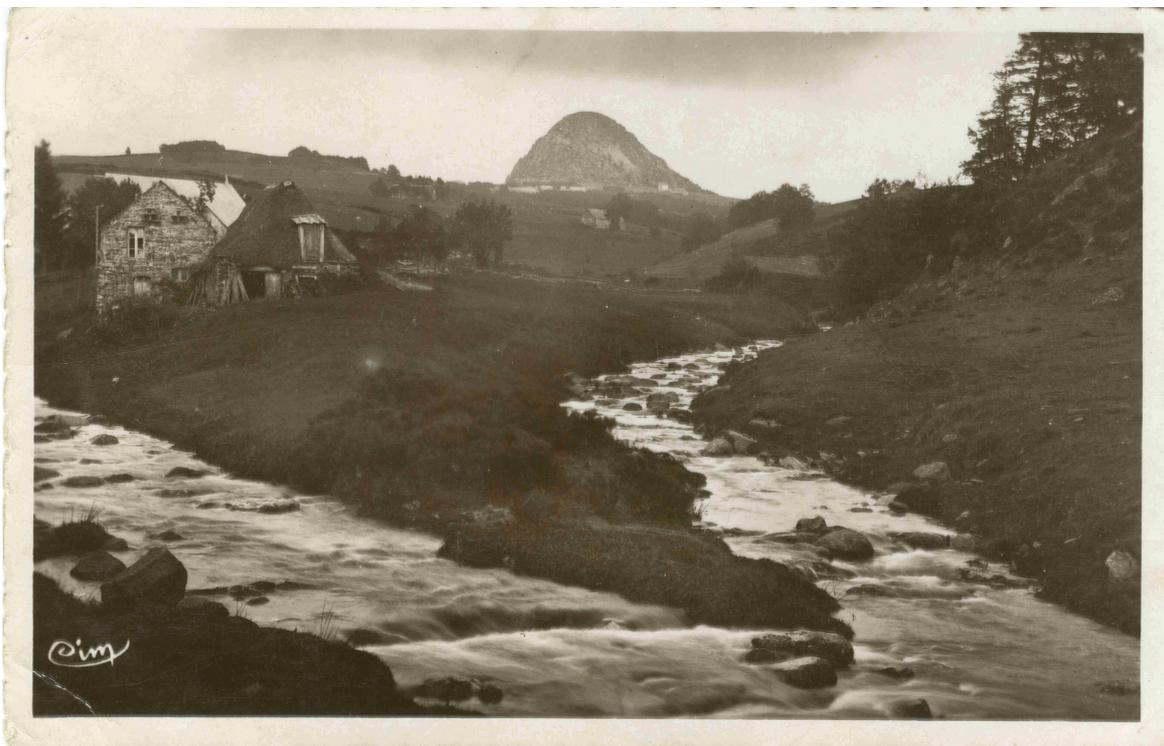


Avant-propos



Le Mont gerbier des joncs (Ardèche) – Altitude 1554 m.
Sources véritables et uniques de la Loire. Début du XX^e siècle.

« Je n'ai pas de souvenirs d'enfance. »

George Perec, *W ou le souvenir d'enfance*, 1975.

*À la mémoire de ma mère
A mon père*

Je me pose souvent la question: « Qu'est ce que tu fais là ? » et me demande l'origine de ma passion pour l'hydrologie. Je crois que mon attachement pour l'eau et cette science vient du fait qu'elles me permettent de retrouver quelques souvenirs d'enfance, liés à l'eau, aux sources et aux rivières.

Je me souviens des visites des sources de la Loire au Mont Gerbier des Joncs ; des parties de pêche infructueuses à la Rèche, dans les bois du grand-père ; des récits incroyables de ma mère qui allait chercher de l'eau aux sources minérales et ferrugineuses de César et Moïse, à Désaignes; des nombreuses après-midi passées à la source et à la mare de St Agrève, au pied du Chiniac; des trois sources du Pouzat, où « la bonne vierge vint se désaltérer »; des eaux minérales pétillantes de l'Ardèche, dont je ne peux me passer : la Vals, l'Arcens et la Reine des Basaltes; de l'observation des variations intrigantes du débit de la source du Fraysse; des crues hivernales du Rhône, des après-midi à construire des barrages en galets dans la Drôme, des eaux froides du Doux à Désaignes, de l'Eyrieux; enfin, de mes nombreuses baignades avec mes cousines dans les profonds canyons granitiques et basaltiques de la Glueyre et du Glot ...

Maintenant, je crois plutôt que mon attachement pour l'hydrologie (au sens large) vient de la prise de conscience du fait que les ressources en eau sont indispensables aux développements des hommes et de leurs sociétés, que leur gestion globale impose la rencontre et la discussion avec d'autres hommes. Je peux dire ainsi que ce qui me passionne dans l'hydrologie, c'est l'hydrologie sociale.

Enfin, comme disait Michel Leiris (L'Afrique fantôme, 1934): « Finalement ce n'est plus du tout le but qui me passionne mais la recherche, seulement en tant que recherche. L'océan de poésie dans lequel nous sommes plongés». Je dois dire que mon engagement vers la recherche tient de certains de mes professeurs, Michel Lepiller, Alain Mangin, Ghislain de Marsily. Mon frère, Renaud, m'a également entraîné dans cette aventure, par sa culture scientifique, sa curiosité inégalable et grâce aux derniers étés passés à côté de sa bétonnière ou sur son toit. Tant bien que mal, il fait parti de ces modèles, dont j'ai essayé de faire quelque chose pour trouver ma voie/voix.

A force de me demander ce que je fais là, j'en arrive à épuiser l'ensemble des raisons plus ou moins obscures pour lesquelles j'ai choisi l'hydrologie et pourquoi je me suis laissé entraîner vers elle... Il me reste alors la passion, le plaisir de la recherche, de l'imagination et de la création !

Bonne lecture.

Remerciements

Je tiens à exprimer ma profonde reconnaissance à Claude Michel, Charles Perrin et Vazken Andréassian, hydrologues au Cemagref d'Antony. Par leurs encouragements et remarques constructives, ils ont contribué significativement aux travaux de cette thèse et à ma motivation quotidienne à 200% pour la recherche.

Claude Michel a encadré ce travail de recherche. Son expérience, son déterminisme et sa rigueur ont été une source de motivation. Il m'a beaucoup aidé à me diriger vers la voie parfois frustrante de l'empirisme. Je tiens à le remercier également pour m'avoir appris à manier le rasoir d'Occam.

Charles Perrin et Vazken Andréassian ont eu la lourde tâche de me suivre au quotidien, de m'écouter, de relancer la machine de temps en temps et de m'empêcher de me disperser... Avec patience et attention, ils ont fait de cette thèse une belle aventure scientifique et humaine.

Mes sincères remerciements à l'ensemble des membres de ce jury de thèse de m'honorer de leur présence, d'avoir accepté de lire et juger mes travaux.

J'adresse ainsi toute ma gratitude à Claude Thirriot, Professeur émérite de l'École Supérieure d'Électrotechnique, d'Électronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT) pour avoir accepté d'être rapporteur de mes travaux et d'avoir animé cette soutenance par ses questions teintées de cet accent toulousain.

J'adresse également ma profonde reconnaissance à François Anctil, Professeur à l'Université Laval (Québec), qui a également accepté d'être rapporteur. Par ses cours, sa sympathie et son ouverture d'esprit, je lui dois mon orientation vers la modélisation hydrologique et ce sujet de thèse ! Du Québec, il a été une source de motivation par ses contributions originales à l'hydrologie.

Je tiens à remercier très chaleureusement Pierre Hubert, Maître de Recherche à l'École des Mines de Paris, qui a présidé à sa manière ce jury de thèse. Ses engagements en faveur du développement des relations au sein de la communauté hydrologique ont remarquablement contribué à mes recherches et à ma volonté de développer le concept d'hydrologie sociale.

Je remercie également Emmanuel Paquet, ingénieur à EDF-DTG Grenoble, pour avoir accepté de lire et d'examiner cette thèse, malgré son poignet défectueux. Notre collaboration m'a permis d'approfondir certaines de mes recherches, notamment du côté obscur de MORDOR et de son algorithme génétique...

Je tiens à remercier également Christophe Bouvier, Chargé de Recherche à Hydrosociences Montpellier, d'examiner cette thèse. Je suis convaincu que son expérience des Cévennes apportera des discussions intéressantes sur le pas de temps horaire.

Enfin, je souhaite exprimer toute ma reconnaissance à Georges-Marie Saulnier qui a accepté d'examiner cette thèse. Quelques unes de nos discussions enflammées en Macédoine ou à Paris ont été une source d'inspirations pour ce travail de modélisation hydrologique. J'espère que ce n'est que le début !

Je tiens à remercier certaines personnes pour les conseils qu'elles m'ont apportés au cours de différentes rencontres : Hocine Bendjoudi, Patrick Lavabre, Bruno Rambaldelli, Eric Parent, Edijatno, Joël Gaillard et Rémy Garçon, ainsi que les membres du MOPEX.

Cette thèse s'appuie sur des données qui ont été recueillies et mises à disposition par des organismes divers. Je tiens à exprimer ma reconnaissance aux personnes qui m'ont donné accès à ces données, sans lesquels ce travail de thèse n'aurait pu être fait :

- Christian Scherrer, du Ministère de l'Écologie et du Développement Durable, pour les données de la banque HYDRO ;
- Bruno Rambaldelli, de Météo France, pour les données de la banque PLUVIO ;
- Les organismes américains de l'Agricultural Research Service et de l'United State Department of Agriculture, pour les données des bassins versants américains ;
- Muguresu Sivapalan et Christian Zammit (University of Western Australia), pour les données des bassins versants australiens ;
- Mira Kobold (Environmental Agency of the Republic of Slovenia), pour les données des bassins versants slovènes ;
- Jérôme Latron et Fransesc Gallart (Institut de Ciències de la Terra Jaume Almera, CSIC), pour les données des bassins versants catalans ;
- Marine Riffard et Catherine Foucher, pour les données des bassins versants expérimentaux du Cemagref (Orgeval et Réal Collobrier) ;
- Patrick Ansart, technicien du bassin versant de l'Orgeval, et toutes les personnes qui contribuent à la collecte des données sans lesquelles la recherche ne pourrait pas être menée ;

- Jérôme Molénat (INRA Rennes) pour les données du bassin versant expérimental de Naizin-Kervidy ;
- Guy Almensa pour les données des bassins versants de l'Eygues et de l'Ouvèze ;
- Joël Gaillard, Emmanuel Paquet et Rémy Garçon (EDF-DTG Grenoble) pour les données des bassins versants suivis par EDF-DTG ;

Je souhaite remercier Jean-Louis Rosique pour s'être occupé avec enthousiasme de la partie cartographique de cette thèse et m'avoir fournis de beaux contours de mes bassins versants. Je remercie également Gilles Bonnet, qui m'a donné accès au Cluster de Clermont-Ferrand, sans lequel les 36 000 heures de calculs n'auraient pu être possibles.

Un grand merci à l'ensemble des permanents ou stagiaires de HBAN du Cemagref : Marine, Yves, Mamoutou, Claudia, Cécile, Jean-Luc, Marie, Laetitia, Rémy, Sébastien, Georges, Thaïs, Julien, les équipes de foot, basket et frisbee... Ludo et Nanée, pour les discussions de fin de thèse sur la distribution de la pluie et les multi-modèles ; Sylvie Tonachella, Sophie Morin et Valérie Dansin pour avoir animé avec bonne humeur la partie administrative de cette thèse ; Je souhaite adresser un clin d'œil aux jeunes filles du 2^{ème}, qui ont égayé d'une manière ou d'une autre ces recherches : Claire et Béné. Merci également à Jean-Luc Pujol, Cécile Loumagne et Gildas Le Bozec, successivement chefs de l'Unité de Recherche HBAN où ces recherches se sont déroulées. Merci enfin à Névine Kocher, Lucia Belloni et l'ensemble du SREI pour avoir financé mes participations à des congrès en Macédoine et à Foz do Iguaçu. Ces congrès ont particulièrement stimulé mes travaux de recherche en me donnant l'occasion de rencontrer d'éminents hydrologues (Georges-Marie Saulnier, Hoshin Gupta, Victor Koren,...).

Je souhaite remercier chaleureusement Bruno Sellier et tous les membres de l'imprimerie du Cemagref pour l'aide et le soin qu'ils ont apportés lors de l'impression de cette thèse.

Enfin, un dernier merci à l'équipe courageuse (et sans pitié !) de relecteurs : Ludo, Laetitia et Charles.

Ces recherches ont reçu le soutien de la Direction de l'Eau du Ministère de l'Ecologie et du Développement Durable sur la période 2003-2005.

Résumé

La simulation et la prévision des débits à l'échelle du bassin versant est un exercice délicat, d'autant plus difficile que le temps de réponse d'un bassin versant est court. Il existe aujourd'hui un très grand nombre de modèles hydrologiques globaux qui, selon leurs auteurs, sont efficaces. Dans la pratique, face à cette diversité, l'étude comparée des meilleurs modèles existants sur un vaste échantillon de bassins versants s'est révélée indispensable pour discriminer leurs performances respectives. Curieusement, ce travail exhaustif de comparaison n'avait jamais été effectué auparavant au pas de temps horaire alors que rien ne garantissait la validité à ce pas de temps, des conclusions obtenues à un pas de temps supérieur, le pas de temps journalier.

Le premier travail fut de montrer l'intérêt d'utiliser un large échantillon de bassins versants pour classer de façon relativement fiable les performances de modèles pluie-débit. Ce faisant, il a été proposé une autre formulation du critère de *Nash et Sutcliffe* [1970] (NS), appelée C_{2M} , mieux adaptée aux grands échantillons de bassins versants, lorsque certains bassins apparemment difficiles à modéliser n'en sont pas expurgés. En effet, ces bassins difficiles engendrent des valeurs exceptionnellement basses pour le critère de NS interdisant toute signification au calcul d'une moyenne. De plus, pour que la comparaison ne soit pas biaisée par la méthode de calage nécessairement simple qui a été employée, sa capacité à optimiser efficacement les paramètres a été confrontée et démontrée, par rapport aux meilleures méthodes développées dans la dernière décennie.

C'est dans ce cadre rigoureux qu'a été effectuée l'étude des performances de 20 structures de modèles au pas de temps horaire, selon une procédure de calage-contrôle. Pour cela, un échantillon de 313 bassins versants très variés, situés dans des pays aussi contrastés que la France, les Etats-Unis, l'Australie, la Slovénie et l'Espagne. Deux structures sont apparues comme les plus efficaces : GR4H et MORDOR6. Ces structures ont été analysées, modifiées, simplifiées, complexifiées, pour mieux comprendre, parmi les nombreux ingrédients possibles de modèles pluie-débit, lesquels semblaient rejetés par les données et lesquels semblaient, au contraire, justifiés par elles. Cette démarche empirique et comparative a débouché sur deux modèles respectivement à 4 et 6 paramètres qui semblent avoir les meilleures structures pour reproduire la transformation de la pluie en débit à l'échelle d'un bassin et au pas de temps horaire. Un travail reste à faire pour extraire de cette dualité un modèle qui pourrait les surclasser tous les deux.

On a mis en évidence tout l'intérêt qu'il y a à relier les meilleurs modèles issus de l'étude de la transformation pluie-débit à différents pas de temps. Cette analyse conduite avec la structure à 4 paramètres permis d'élaborer une variante commune, également efficace aux pas de temps horaire et journalier. Cette structure s'est révélée tellement bien adaptée aux deux pas de temps qu'une prédétermination robuste des paramètres du modèle horaire a été possible à partir de son seul calage au pas de temps journalier.

Ces travaux ne mettent certainement pas un point final à la quête du meilleur modèle au pas de temps horaire, mais ont permis de déblayer considérablement le terrain et de permettre de futures avancées dans les meilleures conditions possible.

Abstract

The simulation and forecast of flows at the catchment scale is an all the more difficult modelling exercise as the response time of the catchment is short. A large number of empirical or conceptual models exist today and prove to be efficient in an operational context. Given this diversity of models, the intercomparison of various models on large samples of basins appears to be a good mean to assess their performances.

However an exhaustive comparative assessment has never been carried out at the hourly time step and nothing guarantees at short time steps the validity of conclusions obtained at a daily time step. So to determine the best performing models at the hourly time step, we placed our research in an empirical framework of model development and intercomparison.

We first showed the usefulness of using large samples of basins to compare the performances of rainfall-runoff models with good reliability. Then we proposed a new formulation of the *Nash and Sutcliffe* [1970] criterion, called C_{2M} , better suited for such large samples. We also confirmed that the local 'step-by-step' optimization method we used was equally effective to identify model parameters with different levels of complexity.

We assessed the performances of twenty structures of rainfall-runoff models at the hourly time-step, using a split-sample test procedure on a sample of 313 basins situated in France, the United-States, Australia, Slovenia and Spain. Two structures appeared to be the most successful. These structures (four and six free parameters respectively) are the results of empirical developments and were built by gradually complexifying or simplifying existing structures. The empirical development and intercomparison of models showed that four to six parameters are necessary to represent the rainfall-runoff transformation at the hourly time-step.

The value of testing models at several time steps was also demonstrated. We proposed a new model structure with four free parameters that performs equally well at the daily and hourly time-steps. We developed a robust method to estimate model parameters at the hourly time-step, using the calibrated parameters at the daily time step. Last, we proposed strategies to account for the spatial heterogeneity of rainfall that provide significant improvements in model efficiency.

Sommaire

<i>Avant-propos</i>	3
<i>Remerciements</i>	5
<i>Résumé</i>	9
<i>Abstract</i>	10
<i>Introduction générale</i>	15
PREMIERE PARTIE : METHODOLOGIE DE COMPARAISON ET DE DEVELOPPEMENT DE MODELES PLUIE-DEBIT	21
CHAPITRE 1 Modélisation hydrologique à l'échelle du bassin versant	25
CHAPITRE 2 Présentation de l'échantillon de bassins versants	51
CHAPITRE 3 Stratégie de comparaison des performances de modèles pluie-débit	85
CHAPITRE 4 Optimisation automatique des modèles pluie-débit : comparaison exhaustive des performances de méthodes locale et globales.	115
DEUXIEME PARTIE : COMPARAISON ET DEVELOPPEMENT DE MODELES PLUIE-DEBIT AU PAS DE TEMPS HORAIRE	157
CHAPITRE 5 Comparaison des performances de modèles pluie-débit au pas de temps horaire	161
CHAPITRE 6 Approche empirique de mise au point d'un modèle pluie-débit au pas de temps horaire	203
TROISIEME PARTIE : QUELS AUTRES DEVELOPPEMENTS PEUT NOUS APPORTER LE PAS DE TEMPS HORAIRE ?	245
CHAPITRE 7 Amélioration d'une méthode d'optimisation locale : prise en compte de l'information apportée par la distribution a priori des paramètres	249
CHAPITRE 8 Apports de la comparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier	279
CHAPITRE 9 Amélioration des performances du modèle GR4H par une meilleure valorisation de l'information pluviométrique	313
Conclusion générale	333
Références bibliographiques	337
Epilogue	353
ANNEXES	355

Introduction générale

« Dans la vie, l'essentiel est de porter sur tout des jugements a priori. Il apparaît, en effet, que les masses ont tort, et les individus toujours raison. Il faut se garder d'en déduire des règles de conduite : elles ne doivent pas avoir besoin d'être formulées pour qu'on les suive. »

Boris Vian, *L'écume des jours*, 1963.

Introduction générale

Contexte

Nous avons tous en mémoire les dernières crues dévastatrices qui ont touché la France : crue de la Borne (Grand-Bornand, 1987), crues de l'Ouvèze (Vaison la Romaine, 1992, 1993), crues de l'Aude (1999), crues de la Laïta à Quimperlé (2000), crues des Gardons (2002), crues de Nîmes (1988, 2005), etc. Ces événements ont tous en commun d'avoir concerné des bassins versants aux réactions particulièrement rapides et violentes : ils ont répondu en quelques heures à des précipitations intenses. Ces phénomènes ne sont d'ailleurs pas particuliers à la France, puisque de nombreux pays, en Europe et ailleurs, subissent les conséquences de ces événements, comme nous l'avons vu cet été 2005 sur l'arc alpin.

Les crues des 20 dernières années ont fait plus de 200 victimes et coûté plusieurs milliards d'euros en France. Elles illustrent la vulnérabilité humaine et économique face aux aléas naturels. Elles ont également mis en évidence une augmentation des exigences des citoyens en matière de prévention et de prévision des inondations. Cette exigence a notamment incité la mise en place des Plans de Prévention des Risques d'Inondations (PPRI), la création du Service Central Hydrométéorologique d'Aide à la Prévision des Inondations (SCHAPI) et la réforme des Services de Prévision des Crues (SPC), qui ont aujourd'hui un rôle réel de prévision des événements de crue.

La prédiction (prévision et simulation) opérationnelle et quantitative de la réponse des bassins versants aux événements pluvieux est un exercice délicat. En effet, les bassins versants sont des systèmes naturels complexes et hétérogènes, dont l'état hydrique et la réponse aux pluies sont conditionnés par de nombreux facteurs variables dans le temps. Ainsi, pour appréhender leurs réponses, on a de plus en plus recours à la modélisation hydrologique. Dans ce cadre, les modèles hydrologiques pluie-débit, conceptuels ou empiriques, sont des outils efficaces qui trouvent toute leur légitimité dans une application opérationnelle. Aujourd'hui, de nombreux services opérationnels de prévision des crues utilisent ces outils dans différents pays, pour simuler et prévoir les débits des rivières de nombreux bassins versants, notamment pour les petits bassins versants amont les plus réactifs.

Problématique scientifique

A l'heure actuelle, la simulation et la prévision des débits sur les bassins versants particulièrement réactifs sont des enjeux très importants. Pour faire confiance aux simulations et aux prévisions fournies par les modèles hydrologiques, il est indispensable d'utiliser des modèles performants, robustes, fiables et adaptés à ce type de bassins versants. Cette thèse s'intéresse à la modélisation des bassins versants à réponse rapide et son principal objectif est de répondre à la question :

« Quels modèles hydrologiques sont les plus intéressants pour le pas de temps horaire ? »

Pour comprendre la raison qui nous pousse à poser la question en ces termes, il faut réaliser que les modèles hydrologiques sont des simplifications toujours extrêmes du système bassin versant réel. A ce titre, ils opèrent une triple globalisation du système pour le représenter : globalisation des processus mis en jeu, globalisation de l'échelle d'espace et globalisation de l'échelle de temps. A l'heure actuelle, les effets de la globalisation des processus (recherche des processus dominants) et de la représentation spatiale du bassin versant (notamment, grâce à la croissante disponibilité des données spatialisées) sont largement étudiés, alors que l'effet de la globalisation temporelle est souvent négligé. Bien au contraire, nous pensons que le rôle de la globalisation temporelle est au moins aussi important dans le choix de modélisation, que celui des processus et de l'échelle d'espace.

Les modèles hydrologiques aujourd'hui disponibles sont encore tous loin d'être satisfaisants et leurs faibles performances reflètent la difficulté d'appréhender la complexité du système réel. Les inconnues sur les déterminants réels du fonctionnement des bassins versants sont à l'origine de nombreuses incertitudes qui affectent les simulations du comportement des bassins versants. Ces incertitudes proviennent principalement de la structure des modèles, de la manière d'estimer les paramètres des modèles, de la manière de prendre en compte la variabilité des pluies et des caractéristiques du bassin versant, des erreurs dans les données utilisées... Ces incertitudes semblent encore plus importantes lorsqu'on s'intéresse à la simulation du débit à pas de temps fin, tel que le pas de temps horaire. Ainsi, dans cette thèse, notre objectif a été d'améliorer les performances et de réduire les incertitudes de la modélisation hydrologique au pas de temps horaire. Pour cela nous avons exploré différentes voies de recherche.

Le cœur de notre travail porte sur la question de l'adéquation de la structure du modèle avec le pas de temps de modélisation (Chapitre 1). Ce questionnement, tout à fait légitime, trouve ses fondements dans les réflexions sur l'approche descendante (*Downward approach*) de *Klemeš* [1983] et de *Sivapalan et al.* [2003], dans les

résultats de *Jothityangkoon et al.* [2001], *Atkinson et al.* [2002] et *Eder et al.* [2003] et au Cemagref de *Mouelhi* [2003]. Pour répondre à cette question, nous nous sommes placés dans un cadre de développement empirique (Chapitre 1) et de comparaison de modèles (Chapitre 5) sur un large échantillon de bassins versants (Chapitre 2). Cette démarche a d'ailleurs déjà été adoptée au Cemagref dans le cadre des thèses de *Perrin* [2000] et *Mouelhi* [2003]. L'originalité de cette démarche est de soumettre les modèles hydrologiques à des conditions hydro-climatiques très variées, afin de mettre en évidence leurs capacités à être performants, robustes et généraux. Cet échantillon nous permet en fait de tester de nombreuses hypothèses sur la manière de représenter le bassin versant, dans une approche s'apparentant finalement à du *data-mining* [*Babovic*, 2005].

La question des erreurs et incertitudes induites par la méthode d'optimisation des paramètres des modèles hydrologiques (Chapitre 4 et 7) a également été posée. Enfin, les intérêts de l'utilisation de données horaires pour une modélisation hydrologique au pas de temps journalier (Chapitre 8) et de la prise en compte de la variabilité spatiale des pluies (Chapitre 9) ont été explorés.

Organisation du mémoire

Ce mémoire s'organise en trois parties et neuf chapitres.

La première partie est méthodologique (quatre chapitres)

Le premier chapitre est introductif. Il vise à préciser le cadre de modélisation hydrologique dans lequel on se place. C'est-à-dire, les particularités liées à la fois au choix du pas de temps horaire et aussi au choix de la philosophie de développement et de comparaison modèles hydrologiques suivie au cours de cette thèse.

Le deuxième chapitre présente l'échantillon de données utilisé pour tester les modèles. Cet échantillon regroupe les données au pas de temps horaire de 313 bassins versants principalement situés en France, mais également aux États-unis, en Australie, en Slovénie et en Espagne. Pour le pas de temps horaire, cet échantillon est inédit : aucun modèle n'a jamais été testé de manière aussi exhaustive au pas de temps horaire.

Le troisième chapitre présente la stratégie de comparaison des performances de modèles pluie-débit. Ce chapitre vise à démontrer l'intérêt des grands échantillons de bassins versants pour comparer significativement les performances de différents modèles hydrologiques. Ce chapitre présente également une formulation alternative au critère de *Nash et Sutcliffe* [1970], appelée C_{2M} , mieux adaptée à la comparaison

des performances de différents modèles sur un grand échantillon de bassins versants.

Le quatrième chapitre présente une comparaison des performances de méthodes d'optimisation automatique des paramètres de modèles hydrologiques. Cette comparaison permet de juger les performances d'une méthode locale simple utilisée au cours de cette thèse, par rapport à celles de deux méthodes globales puissantes. Ce chapitre explore les performances des différents vecteurs optimisés par ces trois méthodes, en calage et en contrôle. Ces résultats nous permettent également de nous prononcer sur d'éventuelles faiblesses de la structure des modèles hydrologiques testés ici, en étudiant la structure de leurs surfaces de réponse.

La deuxième partie présente les principaux résultats (deux chapitres)

Le cinquième chapitre donne les résultats des tests comparatifs des 20 structures de modèles pluie-débit sélectionnées. Nous explorons les performances des modèles, leur robustesse, leur fiabilité et leur complémentarité. Nous étudions également le niveau de complexité nécessaire et suffisant pour obtenir de bonnes performances au pas de temps horaire. Nous essayons ensuite de déterminer des facteurs physiographiques discriminant les performances des modèles. Enfin, nous étudions l'influence de la durée de calage sur les performances et la robustesse des modèles.

Le sixième chapitre présente le développement empirique de meilleures structures de modèles au pas de temps horaire. Nous suivons une démarche de complexification progressive de la structure, à partir de la structure d'un modèle à quatre paramètres (GR4J). Cette démarche nous a permis de proposer deux structures plus performantes au pas de temps horaire : GR4H et GR5H. Nous suivons également une démarche de simplification progressive de structures de modèles à partir de deux structures différentes ayant huit et dix paramètres.

***La troisième partie présente des compléments opérationnels originaux :
les bonus (trois chapitres)***

Le septième chapitre détaille une méthode originale de pré-filtrage de l'espace des paramètres d'un modèle, en valorisant l'information contenue dans la distribution *a priori* des valeurs des paramètres d'un modèle. Cette méthode permet de choisir un point de départ de l'algorithme de recherche pas à pas, généralement bien situé, c'est-à-dire à proximité de l'optimum de la surface de réponse. Nous présentons également une analyse de sensibilité du critère de terminaison de cette méthode d'optimisation.

Le huitième chapitre donne les résultats obtenus en testant de manière comparative les modèles aux pas de temps horaire et journalier. Nous présentons les améliorations que l'on peut espérer au pas de temps journalier, en utilisant un modèle au pas de temps horaire. Nous présentons également la comparaison de différentes versions de modèles de la famille GR aux pas de temps horaire et journalier. Cette comparaison nous permet de présenter une structure (GR4X1) également performante aux deux pas de temps. Enfin, nous présentons différentes stratégies de prédétermination du vecteur de paramètres au pas de temps horaire du modèle GR4X1, à partir de vecteurs optimisés avec des données journalières.

Le dernier et neuvième chapitre présente des résultats obtenus en testant quelques stratégies de prise en compte de la variabilité spatiale de la pluie. Sur un sous-échantillon de 145 bassins versants français, nous testons quatre stratégies de complexité croissante, qui permettent d'améliorer les simulations des modèles par une meilleure valorisation de l'information pluviométrique.

PREMIERE PARTIE :

METHODOLOGIE DE COMPARAISON ET DE DEVELOPPEMENT DE MODELES PLUIE-DEBIT

Chapitre 1

« DE LA RIGEUR DE LA SCIENCE

En cet empire, l'Art de la Cartographie fut poussé à une telle Perfection que la Carte d'une seule Province occupait toute une Ville et la Carte de l'Empire toute une Province. Avec le temps, ces Cartes Démesurées cessèrent de donner satisfaction et les Collèges de Cartographes levèrent une Carte de l'Empire, qui avait le Format de l'Empire et qui coïncidait avec lui, point par point. Moins passionnées pour l'Étude de la Cartographie, les Générations Suivantes réfléchirent que cette Carte Dilatée était inutile et, non sans impiété, elles l'abandonnèrent à l'Inclémence du Soleil et des Hivers. Dans les Déserts de l'Ouest, subsistent des Ruines très abîmées de la Carte. Des Animaux et des Mendiants les habitent. Dans tout le Pays, il n'y a plus d'autre trace des Disciplines Géographiques. »

Suarez Miranda, *Viajes de Varones Prudentes*,
Lib. IV, Cap. XIV, Lerida, 1658

Cité par Jorge Luis Borges, *Histoire universelle de l'infamie*, 1935.

CHAPITRE 1 Modélisation hydrologique à l'échelle du bassin versant

1.1 Introduction

Cette thèse porte sur la modélisation hydrologique pluie-débit, au pas de temps horaire. Avant d'entrer dans le vif du sujet, nous allons faire quelques rappels sur l'hydrologie, le système bassin versant et la modélisation hydrologique. Les enjeux et les applications de la modélisation hydrologique, les philosophies de développement et d'évaluation des modèles hydrologiques seront également évoquées. Enfin, le cas particulier du pas de temps horaire, qui est à l'origine de cette thèse, sera présenté. Nous souhaitons ainsi replacer ce travail dans le contexte actuel de modélisation hydrologique.

1.2 Objets de l'hydrologie

L'hydrologie est la science qui s'intéresse à l'eau. Néanmoins, l'eau étant une ressource naturelle tellement répandue à l'échelle de la planète, nécessaire à la vie des êtres vivants et à la fois indispensable au développement des sociétés et des civilisations, l'objet de l'hydrologie est particulièrement difficile à définir avec précision. Le principal sujet d'étude de l'hydrologie est le cycle de l'eau. Il existe de très nombreuses définitions de l'hydrologie et autant de représentations du cycle hydrologique. Selon Penman [1961], « *l'hydrologie se définit comme la science qui répond à la question : Qu'arrive t'il à la pluie ?* ». Selon le glossaire international d'hydrologie [UNESCO-OMM, 1992], l'hydrologie est :

1. « *la science qui traite des eaux que l'on trouve à la surface de la Terre, ainsi qu'au-dessus et en dessous, de leur formation, de leur circulation et de leur distribution dans le temps et dans l'espace, de leurs propriétés biologiques, physiques et chimiques et de leur interaction avec leur environnement, y compris avec les êtres vivants.* »
2. « *la science qui étudie les processus qui régissent les fluctuations des ressources en eau des terres émergées et traite des différentes phases du cycle hydrologique.* »

L'eau étant présente sous de nombreuses formes, dans de nombreux milieux, l'hydrologie est une science très vaste, qui fait appel à de nombreuses connaissances. Comme le dit J. Rodier en préface de l'ouvrage de Roche [1963] : « *Le parfait hydrologue devrait avoir acquis des connaissances approfondies dans des domaines très divers tels que : hydraulique, hydrodynamique fluviale,*

calcul des probabilités, climatologie, géologie, pédologie, géomorphologie ». L'hydrologie est alors souvent segmentée en différents domaines, qui s'intéressent plus particulièrement à une composante du cycle de l'eau : hydrologie de surface, hydrogéologie, glaciologie, etc. Cette compartimentation de l'hydrologie en sous-domaines est parfois regrettable. Il ne faut pas oublier que l'objet d'étude de l'hydrologie, le cycle de l'eau, ignore ces cloisonnements et que le débit d'une rivière est le résultat de nombreux processus qui se déroulent dans les différents compartiments du cycle de l'eau. Ce type de compartimentation est parfois à l'origine d'erreurs conceptuelles graves, comme celles que font souvent les hydrologues dits *de surface*, qui assimilent le bassin topographique au bassin versant de surface d'une rivière en un point et oublient par exemple l'influence que peut avoir l'hydrogéologie (Figure 1-1). Cette erreur conceptuelle révèle d'une certaine manière l'oubli de l'interaction souterrain-surface, qu'il est possible de faire lorsqu'on s'intéresse à la simulation du débit des rivières.

1.3 Le système bassin versant

1.3.1 Définition du bassin versant

Le bassin versant en un point, ou plus précisément dans une section droite d'un cours d'eau, représente la totalité de la surface topographique et géologique drainée par ce cours d'eau et ses affluents à l'amont de cette section. Le bassin versant est un des objets d'étude de base en hydrologie. Il est considéré comme une unité fonctionnelle d'étude des flux d'eau, de matière et d'énergie.

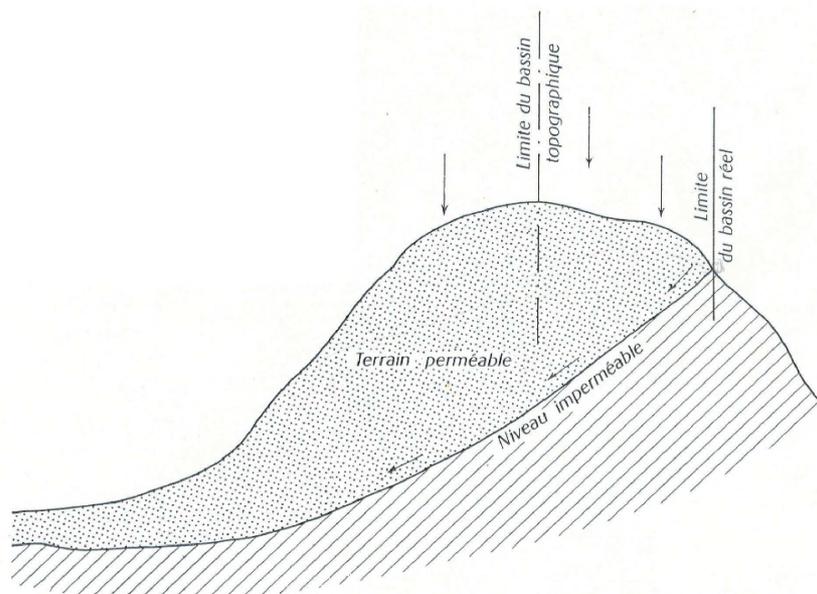


Fig. 64.

Figure 1-1 : Frontières d'un bassin versant [Roche, 1963] : différenciation du bassin versant topographique et du bassin versant hydrogéologique.

Ainsi, les bassins versants représentent une infinie diversité de tailles, de formes et de caractéristiques, allant des petits bassins versants (quelques hectares) à ceux de quelques millions de km² (Volga, Nil, Ob, Mississipi, Amazone, ...).

1.3.2 Approches de représentation du bassin versant

Pour étudier les processus hydrologiques à l'échelle d'un bassin versant, l'approche systémique est couramment utilisée. Selon Dooge [1973] « *un système est n'importe quel structure, dispositif, schéma ou procédure, réel ou abstrait, qui met en relation, dans une base de temps donnée, une entrée, une cause, ou un stimulus de matière, d'énergie ou d'information, et une sortie, un effet ou une réponse, de matière, d'énergie ou d'information* ». Ainsi, le bassin versant peut être considéré comme un système, qui est principalement caractérisé par :

- sa géométrie, sa composition et ses limites,
- ses entrées et ses sorties,
- ses conditions initiales et aux limites,
- ses états internes.

La diversité des bassins versants est infinie et les processus hydrologiques à l'origine des flux en leur sein sont également très nombreux et spécifiques à chaque bassin versant. Quelles que soient la taille d'un bassin versant et les échelles spatiales et temporelles auxquelles on souhaite étudier un tel objet, la connaissance de toutes les composantes du système 'Bassin Versant' est en pratique impossible. De plus, des incertitudes difficilement quantifiables résident dans la mesure et la représentativité de certaines des caractéristiques de ce système.

A l'échelle du bassin versant, le bilan d'eau peut s'exprimer de la manière suivante :

$$P = Q + ETR \pm \Delta S \pm L \quad [1.1]$$

où, P est la pluie reçue, Q est le débit (jaugé) à l'exutoire, ETR est l'évapotranspiration réelle, ΔS est la variation de stock d'eau et L représente les échanges non-atmosphériques avec l'extérieur du système.

A l'heure actuelle, il est seulement possible de mesurer la pluie reçue et le débit. L'évapotranspiration réelle est très difficilement mesurable à l'échelle du bassin versant. Elle est plutôt estimée à partir d'un modèle d'évapotranspiration potentielle (ETP). La variation de stock et les pertes du bassin versant ne sont pas mesurables non plus et sont souvent abusivement négligées [Le Moine, 2005]. La conséquence de ces problèmes de mesure est qu'il est souvent difficile de boucler le bilan d'eau à l'échelle du bassin versant [Beven, 2001b]. Néanmoins, il n'y a aucune raison pour considérer que le bassin versant est un système fermé par le bas, dont le bilan se

limite à $P = Q + ETR + \Delta S$ [Mouelhi, 2003; Le Moine, 2005]. Cette hypothèse a une forte influence sur les structures et les fonctions des modèles hydrologiques.

En caricaturant légèrement cette situation, l'étude des processus hydrologiques à l'échelle du bassin versant est quelque chose d'assez extraordinaire, étant données toutes les incertitudes associées à cette pratique et la difficulté à appréhender cet objet :

- ses limites sont globalement mal connues, la majeure partie des phénomènes qui s'y passent est souterraine ;
- la pluie est mesurée en quelques points du bassin versant par des pluviomètres dont la surface de captation est de $2 \cdot 10^{-8}$ à $4 \cdot 10^{-8}$ km², ce qui représente dans des cas très favorables à peine un milliardième de la surface d'un bassin de quelques centaines de km²
- l'évapotranspiration réelle n'est pas mesurée ;
- les débits extrêmes que l'on essaye souvent de simuler ou de prévoir, sont généralement connus avec une forte incertitude.

Face à une telle diversité de systèmes 'Bassin versant' et face aux incertitudes inhérentes à l'hydrologie, il est possible d'appréhender le bassin versant selon deux principales approches (Figure 1-2) :

- **une approche structuraliste** (également appelée holistique, non réductionniste, systémique, phénoménologique, etc.) : elle vise à étudier le bassin versant dans son ensemble, de manière systémique à partir du développement de relations entre les entrées et les sorties du système. Cette approche n'a aucune prétention sur le lien entre les paramètres optimisés et ses variables internes, avec les particularités d'un bassin versant ;
- **une approche réductionniste** (également appelée physique, représentationnelle, etc.). Elle vise à décrire et étudier le bassin versant en établissant une chaîne causale, basée sur les théories de la physique des écoulements, capable d'expliquer le comportement du bassin versant.

La manière dont est schématisé le bassin versant dépend du type d'approche suivie, structuraliste ou réductionniste. Les tenants de l'approche structuraliste commencent par schématiser le système bassin versant dans son unicité puis complexifient progressivement leur représentation, au fur et à mesure qu'ils cherchent à reproduire avec plus d'exactitude son fonctionnement. Les tenants de l'approche réductionniste cherchent plutôt à représenter le bassin versant dans sa diversité, de manière aussi exhaustive que possible, pour expliquer son fonctionnement à l'aide de la description physique des différents processus, dont ils supposent l'influence.

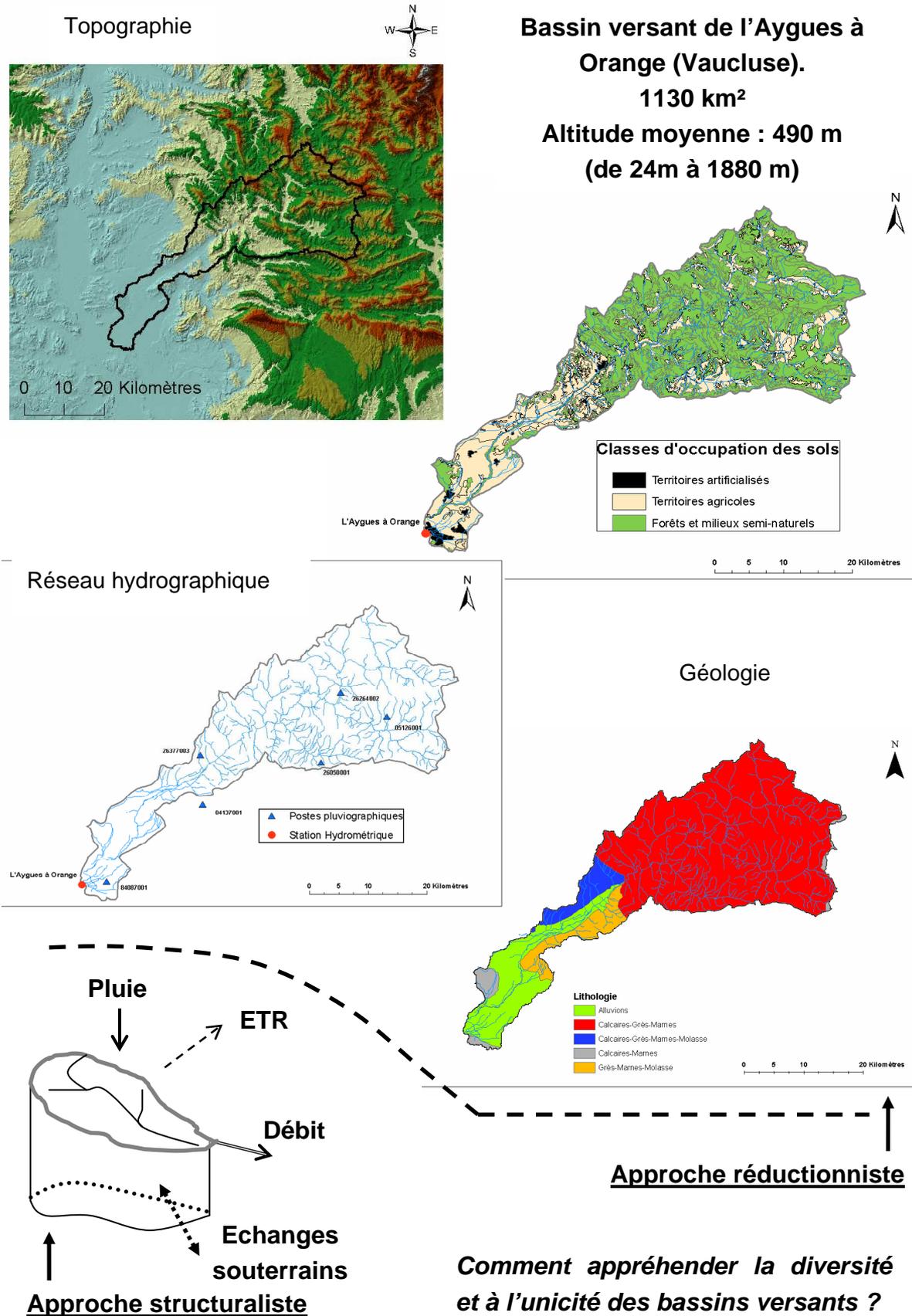


Figure 1-2 : Diversité et unicité des bassins versants. Quelle approche pour appréhender leurs fonctionnements ?

Bien évidemment, il n'existe pas de réponse dogmatique sur la bonne manière de décrire le système et la solution se trouve peut être entre les deux approches [Klemeš, 1983; Sivapalan et al., 2003]. Néanmoins, dans cette thèse, on comprendra rapidement que l'approche retenue est structuraliste et qu'en première approximation, je préfère regarder le bassin versant avec une carte à l'échelle 1/25000, plutôt qu'à l'échelle 1/1 (citation de J. L. Borges).

1.4 Modélisation hydrologique

Un modèle est une représentation sous une forme quelconque d'un objet, d'un processus, ou d'un système [UNESCO-OMM, 1992]. Un modèle hydrologique est défini comme la représentation simplifiée de tout ou partie des processus du cycle hydrologique par un ensemble de concepts hydrologiques, exprimés en langage mathématique et reliés entre eux dans des séquences temporelles et spatiales correspondant à celles que l'on observe dans la nature.

Dans cette thèse, les modèles hydrologiques que nous considérons ont pour unique ambition la reproduction des débits observés à l'exutoire d'un bassin versant. Accroître cette ambition, en essayant de reproduire d'autres variables hydrologiques (ETR, niveaux piézométriques, autres flux, etc.), accroît immédiatement la complexité de l'exercice de modélisation, ce qui ne peut être envisagé que si les données observables correspondantes sont disponibles, ce qui est rarement le cas.

L'action de modéliser sous-entend un certain nombre d'hypothèses implicites. Un modèle (sa construction, sa structure) est défini par l'objectif auquel il doit répondre, ce qu'il doit modéliser. Un modèle (aussi bien analogique, réduit, numérique, conceptuel, etc.), comme son nom l'indique, est une certaine image d'une réalité perçue par le modélisateur.

Ainsi, modéliser nécessite également de réaliser un compromis entre 'généralité' (la propriété du modèle à être transposable à d'autres systèmes de même type), 'réalisme' (capacité à décrire réellement le système) et 'précision', pour représenter le système modélisé [Kauark-Leite et Nascimento, 1993]. D'une certaine manière, un modèle hydrologique est une hypothèse proposée de relation entre la pluie et le débit observé.

En hydrologie, les modèles constituent des outils de base, de simulation et de prévision, essentiels à la gestion adéquate des ressources en eau, afin de répondre de manière optimale aux différents besoins et usages des ressources en eau [Hubert, 1984].

1.4.1 Typologie des modèles hydrologiques

Un modèle hydrologique pluie-débit est généralement défini par (Figure 1-3) :

- **ses variables d'entrée** (variables indépendantes) : il s'agit des entrées du modèle, qui sont essentiellement les chroniques de pluie, d'ETP ou de température,
- **ses variables de sortie** (variables dépendantes) : il s'agit des sorties du modèle, qui sont généralement les débits simulés à l'exutoire du bassin versant, mais qui peuvent parfois être aussi l'ETR, des niveaux piézométriques, etc.,
- **ses variables d'état** : il s'agit des variables internes au système, qui évoluent en fonction du temps et rendent compte de l'état du système à un moment donné. Typiquement, ces variables sont les niveaux de remplissage des différents réservoirs (neige / production / routage),
- **ses paramètres** : les paramètres des modèles hydrologiques, qu'ils aient une pseudo-signification physique ou qu'ils soient calés, servent à adapter la paramétrisation des lois régissant le fonctionnement du modèle, au bassin versant étudié,
- **ses performances** : il s'agit d'estimer l'amplitude des erreurs de modélisation, calculées généralement sur la base d'une mesure de l'écart entre les valeurs simulées et les valeurs mesurées.

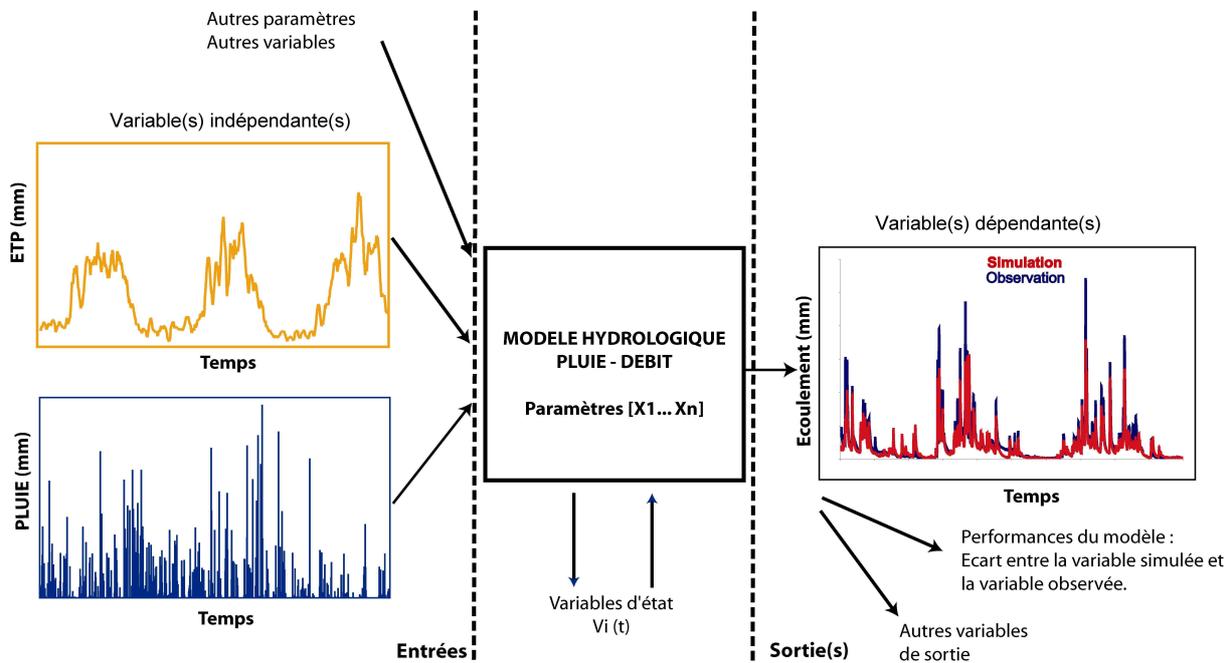


Figure 1-3 : Représentation du fonctionnement d'un modèle hydrologique

La typologie des modèles hydrologiques vise à classer les différents types de modèles disponibles selon quelques critères comme (Tableau 1-1) :

- la philosophie de développement,

- le pas de temps de simulation,
- le mode de fonctionnement (continu, événementiel).

Ainsi, un modèle hydrologique est défini par la manière dont il représente le système 'bassin versant' (sa structure), les équations qui établissent les relations entre les différents éléments du modèle et les paramètres à prendre en compte.

Tableau 1-1 : Critères de classification des modèles hydrologiques globaux. Les critères soulignés représentent les thèmes abordés pendant cette thèse.

Critère de classification	Type de modèle
Fenêtre temporelle de simulation	Événementiel <u>Continu</u>
Pas de temps	<u>Horaire</u> <u>Journalier</u> Mensuel Annuel Pluri-annuel
Fondement théorique	Mathématique (Modèle boîte noire) <u>Conceptuel (Modèle à réservoirs ou boîte grise)</u> <u>Empirique (Modèle à réservoirs ou boîte grise)</u> Théorique (Modèle fondé sur la physique)
Paramètres	<u>Estimés par calage</u> Mesurés sur le terrain

De manière générale, les grandes distinctions entre les différents types de modèles sont leurs fondements théoriques et leur caractère distribué ou non. En effet, la modélisation hydrologique est menée depuis son origine par deux approches de pensée, généralement opposées, l'une basée sur l'approche structuraliste et l'autre basée sur l'approche réductionniste. La prolifération actuelle de modèles, synthétisant les connaissances des personnes qui les développent, souligne l'état parcellaire des connaissances en hydrologie, puisque chaque modèle est une hypothèse proposée pour représenter (expliquer ?) la relation entre la pluie et les débit, à l'échelle du bassin versant. Malgré les bonnes performances de certains de ces modèles, aucun modèle, aucune philosophie de modélisation, ne s'impose de manière inconditionnelle par rapport aux autres aujourd'hui, ce qui souligne la connaissance limitée du fonctionnement hydrologique des bassins versants [Klemeš, 1986a].

1.4.2 Applications des modèles hydrologiques

Les modèles hydrologiques sont un des outils dont disposent les hydrologues pour essayer de comprendre, de simuler et de prévoir le fonctionnement d'un bassin versant. En fonction du pas de temps choisi, les modèles globaux permettent de répondre à différentes questions s'intéressant à la gestion des risques hydrologiques et des ressources en eau.

En se limitant aux applications quantitatives des modèles hydrologiques, leurs domaines d'applications sont vastes et concernent aussi bien la recherche que l'ingénierie. Ils sont notamment utilisés pour :

- **la reconstitution ou la simulation** : le but du modèle est le comblement des lacunes contenues dans une série chronologique de débit ou l'extension de séries de débit ;
- **la prévision** : le but du modèle est essentiellement l'anticipation des évolutions futures du débit d'une rivière, ce qui permet dans un contexte opérationnel, une meilleure gestion des ressources en eau en période d'étiage et l'anticipation des risques d'inondations en période de crue ;
- **la prédétermination** : le modèle est utilisé pour la simulation de séries chronologiques de débits, dans un but d'analyse fréquentielle des extrêmes : débits associés à une période de retour pour un étiage ou une crue [Lamb, 1999; Perrin et al., 2001b] (dimensionnement d'ouvrages d'art, déversoirs de crue, ponts, etc.) ou la délimitation de zones inondables (Plan de Prévention de Risques d'Inondation, etc.). Pour étudier les événements particulièrement rares, les modèles pluie-débit peuvent être couplés à des générateurs (stochastiques) de pluie, capables de générer de longues chroniques de pluie [Boughton et Droop, 2003] ;
- **l'analyse de la non-stationnarité d'un bassin versant** : dans un contexte de variabilité climatique et/ou de modifications anthropiques des caractéristiques d'un bassin versant, les impacts sur le fonctionnement hydrologique d'un bassin versant peuvent éventuellement être recherchés en effectuant des tests sur la stationnarité des vecteurs de paramètres et/ou des sorties du modèle, comparativement aux observations [Andréassian, 2002; Schreider et al., 2002; Niel et al., 2003; Le Lay et al., 2005].
- **L'utilisation de variables d'état du bassin versant pour la modélisation annexe** : le modèle est utilisé pour générer des chroniques d'état du bassin versant (taux de remplissage d'un des réservoirs, stock de neige, etc.) qui sont ensuite utilisées par d'autres modèles (flux de nutriments ou de matières en suspension), prévisions d'apport par fonte nivale, ou pour améliorer les performances d'autres modèles [Anctil et al., 2004b ; Oudin et al., soumis].

Dans un contexte opérationnel, les modèles hydrologiques sont utilisés pour répondre à deux problématiques principales : l'évaluation des risques hydrologiques et la gestion de la ressource en eau. La pertinence des réponses apportées par ces modèles à l'une des deux problématiques est intimement liée à la structure du modèle et à la philosophie de développement qui a conduit à sa structure. Ainsi, lors de l'utilisation d'un modèle hydrologique dans un cas concret, il est impératif de connaître le domaine d'application et les hypothèses conditionnant l'utilisation d'un modèle en particulier.

1.4.3 Qualités requises des modèles hydrologiques

Les modèles auxquels on s'intéresse dans cette thèse font partie des modèles hydrologiques pluie-débit, conceptuels ou empiriques. Face à la diversité des bassins versants et des applications possibles des modèles hydrologiques, de nombreux auteurs [*Nash et Sutcliffe, 1970; Bergström et Forsman, 1973; Linsley, 1982; Klemeš, 1986b; Bergström, 1991; Lindström et al., 1997*] ont fait la liste des qualités attendues des modèles hydrologiques, ce sont principalement les suivantes :

- Le modèle doit être basé sur des fondements scientifiques sains ;
- Les données nécessaires au bon fonctionnement du modèle doivent être en accord avec la disponibilité de celles-ci ;
- La complexité du modèle doit être aussi limitée que possible pour atteindre un objectif de performance donné ;
- Le modèle doit être correctement validé, général, transposable et robuste ;
- Le modèle doit être compréhensible par les utilisateurs.

L'évaluation de ces qualités peut se faire, soit de manière qualitative (appréciation), soit de manière quantitative et objective (critères numériques). Ainsi, un des moyens d'évaluer la capacité d'un modèle à produire de bonnes simulations, est de conduire une étude comparative des performances de différentes structures de modèles, sur un certain nombre de bassins versants [*Perrin, 2000*].

1.5 Le cas du pas de temps horaire

1.5.1 Des pluies et des crues au pas de temps horaire

La pluie est un phénomène discontinu dans le temps et dans l'espace. Le débit mesuré en un point d'un cours d'eau représente l'intégration des processus hydrologiques à l'origine des écoulements en ce point du bassin versant. Pour être estimées, ces deux variables hydrologiques sont mesurées, selon une certaine fréquence. Ainsi, entre deux dates de mesures, les pluies sont cumulées et les débits

sont moyennés. Bien évidemment, cette moyenne a pour effet de réduire l'intensité (Figure 1-4) et la variabilité du phénomène mesuré.

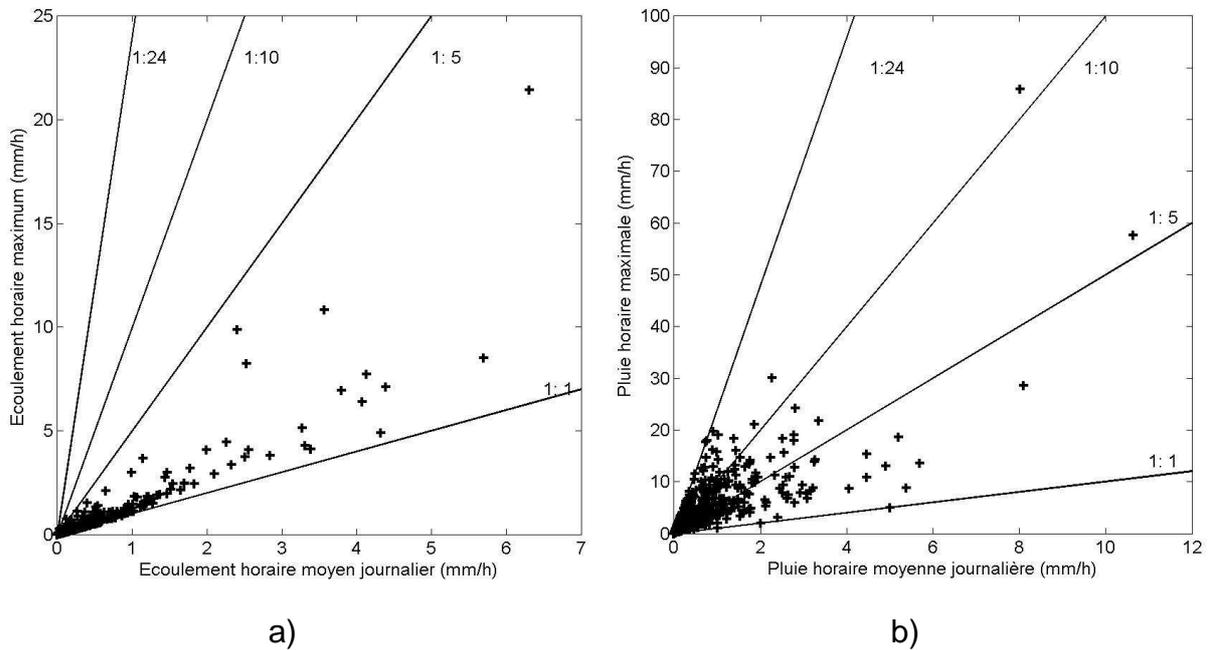


Figure 1-4 : Comparaison entre les intensités horaires maximales et journalières. a) pluie et b) débit de la Goudèche à Saint.-Maurice de Ventalon.

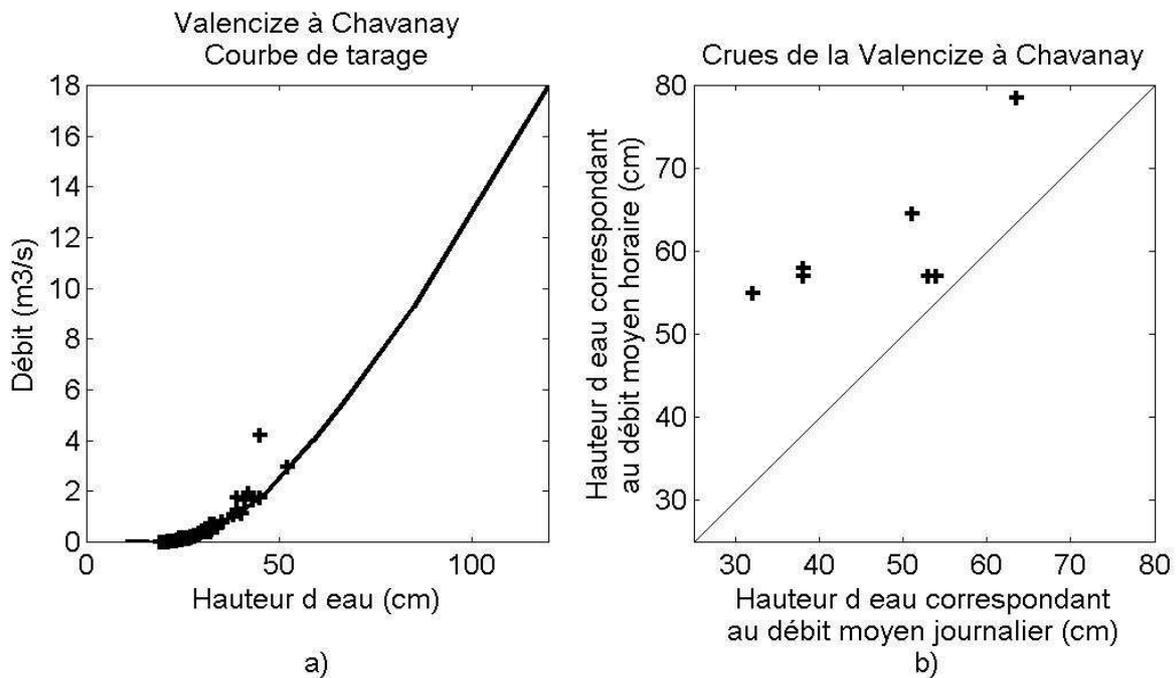


Figure 1-5 : a) courbe de tarage de la Valencize à Chavanay, V3315010, 36 km². b) comparaison des cotes de la rivière atteintes pendant quelques crues, correspondant au débit moyen journalier et au débit moyen horaire maximum.

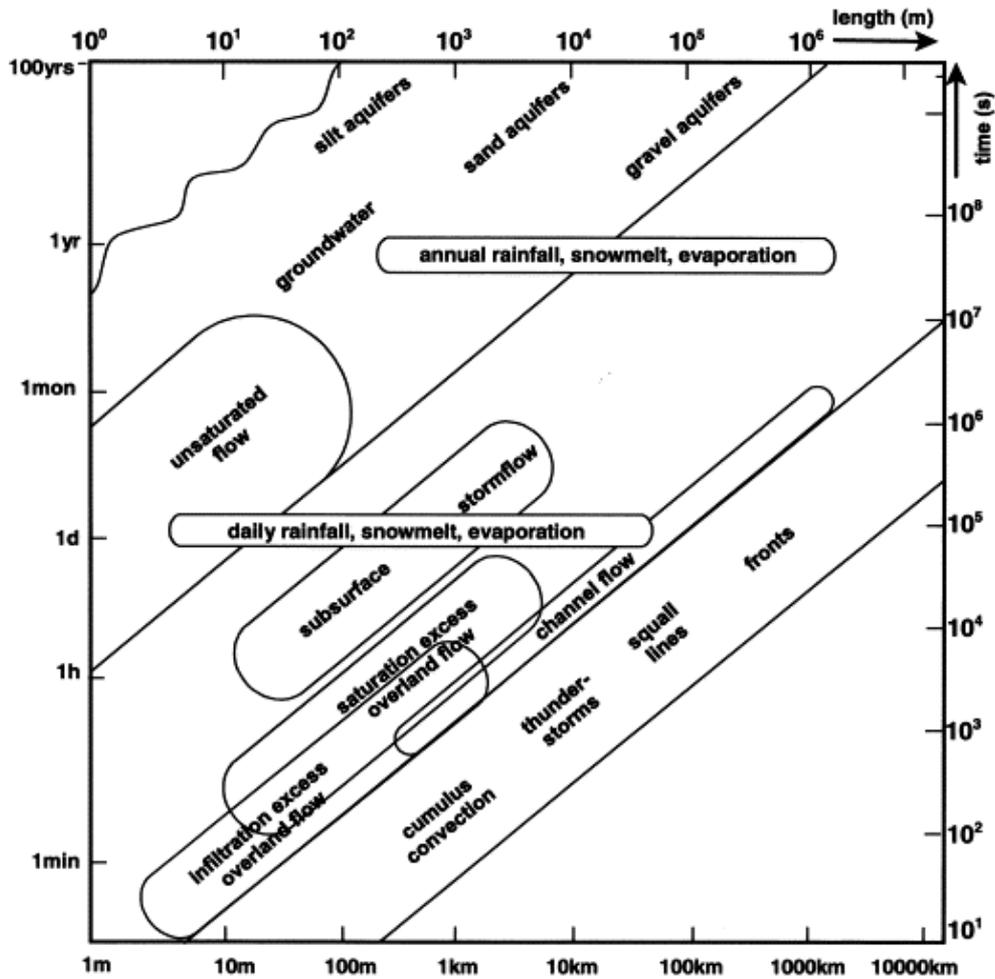


Figure 1-6 : Échelles spatio-temporelle de différents processus [Blöschl et Sivapalan, 1995].

Si la variabilité du phénomène étudié a une fréquence inférieure à 2 fois le pas de temps d'échantillonnage, une partie de l'information disparaît (*Théorème de Shannon*). Comme le montre la Figure 1-4, qui concerne le bassin de la Goudèche à Saint-Maurice de Ventalon (Lozère), le rapport entre le débit maximum horaire et le débit journalier est autour de 2, et peut aller jusqu'à un facteur 4.

Pour la pluie ce rapport oscille en moyenne autour d'un facteur 5 à 10. D'un point de vue local, cette différence peut représenter des hauteurs d'eau sans commune mesure. Ainsi, être capable de simuler correctement le débit moyen journalier d'un bassin versant où son pic de crue peut signifier des situations locales très différentes.

A titre d'exemple, la Figure 1-5 a) montre la courbe de tarage de la Valencize à Chavanay, dont le bassin versant est de 36 km². La Figure 1-5 b) montre la comparaison des hauteurs d'eau au niveau de la station de jaugeage obtenues, soit à partir de la valeur de débit maximum horaire, soit à partir du débit journalier. Pour les sept crues retenues, le rapport entre le débit maximum horaire et le débit

journalier est entre 1.6 et 5.4. D'un point de vue local, cet écart entre valeurs horaires et valeurs journalières se traduit par un rapport de hauteurs d'eau allant de 1.1 à 1.7. Ainsi, d'un point de vue local, le suivi d'une rivière à un pas de temps horaire ou journalier ne se traduit pas du tout par la même configuration. Même si dans l'exemple choisi les variations d'amplitudes des hauteurs peuvent paraître faibles, pour d'autres bassins versants, elles peuvent être considérables (elles dépendent fortement de la configuration géométrique de la section d'écoulement).

Ainsi, lorsqu'on s'intéresse à la relation pluie-débit, et à la simulation ou la prévision des crues en particulier, il peut être très important d'étudier le comportement des petits bassins versants à un pas de temps fin, compatible avec leur temps de réponse. Le pas de temps horaire apporte alors beaucoup plus d'informations sur le fonctionnement des bassins versants que le pas de temps journalier. En effet, ce pas de temps, représente un long pas de temps d'intégration par rapport aux échelles de temps caractéristiques des processus hydrologiques, pour des échelles spatiales allant jusqu'à plusieurs centaines de km², voire quelques milliers de km², suivant les caractéristiques du bassin versant (Figure 1-6).

1.5.2 Applications des modèles hydrologiques au pas de temps horaire

Dans le contexte actuel d'augmentation de la disponibilité des données à pas de temps fin et de l'augmentation du besoin de prévisions hydrologiques [Takeuchi, 2002], le passage au pas de temps horaire devient un atout considérable. En effet, pour un bon nombre de bassins versants aux temps de réaction rapides, l'utilisation du pas de temps horaire est indispensable pour réaliser des prévisions des crues ou des études de prédétermination, pertinentes par rapport à la réactivité des bassins versants. Pour ces bassins versants aux superficies inférieures à quelques centaines de km², le pas de temps horaire s'impose.

Par exemple en France, une bonne partie des bassins versants de l'arc méditerranéen est soumise à des crues intenses lors d'évènements de type cévenol. Les bassins de la Lèze (Haute-Garonne, Ariège), de l'Orbieu, de l'Aude, de l'Hérault, des Gardons, de l'Ardèche, de la Loire Amont et du Var sont connus pour leurs comportements violents et sont plus particulièrement suivis par les services de l'Etat (SPC et SCHAPI). Le pas de temps horaire est utilisé par un bon nombre de systèmes opérationnels de prévision des crues. C'est notamment le cas des systèmes développés au Cemagref pour les bassins versants du Nord-Pas de Calais, de l'Oise ou de la Sarthe [SAFEGE et CEMAGREF, 2002; Perrin, 2003; SAFEGE et CEMAGREF, 2003] ou des services opérationnels d'EDF-DTG. Dans des comparaisons de méthodes de prévision des crues, le pas de temps horaire est également le pas de temps utilisé pour les bassins versants réactifs (études hydro-

météorologiques de l'OHMCV ; expérience du BVNE, mise en place par le SCHAPI et le MEDD ; projet DMIP, de l'American Weather Service [Smith et al., 2004]).

Pour la prédétermination des crues, le pas de temps horaire reste encore assez peu utilisé [Cameron et al., 1999]. Dans la revue de Boughton et Droop [2003] sur l'utilisation des modèles hydrologiques pour l'estimation des crues de projets, il semble que la plupart des études aient été réalisées au pas de temps journalier. Néanmoins, pour des bassins versants de quelques centaines de km², le rapport entre le débit moyen journalier et le débit maximum instantané est en général de l'ordre de 2 et assez variable pour un même bassin versant (Figure 1-4 et Figure 2-12 d), en fonction des conditions hydro-climatiques. Ainsi, l'utilisation d'un coefficient de sécurité en fonction du bassin versant n'est pas forcément judicieuse et il peut s'avérer plus intéressant d'utiliser un modèle au pas de temps horaire, pour avoir une meilleure estimation des débits de crue.

En modélisation hydrologique, la littérature est prolifique au pas de temps journalier. Néanmoins, on trouve de plus en plus d'applications au pas de temps horaire. Notamment, une grande partie des applications de TOPMODEL a été faite à pas de temps fin et/ou sur des événements de crue (entre autres, Franchini et Pacciani [1991], Franchini et al. [1996], Saulnier [1996], Datin [1998], Zin [2002]). L'étude de Franchini et Pacciani [1991] est la seule comparaison de différentes structures de modèles pluie-débit conceptuels, testés au pas de temps horaire. Ces auteurs ont testés sept modèles, malheureusement sur un seul bassin versant pour lesquels ils disposaient d'une chronique horaire de quelques mois. Néanmoins, ces auteurs ne se sont pas posés la question de la particularité du pas de temps, par rapport à la modélisation pluie-débit. Schreider et Jakeman [2001] présentent l'application du modèle IHACRES au pas de temps de 4 heures, sur une dizaine de bassins versants australiens. Quelques applications du modèle HBV ont également été faites à ce pas de temps [Kobold et Brilly, 2004]. On trouve par ailleurs de nombreux articles sur la prévision des crues utilisant le pas de temps horaire, mais ces applications sont généralement événementielles.

Dans l'ensemble, on trouve dans la littérature des applications de modèles au pas de temps horaire, mais il ne semble pas que les structures de ces modèles aient été exhaustivement testées pour ce pas de temps, comme elles ont pu l'être au pas de temps journalier [WMO, 1975; WMO, 1992; Perrin, 2000; Perrin et al., 2001a]. Les modèles sont généralement utilisés de la même manière aux pas de temps horaire ou journalier.

1.6 Philosophie de développement des modèles hydrologiques : sept propositions pour le développement empirique de modèles hydrologiques^{*1}

Depuis plus de 20 ans, des modèles hydrologiques fonctionnant à différents pas de temps et pour différents objectifs ont été développés au Cemagref de manière progressive. Le développement de ces structures de modèles s'est principalement appuyé sur les philosophies de *Nash et Sutcliffe* [1970], *Bergström et Forsman* [1973], *Klemeš* [1986b], c'est-à-dire une approche structuraliste (non-réductionniste) de modélisation pluie-débit et une amélioration empirique des performances des modèles.

Le développement des modèles du Génie Rural (GR) et celui d'un modèle au pas de temps horaire que nous proposons dans cette thèse, repose sur sept propositions relatives à la modélisation pluie-débit. Ces sept propositions^{*} visent à permettre aux modèles d'acquérir les qualités attendues d'un 'bon' modèle hydrologique et s'appuient globalement sur une démarche basée sur la complexification progressive et raisonnable de la représentation de la relation pluie-débit à l'échelle du bassin versant [*Perrin et al.*, 2004].

1.6.1 Proposition n°1 : Une représentation globale du bassin versant

Comme nous l'avons vu précédemment, le bassin versant est une entité particulièrement hétérogène et complexe à appréhender. L'approche que nous avons décidé de suivre est de considérer globalement le bassin versant comme une seule entité, un système, caractérisé par les chroniques de données qui le décrivent. Ce choix est justifié par les raisons suivantes :

- A l'heure actuelle, malgré une bonne compréhension d'une majorité des processus hydrologiques à de petites échelles d'espace, de larges doutes subsistent quant aux processus dominants et ce, particulièrement, à l'échelle du bassin versant (supérieure à quelques km²),
- De plus, les modèles hydrologiques semi-distribués ou distribués sont composés de cellules élémentaires, dont le fonctionnement est celui d'un modèle pluie-débit global. Ainsi, avant d'assembler un grand nombre de

^{*1} Le 'Hall of Fame' de la modélisation hydrologique empirique est rempli de sept sages dont s'inspirent ces travaux et réflexions : John Eamon Nash, Vit Klemeš, Masami Sugawara, Walter Boughton, Ray Linsley, Stein Bergström et Claude Michel.

^{*2} Ces sept propositions ne sont pas des postulats mais bien des choix qui se sont révélés nécessaires pour tendre vers l'objectif de construire des modèles fiables et robustes. Nous nous efforcerons dans les chapitres qui suivent de montrer tout l'intérêt de ces propositions, et éventuellement leur spécificité dans le cadre de notre travail sur le pas de temps horaire.

cellules pour décrire un bassin versant, nous préférons d'abord atteindre les limites des améliorations possibles des performances d'une cellule unique (échelle globale),

- Enfin, la supériorité des approches distribuées par rapport aux approches globales est encore loin d'avoir été clairement démontrée [Loague et Freeze, 1985; Michaud et Sorooshian, 1994; Loumagne et al., 1996; Refsgaard et Knudsen, 1996; Reed et al., 2004; Smith et al., 2004].

Par ailleurs, nous pensons que les progrès réalisés sur une structure globale se peuvent se reporter ensuite sur un modèle distribué basé sur cette structure.

1.6.2 Proposition n°2 : Aucun recours *a priori* à la physique des écoulements

A l'heure actuelle et comme le déplore depuis longtemps [Klemeš, 1986a; Klemeš, 1988], il n'existe pas de théorie générale en hydrologie et, en particulier, pour expliquer la transformation pluie-débit à l'échelle d'un bassin versant. Il est légitime de se demander le sens de l'utilisation, à de grandes échelles de temps et d'espace, des équations de la physique des écoulements en milieux poreux ou à surface libre (Richards, Boussinesq, Saint-Venant, etc.), validées seulement à l'échelle locale. Par ailleurs, il est difficile de déterminer les processus dominants à l'échelle du bassin versant.

Il semble donc préférable de s'orienter vers le développement empirique d'un modèle, qui est guidé par l'information que nous apportent les données (la vraisemblance d'une hypothèse de modélisation, étant mesurée par la performance d'un modèle sur les données d'un certain nombre de bassins versants) et non par les conceptions *a priori*, subjectives et arbitraires, de ce que devraient (pourraient ?) être la relation entre la pluie et le débit à l'échelle du bassin versant. Comme le souligne [Klemeš, 1983], il est regrettable de s'appuyer trop fortement sur les diagrammes convenus sur le cycle hydrologique : « *instead of searching for feasible ways of conceptualization of hydrological processes, they [hydrologists] postulate the structures of their models on the basis of arbitrarily embroidered high-school diagrams of the hydrologic cycle with little concern for testability* ». Ainsi, pour décrire ce mode empirique de développement d'un modèle, on pourrait reprendre le terme anglais '*data driven approach*', généralement utilisé pour les techniques de traitement du signal [Young et al., 1997; Young, 2003].

1.6.3 Proposition n°3 : Augmentation progressive de la complexité d'une structure

Cette proposition veut que pour le développement d'une structure d'un modèle, à un pas de temps donné, pour un objectif donné, on parte d'une structure simple, éventuellement mise au point à un pas de temps supérieur. Cette idée s'inspire fortement de l'approche de modélisation descendante ('*downward approach*')

introduite par *Klemeš* [1983] et *Sivapalan et al.* [2003] dont on trouve récemment de nombreux exemples [*Jothityangkoon et Sivapalan*, 2001, *Jothityangkoon et al.*, 2001, *Atkinson et al.*, 2002, *Eder et al.*, 2003]. Cette stratégie propose que le développement d'un modèle hydrologique se fasse en augmentant progressivement la complexité du modèle, au fur et à mesure qu'on diminue le pas de temps simulé par le modèle. L'idée est que, lorsque l'on veut représenter la relation pluie-débit, pour représenter correctement la variabilité des débits à un pas de temps plus fin, il est déjà important de pouvoir la représenter à un pas de temps supérieur.

Ainsi, dans les travaux cités, les modèles commencent à être développés au pas de temps annuel, puis des fonctions sont incorporées pour représenter la variabilité mensuelle, puis journalière, des débits. Cette stratégie permet ainsi de conserver un niveau de complexité minimal par rapport au pas de temps des données utilisées.

Les travaux du Cemagref sur le développement de modèles hydrologiques, du pas de temps pluri-annuel, jusqu'au pas de temps journalier [*Michel*, 1983; *Edijatno et Michel*, 1989b; *Edijatno*, 1991; *Nascimento*, 1995; *Edijatno et al.*, 1999; *Perrin*, 2000; *Mouelhi*, 2003], ont ainsi visé à trouver une cohérence entre les modèles fonctionnant à différents pas de temps, tout en recherchant le niveau de complexité minimal.

Par ailleurs, la difficulté à comprendre le fonctionnement d'un modèle est généralement proportionnelle à la complexité de sa structure, au nombre de réservoirs utilisés et de paramètres libres. Ainsi, lorsqu'on cherche à améliorer les performances d'une structure de modèle, il semble plus facile de partir d'une structure simple que l'on comprend, plutôt que d'une structure complexe dont on a du mal à appréhender le fonctionnement réel.

1.6.4 Proposition n°4 : Justification de la complexité de la structure d'un modèle par ses performances

Cette proposition est basée sur la nécessaire simplicité d'un modèle. En effet, les modèles hydrologiques étant déjà difficilement falsifiables, il est important de limiter au maximum l'arbitraire dans leur structure. Le principe de parcimonie se traduit par le fait de ne conserver que les composantes du modèle qui sont nécessaires et justifiées par de bonnes performances. Ce principe de minimalité, similaire au fameux rasoir d'Occam (i.e., si un modèle simple suffit, aucun autre modèle complexe n'est nécessaire) permet de réduire l'arbitraire dans la construction des structures des modèles hydrologiques [*Kauark-Leite et Nascimento*, 1993]. Cette approche permet de réduire les risques d'équifinalité que peut générer une structure de modèle.

Cette position :

- a été introduite en modélisation hydrologique par Nash et Sutcliffe [1970] : « *Each additional part of a model must substantially extend the range of application of the whole model. In other words, we are prepared to accept additional parts and hence greater difficulty in determining parametric values only if increased versatility of the model makes it much more likely to obtain a good fit between observed and computed outputs* »;
- elle est soutenue par de nombreux auteurs, dont Bergström [1991] : « *Going from complex to simpler model structures requires an open mind, because it is frustrating to have to abandon seemingly elegant concepts and theories. It is normally much more stimulating, from an academic point of view, to show significant improvement of the model performances by increasing complexity. [...] The idea is that only modifications that significantly improve the results are accepted. [...] We have found that the strategy of going from the simple to the complex is still valid and very useful, in particular for the development of operational basin-oriented models. [...] The point of diminishing returns (no model improvement) was reached surprisingly soon when increasing model complexity* ».

Par ailleurs, un niveau de complexité supérieur au niveau de complexité minimum se traduit souvent par une plus faible identifiabilité des paramètres et des interactions entre les différents paramètres d'un modèle.

A l'heure actuelle, de nombreux auteurs [Beven, 1989; Jakeman et Hornberger, 1993; Perrin et al., 2001a] s'accordent pour dire qu'il suffit de 3 à 6 paramètres libres pour identifier correctement la relation pluie-débit à l'échelle du bassin versant, au pas de temps journalier. La supériorité des modèles plus riches en paramètres n'a jamais été démontrée. Néanmoins, d'après Gupta et al. [1998], Yapo et al. [1998] et Gupta [2000] les plus faibles performances des modèles complexes viennent des stratégies de calage mono-objectif utilisées dans ces études comparatives, qui ne permettent pas de tirer profit de la versatilité des modèles ayant un grand nombre de degrés de liberté.

1.6.5 Proposition n°5 : Recherche de structures de modèle générales

Un problème rencontré en modélisation est que certains modèles ont été développés et testés sur un petit nombre de bassins versants, ayant généralement des conditions hydro-climatiques fortement homogènes. Une conséquence est qu'il arrive que certaines structures manquent de robustesse et de généralité. Ainsi, leurs performances sont fortement dégradées lorsqu'elles sont appliquées à d'autres bassins versants. Néanmoins, une des qualités de base attendues d'un modèle est

justement d'être général et transposable. En effet, que ses applications relèvent de la recherche ou de l'ingénierie, il est important qu'un modèle soit capable d'être efficace pour différents types de bassins versants et de climats. A priori, il est généralement difficile de savoir si un modèle va être efficace ou non sur un bassin versant. Ainsi, il est important que le choix d'un modèle s'oriente vers celui qui a statistiquement de meilleures performances que les autres*.

Par exemple, si on reprend l'exemple du développement des modèles de la famille GR, l'utilisation de bassins versants intermittents et semi-arides a permis d'améliorer la structure du modèle en incluant une fonction capable de simuler correctement les débits de ces bassins (fonction d'échanges souterrains, paramètre X_3). Cette fonction a d'ailleurs permis de faire progresser les performances du modèle sur l'ensemble des autres types de bassins versants. Si ces bassins versants intermittents n'avaient pas été utilisés dans le développement du modèle, il n'est pas sûr que le débit d'un bon nombre de rivières françaises puisse être correctement simulé pour les années de sécheresse 1976 ou 2003, pendant lesquelles les étiages ont été très marqués.

D'une certaine manière, les travaux du Cemagref sur le développement de structures de modèles à différents pas de temps représentent une approche descendante à deux dimensions. Au lieu de se concentrer sur le développement d'une structure de modèle représentant au mieux les propriétés émergentes d'un système particulier (un bassin versant), ces travaux ont cherché à trouver la chaîne de structures représentant au mieux les propriétés émergentes du système bassin versant en général, à différents pas de temps.

1.6.6 Proposition n°6 : Utilisation de larges échantillons de bassins versants pour tester les performances de différentes structures de modèles

On trouve généralement dans la littérature de nombreuses comparaisons des performances de différentes structures de modèles. Sans remettre en cause la justesse des résultats de ces travaux, on peut regretter que ces comparaisons n'aient concerné que peu de structures de modèles, comparées sur peu de bassins versants aux conditions généralement homogènes. Ainsi, il devient difficile d'en déduire des résultats généraux, tellement la variabilité des procédures de comparaison est grande.

Ainsi, pour être consistante, nous pensons qu'il faut que la comparaison des performances de différentes structures de modèles porte sur un grand nombre de bassins versants aux caractéristiques hydrologiques, physiographiques et

* Dans le cadre d'une opération délicate et douloureuse d'un genou, vers quel médecin s'orienter ? Celui dont je sais qu'il est capable de réussir une opération de temps en temps ou bien celui dont je sais qu'il loupe rarement une opération ? C'est votre genou, à vous de choisir...

climatiques variées. Par ailleurs, il est important d'utiliser des critères dont le traitement est compatible avec un grand nombre de bassins versants.

1.6.7 Proposition n°7 : Évaluation d'un modèle par comparaison de différentes structures de modèle

Cette dernière proposition porte sur l'évaluation des performances et la validation d'un modèle. En pratique, la validation d'un modèle hydrologique est impossible de manière absolue. A l'heure actuelle, il n'existe pas de théorie de la relation pluie-débit à l'échelle du bassin versant qui permette de conclure qu'un modèle est 'vrai' ou 'faux', 'bon' ou 'mauvais'. En fait, tous les modèles sont faux, il y en a néanmoins quelques-uns qui sont un peu moins mauvais que d'autres. Encore faut-il pouvoir les discriminer.

Pour contourner ce problème de falsification, un des moyens de juger des qualités d'un modèle est d'évaluer ses performances non plus de manière absolue, mais plutôt de manière relative, comparativement à d'autres modèles. La mesure absolue des performances d'un modèle apporte en effet assez peu d'information. Que signifie qu'un modèle soit capable d'expliquer 85% de la variance d'une chronique hydrologique, si un autre modèle plus simple en explique 90% ? Il semble ainsi que, dans l'état actuel des connaissances, un des seuls moyens de valider provisoirement un modèle hydrologique est de le confronter à d'autres modèles hydrologiques.

Cette évaluation comparative des performances d'un modèle peut se faire par rapport à des modèles plus simples [Wagener, 2003], afin d'évaluer l'intérêt de passer à une modélisation plus complexe. La comparaison des performances de différents modèles doit être basée sur la même procédure de tests (mêmes données d'entrée, même procédure de calage, qui doit être également efficace pour des modèles ayant différents degrés de complexité, etc.).

1.7 Pourquoi travailler au pas de temps horaire ?

Le principal objectif de cette thèse est d'explorer les performances de modèles hydrologiques pluie-débit au pas de temps horaire^{*}. Cette démarche trouve sa justification dans les points détaillés ici.

1.7.1 « Encore un nouveau modèle ? »

Lorsque j'ai commencé cette thèse, un des premiers commentaires que j'ai entendu était : « *Encore un nouveau modèle ?* ». A l'époque je n'ai pas très bien compris ce

^{*} pour éviter toutes confusions, je me permets de rappeler qu'il n'est pas de montrer que les modèles GR sont les meilleurs.

que signifiaient les doutes soulevés par cette question. Il est probable que je n'avais pas été capable de montrer l'intérêt de développer et tester les performances de modèles au pas de temps horaire. Néanmoins, les avancées scientifiques se font rarement sur le court terme et le développement d'un modèle hydrologique est généralement long et étalé sur plusieurs décennies. Ceci est le cas, entre autres, des modèles TANK [Sugawara, 1974; Sugawara, 1995], IHACRES [Jakeman et al., 1990; Croke et Jakeman, 2004], HBV [Bergström et Forsman, 1973; Lindström et al., 1997], XINANJIANG [Zhao et al., 1980; Jayawardena et Zhou, 2000], GR4J [Michel, 1983; Perrin et al., 2003], etc.

Les modifications et les améliorations progressives des structures de modèles hydrologiques viennent de la synthèse du grand nombre de tests effectués par leurs concepteurs. Généralement, les modifications proposées sont minimales et visent à améliorer une composante d'un modèle. Les gains de performances sont souvent modestes, mais témoignent de l'augmentation de la confiance que les modélisateurs ont dans la structure du modèle qu'ils développent.

1.7.2 « Tester un modèle au pas de temps horaire, à quoi ça sert ? »

La deuxième incompréhension que j'ai ressentie a été : « Tester un modèle au pas de temps horaire, à quoi ça sert ? ». Il est vrai que cette question peut être légitime, après les 20 années passées à développer un modèle, GR4J [Michel, 1983; Edijatno et Michel, 1989b; Edijatno, 1991; Nascimento, 1995; Edijatno et al., 1999; Perrin, 2000], dont les performances au pas de temps journalier sont très bonnes. En fait, cette question révèle une faible connaissance des processus de développement des modèles hydrologiques empiriques et un fort recours à des *a priori* 'physiques'.

En effet, à l'heure actuelle, beaucoup de structures n'ont été testées que sur quelques bassins versants, alors que d'autres, largement utilisées ou non, ont subi la confrontation face à plusieurs centaines de bassins versants (c'est le cas notamment, TOPMODEL [Beven et al., 1995], HBV [Bergström, 1995], IHACRES [Jakeman et al., 1990] ou GR4J [Perrin, 2000; Rojas-Serna, 2005]). À part les modèles dérivés de TOPMODEL, la plupart des autres modèles sont généralement testés au pas de temps journalier.

Néanmoins, ces modèles, dont les structures ont été développées de manière conceptuelle ou empirique, ont été mis au point par rapport aux données disponibles, qui étaient principalement au pas de temps journalier à cette époque. La structure des modèles hydrologiques développés de manière empirique dépend donc de l'information contenue dans les données. Plus le pas de temps d'agrégation des données augmente, plus on perd de l'information. De cette manière, la variabilité de

la pluie et des débits est diminuée et les fortes intensités, à l'origine des processus hydrologiques extrêmes, peuvent être filtrées. Si on postule que la structure d'un modèle empirique ou conceptuel représente les propriétés du système bassin versant à un pas de temps, rien ne nous assure qu'elle sera capable de représenter également les propriétés de ce système à un pas de temps plus fin.

Ainsi, rien ne prouve *a priori* qu'un modèle efficace à un pas de temps donné, soit encore efficace à un pas de temps plus fin, la complexité des processus à représenter ayant normalement augmenté. Par exemple, dans le cadre de la modélisation pluie-débit aux pas de temps mensuels et journaliers, il est largement accepté et prouvé que les modèles mensuels sont plus simples que les modèles journaliers [Mouelhi, 2003; Mouelhi et al., 2005]. On peut se demander pour quelles raisons, on ne devrait pas vérifier les performances des modèles au pas de temps horaire et essayer de les améliorer.

1.7.3 « Pour la prévision des crues, quel est l'intérêt d'utiliser un modèle continu ? »

Enfin, la troisième incompréhension que j'ai pu identifier vient du mode de fonctionnement du modèle. En effet, lorsqu'on s'intéresse à la simulation ou à la prévision des crues, les modèles sont souvent utilisés de manière événementielle.

L'utilisation de modèles événementiels vient des débuts de l'hydrologie, lorsque la disponibilité des données était faible, la mémoire et la puissance des ordinateurs très limitées. Ainsi, dans ces conditions techniques, la seule solution était de ne s'intéresser qu'aux événements de crue. Depuis, les moyens technologiques ont évolué, mais cette pratique est restée.

Par rapport à la simulation en mode continu, l'utilisation événementielle des modèles pose quelques problèmes [Boughton et Droop, 2003] :

- la manière de choisir un événement et un non-événement est subjective et dépend de l'utilisateur ;
- l'optimisation des paramètres d'un modèle sur quelques événements peut poser des problèmes d'identifiabilité et de forte variabilité des valeurs de paramètres en fonction de l'intensité des crues de l'échantillon de calage ;
- l'initialisation des réservoirs d'un modèle a une large influence sur la qualité de la simulation d'une crue. Ainsi, dans un cadre de comparaison de différentes structures de modèles, l'initialisation des réservoirs des différents modèles a une influence capitale sur les performances des modèles.

Le mode continu présente quant à lui l'avantage d'être plus efficace, plus objectif et plus facile à mettre en place, dans le cadre comparatif que nous suivons :

- il n'y a pas sélection d'évènements ;
- il n'a y a pas de problèmes d'initialisation du niveau de remplissage des différents réservoirs.

Le mode de fonctionnement continu présente en outre l'intérêt de simuler le fonctionnement du bassin versant dans l'ensemble des situations hydrologiques qu'il rencontre, ce qui est nécessaire lorsqu'on a pour objectif de développer un modèle robuste et général.

1.8 Synthèse et objectif de la thèse

Nous venons de présenter le contexte général de modélisation pluie-débit dans lequel se placent les recherches entreprises au cours de cette thèse. La modélisation pluie-débit s'intéresse à la compréhension et à la représentation de la relation entre la pluie et les débits, à l'échelle du bassin versant. La simulation, la prévision et la prédétermination des débits sont des besoins pour lesquels les hydrologues développent des outils qu'ils cherchent à améliorer afin d'augmenter la précision et la validité de leurs estimations. Alors que le besoin de prévisions hydrologiques augmente, les hydrologues sont de plus en plus sollicités pour travailler à des pas de temps plus fins (horaire) afin de mieux rendre compte des débits des bassins versants réactifs, les débits de pointes pouvant être bien supérieurs aux débits moyens journaliers.

Parmi les différents types de modèles hydrologiques à notre disposition, les modèles conceptuels ou empiriques globaux ont été identifiés comme étant des outils propices pour répondre de manière robuste et efficace à certains des besoins de la recherche et de l'ingénierie en hydrologie, tout en nécessitant une quantité raisonnable de données [Nash et Sutcliffe, 1970; Bergström et Forsman, 1973; Linsley, 1982; Klemeš, 1986b; Bergström, 1991; Lindström et al., 1997; Perrin, 2000]. Par rapport aux modèles fondés sur la description fine du bassin versant et l'utilisation des équations qui régissent les processus hydrologiques, ces modèles nous semblent davantage en adéquation avec le niveau de compréhension du fonctionnement hydrologique des bassins versants et avec les données dont on dispose actuellement. Ils ont notamment toute leur légitimité dans un contexte d'hydrologie appliquée.

Il existe aujourd'hui un très grand nombre de modèles hydrologiques conceptuels ou empiriques. Cette diversité des représentations de la relation pluie-débit reflète d'ailleurs le niveau de connaissance parcellaire du fonctionnement des bassins versants. Cette diversité reflète également la croyance qu'un modèle n'est bon que sur le type de bassins sur lequel il a été développé. Face à cette diversité de

modèles, des comparaisons de structures de modèles ont été effectuées afin d'en évaluer les performances dans un cadre unifié. Ces comparaisons ont notamment permis de dégager des résultats généraux sur la modélisation pluie-débit, d'identifier quelques structures de modèles particulièrement robustes et efficaces et des propositions sur une philosophie de développement des modèles hydrologiques. Néanmoins, ce travail exhaustif de comparaison n'a jamais été effectué auparavant au pas de temps horaire et rien ne nous assure de la validité des conclusions obtenues au pas de temps journalier.

Ainsi, dans le cadre de cette thèse, nous nous proposons de répondre à la question : « **Quels modèles pour le pas de temps horaire ?** » et ce, au travers d'une approche de développement empirique de structures de modèles et au travers d'une approche comparative de différentes structures de modèles, au pas de temps horaire. Nous chercherons également à nous assurer de la robustesse de la méthodologie de comparaison mise en place par [Perrin, 2000], et à identifier la spécificité du pas de temps horaire par rapport au pas de temps journalier en terme de modélisation (niveau de complexité requis, robustesse, complémentarité des modèles).

Chapitre 2

***« Pour trouver une source minérale en Ardèche, pas
besoin de sourcier ou d'ingénieur hydrologue, il suffit de
laisser faire les chèvres. »***

Jean Durand, 1992, *La légende des Sources*

CHAPITRE 2 Présentation de l'échantillon de bassins versants

2.1 Introduction

La modélisation hydrologique pluie-débit a pour objectif de représenter la relation entre pluie et débit établie par le système bassin versant. Les modèles, qui ne sont que des représentations intellectuelles des flux d'eau au sein du bassin versant, doivent être confrontés à des données de systèmes réels pour être construits et validés.

Ayant choisi de suivre une démarche empirique pour construire un modèle pluie-débit et ayant pour objectif de développer et/ou de mettre en évidence la généralité de modèles pluie-débit, nous devons constituer un échantillon de données permettant de confronter les modèles à des bassins versants représentant une grande diversité de conditions hydro-météorologiques, pédologiques, géologiques, d'occupation du sol, etc..

A l'heure actuelle, l'utilisation en hydrologie de grands échantillons de bassins versants est de plus en plus courante aux pas de temps mensuel ou journalier, mais encore très rare à des pas de temps fins (horaire, notamment), pour lesquels les données sont encore difficiles à rassembler. Par exemple, en France, l'installation de nombreuses stations pluviométriques automatiques par Météo France n'a commencé qu'au début des années 1990. Par contre, aux États-unis, où les données sont généralement très disponibles, ce sont les débits au pas de temps horaire qui ne sont pas enregistrés, malgré une large couverture de stations pluviométriques automatiques... Ainsi, notre base de données qui regroupe plus de 300 bassins versants situés en France, aux États-unis, en Australie, en Slovénie et en Espagne est absolument inédite pour ce pas de temps.

2.2 Sources et constitution de la base de données

2.2.1 Sources des données

Les données au pas de temps horaire étant rares, celles-ci proviennent de nombreux producteurs de données ayant des systèmes d'acquisition, de stockage et des unités de mesures très hétérogènes. Le rassemblement de cette base de données

représente ainsi un long travail de collecte, de traitement et de vérification des données de pluie, débit et ETP reçues.

En France, la majeure partie des données hydrométriques provient de la banque HYDRO du MEDD. Nous avons utilisé les fichiers de base d'enregistrement des débits, à pas de temps variable, que nous avons transformé au pas de temps horaire. La majeure partie des données pluviométriques provient des stations automatiques de la banque PLUVIO de Météo France. Les données des autres bassins versants français proviennent des sites expérimentaux de l'Orgeval (géré par le Cemagref d'Antony, situé en Seine et Marne), du Réal Collobrier (géré par le Cemagref d'Aix en Provence, situé dans les Bouches du Rhône), de Naizin-Kervidy (géré par l'INRA de Rennes, situé dans le Morbihan), de la Direction Technique Générale d'EDF (Grenoble) et de la DDE d'Avignon (bassins de l'Eygues et l'Ouvèze, suivis dans le cadre de l'annonce des crues).

Aux États-unis, la majeure partie des données provient de la base de données de l'ARS (Agricultural Research Service). Ces données sont gérées par le WDC (Water Data Center) de l'USDA (United States Department of Agriculture). Les autres données proviennent du bassin versant expérimental de Caspar Creek (Californie), géré par le Forest Service de l'USDA. En Australie, les données proviennent du bassin versant de Collie River. Ce bassin versant est suivi par l'Université de Perth (M. Sivapalan) et le WRC-WA (Water Resources Commission of Western Australia). En Slovénie, les données du bassin versant de Savinja proviennent de l'Agence de l'environnement de Slovénie. En Espagne, les données du bassin versant expérimental de Vallcebre proviennent du CSIC (Consejo Superior de Investigaciones Cientificas).

2.2.2 Critères de sélection des bassins versants

La sélection de chaque bassin versant s'est faite en essayant de répondre aux critères suivants :

1. **Temps de monté du bassin versant inférieur à 24 heures.** N'ayant pas la connaissance a priori du temps de réponse de chaque bassin versant, nous nous sommes limité à des bassins versants de petite superficie, inférieure à 2000 km². Ensuite, nous avons retenu quelques bassins versants de superficies supérieures, mais dont les comportements semblaient plutôt rapide. Nous avons également retenu quelques très petits bassins versants (< 5 km²) pour lesquels le pas de temps horaire est certainement de l'ordre ou légèrement supérieur au temps de réponse du bassin versant ;
2. **Présence de plusieurs pluviomètres pour décrire la pluie tombant sur le bassin.** Les bassins versants sélectionnés possèdent au moins un pluviomètre à l'intérieur du bassin versant et d'autres situés dans ou à proximité du bassin

versant. Bien sûr, plus le bassin versant est grand, plus nous avons essayé d'inclure un nombre élevé de pluviomètres, dans la limite des données disponibles. Autant que possible, nous avons essayé de suivre les recommandations de *Schaake et al.* [2000], utilisées pour les données du projet MOPEX. Dans notre base de données, 55 % des bassins versants suivent ces recommandations ;

Tableau 2-1 : Nombre de pluviomètres nécessaires pour avoir une connaissance suffisante de la pluie, en fonction de la superficie du bassin versant (d'après *Schaake et al.* [2000])

Superficie du bassin (km ²)	50	200	500	1200	2000	3000	6000	
Nombre de stations pluviométriques	2	3	4	5	6	7	8	9

3. **Faible influence de la neige.** En fonction du climat auquel est soumis chaque bassin versant, nous avons limité l'altitude de la station de jaugeage du bassin à 1200 m, afin de nous affranchir d'une trop grande influence de la neige sur le fonctionnement des bassins versants ;
4. **Chroniques de données d'au moins trois années.** Les chroniques horaires continues de pluie et de débit étant rares, nous avons accepté les bassins versants avec des chroniques de faible durée : au minimum trois ans (une année de mise en route, une année de calage et une année de contrôle du modèle).

2.2.3 Constitution de la base de données

Au début de cette thèse, peu de jeux de données de bassins versants étaient disponibles (moins de 10 bassins versants). A l'aide d'un système d'information géographique, nous avons constitué les jeux de données des bassins versants français. Pour cela, nous disposons des cartes des stations de jaugeage de la banque HYDRO (Figure 2-1 a), des stations pluviométriques automatiques (horaire) de Météo France (Figure 2-1 b), du réseau hydrographique et des limites de bassins versants élémentaires.

Dans un premier temps, nous avons retenu toutes les stations limnimétriques jaugeées répondant aux critères de sélection des bassins versants (altitude de la station limnimétrique et superficie du bassin versant). Ensuite, pour chacune des 1700 stations retenues, nous avons regardé quelles étaient les stations pluviométriques situées sur et autour du bassin versant. Chaque fois que l'information de pluie a été jugée suffisante, le bassin versant a été sélectionné. Pour les 300 bassins versants retenus initialement, les données de pluie ont été commandées et les jeux de données de bassins versants ont été constitués. Ensuite, à l'aide des données interannuelles d'ETP de la thèse de *Oudin* [2004], les triplets

pluie-ETP-débit ont été constitués. En fonction des données réellement disponibles et des lacunes de pluie, nous avons retenu au total 227 bassins versants français. Pour les bassins versants des autres pays, une procédure similaire de constitution des jeux de données a été adoptée, ou les jeux de données déjà constitués ont été fournis directement par les gestionnaires.

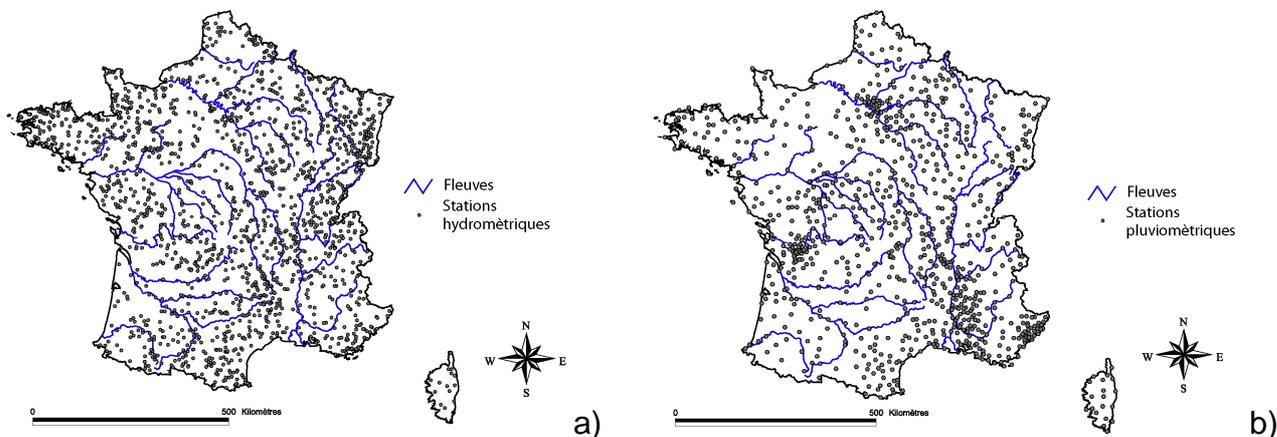


Figure 2-1 : a) stations hydrométriques de la banque HYDRO, b) postes pluviométriques horaires de Météo France, répondant aux critères de sélection des bassins versants.

2.2.4 Quelles pluie et ETP de bassin versant ?

Lorsque plusieurs postes pluviométriques et d'ETP étaient disponibles sur un bassin versant, la pluie et l'ETP de bassin (celles qui sont fournies en entrée du modèle) ont été calculées comme une simple moyenne arithmétique équipondérée des postes disponibles.

2.3 Description des 313 bassins versants

2.3.1 Bassins versants situés en France

Dans cet échantillon, nous disposons de 227 bassins versants, répartis sur le territoire de France métropolitaine (Figure 2-2). Pour la plupart, ces données n'ont jamais été utilisées dans des travaux de recherche en hydrologie, mises à part les données des bassins versants expérimentaux de l'Orgeval, du Réal Collobrier, de Naizin.

Ces bassins versants sont principalement situés sur les grands bassins français Loire-Bretagne, Seine-Normandie et Rhône-Méditerranée-Corse. Les grands bassins Adour-Garonne, Rhin-Meuse et Artois-Picardie sont assez peu représentés à cause d'une faible densité du réseau pluviométrique. La répartition des bassins versants par grand bassin français est la suivante :

- 7 stations sur le bassin Rhin-Meuse,
- 8 stations sur le bassin Artois-Picardie

- 60 stations sur le bassin Seine-Normandie,
- 82 stations sur le bassin Loire-Bretagne,
- 11 stations sur le bassin Adour-Garonne,
- 59 stations sur le bassin Rhône-Méditerranée-Corse.

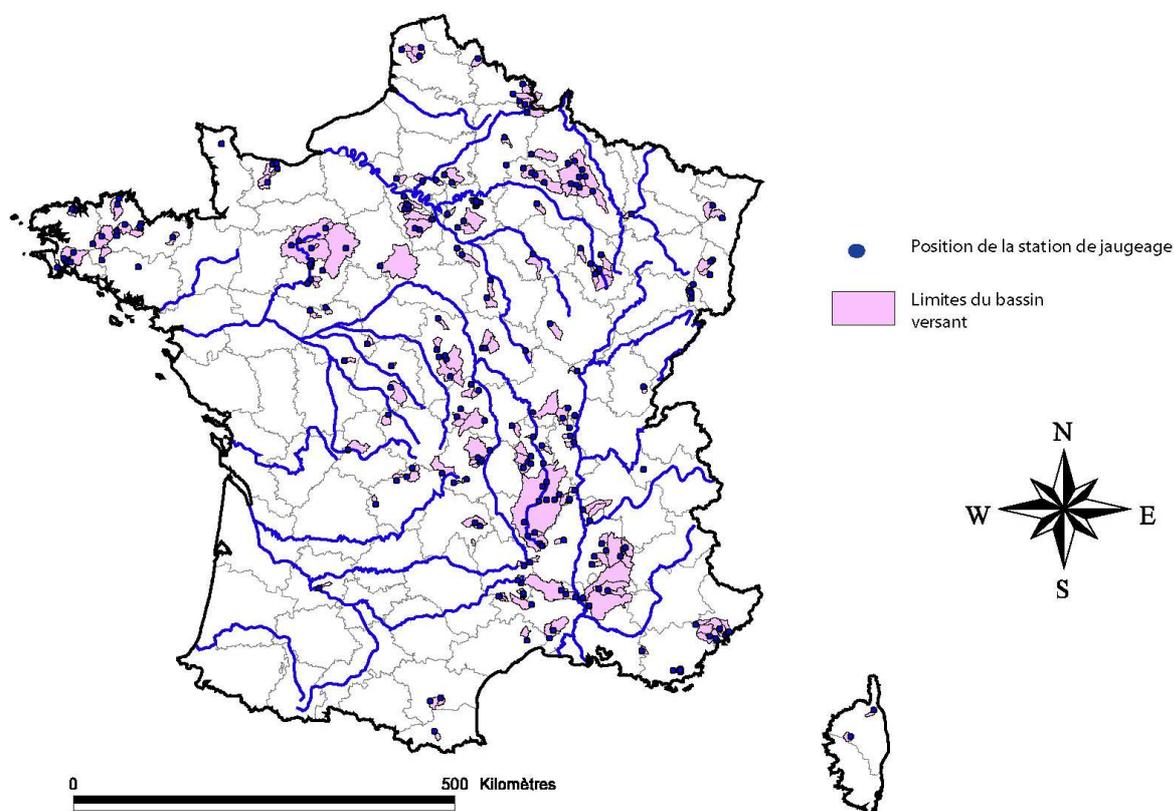


Figure 2-2 : Localisations des 227 bassins versants français

Pour ces 227 bassins versants, nous avons utilisé l'information de 370 postes pluviométriques. Ce nombre de pluviomètres peut paraître assez faible, par rapport aux données disponibles à des pas de temps supérieurs. Néanmoins, de nombreux postes ont été utilisés plusieurs fois dans des bassins versants adjacents (petits bassins versants amont) ou emboîtés.

2.3.2 Bassins versants situés aux États-unis

Nous disposons de 70 bassins versants américains (Figure 2-3). Il s'agit de bassins versants expérimentaux, généralement de très petite surface. Les bassins versants forestiers de Caspar Creek (CA) ont subi des modifications rapides de leur couvert forestier (coupe). Les périodes d'avant et après coupe ont été séparées, comme si les bassins versants étaient différents, afin de prendre en compte la forte non stationnarité des conditions de surface des bassins versants.

Pour ces 70 bassins versants, nous disposons de 311 postes pluviométriques. Certains des bassins américains possèdent une information de pluie extrêmement

détaillée, notamment dans le cas des bassins versants arides de Santa Rosa (NM) et Walnut Gulch (AZ). Les bassins versants américains sont soumis à un climat aride (Safford, Walnut Gulch, Santa Rosa, Sonora), sub-tropical (Riesel, Tifton), océanique (Caspar Creek) et tempéré humide (Oxford, Goodwin Creek, Blacksburg). Les données d'ETP proviennent de la thèse de *Oudin* [2004]. Elles dérivent de données climatiques obtenues sur le CD-ROM Solar and Meteorological Surface Observation Network [NOAA, 1993], qui fournit des données atmosphériques journalières pour 237 stations réparties aux États-unis. Certaines données de cet échantillon ont été utilisées dans de nombreux travaux de recherche.

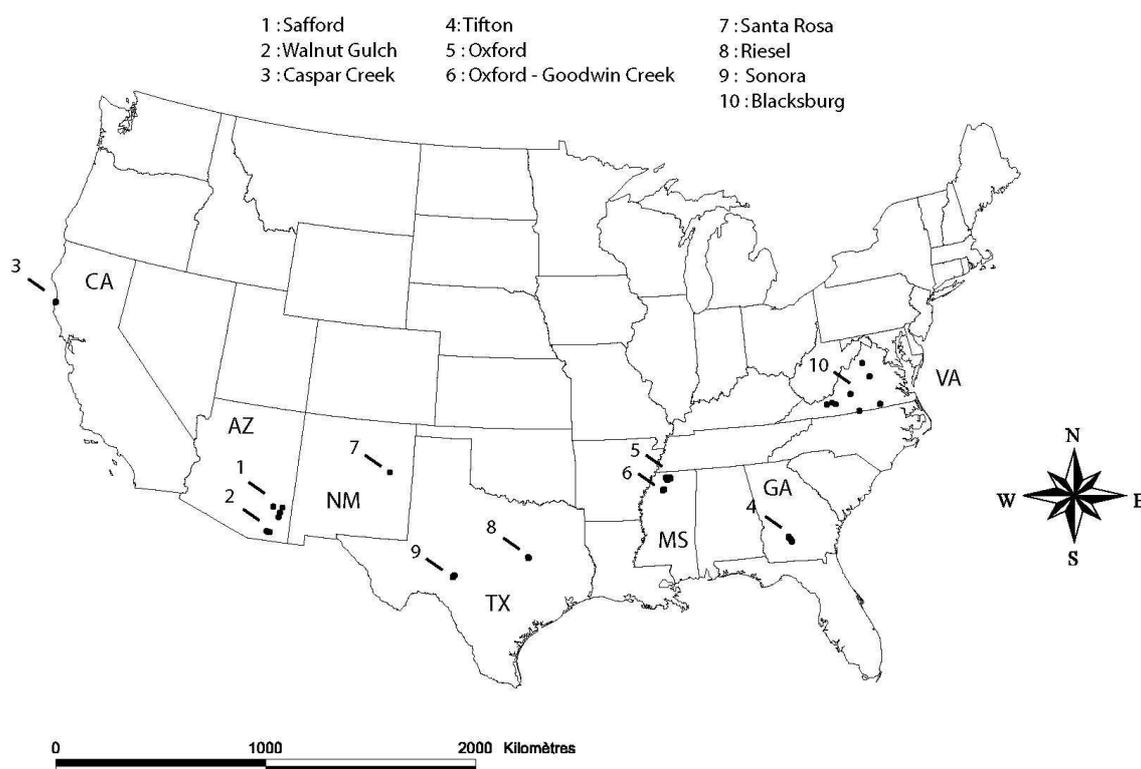


Figure 2-3 : Localisation des 70 bassins versants expérimentaux américains

2.3.3 Bassins versants situés en Australie

Nous disposons des données de 12 bassins versants australiens (Figure 2-4) et de 16 pluviomètres. Ces données concernent le bassin versant de Collie River, situé au sud-ouest de l'Australie, près de la ville de Perth. Le bassin versant de Collie River est caractérisé par un climat méditerranéen, une végétation dominée par l'eucalyptus, des sols latéritiques et des vallées particulièrement encaissées. Ce bassin versant est suivi depuis les années 1950, notamment à cause d'une forte période de déforestation (jusque dans les années 1980) et une salinisation progressive des eaux souterraines. Ces données ont été utilisées dans de nombreux travaux [Jothityangkoon et Sivapalan, 2001; Jothityangkoon et al., 2001].

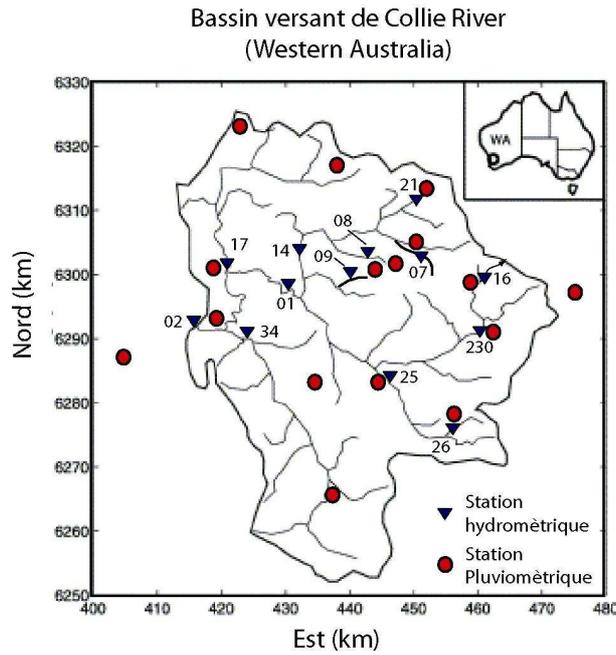


Figure 2-4 : Localisation des 12 bassins versant australiens

2.3.4 Bassins versants situés en Slovénie

Nous disposons des données de deux bassins versants slovènes (Figure 2-5) et de 5 pluviomètres. Ces données concernent le bassin de la rivière Savinja. Ce bassin est caractérisé par un climat de montagne, avec de fortes crues. La contribution de la neige sur ce bassin n'est certainement pas négligeable, néanmoins de bons niveaux de performance des modèles sont obtenus sans la prendre en compte. Les performances obtenues sont d'ailleurs comparables à celles obtenues avec le modèle HBV [Kobold et al., 2003; Kobold et Brilly, 2004], qui prend en compte la neige et le gradient altimétrique des précipitations.

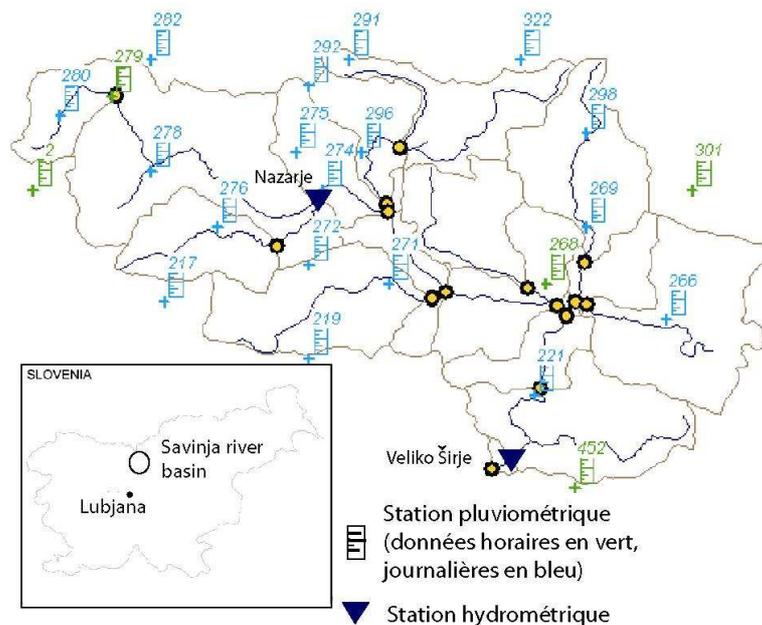


Figure 2-5 : Localisation des deux bassins versants slovènes

2.3.5 Bassins versants situés en Espagne (Catalogne)

Nous disposons des données de deux bassins versants expérimentaux et d'un pluviomètre, situés en Catalogne espagnole. Ces données concernent le bassin de Vallcebre, aux stations de jaugeage de Cal Rodo et Can Vila. Ce bassin est caractérisé par un climat méditerranéen de montagne, mais avec assez peu de neige à cette altitude. Le substrat géologique est composé de marnes et de calcaires. Des pertes karstiques sont constatées au niveau des calcaires. Ce bassin versant expérimental a fait l'objet de nombreuses études et publications [Gallart et al., 1994; Gallart et al., 1997; Anderton et al., 2002; Gallart et al., 2002].

The Vallcebre research catchments

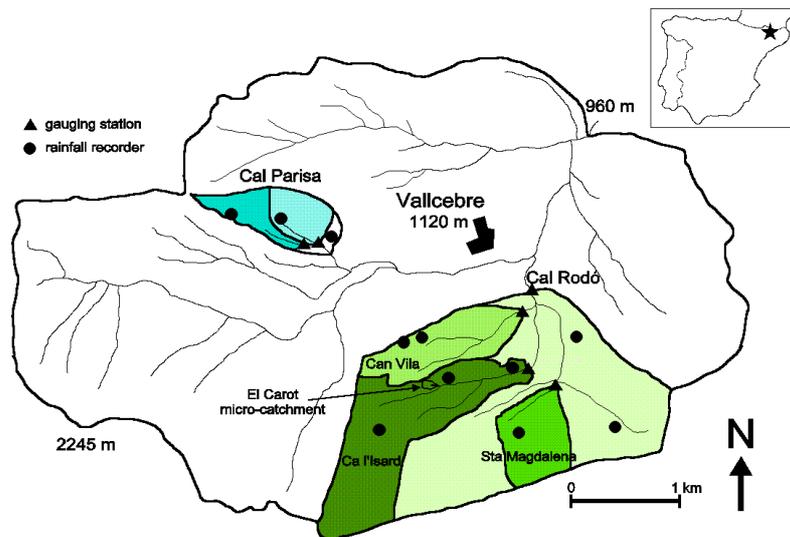


Figure 2-6 : Localisation des deux bassins versants espagnols (catalans).

2.4 Caractéristiques des bassins versants

Cet échantillon de bassins versants représente une très grande quantité d'information et une grande diversité de situations hydro-climatiques. Ces données, environ 23 millions de triplets pluie-ETP-débit (2600 années), doivent nous permettre de tester les modèles dans une large gamme de situations hydrologiques. Ne pouvant décrire chaque bassin versant dans les détails, nous présentons ici une vue d'ensemble des caractéristiques de cet échantillon (Figure 2-8 à Figure 2-13).

2.4.1 Caractéristiques physiques

La superficie des bassins versants varie entre 0.6 et 5000 km², avec une superficie médiane de 85 km² (Figure 2-8 a). 10 % des bassins versants ont une superficie inférieure à 3.6 km² et 10 % une superficie supérieure à 525 km². Par rapport aux échantillons utilisés au Cemagref au pas de temps journalier [Perrin, 2000; Oudin, 2004; Rojas-Serna, 2005], cet échantillon regroupe des bassins versants aux

superficiés plus petites, qui justifient l'usage du pas de temps horaire pour modéliser leur comportement. Les bassins versants les plus petits sont généralement des bassins versants expérimentaux.

Dans l'ensemble, les bassins versants sont situés à une faible altitude : 90 % des bassins versants ont leur station de jaugeage à une altitude inférieure à 625 m (Figure 2-8 c). La médiane de l'altitude moyenne des bassins versants français est seulement de 275 m et 80 % de ces bassins ont leur altitude moyenne entre 110 et 1000 m (Figure 2-8 b). Les altitudes maximales de ces bassins sont assez peu élevées et seulement 10 % des bassins versants ont une altitude maximale supérieure à 1560 m (Figure 2-8 d). Cela ne garantit pas l'absence de neige sur ces bassins, mais il est raisonnable de penser que peu de bassins versants sont très affectés par la neige, et qu'elle n'a qu'une influence limitée sur leur fonctionnement. A l'aide de traitements en ondelettes [Torrence et Compo, 1998; Gauchere, 2002; Mathevet et al., 2004], pour les bassins versants situés dans les zones susceptibles d'imposer une influence nivale, des composantes diurnes (24 h) ont parfois été trouvées (moins de 5% de l'échantillon de bassins versants). Un exemple d'analyse pour un bassin où peut être détecté cette influence est présenté à la Figure 2-7.

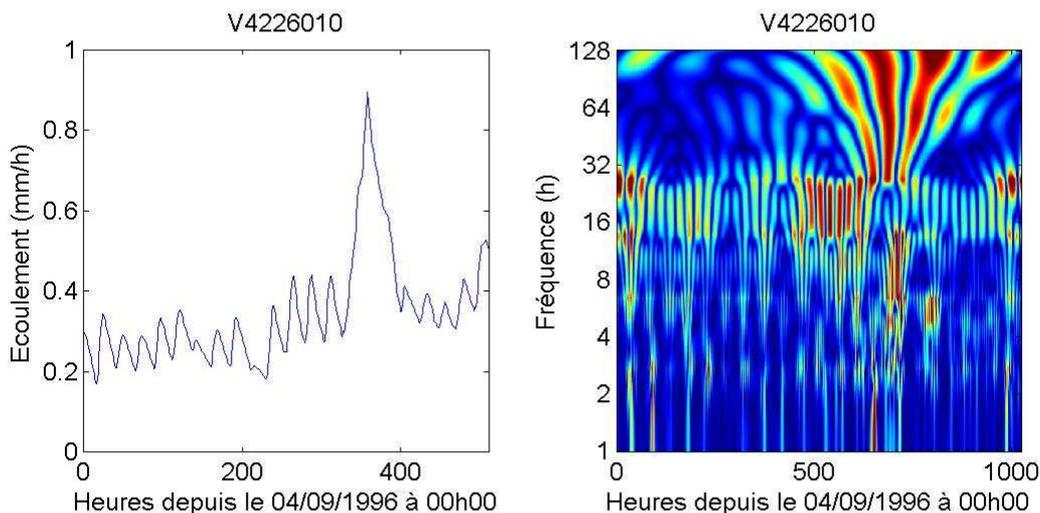


Figure 2-7 : Episode de fonte nivale sur le bassin de l'Archiane à Menée. La présence de période de 24 heures est représentée par les couleurs vert à rouge entre les fréquences 16h et 32h (axe y).

A l'aide de la base de données CORINE Land Cover (CLC), les caractéristiques de l'occupation des sols des 227 bassins versants français ont été extraites. Cette base de données, diffusée par l'IFEN (Institut Français de l'Environnement), permet d'estimer pour chaque bassin versant, les surfaces occupées par différents types de couverts. Chaque type de couvert est ensuite divisé en différents niveaux de précision (de 1 à 3).

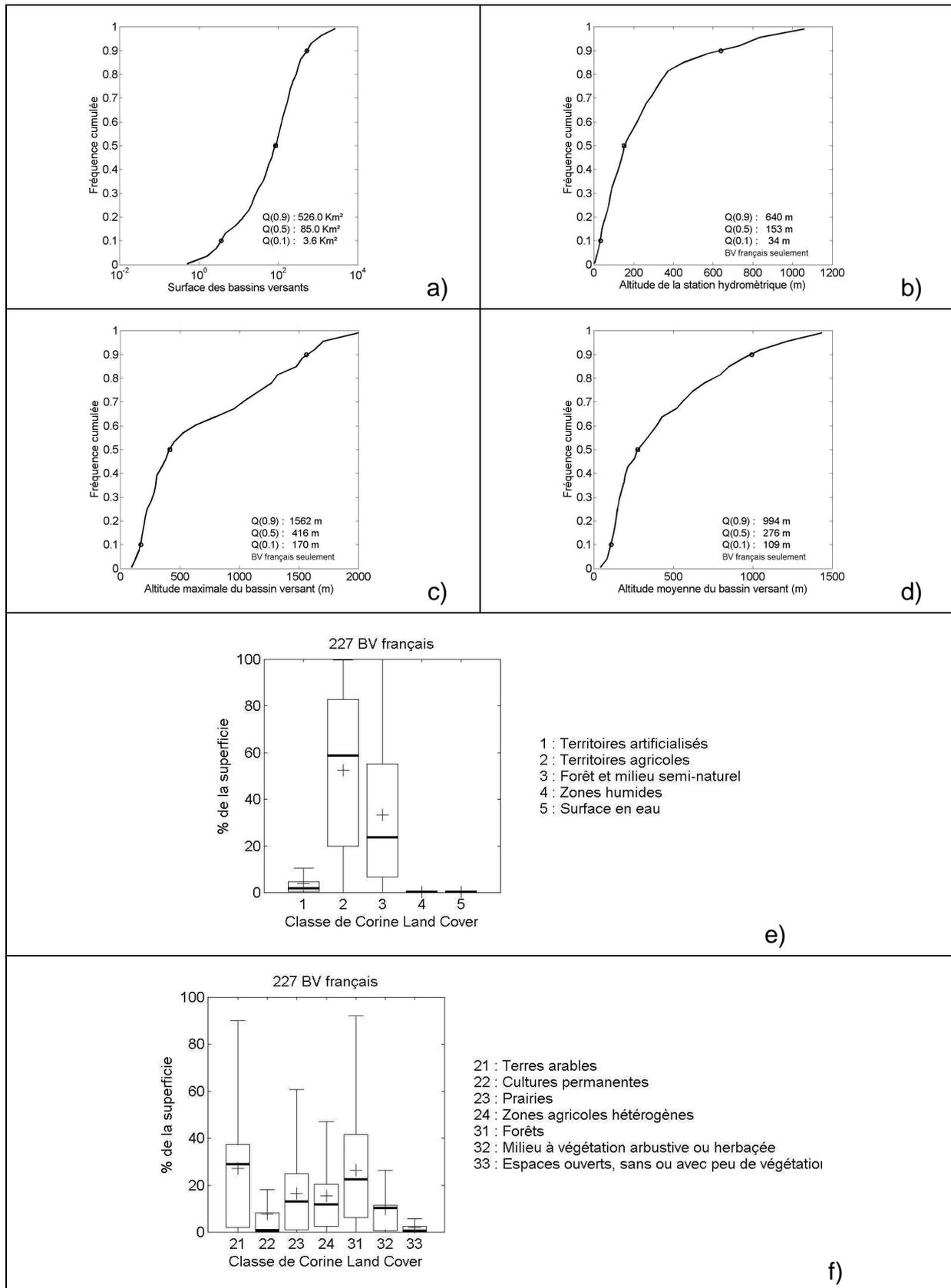


Figure 2-8 : Caractéristiques physiques de l'échantillon de 313 bassins versants : (a) distribution des superficies, (b) distribution de l'altitude de la station hydrométrique, (c) distribution de l'altitude maximale du bassin versant, (d) distribution de l'altitude moyenne du bassin versant, (e) occupation des sols (classe 1), (f) occupation des sols (classe 2).

Les classes de niveau 1 correspondent aux grandes catégories d'occupation des sols repérables à l'échelle de la planète, les classes de niveau 2 correspondent à des échelles de 1/500 000 à 1/1 000 000 et les classes de niveau 3 correspondent à des échelles de 1/100 000. Dans notre analyse, nous nous contentons des classes de niveau 1 et de niveau 2. Les Figure 2-8 e) et f) présentent les pourcentages de superficie occupée par les types de couverts de niveau 1 et 2, pour l'ensemble des 227 bassins versants français. Ces superficies ont été extraites à l'aide d'un système d'information géographique en superposant les contours des bassins versants et la base de données CORINE Land Cover.

Les différentes classes utilisées sont les suivantes :

1. Territoires artificialisés,
2. Territoires agricoles,
 - 2.1. Terres arables,
 - 2.2. Cultures permanentes,
 - 2.3. Prairies,
 - 2.4. Zones agricoles hétérogènes,
3. Forêts et milieux semi naturels,
 - 3.1. Forêts,
 - 3.2. Milieux à végétation arbustive ou herbacée,
 - 3.3. Espace ouvert, sans ou avec peu de végétation,
4. Zones humides,
5. Surfaces en eau.

La Figure 2-8 e) montre que la surface des 227 bassins versants français est principalement recouverte par des territoires agricoles et des forêts. Ces deux couverts représentent en moyenne plus de 95 % de la superficie des bassins versants, avec respectivement 59.5% et 38%. Les autres types de couverts sont marginaux, les territoires artificialisés représentent en moyenne 4.6% de la superficie des bassins versant et les zones humides et surface en eau sont inférieures à 0.4%. Ainsi, on peut conclure que les bassins versants sont principalement situés en zone naturelle ou semi-naturelle. Seuls 10 % des bassins versants ont une part élevée de territoires artificialisés (entre 10% et 38%). La Figure 2-8 f) donne plus de détails sur les types de couverts prépondérants. Les bassins versants sont principalement recouverts par des terres arables, des prairies, des zones agricoles hétérogènes et des forêts, qui représentent en moyenne respectivement 31%, 19%, 18% et 30% de leurs superficies.

2.4.2 Caractéristiques hydro-climatiques générales

D'un point de vue climatique, cet échantillon possède une très grande variété de conditions : arides, tempérées, méditerranéennes, sub-tropicales, océaniques et montagnardes. L'écoulement moyen annuel varie entre $0.001 \text{ mm.an}^{-1}.\text{km}^{-2}$ et $1650 \text{ mm.an}^{-1}.\text{km}^{-2}$ avec une valeur médiane de $390 \text{ mm.an}^{-1}.\text{km}^{-2}$, la pluie moyenne annuelle varie entre 200 et 3000 mm.an^{-1} , avec une valeur médiane de 900 mm.an^{-1} et l'ETP moyenne annuelle varie entre 600 et 2100 mm.an^{-1} , avec une valeur médiane de 770 mm.an^{-1} . Le rendement des bassins (rapport entre la lame d'eau écoulée et la lame d'eau précipitée) est entre 0.4 et 235 %. 80 % des bassins versants ont un rendement compris entre 7 et 65 %. Certains bassins (1.6 % de l'échantillon) possèdent un rendement supérieur à 100 %, ce qui indique que le volume écoulé est supérieur à la pluie reçue par le bassin. Cette situation est bien sûr problématique et soulève l'épineuse question de la qualité des données. Pour ces bassins, cette valeur s'explique principalement par une sous-estimation de la pluie et pour un bassin en particulier, une forte sous-estimation de la pluie et une surestimation des débits. Remarquons cependant que, au-delà des problèmes de qualité ou de représentativité des données, il existe de nombreux bassins pour lesquels il est difficile de boucler le bilan en eau sans prendre en compte des échanges souterrains. C'est ce que montre O'Connor et al. [2005] sur des bassins versants dits 'non conservatifs'.

La principale différence entre les bassins versants français et les bassins versants étrangers de cet échantillon est le caractère beaucoup plus aride de l'échantillon des bassins versants étrangers (USA, Australie et Espagne). Globalement, pour des pluies annuelles supérieures ou égales, leurs ETP sont beaucoup plus fortes et leurs écoulements plus faibles. Ceci est mis en évidence par la grande différence entre les distributions du coefficient de rendement (Figure 2-9 d) et de l'indice d'aridité (Figure 2-9 e) des bassins versants français et étrangers. Une grande partie (65%) des bassins versants de l'échantillon possède un écoulement continu (Figure 2-11 b). Certains des bassins versants, principalement ceux situés dans les zones méditerranéennes ou arides, ont un écoulement intermittent.

Seulement 9 % des bassins français sont intermittents, alors que 80 % des bassins versants étrangers le sont. Le caractère intermittent de ces bassins est d'ailleurs beaucoup plus marqué, avec des probabilités beaucoup plus faibles d'avoir des débits non nuls.

La Figure 2-8 et la Figure 2-9 illustrent la diversité des caractéristiques hydro-climatique des bassins versants de cet échantillon.

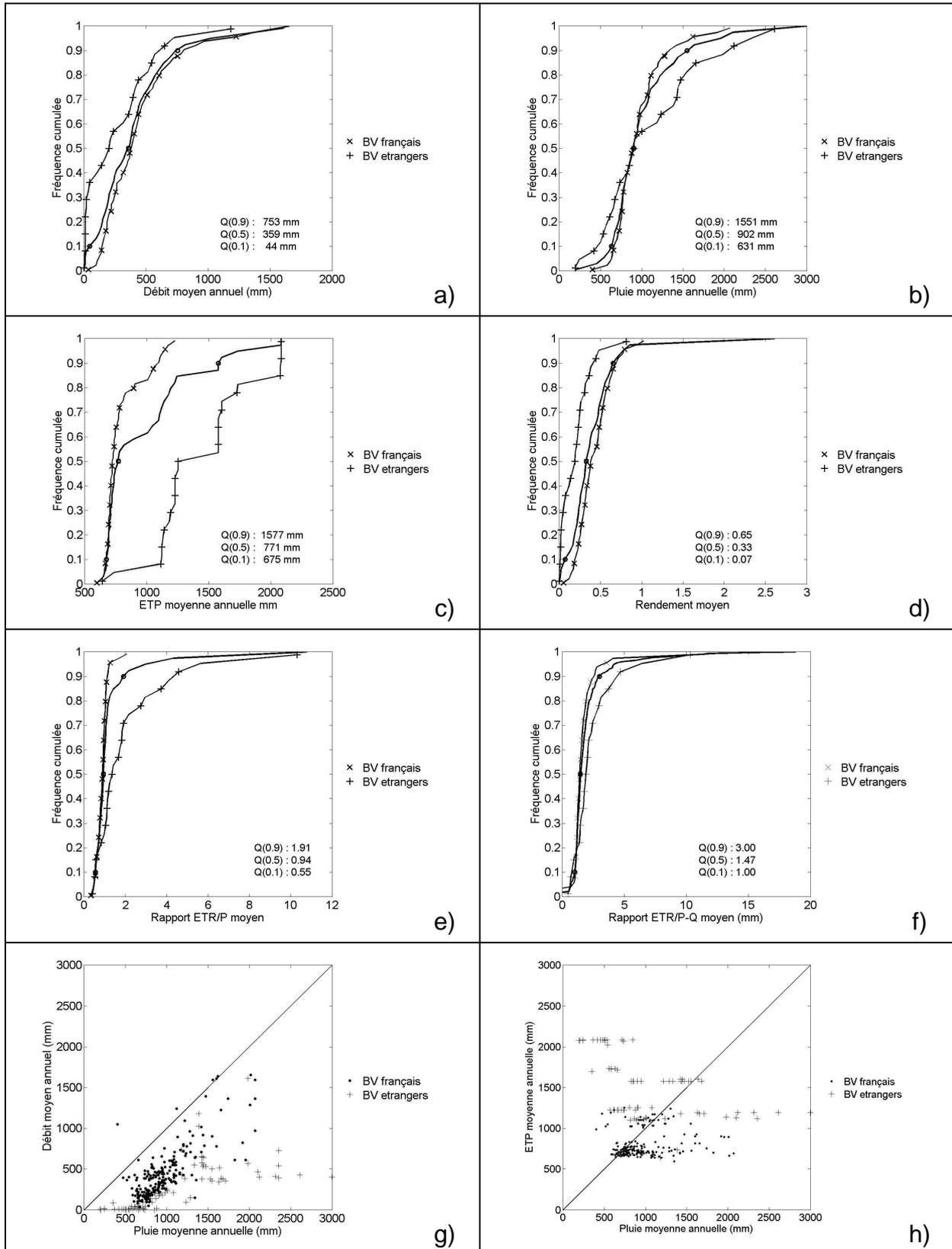


Figure 2-9 : Caractéristiques hydro-climatiques de l'échantillon de 313 bassins versants : (a) distribution des débits, (b) distribution des pluies, (c) distribution des ETP, (d) distribution des coefficients de rendement, (e) distribution des indices d'aridité ETP/P, (f) distribution des déficits d'écoulement ETR/(P-Q), (g) pluie et débit moyen des bassins versants, (h) pluie et ETP moyenne des bassins versants.

Cette diversité, ainsi que celles des caractéristiques physiques (couverts végétaux, caractéristiques géologiques, pédologiques, topographiques), nous permet d'offrir aux modèles des conditions hydrologiques aussi contrastées que possible, afin de tester la généralité et la robustesse de ces modèles.

2.4.3 Quelques caractéristiques de la pluie et des débits au pas de temps horaire et journalier

L'objet de cette thèse est la modélisation hydrologique des bassins versants dont le temps de réponse est rapide, généralement inférieur à 24 h. Pour cela, on utilise des données au pas de temps horaire. Pour illustrer la différence qu'il y a entre les pas de temps horaire et journalier, nous allons nous intéresser à quelques caractéristiques statistiques de la pluie et des débits de notre échantillon à ces deux pas de temps.

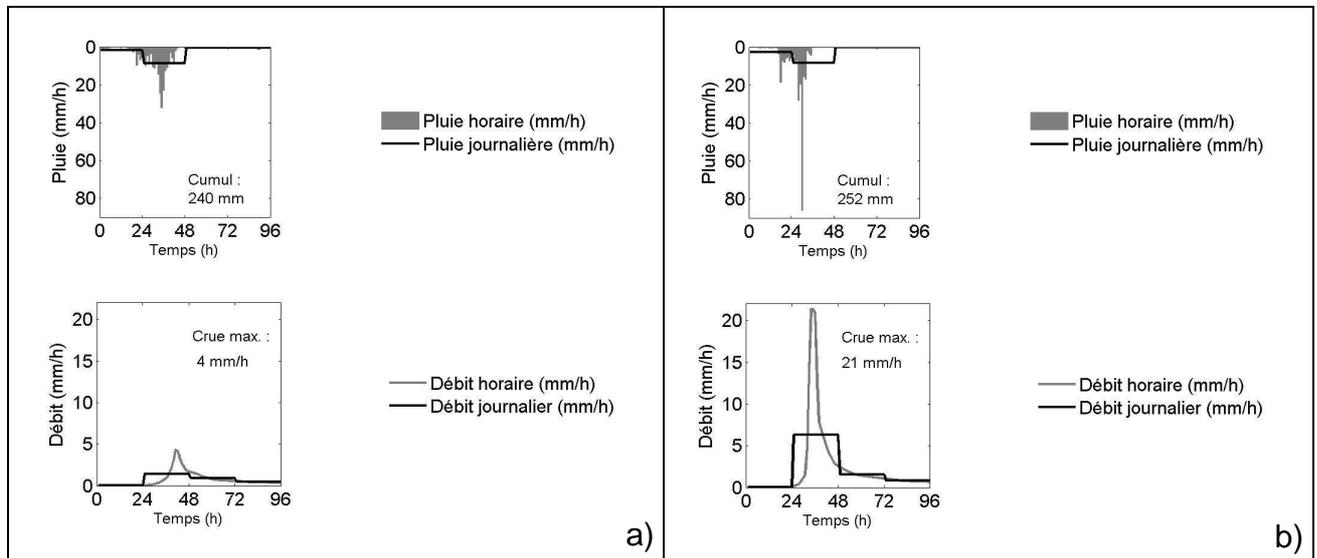


Figure 2-10 : Exemple de deux événements cévenols dont le cumul de pluie journalier est du même ordre de grandeur (La Goudèche à Saint-Maurice-de-Ventalon, 10 km²). a) crue du 18/09/1996 et b) crue du 21/09/1992

Comme l'illustre la Figure 2-10, le principal problème lié au pas de temps journalier est qu'il filtre la dynamique des événements de pluie et la réponse des bassins versants. Les moyennes journalières (traits noirs) représentent très mal le déroulement des événements à un pas de temps plus fin (traits gris). Les bassins versants de cet échantillon ont des durées caractéristiques des pluies et des crues qui sont de quelques heures et pour des événements de pluie usuels et intenses, le cumul journalier de la pluie ne permet pas de qualifier le type d'événement : 10 mm/h pendant 24 h et 240 mm en quelques heures correspondent à des événements très différents, qui donnent lieu à des réponses hydrologiques également différentes (Figure 2-10). Le pas de temps horaire permet ainsi de mieux connaître la dynamique des événements pluvieux, de mieux évaluer les intensités reçues par le

bassin versant et de suivre la dynamique des débits des bassins versants dont les temps de réponse sont inférieurs à un jour.

Au pas de temps journalier, la médiane de la probabilité d'occurrence de la pluie (>1 mm) sur les bassins versants de cet échantillon est de 32% (Figure 2-11 a). Pour 90% des bassins versants, cette probabilité est entre 18 et 40%. Au pas de temps horaire (Figure 2-11 a), la probabilité d'occurrence de la pluie est très faible. La valeur médiane est seulement de 4%, et pour 90 % des bassins versants, cette probabilité est entre 2 et 5%. Par contre, le coefficient de variation de la pluie est bien plus fort au pas de temps horaire qu'au pas de temps journalier (de l'ordre de 3 à 5 fois), ce qui indique le caractère beaucoup plus intense et chaotique de la pluie quand le cumul est fait à un pas de temps fin (Figure 2-11 c). Concernant les débits, les différences entre les probabilités d'occurrence et les coefficients de variation aux pas de temps horaire et journalier sont assez faibles. En effet, les événements de crues sont assez rares et le débit est une variable fortement auto-corrélée. De plus, au pas de temps horaire, le poids de la variabilité apportée par les événements de crues est écrasé par celui des longues récessions.

Pour rendre compte des événements extrêmes (pluie horaire et journalière maximale annuelle, débit horaire et journalier maximal annuel), les études fréquentielles des chroniques hydrologiques nous permettent de calculer les valeurs des événements associés à certaines périodes de retour. Pour être significatives, ces statistiques doivent s'appuyer sur des chroniques suffisamment longues pour pouvoir extrapoler les valeurs vers des périodes de retours élevées. Malheureusement, la plus grande partie des chroniques de notre base de données ont des durées trop courtes pour pouvoir calculer sans trop d'incertitudes les valeurs des extrêmes associés à des périodes de retour de 5 ans, 10 ans ou 50 ans (valeurs servant généralement de référence). Pour pallier ce problème, les chroniques ont été traitées de la manière suivante : pour chaque bassin versant, nous avons calculé la distribution des valeurs de pluie non nulles et de débit non nuls. Ensuite, pour synthétiser cette information, nous présentons la distribution des quantiles 25, 50, 75 et 100 de chacune des distributions (Figure 2-11 e et f).

Ainsi, l'intérêt du pas de temps horaire est de pouvoir connaître les intensité des fortes pluies et d'approcher la valeur des pointes de crues, non biaisées par les cumuls ou les moyennes journalières. La Figure 2-12 indique les caractéristiques des pluies et débits maximaux annuels, aux pas de temps horaire et journalier. Pour mettre en évidence les différences d'intensité entre le pas de temps horaire et journalier, nous avons comparé les valeurs maximales horaires ('instantanées') de la pluie ou du débit, aux moyennes horaires issues du cumul journalier des pluies ou du

débit moyen journalier, le jour de ce maximum. Ainsi, la Figure 2-12 a) et b) indique la distribution des maxima horaire et journalier de la pluie et des débits

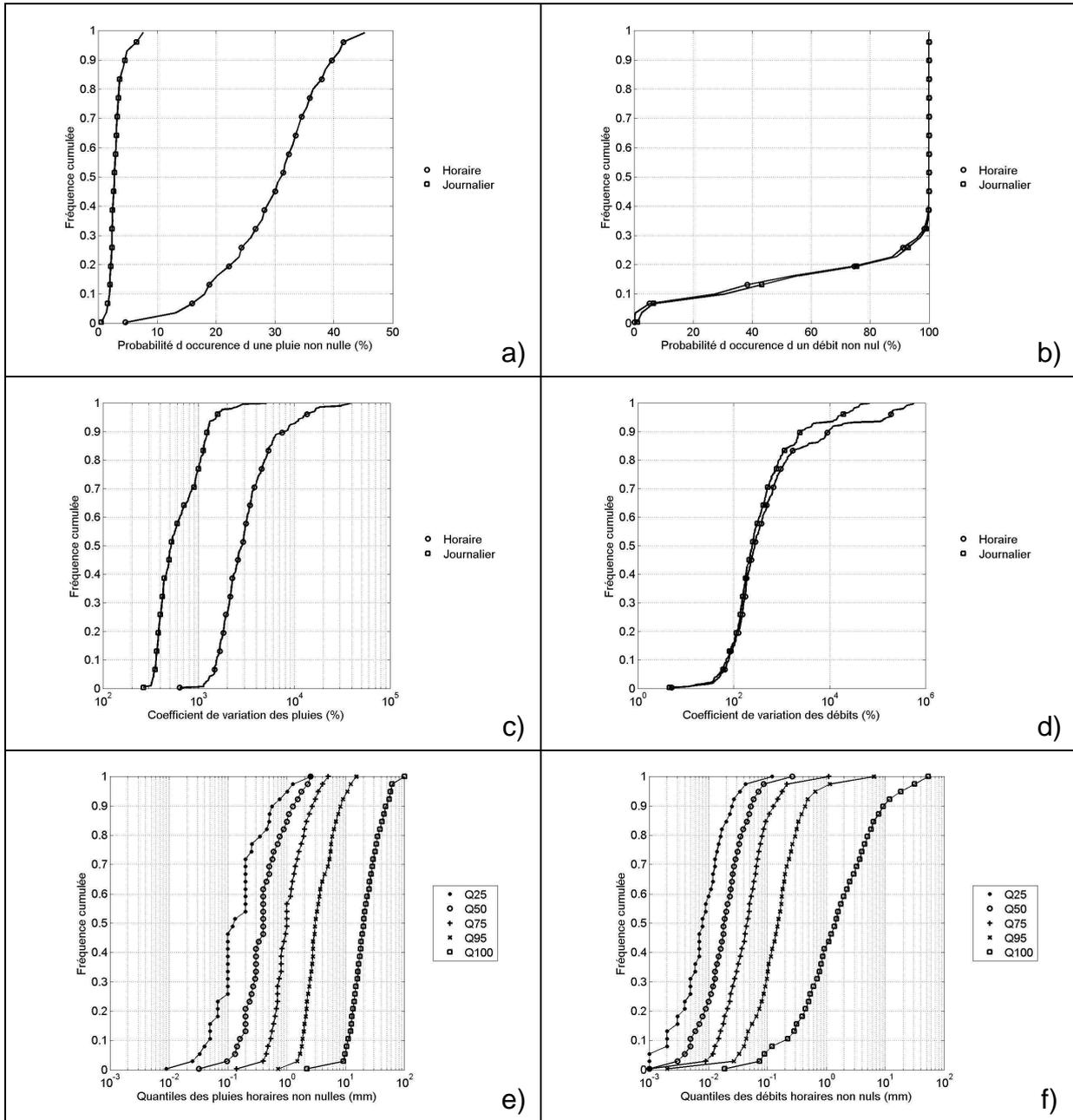


Figure 2-11: Distribution des caractéristiques des pluies et débits aux pas de temps horaire et journalier sur l'échantillon des 313 bassins versants, (a) probabilité d'occurrence d'une pluie non nulle, (b) probabilité d'occurrence d'un débit non nul, (c) coefficient de variation des pluies, (d) coefficient de variation des débits, (e) distribution des quantiles des pluies horaires non nulles, (f) distribution des quantiles des débits horaires non nuls .

Concernant la pluie (Figure 2-12 a), les distributions ont la même allure mais sont décalées d'un ordre de grandeur. La valeur médiane de la pluie horaire maximale annuelle est de 20 mm/h et 80 % des pluies maximales horaires sont comprises entre 9 et 40 mm/h. Par contre, au pas de temps journalier, ces intensités sont beaucoup plus faibles. La valeur médiane de la pluie journalière maximale annuelle

est de 2 mm/h (un cumul journalier de 48 mm) et 80 % des pluies maximales journalières sont comprises entre 1 et 4 mm/h (un cumul journalier de 24 à 96 mm).

Tableau 2-2 : Valeur des écoulements moyens annuels et décennaux pour trois grande rivières françaises : La Seine à Paris, la Loire à Orléans, le Rhône à Lyon (d'après Andréassian [2005]) et comparaison avec l'échantillon de 313 bassins versants (: médiane de l'écoulement maximal annuel).*

Cours d'eau	Écoulement annuel moyen (mm/an)	Écoulement annuel moyen (mm/h)	Écoulement annuel décennal (mm/h)
Seine à Paris (43 800km ²)	210	0.023	0.14
Loire à Orléans (37 000 km ²)	320	0.035	0.28
Rhône à Lyon (20 000 km ²)	930	0.10	0.53
Échantillon de 313 bassins versants	390	0.045	1.0 *

Concernant les débits (Figure 2-12 b), le contraste est beaucoup plus limité. Les distributions ont presque la même allure, mais ne sont décalées que de 2 à 3 unités. La valeur médiane de l'écoulement horaire maximal annuel est de 1 mm/h et 80 % des écoulement horaires maximaux sont compris entre 0.15 et 7 mm/h. Au pas de temps journalier, les différences sont moindres. La valeur médiane de l'écoulement horaire maximal annuel est de 0.4 mm/h et 80 % des écoulements horaires maximaux sont compris entre 0.07 et 2 mm/h. Par rapport aux valeurs annuelles et décennales des écoulements de grands fleuves français, les bassins versants de cet échantillon ont des écoulements maximaux annuels plus forts (Tableau 2-2).

Ces différences de comportement entre la pluie et les débits, sont également illustrées par les Figure 2-12 c) et d) où sont représentées les intensités maximales annuelles de la pluie et des débits aux pas de temps horaire et journalier. Enfin, la Figure 2-12 e) montre la distribution des rapports entre les intensités horaires et journalières des pluies et débits maximaux annuels. Pour la pluie, la valeur de ce rapport est assez uniformément distribuée et la valeur médiane est de 12. Par contre, pour les débits, la valeur de ce rapport est beaucoup plus asymétrique. La valeur médiane est de 2 et 80 % des bassins versants ont un rapport compris entre 1.2 et 10. Néanmoins, même si ce rapport peut sembler faible, il ne faut pas oublier que d'un point de vue local (à l'échelle d'une section de cours d'eau), la sous-estimation des débits d'un facteur 2 ou 3 peut avoir des conséquences importantes au cours des crues, notamment en ce qui concerne les hauteurs d'eau atteintes et les risques d'inondations.

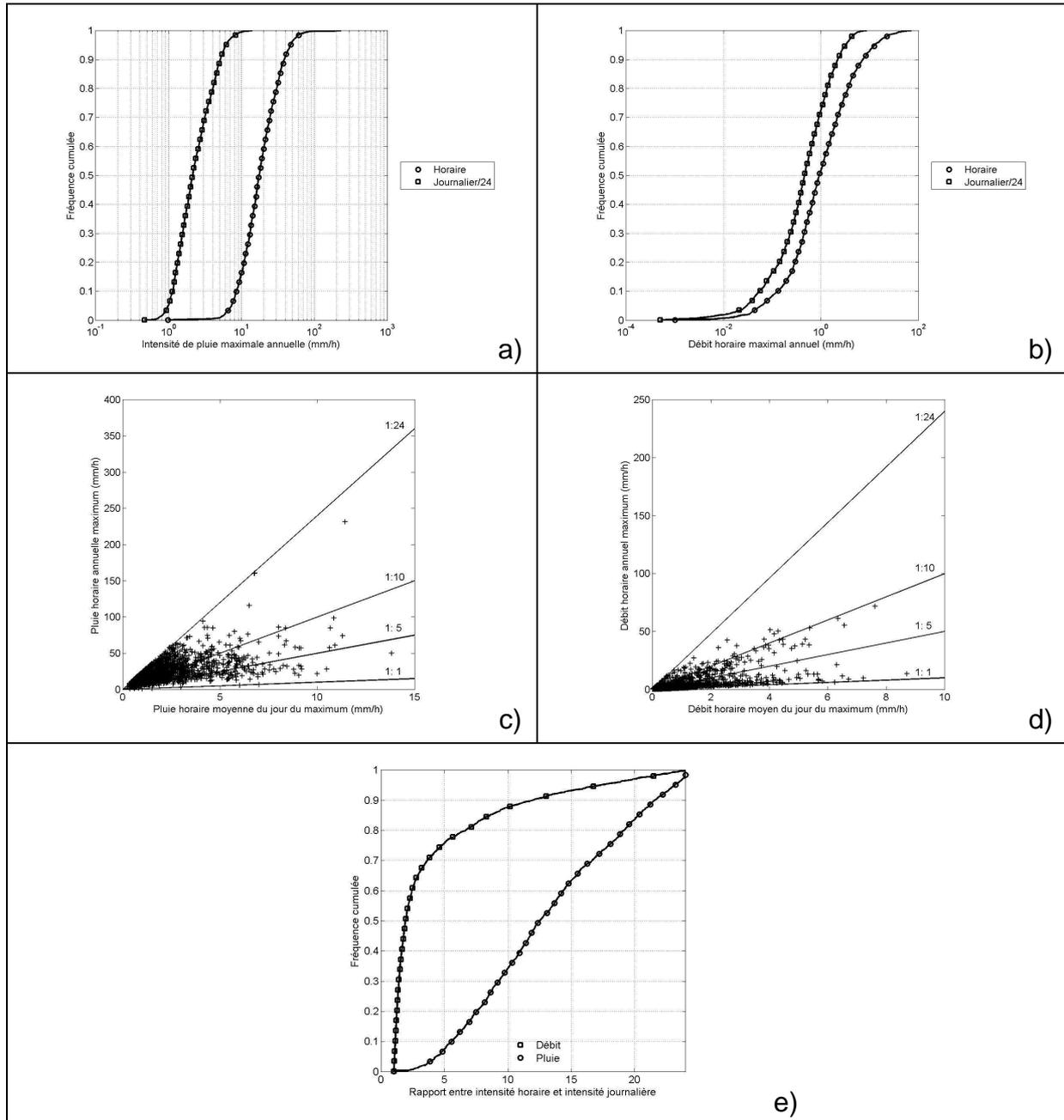


Figure 2-12 : Caractéristiques des pluies et débits horaires maximaux annuels, (a) distribution de l'intensité de la pluie maximale (pluie horaire 'vraie' et pluie horaire issue de la pluie moyenne journalière), (b) distribution du débit horaire maximal annuel (débit horaire 'vrai' et débit horaire issu du débit moyen journalier), (c) comparaison de l'intensité horaire maximale annuelle et de la pluie horaire moyenne issue du cumul du jour de ce maximum, (d) comparaison du débit horaire maximal annuel et du débit horaire moyen issu du débit moyen du jour de ce maximum, (e) distribution du coefficient d'abattement entre la valeur horaire et horaire issue de la moyenne journalière, de la pluie et du débit maximal annuel.

2.4.4 Caractéristiques des chroniques de données

Les chroniques de données des 313 bassins versants de notre échantillon ont une durée comprise entre 3 et 33 ans, avec pour durée médiane 6 ans (Figure 2-13 a).

80 % des bassins versants ont des chroniques dont la durée est comprise entre 4 et 15 ans. Dans ces chroniques, les lacunes de pluie sont assez rares : plus de 70 % des bassins sont sans lacune de pluie (Figure 2-13 c). Les bassins versants restants possèdent une quantité limitée de lacunes de pluie, de telle manière que cela n'empêche pas l'utilisation des modèles pluie-débit. Concernant les chroniques de débit, la quantité de lacunes est légèrement plus importante : 42% des bassins versants sont sans lacune de débit et 90% possèdent moins de 18% de lacunes de débits (Figure 2-13 d). La présence d'une plus grande quantité de lacunes dans les chroniques de débit vient notamment du fait qu'il arrive que des stations de jaugeage soient détruites lors de fortes crues. Il faut ensuite du temps pour les reconstruire. De plus, par rapport à la mesure de la pluie qui est centralisée par Météo France, les moyens dont disposent les producteurs de données hydrométriques (DIREN, DDE, DDA, etc.) sont généralement plus limités.

