ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

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

(AgroParisTech)

Spécialité : Hydrologie

présentée et soutenue publiquement par

Ioanna ZALACHORI

le 19 Avril 2013

Prévisions hydrologiques d'ensemble : développements pour améliorer la qualité des prévisions et estimer leur utilité

Jury

Mme. Anne-Catherine FAVRE, Professeur, INP, LTHE, Grenoble	Rapporteur
M. Massimiliano ZAPPA, Chercheur, WSL, Birmensdorf	Rapporteur
M. Eric MARTIN, Ingénieur en chef, CNRM – GAME, Météo-France, Toulouse	Examinateur
M. Jocelyn GAUDET, Hydrologue expert, Hydro-Québec, IREQ, Montréal	Examinateur
M. Cyril KAO, Ingénieur en chef, AgroParisTech-ENGREF, Paris	Examinateur
M. Rémy GARCON, Ingénieur expert, EDF DTG Grenoble	Examinateur
Mme. Cécile LOUMAGNE, Directrice de recherche, Irstea, Antony	Co-directrice de thèse
Mme. Maria-Helena RAMOS, Chargée de recherche, Irstea, Antony	Co-directrice de thèse



Le réservoir du barrage de Bramianon ($15*10^6 m^3$) à Ierapetra (Crète), le lieu de mon " baptême " en hydrologie en attendant que mon père finisse sa construction.

''Λοιπόν, ας τελείωνουμε όλοι/ες με τις ματαιοδοξίες μας, με τα μεταδιδακτορικά, τα διδακτορικά, τις δουλειές και τις καριέρες και ας επιστρέψουμε όσο πιο γρήγορα πίσω στην γεμάτη "φως", μικρή πατρίδα... Οι αγώνες, όπως και αν τους ορίζει καθένας/καθεμία μας, τώρα αρχίζουν, και οι δυσκολίες είναι οι πρώτες σοβαρές εξετάσεις της γενιάς μας (και όχι τα qualifying ή τα journal papers)...

Το μόνο σίγουρο είναι ότι το "φως" ήταν, είναι και θα είναι πάντα εκεί...''

Στην Λένγκω λοιπόν,

A Lengo,

Remerciements

Le travail d'un doctorat est une épreuve toujours mentale, parfois physique et plus souvent psychologique. C'est pourquoi je pense que je n'aurais pas été capable d'arriver jusqu'ici sans l'aide précieuse de plusieurs personnes.

Je tiens d'abord à remercier profondément Maria-Helena Ramos pour sa confiance en moi et son encadrement rigoureux. Ces idées et sa façon de penser m'ont ouvert des portes de secours aux moments où je me trouvais face à des impasses. Un grand merci !

Un grand merci également à Rémy Garçon pour son encadrement, sa disponibilité et sa vision hydrologique plus large que tous les ensembles combinés ! C'était une expérience rare ! Un grand merci aussi à Cecile Loumagne d'avoir accepté de diriger et suivre mon travail de thèse. Merci aussi à Thibault Mathevet pour m'avoir initié à la fois à la modélisation hydrologique en opérationnel et à la vie grenobloise avec ses Barbarins Fourchus ! Merci à Joël Gaillard et Mathieu Lelay pour m'avoir aidé à me familiariser avec les données et la prévision opérationnelle chez DTG.

Je tiens aussi à remercier Dimitris Koutsoyiannis de m'avoir ouvert le monde de l'hydrologie et de m'avoir appris que la beauté de notre science se trouve souvent dans la simplicité. A partir de là on peut tout construire.

Mes sincères remerciements à l'ensemble des membres de ce jury de thèse de m'avoir honoré de leur présence et d'avoir accepté de lire et juger mes travaux malgré mon français imparfait. Je tiens à remercier particulièrement Anne-Catherine Favre et Massimiliano Zappa pour avoir beaucoup contribué à l'amélioration de ce manuscrit.

Merci également aux membres de mes comités de suivi : Benjamin Renard et Celine de Saint Aubin pour leurs directions et remarques plus qu'utiles.

J'ai eu la chance pendant ces dernières années d'effectuer mon travail de thèse dans l'équipe jeune et très dynamique d'Irstea. Même si souvent on l'oublie, la recherche est un travail d'équipe. J'espère alors que cet esprit d'équipe règnera toujours à Antony. Je tiens à remercier Vazken Andréassian, responsable de l'équipe Hydrologie qui m'a accueillie en thèse. Un grand merci à Charles Perrin pour son soutien permanent et sa gentillesse, ainsi qu'à toute l'équipe intercontinentale de soutien : Annie Randrianasolo, Raji Pushpalatha, Hajer Dhouioui, Carina Furusho, Louisse Crossmore, Anna Kuenz, Beth Allen, Marine Riffard, Tangara Mamoutou, Damien Lilas, Florent Lobligeois, Laurent Coron, Pierre-Yves et François Bourgin, Pierre Nicolle, Federico Garavaglia, Gianluca Boldetti, Guillaume Thirel, Julien Peschard. Un merci particulier à Laure Lebecherel que le destin a mis dans le même bureau que moi ; presque un siècle après que nos arrières grands-mères se racontaient les ragots du

jour à Smyrne. Merci à tous pour le soutien, les moments passés ensemble et les soirées arrosées chez Mamane – la Mecque de tous les hydrologues !

Et si je suis arrivée à soutenir cette thèse, c'est surtout grâce à l'insistance de mes parents qui nous ont toujours protégés avec tous leurs moyens possibles et qui nous ont toujours encouragés même quand ils n'avaient plus leur propre courage psychique. Et c'est pour honorer leur effort que j'ai trouvé le courage de continuer dans les moments les plus difficiles.

Une grande partie de ma motivation vient de mon frère charismatique qui m'inspire avec sa pensée analytique et sa vision globale de la science depuis l'autre côte de l'Atlantique.

Un grand merci à Maxime, pas seulement pour avoir lu et relu cette thèse tard dans la nuit, mais surtout pour son soutien courageux même dans mes décisions les plus déraisonnables.

Enfin un grand merci à ma famille parisienne : Smaragda, Vassia, Chara, Ino, Gibet, Aude, Elisa, Aurore, Alice, Séverine, Carole, Yannick, Thierry, Louis, Thierry le kilt, Pierre, ……. pour tous les moments passés ensemble (et pas seulement nos soirées 24) !!! Et un souvenir doux pour la villa !

Résumé

La dernière décennie a vu l'émergence de la prévision probabiliste de débits en tant qu'approche plus adaptée pour l'anticipation des risques et la mise en vigilance pour la sécurité des personnes et des biens. Cependant, au delà du gain en sécurité, la valeur ajoutée de l'information probabiliste se traduit également en gains économiques ou en une gestion optimale de la ressource en eau disponible pour les activités économiques qui en dépendent.

Dans la chaîne de prévision de débits, l'incertitude des modèles météorologiques de prévision de pluies joue un rôle important. Pour pouvoir aller au-delà des limites de prévisibilité classiques, les services météorologiques font appel aux systèmes de prévision d'ensemble, générés sur la base de variations imposées dans les conditions initiales des modèles numériques et de variations stochastiques de leur paramétrisation. Des scénarios équiprobables de l'évolution de l'atmosphère pour des horizons de prévision pouvant aller jusqu'à 10-15 jours sont ainsi proposés.

L'intégration des prévisions météorologiques d'ensemble dans la chaîne de prévision hydrologique se présente comme une approche séduisante pour produire des prévisions probabilistes de débits et quantifier l'incertitude prédictive totale en hydrologie. Différents défis apparaissent néanmoins quand il s'agit de fournir des prévisions hydrologiques bonnes et utiles aux utilisateurs. D'une part, l'évaluation des prévisions d'ensemble montre souvent l'existence de différents biais qui nécessitent un traitement statistique avant leur utilisation. D'autre part, les rares études déjà réalisées pour évaluer la valeur économique des prévisions hydrologiques laissent percevoir la valeur économique potentielle des prévisions probabilistes, mais se limitent souvent à des cas particuliers. La réalisation d'études plus approfondies s'impose pour mieux comprendre comment l'utilisateur des prévisions peut en faire le meilleur usage dans la prise de décision.

Dans cette thèse, nous nous sommes fixés deux grands objectifs généraux : 1) évaluer des approches disponibles pour améliorer la *qualité* des prévisions pour les applications opérationnelles de la prévision hydrologique d'ensemble ; 2) développer un modèle capable d'estimer la valeur (*utilité*) d'une prévision d'ensemble.

En termes de qualité, nous nous sommes intéressés en particulier aux attributs de fiabilité et discrimination. En termes d'utilité, nous nous sommes intéressés aux gains économiques de la gestion des apports en eau aux retenues de production hydroélectrique.

Nous avons conceptualisé une chaîne de prévision hydrologique d'ensemble basée sur les sorties de la prévision d'ensemble des modèles météorologiques. Cette chaîne s'étend jusqu'à la valorisation des prévisions hydrologiques par un utilisateur-décideur final, en passant par l'évaluation de la qualité des prévisions et le traitement statistique de leurs biais.

Nous avons constitué une large base de données en s'appuyant sur un total de 85 bassins versants, 4 années de prévisions d'ensemble provenant de 8 centres météorologiques différents et 2 modèles hydrologiques utilisés en prévision opérationnelle en France (*GRP* et *MORDOR*).

Les principaux résultats de cette thèse ont montré que les prévisions hydrologiques affichent un défaut de qualité lié notamment à la sous-dispersion des prévisions d'ensemble. L'utilisation d'un grand ensemble peut conduire à une prévision multi-modèle avec une qualité globale des prévisions considérablement améliorée. En mono-modèle, le traitement statistique des prévisions de débits reste une démarche pratiquement obligatoire pour rendre les prévisions fiables. De plus, l'intérêt d'un post-traitement statistique à la sortie du modèle hydrologique par rapport à un prétraitement des forçages en amont a été clairement démontré.

Il a été également montré qu'il n'y a pas une meilleure méthode de traitement statistique qui s'applique partout et qui permet d'améliorer tous les attributs d'une prévision pour toutes les échéances. Selon les objectives du traitement appliqué, il y a des approches statistiques qui arrivent à améliorer soit la qualité globale des prévisions, soit leur fiabilité, soit leur discrimination.

En ce qui concerne l'utilité des prévisions, le modèle d'optimisation heuristique de la gestion d'une retenue du type éclusée développé dans cette thèse a montré que les prévisions d'ensemble apportent un gain économique supplémentaire par rapport à une prévision uni-scénario (déterministe ou moyenne d'ensemble). De plus, nous avons mis également en valeur les gains économiques obtenus quand un traitement statistique de correction de biais est appliqué aux prévisions d'ensemble avant son utilisation pour la gestion hydroélectrique.

Enfin, nous avons montré que les gains économiques de gestion reflètent en général la qualité de la prévision telle qu'elle peut être évaluée par un critère numérique de qualité. En utilisant des prévisions d'ensemble traitées, le gain économique de gestion s'est montré plus élevé par rapport au gain obtenu à partir des prévisions brutes (sans aucun traitement statistique), de moindre qualité. Cette conclusion est fondamentale, car elle montre l'intérêt de poursuivre des travaux qui cherchent à améliorer la qualité des prévisions hydrométéorologiques, sachant que celle aura un impact également en termes d'utilité et gains économiques.

Mots clés : prévision hydrologique d'ensemble, qualité, utilité, évaluation, traitement statistique des biais, valorisation, gain économique, hydroélectricité.

Abstract

The last decade has seen the emergence of streamflow probabilistic forecasting as the most suitable approach to anticipate risks and provide warnings for public safety and property protection. However, beyond the gains in security, the added-value of probabilistic information also translates into economic benefits or an optimal management of water resources for economic activities that depend on it.

In streamflow forecasting, the uncertainty associated with rainfall predictions from numerical weather prediction models plays an important role. To go beyond the limits of classical predictability, meteorological services developed ensemble prediction systems, which are generated on the basis of perturbations of the initial conditions of the models and stochastic variations in their parameterization. Equally probable scenarios of the evolution of the atmosphere are proposed for forecasting horizons up to 10-15 days.

The integration of weather ensemble predictions in the hydrological forecasting chain is an interesting approach to produce probabilistic streamflow forecasts and quantify the total predictive uncertainty in hydrology. However, several challenges arise when it comes to providing good and useful forecasts to users. On the one hand, the evaluation of ensemble forecasts often shows the existence of different biases in the forecasts, which require statistical correction before their use. On the other hand, the few existing studies that show the potential economic value of probabilistic hydrological forecasts are often limited to specific events or catchments. More in-depth studies are needed to better understand how probabilistic forecasts can be used in the decision-making process in operational contexts.

This thesis is developed around two main objectives: 1) to evaluate existing statistical correction approaches used in operational applications to improve the *quality* of hydrological ensemble predictions, and 2) to develop a model able to estimate the value (*utility*) of an ensemble forecast.

In terms of quality, we focused on the attributes of reliability and discrimination. In terms of utility, we targeted the economic benefits of the management of water inflows to reservoirs of hydroelectric power production.

We conceptualized a hydrological ensemble prediction chain based on meteorological ensemble predictions from numerical weather prediction models. This chain extends up to the value of the forecasts for an end-user, and includes the evaluation of the quality of the hydrological forecasts and their statistical calibration.

In this thesis, we built a large database consisting of data from a total of 85 catchments in France and 4 years of ensemble forecasts from 8 different meteorological centers, as well as 2 hydrological models that are used in operational forecasting in France (*GRP* and *MORDOR*).

The main results of this thesis showed that the quality of a hydrological ensemble forecast is significantly affected by the under-dispersion of its members. Combining ensemble predictions from individual meteorological centers can lead to a 'grand ensemble' multi-model forecast that displays better forecast quality. On the other hand, statistical forecast calibration is practically a mandatory step when dealing with a single-model ensemble forecasting system to guarantee forecast reliability. In addition, the added-value of applying a statistical post-processing technique to the outputs of the hydrological model, compared with applying it to the meteorological forcings used as input to the hydrological model, has been clearly demonstrated.

It was also shown that there is no 'best' statistical correction method that outperforms the others for all catchments and improves all the attributes of a forecast at all leadtimes. According to the objectives of the post-processing, there are statistical approaches that may better contribute to improving the overall quality of the forecasts, their reliability, or their discrimination.

Regarding the usefulness of the forecasts, the heuristic optimization model for the management of reservoir inflows developed during this thesis showed that ensemble forecasts provide additional economic gains compared to a single-value forecast (a deterministic forecast or the ensemble mean). We also showed that statistical forecast calibration techniques may further improve the economic gains obtained when using hydrological ensemble forecasts for hydroelectric power production management.

Finally, we have shown that the economic gains in inflows management generally reflect the quality of the forecasts, as it can be evaluated by a numerical criterion of forecast quality. Economic gains are higher using statistically bias-corrected ensemble forecasts, comparatively to the gains obtained with raw (non-corrected) forecasts. This finding is crucial because it shows the importance of research and operational studies seeking to improve the quality of hydrometeorological forecasts, since this may also have an impact on forecast utility and economic gains.

Keywords: hydrological ensemble prediction, forecast quality, utility, evaluation, statistical post-processing, bias correction, forecast value, economic benefit, hydroelectricity.

Περίληψη

Κατά τη διάρκεια των τελευταίων δεκαετίων οι πιθανοτικές υδρολογικές προβλέψεις εμφανίζονται ως η πιο κατάλληλη μεθοδολογία τόσο για την πρόβλεψη των πλημμυρών, όσο και για την προστασία των ανθρώπινων ζωών και περιουσιών. Αυτού του είδους η πιθανοτική πληροφορία μπορεί να μεταφραστεί σε μια βέλτιστη διαχείρηση των υδατικών πόρων και κατά συνέπεια σε αυξημένα οικονομικά οφέλη. Εντούτοις, η μετεορολογική αβεβαιότητα εξακολουθεί και παιζεί ένα καθοριστικό ρόλο μέσα στην υδρολογική αλυσίδα. Για το λόγο αυτό οι μετεορολογικές υπηρεσίες χρησιμοποιούν τις προβλέψεις πολλαπλών σεναρίων («*ensemble forecasts*»), οι οποίες προέρχονται είτε από αμυδρά διαφορετικές αρχικές συνθήκες, είτε από στοχαστικές διαφοροποιήσεις στην παραμετροποίηση των εξισώσεων. Μ΄αυτό τον τρόπο παράγονται πολλαπλά ισοπίθανα σενάρια σχετικά με την εξέλιξη των ατμοσφαιρικών συνθηκών. Η ενσωμάτωση τέτοιων μετεορολογικών πιθανοτικών προβλέψεων μέσα στην υδρολογική αλυσίδα αποτελεί ένα πολλά υποσχόμενο εργαλείο για την ποσοτικοποίηση της συνολικής υδρολογικής αβεβαιότητας.

Παρόλα αυτά πολλές ερωτήσεις μένουν αναπάντητες σχετικά με την ποιότητα και τη χρησιμότητα των υδρολογικών προβλέψεων. Αφενός, οι πιθανοτικές προβλέψεις πολλαπλών σεναρίων αποκαλύπτουν τις αστοχίες μιας πρόβλεψης, οι οποίες απαιτούν στατιστική διόρθωση. Αφετέρου, προς το παρόν ελάχιστες έρευνες έχουν ασχοληθεί με την οικονομική αξία των πιθανοτικών υδρολογικών προβλέψεων πολλαπλών σεναρίων. Όλα τα παραπάνω μας οδηγούν στο συμπέρασμα ότι απαιτείται περαιτέρω διερεύνηση για την καλύτερη κατανόηση του μηχανισμού λήψης αποφάσεων σε επιχειρησιακό επίπεδο.

Η παρούσα διδακτορική εργασία δομείται γύρω από δύο βασικούς στόχους: 1) την αξιολόγηση των υπαρχόντων στατιστικών μεθοδολογιών διόρθωσης που χρησιμοποιούνται σε επιχειρησιακές εφαρμογές και 2) τη δημιουργία ενός κατάλληλου μοντέλου εκτίμησης της αξίας κάθε πρόβλεψης.

Ως προς την ποιότητα των προβλέψεων, μας ενδιέφερε κυρίως η δυνατότητα των προβλέψεων να είναι αξιόπιστες (*«reliability»*) και διακριτές (*«discrimination»*). Ως προς τη χρησιμότητα, επικεντρωθήκαμε κυρίως στο παραγόμενο οικονομικό όφελος από τη διαχείρηση ενός ταμιευτήρα υδροηλεκτρικού σκοπού.

Η υδρολογική αλυσίδα, που ορίστηκε κατά τη διάρκεια της παρούσας διδακτορικής εργασίας, εκτείνεται από το μετεορολογικό μοντέλο έως τον τελικό χρήστη, περνώντας από την αξιολόγηση της ποιότητας των υδρολογικών προβλέψεων και τη στατιστική τους διόρθωση.

Η βάση δεδομένων που χρησιμοποιήσαμε αποτελείται από: 85 υδρολογικές λεκάνες στη Γαλλία, 4 χρόνια προβλέψεων προερχόμενα από 8 μετεορολογικά κέντρα σε όλα τον κόσμο και 2 υδρολογικά μοντέλα ήδη χρησιμοποιούμενα σε επιχειρησιακες εφαρμογές (*GRP* και *MORDOR*).

Η έρευνά μας κατέληξε ότι η ποιότητα των υδρολογικών προβλέψεων επηρεάζεται σημαντικότατα από την ελλειπή διασπορά των υδρολογικών σεναρίων. Επιπλέον, δείξαμε ότι συνδιάζοντας σενάρια προερχόμενα από διαφορετικά μετεορολογικά μοντέλα (*«grand ensemble»*) βελτιώνουμε την ποιότητα των υδρολογικών προβλέψεων. Παράλληλα, αποδείξαμε ότι η στατιστική διόρθωση είναι απαραίτητη στις προβλέψεις που προέρχονται από ένα μοναδικό μετεορολογικό μοντέλο.

Πιο συγκεκριμένα η έρευνά μας απέδειξε ότι η εφαρμόγη μιας στατιστικής διόρθωσης σε επίπεδο υδρολογικών πιθανοτικών προβλέψεων είναι λιγότερο αποτελεσματική από ότι σε επίπεδο μετεορολογικών προβλέψεων. Επιπλέον, από όλες τις στατιστικές μεθοδολογίες διόρθωσης που δοκιμάστικαν, δεν μπορέσαμε να διακρίνουμε μία ως την πιο αποτελεσματική για όλες τις λεκάνες και για όλους του οριζοντες πρόβλεψης. Ανάλογα με τα χαρακτηριστικά κάθε μεθοδολογίας, κάποιες διορθώνουν είτε την αξιοπιστία των προβλέψων, είτε την ικανότητά τους να διακρίνουν διαφορετικά γεγονότα, είτε τέλος την συνολική ποιότητα της πρόβλεψης.

Κατά τη διάρκεια της παρούσας διδακτορικής εργασίας δημιουργήσαμε ένα ευριστικό αλγόριθμο βελτιστοποίησης για τη διαχείρηση ενός μικρού ταμιευτήρα υδροηλεκτρικού σκοπού. Λόγω του μεγέθους του, η διαχείρηση του ταμιευτήρα γίνεται σε εβδομαδιαία βάση. Με βάση τα αποτελέσματά μας, οι υδρολογικές προβλέψεις πολλαπλών σεναρίων αποφέρουν ένα επιπλέον οικονομικό όφελος σε σχέση με μια πρόβλεψη μοναδικού σεναρίου. Επιπλέον, δείξαμε ότι οι διάφορες μέθοδοι στατιστικής διόρθωσης βελτιώνουν περαιτέρω τον οικονομικό όφελος των υδρολογικών προβλέψεων πολλαπλών σεναρίων.

Η παρούσα διδακτορική εργασία εντόπισε μια υπάρχουσα συσχέτιση μεταξύ της ποιότητας και της χρησιμότητας μιας πρόβλεψης. Κατά συνέπεια, ελλείψει κατάλληλου αλγορίθμου, η χρησιμότητα μιας πρόβλεψης μπορεί να μετρηθεί μέσα από ένα υπάρχον κριτήριο αξιολόγησης της ποιότητας μιας πρόβλεψης. Το συμπέρασμα αυτό είναι ιδιαίτερης σημασίας καθώς αποδεικνύει την αναγκαιότητα των ερευνητικών εργασιών πάνω στη διόρθωση της ποιότητας υδρολογικών προβλέψεων.

Λέξεις κλειδιά: υδρολογικές προβλέψεις πολλαπλών σεναρίωνs, ποιότητα, χρησιμότητα, αξιολόγηση προβλεψης, στατιστική διόρθωση πρόβλεψης, αξιοποίησης πρόβλεψης, οικονομικό όφελος, υδροηλεκτρική παραγωγή.

Table des matières

Remerciements ..... **iv**

Résumé ..... **iv**

Abstract ..... **vi**

Περίληψη ..... **viii**

Table des matières ..... **x**

1. Introduction **14**

1.1 La prévision en hydrologie : enjeux et incertitudes 16
1.2 "*De l'incertitude des prévisions à la prévision des incertitudes*" 18
1.2.1 La prévision numérique du temps : un peu d'histoire 19
1.2.2 La prévision météorologique d'ensemble 20
1.2.3 La prévision hydrologique d'ensemble 22
1.3 Aspects clés de la chaine de prévision hydrologique d'ensemble 25
1.4 Défis actuels dans la recherche de la qualité et de l'utilité des prévisions 27
1.4.1 La qualité des prévisions 27
1.4.2 L'utilité des prévisions 29
1.5 Objectifs de la thèse et structure du mémoire 30

2. Données et chaînes de modélisation hydrologique **32**

2.1 Introduction 34
2.2 Prévisions météorologiques 35
2.2.1 Prévisions d'ensemble du CEPMMT 35
2.2.2 Prévisions d'ensemble de la base TIGGE 37
2.2.3 Prévisions probabilistes par analogie 40
2.3 Bassins versants étudiés 42
2.3.1 74 grands bassins versants en France 42
2.3.2 11 bassins versants de fort intérêt pour la production hydro-électrique 45
2.4 Modélisation hydrologique 48
2.4.1 Modèle hydrologique GRP 48
2.4.2 Modèle hydrologique MORDOR 51
2.4.3 Calage des modèles hydrologiques 52
2.5 Synthèse 56

3. La qualité des prévisions: critères d'évaluation et méthodes de correction de biais..... 58

3.1 Introduction 60
3.1.1 Cadre général de l'évaluation des prévisions d'ensemble 60
3.1.2 Attributs de la qualité des prévisions 62
3.2 Les critères d'évaluation des prévisions 66
3.2.1 Biais inconditionnel 68
3.2.2 Diagramme de PIT 69
3.2.3 Dispersion des membres 71
3.2.4 Courbe ROC 72
3.2.5 Score de Brier 73
3.2.6 Score RPS 74
3.2.7 Score CRPS 75
3.2.8 Score de Divergence 76
3.2.9 Remarques générales sur l'évaluation de la qualité des prévisions 77
3.3 Approches de traitement statistique des prévisions 78
3.3.1 Familles de méthodes de traitement statistique 80
3.3.2 Méthodes appliquées aux prévisions de précipitation 82
3.3.3 Méthodes appliquées aux prévisions de débits 86
3.3.4 Approches pour le calage des paramètres des méthodes appliquées 91
3.4 Synthèse 93

4. L'utilité d'une prévision 96

4.1 Pourquoi associer une valeur économique (utilité) aux prévisions ? 98
4.2 Utilité d'une prévision et prise de décision 99
4.3 Maximiser l'utilité pour une prise de décision optimale 102
4.4 Comment mesurer l'utilité des prévisions ? 103
4.5 Etudes d'évaluation économique en prévision hydrologique d'ensemble ... 107
4.6 Synthèse 111

5. Prévisions d'ensemble multi-modèles météorologiques pour la prévision hydrologique 112

5.1 Etudes hydrologiques basées sur les prévisions TIGGE 114
5.2 Evaluation des prévisions d'ensemble 116
5.2.1 Évaluation des prévisions d'ensemble de pluies 116
5.2.2 Évaluation des prévisions d'ensemble de débits 143

5.3 Combinaison de centres 165
5.3.1 Approches étudiées 165
5.3.2 Approches appliquées 167
5.3.3 Évaluation des prévisions d'ensemble de pluies du Grand Ensemble 168
5.3.4 Évaluation des prévisions d'ensemble de débits du Grand Ensemble 175
5.4 Synthèse 181

6 Traitement statistique des prévisions 184

6.1 Introduction 186
6.2 Scénarios de traitement statistique 187
6.3 Evaluation des prévisions brutes de pluies et de débits 190
6.3.1 Prévisions de pluie 190
6.3.2 Prévisions de débit : modèles MORDOR et GRP 198
6.4 Prétraitement des prévisions de pluie 209
6.4.1 Evaluation de la qualité de l'habillage des prévisions de pluie 209
6.4.2 Quel est l'effet du traitement des pluies sur les débits ? 216
6.5 Post-traitement des prévisions 223
6.5.1 Prise en compte de l'incertitude totale de modélisation 223
6.5.2 Approche combinée : prétraitement et post-traitement 232
6.5.3 Prise en compte de l'incertitude totale de prévision 236
6.5.4 Y-a-il une méthode de correction statistique qui soit la meilleure partout ? . 272
6.5.5 Influence de la longueur de calage des méthodes de correction de biais 274
6.6 Synthèse 278

7 Valorisation des prévisions d'ensemble 282

7.1 La problématique en jeu : la gestion d'une réserve hydroélectrique 284
7.2 Un modèle heuristique de gestion des réserves hydroélectriques 286
7.2.1 Cas d'application : les éclusées hebdomadaires 286
7.2.2 Concept général de l'approche développée 287
7.2.3 Etape de gestion 291
7.2.4 Etape de simulation 298
7.2.5 Etape d'évaluation 299
7.2.6 Prise en compte des prévisions 301
7.3 Résultats 302
7.3.1 Illustration d'un cas de gestion durant un événement de crue 303
7.3.2 Analyse de sensibilité aux caractéristiques de l'aménagement 309
7.3.3 Intérêt des prévisions d'ensemble 313
7.3.4 Impact de la qualité des prévisions sur les gains de gestion 317
7.4 Synthèse 323

8 Conclusions et perspectives 326

8.1 Les objectifs, données et outils de cette thèse 328
8.2 Conclusions 330
8.3 Perspectives 335

Bibliographie 338

Liste des illustrations 356

Annexes 376

ANNEXE A : Les modèles utilisés 377

ANNEXE B : Extrait du modèle heuristique de gestion 390

ANNEXE C : Conférences et ateliers 392

ANNEXE D : Article 394

"Φύσις κρύπτεσθαι φιλεῖ"

(La nature aime se cacher)

Héraclite d'Ephèse

1. Introduction

Dans ce chapitre nous présentons les enjeux de la prévision hydrologique et le cadre général dans lequel se place cette thèse : la prévision hydrologique d'ensemble. Nous présentons les questions qui reflètent les défis de cette thèse, qui s'intéresse à la fois à la qualité et à l'utilité des prévisions. Enfin, nous présentons comment ce mémoire a été structuré pour apporter des éléments de réponse à nos interrogations.

1.1 La prévision en hydrologie : enjeux et incertitudes

Selon l'Agence Européenne de l'Environnement (EEA), entre les années 1998 et 2009, les crues observées en Europe ont causé 1126 morts, 500 000 déplacés et 52 milliards d'euros de pertes pour les assurances. De plus, l'Agence souligne que le risque d'observer des événements de crue plus fréquemment à l'avenir est élevé. Pour éviter les pertes de vies humaines et minimiser les dégâts économiques liés à ces événements extrêmes, la mise en place d'un système de prévision et alerte des crues solide et fiable est essentiel. L'objectif général est de pouvoir anticiper l'arrivée d'un événement potentiellement dangereux, ainsi que de bien capturer son intensité et localisation. Prévision et prévention des risques vont ainsi de paire. La Directive Européenne relative à l'évaluation et à la gestion des risques d'inondation de 2007 (DIRECTIVE 2007/60/CE) rappelle ainsi que " les plans de gestion des risques d'inondation englobent tous les aspects de la gestion des risques d'inondation, en mettant l'accent sur la prévention, la protection et la préparation, y compris la prévision des inondations et les systèmes d'alerte précoce (…) " (Chapitre IV, Article 7, §3).

Cependant, l'importance de la prévision hydrologique ne se restreint pas à l'anticipation des événements extrêmes, qu'ils soient caractérisés par des situations de crues ou d'étiages. Un système de prévision hydrologique efficace peut également contribuer à la gestion optimale de la ressource en eau pour les activités économiques qui en dépendent (production hydroélectrique, irrigation, tourisme, navigation, etc.). Il s'agit de mieux répartir l'utilisation de l'eau dans le temps et dans l'espace, à partir de l'information apportée par le système de prévision sur les apports en eau à venir. Au delà du gain en sécurité face aux risques des événements extrêmes, la valeur ajoutée de cette connaissance se traduira en gains économiques ou de gestion des ouvrages

Idéalement, pour que les services de protection civile ou les gestionnaires d'ouvrages aient le temps suffisant pour intervenir, une alerte doit être déclenchée au minimum dans les 12 à 24 heures précédant l'événement. De plus, les services opérationnels de prévision doivent pour cela avoir des procédures efficaces de prévision, interprétation et communication des alertes pour répondre aux besoins liés à ces temps de réaction.

Dans ces cas, il est évident qu'une prévision hydrologique à moyen terme (3 à 10 jours en avance) a toute sa place. Pour cela, la prévision météorologique du temps à moyen terme est indispensable comme source d'information sur les conditions météorologiques à venir qui seront utilisées en entrée des modèles hydrologiques.

Pour pouvoir faire une prévision du temps à moyen terme et aller au-delà des limites de prévisibilité classiques, la prise en compte des incertitudes est incontournable. La prévision déterministe, à scénario unique, n'est pas en mesure de répondre à ces objectifs. Pour cela, les services météorologiques font appel aux systèmes de prévision d'ensemble, générés sur la

base de variations imposées dans les conditions initiales des modèles numériques et de variations stochastiques de leur paramétrisation (Palmer *et al.*, 2005). La prévision météorologique d'ensemble propose des scénarios équiprobables de l'évolution de l'atmosphère pour des horizons de prévision pouvant aller jusqu'à 10-15 jours, ce qui serait en dehors de la place de prévisibilité des solutions traditionnelles déterministes. Le recours à la technique de la prévision d'ensemble permet ainsi, d'une part, d'étendre les échéances de prévision des variables à faible prédictibilité, telles les précipitations, et, d'autre part, de donner une mesure de l'incertitude des prévisions.

Par contraste à une solution unique déterministe, la prévision probabiliste est vue comme mieux adaptée à l'analyse de risque et la prise de décision (Rodwell, 2005). La valeur économique potentielle des prévisions probabilistes ressort notamment face aux utilisateurs exposés aux risques climatiques et hydrologiques à différentes échelles ; ainsi qu'aux gestionnaires dont les décisions dépendent de la disponibilité nécessaire en eau : pour la génération d'électricité, agriculture et irrigation, navigation, sécurité et santé publiques, etc. (Buizza, 2008, 2001 ; Houdant, 2004).

L'approche probabiliste ou d'ensemble est actuellement encore considérée comme faisant partie de l'état de l'art et plusieurs services météorologiques (France, Suisse, Belgique, Allemagne, Suède, etc.) développent ou envisagent de constituer un système de prévisions hydrologiques basé sur les prévisions d'ensemble (Cloke et Pappenberger, 2009b). La prévision hydrologique d'ensemble est au cœur des initiatives internationales regroupant plusieurs centres de recherche et services opérationnels (HEPEX *Hydrological Ensemble Prediction Experiment*, EFAS *European Flood Awareness System*), ainsi que dans les actions de recherche et développement de plusieurs services nationaux (Centre National de Recherche Météorologique, Météo-France, EDF, Institut Royal Météorologique de Belgique, etc.). Les résultats de prévision des crues à partir de la prévision d'ensemble de pluies sont encourageants.

En météorologie, les systèmes de prévision d'ensemble sont développés pour prendre en compte l'incertitude liée aux prévisions. L'application de ces prévisions dans un modèle pluie-débit pour la prévision hydrologique constitue une étape importante pour la prévision des débits, puisque les précipitations sont une des sources principales d'incertitude.

Néanmoins, l'incertitude se trouve dans toute la chaîne de prévision hydrométéorologique et se propage (cascade des incertitudes) de la donnée d'entrée jusqu'à la sortie du modèle. Les incertitudes de prévision sont des sources d'erreur, mais le manque de connaissances des incertitudes constitue également une source d'erreur. La prise en compte de l'ensemble de sources d'incertitude est nécessaire afin d'améliorer la qualité des prévisions et de fournir des informations fiables et utiles.

Le fait de disposer d'une prévision sous forme probabiliste, et non pas de manière déterministe, est sans doute un avantage considérable pour les services opérationnels qui font face à la prise de décision en contexte incertain. Cependant, une telle approche soulève également de nouvelles questions : comment interpréter les sorties probabilistes d'un modèle de prévision ? Comment communiquer les prévisions avec leurs incertitudes ? Comment prendre en compte toutes les sources d'incertitude dans un système de prévision? Comment évaluer la qualité et l'utilité d'un système de prévision ?

Certaines de ces questions seront abordées dans le cadre de cette thèse. Le cadre général dans lequel se développe cette étude est présenté ci-après (§1.2 et §1.3), suivi des défis actuels auxquels s'intéresse cette thèse (§1.4) et des objectifs spécifiques de nos travaux (§1.5).

1.2 "*De l'incertitude des prévisions à la prévision des incertitudes*"

Cette thèse s'intéresse aux prévisions hydrologiques probabilistes (ou multi-scénario) à moyenne échéance. Le système de prévision hydrologique est une chaîne qui a pour principaux éléments (Figure 1) : les données observées (précipitation, température et débits en temps réel observés jusqu'à l'instant de prévision), les prévisions météorologiques (scénarios futurs jusqu'à l'échéance cible) et le modèle hydrologique de transformation de la précipitation en débit. Le but est de fournir des prévisions hydrologiques pour une des utilisations cibles (génération d'électricité, agriculture et irrigation, navigation, sécurité et santé publiques, etc.).

Dans cette thèse, nous avons utilisé l'approche de la prévision d'ensemble pour prendre en compte des incertitudes de prévision et nous avons cherché à explorer l'usage de la prévision météorologique d'ensemble pour la prévision du temps à venir dans la prévision hydrologique des débits des rivières. Pour mieux comprendre nos objectifs de thèse, quelques notions introductives en prévision du temps et approches d'ensemble sont présentées par la suite.

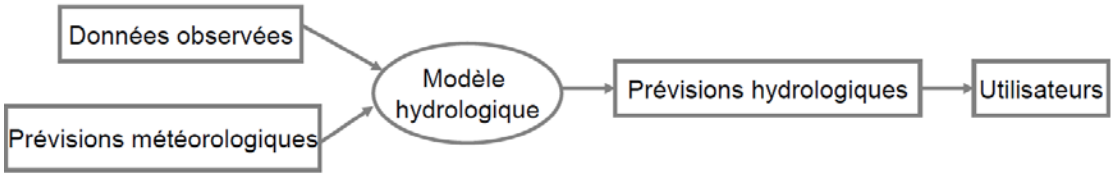


Figure 1 : Schéma général d'une chaîne de prévision hydrologique.

1.2.1 La prévision numérique du temps : un peu d'histoire

Les efforts pour prévoir le temps sont plus que centenaires. En 1904, le physicien V. Bjerknes était le premier à écrire sur la modélisation et la prévision du temps, en utilisant des équations de l'hydrodynamique basées sur des lois de Newton et de la thermodynamique. Le principe général de la prévision y était formulé : " connaissant l'état de l'atmosphère à un instant donné, et les équations qui régissent son évolution, il était possible de connaître l'état de l'atmosphère à tout instant ultérieur " (Rochas, 2000). Le mathématicien L.F. Richardson a tenté de résoudre ces équations par la méthode des différences finies et a été ainsi le précurseur de la modélisation numérique de prévision du temps. Après avoir consacré plusieurs années à la prévision du temps, il est arrivé à la conclusion que pour prévoir le temps plus vite que la vitesse avec laquelle le temps évolue il lui fallait 64 000 personnes travaillant simultanément ! En effet, de nombreux progrès étaient encore nécessaires pour que les méthodes de prévision du temps arrivent à être appliquées avec succès : il fallait des ordinateurs plus puissants, des moyens de mesure des paramètres atmosphériques en altitude couvrant une bonne partie de la planète, ainsi qu'un réseau de communication rapide de ces mesures vers les centres de prévision.

C'est dans les années 1950/60 que les premiers modèles opérationnels de prévision ont vu le jour. Depuis, les modèles ont considérablement évolué grâce à la meilleure représentation des processus physiques, à la meilleure connaissance de la microphysique des nuages, à la prise en compte de mesures nouvelles provenant des réseaux radar et des satellites, et à l'amélioration de la capacité de calcul des ordinateurs, qui a permis, notamment, d'augmenter la résolution spatiale des modèles.

Néanmoins les modèles numériques sont encore soumis à des erreurs liées aux conditions initiales, à leur paramétrage et structure (Murphy *et al.*, 2004 ; Bell *et al.*, 2004 ; Zupanski et Zupanski, 2006 ; Johnson et Swinbank, 2009 ; Goldstein et Rougier, 2009). Les observations qui sont assimilées dans le modèle afin de produire une bonne estimation de l'état actuel de l'atmosphère ont des incertitudes considérables (Leutbecher et Palmer, 2007). Ces erreurs se propagent et s'y ajoutent les erreurs du modèle, notamment sur les régions où la grille spatiale reste trop large pour modéliser les phénomènes locaux et où la représentation physique des phénomènes atmosphériques est imparfaite (Buizza et Palmer, 1995). Les principales sources des fléaux de la prévision se trouvent donc exactement dans le berceau de la prévision : les données, les modèles et le caractère chaotique de l'atmosphère.

En effet, c'est pour cette raison qu'il est impossible de prédire son état exact (Lorenz, 1969). A cause du caractère chaotique, non-linéaire et complexe de l'atmosphère, l'évolution temporelle de deux systèmes météorologiques ressemblant au départ à un système observé ne sera pas nécessairement la même. Ceci est expliqué par les petites différences dans leurs conditions initiales, lesquelles grandissent pendant l'évolution temporelle du système (Lorenz, 1982).

Ainsi, plus la prévision que l'on souhaite fournir est lointaine, plus l'état futur devient incertain. Cette incertitude est aussi dépendante du flux atmosphérique et, par conséquent, elle est considérablement plus importante pour les phénomènes extrêmes, qui sont alimentés par de grands flux d'énergie.

Le caractère chaotique de l'atmosphère contrôle l'imprédictibilité des événements naturels. Selon Koutsoyiannis (2010), dans les processus physiques, les composantes déterministes (qui sont équivalentes à la prédictibilité) et chaotiques (qui sont équivalentes à l'imprédictibilité) coexistent. Selon la composante dominante, la prédictibilité d'un phénomène change. Cela veut dire que les processus très chaotiques peuvent avoir une prédictibilité de quelques jours ou même de quelques heures, alors que, pour des conditions plus stables, la prédictibilité peut s'éteint, par exemple, à plusieurs jours ou même des semaines.

Joly (2008) offre la définition suivante de la prévisibilité : "*l'échéance au-delà de laquelle une prévision s'écarte trop de l'évolution observée pour rester utile*". Nous notons que dans cette définition, les notions de qualité (proximité ou écart de l'observation) et utilité se côtoient. De plus, nous observons que la notion d'incertitude devient grandissante avec l'horizon de prévision. Le besoin de quantifier cette incertitude pour que les prévisions à horizon lointain restent utiles devient une évidence.

1.2.2 La prévision météorologique d'ensemble

Pendant plusieurs décades, les centres opérationnels de prévision du temps lançaient leurs modèles numériques de prévision à partir de la meilleure estimation des conditions initiales. La prévision était ainsi limitée à une seule indication de l'état futur possible de l'atmosphère, le " *best guess* ". En 1992, des modèles de prévision ont commencé à fonctionner en considérant des conditions initiales légèrement perturbées. Le service national de prévision des Etats-Unis (NCEP) et le centre européen de prévision météorologique à moyen terme (CEPMMT) ont été pionniers dans cette démarche (Molteni et Palmer, 1993 ; Toth et Kalnay, 1993). La prévision d'ensemble est ainsi née.

Palmer *et al.* (2002) définissent la prévision d'ensemble comme "une technique de prévision météorologique dans laquelle on fait tourner un modèle numérique de prévision plusieurs fois, pour une même situation à prévoir, à partir de conditions initiales différant entre elles par de petites quantités [les perturbations initiales] compatibles avec les incertitudes existant sur la connaissance de l'état initial de l'atmosphère". Le problème fondamental de la prévision d'ensemble est d'effectuer un choix judicieux des conditions initiales de façon à pouvoir obtenir un maximum de solutions éloignées les unes des autres avec un minimum d'états initiaux. Il faut construire un ensemble initial qui sera un échantillon représentatif de la loi de probabilité qui décrit l'incertitude des conditions initiales.

L'idée est d'échantillonner la loi de probabilité des conditions initiales et de suivre la propagation des différentes perturbations dans l'évolution de l'état de l'atmosphère pour ainsi avoir différents scénarios ou trajectoires de prévision (Figure 2). La prévision d'ensemble est ainsi souvent faite d'une trajectoire provenant de la meilleure estimation de l'état de l'atmosphère au moment de la prévision (trajectoire ou membre dit " contrôle ") et des autres trajectoires obtenues en perturbant les conditions initiales. Idéalement les perturbations de la meilleure estimation de l'état initial conduiront à un ensemble avec une dispersion similaire à la dispersion des erreurs de prévision au bout d'un certain intervalle de temps appelé "période d'optimisation" (Ehrendorfer et Tribbia, 1997).

Plusieurs méthodes permettant de créer les perturbations des conditions initiales existent. Parmi les méthodes opérationnelles les plus courantes se trouvent la méthode des Vecteurs Singuliers (Buizza et Palmer, 1995), la méthode des erreurs croissantes (en anglais *"Breeding of Growing Modes"*), (Toth et Kalnay, 1993, 1997) et la méthode du filtre Kalman d'ensemble (Bishop *et al.*, 2001 ; Kalnay, 2003). De manière générale, toutes ces méthodes visent à appréhender l'incertitude liée aux conditions initiales pour la production des prévisions d'ensemble correctement dispersées. Néanmoins, comme le remarque bien Smith (2007), l'incertitude des conditions présentes ne va se traduire en incertitudes bien quantifiées dans le futur que si le modèle est parfait. De ce fait, une approche complémentaire dans la création des ensembles a été introduite, celle de la perturbation des paramétrisations physiques des modèles (Buizza *et al.*, 1999 ; Houtekamer et Lefaivre, 1997). Il s'agit d'introduire des perturbations à la meilleure estimation de la paramétrisation physique (celle de la trajectoire de contrôle), afin de prendre en compte également les incertitudes liées à la formulation du modèle. Souvent il s'agit de perturbations stochastiques des paramètres.

De plus, on notera l'existence également d'approches qui combinent les sorties de différents modèles numériques du temps pour créer des prévisions d'ensemble :

- l'ensemble du pauvre (en anglais, "*poor man's ensemble*") : combinaison de prévisions déterministes (uni-valeurs) provenant de différents centres météorologiques pour former un ensemble,
- le grand ensemble TIGGE (acronyme en anglais de *THORPEX Interactive Grand Global Ensemble*) : combinaison de prévisions d'ensemble provenant de différents centres météorologiques autour du globe. L'objectif derrière cette initiative est de capturer les incertitudes liées aux conditions initiales, à la paramétrisation, à la structure et à l'assimilation des données des différents modèles numériques du temps utilisés par les différents services météorologiques (Park *et al.*, 2008).

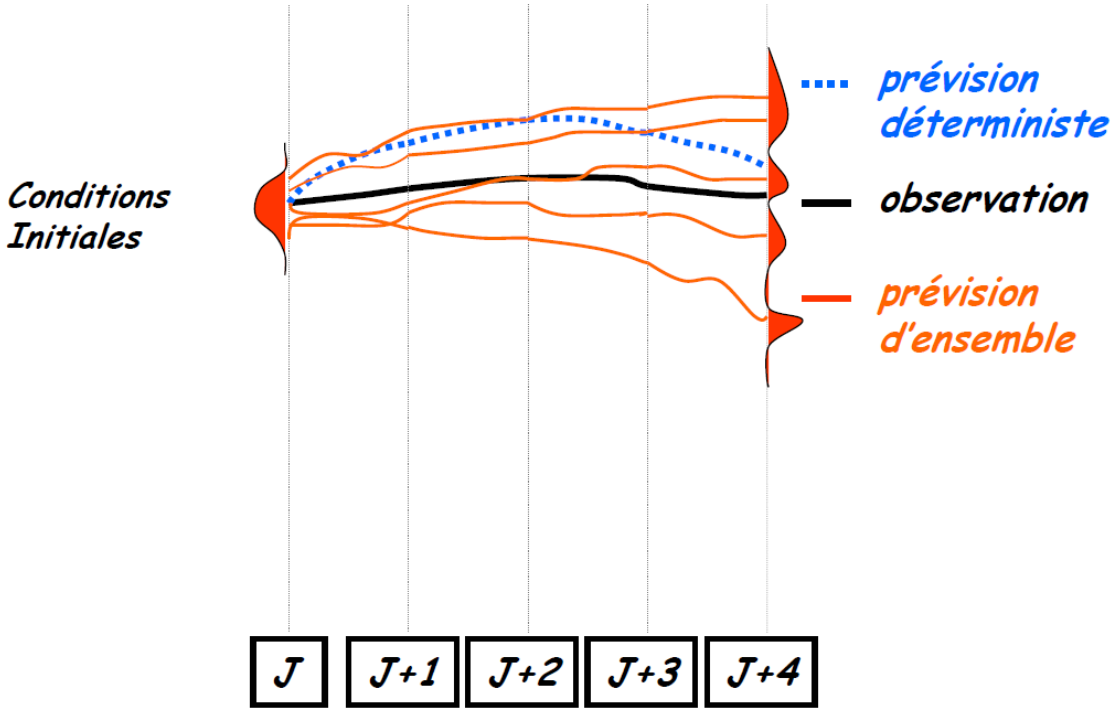


Figure 2: Schéma des prévisions de l'évolution de l'atmosphère pour le cas d'une prévision déterministe et des prévisions d'ensemble et des échéances allant jusque J+4.

Dans cette thèse, nous nous intéressons plus particulièrement à l'utilisation des prévisions météorologiques d'ensemble pour la prévision hydrologique d'ensemble. La description plus détaillée de ces techniques de modélisation d'ensemble n'est pas l'objectif de cette thèse. Nous retenons seulement qu'il existe plusieurs méthodes de génération d'ensembles actuellement utilisées par les centres météorologiques, sans qu'aucune ne soit universellement considérée comme la meilleure.

1.2.3 La prévision hydrologique d'ensemble

Même si les prévisions météorologiques d'ensemble à moyenne échéance sont produites par les services opérationnels depuis 20 ans, ce n'est que pendant la dernière décennie que les premiers systèmes de prévision hydrologique les utilisant ont vu le jour (De Roo et al., 2003).

En effet, l'utilisation des ensembles en hydrologie était plutôt répandue dans les cas de la prévision hydrologique à longue échéance (> 15 jours). La méthode ESP ("Ensemble Streamflow Prediction") a été développée par le service météorologique national des Etats-Unis ("National Weather Service") dans les années 1970-1980 (Day, 1985) dans le but de repousser l'horizon des prévisions hydrologiques. Il s'agit d'utiliser des données météorologiques historiques afin de générer un ensemble des scénarios possibles des débits futurs, conditionnellement aux états initiaux du bassin versant.

L'utilisation de la prévision météorologique d'ensemble pour la prévision hydrologique à des horizons plus courts est plus récente. Elle a été néanmoins rapidement répandue, notamment grâce à des initiatives significatives qui ont été proposées dans un large cadre de coopération entre instituts de recherche et services opérationnels. Dans ce contexte, il ressort notamment:

- l'EFAS (acronyme en anglais de *European Flood Alert System,* récemment rebaptisé *European Flood Awareness System*) : il s'agit d'un système européen d'alerte aux crues dont le développement a été initié en 2003 par le centre commun de recherche de la commission européenne (*JRC*) afin d'améliorer la prévision dans les grands bassins versants transnationaux et d'encourager la coopération entre Etats-Membres. L'un des objectifs de cette initiative est de fournir des pré-alertes (échéances supérieures à 3 jours) aux services nationaux de prévision à partir de l'utilisation de la prévision météorologique d'ensemble en entrée à un modèle hydrologique distribué, mis en place sur des bassins européens transnationaux, et ainsi contribuer à augmenter le temps de préparation des services opérationnels en cas de crues potentiellement dangereuses (Thielen *et al.*, 2009 ; Bartholmes *et al.*, 2009) ;

- l'HEPEX (acronyme en anglais de *Hydrological Ensemble Prediction Experiment*) : il s'agit d'une initiative internationale démarrée en 2003, suscitant fomenter la coopération entre chercheurs en météorologie et hydrologie et services opérationnels dans le domaine de la prévision hydrologique d'ensemble (Schaake *et al.*, 2005, 2007). HEPEX a pour but de réunir les communautés de météorologues, hydrologues et gestionnaires de l'eau pour promouvoir la mise en place des systèmes de prévision hydrologique d'ensemble fiables, ainsi que la communication et l'utilisation des prévisions dans la prise de décision.

Ces dernières années, des nombreuses études ont ainsi été dédiées à l'évaluation de l'intérêt de la prévision d'ensemble pour la prévision hydrologique, qu'il s'agit d'étudier :

- la viabilité de l'approche dans un contexte opérationnel de prévision à moyenne échéance (Rousset-Regimbeau *et al.*, 2007 ; Zappa *et al.*, 2008 ; Addor *et al.*, 2011) ou pour la prévision des crues rapides (Younis *et al.*, 2008a ; Alfieri et Thielen, 2012 ; Marty *et al.*, 2012),
- la qualité des prévisions émises soit sur la base de longues séries prévision-observation (Roulin et Vannitsem, 2005 ; Rousset-Regimbeau *et al.*, 2008 ; Olsson et Lindstrom, 2008 ; Jaun et Ahrens, 2009 ; Demargne *et al.*, 2009 ; Bartholmes *et al.*, 2009 ; Randrianasolo *et al.*, 2010), soit sur des événements particuliers (Bartholmes et Todini, 2005 ; Pappenberger *et al.*, 2005 ; Komma *et al.*, 2007 ; Jaun *et al.*, 2008 ; Kalas *et al.*, 2008 ; Younis *et al.*, 2008b),
- l'impact de la prévision d'ensemble pour la courte échéance et des petites échelles spatiales (Thirel *et al.*, 2008 ; Marsigli *et al.*, 2008),

- l'apport des approches multi-modèles hydrologiques (Velazquez *et al.*, 2009) ou multi-modèles météorologiques (Pappenberger *et al.*, 2008 ; Jasper *et al.*, 2002 ; Davoglio *et al.*, 2008),
- le développement de méthodes de correction de biais (Krzysztofowicz et Kelly, 2000 ; Reggiani et Weerts, 2008 ; Brown et Seo, 2010 ; Coccia et Todini, 2011),
- la valeur économique espérée des prévisions (Buizza, 2008 ; Verkade et Werner, 2011 ; Boucher *et al.*, 2012), ou
- la communication des incertitudes et l'utilisation des prévisions incertaines pour la prise de décision (Houdant, 2004 ; Norbert *et al.*, 2010 ; Ramos *et al.,* 2010 ; 2012).

Toutes ces études ont derrière elles au moins deux points en commun : elles s'intéressent à l'application opérationnelle des prévisions hydrologiques d'ensemble et confirment la valeur ajoutée apportée par la prévision d'ensemble à un système de prévision et prise de décision.

Actuellement, plusieurs services opérationnels en France (SCHAPI - Météo-France/CNRM, EDF) et ailleurs (Cloke et Pappenberger, 2009b) ont développé ou envisagent de constituer un système de prévision hydrologique basé sur les prévisions d'ensemble. Pour les services opérationnels, le fait que la prévision des débits (en crues ou périodes normales) à partir de la prévision météorologique d'ensemble soit présentée sous forme probabiliste ou multi-scénarios, et non pas de manière déterministe (scénario futur unique), est un avantage considérable, notamment pour la prise de décision en contexte incertain (Ramos *et al.*, 2010). En prenant en compte l'incertitude de la prévision météorologique, les prévisions d'ensemble peuvent couvrir des échéances de prévision plus lointaines et augmenter ainsi le temps de préparation (Cloke et Pappenberger, 2009a). De plus, la mise à disposition d'une information probabiliste permet aux décisionnaires de prendre leur décision eux-mêmes, au regard des risques qu'ils souhaitent prendre, au lieu de laisser cette décision dans les mains des modélisateurs ou des prévisionnistes, parfois éloignés des objectifs de l'utilisateur des prévisions (Weerts *et al.*, 2010).

Enfin, on notera que toutes les prévisions d'ensemble ne constituent pas de "bons" ensembles. Par exemple, souvent les ensembles n'arrivent pas à bien décrire toutes les évolutions possibles de l'atmosphère et, par conséquent, l'observation n'est pas capturée par le faisceau des prévisions défini par les trajectoires de l'ensemble. Des traitements statistiques sont alors nécessaires pour corriger les éventuels biais. De plus, afin de juger si une prévision d'ensemble est "bonne" ou pas, il y a deux points de vue différents, même si pas forcement disjoints : celui du prévisionniste et celui de l'utilisateur. Le prévisionniste est souvent intéressé par la qualité des prévisions en termes de proximité par rapport à l'observation, tandis que l'utilisateur focalise l'utilité de l'information pour la décision qu'il doit prendre. Par conséquent, l'évaluation des prévisions d'ensemble concernera souvent deux axes principaux : i) trouver où se situe l'observation par rapport à la gamme de valeurs prévues des différents membres de l'ensemble (*évaluation de la qualité*) et ii) déterminer

l'utilité de la prévision d'ensemble pour un utilisateur-décideur (valeur décisionnelle ou économique) par rapport aux prévisions déterministes et/ou à une prévision dite naïve, telle que la climatologie, c'est-à-dire comparer chaque observation à une prévision égale à la moyenne climatologique, ou la persistance, c'est-à dire comparer chaque observation à une prévision égale au dernier débit observé (*évaluation de l'utilité*) (Jolliffe et Stephenson, 2003 ; Buizza *et al*., 2007b).

1.3 Aspects clés de la chaine de prévision hydrologique d'ensemble

Même si l'incertitude des prévisions météorologiques est souvent considérée comme la principale source d'erreur pour les prévisions hydrologiques, notamment pour les échéances supérieures à 2-3 jours, d'autres facteurs influent également sur la qualité des prévisions hydrologiques. Les principaux d'entre eux sont les erreurs liées à la modélisation hydrologique (structure et paramétrage du modèle) et aux observations utilisées en temps réel pour la mise à jour des conditions initiales du modèle hydrologique avant l'instant de prévision (Pierce *et al.*, 2005 ; Wood et Lettenmaier, 2008 ; Bogner et Kalas, 2008 ; Zhao *et al.* , 2011 ; Pagano *et al.*, 2012 ; Zalachori *et al.*, 2012). La compatibilité entre les échelles spatio-temporelles des sorties des modèles météorologiques et celles requises par les modèles hydrologiques, souvent plus fines, peut également devenir une source d'incertitude et mériter un traitement préalable de désagrégation ou descente d'échelle (Marty *et al.*, 2012).

Dans la réalité opérationnelle, le prévisionniste se trouve souvent face à une combinaison de plusieurs de ces sources d'incertitude (Pappenberger *et al.*, 2005). Etant donné qu'il est devenu évident qu'aucune amélioration de modèle n'éliminera totalement toutes les incertitudes de prévision, il est indispensable de prendre en compte cette incertitude dans la procédure de prévision, même si on doit le faire de manière globale et/ou implicite. Une chaîne de prévision hydrologique doit en effet être capable de gérer les différentes sources d'incertitudes et fournir des prévisions précises et fiables à des utilisateurs-décideurs exposés aux risques hydrologiques à différentes échelles. Elle doit en plus permettre d'évaluer l'impact de la prise en compte des incertitudes sur la qualité et l'utilité opérationnelle des prévisions hydrologiques.

Pour cela, plusieurs aspects doivent être considérés au sein d'une chaîne de prévision hydrologique d'ensemble, dont (au moins) les quatre points-clés suivants (Figure 3):

I. l'*évaluation de la qualité* des prévisions hydrométéorologiques émises : la qualité d'une prévision est en général jugée par rapport à l'observation. Pour cela il est essentiel de comparer une longue série de prévisions avec la série correspondante des observations. Selon Jolliffe et Stephenson (2003) cette évaluation de la qualité d'une prévision d'ensemble consiste à trouver où se situe l'observation par rapport à la gamme des valeurs prévues des différents membres de l'ensemble ; et estimer ainsi

l'écart entre les valeurs prévues et celles réellement observées ("biais"). Cette étape précède toute utilisation des prévisions hydrométéorologiques d'ensemble,

II. le *prétraitement* des prévisions utilisées en entrée du modèle hydrologique : il comprend toutes les approches de correction statistique de biais mises en place en amont du modèle hydrologique afin de réussir au mieux à prendre en compte l'incertitude liée aux prévisions météorologiques,

III. le *post-traitement* statistique des prévisions : il s'agit de toutes les approches de correction statistiques employées à la sortie du modèle hydrologique pour corriger les biais éventuels des prévisions des débits,

IV. l'*évaluation de l'utilité* des prévisions : il s'agit d'examiner si à l'aide de l'information probabiliste fournie par les prévisions d'ensemble, les utilisateurs peuvent mieux décider des actions à mener pour maximiser les gains ou limiter les pertes liées à l'impact d'un événement à venir sur leurs activités. L'évaluation objective de l'utilité d'une prévision sur le long terme fait appel à des techniques économiques, qui cherchent à proposer une mesure de l'utilité, "valeur économique", d'une prévision pour aider à la prise de décision.

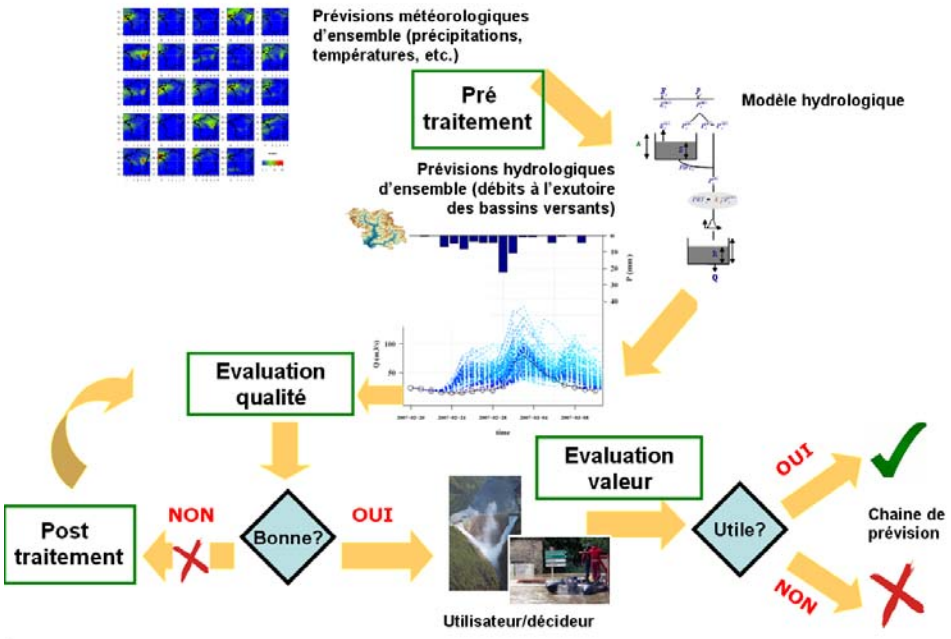


Figure 3: Aspects clés de la chaîne de prévision hydrologique d'ensemble

Ces quatre aspects constituent le cœur des travaux de cette thèse. Ils comportent de nombreux défis à relever pour que le système de prévision puisse émettre des prévisions de bonne qualité et utiles. Ils sont listés dans le §1.4 ci-dessous.

1.4 Défis actuels dans la recherche de la qualité et de l'utilité des prévisions

Dans l'objectif de fournir des prévisions hydrologiques *bonnes* et *utiles* aux utilisateurs, il est essentiel de répondre aux différents défis de la chaîne de prévision hydrologique. Pour cette raison, la première étape consiste à les identifier.

1.4.1 La qualité des prévisions

L'évaluation des prévisions d'ensemble montre souvent l'existence de différents biais dans les prévisions. Une fois identifiés, une correction statistique appropriée doit se mettre en place afin de corriger ces biais et minimiser les différences entre la série de données observées et les prévisions d'ensemble émises.

Il y a plusieurs raisons qui expliquent ces différences :

- les erreurs d'échantillonnage liées à la trop petite taille de l'ensemble ;
- les biais ou erreurs systématiques liés à l'observation ou au(x) modèle(s) hydrométéorologiques (paramètres, structures, conditions initiales, etc.) ;
- la dispersion inappropriée des membres, venant soit d'une mauvaise prise en compte des incertitudes pendant la génération des perturbations qui vont produire les ensembles météorologiques, soit des modifications induites lors du passage par le modèle hydrologique de prévision.

Afin d'améliorer ces défauts des prévisions, la solution idéale est de s'attaquer à la racine de ces problèmes. Une telle approche se concentrera alors sur l'élargissement des ensembles (augmentation de leur dispersion) et/ou sur l'amélioration des modèles hydro-météorologiques et leur façon de quantifier et propager les différentes sources d'incertitude qui affectent leurs simulations. Néanmoins, ces démarches sont considérablement chronophages : il faut bien identifier les sources d'incertitude, les quantifier de façon appropriée et s'assurer de la robustesse de la démarche de propagation des incertitudes. Une alternative plus rapide et efficace est de faire appel aux techniques de correction statistique des biais avant (pré-) ou après (post-) le modèle hydrologique, avec pour but d'améliorer la qualité des prévisions en termes de volume, d'évaluation temporelle et de distribution statistique. L'objectif final se résume à obtenir des prévisions fiables et précises en même temps.

Un nombre considérable d'approches de correction statistique de biais existent dans la littérature. Nous pouvons séparer ces approches en deux principales familles :

- les approches qui visent à corriger l'erreur totale de la prévision (Krzysztofowicz, 2002 ; Roulston et Smith, 2003 ; Gneiting *et al.*, 2005 ; Raftery *et al.*, 2005 ; Wang et Bishop, 2005 ; Fortin *et al.*, 2006 ; Seo *et al.*, 2006 ; Sloughter *et al*., 2007 ; Duan *et al.*, 2007 ; Hamill *et al.*, 2008 ; Vrugt *et al.*, 2008 ; Regianni *et al.*, 2009 ; WIlks, 2009 ; Bao *et al.*, 2010 ; Brown et Seo, 2010, 2012 ; Weerts *et al.*, 2011 ; Roulin et Vannitsem, 2012) ;
- les approches qui traitent chaque source d'erreur séparément : conditions initiales, structure du modèle, paramétrage, etc. (Beven et Binley, 1992 ; Kuczera et Parent, 1998 ; Seo *et al.*, 2003 ; Vrugt *et al.*, 2003 ; Goswani *et al.*, 2005 ; Pappenberger *et al.*, 2005 ; Andreatis et Lettenmaier, 2006 ; Bogner et Kalas, 2008 ; Zhao *et al.* ,2011 ; Pagano *et al.*, 2012 ; Zalachori *et al.*, 2012).

Dans cette thèse nous nous sommes principalement intéressés par les deux familles, mais en ciblant particulièrement les approches de correction de biais de caractère empirique. Ce choix se justifie par le contexte général de la thèse qui se situe dans le cadre des applications opérationnelles de la prévision d'ensemble. Grâce à leur simplicité, les approches empiriques sont considérées comme des approches simplifiées d'agrégation des erreurs et elles sont souvent préférées par les services opérationnels. Dans la plupart des cas, il s'agit d'ajustements des quantiles ou d'habillage des membres (Hashino *et al.*, 2006 ; Seo *et al.*, 2006 ; Wood et Schaake, 2008 ; Olsson et Lindström, 2008 ; Kang *et al.*, 2010 ; Weerts *et al.,* 2011 ; Pagano *et al.*, 2012 ; Boucher *et al.*, 2012 ; Zalachori *et al.*, 2012).

Etant donné que les approches de traitement statistique constituent actuellement un sujet clé de la recherche sur les prévisions hydrologiques d'ensemble, la communauté scientifique se trouve à présent en plein débat sur les questions soulevées. Lors des derniers colloques HEPEX (Delft, 2011 et Beijing, 2012), quelques défis ont été identifiés :

I. Comment évaluer les différentes approches de traitement statistique existantes sur une base comparable de calage/ validation ?
II. Comment évaluer ces approches en utilisant des critères multiples d'évaluation de façon à faire ressortir leurs points forts et leurs points faibles ?
III. Comment combiner les techniques d'assimilation de données et les approches de correction de biais appliquées en amont du modèle hydrologique (prétraitement) de manière efficace ?
IV. Comment les approches de traitement statistique peuvent-elles contribuer à l'amélioration de prévisions probabilistes des événements extrêmes rares ?
V. Comment évaluer ces approches en vue du gain qu'elles apportent à leur utilisateur final dans un contexte de prise de décision ?

VI. Comment guider les utilisateurs vers des approches de post-traitement les plus adaptées à leurs besoins ?
VII. Comment éviter que les approches de traitement statistique ne deviennent des boîtes-noires pour les utilisateurs finaux ? Comment impliquer ces derniers dans la conception de ces approches ?

Parmi ces nombreux défis, cette thèse s'intéressera aux questions suivantes :

- L'utilisation d'une approche multi-modèle peut-elle améliorer la qualité des prévisions ?
- L'utilisation des approches empiriques de correction de biais améliore-t-elle la qualité des prévisions hydrologiques d'ensemble ? Existe-t-il dans ce domaine une méthode plus performante que les autres ?
- Un prétraitement est-il essentiel ou un post-traitement est-il préférable?
- La longueur de calage influence fortement la performance des approches de correction statistique de biais ?

1.4.2 L'utilité des prévisions

L'intérêt de maximiser l'utilité d'une prévision est lié à la recherche d'une prise de décision optimale, à moindre coût et un maximum de bénéfice pour l'utilisateur des prévisions. L'utilité d'une prévision est étroitement liée au concept de la valeur économique de cette prévision. Afin de mesurer et optimiser la valeur économique d'une prévision, il est primordial de recourir à des modèles de décision et des techniques d'optimisation, permettant de lier les événements prévus à ceux observés, les actions des décideurs à leurs conséquences. Ces modèles peuvent être basés sur des objectifs divers : par exemple, minimiser les fausses alertes pour la prévision des crues, maximiser la production d'énergie d'une réserve en eau, minimiser les coûts d'une mesure de protection, etc.

En météorologie, le concept de valeur économique (utilité) des prévisions a été très étudié dans les années 1970-80, suivi par une réapparition de l'intérêt porté à ce sujet à la fin des années 1990. En hydrologie, des études sur l'évaluation économique des prévisions de débits sont plus rares (Roulin, 2007 ; Kim *et al.*, 2007 ; McCollor et Stull, 2008 ; Van de Bergh et Roulin, 2010 ; Muluye, 2011 ; Verkade et Werner, 2011 ; Weijs, 2011 ; Boucher *et al.*, 2012), même si l'utilisation des prévisions d'ensemble pour la prévision des crues et la gestion des ressources en eau est déjà présente depuis une dizaine d'années (Cloke et Pappenberger, 2009). En effet, les études portant sur la qualité des prévisions hydrologiques sont beaucoup plus courantes que celles dédiées à l'utilité (valeur) des prévisions.

Les études déjà réalisées laissent percevoir la valeur économique potentielle des prévisions probabilistes et encouragent les recherches futures sur ce sujet. Néanmoins, elles se limitent à des événements particuliers ou des séries de prévisions-observation courtes, ou encore des aménagements de gestion à caractéristiques fixes. Les forts intérêts économiques liés à la gestion des ressources en eau imposent la réalisation d'études plus approfondies, ainsi que de nouveaux développements pour mieux comprendre comment l'utilisateur des prévisions peut en faire le meilleur usage dans la prise de décision.

Dans ce contexte, nous proposons dans cette thèse notre propre approche d'optimisation de la gestion d'un réservoir. Plus particulièrement, notre objectif est de construire un modèle de gestion qui permette d'examiner l'apport des prévisions, et notamment des prévisions probabilistes d'ensemble, à la gestion optimale d'une réserve en eau superficielle destinée à la production hydroélectrique.

Plus particulièrement, cette thèse s'intéressera aux questions suivantes :

- Quel est le gain apporté en valeur économique par les prévisions d'ensemble par rapport aux prévisions déterministes et climatologiques à la production hydroélectrique ?
- Quelle est la sensibilité du gain économique aux caractéristiques du réservoir ?
- Quel est l'impact de l'amélioration de la qualité des prévisions par des approches de traitement statistiques de biais sur le gain économique ?

1.5 Objectifs de la thèse et structure du mémoire

Cette thèse a deux grands objectifs généraux :

- évaluer des approches empiriques sélectionnées pour garantir une bonne qualité des prévisions pour les applications opérationnelles de la prévision hydrologique d'ensemble,
- développer une méthode simple pour pouvoir estimer la valeur d'une prévision d'ensemble pour le cas de la gestion des apports en eau aux retenues de production hydroélectrique.

Ce mémoire s'organise autour de sept chapitres, décrits sommairement ci-après.

Le *Chapitre 2* présente les données et modèles hydrologiques utilisés : les prévisions météorologiques, les jeux de bassins versants qui ont servi de base à cette étude et leurs séries historiques de données observées. Ces données sont utilisées en entrée de la modélisation hydrologique, effectuée par deux modèles pluie-débit globaux : le modèle *GRP* et le modèle *MORDOR*.

Le *Chapitre 3* examine la qualité des prévisions d'ensemble et présente ses principaux attributs. Nous présentons les différentes mesures de qualité utilisées lors de cette thèse et les principales sources de biais de la prévision d'ensemble. Les différentes approches de pré/post-traitement statistique existantes et celles exploitées lors de cette thèse pour repondre à des questions liées à l'amélioration de la qualité des prévisions sont aussi présentées.

Le *Chapitre 4* présente une autre caractéristique des prévisions d'ensemble : l'utilité, qui est souvent mesurée en termes monétaires (gain ou pertes). Un résumé des techniques les plus courantes est présenté ici, accompagné d'une synthèse des études menées en hydrologie pour évaluer la valeur économique des prévisions.

Dans le *Chapitre 5* nous évaluons la qualité des prévisions tirées de la base TIGGE en utilisant divers critères existants dans la littérature. Nous nous sommes concentrés sur la qualité globale des prévisions, ainsi que sur les attributs de fiabilité et discrimination. Nous évaluons les prévisions d'ensemble de pluie et de débit séparément en examinant leurs performances en fonction de l'horizon de prévision, la surface et la localisation du bassin versant. Différentes approches de combinaison des prévisions de la base TIGGE sont également examinées dans ce chapitre.

Dans le *Chapitre 6* nous utilisons les prévisions météorologiques d'ensemble du centre européen (CEPMMT) et les deux modèles hydrologiques à notre disposition (*MORDOR* et *GRP*) pour examiner différentes approches de traitement statistique appliquées soit sur les prévisions de pluie (pré-), soit sur les prévisions de débit (post-traitement).

Dans le *Chapitre 7* nous nous intéressons à évaluer le gain en termes d'impact économique, apporté par les prévisions, y compris par les prévisions d'ensemble, sur la gestion de la production d'énergie hydroélectrique d'une retenue. Le but est de quantifier la valeur d'une prévision pour la gestion des apports en eau aux barrages-réservoirs. Nous rappelons d'abord les concepts autour de la gestion des apports aux aménagements hydroélectriques. Ensuite, nous décrivons les étapes de construction d'un modèle de gestion heuristique des apports en débit. Enfin, les résultats de l'application du modèle de gestion aux prévisions hydrologiques du modèle *MORDOR* sont présentés et discutés.

Le *Chapitre 8* présente les principales conclusions et les perspectives qui découlent de ce travail de thèse.

"Models are lies that help us see the truth."

Chaitin, after Picasso

2. Données et chaînes de modélisation hydrologique

Dans ce chapitre, nous présentons les bases de données utilisées : les prévisions météorologiques, les jeux de bassins versants qui ont servi de base à cette étude et leurs séries historiques de données observées. Ces données servent comme données d'entrée pour la modélisation hydrologique, effectuée par deux modèles pluie-débit globaux : le modèle GRP et le modèle MORDOR. Les fonctionnalités de ces modèles de prévision hydrologique sont également exposées ici.

2.1 Introduction

Avant de commencer toute opération liée aux prévisions, il est essentiel de bien connaitre les bases de données et les outils de modélisation disponibles. Les travaux de cette thèse utilisent :

- les prévisions météorologiques d'ensemble du Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT ou ECMWF en anglais) ;
- les prévisions météorologiques d'ensemble de la base TIGGE, mise en place dans le cadre du programme THORPEX de l'Organisation Mondial de la Météorologie ;
- les prévisions par analogie, issues de la méthode de recherche de situations analogues utilisée dans les centres de prévision opérationnelle d'EDF.

L'application de ces prévisions est effectuée sur deux jeux de bassins versants :

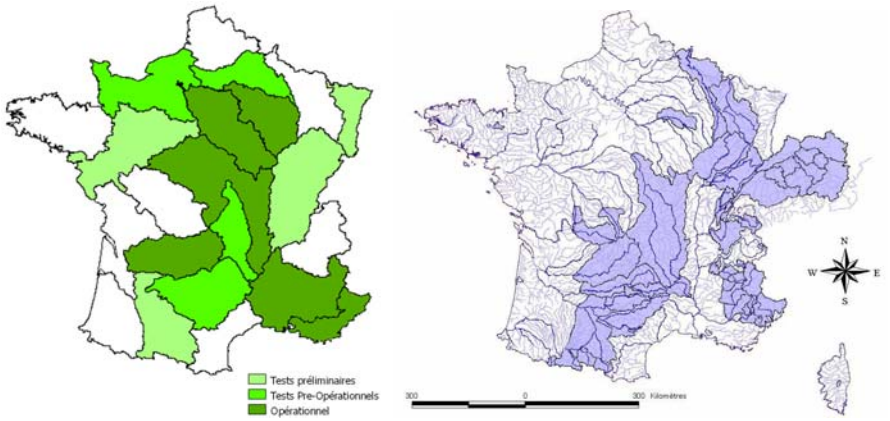
- un échantillon de 74 grands bassins versants en France,
- un échantillon de 11 bassins versants français de fort intérêt pour la production hydroélectrique.

Sur l'ensemble des bassins versants étudiés, nous disposons de données historiques de précipitation et température moyennes de bassin issues de la ré-analyse SAFRAN de Météo-France et des observations de débits journaliers issues de la banque HYDRO du Ministère en charge de l'environnement.

Deux modèles de prévision hydrologique sont ici utilisés au pas de temps journalier :

- le modèle *GRP*, développé par Irstea depuis une dizaine d'années et actuellement utilisé opérationnellement en France par 14 Services de Prévision de Crues (SPC) (Figure 4, à gauche),
- le modèle *MORDOR* développé à EDF et également utilisé en prévision opérationnelle dans ses centres de prévision (Figure 4, à droite).

Dans les paragraphes qui suivent, nous décrivons plus en détail les prévisions météorologiques, les bassins versants et les modèles hydrologiques qui composent la base des travaux de cette thèse.



(a) (b)

Figure 4 : Carte des régions où : (a) le modèle GRP est utilisé en opérationnel/ pré-opérationnel ou des tests préliminaires sont effectués par les services de prévision de crue SPC (source : SCHAPI) et (b) le modèle MORDOR est utilisé en opérationnel (source : EDF-DTG).

2.2 Prévisions météorologiques

2.2.1 Prévisions d'ensemble du CEPMMT

Le CEPMMT est une organisation intergouvernementale soutenue par 34 états membres (18 états européens) ayant pour objectif principal l'émission de prévisions météorologiques à moyen terme sur l'Europe et l'avancement de la recherche scientifique dans le domaine. Les prévisions sont faites par des modèles numériques à partir de données tirées des observations atmosphériques ou au sol. Les principaux types de prévisions émises sont :

- les prévisions de temps à moyen terme jusqu'à 15 jours en avance ;
- les prévisions mensuelles, émises toutes les semaines, allant jusqu'à 30 jours d'échéance ;
- les prévisions saisonnières avec un horizon de 12 mois.

Ces produits alimentent ensuite les services météorologiques nationaux, où ils servent principalement comme un complément aux prévisions à plus court terme.

Le modèle global du CEPMMT émet une prévision déterministe et des prévisions probabilistes d'ensemble. Les prévisions d'ensemble émises par le système de prévision d'ensemble (EPS, *Ensemble Prediction System*) sont constituées d'une prévision dite "*contrôle*", obtenue à partir de l'état initial issu de l'assimilation des données (condition initiale non-perturbée), et de membres qui correspondent à des trajectoires obtenues à partir des perturbations de cet état initial.

Les perturbations sont construites à partir de la méthode des vecteurs singuliers et choisies afin d'optimiser la croissance d'énergie totale sur un intervalle de 48h (Buizza et al., 2005). Plus récemment, des méthodes d'assimilation d'ensemble ont également été introduites (Buizza, 2010). Ces perturbations sont ensuite introduites dans le modèle numérique de temps, où certains processus physiques sont paramétrés. Les incertitudes liées au modèle numérique sont prises en compte en ajoutant des perturbations stochastiques dans les schémas de paramétrisation physique du modèle. Les trajectoires de l'évolution de l'atmosphère obtenues forment ainsi la prévision d'ensemble EPS.

La prévision d'ensemble du CEPMMT est émise deux fois par jour (à 00h et à 12h UTC) et fait partie de son système opérationnel de prévision depuis 1992. Au fil des années plusieurs modifications ont été apportées au système afin notamment d'améliorer les résolutions horizontales et verticales des prévisions et mieux capturer les phénomènes météorologiques à prévoir. Au début, chaque ensemble était construit avec 32 membres sur un horizon maximal de 10 jours et une résolution spatiale qui s'est améliorée au fil des années (Palmer *et al.*, 1993 ; Molteni *et al.*, 1996). En 2005, les prévisions d'ensemble étaient constituées de 51 membres et émises sur une grille spatiale d'environ 80 [km] x 80 [km]. A partir de septembre 2006, les 51 prévisions d'ensemble sont émises avec une résolution spatiale variable : 32 [km] x 32 [km] jusqu'à 10 jours et 64 [km] x 64 [km] au-delà (Buizza *et al.*, 2007a ; Leutbecher et Palmer, 2007). Par ailleurs, depuis mars 2008 les prévisions d'ensemble à moyen terme sont fusionnées avec les prévisions mensuelles : une fois par semaine, les jeudis, les prévisions d'ensemble émises sont ainsi prolongées jusqu'à 32 jours.

Dans notre étude, nous avons utilisé une archive de prévisions météorologiques du CEPMMT qui s'étend sur la période 2005-2008. Les 50 membres perturbés des prévisions journalières des précipitations totales et de températures émises à 12h ont été utilisés. Les données nous ont été fournies à la résolution spatiale de 0.5° lon x 0.5° lat, jusqu'à un horizon de 10-15 jours selon la période.

La Figure 5 présente un exemple des données CEPMMT utilisées dans cette étude. Les prévisions émises sur la France par les 50 membres de la prévision d'ensemble et la prévision contrôle (la dernière figure à droite) pour le 30 mai 2008, à 7 jours d'avance (c'est-à-dire, prévisions émises le 23 mai 2008), sont représentées. Ce jour là, un événement pluvieux sévère a touché le sud-est de la France. Les bassins versants de la Durance à Serre Ponçon et la Drôme à Saillans ont enregistré respectivement 35,7 [mm] et 21,1 [mm] de cumul de pluie journalier en moyenne sur la surface du bassin. Dans la figure, nous observons que la plupart des membres ont bien localisé la région affectée avec 7 jours d'avance. De plus, un tiers des membres ont bien prévu l'intensité de l'événement.

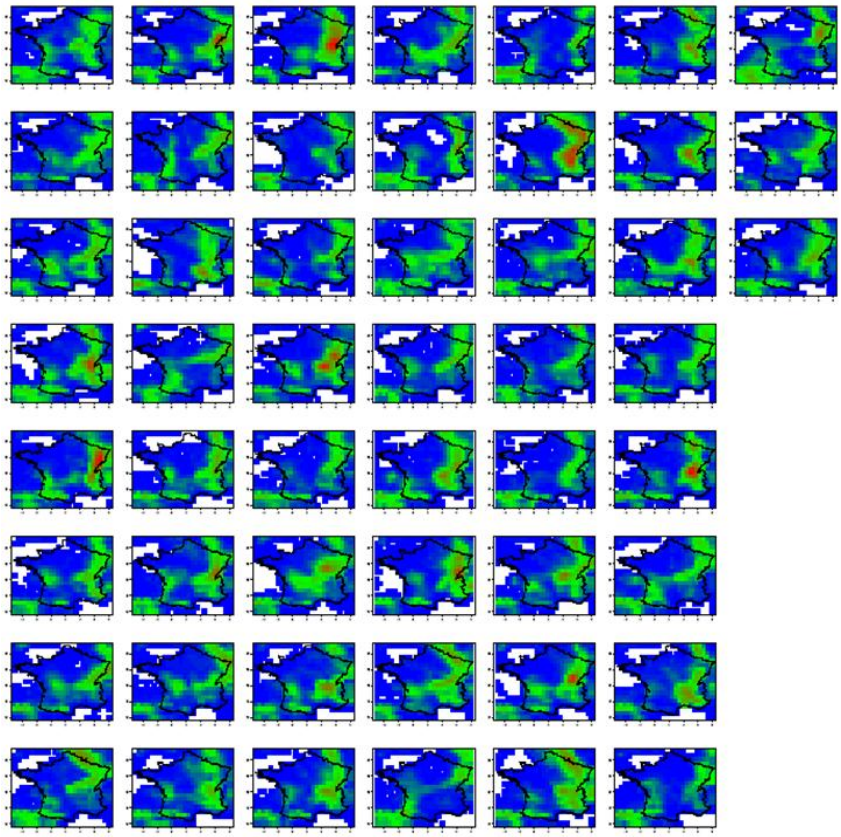


Figure 5 : Cartes de France illustrant les précipitations prévues par les 51 membres de la prévision d'ensemble du centre météorologique européen (CEPMMT) pour le 30 mai 2008 et un horizon de prévision de 7 jours (prévisions émises le 23 mai 2008).

2.2.2 Prévisions d'ensemble de la base TIGGE

La base de prévisions TIGGE fait partie du programme THORPEX de l'Organisation Météorologique Mondiale ("*The Observing System Research and Predictability Experiment*"). Ce programme vise à mieux prévoir le temps et amenuiser ses effets négatifs sur la population. La base TIGGE est constituée de prévisions d'ensemble émises à partir d'octobre 2006, provenant de différents centres météorologiques (Richardson, 2005 ; Park *et al.*, 2008 ; Bougneault *et al.*, 2010). En effet, plusieurs études (ex., Pappenberger *et al.*, 2008, Roulin, 2007, Goswami *et al.*, 2007) ont montré que les prévisions d'ensemble provenant d'un centre météorologique ne prennent en compte que partiellement les incertitudes des conditions initiales. L'idée de base derrière l'initiative TIGGE est que l'utilisation combinée des prévisions fournies par les différents centres, avec leurs différentes méthodes de perturbation et paramétrisation stochastique, permettrait une meilleure description de l'incertitude prédictive (Park *et al.*, 2008).

L'archive mise en place au début des travaux de cette thèse est composée des prévisions de précipitation de 8 centres météorologiques internationaux, couvrant la période d'octobre 2006 jusqu'à septembre 2008 (Figure 6). Les centres météorologiques qui composent notre base de données sont : le centre européen (ci-après *ECMF*), britannique (*EGRR*), japonais (*RJTD*), américain (*KWBC*), chinois (*BABJ*), australien (*AMMC*), canadien (*CWAO*), brésilien (*SBSJ*). Nous notons que, même si nous disposons de données sur environ 2 ans, la longueur de la période pendant laquelle les prévisions d'ensemble sont disponibles varie pour chaque centre météorologique (indiquée sur la Figure 6). Nous observons, par exemple, que le centre brésilien (SBSJ), le dernier à rejoindre la base TIGGE, n'y contribue qu'avec 242 jours.

Le Tableau 1 résume les principales caractéristiques des prévisions TIGGE utilisées dans cette étude. On observe que les prévisions d'ensemble de tous les centres météorologiques combinées composent un grand ensemble d'un total maximum de 232 membres d'ensemble par bassin versant et par jour de prévision. L'horizon maximal des prévisions est de 15 jours.

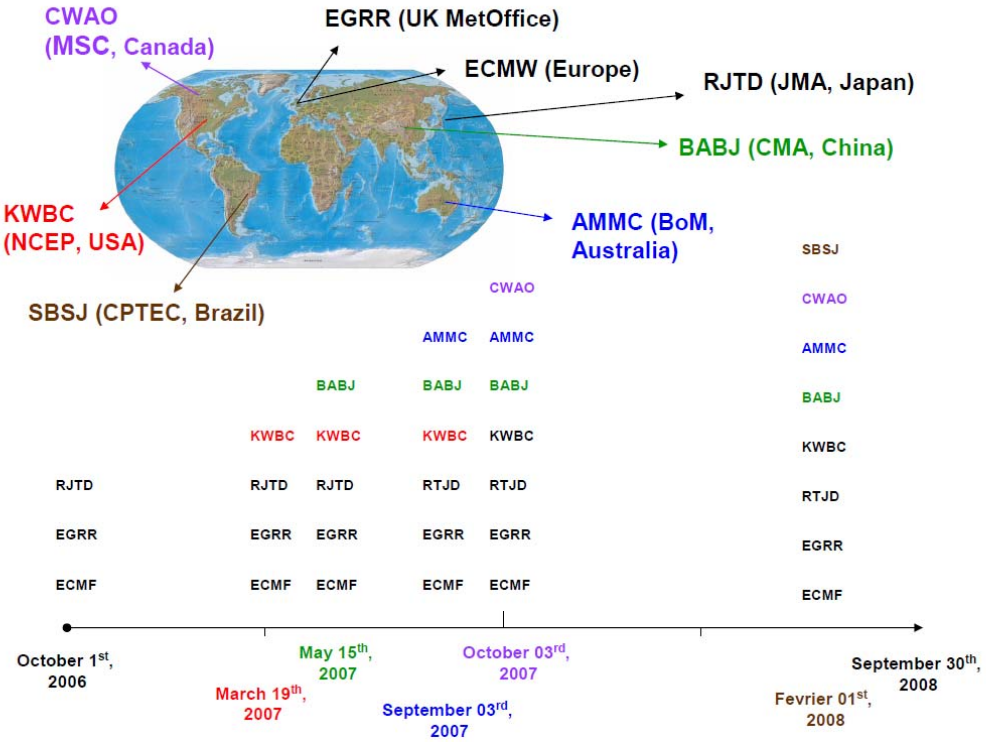


Figure 6 : Centres météorologiques contribuant à la base de prévisions de précipitation TIGGE utilisée dans cette étude, avec indication de la date de début de participation de chaque centre.

Tableau 1 : Caractéristiques principales des centres météorologiques qui forment la base de prévisions de précipitation TIGGE utilisée dans cette étude (d'après : http://tigge.CEPMMT.int/models.html).

	Pays	Nombre de membres	Horizon maximal (jours)	Résolution Spatiale (degrés °)	Méthode de perturbation utilisée	Nombre de jours	Référence
ECMF	Organisation avec 34 Etats Membres (ECMWF)	50	10-15	0.5° x 0.5°	Vecteurs Singulieurs et Filtre d'ensemble Kalman	731	www.CEPMMT.int/
EGRR	Royaume Uni	23	15	0.8333° x 0.5555°	Filtre d'Ensemble Kalman Transformé	731	www.metoffice.gov.uk/
RJTD	Japon	50	9	0.5625° x 0.5625°	Méthode des erreurs croissantes BGM	731	www.jma.go.jp/jma/indexe.html
KWBC	Etats-Unis	20	15	1.0° x 1.0°	Méthode des erreurs croissantes BGM	561	www.ncep.noaa.gov/
BABJ	Chine	14	10	1.0° x 1.0°	Méthode des erreurs croissantes BGM	504	www.cma.gov.cn/cma_new/
AMMC	Australie	32	10	1.5 x 1.5°	Vecteurs Singulieurs	393	www.bom.gov.au/
CWAO	Canada	20	15	~ 1.0° x 1.0°	Filtres d'ensemble Kalman	363	www.weatheroffice.gc.ca/canada_e
SBSJ	Brésil	14	15	0.9375 °x0.9375 °	Fonction orthogonale empirique	242	www.cptec.inpe.br/

2.2.3 Prévisions probabilistes par analogie

Au sein de la Division Technique Générale d'EDF, la notion de prévision probabiliste est présente depuis les années 1950, sous le terme " prévisions probabilistes expertisées " (Garçon *et al.*, 2009). Ce terme signalait la prise en compte de l'expertise humaine pendant le processus de prévision. La première approche probabiliste était basée sur des simples analyses statistiques de pluie et de débit prévus en ajoutant des intervalles de confiance de régression autour de l'estimation initiale. A peu près en même temps (~1953) commence l'effort d'un archivage journalier des champs de pression observés sur chaque bassin versant d'intérêt, conjointement à l'archivage déjà existant des précipitations et températures observées. Dans les années 1970, cette archive comptait déjà vingt ans de données et commençait à devenir opérationnelle. Duband (1970, 1980) a alors formalisé sous forme d'algorithmes la démarche de prévision par analogie déjà utilisée par les prévisionnistes.

La prévision par analogie consiste à mieux comprendre les situations futures à l'aide des situations synoptiques observées dans le passé et mises en relation avec les conditions locales, afin d'obtenir des prévisions de précipitation à partir des prévisions des champs synoptiques. L'hypothèse derrière cette approche est basée sur les travaux de Lorenz (1969), selon lesquels deux situations synoptiques suffisamment similaires doivent conduire à des effets locaux comparables. Nous parlons de prévision probabiliste par analogie, car il ne s'agit pas seulement de trouver la situation la plus ressemblante, mais de construire un sous-ensemble de situations similaires permettant d'obtenir une distribution des scénarios futurs possibles (Zorita et von Storch, 1999 ; Obled *et al.* 2002). Le prévisionniste cherche alors à lier la situation météorologique prévue aux situations analogues observées dans le passé pour expertiser sa prévision.

L'archivage de plusieurs variables comme le champ de pression, l'humidité, etc. est proposé dans la littérature concernant les méthodes par analogie (Zorita et von Storch, 1999 ; Obled *et al.,* 2002 ; Bontron, 2004 ; Ben Daoud, 2009). Dans cette étude, seulement les champs de pression observés à 0h00 et 24h00, à 700hPa et 1000hPa sont utilisés. Les champs prévus proviennent des analyses du modèle *ARPEGE* de Météo-France (Geleyn *et al*., 1994) pour les champs à 00 hTU, ensuite des prévisions *ARPEGE* (champs à 24, 48 et 72 hTU) et enfin des prévisions du modèle du Centre Européen pour les Prévisions Météorologiques à Moyen Terme CEPMMT (Simmons *et al.*, 1989) (pour les échéances de 72 à 168 hTU). Il faut noter que les approches par analogie sont également une forme de descente d'échelle, car elles ajoutent des informations locales, à résolution fine, à partir des sorties des modèles numériques de temps. Ces derniers tournent sur l'ensemble du globe ou sur des fenêtres régionales larges, avec un maillage qui ne leur permet pas de bien capturer la topographie ou les phénomènes locaux.

Le critère utilisé pour trouver le sous-ensemble de situations analogues au champ de géopotentiel est proposé par Teweless et Wobus (1954). Ce critère est calculé entre chacune des situations de l'archive météorologique et la situation de la journée cible sur la base des gradients horizontaux d'altitude pour plusieurs niveaux de pression atmosphérique. Les 50 situations les plus ressemblantes à la journée cible sont identifiées dans l'archive météorologique. Une fois ces 50 situations définies les précipitations associées aux bassins versants de l'étude sont retrouvées. Par conséquent, pour chaque champ de géopotentiel prévu à une échéance donnée, 50 scénarios (valeurs) de précipitation observée par bassin versant, qui forment la prévision probabiliste par analogie pour la journée cible et le bassin versant en question, sont obtenus. La Figure 7 présente schématiquement l'approche de l'analogie pour l'association des valeurs observées du géopotentiel à des situations similaires au passé et leur transformation en précipitation.

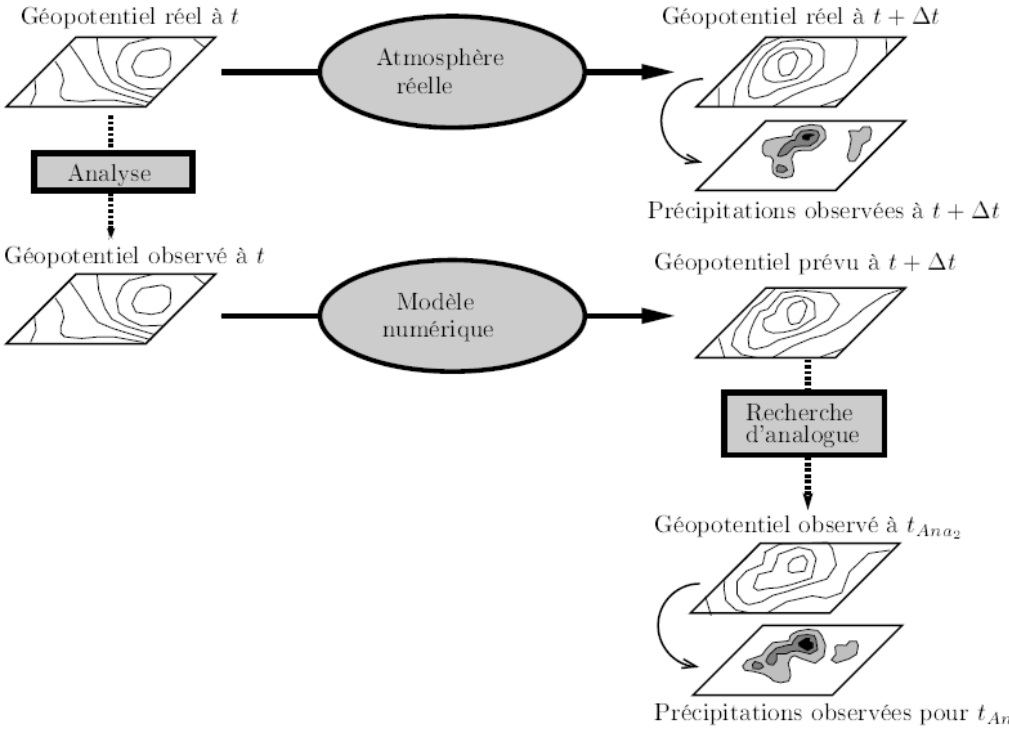


Figure 7 : Description schématique de l'approche de prévisions par analogie (source : Bontron, 2004).

Actuellement l'archive des analogues est bien plus robuste et compte presque 60 ans de données. Dans cette thèse, nous avons utilisé cette archive pour former notre base de prévisions par analogie. Pour chaque échéance et un horizon maximal de 7 jours, les prévisions d'ensemble du CEPMMT ont été utilisées dans la recherche des situations analogues. Cette information sera utilisée pour la correction statistique des biais des prévisions EPS du CEPMMT. La méthodologie adoptée est décrite en détail dans le chapitre 3.

2.3 Bassins versants étudiés

Cette thèse s'appuie sur la base de plus de 3000 bassins versants mise en place dans l'equipe hydrologique d'Irstea à Antony dans le cadre des travaux de thèse passés (Mathevet, 2005 ; Le Moine, 2008), qui était disponible au début de nos travaux. Aujourd'hui cette base, qui regroupe des données météorologiques et hydrologiques, présente déjà près de 3500 bassins versants français. Il s'agit de bassins versants pour lesquels nous disposons de données journalières de pluie et de débit, et qui sont à signification hydrologique et sans influence majeure d'ouvrages hydrauliques selon la Banque de données hydrologiques HYDRO.

Sur chaque bassin versant, les précipitations journalières de bassin sont calculées à partir des données de la ré-analyse SAFRAN de Météo-France (Quintana-Ségui *et al.*, 2008 ; Vidal *et al.*, 2009), disponibles sur la période 1958-2009. Les débits observés ont été extraits de la banque nationale HYDRO. Ils y sont également au pas de temps journalier, avec une longueur des séries historiques variable selon le bassin versant. Enfin, l'information sur l'évapotranspiration potentielle (ETP) est également fournie dans cette base de données. Cette donnée est calculée à partir de la formule d'Oudin *et al.* (2005), qui utilise la température de l'air, le rayonnement extraterrestre, la chaleur latente de vaporisation de l'eau et sa masse volumique dans le calcul de l'évapotranspiration.

Dans notre étude, nous avons cherché à construire deux échantillons de bassins versants, chacun destiné à des buts spécifiques de notre recherche :

- un échantillon de 74 grands bassins versants (surface supérieure à 1000 km²), et
- un échantillon de 11 bassins versants à fort intérêt pour la production hydroélectrique (bassins plutôt de montagne).

Les deux jeux de bassins versants utilisés dans cette etude sont présentés ci-après.

2.3.1 74 grands bassins versants en France

Ce jeu de bassins versants a été mis en place pour répondre à nos questions concernant l'utilisation de l'approche multi-modèle en prévision hydrologique d'ensemble, à partir des prévisions de précipitation de la base TIGGE (Chapitre 5). La sélection de cet échantillon a été alors faite sur la base de trois critères :

- le premier critère a été la surface : les bassins versants devaient avoir une surface supérieure à 1000 km², choisie en fonction de la résolution spatiale des prévisions TIGGE (archive fournie avec une résolution spatiale d'environ 45 [km] x 45 [km] sur la France), une limite de surface plus petite aurait amené à des bassins le plus souvent couverts par une seule maille de la grille des prévisions. D'autre part, une limite de surface plus grande aurait amené à un nombre trop petit de bassins versants, étant donné les échelles d'espace sur le territoire français,

- le deuxième critère a été la disponibilité de données observées : étant donné la courte période dont nous disposons des prévisions TIGGE (2006 à 2008), afin de préserver la robustesse statistique de nos résultats, seuls les bassins avec au moins 95% de données observées disponibles sur cette période ont été retenus. De plus, nous avons également veillé à ce qu'une période antérieure d'au moins 10 années soit également disponible pour le calage du modèle hydrologique utilisé,
- le troisième critère a été la localisation des bassins versants : nous avons cherché à avoir une certaine diversité des bassins versants en termes de caractéristiques physiographiques, climatiques et hydrologiques, et avons évité, dans la mesure du possible, d'avoir beaucoup de bassins versants sur une même rivière ou un tributaire proche.

La localisation des 74 bassins versants sélectionnés est présentée à la Figure 8. La surface moyenne de l'échantillon de bassins versants est d'environ 3 400 km², avec une surface minimale égale à 1 070 km² et maximale égale à 9 387 km² (à l'exception du bassin versant de la Seine à Paris qui ressort de l'échantillon, avec 43 800 km²). Dans l'échantillon, la longueur des séries temporelles des débits varie entre 16 et 51 années, avec une majorité de bassins versants (65 %) disposant d'une période de 40 à 44 années de données complètes.

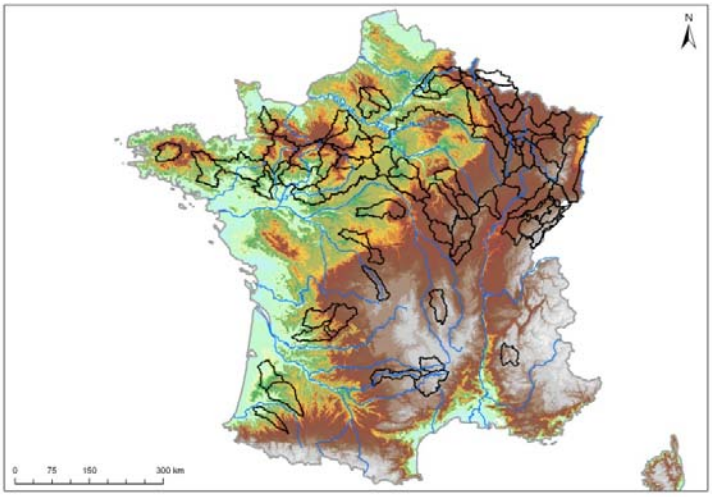


Figure 8 : Localisation du jeu de 74 larges bassins versants français.

