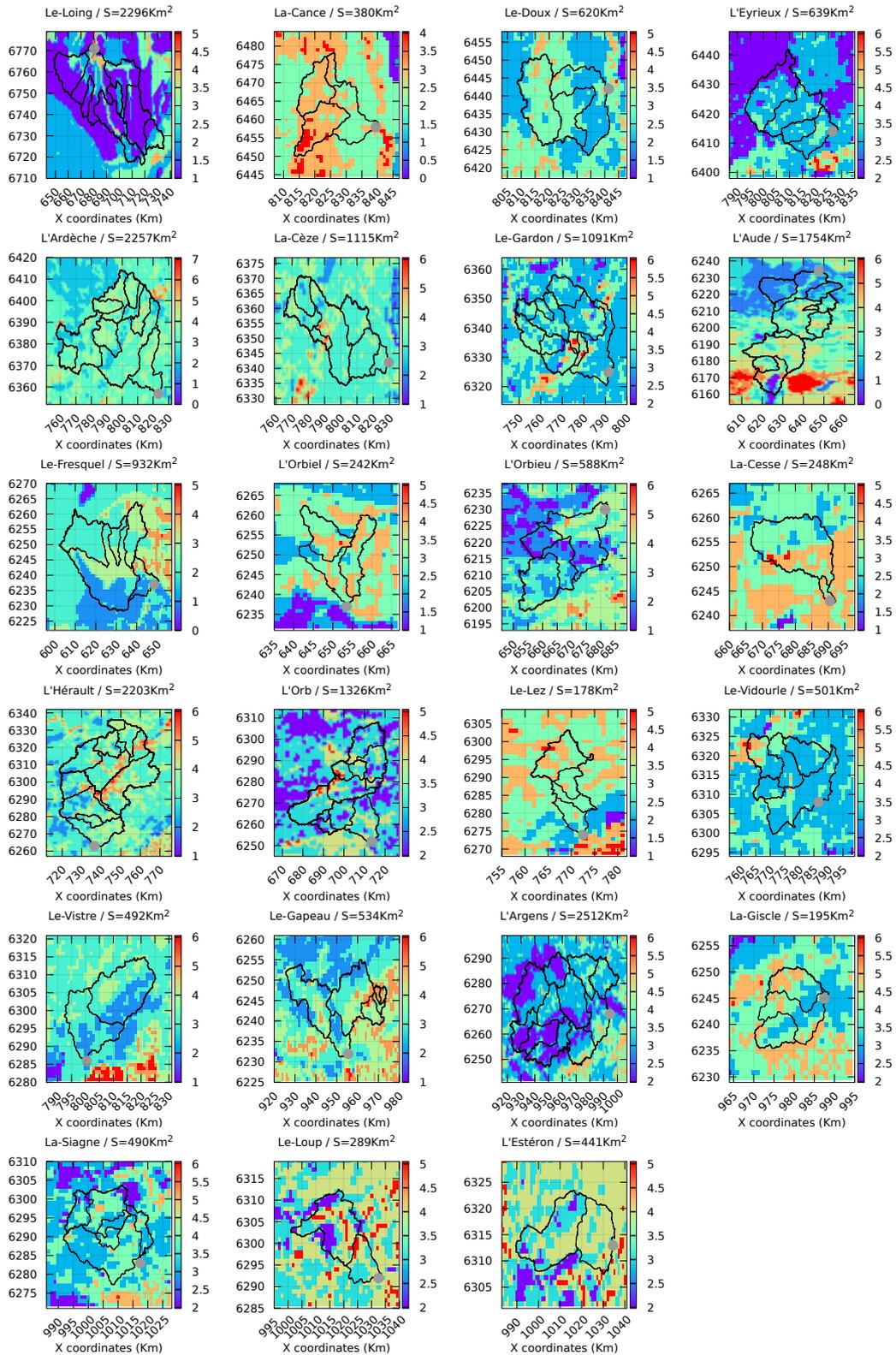


FIGURE C.3 – Masques de réduction du vecteur de contrôle basés sur l'indice mixte de production pour les 23 bassins-versants et utilisés pour les réservoirs de production.