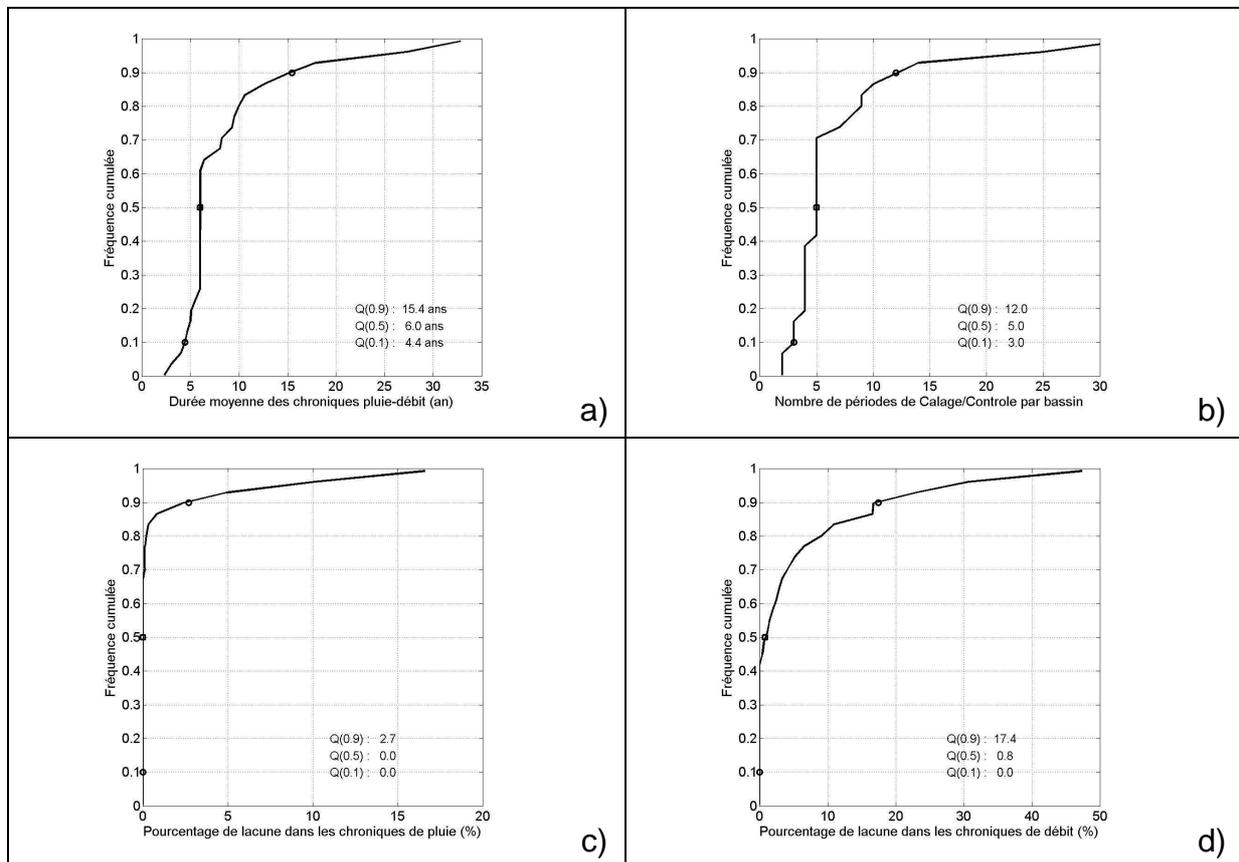


Figure 2-13 : Distribution des caractéristiques des chroniques de données sur les 313 bassins versants, (a) durée des chroniques pluie-débit, (b) nombre de périodes de calage-contrôle, (c) pourcentage de lacunes dans les chroniques de pluie, (d) pourcentage de lacunes dans les chroniques de débit.

Néanmoins, dans le cadre de la modélisation pluie-débit, les lacunes de débit sont beaucoup moins gênantes que les lacunes de pluie. En effet, une lacune de pluie

correspond à l'absence d'une donnée d'entrée dont a besoin le modèle pour fonctionner, alors qu'une lacune de débit qui correspond à la variable de sortie du modèle n'empêche pas son fonctionnement, mais empêche seulement le calcul de l'erreur aux pas de temps en question. Une lacune de pluie, lorsque la pluie n'est pas nulle, peut représenter une non-alimentation en eau du modèle. Si la lacune correspond à un fort épisode de crue, cela peut fausser durablement le niveau de remplissage des réservoirs du modèle et ainsi le fonctionnement du modèle. Lorsqu'une lacune de pluie est rencontrée, le modèle est initialisé à nouveau pour une période de un an*. Généralement, les lacunes de débit sont, soit longues (la station ne fonctionne pas), soit aléatoires et courtes (défaut d'enregistrement ou de transmission de la donnée). Lorsque les lacunes de débit sont longues, la période considérée n'est pas testée. Lorsque les lacunes de débit sont limitées, elles ont très peu d'influence le calage ou l'évaluation du modèle. Elles empêchent seulement la comparaison entre un débit observé et un débit simulé (c'est à dire, le calcul de l'erreur du modèle à un pas de temps).

2.4.5 Découpage des séries en sous-périodes pour la procédure de test en calage-contrôle (*split-sample test*, Klemeš [1986b]) :

La comparaison des performances de modèles hydrologiques est assez complexe, notamment par la diversité des usages possibles de ces modèles (Chapitre 1). D'une part, les modèles peuvent être utilisés (et donc évalués) pour différents modes d'utilisation : la simulation, la prévision, la prédétermination, ou la transposition à des bassins non jaugés, etc.. D'autre part, il existe une très grande diversité de critères d'évaluation, correspondant aux différentes performances des modèles que l'on souhaite tester. Néanmoins, malgré cette diversité, la procédure de calage-contrôle (*split-sample test*) proposée par Klemeš [1986b] permet d'évaluer les performances des modèles dans des conditions proches des conditions d'utilisation opérationnelle. Cette procédure correspond à la séparation de la période disponible en deux sous-périodes indépendantes et au calage du modèle sur la première sous-période, puis au contrôle du modèle sur la deuxième sous-période. Cette procédure permet en outre d'évaluer la robustesse des modèles, c'est-à-dire leur capacité à garder leur niveau de performances dans des conditions différentes des conditions d'apprentissage (calage).

Pour chaque bassin versant, la chronique est séparée en deux périodes indépendantes :

1. **La période de calage** est composée (i) d'une année de mise en route du modèle (*warm-up*) où les niveaux des réservoirs (variables internes) se stabilisent

* Cette durée d'initialisation de un an aurait pu être diminuée à 3 à 6 mois (suivant la complexité des modèles). Dans un souci d'homogénéité des tests, nous avons conservé cette durée.

progressivement, de manière à ce qu'ils soient devenus indépendants, à l'issue de cette période, des niveaux initiaux choisis arbitrairement et (ii) d'une année de calage. Cette durée de calage peut sembler relativement courte par rapport à la pratique de la modélisation pluie-débit avec des données journalières. Néanmoins, cette durée d'un an est réaliste par rapport à la disponibilité actuelle des données de pluie et de débit*. De plus, cette courte durée permet également de tester les modèles dans des conditions difficiles et d'assurer que les modèles qui sortiront de cette comparaison seront capables d'être calés avec une faible quantité d'information.

2. **La période de contrôle** est composée du reste de la chronique. La mise en route du modèle est d'au moins un an. Elle est parfois plus longue lorsqu'elle inclut la période de calage, quand les périodes de calage et de contrôle se suivent.

Ensuite, à l'issue de la première phase de calage-contrôle, d'autres procédures de calage-contrôle sont effectuées en décalant la période de calage d'une année et en effectuant le contrôle sur le reste de la chronique (Figure 2-14). Cette procédure permet ainsi d'échantillonner au mieux l'information contenue dans la chronique hydrologique et la variabilité climatique.

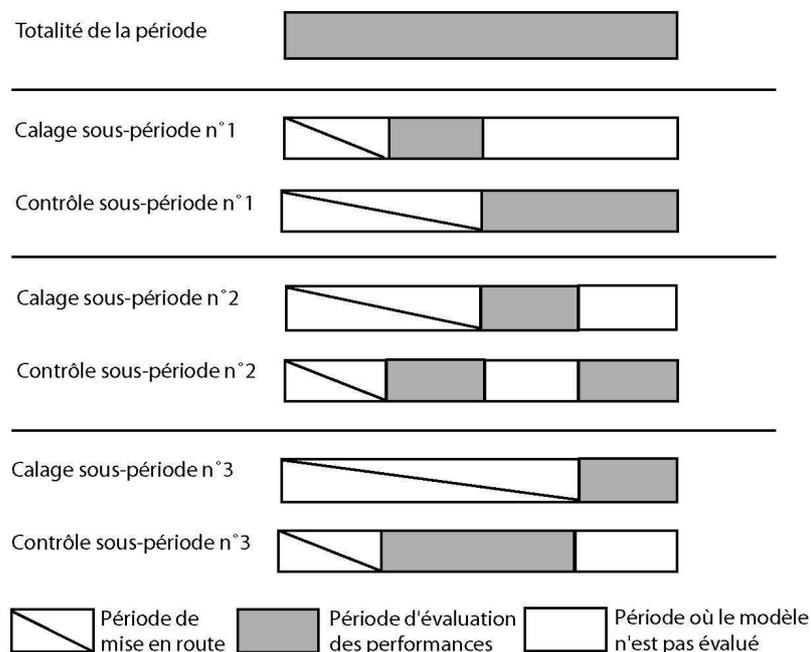


Figure 2-14 : Procédure de Split-sample test [Klemeš, 1986b]. Exemple de division d'une période de 4 ans, en trois sous-périodes indépendantes.

* à quoi ça servirait de développer un modèle qui nécessiterait pour être correctement calé une quantité de données non disponible ?

Ainsi, les 313 bassins versants de notre échantillon donnent lieu à une série de 2093 périodes de calage d'une durée de 1 an et 2093 périodes de validation d'une durée de 1 à 31 ans. En moyenne, cela représente 6 sous-périodes par bassin versant (Figure 2-13 b).

Le Tableau 2-3 présente la répartition du nombre de bassins versants et de périodes de calage-contrôle, par pays. Ce tableau met en évidence que les bassins américains ont un poids assez fort sur cet échantillon, notamment parce que leurs chroniques sont plus longues. Alors qu'ils représentent 22 % des bassins versants, ils représentent 33 % des périodes. Afin de garder un échantillon équilibré et d'étudier les performances des modèles sur la diversité des conditions hydro-climatiques de cet échantillon, les performances des modèles seront jugées par rapport aux performances moyennes par bassin, plutôt que par période (pour ne pas donner un poids trop important aux bassins arides américains).

Tableau 2-3 : Répartition des bassins versants et des périodes de calage-contrôle en fonction des différents pays d'origine (entre parenthèse, proportion du total en %).

	Nombre de bassins versants	Nombre de périodes de calage-contrôle
France	227 (72.5)	1329 (63.5)
États-unis	70 (22.4)	704 (33.6)
Australie	12 (3.9)	46 (2.2)
Slovénie	2 (0.6)	8 (0.4)
Espagne	2 (0.6)	6 (0.6)
Total	313	2093

2.5 Quelle position adopter face à des données de qualité imparfaite ?

Tout travail de modélisation hydrologique doit passer par la connaissance du contexte hydro-climatique de chaque bassin versant, son évolution au cours du temps (stationnarité de son fonctionnement) et l'évaluation de la qualité des données utilisées. Lorsqu'on s'intéresse au développement de modèles hydrologiques et à l'évaluation de leurs performances, il est clair que de larges échantillons de bassins versants sont très utiles. Malheureusement, un des inconvénients de l'utilisation de larges échantillons de bassins versants est qu'il devient impossible de vérifier de manière exhaustive la qualité des données de chaque bassin versant.

2.5.1 Quel type d'erreurs rencontre-t-on dans les données ?

Classiquement, la modélisation hydrologique utilise des données de pluie, de débit et d'ETP. Les données de pluie sont directement mesurées et lorsque plusieurs pluviomètres sont disponibles sur un bassin versant, elles sont assemblées pour en déduire la pluie du bassin versant. Les données de débit ne sont pas mesurées directement mais dérivées de la mesure des hauteurs d'eau au niveau d'une station limnimétrique. Les débits sont ensuite obtenus grâce à une courbe de tarage qui établit la relation entre la hauteur d'eau et les débits. La courbe de tarage est un modèle d'interpolation et d'extrapolation et est donc une source d'incertitudes. Les données d'ETP ne sont pas non plus mesurées directement, mais dérivées d'un modèle d'ETP. Le modèle le plus souvent utilisé, celui de *Penman* [1948], estime l'ETP à partir de la mesure de paramètres climatiques et énergétiques.

Les erreurs et les incertitudes dans les données sont inhérentes à tout processus de mesures et dépendent aussi bien de problèmes métrologiques, que de problèmes méthodologiques. Les erreurs rencontrées dans les données sont soit aléatoires, soit systématiques. Concernant la mesure de la pluie, les erreurs généralement rencontrées sont liées à des problèmes de captation du pluviomètre et au manque de suivi et d'entretien des pluviomètres (cône obstrué par des feuilles, etc.). Concernant la mesure des débits, les erreurs sont généralement liées à la qualité et à la représentativité de la courbe de tarage dans le temps. Les rivières sont 'vivantes' et leurs sections de contrôle ne sont pas toujours stationnaires au fil du temps. De plus, l'estimation des débits est un exercice difficile, particulièrement dans la mesure des extrêmes (étiage, crue). Cette estimation est biaisée à cause de problèmes de précision des systèmes de mesure, de robustesse des stations de jaugeage et d'extrapolation de la courbe de tarage*. Enfin, concernant l'estimation de l'ETP, les erreurs sont également d'ordre métrologique, puisque les valeurs d'ETP de *Penman* [1948] sont données par un modèle utilisant quatre données mesurées.

Dans le cadre de la modélisation hydrologique, après les erreurs liées à la mesure des entrées ou des sorties, s'ajoutent les erreurs liées à la représentativité spatiale des données d'entrée. En effet, si le débit en un point représente l'intégration des processus hydrologiques dans le bassin versant amont, la pluie et l'ETP ne sont mesurées que ponctuellement, alors que l'ensemble du bassin versant est soumis à un champ de pluie et un champ d'ETP. Ce deuxième niveau d'erreur vient de la manière dont sont estimés les champs de pluie et d'ETP à partir de valeurs ponctuelles. Concernant l'estimation du champ de pluie auquel est soumis le bassin versant, les erreurs dépendent généralement de la manière dont est calculée la pluie

* Les événements extrêmes sont rares !

moyenne et de la représentativité des pluviomètres (gradient altimétrique, position par rapport aux limites du bassin versant). Le thème de l'estimation de la pluie de bassin versant a été très abordé dans la littérature. Néanmoins, nous nous contenterons des résultats de *Andréassian et al.* [2001] qui mettent en évidence que les modèles pluie-débit globaux sont généralement capables de s'adapter à l'information de la chronique de pluie fournie au modèle, quelle que soit la méthode utilisée pour calculer la pluie moyenne. Nous plaçant dans un cadre de comparaison de modèles hydrologiques, nous avons choisi de calculer la pluie de bassin de la manière la plus simple, en faisant une moyenne équipondérée des valeurs de chaque pluviomètre. Nous pensons d'ailleurs que différents modèles progresseraient de manière similaire si on leur fournissait des chroniques de pluie pour lesquelles la représentativité réelle de chaque pluviomètre aurait été estimée. L'estimation d'une 'meilleure' pluie de bassin n'aurait pour conséquence que de faire progresser légèrement les performances des modèles. Concernant l'estimation du champ d'ETP, les mêmes problèmes de calcul de valeur moyenne et de représentativité de chaque station existent. Néanmoins, *Oudin* [2004] a mis en évidence la faible sensibilité des modèles pluie-débit aux valeurs ponctuelles de l'ETP. Vu la taille des bassins versants, le comportement des stations d'ETP à long terme sont généralement similaires et nous avons également choisi de calculer l'ETP de bassin de la manière la plus simple, en faisant une moyenne équipondérée des valeurs de chaque station d'ETP.

2.5.2 Impact des erreurs dans les données d'entrées sur l'efficacité et les paramètres des modèles pluie-débit (contribution dans *Oudin et al.* [2005])

En modélisation hydrologique, la connaissance de l'impact des erreurs des données d'entrée (pluie et ETP) sur les modèles est très importante. C'est une étape majeure de l'analyse du fonctionnement d'un modèle hydrologique. Les études de sensibilité doivent mettre en évidence la manière dont les sorties d'un modèle (dans notre cas : le débit) sont affectées, qualitativement et quantitativement, par les incertitudes sur les données d'entrée. Étrangement, les études de sensibilité aux données d'entrée sont rares dans la littérature [*Paturel et al.*, 1995]. De même, l'impact des erreurs dans les débits ne semble pas avoir été étudié en détail. Pourtant, ces erreurs peuvent affecter (1) les performances, (2) les paramètres et (3) l'incertitude sur les paramètres des modèles. Dans la littérature, on ne trouve d'ailleurs pas de conclusions claires. La diversité des résultats est due entre autres aux différentes stratégies de simulation d'erreurs dans les données (corruption), aux études sur données synthétiques ou réelles, aux différents types d'approches utilisées (non recalage ou recalage du modèle), etc.

Concernant l'ETP, de nombreux travaux ont montré la faible sensibilité des modèles aux variations à court terme et ainsi aux erreurs aléatoires de l'ETP [Edijatno et Michel, 1989b; Andersson, 1992; Joukainen, 2000; Kokkonen et Jakeman, 2001; Andréassian et al., 2004b; Oudin, 2004; Oudin et al., 2005b; Oudin et al., 2005c; Oudin et al., 2005d].

Pour ce qui est de la pluie, les études de sensibilité portant sur l'impact d'erreurs aléatoires et systématiques sont également rares. Il est plus facile de trouver des études portant sur les impacts de la variabilité spatiale et temporelle de la pluie sur les débits. Les études utilisant une approche statique (non recalage du modèle après corruption des entrées) concluent que les erreurs aléatoires et systématiques affectent très fortement les performances des modèles [Paturel et al., 1995; Nandakumar et Mein, 1997]. Par contre, les études utilisant une approche dynamique (recalage du modèle après corruption des entrées) concluent que les erreurs aléatoires ou systématiques n'affectent pas ou peu les performances des modèles [Dawdy et Bergmann, 1969; Ibbitt, 1972; Troutman, 1982; Troutman, 1983; Andréassian et al., 2001; Andréassian et al., 2004a; Oudin et al., 2005d]. Seules les erreurs systématiques affectent fortement les valeurs des paramètres.

2.5.3 Influence d'ouvrages sur le fonctionnement des bassins versants

Les barrages sont susceptibles de modifier le fonctionnement naturel des rivières. En fonction de leur distance par rapport à la station de jaugeage en aval, leur influence peut être transparente comme prépondérante sur le débit des rivières. Les modèles pluie-débit utilisés ne prennent pas en compte le fonctionnement des barrages et peuvent ainsi être pénalisés par l'existence d'un barrage sur le bassin versant. Dans la base de données, quelques bassins versants (10 à 15 en France, Figure 2-15 ; 1 en Australie) semblent influencés par le fonctionnement de barrages, situés en amont de la station de jaugeage. Ne connaissant pas le mode de gestion de ces barrages, il n'est pas possible de connaître exactement leur impact sur le fonctionnement des rivières.

Néanmoins, l'inspection visuelle de quelques chroniques a surtout révélé des impacts de lâchures de barrages au cours des étiages. Pendant ces périodes, le fonctionnement de ces barrages se traduit notamment par la présence de pics de débits journaliers, ayant lieu du lundi au vendredi. Cette influence, généralement limitée et détectable seulement pendant les périodes d'étiages, est clairement mise en évidence par des analyses en ondelettes de la chronique de débit (Figure 2-16).

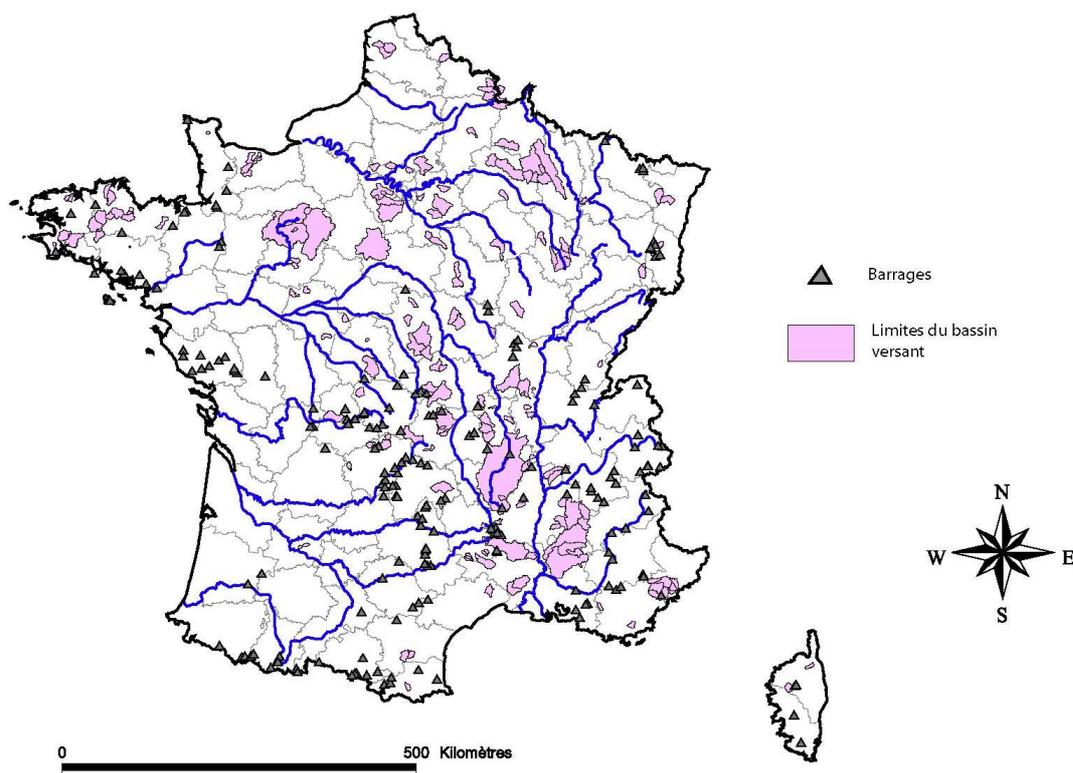


Figure 2-15 : Positions des principaux barrages, par rapport aux 227 bassins versants français.

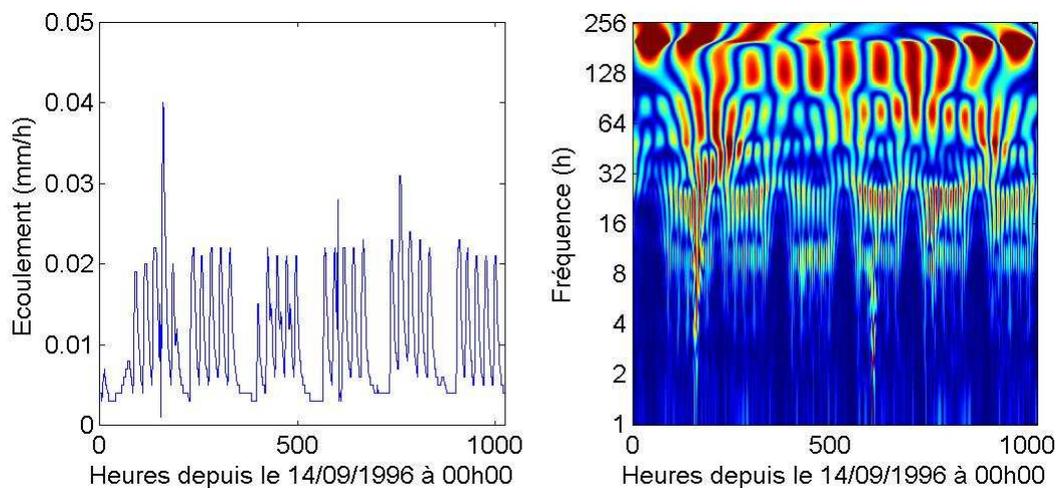


Figure 2-16 : Impact du fonctionnement de barrages sur le débit de la Loire à Feurs. Écoulement horaire et analyse en ondelettes. La présence d'une périodicité est représentée par les couleurs vert à rouge entre les fréquences (axe y).

2.5.4 Stationnarité des bassins versants*

De Marsily [1994] discute des conditions d'utilisation des modèles qu'il définit comme « Boîte noire », qui sont les trois unités de la tragédie classique : (1) unité de lieu, (2) unité de temps et (3) unité d'action. Ces conditions contraignent les modèles calés par apprentissage sur une période à être utilisés sur d'autres périodes (1) strictement pour le domaine d'espace sur lequel ils ont été calés par apprentissage, (2) si les processus modélisés n'ont pas changé, (3) si le système ne s'est pas modifié au cours du temps (non stationnarité du système, modification d'origine anthropique). D'après cet auteur, ces trois conditions encadrent l'utilisation stricte de tels modèles.

Dans notre cas, les paramètres des modèles sont calés sur une série de données. Ces paramètres sont censés fournir au modèle l'information suffisante pour rendre compte des caractéristiques du comportement du bassin versant (néanmoins, il existe un lien certain entre le vecteur de paramètres optimisé et les caractéristiques de la chronique hydrologique utilisée pour cela). Les trois contraintes précédentes limitent la validité du jeu de paramètres identifié au cours du calage. En effet, même si la première condition est toujours respectée, les autres conditions supposent la stationnarité du comportement du bassin sur le long terme, qui est largement contrainte par la stationnarité des conditions climatiques (condition 2) et par la stationnarité du contexte physique du bassin (condition 3).

A propos de la stationnarité ou non des conditions climatiques (condition 2), il est souvent assez difficile de l'évaluer car elle n'est observable qu'à l'échelle de quelques décennies et il n'est pas impossible que les bassins versants de notre base de données aient été soumis à une évolution régionale de leur climat. Néanmoins, la plupart de nos bassins versants ont des chroniques d'une durée de quelques années. Seulement 10 % des bassins versants ont des chroniques relativement longues, de 15 ans à 30 ans (en France, l'Orgeval et le Réal Collobrier ; bassins des États-Unis). Ainsi, pour les durées des chroniques de cet échantillon, il ne semble pas qu'une non stationnarité des conditions climatiques (pluies, ETP) soit détectable et suffisante pour induire une modification du contexte physique du bassin versant.

A propos de la non stationnarité du contexte physique du bassin versant (condition 3), celle-ci peut être induite par des modifications naturelles (croissance de la végétation [Andréassian, 2002]), accidentelles (feu de forêt : bassin versant du Réal Collobrier, Le Rimbaud, Y4616210, ravagé par un incendie en Août 1990), ou d'origine anthropiques (entre autres, coupe forestière : bassins versants de Caspar

* le terme stationnarité est utilisé dans le sens de la stationnarité à long terme des processus météorologiques, hydrologiques et des caractéristiques des bassins versants.

Creek, USA). Ces modifications sont, soit relativement progressives (modifications naturelles ou climatiques), soit relativement rapides (modifications accidentelles ou anthropiques). Dans le cadre des modifications progressives du contexte physique du bassin versant, il n'est pas facile de mettre en évidence une non-stationnarité dans le fonctionnement du bassin versant. De plus pour les mêmes raisons que précédemment, la durée des chroniques disponibles ne le permettrait pas.

Par contre, dans le cadre des modifications rapides, du contexte physique du bassin versant, l'influence est généralement plus nette sur les débits, notamment pour les modifications d'origine anthropique. Dans notre base de données, il semble que seuls quelques bassins versants ont subi ce type de modifications : Le Rimbaud et les cinq bassins versants de Caspar Creek. Les chroniques de Caspar Creek ont été séparées en deux périodes, avant et après la coupe forestière. La chronique du Rimbaud n'a pas été scindée, l'impact majeur du feu de forêt sur le fonctionnement du bassin versant n'ayant pas été clairement détectée.

2.5.5 Quelle position adopter face à des données de qualité imparfaite ?

Quelles qu'en soient les justifications, l'évaluation de la qualité des données d'un échantillon aussi large est une tâche très longue, difficile et certainement subjective. Notre position, contestée et rarement adoptée mais à notre sens justifiée, a été de garder tous les bassins versants.

Les raisons sont les suivantes :

1. **Absence de méthode globale d'évaluation de la qualité des données** : à l'heure actuelle, il n'existe pas de méthode rapide permettant l'évaluation objective de la qualité des données hydrologiques. Seuls les producteurs de données peuvent se prononcer sur la qualité de leurs données. Dans un tel travail de recherche, où les données rassemblées proviennent d'un grand nombre de sources différentes, il est impossible d'obtenir des informations claires et cohérentes de tous les producteurs de données ;
2. **Impossibilité de juger les données au travers de la modélisation pluie-débit** : la qualité des données est parfois appréciée en fonction des faibles performances obtenues par un modèle hydrologique. Même s'il est probable que l'échec de nombreux modèles sur un bassin versant vienne de problèmes dans les données, cette pratique peut à l'évidence conduire à de nombreux abus. Notamment à ne choisir que les bassins versants où le modèle testé est bon. Dans l'objectif d'une comparaison des performances de modèles hydrologiques, la sélection des bassins versants testés peut fortement biaiser l'interprétation des résultats. De plus, lorsqu'un modèle est utilisé pour juger la qualité des données

d'un bassin versant, il ne faut pas oublier que certains modèles peuvent avoir des difficultés à représenter le fonctionnement de certains types de bassins versants (cas par exemple des bassins versants intermittents ou non-conservatifs). Ainsi, la sélection des bassins versants en fonction des performances obtenues par des modèles est très subjective et peut conduire à juger de la mauvaise qualité de certaines données pour de mauvaises raisons ;

3. **Mise dans des conditions d'utilisation courante des données :** dans la plupart des applications des modèles hydrologiques, les données utilisées sont celles des réseaux de mesures hydro-climatiques nationaux, dont la densité et les moyens sont limités. Par exemple, la couverture des pluies ou la qualité des courbes de tarage de ces réseaux est probablement moins bonne que celles des bassins versants expérimentaux, qui sont souvent de superficie limitée et pour lesquels les équipes de recherche possèdent plus de moyens. Mais, les travaux des hydrologues et l'application des modèles hydrologiques doivent-ils se limiter aux cas exceptionnels où les données sont de très bonne qualité ?;
4. **Impact équivalent des erreurs de données sur les différents modèles :** plusieurs types d'erreurs sont présents dans les données, des erreurs aléatoires et des erreurs systématiques. Les erreurs aléatoires, sont, par nature, difficiles à détecter). Par contre, les erreurs systématiques sont plus facilement détectables ou corrigées (étalonnage des pluviomètres, mise à jour régulière de la courbe de tarage). Néanmoins, quel que soit le type d'erreurs, les erreurs sur la pluie, les erreurs sur les débits et les erreurs des modèles sont généralement indépendantes entre elles et indépendantes des erreurs des modèles pluie-débit. On peut raisonnablement penser que différents modèles sont affectés de la même manière par ces différents types d'erreurs. La conséquence principale de données de mauvaise qualité est de réduire les performances des modèles testés. Mais, comme l'a dit *Linsley* [1982] : "*if the data are too poor for the use of a good simulation model they are also inadequate for any other model (p. 13).*". Enfin, de nombreux travaux montrent que les modèles sont capables de s'adapter (à travers le calage) à des erreurs contenues dans les données, lorsque celles-ci restent raisonnables (entre autres, *Boughton* [2005a] ;
5. **Limitation de l'impact des erreurs par l'utilisation d'un grand nombre de bassins versants :** nous pensons que tous les modèles testés vont souffrir de manière équivalente de données de mauvaise qualité. De plus, nous pensons que sur un large échantillon de bassins versants, l'influence des quelques bassins versants dont les données sont mauvaises sera plus limitée. Cette opinion est d'ailleurs partagée par *Boughton* [2005b] : "*There is no way to separate the*

quality of the modelling performance from the quality of the input data on the quality of results except by calibration of a number of models on a substantial number of catchments for direct comparison of results. Unfortunately, there are no such studies reported in the literature (p. 103).”;

Pour toutes ces raisons, nous avons donc choisi de garder les données de tous les bassins versants dont nous disposons. Nous pensons que la sélection ou l'élimination de bassins versants est certainement subjective et ne peut que biaiser l'analyse des performances des modèles. Nous nous sommes cependant adaptés aux bassins versants connus comme étant non-stationnaires (bassins expérimentaux) en séparant les séries de données en morceaux stationnaires.

2.6 Synthèse

Dans ce chapitre, nous avons présenté les 313 bassins versants que nous avons utilisés pour développer et tester des structures de modèles hydrologiques pluie-débit. Le regroupement d'un tel échantillon est inédit pour le pas de temps horaire et représente plus d'une année de travail. La richesse de cet échantillon est d'être composé de bassins versants aux caractéristiques climatiques, hydrologiques et physiographiques très variées. Les bassins versants sont principalement situés en France, mais également aux États-Unis, en Australie, en Slovénie et en Espagne. Par ailleurs, pour les bassins versants français, nous disposons de quelques données géo-référencées (occupation des sols, géologie, limite des bassins versants) dont nous nous sommes servis pour l'analyse des performances des modèles.

Le regroupement d'un tel échantillon permet ainsi de s'affranchir de la dépendance des conclusions de cette étude, par rapport aux caractéristiques de quelques bassins versants. Cet échantillon permet également de tester la robustesse et la généralité des performances des modèles pluie-débit et d'étendre la validité du travail de comparaison de différentes structures de modèles. De plus, cet échantillon nous permettra de donner plus de confiance aux modifications des modèles, proposées pour améliorer leurs performances au pas de temps horaire.

Enfin, tous les bassins versants pour lesquels nous avons collecté des données ont été conservés pour les tests. Nous n'avons pas voulu introduire de sélection subjective des 'bons' et des 'mauvais' bassins versants. De plus, le temps de vérification d'une telle quantité de chroniques hydrologiques aurait fortement rallongé la durée de construction de cet échantillon, qui était déjà assez longue. Plaçant cette étude dans le cadre d'une étude comparative de différents modèles, l'intérêt de

mener une critique exhaustive des données est limité. Nous pensons que les performances des modèles hydrologiques sont également affectées par les bassins versants dont les données sont mauvaises. L'élimination des données de quelques bassins versants n'aurait pour effet que d'augmenter le niveau de performance moyen des modèles testés sur cet échantillon.

Chapitre 3

« L'Homme est enfin, on peut le supposer, le seul animal qui sache qu'il doit mourir. Ses luttes journalières compétitives, sa recherche du bien-être à travers l'ascension hiérarchique, son travail machinal accablant, lui laissent peu de temps pour penser à la mort, sa mort. C'est dommage, car l'angoisse qui en résulte est sans doute la motivation la plus puissante à la créativité. Celle-ci n'est elle pas en effet une recherche de la compréhension, du pourquoi et du comment du monde et chaque découverte ne nous permet elle pas d'arracher un lambeau au linceul de la mort ? N'est ce pas ainsi que l'on peut comprendre qu'en son absence, celui qui 'gagne' sa vie la perd ? »

Henri Laborit, *Éloge de la fuite*, 1985.

CHAPITRE 3 Stratégie de comparaison des performances de modèles pluie-débit

3.1 Introduction

Une partie de ce chapitre a été soumise à Water Resources Research (Annexe 11).

Ces dernières décennies, la connaissance des processus hydrologiques et la disponibilité des données hydro-météorologiques se sont largement améliorées, les puissances de calculs des ordinateurs ont explosé. Ces progrès ont permis aux hydrologues de développer un grand nombre de modèles hydrologiques. Face à cette prolifération de modèles, le besoin de comparer leurs performances respectives s'est fait sentir assez tôt [WMO, 1975]. En fait, à partir des années 1960, le développement des modèles hydrologiques a commencé par une phase de croyance en la physique de l'hydrologie (Modèle de Stanford à 35 paramètres, Crawford et Linsley [1966]), laissant penser que la description la plus exhaustive possible du fonctionnement hydrologique des bassins versants et l'outil informatique permettraient de résoudre tous les problèmes. Cela a conduit au développement de modèles visant à représenter de façon aussi exhaustive que possible les processus hydrologiques, avec un niveau de complexité très élevé. Le but initial de ces comparaisons était d'évaluer la validité des structures de ces différents modèles, notamment pour fournir des conseils aux utilisateurs sur leurs domaines d'application. Ces travaux de comparaison nécessitent d'évaluer objectivement les performances d'un modèle. Or, dans le domaine des sciences de l'environnement et en particulier en hydrologie (où il n'existe pas actuellement de théorie scientifique formalisée à l'échelle du bassin versant), il est très subjectif de dire qu'un modèle est *bon* ou *mauvais* de manière absolue. Par contre, comme nous l'avons vu au premier chapitre, il est beaucoup plus objectif de dire qu'un modèle est *meilleur* ou *moins bon* qu'un autre modèle, ce qui peut être fait lors des comparaisons.

L'objectif de ce chapitre est de démontrer que pour mettre en évidence les qualités souhaitées des modèles hydrologiques (généralité, transposabilité, robustesse ; voir section 1.4.3), les stratégies de comparaison doivent :

1. s'appuyer sur un large échantillon de bassins versants, regroupant une vaste gamme de conditions hydrométéorologiques et de caractéristiques physiques ;
2. avoir recours à une formulation du critère d'évaluation des performances des modèles adaptée aux larges échantillons de bassins versants.

La première condition permet de rendre les conclusions d'une comparaison plus générale (proposition n°6 du Chapitre 1) : les conclusions ne sont plus limitées aux caractéristiques de quelques bassins versants, elles deviennent alors transposables. L'utilisation de larges échantillons de bassins versants a d'ailleurs été recommandée par de nombreux auteurs [Moore et Mein, 1975; Linsley, 1982; Klemeš, 1986b; Andersson, 1992; Perrin et al., 2001a; Boughton, 2005b]. Par exemple, Klemeš [1986b] considère que seul ce type d'échantillons permet de tester et de valider les performances des modèles pluie-débit. Cependant, il semble que ces recommandations ont été peu suivies, puisque la plupart des études de comparaison comportent peu de bassins versants (qui sont généralement homogènes) (Tableau 3-1).

La deuxième condition est imposée par l'utilisation de larges échantillons de bassins versants. En effet, un des problèmes spécifique de ces grands échantillons (en plus du temps qu'il faut pour réunir les données !) est le calcul de statistiques de performances significatives. Tant que la comparaison ne porte que sur quelques bassins versants, il est toujours possible d'utiliser différents types de critères (qualitatifs et quantitatifs) pour comparer les performances respectives et discuter des résultats. En revanche, avec un large échantillon, il est beaucoup plus long de comparer des distributions de performances de plusieurs modèles, performances dont les domaines de variation peuvent être très larges. Aussi, pour synthétiser cette grande quantité de résultats (valeurs de critères), il est important d'utiliser une formulation de critère qui permette d'obtenir des distributions de résultats relativement symétriques sur l'ensemble de l'échantillon. En effet, l'asymétrie rend délicate l'extraction d'une statistique résumant l'ensemble des performances d'un modèle. Dans sa formulation, le critère de *Nash et Sutcliffe* [1970] (NS), classiquement utilisé en hydrologie, n'est pas adapté aux larges échantillons, à cause de son domaine de variation qui n'est pas borné négativement. La formulation du critère proposé dans la suite de ce travail, appelé C_{2M} , permet de compenser certains des problèmes du critère de NS.

3.2 Limitations des comparaisons existantes et intérêt des larges échantillons

3.2.1 Limitations des comparaisons existantes

Dans la littérature, de nombreuses comparaisons des performances de modèles pluie-débit sont disponibles (Tableau 3-1). Généralement, ces études rapportent plutôt les résultats de quelques modèles, testés sur quelques bassins versants. La revue de ces comparaisons a été faite par *Michaud et Sorooshian* [1994], *Refsgaard et Knudsen* [1996] et *Perrin* [2000]. Ces études ont testé différents types de modèles

hydrologiques, fondés sur des approches conceptuelles, empiriques, physique ou boîte-noire suivant une grande variété de méthodologies variant suivant :

- Le nombre, de la taille et du climat des bassins versants,
- Les représentations spatiales des bassins versants : global, semi-distribuée, distribuée,
- Les fonctionnements des modèles testés : continu ou événementiel.
- Les pas de temps testés : horaire, journalier, mensuel ou annuel,
- Les méthodes de calage, les critères de calage, les durées de calage et les stratégies d'évaluation des résultats (en calage, en validation, en suivant différents types de tests préconisés par *Klemeš* [1986b], etc.),
- Les critères d'évaluation des performances : critères qualitatifs ou de nombreux critères quantitatifs.

Cette grande diversité de tests, effectués généralement avec peu de modèles et sur peu de bassins versants, rend l'interprétation et la synthèse de ces études assez difficile, comme le regrettent [*Perrin et al.*, 2001a; *Smith et al.*, 2004]. Seules les études de quelques auteurs permettent d'avoir une vue d'ensemble des performances de plusieurs modèles sur un large échantillon de bassins versants. Néanmoins, dans ces études les performances des différents modèles ont été synthétisées sous forme de tableaux et/ou de distributions de performances, qui sont parfois difficile à interpréter quand elles sont trop nombreuses. Il semble ainsi qu'aucun de ces auteurs n'ait cherché un critère de performance pertinent qui permette de synthétiser clairement une telle quantité de résultats.

3.2.2 Intérêt des larges échantillons de bassins versants

Parmi les comparaisons de modèles pluie-débit, seulement quelques-unes s'intéressent aux performances de différents modèles sur un large ensemble de bassins versants aux caractéristiques variées. Les études de *Vandewiele et al.* [1992] et *Xu et Vandewiele* [1995] portent sur la comparaison plusieurs modèles mensuels, testés sur 79 et 81 bassins versants situés en Belgique, Chine et Birmanie. L'étude de *Makhlouf et Michel* [1994] porte sur la comparaison de cinq modèles mensuels, testés sur 81 bassins versants situés en France. L'étude de *Perrin et al.* [2001a] porte sur 19 modèles journaliers, testés sur 429 bassins versants, situés en France, États-unis, Australie, Brésil et Côte d'Ivoire. Enfin, *Boughton* [2005b] présente une revue d'études impliquant plusieurs modèles, testés aux pas de temps mensuel et journalier, sur plusieurs centaines de bassins versants en Australie.

Tableau 3-1 : synthèse de quelques comparaisons de modèles pluie-débit : nombre et types de modèles, pas de temps des données et nombre de bassins versants (H : horaire, J :

journalier, M : mensuel ; CRR : modèles conceptuel, PB : modèle basé sur la physique, BB : modèle boîte noire, GB : modèle basé sur la géomorphologie ; L : modèle global, SD : modèle semi-distribué ; D : modèle distribué)

Référence	Nombre de modèles	Type de modèle	Bassin versant	Pas de temps testé	Nombre de bassins versants
Moore et Mein [1975]	3	CRR	L	J, M	4
WMO [1975]	10	CRR, BB	L	J	2 à 6
Weeks et Hebbert [1980]	5	CRR, BB	L	J, M	3
Naef [1981]	6	CRR, BB	L	<H	3
Loague et Freeze [1985]	3	BB, CRR, PB	L	H	3
Moussavi et Feyen [1990]	6	CRR	L	M	1
Wilcox et al. [1990]	2	CRR	L	M, A	6
Franchini et Pacciani [1991]	7	CRR	L	H	4
Vandewiele et al. [1992]	2 à 8	CRR	L	3*M, A	1 à 79
Chiew et al. [1993]	6	BB, CRR	L	J, M, A	8
Makhlouf et Michel [1994]	5	CRR	L	M	81
Xu et Vandewiele [1995]	7	CRR	L	M	81
Refsgaard et Knudsen [1996]	3	CRR,PB,	L, D	J, M, A	3
Zhang et Lindström [1996]	2	CRR	L	J	2
Ye et al. [1997]	3	CRR, PB	D	J, M	3
Gan et al. [1997]	4 à 5	CRR	L	J	3
Yang et al. [2000]	3	PB,CRR,GB	D	J	1
Kokkonen et Jakeman [2001]	2	CRR	L	D	2
Perrin et al. [2001a]	19	CRR	L	D	429

Ces auteurs sont les précurseurs dans l'utilisation de larges échantillons de bassins versants pour développer, tester et comparer les performances de modèles pluie-débit.

Nous pensons que seul ce type de démarche permet de tirer des conclusions générales sur les modèles pluie-débit qui soient relativement indépendantes des conditions hydro-climatiques des bassins versants testés. Comme on le verra par la suite, les 'bons' modèles généralistes sont suffisamment souples et efficaces pour avoir de bonnes performances sur une large gamme de bassins versants. Leurs performances ne sont d'ailleurs pas forcément inférieures à celles de modèles développés pour ou sur un seul type de bassins versants. D'ailleurs, nous pensons que dans un contexte d'approche 'descendante' (*downward approach*, Klemeš [1983], Sivapalan et al. [2003]) seuls les larges échantillons de bassins versants permettent de tester si une philosophie de modélisation est capable de représenter les propriétés émergentes des bassins versants.

Dans la littérature, certains auteurs [Nash et Sutcliffe, 1970; Bergström et Forsman, 1973; Linsley, 1982; Klemeš, 1986b; Bergström, 1991; Lindström et al., 1997] insistent sur les qualités requises par les modèles pluie-débit. Ainsi, si la généralité d'un modèle est une des qualités que l'on recherche, il est clair que ce modèle ne peut pas être évalué sur peu de bassins versants. Bien au contraire, connaissant la variété naturelle des bassins versants, la variabilité des philosophies de modélisation de la relation pluie-débit (représentée par les différents modèles disponibles), un nombre croissant d'auteurs [Moore et Mein, 1975; Linsley, 1982; Klemeš, 1986b; Bergström, 1991; Andersson, 1992; Perrin et al., 2001a] encouragent l'utilisation de larges échantillons de bassins versants (Tableau 3-2).

Enfin, les grands échantillons de bassins versants ont deux autres intérêts. D'une part, ils permettent de réaliser des études de sensibilité et de régionalisation beaucoup plus exhaustives, ce qui rend les résultats plus robustes [Merz et Blöschl, 2004; Oudin, 2004; Merz et Blöschl, 2005; Parajka et al., 2005; Rojas-Serna, 2005]. D'autre part, les grands échantillons de bassins versants permettent de définir une distribution *a priori* des paramètres d'un modèle, une information très précieuse qui peut être mise à profit pour améliorer notamment les méthodes d'optimisation des paramètres des modèles pluie-débit (voir Chapitre 8).

Bien entendu, l'utilisation de tels échantillons de bassins versants pose un certain nombre de problèmes, notamment celui de la critique de données (ce point a été discuté au Chapitre 2).

Tableau 3-2 : Extraits des références encourageant l'utilisation de larges échantillons de bassins versants.

Référence	Citation
Moore et Mein [1975]	<p><i>“The following criteria formed the basis for the selection of catchments used in this study:</i></p> <p><i>(i) Different climatic zones should be represented to enable a variety of flow conditions to be studied, (...)</i></p> <p><i>(iv) The catchments on which the original versions of the models were developed should not be included to ensure independence of the test. (p. 123)”</i></p>
Linsley [1982]	<p><i>“because almost any model with sufficient free parameters can yield good results when applied to a short sample from a single catchment, effective testing requires that models be tried on many catchments of widely differing characteristics, and that each trial cover a period of many years (p. 14-15)”.</i></p> <p><i>“As long as some error may result from a less than precise representation of one or more of the hydrologic processes, many different catchments may be our only means of detecting these errors and, hence, devising modifications to correct the problem (p. 15).”</i></p> <p><i>“There should be data (...) for at least 50 and preferably 100 sites distributed around the world. With such information it will be possible to test model performance and compare different models to determine whether one algorithm is superior to another (p. 17)”</i></p>
Klemeš [1986b]	<p><i>“Use of more test basins, more extensive split-sample schemes, etc., would increase the credibility standing of a model, and an accumulation of test results may lead to meaningful generalizations (p. 22)”.</i></p>
Bergström [1991]	<p><i>“Growing confidence are thus key words in hydrological modeling. This can be obtained by applying model under a span of different geographical, climatological and geological conditions (p. 127)”</i></p>
Andersson [1992]	<p><i>“Another conclusion that could be drawn from the attempts of model improvement is that a certain change of model structure can improve the model performance for some basins whereas it is unchanged or deteriorated for other basins. Improvements can also occur only for certain periods. It is therefore important to test the new model for a large set of basins and for long time series before drawing conclusions of a general model improvement (p. 330).”</i></p>
Perrin et al. [2001a]	<p><i>“we believe that the robustness and reliability of a model resides primarily in its ability to perform under as varied a set of hydrological conditions as possible (p. 276).”</i></p> <p><i>“Restricting the study to a smaller number of catchments, e. g. 20 or 30, would have meant becoming overly dependent on an arbitrary selection of catchments. The main value of this research lies in including as many catchments as were available when the work began. A more in depth analysis of a few catchment might not have provided a better insight into the relative merits of the tested models. We strongly believe that a really proficient model should go beyond catchments peculiarities (p. 277)”</i></p>

3.3 Données et méthodes

3.3.1 Échantillon et sous-échantillons de bassins versants

Pour démontrer l'intérêt des grands échantillons de bassins versants pour étudier les performances de différents modèles pluie-débit, nous utilisons notre échantillon de 313 bassins versants. On peut raisonnablement considérer que cet ensemble représente déjà un bon échantillonnage des bassins versants existants. Nous considérons que cet échantillon nous permet de mesurer les performances moyennes 'vraies' des modèles pluie-débit, vers lesquelles devraient tendre tout sous-échantillon aléatoire de bassins versants.

Pour tester l'influence de la taille de l'échantillon de bassins versants sur la comparaison des performances de deux modèles pluie-débit, nous avons procédé à des tirages aléatoires de n sous-échantillons de m bassins versants parmi ces 313 bassins versants. Ensuite nous avons comparé les n performances moyennes de deux modèles, pour chaque taille (m) de sous-échantillon. Enfin, nous avons formulé une hypothèse à tester :

Hypothèse H_0 : sur l'échantillon de m bassins versants, la performance moyenne du modèle A n'est pas différente de celle du modèle B,

Pour cela, nous avons comparé les moyennes des différents modèles sur chaque sous-échantillon et nous avons mesuré la probabilité de rejeter/accepter H_0 , connaissant le résultat de ce test sur l'échantillon total, c'est-à-dire, la probabilité de conclure que la performance moyenne du modèle A est (ou n'est pas) significativement différente de celle du modèle B, alors que le contraire est vrai (ou faux) sur l'échantillon de 313 bassins versants. Autrement dit, ce test nous permet d'étudier l'influence de la taille d'un échantillon de bassins versants sur la probabilité de conclure qu'un modèle est moins bon qu'un autre, alors que le contraire serait 'vrai'.

3.3.2 Modèles pluie-débit utilisés

Nous avons utilisé quatre modèles pluie-débit (Tableau 3-3 ; description en Annexes Annexe 1) : GR4J, IHAC6, TOPMO8, MORDOR10. Ces quatre modèles ont été sélectionnés parce qu'ils sont largement utilisés. Ce sont des modèles conceptuels ou empiriques, mais ils reposent tous sur des philosophies de modélisation et des niveaux de complexité différents (approches de développement, nombre de paramètres libres, nombre de réservoirs, formulations mathématiques du modèle). Ces quatre modèles sont testés au pas de temps horaire, sur les mêmes périodes, les mêmes données d'entrée (pluie, ETP) et de calage (débits).

Tableau 3-3 : Caractéristiques des 4 modèles pluie-débit testés

Modèle	Nombre de paramètres libres	Nombre de réservoirs	Références
MORDOR10	10	4	Dérivé de <i>Garçon</i> [1999] et <i>Paquet</i> [2004]
TOPMO8	8	3	Dérivé de <i>Beven et Kirkby</i> [1979], <i>Beven et al.</i> [1995] et <i>Edijatno et al.</i> [1999]
IHAC6	6	3	Dérivé de <i>Jakeman et al.</i> [1990]
GR4J	4	2	<i>Perrin et al.</i> [2003]

3.3.3 Méthode d'évaluation des performances des modèles et de comparaison des résultats

Pour estimer la performance des modèles, la procédure de calage-contrôle croisée (*split-sample test*) préconisée par *Klemeš* [1986b] et décrite à la section 2.4.5, a été appliquée.

Les paramètres des modèles ont été optimisés avec la même méthode de calage, la méthode pas-à-pas [*Michel*, 1989; *Edijatno*, 1991]. Cette méthode d'optimisation locale est simple et particulièrement efficace (chapitre 4). Pour optimiser les paramètres des modèles, la fonction objectif utilisée est la racine carrée de l'erreur quadratique moyenne (RMSE). Les critères utilisés pour évaluer la performance des modèles en validation sont le critère de NS et sa version bornée, le critère C_{2M} . Ces deux critères sont introduits dans la partie suivante.

Pour comparer les performances de deux modèles sur un échantillon de m bassins versants, il est possible de faire un test statistique de comparaison de moyenne. Un test, comme le test de Student, permet de tester si la différence de performance moyenne entre deux modèles est statistiquement significative. Le test de Student fait l'hypothèse que la distribution des performances d'un modèle est approximativement gaussienne. La statistique T de Student se définit ainsi :

$$T_{A,B} = \frac{\mu_A - \mu_B}{\sqrt{\frac{\sigma_A^2}{m} + \frac{\sigma_B^2}{m}}} \quad [3.1]$$

où, μ_A et σ_A , et μ_B et σ_B représentent la moyenne et l'écart-type des performances du modèle A et B respectivement, sur l'échantillon de m bassins versants. La différence

de performance moyenne entre les modèles A et B peut être considérée comme significative, à un seuil de 5%, si $|T_{A,B}| > 1.96$.

3.4 Quelle formulation de critère est appropriée pour évaluer les performances d'un modèle pluie-débit sur un large échantillon de bassins versants ?

3.4.1 Quels types de critère pour comparer les performances des modèles pluie-débit ?

Dans le domaine de la modélisation hydrologique, il existe un grand nombre de critères d'évaluation des performances des modèles. Comme l'ont montré de nombreux auteurs [Green et Stephenson, 1986; Weglarczyk, 1998] et comme le démontre la diversité des formulations des critères existants, il n'existe pas de critère idéal ou unique qui permette de juger des performances d'un modèle, activité qui relève d'ailleurs des seuls utilisateurs des modèles. La grande variété de critères d'évaluation tient principalement à deux raisons [Perrin et al., 2005] :

1. la plupart des variables hydrologiques ont des domaines de variation très larges, qui sont typiquement de plusieurs ordres de grandeurs (3 à 4), avec un niveau d'erreur sur le débit simulé dépendant du niveau de débit. Une formulation de critère ne peut pas être également pertinente pour toutes les gammes de débits,
2. les intérêts des utilisateurs sont larges et les modèles sont généralement calés pour différentes gammes de débits, qui ont des ordres de variation différents.

Dans la longue liste des critères existants, il est possible de distinguer deux types de critères : les critères qualitatifs et les critères quantitatifs. Les critères qualitatifs sont principalement basés sur l'inspection visuelle des hydrogrammes et/ou des résidus du modèle. Ce type de critère est très couramment utilisé pour comparer les performances des modèles (voir la revue de Seibert [2001] et Perrin et al. [2005]) et sont très utiles pour comparer deux simulations une à une. Néanmoins, ce type de critère d'évaluation présente deux inconvénients majeurs. D'une part, la comparaison de graphiques (hydrogrammes simulés, QQ plots, distribution des débits simulés, etc.) est très subjective. Sur la Figure 3-1, on peut d'ailleurs se demander comment classer les deux modèles testés ? Comme l'a montré Houghton-Carr [1999], deux personnes qui ont des expériences et des sensibilités différentes peuvent classer différemment des modèles sur le même cas d'étude.

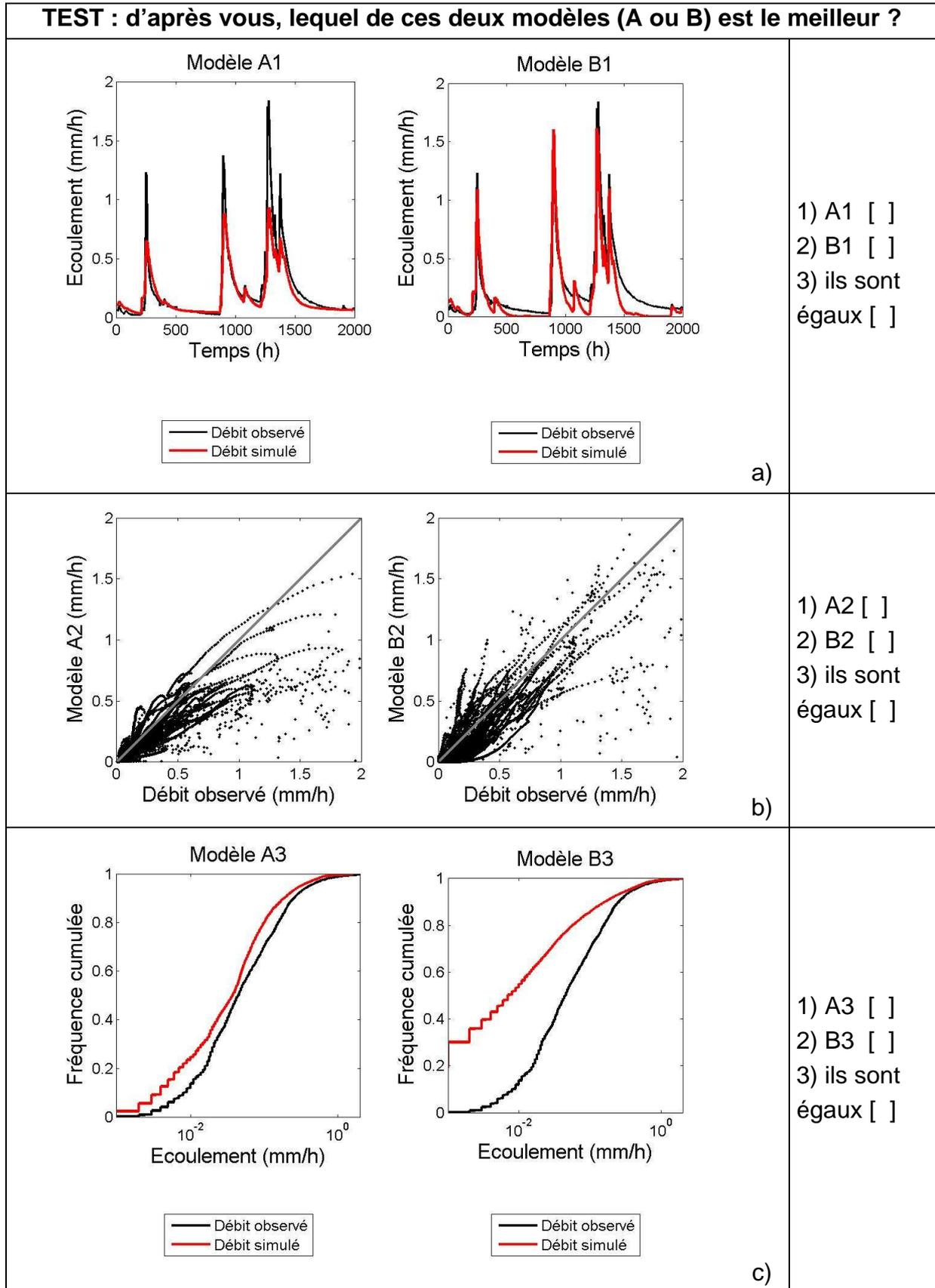


Figure 3-1 : Test d'évaluation des performances de deux modèles (A et B), à partir de (a) hydrogrammes de crues, (b) QQ-plots, (c) distribution cumulée des débits (réponse : $B1 > A1$; $A2 = B2$; $B3 > A3$, d'après le critère numérique $NS(Q)$).

D'autre part, vu le nombre de modèles testés et le nombre de bassins versants de cette étude, il est impossible d'évaluer et de comparer facilement les performances de nombreux modèles par le traçé des hydrogrammes, mais cela est rendu possible par l'utilisation de critères quantitatifs.

Les techniques d'évaluation des performances des modèles pluie-débit à l'aide de critères numériques quantitatifs se sont développées en même temps que les modèles de simulation. Les premiers critères quantitatifs et les plus utilisés ont été ceux basés sur le coefficient de régression linéaire. Mais, au cours des expériences de comparaison de modèles menées par l'OMM [WMO, 1975, 1986, 1992], le besoin de développer un critère d'évaluation. Le critère recherché ne doit pas seulement permettre la comparaison de plusieurs modèles sur un même bassin, mais également sur plusieurs bassins aux caractéristiques différentes. Ainsi, le besoin d'un critère adimensionnel d'évaluation des erreurs des modèles a été clairement exprimé. Ce critère peut être obtenu en comparant les performances du modèle à une référence. Cette référence peut être une grandeur statistique de la variable observée (e. g., sa déviation standard) ou une autre mesure de performance obtenue par un autre modèle [Weglarczyk, 1998; Seibert, 2001; Perrin et al., 2005]. Cette approche a été suivie dans les travaux pionniers de Nash et Sutcliffe [1970].

3.4.2 Le critère de NS : avantages et inconvénients

Dans leur article de référence Nash et Sutcliffe [1970] discutent du calage et de l'évaluation des modèles pluie-débit. Leur méthode pour construire un index de concordance entre les débits observés et les débits simulés est résumée ici. Ils ont commencé par la construction d'un index basé sur le modèle d'erreur quadratique, suggéré par la régression linéaire :

$$SE = \sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2 \quad [3.2]$$

où, SE est la somme des erreurs quadratiques du modèle, $Q_{obs,i}$ et $Q_{sim,i}$ sont respectivement le débit observé et simulé au pas de temps i et N est le nombre de pas de temps. SE est analogue à la variance résiduelle d'une régression linéaire. Cependant, il est difficile de comparer de manière absolue les valeurs de SE d'un même modèle pour des périodes différentes car SE est généralement une fonction croissante de la valeur des débits*. Ainsi, pour que ce critère de performance soit comparable d'une période à une autre, ou d'un bassin à un autre, il est important de le normer. Dans la formulation de NS, SE est normée par la variance V du débit observé (ce qui revient à comparer le modèle testé au modèle $Q = \overline{Q_{obs}}$) :

* Qu'on le veuille ou non, les modèles se trompent d'autant plus que les débits sont forts !

$$V = \sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2 \quad [3.3]$$

où, $\overline{Q_{obs}}$ est la moyenne des débits observés.

Ainsi, le critère de NS est défini comme la part de variance des débits expliquée par le modèle :

$$NS = 1 - \frac{SE}{V} \quad [3.4]$$

Le critère de NS varie dans l'intervalle $]-\infty, 1]$. Il estime l'amélioration obtenue par le modèle testé, par rapport au modèle de référence (dans ce cas $Q = \overline{Q_{obs}}$). Ainsi, si le critère de NS est négatif, le modèle utilisé est plus mauvais que le modèle de référence, si le critère est positif il est meilleur. Une valeur de zéro indique que les deux modèles ont les mêmes performances. Une valeur de 1 indique que le modèle utilisé est parfait. La valeur de $\overline{Q_{obs}}$ est généralement déterminée sur la période de calage ou sur l'ensemble de la chronique. Ce choix est important, car V est très sensible à la valeur de $\overline{Q_{obs}}$. En toute rigueur, sur une période de validation, on devrait utiliser la valeur $\overline{Q_{obs}}$ déterminée sur la période de calage, puisque c'est le paramètre du modèle de référence qui a servi à déterminer le jeu de paramètres sur la période de calage.

3.4.3 Le problème de la formulation du critère pour l'évaluation extensive des performances des modèles hydrologiques

Comme nous l'avons rappelé précédemment, le problème de l'évaluation extensive des performances des modèles hydrologiques est difficile, car il dépend en fait des besoins des utilisateurs potentiels. Si on s'intéresse à l'évaluation comparative des performances de modèles hydrologiques, le cas le plus simple est celui de la comparaison des performances de deux modèles sur un bassin versant. Dans ce cas, il y a de nombreuses manières de comparer leurs performances (hydrogrammes, critères numériques, etc.). En revanche, quand le nombre de modèles et le nombre de bassins versants augmentent, la classification des modèles devient beaucoup plus complexe. En prenant n modèles et m bassins versants, cette comparaison revient à traiter n listes de m critères de NS (évoluer dans un espace à $n \times m$ dimensions). Pour synthétiser cette importante quantité d'informations, on utilise couramment :

- Les distributions de différents critères obtenues pour chaque modèle, sur un échantillon de bassins versants (Figure 3-2) ;

- quelques indicateurs de performances (valeurs moyennes, médianes, et quantiles) de différents critères obtenus pour chaque modèle, sur un échantillon de bassins versants. Il s'agit en fait de statistiques résumant les distributions de résultats mentionnées précédemment.

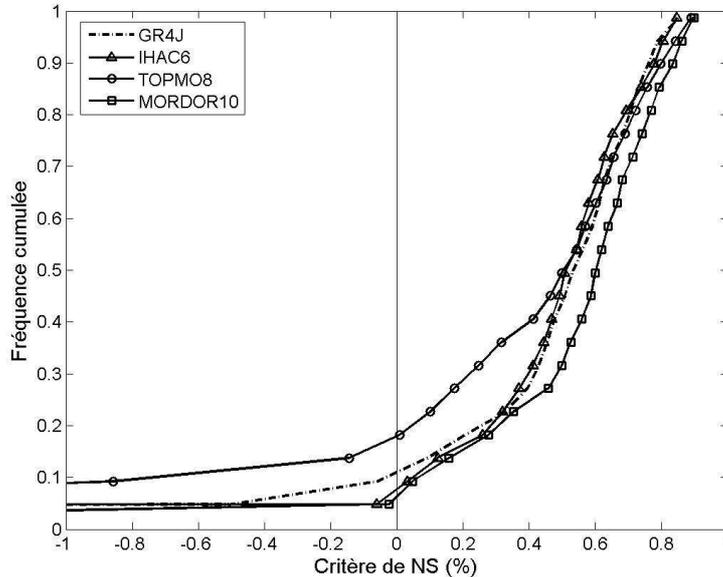


Figure 3-2 : Exemple de distribution des performances obtenues au pas de temps horaire avec quatre modèles, sur l'échantillon de 313 bassins versants

Ces informations sont très intéressantes, mais elles ne sont pas toujours cohérentes. D'une part, lorsqu'on utilise les distributions de performances de plusieurs modèles, il devient rapidement impossible de classer différents modèles, étant donné que les distributions se croisent (comment interpréter la Figure 3-2 ?).

D'autre part, comme le montre le Tableau 3-4, le classement de plusieurs modèles est également très difficile et dépend généralement de l'indicateur de performance retenu. Ce tableau montre par ailleurs que le critère de NS peut prendre des valeurs très négatives et que l'écart-type des résultats peut être également extrêmement grand. Par contre, quand on compare la valeur moyenne du critère de NS à sa médiane, on se rend compte que cette très basse moyenne et ce fort écart-type sont causés par seulement quelques bassins versants dont les critères de NS sont très bas. Ces quelques bassins biaisent artificiellement les performances moyennes des modèles et les conclusions de la comparaison. En pratique, il est courant que des bassins versants, pour lesquels les performances sont très mauvaises, soient éliminés de l'échantillon testé, ce qui améliore artificiellement les performances des modèles. Ces mauvais résultats sont obtenus pour de nombreuses raisons, par exemple manque de pertinence du modèle, erreur dans les données, échec de la méthode d'optimisation, etc...

Néanmoins, comme nous l'avons dit précédemment (section 2.5.5), dans un souci d'objectivité, nous préférons ne pas filtrer ces bassins et nous ne rejetons pas ces bassins, à cause des faibles performances de certains modèles. Au contraire, nous préférons proposer une méthode simple pour faire face à ces valeurs de critère de NS très négatives.

Tableau 3-4 : valeurs de critères (valeur minimale, maximale, moyenne, médiane, quantile et écart-type) en utilisant le critère de NS et le C_{2M} sur l'ensemble des bassins versants. Classement des quatre modèles en fonction de ce critère (entre crochets).

Indicateur de performance	GR4J	IHAC6	TOPMO8	MORDOR10
NS minimum	-6050 [3]	-369 [2]	-180.2 [1]	-30103 [4]
NS moyen	6 [2]	44.7 [1]	7.1 [3]	-81.6 [4]
NS maximum	87.6 [3]	87.1 [4]	91.8 [2]	92.1 [1]
NS quantiles				
0.1	-3.8 [3]	5.6 [2]	-56.6 [4]	6.8 [1]
0.3	41.8 [2]	40.2 [3]	22.7 [4]	48.8 [1]
0.5	53.3 [2]	51.2 [3]	50.8 [4]	60.5 [1]
0.9	76.0 [4]	78.0 [3]	79.9 [2]	83.7 [1]
% de NS				
< 0 %	10 % [3]	6 % [1]	17 % [4]	6 % [1]
< -100 %	2 % [2]	< 1 % [1]	8% [4]	2 % [2]
Ecart-type du NS	417.0 [3]	39.4 [1]	164.2 [2]	1794.3 [4]
C _{2M} moyen	32.7 [3]	33.8 [2]	26.4 [4]	39.5 [1]
C _{2M} quantiles				
0.1	-1.8 [3]	2.9 [2]	-22.0 [4]	3.5 [1]
0.3	26.4 [2]	25.1 [3]	12.8 [4]	32.2 [1]
0.5	36.3 [2]	34.4 [3]	34.0 [4]	43.3 [1]
0.9	61.2 [4]	63.9 [3]	66.4 [2]	71.9 [1]
Ecart-type du C _{2M}	27.2 [2]	22.5 [1]	36.2 [4]	27.9 [3]

Le caractère dissymétrique de la distribution des critères de NS pose problème dans le calcul de certaines statistiques qui nécessitent une distribution de valeurs symétrique. Ainsi, des tests de comparaison de moyennes (comme le test de Student) sont impossibles avec la formulation du critère de NS, telle qu'elle est. En fait, ces problèmes de valeurs de du critère de NS très négatives et de distribution dissymétriques proviennent de l'absence de borne négative pour le critère de NS, qui peut varier dans l'intervalle $]-\infty, 1]$. Pour éviter ce problème et pouvoir obtenir des

statistiques plus significatives sur un échantillon important de bassins versants, nous proposons de modifier la formulation du critère de NS, afin que son domaine de variation soit borné et qu'il varie entre -1 et $+1$.

3.4.4 C_{2M} , une formulation bornée du critère de NS

Pour donner au critère de NS une borne négative et garder la valeur 0 comme même valeur de référence ($Q = \overline{Q_{obs}}$), nous proposons la formulation suivante, que nous avons appelé C_{2M} :

$$C_{2M} = \frac{1 - \frac{SE}{V}}{1 + \frac{SE}{V}} \quad [3.5]$$

où, SE est la somme des erreurs quadratiques du modèle (Eq. [3.2]) et V est la variance des débits observés (Eq. [3.3]).

NS et C_{2M} sont reliés de la manière suivante (Figure 3-3) :

$$NS = \frac{2 \cdot C_{2M}}{1 + C_{2M}} \quad [3.6]$$

$$C_{2M} = \frac{NS}{2 - NS} \quad [3.7]$$

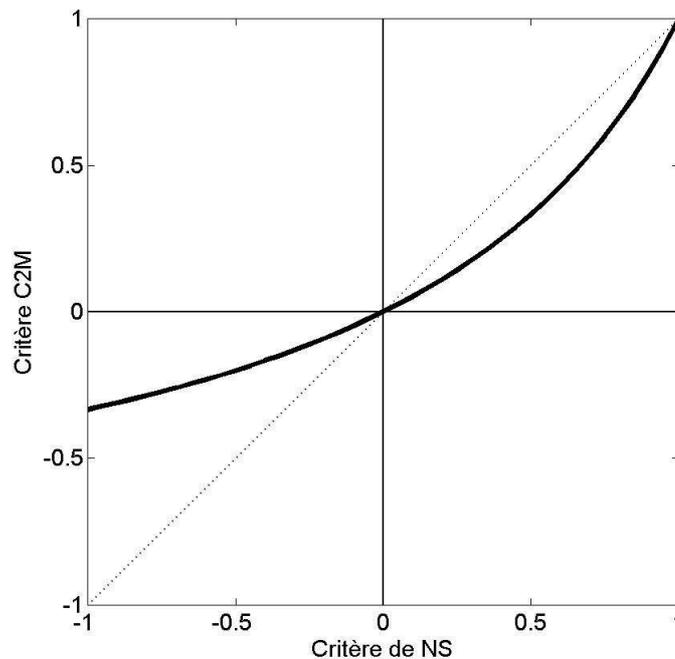


Figure 3-3 : Relation entre le critère de NS et le critère C_{2M}

L'intérêt du critère C_{2M} est multiple :

1. Le domaine de variation du C_{2M} est borné, entre -1 et $+1$;
2. Le signe et le zéro du C_{2M} coïncident avec le signe et zéro du critère de NS ;
3. Dans sa partie positive, le critère C_{2M} est sensiblement inférieur au critère de NS. Ce critère donne ainsi une image moins optimiste, de la qualité des simulations du modèle, que le critère de NS. Cela va dans le sens du point de vue de *Garrick et al.* [1978], qui faisaient remarquer que les valeurs du critère de NS étaient généralement trop fortes^{*1} ;
4. Grâce à des distributions de résultats moins asymétriques, le C_{2M} permet de calculer des statistiques plus significatives, dont l'interprétation est plus facile sur un grand échantillon de bassins versants ;
5. Par rapport à l'utilisation d'un quantile de la distribution des critères de NS (par exemple la médiane), qui ne donne qu'une vue partielle de l'ensemble des performances, le calcul de la moyenne des critères C_{2M} repose sur l'ensemble des résultats, tout en restant facilement interprétable par rapport à une moyenne calculée sur le critère de NS.

3.5 Comparaison des performances de modèles hydrologiques en utilisant les critères de NS et C_{2M}

Dans cette partie, les résultats des modèles hydrologiques vont être utilisés pour démontrer (1) l'intérêt d'utiliser des échantillons regroupant un grand nombre de bassins versants, et (2) l'intérêt du critère C_{2M} pour effectuer de telles comparaisons.

3.5.1 Comparaison des performances moyennes de sous-échantillons aléatoires de 10, 30 et 100 bassins versants, en utilisant le critère de NS et le critère C_{2M}

Dans cette partie, nous souhaitons mettre en évidence la grande variabilité des performances moyennes d'un modèle, quand il est testé sur un petit nombre de bassins versants, ce qui peut conduire à différentes appréciations des performances relatives des modèles. Pour démontrer l'intérêt d'utiliser de grands échantillons de bassins versants, nous avons procédé à des tirages aléatoires d'un grand nombre de sous-échantillons de bassins versants, de tailles différentes. Nous avons effectué le tirage de 10 000 échantillons^{*2} différents, de 10, 30 et 100 bassins choisis parmi les 313 bassins versants de l'échantillon total. En utilisant le critère de NS et le critère C_{2M} , nous avons ensuite calculé les performances moyennes de chaque sous-

^{*1} et que les hydrologues sont trop optimistes quant aux performances réelles de leurs modèles !

^{*2} Il est certain que ce nombre élevé d'échantillons introduit une certaine redondance et que moins d'échantillons auraient suffi pour cette démonstration. Oublier le principe de parcimonie ne fait pas de mal de temps en temps...

échantillon pour les différents modèles, que nous avons comparés deux à deux et par rapport à la moyenne des performances obtenues avec les 313 bassins versants.

La Figure 3-5 représente les comparaisons deux à deux des performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10, en utilisant le critère de NS. Sur cette figure, on observe que :

- La performance moyenne des modèles, obtenue sur l'échantillon de 313 bassins versants (symbolisée par une croix), n'a généralement pas de sens (valeurs très basses). Elle est très différente des tendances mises en évidence par les sous-échantillons de bassins versants ;
- La dispersion des performances moyennes des modèles est très grande, mais se réduit néanmoins avec la taille croissante de l'échantillon ;
- Il est toujours possible de trouver des échantillons de bassins versants (même avec 100 bassins versants) pour lesquels n'importe quel modèle est meilleur qu'un autre. Il est ainsi très difficile de classer deux modèles.

La forte variabilité des performances moyennes d'un modèle, pour une taille d'échantillon donnée, est généralement causée par la présence de quelques bassins versants sur lesquels le modèle obtient de très mauvaises performances. Quand la taille de l'échantillon augmente, le poids de ces quelques bassins versants diminue, et donc la variabilité des performances du modèle également.

La Figure 3-4 montre l'évolution de la moyenne et de l'écart-type de la performance moyenne de 10 000 sous-échantillons de bassins versants, en fonction de leur taille (la décroissance de l'écart-type est logiquement proportionnelle à la racine carrée de la taille du sous-échantillon). Cette figure montre que pour les modèles GR4J, TOPMO8 et MORDOR10, la valeur de la performance moyenne n'a pas de sens car elle est beaucoup trop variable et est influencée par seulement quelques bassins versants.

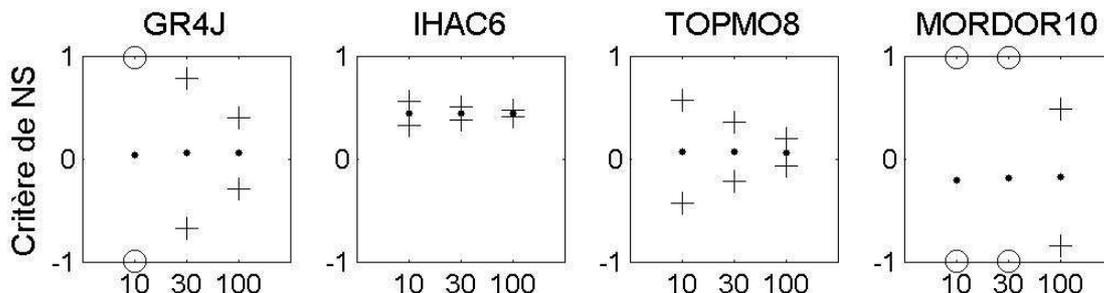


Figure 3-4 : Moyenne (point) et écart-type (plus ; rond quand elle dépasse les bornes $-1 / 1$) des performances moyennes des 10000 sous-échantillons aléatoires de 10, 30 et 100 bassins versants, en utilisant le critère de NS.

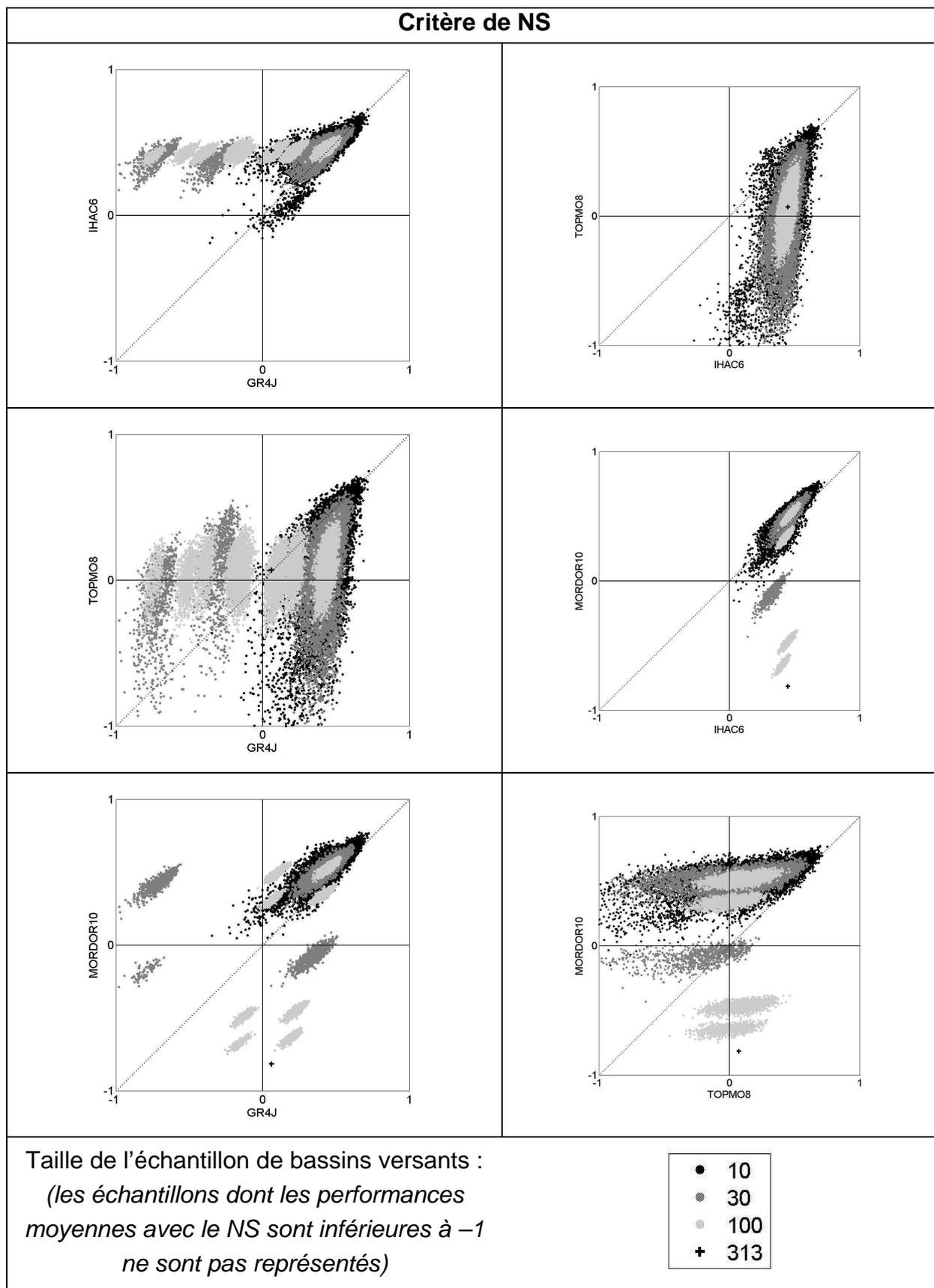


Figure 3-5 : Comparaison des performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10 pour 10000 sous-échantillons aléatoires différents de 10, 30 et 100 bassins versants, en utilisant le critère de NS.

La Figure 3-6 représente les comparaisons deux à deux des performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10, en utilisant le critère C_{2M} . Par rapport à la Figure 3-5, on se rend compte que le critère C_{2M} facilite considérablement la comparaison des performances de différents modèles :

- La performance moyenne des modèles, obtenue sur l'échantillon de 313 bassins versants (croix) est bien représentée par les tendances mises en évidence par les sous-échantillons de bassins versants,
- La dispersion des performances moyennes des modèles est beaucoup plus limitée que celle obtenue en utilisant le critère de NS, ce qui permet de comparer plus facilement les performances de deux modèles,
- Peu de bassins versants sont nécessaires pour différencier les performances de deux modèles. Par exemple, pour comparer GR4J et TOPMO8, 100 bassins versants sont suffisants (aucun échantillon de 100 bassins versants ne traverse la bissectrice), pour comparer TOPMO8 et MORDOR10, seulement 30 bassins versants sont nécessaires.

De plus, on peut se demander si l'allure de ces nuages de points de forme ellipsoïdale (comète) ne peut pas nous renseigner sur le comportement d'un modèle et sur sa complémentarité avec un autre modèle ? On peut distinguer deux cas de figure : soit l'axe d'allongement du nuage est grossièrement parallèle à la bissectrice, soit l'axe d'allongement du nuage en diffère très sensiblement. Lorsque l'axe est quasi-parallèle à la bissectrice, cela signifie que les deux modèles comparés ont des performances uniformément différentes sur l'échantillon. Lorsque l'axe est incliné sur la première bissectrice, cela signifie qu'il y a en plus une différence de robustesse entre les deux modèles : sur certains groupes de bassins versants, un modèle est sensiblement meilleur qu'un autre modèle et inversement. D'autre part, la largeur des nuages nous renseigne également sur la plus ou moins grande indépendance entre les performances des deux modèles comparés.

La Figure 3-7 montre l'évolution de l'écart-type de la performance moyenne de 10000 sous-échantillons de bassins versants, en fonction de leur taille. Cette figure montre tout d'abord qu'avec le critère C_{2M} les modèles obtiennent des niveaux de performances et un écart-type de leurs performances moins sensibles aux quelques bassins versants pour lesquels les performances sont mauvaises. La comparaison des écart-types des performances entre plusieurs modèles permet également de mettre en évidence la plus faible robustesse de certains modèles (TOPMO8, par exemple).

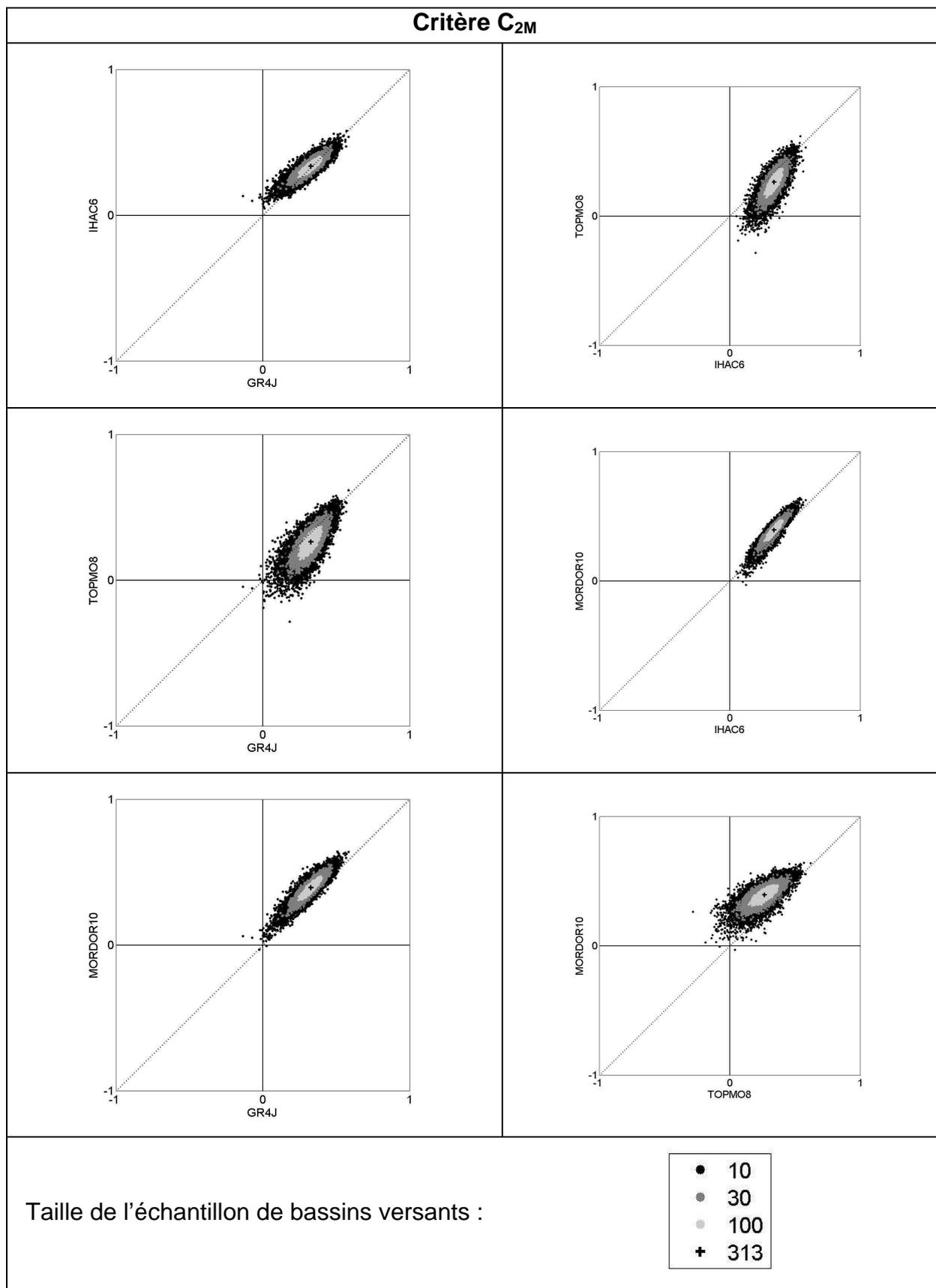


Figure 3-6 : Comparaison des performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10 pour 10000 sous-échantillons aléatoires différents de 10, 30 et 100 bassins versants, en utilisant le critère C_{2M} .

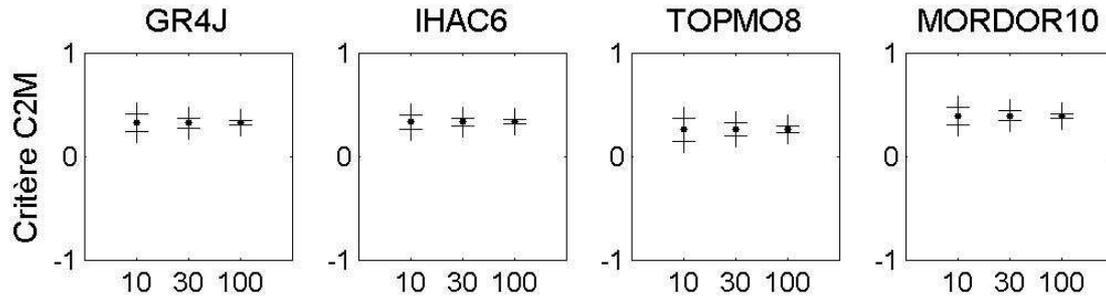


Figure 3-7 : Moyenne (point) et écart-type (plus) des performances moyennes des 10000 sous-échantillons aléatoires de 10, 30 et 100 bassins versants, en utilisant le critère C_{2M} .

Les Figure 3-5 à Figure 3-7 ont mis en évidence la diminution de la variabilité des performances des modèles sur des échantillons de taille croissante et l'intérêt du critère C_{2M} , pour réduire le poids des bassins versants où certains modèles sont très mauvais. Ce que les figures montrent moins bien, c'est l'évolution de la densité des performances en fonction de la taille de l'échantillon, c'est-à-dire la répartition des performances des modèles, à l'intérieur d'une région de l'espace des performances (l'espace des performances de deux modèles représente les performances du modèle B en fonction de celles du modèle A). Pour répondre à cette question, la Figure 3-8 représente la densité des performances de deux modèles (IHAC6 et TOPMO8) en fonction de la taille des sous-échantillons aléatoires. La densité est représentée par le pourcentage des sous-échantillons dont les performances des modèles A et B se situent dans une certaine unité de l'espace (un carré de 5 x 5 points de NS ou C_{2M}).

Cette figure montre tout d'abord qu'avec le critère de NS, la dispersion des performances est très grande et qu'avec des échantillons de 10 ou 30 bassins versants, la densité des performances est quasiment la même dans tout l'espace. A partir d'une centaine de bassins versants, on commence à distinguer une zone où les performances sont plus concentrées. Néanmoins, cette zone est très étendue.

Avec le critère C_{2M} la dispersion des performances est moins grande et les densités sont plus fortes. A partir d'échantillons de seulement 10 bassins versants, on peut distinguer une zone où les performances sont plus concentrées. A partir de 30 et avec 100 bassins versants, cette zone est peu étendue et très bien définie dans l'espace. Ainsi, cette figure illustre d'une autre manière l'intérêt du critère C_{2M} par rapport au critère de NS et l'intérêt des grands échantillons de bassins versants, pour réduire la variabilité de la différence de performances entre deux modèles et mettre en évidence la zone de performance la plus dense (la plus probable). Elle illustre également que sur de grands échantillons, les différences de performances entre

modèles peuvent être exagérées par un petit nombre de bassins versants, sur lesquels un modèle a de très mauvaises performances.

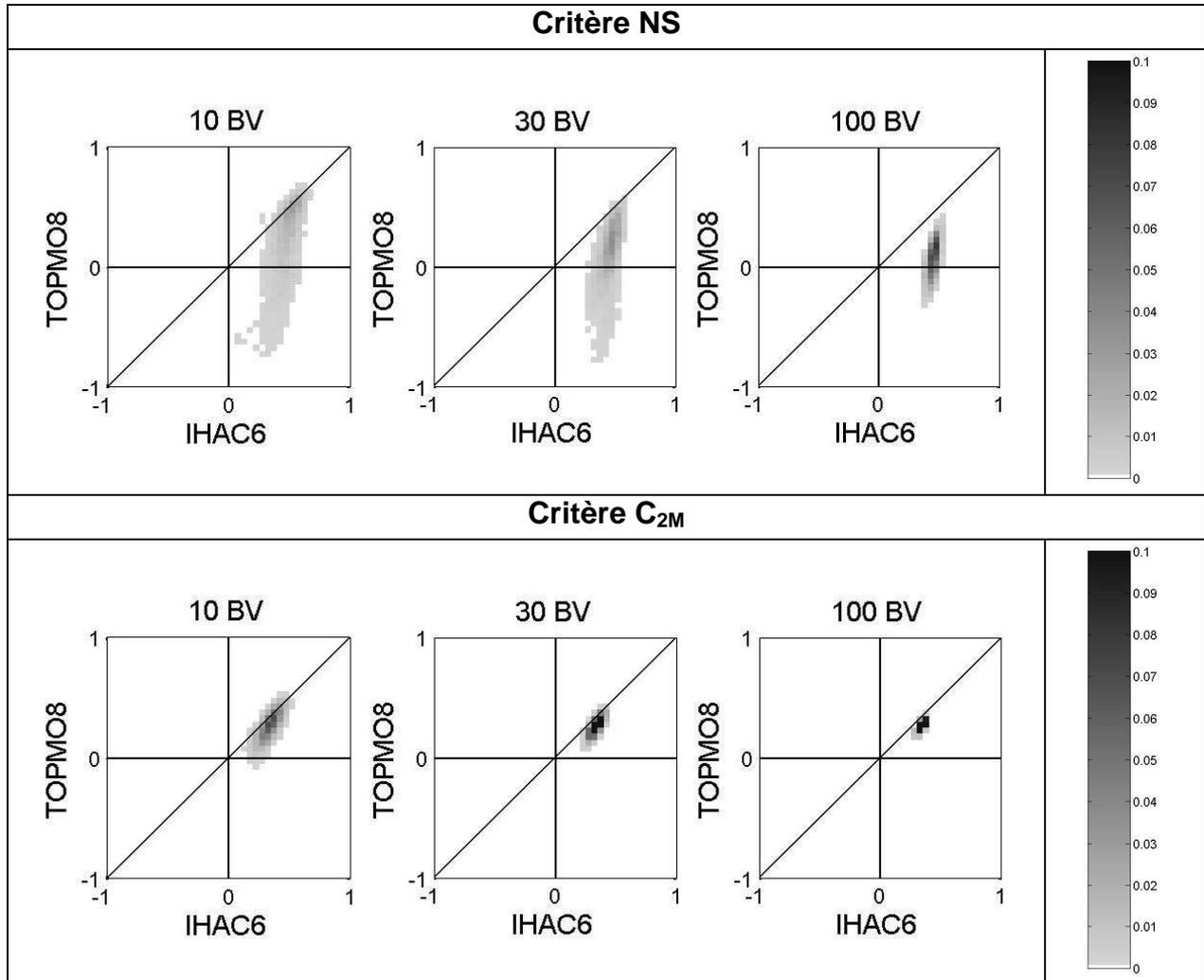


Figure 3-8 : Évolution de la densité des performances de IHAC6 et TOPMO8, en fonction de la taille des sous-échantillons de bassins versants (10000 tirages aléatoires différents). La couleur blanche correspond à une densité nulle et la couleur noire à une densité supérieure à 10 % de l'échantillon total.

Une autre manière d'utiliser cette procédure de sous échantillonnage de l'échantillon pour comparer deux modèles est proposée en Annexe 3. Cette procédure permet d'exprimer la différence de performances entre deux modèles, non plus de manière absolue, mais plutôt de manière probabiliste.

3.5.2 Combien de bassins versants sont nécessaires pour différencier significativement les performances de deux modèles hydrologiques ?

Dans cette partie, un test statistique (test de Student) est utilisé pour détecter si la différence de performances entre deux modèles est significative ou non. Ce test est complémentaire de la simple comparaison des performances moyennes de deux

modèles. Il permet de prendre en compte la grande variabilité des performances d'un modèle sur un échantillon de bassins versants différents et la taille de cet échantillon. Ce test de Student fait l'hypothèse de la distribution gaussienne des résultats (critères de performance), ce qui est à la limite acceptable pour le critère C_{2M} (Figure 3-9), mais ne l'est pas du tout pour le critère de NS qui n'est pas borné. Comme nous l'avons vu précédemment, sur l'échantillon de 313 bassins versants, la différence de performances entre GR4J et IHAC6 est d'environ 1 point, entre GR4J et MORDOR10, IHAC6 et MORDOR10, TOPMO8 et GR4J, TOPMO8 et IHAC6 d'environ 7 à 8 points, entre TOPMO8 et MORDOR10 d'environ 13 points. Le Tableau 3-5 résume les résultats des tests de Student effectués sur les résultats des modèles pris deux à deux, sur l'échantillon de 313 bassins versants. Seule la différence de moyenne entre GR4J et IHAC6 n'est pas statistiquement significative. Pour les autres modèles comparés deux à deux, cette différence est statistiquement significative.

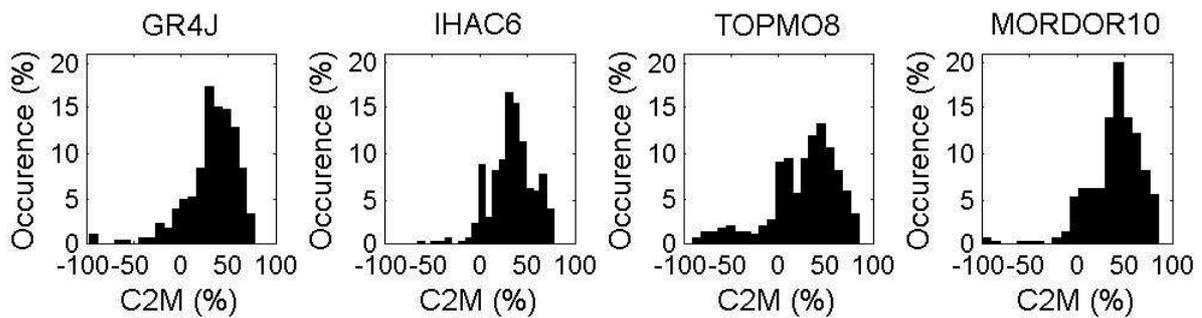


Figure 3-9 : Distributions des critères C_{2M} des modèles GR4J, IHAC6, TOPMO8 et MORDOR10, testés sur l'échantillon de 313 bassins versants.

Tableau 3-5 : Comparaison de la performance moyenne de deux modèles, en suivant un test de Student à un seuil de 5% (hypothèse H_0 : les performances moyennes ne sont pas significativement différentes).

313 bassins versants	IHAC6	TOPMO8	MORDOR10
GR4J	GR4J \approx IHAC6	TOPMO8 < GR4J	GR4J < MORDOR10
IHAC6	/	TOPMO8 < IHAC6	IHAC6 < MORDOR10
TOPMO8	/	/	TOPMO8 < MORDOR10

Le résultat d'un test de Student dépend de la différence de performance entre deux modèles, mais également l'écart-type de ces différences. Par ailleurs, comme on l'a vu précédemment, cet écart-type est très influencé par la taille de l'échantillon de bassins versants. Ainsi, on peut se demander quelle est l'influence de la taille de l'échantillon de bassins versants sur la probabilité de rejeter ou d'accepter H_0 . Pour cela, nous avons procédé à des tirages aléatoires de 10000 sous-échantillons,

composés de 10, 30, 100 et 200 bassins versants. Ensuite, pour chaque sous-échantillon, nous avons effectué un test de Student. Les résultats, synthétisés au Tableau 3-6 et à la Figure 3-10, présentent la probabilité de rejeter le test de Student, pour les modèles comparés deux à deux, en fonction de la taille du sous-échantillon de bassins versants. Cette probabilité représente le nombre de cas où le test de Student est rejeté, sur les 10000 tests effectués.

Tableau 3-6 : Comparaison des performances moyennes de deux modèles. Probabilité de rejeter H_0 en fonction de la taille de l'échantillon de bassins versants (tirage aléatoire de 10000 échantillons différents).

Nombre de bassins versants	Différence de C_{2M} moyen (%)	10	30	100	200	313
GR4J / IHAC6	- 1.1	0.00	0.00	0.00	0.00	0.00
GR4J / TOPMO8	+ 6.3	0.00	0.00	0.09	0.51	1.00
GR4J / MORDOR10	- 6.7	0.00	0.01	0.25	0.99	1.00
IHAC6 / TOPMO8	+ 7.3	0.00	0.02	0.29	0.95	1.00
IHAC6 / MORDOR10	-5.6	0.00	0.01	0.17	0.84	1.00
TOPMO8 / MORDOR10	+ 13	0.04	0.25	0.97	1.00	1.00

Ce tableau montre tout d'abord qu'il n'est pas possible de distinguer statistiquement les performances de modèles dont la différence de performance moyenne est trop faible (GR4J et IHAC6). Ensuite, ce tableau montre également que, quand la différence de performances est de l'ordre de 6 à 7 points, même une centaine de bassins versants n'est pas suffisante pour distinguer clairement les performances de deux modèles. La probabilité de rejeter H_0 , alors que H_0 est faux (conclure que les modèles ne sont pas statistiquement différents), est seulement entre 9 % et 29 %. Pour ce niveau d'écart de performances, il faut un échantillon d'au moins 200 bassins versants pour pouvoir distinguer les performances de deux modèles. Néanmoins, il faut noter le cas de GR4J et TOPMO8, pour lequel la probabilité de rejeter H_0 n'est que de 51%, à cause du fort écart-type de leurs performances. Enfin, lorsque la différence de performance est grande (TOPMO8 et MORDOR10), une centaine de bassins versants semble suffisants.

Le Tableau 3-6 et la Figure 3-10 montrent que, étant donnée la variabilité des performances des modèles hydrologiques, lorsque la différence de performance est moyenne (de l'ordre de 5 points), il est difficile de conclure qu'un modèle est meilleur qu'un autre, même avec un échantillon d'une centaine de bassins versants (ce qui est déjà rarement utilisé). Néanmoins, on peut essayer d'explorer cette probabilité de rejet de H_0 , en fonction de la différence de performance moyenne entre deux modèles, de l'écart-type de leurs résultats et de la taille de l'échantillon de bassins

versants. Par exemple, pour chaque comparaison de deux modèles, on peut reprendre l'équation [3.1] et fixer la différence de performance, l'écart-type de leurs performances et rechercher la taille minimale de l'échantillon qu'il faudrait pour rejeter H_0 .

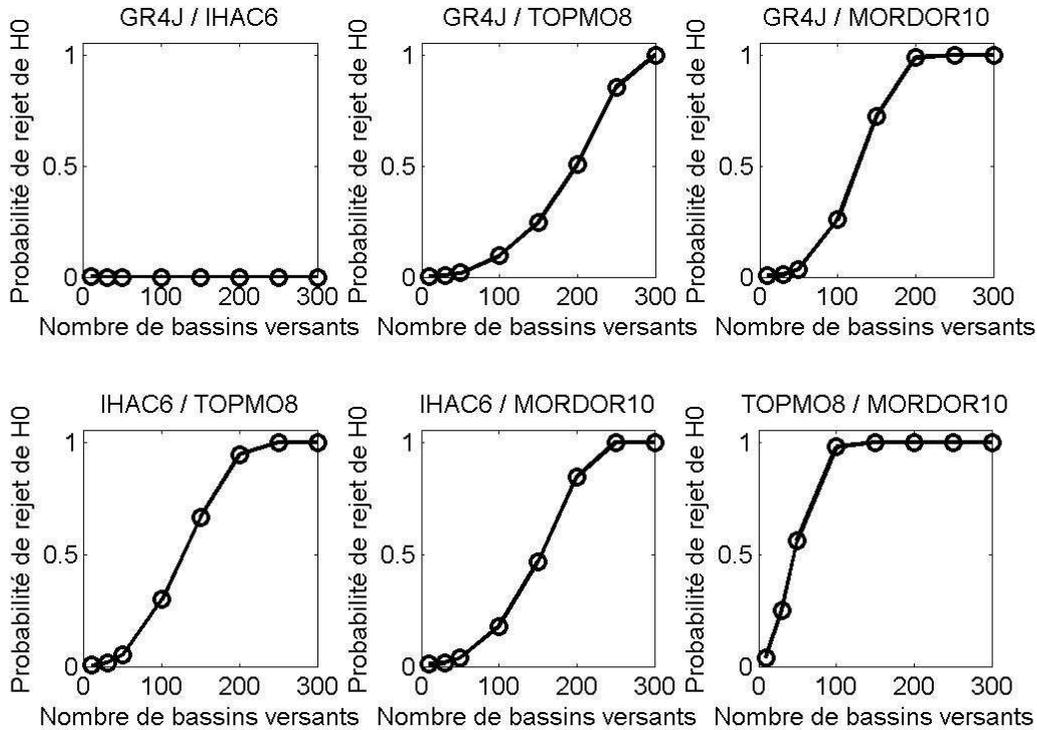


Figure 3-10 : Probabilité de rejet de H_0 en fonction de la taille de l'échantillon de bassins versants (10000 tirages aléatoires de sous-échantillons).

Tableau 3-7 : Taille minimale d'échantillon nécessaire pour que la différence de performances entre deux modèles soit significative, en suivant un test de Student avec un seuil à 5%.

	IHAC6	TOPMO8	MORDOR10
GR4J	∞	200	150
IHAC6	/	150	200
TOPMO8	/	/	50

Le Tableau 3-7 montre la taille minimale qu'il faudrait pour distinguer les performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10 comparés deux à deux. Ces résultats corroborent les tendances des résultats précédents (Figure 3-10 et Tableau 3-7) et confirment que, plus la différence de performances entre deux modèles est petite et l'écart-type des performances d'un modèle est grand, plus il faut un grand nombre de bassins versants pour distinguer leurs performances.

Dans ce cadre théorique, on peut reprendre l'équation [3.1], en fixant la différence de performance moyenne, l'écart-type des performances d'un modèle et rechercher la

taille minimale d'échantillon qu'il faudrait pour rejeter l'hypothèse H_0 . De la même manière, le Tableau 3-8 donne une idée de la taille de l'échantillon qu'il faudrait pour différencier deux modèles, dont l'écart de performances est entre 2.5 et 10 points. Ce tableau montre que, pour une différence de performance donnée, la taille minimale de l'échantillon nécessaire est extrêmement dépendante de l'écart-type des performances des modèles, c'est-à-dire leur robustesse. Par exemple, si la différence de performance est de 10 points (ce qui est énorme !), la taille minimale de l'échantillon nécessaire peut varier entre 35 et 100 bassins versants (pour des valeurs usuelles d'écart-type de performance). Ce test a également été effectué pour des valeurs d'écart-type plus faibles (15 et 20 points), qui correspondent au cas hypothétique où les bassins versants où les modèles ne marchent pas auraient été éliminés*. Avec ces quatre modèles et l'échantillon de 313 bassins versants, si on élimine les bassins pour lesquels les C2M sont négatifs, les écart-types varient entre 16 et 21 points. Ces résultats montrent néanmoins que le fait de filtrer arbitrairement l'échantillon de bassins versants (et d'améliorer artificiellement le niveau de performance des modèles) ne facilite pas la comparaison des performances de deux modèles.

Tableau 3-8: Nombre minimal de bassins versants nécessaires pour différencier les performances de deux modèles, basé sur un test de Student avec un seuil à 5%.

Différence de performance moyenne	Ecart-type de la performance				
	15	20	25	30	35
2.5	300	∞	∞	∞	∞
5	70	125	200	300	∞
7.5	35	70	100	125	175
10	20	35	50	70	100

D'un point de vue pratique, ce test montre que, quand la différence de performance entre deux modèles est faible (de 5 à 10 points), il est généralement impossible de conclure significativement qu'un modèle est meilleur qu'un autre avec un petit échantillon de bassins versants, comme c'est fait généralement.

* Ces valeurs d'écart-type des performances des modèles pluie-débit correspondent également au cas du pas de temps journalier, pour lequel les modèles sont légèrement plus robustes. Des valeurs d'écart-type comprises entre 15 et 20 points ont été trouvées sur un échantillon de 308 bassins versants et pour quatre modèles testés au pas de temps journalier (thèse de *Oudin* [2004]).

3.6 Synthèse

Nous avons vu dans ce chapitre (1) l'intérêt d'utiliser de grands échantillons de bassins versants pour comparer les performances des modèles pluie-débit, (2) que la formulation du critère de NS ne permet pas de tirer des conclusions significatives de ces études et (3) qu'une formulation pertinente du critère d'évaluation des performances, le critère C_{2M} , permet de mieux comparer les performances de différents modèles sur un grand échantillon de bassins versants. Nos résultats vont dans le sens des propositions n°6 et n°7, formulées au Chapitre 1. Ils montrent l'intérêt de l'approche comparative, suivant des critères bien choisis, sur de nombreux bassins versants.

Le critère de NS, bien qu'il soit très utilisé en hydrologie, a un inconvénient : il n'est pas borné négativement et a donc une distribution fortement asymétrique. Par conséquent, le critère de NS ne permet pas de calculer des statistiques significatives pour caractériser les distributions des performances de modèles. En effet, pour de nombreuses raisons, des valeurs très négatives de critère de NS sont obtenues sur quelques bassins versants*. Le critère C_{2M} permet de donner une formulation bornée du critère de NS, variant dans l'intervalle $]-1, +1]$. La valeur du zéro et le signe du critère C_{2M} , sont les mêmes que ceux du critère de NS. En empêchant les trop fortes valeurs négatives dans les cas où un modèle ne marche pas, le critère C_{2M} permet de donner à ces bassins versants un poids raisonnable dans le calcul de statistiques caractérisant la distribution des performances d'un modèle, notamment le calcul de la moyenne.

Nous avons utilisé une stratégie de sous-échantillonnage des performances de quatre modèles pour évaluer leurs performances respectives sur un grand nombre de sous-échantillons générés aléatoirement. Cette procédure nous a permis de clairement démontrer l'intérêt du critère C_{2M} par rapport au critère de NS et l'intérêt des larges échantillons de bassins versants pour comparer de manière plus significative les performances de deux modèles. Cette stratégie a montré qu'un échantillon d'une centaine de bassins versants permettait de différencier distinctement les performances de deux modèles.

Enfin, un test statistique de comparaison de deux moyennes a montré qu'avec le niveau de robustesse des modèles testés, il faut au moins entre 100 et 300 bassins versants pour différencier de manière significative les performances moyennes de

* La pratique courante veut que les bassins versants sur lesquels les modèles ont des critères négatifs soient éliminés. Ce qui est inacceptable ! C'est comme si on jugeait la qualité d'un enseignant par la moyenne des notes des élèves et que l'enseignant éliminait les mauvais élèves de sa classe.

deux modèles. Bien que très exigeant, ce test indique que, quand les différences de performances sont inférieures à 5 points, elles sont à peine suffisantes pour différencier deux modèles, même sur un échantillon assez grand.

Chapitre 4

« Que peut-on connaître du monde ? De notre naissance à notre mort, quelle quantité d'espace notre regard peut-il espérer balayer ? Combien de centimètres carrés de la planète Terre nos semelles auront-elles touché ? »

Georges Perec, *Espèces d'espaces*, 1974

CHAPITRE 4 Optimisation automatique des modèles pluie-débit : comparaison exhaustive des performances de méthodes locale et globales.

4.1 Introduction

Nous avons vu au chapitre CHAPITRE 3 qu'avoir un échantillon test d'une taille suffisante et que bien choisir le critère d'évaluation des performances d'un modèle étaient des pré-requis pour mener une comparaison fiable de modèles.

Dans ce chapitre, nous allons étudier l'importance du choix de la méthode d'optimisation pour caler les paramètres des modèles testés.

Les modèles hydrologiques actuels, qu'ils soient fondés physiquement, empiriques ou conceptuels, sont caractérisés par un certain nombre de fonctions mathématiques qui représentent le lien entre pluie et débit. Dans la plupart des cas, ces fonctions contiennent des paramètres à déterminer, qui ne sont généralement pas mesurables à l'échelle du bassin versant. Ainsi, pour être utiles et donner des simulations ou des prévisions de débit de bonne qualité, les modèles hydrologiques doivent être nécessairement calés. Le processus de calage signifie que, d'une manière manuelle ou automatique, la distance entre l'hydrogramme simulé par le modèle et l'hydrogramme observé est minimisée, en cherchant un vecteur de paramètres qui donne la meilleure simulation possible.

Depuis les années 1970, grâce au fort développement des capacités de calcul et des modèles pluie-débit, de nombreux auteurs ont montré la difficulté de l'exercice de calage, en rapportant les nombreux échecs qu'ils ont subis au cours de longues et fastidieuses tentatives de calibration. On peut citer, pour l'exemple, les travaux pionniers de *Johnston et Pilgrim* [1976], qui ont quand même passé deux années sur l'optimisation des huit paramètres d'un modèle (modèle de Boughton) sur un bassin versant, sans pouvoir trouver un vecteur de paramètres optimal ! L'exercice de calage est un exercice d'autant plus difficile que les processus hydrologiques sont caractérisés par un grand nombre de composantes dont les contributions sont mal connues, qu'il existe de fortes incertitudes sur les variables d'entrée et de sortie du système bassin versant et que les modèles hydrologiques sont encore loin de représenter correctement le fonctionnement des bassins versants. De plus, la plupart

des modèles hydrologiques couramment utilisés sont caractérisés par l'hyper-sophistication de leurs structures conceptuelles, un grand nombre de paramètres à optimiser, des interactions et des compensations entre paramètres. Les échecs rencontrés dans les années 1970 dans l'optimisation des modèles ne semblent pas étonnants *a posteriori*. A cette époque, les hydrologues croyaient peut être encore à l'existence du *modèle conceptuel* et d'un vecteur de paramètres unique capable de décrire la physique du fonctionnement des bassins versants, qu'ils avaient postulé *a priori*. Depuis les travaux de *Beven* [1993, 2001a] sur les problèmes d'équifinalité et les recherches actuelles sur la thématique des incertitudes dans la modélisation hydrologique, on peut penser que d'une certaine manière le vecteur optimal n'existe probablement pas, puisque le modèle hydrologique idéal n'existe pas. Ainsi, sa recherche revêt peut être un caractère infondé.

Néanmoins, en dépit de problèmes dans les structures des modèles hydrologiques, la tendance actuelle en hydrologie est au développement et à l'utilisation de stratégies d'optimisation sophistiquées (globales), qui sont réputées pouvoir trouver les meilleurs vecteurs de paramètres possibles, quelle que soit la dimension de l'espace de recherche. Nous pensons cependant que des méthodes d'optimisation simples (mais pas *simplistes*) restent toujours d'actualité. En effet, grâce à leur rapidité, elles restent indispensables dans les travaux comportant un grand nombre de bassins versants [*Perrin*, 2000; *Oudin*, 2004; *Rojas-Serna*, 2005; *Tangara*, 2005], mais elles sont également nécessaires pour calibrer des modèles distribués, comme le soulignent *Lindström* [1997] et *Kuzmin et al.* [2004]. Convaincus que la structure d'un modèle est la pierre d'angle d'un système de simulation/prévision hydrologique*, nous pensons que l'hyper-sophistication des méthodes d'optimisation des modèles hydrologiques s'apparente plutôt à un pansement sur une jambe de bois, tant que les structures des modèles ne seront pas meilleures.

Ainsi, comme nous le proposons, une des voies possibles pour améliorer les performances des modèles hydrologiques est le développement empirique et la comparaison de nombreux modèles, sur un large échantillon de bassins versants. Un des inconvénients de cette approche est néanmoins le temps qu'elle coûte, malgré l'explosion des puissances de calcul. Pour pouvoir réaliser ce développement empirique et cette comparaison de modèles, l'utilisation d'une méthode d'optimisation rapide est absolument indispensable. Il faut donc avoir recours aux méthodes locales. Néanmoins, un des talons d'Achille des méthodes locales est leur manque de fiabilité présumé qui est d'autant plus probable que le nombre de

* également composé de l'algorithme d'optimisation, de méthodes d'estimation des incertitudes, de mise à jour des paramètres du modèles et des variables internes du modèle, d'une stratégie de distribution du modèle, etc.

paramètres à optimiser est grand. De ce fait, on peut se demander si une méthode d'optimisation locale est aussi efficace pour les modèles parcimonieux (ayant peu de paramètres, ≤ 6) que pour les modèles moins parcimonieux (>6 paramètres). Dans le cadre d'une comparaison de modèles hydrologiques ayant des niveaux de complexité différents, on pourrait craindre que le manque de robustesse de la méthode d'optimisation puisse fausser la comparaison, au détriment des modèles complexes. Pour lever ce doute, nous avons donc souhaité comparer les performances de méthodes locales (simples) et de méthodes globales (plus sophistiquées), ces dernières étant reconnues pour leurs performances, quelle que soit la dimension de l'espace de recherche.

L'objectif de ce chapitre est de présenter la comparaison des performances d'une méthode locale (PAP-GR, développée au Cemagref par *Michel* [1989], *Edijatno* [1991]) et de deux méthodes globales (SCE-UA, développée par *Duan et al.* [1992, 1993, 1994] ; GA-EDF, développée par EDF-DTG et basée sur *Wang* [1991, 1997]). Pour cela nous proposons d'étudier les performances en calage et en validation de 4 modèles hydrologiques, dont le nombre de paramètres est croissant (4, 6, 8 et 10), testés sur l'échantillon de 313 bassins versants, en utilisant les trois méthodes d'optimisation. A notre connaissance, ce type de comparaison entre méthodes locales et globales est inédit sur un échantillon aussi large de bassins versants et avec autant de modèles. Seul le travail de *Gan et Biftu* [1996] a exploré cette thématique, en testant trois méthodes d'optimisation, avec quatre modèles et huit bassins versants, sans pouvoir en tirer des conclusions générales comme notre approche le permet.

4.2 Qualité recherchées des stratégies d'optimisation

Les deux qualités majeures que doit posséder une méthode d'optimisation sont :

- 1. l'efficacité** (dans la littérature anglophone, *effectiveness*), c'est-à-dire la capacité à trouver l'optimum de la surface de réponse, ou le meilleur vecteur de paramètre possible ;
- 2. l'efficience** (dans la littérature anglophone, *efficiency*), c'est-à-dire la capacité à trouver un optimum de la surface de réponse rapidement, avec un nombre d'itérations limité (une itération correspond à un calcul de la valeur de la fonction objectif dans l'espace des paramètres sur la chronique de calage).

Une autre qualité, qui est une composante de la première, est la robustesse, c'est-à-dire la capacité à trouver des vecteurs de paramètres robustes, autrement dit, des vecteurs de paramètres généraux et bien adaptés à des conditions hydrologiques différentes de celles de la période de calage. La robustesse d'un vecteur de

paramètre dépend finalement de la méthode d'optimisation de la fonction objectif et des données utilisées pour le calage des paramètres du modèle.

4.3 Stratégies d'optimisation des modèles hydrologiques : un état de l'art

Un algorithme d'optimisation des paramètres d'un modèle hydrologique est défini par trois caractéristiques :

1. **une fonction objectif**, qui mesure l'erreur du modèle, c'est-à-dire la différence entre les débits simulés par le modèle et les débits observés ;
2. **une stratégie de recherche de l'optimum** de la fonction objectif dans l'espace des paramètres ;
3. **un critère d'arrêt**, qui permet de stopper l'algorithme de recherche.

La majeure partie des travaux de recherche a porté sur les stratégies de recherche de l'optimum, mais la réussite d'une optimisation dépend des interactions complexes entre la fonction objectif, la stratégie de recherche, l'information contenue dans les données et la structure du modèle testé. Ce chapitre porte uniquement sur la comparaison de stratégies de recherche de l'optimum, pour une fonction objectif (basée sur la minimisation de l'erreur quadratique du modèle) et différentes structures de modèles.

4.3.1 Optimisation des modèles hydrologiques : un état de l'art

L'optimisation manuelle des paramètres des modèles hydrologiques permet généralement de trouver de bons vecteurs de paramètres et de produire de bonnes simulations de débit. Néanmoins, le calage manuel n'est pas à la portée de tout le monde : elle est compliquée, fastidieuse et la réussite de cet exercice dépend souvent de l'expérience de la personne qui l'entreprend. Le procédé de calage manuel est donc subjectif, difficile à transférer et impossible à mettre en œuvre dans le cadre de la méthodologie adoptée dans cette thèse. En revanche, les stratégies d'optimisation automatiques, qu'elles soient locales ou globales, sont plus rapides, objectives et très faciles à utiliser pour le type de modèles hydrologiques étudiés ici.

Néanmoins, les stratégies automatiques d'optimisation des modèles hydrologiques sont loin d'être parfaites et souffrent, soit d'un manque d'efficacité (difficulté à trouver le 'meilleur' vecteur de paramètres), soit d'un manque d'efficacité (nécessité de tester un très grand nombre de vecteurs de paramètres avant d'en trouver un bon). Ces imperfections proviennent généralement de la complexité de la surface de réponse produite par les modèles hydrologiques. Comme l'ont rapporté de nombreux

auteurs (dont *Johnston et Pilgrim* [1976], *Duan et al.* [1992], *Sorooshian et Gupta* [1995]), les surfaces de réponse des modèles hydrologiques peuvent présenter :

1. **de nombreuses régions d'attraction**, ce qui rend le vecteur optimal de paramètres dépendant du point de départ de l'algorithme d'optimisation ;
2. **des optima locaux mineurs** (petits pics sur la surface de réponse où peut être piégé l'algorithme d'optimisation) ;
3. **une surface de réponse irrégulière** et avec des points singuliers (où la surface de réponse est non-dérivable) ;
4. **des zones d'insensibilité dans la région de l'optimum**, avec des interactions entre les paramètres du modèle ;
5. **des surfaces non convexes** avec de longues vallées.

Ces cinq caractéristiques expliquent que les stratégies d'optimisation locales, qu'elles soient basées sur une méthode de gradient ou pas, ont généralement des difficultés à converger jusqu'à l'optimum de la surface de réponse. Les méthodes locales sont particulièrement dépendantes du point de départ de leur algorithme de recherche, sa distance à l'optimum et la dimension de l'espace dans lequel elles évoluent. Par contre, si le point de départ est judicieusement choisi, les méthodes locales sont particulièrement robustes et efficaces. Par exemple, *Kuzmin et al.* [2004] ont amélioré une méthode de recherche locale en prédéterminant le point de départ d'une méthode locale par l'utilisation de relations de régionalisation donnant des valeurs de vecteurs de paramètres *a priori*. Ces vecteurs de paramètres *a priori* sont moins bons que des vecteurs obtenus par optimisation, mais sont généralement situés dans la région d'attraction de l'optimum de la surface de réponse, ce qui rend particulièrement efficace et robuste la méthode de recherche locale.

Dans l'objectif de corriger les défauts des méthodes locales et de garantir l'indépendance du résultat vis à vis du point de départ de l'algorithme de recherche et des optima locaux, des méthodes d'optimisation globales ont été développées. Ces méthodes visent à explorer la surface de réponse de manière aussi exhaustive que possible. On peut classer ces méthodes selon leur mode d'exploration de l'espace, soit déterministe, soit stochastique. Les stratégies déterministes discrétisent l'espace des paramètres de manière systématique et testent chaque nœud d'un maillage déterminé d'une certaine manière (par exemple, *Exhaustive Gridding* [*Duan et al.*, 1992; *Nascimento*, 1995]). D'un point de vue théorique, ces méthodes sont capables de localiser parfaitement l'optimum d'une surface de réponse, mais pour cela elles requièrent une infinité d'itérations, ce qui les rend ni efficaces ni intéressantes pour la démarche adoptée dans cette thèse. Les stratégies stochastiques explorent l'espace des paramètres de manière aléatoire. Ces stratégies sont également peu efficaces car elles nécessitent un grand nombre

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

d'itérations pour trouver un optimum et parce que leur exploration de l'espace n'est pas guidée par les résultats des itérations précédentes (manque de mémoire).

Tableau 4-1 : Description des principales méthodes d'optimisation citées

Méthode	Abreviation	Stratégie	Type	Référence
Simplex	Simplex	Locale	Directe	<i>Nelder et Mead</i> [1965]
Pas à pas	PAP-GR	Locale	Directe	<i>Michel</i> [1989], <i>Edijatno</i> [1991]
Stepwise Line Search	SLS	Locale	Directe	<i>Kuzmin et al.</i> [2004]
Pattern Search	PS	Locale	Directe	<i>Hooke et Jeeves</i> [1961]
Powell Conjugate direction	Powell	Locale	Gradient	-
Exhaustive Gridding	EG	Globale	Déterministe	<i>Kuczera</i> [1997]
Multistart simplex	MSX	Globale	Multi-départ	<i>Johnston et Pilgrim</i> [1976] <i>Duan et al.</i> [1992]
Uniform Random Search	URS	Globale	Multi-départ	<i>Duan et al.</i> [1992]
Adaptive Random Search combined with Simplex	ARS + Simplex	Globale	Stochastique	<i>Duan et al.</i> [1992]
Simulated Annealing	SA	Globale	Stochastique	<i>Kirkpatrick et al.</i> [1983] <i>Summer et al.</i> [1997] <i>Thyer et al.</i> [1999]
Genetic Algorithm	GA	Globale	Stochastique	<i>Wang</i> [1991, 1997]
Shuffle Complex Evolution algorithm – University of Arizona	SCE-UA	Globale	Stochastique	[<i>Duan et al.</i> [1992, 1993, 1994]

Ainsi, d'autres méthodes basées sur ces approches ont essayé de tirer profit de l'information acquise au cours des itérations. Ces méthodes sont notamment basées sur des concepts d'évolution de populations de systèmes physiques ou naturels (recuit simulé [*Kirkpatrick et al.*, 1983] ; algorithmes génétiques [*Wang*, 1991; *Wang*, 1997]). Enfin, dans les dernières méthodes développées, *Duan et al.* [1992, 1993, 1994] a proposé la méthode du SCE-UA (*Shuffle Complex Evolution Algorithm, University of Arizona*). Cette méthode fait la synthèse des points forts de nombreuses stratégies d'optimisation, en combinant des stratégies de multi-départ,

des algorithmes génétiques, une recherche locale pour affiner la recherche et un échange périodique d'informations au cours de l'optimisation. Parmi toutes les stratégies du Tableau 4-1, la méthode du SCE-UA est la plus connue, la plus utilisée et, a priori, la plus efficace et efficiente des méthodes globales.

4.3.2 Comparaison de méthodes d'optimisation

A notre connaissance, peu d'études ont porté sur la comparaison de méthodes locales et globales. Parmi les études disponibles (Tableau 4-2), il faut distinguer les études portant (1) sur la résolution de fonctions usuelles, (2) sur l'optimisation de modèles hydrologiques testés sur des données synthétiques (sans erreur) et (3) sur l'optimisation de modèles hydrologiques testés sur des données réelles.

La différence entre les études sur données synthétiques ou réelles est que les données synthétiques sont simulées par le modèle à partir d'un jeu de paramètres défini, donc le 'vrai' optimum de la fonction objectif est connu. En revanche, les études sur données réelles permettent de tester les modèles dans des conditions plus réalistes (surface de réponse plus chaotique, modèle non parfait, influence des erreurs dans les données), de tester l'efficacité des algorithmes en calage et de tester la robustesse du vecteur de paramètres trouvé. La plupart des études montrent généralement la supériorité d'une méthode par rapport à une autre en calage, mais le plus important n'est il pas la robustesse des vecteurs trouvés en validation ?*

Parmi les études portant sur des données synthétiques :

- *Duan et al.* [1992] ont comparé les performances de quatre stratégies d'optimisation globales (ARS, ARS/Simplex, MSX et SCE-UA) utilisées pour calibrer les paramètres du modèle SIXPAR (6 paramètres). Il a été prouvé que les méthodes MSX et SCE-UA étaient toutes les deux capables de trouver avec succès la valeur de l'optimum 'vrai' de la surface de réponse et que le SCE-UA était particulièrement plus rapide que le MSX ;
- *Sorooshian et al.* [1993], *Tanakamaru* [1995], *Cooper et al.* [1997], *Franchini et al.* [1998] ont montré que les stratégies MSX et GA donnent de bons résultats lorsqu'elles sont couplées avec des méthodes de recherche locale (pour affiner la terminaison de la recherche) ;
- *Duan et al.* [1992], *Cooper et al.* [1997], *Kuczera* [1997], *Ndiritu et Daniell* [2001] ont montré que la stratégie SCE-UA est plus robuste et efficace que d'autres méthodes : MSX, GA et SA ;

* d'autant plus si on ne croit pas en l'existence d'un optimum de la surface de réponse !

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

- Enfin, *Tanakamaru* [1995], *Cooper et al.* [1997], *Franchini et Galeati* [1997] ont également montré que les méthodes URS, ARS, et GA ne donnent pas des résultats satisfaisants.

Tableau 4-2 : Synthèse d'études portant sur la comparaison de méthodes d'optimisation des modèles hydrologiques

Référence	Algorithme d'optimisation	Modèle(s)	Type de données	Nombre de bassins versants
<i>Duan et al.</i> [1992]	ARS ARS/Simplex MSX SCE-UA	SIXPAR	Synthétique	/
<i>Sorooshian et al.</i> [1993]	MSX SCE-UA	SAC-SMA	Synthétique Réelle	/ 1
<i>Luce et Cundy</i> [1994]	SCE-UA Simplex	1	Réelle	1
<i>Tanakamaru</i> [1995]	2 Powell 2 Simplex SCE-UA GA	TANK	Synthétique	/
<i>Gan et Biftu</i> [1996]	SCE-UA MSX Local Simplex	SMA NAM SMAR XNJ	Réelle	8
<i>Kuczera</i> [1997]	GA SCE-UA MSX	SFB	Synthétique	/
<i>Cooper et al.</i> [1997]	SCE-UA GA SA	TANK	Synthétique	/
<i>Franchini et al.</i> [1998]	11 GA PS	ADM	Synthétique Réelle	/ 2
<i>Ndiritu et Daniell</i> [1999]	4 GA	MOD- HYDROLOG	Réelle	1
<i>Thyer et al.</i> [1999]	SCE-UA SA/Simplex	SFB	Réelle	2
<i>Ndiritu et Daniell</i> [2001]	GA SCE-UA	SIXPAR	Synthétique	/

Parmi les études portant sur données réelles :

- *Sorooshian et al.* [1993] a montré que le SCE-UA était capable de trouver un meilleur vecteur de paramètre en calibration que celui trouvé par le MSX. En revanche, en validation, aucun des vecteurs de paramètres trouvés par les deux méthodes n'est significativement meilleur que l'autre !
- *Gan et Biftu* [1996] ont testé le SCE-UA, MSX et Local Simplex pour optimiser les paramètres de quatre modèles (ayant de 9 à 15 paramètres), testés sur huit bassins versants aux caractéristiques variées. Leurs résultats ont montré que les vecteurs de paramètres trouvés par le SCE-UA n'étaient que très légèrement meilleurs que ceux trouvés par un MSX et un Local Simplex, en calibration et en validation. De plus, ces auteurs ne constatent aucune influence du nombre de paramètres à optimiser sur la robustesse du SCE-UA et du Local Simplex ! La différence moyenne du critère de NS est seulement de 3% pour un modèle à 9 paramètres et de 0.5% pour un modèle à 15 paramètres. Ces ordres de grandeurs de différence de critères de NS sont bien inférieurs à ceux obtenus sur différentes périodes d'un bassin soumis à une importante variabilité climatique.
- *Franchini et Galeati* [1997] ont testé 11 GA et un PS pour optimiser le modèle ADM (11 paramètres), sur deux bassins versants. Ils ont montré que le PS était clairement meilleur que les 11 GA, en dépit du raffinement des GA.
- *Summer et al.* [1997] ont utilisé un SA/Simplex pour calibrer une version modifiée du modèle SFB (quatre à sept paramètres) sur 25 bassins versants australiens. Ils ont montré certaines des difficultés de l'algorithme testé pour optimiser correctement les paramètres du modèle. Ils ont également remarqué que les SA nécessitent d'être finement réglés pour être efficaces et que dans cet espace à seulement six à huit dimensions le processus de refroidissement s'arrête généralement très rapidement.
- *Ndiritu et Daniell* [1999] ont utilisé quatre GA pour optimiser le modèle MORDHYDROLOG (10 paramètres) sur un bassin versant. Ils ont également montré que, même si le GA le plus sophistiqué était capable de donner de meilleurs vecteurs de paramètres en calage, ces vecteurs n'étaient pas nécessairement meilleurs en validation, que ceux trouvés avec des GA plus simples.
- *Thyer et al.* [1999] ont utilisé le SCE-UA et SA/Simplex pour calibrer le modèle SFB (8 paramètres) sur deux bassins versants. Ils ont montré que le SCE-UA était bien meilleur que le SA/Simplex, parce que les méthodes SA testées sont peu efficaces dans des espaces de fortes dimensions.
- Enfin, *Kuzmin et al.* [2004] ont utilisé une méthode locale (SLS) et le SCE-UA pour calibrer le modèle SAC (11 paramètres) sur 20 bassins versants. Ils ont montré que le SLS était capable de trouver de bons vecteurs de paramètres

rapidement (situés dans la région du vecteur de paramètre initial) et que le SCE-UA était capable de trouver des vecteurs de paramètres seulement légèrement meilleurs en calage et donnant de moins bons résultats en validation. Le SCE-UA nécessite cependant de 20 à 300 fois plus d'itérations !

Pour résumer ces résultats, il semble que les méthodes locales sont capables d'identifier des vecteurs de paramètres performants en calage et donnant des résultats également intéressants en validation, quand elles sont comparées à des méthodes globales, testées sur des données réelles. Mais, il semble que les méthodes locales sont plus efficaces et robustes que les méthodes globales quand le point de départ de leur recherche est situé dans la région d'attraction de l'optimum. Néanmoins, cette démonstration n'a pas été clairement réalisée et les différentes études sont trop hétérogènes pour pouvoir généraliser ce résultat.

4.4 Modèles et stratégies d'optimisation

4.4.1 Modèles pluie-débit utilisés

Pour réaliser notre comparaison nous avons utilisé quatre modèles pluie-débit présentés à la 3.3.2 (GR4H, IHAC6, TOPMO8, MORDOR10), le modèle GR4J étant remplacé par le modèle GR4H, développé pendant cette thèse (Chapitre 6).

Dans un premier temps, ces quatre modèles ont été testés au pas de temps horaire, sur nos 313 bassins versants. Dans un deuxième temps, ils ont été également testés sur un sous-échantillon de 60 bassins versants sur des chroniques de calage de 1 an et 4 ans et validés sur les mêmes périodes de validation. Ils ont enfin été testés sur un sous-échantillon de 50 bassins versants au pas de temps journalier.

La deuxième série de tests, effectués sur des chroniques de 4 ans au pas de temps horaire ou au pas de temps journalier, a eu pour objectif d'étudier l'influence de la quantité de données sur l'efficacité des différentes méthodes d'optimisation. Ces tests apportent les mêmes conclusions que les premiers tests effectués au pas de temps horaire, sur des chroniques de un an. Les résultats concernant la deuxième série de tests ne sont pas présentés.

4.4.2 Méthode Pas à pas (PAP-GR) (Figure 4-1)

La méthode 'pas-à-pas' a été développée par *Michel* [1989] et *Edijatno* [1991] au Cemagref d'Antony. C'est une méthode locale directe, très simple, qui optimise (maximise, dans notre cas) une fonction objectif (critère de NS). Cette méthode a été utilisée et testée par *Edijatno* [1991], *Nascimento* [1995] et *Perrin* [2000]. La méthode

PAP-GR, dont l'algorithme est représenté à la Figure 9-20, est décrite exhaustivement par *Perrin* [2000] et en Annexe 9.1.

La méthode PAP-GR a été proposée par [*Michel*, 1989] suite aux échecs de la méthode de *Rosenbrock* [1960] pour caler le modèle GR3J (seulement trois paramètres), constatés par *Edijatno* [1991]. Ensuite, *Nascimento* [1995] a réalisé une évaluation de cette méthode en l'appliquant au modèle GR4J (quatre paramètres). Pour cela, il l'a comparée aux méthodes globales URS et EG (Tableau 4-1), en suivant la méthodologie de *Duan et al.* [1992] et à une méthode multi-départ, sur des données du bassin versant de l'Orgeval. Les résultats montrent que, malgré les problèmes classiquement rencontrés dans les surfaces de réponse (existence d'optima locaux, régions de la surface de réponse faiblement sensibles à la variation des paramètres), la méthode PAP-GR donne d'assez bons résultats par rapports aux trois méthodes testées. Le faible nombre de paramètres du modèle GR4J et le bon degré d'indépendance entre eux permettent de rendre les surfaces de réponse facilement explorables par la méthode PAP-GR. Même si les résultats ne concernent qu'un bassin, la méthode PAP-GR paraît fiable à plus de 90 % (même optimum que ceux trouvés par les méthodes globales dans 90 % des cas).

[*Perrin*, 2000] a suivi les recommandations de *Nascimento* [1995] et utilisé la méthode PAP-GR avec une procédure de multi-départs, pour calibrer quatre modèles pluie-débit (GR4J, IHAC, TOPM et HBV0 ; de quatre à huit paramètres) sur un échantillon de 307 bassins versants français. Classiquement, les méthodes multi-départs utilisent des points de départ aléatoires, ce qui les rend généralement peu efficaces et peu robustes, sachant qu'un point de départ bien choisi pour un ensemble de bassins versants peut donner de bons résultats. Pour éviter cet inconvénient, cet auteur a adopté une stratégie originale de classification hiérarchique ascendante pour déterminer différents points de départs. Cette étude a montré la robustesse de la méthode PAP-GR, car l'utilisation de deux points de départ n'améliore que faiblement les performances en calage des modèles et presque pas les performances en validation.

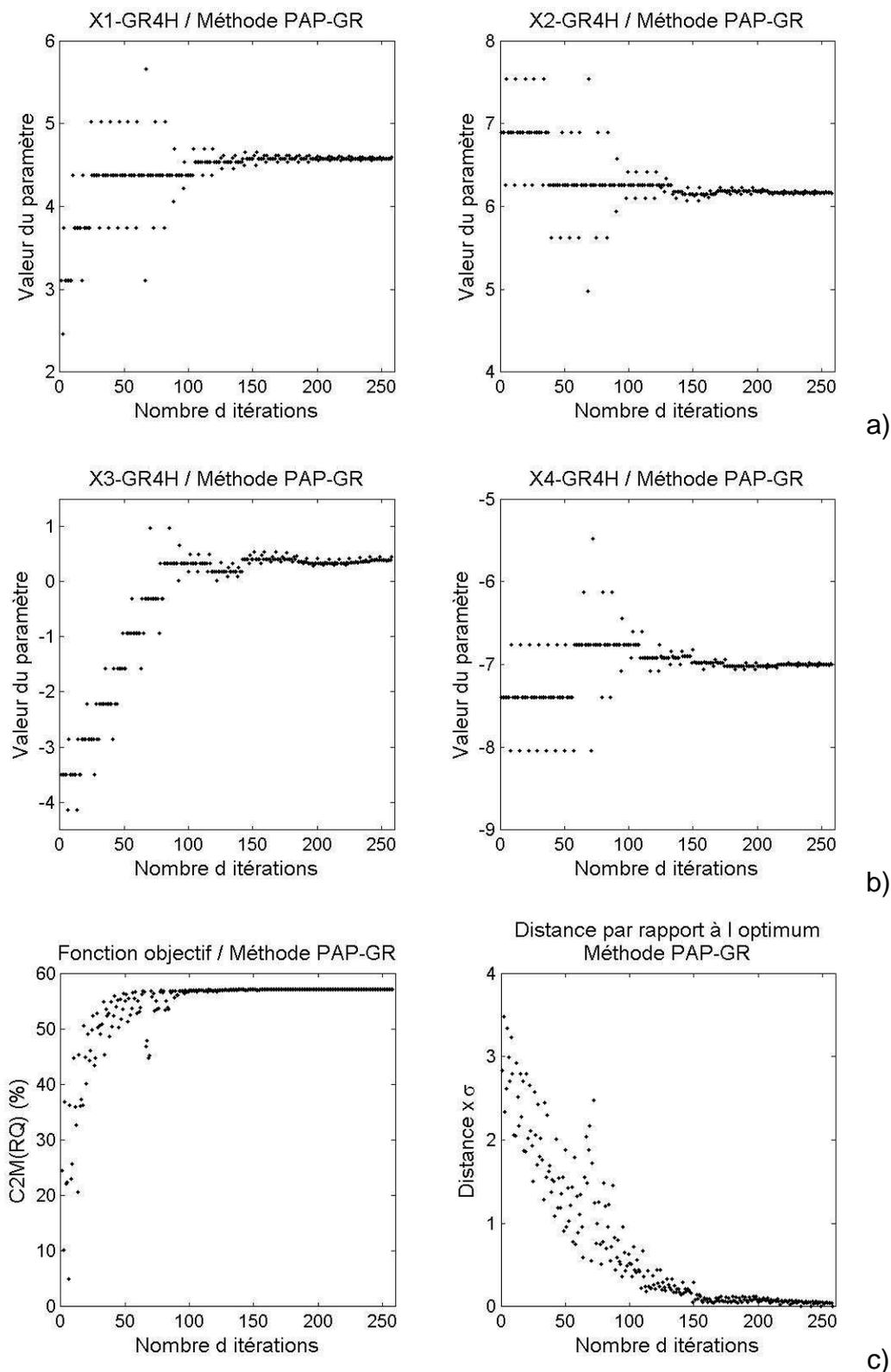


Figure 4-1 : Illustration du fonctionnement de l'algorithme d'optimisation PAP-GR. a) et b) évolution de la valeur des paramètres X_1 , X_2 , X_3 et X_4 du modèle GR4H, au cours de l'optimisation. c) évolution de la valeur de la fonction objectif et de la distance des vecteurs par rapport au vecteur optimal au cours de l'optimisation (Bassin versant de la Doller à Sewen, calage : 1995-1996, **258 itérations**).

4.4.3 Méthode SCE-UA (Figure 4-2)

La méthode du SCE-UA *Duan et al.* [1992, 1993, 1994] est une méthode d'optimisation globale, qui traite le problème de l'optimisation comme un processus d'évolution naturelle d'une population, comme les algorithmes génétiques. L'échantillon de points constitue une population (un point ou un individu représente un vecteur de paramètres). La population est divisée en plusieurs sous-groupes de points, les complexes. Chaque complexe a la possibilité d'évoluer de manière indépendante des autres complexes (i.e., chaque complexe peut évoluer dans l'espace des paramètres dans différentes directions). Après un certain nombre de générations (évolutions), les complexes sont mélangés périodiquement pour former de nouvelles populations et le processus itératif d'évolution des populations est recommencé. Cette procédure permet d'augmenter les probabilités de succès de l'algorithme en partageant les informations gagnées indépendamment par chaque complexe.

L'algorithme d'évolution des complexes (CCE, *competitive complex evolution*) est le suivant. Chaque complexe est composé de plusieurs individus qui ont la possibilité de participer au processus de reproduction et de génération de nouveaux individus. Un sous-complexe, sélectionné à partir d'un complexe, est comme une paire de parents, à part qu'il peut être composé de plus de deux individus. Pour s'assurer que le processus d'évolution est compétitif, la probabilité que de 'meilleurs' parents (vecteurs de paramètres ayant de 'meilleures' valeurs de fonction objectif) donne des enfants est plus grande, que celle des 'mauvais' parents. Pour cela, une fonction de probabilité triangulaire est utilisée *Nelder et Mead* [1965]. Cette procédure est utilisée pour générer la plupart des générations enfant. Cette stratégie d'évolution des complexes utilise l'information contenue dans chaque sous-complexe pour diriger l'évolution des générations, dans la direction d'une amélioration de la fonction objectif. Par ailleurs, au cours de l'évolution des générations, des points (enfants) tirés aléatoirement dans l'espace des paramètres sont introduits, afin d'empêcher que le processus d'évolution des générations ne soit piégé dans une région limitée de l'espace. Ces points tirés aléatoirement représentent un processus de mutation. Chaque mutation doit aider à augmenter la quantité d'information contenue dans l'échantillon de points. Enfin, chaque nouveau point remplace le plus mauvais point du sous-complexe en cours d'évolution. Cela permet à chaque point du sous-complexe d'avoir une chance de contribuer au processus de reproduction avant d'être remplacé ou éliminé. Ainsi, aucun des points du sous-complexe n'est ignoré. Le processus CCE et de mélange de complexes de l'algorithme du SCE-UA permet à l'information contenue dans chaque échantillon de points d'être complètement et efficacement exploitée. Cette méthode d'optimisation paraît très efficace pour optimiser les vecteurs de paramètres des modèles hydrologiques.

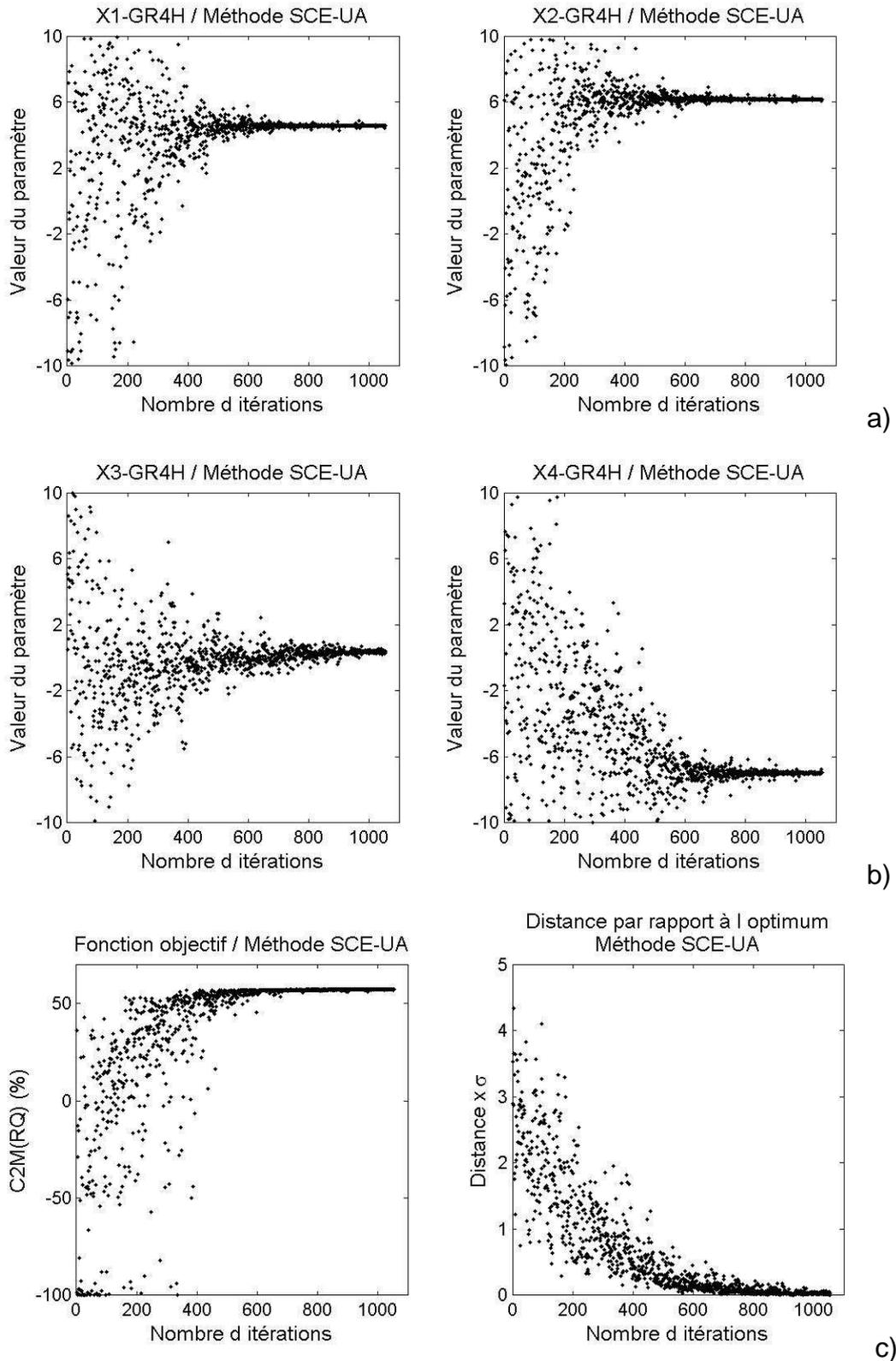


Figure 4-2 : Illustration du fonctionnement de l'algorithme d'optimisation SCE-UA. a) et b) évolution de la valeur des paramètres X_1 , X_2 , X_3 et X_4 du modèle GR4H, au cours de l'optimisation. c) évolution de la valeur de la fonction objectif et de la distance des vecteurs par rapport au vecteur optimal au cours de l'optimisation (Bassin versant de la Doller à Sewen, calage : 1995-1996, **1055 itérations**).

4.4.4 Méthode GA-EDF (Figure 4-3)

Les méthodes d'optimisation basées sur les algorithmes génétiques ont été introduites par *Holland* [1975]. Leur principe est d'effectuer la recherche d'un vecteur optimal de paramètres, en se basant sur les mécanismes de sélection naturelle et de sélection génétique. Le principe de Darwin considère qu'une espèce évolue par sélection naturelle, de façon à s'adapter aux conditions qui lui sont imposées et que seuls les individus qui se sont adaptés survivent et peuvent avoir une descendance.

Un algorithme génétique est basé sur :

- Une population : un ensemble d'individus (un vecteur de paramètres), constitués de gènes (un des paramètres du vecteur) ;
- La duplication : pendant cette phase, les gènes d'un individu sont dupliqués pour être transmis à une génération ;
- Le croisement des gènes : pendant cette phase, des gènes sont choisis aléatoirement chez des individus parents, pour être transmis à un individu d'une génération enfant ;
- La mutation : pendant cette phase, des gènes sont altérés aléatoirement pour introduire de nouveaux éléments dans la population.

L'algorithme génétique part d'une population de vecteurs de paramètres et opère des croisements entre individus (parents) pour créer de nouvelles générations (enfants). A chaque croisement, les nouveaux individus sont composés des gènes de leurs parents et de quelques gènes qui peuvent être mutés. Les contraintes de survie des individus sont matérialisées par le critère à optimiser (minimiser ou maximiser). Pour chaque individu, on calcule la valeur de la fonction objectif et on sélectionne les n individus qui répondent au mieux à ce critère. Ensuite les individus sélectionnés sont utilisés à nouveau pour construire de nouvelles générations et les faire évoluer. D'une génération à la suivante, le seuil de survie est de plus en plus sélectif, sans jamais entraîner d'extinction de la population, puisque les n meilleurs individus sont toujours conservés. Ainsi, après un grand nombre de générations, on obtient un groupe d'individus qui se sont adaptés à un critère de plus en plus sélectif.

Dans les domaines de la gestion des ressources en eau et de la modélisation hydrologique, les algorithmes génétiques ont été très utilisés *Anctil et al.* [2004]. En modélisation pluie-débit, ils ont été initialement développés et testés au cours des années 1990 [*Wang, 1991; Franchini, 1996; Franchini et Galeati, 1997; Wang, 1997; Franchini et al., 1998*]. Un de leurs avantages est d'être indépendants du point de départ de leur recherche, de ne pas être piégés par des optima locaux et de localiser facilement la région de l'optimum.

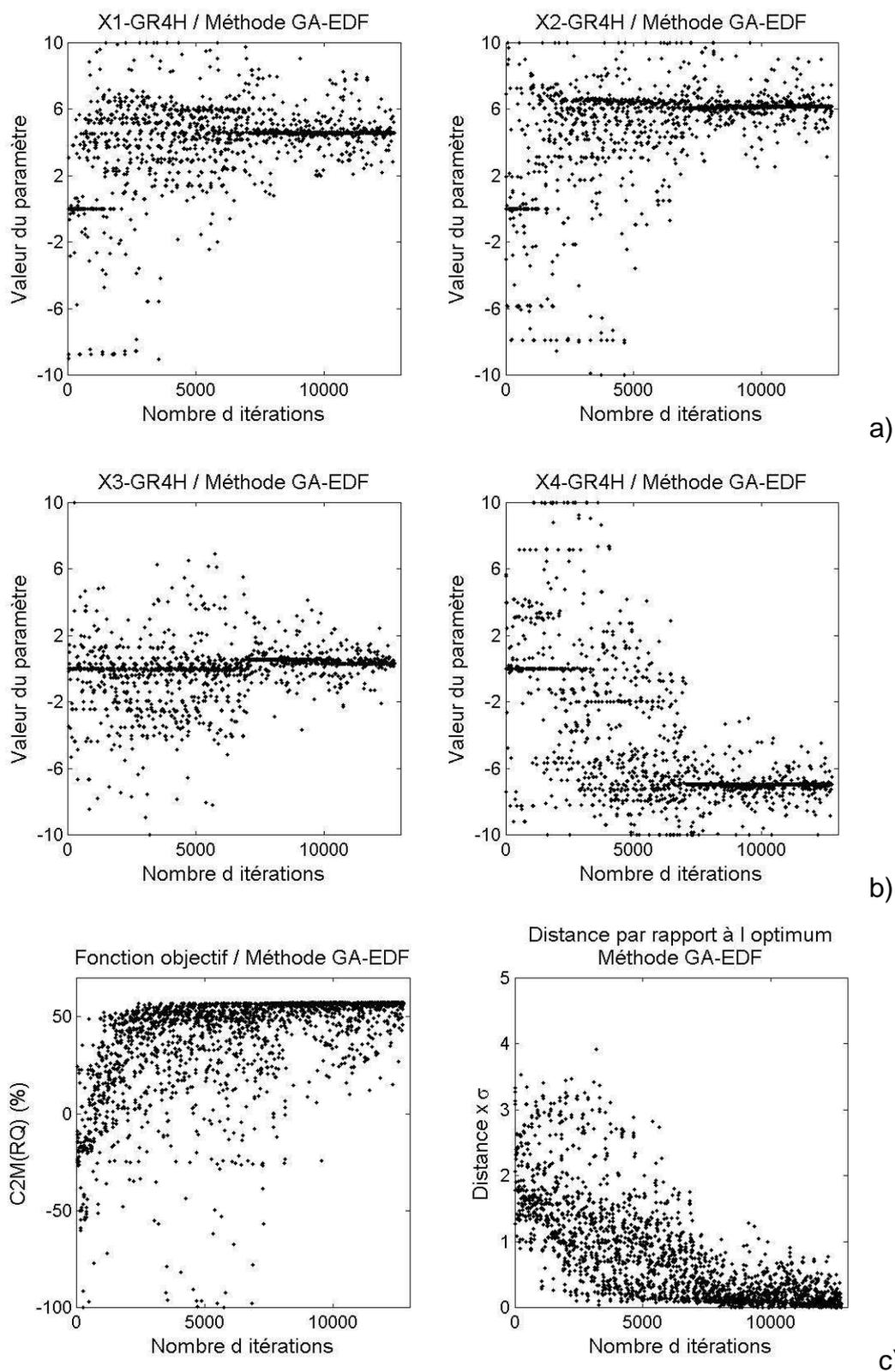


Figure 4-3 : Illustration du fonctionnement de l'algorithme d'optimisation GA-EDF. a) et b) évolution de la valeur des paramètres X_1 , X_2 , X_3 et X_4 du modèle GR4H, au cours de l'optimisation. c) évolution de la valeur de la fonction objectif et de la distance des vecteurs par rapport au vecteur optimal au cours de l'optimisation (Bassin versant de la Doller à Sewen, calage : 1995-1996, **12767 itérations**).

Un de leurs désavantages est leur manque d'efficacité (ce qui est inhérent à toute méthode globale) et la difficulté qu'ils ont à localiser avec précision l'optimum. Par contre, ils sont performants lorsqu'ils sont associés avec une méthode locale.

L'algorithme génétique testé dans cette thèse est celui utilisé par les services opérationnels d'EDF-DTG [Paquet, 2004]. L'algorithme a été utilisé tel quel, sans d'autres modifications que celle de l'ajout d'un critère de convergence qui permet de limiter le nombre de générations lorsque les performances de la population des vecteurs de paramètres ne progressent plus.

4.5 Résultats

Dans cette étude nous comparons les performances en calage et en validation des modèles et leurs vecteurs de paramètres obtenus avec les méthodes d'optimisation PAP-GR, SCE-UA et GA-EDF, pour nos quatre modèles pluie-débit (Tableau 4-3). Afin de limiter le nombre de figures, nous ne présenterons les résultats détaillés que pour les modèles GR4H et MORDOR10. Les autres résultats se trouvent en Annexe 5. Dans les tableaux nous présentons l'intégralité des résultats. Par rapport à l'étude de Gan et Biftu [1996], notre étude est exhaustive, dans le sens où elle teste ces trois méthodes d'optimisation sur un large échantillon de bassins versants, avec quatre modèles de complexité croissante et qu'elle ne s'intéresse pas seulement aux performances des méthodes d'optimisation, mais aussi aux vecteurs de paramètres optimisés.

Tableau 4-3 : Coefficients de corrélation (r) entre performance et vecteurs de paramètres obtenus selon 2 méthodes d'optimisation différentes.