Le Tableau 2 présente les principales caractéristiques du jeu de 74 bassins versants étudiés. Ces caractéristiques sont représentées par leur valeur moyenne avec, entre crochets, les valeurs minimales et maximales.

Tableau 2 : Principales caractéristiques des 74 bassins versants étudiés en termes de valeurs moyennes (entre les crochets sont affichées les valeurs minimales et maximales).

Altitude moyenne (m)	Moyenne interannuelle des précipitations (mm/an)	Fraction solide des précipitations (%)	Moyenne interannuelle d'ETP - Oudin (mm/an)	Moyenne interannuelle des débits (mm/an)
313	962	5.10	677	389
[74, 1046]	[638, 1565]	[0.6, 20.4]	[590, 793]	[99, 1304]

Nous remarquons que l'échantillon des bassins sélectionnés inclut surtout des bassins de plaine, peu influencés par la neige. La plupart des bassins présentent également une alimentation pluviale dominante avec une hauteur des précipitations par année assez élevée. Par conséquent, l'influence de la température sur la formation des débits des fontes nivales, ainsi que l'influence de la topographie sont modérées pour la majorité des bassins versants étudiés dans cet échantillon.

Ces conclusions sont confirmées par la classification proposée par Sauquet (2006) concernant les régimes hydrologiques des bassins versants français. Selon cette classification, la grande majorité des bassins versants sélectionnés dans notre échantillon font partie des groupes caractérisés par un régime hydrologique pluvial, fortement influencé par la pluie d'hiver et d'automne, et avec des écoulements faibles pendant les mois d'été. Seul un petit nombre de nos bassins (3 parmi les 74) est dominé par un régime hydrologique plutôt identifié à un régime méditerranéen : avec des événements de précipitations intenses surtout en automne et des étés particulièrement secs, comme il en est pour le Luy à Saint-Pandelon et l'Aveyron à Laguépie. Enfin, nous retrouvons également des régimes hydrologiques de transition, où les variations saisonnières de l'écoulement sont affectées aussi bien par les précipitations que par la température et l'influence topographique, dans un nombre limité de nos bassins versants (6 parmi les 74), comme il en est pour le Lot à Basannac et la Drôme à Saillans, où les influences nivale et pluviale coexistent.

2.3.2 11 bassins versants de fort intérêt pour la production hydro-électrique

Ce deuxième échantillon de bassins versants a été construit dans le but d'étudier la qualité et la valeur des prévisions hydrologiques d'ensemble dans le cas des bassins versants de fort intérêt économique pour la production hydroélectrique en France (Chapitres 6 et 7). Le choix a été basé sur les critères de sélection suivants :

- des bassins situés en amont des retenues d'eau, dont les apports alimentent les lacs-réservoirs des barrages géreés par EDF, ou font figure de témoins importants pour la génération d'électricité,
- des bassins versants de superficies et de régimes hydrologiques variés,
- des bassins versants distribués dans les principaux massifs français : les Alpes, le Massif Central, les Cévennes et le Jura.

La localisation des bassins versants sélectionnés pour cet échantillon est présentée dans la Figure 9.

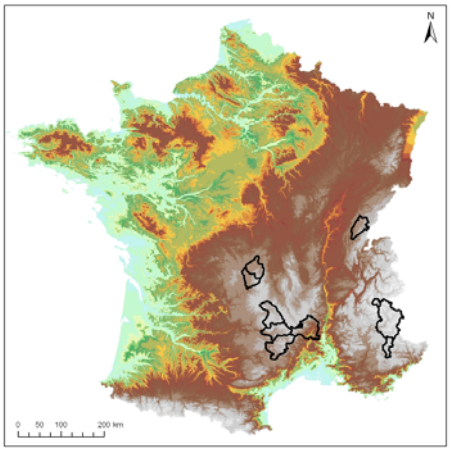


Figure 9 : Localisation du jeu de 11 bassins versants de fort intérêt en hydroélectricité de cette étude.

La plupart des bassins versants examinés ici sont des bassins de montagne avec pour la plupart des fortes influences nivales, surtout pour les bassins alpins. En suivant la classification des régimes hydrologiques de Sauquet (2006), nous notons que ces bassins versants appartiennent aux groupes caractérisés par des écoulements forts liés à la fonte de neige, avec un maximum de débit au printemps et en été. Les bassins versants moins influencés par la neige sont ceux situés dans le Massif Central et le Jura, citons les exemples du Tarn à Millau, de l'Ardèche à Sauze, de la Sioule aux Fades et de la Truyère à Grandval. Ces bassins versants sont caractérisés par des écoulements forts dès la fin d'automne, jusqu' au printemps, ainsi que par des étés particulièrement secs.

Le Tableau 3 présente les principales caractéristiques du jeu de 11 bassins versants étudiés. Les caractéristiques sont représentées par leur valeur moyenne avec, entre crochets, leurs valeurs minimales et maximales. Le Tableau 4 résume également les principales caractéristiques des retenues liées aux 11 bassins versants étudiés.

Tableau 3 : Principales caractéristiques des 11 bassins versants étudiés en termes de valeurs moyennes (entre crochets sont affichées les valeurs minimales et maximales).

Altitude moyenne (m)	**Moyenne interannuelle des précipitations (mm/an)**	**Fraction solide des précipitations (%)**	**Moyenne interannuelle d'ETP - Oudin (mm/an)**	**Moyenne interannuelle des débits (mm/an)**
1170	1257	19.9	673	775
[490, 2362]	[955, 1838]	[5, 61]	[570, 826]	[430, 1352]

Tableau 4 : Principales caractéristiques des retenues liées aux 11 bassins versants de fort intérêt en hydroélectricité de cette étude.

Bassin Versant	Surface (km^2)	Localisation dans les massifs montagneux français	Volume utile de la retenue (hm^3)	Vocation	Puissance hydroélectrique installée (MW)
Durance à Serre Ponçon	3580	Alpes	1030	Hydroélectricité et Irrigation	380
Borne au Pont Nicoulaud	63	Cévennes	35	Hydroélectricité	100
Verdon à Castillon	656	Alpes	113	Hydroélectricité et Irrigation	50
Ain à Vouglans	1195	Jura	419	Hydroélectricité	262
Truyère à Grandval	1795	Massif Central	225	Hydroélectricité	68
Dordogne à Bort	1005	Massif Central	408	Hydroélectricité et Tourisme	240
Sioule aux Fades	1300	Massif Central	56	Hydroélectricité	29
Tarn à Millau	2143	Massif Central	4	Hydroélectricité	40
Lot à Castelnau	1158	Massif Central	20	Hydroélectricité	41
Ardèche à Sauze	2261	Massif Central	-	-	-
Romanche aux Chambons	223	Alpes	50,8	Hydroélectricité	116

2.4 Modélisation hydrologique

Une fois définies les prévisions météorologiques qui seront utilisées dans le système de prévision hydrologique d'ensemble, les données disponibles (précipitations, débits et températures observés) et les bassins versants qui seront étudiés, il reste à choisir le modèle hydrologique qui permettra de transformer les précipitations en débit et effectuer le calage de ses paramètres libres.

Dans cette étude, deux modèles hydrologiques globaux sont utilisés : le modèle *GRP* dans sa version sans module de prise en compte de stockage et fonte de la neige (Tangara, 2005) et avec module de prise en compte de la neige (Valery, 2010 ; Nicolle *et al.*, 2011), et le modèle *MORDOR* (Garçon, 1999 et Paquet, 2004). D'une part, le modèle *GRP* sans module neige est utilisé pour la modélisation du jeu de 74 grands bassins versants français peu influencés par la neige. D'autre part, le modèle *GRP* avec module neige et le modèle *MORDOR* sont utilisés pour la modélisation du jeu de 11 bassins versants situés dans les principaux massifs montagneux français.

Dans un premier temps, nous présentons les principales caractéristiques de chaque modèle. Ensuite, nous présentons la procédure et les résultats du calage effectué sur les deux jeux de bassins versants de notre étude.

2.4.1 Modèle hydrologique GRP

Le modèle de prévision *GRP*, développé par l'équipe Hydrologie d'Irstea à Antony, est un modèle global à 3 paramètres, dérivé du modèle *GR4J* (Perrin, 2000) et adapté pour la prévision des crues (Tangara, 2005). Dans cette étude, le modèle *GRP* est utilisé au pas de temps journalier, dans sa version adaptée aux prévisions d'ensemble (Ramos *et al.*, 2008). Pendant son développement, le modèle a été testé sur plus d'un millier de bassins versants français, ce qui lui permet de s'adapter à une grande variété de conditions hydroclimatiques, géologiques et d'occupation du sol.

GRP modélise le bassin versant comme un assemblage de réservoirs représentant les stocks d'eau dans les sols et les nappes et leur transfert vers l'exutoire du bassin. La structure du modèle repose sur un réservoir de production qui permet d'assurer un suivi des conditions d'humidité du bassin, une fonction d'ajustement de la pluie efficace, un hydrogramme unitaire et un réservoir de routage quadratique (Figure 10).

Le premier paramètre du modèle représente la capacité à un jour du réservoir de routage (X1, en mm), le deuxième représente le facteur d'ajustement de la pluie (X2, paramètre adimensionnel supérieur ou inférieur à 1) et le troisième correspond au temps de base de l'hydrogramme unitaire (X3, en jours). La mise à jour du modèle est faite au niveau du réservoir de routage : son niveau est recalculé de manière à pouvoir générer un débit exactement égal au dernier débit observé à l'instant de la prévision (Tangara, 2005). Berthet (2010) a comparé la performance de l'approche de mise à jour à pas de temps horaire utilisée avec d'autres méthodes existantes (filtre de Kalman, réseaux neurones, etc) dans la littérature. Son étude a montré que :

- il y a un intérêt à appliquer la mise à jour au niveau du réservoir de routage, car le modèle est peu sensible aux mises à jour des autres états ;
- la mise à jour ajustant le niveau du réservoir de routage avec le dernier débit observé présente une meilleure performance par rapport aux autres approches examinées, notamment pour les plus longues échéances.

Les données d'entrée du modèle *GRP* sont les précipitations et l'évapotranspiration potentielle (ETP), tandis que la mise à jour a besoin du dernier débit observé pour être effective.

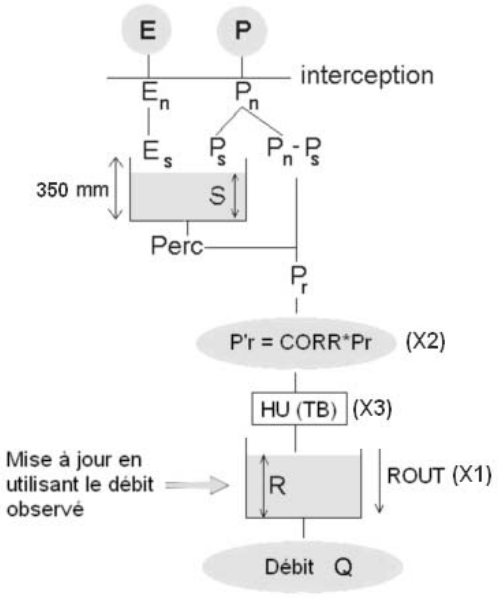


Figure 10 : Schématique de la structure du modèle GRP sans module neige (Tangara, 2005)

Récemment un module de stockage et fonte de la neige a été développé pour la famille de modèles *GR*. Ce module neige, appelé Cemaneige a été développé par Valery (2010), sur la base du modèle *GR4J* et intégré au modèle *GRP* par Nicolle (2010, 2012). L'intégration d'un module neige apporte deux paramètres supplémentaires au modèle : le facteur de fonte (*Kf*) et le coefficient de pondération de l'état thermique du manteau (CT_G).

La Figure 11 présente une schématique de l'intégration du module neige Cemaneige dans la structure du modèle *GRP*. Le module est inséré en amont du modèle hydrologique. La séparation de la pluie en liquide ou solide est basée sur les températures minimales et maximales, ou, à défaut, sur la température moyenne journalière. Le modèle a ainsi besoin, en plus des précipitations et de l'évapotranspiration potentielle, de la température de l'air en entrée. La neige est stockée au sein d'un réservoir conceptuel qui représente le manteau neigeux. Ce réservoir est alimenté uniquement par la fraction solide des précipitations. Le manteau neigeux présente un état thermique qui permet de retarder ou déclencher la fonte. Une fonction de transfert permet ensuite de calculer la lame de fonte écoulée, issue du stock de neige, qui va être intégrée au modèle hydrologique, en plus de la fraction liquide des précipitations (Nicolle, 2010).

Le logiciel *GRP* est constitué de deux utilitaires :

- l'utilitaire de calage, qui permet, à partir des chroniques des données historiques, de déterminer ses paramètres et d'évaluer les performances du modèle ;
- l'utilitaire de prévision, permettant de calculer les prévisions de débit.

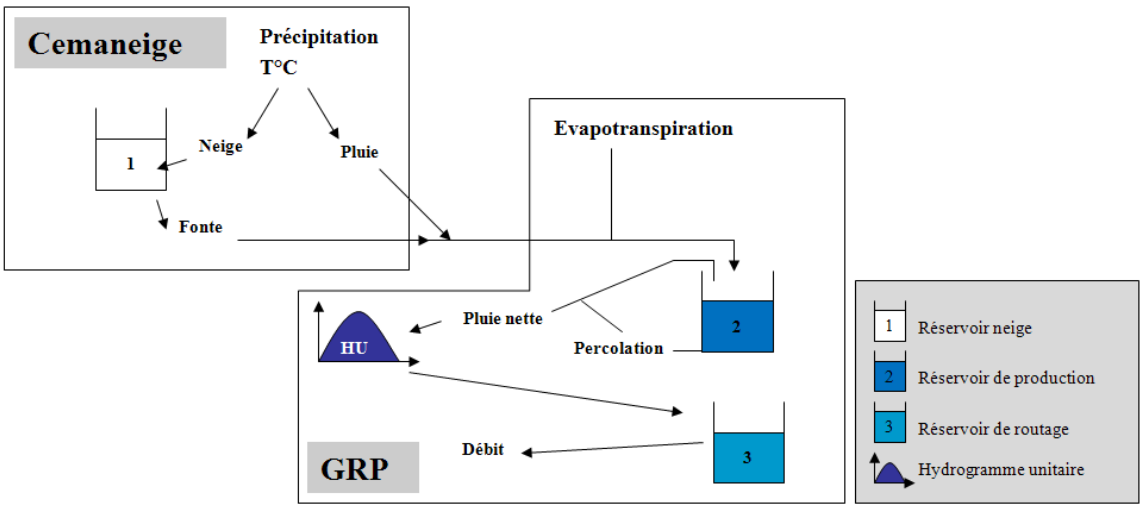


Figure 11 : Schématique de l'intégration du module neige Cemaneige dans la structure du modèle GRP (Valéry, 2010 ; Nicolle et al., 2012).

Les détails du fonctionnement des modèles *GRP* sans et avec module neige est présentée dans l'ANNEXE A.

2.4.2 Modèle hydrologique MORDOR

Le modèle hydrologique MORDOR (abréviation pour : MOdèle à Réservoirs de Détermination Objective du Ruissellement) est un modèle conceptuel global développé à EDF (Garçon, 1996 &1999 ; Paquet, 2004). Les variables de forçage du modèle MORDOR sont la température de l'air et la précipitation totale journalière. La Figure 12 présente une représentation schématique simplifiée du modèle MORDOR. Pour la représentation des processus physiques, 4 réservoirs sont utilisés :

- le réservoir "*Z*" : représente les prélèvements qui ne contribuent pas aux écoulements ;
- le réservoir de surface "*U*" : représente l'état hydrique superficiel. Il permet la prise en compte de l'évaporation directe et des effets de débordement rapide ;
- le réservoir "*L*" : simule la percolation sub-superficielle et permet également un débordement semi-rapide ;
- le réservoir profond "*N*" à capacité constante : représente les processus de percolation profonde au sein du bassin versant.

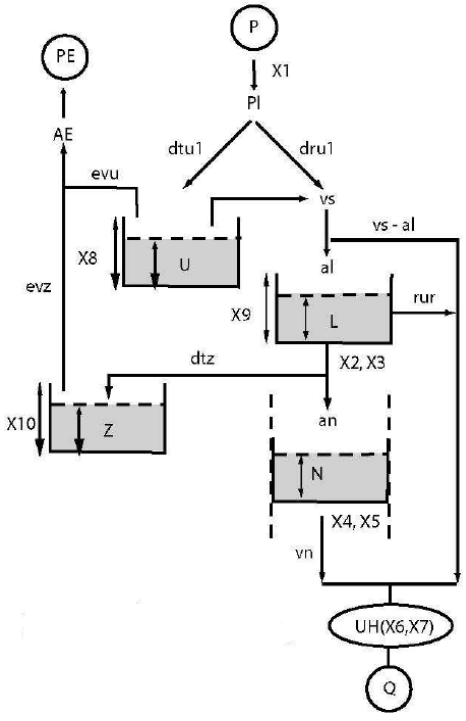


Figure 12: Schématique du fonctionnement du modèle hydrologique MORDOR (source : Paquet, 2004).

La version du modèle MORDOR utilisée dans cette thèse présente au total 10 paramètres libres à caler : le coefficient de correction de pluie (X1), les 3 capacités de réservoirs (X8, X9, X10), les 4 paramètres pour la vidange des réservoirs L et N (respectivement : X2, X3 et X4, X5), ainsi que les 2 paramètres pour fixer la forme de l'hydrogramme unitaire (X6, X7). A noter que la correction de la précipitation en entrée du modèle permet un premier débiaisage des pluies. La capacité du réservoir profond N est considérée constante.

Le modèle MORDOR a également un module neige intégré dans sa structure. Ce module fonctionne autour du réservoir "*NEIGE*", qui permet la gestion du stock et de la fonte de neige. Dans ce module, une partie de la lame de fonte est directement dirigée vers le ruissellement tandis qu'une autre partie est ajoutée soit au réservoir de surface *U*, soit au réservoir profond *N*. La répartition se fait à l'aide d'un paramètre à optimiser lors de la phase de calage.

Les détails du fonctionnement du modèle MORDOR, ainsi que de son module neige sont présentés dans l'ANNEXE A.

2.4.3 Calage des modèles hydrologiques

Pour chaque bassin versant, les modèles hydrologiques ont été calés à partir des données observées disponibles sur la période du 01 août 1958 au 10 mars 2005. Nous utilisons l'ensemble de données disponibles avant le début de la période de prévision qui servira de validation des approches étudiées en prévision d'ensemble. Nous présentons ici les performances en calage. La performance des modèles en validation, c'est-à-dire, pendant la période de prévision est au cœur de cette thèse et sera présentée dans les chapitres 5 et 6, où la qualité des prévisions émises par les modèles est examinée.

L'objectif d'une procédure de calage d'un modèle hydrologique consiste à optimiser les paramètres libres du modèle en s'appuyant sur des données connues, comme le débit à l'exutoire. La fonction objectif utilisée pour l'optimisation des paramètres libres vise à minimiser la distance entre l'observation (valeur connue) et la simulation (valeur calculée). En résultat, il s'agit d'identifier le jeu de paramètres qui minimise l'écart entre l'observation et la simulation en optimisant la fonction objectif.

Les modèles hydrologiques *GRP* et *MORDOR* ont des fonctions objectifs déjà codées au préalable dans leur module de calage automatique. Nous avons utilisé ces fonctions pour caler les modèles. La performance en calage est ainsi évaluée par :

- le critère d'efficacité de Nash et Sutcliffe (1970) pour les deux modèles *GRP* et *MORDOR*,
- le critère de *C2MP*, introduit par Mathevet (2005) afin de donner au critère de persistance (Kitanidis et Bras, 1980) une borne négative à -1, pour le modèle *GRP*,
- le critère de *Kling-Gupta* (Gupta *et al.*, 2009) pour le modèle *MORDOR*.

Le critère d'efficacité de *Nash et Sutcliffe* (1970) est défini comme la part de variance des débits expliquée par le modèle. Il varie dans l'intervalle $[-\infty,1]$ et estime sur *N* jours la performance du modèle par rapport à un modèle de référence qui donnerait comme débit simulé Q^{sim} le débit moyen observé $\overline{Q}^{obs}$:

$$NASH = 1 - \frac{\sum_{i=1}^{N}(Q_i^{sim} - Q_i^{obs})^2}{\sum_{i=1}^{N}(Q_i^{obs} - \overline{Q}^{obs})^2}) \qquad \text{Eq. 1}$$

Ainsi, si le critère de *Nash-Sutcliffe* est négatif, le modèle utilisé est plus mauvais que le modèle de référence ; si le critère est positif, il est meilleur.

Le critère de persistance (*PERS*) a été introduit pour permettre de comparer la performance du modèle à un modèle naïf qui donnerait comme débit simulé le dernier débit observé. Ce critère est aussi borné aux valeurs positives et varie dans l'intervalle $[-\infty,1]$:

$$PERS = 1 - \frac{\sum_{i=1}^{N}(Q_i^{sim} - Q_i^{obs})^2}{\sum_{i=1}^{N}(Q_i^{obs} - Q_{i-1}^{obs})^2} \qquad \text{Eq. 2}$$

Comme le critère de *Nash-Sutcliffe*, le critère de persistance est marqué par une absence de borne négative. Le critère de *C2MP* a été introduit par Mathevet (2005) afin de donner au critère de persistance une borne négative à -1. Le critère *C2MP* varie dans l'intervalle [-1, 1], ou [-100, 100] quand ceci est donné en pourcentage :

$$C2MP = \frac{PERS}{2 - PERS} \qquad \text{Eq. 3}$$

Le critère de *Kling-Gupta* (Gupta *et al.*, 2009), quant à lui, est une modification du critère de *Nash-Sutcliffe* qui améliore la performance des modèles pluie-débit en simulation. Il s'agit d'une décomposition du critère de *Nash-Sutcliffe* en trois termes :

$$NASH = -\beta + \alpha^2 + 2r\alpha \quad \text{Eq. 4}$$

, ou :

$$\beta = \frac{(\overline{Q}^{obs} - \overline{Q}^{sim})^2}{(\sigma^{obs})^2} \quad \text{Eq. 5}$$

$$\alpha = \frac{\sigma^{sim}}{\sigma^{obs}} \quad \text{Eq. 6}$$

$$r = \frac{1}{n}\sum_{i=1}^{N} \frac{(Q_i^{obs} - \overline{Q}^{obs}) * (Q_i^{sim} - \overline{Q}^{sim})}{\sigma^{sim}\sigma^{obs}} \quad \text{Eq. 7}$$

La composante *β* est équivalente à l'erreur du bilan normée entre les moyennes des débits simulés et observés, la composante *α* représente le rapport des variances des debits simulés et observés et la composante *r* est le coefficient de corrélation entre débits simulés et observés.

Le critère de *Kling-Gupta* (*KGE*) est un critère de type distance euclidienne entre les valeurs prises par ces trois composantes et leurs valeurs optimales. Plus particulièrement le critère de *Kling-Gupta* est la différence entre 1 et cette distance euclidienne. En combinant ces termes, l'expression mathématique du critère de *Kling-Gupta* est :

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (\frac{Q^{sim}}{Q^{obs}} - 1)^2} \quad \text{Eq. 8}$$

En calage, l'objectif étant d'atteindre les valeurs optimales des composantes, la distance euclidienne doit tendre vers zéro et maximiser ainsi la valeur du critère *KGE* (Eq. 8).

Nous rappellons que le modèle *GRP* a été calé sur les deux jeux de bassins versants de cette étude : les 74 grands bassins versants (modèle *GRP* sans module neige) et les 11 bassins versants à intérêt en hydroélectricité (modèle *GRP* avec module neige) et que le modèle *MORDOR* n'a été calé que sur ce dernier jeu de bassins versants.

Les Figure 13 et Figure 14 présentent les performances obtenues en calage selon le critère de *NASH* et *C2MP*, respectivement, pour le modèle GRP. La Figure 15 présente les performances obtenues en calage selon le critère de *NASH* et *Kling-Gupta* pour le modèle *MORDOR*.

Dans la Figure 13 nous observons que le modèle GRP montre de bonnes performances pour la majorité des bassins versants étudiés en termes de critère de *NASH*. Le modèle *GRP* obtient des performances plus faibles seulement sur quelques bassins versants situés en Alsace-Lorraine et dans les Cévennes. Cependant les valeurs du score de *NASH* sur ces bassins versants restent supérieures à 80%.

En ce qui concerne le score C2MP (Figure 14) les valeurs varient entre 15 et 70 pendant la période de calage. Rappelons que les valeurs les plus élevées sont souhaitées. Nous remarquons également que les bassins versants présentant les plus faibles performances (*C2MP* < 30) se situent dans le Nord du pays (bassins de la Seine aval et de l'Eure), ainsi que dans les Alpes et le Massif Central. Nous ne remarquons pas une tendance claire entre la surface du bassin versant et la valeur du critère *C2MP*.

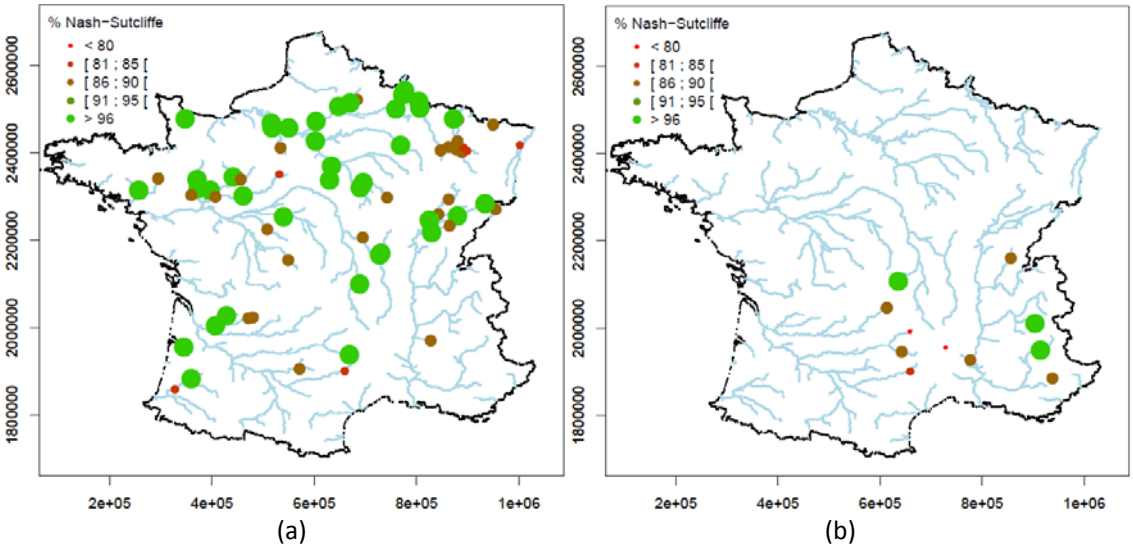


Figure 13 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) le module Cemaneige selon le critère de NASH et pour l'échantillon de : (a) 74 et (b) 11 bassins versants.

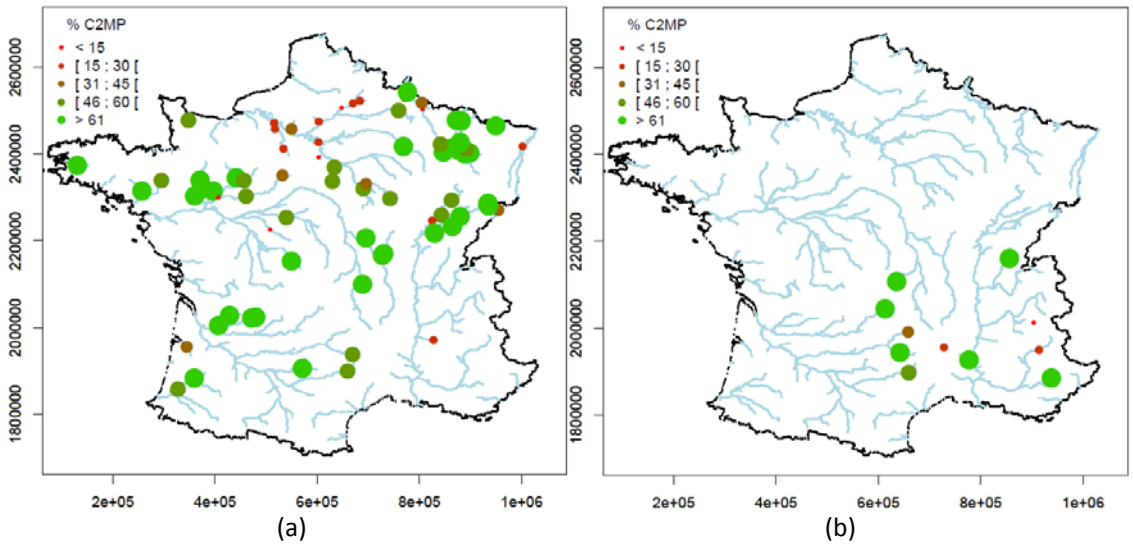


Figure 14 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) selon le critère de C2MP et pour l'échantillon de : (a) 74 et (b) 11 bassins versants.

Pour la performance en calage du modèle *MORDOR* (Figure 15), en ce qui concerne le critère de *NASH,* les valeurs obtenues varient entre 78 et 91. Deux bassins versants situés dans les Alpes et deux dans le Massif Central (Verdon à Castillon, Romanche aux Chambons, Sioule aux Fades et Truyère à Grandval) présentent les plus faibles performances (*NASH* < 85). Nous remarquons néanmoins que les valeurs du critère de *NASH* restent élevées et satisfaisantes pour tous les bassins versants étudiés. Quant au calage basé sur le critère *KGE,* nous remarquons que les bassins versants avec les plus faibles performances correspondent aux bassins versants de faible performance également avec le critère de *NASH*.

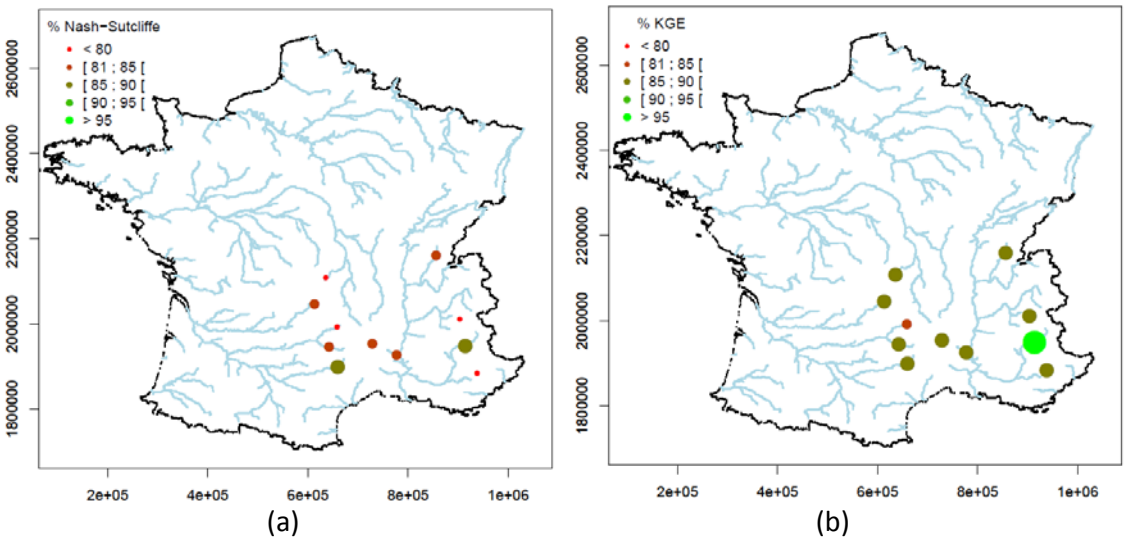


Figure 15 : Performance au calage du modèle MORDOR sur un échantillon de 11 bassins versants selon le critère de: (a) Nash-Sutcliffe (NASH), (b) Kling-Gupta (KGE).

2.5 Synthèse

Ce chapitre a présenté les données et les modèles qui ont été utilisés dans cette thèse. Pour répondre à nos objectifs, nous avons mis en place une large base de prévisions météorologiques (précipitation et températures) :

- les prévisions météorologiques d'ensemble du centre européen CEPMMT avec au total 51 membres sur un horizon de prévision de 10 à 15 jours et pour 4 années de prévisions disponibles (2005 à 2008),
- les prévisions d'ensemble de la base TIGGE avec 8 centres météorologiques disposant au total de 232 membres sur un horizon maximal de prévision de 15 jours et pour 2 années de prévisions (2006 à 2008),
- les prévisions par analogie avec 50 membres sur un horizon de prévision de 7 jours et pour 4 années de données (2005 à 2008).

Ces prévisions seront utilisées avec deux modèles hydrologiques : le modèle *GRP* et le modèle *MORDOR*, calés sur deux échantillons de bassins versants français : un jeu de 74 grands bassins versants et un jeu de 11 bassins versants à fort intérêt pour la production hydroélectrique. La constitution de ces échantillons de bassins versants s'est centrée sur deux objectifs principaux :

- d'un côté, avoir un échantillon suffisamment large afin de produire des conclusions d'une représentativité générale,
- de l'autre côté, avoir un échantillon suffisamment diversifié pour que les conclusions soient pertinentes et représentatives d'une large gamme de conditions climatiques et hydrologiques rencontrées en France.

L'utilisation des prévisions de plusieurs modèles météorologiques et l'application à deux modèles hydrologiques, qui ont des traits en commun (les deux sont des modèles globaux à vocation opérationnelle) mais se distinguent par des représentations différentes des processus physiques et des paramétrages différentes - avec un nombre de paramètres libres différent - d'un côté confère à cette étude une diversité de configurations à examiner lors de l'évaluation de la qualité et de la valeur des prévisions hydrologiques d'ensemble ; de l'autre côté fournit aux conclusions une généralité et une robustesse nécessaire. Ces deux aspects sont examinés dans les prochains chapitres.

"Goodness like beauty lies in the eye of the beholder and has many facets"

Jolliffe, I. et D. Stephenson, 2003

3. La qualité des prévisions: critères d'évaluation et méthodes de correction de biais

Dans ce chapitre, nous présentons les méthodes utilisées pour examiner la qualité des prévisions d'ensemble et ses principaux attributs. Nous discutons également des différentes approches statistiques existantes dans la littérature pour améliorer la qualité des prévisions. Les approches exploitées dans cette thèse sont présentées.

3.1 Introduction

Selon Murphy (cité dans Jolliffe et Stephenson, 2003), une " bonne prévision " peut l'être selon deux aspects : i) la qualité, qui examine la correspondance entre observations et prévisions, et ii) la valeur (ou utilité), qui concerne la valeur économique de la prévision pour un utilisateur décideur.

Dans ce chapitre, nous nous focalisons sur l'évaluation de la qualité des prévisions (l'utilité sera abordée au Chapitre 4). Le but est de comprendre les attributs que l'on souhaite retrouver dans les prévisions et d'introduire les critères appropriés qui seront utilisés dans cette thèse.

Une fois les critères d'évaluation définis, des approches statistiques qui cherchent à améliorer la qualité des prévisions sont présentées. Nous nous focalisons également sur les approches qui seront utilisées dans le cadre de cette thèse.

3.1.1 Cadre général de l'évaluation des prévisions d'ensemble

La qualité d'une prévision est en général jugée par rapport à l'observation. Dans le cadre simple de l'évaluation d'une prévision à scénario unique, on peut directement dire que la prévision est "correcte" ou "fausse" une fois l'événement observé. En revanche, dans le cadre des prévisions probabilistes ou des prévisions d'ensemble certaines particularités sont à prendre en compte dans l'évaluation de la qualité des prévisions. Puisque ces prévisions attribuent une probabilité à l'occurrence et à la magnitude d'un événement, les prévisionnistes doivent également évaluer cette information sur l'incertitude de la prévision émise. Il s'agit de déterminer comment la probabilité des événement s prévus dans l'ensemble correspond à la fréquence à laquelle les événement s sont observés (évaluation de la *cohérence* statistique des prévisions). Pour cela, il est nécessaire de comparer une longue série de prévisions avec la série correspondante des observations. Selon Jolliffe et Stephenson (2003), cette évaluation de la qualité d'une prévision d'ensemble consiste à trouver où se situe l'observation par rapport à la gamme de valeurs prévues par les différents membres de la prévision d'ensemble.

Ces différents membres de la prévision d'ensemble déterminent la densité de probabilité prédictive. La fonction de répartition correspondante donne les probabilités au non-dépassement pour différentes valeurs de la variable. L'évaluation de la prévision s'effectue en comparant, sur une longue série de paires " prévision-observation ", la probabilité de prévision affectée par la prévision d'ensemble à l'observation et la fréquence empirique d'occurrence de l'observation.

Murphy et Winkler (1987) ont ainsi proposé une approche d'évaluation des prévisions centrée sur la distribution conjointe des prévisions et des observations. Cette approche est basée sur le principe que, si les prévisions et les observations sont des variables indépendantes et identiquement distribuées, la relation entre prévision et observation est décrite intégralement par leur distribution conjointe.

Deux lectures, dites " factorisations ", de la distribution conjointe selon une distribution conditionnelle et une distribution marginale sont possibles :

- *Factorisation 1 (en anglais "calibration-refinement factorisation") :*

$$p(f,o) = p(o|f)\,p(f) \qquad \text{Eq. 9}$$

, où $p(o|f)$: est la distribution de probabilité conditionnelle de l'observation o étant donnée la prévision f ,

$p(f)$: est la distribution de probabilité marginale de la prévision f .

Les probabilités conditionnelles $p(o|f)$ sont les probabilités prédictives, qui permettent de répondre, par exemple, à la question " quelle observation est la plus probable pour une prévision donnée ? ". Cette distribution de probabilité conditionnelle est également appelée " distribution de probabilité *a posteriori* ". Elle donne les valeurs possibles de l'observation en fonction des valeurs prévues par la prévision *f*.

- *Factorisation 2 (en anglais "likelihood-base rate factorization") :*

$$p(f,o) = p(f|o)\,p(o) \qquad \text{Eq. 10}$$

, où $p(f|o)$: est la distribution de probabilité conditionnelle de la prévision f étant donnée l'observation o ,

$p(o)$: est la distribution de probabilité marginale de l'observation o .

La distribution de probabilité conditionnelle $p(f|o)$ est la fonction de vraisemblance. Elle permet de répondre à la question : " quelle prévision ressemble le plus à l'observation ? ". C'est à cette question que nous nous adressons lors du traitement statistique des prévisions pour la correction des erreurs de prévision. Nous cherchons à identifier la distribution prédictive qui capture l'observation. Les erreurs des prévisions brutes sont corrigés de façon à aboutir, par exemple, à un maximum de vraisemblance (les méthodes de traitement statistique de correction de biais sont présentées dans §3.3).

Les deux factorisations de la distribution de probabilité conjointe sont équivalentes et reliées par le Théorème de Bayes :

$$p(o|f) = \frac{p(f|o)\,p(o)}{p(f)} \qquad \text{Eq. 11}$$

La distribution de probabilité *a posteriori* (la distribution prédictive) $p(o|f)$ résume l'incertitude d'estimation de l'observation étant donnée la prévision. Elle est proportionnelle au produit de la connaissance *a priori* de la variable à prévoir (l'incertitude autour de l'observation, $p(o)$, sans connaissance de la prévision) et de la fonction de vraisemblance $p(f|o)$ (la fonction qui traduit combien le système de prévision est vraisemblable par rapport aux observations). A chaque prévision, la connaissance *a priori* est ainsi mise à jour par l'information supplémentaire qui vient des prévisions $p(f)$ pour fournir la distribution prédictive qui est celle qui intéresse le prévisionniste. C'est à partir de cette distribution prédictive (la distribution *a posteriori*) que le prévisionniste aura l'information sur la probabilité d'occurrence d'un événement donné, par exemple.

3.1.2 Attributs de la qualité des prévisions

Les deux factorisations des probabilités conjointes présentées peuvent être utilisées pour déterminer les attributs définissant la qualité des prévisions. Murphy (1997) a montré que les moments de la distribution conjointe, des distributions marginales, ainsi que des distributions conditionnelles décrivent les différents attributs de la qualité des prévisions. La qualité d'une prévision est pourtant un concept multidimensionnel.

Par la factorisation 1, le biais conditionnel des observations *o* étant donnée la prévision *f* peut être mesuré. Dans la littérature, ce biais est souvent connu comme " Biais du Type I ". Il est une mesure globale de la fiabilité des prévisions (Bradley *et al.*, 2004). Mathématiquement, il est exprimé par :

$$BiaisI = E_f[(\mu_{o|f} - f)^2] \qquad \text{Eq. 12}$$

, où *E[]* : est l'espérance,

$\mu_{o|f}$: est l'espérance de l'observation conditionnée à la prévision.

Les prévisions conditionnellement non-biaisées, c'est-à-dire, fiables, sont désirables. La fiabilité correspond à la ressemblance entre les probabilités associées à l'observation par la prévision et les fréquences relatives d'occurrence des valeurs observées. Un système de prévision fiable est capable de bien représenter les incertitudes de prévision et de fournir des intervalles de confiance qui correspondent bien aux niveaux de confiance réels (ou qui sont proches des niveaux nominaux) des observations. Remarquons que Murphy (1973) a utilisé le terme de " validité " au lieu de " fiabilité ", avec une notion plus générale, liée à la prévision dans sa totalité.

Un autre attribut de la qualité des prévisions extrait par la factorisation 1 est la "résolution". Mathématiquement cette propriété est exprimé par :

$$RES = E_f[(\mu_{o|f} - \mu_o)^2] = \mathrm{var}(\mu_{o|f}) \qquad \text{Eq. 13}$$

, où *E[]* : est l'espérance,

μ_o : est l'espérance des observations,

$\mu_{o|f}$: est l'espérance de l'observation conditionnée à la prévision,

var() : est la variance.

La résolution est une qualité de la prévision à elle seule, indépendante de la fiabilité des prévisions. Un système de prévision d'ensemble possède une forte " résolution " (souhaitable) si la distribution prédictive est capable de fournir des probabilités distinctes des probabilités climatologiques ou, autrement dit, la résolution mesure quelle part de l'incertitude de l'observation peut être résolue par les prévisions (Weijs, 2010). Une prévision basée sur la climatologie possède ainsi une bonne fiabilité, mais une très faible résolution, tandis qu'une prévision déterministe ne contiendra aucune information sur la résolution.

Brown et Seo (2012) ont fait une remarque intéressante sur ces deux attributs de la qualité des prévisions, fiabilité et résolution : la probabilité conditionnelle $p(o|f)$ est à la fois une mesure de la fiabilité des prévisions quand elle est comparée à l'observation et une mesure de la résolution quand nous prenons en compte seulement sa sensibilité à la prévision *f*.

Si nous revenons à la factorisation 2, nous pouvons estimer le biais conditionnel des prévisions *f* étant donnée l'observation *o*. Dans la littérature, ce biais est connu comme " Biais Type II " (Bradley *et al.*, 2004 ; Brown et Seo, 2012). Mathématiquement, ce biais conditionnel est exprimé par :

$$BiaisII = E_o[(\mu_{f|o} - o)^2] \quad \text{Eq. 14}$$

, où *E[]* : est l'espérance,

$\mu_{f|o}$: est l'espérance de la prévision conditionnée à l'observation.

Idéalement, les prévisions doivent être non-biaisées, c'est-à-dire, la valeur de ce biais doit être la plus petite possible.

L'autre attribut issu de la factorisation 2 est la " discrimination ". Cet attribut représente la capacité du système à discriminer entre deux événements distincts (ex., occurrence ou non d'un dépassement d'un seuil critique) et à attribuer des probabilités fortes à un événement lorsque celui-ci est observé. Par conséquent, les prévisions avec une discrimination plus élevée sont souhaitables, car cela veut dire que des prévisions différentes seront suivies par des observations différentes. L'expression mathématique de cet attribut est donnée par :

$$DIS = E_o[(\mu_{f|o} - \mu_f)^2] \quad \text{Eq. 15}$$

Indépendamment de la factorisation suivie, la multi-dimensionnalité de la qualité des prévisions est souvent difficile à synthétiser. Pour cette raison, il est souvent essentiel de faire référence à des concepts familiers. La notion de " justesse " en est un exemple.

Globalement, une prévision est considérée comme bonne si elle diffère peu de l'observation. Plus particulièrement pour une prévision d'ensemble, une prévision sera qualifiée comme bonne si tous les membres sont proches de l'observation, c'est-à-dire, si le faisceau de prévision est serré autour de l'observation (Figure 16).

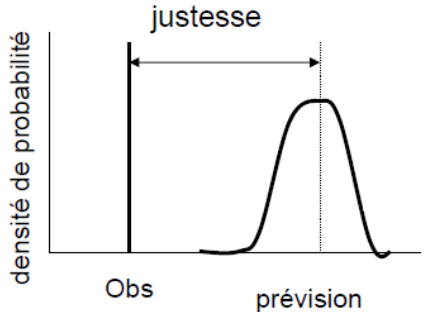


Figure 16 : Représentation schématique de la justesse d'une prévision.

La justesse est facilement calculée par une différence quadratique entre la prévision et l'observation. L'Eq. 16 exprime mathématiquement la justesse.

$$JUST = E[(f - o)^2] \quad \text{Eq. 16}$$

Dans le cas de prévisions déterministes, cette équation correspond à l'erreur quadratique moyenne des prévisions. En décomposant l'Eq. 16, nous obtenons :

$$JUST = (\sigma_f{}^2 + \mu_f{}^2) + \mu_o(1 - 2\mu_{f|o}) \quad \text{Eq. 17}$$

, où μ_f : est l'espérance de la prévision,

σ_f : est la variance de la prévision,

μ_o : est l'espérance de l'observation,

$\mu_{f|o}$: est l'espérance de la prévision conditionnée à l'observation.

En suivant la factorisation 1 et en prenant en compte les Eq. 12 et Eq. 13, nous obtenons :

$$JUST = BiaisI - RES + \sigma_o{}^2 \quad \text{Eq. 18}$$

, où σ_o : est la variance de l'observation.

De même en suivant la factorisation 2 et en combinant les Eq. 14 et Eq. 15, la justesse s'exprime par :

$$JUST = BiaisII - DIS + \sigma_f{}^2 \quad \text{Eq. 19}$$

, où σ_f : est la variance de la prévision.

Les Eq. 18 et Eq. 19 montrent alors la composition de la justesse par les autres attributs de la qualité des prévisions. Nous remarquons que les deux variances, σ_o et σ_f, sont des propriétés statistiques des observations et des prévisions, respectivement. La variance des observations σ_o correspond à l'incertitude intrinsèque de la variable à prévoir (température, précipitation, débit, etc.). De l'autre côté, la variance des prévisions σ_f correspond à la dispersion des membres de la prévision d'ensemble entre eux, autrement appelée *"finesse"* (en anglais *sharpness*). Pour une prévision parfaitement fiable, la distance moyenne entre la moyenne des ensembles et l'observation est égale à la distance moyenne entre la moyenne des membres et chaque membre individuel. Dans ce cas la dispersion des ensembles est une mesure de la résolution (Jolliffe et Stephenson, 2003).

Dans le cas de l'évaluation des prévisions d'ensemble, le but est d'évaluer, d'une part, la justesse de l'ensemble (et ses attributs) et, d'autre part, la précision des membres de l'ensemble. Nous souhaitons, par exemple, évaluer si les valeurs prévues sont proches ou non des valeurs observées ainsi que si la probabilité d'occurrence de l'événement est respectée (c'est-à-dire que si un système de prévision d'ensemble donne, en moyenne, une probabilité d'occurrence de 30% pour un événement donné, cet événement est, en moyenne, observé dans 30% des cas). Graphiquement, ces propriétés sont schématisées dans la Figure 17.



(a) (b)

Figure 17 : Illustration de : (a) la justesse et (b) la précision d'une prévision pour attendre sa cible, l'observation.

3.2 Les critères d'évaluation des prévisions

Il existe deux principales démarches pour l'évaluation de la qualité des prévisions probabilistes : celles qui évaluent les propriétés intrinsèques à l'ensemble et qui, strictement parlant, sont indépendantes des observations (ex., la dispersion des membres) et celles qui utilisent les scores ou critères de performance pour évaluer des paires '' prévision-observation '' sur une longue série de données (Jolliffe et Stephenson, 2003 ; Wilks, 2011).

Parmi les critères les plus utilisés pour l'évaluation de la qualité de la dispersion de l'ensemble et de son rapport à l'observation, il ressort : le diagramme de rang ou diagramme de Talagrand (Talagrand *et al.,* 1997), le diagramme de PIT (Hamil, 2001 ; Gneiting *et al.*, 2007), la dispersion (ou écart-type de l'ensemble) et le diagramme de fiabilité (Wilks, 2011).

En ce qui concerne les scores probabilistes d'évaluation des paires '' prévision-observation '', le Score de Brier ou *BS* (Brier, 1950 ; Epstein et Murphy, 1965), le Score de Probabilité Ordonnée ou *RPS* (en anglais *Ranked Probability Score,* Epstein, 1969 ; Murphy, 1971), et sa forme continue, le Score Continu de Probabilité Ordonnée ou *CRPS* (en anglais *Continuous Ranked Probability Score,* Matheson et Wilkler, 1976 ; Hersbach., 2000), le Score de Divergence (Weijs, 2010) et les courbes ROC (en anglais *Receiver operating characteristic,* Mason, 1982 ; Ahrens et Jaun, 2007) sont parmi les scores les plus largement utilisés pour les

prévisions probabilistes (Murphy et Epstein, 1967 ; Epstein, 1969 ; Mason, 1982 ; Parker et Fordham, 1996 ; Wood *et al.*, 2002 ; Jolliffe et Stephenson, 2003 ; Laio et Tamea, 2007 ; Roulin, 2007 ; Zappa *et al.*, 2008 ; Casati *et al.*, 2008 ; Cloke et Pappenberger, 2009a ; Demargne *et al.*, 2009 ; Jaun et Ahrens, 2009 ; Brown *et al.*, 2010a ; Wilks, 2011 ; etc.).

De plus, des scores traditionnellement utilisés en prévision déterministe (tableau de contingence, Taux de Bonne Détection *POD*, *Taux de Fausses Alertes FAR*, *RMSE,* etc.) peuvent également être utilisés dans l'évaluation des prévisions probabilistes (ou d'ensemble). Dans ce cas, c'est soit la moyenne des membres (ou moyenne d'ensemble) qui est souvent considérée, soit nous évaluons le tableau de contingence pour différentes valeurs de probabilité (pourcentage de membres de l'ensemble qui prévoient l'événement en question).

Souvent les scores probabilistes sont exprimés par rapport à une référence, donnant lieu aux scores de performance (en anglais, *Skill Scores*). Ils indiquent la qualité d'une prévision par rapport à une référence, qui est définie par l'utilisateur. Il s'agit souvent d'une prévision naïve que l'utilisateur aurait utilisé s'il ne disposait pas d'un système hydrométéorologique de prévision. Les références les plus courantes sont : la climatologie, qui peut correspondre à la valeur moyenne des observations historiques ou à un ensemble de valeurs tirées de la distribution climatologique, la climatologie conditionnelle, qui recherche dans le passé des périodes analogues à celle qui a été observée à l'instant de la prévision (Wilson, 1999), et la persistance, qui correspond au résultat d'un modèle naïf pour lequel la quantité prévue est égale à la dernière valeur observée.

Dans cette pléthore de critères d'évaluation, il est souvent difficile de choisir le score le plus adapté à chaque cas. Une façon de réduire les options possibles est d'examiner si un score d'évaluation est propre ou pas. Un score est strictement propre si, et seulement si, la prévision *F* correspond au meilleur jugement *Q* du prévisionniste (ex., Score de Brier, Score de Divergence, etc.) et, par conséquent, elle ne peut pas être améliorée en fournissant une prévision différente. Ce fait est en contraste avec les cas où le prévisionniste sait à l'avance que le score sera amélioré si la prévision diffère de son meilleur jugement. Mathematiquement, la definition d'un score propre est exprimée par :

$$S(Q,Q) \geq S(F,Q) \quad quand \quad Q \geq F \qquad \text{Eq. 20}$$

, où *S(F,O)* : est la récompance du prévissioniste pour sa prévision *F* quand l'observation *O* se réalise.

Puisque chaque critère apporte des informations supplémentaires il est essentiel d'examiner plusieurs critères d'évaluation afin d'obtenir une vision complète de la qualité des prévisions selon ses différents attributs. Dans cette thèse, après avoir synthétisé les principaux scores rencontrés dans la littérature - tout en précisant la question à laquelle chaque score cherche à répondre, ainsi que les attributs qu'il identifie - nous avons sélectionné quelques mesures que nous jugeons le mieux adaptées à l'évaluation de la qualité des prévisions hydrologiques d'ensemble. Cette sélection est une combinaison de mesures graphiques et de scores numériques. L'objectif est de rassembler les informations nécessaires pour identifier la qualité des faisceaux de nos prévisions, et, par conséquent, pour évaluer l'apport des méthodes de post-traitement adoptées. Ci-après, nous présentons avec plus de détails les critères choisis.

3.2.1 Biais inconditionnel

La comparaison des prévisions à l'observation permet l'identification des biais éventuels (écarts à l'observation) et plus particulièrement du biais inconditionnel. Ce biais se réfère au degré de correspondance entre les valeurs moyennes (ou la moyenne des valeurs) de la prévision et de l'observation. Par exemple, les graphiques de type '' nuages de points '' affichant la relation entre les prévisions (en abscisses) et les observations (en ordonnées) fournissent une première estimation visuelle des biais (inconditionnels) existants. Cette première estimation peut être confirmée par des graphiques montrant les erreurs (ou résidus) de prévision en fonction des valeurs prévues. De tels graphiques montrent aussi la tendance à la sur ou sous-estimation des observations et ce pour différentes catégories (classes) de débits.

Une mesure souvent utilisée pour évaluer le biais inconditionnel des prévisions est la racine carrée de l'erreur quadratique moyenne RMSE (ou l'erreur absolue moyenne, MAE). Dans cette thèse, nous utiliserons l'erreur RMSE, calculée par la formule :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2} \qquad \text{Eq. 21}$$

, où o_i : est la valeur observée à l'instant de prévision *i*,

f_i : est la valeur prévue à l'instant de prévision *i*,

N : est le nombre total d'instants de prévisions utilisées pour calculer le score.

La RMSE varie entre 0 et ∞ ; score parfait = 0. Elle est exprimée en unités de la variable évaluée.

3.2.2 Diagramme de PIT

L'identification du biais inconditionnel est essentielle pour la correction des prévisions, mais elle doit être accompagnée d'une estimation du biais conditionnel. Les deux principaux types de biais conditionnels sont le Type I et le Type II (Bradley *et al.*, 2004 ; Brown et Seo, 2012). Ils sont liés aux factorisations 1 et 2 présentées ci-dessus. Nous rappelons que :

- le biais conditionnel du Type I est lié à la probabilité de l'observation o conditionnée à la prévision f : $p(o|f)$,
- le biais du Type II est liée à la probabilité de prévision f conditionnée à l'observation o : $p(f|o)$.

Les mesures qui sont souvent utilisées pour l'identification du biais du Type I sont le diagramme de fiabilité, le diagramme de rang (au diagramme de Talagrand ; Talagrand, 1997) et l'histogramme de PIT (en anglais, *Probability Integral Transform* ; Diebolt *et al.*, 1998). La principale différence entre ces mesures est que le diagramme de fiabilité est construit pour un événement donné, et donc répond à un problème particulier de prise de décision (dépassement ou non d'un seuil critique), tandis que l'histogramme de PIT et le diagramme de rang évaluent la fiabilité des prévisions sur l'ensemble des observations. L'histogramme de PIT est visuellement similaire au diagramme de rang (Glahn, 2009).

Dans cette thèse, nous utiliserons l'histogramme de PIT ou sa version cumulée, le diagramme de PIT, pour évaluer la fiabilité des prévisions. L'histogramme de PIT considère la localisation de l'observation dans la fonction de répartition des prévisions (Hamill, 2001). A chaque observation, nous évaluons la probabilité de prévision associée à cette valeur observée. L'ensemble de probabilités peut ensuite être analysé selon un histogramme de fréquence (histogramme de PIT) ou une fonction de répartition (diagramme de probabilités cumulées).

La Figure 18 présente les différentes formes que l'on peut obtenir à partir de l'analyse des PIT. Un histogramme uniforme ou un diagramme de probabilité cumulée aligné à la diagonale est une indication de prévisions fiables. Comme pour le diagramme de rang, un histogramme sous forme d'arche correspond à des prévisions plutôt sur-dispersées, tandis qu'un diagramme sous forme de "U" est indicatif de prévisions sous-dispersées (Hamill, 2001). Ainsi, quand les fréquences des observations sont supérieures aux probabilités cumulées des prévisions le système de prévision a tendance à sous-estimer l'observation et vice versa. Des diagrammes asymétriques illustrent la présence de biais particuliers dans les prévisions. Le diagramme de probabilités cumulées représente l'accord entre les fréquences d'observation et la probabilité cumulée des prévisions probabilistes correspondantes. La correspondance idéale est donnée par la diagonale.

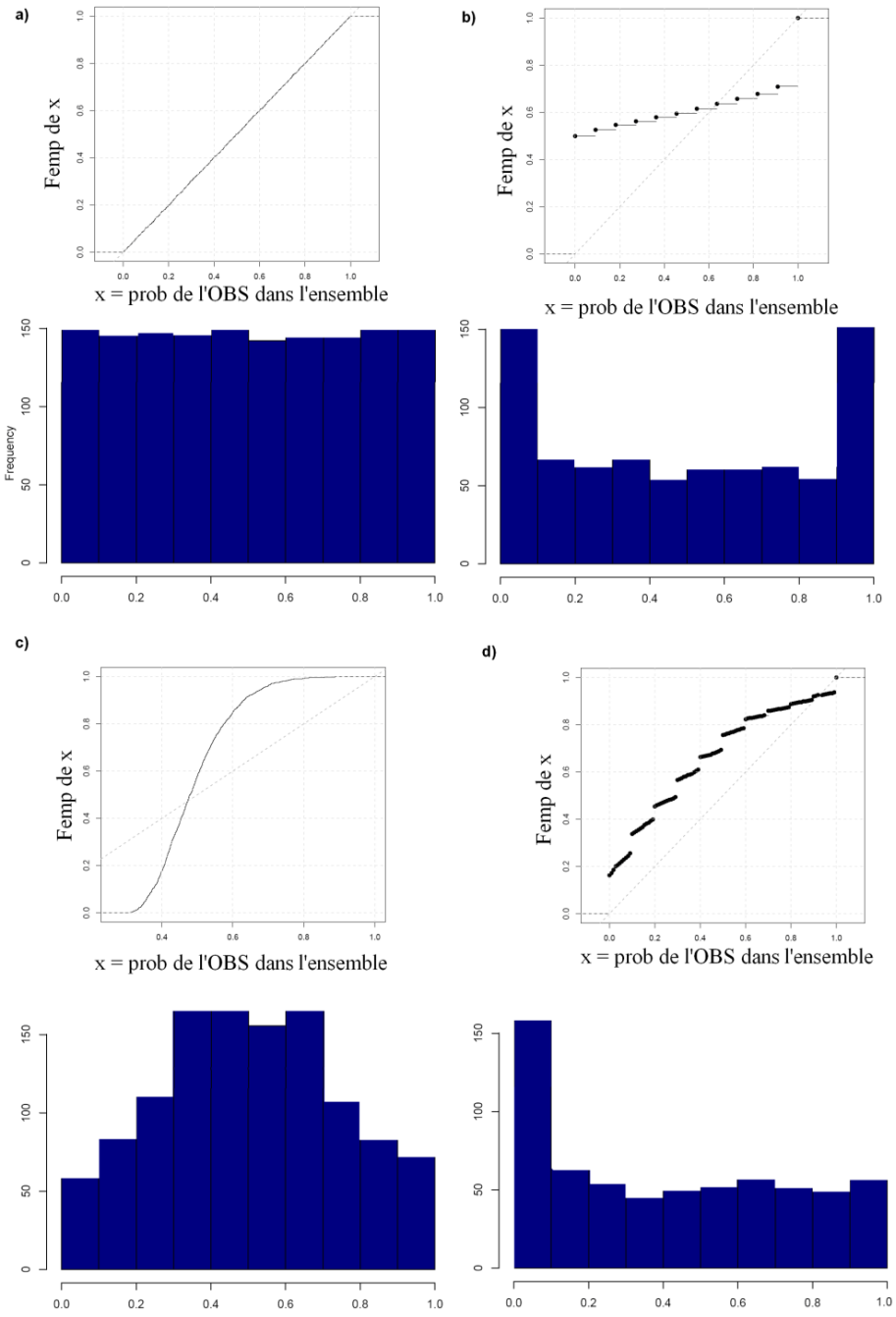


Figure 18 : Interprétation du critère '' PIT '', a) prévision fiable, b) prévision sous-dispersée, c) prévision sur-dispersée, d) prévision biaisée. Diagramme de PIT en haut et histogramme de '' PIT '' en bas.

3.2.3 Dispersion des membres

Souvent la sur- ou sous- dispersion des membres est responsable des événements manqués ou des fausses alarmes. La dispersion de l'ensemble représente l'écart moyen des membres de l'ensemble par rapport à la moyenne de l'ensemble. Elle est une mesure de la précision de l'ensemble.

La sous-dispersion en particulier constitue souvent un défaut important des systèmes de prévision d'ensemble, surtout pour les prévisions aux plus courtes échéances. Dans ce cas-là, la dispersion est trop faible et l'ensemble risque de ne pas capturer suffisamment la réalité. D'un autre côté quand un ensemble est sur-dispersif l'observation est en moyenne "noyée" entre les membres de l'ensemble et ainsi la prévision fournit par l'ensemble n'est pas pertinente.

La mesure de la dispersion des membres de l'ensemble la plus utilisée est l'écart-type σ :

$$\sigma_i = \sqrt{\frac{1}{\mathrm{K}} \sum_{j}^{\kappa} (f_{ij} - \overline{f_i})^2} \qquad \text{Eq. 22}$$

, où K : le nombre de membres de l'ensemble,

f_{ij} : la valeur prévue par le membre *j* à l'instant *i,*

$\overline{f_i}$: la moyenne des valeurs prévues par les membres à l'instant *i*.

Souvent la dispersion des membres de la prévision d'ensemble est confrontée à l'erreur quadratique moyenne (RMSE), calculée à partir de la moyenne des membres de l'ensemble. Pour un bassin versant et pour chaque échéance, le graphique " précision-dispersion " montre ainsi la relation entre la RMSE (en abscisses) et la dispersion (en ordonnées) pour un instant de prévision *i*. La diagonale représente une erreur quadratique moyenne égale à la dispersion des ensembles, ce qui correspond à des erreurs de prévision expliquées totalement par la dispersion de l'ensemble.

Il faut souligner que l'interprétation de la relation entre la dispersion et l'erreur de la prévision n'est pas toujours évidente (Mason et Stephenson, 2008). Plusieurs études ont montré une faible corrélation entre la dispersion et l'erreur de prévision (Atger, 1999 ; Hamill *et al.*, 2004). Ce fait est partiellement expliqué par la faible variabilité de la dispersion (Whitaker et Loughe, 2003) rencontrée souvent dans les prévisions d'ensemble.

3.2.4 Courbe ROC

La courbe ROC (Mason, 1982) consiste à tracer les taux de fausses alertes *POFD* (Eq. 26) contre les probabilités de bonne détection *POD* (Eq. 27), les deux mesures issues de l'analyse des tableaux de contingence (Figure 19**)** pour un seuil critique donné qui définit un événement observé et pour différents seuils de probabilités de prévision de l'événement.

		Evénement observé		
		Oui	Non	
Evénement prévu	Oui	Succès (a)	Fausse Alerte (b)	Prévu
	Non	Manqué (c)	Rejet Correct (d)	Non prévu
		Observé	Non Observé	Total

Figure 19 : Tableau de contingence pour une prévision.

En pratique, pour les prévisions d'ensemble, nous calculons la probabilité de dépasser le seuil critique par au moins 1, 2, 3… ou *K* membres de la prévision d'ensemble. Pour chaque probabilité de dépassement du seuil, on calcule un tableau de contingence pour déduire les scores POFD et POD. Par conséquent, si on a *K* membres de prévision, l'ensemble des *K* couples (POFD, POD) constitue la courbe ROC (Atger, 2001). Plusieurs systèmes (d'ensemble ou déterministe) peuvent être graphiquement comparés par rapport à leur position dans le graphique de la courbe ROC. Une prévision déterministe figurera comme un point dans le graphique, tandis que pour les prévisions d'ensemble une courbe peut être construite en fonction d'un ensemble croissant de probabilités de prévision (Figure 20).

Dans le graphique ROC, une prévision parfaite donnerait un taux de fausses alertes nul (POFD=0), alors que le taux de bonnes prévisions serait de 100% (POD=1). Ainsi, plus le point ou la courbe ROC sera proche du couple (0,1), c'est-à-dire, proche du coin supérieur gauche du graphique, meilleures seront les prévisions. Une courbe qui approche la diagonale montre une capacité nulle du système de prévision à discriminer entre bonne détection et fausses alertes de l'événement en question.

$$POD = \frac{a}{a+c} \qquad \text{Eq. 23}$$

$$POFD = \frac{b}{b+d} \qquad \text{Eq. 24}$$

L'aire sous la courbe ROC est aussi fréquemment utilisée pour comparer différents systèmes de prévision. Une courbe ROC confondue avec la 1ère bissectrice indique que la prévision n'a pas de performance et son aire est ainsi égale à 0.5. Dans ce cas, les valeurs de POFD et de POD sont confondues ; le système de prévision serait fait au hasard puisqu'il présente des probabilités de détection et de fausses alertes égales (le pourcentage d'événements prévus quand ils sont observés est le même des événements prévus quand ils ne sont pas observés).

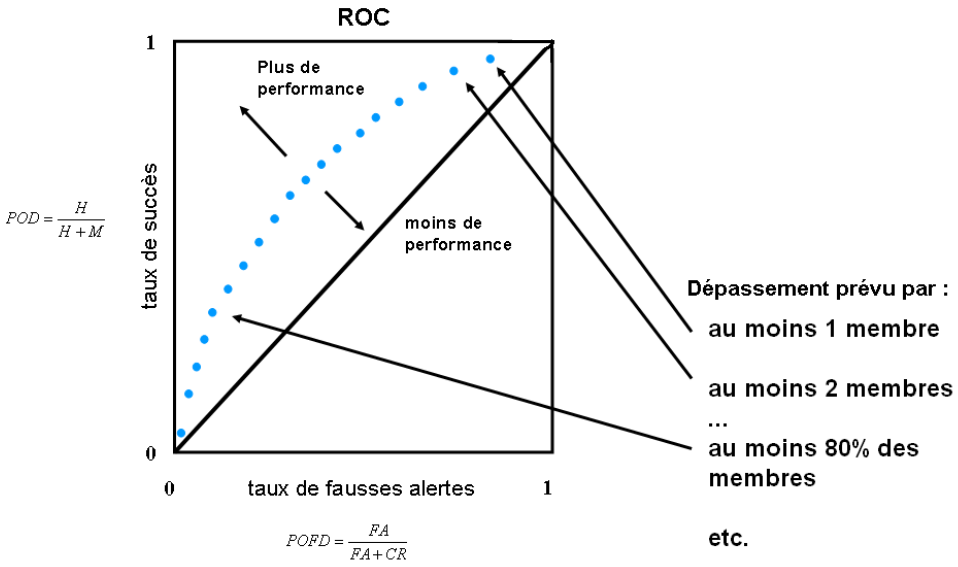


Figure 20 : Schéma représentant la construction d'une courbe ROC (H=a, M=c, FA=b, CR=d du Tableau de Contingence).

Les courbes ROC permettent d'évaluer la capacité des prévisions à discriminer entre les événements (discrimination). Cette mesure n'évalue pas la fiabilité des prévisions : par exemple, une prévision qui n'est pas fiable peut donner une bonne courbe ROC, tandis qu'une prévision basée sur la climatologie, donc fiable, peut donner une courbe ROC proche de la diagonale. Pour cette raison, la courbe ROC doit être utilisée en combinaison avec d'autres mesures de fiabilité.

3.2.5 Score de Brier

Si on s'intéresse à une catégorie spécifique (ex., la prévision du dépassement d'un seuil d'alerte donné), un autre score adapté aux prévisions probabilistes est le Score de Brier, qui a été introduit par Brier en 1950. Ce score est simplement l'erreur quadratique moyenne des probabilités de prévision et d'observation de l'événement en question. Il y a deux approches principales de présentation du Score de Brier : en considérant la prévision probabiliste

directement comme une quantité scalaire (une probabilité) ou comme un vecteur (un ensemble de valeurs de prévision). Selon Murphy *et al.* (1972a, 1972b, 1973), du point de vue scientifique et économique, la définition sous la forme de vecteur est avantageuse. Leur argumentation est basée sur le fait que si on s'intéresse principalement à la prévision, étant donné que la prévision est un vecteur, il est préférable de la traiter sous la forme de vecteur. Par contre, si on s'intéresse principalement aux probabilités fournies par la prévision, étant donné que la probabilité est une quantité scalaire, il est plus simple de la traiter sous la forme scalaire. De ce fait, le Score de Brier s'exprime souvent sous sa forme scalaire.

Pour un événement donné, le Score de Brier est calculé pour deux catégories (occurrence ou non de l'événement) sur un échantillon de *N* paires prévision-observation par :

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \qquad \text{Eq. 25}$$

, où f_i : est la probabilité associée à la paire prévision-observation à l'instant de prévision *i*,

o_i : = 1, si l'événement a été observé à l'instant de prévision *i*.

= 0, sinon.

Le Score de Brier est donc estimé en évaluant la somme des différences au carré entre les probabilités de prévision (pourcentage des membres de l'ensemble qui dépassent le seuil critique) et l'occurrence de l'observation. Pour améliorer ce critère, il faut alors augmenter la probabilité de prévision quand l'événement se produit et la diminuer quand l'événement ne se produit pas.

Le Score de Brier (*BS)*, comme les erreurs du type erreur moyenne absolue (*MAE)* ou erreur quadratique moyenne (*RMSE)*, est négativement orienté : plus petite est sa valeur, meilleure est la performance du système évalué. Le score optimal (*BS* = 0) est atteint pour une prévision déterministe " parfaite ", c'est-à-dire, toujours correcte.

3.2.6 Score RPS

Le score de probabilité ordonée *RPS* a été introduit par Epstein (1969), suivi par une modification majeure proposée par Murphy (1971). Le principe de ce score consiste à comparer la fonction de répartition de la prévision (la distribution prédictive donnée par l'ensemble) à la fonction de répartition de l'observation (qui, en fait, est une fonction cumulée d'une densité de probabilité qui vaut 1 au point de l'observation et 0 ailleurs). Le score RPS est équivalent à la somme des scores de Brier, évalués sur différentes catégories *m*.

Pour un échantillon de *N* paires prévision-observation et *M* états mutuellement exclusifs et exhaustifs (composés des sous–catégories, définies, par exemple, par les percentiles 10% à 90% des observations), le score est donné par :

$$RPS = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{M}\sum_{m=1}^{M}\left(f_{i,m} - o_{i,m}\right)^2\right) \qquad \text{Eq. 26}$$

, où $f_{i,m}$: la probabilité prédictive pour la paire prévision-observation à l'instant de prévision *i* et, pour la sous - catégorie *m*,

$o_{i,m}$: = 1, si l'observation appartient à la sous-catégorie *m* à l'instant de prévision *i.*

= 0, si l'observation n'appartient pas à la sous-catégorie *m.*

Le score parfait est égal à 0 et le score 1 est attribué à la plus mauvaise performance du système. Le *RPS* est sensible à la distance entre la probabilité prédictive et la fréquence d'observation.

L'avantage du score *RPS* est sa formulation adaptée au cas des prévisions exprimées en catégories ordonnées : par exemple, le cas où nous nous intéressons à la capacité d'un système à prévoir : 1) les débits faibles (inférieurs au quantile 20%), 2) les débits moyens (entre les quantiles 20% et 80%) et 3) les débits forts (supérieurs au quantile 80%). Néanmoins, un inconvénient majeur de ce score est le fait que sa valeur peut varier fortement selon le découpage en catégories (le choix des catégories et de la largeur de chaque catégorie).

3.2.7 Score CRPS

Si on considère un nombre de catégories infinies avec une largeur négligeable, le *RPS*, un score qui s'exprime sous la forme discrète, tend vers la forme continue donnée par le *CRPS* (Matheson et Winkler, 1976 ; Hersbach, 2000). De plus, le *CRPS* a un caractère dimensionné : il s'exprime dans les mêmes unités de la variable évaluée (ici, m^3/s pour les débits et mm pour les pluies).

Le *CRPS* a été interprété comme la version probabiliste de l'erreur absolue moyenne *MAE* (Gneiting et Raftery, 2007) et utilisé pour faire des comparaisons entre la performance d'un système de prévision d'ensemble (exprimée par le *CRPS*) et celle d'un système déterministe (exprimée par la *MAE*), pour une même variable évaluée (ex., Velázquez *et al.*, 2011). Enfin, nous notons que le *CRPS* peut être aussi compris comme l'intégrale du Score de Brier sur toutes les sous-catégories et il est exprimé en mm.

$$\overline{CRPS} = \int_{-\infty}^{+\infty} BS(x)dx \qquad \text{Eq. 27}$$

$$CRPS = \frac{1}{N}\sum_{i=1}^{N} \int_{x=+\infty}^{x=-\infty} \left(F_i^{\,f}(x) - F_i^{\,o}(x)\right)^2 \, dx \qquad \text{Eq. 28}$$

, où $F_i^{\,f}(x)$: est la fonction de répartition des prévisions à l'instant de prévision *i*,

$F_i^{\,o}(x)$: est l'observation, exprimée comme une fonction de répartition, à l'instant de prévision *i*,

N : est le nombre total d'instants de prévision utilisé pour calculer le score.

Si l'observation est une valeur unique, alors sa fonction de répartition est une fonction du type " marche d'escalier " ou indicatrice (fonction Heaviside), prenant la valeur 0 pour tous les réels inférieurs à la valeur observée et la valeur 1 partout ailleurs :

Afin de calculer numériquement le *CRPS*, Gneiting et Raftery (2007) proposent l'adoption d'une loi de distribution gaussienne pour l'estimation de la fonction de probabilité. Une telle hypothèse peut être satisfaite pour les prévisions de température ou de pression atmosphérique, par contre ceci n'est pas le cas des séries hydrologiques. Dans ce cas, le calcul du *CRPS* est souvent fait par une approximation discrète de l'intégrale de l'Eq. 28 (Hersbach, 2000). Dans cette thèse nous avons appliqué une loi empirique pour l'estimation de la densité de probabilité cumulée et une discrétisation suffisamment importante de cette distribution prédictive.

3.2.8 Score de Divergence

Sous un principe similaire à celui du Score de Brier et du Score RPS, est fondé le Score de Divergence (en anglais *"Kullback-Leibler Divergence Score"*), qui quantifie le gain en information supplémentaire obtenu par la prévision d'ensemble par rapport à l'observation (Weijs et al., 2010). Ce score *DS* calcule la divergence entre la loi de fréquence de l'observation o_i et la distribution prédictive f_i à l'instant de prévision *i* (Eq. 29 et Eq. 30). Pour un événement binaire (ex., dépassement ou non d'un seuil critique) et *N* paires " prévision-observation ", le score de divergence est défini par les équations suivantes :

$$DS(o_t|f_i) = \sum_{i=1}^{N} [o_t] \log\left(\frac{[o_t]}{[f_t]}\right) \quad \text{Eq. 29}$$

$$DS = \frac{1}{N}\sum_{i=1}^{N} DS(o_t|f_i) \quad \text{Eq. 30}$$

, où f_i : est la probabilité associée à la paire prévision-observation à l'instant de prévision *i*,

o_i : = 1, si l'événement a été observé à l'instant de prévision *i*.

= 0, sinon.

Ce score peut être interprété comme le gain d'information en passant de la distribution prédictive à la fréquence de l'observation. Si la prévision contient déjà toute l'information existante dans l'observation, le gain d'information est 0 et le score est parfait. Pour le cas des événements multi-catégoriques (*M* étant le nombre de catégories), le Score de Divergence Ordonnée *RDS* (Weijs *et al.*, 2010) est défini par :

$$RDS = \frac{1}{M}\sum_{m=1}^{M}(\frac{1}{N}\sum_{i=1}^{N} DS(o_t|f_{i,m}) \quad \text{Eq. 31}$$

, où $f_{i,m}$: la probabilité prédictive pour la paire prévision-observation à l'instant de prévision *i* et, pour la sous - catégorie m,

$o_{i,m}$: = 1, si l'observation appartient à la sous-catégorie *m à* l'instant de prévision *i.*

= 0, si l'observation n'appartient pas à la sous-catégorie *m.*

3.2.9 Remarques générales sur l'évaluation de la qualité des prévisions

Un des problèmes qui émerge pendant l'évaluation d'un système de prévision est lié à la taille de l'échantillon. Souvent les scores statistiques utilisés demandent une série de données suffisamment longue pour la vérification des prévisions. Le recueil d'un grand nombre de paires " prévision-observations " n'est cependant pas toujours faisable pour plusieurs raisons : le coût important des simulations des systèmes de prévision, l'indisponibilité ou bien l'hétérogénéité des données, les erreurs et lacunes dans les séries des données observées, etc.

Candille (2003) a calculé le nombre théorique de réalisations nécessaires pour la vérification des prévisions d'ensemble, en prenant en compte le nombre des membres de l'ensemble. L'auteur a indiqué que, afin de réaliser une évaluation fiable des prévisions pour un ensemble de 10 membres, il faudrait plus de 5000 prévisions indépendantes. En pratique cette taille de séries n'est pas disponible et l'évaluation des prévisions s'effectue sur des échantillons d'une longueur de quelques années.

En particulier, dans le cas des événements rares, quand la prévision et sa vérification deviennent cruciales, les échantillons de données sont encore plus courts à cause du nombre limité d'occurrences de ces phénomènes. Dans ce cas, l'incertitude liée à l'échantillonnage devient significativement plus importante. Il est néanmoins difficile d'associer une valeur d'incertitude à un score moyen calculé, comme le score de Brier *BS* et le *CRPS*, sur un échantillon. Dans la littérature, des ré-échantillonnages du type Monte Carlo sont proposés (Brown et Seo, 2010), mais la procédure reste coûteuse en temps de calcul et dépendante de la longueur des séries disponibles.

Un deuxième aspect important affectant l'évaluation de la qualité des prévisions d'ensemble est lié au nombre de membres de l'ensemble. La performance de plusieurs mesures de vérification est fortement liée au nombre de membres de l'ensemble. Comme indiqué par Muller *et al.* (2004) et Weigel *et al.* (2006), pour un nombre de membres d'ensemble limité (inférieur à 40 membres), le Score de Brier (*BS*), ainsi que le Score de Probabilité Ordonnée (*RPS*) sont négativement biaisés. Dans ce cas, une correction est proposée dans la littérature (Stephenson *et al.*, 2007).

Enfin, remarquons que si parfois les prévisions sont bonnes et capturent bien les événements à prévoir, nous sommes souvent en présence d'un biais dans le système de prévision qu'il faut corriger. Cette procédure est connue sous le terme *traitement statistique* de correction des sorties d'un modèle, nécessaire afin de débiaiser les prévisions et les rendre bonnes. Les principales familles de méthodes de traitements statistiques sont présentées ci-après, et les méthodes qui seront utilisées dans cette thèse y sont décrites.

3.3 Approches de traitement statistique des prévisions

Pour définir la notion de " traitement statistique des prévisions ", il faut d'abord revenir aux deux factorisations de la probabilité conjointe des prévisions-observations (Eq. 9 et Eq. 10). Rappelons que la factorisation 2 fait appel à la fonction de vraisemblance et cherche à répondre à la question "quelle prévision ressemble le plus à une observation donnée ?".

C'est à cette question que nous nous adressons lors de la procédure de traitement statistique des prévisions. Nous cherchons à identifier la distribution prédictive ayant engendré une observation donnée. Sous l'hypothèse que, à n'importe quel instant *i* la fréquence d'une observation vient d'une fonction de répartition F_i^o et, en même temps, la prévision probabiliste vient d'une fonction de répartition F_i^f, la prévision idéale vérifierait :

$$F_i^{\ f} = F_i^{\ o} \qquad \text{Eq. 32}$$

Néanmoins, la fonction de répartition des observations n'est, au mieux, qu'une hypothèse, tandis que la fonction de répartition prédictive des prévisions reste souvent une estimation d'une loi de probabilité, basée sur un nombre limité de membres d'un ensemble.

L'objectif du traitement statistique des prévisions est de minimiser la différence entre la fonction de répartition des observations et la fonction de répartition prédictive des prévisions. La démarche souvent entreprise est celle qui fait appel aux techniques de correction statistique avant (pré-) ou après (post-) le modèle hydrologique, avec pour but d'améliorer la qualité de prévisions en termes de volume, temps et distribution.

De manière générale, il s'agit d'examiner les erreurs qui se réfèrent au degré de correspondance entre l'observation et la prévision, ainsi que les biais provenant d'une mauvaise dispersion des membres. Pour les prévisions météorologiques, la dispersion des membres est surtout liée à la génération des ensembles et ainsi aux perturbations des conditions initiales considérées. Si la variation initiale introduite par les perturbations n'est pas suffisante, cela conduit à des membres de l'ensemble peu dispersés entre eux et à des prévisions manifestant un excès de confiance (Buizza *et al.*, 2005 ; Wang et Bishop, 2005 ; Zhu, 2005). En même temps, une dispersion insuffisante des membres peut aussi être liée à une incapacité du modèle numérique à représenter proprement l'incertitude du climat et des processus physiques.

Pour ce qui est des prévisions hydrologiques, la dispersion des membres peut être perturbée par la procédure de mise à jour qui, selon l'approche adoptée - même si elle n'introduit pas un biais, elle peut "écraser" l'ensemble. Les approches statistiques de post-traitement visent particulièrement à s'attaquer à ces problèmes, de façon à corriger les biais de la prévision et à mieux ajuster la dispersion des membres de l'ensemble de façon à les rendre plus fiables.

3.3.1 Familles de méthodes de traitement statistique

Un nombre considérable d'approches statistiques pour la correction des prévisions est présenté dans la littérature. En faisant une généralisation, nous pouvons séparer ces approches en deux familles principales :

- les approches qui visent à corriger l'erreur totale de la prévision,
- les approches qui traitent les sources d'erreur séparément.

Parmi les approches qui traitent les sources d'erreur séparément se trouvent : les approches de mise à jour du modèle hydrologique par une technique d'assimilation de données (Seo *et al.*, 2003 ; Goswani *et al.*, 2005 ; Andreatis et Lettenmaier, 2006 ; Liu *at al.*, 2012), les approches d'équiprobabilité qui utilisent des jeux de paramètres du modèle (Kuczera et Parent, 1998 ; Vrugt *et al.*, 2003 ; Beven et Binley, 1992 ; Pappenberger *et al.*, 2005), l'habillage des simulations par la prise en compte exclusivement des erreurs passées du modèle hydrologique (Bogner et Kalas, 2008 ; Zhao *et al.*, 2011 ; Pagano *et al.*, 2012).

Néanmoins, Krzysztofowicz (2002) a souligné l'importance de prendre en compte l'incertitude totale lors du traitement statistique des prévisions. Dans cette direction sont dirigées les approches proposées par Seo *et al.* (2006) et Regianni *et al.* (2009), qui ont utilisé l'inférence bayésienne afin de calculer la probabilité d'occurrence d'un événement basé sur l'expérience du passé. Raftery *et al.* (2005) ont également proposé une méthodologie – dite *Bayesian Model Averaging* (BMA) – qui traite l'incertitude totale en utilisant le cadre bayésien afin de former un ensemble de prévisions de bonne qualité à partir de prévisions de différentes sources. Cette approche est également appliquée par Sloughter *et al.* (2007), Duan *et al.* (2007), Vrugt *et al.* (2008), Bao *et al.* (2010) sur des prévisions hydrométéorologiques d'ensemble. L'hypothèse principale derrière la BMA est que la distribution de probabilité conditionnelle des prévisions sachant les observations suit la loi normale ou gamma. Sur ce point Rings *et al.* (2012) ont récemment proposé une modification à la BMA afin de rendre la représentation de la distribution de probabilité conditionnelle plus flexible en utilisant un filtre particulaire dans un contexte gaussien. Une autre sous-catégorie de méthodes qui visent le traitement statistique de l'incertitude totale des prévisions est la famille de méthodes de régression linéaires et non-linéaires (Gneiting *et al.*, 2005 ; Hamill *et al.*, 2008 ; WIlks, 2009 ; Weerts *et al.*, 2011).

Dans la plupart des méthodes, quelle que soit l'approche adoptée, la loi de probabilité utilisée pour décrire les erreurs de prévision n'est pas toujours connue ou facile à estimer. Les transformations " normalisatrices " des données sont souvent appliquées pour ajuster une loi normale, plus facile à manipuler, aux erreurs de prévisions. Des techniques alternatives (paramétrées et non-paramétrées) sont proposées à partir : d'une estimation bayésienne

linéaire optimale des variables indicatrices (Brown et Seo, 2010), des fonctions copules (Madadgar *et al.*, 2012), des transformations dite -" *wavelet* " - (Bogner et Pappenberger, 2011), ou de " méthodes de noyaux " (Roulston et Smith, 2003 ; Wang et Bishop, 2005 ; Fortin *et al.*, 2006 ; Boucher *et al.*, 2012).

Naturellement, l'objectif des deux approches principales est de rendre les prévisions plus précises et fiables en utilisant l'information contenue dans les observations. Pour chaque approche, il y a plusieurs groupes de méthodes de traitement statistique qui ont été proposés dans la littérature. Pour les méthodes déjà testées en hydrologie, celles-ci ont été appliquées soit sur un petit échantillonnage de bassins versants (Seo *et al.*, 2006 ; Hashino *et al.*, 2006 ; Olsson et Lindstrôm, 2008 ; Reggiani *et al.*, 2009 ; Fortin *et al.*, 2006 ; Brown and Seo, 2010, 2012 ; Weerts *et al.*, 2011 ; Boucher *et al.*, 2012), soit pour évaluer la capacité d'un système de prévision lors d' événements particuliers (Pappenberger *et al.*, 2008).

Parmi les techniques existantes, nous avons choisi d'appliquer des approches empiriques dites " méthodes d'habillage " ou " méthodes de noyaux ", qui, de manière générale, font appel à la loi de distribution des erreurs de prévision observées dans le passé pour corriger l'incertitude prédictive (Smith *et al.*, 1992 ; Roulston et Smith, 2003 ; Wang et Bishop, 2005 ; Fortin *et al.*, 2006 ; Hashino *et al.*, 2006 ; Olsson and Lindström, 2008 ; Kang *et al.*, 2010 ; Weerts *et al.*, 2011 ; Boucher *et al.*, 2012 ; Pagano *et al.*, 2012).

Ces approches peuvent corriger soit chaque membre de l'ensemble séparément, soit la distribution prédictive de l'ensemble. La stratégie de ces méthodologies se résume néanmoins au même principe : la correction des prévisions en utilisant une série historique des erreurs observées dans une période passée (séries historiques de prévisions archivées ou de prévisions rétrospectives). Leur avantage principal se résume à la simplicité de leur application. Même si les approches bayésiennes proposent un cadre théorique solide pour l'estimation de l'incertitude prédictive, néanmoins leur implémentation n'est pas toujours facile.

Dans notre étude, nous avons également appliqué une technique d'habillage empirique traitant spécifiquement les erreurs du modèle hydrologique. Enfin, des méthodes d'habillage spécifiquement destinées à traiter les prévisions de précipitation qui seront utilisées en entrée des modèles hydrologiques ont également été examinées. L'ensemble des méthodes de traitement statistique que nous avons étudiées est présentée en détail ci-après.

3.3.2 Méthodes appliquées aux prévisions de précipitation

3.3.2.1 Méthode empirique par facteur de correction

Les méthodes empiriques par facteur de correction sont des méthodes simples, souvent utilisées intuitivement par les services opérationnels. Ces approches de débiaisage prennent la forme d'une correction additive ou multiplicative, appliquée à chaque prévision. Le facteur de correction est en général estimé à partir de la moyenne de l'erreur calculée sur une série de paires "prévisions-observations" disponible. La correction multiplicative est la plus courante, car elle permet d'éviter la génération de valeurs négatives. Mathématiquement, elle est exprimée par :

$$b = E_f\left[\frac{o}{f}\right] \qquad \text{Eq. 33}$$

, où E [] : correspond à l'espérance mathématique,

b : est le facteur de correction multiplicatif,

o : est la valeur observée,

f : est la valeur prévue.

Le facteur de correction multiplicatif peut être évalué de manière générale ou selon des catégories prédéfinies (saisons, types de temps, classes de valeurs observées, etc.). Lors de la prévision, il sera appliqué à chaque nouvelle prévision :

$$\hat{f} = b * f \qquad \text{Eq. 34}$$

, où f : est la prévision brute (non-corrigée),

$\hat{f}$: est la prévision corrigée.

Un exemple d'une simple correction multiplicative, proposée à l'origine par Smith *et al.* (1992), est présenté par Hashino *et al.* (2006) dans un cadre d'application à des prévisions mensuelles de volumes écoulés dans le bassin Des Moines à Iowa. Dans leur étude, l'hypothèse principale est que toutes les prévisions ont le même biais multiplicatif à chaque fois que la même séquence climatique est utilisée pour générer les ensembles. En conséquence, le facteur de correction appliqué à la prévision brute dépend seulement de la séquence climatique (c'est-à-dire, du mois et de l'année en question).

Hashino *et al.* (2006) proposent également une approche équivalente. Il s'agit de corriger (de façon multiplicative ou additive) soit les moments de la distribution prédictive, et plus particulièrement la moyenne et l'écart-type (moments de 1er et 2ème ordre, respectivement), soit les quantiles de la distribution prédictive. Cette méthode consiste à remplacer la variable prévue par une variable observée avec la même probabilité au non-dépassement.

D'autres auteurs comme Olsson et Lindstrôm (2008) proposent une méthodologie semblable qui consiste à ajuster les quantiles 25% et 75% de la distribution prédictive de l'ensemble, ainsi que les valeurs minimales et maximales des prévisions de l'ensemble, en les multipliant par des coefficients correctifs calculés sur une période de calage pour chaque bassin versant pris individuellement.

Dans cette thèse, nous avons examiné la méthode de correction par facteur multiplicatif pour corriger des biais des précipitations prévues (voir Chapitre 6, §6.3). Plus particulièrement, nous appliquons une correction multiplicative sur chaque membre de la prévision d'ensemble calculée sur une période de calage.

L'équation Eq. 35 décrit mathématiquement l'erreur de la moyenne de la prévision d'ensemble ε par bassin versant *ibv* en fonction du jour *i* et du membre *k* pour la première échéance de prévision.

$$\varepsilon_{ibv,i,1} = \left(\frac{1}{K}\sum_{k=1}^{K} f_{ibv,i,1,k}\right) - o_{ibv,i,1} \qquad \text{Eq. 35}$$

Une fois les erreurs de la moyenne de la prévision d'ensemble calculées par jour *i*, nous constituons une archive de ces erreurs pour chaque bassin versant. Le facteur correctif s'exprime en pourcentage d'augmentation ou de réduction de la valeur prévue. Le calage de ces facteurs correctifs se fait année par année sur l'ensemble de la période 2005-2008 et leurs valeurs moyennes sont retenues à la fin. Pour l'instant le même facteur correctif est appliqué pour toutes les échéances *j*.

Les principaux avantages de la méthode de correction par facteur multiplicatif sont :

- sa simplicité,
- l'élimination du biais inconditionnel,
- la bonne performance par rapport au biais conditionnel et l'amélioration de la dispersion des membres quand la méthode est appliquée aux moments ou aux quantiles de la distribution prédictive.

La principale limitation de la méthode est :

- la dépendance à la longueur de la période de calage, notamment quand les facteurs de correction doivent être évalués sur des sous-échantillons de l'échantillon total (manque de robustesse des estimations si la longueur n'est pas suffisante).

3.3.2.2 Méthode d'habillage à l'aide des situations analogues

Dans cette thèse, nous avons pris en main l'archive disponible à EDF pour la prévision par analogie. L'information extraite par les archives des analogues est utilisée afin de produire des prévisions de précipitation. Les variables météorologiques, prises dans la recherche de situations analogues, sont les champs d'altitude des niveaux de pression 700hPa et 1000hPa à 0h00 et 24h00, qui correspondent en moyenne à une altitude de 3000m et 100m, respectivement. Afin de définir la ressemblance entre la situation prévue et les situations observées dans le passé, un critère de distance est utilisé (Score de Teweles-Wobus, 1954) et les 50 situations analogues qui ont les meilleurs scores sont retenues.

La méthode par analogie fonctionne de la même façon dans le cadre des prévisions déterministes et probabilistes. La seule différence repose sur le nombre de scénarios initiaux à la base desquels se fait la recherche des situations analogues. Dans cette thèse, nous avons suivi la procédure utilisée à EDF, qui prévoit la sélection de 50 situations analogues par scénario initial de prévision. La méthode a été appliquée à la prévision d'ensemble du CEPMMT. Chaque scénario initial est formé par un des 50 membres de la prévision d'ensemble. L'application de la méthode des analogues a ainsi généré un total de 2500 membres ou scénarios de prévision (50 membres x 50 situations analogues) par jour de prévision.

Derrière chaque scénario nous avons ainsi une paire composée d'une situation analogue et du membre correspondant de la prévision d'ensemble qui a déclenché la recherche de la situation analogue. Une correction est ensuite appliquée à chaque membre de la prévision d'ensemble, et ce pour chaque échéance de prévision. Elle se traduit par :

$$\widehat{f} = \beta * f^{ECMWF} (f^{ANALOG} / \bar{f}^{ANALOG})^{\alpha} \quad \text{Eq. 36}$$

, où $\hat{f}$: est la prévision corrigée du membre de la prévision d'ensemble,

f^{ECMWF} : est la valeur de la prévision brute (avant correction) du membre de la prévision d'ensemble,

f^{ANALOG} : est la valeur de la variable de prévision donnée par la situation analogue,

$\bar{f}^{ANALOG}$: est la moyenne des 50 valeurs de la variable de prévision, données par les 50 situations analogues associées au membre de la prévision d'ensemble qui est corrigé,

α , β : sont des paramètres libres de la correction (paramètres à caler).

Les valeurs de paramètres α et β peuvent varier selon le bassin versant et l'échéance de prévision. Dans cette thèse, nous avons utilisé les valeurs qui étaient utilisées par le système de prévision opérationnelle à EDF. Il s'agit de valeurs fixes et égales à 0.3 et 1.2, pour les paramètres α et β, respectivement.

Etant donné le grand nombre de scénarios (2500) générés par la méthode, ce qui correspond à un temps de calcul proportionnellement important, un ré-échantillonnage est souvent appliqué dans les procédures opérationnelles à EDF. La technique de ré-échantillonnage consiste à trier les 2500 scénarios obtenus en ordre croissant et, parmi eux, choisir 50 scénarios correspondant à des quantiles équidistants dans la distribution empirique obtenue. Ces 50 scénarios de prévision de précipitation sont ensuite utilisés en entrée du modèle hydrologique.

Les principaux avantages de l'habillage des prévisions avec l'information de situations analogues sont :

- sa simplicité d'application, notamment quand la méthode est déjà mise en place,
- la bonne performance par rapport à la correction des biais conditionnels et non-conditionnels,
- la correction de la dispersion des membres.

Les principales limitations se résument aux points suivants :

- les scénarios originaux de la prévision d'ensemble (les membres de la prévision d'ensemble) ne sont pas forcément tous gardés par la méthode,
- le pré-requis d'avoir une archive la plus longue possible disponible pour la recherche des situations analogues.

3.3.3 Méthodes appliquées aux prévisions de débits

3.3.3.1 Méthode du meilleur membre et ses variantes

Une approche typique de la famille des techniques d'habillage est la méthode du meilleur membre. Développée initialement par Roulston et Smith (2003), elle a subi des modifications proposées par Wang et Bishop (2005) et par Fortin *et al.* (2006). Cette méthode propose d'assigner à chaque membre de l'ensemble une distribution de probabilité d'erreur, estimée à l'aide des erreurs du meilleur membre de l'ensemble. Le meilleur membre est celui qui a la plus petite erreur absolue par rapport à l'observation. Lors du calage de la méthode, ces erreurs minimales sont évaluées et une distribution de probabilité est ajustée aux valeurs obtenues. Cette distribution servira pour le tirage aléatoire des erreurs qui iront corriger les prévisions d'ensemble brutes lors de l'application de la méthode. L'hypothèse souvent adoptée est que les erreurs du meilleur membre suivent une loi de probabilité normale. Afin d'assurer que cette hypothèse soit respectée, une transformation logarithmique des valeurs prévues et observées est appliquée lors du calcul de l'erreur du meilleur membre.

L'ensemble corrigé est alors obtenu par :

$$\widehat{f}_{i,k,n} = f_{i,k} * \exp(\varepsilon_{i,k,n}) \qquad \text{Eq. 37}$$

, où $f_{i,k}$: est la valeur de la prévision brute (avant correction) du membre *k* de la prévision d'ensemble à l'instant *i,*

$\widehat{f}_{i,k,n}$: est la valeur de la prévision corrigée du membre *k* de la prévision d'ensemble à l'instant *i,* habillé avec l'erreur, correspondant au tirage *n* aléatoire de la distribution d'erreur estimée en calage*,*

$\varepsilon_{i,k,n}$: est l'erreur qui est attribuée à la prévision brute*,* correspondant au tirage *n* aléatoire de la distribution d'erreur estimée en calage.

Une des limites de l'approche originalement proposée par Roulston et Smith (2003) est qu'elle présente parfois une efficacité limitée dans la correction des ensembles sous-dispersés, car elle n'arrive pas à bien capturer la variance des observations. Wang et Bishop (2005) ont proposé une première modification pour augmenter la correction de la dispersion dans le cas des ensembles sous-dispersés. Ils ont proposé d'habiller les ensembles avec une contrainte posée sur le moment de deuxième ordre (variance ou covariance dans le cas multi-varié).

En pratique, il s'agit de multiplier la distribution de probabilité d'erreur par un facteur de réduction *ω*. L'objectif est qu'en multipliant par le facteur de réduction *ω* nous arrivons à une distribution d'erreur avec une moyenne égale à zéro et une variance égale à la covariance entre la moyenne des membres de l'ensemble et l'observation. Le facteur de réduction est donné par :

$$\omega = \sqrt{\frac{s^2}{s_\varepsilon{}^2}} \qquad \text{Eq. 38}$$

, où s^2 est la covariance entre la moyenne d'ensemble et l'observation,

$s_\varepsilon{}^2$ est la variance des membres de l'ensemble.

Si l'ensemble à corriger est sur-dispersé, *ω* prend des valeurs inférieures à 1. Par contre, si l'ensemble est sous-dispersé à l'origine la valeur de ω sera supérieure à 1.

Fortin *et al.* (2006) ont proposé une deuxième modification de la méthode du meilleur membre, en introduisant le facteur de réduction comme un paramètre de calage qui prendra des valeurs différentes pour chaque membre de la prévision d'ensemble et selon sa position (ou rang) par rapport au meilleur membre. Les erreurs du meilleur membre sont séparées en sous-groupes selon la position du meilleur membre au sein de l'ensemble. Par la suite un vecteur d'erreurs est formé pour chaque rang. Le but était de pouvoir ainsi donner un poids de correction différent ω_k à chaque membre *k* de la prévision d'ensemble, en fonction de sa position par rapport au membre le plus proche de l'observation. Cette approche utilise une fonction de paramétrage avec un seul paramètre *α*, qui est calé avec l'objectif de mieux rapprocher la variance des prévisions à la variance des observations et de mieux estimer les probabilités des événements extrêmes.

$$\omega_k \propto \int_{\frac{k-1}{K}}^{\frac{k}{K}} u^{\alpha-1}(1-u)^{\alpha-1}du \qquad \text{Eq. 39}$$

Dans cette thèse, nous avons examiné la méthode originale proposée par Roulston et Smith (2003) et ses modifications proposées par Wang et Bishop (2005), Fortin *et al.* (2006), en les appliquant à la correction des prévisions de débits. Ces approches sont d'abord appliquées en météorologie (températures et précipitations) et ensuite Boucher *et al.* (2012) les ont appliquées en hydrologie sur des bassins versants québécois avec des prévisions d'ensemble fournies par le centre météorologique canadien ou avec des données synthétiques.

Les principaux avantages de la méthode du meilleur membre et ses modifications sont :

- la simplicité d'application,
- la bonne correction de la dispersion dans les cas des ensembles sous- et sur-dispersés (ceci dépendant de l'approche appliquée),
- la correction du biais conditionnel de Type I.

Les principales limitations sont :

- la faible correction du biais inconditionnel,
- la faible correction de la capacité des prévisions à discriminer entre les événements,
- la dépendance à la longueur de la période du calage pour l'estimation des paramètres de la distribution de probabilité de l'erreur, notamment lorsque des paramètres doivent être estimés à partir de sous-échantillons de la série disponible.

3.3.3.2 Modifications méthodologiques de la méthode du meilleur membre

Nous avons apporté trois modifications méthodologiques supplémentaires à la méthode du meilleur membre originale et nous les avons testées sur nos données.