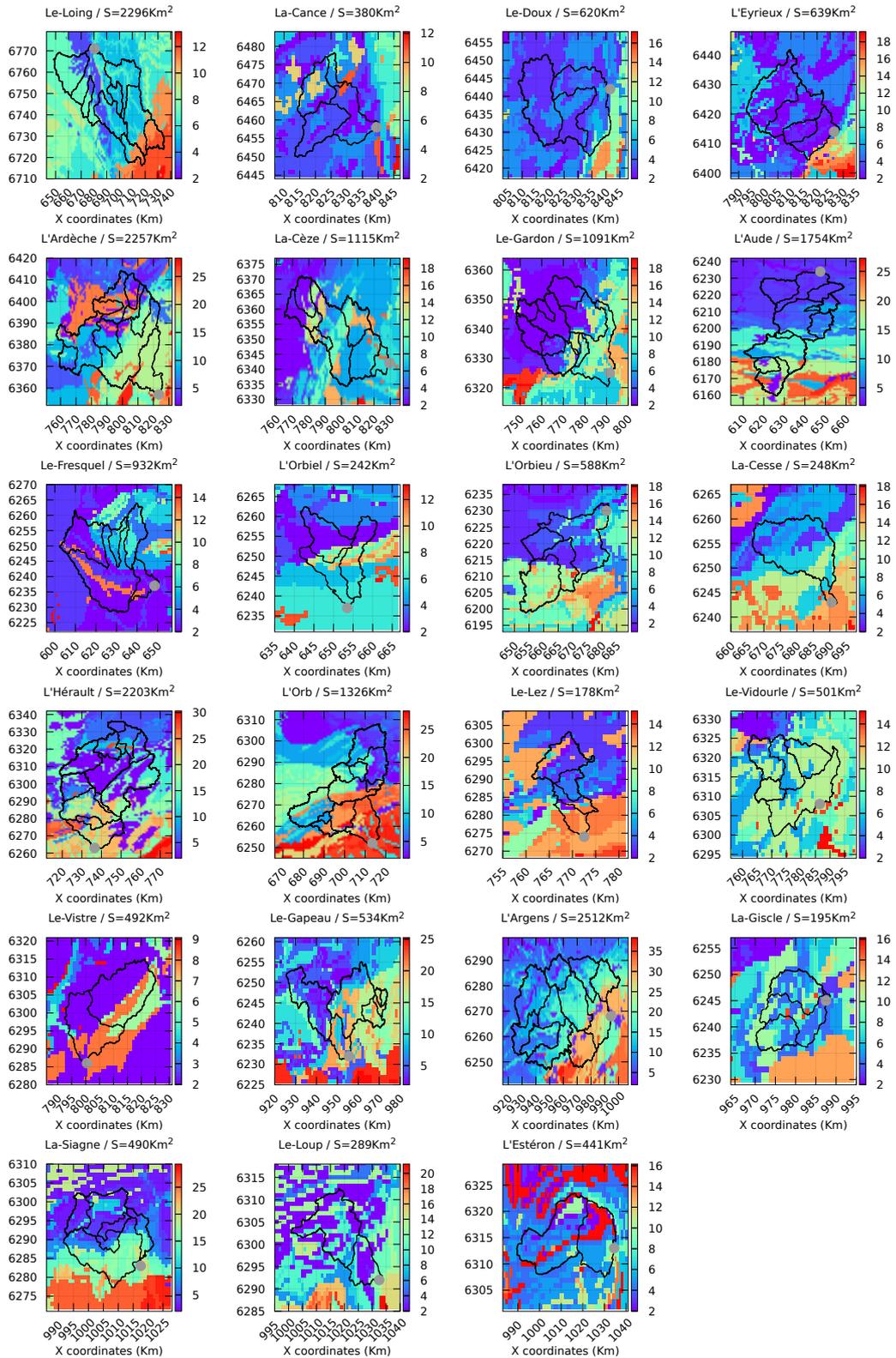


FIGURE C.4 – Masques de réduction du vecteur de contrôle basés sur les données des couches géologiques de surface pour les 23 bassins-versants et utilisés pour les réservoirs de transfert.