Modèle	PAP-GR vs SCE-UA				SCE-UA vs GA-EDF			
	GR4H	IHAC6	TOPMO	MORDOR10	GR4H	IHAC6	TOPMO	MORDOR10
Calage	0.95	0.98	0.88	0.93	0.99	0.99	0.88	0.98
Validation	0.93	0.97	0.80	0.91	0.96	0.95	0.78	0.95
X_1	0.44	0.03	0.19	0.57	0.67	0.87	0.48	0.73
X_2	0.19	0.63	0.25	0.35	0.73	0.75	0.59	0.40
X_3	0.88	0.63	0.49	0.42	0.92	0.69	0.57	0.54
X_4	0.73	0.74	0.44	0.14	0.83	0.89	0.47	0.23
X_5	-	0.61	0.26	0.65	-	0.75	0.25	0.65
X_6	-	0.85	0.05	0.64	-	0.87	0.29	0.66
X_7	-	-	0.27	0.69	-	-	0.28	0.72
X_8	-	-	0.39	0.47	-	-	0.34	0.57
X_9	-	-	-	0.56	-	-	-	0.57
X_{10}	-	-	-	0.52	-	-	-	0.55

4.5.1 Performances des modèles pluie-débit

Performances en calage

Au calage, les méthodes globales sont capables de trouver de meilleurs vecteurs de paramètres que la méthode locale (Tableau 4-4). Les différences de C_{2M} moyen sur l'ensemble des 313 bassins versants vont de 1 à 7 points. Les méthodes globales sont d'autant meilleures que le nombre de paramètres à optimiser est grand. Ceci s'explique dans la mesure où ces méthodes ont été développées pour ne pas être trop sensibles au point de départ de l'algorithme de recherche et aux optima locaux, alors que les méthodes locales le sont d'autant plus que le nombre de paramètres à optimiser est grand. Par ailleurs, on remarque une légère différence de performances entre la méthode SCE-UA et GA-EDF. Cette différence provient essentiellement du fait que les utilisateurs peuvent choisir quelques paramètres, qui vont guider la puissance de recherche de ces algorithmes. Afin de limiter le nombre d'itérations de la méthode SCE-UA, ses paramètres ont été fixés à leurs valeurs basses (par rapport à celles préconisées par [Duan et al., 1994]). Par contre, la méthode GA-EDF a été utilisée avec les paramètres fournis par EDF-DTG. Aucune étude de sensibilité n'a été faite sur les paramètres de la méthode GA-EDF. Il en résulte que sur les tests effectués, la méthode GA-EDF est légèrement meilleure que la méthode SCE-UA. Cette différence n'est probablement qu'un artéfact lié à la différence de paramétrisation de ces deux méthodes, la méthode GA-EDF nécessitant en moyenne trois fois plus d'itérations que la méthode SCE-UA.

Tableau 4-4 : Performances moyennes en calage et validation, obtenues pour les méthodes PAP-GR, SCE-UA et GE-EDF, avec les modèles GR4H, IHAC6, TOPMO8 et MORDOR10

	CALAGE			VALIDATION		
	PAP-GR	SCE-UA	GA-EDF	PAP-GR	SCE-UA	GA-EDF
GR4H	57.3	59.1	59.1	33.7	33.5	33.6
IHAC6	52.6	53.5	53.8	29.5	29.3	30.2
TOPMO8	55.2	58.3	62.3	29.4	26.9	27.1
MORDOR10	59.5	62.8	63.0	36.2	35.3	35.2

Par rapport aux modèles utilisés, on peut également remarquer que les performances des méthodes d'optimisation dépendent des modèles optimisés, c'est-à-dire de la structure de la surface de réponse qui résultent de trois facteurs essentiels : la structure du modèle, la fonction objectif et la période de calage sélectionnée. Par exemple, le modèle IHAC6 semble plutôt assez facile à optimiser, dans le sens où les trois méthodes ont des niveaux de performance assez proches. Ce comportement est peut-être dû à la simplicité de sa structure et des fonctions mathématiques utilisées. En revanche, le modèle TOPMO8 semble quant à lui assez

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

difficile à optimiser (et même plus difficile à optimiser que MORDOR10). La méthode locale subit de nombreux échecs et la différence de performances entre les deux méthodes globales est grande. Ce comportement peut éventuellement venir du manque de robustesse constaté au pas de temps horaire des modèles de la famille TOPMO.

Tableau 4-5 : Pourcentage d'optima trouvé par une méthode d'optimisation pour un modèle (l'optimum est le vecteur de paramètres qui donne le meilleur critère de calage, à +/- 0.1 points). En validation, l'optimum est le meilleur des trois vecteurs.

	CALAGE			VALIDATION		
	PAP-GR	SCE-UA	GA-EDF	PAP-GR	SCE-UA	GA-EDF
GR4H	78	93	93	62	62	62
IHAC6	78	94	97	58	63	65
TOPMO8	18	54	52	46	29	35
MORDOR10	53	82	69	46	42	49

Le Tableau 4-5 montre le pourcentage d'optima trouvés par une méthode de calage (l'optimum est le vecteur de paramètres qui donne le meilleur critère de calage, à +/- 0.1 points). Les méthodes globales sont particulièrement efficaces et trouvent entre 70% et 90% des optima, alors que la méthode locale en trouve entre 50% et 80% (sans prendre en compte le cas de TOPMO8, qui pose des problèmes de calage).

Tableau 4-6 : Pourcentages de calages où deux méthodes de calage donnent des critères C_{2M} égaux (à +/- 0.1 points), et où l'une est supérieure ou inférieure à l'autre (modèle GR4H)

GR4H	CALAGE			VALIDATION		
Méth.1 / Méth.2	PAPGR / SCEUA	PAPGR / GAEDF	SCEUA / GAEDF	PAPGR / SCEUA	PAPGR / GAEDF	SCEUA / GAEDF
Méth.1 ≈ Méth. 2	77	78	88	56	37	45
Méth.1 > Méth.2	2	2	7	20	30	27
Méth.2 > Meth.1	21	20	6	24	33	28

Tableau 4-7 : Pourcentages de calages où deux méthodes de calage donnent des critères C_{2M} égaux (à +/- 0.1 points), et où l'une est supérieure ou inférieure à l'autre (modèle MORDOR10).

MORDOR10	CALAGE			VALIDATION		
Méth.1 / Méth.2	PAPGR / SCEUA	PAPGR / GAEDF	SCEUA / GAEDF	PAPGR / SCEUA	PAPGR / GAEDF	SCEUA / GAEDF
Méth.1 ≈ Méth. 2	53	42	54	27	16	21
Méth.1 > Méth.2	7	19	29	38	41	38
Méth.2 > Meth.1	40	39	17	34	43	41

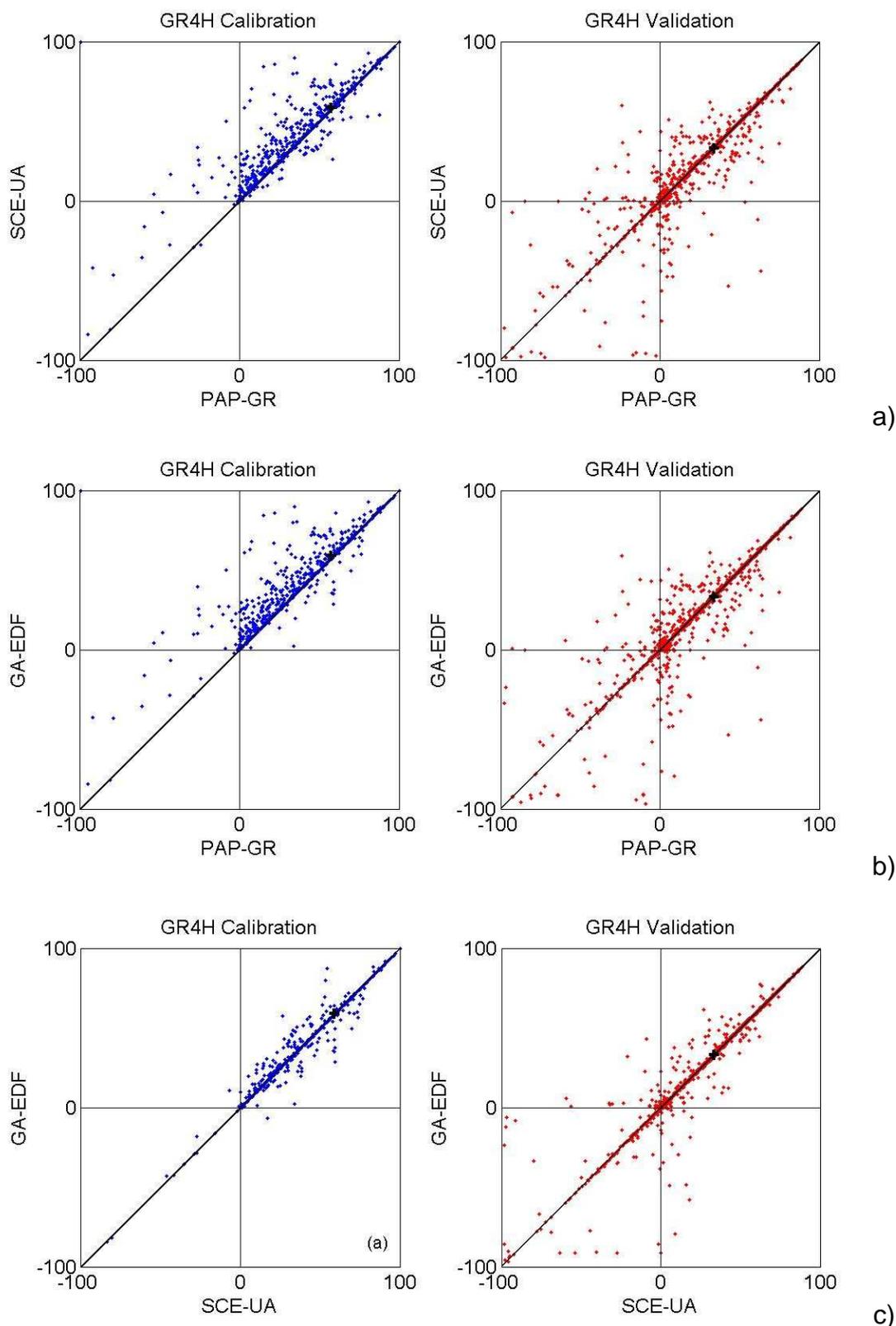


Figure 4-4 : Comparaison des performances en calibration et en validation des vecteurs trouvés par les méthodes PAP-GR, SCE-UA et GA-EDF, pour le modèle GR4H. a) PAP-GR et SCE-UA, b) PAP-GR et GA-EDF, c) SCE-UA et GA-EDF.

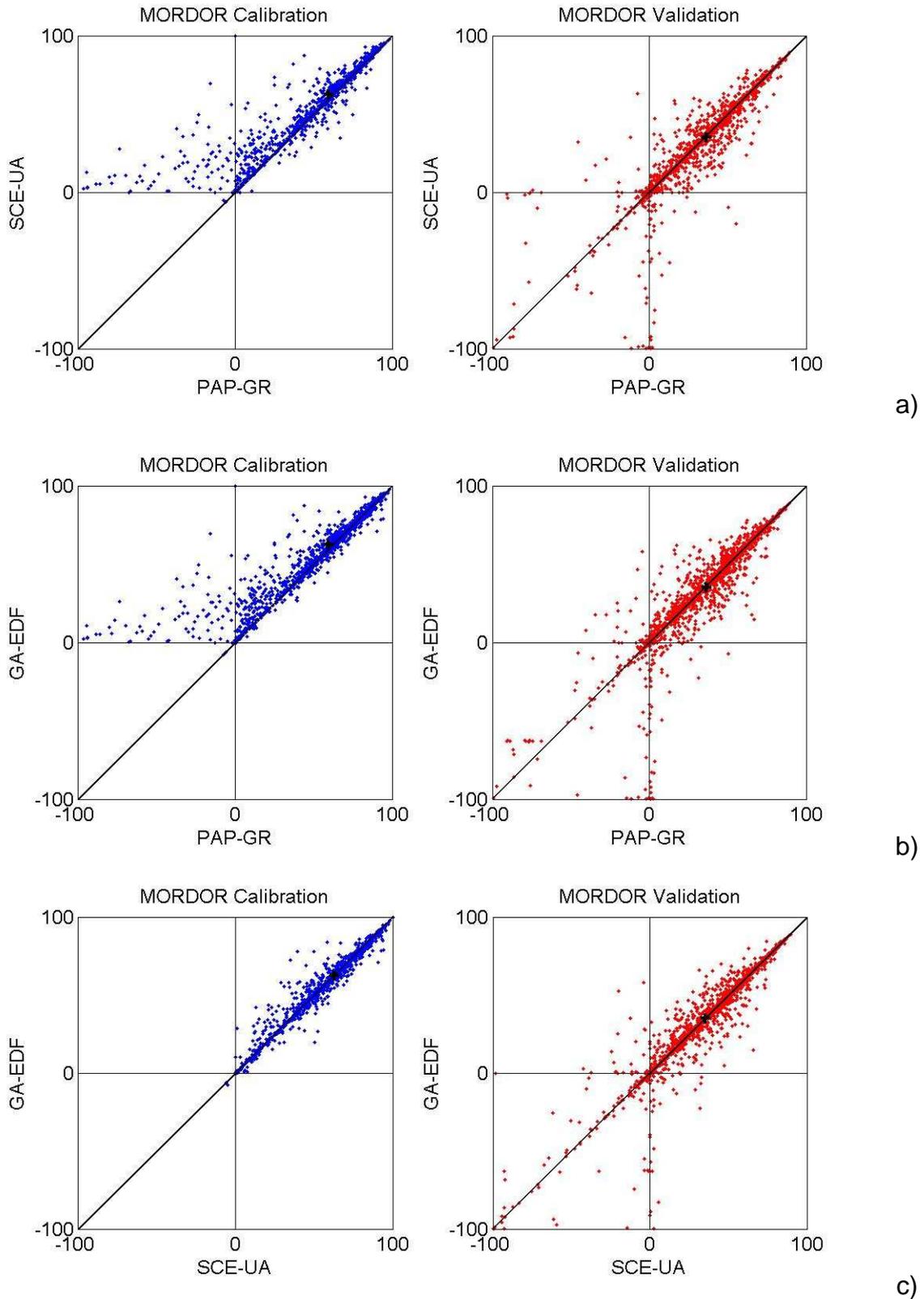


Figure 4-5 : Comparaison des performances en calibration et en validation des vecteurs trouvés par les méthodes PAP-GR, SCE-UA et GA-EDF, pour le modèle MORDOR10. a) PAP-GR et SCE-UA, b) PAP-GR et GA-EDF, c) SCE-UA et GA-EDF.

La Figure 4-4 et la Figure 4-5 comparent les performances en calage obtenues pour chaque période par deux méthodes d'optimisation différentes. Ces figures illustrent la supériorité apparente des méthodes globales, la majorité des points se trouvant sur ou au-dessus de la première bissectrice.

Le Tableau 4-6 et le Tableau 4-7 montrent le pourcentage de calages où deux méthodes donnent des critères C_{2M} égaux, ou bien où l'une est supérieure ou inférieure à l'autre, pour les modèles GR4H et MORDOR10. On se rend compte encore une fois que les méthodes globales sont généralement supérieures à la méthode locale. Néanmoins pour le modèle GR4H, la différence est faible et les méthodes locales et globales obtiennent le même niveau de performance dans presque 80% des périodes de calage. Pour MORDOR10, la différence entre méthodes globales et locales s'accroît et elles ont le même niveau de performance pour 50% des périodes de calage. Entre les deux méthodes globales, les résultats sont équilibrés.

Pour GR4H, lorsque l'on compare les performances des vecteurs obtenus avec deux méthodes différentes, une majorité des critères de calage sont égaux, alors que pour MORDOR10 seulement la moitié de ces critères de calage sont égaux. Enfin, ces tests montrent que les méthodes globales (notamment le SCE-UA) ne sont pas infaillibles, contrairement à ce que croient de nombreux utilisateurs : dans de faibles proportions (de 2% à 20 %), la méthode locale est capable de trouver de meilleurs optima que les méthodes globales ! Notamment par rapport à la méthode GA-EDF, qui est moins bonne que la méthode PAP-GR dans 20% des cas, pour le modèle MORDOR. En fait, malgré les très bons résultats de la méthode GA-EDF, il ne faut pas oublier que les méthodes basées sur des algorithmes génétiques ont généralement du mal à affiner leurs recherches, une fois qu'elles ont localisé la région de l'optimum. On voit ici l'intérêt de SCE-UA qui est capable d'affiner ses recherches avec la procédure du Simplex. Les cas où la méthode locale est supérieure au SCE-UA sont plus rares (néanmoins, jusqu'à 7%, pour MORDOR10).

L'étude des performances de ces trois méthodes d'optimisation en calage montre que les deux méthodes globales sont généralement meilleures que la méthode locale. Néanmoins, avec des modèles de 4 à 10 paramètres, dans au moins 50% des cas, méthodes globales et locale ont le même niveau de performance. Par ailleurs, de façon surprenante, dans une faible proportion des cas, la méthode locale est meilleure que les méthodes globales. Ceci démontre finalement que (1) les méthodes locales peuvent être performantes en calage et (2) que les méthodes globales ne sont pas infaillibles pour localiser l'optimum, sur des données réelles.

Performances en validation

En validation, l'avantage des méthodes globales par rapport à la méthode locale s'efface. Les performances de la méthode PAP-GR sont équivalentes ou meilleures à celles des méthodes globales (jusqu'à deux points de C_{2M} gagnés en moyenne). Ce constat, déjà fait dans la littérature mais rapporté avec prudence [Gan et Biftu, 1996; Kuzmin et al., 2004], est clairement mis en évidence grâce au nombre de tests effectués ici. En effet, comme le montrent la Figure 4-4 et la Figure 4-5, le nuage de points qui était sur ou au-dessus de la première bissectrice en calage, se trouve centré sur la première bissectrice.

La Figure 4-6 montre la différence de performances obtenue entre deux méthodes en validation, en fonction de la différence de performances obtenue en calage. Sur cette figure, on peut distinguer 4 zones :

- [i] zone où la méthode 1 est supérieure à la méthode 2 en calage et en validation ;
- [ii] zone où la méthode 1 est supérieure à la méthode 2 en calage et inférieure en validation ;
- [iii] zone où la méthode 1 est inférieure à la méthode 2 en calage et en validation ;
- [iv] zone où la méthode 1 est inférieure à la méthode 2 en calage et supérieure en validation.

Quand on compare une méthode globale à la méthode locale, les Figure 4-6 a), b), d), e) montrent que les vecteurs de paramètres trouvés avec les méthodes globales dont les performances étaient supérieures en calage, ont presque autant de chance d'être meilleurs ou inférieurs en validation, par rapport à ceux trouvés avec la méthode locale. Par exemple, avec le modèle GR4H, 95 % des vecteurs optimisés par la méthode SCE-UA donnent des performances égales ou meilleures en calage, que celles des vecteurs trouvés par la méthode PAP-GR. Par contre, en validation, seulement 55 % de ces vecteurs conduisent à de meilleures performances et 45 % conduisent à des performances inférieures. Pour le modèle MORDOR10, 80 % des vecteurs optimisés avec la méthode SCE-UA donnent des performances égales ou meilleures en calage et 50 % de ces vecteurs conduisent à des performances meilleures ou inférieures en validation ! Ces résultats indiquent donc que trouver un optimum au cours du calage ne garantit pas d'obtenir les meilleurs résultats en validation.

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

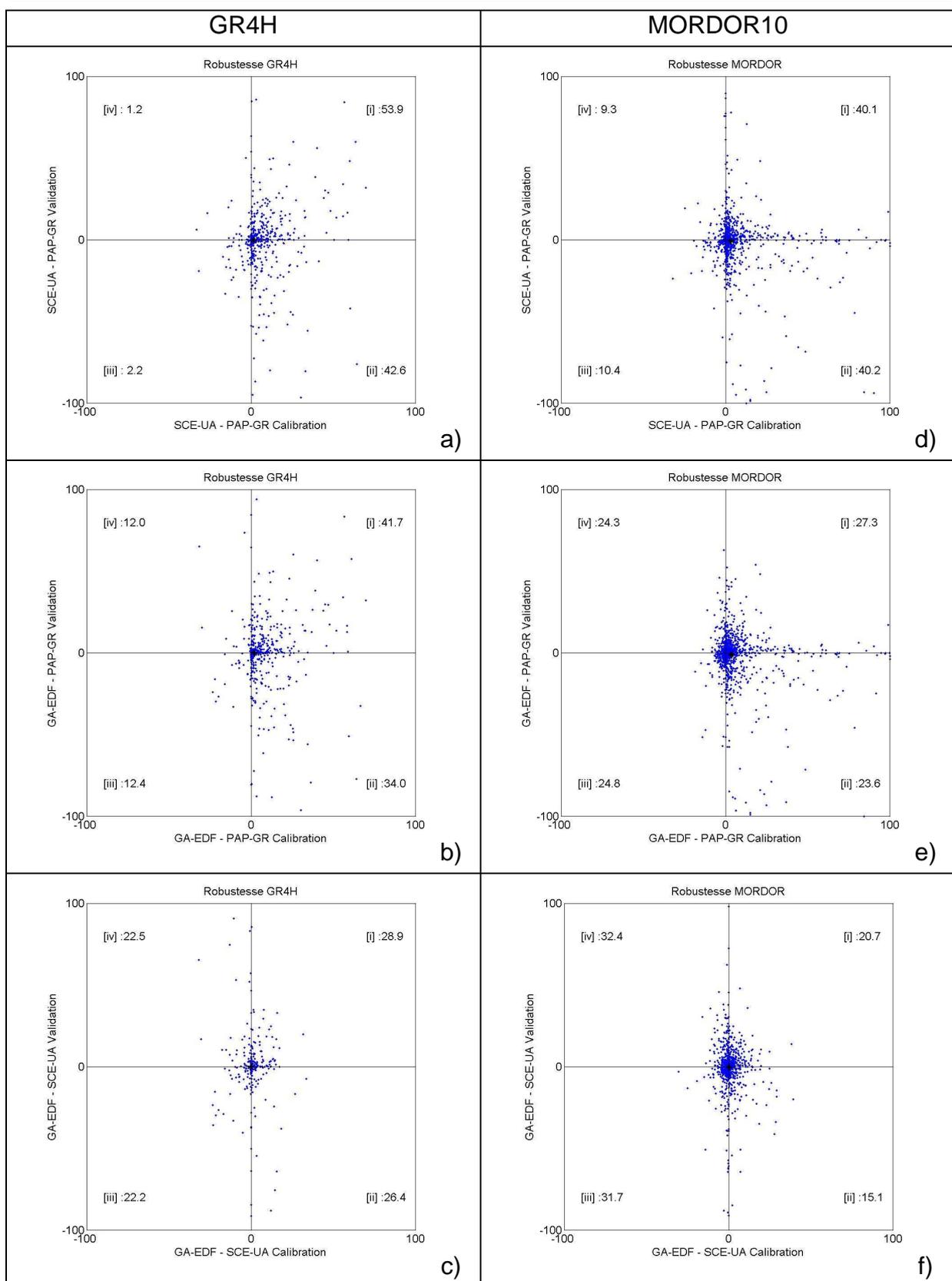


Figure 4-6 : Comparaison de la différence de performances obtenues en calibration et en validation, entre deux méthodes d'optimisation. a) et d) PAP-GR et SCE-UA, b) et e) PAP-GR et GA-EDF, c) et f) SCE-UA et GA-EDF (respectivement modèles GR4H et MORDOR10).

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

Cette tendance est également mise en évidence par le Tableau 4-5, le Tableau 4-6 et le Tableau 4-7. D'une part, excepté pour le cas particulier de TOPMO8, le pourcentage d'optima trouvés par une méthode diminue fortement (d'au moins 20%) entre le calage et la validation. Ceci indique encore une fois que trouver un optimum au cours du calage ne garantit pas d'obtenir les meilleurs résultats pendant la validation. D'autre part, le pourcentage de cas où deux méthodes donnent des résultats égaux en calage diminue fortement en validation. Ceci indique que dans des conditions hydrologiques différentes, deux vecteurs de paramètres qui pouvaient donner des performances du même ordre en calage donnent des performances sensiblement différentes en validation. Cet effet est certainement dû à l'existence de nombreux vecteurs équifinaux. Par ailleurs, le pourcentage de cas où la méthode locale est supérieure aux méthodes globales augmente fortement du calage à la validation.

Le Tableau 4-9 montre les performances moyennes pour chaque modèle :

- en calage *a priori* : lorsqu'on prend pour chaque période de calage la performance obtenue par le meilleur des trois vecteurs obtenu en calage par les différentes méthodes d'optimisation ;
- en validation *a posteriori* : lorsqu'on prend pour chaque période de validation la performance en validation obtenue par le meilleur des trois vecteurs ;
- en validation *a priori* : lorsqu'on prend pour chaque période de validation la performance obtenue en validation par le meilleur des trois vecteurs en calage.

Pour clarifier ces définitions, le Tableau 4-8 explique comment sont choisis les performances en calage *a priori*, validation *a posteriori* et validation *a priori*.

Tableau 4-8 : Explication des différents types de manière de calculer les performances des modèles en calage a priori, en validation a posteriori et en validation a priori. Exemple de l'application d'un modèle sur un bassin.

	Calibration			Validation		
	PAP-GR	SCE-UA	GA-EDF	PAP-GR	SCE-UA	GA-EDF
GR4H	69.7	73.5	71.2	57.8	55.0	56.6
		Meilleur calage <i>a priori</i>		Meilleure validation <i>a posteriori</i>	Meilleure validation <i>a priori</i>	

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

Tableau 4-9 : Performances moyennes du meilleur calage a priori, de la meilleure validation a posteriori et de la meilleure validation a priori, obtenues avec la meilleure des trois méthodes d'optimisation (PAP-GR, SCE-UA et GE-EDF), avec les modèles GR4H, IHAC6, TOPMO8 et MORDOR10.

	C _{2M} moyen du meilleur calage a priori	C _{2M} moyen de la meilleure validation a posteriori	C _{2M} moyen de la meilleure validation a priori
GR4H	59.4	35.5	33.6
IHAC6	53.9	30.9	29.9
TOPMO8	63.9	36.7	27.9
MORDOR10	63.6	38.8	35.2

On remarque d'abord que les moyennes des meilleurs calages *a priori* sont assez proches des performances moyennes obtenues avec les méthodes globales (le gain est à peine d'un point). Ceci indique que, (1) quel que soit le modèle, les deux méthodes globales ont à peu près le même niveau d'efficacité, (2) que les vecteurs de paramètres trouvés ont des performances assez proches et (3) qu'elles sont capables de trouver à elles deux l'ensemble des vecteurs optima (cas de IHAC6 par exemple).

Ensuite, les performances moyennes en validation *a posteriori* sont sensiblement meilleures que les performances moyennes en validation *a priori*. Cette différence peut aller de 1 à 9 points. Elle indique le manque de robustesse de certains des vecteurs trouvés au cours du calage, ce qui est notamment le cas pour le modèle TOPMO8. Le modèle IHAC6, lui, est particulièrement robuste puisque cette différence de performances n'est que d'un point. Cette différence de performance illustre encore une fois que, pour certains modèles, trouver de meilleurs optima au cours du calage ne garantit absolument pas d'obtenir les meilleurs résultats en validation !

Ces différences de résultats entre calage et validation permettent d'illustrer la robustesse des vecteurs de paramètres trouvés par les différentes stratégies d'optimisation des paramètres. Il semble que les vecteurs de paramètres trouvés par les méthodes globales soient moins robustes que ceux trouvés par la méthode locale, dans le sens où ils conduisent généralement à des performances meilleures en calage et inférieures en validation. Ceci est certainement dû (1) à la capacité des méthodes globales à localiser un optimum dans n'importe quelle région de l'espace des paramètres et (2) au choix de la fonction objectif qui est peut être trop sensible aux conditions hydrologiques de la période de calage, ce qui produit un sur-calage des vecteurs de paramètres (même avec des modèles parcimonieux). Ce problème

de sur-calage du vecteur de paramètres a été testé en comparant les performances des méthodes de calage avec des durées de calage de 1 ou 4 ans.

4.5.2 Vecteurs de paramètres

Pour limiter le nombre de figures, seuls les résultats de la méthode PAP-GR et SCE-UA sont montrés dans cette partie. Une synthèse est présentée au Tableau 4-3 et les autres figures se trouvent en Annexe 5.

La comparaison des performances des vecteurs de paramètres optimisés selon les trois méthodes de calage montre que, dans 50 à 90 % des cas, les vecteurs de paramètres donnent le même niveau de performance. On peut alors se demander si les vecteurs de paramètres identifiés par les différentes méthodes se ressemblent. La Figure 4-7 et la Figure 4-8 (et celles en Annexe 5) comparent les valeurs des paramètres optimisés selon deux méthodes différentes. Pour le modèle GR4H, la Figure 4-7 montre une certaine dispersion des valeurs des différents paramètres optimisés. Cependant, une bonne partie des valeurs de ces paramètres se trouve proche de la première bissectrice.

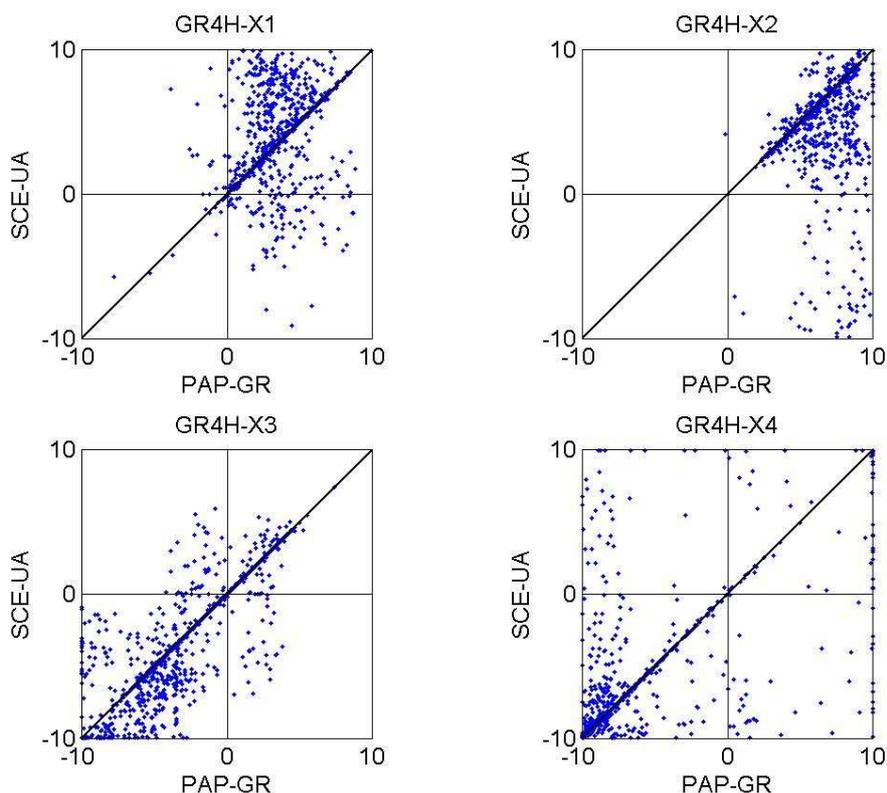


Figure 4-7 : Comparaison des valeurs des paramètres de GR4H optimisés avec les méthodes PAP-GR et SCE-UA

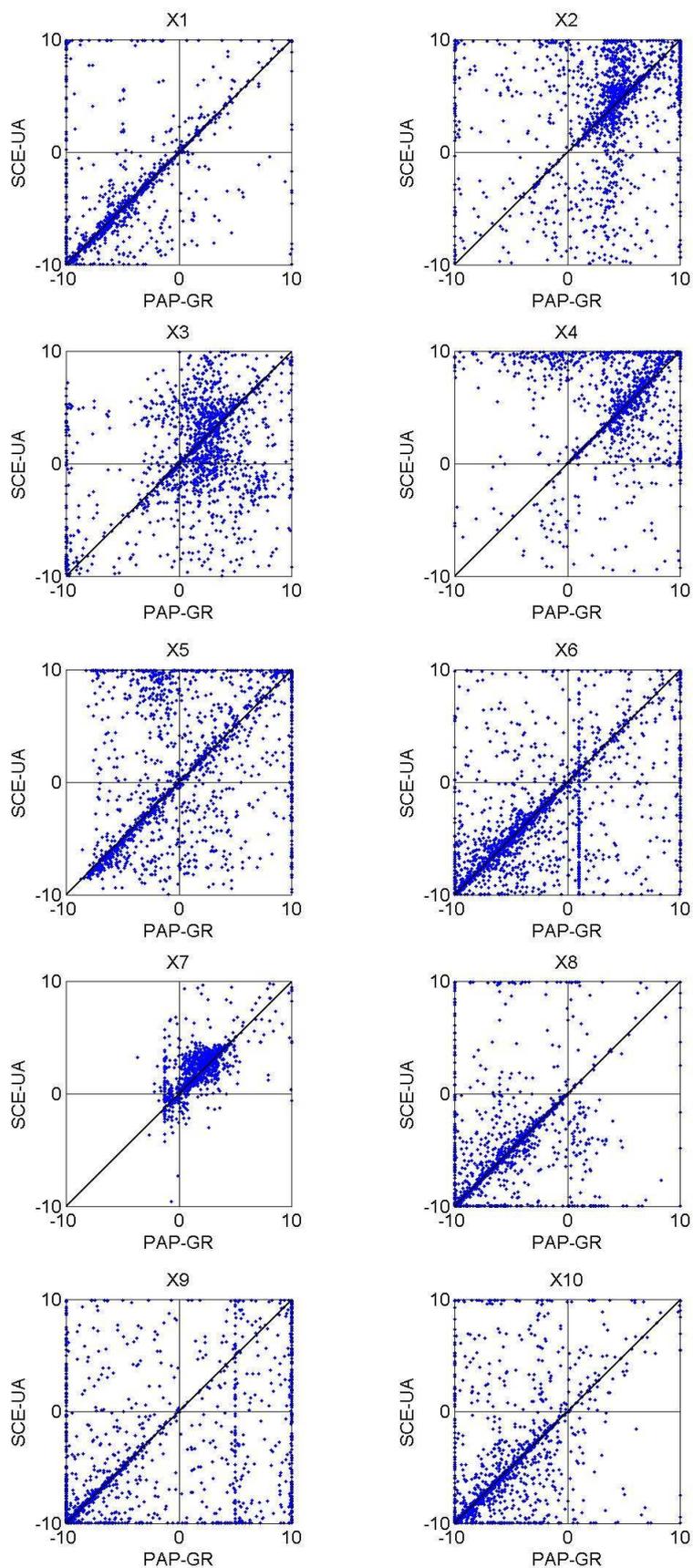


Figure 4-8 : Comparaison des valeurs des paramètres de MORDOR10 optimisés avec les méthodes PAP-GR et SCE-UA

En revanche, la Figure 4-8 montre une plus grande dispersion des valeurs des différents paramètres optimisés du modèle MORDOR10. On peut également remarquer que cette dispersion n'est pas la même pour tous les paramètres et que certains semblent plus contraints (définis) que d'autres. Par exemple, les valeurs des paramètres X_2 , X_3 , X_5 , X_6 et X_9 sont très dispersées et semblent particulièrement mal définies. Avec les modèles IHAC6 et TOPMO8, on rencontre à peu près les mêmes caractéristiques, le modèle IHAC6 ayant ses paramètres assez bien définis (comme GR4H) et le modèle TOPMO8 ayant ses paramètres très mal définis (encore plus dispersés que MORDOR10).

Sur ces figures et pour chaque modèle, on peut néanmoins remarquer que le domaine de recherche des paramètres par la méthode locale est plus limité que ceux des méthodes globales. En effet, les méthodes locales explorent une région assez limitée de l'espace des paramètres.

Pour synthétiser ces informations, on peut calculer la distance moyenne entre deux vecteurs de paramètres. Pour cela nous avons calculé, pour chaque période et couple de paramètres optimisés selon deux méthodes différentes, la distance euclidienne entre deux paramètres, comparée à l'écart-type de la distribution de chaque paramètre :

$$d_{AB,i} = \frac{|X_{A,i} - X_{B,i}|}{\sigma_{X_i}} \quad [4.1]$$

où, $X_{A,i}$ et $X_{B,i}$ sont respectivement les paramètres trouvés par deux méthodes différentes (A et B), et σ_{X_i} est l'écart-type de la distribution du paramètre X_i .

Nous avons normé la distance euclidienne par l'écart-type, pour pouvoir comparer les distances de différents paramètres d'un même modèle, qui ont des distributions dont les écarts-types sont sensiblement différents. Ensuite, la distance entre deux vecteurs de paramètres est définie comme la moyenne des distances entre chaque paramètre :

$$d_{AB} = \frac{1}{N} \sum_{i=1}^N d_{X_{AB,i}} \quad [4.2]$$

où, N est le nombre de paramètres optimisés.

La Figure 4-9 a) et la Figure 4-9 b) montrent, pour les modèles GR4H et MORDOR10, les distributions cumulées des distances moyennes entre vecteurs optimisés selon deux méthodes différentes. La Figure 4-9 a) montre que pour le modèle GR4H environ 80% des vecteurs de paramètres sont les mêmes. Cela

signifie que les trois méthodes convergent vers la même région de l'espace des paramètres, quelle que soit leur point de départ ou leur stratégie pour évoluer dans l'espace des paramètres.

En revanche, pour le modèle MORDOR10, la Figure 4-9 b) montre que seulement 10% à 20% des vecteurs sont très proches ! Au moins 80% des vecteurs sont relativement différents. De plus, cette figure montre qu'au moins 50% des vecteurs sont distants d'au moins un écart-type. Cette figure illustre ce qu'on pouvait pressentir précédemment, les problèmes d'optima locaux et d'équifinalité, dont souffrent les modèles ayant de nombreux paramètres dont certains sont peu ou mal définis. Dans un tel cas, on comprend bien que tous les efforts portés sur le perfectionnement des méthodes de calage n'ont qu'une portée limitée : c'est la structure mathématique même de ces modèles qu'il faut mettre en cause !

Afin d'approfondir notre analyse, nous avons étudié les distances entre les valeurs obtenues par deux méthodes d'optimisation différentes, pour chaque paramètre d'un modèle. Il est probable que de cette manière nous puissions discriminer les paramètres bien ou mal définis.

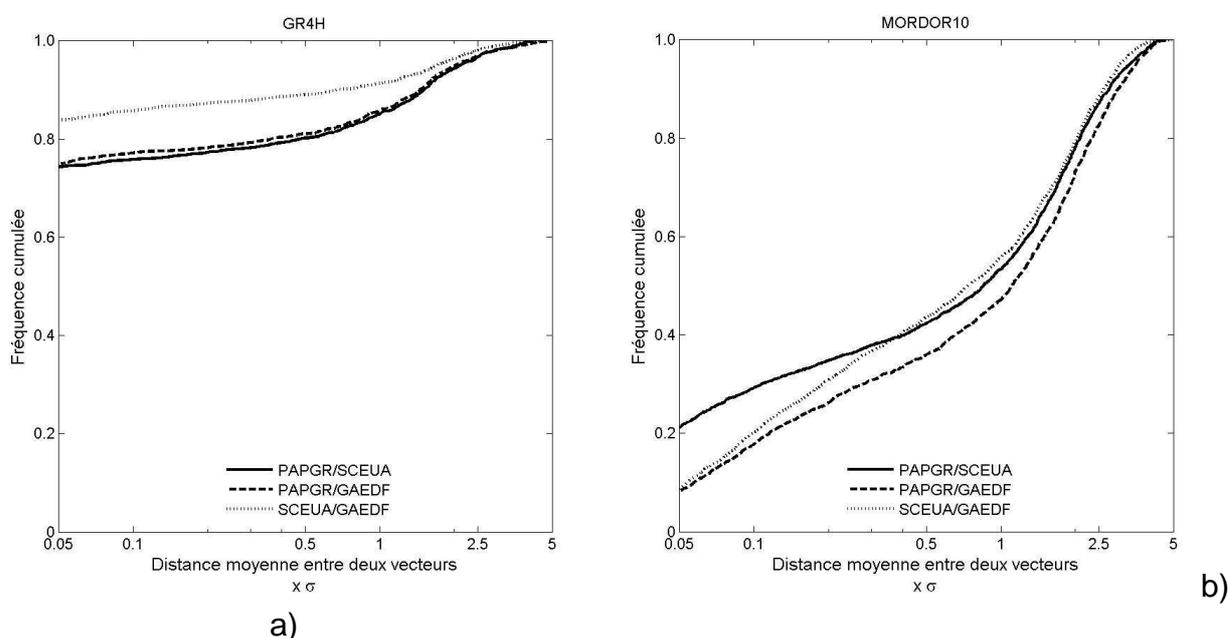


Figure 4-9 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par les deux méthodes d'optimisation PAP-GR et SCE-UA, a) GR4H et b) MORDOR10.

Les Figure 4-10 et Figure 4-11 montrent pour les modèles GR4H et MORDOR10 les distributions cumulées des distances pour chaque paramètre des modèles, entre les valeurs optimisées selon la méthode PAP-GR et SCE-UA. Pour le modèle GR4H, la Figure 4-10 montre que chaque paramètre est très bien défini et que dans plus de 80% des cas la distance entre les deux valeurs d'un paramètre est inférieure à 5%

de l'écart-type de la distribution de ce paramètre, obtenue sur l'ensemble des bassins.

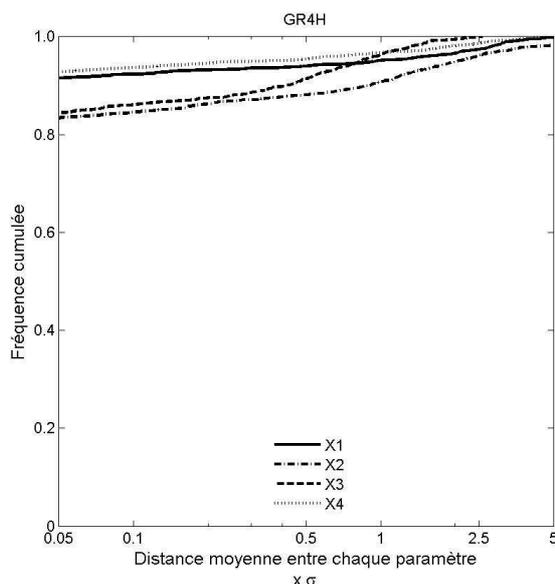


Figure 4-10 : Distribution cumulée des distances entre chaque paramètre trouvés par les deux méthodes d'optimisation PAP-GR et SCE-UA (GR4H).

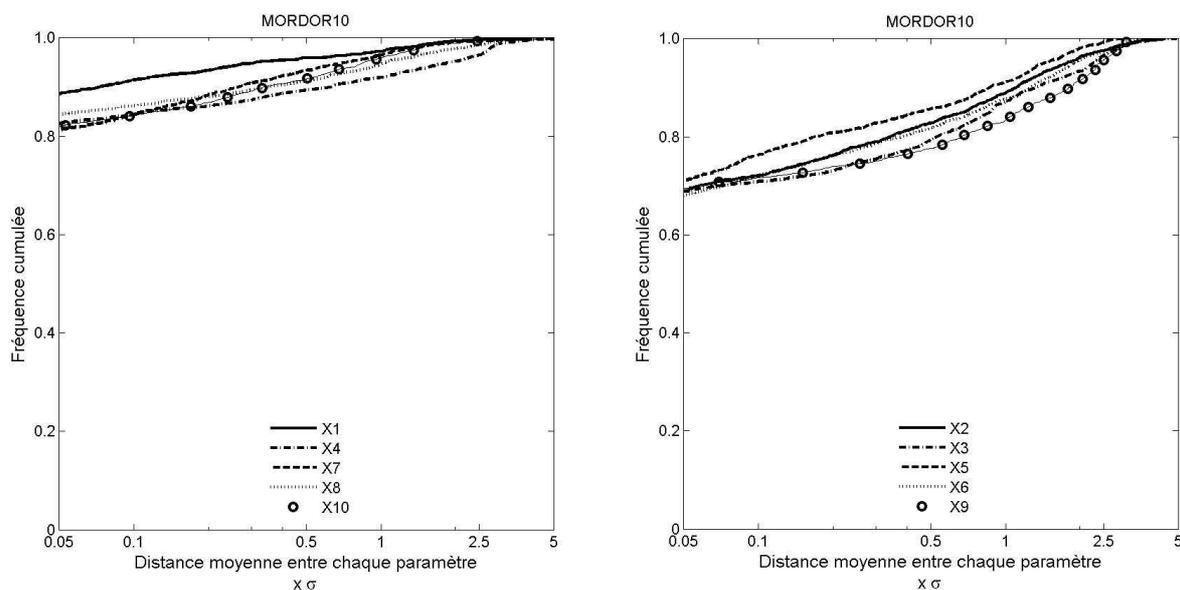


Figure 4-11 : Distribution cumulée des distances entre chaque paramètre trouvés par les deux méthodes d'optimisation PAP-GR et SCE-UA (MORDOR10).

Pour le modèle MORDOR10 les résultats sont différents. En effet, la Figure 4-11 montre que certains paramètres sont bien définis alors que d'autres ne le sont pas du tout. Par exemple, les paramètres X_1 , X_4 , X_7 , X_8 et X_{10} sont distants de moins de 5% de l'écart-type de leur distribution dans 80% à 90% des cas. Les autres paramètres, X_2 , X_3 , X_5 , X_6 et X_9 , ne sont distants de 5% de l'écart-type que dans 70% des cas. Cette diversité des vecteurs de paramètres optimisés met en évidence (1) la complexité de la surface de réponse du modèle MORDOR10 où de nombreux optima

existent et (2) la faible définition de certains des paramètres (de certaines des fonctions ?) de ce modèle, auxquels le modèle semble être faiblement sensible. Malgré les bonnes performances du modèle MORDOR10, ces caractéristiques sont celles d'un modèle dont certaines composantes de la structure sont mal définies.

Les résultats ci-dessus ont mis en évidence l'existence d'optima locaux, le caractère équifinal de certains vecteurs de paramètres et/ou la mauvaise définition de certains paramètres.

On peut donc se demander quel est le degré d'équifinalité des différents modèles. Pour cela, nous avons étudié la distance entre deux vecteurs de paramètres dont les performances en calage étaient égales (même critère C_{2M} à ± 0.1 points).

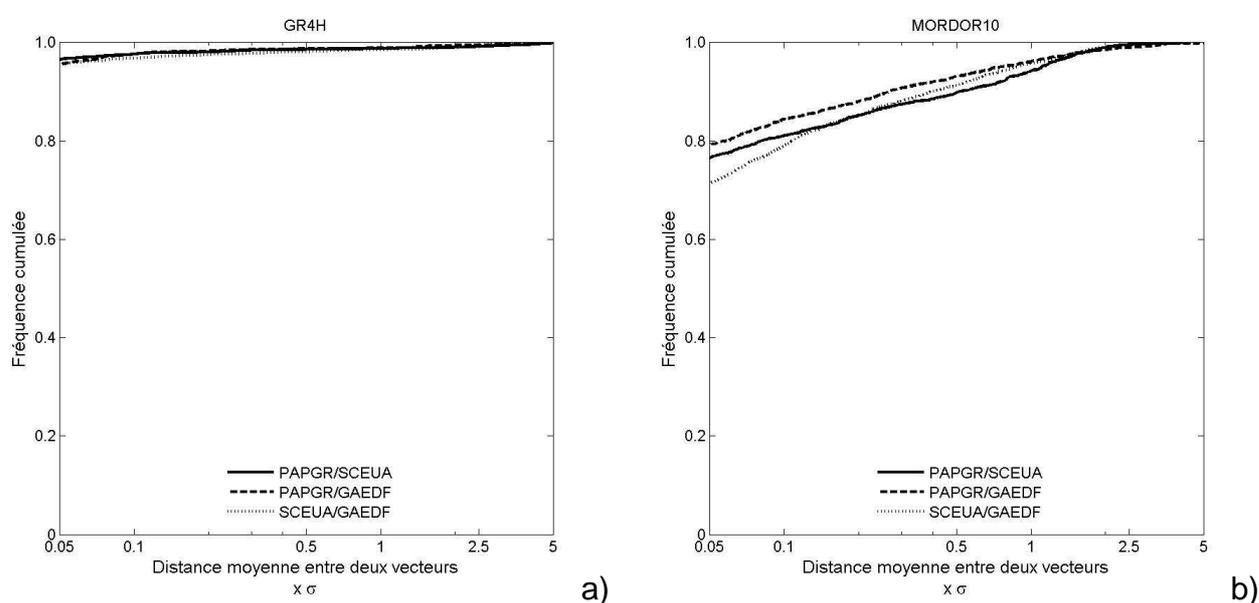


Figure 4-12 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par les deux méthodes d'optimisation PAP-GR et SCE-UA, dont les performances sont égales. a) GR4H et b) MORDOR.

Pour le modèle GR4H, la Figure 4-12 a) indique que des vecteurs équifinaux appartiennent à la même région de l'espace, dans plus de 95% des cas. En revanche, pour le modèle MORDOR10, la Figure 4-12 b) indique que des vecteurs équifinaux sont les mêmes dans seulement 70% à 80% des cas. Ce résultat indique que l'on peut s'attendre à ce que le modèle GR4H ait une surface de réponse présentant un optimum global identifiable, alors que le modèle MORDOR10 a sa surface de réponse caractérisée par l'existence de plusieurs optima locaux équivalents et assez distants.

4.6 Quelle précision dans la recherche de l'optimum de la surface de réponse ?

Les résultats obtenus précédemment soulèvent de nombreuses remarques et questions. Dans un premier temps, il a été clairement démontré que la méthode locale PAP-GR est au minimum aussi performante (voire meilleure) que les méthodes globales. Ce résultat, à contre-courant de la tendance actuelle en matière d'optimisation, signifie tout simplement que les médiocres résultats obtenus par les modèles hydrologiques proviennent rarement de la mauvaise définition des vecteurs de paramètres, mais plutôt de la mauvaise représentation de la relation pluie-débit par un modèle hydrologique (quel que soit le niveau de complexité de la méthode d'optimisation utilisée). Ces résultats confirment les conclusions de *Gan et Biftu* [1996] : *“In conclusion, other than using an advanced optimization technique, a successful application of a CRR model also requires models with sound physical bases, adequate model structure, model parameters that are readily identifiable, and representative data that have minimal measurement errors”* et celles de *Thyer et al.* [1999]: *“Nevertheless, the results suggest that SCE-UA should not be viewed as a panacea capable of satisfactorily calibrating poor CRR model with unrepresentative and error-laden data. Even the most powerful optimization algorithm will experience difficulties under such circumstances.”*.

Ces résultats mettent également en évidence que le plus important n'est pas de localiser l'optimum de la surface de réponse avec une grande précision, mais plutôt de se trouver dans la région de cet optimum. En effet, les paramètres des modèles hydrologiques restent sensibles aux caractéristiques hydro-climatiques des périodes de calage et il en résulte que lorsqu'un optimum est trouvé, rien ne garantit qu'il soit encore un optimum dans des conditions hydro-climatiques différentes. Si on imagine que la surface de réponse ressemble à une montagne, les méthodes globales ont la puissance nécessaire pour se trouver presque toujours sur le sommet de cette montagne, alors qu'il est seulement nécessaire et suffisant de se trouver sur la montagne, pas trop loin du sommet, ce que permet la méthode locale. Si par contre, la surface de réponse ressemble à un ensemble de montagnes, il est vain de chercher celle qui culmine le plus haut. Il vaudrait mieux remettre en cause l'architecture du modèle et notamment le nombre de paramètres à optimiser, sans lui faire perdre de son efficacité (Chapitre 6).

Pour tester s'il n'est pas important de localiser avec exactitude l'optimum de la surface de réponse, mais plutôt de se trouver dans la région de cet optimum, nous avons réalisé une étude de sensibilité des vecteurs de paramètres optimisés. Pour cela nous avons effectué le test suivant :

- Phase 1 : Optimisation du vecteur $\vec{V}_N^* = [X_1^*, \dots, X_N^*]$
- Phase 2 : Tirage aléatoire de m vecteurs $(\vec{V}_N)_{k=1,m}$ situés dans la région du vecteur optimal \vec{V}_N^* . Les vecteurs $(\vec{V}_N)_{k=1,m}$ sont définis de la manière suivante :
 Pour $i=1,N$, $X_i = X_i^* + U \times \alpha \cdot \sigma_{X_i}$
 où, N est le nombre de paramètres optimisés, U est un tirage aléatoire dans une loi uniforme sur l'intervalle $[-1,1]$, α est le taux de déviation du point aléatoire, exprimé en fonction de l'écart-type de la distribution de chaque paramètres (σ_{X_i}), autour de cet optimum.
- Phase 3 : analyse des performances des m vecteurs aléatoires sur la période de calage et la période de validation.

Nous avons réalisé cette étude de sensibilité pour les 2093 périodes de calage, pour lesquelles nous avons généré 50 vecteurs de paramètres aléatoires situés autour de l'optimum, avec des déviations représentant 5%, 10%, 15% et 20% de l'écart-type de la distribution des paramètres.

Tableau 4-10 : Ecart-types σ_{X_i} des distributions des paramètres transformés de GR4H et MORDOR10 (valeurs transformées entre]-10, 10[).

Modèle	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
GR4H	1.6	1.3	3.7	4.5						
MORDOR10	4.3	4.3	3.9	3.3	6.3	5.5	1.5	3.7	7.7	3.9

Le Tableau 4-10 présente les valeurs des écart-types des dispersions des paramètres des modèles GR4H et MORDOR10. Ce tableau montre la faible dispersion des paramètres de GR4H et la plus forte dispersion des paramètres de MORDOR10. Cette différence vient d'une part de la parcimonie du modèle GR4H, qui contraint les paramètres à être bien définis, des problèmes structuraux de MORDOR10 (interactions entre paramètres, faible sensibilité à certaines fonctions, etc.). Les transformations mathématiques des paramètres contraignent également leurs domaines de variation. Par exemple, dans l'analyse de sensibilité des vecteurs de paramètres optimisés, un taux de déviation de 10% signifie, pour le modèle GR4H, que les quatre paramètres sont tirés aléatoirement autour de leur optimum, avec respectivement un intervalle de +/- 0.16, +/- 0.13, +/- 0.37 et +/- 0.45. Pour MORDOR, un taux de déviation de 10%, signifie que les paramètres peuvent varier de +/- 0.15 à +/- 0.77 et varient en moyenne de +/- 0.44 (ce qui est très grand par rapport à leurs domaines de variation).

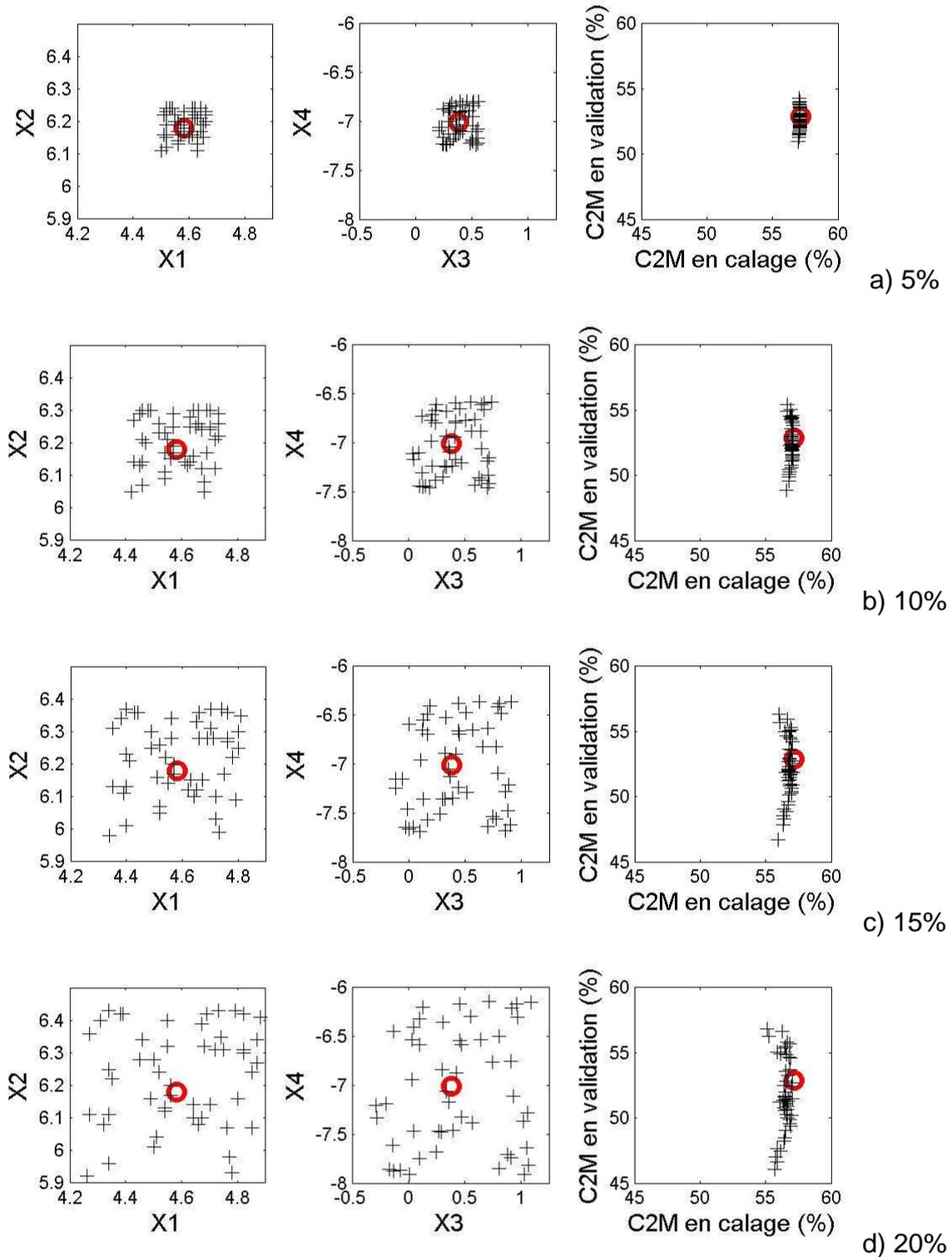


Figure 4-13: Analyse de sensibilité autour de l'optimum de la réponse de surface. Coordonnées de l'optimum (rond) et des vecteurs aléatoires (croix), comparaison des critères en calage et en validation. Taux de déviation a) :5%, b) 10%, c) 15%, d) 20%. Exemple du bassin versant de La Doller à Sewen (A1202030), période de calage : 1995-1996, modèle GR4H.

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

La Figure 4-13 représente les résultats typiques d'une analyse de sensibilité de l'optimum de la surface de réponse, réalisée pour une période de calage. Cette figure illustre la dispersion des vecteurs aléatoires autour de l'optimum de la surface de réponse, au fur et à mesure que le taux de déviation augmente.

Néanmoins, même pour un faible taux de déviation (5%, ce qui représente la distance entre 70% à 80% des vecteurs optimisés selon deux méthodes différentes, pour le modèle GR4H), on remarque que les performances en calage varient assez peu (diminution de seulement 1 à 2 points) alors qu'elles peuvent varier de 2 à 4 points en validation. Par ailleurs, plus on augmente le taux de dispersion, plus les variations de critères C_{2M} en validation sont grandes, alors qu'elles restent limitées en calage.

Cette analyse de sensibilité représente un grand nombre de tests et la génération de plus de 100 000 vecteurs aléatoires, pour chaque modèle sur l'ensemble de l'échantillon. Le Tableau 4-11 synthétise les résultats de cette étude de sensibilité, réalisée avec les modèles GR4H et MORDOR10, pour quatre taux de déviation.

Tableau 4-11 : Performances moyennes et écart-types des 2093 périodes de calage et des 2093 groupes de 50 vecteurs aléatoires, en calage et en validation, pour les modèles GR4H et MORDOR10.

Modèle	Moyennes	Calage		Validation		
		C_{2M}	Ecart-type	C_{2M}	Ecart-type	$[C_{2M(max)} - C_{2M(min)}]$
GR4H	2093 périodes	57.3	26.3	33.7	29.8	X
	$\alpha.\sigma_{V_i} = 5\%$	56.1	27.6	33.4	29.8	10.0
	$\alpha.\sigma_{V_i} = 10\%$	53.8	28.8	32.0	30.0	20.0
	$\alpha.\sigma_{V_i} = 15\%$	51.2	30.2	30.0	30.3	30.3
	$\alpha.\sigma_{V_i} = 20\%$	48.5	30.9	28.9	30.5	41.2
MORDOR10	2093 périodes	60.0	30.6	36.2	28.2	X
	$\alpha.\sigma_{V_i} = 5\%$	58.0	32.6	35.8	28.2	10.3
	$\alpha.\sigma_{V_i} = 10\%$	55.6	33.5	35.1	28.0	19.4
	$\alpha.\sigma_{V_i} = 15\%$	52.9	34.2	33.9	27.8	28.5
	$\alpha.\sigma_{V_i} = 20\%$	49.7	34.8	32.5	27.6	36.9

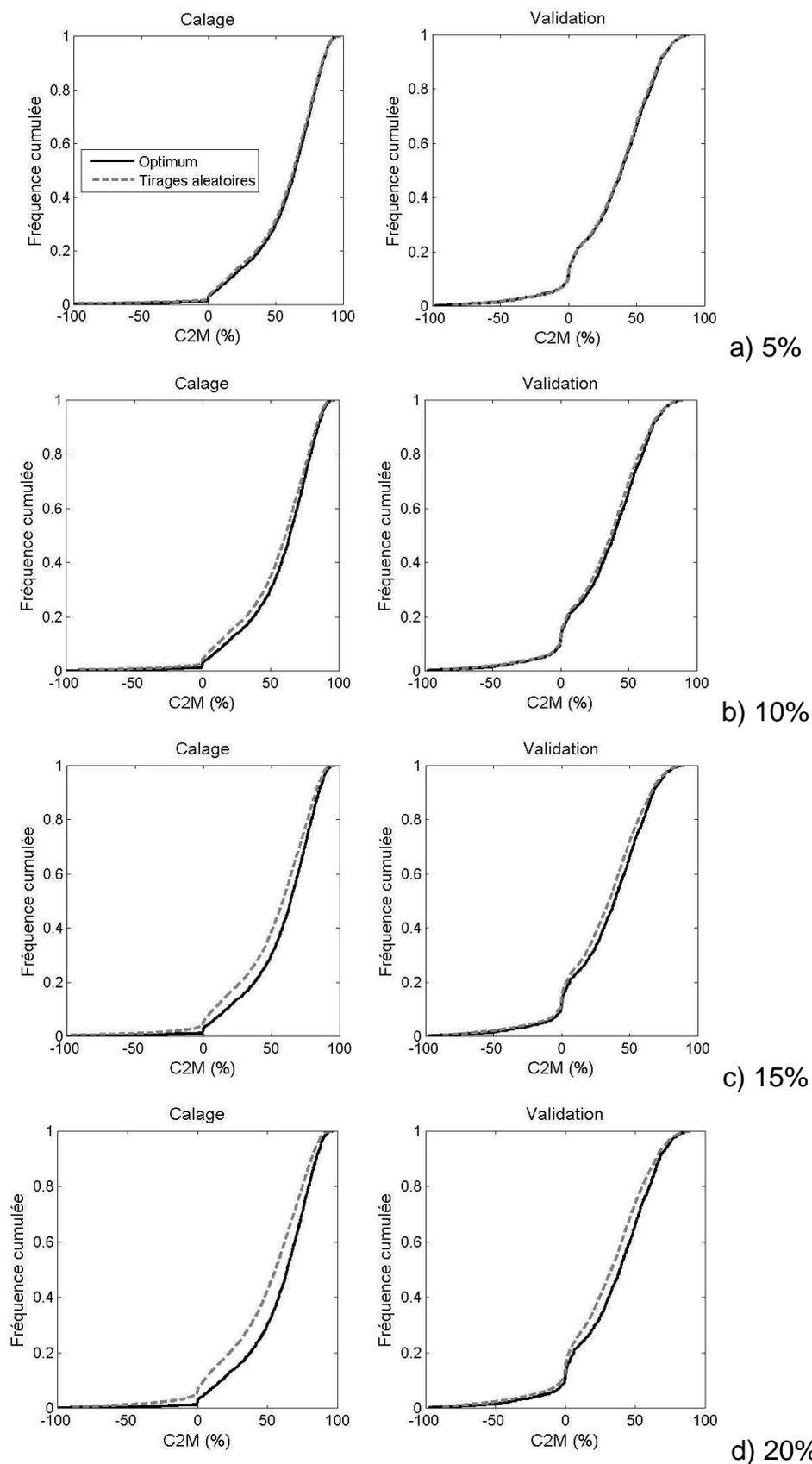


Figure 4-14 : Distribution des performances obtenues sur les 2093 périodes de calage, avec le vecteur optimum (ligne noire) et les 50 vecteurs générés de manière aléatoire pour chaque période de calage. Taux de déviation a) : 5%, b) : 10%, c) : 15%, d) : 20% (GR4H).

Chapitre 4 : Optimisation automatique des modèles pluie-débit : intercomparaison exhaustive des performances de méthodes locales et globales

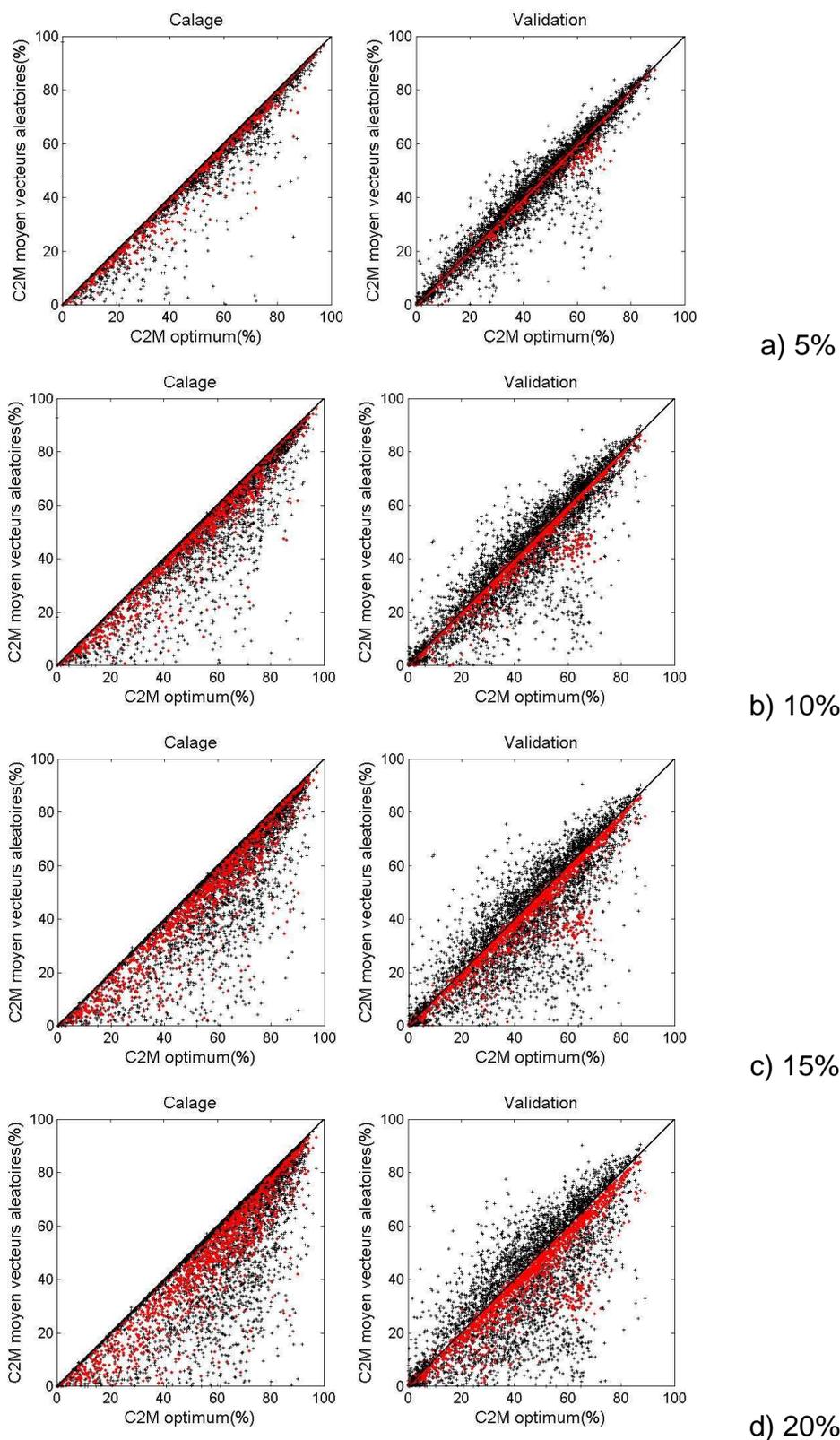


Figure 4-15 : Synthèse de l'analyse de sensibilité des vecteurs optimum des 2093 périodes de calage, pour le modèle GR4H. Comparaison de la moyenne des performances obtenues avec les 50 vecteurs aléatoires, par rapport au critère obtenu en calage ou en validation. Les croix représentent le minimum ou le maximum des valeurs de critères obtenues avec 50 vecteurs aléatoires. Taux de déviation a) : 5%, b) : 10%, c) : 15%, d) : 20%.

Pour le modèle GR4H et le modèle MORDOR10, ce tableau montre que lorsque le taux de déviation reste limité (5%), les performances moyennes en calage diminuent légèrement et les performances en validation ne sont pas affectées significativement. Ce résultat est mis en évidence pour le modèle GR4H par la Figure 4-14 a), puisque les distributions des performances des vecteurs optimaux et des vecteurs aléatoires sont presque superposées en calage et en validation. Pour les deux modèles, les écart-types des distributions des performances en calage et en validation sont faiblement affectés. Ce résultat signifie donc que, pour un taux de déviation de 5% autour de l'optimum, les groupes de 50 vecteurs aléatoires ont en moyenne la même performance que l'optimum en validation ! Ce qui est intéressant dans ce résultat, c'est que les deux modèles semblent avoir la même sensibilité autour de leur optimum, alors qu'un taux de déviation de 5% pour MORDOR10 signifie un espace autour de l'optimum beaucoup plus grand que celui de GR4H.

Lorsque le taux de déviation augmente, pour les deux modèles, les performances moyennes en calage diminuent fortement, les écart-types de la distribution des performances en calage augmentent. En revanche, en validation, les performances moyennes diminuent un peu moins fortement et les écart-types de la distribution des performances ne sont presque pas affectés. Les Figure 4-14 b), c), d) montrent, pour le modèle GR4H que la distribution des performances obtenues avec les différents taux de déviation s'éloigne progressivement de celle des performances optimales.

La Figure 4-15 confirme ces résultats et montre pour le modèle GR4H les performances des optima (en calage et en validation), comparées à la moyenne des performances des 50 vecteurs aléatoires générés pour chaque période. La performance minimale et maximale de chaque groupe de vecteurs aléatoires a été également indiquée pour illustrer la variabilité des performances. Par rapport aux résultats précédents, cette figure montre que pour une même période de calage, la variabilité des résultats en validation augmente fortement avec le taux de déviation des vecteurs aléatoires, de +/- 5 points à +/- 20 points par rapport à la performance de l'optimum.

Ces résultats semblent confirmer nos doutes. En effet, l'estimation du vecteur optimal de paramètres d'un modèle, pour un bassin et une période, est très dépendante des caractéristiques hydrologiques de cette période. De plus, la variabilité hydrologique naturelle et les erreurs des modèles hydrologiques, sont très grandes. Pour ces raisons, il ne semble pas nécessaire de situer avec trop de précision l'optimum de la surface de réponse.

Ces tests, où nous avons générés un grand nombre de vecteurs aléatoires autour des valeurs optimales, confirment que les performances du modèle en validation sont faiblement affectées lorsqu'on tire aléatoirement le vecteur de paramètres dans la région proche de l'optimum (écart-type de 5% à 10% de la distribution des paramètres)

4.7 Synthèse

Nous avons présenté dans ce chapitre une comparaison des performances de trois méthodes d'optimisation des modèles pluie-débit. Ces méthodes utilisent une stratégie de recherche de l'optimum qui est, soit locale (PAP-GR), soit globale (SCE-UA et GA-EDF). Ces trois méthodes d'optimisation ont été utilisées pour optimiser les vecteurs de paramètres de quatre modèles ayant de 4 à 10 paramètres, sur les 313 bassins versants de la base de données. Les performances des vecteurs optimisés par les différentes méthodes ont été comparées, pour les phases de calage et de validation. Les distances entre les différents vecteurs ont également été comparées.

Les tests effectués montrent que :

- les méthodes globales ne sont que marginalement plus efficaces en calage que la méthode locale testée, alors qu'elles sont beaucoup moins efficaces. En effet, pour le modèle le plus simple, les performances en calage sont les mêmes dans plus de 80% des cas ; pour le modèle le plus compliqué, elles restent les mêmes dans 50% des cas. Ces résultats montrent que le raffinement des méthodes globales n'apporte finalement qu'un faible progrès des performances en calage, pour une recherche dans l'espace des paramètres nécessitant d'augmenter le nombre d'itérations de un à deux ordres de grandeur.
- les performances en validation des vecteurs de paramètres optimisés par la méthode PAP-GR sont au moins aussi bonnes, voire meilleures, que celles des vecteurs de paramètres optimisés par les méthodes globales. Ces résultats indiquent que les vecteurs trouvés par les méthodes globales sont généralement moins robustes que ceux trouvés par la méthode locale. En effet, les différences de performances entre calage et validation sont supérieures. Ce phénomène vient certainement de la puissance de ces méthodes qui sont capables de localiser un optimum dans n'importe quelle région de l'espace des paramètres, au détriment de la stabilité des paramètres, alors que la méthode locale est efficace lorsque l'optimum de la surface de réponse est bien marqué. Ce manque de robustesse vient peut-être de la fonction objectif utilisée. Néanmoins, le fait est qu'en l'état actuel

des choses, la puissance des algorithmes globaux pousse le modèle au sur-calage : les vecteurs de paramètres identifiés par ces méthodes sont généralement sur-adaptés aux conditions hydrométéorologiques de cette période.

- La méthode d'optimisation globale utilisant une stratégie de recherche par algorithmes génétiques, GA-EDF, donne d'excellents résultats. Elle est très efficace (mais très peu efficiente), contrairement aux résultats reportés dans la littérature sur les méthodes d'optimisation basée uniquement sur des algorithmes génétiques [Tanakamaru, 1995; Cooper et al., 1997; Franchini et Galeati, 1997]. Cette méthode est encore peu efficiente, mais elle pourrait être améliorée facilement de manière empirique avec notre base de données.
- Enfin, il n'est ni nécessaire, ni très utile de localiser avec précision l'optimum de la surface de réponse ! En effet, avoir localisé avec précision l'optimum de la surface de réponse ne garantit pas que cet optimum le soit encore au cours de la validation. Lorsqu'on se situe dans la région proche de l'optimum, il est équiprobable de trouver un vecteur qui a de meilleures ou moins bonnes performances en validation que celles du vecteur optimum obtenu en calage.

Ces résultats permettent d'affirmer que, contrairement à ce que de nombreux hydrologues pensent, une méthode locale est capable de trouver des optima au moins aussi bons que ceux trouvés par des méthodes globales raffinées, de manière très efficiente et ce quel que soit le nombre de paramètres de la surface de réponse. Ces résultats rappellent le constat amer de Klemeš [1986a] sur l'orientation technologique de certains domaines de l'hydrologie. En effet, même si le développement de méthodes telles que le SCE-UA ou d'autres méthodes *compliquées* est intellectuellement stimulant et leur utilisation particulièrement *chic*, lorsque la structure du modèle est mal définie, ses performances ne seront pas améliorées et les problèmes d'équifinalité ne seront pas réduits ! Ce constat nous ramène à Gupta et Sorooshian [1983] qui concluaient il y a plus de 20 ans « *the problems are not all due to inefficiencies in the calibration techniques used but are caused by the manner in which the model is structurally formulated* ».

Enfin, ces tests renforcent notre confiance dans la méthode PAP-GR. Son efficacité, même pour des modèles complexes, indique qu'elle peut être utilisée pour la comparaison de modèles et qu'elle ne biaisera pas les résultats de la comparaison. C'est donc cette méthode que nous avons choisi d'utiliser dans la comparaison de modèles présentée au chapitre suivant.

DEUXIEME PARTIE :

COMPARAISON ET DEVELOPPEMENT DE MODELES PLUIE-DEBIT AU PAS DE TEMPS HORAIRE

Chapitre 5

« Bien des années plus tard, face au peloton d'exécution, le colonel Aureliano Buendia devait se rappeler ce lointain après-midi au cours duquel son père l'emmena faire connaissance avec la glace. Macondo était alors un village d'une vingtaine de maisons en glaise et en roseaux, construites au bord d'une rivière dont les eaux diaphanes roulaient sur un lit de pierres polies, blanches, énormes comme des œufs préhistoriques. Le monde était si récent que beaucoup de choses n'avaient pas encore de nom et pour les mentionner, il fallait les montrer du doigt. Tous les ans, au mois de mars, une famille de gitans déguenillés plantait sa tente près du village et, dans un grand tintamarre de fifres et de tambourins, faisait part des nouvelles inventions. Ils commencèrent par apporter l'aimant. Un gros gitan à la barbe broussailleuse et aux mains de moineau, qui répondait au nom de Melquiades, fit en public une truculente démonstration de ce que lui-même appelait la huitième merveille des savants alchimistes de Macédoine. »

Gabriel Garcia Marquez
Cent ans de solitude, 1967.

CHAPITRE 5 Comparaison des performances de modèles pluie-débit au pas de temps horaire

5.1 Introduction

L'objectif de ce chapitre est de présenter les performances de 20 modèles pluie-débit testés au pas de temps horaire. Ces 20 structures de modèle (voir détails en Annexe 1) proviennent essentiellement des 38 structures de modèles déjà testées par Perrin [2000], au pas de temps journalier. Dans un premier temps nous avons repris ces 38 structures et éliminé celles qui donnaient de trop mauvais résultats au pas de temps horaire. A ces structures déjà testées, nous avons ajouté la structure du modèle MORDOR10 qui n'avait jamais été testée exhaustivement et d'autres structures développées pendant cette thèse.

Dans ce chapitre, nous procédons à l'analyse des performances, de la robustesse et de la fiabilité des modèles testés. Ensuite, nous essayons de mettre en évidence une éventuelle complémentarité entre certaines structures de modèles et nous nous posons la question de l'existence de facteurs physiographiques pertinents pour la discrimination des performances de certains modèles.

5.2 Synthèse et présentation des résultats

Les 20 structures de modèles présentées ont été testées sur l'échantillon de 313 bassins versants, ce qui représente 2093 périodes de calage et de validation, en suivant une procédure de *split sample test*. Pour chaque modèle, nous disposons d'une grande quantité de résultats qui sont ses performances mesurées pour chaque période, en calage et en validation, avec différents critères de performances.

Pour chaque modèle, la procédure de test est représentée à la Figure 5-1 et le processus d'agrégation des résultats (Figure 5-2) est le suivant :

- Représentation des performances obtenues pour chaque période de calage (resp. de validation),
- Histogrammes des performances par périodes,
- Histogrammes des performances par bassins versants,
- Distribution des performances par bassins versants,
- Boxplot de la distribution des performances (moyenne, médiane, quantiles).

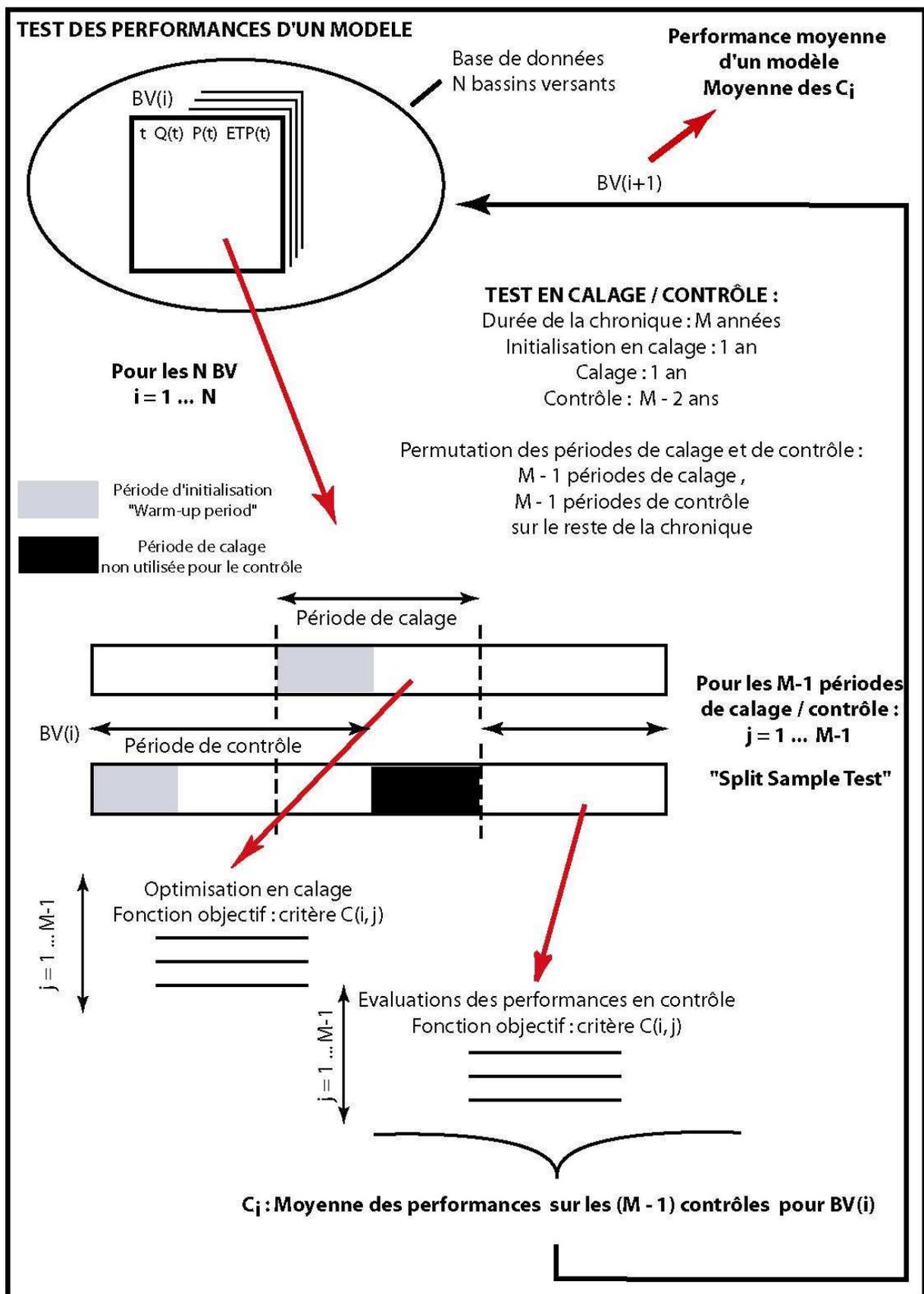
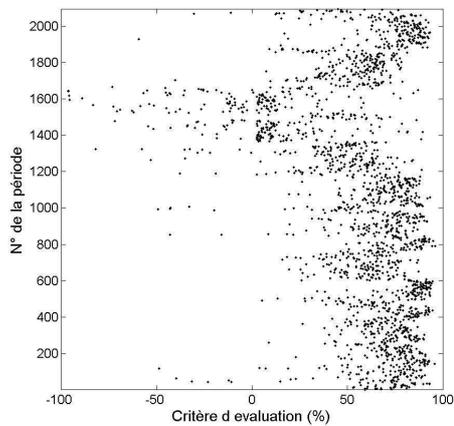


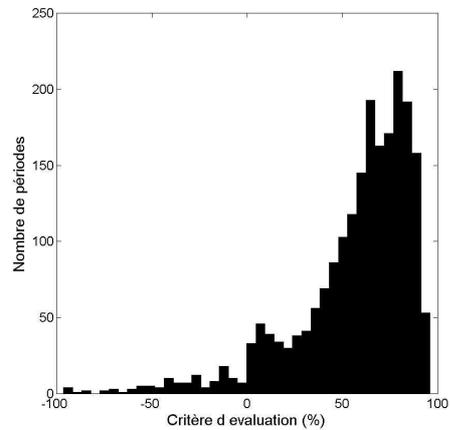
Figure 5-1 : Procédure de calage-contrôle effectuée pour chaque modèle sur l'ensemble des bassins versants de la base de données.

[1] : Résultats bruts sur l'ensemble des périodes



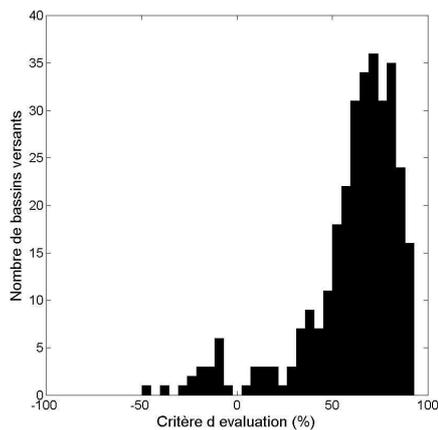
a)

[2] : Histogramme des résultats par période



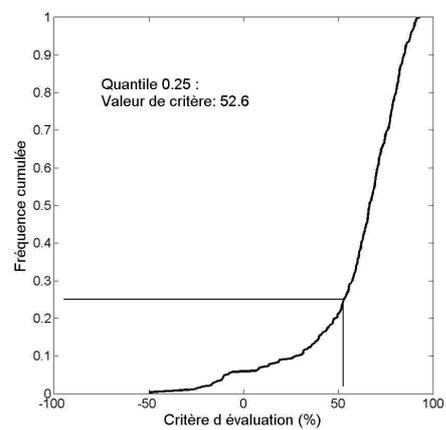
b)

[3] : Histogramme des résultats par bassin versant



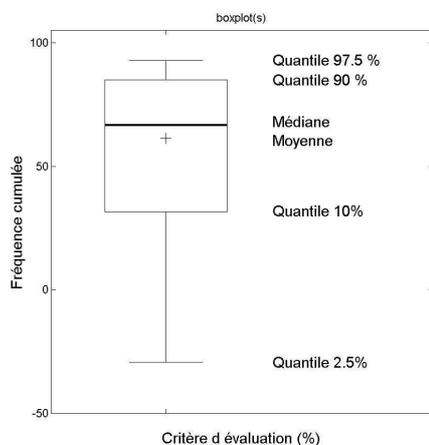
c)

[4] : Distribution des résultats par bassins versants



d)

[5] : Boxplot de la distribution des résultats par bassins versants



e)

Figure 5-2 : Exemple de représentation et d'agrégation des résultats d'un modèle sur l'ensemble de l'échantillon de bassins versants.

Nous agrégeons les performances de plusieurs périodes d'un même bassin afin de donner le même poids à tous les bassins, quelle que soit la durée de leur chronique de données.

La base de l'analyse des résultats d'un modèle est la distribution de ses performances obtenues sur chaque bassin versant (Figure 5-2 d, Figure 5-3) . Une distribution est obtenue en classant les performances d'un modèle par ordre croissant. Comme sur l'exemple de la Figure 5-2 d), l'axe des ordonnées représente la fréquence cumulée au non dépassement. Ainsi, la lecture d'un quantile correspond à la proportion des tests dont le critère d'évaluation est inférieur à la valeur donnée par la courbe : sur cet exemple, le modèle testé a ses performances inférieures à 52.6 points, pour 25% des bassins versants. Lorsqu'on représente les résultats avec des courbes de distribution des performances, les meilleurs modèles ont leurs distributions vers la droite du graphique, ce qui veut dire que, pour une fréquence au non dépassement donnée, le critère correspondant est plus élevé.

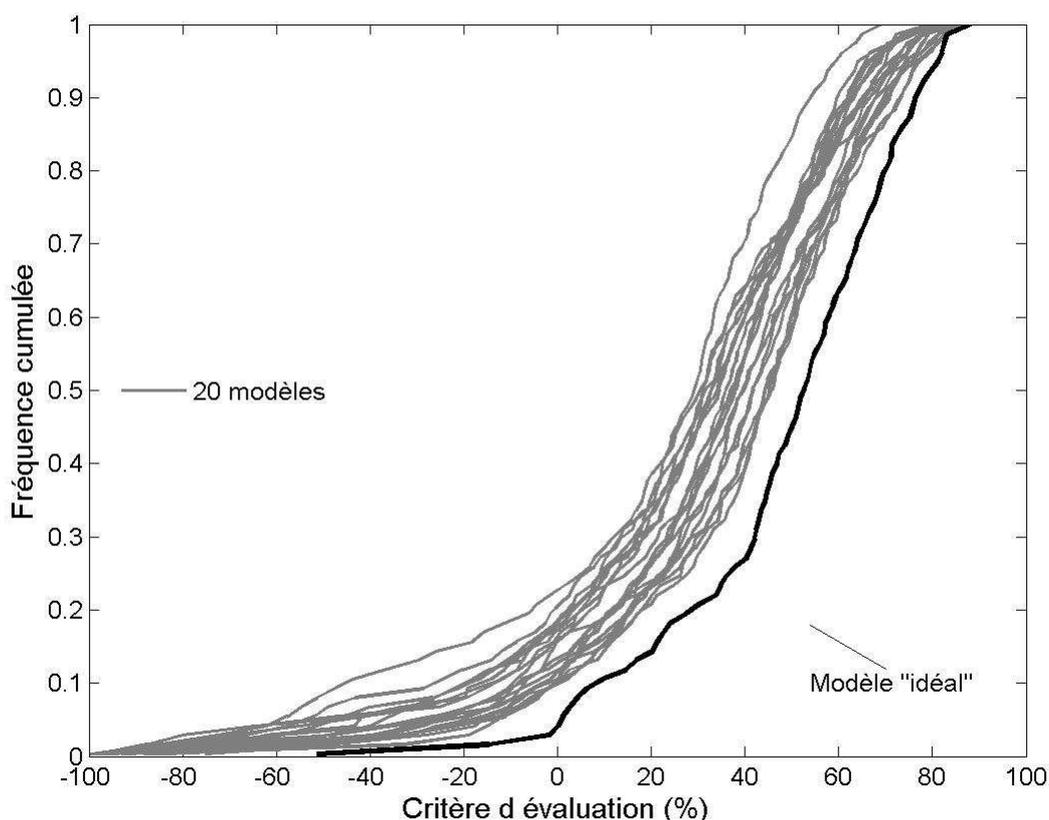


Figure 5-3 : Exemple de distribution expérimentale des résultats des 20 modèles, testés sur 313 bassins versants. Chaque courbe grise représente les résultats d'un modèle, la courbe noire représente les résultats du modèle 'idéal', c'est-à-dire le modèle qui obtiendrait sur chaque bassin la performance du meilleur modèle.

Néanmoins, comme l'illustre la Figure 5-3, la comparaison des performances obtenues par un grand nombre de modèles (via leurs distributions) est délicate à partir du moment où leurs distributions se croisent. Entre deux modèles, il faut toujours faire un compromis entre le choix d'une bonne robustesse (faible nombre d'échecs) ou de bonnes performances en moyenne. Ainsi, pour résumer ces distributions, nous utiliserons des représentations sous forme de boxplot (Figure 5-2 e). Un boxplot synthétise la distribution des performances d'un modèle en indiquant : les quantiles 2.5%, 10%, 50%, 90% et 97.5% et la moyenne des résultats. De plus, lorsque la différence entre la médiane et la moyenne est grande, cette représentation illustre également les problèmes de robustesse d'un modèle (la moyenne est affectée par de très faibles performances sur quelques bassins versants). Par la suite, nous présenterons les résultats sous forme de boxplot, classés par ordre croissant de la moyenne des résultats.

5.3 Performances des modèles testés

Dans cette partie, nous présentons les performances des 20 modèles (Tableau 5-1), testés sur les 313 bassins versants de la base de données. L'évaluation des performances des modèles hydrologiques se fait en deux étapes. Dans un premier temps, on évalue les performances au cours du calage, ce qui indique la capacité du modèle à reproduire une chronique hydrologique d'apprentissage. Dans un deuxième temps, on évalue les performances au cours de la validation, ce qui indique la capacité d'un modèle à simuler le débit d'un bassin, dans des conditions hydrologiques différentes de celles de la phase d'apprentissage, sans changement de paramètres.

Les modèles ont été calés en utilisant pour fonction objectif le critère $C_{2M}(Q)$ ou le critère $C_{2M}(\sqrt{Q})$ (que nous appellerons $C_{2M}(RQ)$) et leurs performances ont été évaluées en validation également avec ces deux critères*. Les limites de la boîte des boxplots indiquent également le domaine de variation de 80% des performances des modèles. Ces figures comportent la valeur du C_{2M} moyen à gauche et également du NS moyen à droite. Cette double indication permet aisément de faire référence aux niveaux de performance habituellement obtenus en utilisant le critère de NS. Il faut néanmoins noter que ce NS moyen n'est pas la moyenne des NS, mais le critère de NS correspondant à la valeur du C_{2M} moyen, déduit de la relation entre NS et C_{2M} (Eq. [3.6]).

* Dans l'évaluation des performances d'un modèle, les transformations sur la variable cible, Q , sont courantes. Elles permettent d'insister sur les forts débit (Q) ou les débits les plus fréquent (RQ).

Chapitre 5 : Intercomparaison des performances de modèles pluie-débit au pas de temps horaire

Les modèles ont été calés avec la méthode PAP-GR, dont nous avons vérifié au CHAPITRE 4 la fiabilité, même pour les modèles les plus complexes. Dans un premier temps l'évaluation des performances des modèles se fait en calage. Les performances sont présentées à la Figure 5-4 ($C_{2M}(Q)$) et la Figure 5-5 ($C_{2M}(RQ)$).

Tableau 5-1 : Liste des modèles sélectionnés.

Modèle	Abréviation	Modèle	Abréviation
Institute of Hydrology lumped model [Blackie et Eeles, 1985]	BLAE	Model B [Bonvoisin et Boorman, 1992]	MODB
CREC [Cormary et Guilbot, 1973]	CREC	MORDOR6 (cette thèse)	MOR6
GARDENIA [Thiery, 1982]	GARD	MORDOR [Garçon, 1999; Paquet, 2004]	MORD
GR3H (cette thèse)	GR3H	NAM [Nielsen et Hansen, 1973]	NAM0
GR4H (cette thèse)	GR4H	PDM [Moore et Clarke, 1981]	PDM
GR5H (cette thèse)	GR5H	Bucket model [Thornthwaite et Mather, 1955]	THOR
GR4J [Perrin et al., 2003]	GR4J	TOPMO6 (cette thèse)	TOP6
HBV [Bergström et Forsman, 1973; Bergström, 1995]	HBV0	TOPMODEL [Beven et Kirkby, 1979; Beven et al., 1995]	TOP8
IHACRES [Jakeman et al., 1990]	IHAC	WAGENINGEN [Warmerdam et al., 1997]	WAGE
MARTINE [Mazenc et al., 1984]	MART	XINANJIANG [Zhao et al., 1980; Zhao et Liu, 1995]	XINJ

Quel que soit le critère de calage utilisé, les modèles obtiennent de bons niveaux de performance en calage. Les performances moyennes des modèles varient entre 50 et 70 points de C_{2M} (respectivement, 67 et 82 points de NS).

Le Tableau 5-2 indique les écarts maximaux entre les moyennes et les quantiles 0.1 et 0.9 des distributions des performances des 20 modèles. En utilisant le critère de calage $C_{2M}(Q)$, les écarts entre les modèles sont de 15 à 20 points, pour la moyenne et les quantiles 0.1 et 0.9. En utilisant le critère de calage $C_{2M}(RQ)$, les écarts entre les modèles sont d'une vingtaine de points pour la moyenne, de 16 points pour le quantile 0.9 et de 31 points pour le quantile 0.1. Cette différence d'écarts entre le quantile 0.9 et le quantile 0.1 indique la capacité des modèles à atteindre de bons niveaux de performances et quelques problèmes de calage de certains modèles sur certains bassins versants.

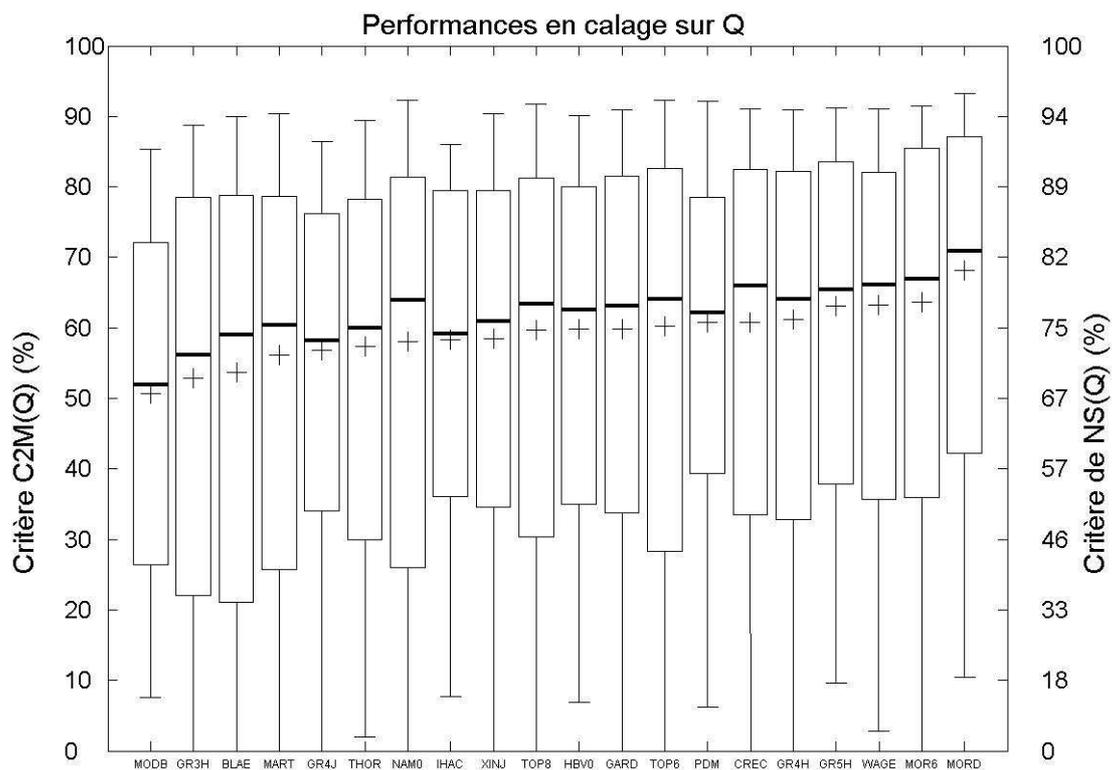


Figure 5-4 : Performances en calage, calculées sur les débits, des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$. Classement croissant selon le moyenne (+).

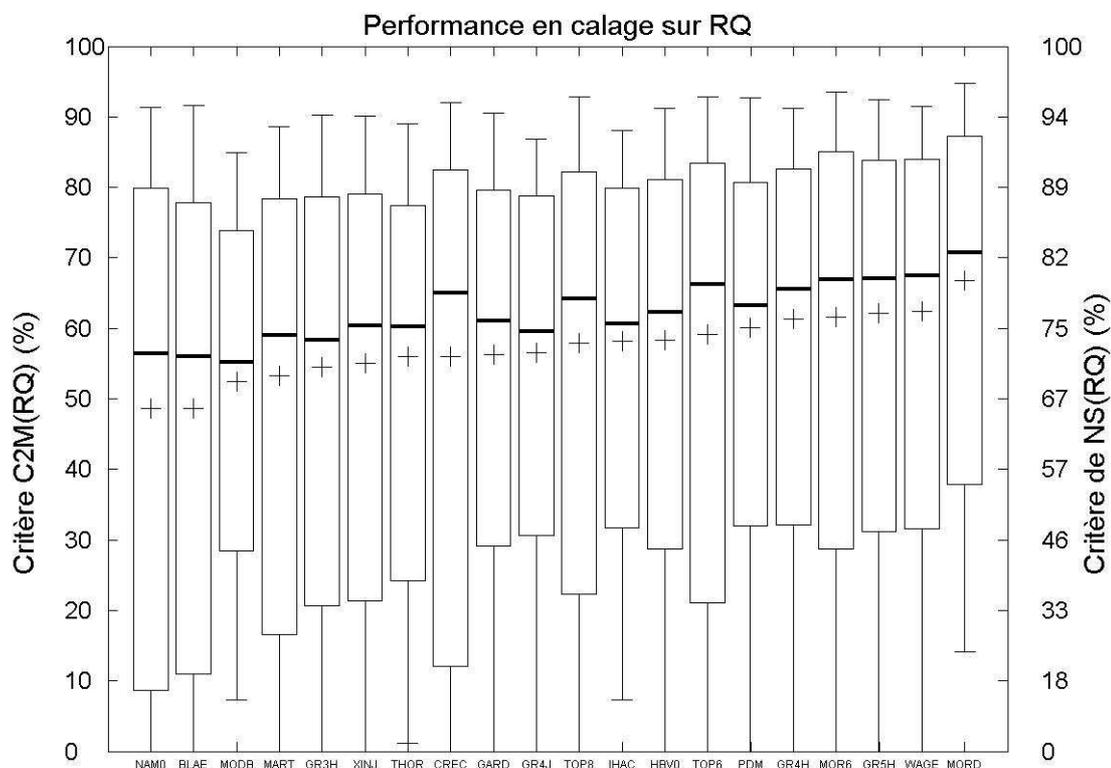


Figure 5-5 : Performances en calage, calculées sur la racine des débits, des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$. Classement croissant selon le moyenne (+).

Chapitre 5 : Intercomparaison des performances de modèles pluie-débit au pas de temps horaire

Tableau 5-2 : Moyenne de la performance moyenne et des quantiles 0.1 et 0.9, et écarts maximaux entre la moyenne et les quantiles 0.1 et 0.9 des 20 modèles testés.

	Moyenne			Ecart		
	Quantile 0.1	Moyenne	Quantile 0.9	Quantile 0.1	Moyenne	Quantile 0.9
Calage $C_{2M}(Q)$	32	59	80	21	20	15
Calage $C_{2M}(RQ)$	25	57	81	21	17	15
Calage $C_{2M}(Q)$ / Validation $C_{2M}(Q)$	-10	31	65	29	18	13
Calage $C_{2M}(Q)$ / Validation $C_{2M}(RQ)$	-21	29	65	45	18	19
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(Q)$	-5	32	64	35	19	21
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(RQ)$	-4	38	70	30	17	20

Le modèle le plus complexe a les meilleures performances en calage, mais comme l'indique la Figure 5-6, il n'y pas une classification claire des performances des modèles en fonction de leur degré de complexité (nombre de paramètres à optimiser). Cela veut dire qu'en phase de calage (apprentissage), les modèles ont a peu près la même capacité à simuler les débits, quels que soient leurs niveaux de complexité et que les modèles se différencient surtout par leur structure.

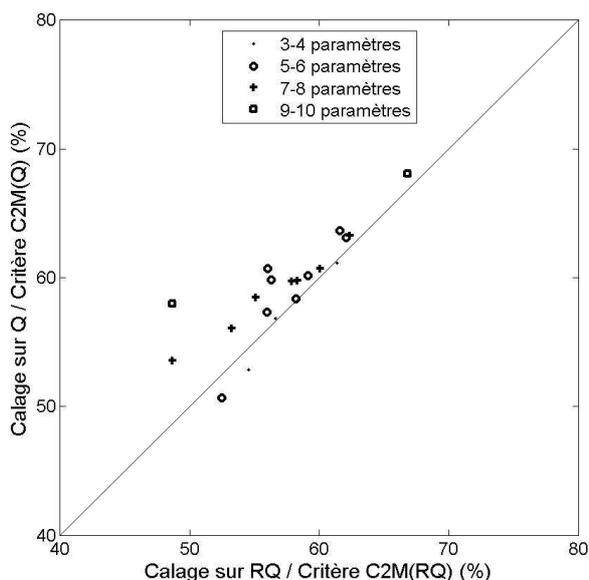


Figure 5-6 : Comparaison des performances en calage, en fonction de la fonction objectif choisie, $C_{2M}(Q)$ ou $C_{2M}(RQ)$, et de la complexité des modèles.

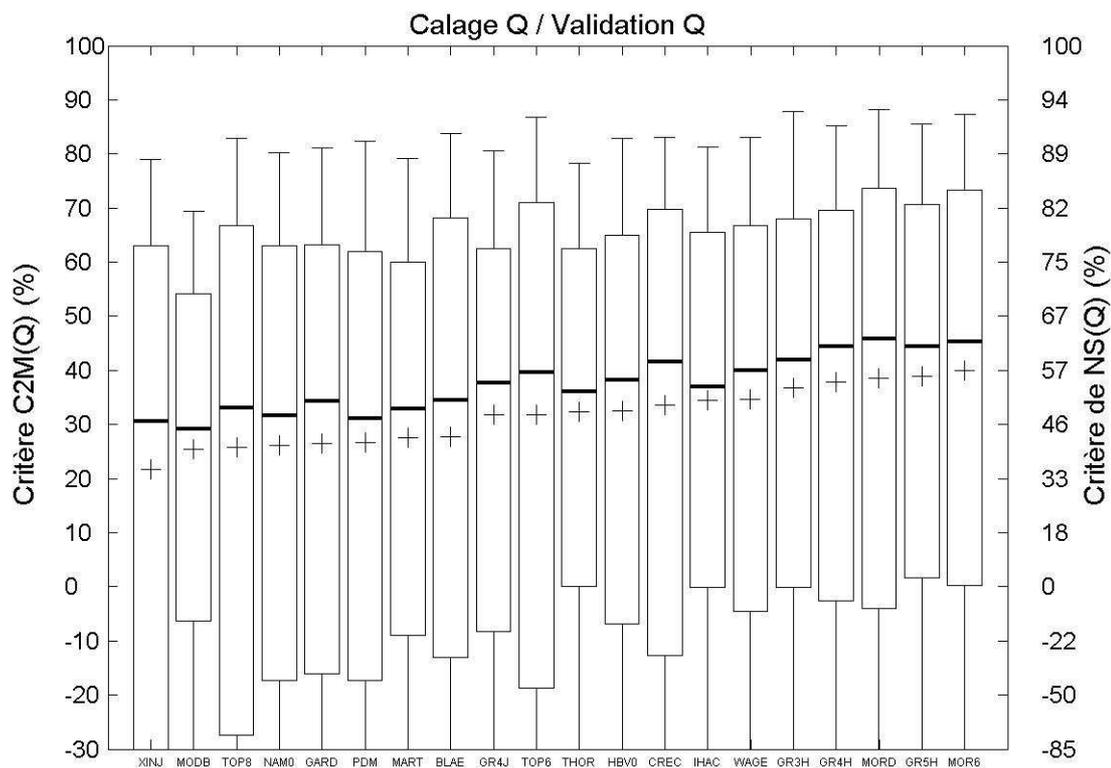


Figure 5-7 : Performances en validation des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(Q)$. Classement croissant selon le moyenne (+).

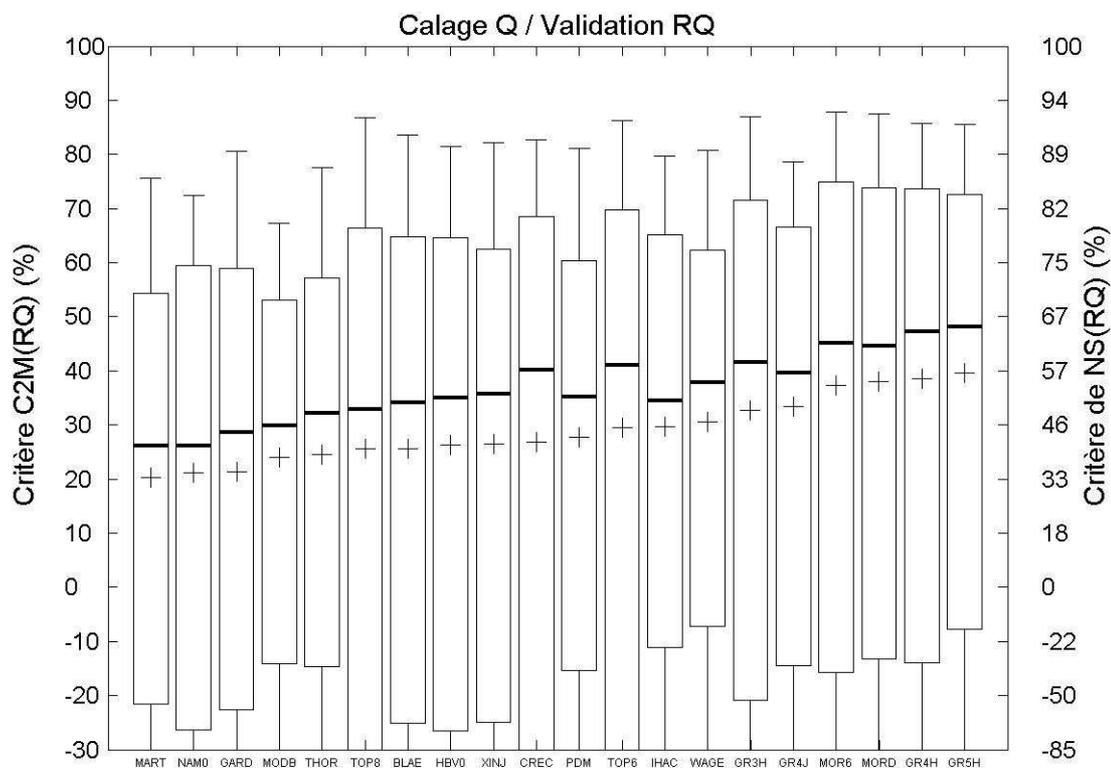


Figure 5-8 : Performances en validation des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(RQ)$. Classement croissant selon le moyenne (+).

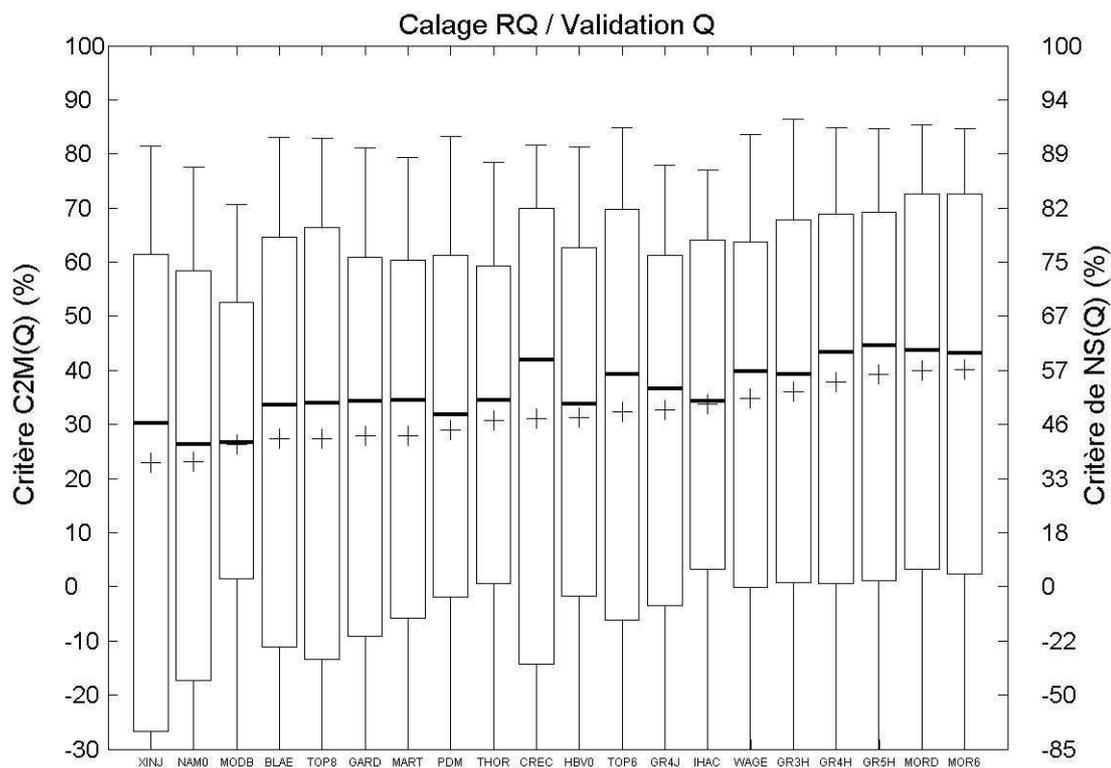


Figure 5-9 : Performances en validation des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(Q)$. Classement croissant selon le moyenne (+).

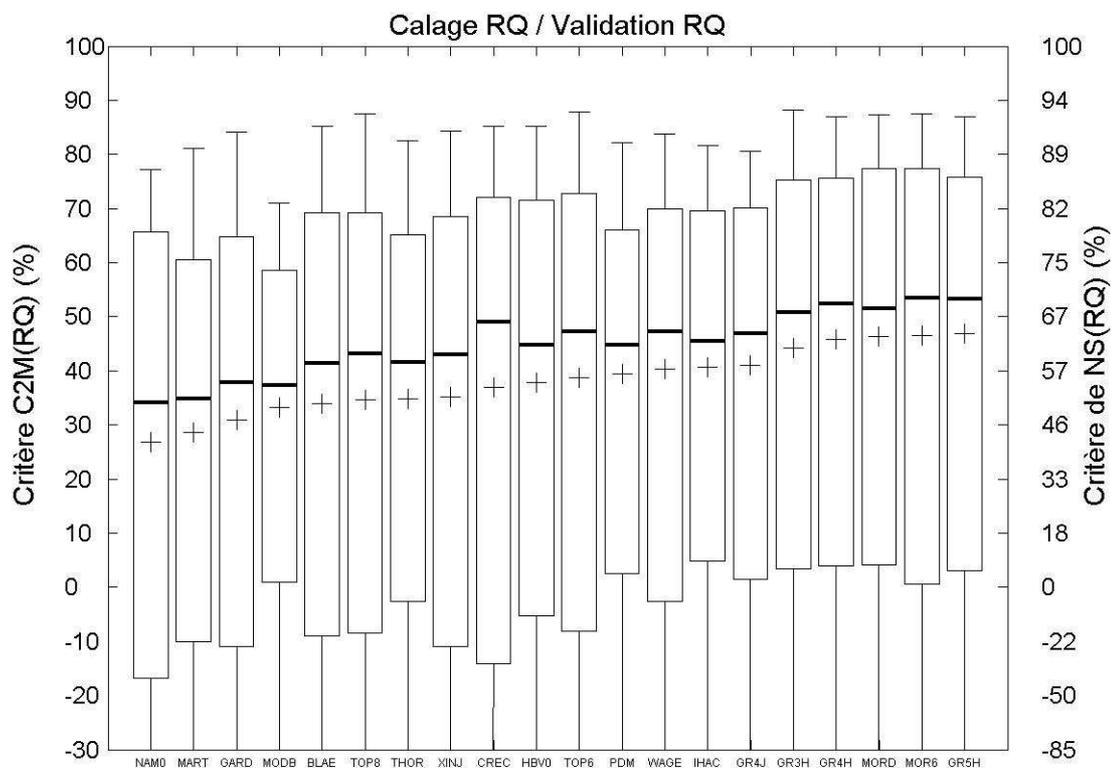


Figure 5-10 : Performances en validation des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(RQ)$. Classement croissant selon le moyenne (+).

Lorsqu'on utilise la fonction objectif $C_{2M}(RQ)$, il semble que les critères de performances des modèles soient plus dispersés. Les valeurs du quantile 0.1 diminuent assez fortement (de 10 à 20 points) pour une majorité des modèles. Cet effet est également mis en évidence par l'augmentation de l'écart entre la valeur de la médiane et celle de la moyenne des performances des modèles (Figure 5-4 et Figure 5-5).

Dans un second temps, l'évaluation des modèles se fait en validation. C'est d'ailleurs en validation que l'évaluation des performances est la plus importante, puisqu'elle correspond bien aux conditions d'utilisation classique des modèles. Nous avons évalué les performances des modèles avec les deux critères utilisés précédemment, en fonction des deux critères de calage utilisés, ce qui fait en tout quatre combinaisons :

- Q/Q : calage $C_{2M}(Q)$ / validation $C_{2M}(Q)$, Figure 5-7,
- Q/RQ : calage $C_{2M}(Q)$ / validation $C_{2M}(RQ)$, Figure 5-8
- RQ/Q : calage $C_{2M}(RQ)$ / validation $C_{2M}(Q)$, Figure 5-9
- RQ/RQ : calage $C_{2M}(RQ)$ / validation $C_{2M}(RQ)$, Figure 5-10.

Ces figures montrent tout d'abord, une forte baisse des performances moyennes des modèles en validation. Pour les cas Q/Q, Q/RQ et RQ/Q, les performances moyennes des modèles varient entre 20 et 40 points de C_{2M} . Pour le cas RQ/RQ, les performances moyennes sont légèrement meilleures et elles varient entre 27 et 46 points de C_{2M} . Par ailleurs, les valeurs des quantiles 0.9 ont baissé d'environ 10 à 15 points et celles des quantiles 0.1 de plus de 40 points : les modèles atteignent des niveaux de performances plus faibles et connaissent de forts échecs sur certains bassins versants. Par rapport aux résultats en calage, les écarts maximaux entre quantiles (Tableau 5-2) et les écarts entre médianes et moyennes ont également augmenté. Tous ces éléments indiquent, de manière générale, les fortes pertes de performances des modèles à cause de problèmes de robustesse. Rappelons cependant ici que nous plaçons les modèles dans des conditions difficiles, puisqu'ils ne disposent que d'une seule année de calage pour optimiser les valeurs de leurs paramètres.

Comme le montre la Figure 5-11, le rang des modèles peut être assez variable, en fonction du critère d'évaluation choisi. Néanmoins, pour les modèles de 'tête', les résultats et classements sont assez stables et on peut facilement identifier quelques structures comme étant performantes et robustes : MORD, MOR6, GR5H, GR4H, GR3H, WAGE et IHAC. Par ailleurs, certaines structures de modèles souffrent particulièrement de problèmes de robustesse et ont des écarts entre performance

moyenne et médiane assez grands. Ces structures, potentiellement intéressantes, sont TOP6 et CREC.

Nous n'avons utilisé que ces quatre critères pour juger des performances des modèles, *Perrin* [2000] ayant montré que le classement des modèles dépendait fortement du critère utilisé, mais les tendances entre modèles sont globalement conservées en fonction des différents critères. Un des résultats étonnant de ces tests est le faible niveau de performance des modèles HBV0 et TOP8, qui donnaient de bons résultats au pas de temps journalier dans la thèse de *Perrin* [2000].

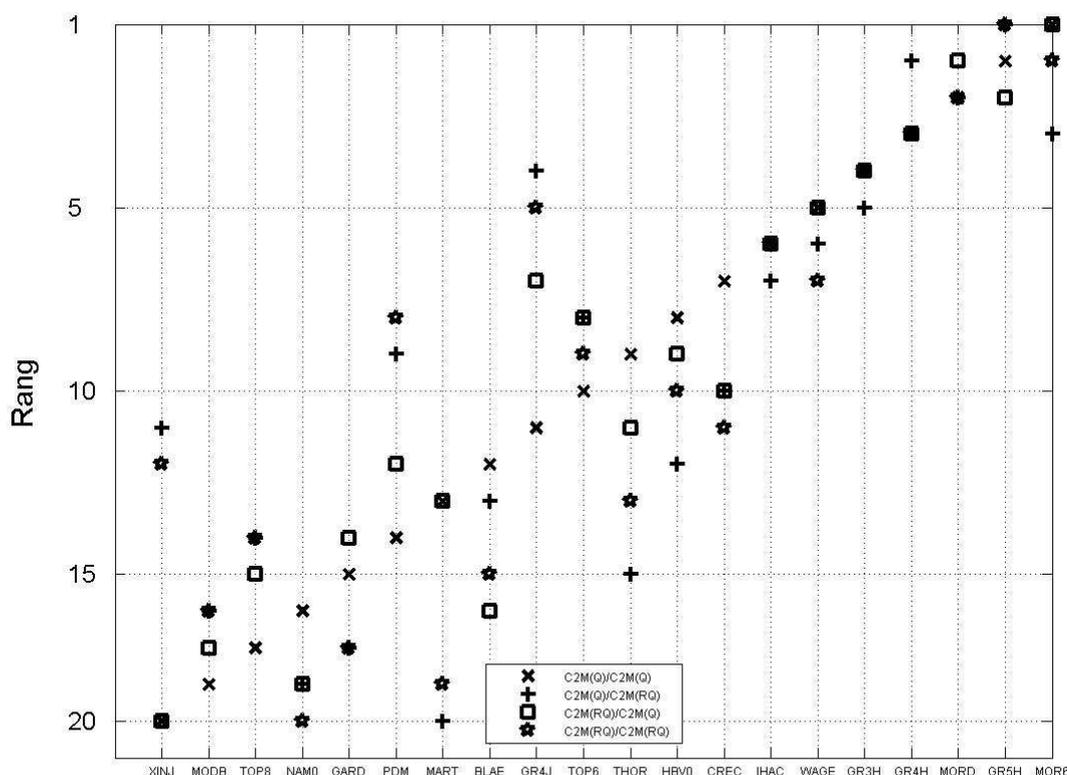


Figure 5-11 : Rang en validation des structures suivant les quatre cas testés de calage/contrôle.

La Figure 5-12 met en évidence l'influence de la fonction objectif sur les performances en validation, mesurées avec le critère $C_{2M}(Q)$ ou $C_{2M}(RQ)$. Cette figure montre que si le critère d'intérêt en validation est le $C_{2M}(Q)$, il vaut mieux calibrer le modèle avec ce critère pour 2/3 des modèles. Néanmoins, pour certains modèles (XINJ, PDM, GR4J, GR4H), l'autre critère de calage fournit de meilleurs résultats. Dans les deux cas, les différences de performances entre les modèles calibrés avec $C_{2M}(Q)$ ou $C_{2M}(RQ)$ sont limitées. Par contre, si le critère d'intérêt en validation est le $C_{2M}(RQ)$, il vaut mieux calibrer les modèles avec ce critère et les différences de performances entre les modèles calibrés avec $C_{2M}(Q)$ ou $C_{2M}(RQ)$ peuvent être un peu plus importantes (cinq à dix points quand même).

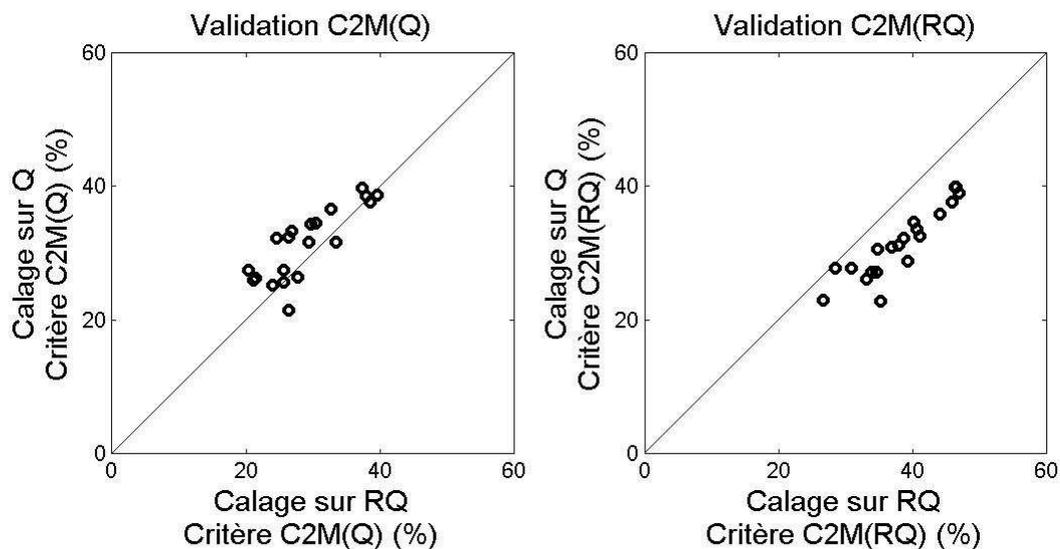


Figure 5-12 : Influence de la fonction objectif de calage sur les performances en validation pour les 20 modèles testés. Fonctions objectifs : $C_{2M}(Q)$ et $C_{2M}(RQ)$.

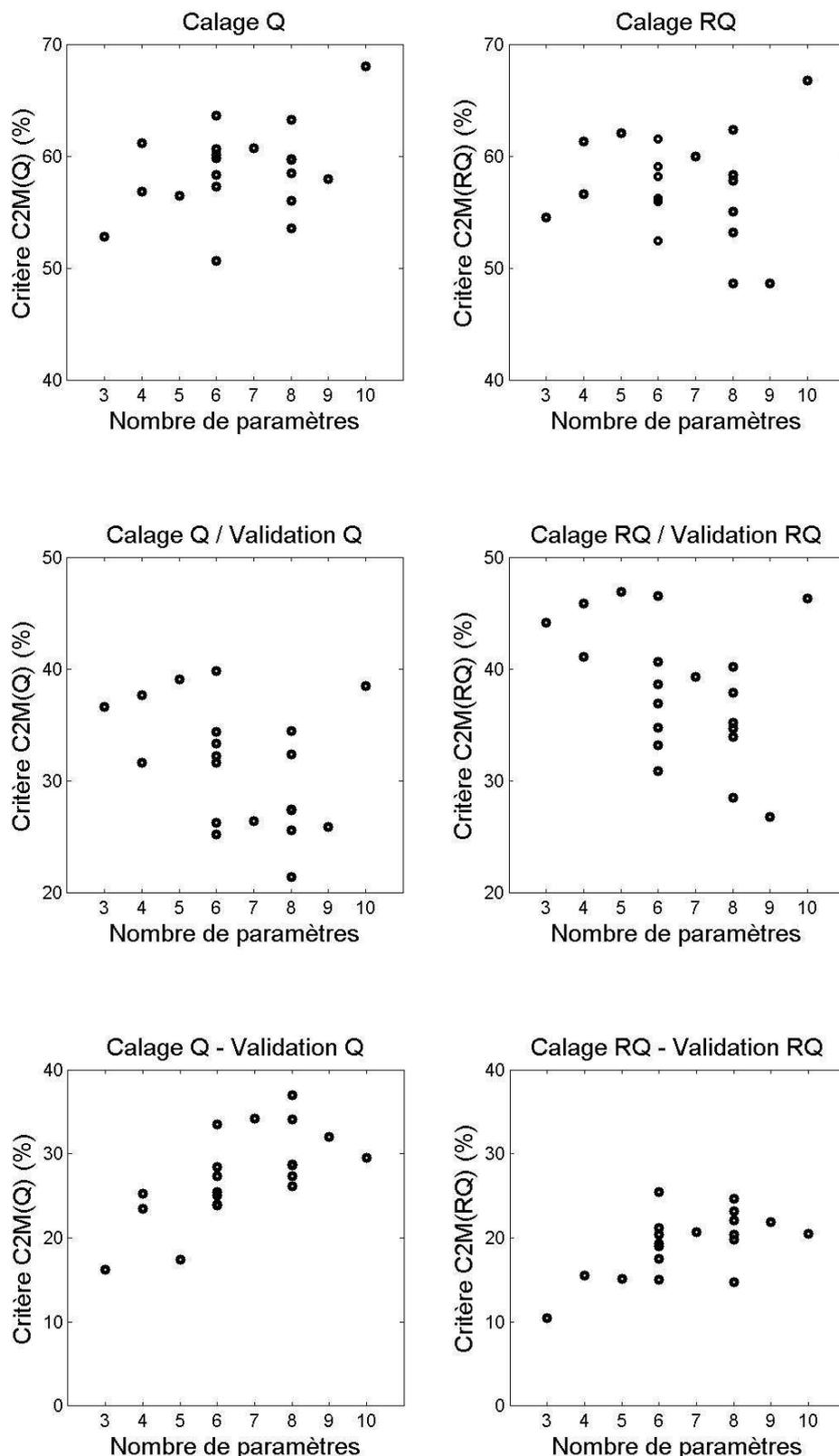
On peut définir un groupe de quelques modèles ayant probablement les meilleures chances de donner les meilleurs résultats. Le Tableau 5-3 présente le pourcentage de bassins versants où un modèle est meilleur qu'un autre. Ce tableau confirme bien la domination des modèles cités précédemment, qui sont généralement meilleurs que les autres modèles pour plus de 50% à presque 90 % des bassins versants. Par contre, quand on les compare entre eux, la domination d'un modèle est plus réduite et un modèle peut être meilleur qu'un autre dans seulement 50% à 70% des cas. Ces résultats confirment que le modèle unique, parfait, n'existe pas. Si on ne doit retenir qu'un ou deux modèles, il faut garder ceux qui ont les meilleures performances en moyenne et qui ont la plus grande probabilité d'être meilleurs qu'un autre modèle : les modèles MOR6 et GR5H. Le modèle MOR6 donne de meilleures performances sur les forts débits (critère $C_{2M}(Q)$), alors que le modèle GR5H donne de meilleures performances sur les débits moyens (critère $C_{2M}(RQ)$).

5.4 Robustesse des modèles testés

La robustesse est une des qualités des modèles hydrologiques la plus recherchée. En effet, la robustesse permet de juger de la capacité des modèles à fournir en phase de validation, des performances d'un même niveau que celui atteint au cours du calage. En fait, la robustesse mesure la capacité d'un modèle à être performant dans des situations hydro-climatiques différentes de celles qui ont servies à caler son vecteur de paramètres. La robustesse est calculée par la différence de performances entre le calage et la validation.

Tableau 5-3 : Pourcentage de bassins versants où le modèle A est supérieur au modèle B (modèles classés dans l'ordre croissant des performances moyennes selon le critère Q/Q).

		Modèle A																		
		MODB	TOP8	NAM0	GARD	PDM	MART	BLAE	GR4J	TOP6	THOR	HBV0	CREC	IHAC	WAGE	GR3H	GR4H	MORD	GR5H	MOR6
Modèle B	XINJ	42	57	55	56	49	52	61	69	74	62	71	79	70	75	78	81	82	82	84
	MODB		59	61	65	58	66	65	72	71	80	77	78	83	82	81	83	84	86	88
	TOP8			46	46	43	43	51	58	73	52	61	75	60	63	72	75	74	74	79
	NAM0				51	43	48	57	62	72	57	69	76	62	70	75	77	80	78	81
	GARD					42	47	53	61	67	61	61	78	63	66	74	77	75	78	81
	PDM						57	60	63	70	63	69	77	68	76	75	78	82	81	83
	MART							59	60	67	64	64	74	65	73	73	79	77	80	79
	BLAE								57	68	52	58	70	59	63	72	73	76	74	79
	GR4J									57	50	53	64	57	59	67	70	74	78	76
	TOP6										44	42	57	49	48	59	58	65	60	67
	THOR											58	68	56	64	66	71	72	71	74
	HBV0												70	56	55	63	68	70	73	75
	CREC													40	39	53	52	58	56	61
	IHAC														56	62	66	73	68	76
	WAGE															65	63	67	62	67
	GR3H																50	58	54	64
	GR4H																	57	58	60
MORD																		46	59	
GR5H																			55	



a)

b)

c)

Figure 5-13 : Performances moyennes en calage (a) et en validation (b) et diminution des performances entre calage et validation (c) des 20 structures de modèles testées, en fonction du nombre de paramètres optimisés.

Chapitre 5 : Intercomparaison des performances de modèles pluie-débit au pas de temps horaire

Tableau 5-4 : Différences moyennes, maximales et minimales des performances des 20 modèles entre le calage et la validation. Diminution exprimée en points de C_{2M} .

	Q/Q	RQ/RQ
Différence moyenne	28	19
Différence maximale	37	25
Différence minimale	16	10

Le Tableau 5-4 donne les diminutions moyennes, maximales et minimales des performances des 20 modèles, entre la phase de calage et la phase de validation. Ce tableau met en évidence les fortes pertes de performances en validation et laisse peut être à penser que la durée de calage est trop courte. En effet, les pertes de performances moyennes sont de 20 points de C_{2M} , pour le test RQ/RQ. Elles peuvent monter jusqu'à 30 points pour le test Q/Q. Cette configuration (Q/Q) rend les vecteurs de paramètres moins robustes, car ils sont plus sensibles aux extrêmes de la chronique utilisée pour le calage du modèle. Les pertes maximales peuvent même aller jusqu'à une quarantaine de points. Les écarts entre perte de performance minimale et maximale sont grands et vont de 15 à 21 points, ce qui indique que les modèles testés ont des niveaux de robustesse assez différents. Il semble également que les modèles soient légèrement plus robustes quand on utilise la fonction objectif $C_{2M}(RQ)$ pour les caler. Ce résultat n'est peut être pas étonnant, puisque cette fonction met l'accent sur les débits les plus fréquents, ce qui rend le vecteur de paramètres calés moins sensible aux événements rares d'une période.

Par rapport aux résultats de *Perrin* [2000], il semble que la robustesse des modèles testés soit plus faible. Cette perte de robustesse peut venir soit du pas de temps du modèle (plus grande complexité au pas de temps horaire), soit de la durée de calage (seulement 1 an dans ces tests, 4 à 5 ans pour *Perrin* [2000]), soit de la différence de base de données. L'hypothèse la plus plausible, celle de l'influence de la durée de calage, est testée au paragraphe 5.8.

La robustesse vient principalement de la structure des modèles (complexité et nombre de paramètres à optimiser). En effet, plus les modèles possèdent de degrés de liberté, plus ils peuvent s'adapter précisément aux données utilisées pour caler leurs paramètres, comme l'indique la Figure 5-13 a). Néanmoins, avec le même nombre de paramètres libres, on remarque la diversité possible des performances, ce qui indique que le nombre de paramètres n'est pas le seul facteur influençant le niveau de robustesse. La Figure 5-13 b) indique les performances des 20 modèles en validation. Par rapport au calage, la tendance précédente disparaît et on

remarque que des modèles ayant 4, 6 ou 10 paramètres sont capables de fournir le même niveau de performances. Même un modèle à 3 paramètres est capable d'être aussi bon ou meilleur que de nombreux modèles plus complexes. Enfin, la Figure 5-13 c) fait la synthèse des deux figures précédentes et indique la perte de performance entre calage et validation. Cette figure donne une image de la robustesse des modèles testés. La tendance est assez claire et montre qu'en général plus les modèles sont complexes, moins ils sont robustes. En effet, un grand nombre de paramètres permet aux modèles de s'adapter plus aisément aux particularités de la période de calage, mais diminue la généralité du vecteur de paramètres optimisé. C'est également pour cette raison que GR3H est performant en validation, car il ne peut s'adapter que faiblement aux caractéristiques particulières de chaque période de calage.

5.5 Fiabilité des modèles testés

La fiabilité est une des autres qualités recherchées. Elle représente la capacité d'un modèle à être régulièrement parmi les plus performants. Nous reprenons ici la démarche de *Perrin* [2000] et nous allons quantifier la fiabilité des modèles de manière relative. Pour chaque modèle, nous mesurons le nombre de bassins versants où il fait partie des meilleurs modèles. Il est de classe 1, s'il fait partie des 3 meilleurs modèles (premier sixième de l'ensemble des modèles) et de classe 2 s'il fait partie des 7 meilleurs modèles (premier tiers de l'ensemble des modèles).

La Figure 5-14 et la Figure 5-15 indiquent le pourcentage de bassins pour lesquels un modèle est meilleur que la moyenne des performances des 20 modèles, de classe 1 ou de classe 2. Ces figures correspondent respectivement aux configurations calage/validation : Q/Q et RQ/RQ.

Ces figures montrent clairement les différences de fiabilité des différents modèles, qui correspondent globalement aux différences de performances. Les deux configurations testées (Q/Q, RQ/RQ) donnent à peu près les mêmes résultats, qui sont résumés au Tableau 5-5. La proportion des bassins pour lesquels un modèle est supérieur à la moyenne des performances des 20 modèles varie globalement entre 20% et 90%. Le pourcentage de bassins pour lesquels un modèle peut être de classe 1 ou de classe 2 peut aller jusqu'à une quarantaine de pourcents.

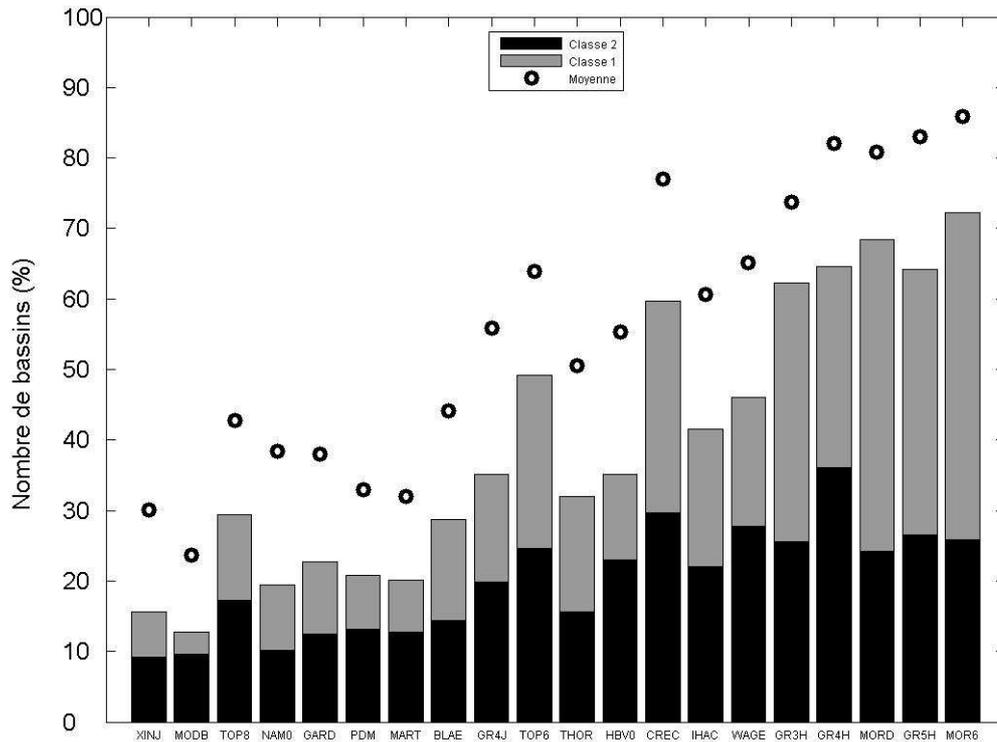


Figure 5-14 : Pourcentage de bassins pour lesquels un modèle est de classe 1, de classe 2, supérieur à la performance moyenne des 20 modèles (calage : $C_{2M}(Q)$ / validation : $C_{2M}(Q)$).

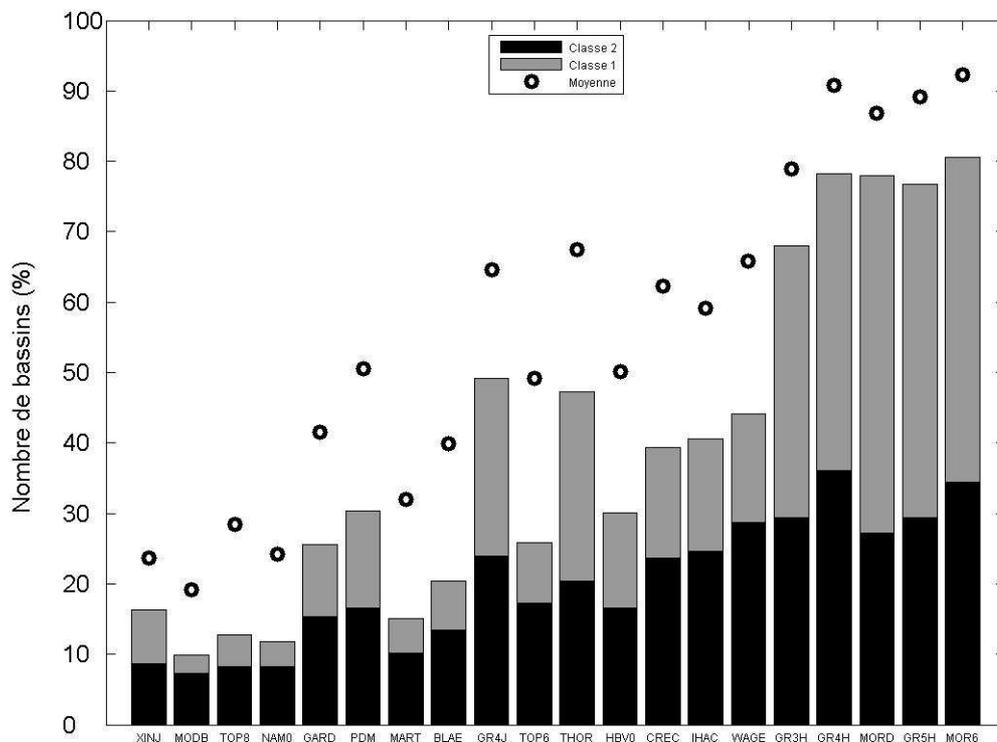


Figure 5-15 : Pourcentage de bassins pour lesquels un modèle est de classe 1, de classe 2, supérieur à la performance moyenne des 20 modèles (calage : $C_{2M}(RQ)$ / validation : $C_{2M}(RQ)$).

Chapitre 5 : Intercomparaison des performances de modèles pluie-débit au pas de temps horaire

Tableau 5-5 : Valeurs moyennes, minimales, maximales du pourcentage de bassins où un modèle est meilleur que la moyenne des 20 modèles, de classe 1 (C1), de classe 2 (C2).

	Q/Q				RQ/RQ			
	C1	C2	C1+C2	≥moy.	C1	C2	C1+C2	≥moy.
Moyenne	15	21	36	55	15	21	36	55
Maximum	42	37	70	86	44	41	78	91
Minimum	2	9	12	25	2	7	9	16

Ces résultats illustrent la grande variabilité des performances des modèles, puisque avec certaines structures (GR4H, GR5H, MORD, MOR6), il est probable d'obtenir les meilleures performances sur un bassin versant (classe 1 + classe 2) dans 65% à 80% des cas. De la même manière, certains modèles sont particulièrement peu fiables, puisqu'ils appartiennent aux meilleurs modèles dans moins de 20% des cas. Néanmoins, pour chacun des 20 modèles, il existe toujours au moins un bassin versant pour lequel ce modèle est le meilleur de tous.

5.6 Quelle complémentarité entre différentes structures de modèles ?

L'étude des performances des modèles a mis en évidence la grande variabilité de leurs performances moyennes et de celles obtenues sur un bassin versant. Chaque modèle a la possibilité d'être le meilleur modèle sur au moins un bassin versant. Ce résultat soulève la question de la complémentarité entre différentes structures de modèles, c'est-à-dire, identifier s'il est possible d'obtenir de meilleures performances sur un ensemble de bassins versants si on associe n modèles entre eux.

La Figure 5-16 présente les distributions des performances de chaque modèle, pour les quatre cas testés (Q/Q, Q/RQ, RQ/Q, RQ/RQ). Sur ces figures, les distributions du modèle 'idéal' *a priori* (pointillé noir) et *a posteriori* (trait noir) sont ajoutées. Le modèle idéal *a priori* obtient sur chaque bassin les performances en validation du meilleur modèle en calage. Le modèle idéal *a posteriori* obtient sur chaque bassin les performances en validation du meilleur modèle en validation.

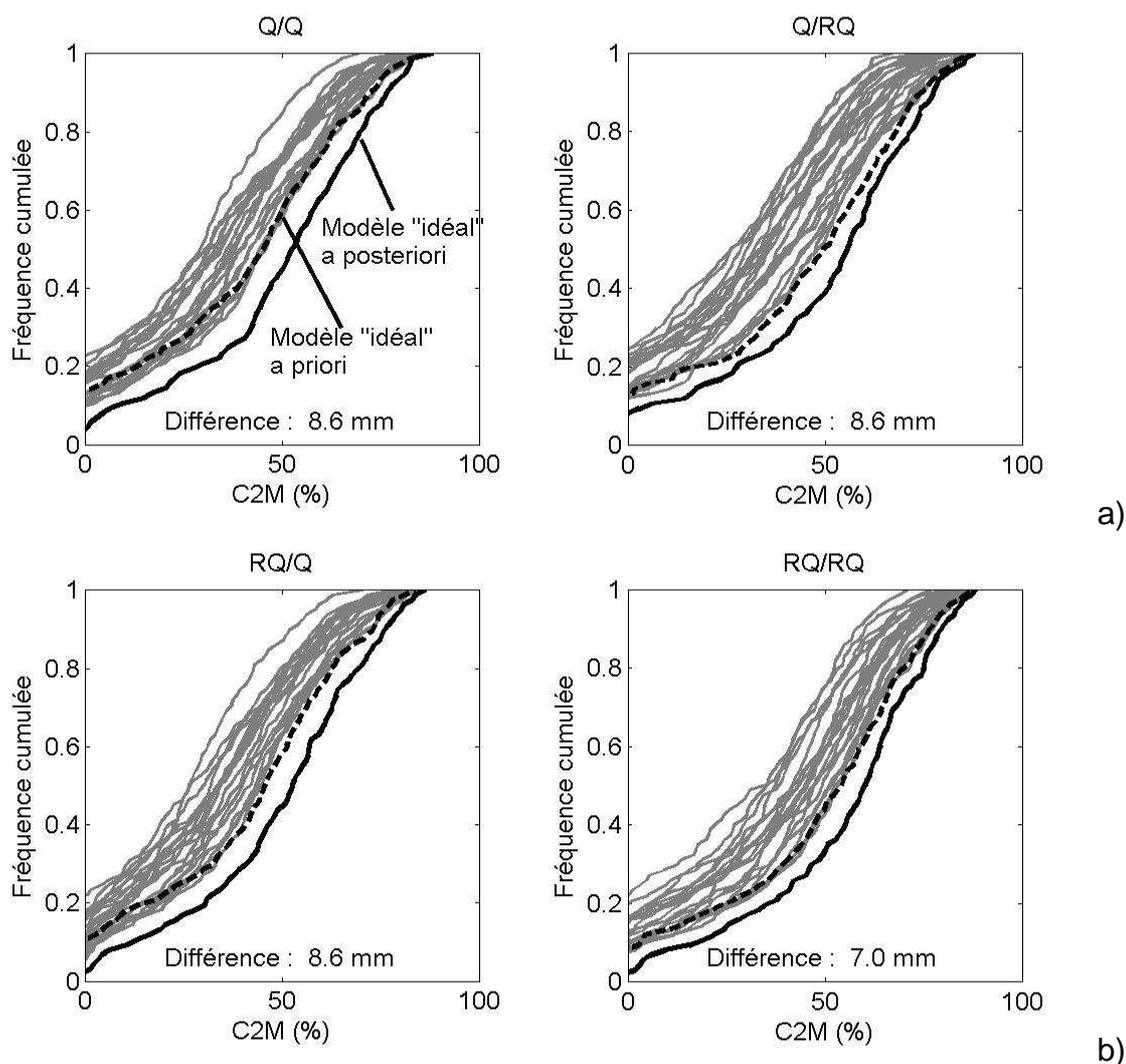


Figure 5-16 : Distribution des performances en validation des 20 modèles testés et distribution des modèles idéaux a priori et a posteriori. La différence correspond à la différence de performance moyenne entre le modèle idéal a posteriori et le meilleur modèle. a) modèles calés avec la fonction objectif $C_{2M}(Q)$, b) modèles calés avec la fonction objectif $C_{2M}(RQ)$.

La performance du modèle idéal *a posteriori* correspond à la meilleure performance qu'il est possible d'obtenir avec l'ensemble des modèles hydrologiques. Sur cette figure, on remarque que cette distribution est assez éloignée des distributions des 20 modèles testés. Cela veut dire que les structures testées sont loin de posséder une structure de modèle 'idéale', c'est-à-dire s'adaptant toujours au mieux sur tous les bassins. La différence de performance moyenne entre le modèle 'idéal' et la meilleure structure est de l'ordre d'une dizaine de points, pour les quatre cas testés. Cela indique que la marge de progrès des modèles est encore grande. Cette marge de progrès est du même ordre de grandeur que celle obtenue par Perrin [2000] pour des modèles au pas de temps journalier. Les deux meilleurs modèles testés peuvent être considérés comme équivalents au modèle 'idéal' *a priori*.

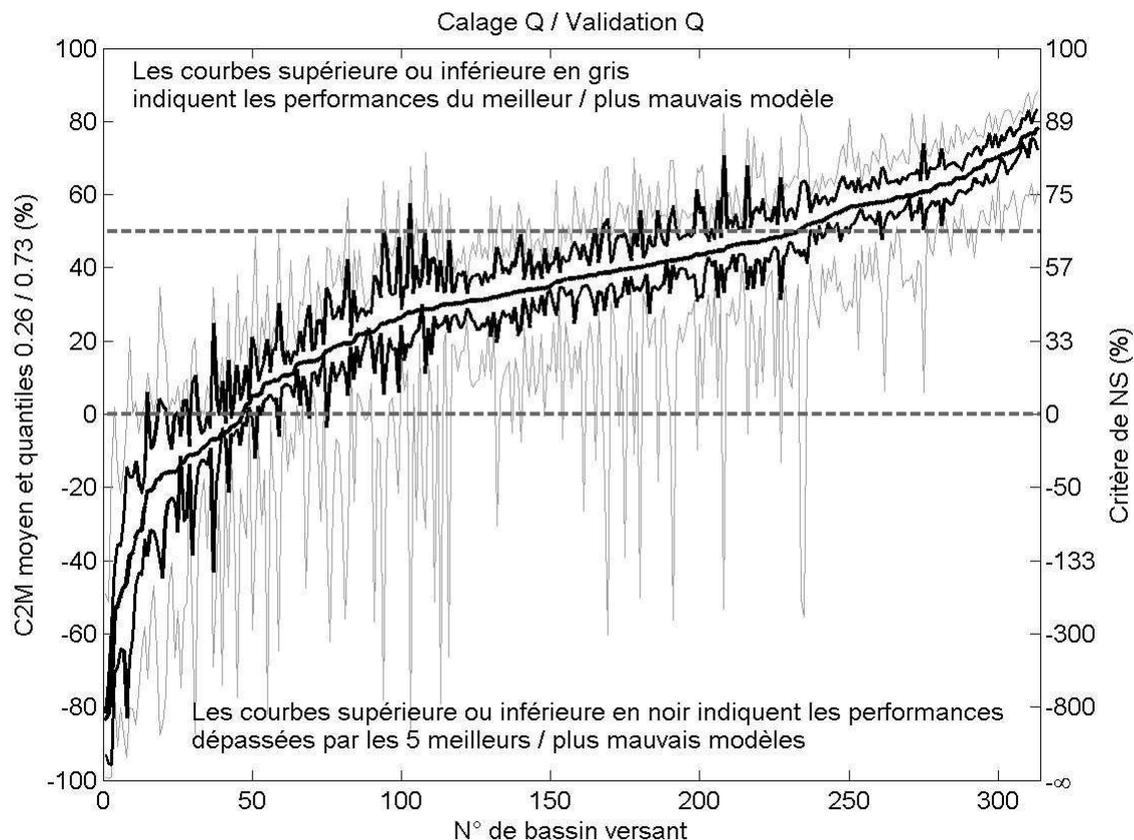


Figure 5-17 : Performances obtenues par les 20 modèles sur les 313 bassins versants : moyenne, valeurs des quantiles 0.26 et 0.73 (5 meilleurs/plus mauvais modèles), minimum et maximum.

En effet, les distributions des performances de MOD6 ou GR5H sont généralement un peu plus à droite que celles du modèle idéal *a priori*. Ainsi, on peut conclure que ces deux modèles peuvent être recommandés comme les meilleurs des modèles, à l'issue de cette comparaison.

La Figure 5-17 détaille la moyenne des performances obtenues sur chaque bassin versant par les 20 modèles. Sur cette figure sont également indiquées les performances dépassées par les cinq meilleurs / plus mauvais modèles (quantiles 0.26 et 0.76) et les performances minimales et maximales obtenues sur chaque bassin versant. Cette figure souligne la grande variabilité des résultats que l'on peut obtenir sur chaque bassin versant et les nombreux cas où un modèle n'est absolument pas robuste (pics vers le bas). Sur cette figure, le modèle 'idéal' *a posteriori* est celui qui correspond à la courbe supérieure grise.

La Figure 5-18 indique le nombre de bassins versants où un modèle est le meilleur en validation. Cette figure illustre la grande variabilité des résultats possibles des modèles.

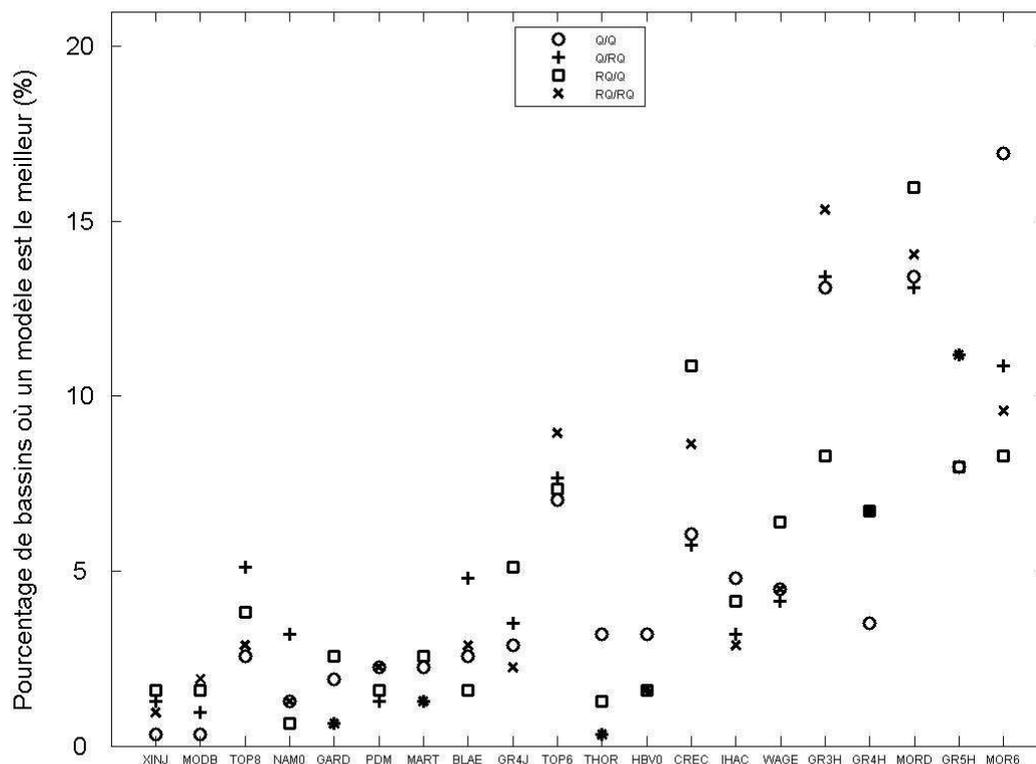


Figure 5-18 : Pourcentage de bassins où un modèle est le meilleur.

Les modèles les moins performants, sont généralement les meilleurs sur assez peu de bassins versants (moins de 5%). Par contre, les modèles les plus performants sont capables d’être les meilleurs sur 10% à 18% des bassins versants.

Enfin, le Tableau 5-3, qui indique le pourcentage de bassins versants où un modèle est meilleur qu’un autre, montre la complémentarité possible entre modèles. En effet, comme on l’a vu précédemment, un modèle peut être meilleur qu’un autre dans 50% à 90% des cas. Ainsi, on peut s’attendre à ce que deux modèles complémentaires permettent de faire progresser la moyenne des performances sur un groupe de bassins versants.

Ainsi, les propriétés des modèles mises en évidence par la Figure 5-16 à la Figure 5-18 et le Tableau 5-3 montrent qu’un gain de performances des modèles hydrologiques est envisageable et qu’il est éventuellement possible de tirer partie d’une complémentarité entre structures de modèles. Pour cela, nous présentons au Tableau 5-6 et aux Tableau 9-3 à, Tableau 9-5 en Annexe 6) les performances moyennes obtenues par un couple de deux modèles, pour lequel nous avons retenu *a posteriori* les meilleures performances en validation. Ces résultats (synthétisés au Tableau 5-7) montrent l’intérêt d’utiliser plusieurs modèles :

- l’approche complémentaire permet de faire progresser fortement la performance moyenne minimale d’un couple de deux modèles (d’au moins 10 points). Cela

veut dire que cette approche, même avec des modèles peu robustes et peu fiables, permet de faire progresser très significativement leurs performances ;

- l'association de deux modèles permet d'obtenir de meilleures performances, par rapport aux performances du meilleur modèle, dans 30% à 35% des combinaisons de deux modèles,
- les meilleures associations de modèles sont toutes basées sur la combinaison des modèles de la famille GR et MORDOR avec un autre modèle.

En revanche, les résultats de cette approche peuvent sembler légèrement décevants car les gains restent limités, les gains moyens maximaux sont seulement de 3 à 4 points de C_{2M} . Les meilleures associations de modèles sont celles des modèles MORD ou MOR6 avec les modèles GR3H, TOP6 ou THOR. Ce résultat n'est finalement pas étonnant car ces modèles sont généralement parmi les plus performants : ils sont classés parmi les meilleurs modèles pour plus de 60% à 80% des bassins versants. Ainsi, même si un de ces modèles est meilleur qu'un autre sur un bassin versant, la différence de performance entre l'un des deux modèles (donc le gain) est généralement limitée.

Ces résultats ne concernent que l'association *a posteriori* de deux modèles. Néanmoins, dans un cadre opérationnel, le choix du meilleur des n modèles doit se faire *a priori* (au calage). Nous avons testé une stratégie très simple de choix *a priori*, où nous retenons le meilleur modèle au calage, parmi n modèles testés. Dans notre cas, nous n'avons testé que la combinaison de deux modèles, parmi ceux qui donnaient de bons résultats au cours de la combinaison *a posteriori* : TOP6, THOR, CREC, IHAC, WAGE, GR3H, GR4H, MOR6. Les résultats de cette approche de combinaison *a priori* se trouvent en Annexe (Tableau 9-6, Tableau 9-7, Tableau 9-8, Tableau 9-9) et sont synthétisés au Tableau 5-7. Cette approche donne également des gains limités pour les tests sur Q/Q et RQ/Q, les gains étant inférieurs à 1 points. En revanche pour les tests Q/RQ et RQ/RQ, les gains commencent à être intéressants. Ils sont de 1.4 à 2 points pour des combinaisons *a priori* de GR4H et MOR6.