Dans un premier temps, une discrétisation par quantile des débits observés a été introduite. Au lieu d'avoir une distribution unique des erreurs de prévision, une distribution est estimée pour chaque gamme de débits étudiée (ici, les quantiles 10%, 20%, …., 90% des débits observés). De cette façon, nous avons espéré améliorer la performance de la méthode en prenant en compte les différences entre les propriétés statistiques des erreurs commises en, par exemple, bas, moyen ou hauts débits. Toutefois, il faut noter qu'une importante limitation de cette modification par quantiles est la longueur restreinte de la période de calage, en particulier pour les quantiles extrêmes pour lesquels il y a peu de données disponibles. Par conséquent, étant donné que nous avons à notre disposition que 4 ans de prévisions, le calage des paramètres de la distribution des erreurs par gamme de débits n'était pas toujours évident et l'amélioration apportée par rapport à la méthode originale restait limitée.

Dans un deuxième temps, nous avons voulu répondre à un problème fréquent dans une grande majorité de méthodes de correction statistique de biais : le maintien de la cohérence temporelle entre les échéances de prévision. L'application d'un traitement statistique par habillage empirique ne garantit pas que les trajectoires des membres seront conservées au long des échéances. L'habillage est en effet souvent calé et appliqué à des échéances fixes, prises séparément. Néanmoins, quelques études ont essayé de répondre à cette

problématique dans un contexte de prévision (Krzysztofowicz et Maranzano, 2004 ; Coccia et Todini, 2011 ; Zhao *et al.*, 2011).

Pour répondre à cette question dans cette thèse, tout en gardant une approche simple d'habillage par la méthode du meilleur membre, nous avons remplacé dans la méthode modifiée par Wang et Bishop (2005) la variance des erreurs du meilleur membre σ^2 à l'échéance *j* par la covariance des erreurs du meilleur membre entre les échéances *j* et *j-1*. Nous avons également introduit une modification similaire à la méthode originale (Roulston et Smith, 2003). L'expression mathematique de la modification proposée est donnée dans l'equation suivante :

$$\sigma^2 = \overline{(\bar{f}_j - o_j)(\bar{f}_{j-1} - o_{j-1})} - (1 - \frac{1}{N})\overline{(\sum(\bar{f}_j - f_j)(f_{j-1} - f_{j-1}))} \quad \text{Eq. 40}$$

Dans un troisième temps, nous avons examiné une modification méthodologique simple de la méthode du meilleur membre, consistant à prendre en compte l'erreur de plusieurs membres et non pas seulement du meilleur membre. En effet, parfois la méthode du meilleur membre n'arrive pas à cerner l'incertitude totale des prévisions et les prévisions d'ensemble corrigées restent sous-dispersées. Ceci est notamment le cas quand le meilleur membre est très proche de l'observation, conduisant à des erreurs faibles de prévision.

Pour cette raison, et afin d'élargir le faisceau des ensembles corrigés, nous avons introduit dans la construction de la distribution des erreurs de prévision, l'erreur de plusieurs membres. Pour définir le nombre optimal de " meilleurs membres " à considérer, nous avons au préalable conduit quelques tests préliminaires. De ces tests, l'utilisation des 30% des " meilleurs membres " (c'est-à-dire, 15 membres sur un ensemble de 50 membres) a été retenue. Similairement à la méthode du meilleur membre originale, decrite dans le paragraphe précédent, une distribution de probabilité est ajustée aux valeurs des erreurs obtenues. Cette distribution servira à nouveau pour le tirage aléatoire des erreurs sous l'hypothèse que les erreurs de 30% des meilleurs membres suivent une loi de probabilité normale.

En plus des principaux avantages de la méthode du meilleur membre présentés précédemment, nous espérons que les modifications méthodologiques proposées apporteront :

- une cohérence temporelle améliorée,
- une amélioration en termes du biais inconditionnel,

- une correction améliorée du biais conditionnel de Type I,
- une correction de la capacité de discrimination des prévisions,
- une correction améliorée de la dispersion des membres.

Néanmoins, nous sommes conscients que les principales limitations restent les mêmes, notamment celle liée à la forte dépendance à la longueur de la période de calage.

3.3.3.3 Méthode empirique d'habillage avec les erreurs de modélisation

Nous avons également examiné une approche de traitement statistique qui prend en compte exclusivement les erreurs liées au modèle hydrologique de prévision. Ce type d'approche est une méthode qui peut être assimilée à la famille des méthodes empiriques d'habillage des prévisions (Mathevet, 2010 ; Pagano *et al.,* 2012), ayant pour but spécifique la prise en compte de l'incertitude de modélisation hydrologique dans les prévisions de débits.

La particularité de cette technique réside dans le fait que, au lieu d'évaluer les erreurs commises en situation de prévision, c'est-à-dire, à partir d'une archive de paires " prévisions-observations ", nous évaluons la distribution de probabilité de l'erreur de simulation du modèle hydrologique, c'est-à-dire, quand ce dernier est alimenté par des prévisions dites " parfaites " (la pluie prévue est égale à la pluie observée).

Le post-traitement des prévisions de débits est donc basé sur la modélisation de l'erreur empirique du modèle hydrologique et cette modélisation peut être faite par gamme (ou catégorie) de débits et par échéance de prévision. Dans cette thèse, nous avons suivi la méthode appliquée dans Mathevet (2010), utilisée déjà opérationnellement à EDF. Les gammes de débits ont été définies par des quantiles choisis par un pas de 8% pour les quantiles entre 10%-90% et par un pas de 2% pour le restant des quantiles. Ce choix de découpage par gamme de débits vise à obtenir un nombre limité (20) de sous-populations pour lesquelles les caractéristiques statistiques de l'erreur doivent être estimées.

Pour chaque gamme de débits et pour chaque échéance, les débits simulés par le modèle sur une longue série de simulations passées sont comparés aux débits observés. Les erreurs commises par le modèle hydrologique sont normalisées à partir d'une transformation logarithmique des ratios entre débits observés et débits prévus. Chaque erreur est ainsi calculée par la formule :

$$\varepsilon = \ln\left(\frac{Q}{\hat{Q}}\right) \quad \text{Eq. 41}$$

, où ε : est l'erreur du modèle,

Q : est le débit observé,

$\hat{Q}$: est le débit simulé.

La moyenne et l'écart-type des erreurs obtenues sont ensuite utilisés pour caler une loi normale à la distribution de probabilité de l'erreur. Le tirage aléatoire de plusieurs valeurs d'erreurs permet ensuite de corriger chaque membre de la prévision hydrologique d'ensemble. Cette correction se fait de la même manière que celle présentée dans l'Eq. 45. La correction est appliquée pour chaque gamme de débit et pour chaque échéance de prévision.

Les principaux avantages de cette méthode sont :

- la simplicité d'application,
- la correction du biais inconditionnel,
- la correction du biais conditionnel de Type I.

Les principales limitations sont :

- la faible correction de la capacité des prévisions à discriminer entre les événements,
- l'introduction parfois d'une dispersion non-appropriée (trop importante),
- la nécessité d'une longue archive des erreurs passées du modèle hydrologique.

3.3.4 Approches pour le calage des paramètres des méthodes appliquées

La plupart des approches testées lors de cette thèse sont paramétriques – surtout bi-paramétriques quand on considère l'estimation de la moyenne et de l'écart-type des distributions normales des erreurs de prévision. La procédure habituelle pour estimer les paramètres libres d'une méthode consiste à diviser les séries temporelles disponibles en deux périodes indépendantes : période de calage et période de validation (ou de contrôle). Les études en hydrologie s'appuient souvent sur cette méthodologie, connue par le nom de " split-sample test " et introduite en hydrologie par Klemeš (1986). Par cette approche, les périodes de calage et de validation sont souvent de même longueur pour que les résultats provenant de ces deux sous-échantillons puissent être comparables. La faiblesse de toute procédure de calage/ contrôle est notamment liée à la taille de l'échantillon : plus la série de données disponible est longue, plus performante sera le calage des paramètres.

De plus, si les deux périodes sont très contrastées climatologiquement, la performance en validation peut s'avérer très faible.

Une façon de contourner ce problème consiste à effectuer une validation croisée. Lors des travaux de cette thèse nous avons testé l'approche de validation croisée dite " leave one out " (Kearns et Ron, 1999). Le principe de cette approche est d'enlever de la période entière disponible une sous-période sur laquelle nous effectuons la validation de l'approche. Le calage est ainsi effectué sur le reste de la période.

Dans un premier temps nous avons divisé la période entière en deux sous periodes égales de 24 mois. Nous avons effectué le calage sur la première période et la validation sur la deuxième, suivie du calage sur la deuxième période et de la validation sur la première. De ce fait, nous disposons d'une période plus longue de résultats de validation (24+24=48 mois).

Nous avons également appliqué la technique de validation croisée en divisant la période entière (48 mois) en sous-périodes de trois mois. Nous avons calé les paramètres de différentes méthodes sur les 45 mois restants et nous avons fait la validation sur les sous périodes de 3 mois. Nous avons appliqué cette technique en répétition pour toutes les sous-périodes (fenêtres glissantes) de façon à ce que la validation et le calage soient faits sur la période entière.

Cependant nous avons constaté que pour notre jeu de données la différence en termes d'estimation de paramètres entre les deux divisions étudiés (24 et 3 mois) était presque négligeable. De ce fait, dans le Chapitre 6, nous présentons les résultats de la validation croisée appliquée à deux sous-périodes de 24 mois.

Le Tableau 5 illustre les différentes périodes utilisées selon la méthode de correction statistique appliquée. Pour la famille de méthodes du meilleur membre – étant donné que ces approches utilisent la série temporelle des erreurs de prévision, qui est bien plus courte que les séries historiques d'observations et simulations des modèles – la validation croisée est appliquée avec des périodes de 24 mois. Pour l'habillage avec les analogues et avec l'erreur du modèle hydrologique, le calage s'est effectué sur des périodes plus longues grâce à la disponibilité des archives respectives.

Tableau 5 : Périodes de calage et de validation pour chaque approche de traitement statistique testée.

Approche de traitement statistique	Période de calage	Période de validation
Meilleur membre et ses modifications ; autres approches empiriques	01/01/2005-31/12/2006	01/01/2007-26/12/2008
	01/01/2007-26/12/2008	01/01/2005-31/12/2006
Habillage avec les analogues	01/01/1953*-31/12/2007	01/01/2005*-31/12/2008
Habillage avec l'erreur du modèle hydrologique	01/01/1958*-31/12/2004	01/01/2005-26/12/2008

* selon le bassin versant

3.4 Synthèse

Ce chapitre a présenté les principaux concepts autour de l'évaluation de la qualité des prévisions et des méthodes de traitement statistique pour la correction des biais de prévision. Nous avons détaillé les critères et méthodes qui seront utilisées dans les prochains chapitres pour évaluer et corriger les prévisions hydrologiques d'ensemble des échantillons de bassins versants étudiés.

La présentation faite n'est sûrement pas exhaustive, car des nombreux critères et méthodes existent dans la littérature et sont continuellement développés pour améliorer les systèmes de prévision hydrométéorologiques. Nous retiendrons que quels que soient les avantages et les faiblesses de chaque approche, il est essentiel de garder à l'esprit que l'objectif d'un traitement statistique des prévisions est de rendre les prévisions plus fiables, plus justes et précises en termes de volumes prévus, ou de temps et de pointe des crues. La question clé reste à savoir s'il existe une méthode plus performante que toutes les autres et si ceci est valable pour tous les bassins versants, toutes les saisons et tous les événements de prévision. Bien évidement, il se peut que cette question reste rhétorique, néanmoins c'est cette quête qui inspire plusieurs travaux d'application et de développement de méthodes. Dans cette thèse, nous nous intéressons également à cette question et, plus particulièrement, nous cherchons à comprendre l'impact que l'amélioration de la qualité des prévisions peut avoir sur sa valeur économique, face à une utilité définie par un utilisateur-décideur.

Dans le prochain chapitre (Chapitre 4), nous nous penchons sur les notions d'utilité des prévisions et sur les méthodes existantes pour mesurer la valeur des prévisions. Nos développements et résultats numériques sont ensuite présentés dans le Chapitre 5, Chapitre 6 et Chapitre 7 pour ce qui est, respectivement, de la qualité, du traitement statistique et la valorisation économique des prévisions.

"Gérer, c'est prévoir"

4. L'utilité d'une prévision

Dans ce chapitre, nous présentons une autre caractéristique des prévisions d'ensemble : l'utilité, qui est souvent mesurée en termes monétaires (gains ou pertes). Ceci dit, la maximisation d'un gain (ou la minimisation des coûts) n'est pas l'objectif majeur des centres de prévision, qu'ils soient responsables de la protection des populations et biens ou de la gestion de la ressource en eau. Le recours aux techniques d'optimisation est souvent nécessaire pour aider la prise de décision dans le contexte probabiliste et incertain de la prévision. Un résumé des techniques les plus courantes est présenté ici, accompagné d'une synthèse des études menées en hydrologie pour évaluer la valeur économique des prévisions. Ce chapitre introduit la démarche entreprise dans cette thèse pour la construction d'un outil de gestion et d'évaluation économique des prévisions, présenté dans le Chapitre 7.

4.1 Pourquoi associer une valeur économique (utilité) aux prévisions ?

Nous avons vu dans le Chapitre 1 que la prévision d'ensemble est actuellement considérée comme "l'état de l'art" en matière de prévision probabiliste en hydrométéorologie. Plusieurs services opérationnels - tant d'intérêt public que d'intérêt privé - ont décidé de produire et/ou utiliser ces prévisions (Cloke et Pappenberger, 2009b). Le but est, d'une part, de pouvoir étendre l'échéance de prévision et, d'autre part, de fournir une quantification des incertitudes de prévision à partir des différents scénarios qui sont proposés par la prévision d'ensemble.

En effet, une meilleure anticipation des événements extrêmes permettrait d'avertir les populations en avance pour une meilleure protection des vies humaines et des biens, de même qu'elle contribuerait à mieux gérer les risques et les ressources en eau utilisées par divers secteurs économiques (par exemple, hydroélectricité, distribution d'eau potable, irrigation, navigation, etc.). Palmer (2002) indiquait déjà dans son article sur l'avenir de la prévision d'ensemble que " (…) si l'on développait l'usage du couplage entre les sorties de la prévision d'ensemble et l'entrée de modèles d'application, on se rendrait compte de la véritable valeur de la prévision d'ensemble en tant qu'outil de gestion du risque ".

La chaîne typiquement suivie en prévision pour la gestion du risque hydrologique relie le système d'émission des prévisions hydrométéorologiques à une étape de prise de décision qui, de son côté, alimente un système d'alerte ou d'actions de gestion de la ressource (Figure 21).

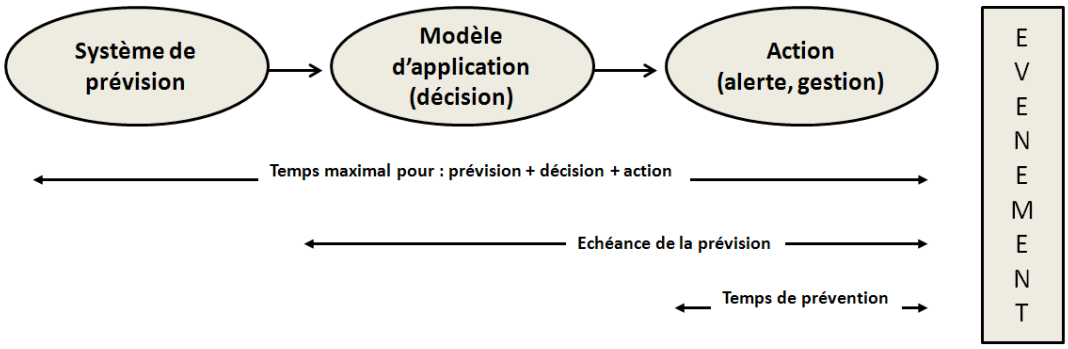


Figure 21 : La chaîne de la prévision du point de vue de l'utilisateur à l'arrivée d'un événement (d'après Parker et Fordham, 1996).

Le temps maximal disponible pour accomplir ces étapes est défini par l'intervalle de temps entre le moment où le système de prévision initie ses simulations et le moment où l'événement a lieu. L'échéance de prévision est l'intervalle de temps entre le moment où les prévisions sont émises et le moment ou l'événement a lieu, tandis que le temps de prévention est celui entre le moment où l'action est déclenchée et l'arrivée de l'événement. Il est donc dans l'intérêt de chaque service d'utiliser efficacement les temps disponibles pour disposer du maximum de temps possible pour l'action de prévention envisagée.

La Figure 21 montre que le temps maximal disponible pour l'action complète, jusqu'à l'arrivée d'un événement est fonction de l'horizon de prévision (échéance), est une fonction du temps passé pour produire les prévisions et du temps passé pour prendre la décision. L'utilité d'une prévision d'ensemble est intrinsèquement liée à ces aspects : la prise en compte des incertitudes de prévision permet d'allonger les échéances du système de prévision et de : i) soit fournir aux décideurs plus de temps pour qu'ils puissent prendre leur décision (à condition que les informations soient bien communiquées, voir les discussions dans Demeritt *et al.*, 2007 ; Ramos *et al.*, 2010, 2012, par exemple), ii) soit augmenter le temps de prévention. De même, si on diminue le temps passé à prendre la décision, on augmente le temps de prévention. D'où l'importance des modèles d'application qui fournissent de l'aide à la prise de décision.

La question qui se pose ici est : " Est-ce qu'à l'aide de l'information probabiliste fournie par les prévisions d'ensemble, les utilisateurs peuvent en effet mieux décider des actions à mener pour maximiser les gains ou limiter les pertes liés à l'impact d'un événement à venir sur leurs activités ? ". Dans le but d'apporter une réponse à cette question, des études ont été menées pour évaluer le gain économique (effectif ou potentiel) apporté par l'utilisation de la prévision probabiliste. L'évaluation objective de l'utilité d'une prévision sur le long terme fait appel à des techniques économiques, qui cherchent à proposer une mesure de l'utilité, " valeur économique ", d'une prévision pour aider à la prise de décision.

4.2 Utilité d'une prévision et prise de décision

Selon la théorie économique de la valeur d'une information, la valorisation d'une information est, en réalité, une estimation de l'utilité de l'information en termes monétaires dans un contexte décisionnel (Pratt, 1964 ; Arrow, 1965). Nous faisons ainsi indirectement référence à la théorie de l'information et plus particulièrement à une autre branche des mathématiques appliquées, la théorie de la décision (von Neumann et Morgenstern, 1944 ; Winkler, 1972).

Supposons que *e* dans {1,…, *N*} sont les différents événements possibles pour le jour *i* et *a* dans {1,…,*M*} sont les différentes actions possibles. Chaque action *a* un certain coût et/ou gain. Dans ce cas, l'ensemble des conséquences *c* de chaque décision prise est une fonction des événements *e* et des actions *a* : *c* = *c*(*e*, *a*). L'utilité des prévisions *u* est estimée en fonction des conséquences des actions considérées (c'est-à-dire, du gain et/ou coût qui les accompagne) : *u* = *u*(*c*). Les conséquences de chaque action sont évaluées sur la base de la maximisation (ou minimisation) de l'espérance du gain (ou du coût). L'utilisateur est alors invité à prendre la décision *D* avec l'utilité maximale :

$$D(u[c(a,e)]) = \max(E\left[u\left[c(a,e)\right]\right]) \text{ , si gain} \quad \text{Eq. 42}$$

$$D(u[c(a,e)]) = \min(E\left[u\left[c(a,e)\right]\right]) \text{ , si coût} \quad \text{Eq. 43}$$

Une fois la fonction d'utilité définie quantitativement, la meilleure façon de l'optimiser est de passer par des décisions rationnelles, qui prennent en compte toutes les informations disponibles. Néanmoins, très souvent les informations disponibles sont imparfaites et/ou manquantes. Ce fait génère de l'incertitude. Dans ce contexte incertain, l'utilisateur doit prendre une décision sans avoir à sa disposition des observations, mais seulement des estimations. Il doit ainsi choisir l'option ayant l'utilité espérée la plus élevée. Ce principe est à la base de la théorie bayésienne de la décision.

Dans ce contexte, les utilisateurs attribuent une probabilité subjective *a priori* à chaque scénario d'action. Dans le cas d'une décision prise avec l'utilité maximale (Eq. 42) l'utilité est donnée par le maximum du produit de cette probabilité *a priori* :

$$D(u[c(a,e)]) = \max(\sum_{e} p_{apriori} * u[c(a,e)]) \quad \text{Eq. 44}$$

En pratique, la probabilité subjective *a priori* est souvent estimée à partir des données historiques (la climatologie, par exemple). Néanmoins, elle pourrait être modifiée avec l'acquisition d'informations supplémentaires, comme, par exemple, celles apportées par les prévisions d'ensemble f {1,…,*K*}, où *K* sont les différents scénarios de la prévision d'ensemble. Ces informations supplémentaires aident l'utilisateur à réviser : (a) la probabilité subjective *a priori* définie initialement et (b), en conséquence, sa décision optimale. La prise en compte d'une information complémentaire dans une prévision fait apparaître la notion de probabilité conditionnelle.

En effet, un système de prévision probabiliste a pour but de fournir à l'utilisateur la probabilité de l'événement *e*, sachant les informations supplémentaires *f*, p*(e|f)*. Par le Théorème de Bayes, ceci s'écrit en fonction de la probabilité *a priori* d'observer l'événement, *e = p(e)*, de la probabilité des informations supplémentaires *p(f)* et de la probabilité conditionnelle des informations *f* étant donné l'événement *e*, p*(f|e)* (fonction de vraisemblance) :

$$p(e|f) = \frac{p(f|e)}{p(f)} p(e) \qquad \text{Eq. 45}$$

Les systèmes de prévision probabiliste produisent directement la probabilité *p(e|f),* appelée probabilité *a posteriori*. Pour la prise de décision optimale, il conviendra de maximiser la fonction de vraisemblance, c'est-à-dire, la probabilité que la prévision représentée par *f* soit égale à l'événement *e*, sur des données d'apprentissage (paires prévision/observation). Nous revenons ici au Théorème de Bayes afin de démontrer l'utilité de l'information apportée par une prévision à la prise de décision.

En pratique, deux voies sont souvent indiquées pour identifier la meilleure décision : les études prescriptives et les études descriptives. Une étude prescriptive sert à identifier la meilleure décision sous l'hypothèse d'un décideur idéal (c'est-à-dire, juste et rationnel) et parfaitement informé (Katz et Murphy, 2000). Une analyse de décision qui suit une telle étude décrit ainsi, théoriquement, comment les décideurs *doivent* prendre leurs décisions. Les outils opérationnels de gestion sont souvent développés sur ce principe.

Or, en réalité, les décideurs s'éloignent, pour des raisons diverses, de l'optimalité. Par exemple, selon Houdant (2004) et Ramos *et al.* (2010), quand l'incertitude sur les scénarios météorologiques futurs devient très importante (en cas d'événements rares, par exemple), les prévisionnistes ont souvent du mal à imaginer un scénario météorologique radicalement différent de leur estimation *a priori.* Dans ce cas de figure, des études descriptives, qui cherchent à décrire le comportement réel des décideurs, sont mises en place. Les deux approches, prescriptives et descriptives, sont, bien évidement, liées entre elles : des hypothèses sur le comportement des décideurs peuvent être énoncées et testées face au comportement réel des décideurs devant le processus de prise de décision.

Cette thèse se situe dans le cadre des études prescriptives, dans le sens où nous nous intéressons à évaluer le gain apporté par les prévisions étant donnée l'action optimale d'un décideur, sans prendre en compte son comportement réel (nous verrons que le modèle développé dans le Chapitre 7 suit ce cheminement méthodologique).

4.3 Maximiser l'utilité pour une prise de décision optimale

L'intérêt à maximiser l'utilité d'une prévision est lié à la recherche d'une prise de décision optimale, avec moindre coût et maximum de bénéfice pour l'utilisateur des prévisions. Un problème de décision peut être un problème de décision "statique" ou "dynamique". Un *problème de décision statique* concerne des décisions isolées, qui ne s'inscrivent pas dans la durée, ou des décisions séquentielles qui ne s'affectent pas mutuellement. Un *problème de décision dynamique* implique, au contraire, des décisions séquentielles liées les unes aux autres. Il s'agit de plusieurs décisions en séquence, chaque décision courante influençant la résolution des problèmes de décision qui suivent. L'arbre de décision est un outil de représentation visuelle usuellement employé dans le cas d'un problème de décision dynamique.

La Figure 22 présente un exemple d'arbre de décision pour un problème simple de gestion d'une réserve hydraulique. Dans la figure, les carrés représentent des points de décisions et ils sont suivis par des branches. Chaque branche de l'arbre représente une décision différente. A la fin de chaque branche il y a un cercle qui représente un événement. A chacun de ces cercles les décideurs attribuent une probabilité, selon les informations dont ils disposent. Au bout de chaque branche il y a souvent des triangles qui représentent les nœuds finaux. Ce schéma illustre comment chaque décision finale est liée au chemin des décisions qui la précèdent. Par conséquent, le problème devient de plus en plus complexe quand les variables de décision se multiplient. Cet aspect est connu comme la "malédiction de dimensionnalité". A cause de cet effet, la solution analytique dans l'optimisation d'un problème de décision dynamique peut devenir difficile et les solutions numériques, nécessaires.

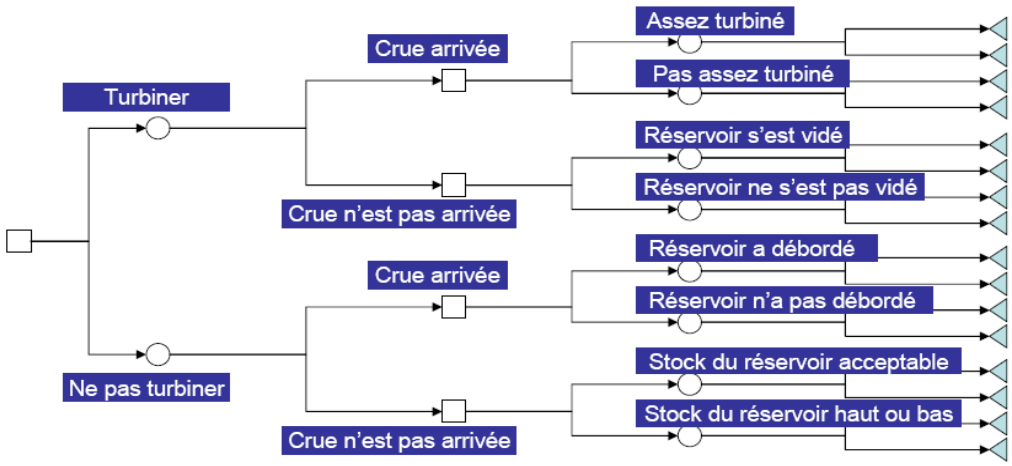


Figure 22: Illustration d'un arbre de décision.

Afin de résoudre numériquement un problème d'optimisation, plusieurs techniques de programmation sont proposées dans la littérature. Labardie (2004) présente une synthèse des approches les plus courantes et de leurs caractéristiques principales. En résumé, les techniques d'optimisation sont caractérisées par plusieurs aspects qui décrivent le problème de décision : la linéarité ou non-linéarité des équations liant les diverses variables du problème, le traitement statique ou dynamique (séquentiel) du problème, le contexte déterministe ou probabiliste (décision en présence d'incertitudes), la recherche d'une décision optimale mono- ou multi-objectifs, etc.

En contraste avec les algorithmes d'optimisation qui assurent une convergence mathématique vers la solution optimale, s'inscrivent les approches heuristiques. Une programmation heuristique propose une optimisation basée sur les connaissances empiriques et l'expérience des décideurs. Ces techniques ne garantissent pas une convergence vers la décision optimale, globale ou locale, mais proposent une solution approchée, qui souvent accélère le processus de résolution.

Malgré leur simplicité, les approches heuristiques peuvent néanmoins être compétitives par rapport à des algorithmes de programmation simples. Ilich (2001), par exemple, propose le remplacement de la programmation linéaire par des méthodes heuristiques dans les modèles d'optimisation de la gestion des aménagements pour certains bassins versants. Néanmoins, la simplicité des algorithmes heuristiques ne cache pas leur défaut principal qui est leur manque de généralité. Conçues pour des problèmes particuliers, les approches heuristiques ont souvent une structure propre au problème traité, même si des principes plus généraux sont utilisés pour leur construction.

Dans cette thèse, nous allons explorer l'approche heuristique pour construire notre modèle de valorisation des prévisions dans la gestion des retenues destinées à la production hydroélectrique, présenté dans le Chapitre 7.

4.4 Comment mesurer l'utilité des prévisions ?

Weijs (2011) décrit l'information comme un bien économique étrange : une fois produite, elle n'a plus aucun coût marginal, mais possède encore une valeur pour celui qui la reçoit. La valeur d'une prévision hydrométéorologique peut varier en fonction de son utilisateur, notamment des coûts, pertes ou bénéfices qu'il peut subir quand cette information est (ou n'est pas) disponible.

Pour mesurer la valeur économique d'une information de prévision, des modèles de décision liant les événements prévus et observés, les actions des décideurs et leurs conséquences doivent être mis en place. Ces modèles peuvent être basés sur des objectifs divers : par exemple, minimiser les fausses alertes pour la prévision des crues, maximiser la production d'énergie d'une réserve en eau, minimiser les coûts d'une mesure de protection, etc.

Un modèle de décision simple et très répandu dans la littérature est le *modèle Coût-Pertes* (Thompson, 1952). Ce modèle cherche à estimer la relation entre le coût d'une action et les pertes subies au cas où un événement dommageable ait lieu sans qu'une action soit prise. La valeur économique est calculée en fonction du rapport coût/perte (C/L, pour "cost/loss" en anglais) de l'utilisateur. Ce rapport C/L caractérise la sensibilité de l'utilisateur : typiquement un décideur aura un coût d'action C pour éviter une perte L (L > C) face à un événement non prévu.

Le modèle Coût-Pertes de décision peut être couplé à un tableau de contingence (voir Chapitre 3) et, de ce couplage, des critères peuvent être calculés pour déterminer la valeur économique potentielle d'une prévision (Murphy, 1985 ; Wilks et Hamill, 1995 ; Richardson, 2000 ; Buizza *et al.*, 2007b). La Figure 23 illustre un exemple de modèle Coût-Pertes couplé à un tableau de contingence.

	Evénement observé	***Evénement non observé***	
Action prise	*C*	*C*	*Evénement prévu*
	-a-	*-b-*	
Action non prise	*L*	*0*	*Evénement non-prévu*
	-c-	*-d-*	
			Total
			N = a+b+c+d

Figure 23: Modèle Coût-Pertes de décision couplé à un tableau de contingence des événements prévus/observés.

Selon ce tableau, sur *N* prévisions au total, le nombre de succès de la prévision (événement prévu et observé) est égal à *a*, le nombre de fausses alertes est égal à *b*, le nombre d'alertes manquées est égal à *c*, et le nombre d'événements non-prévus et non-observés est égal à *d*. Chaque case est associée à une action, et chaque action suivie par un coût des mesures protectrices *C* ou une perte *L* en cas d'occurrence d'un événement non prévu.

Des scores de mesure de la valeur économique des prévisions ont été proposés pour les systèmes décrits par un modèle du type Coût-Pertes. La valeur économique V_{forc} des prévisions, par exemple, peut être calculée en prenant la somme des produits entre la fréquence de chaque option valable du tableau de contingence (fréquence des succès, fausses alertes et événements manqués) et le coût ou perte respectif qui y est associé dans la matrice Coûts-Pertes :

$$V_{forc} = \frac{a}{N}C + \frac{b}{N}C + \frac{c}{N}L \qquad \text{Eq. 46}$$

Un score de performance, nommé Valeur Economique Relative (*REV)* adapté en météorologie par Richardson (2000), peut être estimé en considérant la valeur économique d'une prévision de référence (la prévision "parfaite", égale aux observations, ou la climatologie, par exemple). Ainsi ce score est exprimé mathématiquement :

$$REV = (V_{forc} - V_{c\lim}) / (V_{parf} - V_{c\lim}) \qquad \text{Eq. 47}$$

Pour ce qui est des prévisions de référence, la valeur économique d'une prévision parfaite, par exemple, est donnée par le produit du coût des mesures protectrices par la fréquence climatologique de l'événement $p_{c\lim}$, étant donné que dans ce cas il n'y aurait pas de fausses alertes ou d'événements manqués.

En ce qui concerne la valeur économique de prévisions climatologiques, Richardson (2000) l'a définie comme le minimum entre : a) prendre toujours une mesure protectrice qui coûte *C* et alors éviter une perte *L*, ou b) ne jamais prendre de mesures protectrices, ce qui résulte à prendre le produit des pertes *L* par la fréquence climatologique de l'événement $p_{c\lim}$. En prenant en compte cette définition, l'Eq. 47 devient :

$$REV = \frac{(\frac{a}{n}C + \frac{b}{n}C + \frac{c}{n}L) - \min(C,\, p_{c\lim}L)}{p_{c\lim}C - \min(C,\, p_{c\lim}L)} \qquad \text{Eq. 48}$$

Le modèle Coût-Pertes peut s'appliquer à des problèmes de complexités diverses. Dans le cas des prévisions probabilistes d'ensemble, la construction du tableau de contingence est faite par seuils de probabilité (voir Chapitre 3, §3.2). Pour chaque seuil de probabilité, et pour un événement particulier avec une certaine fréquence climatologique, la valeur économique des prévisions d'ensemble est estimée (ex., Atger, 2001 ; Buizza, 2001). En outre, pour que les résultats puissent être utiles à des décideurs sensibles à des ratios coûts/pertes différents, les valeurs économiques obtenues sont souvent représentées de façon normalisée, par une courbe donnée en fonction du rapport *C/L* (Figure 24). Dans le cas d'une prévision d'ensemble, la valeur économique de la prévision d'ensemble est celle donnée par la courbe enveloppe des valeurs maximales du score *REV* (Buizza, 2001).

Ainsi, le décideur peut optimiser la valeur de la prévision d'ensemble en choisissant le ratio *C/L* approprié à son problème. Bien évidemment, cette procédure peut se faire à l'inverse : pour un ratio *C/L* donné, le décideur peut maximiser la valeur des prévisions probabilistes en choisissant le seuil de probabilité le plus approprié.

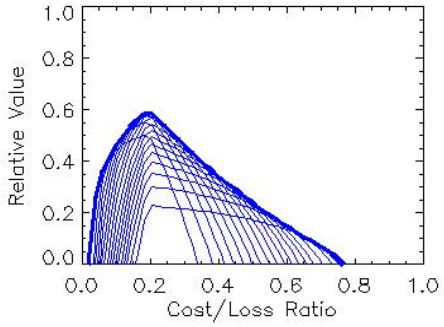


Figure 24: Courbes de valeur économique relative pour une prévision probabiliste (tiré de http://www.cawcr.gov.au/projects/verification/)

Un autre modèle simple de décision est celui qui maximise les taux de bonne détection et minimise les fausses alertes. Dans ce cas, les courbes ROC (voir Chapitre 3, §3.2) sont un exemple de score qui peut être utilisé à cet effet. Rappelons qu'une courbe ROC qui est proche du coin haut gauche du graphique représente un système qui a une bonne capacité à discriminer l'événement ciblé. En termes économiques, le positionnement d'un système sur le graphique équivaut à un coût (ou un gain) plus ou moins important, selon les valeurs de taux de fausses alertes (POFD) et de probabilité de bonnes détections (POD) obtenues (ex., une valeur de POFD élevée est souvent accompagnée d'un coût *C* et/ou des pertes *L* plus élevés également dû au nombre élevé de fausses alertes). Les courbes ROC deviennent alors un outil visuel de mesure de la valeur d'une prévision.

A ces modèles simples de décision utilisés pour mesurer la valeur économique d'une information de prévision, il s'ajoute des modèles basés sur des algorithmes d'optimisation pour aider la prise de décision dans la gestion de l'eau et des risques hydrologiques. Ces modèles peuvent servir à des objectifs divers et cibler des horizons de prévision limités (ex., cas des prévisions à quelques jours ou semaines) ou illimités (ex., cas des applications visant la planification des ouvrages à long terme ou leur adaptation aux changements climatiques).

Quel que soit le choix de modélisation économique, on notera que dans le domaine de la prévision hydrologique d'ensemble, les études portant sur la valorisation économique des prévisions sont récentes et révèlent un large spectre d'applications, comme on le verra par la suite.

4.5 Etudes d'évaluation économique en prévision hydrologique d'ensemble

Selon Atger (2001), en météorologie, le concept de valeur économique (utilité) des prévisions a été très étudié dans les années 1970-80, suivi par une réapparition de l'intérêt porté au sujet à la fin des années 1997/1998. Les modèles du type Coût-Pertes ont été ainsi utilisés pour montrer la valeur économique potentielle des systèmes de prévision d'ensemble de précipitations par rapport aux prévisions déterministes (Atger, 2001 ; Buizza, 2001, 2008). Tsikalakis *et al.*, (2009) ont étudié l'impact économique des prévisions de vent sur la production éolienne à l'aide de la programmation séquentielle et ont montré l'importance d'avoir des prévisions fiables pour un meilleur gain économique.

En hydrologie, des études sur l'évaluation économique des prévisions de débits sont plus rares, même si l'utilisation des prévisions d'ensemble pour la prévision des crues et la gestion des ressources en eau est déjà présente depuis une dizaine d'années (Cloke et Pappenberger, 2009a). En effet, les études portant sur la qualité des prévisions hydrologiques (voir Chapitre 3) sont beaucoup plus courantes que celles dédiées à l'utilité (valeur) des prévisions.

Le modèle de décision statique du type Coût-Pertes est également celui qui reçoit le plus d'attention dans la littérature hydrologique. Roulin (2007) l'a étudié sur deux bassins versants en Belgique pour évaluer la valeur des prévisions de crues pour la prise de décision. Les prévisions hydrologiques à moyen-terme (1 à 10 jours), obtenues à partir des prévisions météorologiques d'ensemble du CEPMMT et d'un modèle hydrologique conceptuel à 10 paramètres, ont été évaluées sur une période de six ans. L'auteur propose également l'utilisation d'un modèle Coût-Pertes dynamique, à horizon limité, qui permet de prendre en compte différentes décisions successives en fonction de l'échéance de prévision. Dans les deux cas, la valeur économique relative *REV* des prévisions d'ensemble a été supérieure à celle des prévisions déterministes pour une large gamme de ratio *C*/*L*, le choix d'un seuil de probabilité pour la détection des événements jouant un rôle important sur la valeur économique estimée.

Des conclusions semblables quant aux performances comparées des prévisions déterministes et probabilistes, basées sur ce type de modèle, ont été obtenues par McCollor et Stull, (2008), Van de Bergh et Roulin (2010), Muluye (2011) et Verkade et Werner (2011) :

- Verkade et Werner (2011) ont étudié la valeur économique relative *REV* pour évaluer la valeur d'une alerte anticipée dans une chaîne de prévision du risque d'inondation. Ils ont notamment montré le lien entre l'échéance optimale pour l'alerte et le ratio C/L accepté par le décideur.

- McCollor et Stull (2008) ont évalué les prévisions de précipitation sur un bassin versant au Canada et ont ciblé, à leur tour, un utilisateur ayant pour objectif la production hydroélectrique d'une réserve en eau. Les auteurs ont utilisé deux types de modèle de décision : un modèle Coût-Pertes statique appliqué sur les précipitations et un modèle de maximisation d'une fonction d'utilité liée au gain en production d'énergie d'une usine-réservoir simplifiée. Ils ont montré que les prévisions probabilistes d'ensemble sont plus performantes pour une gamme plus large d'utilisateurs par rapport aux prévisions déterministes, et ce pour les deux types de modèle de décision utilisés. Ils ont également montré que la prévision donnée par la moyenne d'ensemble est légèrement plus performante que les prévisions de chaque membre pris indépendamment.
- L'application sur un bassin versant au Québec (Canada) a été aussi au cœur de l'étude de Muluye (2011). Les prévisions météorologiques du centre américain NCEP (15 membres, échéance maximale de 15 jours et 23 ans de prévisions journalières) ont été utilisées avec le modèle hydrologique semi-distribué *HBV* dans deux cas de figure : sans et avec un prétraitement de désagrégation des champs des précipitations prévues. Les auteurs ont montré que la technique de désagrégation employée permettait d'obtenir des prévisions à plus forte valeur économique sur une large gamme de ratios C/L. En effet, seulement sur une petite gamme de ratios C/L les prévisions climatologiques ont montré une valeur de *REV* supérieure aux prévisions probabilistes.
- A partir d'un modèle statique Coût-Pertes appliqué sur 14 bassins versants de la Meuse et de la rivière Scheldt en Belgique, Van de Bergh et Roulin (2010) ont montré que la valeur économique des prévisions d'ensemble se réduit en fonction de l'horizon de prévision (comme c'est également le cas pour la qualité des prévisions) : plus l'horizon est lointain de l'instant de prévision, moins les prévisions possèdent de valeur.

D'autres travaux comme celui de McCollor et Stull (2008) cités ci-dessus ont également étudié la valeur économique des prévisions hydrologiques à l'aide de méthodes d'optimisation numérique :

- Kim *et al.* (2007), par exemple, ont utilisé un outil d'aide à la décision basé sur la programmation dynamique stochastique pour évaluer les prévisions d'ensemble d'un bassin versant en Corée du Sud. Ils ont montré que les prévisions d'ensemble sont plus performantes que les prévisions déterministes et que la justesse des prévisions est un facteur important vers une gestion optimisée. De plus ils ont trouvé que la combinaison de la programmation dynamique avec la programmation stochastique aide à surpasser la "malédiction de la dimensionnalité".

- Weijs (2011), en utilisant un outil de programmation dynamique stochastique sur un bassin versant et en évaluant 55 ans de débits modélisés, a montré que l'information supplémentaire provenant, par exemple, des prévisions probabilistes augmente le gain de gestion. De plus, il a été montré que cette information supplémentaire a une valeur significative en temps réel pour des opérations à courts horizons de prévision et à des niveaux de réservoirs élevés.
- Boucher *et al.* (2011) ont pris en main un outil d'aide à la décision (*SOHO*) qui constitue une application de la programmation stochastique. L'application porte sur un bassin versant au Québec (Canada) avec 1,5 an de calage et 4 mois de validation. Les auteurs ont conclu que les prévisions d'ensemble brutes conduisent à une production d'énergie moins importante (à cause de déversements plus élevés) par rapport aux prévisions déterministes. Cependant ils ont montré que les prévisions post-traitées sont plus performantes par rapport aux prévisions brutes et aux prévisions déterministes.

Enfin, dans une perspective à plus long-terme et dans le but d'étudier les impacts du changement climatique sur la gestion de la ressource en eau et de la ressource énergétique associée, François *et al.* (2012) ont utilisé la programmation dynamique multi-objectif pour développer un modèle de gestion optimisée pour les rivières françaises de la Durance et de la Loire. Les premiers résultats ont montré une variation non-linéaire de la gestion optimale en fonction des changements de précipitation et température, affectant la performance du système en termes de disponibilité d'eau dans les réservoirs.

En résumé, on soulignera que la valorisation des prévisions hydrologiques a été étudiée à l'aide :

- d'un modèle Coût-Pertes statique ou dynamique destiné à évaluer la valeur des prévisions pour des objectifs de protection vis-à-vis des crues et inondations (Verkade et Werner, 2011 ; Muluye, 2011 ; Roulin, 2006 ; Van de Bergh et Roulin, 2010) ;

- d'un outil d'aide à la décision déjà mis en place par un service opérationnel pour la gestion de la ressource en eau (Boucher *et al.*, 2011 ; Kim *et al.*, 2007 ; Weijs, 2011a,b) ;

- d'un modèle Coût-Pertes statique en parallèle avec un modèle maximisant l'espérance de l'utilité d'une prévision (McCollor et Stull, 2008) ;

- d'un outil d'optimisation développé spécialement pour des besoins spécifiques de planification à long-terme (François *et al.*, 2012).

Le point commun de toutes ces approches se trouve dans le fait qu'elles ont été développées et appliquées pour étudier la valeur des prévisions d'ensemble ou des multi-scénarios climatiques pour des applications hydrologiques. Les objectifs des utilisateurs cependant sont variés, illustrant la diversité des intérêts opérationnels envers les prévisions hydrologiques : par exemple, Boucher *et al.* (2011) et Weijs (2011) ciblent l'optimisation de la production d'énergie, tandis que Kim *et al.* (2007) s'appuient sur la satisfaction de la demande dans la distribution de l'eau. Weijs (2011) étudie également la gestion optimisée d'un système de drainage qui, d'un côté, sert à sécuriser l'évacuation des eaux et, de l'autre côté, contribue à minimiser les dommages associés aux hautes et basses eaux. François *et al.* (2012) sont plus orientés vers une approche multi-objectifs (production d'énergie, protection de crues, tourisme, etc.), en prenant en compte les pertes économiques dues au non-respect des objectifs préfixés.

Les conclusions principales que nous pouvons tirer des études existantes peuvent être résumées aux points suivants :

- souvent l'existence d'outils déjà mis en place est limitée et, dans ce cas, l'application d'un modèle Coût-Pertes statique ou dynamique, par sa simplicité de mise en œuvre et son domaine large d'applicabilité, est privilégiée ;
- la valeur économique des prévisions probabilistes d'ensemble est souvent plus importante que celle des prévisions déterministes ;
- la valeur économique des prévisions d'ensemble décroît avec l'horizon de prévision et peut même, au delà d'un certain horizon, n'être plus significativement meilleure que la valeur apportée par une prévision climatologique ;
- la moyenne d'ensemble peut parfois être légèrement plus performante que les prévisions de chaque membre de l'ensemble prises individuellement ;
- les prévisions probabilistes d'ensemble peuvent être plus utiles que les prévisions déterministes pour une gamme plus large d'utilisateurs (représentés par des ratios C/L différents) ;
- la qualité des prévisions (justesse et fiabilité) peut affecter la valeur économique des prévisions probabilistes ;
- la valeur potentielle des prévisions d'ensemble peut être affectée par les erreurs de modélisation, ainsi que par la taille de l'ensemble ;
- la valeur économique des prévisions hydrologiques corrigées (soumises à un prétraitement des entrées ou à un post-traitement des sorties) est souvent supérieure à celle des prévisions brutes et des prévisions déterministes.

Le fait que les conclusions des différentes études, menées avec des données et modèles hydrométéorologiques variés, convergent est encourageant. Les forts intérêts économiques liés à la gestion des ressources en eau justifient ces études et de nouveaux développements pour mieux comprendre comment l'utilisateur des prévisions peut en faire le meilleur usage dans la prise de décision.

4.6 Synthèse

Ce chapitre a présenté une autre caractéristique des prévisions d'ensemble à savoir : leur utilité. Afin de valoriser et optimiser cette caractéristique, il est primordial de recourir aux techniques d'optimisation. Un résumé des techniques les plus courantes a été présenté, accompagné d'une synthèse des études menées en hydrologie pour évaluer la valeur économique des prévisions.

Cette introduction s'insère dans le cadre de la construction de notre propre approche d'optimisation de la gestion d'un réservoir. Plus particulièrement, notre objectif est d'examiner l'apport des prévisions, et notamment des prévisions probabilistes d'ensemble, dans la gestion optimale d'une réserve en eau superficielle destinée à la production hydroélectrique. L'outil développé utilise une approche heuristique basée sur l'expérience opérationnelle et est présenté en détail dans le Chapitre 7 de ce manuscrit.

"There is no such thing as bad weather. Only different kinds of good weather." *John Ruskin 1819-1900*

5. Prévisions d'ensemble multi-modèles météorologiques pour la prévision hydrologique

Dans ce chapitre, nous nous intéressons à évaluer l'apport d'une approche multi-modèles de prévision d'ensemble météorologique à la prévision de débits. La base TIGGE de prévisions d'ensemble de 8 centres météorologiques et le modèle hydrologique GRP (sans module neige) sont ici utilisés. Nous nous focalisons sur la qualité globale des prévisions, notamment sur les attributs de fiabilité et discrimination. Les performances des prévisions obtenues sur le jeu de 74 grands bassins versants sont évaluées en fonction de l'horizon de prévision et de la surface et de la localisation des bassins versants. Enfin, nous examinons comment combiner les informations extraites de chaque centre météorologique pour former un grand ensemble composé de tous les membres des différents centres qui soit plus performant que chaque centre pris individuellement. Différentes approches de combinaison sont ici proposées.

5.1 Etudes hydrologiques basées sur les prévisions TIGGE

L'utilisation de la base TIGGE en météorologie est assez répandue. Selon Roulin (2007) l'utilisation des prévisions d'un seul centre contribue à la prise en compte d'une partie limitée de l'incertitude liée aux conditions initiales. Par contre, l'utilisation des prévisions provenant d'un multi-modèle (météorologique) permet de mieux prendre en compte pas non seulement les incertitudes liées aux conditions initiales mais aussi – partiellement au moins – celles liées à la modélisation numérique et l'assimilation de données (Goswami *et al.*, 2007 ; Park *et al.*, 2008 ; Bougeault *et al.*, 2010 ; Hagedorn *et al.*, 2012 ; Hamill *et al.*, 2012). Par conséquent, les prévisions tirées d'un multi-modèle contribuent à une meilleure représentation de la loi de probabilité prédictive (Hagedorn *et al.*, 2005 ; Zhi *et al.*, 2012).

Dans la littérature, l'utilisation de la base de prévisions météorologiques d'ensemble TIGGE en hydrologie est récente et a concerné notamment quelques bassins versants et/ou périodes courtes d'évaluation.

- L'étude de Pappenberger *et al.* (2008) est l'une des premières à examiner l'utilisation de la base TIGGE pour la prévision de crues. Les auteurs présentent un retour d'expérience concernant la crue d'octobre 2007 qui a touché les bassins versants de divers affluents du Danube en Roumanie. Les prévisions météorologiques de 7 centres météorologiques sont utilisées en entrée du modèle hydrologique LISFLOOD du système EFAS (Thielen *et al.,* 2009). Les hypothèses principales de cette étude étaient l'équiprobabilité des membres et l'indépendance de la prévision à la taille de chaque ensemble. Par conséquent, le grand ensemble, avec au total 216 membres, est construit à partir de la combinaison directe des prévisions d'ensemble des centres individuels sous le principe de l'équiprobabilité de membres. De plus, les prévisions numériques du temps sont considérées comme la source principale d'incertitude des prévisions hydrologiques. Les prévisions de précipitation et de débit sont évaluées à partir de l'analyse des scores de Brier *BS* et de probabilité ordonnée *RPS* et à l'aide de scores graphiques. Cette étude a démontré que l'utilisation d'un grand ensemble améliore la fiabilité des prévisions hydrologiques par rapport à celle des prévisions tirées des centres individuels et apporte une valeur ajoutée aux systèmes opérationnels d'alerte aux crues.

- L'étude de He *et al.* (2009), quant à elle, est centrée sur le bassin versant de la Severn (4062 km²) en Angleterre et pour une période de prévision allant du 09/2007 au 04/2008, avec un seul événement de crue. La sélection du bassin versant est faite sur l'hypothèse que, pour que les prévisions soient fiables dans les cas des petits bassins, un grand nombre de membres d'ensemble est nécessaire. La combinaison des

prévisions des différents centres météorologiques est également basée sur le principe de l'équiprobabilité. L'évaluation de prévisions d'ensemble de pluie s'effectue en utilisant le score *CRPS* et la dispersion de membres. Les prévisions de débits pour l'événement de crue examiné sont évaluées à l'aide d'un tableau de contingence, des courbes ROC et des hydrogrammes. Les auteurs ont conclu que les incertitudes des prévisions de pluie dominent toute la chaîne de prévision et ainsi les prévisions hydrologiques tirées du grand ensemble se caractérisent par une meilleure chance de détecter la crue examinée.

- Une autre étude faite par He *et al.* (2010) se concentre sur trois événements de crue (de juillet à octobre 2008) qui ont eu lieu dans un bassin versant plus grand, celui de Huai en Chine (30.672 km²). Cette étude est également basée sur l'hypothèse d'équiprobabilité des membres des prévisions et les résultats ont conduit à des conclusions similaires.

- Enfin, deux études plus récentes ont été conduites par Bao *et al.* (2011 ; 2012) sur les bassins versants de Xixian et de Huaihe en Chine (8.826 et 920 km² respectivement). Les prévisions d'ensemble de 5 centres météorologiques de la base TIGGE ont été utilisées (ECMF, EGRR, KWBC, BABJ et CWAO). Cette étude se focalise sur l'événement de crue de juillet 2007 et vise à évaluer la performance du grand ensemble, construit également avec le regroupement des membres de tous les centres étudiés. Les prévisions d'ensemble de pluie et de débits sont évaluées sur la base de seuils de dépassement à l'aide de l'indice critique de succès *CSI* (en anglais critical succes index) et des hydrogrammes respectivement. Les auteurs ont conclu que les résultats sont encourageants et le grand ensemble permet de produire une alerte aux crues avec 10 jours d'anticipation.

Toutes ces études nous ont clairement démontré la valeur ajoutée des prévisions hydrologiques tirées d'un grand ensemble construit à partir de prévisions de différents centres météorologiques. Néanmoins, les résultats manquent de généralité, car ces études sont basées pour la plupart sur un seul bassin versant et sur un nombre limité d'événements. Les perspectives énoncées par les différents auteurs (Pappenberger *et al.*, 2008 ; He *et al.*, 2009 ; Bao *et al.*, 2011) ont par ailleurs souligné l'importance d'appliquer la base de prévisions TIGGE sur un échantillon plus large de bassins versants avec des conditions hydro-climatiques variées. Ils ont également suggéré d'examiner des combinaisons pondérées des prévisions tirées des différents centres météorologiques pour la construction du grand ensemble. He *et al.* (2009) proposent en particulier de pondérer les différents centres selon leur performance en tant que centres individuels.

Nous exploitons la base de prévisions TIGGE sur l'échantillon de 74 bassins versants français constitué dans ce but (voir Chapitre 2, §2.3) et examinons différentes approches de construction du grand ensemble. La transformation de prévisions météorologiques en prévisions hydrologiques est réalisée à l'aide du modèle conceptuel pluie-débit *GRP* (voir Chapitre 2, §2.4) dans sa version journalière, adaptée à la prévision d'ensemble et sans module de prise en compte de la neige.

Dans l'analyse présentée ici, les prévisions de précipitation émises par 8 centres TIGGE (ECMF, EGRR, JRTD, KWBC, BABJ, AMMC, CWAO, SBSJ) ont été utilisées, avec un maximum de 232 membres d'ensemble par bassin versant et par jour, pendant 2 ans (01/10/2006 à 30/09/2008). Nous présentons ci-après, l'évaluation des prévisions d'ensemble basées sur les prévisions météorologiques de la base de données TIGGE sur les 74 bassins étudiés.

5.2 Evaluation des prévisions d'ensemble

5.2.1 Évaluation des prévisions d'ensemble de pluies

5.2.1.1 Illustration d'un événement particulier

Avant de procéder à l'évaluation de la qualité des prévisions de précipitations sur l'ensemble de la période disponible (01/10/2006 à 30/09/2008), nous nous concentrerons d'abord sur un exemple particulier d'événement de pluies fortes, afin d'illustrer les prévisions tirées des différents centres météorologiques. Pour cela, nous avons choisi l'événement du 25 mai 2008. Cet événement a été caractérisé par de fortes précipitations dans le Sud-Est de la France, qui sont arrivées en même temps que le pic de la fonte, avec 38,4 et 44,4 [mm] de pluies enregistrées respectivement sur les bassins versants de la Durance à Serre Ponçon (3580 km²) et de la Drôme à Saillant (1123 km²). La Figure 25 et la Figure 26 illustrent les prévisions de 4 membres de l'ensemble choisis aléatoirement dans chaque centre météorologique de la base TIGGE pour les horizons de prévision de 3 et 1 jours respectivement, et sur l'ensemble de la France. Sur les Figure 25 et Figure 26 nous remarquons que les 8 centres météorologiques ont eu au moins un membre qui a détecté l'arrivée d'un événement important sur la France et a bien localisé l'événement sur la région où l'événement s'est produit (Sud-Est). De plus, nous remarquons que le plus lointain l'horizon de prévision, le moins bien capté est l'événement.

La Figure 27 présente la distribution (sous forme de boîtes à moustaches) des valeurs prévues par les membres de l'ensemble à 5 jours d'échéance sur le bassin versant de la Drôme à Saillant (1123 km²) et pour la période du 23 à 30 mai 2008. Chaque figure (1 à 8) correspond à chacun des centres météorologiques de la base TIGGE.

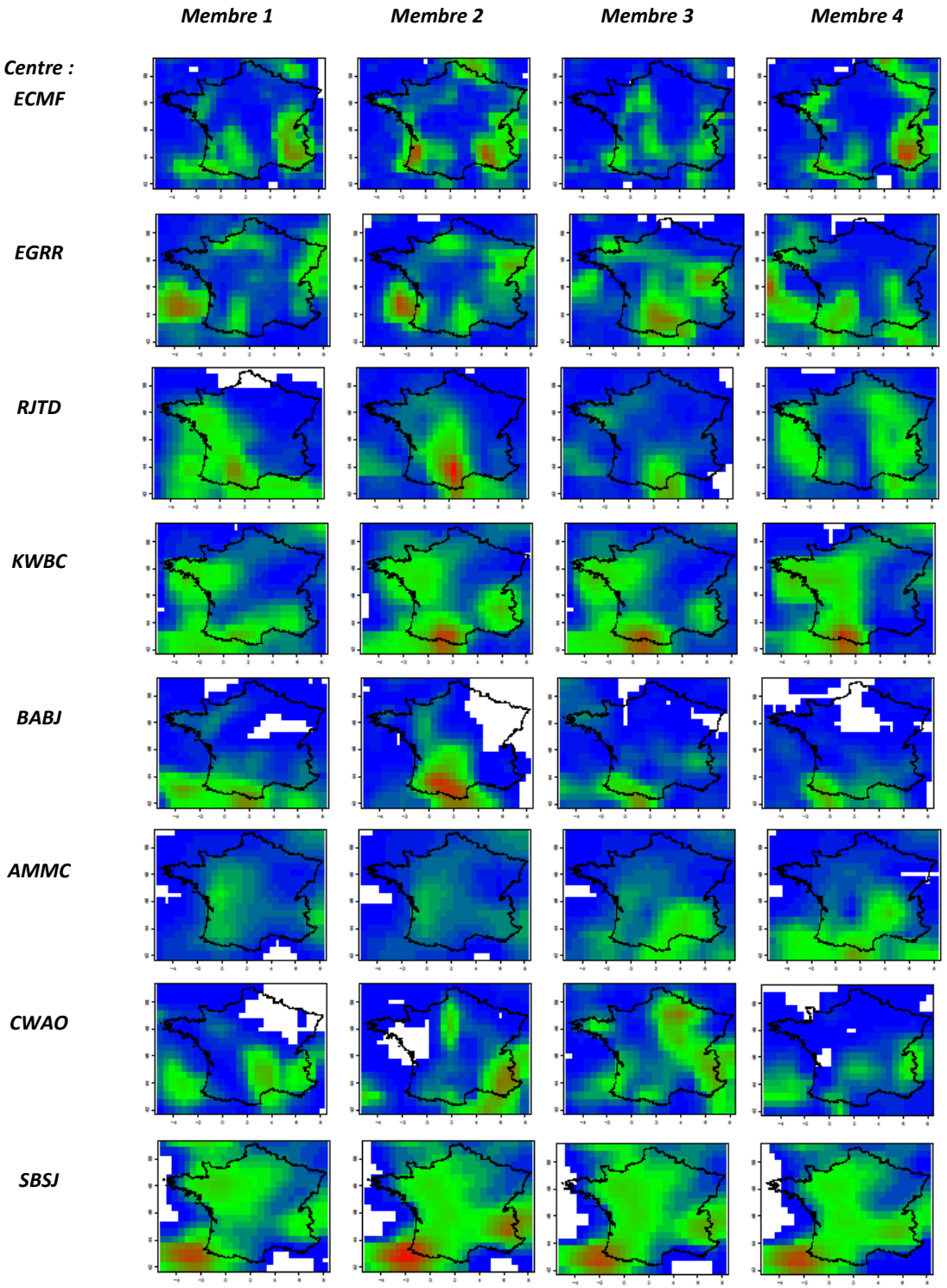


Figure 25 : Prévisions de précipitation valables pour le 25 mai 2008 sur la France des 8 centres météorologiques de la base TIGGE de cette étude et à 3 jours d'échéance : répresentation de 4 membres de chaque prévision d'ensemble.

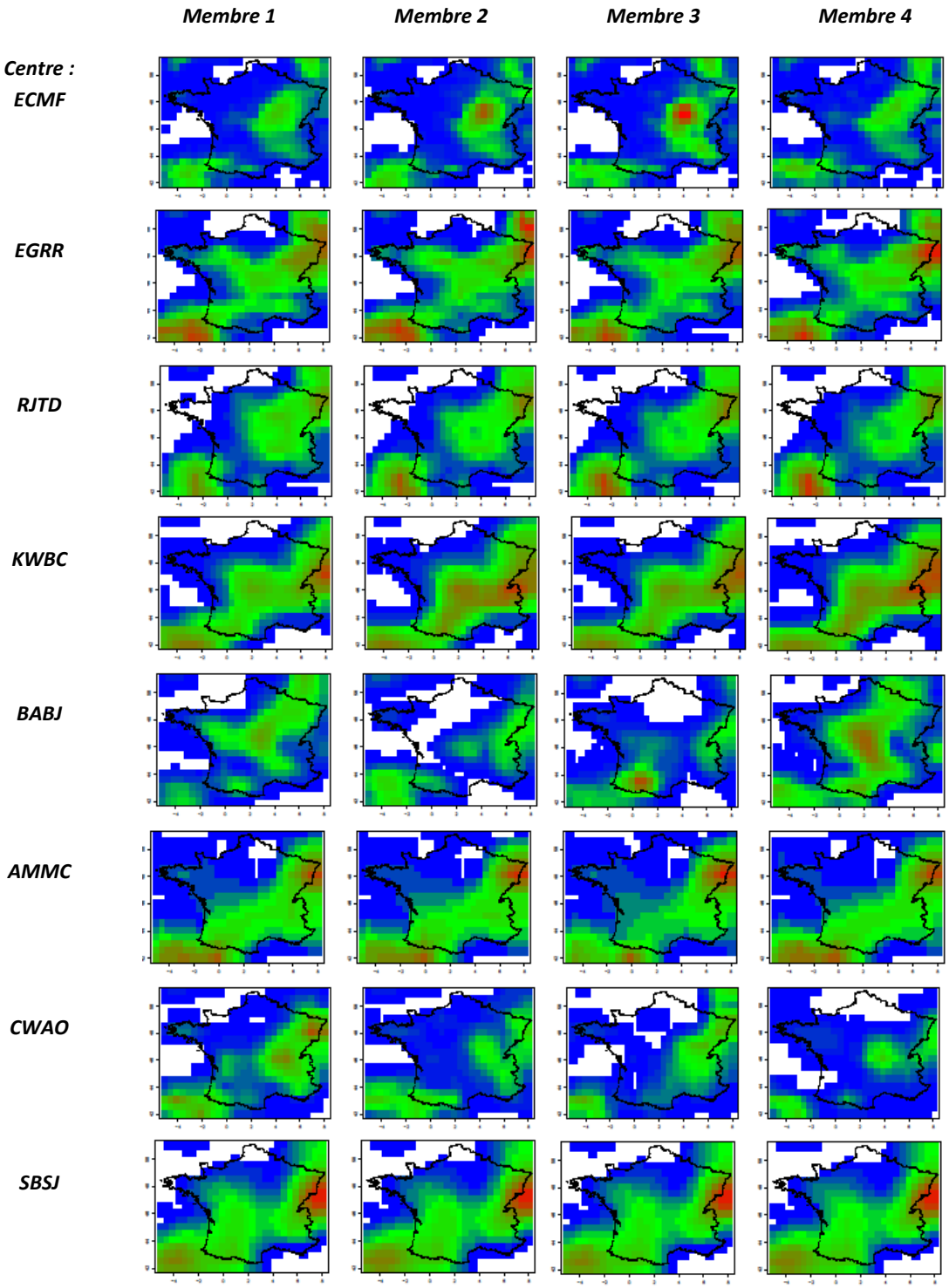


Figure 26 : Prévisions de précipitation valables pour le 25 mai 2008 sur la France des 8 centres météorologiques de la base TIGGE de cette étude et à 1 jour d'échéance : répresentation de 4 membres de chaque prévision d'ensemble.

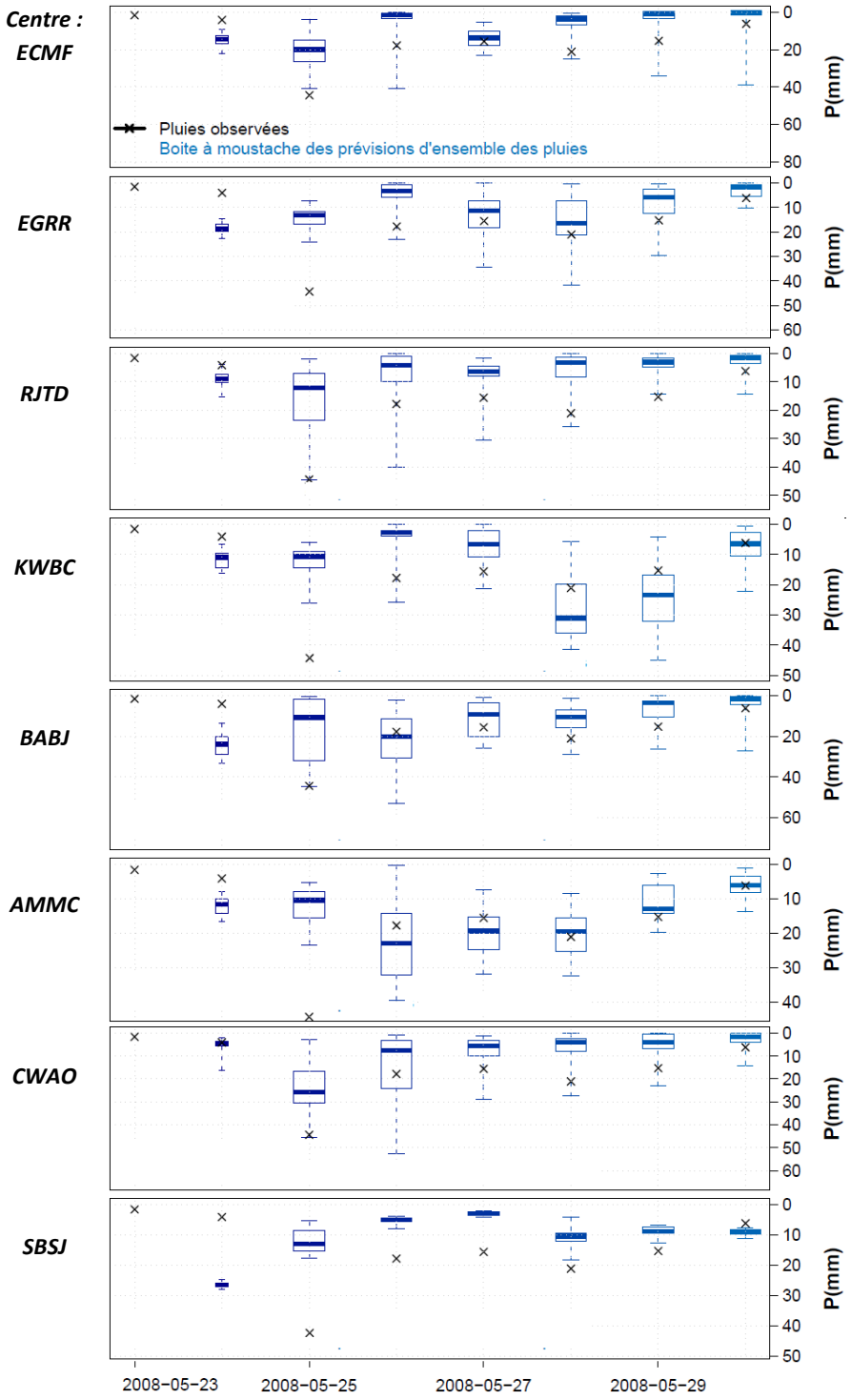


Figure 27 : Boîtes à moustaches pour les prévisions d'ensemble de précipitation émises le 23 mai 2008 sur le bassin versant de la Drôme à Saillant pour les 5 jours à venir et selon les 8 centres météorologiques de la base TIGGE de cette étude. Les précipitations observées sont indiquées à l'aide du symbole "x".

Sur la Figure 27, la ligne bleue épaisse montre la médiane, les bords bas et hauts de la boîte correspondent aux quantiles 25% et 75% et les extrémités basses et hautes de l'intervalle aux valeurs minimales et maximales. Sur cette figure les points noirs représentés par un "x" indiquent la précipitation observée. Notons que la direction de l'axe des précipitations va du haut vers le bas (les valeurs proches de zéro sont localisées vers le haut du graphique).

En termes des intensités de la précipitation nous remarquons dans la Figure 27 que, pour 6 sur ces 8 centres météorologiques, l'observation se retrouve en général dans l'intervalle donné par les quantiles maximaux et minimaux des ensembles. Néanmoins, nous remarquons que la plupart des centres ont pourtant tendance à sous-estimer les précipitations observées.

5.2.1.2 Critères graphiques pour deux bassins versants représentatifs

Nous évaluons ici la qualité des prévisions d'ensemble de pluie en examinant la période entière, ainsi que différents attributs de la qualité d'une prévision qui sont mesurés par des critères graphiques. Parmi ces différents attributs nous examinons par la suite : la fiabilité, la discrimination et la dispersion des membres (voir Chapitre 3, §3.2). Nous considerons que avec ces trois attributs nous arrivons à bien capturer les composants principaux de la qualité d'une prévision. Il faut noter que l'évaluation des prévisions extraites de chaque centre météorologique est faite sur le nombre de jours des prévisions disponibles pour chaque centre séparément. Par conséquent, la période d'évaluation est de longueur variable selon le centre considéré.

Fiabilité : diagramme de PIT

Les Figure 28 et Figure 29 présentent les histogrammes de PIT pour les prévisions d'ensemble de pluie de bassin tirées des 8 centres météorologiques constituant la base TIGGE, pour l'échéance de 7 jours et pour les bassins versants de la Drôme à Saillant (1123 km²) et de l'Oise à Sempigny (4320 km²) respectivement. Ces deux bassins versants ont été choisis pour les contrastes qu'ils présentent : régimes hydro-climatiques, surfaces et localisations. Le bassin versant de la Drôme s'agit d'un bassin méditerranéen : dominé par des étés chauds et des pluies intenses (surtout en automne) et légèrement influencé par la neige surtout aux mois d'hiver et de printemps. Par contre le bassin versant de l'Oise à Sempigny est plutôt caractérisé par des écoulements périodiques concentrés pendant les mois d'hiver - automne et avec des écoulements souterrains assez limités. Leurs comportements en ce qui concerne la performance de la base TIGGE en termes de fiabilité sont qualitativement représentatifs de l'échantillon des 74 bassins versants examinés.

Rappelons que la fiabilité représente la correspondance entre la probabilité prédictive et la fréquence observée. Un des critères les plus utilisés pour évaluer la fiabilité de prévision est l'histogramme de PIT : un histogramme uniforme est une indication de prévisions fiables et, au contraire, un histogramme sous forme d'arche correspond à des prévisions plutôt sur-dispersées, tandis qu'un diagramme sous forme de "U" est indicatif de prévisions sous-dispersées (Hamill, 2001).

De l'ensemble de notre analyse, nous remarquons que :

- la fiabilité de prévisions météorologiques est très dépendante de l'horizon de prévision. En examinant tous les horizons de prévisions, nous constatons que plus l'horizon de prévision est lointain, plus la fiabilité de la prévision s'améliore : les prévisions à un horizon de 3 jours sont beaucoup moins fiables que celles aux horizons 7 ou 9 jours. La raison principale est liée à la dispersion des prévisions d'ensemble, car le faisceau de membres s'élargit en fonction de l'échéance. Cette tendance est valable pour tous les centres météorologiques de la base TIGGE, cependant pour certains centres - australien (6) et brésilien (8) - l'amélioration de la fiabilité en fonction de l'horizon de prévision est beaucoup moins forte par rapport à celle du centre européen (1) et du canadien (7).

- certains centres présentent des prévisions fiables au-délà des échéances de 5 à 7 jours. C'est le cas du centre européen (1), du centre japonais (3), du centre chinois (5) et du centre canadien (7). D'autres centres montrent une forte tendance à produire de prévisions biaisées. Les sources de ces biais varient selon le centre météorologique : pour le centre australien (6) et brésilien (8) elles sont liées à la sous-dispersion de membres. Pour le centre américain (4) ce biais est plutôt lié à la sous-estimation de l'observation.

- sur les 74 bassins versants et jusqu'à un horizon de prévision de 7 jours, le centre européen (1) se montre le plus performant avec une fiabilité qui s'améliore considérablement en fonction des échéances. Les centres japonais (3), chinois (5) et canadien (7) présentent également de bonnes performances. Nous constatons que la performance des centres est indépendante du nombre de jours de prévisions disponibles, ainsi que des caractéristiques du bassin versant.

Discrimination : courbes ROC

Les Figure 30 et Figure 31 présentent les courbes ROC obtenues pour la prévision à l'échéance de 7 jours d'un événement défini par le dépassement d'un seuil égal au quantile 70% des précipitations observées sur la période d'évaluation. Ces figures illustrent les résultats pour les 8 centres de la base TIGGE et concernent les prévisions de pluie sur le bassin versant de la Drôme à Saillant et l'Oise à Sempigny respectivement.

Rappelons que le principe des courbes ROC probabiliste est que chaque événement considéré (dépassement d'un seuil égal au quantile 40%, 50%, 70%, etc) est prévu avec différentes probabilités de bonne détection : la probabilité de dépasser le seuil critique par au moins i = 1, 2, 3… ou M membres de la prévision d'ensemble. Pour chaque probabilité de dépassement du seuil i, nous calculons les valeurs des scores $POFD_i$ et POD_i. Idéalement, tous les points de la courbe doivent être regroupés dans le coin en haut à gauche du diagramme. De plus, l'aire sous la courbe ROC est aussi utilisée pour comparer différents systèmes de prévision. Une courbe ROC confondue avec la 1ère bissectrice indique que la prévision n'apporte rien. De l'analyse des Figure 30 et Figure 31, nous observons que :

- le dépassement du quantile 70% avec une probabilité de détection supérieure à 60% est prévu par la majorité de membres par le centre européen (1), japonais (3), australien (6) et canadien (7). Cette conclusion est confirmée également par l'aire sous la courbe ROC qui est plus elevée pour ces 4 centres. Le comportement est influencé par l'horizon de prévision : par exemple dans le cas du bassin versant de la Drôme à Saillant et pour les prévisions tirées du centre européen *ECMF* le nombre de membres, qui prévoient le dépassement du quantile 70% avec une probabilité de détection 60%, passe de 35 membres sur 50 pour l'horizon de 3 jours à 10 membres pour celui de 15 jours. En ce qui concerne les autres centres météorologiques, les sytèmes se montrent peu performants au-delà d'un horizon de prévision de 9 jours (centres britannique *EGRR* et japonais *RJTD*) et de 7 jours pour les autres centres. Au-delà de cet horizon les prévisions présentent des probabilités de détection et de fausses alertes égales. Par conséquent le pourcentage d'événements prévus quand ils sont observés est le même que celui des événements prévus quand ils ne sont pas observés. En examinant les autres seuils de non-dépassement, nous constatons que le nombre de membres qui prévoient le dépassement avec une probabilité de détection de 60% diminuent en augmentant le seuil de 10% à 90%. Plus particulièrement dans le cas du centre européen *ECMF*, quand le seuil de non-dépassement passe de 50% à 90% le nombre de membres, qui prévoient l'événement avec une probabilité de détection de 60%, varient de 32 à 9 (sur un total de 50 membres). Cette remarque reflète la tendance reproduite par tous les centres.

- sur l'ensemble des 74 bassins versants, nous remarquons que les prévisions tirées du centre européen *ECMF* se montrent les plus performantes à un horizon de prévision jusqu'à 7 à 9 jours. Au-delà de cet horizon presqu'aucun centre n'arrive à discriminer entre les événement s observés et non observés. De plus, selon le bassin versant, les centres japonais *RJTD*, chinois *BABJ* et canadien *CWAO* se montrent également performants. Nous constatons à nouveau que la performance de centres est indépendante du nombre de jours de prévisions disponibles, ainsi que des caractéristiques du bassin versant.

Dispersion des membres

Les Figure 32 et Figure 33 illustrent cette relation entre la dispersion des membres et la RMSE pour les bassins versants de la Drôme à Saillant et de l'Oise à Sempigny à un horizon de prévision de 7 jours. Dans ce graphique nous présentons les résultats pour les 8 centres météorologiques de la base TIGGE.

Rappelons qu'une façon d'examiner l'effet de la dispersion des membres sur la qualité des prévisions est en effet de tracer la dispersion de l'ensemble en fonction de l'erreur quadratique moyenne (*RMSE*). Une forte corrélation entre la dispersion des membres et la *RMSE* indiquera qu'une partie de l'erreur de prévision est attribuée à la dispersion non-appropriée des membres de l'ensemble –sous ou sur-dispersifs.

Nous constatons que :

- la corrélation entre la dispersion et la racine carrée de l'erreur quadratique moyenne *RMSE* est plus forte jusqu'à un horizon de 3-4 jours. Aux horizons au-delà de 3 jours, cette corrélation devient plus faible pour la plupart de centres météorologiques. Plus particulièrement, nous constatons que la dispersion des prévisions tirées du centre européen (1) et du centre japonais (3) est peu corrélée avec la racine carrée de l'erreur quadratique moyenne *RMSE*. Par contre dans le cas des centres britannique (2), américain (4) et chinois (5) quand la dispersion de membres est faible (<5 mm), sa corrélation avec l'erreur quadratique moyenne est plus forte. Ceci signifie que l'erreur de prévision peut être expliquée par une faible dispersion de l'ensemble. Par contre quand la dispersion de membres est plus importante (>5 mm), les erreurs de prévisions (exprimées par le *RMSE*) sont en général moins importantes en termes de valeurs et peu corrélées avec la dispersion des membres. Pour le reste des centres météorologiques de la base TIGGE, le centre australien (6) et le centre brésilien (8), la corrélation des erreurs de prévision reste forte jusqu'à un horizon de prévision de 7 à 9 jours selon le bassin versant.

Afin d'expliquer ces conclusions nous présentons l'évolution de la dispersion de membres (Figure 34 et Figure 35) - sous forme de boîtes à moustaches calculées sur les 242 jours disponbles - pour deux centres démontrant des performances en termes de corrélation de l'erreur quadratique moyenne avec la dispersion totalement différentes : le centre européen (1) et le centre brésilien (8). Ces figures représentent les résultats pour les bassins versants de la Drôme à Saillant et l'Oise à Sempigny respectivement.

Nous remarquons que :

- dans le cas du centre européen *ECMF* la dispersion des membres évolue progressivement jusqu'à 9 jours de prévisions et elle se stabilise au-delà de 9 jours. De plus, la dispersion des membres devient plus élevée à partir d'une échéance de 3 jours. A un horizon de 7 jours de prévision la dispersion médiane des membres sur les 242 jours de prévision est égale à 4 mm et 3,5 mm pour la Drôme à Saillant et l'Oise à Sempigny respectivement. Les quantiles à 25% et 75% varient également de 3 à 5 mm pour la Drôme à Saillant et de 2,5 à 4,5 mm pour l'Oise à Sempigny.

- en ce qui concerne le centre brésilien *SBSJ*, nous constatons que la dispersion des membres reste très faible jusqu'à un horizon de 7 jours de prévision. La dispersion medianne des membres sur les 242 jours de prévision est égale à 1 mm et 0.7 mm ; en même temps que les quantiles à 25% et 75% varient entre 0,5 à 2,5 mm et 0,2 mm à 2,0 mm pour la Drôme à Saillant et l'Oise à Sempigny respectivement. Ces observations renforcent le lien potentiel entre une dispersion non-appropriée, provoquée ici par des ensembles sous-dispersifs, et les erreurs de prévision, représentées ici par l'erreur quadratique moyenne *RMSE.*

Les conclusions de cette analyse sur ces deux bassins versants sont ensuite confirmées par le reste de l'échantillon de 74 bassins versants examinés.

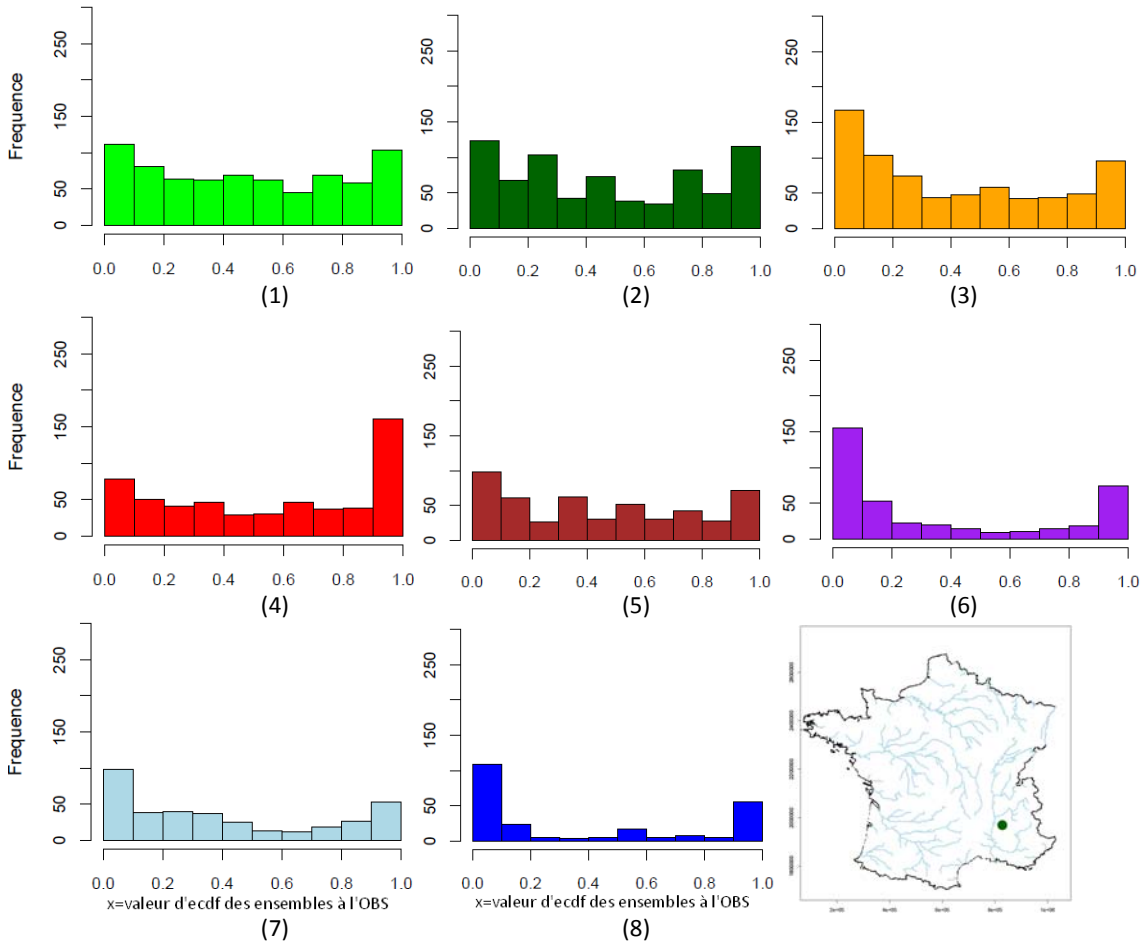


Figure 28 : Histogrammes de PIT pour les prévisions de pluie sur le bassin de la Drôme à Saillant à un horizon de 7 jours des centres météorologiques TIGGE : (1) ECMF, (2) EGRR, (3) RJTD), (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

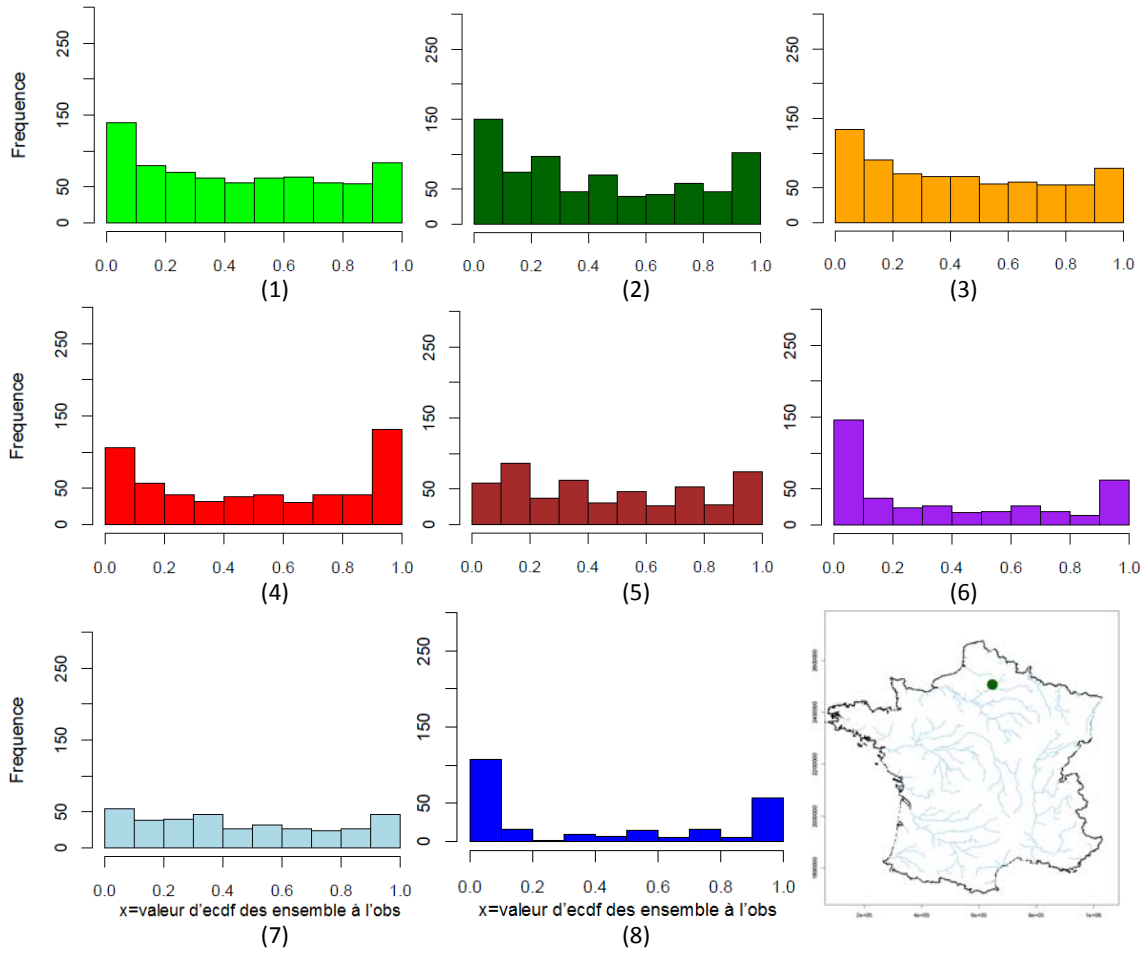


Figure 29 : Histogrammes de PIT pour les prévisions de pluie sur le bassin de l'Oise à Sempigny à un horizon de 7 jours des centres météorologiques TIGGE : (1) ECMF, (2) EGRR, (3) RJTD), (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée.

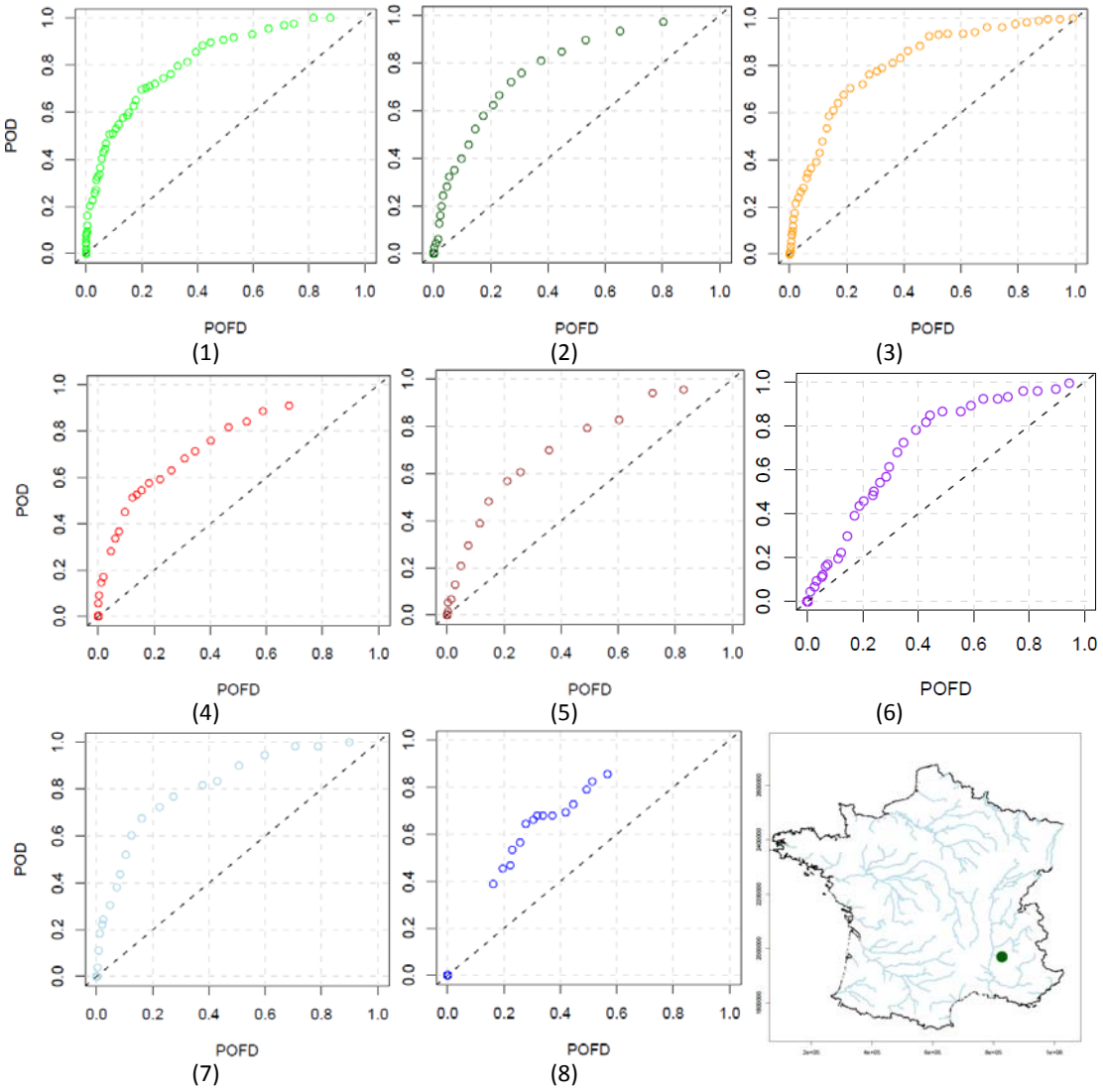


Figure 30 : Les courbes ROC pour le bassin de la Drôme à Saillant et pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – (en parenthèse). La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

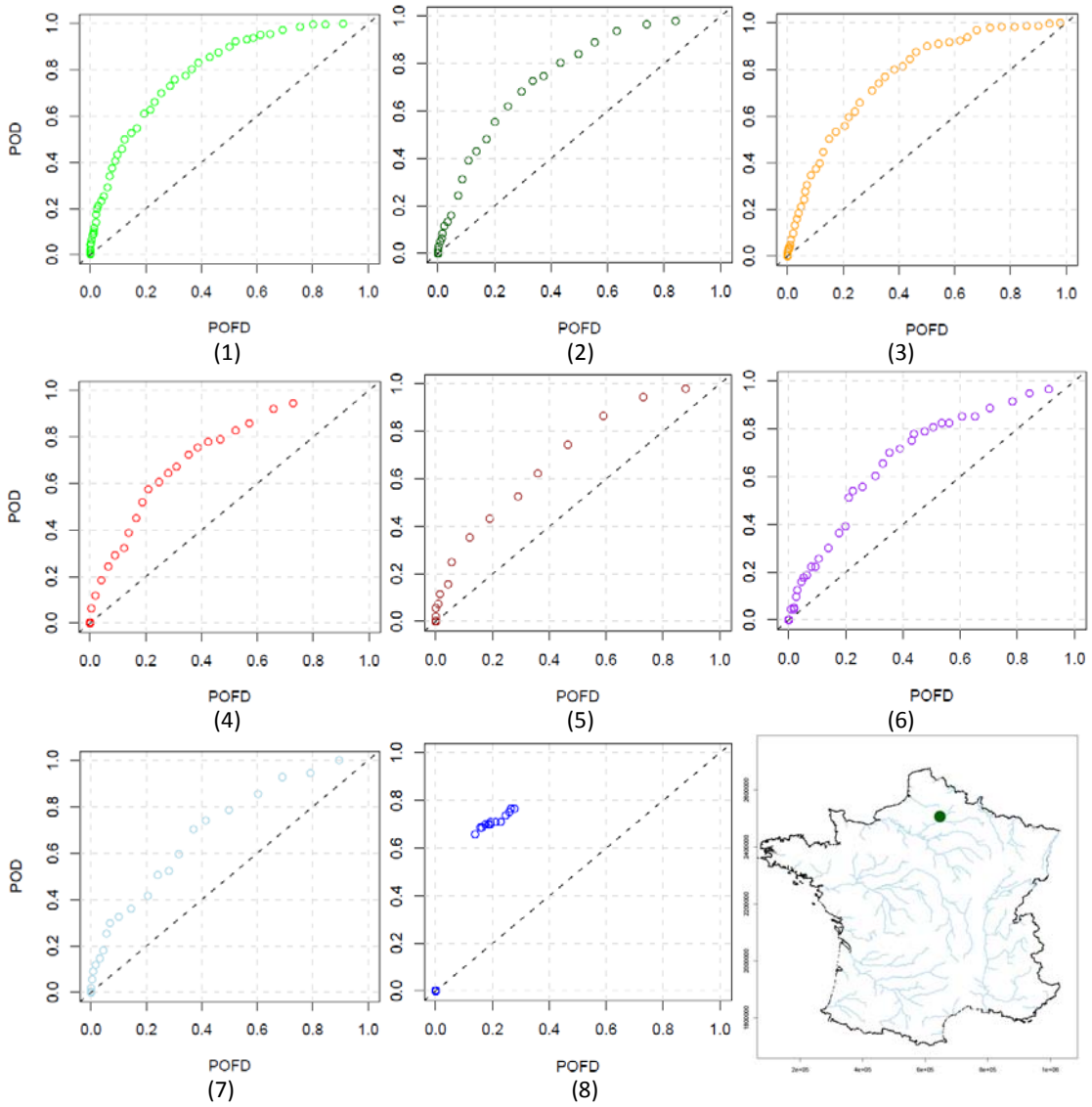


Figure 31 : Les courbes ROC pour le bassin de l'Oise à Sempigny et pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – (en parenthèse). La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée.

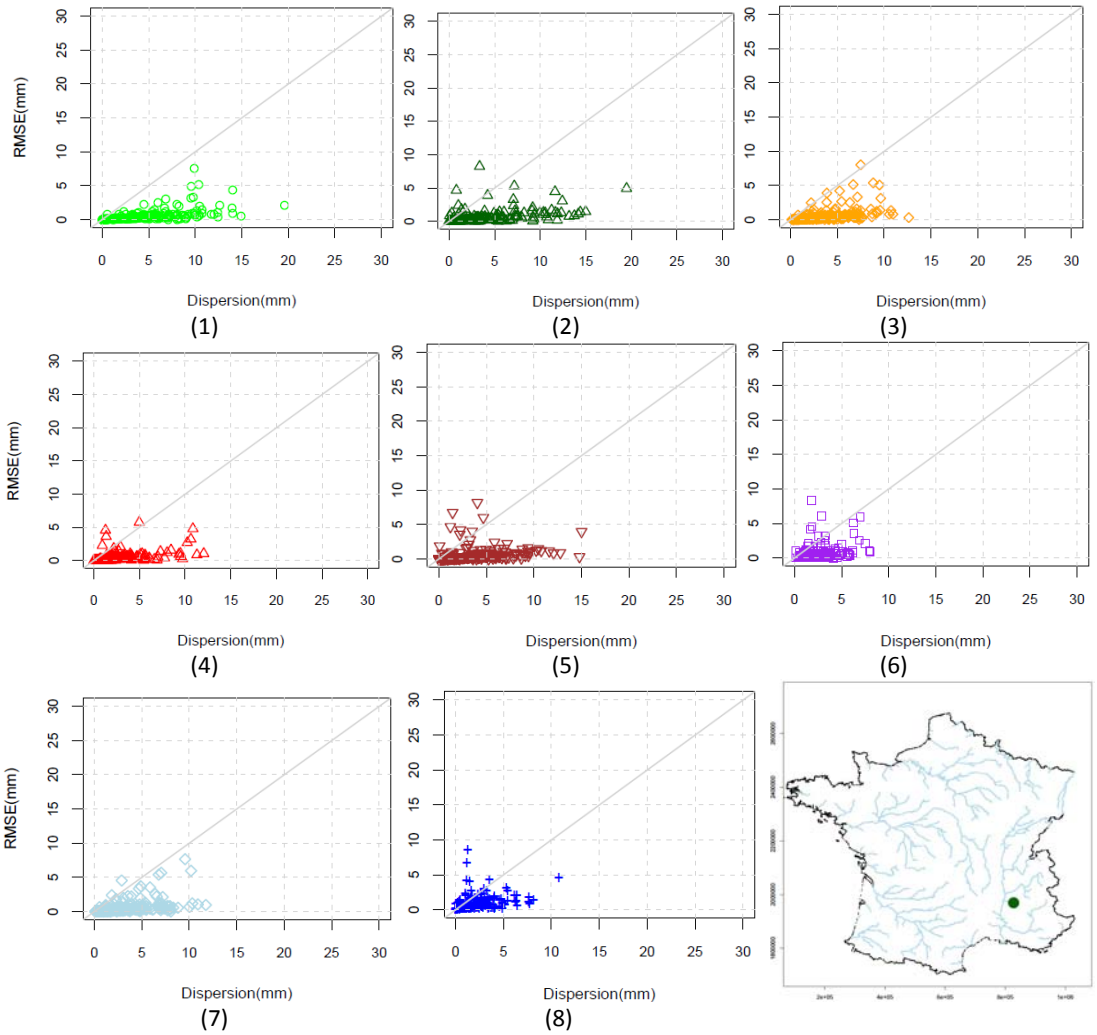


Figure 32 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion de l'ensemble pour le bassin de la Drôme à Saillant pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des diagrammes est faite sur la période d'évaluation commune disponible (242 jours) pour chaque centre. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

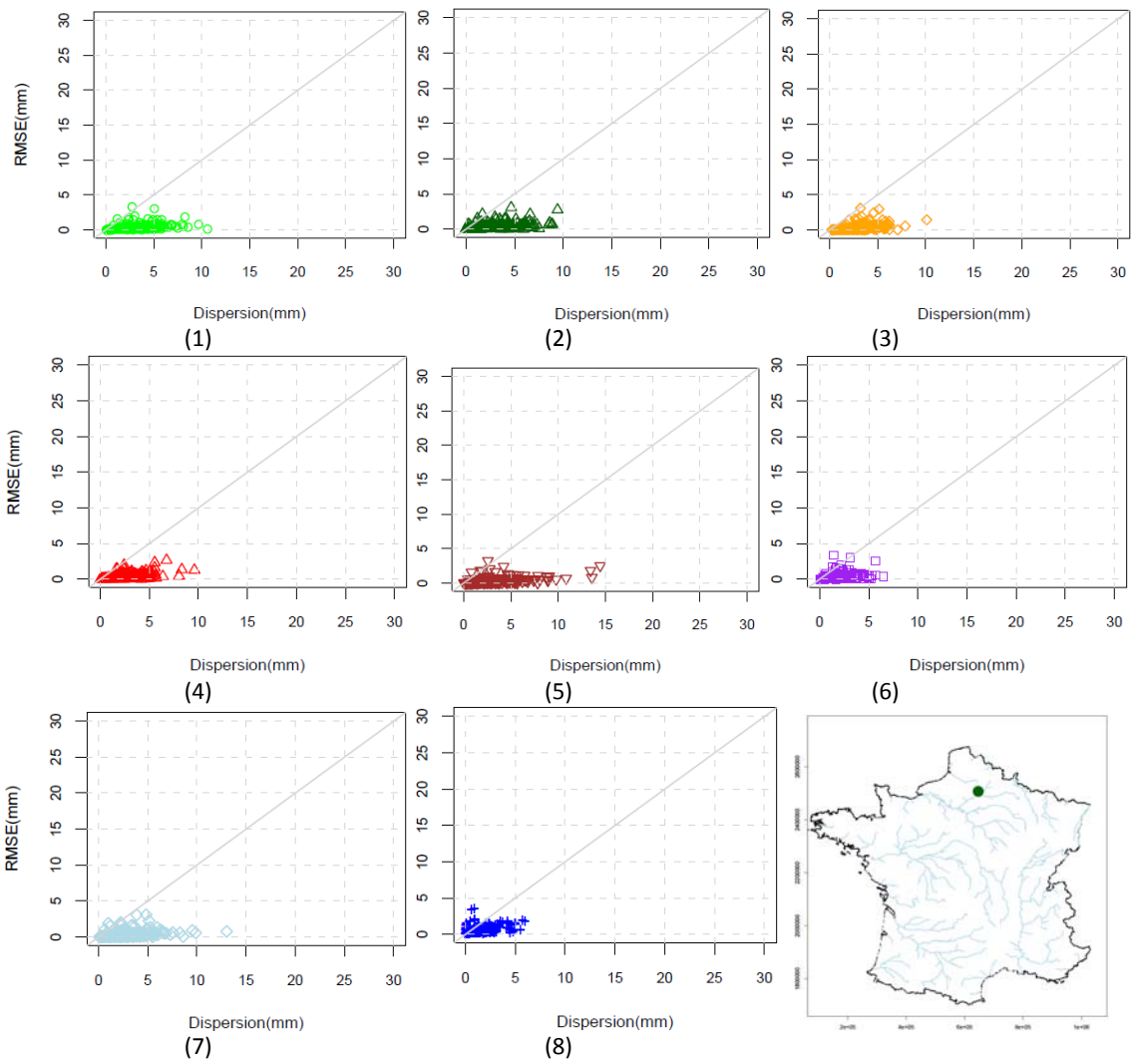


Figure 33 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion de l'ensemble pour le bassin de l'Oise à Sempigny pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des diagrammes est faite sur la période d'évaluation commune disponible (242 jours) pour chaque centre. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée.