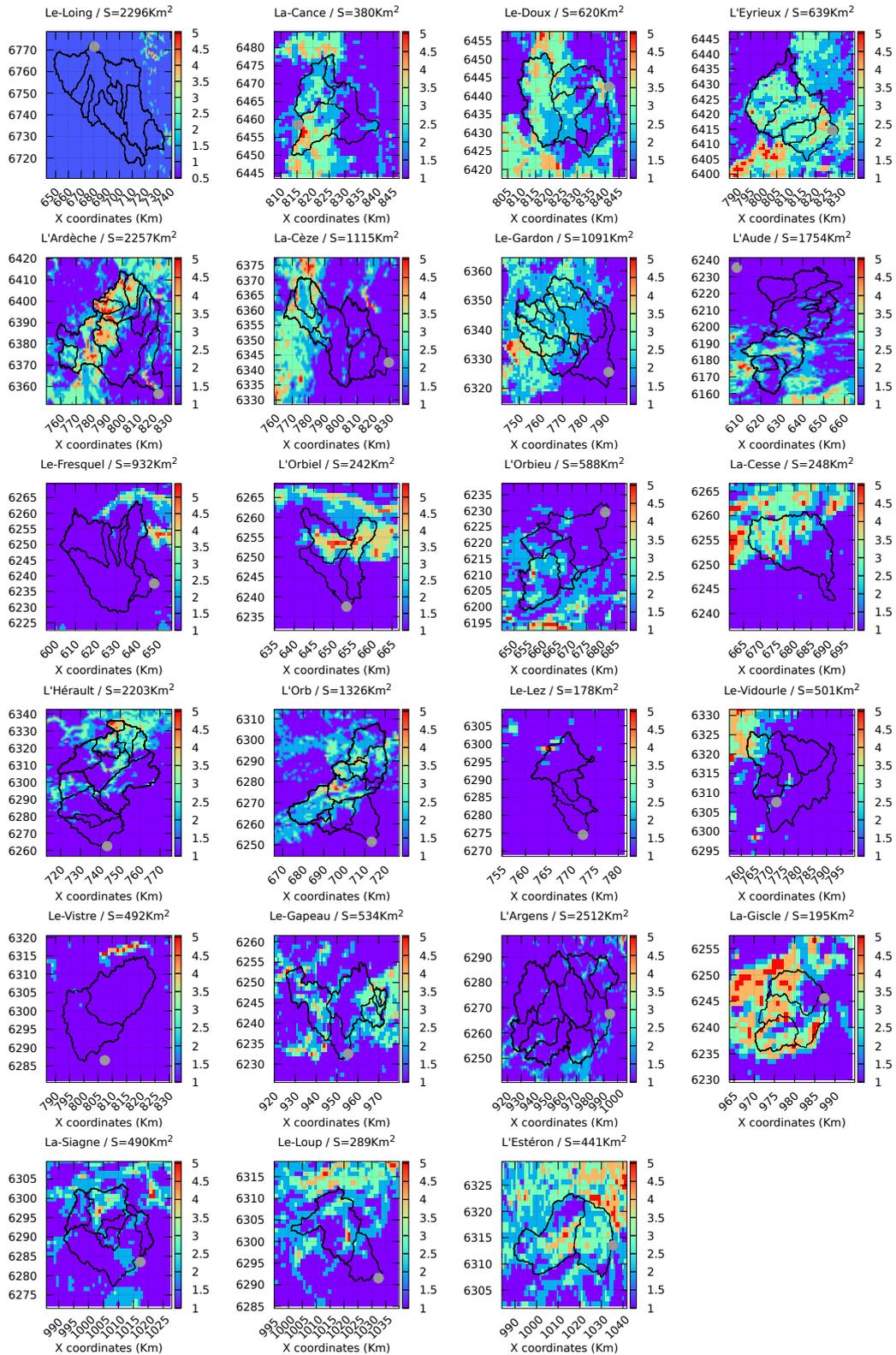


FIGURE C.5 – Masques de réduction du vecteur de contrôle basés sur les données des pentes pour les 23 bassins-versants et utilisés pour les réservoirs de transfert.