Tableau 5-6 : Performances moyennes obtenues lorsqu'on associe a posteriori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures. Performance moyenne du meilleur modèle seul : MOR6, $C_{2M}=39.9$. En gras les combinaison supérieure à cette valeur. Fonction objectif en calage et en validation : $C_{2M}(Q)$.

		Modèle A																		
		MODB	TOP8	NAM0	GARD	PDM	MART	BLAE	GR4J	TOP6	THOR	HBV0	CREC	IHAC	WAGE	GR3H	GR4H	MORD	GR5H	MOR6
Modèle B	XINJ	32.5	31.3	31.9	31.8	32.8	32.3	32.8	35.5	34.8	34.7	34.8	36.2	37.3	36.6	39.5	39.4	40.3	40.1	41.3
	MODB		33.8	32.8	32.4	32.8	32.4	33.8	34.9	37.0	34.1	35.7	37.8	35.8	36.8	39.4	39.9	41.0	40.3	41.6
	TOP8			32.2	33.3	34.8	33.9	33.5	36.5	33.1	36.6	35.4	36.7	38.4	37.2	39.6	39.7	40.9	40.8	41.7
	NAM0				33.4	34.3	33.3	33.3	36.9	35.1	35.6	35.4	37.3	37.8	36.9	39.4	40.1	41.4	40.8	42.1
	GARD					33.7	32.4	34.0	35.4	36.1	35.7	36.8	36.3	37.5	37.4	40.2	39.9	40.8	40.5	41.4
	PDM						33.5	34.9	36.4	37.7	35.7	36.5	38.3	37.0	37.0	40.4	39.8	40.3	40.3	41.3
	MART							34.1	36.1	36.7	34.9	35.9	36.9	37.6	37.0	40.2	39.9	41.0	40.4	41.7
	BLAE								36.7	35.9	37.1	36.4	37.5	38.4	37.4	39.7	39.7	40.7	40.7	41.6
	GR4J									38.6	38.2	38.1	39.5	38.0	38.8	41.0	39.9	41.0	40.0	41.4
	TOP6										38.9	37.5	38.0	40.3	39.0	40.9	41.0	41.9	41.8	42.8
	THOR											37.6	38.9	38.1	38.1	41.3	41.4	42.6	41.8	42.9
	HBV0												38.8	38.9	38.6	40.3	40.5	41.6	41.2	42.1
	CREC													40.6	39.8	41.7	41.2	42.1	41.9	42.6
	IHAC														39.7	41.8	41.4	42.1	41.7	42.3
	WAGE															42.3	41.7	42.2	42.1	42.6
	GR3H																		43.1	42.0
GR4H																		42.1	40.1	42.6
MORD																			42.5	41.7
GR5H																				42.9

Tableau 5-7 : Synthèse des tests de complémentarité entre deux structures de modèles, en fonction des stratégies de combinaison de modèles a posteriori et a priori (gain entre parenthèses).

	Q / Q	Q / RQ	RQ / Q	RQ / RQ
Meilleur modèle	MOR6	GR5H	MOR6	GR5H
C _{2M} moyen	39.9	39.5	40.0	47.0
Meilleur couple a posteriori	GR3H / MOR6	GR4H / MORD	TOP6 / MORD	GR3H / MORD GR5H / MORD
C _{2M} moyen	43.3 (+3.4)	42.5 (+3.0)	43.5 (+3.5)	50.0 (+3.0)
Meilleur couple a priori	IHAC / MOR6	GR4H / GR5H	GR5H / MOR6	GR4H / MOR6 GR5H / MOR6 GR4H / GR5H
C _{2M} moyen	40.5 (+0.6)	41.0 (+1.5)	41.1 (+1.1)	48.0 (+1.0)

Par rapport aux résultats de Perrin [2000], cette approche de complémentarité entre structures semble un peu plus limitée. En effet, nos tests ont permis de discriminer plusieurs structures ayant de bons niveaux de performances. En revanche, les études complémentaires sur la robustesse et la fiabilité des modèles a mis en évidence la domination des performances des structures de GR4H et MOR6. Les autres structures (WAGE, IHAC, CREC, TOP6) semblent légèrement moins performantes. Ainsi, par rapport aux performances des modèles au pas de temps journalier, il existe moins de structures ayant un niveau élevé de performances au pas de temps horaire, ce qui limite la diversité des philosophies de modélisation et donc l'exploitation de la complémentarité de certaines structures de modèles.

5.7 Existe-t-il certains facteurs discriminant les performances de certains modèles ?

La recherche de relations entre les performances d'un modèle et les caractéristiques des bassins versants est une thématique de recherche très active en hydrologie, en relation avec les tentatives de régionalisation des paramètres des modèles hydrologiques et d'application (efficace) des modèles sur des bassins versants non jaugés. De nombreux hydrologues nourrissent l'espoir de trouver les bassins versants pour lesquels un modèle est performant. D'ailleurs au cours du développement progressif des modèles de type GR et des comparaisons de

structures de modèles à différents pas de temps ou sur différents échantillons de bassins versants, la question : « **sur quels bassins ce modèle fonctionne t'il bien ?** » a été très souvent posée.

A l'heure actuelle, les résultats restent timides et les caractéristiques permettant de discriminer significativement les performances des modèles hydrologiques sont encore à trouver. Néanmoins, nous allons explorer ici les relations entre les performances des modèles et certaines caractéristiques hydro-climatiques et géographiques des bassins versants de notre échantillon. L'autre intérêt d'une recherche de facteurs discriminant les performances des modèles hydrologique est de réaliser une typologie 'bassin/modèle', c'est-à-dire trouver les types de bassins sur lesquels certains modèles sont significativement plus performants.

Pour chaque descripteur testé, nous étudions (1) les performances moyennes des 20 modèles sur des sous classes de bassins versants (caractérisées par une valeur faible, moyenne ou forte pour chaque caractéristique) et (2) la relation entre la performance sur un bassin versant des modèles GR4H et MOR6 et la valeur de cette caractéristique.

Nous utilisons les caractéristiques des bassins versants décrites au Chapitre Chapitre 2. Les caractéristiques hydro-climatiques que nous avons retenues sont : le débit moyen, la pluie moyenne, l'ETP moyenne, le rendement moyen, le coefficient de variation des pluies non nulles et des débits non nuls. Les caractéristiques géographiques que nous avons retenues sont la surface du bassin versant et pour les bassins versants français : le type de couvert végétal dominant (4 classes) et des indicateurs de la géologie de surface.

Avant de tester l'influence de certaines caractéristiques des bassins versants sur les performances des modèles, nous avons voulu tester l'influence éventuelle des bassins versants dont les données pourraient être de mauvaise qualité. Pour cela, nous avons divisé l'échantillon de 313 bassins versants en fonction des performances moyennes des modèles sur chaque bassin versant. Nous avons pris, soit trois classes (Figure 5-19 ; Q : quantile ; $\overline{C_{2M}} < Q(0.33)$, $\overline{C_{2M}} < Q(0.66)$, $\overline{C_{2M}} > Q(0.66)$), soit en deux classes (Figure 5-20 ; $\overline{C_{2M}} < Q(0.1)$, $\overline{C_{2M}} > Q(0.1)$). Ce test doit nous permettre d'évaluer si les bassins versants pour lesquels les modèles ont de faibles performances ne sont pas capables d'influencer fortement le classement des modèles.

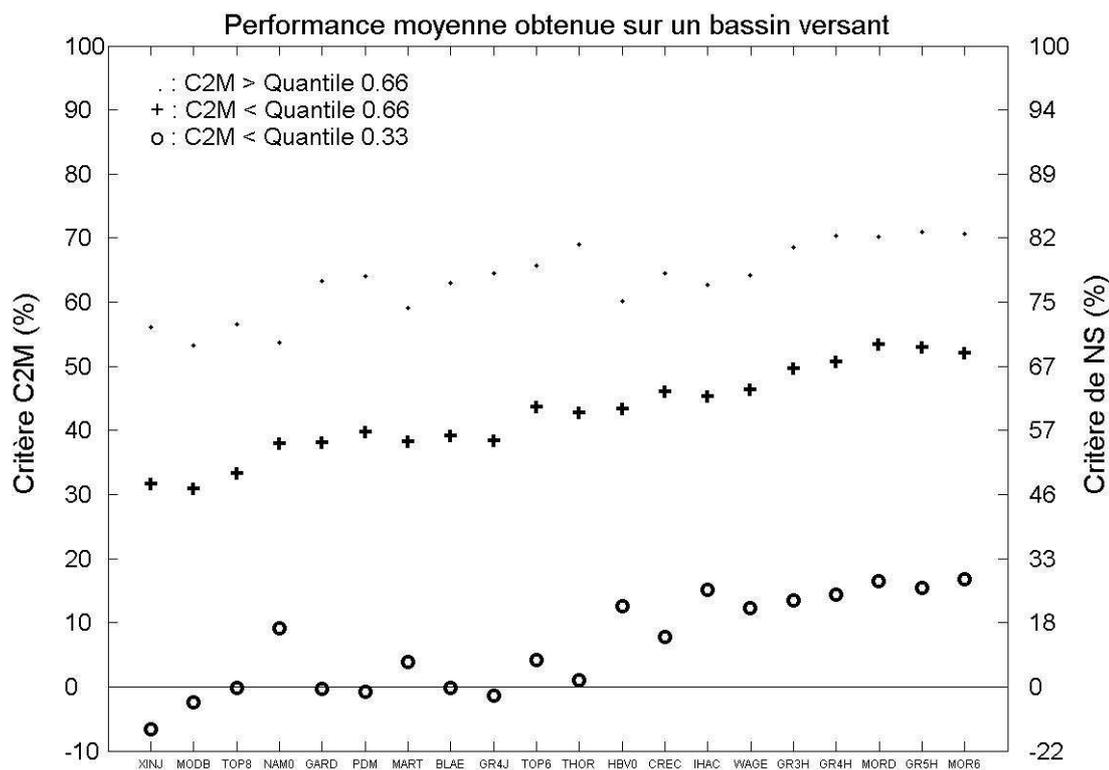


Figure 5-19 : Discrimination du classement des modèles en fonction des performances moyennes des modèles sur les bassins versants (3 classes).

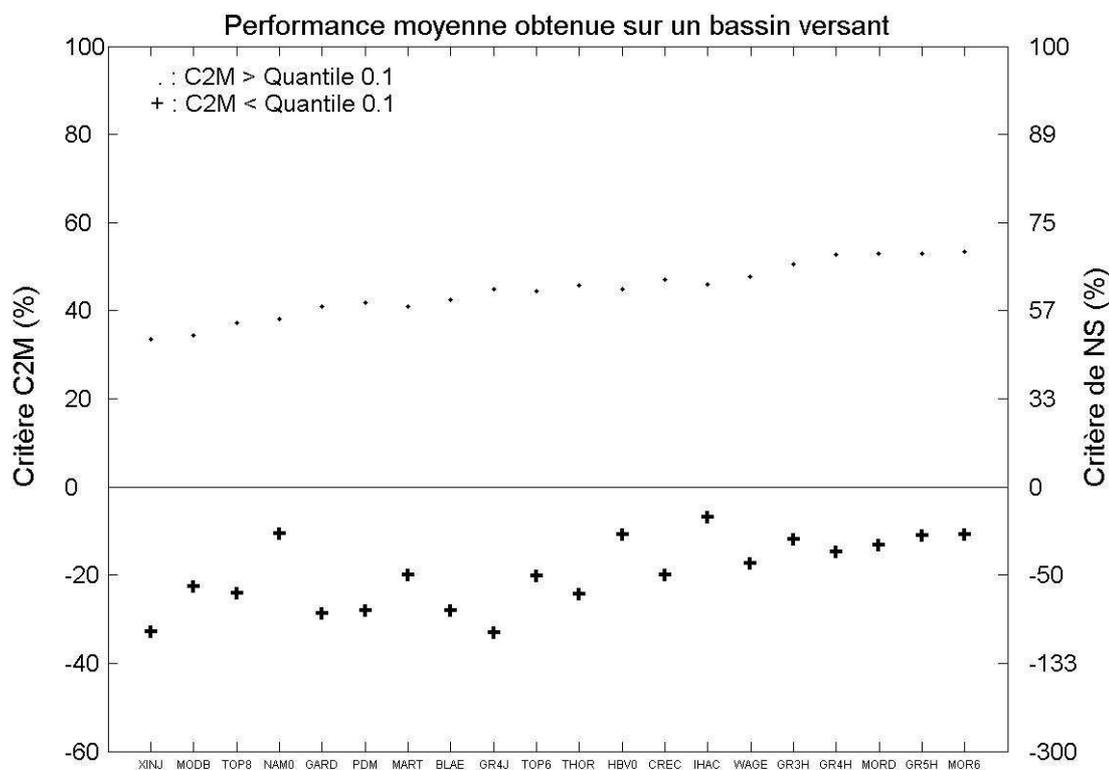


Figure 5-20 : Discrimination du classement des modèles en fonction des performances moyennes des modèles sur les bassins versants (2 classes).

Ces deux figures montrent que pour la classe moyenne des performances des modèles et les 90% des bassins versants ayant les meilleures performances moyennes, le classement des modèles n'est pas significativement modifié. Cela nous permet de conclure que, même si notre base de données contient un certain nombre de bassins versants ayant des données de mauvaise qualité, le classement des modèles n'en est pas affecté. Par contre, les classements avec la classe faible et la classe forte, montrent la capacité à avoir de très bonnes performances et des problèmes de robustesse de certains modèles (THOR, TOP8, PDM).

Les relations entre les performances moyennes des modèles et quelques caractéristiques des bassins versants sont illustrées par les Figure 5-21, à, Figure 5-24, et en Annexe 7. De ces figures, il ne ressort pas de facteurs très discriminants. Généralement, on peut observer une tendance entre les performances moyennes des modèles (Tableau 5-8), qui sont croissantes ou décroissantes avec le type de classe du descripteur (faible, moyenne ou forte). Ces figures montrent que les performances sont croissantes avec le débit, la pluie, le rendement et la surface des bassins versants et décroissantes avec l'ETP.

Tableau 5-8 : Relations entre évolution des performances moyennes et évolution des caractéristiques des bassins versants. Synthèse des tendances : +, performances croissantes avec la valeur de la caractéristique ; -, performances décroissantes avec la valeur de la caractéristique ; 0, pas de tendance.

Caractéristique	Tendance
Débit	+
Pluie	+
ETP	-
Rendement	+
Coefficient de variation de la pluie	0
Coefficient de variation des débits	0
Surface	+
Végétation	0
Géologie	0

Ce résultat est nullement étonnant, et rappelle les faibles niveaux de performances des modèles sur les bassins à faible rendement ou intermittents. Cette tendance vient de deux facteurs, d'une part la faible capacité de certains modèles à simuler les bassins versants intermittents et d'autre part, les plus faibles valeurs du critère de NS (et donc du C_{2M}) sur ce type de bassins [Perrin, 2000]. Comme nous l'avons vu

précédemment, les plus faibles performances des modèles sur les bassins versants intermittents viennent de la formulation du critère qui est plus exigeant pour ce type de bassins versants. Néanmoins, certains modèles sont peu performants sur les bassins intermittents. Ces difficultés proviennent soit de problèmes d'initialisation des réservoirs de certains modèles et/ou à l'absence de fonctions qui permettent une perte d'eau par ailleurs que la voie aérienne (sous forme d'ETR).

La Figure 5-21 met en évidence un résultat intéressant. Sur cette figure les performances moyennes des modèles sont classées en fonction du pays d'origine. Les bassins versants australiens sont tous de la même région, soumis à un climat méditerranéen et ont un régime d'écoulement intermittent, avec un assez faible rendement. Ils sont néanmoins différents des bassins versants américains arides, qui reçoivent encore moins de précipitations. Généralement, les bassins intermittents posent des problèmes en modélisation, notamment pour les modèles qui ne bouclent leur bilan en eau que par l'atmosphère. Ainsi, sur ce sous échantillon, des modèle aux performances moyennes assez faibles donnent d'excellents résultats : c'est le cas de XINJ et TOP8. Les bonnes performances du modèle XINJ sur les bassins versants arides avaient déjà été constatée [Gan et al., 1997]. Par ailleurs d'autres modèles se distinguent également : THOR et CREC. Enfin, certains modèles sont particulièrement mauvais sur ces bassins versants : NAM0, MODB et PDM.

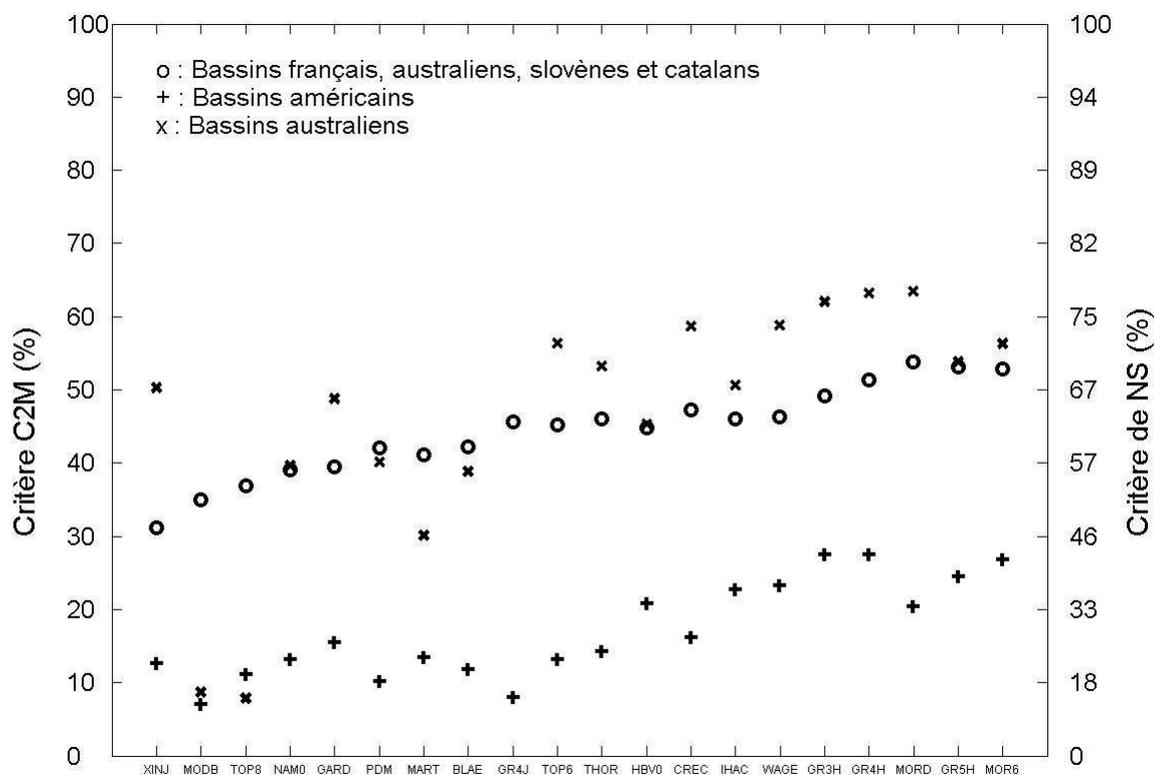


Figure 5-21 : Performances moyennes des 20 modèles en fonction du pays d'origine des bassins versants.

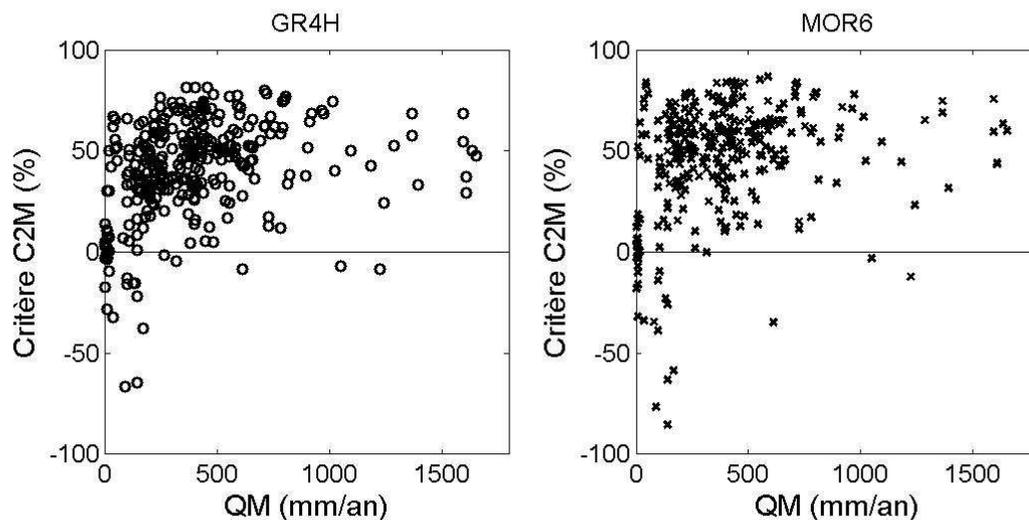


Figure 5-22 : Performances des modèles GR4H et MOR6 en fonction du débit moyen annuel des bassins versants.

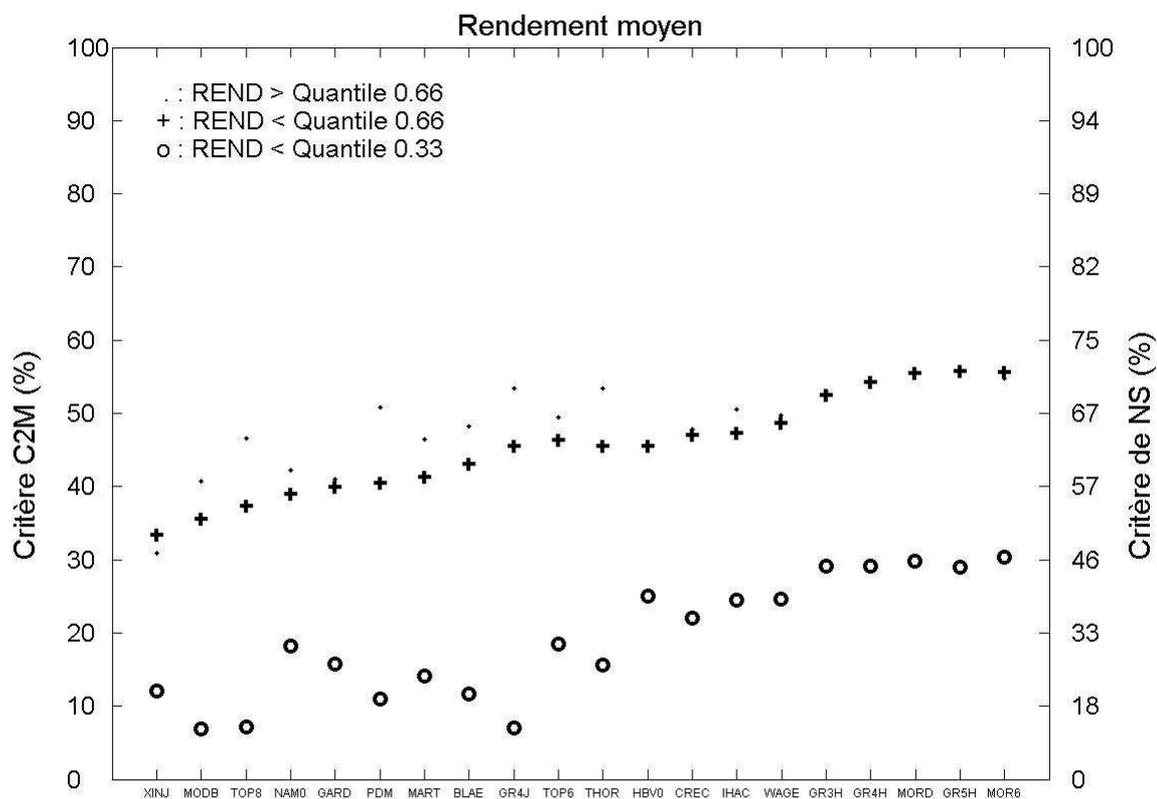


Figure 5-23 : Performances moyennes des 20 modèles en fonction de la classe du rendement moyen annuel des bassins versants.

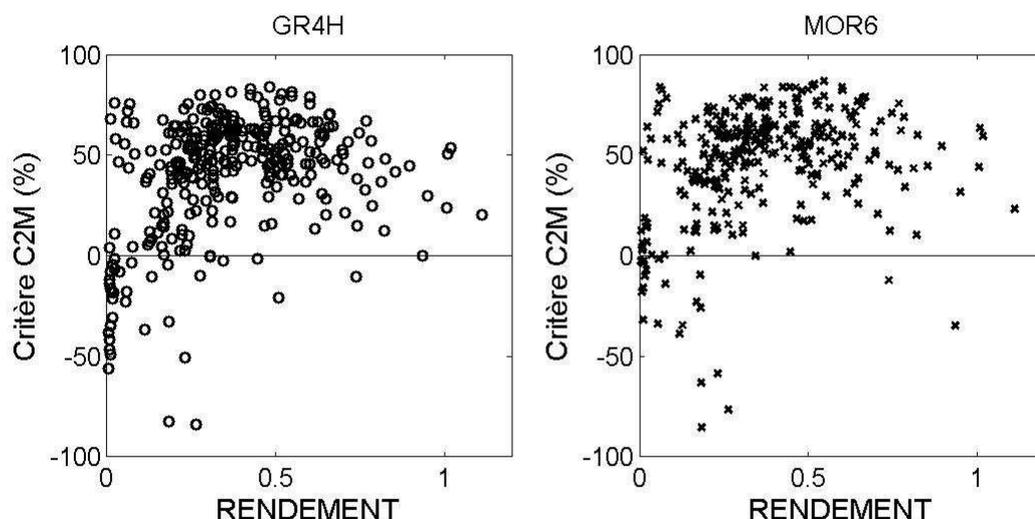


Figure 5-24 : Performances des modèles GR4H et MOR6 en fonction du rendement des bassins versants.

5.8 Impact de la durée de calage sur les performances du modèle

5.8.1 Problématique

Cette base de données au pas de temps horaire, du fait de la difficulté et du coût de se procurer les données de pluie, possède des chroniques de données d'assez courtes durées (en moyenne, de 6 à 8 ans). Ainsi, par rapport aux tests classiques de modèles au pas de temps journalier, la durée de calage peut sembler très courte. Il devient alors légitime de se demander quel est l'impact de cette plus courte durée de calage sur les performances, le classement des modèles et les vecteurs de paramètres optimisés. En effet, l'intérêt d'une plus grande durée d'une chronique de données est de potentiellement fournir au modèle une plus grande variabilité de situations hydro-climatiques, et d'activer ainsi différentes composantes de son fonctionnement. Un modèle ayant un grand nombre de paramètres (c'est-à-dire, plus de fonctions) à ainsi peut-être besoin de plus de données, ou de plus d'information hydrologique contenue dans les données, pour caler (contraindre) ses différents paramètres.

Dans la littérature, cette thématique a été abordée par quelques auteurs [Allred et Haan, 1991; Sorooshian et al., 1993; Yapo et al., 1996; Gan et al., 1997; Perrin, 2000], dont les conclusions ne sont pas toujours concordantes. Globalement, ces auteurs s'accordent pour dire que ce n'est pas la durée des données qui est le plus important mais l'information contenue dans les données, la présence d'événements rares notamment. Certains auteurs insistent également sur l'intérêt de calibrer les modèles sur des années plutôt humides [Sorooshian et al., 1993; Yapo et al., 1996; Gan et al., 1997], qui permettent de mieux activer le fonctionnement des différentes

composantes du modèle. En revanche, certains auteurs ne s'accordent pas sur la durée optimale des chroniques de données qu'il faut pour correctement optimiser un modèle. D'après les travaux de *Yapo et al.* [1996], en utilisant une version du modèle SAC-SMA (13 paramètres) testée sur un bassin versant, une durée d'environ 8 années permet d'obtenir des performances et des vecteurs de paramètres relativement insensibles aux caractéristiques des années sélectionnées. Par contre, pour *Gan et al.* [1997], en utilisant 5 modèles (de 9 à 21 paramètres) testés sur trois bassins versants, leurs résultats n'indiquent pas qu'il existe une relation entre la durée de la période de calage et les performances des modèles testés. *Perrin* [2000] a testé l'influence de deux découpages différents des périodes de calage / contrôle sur les performances de sept modèles (3 à 9 paramètres), testés au pas de temps journalier sur 231 bassins versants. Les deux découpages ont pour durées moyennes de calage 5.1 et 8.6 ans. Il constate que les performances des sept modèles progressent uniformément lorsque la durée de calage augmente, ce qui ne remet pas en cause le classement des modèles.

Enfin, dans le domaine des réseaux de neurones artificiels (RNA), certaines études [*Dawson et Wilby*, 1998; *Anctil et al.*, 2004b] soulignent également l'intérêt des périodes humides pour correctement optimiser des RNA. Par ailleurs, *Anctil et al.* [2004b] ont comparé l'influence de la durée de chroniques de calage sur les performances du modèle GR4J et d'un modèle basé sur des RNA, testés sur un bassin versant. Ces travaux démontrent que les performances du modèle GR4J atteignent un optimum à partir d'une durée de 4 années de calage, alors que les performances du RNA progressent jusqu'à une durée de 9 années.

En fait, la durée optimale de chronique de calage dépend certainement de la structure du modèle et de sa complexité et des conditions hydro-climatiques des bassins versants. Certains auteurs s'accordent pour dire qu'une durée de 8 à 10 ans est nécessaire. Il est probable que cette durée est suffisante pour confronter le modèle à une bonne variabilité hydro-climatique et à des événements rares qui vont activer les différentes composantes du modèle. Cette période de 8 à 10 ans peut être mise en parallèle à la variabilité climatique décrite par *Mandelbrot et Wallis* [1968] comme *l'effet Joseph*, qui s'approche également de celle de l'activité solaire dont les systèmes hydrologiques semblent être dépendants.

5.8.2 Tests et résultats

Pour tester l'influence de la durée du calage sur les performances du modèle, nous avons repris les 20 modèles testés précédemment. La procédure suivie au cours de ce test est illustrée par la Figure 5-25. Nous avons calé les modèles sur des chroniques de 1, 2, 3 et 4 années, en utilisant comme fonction objectif $C_{2M}(Q)$ au

calage. Nous les avons évalué sur les mêmes périodes de validation (quelle que soit la période de calage) avec les fonctions $C_{2M}(Q)$ et $C_{2M}(RQ)$. Malheureusement, notre base de données ne nous a pas permis de tester des durées plus longues, le nombre de bassins versants ayant suffisamment de données devenant trop réduit. Nous avons effectué ce test sur un sous-échantillon de 70 bassins versants (47 français et 23 américains).

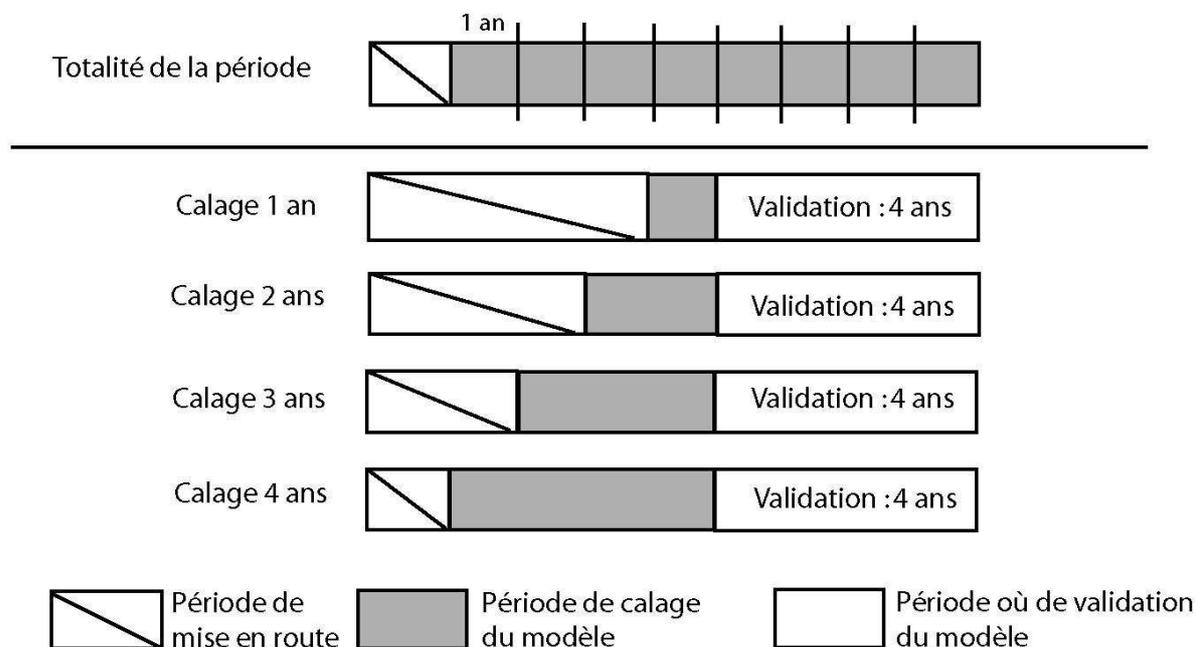


Figure 5-25 : Procédure d'évaluation de l'influence de la durée de calage sur les performances des modèles.

La Figure 5-26 montre l'évolution des performances des 20 modèles testés, en fonction de la durée de la chronique de calage. Sur cet exemple, les modèles sont calés avec la fonction objectif $C_{2M}(Q)$ et les performances en calage sont évaluées avec les fonctions $C_{2M}(Q)$ et $C_{2M}(RQ)$. Cette figure indique clairement qu'en calage, les performances diminuent progressivement avec l'augmentation de la durée de la période de calage. Cette tendance est très claire et concerne l'ensemble des modèles, pour les deux critères d'évaluation utilisés. De la même manière, la Figure 5-27 montre les performances en validation de ces 20 modèles, évaluées avec les fonctions $C_{2M}(Q)$ et $C_{2M}(RQ)$. La tendance mise en évidence en calage s'inverse et on voit clairement une augmentation progressive des performances en validation avec la durée de la période de calage. En fait ce résultat vient certainement du fait que plus la durée de calage est longue, plus la probabilité de rencontrer une grande diversité de conditions hydro-climatiques est grande. Ceci diminue les performances en calage, car il est plus difficile de trouver un vecteur de paramètre qui s'adapte aussi bien que pour des périodes de calage courtes. En validation, l'effet est inverse, car des périodes plus longues en calage permettent aux vecteurs de paramètres

d'être moins adaptés aux conditions typiques d'une année hydrologique et ainsi d'être plus robuste.

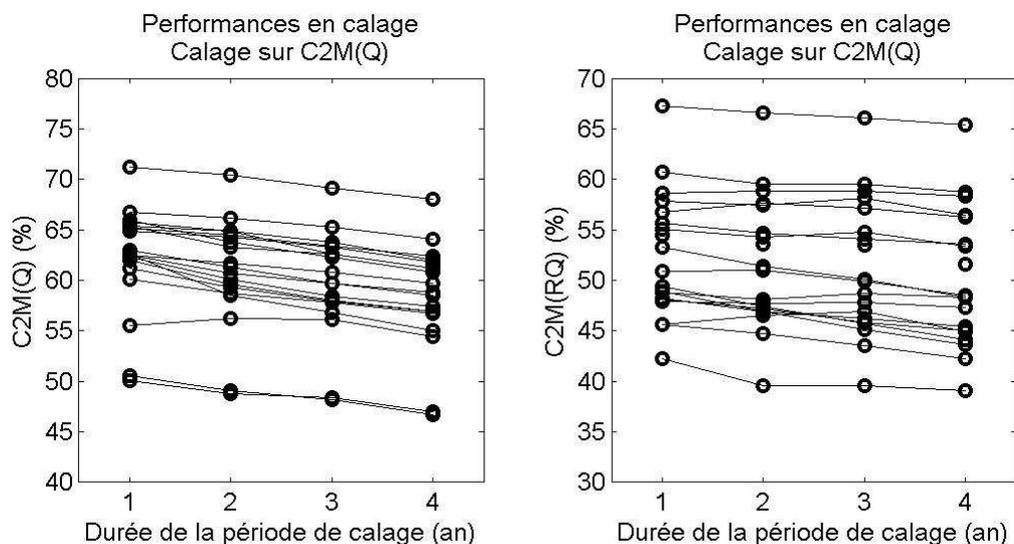


Figure 5-26 : Influence de la durée des périodes de calage sur les performances moyennes des modèles en calage. Calage sur $C_{2M}(Q)$ et évaluation des performances en calage sur $C_{2M}(Q)$ et $C_{2M}(RQ)$.

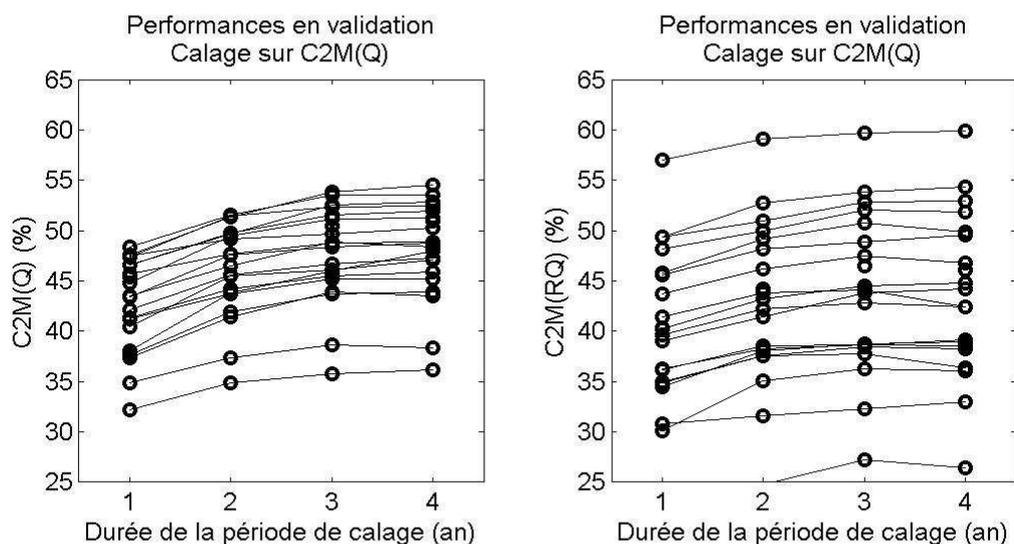


Figure 5-27 : Influence de la durée des périodes de calage sur les performances moyennes des modèles en validation. Calage sur $C_{2M}(Q)$ et évaluation des performances en validation sur $C_{2M}(Q)$ et $C_{2M}(RQ)$.

La Figure 5-28 synthétise ces résultats et compare les performances moyennes en validation des modèles, pour une période de calage de 1 an et 4 ans. Pour les deux critères d'évaluation étudiés, on voit clairement la supériorité des performances en validation lorsque la durée de calage est de 4 ans. Pour le critère d'évaluation $C_{2M}(Q)$, les gains moyens se situent autour de 5 points de C_{2M} et peuvent atteindre 8 à 10 points pour certains modèles. Pour le critère d'évaluation $C_{2M}(RQ)$, les gains

sont logiquement inférieurs (ce n'est pas le critère de calage) mais sont de l'ordre de 2 à 3 points en moyenne. Ceci indique donc qu'une durée de 4 ans permet d'améliorer les performances du modèle pour le critère de calage, $C_{2M}(Q)$, mais également pour d'autres critères, ce qui signifie pour différentes gammes de débit. Une plus longue durée de calage améliore donc la robustesse générale du modèle.

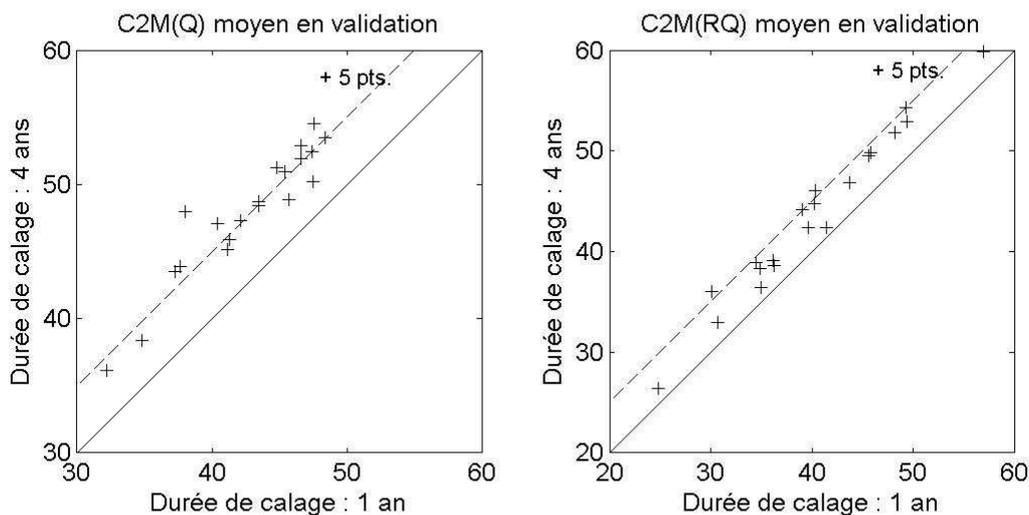


Figure 5-28 : Comparaison des performances en validation des 20 modèles pour une durée de calage de 1 an et 4 ans. Calage sur $C_{2M}(Q)$, validation sur $C_{2M}(Q)$ ou validation sur $C_{2M}(RQ)$.

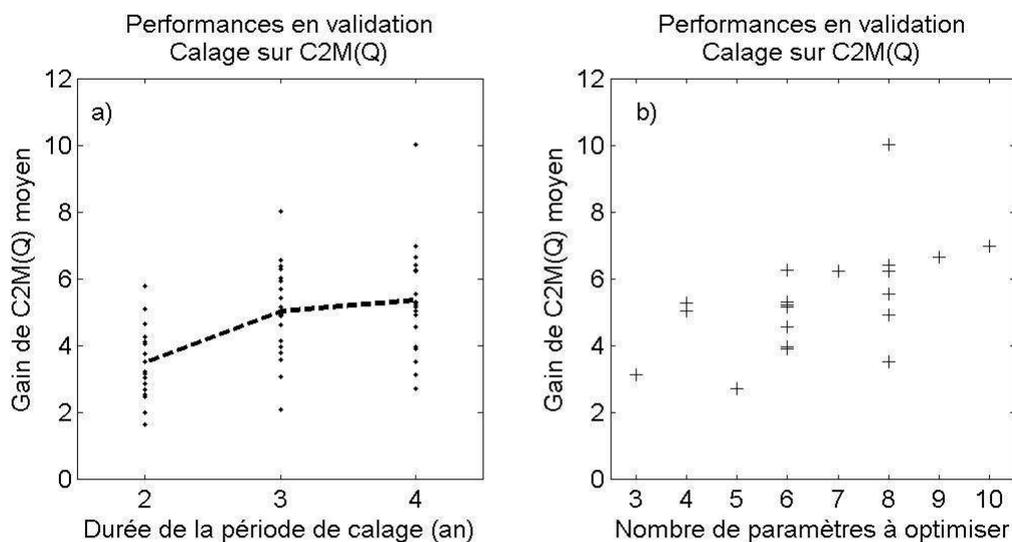


Figure 5-29 : a) Gains de performance en validation des 20 modèles pour une durée de calage de 2, 3 et 4 ans, par rapport à une durée de calage de 1 an. b) Gains de performances en validation des 20 modèles pour une durée de calage de 4 ans, par rapport à une durée de calage de 1 an, en fonction du nombre de paramètres à optimiser.

Enfin, la Figure 5-29 a) indique les différents gains de performances des 20 modèles pour des durées de calage de 2, 3, 4 ans par rapport à la durée initiale de calage de

1 an. Cette figure montre la diversité des gains possibles des modèles, de 2 à 6 points pour une durée de 2 ans et de 2.5 à 10 points pour une durée de 4 ans.

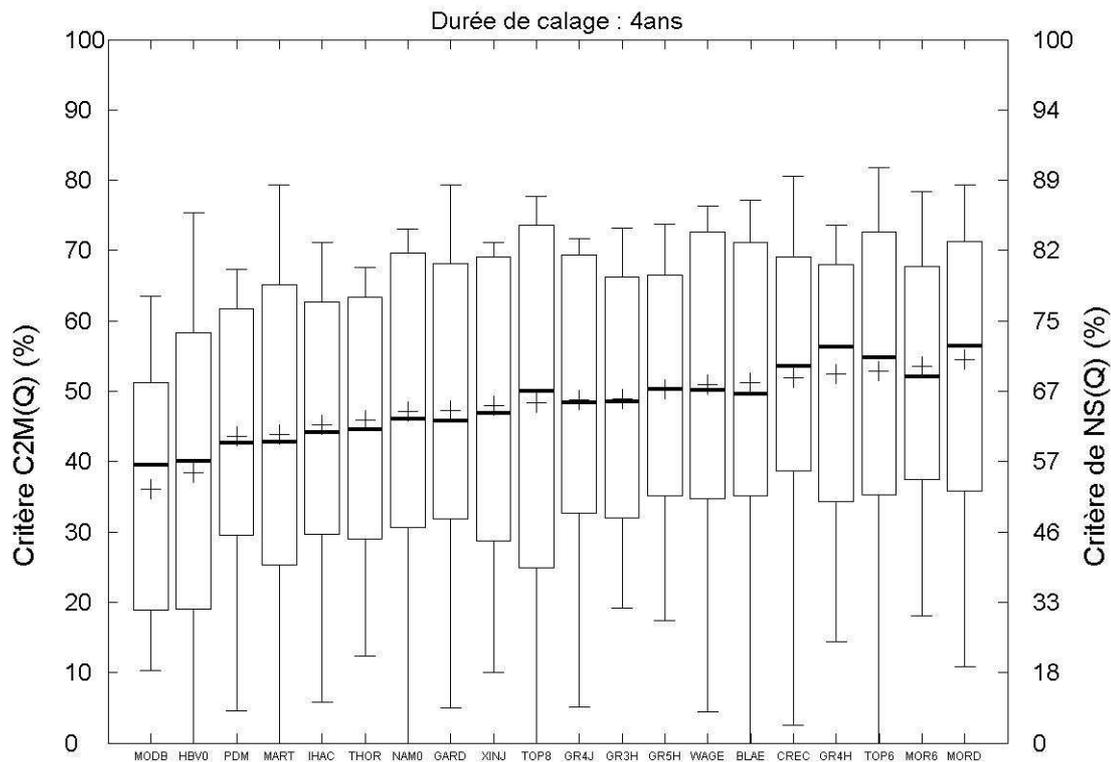


Figure 5-30 : Performances en validation des 20 structures de modèle testées, pour une durée de calage de 4 ans. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(Q)$.

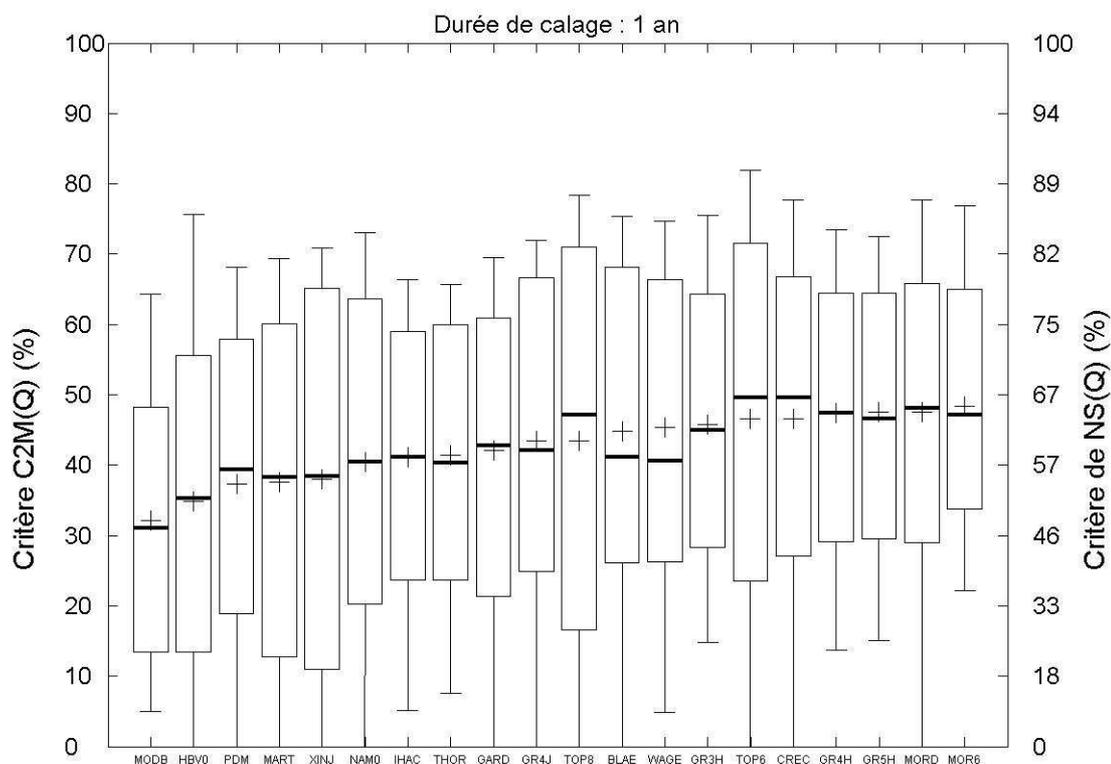


Figure 5-31 : Performances en validation des 20 structures de modèle testées, pour une durée de calage de 1 an. Calage sur C2M(Q) et validation sur C2M(Q). Classement croissant selon le moyenne (+).

En moyenne, les gains sont de 3 points pour une durée de 2 ans à 5 points pour une durée de 4 ans. En fait il semble qu'une durée de 2 ans suffise déjà pour améliorer les performances et la robustesse des modèles. La Figure 5-29 b) montre les gains de performances entre des durées de calage de 1 an et 4 ans, en fonction du nombre de paramètres. Cette figure montre qu'il semble bien que les modèles parcimonieux progressent plus modestement que les modèles plus complexes. Par exemple, pour les modèles ayant de 3 à 6 paramètres, les gains de performances sont de 2.5 à 6 points. Pour les modèles ayant plus de 6 paramètres, les gains sont de 3 à 10 points. Cette figure montre que le classement des modèles de complexité différentes peut être éventuellement modifié, en fonction de la durée de calage des modèles : plus la durée étant grande, plus les modèles complexes semblent capable de tirer profit de cette information.

Ainsi, la Figure 5-30 et la Figure 5-31 montrent les performances des modèles classés par ordre croissant de leurs performances, pour une durée de calage de 4 ans et 1 an. Sur ces deux figures on remarque la progression des performances avec la durée de calage, que nous avons déjà constatée. On voit également l'intérêt des plus longues périodes de calage, qui permettent de faire progresser la robustesse (les valeurs des quantiles 0.1 ont progressé en moyenne de 11 points) et les fortes performances des modèles (les valeurs des quantiles 0.9 ont progressé en moyenne de 3 points seulement, mais pour certains modèles de 10 points).

Par rapport aux résultats de la comparaison précédente, les classements des modèles ne sont pas bouleversés dans l'ensemble. Ce test sur un sous-échantillon réduit de bassins versants et pour des durées de 1 an et 4 ans de calage permet de mettre mieux en évidence les bonnes performances de certains modèles, pour lesquels la comparaison précédente était plus exigeante. On remarque ainsi les structures de TOP6, CREC, BLAE et WAGE.

5.9 Synthèse

Le cadre de comparaison que nous avons suivi [Perrin, 2000], dont nous avons montré l'intérêt et l'objectivité (Chapitres 3 et 4), nous a permis d'évaluer les performances de 20 structures de modèles pluie-débit, sur un échantillon hétérogène de 313 bassins versants. Ces tests représentent la première comparaison de différentes structures de modèles, testées au pas de temps horaire, sur un nombre conséquent de bassins versants.

Par rapport aux résultats de *Perrin* [2000], au pas de temps journalier sur un échantillon de 429 bassins versants, les meilleurs des modèles aux pas de temps horaire obtiennent un niveau de performance moyenne proche de celui atteint au pas de temps journalier. Cependant, il semble qu'au pas de temps horaire les différences de performances entre modèles soient plus grandes et plus discriminantes : un petit nombre de modèles obtiennent un très bon niveau de performance et ensuite les performances des autres modèles diminuent assez rapidement.

En étudiant la robustesse des modèles (différence de performance entre la phase de calage et la phase de validation), en fonction de leur nombre de paramètres optimisés, il apparaît clairement que 4 à 6 paramètres semblent suffisants pour identifier la relation pluie-débit correctement au pas de temps horaire. En effet, quelques structures de modèles ayant 4, 5 et 6 paramètres libres, sont capables d'être au moins aussi bons que d'autres structures ayant jusqu'à 10 paramètres libres. Ces modèles plus complexes ne sont pas capables de tirer profit robustement de leurs plus grandes capacités d'adaptation au cours du calage et voient leurs performances fortement diminuer en validation. Ils ont généralement des problèmes de sur-calage et de pauvre définition de leur surface de réponse (Chapitre 0). Ces résultats vont également dans le sens de *Beven* [1989], *Jakeman et Hornberger* [1993] et *Perrin* [2000] qui préconisent de 3 à 6 paramètres libres pour reproduire la relation pluie-débit au pas de temps journalier. A niveau de performance égal, nous pensons que la parcimonie d'un modèle est un facteur très important, qui doit être recherché. En effet, l'étude de la surface de réponse du modèle MORDOR l'a montré : la faible définition de sa surface de réponse et la faible robustesse des vecteurs de paramètres optimisés (et également, la faible confiance que l'on peut accorder à ses vecteurs !). Dans une perspective de prévision des crues, de maîtrise des incertitudes associées à la simulation/prévision hydrologique ou de régionalisation, la parcimonie des modèles est un enjeu capital.

Parmi les 20 structures de modèles testées, les modèles de la famille GR (GR4H et GR5H) et le modèle MOR6 sont les plus efficaces et les plus parcimonieux. Ces résultats soulignent l'intérêt du développement continu de structures de modèles et du test de nouvelles structures, car ce sont ces nouvelles structures qui ont les meilleures performances au pas de temps horaire (Figure 5-32). La structure de MORDOR, qui n'avait jamais été comparée sur un échantillon aussi large de bassins versants, donne d'excellents résultats.

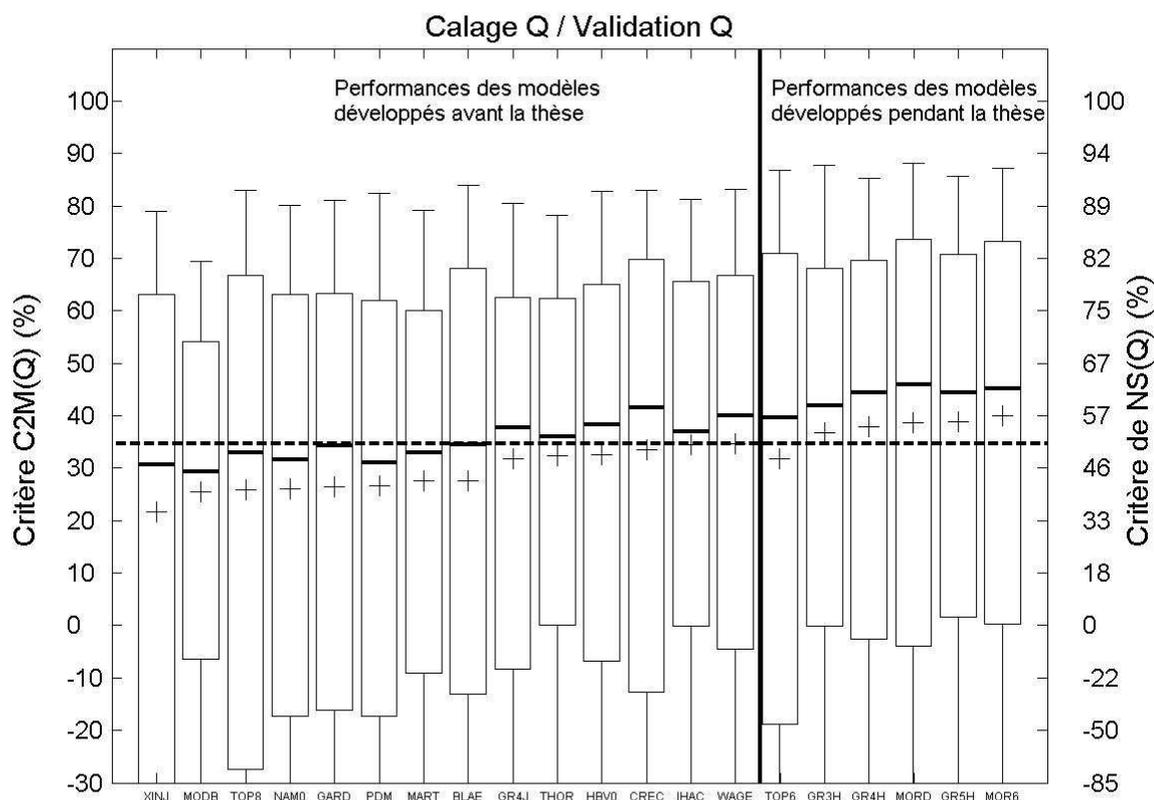


Figure 5-32 : Performances en validation des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(Q)$. Comparaison des performances des structures déjà développées et testées avant la thèse et de celles de structures développées pendant cette thèse.

Les études sur la complémentarité des modèles montrent qu'il est possible d'espérer un gain de performances moyen d'environ 3 à 4 points, en couplant deux structures de modèles. Les modèles efficaces au pas de temps horaire, basés sur des philosophies de modélisation différentes, étant peu nombreux (famille GR, famille MORDOR, IHAC6, WAGE), il serait judicieux de poursuivre cette étude de complémentarité des modèles, par une approche multi-modèle, notamment comme le proposent *Oudin et al.* [soumis]. Il serait intéressant de comparer les gains potentiels de cette approche multi-modèles, à ceux obtenus grâce à l'amélioration de la structure d'un modèle (Chapitre 1) ou grâce à un meilleur traitement de la variabilité des pluies (Chapitre 9).

Enfin, les études que nous avons menées sur la recherche de relations entre les performances des modèles pluie-débit et des descripteurs basiques des bassins versants n'ont pas mis en évidence de facteurs significativement discriminants. Ces résultats vont dans le sens de ceux de *Merz et Blöschl* [2004], qui ont tenté de régionaliser les paramètres du modèle HBV pour un grand nombre de bassins versants d'Autriche. Ce résultat peut étonner car de nombreux modélisateurs croient

(encore !) qu'il faut développer une structure de modèle pour différents types de bassins versants. Au contraire, ces résultats ont montré que quelques structures (famille GR, famille MORDOR, IHAC6, WAGE) sont capables d'être les plus efficaces sur l'ensemble des 313 bassins versants. A part les modèles de la famille GR, les modèles de la famille MORDOR ont été développés plutôt pour des bassins montagneux de France, le modèle WAGE a été développé et testé sur un bassin versant des Pays-Bas et le modèle IHACRES a été développé sur des bassins versants anglais. Rien ne destinait ces trois derniers modèles à avoir de bonnes performances sur un échantillon aussi hétérogène de bassins versants. Parmi les 20 structures de modèles testées, seul le modèle XINANJIANG a ses performances clairement meilleures sur les bassins versants semi-arides, comme cela a déjà été constaté par *Gan et al.* [1997].

Chapitre 6

**« 1947 : Il abandonne le métier d'ingénieur
pour se consacrer uniquement aux arts. »**

Extrait de la biographie de Boris Vian.

CHAPITRE 6 Approche empirique de mise au point d'un modèle pluie-débit au pas de temps horaire

6.1 Introduction

Parmi les 20 structures de modèles dont les performances ont été comparées au Chapitre 5, quatre structures sont issues de l'amélioration empirique de modèles au pas de temps horaire : GR3H, GR4H, MOR6 et TOP6. Si on reprend cette comparaison des performances des 15 structures de modèles initialement disponibles, on se rend compte que (Figure 6-1) :

- quelques structures à 6, 8 et 10 paramètres ont des performances sensiblement meilleures que la structure la plus parcimonieuse (GR4J) ;
- il semble exister une tendance entre performances maximales et niveau de complexité des modèles (nombre de paramètres) (Figure 6-1).

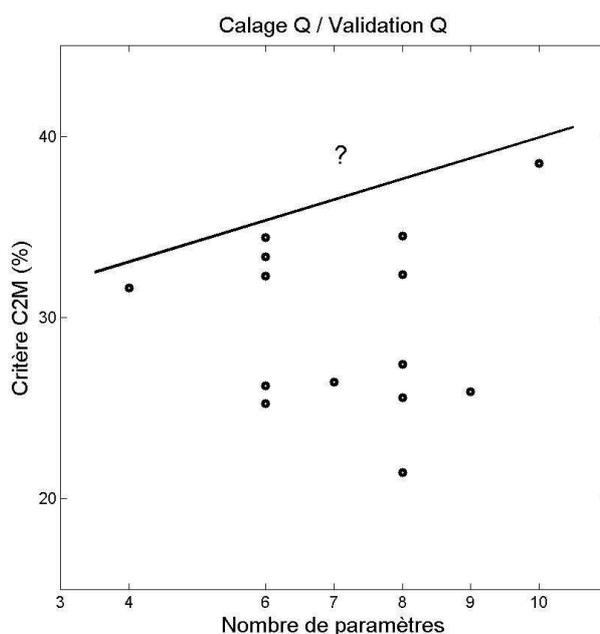


Figure 6-1 : Bilan des performances au pas de temps horaire des 15 structures de modèles disponibles au début de cette thèse.

Ce constat est le point de départ de notre démarche empirique d'amélioration de la structure de modèles au pas de temps horaire. La différence de performance entre le modèle GR4J et le modèle MORD, les résultats des tests de robustesse, de fiabilité et de recherche de complémentarité, nous laissent espérer une amélioration des performances de la structure de GR4J.

Dans ce chapitre, nous allons tout d'abord présenter l'ensemble des tests effectués à partir de la structure de GR4J et la structure vers laquelle les résultats nous ont emmenés. Ensuite, nous présenterons une démarche analogue, que nous avons conduit sur les modèles MORD et TOP8, mais qui peut être appliquée à d'autres modèles.

6.2 Quelle démarche de développement de modèle suivre ?

6.2.1 Synthèse des travaux réalisés au Cemagref

Le modèle GR4J est le résultat de plus de 20 ans d'améliorations progressives, d'un modèle fonctionnant au pas de temps journalier. Ce long processus de mise au point d'une structure est assez courant et d'autres centres de recherche travaillent à l'amélioration des structures de modèles depuis de nombreuses d'années : cas des modèles TANK, SMAR, HBV, IHACRES, TOPMODEL, etc.. La structure de GR4J résulte d'une structure à 1 paramètre [Michel, 1983] et des améliorations successives de *Edijatno et Michel* [1989a], *Edijatno* [1991], *Nascimento* [1995], *Edijatno et al.* [1999] et finalement de *Perrin* [2000]. La structure actuelle, à 4 paramètres, a été testée et développée sur plusieurs centaines de bassins versants aux caractéristiques variées, situés dans le monde entier. Actuellement, *Rojas-Serna* [2005], utilise cette structure sur plus de 1000 bassins versants pour développer une méthode d'application à des bassins versants dépourvus d'une station hydrométrique pérenne.

Au Cemagref, les développements de modèles hydrologiques de la famille GR se sont appuyés sur une démarche empirique, qui trouve ses sources dans les travaux fondateurs de *Nash et Sutcliffe* [1970], *Bergström et Forsman* [1973] et *Klemeš* [1986b], comme nous l'avons vu au Chapitre 1. Au cours des développement successifs de la structure d'un modèle, le respect de la parcimonie et de la simplicité de la structure a été une règle de base et une complexification de la structure du modèle n'a été acceptée que si elle était justifiée par une amélioration significative des performances du modèle, confirmée par un échantillon substantiel de bassins versants (propositions n°3 et n°4 du Chapitre 1).

Dernièrement, les travaux de *Mouelhi* [2003] ont porté sur la modélisation de la relation pluie-débit aux pas de temps pluri-annuel, annuel et mensuel (Figure 6-2). En suivant la méthodologie précédente, cet auteur a testé et développé des modèles à ces différents pas de temps, sur le même échantillon de bassins versants que celui utilisé pour développer GR4J, au pas de temps journalier. Pour chaque pas de temps, la structure de modèle la plus performante a été recherchée. Ces travaux ont aussi concerné la cohérence de cette famille de modèles et mis en évidence la

Chapitre 6 : Approche empirique de mise au point d'un modèle pluie-débit au pas de temps horaire

complexification progressive de la structure des modèles hydrologiques lorsque le pas de temps des données diminue (Tableau 6-1).

Tableau 6-1 : Synthèse des travaux de Mouelhi [2003] sur la chaîne de modèles hydrologiques développés, du pas de temps pluri-annuel jusqu'au pas de temps journalier.

Pas de temps	Pluri-annuel	Annuel	Mensuel	Journalier
Modèle	GR0P	GR1A	GR2M	GR4J
Nombre de paramètres libres	0	1	2	4
Nombre de réservoirs	0	0	2	2
Échanges souterrains	Non	Non	Oui	Oui
Hydrogramme unitaire	Non	Non	Non	Oui

Ces travaux mettent en évidence la nécessité d'une fonction de perte, que l'on appelle 'échanges souterrains'. Une fonction mémoire apparaît, sous forme simplifiée, dans la représentation du système hydrologique à partir du pas de temps annuel. Cette fonction mémoire se développe ensuite au pas de temps mensuel (en nécessitant l'utilisation d'un réservoir). Ensuite au pas de temps journalier, la fonction de routage comprend un hydrogramme unitaire. En suivant cette approche de complexification progressive d'une structure de modèle, nous faisons également l'hypothèse qu'il est plus fructueux de complexifier une structure de modèle en faisant progresser ses performances (propositions n°3 du Chapitre 1), plutôt que de simplifier une structure de modèle, tout en faisant progresser ses performances (au moins, sans les faire diminuer). Ces deux approches seront testées par la suite.

Par ailleurs, les travaux de *Mouelhi et al.* [2005] montrent bien qu'une structure de modèle mensuel à deux paramètres est capable d'être plus performante que 8 autres structures tirées de la littérature, comportant jusqu'à 5 paramètres libres. De la même manière, *Perrin* [2000] a également démontré qu'une structure à 4 paramètres au pas de temps journalier était au moins aussi performante ou meilleure que d'autres structures plus complexes. Par ailleurs, comme le montrent les nombreuses publications dans la littérature, il est largement accepté que les modèles mensuels sont plus simples que les modèles journaliers. On peut se demander ainsi pour quelle raison un modèle développé au pas de temps journalier devrait être nécessairement adapté au pas de temps horaire.

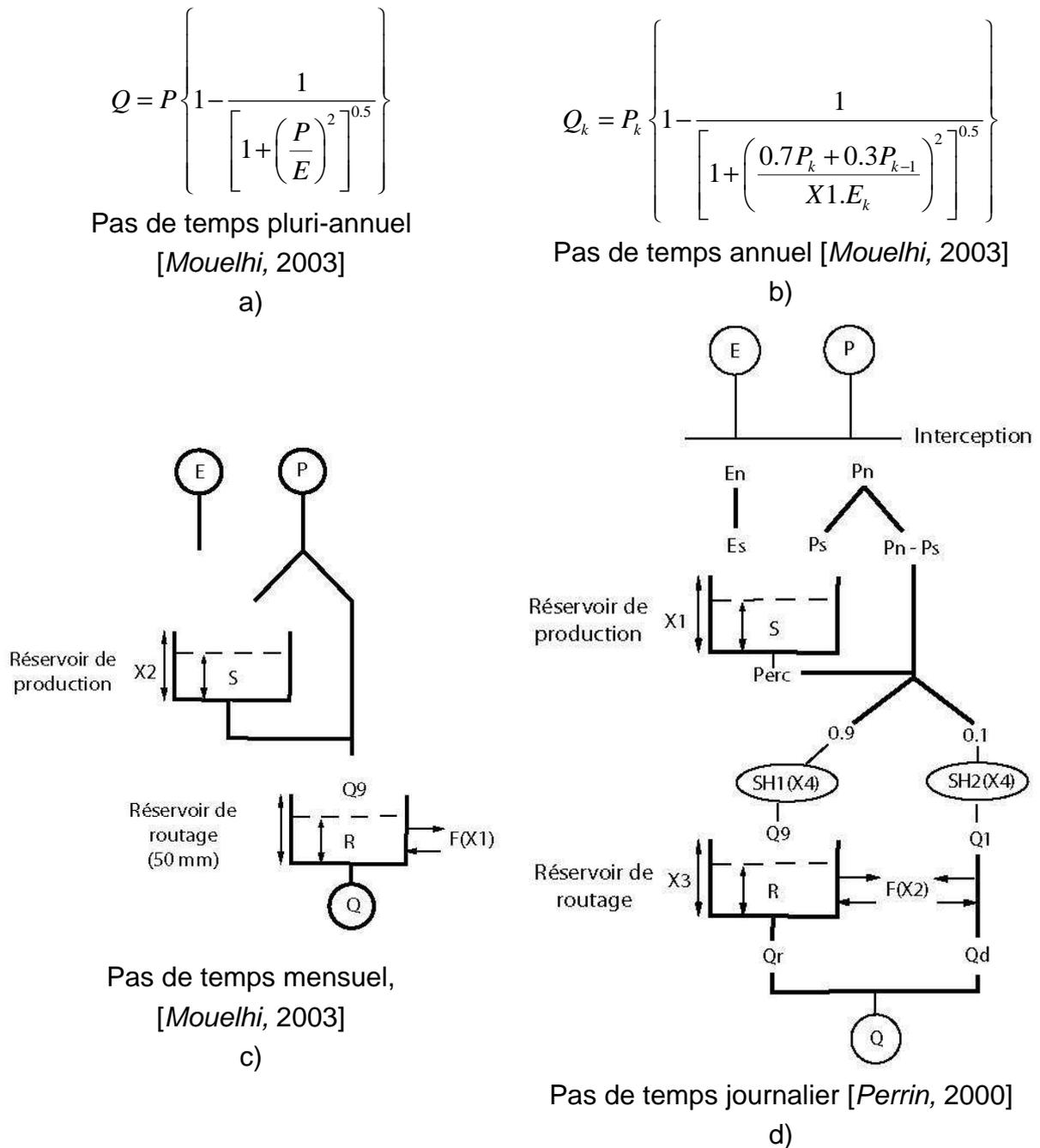


Figure 6-2 : Chaîne de modèles développés au Cemagref. Pas de temps, a) pluri-annuel, b) annuel, c) mensuel et d) journalier.

Nous pensons donc qu'il n'y a aucune raison *a priori* de ne pas s'interroger sur la structure et le niveau de complexité le plus adapté qu'un modèle devrait avoir pour représenter efficacement la relation pluie-débit au pas de temps horaire.

6.2.2 Quels progrès attendre pour un paramètre de plus ?

Cette question est une question assez difficile, pour laquelle il n'existe pas de réponse précise. L'idée est de se fixer un seuil minimal de gain de performance (en points de C_{2M}) qu'il faut dépasser, pour accepter de complexifier davantage la structure d'un modèle. Ce seuil correspond au gain que l'on considère significatif

pour valider une modification de la structure d'un modèle. L'idée est que chaque paramètre rajouté doit apporter une fonction qui est nécessaire au bon fonctionnement du modèle, validée par un gain significatif de performance.

D'après les tests effectués au Chapitre 3, il faudrait que la différence de performance entre deux modèles soit d'au moins 5 points pour que les performances de deux modèles puissent être considérées comme significativement différentes. D'après l'expérience acquise au Cemagref, au cours du développement successif de structures de modèles, il semble que cette valeur de cinq points soit très forte. La valeur du seuil de gain de performances entre deux structures serait plutôt de deux points pour le critère moyen sur un grand échantillon de bassins versants. Cette valeur peut sembler faible, mais en pratique cette contrainte est déjà assez forte. Cette contrainte permet de valider seulement les modifications de structures qui sont significatives. De plus, en rendant l'exercice plus difficile, elle permet d'aiguiser la créativité et l'imagination du modélisateur. Le nombre de structures possibles étant infini, le recours à une complexification de la structure d'un modèle est généralement un recours en désespoir de cause (fuite en avant), lorsque aucun progrès n'est constaté. De plus, un très grand nombre de structures doivent être testées à niveau de complexité égal ! Imposer une telle contrainte sur le progrès d'une structure sur un large échantillon de bassins versants permet de garantir que la solution proposée soit générale et robuste, et non particulière à quelques bassins versants. Cette démarche, vivement encouragée par *Andersson* [1992] notamment, permet au modélisateur de ne pas se limiter aux idées *a priori* qu'il a sur la structure que *devrait* avoir un modèle hydrologique.

Par exemple, *Perrin* [2000] et *Perrin et al.* [2003] ont accepté d'ajouter un paramètre au modèle journalier, pour le faire progresser de 2 à 4 points sur un échantillon de 429 bassins versants, suivant le critère d'évaluation des performances choisi. De la même manière, *Lindström et al.* [1997] proposent une nouvelle structure du modèle HBV, pour un gain de 3 points sur 7 bassins versants (seulement !).

6.2.3 Méthodologie de recherche d'une meilleure structure et présentation des tests effectués

Les travaux qui suivent une approche descendante à une dimension (plusieurs pas de temps, un bassin ou un type de bassins) [*Jothityangkoon et al.*, 2001; *Atkinson et al.*, 2002; *Eder et al.*, 2003] soulignent la complexification progressive de la structure des modèles, lorsque le pas de temps représenté diminue. De la même manière, les travaux de *Mouelhi* [2003] et de *Perrin* [2000], qui suivent finalement une approche descendante à deux dimensions (plusieurs pas de temps, un grand nombre de

bassins hétérogènes) mettent également en évidence cette complexification progressive de la structure des modèles.

Première voie

Notre stratégie d'amélioration d'une structure de modèle relève donc d'une approche descendante. Par ailleurs, le pas de temps horaire permettant de mieux représenter les intensités de la pluie, nous tenterons également de mieux modéliser une capacité maximale d'infiltration. Pour tester ces hypothèses, nous sommes partis de la structure du modèle GR4J, ce modèle étant le plus simple des 15 structures testées et suffisamment performant.

Notre tentative d'amélioration d'un modèle au pas de temps horaire suit trois directions :

1. Remise en cause des paramètres fixes du modèle initial (GR4J),
2. Tentatives de complexification de la fonction de routage du modèle,
3. Tentatives de prise en compte des intensités de la pluie dans le modèle.

Seconde voie

La seconde voie suivie est la voie alternative, opposée, celle de la réduction progressive de la complexité d'un modèle. Cette approche s'apparente à une analyse de sensibilité à grande échelle de ces modèles. Nous présenterons les tests que nous avons effectués sur les structures des modèles MORDOR et TOPMO8. La réduction du nombre de paramètres de modèles complexes a déjà été pratiquée par de nombreux auteurs [*Nielsen et Hansen, 1973; Mein et Brown, 1978; Chiew et McMahon, 1994; Zhao et Liu, 1995; Tan et O'Connor, 1996; Boughton, 2005b*]. Ils ont montré que la réduction du nombre de paramètres, parfois drastique, ne diminuait que très légèrement les performances de ces modèles.

Comme nous l'avons dit précédemment, l'éventail des modifications possibles est très vaste. Toutes les modifications peuvent être testées seules ou combinées avec d'autres. Ainsi, la démarche empirique est généralement longue et fastidieuse puisque nous procédons à des tests aussi exhaustifs que possibles de nombreuses structures de modèle (qui sont autant d'hypothèses faites sur la relation pluie-débit). La progression par essais et erreurs est généralement très lente et il convient de rester humble face aux résultats et aux échecs, car une très petite partie du champ des possibles est testée. Dans notre démarche, nous avons testé chaque version de modèle sur l'échantillon de 313 bassins versants. Nous avons calé les modèles avec le critère $C_{2M}(RQ)$ et évalué le modèle en contrôle avec le critère $C_{2M}(Q)$. Nous avons également vérifié que l'écart-type des performances du modèle reste stable. En effet, l'écart type nous renseigne sur la robustesse du modèle.

L'ensemble des tests réalisés pour tenter d'améliorer la structure de GR4J au pas de temps horaire représente plus d'une année de travail, plusieurs milliers d'heures de calculs et le test de 400 à 500 structures de modèles différentes.

6.3 Tentatives d'amélioration d'une structure au pas de temps horaire par une approche empirique de complexification progressive

Les grandes lignes de cette démarche sont résumées à la Figure 6-3.

6.3.1 Remise en cause au pas de temps horaire des paramètres fixes de GR4J

L'objectif de ces tests est de vérifier s'il ne faut pas adapter certaines valeurs des paramètres fixes de GR4J (valeurs des puissances et coefficients des différentes fonctions du modèle), qui ont été déterminées sur une base de données aux pas de temps journalier.

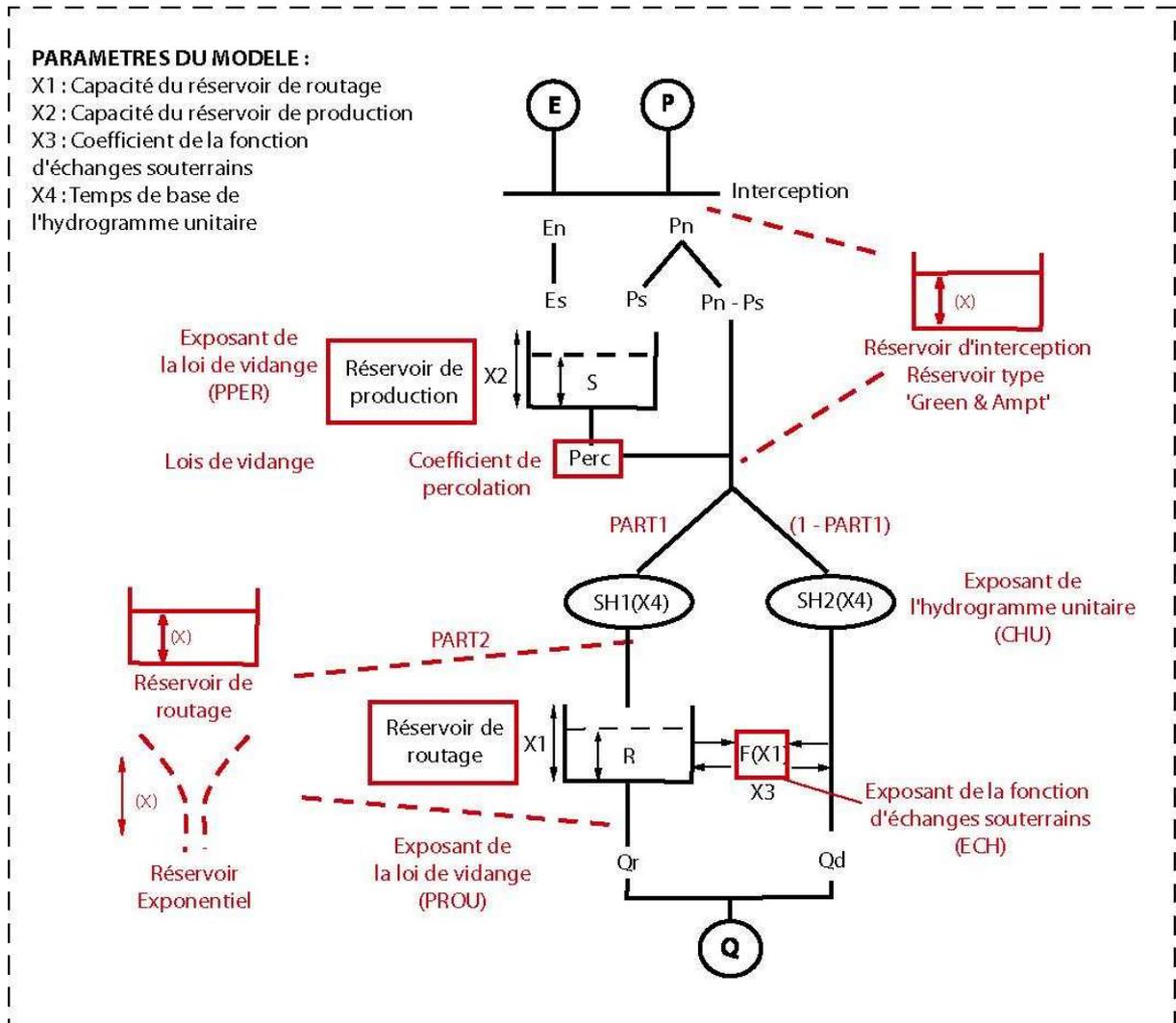


Figure 6-3 : Synthèse de l'ensemble des tests effectués.

Ces tests ont porté sur (nom du paramètre et valeur initiale entre parenthèses) :

- Coefficient de partage (PART1 : 0.9) ;
- Coefficient de la fonction de percolation (PERC : 2.25),
- Exposants de la fonction vidange du réservoir de production (PPER : 3), du réservoir de routage (PROU : 4), de la fonction d'échanges souterrains (PECH : 4), des hydrogrammes unitaires (CHU : 2.5).

Rappelons que la structure du modèle GR4J obtient au pas de temps horaire sur l'échantillon de 313 bassins versants (calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(Q)$) une performance moyenne $\overline{C_{2M}} = 32.6$ avec un écart-type $\sigma_{\overline{C_{2M}}} = 26.6$.

Chaque paramètre fixe (PART1, PERC, PPER, PROU, PECH et CHU) a été examiné. Pour cela nous avons testé plusieurs valeurs autour de la valeur initiale et nous avons également tenté d'optimiser cette valeur. Parmi les modifications qui donnent des résultats significatifs, les plus importantes sont celles de la modification du coefficient de la loi de percolation (PERC : 4 au lieu de 2.25) et de la puissance de la fonction de l'hydrogramme unitaire (CHU : 1.25 au lieu de 2.5). Par contre la valeur des puissances des lois de vidange reste inchangée ou très légèrement modifiée (PECH : 3.5 au lieu de 4.0). L'optimisation, bassins par bassins, de ces paramètres n'apporte aucun gain de performances, mais diminue au contraire la robustesse du modèle. Le cas du coefficient de partage (PART1) est particulièrement intéressant car dans de nombreuses structures de modèles, ce coefficient de partage est souvent optimisé (IHACRES, par exemple), alors que dans le cas de GR4J, son optimisation diminue sensiblement les performances en validation.

En fait le premier résultat n'est pas étonnant car la valeur du coefficient de la fonction de percolation dépend du pas de temps d'intégration. Ainsi, la valeur trouvée empiriquement est proche de celle trouvée théoriquement, en passant du pas de temps journalier à horaire (Annexe 8). La modification de la puissance de la fonction de l'hydrogramme unitaire est plus surprenante. En effet, la meilleure valeur trouvée est de 1.25, alors qu'elle était initialement de 2.5. Cela signifie que la forme de l'hydrogramme unitaire devient particulièrement écrasée (Figure 6-4).

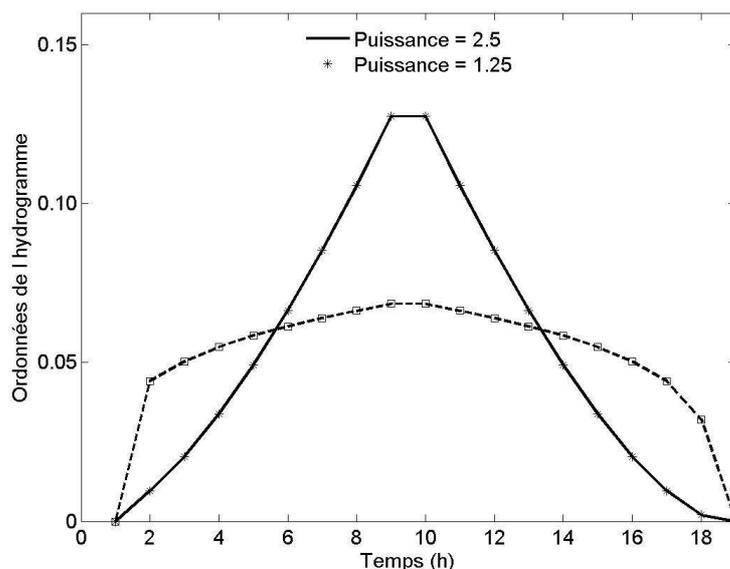


Figure 6-4 : Forme de l'hydrogramme unitaire HU2, pour deux valeurs de puissances (l'ancienne : 2.5 et la nouvelle : 1.25). Temps de montée : 9h.

En fait, cette modification vient peut être du fait qu'au pas de temps horaire, il est moins nécessaire de répartir les évènements de pluie de manière aussi contrastée qu'au pas de temps journalier. En effet, le détail horaire des évènements pluvieux apporte peut être suffisamment de nervosité au modèle et il n'est peut être pas nécessaire d'introduire de la variation dans les évènements pluvieux, comme au pas de temps journalier. Nous avons également essayé de simplifier l'utilisation de deux hydrogrammes unitaires. Effectivement, l'utilisation d'un seul hydrogramme unitaire à deux branches ne modifie pas les performances du modèle et simplifie légèrement sa structure.

Cette première phase de tests a permis de faire le point sur les performances au pas de temps horaire de la structure de GR4J. Les modifications simples proposées permettent de faire progresser très fortement les performances du modèle, sans complexifier davantage sa structure et même en simplifiant la fonction de transfert par l'utilisation d'un seul hydrogramme (Modèle M69, performance moyenne et écart-type : $\overline{C_{2M}} = 37.5 / \sigma_{\overline{C_{2M}}} = 28.1$).

6.3.2 Simplification de la structure M69

Bien que les travaux de *Mouelhi* [2003] nous aient suggéré la nécessité de complexifier la fonction de routage du modèle, il nous a semblé nécessaire de vérifier si une simplification de la structure du modèle pouvait être acceptée par les données*. En effet, dans la fonction de routage, il y a une interaction entre la fonction

* ce qui revient à reculer pour mieux sauter

d'échanges souterrains et le réservoir puissance. Voulant tester l'intérêt de rajouter un second réservoir et ne sachant pas comment gérer l'interaction entre ce nouveau réservoir et la fonction d'échanges souterrains, nous avons essayé d'insérer la fonction d'échanges à d'autres endroits du modèle. Ces tests ont confirmé au passage la nécessité de cette fonction. Lorsqu'on la supprime le modèle perd plus de 15 points de C_{2M} . Lorsqu'on la modifie pour utiliser seulement un coefficient correctif (multiplicatif) appliqué à la pluie efficace (P_n), les pertes de performances sont plus limitées. Dans un premier temps, nous utiliserons cette structure pour les tests qui suivront (Modèle M72 : $\overline{C_{2M}} = 35.7 / \sigma_{C_{2M}} = 26.3$).

6.3.3 Complexification de la fonction de routage : mise en parallèle de deux réservoirs aux capacités optimisées (Figure 6-6)

Ces tests concernent la mise en parallèle de deux réservoirs dans la fonction de routage, dans la structure du modèle M72. Nous avons utilisé deux réservoirs dont on optimise les capacités et dont le partage des alimentations est fixé. Dans un souci d'exhaustivité, nous avons testé toutes les combinaisons possibles de deux réservoirs dont les lois de vidange sont :

- Linéaire,
- Puissance (2, 3, 4, 5),
- Exponentielle.

Entre ces deux réservoirs, nous avons testé des valeurs de partage allant de 0.1 à 0.9 par pas de 0.1, ou de 0.2 à 0.8 par pas de 0.2. Nous avons également testé certaines de ces modifications sur la structure du modèle M69 et pour différentes valeurs de PART1. Ce type de fonction de routage avec au moins deux réservoirs en parallèle est assez courant en hydrologie. On les retrouve dans une vingtaine des structures recensées par [Perrin, 2000].

Introduction d'un réservoir exponentiel

Dans un premier temps, nous avons testé l'ajout d'un réservoir exponentiel en parallèle à un deuxième réservoir (linéaire ou puissance). Ce type de réservoir, dont le fonctionnement a été présenté par Michel *et al.* [2003], n'est pas d'un usage très courant. Néanmoins, il peut s'avérer être un excellent outil pour représenter les étiages, avec un pas de temps aussi faible qu'une heure. On le retrouve dans la structure de TOPMODEL et Perrin [2000] avait déjà remarqué qu'il donnait potentiellement de bons résultats, dans les modifications qu'il avait testées au pas de temps journalier.

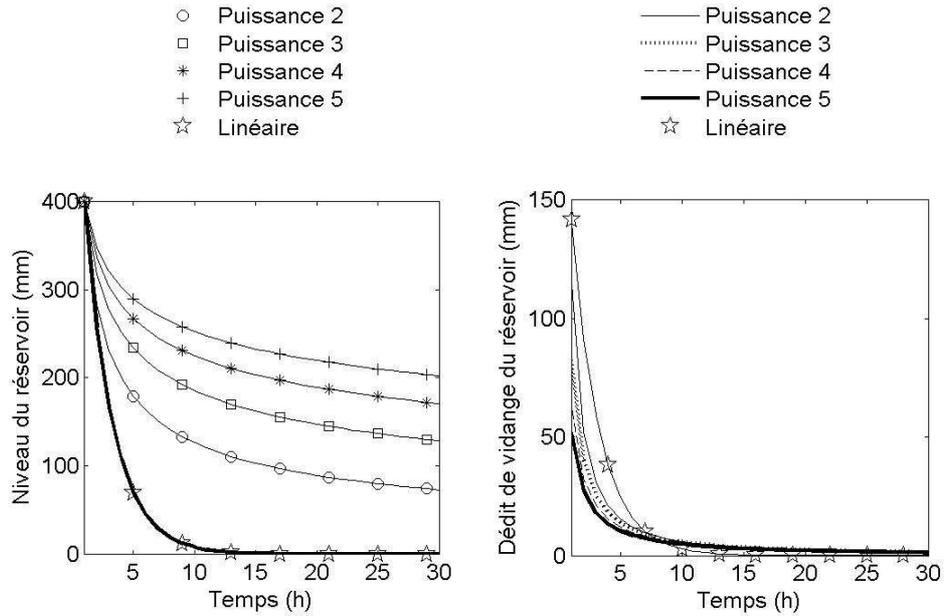


Figure 6-5 : Exemple de vidange d'un réservoir saturé, de capacité maximale 400 mm, initialement rempli à 400 mm, dont la vidange suit une loi puissance (2, 3, 4, 5) ou linéaire (Constante de vidange : 2.8 h.).

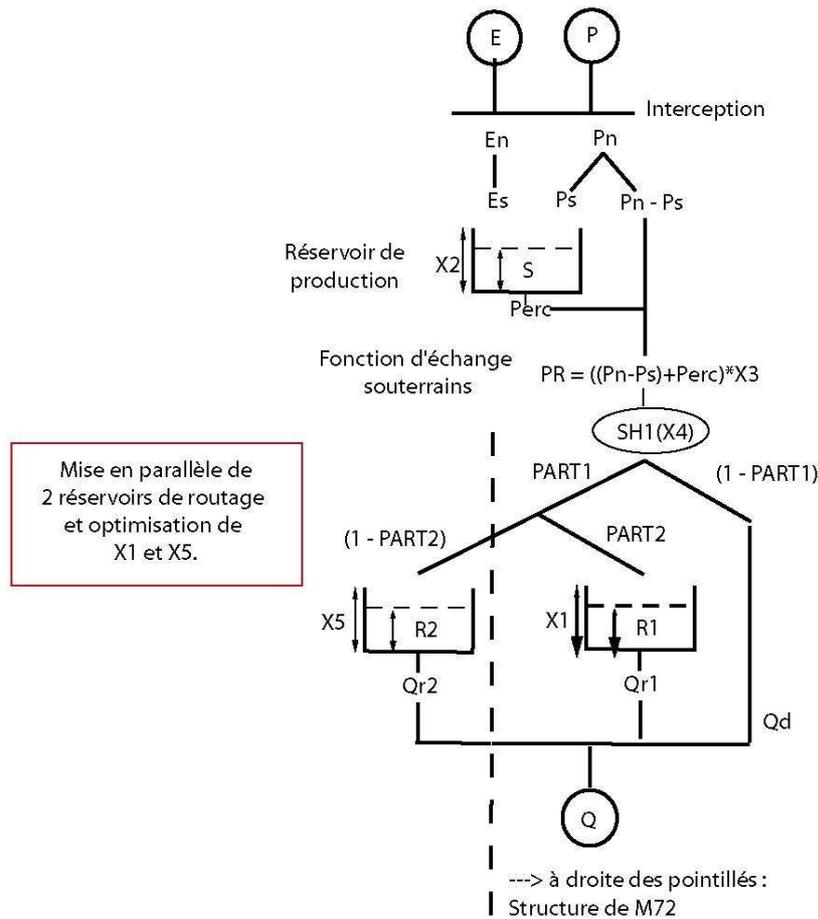


Figure 6-6 : Structure du modèle avec mise en parallèle de deux réservoirs de routage aux capacités optimisées (X1, X5), mais au partage (PART2) fixe.

A partir de la structure M72 (Figure 6-6), nous avons testé la combinaison d'un réservoir exponentiel avec un autre réservoir (linéaire ou puissance), le partage entre ces deux réservoirs étant toujours fixe (PART2). Ces tests n'ont donné aucun résultat probant. Malgré l'ajout d'un degré de liberté, les performances moyennes n'ont pas progressé et la robustesse a diminué. Par ailleurs, lorsqu'on s'intéresse à la distribution des valeurs du coefficient de vidange du réservoir exponentiel, celui-ci est nul dans plus de 50% des cas. Cette valeur signifie que le réservoir exponentiel fonctionne comme une perte, car tous les écoulements reçus sont absorbés. Ces résultats semblent indiquer un rejet du réservoir exponentiel (meilleurs modèles : M156 et M161).

Avant d'exclure définitivement l'utilisation d'un réservoir exponentiel, nous avons effectué les mêmes tests sur la structure du modèle M69 (avec la fonction d'échanges souterrains initiale). Ces tests permettent de constater une légère amélioration des performances de la structure avec un gain de 1 point de C_{2M} (Modèle M124). Néanmoins, ce progrès est insuffisant pour justifier l'ajout d'un paramètre optimisable supplémentaire et l'examen de la distribution des valeurs du coefficient de vidange du réservoir exponentiel indique que celui-ci est infini dans plus de 50% des cas. Cette valeur signifie que le réservoir exponentiel fonctionne comme un apport quasi constant (Modèle M124 : $\overline{C_{2M}} = 38.4 / \sigma_{C_{2M}} = 27.3$). Une autre hypothèse est que ce nouveau réservoir n'apporte pas à la structure du modèle horaire une fonction suffisamment complémentaire à ses fonctions déjà existantes, mais qu'il est au contraire redondant.

Dans les deux séries de tests que nous avons effectuées avec le réservoir exponentiel, il semble que la faible progression des performances et que le rejet de la fonction de réservoir par l'adoption de coefficients de vidange extrêmes (0 ou l'infini) indiquent probablement que cet outil n'est pas approprié, pour la modélisation au pas de temps horaire, malgré les gains de performances constatés.

Par ailleurs, ces deux séries de tests montrent un exemple typique d'interactions entre plusieurs fonctions d'un modèle (ici, la fonction d'échanges souterrains et le réservoir exponentiel), dont les effets sont assez difficiles à appréhender *a priori*. Cet exemple doit permettre d'illustrer l'intérêt des modèles parcimonieux, où le nombre limité de composantes dans le modèle permet de mieux comprendre le fonctionnement général.

Introduction d'un réservoir linéaire ou puissance

Suite à la tentative infructueuse d'introduction d'un réservoir exponentiel, à partir de la structure M72, nous avons testé la mise en parallèle systématique de deux réservoirs (linéaires ou puissance, de 2 à 5), pour différentes valeurs de partage fixe (PART2). Ces tests exhaustifs, ont permis de constater quelques progrès lorsqu'un réservoir quadratique est mis en parallèle à un réservoir puissance 3 ou 4 (Modèle M204 et M211 ; Tableau 6-2). Ces résultats sont assez étonnants, car notre hypothèse *a priori* était qu'il était plus intéressant pour un modèle d'utiliser des réservoirs aux fonctionnements différents, afin d'offrir plus de diversité de fonctionnement au modèle.

Tableau 6-2 : Moyenne et écart-type du meilleur modèle obtenu lorsqu'on met en parallèle deux réservoirs, dont on optimise les capacités. Entre parenthèses, valeur de PART2, de la meilleure combinaison. RL : réservoir linéaire ; R2 à R5 : réservoir puissance 2 à 5 ; EXP : réservoir exponentiel.

PART1=0.9 M72 35.7±26.3		1-PART2					
	Type de Réservoir	RL	R2	R3	R4	R5	EXP
PART2	RL	30.5±24.4 (0.6)	35.2±24.7 (0.4)	36.2±25.8 (0.4)	36.5±26.4 (0.2)	36.5±25.7 (0.4)	33.1±25.8 (0.4)
	R2	X	36.7±26.0 (0.6)	M204 37.0±25.7 (0.6)	M211 37.0±26.1 (0.4)	36.6±25.5 (0.2-0.6)	M156 36.1±26.2 (0.6)
	R3	X	X	35.9±26.5 (0.4)	36.7±26.5 (0.6)	36.6±26.6 (0.4)	M161 36.1±26.7 (0.8)
	R4	X	X	X	35.9±26.7 (0.8)	36.6±26.1 (0.6)	35.8±26.8 (0.8)
	R5	X	X	X	X	35.6±26.7 (0.6)	35.2±27.3 (0.2)

Réintroduction de la fonction d'échanges souterrains

Les tests effectués à partir des structures M69 ou M72 ont souligné l'importance de cette fonction pour le modèle et la forte interaction qu'elle a avec la fonction de routage du modèle. A partir des meilleurs modèles dérivés de la structure M72 et celles ayant deux réservoirs en parallèle (M156 et M211), nous avons testé différentes formulations des échanges souterrains et différents coefficients de report des échanges sur le nouveau réservoir. Lorsqu'on réintroduit cette fonction, le modèle progresse légèrement, mais le gain de performance n'est toujours pas suffisant pour justifier l'ajout d'un réservoir supplémentaire. Par ailleurs, quand on reporte cet échange souterrain sur le réservoir exponentiel (M156), le coefficient (multiplicatif) de report qui donne les meilleurs résultats est très grand, ce qui indique que le réservoir exponentiel joue encore un rôle de perte dans le modèle. Ce résultat semble confirmer le rejet de ce type de réservoir par le modèle.

Utilisation d'un coefficient de partage non-stationnaire

La comparaison des modèles réalisée au Chapitre 5 a mis en évidence les bonnes performances des modèles WAGENIGEN et MORDOR10. Ces deux modèles utilisent dans leur structure une fonction qui permet d'établir le partage entre deux types d'écoulements dans le modèle. Les deux fonctions utilisées sont dépendantes de l'état (taux de remplissage) d'un des réservoirs du modèle, afin d'adapter le partage entre une branche 'rapide' et une branche 'lente' en fonction des conditions hydrologiques passées.

Nous avons tenté d'introduire ce type de fonction dans la structure M69. Pour cela, nous avons indexé la valeur du coefficient de partage PART1 sur le taux de remplissage du réservoir de production ou de routage (Tableau 6-3).

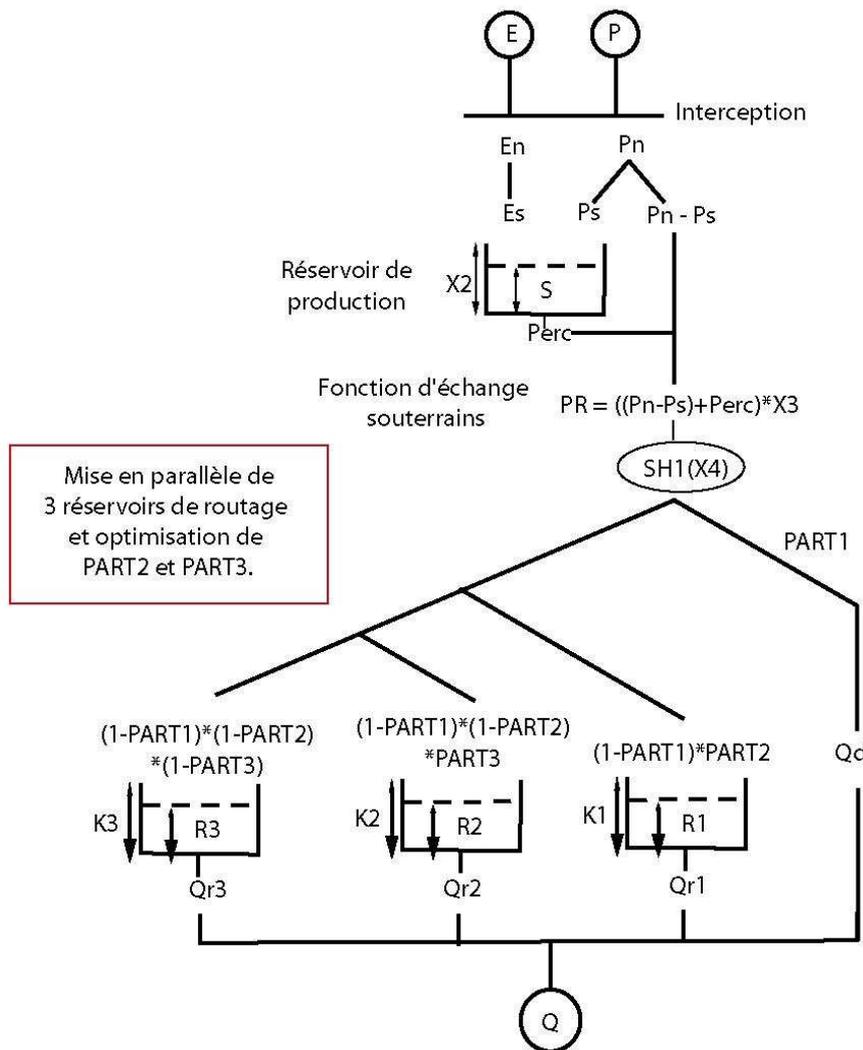
Tableau 6-3 : Différents types de fonctions de partage non-stationnaires testées

Test	Type de fonction de partage
T72 / M319	$PART1 = 1 - (TX(R).X_5)$
T73 / M320	$PART1 = 1 - (TX(S).X_5)$
T74 / M321	$PART1 = 0.7 + 0.3(TX(R).X_5)$
T75 / M322	$PART1 = 1 - 0.3(TX(R).X_5)$
TX(R) : taux de remplissage du réservoir de routage TX(S) : taux de remplissage du réservoir de production X ₅ : paramètre optimisé entre 0 et 1	

Ces fonctions, qui laissaient espérer une meilleure souplesse du modèle, ne semblent pas très intéressantes pour le modèle. En effet, soit les performances restent inchangées, soit le gain de performances est minime (M319). Par ailleurs, l'écart type de la distribution du paramètre X_5 est très grand (ce paramètre est très variable, pour un même bassin), ce qui prouve la faible stabilité de cette fonction.

6.3.4 Complexification de la fonction de routage : mise en parallèle de trois réservoirs aux capacités fixes, mais aux partages optimisés

La tentative précédente de complexification de la structure de GR4J n'a pas donné des gains de performance significatifs. Les modèles testés sont à peine arrivés à être légèrement meilleurs que la structure M69. La deuxième tentative que nous avons effectuée a été de mettre en parallèle trois réservoirs aux capacités fixées, mais aux partages optimisés.



Mise en parallèle de 3 réservoirs de routage et optimisation de PART2 et PART3.

Tableau 6-4 : Structure du modèle sur lequel est testée la mise en parallèle de trois réservoirs de routage aux capacités fixées ($K1, K2, K3$), mais aux partages optimisés ($PART2, PART3$).

Chapitre 6 : Approche empirique de mise au point d'un modèle pluie-débit au pas de temps horaire

L'idée initiale de cette structure est de fixer la valeur de K1, K2, K3 par une loi géométrique (par exemple : $K2=10*K1$, $K3=100*K1$). L'objectif est d'avoir des réservoirs qui répondent rapidement (échelle de l'événement), lentement (échelle saisonnière), très lentement (échelle annuelle). Pour cela nous avons testé de nombreuses combinaisons avec différents types de réservoirs, différents types de lois entre les capacités des réservoirs, etc.

Tableau 6-5 : Meilleurs critères C_{2M} obtenus lorsqu'on met en parallèle trois réservoirs de routage aux capacités fixées (K1, K2, K3), mais aux partages optimisés (PART2, PART3).

Raison de la suite géométrique : 10 M72 35.7±26.3		Type de réservoir de vidange				
		L	2	3	4	5
K1 : capacité initiale	1	25.7±21.7	34.0±26.6	35.0±26.2	35.1±26.4	35.1±26.4
	5	30.9±23.6	M256 36.2±26.4	35.8±26.7	35.0±26.7	34.5±26.8
	10	31.9±23.6	36.1±26.6	34.9±27.0	34.2±26.1	33.3±26.3
	25	32.3±23.6	34.8±26.7	32.5±26.7	30.2±27.6	28.5±28.4
	50	32.6±23.4	33.0±26.1	27.6±28.0	23.6±28.2	20.7±29.0
	optimisé	29.6±23.1	M243 36.4±26.4	35.7±27.2	35.5±27.4	35.3±27.2

Par rapport à la structure initiale (M72), ce type de structure apporte un léger gain. Néanmoins, ce gain n'est pas suffisant pour justifier l'ajout d'un paramètre optimisé.

6.3.5 Complexification de la fonction de routage : modifications de l'hydrogramme unitaire

Les deux tentatives précédentes de modifications de la fonction de routage n'ayant pas donné d'amélioration probante des performances, nous avons également essayé de modifier la forme de l'hydrogramme unitaire. Au cours de la vérification des paramètres fixes de GR4J, nous avons remarqué qu'une puissance de valeur 1.25 permettait d'améliorer les performances du modèle (+ 1 point de C_{2M}). Nous avons donc voulu tester d'autres fonctions, comme celles présentées à la Figure 6-7. L'idée est de tester la sensibilité du modèle à d'autres formes d'hydrogrammes unitaires, plus ou moins pointues, symétriques ou asymétriques, etc. Les fonctions utilisées peuvent être paramétrées par un ou deux paramètres. Il y en a généralement un pour le temps caractéristique de l'hydrogramme et un pour la forme.

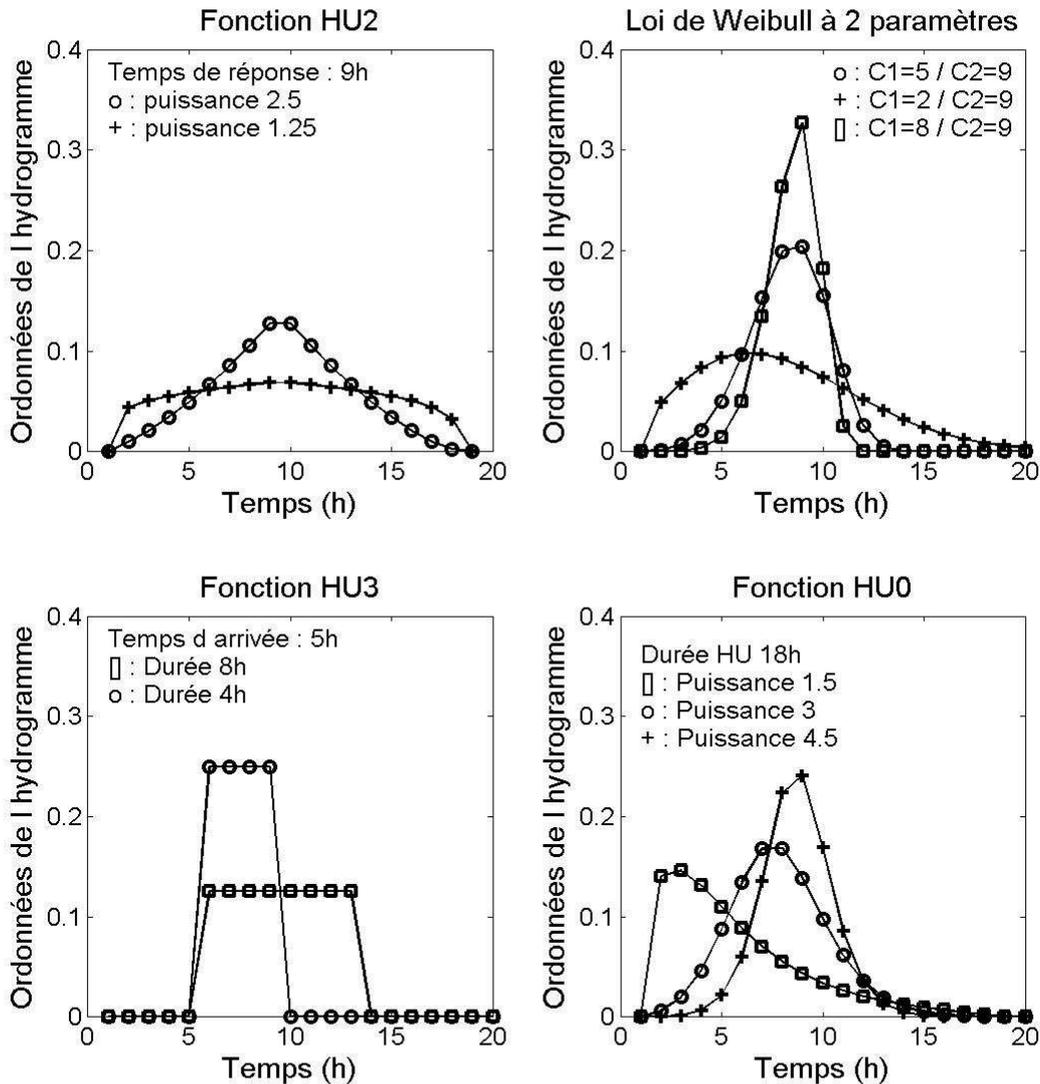


Figure 6-7: Exemples de formes d'hydrogramme unitaire qu'il est possible d'obtenir avec différentes fonctions (HU2, loi de Weibull, HU3, HU0).

Parmi tous les tests effectués, nous avons utilisé des fonctions mises au point pour les modèles GR (HU0, HU2, HU3) et une loi de Weibull (à deux ou un paramètres, utilisée par le modèle MORDOR10). En fait, les premiers tests ont mis en évidence l'intérêt d'un hydrogramme de forme écrasée (HU2, puissance 1.25). Le problème de cette forme d'hydrogramme est qu'elle ne permet pas d'avoir une montée progressive et un retard, avec quelques pas de temps où le bassin versant ne réagit pas (ordonnées nulles). Pour ces raisons, nous avons testé l'hydrogramme HU3. Cette fonction est une fonction créneau caractérisée par deux paramètres : le temps où l'hydrogramme est nul (temps d'arrivée) et la durée de la fonction créneau. Nous avons testé cet hydrogramme à deux paramètres et différentes relations entre le temps d'arrivée et la durée du créneau (généralement le temps d'arrivée est une fraction fixe de la durée du créneau). Ces tests n'ont mis en évidence aucun progrès significatif.

Nous avons également testé une fonction de Weibull à deux paramètres et à un paramètre. De la même manière, ces tests n'ont mis en évidence aucun progrès significatif, malgré la flexibilité de cette fonction. Enfin, Nous avons également testé la fonction HU0 à deux paramètres et à un paramètre. Cette fonction a été utilisée par *Tangara* [2005], parce qu'elle donnait de bons résultats dans un contexte de prévision des crues au pas de temps journalier. Néanmoins, de la même manière, ces tests n'ont mis en évidence aucun progrès significatif, malgré la flexibilité de cette fonction.

Enfin, nous avons testé d'autres modifications de la fonction HU2. L'objectif était d'essayer de prendre en compte le temps de réponse de l'hydrogramme pour déterminer sa forme, un effet de saisonnalité dans le fonctionnement du bassin ou de faire évoluer certains paramètres en fonction de certains états du système. Certains auteurs [*Aubert et al.*, 2003; *Anctil et al.*, 2004b; *Oudin et al.*, 2005a] ont d'ailleurs déjà montré l'intérêt d'utiliser le taux de remplissage d'un des réservoirs d'un modèle, pour améliorer les simulations de modèles hydrologiques. Par exemple, nous avons testé :

- La puissance dépend du temps de base de l'hydrogramme unitaire,
- La puissance de l'hydrogramme unitaire dépend du taux de remplissage du réservoir de production,
- Le temps de réponse de l'hydrogramme unitaire dépend du taux de remplissage du réservoir de production.

Ces idées, *a priori* intéressantes, n'ont malheureusement pas donné de progrès des performances des structures de modèles testés. Les modèles testés ont été capables de s'adapter à la nouvelle formulation de l'hydrogramme unitaire et conserver ainsi leur niveau de performance.

6.3.6 Complexification de la fonction de production : tentatives de prise en compte de l'intensité de la pluie et autres tests

Contre toute attente, les tentatives précédentes de complexification de la fonction de routage n'ont pas donné d'amélioration significative des performances des modèles. Le modèle le plus efficace et robuste est le modèle M69, dont la structure est très proche de celle de GR4J et même légèrement plus simple.

Introduction d'un réservoir d'interception (Figure 6-8)

Nous avons commencé par tester l'ajout d'un réservoir d'interception, en tête de modèle (comme dans la structure de TOPMO8). L'objectif de ce réservoir est d'intercepter une partie de la pluie par l'ETP. Le traitement de l'ETP étant complexe

au pas de temps horaire (l'ETP est toujours inférieure à la pluie, à partir du moment où il pleut), nous avons voulu tester ce réservoir, avec de faibles valeurs de capacités pour qu'il soit comparable avec celles de l'ETP.

Nous avons testé une capacité de réservoir optimisée tout d'abord et différentes valeurs fixées (entre 0.25 et 200 mm). Les résultats indiquent que cette fonction ne semble pas nécessaire pour le modèle. En effet, il n'y a pas de progrès significatif des performances (léger progrès lorsque la capacité du réservoir d'interception est de 0.25 mm).

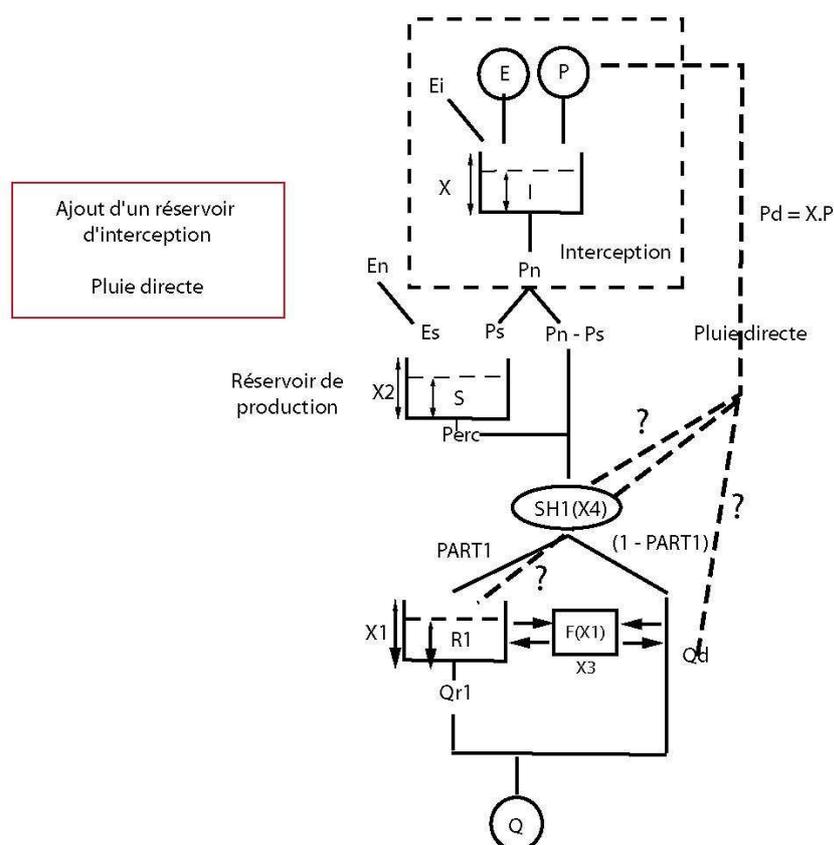


Figure 6-8 : Tentatives de modifications de la fonction de production du modèle M69. Ajout d'un réservoir d'interception et interception d'une partie de la pluie pour l'introduire directement dans la fonction de routage.

Détermination d'une pluie directe (Figure 6-8)

Nous avons testé l'introduction d'une pluie directe, au niveau de la fonction de routage. L'idée est de tester l'hypothèse qu'une partie du bassin répond directement à la pluie (surfaces imperméables ?), de manière indépendante de l'état d'humidité du bassin versant (taux de remplissage du réservoir de production).

Pour cela, nous avons intercepté une partie de la pluie, avant qu'elle ne rentre dans le réservoir de production, que nous avons réintroduite ensuite dans différentes parties de la fonction de routage du modèle. Nous avons testé différents taux

d'interception de la pluie (taux optimisé, ou fixé à 1%, 5%, 10% ou 15% de la pluie). Nous avons testé également différents points d'insertion de cette pluie directe (réservoir de routage, branche directe, hydrogramme unitaire). Ces modifications ne donnent aucun progrès significatif. Lorsqu'on optimise ce taux de pluie directe, pour plus de 75% des bassins versants, il est inférieur à 5% de la pluie. Néanmoins, pour 10% des bassins versants, le taux de pluie directe est supérieur à 50% de la pluie. Ces valeurs peu probables ou peu réalistes de pluie directe, indiquent soit un rejet de la fonction de production utilisée par le modèle, soit un mauvais fonctionnement du modèle (très mauvaises performances des modèles sur ces bassins versants) et/ou des erreurs dans les données.

Utilisation d'autres fonctions de production (Figure 6-9)

Les tentatives précédentes de complexification de la fonction de routage et les essais sur la fonction de production n'ayant rien donné jusque là, nous avons essayé d'introduire une fonction prenant en compte les intensités de la pluie et une limitation à leur infiltration.

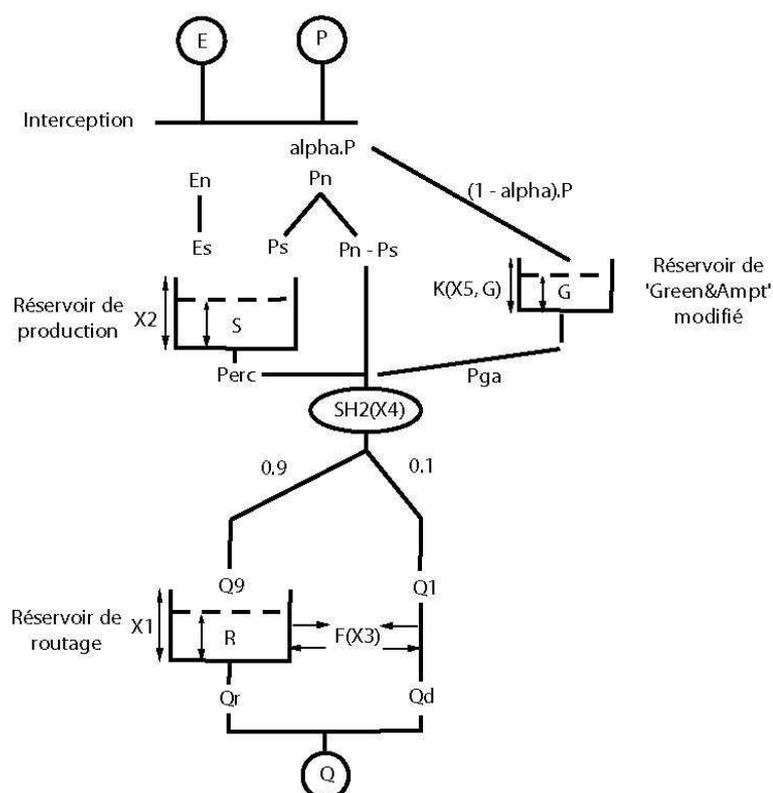


Figure 6-9 : Tentatives de modifications de la fonction de production du modèle M69. Ajout d'un réservoir de Green & Ampt modifié, dont la capacité est variable et permet de prendre en compte une limitation des capacités d'infiltration.

Pour cela, nous nous sommes inspirés d'une fonction de type 'Green et Ampt', que nous avons modifiée. Nous avons ainsi essayé d'introduire un réservoir en parallèle du réservoir de production, dont la capacité dépend de l'intensité de la pluie reçue.

Nous avons testé différentes variantes (taux d'interception, lois de fonctionnement du réservoir, introduction des écoulements), qui n'ont pas donné non plus de progrès significatifs. Néanmoins, les tests sur cette partie du modèle ont été les moins nombreux et les moins exhaustifs. La suite des développements au pas de temps horaire devra passer par un approfondissement des tests sur la fonction de production.

6.3.7 Synthèse des tests sur la structure de GR4H

Dans cette partie nous avons essayé de synthétiser aussi clairement que possible la multitude de tests effectués. A partir des résultats obtenus par *Perrin* [2000] et *Mouelhi* [2003], nous sommes partis de la structure du modèle GR4J, que nous avons confrontée aux données horaires d'un échantillon de 313 bassins versants. Nous avons remis en cause la structure de GR4J, (1) en testant les paramètres fixes de ce modèle, (2) en complexifiant la fonction de routage du modèle et (3) en prenant en compte les intensités de pluie dans la fonction de production.

La première phase de tests a permis d'améliorer fortement les performances au pas de temps horaire, en simplifiant légèrement la structure du modèle. Nous avons modifié la valeur du coefficient de la fonction de percolation (qui dépend du pas de temps des données), la puissance de l'hydrogramme unitaire et nous proposons de n'utiliser plus qu'un hydrogramme unitaire à deux branches. Cette structure, à quatre paramètres, représente un gain de performances de plus de 5 points de critère $C_{2M}(Q)$, nous l'avons appelée GR4H (M69).

La deuxième et la troisième phase de tests nous ont amenés à tester différentes complexifications de la fonction de routage et de la fonction de production du modèle. Contre toute attente, ces résultats indiquent avant tout qu'il ne semble pas nécessaire d'introduire trois branches en parallèle dans le modèle. Dans le meilleur des cas, les gains obtenus sont modestes et ne semblent pas justifier l'ajout de cette branche supplémentaire. Néanmoins, si nous devons conserver un modèle à cinq paramètres, nous conserverons la structure M124 (réservoir exponentiel) que nous appelée GR5H. Cette structure soulève quelques interrogations quant aux interactions entre la fonction d'échanges souterrains et le réservoir exponentiel, qui devront être examinées par la suite. Enfin, les différentes tentatives de prise en compte de l'intensité de la pluie dans la fonction de production n'ont pas apporté de résultats satisfaisants. Cette fonction devra également être examinée plus en détail par la suite.

6.4 Tentatives d'amélioration d'une structure au pas de temps horaire par une approche empirique de simplification progressive

L'objet de cette partie est la présentation des résultats concernant l'analyse des structures de TOPMO8 et MORDOR10. Ces deux modèles ont été choisis à titre d'exemple, pour illustrer ce que peut apporter la simplification empirique et progressive de la structure d'un modèle, comportant de nombreux paramètres libres. Nous le rappelons, l'intérêt de réduire la complexité d'un modèle est multiple :

- Améliorer la compréhension de son fonctionnement,
- Supprimer les fonctions qui ne sont pas nécessaires au bon fonctionnement du modèle,
- Améliorer la procédure de calage du modèle en réduisant les interactions entre paramètres et en supprimant les paramètres peu, mal ou non définis,
- Améliorer la robustesse du modèle.

A partir de ces deux structures, nous avons testé différentes hypothèses de réduction de la complexité du modèle. Nous avons validé une simplification seulement si elle ne provoquait pas de réduction des performances. Nous le verrons, contre toute attente, cette démarche a même permis d'améliorer les performances de ces modèles.

6.4.1 Modèle MORDOR10 (Figure 6-10)

Nous avons souhaité réduire la complexité de ce modèle, car au cours de la comparaison des performances des différentes structures de modèles (Chapitre 5), il est apparu comme un des meilleurs modèles. Malheureusement, ce modèle est très long et difficile à optimiser (plus de 80 h., pour les 313 bassins versants). De plus, la comparaison des vecteurs de paramètres optimisés par différentes méthodes de calage (Chapitre 4) a mis en évidence le caractère mal défini (aléatoire ?) de certains paramètres. Comme le montrent la Figure 4-8 et la Figure 4-11, certains des paramètres de MORDOR10 semblent moins bien définis et il en va peut être de même pour les fonctions qui leur sont associées (X_2 , X_3 , X_5 , X_6 , X_9). Par ailleurs, l'étude des différents vecteurs de paramètres optimisés a également mis en évidence l'existence de nombreux optima locaux ou de régions de vecteurs équifinaux dans la surface de réponse de MORDOR10, alors que cela n'existe pas pour le modèle parcimonieux GR4H.

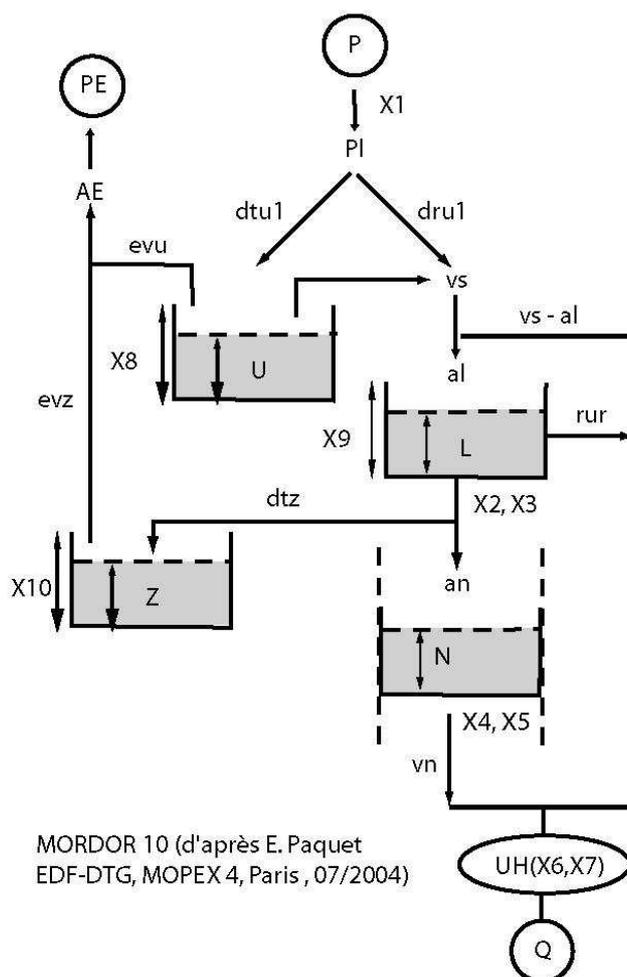


Figure 6-10 : Schéma de la structure initiale du modèle MORDOR10.

A partir de ce constat, nous avons cherché à réduire le niveau de complexité de MORDOR10, à mettre en évidence les paramètres libres auquel MORDOR10 n'est pas sensible et éventuellement à simplifier les fonctions associées, tout ça en ne réduisant pas les performances de la structure initiale.

Le modèle MORDOR10 possède 10 paramètres libres :

- 1 coefficient correcteur sur la pluie : X_1 ;
- 3 capacités de réservoirs (U, L, Z) : X_8 , X_9 , et X_{10} ;
- 2 paramètres pour fixer la vidange du réservoir L : X_2 et X_3 ;
- 2 paramètres pour fixer la vidange du réservoir N : X_4 et X_5 ;
- 2 paramètres pour fixer la forme de l'hydrogramme unitaire : X_6 et X_7 ;

Par ailleurs, ce modèle peut générer des sorties d'eau seulement sous forme de débit et d'ETR. Ainsi, le seul moyen pour ce modèle d'adapter son bilan est à l'ajustement de l'ETR et du coefficient correcteur sur les pluies (X_1). Il semble intéressant de noter que certains cumuls annuels d'ETR, simulés par MORDOR10 et

GR4H ont été comparés pour certains bassins versants. Entre les deux modèles, un facteur 2 a été trouvé, GR4H ayant la souplesse de générer d'autres flux sortant, sous forme d'échanges souterrains.

Suite aux résultats du Chapitre 0, nous avons commencé par explorer des modifications sur les paramètres X_2 , X_3 , X_5 , X_6 et X_9 et leurs fonctions associées. Il s'agit en fait d'une analyse de sensibilité de la structure du modèle MORDOR10. A première vue, il semble que certaines fonctions soient paramétrées de manière maladroite (avec finalement trop de degrés de liberté) :

- la vidange du réservoir N, avec deux paramètres (capacité du réservoir et puissance de la loi de vidange), dont les valeurs peuvent être assez fortement interdépendantes. Généralement, afin d'assurer une meilleure robustesse à un modèle, il est bien plus efficace de fixer la valeur de la puissance, la valeur de la capacité pouvant s'adapter ;
- La loi de Weibull à deux paramètres pour l'hydrogramme unitaire. Cette loi possède deux paramètres (facteur de forme et temps caractéristique) pour pouvoir adopter une forme très souple, alors que les modèles hydrologiques semblent relativement peu sensibles à la forme de l'hydrogramme unitaire ;
- la vidange du réservoir L, avec deux paramètres dont les valeurs peuvent être assez fortement interdépendantes. De la même manière, il peut être plus intéressant de fixer la valeur d'un des deux paramètres, la valeur de l'autre pouvant s'adapter ;

Par ailleurs, nous avons tenté de remettre en cause chaque paramètre libre de la structure de MORDOR10. Pour cela, nous avons fixé un à un les paramètres libres à la valeur de la médiane de leurs distributions.

Ainsi, nous avons suivi une démarche empirique, en fonction des résultats successifs apportés au cours des différents tests. Ces tests ont confirmé certains de nos *a priori* :

- La structure de MORDOR10 n'est pas sensible à la forme d'hydrogramme unitaire utilisée (trop sophistiquée). Ses performances ne sont pas modifiées, lorsque la valeur du facteur de forme de l'hydrogramme unitaire (X_6) est fixée ou lorsque l'hydrogramme est remplacé par la fonction HU2 de GR4H ;
- Il n'est pas nécessaire d'optimiser la valeur de la puissance de la loi de vidange du réservoir N (X_5). Lorsque sa valeur est fixée, ses performances sont même légèrement améliorées (Test M6) ;

Cette première série de simplifications assez faciles à tester a déjà permis d'obtenir un modèle à 8 paramètres, aux performances aussi bonnes que celles de la

structure initiale (X_6 fixé à 3 ou utilisation de HU2 ; X_5 fixé à 3). Ces tests ont également mis en évidence l'extrême sensibilité du modèle à l'optimisation du paramètre X_1 , qui sert à corriger la pluie. Si on supprime ce paramètre, le modèle perd plus d'une dizaine de points, ce qui le rend particulièrement mauvais. Cette extrême sensibilité laisse d'ailleurs penser qu'il serait intéressant d'investir ce paramètre dans une fonction d'échanges souterrains, similaire à celle de la famille des modèles GR. Ces tests ont également mis en évidence une forte sensibilité à l'optimisation de la capacité du réservoir L (X_9).

Les tests suivants ont porté sur la simplification de la structure pour obtenir une structure à 7 paramètres. Nous avons fixé certains paramètres à la valeur médiane de leurs distributions ($X_2, X_3, X_8, X_9, X_{10}$). Lorsqu'on fixe les valeurs de X_2, X_3, X_8, X_{10} les performances sont pas ou peu affectées. Seul le paramètre X_9 a besoin d'être optimisé. Le meilleur résultat est obtenu lorsqu'on fixe le paramètre X_2 . Ce résultat permet de remettre légèrement en cause le fonctionnement de ce réservoir, puisqu'il déterminait un partage entre différents écoulements (rur et an, Figure 6-10). Lorsqu'on fixe la valeur de X_2 , le caractère non-stationnaire de ce partage est peut être remis en cause. Nous avons alors testé différents types de lois de partage (non stationnaire mais indexée sur X_3 , ou partage fixe). Le réservoir L est toujours caractérisé par sa capacité X_9 et une constante de vidange X_3 ; en revanche, le partage entre un écoulement direct (rur) et un écoulement vers le réservoir N (an) n'est plus déterminé par une autre constante de vidange sur le réservoir L, mais par un partage fixe. Le meilleur partage est de 80% vers le réservoir N (an) et 20% vers l'écoulement direct (rur). Ainsi, cette structure est la meilleure à sept paramètres (M8).

Enfin la dernière série de tests a porté sur la simplification de la structure pour obtenir une structure à 6 paramètres. Nous avons remis en cause la capacité du réservoir Z, dont la position est périphérique dans le fonctionnement du modèle. Lorsque sa capacité est fixée à 90 mm, les performances ne sont pas affectées et le modèle ne possède plus que six paramètres libres. Nous avons effectué quelques tests pour réduire la structure du modèle à cinq paramètres. Les performances des versions obtenues sont légèrement diminuées et montrent notamment l'extrême rigidité de la fonction de production, qui doit être capable d'adapter son bilan. Par exemple, lorsqu'on fixe la capacité du réservoir de production (U), les performances sur les bassins français ne sont pas affectées, alors qu'elles chutent fortement sur les bassins arides américains et australiens. Cette différenciation très nette des résultats sur cette version à 5 paramètres rappelle l'intérêt de larges échantillons de bassins versants, aux caractéristiques hydro-climatiques variées.

- Constante de vidange du réservoir L : X_2 ;
- Constante de vidange du réservoir N : X_3 ;
- Temps de réponse de l'hydrogramme unitaire HU2 : X_4 ;
- Capacité du réservoir U : X_5 ;
- Capacité du réservoir L : X_6 .

L'ensemble des résultats des tests effectués est résumé à la Figure 6-12. Cette figure montre l'évolution des performances des versions testées ayant de 10 à 5 paramètres optimisables. Les dernières versions testées n'ont plus que 5 paramètres et ont leurs performances toujours proches de celles de la version initiale.

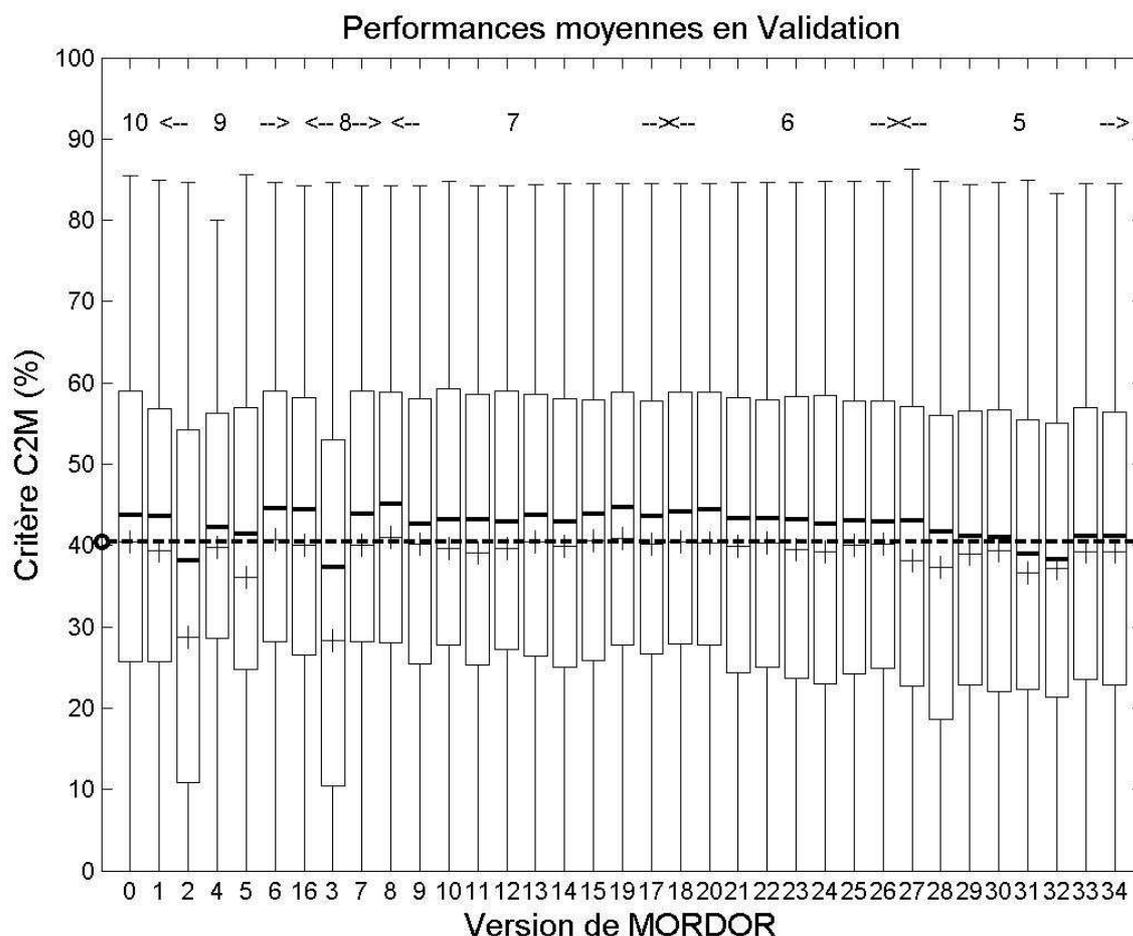


Figure 6-12 : Synthèse des performances moyennes de l'ensemble des structures de MORDOR testées, ayant de 10 à 5 paramètres libres. Le nombre de paramètres est indiqué sur la figure, le pointille indique le niveau de performance initial.

Pour l'instant, les tests sur la structure des modèles de la famille MORDOR s'arrêtent là, mais les bons résultats de la nouvelle version à 6 paramètres révèlent l'intérêt d'étudier plus en profondeur les détails du fonctionnement de ce modèle. Notamment, le paramètre X_1 , coefficient correcteur sur les pluies, devrait pouvoir être investi de manière plus judicieuse dans une fonction d'échanges souterrains, comme on peut la trouver dans les modèles de la famille GR. En effet, ce type de fonction

est généralement plus simple et plus efficace qu'un simple coefficient multiplicateur sur les pluies.

L'intérêt d'une telle approche de réduction empirique de la complexité d'un modèle est clairement mis en évidence par la Figure 6-13. Comme au 0, la Figure 6-13a) montre la distribution cumulée des distances entre les vecteurs de paramètres du modèle MOR6, trouvés par deux méthodes d'optimisation différentes. Cette figure montre que la proportion de vecteurs, optimisés selon deux méthodes différentes et situés dans une même région de l'espace, augmente fortement lorsqu'on passe de MORDOR10 à MOR6. Cette proportion passe de 15% à 55% des vecteurs optimisés et se rapproche de celle de GR4H (80%). Celle de IHAC6, modèle qui a aussi six paramètres est entre 50% et 70% selon les méthodes d'optimisation comparées.

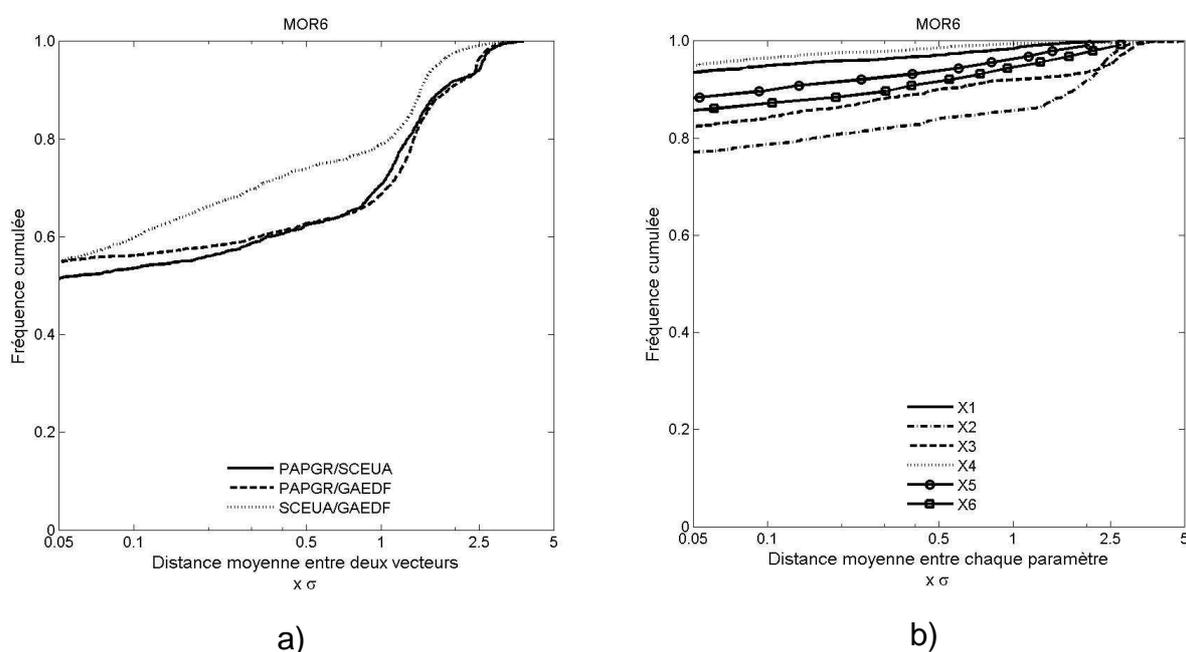


Figure 6-13 : Modèle MOR6 a) Distribution cumulée des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différentes, b) Distribution cumulée des distances entre chaque paramètre trouvé par deux méthodes d'optimisation différentes (PAP-GR et SCE-UA).

La Figure 6-13 b) montre la distribution cumulée des distances entre les valeurs de chaque paramètre du modèle MOR6, trouvées par deux méthodes d'optimisation différentes (PAP-GR/SCE-UA). Cette figure montre que les paramètres du modèle MOR6 sont très bien définis et que dans plus de 80% des cas la distance entre les deux valeurs d'un paramètre est inférieure à 5% de l'écart-type de la distribution de ce paramètre, obtenue sur l'ensemble des bassins. Cette figure montre d'ailleurs que les paramètres les mieux définis sont X_1 et X_4 (X_1 est le coefficient correcteur de la pluie et permet de régler le bilan en eau du modèle et X_4 est le temps de réponse du l'hydrogramme unitaire). Pour le modèle GR4H, le paramètre du temps de réponse

de l'hydrogramme unitaire était également le paramètre le mieux défini. Ces niveaux de définition des six paramètres de MOR6 sont du même ordre et même un peu meilleurs que ceux des six paramètres les mieux définis de MORDOR10.

Ainsi, la Figure 6-13 a permis de mettre en évidence l'intérêt de la simplification de la complexité de la structure de MOR6, en facilitant le calage de ce modèle et permettant une bien meilleure définition des vecteurs de paramètres calés selon différentes méthodes. Nous pensons que ces caractéristiques permettent, à niveaux de performance égaux, d'accroître la robustesse des vecteurs de paramètres optimisés et la confiance qu'on peut avoir dans les capacités de ce modèle à reproduire les débits d'un bassin versant.

6.4.2 Modèle TOPMO8

Nous avons également choisi de travailler sur la structure du modèle TOPMO8 parce que, dans les travaux de Perrin [2000] et Oudin [2004], ce modèle donne d'excellents résultats au pas de temps journalier. Comme nous l'avons remarqué, au Chapitres 4 et 5, ce modèle donne d'assez mauvais résultats au pas de temps horaire, sur cette base de données. Ces faibles performances peuvent provenir de nombreux facteurs, entre autres :

- Effet du pas de temps des données,
- Effet de la différence de base de données et notamment, de la plus faible taille des bassins versants,
- Effet de la durée de calage,

L'hypothèse de l'effet du pas de temps des données et de la petite taille des bassins versants semble assez peu plausible, puisque TOPMODEL (ici, fortement modifié par rapport à la version originale) était justement conçu pour la modélisation de la relation pluie-débit sur de petits bassins versants, à des pas de temps fins.

Ainsi, ces résultats sont intrigants et les hypothèses formulées pour essayer de comprendre l'origine des mauvaises performances de cette structure, ont motivé les tests sur la structure de TOPMO8, qui est particulièrement peu robuste et difficile à caler, dans sa version à 8 paramètres.

Le modèle TOPMO8 possède 8 paramètres libres :

- 2 capacités de réservoirs (R, S) : X_1, X_3 ;
- 1 paramètre de vidange du réservoir exponentiel (T) : X_2 ;
- 2 paramètres pour la fonction logistique approximant la distribution de l'indice topographique : X_5, X_6 ;
- 2 paramètres pour la fonction d'ETR : X_7, X_8 ;
- 1 paramètre de délai pur : X_4 ;

L'ensemble des résultats des tests effectués est présenté à la Figure 6-14. Nous avons suivi la même démarche de simplification progressive de la structure que pour MORDOR, en commençant par la fonction logistique représentant la distribution de l'indice topographique et la fonction d'ETR. Ces deux fonctions sont toutes deux caractérisées par deux paramètres libres, qui peuvent être interdépendants et poser donc des problèmes d'identification (calage) et de robustesse.

Les premiers tests ont porté sur l'identification d'une structure à 7 paramètres. Tout d'abord, nous avons remis en cause les fonctions associées à la distribution de l'indice topographique et à la détermination de l'ETR. Les fonctions testées ont donné de bons résultats au pas de temps journalier dans le cadre de la thèse de [Perrin, 2000; Oudin, 2004]. Néanmoins, dans notre cas, ces simplifications donnent d'assez mauvais résultats, surtout ceux portant sur la distribution de l'indice topographique. Nous avons alors effectué d'autres tests, en gardant les fonctions initiales telles quelles, mais en fixant un des deux paramètres de ces fonctions. Nous avons donc fait des tests en fixant successivement X_5 , X_6 , X_7 et X_8 , à la valeur de la médiane de leurs distributions. Le test sur X_5 entraîne une légère perte de performance. Les tests portant sur X_6 , X_7 et X_8 permettent un gain significatif de performances, de trois à quatre points de C_{2M} en moyenne.

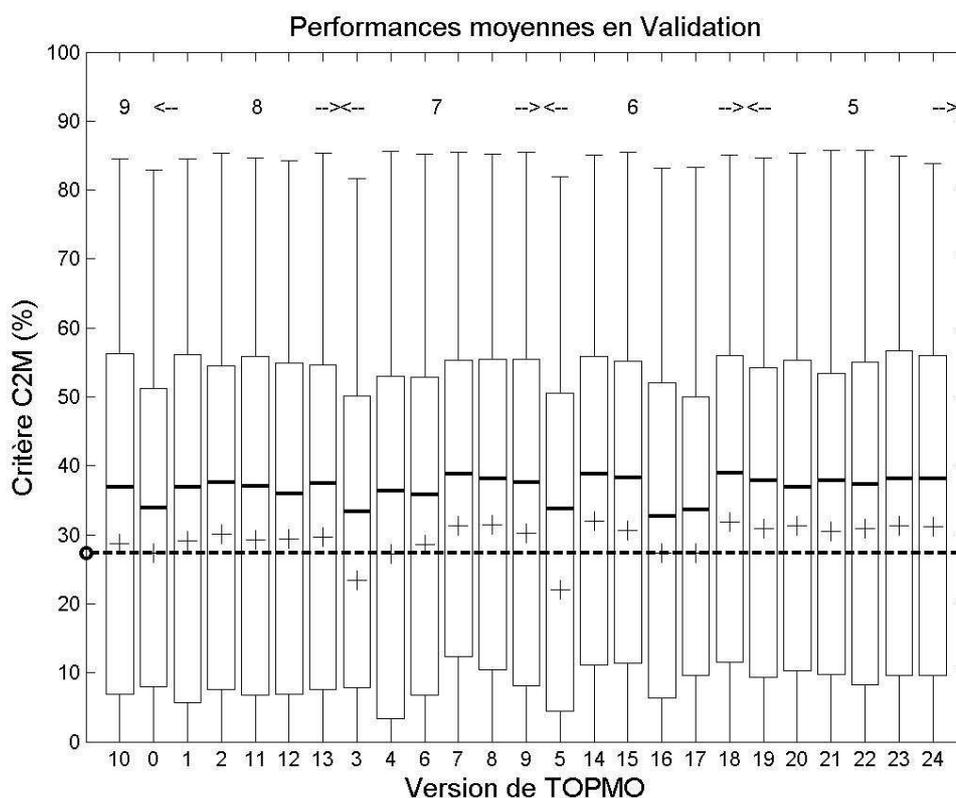


Figure 6-14 : Synthèse des performances moyennes de l'ensemble des structures de TOPMO testées, ayant de 9 à 5 paramètres. Le nombre de paramètres est indiqué sur la figure. La version 0 est la version initiale, le pointille indique le niveau de performance initial.

Ensuite, nous avons poursuivi les tests sur des versions à six paramètres. Nous avons essayé de combiner les progrès obtenus avec les simplifications précédentes. Les meilleurs résultats sont obtenus lorsque les paramètres X_6 et X_7 sont fixés à leur valeurs médianes ($XV_6 = -4.1$, $X_7 = 2.2$). Par ailleurs, la fonction de délai a été remplacée par un hydrogramme unitaire (HU2), qui permet également de faire progresser les performances moyennes du modèle. Ainsi, cette version à six paramètres est significativement meilleure que la version initiale à huit paramètres. Le progrès des performances en validation est de 4.7 points.

Nous avons enfin testé quelques versions à cinq paramètres, en fixant la valeur de la capacité du réservoir d'interception. Cette hypothèse a été testée suite aux résultats sur GR4H, qui ont montré que le modèle est plus robuste lorsqu'on fixe ce paramètre et aux tests de *Perrin* [2000] qui fixe la capacité de ce réservoir à 3 mm (pas de temps journalier). Nous avons cherché de manière empirique la meilleure capacité de ce réservoir, au pas de temps horaire. Nous avons testé différentes valeurs entre 1 mm et 9 mm. Les meilleurs résultats sont obtenus pour des valeurs de 5 mm à 7 mm, mais cela entraîne une légère perte de performances de 0.7 points de C_{2M} .

Ainsi, cette simplification empirique de TOPMO8 a permis d'améliorer les performances de ce modèle, tout en réduisant son niveau de complexité. Le niveau de performance de ce modèle reste encore inférieur à celui de nombreuses autres structures de modèles, mais devient un peu plus acceptable pour la version à six paramètres (TOP6).

6.4.3 Synthèse des tests sur les structures de MORDOR et TOPMO8

Les exemples des modèles MORDOR et TOPMO8 montrent clairement l'intérêt de la démarche empirique d'analyse de sensibilité structurelle qui permet de réduire le niveau de complexité de ces modèles tout en améliorant leurs performances. De cette manière, le test exhaustif de ces structures progressivement modifiées et simplifiées permet de réaliser une analyse de sensibilité des différentes composantes de ces modèles. La simplification progressive permet de réduire les idées *a priori* sur la structure des modèles et de ne garder que les fonctions nécessaires aux modèles pour obtenir de bons niveaux de performances. Néanmoins, l'inconvénient de cette démarche par rapport à celle de la complexification progressive d'une structure de modèle, est que l'on part d'une structure arbitraire postulée *a priori*. Par opposition, l'approche descendante part d'une structure simple, justifiée par ses performances, obtenues à un pas de temps plus grand.

6.5 Quelques éléments d'analyse des performances de GR4H et MORDOR6

6.5.1 Influence de la fonction objectif de calage

Au CHAPITRE 5 , nous avons mis en évidence l'influence de la fonction objectif (FO) utilisée pour le calage des paramètres du modèle ($C_{2M}(Q)$ ou $C_{2M}(RQ)$) sur les performances en validation du modèle. La FO $C_{2M}(Q)$ est plus adaptée pour représenter les crues, alors que la fonction $C_{2M}(RQ)$ est plus adaptée pour représenter la gamme des bas et moyens débits [pour une synthèse sur le sujet, voir *Oudin et al.*, 2006]. La fonction objectif utilisée, conduisant à l'adaptation du modèle pour certaines gammes de débits, doit également avoir une influence sur les vecteurs de paramètres optimisés.

Les Figure 6-15 et Figure 6-16 montrent l'influence de la FO sur les performances en validation et les vecteurs de paramètres de GR4H et MORDOR6. Les Figure 6-15 a) et Figure 6-16 a) montrent que lorsque que le critère $C_{2M}(Q)$ est utilisé en validation, les performances des modèle semblent insensibles à la FO utilisée en calage, alors que lorsque le critère $C_{2M}(RQ)$ est utilisé en validation, la FO $C_{2M}(RQ)$ donne des performances sensiblement meilleures. Par ailleurs, il semble que la FO $C_{2M}(RQ)$ permettent d'identifier des vecteurs plus robustes, dont les performances en validation sont plus rarement très mauvaises, comparativement à celles obtenues avec la FO $C_{2M}(Q)$. Les Figure 6-15 b) et Figure 6-16 b) montrent la comparaison des vecteurs de paramètres obtenus avec les deux FO. Ces figures indiquent une certaines dispersion des paramètres autour de la bissectrice, avec une plus ou moins grande dispersion selon les paramètres. Pour le modèle GR4H, la dispersion de chaque paramètre est relativement uniforme et du même ordre de celle obtenue lorsqu'on utilise deux méthodes d'optimisation différentes (CHAPITRE 4). Pour le modèle MORDOR6, la dispersion des paramètres est moins uniforme et certains paramètres semblent s'adapter un peu plus à la FO (X_1, X_2, X_3). C'est notamment le cas de X_1 , coefficient correcteur de la pluie, qui a besoin systématiquement de surestimer légèrement la pluie, lorsque que la FO $C_{2M}(Q)$ est utilisée.

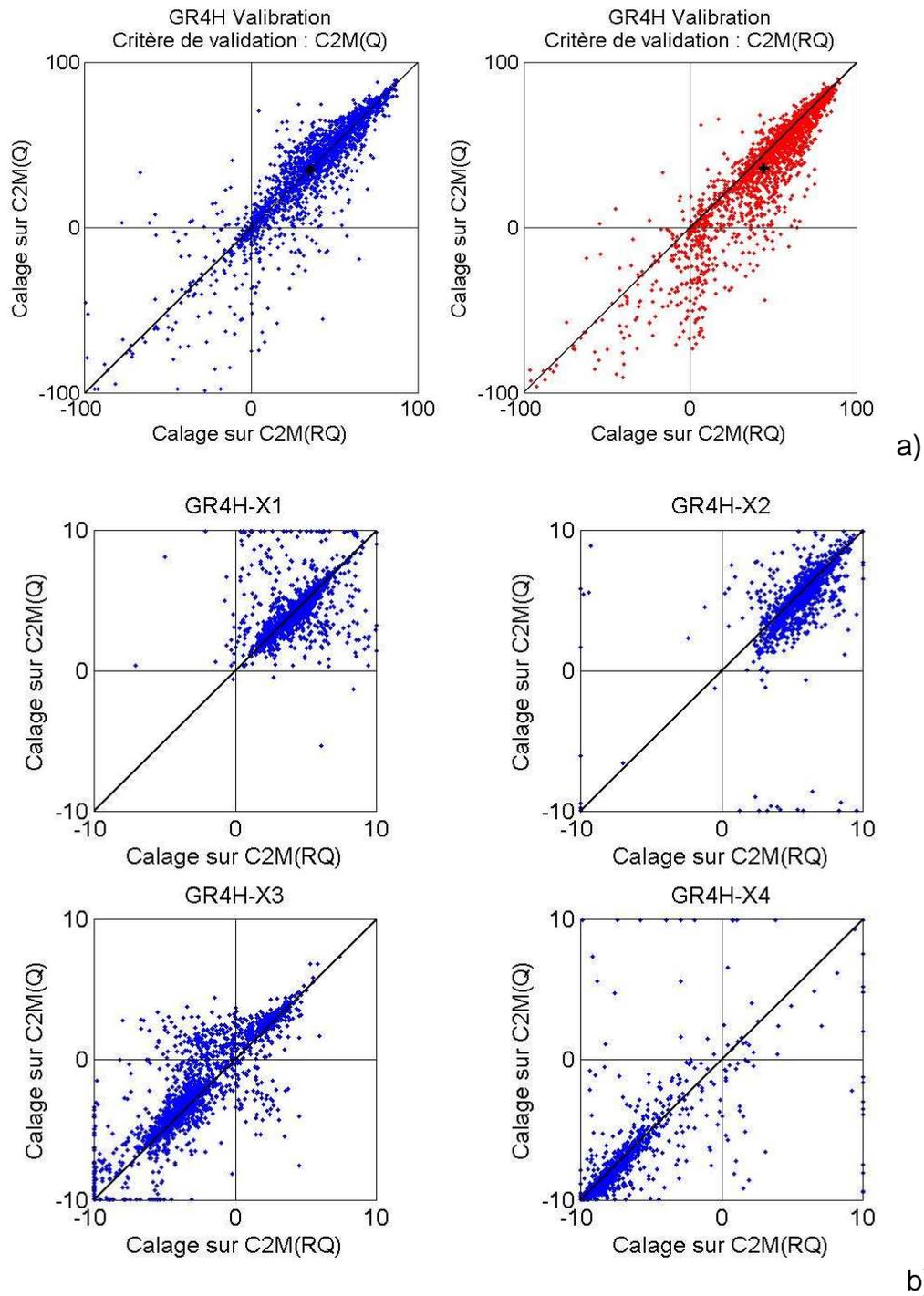


Figure 6-15 : Influence de la fonction objectif utilisée en calage sur (a) les performances en validation et (b) les vecteurs de paramètres de GR4H [$r(C_{2M}(Q)) = 0.89$; $r(C_{2M}(RQ)) = 0.90$; $r(X_1) = 0.48$; $r(X_2) = 0.59$; $r(X_3) = 0.85$; $r(X_4) = 0.83$; r : coefficient de corrélation].