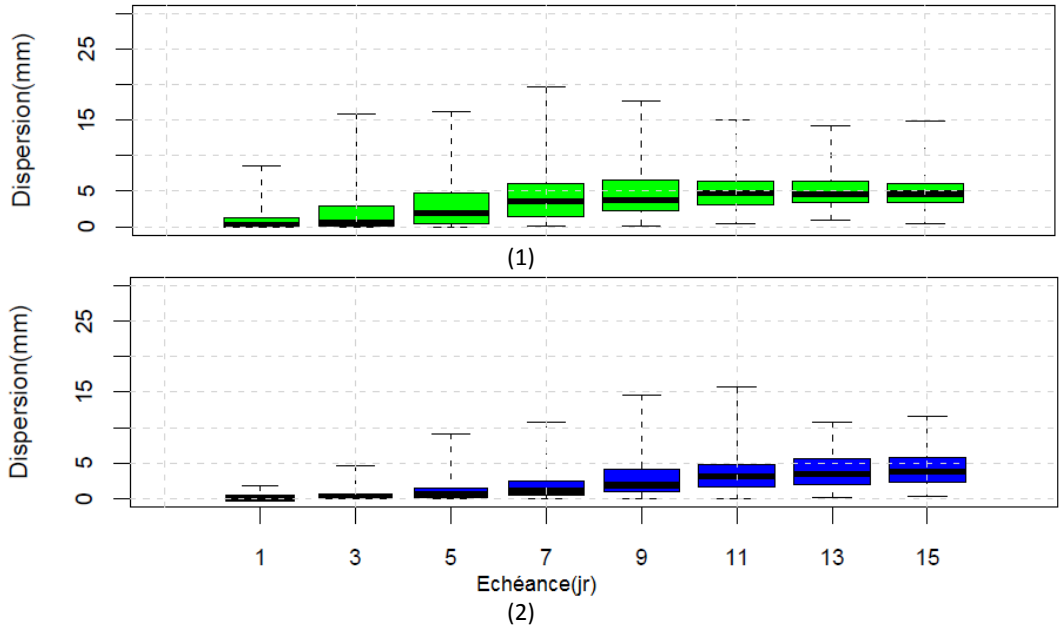


Figure 34 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision pour le bassin de la Drôme à Saillant. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF et (2) SBSJ. La construction des boîtes à moustaches est faite sur la période d'évaluation commune disponible (242 jours).

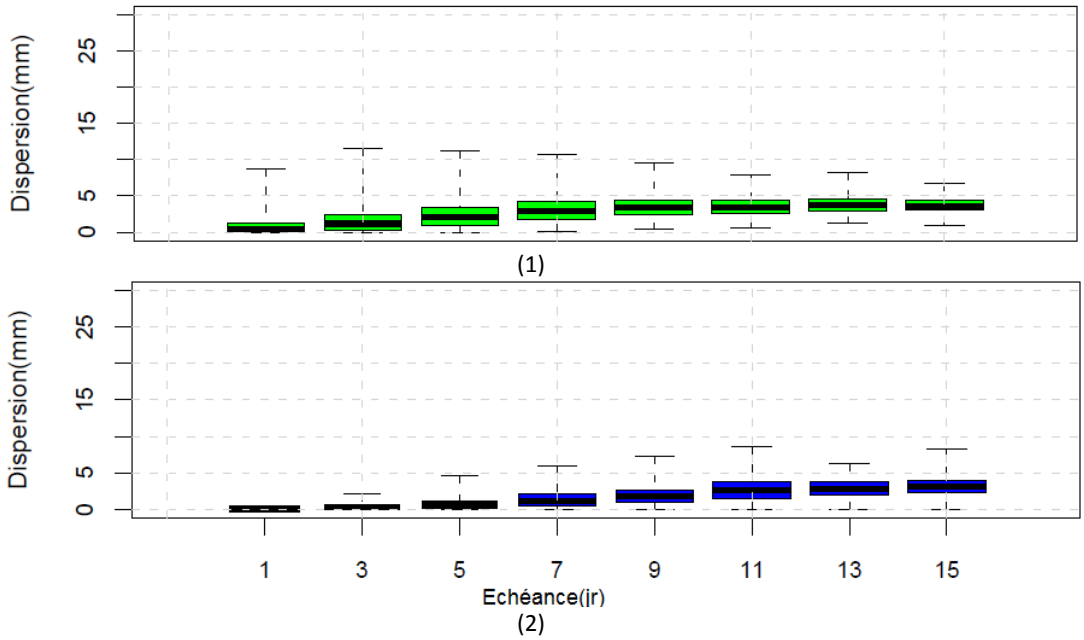


Figure 35 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision pour le bassin de l'Oise à Sempigny. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF et (8) SBSJ. La construction des boîtes à moustaches est faite sur la période d'évaluation commune disponible (242 jours).

5.2.1.3 Critères numériques sur l'échantillon de 74 bassins versants

Nous examinons ici la performance des prévisions d'ensemble de précipitation des 8 centres constituants la base TIGGE pour l'ensemble des 74 bassins versants de notre échantillon à partir des critères numériques suivants : le score *RPS* (Figure 36, Figure 38, Figure 39), le score de divergence *RDS* (Figure 37) et le score de Brier (Figure 40). Le calcul des scores est effectué sur le nombre de jours de prévision disponibles pour tous les centres - 242 jours (voir Chapitre 2, §2.2).

Dans la Figure 36 les performances en termes de score de probabilité ordonnée *RPS* sont présentées sous forme de boîtes à moustache. Dans cette figure, nous examinons en parallèle la distribution des valeurs obtenues sur les 74 bassins versants pour chaque centre météorologique, calculée sur les jours de prévision communs (242 jours). Trois échéances sont présentées : 3, 7 et 9 jours. Rappelons que plus la valeur du score RPS est élevée, moins la prévision correspondante est performante et que le score parfait est égal à zéro.

Nous remarquons que :

- pour toutes les échéances le centre européen *ECMF* et le centre japonais *RJTD* possèdent les meilleurs scores. Les performances de ces deux centres sont comparables, avec une performance légèrement supérieure pour le centre *ECMF* aux horizons plus lointains (7 et 9 jours). Ces deux centres météorologiques utilisent des approches différentes pour la perturbation des conditions initiales, mais ils partagent le même nombre de membres (50) ainsi qu'une résolution spatiale similaire (0.5° x 0.5°). La combinaison de ces deux facteurs (nombre de membres et résolution spatiale) contribue potentiellement à la bonne performance de prévisions de ces centres.

- les centres britannique *EGRR* et australien *AMMC* montrent également de bonnes performances, très proches de celle des meilleurs centres. Cependant, en examinant les caractéristiques principales de ces centres, nous remarquons que les prévisions de pluie sont issues d'une grille spatiale moins dense (0.833° x 0.555° et 1.5° x 1.5° respectivement). De plus, chaque centre fournit les prévisions d'ensemble à partir de 24 et 33 perturbations des conditions initiales respectivement.

- les centres américains *KWBC*, canadien *CWAO* et brésilien *SBSJ* sont les centres les moins performants sur l'échantillon de 74 bassins versants. Cette conclusion est confirmée pour toutes les échéances examinées. Ces centres disposent des ensembles avec un nombre plus limité de membres (21, 21 et 15 respectivement). Ils fournissent également les prévisions de pluie sur une grille plus large (autour de 1°*1° pour tous les trois centres). Par conséquent, leur performance peut être influencée par ces deux paramètres.

Plus généralement, nous n'avons pas pu identifier une corrélation claire entre la methode utilisée par chaque centre pour la perturbation des conditions initiales et la performance du centre en termes de *RPS*. C'est plutôt le nombre de membres, ainsi que la résolution spatiale qui vont de pair avec la performance des prévisions en termes de *RPS*.

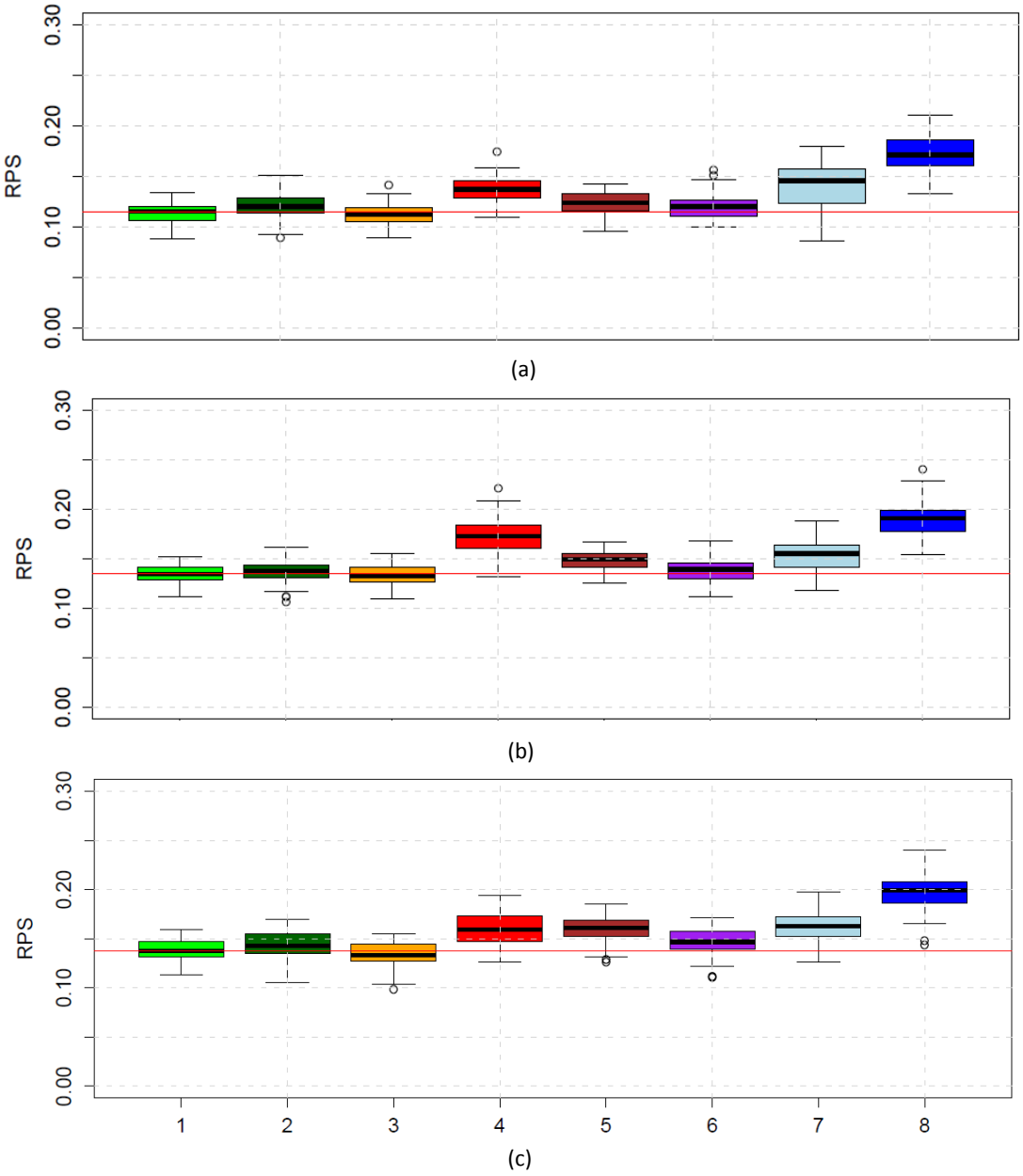


Figure 36 : Boîtes à moustache pour le score de probabilité ordonnée RPS sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les boîtes à moustache sont calculées à partir des prévisions d'ensemble de pluie tirées des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ.

En parallèle au score de probabilité ordonnée, nous examinons également le score de divergence *RDS* (voir Chapitre 3, §3.2). Rappelons qu'il s'agit d'un critère logarithmique quantifiant le gain en information obtenu par chaque prévision d'ensemble par rapport à l'observation. La Figure 37 présente les valeurs obtenues sous forme de boîtes à moustache pour les prévisions de pluie des 74 bassins versants étudiés et pour les échéances 3, 7 et 9 jours. Chaque centre météorologique de la base TIGGE est représenté séparément. Nous remarquons que :

- la performance des différents centres est similaire à celle donnée par le score de probabilité ordonnée *RPS* : les centres les plus performants sont le centre européen *ECMF* et le centre japonais *RJTD*, suivi des centres australien *AMMC* et britannique EGRR. Les centres les moins performants sur l'échantillon de 74 bassins versants français sont les centres américains *KWBC*, canadien *CWAO* et brésilien *SBSJ.* Nous observons également que la performance relative des différents centres entre eux dépend peu de l'échéance de prévision.

Dans un deuxième temps nous examinons la sensibilité de la performance des prévisions d'ensemble de pluie à la surface du bassin versant. Il est important d'investiguer s'il existe une relation entre la performance des prévisions d'un centre et les caractèristiques du bassin versant, comme la surface. La Figure 38 présente la performance en termes de *RPS* pour les échéances 3 jours et 7 jours. Sur l'abscisse x, l'échelle logarthmique est utilisée. Nous remarquons que :

- ces graphiques montrent qu'une légère variation de la valeur du *RPS* avec la surface des bassins versants quel qui soit le centre météorologique TIGGE. Les performances semblent diminuer avec la surface pour les bassins versants au-dessous de 2000 km². A ces échelles les aggravations locales de la convection, mal représentées par les modèles météorologique à maille large, commencent à devenir sensibles. Pour les surfaces au-dessus de 2000 km², les performances en termes de RPS varient peu avec la surface. Néanmoins, il faut noter que l'échantillon de 74 bassins versants est constitué à partir de grands bassins avec une surface supérieure à 1000 km². Par conséquent, la plupart des bassins versants disposent de plusieurs points de grille sur lesquelles les prévisions sont issues. Ce fait peut potentiellement expliquer la faible variabilité du score RPS en fonction de la surface, surtout pour les plus grands bassins versants.

- les performances en termes de *RPS* varient également selon les centres en raison de différents nombres de membres et de résolutions spatiales, mais elles varient peu en fonction de la surface du bassin versant.

Nous obtenons des résultats similaires pour la variabilité du score de divergence *RDS* avec la surface.

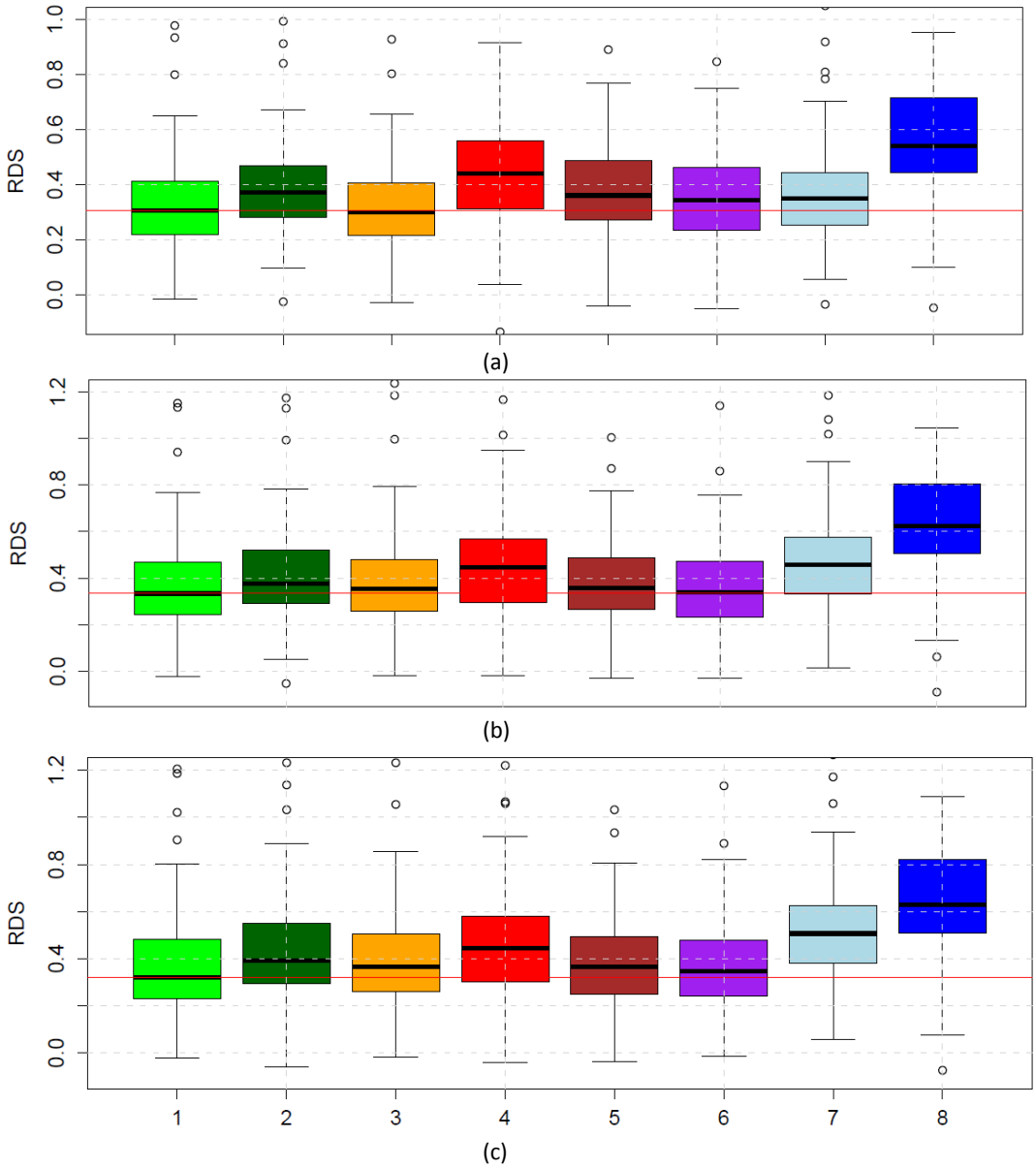


Figure 37 : Boîtes à moustache pour le score de divergence sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions d'ensemble de pluie sont extraits des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ.

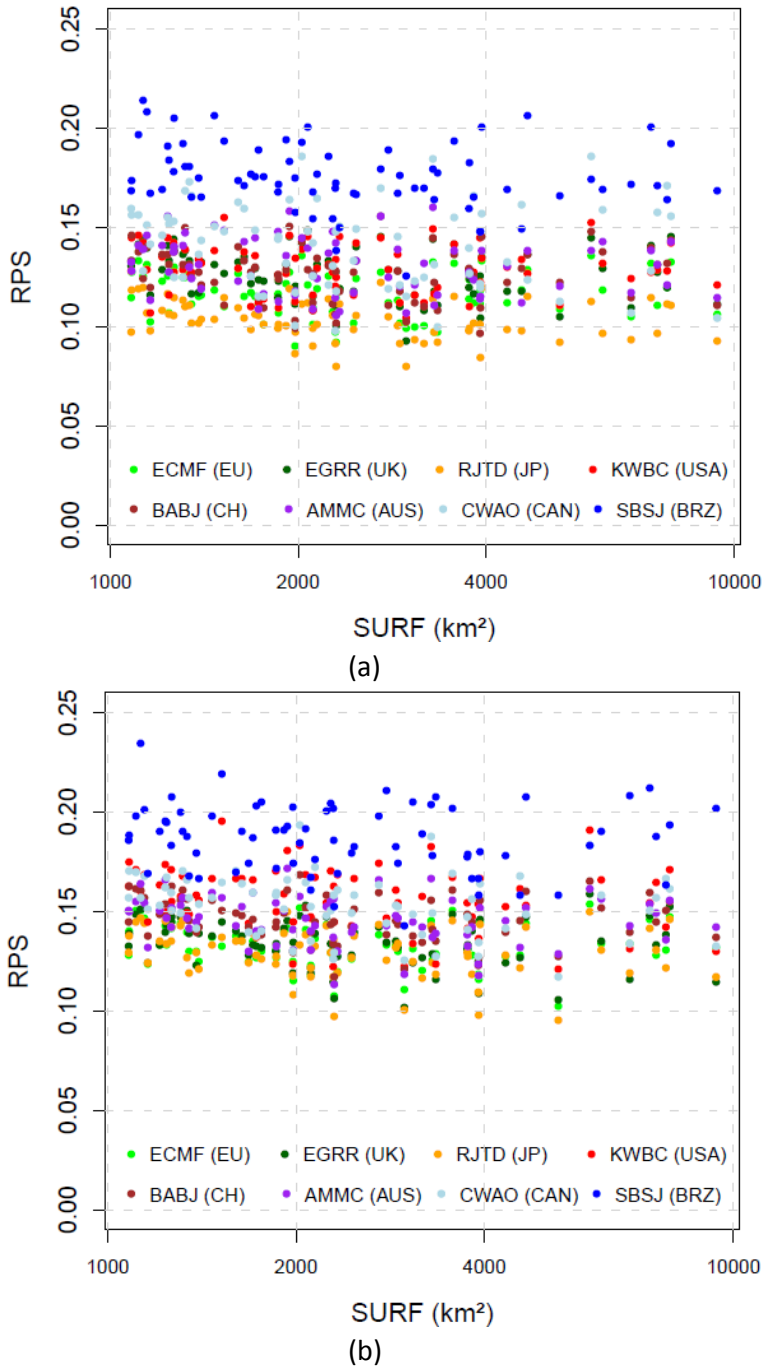


Figure 38 : Graphique du score de probabilité ordonnée RPS en fonction de la surface du bassin versant pour les 74 bassins versants étudiés et pour les échéances 3 (a) et 7 (b) jours. Les prévisions d'ensemble de pluie sont extraites des centres : ECMF (vert clair), EGRR (vert foncé), RJTD (orange), KWBC (rouge), BABJ (marron), AMMC (mauve), CWAO (ciel) et SBSJ (bleu).

Afin d'examiner quel centre est le plus performant, bassin par bassin, et s'il y a une localisation géographique en France où un centre présente une meilleure performance, nous avons tracé une carte indiquant, pour chaque point géographique d'un bassin versant, le centre météorologique avec la meilleure performance. Le critère utilisé ici est le score de probabilité ordonnée *RPS* (Figure 39) et le score de Brier *BS* pour le non-dépassement du quantile 90% (BSQ90) (Figure 40). Les cartes sont présentées pour les échéances de 3, 5, 7 et 9 jours. Puisqu'il est possible que plusieurs centres aient une bonne performance ou une performance équivalente pour un même bassin, nous affichons les résultats sous forme d'un diagramme circulaire avec les deux meilleurs centres par bassin versant. Ce diagramme circulaire représente également la proportion obtenue par chaque centre : si les deux meilleurs centres ont par exemple des valeurs identiques de score, le cercle est à moitié coloré par la couleur d'un centre et à moitié coloré par la couleur d'un autre centre.

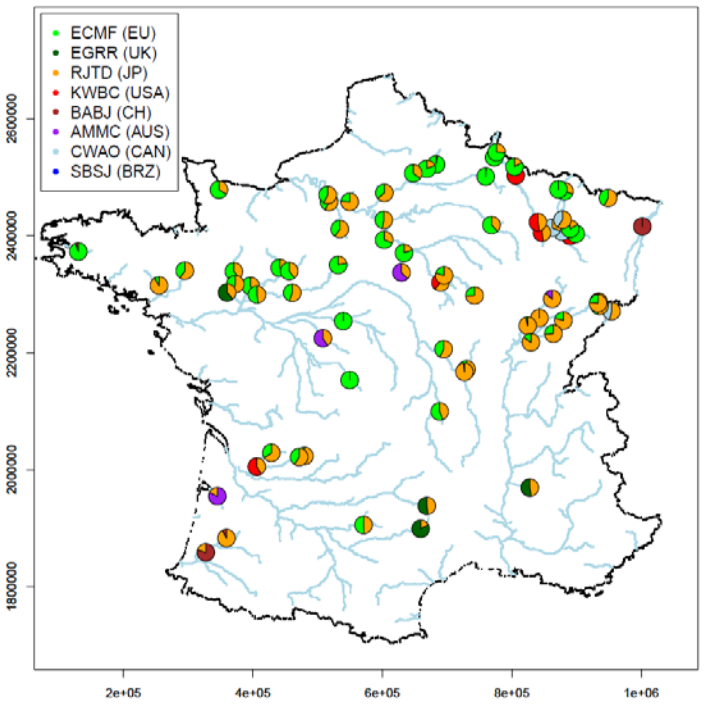
De l'analyse des Figure 39 et Figure 40 nous remarquons que :

- pour un horizon de prévision jusqu'à 5 jours, le centre européen *ECMF* et le centre japonais *RJTD* partagent la meilleure performance sur la plupart de bassins versants de notre base de données : 87% des bassins versants partagent la meilleure performance entre ces deux centres pour l'échéance 3 jours et 5 jours. Pourtant les prévisions de pluie extraites du centre australien *AMMC,* ainsi que celles tirées du centre canadien *CWAO* apportent le meilleur score pour certains bassins versants : 12 à 19% des bassins versants – selon l'échéance de prévision - ont un de ces deux centres parmi les deux meilleurs. Ces bassins sont plutôt caractérisés par des précipitations fortes durant les mois d'automne et d'hiver ; ainsi que par des écoulements plus uniformes avec des pics durant les mois d'hiver et de printemps, et sont notamment situés dans les bassins de l'Isle, pour le centre *AMMC* et dans les bassins de la Meuse, pour le centre *CWAO*. Ces observations sont confirmées également par la Figure 38, où nous remarquons que pour les bassins de la Meuse par exemple, (surfaces entre 3000 et 7000 km²), le centre canadien obtient la deuxième meilleure valeur de *RPS* après le centre japonais *RJTD*. Parmi les deux centres les plus performants, nous observons également une présence du centre canadien *CWAO* également dans le bassin de l'Eure à Saulles (avec une surface de 1700 km²) pour l'échéance de 3 jours, qui néanmoins disparaît avec l'augmentation de l'échéance, et dans le bassin de l'Isle pour l'échéance de 5 jours.

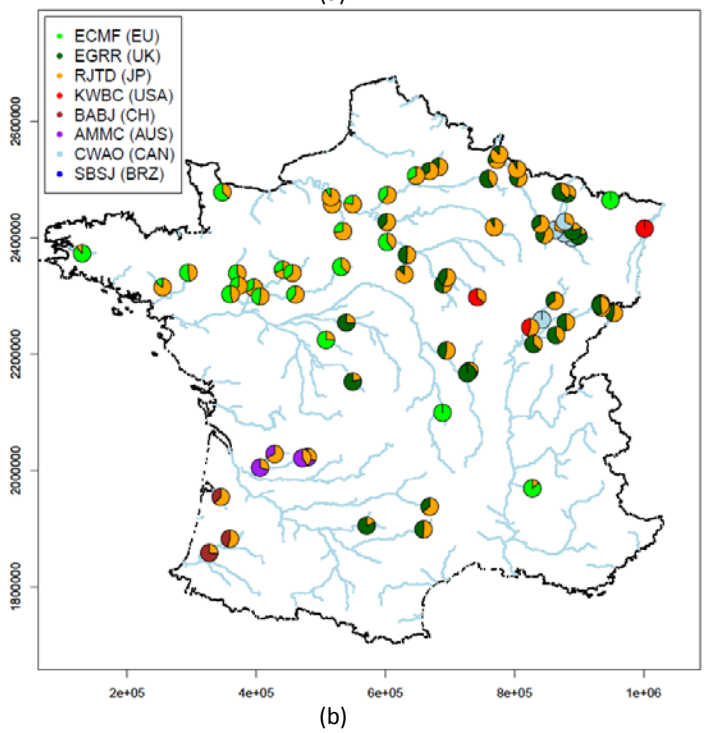
- pour les horizons plus lointains, le centre européen *ECMF* reste parmi les deux plus performants dans la majorité des bassins (51%), même si la présence d'autres centres apparaît de façon plus marquée. A noter également que le nombre de bassins versants pour lesquels le centre japonais *RJTD* apporte le meilleur score se réduit avec l'augmentation de l'échéance. De plus, le centre australien *AMMC*

apparaît avec plus de fréquence, notamment sur un axe SO-NE coupant la France. Nous constatons que le centre australien montre la meilleure performance sur les bassins de l'Isle, de la Sauldre, de l'Indre et de la Meuse.

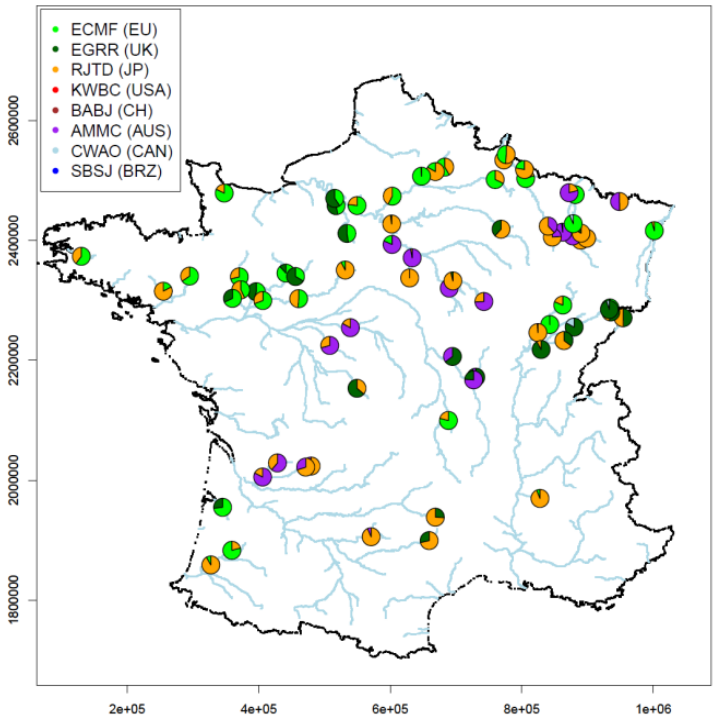
- Quand nous nous focalisons sur les événements plus forts (Figure 40) nous remarquons que pour les horizons de prévision courts (3 jours), le centre européen *ECMF* reste celui avec la meilleure performance en termes de score de Brier *BS* pour la majorité des bassins versants examinés (58 parmi les 74 bassins versants). Néanmoins, sa performance se dégrade aux horizons plus lointains, au bénéfice d'autres centres qui voient leur performance figurer parmi les deux meilleures : c'est le cas du centre australien *AMMC* et du centre canadien *CWAO*. Globalement, lors de cette analyse centrée sur les événements forts, nous observons une plus grande variabilité des centres qui figurent parmi les deux meilleurs. Nous remarquons aussi une séparation Nord-Sud plus nette quand nous regardons les échéances plus lointaines : le centre européen *ECMF* et le centre japonais *RJTD* se partagent les performances dans la moitié sud de la France, le centre australien *AMMC*, canadien *CWAO* et brésiliens *SBSJ* dominent les performances des bassins localisés dans la moitié Nord du pays. Les bassins du sud de la France sont caractérisés plutôt par des étés secs et des précipitations fortes surtout dans les mois d'automne. Par contre les bassins dans le nord de la France se caractérisent par une distribution plus homogène des précipitations tout au long de l'année.


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

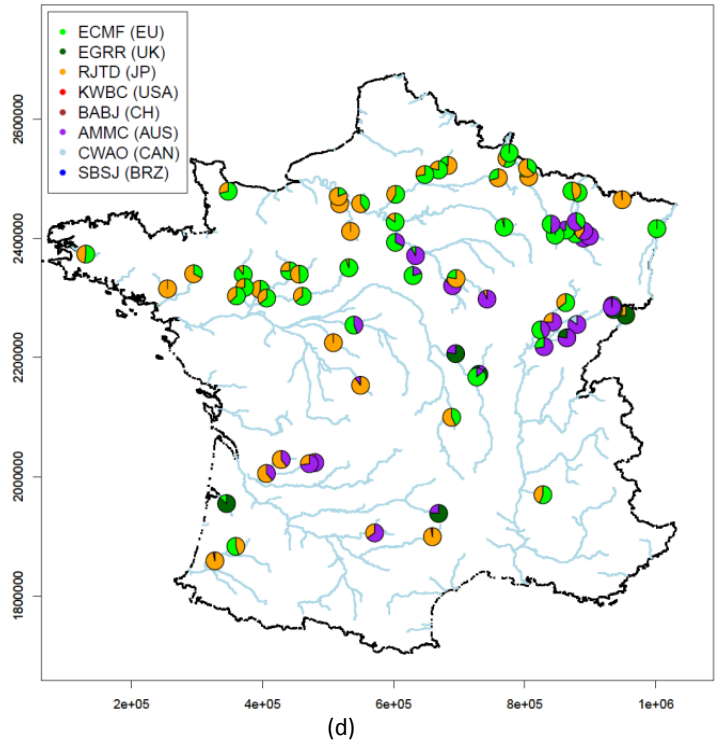
(a)


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

(b)

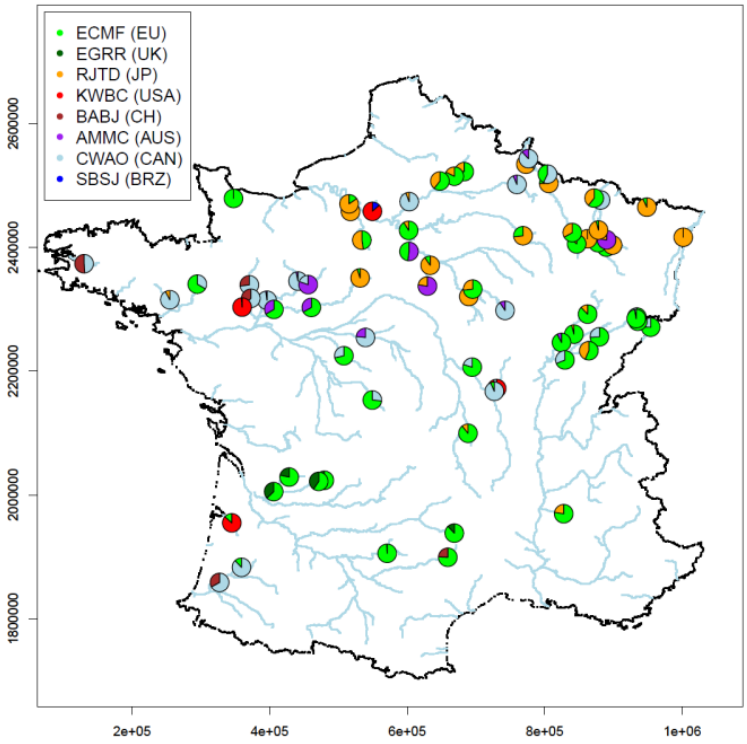


(c)

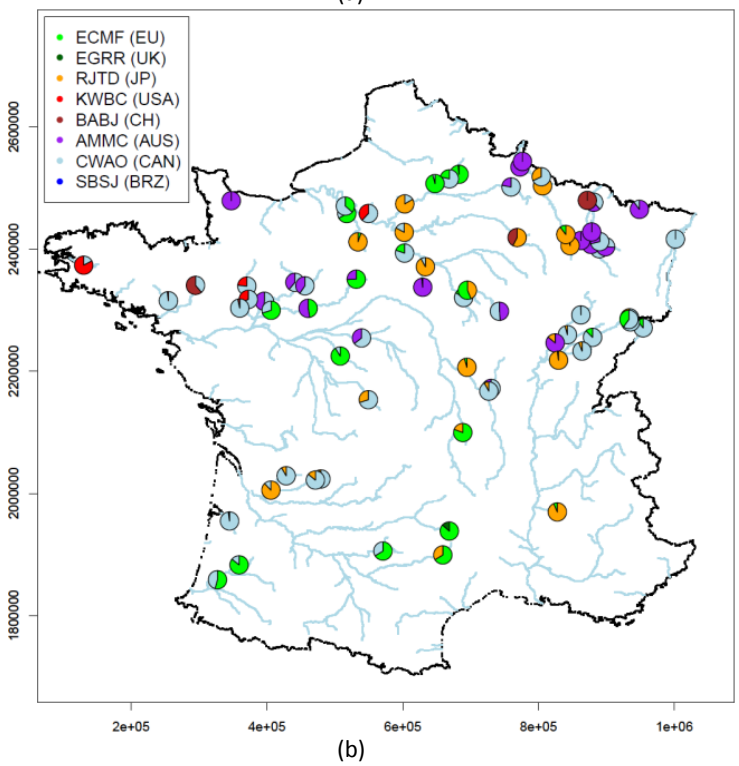


(d)

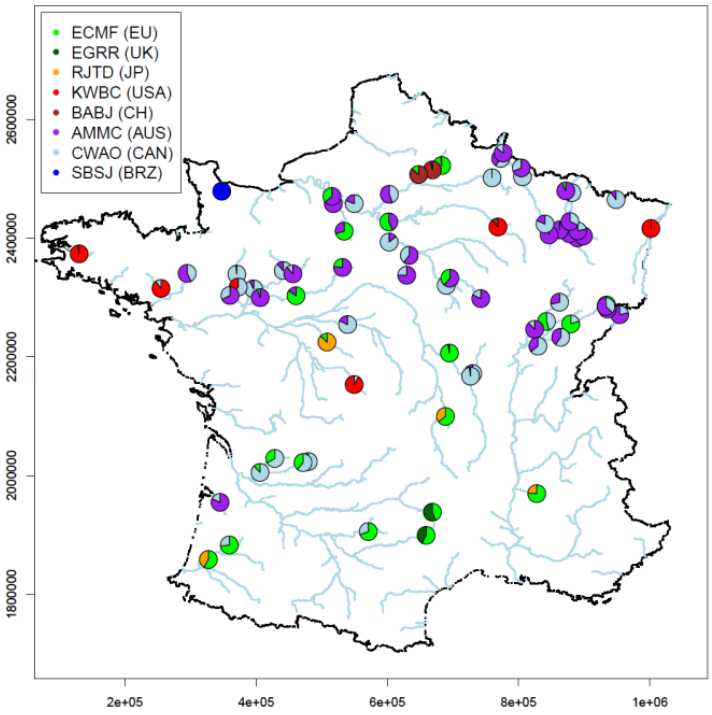
Figure 39 : Carte de localisation des centres qui donnent la meilleure performance selon le critère RPS pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d).


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

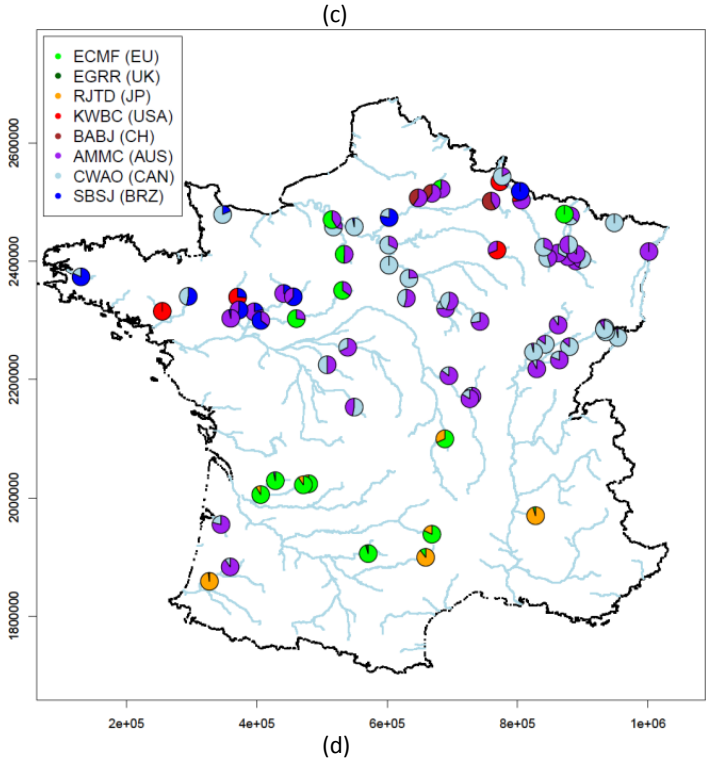
(a)


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

(b)



(c)



(d)

Figure 40 : Carte de localisation des centres qui donnent la meilleure performance selon le critère BSQ90 pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d).

5.2.2 Évaluation des prévisions d'ensemble de débits

Un des axes principaux de cette étude se trouve dans l'évaluation de la qualité des prévisions d'ensemble de débits. Rappelons qu'ici nous utilisons le modèle hydrologique *GRP*, alimenté par les prévisions d'ensemble de pluies des différents centres météorologiques de la base TIGGE.

5.2.2.1 Illustration pour un événement particulier

Nous reprenons ici l'événement du 25 mai 2008 sur les bassins versants de la Durance à Serre Ponçon et de la Drôme à Saillants. Les débits observés dans ces bassins versants ont atteint, en moyenne journalière, les 775 m3/s et 59 m3/s respectivement. Nous notons que lors de cet événement le barrage de Serre Ponçon a deversé, pour la deuxième fois de son histoire, des quantités d'eau importantes.

La Figure 41 présente les hydrogrammes de la prévision d'ensemble sous forme d'intervalle inter-quantile Q20%-Q80% des valeurs prévues émises le 28 mai 2008 à 3 jours d'échéance sur le bassin versant de la Drôme à Saillant. Les hydrogrammes ont été obtenus à partir du modèle *GRP*, utilisant en entrée les précipitations prévues par chacun des centres météorologiques de la base TIGGE. La ligne noire épaisse représente le débit observé et le faisceau en gris clair correspond à l'interquartile entre 20 et 80%.

Nous observons que les faisceaux de prévisions de débits basés sur les prévisions de précipitation des centres européen *ECMF*, japonais *RJTD*, chinois *BABJ*, canadien *CWAO* et brésilien *SBSJ* englobent totalement ou en partie l'observation. Notons, que le centre européen produise le faisceaux de prévisions le plus large. Il y a pourtant une tendance à sous-estimer les débits observés. Ce fait peut s'expliquer en partie par la sous-estimation de la pluie prévue par les prévisions d'ensemble météorologiques (Figure 27).

Nous notons cependant que ceci n'est qu'un exemple de visualisation de la prévision d'ensemble de débits et ne concerne qu'un jour de prévision pour une échéance et un bassin versant donnés. L'évaluation des prévisions hydrologiques d'ensemble présentée par la suite concernera l'ensemble des jours de la période d'évaluation, des échéances examinées et des bassins versants étudiés.

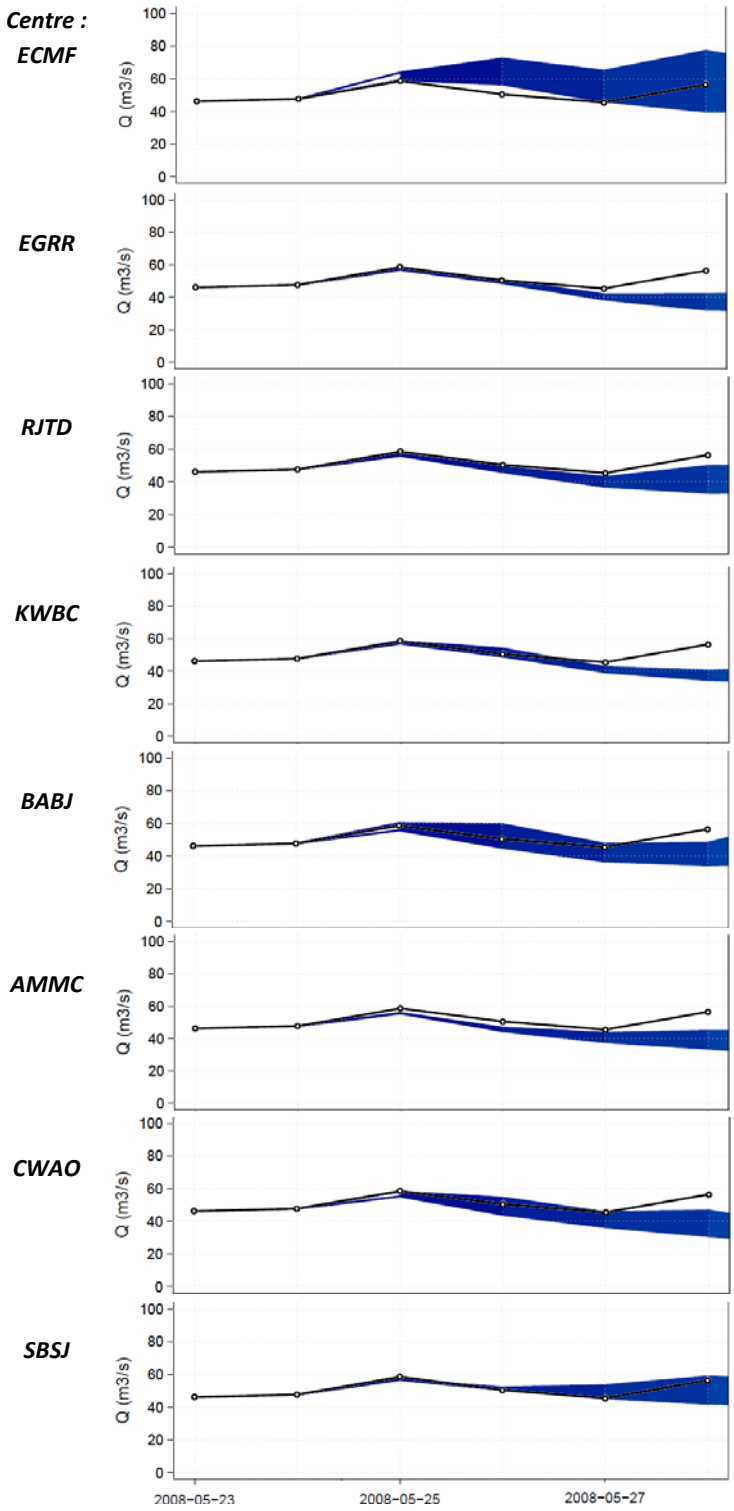


Figure 41 : Prévisions d'ensemble de débits émises le 28 mai 2008 par le modèle GRP sur le bassin versant de la Drôme à Saillant pour les 5 jours à venir et selon les 8 centres météorologiques de la base TIGGE de cette étude (intervalle inter-quantile Q20%-Q80%). En trait plein, les débits observés.

5.2.2.2 Critères graphiques pour deux bassins versants représentatifs

Ainsi que pour les prévisions de précipitation, nous examinons ici les attributs : fiabilité (Figure 42 et Figure 43), discrimination (Figure 44 et Figure 45) et relation erreur-dispersion des membres (Figure 46 et Figure 47). Les deux bassins versants représentatifs de notre échantillon sont utilisés à titre d'illustration : le bassin versant de la Drôme à Saillant et le bassin versant de l'Oise à Sempigny. Les conclusions que nous tirons de l'analyse sont cependant généralisables sur l'ensemble de notre échantillon de 74 bassins versants.

Fiabilité des prévisions : histogramme de PIT

Les principales remarques que nous tirons de ces résultats sont listées ci-après.

- Pour le bassin versant de la Drôme à Saillant les prévisions de précipitations des centres européen *ECMF* (1), japonais *RJTD* (3), chinois *BABJ* (5) et canadien *CWAO* (7) se sont montrées relativement fiables, en même temps que les prévisions des autres centres ont montré une tendance à la sous- ou surestimation de l'observation (Figure 28). Pour le cas du bassin versant de l'Oise à Sempigny le manque de fiabilité est très marqué, avec des prévisions fortement sous-dispersives (Figure 29).

- Sur notre échantillon de 74 bassins versants, et quel qui soit le centre météorologique de la base de données TIGGE, les prévisions d'ensemble de débit se montrent considérablement moins fiables par rapport aux prévisions de précipitation (Figure 42 et Figure 43). Nous observons des histogrammes de PIT fortement marqués par une forme en " U ", c'est-à-dire, que les prévisions sont sous-dispersives et souvent biaisées. Ce fait est observé pour tous les centres météorologiques de la base de données TIGGE et quel que soit le bassin versant examiné. Il est notamment plus marqué pour les horizons de prévision plus courts. Pour des horizons de prévisions plus lointains, les prévisions d'ensemble de débit deviennent plus dispersées et donc plus fiables.

Cette remarque nous conduit à la conclusion que le modèle hydrologique affecte la fiabilité des prévisions, notamment sur les échéances plus courtes, en augmentant le nombre de fois où les membres de la prévision d'ensemble sont i) soit plus petits que l'observation (sous-estimation de l'observation), occasionnant une augmentation de la fréquence des probabilités de prévision faibles (barres hautes à gauche des histogrammes PIT), ii) soit plus élevés que l'observation (surestimation de la prévision d'ensemble), occasionnant une augmentation de la fréquence des probabilités de prévision élevées (barres hautes à droite des histogrammes PIT). Il est aussi probable que cette erreur d'estimation est observée dans le cas de pluies observées et prévues nulles.

Discrimination : courbes ROC

D'après l'analyse des Figure 44 et Figure 45 nous observons que :

- sur l'ensemble des 74 bassins versants, le passage des prévisions d'ensemble de précipitation par le modèle hydrologique GRP influence l'attribut discrimination. Le dépassement du quantile 70% avec une probabilité de bonne détection supérieure à 60% est prévu par presque tous les centres de la base TIGGE. Nous tirons la conclusion que les prévisions d'ensemble de débit présentent une bonne capacité à discriminer entre les différents événements. De plus cette capacité est significativement améliorée en passant par le modèle hydrologique, car les prévisions d'ensemble de pluie se caractérisent par des taux plus faibles de bonne détection POD. Cette conclusion est confirmée également par l'aire sous la courbe ROC qui est plus elevée par rapport à celle des courbes ROC de pluie pour tous les centres. Nous remarquons également que les prévisions tirées du centre européen *ECMF* se montrent les plus performantes en termes de courbes ROC à un horizon de prévision jusqu'à 10 à 15 jours selon le bassin versant. De plus, selon le bassin versant, d'autres centres se montrent également performants, comme le centre japonais *RJTD* et canadien *CWAO*. Nous constatons à nouveau que la performance des centres est indépendante du nombre de jours de prévisions disponibles.

- en examinant tous les horizons de prévision, nous constatons aussi que ce comportement est influencé par l'horizon de prévision : par exemple dans le cas du bassin versant de la Drôme à Saillant et pour les prévisions de débits *GRP* basées sur les prévisions de précipitation du centre européen *ECMF* le nombre de membres, qui prévoient le dépassement du quantile 70% avec une probabilité de bonne détection 60%, passe de 0 membres sur 50 pour l'horizon de 3 jours à 27 membres pour celui de 15 jours. En ce qui concerne les autres centres météorologiques, les systèmes se montrent moins performants aux horizons plus lointains, par exemple pour un horizon de prévision de 15 jours (centres britannique *EGRR* et brésilien *SBSJ*), de 11 jours (centres américain *KWBC* et canadien *CWAO*) et de 9 jours pour les autres centres. Au-delà de cet horizon les prévisions présentent des probabilités de bonne détection et de fausses alertes égales. Par conséquent le pourcentage d'événements prévus quand ils sont observés est le même que celui des événements prévus quand ils ne sont pas observés. En examinant les autres seuils de non-dépassement, nous constatons que le nombre de membres qui prévoient les dépassements avec une probabilité de détection 60% diminue en augmentant le seuil de 10% à 90% et ce quel qui soit le centre météorologique à la base des prévisions de précipitation : dans le cas du bassin versant de la Drôme à Saillant et du centre européen *ECMF* par exemple, quand le seuil de dépassement passe du quantile 50% des débits observés

(médiane des valeurs observées) au quantile 90%, le nombre de membres qui prévoient l'événement avec une probabilité de détection 60%, varie de 50 à 13 membres sur le total de 50 membres. Cette conclusion reflète la tendance reproduite par tous les centres.

- l'allure des courbes ROC peut varier selon le bassin versant : quand on compare le cas du bassin versant de la Drôme à Saillant (Figure 44) à celui du bassin de l'Oise à Sempigny (Figure 45), nous constatons que les prévisions de débit de l'Oise montrent une meilleure performance en termes de discrimination. Plus particulièrement, nous observons que les prévisions de tous les centres prévoient le dépassement du quantile 70% avec une probabilité de détection POD supérieure à 70%, voire 80%. Même si nous remarquons une légère baisse de la POD sur ce bassin en fonction de l'échéance, toutefois cette probabilité reste supérieure à 60% pour toutes les échéances examinées.

Dispersion des membres

Un autre attribut que nous examinons concerne la dispersion de membres de l'ensemble. Dans les Figure 46 et Figure 47 nous traçons la dispersion de l'ensemble en fonction de la racine carrée de l'erreur quadratique moyenne *RMSE* pour le bassin versant de la Drôme à Saillant et de l'Oise à Sempigny sur un horizon de 7 jours. Nous remarquons en général que :

- la corrélation entre la dispersion des membres et la racine carrée de l'erreur quadratique moyenne est plus significative pour les prévisions de débits que pour les prévisions de pluie. Une explication probable est liée à l'influence du modèle hydrologique, qui a tendance à écraser la dispersion des membres même quand celle-ci est présente dans les prévisions météorologiques, ainsi qu'à la sensibilité des bassins aux petites pluies. Il peut aussi être lié à la nature même du bassin versant et de son rôle intégrateur des processus, mais également à la procédure de mise à jour du modèle de prévision *GRP*. La mise à jour du modèle *GRP* est bassé sur le dernier débit observé avant l'émission de la prévision (voir description du modèle dans Chapitre 2, §2.4). Ces conclusions sont valables pour l'ensemble de 74 bassins versants.

- nous observons également que le comportement de cette relation peut varier selon le bassin versant et l'horizon de prévision. Dans le cas du bassin versant de la Drôme à Saillant, par exemple, la corrélation entre la dispersion et l'erreur quadratique moyenne à l'échéance de 7 jours n'est pas forte pour la majorité des centres (Figure 46), ce qui fait que la partie de l'erreur de prévision expliquée par la dispersion des membres reste limitée. Au contraire, le cas du bassin de l'Oise à Sempigny (Figure 47) montre que la racine carrée de l'erreur quadratique moyenne et la dispersion des

membres de l'ensemble sont fortement corrélées, et ce quel qui soit le centre météorologique et l'horizon de prévision consideré. Nous remarquons également que la dispersion des membres augmente plus rapidement avec l'échéance sur certains bassins par rapport à d'autres (comme c'est le cas du bassin de la Drôme par rapport à celui de l'Oise, Figure 48 et Figure 49).

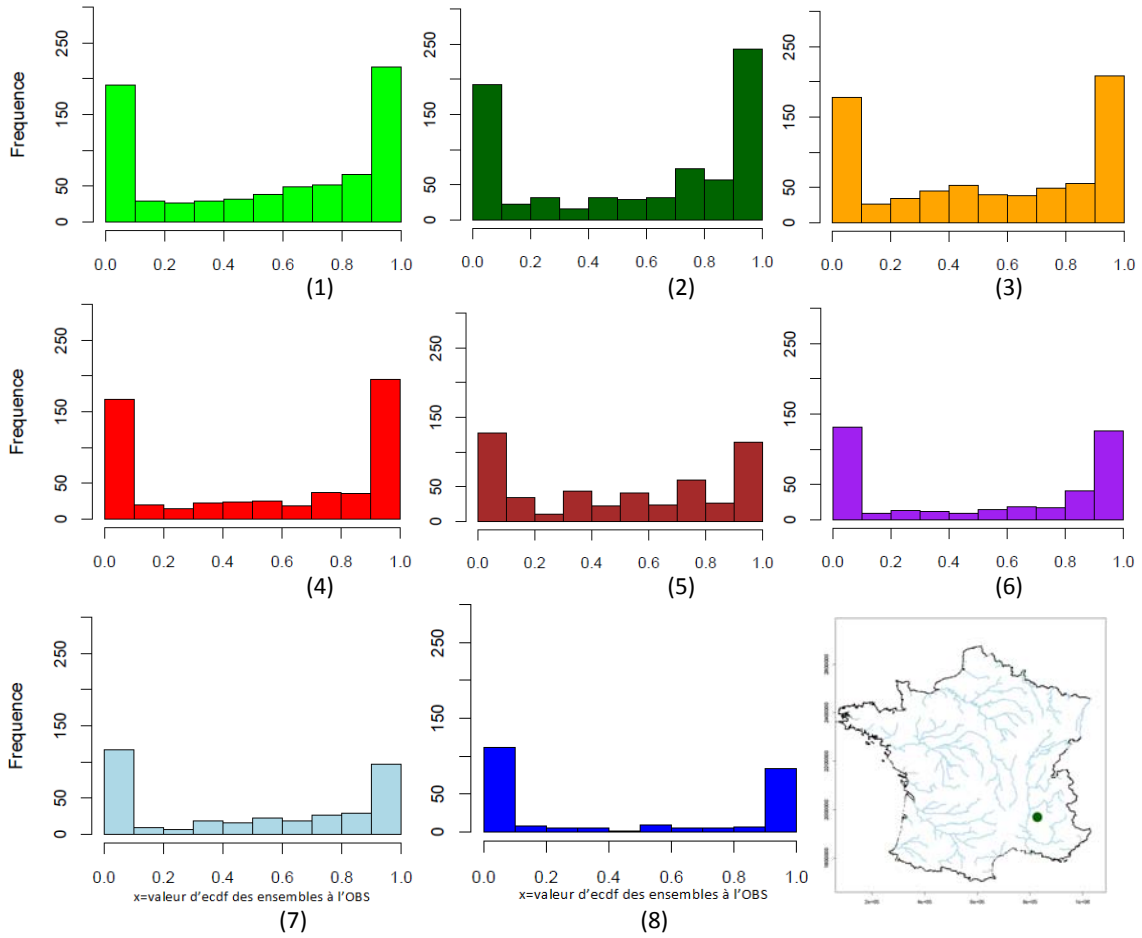


Figure 42 : Histogrammes de PIT pour le bassin de la Drôme à Saillant à partir des prévisions de débit à un horizon de 7 jours. Notons que les prévisions de pluie sont issues des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

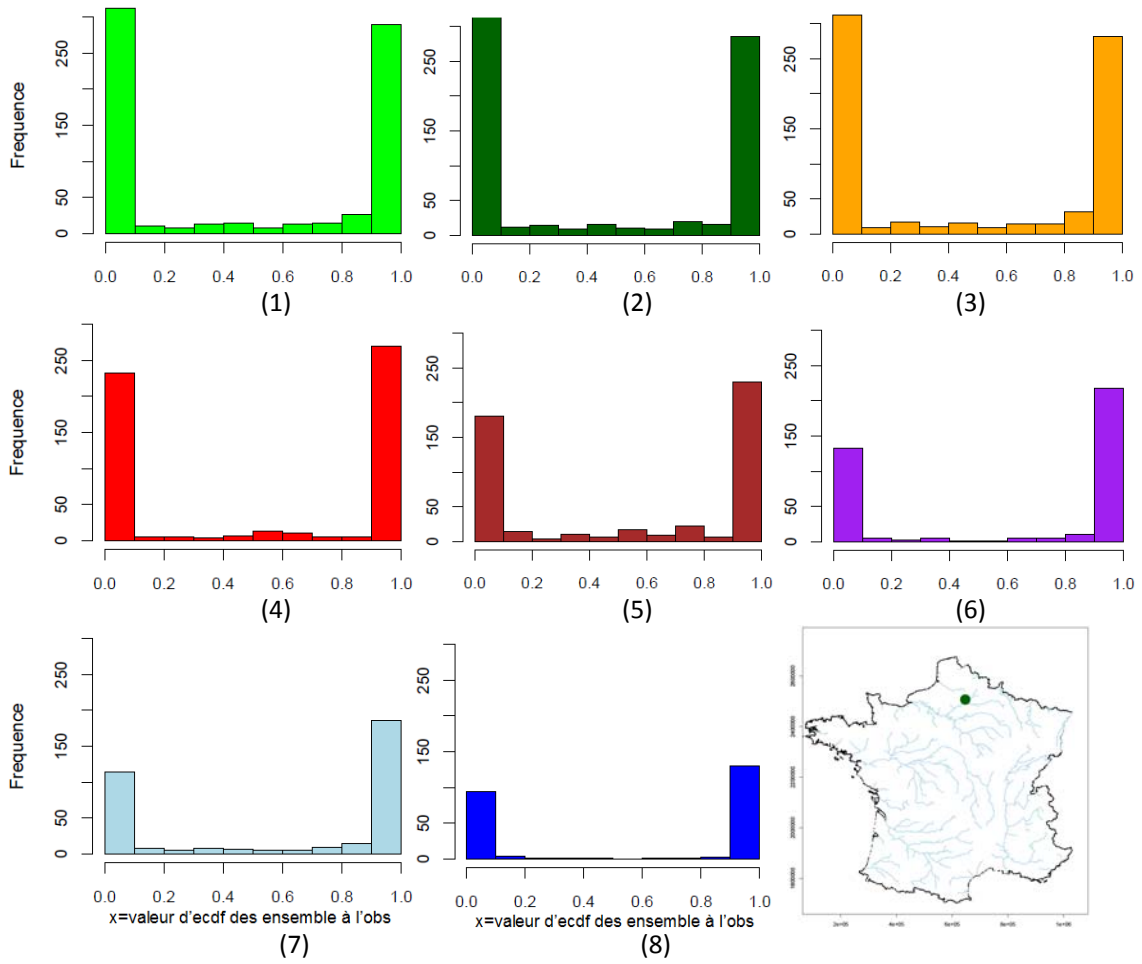


Figure 43 : Histogrammes de PIT pour le bassin de l'Oise à Sempigny à partir des prévisions de débit à un horizon de 7 jours. Notons que les prévisions de pluie sont issues des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempigny sur la carte de France est également présentée.

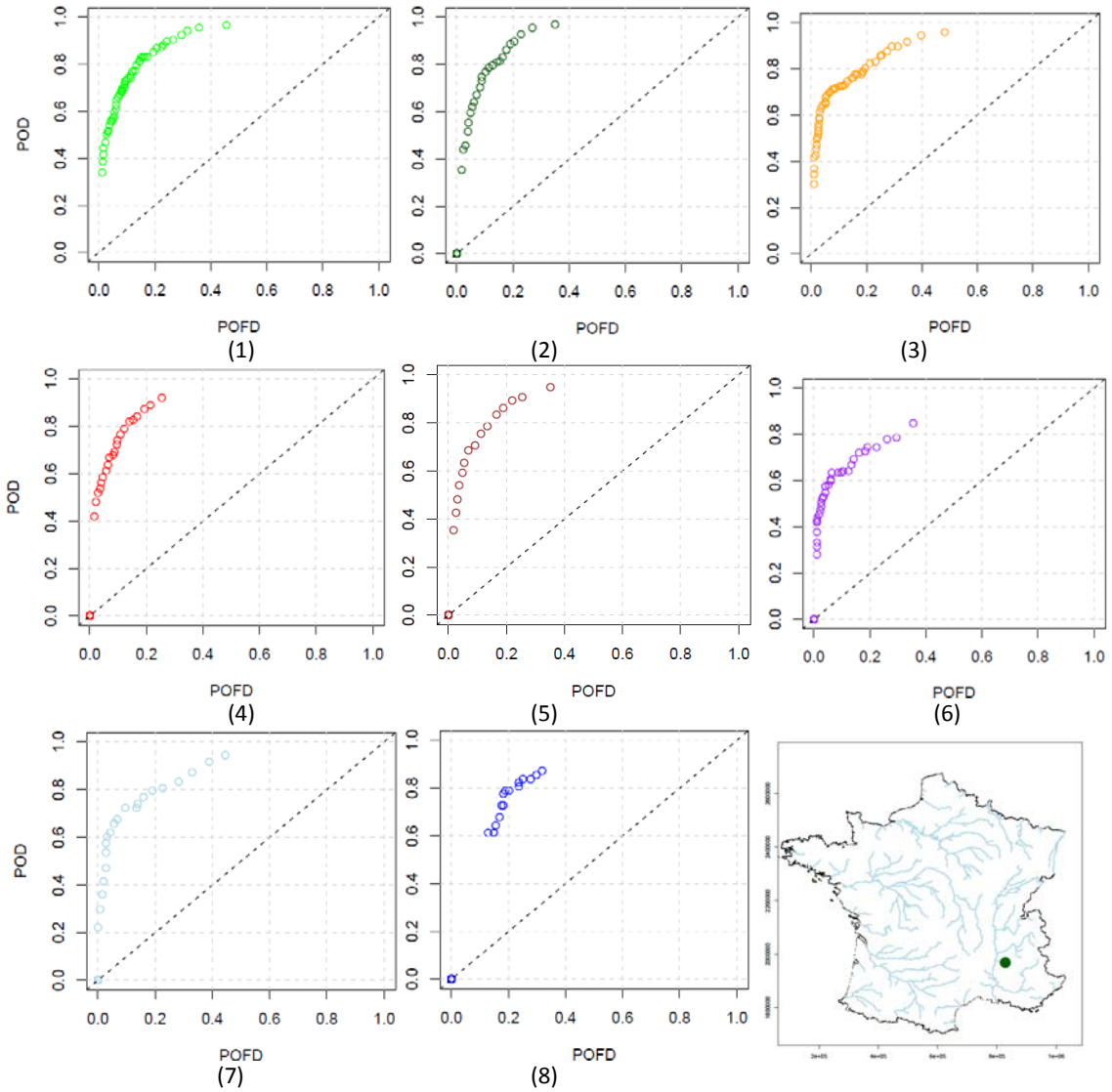


Figure 44 : Les courbes ROC pour les dé bits du bassin de la Drôme à Saillant pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

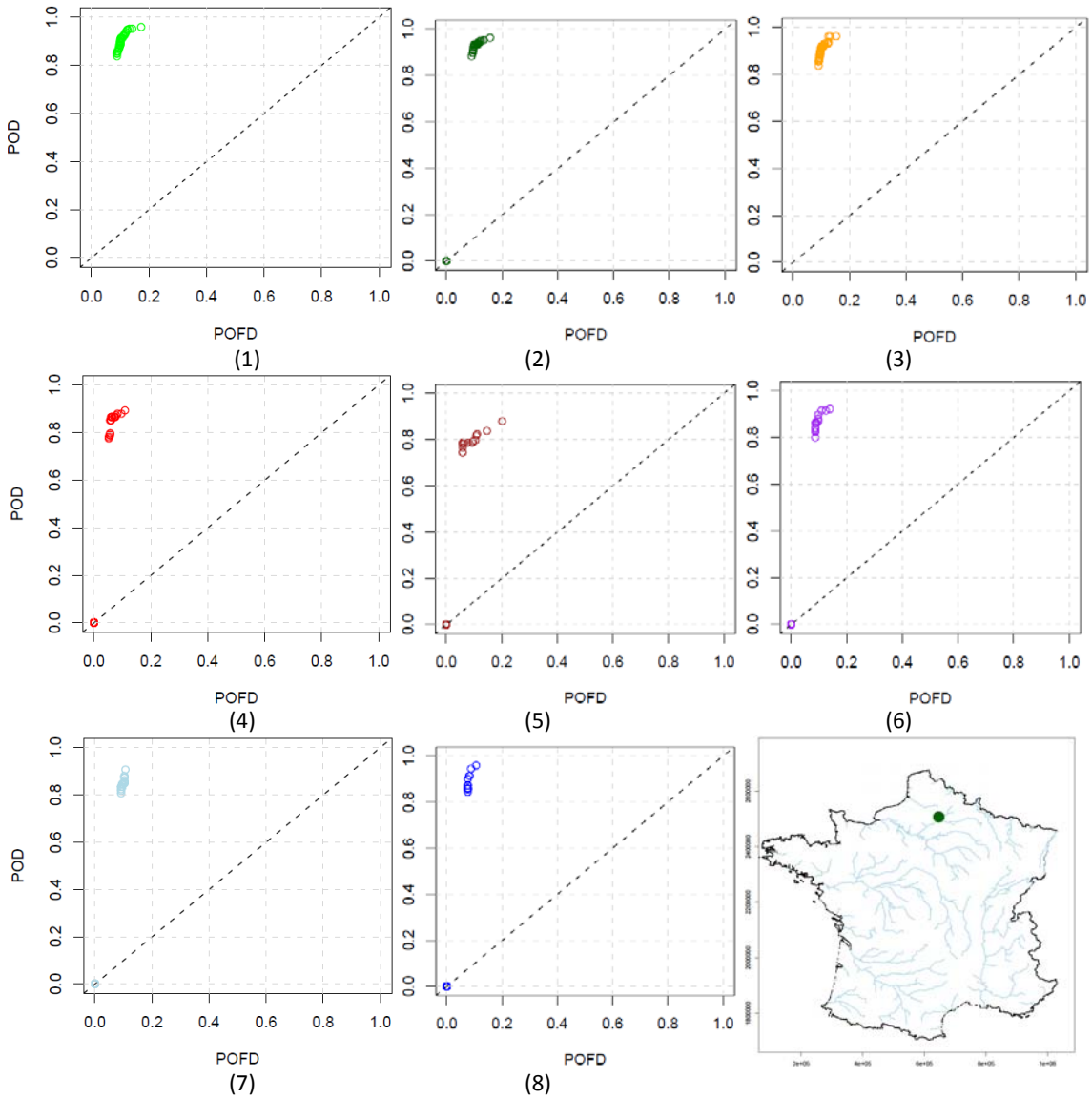


Figure 45 : Les courbes ROC pour les débits du bassin de l'Oise à Sempigny pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempigny sur la carte de France est également présentée.

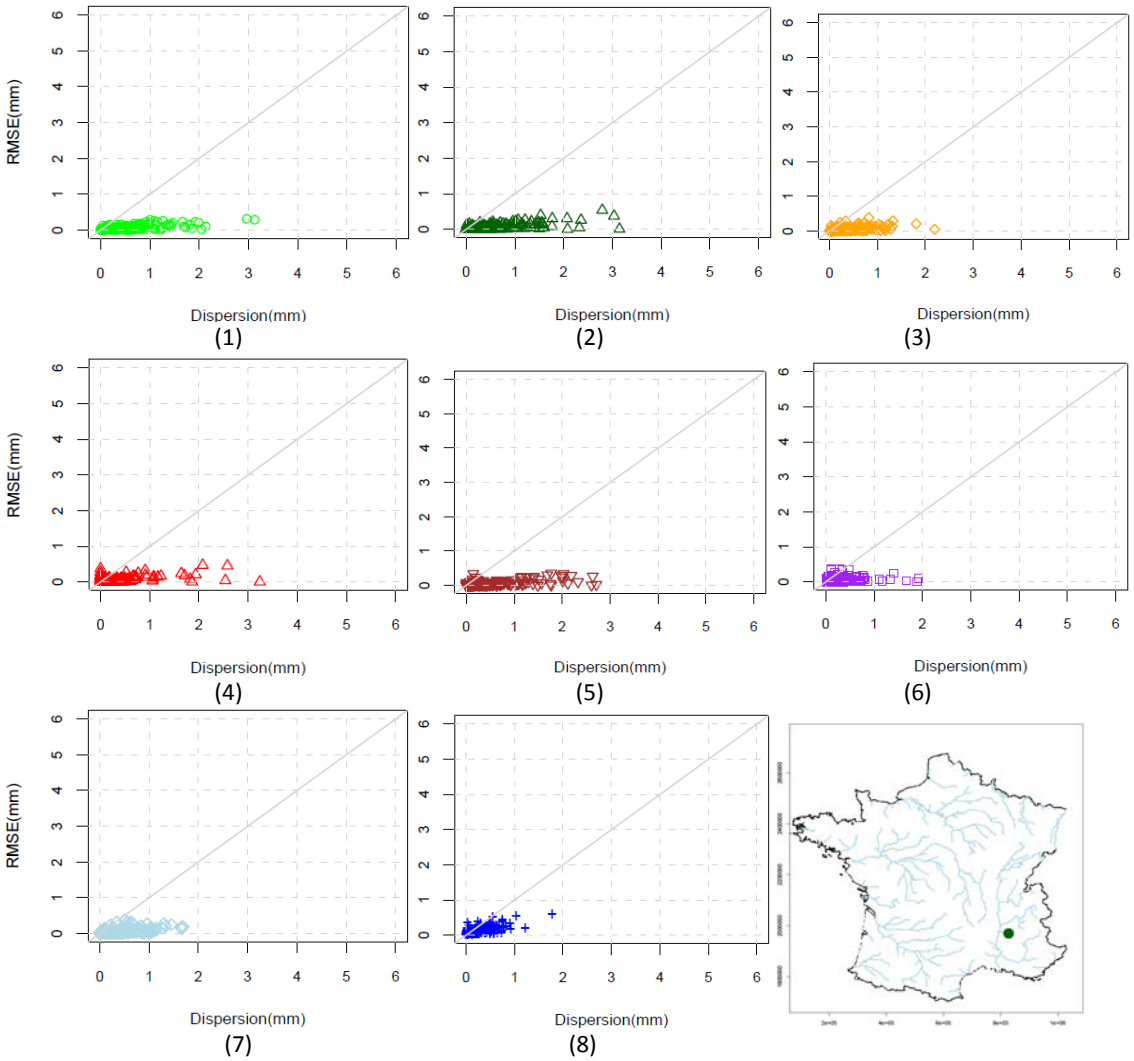


Figure 46 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion des prévisions d'ensemble de débit pour le bassin de la Drôme à Saillant pour l'échéance 7 jours. Les prévisions de pluies sont issues des centres : (1) européen ECMF, (2) britannique EGRR, (3) japonais RJTD, (4) américain KWBC, (5) chinois BABJ, (6) australien AMMC, (7) canadien CWAO et (8) brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée.

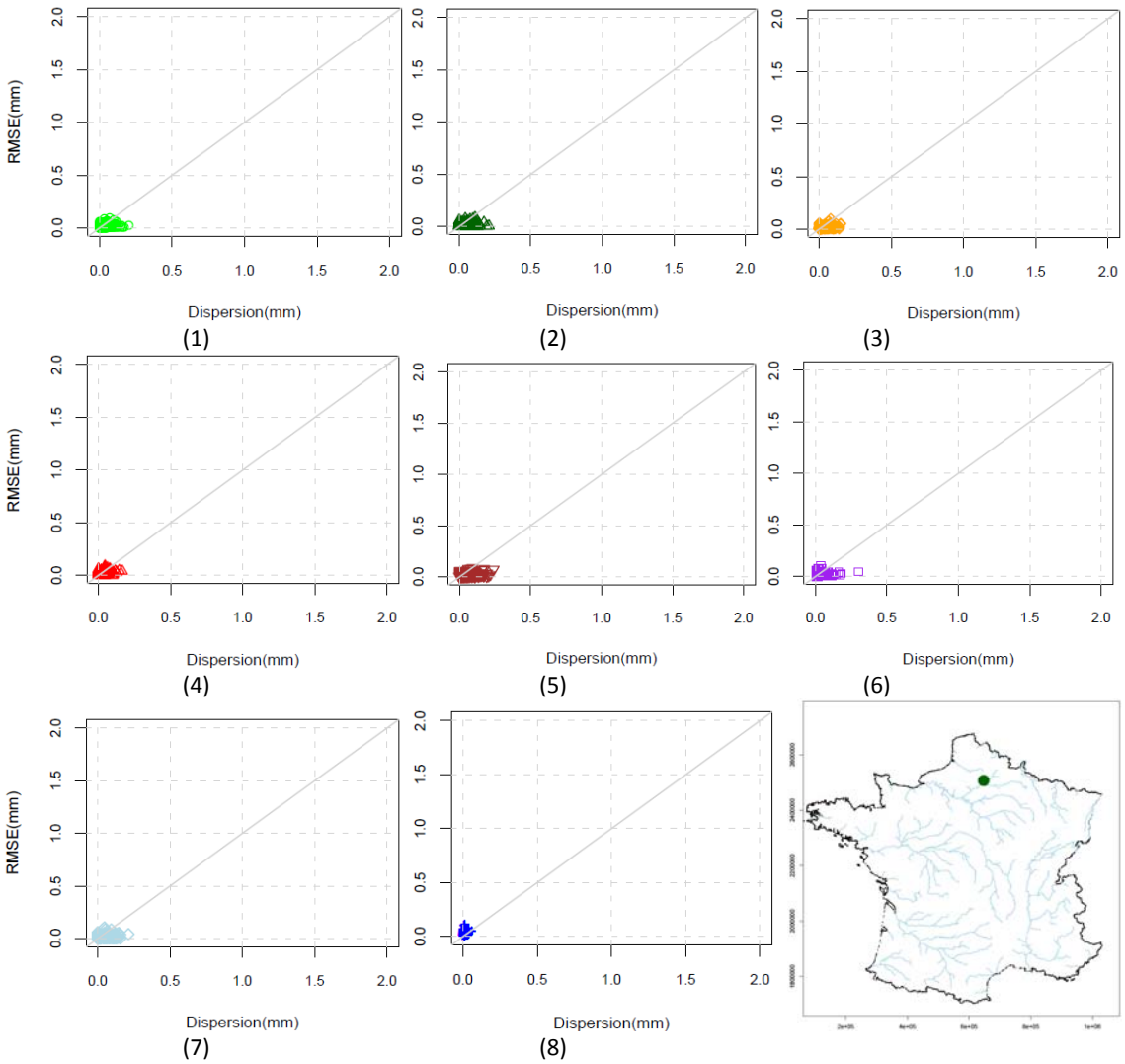


Figure 47 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion des prévisions d'ensemble de débit pour le bassin de l'Oise à Sempigny pour l'échéance 7 jours. Les prévisions de pluie sont issues des centres : (1) européen ECMF, (2) britannique EGRR, (3) japonais RJTD, (4) américain KWBC, (5) chinois BABJ, (6) australien AMMC, (7) canadien CWAO et (8) brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée.

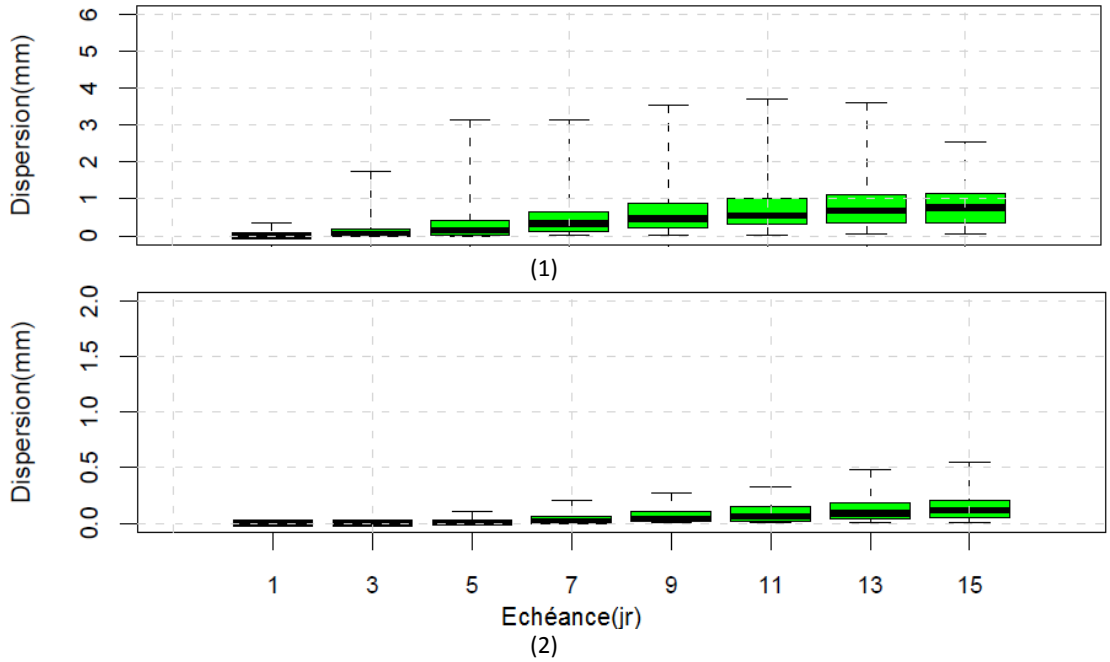


Figure 48 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision sous forme de boîtes à moustache pour le bassin de la Drôme à Saillant (1) et de l'Oise à Sempigny (2). Les prévisions de pluie sont issues du centre européen ECMF. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre.

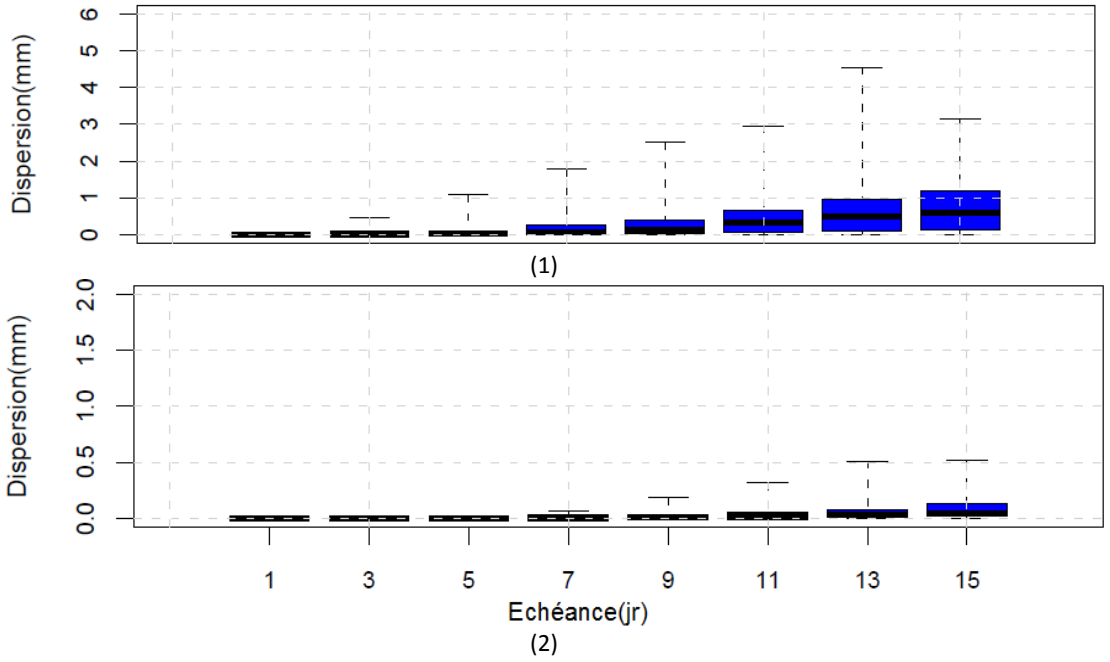


Figure 49 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision sous forme de boîtes à moustache pour le bassin de la Drôme à Saillant (1) et de l'Oise à Sempigny(2). Les prévisions d'ensemble de pluie sont issues du centre brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre.

5.2.2.3 Critères numériques sur l'échantillon de 74 bassins versants

Ci-après nous examinons la performance globale des prévisions de débits provenant du modèle *GRP* alimenté par les prévisions de pluie des différents centres de la base TIGGE en considérant l'ensemble des 74 bassins versants de l'étude.

Différents scores et représentations sont examinés pour avoir une information la plus complète possible sur la qualité des prévisions. Nous examinons le score RPS (Figure 50,Figure 52 et Figure 53), le score de divergence RDS (Figure 51) et le score de Brier BSQ90 pour les dépassements du quantile 90% (Figure 54) des débits observés sur chaque bassin versant.

Les graphiques représentent soit la distribution des valeurs du score sur tous les bassins versants, pour chaque centre météorologique étudié, et pour les échéances 3, 7 et 9 jours (Figure 50 et Figure 51, pour le RPS et le RDS, respectivement), soit l'évolution du score selon la surface du bassin pour chaque centre météorologique et les échéances 3 et 7 jours (Figure 52 pour le RPS), soit les cartes de France avec les résultats des deux meilleures centres pour chaque bassin versant, affichés sous forme de diagramme circulaire, pour les échéances 3, 7 et 9 jours (Figure 53 et Figure 54 pour le RPS et le BSQ90, respectivement).

De l'ensemble des résultats, nous dégageons les conclusions listées ci-dessous. Nous constatons que :

- le centre européen *ECMF* et le centre japonais *RJTD* montrent les meilleurs scores en terme de RPS pour tous les bassins versants étudiés, alors que le centre britannique *EGRR*, le centre *AMMC* et le centre *SBSJ* se montrent moins performants. La qualité des prévisions se dégrade en fonction de l'échéance, toutefois les performances des différents centres restent comparables entre elles pour toutes les échéances examinées. En examinant le score de divergence RDS (Figure 51) le centre australien *AMMC* ressort comme le centre le plus performant parmi les centres examinés. Tandis que les différences de performances des centres restent limitées pour les premières échéances de prévision, aux horizons plus lointains, elles augmentent et deviennent plus marquantes.

- de plus, nous constatons également que les performances de chaque centre sont plus robustes par rapport à l'ensemble des bassins versants examinées pour les horizons de prévision plus courts : les intervalles entre quantiles de la distribution représentée par les boîtes à moustaches sont plus petits que pour les échéances plus lointaines (Figure 50 et Figure 51).

- comme pour les pluies, nous remarquons qu'il n'y a qu'une relation limitée entre la valeur du score RPS et la surface du bassin versant (Figure 52). Cette relation se limite aux bassins versants de surface inférieure à 2000 km². Au-delà, la performance est presque indépendante de la surface du bassin versant.

- considérant les centres météorologiques les plus performants selon le score RPS (Figure 53), nous remarquons que pour un horizon de prévision de 3 jours il n'y a pas un centre particulier pour lequel les prévisions hydrologiques d'ensemble se montrent considérablement meilleures que les autres. Par contre, au-delà d'un horizon de 5 jours le centre européen *ECMF* est le plus performant pour la plupart des bassins versants, suivi par le centre canadien *CWAO*. Nous constatons également que pour les bassins versants pour lesquels ces deux centres sont les plus performants, le deuxième meilleur centre se montre significativement moins performant. De plus, la performance du centre européen *ECMF* s'améliore considérablement aux horizons plus lointains de prévision et le nombre de bassins où ce centre obtient le meilleur score augmente également. Nous notons que les prévisions hydrologiques d'ensemble basées sur les prévisions de pluie du centre canadien *CWAO*, accompagnées par celles du centre japonais *RJTD*, apportent le meilleur score pour toutes les échéances de prévision sur certains bassins versants. C'est le cas des bassins de la Villaine, de la Sarthe, de la Moselle et du Doubs - plutôt caractérisés par des écoulements périodiques concentrés pendant les mois d'hiver/ automne et avec des écoulements souterrains assez limités. Dans le cas de ces bassins versants, nous pouvons confirmer la performance du centre canadien *CWAO* en examinant en parallèle les Figure 52 et Figure 53. Nous remarquons également que les prévisions de pluie tirées du centre canadien *CWAO* ont démontré des performances comparables dans les bassins de la Moselle et du Doubs (Figure 39 et Figure 40).

- En ce qui concerne le critère de Brier *BS* pour un seuil de dépassement du quantile 90% (Figure 54), nous remarquons qu'aucun centre ne se montre considérablement meilleur que les autres sur une majorité de bassins versants. En effet, les meilleures performances sont distribuées entre les centres et varient selon le bassin versant. Ceci est notamment observé pour les échéances plus courtes (3 et 5 jours). Pour les longues échéances, nous observons une apparition plus fréquente des centres européen *ECMF*, japonais *RJTD* et australien *AMMC* parmi les deux centres les plus performantes sur l'ensemble des bassins versants.

La meilleure performance générale du centre européen (ECMF) sur notre échantillon de bassins versants est confirmée par les résultats affichés dans les Tableau 6 et Tableau 7, qui résument les pourcentages des bassins versants où chaque centre est le premier ou le

deuxième meilleur centre en performance pour les scores *RPS* et *BSQ90*, respectivement, et pour les échéances de 3, 5, 7 et 9 jours. Nous remarquons que les pourcentages les plus élevés sont associés aux prévisions du centre européen notamment pour les échéances plus lointaines. Le centre japonais (*RJTD*) affiche également un bon pourcentage sur notre échantillon pour les échéances de 5 à 9 jours, tandis que pour les échéances plus courtes c'est le centre chinois (*BABJ*) qui ressort le plus.

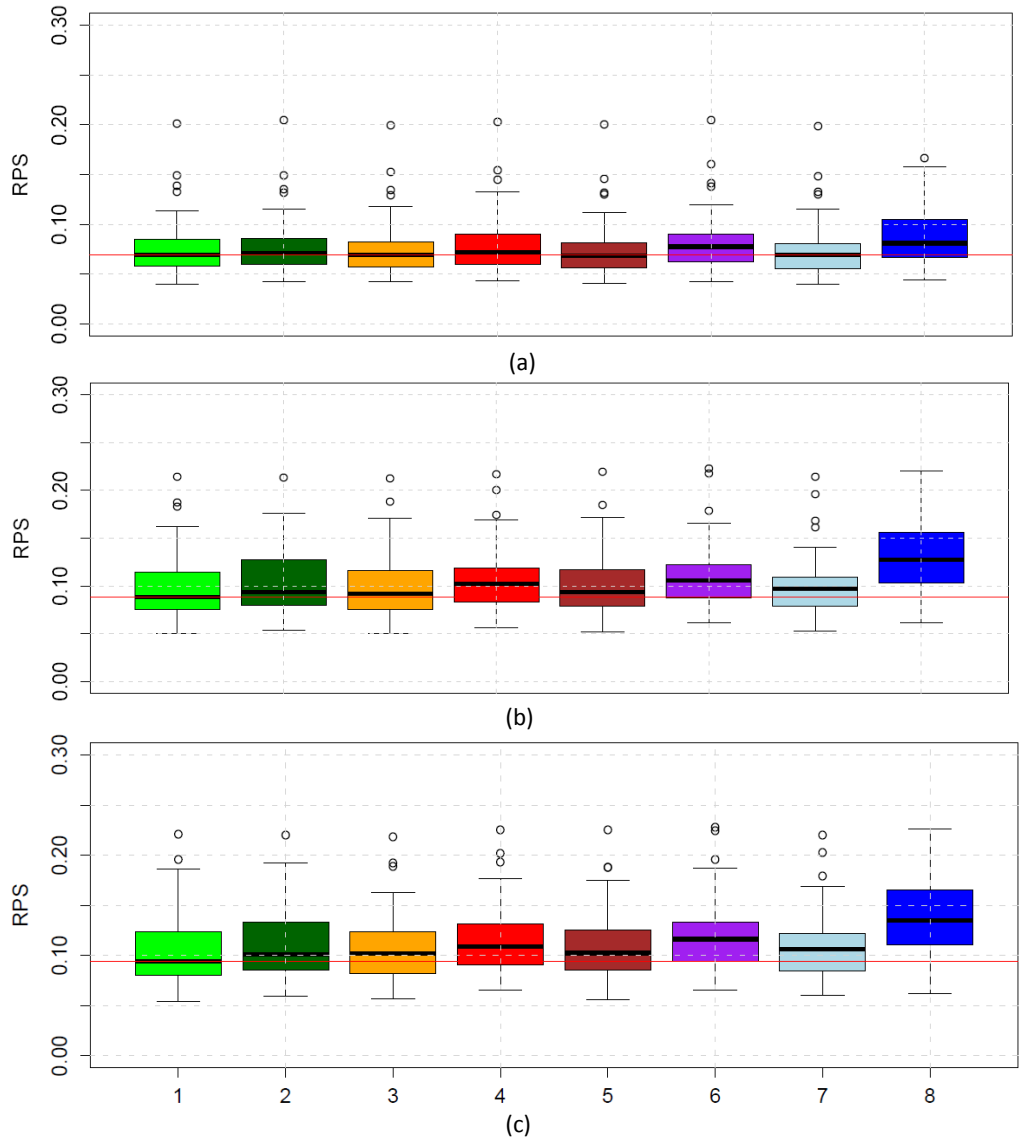


Figure 50 : Boîtes à moustache pour le score de probabilité ordonnée RPS sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions de pluie sont issues des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ.

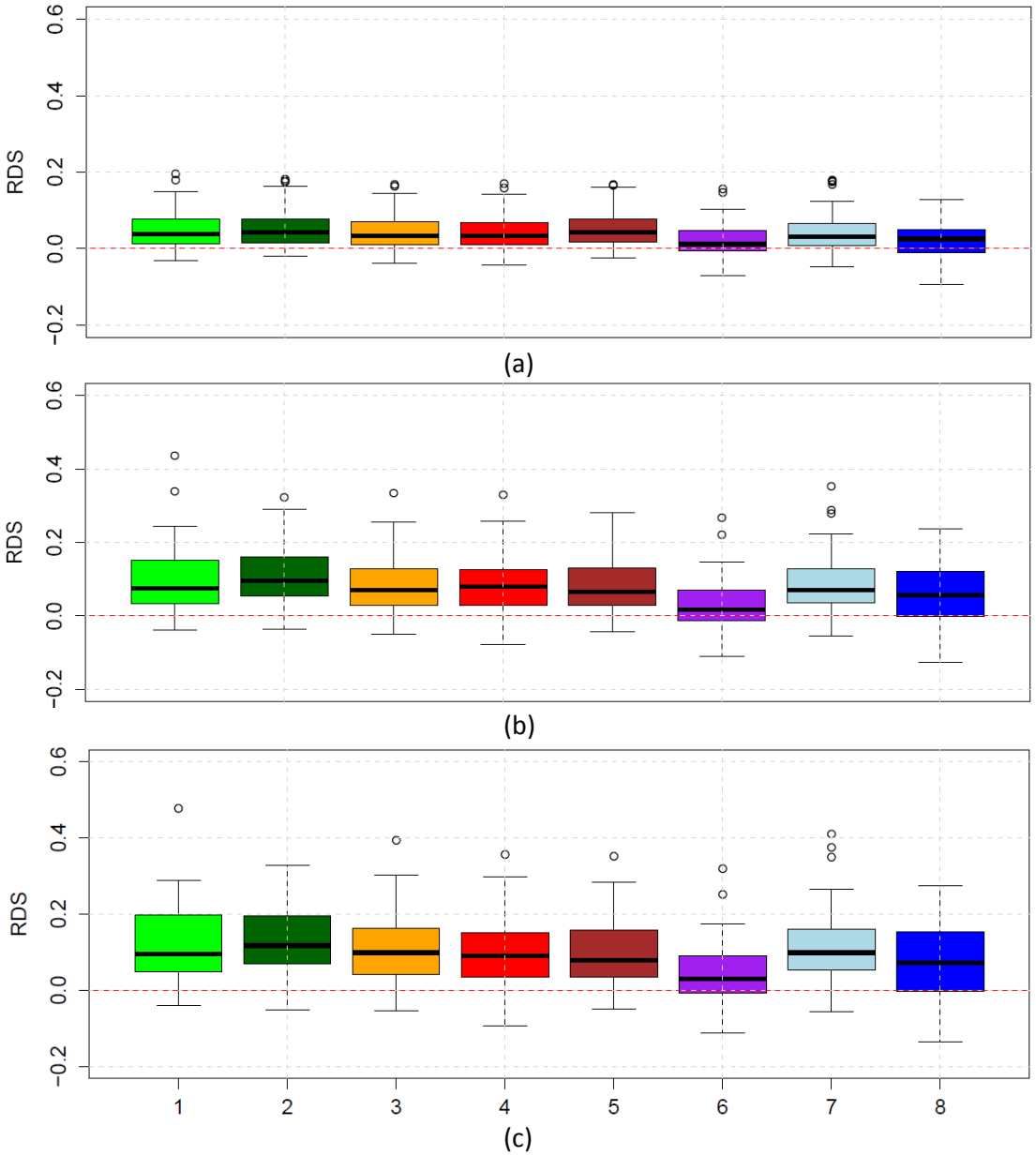


Figure 51 : Boîtes à moustache pour le score de divergence sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions de pluie sont issues des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ.

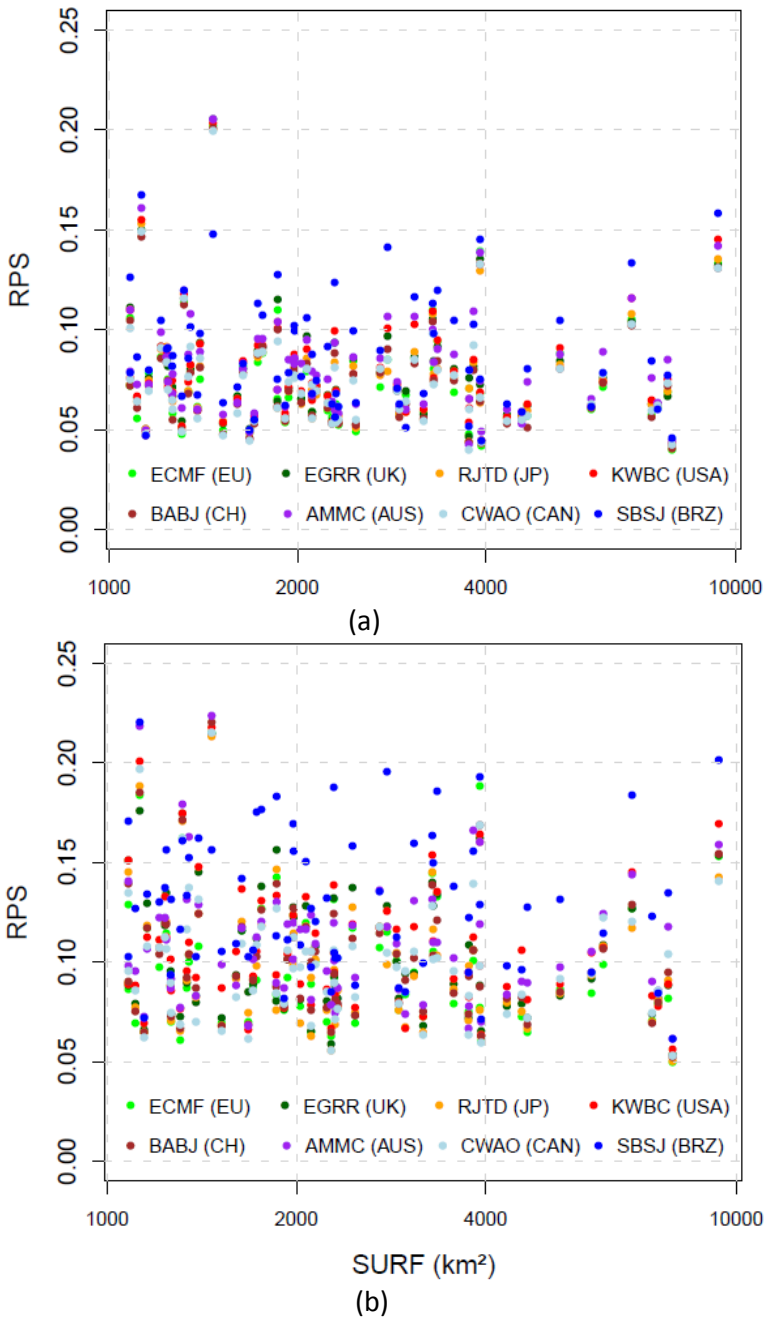
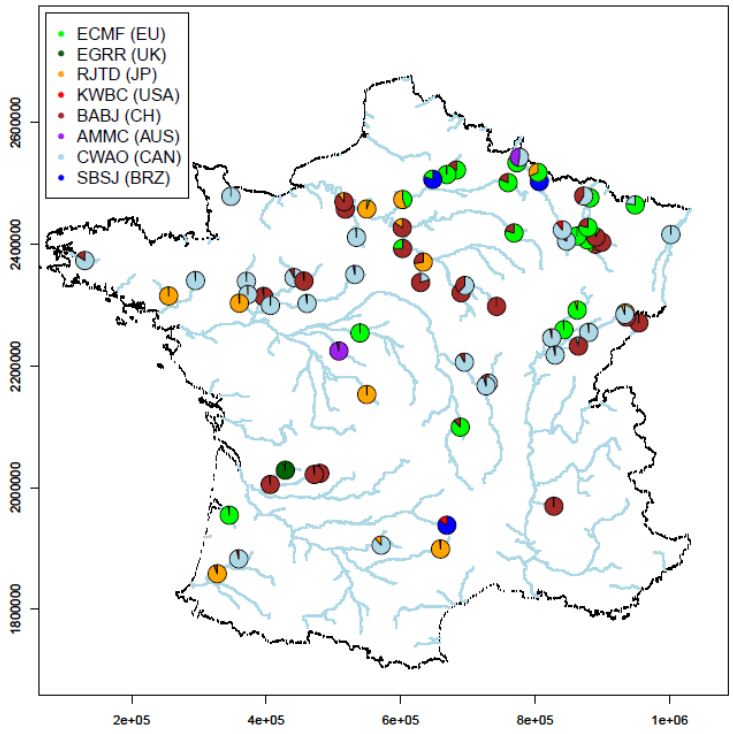
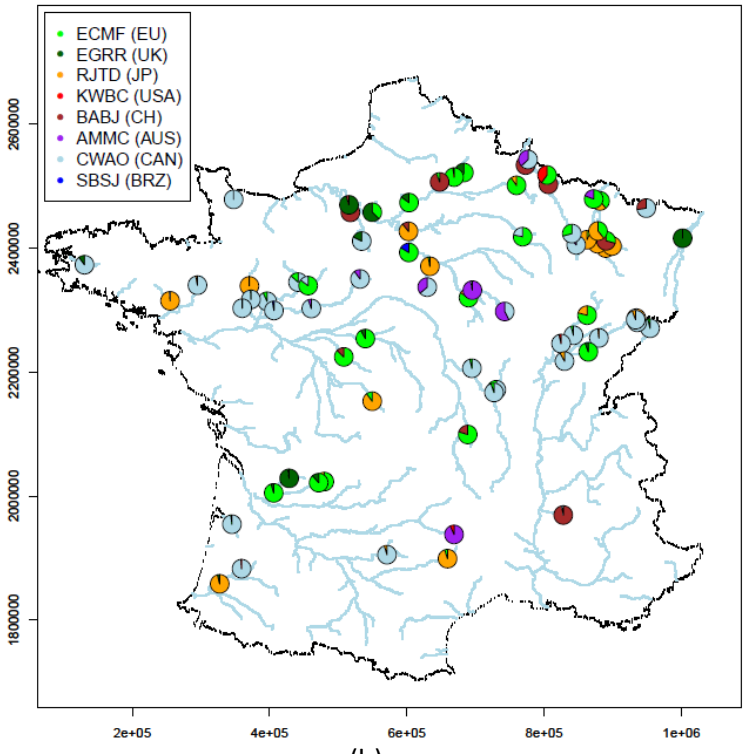


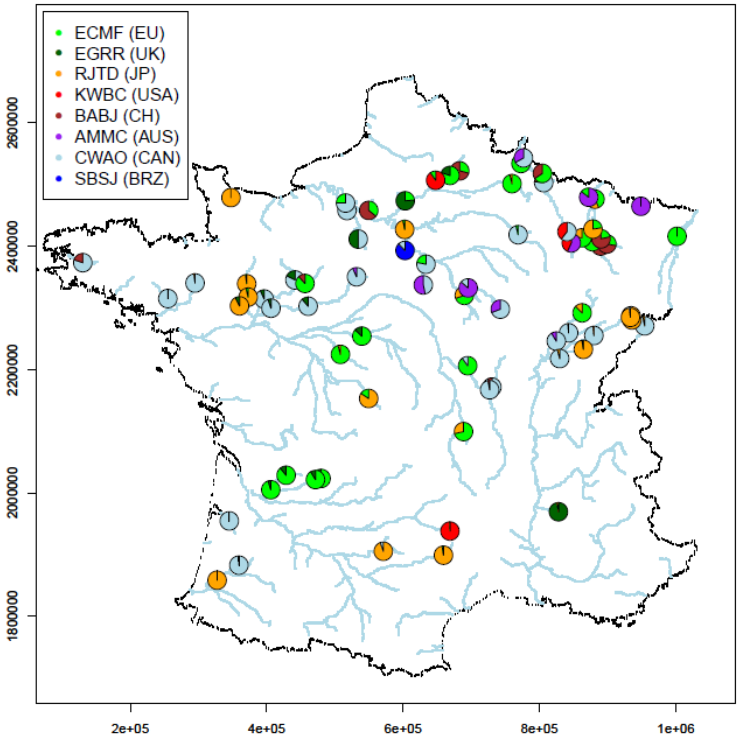
Figure 52 : Graphique du score de probabilité ordonnée RPS, des prévisions d'ensemble de débit en fonction de la surface du bassin versant pour les 74 bassins versants étudiés et pour les échéances 3 (a) et 7 (b) jours. Les prévisions de pluie sont issues des centres : ECMF (vert clair), EGRR (vert foncé), RJTD (orange), KWBC (rouge), BABJ (marron), AMMC (mauve), CWAO(ciel) et SBSJ (bleu).