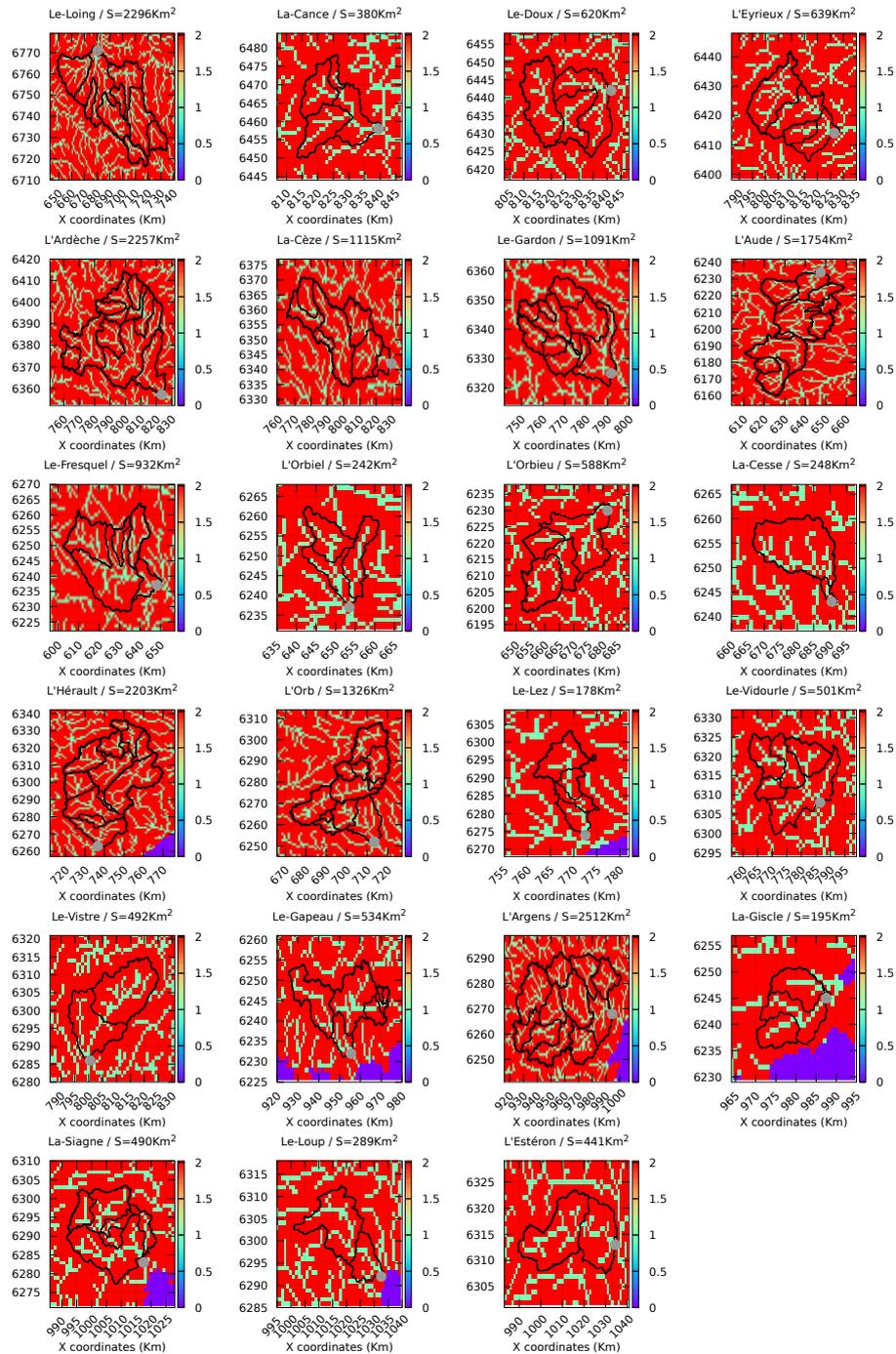


FIGURE C.6 – Masques de réduction du vecteur de contrôle basés sur les surfaces cumulées pour les 23 bassins-versants et utilisés pour les paramètres de routage. La classe 0 correspond aux zones non drainées (zones maritimes); la classe 1 correspond au maille situées sur le réseau hydrographique; la classe 2 correspond aux mailles situées sur les versants.

Bibliographie

- A. Abbaris, H. Dakhlaoui, S. Thiria, and Z. Bargaoui. Variational data assimilation with the yao platform for hydrological forecasting. *Proceedings of the International Association of Hydrological Sciences, Volume 364, 2014, pp.3-8*, 364 :3–8, sep 2014. doi : 10.5194/piahs-364-3-2014.
- K. C. Abbaspour, J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan. Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using swat. *Journal of Hydrology*, 333 :413–430, feb 2007.
- L. Alfieri, P. Salamon, F. Pappenberger, F. Wetterhall, and J. Thielen. Operational early warning systems for water-related hazards in europe. *Environmental Science & Policy*, 21 :35–49, 2012.
- O. M. Alifanov, E. A. Artyukhin, and S. Rumyantsev. *Extreme methods for solving ill-posed problems with applications to inverse heat transfer problem*. Begell house New York, 1995.
- R. M. Anderson, V. I. Koren, and S. M. Reed. Using ssurgo data to improve sacramento model a priori parameter estimates. *Journal of Hydrology*, 320(1) :103 – 116, 2006. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2005.07.020>. URL <http://www.sciencedirect.com/science/article/pii/S0022169405003306>. The model parameter estimation experiment.
- P. Arnaud, J. Lavabre, B. Sol, and C. Desouches. Cartographie de l'aléa pluviographique de la france. *La houille blanche*, (5) :102–111, 2006.
- V. A. Bell, A. L. Kay, R. G. Jones, and R. J. Moore. Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrology and Earth System Sciences Discussions*, 11(1) :532–549, Jan 2007. URL <https://hal.archives-ouvertes.fr/hal-00305636>.
- K. Beven. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1) :41–51, 1993. ISSN 0309-1708. URL <http://www.sciencedirect.com/science/article/pii/030917089390028E>.