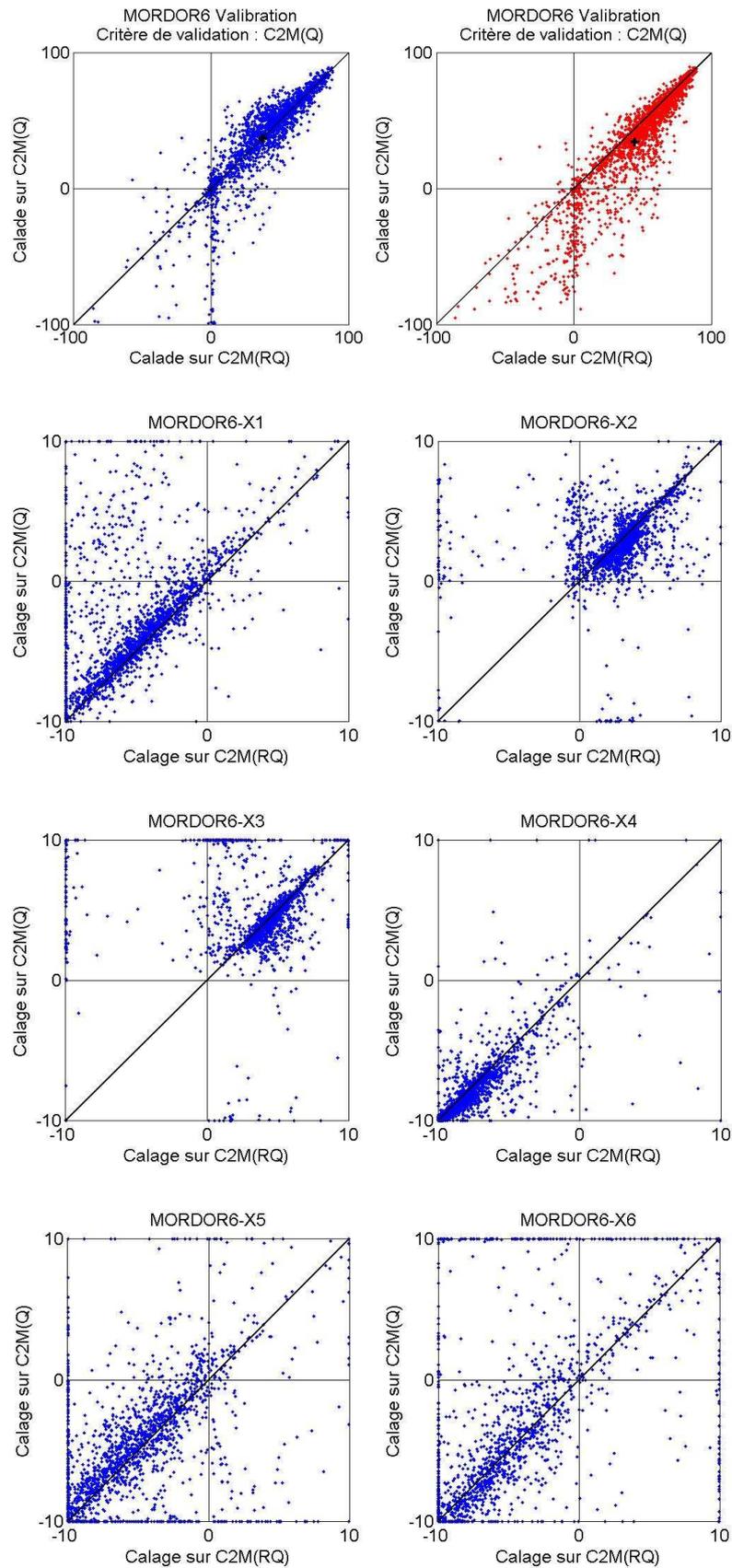


Figure 6-16 : Influence de la fonction objectif utilisée en calage sur (a) les performances en validation et (b) les vecteurs de paramètres de MORDOR6 [$r(C_{2M}(Q)) = 0.88$; $r(C_{2M}(RQ)) = 0.89$; $r(X_1) = 0.58$; $r(X_2) = 0.31$; $r(X_3) = 0.45$; $r(X_4) = 0.87$; $r(X_5) = 0.53$; $r(X_6) = 0.62$;]

6.5.2 Hydrogrammes

Des cette partie nous présentons les simulations obtenues par GR4H et MORDOR6 sur le bassin versant de la Glueyre à Tisonèche, situé en Haute-Ardèche. Un seul bassin versant a été retenu a titre d'exemple, pour illustrer les analyses que l'on peut faire à partir des simulations d'un modèle.

La Glueyre à Tisonèche, 71 km² (V4145210) :

La Glueyre est une magnifique rivière ardéchoise située dans la région du Vivarais et des Boutières. Elle prend sa source à l'est du Mont Mezenc et s'écoule de l'ouest vers l'est, jusqu'à St Sauveur de Montagut dans la vallée de l'Eyrieux. Le bassin versant de la Glueyre à Tisonèche a une superficie de 71 km², son altitude moyenne est de 850 m et elle est comprise entre 411 m et 1343 m (Le champ de Mars). Le bassin versant se développe en montagne, avec des vallées encaissées et des canyons taillés dans le substrat géologique cristallin et quelques zones de roches métamorphiques et volcaniques. Le bassin versant de la Glueyre est caractérisé par sa forêt de châtaigniers qui a marqué l'identité de cette région, les terrasses de cultures (en patois, *les chambas*), ses moulins et moulinages pour la soie et quelques sources d'eaux minérales.

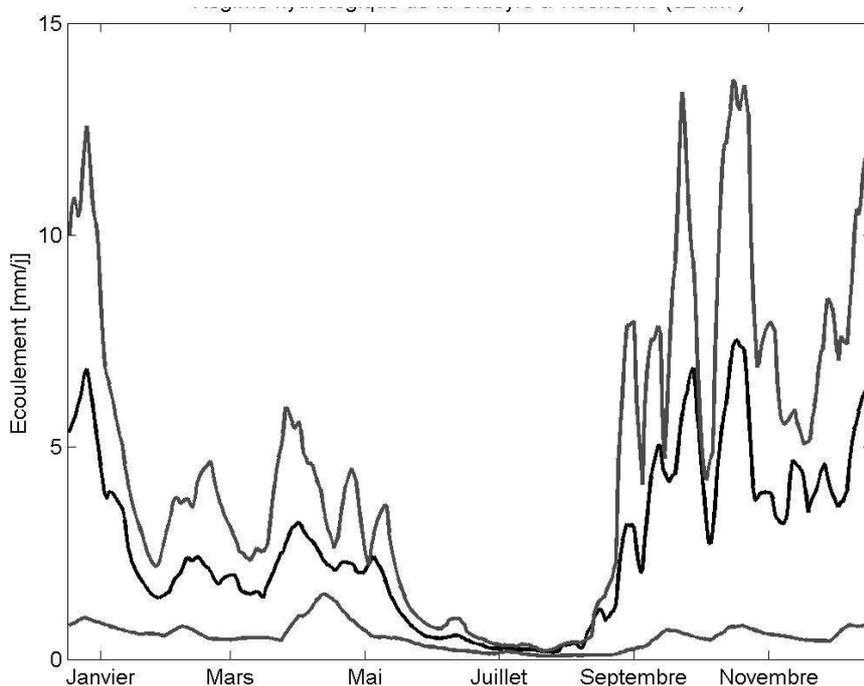


Figure 6-17 : Régime hydrologique de la Glueyre à Tisonèche (71 km²). La courbe en noir indique le débit médian interannuel de chaque jour, les courbes en gris indiquent les quantiles 0.1 et 0.9 de chaque jour. Chronique : 1993 – 2002.

Le régime hydrologique de la Glueyre (Figure 6-17) est pluvio-nival méditerranéen avec un étiage très marqué en été, de fortes crues durant des épisodes cévenols,

lorsque les fortes pluies qui viennent lécher le Mézenc en début d'automne débordent au nord et à l'est. Le régime de la Glueyre possède également une certaine composante nivale, du fait de son altitude, du climat rigoureux et de la burle* qui sévissent sur le plateau ardéchois.

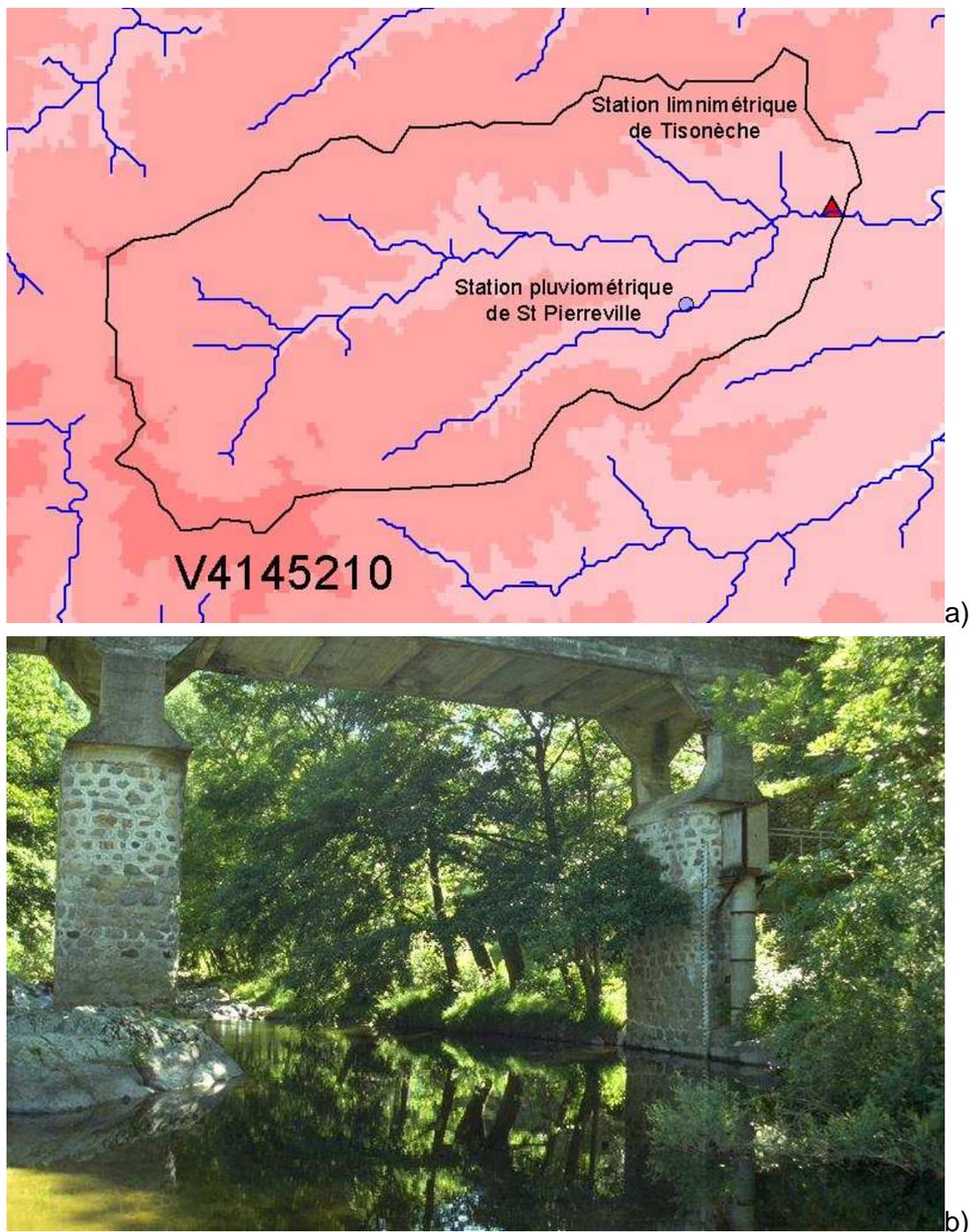


Figure 6-18 : La Glueyre à Tisonèche (71 km²), a) limites du bassin versant et réseau hydrographique, b) station de jaugeage.

* le climat et la burle sont également à l'origine du caractère bien forgé des Ardéchois...

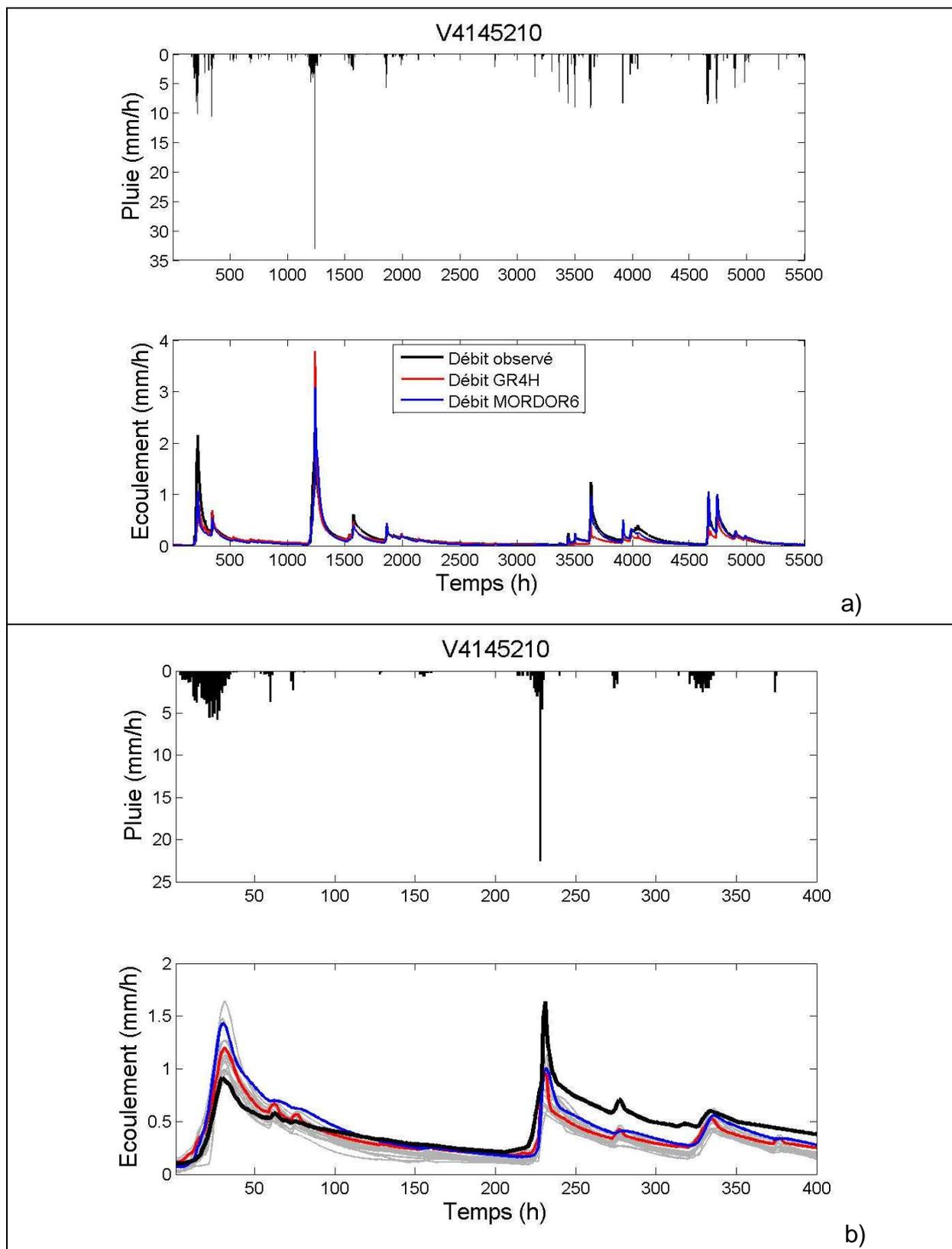


Figure 6-19 : La Glueyre à Tisonèche, a) Hydrogramme et simulations obtenues avec les modèles GR4H (rouge) et MORDOR6 (bleu), pour une durée de 8 mois, b) Hydrogramme, simulations obtenues avec les modèles GR4H (rouge) et MORDOR6 (bleu) et simulations obtenues par les 18 autres modèles (gris) pour deux crues successives [critère de calage $C_{2M}(Q)$].

La Figure 6-19 présente une chronique de pluie, de débits observés et les simulations des différents modèles. Sur ce bassin versant, les modèles sont assez bons, leurs critères de performances vont de 55% à 77% ($C_{2M}(Q)$). La Figure 6-19 a) présente des hydrogrammes et les simulations obtenues par les modèles GR4H (rouge ; $C_{2M}(Q) = 65\%$), MORDOR6 (bleu ; $C_{2M}(Q) = 72\%$). Cette chronique est caractérisée par quatre crues notables. Globalement, les deux modèles simulent correctement la première et la quatrième crue (synchronisation des crues et valeur du pic). Par contre, ils surestiment assez la deuxième crue (d'un facteur 2). Les modèles semblent réagir un peu fortement aux fortes intensités. De plus, sur ce bassin versant, nous ne disposons que d'un pluviomètre et il est possible que la forte intensité pluvieuse [30 mm/h] ne soit pas suffisamment représentative de cet événement. Enfin, pour la troisième crue, les deux modèles réagissent différemment. Il semble que GR4H n'arrive pas à 'repartir' après une période d'étiage prolongée, alors que MORDOR6 réagit assez bien et simule correctement cette crue.

La Figure 6-19 b) présente des hydrogrammes et les simulations obtenues par les modèles GR4H (rouge), MORDOR6 (bleu) et les 18 autres modèles (gris). L'ensemble des modèles a un comportement relativement similaire. Ils simulent à peu près correctement la première crue et loupent la deuxième. Ces deux crues sont intéressantes car elles ont été causées par des averses, dont les cumuls sont comparables [de 60 à 70 mm], mais dont les intensités maximales sont très différentes [5 et 23 mm/h]. Cette figure montre différents types de réactions du bassin versant et des modèles.

De manière plus globale il est possible d'appréhender autrement les résultats des modèles (Figure 6-20). Par exemple, la Figure 6-20 a) présente la comparaison des débits observés et simulés (QQ Plot), la Figure 6-20 b) présente la comparaison du biais multiplicatif entre débits observés et simulés en fonction du débit observé, la Figure 6-20 c) présente la distribution des débits observés et simulés. Ces figures permettent de qualifier les simulations d'un modèle par rapport au comportement réel d'un bassin versant. Elles permettent notamment de vérifier si il existe un biais systématique du modèle, ou bien si le modèle semble plus adapté pour certaines gammes de débit. Par exemple, la Figure 6-20 a) montre que les erreurs des deux modèles sont relativement bien dispersées autour de la bissectrice et indépendante du débit simulé, à part un léger biais systématique pour les très bas débits. Cette figure montre également que les erreurs du modèle GR4H semblent être plus dispersées que celles de MORDOR6. La Figure 6-20 b) indique le biais multiplicatif des débits simulés par rapport aux débits observés.

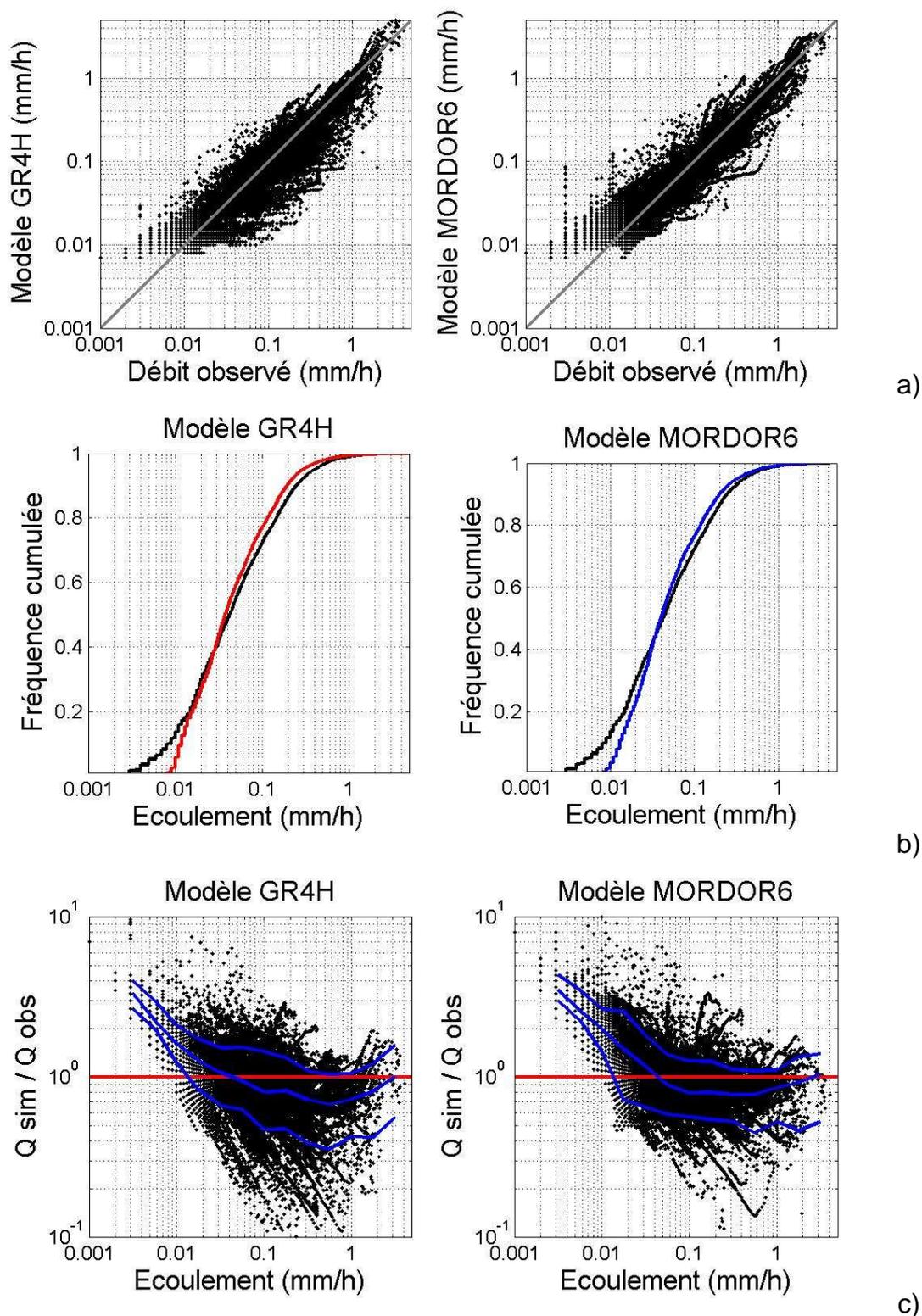


Figure 6-20 : Analyse des performances des modèles : a) comparaison des débits observés et simulés (diagramme logarithmique) ; b) comparaison de la distribution des débits observés et simulés (noir : observé ; couleur : simulé). c) biais multiplicatif entre débit simulé et observé en fonction du débit observé, lignes bleues : quantiles 10%, 50% et 90% [critère de calage $C_{2M}(Q)$] ;

Cette figure montre que les deux modèles surestiment les bas débits et sous estiment les haut débits. Les quantiles 10% et 90% des biais par classes de débit montrent que le modèle GR4H a une gamme de biais plus importante que celle du modèle MORDOR6. La Figure 6-20 c) montre que les modèles ont la même distribution des débits simulés. Ils sur-estiment (en fréquence) tous les deux les bas-débits, sous-estiment légèrement la gamme moyenne de débit et estiment à peu près bien les forts débits.

L'influence de la FO utilisée pour optimiser les paramètres du modèle est mise en évidence en comparant les Figure 6-20 et Figure 6-21. Par rapport à la Figure 6-20, la Figure 6-21 montre que lorsque la FO $C_{2M}(RQ)$ est utilisée les biais sur les bas débits sont plus limitées, mais les biais sur les moyens et hauts débits sont supérieurs à ceux obtenus avec la FO $C_{2M}(Q)$.

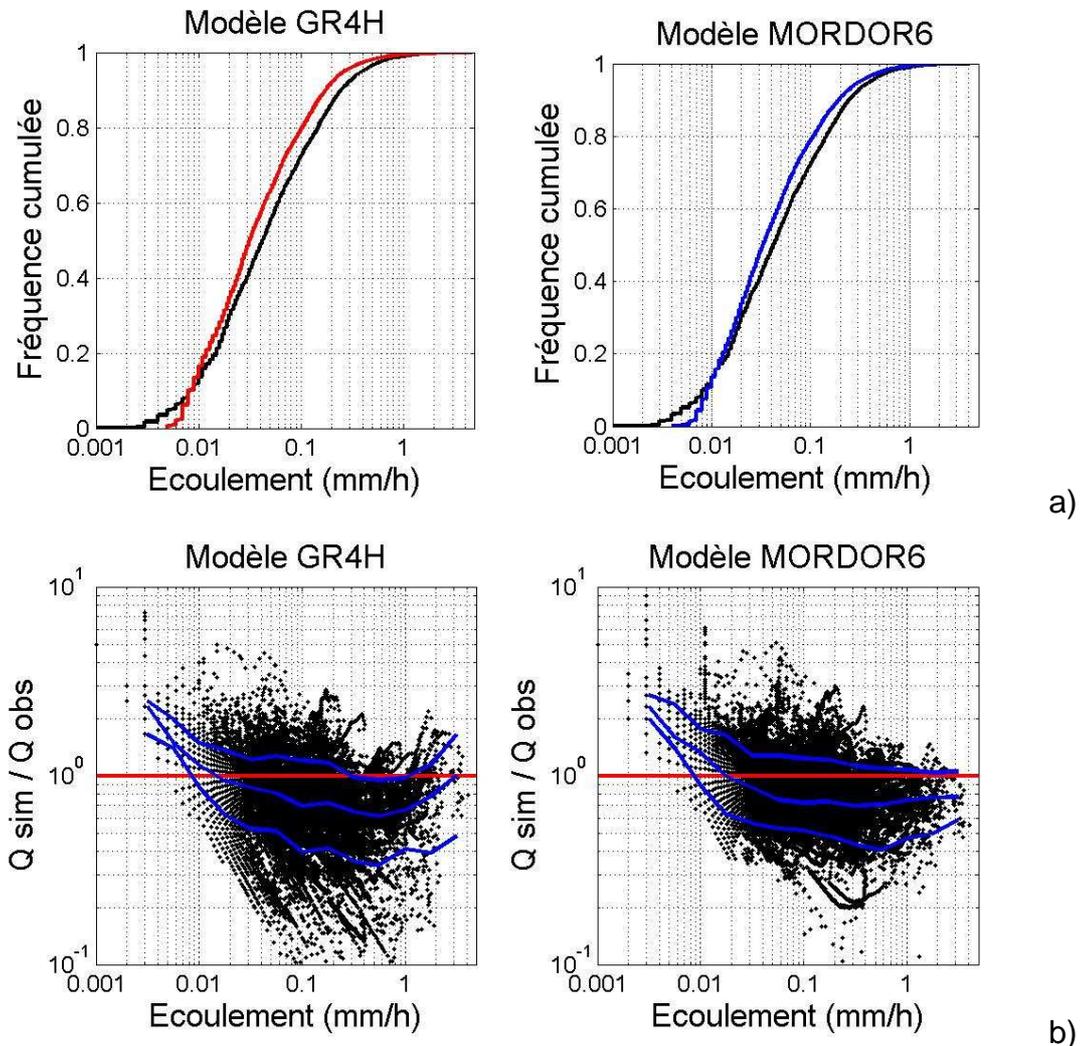


Figure 6-21 : Analyse des performances des modèles : b) comparaison de la distribution des débits observés et simulés (noir : observé ; couleur : simulé). c) biais multiplicatif entre débit simulé et observé en fonction du débit observé, lignes bleues : quantiles 10%, 50% et 90% [critère de calage $C_{2M}(RQ)$];

6.6 Synthèse

Ce chapitre synthétise les modifications de plusieurs centaines de structures de modèles testées au pas de temps horaire. La recherche de meilleures structures au pas de temps horaire a permis d'obtenir des améliorations significatives des performances de certaines structures de modèles. La modification empirique des différentes structures de modèles s'est faite selon deux voix : ascendante et descendante.

Nous avons commencé par une approche ascendante, de complexification progressive de la structure du modèle GR4J [Perrin, 2000], en nous basant sur les directions que semblaient indiquer les résultats de Mouelhi [2003]. Cette approche nous a permis de faire progresser fortement les performances de la structure initiale au pas de temps horaire. Dans l'état actuel des choses, nous proposons d'utiliser au pas de temps horaire les structures GR4H et GR5H. Au travers d'une approche comparative, ces structures se sont révélées particulièrement performantes et robustes. Leurs niveaux de performance sont du même ordre que celui des structures de MOR6 et MORDOR10.

Nous avons également testé l'approche opposée, celle de la simplification progressive d'une structure de modèle complexe. Cette approche s'apparente également à une étude de sensibilité structurelle à grande échelle. Elle nous a permis de proposer quelques modifications de ces structures. Nous nous sommes basés pour cela sur le modèle MORDOR10 à dix paramètres et le modèle TOPMO8 à huit paramètres. MORDOR10 est caractérisé par de très bonnes performances, mais ses paramètres sont mal définis et il est très difficile à caler. Le modèle TOPMO8 est caractérisé par de mauvaises performances au pas de temps horaire, alors qu'il donne de bons résultats au pas de temps journalier [Perrin, 2000; Oudin, 2004]. Avec ces deux modèles, nous avons donc testé de manière empirique des modifications progressives de sa structure, en diminuant le nombre de paramètres à optimiser. Ces tests ont permis de proposer une structure de MORDOR à six paramètres, MOR6. Cette structure est au moins aussi efficace que la structure initiale, son calage est beaucoup plus rapide et ses paramètres bien mieux définis. Ces résultats sont exemplaires car ils illustrent la réduction drastique des problèmes d'équifinalité de la structure de MORDOR. De la même manière, nous avons proposé une structure de TOPMO à six paramètres, dont les performances ont même progressé !

Pour conclure, ces tests ont montré la difficulté de cet exercice et comme le dit Bergström [1991] : « *The point of diminishing returns (no model improvement) was*

reached surprisingly soon when increasing model complexity », ce qui n'empêche pas que cette procédure de développement apporte de nombreux résultats très intéressants sur le fonctionnement des modèles hydrologiques et les interactions entre définition des vecteurs de paramètres et complexité de la structure d'un modèle.

TROISIEME PARTIE :

**QUELS AUTRES DEVELOPPEMENTS PEUT
NOUS APPORTER LE PAS DE TEMPS
HORAIRE ?**

Chapitre 7

**« Une mathématique bleue,
Sur cette mer jamais étale
D'où me remonte peu à peu
Cette mémoire des étoiles »**
Léo Ferré, *La mémoire et la mer.*

CHAPITRE 7 Amélioration d'une méthode d'optimisation locale : prise en compte de l'information apportée par la distribution *a priori* des paramètres

7.1 Introduction

Les résultats de nos recherches sur la comparaison de différentes stratégies d'optimisation a mis en évidence le bon niveau d'efficacité de la méthode PAP-GR que nous utilisons. Mais cette méthode d'optimisation peut encore rester piégée sur un optimum local (Figure 4-4 et Figure 4-5). De plus, les tests au pas de temps horaire nécessitant de longs temps de calcul, il serait intéressant d'améliorer l'efficience (rapidité) de cette méthode.

Nos tests ont montré que le risque d'être piégé par un optimum local est d'autant plus faible que le point de départ de cet algorithme est correctement choisi, c'est-à-dire dans la région de l'optimum. Inversement, lorsque le point de départ est mal choisi (mal situé, trop distant par rapport à l'optimum ou aléatoirement choisi) l'efficacité et l'efficience de cet algorithme de recherche locale sont généralement fortement diminuées. Ainsi, pour trouver un bon optimum, il faut donc identifier au préalable un bon point de départ de l'algorithme de recherche. Comment faire ?

Les tests effectués précédemment ont révélé que, globalement, quel que soit le modèle choisi ou le nombre de paramètres à optimiser, la médiane de la distribution de chaque paramètre est un bon point de départ de recherche, pour cet algorithme. On peut se demander alors comment il serait possible d'améliorer notre algorithme de recherche de l'optimum. Souvent, les algorithmes de recherche locale utilisent des méthodes multi-départs pour tenter de mieux explorer l'espace des paramètres et diminuer les probabilités de rester piégé sur un optimum local. Ensuite, les différents vecteurs optimisés sont mis en compétition et celui dont la performance en calage est la meilleure est conservé. Néanmoins, il reste toujours le problème du choix judicieux du point de départ de l'algorithme d'optimisation.

Nous proposons d'explorer ici diverses manières de déterminer le plus judicieusement possible le point de départ de l'algorithme d'optimisation par recherche locale PAP-GR. Pour cela, nous allons tirer profit de la distribution *a posteriori* des paramètres d'un modèle, obtenue grâce au calage du modèle sur un grand échantillon de bassins versants. L'objectif de ce travail sur la méthode PAP-

GR est d'améliorer aussi bien son efficacité (meilleures performances en calage) que son efficacité (augmentation de la rapidité de l'algorithme de recherche). En revanche, les gains de performance attendus ne sont pas forcément très grands, mais l'intérêt de cette méthode est d'essayer d'éliminer les quelques cas où l'algorithme d'optimisation s'arrête sur un optimum local, du fait d'un mauvais point de départ.

7.2 Choix d'un point de départ pour une méthode d'optimisation par recherche locale

La méthode mise au point est fondée sur la génération d'un certain nombre de points de départs possibles, bien répartis dans l'espace des paramètres. Les performances de chaque point de départ possible sont évaluées et le point dont les performances sont les meilleures est conservé comme point de départ de la méthode locale d'optimisation. L'originalité de cette méthode est la manière dont sont générés les points de départs possibles, à l'aide des distributions des paramètres d'un modèle. Ainsi, pour utiliser cette méthode, il faut déjà disposer de la distribution des paramètres d'un modèle, c'est-à-dire qu'il faut déjà avoir optimisé un modèle sur un gros échantillon de bassins versants.

7.2.1 Exploitation de la distribution *a priori* des paramètres d'un modèle pluie-débit

En calant un modèle sur un grand échantillon de bassins versants, aux caractéristiques variées, il est possible de définir la distribution *a posteriori* des paramètres d'un modèle (Figure 7-1 a). Cette information est très précieuse, car elle permet de connaître le domaine de variation des paramètres d'un modèle. L'idée est d'utiliser cette information pour explorer ponctuellement l'espace probable des paramètres, avant d'effectuer une optimisation sur un nouveau bassin versant en utilisant l'algorithme de recherche locale.

La méthode d'optimisation initiale (PAP-GR), commence sa recherche à partir d'un vecteur de départ, dont chaque composante est la médiane de la distribution de chaque paramètre. Néanmoins, il est possible d'exploiter plus l'information contenue dans les distributions des paramètres d'un modèle, en ne résumant pas chaque distribution à la médiane, mais en prenant n quantiles pour la caractériser (Figure 7-1 b) à f); modèle GR4H).

Chapitre 7 : Amélioration d'une méthode d'optimisation locale: prise en compte de l'information apportée par la distribution a priori des paramètres

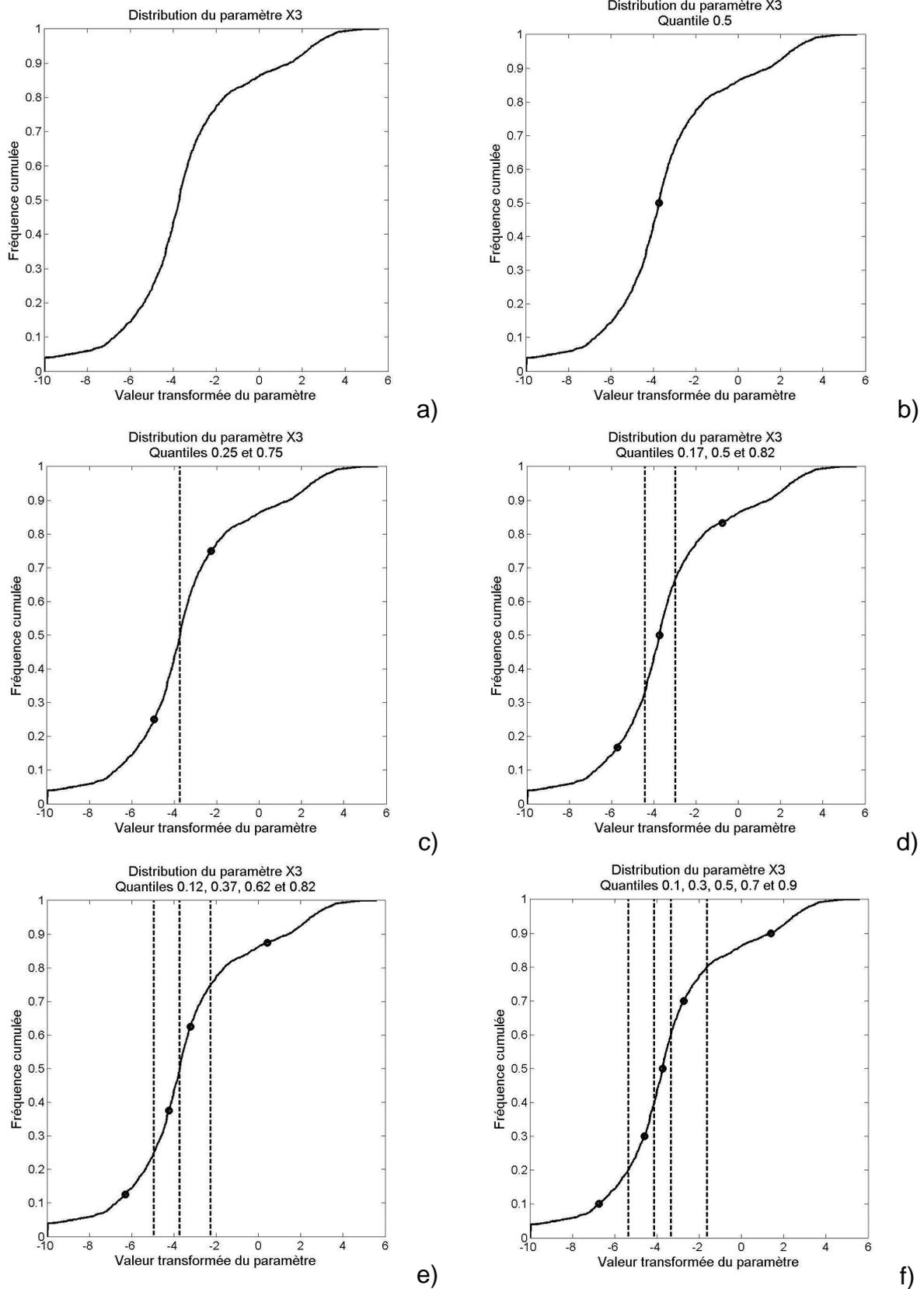


Figure 7-1 : Exemple de distribution d'un paramètre obtenue sur l'échantillon de 313 bassins versants (paramètre X_3 , modèle GR4H). a) distribution ; division de la distribution en 1 (b), 2 (c), 3 (d), 4 (e) et 5 (f) parties, avec les frontières et le point central de chaque partie.

L'idée est de diviser la distribution de chaque paramètre en n parties égales, dont on prend la partie centrale comme valeur représentative. Ainsi, pour les m paramètres d'un modèle, il est possible de générer les vecteurs qui sont la combinaison des n points définissant les m paramètres, ce qui représente n^m vecteurs possibles, au lieu de un vecteur lorsqu'on se limite à la valeur de la médiane de chaque distribution.

Chaque partie est donc caractérisée par sa frontière (valeur du quantile limitant deux parties consécutives) et sa valeur représentative (valeur du quantile central), dont les valeurs sont résumées au Tableau 7-1 :

- Valeurs des quantiles des n frontières :

Pour $j = 1, n-1$

$$Front_{xi} = \frac{j}{n} \quad [4.3]$$

- Valeurs des quantiles des n points centraux :

Pour $j = 1, n$

$$Centr_{xi} = \frac{2.(j-1)+1}{n} \quad [4.4]$$

Tableau 7-1 : Valeurs des quantiles des frontières et des points centraux des n parties d'une distribution, lorsqu'on utilise un à cinq points pour représenter une distribution.

Méthode PAP-GR n	Nombre de divisions (n)	Quantiles des frontières des n parties	Quantiles des valeurs représentatives des parties
PAP-GR $_1$	1	/	0.5
PAP-GR $_2$	2	0.5	0.25 / 0.75
PAP-GR $_3$	3	0.33 / 0.66	0.17 / 0.5 / 0.82
PAP-GR $_4$	4	0.25 / 0.5 / 0.75	0.12 / 0.37 / 0.62 / 0.87
PAP-GR $_5$	5	0.2 / 0.4 / 0.6 / 0.8	0.1 / 0.3 / 0.5 / 0.7 / 0.9

Dans le cadre de la méthode initiale, la médiane représente le cas où $n=1$, ce qui revient à prendre la valeur centrale de la distribution. Mais par exemple, si on souhaite diviser la distribution en trois parties égales ($n=3$), cela revient à retenir pour chaque paramètre une valeur plutôt faible, une valeur moyenne et une valeur plutôt forte pour résumer la distribution. Les trois parties de cette distribution ont pour frontières les valeurs associées aux quantiles 0.33 et 0.66. Par exemple, sur la Figure 7-1 d), les frontières entre ces parties sont représentées par les traits verticaux dont les coordonnées sont $X_3=-4.4$ et $X_3=-2.9$. Les points centraux de

chaque partie de cette division, résumant cette distribution, sont $X_3=-5.7$, $X_3=-3.7$ et $X_3=-0.7$. Ainsi, pour les quatre paramètres du modèle GR4H, la Figure 7-2 montre les distributions et les points centraux des trois parties utilisées pour synthétiser la distribution des quatre paramètres du modèle GR4H.

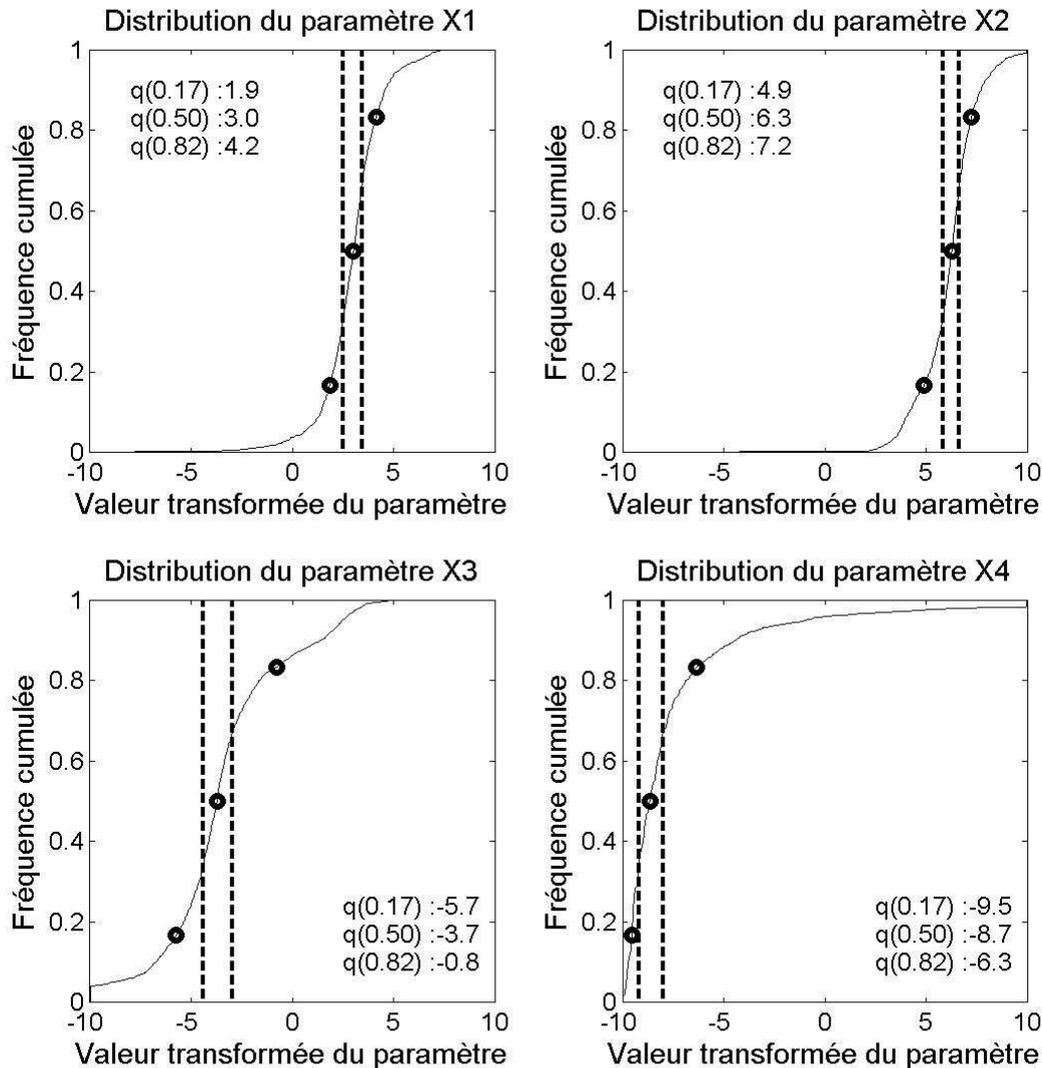


Figure 7-2 : Division des distributions des valeurs des paramètres de GR4H en trois parties et valeurs des trois points résumant chaque distribution.

Ainsi, avec quatre paramètres et trois valeurs pour chaque distribution, $81 (3^4)$ vecteurs différents peuvent être générés. Ces 81 vecteurs sont autant de points de départ possibles qui échantillonnent de manière parcimonieuse, mais exhaustive, la surface de réponse des paramètres. Par exemple, la Figure 7-3, représente en trois dimensions les 27 vecteurs possibles de l'espace de $\{X_1, X_2, X_3\}$ qu'il suffit de combiner avec les trois valeurs possibles pour X_4 , pour avoir les 81 vecteurs possibles.

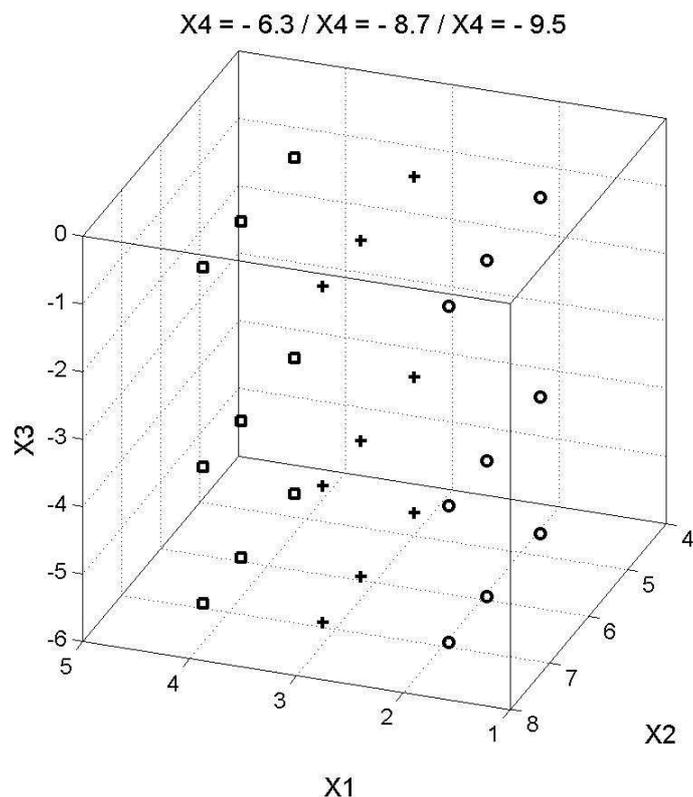


Figure 7-3: Représentation en trois dimensions du sous-échantillonnage de l'espace des paramètres $X_1X_2X_3$, par 27 vecteurs de paramètres obtenus à partir de la division des distributions de chaque paramètre en trois parties.

L'intérêt de ces n^m vecteurs est qu'ils sont répartis dans l'espace des paramètres en fonction des distributions de chaque paramètre et qu'ils échantillonnent de manière parcimonieuse la surface de réponse d'un modèle. L'intérêt de cette méthode est que, grâce à la connaissance des distributions de chaque paramètre, des vecteurs de paramètres peu probables ne sont pas utilisés inutilement comme point de départ possible. La faible efficacité des méthodes globales vient en partie de leur utilisation d'une loi de distribution uniforme des paramètres pour générer aléatoirement des vecteurs de départ. Par exemple, avec le modèle GR4H, la Figure 7-2 montre qu'il est inutile de générer des vecteurs ayant des valeurs négatives pour X_1 , X_2 ou positives pour X_4 .

7.2.2 Méthode de filtrage des points de départ possibles

Nous avons vu précédemment qu'il est possible de générer un grand nombre de vecteurs de paramètres, qui sont répartis sur la surface de réponse d'un modèle. Suivant le nombre de paramètres à optimiser (m), suivant le nombre de parties que l'on souhaite choisir pour résumer la distribution de chaque paramètre (n), il est possible de générer n^m vecteurs possibles. Ainsi, comme le montre le Tableau 7-2, le nombre de vecteurs possibles peut devenir très grand, dès que le modèle possède un grand nombre de paramètres. Or l'objectif de cette tentative d'amélioration de la

Chapitre 7 : Amélioration d'une méthode d'optimisation locale: prise en compte de l'information apportée par la distribution a priori des paramètres

méthode PAP-GR, est d'améliorer l'efficacité (trouver un meilleur vecteur optimum) et également l'efficience (diminuer le nombre d'itérations nécessaires pour optimiser les paramètres d'un modèle ; une itération représente un calcul de la fonction objectif sur la période de calage). Ainsi, on remarque que, même en divisant la distribution des paramètres en deux ou en trois parties, le nombre de vecteurs possibles est rapidement de l'ordre du nombre moyen de simulations nécessaires pour optimiser un modèle. Ainsi, tant que tous les points de départ sont conservés, cette méthode n'est pas intéressante pour optimiser les paramètres d'un modèle, puisqu'un nombre trop grand de vecteurs sont générés.

Tableau 7-2 : Nombre de vecteurs de paramètres possibles, en fonction du nombre de divisions de la distribution des paramètres et du nombre de paramètres à optimiser. Nombre moyen de simulations nécessaires pour optimiser un modèle sur une période.

n : nombre de divisions de chaque distribution	m : nombre de paramètres à optimiser			
	4	6	8	10
	n ^m : nombre de points de départ possibles			
2	16	64	256	1024
3	81	729	6561	≈ 60000
4	256	4096	≈ 65000	≈ 10 ⁶
5	625	≈ 15000	≈ 390000	≈ 9.10 ⁶
Nombre moyen d'itérations nécessaire pour optimiser les paramètres d'un modèle sur une période	GR4H ≈ 340	IHAC6 ≈ 940 MOR6 ≈ 710	TOPMO8 ≈ 2440	MORDOR10 ≈ 3860

Cette situation nous a amené à imaginer une méthode de filtrage des points de départs possible. Cette méthode de filtrage très simple est fondée sur l'analyse des vecteurs de paramètres d'un modèle rencontrés sur un large échantillon de bassins versants. L'idée est de comparer chaque point de départ possible avec les vecteurs déjà optimisés. Si un point de départ possible appartient à une région de l'espace dans laquelle est déjà situé un vecteur de paramètres optimisé, alors on conserve ce point de départ possible. En revanche, si le point de départ possible appartient à une région vers laquelle la méthode d'optimisation n'est jamais allée, ce point de départ possible est éliminé. Une région est une zone de l'espace des paramètres délimitée par les frontières des différents intervalles délimités pour chaque paramètre.

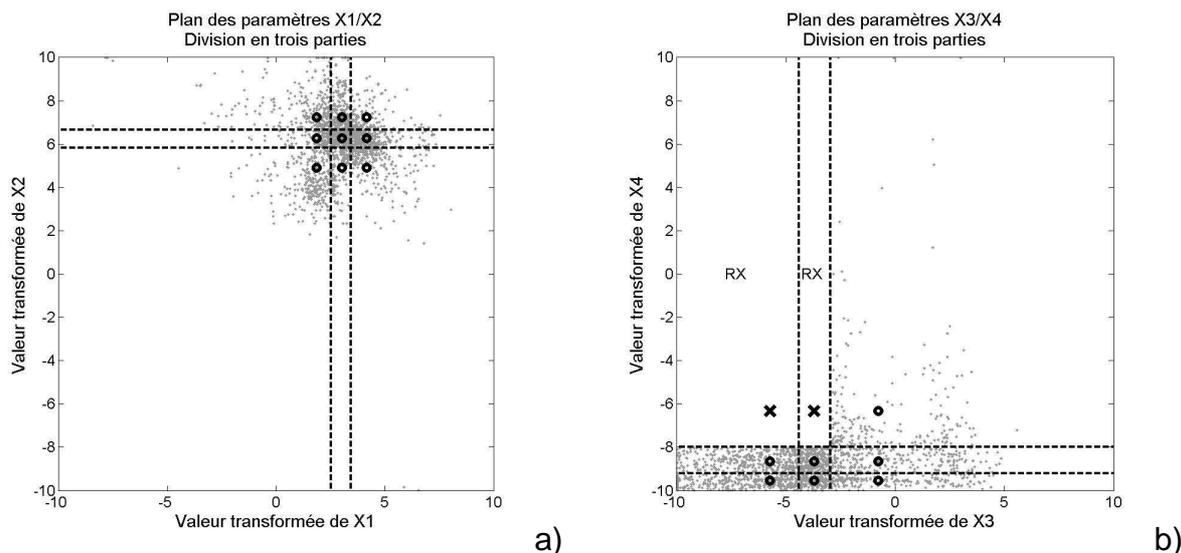


Figure 7-4 : Exemple de filtrage des points de départ possibles (symbolisés par 'o' ou 'x'), dans un espace à deux dimensions. 'o' signifie que les points de départs situés dans une région sont conservés, alors que 'x' signifie qu'ils sont éliminés. a) espace $\{X_1, X_2\}$, b) espace $\{X_3, X_4\}$. Les points gris représentent tous les vecteurs $[(X_1, X_2)$ ou $(X_3, X_4)]$ déjà optimisés.

Par exemple, dans un espace à deux dimensions (Figure 7-4 a, b), dont on a divisé les distributions des paramètres en trois parties, il est possible de délimiter neuf régions. Les signes 'o' ou 'x' représentent les 9 points de départ possibles de chaque espace. Dans le cas de l'espace $\{X_1, X_2\}$ (Figure 7-4 a), on remarque que les vecteurs $[X_1, X_2]$ déjà optimisés sont situés dans les régions des neuf points de départ possibles, qui sont donc conservés. Par contre, dans le cas de l'espace $\{X_3, X_4\}$ (Figure 7-4 b), on remarque qu'aucun des vecteurs $[X_3, X_4]$ n'est situé dans les deux régions marquées RX. On élimine donc ces régions, où aucun vecteur n'a été optimisé.

Pour compléter notre stratégie de filtrage, il est aussi possible de prendre en compte les performances des vecteurs de paramètres optimisés. En effet, il est possible de considérer un seuil de performance minimale à dépasser pour qu'un des vecteurs optimisé occupe une des régions définies précédemment. Ainsi, si tous les vecteurs d'une région ont des performances inférieures à ce seuil, on rejette le point de départ possible de cette région, ce qui permet de réduire encore un peu le nombre de points de départ.

De cette manière, le nombre maximal théorique de points de départ possibles correspond au nombre maximal de vecteurs optimisés de notre base de données, c'est-à-dire, 2093 (nombre de périodes). En pratique, le nombre de points de départs possibles, après les deux étapes de filtrage, est bien inférieur car de nombreux vecteurs de paramètres se situent dans la même région de l'espace.

Tableau 7-3 : Nombre de points de départ possibles après les deux étapes de filtrage. Cas des modèles GR4H, IHAC6, MOR6 TOPMO8 et MORDOR10.

<i>n</i> : nombre de divisions de chaque distribution	Nombre de points de départ possibles après filtrage (seuil $C_{2M} > 50$)				
	GR4H	IHAC6	MOR6	TOPMO8	MORDOR10
2	16	64	63	207	559
3	77	398	373	874	1236
4	208	777	808	1226	1415
5	396	1044	1059	1346	1468
Nombre moyen de simulations pour l'optimisation du modèle	≈ 340	≈ 940	≈ 710	≈ 2440	≈ 3860

Le Tableau 7-3 indique le nombre de points de départs obtenus après les deux étapes de filtrage de tous les points de départ possibles. On remarque que cette méthode permet de réduire fortement le nombre de points de départs possibles. Lorsque deux divisions sont choisies, presque toutes les régions de l'espace des paramètres sont visitées par les vecteurs des modèles GR4H, IHAC6, MOR6 et TOPMO8. Par contre, pour le modèle MORDOR10, la moitié des points de départs possibles est déjà éliminée. Lorsque le nombre de parties est limité à deux ou trois, le nombre de points de départs possibles reste raisonnable par rapport au nombre moyen d'itérations des modèles. Ce sont donc ces nombres de divisions qui sont *a priori* intéressants, puisqu'il est probable qu'ils n'affectent pas l'efficacité de la méthode d'optimisation.

7.2.3 Synthèse de la méthode de pré-filtrage de la surface de réponse du modèle

L'amélioration proposée de la méthode PAP-GR₁ (1 : pour indiquer qu'elle utilise seulement la médiane de la distribution des paramètres comme point de départ possible) est basée sur deux composantes (Figure 7-5) :

1. La génération d'un grand nombre (n^m) de points de départ possibles, en exploitant la distribution des paramètres optimisés du modèle,
2. Le filtrage des n^m points de départ possibles, en utilisant la réserve de vecteurs de paramètres déjà optimisés,
3. La sélection du meilleur des points de départ possibles restant après filtrage, en fonction des performances en simulation du modèle.

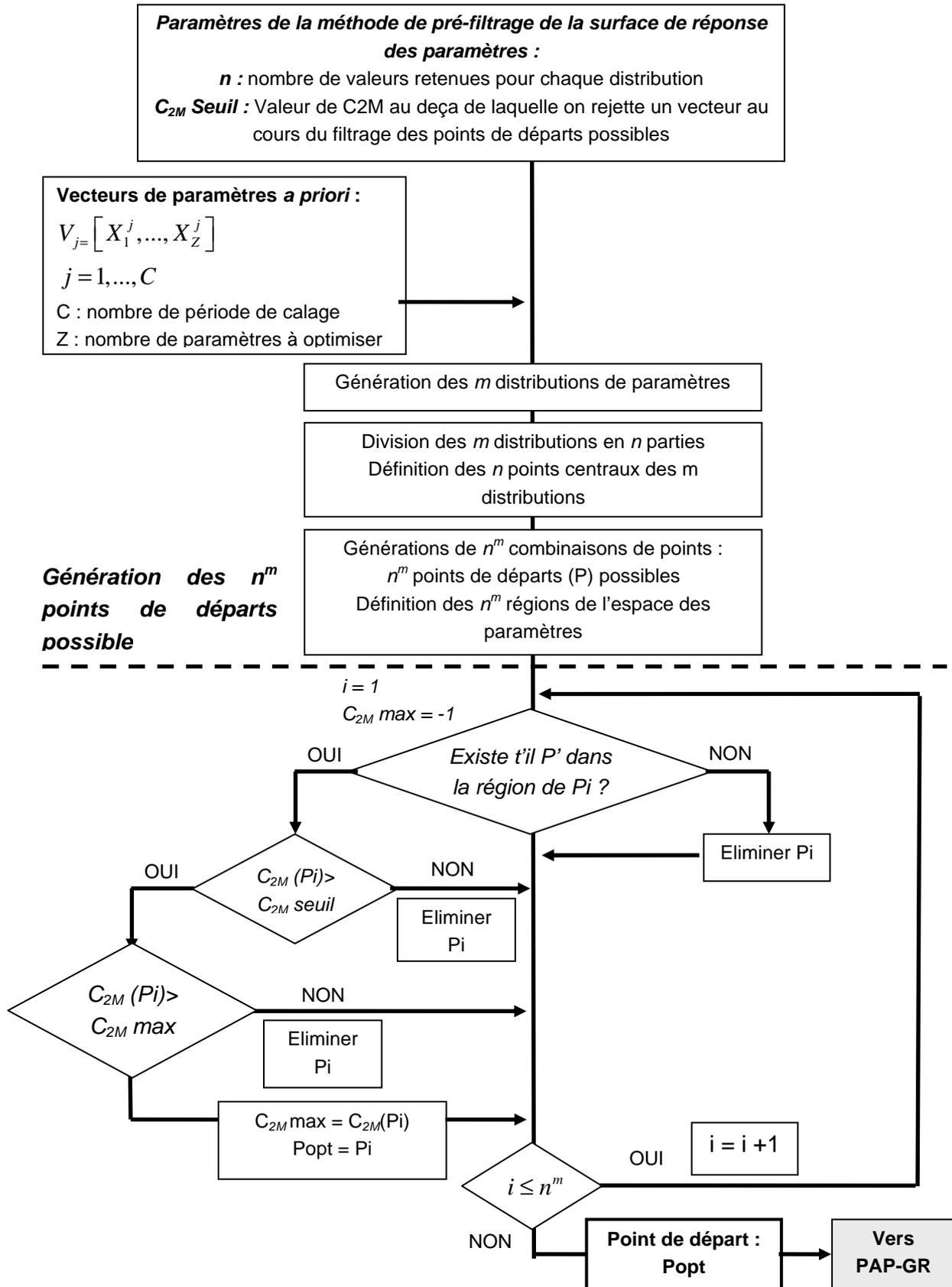


Figure 7-5 : Algorithme de la méthode de génération et de filtrage des points de départ possibles.

7.3 Evaluation de la méthode de calage PAP-GR_n avec pré-filtrage de la surface de réponse des paramètres

7.3.1 Méthodologie

L'évaluation de la méthode de calage PAP-GR_n avec pré-filtrage de la surface de réponse des paramètres est assez simple. Nous allons comparer l'efficacité (performance en calage), l'efficacité (nombre moyen de simulations nécessaires pour optimiser les paramètres d'un modèle) et la robustesse (stabilité des performances du modèle du calage au contrôle) des méthodes PAP-GR₂ à PAP-GR₅, à celle de la méthode PAP-GR₁. Nous n'avons pas testé plus de divisions des distributions des paramètres, le nombre de points de départ à tester devenant prohibitif. Nous avons utilisé ces méthodes pour calibrer les modèles GR4H, IHAC6, TOPMO8 et MORDOR10, sur les 313 bassins versants de la base de données. Il faut noter que la connaissance *a priori* sur les paramètres intègre le bassin dont on cherche à optimiser les paramètres. On aurait pu l'exclure avant d'effectuer les statistiques permettant de définir la distribution a priori des paramètres. Cependant, vues la simplicité du découpage et surtout la taille de l'échantillon, cette opération a semblé entièrement superflue.

7.3.2 Fonctionnement de l'algorithme

Le fonctionnement de l'algorithme de recherche PAP-GR_n est illustré à la Figure 7-6 et à la Figure 7-7.

La Figure 7-6 montre l'évolution des performances des vecteurs testés pendant la phase de pré-filtrage et la phase d'optimisation pas-à-pas. Cette figure illustre bien cette première phase de pré-filtrage, où les vecteurs testés obtiennent une grande variabilité de performances, et où quelques vecteurs ayant de bonnes performances peuvent être identifiés. Ensuite, le meilleur des vecteurs est utilisé comme point de départ de la phase de calage pas à pas. On voit que cette phase de pré-filtrage permet à la méthode de calage de partir d'un vecteur dont la performance est très proche de celle de l'optimum.

De la même manière, la Figure 7-7 illustre ce processus mais dans l'espace des paramètres, en montrant l'évolution de la distance à l'optimum des vecteurs testés au cours du calage. Cette figure montre bien l'exploration de l'espace des paramètres au cours du pré-filtrage, avec de nombreux vecteurs dont les distances à l'optimum peuvent être très variables. De la même manière, la phase de pré-filtrage permet de fournir à la méthode de calage pas à pas un point de départ bien situé dans la région de l'optimum.

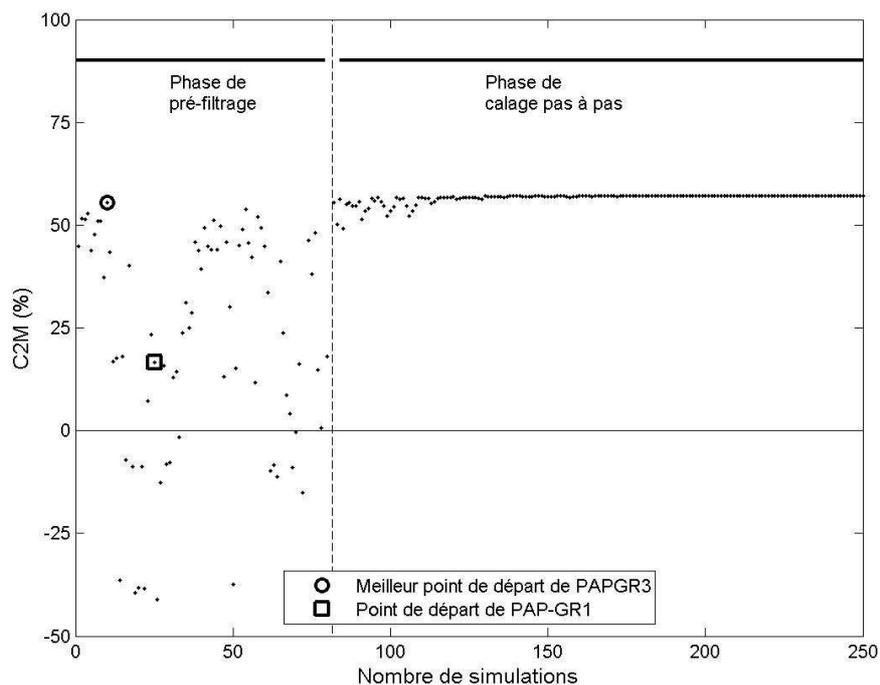


Figure 7-6 : Évolution des performances des vecteurs de paramètres testés au cours d'une procédure de calage. Mise en évidence de la phase de pré-filtrage avec son meilleur départ (rond) et avec le départ de la méthode PAP-GR₁ (carré).

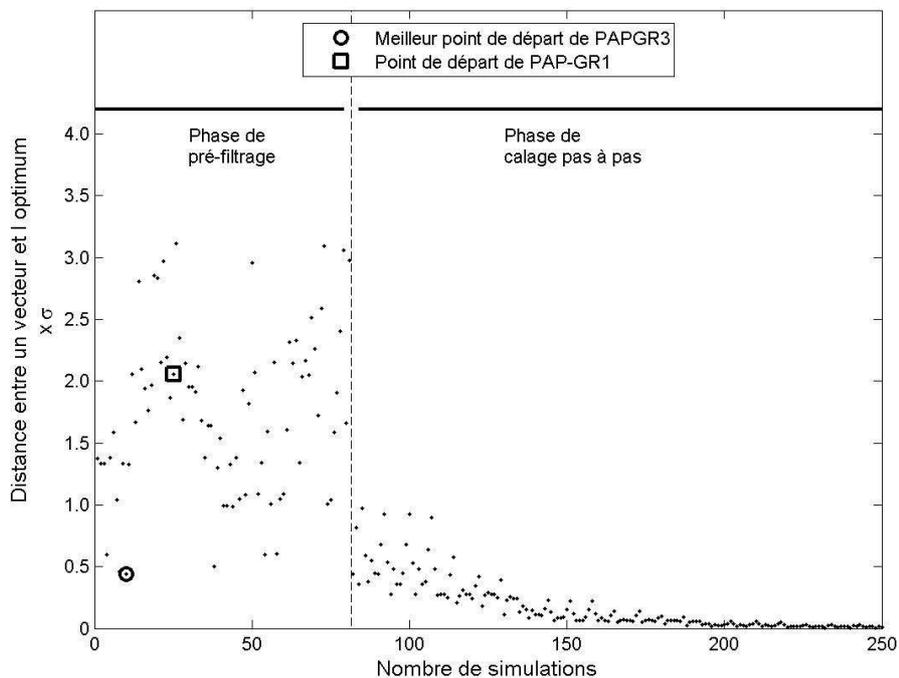


Figure 7-7 : Évolution de la distance à l'optimum des vecteurs de paramètres testés au cours d'une procédure de calage. Mise en évidence de la phase de pré-filtrage avec son meilleur départ (rond) et avec le départ de la méthode PAP-GR₁ (carré).

Ces deux figures illustrent clairement l'intérêt de la phase de pré-filtrage qui permet d'explorer de manière efficace la surface de réponse du modèle et de choisir ensuite un point de départ situé dans la région de l'optimum de la surface de réponse.

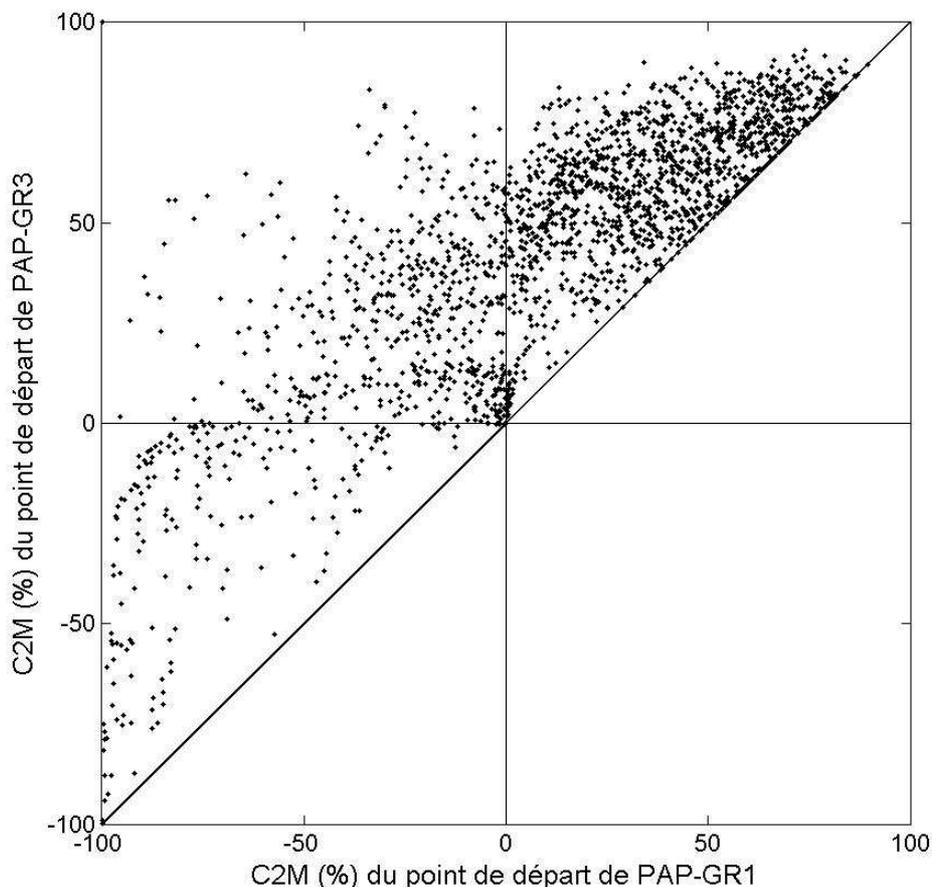


Figure 7-8 : Comparaison des performances du point de départ des algorithmes PAP-GR₁ et PAP-GR₃.

La Figure 7-8 illustre la différence de performances entre le point de départ de la méthode PAP-GR₁ et celui de la méthode PAP-GR₃. Cette figure montre la très forte différence possible de performances entre ces points de départ. La performance moyenne ($C_{2M}(RQ)$) du point de départ de la méthode PAP-GR₁ est de 12 points, alors qu'elle est de 44 points pour la méthode PAP-GR₃. Cette forte différence de performances, qui signifie un mauvais positionnement du point de départ de PAP-GR₁, peut d'ailleurs être à l'origine de l'échec de la méthode PAP-GR₁ dans certains cas. En effet, lorsque le point de départ est très mauvais et très éloigné de la région de l'optimum, il peut se trouver dans une zone où la surface de réponse est tellement insensible, que la méthode pas à pas ne sait pas dans quelle direction évoluer.

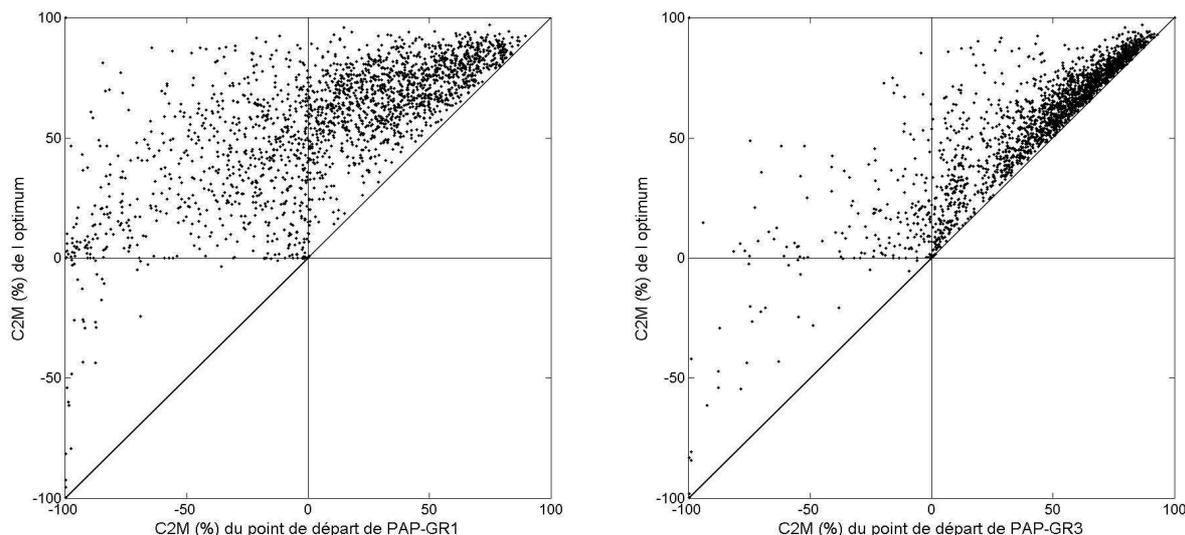


Figure 7-9 : Comparaison des performances du point de départ des algorithmes $PAP-GR_n$ et de l'optimum trouvé par la méthode. a) $PAP-GR_1$, b) $PAP-GR_3$.

La Figure 7-9 compare la différence de performances entre le point de départ de la méthode $PAP-GR_n$ et l'optimum identifié par cette méthode, pour les méthodes $PAP-GR_1$ et $PAP-GR_3$. La Figure 7-9 a) montre la très grande différence de performances entre le point de départ de la méthode $PAP-GR_1$ et l'optimum, qui est en moyenne de 45 points de C_{2M} . En revanche, grâce à la méthode de pré-filtrage, la différence de performances moyennes entre le point de départ de la méthode $PAP-GR_3$ et l'optimum est réduite, elle n'est plus que de 14 points de C_{2M} (Figure 7-9 b).

Enfin, l'intérêt de la méthode de pré-filtrage est illustré par la Figure 7-10, qui compare la distance à l'optimum du point de départ de la méthode $PAP-GR_1$ et de la méthode $PAP-GR_3$. Cette figure montre dans une majorité des cas (85%), la distance à l'optimum de la méthode $PAP-GR_1$ est bien plus grande que celle de la méthode $PAP-GR_3$. Ainsi, la méthode de pré-filtrage est capable de cibler généralement la région de l'optimum et proposer un vecteur qui se situe proche de l'optimum. Cela permet donc à cette méthode de limiter le nombre de simulations dont elle a besoin pour localiser l'optimum, ce qui diminue les probabilités de rester piégé sur un optimum secondaire.

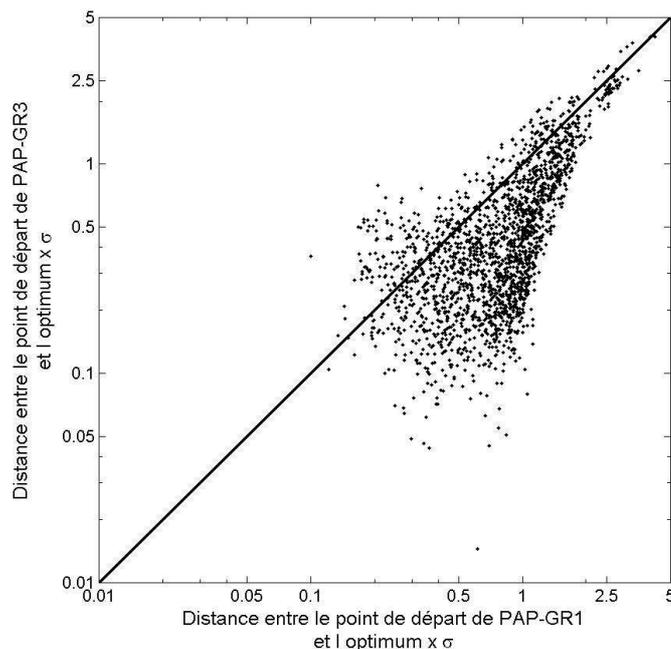


Figure 7-10 : Comparaison des distances à l'optimum du point de départ des algorithmes PAP-GR₁ et PAP-GR₃ (distance exprimée par rapport à l'écart-type de la distribution de chaque paramètre : 1 signifie une fois l'écart type moyen de chaque paramètre).

7.3.3 Performances des méthodes

Dans cette section, l'analyse des résultats est basée sur la comparaison des performances en calage, en validation et du nombre moyen de simulations nécessaires pour calibrer les paramètres d'un modèle.

Performance en calage :

A titre d'exemple, les performances en calage des vecteurs de paramètres optimisés avec les méthodes PAP-GR₁ et PAP-GR₃ sont comparées pour les modèles GR4H (Figure 7-11) et MORDOR10 (Figure 7-12). Ces deux figures illustrent les deux types de résultats que l'on peut obtenir pour les quatre modèles optimisés avec les différentes méthodes d'optimisation.

Sur ces figures, on voit qu'une bonne partie des vecteurs de paramètres conduit aux mêmes performances : environ 80% pour le modèle GR4H (Tableau 7-4) et 55% pour MORDOR10 (Tableau 7-5). Néanmoins, il existe une certaine dispersion des performances des vecteurs de paramètres restant. Dans le cas du modèle GR4H, la tendance est d'avoir de meilleurs vecteurs de paramètres avec la méthode PAP-GR₃ (10% des vecteurs). Sur cette figure, on remarque de nombreux cas où les différences de performances sont de plus de 10 points, ce qui indique bien l'arrêt de la méthode PAP-GR₁ sur un optimum local et l'intérêt de la méthode PAP-GR₃.

Lorsque les vecteurs de paramètres optimisés avec la méthode PAP-GR₃ ont des performances inférieures à celles obtenues avec la méthode PAP-GR₁ (5% des vecteurs), ces différences de performances sont généralement limitées et il existe peu de cas de fort échec sur un optimum local. En revanche, dans le cas de MORDOR10, on voit que la dispersion des performances en calage est beaucoup plus équilibrée autour de la bissectrice, avec un léger avantage pour la méthode d'optimisation initiale. Cependant, on voit qu'avec ce modèle, la méthode initiale a de forts échecs sur les bassins versants arides des États-unis, que n'a plus la méthode PAP-GR₃.

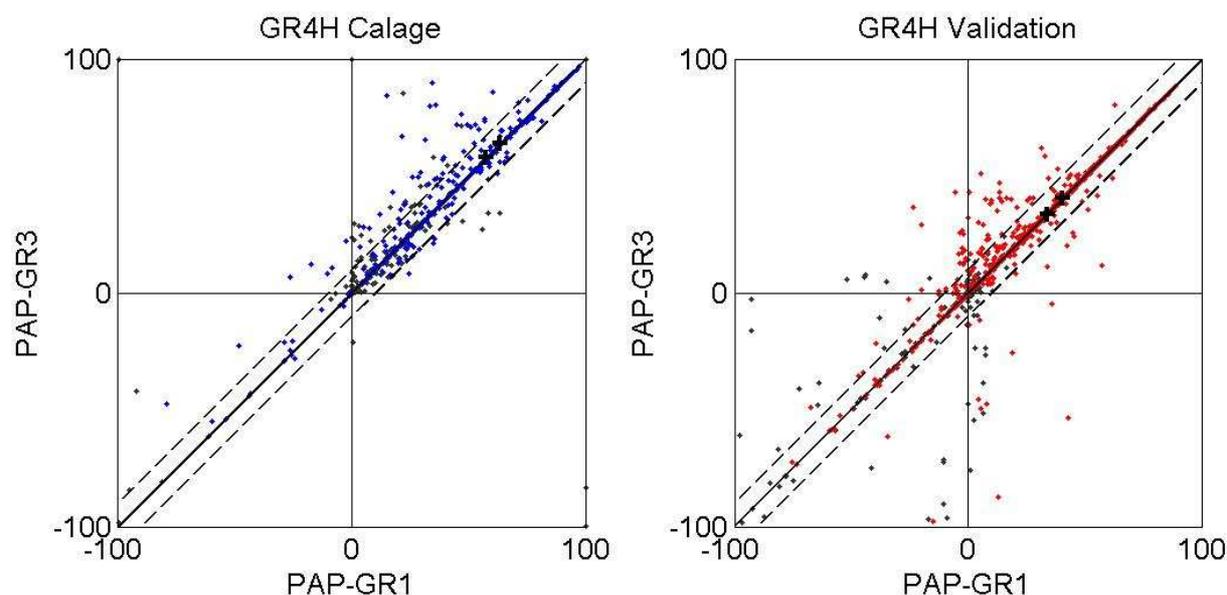


Figure 7-11 : Comparaison des performances ($C_{2M}(Q)$) en calage et en validation des vecteurs trouvés par les méthodes PAP-GR₁ et PAP-GR₃, pour le modèle GR4H (en noir sont indiqués les bassins versants arides des États-unis). Les tiretés représentent les courbes enveloppe +10 / -10.

Tableau 7-4 : Pourcentages de calages où deux méthodes donnent des critères C_{2M} égaux (à +/- 0.1 points), et où l'une est supérieure ou inférieure à l'autre (modèle GR4H)

GR4H Méth.1 / Méth.2	CALAGE		VALIDATION	
	PAPGR ₁ / PAPGR ₂	PAPGR ₁ / PAPGR ₃	PAPGR ₁ / PAPGR ₂	PAPGR ₁ / PAPGR ₃
Méth.1 ≈ Méth. 2	84.5	83.4	57.6	57.0
Méth.1 > Méth.2	5.8	6.1	18.2	18.7
Méth.2 > Meth.1	9.7	10.5	24.2	24.3

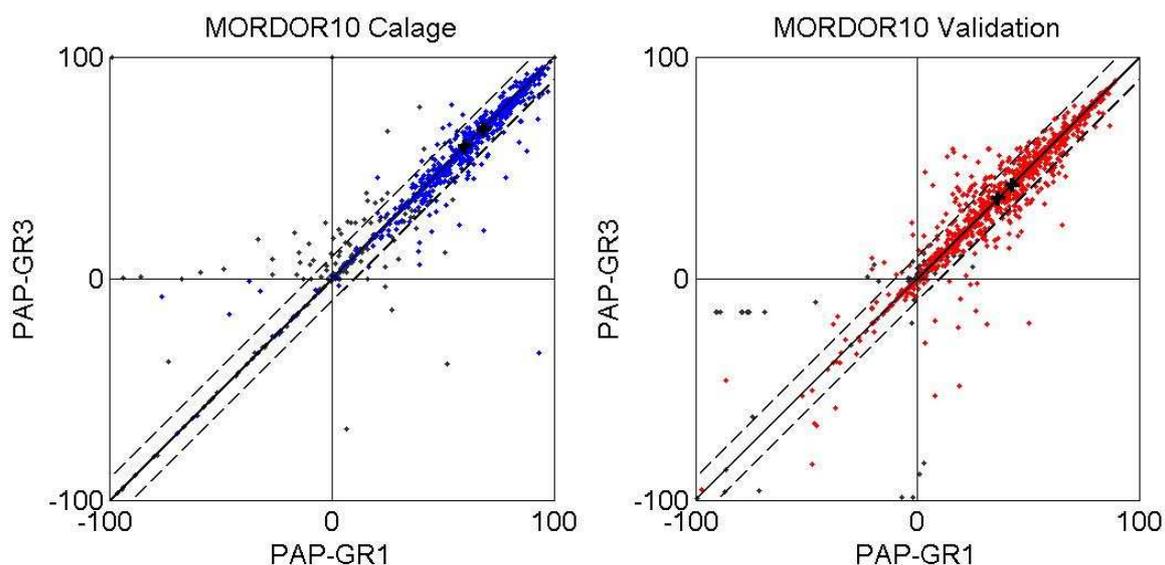


Figure 7-12 : Comparaison des performances ($C_{2M}(Q)$) en calibration et en validation des vecteurs trouvés par les méthodes $PAP-GR_1$ et $PAP-GR_3$, pour le modèle MORDOR10 (en noir sont indiqués les bassins versants arides des États-unis). Les tiretés représentent les courbes enveloppe +10 / -10.

Tableau 7-5 : Pourcentages de calages où deux méthodes donnent des critères C_{2M} égaux (à +/- 0.1 points), et où l'une est supérieure ou inférieure à l'autre (modèle MORDOR10).

MORD	CALAGE		VALIDATION	
	PAPGR ₁ / PAPGR ₂	PAPGR ₁ / PAPGR ₃	PAPGR ₁ / PAPGR ₂	PAPGR ₁ / PAPGR ₃
Méth.1 ≈ Méth. 2	52.0	57.6	35.5	38.0
Méth.1 > Méth.2	23.3	25.7	34.2	33.0
Méth.2 > Meth.1	24.6	16.7	30.3	29.0

La comparaison des performances de la méthode $PAP-GR_2$ et $PAP-GR_3$ (Tableau 7-4, Tableau 7-5) montre également que ces deux méthodes donnent à peu près les mêmes résultats pour le modèle GR4H, alors qu'ils sont sensiblement différents pour le modèle MORDOR10. La variabilité des pourcentages où les deux méthodes sont égales, ou bien supérieures ou inférieures à la méthode $PAP-GR_1$, révèle la sensibilité de ce modèle à la méthode d'optimisation. Ce constat est renforcé par le caractère équilibré de la distribution des performances autour de la bissectrice, qui indique qu'avec ce modèle, il y a de fortes probabilités de tomber sur un optimum secondaire. Ainsi, l'optimisation de MORDOR10 semble fortement sensible au point de départ et à l'algorithme de calage, ce qui révèle encore une fois la grande complexité de sa surface de réponse et éventuellement certains problèmes structuraux dans sa formulation.

Les gains de performance en calage des méthodes PAP-GR₂ à PAP-GR₅, par rapport à la méthode PAP-GR₁, sont présentés à la Figure 7-13. Sur cette figure, on remarque tout d'abord que les gains en calage sont limités (inférieurs à 1 point) pour les modèles parcimonieux (GR4H et IHAC6) et qu'ils peuvent être un peu plus importants (de 2 à 3 points) pour les modèles TOPMO8 et MORDOR10. Par rapport aux performances des méthodes d'optimisation testées au chapitre 0, les performances en calage des méthodes PAP-GR₂ à PAP-GR₅, sont intermédiaires entre les performances de la méthode PAP-GR₁ et celles des méthodes globales.

Les gains en calage des modèles GR4H et IHAC6, qui sont limités et indépendants du nombre de divisions choisies, indiquent l'insensibilité de ces modèles au point de départ choisi, ce qui souligne encore une fois la facilité à caler ces modèles. Cette facilité provient certainement de leurs surfaces de réponse particulièrement simples et marquées par un optimum principal facilement identifiable. En revanche, les gains en calage des modèles TOPMO8 et MORDOR10, qui peuvent être assez forts et assez variables, indiquent bien la plus grande difficulté, déjà constatée, à caler ces modèles. La sensibilité au nombre de divisions choisies, et donc au point de départ, confirme bien l'existence de nombreux optima locaux dans la surface de réponse de ces modèles.

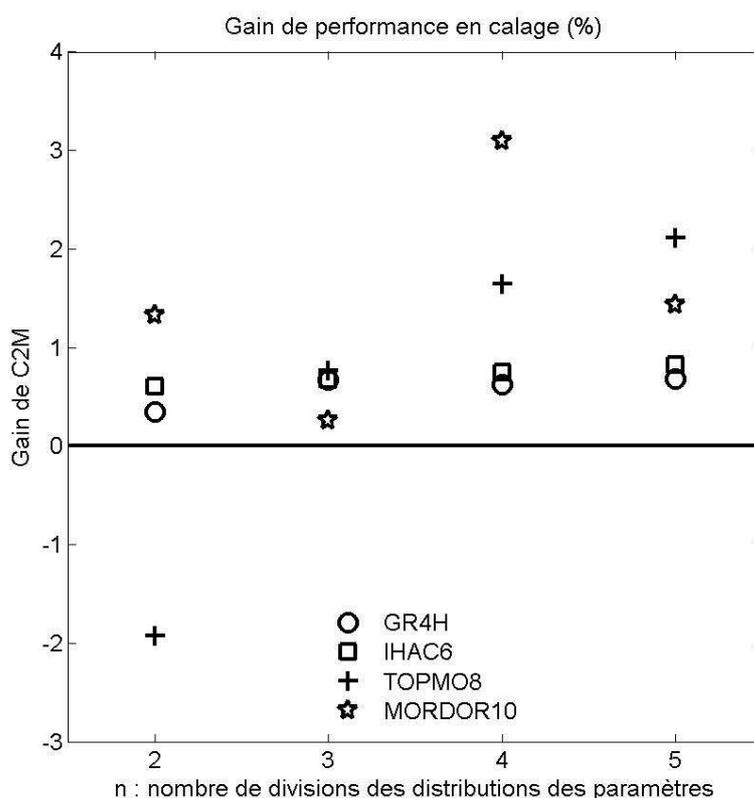


Figure 7-13 : Gains de performance en calage des modèles GR4H, IHAC6, TOPMO8 et MORDOR10, en fonction du nombre de divisions des distributions des paramètres.

Performances en validation :

Au cours de la validation, le pourcentage de vecteurs qui ont les mêmes performances diminue pour les deux modèles de 20 à 30 points (Tableau 7-4, Tableau 7-5). La dispersion des performances des vecteurs autour de la bissectrice augmente (Figure 7-11, Figure 7-12). Pour le modèle GR4H, l'avantage en calage de la méthode PAP-GR₃ se conserve nettement en validation, alors que pour le modèle MORDOR10, la probabilité que la méthode initiale soit supérieure ou inférieure aux autres méthodes est la même.

Les gains de performances en validation des méthodes PAP-GR₂ à PAP-GR₅, par rapport à celle de la méthode PAP-GR₁, sont présentés à la Figure 7-14. Cette figure met en évidence que les gains de performance en validation des modèles GR4H et IHAC6 sont à peu près du même ordre, quel que soit le nombre de divisions choisies.

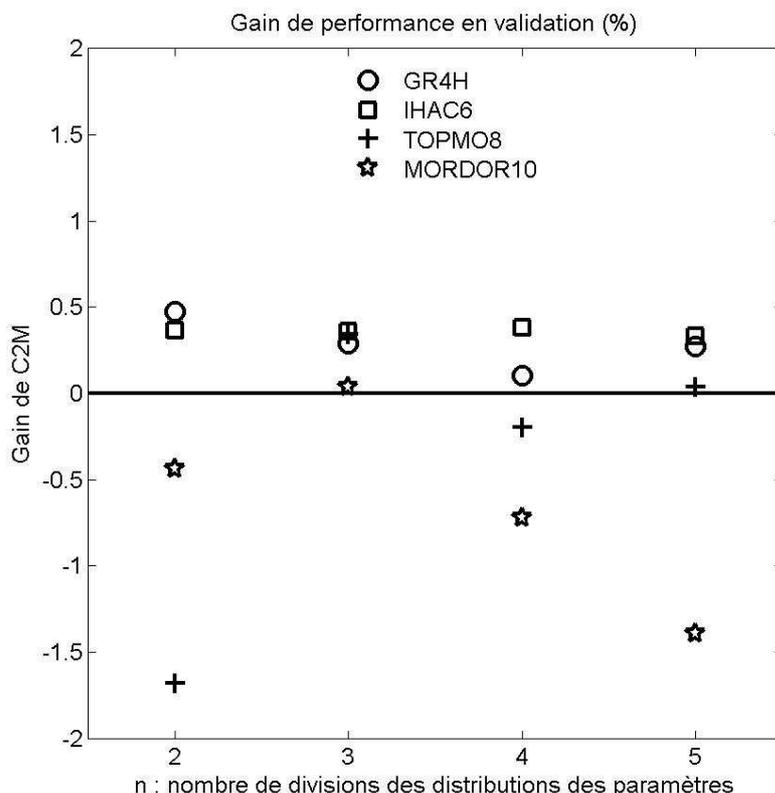


Figure 7-14 : Gains de performance en validation des modèles GR4H, IHAC6, TOPMO8 et MORDOR10, en fonction du nombre de divisions des distributions des paramètres.

En revanche, pour les modèles TOPMO8 et MORDOR10, dans trois cas sur quatre, les performances en validation sont égales ou inférieures aux performances initialement obtenues avec la méthode PAP-GR₁. Il est néanmoins possible de trouver un cas (trois divisions) où les performances en validation de ces modèles progressent légèrement.

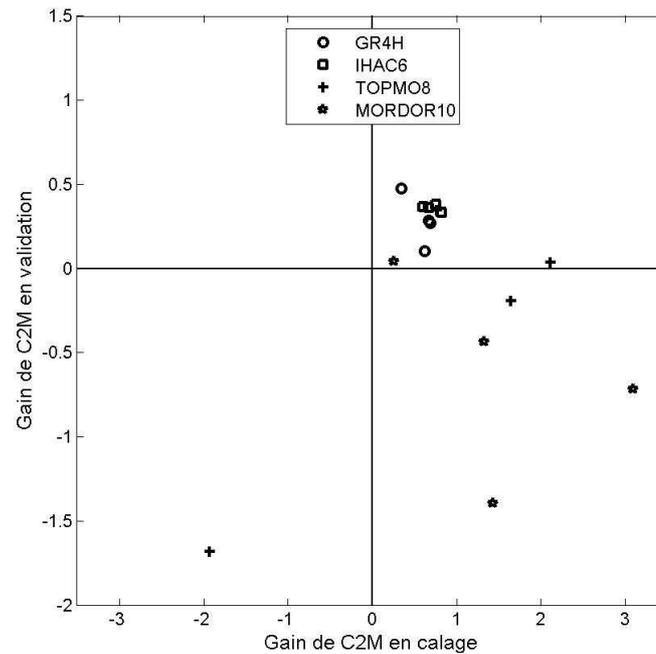


Figure 7-15 : Comparaison des gains de performance en calage et en validation des modèles GR4H, IHAC6, TOPMO8 et MORDOR10.

Ces résultats indiquent paradoxalement que, pour les modèles parcimonieux, la proposition d'un meilleur point de départ permet au modèle de limiter le nombre de cas où la méthode d'optimisation s'arrête sur un optimum secondaire. Ce qui entraîne une amélioration robuste des performances de ces modèles, aussi bien en calage qu'en validation.

En revanche, pour les modèles TOPMO8 et MORDOR10, une amélioration des performances en calage ne signifie pas une amélioration des performances en validation, au contraire. Le même phénomène que celui déjà constaté avec les méthodes globales a lieu, lorsqu'on améliore le point de départ de la méthode PAP-GR et ainsi les performances en calage. La Figure 7-15 met en évidence la bonne robustesse des modèles GR4H et IHAC6 et la faible robustesse des modèles TOPMO8 et MORDOR10.

Efficiences :

L'efficacité représente la rapidité d'une méthode à trouver le bon optimum. Ainsi, pour valider la méthode de pré-filtrage de la surface de réponse des paramètres d'un modèle, il faut que son efficacité et son efficacité aient progressé. La Figure 7-16 montre le pourcentage de simulations gagnées ou perdues par la méthode utilisant le pré-filtrage, par rapport à la méthode initiale, en fonction du nombre de divisions. Ce gain est exprimé en pourcent : - 15 % correspond à une diminution du nombre moyen de simulations de 15% et + 100%, correspond à un doublement du nombre de simulations.

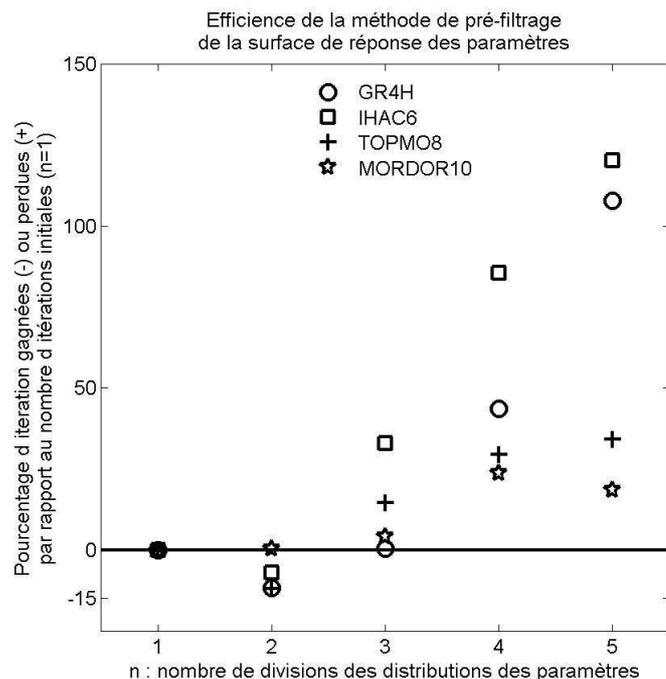


Figure 7-16 : Gains d'efficience (pourcentage d'itérations gagnées ou perdues) des modèles GR4H, IHAC6, TOPMO8 et MORDOR10, en fonction du nombre de divisions des distributions des paramètres.

Sur cette figure, on remarque donc que seule la division en deux parties permet d'améliorer légèrement l'efficience de la méthode PAP-GR pour les quatre modèles. A partir de trois divisions, l'efficience est diminuée. L'efficience des modèles parcimonieux (GR4H, IHAC6) diminue plus vite que celle des autres modèles, car une faible proportion des points de départ possibles est éliminée par la première partie de la méthode de filtrage. En effet, ces modèles étant parcimonieux, ils sont capables d'utiliser des vecteurs de paramètres issus de presque toutes les régions de l'espace des paramètres.

Dans le cas du modèle GR4H, dont nous avons amélioré la structure (Chapitre 1), un gain d'efficience de 15% est loin d'être négligeable, bien au contraire. Il correspond à la diminution de la durée du test d'une structure à quatre paramètres sur les 313 bassins versants d'une heure (ce qui représente, à l'échelle de la durée des calculs de cette thèse, une durée cumulée de 3 à 4 mois !).

7.3.4 Vecteurs de paramètres

A titre d'exemple, les vecteurs de paramètres optimisés avec les méthodes PAP-GR₁ et PAP-GR₃ sont comparées pour les modèles GR4H (Figure 7-17) et MORDOR10 (Figure 7-19). Ces figures précisent l'analyse des résultats obtenus au cours du calage des modèles. De la même manière que pour les performances en calage, la

Figure 7-17 et la Figure 7-19 indiquent la plus ou moins grande dispersion des vecteurs de paramètres optimisés par les méthodes PAP-GR₁ et PAP-GR₃.

Le modèle GR4H est caractérisé par une assez bonne cohérence des paramètres optimisés par différentes méthodes de calage, comme nous l'avons déjà remarqué au chapitre 0. Environ 80% des vecteurs optimisés (Figure 7-20) sont dans la même région de l'espace des paramètres. Cette dispersion des paramètres optimisés est due en partie aux bassins versants arides des États-unis, sur lesquels les performances des modèles sont assez mauvaises et où le processus d'optimisation s'arrête souvent un peu n'importe où. On remarque que pour le modèle GR4H, la valeur du paramètre X1 optimisée avec la méthode PAP-GR₃ a tendance à être légèrement supérieure et que la valeur du paramètre X2 optimisée avec la méthode PAP-GR₃ a tendance à être légèrement inférieure, aux valeurs des paramètres obtenus avec la méthode PAP-GR₁. La Figure 7-18 met en évidence cette éventuelle tendance et la possible corrélation entre les valeurs des paramètres X₁ et X₂, qui semblent se compenser dans ces cas là.

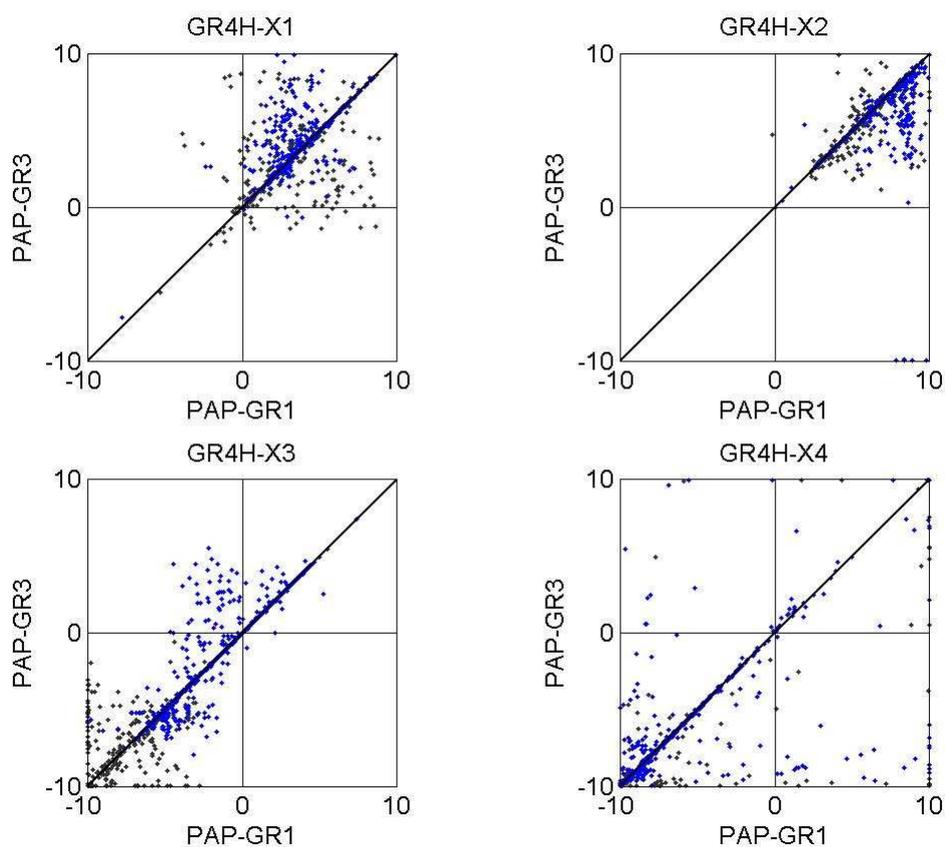


Figure 7-17: Comparaison des valeurs des paramètres de GR4H optimisés avec les méthodes PAP-GR₁ et PAP-GR₃ (en noir sont indiqués les bassins versants arides des États-unis).

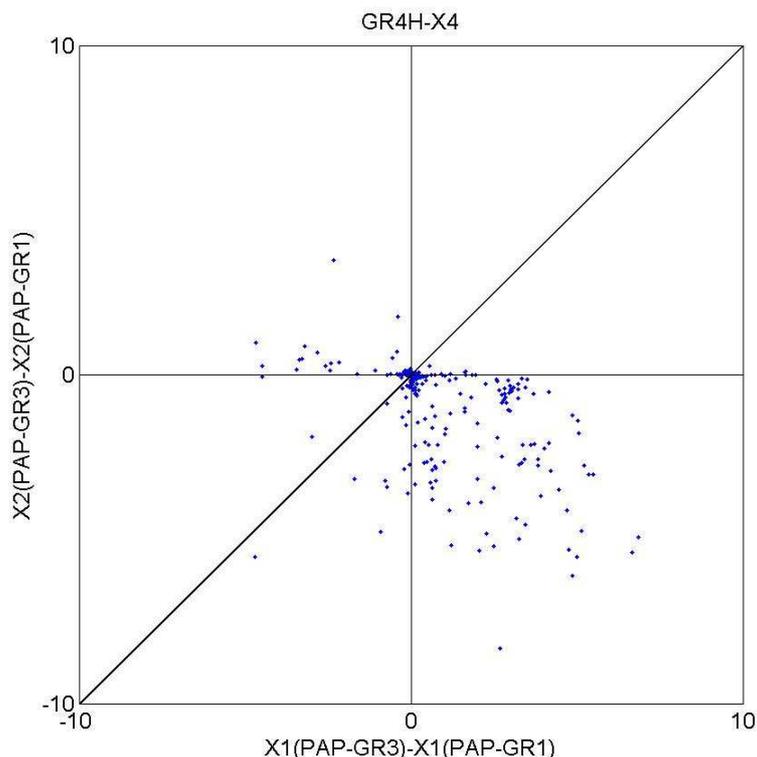


Figure 7-18 : Relation entre les différences de valeurs obtenues pour le paramètre X1 et le paramètre X2, optimisés avec la méthode PAP-GR₁ et PAP-GR₃.

Les paramètres du modèle MORDOR10 observent le même comportement que celui constaté précédemment au chapitre 4. En effet, comme l'indique la Figure 7-19, la dispersion de leurs valeurs est globalement assez grande et plus particulièrement pour certains paramètres (X_2 , X_3 , X_5 , X_6 , X_9). Quel que soit le point de départ choisi, seulement 30% des vecteurs de paramètres (Figure 7-20) se trouvent dans la même région de la surface de réponse des paramètres.

Ce résultat est néanmoins assez étonnant puisque la méthode d'optimisation est ce coup-ci la même, seul le point de départ change. Ceci indique finalement la complexité et le caractère chaotique de la surface de réponse de ce modèle. Cela révèle par ailleurs le caractère mal défini de certaines composantes de ce modèle, comme nous l'avons vu au chapitre Chapitre 4. Même si ses performances sont très bonnes, la grande diversité des vecteurs de paramètres 'optimaux' que l'on peut obtenir est inquiétante et révèle quand même une certaine faiblesse de ce modèle.

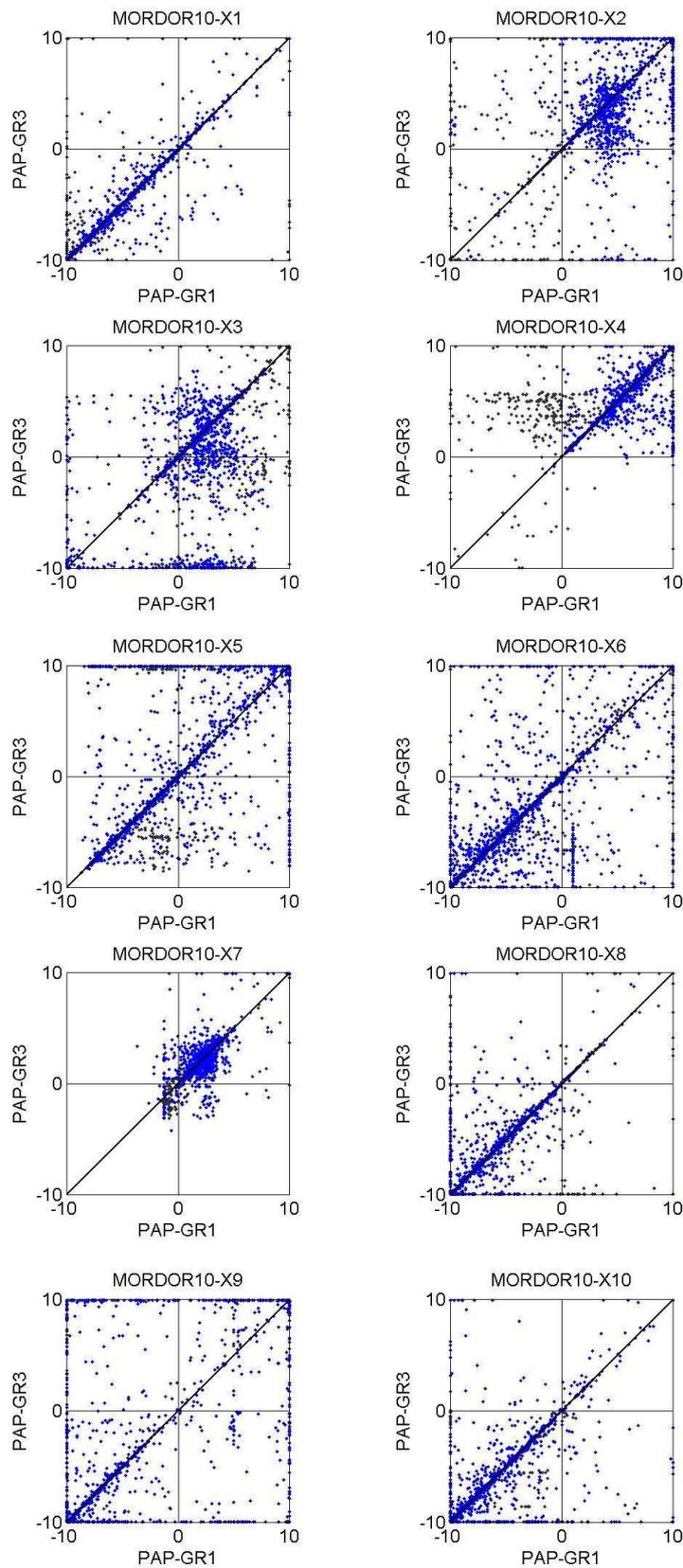


Figure 7-19 : Comparaison des valeurs des paramètres de MORDOR10 optimisés avec les méthodes PAP-GR1 et PAP-GR3 (en noir sont indiqués les bassins versants arides des États-unis).

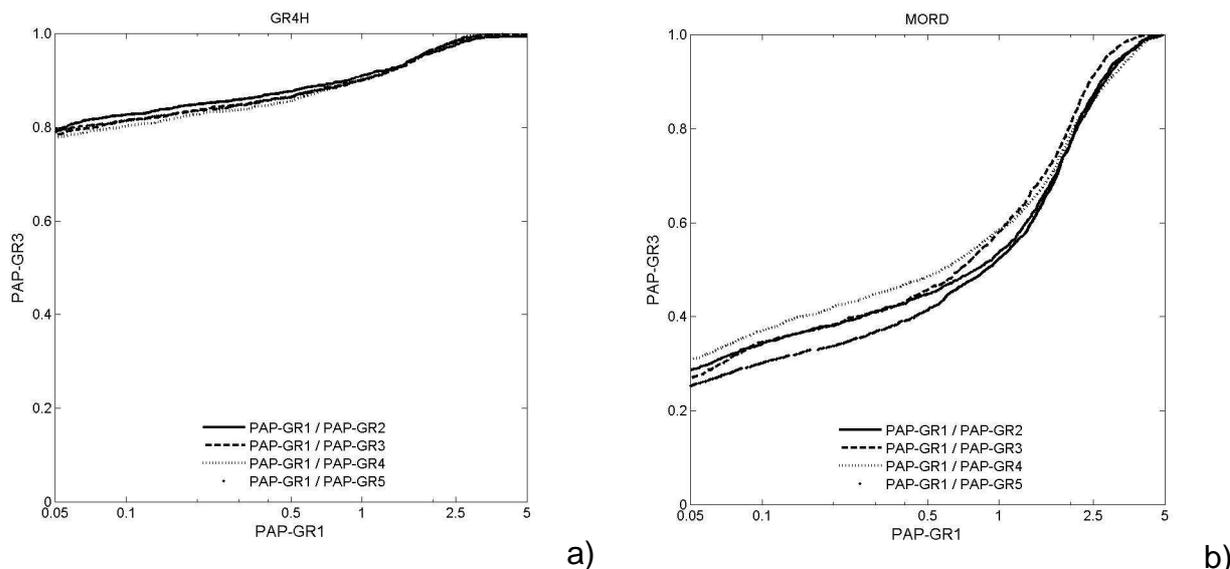


Figure 7-20 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différentes, a) modèle GR4H, b) MORDOR10.

7.4 Amélioration de l'efficacité de PAP-GR_n : étude de sensibilité à la valeur du critère de terminaison (déviations : ΔX_{MIN})

La méthode d'optimisation PAP-GR_n, dont l'algorithme est présenté en Annexe 9.1, possède un critère d'arrêt qui permet de stopper la recherche d'un meilleur vecteur, lorsqu'aucune amélioration du critère de calage n'est constatée. Ce critère peut être, soit un nombre maximal d'itérations autorisée (pour PAP-GR_n : 50 000 itérations), soit une valeur minimale de la déviation ΔX_{MIN} testée. En effet, la méthode PAP-GR_n, opère sa recherche pas-à-pas en testant des valeurs de $X_i \pm \Delta X_i$, ΔX étant la valeur de la déviation du paramètre X_i . Au cours de l'optimisation, cette valeur de déviation diminue au fur et à mesure que l'algorithme se rapproche de la région de l'optimum, afin d'augmenter sa précision pour estimer la valeur de chaque paramètre. Dans l'algorithme initial, la déviation minimale est $\Delta X_{MIN} = 0.01$ (valeur transformée). Les résultats du Chapitre 0 ayant montré que de meilleures performances en calage n'assuraient pas de meilleures performances en validation, il est peut être intéressant de réduire la précision de l'algorithme de recherche PAP-GR_n, en augmentant la valeur de la déviation minimale.

Dans cette partie, nous avons testé différentes valeurs de ΔX_{MIN} (0.02, 0.04, 0.08, 0.16) et nous avons comparé les performances en calage, en validation et le nombre de simulations nécessaires, à celles de la méthode initiale ayant pour déviation minimale 0.01. La précision de chaque paramètre, associée à un taux de déviation défini, est présentée au Tableau 7-6. Par exemple, une précision de +/- 4% signifie que l'on ne peut faire varier la valeur d'un paramètre que par pas de 4%. Ainsi, plus

le taux de déviation minimale est grand, moins on est exigeant sur la précision de l'optimum trouvé au calage.

Tableau 7-6 : Précision sur les paramètres X_1, X_2, X_3, X_4 en fonction du taux d'auscultation. 0.64 est le taux initial et 0.01 est le taux final de l'algorithme PAP-GR_n. Performances en calage, validation et nombre d'itérations moyennes en fonction des différents taux de déviation testés.

		Pas d'auscultation : ΔX					
		0.01	0.02	0.04	0.08	0.16	0.64
Précision sur X_i	X1	+/- 1%	+/- 2%	+/- 4%	+/- 8%	+/- 17%	+/- 90%
	X2	+/- 1%	+/- 2%	+/- 4%	+/- 8%	+/- 17%	+/- 90%
	X3	+/- 1%	+/- 2%	+/- 4%	+/- 8%	+/- 15%	+/- 47%
	X4	+/- 0.4%	+/- 0.7%	+/- 1%	+/- 3%	+/- 6%	+/- 25%
C2M en calage		61.3	61.2	60.9	60.6	60.0	/
C2M en validation		37.9	37.9	37.8	37.6	37.5	/
Nombre d'itérations		338	295	252	212	175	/

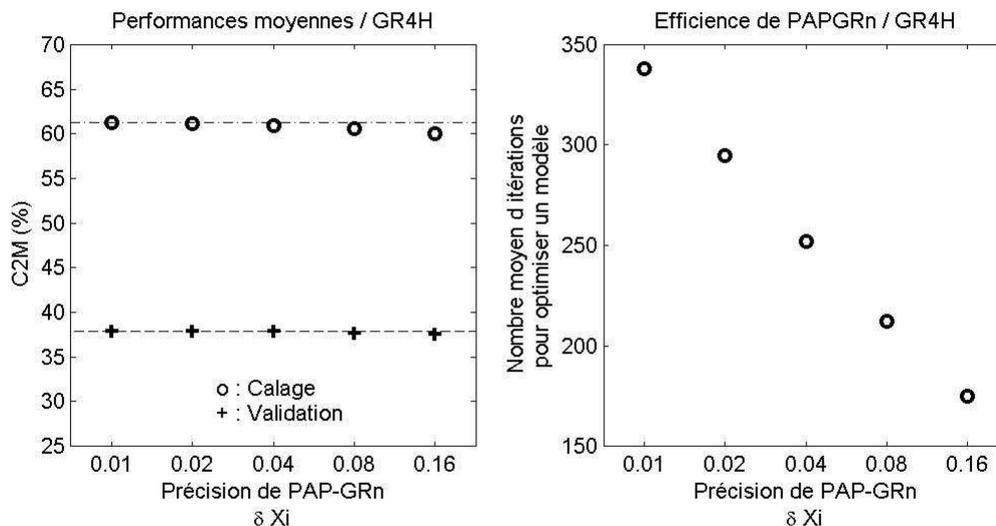


Figure 7-21 : Influence sur les performances en calage, en validation et sur l'efficacité de la valeur du critère d'arrêt de la méthode PAP-GR_n (déviation minimale), pour le modèle GR4H.

La Figure 7-2 et le Tableau 7-6 montrent l'influence de la valeur du critère d'arrêt choisie sur les performances en calage, en validation et sur l'efficacité du calage. Lorsque le taux de déviation passe de 0.01 à 0.16, les performances en calage perdent seulement 1.3 points en calage et 0.4 points en validation, pour un gain d'efficacité de 50% (diminution du nombre d'itérations par deux). Un taux de déviation de 0.04 permet de ne perdre que 0.1 points de C_{2M} par rapport aux résultats de la méthode initiale, tout en diminuant le nombre d'itérations de 25%.

7.5 Synthèse

L'évaluation des méthodes PAP-GR₂ à PAP-GR₅ a montré que la méthode proposée pour le pré-filtrage de la surface de réponse des paramètres d'un modèle donnait de bons résultats pour les modèles parcimonieux testés (GR4H et IHAC6). Pour ces deux modèles, il est possible d'améliorer modestement, mais robustement, leurs performances en calage et en validation, tout en diminuant le temps de calcul de la méthode de calage de 10% à 15%. La comparaison des vecteurs de paramètres optimisés par les méthodes PAP-GR₁ et PAP-GR₃, a révélé encore une fois la facilité à caler ces modèles et la bonne structure de leur surface de réponse qui permet à n'importe quelle méthode d'optimisation, utilisant différents points de départ de tendre vers le même optimum dans plus de 80% des cas.

Pour les modèles ayant huit à dix paramètres à optimiser (TOPMO8 et MORDOR10), l'intérêt de cette méthode est plus limité : les gains de performances en validation et d'efficacité en calage sont faibles. Le moindre intérêt de cette méthode vient de la complexité de la surface de réponse de ces modèles, qui pose des problèmes d'optimisation (quelles que soient les méthodes d'optimisation utilisées) et de la mauvaise définition de l'optimum de la surface de réponse de ces modèles. D'ailleurs, comme l'ont montré le Chapitre 4 et la comparaison des vecteurs de paramètres optimisés par les méthodes PAP-GR₁ et PAP-GR₃, ce n'est pas la méthode d'optimisation qu'il faut améliorer, mais la structure de ce type de modèle. En effet, les vecteurs de paramètres 'optimaux' de ces modèles dépendent beaucoup trop de la méthode d'optimisation et du point de départ choisi.

Ces résultats suggèrent qu'il est intéressant d'utiliser cette méthode avec deux divisions des distributions de paramètres pour les modèles ayant jusqu'à six paramètres à optimiser, et trois divisions pour les modèles ayant plus de six paramètres à optimiser.

Par ailleurs, l'intérêt potentiel de cette méthode pour les utilisateurs qui ne disposent pas des distributions des paramètres d'un modèle, est de générer un grand nombre de points de départ, en proposant par exemple une valeur forte et une valeur basse pour chaque paramètre. Avec seulement deux valeurs par paramètres, cela permet de limiter le nombre de points de départ possibles, mais de tester différents points de départ en mettant à profit l'expérience de la personne qui utilise cette méthode.

L'autre application possible de cette méthode est d'utiliser une approche multi-départs. Même si nos résultats montrent que ce type d'approche n'est pas nécessaire (soit l'optimum est identifiable et n'importe quelle méthode y arrive, soit il

n'est pas identifiable et toutes les méthodes convergent vers des optima différents), c'est une éventualité intéressante. L'application de cette méthode selon un mode multi-départs est très simple : il suffirait de sélectionner les P meilleurs points de départs et de réaliser plusieurs optimisations en parallèle, avec la méthode de recherche locale.

Enfin, l'étude de sensibilité sur le pas minimum d'auscultation a montré qu'il était possible, en prenant un critère de terminaison moins exigeant, de diminuer la durée du calage d'un modèle d'au moins 25% sans limiter les performances en validation. Ces résultats confirment également qu'il n'est pas nécessaire de définir un vecteur de paramètres avec trop de précision, pour obtenir de bonnes performances d'un modèle.

Chapitre 8

***« para mi las deudas del juego son siempre
deudas de honor. Te entrego lo que más
quiero, pero te la entrego muerta, aunque me
destroze el alma, de sentimiento y dolor »***

Los Tigres del Norte

«Ahí llegó un fulano »

Miguel Antonio Tapia,

Al pueblo llego un fulanó, 2004

CHAPITRE 8 Apports de la comparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier

8.1 Introduction

Le premier objectif de ce chapitre est de présenter les performances au pas de temps journalier des 20 modèles testés précédemment. L'intérêt est de comparer ces résultats à ceux obtenus au pas de temps horaire et d'étudier l'influence du pas de temps sur le classement et les performances des modèles. Il ne faut pas oublier qu'une des hypothèses initiales de ce travail de thèse est que la structure d'un modèle pluie-débit dépend du pas de temps des données utilisées pour son développement, et donc de ses performances à un pas de temps donné. Cette comparaison des performances aux pas de temps horaire et journalier doit nous permettre de discriminer si des structures de modèle sont typiquement adaptées pour un pas de temps, ou au contraire, s'il existe des structures de modèles également efficaces aux deux pas de temps.

Le deuxième objectif de ce chapitre est d'étudier l'amélioration des performances des modèles pluie-débit que l'on peut espérer au pas de temps journalier en valorisant une information de pluie ayant une résolution temporelle plus détaillée (dans notre cas, horaire). D'un point de vue opérationnel, cette stratégie d'amélioration de la simulation du débit au pas de temps journalier peut être intéressante quand des pluies horaires sont disponibles.

8.2 Performances au pas de temps journalier : utilisation de données journalières de pluie et débits (PJQJ)

Dans cette partie, nous présentons les performances des 20 modèles testés, au pas de temps journalier (pluie et débit), sur les 313 bassins versants de la base de données (symbole de ce test : PJQJ). Comme précédemment au pas de temps horaire, les modèles ont été calés en utilisant pour fonction objectif le critère $C_{2M}(Q)$ ou le critère $C_{2M}(RQ)$ et leurs performances ont été évaluées en validation également avec ces deux critères.

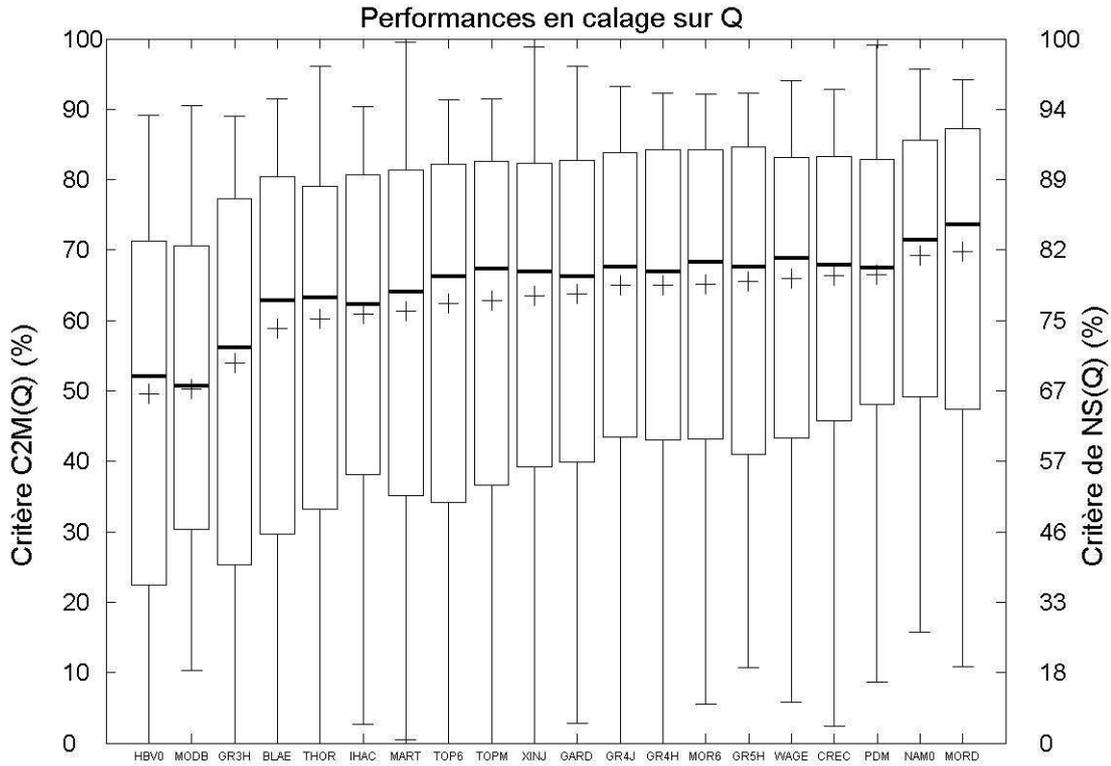


Figure 8-1 : Performances en calage au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$. Classement croissant selon la moyenne (+).

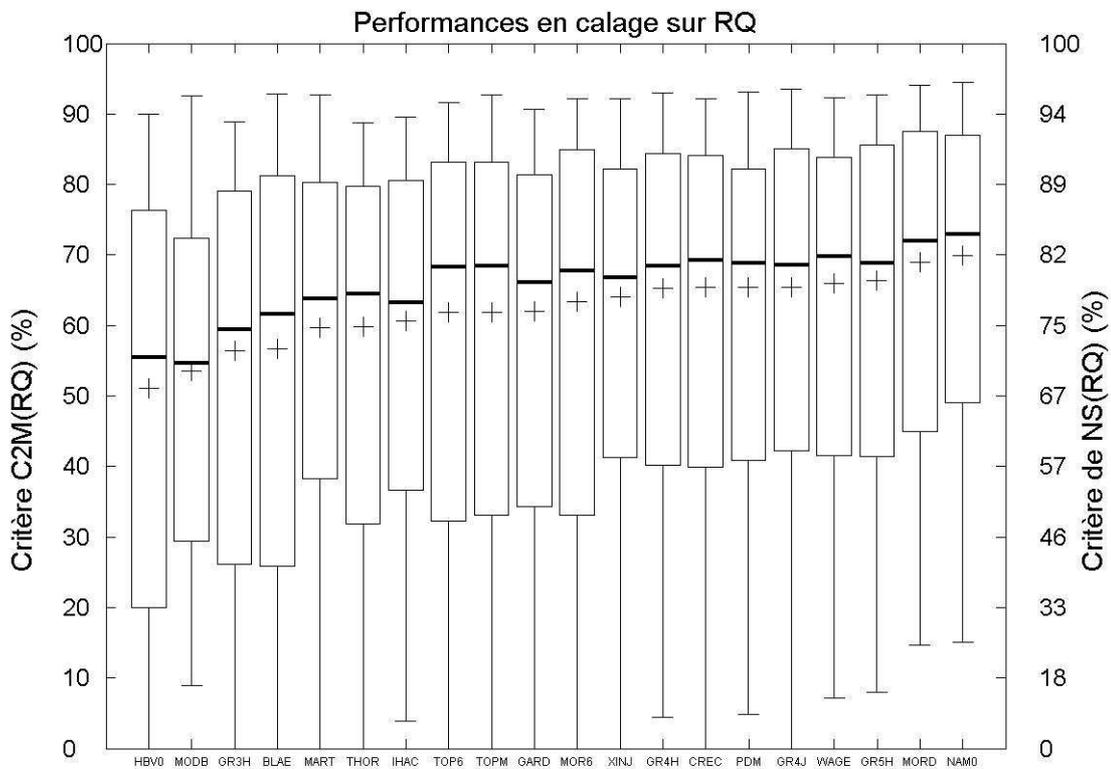


Figure 8-2 : Performances en calage au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$. Classement croissant selon la moyenne (+).

Les performances en calage sont indiquées à la Figure 8-1 ($C_{2M}(Q)$) et à la Figure 8-2 ($C_{2M}(RQ)$). Les niveaux de performance moyenne atteints en calage sont assez élevés et varient entre 50% et 70% de $C_{2M}(Q)$ ou $C_{2M}(RQ)$.

La majorité des modèles obtiennent des performances moyennes très proches, tandis que les performances de NAM0 et MORD sont légèrement supérieures et celles de HBV0, MODB et GR3H sont sensiblement inférieures à celles des autres modèles. La performance moyenne des 20 modèles est légèrement supérieure à 60% de C_{2M} (75% de NS). En moyenne 90% des bassins versants ont un critère de calage supérieur à 36% de C_{2M} (50% de NS) et 10% des bassins versants ont un critère de calage supérieur à 80% de C_{2M} (90% de NS) (Tableau 8-1). Les écarts de performance moyenne et entre les quantiles 0.9 restent limités (inférieurs à 20 points de C_{2M}), alors que les écarts entre quantiles 0.1 sont plus importants. Ceci indique que les modèles ont tous la capacité d'obtenir de très bonnes performances sur une partie des bassins versants, mais que certains souffrent de problèmes de calage sur quelques bassins versants (Tableau 8-1).

Tableau 8-1 : Moyenne de la performance moyenne et des quantiles 0.1 et 0.9, et écarts maximaux entre la moyenne et les quantiles 0.1 et 0.9 des 20 modèles testés.

	Moyenne			Écarts maximaux		
	Quantile 0.1	Moyenne	Quantile 0.9	Quantile 0.1	Moyenne	Quantile 0.9
Calage $C_{2M}(Q)$	38	62	81	21	17	15
Calage $C_{2M}(RQ)$	36	62	82	29	18	13
Calage $C_{2M}(Q)$ / Validation $C_{2M}(Q)$	-6	36	68	45	18	19
Calage $C_{2M}(Q)$ / Validation $C_{2M}(RQ)$	-8	36	69	35	19	21
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(Q)$	-2	37	68	30	17	20
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(RQ)$	4	44	73	21	20	19

Les deux modèles les plus complexes (9 et 10 paramètres) ont les meilleures performances en calage, avec presque 4 points d'avance sur les autres modèles (Figure 8-3). Par contre, pour les autres modèles, il n'y pas une classification claire des performances des modèles en fonction de leur degré de complexité. Des modèles à seulement quatre paramètres obtiennent d'aussi bonnes performances que d'autres modèles à six et huit paramètres. Au cours du calage, la plupart des

modèles ont la même capacité à simuler les débits, quels que soient leurs niveaux de complexité.

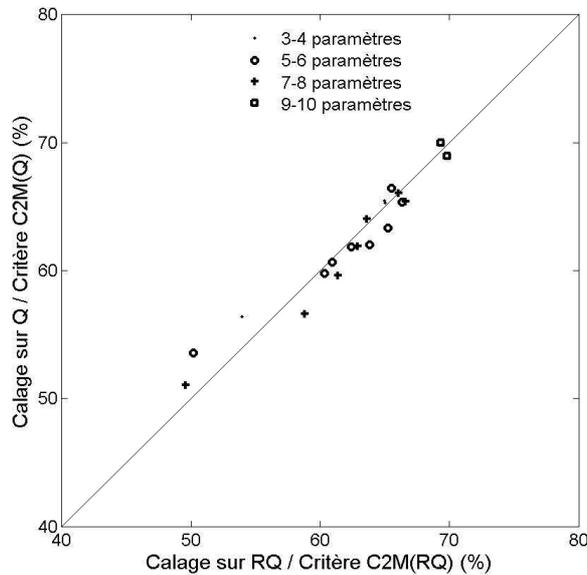


Figure 8-3 : Comparaison des performances en calage, en fonction de la fonction objectif choisie, $C_{2M}(Q)$ ou $C_{2M}(RQ)$, et de la complexité des modèles.

En validation, nous avons évalué les performances des modèles avec les deux critères utilisés précédemment, en fonction des deux critères de calage utilisés :

- Q/Q : calage $C_{2M}(Q)$ / validation $C_{2M}(Q)$, Figure 8-4
- Q/RQ : calage $C_{2M}(Q)$ / validation $C_{2M}(RQ)$, Figure 8-5,
- RQ/Q : calage $C_{2M}(RQ)$ / validation $C_{2M}(Q)$, Figure 8-6,
- RQ/RQ : calage $C_{2M}(RQ)$ / validation $C_{2M}(RQ)$, Figure 8-7.

En validation, les performances des modèles chutent d'environ 20 à 30 points de C_{2M} , les écarts entre les performances moyennes, entre les quantiles 0.1 ou 0.9 des différents modèles augmentent (Tableau 8-1). Les plus fortes baisses de performances et de robustesse sont pour les cas Q/Q, Q/RQ, RQ/Q. Les modèles sont légèrement plus performants et robustes dans le cas RQ/RQ. Par rapport à la phase de calage, le classement des modèles est sensiblement modifié. Il y a notamment le cas du modèle NAM0 qui est très peu robuste et dont les performances chutent fortement. Par contre, les modèles de la famille GR sont assez robustes et voient leurs classements améliorés. De ces tests, il en ressort quelques modèles de tête : les modèles de la famille GR (GR4J, GR4H, GR5H) , les modèles de la famille MORDOR (MORD, MOR6), WAGE, IHAC et CREC. A ces modèles, on peut éventuellement ajouter le modèle THOR (plutôt pour Q) et le modèle PDM (plutôt pour RQ). Par rapport au pas de temps horaire, le modèle MORD est moins performant au pas de temps journalier.

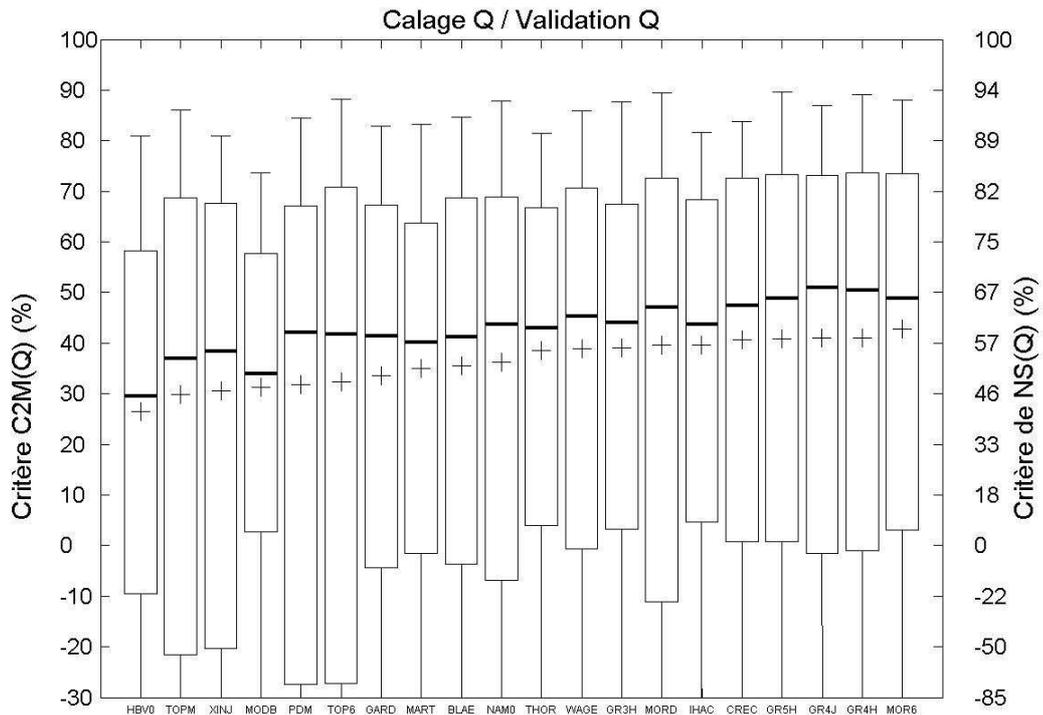


Figure 8-4 : Performances en validation au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(Q)$. Classement croissant selon la moyenne (+).

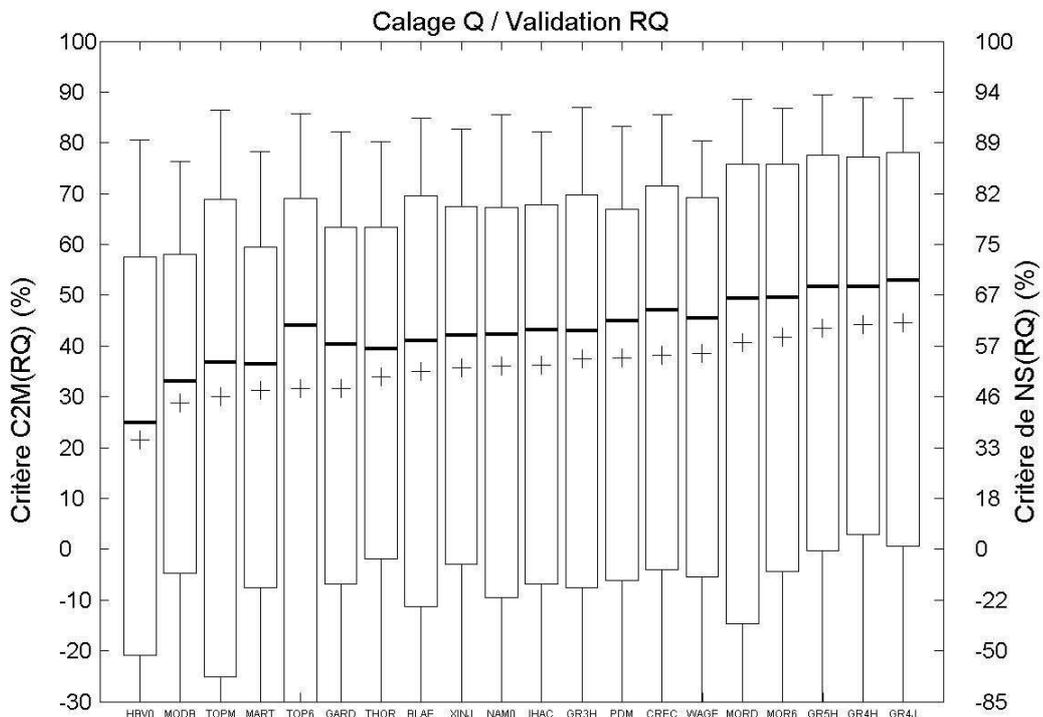


Figure 8-5 : Performances en validation au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(RQ)$. Classement croissant selon la moyenne (+).

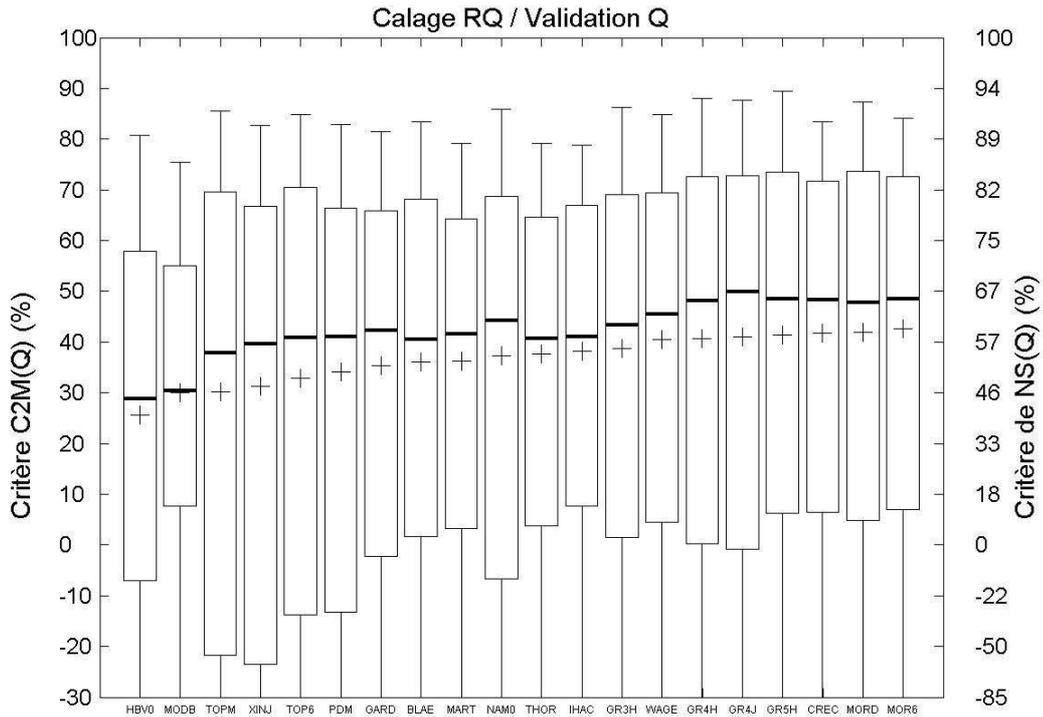


Figure 8-6 : Performances en validation au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(Q)$. Classement croissant selon le moyenne (+).

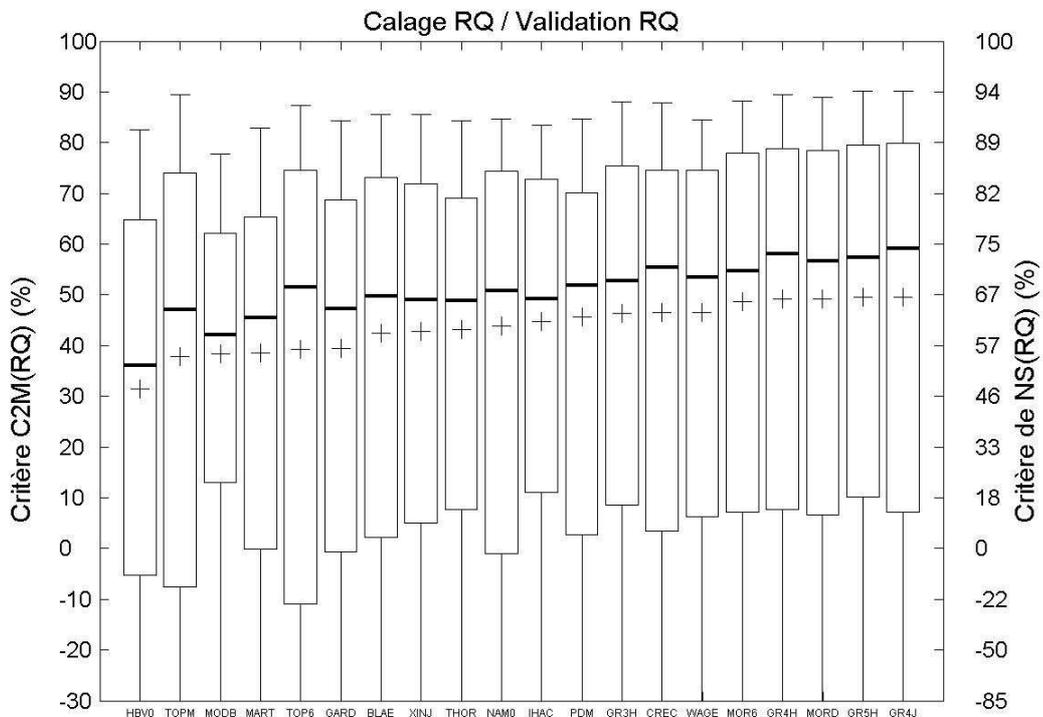


Figure 8-7 : Performances en validation au pas de temps journalier des 20 structures de modèle testées. Calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(RQ)$. Classement croissant selon le moyenne (+).

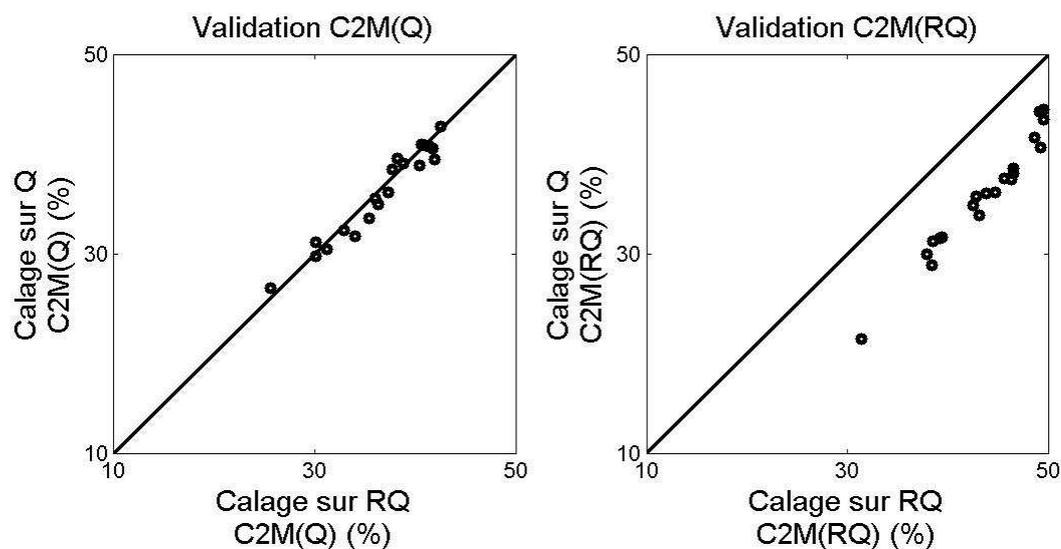


Figure 8-8 : Influence de la fonction objectif de calage sur les performances au pas de temps journalier en validation pour les 20 modèles testés. Fonctions objectifs : $C_{2M}(Q)$ et $C_{2M}(RQ)$.

La Figure 8-8 présente l'influence de la fonction objectif utilisée pour caler un modèle, sur ses performances en validation. Les résultats sont assez proches de ceux obtenus au pas de temps horaire. Cette figure montre très clairement que, si on s'intéresse au critère $C_{2M}(RQ)$, il faut utiliser cette fonction objectif pour le calage. Par contre, pour le critère $C_{2M}(Q)$, l'influence de la fonction objectif utilisée au cours du calage est assez limitée et les différences de performances sont assez faibles. Il semble presque qu'il vaut mieux utiliser la fonction objectif $C_{2M}(RQ)$. On retrouve ici le même résultat que celui obtenu par Perrin [2000] sur 429 bassins versants, testés au pas de temps journalier.

La Figure 8-9 présente les performances en calage, en validation et les différences de performances entre calage et validation, pour les 20 modèles testés, en fonction de leur degré de complexité (nombre de paramètres). La Figure 8-9 a) confirme la supériorité en calage des deux modèles les plus complexes et le même niveau de performance atteint par des modèles ayant de 4 à 8 paramètres. Par contre, la Figure 8-9 b) met en évidence que certains modèles ayant de 4 à 10 paramètres peuvent atteindre le même niveau de performances. Comme au pas de temps horaire, des modèles ayant le même niveau de complexité peuvent avoir des niveaux de performances assez différents. Ceci confirme l'importance de la structure des modèles sur le nombre de paramètres, comme l'ont déjà constaté Gan et al. [1997], Ye et al. [1997] et Perrin [2000]. Avec le critère $C_{2M}(Q)$, un modèle à 6 paramètres est légèrement meilleur que d'autres modèles à 4 et 5 paramètres. Par contre avec le critère $C_{2M}(RQ)$, les modèles les plus performants ont 4 et 5 paramètres.

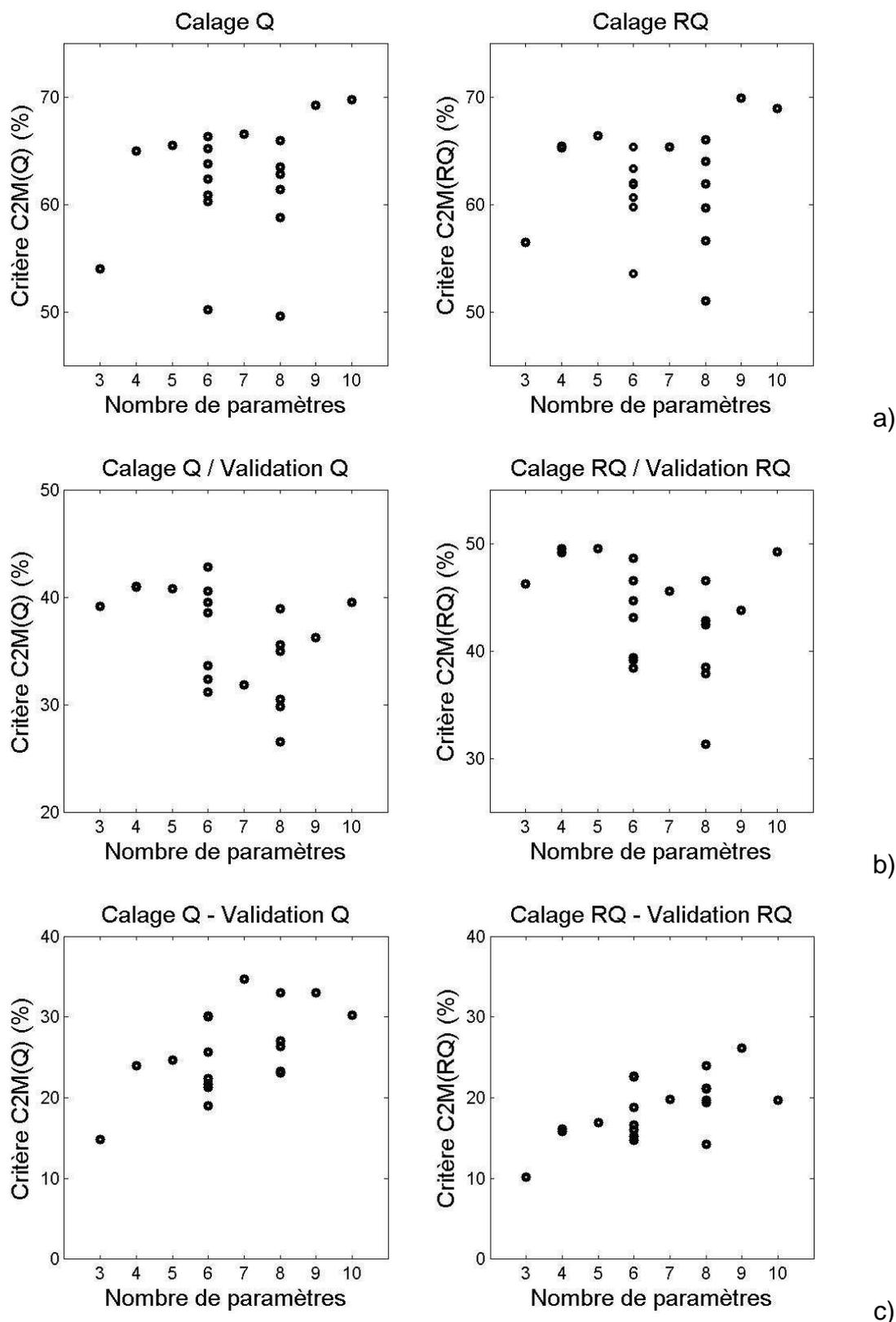


Figure 8-9 : Performances moyennes en calage (a) et en validation (b) et diminution des performances entre calage et validation (c) des 20 structures de modèles testées, en fonction du nombre de paramètres optimisés.

Ainsi, la faible robustesse des modèles peu parcimonieux est clairement mise en évidence. La Figure 8-9 c) illustre cette forte tendance entre perte de performances et nombre de paramètres à optimiser.

Au pas de temps journalier, ces tests confirment les résultats de travaux antérieurs [Beven, 1989; Jakeman et Hornberger, 1993; Perrin et al., 2001a] qui indiquent qu'il suffit de 3 à 6 paramètres pour qu'un modèle identifie correctement la relation pluie-débit, selon une seule fonction objectif. Si l'on compare la Figure 8-9 à la Figure 5-13, qui synthétise les performances des mêmes modèles testés dans exactement les mêmes conditions, mais aux pas de temps horaire et journalier, il semblerait que le niveau de complexité minimal pour correctement représenter la relation pluie débit au pas de temps horaire soit légèrement supérieur. En effet, au pas de temps journalier, la domination des modèles MORDOR et MOR6 est moins forte qu'au pas de temps horaire. Ainsi, ces résultats tendent peut être à prouver qu'il faudrait plutôt 5 à 6 paramètres au minimum pour correctement représenter la relation pluie-débit au pas de temps horaire.

Ces résultats ont une portée assez générale et ils représentent la deuxième comparaison d'un grand nombre de modèles pluie-débit, sur un grand échantillon de bassins versants. Par rapport aux travaux de Perrin [2000], l'échantillon réuni au cours de cette thèse est vraiment très différent. Il y a seulement quelques bassins versants en commun : certains bassins versants des États-unis et quelques bassins versants français (mais, pour ceux là, les périodes sont différentes). Ces échantillons ont été constitués avec la même approche et le cas des données, dites de mauvaise qualité, a été traité de la même manière. La forte différence entre ces deux échantillons de bassins versants réside également dans la superficie, dans la concordance des pas de temps (qui ne débute pas au même instant pour les pluies et débits au pas de temps journalier) et le temps de réponse des bassins versants et dans la durée des chroniques de calage et contrôle qui sont plus longues pour la base de données utilisée par Perrin [2000].

Au pas de temps journalier, les niveaux de performance atteints par les 20 modèles testés au cours de cette thèse sur 313 bassins versants et les niveaux de performance atteints par les 39 modèles testés par Perrin [2000] sur 429 bassins versants sont du même ordre. Au cours du calage, les modèles obtiennent un critère de NS médian d'environ 80%. En validation, les modèles obtiennent un critère de NS(Q) d'environ 60 % et un critère de NS(Q) d'environ 70 %. Ces niveaux de performances sont également atteints par le modèle HBV testé sur 320 bassins versants situés en Autriche [Parajka et al., 2005].

Par contre, sur cet échantillon de 313 bassins versants, certains résultats sont assez intrigants par rapport à ceux obtenus par *Perrin* [2000]. Tout d'abord, il y a les très mauvaises performances (au pas de temps horaire et journalier) de quelques modèles pluie-débit qui donnaient de bons résultats dans les travaux de [*Perrin*, 2000]: c'est le cas notamment de TOPM et HBV0. A l'inverse d'autres modèles, qui n'avaient pas été trop remarqués précédemment, donnent de bons résultats : c'est le cas de IHAC6 et WAGE. Les codes de ces modèles (TOPM, HBV0) ont été vérifiés et l'origine de cette perte de performances reste encore inconnue : est-ce l'effet de l'échantillon, de la durée de calage, etc. ?

8.3 Performances au pas de temps journalier : utilisation de données de pluie horaires et de débits journaliers (PHQJ)

Dans cette partie, nous présentons les performances des 20 modèles que l'on fait fonctionner au pas de temps horaire (pluie et ETP) et dont on agrège les simulations de débit horaire au pas de temps journalier, que l'on compare au débit moyen journalier (symbole de ce test : PHQJ). On évalue ainsi la capacité des modèles à valoriser une information de pluie temporellement plus détaillée (horaire), pour améliorer la simulation du débit moyen journalier. Les modèles ont été calés en utilisant pour fonction objectif le critère $C_{2M}(Q)$ ou le critère $C_{2M}(RQ)$ et leurs performances ont été évaluées en validation également avec ces deux critères.

Les performances en calage sont indiquées à la Figure 8-11 et à la Figure 8-12. Sur ces figures, la performance moyenne obtenue par chaque modèle avec la simulation classique au pas de temps journalier (PJQJ) est indiquée par une étoile. Si l'étoile est située en dessous du symbole '+' (performance moyenne de la simulation PHQJ), cela veut dire que le modèle est capable d'améliorer ses performances au pas de temps journalier, en utilisant une entrée de pluie horaire. Par contre, si l'étoile est située au dessus du symbole '+', cela veut dire que le modèle n'est pas capable d'améliorer ses performances avec une pluie horaire, c'est-à-dire que son pas de temps de prédilection est le pas de temps journalier. Lorsque les modèles sont calés avec le critère $C_{2M}(Q)$, leurs performances en calage (moyenne, quantile 0.9) progressent légèrement (Tableau 8-2). La moitié des modèles voient leurs performances améliorées, alors que pour les autres modèles les performances moyennes restent les mêmes ou diminuent (cas notamment du modèle NAM0). Lorsque les modèles sont calés avec le critère $C_{2M}(RQ)$, leurs performances en calage (moyenne, quantile 0.1) diminuent clairement (Tableau 8-2). Ces résultats montrent l'interaction entre le pas de temps des données, les structures de modèles et la fonction objectif utilisée pour caler les modèles. La principale différence entre le critère $C_{2M}(Q)$ et le critère $C_{2M}(RQ)$, réside dans la plus ou moins grande importance

accordée aux crues. Le critère $C_{2M}(Q)$ accorde logiquement une plus grande importance aux erreurs des modèles sur les crues. Ainsi, il semble logique qu'en calant un modèle avec ce critère, il soit plus facile de voir l'influence d'une entrée de pluie horaire sur la simulation d'un débit journalier. Par contre, avec le critère $C_{2M}(RQ)$, il semble qu'il ne soit pas intéressant d'utiliser une entrée de pluie plus détaillée pour simuler un débit journalier, ce critère étant moins sensible aux crues. En fait, les modèles faisant plus d'erreurs au pas de temps horaire, les gains de performances apportés par une simulation horaire ne deviennent intéressants qu'à partir du moment où on met l'accent sur les crues, qui sont moins bien traduites avec des pluies au pas de temps journalier.

Tableau 8-2 : Moyenne de la performance moyenne et des quantiles 0.1 et 0.9, et écarts maximaux entre la moyenne et les quantiles 0.1 et 0.9 des 20 modèles testés (PHQJ). En italique sont indiqués les résultats du test PJQJ.