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

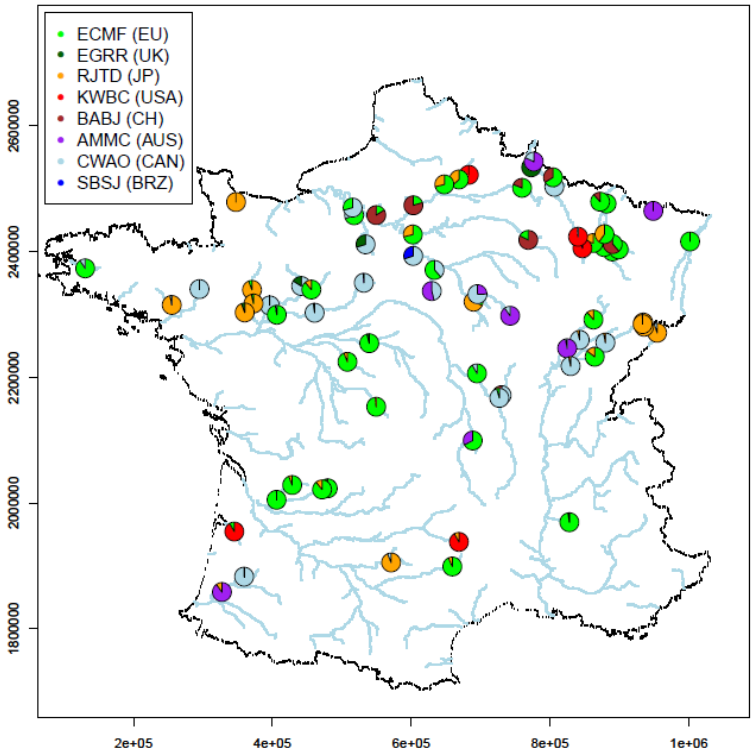
(a)


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

(b)

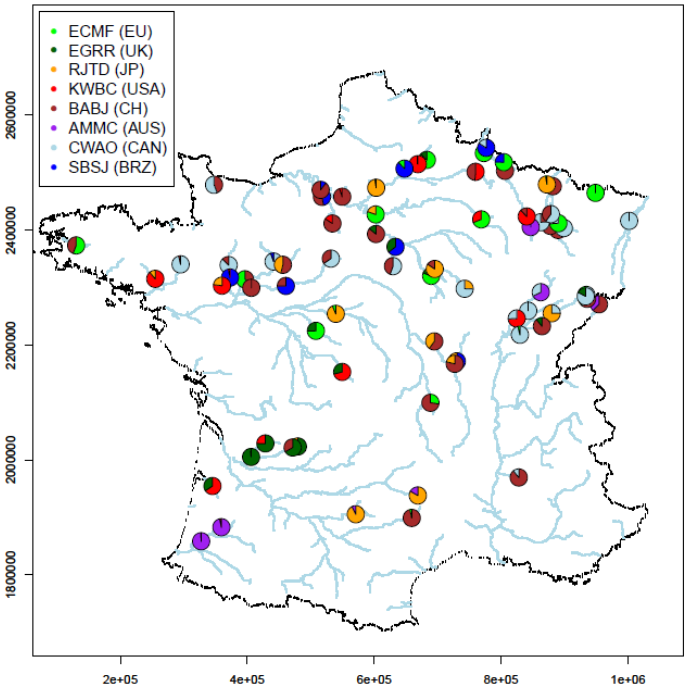


(c)

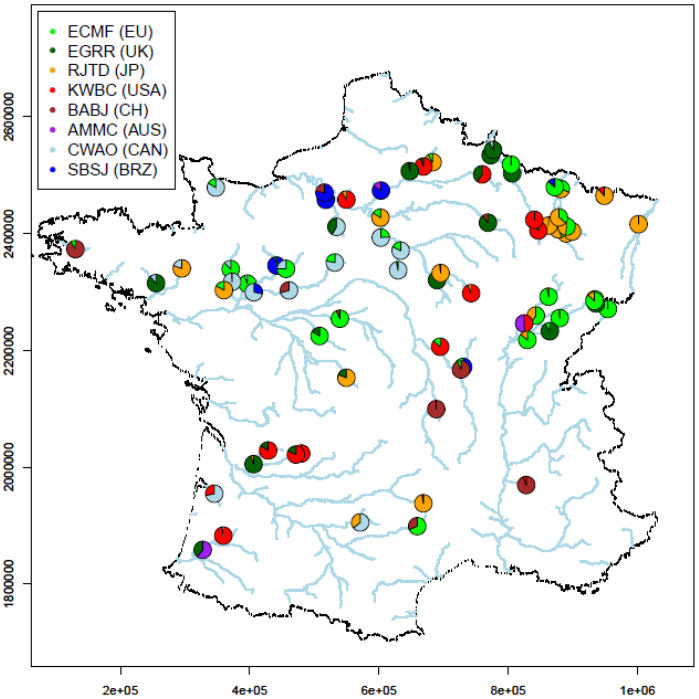


(d)

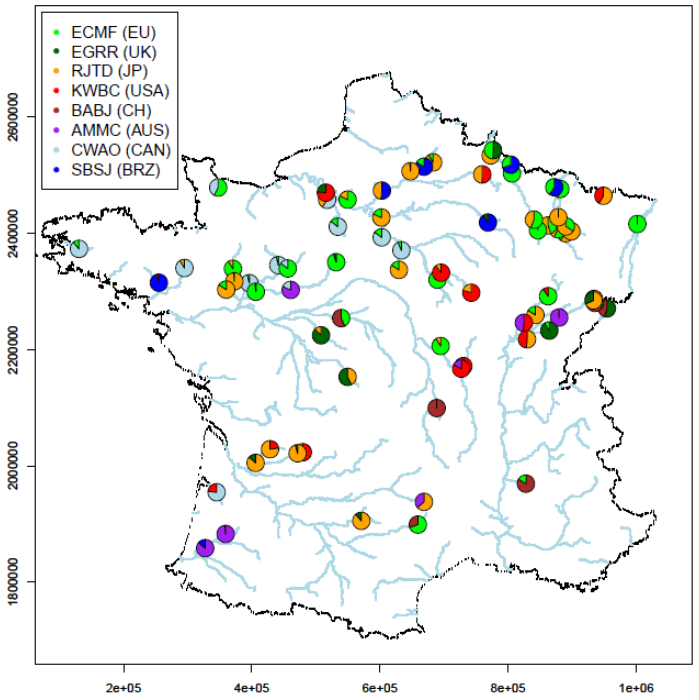
Figure 53 : Carte de localisation des centres qui donnent la meilleure performance selon le critère RPS pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d).


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

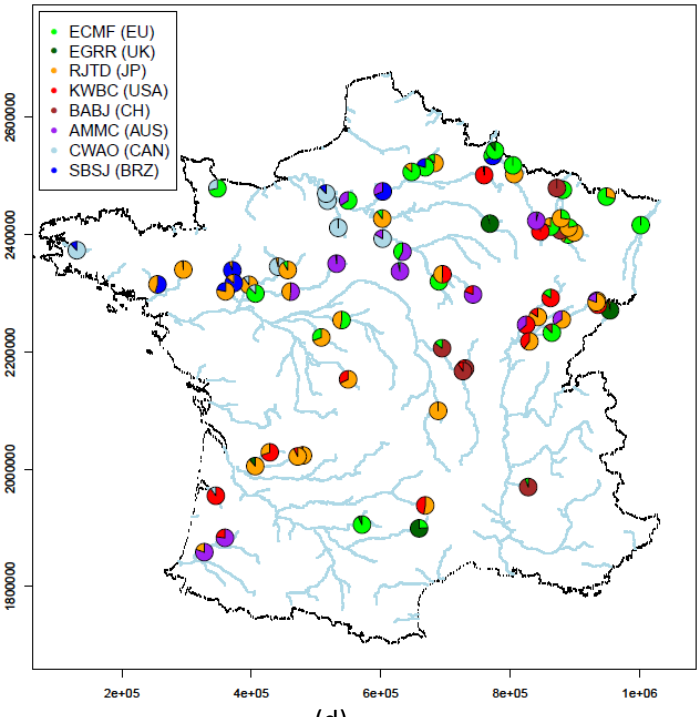
(a)


ECMF (EU)
EGRR (UK)
RJTD (JP)
KWBC (USA)
BABJ (CH)
AMMC (AUS)
CWAO (CAN)
SBSJ (BRZ)
2600000
2400000
2200000
2000000
1800000
2e+05
4e+05
6e+05
8e+05
1e+06

(b)



(c)



(d)

Figure 54 : Carte de localisation des centres qui donnent la meilleure performance selon le score de Brier pour le non-dépassement du quantile 90% pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d).

Tableau 6 : Pourcentage de bassins versants pour lesquels chaque centre démontre les meilleures (1ère et 2ème) performances selon le critère RPS par échéance de prévision.

	3 jrs	5 jrs	7 jrs	9 jrs
ECMF	32%	53%	46%	54%
EGRR	8%	19%	18%	5%
RJTD	20%	28%	28%	35%
KWBC	6%	3%	5%	7%
BABJ	46%	15%	12%	11%
AMMC	3%	12%	12%	11%
CWAO	41%	46%	42%	31%
SBSJ	4%	1%	1%	1%

Tableau 7 : Pourcentage de bassins versants pour lesquels chaque centre demontre les meilleures (1ère et 2ème) performances selon le critère BSQ 90 par échéance de prévision.

	3 jrs	5 jrs	7 jrs	9 jrs
ECMF	23%	51%	51%	41%
EGRR	18%	24%	12%	8%
RJTD	20%	30%	46%	47%
KWBC	15%	18%	18%	24%
BABJ	46%	8%	9%	9%
AMMC	11%	3%	10%	20%
CWAO	5%	16%	16%	14%
SBSJ	14%	14%	9%	9%

5.3 Combinaison de centres

Jusqu'ici nous avons traité chaque centre séparément et nous avons réalisé que même si certains centres obtiennent de meilleurs scores que d'autres, ils apportent tous une information supplémentaire sur l'échantillon de 74 bassins versants examinés. Il devient alors intéressant d'examiner si la combinaison des prévisions provenant des différents centres arrive à améliorer la qualité des prévisions émises.

Nous avos recensé un nombre limité d'études récentes sur l'utilisation d'un grand ensemble composé de différents centres de la base TIGGE (Pappenberger *et al.*, 2008 ; He *et al.*, 2009 ; Bao *et al.*, 2011 ; Keller *et al.*, 2011 ; Hagedorn *et al.*, 2012 ; Hamill, 2012 ; Yamaguchi *et al.*, 2012). Cependant peu d'études se concentrent sur une combinaison pondérée des différents centres (Johnson et Swinbank, 2009 ; Yang *et al.*, 2012 ; Zhi *et al.*, 2012). De plus, pratiquement toutes ces études jusqu'à present se concentrent sur les prévisions de pluie ou/et de température de la base TIGGE. A notre connaissance la combinaison de différentes prévisions de debits n'est pas encore abordée. Dans cette partie de la thèse nous essayons d'aborder ce sujet.

5.3.1 Approches étudiées

Avant de décrire en détails les approches de combinaison de centres que nous avons appliquées, nous présentons brièvement les méthodes de combinaison de centres déjà utilisées dans la littérature pour les prévisions de pluie.

- la pondération bayésienne de modèles (en anglais "bayesian model averaging") ou BMA.

Yang *et al.*, (2012) ont appliqué cette approche sur les prévisions des précipitations du mois de juillet 2007 dans le bassin versant de Yishusi en Chine. Selon les auteurs, le grand ensemble pondéré a démontré de meilleures performances - en termes de scores de performance - que les différents centres individuels, ainsi que le grand ensemble équiprobable. Cette amélioration est également évidente sur les précipitations extrêmes.

Dans l'approche de *BMA*, nous notons que la probabilité *a priori* attribuée *p(x)* à un événement *x* est estimée par la somme pondérée de la probabilité prédictive conditionnelle de la valeur finale de l'événement *x*. Le poids *wk* attribué à chaque centre *k* est égal à la probabilité *a posteriori* $p(x|y)$, une fois la valeur réele *y* de l'événement est connue, calculée en période de calage. Quand le poids *wk* correspondant au centre *k* est petit, cela nous permet (dans le cas où le temps de calcul est limité) d'identifier les centres qui apportent

peu à l'ensemble et les éliminer. L'estimation de poids *wk* est faite en utilisant un algorithme d'espérance-maximisation [EM], qui permet de trouver le maximum de vraisemblance des paramètres (ici poids *wk*) en utilisant des données de calage disponibles.

- la régression linéaire multiple (Krishnamurti et al., 1999 ; Kharin et Zwiers, 2002 ; Doblas-Reyes et al., 2005).

Il s'agit d'une généralisation à plusieurs variables de la régression linéaire simple. En ce qui concerne la combinaison de différents modèles, cette approche permet le calcul d'une moyenne pondérée des valeurs des différents modèles. Plusieurs auteurs trouvent néanmoins que l'extension de cette approche à la pondération de probabilités – comme c'est le cas pour les prévisions probabilistes - n'est pas toujours évidente (Stefanova et Krishnamurti, 2002 ; Pena et van den Dool, 2008 ; Johnson et Swinbank, 2009).

Ce type de méthode est souvent appliquée lors de la combinaison des différentes prévisions, issues de divers modèles. Dans ce cas de prévisions multi-modèles, la méthode peut être assimilée à une approche d'habillage empirique de chaque centre individuellement, à partir de la distribution de probabilité de l'erreur préalablement calée. En considérant chaque centre comme indépendant nous pouvons lui assigner un poids, qui est calculé à partir d'une régression linéaire.

La théorie de régression propose la construction des probabilités prédictives et des densités de probabilité prédictives à partir d'une loi de distribution gaussienne (Glahn et Lowry, 1992 ; Stefanova et Krishnamurti, 2002). L'estimation des deux moments de cette distribution (moyenne et variance) dans le cas d'une régression (multi-)linéaire est faite en utilisant soit la méthode des moments, soit la méthode de maximum de vraisemblance.

Comme pour l'approche BMA, une longue série de données est requise pour le calage des paramètres de la méthode de combinaison (Wilson et Vallée, 2002). Sans les données nécessaires pour le bon calage des paramètres, les performances de ces approches ne diffèrent pas considérablement de celles d'une méthode plus simple de pondération (Johnson et Swinbank, 2009).

- la pondération de centres selon leur performance pour un critère d'evaluation donné (Woodcock et Engel, 2005).

Cette approche calcule les poids attribués à chaque centre en fonction de sa performance selon un critère d'évaluation. Les études existantes ont utilisé l'erreur quadratique moyenne *RMSE* ou le score de Brier (Woodcock et Engel, 2005 ; Johnson et Swinbank, 2009).

De plus, Johnson et Swinbank (2009) ont montré dans leurs cas que les valeurs de poids obtenués sont similaires à celles calculées par la régression linéaire multiple ou la pondération bayésienne de modèles. Parmi les avantages de cette approche se trouve sa simplicité, ainsi que le fait qu'elle s'adapte à n'importe quelle longueur de données.

Dans leur étude Johnson et Swinbank (2009) ont appliqué cette approche sur des prévisions de températures tirées des centres européen, britannique et américain sur la période du 23 août 2007 au 29 avril 2008. Ils ont montré que le grand ensemble est plus performant que chaque centre individuel. Ils ont obtenu des résultats similaires en considerant des prévisions débiaisées de chaque centre individuel. De plus, les auteurs ont conclu que le grand ensemble offre une prévisibilité améliorée pour 2 jours – c'est-à-dire, les prévisions sont plus fiables jusqu'à un horizon de 2 jours supplémentaires.

Dans notre étude nous avons essayé de combiner les prévisions de différents centres de façon à élargir la distribution prédictive en prenant en compte les membres des ensembles individuels. La pondération des differents centres est basée sur leur performance en termes de score de probabilité ordonnée *RPS*. Afin que la comparaison soit faite dans les mêmes termes pour tous les centres, la période d'évaluation examinée est limitée aux 242 jours disponibles pour tous les centres.

5.3.2 Approches appliquées

Nous étudions plus en détails la construction d'un "grand ensemble" constitué des prévisions de l'ensemble des centres disponibles dans la base TIGGE. Deux approches sont considérées :

1. le grand ensemble basé sur l'équiprobabilité des centres et des membres de chaque centre. Selon cette approche, le grand ensemble est formé en regroupant, à chaque jour de prévision, tous les membres de tous les centres disponibles. Au total, nous avons ainsi une prévision d'ensemble avec 232 membres,

2. le grand ensemble basé sur la combinaison des différents centres, pondérés selon leur performance en termes de score de probabilité ordonnée *RPS*. Nous supposons ici qu'aucun centre a un score parfait (*RPS* = 0, au cas où il ne serait pas question de pondérer les centres mais d'adopter directement le "centre parfait"). Les poids w_k attribués à chaque centre *k* s'exprime comme :

$$w_k = \frac{1 - \min(RPS_{1:k})}{RPS_k} \qquad \text{Eq. 49}$$

Pour la construction du grand ensemble, chaque w_k est considéré en tant que facteur multiplicatif du nombre de membres N_k avec lequel chaque centre participe dans le grand ensemble. Nous avons examiné deux options possibles :

(A) l'application des poids w_k à la moyenne d'ensemble de chaque centre : le grand ensemble est ainsi constitué de w_k fois la moyenne d'ensemble de chaque centre k,

(B) l'application des poids *wk* en tant que facteurs multiplicatifs du nombre de membres *Nk* avec lesquels chaque centre participe dans le grand ensemble.

Le calage des poids a suivi le principe "split-sample" (Klemeš, 1986) : la période de 242 jours disponibles a été divisée en deux sous-périodes de calage-validation et inversément. Les résultats en validation ont été ensuite combinés et sont présentés ci-après.

Notons enfin que les deux approches par pondération selon le *RPS* nous permettent de construire la distribution prédictive du grand ensemble similairement à la construction du grand ensemble basé sur l'équiprobabilité. Ces trois approches de combinaison seront comparées entre elles, mais également comparées aux performances de deux centres individuels, à savoir le centre européen (*ECMF*) et le centre brésilien (*SBSJ*), les deux calculées sur la période de 242 jours utilisés pour la construction des trois grands ensembles.

5.3.3 Évaluation des prévisions d'ensemble de pluies du Grand Ensemble

Afin de procéder à l'évaluation de la qualité du grand ensemble, nous nous appuyons à nouveau sur les différents attributs qui qualifient une prévision d'ensemble : la fiabilité, la dispersion des membres, la discrimination et la qualité globale.

Nous examinons d'abord les prévisions de précipitation : la fiabilité des prévisions exprimée par l'histogramme de PIT (Figure 55 et Figure 56), les courbes ROC (Figure 57 et Figure 58), l'évolution de la dispersion des membres en fonction de l'échéance de prévision (Figure 59) et les performances sur l'ensemble de 74 bassins versants selon les critères *RPS* et *RDS*.

Pour les critères graphiques, nous retenons les mêmes bassins versants utilisés à titre illustratif : la Drôme à Saillant et l'Oise à Sempigny. Ces graphiques illustrent les performances à une échéane de prévision de 7 jours des grands ensembles construits : (a) à partir de l'equiprobabilité de membres, (b) avec la pondération des moyennes d'ensemble de différents centres en fonction du score *RPS* et (c) avec la pondération des prévisions de différents centres en fonction du score *RPS*.

L'analyse a montré que :

- les performances du grand ensemble de pluie brut (construit sans pondération, sur le principe d'équiprobabilité) et du grand ensemble pondéré montrent une fiabilité des prévisions de pluie significativement améliorée par rapport à celle des prévisions tirées des centres individuels (Figure 55 et Figure 56). Cette conclusion est valable surtout pour les prévisions du grand ensemble basé sur l'équiprobabilité (a) ou pondéré en fonction du score *RPS* sur l'ensemble des membres de chaque centre (c).
- en termes de discrimination (Figure 57 et Figure 58), les courbes ROC pour un seuil de dépassement égal au quantile 70%, et pour l'échéance de 7 jours, montrent que le dépassement de ce quantile avec une probabilité de bonne détection POD supérieure à 60% est prévu par la majorité de membres (plus de 60% du nombre total de membres) des trois grands ensembles examinés. Nous remarquons également que les performances de ces trois grands ensembles sont similaires à celle du centre européen, avec une aire sous la courbe approximativement équivalente. Cependant, les performances des grands ensembles sont considérablement améliorées par rapport aux performances des autres centres individuels, comme par exemple le centre américain ou le centre brésilien.
- toutefois, en examinant tous les horizons de prévision, nous constatons aussi que le comportement en termes de discrimination est fortement influencé par l'horizon de prévision. Les courbes ROC des prévisions des trois approches de combinaison du grand ensemble ici testées montrent que le nombre de membres qui prévoient le dépassement du quantile 70% avec une probabilité de détection POD de 60% diminue considérablement avec l'échéance. De plus, aux horizons plus lointains de prévision (supérieurs à 10-13 jours), l'aire sous la courbe ROC approche la valeur minimale de 0,5. En ce qui concerne les autres seuils de dépassement, nous constatons que le nombre de membres qui prévoient les dépassements avec une probabilité de bonne détection POD de 60% diminue considérablement quand le seuil varie de 10% à 90%. Ce fait devient plus apparent dans le cas du grand ensemble construit à partir des moyennes d'ensemble pondérées (b).
- Le grand ensemble construit à la base de la pondération des moyennes d'ensemble (b) produit des prévisions peu dispersées. Une explication possible est liée à la dispersion de membres qui est moins forte dans ce cas, étant donné que l'on réduit la variabilité des membres en prenant seulement leur moyenne pour la construction du grand ensemble avec cette approche. Ceci est confirmé par la Figure 59, qui montre en parallèle l'évolution de la dispersion des membres de l'ensemble de pluie en fonction de l'horizon de prévision pour le grand ensemble constitué par la pondération des moyennes et celle du centre européen pris individuellement.

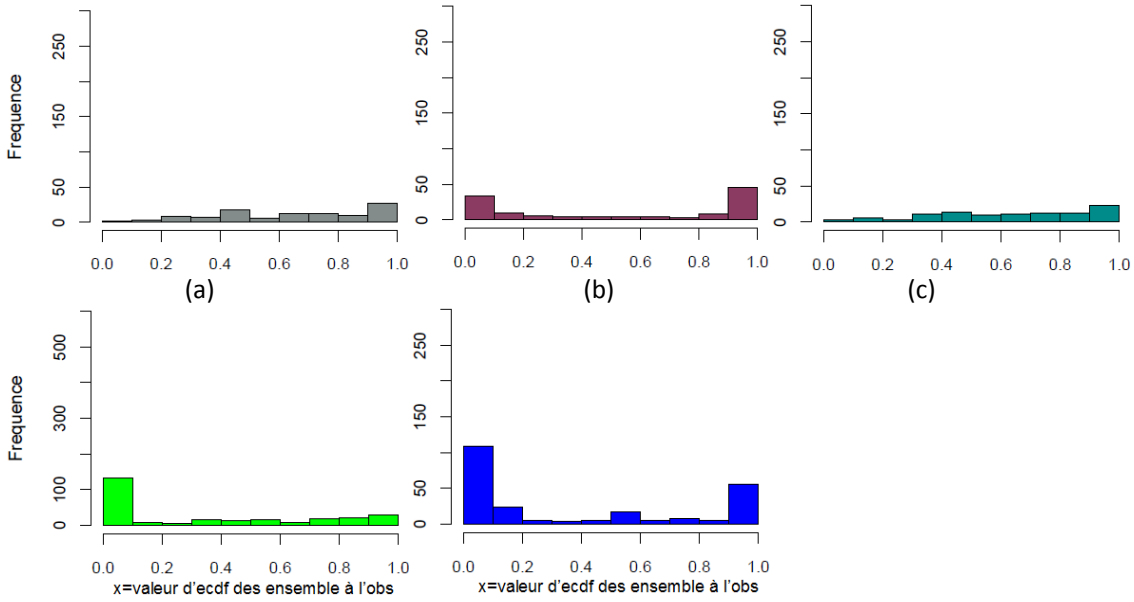


Figure 55 : Histogrammes de PIT pour les prévisions de pluie pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées des centres européen et brésilien évaluées sur les 242 jours sont présentées dans la figure (d) & (e).

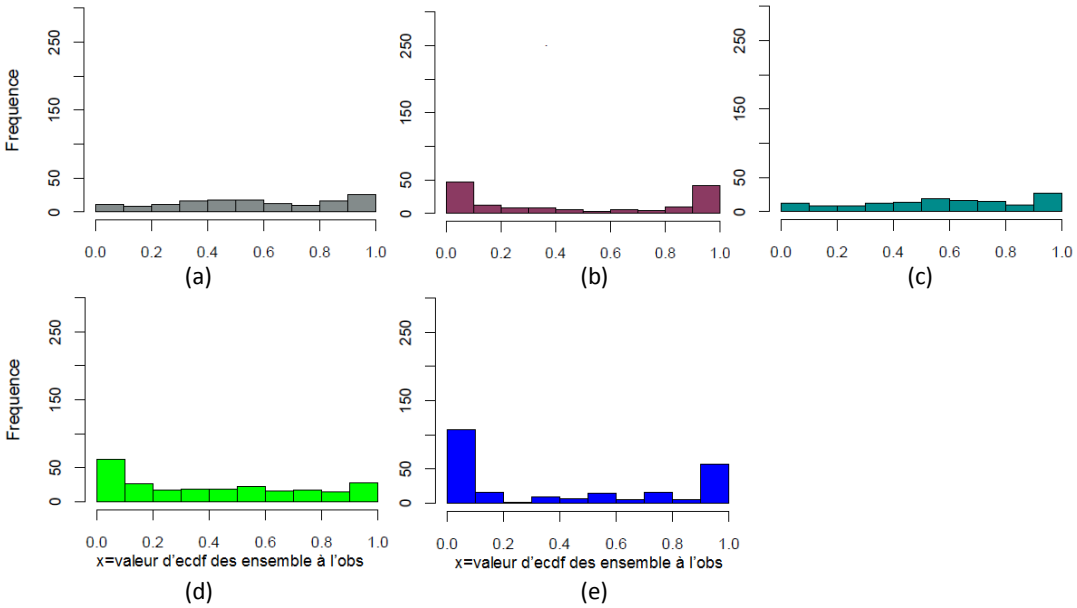


Figure 56 : Histogrammes de PIT pour les prévisions de pluie pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées des centres européen et brésilien évaluées sur les 242 jours sont présentées dans la figure (d) & (e).

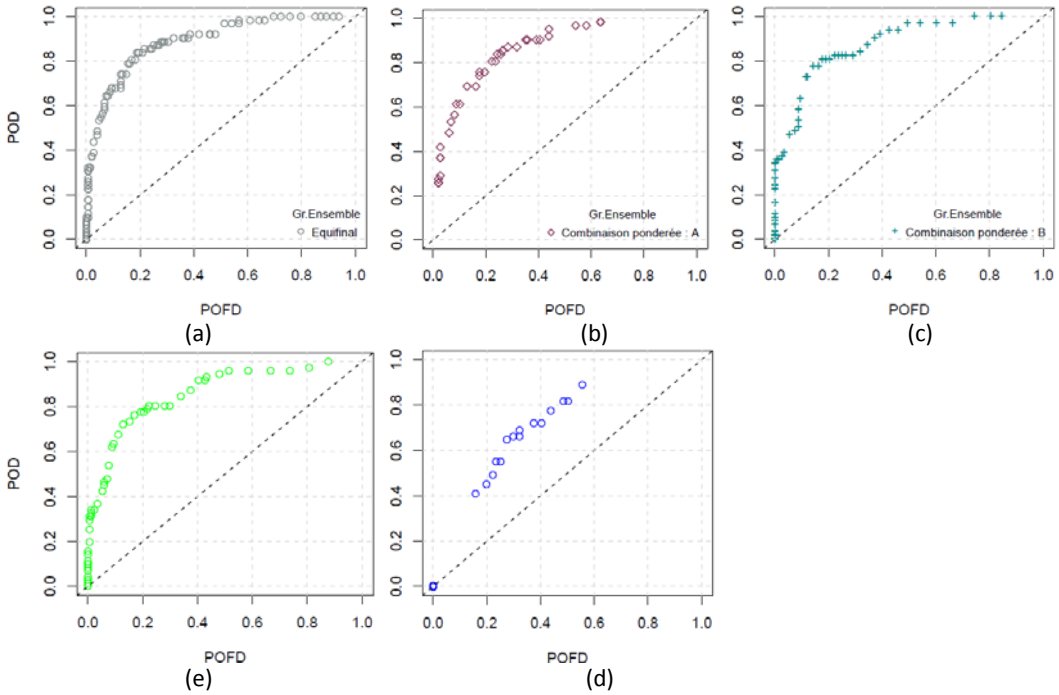


Figure 57 : Courbes ROC pour les prévisions de pluie pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées du grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres selon le score RPS, (c) la pondération en fonction de leur performance en fonction du RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien (d) & (e).

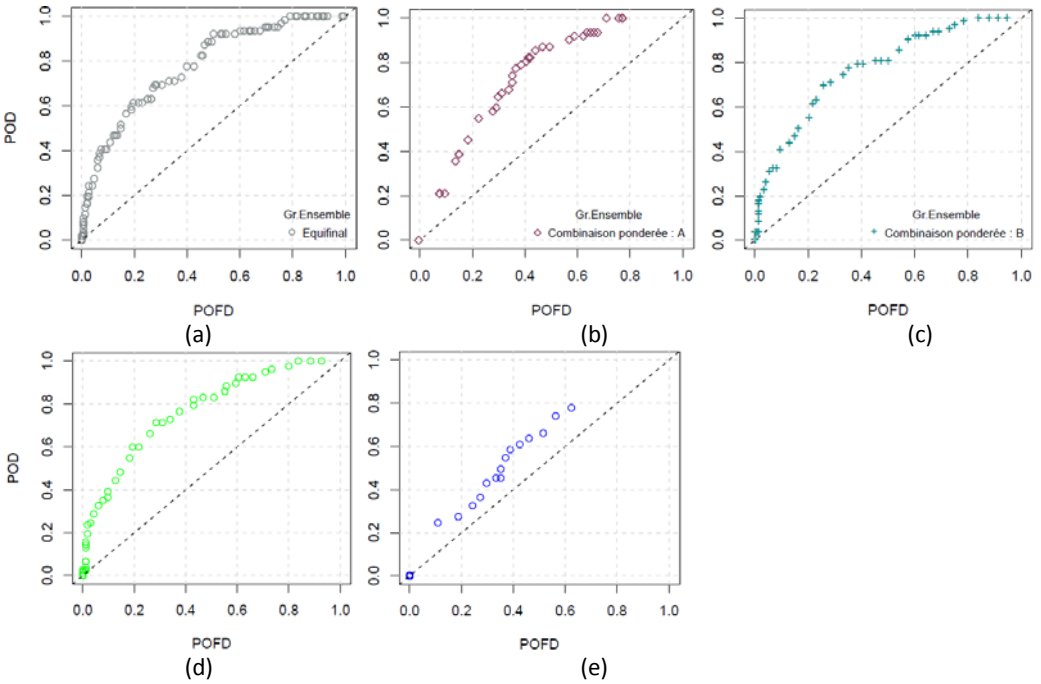


Figure 58 : Courbes ROC pour les prévisions de pluie pour le bassin de l'Oise à Sempigny à un horizon de 7 jours.

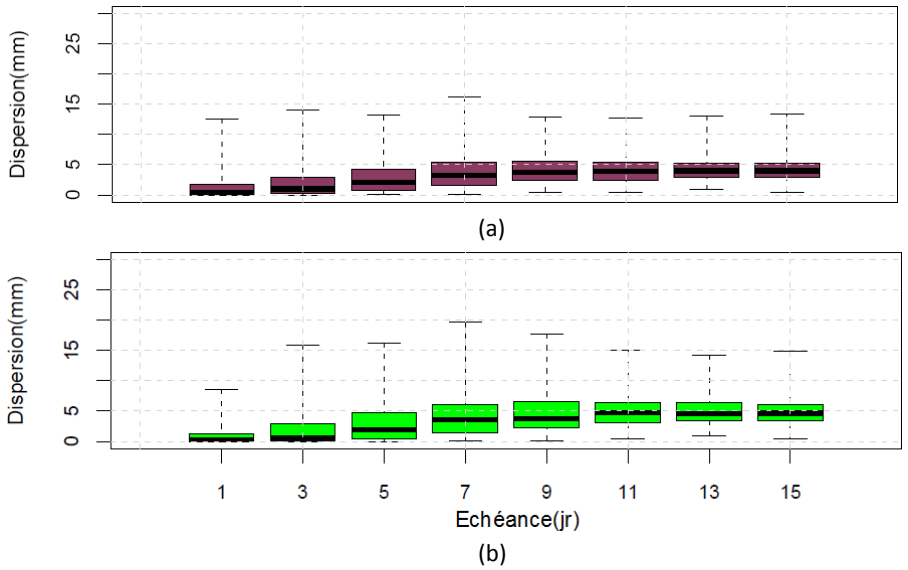


Figure 59 : Boîtes à moustaches représentant l'évolution de la dispersion des membres de l'ensemble de pluie en fonction de l'échéance de prévision pour le bassin de la Drôme à Saillant. Les prévisions sont tirées : (a) du grand ensemble utilisant une pondération de la moyenne d'ensembles des différents centres en fonction du score RPS et (b) du centre européen.

Par la suite nous examinerons la performance de la qualité globale des prévisions météorologiques d'ensemble. Sur l'échantillon de 74 bassins versants examinés, nous présentons des boîtes à moustaches correspondant à la distribution de valeurs du score de probabilité ordonnée *RPS* et du score de divergence (Figure 60 et Figure 61). La ligne rouge correspond à la médiane des valeurs du score de probabilité ordonnée *RPS* pour les prévisions tirées du centre européen ECMF. Les graphiques en haut et en bas présentent les échéances de prévisions 3 et 7 jours respectivement. Nous constatons que :

- en termes de score RPS (Figure 60), la performance de la qualité globale des prévisions des grands ensembles montre que lors des premières échéances de prévision les différentes approches de construction du grand ensemble – avec l'exception du grand ensemble construit en pondérant les membres de l'ensemble (4) – se montrent plus performantes par rapport aux prévisions tirées d'un centre individuel (dans la figure nous les comparons au centre européen - 1). Pour la plupart des bassins versants, le grand ensemble construit avec les moyennes d'ensemble pondérées en fonction du score RPS (3) obtient des scores comparables à ceux du grand ensemble basé sur l'équiprobabilité (2).

- cependant, aux horizons plus lointains de prévision, la performance des différentes combinaisons du grand ensemble proposées se dégrade considérablement. Le grand ensemble basé sur l'équiprobabilité (2) se montre considérablement plus performant par rapport aux autres combinaisons proposées et légèrement meilleur par rapport aux prévisions du centre européen (1) pris individuellement. Une explication possible est liée à la dispersion importante que les deux combinaisons par pondération proposées apportent. D'un côté, la pondération des moyennes d'ensemble (3) améliore la fiabilité des prévisions, mais produit des ensembles sous dispersés aux horizons plus lointains. D'un autre côté, la pondération qui considère tous les membres des prévisions des différents centres (4) introduit plus de dispersion que le nécessaire et, par conséquent, les ensembles sont fiables mais trop dispersés.

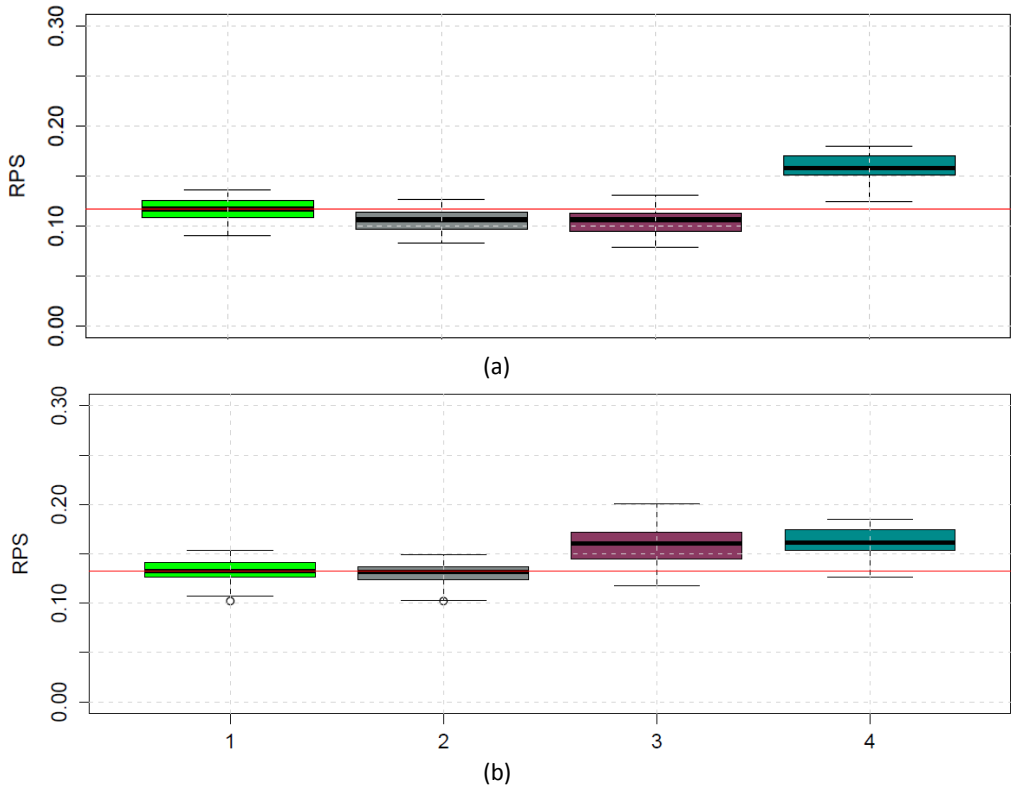


Figure 60 : Boîtes à moustaches représentant la performance des prévisions de pluie en termes de scores de probabilité ordonnée RPS pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours tirées sur les prévisions du : (1) centre européen, (2) grand ensemble basé sur l'équiprobabilité de membres, (3) le grand ensemble construit en pondérant la moyenne d'ensemble des différents centres en fonction du score RPS, (4) le grand ensemble en pondérant les membres en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie sont évaluées sur 242 jours.

- les performances globales des grands ensembles évaluées selon le score de divergence RDS (Figure 61) montrent que, aux horizons courts de prévision, les grands ensembles construits avec les différentes approches de combinaison des centres sont comparables à celle du centre européen (1) individuellement. Ce score permet également de tirer une autre conclusion générale de l'analyse : les prévisions de précipitation des grands ensembles se caractérisent par une tendance à plutôt surestimer l'observation (valeurs RDS inférieures à zéro). Néanmoins, la tendance à la bonne performance des grands ensemble diminue, voire disparaît, aux horizons lointains de prévision. A ces horizons, il devient évident que, en combinant les différents centres avec la méthode de pondération des moyennes d'ensemble, nous obtenons un grand ensemble moins performant et qui a une forte tendance à sous-estimer l'observation (valeurs RDS supérieures à zéro). Similairement à la Figure 60, qui analyse la qualité selon le critère RPS, cette approche de combinaison de centres se présente comme l'approche la moins performante.

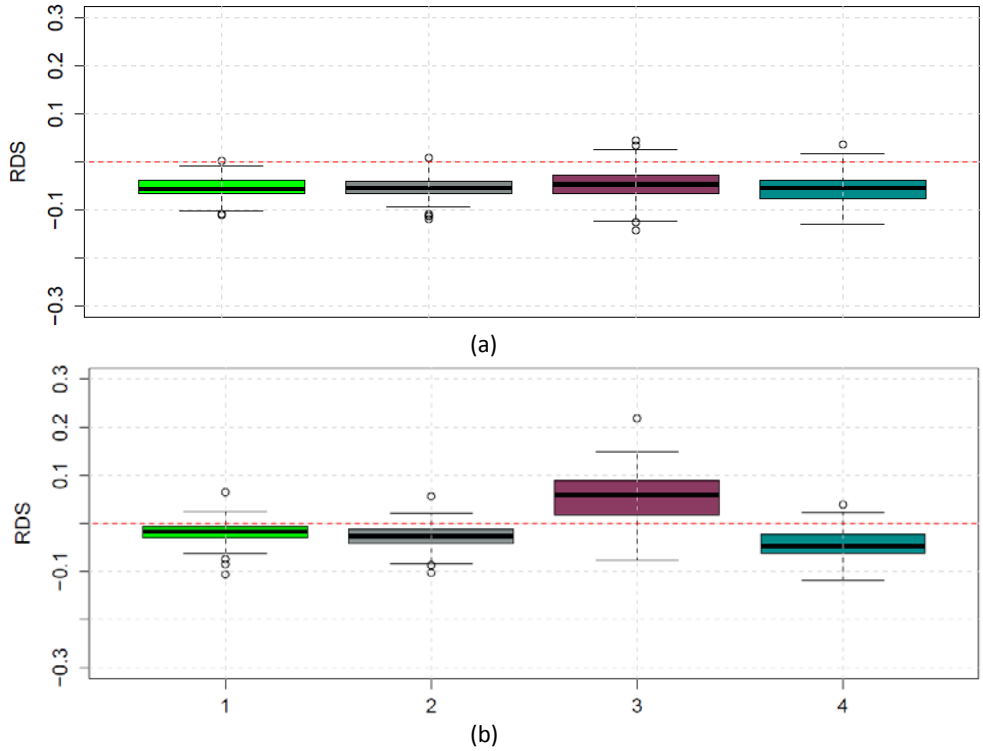


Figure 61 : Boîtes à moustaches représentant la performance des prévisions de pluie en termes de scores de divergence pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours tirées sur les prévisions du : (1) centre européen, (2) grand ensemble basé sur l'équiprobabilité de membres, (3) le grand ensemble construit en pondérant la moyenne d'ensembles des différents centres en fonction du score RPS, (4) le grand ensemble en pondérant les membres en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie sont évaluées sur 242 jours.

5.3.4 Évaluation des prévisions d'ensemble de débits du Grand Ensemble

Nous évaluons maintenant la performance des grands ensembles par rapport à la qualité des prévisions de débits du modèle *GRP*. Les mêmes attributs sont examinés : la fiabilité, la dispersion des membres, la discrimination et la qualité globale. Ces attributs sont évalués à l'aide des mêmes outils graphiques et scores : la fiabilité des prévisions exprimée par l'histogramme de PIT (Figure 62 et Figure 63), les courbes ROC (Figure 65 et Figure 66), l'évolution de la dispersion des membres en fonction de l'échéance de prévision (Figure 64) et les performances sur l'ensemble de 74 bassins versants selon les critères RPS et RDS. Pour les critères graphiques, nous retenons à nouveau les bassins versants de : la Drôme à Saillant et l'Oise à Sempigny.

Sur la qualité générale des prévisions de débits des grands ensembles, nous remarquons que :

- les grands ensembles arrivent à améliorer la fiabilité des prévisions dans les trois cas, avec l'approche par équiprobabilité et avec les deux approches de pondération. Selon le bassin versant, l'amélioration est plus évidente : c'est le cas du bassin de la Drôme à Saillant (Figure 62), où nous remarquons que les prévisions hydrologiques à partir des centres individuels sont soit fortement sous dispersées, soit biaisées et les grands ensembles améliorent cette tendance. Sur d'autres bassins cependant, comme pour le bassin de l'Oise à Sempigny (Figure 63), les grands ensembles n'arrivent à améliorer que légèrement la fiabilité des centres individuels en augmentant la dispersion des membres.
- sur l'ensemble de 74 bassins versants, nous constatons que les approches de combinaisons basées sur l'équiprobabilité (a) et sur la pondération des membres des centres en fonction de leur performance en termes de score RPS (c) sont celles qui améliorent le plus la fiabilité des prévisions. Ce fait est valable pour la majorité des bassins versants, avec la combinaison pondérée se montrant légèrement plus performante. Nous remarquons que cette amélioration est combinée avec une dispersion plus forte des membres de l'ensemble, surtout pour les horizons jusqu'à 5-7 jours (Figure 64).
- En ce qui concerne la performance en termes de discrimination, les courbes ROC pour un seuil de dépassement égal au quantile 70% montrent que les grands ensembles apportent une amélioration en termes de discrimination par rapport aux prévisions de débits obtenues à partir de chaque centre individuellement : par exemple l'aire sous la courbe qui est plus importante (Figure 65 et Figure 66, pour les bassins de la Drôme à Saillant et l'Oise à Sempigny, respectivement). Comme pour les pluies, les capacités de discrimination varient d'un bassin à l'autre : par exemple, le bassin versant de l'Oise à Sempigny (Figure 66) se caractérise par une capacité plus

élevée des prévisions à discriminer entre dépassement et non-dépassement du quantile 70% par rapport au bassin de la Drôme (Figure 65). Cependant, sur l'ensemble de 74 bassins versants, pour la majorité des cas, les grands ensembles améliorent la capacité des prévisions à discriminer entre les événements. De plus, parmi les trois approches de combinaison de centres examinées, celle basée sur l'équiprobabilité des membres (a) est la plus performante, suivie de près de celle construite sur la base de la pondération des différents membres des centres en fonction du score RPS (c).

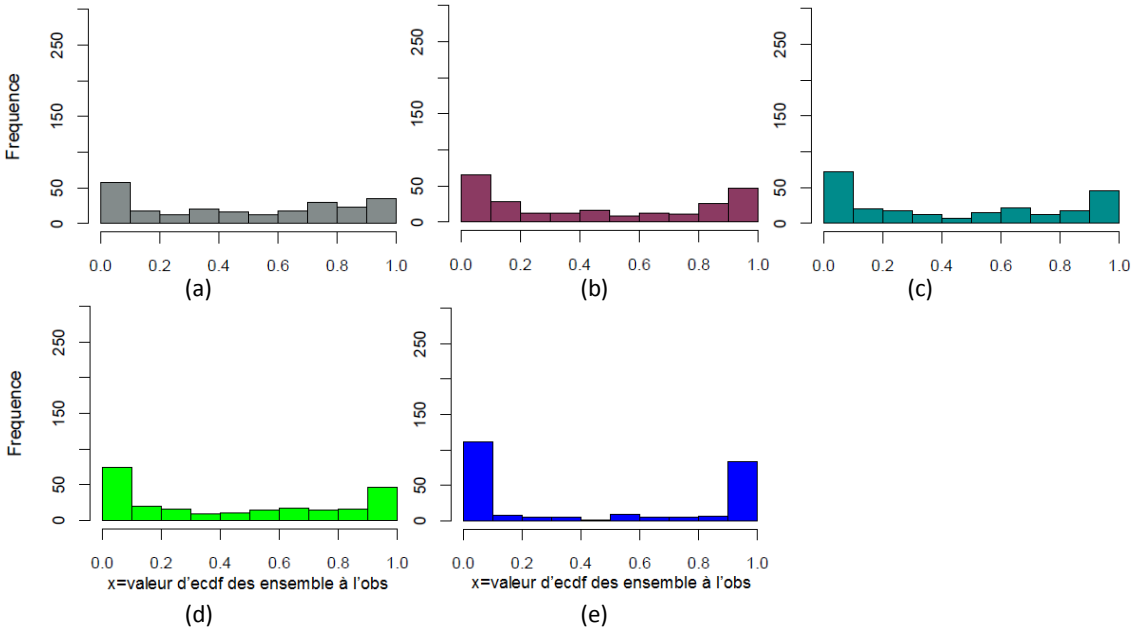


Figure 62 : Histogrammes de PIT pour les prévisions de débit pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le du RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien évaluées sur 242 jours pour tous les centres sont présentées dans la figure (d) & (e).

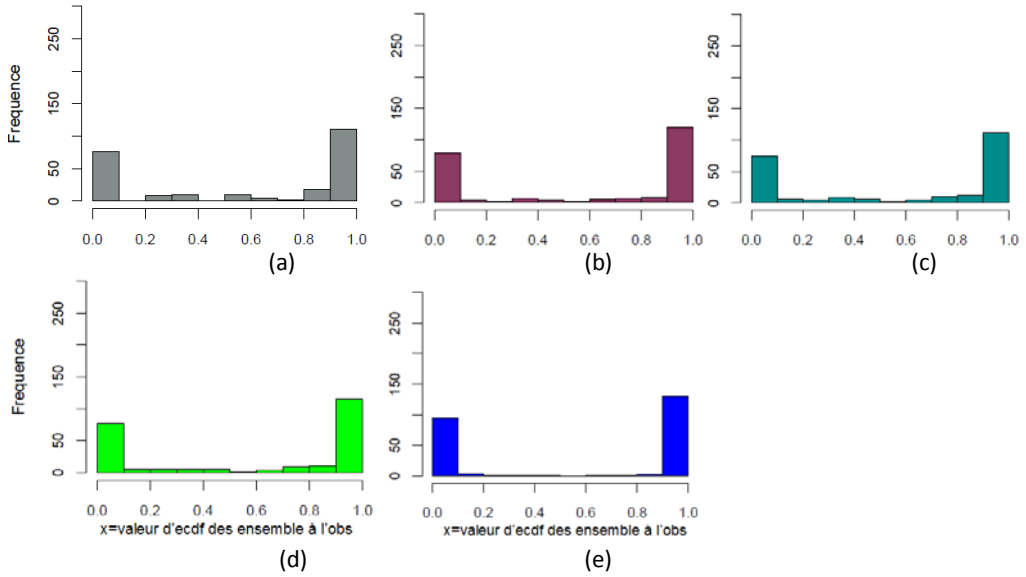


Figure 63 : Histogrammes de PIT pour les prévisions de débit pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité des membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien évaluées sur 242 jours pour tous les centres sont présentées dans la figure (d) & (e).

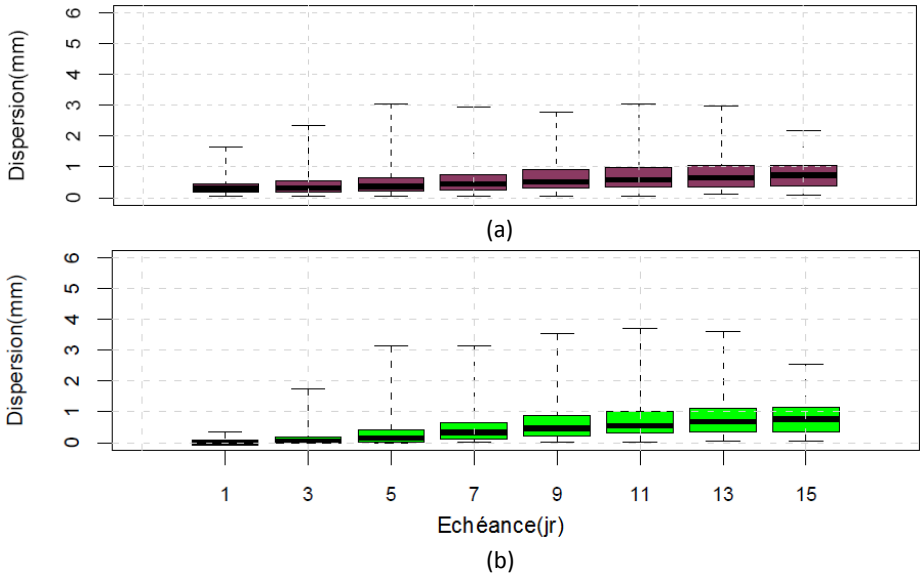


Figure 64 : Boîtes à moustaches représentant l'évolution de la dispersion des membres de l'ensemble de débit en fonction de l'échéance de prévision pour le bassin de la Drôme à Saillant. Les prévisions sont tirées : (a) du grand ensemble utilisant une pondération en fonction du nombre de membres et (b) du centre européen.

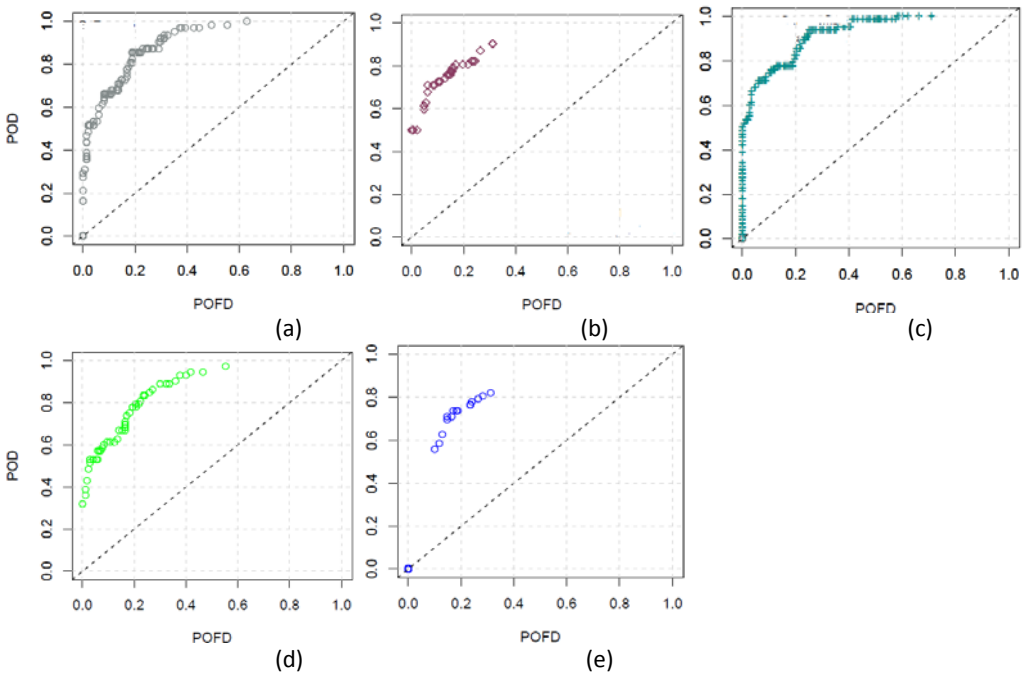


Figure 65 : Courbes ROC pour les prévisions de débit pour le bassin de la Drôme à Saillant à un horizon de 7 jours.

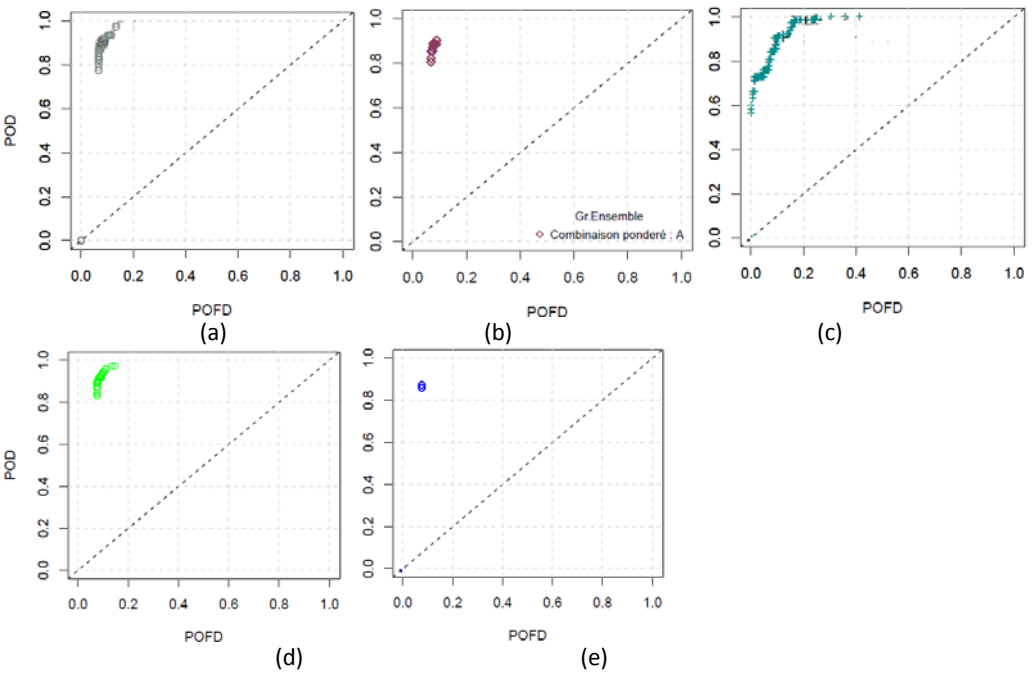


Figure 66 : Courbes ROC pour les prévisions de débit pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées du grand ensemble basé sur :(a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de débit tirées du centre européen et brésilien (d) & (e).

Afin d'obtenir une vision générale sur l'ensemble de l'échantillon de 74 bassins versants examinés, nous présentons des boîtes à moustache correspondant à la distribution des valeurs du score de probabilité ordonnée *RPS* (Figure 67). Les graphiques en haut et en bas présentent les échéances de prévisions 3 et 7 jours respectivement. De plus ce graphique présente en parallèle les performances de prévisions hydrologiques tirées du centre européen ECMF (1), ainsi que celles du grand ensemble basé sur l'equifinalité de membres (2), construit à partir des moyennes d'ensemble pondérées (3) ou en combinant les différents centres en fonction du score RPS (4). De l'analyse de performance globale selon le score RPS, nous remarquons que :

- les grands ensembles montrent une performance améliorée par rapport à celle du centre européen pris individuellement. Nous constatons qu'à la différence des prévisions d'ensemble de précipitation, où seulement le grand ensemble basé sur l'équiprobabilité (2) était plus performant pour toutes les échéances de prévision, dans le cas des prévisions de débit, toutes les différentes combinaisons proposées se montrent clairement plus performantes que les prévisions tirées des centres individuels. Ceci est d'autant plus marquant pour les horizons plus courts, tandis qu'aux horizons de prévisions plus lointains la différence entre les prévisions tirées du centre européen et les différentes combinaisons du grand ensemble se réduisent considérablement.
- les performances globales des prévisions hydrologiques en utilisant le score de divergence RDS (Figure 68) confirment la valeur ajoutée du grand ensemble indépendamment de l'approche utilisée pour combiner les centres. De plus, nous constatons que la combinaison de centres pondérée en fonction du score *RPS* (4) se montre à nouveau la plus performante. L'information supplémentaire, que le score de divergence apporte vis-à-vis des différentes combinaisons proposées, est liée à la sous- ou surestimation de l'observation par les prévisions fournies. Sur cet aspect nous constatons que les trois approches de combinaison grand ensemble ont tendance à surestimer l'observation. Cette tendance est plus forte pour les grands ensembles construits à la base d'équiprobabilité (2) ou de pondération des moyennes d'ensemble (3). En général, cette tendance diminue aux horizons de prévision plus lointains.

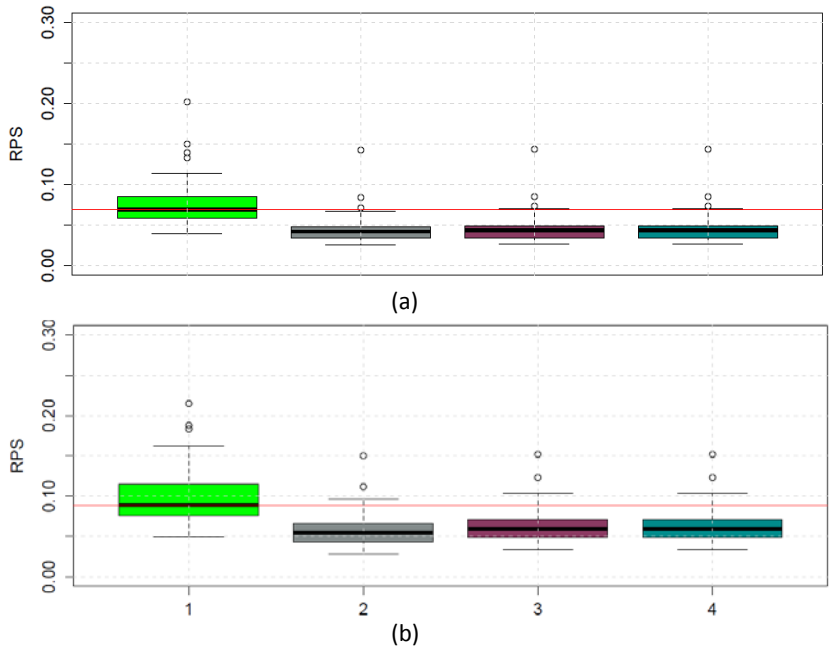


Figure 67 : Boîtes à moustaches représentant la performance des prévisions de débit en termes de scores de probabilité ordonnée RPS pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours pour les prévisions tirées du : (1) centre européen, (2) grand ensemble basé sur l'equifinalité de membres, (3) grand ensemble construit en pondérant la moyenne d'ensemble des différents centres en fonction du score RPS, (4) grand ensemble en pondérant en fonction de leur performance selon le RPS.

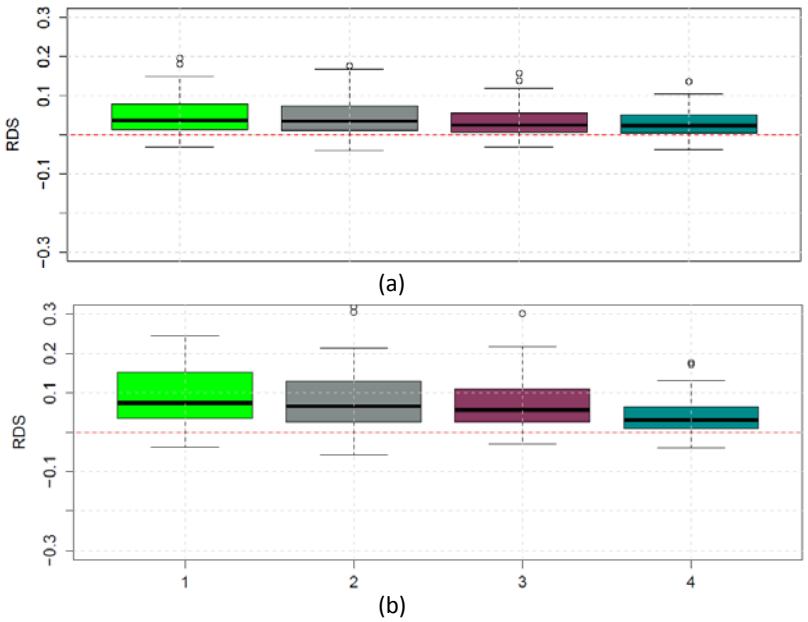


Figure 68 : Boîtes à moustaches représentant la performance des prévisions de débit en termes de scores de divergence pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours.

5.4 Synthèse

Dans ce chapitre nous avons évalué la qualité des prévisions tirées de la base de données TIGGE en utilisant divers critères existant dans la littérature. Nous nous sommes centrés sur la qualité globale des prévisions, ainsi que sur les attributs de la fiabilité et de la discrimination. Nous avons également examiné la relation entre la dispersion des membres de l'ensemble et la racine carrée de l'erreur quadratique moyenne normalisée de prévisions. Cette démarche est effectuée pour les prévisions tirées à partir de 8 centres météorologiques formant la base de données TIGGE. Nous avons évalué les prévisions d'ensemble de pluie et de débit pour chaque centre météorologique séparément en examinant leurs performances en fonction de l'horizon de prévision, la surface et la localisation du bassin versant. Par la suite nous avons essayé de combiner les informations extraites de chaque centre météorologique en formant un grand ensemble composé de tous les membres des différents centres. Différentes approches de combinaison de centres ont été proposées.

Notons que lors de cette étude nous avons examiné des prévisions d'ensemble jusqu'à un horizon de 15 jours. Cependant dans ce chapitre nous avons présenté des résultats jusqu'à une échéance de 9 jours. Ce choix a été fait afin d'introduire une base commune pour la comparaison de différents centres et il était aussi justifié par le fait que la performance de prévisions en termes de qualité se dégrade considérablement au-delà de 9 jours (résultats non présentés ici).

Les conclusions principales de ce chapitre sont les suivantes :

- l'effet écrasant du modèle hydrologique : les prévisions d'ensemble de pluie tirées des 8 centres météorologiques examinés se montrent considérablement plus fiables et dispersées – surtout aux horizons plus lointains de prévision- par rapport aux prévisions hydrologique d'ensemble,
- le passage par le modèle hydrologique améliore significativement la capacité de prévision à discriminer entre deux événements (discrimination),
- pour des bassins versants avec des surfaces supérieures à 2000 km² les performances des prévisions d'ensemble de pluies et de débits dépendent peu de la surface du bassin versant considéré,
- la variation de la performance des prévision d'ensemble de pluie et de débits selon la localisation du bassin versant avec certains centres météorologiques (*ECMF, RJTD, CWAO*) bien plus performants que d'autres pour la plupart de bassins versants,

- la conclusion précédente se confirme également pour les cas des événements extrêmes de pluie ; quant aux prévisions de débits aucun centre n'arrive à être performant pour l'ensemble des bassins versants de notre échantillon,
- la combinaison de prévisions extraites de différents centres météorologiques apporte une amélioration à la qualité de prévisions d'ensemble de pluie et de débit en termes de qualité globale. Toutefois, cette amélioration est plus significative pour les prévisions de débit et elle se concentre surtout aux horizons courts de prévisions,
- le grand ensemble contribue à la fois à l'amélioration de la discrimination et de la fiabilité pour les prévisions de pluie et de débit.

Dans les chapitres qui suivent nous utiliserons les prévisions d'ensemble de pluie du centre européen *ECMF,* car elles disposent d'une archive plus longue ce qui nous permet des mieux caler nos méthodes de post-traitement. Plusieurs raisons justifient ce choix :

- le centre européen se révèle être le centre le plus performant en termes de qualité globale pour la plupart des bassins versants dans la base des 74 bassins versants à notre disposition et pour tous les horizons de prévision examinés,
- les prévisions d'ensemble du centre européen se sont montrées plus fiables par rapport aux prévisions de la plupart de centres examinés, ainsi qu'elles ont montré une capacité élevée à discriminer entre des événements différents (discrimination),
- la longueur de l'archive de prévision disponible : le centre européen possède l'archive de prévision d'ensemble la plus longue à notre disposition.

Pour la suite nous retenons alors les prévisions d'ensemble du centre européen *ECMF* afin de corriger les prévisions d'ensemble par différents traitements statistiques (Chapitre 6) et de valoriser l'utilité de prévisions d'ensemble dans un contexte opérationnel (Chapitre 7).

"Tout doit être construit du haut vers le bas, sauf la première fois."

« Epigrams on Programming », Alan J. Perlis (1982)

6 Traitement statistique des prévisions

Dans ce chapitre[1] nous examinons différentes approches pour le traitement statistique des prévisions et la correction d'éventuels biais présents dans celles-ci. L'étude présentée ici est basée sur les prévisions météorologiques d'ensemble du centre européen (ECMWF) et les deux modèles hydrologiques à notre disposition : MORDOR et GRP. Le jeu de 11 bassins versants est utilisé. Dans un premier temps, nous évaluons les prévisions hydrométéorologiques d'ensemble brutes (sorties directes des modèles) afin de déterminer les différents erreurs de prévision. Ensuite, pour traiter ces erreurs, nous examinons différentes approches de traitement statistique appliquées soit sur les prévisions de pluie, soit sur les prévisions de débit. Enfin, nous abordons quelques questions clés liées à la performance des approches de traitement statistique et aux stratégies à suivre pour la mise en place d'un module de correction de biais des prévisions dans une chaîne de prévision hydrométéorologique.

[1] *Une partie de ce chapitre est publieé dans* :

- Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research*, 8, 135-141.

6.1 Introduction

Dans le chapitre précédent (Chapitre 5), nous avons montré que, parmi les différents centres météorologiques constituant la base de prévisions d'ensemble TIGGE, le centre européen (*ECMF*) était celui qui produisait des prévisions hydrométéorologiques les mieux adaptées en termes de qualité sur la France. Par conséquent, et pour faciliter l'investigation des approches de traitement statistique présentée dans ce chapitre, nous avons choisi de poursuivre les analyses avec les seules prévisions d'ensemble du centre européen (*ECMF*).

De plus, l'étude proposée dans ce chapitre vise à examiner quelles méthodes et stratégies de correction de biais seraient les plus intéressantes pour la prévision opérationnelle. De ce fait, nous utilisons ici l'échantillon de 11 bassins versants de fort intérêt économique pour la production hydroélectrique à EDF (voir Chapitre 2, §2.3). Ces bassins versants ont des superficies plus petites par rapport aux 74 bassins versants étudiés dans le Chapitre 5 et sont plus adaptés pour étudier l'influence d'un traitement statistique sur les précipitations moyennes de bassin, donc en amont du modèle hydrologique, car à ces échelles les aggravations locales mal représentées par les modèles météorologiques deviennent sensibles.

L'archive de prévisions d'ensemble du CEPMMT utilisée ici est basée sur celle disponible à EDF. Elle s'étend de janvier 2005 jusqu'à fin décembre 2008 (environ 4 années de prévisions). Il s'agit des prévisions de précipitation et de température des 50 membres de la prévision d'ensemble pour un horizon maximal de 7 jours. Ces prévisions sont utilisées dans les deux modèles conceptuels pluie-débit de cette étude : le modèle *MORDOR* et le modèle *GRP* (voir Chapitre 2, §2.4). Nous rappelons que les deux modèles sont couplés avec un module neige afin de représenter les écoulements liés à la fonte du manteau neigeux. Les deux modèles hydrologiques ont été calés sur les 11 bassins versants à partir des longues périodes de données historiques disponibles et de leurs propres fonctions objectifs de calage (voir Chapitre 2, §2.4). L'intérêt de l'utilisation en parallèle de deux modèles est de tester la robustesse des approches utilisées et de consolider les conclusions tirées. Nous avons ainsi un modèle à 11 paramètres (*MORDOR*) et un modèle plus parcimonieux à 5 paramètres (*GRP*).

Dans ce chapitre, nous évaluons les prévisions émises sur une période ultérieure au calage et, à travers la mise en place de différentes approches de traitement statistique, nous examinons l'impact de ces méthodes sur la qualité des prévisions. Plus particulièrement, nous visons à répondre aux questions suivantes :

- quels problèmes retrouvons-nous le plus souvent dans les prévisions brutes de pluie et de débit ?
- est-ce qu'il existe une méthode de traitement statistique à privilégier ?
- quelle est la meilleure stratégie à adopter : est-il suffisant d'effectuer le traitement statistique seulement sur les prévisions de pluie, seulement sur les prévisions de débit ou faut-il effectuer les deux ?
- quel est l'impact de la longueur de la période de calage des paramètres des méthodes de traitement statistique sur la performance des approches de correction de biais ?

6.2 Scénarios de traitement statistique

Nous avons vu qu'il est essentiel d'évaluer les prévisions en termes des deux axes composant la valeur d'une prévision : la qualité et l'utilité. Les objectifs et les outils d'évaluation de la qualité et de l'utilité d'une prévision ont été présentés en détail dans les Chapitres 3 et 4. Dans ce chapitre, nous nous focalisons sur la qualité des prévisions et le traitement statistique de celles-ci. Notre point de vue est centré sur l'hydrologie. En laissant de côté l'évaluation du modèle hydrologique dans la chaîne hydrométéorologique de prévision, telle que nous l'avions schématisée (Chapitre 1, §1.3, Figure 3), nous pouvons évaluer, d'un coté, les entrées du modèle hydrologique (prévisions météorologiques) et, d'un autre côté, les sorties du modèle (prévisions hydrologiques). Des erreurs de prévision peuvent exister dans ces deux cas et, par conséquent, l'application d'un traitement statistique peut être nécessaire soit sur les prévisions météorologiques (nous parlerons alors de " prétraitement " des entrées du modèle hydrologique), soit sur les prévisions hydrologiques (on parlera de " post-traitement " des sorties du modèle hydrologique), ou encore sur les deux.

Les techniques de traitement statistique examinées dans cette thèse ont été présentées dans le Chapitre 3. Pour rappel, elles sont les suivantes :

- méthode empirique basée sur un facteur de correction multiplicative,
- habillage à l'aide de l'information de la prévision par analogie,
- habillage par la méthode du meilleur membre développée par Roulston et Smith (2003),
- habillage par la méthode du meilleur membre modifiée par Wang et Bishop (2005) et Fortin *et al.* (2006),

- habillage par la méthode du meilleur membre modifiée pour tenir compte de la cohérence temporelle entre les échéances de prévision,
- habillage avec l'erreur de 30% des meilleurs membres,
- habillage empirique avec les erreurs passées du modèle hydrologique.

Les deux premières techniques sont plutôt appliquées aux prévisions météorologiques (prétraitement), tandis que les autres techniques sont appliquées aux prévisions de débits des deux modèles hydrologiques étudiés (post-traitement). Toutes les techniques, à l'exception de la dernière, font partie des approches de traitement de l'incertitude totale de prévision (qu'elle soit de précipitation ou de débit, selon que son application se fasse sur les précipitations ou les débits prévus). En revanche, la dernière technique ne traite que l'incertitude au niveau du modèle hydrologique, ce qui nous permet d'examiner l'influence de la prise en compte des erreurs seules du modèle hydrologique sur la performance des prévisions de débit.

Ces techniques ont été utilisées pour composer les scénarios de pré- et post- traitement ici évalués. Le Tableau 8 présente la composition de ces scénarios. Nous y retrouvons les techniques mentionnées seules ou combinées.

Ces scénarios vont être appliqués au jeu de 11 bassins versants de cette étude et nous illustrerons quelques-uns des résultats à l'aide de 3 bassins versants (Figure 69) qui serviront d'exemple, car ils présentent des comportements différents :

- la Durance à Serre Ponçon (3580 km²),
- le Verdon à Castillon (656 km²), et
- la Romanche à Chambons (223 km²).

La qualité des prévisions hydrométéorologiques est évaluée à l'aide des critères présentés dans le Chapitre 3 (§3.2).

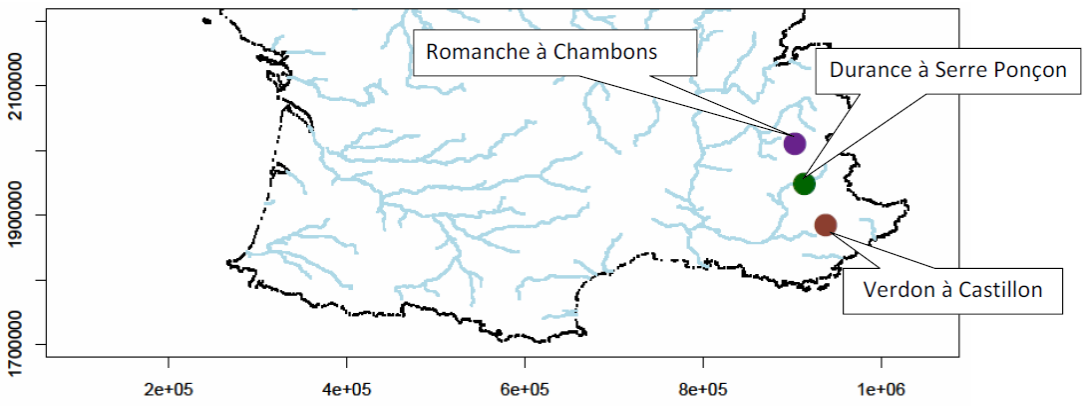


Figure 69 : Localisation des trois bassins versants utilisés en exemple.

Tableau 8 : Liste de scénarios de traitement statistique évalués.

Scénario	Technique de correction de biais	Prévision traitée	Stratégie de traitement statistique dans la chaîne hydrométéorologique
1	NA	NA	NA (Prévisions brutes)
2A	Empirique par facteur de correction	Précipitation	Prétraitement
2B	Habillage avec analogues	Précipitation	Prétraitement
3	Habillage avec les erreurs passées du modèle hydrologique	Débit	Post-traitement
4	Combinaison des techniques des scénarios : 2B et 3	Précipitation et débit	Prétraitement et post-traitement
5	Habillage avec la méthode du meilleur membre (2003)	Débit	Post-traitement
6	Habillage avec la méthode du meilleur membre modifiée (2005)	Débit	Post-traitement
7	Habillage avec la méthode du meilleur membre modifiée avec la covariance (cohérence temporelle)	Débit	Post-traitement
8	Habillage avec les erreurs des 30% des meilleurs membres	Débit	Post-traitement
9	Habillage avec la méthode du meilleur membre modifiée (2006)	Débit	Post-traitement

6.3 Evaluation des prévisions brutes de pluies et de débits

Pour l'évaluation de la qualité des prévisions, nous avons fait appel aux critères graphiques et scores numériques suivants : l'histogramme et diagramme de PIT, la dispersion des membres de l'ensemble, les courbes ROC et le score CRPS. Nous tenons à garder les mêmes scores d'évaluation pour les prévisions de pluies et de débits afin de rendre l'analyse homogène et de pouvoir étudier la variabilité des performances qu'il s'agisse d'évaluer les entrées ou les sorties des modèles hydrologiques.

6.3.1 Prévisions de pluie

Fiabilité : l'histogramme et diagramme de PIT

Nous avons utilisé l'histogramme de PIT pour évaluer la fiabilité des prévisions brutes de pluies des 11 bassins versants de notre échantillon. La Figure 70 présente les histogrammes de PIT pour les prévisions de pluies sur les 3 bassins versants utilisés ici à titre d'exemple et pour deux horizons de prévision : 3 jours (à gauche) et 7 jours (à droite). Les diagrammes cumulés de PIT correspondants sont également présentés au coin en haut à gauche de chaque histogramme.

Nous observons que pour l'horizon de prévision plus court (3 jours), l'histogramme de PIT montre des prévisions avec un défaut de fiabilité et l'existence d'un biais, plus particulièrement, d'un biais de sous-estimation des valeurs observées (biais indiqué par les fréquences élevées des fortes valeurs de probabilité prévue au non-dépassement associées aux observations). La présence de ce biais devient moins évidente à l'horizon plus lointain (7 jours) et, pour certains bassins comme le cas du Verdon à Castillon (c2, Figure 70), les prévisions sont très proches d'un système parfaitement fiable.

De manière générale, dans l'ensemble de notre échantillon de 11 bassins versants, l'analyse des histogrammes de PIT calculés sur les prévisions de pluie brutes a montré :

- pour les horizons courts de prévision (3-4 jours) : les prévisions de pluie sont sous-dispersées pour presque tous les bassins versants examinés,
- pour les horizons plus lointains de prévision (5-7 jours) : des prévisions de pluie peuvent êtres considerées fiables pour la majorité de bassins versants (8 bassins sur 11), et des prévisions sous-estimant les pluies pour une minorité de bassins versants (3 bassins sur 11).

Ces conclusions sont cohérentes avec les conclusions du Chapitre 5 et confirment que les prévisions de pluie ont la tendance d'être sous-dispersives lors des premières échéances de prévisions, mais elles deviennent fiables aux horizons de prévision plus lointains.

Dispersion des membres d'ensemble

En ce qui concerne la dispersion de l'ensemble, la Figure 71 présente la distribution (sous la forme de boîtes à moustaches) des valeurs de dispersion des membres, calculées pour chaque jour de la période d'évaluation, pour les différentes échéance de prévision et pour la période entière d'évaluation, pour les trois bassins versants utilisés à titre d'exemple. Dans la Figure 72, nous présentons la relation entre les valeurs de la dispersion des membres et les erreurs RMSE calculées également pour chaque jour de la période d'évaluation. Ici, deux échéances de prévision sont présentées : 3 et 7 jours.

Nous remarquons que la dispersion des membres augmente en fonction de l'échéance, comme espéré : par exemple, aux horizons courts de prévision les quantiles 25%-75% de la distribution des dispersions calculées à chaque jour de prévision (les extrémités en haut et en bas de chaque boîte à moustache), ainsi que les quantiles 50% (trait au milieu des boîtes) sont très proches, tandis que pour les échéances plus lointaines, les longueurs des boîtes à moustache deviennent de plus en plus larges et les médianes, plus dissemblables. En moyenne sur l'ensemble des bassins versants, la dispersion médiane à 7 jours est de l'ordre de 3 à 4 fois plus grands que la dispersion à 3 jours.

Une autre conclusion tirée de l'analyse et illustrée par la Figure 71 est que la dispersion des membres diffère peu d'un bassin versant à l'autre : par exemple, nous remarquons que la dispersion des membres suit une évolution très proche pour les bassins versants de la Durance à Serre Ponçon et du Verdon à Castillon (graphiques a et b de la figure), avec peu de dispersion dans les premières échéances. En revanche, dans le cas du bassin versant de la Romanche à Chambon (graphique c), la dispersion des membres croît déjà de manière importante dès les premières échéances. Ceci, en combinaison avec les conclusions tirées de l'analyse de fiabilité (Figure 70), renforce le fait qu'une plus large dispersion des membres conduit à plus de chance d'avoir l'observation dans le faisceau des trajectoires de l'ensemble et, par conséquent, d'atteindre une meilleure fiabilité.

Enfin, l'analyse de la dispersion des membres des prévisions d'ensemble de pluie brutes de notre échantillon de 11 bassins versants, ainsi que la comparaison de la dispersion de l'ensemble avec la racine carrée de l'erreur quadratique moyenne (*RMSE*) a montré que :

- sur l'ensemble des bassins versants, la dispersion augmente avec l'échéance de prévision,

- sur 11 bassins versants la dispersion médiane à 7 jours est au moins 3 à 4 fois plus élevée que celle à 3 jours,
- quand la dispersion de l'ensemble est petite, il existe une corrélation forte entre la racine carrée de l'erreur quadratique moyenne et la dispersion. Ceci signifie qu'une partie importante de l'erreur des prévisions est expliquée par la dispersion,
- la corrélation entre la dispersion et la racine carrée de l'erreur quadratique moyenne est plus forte aux horizons plus courts de prévision. Cette remarque est valable pour les 11 bassins évalués, néanmoins la valeur à laquelle la racine carrée de l'erreur quadratique moyenne n'est plus expliquée par la dispersion varie selon le bassin versant.

Discrimination : Courbes de ROC

La Figure 73 présente les courbes ROC de la prévision d'ensemble des pluies de bassin brutes pour un seuil de dépassement égal au quantile 70% des pluies observées. Cette figure illustre les résultats pour les trois bassins versants choisis comme exemples et pour les échéances de 3 et 7 jours. Rappelons que la courbe ROC trace les taux de fausses alarmes (POFD) du tableau de contingence contre les probabilités de bonne détection (POD) pour différents seuils de risque (probabilités de prévision données par le nombre de membres de l'ensemble dépassant le seuil de discrimination choisi). Idéalement, l'aire sous la courbe ROC doit être égale à 1, avec les points de la courbe proches du coin en haut et à gauche du diagramme.

Nous observons que, pour les trois bassins versants illustrés à la Figure 73, les prévisions arrivent à bien discriminer les dépassements ou non du quantile 70% quand l'horizon de prévision est court (jusqu'à 3 à 5 jours selon le bassin versant). Jusqu'à cet horizon le non-dépassement du quantile 70% est prévu avec une probabilité de bonne détection superieure à 60% par la majorité de membres (25 à 40/50 selon le bassin versant et l'échéance). Au-delà de cet horizon, le nombre de membres qui prévoient le non-dépassement diminue progressivement (15 à 25/50 selon le bassin versant). Ce comportement s'accentue quand le quantile au non-dépassement passe de 50 à 90%.

Les résultats illustrés reflètent les tendances retrouvées sur l'ensemble des 11 bassins versants étudiés et confirment également les conclusions tirées sur les 74 bassins versants.

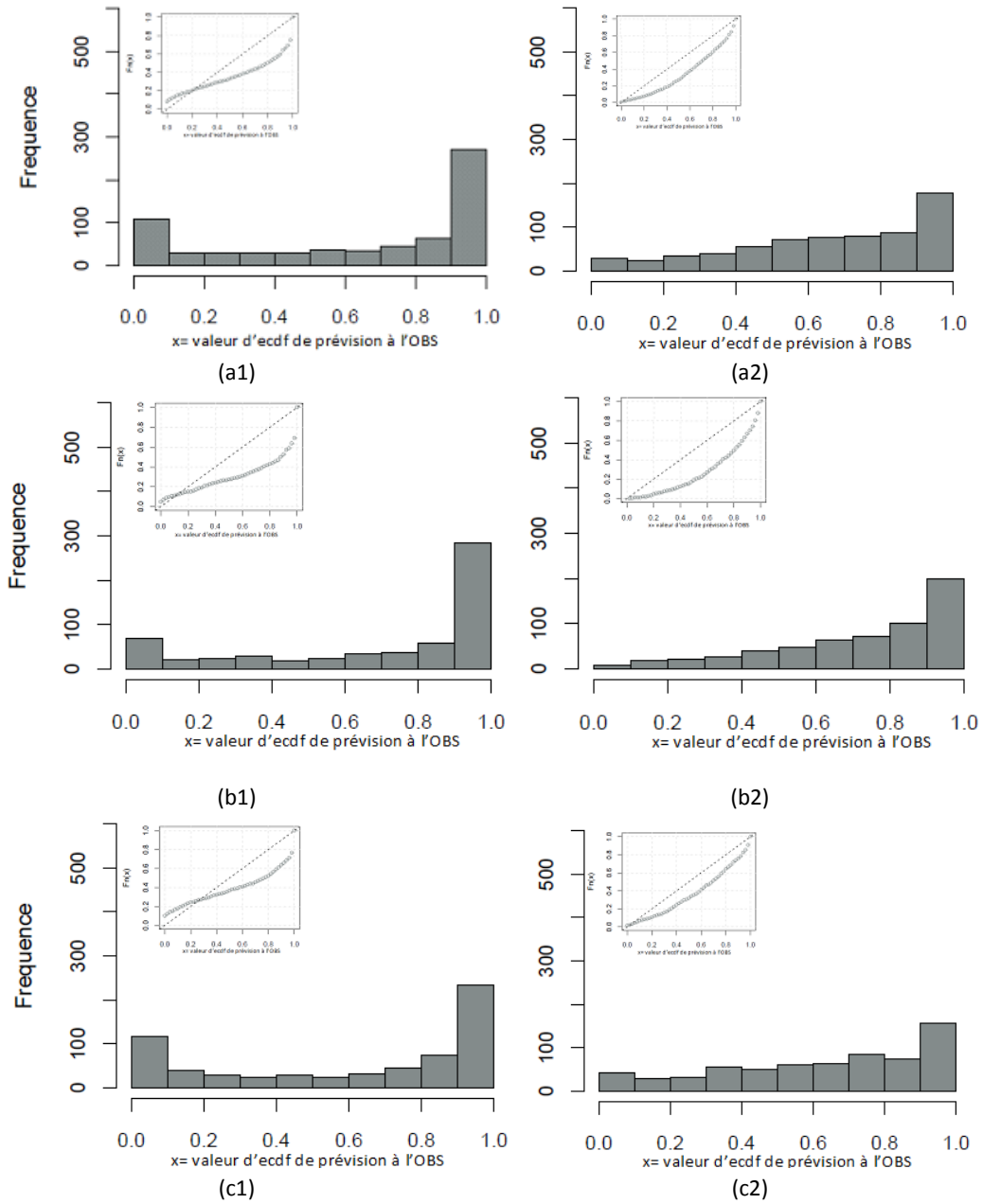
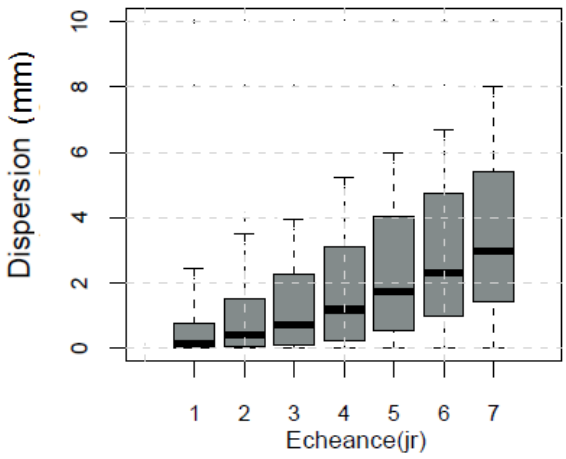
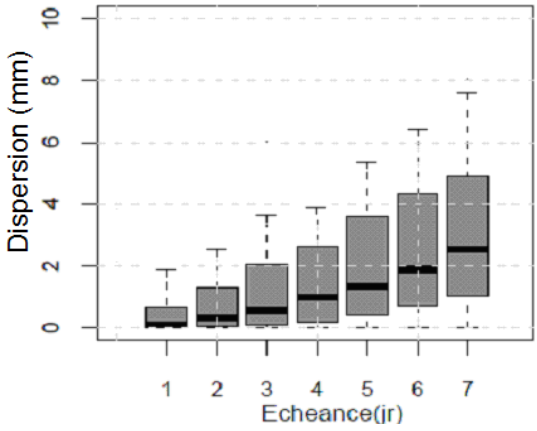


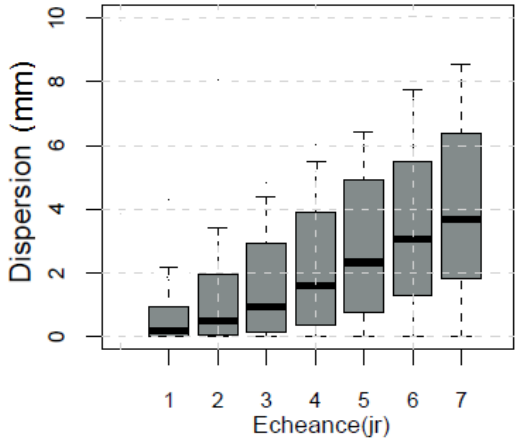
Figure 70 : Histogrammes de PIT – avec le diagramme de probabilités cumulées en haut gauche - pour les prévisions d'ensemble de pluies du CEPMMT brutes pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite).



(a)



(b)



(c)

Figure 71 : Evolution de la dispersion des membres de la prévision d'ensemble du CEPMMT brutes des pluies en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

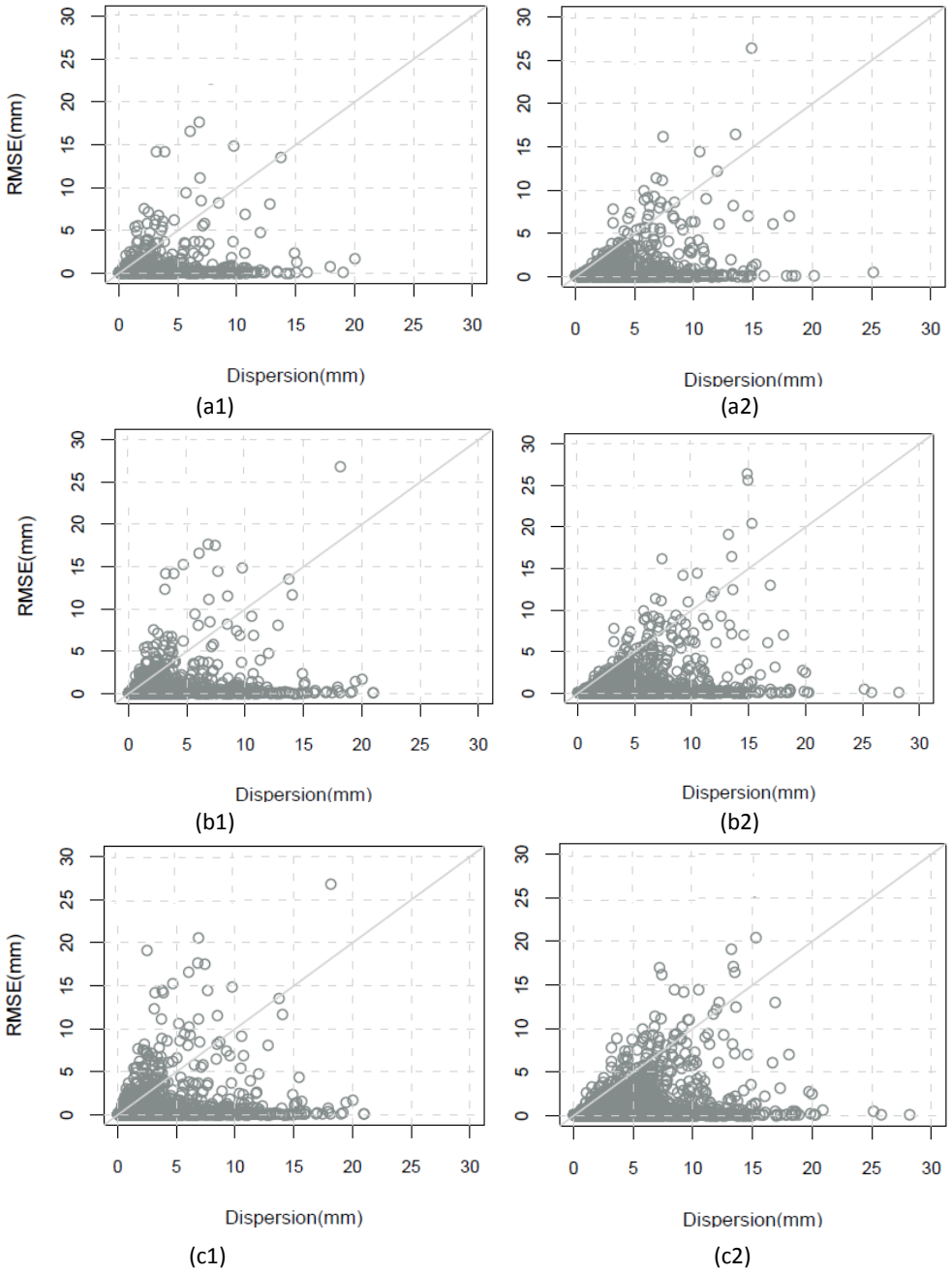


Figure 72 : Relation de la racine carrée de l'erreur quadratique moyen (RMSE) et la dispersion de la prévision d'ensemble des pluies du CEPMMT brutes pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours.

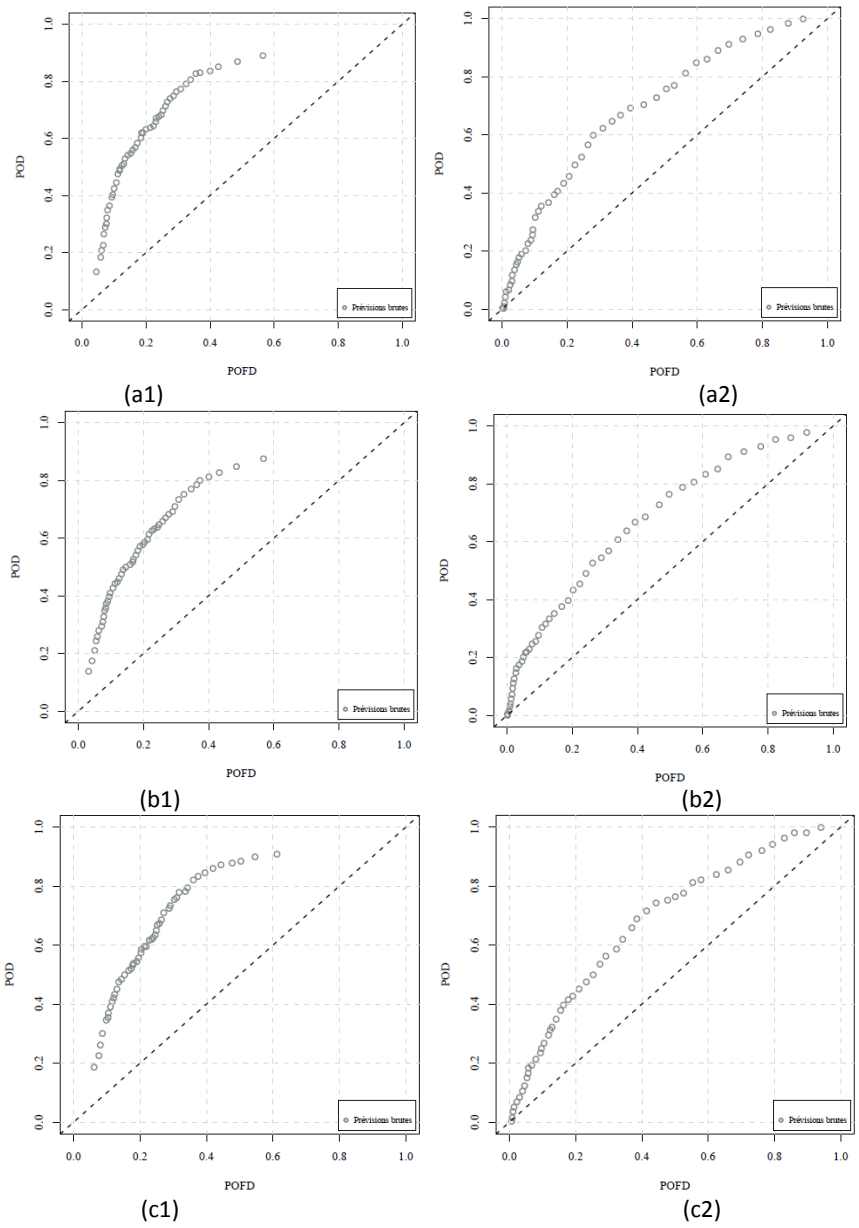


Figure 73 : Courbes ROC de la prévision d'ensemble des pluies de bassin brutes pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite).

Qualité globale : score *CRPS*

Afin d'avoir une vision plus globale sur la performance des prévisions de précipitation sur l'échantillon de 11 bassins versants, nous avons également calculé le score *CRPS* (Chapitre 3, §3.2). Même si ce critère numérique d'évaluation ne donne pas un aperçu sur chacun des attributs de la qualité des prévisions séparément, il reste pourtant un indice de la qualité globale des prévisions et surtout une façon pratique de comparer différents bassins versants entre eux.

La Figure 74 illustre les valeurs moyennes du score numérique *CRPS* pour les 11 bassins versants testés en fonction de la surface du bassin (en km²) et pour les échéances de 3 jours et 7 jours. Rappelons que la valeur parfaite pour le *CRPS* est égale à 0 et que le score n'est pas borné vers les valeurs maximales. Nous remarquons qu'il y a une tendance à avoir des valeurs plus élevées de CRPS pour les bassins plus petits, avec une stabilisation de la valeur du score à partir de 1000 km² de surface.

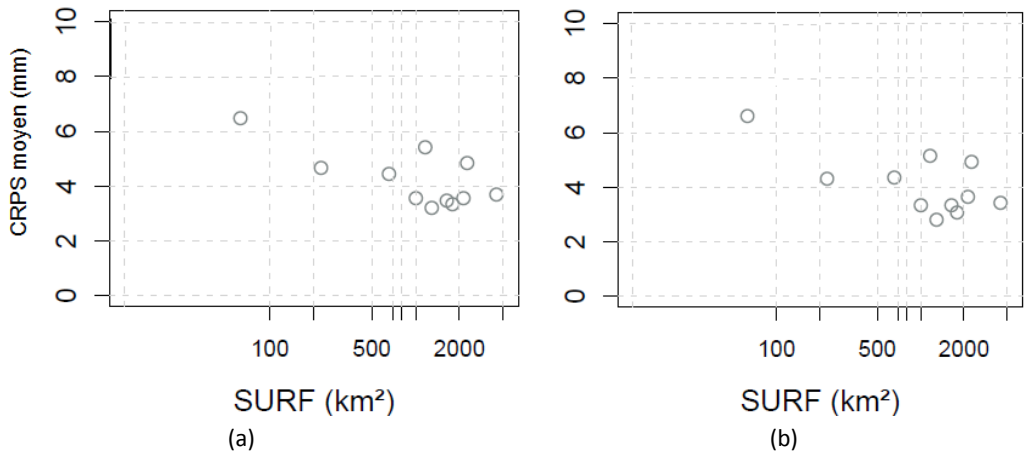


Figure 74 : CRPS moyens de la prévision d'ensemble des précipitations de bassin brutes en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours.