- K. Beven. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences Discussions*, 5(1) :1–12, 2001. URL <https://hal.archives-ouvertes.fr/hal-00304564>.
- M. Borga, E. N. Anagnostou, G. Blöschl, and J. D. Creutin. Flash flood forecasting, warning and risk management : the hydrate project. *Environmental Science & Policy*, 14(7) :834–844, 2011. doi : 10.1016/j.envsci.2011.05.017.
- P. Brigode, L. Oudin, and C. Perrin. Hydrological model parameter instability : A source of additional uncertainty in estimating the hydrological impacts of climate change? *Journal of Hydrology*, 476 :410 – 425, 2013. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2012.11.012>. URL <http://www.sciencedirect.com/science/article/pii/S002216941200964X>.
- J.-M. Carrière, P. Brovelli, and R. Tzanos. Apic : service d’avertissement de pluie intense. *Météo : le magazine*, (14) :18–19, 2011.
- W. Castaings, D. Dartus, F.-X. Le Dimet, and G.-M. Saulnier. Sensitivity analysis and parameter estimation for distributed hydrological modeling : potential of variational methods. *Hydrology and Earth System Sciences*, 13(4) :503–517, 2009. doi : 10.5194/hess-13-503-2009. URL <https://www.hydrol-earth-syst-sci.net/13/503/2009/>.
- Champeaux, Jean-Louis, Dupuy, Pascale, Laurantin, Olivier, Soulan, Isabelle, Tabary, Pierre, and Soubeyroux, Jean-Michel. Les mesures de précipitations et l’estimation des lames d’eau à météo-france : état de l’art et perspectives. *La Houille Blanche*, (5) :28–34, 2009. doi : 10.1051/lhb/2009052. URL <https://doi.org/10.1051/lhb/2009052>.
- M. P. Clark, D. E. Rupp, R. A. Woods, X. Zheng, R. P. Ibbitt, A. G. Slater, J. Schmidt, and M. J. Uddstrom. Hydrological data assimilation with the ensemble kalman filter : Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, 31(10) :1309 – 1324, 2008. ISSN 0309-1708. doi : <https://doi.org/10.1016/j.advwatres.2008.06.005>. URL <http://www.sciencedirect.com/science/article/pii/S0309170808001012>.
- H. Dakhlaoui, D. Ruelland, Y. Trambly, and Z. Bargaoui. Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern tunisia. *Journal of Hydrology*, 550 :201 – 217, 2017. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2017.04.032>. URL <http://www.sciencedirect.com/science/article/pii/S0022169417302512>.
- R. Dankers and L. Feyen. Climate change impact on flood hazard in europe : An assessment based on high-resolution climate simulations. *Journal of Geophysical Research : Atmospheres, Volume 113, Issue D19, CiteID D19105*, 113 :D19105, oct 2008. doi : 10.1029/2007JD009719.

- A. De Lavenne, V. Andréassian, G. Thirel, M. H. Ramos, and C. Perrin. A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research*, 2019.
- C. de Saint-Aubin, L. Garandeau, B. Janet, and P. Javelle. A new french flash flood warning service. In *E3S Web of Conferences*, volume 7, page 18024, 2016.
- J. Dehotin and P. Breil. Projet irip : Rapport technique. *Cartographie de l'aléa inondation par ruissellement (Rapport Irstea)*. Irstea Unité de recherche Hydrologie-Hydraulique, 2011.
- D. Dorchies, J.-P. Baume, and P.-O. Malaterre. Sic-2, un logiciel pour la gestion des canaux, rivières et fleuves. *Sciences Eaux & Territoires*, (2) :48–49, 2013.
- Edijatno. *Mise au point d'un modèle élémentaire pluie-débit au pas de temps journalier*. PhD thesis, Université Louis Pasteur, ENGEES, Cemagref Antony, France, 1991.
- R. M. Errico. What is an adjoint model? *Bulletin of the American Meteorological Society*, 78(11) :2577–2591, 1997.
- G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research : Oceans*, 99(C5) :10143–10162, 1994. doi : 10.1029/94JC00572. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC00572>.
- A. Ficchi, C. Perrin, and V. Andréassian. Impact of temporal resolution of inputs on hydrological model performance : An analysis based on 2400 flood events. *Journal of Hydrology*, 538 :454–470, 2016. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2016.04.016>. URL <http://www.sciencedirect.com/science/article/pii/S0022169416301974>.
- M. A. Fischler and R. C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24 (6) :381–395, June 1981. ISSN 0001-0782. doi : 10.1145/358669.358692. URL <https://doi.org/10.1145/358669.358692>.
- L. Frédéric, V. Freddy, and D. Nancy. *Aléas naturels et gestion des risques*. 01 2010. ISBN 978-2-13-057432-3.
- E. Gaume, V. Bain, P. Bernardara, O. Newinger, M. Barbuc, A. Bateman, L. Blaškovičová, G. Blöschl, M. Borga, A. Dumitrescu, et al. A compilation of data on european flash floods. *Journal of Hydrology*, 367(1-2) :70–78, 2009.
- I. Gejadze and P.-O. Malaterre. Discharge estimation under uncertainty using variational methods with application to the full saint-venant hydraulic network model. *International Journal for Numerical Methods in Fluids*, 83(5) :405–430, 2017.