	Moyenne			Écarts		
	Quantile 0.1	Moyenne	Quantile 0.9	Quantile 0.1	Moyenne	Quantile 0.9
Calage $C_{2M}(Q)$	35 38	63 62	84 81	25 21	17 17	12 15
Calage $C_{2M}(RQ)$	28 36	60 62	83 82	31 29	18 18	12 13
Calage $C_{2M}(Q)$ / Validation $C_{2M}(Q)$	-3 -6	39 36	72 68	30 45	17 18	15 19
Calage $C_{2M}(Q)$ / Validation $C_{2M}(RQ)$	-10 -8	36 36	69 69	40 35	24 19	20 21
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(Q)$	-1 -2	39 37	70 68	27 30	16 17	15 20
Calage $C_{2M}(RQ)$ / Validation $C_{2M}(RQ)$	2 4	44 44	74 73	26 21	20 20	15 19

Les résultats en validation (Figure 8-12, Figure 8-13, Figure 8-14, Figure 8-15, Tableau 8-2) révèlent les mêmes tendances que celles obtenues au cours du calage : les modèles sont globalement capables de tirer profit d'une entrée de pluie horaire pour simuler des débits journaliers, lorsqu'on utilise la fonction objectif $C_{2M}(Q)$. En effet, avec cette fonction objectif de validation, la performance moyenne des modèles et la valeur des quantiles 0.1 et 0.9 augmentent et les écarts entre ces valeurs diminuent, sur l'ensemble des modèles, quelle que soit la fonction objectif de calage.

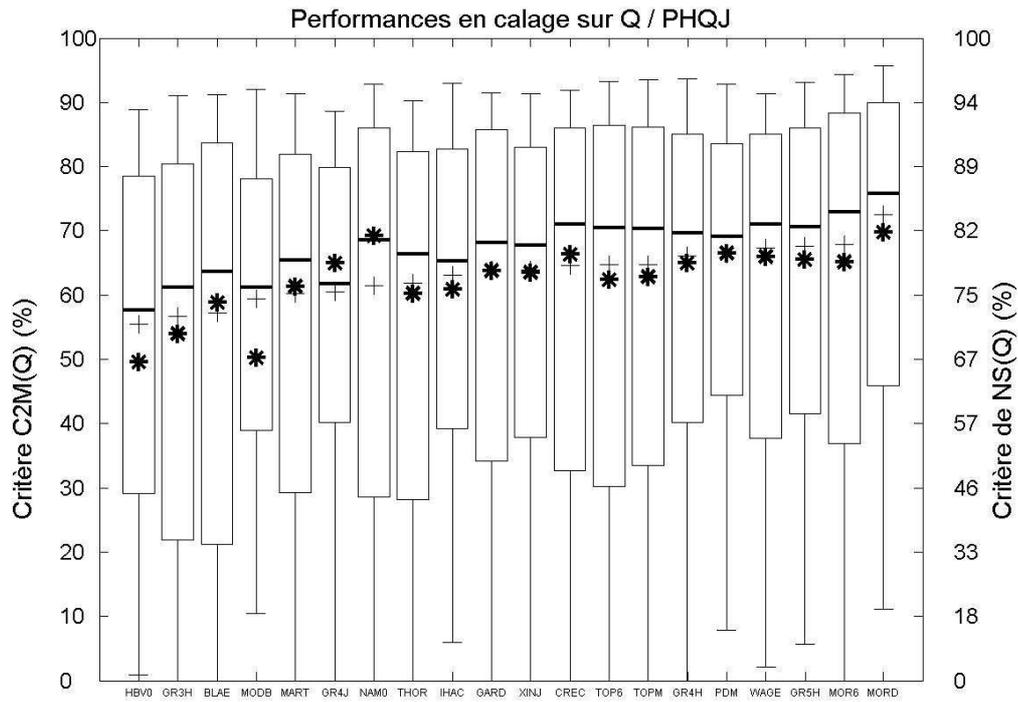


Figure 8-10 : Performances en calage au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(Q)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJQJ).

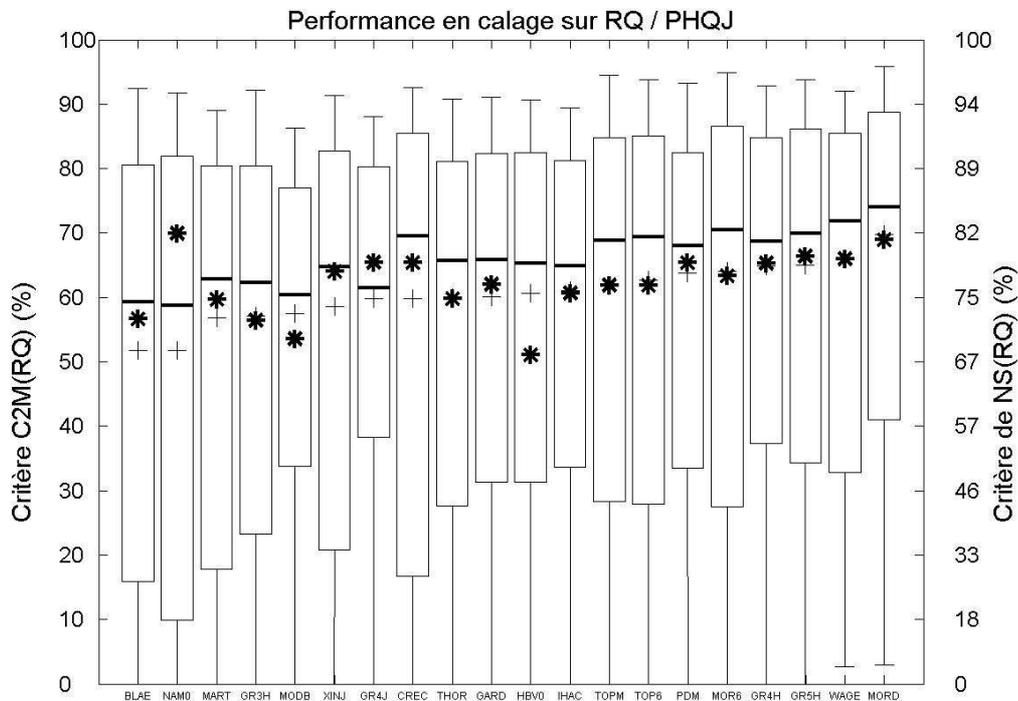


Figure 8-11 : Performances en calage au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(RQ)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJQJ).

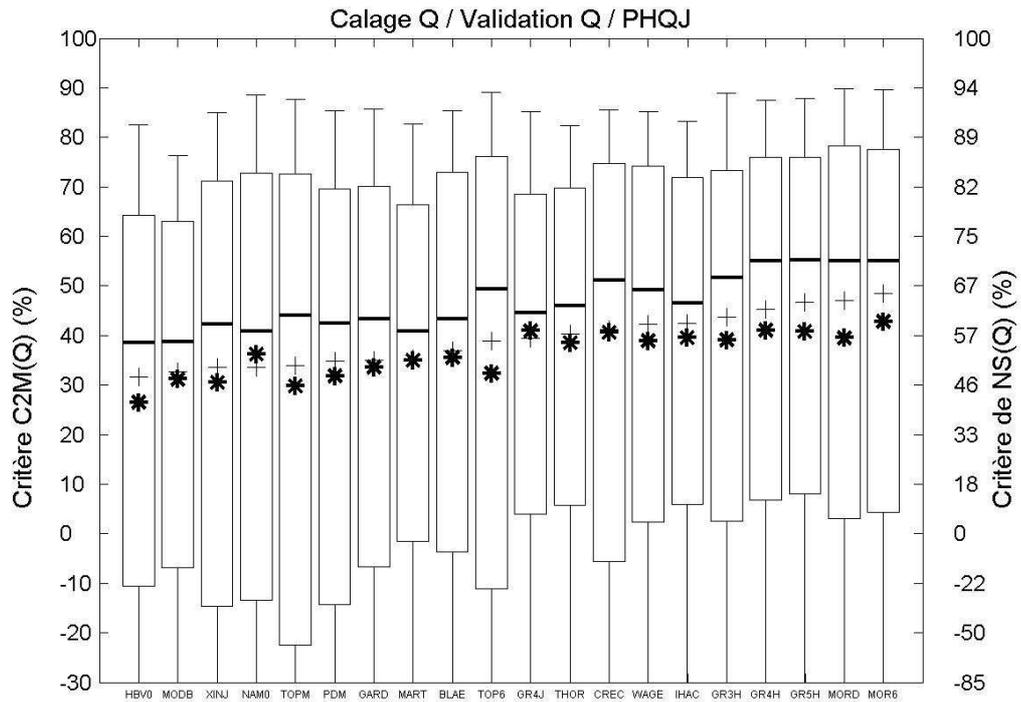


Figure 8-12 : Performances en validation au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(Q)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJQJ).

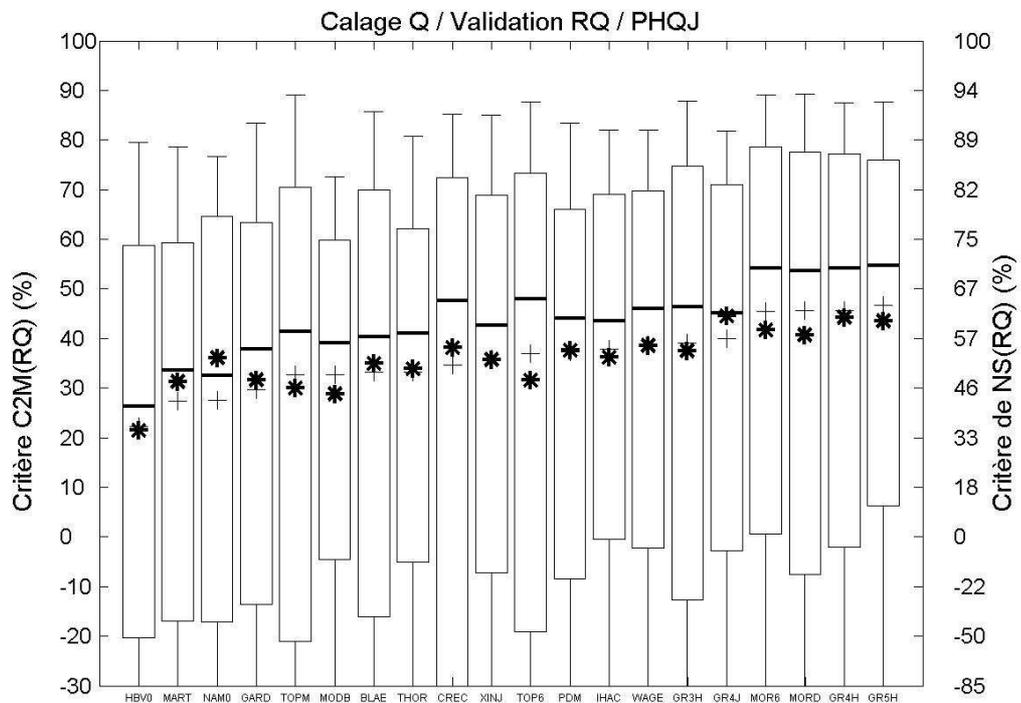


Figure 8-13 : Performances en validation au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(Q)$ et validation sur $C_{2M}(RQ)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJQJ).

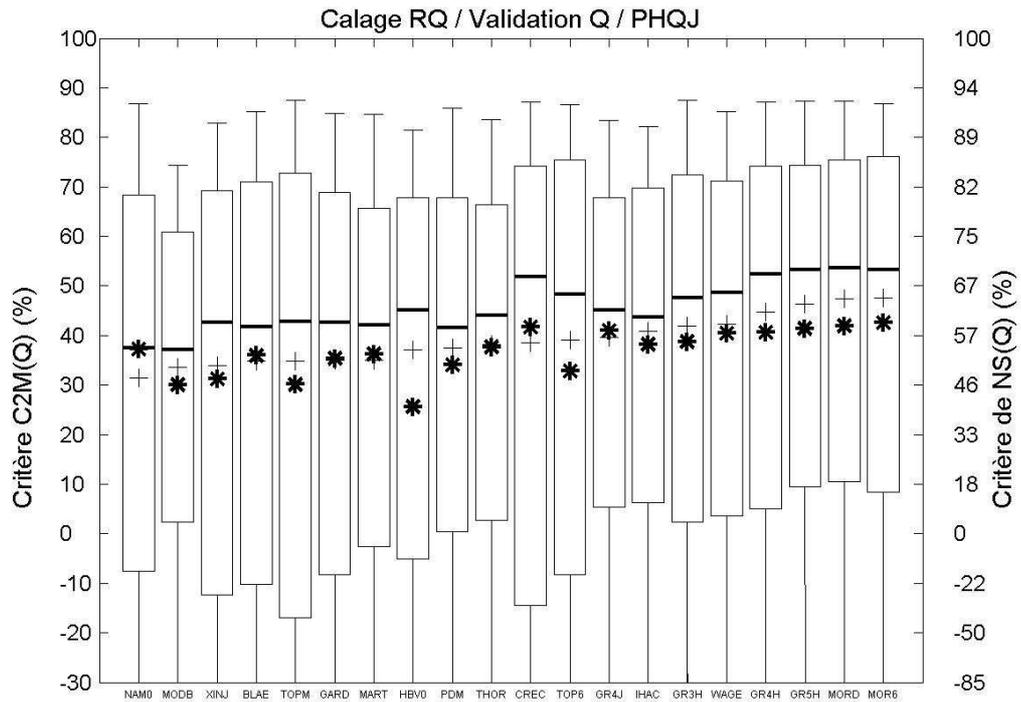


Figure 8-14 : Performances en validation au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(Q)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJJ).

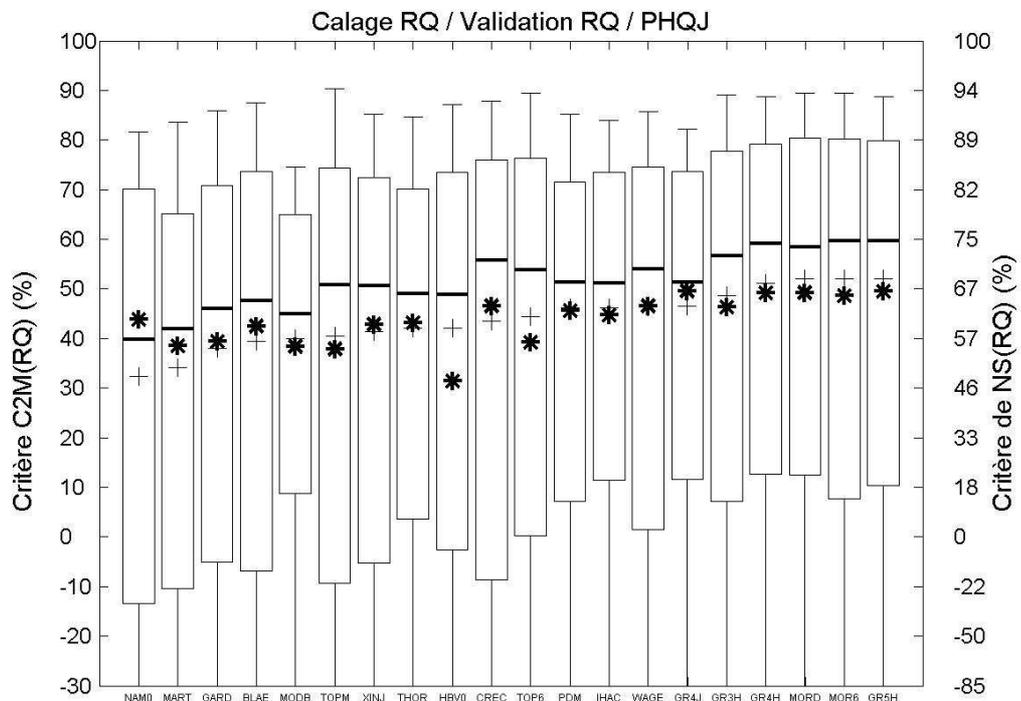


Figure 8-15 : Performances en validation au pas de temps journalier des 20 structures de modèle testées, alimentées par une entrée de pluie horaire (PHQJ), calage sur $C_{2M}(RQ)$ et validation sur $C_{2M}(RQ)$. L'étoile représente la performance moyenne du même modèle, alimenté par une entrée de pluie journalière (PJJ).

Chapitre 8 : Apports de l'intercomparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier

Les modèles qui progressent le plus sont ceux qui donnent de bons résultats au pas de temps horaire : TOP6, IHAC, WAGE, GR3H, GR4H, GR5H et famille MORDOR. Les gains de performance moyenne sur les 313 bassins versants peuvent atteindre pour un modèle +7.4 points de C_{2M} (MORD). Le gain de performance moyenne entre le meilleur modèle du test PJQJ et celui du test PHQJ est de 5.5 points (MOR6), et la meilleure performance moyenne au pas de temps journalier atteint la valeur de 48.4% de $C_{2M}(Q)$. Par contre, les modèles NAM0 et GR4J voient leurs performances diminuées, ce qui indique qu'ils sont mieux adaptés pour le pas de temps journalier.

Lorsqu'on utilise la fonction objectif $C_{2M}(RQ)$, les performances en validation progressent pas ou peu. De manière globale, la valeur moyenne du quantile 0.1 diminue (et les écarts augmentent) et la performance moyenne des 20 modèles et la moyenne du quantile 0.9 ne changent pas. La performance du meilleur modèle du test PJQJ (GR5H) progresse de 2.5 points de $C_{2M}(RQ)$, pour atteindre un $C_{2M}(RQ)$ en validation de 52.1%.

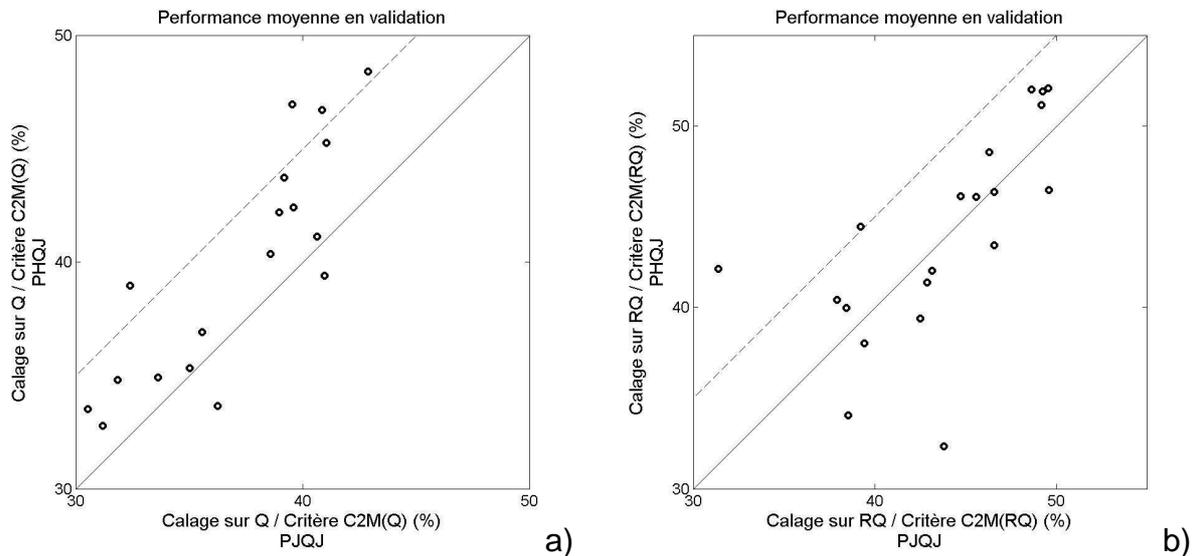


Figure 8-16 : Comparaison des performances obtenues en validation au pas de temps journalier, lorsqu'on alimente le modèle avec des pluies journalière (abscisse) ou des pluies horaires (ordonnée). a) Critère de calage et de validation : $C_{2M}(Q)$, b) Critère de calage et de validation : $C_{2M}(RQ)$. La ligne en pointillé indique un gain de 5 points.

La Figure 8-16 synthétise les résultats de ce test. Cette figure montre que, si on s'intéresse au débit moyen journalier plutôt pendant les crues (calage et validation avec $C_{2M}(Q)$), il est très intéressant d'utiliser une information de pluie horaire, si elle est disponible (Figure 8-16 a). Les gains de performance moyenne des modèles peuvent être assez importants et le meilleur des modèles au pas de temps journalier progresse ainsi de 5 points de C_{2M} (MOR6). Par contre, si on s'intéresse au débit moyen journalier sur l'ensemble de la chronique hydrologique (calage et validation avec $C_{2M}(RQ)$), les résultats montrent qu'il n'est pas forcément nécessaire d'utiliser

une information de pluie détaillée (Figure 8-16 b). En effet, il y a autant de modèles dont les performances progressent ou régressent et les gains de performance pour le meilleur des modèles restent limités : 2.5 points (GR5H). En fait, avec cette fonction objectif, il semble que seuls les bons modèles au pas de temps horaire soient encore capable de tirer profit de cette information de pluie plus détaillée.

8.4 Vers une structure de GR aussi efficace au pas de temps horaire et journalier ?

La comparaison des performances des 20 modèles aux pas de temps horaire et journalier a montré que les structures de la famille GR sont particulièrement efficaces à ces deux pas de temps. Un résultat intéressant de cette comparaison est que la structure de GR4J est clairement moins bonne au pas de temps horaire que la structure de GR4H ou GR5H. Par contre, au pas de temps journalier, les structures de GR4H ou GR5H sont globalement aussi performantes que celle de GR4J. Les différences entre les structures des modèles GR4J et GR4H sont :

- La valeur du coefficient de percolation ($PERC_H=4$, $PERC_J=2.25$),
- La valeur de la puissance de l'hydrogramme unitaire ($CHU_H=1.25$, $CHU_J=2.5$),
- La simplification de la fonction de routage au pas de temps horaire (1 seul hydrogramme à deux branches).

Ainsi, nous avons testé s'il était possible de transférer une structure du pas de temps horaire vers le pas de temps journalier, en adaptant seulement la valeur du coefficient de percolation, en fonction du pas de temps. Pour cela, nous avons repris les deux structures horaires GR4H et GR5H, que nous avons testé au pas de temps journalier en changeant seulement la valeur du coefficient de percolation (PERC). Nous appellerons ces structures indépendantes du pas de temps GR4X1 et GR5X1.

Nous avons comparé les performances des modèles GR3H, GR4H, GR5H, GR4X1, GR5X1 et GR4J, testés au pas de temps journalier sur l'échantillon de 313 bassins versants. Les résultats en calage et en validation sont présentés à la Figure 8-17. Cette figure montre que les performances en calage et en validation des modèles GR4H, GR5H, GR4X1, GR5X1 et GR4J sont très proches, celles de GR3H étant légèrement inférieures. Ce qui est intéressant, c'est que les structures de GR4X1 et GR5X1 ont toujours de meilleures performances en validation que celles de leurs structures d'origine (GR4H et GR5H) et celles du modèle GR4J (Figure 8-17 c), d), e) et f), même si leurs gains de performances sont assez limités (maximum : + 1 points). Les différences de performances entre GR4X1 et GR5X1 restent assez limitées et ne justifient pas le degré de complexité supérieur de la structure de GR5X1.

Chapitre 8 : Apports de l'intercomparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier

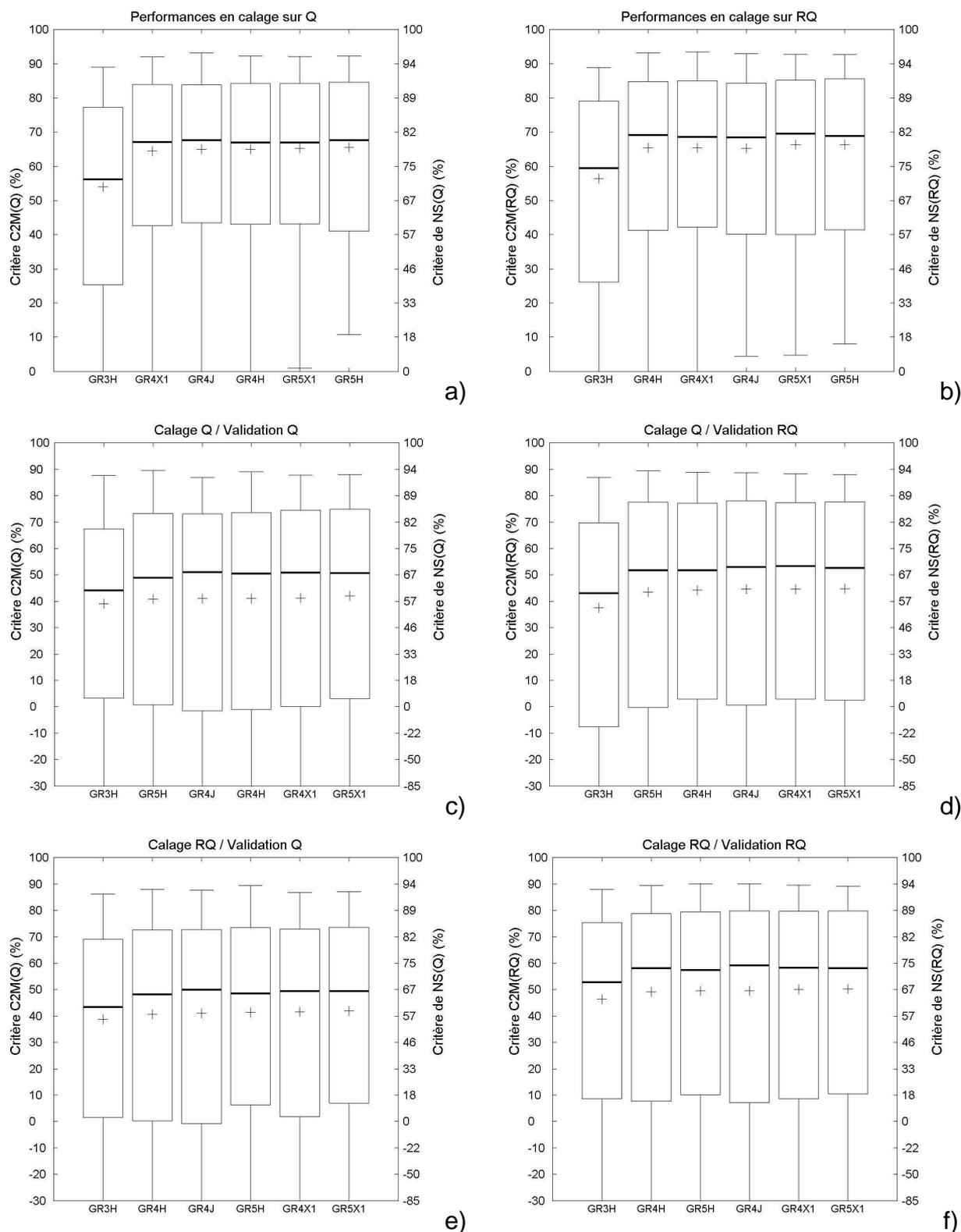


Figure 8-17 : Comparaison des performances en calage et validation de différentes version de modèles de la famille GR, testés au pas de temps journalier, sur l'échantillon de 313 bassins versants. Performance : a) calage sur Q, b) calage sur RQ, c) calage sur Q, validation sur Q, d) calage sur Q, validation sur RQ, e) calage sur RQ, validation sur Q, f) calage sur RQ, validation sur RQ.

Ces tests démontrent ainsi l'intérêt d'étudier les performances de différents modèles à différents pas de temps. Finalement, contrairement à notre hypothèse de départ, les développements effectués au pas de temps horaire nous ont conduits à une structure (GR4X1 ou GR5X1) capable d'obtenir d'aussi bons niveaux de performance aux pas de temps horaire et journalier. Les performances de la structure GR4X1 ont été comparées à celles de GR4J, sur les 429 bassins versants de la thèse de *Perrin* [2000]. Les résultats montrent que les performances de GR4X1 sont au moins aussi bonnes ou meilleures que celles du modèle GR4J.

Ne conserver qu'une structure de modèle, pour les pas de temps horaire et journalier, permet ainsi de gagner en lisibilité du fonctionnement des modèles et peut également faciliter les procédures de transferts de paramètres et d'états du système d'un pas de temps à un autre. Par ailleurs, dans un contexte de prévision opérationnelle des crues, il peut être intéressant d'avoir une même structure de modèle et de faciliter le passage d'un pas de temps journalier à un pas de temps horaire de simulation/prévision, pendant les évènements pluvieux.

8.5 Transfert de paramètres du pas de temps journalier vers le pas de temps horaire

Pour optimiser les paramètres d'un modèle à un certain pas de temps, il faut normalement disposer des chroniques de pluie et de débit à ce pas de temps. Ainsi, pour optimiser les paramètres d'un modèle horaire, il faut des chroniques horaires de pluie et de débit. Parfois, la disponibilité de chroniques de débit ou de pluie au pas de temps horaire peut être limitée, alors que l'on dispose de longues chroniques au pas de temps journalier. Par exemple, aux États-unis, l'USGS, qui gère les stations hydrométriques du pays, est capable de fournir gratuitement les chroniques de débits journaliers. Par contre, les débits horaires sont plus difficiles à obtenir et rarement conservés. Dans d'autres cas, il peut être possible de disposer de longues chroniques au pas de temps journalier, mais de chroniques horaires de débit ou de pluie trop courtes pour pouvoir définir de manière robuste les paramètres d'un modèle au pas de temps horaire.

Néanmoins, pour certaines études, dans le cas de petits bassins versants, il peut être important d'appliquer un modèle au pas de temps horaire. Les chroniques de pluie au pas de temps horaire devenant de plus en plus disponibles, le transfert de paramètres estimés au pas de temps journalier, pour faire tourner un modèle au pas de temps horaire peut devenir très utile. De plus, il est probable que des vecteurs de paramètres transférés du pas de temps journalier vers le pas de temps horaire soient plus robustes que des vecteurs optimisés en utilisant de très courtes chroniques de

données au pas de temps horaire. Dans ce contexte, nous allons tester des stratégies de transfert de paramètres du pas de temps journalier vers l'horaire.

8.5.1 Comparaison des vecteurs de paramètres optimisés à différents pas de temps

La Figure 8-18 présente la relation entre les paramètres du modèle GR4X1, optimisés sur les mêmes périodes, aux pas de temps horaire et journalier.

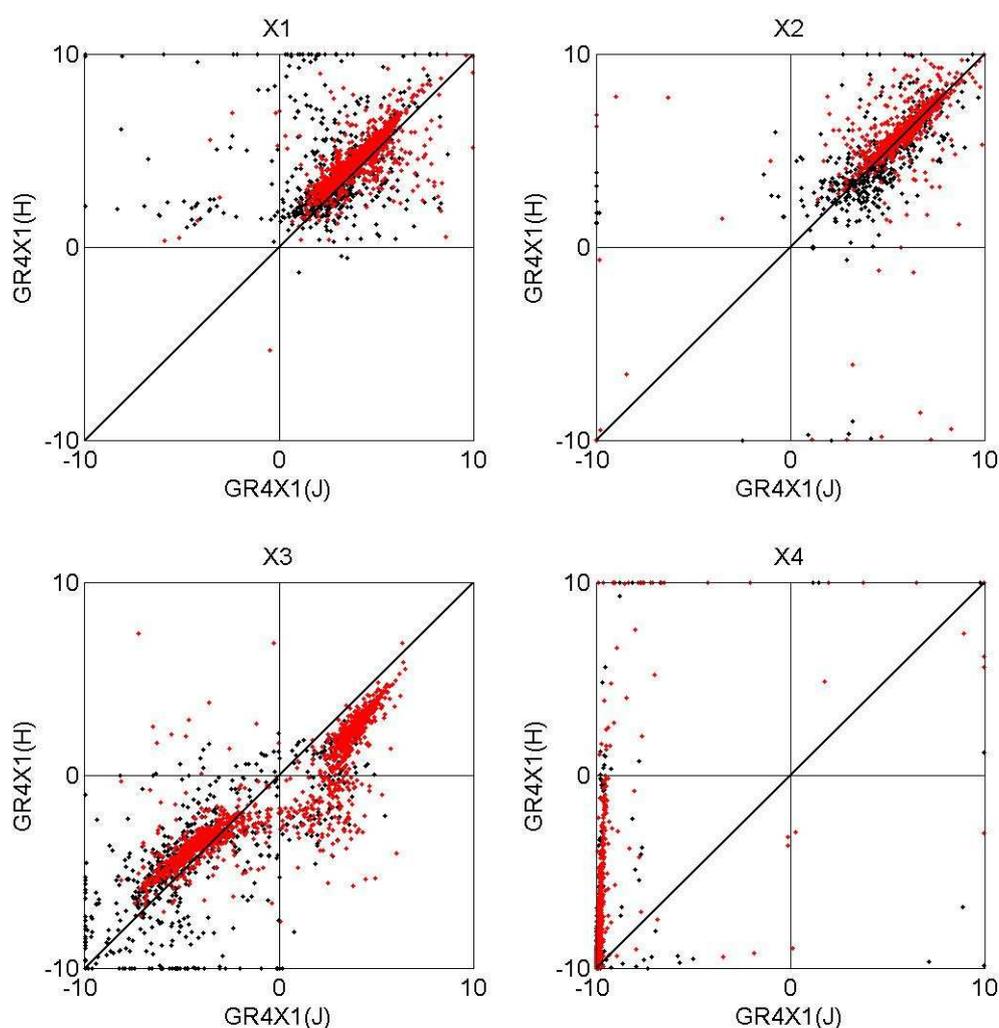


Figure 8-18 : Relations entre les paramètres transformés du modèle GR4X1, estimés au pas de temps journalier et horaire. En rouge, bassins versants français ; en noir, bassins versants étrangers.

Sur cette figure ont été distingués les paramètres des bassins versants français (en rouge) et étrangers (en noirs). Les paramètres des bassins versants secs étant plus difficiles à optimiser, les valeurs de leurs paramètres sont souvent fortement dispersées. La Figure 8-18 montre une relativement bonne relation entre les paramètres X_1 et X_2 , aux pas de temps horaire et journalier. Les valeurs de ces paramètres peuvent être assez dispersées, mais elles sont situées autour de la première bissectrice. Le paramètre X_4 présente également une très bonne relation

linéaire aux pas de temps horaire et journalier. Les valeurs dispersées sont celles des périodes où les modèles sont particulièrement mauvais. Enfin, le paramètre X_3 présente un comportement un peu plus difficile à appréhender. Pour ses valeurs positives, la valeur journalière est légèrement supérieure à la valeur horaire, alors que c'est le contraire pour ses valeurs négatives. De plus, il présente également une assez forte dispersion. Le paramètre X_3 est le paramètre de la fonction d'échanges souterrains, dont le fonctionnement et le rôle dans les modèles de la famille GR, à travers différents pas de temps, est encore mal cerné [Le Moine, 2005].

La Figure 9-58 montre une très bonne relation entre les paramètres d'un modèle fonctionnant au pas de temps horaire, mais dont les valeurs ont été optimisées en fonction d'un débit objectif horaire ou journalier. La majeure partie de la dispersion vient des paramètres des bassins versants étrangers. Il semble que le modèle utilise principalement le paramètre X_4 pour s'adapter au pas de temps du débit objectif.

8.5.2 Première stratégie de transfert des paramètres : régressions linéaires

La première stratégie de transfert de paramètres testée est particulièrement simple, puisque nous avons calé une relation linéaire entre les paramètres optimisés au pas de temps journalier et horaire. Nous sommes partis des relations linéaires $X(H) = a.X(J) + b$ ayant le meilleur coefficient de détermination. Ensuite, à partir de ces relations, pour chaque paramètre du modèle (X_1, X_2, X_3, X_4), nous avons essayé de chercher les paramètres ($a_i, b_i / X_i$), situés autour des valeurs données par la régression linéaire, donnant les meilleures performances au pas de temps horaire lorsqu'on détermine les paramètres d'un bassin versant à partir de ses paramètres journaliers et de ces relations. Cette démarche empirique représente une cinquantaine de couples ($a_i, b_i / X_i$) testés pour les quatre paramètres du modèle GR4X1. Ces essais ont permis de faire progresser sensiblement les performances du modèle au pas de temps horaire, ce qui signifie que ce ne sont pas les meilleures relations linéaires entre paramètres journaliers et horaires qui donnent les meilleures simulation horaires.

Les relations que nous avons déterminées à partir des valeurs transformées des paramètres sont les suivantes :

$$X_1(H) = X_1(J) * 0.75 + 1.55 \quad [8.1]$$

$$X_2(H) = X_2(J) * 0.6 + 2.8 \quad [8.2]$$

$$X_3(H) = X_3(J) * 0.69 - 1.15 \quad [8.3]$$

$$X_4(H) = X_4(J) * 18.5 + 175.0 \quad [8.4]$$

8.5.3 Deuxième stratégie de transfert des paramètres : transfert de régions

La deuxième stratégie de transferts de paramètres que nous avons testée est inspirée de la méthode d'optimisation par pré-filtrage de l'espace des paramètres. L'idée de cette méthode est assez simple :

1. identifier la région de l'espace dans laquelle se situe chaque vecteur optimisé au pas de temps journalier pour un bassin versant,
2. identifier la région de l'espace dans laquelle se situe chaque vecteur optimisé au pas de temps horaire, pour un bassin correspondant,
3. calculer la matrice de transition qui permet de passer d'une région de l'espace identifiée au pas de temps journalier, vers une région de l'espace identifiée au pas de temps horaire. La matrice de transition nous fournit la probabilité de passer de la région i au pas de temps journalier, vers les régions j, k, l , etc. au pas de temps horaire, avec les probabilités associées $(P_{ij}, P_{ik}, P_{il}, etc.)$.

Pour calculer la matrice de transition (Figure 8-19), nous nous sommes basés soit sur la distribution des 313 vecteurs de paramètres obtenus en calant les vecteurs sur l'ensemble de la période, soit sur la distribution des 2093 vecteurs de paramètres obtenus en calant les vecteurs de paramètres sur des périodes d'une année. Ensuite, nous avons identifié les régions de l'espace en retenant trois ou cinq divisions par vecteurs, ce qui revient à 81 ou 625 régions de l'espace.

Ainsi, chaque bassin versant est caractérisé par :

- la région de l'espace occupée par le vecteur de paramètres obtenu au pas de temps journalier, lorsqu'on effectue un calage sur toute la période de données disponible,
- les régions de l'espace occupées par les n vecteurs de paramètres obtenus au pas de temps journalier, lorsqu'on effectue un calage sur n périodes de 1 an.

La matrice de transition fait le lien entre la région d'un vecteur optimisé au pas de temps journalier et ses régions correspondantes au pas de temps horaire. Pour faire une simulation horaire pour un bassin versant, il est possible d'utiliser cette information de deux manières :

- en créant un nouveau vecteur de paramètres horaires, recomposé à partir de la somme des vecteurs représentant chaque région correspondante, pondérée par chaque probabilité de transition d'une région journalière vers une région horaire (approche 'vecteur composite') ;

Chapitre 8 : Apports de l'intercomparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier

- en faisant tourner un modèle par vecteurs représentant chaque région correspondante et en créant un débit composite, qui est la somme de tous les débits simulés par les différents vecteurs de paramètres, pondérés par chaque probabilité de transition d'une région journalière vers une région horaire (approche 'multi-modèle').

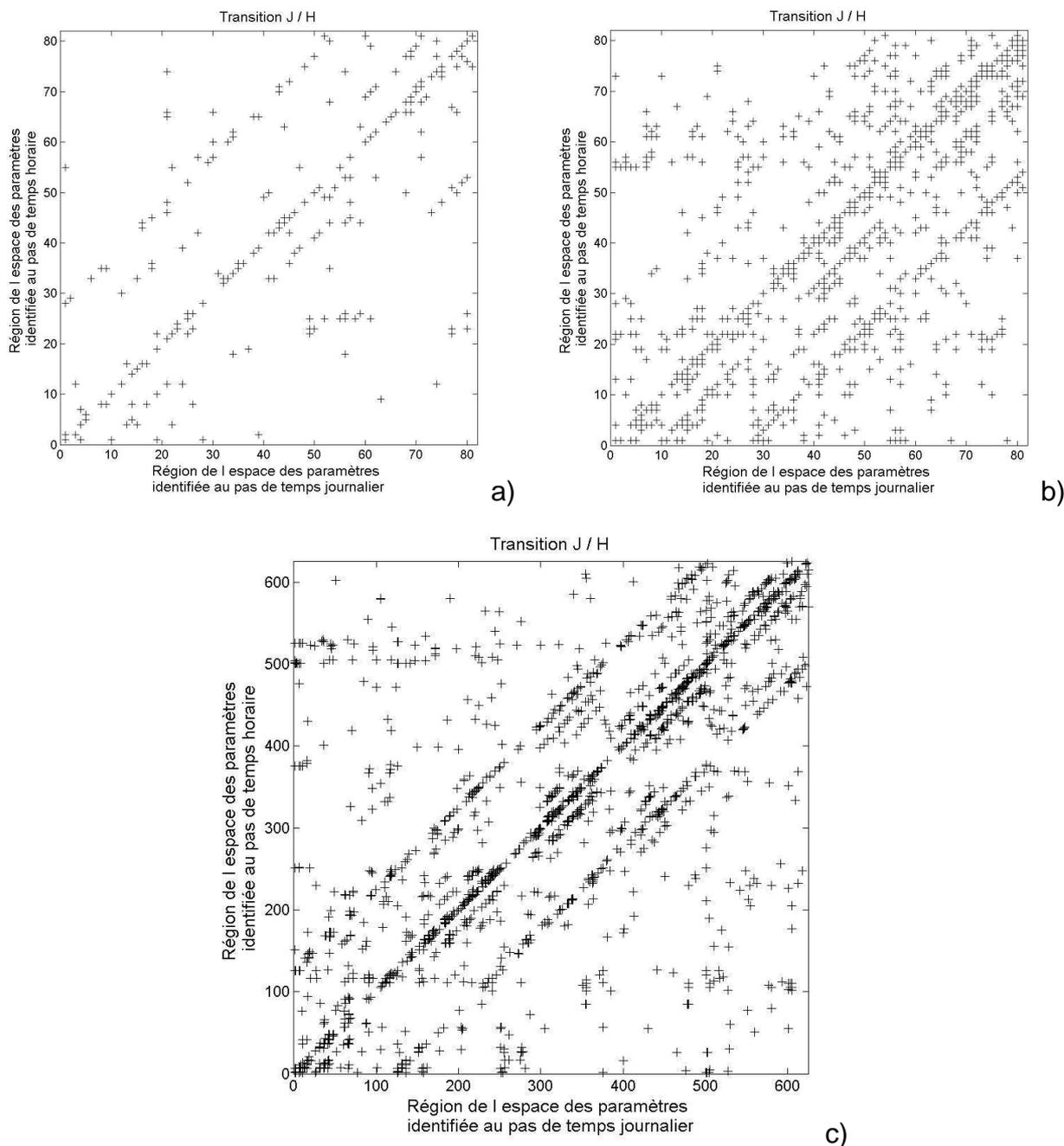


Figure 8-19 : Correspondances entre les régions de l'espace des paramètres identifiées aux pas de temps journalier et horaire (modèle GR4X1). a) Correspondances obtenues avec 313 vecteurs optimisés, b) correspondances obtenues avec 2093 vecteurs optimisés et trois régions, c) correspondances obtenues avec 2093 vecteurs optimisés et cinq régions.

Chapitre 8 : Apports de l'intercomparaison des performances et des paramètres des modèles à différents pas de temps : cas des pas de temps horaire et journalier

Tableau 8-3 : Synthèse des différents types de transferts de paramètres effectués pour déterminer les paramètres d'un modèle horaire, à partir des paramètres optimisés au pas de temps journalier.

Information initiale	Approche 'vecteur composite'	Approche 'multi-modèle'
313 vecteurs de paramètres calés au pas de temps journalier sur l'ensemble de la chronique disponible. 1 région identifiée par bassin versant.	TRSF313	TRSF313 m
2093 vecteurs de paramètres calés au pas de temps journalier sur des périodes de 1 an. Distributions des paramètres divisés en 3 parties : 81 régions. n régions identifiées par bassin versant (n : nombre de périodes de 1 an).	TRSF3R	TRSF3R m
2093 vecteurs de paramètres calés au pas de temps journalier sur des périodes de 1 an. Distributions des paramètres divisés en 5 parties : 625 régions. n régions identifiées par bassin versant.	TRSF5R	TRSF5R m

8.5.4 Synthèse des performances des différentes stratégies de transfert des vecteurs de paramètre

Performances en calage sur toute la période :

Nous avons testé l'approche basée sur des régressions linéaires entre paramètres optimisés au pas de temps journalier et horaire et les six approches de transferts de vecteurs de paramètres (Figure 8-20). La performance moyenne du modèle GR4X1, lorsqu'on le cale au pas de temps horaire sur l'ensemble de la chronique disponible de chaque bassin versant, est de 58.7 points de $C_{2M}(RQ)$. Les performances moyennes des sept approches varient entre 38 et 45 points de $C_{2M}(RQ)$. Ces pertes de performances représentent une baisse de 22% à 35% du critère $C_{2M}(RQ)$. Les approches 'multi-modèle' sont meilleures que les approches de 'vecteur composite'. Les deux meilleures approches sont l'approche 'multi-modèle' (TRSF5R m) basée sur 2093 vecteurs de paramètres et cinq régions et les régressions linéaires. Ces deux meilleures approches ont des performances moyennes assez proches, mais l'approche par régression linéaire est plus robuste que l'approche 'multi-modèle'. Ces résultats étant assez proches, l'approche 'multi-modèle' ne semble pas justifiée, vues sa complexité et sa lourdeur.

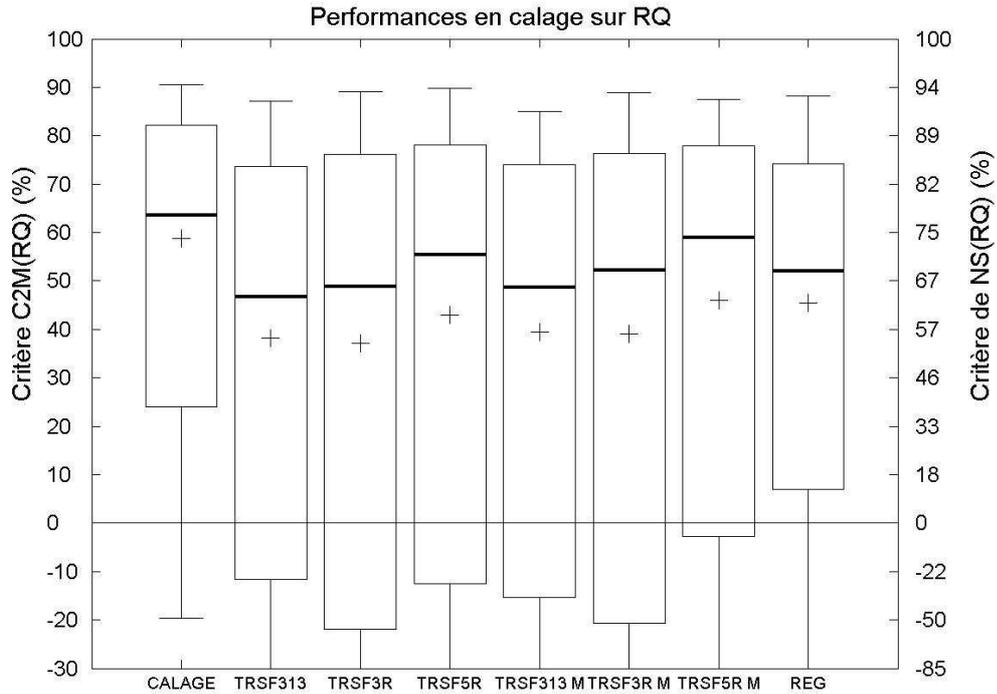


Figure 8-20 : Performances moyennes ($C_{2M}(RQ)$) obtenues en calage au pas de temps horaire et par les différentes approches d'estimation des vecteurs de paramètres horaire, à partir des vecteurs optimisés au pas de temps journalier.

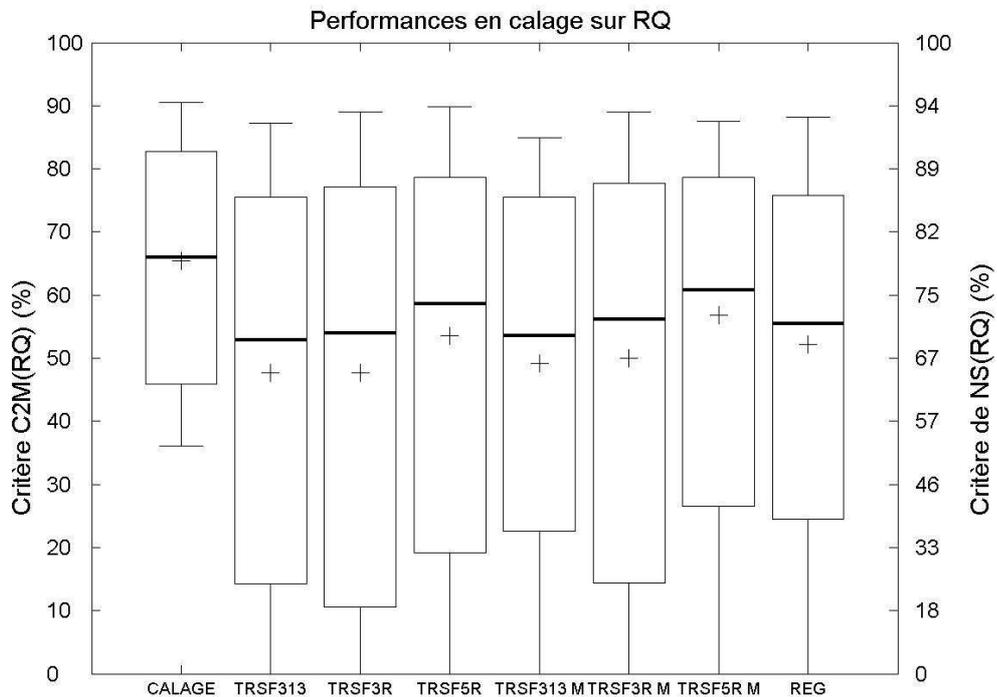


Figure 8-21 : Performances moyennes ($C_{2M}(RQ)$) obtenues en calage au pas de temps horaire et par les différentes approches d'estimation des vecteurs de paramètres horaires, à partir des vecteurs optimisés au pas de temps journalier. Bassins versants dont les performances au pas de temps journalier sont telles que $C_{2M}(RQ) > 35\%$ ($NS > 50\%$).

Par contre, si on ne retient que les bassins versants dont le niveau de performance est acceptable (critère de NS > 50% au pas de temps journalier ; 87% de l'échantillon initial), les résultats sont sensiblement différents (Figure 8-21). Le recours aux bassins versants dont le niveau de performance est acceptable semble justifié dans ce cas, car nous comparons deux méthodes de transferts de paramètres. Ainsi, si le vecteur initial de paramètres donne de mauvaises performances, ce n'est pas forcément la méthode de transfert de paramètres qui est mauvaise, mais l'information initiale. Dans ce cas, la performance moyenne en calage au pas de temps horaire est de 65% de $C_{2M}(RQ)$ et les sept approches testées ont leurs performances moyennes qui varient entre 47% et 56% de $C_{2M}(RQ)$. Cela représente une diminution de la perte de performance, qui n'est plus qu'entre 14% et 28%. Ces bassins versants, dont les performances étaient très mauvaises, étaient clairement à l'origine de la faible robustesse des méthodes de transfert, ce qui masquait légèrement les performances de ces méthodes. Dans ce contexte, les meilleures méthodes sont l'approche de vecteur composite (TRSF5R) et l'approche 'multi-modèle' (TRSF5R m). L'approche 'multi-modèle' devient clairement meilleure et plus robuste que la méthode basée sur les régressions linéaires.

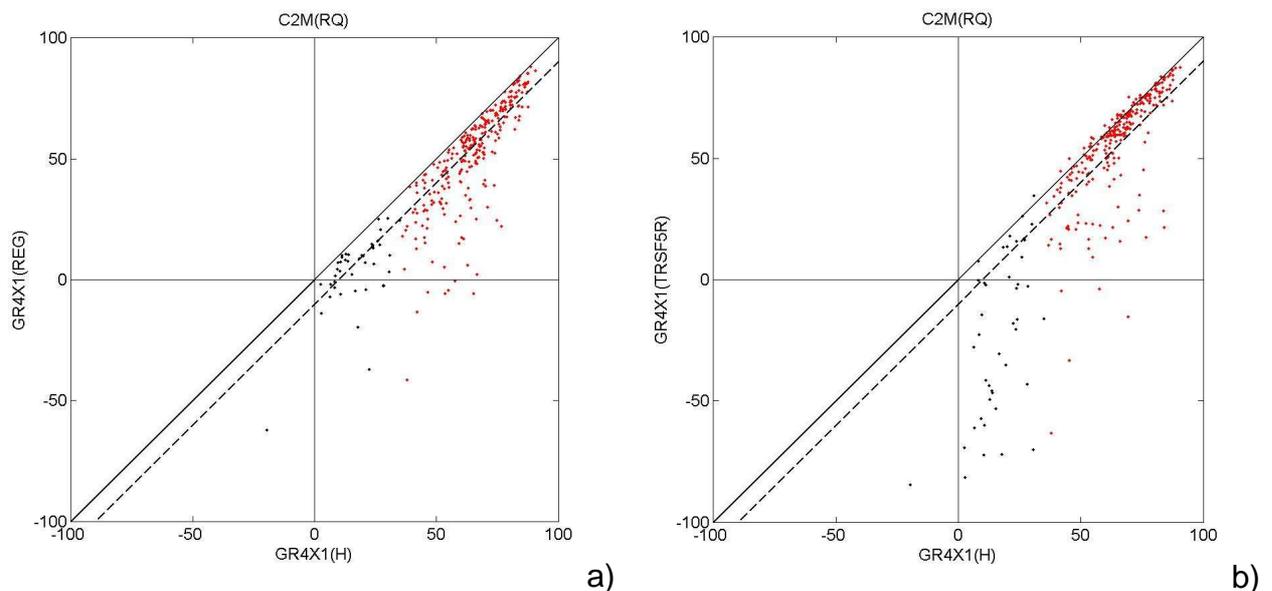


Figure 8-22 : Performances obtenues au pas de temps horaire sur les 313 bassins versants (Modèle GR4X1). a) Comparaison entre les performances obtenues avec des vecteurs calés et des vecteurs déterminés par les régressions linéaires ; b) Comparaison entre les performances obtenues avec des vecteurs calés et des vecteurs déterminés par l'approche de vecteur composite (TRSF5R). En rouge et en noir : paramètres des bassins versants dont les performances sont respectivement supérieures et inférieures à 35 points de $C_{2M}(RQ)$, Les pointillés indiquent la ligne -10 par rapport à la bissectrice.

La Figure 8-22 montre la comparaison des performances obtenues par (a) la méthode basée sur des régressions linéaires et (b) l'approche 'multi-modèle'

(TRSF5). Cette figure montre la différence de résultats de ces deux méthodes. La méthode basée sur des régressions linéaires est particulièrement robuste et les pertes de performances sur les bassins versants pour lesquels elle est mauvaise restent limitées. Par contre, la différence de performances avec les vecteurs calés est légèrement supérieure à 10 points de $C_{2M}(RQ)$.

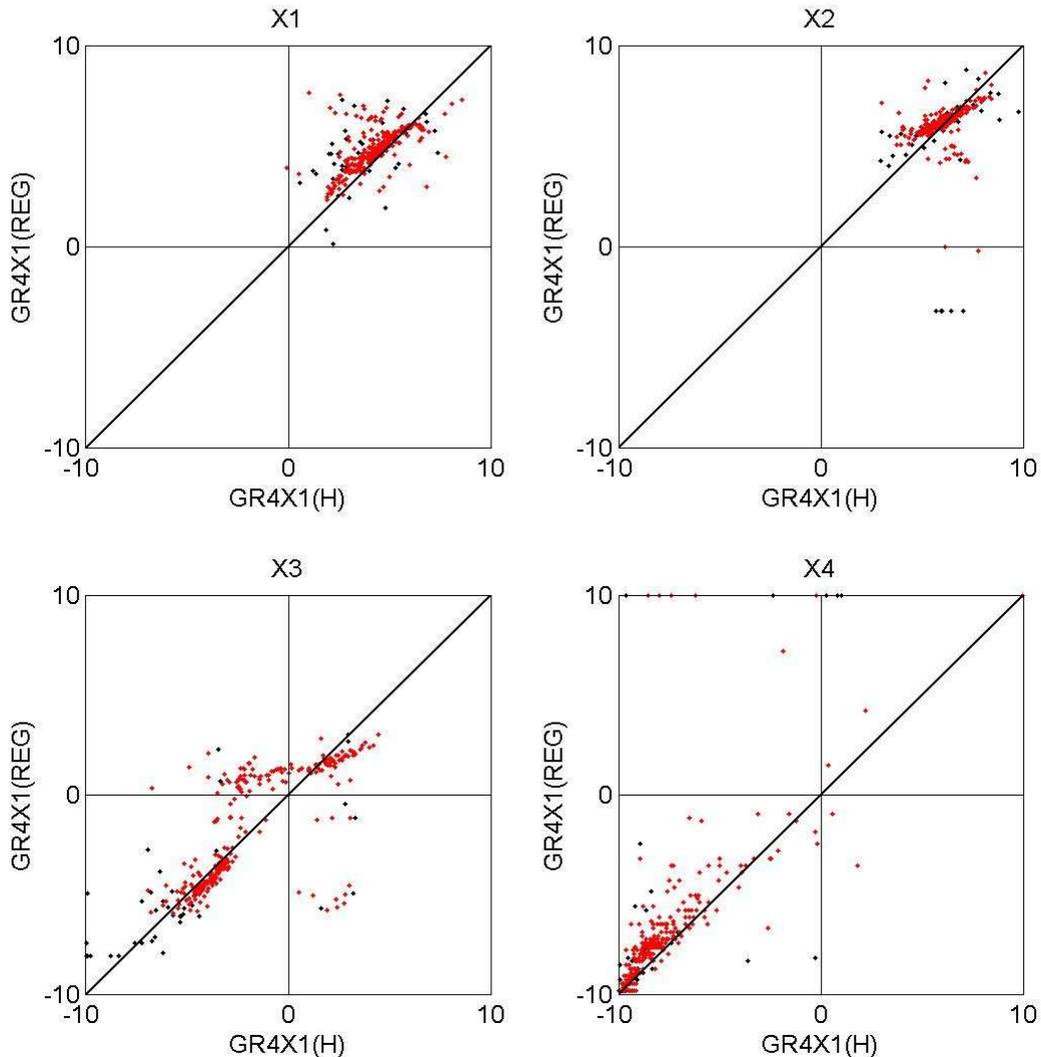


Figure 8-23 : Relations entre les paramètres transformés du modèle GR4X1, optimisés au pas de temps horaire et estimés à partir des régressions linéaires. En rouge et en noir : paramètres des bassins versants dont les performances sont respectivement supérieures et inférieures à 35 points de $C_{2M}(RQ)$,

L'approche 'multi-modèle' (TRSF5) est caractérisée par de bonnes performances pour les bassins dont les performances sont déjà bonnes. La différence de performances entre les vecteurs calés et les vecteurs estimés est de 5 à 10 points de $C_{2M}(RQ)$. Par contre, cette méthode est peu robuste, pour estimer les vecteurs des bassins versants qui sont mauvais en calage. En fait, étant donné la complexité de cette approche (nombreux vecteurs de paramètres combinés selon une approche

'multi-modèle'), si le vecteur de paramètres à transférer au pas de temps journalier est déjà mal défini, les conséquences sont amplifiées et le transfert est très mauvais. En même temps, ça n'a peut être pas de sens d'essayer d'estimer les paramètres d'un modèle au pas de temps horaire, alors que ses performances au pas de temps journalier sont déjà mauvaises.

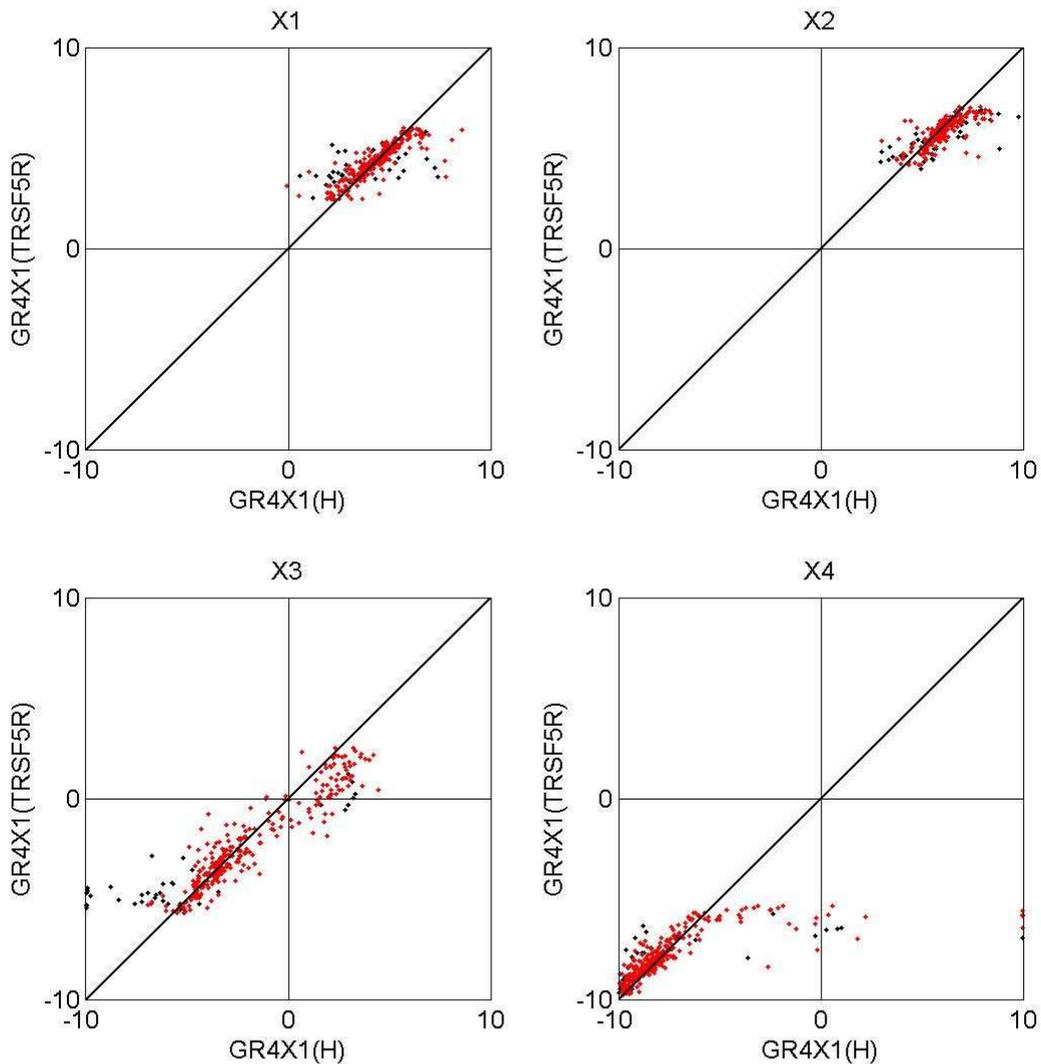


Figure 8-24 : Relations entre les paramètres transformés du modèle GR4X1, optimisés au pas de temps horaire et estimés à partir de l'approche 'vecteur composite' basée sur les 2093 vecteurs de paramètres et les 625 régions de l'espace des paramètres. En rouge et en noir : paramètres des bassins versants dont les performances sont respectivement supérieures et inférieures à 35% de $C_{2M}(RQ)$,

Enfin, cette figure montre que dans une faible proportion des cas (12%), la méthode basée sur l'approche 'multi-modèle' est capable de tirer profit de la diversité des vecteurs de paramètres utilisés et de fournir des simulations légèrement meilleures que celles obtenus avec les vecteurs de paramètres calés.

Les différents vecteurs de paramètres obtenus par la méthode basée sur les régressions linéaires et l'approche de vecteur composite (TRSF5R) sont présentés respectivement à la Figure 8-23 et à la Figure 8-24. Par rapport aux paramètres calés, ces figures montrent une plus grande dispersion des paramètres estimés par les régressions linéaires que ceux estimés par l'approche de 'vecteur composite'. D'ailleurs, les paramètres estimés par l'approche de 'vecteur composite' sont étonnamment bien situés autour de la première bissectrice et sont généralement peu différents des paramètres calés. Ce résultat démontre également l'intérêt de cette approche de transfert de paramètres, malgré sa plus grande complexité par rapport aux régressions linéaires.

Performances en validation :

Nous avons voulu comparer les performances en validation des vecteurs de paramètres estimés selon ces différentes méthodes de transfert de paramètres. Pour cela, nous avons repris la procédure de *split sample test* utilisée, et nous avons comparé les performances obtenues en validation avec des vecteurs optimisés pour chaque période, à celles obtenues avec des vecteurs estimés à partir des données journalières.

Nous avons voulu tester si, malgré l'incertitude liée au transfert de paramètres, l'utilisation de la méthode de transfert basée sur l'approche de 'vecteur composite' permettait de créer des vecteurs de paramètres plus robustes, car composés de nombreux vecteurs de paramètres.

La Figure 8-25 (tout l'échantillon de bassins versants) et la Figure 8-26 (échantillon de bassins versants dont les performances sont supérieures à 35.0% de $C_{2M}(RQ)$) montrent les performances en validation des approches basées sur des régressions linéaires entre paramètres optimisés au pas de temps journalier et horaire et les trois approches de création de 'vecteur composite'. La Figure 8-25 montre que les performances moyennes des quatre approches d'estimation des paramètres sont inférieures à la performance en validation obtenue lorsqu'on cale les vecteurs de paramètres au pas de temps horaire. Les performances des vecteurs estimés à partir de l'approche composite sont également caractérisées par une faible robustesse. Néanmoins, on se rend compte que les performances obtenues par la méthode basée sur les régressions linéaires et l'approche de 'vecteur composite' (TRSF5) sont assez proches et comparables à celles obtenues avec un calage au pas de temps horaire.

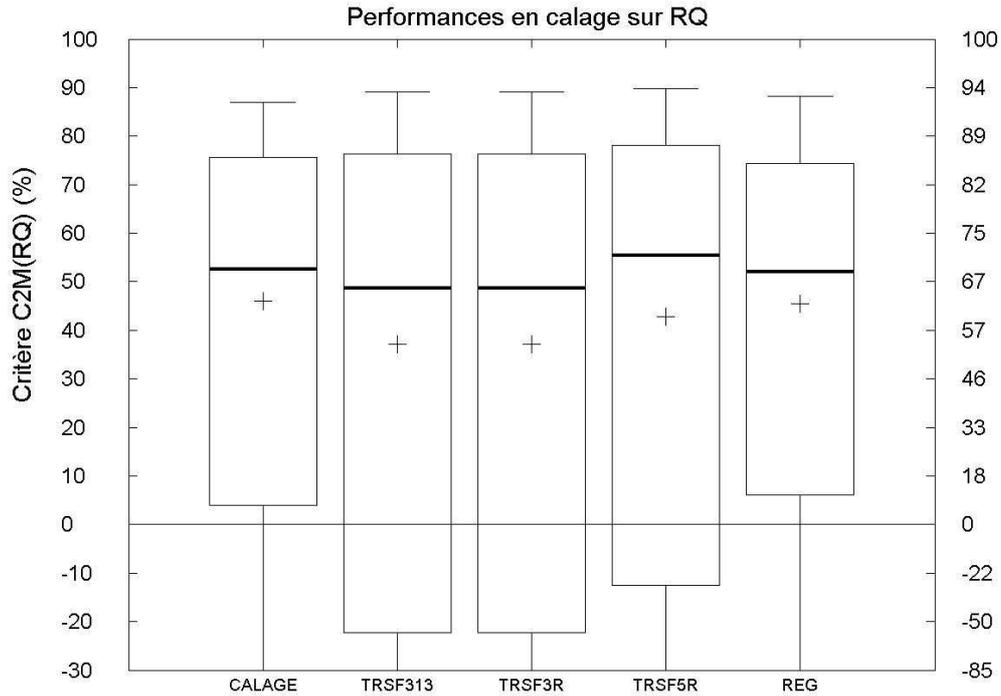


Figure 8-25 : Performances moyennes ($C_{2M}(RQ)$) obtenues en validation au pas de temps horaire et par les différentes approches d'estimation des vecteurs de paramètres horaires, à partir des vecteurs optimisés au pas de temps journalier.

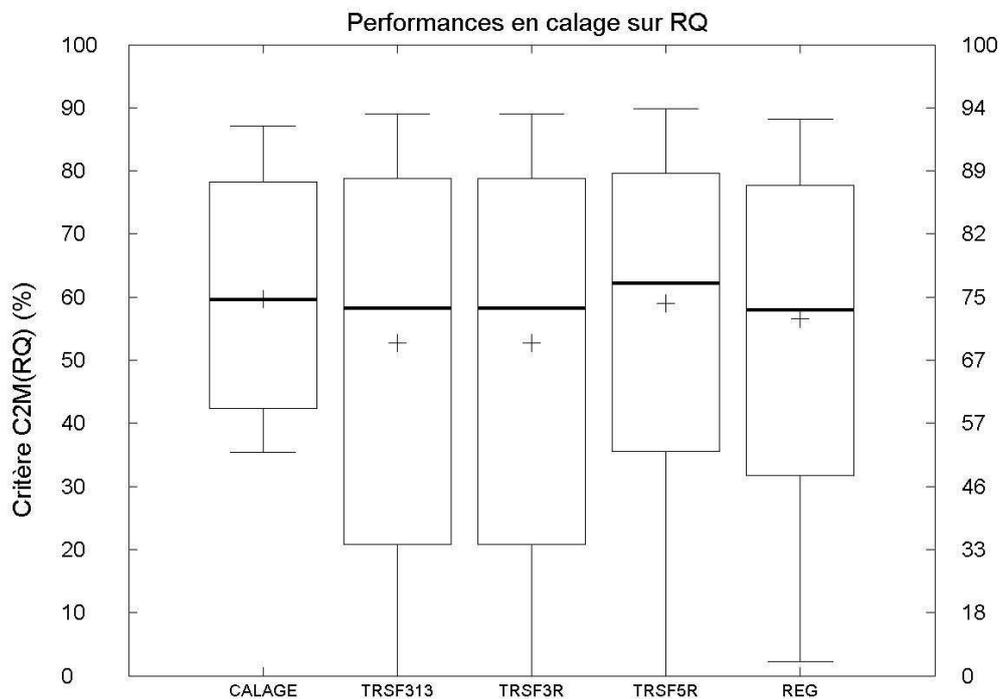


Figure 8-26 : Performances moyennes ($C_{2M}(RQ)$) obtenues en validation au pas de temps horaire et par les différentes approches d'estimation des vecteurs de paramètres horaires, à partir des vecteurs optimisés au pas de temps journalier. Bassins versants dont les performances au pas de temps journalier sont telles que $C_{2M}(RQ) > 35\%$ ($NS > 50\%$).

Ceci signifie que l'incertitude sur la qualité d'un vecteur de paramètres utilisé en validation au pas de temps horaire (conditions hydro-climatiques différentes des conditions d'apprentissage) est aussi grande lorsqu'on cale ce vecteur sur des données horaires, que lorsqu'on détermine ce vecteur à partir d'un vecteur calé au pas de temps journalier sur des chroniques plus longues. Et ce, malgré les incertitudes et erreurs liées au transfert de paramètres d'un pas de temps à un autre.

La Figure 8-26 souligne encore ce constat et met en évidence l'intérêt de ce résultat. Lorsqu'on élimine les bassins versants dont les performances ne sont pas acceptables pour tester une procédure de transfert, les performances moyennes des quatre approches d'estimation des paramètres sont proches ou supérieures (régressions linéaires et TRSF5) à la performance en validation obtenue lorsqu'on cale les vecteurs de paramètres au pas de temps horaire. Les performances obtenues par l'approche utilisant les 'vecteurs composites' (TRSF5) sont les meilleures. Ce résultat confirme ainsi notre hypothèse de départ. Malgré l'incertitude liée au transfert de paramètres d'un pas de temps à un autre, cette approche permet de donner d'aussi bons résultats en validation au pas de temps horaire. Ainsi, si peu de données horaires sont disponibles et qu'une longue chronique journalière existe, il est possible de tirer profit de cette méthode de transfert de paramètres !

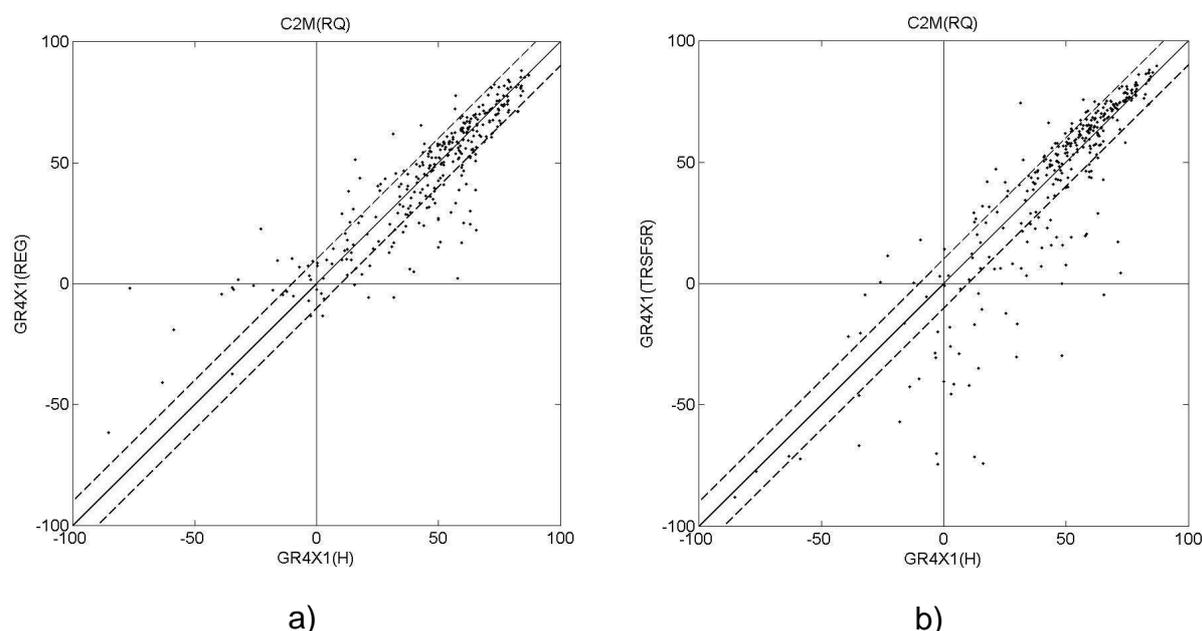


Figure 8-27 : Performances obtenues au pas de temps horaire sur les 313 bassins versants (Modèle GR4X1). a) Comparaison entre les performances en validation des vecteurs calés et des vecteurs déterminés par les régressions linéaires ; b) Comparaison entre les performances en validation des vecteurs calés et des vecteurs déterminés par l'approche de vecteur composite (TRSF5R).

La Figure 8-27 illustre ce résultat. Elle montre la comparaison des performances en validation par bassin, obtenues avec les vecteurs calés au pas de temps horaire et les vecteurs estimés selon les méthodes de transfert de paramètres (régressions linéaires et TRSF5). Ces figures montrent que les résultats sont à peu près les mêmes entre les vecteurs calés et les vecteurs estimés. On remarque néanmoins un léger avantage de la méthode TRSF5 par rapport à la méthode basée sur les régressions linéaires et le calage au pas de temps horaire, quand le $C_{2M}(RQ)$ est supérieur à 50 (Figure 8-27 b)

8.6 Synthèse

Ce chapitre apporte de nombreux résultats, obtenus en testant les modèles hydrologiques en faisant des allers-retours entre les pas de temps horaire et journalier. Dans la littérature, d'autres auteurs ont déjà pratiqué ce système d'analyse des performances entre deux pas de temps, mais c'était généralement le cas entre le pas de temps journalier et mensuel [Moreda, 1999; Mouelhi, 2003]. Cette approche apporte de nombreux résultats particulièrement intéressants et devrait être systématisée.

Le test des 20 modèles au pas de temps journalier apporte des résultats proches de ceux de [Perrin, 2000] sur un échantillon aux caractéristiques différentes. Les conclusions de ces tests par rapport au degré de complexité nécessaire pour représenter correctement la transformation pluie-débit au pas de temps journalier sont les mêmes. Les modèles de la famille MORDOR ont été testés pour la première fois au pas de journalier et donnent d'excellents résultats. Les seules différences de résultats viennent du classement de certains modèles, qui obtiennent de manière étonnante d'assez mauvais résultats sur cet échantillon de bassins versants et avec la méthodologie suivie au cours de cette thèse. Il serait intéressant de déterminer si cela provient des caractéristiques intrinsèques de ces bassins versants plus petits ou bien de l'effet du faible durée de la période de calage (1 an).

Nous avons ensuite démontré l'intérêt d'utiliser une information temporellement distribuée (horaire), pour améliorer les simulations de débit à un pas de temps supérieur (journalier). Comme l'a montré Mouelhi [2003], en améliorant des simulations mensuelles par agrégation de simulations journalières, il est possible d'améliorer fortement les performances moyennes de certains modèles au pas de temps journalier, en agrégeant leurs simulations au pas de temps horaire.

Nous avons également testé différentes versions de modèles de la famille GR, qui ont été développées durant cette thèse, soit au pas de temps horaire, soit au pas de

temps journalier. Ces tests ont révélé qu'il est possible de proposer une structure de modèle également efficace au pas de temps horaire et journalier. Cette version de modèle, appelée GR4X1 (X, pour indiquer son indépendance par rapport au pas de temps) est la même que la structure de GR4H et a seulement la valeur de son paramètre de percolation qui dépend du pas de temps (horaire : 4 ; journalier : 2.25). Cette structure de modèle a été testée au pas de temps journalier sur les 313 bassins versants de cette thèse et également sur les 429 bassins versants de la thèse de *Perrin* [2000]. Les résultats montrent que cette version, à quatre paramètres, est au moins aussi bonne ou meilleure que la structure du modèle GR4J [*Perrin*, 2000]. Ceci confirme l'intérêt de faire sans cesse des aller-retours entre différents pas de temps, différentes stratégies de modélisation, pour faire progresser une structure de modèle.

Enfin, ce chapitre traite des stratégies de prédétermination des paramètres d'un modèle, fonctionnant au pas de temps horaire. Nous avons testé deux approches de transfert de paramètres à travers les pas de temps :

- une approche classique, basée sur des régressions entre paramètres optimisés aux pas de temps journalier et horaire,
- une approche originale, basée sur l'identification des relations entre le positionnement des vecteurs dans l'espace de paramètres, optimisés aux pas de temps journalier et horaire (appelée transfert de régions).

Par rapport aux performances des vecteurs calés sur l'ensemble de la période disponible sur un bassin versant, ces deux approches montrent qu'il est possible de prédéterminer des vecteurs de paramètres ayant de bons niveaux de performance dans 70% des cas.

Par rapport aux performances en validation au pas de temps horaire, ces résultats montrent que les vecteurs de paramètres identifiés par transfert de régions sont capable d'obtenir le même niveau de performance. Cela veut dire qu'il semble que l'incertitude sur les paramètres d'un modèle optimisés au pas de temps horaire avec une chronique d'un an est du même ordre que celle résultant d'un transfert de paramètres optimisés sur de plus longues périodes au pas de temps journalier. Ainsi, d'un point de vue pratique, l'approche de transfert de régions peut permettre de mieux estimer les paramètres du modèle horaire d'un bassin versant, si l'information horaire est peu disponible et qu'il existe de longues chroniques journalières !

Chapitre 9

« L'Errance, terme à la fois explicite et vague , est d'ordinaire associé au mouvement , et singulièrement à la marche, à l'idée d'égarement, à la perte de soi-même. Pourtant le problème principal de l'errance n'est rien d'autre que celui du lieu acceptable »
Alexandre Laumonier, *Errance ou la pensée du milieu.*

CHAPITRE 9 Amélioration des performances du modèle GR4H par une meilleure valorisation de l'information pluviométrique

9.1 Introduction

Dans ce chapitre nous nous demandons s'il est possible d'améliorer les simulations de débit au pas de temps horaire, en utilisant d'autres manières de traiter l'information des différents postes pluviométriques disponibles sur un bassin versant. Classiquement, lorsqu'on utilise un modèle conceptuel à l'échelle globale, les différentes séries pluviométriques sont agrégées pour fournir une seule série de pluie en entrée du modèle. Il existe de nombreuses manières de calculer cette pluie de bassin, que nous n'aborderons pas. Les travaux de *Andréassian et al.* [2001] ont montré que les modèles pluie-débit globaux sont généralement capables d'adapter leurs vecteurs de paramètres à différentes entrées de pluie, pour fournir des simulations de débit de même qualité, quelle que soit la manière d'agréger la pluie disponible.

D'autres travaux au Cemagref concernent la manière d'intégrer la connaissance spatiale de la pluie au sein d'un modèle global [*Oddos, 2002; Andréassian et al., 2004a; Nejmeddine, 2004*]. Ces travaux ont testé différentes approches de prise en compte de l'hétérogénéité de la pluie et de distribution de la pluie, au sein de modèles pluie-débit, au pas de temps journalier. Ces travaux ont porté, soit sur des bassins versants réels, soit sur des bassins versants 'chimères', bassins virtuels issus de l'agrégation de deux bassins différents. Ces travaux n'ont pas mis en évidence l'intérêt de distribuer la pluie au sein d'un modèle pluie-débit tel que GR4J, au pas de temps journalier, sur des bassins versants réels. La raison avancée est que les séries pluviométriques des différents postes disponibles sur un bassin versant sont généralement très homogènes. En effet, la distance de corrélation d'un champ de pluie est généralement assez grande au pas de temps journalier (portée du variogramme de la pluie de plusieurs centaines de kilomètres). En fait, bien que cette idée se retrouve assez peu dans la littérature, il semble exister une certaine taille critique de bassin versant, en dessous de laquelle les champs de pluie sont fortement homogènes et au dessus de laquelle le système bassin versant joue un rôle de filtre passe-bas assez fort, ce qui atténue les variabilités à court terme des débits. Sur des bassins versants chimères, où l'hétérogénéité de la pluie a été artificiellement augmentée, il semble que la distribution de la pluie soit responsable de 75% des gains de performances des modèles et que la distribution des

paramètres des modèles ne soit responsable que de 25% des gains de performances, par rapport à une approche globale.

Lorsqu'on s'intéresse à la simulation des débits au pas de temps horaire, la distance de corrélation d'un champ de pluie devient certainement plus petite qu'au pas de temps journalier. Ainsi, la question de l'intérêt de la distribution se pose à nouveau. Au pas de temps horaire, en France, il semble que la distance de corrélation d'un champ de pluie se situe entre 15 km et jusqu'à 150 km en plaine (Jean-Paul Goutorbe, Météo France, Communication personnelle). *Datin* [1998] a estimé pour les Cévennes ardéchoises que la portée du variogramme de la pluie suit empiriquement la relation :

$$\beta(\Delta t) \approx 25 \cdot \sqrt{\Delta t} \quad [8.5]$$

avec, β la portée du variogramme et Δt le pas de temps considéré en heure. Ainsi, la portée du variogramme des pluies horaires est d'environ 25 km et celle des pluies journalières est d'environ 120 km.

Les bassins versants de notre base de données ont une superficie médiane de 85 km², ce qui les situe plus ou moins dans cette distance de corrélation, suivant leur régime climatique. Dans ce chapitre, nous allons présenter les résultats de différentes stratégies de prise en compte de la répartition spatiale de la pluie.

9.2 Méthodologie

9.2.1 Bassins versants

Pour ces tests, nous avons retenu 145 bassins versants français (Figure 9-1), pour lesquels nous disposons de 2 à 8 pluviomètres (Figure 9-2 a). Les superficies de ces bassins versants vont de 1.1 à 2700 km² (Figure 9-2 b).

La durée des chroniques va de 3 ans à 34 ans, avec une durée de 6 ans en moyenne. Ces bassins versants se retrouvent bien dispersés sur le territoire français, dans des zones climatiques contrastées, avec des types de pluies bien différents (contraste climatique de la Bretagne et de la Méditerranée, par exemple).

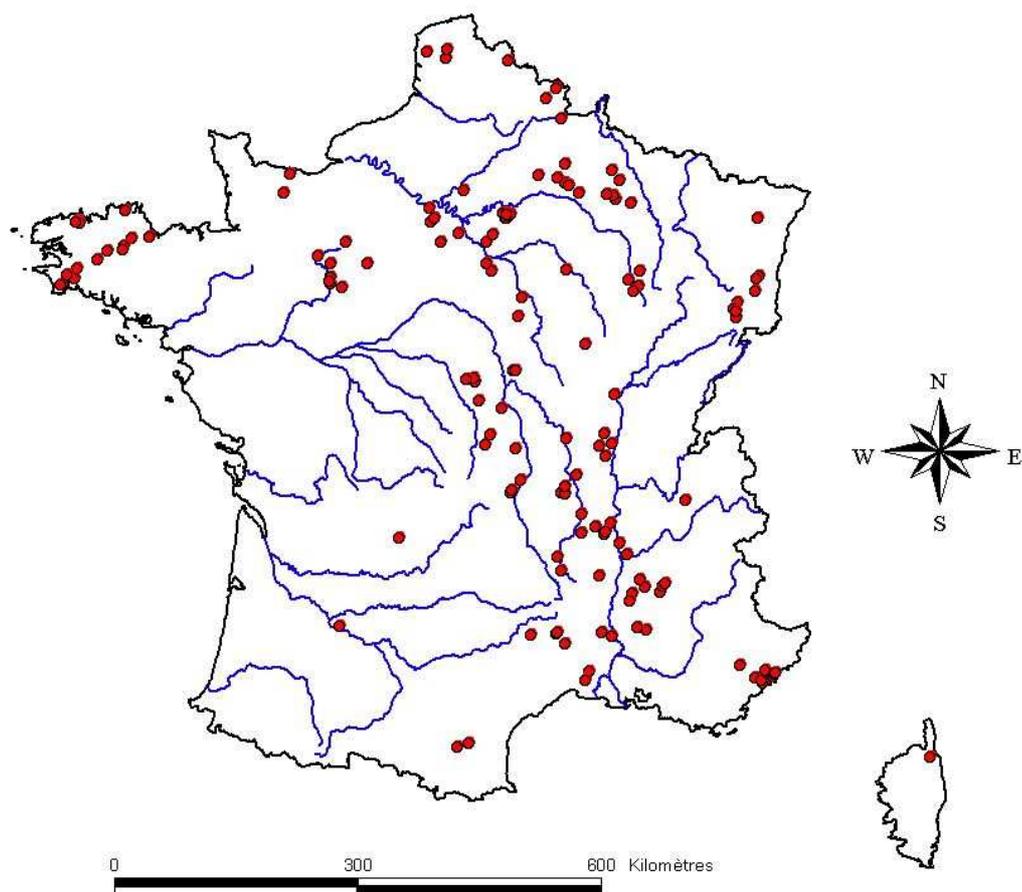


Figure 9-1 : Localisation des exutoires des bassins versants pour lesquels l'approche de distribution a été testée.

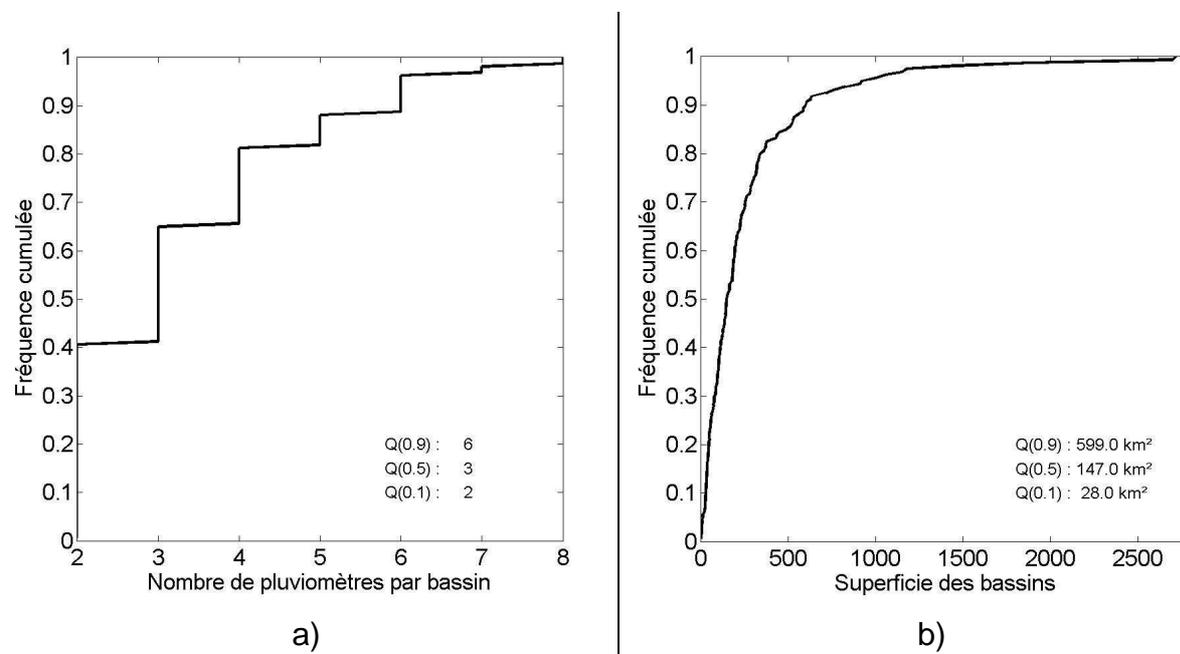


Figure 9-2 Distributions cumulées du nombre de pluviomètres (a) et de la superficie (b) des 160 bassins versants utilisés pour tester l'intérêt de la distribution de la pluie.

9.2.2 Stratégies de valorisation de l'information pluviométrique

Ces tests étant une première approche de prise en compte de l'information spatiale pluviométrique au pas de temps horaire, nous avons testé quelques approches assez simples. L'idée de base est que, lorsqu'on calcule une pluie moyenne à partir de plusieurs postes pluviométriques, beaucoup d'information est perdue, dont (1) les intensités maximales, qui sont fortement diminuées à cause de l'effet moyennage et (2) la variabilité de la pluie. Cet effet de réduction de la variance d'une pluie moyenne par rapport à la variance de la pluie mesurée en un poste a été mis en évidence par *Yevjevich [1972]* et illustré récemment par *Dong [2005]* et *Dong et al. [2005]*. Ainsi, au pas de temps horaire et par rapport au pas de temps journalier, il peut être intéressant de conserver les fortes intensités de pluie pour bien représenter les fortes crues. Les chroniques de pluie de chaque pluviomètre possèdent parfois des lacunes aléatoires dans leurs enregistrements. Pour les heures concernées, nous nous sommes permis de combler cette lacune par la pluie moyenne calculée avec les autres postes pluviométriques disponibles.

L'idée de base des tests effectués est de faire tourner un modèle par poste pluviométrique disponible et d'assembler ensuite les débits. L'intérêt de faire tourner différents modèles en parallèle est de profiter d'une diversité d'états internes des modèles qui représentent l'état d'humidité du bassin versant. Les différentes variantes testées concernent la manière d'optimiser les paramètres des modèles associés à chaque pluviomètre et d'assembler les débits. Cette approche est illustrée à la Figure 9-3 b). Dans notre approche, la modélisation pluie-débit reste globale et la prise en compte de l'hétérogénéité de la pluie se fait au travers d'une approche multi-modèle.

Nous avons testé quatre stratégies différentes, basées sur deux lois de composition des débits et deux types d'optimisation des paramètres du modèle (Tableau 9-1) :

- Composition des débits :
 - Option E : somme équipondérée des débits simulés par les n modèles en parallèle,
 - Option P : somme pondérée par les performances en calage de chaque débit simulé. Elimination des chroniques de débits simulés en deçà d'un seuil de performance ($C2M \leq 0.$).

- Optimisation des paramètres des modèles :
 - Option G : un seul vecteur de paramètres optimisé pour les n modèles
 - Option SG : un vecteur de paramètres optimisé par modèle, optimisé par rapport au débit total (soit n vecteurs de paramètres).

Chapitre 9 : Amélioration des performances du modèle GR4H par une meilleure valorisation de l'information pluviométrique

Tableau 9-1 : Synthèse des différentes stratégies de prise en compte de l'hétérogénéité de la pluie.

Symbole	Composition des débits	Paramètres du modèle
GLOBAL	Fonctionnement global du modèle	
EG	Somme équipondérée des n débits simulés n : nombre de postes pluviométriques disponibles	Paramétrage Global : Un seul vecteur de paramètres optimisé pour les n modèles
ESG	Somme équipondérée des n débits simulés n : nombre de postes pluviométriques disponibles	Paramétrage Semi-Global : Un vecteur de paramètres optimisé par modèle, optimisé par rapport au débit total.
PG	Somme pondérée par les performances en calage de chaque débit simulé. Elimination des débits simulés en deça d'un seuil de performance ($C2M \leq 0.$)	Paramétrage Global : Un seul vecteur de paramètres optimisé pour les n modèles
PSG	Somme pondérée par les performances en calage de chaque débit simulé. Elimination des débits simulés en deça d'un seuil de performance ($C2M \leq 0.$)	Paramétrage semi-global : Un vecteur de paramètres optimisé par modèle, optimisé par rapport au débit total.

La variante la plus simple est caractérisée par l'optimisation d'un seul vecteur de paramètres pour les n modèles en parallèle et le débit simulé est la somme équipondérée des débits. La variante la plus complexe est caractérisée par l'optimisation de n vecteurs de paramètres (un par modèle) et le débit simulé est la somme des débits, pondérés par les performances de chaque modèle (pluviomètre) par rapport au débit du bassin versant.

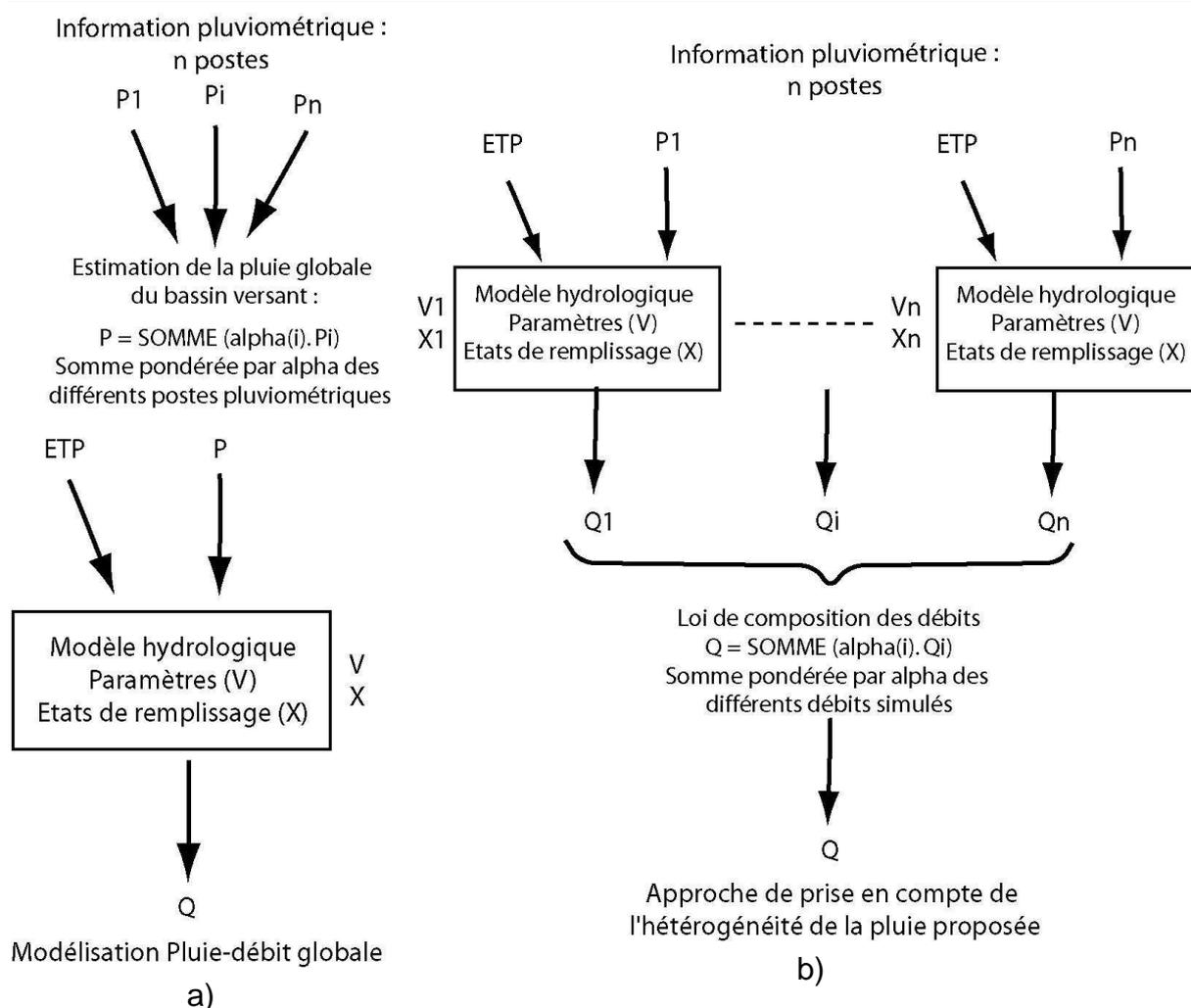


Figure 9-3 : Modélisation pluie-débit, a) schéma représentant l'approche globale classique et b) schéma représentant l'approche de prise en compte de la diversité de l'information pluviométrique.

9.3 Résultats

Nous présentons ici les résultats des quatre approches de prise en compte de l'information pluviométrique spatialisée. Nous comparons ces résultats à la simulation de base du modèle en mode global. Nous ne présentons que les résultats du modèle GR4H. Nous avons testé également cette approche sur la structure du modèle GR4J. Quelles que soient les approches testées, les différences moyennes de performances entre ces deux modèles ont toujours été de l'ordre de 4 à 5 points. Ceci permet de confirmer qu'il est toujours intéressant de progresser sur la structure d'un modèle hydrologique, car les progrès réalisés sur une structure se répercutent ensuite, quel que soit le mode de fonctionnement du modèle.

Performances en calage :

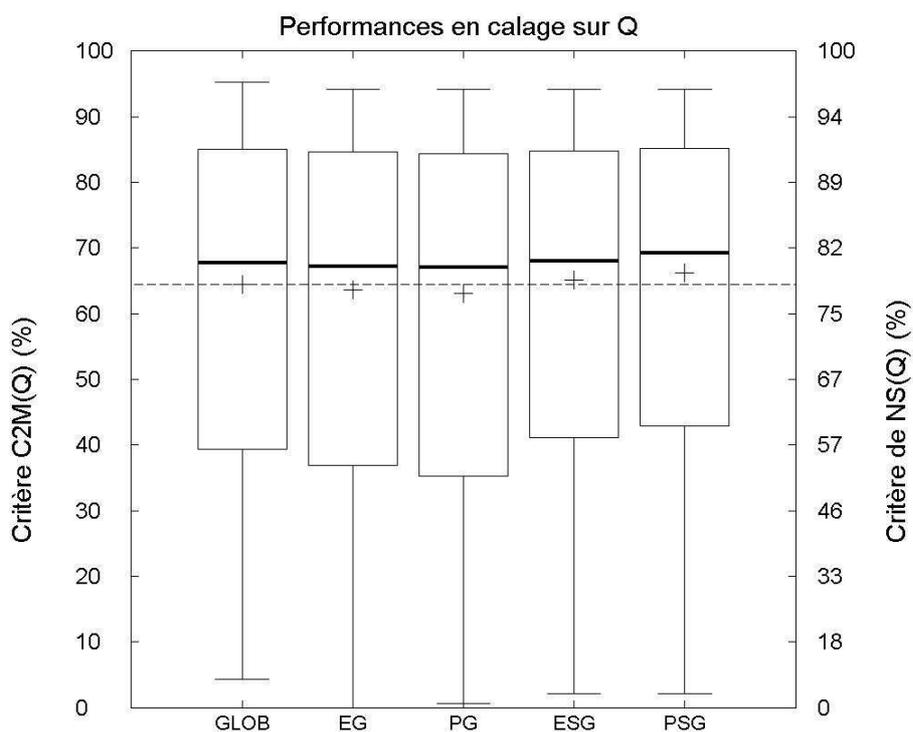


Figure 9-4 : Synthèse des performances en calage ($C_{2M}(Q)$) des différentes stratégies de prise en compte de l'hétérogénéité de la pluie et comparaison par rapport à l'approche globale (modèle GR4H). Les pointillés indiquent la performance moyenne de l'approche globale.

Les performances en calage sur 145 bassins versants sélectionnés sont du même ordre que celles obtenues sur l'ensemble des 313 bassins versants (Figure 9-4). La fonction objectif de calage utilisée est le critère $C_{2M}(Q)$. Par rapport aux performances de la méthode globale, les performances des stratégies EG et PG (un vecteur de paramètre global et recomposition des débits) sont légèrement inférieures. En effet, il est certainement plus difficile pour les modèles en parallèle de trouver un vecteur de paramètres qui convienne aux n séries pluviométriques et modèles tournant en parallèle. Par contre, les performances des stratégies ESG et PSG (un vecteur de paramètre par modèle et recomposition des débits) sont légèrement meilleures que celles de la méthode globale. Effectivement, ces stratégies possèdent un plus grand nombre de degrés de libertés (n vecteurs de 4 paramètres), ce qui permet aux modèles de fournir de meilleures simulations en calage.

Performances en validation :

Les performances des différentes stratégies de prise en compte de la connaissance spatiale de la pluie en validation vont nous servir à juger de leur intérêt. Leurs

performances ont été évaluées avec les critères $C_{2M}(Q)$ (Figure 9-5) et $C_{2M}(RQ)$ (Figure 9-6).

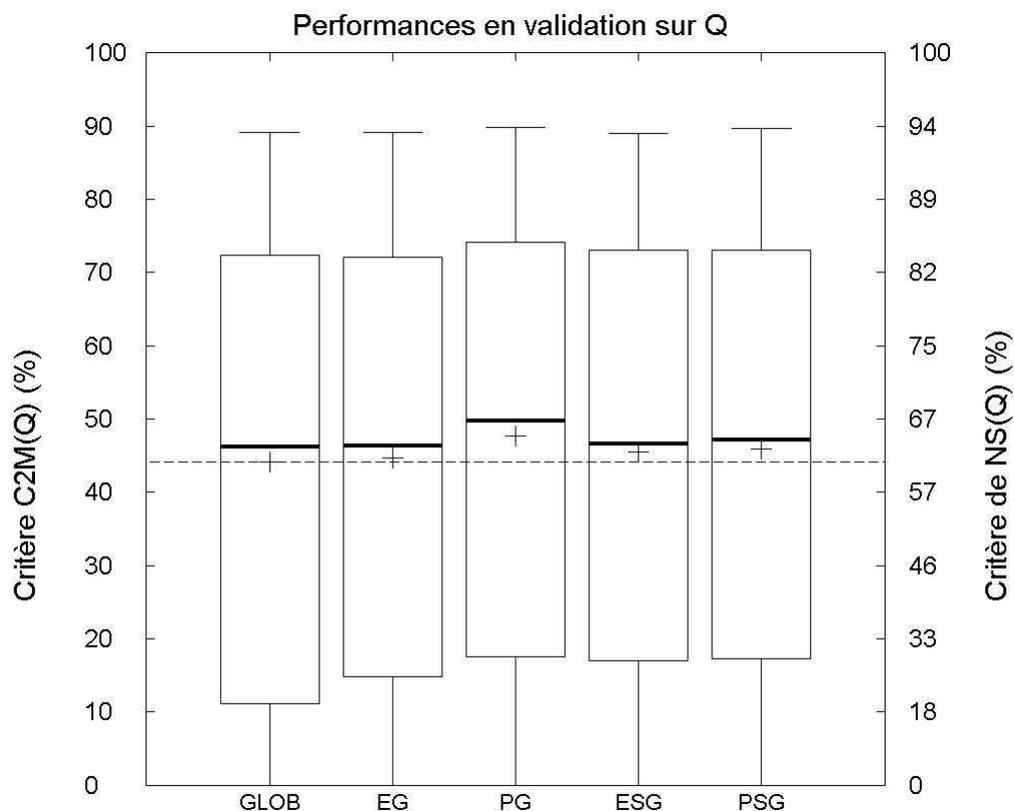


Figure 9-5 : Synthèse des performances en validation ($C_{2M}(Q)$) des différentes stratégies de prise en compte de l'hétérogénéité de la pluie et comparaison par rapport à l'approche globale (modèle GR4H). Les pointillés indiquent la performance moyenne de l'approche globale.

La Figure 9-5 et la Figure 9-6 montrent les performances des différentes stratégies. Elles montrent que la stratégie la plus simple (EG) n'apporte pas de progrès significatif et ses performances moyennes sont égales à celles de la méthode globale. Par contre, les autres méthodes PG, ESG et PSG, sont meilleures en validation que la méthode globale. Les gains moyens de performances sont similaires pour les deux critères de performances et vont de +1.4 à +3.5 points pour le critère $C_{2M}(Q)$ * et de +1.1 à 2.1 points pour le critère $C_{2M}(RQ)$. La meilleure stratégie est finalement la plus simple des trois, la stratégie PG : utilisation d'un seul vecteur pour tous les n modèles en parallèle et pondération des différentes simulation par le critère de calage associé à chaque modèle. Les deux méthodes pondérées donnent les meilleures performances.

* D'autres tests que je n'ai pu m'empêcher de faire pendant la rédaction de mon mémoire de thèse permettent de gagner jusqu'à 4 points pour le critère $C_{2M}(Q)$, en utilisant d'autres pondérations.

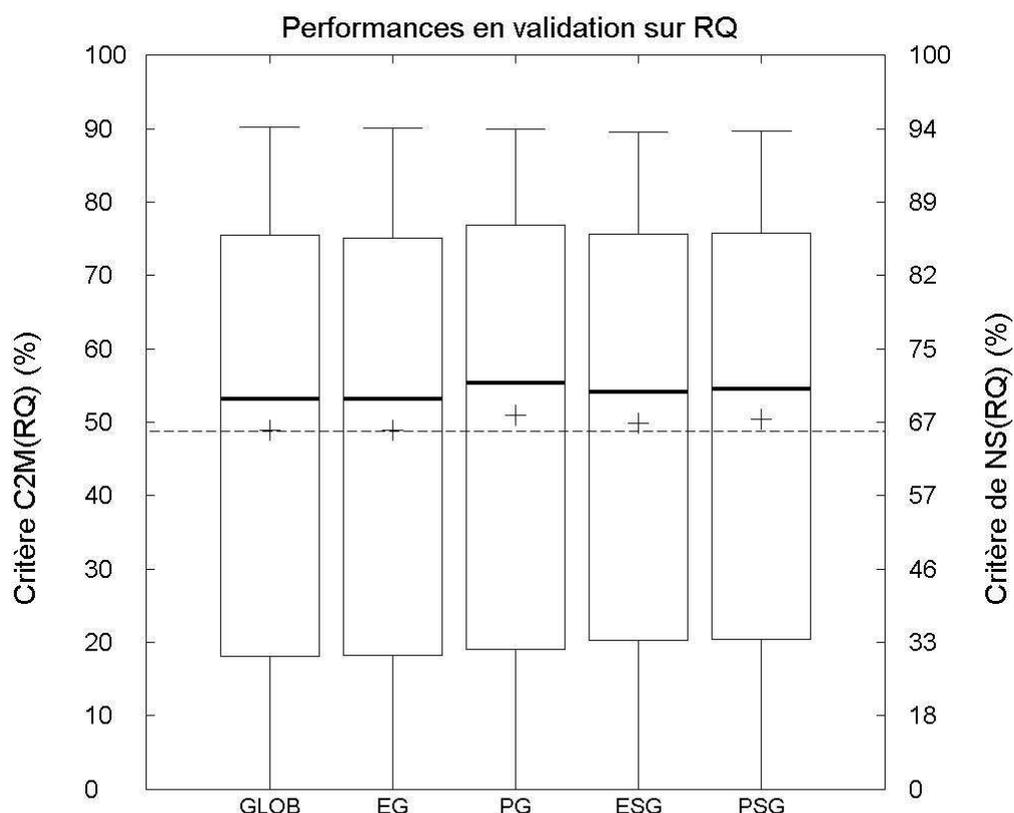


Figure 9-6 : Synthèse des performances en validation ($C_{2M}(RQ)$) des différentes stratégies de prise en compte de l'hétérogénéité de la pluie et comparaison par rapport à l'approche globale (modèle GR4H). Les pointillés indiquent la performance moyenne de l'approche globale.

Un gain moyen de 3.5 points sur un échantillon de 145 bassins versants est remarquable. Il souligne l'intérêt et la robustesse de cette approche. La robustesse de la stratégie PG est également illustrée par la comparaison des distributions des performances de l'approche globale et des différentes stratégies (Figure 9-7). Cette figure montre un assez fort décalage entre la distribution des performances de la méthode globale et de la stratégie PG, grâce à la meilleure robustesse de cette stratégie.

La Figure 9-8, la Figure 9-9 et la Figure 9-10 présentent les comparaisons respectives des performances en validation (critère $C_{2M}(Q)$) de la méthode globale et des méthodes PG et PSG. Ces figures montrent clairement l'intérêt de la méthode PG par rapport à la méthode globale et à la stratégie PSG.

Pour une bonne partie des périodes de validation (75%), les performances de cette méthode sont supérieures à celle de la méthode globale.

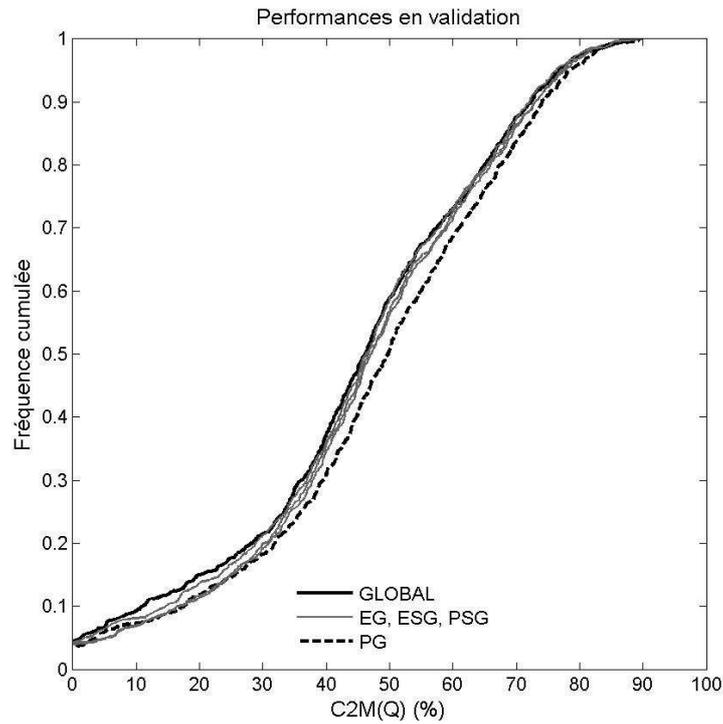


Figure 9-7 : Distribution des performance en validation ($C_{2M}(Q)$) de la méthode globale et des stratégies de prise en compte de l'hétérogénéité de la pluie.

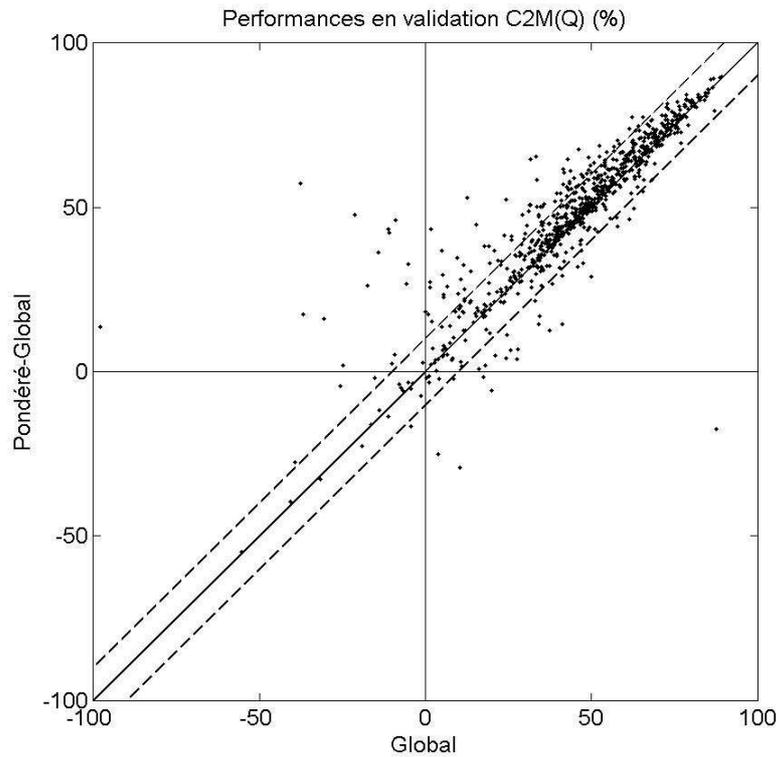


Figure 9-8 : Comparaison des performances en validation (critère $C_{2M}(Q)$) de l'approche globale et de la stratégie PG. Pointillés : lignes +/- 10 points par rapport à la bissectrice.

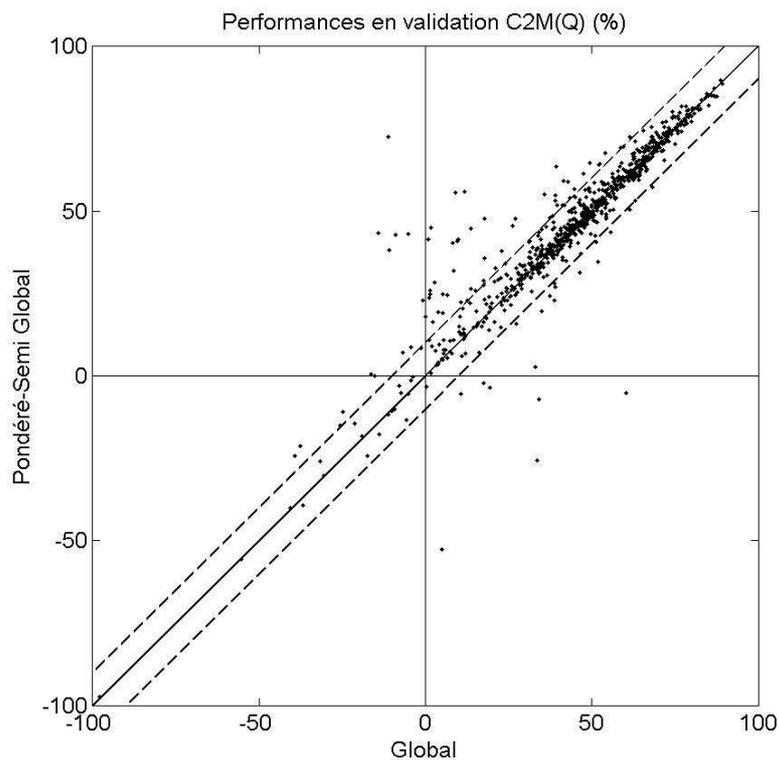


Figure 9-9 : Comparaison des performances en validation (critère $C_{2M}(Q)$) de l'approche globale et de la stratégie PSG. Pointillés : lignes +/- 10 points par rapport à la bissectrice.

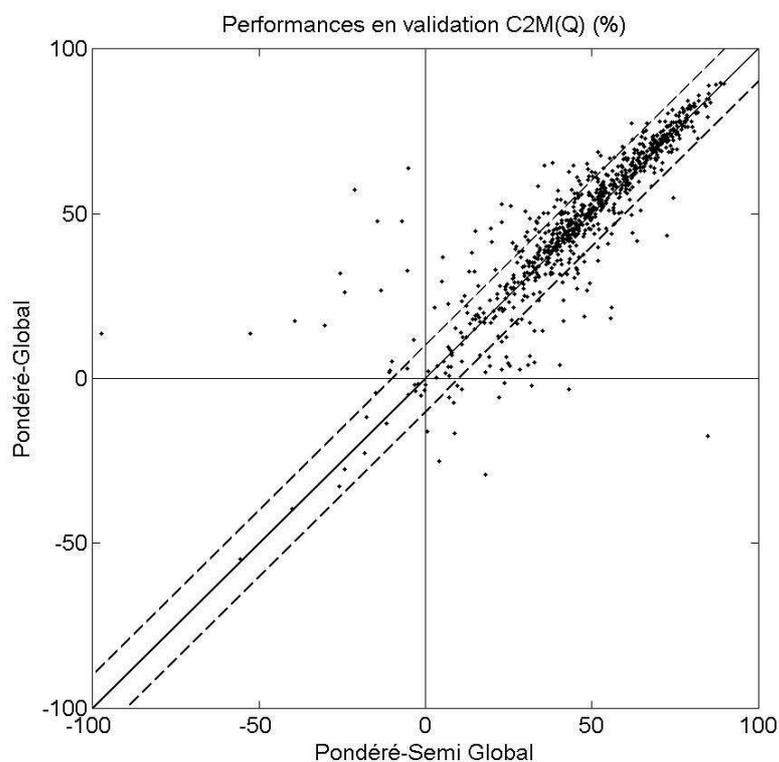


Figure 9-10 : Comparaison des performances en validation (critère $C_{2M}(Q)$) de la stratégie PG et de la stratégie PSG. Pointillés : lignes +/- 10 points par rapport à la bissectrice.

Dans certains cas (13%), les gains de performances peuvent même être supérieurs à plusieurs dizaines de points. Ces cas rares proviennent probablement du filtrage de certains pluviomètres qui donnaient des simulations peu satisfaisantes et pour lesquelles le poids des modèles correspondants était faible dans le calcul du débit total.

La Figure 9-11 présente la comparaison des performances de la stratégie PG et de l'approche globale par bassin versant. Cette figure met en évidence la robustesse de cette méthode, puisqu'elle est meilleure pour la majorité des bassins versants.

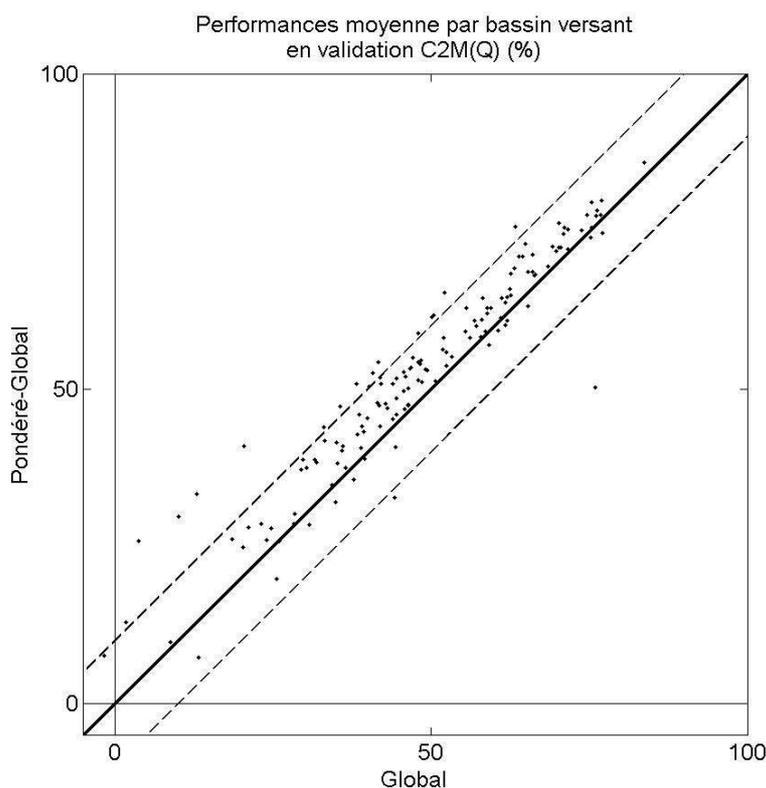


Figure 9-11 : Comparaison des performances en validation par bassin versant (critère $C_{2M(Q)}$) de l'approche globale et de la stratégie PG. Pointillés : lignes +/- 10 points par rapport à la bissectrice.

Les gains de la stratégie PSG par rapport à la méthode globale sont légèrement inférieurs à ceux de la méthode PG (Figure 9-9). Ceci vient peut être du manque de robustesse des vecteurs de paramètres identifiés au pas de temps horaire. Néanmoins lorsqu'on compare les deux stratégies PG et PSG (Figure 9-10), la stratégie PG est supérieure dans 66% des cas. Ainsi, il est peut être intéressant de ne pas abandonner la stratégie PSG. Il existe peut-être d'autres stratégies intermédiaires à tester. Si la différence de performances vient du manque de robustesse des vecteurs identifiés, il pourrait être assez facile de tester le passage de la stratégie PG vers PSG, en optimisant certains paramètres pour chaque modèle et en gardant les autres paramètres en commun.

Vecteurs de paramètres :

Les valeurs des paramètres optimisés* selon la méthode globale et l'approche PG sont comparées à la Figure 9-12. Cette figure met en évidence la plus grande variabilité des paramètres X_1 , X_2 et dans une moindre mesure du paramètre X_3 . Le paramètre X_4 semble peu affecté par la stratégie de distribution de la pluie. Ce résultat peut faire penser aux résultats obtenus lorsqu'on étudie l'influence d'un biais systématique contenu dans les données d'entrées [Oudin et al., 2005d]. Les paramètres X_1 , X_2 , X_3 peuvent être utilisés pour compenser des erreurs de volume, alors que le paramètre X_4 (temps de réponse de l'hydrogramme unitaire) n'a logiquement que peu d'influence sur ce phénomène. Les valeurs des paramètres X_1 et X_2 de la stratégie PG sont différentes de celles des paramètres globaux dans 25% à 35% des cas. L'augmentation du paramètre X_1 est généralement corrélée à une diminution du paramètre X_2 (Figure 9-12 c). Le modèle semble donc utiliser prioritairement les paramètres X_1 et X_2 pour s'adapter aux séries pluviométriques. Dans la perspective d'une approche intermédiaire entre les stratégies PG et PSG, il pourrait donc être intéressant d'essayer de ne distribuer que les paramètres X_1 et X_2 pour chaque modèle associé à un pluviomètre et de garder X_3 et X_4 identiques pour tous les modèles.

Existe t'il un déterminant dans les gains de performance des stratégies de prise en compte de l'hétérogénéité de la pluie ?

Une idée *a priori* commune à ces tests de prise en compte de la variabilité de la pluie est qu'il est plus intéressant de distribuer la pluie lorsque les champs de pluie sont très variés. Ainsi, on pourrait éventuellement s'attendre à ce que cette approche donne de meilleurs résultats dans la région Méditerranéenne qu'en Bretagne. Néanmoins, comme le montre la Figure 9-13 les stratégies de distribution sont généralement les meilleures et il ne semble pas exister de déterminisme régional permettant de discriminer quelle stratégie est la meilleure. Il semble exister une légère tendance sur les bonnes performances de la stratégie PSG, qui semblent plus concentrées dans l'aire méditerranéenne. La Figure 9-14 et la Figure 9-15 comparent respectivement la méthode globale et la stratégie PG, et les stratégies PG et PSG.

* Pour rappel : X_1 et X_2 sont respectivement les capacités des réservoirs de routage et de production ; X_3 est le coefficient de la fonction d'échanges souterrains ; X_4 est le temps de base de l'hydrogramme unitaire.

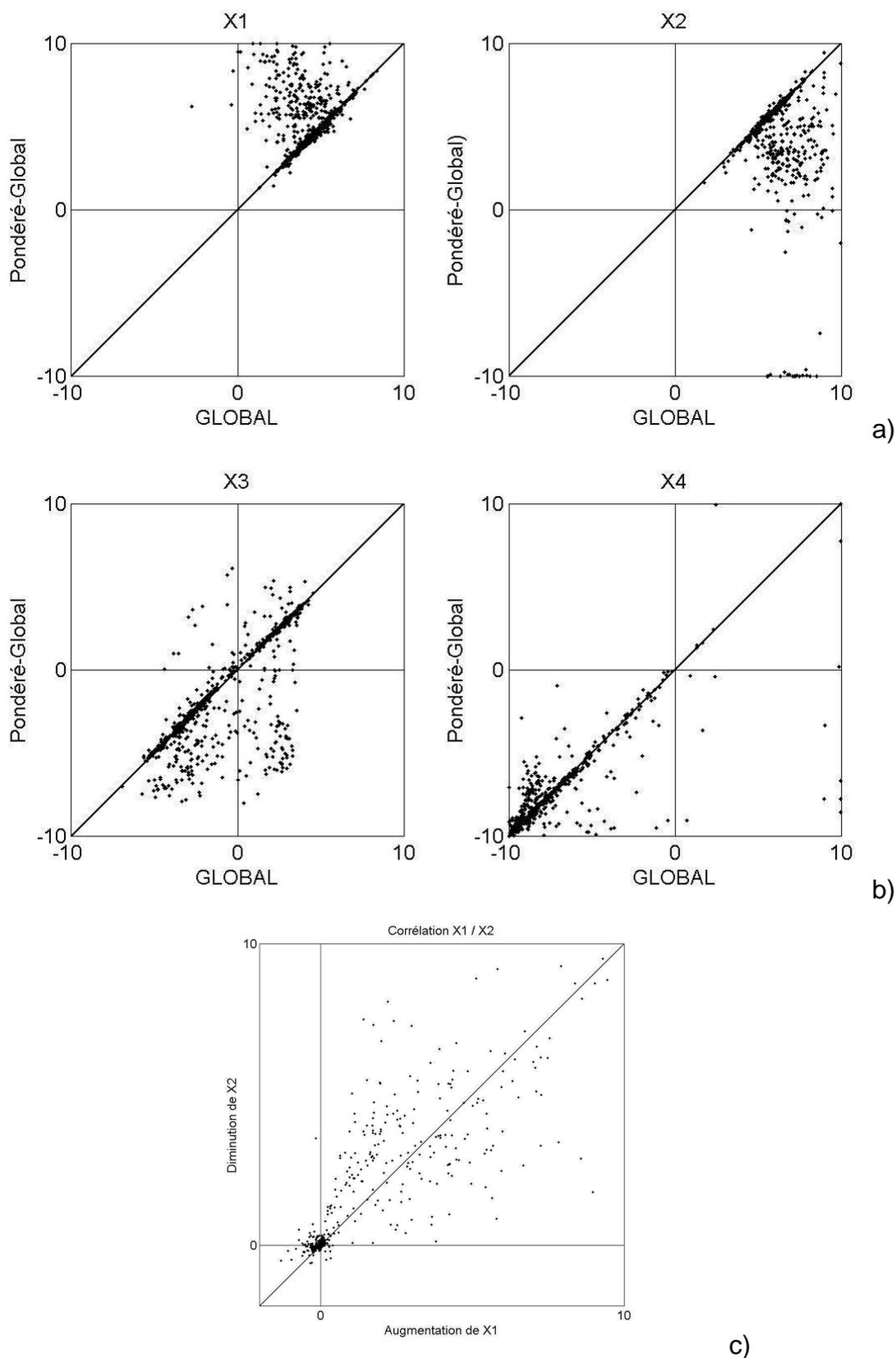


Figure 9-12 : Comparaison des valeurs des paramètres optimisés lorsque le modèle est utilisé avec une pluie globale ou la stratégie Pondéré-Global (PG). a) paramètres X_1 , X_2 , b) paramètres X_3 , X_4 , c) relations entre les différences de valeurs des paramètres X_1 et X_2 , entre la méthode globale et la stratégie PG.

Sur ces figures ont été différenciés les cas où la différence de performances est supérieure à 2.5 points de $C_{2M}(Q)$ en validation (70% des cas). Ces figures montrent également qu'il n'existe pas de fort déterminisme régional discriminant les différences de performances des stratégies de distribution. L'intérêt de la stratégie PG est renforcé car la méthode globale est très rarement significativement supérieure à cette méthode. Par rapport à la méthode PSG, il semble y avoir une petite tendance, avec une majorité des cas où elle est meilleure dans l'aire d'influence méditerranéenne (Cévennes, haut bassin de la Loire et Alpes maritimes).

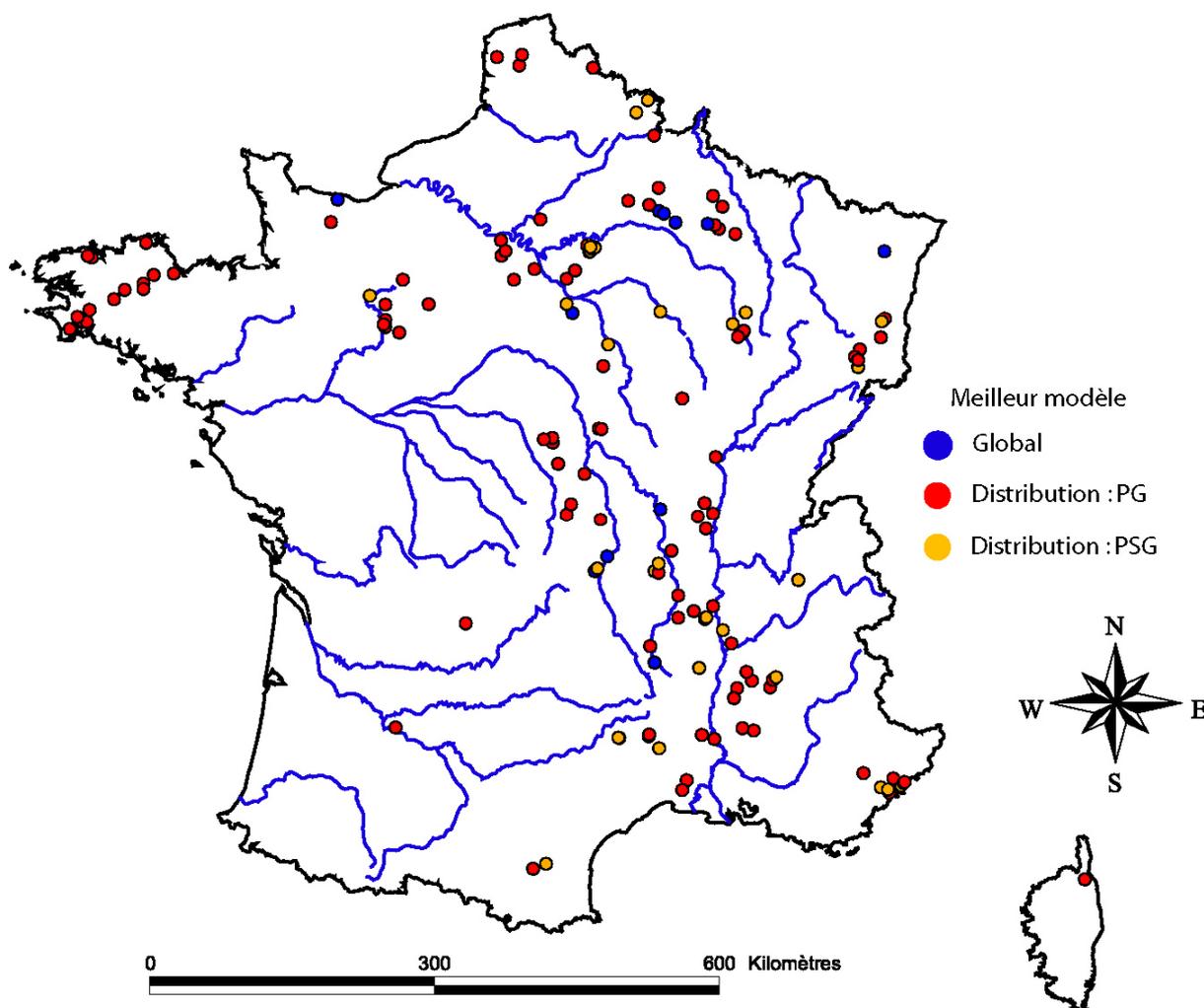


Figure 9-13 : Répartition géographique de la meilleure stratégie de prise en compte de l'hétérogénéité de la pluie par bassin versant. Approche globale, stratégie PG (pondéré globale), PSG (pondéré semi-global).

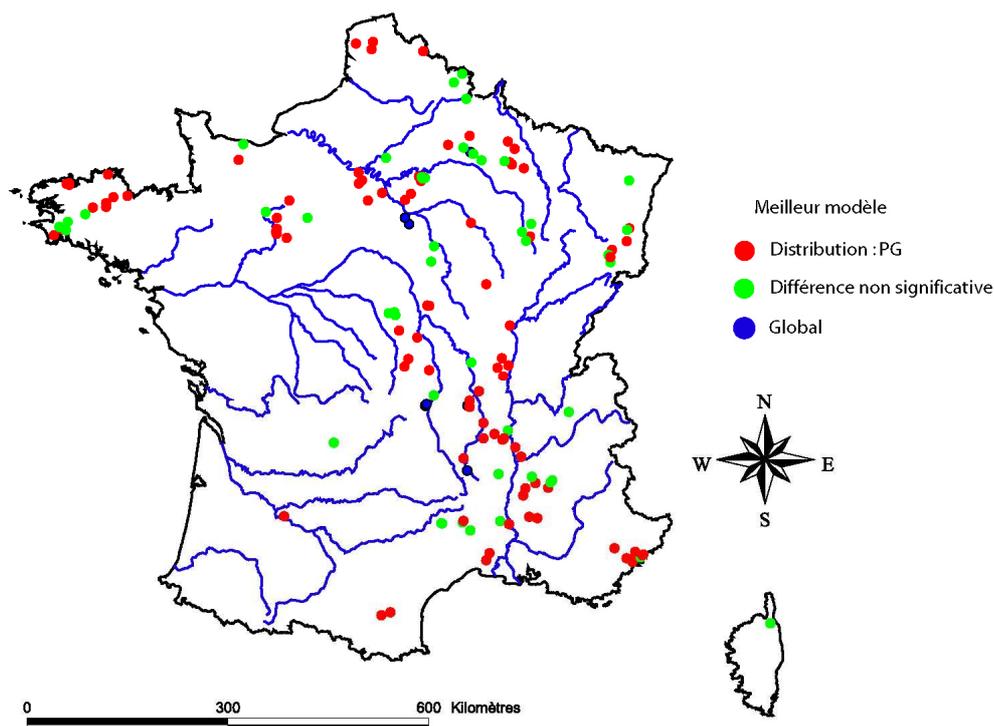


Figure 9-14 : Comparaison de l'approche globale et de la stratégie PG. Une stratégie est jugée meilleure si la différence de performances est supérieure à 2.5 points de $C_{2M}(Q)$. La différence de performances est jugée non-significative si elle inférieure à 2.5 points.

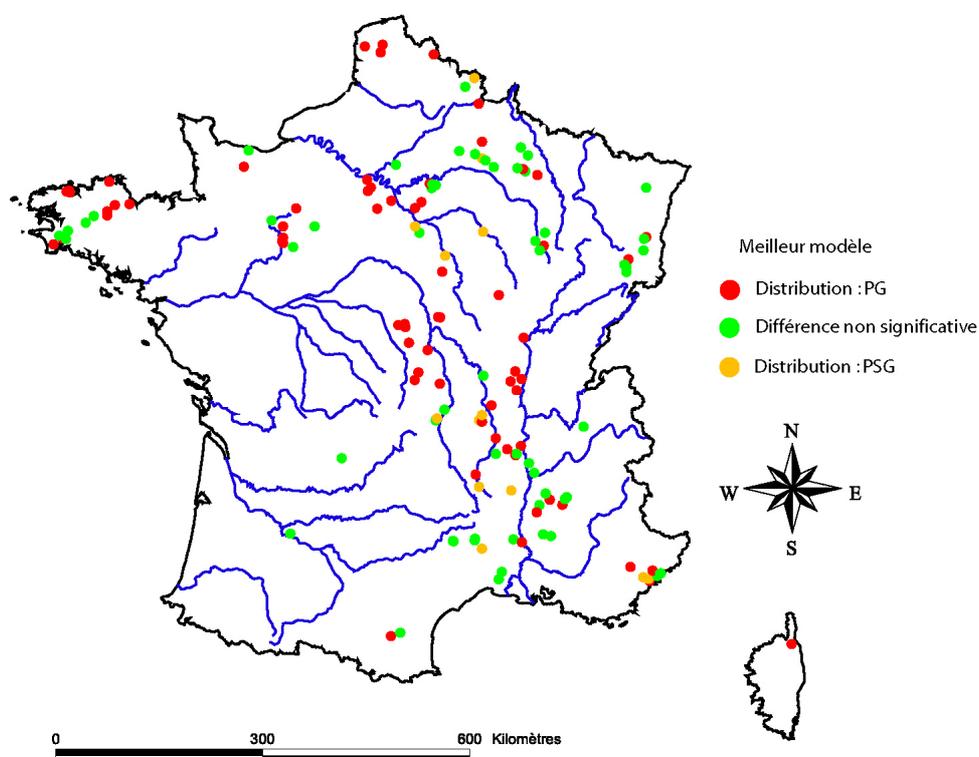


Figure 9-15 : Comparaison des stratégies PG et PSG. Une stratégie est jugée meilleure si la différence de performances est supérieure à 2.5 points de $C_{2M}(Q)$. La différence de performances est jugée non-significative si elle inférieure à 2.5 points.

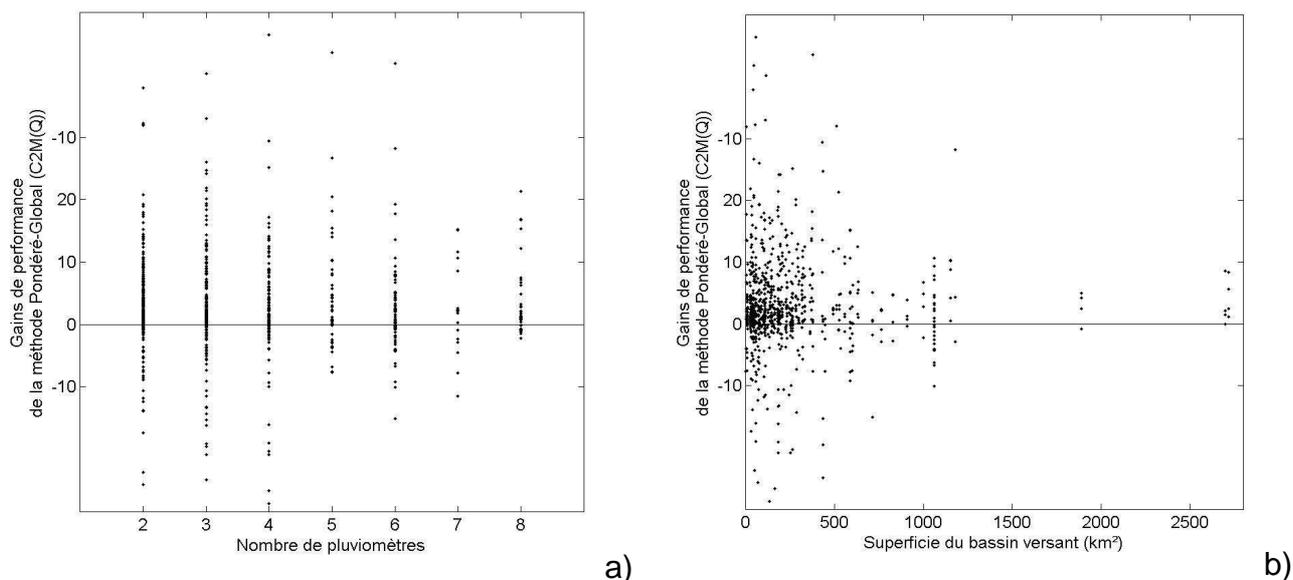


Figure 9-16 : Gains de performances de la méthode PG par rapport à l'approche globale, en fonction a) du nombre de pluviomètres et de b) la superficie du bassin versant.

Enfin, nous avons voulu vérifier si des facteurs comme le nombre de pluviomètres disponibles ou la superficie du bassin versant pouvaient être corrélés aux gains de performance de la stratégie PG, par rapport à l'approche globale. Comme le montre la Figure 9-16, il ne semble pas exister de tendances très claires entre ces facteurs et les gains de performances. Une analyse plus détaillée des résultats, prenant en compte la configuration pluviométrique devra être menée pour essayer de comprendre l'origine de ces gains de performances.

9.4 Conclusion

Dans ce chapitre, nous avons présenté quelques stratégies simples de prise en compte de l'hétérogénéité de la pluie. Les premiers résultats que nous avons obtenus montrent l'intérêt de faire tourner un modèle par pluviomètre dont les données sont disponibles et de ne chercher à optimiser qu'un seul vecteur de paramètres pour tous les modèles utilisés en parallèle. Ensuite, la manière la plus efficace de composer les débits simulés par les différents modèles en parallèle est de faire une moyenne, pondérée par les performances en calage associée à chaque pluviomètre. Les résultats de cette approche, appelée PG (pondéré-global), sont significativement meilleurs en validation et plus robustes que ceux de l'approche globale et d'une approche utilisant un vecteur de paramètres par pluviomètres.

Ces résultats sont préliminaires et leur analyse devra être approfondie. Se pose notamment la question du sens de cette approche proposée. Est-ce réellement une approche de distribution ou de post-filtrage de la qualité, dans le sens de la

modélisation pluie-débit, de l'information pluviométrique ? D'autres tests devraient le déterminer. Il serait notamment important d'examiner le coefficient de corrélation de tous les postes pluviométriques d'un bassin versant et la relation entre les bassins à pluie 'hétérogène' et les performances de ces stratégies. Par ailleurs, il serait intéressant d'explorer d'autres stratégies de semi-globalisation / semi-distribution des vecteurs de paramètres en ne distribuant que certains, afin d'assurer une bonne robustesse des vecteurs optimisés.

Enfin, de tels tests sur la valorisation de l'information contenue dans diverses séries de pluie doivent s'appuyer sur des chroniques de bonne qualité. Ainsi, il faudrait également vérifier de manière plus exhaustive ces chroniques. Idéalement, il faudrait rallonger la durée de ces chroniques afin de pouvoir optimiser les paramètres des modèles sur des durées plus longues. Enfin, ces tests devraient être réalisés au pas de temps journalier afin de vérifier si la pertinence de la distribution de la pluie ne dépend pas du pas de temps des données et de la distance de corrélation du champ de pluie.

Conclusion générale

« *Tout objet sorti de son contexte est une œuvre d'art* »
d'après Marcel Duchamp

Conclusion générale

Au cours des vingt dernières années, une chaîne de modèles fiables, robustes et à vocation opérationnelle a été développée au Cemagref pour les pas de temps annuel, mensuel et journalier. Cette thèse a permis de compléter cette chaîne et de proposer un modèle pour un pas de temps plus fin, le pas de temps horaire.

Nos recherches s'inscrivent dans le cadre comparatif précédemment utilisé par *Perrin* [2000]. Notre approche s'est focalisée sur le développement empirique de structures de modèles et sur la comparaison de modèles conceptuels ou empiriques au pas de temps horaire. Face à la diversité des bassins versants, à la diversité de leurs représentations mathématiques (modèles) et aux incertitudes inhérentes à la modélisation hydrologique, la comparaison de différents modèles hydrologiques sur un grand nombre de bassins versants s'est révélée être un outil efficace pour évaluer la capacité des modèles à représenter la transformation des pluies en débit à l'échelle du bassin versant. Les performances de différents modèles ont été comparées, au pas de temps horaire, sur un large échantillon de bassins versants, aux caractéristiques climatiques, hydrologiques et physiographiques très variées. Nous avons constitué à cet effet un échantillon inédit de 313 bassins versants situés principalement en France, mais également aux États-Unis, en Australie, en Slovénie et en Espagne. Les bassins versants de cet échantillon sont essentiellement de petite taille (médiane de 85 km²) et leurs temps de réponse aux pluies sont de quelques heures.

Cette thèse apporte des contributions significatives et inattendues dans le développement de modèles, leur optimisation, la recherche du plus performant, la cohérence des modèles retenus pour différentes échelles de temps et la recherche d'une exploitation maximale de l'information contenue dans les données de pluie.

Comment rechercher de meilleures structures de modèles ?

A partir du cadre comparatif développé par *Perrin* [2000], nous avons montré l'intérêt d'utiliser de grands échantillons de bassins versants pour comparer les performances de modèles pluie-débit. A l'aide d'un test statistique de comparaison de moyennes, nous avons ainsi montré qu'il faut souvent un échantillon d'au moins 100 bassins versants pour discriminer les performances de différents modèles pluie-débit.

Nous avons également montré que la formulation du critère de *Nash et Sutcliffe* [1970] ne permet pas de comparer les performances de différents modèles, du fait de valeurs très fortement négatives. Nous avons proposé une autre formulation de ce critère, appelée le C_{2M} , dont la formulation permet de mieux comparer les performances de différents modèles sur un grand échantillon de bassins versants.

Comment passer du pas de temps journalier au pas de temps horaire ?

Nous nous sommes appuyés sur la structure du modèle GR4J [Perrin, 2000] et sur les voies de recherche ouvertes par les travaux de *Mouelhi* [2003] et de *Jothityangkoon et al.* [2001]. Les travaux de *Mouelhi* [2003], montrent une modification drastique de la structure des modèles lorsqu'on réduit le pas de temps et constatent une plus grande sophistication de la partie transfert des modèles à pas de temps plus fin.

Nous avons donc commencé par diversifier la fonction de routage du modèle. Un premier résultat surprenant a été le relatif échec de cette tentative. Nous avons alors estimé que le pas de temps horaire pouvait être l'occasion de prendre en compte le rôle de l'intensité des pluies. De façon également inattendue, les solutions que nous avons testées ont échoué et cette hypothèse de travail a dû être écartée. Finalement, contre toute attente, c'est une légère simplification de la structure de GR4J qui s'est révélée l'approche la plus efficace pour aborder le pas de temps horaire. Pourtant, dans le cadre de la comparaison de modèles issus de la littérature, c'est un modèle beaucoup plus complexe (dix paramètres et quatre réservoirs) qui apparaissait initialement comme le plus efficace. Mais, dans ce cas aussi, une réduction à six du nombre de paramètres à caler a permis de maintenir ou d'améliorer ses performances. Les travaux réalisés au pas de temps horaire nous permettent donc de confirmer que quatre à six paramètres semblent suffisants dans un modèle pluie-débit à l'échelle du bassin versant.

Ensuite, nous sommes allé du pas de temps horaire vers le pas de temps journalier. De manière inattendue, la structure proposée au pas de temps horaire, GR4H, s'est révélée également performante au pas de temps journalier. Ainsi, nous proposons d'utiliser une structure similaire pour les deux pas de temps avec seulement une constante à adapter en fonction du pas de temps. Cette nouvelle structure, appelée GR4X1, est la plus efficace parmi les modèles à quatre paramètres, à la fois au pas de temps horaire et au pas de temps journalier.

Entre ces deux pas de temps, les valeurs de paramètres diffèrent. Cependant, il a été montré qu'il est possible d'estimer les paramètres d'un modèle horaire, à partir

de ceux calés au pas de temps journalier. C'est d'ailleurs la première fois qu'un transfert de paramètres d'un pas de temps à un autre a été réussi avec une telle efficacité.

Faut-il recourir à des méthodes de calage plus sophistiquées ?

Ayant eu affaire à des modèles de niveaux de complexité très différents, on peut se demander si la méthode de calage n'est pas responsable de certains échecs. Pour cela, nous avons comparé les performances en calage et en validation de modèles de complexité croissante, dont les vecteurs de paramètres ont été optimisés d'une part avec notre méthode locale d'optimisation et d'autre part, avec deux méthodes globales très puissantes. Contre toute attente, nos résultats montrent que les modèles obtiennent des performances sensiblement meilleures en validation lorsque leurs paramètres sont optimisés avec notre méthode locale, alors que les méthodes globales testées, nécessitent un temps de calcul 10 à 100 fois supérieur.

Nos tests ont également révélé que l'intérêt opérationnel des méthodes globales était largement surévalué. En effet, comparée à l'incertitude sur la structure des modèles hydrologiques, l'incertitude sur les paramètres est négligeable. Le vecteur optimal n'est pas plus représentatif du comportement d'un bassin versant que d'autres vecteurs situés dans une proche région de l'espace des paramètres. Ces résultats ont clairement mis en évidence que la difficulté à définir les paramètres d'un modèle ne venait pas de la méthode d'optimisation utilisée, mais tout simplement de la plus ou moins bonne adéquation de sa structure.

Que reste-t-il à faire ?

- Les différents modèles de la chaîne GR sont maintenant disponibles pour les principaux pas de temps qui intéressent les utilisateurs. Il conviendrait d'augmenter la cohérence de cette chaîne de modélisation et d'approfondir son point le plus faible qui semble être la modélisation des échanges non atmosphériques.
- La deuxième voie de recherche devrait aborder la distribution des pluies. Il faudra identifier si la stratégie de prise en compte de l'hétérogénéité spatiale de la pluie que nous avons proposé est un début de modélisation distribuée ou simplement une méthode de filtrage des données de pluie en fonction de leur pertinence pour la modélisation. Il sera également intéressant de comparer l'amplitude des gains apportés par la modification de la structure d'un modèle, à ceux apportés par la distribution temporelle et par la

distribution spatiale. Il sera également nécessaire de tester si ces gains de performance apportés par différents types de distributions sont additifs ou non.

- Enfin, une autre voie de recherche devrait examiner la pertinence des modèles développés au pas de temps horaire dans le cadre de la prévision des crues. Il sera important d'évaluer leur applicabilité pour cet objectif et de confronter ces résultats à ceux obtenus par *Tangara* [2005] au pas de temps journalier. De plus, en suivant les travaux de *Rojas-Serna* [2005], il serait intéressant de tester l'application d'un modèle horaire à des bassins non-jaugés (ou partiellement), en utilisant l'approche de pré-détermination des paramètres à partir du pas de temps journalier et en introduisant progressivement de l'information horaire, issue de quelques jaugeages ou de courtes chroniques de données.

Les retours d'expérience suite à la future mise en oeuvre opérationnelle des modèles que nous avons mis au point qui permettra d'identifier les points faibles sur lesquels il nous faudra faire porter nos efforts de recherche. Une modélisation empirique comme la nôtre ne peut se justifier que par ses résultats : elle doit donc rester toujours attentive aux messages et critiques que ne manqueront pas de lui faire parvenir ses utilisateurs.

Enfin, la plus grande leçon que m'a apportée cette approche empirique de développement de modèles hydrologiques est que les idées *a priori*, que l'on peut se faire sur la manière de représenter la relation pluie-débit, sont rarement celles qui conduisent aux meilleurs résultats. Il convient ainsi de rester humble, d'écouter les données, de remettre en cause ses hypothèses et de recommencer...

Références bibliographiques

- [1] Allred, B. et Haan, C.T., 1991. Variability of optimized parameter estimates based on observed record length. *Trans. of Amer. Soc. Agr. Eng.* 34(6), 2421-2426.
- [2] Anctil, F., Lauzon, N., Andréassian, V., Oudin, L. et Perrin, C., 2004a. Improvement of rainfall-runoff forecasts through mean areal rainfall optimization. *Journal of Hydrology*(submitted).
- [3] Anctil, F., Michel, C., Perrin, C. et Andréassian, V., 2004b. A soil moisture index for ANN stream flow forecasting. *Journal of Hydrology* 286(1-4), 155-167.
- [4] Andersson, L., 1992. Improvement of runoff models. What way to go? *Nordic hydrology* 23, 315-332.
- [5] Anderton, S., Latron, J., White, S., Llorens, P., Salvany, M.C., Gallart, F. et O'Connell, E., 2002. Internal validation of a physically-based distributed model using data from a Mediterranean mountain catchment. *Hydrology and Earth System Sciences* 6(1), 67-83.
- [6] Andréassian, V., 2002. Impact de l'évolution du couvert forestier sur le comportement hydrologique des bassins versants, Thèse de Doctorat, Thèse de Doctorat, Ecole Doctorale Géosciences et Ressources Naturelles, Université Pierre et Marie Curie Paris VI, Cemagref (Antony), Paris, France, 276 p.
- [7] Andréassian, V., 2005. Pourquoi les rivières débordent-elles ? Les Petites Pommés du Savoir. Le Pommier, Paris.
- [8] Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C. et Loumagne, C., 2004a. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: a theoretical study using chimera watersheds. *Water Resources Research* 40(5), W05209, doi: 10.1029/2003WR002854.
- [9] Andréassian, V., Perrin, C. et Michel, C., 2004b. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology* 286(1-4), 19-35.
- [10] Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. et Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* 250(1-4), 206-223.
- [11] Atkinson, S.E., Woods, R.A. et Sivapalan, M., 2002. Climate and landscape controls on water balance model complexity over changing

- timescales. *Water Resources Research* 38(12), 1314, doi:10.1029/2002WR001487.
- [12] Aubert, D., Loumagne, C., Oudin, L. et Le Hegarat-Masclé, S., 2003. Assimilation of soil moisture into hydrological models: the sequential method. *Canadian Journal of Remote Sensing* 29(6), 711-1717.
- [13] Babovic, V., 2005. Data mining in hydrology. *Hydrol. Process.* 19(7), 1511-1515.
- [14] Bergström, S., 1991. Principles and confidence in hydrological modelling. *Nordic Hydrology* 22, 123-136.
- [15] Bergström, S., 1995. The HBV model. In: V.P. Singh (Editor), *Computer Models of Watershed Hydrology*, Chapter 13. Water Resources Publications, pp. 443-476.
- [16] Bergström, S. et Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff model. *Nordic Hydrology* 4, 147-170.
- [17] Beven, K., 1989. Changing ideas in hydrology - The case of physically-based models. *Journal of Hydrology* 105, 157-172.
- [18] Beven, K., Lamb, R., Quinn, P., Romanovicz, R. et Freer, J., 1995. TOPMODEL. In: V.P. Singh (Editor), *Computer Models of Watershed Hydrology*, Chapter 18. Water Resources Publications, pp. 627-668.
- [19] Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* 16, 41-51.
- [20] Beven, K.J., 2001a. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences* 5(1), 1-12.
- [21] Beven, K.J., 2001b. On hypothesis testing in hydrology. *Hydrological Processes* 15(9), 1655-1657.
- [22] Beven, K.J. et Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* 24(1), 43-69.
- [23] Blackie, J.R. et Eeles, C.W.C., 1985. Lumped catchment models - Chapter 11. In: M.G. Anderson et T.P. Burt (Editors), *Hydrological forecasting*. John Wiley & Sons Ltd, pp. 311-345.
- [24] Blöschl, G. et Sivapalan, M., 1995. Scale issues in hydrological modelling - a review. *Hydrological Processes* 9(3-4), 251-290.
- [25] Bonvoisin, N.J. et Boorman, D.B., 1992. Daily rainfall-runoff modelling as an aid to the transfer of hydrological parameters, Report to MAFF, Institute of Hydrology, Wallingford, UKp.
- [26] Boughton, W., 2005a. Calibrations of a daily rainfall-runoff model with poor quality data. *Environmental Modelling & Software* In press.
- [27] Boughton, W., 2005b. Catchment water balance modelling in Australia 1960-2004. *Agricultural water management* 71, 91-116.

- [28] Boughton, W. et Droop, O., 2003. Continuous simulation for design flood estimation--a review. *Environmental Modelling & Software* 18(4), 309-318.
- [29] Cameron, D.S., Beven, K.J., Tawn, J., Blazkova, S. et Naden, P., 1999. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *Journal of Hydrology* 219(3-4), 169-187.
- [30] Chiew, F. et McMahon, T., 1994. Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. *Journal of Hydrology* 153, 383-416.
- [31] Chiew, F.H.S., Stewardson, M.J. et McMahon, T.A., 1993. Comparison of six Rainfall-Runoff Modeling Approaches. *Journal of Hydrology* 147(1-4), 1-36.
- [32] Cooper, V.A., Nguyen, V.T.V. et Nicell, J.A., 1997. Evaluation of global optimization methods for conceptual rainfall-runoff model calibration. *Water Sci. Tech.* 36(5), 53-60.
- [33] Cormary, Y. et Guilbot, A., 1973. Etude des relations pluie-débit sur trois bassins versants d'investigation, IAHS Madrid Symposium. IAHS Publication n°108, pp. 265-279.
- [34] Crawford, N.H. et Linsley, R.K., 1966. Digital simulation in hydrology: Stanford Watershed Model IV.p.
- [35] Croke, B.F.W. et Jakeman, A.J., 2004. A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling & Software* 19(1), 1-5.
- [36] Datin, R., 1998. Outils opérationnels pour la prévision des crues rapides : traitement des incertitudes et intégration des prévisions météorologiques. Développements de TOPMODEL pour la prise en compte de la variabilité spatiale de la pluie. Application au bassin versant de l'Ardèche, Thèse de Doctorat, Ecole Doctorale Terre, Univers, Environnement, INPG, Grenoble, 370 p.
- [37] Dawdy, D.R. et Bergmann, J.M., 1969. Effect of rainfall variability on streamflow simulation. *Water Resources Research* 5(5), 958-966.
- [38] Dawson, C.W. et Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal* 43(1), 47-66.
- [39] de Marsily, G., 1994. Quelques réflexions sur l'utilisation des modèles en hydrologie. *Revue des Sciences de l'Eau* 7(3), 219-234.
- [40] Dong, X., 2005. Appropriate flow forecasting for reservoir operation. Ph. D. Thesis, University of Twente, Netherlands., 132 pp.
- [41] Dong, X., Dohmen-Janssen, M. et Booij, M., 2005. Appropriate spatial sampling of rainfall for flow simulation. *Hydrological Sciences - Journal- des Sciences Hydrologiques* 50(2), 279-298.
- [42] Dooge, J.C., I., 1973. Linear Theory of Hydrologic Systems. U. S. Department of Agriculture, Tech. Bull. n°1468, 327 p.