6.3.2 Prévisions de débit : modèles MORDOR et GRP

Les prévisions de précipitation ont été utilisées dans deux modèles hydrologiques, où elles ont été transformées en prévisions de débit. Les deux modèles globaux pluie-débit utilisés sont : le modèle *MORDOR* et le modèle *GRP* (voir Chapitre 2, §2.4).

Fiabilité : Histogrammes et diagrammes de PIT

Les Figure 75 et Figure 76 présentent les histogrammes de PIT (et digrammes cumulés correspondants) pour les prévisions hydrologiques fournies par le modèle *MORDOR* et *GRP* respectivement, et pour les horizons de prévision de 3 et 7 jours. Dans ces graphiques nous affichons les résultats pour les mêmes trois bassins versants utilisés dans l'évaluation des pluies : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons.

En ce qui concerne le modèle MORDOR (Figure 75), nous remarquons que les prévisions hydrologiques sont fortement sous-dispersées (histogrammes en forme de "U" et diagrammes éloignés de la diagonale) quel que soit le bassin versant. Sur certains bassins (5 parmi les 11 bassins analysés), les prévisions ont tendance à sous-estimer l'observation (comme il en est pour les bassins de la Durance à Serre Ponçon et du Verdon à Castillon), tandis que sur d'autres bassins (3 parmi les 11 bassins analysés), cette tendance s'inverse légèrement (comme pour le bassin versant de la Romanche aux Chambons et la Truyère à Grandval).

Pour les prévisions hydrologiques du modèle *GRP* (Figure 76), nous remarquons que, même si la tendance à sous-estimer l'observation est toujours présente, la modélisation des débits est différente avec les deux modèles hydrologiques utilisés. Nous remarquons que les prévisions issues du modèle *GRP* sont legèrement plus sous-dispersées que celles du modèle *MORDOR*. La différence la plus frappante se trouve sur le bassin de la Romanche aux Chambons, où nous observons que la fiabilité des prévisions hydrologiques est nettement dégradée par rapport à celle de prévisions météorologiques, avec la présence d'un biais plus forte dans les débits prévus par *GRP*.

En ce qui concerne la qualité des prévisions hydrologiques en fonction de l'horizon de prévision, quel que soit le modèle hydrologique évalué, nous observons une légère amélioration de la fiabilité aux horizons de prévision plus lointains. Toutefois, la fiabilité reste un attribut à améliorer pour toutes les échéances pour rendre les prévisions de débits des deux modèles fiables.

Dispersions des membres de l'ensemble

L'évolution de la dispersion des membres de l'ensemble des prévisions hydrologiques fournies par les modèles *MORDOR* et *GRP* est présentée dans la Figure 77. De plus, la relation entre dispersion de l'ensemble et la racine carrée de l'erreur quadratique moyenne (*RMSE*) est montrée dans les Figure 78 et Figure 79 pour les deux modèles hydrologiques respectivement.

De manière générale, nous observons que, comparativement aux pluies, la dispersiosn des membres de la prévision d'ensemble de débit augmente moins rapidement avec l'échéance. Nous constatons que les modèles hydrologiques affectent la dispersion des membres de l'ensemble, car la dispersion des prévisions de précipitation est, en général, de l'ordre de deux fois supérieure à celle des débits. De manière générale, sur l'ensemble des bassins versants étudiés, la dispersion moyenne des débits à 7 jours d'échéance varie entre 2 et 5 fois la dispersion à 3 jours.

Dans les exemples illustrés à la Figure 77, nous observons que pour les deux modèles hydrologiques et pour les bassins versants de la Durance à Serre Ponçon et du Verdon à Castillon la dispersion des membres est très faible et elle n'augmente que légèrement en fonction de l'échéance. Par contre, sur le bassin versant de la Romanche à Chambon, des différences entre les modèles sont plus visibles : pour le modèle *MORDOR*, les membres sont beaucoup plus dispersés dès les premières échéances, tandis que pour le modèle *GRP* la dispersion reste très faible sur l'ensemble des échéances examinées.

Sur l'ensemble des 11 bassins versants étudiés, nous remarquons que, en général, la dispersion du modèle *GRP* n'est comparable à la dispersion du modèle *MORDOR* que jusqu'à l'échéance de 3 jours. Au-delà de cette échéance de prévision, la différence entre les dispersions produites par les deux modèles s'élargit : à l'horizon de 7 jours, la dispersion du modèle *MORDOR* varie entre 1.5 à 3 fois la dispersion du modèle *GRP,* selon le bassin versant.

En ce qui concerne la relation entre la RMSE et la dispersion, nous remarquons que :

- pour le modèle *MORDOR* et les trois bassins testés, lors des horizons courts de prévision (à gauche) une partie de l'erreur de prévision est expliquée par la dispersion. Cela est vrai surtout pour le bassin versant de la Romanche à Chambons. Pour les deux autres bassins – la Durance à Serre Ponçon et le Verdon à Castillon – la partie de l'erreur de prévision expliquée par la dispersion est plus limitée. Aux horizons de prévision plus lointains (à droite) et pour les trois bassins présentés la corrélation entre la racine carrée de l'erreur quadratique et la dispersion reste faible. Cette remarque est valable pour les 11 bassins évalués.

- pour le modèle *GRP* la Figure 79 montre qu'une partie importante de la racine carrée de l'erreur quadratique est expliquée par la dispersion pour le cas de la Durance à Serre Ponçon (a), du Verdon à Castillon (b) et de la Romache aux Chambons (c). Cette corrélation existe pour toutes les échéances, mais elle devient plus forte pour les horizons plus courts. De plus, cette remarque est valable pour les 11 bassins de notre échantillon. Dans le cas de la Romanche aux Chambons la partie de la racine carrée de l'erreur quadratique moyenne normalisée (*RMSE*) expliquée par la dispersion non-appropriée est considérablement plus important que dans le cas de prévisions de débits produites par le modèle *MORDOR*. Cette conclusion est soutenue par la très faible dispersion des membres apparente dans la Figure 77. Enfin, de même que pour le modèle *MORDOR*, la valeur par laquelle la racine carrée de l'erreur quadratique moyenne n'est plus expliquée par la dispersion varie selon le bassin versant.

En comparant les Figure 78 et Figure 79 avec les figures correspondantes sur les prévisions de précipitations (Figure 72) nous remarquons que pour les prévisions de précipitations il y a une corrélation forte entre la racine carrée de l'erreur quadratique moyenne et la dispersion pour des horizons courts. Le fait que cette corrélation devient considérablement moins forte pour les prévisions de débits est indicatif d'une source d'erreur supplémentaire (autre que la dispersion non-appropriée) qui affecte les prévisions hydrologiques.

Discrimination : courbes de ROC

Les courbes ROC sont examinées, toujours pour un seuil égal au quantile 70% des débits observés (Figure 80 et Figure 81), pour les modèles *MORDOR* et *GRP,* respectivement, et les trois bassins versants ici utilisés comme exemple d'illustration. Nous remarquons que les résultats sont, en général, similaires pour les prévisions des modèles *GRP* et *MORDOR* :

- aux horizons courts de prévision, presque tous les membres prévoient le dépassement du quantile 70% avec une probabilité de détection élevée (supérieure à 60%) et un faible taux de fausses alertes (inférieur à 20%),

- aux horizons plus lointains, les résultats des deux modèles montrent une réduction de l'aire sous la courbe ROC et, par conséquent, une capacité des prévisions à discriminer entre deux événements dégradée. Nous remarquons néanmoins que la capacité du modèle *GRP* à discriminer entre deux événements se détériore plus rapidemement avec l'échéance que celle du modèle *MORDOR*. Ce fait est bien illustré par le bassin de la Romanche à Chambons (Figure 80 et Figure 81). Une explication possible est liée à la forte sous dispersion des membres (Figure 77), ainsi qu'aux valeurs plus élevées de la racine carrée de l'erreur quadratique moyenne *RMSE* des prévisions (Figure 79) pour le modèle *GRP* par rapport au modèle *MORDOR*. Ces conclusions se confirment sur l'ensemble des 11 bassins versants.

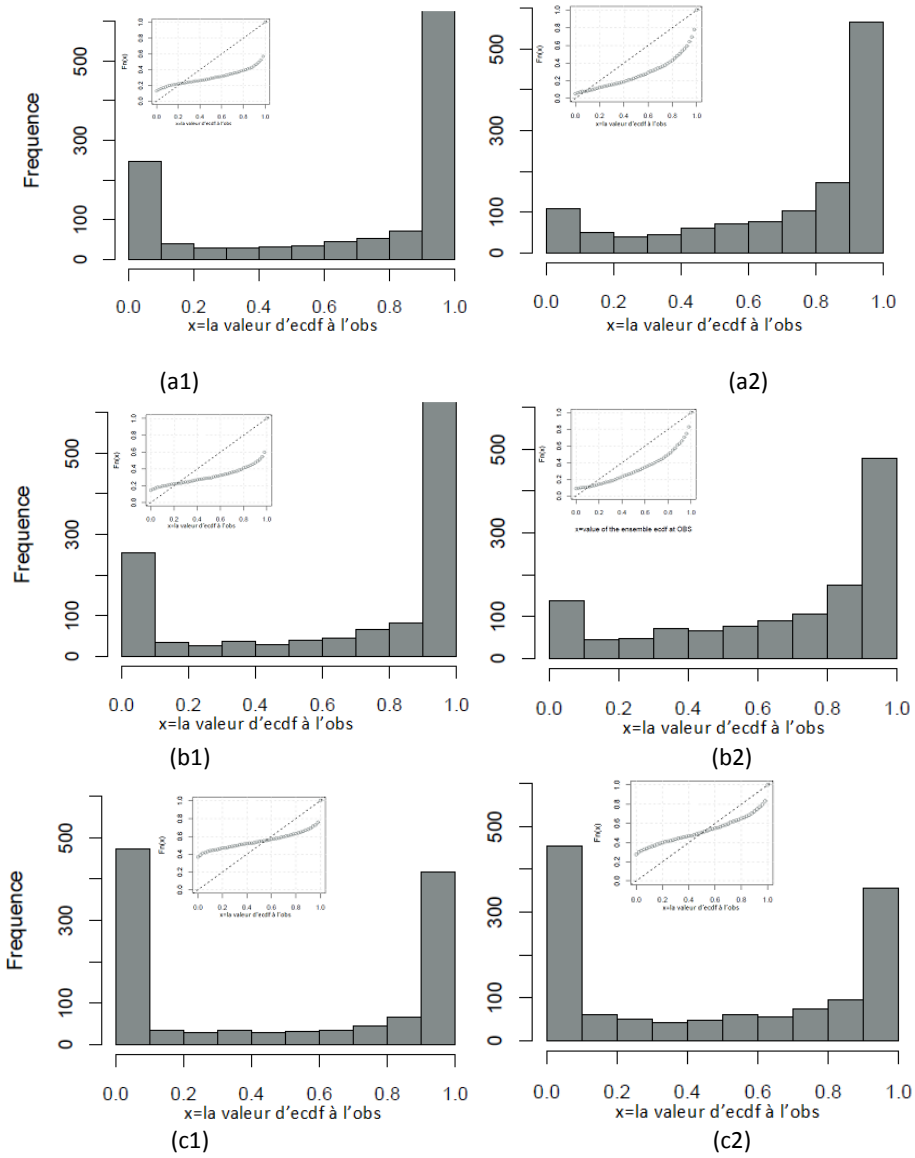


Figure 75 : Histogrammes et diagrammes de PIT pour les prévisions d'ensemble de débit brutes du modèle MORDOR pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite). Prévisions d'ensemble de pluie du CEPMMT utilisées en entrée du modèle hydrologique.

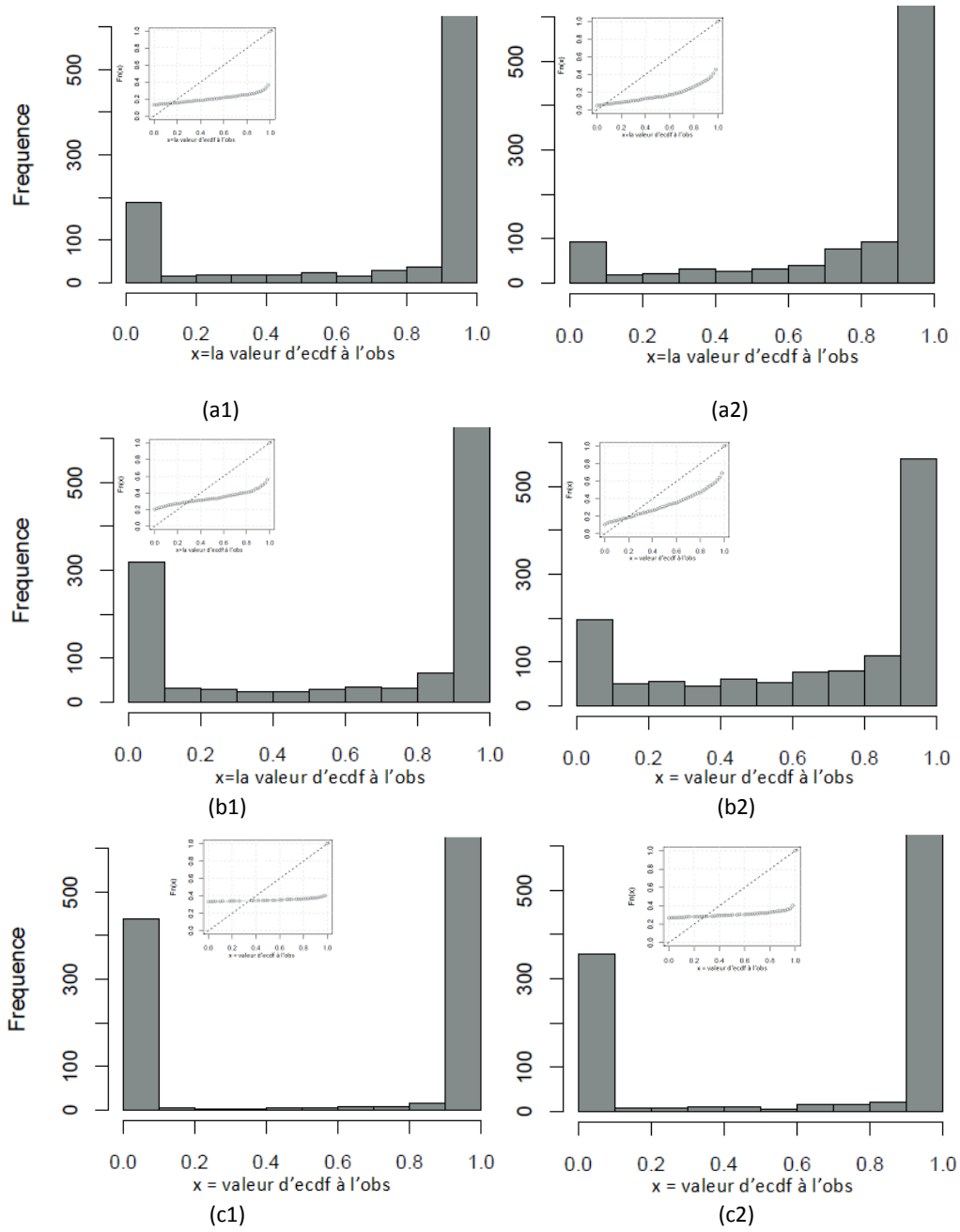


Figure 76 : Histogrammes et diagrammes de PIT pour les prévisions d'ensemble de débit brutes du modèle GRP pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite). Prévisions d'ensemble de pluie du CEPMMT utilisées en entrée du modèle hydrologique.

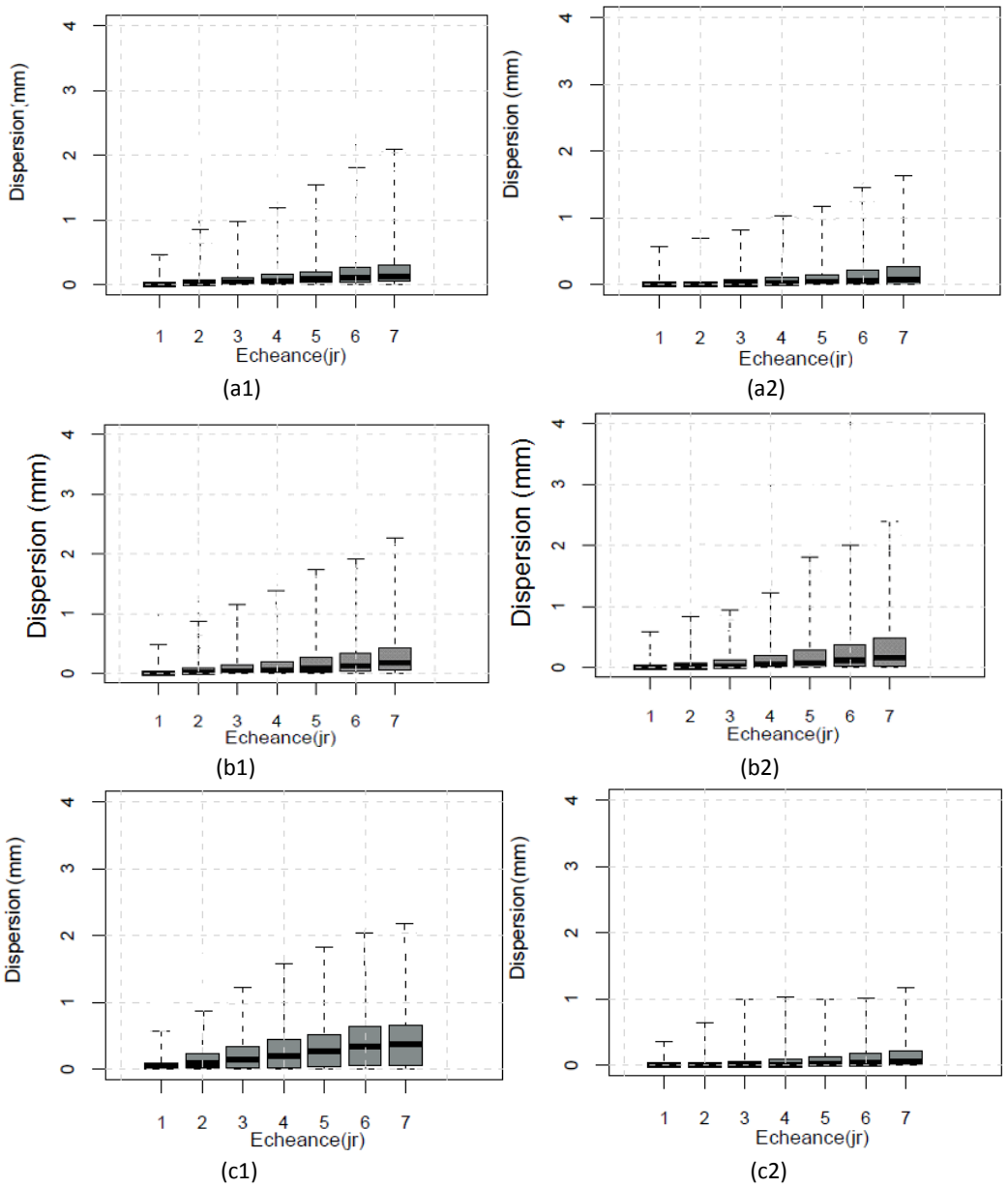


Figure 77 : Evolution de la dispersion des membres de la prévision hydrologique d'ensemble fournies par les modèles MORDOR (1) et GRP (2) de bassin brutes en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noire dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

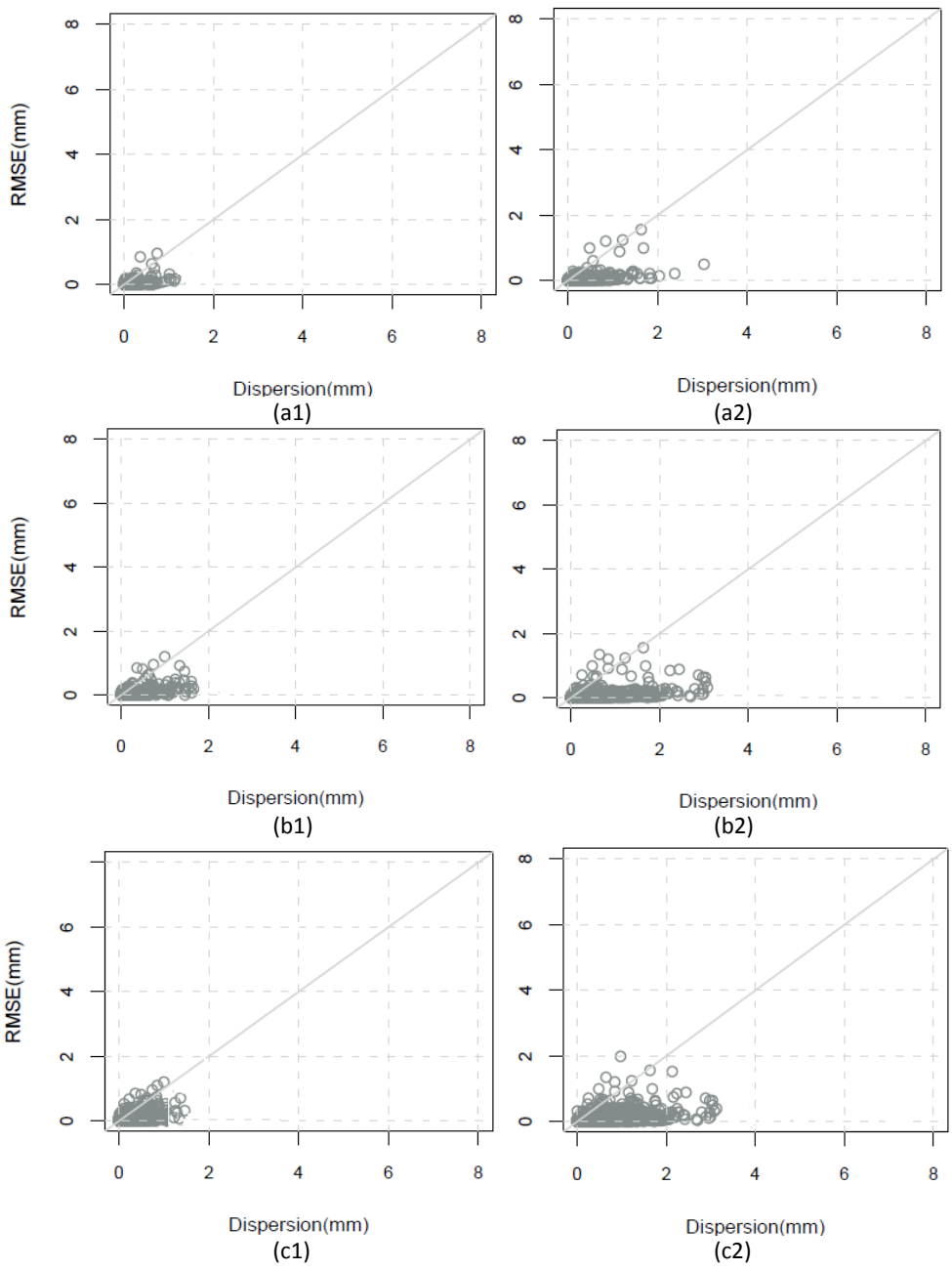


Figure 78 : La racine carrée de l'erreur quadratique moyen (RMSE) en fonction de la dispersion de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours.

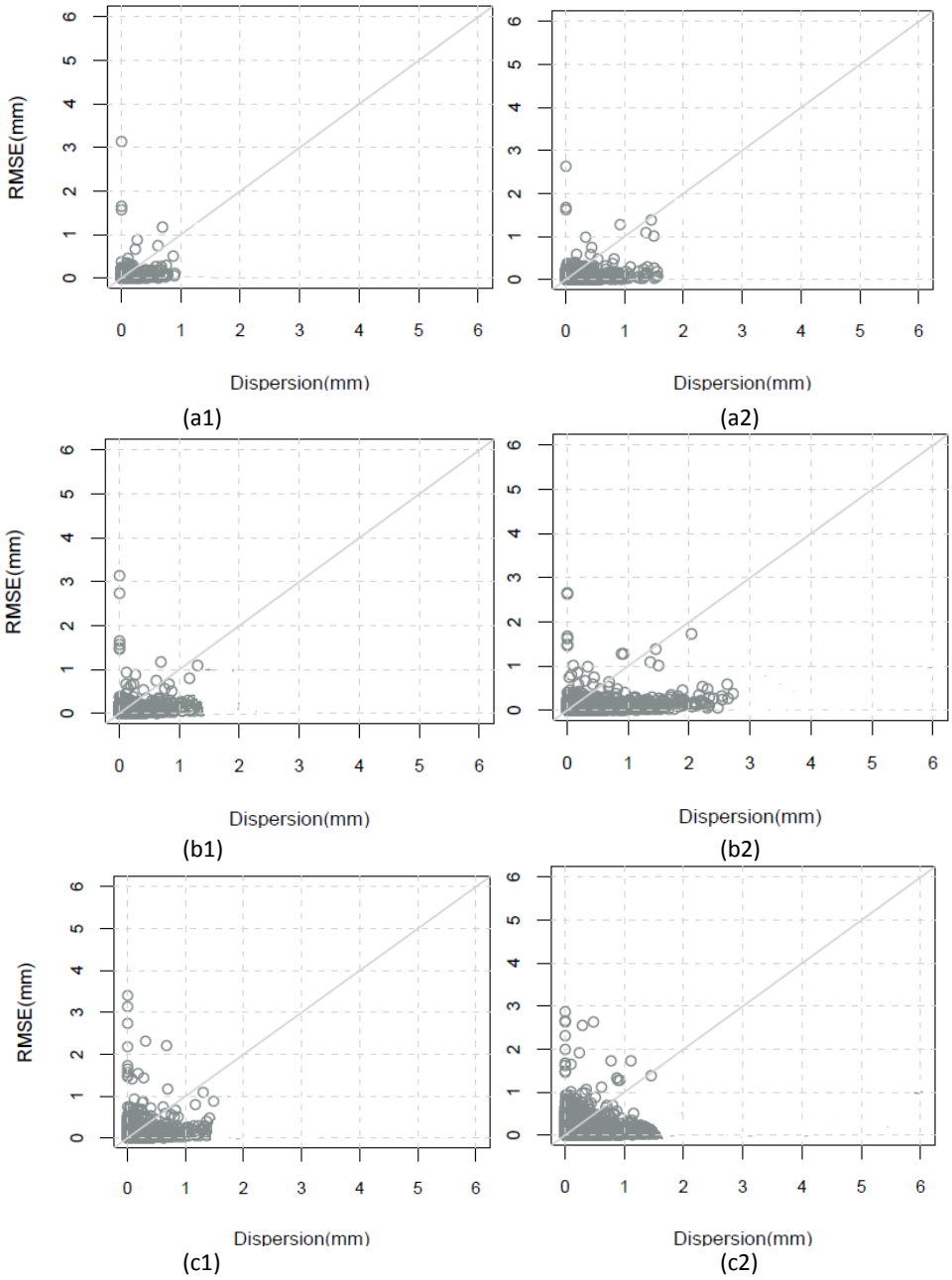


Figure 79 : Erreur quadratique moyen (RMSE) en fonction de la dispersion de la prévision d'ensemble des débits brutes fournies par le modèle GRP pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours.

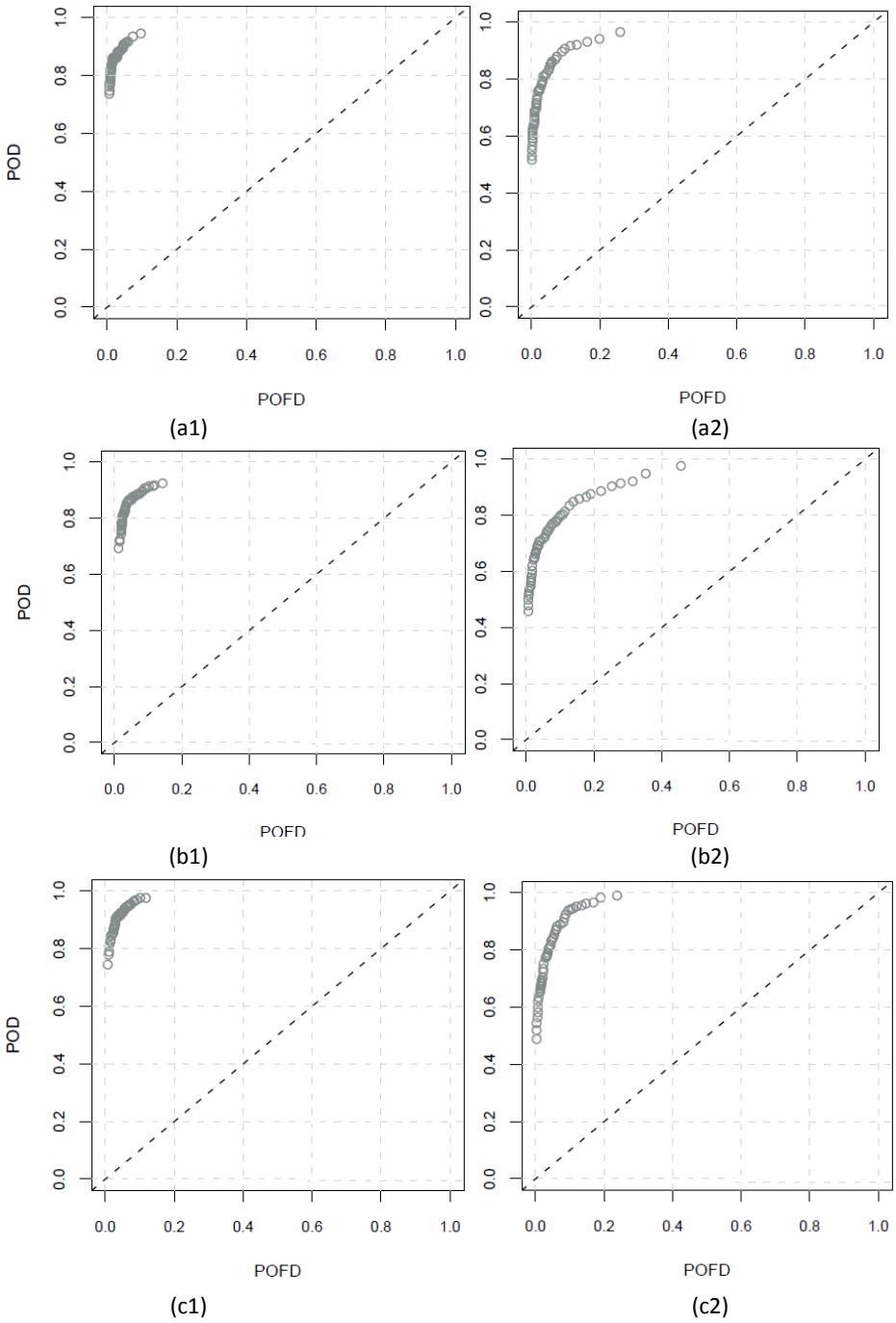


Figure 80 : Courbes ROC de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite).

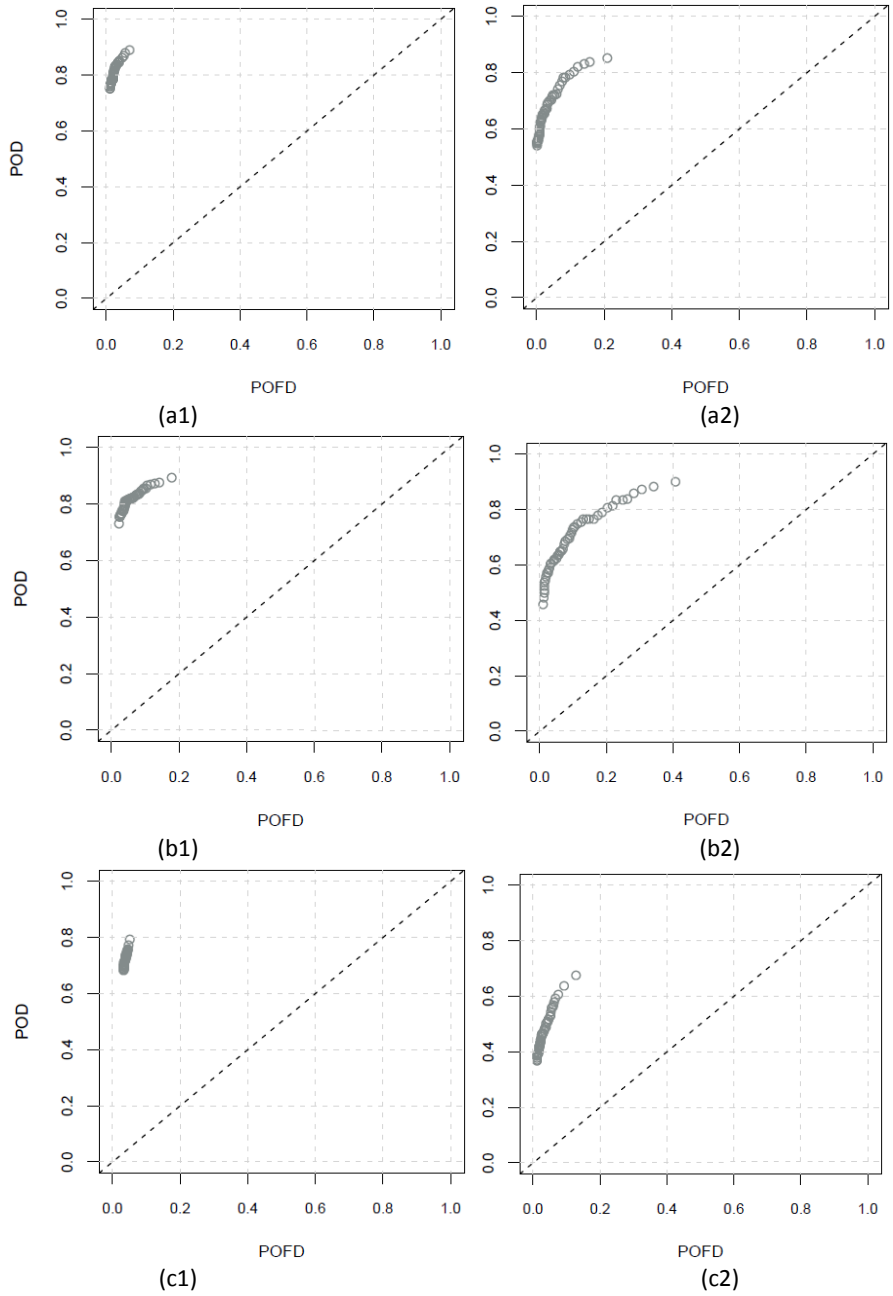


Figure 81 : Courbes ROC de la prévision d'ensemble des débits brutes fournies par le modèle GRP pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite).

Qualité globale

La qualité globale des prévisions hydrologiques d'ensemble brutes des modèles *GRP* et *MORDOR* est présentée à travers le score *CRPS* moyen dans les Figure 82 et Figure 83 en fonction de la surface (km²) des 11 bassins versants de notre échantillon d'étude.

Nous remarquons qu'il y a une tendance similaire à celle observée pour les prévisions brutes de précipitation : les valeurs plus élevées (moins bonne qualité des prévisions) sont associées aux bassins plus petits. Par rapport à l'échéance de prévision, nous remarquons une légère augmentation du score avec l'augmentation de l'échéance.

En comparant les performances des deux modèles nous constatons que les prévisions hydrologiques tirées du modèle *GRP* se montrent moins performantes par rapport aux prévisions du modèle *MORDOR*, car elles obtiennent des valeurs de *CRPS* plus élevées. Ce fait est plus marquant pour certains bassins versants, comme par exemple pour les bassins de la Borne au Pont de Nicoulaud (62 km²) et de la Dordogne à Bort (1005 km²) : pour ces bassins, les valeurs moyennes du score *CRPS* varient respectivement de 0.85 et 0.6 pour le modèle *MORDOR* à 1.1 et 0.95 pour le modèle *GRP* pour l'échéance de 7 jours de prévision.

En général nous remarquons que les prévisions hydrologiques fournies par les deux modèles se montrent fortement sous-dispersives et avec une tendance à sous-estimer l'observation. Ce fait démontre la nécessité d'appliquer un traitement statistique.

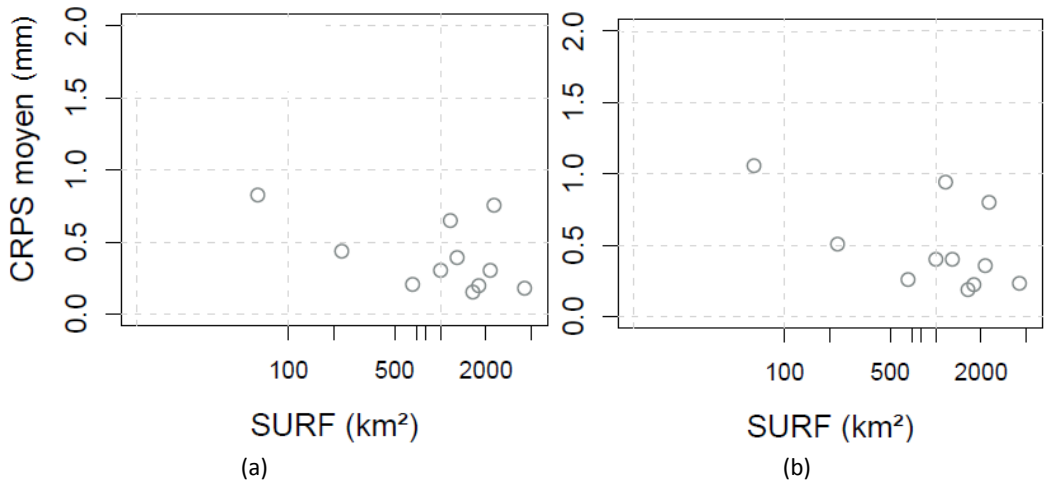


Figure 82 : CRPS moyens de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours.

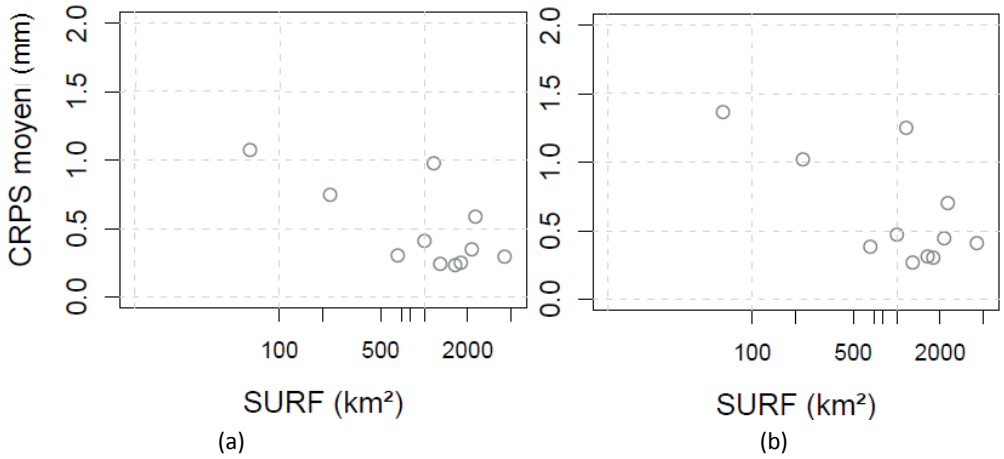


Figure 83 : CRPS moyens de la prévision d'ensemble des débits brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours.

6.4 Prétraitement des prévisions de pluie

L'analyse des prévisions brutes de pluies et de débits a montré que certains biais sont présents et pourraient être corrigés. Nous examinons ici l'apport des différents scénarios de traitement statistique définis dans le paragraphe §6.2, appliqués soit sur les prévisions de pluie, soit sur les prévisions de débits, soit sur les deux de manière combinée.

6.4.1 Evaluation de la qualité de l'habillage des prévisions de pluie

Dans notre étude, le traitement statistique des prévisions d'ensemble de pluie concerne les scénarios de prétraitement suivants (Tableau 8) :

- correction par l'application d'une méthode empirique, utilisant un facteur de correction (scénario 2A) ;
- habillage des prévisions à l'aide l'information issue des situations analogues (scénario 2B).

En utilisant les critères de diagnostic de la qualité des prévisions adoptés dans le cadre de cette thèse (Chapitre 3), nous évaluons l'apport de ces deux approches à la qualité des prévisions de pluie. Nous comparons ainsi l'apport de ces deux méthodes à l'amélioration de la qualité des prévisions brutes (sans aucun prétraitement). Ici également notre analyse porte sur les 11 bassins versants d'intérêt pour l'hydroélectricité et nous utilisons les trois bassins versants séléctionnés dans ce chapitre à titre d'exemple pour les critères graphiques : la Durance à Serre Ponçon, le Verdon à Castillon et la Romanche à Chambons.

Nous avons analysé l'impact du prétraitement sur les critères suivants : les diagrammes de PIT (Figure 84), la dispersion des membres (Figure 85), les courbes ROC (Figure 86) et le score de performance globale *CRPS* (Figure 87).

Fiabilité : diagramme de probabilité cumulée PIT

Par l'analyse comparée des performances, nous observons que :

- la correction multiplicative arrive à améliorer la fiabilité des prévisions, dans les bassins versants où les prévisions d'ensemble montrent une tendance à sous-estimer l'observation (la Durance à Serre Ponçon et le Verdon à Castillon), en approchant les courbes obtenues avec les prévisions corrigées de la diagonale. Par contre, dans certains bassins de notre échantillon, comme c'est le cas de la Romanche aux Chambons, la correction multiplicative n'est pas autant efficace pour améliorer la fiabilité des prévisions.

- la correction à l'aide des analogues ne permet pas une amélioration similaire de la fiabilité. Dans notre échantillon, ceci est le cas pour 6 bassins versants. De plus, dans les bassins versants qui démontrent une légère tendance soit à sous-estimer, soit à surestimer l'observation (5 bassins de notre échantillon, comme l'Ain à Vouglans et la Truyère à Grandval), l'approche des analogues arrive à améliorer considérablement la fiabilité des prévisions.

Enfin, nous remarquons que l'effet de la correction multiplicative sur l'amélioration de la fiabilité est considérablement plus fort aux horizons plus courts de prévision. Ceci est expliqué par le fait que le facteur de correction est calé pour la première échéance de prévision. En ce qui concerne l'habillage des prévisions de précipitation avec l'information des analogues, nous remarquons également que la correction est en général plus efficace pour les horizons de prévisions courts.

Dispersion des membres

Nous remarquons que :

- la correction multiplicative (scénario 2A) parvient à augmenter la dispersion des membres de l'ensemble pour les bassins versants de la Durance à Serre Ponçon et du Verdon à Castillon et pour les premières échéances de prévision (jusqu'à 4 jours). Ce fait justifie en partie l'amélioration de la fiabilité des prévisions observée dans la Figure 84. Par contre dans le cas de la Romanche aux Chambons, cette approche de correction n'est pas efficace et par conséquent elle conduit également à la faible amélioration de la fiabilité remarquée dans la Figure 84.

- la correction avec l'information des analogues (scénario 2B) contribue à une dispersion de l'ensemble trop élevée dès les premières échéances dans les bassins versants de la Durance à Serre Ponçon et du Verdon à Castillon (Figure 85). Cette dispersion non appropriée explique que la correction avec l'information des analogues est peu efficace sur la fiabilité des prévisions (Figure 84). En ce qui concerne le cas de la Romanche aux Chambons, la correction avec les analogues contribue à une dispersion moins importante. Des résultats similaires se retrouvent pour 5 autres bassins versants dans notre échantillon (comme l'Ain à Vouglans et la Tryuère à Grandval). Dans ces cas la méthode des analogues arrive à améliorer la fiabilité de prévision, car elle introduit une dispersion plus appropriée.

Discrimination : courbes ROC

Dans la Figure 86, nous observons que :

- les courbes ROC des prévisions de pluie corrigées comparées aux courbes obtenues avec les prévisions brutes, montrent que les approches de correction améliorent considérablement la capacité de la prévision à discriminer entre des événements lors des premières échéances de prévision. En général, l'amélioration est toujours observée au-delà d'un horizon de 3-4 jours, mais elle est moins importante. A ces horizons plus lointains, la correction multiplicative (2A) montre, en général, une meilleure performance que l'approche de correction avec l'information des analogues (2B). L'amélioration apportée par les approches de correction devient néanmoins moins importante quand le seuil de discrimination passe du quantile 70% au quantile 90%. Les conclusiosn tirées des courbes ROC sont similaires sur l'ensemble de 11 bassins versants examinés.

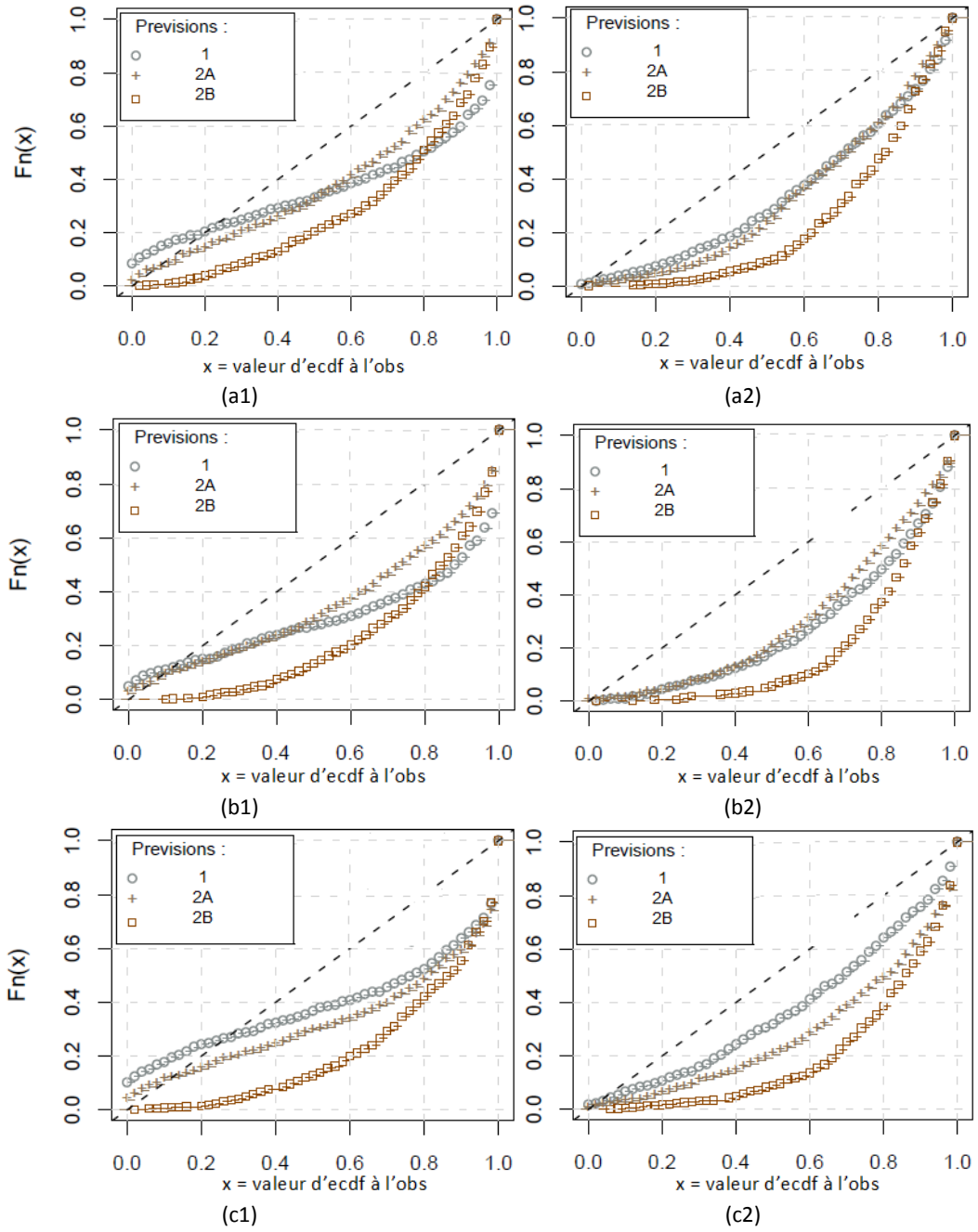


Figure 84 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de pluie pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiques sont calculés à partir des pluies du centre européen CEPMMT brutes (1) et corrigées en utilisant un facteur multiplicatif (2A) et la méthode des analogues (2B).

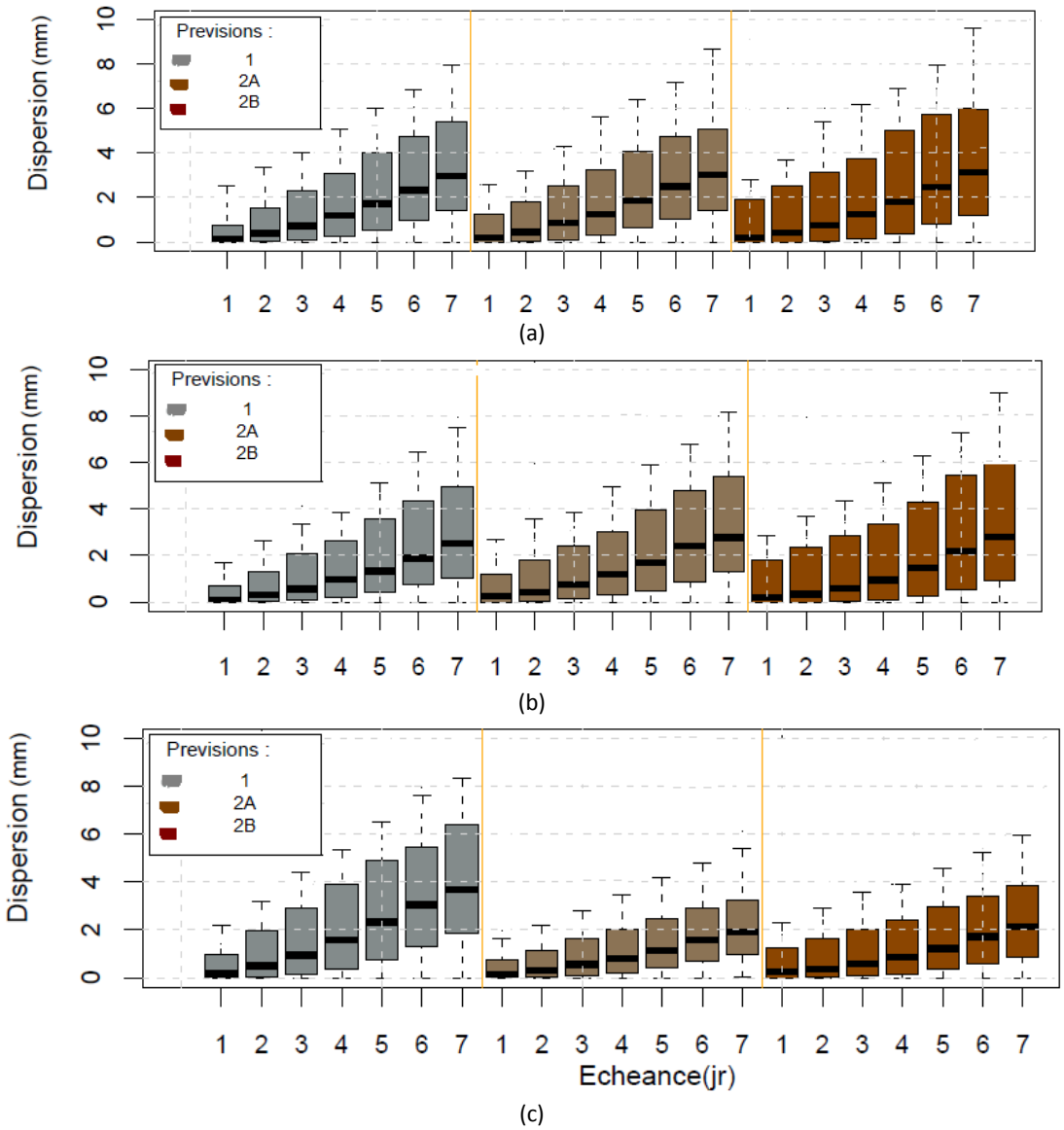


Figure 85 : Evolution de la dispersion des membres de la prévision d'ensemble de précipitations : brute (1), biaisée avec la correction multiplicative (2A) et corrigée avec l'information des analogues (2B) en fonction de l'échéance de prévision pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

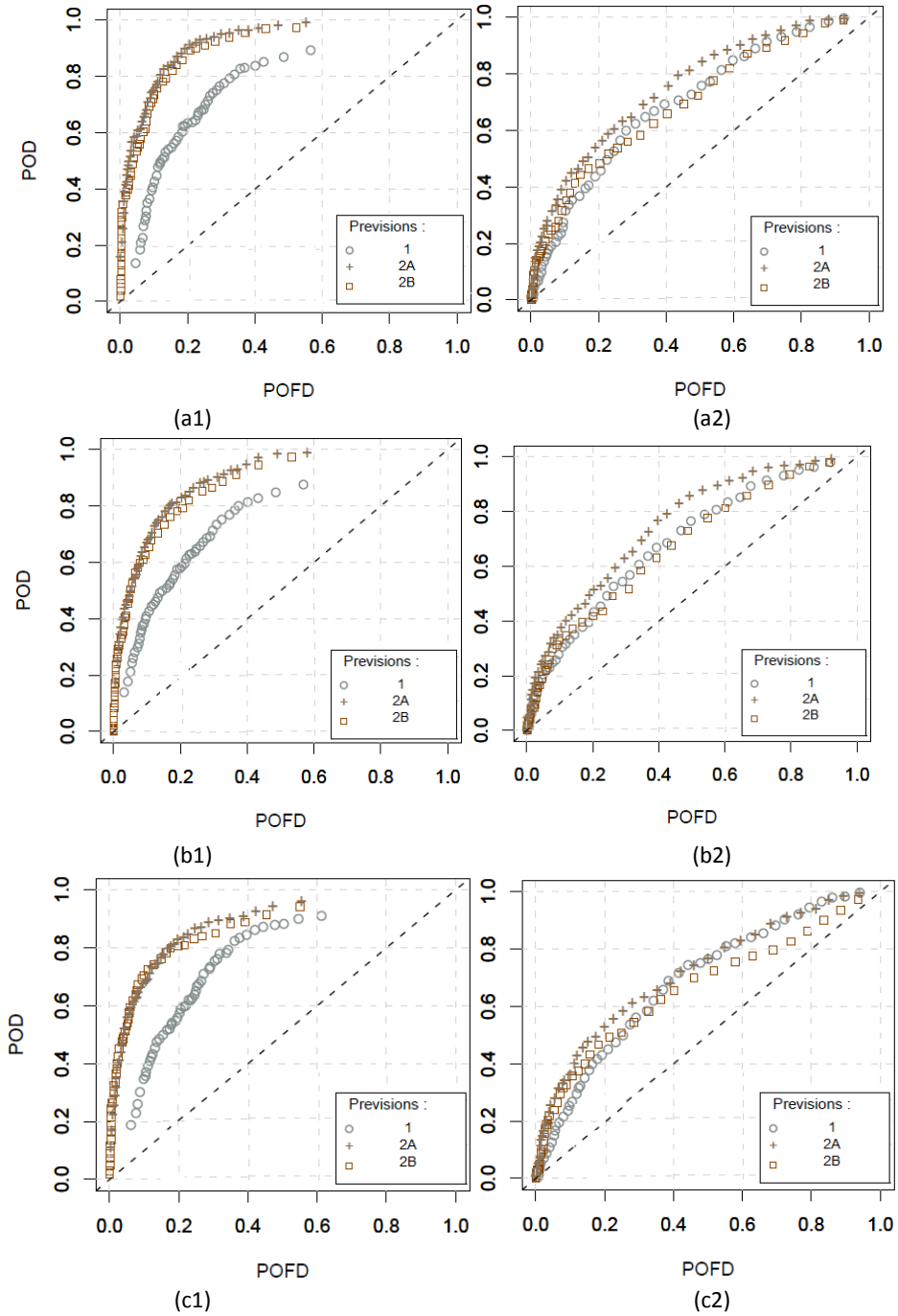


Figure 86 : Courbes ROC de la prévision d'ensemble des précipitations pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). Les courbes ROC sont calculées à partir des prévisions de pluies du centre européen brutes (1), corrigées en utilisant un facteur multiplicatif (2A) et l'information des analogues (2B).

Qualité globale

Les valeurs moyennes du score numérique *CRPS* sont examinées dans la Figure 87 en fonction du la surface du bassin versant pour les échéances de 3 jours (à gauche) et de 7 jours (à droite). Nous remarquons une nette amélioration des valeurs moyennes du *CRPS* avec l'application du traitement statistique de correction de biais, que ce soit avec la correction empirique par facteur multiplicatif (2A) ou avec l'habillage à l'aide des analogues (2B), pour tous les bassins versants étudiés. Cette amélioration est plus importante pour les horizons courts de prévision et dans le cas des petits bassins versants, où les prévisions d'ensemble brutes sont moins performantes.

De plus, nous observons que l'habillage avec l'information des analogues apporte une amélioration supplémentaire à la qualité des prévisions d'ensemble brutes pour la plupart de bassins versants, notamment lors des courtes échéances des prévisions. Cependant sur les horizons de prévision plus longs (7 jours) la performance d'habillage avec l'information des analogues est similaire à celle de prévisions corrigées avec un facteur multiplicatif. Les différences sont néanmoins peu significatives et peuvent être liées à des effets d'échantillonage, par exemple.

En résumé, nous concluons qu'en termes de qualité globale et de discrimination les deux approches de prétraitement examinées apportent des améliorations considérables surtout pour les premières échéances de prévision. En termes de fiabilité, nous constatons que les performances de ces deux approches dépendent fortement du bassin versant examiné.

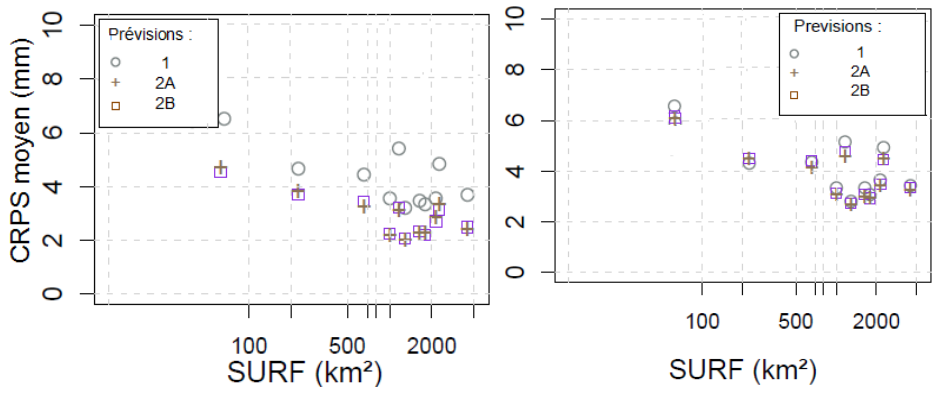


Figure 87 : CRPS moyens de la prévision d'ensemble des précipitations de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions de pluies du centre européen brutes (1) et corrigées en utilisant un facteur multiplicatif (2A) et l'information des analogues (2B).

6.4.2 Quel est l'effet du traitement des pluies sur les débits ?

Dans ce paragraphe nous évaluons l'effet au niveau des prévisions de débits d'une correction statistique dans les prévisions d'ensemble de pluie. Plus particulièrement, nous examinons le "résidu" de l'habillage de prévisions d'ensemble de pluie avec l'information des analogues en aval du modèle hydrologique. De l'analyse des pluies corrigées, nous retiendrons le scénario 2B (habillage des pluies par la méthode des analogues) pour le prétraitement des entrées de pluie aux modèles hydrologiques. Etant donné que les deux approches de prétraitement (2A et 2B) ont des performances similaires, nous avons retenu l'habillage avec l'aide des analogues (2B), car il s'agit d'une approche très répandue parmi les opérationnels. Pour cette application le modèle hydrologique *MORDOR* est utilisé. Les critères PIT, dispersion, courbes ROC, score CRPS et score RPS sont montrés dans les Figure 88, Figure 89 , Figure 90, Figure 91et Figure 92, respectivement.

Fiabilité : diagrammes de PIT

Nous remarquons que l'habillage des prévisions de pluie avec l'information des analogues n'apporte pas une nette amélioration à la fiabilité des prévisions hydrologiques. Celles-ci restent marquées par une sous-dispersion des membres. Ceci est observé sur tous les bassins versants examinés et pour toutes les échéances de prévision. De plus, nous constatons que, pour les deux scénarios étudiés (1 et 2B), les prévisions hydrologiques ont tendance à surestimer l'observation aux basses fréquences et à la sous-estimer aux hautes fréquences.

Dispersion des membres

En complément du diagnostic de fiabilité présenté ci-dessus, nous examinons la dispersion des membres dans la Figure 89. Nous constatons que l'habillage des prévisions de pluie avec l'information des analogues contribue à une légère augmentation de la dispersion des membres des prévisions hydrologiques. Cet effet est plus apparent aux échéances plus lointaines de prévision. Sur l'ensemble des 11 bassins versants et pour un horizon de prévision égal à 7 jours, nous remarquons qu'en habillant les prévisions hydrologiques avec l'information des analogues la dispersion des membres augmente entre 10 et 30% selon le bassin.

Discrimination : courbes ROC

Les courbes ROC (Figure 90) montrent que la correction des pluies n'influence pas la capacité des prévisions de débits à discriminer deux événements entre eux. Ce comportement se retrouve dans tous les bassins versants étudiés et il se répète sur tous les horizons de prévision. En accord avec les résultats sur les prévisions de précipitation, la discrimination

des prévisions hydrologiques varie également en fonction de l'échéance de prévision et le seuil de non-dépassement examiné.

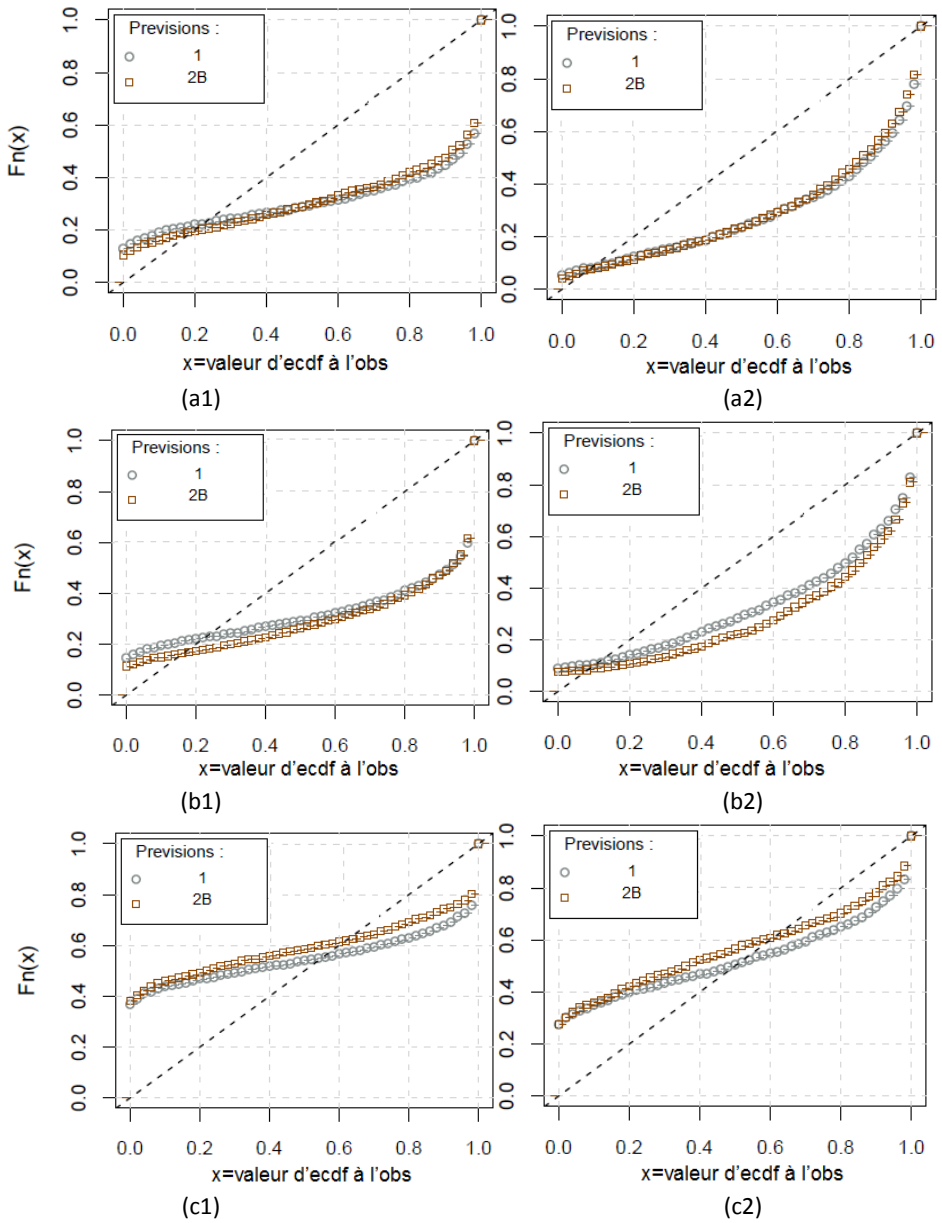


Figure 88 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de débit pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiques sont calculés à partir des pluies du centre européen CEPMMT brutes (1) et corrigées en utilisant la méthode des analogues (2B).

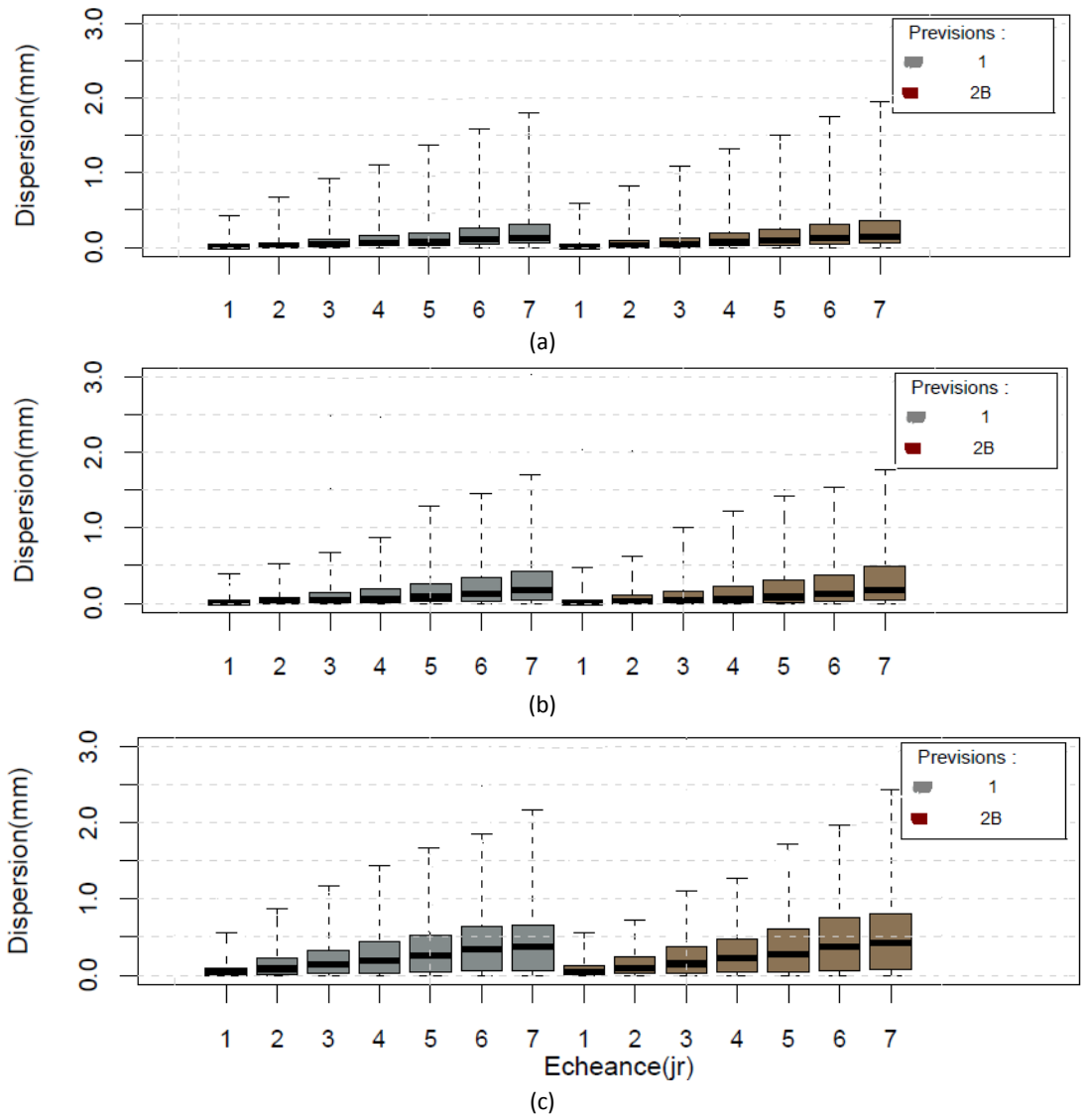


Figure 89 : Evolution de la dispersion des membres de la prévision d'ensemble de débits : brutes (1) et à partir des prévisions de pluie corrigées avec l'information des analogues (2B) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

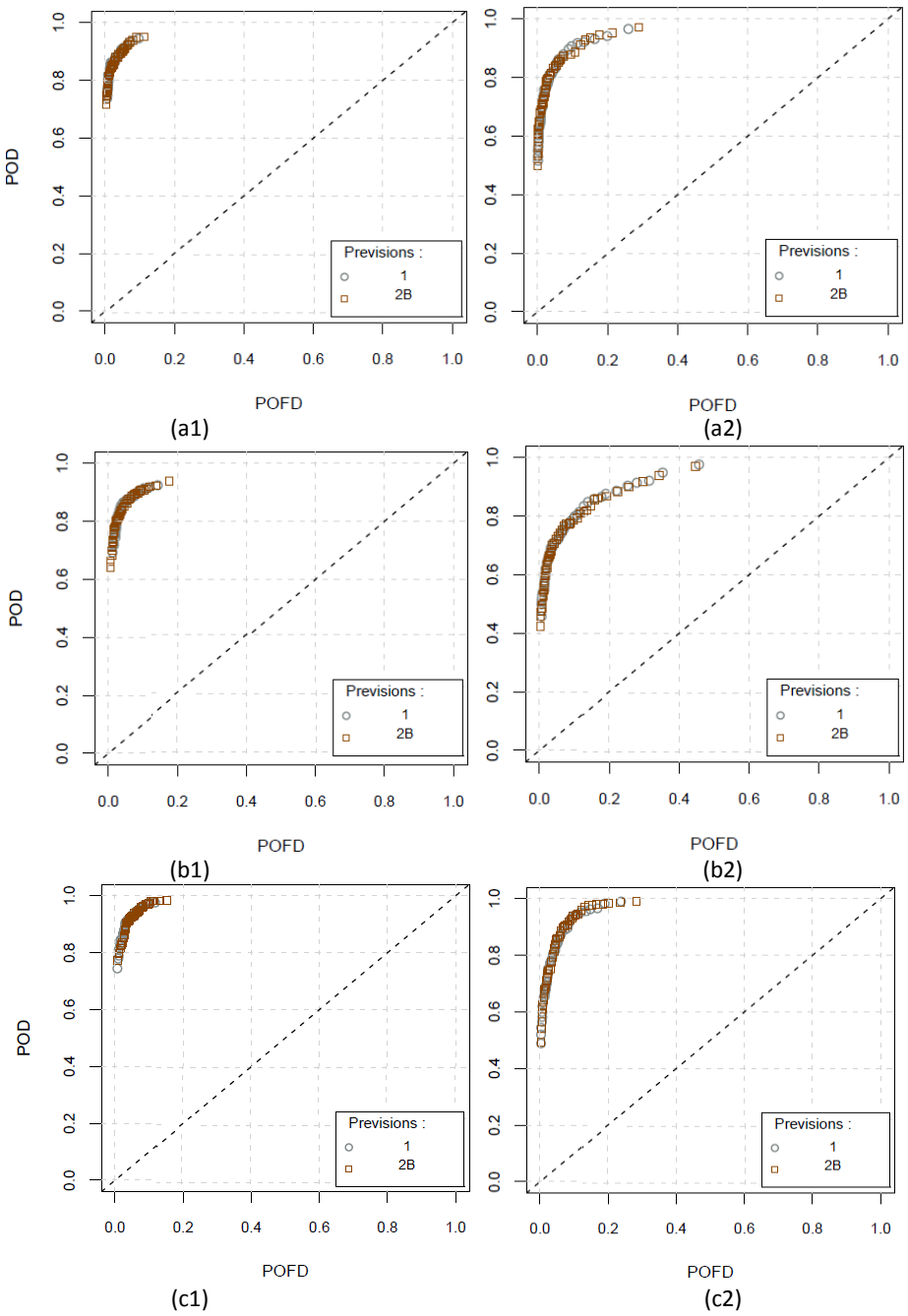


Figure 90 : Courbes ROC de la prévision hydrologiques d'ensemble pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). Les courbes ROC sont calculées à partir des prévisions de pluies du centre européen brutes (1) et l'information des analogues (2B).

Qualité globale

La Figure 91 montre les valeurs moyennes du *CRPS* des bassins versants en fonction de la surface des bassins pour les prévisions de débits brutes et pour les prévisions de débits obtenues après habillage des prévisions des pluies. Cette figure montre également que l'amélioration apportée aux prévisions hydrologiques par l'habillage des pluies avec l'information des analogues est très limitée.

Ces résultats sont similaires pour toutes les échéances étudiées et pour tous les bassins versants de notre échantillon. Nous observons que - similairement à la tendance observée pour les prévisions hydrologiques brutes - les valeurs plus élevées sont en général associées aux bassins plus petits ; ce qui signifie une moins bonne qualité de prévisions.

Pour résumer l'ensemble des observations précédentes, nous constatons que, même si l'habillage avec l'information des analogues apporte une amélioration à la qualité globale de prévisions d'ensemble de pluie, au niveau des prévisions hydrologiques d'ensemble cet effet reste limité. Comme pour les pluies, l'amélioration se trouve surtout aux horizons courts. Une explication possible pour ceci est que le modèle hydrologique écrase les corrections apportées au niveau pluie et introduit de nouvelles erreurs dans les prévisions.

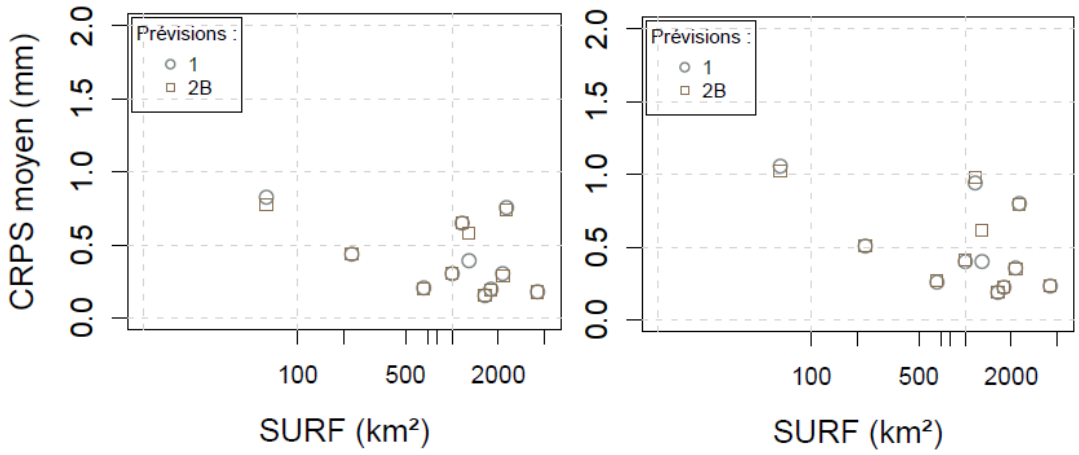


Figure 91 : CRPS moyens de la prévision d'ensemble des débits fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculés à partir des prévisions de pluies du centre européen brutes (1) et corrigées en utilisant l'information des analogues (2B).

Afin de visualiser l'effet de l'habillage des pluies avec l'information des analogues sur les débits sur chaque bassin versant de notre échantillon, nous calculons l'amélioration apportée selon le critère *RPS*. Cette amélioration est jugée à partir de la différence entre le score *RPS* correspondant aux prévisions hydrologiques brutes et celui des prévisions corrigées avec l'aide des analogues :

$$amelioration(ibv,t) = RPS_{brutes} - RPS_{corrigé} \qquad \text{Eq. 50}$$

Cette différence est calculée par bassin versant *ibv* et par échéance *t*. Rappelons que la valeur idéale du score *RPS* est égale à 0. Si la différence est positive, l'habillage avec l'information des analogues apporte une amélioration à la qualité des prévisions hydrologiques brutes. Si la différence est négative, les prévisions corrigées n'améliorent pas la qualité globale des prévisions de débit. Nous considérons qu'une différence supérieure à 20% donne une amélioration forte, tandis qu'une différence entre 0 et 9% apporte une amélioration faible.

Dans la Figure 92 nous présentons les résultats pour l'échéance à 3 (à gauche) et 7 jours (à droite) pour les prévisions hydrologiques d'ensemble émises par le modèle *MORDOR*. Nous remarquons que l'amélioration apportée par l'habillage avec l'information des analogues se concentre principalement aux premières échéances de prévision et elle reste limitée à un petit nombre de bassins versants. Cette amélioration est surtout observée sur les bassins versants des Alpes et du Massif Central pour des horizons courts de prévision et sur le Massif Central pour les horizons plus lointains de prévision.

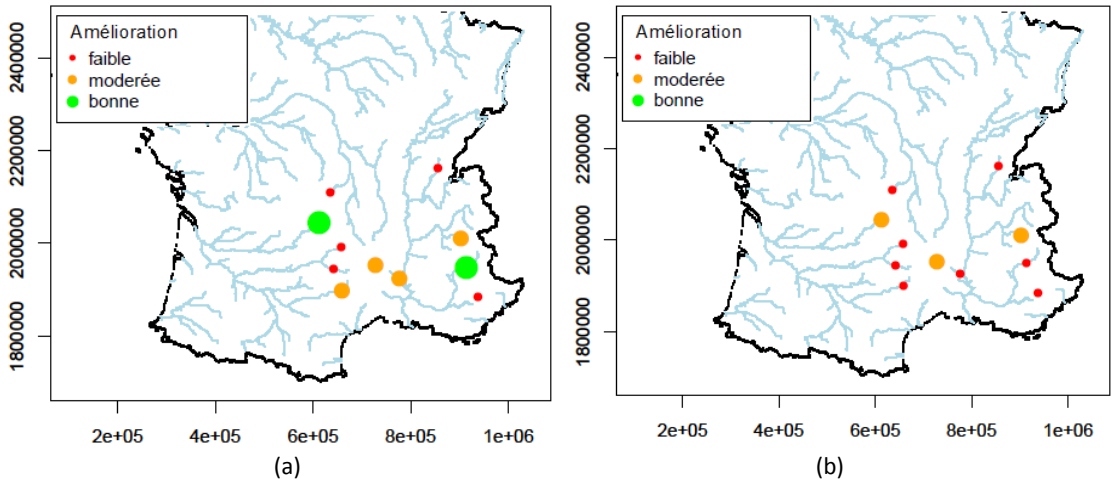


Figure 92 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS selon l'habillage avec l'information des analogues. Les graphiques sont calculés à partir des prévisions hydrologiques produites par le modèle MORDOR et à une échéance de prévisions égale à (a) 3 et (b) 7 jours.

En résumé, à propos des corrections des pluies brutes examinées, nous concluons que :

- en général les prévisions de précipitation ont un défaut de fiabilité (faible dispersion des membres) dans les premières échéances (prévisions sous-dispersées), avec une fiabilité améliorée avec l'augmentation de l'échéance,
- sur les 11 bassins, les deux approches de correction statistique examinées (la correction avec un facteur multiplicatif -2A- et l'habillage avec l'information des analogues - 2B) apportent une nette amélioration à la qualité globale des prévisions en termes de critère de *CRPS* pour les échéances courtes essentiellement,
- la correction multiplicative et l'habillage avec les analogues améliorent la fiabilité des prévisions de précipitation, mais cette amélioration dépend du bassin versant,
- en ce qui concerne la discrimination des prévisions, les deux approches démontrent une performance améliorée surtout pour les horizons courts de prévision.

En ce qui concerne les prévisions hydrologiques, nous constatons que, même si le traitement statistique des prévisions de pluie avec l'information des analogues apporte une amélioration à la qualité des prévisions d'ensemble de pluie, toutefois cette amélioration ne se répercute pas dans les prévisions des débits. L'application de l'habillage avec l'information des analogues n'améliore que légèrement la qualité globale des prévisions et surtout aux courtes échéances de prévision. La fiabilité des prévisions des débits reste insuffisante pour toutes les échéances étudiées.

Il semblerait que le passage par le modèle hydrologique affecte fortement la dispersion des membres de la prévision d'ensemble en la diminuant considérablement. Ceci peut aussi être un effet naturel du rôle intégrateur du bassin versant. C'est ce qui sera examiné dans les paragraphes suivants.

6.5 Post-traitement des prévisions

Dans cette partie nous explorons la performance de différentes approches de correction statistique sur les prévisions d'ensemble hydrologiques. Nous examinons la qualité des prévisions de débits traitées par les techniques de correction de biais que nous avons mis en place. Pour cela, nous procédons ici en trois étapes :

- d'abord, nous évaluons les scénarios de post-traitement qui visent à traiter l'incertitude du modèle hydrologique séparément de l'incertitude des prévisions de pluie (scénarios 3 et 4, Tableua 8, §6.2 : seulement post-traitement des débits, à partir des erreurs de simulation hydrologique) §6.5.1,
- ensuite, nous examinons les scénarios de post-traitement qui visent à traiter directement l'incertitude totale de prévision (scénarios 5 à 8, Tableau 8, §6.2 : seulement post-traitement des débits, à partir des erreurs de prévision hydrologique) §6.5.2.

6.5.1 Prise en compte de l'incertitude totale de modélisation

Nous examinons la méthode de post-traitement basée sur l'habillage avec les erreurs empiriques du modèle hydrologique (scénario 3). Les prévisions de précipitation brutes (sans prétraitement) sont ici considérées en entrée des modèles hydrologiques *MORDOR* et *GRP.*

Rappelons que, pour chaque bassin versant examiné, les erreurs du modèle hydrologique observées pendant une longue période historique de modélisation, appelée *période de calage* de la méthode d'habillage, sont modélisées statistiquement par une loi normale ajustée à l'échantillon des erreurs observées. Ensuite, pendant la période de prévision, des valeurs d'erreur sont tirées de cette loi et utilisées pour habiller les prévisions hydrologiques brutes. La modélisation des erreurs est faite par gamme (ou catégorie) de débit simulé. L'habillage des prévisions est fait pour chaque échéance et en fonction de la valeur du débit simulé en prévision.

Nous examinons les différents résultats des critères d'évaluation de la qualité des prévisions : les diagrammes de PIT (Figure 93), l'evolution de la dispersion (Figure 94 et Figure 95), les courbes ROC (Figure 96), le score CRPS (Figure 97 et Figure 98), cartes de l'amélioration du score RPS par rapport aux prévisions hydrologiques sans post-traitement (Figure 99).

Fiabilité : diagrammes de PIT

La Figure 93 présente le diagramme de probabilité cumulée "PIT" pour l'échéance de 7 jours et pour les 3 bassins versants utilisés à titre illustratif. Nous constatons que l'habillage avec les erreurs du modèle hydrologique améliore considérablement la fiabilité des prévisions : les points du diagramme de probabilité cumulée de PIT sont très proches de la diagonale. Cette amélioration est observée pour tous les bassins versants et sur toutes les échéances étudiées. Elle est considérablement plus marquante pour le modèle hydrologique MORDOR. Rappelons que les prévisions hydrologiques brutes du modèle *GRP* sont considérablement moins performantes que les prévisions du modèle *MORDOR*. L'approche de post-traitement des débits par habillage avec les erreurs empiriques de modélisation est donc aussi efficace pour le modèle *GRP*. Ces conclusions pour le modèle *GRP* sont valables pour la plupart de bassins versants examinés avec l'exception de la Romanche à Chambons, la Sioule à Fades et de l'Ardeche à Sauze.

Dispersion des membres

Nous remarquons également que cette approche de correction de biais contribue globalement à l'augmentation de la dispersion des membres, et ce pour les deux modèles (Figure 94 et Figure 95). Pour les cas de la Durance à Serre Ponçon et du Verdon à Castillon, où la performance de l'habillage avec l'erreur du modèle est très bonne, il s'agit d'une augmentation progressive (c'est le cas également pour les autres bassins versants de notre échantillon).

Par contre, dans le cas du modèle *GRP* et pour le bassin de la Romanche aux Chambons, nous observons plutôt une augmentation abrupte, qui conduit à une dispersion moins appropriée. Ce fait est également confirmé par la relation entre l'erreur quadratique moyenne et la dispersion (graphiques non présentés ici), qui montrent que pour les cas de bassins versants de la Durance à Serre Ponçon et du Verdon à Castillon il y a une corrélation faible entre la dispersion des membres et l'erreur quadratique moyenne, tandis que pour la Romanche aux Chambons il existe une forte corrélation.

Discrimination : courbes de ROC

La Figure 96 illustre les courbes ROC pour les prévisions hydrologiques brutes (1) et traitées (3) fournies par les modèles *MORDOR* (à gauche) et *GRP* (à droite) pour l'échéance de 7 jours. Nous remarquons que dans le cas du modèle *MORDOR* le post-traitement en prenant en compte les erreurs du modèle hydrologique n'améliore pas significativement la capacité des prévisions à discriminer entre les événements (ici le dépassement/non-dépassement du quantile 70% des débits observés). Cette conclusion est généralisée sur tous les bassins versants et échéances étudiés.

Par contre, pour le cas du modèle *GRP*, les prévisions habillées avec l'erreur du modèle hydrologique sont légèrement améliorées en termes de discrimination par rapport aux prévisions brutes. Cette différence entre les deux modèles peut être attribuée à la très faible dispersion des prévisions hydrologiques du modèle *GRP*.

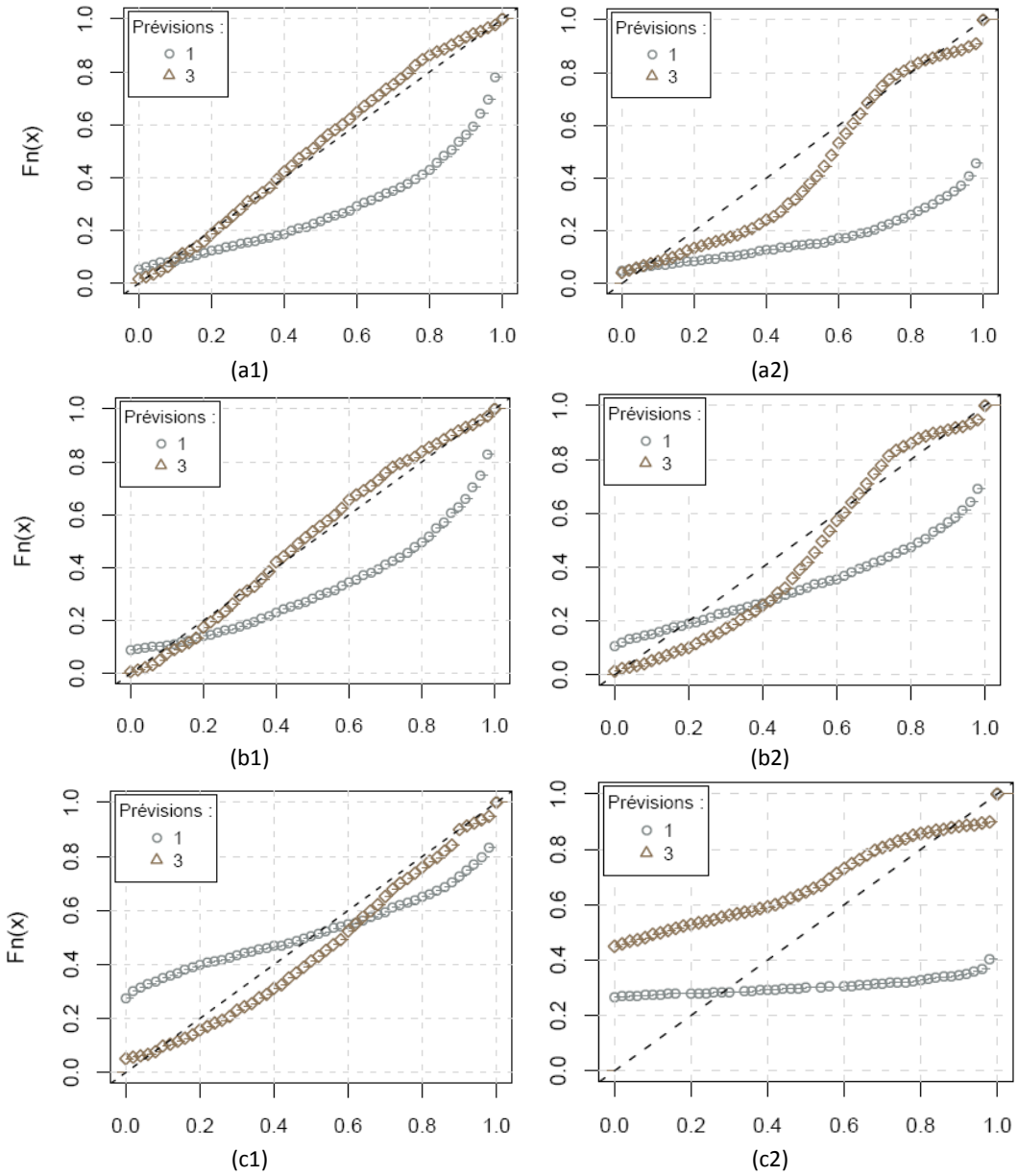


Figure 93 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de débits pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiquess sont calculés à partir des prévisions hydrologiques habillées avec l'erreur du modèle hydrologique (3) : MORDOR (1) et GRP (2).

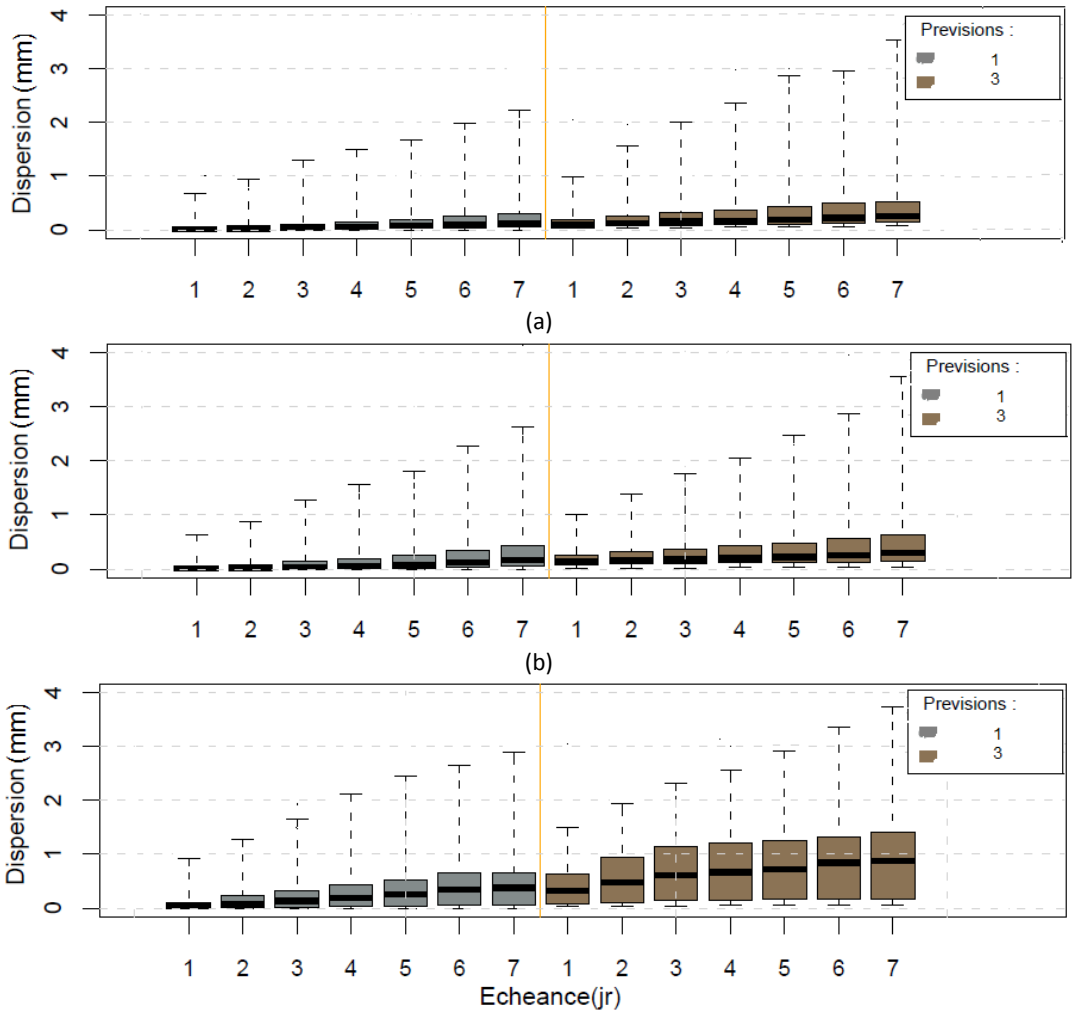


Figure 94 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec l'erreur du modèle hydrologique (3) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

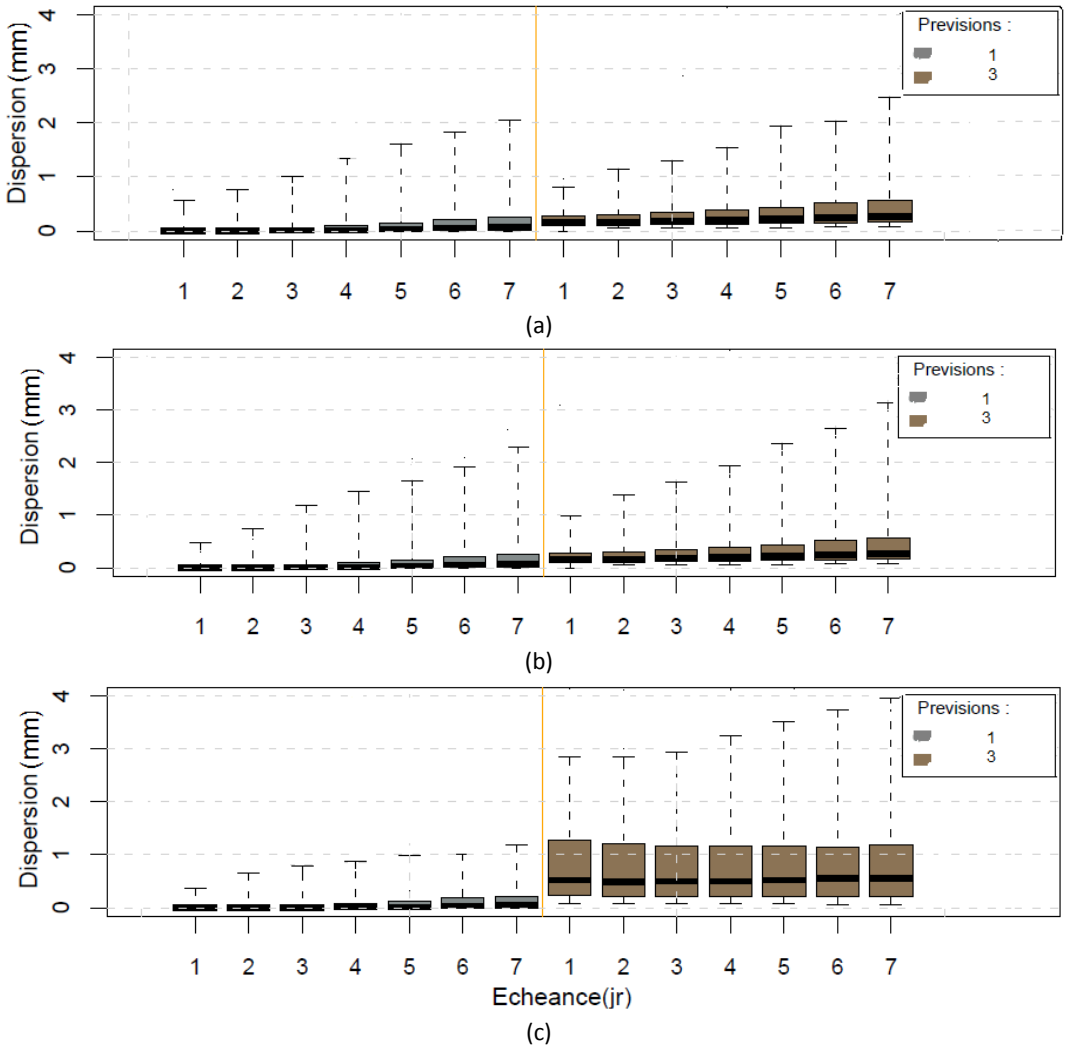


Figure 95 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec l'erreur du modèle hydrologique (3) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

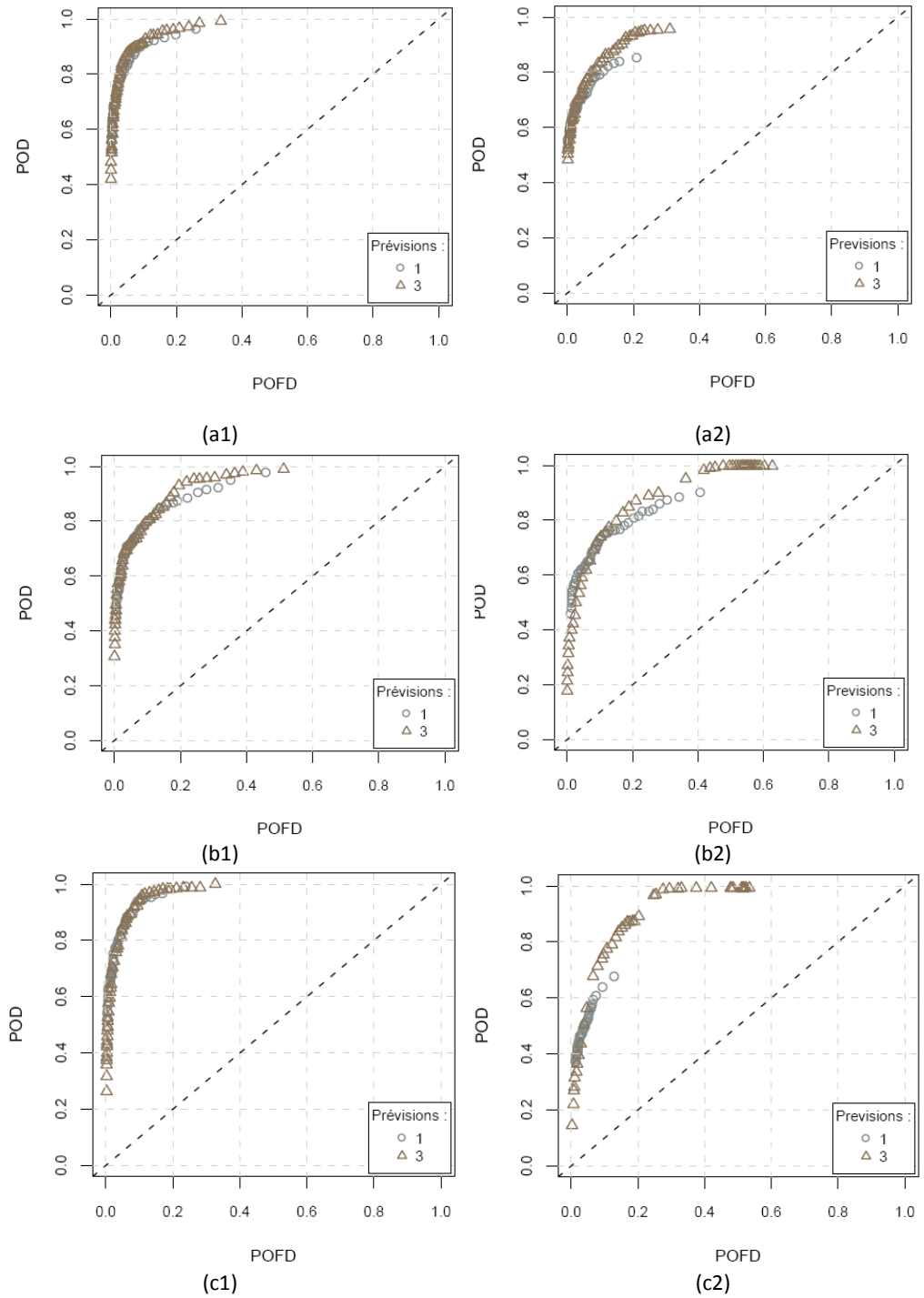


Figure 96 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle (3).

Qualité globale

La performance globale de l'habillage avec l'erreur du modèle est présentée dans les Figure ***97*** et Figure 98 en termes des valeurs moyennes du critère CRPS obtenues pour les échéances de 3 et 7 jours avec les prévisions du modèle *MORDOR* et *GRP*, respectivement.

Nous remarquons que la qualité globale des prévisions habillées est souvent pratiquement égale ou légèrement améliorée par rapport à celle des prévisions brutes ; l'amélioration étant plus particulièrement observée pour les bassins versants plus petits. Ceci est observé pour l'ensemble des échéances étudiées. En effet, les CRPS obtenus avec le modèle *GRP* (Figure 98) montrent que les performances de *GRP* sont, en général, moins bonnes que celles de *MORDOR* : pour un même bassin versant, la valeur du *CRPS* de *MORDOR* est plus petite que celle de *GRP*, autant pour les prévisions brutes que pour les prévisions post-traitées.