- I. Gejadze and V. Shutyaev. A method for solving variational data assimilation problems involving linear inequality constraints. *SIAM J. Sci. Comp.*, in review, 2020.
- I. Gejadze, F.-X. Le Dimet, and V. Shutyaev. On optimal solution error covariances in variational data assimilation problems. *Journal of Computational Physics*, 229(6) :2159 – 2178, 2010. ISSN 0021-9991. doi : <https://doi.org/10.1016/j.jcp.2009.11.028>. URL <http://www.sciencedirect.com/science/article/pii/S0021999109006573>.
- J. C. Gilbert. SQPpro - A solver of nonlinear optimization problems, using an SQP approach. Technical Report RT-0378, INRIA, 2009. URL <https://hal.inria.fr/inria-00442314>.
- J. J. Gourley, Z. L. Flamig, H. Vergara, P.-E. Kirstetter, I. Clark, Robert A., E. Argyle, A. Arthur, S. Martinaitis, G. Terti, J. M. Erlingis, Y. Hong, and K. W. Howard. The FLASH Project : Improving the Tools for Flash Flood Monitoring and Prediction across the United States. *Bulletin of the American Meteorological Society*, 98(2) :361–372, 03 2017. ISSN 0003-0007. doi : 10.1175/BAMS-D-15-00247.1. URL <https://doi.org/10.1175/BAMS-D-15-00247.1>.
- P. C. Hansen and D. P. O’Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6) :1487–1503, 1993.
- L. Hascoet and V. Pascual. The tapenade automatic differentiation tool : principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, 39 (3) :20, 2013.
- P. Javelle, C. Fouchier, P. Arnaud, and J. Lavabre. Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *Journal of Hydrology*, 394(1-2) :267–274, nov 2010. doi : 10.1016/j.jhydrol.2010.03.032.
- P. Javelle, J. Demargne, D. Defrance, J. Pansu, and P. Arnaud. Evaluating flash-flood warnings at ungauged locations using post-event surveys : a case study with the aiga warning system. *Hydrological sciences journal*, 59(7) :1390–1402, 2014.
- P. Javelle, D. Organde, J. Demargne, C. Saint-Martin, C. de Saint-Aubin, L. Garandeau, and B. Janet. Setting up a french national flash flood warning system for ungauged catchments based on the aiga method. In *3rd European Conference on Flood Risk Management FLOODrisk 2016*, volume 7, page 11, 2016.
- M. Jay-Allemand, P. Javelle, I. Gejadze, P. Arnaud, P.-O. Malaterre, J.-A. Fine, and D. Organde. On the potential of variational calibration for a fully distributed hydrological model : application on a mediterranean catchment. *Hydrology and Earth System Sciences Discussions*, 2019 :1–24, 2019. doi : 10.5194/hess-2019-331. URL <https://hess.copernicus.org/preprints/hess-2019-331/>.

- B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. De Gruyter, Berlin, Boston, 20 May. 2008. ISBN 978-3-11-020827-6. doi : <https://doi.org/10.1515/9783110208276>. URL <https://www.degruyter.com/view/title/32605>.
- O. Laganier, P. A. Ayrat, D. Salze, and S. Sauvagnargues. A coupling of hydrologic and hydraulic models appropriate for the fast floods of the gardon river basin (france). *Natural Hazards and Earth System Sciences*, 14(11) :2899–2920, 2014.
- J. Le Coz, B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves : A bayesian approach. *Journal of Hydrology*, 509 :573–587, 2014.
- F. Ledimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations - theoretical aspects. *Tellus Series A-Dynamic Meteorology And Oceanography*, 38 :97–110, 1986. doi : 10.3402/tellusa.v38i2.11706.
- H. Lee, D.-J. Seo, Y. Liu, V. Koren, P. McKee, and R. Corby. Variational assimilation of streamflow into operational distributed hydrologic models : effect of spatiotemporal scale of adjustment. *Hydrology and Earth System Sciences, Volume 16, Issue 7, 2012, pp.2233-2251*, 16 :2233–2251, jul 2012. doi : 10.5194/hess-16-2233-2012.
- X. Litrico, J.-B. Pomet, and V. Guinot. Simplified nonlinear modeling of river flow routing. *Advances in Water Resources*, 33(9) :1015–1023, 2010.
- Y. Liu, A. H. Weerts, M. Clark, H.-J. Hendricks Franssen, S. Kumar, H. Moradkhani, D.-J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec, and P. Restrepo. Advancing data assimilation in operational hydrologic forecasting : progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10) :3863–3887, 2012. doi : 10.5194/hess-16-3863-2012. URL <https://www.hydrol-earth-syst-sci.net/16/3863/2012/>.
- M. d. C. Llasat Botija, M. Llasat-Botija, M. Prat Sabartés, F. Porcu, C. Price, A. Mugnai, K. Lagouvardos, V. Kotroni, D. Katsanos, S. Michaelides, et al. High-impact floods and flash floods in mediterranean countries : the flash preliminary database. *Advances in Geosciences, 2010, vol. 23, p. 47-55*, 2010.
- F. Lobligeois. *Mieux connaitre la distribution spatiale des pluies améliore-t-il la modélisation des crues ? : Diagnostic sur 181 bassins versants français*. PhD thesis, AgroParisTech, 2014.
- F. Lobligeois, V. Andréassian, C. Perrin, P. Tabary, and C. Loumagne. When does higher spatial resolution rainfall information improve streamflow simulation ? an evaluation