- [43] Duan, Q.Y., Gupta, V.K. et Sorooshian, S., 1993. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and Applications* 76(3), 501-521.
- [44] Duan, Q.Y., Sorooshian, S. et Gupta, V., 1992. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research* 28(4), 1015-1031.
- [45] Duan, Q.Y., Sorooshian, S. et Gupta, V.K., 1994. Optimal Use of the Sca-Ua Global Optimization Method for Calibrating Watershed Models. *Journal of Hydrology* 158(3-4), 265-284.
- [46] Eder, G., Sivapalan, M. et Nachtnebel, H.P., 2003. Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales. *Hydrological Processes*(11), 2125-2149.
- [47] Edijatno, 1991. Mise au point d'un modèle élémentaire pluie-débit au pas de temps journalier, Thèse de Doctorat, Université Louis Pasteur/ENGEEES, Strasbourg, 242 p.
- [48] Edijatno et Michel, C., 1989a. Recherche sur la mise au point d'un modèle journalier pluie-débit pour les petits bassins non jaugés, Rapport interne, Cemagrefp.
- [49] Edijatno et Michel, C., 1989b. Un modèle pluie-débit journalier à trois paramètres. *La Houille Blanche*(2), 113-121.
- [50] Edijatno, Nascimento, N.O., Yang, X., Makhlof, Z. et Michel, C., 1999. GR3J: a daily watershed model with three free parameters. *Hydrological Sciences Journal* 44(2), 263-277.
- [51] Franchini, M., 1996. Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal* 41(1), 21-39.
- [52] Franchini, M. et Galeati, G., 1997. Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal* 42(3), 357-379.
- [53] Franchini, M., Galeati, G. et Berra, S., 1998. Global optimization techniques for the calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal* 43(3), 443-458.
- [54] Franchini, M. et Pacciani, M., 1991. Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology* 122, 161-219.
- [55] Franchini, M., Wendling, J., Oblad, C. et Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. *Journal of Hydrology* 175, 293-338.
- [56] Gallart, F., Latron, J., Llorens, P. et Rabadà, d., 1997. Hydrological functioning of mediterranean mountain basins in Vallcebre, Catalonia: Some

- challenges for hydrological modelling. *Hydrological Processes* 11(9), 1263-1272.
- [57] Gallart, F., Llorens, P. et Latron, J., 1994. Studying the role of old agricultural terraces on runoff generation in a small Mediterranean mountainous basin. *Journal of Hydrology* 159(1-4), 291-303.
- [58] Gallart, F., Llorens, P., Latron, J. et Regüés, D., 2002. Hydrological processes and their seasonal controls in a small Mediterranean mountain catchment in the Pyrenees. *Hydrology and Earth System Sciences* 6(3), 527-537.
- [59] Gan, T.Y. et Biftu, G.F., 1996. Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resources Research* 32(12), 3513-3524.
- [60] Gan, T.Y., Dlamini, E.M. et Biftu, G.F., 1997. Effects of model complexity and structure, data quality and objective function on hydrologic modeling. *Journal of Hydrology* 192, 81-103.
- [61] Garçon, R., 1999. Modèle global Pluie-Débit pour la prévision et la prédétermination des crues (Lumped rainfall-runoff model for flood forecasting and design flood estimation, in french). *La Houille Blanche* 7/8, 88-95.
- [62] Garrick, M., Cunnane, C. et Nash, J.E., 1978. A criterion of efficiency for rainfall-runoff models. *Journal of Hydrology* 38, 375-381.
- [63] Gaucherel, C., 2002. Use of wavelet transform for temporal characterisation of remote watersheds. *Journal of Hydrology* 269(3-4), 101-121.
- [64] Green, I.R.A. et Stephenson, D., 1986. Criteria for comparison of single event models. *Hydrological Sciences Journal* 31(3), 395-411.
- [65] Gupta, V.K. et Sorooshian, S., 1983. Uniqueness and observability of conceptual rainfall-runoff model parameters: the percolation process examined. *Water Resources Research* 19(1), 269-276.
- [66] Holland, J.H., 1975. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, Michigan.
- [67] Hooke, R. et Jeeves, T.A., 1961. Direct search solutions of numerical and statistical problems. *J. Ass. Comput. Mach.* 8(2), 212-229.
- [68] Houghton-Carr, H.A., 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. *Hydrological Sciences Journal* 44(2), 237-261.
- [69] Hubert, P., 1984. *Eaupuscule. Une introduction à la gestion de l'eau*. Ellipses, Paris, 192 pp.
- [70] Ibbitt, R.P., 1972. Effects of random data errors on the parameter values for a conceptual model. *Water Resources Research* 8(1), 70-78.

- [71] Jakeman, A.J. et Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model ? *Water Resources Research* 29(8), 2637-2649.
- [72] Jakeman, A.J., Littlewood, I.G. et Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology* 117, 275-300.
- [73] Jayawardena, A.W. et Zhou, M.C., 2000. A modified spatial soil moisture storage capacity distribution curve for the Xinanjiang model. *Journal of Hydrology* 227, 93-113.
- [74] Johnston, P.R. et Pilgrim, D.H., 1976. Parameter optimization for watershed models. *Water Resources Research* 12(3), 477-486.
- [75] Jothityangkoon, C. et Sivapalan, M., 2001. Temporal scales of rainfall-runoff processes and spatial scaling of flood peaks: space-time connection through catchment water balance. *Advances in Water Resources* 24(9-10), 1015-1036.
- [76] Jothityangkoon, C., Sivapalan, M. et Farmer, D.L., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. *Journal of Hydrology* 254(1-4), 174-198.
- [77] Joukainen, S., 2000. Improving the calculation of potential evapotranspiration of the HBV model: application to the Ounasjoki watershed. In: Torbjörn Nilsson (Editor), *Nordic hydrological conference 2000. Nordic Hydrological Programme report*. Swedish Hydrological Council, Uppsala, pp. 347-354.
- [78] Kauark-Leite, L.A. et Nascimento, N.O., 1993. Développement, utilisation et incertitudes des modèles conceptuels en hydrologie, Séminaire Eaux - Environnement n°3, Modélisation du comportement des polluants dans les hydrosystèmes, Paris, pp. 191-219.
- [79] Kirkpatrick, S., Gelatt, C.D. et Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220(4598), 671-680.
- [80] Klemeš, V., 1983. Conceptualisation and scale in hydrology. *Journal of Hydrology* 65, 1-23.
- [81] Klemeš, V., 1986a. Dilettantism in hydrology: transition or destiny ? *Water Resources Research* 22(9), 177S-188S.
- [82] Klemeš, V., 1986b. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31(1), 13-24.
- [83] Klemeš, V., 1988. A hydrological perspective. *Journal of Hydrology* 100, 3-28.

- [84] Kobold, M. et Brilly, M., 2004. Calibration and verification of the HBV model with daily and hourly data. *Geophysical Research Abstracts* 6.
- [85] Kobold, M., Suselj, K. et Brilly, M., 2003. Sensitivity of the hydrological model to the rainfall error. *Geophysical Research Abstracts* 5.
- [86] Kokkonen, T.S. et Jakeman, A.J., 2001. A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resources Research* 37(9), 2345-2352.
- [87] Kuczera, G., 1997. Efficient subspace probabilistic parameter optimization for catchment model. *Water Resources Research* 33(1), 177-185.
- [88] Kuzmin, V., Seo, D.-J. et Koren, V., 2004. Fast and efficient optimization of hydrologic model parameter using a priori estimates and stepwise line search. *Journal of Hydrology* (Soumis en 09/2004).
- [89] Lamb, R., 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resources Research* 35(10), 3103-3114.
- [90] Le Lay, M., Galle, S., Saulnier, G.-M. et Braud, I., 2005. Detecting changes in watershed behavior: a model-based methodology. Application on Wets African non-stationary conditions. En préparation.
- [91] Le Moine, N., 2005. Déterminants du bilan d'eau à l'échelle du bassin versant : vers une estimation réaliste des pertes des bassins de surface. Cemagref Antony., Master 2 Sciences de l'Univers -Ecologie Environnement. Parcours hydrologie hydrogéologie. Université Pierre et Marie Curie, Paris VI., Paris., 52 pp.
- [92] Lidén, R. et Harlin, J., 2000. Analysis of conceptual rainfall-runoff modelling performance in different climates. *Journal of Hydrology* 238, 231-247.
- [93] Lindström, G., 1997. A simple automatic calibration routine for the HBV model. *Nordic Hydrology* 28, 153-168.
- [94] Lindström, G., Johansson, B., Persson, M., Gardelin, M. et Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* 201, 272-288.
- [95] Linsley, R.K., 1982. Rainfall-runoff models - an overview. In: V. P. Singh (Editor), *Proceedings of the international symposium on rainfall-runoff modelling*. Water Resources Publications, Littleton, CO, USA, pp. 3-22.
- [96] Loague, K.M. et Freeze, R.A., 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research* 21(2), 229-248.
- [97] Loumagne, C., Chkir, N., Normand, M., Otlé, C. et Vidal-Madjar, D., 1996. Introduction of soil/vegetation/atmosphere continuum in a conceptual rainfall-runoff model. *Hydrological Sciences Journal* 41(6), 889-902.

- [98] Luce, C.H. et Cundy, T.W., 1994. Parameter identification for a runoff model for forest roads. *Water Resources Research* 30(4), 1057-1069.
- [99] Mandelbrot, B. et Wallis, J., 1968. Noah, Joseph and Operational Hydrology. *Water Resources Research*, 4 (5), 909-918.
- [100] Makhlouf, Z. et Michel, C., 1994. A two-parameter monthly water balance model for French watersheds. *Journal of Hydrology* 162(3-4), 299-318.
- [101] Martinec, J. et Rango, A., 1989. Merits of statistical criteria for the performance of hydrological models. *Water Resources Bulletin* 25(2), 421-432.
- [102] Mathevet, T., Lepiller, M. et Mangin, A., 2004. Application of time-series analyses to the hydrological functioning of an Alpine karstic system: the case of Bange-L'Eau-Morte. *Hydrology and Earth System Sciences* 8(6), 1051 - 1064.
- [103] Mazenc, B., Sanchez, M. et Thiery, D., 1984. Analyse de l'influence de la physiographie d'un bassin versant sur les paramètres d'un modèle hydrologique global et sur les débits caractéristiques à l'exutoire. *Journal of Hydrology* 69, 97-118.
- [104] Mein, R.G. et Brown, B.M., 1978. Sensitivity of optimized parameters in watershed models. *Water Resources Research* 14(2), 299-303.
- [105] Merz, R. et Blöschl, G., 2004. Regionalisation of catchments parameters. *Journal of Hydrology* 287, 95-123.
- [106] Merz, R. et Blöschl, G., 2005. Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology* 302(1-4), 283-306.
- [107] Michaud, J. et Sorooshian, S., 1994. Comparison of simple versus complex distributed runoff models on a mid-sized semiarid watershed. *Water Resources Research* 30(3), 593-605.
- [108] Michel, C., 1983. Que peut-on faire en hydrologie avec un modèle conceptuel à un seul paramètre ? *La Houille Blanche*(1), 39-44.
- [109] Michel, C., 1989. *Hydrologie appliquée aux petits bassins versants ruraux*, Cemagref, Antony, 528 p.
- [110] Michel, C., Perrin, C. et Andreassian, V., 2003. The exponential store : a correct formulation for rainfall-runoff modelling. *Hydrological Sciences Journal* 48(1).
- [111] Moore, I.D. et Mein, R.G., 1975. An evaluation of three rainfall-runoff models. In: *Inst. Eng. Aust. (Editor), Hydrological Symposium, Sydney*, pp. 122-126.
- [112] Moore, R.J. et Clarke, R.T., 1981. A distribution function approach to rainfall-runoff modeling. *Water Resources Research* 17(5), 1367-1382.
- [113] Moreda, F., 1999. Conceptual rainfall-runoff models for different time steps, with special consideration for semi-arid and arid catchments. Thèse de

- doctorat (en anglais)., Laboratory of Hydrology. Faculty of Applied Sciences. Vrije Universiteit Brussel, Brussels, 229 pp.
- [114] Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier, Thèse de Doctorat, Ecole Doctorale Géosciences et Ressources Naturelles, Ecole Nationale du Génie Rural des Eaux et des Forêts, Cemagref (Antony), Paris, France, 323 p.
- [115] Mouelhi, S., Claude, M., Charles, P. et Vazken, A., 2005. Stepwise development of a two-parameter monthly water balance model. Submitted to Journal of Hydrology.
- [116] Moussavi, M. et Feyen, J., 1990. Comparison of different structures for a monthly water yield model in seasonally snow-covered mountainous watersheds of Iran. *Hydrological Sciences Journal* 35(5), 535-546.
- [117] Naef, F., 1981. Can we model the rainfall-runoff process today ? *Hydrological Sciences Bulletin* 26(3), 281-289.
- [118] Nandakumar, N. et Mein, R.G., 1997. Uncertainty in rainfall-runoff model simulations and the implications for predicting the hydrologic effects of land-use change. *Journal of Hydrology* 192(1-4), 211-232.
- [119] Nascimento, N.d.O., 1995. Appréciation à l'aide d'un modèle empirique des effets d'action anthropique sur la relation pluie-débit à l'échelle du bassin versant, thèse de doctorat, CERGRENE/ENPC, Paris, 550p pp.
- [120] Nash, J.E. et Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I - A discussion of principles. *Journal of Hydrology* 27(3), 282-290.
- [121] Ndiritu, J.G. et Daniell, T.M., 1999. Assessing model calibration adequacy via global optimisation. *Water SA* 25(3), 317-326.
- [122] Ndiritu, J.G. et Daniell, T.M., 2001. An improved algorithm for rainfall-runoff model calibration and function optimisation. *Mathematical and Computer Modelling* 33, 695-706.
- [123] Nejmeddine, M., 2004. Intégration de l'information de la distribution spatiale des pluie dans un modèle pluie-débit global (GR4J), DEA HHGG. Filière Géochimie. Université Paris Sud., Orsay., 85 pp.
- [124] Nelder, J.A. et Mead, R., 1965. A Simplex method for function minimisation. *The Computer Journal* 7(4), 308-313.
- [125] Niel, H., Paturel, J.-E. et Servat, E., 2003. Study of parameter stability of a lumped hydrologic model in a context of climatic variability. *Journal of Hydrology* 278(1-4), 213-230.
- [126] Nielsen, S.A. et Hansen, E., 1973. Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology* 4, 171-190.

- [127] NOAA, 1993. Solar and Meteorological Surface Observation Network 1961-1990. EDIS, Federal Building, Asheville, NC.
- [128] Oddos, A., 2002. Intérêt d'une approche semi-distribuée par rapport à une approche globale en modélisation pluie-débit. DEA Thesis, ENGEEES, Strasbourg, 96 pp.
- [129] Oudin, L., 2004. Recherche d'un modèle d'évapotranspiration potentielle pertinent comme entrée d'un modèle pluie-débit global, Thèse de Doctorat, Ecole Doctorale Géosciences et Ressources Naturelles, Ecole Nationale du Génie Rural des Eaux et des Forêts, Cemagref (Antony), Paris, Francep.
- [130] Oudin, L., Andréassian, V., Mathevet, T., Perrin, C. et Claude, M., 2005a. Can we combine simulations produced by different model parameterizations to produce more accurate streamflow time series? En cours de soumission.
- [131] Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. et Loumagne, C., 2005b. 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.
- [132] Oudin, L., Michel, C. et Anctil, F., 2005c. Which potential evapotranspiration input for a lumped rainfall-runoff model? - Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology* 303(1-4), 275-289.
- [133] Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. et Michel, C., 2005d. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology* In press.
- [134] Oudin, L., Andréassian, V., Mathevet T., Perrin, C., Micel, C., 2006. Dynamic combination of rainfall-runoff models simulation obtained from different parameterizations, soumis à *Water Resources Research*.
- [135] Paquet, E., 2004. Modèle MORDOR à 10 paramètres, MOPEX Workshop n°4., Paris.
- [136] Parajka, A., Merz, R. et Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences* 9, 157-171.
- [137] Paturel, J.E., Servat, E. et Vassiliadis, A., 1995. Sensitivity of conceptual rainfall-runoff algorithms to errors in input data -- case of the GR2M model. *Journal of Hydrology* 168(1-4), 111-125.
- [138] Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. *Proc. R. Soc. London A193*, 120-145.
- [139] Penman, H.L., 1961. Weather, plant and soil factors in hydrology. *Weather* 16, 207-219.

- [140] Perrin, C., 2000. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative, Thèse de Doctorat, Ecole Doctorale Terre, Univers, Environnement, INPG, Grenoble, 530 p.
- [141] Perrin, C., Andréassian, V. et Michel, C., 2005. Simple benchmark models as a basis for model efficiency criteria. Submitted to Large Rivers.
- [142] Perrin, C., Michel, C. et Andréassian, V., 2001a. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology* 242(3-4), 275-301.
- [143] Perrin, C., Michel, C. et Andréassian, V., 2001b. Long-term low flow forecasting for French rivers by continuous rainfall-runoff modelling. In: I.G. Littlewood (Editor), Meeting of the British Hydrological Society on Continuous River Flow Simulation. BHS Occasional Paper n°13, Wallingford, UK, pp. 21-29.
- [144] Perrin, C., Michel, C. et Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* 279, 275-289.
- [145] Perrin, C., Michel, C. et Andréassian, V., 2004. The GR4J model : an empirical approach to rainfall-runoff modelling, MOPEX Workshop n°4., ENGREF, Paris, France.
- [146] Perrin, C., Michel, C. Rosique, J.-L., 2003. Modélisation pluie-débit et prévision des crues sur des bassins versants du Nord-Pas-de-Calais (Helpe Mineure, Helpe Majeure, Solre, Aa et Liane). Rapport d'étude, Cemagref (Antony)p.
- [147] Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. et DMIP Participants, 2004. Overall distributed model intercomparison project results. *Journal of Hydrology* 298(1-4), 27-60.
- [148] Refsgaard, J.C. et Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resources Research* 32(7), 2189-2202.
- [149] Roche, M., 1963. Hydrologie de surface. ORSTOM. Gauthier Villars Editeur., Paris, 430 pp.
- [150] Rojas-Serna, 2005. Quelle connaissance minimale des débits pour définir les paramètres d'un modèle pluie-débit ?, Thèse de Doctorat, Ecole Doctorale Géosciences et Ressources Naturelles, Ecole Nationale du Génie Rural des Eaux et des Forêts, Cemagref (Antony), Paris, Francep.
- [151] Rosenbrock, H.H., 1960. An automatic method for finding the greatest or least value of a function. *The Computer Journal* 3, 175-184.

- [152] SAFEGE et CEMAGREF, 2002. Modélisation hydrologique sur les bassins versants amont de l'Oise, de l'Aisne et de l'Aire. Rapport final., SAFEGE (Nanterre), CEMAGREF (Antony), 156 p. + annexes p.
- [153] SAFEGE et CEMAGREF, 2003. Modélisation hydrologique sur les bassins versants de la Sarthe et de l'Huisne. Rapport final, SAFEGE (Nanterre), CEMAGREF (Antony)p.
- [154] Saulnier, G.-M., 1996. Informations pédologiques spatialisées et traitements topographiques améliorés dans la modélisation hydrologique par TOPMODEL, Thèse de Doctorat, Ecole Doctorale Terre, Univers, Environnement, INPG, Grenoblep.
- [155] Schaake, J., Duan, Q., Smith, M. et Koren, V., 2000. Criteria to select basins for hydrologic model development and testing. Conference on Hydrology, AMS, Long Beach, CA, USA.(4).
- [156] Schreider, S.Y. et Jakeman, A.J., 2001. Streamflow modelling on a subdaily time step in the upper Murray basin. *Mathematical and Computer Modelling* 33(6-7), 659-668.
- [157] Schreider, S.Y., Jakeman, A.J., Letcher, R.A., Nathan, R.J., Neal, B.P. et Beavis, S.G., 2002. Detecting changes in streamflow response to changes in nonclimatic catchment conditions: farm dam development in the Murray-Darling basin, Australia. *Journal of Hydrology* 262(1-4), 84-98.
- [158] Seibert, J., 2001. On the need for benchmarks in hydrological modelling. *Hydrological Processes* 15(6), 1063-1064.
- [159] Sivapalan, M., Blöschl, G., Zhang, L. et Vertessy, R., 2003. Downward approach to hydrological prediction. *Hydrological Processes* 17(11), 2101-2111.
- [160] Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F. et Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology* 298(1-4), 4-26.
- [161] Sorooshian, S., Duan, Q.Y. et Gupta, V.K., 1993. Calibration of Rainfall-Runoff Models - Application of Global Optimization to the Sacramento Soil-Moisture Accounting Model. *Water Resources Research* 29(4), 1185-1194.
- [162] Sorooshian, S. et Gupta, V.K., 1995. Model calibration. In: V.P. Singh (Editor), *Computer models of watershed hydrology*, Chapter 2. Water Resources Publications, pp. 23-68.
- [163] Sugawara, M., 1974. Tank model and its application to Bird Creek, Wollombi Brook, Bikin River, Kitsu River, Sanga River and Nam Mune. Research note, N° 11., National Research Center for Disaster Prevention, Kyoto, Japan., 1-64 p.

- [164] Sugawara, M., 1995. Tank model. In: V.P. Singh (Editor), *Computer Models of Watershed Hydrology*, Chapter 7. Water Resources Publications, pp. 165-214.
- [165] Summer, N.R., Fleming, P.M. et Bates, B.C., 1997. Calibration of a modified SFB model for twenty-five Australian catchment using simulated annealing. *Journal of Hydrology* 197, 166-188.
- [166] Takeuchi, K., 2002. Increasing vulnerability to extreme floods and societal needs of hydrological forecasting. *Hydrological Sciences Journal* 46(6), 869-881.
- [167] Tan, B.Q. et O'Connor, K.M., 1996. Application of an empirical infiltration equation in the SMAR conceptual model. *Journal of Hydrology* 185, 275-295.
- [168] Tanakamaru, H., 1995. Parameter estimation for the Tank Model using global optimisation. *Trans. JSIDRE* 178, 103-112.
- [169] Tangara, M., 2005. Amélioration d'une méthode de prévision de crue utilisant un modèle global pluie-débit., Thèse de Doctorat, EPHE, Paris, Francep.
- [170] Thiery, D., 1982. Utilisation d'un modèle global pour identifier sur un niveau piézométrique des influences multiples dues à diverses activités humaines. *IAHS Publication n°136*, 71-77.
- [171] Thornthwaite, C.W. et Mather, J.R., 1955. The water balance. *Publ. Climatol. Lab. Climatol. Drexel. Inst. Techol.* 8(1), 1-104.
- [172] Thyer, M., Kuczera, G. et Bates, B.C., 1999. Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resources Research* 35(3), 767-773.
- [173] Torrence, C. et Compo, G.P., 1998. A practical guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.* 79, 61-78.
- [174] Troutman, B.M., 1982. An analysis of input errors in precipitation-runoff models using regression with errors in the independent variables. *Water Resources Research* 19(3), 791-810.
- [175] Troutman, B.M., 1983. Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resources Research* 21(8), 1217-1222.
- [176] UNESCO-OMM, 1992. *Glossaire international d'hydrologie*. Deuxième édition quadrilingue (EN,ES,FR,RU)., 413 pp.
- [177] Vandewiele, G.L., Xu, C.Y. et Win, N.L., 1992. Methodology and comparative study of monthly models in Belgium, China and Burma. *Journal of Hydrology* 134, 315-347.
- [178] Wagener, T., 2003. Evaluation of catchment models. *Hydrological Processes* 17(16), 3375-3378.

- [179] Wang, Q.J., 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resources Research* 27(9), 2467-2471.
- [180] Wang, Q.J., 1997. Using genetic algorithms to optimise model parameters. *Environmental Modelling and Software* 12(1), 27-34.
- [181] Warmerdam, P.M.M., Kole, J. et Chormanski, J., 1997. Modelling rainfall-runoff processes in the Hupselse Beek research basin. In: D. Viville and I.G. Littlewood (Editor), *Ecohydrological processes in small basins*. IHP-V, Technical Documents in Hydrology n°14, UNESCO, Paris, Proceedings of the Strasbourg Conference (24-26 September 1996), pp. 155-160.
- [182] Weeks, W.D. et Hebbert, R.H.B., 1980. A comparison of rainfall-runoff models. *Nordic Hydrology* 11, 7-24.
- [183] Weglarczyk, S., 1998. The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology* 206, 98-103.
- [184] Wilcox, B.P., Rawls, W.J., Brakensiek, D.L. et Wight, J.R., 1990. Predicting runoff from rangeland catchments: a comparison of two models. *Water Resources Research* 26(10), 2401-2410.
- [185] WMO, 1975. Intercomparison of conceptual models used in operational hydrological forecasting, *Operational Hydrology Report n° 7*, WMO n°429, World Meteorological Organization, Geneva, Switzerlandp.
- [186] WMO, 1986. Intercomparison of models of snowmelt runoff, *Operational Hydrology Report n° 23*, WMO n°646, World Meteorological Organization, Geneva, Switzerlandp.
- [187] WMO, 1992. Simulated real-time intercomparison of hydrological models, *Operational Hydrology Report n° 38*, WMO 779, World Meteorological Organization, Geneva, Switzerlandp.
- [188] Xu, C.Y. et Vandewiele, G.L., 1995. Parsimonious monthly rainfall-runoff models for humid basins with different input requirements. *Advances in Water Resources* 18, 39-48.
- [189] Yang, D., Herath, S. et Musiak, K., 2000. Comparison of different distributed hydrological models for characterization of catchment spatial variability. *Hydrological Processes* 14(3), 403-416.
- [190] Yapo, P.O., Gupta, H.V. et Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* 181, 23-48.
- [191] Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M. et Jakeman, A.J., 1997. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. *Water Resources Research* 33(1), 153-166.

- [192] Yevjevich, V., 1972. Probability and statistics in Hydrology. Water Resources Publications., Littleton, Colorado, USA.
- [193] Young, P., 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes* 17(11), 2195-2217.
- [194] Young, P.C., Jakeman, A.J. et Post, D.A., 1997. Recent advances in the data-based modelling and analysis of hydrological systems. *Wat. Sci. Tech.* 36(5), 99-116.
- [195] Zhang, X. et Lindström, G., 1996. A comparative study of a Swedish and a Chinese hydrological model. *Water Resources Bulletin* 32(5), 985-994.
- [196] Zhao, R.J. et Liu, X.R., 1995. The Xinanjiang model. In: V.P. Singh (Editor), *Computer models of Watershed Hydrology*, Chapter 7. Water Resources Publications, pp. 215-232.
- [197] Zhao, R.J., Zuang, Y.L., Fang, L.R., Liu, X.R. et Zhang, Q.S., 1980. The Xinanjiang model. *IAHS Publication n°129*, 351-356.
- [198] Zin, I., 2002. Incertitudes et ambiguïté dans la modélisation hydrologique. Discussion, développements méthodologiques et application à l'hydrologie de crue en Ardèche/, Thèse de Doctorat, Ecole Doctorale Terre, Univers, Environnement, INPG, Grenoble.

Epilogue

Cette thèse a commencé par la descente de l'Amazone de Leticia à Belém. Elle s'est poursuivie par la découverte de la Vardar à Skopje, du lac d'Orhid, de quelques baignades dans le Danube à Wien et dans le joli torrent de l'Archiane, de petits tours dans la vallée du Chéran et à la source de Bange-L'Eau-Morte... Enfin, elle s'est terminée par la découverte des chutes de Foz do Iguaçu, du Rio della Plata à Buenos Aeres, des chatoyantes rivières Matapédia et Cascapédia en Gaspésie.

Je souhaiterais remercier toutes les personnes qui m'ont permis de ne pas m'occuper seulement d'eau et d'hydrologie pendant ces trois années de thèse. En premier lieu, mes deux compagnons d'aventures et de galères, Thomas K. et To². A leurs manières, ils ont tous les deux contribué significativement aux résultats de cette thèse. Ils ont subis, avec enthousiasme et application, mon besoin nécessaire de dépenser de l'énergie pour deux, d'une manière ou d'une autre.

Les Renards, illustres bons-vivants et spécialistes du concept de la Randonnée Cassoulet, ont également été de très proches collaborateurs au cours des RAR 1 et 2 : Ungérable JP from Vesoul et Ronan Renard Chambreur, amateurs de one-to-one et de rugby nocturnes... Je regrette encore de leur avoir fait faux bond à la station Odéon... ils auraient pu me sauver de ce mauvais pas néanmoins...

Je souhaite également évoquer ici la création et le développement du Club International d'Hydrologie Social. Mon investissement dévoué pour ce club m'a permis de passer de grands moments d'hydrologie sociale, grâce à la participation de tous et de nos maitres spiritueux : François et Pierre, Noémi, Julio, Matholo Condom, Julien, Anne, Rémy, Toto, Angelbert, Marcos, Edson, l'incroyable Jean Mich, Sarah, Nico et sa Flipette, Véro, Georges-Marie, Ludo (4 figures en couleurs à Buenos Aeres !), Miss Béné, Marine sans les barbelés, Pauline, Pietro et ceux que j'oublie... Un grand Merci au Restaurant Chez Mamane (siège du Club) : Mamane, Majid, Dalila, Bronson et tous les piliers de comptoir ont contribué à animer cette thèse... Tous les judokas sociaux de l'ANDS : JB, Rémy, Greg, Valy et Thierry. Un grand merci à Laetitia pour m'avoir transmis sa passion du Judo et de la prospérité mutuelle. Une pensée pour Claire et son inoubliable mojette.

Mes pensées les plus chaleureuses vont également à mes ami(e)s de France et de Navarre : en premier lieu pour Le Seb, qui a essayé de monter le quintal sur de beaux glaciers... Fabrice de Die, Lucero del Pilar, Riko, Gars Manu, France, mon ami Glandut, Astrid, Anne-Jo, Anne L., Caro & Aurel, Bébert & Romain, Nico de

Bretagne, Niko Carayon, Nadège et Charli, Yann & Annabelle, Rodolphe, Nico, mes cousines du Fraysse, Christian B. et sa guitare. Toute mon affection et mon amour fraternel pour Cuca, Yoshi et Miguel. Un grand merci à tous les Errants et à Jean Leloup pour m'avoir accompagné quotidiennement avec sa Vallée des Réputations.

Enfin, mes dernières pensées vont pour mes proches : Renaud & Armelle, les petits (Théo, Hugo, Margot) ; la famille du club œno : Maher & Marie, Alain (jamais assez de sauce dans tes pâtes) et Mumu ; Alain ; les Courbis ;

Toute mon admiration pour mon papa infatigable.

La dernière ligne, pour ma maman, déjà disparue depuis trop longtemps, dont le manque quotidien est le marteau de mes vaines dépenses d'énergies.

ANNEXES

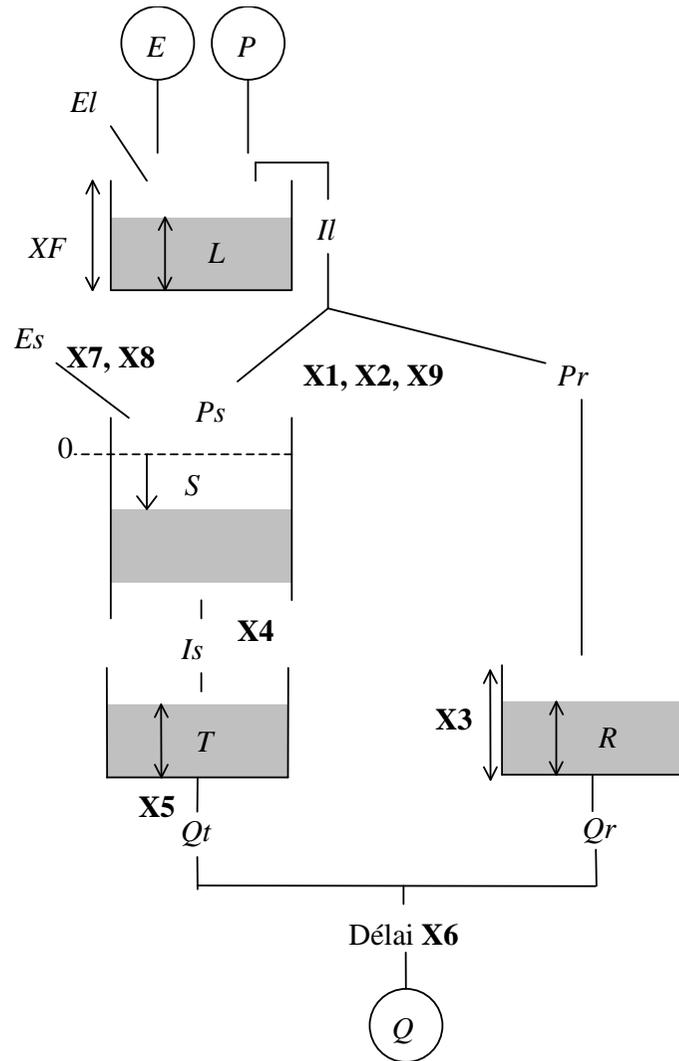
Annexe 1 Modèles hydrologiques

La description détaillée de la plupart des modèles hydrologiques utilisés dans cette thèse se trouve dans le thèse de Perrin [2000], dont sont extraits certaines des description de ces modèles.

Tableau 9-2 : Liste des modèles sélectionnés.

Modèle	Abréviation	Modèle	Abréviation
Institute of Hydrology lumped model [Blackie et Eeles, 1985]	BLAE	Model B [Bonvoisin et Boorman, 1992]	MODB
CREC [Cormary et Guilbot, 1973]	CREC	MORDOR6 (cette thèse)	MOR6
GARDENIA [Thiery, 1982]	GARD	MORDOR [Garçon, 1999; Paquet, 2004]	MORD
GR3H (cette thèse)	GR3H	NAM [Nielsen et Hansen, 1973]	NAM0
GR4H (cette thèse)	GR4H	PDM [Moore et Clarke, 1981]	PDM
GR5H (cette thèse)	GR5H	Bucket model [Thornthwaite et Mather, 1955]	THOR
GR4J [Perrin et al., 2003]	GR4J	TOPMO6 (cette thèse)	TOP6
HBV [Bergström, 1995]	HBV0	TOPMODEL [Beven et Kirkby, 1979]	TOP8
IHACRES [Jakeman et al., 1990]	IHAC	WAGENINGEN [Warmerdam et al., 1997]	WAGE
MARTINE [Mazenc et al., 1984]	MART	XINANJIANG [Zhao, 1992]	XINJ

Structure BLAE



$$L = L + P \qquad El = \min(L, E)$$

$$E' = E - El$$

$$Il = \max(0, L - XF) \qquad L = L - Il$$

$$Fr = X9 \cdot \left(\exp\left(\frac{S}{X1}\right) + \exp\left(\frac{Il}{X2}\right) - 1 \right)$$

$$Fr = \max(0, \min(1, Fr))$$

$$Pr = Fr \cdot P \qquad Ps = P - Pr$$

$$S = S - Ps$$

$$Is = \min(0, -S \cdot X4) \qquad S = S + Is$$

si $S < X7$, $Es = E'$

si $X7 \leq S < X8$, $Es = E' \cdot \frac{X8 - S}{X8 - X7}$

$$S = S + Es$$

$$T = T + Is \qquad Qt = T / X5$$

$$T = T - Qt$$

$$R = R + Pr$$

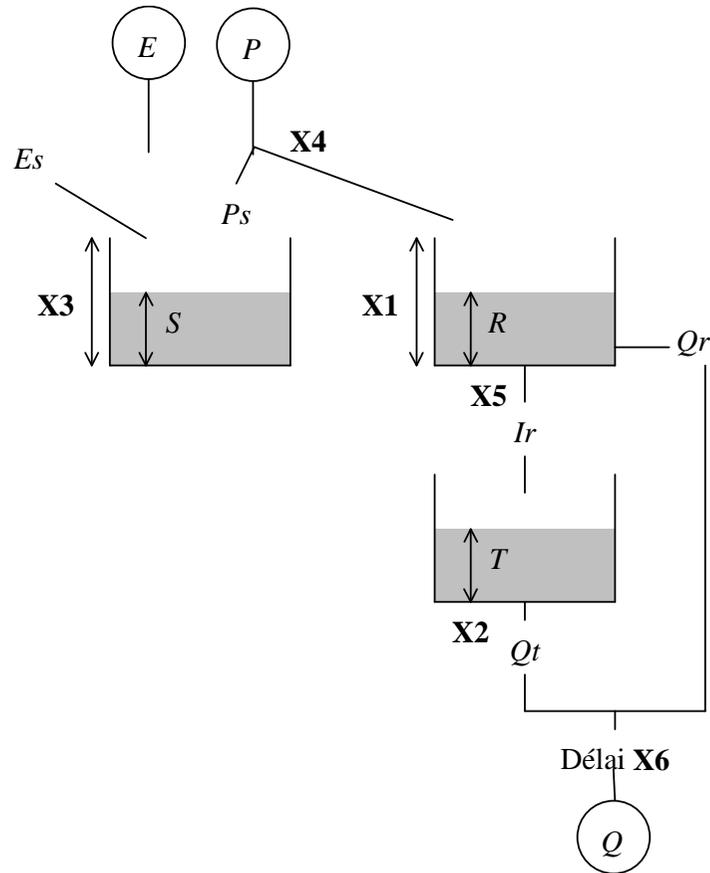
$$Qr = \frac{R^2}{R + X3}$$

$$R = R - Qr$$

$$Q = Qt + Qr$$

(avec translation dans le temps de X6 pas de temps)

Structure CREC



$$Pr = \frac{P}{1 + \exp\left(\frac{X3 - S}{X4}\right)}$$

$$Ps = P - Pr$$

$$S = S + Ps$$

$$Es = \frac{E}{1 - \exp\left(-\frac{S}{XF}\right)}$$

$$S = \max(0, S - Es)$$

$$R = R + Pr$$

$$Qr = \frac{R^2}{R + X1}$$

$$R = R - Qr$$

$$Ir = R / X5$$

$$R = R - Ir$$

$$T = T + Ir$$

$$Qt = T / X2$$

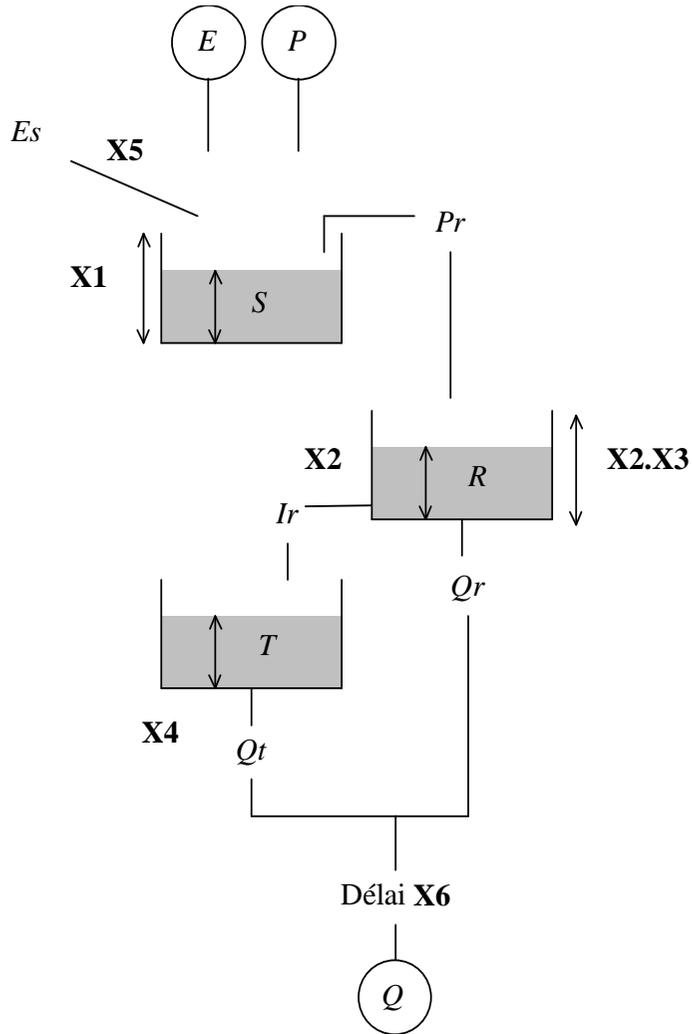
$$T = T - Qt$$

$$Q = Qr + Qt$$

(avec translation dans le temps de X6 pas de temps)

(avec $XF = 245$)

Structure GARD



$$S=S+P$$

$$Pr=\max(0,S-X1) \quad S=S-Pr$$

$$Es=X5.E \quad S=S-Es$$

$$R=R+Pr$$

$$Qr=\frac{R^2}{R+X2.X3} \quad R=R-Qr$$

$$Ir=R/X2 \quad R=R-Ir$$

$$T=T+Ir$$

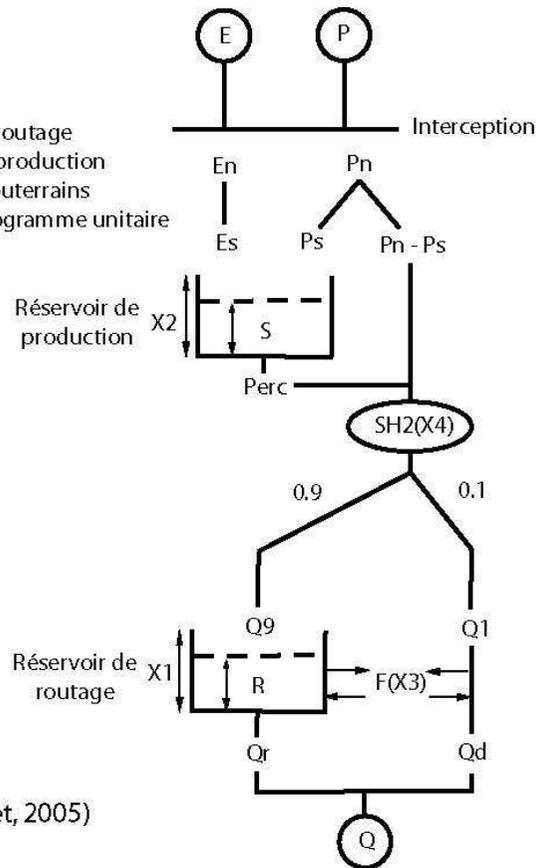
$$Qt=\frac{T}{X4} \quad T=T-Qt$$

$$Q = Qt + Qr$$

(avec translation dans le temps de X6 pas de temps)

Modèle GR3H, GR4H et GR4X1 (Mathevet, 2005)

X1 : Capacité du réservoir de routage
 X2 : Capacité du réservoir de production
 X3 : Coefficient d'échanges souterrains
 X4 : Temps de base de l'hydrogramme unitaire



GR4H (Mathevet, 2005)

Description de GR4H :

si $P \geq E$, $P_n = P - E$, $E_n = 0$, si $P < E$, $E_n = E - P$, $P_n = 0$

$$P_s = \frac{P_n \cdot \left(1 - \left(\frac{S}{X_2}\right)^2\right)}{1 + \frac{P_n}{X_2} \cdot \left(1 + \frac{S}{X_2}\right)}$$

$$E_s = \frac{E_n \cdot \frac{S}{X_2} \left(2 - \frac{S}{X_2}\right)}{1 + \frac{E_n}{X_2} \cdot \left(2 - \frac{S}{X_2}\right)}$$

$CPERC = 4$.

$$Perc = S - \left(S^{-4} + (CPERC \cdot X_2)^{-4}\right)^{-\frac{1}{4}}$$

$$S = S - Perc$$

$$0 \leq j \leq X_4, SH2(j) = \frac{1}{2} \left(\frac{j}{X_4}\right)^{\frac{5}{4}} \quad X_4 < j \leq 2 \cdot X_4, SH2(j) = 1 - \frac{1}{2} \left(2 - \frac{j}{X_4}\right)^{\frac{5}{4}}$$

$$j > 2 \cdot X_4, SH2(j) = 1$$

$$F = X_3 \cdot \left(\frac{R}{X_1}\right)^3 \quad Q_d = \max(0, Q_1 + F)$$

$$R = \max(\varepsilon, R + Q_9 + F)$$

$$Q_r = R - \left(R^{-4} + X_1^{-4}\right)^{-\frac{1}{4}}$$

$$R = R - Q_r$$

$$Q = Q_r + Q_d$$

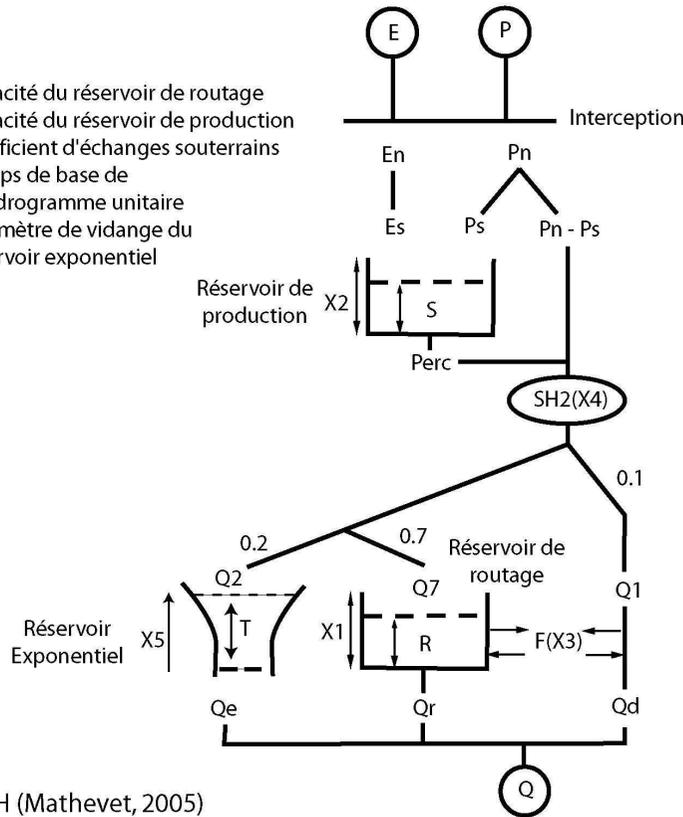
GR3H a la même structure que GR4H, mais X2 est fixé à 400 mm.

GR4X1 a la même structure que GR4H, mais au pas de temps journalier

$CPERC = 9./4$.

Modèle GR5H et GR5X1 (Mathevet, 2005)

- X1 : Capacité du réservoir de routage
- X2 : Capacité du réservoir de production
- X3 : Coefficient d'échanges souterrains
- X4 : Temps de base de l'hydrogramme unitaire
- X5 : Paramètre de vidange du réservoir exponentiel



GR5H (Mathevet, 2005)

si $P \geq E$, $Pn = P - E$, $En = 0$, si $P < E$, $En = E - P$, $Pn = 0$

$$Ps = \frac{Pn \cdot \left(1 - \left(\frac{S}{X2}\right)^2\right)}{1 + \frac{Pn}{X2} \cdot \left(1 + \frac{S}{X2}\right)}$$

$$Es = \frac{En \cdot \frac{S}{X2} \left(2 - \frac{S}{X2}\right)}{1 + \frac{En}{X2} \cdot \left(2 - \frac{S}{X2}\right)}$$

$$CPERC = 4. Perc = S - \left(S^{-4} + (CPERC \cdot X2)^{-4}\right)^{-\frac{1}{4}} \quad S = S - Perc$$

$$0 \leq j \leq X4, SH2(j) = \frac{1}{2} \left(\frac{j}{X4}\right)^{\frac{5}{4}} \quad X4 < j \leq 2 \cdot X4, SH2(j) = 1 - \frac{1}{2} \left(2 - \frac{j}{X4}\right)^{\frac{5}{4}}$$

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

$$F = X3 \cdot \left(\frac{R}{X1}\right)^3 \quad Qd = \max(0, Q1 + F) \quad R = \max(\epsilon, R + Q9 + F)$$

$$Qr = R - \left(R^{-4} + X1^{-4}\right)^{-\frac{1}{4}}$$

$$R = R - Qr$$

$$T = T + Q2 \quad Qt = X5 \cdot \exp\left(\frac{T}{X5}\right)$$

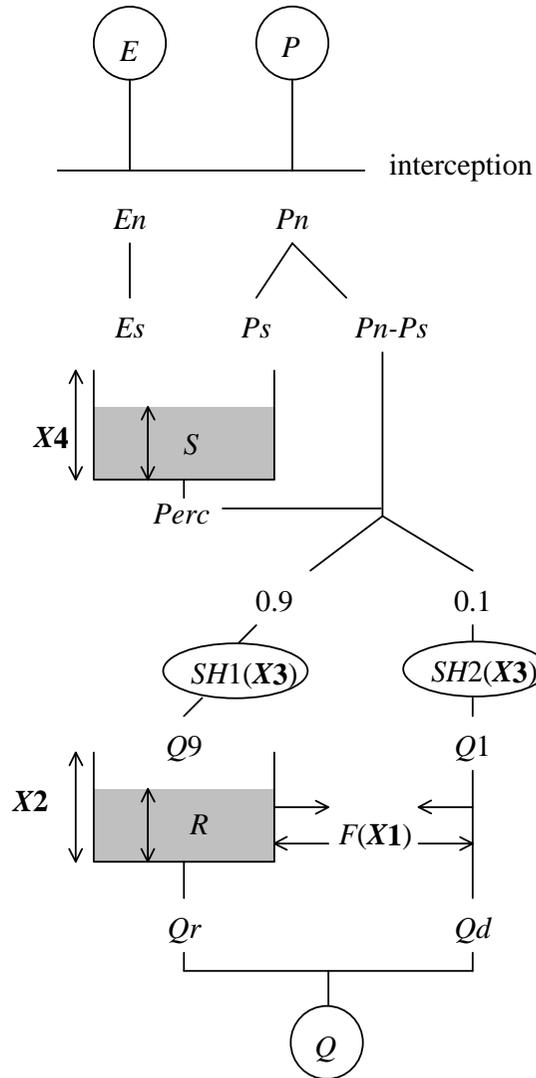
$$T = T - Qt$$

$$Q = Qr + Qd + Qt$$

GR5X1 a la même structure que GR5H, mais au pas de temps journalier

$CPERC = 9./4.$

Modèle GR4J (Perrin, 2000)



si $P \geq E$, $Pn = P - E$, $En = 0$, si $P < E$, $En = E - P$, $Pn = 0$

$$Ps = \frac{Pn \cdot \left(1 - \left(\frac{S}{X4}\right)^2\right)}{1 + \frac{Pn}{X4} \cdot \left(1 + \frac{S}{X4}\right)}$$

$$Es = \frac{En \cdot \frac{S}{X4} \cdot \left(2 - \frac{S}{X4}\right)}{1 + \frac{En}{X4} \cdot \left(2 - \frac{S}{X4}\right)}$$

$$Perc = S - \left(S^{-4} + \left(\frac{9}{4} \cdot X4\right)^{-4}\right)^{-\frac{1}{4}} \quad S = S - Perc$$

$$0 \leq j \leq X3, \quad SH1(j) = \left(\frac{j}{X3}\right)^{\frac{5}{2}} \quad j > X3, \quad SH1(j) = 1$$

$$0 \leq j \leq X3, \quad SH2(j) = \frac{1}{2} \left(\frac{j}{X3}\right)^{\frac{5}{2}} \quad X3 < j \leq 2 \cdot X3, \quad SH2(j) = 1 - \frac{1}{2} \left(2 - \frac{j}{X3}\right)^{\frac{5}{2}}$$

$$j > 2 \cdot X3, \quad SH2(j) = 1$$

$$F = X1 \cdot \left(\frac{R}{X2}\right)^{\frac{7}{2}}$$

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

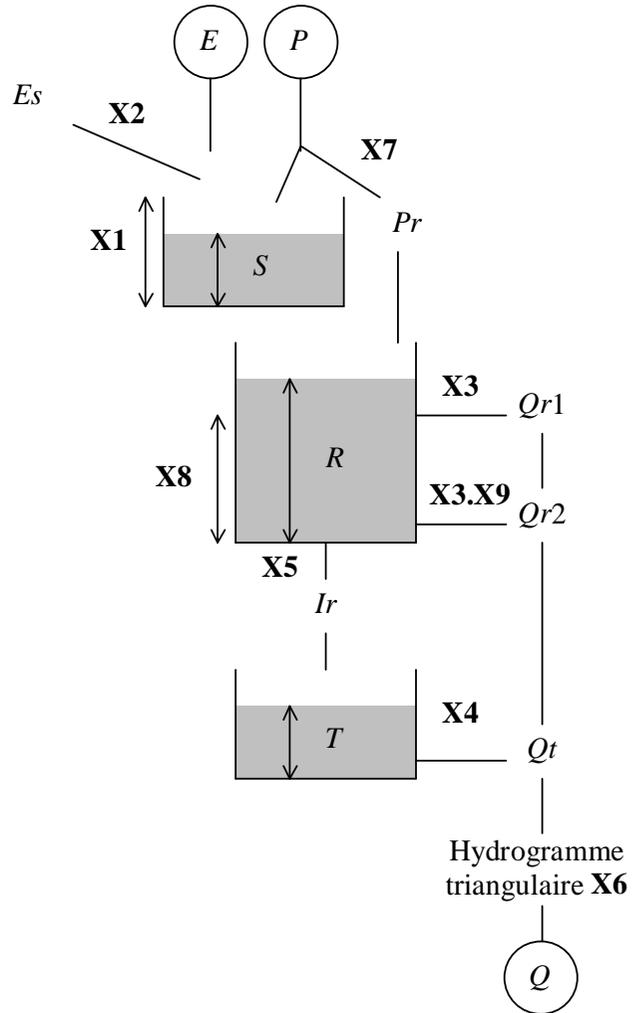
$$R = \max(\varepsilon, R + Q9 + F)$$

$$Qr = R - \left(R^{-4} + X2^{-4}\right)^{-\frac{1}{4}}$$

$$Q = Qr + Qd$$

$$R = R - Qr$$

Structure HBV0



$$Pr = 0$$

on réalise une boucle sur chaque pas de temps en décomposant le pas de temps en cinq sous pas de temps. On considère sur chacun d'eux un cinquième de la pluie et un cinquième de l'ETP

$$P5 = P / 5$$

$$E5 = E / 5$$

$$Pri = P5 \cdot \left(\min \left(1, \frac{S}{X1} \right) \right)^{X7}$$

$$Pr = Pr + Pri$$

$$S = S + (P5 - Pri)$$

$$Esi = \min \left(S, E \frac{S}{X1} \right)$$

$$S = S - Esi$$

(fin de boucle)

$$R = R + Pr$$

$$Qr1 = \max(0, (R - X8) / X3)$$

$$R = R - Qr1$$

$$Qr2 = R / (X3 \cdot X9)$$

$$R = R - Qr2$$

$$Ir = \min(S, X5)$$

$$S = S - Ir$$

$$T = T + Ir$$

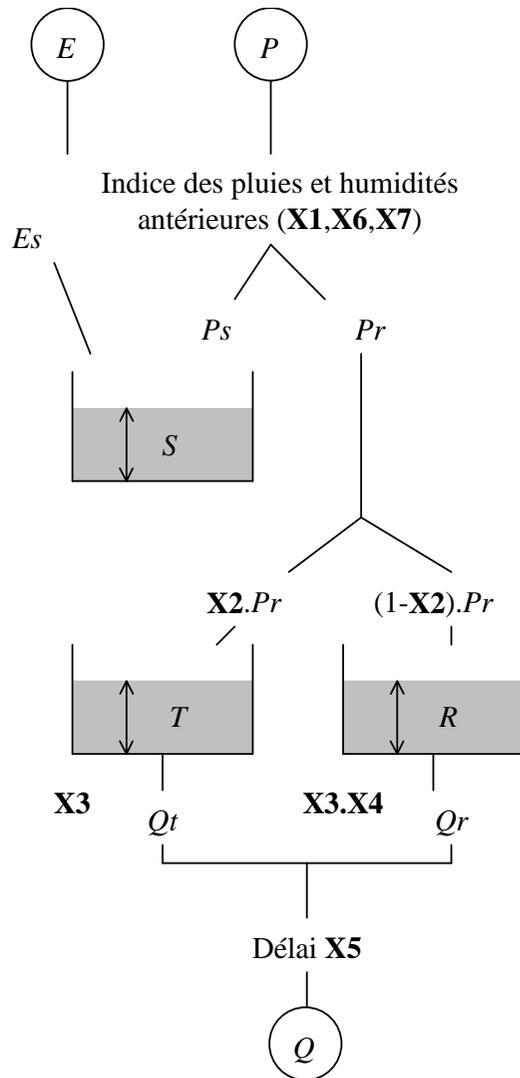
$$Qt = T / X4$$

$$T = T - Qt$$

$$Q = Qr1 + Qr2 + Qt$$

(avec routage par hydrogramme triangulaire de temps de base X6)

Structure IHAC



$$XS = S$$

$$E1 = \max(0., X7 - E / X6)$$

$$S = XS + \frac{P}{X1} - \frac{XS}{\exp(E1)}$$

$$Pr = \frac{1}{2}(XS + S).P$$

$$T = T + X2.Pr$$

$$R = R + (1 - X2).Pr$$

$$Qt = \frac{T}{X3}$$

$$Qr = \frac{R}{X3.X4}$$

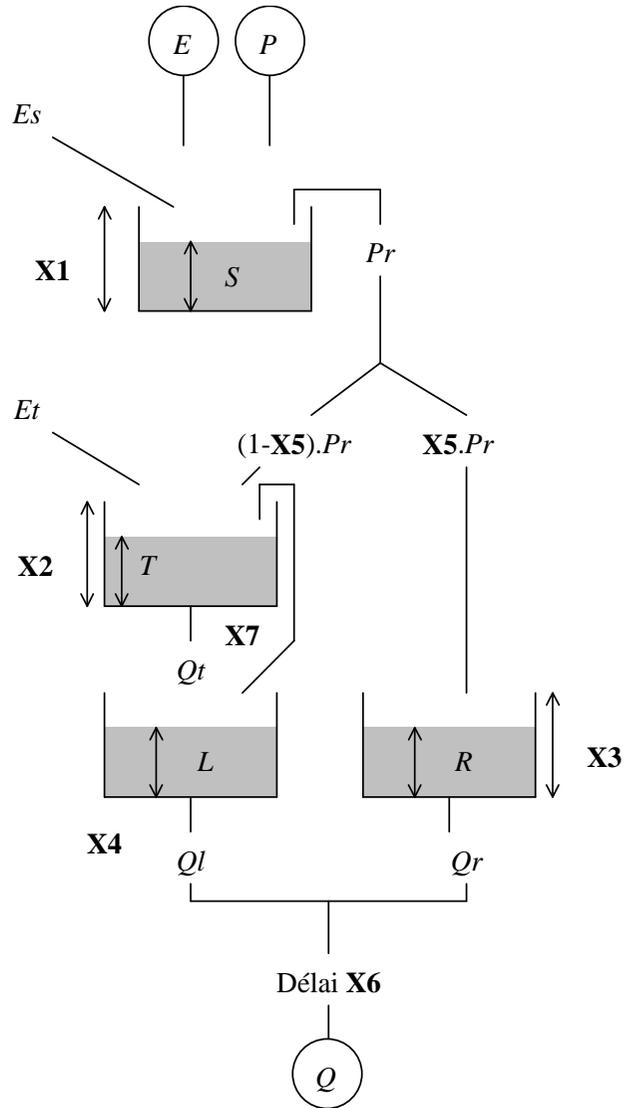
$$T = T - Qt$$

$$R = R - Qr$$

$$Q = Qt + Qr$$

(avec translation dans le temps de $X5$ pas de temps)

Structure MART



$$S = S + P$$

$$Pr = \max(0, S - X1)$$

$$S = S - Pr$$

$$Es = \min(S, E)$$

$$S = S - Es$$

$$Er = E - Es$$

$$R = R + X5.Pr$$

$$Qr = \frac{R^2}{R + X3}$$

$$R = R - Qr$$

$$T = T + (1 - X5).Pr$$

$$Qt1 = \frac{T}{X7}$$

$$T = T - Qt1$$

$$Qt2 = \max(0, T - X2)$$

$$T = T - Qt2$$

$$L = L + Qt1 + Qt2$$

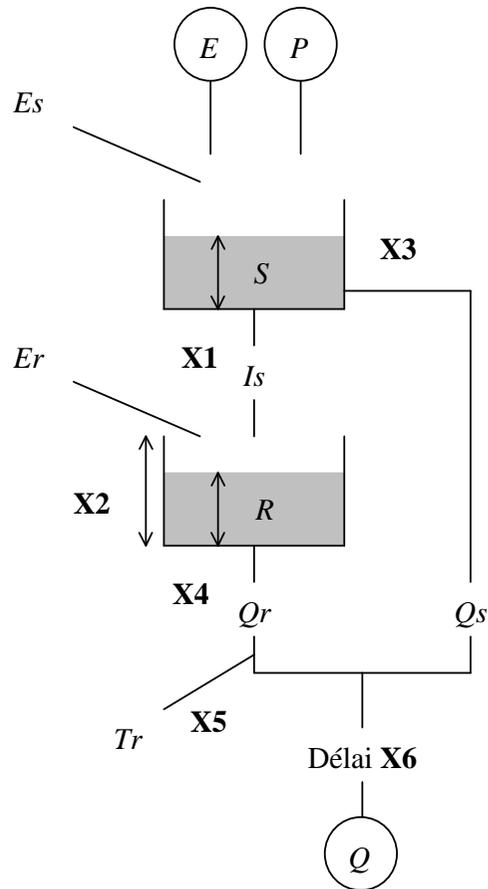
$$Ql = \frac{L}{X4}$$

$$L = L - Ql$$

$$Q = Ql + Qr$$

(avec translation dans le temps de X6 pas de temps)

Structure MODB



$$S = S + P$$

$$E_s = \min(E, S) \quad E' = E - E_s$$

$$Q_s = S / X_3 \quad S = S - Q_s$$

$$I_s = \min(S, X_1) \quad I_s = \min(I_s, X_2 - R) \quad S = S - I_s$$

$$R = R + I_s$$

$$E_r = \min\left(R, E' \cdot \frac{R}{X_2}\right) \quad R = R - E_r$$

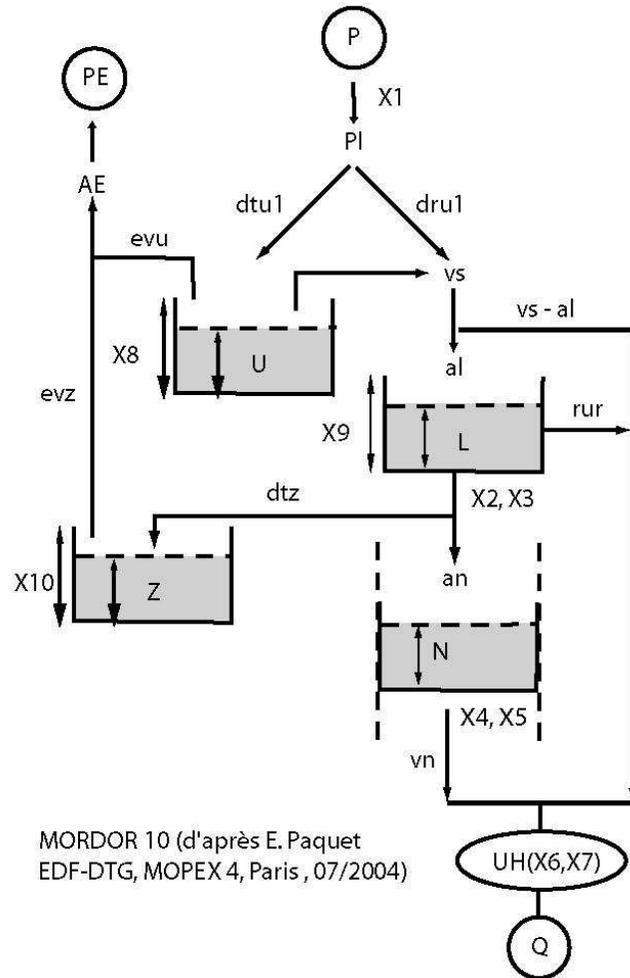
$$Q_r = R / X_4 \quad R = R - Q_r$$

$$T_r = Q_r \cdot X_5 \quad Q_r = Q_r - T_r$$

$$Q = Q_s + Q_r$$

(avec translation dans le temps de X_6 pas de temps)

Structure MORDOR10 (Paquet, 2004)



Correction sur la pluie

$$Pl = P * X1$$

Répartition de la pluie selon U

$$dtr1 = Pl.U / X8$$

$$dtu1 = Pl - dtr1$$

Évolution du réservoir U

$$vs = dtr1 + \max(0, U - X8)$$

$$U = \min(U + dtu1, X8)$$

$$evu = \min(X8, E.U / X8)$$

$$U = U - evu$$

Évolution du réservoir L

$$al = \min(X9 - L, vs.(1 - L / X9))$$

$$L = L + al$$

$$vl1 = L / (X2.X3)$$

$$vl2 = L / X3$$

$$vl = vl1 + vl2$$

Si $vl > L$

$$vl1 = vl1.L / vl$$

$$vl2 = vl2.L / vl$$

$$L = L - vl$$

Évolution du réservoir Z

$$dtz = vl.(1 - Z / X10)$$

$$rur = vl1.Z / X10$$

$$an = vl2.Z / X10$$

$$Z = Z + dtz$$

$$evz = \min(Z, (E - evu).Z / X10)$$

$$Z = \min(X10, Z - evz)$$

Évolution du réservoir N

$$N = N + an$$

$$vn = \min(N, (N / X4)^{X5})$$

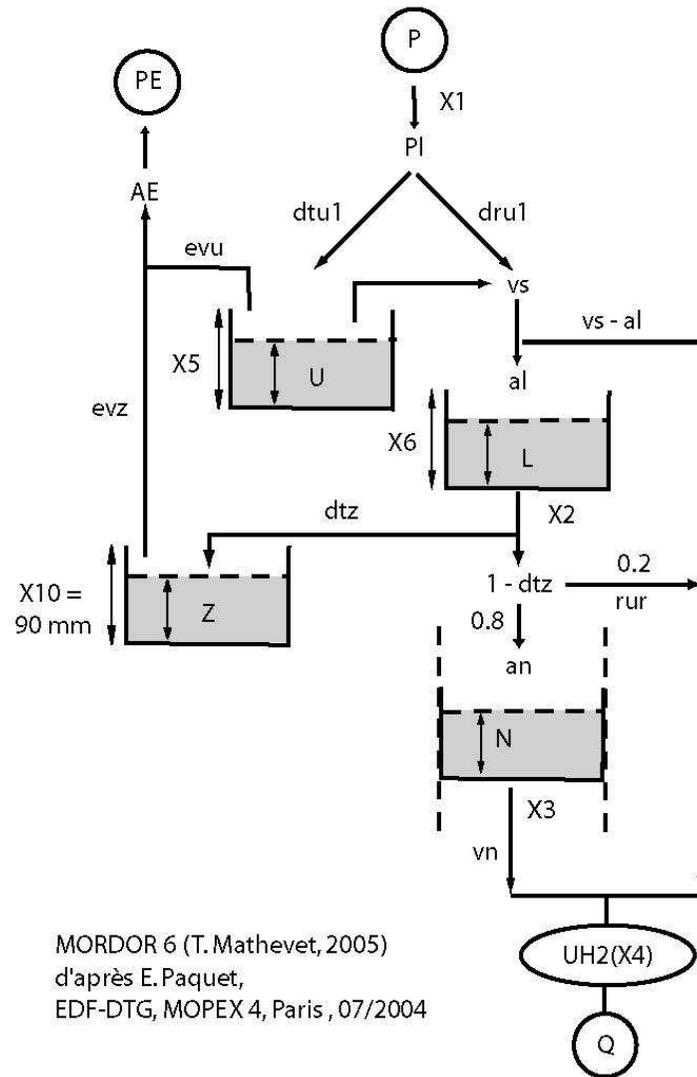
$$N = N - vn$$

Somme et routage des différentes contributions

$$Qt = vs - al + rur + vn$$

$$Q = Weibull(X6, X7, Qt)$$

Structure MORDOR6 (Mathevet, 2005)



Correction sur la pluie

$$Pl = P * X1$$

Répartition de la pluie selon U

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

Évolution du réservoir U

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

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

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

$$U = U - evu$$

Évolution du réservoir L

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

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

Évolution du réservoir Z

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

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

$$Z = Z + dtz$$

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

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

Évolution du réservoir N

$$N = N + an$$

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

$$N = N - vn$$

Somme et routage des différentes contributions

$$Qt = vs - al + rur + vn$$

Hydrogramme unitaire de GR4J (Perrin, 2000)

$$0 \leq j \leq X4,$$

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

$$X4 < j \leq 2.X4,$$

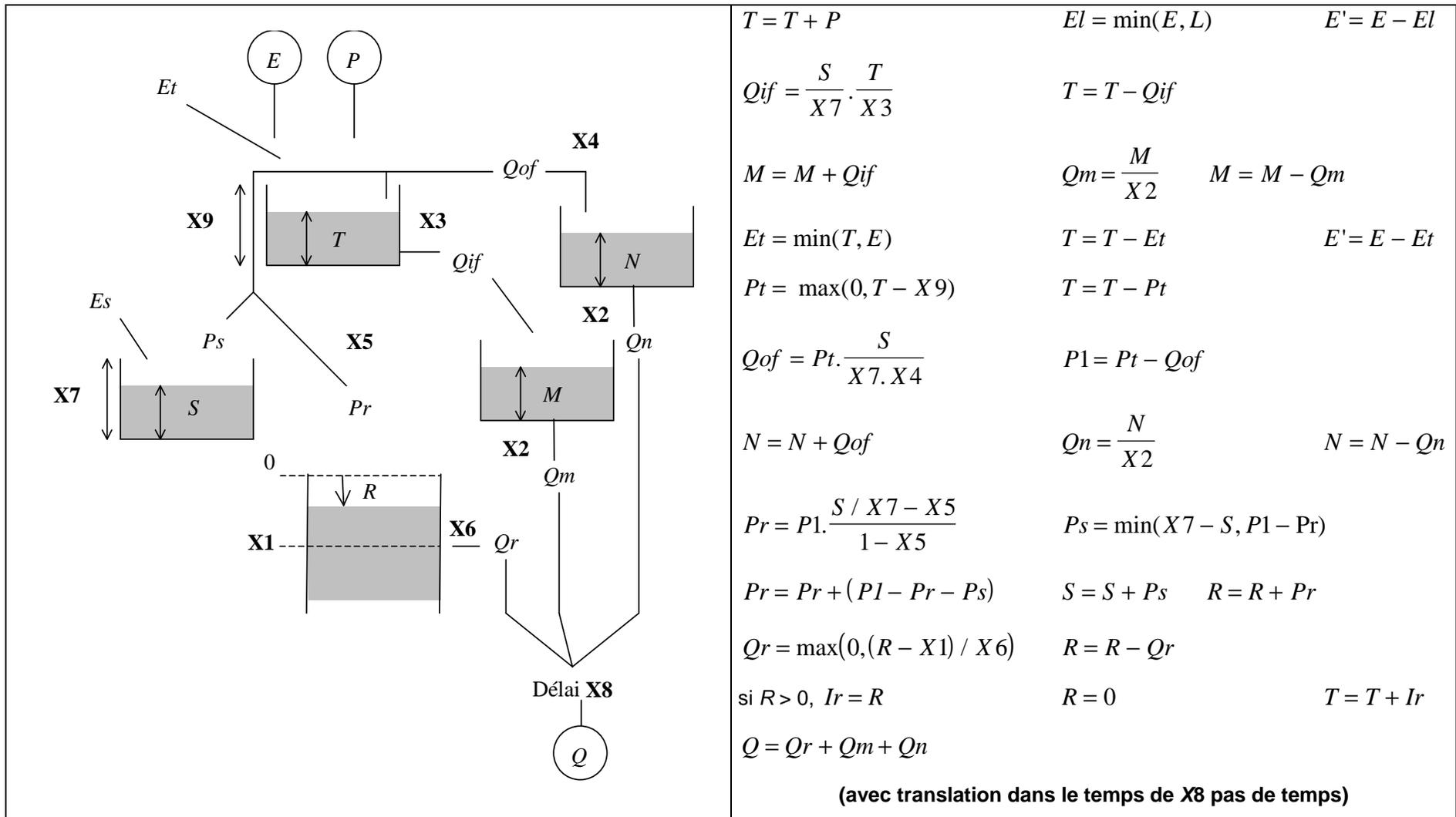
$$SH2(j) = 1 - \frac{1}{2} \left(2 - \frac{j}{X4} \right)^{\frac{5}{4}}$$

$$j > 2.X4,$$

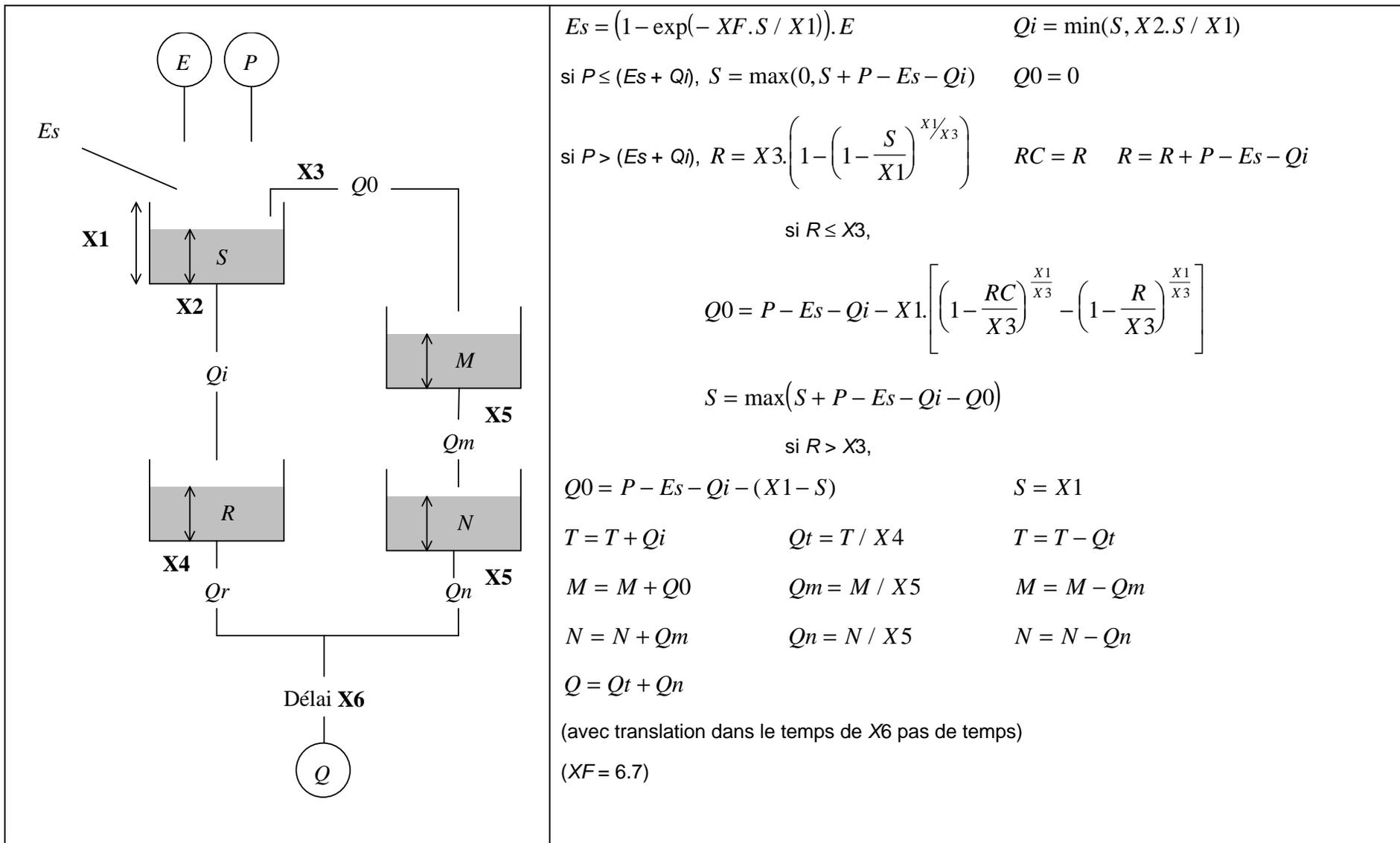
$$SH2(j) = 1$$

$$Q = SH2(X4, Qt)$$

Structure NAMO



Structure PDM



$$Es = (1 - \exp(-XF.S / X1)).E$$

$$Qi = \min(S, X2.S / X1)$$

si $P \leq (Es + Qi)$, $S = \max(0, S + P - Es - Qi)$ $Q0 = 0$

si $P > (Es + Qi)$, $R = X3 \cdot \left(1 - \left(1 - \frac{S}{X1}\right)^{X1/X3}\right)$ $RC = R$ $R = R + P - Es - Qi$

si $R \leq X3$,

$$Q0 = P - Es - Qi - X1 \cdot \left[\left(1 - \frac{RC}{X3}\right)^{\frac{X1}{X3}} - \left(1 - \frac{R}{X3}\right)^{\frac{X1}{X3}} \right]$$

$$S = \max(S + P - Es - Qi - Q0)$$

si $R > X3$,

$$Q0 = P - Es - Qi - (X1 - S)$$

$$S = X1$$

$$T = T + Qi$$

$$Qt = T / X4$$

$$T = T - Qt$$

$$M = M + Q0$$

$$Qm = M / X5$$

$$M = M - Qm$$

$$N = N + Qm$$

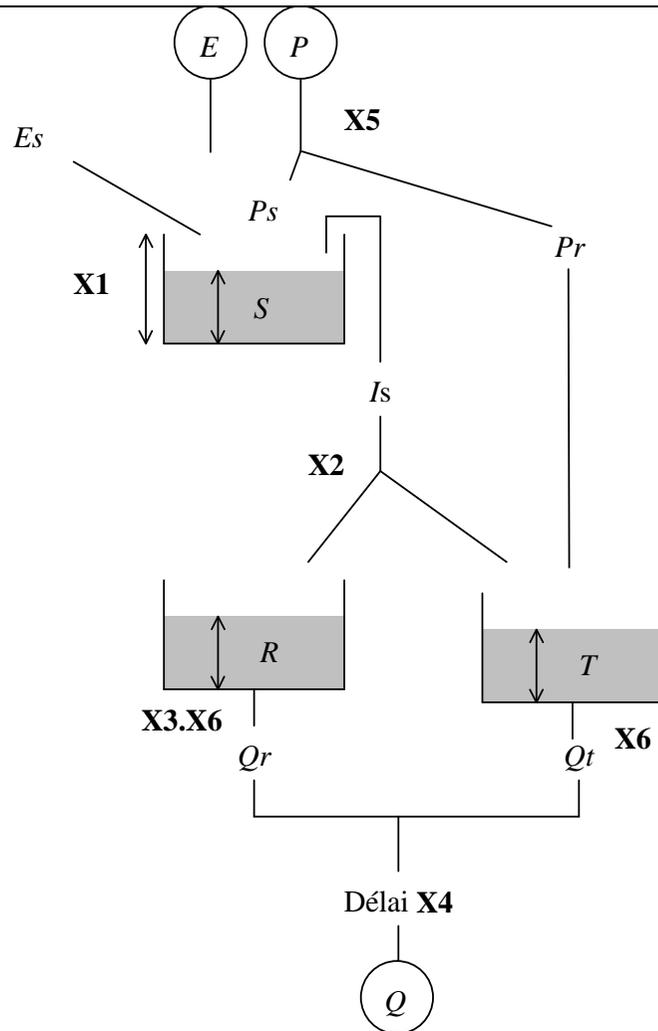
$$Qn = N / X5$$

$$N = N - Qn$$

$$Q = Qt + Qn$$

(avec translation dans le temps de $X6$ pas de temps)
($XF = 6.7$)

Structure BUCK



$$Ps = (1 - X5) \cdot P$$

$$Pr = P - Ps$$

$$\text{si } Ps \geq E, \quad S = S + Ps - E$$

$$Is = \max(0, S - X1)$$

$$S = S - Is$$

$$\text{si } Ps < E, \quad S = S \cdot \exp\left(\frac{Ps - E}{X1}\right)$$

$$Is = 0$$

$$R = R + Is \cdot (1 - X2)$$

$$Qr = R / (X3 \cdot X6)$$

$$R = R - Qr$$

$$T = T + Pr + Is \cdot X2$$

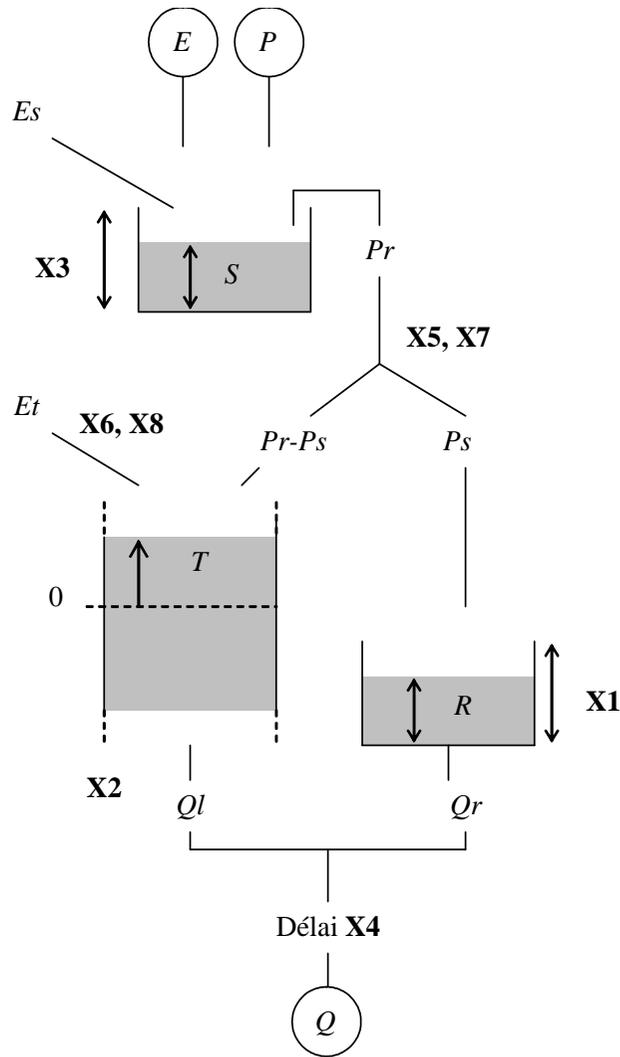
$$Qt = T / X6$$

$$T = T - Qt$$

$$Q = Qt + Qr$$

(avec translation dans le temps de $X4$ pas de temps)

Structure TOPM



$$S = S + P$$

$$Es = \min(S, E)$$

$$S = S - Es$$

$$E' = E - Es$$

$$Pr = \max(0, S - X3)$$

$$S = S - Pr$$

$$Ps = \frac{Pr}{1 + \exp(X7 - T/X5)} \quad T = T + Pr - Ps$$

$$Es = \frac{E'}{1 + \exp(X6 - T/X8)} \quad T = T + Es$$

$$R = R + Ps$$

$$Qr = \frac{R^2}{R + X1}$$

$$R = R - Qr$$

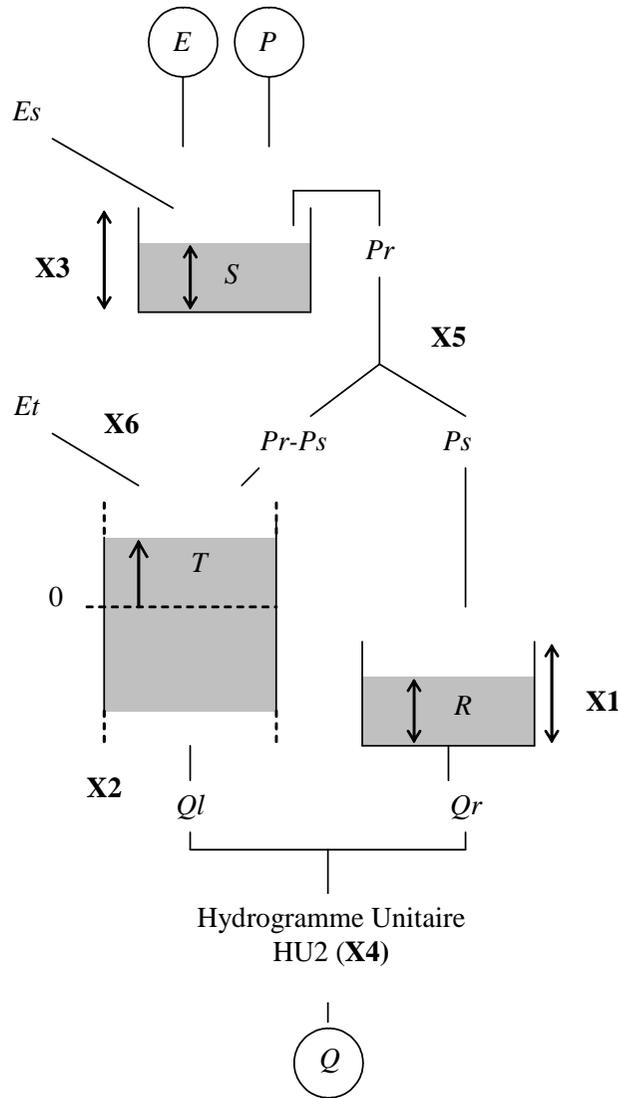
$$Qt = X2 \cdot \exp\left(\frac{T}{X2}\right)$$

$$T = T - Qt$$

$$Q = Qt + Qr$$

(avec translation dans le temps de X4 pas de temps)

Structure TOP6 (Mathevet, 2005)



$$S = S + P$$

$$Es = \min(S, E)$$

$$E' = E - Es$$

$$S = S - Es$$

$$Pr = \max(0, S - X3) \quad S = S - Pr \quad Ps = \frac{Pr}{1 + \exp(165 - T/X5)}$$

$$T = T + Pr - Ps \quad Es = \frac{E'}{1 + \exp(X5 - T/3.6)} \quad T = T + Es$$

$$R = R + Ps \quad Qr = \frac{R^2}{R + X1} \quad R = R - Qr$$

$$Qt = X2 \cdot \exp\left(\frac{T}{X2}\right)$$

$$T = T - Qt$$

$$Q = Qt + Qr$$

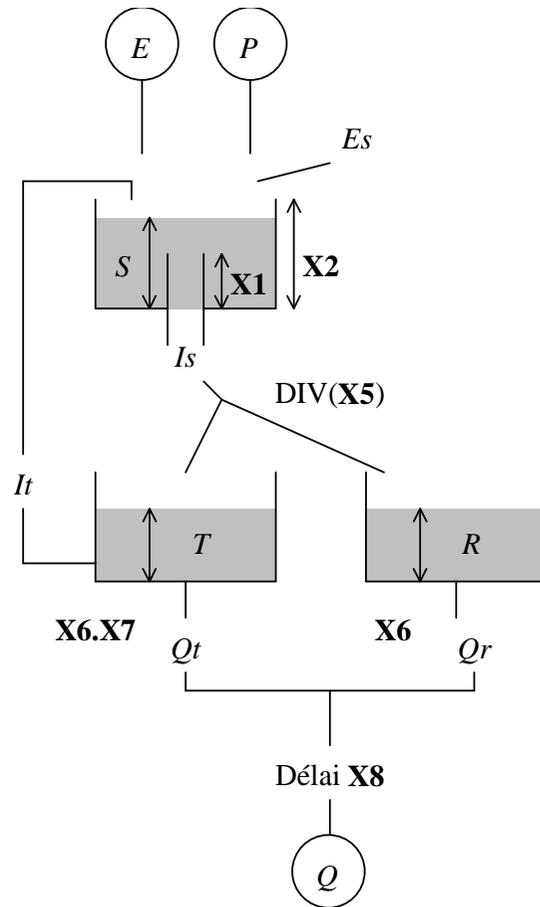
Hydrogramme unitaire de GR4J (Perrin, 2000)

$$0 \leq j \leq X4, SH2(j) = \frac{1}{2} \left(\frac{j}{X4} \right)^{\frac{5}{4}}$$

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

$$j > 2 \cdot X4, SH2(j) = 1 \quad Q = SH2(X4, Qt)$$

Structure WAGE



$$S = S + P$$

$$\text{si } S \geq X1, I_s = \frac{S}{X2} \cdot \frac{S - X1}{X3} \quad I_t = 0$$

$$\text{si } S < X1, I_s = 0 \quad I_t = \frac{T}{X4} \cdot (X1 - S)$$

$$S = S + I_t - I_s$$

$$\text{si } S \geq X1, E_s = E \quad \text{sinon}$$

$$E_s = E \cdot \cos\left(\frac{\pi}{2} \cdot \frac{X1 - S}{X1}\right)$$

$$S = \max(0, S - E_s)$$

$$DIV = \min\left(1, \frac{T}{X5}\right)$$

$$T = T + (1 - DIV) \cdot I_s$$

$$R = R + DIV \cdot I_s$$

$$Q_r = \frac{R}{X6}$$

$$R = R - Q_r$$

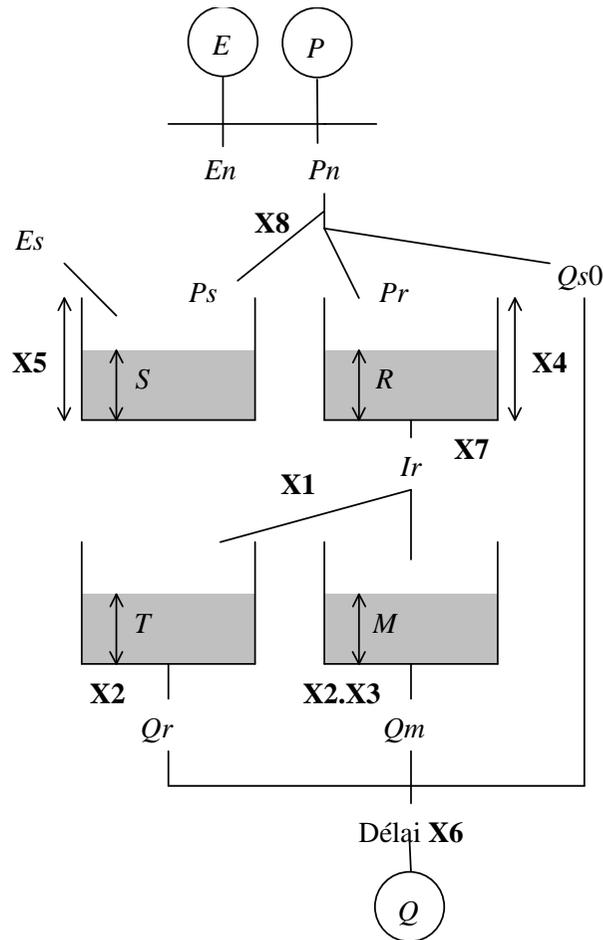
$$Q_t = \frac{T}{X6 \cdot X7}$$

$$T = T - Q_t$$

$$Q = Q_r + Q_t$$

(avec translation dans le temps de X8 pas de temps)

Structure XINJ



$$Pn = \max(0, P - E)$$

$$En = \max(0, E - P)$$

si $En \geq 0$,

$$\text{si } S/X5 \geq 0,9,$$

$$Es = \min(S, En)$$

$$\text{si } 0,09 \leq S/X5 < 0,9,$$

$$Es = \min(S, En \cdot S / (0,9 \cdot X5))$$

$$\text{si } S/X5 < 0,09,$$

$$Es = \min(S, En \cdot 0,1)$$

$$S = S - Es$$

$$Qs0 = 0$$

$$Ir = 0$$

si $Pn > 0$,

$$Fs = [\max(0, 1 - S / X5)]^{\frac{1}{1+XF1}} - \frac{Pn}{(1 + XF1) \cdot X5}$$

$$Fs = [\max(Fs, 0)]^{1+XF1}$$

$$Ps = \max(0, X5 - S - Fs \cdot X5)$$

$$S = \min(X5, S + Ps)$$

$$Pr = \max(0, Pn - Ps)$$

$$Fr = [\max(0, 1 - R / X4)]^{\frac{1}{1+X8}} - \frac{Pr}{(1 + X8) \cdot X4}$$

$$Fr = [\max(Fr, 0)]^{1+X8}$$

$$Pr2 = \max(0, X4 - R - Fr \cdot X4)$$

$$R = \min(X4, R + Pr2) \quad Qs0 = \max(0, Pr - Pr2)$$

$$Ir = R / X7$$

$$R = R - Ir$$

$$T = T + Ir \cdot X1$$

$$Qt = T / X2$$

$$T = T - Qt$$

$$M = M + Ir \cdot (1 - X1)$$

$$Qm = M / X2 \cdot X3$$

$$M = M - Qm$$

$$Q = Qs0 + Qm + Qt$$

(avec translation dans le temps de X6 pas de temps) (avec XF1= 1/4)

Annexe 2 Quelques faiblesses du critère de Nash-Sutcliffe

Un des avantages du critère de NS est que sa formulation est très simple et qu'il est d'un usage très courant en modélisation hydrologique. Néanmoins, sa formulation présente dans certains cas des faiblesses et peut manquer de pertinence :

- *Martinec et Rango* [1989] ont montré que le choix du modèle $Q = \overline{Q_{obs}}$, comme modèle de référence, n'est pas toujours judicieux pour évaluer la qualité d'un modèle. Notamment dans le cas des bassins versants dont les séries comportent une forte variabilité hydrologique, i.e. des années très sèches et des années très humides, dont le débit moyen est très variable. *Perrin* [2000] et *Perrin et al.* [2005] ont également illustré ce problème et montré que la valeur du critère de NS obtenue pour certains bassins versants sur certaines périodes dépend clairement de la valeur du débit moyen choisi : débit moyen de la totalité de la chronique hydrologique ou débit moyen de la période testée. Entre ces deux modèles de références, les résultats sont globalement concordants, mais pour certains cas les critères de NS peuvent être très différents et sont généralement meilleurs lorsqu'on utilise le débit moyen à long terme [*Perrin*, 2000; *Perrin et al.*, 2005]. Ce problème est également illustré par la Figure 9-17.

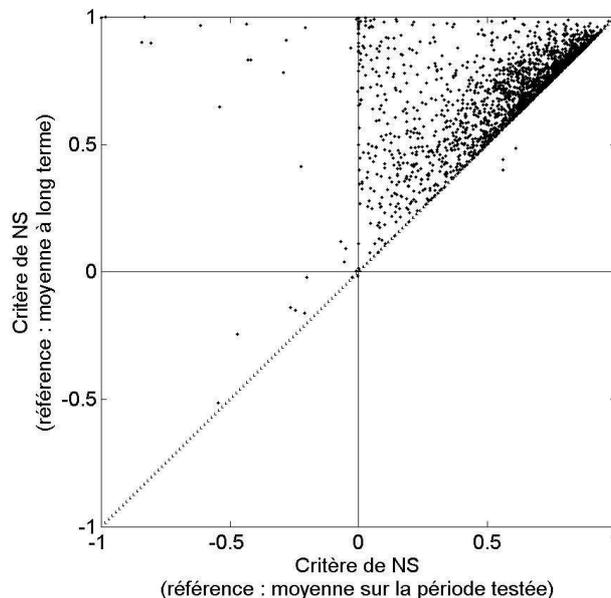


Figure 9-17 : Comparaison du critère de NS obtenu sur l'échantillon de 313 bassins versants, en utilisant deux modèles de référence différents : le débit moyen de la période testée (axe x) et le débit moyen de la série chronologique (axe y)

- D'après les travaux de quelques auteurs [*Chiew et al.*, 1993; *Gan et Biftu*, 1996; *Gan et al.*, 1997; *Lidén et Harlin*, 2000], il semble que les bassins versants secs (à faible coefficient d'écoulement) soient plus difficiles à modéliser. Par ailleurs, il semble également que les performances des modèles hydrologiques soient plus faibles sur ce type de bassins. Sur un large échantillon de bassins versants *Perrin* [2000] et *Perrin et al.* [2005] ont montré que les bassins à faible coefficient d'écoulement ont généralement des valeurs de critère beaucoup plus faibles que les bassins à fort coefficient d'écoulement. Néanmoins, on peut se demander si la relation pluie-débit est réellement plus difficile à modéliser pour ces bassins versants ou bien si la différence de performances ne vient pas de la formulation du critère qui n'est pas également exigeante pour tous les types de climats ? D'après ces auteurs, le modèle de référence est responsable de la forte dépendance des valeurs du critère de NS au climat du bassin versant et des faibles valeurs obtenues sur les bassins à faible coefficient d'écoulement. En effet, sur les bassins à faible coefficient d'écoulement, la variance des débits est très faible et le modèle de référence est déjà un très bon modèle pour ce type de bassins versants. Il est donc plus difficile de faire mieux que ce modèle de référence.
- D'autre part, *Garrick et al.* [1978] et *Martinec et Rango* [1989] ont mis en évidence que le critère de NS était capable de produire des valeurs assez fortes, même pour de mauvais modèles. Ce problème est principalement dû à la simplicité du modèle de référence ($Q = \overline{Q_{obs}}$). Ainsi, il est généralement très facile pour un modèle de dépasser les performances de ce modèle de référence ($SE \ll V$). Ces auteurs recommandent d'utiliser un autre modèle alternatif, presque aussi simple, le modèle 'saisonnier' ou 'paysan'. Ce modèle alternatif fait référence au bon sens paysan et correspond au débit moyen de chaque jour de l'année, calculé sur l'ensemble de la chronique. Ce modèle de référence permet de tester un modèle par rapport au fonctionnement moyen d'un bassin versant. D'autres auteurs *Seibert* [2001] et *Perrin et al.* [2005] ont également proposé d'utiliser d'autres modèles de référence pour évaluer les performances d'un modèle. En effet, bien souvent, on peut vouloir tester l'intérêt d'un nouveau modèle par rapport à un autre, plus simple ou déjà en service. Ainsi, le critère de NS peut être reformulé de la manière suivante, en utilisant un autre modèle de référence :

$$NS_{REF} = 1 - \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - Q_{réf,i})^2} \quad [8.6]$$

où, $Q_{\text{obs},i}$, $Q_{\text{sim},i}$ et $Q_{\text{référence},i}$ sont respectivement le débit observé, le débit simulé par le modèle testé et le débit simulé par le modèle de référence.

Annexe 3 Comparaison des performances de deux modèles en utilisant une procédure de sous-échantillonnage de l'échantillon total

Au début de ce chapitre, nous avons fait l'hypothèse que l'échantillon de 313 bassins versant nous permet d'approcher les performances 'vraies' des modèles pluie-débit. Même si cet échantillon regroupe un grand nombre de bassins versants, avec une forte diversité de situations hydro-climatiques, on peut tout de même se demander de quelle manière les résultats dépendront de cet échantillon. Il pourrait être intéressant de traduire la différence de performances entre deux modèles en densité de probabilité, plutôt que de manière absolue. Pour cela, on peut générer un grand nombre de sous-échantillons de différentes tailles et regarder ensuite la densité et la distribution des performances comparatives de deux modèles.

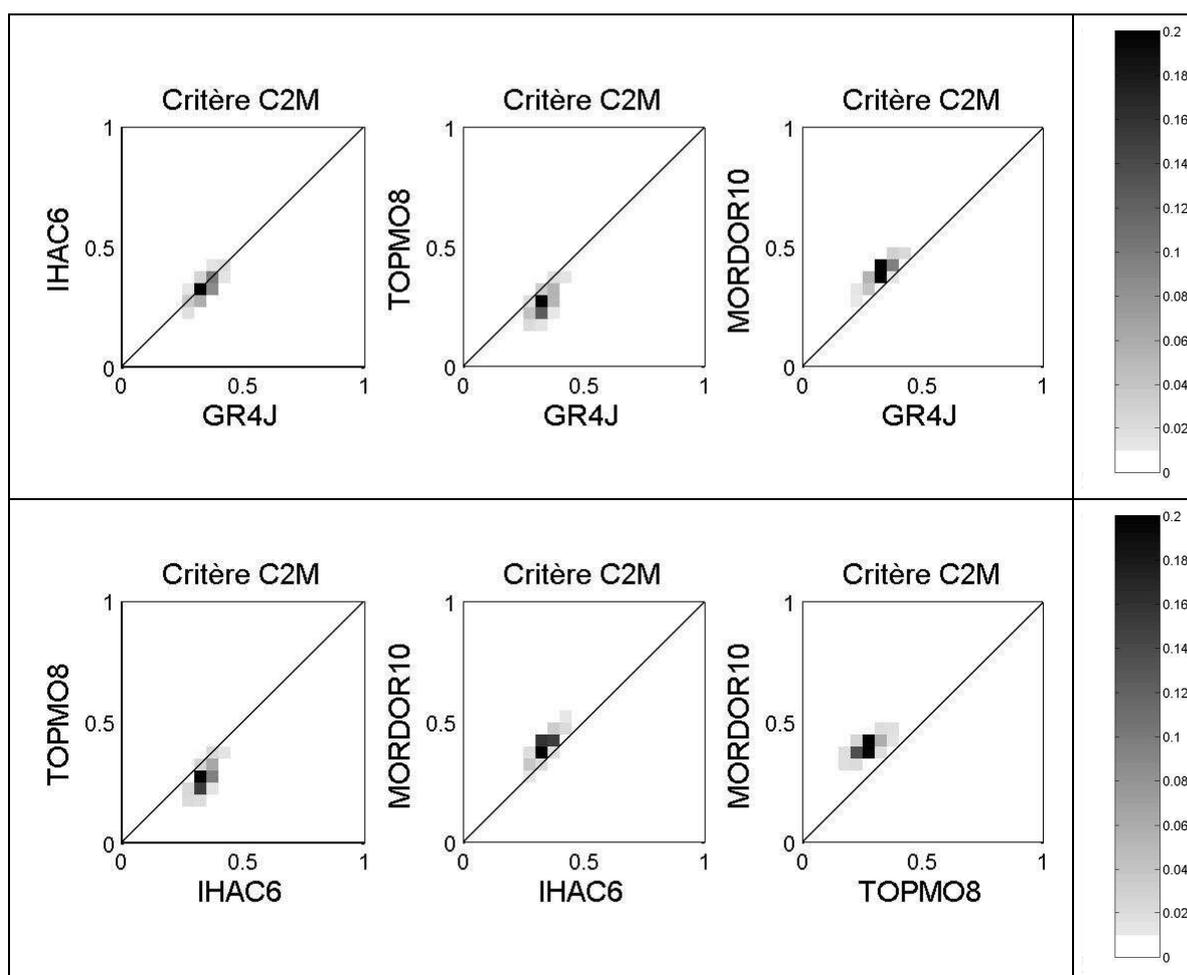


Figure 9-18 : Comparaison des performances des modèles GR4J, IHAC6, TOPMO8 et MORDOR10 pour 10000 sous-échantillons aléatoires différents de 10, 30, 100, 200 et 250

bassins versants, en utilisant le critère C_{2M} . Les performances des 50000 sous-échantillons ont été assemblées pour évaluer les densités de performances.

Par exemple, la Figure 9-18 représente la densité des performances comparatives de deux modèles testés sur 10000 sous-échantillons aléatoires différents de 10, 30, 100, 200 et 250 bassins versants. Cette représentation permet de donner une image des différences de performances entre deux modèles, indépendamment de la taille du sous-échantillon et de celle de l'échantillon total. Enfin, la Figure 9-19 nous permet de revenir à une seule dimension et de nous intéresser seulement aux différences de performances, quelle que soit la performance moyenne de chaque modèle. Cette figure représente la distribution des différences de performances entre deux modèles, calculées à partir des 50000 échantillons tirés précédemment.

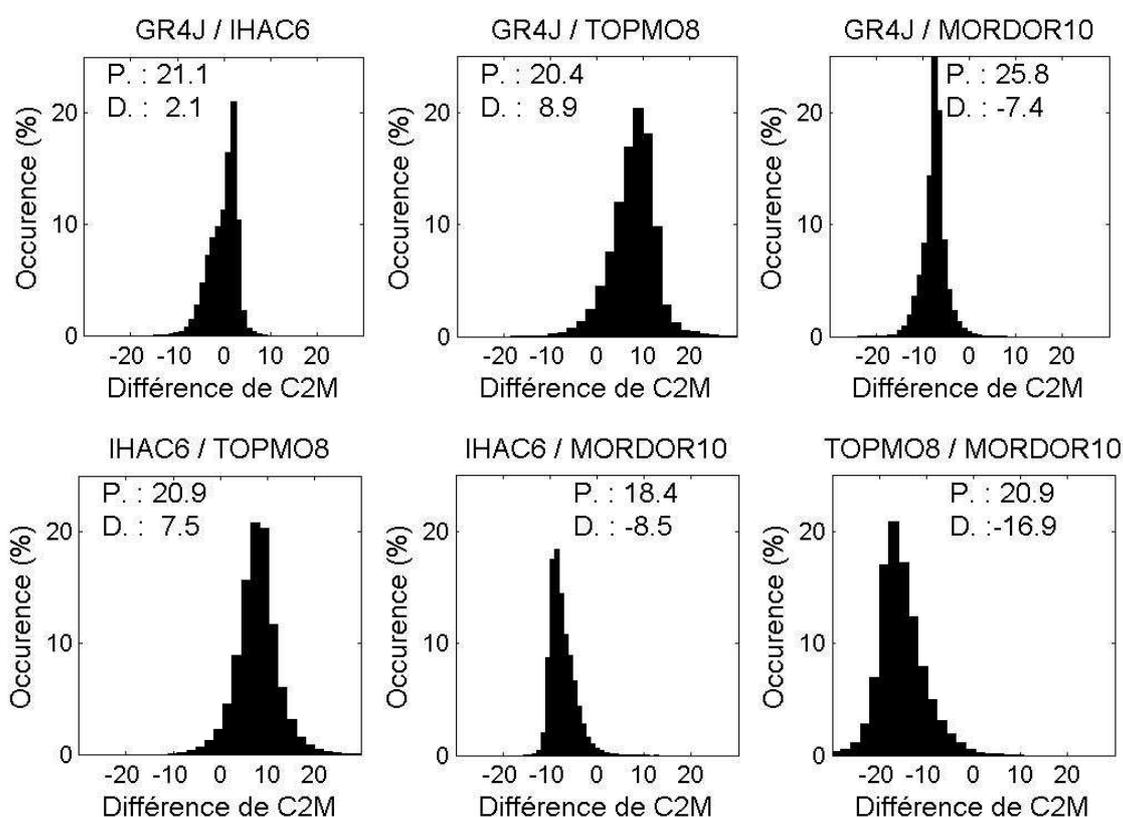


Figure 9-19 : Distribution des différences de performances entre deux modèles (A/B). Tirages aléatoires de 10000 sous-échantillons différents de 10, 30, 100, 200 et 250 bassins versants. (P : pourcentage maximal au pic ; D : différence de performance moyenne. Si $D < 0$, $B > A$. Si $D > 0$, $A > B$).

Ces représentations des performances comparatives de deux modèles donnent une vision un peu plus juste de leurs niveaux de performances. Elles permettent notamment d'exprimer ces différences de performances entre deux modèles non pas d'une manière absolue, mais d'une manière probabiliste, ce qui permet de réduire les dépendances des conclusions vis-à-vis de l'échantillon total des 313 bassins versants. En effet, même si l'échantillon complet n'est pas 'représentatif' de

l'ensemble des situations possibles, on peut penser qu'une partie des sous-échantillons est assez représentative de la variété des situations hydro-climatiques existantes.

Annexe 4 Algorithmes des méthodes d'optimisation

9.1 PAP-GR

Supposons que l'on ait à caler un vecteur x de paramètres ayant n composantes (x_1, x_2, \dots, x_n) . On désire trouver le vecteur x^* qui maximise le critère R^2 dont la valeur dépend, par l'intermédiaire des débits calculés, des valeurs des paramètres (x_1, x_2, \dots, x_n) .

Le processus d'optimisation est itératif. La méthode adopte une stratégie de déplacement, la plupart du temps, le long des axes de l'espace des paramètres, avec un pas de recherche pouvant varier d'une itération à l'autre. L'amplitude du pas de recherche étant ici la même pour tous les paramètres, des transformations mathématiques préalables (transformations logarithmiques ou puissance par exemple) peuvent être appliquées pour garantir des sensibilités grossièrement équivalentes à ce pas de recherche pour tous les paramètres. Ces transformations sur les paramètres sont choisies en tenant compte de la façon dont les paramètres interviennent dans le modèle, et sont propres à celui-ci.

La recherche démarre à partir d'un vecteur initial des paramètres x^0 , de composantes $(x_1^0, x_2^0, \dots, x_n^0)$. On calcule alors la valeur correspondante de la fonction objectif. On fait ensuite varier successivement chacune des valeurs des paramètres d'une déviation initiale $\pm \Delta x_{ini}$ (ici on adopte $\Delta x_{ini} = 0,64$, ce qui revient à multiplier ou diviser par deux la valeur du paramètre lorsque la transformation est logarithmique). Lorsque la valeur du $i^{\text{ème}}$ paramètre est modifiée, les deux vecteurs de paramètres testés sont donc $(x_1^0, x_2^0, \dots, x_i^0 + \Delta x_{ini}, \dots, x_n^0)$ et $(x_1^0, x_2^0, \dots, x_i^0 - \Delta x_{ini}, \dots, x_n^0)$. A chaque fois la valeur de la fonction objectif correspondante est calculée.

Lorsque tous les paramètres ont été modifiés un à un, deux cas de figure peuvent se présenter:

1. il y a une amélioration de la valeur de la fonction objectif pour un ou plusieurs des nouveaux vecteurs des paramètres. On retient dans ce cas le vecteur x^1 qui correspond à la meilleure amélioration de la fonction. Ce vecteur, dont les composantes correspondent à celles du vecteur x^0 sauf la composante i qui correspond à $x_i^0 + \Delta x$ ou $x_i^0 - \Delta x$, devient le nouveau vecteur 'initial' de recherche. S'il y a $2n$ améliorations successives de la fonction dans une même direction, le pas de recherche est alors multiplié par deux pour accélérer la recherche (dans la limite de Δx_{max} , que l'on prendra ici égal à 2). La procédure peut également être affinée et/ou accélérée en utilisant, au delà de $4n$ itérations de calcul, un pas de recherche correspondant à un lissage exponentiel sur les déplacements effectués

aux itérations précédentes. Ceci a pour but d'accélérer et d'améliorer la recherche, en particulier dans le cas où il existe sur la surface de réponse une vallée qui ne soit pas dans la direction de l'axe de variation d'un des paramètres. Ce sont les seuls cas où la méthode adopte un déplacement non parallèle à l'un des axes de l'espace des paramètres.

2. il n'y a amélioration de la valeur de la fonction critère pour aucune des modifications des paramètres. On affine alors la recherche en divisant le pas de recherche par deux, et on réitère les modifications des paramètres sur le même vecteur initial x^0 avec ce nouveau pas de recherche.

A chaque itération, on vérifie que les valeurs des paramètres appartiennent au domaine des paramètres transformés possibles préalablement spécifié, ce domaine correspondant à une zone de sensibilité algébrique des paramètres définie a priori.

La procédure s'arrête lorsque la déviation minimale Δx_{\min} spécifiée sur les paramètres (ici, on adopte $\Delta x_{\min} = 0,01$) n'apporte plus d'amélioration de la fonction critère. On a alors localisé l'optimum x^* (ou un optimum) avec une précision relative à la déviation minimale. Par sécurité, pour ne pas avoir des temps de calcul trop longs, la procédure se termine si le nombre d'itérations est supérieur à $100n$ (cas rare que nous n'avons jamais rencontré). La méthode n'a alors pu localiser un optimum sur la surface de réponse.

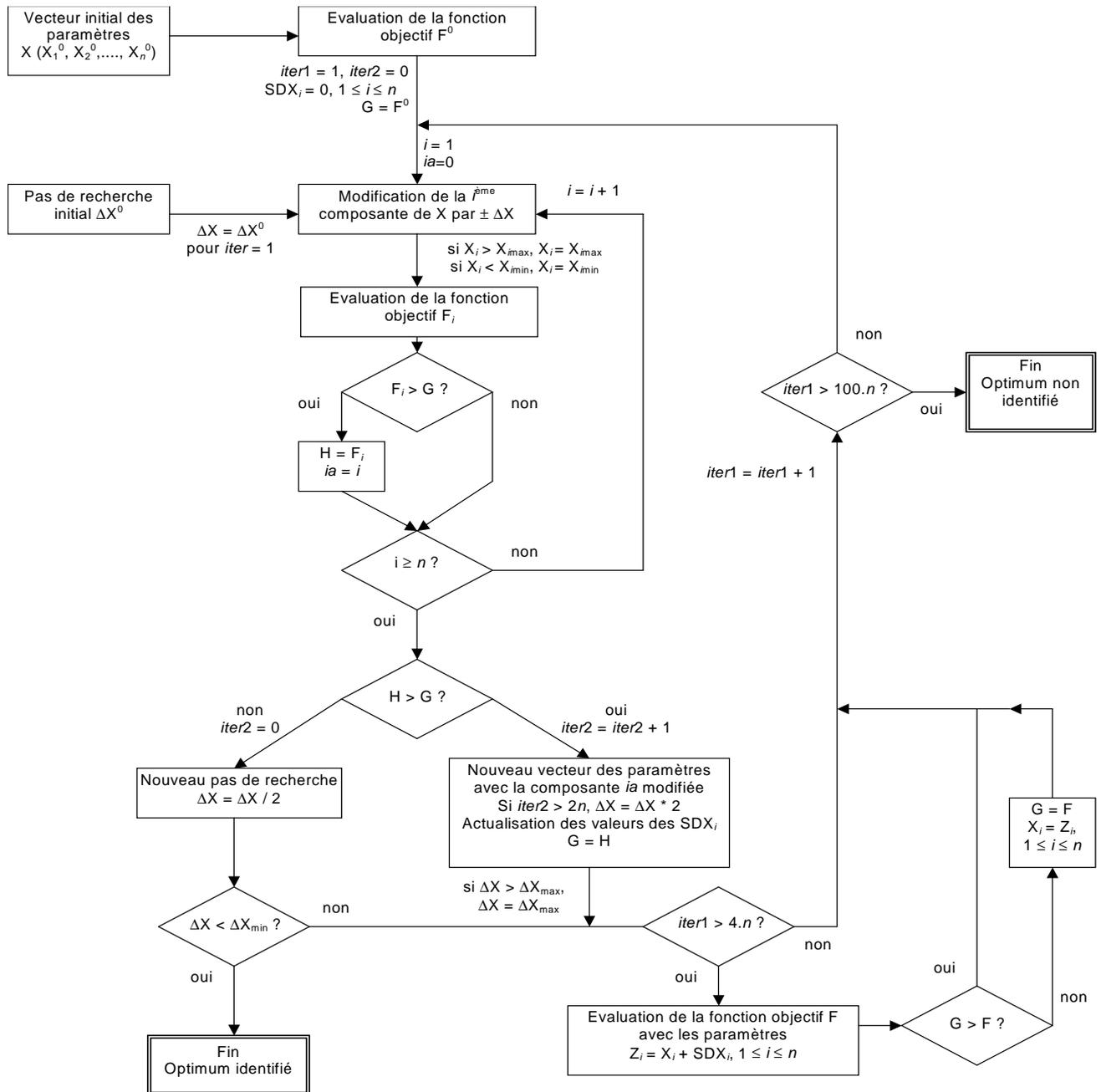


Figure 9-20 : Diagramme de la méthode 'Pas à pas' (PAP-GR)

9.2 SCE-UA

La méthode d'optimisation du SCE-UA suit l'algorithme suivant :

1. Choix des paramètres de l'algorithme :
 - $p \geq 1$, p est le nombre de complexes,
 - $m \geq n+1$, m le nombre de points dans chaque complexe,
 - n , le nombre de paramètres à être optimisé,
 - $s = p \times m$, la taille de l'échantillon de points
2. Génération d'un échantillon de s points, $x_1, \dots, x_i, \dots, x_s$ dans l'espace des paramètres $\Omega \subset \mathbb{R}^n$. Calculer la fonction objectif f_i à chaque point x_i . En absence d'information a priori, générer l'échantillon de points avec une distribution uniforme.
3. Classement des points : classer les s points dans l'ordre croissant de leur fonction objectif. Classer les s points dans un tableau $D = \{x_i, f_i, i = 1, \dots, s\}$, de telle manière que $i = 1$ représente le point avec la fonction objectif la plus petite,
4. Préparation des complexes : diviser l'échantillon de s points en p complexes A^1, \dots, A^p , chacun contenant m points de telle manière que $A^k = \{x_j^k, f_j^k \mid x_j^k = x_{k+p(j-1)}, f_j^k = f_{k+p(j-1)}, j = 1, \dots, m\}$
5. Evolution des complexes : faire évoluer chaque complexe A^1, \dots, A^p en utilisant l'algorithme CCE (Competitive Complex Evolution).
6. Mélanger les complexes : remettre les complexes A^1, \dots, A^p dans D , tel que $D = \{A^k, k = 1, \dots, p\}$. Classer D dans l'ordre croissant de leur fonction objectif.
7. Vérifier la convergence : si le critère de convergence est satisfait, arrêter l'algorithme, sinon retourner à l'étape (4).

Le CCE suit l'algorithme suivant :

1. Choix des paramètres de l'algorithme :
 - i. $2 \leq q \leq m$,
 - ii. $\alpha \geq 1$,
 - iii. $\beta \geq 1$,
2. Assigner une probabilité de distribution trapézoïdale à A^k :
 - i. $\rho_i = \frac{2(m+1-i)}{m(m+1)}, i = 1, \dots, m$

Le point x_1^k a la plus grande probabilité $\rho_1 = \frac{2}{m} + 1$. Le point x_m^k a la plus faible

$$\text{probabilité } \rho_m = \frac{2}{m(m+1)}$$

3. Choix des parents : choisir aléatoirement q points distincts u_1, \dots, u_q dans A^k , en suivant la probabilité de distribution trapézoïdale (les q points définissent un sous-complexe). Classer le q points dans $B = \{u_i, v_i, i = 1, \dots, q\}$, où v_i est la valeur de la fonction objectif du point u_i . Stocker dans L les coordonnées de A^k qui sont utilisées pour construire B .
4. Génération des enfants :
 - i. Classer B et L dans l'ordre croissant de leur fonction objectif. Calculer le centroïde g , en utilisant l'expression suivante :

$$g = \frac{1}{q-1} \sum_{j=1}^{q-1} u_j$$
 - ii. Calculer le nouveau point $r = 2g - u_q$ (phase de réflexion),
 - iii. Si $r \in \Omega \subset \mathbb{R}^n$, calculer la fonction objectif f_r et aller à l'étape (iv). Sinon, calculer le plus petit hypercube $H \subset \mathbb{R}^n$, qui contienne A^k , générer aléatoirement un point z dans H et calculer la fonction objectif f_z et fixer $f_r = f_z$ (Phase de mutation).
 - iv. Si $f_r < f_q$, remplacer u_q par r et aller à l'étape (vi). Sinon, calculer $c = (g + u_q)/2$ et la valeur de la fonction objectif f_c (Phase de contraction).
 - v. Si $f_c < f_q$, remplacer u_q par c et aller à l'étape (vi). Sinon, générer aléatoirement un point z dans H et calculer la fonction objectif f_z et fixer $f_r = f_z$ (Phase de mutation). Remplacer u_q par z .
 - vi. Répéter les étapes (i) à (v) α fois, où α est choisi par l'utilisateur.
5. Remplacer les parents par les enfants : remettre B dans A^k en utilisant les coordonnées originelles stockées dans L . Classer A^k dans l'ordre croissant de leur fonction objectif.
6. Répéter les phases (1) à (5) β fois, où β est choisi par l'utilisateur. β représente le nombre de générations enfants qui doivent être générées (le degré d'évolution de chaque complexes).

Dans la version du SCE-UA utilisées, $m = 2n + 1$, $q = n + 1$, $\alpha = 1$, $\beta = 2n + 1$.

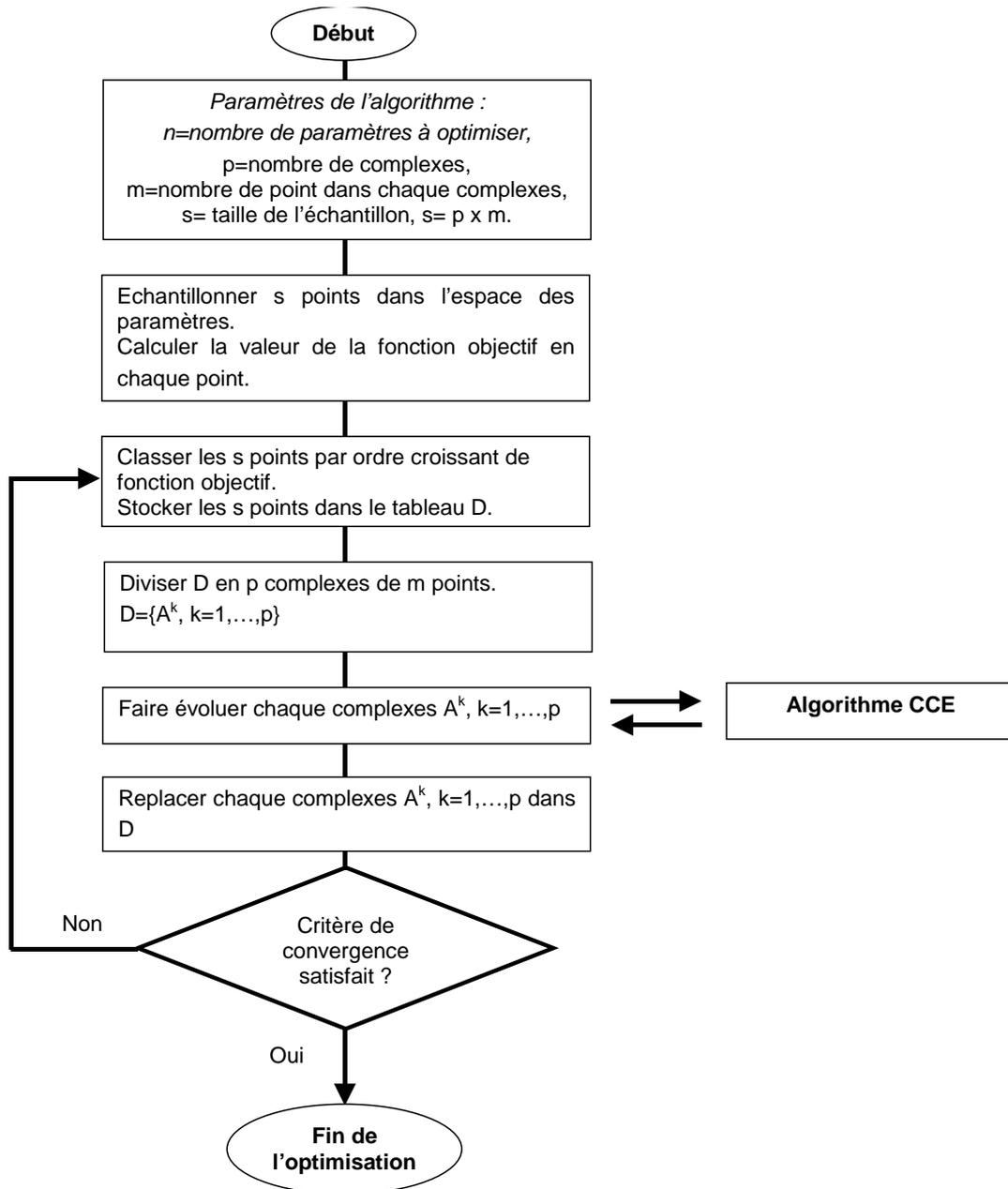


Figure 9-21 : Diagramme de la méthode du SCE-UA [Duan et al., 1992]

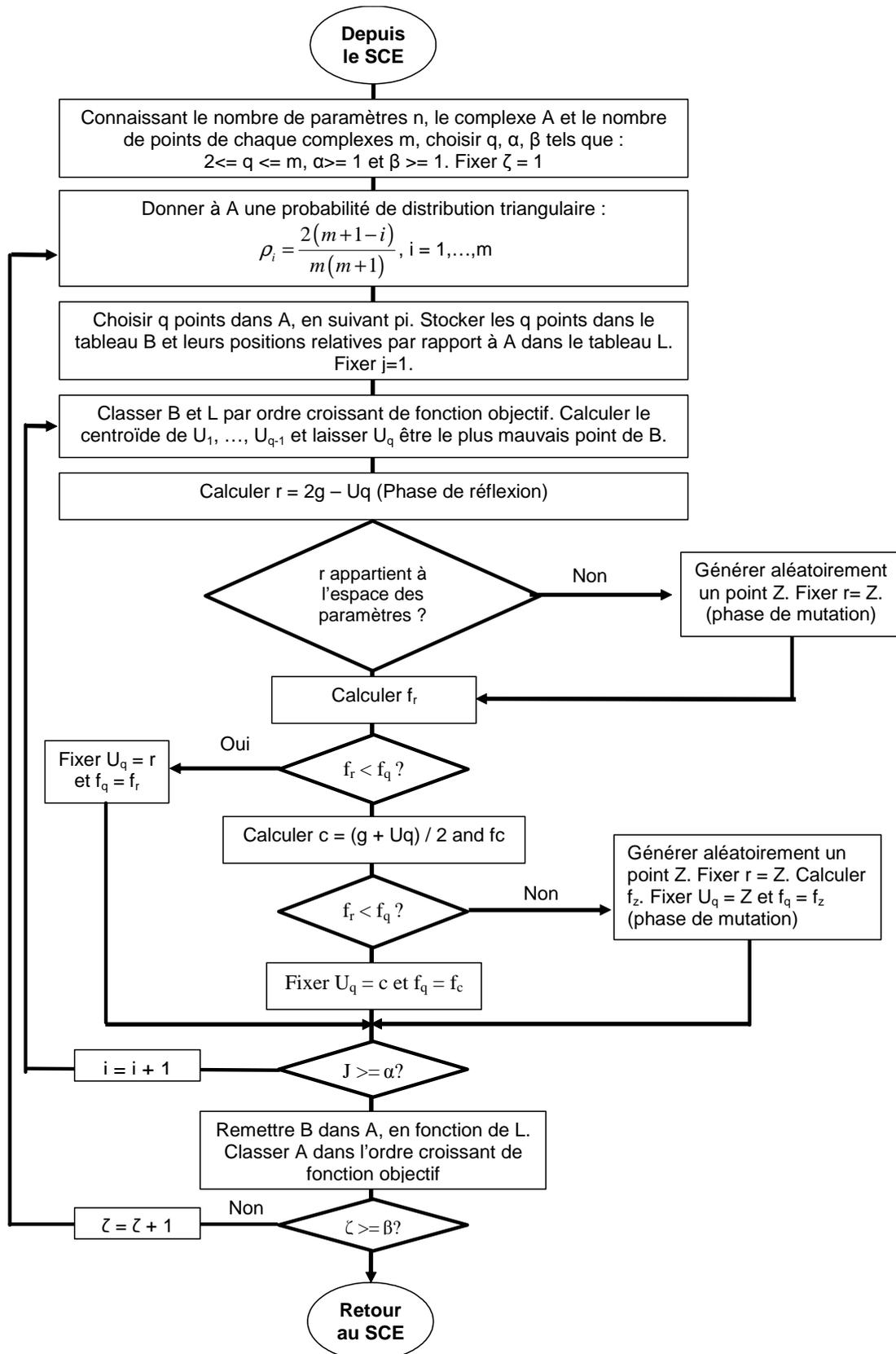


Figure 9-22 Diagramme de l'algorithme CCE, appartenant à la méthode du SCE-UA [Duan et al., 1992]

9.3 GA-EDF

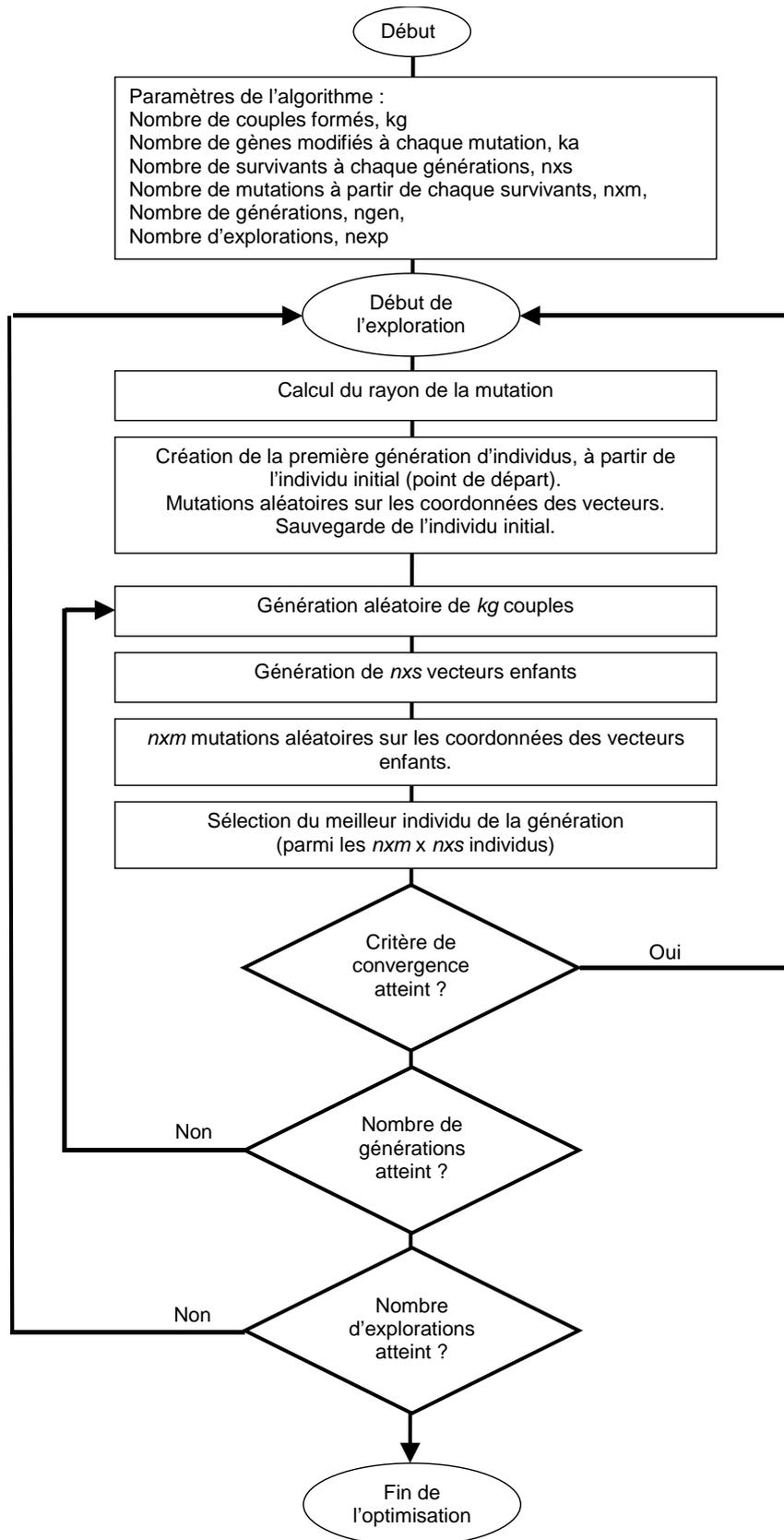


Figure 9-23 : Diagramme de l'algorithme GA-EDF.

Annexe 5 Comparaison de méthodes d'optimisation : résultats complémentaires

9.4 Performances des modèles

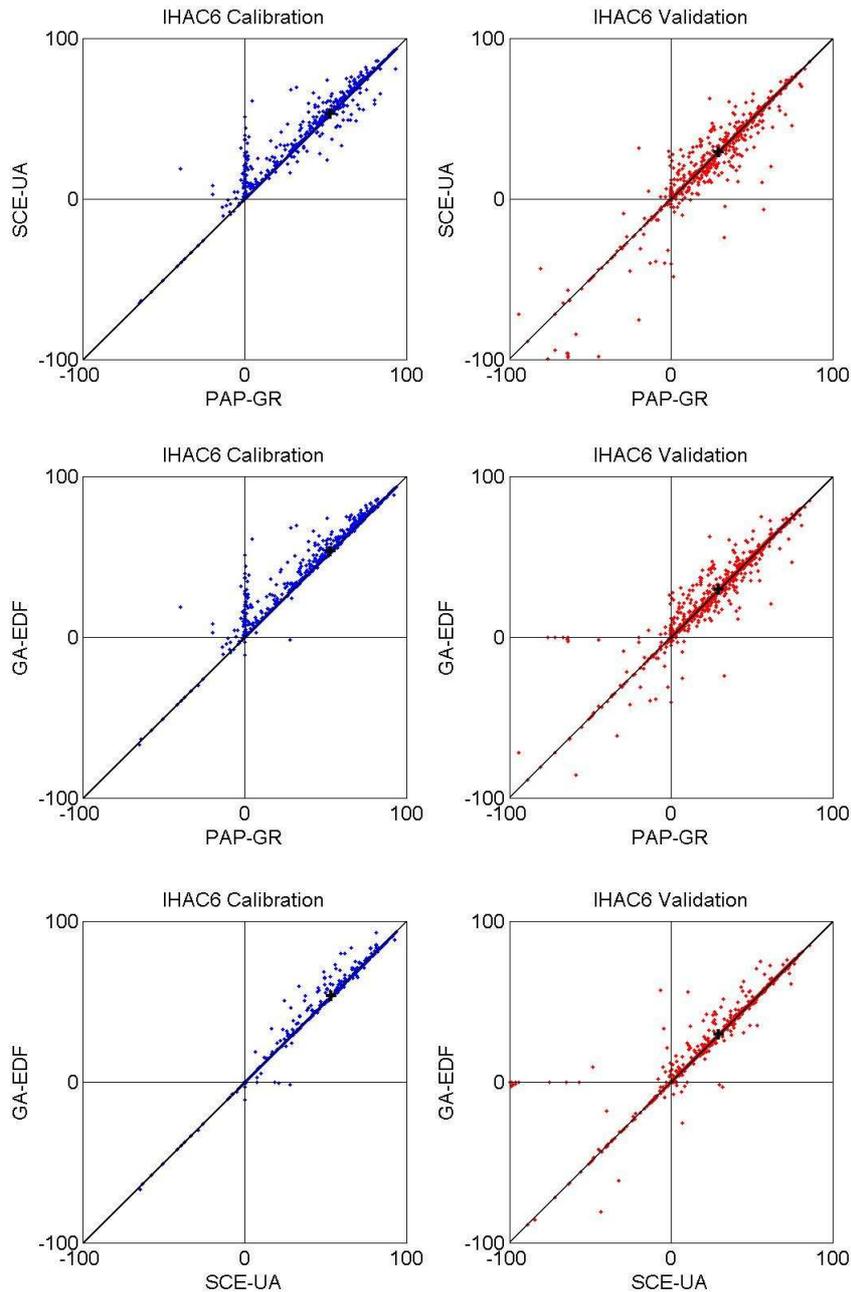


Figure 9-24 : Comparaison des performances en calibration et validation des vecteurs trouvés par les méthodes PAP-GR, SCE-UA et GA-EDF, pour le modèle IHAC6. a) comparaison des méthodes PAP-GR et SCE-UA, b) comparaison des méthodes PAP-GR et GA-EDF, c) comparaison des méthodes SCE-UA et GA-EDF. La croix représente la moyenne.

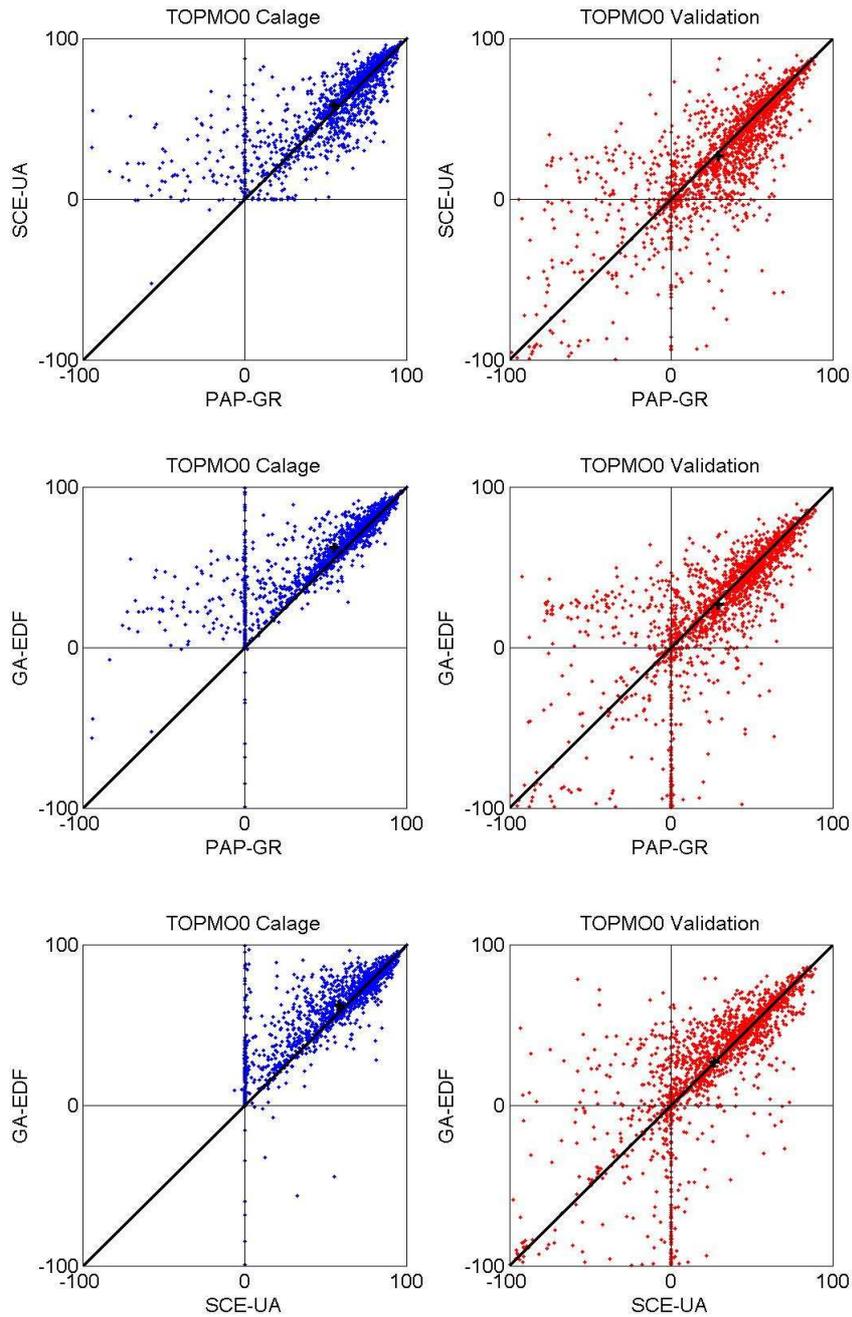


Figure 9-25 : Comparaison des performances en calibration et validation des vecteurs trouvés par les méthodes PAP-GR, SCE-UA et GA-EDF, pour le modèle TOPMO8. a) comparaison des méthodes PAP-GR et SCE-UA, b) comparaison des méthodes PAP-GR et GA-EDF, c) comparaison des méthodes SCE-UA et GA-EDF. La croix représente la moyenne.

9.5 Résidus des performances des méthodes d'optimisation

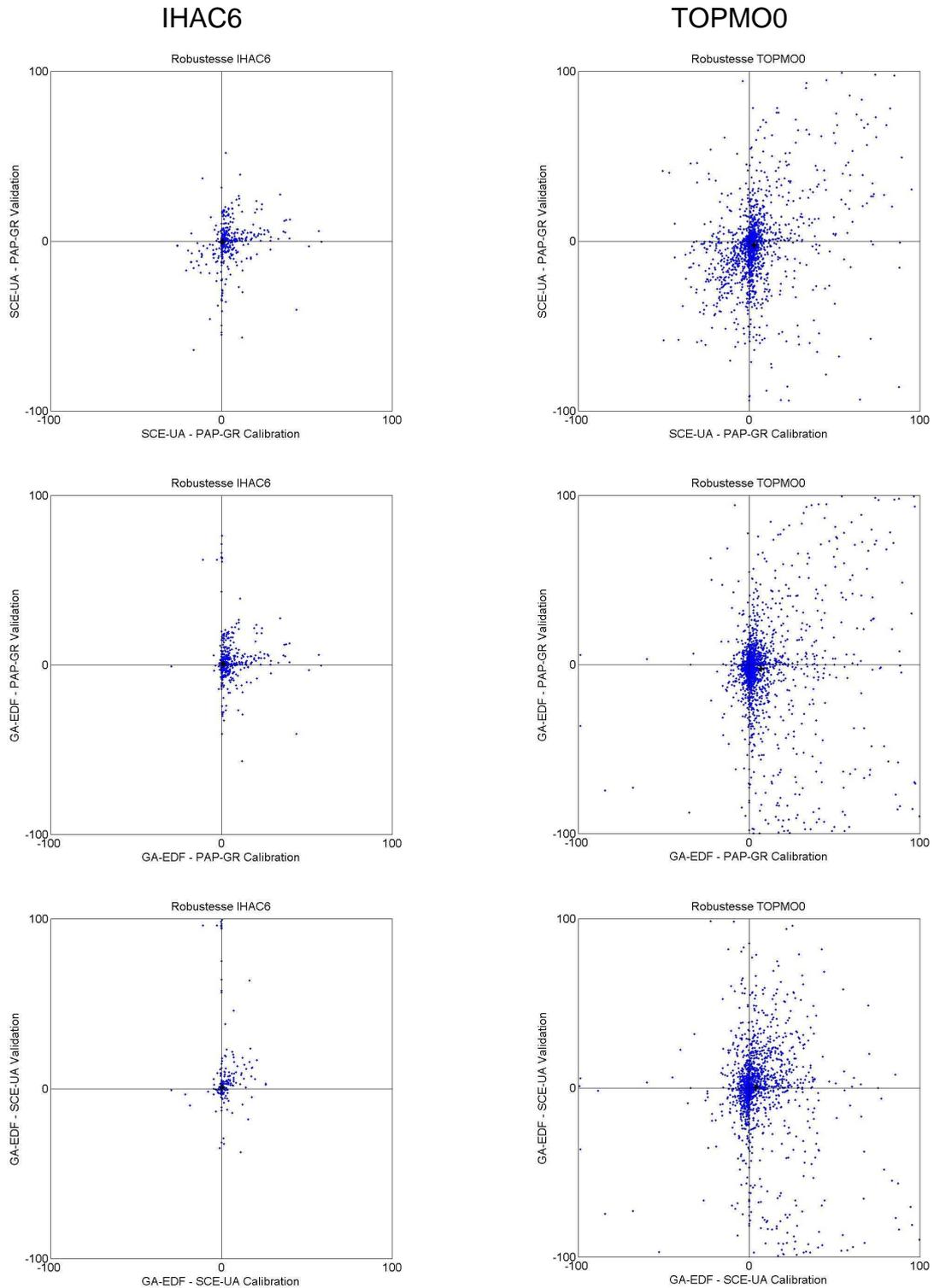


Figure 9-26 : Comparaison de la différence de performances obtenue en calibration et en validation, entre deux méthodes d'optimisation. a) PAP-GR et SCE-UA, b) PAP-GR et GA-EDF, c) SCE-UA et GA-EDF.