Enfin, les Figure 99 et Figure 100 présentent la cartographie de l'amélioration apportée par l'habillage avec l'erreur du modèle hydrologique *MORDOR* et *GRP* respectivement et pour les échéances de prévision 3 (à gauche) et 7 (à droite) jours. Nous remarquons que, selon le critère de *RPS* et pour les prévisions fournies par les deux modèles hydrologiques, en habillant avec l'erreur du modèle hydrologique nous arrivons à améliorer la qualité globale de prévisions d'ensemble pour la plupart de bassins versants étudiés. Cette amélioration est similaire pour toutes les horizons de prévisions et elle se retrouve surtout sur les bassins des Cévennes, du Juras et du Massif Central pour le modèle *MORDOR* et sur les bassins des Alpes du Sud et du Massif Central pour le modèle *GRP*. Paradoxalement, cette amélioration est moins importante pour les bassins alpins pour tous les deux modèles. De plus, l'amélioration apportée par la méthode est peu dépendante de l'échéance de prévision.

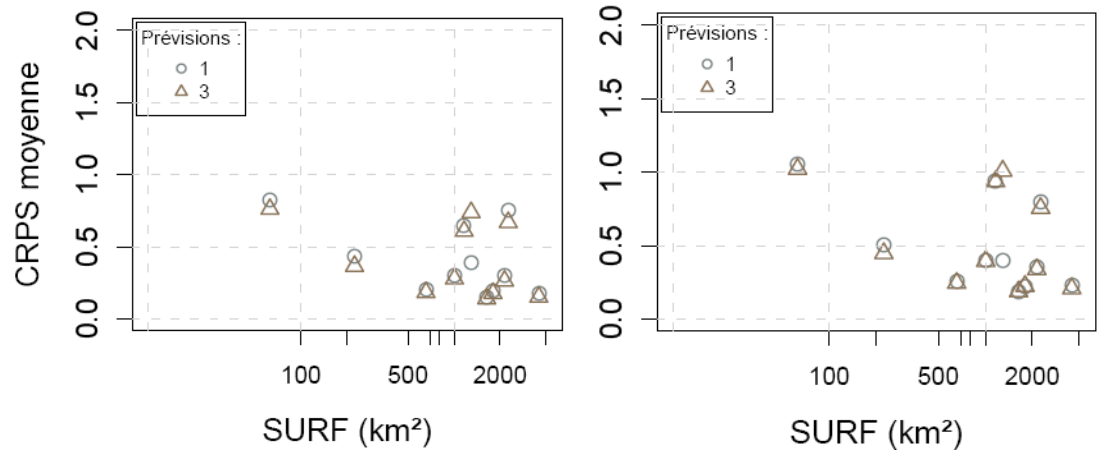


Figure 97 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle hydrologique (3).

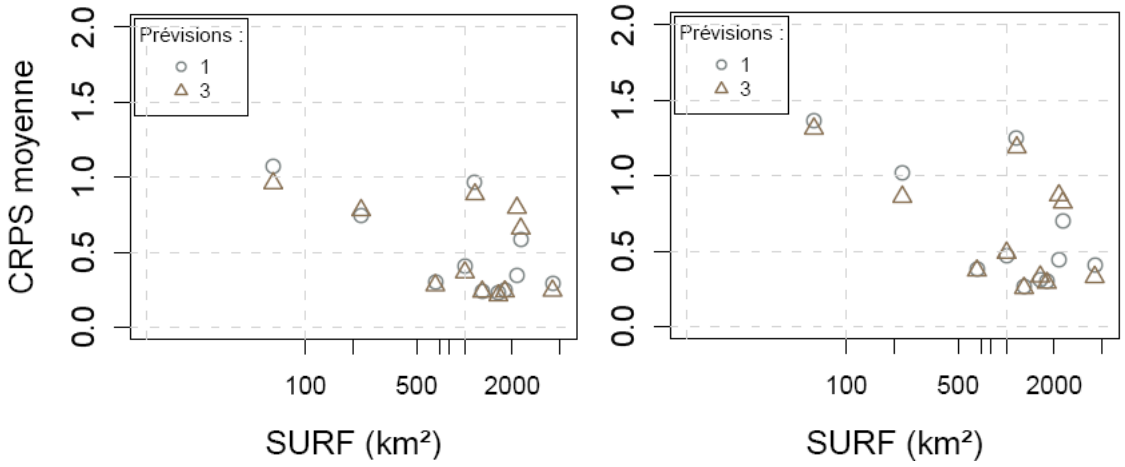


Figure 98 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle hydrologique (3).

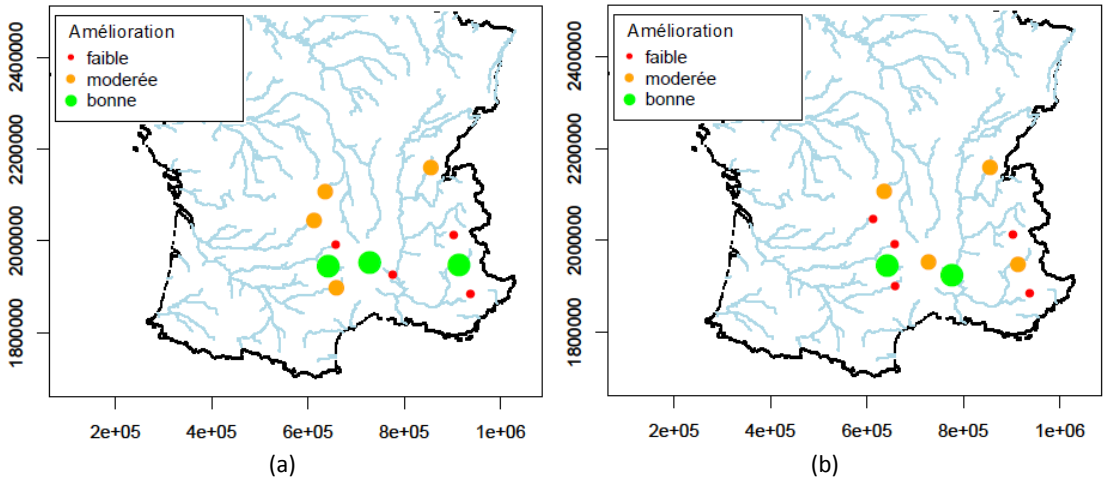


Figure 99 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques du modèle MORDOR par bassin versant en termes de score de RPS par l'habillage avec l'erreur du modèle. Les graphiques sont calculés à partir des prévisions hydrologiques à une échéance de prévisions égale à : 3 (à gauche) et 7 (à droite) jours.

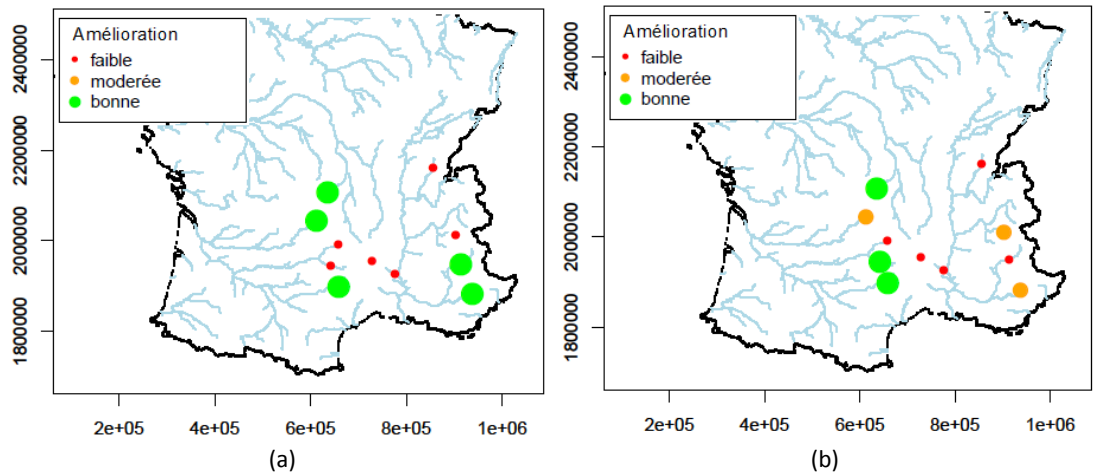


Figure 100 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques du modèle GRP par bassin versant en termes de score de RPS par l'habillage avec l'erreur du modèle. Les graphiques sont calculés à partir des prévisions hydrologiques à une échéance de prévisions égale à : 3 (à gauche) et 7 (à droite) jours.

6.5.2 Approche combinée : prétraitement et post-traitement

Une problématique qui se pose souvent surtout parmi les services opérationnels est celle du choix entre l'application d'un traitement de prévisions d'ensemble en amont (prétraitement) ou en aval (post-traitement) du modèle hydrologique. Afin de pouvoir répondre à cette question, nous devons examiner en parallèle les approches de prétraitement et les approches de post-traitement, avec également une approche combinatoire des deux. Les approches comparées ci-après sont : l'habillage seule des prévisions de pluie avec l'information des analogues (scénario 2), l'habillage seule des prévisions de débits avec l'erreur du modèle hydrologique (scénario 3), et l'approche qui combine les deux (scénario 4).

Nous regardons trois critères de qualité : la fiabilité (Figure 101), la discrimination (Figure 102) et la qualité globale (Figure 103).

Fiabilité : diagrammes de PIT

La Figure 101 permet de comparer la performance en fiabilité de l'habillage seule des débits avec l'erreur du modèle hydrologique (3) à celle des prévisions brutes (1), de l'habillage seule des pluies avec l'information des analogues (2), et celle de la combinaison des deux approches d'habillage (4).

Nous constatons que l'habillage avec l'erreur du modèle hydrologique (sans ou avec le couplage avec le prétraitement en amont du modèle hydrologique) permet d'améliorer considérablement la fiabilité de prévisions. Cette amélioration drastique se retrouve sur tous les bassins versant et les échéances étudiés. L'influence de l'habillage des pluies avec l'information des analogues en amont du modèle hydrologique reste limitée par rapport à celle de l'habillage avec l'erreur du modèle hydrologique. La combinaison d'un prétraitement et un post-traitement apporte une légère amélioration supplémentaire à la qualité des prévisions. Cette amélioration est plus évidente dans les premières échéances de prévision.

Discrimination : courbes de ROC

L'analyse de performance en termes de discrimination des dépassements du quantile 70% des débits observés est présentée par les courbes ROC dans la Figure 102. En examinant ce graphique nous remarquons que l'influence du pré- et du post-traitement sur la discrimination de prévisions reste limitée et comparable à celle des prévisions brutes. Par conséquent nous constatons que l'application d'un post-traitement est plus efficace par rapport à un prétraitement en termes de fiabilité. Par contre sur la discrimination des prévisions les performances des différents scénarios sont en effet comparables à celle des prévisions brutes.

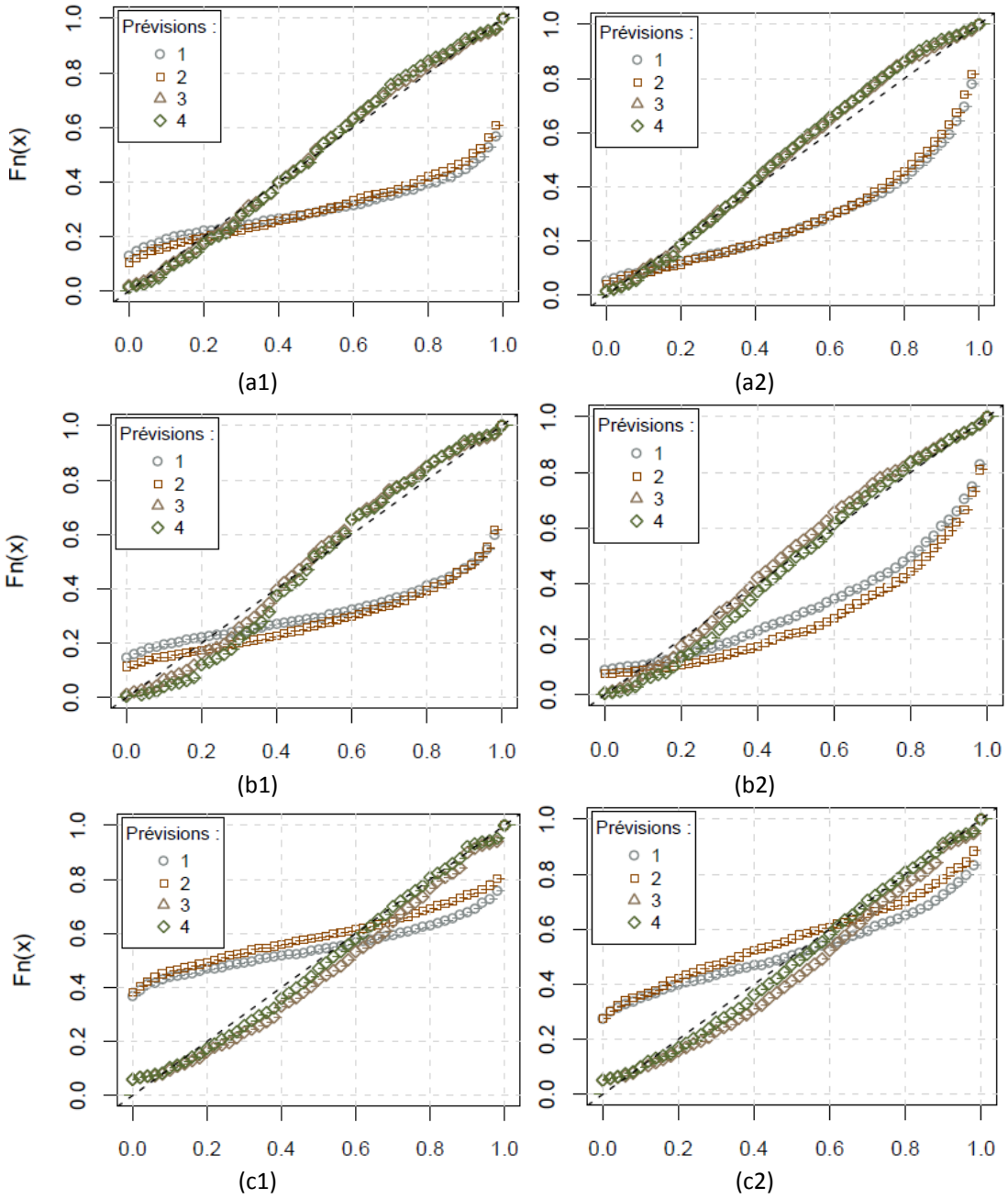


Figure 101 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les histogrammes sont calculés à partir des prévisions hydrologiques habillées avec l'erreur du modèle hydrologique MORDOR – sans et avec la correction à l'aide des analogues.

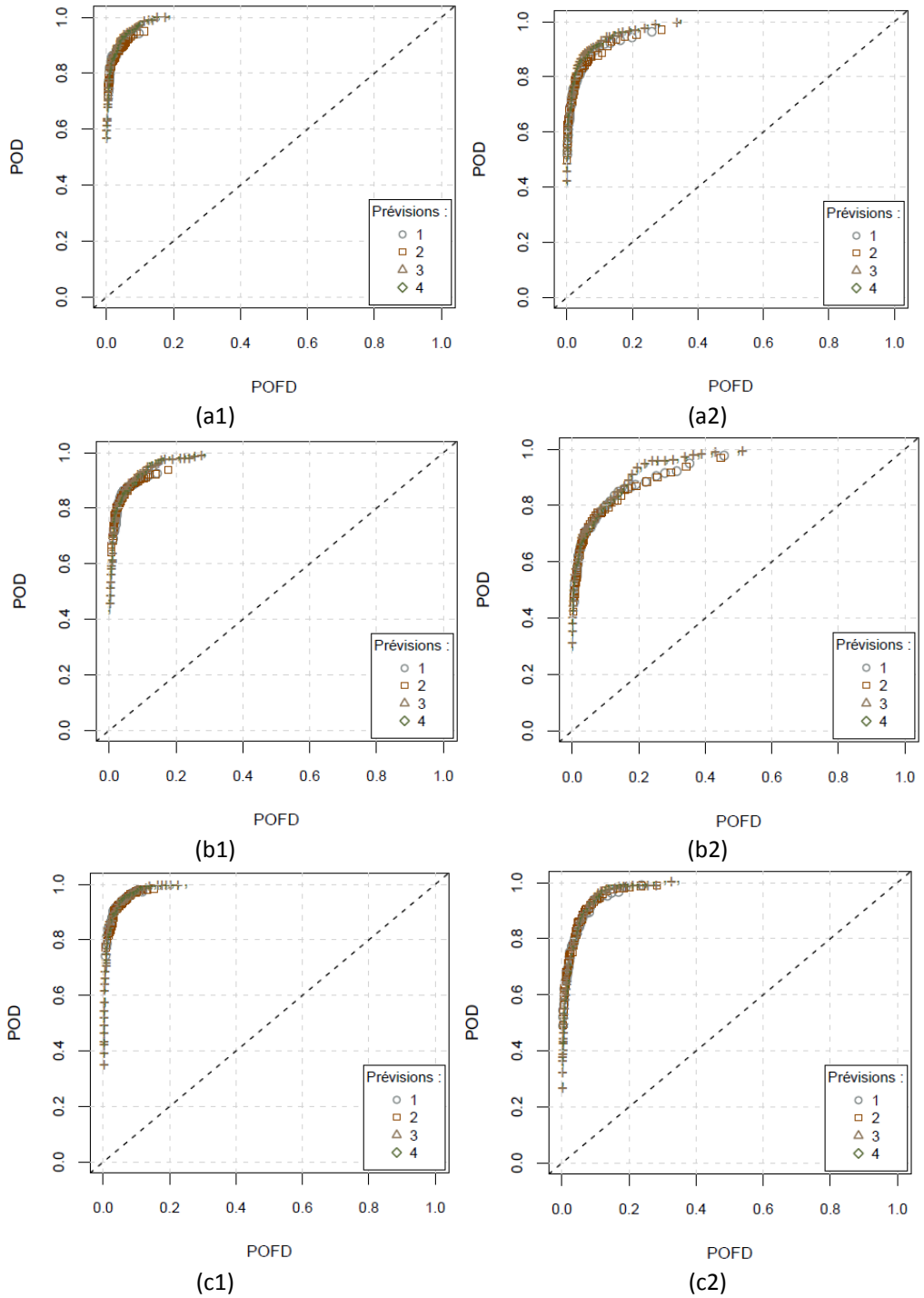


Figure 102 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), des prévisions corrigées avec l'aide des analogues (2), corrigées avec l'erreur du modèle (3) et l'approche combinée (4).

Qualité globale

Pour évaluer la performance des méthodes utilisées en termes de qualité totale, le score de performance *RPSS* est utilisé, calculé avec en référence les prévisions brutes (Figure 103). Ainsi, la performance du scénario 1 s'affiche à zéro, tandis que les performances supérieures et inférieures à 0 indiquent une performance améliorée ou détériorée respectivement par rapport aux prévisions brutes. Ici, l'échantillon de 11 bassins versants a été utilisé et les résultats sont affichés sous la forme de boîtes à moustaches. Les prévisions d'ensemble ici utilisées sont issues du modèle hydrologique *MORDOR.*

La performance de l'habillage avec l'erreur du modèle (scénario 3) est présentée en comparaison avec celle des prévisions brutes (scénario 1), celle de l'habillage avec l'information des analogues au niveau de prévisions de pluie (scénario 2) et de la combinaison de deux approches d'habillage (scénario 4). Nous remarquons dans ces graphiques que l'application d'un post-traitement statistique améliore plus fortement la qualité globale de prévisions pour toutes les échéances examinées que si nous appliquons uniquement un prétraitement statistique au niveau de prévisions de pluie. De plus, la combinaison d'un pré et post-traitement apporte une légère amélioration supplémentaire à la qualité de prévisions post-traitement, et de façon plus évidente pour les premières échéances.

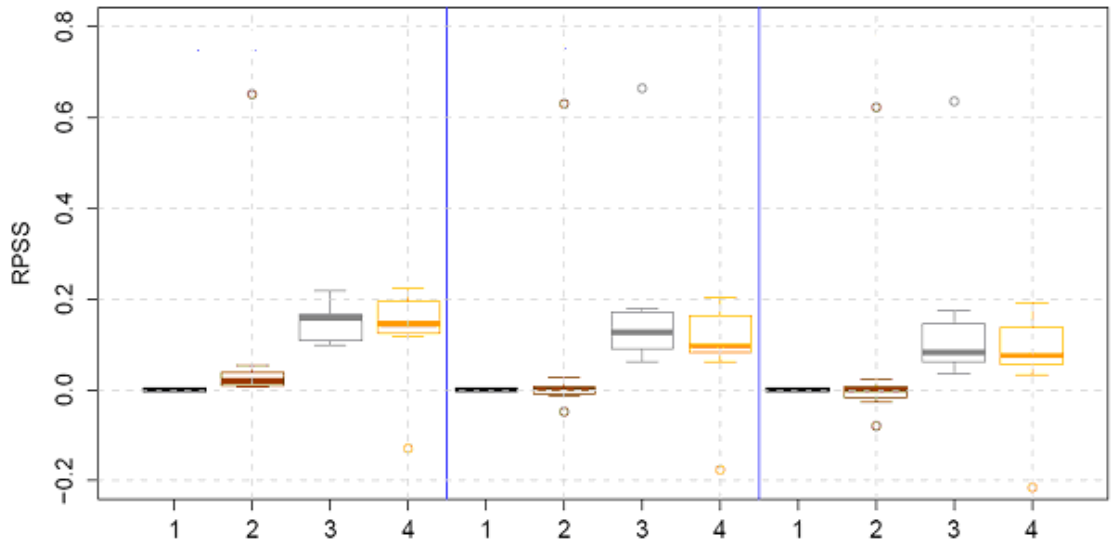


Figure 103 : Boîtes à moustaches du score de performance RPSS (avec en référence les prévisions brutes) pour les prévisions hydrologiques d'ensemble et les échéances 3 (à gauche), 5 (centre) et 7 jours (à droite). Les approches de traitement statistique utilisées sont les : (1) prévisions brutes, (2) l'habillage avec l'information des analogues, (3) l'habillage avec l'erreur du modèle et (4) une combinaison des traitements (2) et (3). Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

Ces conclusions font partie de l'article de Zalachori *et al.* (2012) (Annexe D). Ces résultats sont confirmés par des études similaires effectuées par Kang *et al.* (2010) et Verkade *et al.* (2012) sur des bassins versants coréens et allemands, respectivement. Les auteurs cités ont également conclu que l'effet du prétraitement reste souvent limité en aval du modèle hydrologique, et que les améliorations plus significatives à la qualité des prévisions de débits viennent du post-traitement des prévisions hydrologiques. De plus, ils remarquent également que l'utilisation d'une combinaison de pré- et post-traitement apporte une légère amélioration supplémentaire à la qualité des prévisions hydrologiques d'ensemble.

6.5.3 Prise en compte de l'incertitude totale de prévision

Nous examinons ici le comportement d'une autre famille d'approches de post-traitement statistique qui visent à prendre en compte l'incertitude totale de la prévision et non pas une source particulière, comme l'erreur du modèle hydrologique. Dans ce cas, le calage de la méthode de post-traitement se fait avec les prévisions de débits et non pas avec les débits simulés. Les approches choisies lors de cette thèse appartiennent à la famille des méthodes d'habillage empirique. Il s'agit de la méthode du meilleur membre et ses variantes (Tableau 8, scénarios 5 à 8).

Ces approches, qui appartiennent à la méthode du meilleur membre (en anglais *Best Member Method*), sont présentées dans le paragraphe §3.4. Rappelons que le principe derrière cette famille de méthodes est d'enrichir les prévisions avec l'information provenant de l'erreur minimale (meilleur membre) de prévision qui améliore la qualité de prévisions. Ici nous étudions la méthode originale proposée par Roulston et Smith (2003) (scénario 5), ainsi que ses modifications introduites par Wang et Bishop (2005) (scénario 6) et Fortin *et al.* (2006) (scénario 9). De plus nous proposons deux nouvelles modifications de cette approche :

- la modification de Wang et Bishop (2005) modifiée avec la covariance des échéances afin de prendre en compte la cohérence temporelle (scénario 7),
- l'habillage avec l'erreur de 30% des meilleurs membres (scénario 8), où nous ne sommes pas limités à l'erreur du "meilleur" membre mais nous constituons une archive avec les erreurs de 30% des meilleurs membres, pour chaque échéance de prévision.

Nous ajoutons ici que plusieurs autres modifications de cette approche ont également été testées lors de la thèse. Néanmoins, nous présentons que les modifications les plus performantes. Dans un premier temps nous présentons chaque approche en comparaison avec les prévisions hydrologiques brutes (scénario 1). Les mêmes critères de qualité adoptés précédemment sont étudiés. Nous synthétisons ici les principales conclusions tirées de l'analyse des critères graphiques et numériques.

6.5.3.1 La méthode du meilleur membre et sa modification par Wang et Bishop (2005)

Nous présentons par la suite la performance de la méthode du meilleur membre (scénario 5), ainsi que celle de la modification (scénario 6) faite par Wang et Bishop (2005) selon : les diagrammes de PIT (Figure 104), les courbes ROC (Figure 105 et Figure 106), la dispersion des membres (Figure 107), le scores numériques CRPS (Figure 108 et Figure 109).

Fiabilité : diagrammes de PIT

Dans la Figure 104 nous présentons la performance en termes de fiabilité de prévisions hydrologique d'ensemble brutes. Les résultats produits par le modèle *MORDOR* (à gauche) et *GRP* (à droite) sont présentés en parallèle. Nous remarquons que :

- pour le modèle *MORDOR* (à gauche), l'approche originale (5) de correction arrive à rendre les prévisions légèrement plus fiables. Néanmoins, la modification de 2006 (6) est nettement plus perfomante en termes de fiabilité et réduit la tendance à sous-estimer les observations pour les trois bassins versants examinés. Nous remarquons par ailleurs que cette amélioration est également considérable sur toutes les échéances et sur la plupart de 11 bassins examinés.
- pour les prévisions extraites du modèle *GRP* (à droite), nous observons que la méthode du meilleur membre (5) améliore légèrement la performance des prévisions brutes en termes de fiabilité. Cette amélioration s'applique surtout sur le cas où le modèle hydrologique sous-estime l'observation (pour la Durance à Serre Ponçon et le Verdon à Castillon). Par contre dans le cas où les prévisions montrent une tendance à surestimer l'observation, (par exemple la Romanche aux Chambons), la méthode du meilleur membre (5) ne rend pas les prévisions plus fiables.

De manière générale, c'est la méthode modifiée (6) qui conduit à une amélioration de la fiabilité de prévisions. Cet effet varie d'un bassin versant à l'autre et il est plus évident autour d'un horizon de prévision de 5 jours.

Dispersion des membres

Dans les Figure 105 et Figure 106 nous présentons l'effet de ces approches de correction statistique sur la dispersion de membres de l'ensemble pour les prévisions hydrologiques d'ensemble produites par le modèle *MORDOR* et *GRP* respectivement. La méthode originale du meilleur membre (5) est représentée avec les boîtes à moustache mauves ; ainsi que la méthode modifiée (6) par Wang et Bishop (2005) avec les boîtes à moustache bleues. Nous remarquons que :

- l'influence de ces approches de correction se résume à une évolution progressive de la dispersion de membres en fonction de l'échéance. Néanmoins, l'effet de la méthode modifiée (6) sur la dispersion de membres est beaucoup plus significatif.
- pour le modèle *MORDOR* l'augmentation de la dispersion en fonction de l'échéance reste progressive pour le cas de la Durance à Serre Ponçon. Par contre, sur les deux autres bassins étudiés (le Verdon à Castillon et la Romanche aux Chambons) les membres sont fortement dispersés surtout au-delà de l'échéance de 3 jours. Ce fait explique la performance modérée de la modification de 2006 au niveau fiabilité sur ces deux bassins versants. Sur l'ensemble des 11 bassins versants nous constatons que la dispersion moyenne des membres des prévisions corrigées avec la méthode du meilleur membre (5) à un horizon de 7 jours varie entre 2 et 4 fois la dispersion à 3 jours de prévision. Pour la méthode du meilleur membre modifiée (6) la dispersion des membres à 7 jours est en moyenne 2 fois plus élevée que la dispersion à 3 jours.
- pour les prévisions hydrologiques produites par le modèle *GRP* la performance de la méthode du meilleur membre modifiée (6) introduit une dispersion plus importante aux membres de l'ensemble. Ce comportement est retrouvé sur l'ensemble des 11 bassins versants examinés. De plus, nous remarquons que la dispersion des membres varie peu avec l'application de la méthode du meilleur membre (5) entre l'échéance 3 et 7 jours (augmentation entre 30 et 60%) ; en même temps qu'elle est en moyen multipliée par 2,5 dans le cas de l'approche modifiée (6).

Nous constatons que, pour les deux modèles hydrologiques utilisés, la corrélation entre la racine carrée de l'erreur quadratique moyenne et la dispersion de membres ne devient forte que dans le cas de dispersions faibles.

Discrimination : courbes ROC

En ce qui concerne la capacité des prévisions à discriminer les événements entre eux (Figure 107), nous constatons que :

- pour le modèle *MORDOR*, les prévisions corrigées ne montrent pas une meilleure discrimination par rapport aux prévisions brutes.
- en ce qui concerne le modèle *GRP* nous remarquons que la discrimination des prévisions post-traitées est moins bonne que celle des prévisions brutes pour tous les bassins versants et les échéances étudiées. Cette tendance est plus marquante pour le bassin de la Romanche à Chambon et elle s'accentue quand le seuil de non-dépassement passe de 30 à 90%. Une explication possible est liée à la dispersion trop forte des membres de l'ensemble qui ne permet pas aux prévisions corrigées de discriminer deux événements différents.

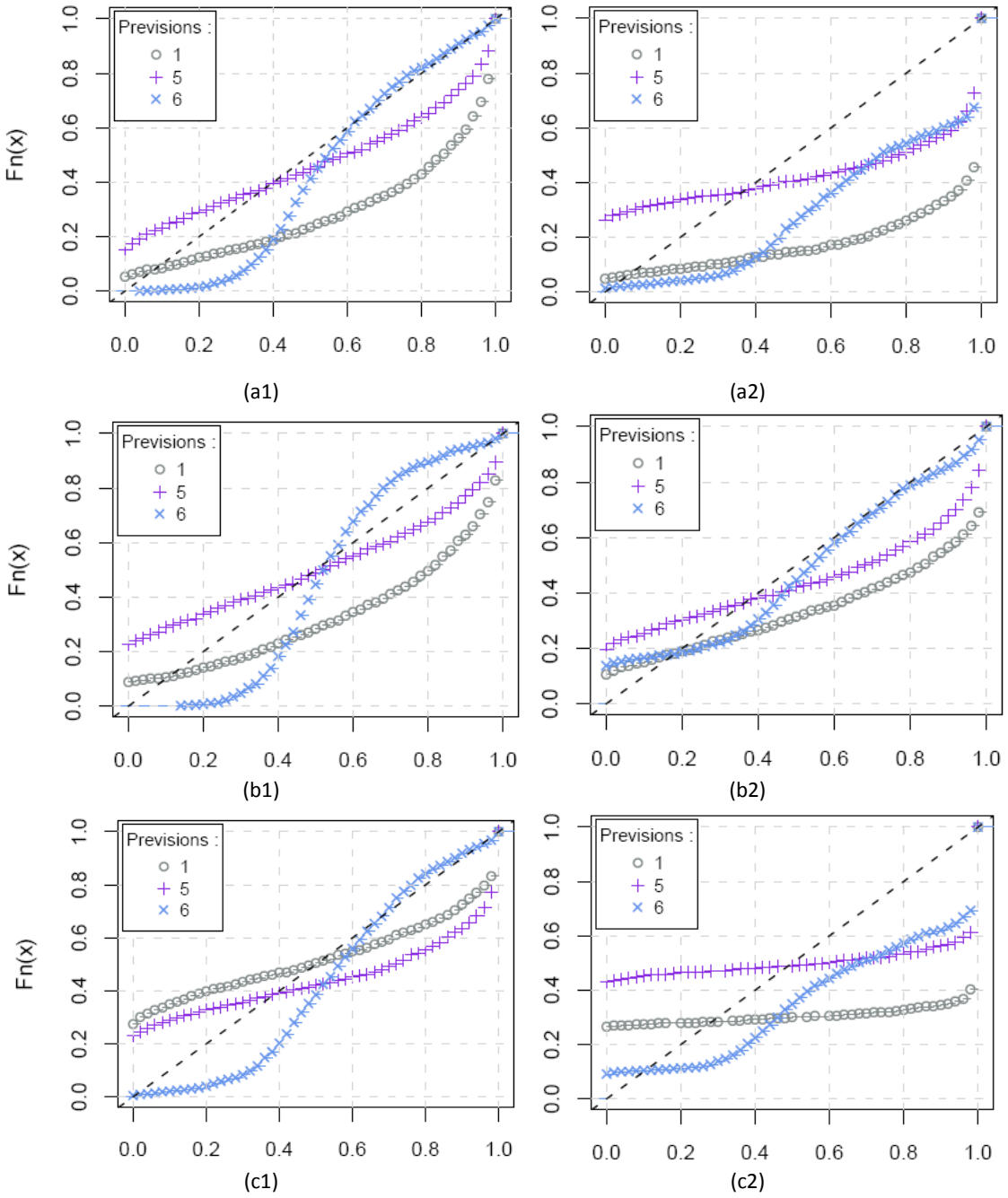


Figure 104 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre et sa modification par Wang et Bishop (2005) et extraites par le modèle hydrologique : (1) MORDOR et (2) GRP.

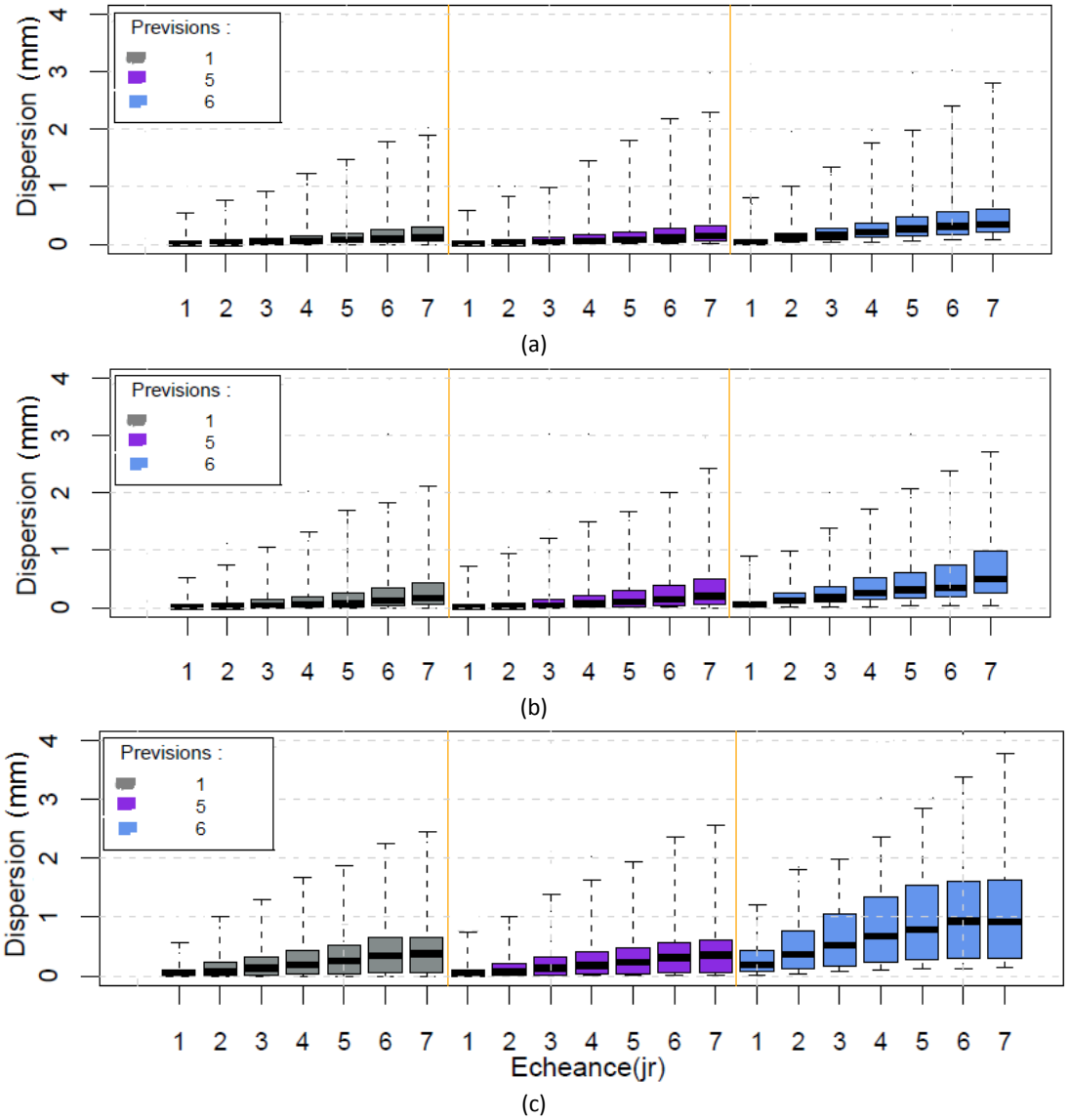
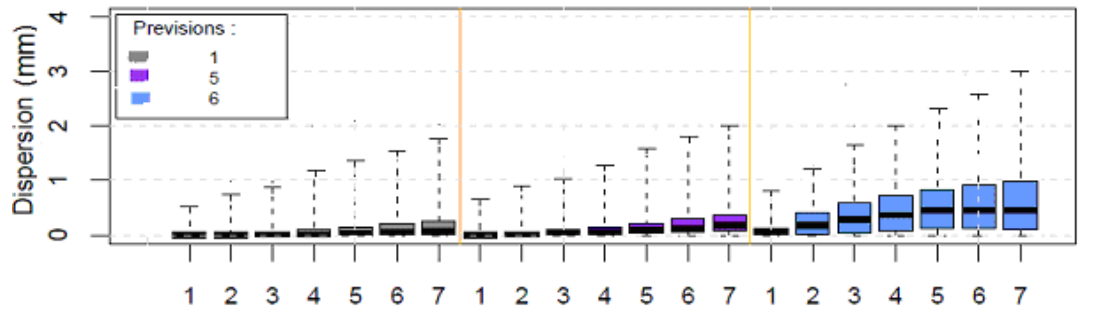
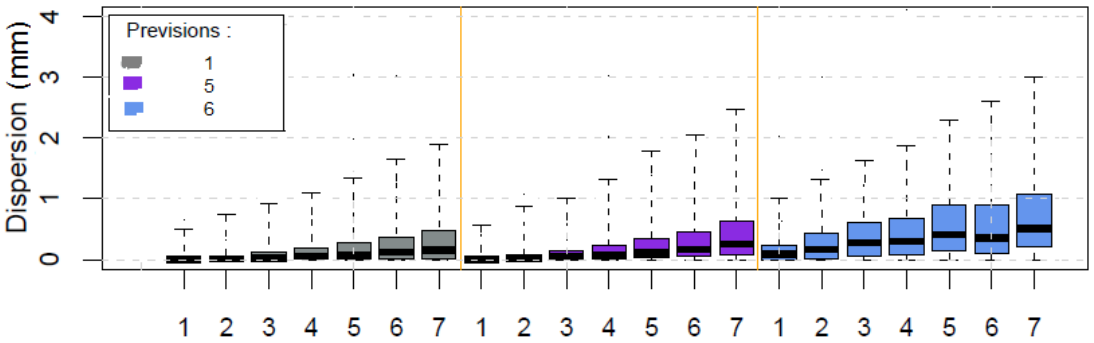


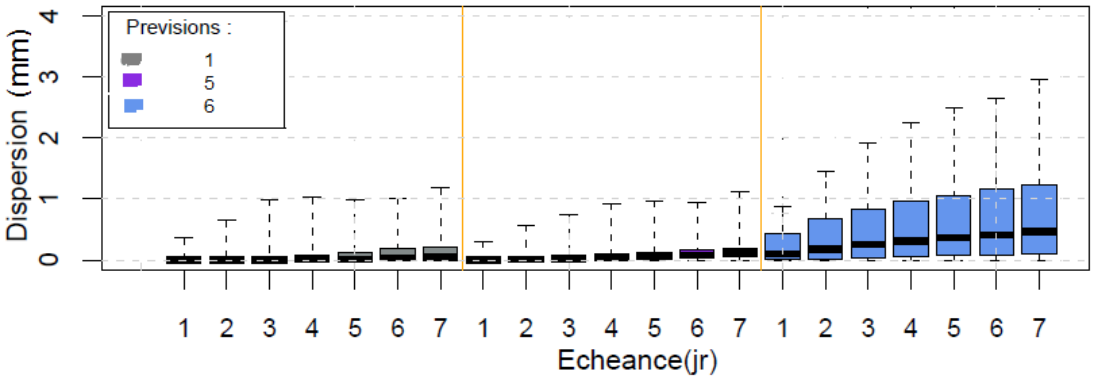
Figure 105 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.



(a)



(b)



(c)

Figure 106 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

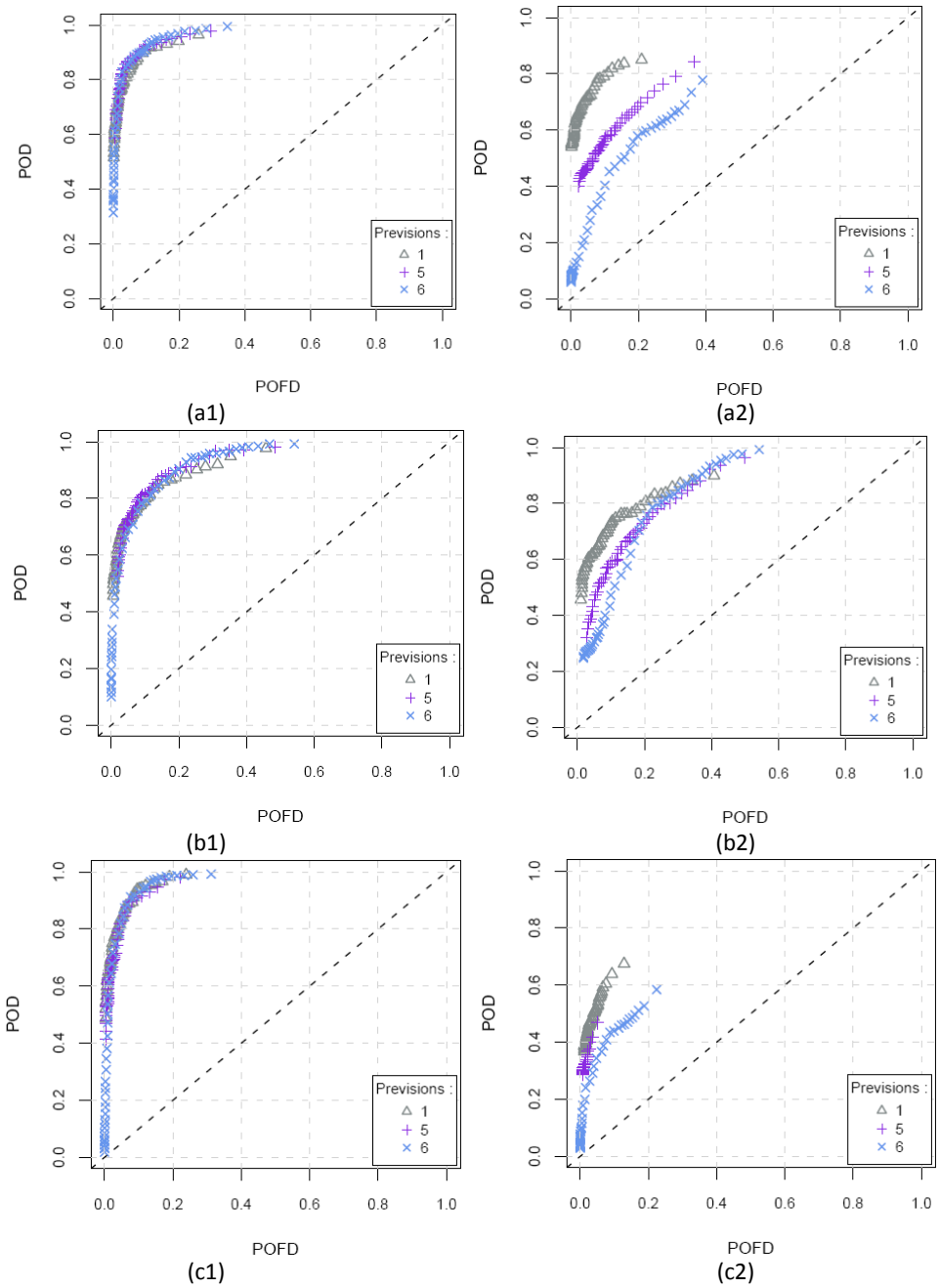


Figure 107 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6).

Qualité globale

Les Figure 108 et Figure 109 présentent les performancess de la méthode du meilleur membre original (5) et modifiée (6) par rapport aux prévisions brutes fournies par le modèle hydrologique *MORDOR* et *GRP* respectivement. Nous constatons que :

- dans le cas du modèle *MORDOR*, l'apport de la méthode du meilleur membre (5) et de sa modification (6) à la qualité globale des prévisions hydrologiques reste limité pour presque tous les bassins versants et échéances testés. Ce comportement est peu influencé par la superficie du bassin.
- pour le modèle GRP, l'apport de la méthode originale du meilleur membre (5) influence peu la performance globale de prévisions hydrologiques. Cependant sa modification (6) n'arrive pas à améliorer la qualité globale des prévisions hydrologiques pour tous les bassins versants examinés. Ceci peut s'expliquer par la mauvaise discrimination des prévisions corrigées, car au niveau fiabilité les prévisions sont considérablement améliorées.

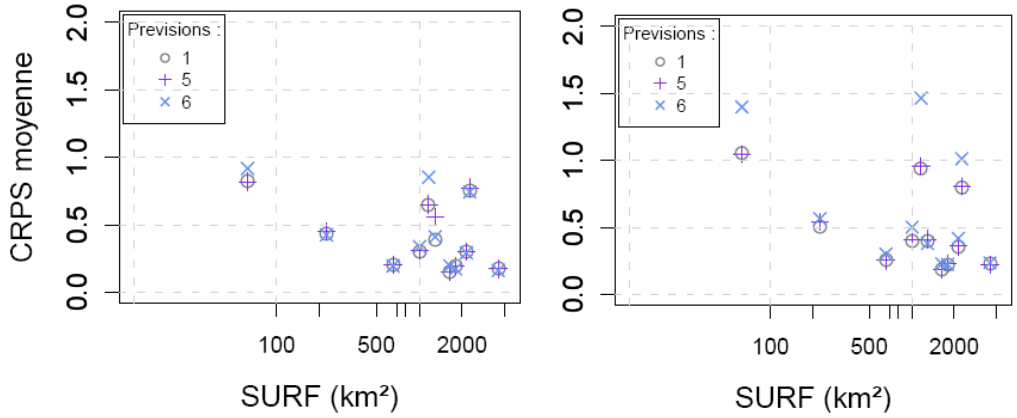


Figure 108 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6).

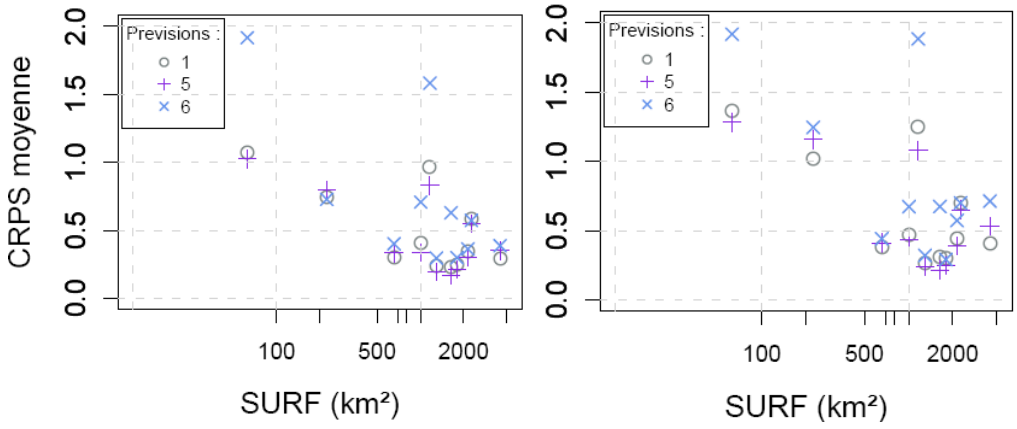


Figure 109 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6).

6.5.3.2 Méthode BMM modifiée pour tenir en compte de la cohérence temporelle

La méthode du meilleur membre modifiée par Wang et Bishop (2005) avait pour but l'augmentation de la dispersion dans le cas des ensembles sous dispersifs en obtenant une covariance entre chaque paire des membres de l'ensemble égale à la covariance d'un membre de l'ensemble et l'observation. Une des limitations de cette approche est qu'elle ne garantit pas que les trajectoires de membres soient conservées. Par conséquent, quand une approche de post-traitement est mise en place pour un horizon de prévision à 3 jours, les trajectoires prévues par les membres changent et la cohérence temporelle n'est pas préservée. Pour faire face à ce problème nous avons introduit dans la méthode du meilleur membre la covariance des échéances.

Les attributs examinés par la suite sont : la fiabilité (Figure 110), la dispersion des membres (Figure 111 et Figure 112), la discrimination (Figure 114 et Figure 115) et la qualité globale (Figure 116 et Figure 117).

Fiabilité : diagrammes de PIT

Sur l'effet de cette approche par rapport à la fiabilité de prévisions (Figure 110), nous remarquons que :

- pour les prévisions produites par le modèle *MORDOR*, la modification proposée (7) limite la tendance des prévisions brutes (1) et modifiées (6) à surestimer l'observation et elle réduit leur tendance à sous-estimer. Ce comportement est répandu pour tous les bassins versants et toutes les échéances examinés. En ce qui concerne les prévisions produites par le modèle *GRP* l'approche proposée (7) améliore considérablement la fiabilité de prévisions brutes (1) et traitées (6) en limitant leur tendance à sous-estimer et surestimer l'observation.

Dispersion des membres

En ce qui concerne l'influence de la méthode du meilleur membre modifiée en 2005 (6), ainsi que de la méthode proposée (7) sur la dispersion de membres pour les modèle hydrologique *MORDOR* et *GRP* (Figure 111 et Figure 112), nous constatons que :

- pour le modèle hydrologique *MORDOR,* la modification proposée (7) réduit légèrement la dispersion des membres en adoucissant les effets de la méthode modifiée du meilleur membre (6). De plus nous remarquons que l'évolution de la dispersion de membres liée aux corrections examinées change d'un bassin à l'autre : pour les bassins de la Durance à Serre Ponçon et le Verdon à Castillon la dispersion est augmentée progressivement ; en même temps que pour le bassin de la Romanche aux Chambons cette augmentation est plus brusque.

Sur l'ensemble de 11 bassins versant nous constatons que la dispersion des pré visions corrigées avec la modification proposée (7) à un horizon de 7 jours est réduite en moyenne entre 10 et 30% par rapport à la dispersion des prévisions de la méthode du meilleur membre modifiée par Wang et Bishop (6). De plus, la dispersion moyenne des membres à un horizon de 7 jours produite par la méthode (7) varie – selon le bassin versant - entre 1,5 et 3 fois la dispersion des membres à 3 jours de prévision.

- en ce qui concerne les prévisions fournies par le modèle *GRP*, l'évolution de la dispersion des membres est plus douce en fonction de l'échéance. Cependant nous observons que, lors des premières échéances de prévision, la dispersion des membres est significativement plus forte par rapport à celle de prévisions habillées avec la méthode du meilleur membre modifiée (6). Sur l'ensemble de 11 bassins versants nous constatons que l'augmentation de la dispersion moyenne lors des premières échéances de prévisions varie entre 50 et 100%. Par contre le ratio entre la dispersion moyenne à 7 jours et celle à 3 jours est légèrement réduite (10 à 20%). Ce qui apporte une amélioration supplémentaire à la fiabilité des prévisions hydrologiques aux courts horizons de prévisions (Figure 113). Cette figure montre les diagrammes de probabilité cumulée pour les deux modèles utilisés, les trois bassins versants étudiés et l'échéance de prévision égale à 3 jours.

Discrimination : courbes ROC

Nous remarquons en examinant les courbes ROC (Figure 114) que :

- pour les prévisions hydrologiques fournies par le modèle *MORDOR* la modification proposée améliore la discrimination légèrement des prévisions - surtout par rapport à la méthode du meilleur membre modifiée par Wang et Bishop en 2005 (6). Néanmoins, l'influence de l'approche proposée reste limitée. Ce comportement est répandu pour tous les bassins versants et toutes les échéances examinées.
- dans le cas du modèle *GRP*, cet effet de la modification proposée (7) devient beaucoup plus évident sur les prévisions hydrologiques. Pour tous les bassins versants et échéances examinés la méthode proposée améliore la capacité de prévisions à discriminer les événements entre eux par rapport aux prévisions brutes (1) et habillées avec la méthode du meilleur membre de Wang et Bishop (6).

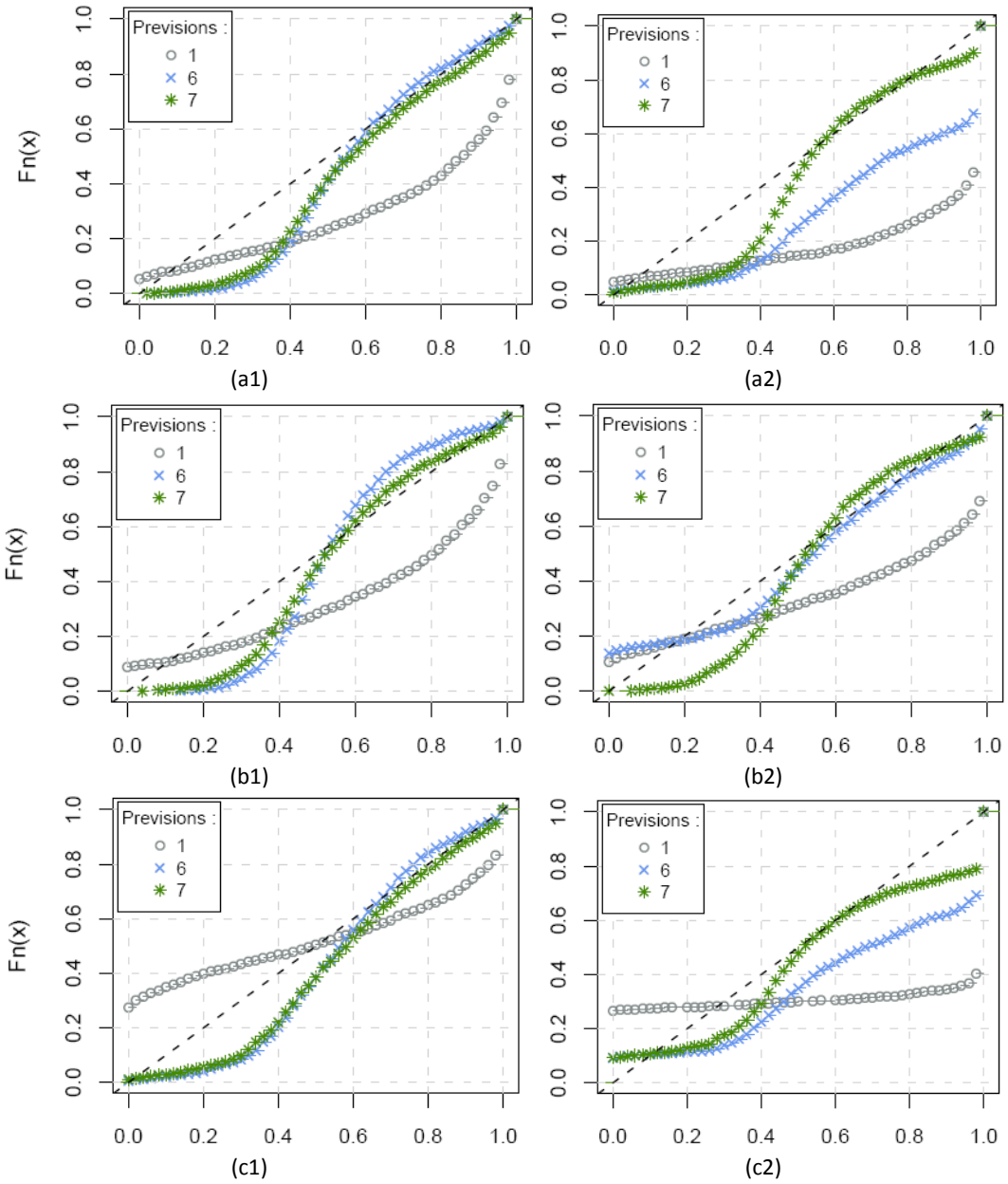


Figure 110 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7), extraites par le modèle hydrologique : (1) MORDOR et (2) GRP.

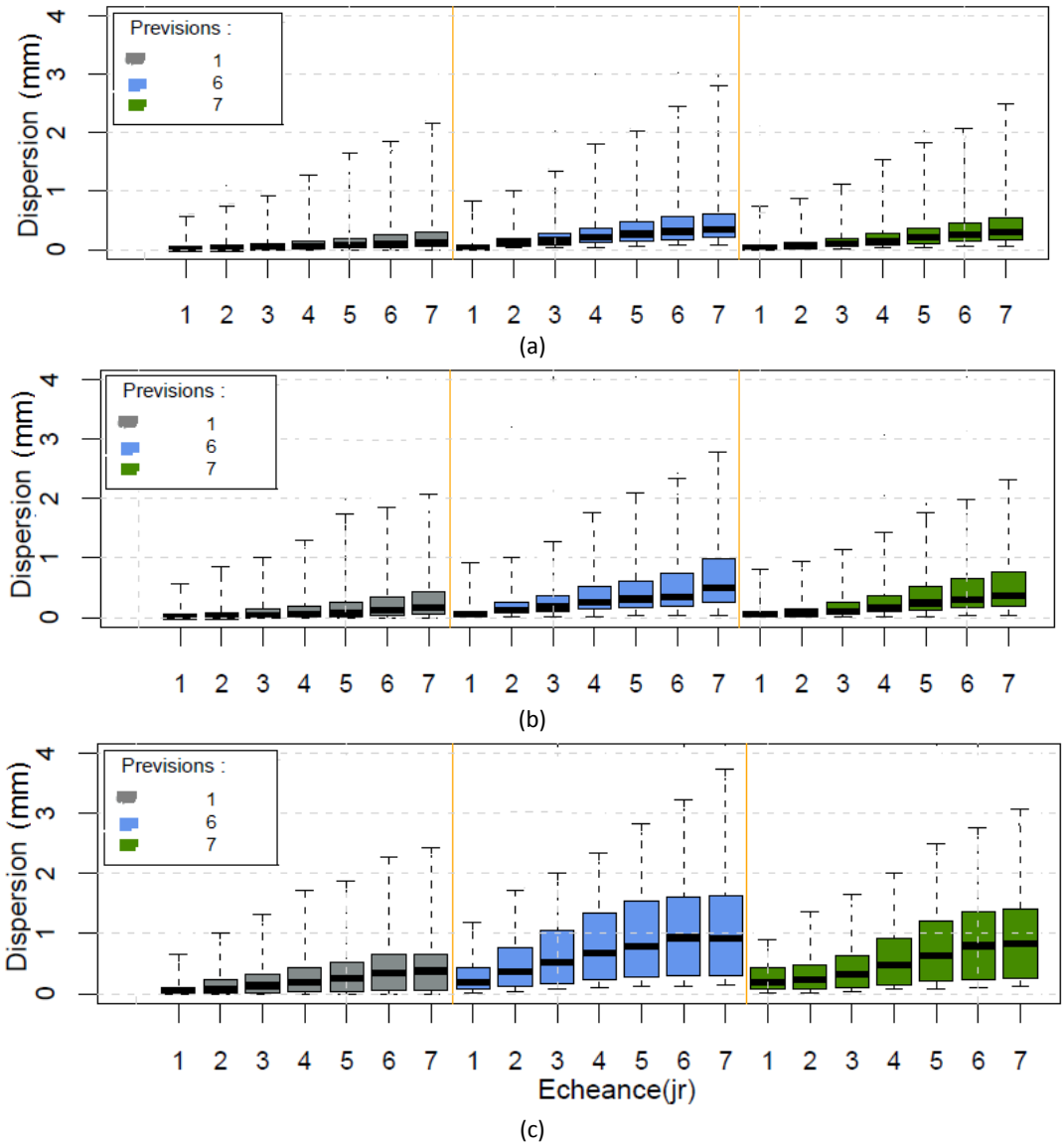


Figure 111 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

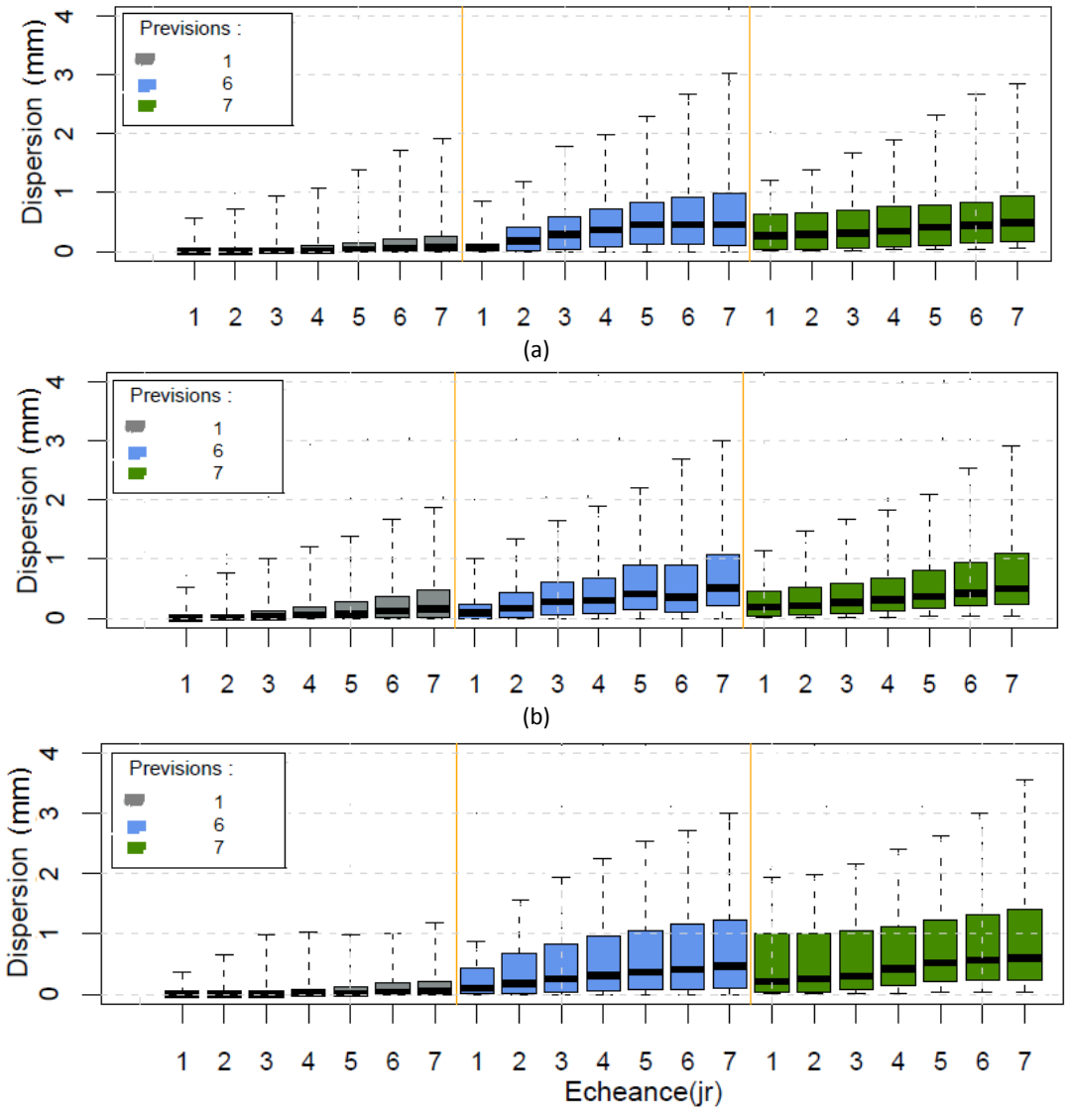


Figure 112 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

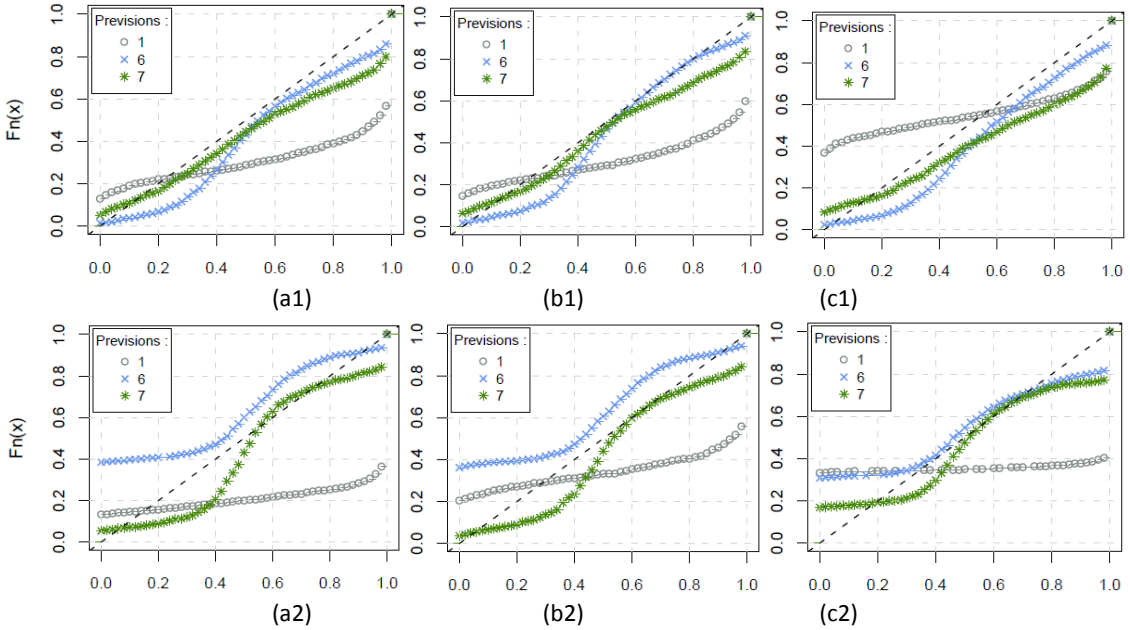


Figure 113 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 3 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (2005) et la modification que nous proposons et extraites par le modèle hydrologique : (1) MORDOR et (2) GRP.

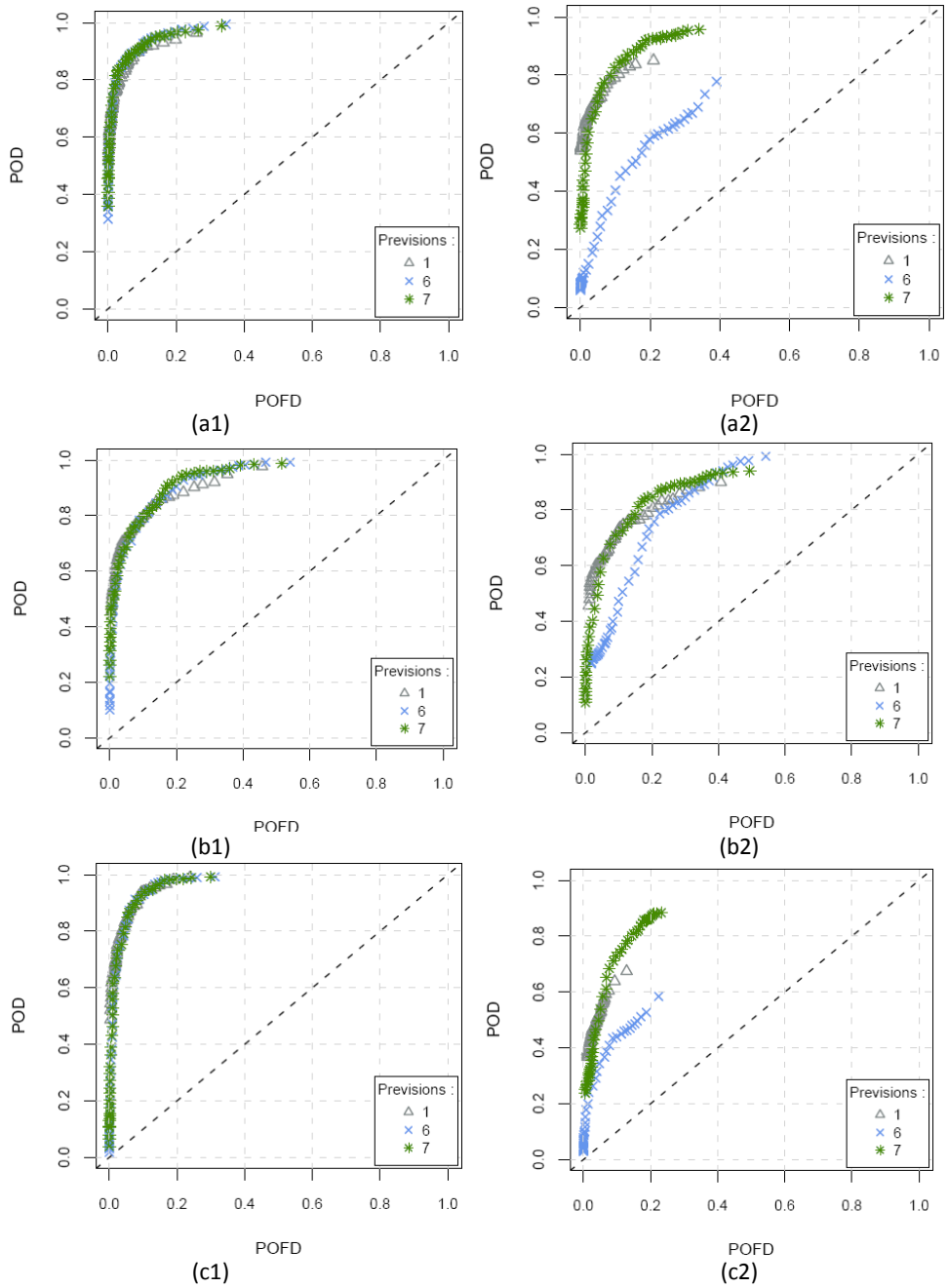


Figure 114 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7).

Qualité globale

Les Figure 115 et Figure 116 démontrent la qualité globale en termes de valeurs moyennes du *CRPS* des prévisions brutes (1) et traitées (6) et (7) fournies par les modèles *MORDOR* et *GRP* respectivement. Nous remarquons que :

- l'approche proposée améliore considérablement la performance globale des prévisions surtout par rapport aux prévisions habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6). Ce comportement se retrouve aux prévisions extraites de deux modèles utilisés et il y a peu de dépendance entre la performance de la méthode et la superficie du bassin. Cependant pour les prévisions produites par le modèle *MORDOR*, l'effet du traitement proposé est plus évident aux échéances au-delà de 5 jours, quand pour le modèle *GRP* ceci est apparent dès les premières échéances. De plus cette performance est le résultat d'une amélioration combinée sur la fiabilité et la discrimination des prévisions.

La Figure 117 présente la carte des bassins versant en fonction de la performance de l'approche statistique proposée.

- Pour les prévisions hydrologiques du modèle *MORDOR* nous constatons que l'approche améliore la qualité globale de prévisions pour la plupart de bassins versants examinés. De plus, la performance de la méthode proposée est fortement influencée par l'échéance de prévision. Par exemple pour les horizons courts de prévision elle fonctionne mieux pour les bassins versants des Cévennes, du Massif Central du Sud et des Alpes du Sud et de l'Ain à Vouglans. Par contre, pour les horizons plus lointains de prévision nous remarquons que sa performance sur les bassins versants alpins et l'Ain à Vouglans reste limitée.
- Examinons les prévisions d'ensemble issues du modèle *GRP*, nous constatons que, comme dans le cas des prévisions fournies par le modèle *MORDOR*, l'amélioration apportée est dépendante de l'échéance de prévision. Nous remarquons également que l'amélioration apportée par la méthode proposée est similaire pour les deux modèles hydrologiques avec une meilleure performance sur les bassins du Massif Central. Une explication possible liée à cette observation est que l'incertitude des prévisions aux bassins versants des Cévennes et du Massif Central est très élevée à cause de l'influence orographique même aux horizons courts de prévisions. Du coup en prenant en compte l'information existante aux précedantes échéances nous arrivons à mieux prendre en compte l'incertitude inhérente.

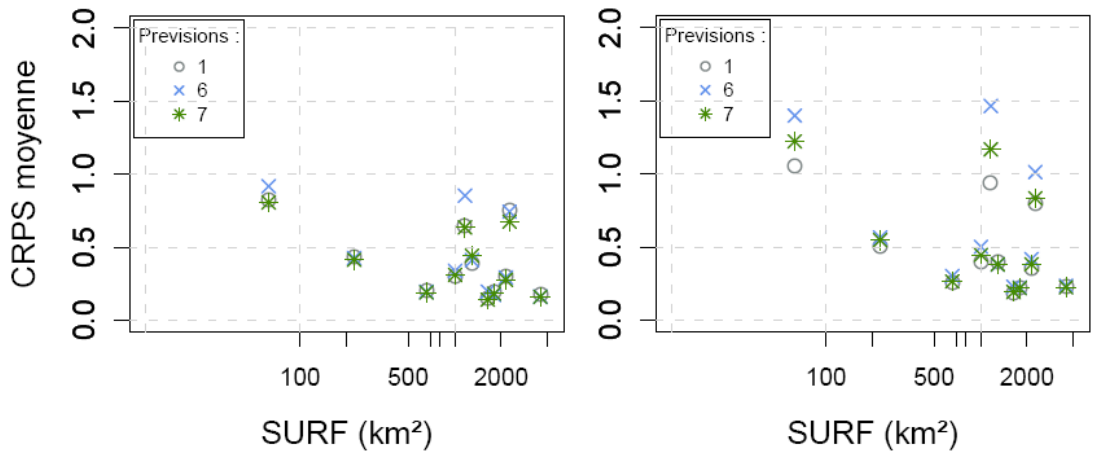


Figure 115 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et modifiée avec covariance des échéances (7).

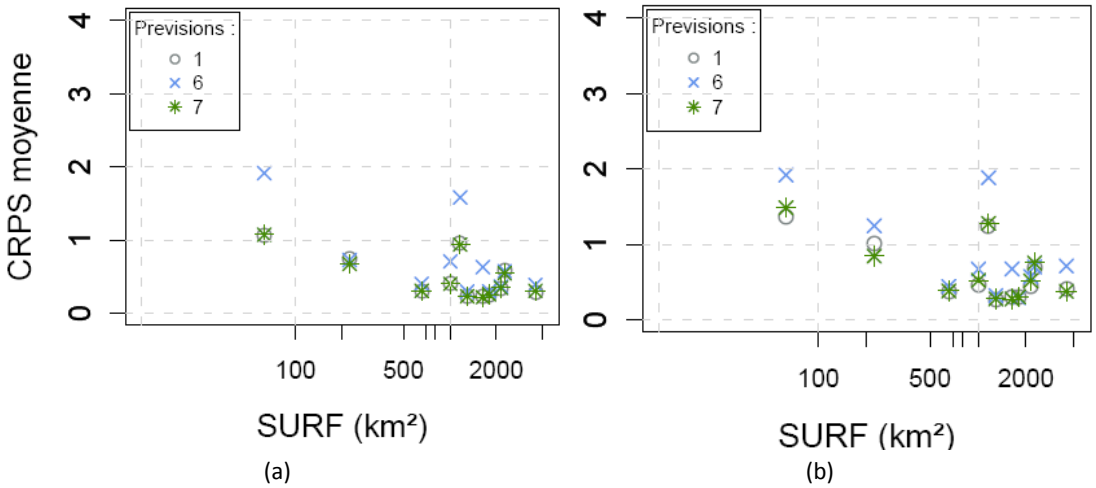


Figure 116 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec la méthode du meilleur membre modifiée (6) et modifiée avec covariance des échéances (7).

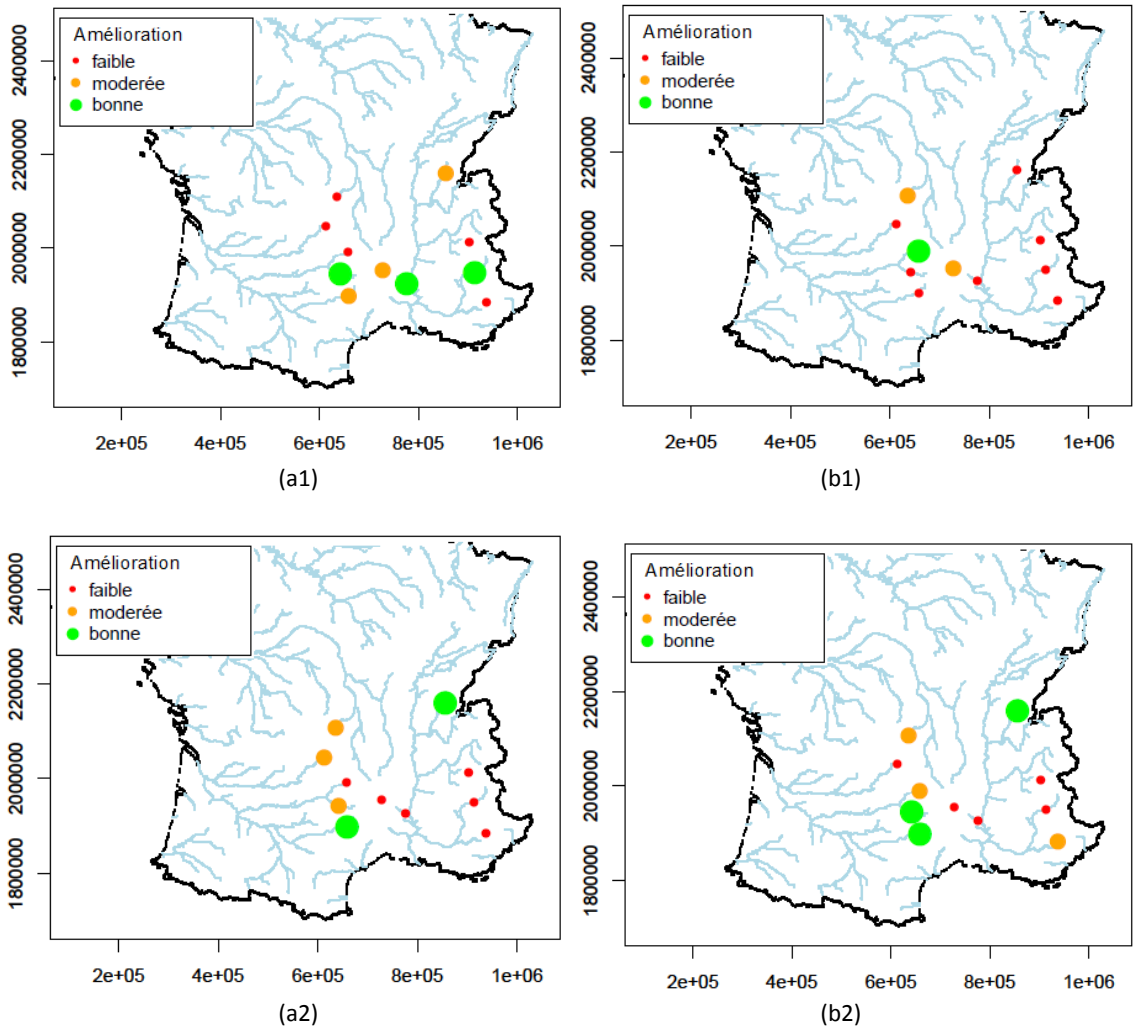


Figure 117 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS la méthode du meilleur membre modifiée avec covariance des échéances. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours.

6.5.3.3 Habillage avec l'erreur de 30% des meilleurs membres

Une autre modification proposée vise à augmenter la dispersion de membres de l'ensemble en prenant en compte l'erreur de 30% des meilleurs membres. Ces erreurs sont évaluées et une distribution de probabilité est ajustée aux valeurs obtenues. Cette distribution servira pour le tirage aléatoire des erreurs qui iront corriger les prévisions d'ensemble brutes lors de l'application de la méthode. L'hypothèse souvent adoptée est que les erreurs de tous les membres suivent une loi de probabilité normale. Afin d'assurer que cette hypothèse est respectée, une transformation logarithmique des valeurs prévues et observées est appliquée lors du calcul de l'erreur de tous les membres.

Fiabilité : diagramme de PIT

Dans la Figure 118 nous présentons les diagrammes de probabilité cumulée pour les prévisions habillées avec l'erreur de 30% des meilleurs membres (scénario 8) pour l'échéance de 7 jours sont tracés en comparaison avec les diagrammes produits par les prévisions brutes (scénario 1) et les prévisions habillées avec la méthode du meilleur membre modifiée en 2005 (scénario 6). Les prévisions sont produites par le modèle hydrologique *MORDOR* (à gauche) et *GRP* (à droite). Nous remarquons que :

- dans les cas de deux modèles, l'habillage avec l'erreur de 30% des meilleurs membres (scénario 8) réduit la tendance des prévisions brutes à sous-estimer l'observation. Cependant elle démontre une influence limitée quand les prévisions surestiment l'observation. Ce comportement est plus apparent sur les horizons lointains de prévision. De plus, nous constatons que dans le cas du modèle *MORDOR* l'amélioration sur la fiabilité des prévisions par la méthode proposée est nettement meilleure par rapport à celle du modèle *GRP*.

Ces conclusions sont valables pour l'ensemble de bassins versants examinés.

Dispersion des membres

En examinant l'influence de l'approche proposée sur la dispersion de membres (Figure 119 et Figure 120), nous constatons que :

- dans le cas de deux modèles la dispersion des membres est augmentée par rapport à la dispersion de l'ensemble brute, cette augmentation reste pourtant modérée par rapport à la dispersion introduite par la méthode du meilleur membre modifiée par Wand et Bishop (6). Nous remarquons également que l'évolution de la dispersion des membres avec la méthode proposée (en noir) reste proportionnelle à celle des prévisions brutes (en gris) pour tous les bassins examinés.

- pour le modèle *MORDOR* et sur l'ensemble de 11 bassins versants, nous constatons que la dispersion moyenne des membres à 7 jours d'échéance pour les prévisions habillées avec l'erreur de 30% des meilleurs membres (8) varie entre 1,5 et 2 fois la dispersion moyenne à 3 jours. De façon similaire, pour le modèle GRP la dispersion moyenne à 7 jours correspond à 2 fois la dispersion à 3 jours. En gérénal, la dispersion des membres habillés avec l'erreur de 30% des meilleurs membres (8) est réduite de 50 à 100% (selon le bassin, le modèle hydrologique et l'échéance de prévision) par rapport à la dispersion des membres habillées avec la méthode du meilleur membre modifiée (6).

Discrimination : courbes de ROC

A partir des courbes ROC (Figure 121) produites par les prévisions brutes et traitées issues des deux modèles hydrologiques, nous observons que :

- dans le cas du modèle *MORDOR*, en appliquant l'habillage avec l'erreur de 30% des meilleurs membres (8) la discrimination des prévisions s'améliore. De plus, la performance de l'habillage avec l'erreur de 30% des meilleurs membres (8) est également améliorée par rapport à celle de la méthode du meilleur membre modifiée en 2005 (6).
- pour les prévisions hydrologiques du modèle *GRP,* l'approche que nous proposée (8) améliore considérablement la capacité des prévisions à discriminer entre deux événements par rapport à celles des prévisions brutes (1) et celle de prévisions traitées avec la méthode du meilleur membre modifiée en 2005 (6). Ce comportement est similaire pour la plupart de bassins versants, ainsi que pour tous les horizons de prévision examinés. Par contre ce comportement devient moins marquant quand le seuil de non-dépassement passe de 30 à 90%. Nous remarquons que l'approche proposée (8) est significativement plus performante pour les prévisions hydrologiques tirées du modèle *GRP*. Cette approche (8) reconstitue une partie de cette capacité inhérente des pluies.

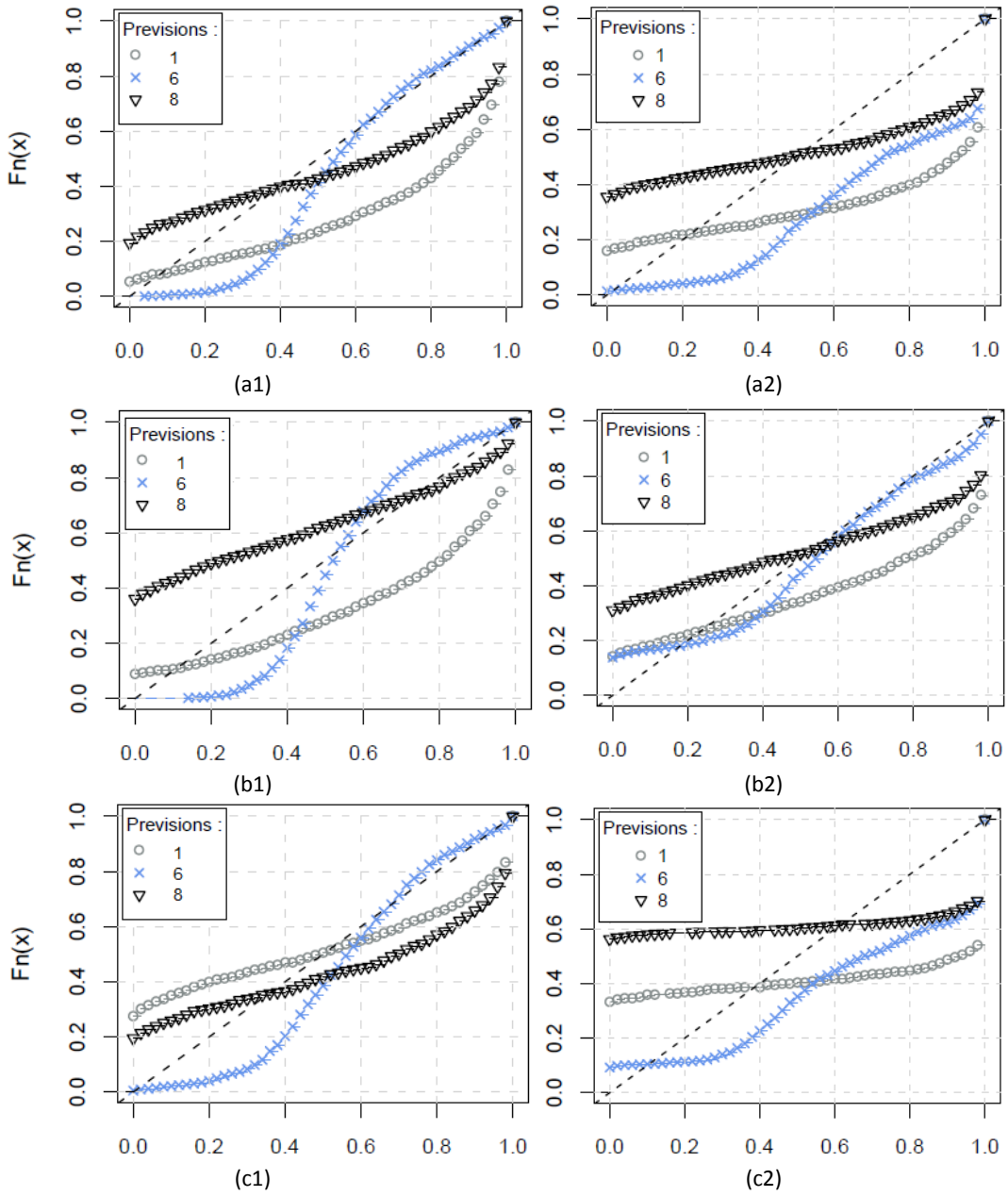


Figure 118 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec l'erreur de 30% de meilleurs membres et fournies par le modèle hydrologique : (1) MORDOR et (2) GRP.

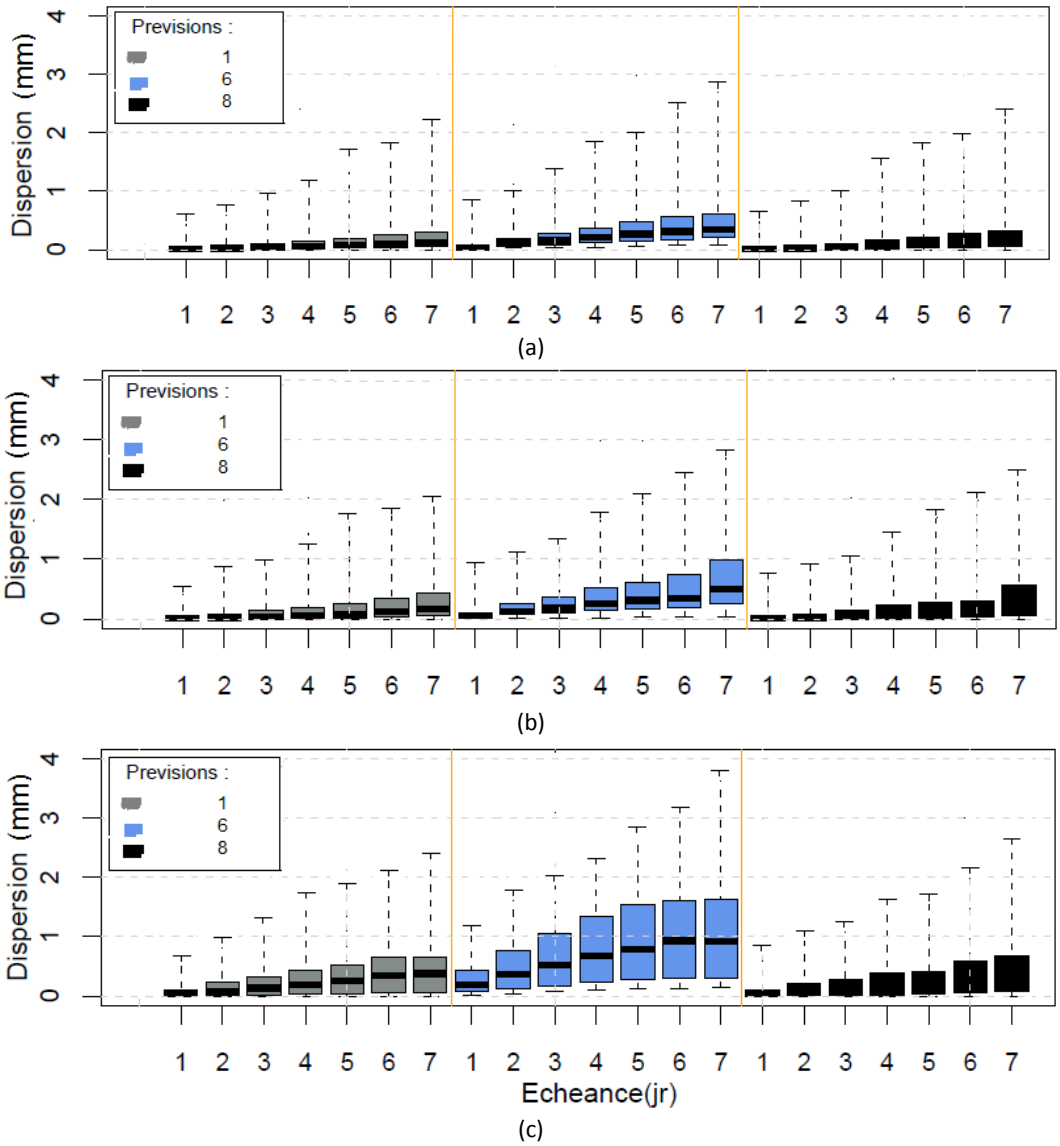


Figure 119 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et avec l'erreur de 30% de meilleurs membres (8) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

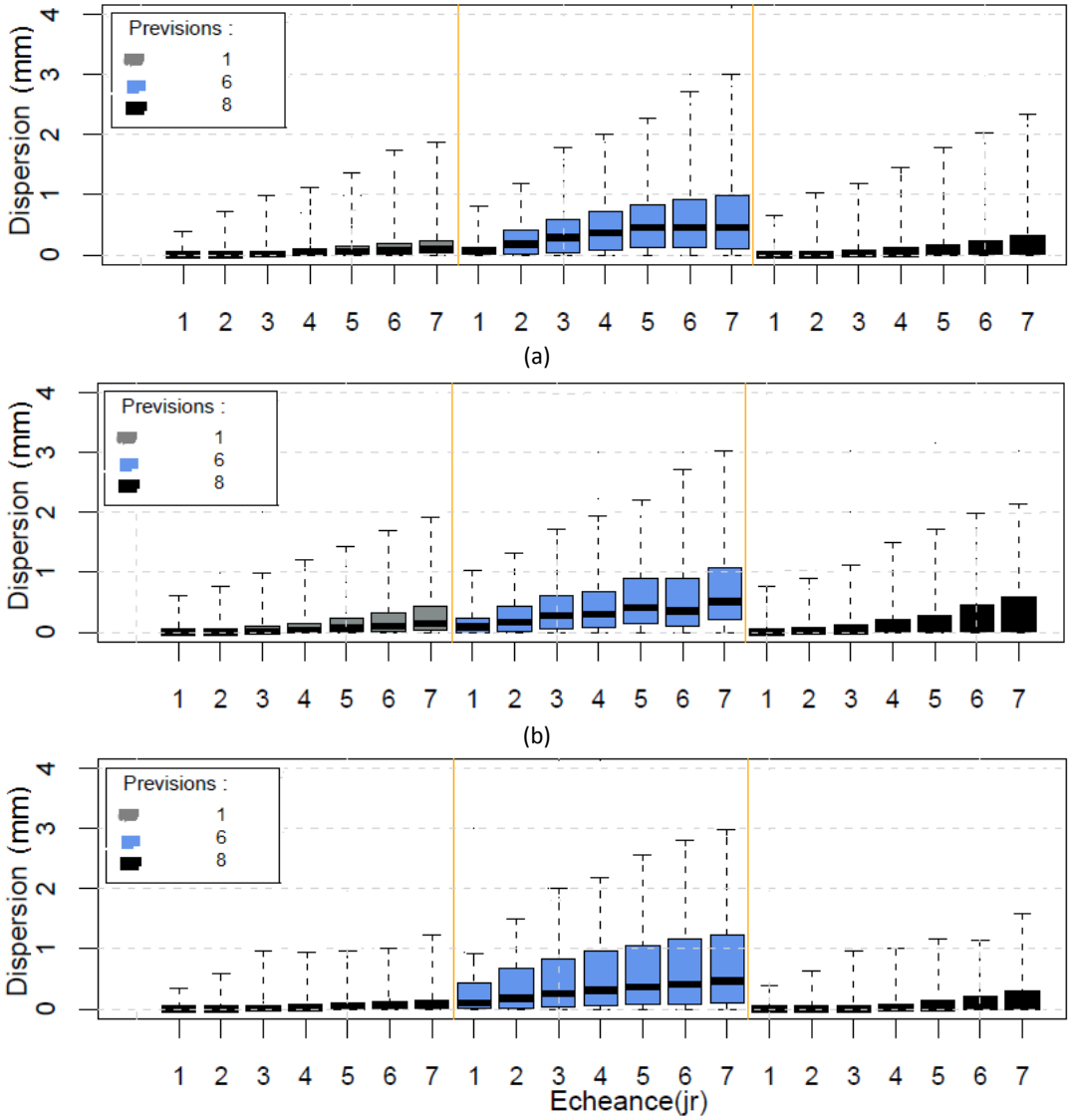


Figure 120 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et avec l'erreur de 30% de meilleurs membres (8) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

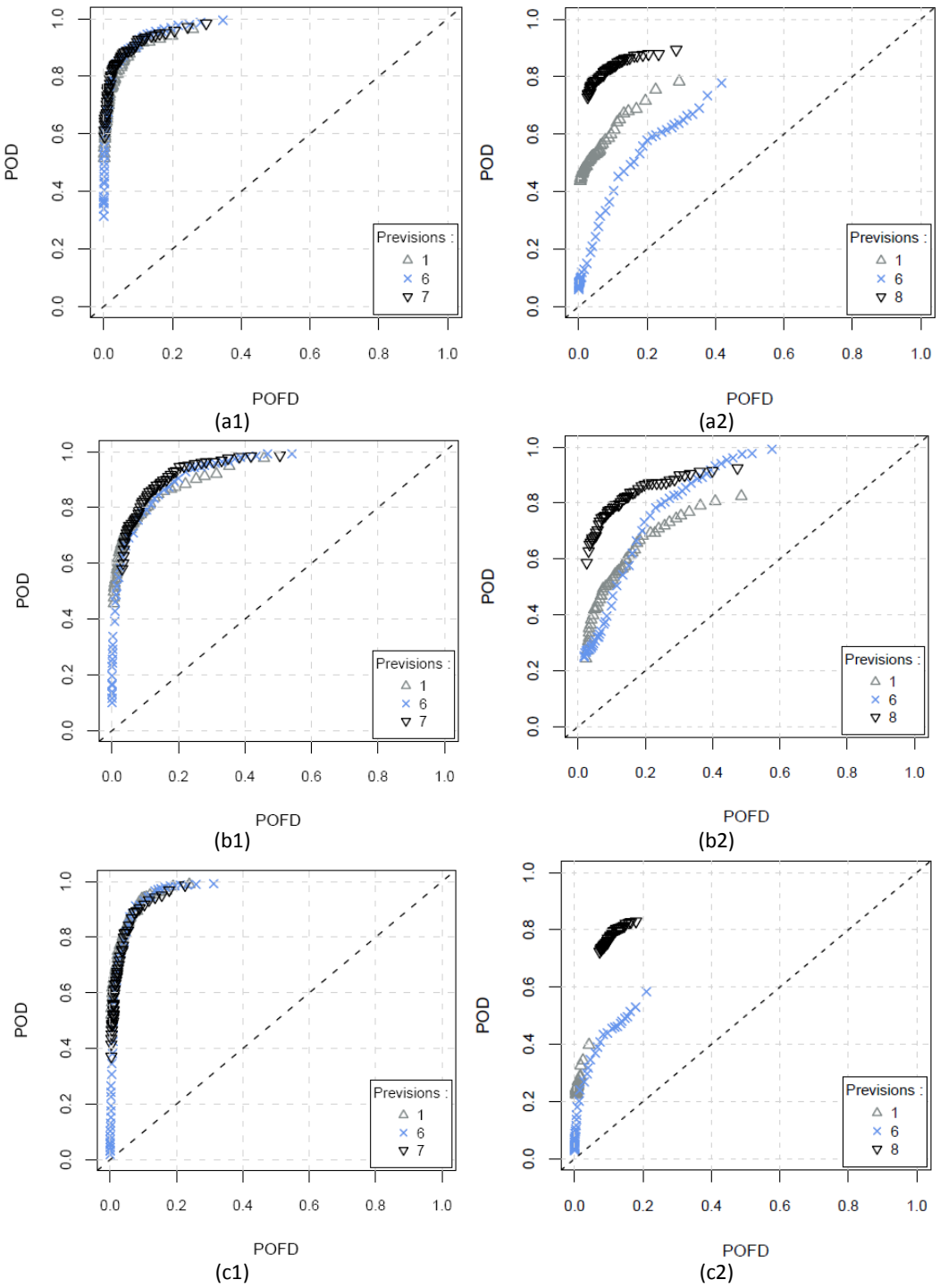


Figure 121 : Les courbes ROC pour un seuil de dépassement égal au quantile de 70% pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les courbes ROC sont calculées à partir des prévisions hydrologiques habillées avec l'erreur de tous les membres produites par le modèle hydrologique : (1) MORDOR et (2) GRP.

Qualité globale

Ensuite, dans les Figure 122 et Figure 123 nous examinons la qualité globale de prévisions traitées avec la méthode proposée (8) et pour les échéances 3 (à gauche) et 7 jours (à droite). Nous remarquons que :

- selon les résultats liés au modèle *MORDOR,* la performance globale des prévisions traitées en termes de valeurs moyenne de *CRPS* est améliorée par rapport à celle de prévisions brutes et des prévisions habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6). Néanmoins, cette amélioration reste limitée. Ce comportement est plus évident aux horizons lointains de prévision.
- en ce qui concerne les prévisions hydrologiques fournies par le modèle *GRP*, l'amélioration de la qualité globale des prévisions est significative pour presque tous les bassins versants et échéances examinés.

Dans le cas de deux modèles nous constatons que ce comportement est peu dépendant de la surface du bassin examiné.

Finalement, la Figure 124 illustre la carte d'amélioration apportée à la qualité des prévisions brutes par l'habillage avec l'erreur de 30% des meilleurs membres (8). Nous constatons que :

- dans le cas des prévisions fournies par le modèle *MORDOR*, cette approche améliore la qualité dans un nombre limité des bassins versants. La localisation des bassins versants où la méthode fonctionne mieux se concentre surtout dans la région du Massif Central et des Cévennes.
- en ce qui concerne le modèle *GRP*, nous observons une amélioration supplémentaire apportée sur les bassins des Alpes du Sud et du Juras ; ce qui en accord avec les figures concernant la qualité globale des prévisions. Cette amélioration se porte surtout sur le pourcentage de l'amélioration apportée et peu sur la localisation des bassins.

Nous remarquons également que la performance de la méthode proposée est dépendante de l'échéance. Néanmoins, à la différence des autres approches examinées auparavant, l'habillage avec l'erreur de 30% des meilleurs membres se montre plus performant aux horizons lointains de prévision.