- using 3620 flood events. *Hydrology and Earth System Sciences, Volume 18, Issue 2, 2014*, pp.575-594, 18 :575–594, feb 2014. doi : 10.5194/hess-18-575-2014.
- A. Mahe. Détermination de descripteurs environnementaux pour la régionalisation d'un paramètre d'un modèle hydrologique. Master's thesis, ENGEES, 2016.
- L. T. Michel C., Maihol J.C. Hydrologie appliquée. Technical report, Cemagref, 1989.
- H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state-parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, 28(2) :135 – 147, 2005. ISSN 0309-1708. doi : <https://doi.org/10.1016/j.advwatres.2004.09.002>. URL <http://www.sciencedirect.com/science/article/pii/S0309170804001605>.
- S. Mouelhi, C. Michel, C. Perrin, and V. Andréassian. Linking stream flow to rainfall at the annual time step : the manabe bucket model revisited. *Journal of hydrology*, 328 (1-2) :283–296, 2006.
- R. Moussa. Analytical hayami solution for the diffusive wave flood routing problem with lateral inflow. *Hydrological Processes*, 10 :1209–1227, 1996. doi : 10.1002/(SICI)1099-1085(199609)10:9<1209::AID-HYP380>3.0.CO;2-2.
- S. Munier, X. Litrico, G. Belaud, and C. Perrin. Assimilation of discharge data into semidistributed catchment models for short-term flow forecasting : Case study of the seine river basin. *Journal of Hydrologic Engineering*, 20(5) :05014021, 2014.
- J. Nash and J. Sutcliffe. River flow forecasting through conceptual models. part i. a conceptual models discussion of principles. *Journal of Hydrology*, 10 :282–290, 1970.
- J. E. Nash. The form of the instantaneous unit hydrograph. *International Association of Scientific Hydrology, Publ*, 3 :114–121, 1957.
- V. T. Nguyen, D. Georges, G. Besançon, and I. Zin. Parameter estimation of a real hydrological system using an adjoint method**this work has been partially supported by the labex persyval-lab (anr-11-labx-0025-01) and the mepiera project, grenoble institute of technology. *IFAC-PapersOnLine*, 49(13) :300 – 305, 2016. ISSN 2405-8963. doi : <https://doi.org/10.1016/j.ifacol.2016.07.978>. URL <http://www.sciencedirect.com/science/article/pii/S2405896316312666>. 12th IFAC Workshop on Adaptation and Learning in Control and Signal Processing ALCOSP 2016.
- D. Organde. Cartographie des débits de crue en métropole - régionalisation du modèle pluie-débit - prise en compte des pixels urbains et en eau dans le calage et la régionalisation. Technical report, HYDRIS hydrologie, 2016.

- D. Organde, P. Javelle, J. Demargne, M. Jay-Allemand, J.-A. Fine, and P. Arnaud. Méthode aiga : vers une modélisation hydrologique au pas de temps infra-horaire continue et totalement distribuée. In *De la prévision des crues à la gestion de crise*. Société Hyddraulique de France Société Hyddraulique de France, Nov. 2018. doi : 0000.
- L. Oudin, F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne. Which potential evapotranspiration input for a lumped rainfall–runoff model? : Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, 303(1-4) :290–306, 2005.
- O. e. a. Payraastre. Améliorer la prévision immédiate des crues soudaines et de leurs impacts : le projet de recherche anr pics, colloque shf “de la prévision des crues à la gestion de crise”. 10 2018.
- H. L. Penman and B. A. Keen. Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 193(1032) :120–145, 1948. doi : 10.1098/rspa.1948.0037. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1948.0037>.
- C. Perrin, C. Michel, and V. Andréassian. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology, Volume 279, Issue 1, p. 275-289.*, 279 : 275–289, aug 2003. doi : 10.1016/S0022-1694(03)00225-7.
- P. Pokhrel and H. V. Gupta. On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, 46(1), 2010.
- C. Poncelet. *Du bassin au paramètre : jusqu’où peut-on régionaliser un modèle hydrologique conceptuel?* Theses, Université Pierre et Marie Curie - Paris VI, Nov. 2016. URL <https://tel.archives-ouvertes.fr/tel-01529196>.
- A. Quesney, C. Francois, C. Otle, S. Hegarat, C. Loumagne, M. Normand, et al. Sequential assimilation of sar/ers data in a lumped rainfall-runoff model with an extended kalman filter. *IAHS-AISH PUBL.*, (267) :495–497, 2000.
- F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteorol. Soc.*, 118 :649–672, feb 1992.
- R. H. Reichle, D. B. McLaughlin, and D. Entekhabi. Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Monthly Weather Review*, 130(1) :103–114, 01 2002. ISSN 0027-0644. doi : 10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2. URL [https://doi.org/10.1175/1520-0493\(2002\)130<0103:HDAWTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2).
- P. Riboust, G. Thirel, N. L. Moine, and P. Ribstein. Revisiting a simple degree-day model for integrating satellite data : Implementation of swe-sca hystereses. *Journal of Hydrology and Hydromechanics*, 67(1) :70–81, 2019.