9.6 Vecteurs de paramètres

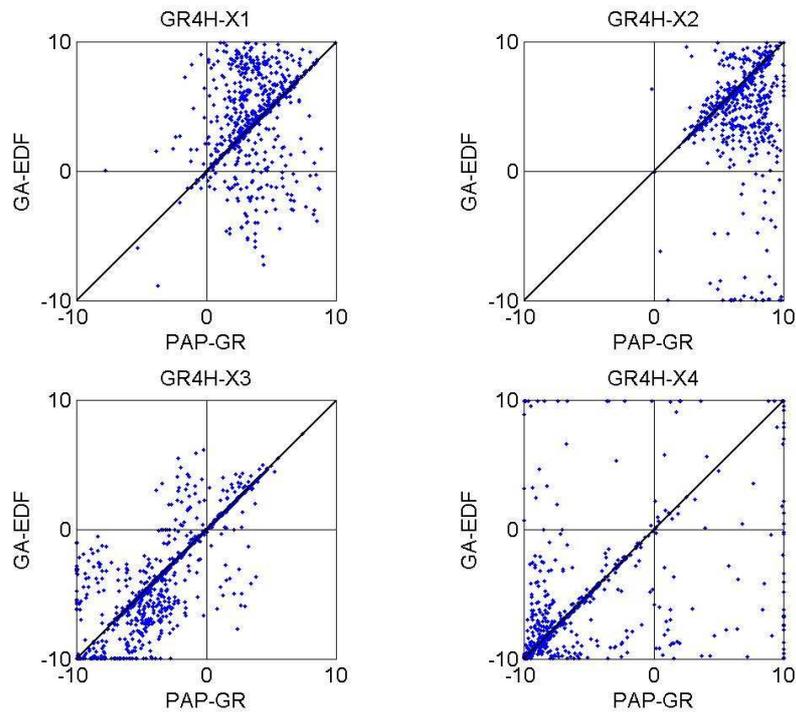


Figure 9-27 : Comparaison des valeurs des paramètres de GR4H optimisés avec la méthode PAP-GR et GA-EDF

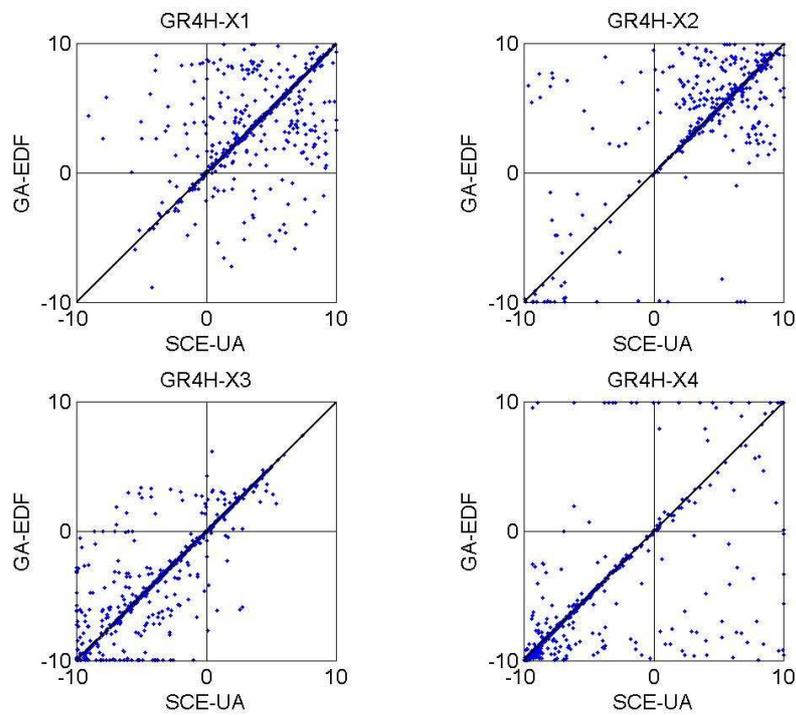


Figure 9-28 : Comparaison des valeurs des paramètres de GR4H optimisés avec la méthode SCE-UA et GA-EDF

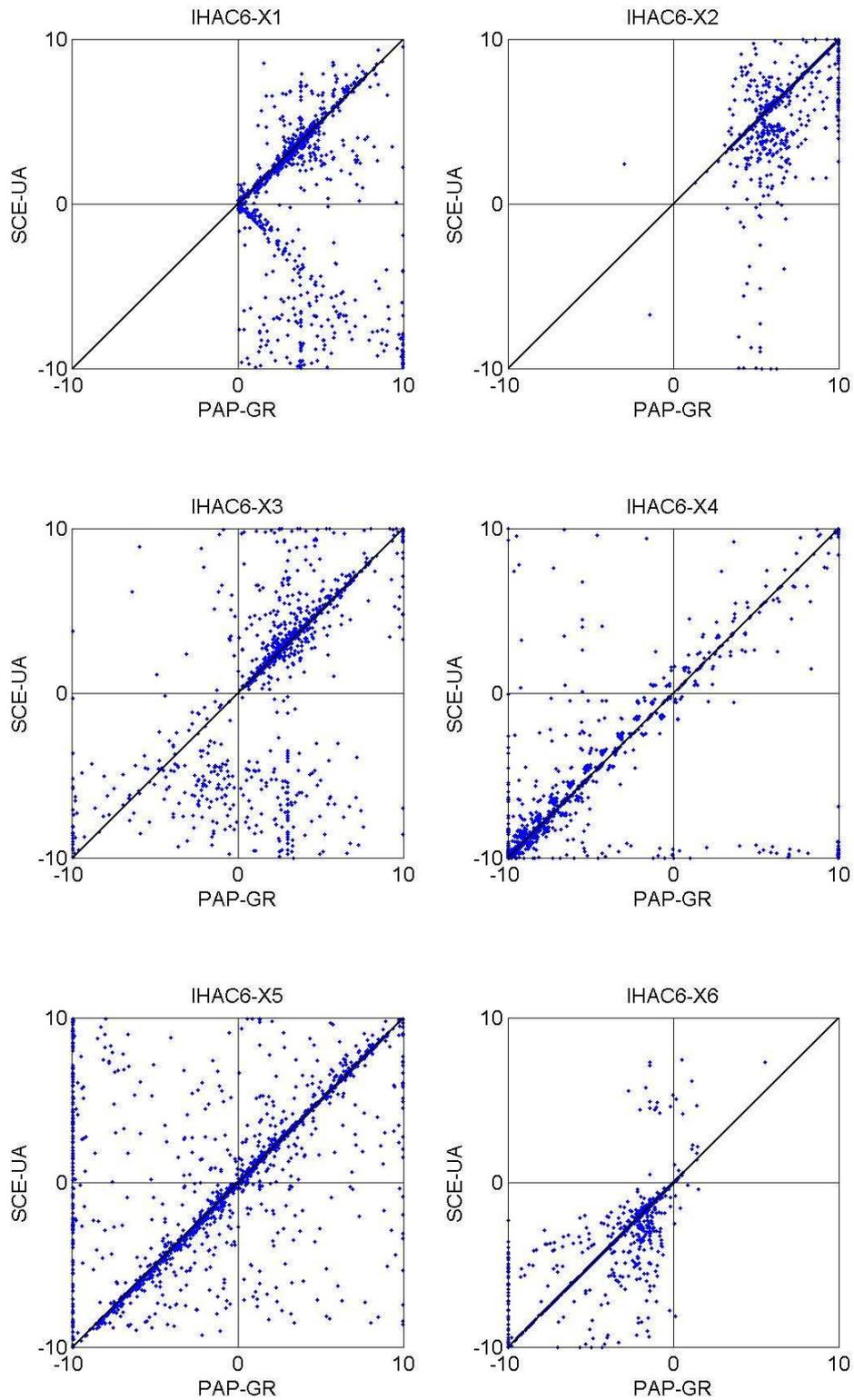


Figure 9-29 : Comparaison des valeurs des paramètres de IHAC6 optimisés avec la méthode PAP-GR et SCE-UA

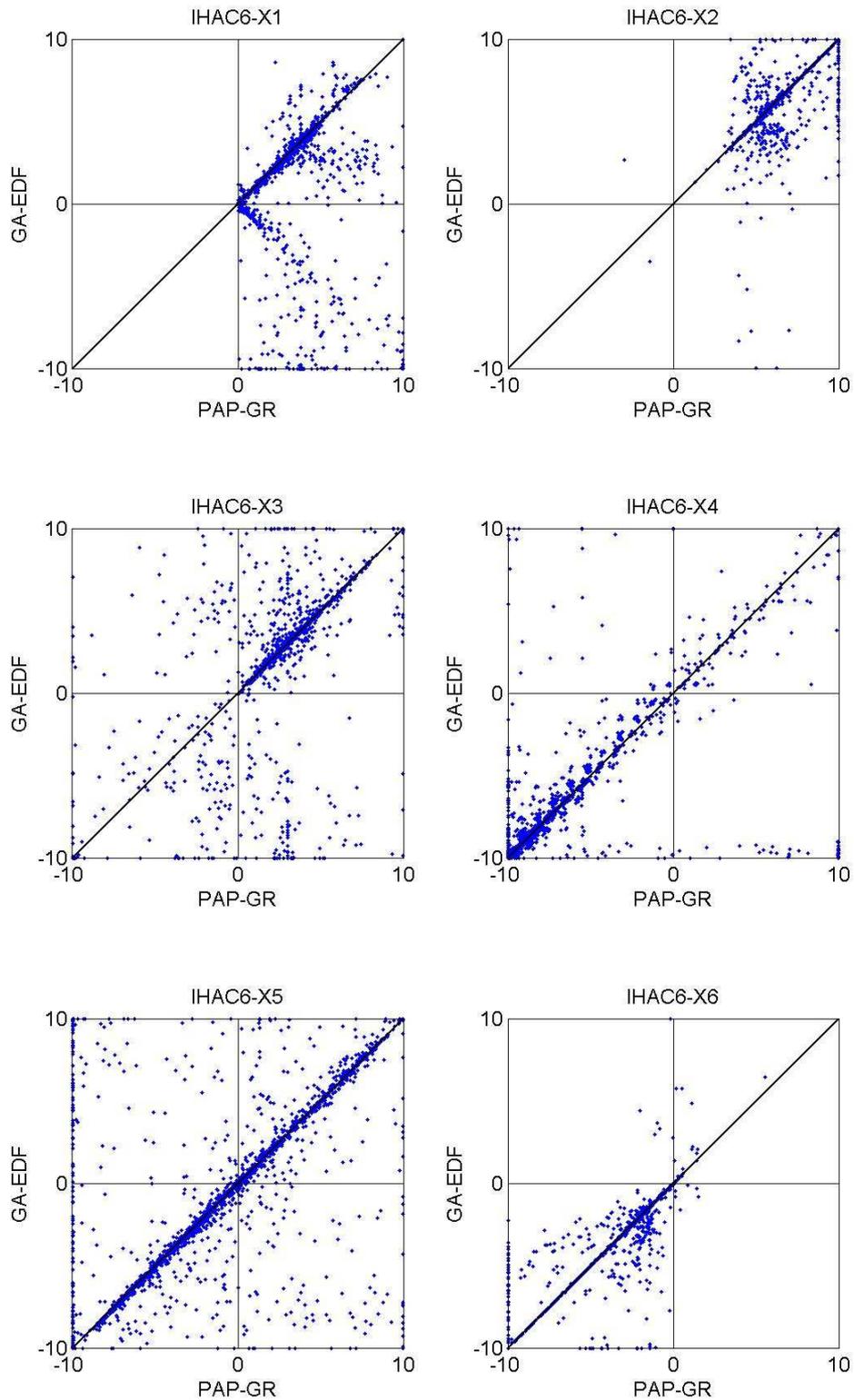


Figure 9-30 : Comparaison des valeurs des paramètres de IHAC6 optimisés avec la méthode PAP-GR et GA-EDF

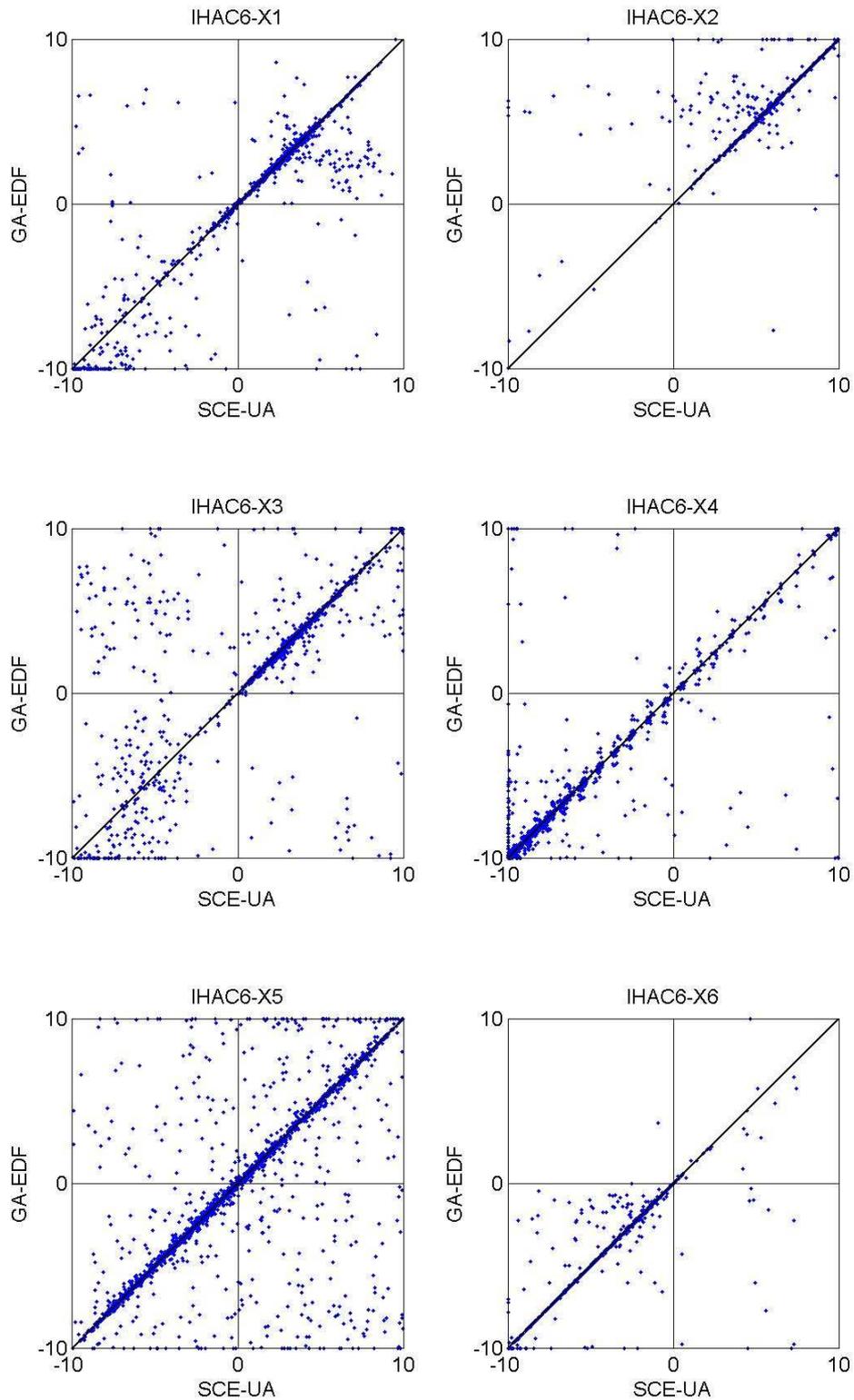


Figure 9-31 : Comparaison des valeurs des paramètres de IHAC6 optimisés avec la méthode SCE-UA et GA-EDF

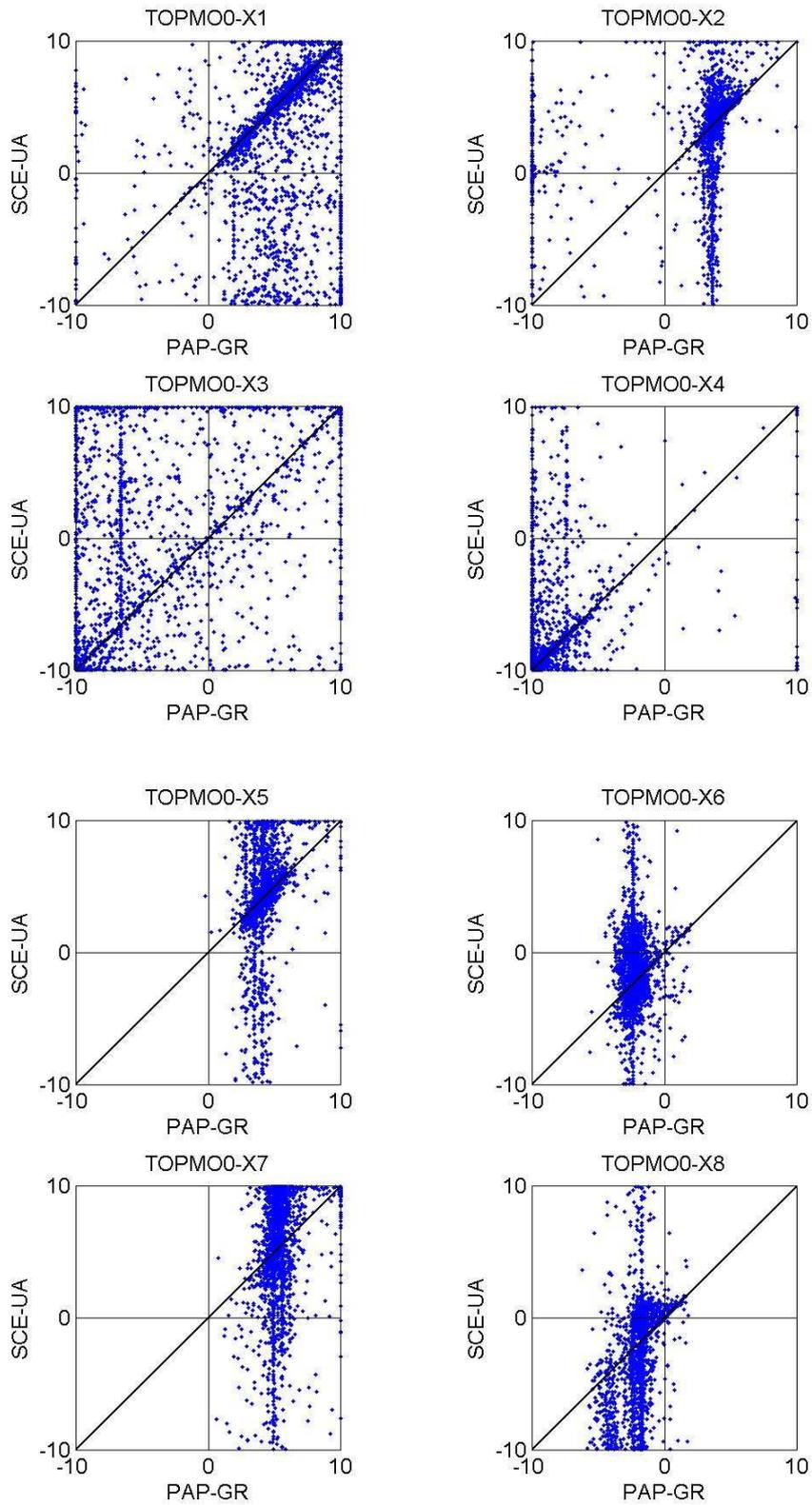


Figure 9-32 : Comparaison des valeurs des paramètres de TOPMO0 optimisés avec la méthode PAP-GR et SCE-UA

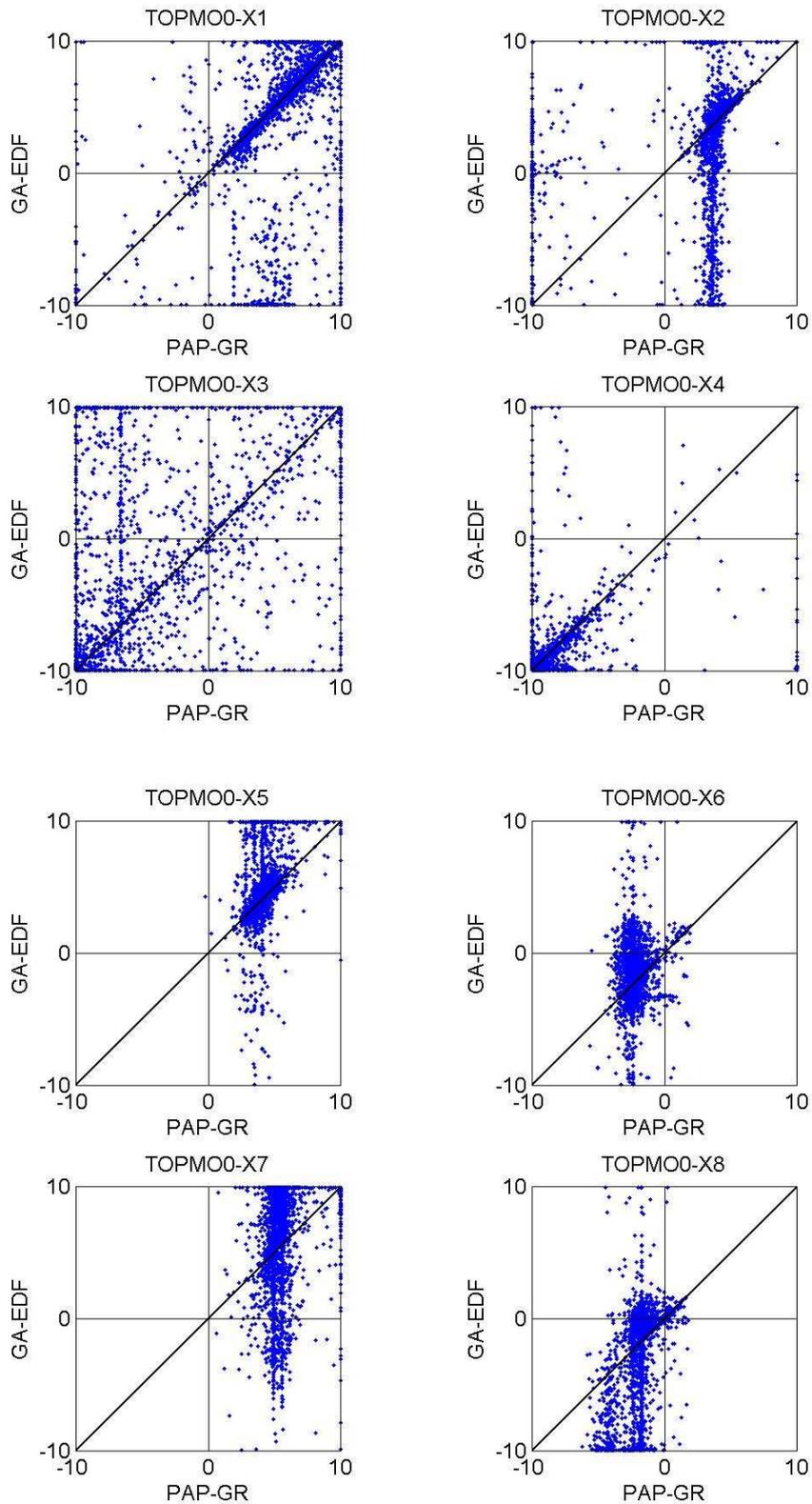


Figure 9-33 : Comparaison des valeurs des paramètres de TOPMO0 optimisés avec la méthode PAP-GR et GA-EDF

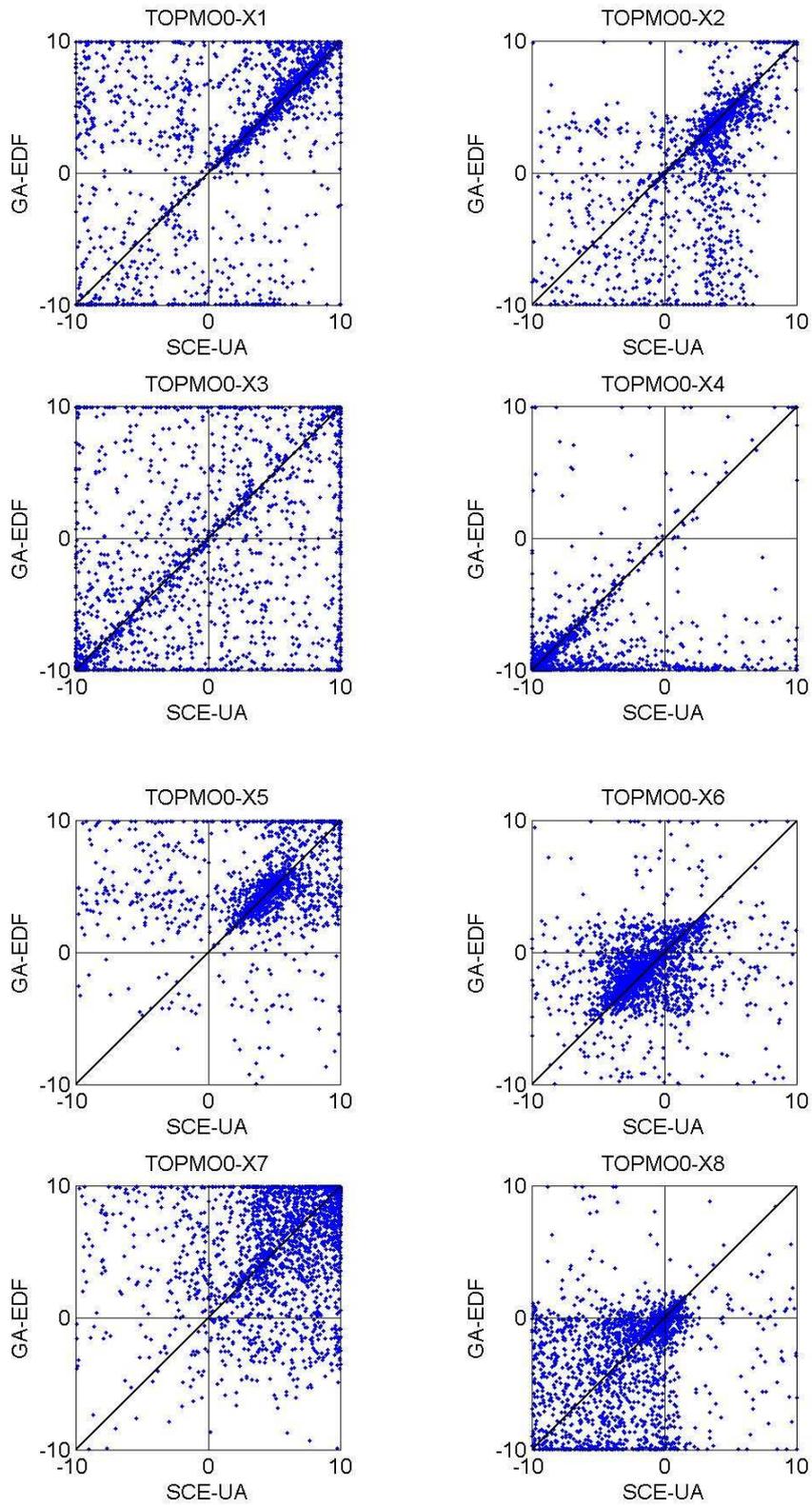


Figure 9-34 : Comparaison des valeurs des paramètres de TOPMO0 optimisés avec la méthode SCE-UA et GA-EDF

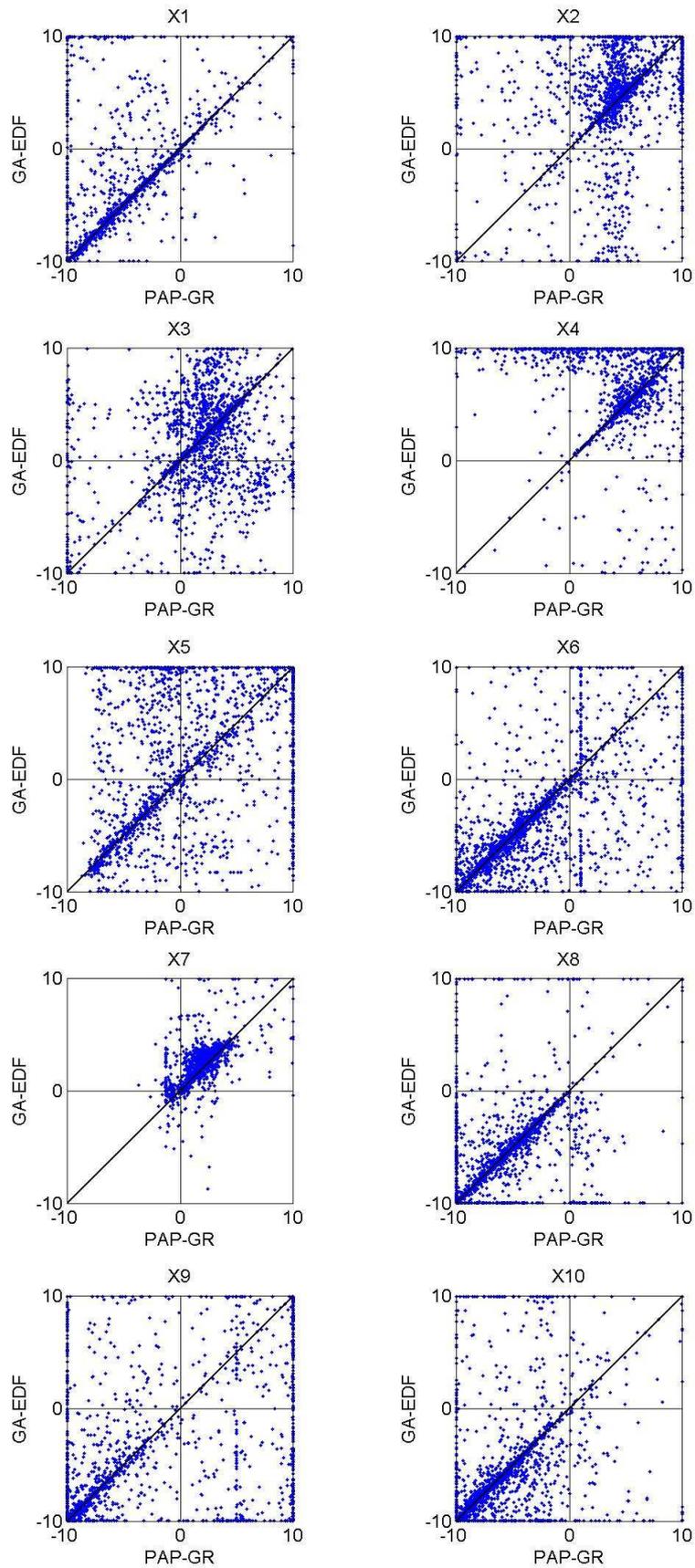


Figure 9-35 : Comparaison des valeurs des paramètres de MORDOR10 optimisés avec la méthode PAP-GR et SCE-UA

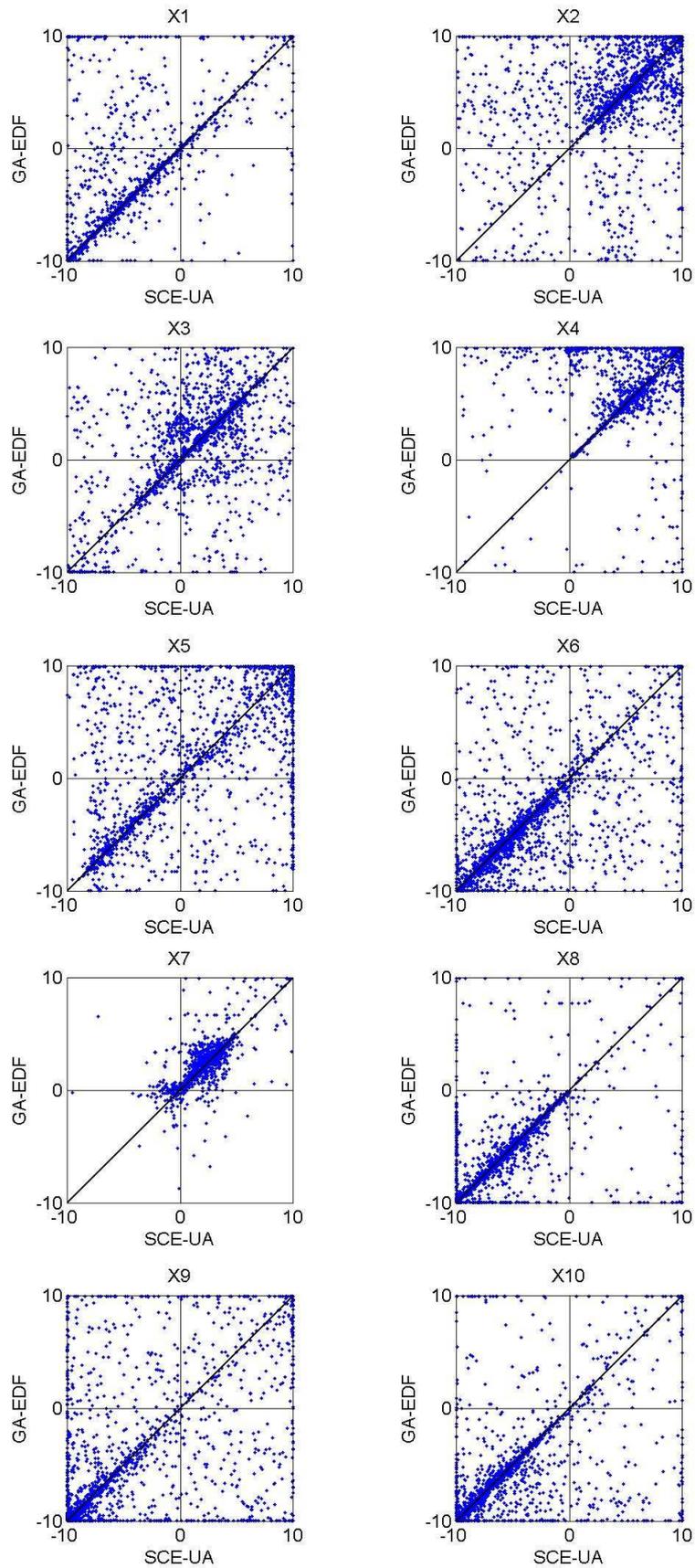


Figure 9-36 : Comparaison des valeurs des paramètres de MORDOR10 optimisés avec la méthode SCE-UA et GA-EDF

9.7 Distances entre paramètres et vecteurs de paramètres

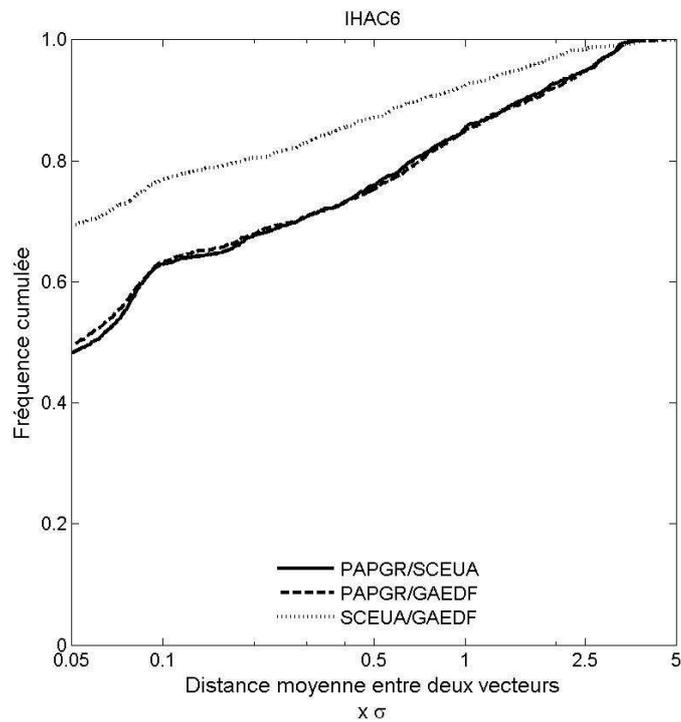


Figure 9-37 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différentes (modèle IHAC6).

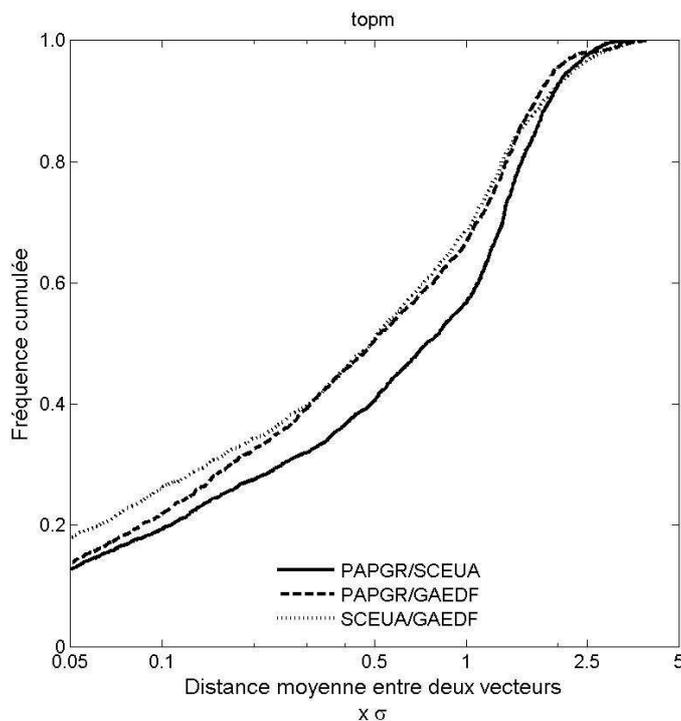


Figure 9-38 : Histogrammes des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différente, pour le modèle TOPMO0.

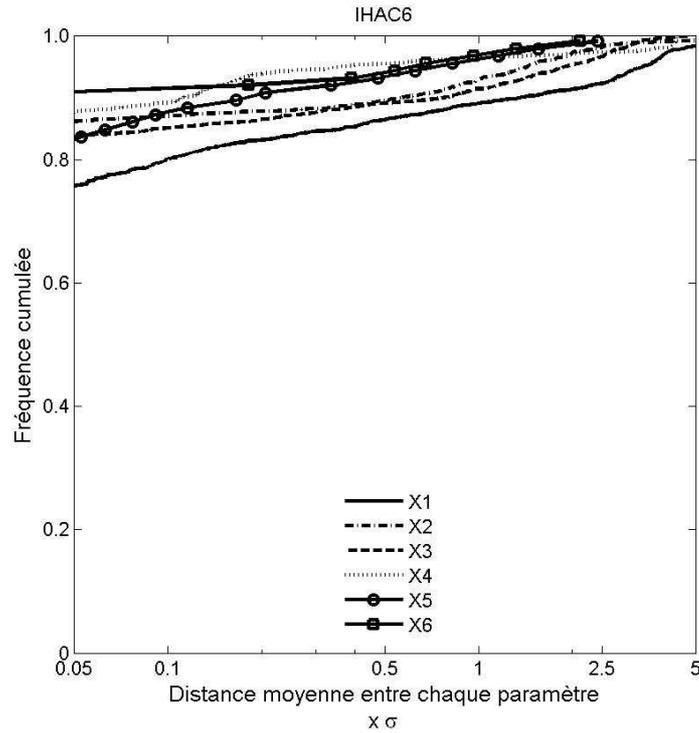


Figure 9-39 : Histogrammes des distances entre paramètres trouvés par les méthodes PAP-GR et SCE-UA, pour le modèles IHAC6.

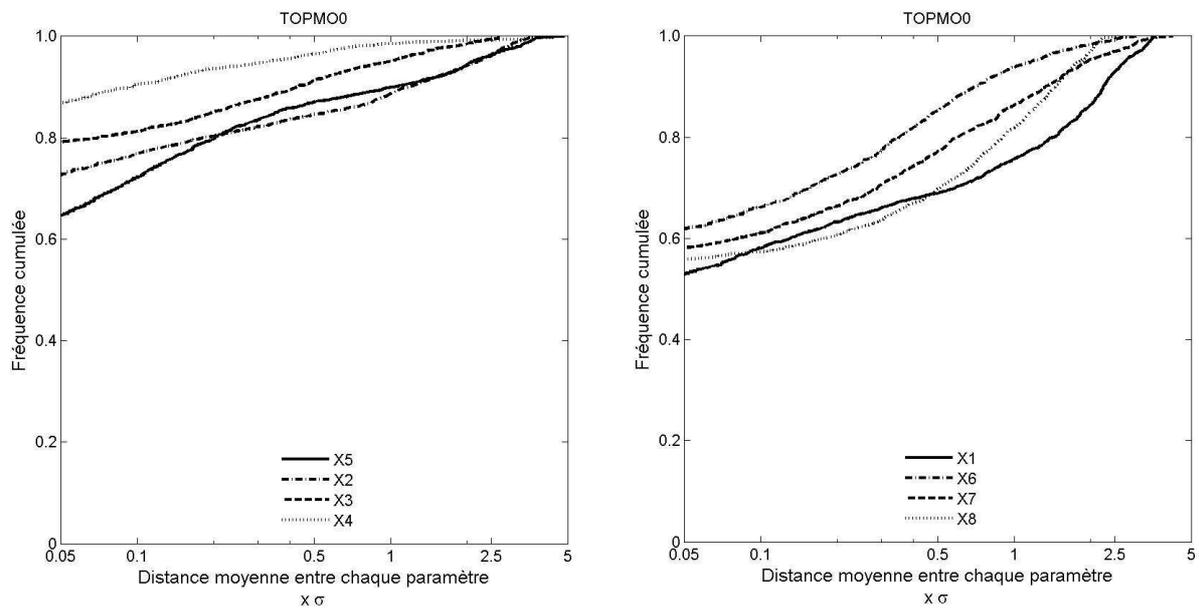


Figure 9-40 : Histogrammes des distances entre paramètres trouvés par les méthodes PAP-GR et SCE-UA, pour le modèle TOPMO0.

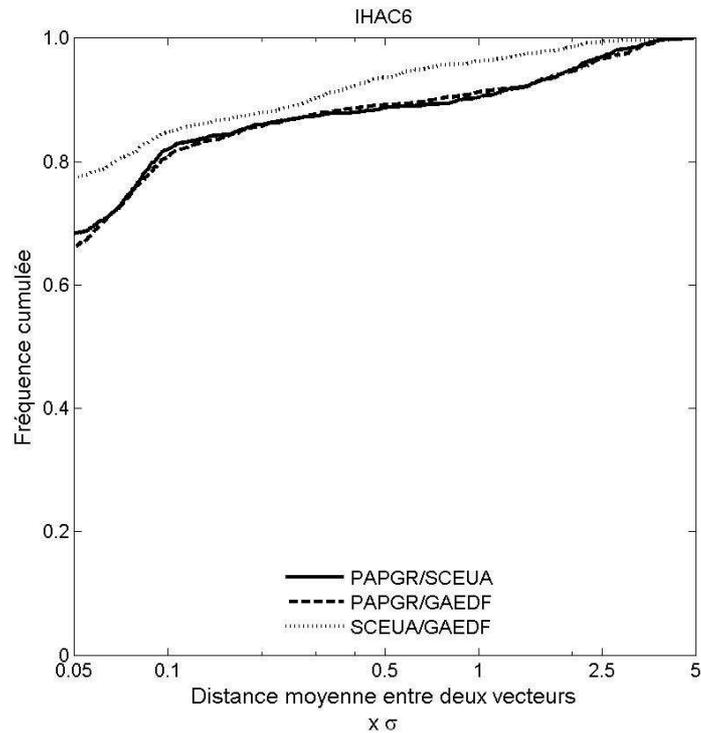


Figure 9-41 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différentes dont les performances sont égales, 91% des vecteurs (modèle IHAC6).

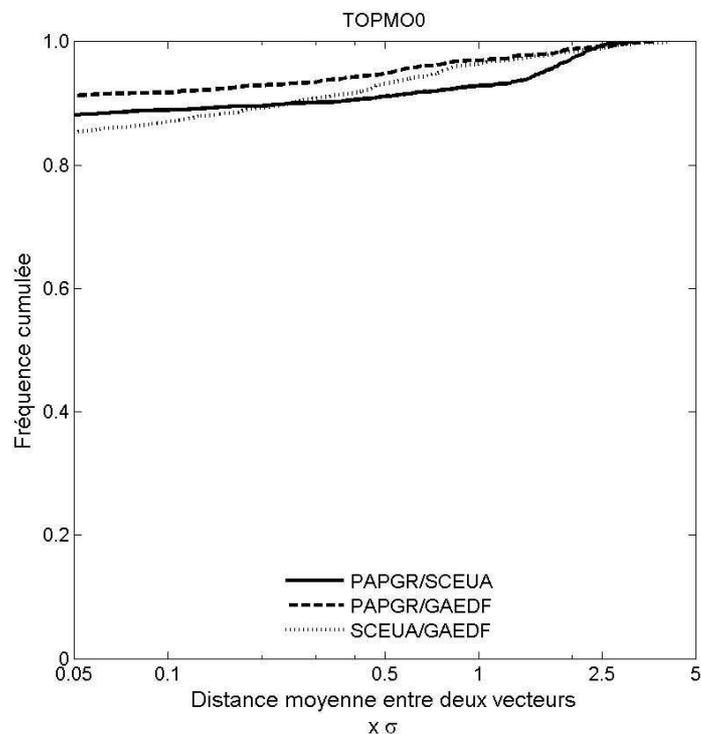


Figure 9-42 : Distribution cumulée des distances entre vecteurs de paramètres trouvés par deux méthodes d'optimisation différentes dont les performances sont égales, 9% des vecteurs (modèle TOPMO0).

Annexe 6 Comparaison de modèles

Tableau 9-3 : Performances moyenne obtenues lorsqu'on associe a posteriori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures. Performance moyenne du meilleur modèle seul : GR5H, $C_{2M}=39.5$. En gras, les combinaison supérieure à cette valeur. Fonction objectif en calage et en validation : $C_{2M}(Q)$ et $C_{2M}(RQ)$.

		Modèle A																			
		MODB	TOP8	NAM0	GARD	PDM	MART	BLAE	GR4J	TOP6	THOR	HBV0	CREC	IHAC	WAGE	GR3H	GR4H	MORD	GR5H	MOR6	
Modèle B	XINJ	32.0	32.9	31.3	30.6	33.7	30.2	32.7	37.1	35.3	30.4	32.4	34.5	34.5	34.8	37.9	40.3	40.1	40.8	39.8	
	MODB		33.0	31.3	28.6	31.0	28.4	32.9	35.3	35.7	29.5	32.1	34.1	32.2	33.4	37.5	40.3	39.4	40.3	38.8	
	TOP8			31.8	31.7	34.7	31.4	33.4	37.4	31.9	32.4	32.2	34.9	35.4	35.6	38.0	40.5	40.9	41.5	40.8	
	NAM0				31.0	33.5	29.1	30.0	36.8	35.0	31.1	32.2	36.2	35.2	34.3	37.7	40.6	41.2	41.7	41.1	
	GARD					31.1	26.2	32.8	35.7	34.2	28.7	30.9	31.4	32.6	33.2	37.3	39.9	39.3	40.3	38.8	
	PDM						31.4	34.9	37.5	36.5	31.9	33.2	34.6	34.0	34.5	38.4	40.9	39.8	40.8	39.5	
	MART							31.6	35.7	34.5	28.1	30.7	32.1	33.3	33.1	37.8	40.2	39.8	40.6	39.5	
	BLAE								37.4	35.9	32.6	33.6	37.2	35.7	35.0	38.3	40.4	41.3	41.7	40.8	
	GR4J									38.7	36.5	37.0	39.3	36.8	38.4	39.7	40.6	41.3	41.1	40.8	
	TOP6										35.5	34.0	35.6	37.2	37.5	39.1	41.2	41.5	42.0	41.3	
	THOR											31.7	34.4	33.1	33.2	38.1	40.3	40.1	40.7	39.6	
	HBV0												33.3	34.6	34.6	37.6	40.5	39.7	41.0	39.3	
	CREC													36.3	37.4	39.1	41.5	40.2	41.5	39.8	
	IHAC														36.0	38.8	40.8	40.1	40.8	39.6	
	WAGE																40.4	41.8	41.4	42.1	41.1
	GR3H																	40.7	41.6	41.9	40.6
	GR4H																		42.5	41.6	42.3
	MORD																			42.4	40.7
	GR5H																				42.1

ANNEXE

Tableau 9-4 : Performances moyenne obtenues lorsqu'on associe a posteriori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures. Performance moyenne du meilleur modèle seul : MOR6, $C_{2M}=40.0$. En gras, les combinaison supérieure à cette valeur. Fonction objectif en calage et en validation : $C_{2M}(RQ)$ et $C_{2M}(Q)$.

		Modèle A																		
		MODB	TOP8	NAM0	GARD	PDM	MART	BLAE	GR4J	TOP6	THOR	HBV0	CREC	IHAC	WAGE	GR3H	GR4H	MORD	GR5H	MOR6
Modèle B	XINJ	33.5	32.8	31.7	34.0	34.3	34.1	33.5	36.8	35.6	33.8	34.2	36.8	36.7	37.3	39.3	39.8	41.9	40.8	41.9
	MODB		35.4	32.5	33.7	33.0	33.9	34.6	36.1	37.9	33.9	35.0	39.1	35.1	37.6	38.9	40.6	41.9	41.2	41.2
	TOP8			32.9	35.4	36.9	35.7	34.2	38.3	34.2	36.2	35.3	37.0	38.5	39.3	39.4	40.7	42.6	41.9	42.3
	NAM0				34.2	33.9	33.7	32.3	36.3	35.5	34.4	34.6	37.2	37.1	37.3	38.3	40.0	41.9	41.1	41.6
	GARD					35.1	33.1	35.0	36.9	37.8	35.0	36.4	36.4	37.4	38.5	39.7	40.5	41.9	41.2	41.6
	PDM						34.7	35.5	37.6	38.7	35.4	36.2	38.8	36.7	37.8	39.6	40.4	41.9	41.1	41.6
	MART							34.9	37.7	37.6	34.5	35.9	36.1	37.7	38.6	40.0	41.1	42.5	41.7	41.9
	BLAE								37.1	36.5	35.6	35.9	37.3	38.0	38.0	39.0	39.9	41.8	40.9	42.1
	GR4J									39.9	37.9	38.5	40.7	38.0	39.0	40.2	39.8	41.7	40.3	42.1
	TOP6										38.2	37.2	38.3	40.2	41.0	40.7	41.5	43.5	42.6	43.1
	THOR											36.0	37.7	37.0	37.4	40.1	41.1	42.5	41.8	42.1
	HBV0												38.2	38.0	38.8	39.3	40.6	42.4	41.5	41.9
	CREC													41.1	40.8	41.7	42.2	43.4	42.8	43.1
	IHAC														39.0	40.6	41.4	42.1	41.8	41.9
	WAGE															41.6	41.7	42.5	42.3	42.7
	GR3H																40.3	42.9	41.7	42.3
	GR4H																	43.0	40.4	43.0
MORD																		43.2	42.5	
GR5H																			43.1	

ANNEXE

Tableau 9-5 : Performances moyenne obtenues lorsqu'on associe a posteriori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures. Performance moyenne du meilleur modèle seul : GR5H, $C_{2M}=47.0$. En gras, les combinaison supérieure à cette valeur. Fonction objectif en calage et en validation : $C_{2M}(RQ)$.

		Modèle A																		
		MODB	TOP8	NAM0	GARD	PDM	MART	BLAE	GR4J	TOP6	THOR	HBV0	CREC	IHAC	WAGE	GR3H	GR4H	MORD	GR5H	MOR6
Modèle B	XINJ	40.3	40.3	39.1	39.1	42.6	38.3	40.8	44.2	41.9	38.6	40.7	41.8	43.7	43.1	46.7	47.1	48.1	47.9	47.7
	MODB		41.2	38.2	37.9	41.0	36.9	40.3	42.9	43.7	38.9	41.9	43.7	42.3	42.9	47.0	47.6	47.6	48.0	47.7
	TOP8			39.5	40.2	44.1	39.5	41.0	44.9	40.1	40.5	41.3	42.6	44.6	44.3	46.5	47.7	48.7	48.7	48.3
	NAM0				38.4	41.5	35.6	36.7	43.2	41.7	38.9	40.9	42.8	43.7	42.6	45.4	46.9	48.3	48.0	48.0
	GARD					41.6	35.3	40.2	43.1	42.7	37.7	41.5	40.9	42.9	42.2	46.6	47.3	47.6	47.9	47.5
	PDM						40.8	42.9	45.0	45.6	41.7	44.0	45.1	44.2	44.1	48.0	48.3	48.6	48.6	48.7
	MART							39.2	43.4	42.4	36.8	40.6	40.4	42.7	42.2	46.5	47.5	47.9	48.1	47.7
	BLAE								43.9	42.9	40.7	42.0	43.7	44.5	42.9	45.9	46.9	48.2	47.9	48.0
	GR4J									46.1	44.1	45.1	46.5	44.8	45.4	47.7	47.3	48.6	47.8	48.4
	TOP6										42.7	42.7	44.0	46.1	45.7	47.6	48.1	49.5	49.3	49.1
	THOR											41.2	41.8	43.2	42.4	46.9	47.5	48.0	48.2	47.8
	HBV0												43.1	44.8	44.4	46.9	47.6	48.3	48.5	48.1
	CREC													46.0	45.4	48.3	48.5	48.5	48.9	48.4
	IHAC														45.7	48.0	48.1	48.3	48.4	48.3
	WAGE															48.7	48.4	49.0	49.0	48.9
	GR3H																48.0	50.0	49.4	49.4
	GR4H																	49.8	48.5	49.6
MORD																		50.0	49.0	
GR5H																			49.6	

Tableau 9-6 : Performances moyenne obtenues lorsqu'on associe a priori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures au calage. Performance moyenne du meilleur modèle seul : MOR6, $C_{2M}=39.9$. Fonction objectif en calage et en validation : $C_{2M}(RQ)$.

QQ	THOR	CREC	IHAC	WAGE	GR4H	GR5H	MOR6
TOP6	35.4	34.8	37.0	36.5	36.7	38.4	38.6
THOR		35.8	35.3	34.8	38.3	39.0	39.7
CREC			37.6	37.5	38.4	39.0	39.6
IHAC				36.5	39.3	39.7	40.5
WAGE					37.8	38.9	39.4
GR4H						38.9	40.4
GR5H							40.5

Tableau 9-7 : Performances moyenne obtenues lorsqu'on associe a priori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures au calage. Performance moyenne du meilleur modèle seul : GR5H, $C_{2M}=39.5$. Fonction objectif en calage et en validation : $C_{2M}(RQ)$.

QR	THOR	CREC	IHAC	WAGE	GR4H	GR5H	MOR6
TOP6	33.9	34.3	34.6	35.8	38.2	40.2	38.4
THOR		33.6	31.5	31.3	38.9	39.5	38.7
CREC			34.5	36.1	39.5	39.4	37.9
IHAC				33.4	39.7	39.8	38.4
WAGE					38.6	39.3	38.8
GR4H						41.0	40.8
GR5H							40.8

Tableau 9-8 : Performances moyenne obtenues lorsqu'on associe a priori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures au calage. Performance moyenne du meilleur modèle seul : MOR6, $C_{2M}=40.0$. Fonction objectif en calage et en validation : $C_{2M}(RQ)$.

RQ	THOR	CREC	IHAC	WAGE	GR4H	GR5H	MOR6
TOP6	35.6	35.2	37.4	38.0	36.6	38.5	39.5
THOR		35.6	35.2	34.6	38.9	39.4	40.3
CREC			38.6	38.8	39.9	40.4	40.5
IHAC				37.1	39.2	40.1	40.3
WAGE					39.5	40.1	39.9
GR4H						39.9	40.7
GR5H							41.1

Tableau 9-9 : Performances moyenne obtenues lorsqu'on associe a priori deux structures et qu'on choisit pour chaque bassin versant la meilleure des deux structures au calage. Performance moyenne du meilleur modèle seul : GR5H, $C_{2M}=47.0$. Fonction objectif en calage et en validation : $C_{2M}(RQ)$.

RR	THOR	CREC	IHAC	WAGE	GR4H	GR5H	MOR6
TOP6	40.7	42.0	43.5	43.9	44.8	46.4	45.8
THOR		40.2	41.0	40.5	45.3	45.9	45.4
CREC			44.2	44.0	46.4	46.8	45.4
IHAC				43.4	46.8	47.2	47.1
WAGE					45.4	45.8	45.6
GR4H						48.0	48.0
GR5H							48.0

Annexe 7 Recherche de caractéristiques discriminant les performances des modèles.

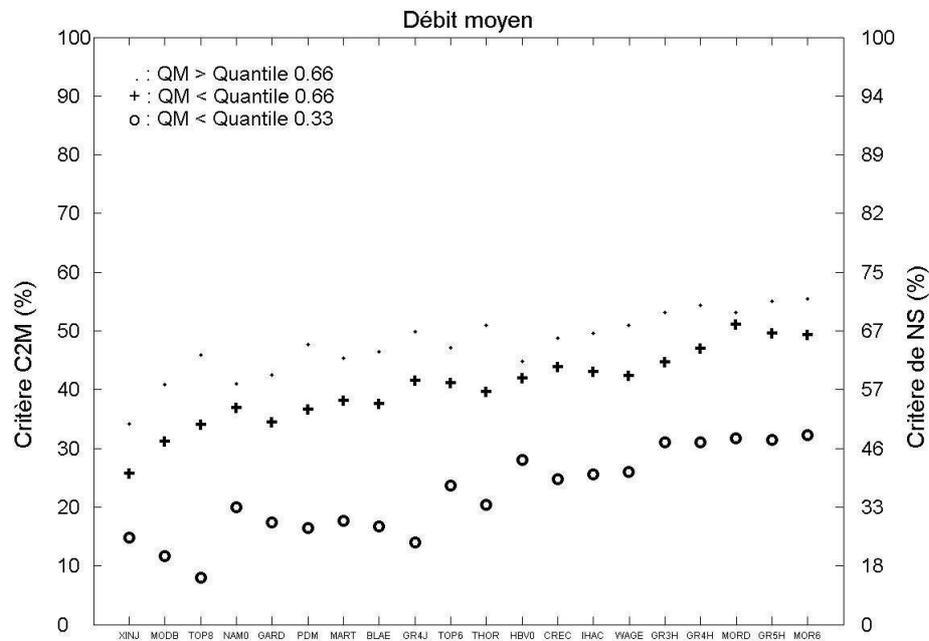


Figure 9-43 : Performances moyennes des 20 modèles en fonction de la classe du débit moyen annuel des bassins versants.

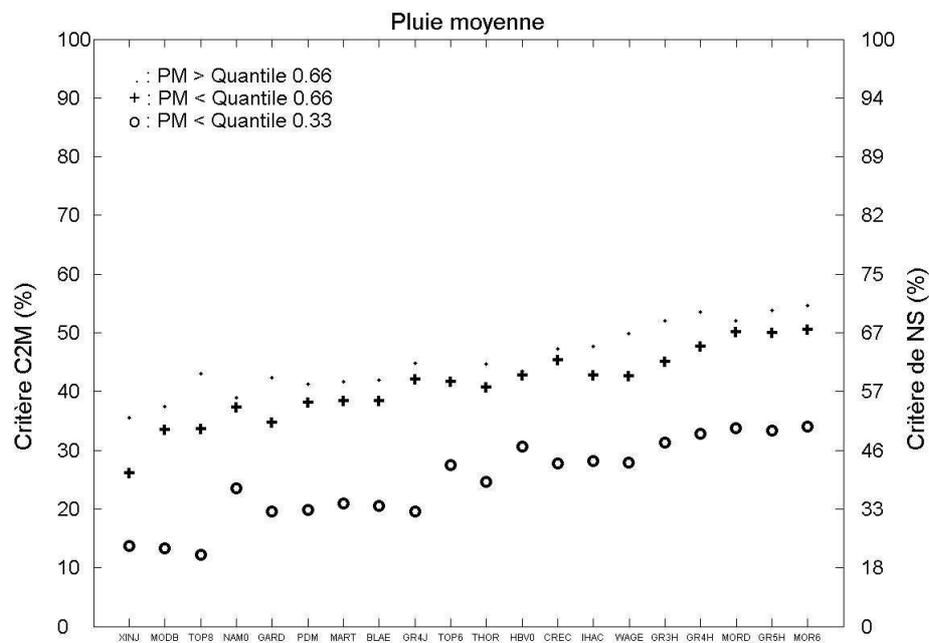


Figure 9-44 : Performances moyennes des 20 modèles en fonction de la classe de la pluie moyenne annuelle des bassins versants.

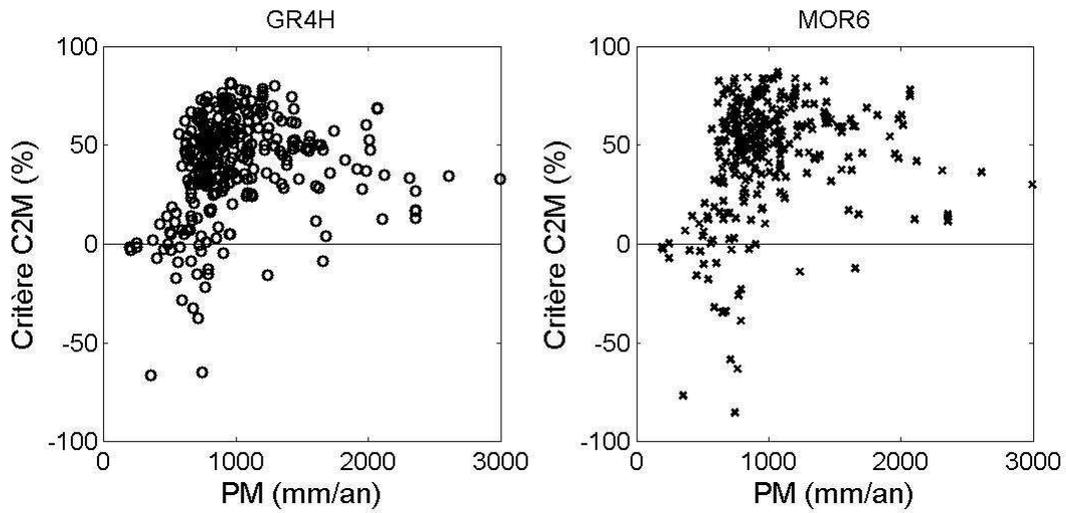


Figure 9-45 : Performances des modèles GR4H et MOR6 en fonction de la pluie moyenne annuelle des bassins versants.

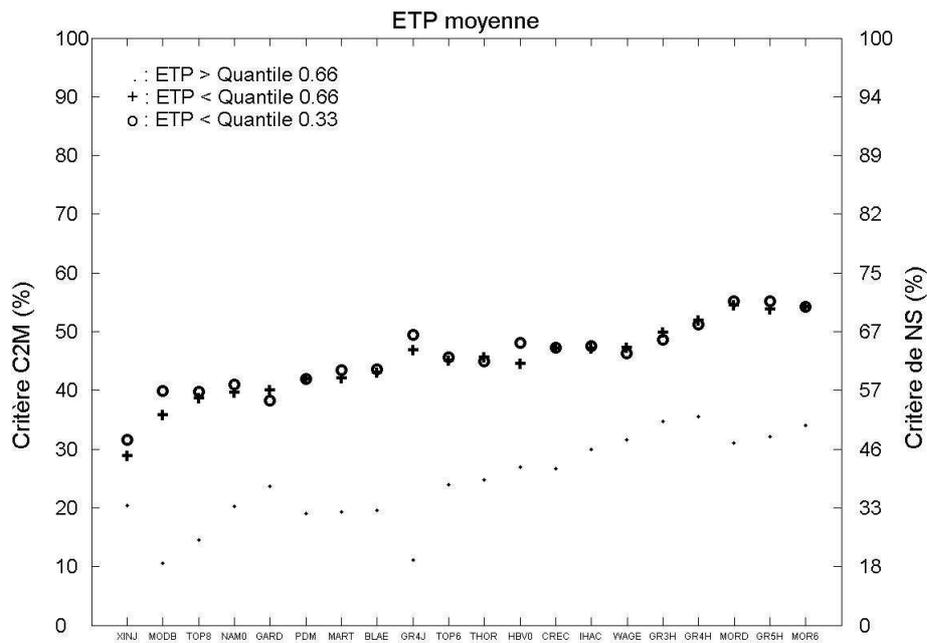


Figure 9-46 : Performances moyennes des 20 modèles en fonction de la classe d'ETP moyenne annuelle des bassins versants.

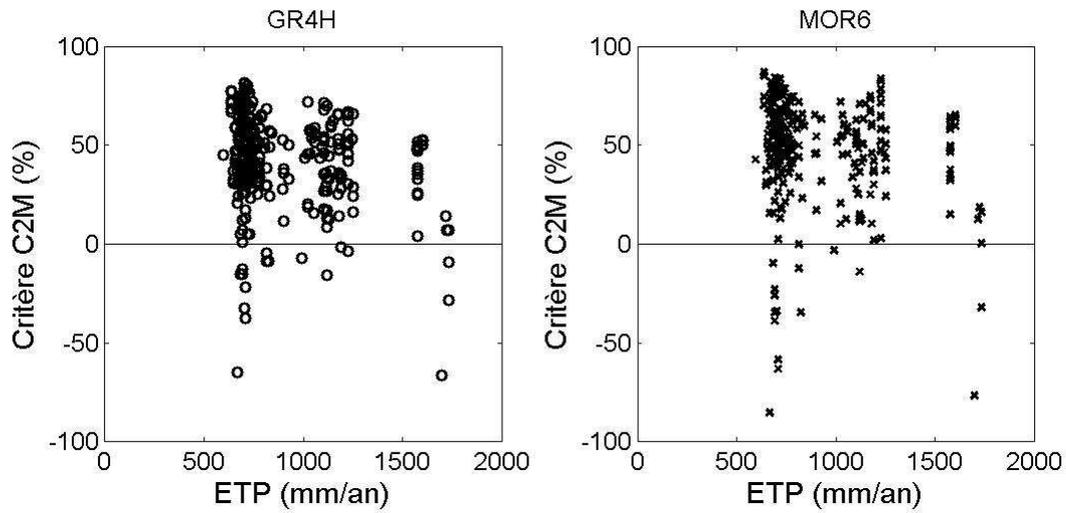


Figure 9-47 : Performances des modèles GR4H et MOR6 en fonction de l'ETP moyenne annuelle des bassins versants.

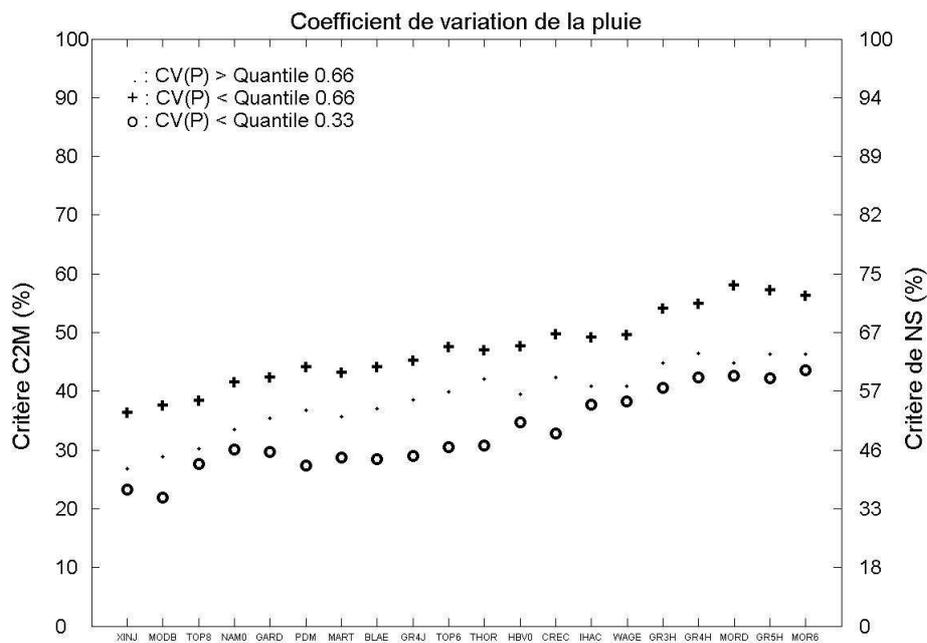


Figure 9-48 : Performances moyennes des 20 modèles en fonction de la classe du coefficient de variation de la pluie non nulle des bassins versants.

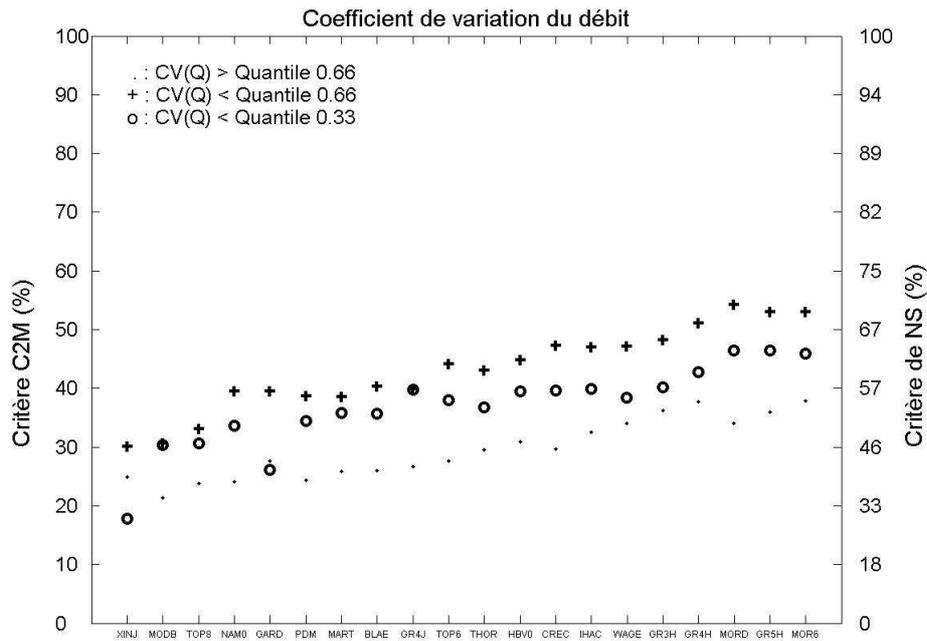


Figure 9-49 : Performances moyennes des 20 modèles en fonction de la classe du coefficient de variation du débit non nul des bassins versants.

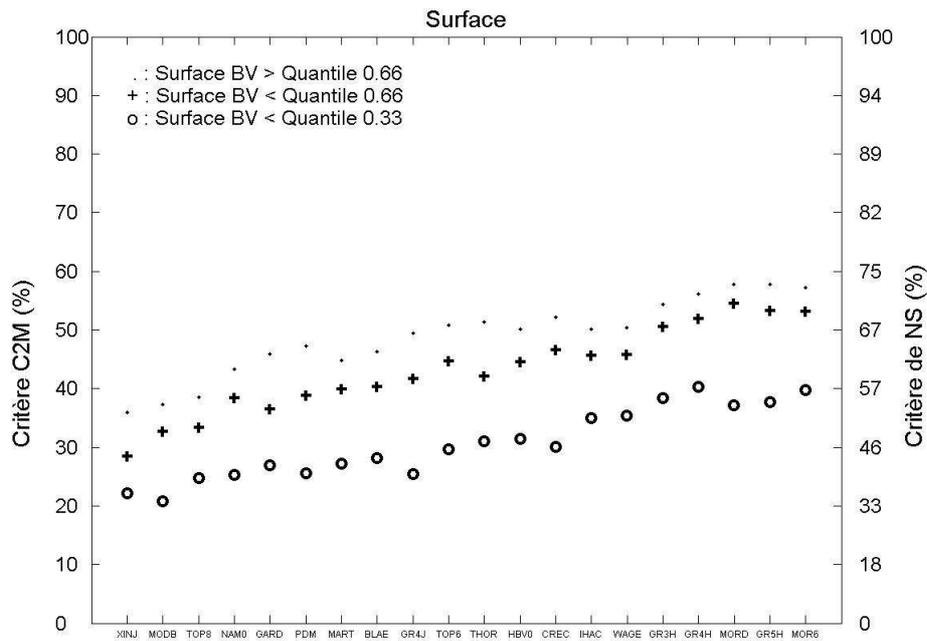


Figure 9-50 : Performances moyennes des 20 modèles en fonction de la classe de surface des bassins versants.

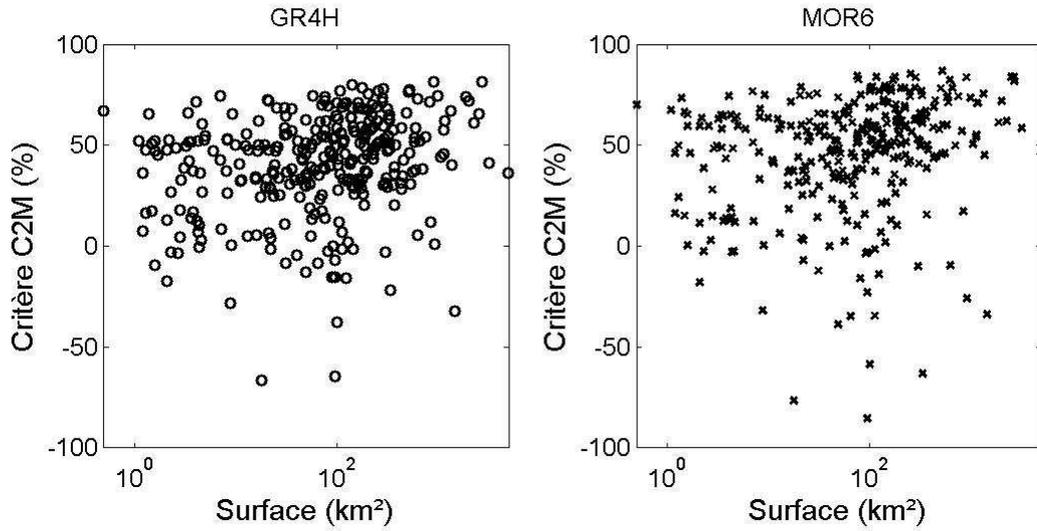


Figure 9-51 : Performances des modèles GR4H et MOR6 en fonction de la surface des bassins versants.

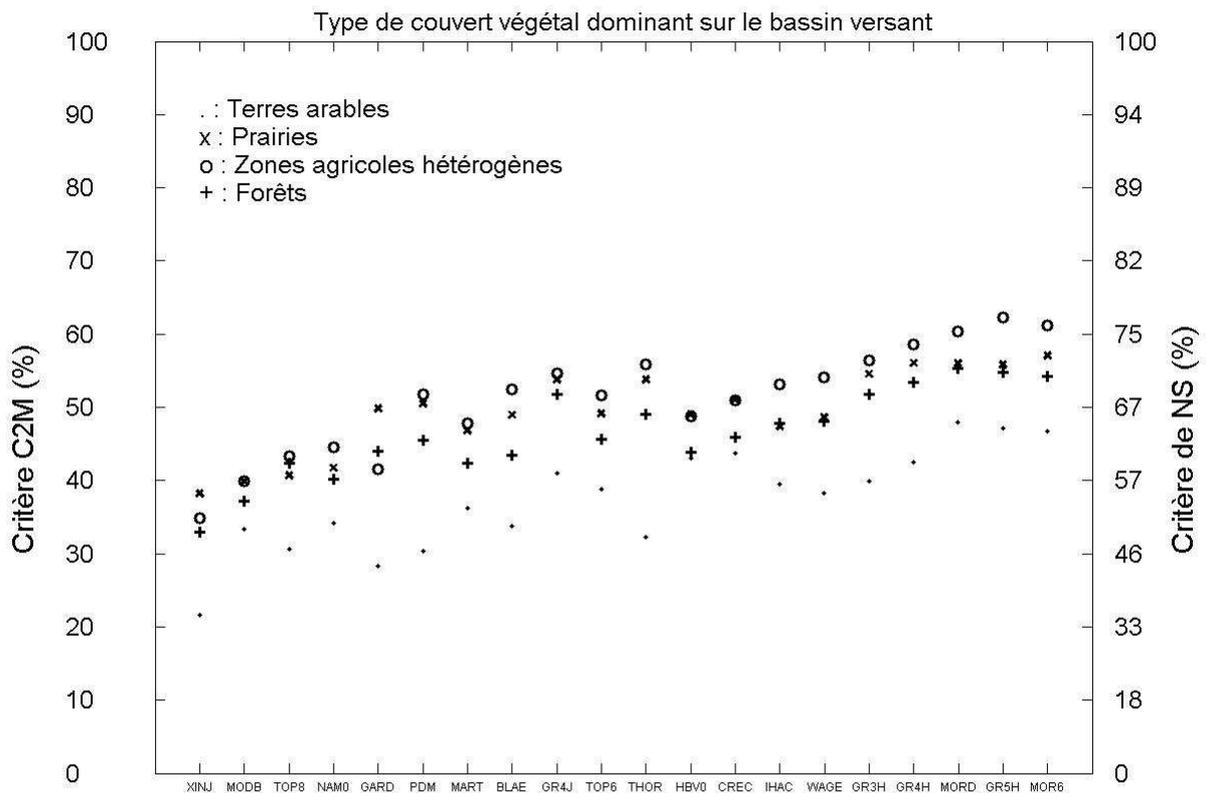


Figure 9-52 : Performances moyennes des 20 modèles en fonction du type de couverture végétale dominante sur chaque bassin versant (227 bassins versants français).

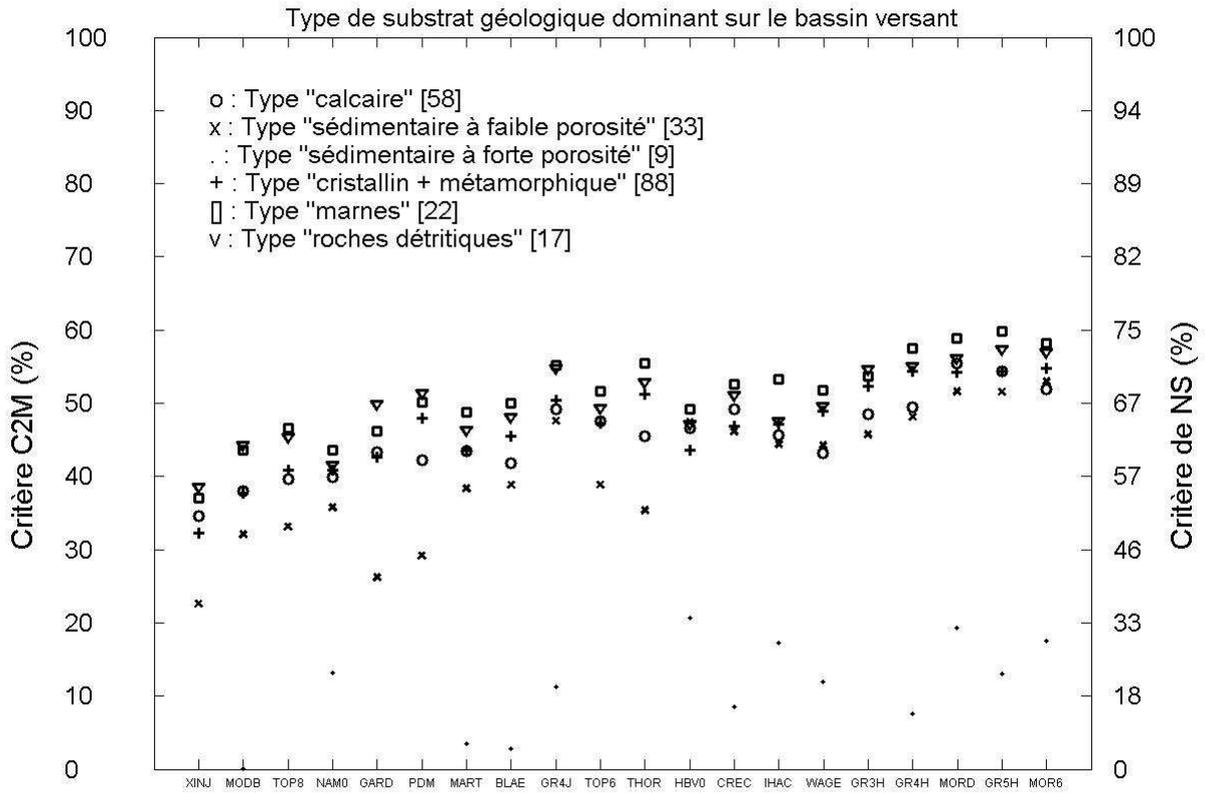


Figure 9-53 : Performances moyennes des 20 modèles en fonction du type de substrat géologique dominant sur chaque bassin versant (227 bassins versants français).

Annexe 8 Tentatives d'amélioration d'un modèles pluie-débit au pas de temps horaire en suivant une approche empirique

9.8 Modification de la fonction de percolation

Problème du coefficient de percolation.

Reprenons la loi instantanée de vidange d'un réservoir puissance :

$$Q = k.S^\alpha \quad [8.7]$$

avec, S : niveau d'eau dans le réservoir, k : constante de vidange du réservoir

Au cours d'une période de récession (non influencée), la variation du niveau du réservoir s'exprime ainsi :

$$dS = -k.S^\alpha .dt \quad [8.8]$$

$$-\frac{dS}{S^\alpha} = k.dt \quad [8.9]$$

$$-S^{-\alpha} .dS = k.dt \quad [8.10]$$

Lorsqu'on intègre cette expression sur un pas de temps on obtient :

$$\left[\frac{S^{-\alpha+1}}{\alpha-1} \right]_1^2 = [k.t]_0^{\Delta t} \quad [8.11]$$

avec, Δt : durée du pas de temps.

D'où :

$$\frac{1}{S_2^{\alpha-1}} = \frac{1}{S_1^{\alpha-1}} + k(\alpha-1)\Delta t \quad [8.12]$$

$$S_2^{\alpha-1} = \frac{1}{\frac{1}{S_1^{\alpha-1}} + \frac{1}{K^{\alpha-1}}} \quad [8.13]$$

avec :

$$K = [k(\alpha-1)\Delta t]^{\frac{1}{1-\alpha}} \quad [8.14]$$

$$S_2^{\alpha-1} = \frac{S_1^{\alpha-1}}{1 + \left(\frac{S_1}{K}\right)^{\alpha-1}} \quad [8.15]$$

$$S_2 = \frac{S_1}{\left[1 + \left(\frac{S_1}{K}\right)^{\alpha-1}\right]^{\frac{1}{\alpha-1}}} \quad [8.16]$$

D'où, dans notre cas, la percolation est égale à la baisse de niveau au cours du pas de temps considéré :

$$Perc = S_1 - S_2 = \int_0^{\Delta t} Q(t).dt \quad [8.17]$$

$$Perc = S_1 \left[1 - \frac{1}{\left[1 + \left(\frac{S_1}{K}\right)^{\alpha-1}\right]^{\frac{1}{\alpha-1}}} \right] \quad [8.18]$$

Ce développement met en évidence de la dépendance du facteur K, en fonction du pas de temps Δt . Si l'on passe de $\Delta t = 24$ h à $\Delta t = 1$ h, on doit théoriquement multiplier K par $24^{1/4}$. La loi de percolation dépend déjà d'un paramètre optimisable (X_1), c'est la raison pour laquelle nous n'avons pas testé le modèle avec le coefficient de percolation optimisable, afin d'éviter l'interaction entre deux paramètres optimisables. On estime K par X_1/β , avec β identique pour tous les bassins versants. Dans notre cas, on devrait 'théoriquement' passer d'un coefficient de percolation β de 4/9, à un coefficient de percolation proche de 1/5.

Annexe 9 Proposition d'une méthode de calage discret

9.9 Méthodologie de calage discret

Tous les tests effectués sur la comparaison de différentes méthodes d'optimisation et la comparaison des performances de différents modèles, ont révélé le manque de robustesse de certains modèles, dont les vecteurs de paramètres sont parfois sur-adaptés à la période de calage, sur laquelle ils sont déterminés.

Ce constat nous a amené à proposer une méthode de calage discret, basée sur la méthode de pré-filtrage précédente. Cette méthode de calage discret a pour objectif d'identifier la région de l'optimum en utilisant seulement la méthode de pré-filtrage. Une fois que le meilleur des points de départ possibles a été identifié, la méthode de calage s'arrête sur ce vecteur et ne procède pas à une recherche par optimisation locale d'un meilleur vecteur de paramètre (calage continu). Ce vecteur de paramètre est ensuite directement testé en validation. De cette manière, en effectuant un calage discret grossier, il est probable que les vecteurs de paramètres identifiés soient plus généraux et ainsi plus robustes.

Cette méthode de calage discret a également été proposée dans une optique de régionalisation des paramètres d'un modèle. A titre d'exemple, elle n'a été testée que sur le modèle GR4H. La méthode de pré-filtrage de la surface de réponse d'un modèle permet de définir des classes type de valeurs de paramètres, par exemple :

- Lorsqu'on divise la distribution en deux parties : valeurs de paramètres plutôt faibles ou plutôt fortes,
- Lorsqu'on divise la distribution en trois parties : valeurs de paramètres plutôt faibles, moyennes ou plutôt fortes,
- Etc..

Ainsi les n^m vecteurs possibles qui en découlent sont autant de type de bassins versants. Dans ce cadre, la régionalisation des vecteurs de paramètres peut devenir plus facile, si les relations entre paramètres et facteurs explicatifs ne se font pas entre les valeurs optimisées typiquement pour un couple bassin/période, mais plutôt avec, soit des classes de valeurs de paramètres (plutôt faible, moyen ou plutôt fort), soit directement avec les n^m vecteurs définissant les bassins types.

Dans un premier temps, nous allons évaluer la potentialité de cette méthode de calage discret, c'est-à-dire, les niveaux de performance obtenue lorsque les vecteurs de paramètres ne sont pas affinés.

9.10 Performances des modèles

La présentation des résultats est très simple et basée sur la comparaison des performances en calage et en validation du modèle GR4H. La Figure 9-54 présente les performances du modèle GR4H en calage et en validation, lorsqu'on utilise un calage discret avec 2, 3, 4 et 5 divisions de la distribution des paramètres (respectivement 16, 71, 148 et 238 points de départs testés à chaque période ; C_{2M} seuil = 70). Sur cette figure, sont aussi indiquées à titre comparatif les performances de la méthode PAP-GR₁ (calage continu). Cette figure montre clairement les plus faibles performances en calage de la méthode de calage discret. Néanmoins, plus le nombre de points de départ augmente, plus les performances augmentent, même si elles restent fortement inférieures à celles obtenues lors d'un calage continu (au minimum, 10 points de C_{2M}). Cependant, en validation, cette figure met en évidence que le niveau de performance atteint, lorsqu'on utilise le plus grand nombre de points possibles, est assez proche de celui atteint lors du calage continu. La différence de performances est seulement de trois points. Ce résultat permet de renforcer les conclusions précédentes, puisque les performances obtenues par un calage discret indiquent qu'il n'est pas nécessaire de raffiner trop le calage d'un vecteur de paramètres, connaissant la variabilité des conditions hydro-climatiques. Ces résultats montrent que des vecteurs calés de manière grossière (on se contente de repérer la région optimale de la surface de réponse) peuvent fournir de bons niveaux de performances au cours d'autres périodes hydrologiques, pour lesquels ils sont suffisamment généraux.

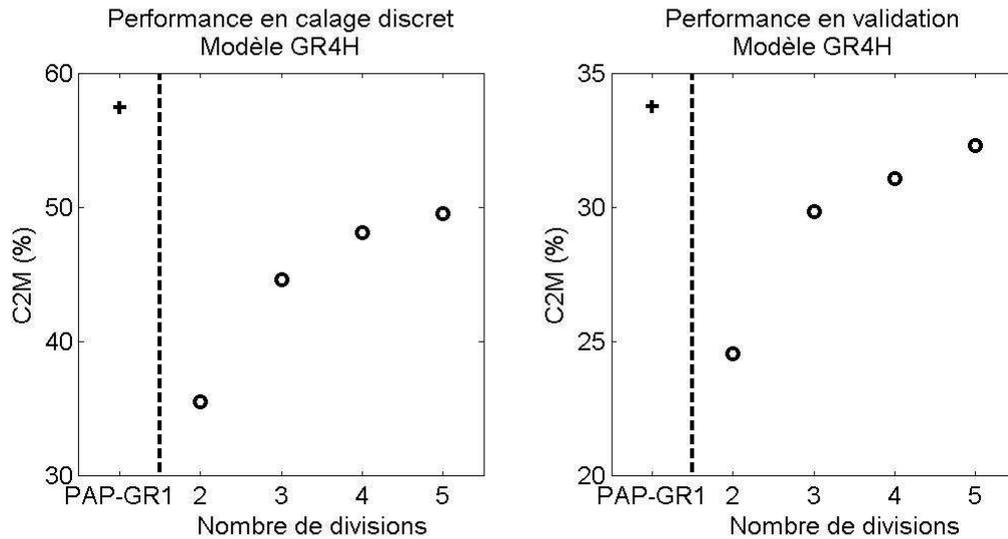


Figure 9-54 : Performances moyennes du modèle GR4H lorsqu'il est calé de manière discrète. Illustration de performances en fonction du nombre de divisions choisies, comparées aux performances lors d'un calage continu (méthode PAP-GR1).

La Figure 9-55 compare les performances obtenues pour chaque période de calage et de validation avec les vecteurs de paramètres optimisés selon un calage discret (points de départ de PAP-GR₃) et un calage continu (PAP-GR₃). Cette figure souligne clairement les résultats obtenus précédemment et montre les plus faibles performances en calage des vecteurs obtenus selon un calage discret. En revanche, en validation, la tendance s'inverse et les performances des vecteurs obtenus selon les deux méthodes de calage sont presque équilibrées. La majeure partie du nuage de points se trouve dans une fourchette de plus ou moins 20 points de C_{2M} autour de la première bissectrice. Parmi les vecteurs optimisés selon la méthode de calage discret, 37% des vecteurs obtiennent de meilleures performances en validation, alors qu'elles étaient inférieures au cours du calage.

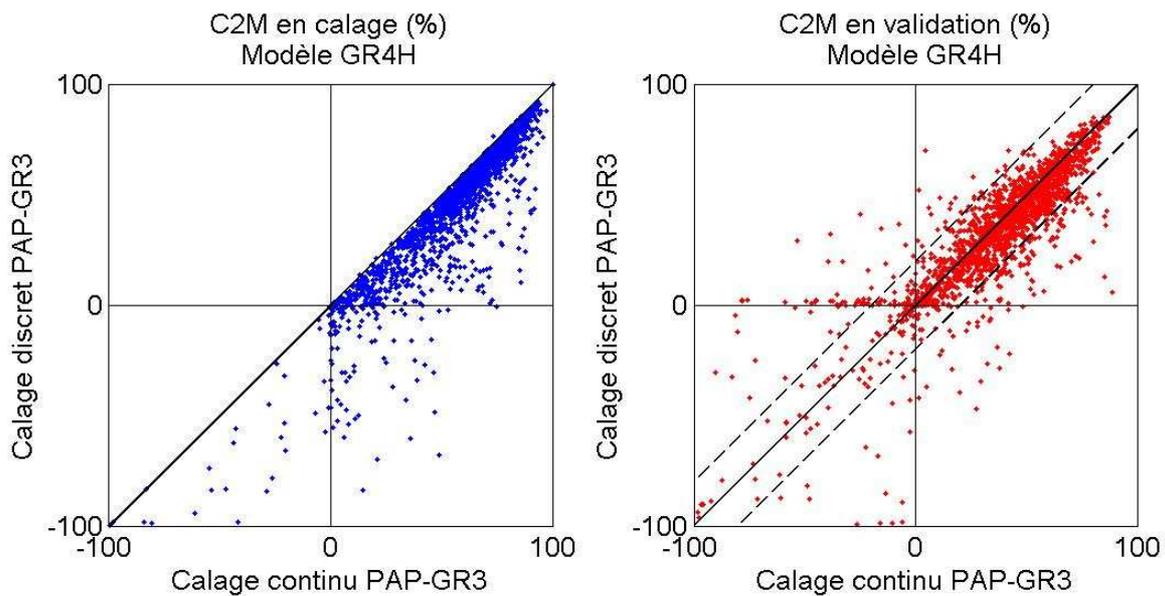


Figure 9-55 : Comparaison des performances en calage et en validation des vecteurs optimisés selon un calage discret (meilleur des points de départ de PAP-GR3) et un calage continu (PAP-GR3). Les pointillés indiquent la limite de +/- 20 points de C_{2M} .

Ainsi, cette méthode de calage discret est très simple, puisqu'elle teste une liste de points issus de la distribution des valeurs de chaque paramètre. Elle est capable de fournir des vecteurs, dont les performances sont relativement faibles en calage, mais dont les performances en validation sont relativement bonnes. Les performances moyennes sont seulement inférieures de 3 points aux performances obtenues au cours d'un calage discret. Par rapport aux 20 modèles testés précédemment, le classement de ce modèle, dont les vecteurs seraient optimisés selon cette méthode de calage discret, serait autour de la septième place, ce qui est tout à fait bien, vu la simplicité de cette méthode de calage.

La Figure 9-56 représente le nombre d'occurrences où un des vecteurs de départ possibles est le meilleur des vecteurs (c'est-à-dire, le vecteur calé), pour les quatre stratégies de calage discret testées (2, 3, 4, 5 divisions). Cette figure montre une distribution assez inégale des occurrences, d'autant plus inégale que le nombre de divisions est grand. Par ailleurs, comme l'indique le pointillé vertical sur cette figure, une faible proportion des vecteurs de départ possibles représente la majorité des meilleurs vecteurs. En effet, suivant le nombre de divisions choisies, 50% à 60% des vecteurs de départ possibles représentent 80% des meilleurs vecteurs. Ce constat peut laisser penser qu'il est encore possible de réduire *a posteriori* les vecteurs de départ possibles.

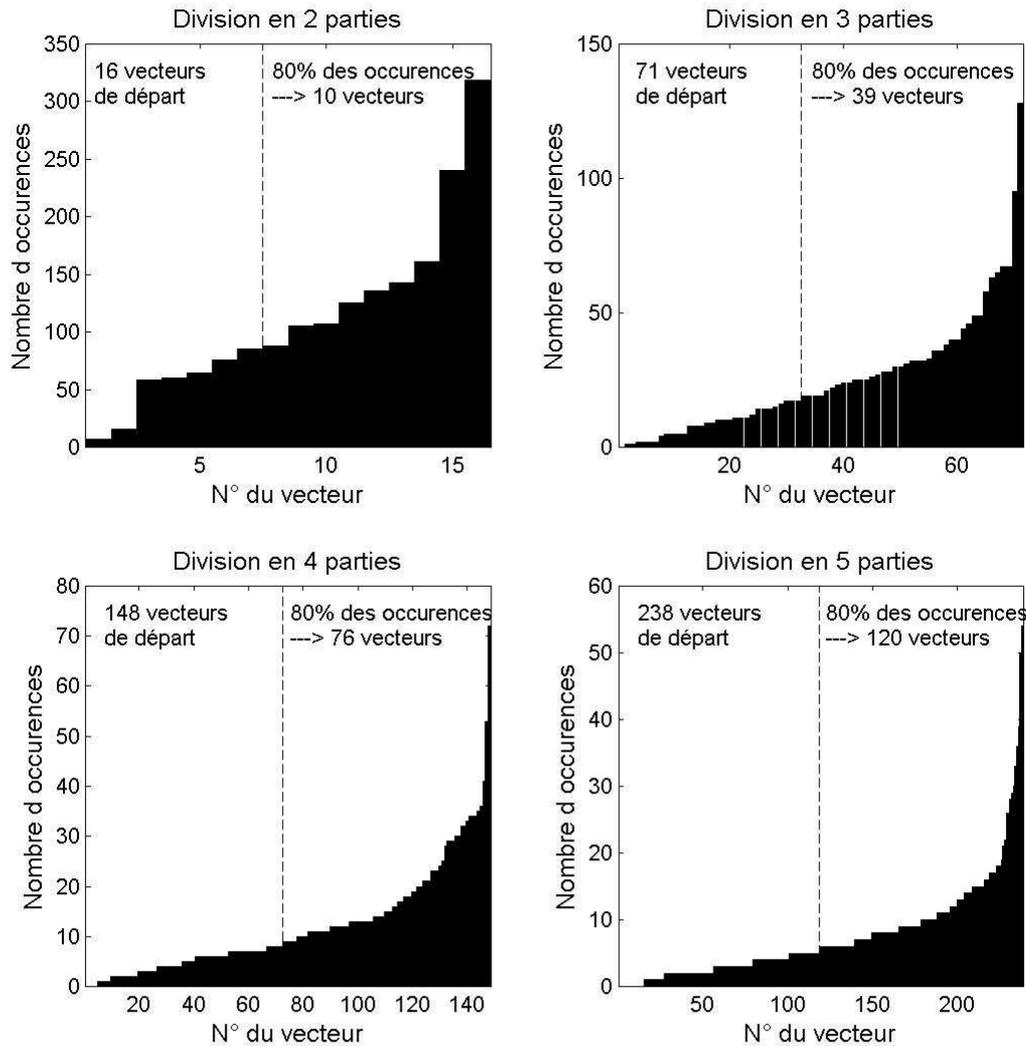


Figure 9-56 : Nombre d'occurrences où un vecteur est le meilleur des vecteurs de départ possibles. A droite de la ligne pointillée, les vecteurs restants représentent au moins 80% des meilleurs vecteurs de départ.

Comme le montre la Figure 9-57, il est encore possible de réduire le nombre de vecteurs possibles pour optimiser le modèle GR4H et obtenir encore un assez bon niveau de performance. Lorsqu'on utilise le groupe de vecteurs regroupant 80% des vecteurs optimisés (occurrences), les performances des modèles en validation ne sont pas réduites, alors que ce groupe de vecteurs représente 50% à 60% des vecteurs possibles, déterminés initialement.

Les résultats obtenus avec cette méthode de calage discret montrent qu'il est donc possible d'obtenir un assez bon niveau de performances avec le modèle GR4H, en le calant de manière très grossière. En effet, la différence de $C_{2M}(Q)$ moyen entre le modèle GR4H calé de manière continue (PAP-GR₁) ou de manière discrète (5 divisions) n'est que de 1.5 points, sur l'échantillon de 313 bassins versants. De plus, pour améliorer l'efficacité de la méthode PAP-GR_n, les échantillons de vecteurs

regroupant 80% des occurrences, pourront être utilisés comme points de départ possibles d'une optimisation continue et réduire ainsi le nombre de ces points de départ de plus de 50%.

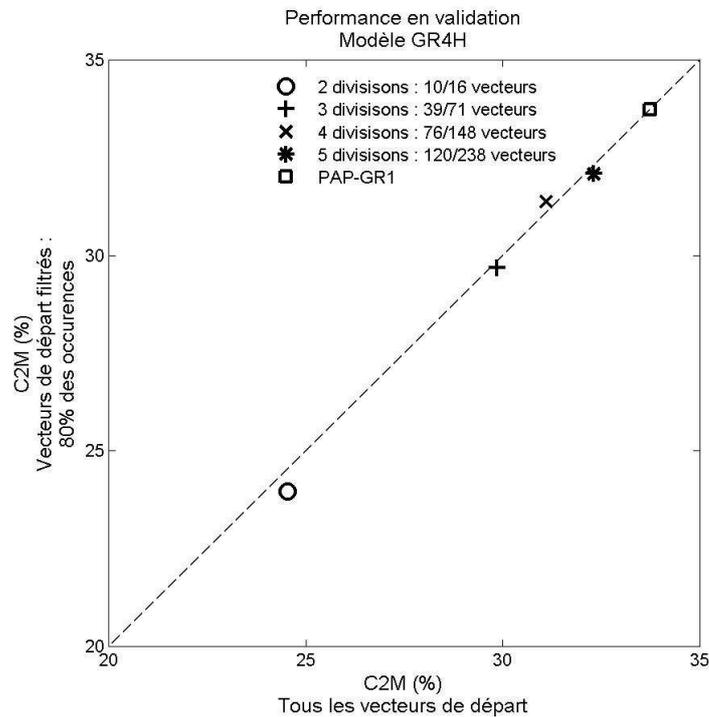


Figure 9-57 : Comparaison des performances en validation du modèle GR4H dont les paramètres ont été optimisés selon la méthode de calage discret, utilisant soit la totalité des vecteurs possibles, soit les vecteurs possibles regroupant a posteriori 80% des occurrences.

Enfin, cette méthode de calage discret permet d'espérer de bons résultats au cours d'une approche de régionalisation, puisqu'il est possible d'obtenir de très bons résultats avec seulement 39 à 120 bassins versants types. Il ne reste plus qu'à les identifier.

Annexe 10 Relations entre vecteurs de paramètres du modèle GR4X1 à différents pas de temps

9.11 Paramètres du modèle GR4X1, estimés au pas de temps horaire, avec pour objectif le débit journalier, et horaire

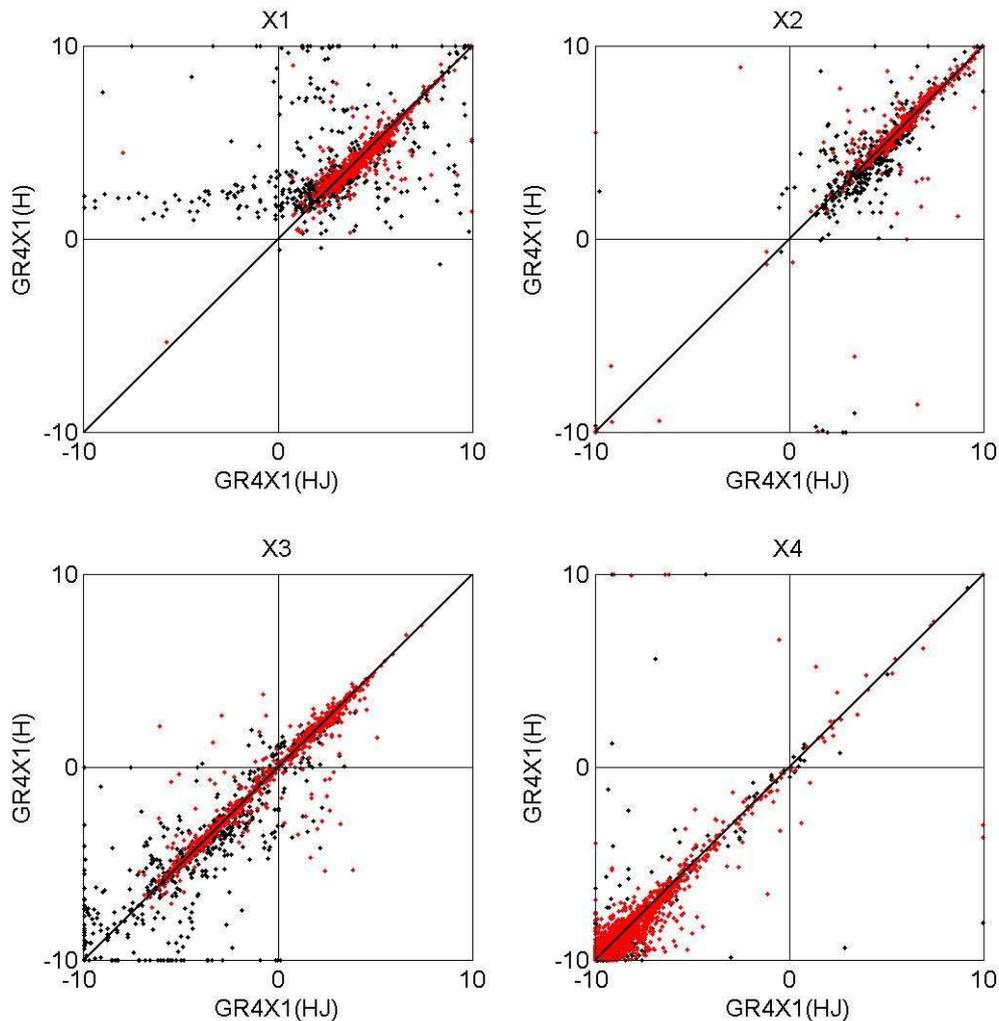


Figure 9-58 : Relations entre les paramètres transformés du modèle GR4X1, estimés au pas de temps horaire, avec pour objectif le débit journalier, et ceux estimés au pas de temps horaire. En rouge, bassins versants français ; en noir, bassins versants étrangers.

9.12 Paramètres du modèle GR4X1, estimés au pas de temps horaire avec pour objectif le débit journalier, et journalier

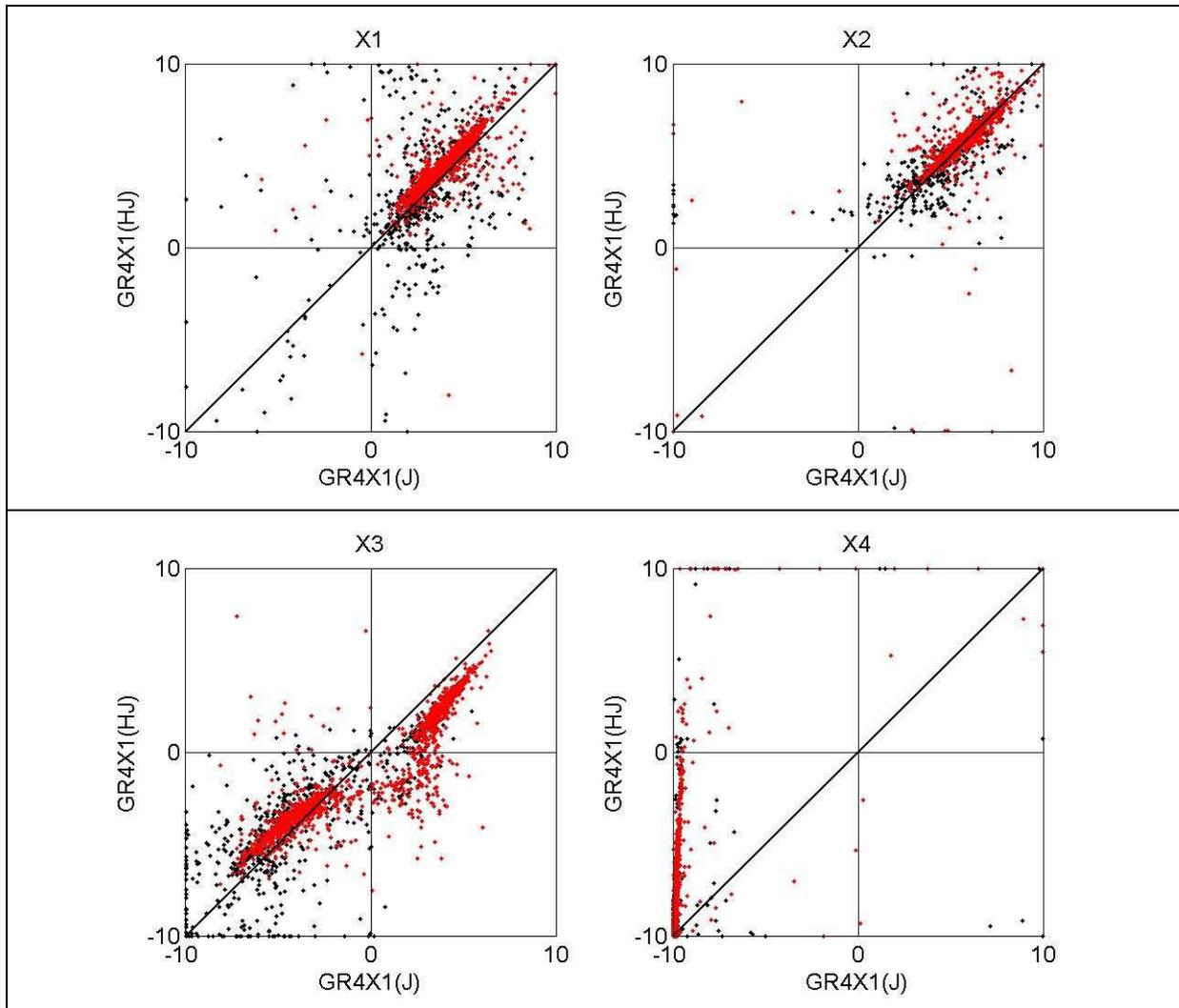


Figure 9-59 : Relations entre les paramètres transformés du modèle GR4X1, estimés au pas de temps horaire, avec pour objectif le débit journalier, et ceux estimés au pas de temps journalier. En rouge, bassins versants français ; en noir, bassins versants étrangers.

Annexe 11 Article soumis à Water Resources Research

Large samples of watersheds: a necessary condition to compare the efficiency of rainfall-runoff models?

Thibault Mathevet^{1,2}, Charles Perrin¹, Vazken Andréassian¹ and Claude Michel¹

1: Cemagref, Parc de Tourvoie, BP 44, 92163 Antony Cedex, France

2 : now at EDF-DTG, 21, Avenue de l'Europe - BP 41 - 38040 Grenoble Cedex 09,
France

E-mail address: thibault.mathevet@edf.fr

Running title: How many watersheds for rainfall-runoff models assessment?

Abstract

Rainfall-runoff models are useful tools in hydrological research, water engineering and environmental studies. Especially in practical applications, they must be as robust and reliable as possible. Many comparative studies have been made to assess the performance of rainfall-runoff models, specify their domains of application and provide guidance to end-users. However, most existing comparative studies have tested only a few models on a few watersheds, under various testing schemes. It is therefore difficult to generalize the results of these studies, as their conclusions are too dependent on the basin characteristics. The main objective of this paper is to demonstrate that comparative studies of rainfall-runoff models must be carried out on large samples of watersheds with appropriate assessment criteria to obtain conclusions with a general reach. To this end, an extensive comparison of the performance of five rainfall-runoff models was carried out on a large sample of 313 watersheds and on random sub-samples of different sizes. The performance was assessed with a bounded version of the Nash-Sutcliffe criterion, which gives less weight to large model failures in the evaluation of mean performances. Given the variability in model performances between watersheds, it is shown that samples need to include at least 50 to 100 watersheds to yield reliable comparisons of model efficiency. A statistical test shows that 100 to 300 watersheds are required to detect significant differences between mean model efficiencies.

Keywords: Rainfall-runoff modeling; large watershed sample; Nash-Sutcliffe criterion; model intercomparison; hourly time step.

Introduction

Thanks to the growing availability of hydro-meteorological data sets and the continuous improvement in computing power over the last decades, a large number of more or less complex rainfall-runoff (RR) models have been developed. It is difficult to say whether this large number corresponds to a blossoming of hydrological science or to the proliferation of similar tools. The need for intercomparison studies appeared quite early [WMO, 1975] to assess the relative performance of these RR models, with the goals to improve them and to eventually provide guidance to end-users. During the last three decades, the need for comparative studies has not faded away, as shown by the existing comparative studies [reviewed e.g. by *Michaud and Sorooshian*, 1994; *Refsgaard and Knudsen*, 1996; *Perrin et al.*, 2001] and by some recent initiatives to assess models [DMIP: *Smith et al.*, 2004; MOPEX: *Schaake and Duan*, in press].

A key issue when doing a comparative study of RR models is the number and the diversity of case studies necessary to effectively assess model performances and draw conclusions on model generality. In the early days of hydrological modelling, studies were limited by data availability and computing power. This time has long been over. More than twenty years ago, pioneers such as *Linsley* [1982] and *Klemeš* [1986] already recommended that models be tested on many catchments of widely differing characteristics, and that each trial cover a period of many years. However, let us recognize that very few modellers seem to have listened to these famous hydrologists, since most of the RR modelling studies reported in the literature present the performance of one model on a single watershed (or on a small number of similar watersheds).

In this paper, we will address the specific problem of assessing RR model performance on large samples of watersheds. Our objective is to show how too small samples of watersheds can seriously bias comparison results, and to draw recommendations on minimal sample sizes. We will also see that the comparison of models on large samples raises the issue of assessment criteria and that some classically used criteria are not well suited for such assessments.

This paper is organized as follows: in section 2, we review the hydrological literature about comparative assessments of RR models and the use of large watershed

samples. In section 3, we present the watershed sample used in this study, the tested models and the proposed comparative assessment methodology with a specific criterion to judge model performances. In section 4, we discuss our capacity to discriminate between five different models for different sizes of test watershed samples. Last, in section 5, we discuss the question of sample homogeneity and its impact on intercomparison results.

1. Limitations of assessing a single model on one or a few watersheds

1.1 What type of validation for RR models?

The question whether a RR model can or cannot be validated is still a matter of debate [*Konikow and Bredehoeft, 1992; de Marsily et al., 1992; Oreskes et al., 1994*]. Without entering the complex debate on the model validation and falsification [see *Refsgaard and Henriksen, 2004* for a recent review of the state of this debate], we can at least state that comparisons are very useful for an objective model assessment. Indeed, it is generally subjective to say that model A is good or bad in absolute terms, but much more objective to say that model A is better or worse than model B. Model B used here as a reference can be a baseline or a benchmark model (see on this point *Seibert [2001]*).

We would also think that the common practice of model assessment in RR modelling (i.e. using the split-sample test scheme [*Klemeš , 1986*] on one or a few watersheds) does not amount to a sufficient validation of a RR model, whatever its performance. When successful, this type of testing simply lets the modeller verify that the errors anticipated after calibration remain of the same order when the model is used on a new set of data.

Whether the proper validation of a RR model is possible or not, it does not alter the fact that, in our view, the two most important conditions for any validation of its structure is:

- a general overview of the applicability of a model over at least a reasonable range of hydro-meteorological conditions and characteristics of geology, soil, vegetation, topography and land-use;
- a testing scheme allowing to compare several alternative RR models (and to learn from the relative performances of these structures).

1.2 Why are large watershed samples needed to assess RR model efficiency?

Most comparative studies rely on a small number of watersheds (*Table 1*), but there are a few exceptions: e.g., the studies by *Vandewiele et al.* [1992], *Makhlouf and Michel* [1994], *Xu and Vandewiele* [1995], *Mouelhi et al.* [2006] at a monthly time step, *Perrin et al.* [2001], *Lee et al.* [2005] and those reported by *Boughton* [2005a] at a daily time step, and recently by *Mathevet* [2005] at an hourly time step. These authors were precursors in the use of large watershed samples to assess RR models and to ensure that they have a sufficient level of generality. On this topic, *Nash and Sutcliffe* [1970], *Moore and Mein* [1975] and *Klemeš* [1986] all stressed that generality should be a fundamental requirement for any RR model. Moreover, nowadays, some hydrological research groups are trying to gather large samples of watersheds to perform large scale intercomparison projects (DMIP, MOPEX). In this paper, we would like to demonstrate why large samples are of prime importance to reach an acceptable level of generality in model assessment.

It could be of course argued that not all models should aim at being general, and that given the extreme complexity of hydrological modelling, given the scarcity of process-related data, modellers should work on more modest goals (such as building a climate- or region-specific models). *Linsley* [1982] thought in contrast that the fundamental processes of hydrology being the same in all catchments, a model should represent them so that irrelevant processes can be “shut off” by optimization and therefore should be applicable in many different catchments. We believe that since some well-known RR models have been successfully applied on various types of watersheds, a model with a certain level of generality is sensible. Such examples are given by *Bergström* [1995], who reported the application of the HBV model on hundreds of watersheds in more than 35 countries (even if this model had originally been developed for Scandinavia); by *Ingram* [1996] who reported the applications of the SAC-SMA model at about 4000 forecast points in the United States; by *Perrin et al.* [2003] who based the improvement of a parsimonious RR model on a comparative study involving a worldwide set of 429 watersheds; and by the many successful applications of other models applied in countries and climates very different from their original development environment [e.g. *Liden and Harlin*, 2000].

If we are to work on the kind of general model advocated by *Linsley* [1982], it seems obvious that model assessment cannot rely only on a single or a few watersheds. Given natural watershed variability, a large number of watersheds is needed to judge the adaptability of a RR model to different conditions [*Moore and Mein*, 1975; *Linsley*, 1982; *Klemeš*, 1986; *Bergström*, 1991; *Andersson*, 1992; *Perrin et al.*, 2001; *Moreda et al.*, 2005]. Large watershed samples may have other advantages: data errors that may exist for a given watershed will have a weaker influence on the overall result of the comparison, than within a small watershed sample.

The main drawback of large samples of watersheds is the difficulty to thoroughly check the quality of rainfall and runoff data. In a large sample, data sets result from heterogeneous collection practices. So some specific features, not accounted for by classical RR models (such as karstic circulations, large groundwater contributions, human activities) and difficult to detect from the analysis of recorded flow series, may affect the quality of simulations. However, as argued by *Perrin et al.* [2001], there is presently no objective and widely accepted way for filtering all possible types of data errors. The selection or rejection of apparently erroneous watersheds should not be based on some modelling prejudice or on *a priori* judgment of hypothetical data quality, as it may potentially bias the comparison process. Anyway, *Linsley* [1982] rightly pointed out that in a model comparison framework, it is likely that all models will suffer equally from data errors. And *Boughton* [2005b] showed that RR models are able to cope with poor quality data to some extent.

2. Data and methods

2.1 Watershed sample test

A sample of 313 watersheds was used (see location in Figure 1) and main characteristics in Table 2). Most watersheds are in France (227) and in the USA (70), the others in Australia (12), Spain (2) and Slovenia (2). Climatic conditions vary from semi-arid to Mediterranean, oceanic, temperate, mountainous and continental. Snowmelt influence is generally limited. Mean annual rainfall varies between 200 and 3000 mm, mean annual potential evapotranspiration (PE) between 600 and 2100 mm and mean annual runoff between less than 1 and 1700 mm. Watershed area ranges between 0.5 km² to about 5000 km², with a median of 85 km². Data were collected at the hourly time step. The records are 8.5-year long on average. This database was

not used in any previous research work and is independent from recent comparative studies carried out at other time steps [e.g. *Perrin et al.*, 2001].

2.2 Rainfall-runoff models

Since our objective here was to show the usefulness of comparing RR model structures on a large sample of watersheds, five simple RR models were used (see Table 3). All models tested here are lumped and continuous. They were run at the hourly time step. These five structures were selected because the original models are quite widely applied and have different degrees of complexity (number of free parameters, number of stores). For the sake of this assessment, to put all models exactly in the same test conditions, some of the model structures were modified/simplified. As the purpose of this study is to discuss methodological aspects and not to compare original models, the models are called M1, M2, M3, M4, M5 hereafter, to avoid undue conclusions about the original models. All five RR models were fed with exactly the same input data, i.e. rainfall time-series and estimated PE, and their free parameters were calibrated using observed runoff.

Assessment methodology

Split-sample test:

To assess the performance of the selected RR models, a split-sample test procedure was used [*Klemeš*, 1986]. The models can thus be tested in simulation mode, under meteorological conditions different from those of the calibration period. For each watershed, the available time-series was split into several independent sub-periods. The models were successively calibrated on each sub-period and tested in validation mode on the remaining ones. The results are based on a total of 2093 validation tests for the 313 watersheds. The average length of a validation period is 2.2 years. The periods are quite short but this is not a problem here since the conditions were strictly the same for all the models.

Optimization technique:

The local optimization technique used here is presented by *Edijatno et al.* [1999]. The optimization procedure starts with an initial set of parameters, identified as the median value of each parameter over the large sample of watersheds. The algorithm uses a steepest descent method to move step by step in the parameter space, toward the optimum parameter set. Comparative tests, not reported here, with global search procedures indicate that the method is efficient to locate an optimum in the

parameter space that will yield good model performances in validation mode [Mathevet, 2005]. The mean square error of the model was used as objective function for parameter calibration.

Assessment criterion:

In the difficult exercise of RR model comparison, assessment criteria must be selected to evaluate and discriminate the performances of different models. As shown by *Weglarczyk* [1998] and by the large number of existing criteria, there is no ideal or unique criterion to judge the efficiency of models, i.e. there might be as many criteria as model end-users. In RR model evaluation, it is a common practice to use the *Nash and Sutcliffe* [1970] criterion, which defines model efficiency (*NS*) as the proportion of the initial variance accounted for by the model:

$$NS = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs}})^2} = 1 - \frac{F}{F_0} \quad \text{Eq. 1}$$

where, $Q_{obs,i}$ and $Q_{sim,i}$ are the observed and simulated discharges at time step i , and $\overline{Q_{obs}}$ is the mean observed discharges, the sum being taken over n time steps of the test period. F is the index of disagreement, analogous to the residual variance of a regression analysis. F_0 may be defined as the value of F for the naïve model ($Q_{sim,i} = \overline{Q_{obs}} \quad \forall i \in [1, n]$).

NS can take values between $-\infty$ and 1. A value of 1 indicates a perfect agreement, a value of zero indicates that the model does not explain any part of the initial variance and a negative value indicates that the model is worst than the naïve model used as a reference. The main advantage of the *NS* criterion is its dimensionless formulation. In spite of its common use in hydrology, the *NS* criterion has some drawbacks, as mentioned by *Garrick et al.* [1978] and *Martinec and Rango* [1989]. For example, these authors found it not demanding enough (it gives too high values) and too dependent on the climatic conditions of the watersheds (it is generally easier to have high performances for catchments with highly variable flows).

Another drawback originates from the formulation of the *NS* criterion that has no lower bound. For large model failures, the criterion values tend towards $-\infty$. This is a problem when one wishes to compute mean model efficiency over large watershed

samples. In such samples, there will always be watersheds where the model fails to simulate the hydrological behavior (see for example the results on the test sample in Table 4). So the distribution of the NS criterion over these watershed samples will be strongly skewed and the negative NS criteria will too heavily influence the calculation of the mean efficiency. The use of percentiles of the result distribution gives only a partial overview of the spread of model efficiencies. To obtain more meaningful statistics representing the mean model performance over large watershed samples and avoid these highly skewed distributions, we propose to adopt here a modified formulation of the NS criterion to make it vary between -1 and $+1$. This new formulation, called here C_{2M} [Mathevet, 2005; Mathevet et al., 2006], is a simple anamorphosis of the NS criterion:

$$C_{2M} = \frac{1 - \frac{F}{F_0}}{1 + \frac{F}{F_0}} \quad \text{Eq. 2}$$

The NS and C_{2M} criteria are related as follows:

$$NS = \frac{2 \cdot C_{2M}}{1 + C_{2M}} \quad \text{Eq. 3}$$

$$C_{2M} = \frac{NS}{2 - NS} \quad \text{Eq. 4}$$

This criterion behaves like the NS criterion: negative C_{2M} values correspond to negative NS values, and the 0 value in C_{2M} corresponds to the same level of model error as in the NS criterion. However the C_{2M} lower bound is -1 instead of $-\infty$. The relation between the NS and C_{2M} criteria is plotted in Figure 2. Note that the C_{2M} is less optimistic than the NS criterion for positive values. This partially answers the criticism made by Garrick et al. [1978] who found that the NS criterion produces too high values. Because the distribution of possible performances is now bounded, it should be possible to derive more meaningful statistics to summarize and compare model performances over large datasets. Giving a lower bound to the NS formulation, the C_{2M} puts less weight on large model failures and thus allows computing mean model efficiency without discarding some watersheds. Indeed, the C_{2M} allows not to subjectively discard watersheds whose NS criteria are lower than a limit, e.g. zero. We believe that this common practice to reject watersheds where model fails is highly detrimental to modeling for at least two reasons:

- first it allows the modeler to choose the sample of watersheds used to test models and therefore often gives a too optimistic picture of model capacities;
- second the criteria used to reject models are always subjective and often based on modeling prejudices (e.g. minimum value of an efficiency index) thus assuming that the model represents a true perception of reality.

Model comparison

To compare the efficiencies of m RR models on a large sample of n watersheds, one must in fact compare m vectors C_n of n criterion values. To see whether the difference between two models is significant, one can perform a Student test as proposed by Mouelhi *et al.* (in press), provided that the distribution of criterion values is approximately Gaussian. The statistic T of Student is defined as:

$$T_{1,2} = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \quad \text{Eq. 5}$$

where μ_1 and σ_1 , and μ_2 and σ_2 are the mean and standard deviation of C_n values for models M_1 and M_2 respectively, over the watershed sample. The difference in the average values of M_1 and M_2 efficiency is considered significant at a 5% level, if $|T_{1,2}|$ is greater than 1.96.

3. Comparative assessment of RR model efficiencies on samples of watersheds

In this section, we wish to demonstrate that it is necessary to use large samples of watersheds to effectively compare RR models. For the sake of simplicity, we compare only the efficiencies of model M2 to models M1, M3, M4 and M5. These models were tested on the whole sample of 313 watersheds and on randomly generated sub-samples as explained below. In the following, only the bounded version of the NS criterion (C_{2M}) was used to present results. Table 4 summarizes the performances of the different models on the whole sample, where the difference of C_{2M} efficiency between M2 and M1, M3, M4 and M5 varies from -0.01 to $+0.1$ (we left NS results in this table to provide a more common reference of model efficiency).

3.1 Variability of models efficiencies for a given watershed

In this section, we will show that the relative efficiencies of different RR models can be highly variable from one catchment to another. Figure compares model efficiencies on the sample of 313 watersheds. It shows a wide spread of efficiencies, which means that the difference between two models can vary much between watersheds. Figure 3 also shows that, even if M2 is better than M3, M4 or M5 on average on the whole sample, there are 15 to 33% of the watersheds where the other model is better. This means that, using a single watershed to compare the efficiency of different models may lead to undue conclusions on their respective values in general, because of the characteristics of the selected watershed. It is therefore impossible to draw conclusions on model generality by comparison on a single watershed.

Figure 4 shows the worst and best efficiencies obtained by one of the M1 to M5 models, as a function of the mean efficiency of the five models obtained on each watershed. It shows a wide spread of the minimum and maximum efficiencies around the mean efficiency. Between the best and the worst models, the mean difference over the whole watershed sample is very high, around +/- 0.2 for the C_{2M} criterion. This means that the ability of a model to represent the transformation of rainfall into runoff may vary much between watersheds.

Figure 5 shows the relation between the mean and the standard deviation of the efficiencies obtained by the five models. It shows that the variability of models efficiency increases when mean models efficiency decreases. For example, the variability of the efficiencies of the five models is low on watersheds where models can be considered as well performing (for $C_{2M} > 0.65$, i.e. for $NS > 0.8$). For watersheds where mean models efficiencies can be considered as acceptable (for $0.35 < C_{2M} < 0.65$, i.e. $0.5 < NS < 0.8$), the variability of models efficiencies is high and the mean standard deviation of models efficiencies represent 0.1 for the C_{2M} criterion. Last, for the watersheds where models can be considered as inefficient on average (for $C_{2M} < 0.35$, i.e. $NS < 0.5$), it is often still possible to find a model with good efficiency. These results show that:

- Differences between models are limited on the watersheds where the models are good on average. The discrimination of models is therefore difficult on these watersheds;

- the watersheds where the models are acceptable are very interesting to discriminate models, because the variability of their efficiencies is high;
- the watersheds where the models are not very efficient must not be discarded from the test sample because it is often possible to find an efficient model.

All these results show that the conclusions of a model comparison on a single watershed are limited to the context of this watershed and cannot be generalized. It is therefore advised to test models on more than one watershed if one wishes conclusions of general reach. They also show that using a lot of different watersheds will multiply the occasions to evaluate the differences between models.

3.2 Impact of the watershed sample size on the comparative assessment of model efficiency

In this section, our aim is to show how the respective value of the tested models may vary when their efficiencies are computed on a small sample of watersheds. To demonstrate the limits of using small sample of watersheds to compare models, we generated 100 random sub-samples of 5, 20 and 50 watersheds from the initial sample of 313 watersheds. To get sufficiently different sub-samples, their size was limited to 50 watersheds and the number of sub-samples was limited to 100. Figure 6 compares the mean efficiencies of model M2 to those of models M1, M3, M4 and M5, for the randomly drawn sub-samples. The mean efficiency on the whole sample is also shown by a cross.

This figure underlines the dispersion of mean model efficiencies of the different sub-samples around the mean efficiency obtained on the whole sample. This dispersion is all the more important that the sample size is reduced. If one considers that the mean performance over the whole watershed sample is the more realistic picture of the relative value of models, it can be seen that the use of small sub-samples may lead to false conclusions in the model comparison. Indeed, in all cases, one can find sub-samples where a model is better than another whereas the opposite is true on the whole sample. When the difference of C_{2M} efficiency between two models over the whole sample is about 0.05 to 0.1 (the case of M2 vs M4 and M2 vs M5), a sample of about 20 to 50 watersheds seems sufficient to differentiate between these models, i.e. to avoid these undue conclusions. However, when the difference is smaller, the size of the required watershed sample increases, but there are less cases of obvious misinterpretation.

To summarize, we believe that Figure 3 to Figure 6 clearly demonstrate that given the variability of model efficiency on small samples of watersheds, a large sample size (at least 50 watersheds) is clearly required to draw statistically significant conclusions on model respective value. The smaller the difference between model efficiencies, the larger the sample required to establish this difference.

3.3 How many watersheds are required in a sample test to make the difference significant between model efficiencies?

In this section, we use the statistical test presented in section 2.3 to detect whether the difference of mean efficiency between two models is significant. This test is complementary to the simple comparison of mean model efficiencies, because it takes into account the standard deviation of efficiencies and the size of the watershed sample. This Student test assumes a Gaussian distribution for the criterion, which is approximately true for the C_{2M} criterion, thanks to its bounded formulation.

As mentioned before, over the whole sample, the difference of C_{2M} efficiency between M2 and M1, M3, M4 and M5 varies from -0.01 to $+0.1$ (Table 4). On this sample, the statistical test (Table 5) rejects the null hypotheses and concludes that the mean efficiency of model M5 is significantly lower than the mean efficiency of M2. However, this test concludes that the difference of efficiency between M2 and M1, M3 and M4 is not significant. This means that, given the variability of model efficiency, the difference between models needs to be large (0.1) to be statistically significant.

To illustrate these results, we applied the T test on the random sub-samples of 5, 20 and 50 watersheds (

Table 6). Results indicate that:

between M2 and M1, M3 and M4, the mean C_{2M} efficiency is not significantly different, whatever the sample size, because their difference of mean efficiency is too small and the standard deviation of the mean efficiency remains limited.

between M2 and M5, the probability to conclude that their difference of efficiency is significant varies from 3 to 50%, when the sample size varies from 5 to 50 watersheds.

These results show that, given the variability of model efficiency in a random sample of watersheds, it is important to use a large sample to robustly compare models.

Even with a sample of 5 to 20 watersheds, the likelihood of the hypothesis “M2 and M5 have an equivalent mean efficiency” is greater than 0.94, whereas the contrary is true on a larger sample!

From a theoretical point of view, if it is assumed in Eq. (5) that the standard deviation of the model efficiencies does not depend on the sample size and is equal to the standard deviation of the efficiency on the 313 watersheds sample, it is possible to compute a Student test for different watershed sample sizes. Given this assumption, a Student test was performed for the five models successively compared to each other, for different watershed sample sizes. Table 7 shows the minimum sample sizes required to conclude that a model is significantly different from another. With the 313 watersheds sample, when the difference of mean efficiency between two models is less than 0.04, there is a 0.95 probability to conclude that the mean model efficiencies are not significantly different; when the difference between two models is around 0.04 to 0.07, 150 to 310 watersheds are required to indicate a significant difference in efficiency; when the difference of efficiency between two models is greater than 0.1, 75 watersheds are required.

By the same token, it is possible to roughly give the minimum sample size required to differentiate between the mean efficiencies of two models. Table 8 shows that, for a given difference of efficiency between two models (from 0.025 to 0.1), the minimum sample size required is extremely sensitive to the standard deviation of model efficiency. For example, if the difference of efficiency between two models is about 0.1, the minimum sample size could vary from 50 to 100 watersheds, depending on the model efficiency standard deviation (the range of model efficiency standard deviation corresponds to the one obtained on this sample and by *Perrin et al.* [2001] or *Oudin et al.* [2004] on other samples). From a practical point of view, Table 7 and Table 8 show that when the difference of efficiency between two RR models is small (from 0.05 to 0.1), it is usually not possible to conclude with confidence that a model is significantly better than another when the sample has less than 10 watersheds, as it is usually done in the hydrological literature.

4. Comparison of RR model efficiencies on homogeneous random samples

It may be argued that the conclusion of a comparison does not only depend on the size of the watershed sample but also on the variability of watershed characteristics. For example, to effectively test the generality of models, one may prefer to have a smaller sample of very different watersheds (in terms of characteristics and climatic conditions) rather than a larger sample with very homogeneous conditions. In this section, we used the previous methodology, but the random samples of watersheds were drawn in more homogeneous samples of watersheds. Simple descriptors of watershed characteristics were chosen to quantify the heterogeneity: mean annual streamflow, mean annual rainfall, mean annual potential evapotranspiration (PE) and watershed area. For each descriptor, the whole watershed sample was divided into three sub-samples, of low, medium and high value using the 0.33 and 0.66 percentiles of the distributions of each descriptor over the whole watershed sample (Table 9). For each descriptor, it yielded three classes of watersheds having low, medium and high characteristics values. In each sub-sample, 100 random samples of 20 watersheds were randomly drawn. This comparison was made to detect if a model could be typically more efficient on one type of watersheds, characterized by a given descriptor.

Figure 6 shows the comparison of model efficiencies of two models on random sub-samples drawn from samples of watersheds characterized by low, medium and high value for mean annual streamflow (a1 to a4), mean annual rainfall (b1 to b4), mean annual PE (c1 to c4) and watershed area (d1 to d4). These figures exhibit the same shape as Figure , which means that the clustering of watershed using their characteristics does not affect the comparison of mean model efficiencies. The only patterns that might appear are:

- the same level and low variability of efficiency of M1 and M2 on watersheds characterized by high annual streamflow (Figure a1),
- the better efficiency of M2 compared to M3, on watersheds characterized by a high annual streamflow (Figure b1),

These figures also show that, contrary to a common belief, it seems that none of the chosen descriptors strongly influence the comparison of model performances. From a practical point of view, these figures show that the correspondence between models and types of descriptors is not straightforward and that models often developed in specific climatic contexts may also provide good results under other contexts. This means that some models are sufficiently general and robust to be

more efficient than the others, on a wide span of watershed characteristics. This result is in agreement with the conclusion of *Lee et al.* [2005] who could not find links between models and types of watersheds.

Conclusion

The aim of this paper was to present a methodological framework for the comparison of rainfall-runoff models. The discussion raised two issues: (i) the need to use large watershed samples to compare model efficiency and (ii) the lack of significance of model assessment without some comparative reference to other models.

To significantly compare model results on large watershed samples, we used a bounded formulation (called C_{2M}) of the well-known Nash-Sutcliffe efficiency index, which avoids large negative values difficult to interpret. The C_{2M} criterion makes it possible to compute meaningful mean model efficiencies on large test samples.

We show that large watershed samples (at least 50 watersheds) yield more reliable conclusions on the comparison of model efficiency, i.e. they reduce the risks of drawing wrong conclusions concerning the respective value of different tested models. From a practical point of view, we show through the implementation of a demanding statistical test that, given the usual variability of model efficiency, between 100 and 300 watersheds are often required to show significant differences between the mean efficiency of two models. We also show that the type of watersheds used for the comparison does not influence much the results of the model comparison. So the size of the watershed sample remains a key point to effectively compare models.

Acknowledgements

The authors acknowledge associate editor and the three anonymous reviewers for their suggestions, which help to improve this paper. The authors would like to thank the institutions and researchers who provided data sets for model testing: Christian Scherer of the French Ministry for Ecology and Sustainable Development, for the hydrological data sets of the French watersheds (from the HYDRO database); Bruno Rambaldelli of Météo France, for the meteorological data sets of the French watersheds (from the PLUVIO database); Dr Christian Zammit at the Water and River Commission of Western Australia, and Dr Muguresu Sivapalan at the Center for

Water Research of the university of Western Australia, for the data sets of the Collie River watershed; Mira Kobold, at the Environmental Agency of the Republic of Slovenia, for the data sets of the Savinja River watershed; Dr. Fransesc Gallart and Dr. Jérôme Latron from the Jaume Almera Institute of Earth Sciences (CSIC, Spain) for the data sets of the Can Vila et Cal Rodo watersheds; Robert Zeimer for the data sets of the Caspar Creek watersheds (CA, USA); the Water Data Center of the Agricultural Research Service (ARS) for the data sets of the watersheds from the USA. The authors also thank Emmanuel Paquet and Joël Gailhard from EDF-DTG (Grenoble, France) for providing MORDOR model algorithms and for data sets. But special thanks are due to all anonymous individuals involved in field data collection. Such work would not have been possible without them.

References

- Andersson, L., Improvement of runoff models. What way to go?, *Nordic hydrology*, 23, 315-332, 1992.
- Andréassian, V., C. Perrin, C. Michel, I. Usart-Sanchez, and J. Lavabre, Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, *Journal of Hydrology*, 250, 206-223, 2001.
- Bergström, S., The HBV model, in *Computer Models of Watershed Hydrology*, Chapter 13, edited by V.P. Singh, pp. 443-476, Water Resources Publications, 1995.
- Bergström, S., and A. Forsman, Development of a conceptual deterministic rainfall-runoff model, *Nordic Hydrology*, 4, 147-170, 1973.
- Bergström, S., Principles and confidence in hydrological modelling. *Nordic Hydrology*, 22, 123-136, 1991.
- Beven, K., R. Lamb, P. Quinn, R. Romanovicz, and J. Freer, TOPMODEL, in *Computer Models of Watershed Hydrology*, Chapter 18, edited by V.P. Singh, pp. 627-668, Water Resources Publications, 1995.
- Beven, K.J., and M.J. Kirkby, A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24 (1), 43-69, 1979.
- Boughton, W., Catchment water balance modelling in Australia 1960-2004, *Agricultural water management*, 71, 91-116, 2005a.
- Boughton, W., Calibrations of a daily rainfall-runoff model with poor quality data. *Environmental Modelling & Software*, 2005b.

- Chiew, F.H.S., M.J. Stewardson, and T.A. McMahon, Comparison of six rainfall-runoff modelling approaches, *Journal of Hydrology*, 147, 1-36, 1993.
- Edijatno, N.O. Nascimento, X. Yang, Z. Makhlof, and C. Michel, GR3J: a daily watershed model with three free parameters, *Hydrological Sciences Journal*, 44 (2), 263-277, 1999.
- Franchini, M., and M. Pacciani, Comparative analysis of several conceptual rainfall-runoff models, *Journal of Hydrology*, 122, 161-219, 1991.
- Gan, T.Y., E.M. Dlamini, and G.F. Biftu, Effects of model complexity and structure, data quality and objective function on hydrologic modeling, *Journal of Hydrology*, 192, 81-103, 1997.
- Garçon, R., Modèle global Pluie-Débit pour la prévision et la prédétermination des crues (Lumped rainfall-runoff model for flood forecasting and design flood estimation, in french), *La Houille Blanche*, 7/8, 88-95, 1999.
- Garrick, M., C. Cunnane, and J.E. Nash, A criterion of efficiency for rainfall-runoff models., *Journal of Hydrology*, 38, 375-381., 1978.
- Jakeman, A.J., I.G. Littlewood, and P.G. Whitehead, Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, *Journal of Hydrology*, 117, 275-300, 1990.
- Klemeš, V., Conceptualisation and scale in hydrology, *Journal of Hydrology*, 65, 1-23, 1983.
- Klemeš, V., Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, 31 (1), 13-24, 1986.
- Kokkonen, T.S., and A.J. Jakeman, A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications, *Water Resources Research*, 37 (9), 2345-2352, 2001.
- Lidén R. and J. Harlin, Analysis of conceptual rainfall-runoff modelling performance in different climate. *Journal of hydrology*, 238, 231-247, 2000.
- Lee, H., McIntyre, N., Wheeler, H. and Young, A., Selection of conceptual models for regionalisation of the rainfall-runoff relationship. *Journal of Hydrology*, 312, 125-147, 2005.
- Linsley, R.K., Rainfall-runoff models - an overview., in *Proceedings of the international symposium on rainfall-runoff modelling.*, edited by V.P. Singh, pp. 3-22, Water Resources Publications, Littleton, CO, USA, 1982.

- Makhlouf, Z., and C. Michel, A two-parameter monthly water balance model for French watersheds, *Journal of Hydrology*, 162 (3-4), 299-318, 1994.
- Martinec, J., and A. Rango, Merits of statistical criteria for the performance of hydrological models, *Water Resources Bulletin*, 25 (2), 421-432, 1989.
- Mathevet, T., Which rainfall-runoff model at the hourly time-step ? Empirical development and intercomparison of rainfall-runoff models on a large sample of watersheds, PhD thesis (in french), ENGREF University, Paris, 2005.
- Mathevet, T., C. Michel, C. Perrin, V. Andréassian, A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sample of watersheds, *Red Book*, XXXX.
- Michaud, J., and S. Sorooshian, Comparison of simple versus complex distributed runoff models on a mid-sized semiarid watershed, *Water Resources Research*, 30 (3), 593-605, 1994.
- Moore, I.D., and R.G. Mein, An evaluation of three rainfall-runoff models, in *Hydrological Symposium*, edited by I.E. Aust., pp. 122-126, Sydney, 1975.
- Moreda, F., V. Koren, Z. Zhang, S. Reed, M. Smith, Parameterization of distributed hydrological models: learning from the experiences of lumped modeling, *Journal of Hydrology*, in press.
- Mouelhi, S., Michel, C., Perrin, C. and Andréassian, V. (2006). Stepwise development of a two-parameter monthly water balance model. *Journal of Hydrology*, 318(1-4), doi:10.1016/j.jhydrol.2005.06.014.
- Mouelhi, S., Michel, C., Perrin, C. and Andréassian, V., Linking stream flow to rainfall at the annual time step: the Manabe bucket model revisited. *Journal of Hydrology*, in press.
- Moussavi, M., and J. Feyen, Comparison of different structures for a monthly water yield model in seasonally snow-covered mountainous watersheds of Iran, *Hydrological Sciences Journal*, 35 (5), 535-546, 1990.
- Nash, J.E., and J.V. Sutcliffe, River flow forecasting through conceptual models. Part I - A discussion of principles., *Journal of Hydrology*, 27 (3), 282-290., 1970.
- Oudin, L., C. Michel, F. Anctil, Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?, *Journal of Hydrology*, 303 (1-4), 275-289.

- Perrin, C., C. Michel, and V. Andréassian, Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *Journal of Hydrology*, 242 (3-4), 275-301, 2001.
- Perrin, C., C. Michel and V. Andréassian, Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, 279 (1-4), 275–289, 2003.
- Refsgaard, J.C., and H.J. Henriksen, Modelling guidelines - terminology and guiding principles, *Advances in Water Resources*, 27, 71-82, 2004.
- Refsgaard, J.C., and J. Knudsen, Operational validation and intercomparison of different types of hydrological models, *Water Resources Research*, 32 (7), 2189-2202, 1996.
- Schaake, J., Q. Duan, The model parameter experiment project (MOPEX), *Journal of Hydrology*, in press.
- Seibert, J., On the need for benchmarks in hydrological modelling, *Hydrological Processes*, 15 (6), 1063-1064, 2001.
- Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy, Downward approach to hydrological prediction, *Hydrological Processes*, 17 (11), 2101-2111, 2003.
- Smith, M., K. Georgakakos, X. Liang, The distributed model intercomparison project (DMIP), *Journal of Hydrology*, 298 (1-4), 1-3, 2004.
- Vandewiele, G.L., C.Y. Xu, and N.L. Win, Methodology and comparative study of monthly models in Belgium, China and Burma, *Journal of Hydrology*, 134, 315-347, 1992.
- Warmerdam, P.M.M., Kole, J. et Chormanski, J., Modelling rainfall-runoff processes in the Hupselse Beek research basin. In: D. Viville and I.G. Littlewood (Editor), *Ecohydrological processes in small basins*. IHP-V, Technical Documents in Hydrology n°14, UNESCO, Paris, Proceedings of the Strasbourg Conference (24-26 September 1996), pp. 155-160, 1997.
- Weeks, W.D., and R.H.B. Hebbert, A comparison of rainfall-runoff models, *Nordic Hydrology*, 11, 7-24, 1980.
- Weglarczyk, S., The interdependence and applicability of some statistical quality measures for hydrological models, *Journal of Hydrology*, 206, 98-103, 1998.
- Wilcox, B.P., W.J. Rawls, D.L. Brakensiek, and J.R. Wight, Predicting runoff from rangeland catchments: a comparison of two models, *Water Resources Research*, 26 (10), 2401-2410, 1990.

- WMO, Intercomparison of conceptual models used in operational hydrological forecasting, World Meteorological Organization, Geneva, Switzerland, 1975.
- Xu, C.Y., and G.L. Vandewiele, Parsimonious monthly rainfall-runoff models for humid basins with different input requirements, *Advances in Water Resources*, 18, 39-48, 1995.
- Yang, D., S. Herath, and K. Musiak, Comparison of different distributed hydrological models for characterization of catchment spatial variability, *Hydrological Processes*, 14 (3), 403-416, 2000.
- Ye, W., B.C. Bates, N.R. Viney, M. Sivapalan, and A.J. Jakeman, Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resources Research*, 33 (1), 153-166, 1997.
- Zhao, R.J., Zuang, Y.L., Fang, L.R., Liu, X.R. et Zhang, Q.S, The Xinanjiang model. IAHS Publication n° 129, 351-356, 1980.
- Zhao, R.J. et Liu, X.R., The Xinanjiang model. In: V.P. Singh (Editor), *Computer models of Watershed Hydrology*, Chapter 7. Water Resources Publications, pp. 215-232, 1995.
- Zhang, X., and G. Lindström, A comparative study of a Swedish and a Chinese hydrological model, *Water Resources Bulletin*, 32 (5), 985-994, 1996.

Study	Number of tested model	Type(s) of model(s)	Number of test watersheds
<i>Moore and Mein</i> [1975]	3	CRR-L	4
<i>WMO</i> [1975]	10	CRR-L, BB-L	2 to 6
<i>Weeks and Hebbert</i> [1980]	5	CRR-L, BB-L	3
<i>Moussavi and Feyen</i> [1990]	6	CRR-L	1
<i>Wilcox et al.</i> [1990]	2	CRR-L	6
<i>Franchini and Pacciani</i> [1991]	7	CRR-L	4
<i>Vandewiele et al.</i> [1992]	2 to 8	CRR-L	1 to 79
<i>Chiew et al.</i> [1993]	6	BB-L, CRR-L	8
<i>Makhlouf and Michel</i> [1994]	5	CRR-L	91
<i>Xu and Vandewiele</i> [1995]	2	CRR-L	91
<i>Refsgaard and Knudsen</i> [1996]	3	CRR-L, PB-D, CRR-D	3
<i>Zhang and Lindström</i> [1996]	2	CRR-L	2
<i>Ye et al.</i> [1997]	3	CRR-D, PB-D	3
<i>Gan et al.</i> [1997]	4 to 5	CRR-L	3
<i>Yang et al.</i> [2000]	3	PB-D, CRR-D, GB-D	1
<i>Kokkonen and Jakeman</i> [2001]	2	CRR-L	2
<i>Perrin et al.</i> [2001]	19	CRR-L	429
<i>Lee et al.</i> [2005]	12	CRR-L	28
<i>Mouelhi et al.</i> [2006]	9	CRR-L	410

Table 1: Comparative studies of RR models, number and type of tested models, size of the watershed sample. Types of model: (CRR) conceptual rainfall-runoff model, (PB) physically-based model, (BB) black box model, (GB) geomorphology-based model, in a (L) lumped mode, (SD) semi-distributed mode, (D) distributed mode.

Country	France	USA	Australia	Spain	Slovenia
Number of watersheds	227	70	12	2	2
Annual runoff [mm]	35 - 44 - 1655	0 - 279 - 1612	9 - 35 - 96	429 - 736	1024 - 1182
Annual rainfall [mm]	403 - 963 - 2067	193 - 1163 - 2996	569 - 674 - 1025	1183	1384
Annual PE [mm]	595 - 791 - 1252	1104 - 1545 - 2085	1226	639	735
Annual runoff - rainfall ratio(%)	0.05 - 0.44 - 2.59	0 - 0.20 - 0.81	0.01 - 0.05 - 0.13	0.36 - 0.62	0.74 - 0.85
Type of climate	Temperate, mediterranean, oceanic, continental	Temperate, oceanic, semi-arid	Semi-arid	Mountainous mediterranean	Mountainous
Watershed area [km ²]	1.1 - 280 - 4978	1.2 - 33 - 334	2.7 - 48 - 2538	0.56 - 4.17	457 - 1385
Length of the time serie [year]	3 - 8 - 34	3 - 11 - 43	6	3 - 4	5

Table 2: Minimum - mean - maximum characteristics of the 313 watersheds.

Tested model	Number of free parameters	Number of reservoirs	Original model	Reference of original model
XINANJ	8	4	XINANJIANG	<i>Zhao et al. [1980; 1995]</i>
WAGE	8	3	WAGENINGEN	<i>Warmerdam, [1997]</i>
MORDOR6	6	4	MORDOR	<i>Garçon [1999]</i>
IHAC	6	3	IHACRES	<i>Jakeman et al. [1990]</i>
GR4H	4	2	GR4J	<i>Perrin et al. [2003]</i>

Table 3: Characteristics of the five RR models used in this study. The details of the modified versions tested here can be found in *Mathevet [2005]*.

	M1	M2	M3	M4	M5
Minimum NS	- 362.9	- 815.2	- 648.5	- 420.5	- 1778.2
Mean NS	58.8	55.7	52.9	54.8	38.5
Maximum NS	89.8	89.5	87.0	82.6	85.4
NS Percentile					
0.1	16.4	10.3	0	12.6	- 11.2
0.5	71.45	71.5	68.6	65.4	62.3
0.9	87.5	87.0	85.0	83.5	82.6
NS standard deviation	48.2	69.4	60.4	41.5	120.9
% of NS					
< -100 %	2	1.5	2.2	1	3.5
< 0 %	6	6	10.5	5.5	11.5
Mean C_{2M}	49.9	48.9	45.5	43.7	38.7
C_{2M} Percentile					
0.1	8.9	5.4	0	6.7	- 5.0
0.5	55.6	55.6	52.2	48.6	45.2
0.9	77.8	77.0	74.0	71.7	70.3
C_{2M} standard deviation	27.8	28.0	28.2	24.1	31.2

Table 4: Characteristics of model efficiencies over the 313 watershed sample using the NS and C_{2M} criteria.

Number of watersheds	313
M2 / M1	$\Delta C_{2M} = - 0.01$ H ₀ rejected
M2 / M3	$\Delta C_{2M} = 0.035$ H ₀ rejected
M2 / M4	$\Delta C_{2M} = 0.05$ H ₀ rejected
M2 / M5	$\Delta C_{2M} = 0.1$ H ₀ accepted

Table 5: Comparison of mean model efficiency (H₀ hypothesis: mean C_{2M} efficiencies are not different from one model to another), following a Student test, at 5%.

	Number of watersheds		
Models compared	5	20	50
M2 / M1	0.	0	0
M2 / M3	< 1	0	0
M2 / M4	2	1	3.5
M2 / M5	3	6	50

Table 6: Percentage of random sub-samples where the H_0 hypothesis is rejected (H_0 hypothesis: mean C_{2M} efficiencies are not different from one model to another), following a Student test at 5% (over 500 random samples).

	M1	M2	M3	M4	M5
M1	-	∞ $\Delta C_{2M} = -0.01$	310 $\Delta C_{2M} = 0.04$	150 $\Delta C_{2M} = 0.06$	75 $\Delta C_{2M} = 0.11$
M2	-	-	∞ $\Delta C_{2M} =$ 0.035	200 $\Delta C_{2M} = 0.05$	75 $\Delta C_{2M} = 0.1$
M3	-	-	-	∞ $\Delta C_{2M} = 0.02$	150 $\Delta C_{2M} = 0.07$
M4	-	-	-	-	250 $\Delta C_{2M} = 0.05$
M5	-	-	-	-	-

Table 7: Minimum sample size required to differentiate between the mean efficiencies of two models, based on a Student test and a threshold at 5%.

Difference of mean efficiency between two models ($\Delta\mu$)	Model efficiency standard deviation (σ)			
	0.20	0.25	0.30	0.35
0.025	∞	∞	∞	∞
0.050	125	200	300	∞
0.075	70	100	125	175
0.100	35	50	70	100

Table 8: Theoretical minimum sample size required to differentiate between the mean efficiencies of two models, based on a Student test and a threshold at 5%.

	0.33 percentile	0.66 percentile
Mean annual streamflow [mm]	220	440
Mean annual rainfall [mm]	780	1050
Mean annual PE [mm]	720	1100
Watershed area [km²]	35	140

Table 9: Value of the different descriptors to generate low, medium and high class samples

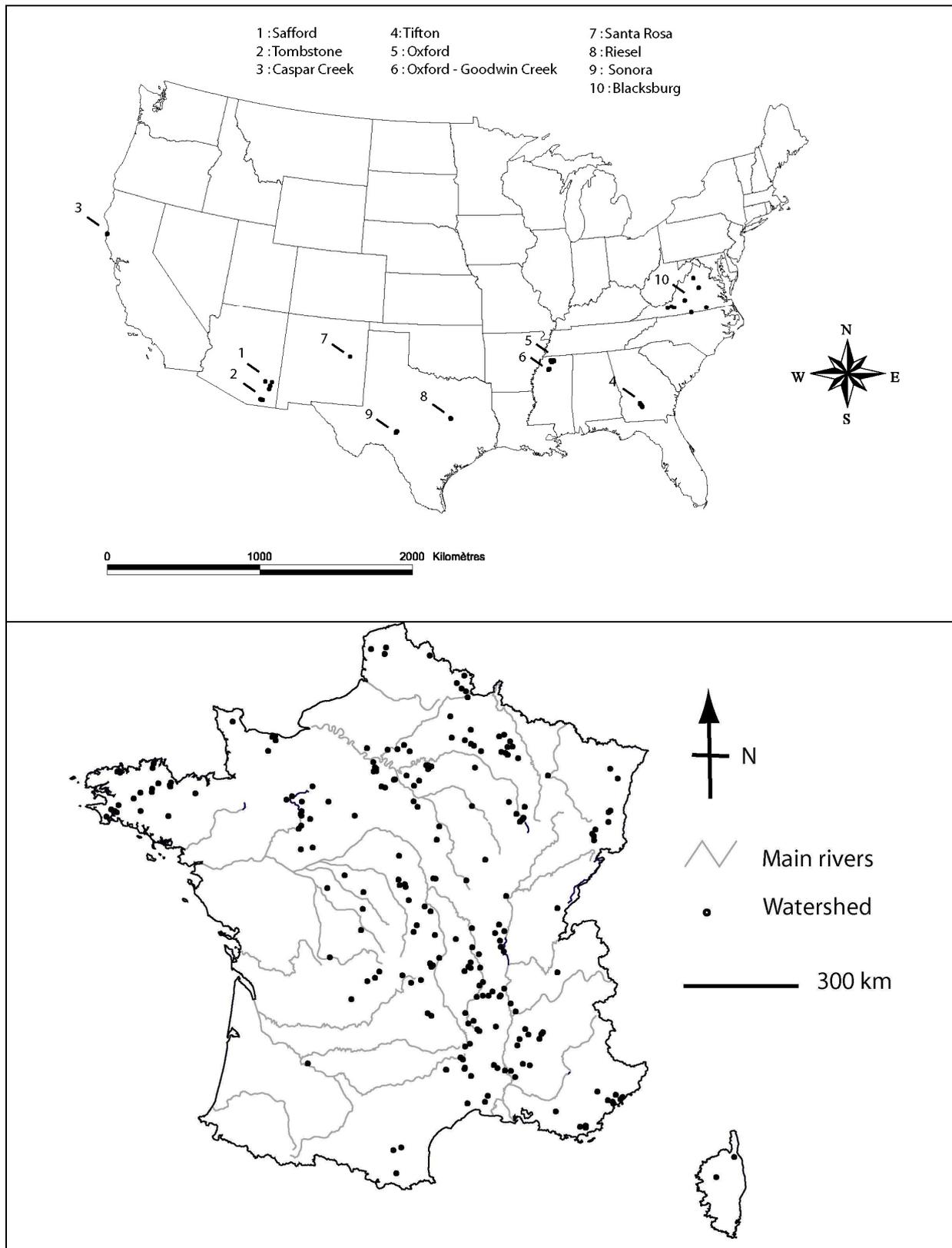


Figure 1 : Location of the watersheds in France and in the USA used in this study

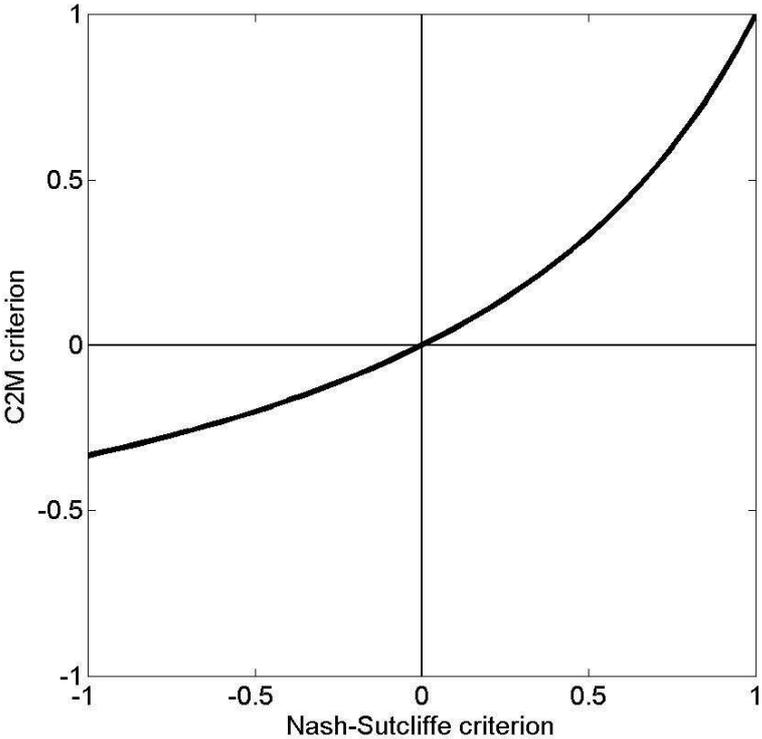


Figure 2 : Relation between the Nash-Sutcliffe and the C_{2M} criteria.

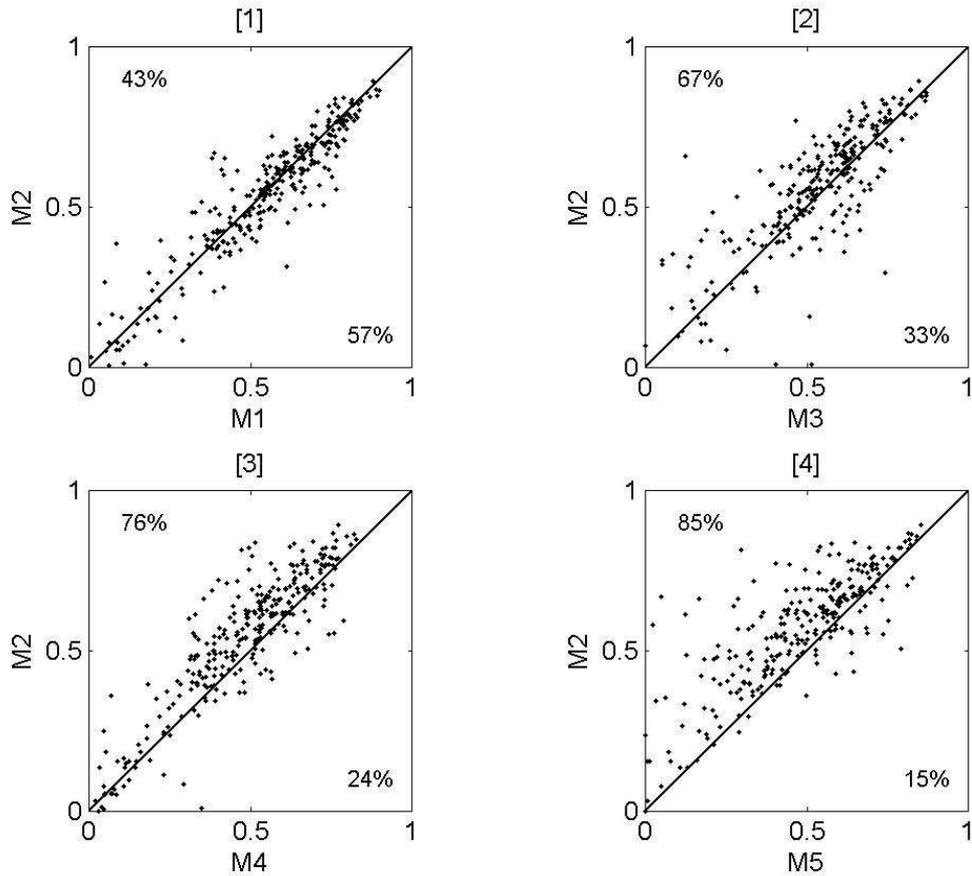


Figure 3 : Comparison of M2 model efficiencies, to M1, M3, M4 and M5 model efficiencies on the sample of 313 watersheds, using the C_{2M} criterion. Watersheds with mean C_{2M} efficiency lower than 0 are not shown. Figures in the corners indicate the percentage of watersheds where a model is better than the other.

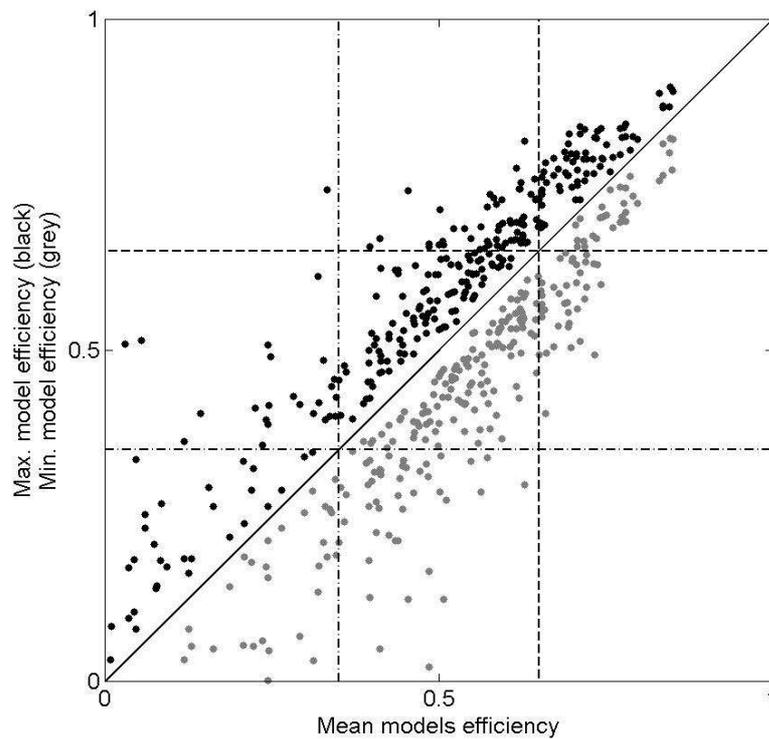


Figure 4 : (a) Minimum (grey) and maximum (black) model efficiency obtained by models M1 to M5 on the sample of 313 watersheds, in function of the mean efficiency obtained by the five models on each watershed, using the C_{2M} criterion. The dashed-dotted lines indicates the lower limit of the domain where the C_{2M} efficiency can be considered as acceptable ($C_{2M} > 0.35$, i.e. $NS > 0.5$); the dotted lines indicates the lower limit of the domain where the C_{2M} efficiency can be considered as good ($C_{2M} > 0.65$, i.e. $NS > 0.8$).

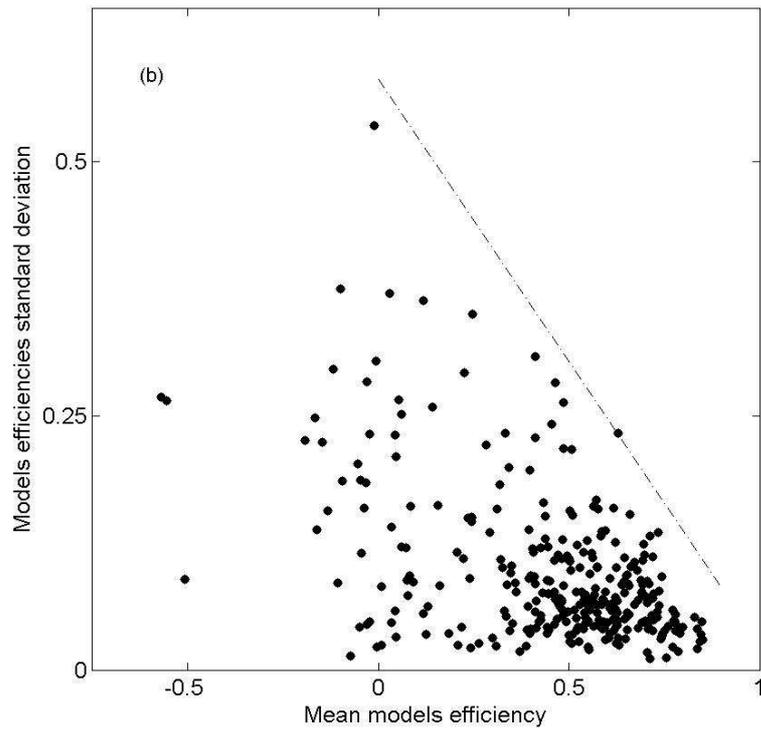


Figure 4 : Standard deviation of the efficiencies obtained by models M1 to M5 on the sample of 313 watersheds, in function of the mean efficiencies obtained by the five models on each watershed using the C_{2M} criterion.

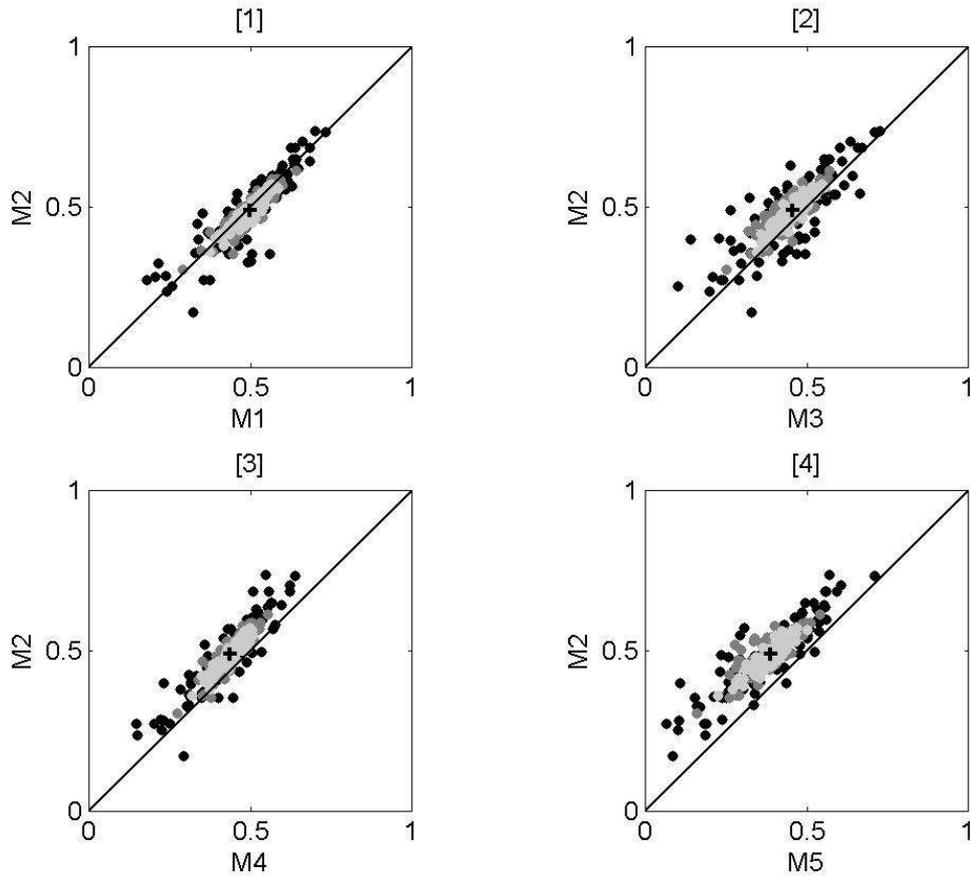


Figure 5 : Comparison of mean efficiencies of model M2, to those of models M1, M3, M4 and M5 for 100 random samples of 5, 20, 50 watersheds, using the C_{2M} criterion. Samples with mean C_{2M} efficiency lower than 0 are not shown. Black dots: 5-watershed samples; Dark grey dots: 20-watershed samples; Grey dots: 50-watershed samples.

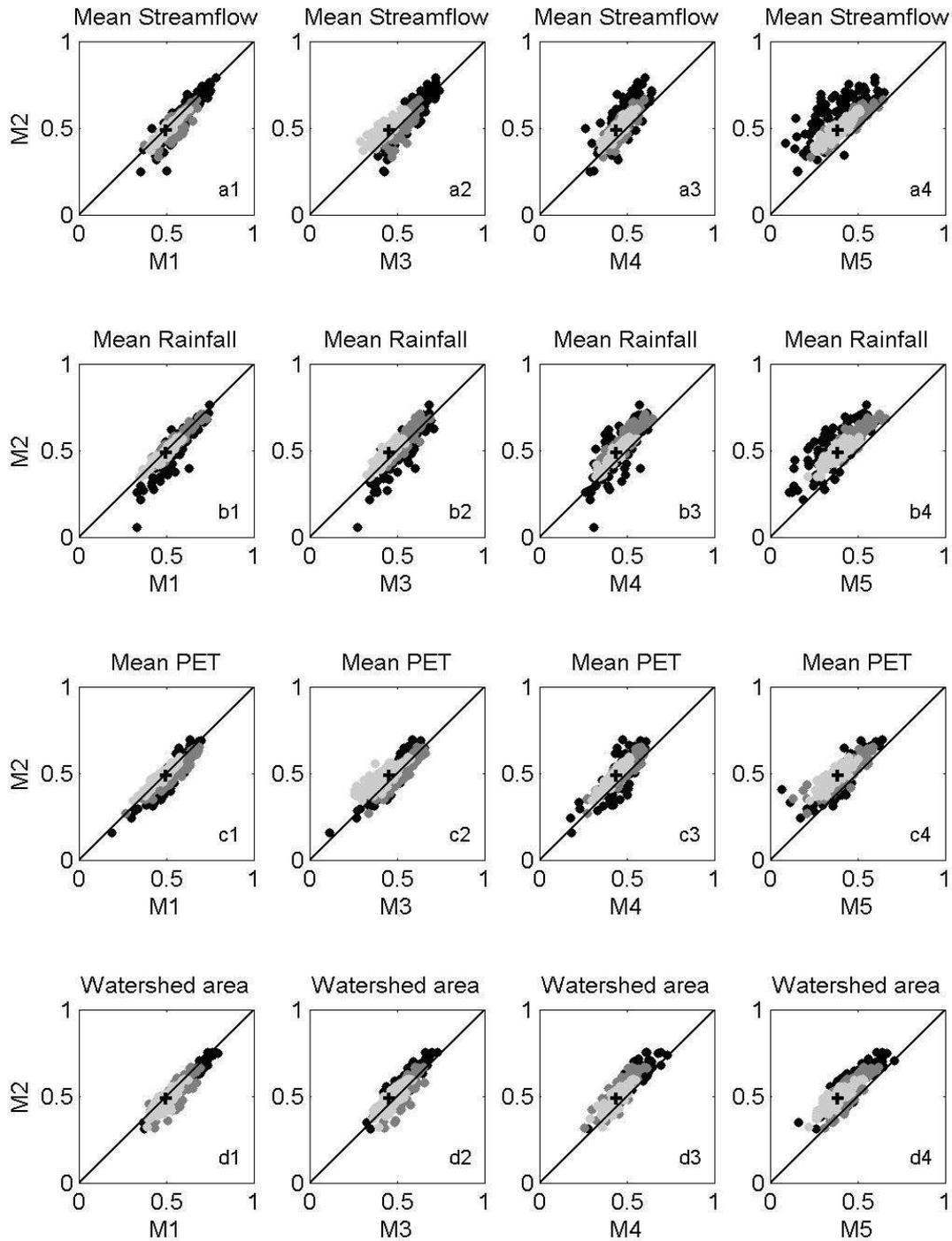


Figure 6 : Comparison of the mean efficiencies of mode M2, to those of models M1, M3, M4 and M5 for 100 random samples of 20 watersheds, using the C_{2M} criterion. Black dots: low descriptor value; Dark grey dots: medium descriptor value; Grey dots: high descriptor value. Descriptors: a1 to a4: streamflow; b1 to b4: rainfall; c1 to c4: PE; d1 to d4: watershed area.