- P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.
- H. Roux, D. Labat, P.-A. Garambois, M.-M. Maubourguet, J. Chorda, and D. Dartus. A physically-based parsimonious hydrological model for flash floods in mediterranean catchments. *Natural Hazards and Earth System Science, Volume 11, Issue 9, 2011, pp.2567-2582*, 11 :2567–2582, sep 2011. doi : 10.5194/nhess-11-2567-2011.
- C. Saint Martin, C. Fouchier, P. Javelle, J. Douvinet, and F. Vinet. Assessing the exposure to floods to estimate the risk of flood-related damage in french mediterranean basins. 2016.
- L. Samaniego, R. Kumar, and S. Attinger. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), 2010. doi : 10.1029/2008WR007327. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008WR007327>.
- M. Sangati and M. Borga. Influence of rainfall spatial resolution on flash flood modelling. *Natural Hazards & Earth System Sciences*, 9(2), 2009.
- L. Santos, G. Thirel, and C. Perrin. Continuous state-space representation of a bucket-type rainfall-runoff model : a case study with the gr4 model using state-space gr4 (version 1.0). *Geoscientific Model Development*, 11(4) :1591–1605, 2018.
- D.-J. Seo, L. Cajina, R. Corby, and T. Howieson. Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, 367 (3) :255 – 275, 2009. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2009.01.019>. URL <http://www.sciencedirect.com/science/article/pii/S002216940900390>.
- L. Sun, I. Nistor, and O. Seidou. Streamflow data assimilation in swat model using extended kalman filter. *Journal of Hydrology*, 531 :671–684, 2015.
- L. Sun, O. Seidou, I. Nistor, and K. Liu. Review of the kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, 61(13) :2348–2366, 2016.
- J. Thielen, J. Bartholmes, M.-H. Ramos, and A. De Roo. The European Flood Alert System – Part 1 : Concept and development. *Hydrology and Earth System Sciences Discussions*, 5(1) :257–287, Feb. 2008. URL <https://hal.archives-ouvertes.fr/hal-00330814>.
- G. Thirel, E. Martin, J. F. Mahfouf, S. Massart, S. Ricci, and F. Habets. 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, 2010.

- A. N. Tikhonov. On the regularization of ill-posed problems. In *Doklady Akademii Nauk*, volume 153, pages 49–52. Russian Academy of Sciences, 1963.
- Y. Tramblay, C. Bouvier, A. Crespy, and A. Marchandise. Improvement of flash flood modelling using spatial patterns of rainfall : A case study in southern france. In *Sixth World FRIEND Conference*, pages 172–178, 2010.
- F. Vinet. Geographical analysis of damage due to flash floods in southern france : The cases of 12–13 november 1999 and 8–9 september 2002. *Applied Geography*, 28(4) : 323–336, 2008. doi : 10.1016/j.apgeog.2008.02.007.
- Vinet, Freddy, Boissier, Laurent, and Saint-Martin, Clotilde. Flashflood-related mortality in southern france : first results from a new database. *E3S Web Conf.*, 7 :06001, 2016. doi : 10.1051/e3sconf/20160706001. URL <https://doi.org/10.1051/e3sconf/20160706001>.
- J. Wang, Y. Hong, L. Li, J. J. Gourley, S. I. Khan, K. K. Yilmaz, R. F. Adler, F. S. Policelli, S. Habib, D. Irwn, A. S. Limaye, T. Korme, and L. Okello. The coupled routing and excess storage (crest) distributed hydrological model. *Hydrological Sciences Journal*, 56(1) :84–98, 2011. doi : 10.1080/02626667.2010.543087. URL <https://doi.org/10.1080/02626667.2010.543087>.
- S. Westrelin, P. Mériaux, P. Tabary, and Y. Aubert. Hydrometeorological risks in mediterranean mountainous areas-rhythmme project : Risk management based on a radar network. In *ERAD 2012 7th European Conference on Radar in Meteorology and Hydrology*, page 6, 2012.
- X. Xie and D. Zhang. Data assimilation for distributed hydrological catchment modeling via ensemble kalman filter. *Advances in Water Resources*, 33(6) :678 – 690, 2010. ISSN 0309-1708. doi : <https://doi.org/10.1016/j.advwatres.2010.03.012>. URL <http://www.sciencedirect.com/science/article/pii/S0309170810000618>.
- C.-Y. Xu, L. Xiong, and V. P. Singh. *Black-Box Hydrological Models*, pages 1–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017. ISBN 978-3-642-40457-3. doi : 10.1007/978-3-642-40457-3_21-1. URL https://doi.org/10.1007/978-3-642-40457-3_21-1.
- C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. L-bfgs-b : a limited memory fortran code for solving bound constrained optimization problems. *EECS Department, Northwestern University, Evanston, IL, Technical Report No. NAM-11*, 1994.