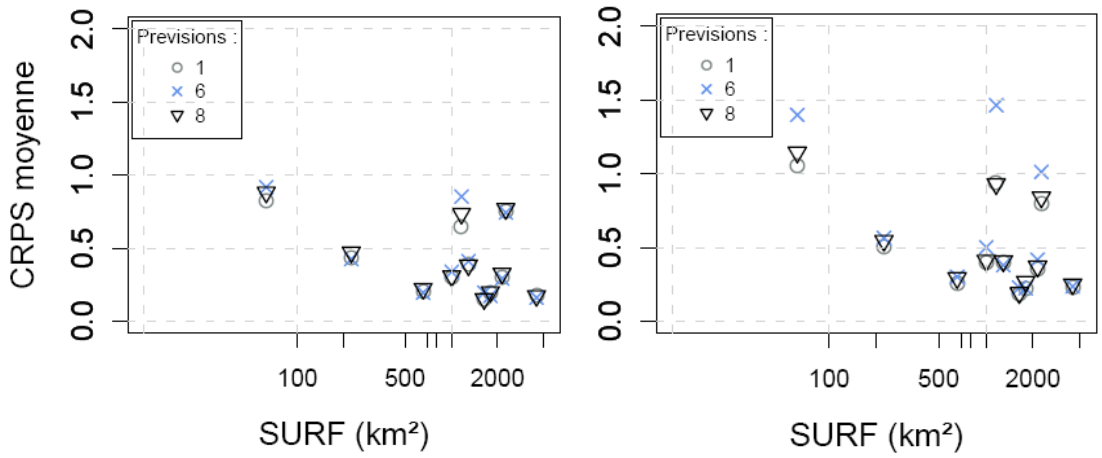


Figure 122 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et habillées avec l'erreur de 30% des meilleurs membres (8).

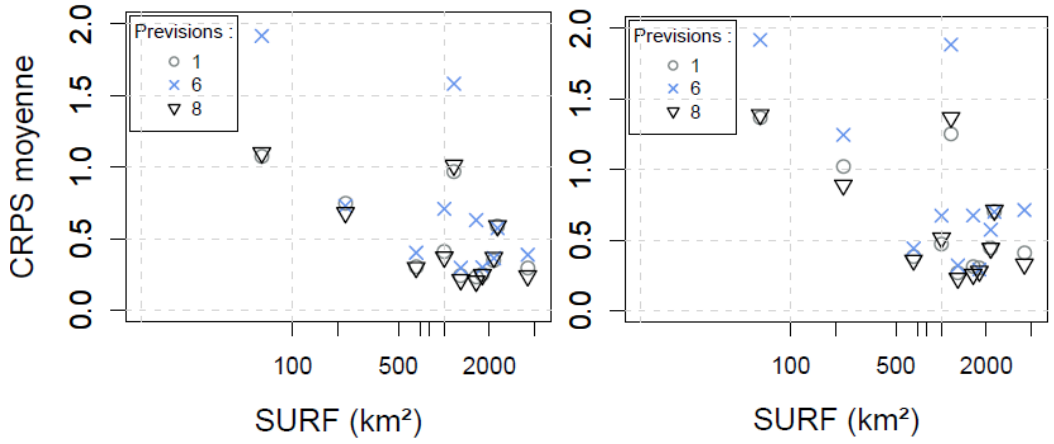


Figure 123 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et habillées avec l'erreur de 30% des meilleurs membres (8).

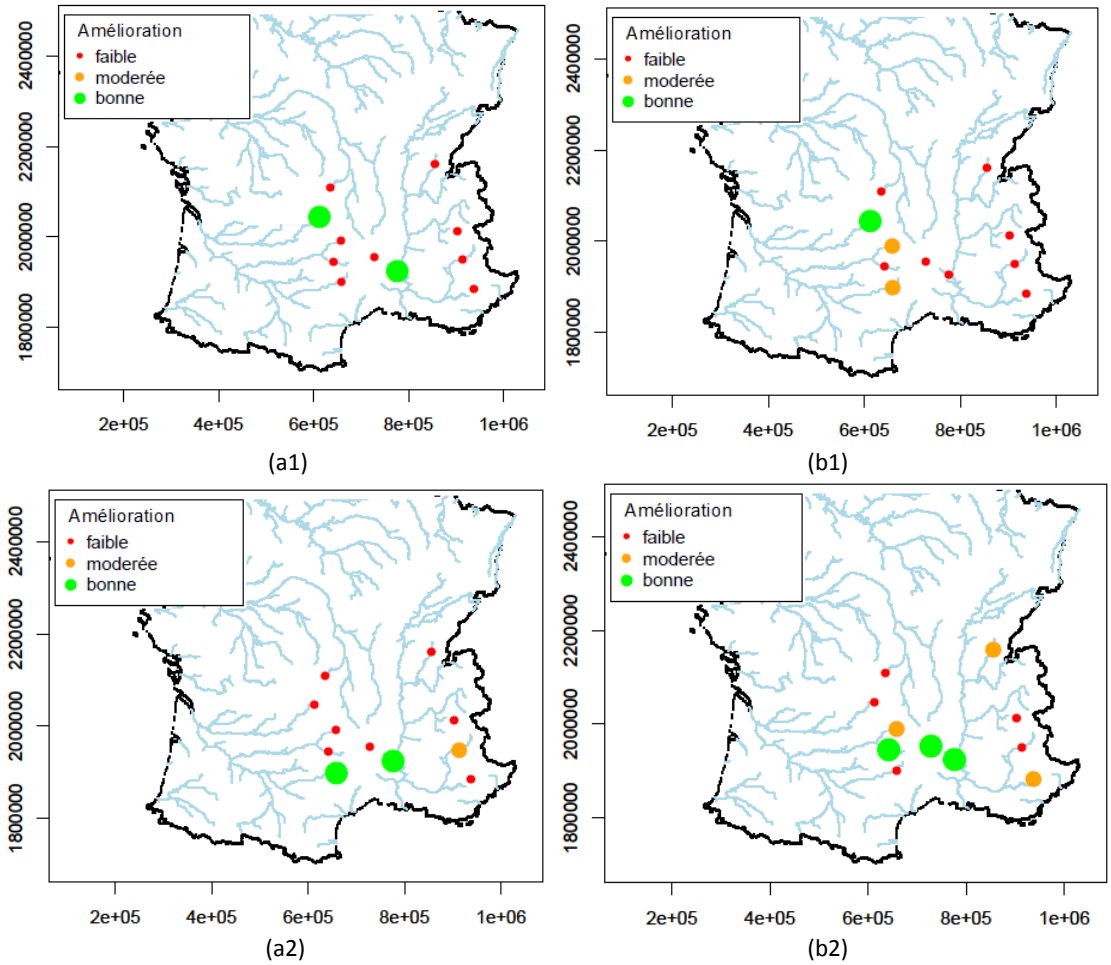


Figure 124 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS par la méthode prenant en compte l'erreur de 30% des meilleurs membres. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours.

6.5.3.4 La méthode du meilleur membre modifiée par Fortin et al. (2006)

Enfin nous avons testé la méthode du meilleur membre modifiée par Fortin *et al.* (2006). Rappelons que la méthode de Fortin *et al.* (2006) propose une pondération à chacun des membres selon son ordre statistique dans l'ensemble. Ceci permet une meilleure estimation des probabilités liées aux événements extrêmes.

Les critères examinés par la suite sont le diagramme de PIT (Figure 125), la dispersion des membres (Figure 126 et Figure 127), les courbes ROC (Figure 128), les scores CRPS (Figure 129 et Figure 130) et RPS (Figure 131).

Fiabilité : diagrammes de PIT

En examinant l'influence de l'approche (9) sur la fiabilité de prévisions en parallèle avec la performance des prévisions brutes (1) et celle de la méthode du meilleur membre modifiée (6), nous constatons que :

- pour les prévisions hydrologiques produites par le modèle hydrologique *MORDOR,* la méthode modifiée (9) n'apporte pas d'amélioration supplémentaire par rapport à la méthode modifiée par Wang et Bishop (6).
- dans le cas du modèle *GRP* cette approche (9) améliore considérablement la fiabilité de prévisions par rapport aux prévisions brutes (1) et corrigées (6). Une explication possible est liée à la forte tendance de prévisions tirées par le modèle *GRP* à sous-estimer l'observation, ainsi qu'à la capacité limitée de la méthode modifiée par Wang et Bihop (6) à réduire cette tendance. Dans le cas du modèle *MORDOR* cette tendance est relativement moins forte et cela peut expliquer la différence en termes d'amélioration apportée.

Les conclusions sont similaires pour presque tous les 11 bassins versants étudiés.

Dispersion des membres

Dans les Figure 126 et Figure 127 nous présentons l'évolution de la dispersion des membres en fonction de l'échéance pour le modèle *MORDOR* et *GRP*.

- Dans le cas du modèle MORDOR nous remarquons que l'évolution de la dispersion avec la méthode modifiée par Fortin *et al.* (9) est similaire à celle de la méthode modifiée par Wang et Bishop (6). Plus particulièrement, nous remarquons que sur l'ensemble des bassins versants la dispersion des membres avec le scénario (9) est en général inférieure de 10 à 20% à celle du scénario (6) relativement sur toutes les échéances – avec l'exception du Verdon à Castillon et de la Sioule aux Fades.

- Dans le cas du modèle *GRP*, lors des premières échéances de prévision, la modification par Fortin *et al.* (9) contribue à une augmentation significative de la dispersion des membres pour tous les 11 bassins versants – entre 20 et 50% selon le bassin. Cet effet se limite pratiquement qu'aux premières échéances (jusqu'à 3 à 4 jours), quand la sous-dispersion de l'ensemble est fortement pénalisante, et il se réduit aux échéances plus longues de prévision. Par conséquent, la dispersion produite par le scénario (9) aux longues échéances est en général 10 à 30% supérieure à celle du scénario (6).

Discrimination : courbes de ROC

En examinant l'influence de la modification par Fortin *et al.* (9) sur la capacité des prévisions à discriminer entre les événements (Figure 128), nous constatons que :

- en ce qui concerne le modèle *MORDOR* la méthode modifiée par Fortin *et al.* (9) influence peu la discrimination des prévisions par rapport à la méthode de Wang et Bishop (6).
- dans le cas du modèle *GRP*, le scénario (9) améliore considérablement la discrimination des prévisions. Cet effet est plus fort aux premières échéances de prévisions et il se réduit quand le seuil de non-dépassement passe de 30 à 90%.

Cette conclusion est répandue sur l'ensemble de 11 bassins versants.

Qualité globale

Dans les Figure 129 et Figure 130 nous examinons la qualité gloabale en termes de valeurs moyenne de *CRPS* des prévisions pour les scénarios (1), (6) et (9), les échéances 3 et 7 jours et les modèles *MORDOR* et *GRP* respectivement. En résumé :

- pour le modèle *MORDOR*, nous constatons que la performance des prévisions traitées avec le scénario (9) n'est pas améliorée par rapport au scénario (6). Ce comportement se répete à toutes les échéances.
- dans le cas du modèle *GRP* l'amélioration à la qualité globale des prévisions est significative pour presque tous les bassins versant et échéances examinés. Ce comportement est plus évident aux horizons lointains de prévision.

Enfin un aperçu de l'amélioration apportée à la qualité des prévisions brutes par l'habillage avec la méthode du meilleur membre modifiée par Fortin *et al.* (9) est présenté dans la Figure 131. En accord avec les conclusions tirées dans la Figure 129, nous constatons que :

- dans le cas du modèle *MORDOR*, le scénario (9) améliore la qualité dans un nombre limité des bassins versants principalement situé dans le Massif Central et les Cévennes. Cette amélioration se limite sur les horizons courts des prévisions.
- en ce qui concerne le modèle *GRP*, nous observons une forte amélioration, qui devient plus considérable aux horizons plus lointains. Cette amélioration se concentre surtout aux bassins du Massif Central, du Juras et des Alpes.

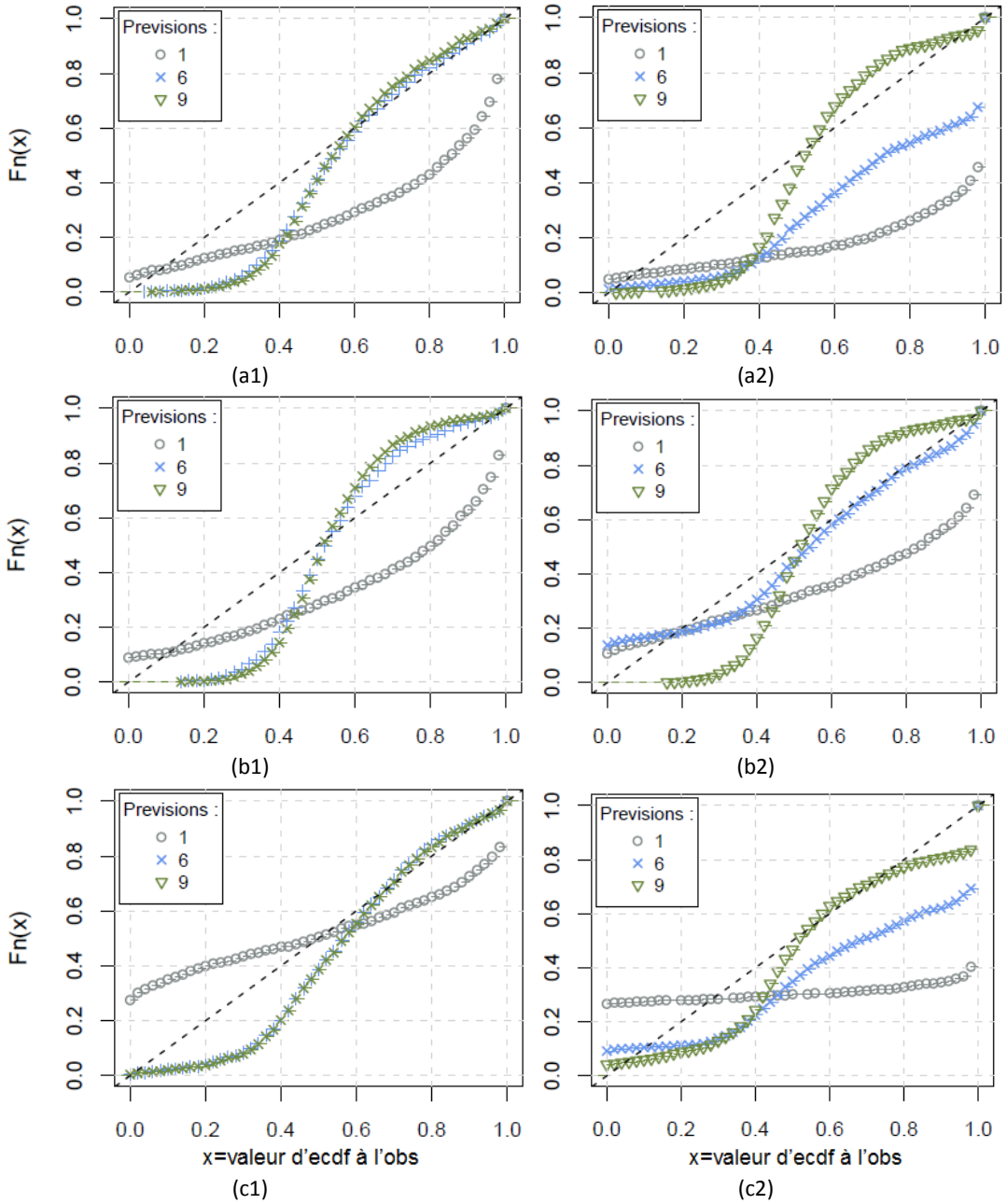


Figure 125 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifée par Fortin et al. (2006) et fournies par le modèle hydrologique : (1) MORDOR et (2) GRP.

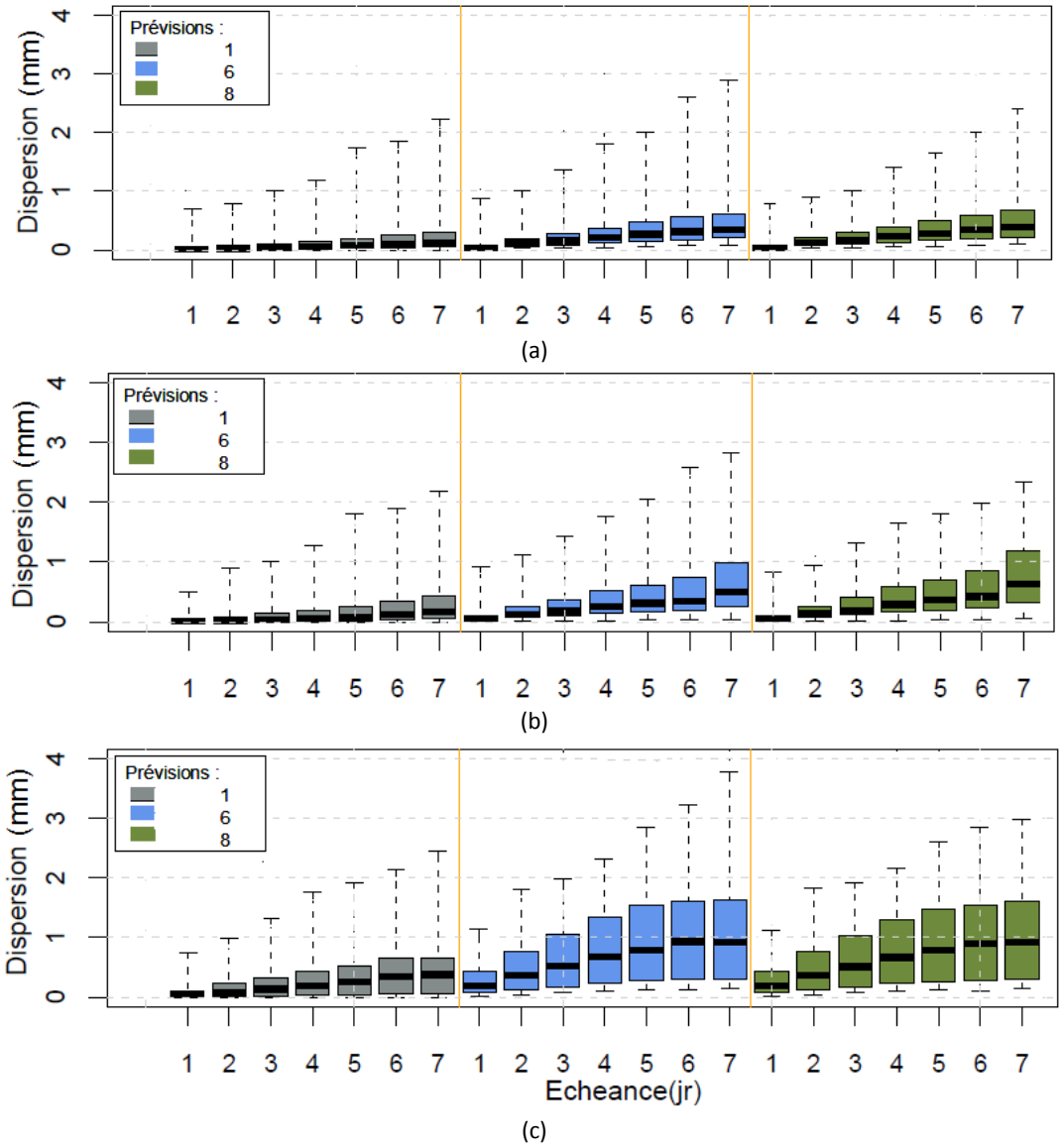


Figure 126 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et Fortin et al. (9) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

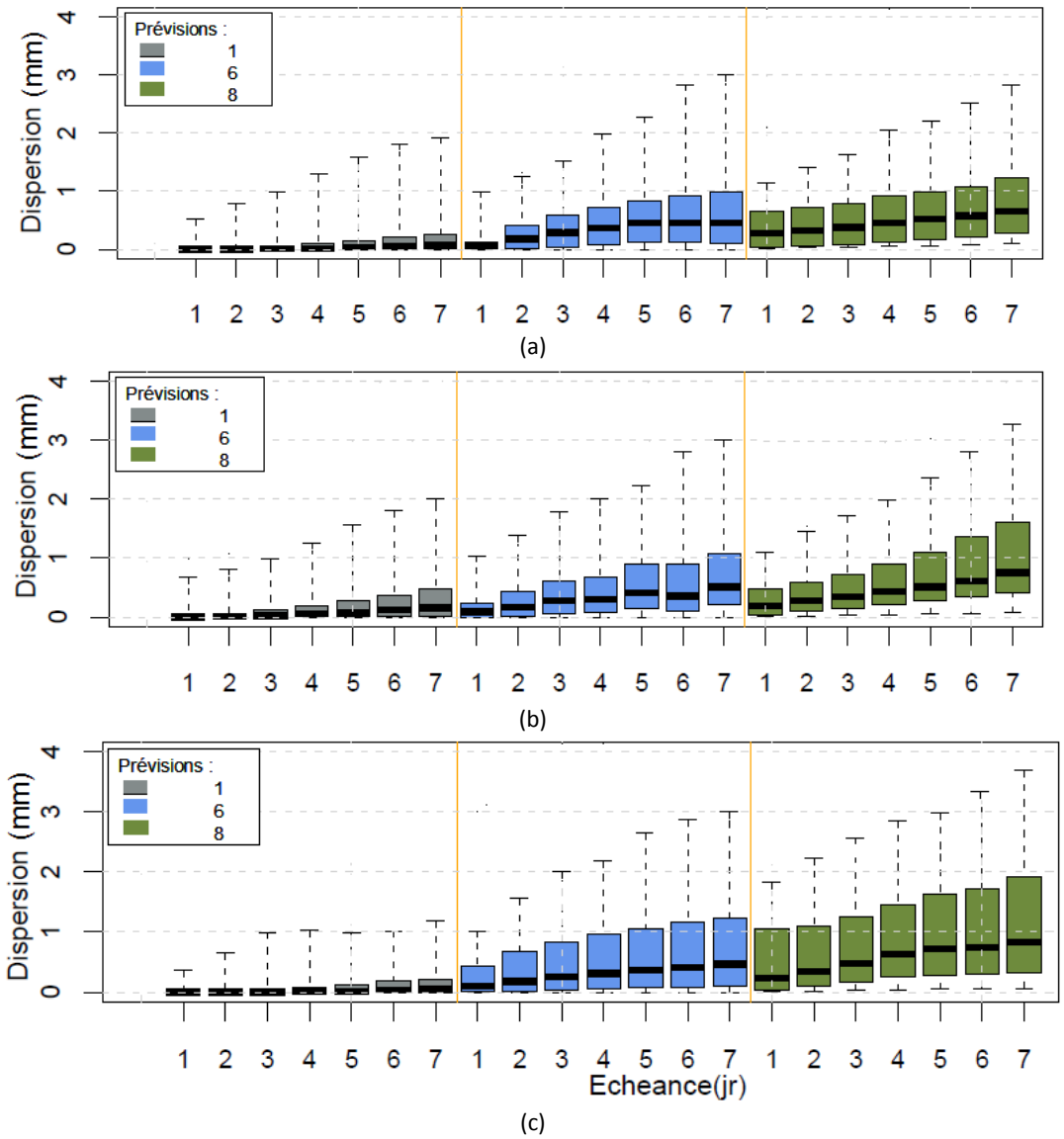


Figure 127 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches.

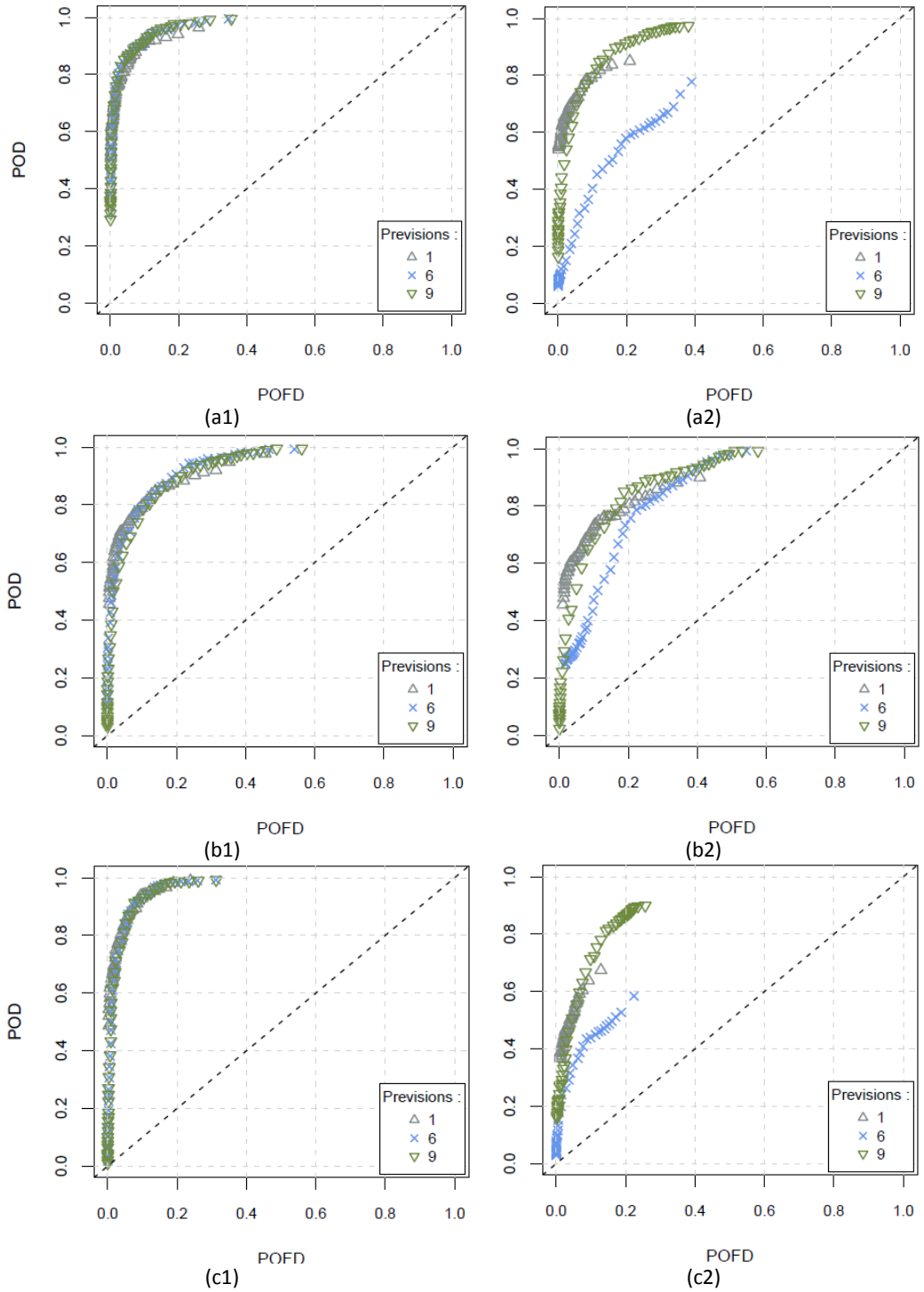


Figure 128 : Les courbes ROC pour un seuil de dépassement égal au quantile de 70% pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les courbes ROC sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Fortin et al. (2006) produites par le modèle hydrologique : (1) MORDOR et (2) GRP.

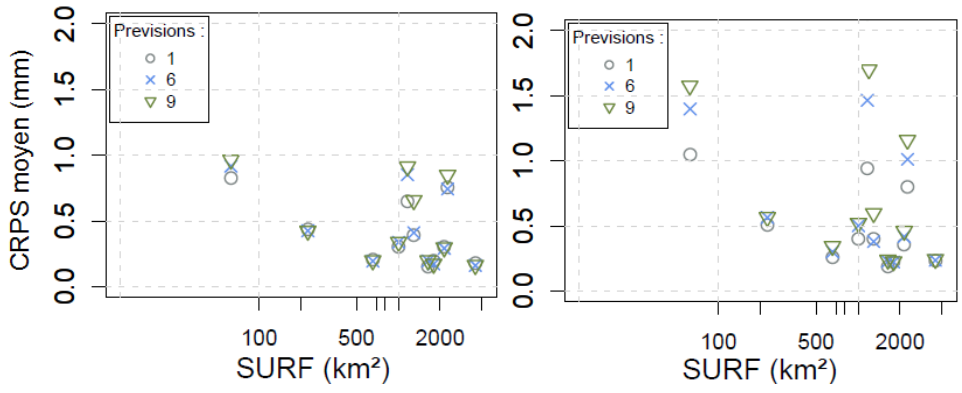


Figure 129 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9).

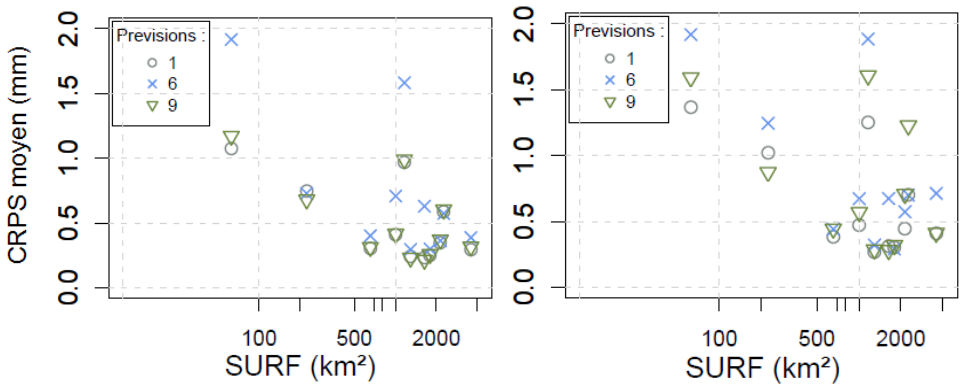


Figure 130 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9).

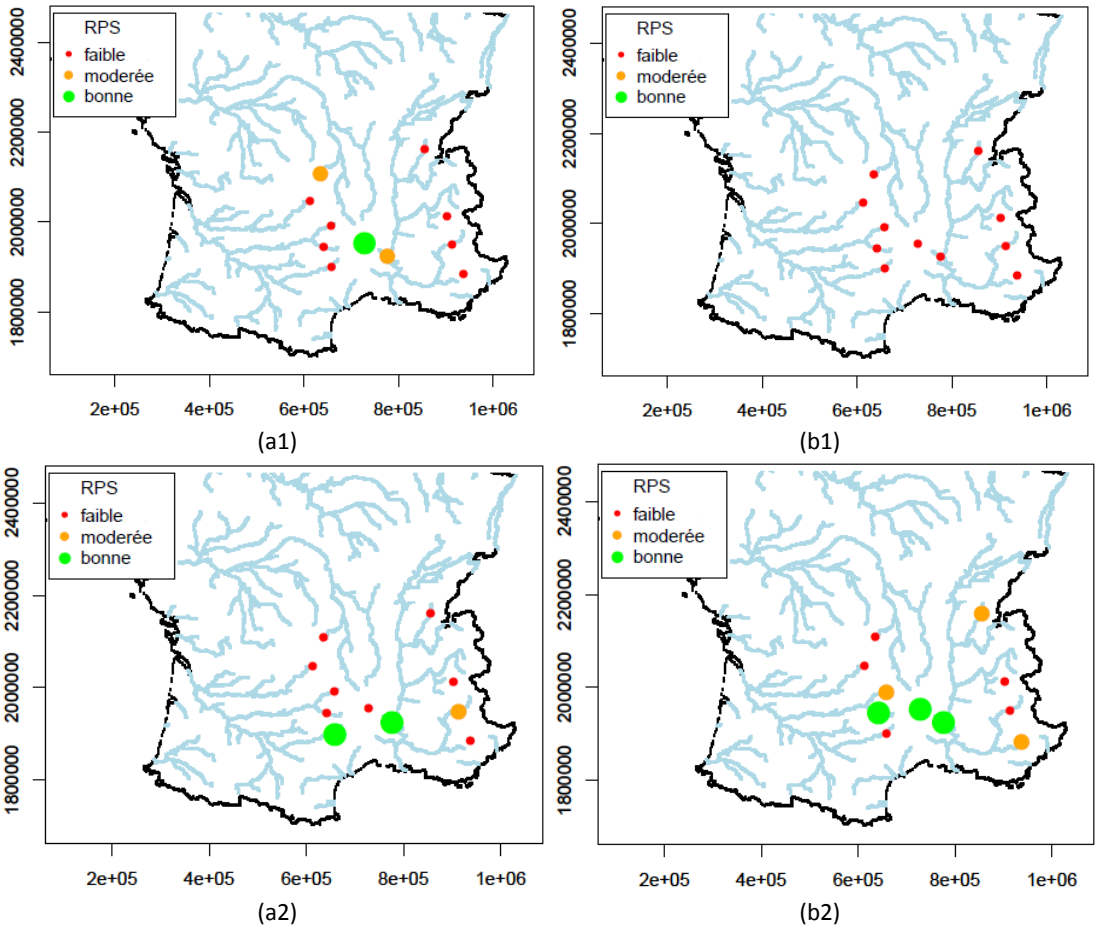


Figure 131 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS par la méthode du meilleur membre modifiée par Fortin et al. (2006). Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours.

6.5.4 Y-a-il une méthode de correction statistique qui soit la meilleure partout ?

Après avoir testé différentes méthodes de traitement statistique des biais des prévisions hydrologiques, nous allons essayer dans ce paragraphe d'examiner s'il y a une méthode notablement plus performante que toutes les autres pour tous les bassins versants et échéances de prévisions examinés. Une façon de répondre à cette question est d'examiner la performance de toutes les approches par rapport à la qualité globale des prévisions. Pour cette raison, nous avons choisi ici d'examiner le critère numérique *RPS*. Rappelons qu'il s'agit d'un critère qui évalue la qualité globale des prévisions et que le score parfait est égal à 0.

La Figure 132 illustre, sous forme de carte, les deux méthodes qui performent le mieux en termes de qualité des prévisions par bassin versant de notre échantillon et pour les prévisions hydrologiques d'ensemble issues, respectivement, du modèle *MORDOR* et du modèle *GRP* pour les échéances de 3, 5 et 7 jours. Nous remarquons d'abord qu'il n'y a pas une "meilleure" méthode pour tous les bassins et toutes les échéances de prévision. En effet selon le critère de probabilité ordonée *RPS*, plusieurs méthodes fonctionnent bien selon le bassin et l'échéance étudiée. Par exemple, pour le modèle *MORDOR* et pour l'échéance de prévision égale à 3 jours : l'habillage avec l'erreur du modèle - combiné ou pas avec l'information des analogues -(scénarios 3 et 4, Tableau 8), la méthode du meilleur membre modifiée par Wang et Bishop (scénario 6, Tableau 8) et ses variantes (scénario 7 et 8, Tableau 8) apparaissent comme des méthodes les plus performantes sur, respectivement, 9, 3 et 7 bassins versants de l'échantillon étudié.

Cependant aux horizons plus lointains de prévisions, les variantes de la méthode du meilleur membre modifiée et ses variantes (6 et 7) apparaisent plus performantes. Nous constatons que ces deux approches se montrent les plus performantes sur quelques bassins versants des Alpes et du Massif Central. De plus, l'habillage avec l'erreur du modèle - combiné ou pas avec l'information des analogues - (scénarios 3 et 4) apparaît être la "meilleure" méthode pour les bassins versants situés dans le Massif Central, les Cévennes et le Jura.

Pour le modèle *GRP*, nous remarquons également qu'il n'y a pas une méthode qui est la plus performante. Lors de premières echeances l'habillage avec l'erreur du modèle - combiné ou pas avec l'information des analogues - (scénarios 3 et 4), la méthode du meilleur membre (scénario 5) et l'habillage avec l'erreur des 30% des meilleurs membres (scénario 8) se montrent les plus performantes. Dans les echeances plus lointaines, l'habillage avec l'erreur de 30% des meilleurs membres (scénario 8) est remplacé par la methode du meilleur membre modifiée en prenant en compte la covariance des échéances (scénario 7).

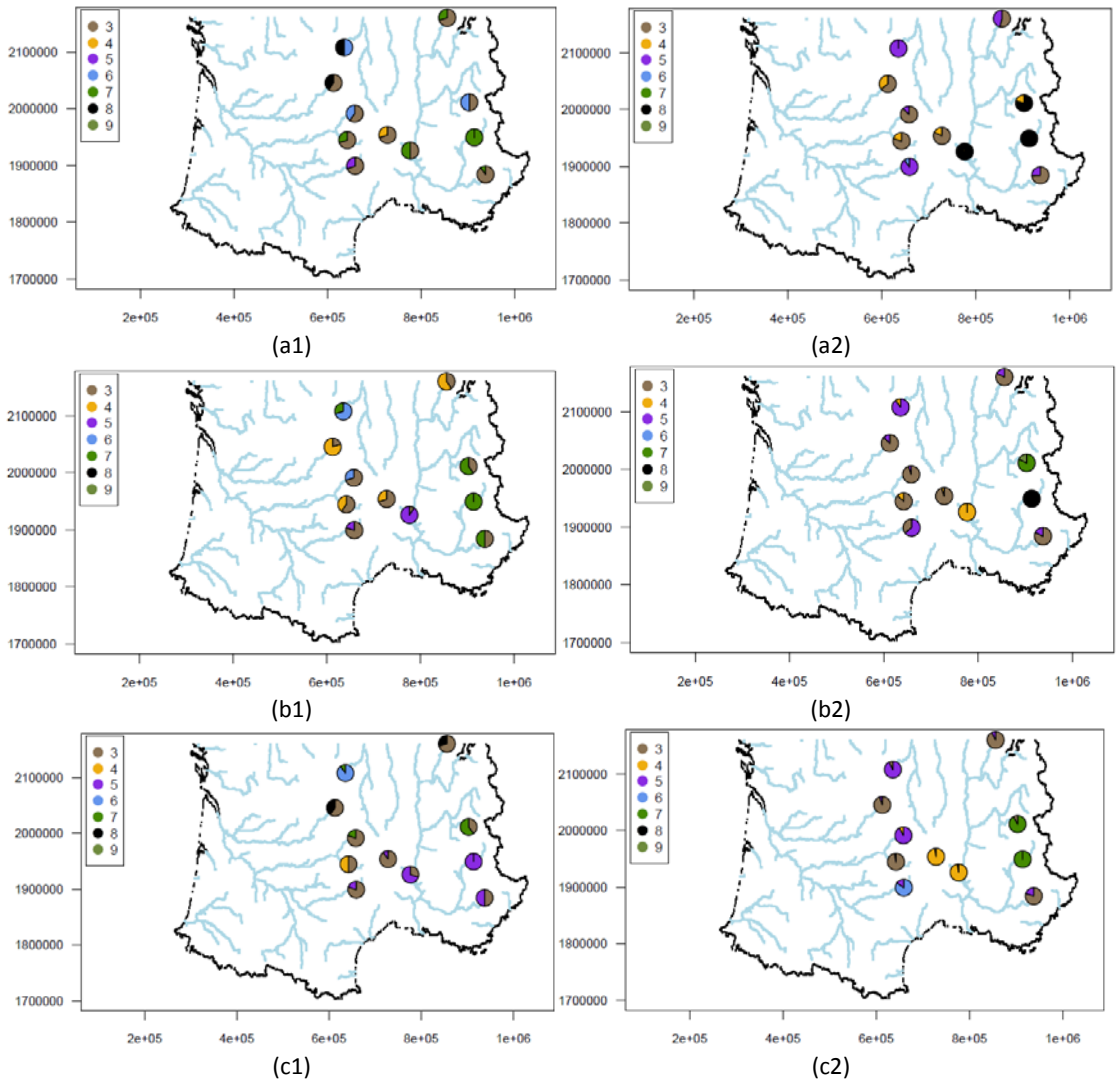


Figure 132 : Carte avec la méthode correspondant à la meilleur performance par bassin versant selon le critère RPS et pour une échéance de prévisions égale à (a) 3, (b) 5 et (c) 7 jours. Ici les prévisions hydrologiques sont fournies par les modèles MORDOR (1) et GRP (2).

En ce qui concerne la méthode modifiée par Fortin *et al.* (2006) nous observons qu'en comparaison avec les autres approches elle se montre parmi les plus performantes que dans le cas du bassin de la Romanche aux Chambons avec des prévisions hydrologiques tirées du modèle *GRP*. Toutefois, nous notons que l'application de méthode n'est pas toujours facile, car - en attribuant un poids à chaque membre - elle nécessite une archive longue. Par conséquent, nous considérons que l'application de cette approche dans une archive plus longue des prévisions apportera des résultats améliorés (voir Boucher *et al.*, 2011).

Pour conclure nous pouvons dire qu'il n'y a pas de "meilleure" méthode de correction statistique présentant la meilleure performance en termes des prévisions de débits pour tous les modèles, tous les bassins versants et toutes les échéances. Plusieurs méthodes ont des performances comparables et la "meilleure" ne surclasse que très légèrement les autres. Toutefois nous constatons qu'en général l'habillage avec l'erreur du modèle - combiné ou pas avec l'information des analogues – se montre plus performant dans les bassins du Massif Central et des Cévennes. Par contre, les variantes de la méthode du meilleur membre (scénarios 7 et 8) sont plus performantes dans les bassins des Alpes et du Jura. Ces observations se confirment pour les deux modèles hydrologiques et pour la plupart des échéances examinées et elles indiquent une tendance existante. Bien évidemment, cette tendance reste à confirmer sur une série plus longue de prévisions et sur d'autres jeux de données.

De plus, comme montré dans les paragraphes précédents, parmi les méthodes testées, il y a quelques approches qui arrivent à améliorer considérablement un attribut de la qualité des prévisions, tandis qu'un autre attribut est plutôt amélioré par une autre méthode. Par exemple, la fiabilité de prévisions est souvent améliorée par : l'habillage avec l'erreur du modèle (scénario 3), la méthode du meilleur membre modifiée (scénario 6), ou la méthode du meilleur membre modifiée avec la covariance des échéances (scénario 7). D'autres approches ont permis d'améliorer la capacité des prévisions à discriminer entre deux événements, comme c'est le cas de l'habillage avec l'erreur de 30% des meilleurs membres (scénario 8 et 9). Enfin, certaines méthodes ont permis d'améliorer significativement la qualité globale des prévisions : c'est le cas de la méthode du meilleur membre modifiée (scénarios 6 et 9). Par conséquent, la méthode à choisir dépend du point de vue de l'utilisateur et selon le barycentre de son intérêt il faut qu'il se dirige vers le traitement statistique le plus approprié.

6.5.5 Influence de la longueur de calage des méthodes de correction de biais

Une question qui se pose quand nous traitons le sujet de la correction des biais est celle de la longueur de la période de calage nécessaire pour définir les paramètres des méthodes de post-traitement statistique. En effet, dans la pratique, il n'est pas toujours évident de pouvoir disposer de longues périodes d'observations et de prévisions archivées pour pouvoir caler les méthodes de correction de biais. Dans ce paragraphe, nous allons examiner la sensibilité de la performance d'une méthode à sa longueur de période de calage. Nous examinons cette question ici en considérant deux des méthodes empiriques de post-traitement appliquées dans ce chapitre :

- la méthode d'habillage avec les erreurs de simulation du modèle hydrologique (méthode qui considère les incertitudes de la simulation hydrologique),

- la méthode d'habillage avec les erreurs de prévision du meilleur membre (méthode qui considère les incertitudes totales de prévision).

Pour la première méthode, nous étudions l'influence de la longueur de la période utilisée pour construire l'archive des erreurs de modélisation sur la valeur du critère de performance globale *CRPS*. Rappelons qu'ici la méthode d'habillage avec les erreurs de simulation du modèle hydrologique utilise une archive des erreurs du modèle hydrologique calculée sur une période de 35 ans, période pour laquelle nous avons des données de débit disponibles. Ensuite les erreurs archivées sont appliquées sur une période indépendante, plus courte, de validation (ici la période de 4 années, 2005 à 2008).

Dans la Figure 133 nous comparons la performance en termes de qualité globale de l'habillage avec l'erreur du modèle hydrologique *GRP* quand la longueur de calage est soit de 35 ans, soit de 4 ans. La comparaison est faite sur la base des valeurs moyennes du score *CRPS* et les scores sont présentés ici en fonction de la superficie de chaque bassin versant. Nous remarquons que pour la plupart des bassins versant examinés la performance de l'habillage avec l'erreur du modèle hydrologique calé sur 4 ans de données est considérablement inférieure à celle de la méthode calée sur 35 ans. Ce fait est plus évident sur les horizons plus lointains de prévision.

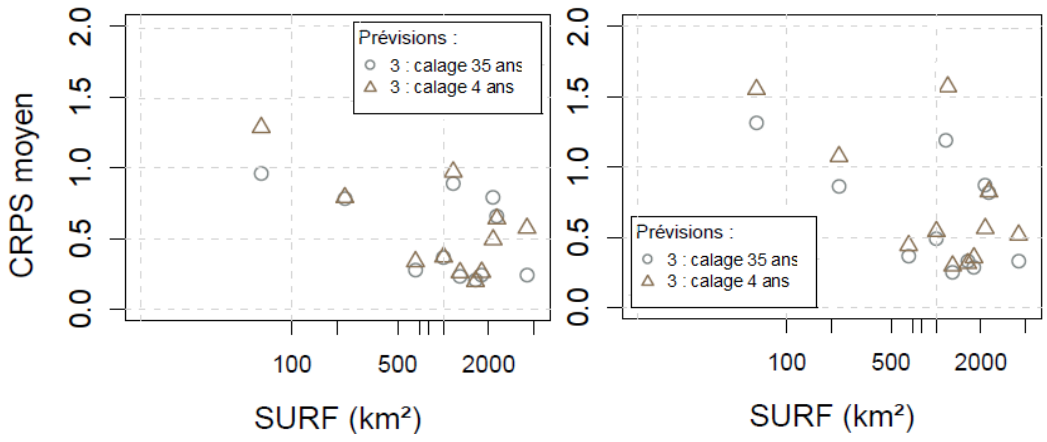


Figure 133 : Valeurs moyennes de CRPS en fonction de la surface du bassin pour l'ensemble de 11 bassins versants testés et pour les échéances : (a) 3 et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec l'erreur du modèle hydrologique avec un calage de longueur de 35 (cercle) et de 4 ans (triangle).

Nous remarquons alors que la longueur du calage influence la performance de la méthode de traitement statistique par habillage empirique, car quand la méthode est calée sur 4 ans les valeurs du score *CRPS* sont en moyenne supérieures de 10% à un horizon de prévision de 3 jours et de 25% à 7 jours. De plus, sur certains bassins versants cela va jusqu'à 40 et 50% pour un horizon de 3 et 7 jours respectivement.

La deuxième méthode ici considérée est celle de l'habillage empirique avec les erreurs passées du meilleur membre (Roulston et Smith, 2003) de la prévision d'ensemble. Rappelons que cette méthode est basée sur deux paramètres à caler : la moyenne et l'écart-type. Nous nous posons la question de l'influence de la période de calage sur l'estimation des paramètres de la méthode. Afin d'examiner la variation des paramètres de calage avec la longueur des séries des données disponibles, nous avons évalué les deux paramètres en utilisant des périodes de calage de différentes longueurs. La période totale de prévision disponible (2005 à 2008) a été divisée en sous-périodes glissantes (suivant le concept de fenêtres glissantes). Le Tableau 9 présente le nombre de sous-périodes dans chaque échantillon ainsi obtenu, chacune avec une longueur définie.

Tableau 9 : Les sous-périodes de calage en utilisant le concept de fenêtres glissantes : leur longueur en mois (en haut) et le nombre de périodes dans chaque sous-échantillon (en bas).

Longueur de la période (mois)	1	3	6	12	18	24	30	36	40	42
Nombre de périodes	42	40	37	31	25	19	13	7	3	1

Rappelons que la méthode du meilleur membre (Roulston et Smith, 2003) utilise l'archive des erreurs commises par le "meilleur" membre pour s'ajuster à une loi normale. Par conséquent, les deux paramètres de la méthode à caler sont les paramètres de la loi normale, à savoir : la moyenne et l'écart-type des erreurs. La Figure 134 présente, sous forme de boîtes à moustache, l'incertitude autour de l'estimation de la moyenne en fonction des différentes longueurs des sous-périodes utilisées, et pour les échéances de prévision de 3 jours et 7 jours. Nous remarquons que les valeurs moyennes des sous-périodes convergent rapidement vers la moyenne, avec des épaisseurs des boîtes à moustaches qui diminuent rapidement et peu de variabilité entre sous-périodes au-delà de 6-12 mois.

De même, dans la Figure 135 nous affichons les résultats concernant l'estimation de l'écart-type. Nous remarquons qu'à partir d'une longueur de calage de 12 mois, nous obtenons la valeur quasi-définitive du paramètre.

Par conséquent, nous pouvons conclure que la période d'évaluation disponible (2005-2008) est suffisamment homogène hydrologiquement pour pouvoir obtenir une bonne estimation sur les paramètres de la méthode du meilleur membre à partir d'une période de calage de 12 mois. Néanmoins, cette conclusion ne nous permet pas de transférer et utiliser ces mêmes paramètres avec de bons résultats sur une autre période d'évaluation hétérogène hydrologiquement.

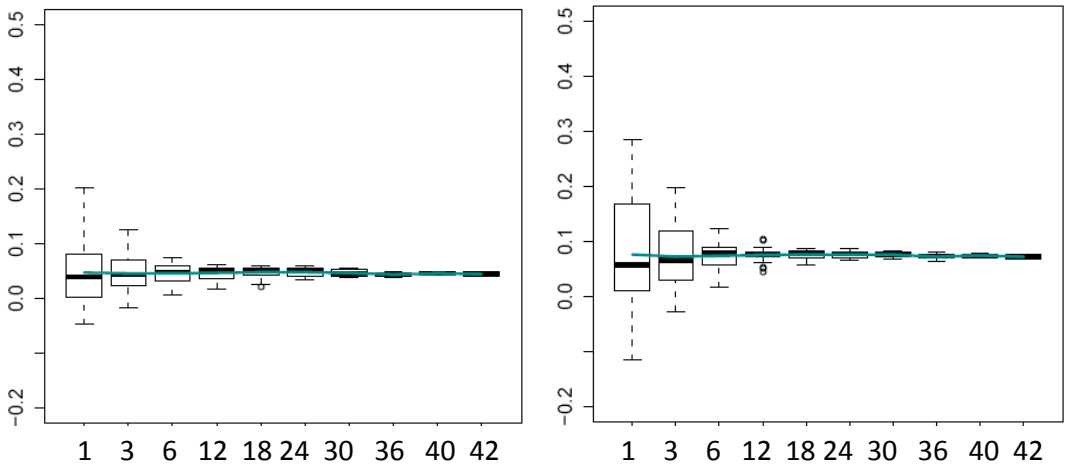


Figure 134 : Boîtes à moustache représentant l'incertitude de l'estimation de la moyenne des erreurs du meilleur membre pour les échéances : (a) 3 et (b) 7 jours. Les boîtes à moustaches sont calculées à partir des prévisions hydrologiques d'ensemble produite par le modèle GRP.

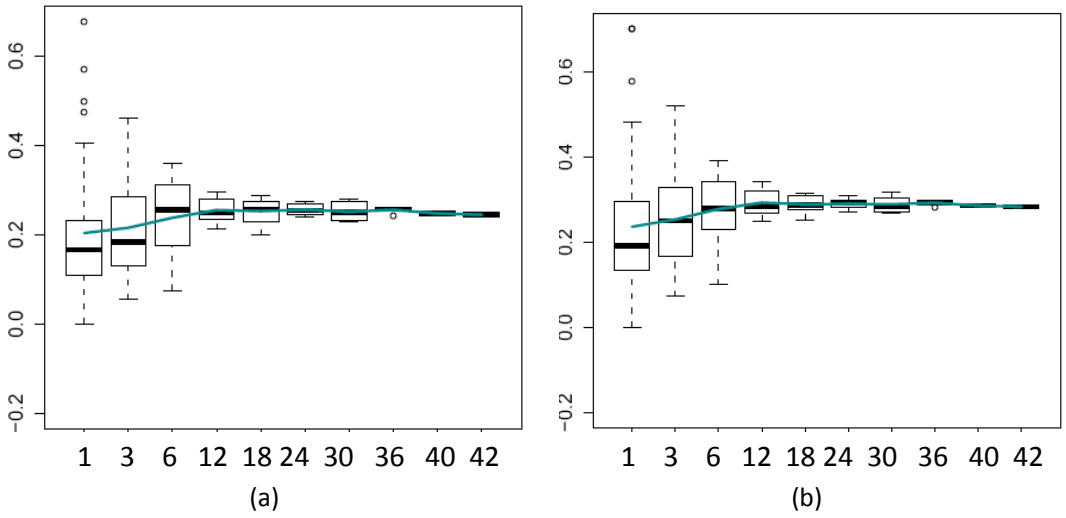


Figure 135 : Boîtes à moustache représentant l'incertitude de l'estimation de l'écart-type des erreurs du meilleur membre pour les échéances : (a) 3 et (b) 7 jours. Les boîtes à moustaches sont calculées à partir des prévisions hydrologiques d'ensemble produite par le modèle GRP.

6.6 Synthèse

Dans ce chapitre nous avons utilisé les prévisions météorologiques d'ensemble du centre européen (*CEPMMT*) et les deux modèles hydrologiques à notre disposition : *MORDOR* et *GRP*. Nous avons évalué les prévisions hydrométéorologiques d'ensemble brutes et ensuite nous avons essayé de répondre aux problématiques liées aux différent biais présents dans les prévisions brutes. Pour cela nous avons mis en place différentes approches de traitements statistiques appliqués soit sur les prévisions de pluie, soit sur celles de débit. Finalement, nous avons abordé quelques questions clés liées à la performance des approches de traitement statistique.

Pour revenir aux conclusions principales de ce chapitre, notre étude a montré en général que les prévisions de précipitation disposent d'une bonne fiabilité. Néanmoins, elles restent sous-dispersives surtout aux échéances courtes de prévision. De plus, nous observons qu'en général les prévisions d'ensemble de précipitation arrivent à bien discriminer entre les différents événements à un horizon court de prévision. Par contre pour les échéances plus lointaines et les seuils de non-dépassement plus élevés le nombre des membres qui arrivent à prévoir un événement avec une probabilité élevée se réduit.

En ce qui concerne les prévisions d'ensemble de débit nous avons basé notre étude sur : la performance des prévisions produites par le modèle *MORDOR* et celle des prévisions produites par le modèle *GRP*. Plus particulièrement, dans le cas du modèle *MORDOR* nous constatons que dans la plupart de bassins versants examinés nous retrouvons une tendance à sous-estimer l'observation. Cependant sur un nombre limité de bassins étudiés cette tendance s'inverse légèrement. Nous remarquons également que les prévisions hydrologiques se caractérisent par une capacité plus élevée de discriminer des événements entre eux par rapport aux prévisions de précipitation. En ce qui concerne la dispersion des prévisions hydrologiques, les conclusions sont similaires à celles tirées pour les prévisions de pluie, car elles restent sous-dispersives. Ce fait démontre que le modèle hydrologique utilisé (*MORDOR*) n'affecte pas considérablement la dispersion des membres de l'ensemble. En comparant la dispersion de membres contre le ratio de la racine carrée de l'erreur quadratique moyenne nous constatons qu'une partie limitée de l'erreur quadratique est expliquée par la dispersion, et cela arrive surtout aux horizons courts de la prévision. Cette observation contraste avec les mêmes diagrammes liés aux prévisions de précipitations et elle est indicative d'une source d'erreur supplémentaire (autre que la dispersion non appropriée) qui affecte les prévisions hydrologiques.

Pour les prévisions produites par le modèle *GRP*, nous remarquons que, même si la tendance de sous-estimer l'observation est toujours présente, la modélisation des débits prévus est différente pour les 2 modèles hydrologiques utilisés. En général nous constatons que la présence d'un biais est plus forte quand la modélisation est faite avec *GRP*. La différence la plus frappante se trouve sur le bassin de la Romanche aux Chambons, où nous observons que la fiabilité de prévisions hydrologiques est nettement dégradée par rapport à celle de prévisions météorologiques, ainsi que celle des prévisions hydrologiques produites par le modèle *MORDOR*. En ce qui concerne la dispersion des membres, nous remarquons la dispersion des membres est également très faible. Cela signifie que le modèle *GRP* écrase la dispersion des membres d'ensemble. Par conséquent la partie de la racine carrée de l'erreur quadratique moyenne normalisée (*RMSE*) expliquée par la dispersion non-appropriée est considérablement plus importante que dans le cas de prévisions de débits produites par le modèle *MORDOR*. Cette conclusion est également confirmée par Randrianasolo et al. (2010) qui a comparé le modèle *GRP* avec le modèle *SIM*.

Une fois les différents biais de prévisions identifiés les différentes approches de traitement statistique ont été mises en place. Plusieurs approches ont été testées traitant soit de l'incertitude globale de prévisions, soit d'une source particulière d'incertitude. Ces approches sont soit tirées de la littérature existante, soit développées lors de cette thèse. Le Tableau 10 présente les approches mises en place et il résume leur performance par rapport aux divers attributs de la qualité de prévision. Nous avons constaté que chaque approche traite un ou plusieurs attributs de la qualité. Sur l'échantillon de 11 bassins versants et avec la période d'évaluation à notre disposition, nous n'avons pas pu conclure sur une meilleure méthode pour tous les modèles, tous les bassins versant et toutes les échéances. Cependant, nous avons identifié quelques tendances. De plus, nous avons constaté que selon les objectifs d'une étude et les données disponibles il y a des méthodes de traitement plus ou moins adaptées.

Dans le cadre de cette étude nous avons essayé aborder un nombre d'autres problématiques liées au traitement statistique de prévisions : l'intérêt d'appliquer plutôt un traitement en amont (pré-) ou en aval (post-) du modèle hydrologique, ou encore la sensibilité de la performance du traitement statistique à la longueur du calage.

Tableau 10 : Illustration comparative entre les différentes approches de traitement statistique utilisées en marquant les attributs de la qualité des prévisions hydrologiques que chaque approche améliore.

Approche de traitement statistique testé	*Qualité globale*	*Fiabilité*	*Discrimination*	*Dispersion*	*Pré- ou post-traitement*	*Sensibilité à la longueur de calage*	*Limitations*
Correction empirique par facteur multiplicatif	+++	+	+++	+	**Pré-**traitement	√	**performance dépendante de l'échéance et des bassins versants*
Habillage avec l'information des analogues	+++	+	+++	+++	**Pré-**traitement	√	**pré-requis d'une longue archive d'analogues* **les scénarios originaux de la prévision d'ensemble ne sont pas tous gardés* **effet limité sur les débits* ** performance influencée par la mise à jour du modèle*
Habillage avec l'erreur du modèle hydrologique (Mathevet, 2010 ; Pagano et al., 2012)	+	+++	+	+++	**Post-**traitement	√	**performance dépendante de l'échéance et des bassins versants*

Méthode du meilleur membre (Roulston et Smith, 2003)	+	+	+	+++	**Post-**traitement	√	**performance dépendante du modèle hydrologique utilisé et du bassin versant*
Méthode du meilleur membre modifiée (Wang et Bishop, 2005)	+	+++	+	+++	**Post-**traitement	√	**performance dépendante du modèle hydrologique utilisé et du bassin versant*
Méthode du meilleur membre modifiée pour prendre en compte la cohérence temporelle	+++	+++	+++	+++	**Post-**traitement	√	**performance dépendante du modèle hydrologique utilisé*
Habillage avec l'erreur de 30% des meilleurs membres	+++	+	+++	+	**Post-**traitement	√	**performance dépendante du modèle hydrologique utilisé*
Méthode du meilleur membre modifiée (Fortin *et al*., 2006)	+	+++	++	+++	**Post-**traitement	√	**performance dépendante du modèle hydrologique utilisé*

+++ : amélioration significative

+ : amélioration faible

√ : vrai

- : faux

"Hydroélectricité : Vous démarrez au quart de tour et, à trop vouloir en faire, vous perturbez votre entourage. Un peu de mesure''

("Énergiescope" publié dans le numéro spécial de Metro du 5 juin 2012)

7 Valorisation des prévisions d'ensemble

Dans ce chapitre, nous nous intéressons à évaluer le gain apporté par les prévisions, plus particulièrement les prévisions probabilistes (d'ensemble), sur la gestion de la production d'énergie hydroélectrique d'une retenue en termes d'impact économique. Le but est de quantifier la valeur d'une prévision pour la gestion des apports en eau aux barrages-réservoirs. Nous allons d'abord rappeler les concepts clés de la gestion des apports aux aménagements hydroélectriques. Ensuite, nous décrivons les étapes de construction d'un modèle de gestion heuristique. Enfin, les résultats de l'application du modèle de gestion aux prévisions d'ensemble sont présentés et discutés.

7.1 La problématique en jeu : la gestion d'une réserve hydroélectrique

D'après Roche et al. (2012), jusqu'en 1950 les réservoirs des aménagements hydrauliques français étaient essentiellement des réservoirs mono-usages. L'usage multi-objectif des réserves en eaux a cependant été progressivement propulsé par les demandes en eaux croissantes et la mise en place de plusieurs grands projets d'aménagement. Le contexte multi-usage introduit forcement des nouvelles contraintes de gestion, non seulement hydraulique, mais aussi financière des aménagements. La gestion intégrée des ressources en eau peut concerner plusieurs applications : la distribution d'eau potable, la fonctionnalité des réseaux de collecte des eaux pluviales, la gestion des eaux usées, etc. Ici, nous nous focalisons sur la gestion d'une réserve hydroélectrique.

La première étape vers une gestion optimisée d'une réserve en eaux passe par la structuration du problème de gestion. Ce problème se résume souvent par un modèle physique basé sur la conservation de la masse. Afin de bien structurer le problème de gestion, les variables descriptives de l'état du système et les contraintes que les gestionnaires sont obligées de respecter doivent être définies. Ici, nous allons considérer un schéma simplifié d'une réserve en eau superficielle destinée à la production hydroélectrique, tel que celle présentée dans la Figure 136. Les principales variables descriptives de l'état du système sont:

- le stock du réservoir (S) ;
- le débit d'entrée ($Q_{entrée}$) ;
- le débit turbiné ($Q_{turbiné}$) ;
- le débit déversé ($Q_{deversé}$) ;
- le débit réservé ($Q_{reservé}$).

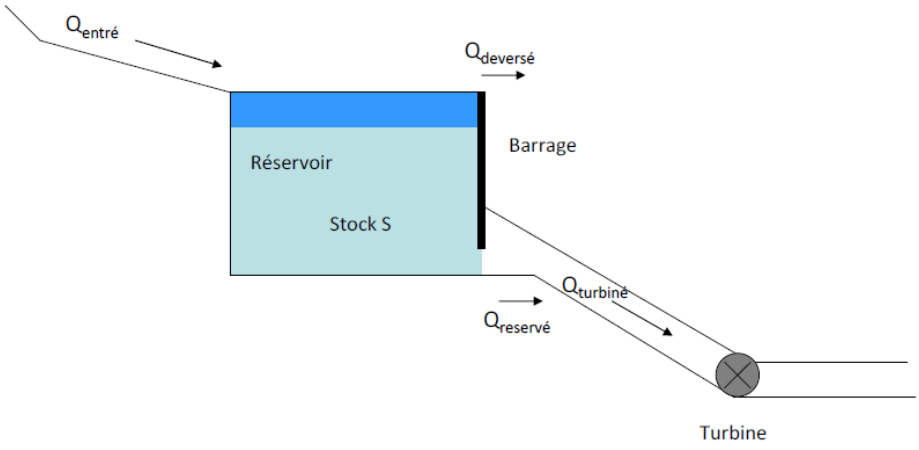


Figure 136: Schématique simplifié d'une réserve hydroélectrique.

A ces variables, diverses contraintes s'imposent pour le bon fonctionnement de l'aménagement. D'abord, pour la sécurité du barrage, le volume du stock maximal ne doit pas dépasser une valeur maximale *Smax*. De plus, des quotas environnementaux, touristiques ou de navigation vont souvent dicter également un stock minimum et/ou maximal (*Smin* et *S'max < Smax*). En outre, il faut aussi respecter la puissance maximale des turbines *Tmax*, ainsi que la limite de débit maximal déversé, pour laquelle l'évacuateur des crues a été dimensionné (*Dmax*). De plus, les quotas environnementaux imposent la sortie d'un débit minimal réservé pour des raisons écologiques (Qréservé). Ici pour de raisons de simplification le débit réservé (Qréservé) est assumé égal à 0.

Généralement, aux variables et contraintes, s'ajoutent les besoins des autres utilisateurs de l'eau réservée (réseaux de distribution d'eau potable, irrigation, etc.), ainsi que des objectifs d'écrêtement de crue et/ou soutien d'étiage. Une particularité du réseau français est liée au parc nucléaire, qui se trouve souvent en aval des aménagements hydroélectriques et qui produit 78% de l'électricité du pays. Pour les besoins du parc nucléaire, les gestionnaires des aménagements peuvent aussi tenir compte d'un débit minimal pour garantir le refroidissement des eaux sortantes des centrales nucléaires et pour maintenir la température de l'eau des rivières à des niveaux acceptables.

Les contraintes citées ci-dessus représentent les contraintes dites physiques de la gestion. Néanmoins, la loi du marché impose des contraintes supplémentaires à respecter. Ce marché, censé satisfaire la demande en énergie au moindre coût global, incite chaque producteur à produire lorsque la marge entre les prix et son propre coût de production est la plus importante.

Bien évidemment, dans un aménagement hydroélectrique, la définition des variables, contraintes et objectifs de gestion dépend de l'utilisateur. Souvent la variable qu'un gestionnaire considère comme variable de décision devient une contrainte pour un autre gestionnaire. Cependant, de manière générale, les problèmes de gestion auxquels sont souvent confrontés les services opérationnels se résument à des problèmes d'optimisation, souvent appelés aussi "problèmes de contrôle". Il s'agit alors de simuler le système (entrées, variables, contraintes) et rechercher les meilleures règles de décision de façon à se rapprocher au mieux des objectifs de gestion. Gérer un aménagement hydroélectrique revient alors à "décortiquer le système contrôlable, ses entrées et les usages, puis à les modéliser" (Roche et al., 2012).

La modélisation du système physique est constituée d'une série d'équations décrivant, à chaque instant, les relations entre les variables du système. Par exemple, un gestionnaire peut souhaiter que le stock du réservoir à un instant t soit identique au stock à l'instant précédent, *t-1*. En appliquant le principe de conservation de la masse - sous forme de débit

entrant et sortant-, nous arrivons à structurer l'équation décrivant une telle gestion d'un aménagement hydroélectrique (sans autres usages) pour un instant t donné :

$$\Delta S = S(j-1) - S(j) = Qentrée(j) - Qturbiné(j) - Qdeversé(j) - \sum Q_{divers} = 0 \qquad \text{Eq. 51}$$

Cette dernière équation résume l'aspect physique du problème de gestion. Elle permet de calculer, à chaque pas de temps, les variables descriptives de l'état du système modélisé. A côté de la modélisation du système physique, il existe le problème de la gestion de l'aménagement hydroélectrique selon les objectifs économiques de gestion, tout en respectant les contraintes qui s'imposent au système. Dans ce cas, à chaque pas de temps, les variables descriptives de l'état du système sont recalculées en fonction des variables d'entrée (apports) et des décisions prises, celles-ci souvent encadrées par une fonction économique (de coût ou de bénéfice). Cette fonction est calculée à chaque pas de temps et la somme engendrée sur une période, suite aux apports et décisions prises, définit la "fonction objectif du problème de contrôle". Les enjeux d'une gestion, avec ses contraintes physiques et économiques, sont ainsi synthétisés par cette fonction que le gestionnaire cherchera à optimiser (Roche et al., 2012)

7.2 Un modèle heuristique de gestion des réserves hydroélectriques

7.2.1 Cas d'application : les éclusées hebdomadaires

Les trois grandes catégories d'aménagements hydroélectriques sont représentées dans la Figure 137, où quelques aménagements français sont caractérisés en fonction de leur capacité utile (indiquée en jours d'apport moyen en débit stockable par le réservoir) et de leur coefficient de turbinage (ratio entre le débit maximum qui peut être turbiné et l'apport moyen en débit). Nous observons que les usines au fil de l'eau correspondent à des aménagements caractérisés par une capacité utile inférieure à un jour (il n'y a pas – ou presque pas – de retenue d'eau et les apports amont sont turbinés au moment où ils arrivent, dans les limites de la capacité de l'usine) et une puissance de turbinage qui varie de 1 à 5 fois le débit moyen. Les aménagements avec retenue (éclusées et lacs) ont une capacité utile plus importante, de l'ordre de 1 à 15 jours pour les éclusées et supérieure à 15 jours pour les lacs. La capacité de turbinage est également légèrement croissante entre éclusées et lacs : entre 2 à 4 fois le débit moyen pour les premiers et entre 4 à 8 fois le débit moyen pour les lacs. Typiquement, les aménagements d'éclusée permettent de stocker l'eau arrivant durant les heures creuses de la semaine (nuit et week-end, par exemple) pour la turbiner pendant les heures pleines. Les capacités de stockage des lacs permettent quant à elles un transfert à l'échelle saisonnière.

Sachant que notre but est d'examiner la valeur potentielle des prévisions journalières, d'échéance maximale de sept jours, nous cherchons alors, pour servir de contexte d'application, des aménagements de taille intermédiaire, pouvant stocker des apports sur une durée de l'ordre de grandeur de la semaine. D'après les chiffres de la Figure 1, on observe que ceci correspond au cas des usines de production électrique du type *éclusée hebdomadaire.* N'ayant pas ici l'ambition de modéliser dans le détail toutes les contraintes d'un aménagement réel, nous travaillerons avec des aménagements fictifs simples, pour lesquels nous ferons varier la capacité utile et de turbinage dans les limites indiquées par les données de la Figure 137. Les apports en débits prévus et observés au droit de ces aménagements fictifs seront ceux des bassins versants réels étudiés dans cette thèse.

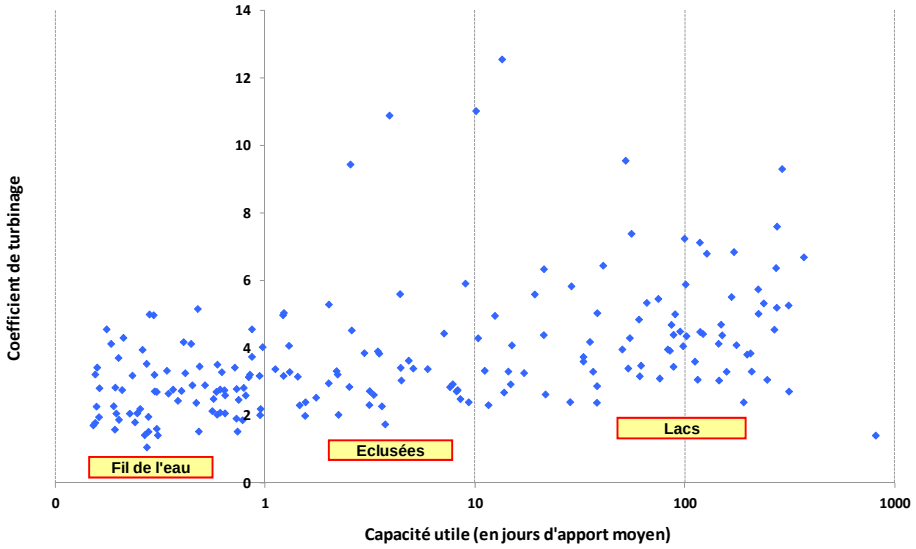


Figure 137: Classification générale des aménagements hydroélectriques en fonction du coefficient de turbinage et de la capacité utile (source des données : EDF).

7.2.2 Concept général de l'approche développée

Dans notre outil de gestion d'une éclusée hebdomadaire, nous nous intéressons à évaluer, à chaque jour de prévision, la meilleure règle de gestion à adopter face aux apports prévus en débits sur les sept jours à venir. Une *règle de gestion* se traduit par une programmation, pour la semaine à venir, des heures de turbinage et d'arrêt, dans le but d'optimiser la production électrique. Dans une logique d'optimisation des gains, il s'agit de choisir les heures de turbinage en fonction du prix horaire de l'électricité. Ce prix traduit la valeur de l'énergie. Il est variable dans le temps et suit les fluctuations de consommation, qu'elles soient journalières (heures de pointe du matin et du soir) ou hebdomadaires (différences entre jours ouvrables et week-end).

Dans notre cas, les *prix de l'énergie* sont supposés connus à l'avance et le coût marginal de production est supposé négligeable. En fonction des apports prévus et de la capacité de turbinage de l'usine, un seuil de prix est choisi. L'évolution temporelle des heures de turbinage ou arrêt est donnée par une règle simple : aux heures où le prix de l'énergie est supérieur à un prix-seuil, l'usine turbine à pleine puissance ; aux heures où le prix de l'énergie est inférieur au même prix-seuil, l'eau est stockée dans la retenue en attendant une remontée des prix pour être turbinée. La gestion est ainsi directement liée à la demande d'électricité : en période de faible demande (la nuit et le week-end), la centrale cesse de turbiner et l'eau est stockée dans la retenue (de l'énergie stockée qui pourra être utilisée plus tard lors des pics de consommation). En période de forte consommation d'électricité, où les prix sont plus élevés, la centrale turbine à sa capacité maximale *(Tmax)*.

La règle de gestion est définie quotidiennement, à chaque arrivée d'une prévision pour les sept jours à venir. Elle dépend également des *contraintes* induites par les caractéristiques de remplissage de l'aménagement : notamment, la capacité maximale du réservoir (*Smax*), au-delà de laquelle il faudra déverser l'eau excédentaire (l'eau déversée ne pourra pas être utilisée pour produire de l'énergie et représente donc une perte - ou un manque de gain - économique), et le volume minimal du réservoir (*Smin*), qui pourrait contraindre l'usine à arrêter le turbinage si le niveau du réservoir atteint cette limite.

Pour automatiser le processus, nous avons développé une méthode simplifiée, du type heuristique, capable de fournir une solution approximative satisfaisante pas nécessairement optimale, mais proche de cet optimum (voir §4.3). Pendant l'*étape de gestion*, la méthode procède par itérations successives, en essayant de garantir le respect des contraintes tout en restant optimale (ou quasi-optimale) dans l'établissement des meilleures heures à turbiner l'eau en stock ou les apports prévus. Une fois la règle de gestion établie, on procède à son application. Pour cela, la gestion doit être simulée avec les débits observés : *combien gagne-t-on réellement lorsque l'on turbine les débits observés aux heures programmées a priori par la gestion établie en fonction des apports prévus ?*

Puisque l'on rafraîchit quotidiennement les prévisions et, par conséquent, la règle de gestion, seules les premières 24 heures sont des *simulation***.** Nous considérons que, si la retenue se vide, la centrale doit arrêter de turbiner (même si le turbinage était programmé) ; si elle se remplit (et le turbinage n'est pas programmé), la centrale doit quand même turbiner de l'eau, même à un prix inférieur à celui programmé, en essayant de le faire juste assez pour qu'il y ait le moins de débordement possible.

La simulation de la gestion avec les débits observés fournit les gains de gestion sur le jour simulé, ainsi que le stock initial pour le prochain jour de prévision, quand une nouvelle règle de gestion sera définie à partir des nouvelles prévisions des apports en eau. A la fin d'une période d'évaluation donnée (une saison ou une année), les gains de gestion, les volumes turbinés et les volumes déversés sont calculés. C'est l'*étape de l'évaluation* de la gestion. Ces variables seront utilisées pour comparer les différents types de prévision entre eux et pour déterminer celui qui présente une valeur économique plus élevée pour la gestion d'une centrale du type éclusée hebdomadaire. Puisque l'aménagement du type éclusée peut présenter des caractéristiques diverses et variées (Figure 2), il peut être intéressant d'examiner la sensibilité des résultats de l'évaluation aux principaux facteurs définissant les capacités de stockage de son réservoir et de turbinage de l'usine.

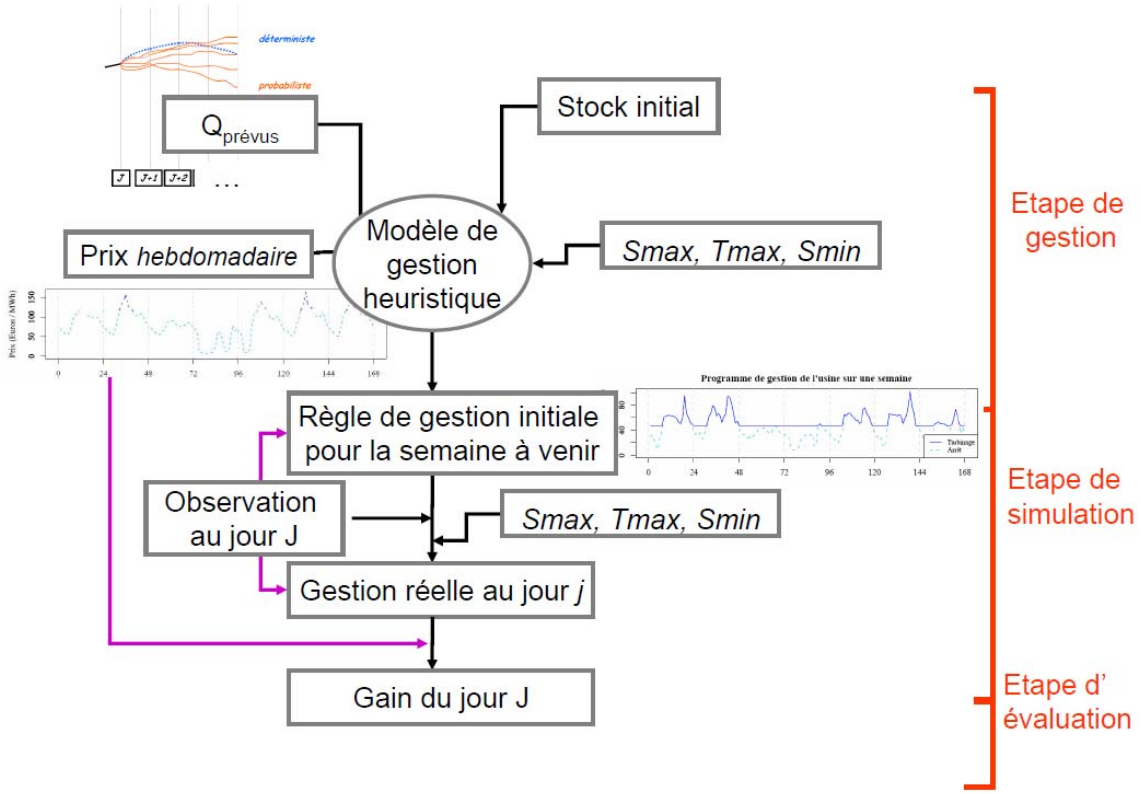


Figure 138: Schématique de la procédure développée pour l'optimisation heuristique d'une réserve hydroélectrique pour un jour J.

Afin de pouvoir optimiser la gestion d'une réserve hydroélectrique de façon heuristique et dans le contexte de la thèse, nous nous sommes basés sur certaines simplifications necessaires. Les principales hypothèses simplificatrices sont :

- les prix de l'énergie sont connus une semaine à l'avance,
- les débits entrants sont considérés constants toute au long de la journée,
- le débit réservé est considéré comme nul,
- le volume minimal du réservoir est considéré zéro,
- l'usine fonctionne à pleine puissance,
- il n'y a aucun autre usage de l'eau à valoriser,
- le coût marginal de production est considéré comme nul,
- et le rendement est considéré comme constant, indépendant du débit turbiné ou du niveau de la réserve.

Dans son ensemble, notre approche d'estimation de la valeur d'une prévision comprend trois étapes de calcul (Figure 138) et une étape d'analyse de sensibilité, résumées ci-après et détaillés par la suite :

- A chaque jour *J* de la période d'évaluation, le modèle de gestion est appliqué pour établir la règle de gestion à considérer : *étape de gestion.*
- Ensuite, les gains de gestion sont calculés en appliquant la règle de gestion aux débits observés au jour *J* : *étape de simulation.*
- A la fin de la période d'évaluation *T*, le gain total, sur l'ensemble des jours *J*, est évalué. Les gains obtenus pour les différents types de prévision étudiés sont alors comparés entre eux : *étape d'évaluation.*

Les étapes précédentes sont appliquées plusieurs fois, en faisant varier les caractéristiques de l'aménagement : *analyse de sensibilité.*

7.2.3 Etape de gestion

Considérons la gestion d'une éclusée hebdomadaire pendant une période d'évaluation *T*. L'éclusée et le bassin versant associé ont les caractéristiques suivantes :

S_{BV} : est la surface du bassin versant en amont de l'éclusée (en km²),

$Q_{o_H}(t)$: est la série de débits observés à l'exutoire du bassin versant, au pas de temps journalier, pendant une période historique *H* de plusieurs années (en m^3/s),

Q_{moy_H} : est le débit journalier moyen sur la période historique *H* (en m^3/s),

Q_{min_H} : est le débit journalier minimal sur la période historique *H* (en m^3/s),

V_{max} : est le volume maximal (volume utile) du réservoir (en Mm^3). Il est évalué en fonction de la capacité utile du réservoir, k_{max} (en jours), et du débit journalier moyen de la période historique Q_{moy_H} exprimé en Mm^3 de volume sur 1 jour :

$$V_{\max} = k_{\max} * (Q_{moy_H} * 0.0036 * 24) \quad \text{Eq. 52}$$

V_{min} : est le volume minimal qui doit être maintenu dans le réservoir (en Mm^3). Il est donné par le produit entre le coefficient de volume minimal, k_{min} (en jours), et le débit journalier minimal de la période historique Q_{min_H}, exprimé en Mm^3 de volume sur 1 jour :

$$V_{\min} = k_{\min} * (Q_{\min_H} * 0.0036 * 24) \quad \text{Eq. 53}$$

Ici, on fixera $k_{min} = 0$, c'est-à-dire, le réservoir peut être complètement vidé ($V_{min} = 0$). Avec, 0.0036 la constante de transformation des unités : [m^3/s] = 3600/1000000 [Mm^3/h].

QT_{max} : est le débit turbinable (maximum que peuvent turbiner les équipements de la centrale) (en m^3/s). Il sera donné par le coefficient de turbinage, C_T, multiplié par le débit journalier moyen de la période historique Q_{moy_H} :

$$QT_{\max} = C_T * Q_{moy_H} \quad \text{Eq. 54}$$

E_{usine} : est le rendement de l'usine. Ici, considéré égal à 1 MWh par m^3/s turbiné pendant 1 heure.

Au jour *J*, *l'hydrogramme des apports prévus en débit* (en m^3/s) pour les échéances journalières J+1 à J+7 est donné par la série chronologique :

$$Q_{7j} = \{Q_{j+1}, Q_{j+2}, Q_{j+3}, Q_{j+4}, Q_{j+5}, Q_{j+6}, Q_{j+7}\} \quad \text{Eq. 55}$$

Le débit moyen prévu sur 7 jours (en m^3/s) est donné par :

$$\overline{Q_{7j}} = \frac{1}{7}\sum_{i=1}^{7} Q_{j+i} \qquad \text{Eq. 56}$$

Nous considérons ici une *gestion hebdomadaire*, c'est-à-dire, dans un premier temps on vise à retrouver au bout d'une semaine le même stock (volume d'eau contenu dans le réservoir de l'éclusée) qu'au départ[2]. Cela veut dire que l'on cherche à turbiner tout le volume prévu arrivant dans la retenue. Dans cette phase de gestion, nous considérons également le *fonctionnement de l'usine à pleine puissance.*

La question qui se pose est alors un simple problème de proportionnalité : *si l'usine à pleine puissance peut turbiner QT_{max} m^3/s pendant 168 heures, combien d'heures est-il nécessaire pour turbiner à pleine puissance le débit moyen d'apport prévu ?*

La solution est dans une simple règle de trois : le *temps de turbinage*, *XT* (en heures), sera :

$$XT = \min(168 * \frac{\overline{Q_{7j}}}{QT_{\max}}, 168), \qquad XT \in [0,168h] \qquad \text{Eq. 57}$$

- si l'apport moyen prévu est nul, il n'y aura pas de turbinage : *XT* = 0 ;
- si l'apport moyen prévu est supérieur ou égal à la capacité maximale de turbinage, l'usine sera programmée à turbiner en continu pendant les 168 heures de la semaine : *XT* = 168 h. Le volume excédentaire sera déversé.
- si l'apport moyen prévu est non-nul et inférieur à la capacité maximale de turbinage, on réduira proportionnellement le temps total de turbinage possible (168 heures).

Connaissant le nombre total d'heures pendant lesquelles il faudra turbiner, il s'agit maintenant de déterminer à quels moments (heures) de la semaine le turbinage sera activé et à quels moments l'usine sera en arrêt. Pour cela, on fait appel à la courbe des *prix de l'énergie* (en €/MWh) de la semaine, donnée par la série chronologique :

$$PRIX(h) = \{p(1), p(2), ..., p(168)\} \qquad \text{Eq. 58}$$

La profitabilité de l'usine, qui ne va produire que pour un nombre d'heures *XT* dans la semaine, va être particulièrement sensible au niveau des prix durant les jours et heures de forte demande. Afin de maximiser les gains, il faudra trouver le *prix seuil* au dessus duquel l'usine sera mise à turbiner et en-dessous duquel, l'usine sera en arrêt.

[2] *Ceci n'est cependant pas une contrainte, et en réalité ne se réalisera pas. Il s'agit juste une considération de départ pour les calculs.*

Nous utilisons ici les données infra-journalières de la bourse des marchés spot de l'électricité EPEX (*European Power Exchange*)[3], qui gèrent les marchés français, allemand, autrichien et suisse, extraites et mises à disposition par la Direction Optimisation Amont & Aval d'EDF à Lyon. Les données s'étendent du 01/01/2003 à 00h au 22/06/2010 à 23h, au pas de temps horaire.

La Figure 139 montre un exemple des prix horaires de l'énergie pour les années 2003 et 2007. Les variations de prix dépendent principalement de la demande d'électricité. En général, les prix atteignent les valeurs les plus élevées dans les années très froides en hiver ou les années très chaudes en été. Lors de la canicule de 2003 et de la vague de froid de 2007, la Figure 139 montre que les prix ont atteint les 1000 €/MWh et 2500 €/MWh, respectivement.

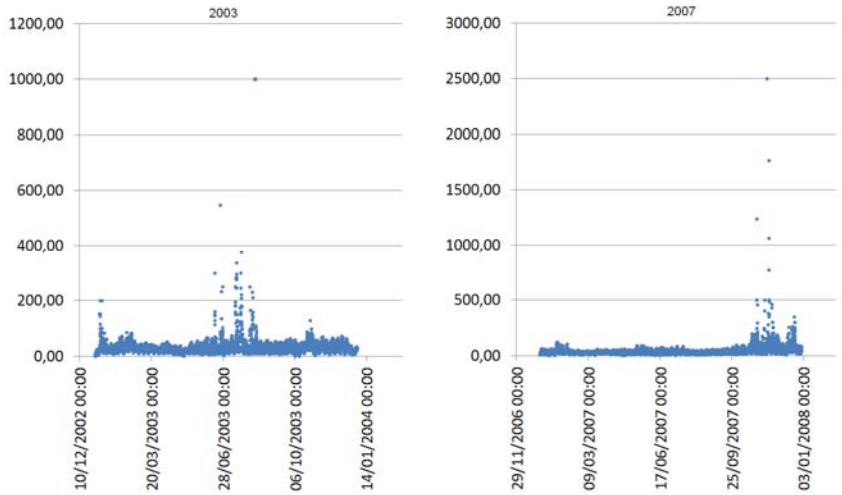


Figure 139 : Prix horaire de l'énergie EPEX-SPOT pour les années 2003 (à gauche) et 2007 (à droite). Prix en €/MWh.

En sus de la variabilité annuelle, le prix de l'énergie évolue également en fonction du jour de la semaine. On observe que la demande en énergie – et par conséquent les prix – sont plus élevés pendant les jours ouvrables lundi au vendredi) et plus faibles durant le week-end. La Figure 140 illustre cette variabilité hebdomadaire pour la semaine du 8 au 14 septembre 2008 (lundi au dimanche). De plus, les prix sont variables aussi sur une journée. La Figure 141 montre les prix observés au long de la journée du 8 septembre 2008, un lundi. Les prix sont plus bas durant la nuit et plus élevés vers midi, quand la demande est la plus haute, avec un pic observé également vers l'heure du dîner (19h-20h).

[3] *Le prix spot (« spot price ») est une tarification en temps réel : le prix au moment où l'on achète. Il est contrasté avec les prix fixés dans les marchés à terme (« forward price »). Le prix spot horaire est à la base des transactions des marchés de l'électricité et peut être défini comme le prix qui équilibre l'offre et la demande d'électricité pour un intervalle de temps donné. Les prix EPEX SPOT sont disponibles sur : http://www.epexspot.com/fr/*

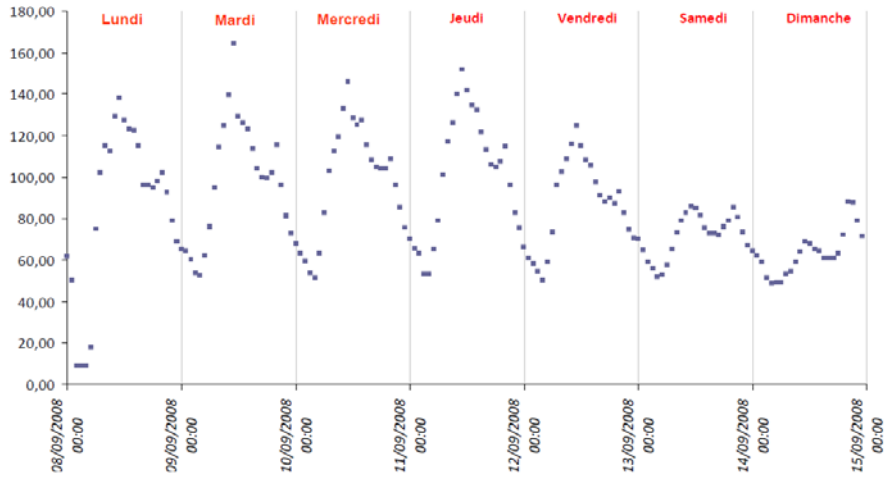


Figure 140 : Prix horaire de l'énergie EPEX-SPOT au long de la semaine du 8 au 14 septembre 2008. Prix en €/MWh.

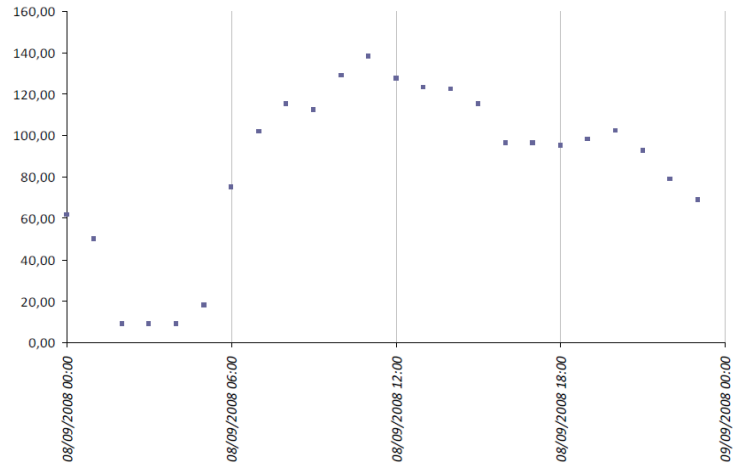


Figure 141 : Prix horaire de l'énergie EPEX-SPOT au long de la journée du 8 septembre 2008 (lundi). Prix en €/MWh.

Pour établir les meilleures heures de turbinage, nous construisons d'abord la fonction monotone des prix horaires en triant les 168 prix horaires de la semaine à venir de la plus haute à la plus basse valeur :

$$PRIX_{monotone} = \{p(1) > p(2) > p(i) \ldots > p(168)\} \qquad \text{Eq. 59}$$

L'ordonnée correspondant à l'heure *XT* fournit le prix-seuil (en €/MWh) à utiliser :

$$PRIX_{seuil} = \begin{cases} PRIXmonotone(truc(XT)) & ,trunc(XT)>0 \\ PRIXmonotone(1)+1 & ,trunc(XT)=0 \end{cases} \quad \text{Eq. 60}$$

Avec *trunc* pour la troncature à l'unité de la valeur de *XT*.

On note que le cas où la troncature à l'unité de *XT* est égale à 0 indique qu'il n'y aura pas de programme de turbinage prévu dans la semaine. Ceci équivaut à attribuer une valeur de prix seuil supérieure à la plus forte valeur de la courbe des prix de la semaine.

Le *programme de gestion de l'usine* sur la semaine est donné par la fonction de Heaviside, indicatrice des heures de turbinage (= 1) et des heures d'arrêt (= 0) :

$$H_{TURB}(h) = \begin{cases} 1 & ,PRIX(h) \geq PRIX_{seuil} \\ 0 & ,PRIX(h) < PRIX_{seuil} \end{cases} \quad \text{Eq. 61}$$

Dans notre cas, on considère que quand l'usine turbine elle le fait à pleine puissance. Dans ce cas, l'évolution temporelle des *volumes turbinés* (en Mm^3), au pas de temps horaire, sur la semaine, est donnée par :

$$V_{TURB}(h) = H_{TURB}(h) * QT_{\max} * 0.0036 \quad \text{Eq. 62}$$

L'évolution temporelle du stock d'eau du réservoir est calculée en fonction du stock initial, des volumes turbinés et des apports prévus. Au premier jour de la période d'évaluation, nous démarrons les calculs avec un choix arbitraire pour le stock initial équivalent à la moitié du volume maximal du réservoir. Pour les jours suivants, le stock initial à *J+1* sera donné par le stock final obtenu en fin de l'étape de simulation de la gestion sur 24 heures avec débits observés au jour *J*.

L'évolution du *stock d'eau horaire* (en Mm^3) sur la semaine est donnée par :

$$S(h) = S(h-1) + Q_p(h) * 0.0036 + V_{TURB}(h-1) \quad \text{Eq. 63}$$

Les *débits prévus horaires* $Q_p(h)$ (en m^3/s) sont évalués à partir des apports prévus en débit journalier, sous l'hypothèse d'une répartition constante dans le temps :

$$Q_p(h) = Q_{j+k}, \text{ avec } k = 24 - [trunc(24 - \frac{h}{24})] \quad \text{Eq. 64}$$

Il s'agit ensuite de vérifier les *contraintes de gestion*. Dans notre cas, cela se traduit en vérifiant si le stock (à toutes les heures de la semaine) est compris entre le volume minimal du réservoir et son volume maximal :

$$V_{\min} \leq S(h) \leq V_{\max} \quad \text{Eq. 65}$$

Si la condition ci-dessous est respectée, c'est-à-dire, le stock est compris entre les limites du réservoir, aucune *rectification* n'est à faire et la règle de gestion est déterminée. C'est la fin de l'étape de gestion pour le jour *J*.

Si la condition n'est pas respectée, deux cas de figures peuvent se présenter (Fig. 2) :

1. $S(h) > V_{max}$
    - Il faut alors *turbiner plus*, avant que le stock devienne supérieur à la capacité maximale du réservoir, pour éviter que la retenue déborde.
    - Pour pouvoir faire plus d'heures de turbinage, il faut accepter de *baisser le prix-seuil* (turbiner à des prix moins élevés).
    - Notamment, il faut turbiner en plus le *volume excédentaire*, celui qui dépasse le seuil de capacité maximale du réservoir.
2. $S(h) < V_{min}$
    - Il faut alors *turbiner moins*, avant que le stock devienne inférieur au seuil de volume minimum du réservoir.
    - Pour pouvoir faire moins d'heures de turbinage de façon optimale, il faut *augmenter le prix-seuil* (turbiner moins d'heures, mais tout en cherchant à turbiner aux heures de prix élevés).
    - Notamment, il ne faut pas turbiner le *volume déficitaire*, celui qui est en-dessous du seuil de volume minimum du réservoir.

Quel que soit le cas, le volume à turbiner doit être recalculé en faisant le bilan d'eau à l'intérieur du réservoir à l'instant où l'on démarre la rectification ($h_{ini} = 0$ pour la première itération) et l'heure (h_{fin}) où le stock atteint sa plus forte (ou basse) valeur. Cet intervalle de temps correspond à la *fenêtre de rectification*, ici indiquée par l'intervalle (h_{ini} : h_{fin}).

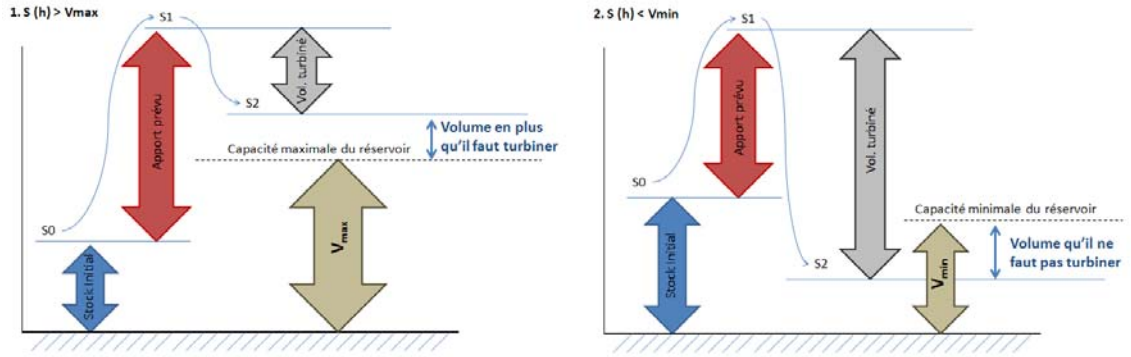


Figure 142 : Illustration des cas où les contraintes de gestion ne sont pas respectées et il faut faire une rectification de la règle de gestion : quand le stock dépasse la capacité maximale du réservoir (à gauche) et quand le stock devient inférieur au volume minimum du réservoir (à droite).

Ce volume représente un *débit de rectification*, Q_{rectif} (en m^3/s) *sur les heures de la fenêtre de rectification* de :

$$Q_{rectif} = \frac{Vol_{rectif}(h_{ini} : h_{fin})}{0.0036} \qquad \text{Eq. 66}$$

Le *nouveau temps de turbinage* nécessaire pour rectifier la gestion est obtenu, comme précédemment, par une règle de proportionnalité. Il vaut *XT'* (en heures) *sur les heures de la fenêtre de rectification* :

$$XT' = trunc[\min(\frac{Q_{rectif}}{QT_{\max}}; (h_{fin} - h_{ini} + 1))] + \Delta \qquad \text{Eq. 67}$$

$$0 \le XT' \le (h_{fin} - h_{ini} + 1) \qquad \text{Eq. 68}$$

, ou *trunc* pour la troncature à l'unité de la valeur obtenue

Δ = 0 heure, pour une rectification quand $S(h) < V_{min}$

Δ = 1 heure, pour une rectification quand $S(h) > V_{max}$

Nous notons que cette constante Δ ajoutée dans le programme de calcul de la gestion rectifiée correspond à une heure ajoutée au temps de turbinage lorsque l'on dépasse la capacité maximale du réservoir. Son rôle n'est autre que celui de garantir plus rapidement le respect de cette contrainte face à la troncature à l'unité de la valeur calculée.

Une fois le nouveau temps de turbinage calculé, il faut ensuite obtenir le *nouveau prix seuil*. Pour cela, une nouvelle courbe monotone de prix est construite, cette fois-ci en triant seulement les prix horaires compris dans la fenêtre temporelle de rectification. Le nouveau prix seuil est obtenu come décrit précédemment et un *nouveau programme de gestion* est défini. Le respect des contraintes est à nouveau vérifié. La procédure de modélisation se répètera jusqu'à ce que les contraintes soient respectées et il n'y ait plus de rectification à faire.

Enfin, dans le modèle de gestion construit, quelques conditions additionnelles ont été ajoutées afin d'assurer que le processus itératif aboutisse à une règle de gestion à chaque jour *J*. Ceci garantit le passage au jour suivant et permet à l'algorithme de parcourir l'ensemble de la période d'évaluation sans interruption. Une première condition concerne la possibilité de définir plusieurs prix-seuils sur l'ensemble de la gestion hebdomadaire. Ainsi, si une rectification qui a été faite pour le respect des contraintes à l'heure *h* entraîne le non-respect des contraintes à une heure *h'* > *h*, l'algorithme fixera d'abord le prix seuil obtenu jusqu'à l'heure *h* et ensuite reprendra les itérations pour faire la rectification ciblant l'heure *h'*, mais cette fois-ci sur une fenêtre de rectification qui débutera à l'heure *h* + 1. L'autre condition ajoutée limite le nombre d'itérations possibles : si l'algorithme n'arrive pas à trouver une règle de gestion au bout de 20 itérations, le processus de rectification s'arrête et on garde la gestion définie (même si les contraintes sont enfreintes) pour la suite des calculs.

Une fois la règle de gestion définie pour le jour *J* sur la base des apports en eau prévus, l'algorithme passe à l'étape de simulation.

7.2.4 Etape de simulation

Dans cette étape, on simule les premières 24 heures de la gestion avec les débits observés sur la journée. On cherche ici à évaluer ce qui se passe en réalité quand on applique la règle de gestion que le modèle a jugée optimale aux quantités d'eau qui ont effectivement été apportées au réservoir de l'éclusée.

Il s'agit d'évaluer l'évolution du stock d'eau horaire (en Mm^3) sur les premières 24 heures de la semaine à partir des apports observés et des volumes turbinés tels que programmés par la règle de gestion :

$$S_{obs}(h) = S_{obs}(h-1) + Q_o(h) * 0.0036 - V_{TURB}(h-1)\ ,\ h \in [1,24] \qquad \text{Eq. 69}$$

Les débits observés horaires $Q_o\ (h)$ (en m^3/s) sont calculés à partir des apports en débit observé journalier, sous l'hypothèse d'une répartition constante dans le temps.

Les contraintes imposées par la capacité maximale et le volume minimum du réservoir sont vérifiées heure par heure (on simule ici une gestion en temps réel, où le gestionnaire adaptera la règle de gestion qui lui a été fournie en fonction des quantités d'eau qui arrivent au réservoir).

Nous considérons que :

- si à l'heure *h*, la retenue se vide ($S_{obs}(h+1) < V_{min}$), la centrale doit turbiner moins que prévu par la gestion :

$$V'_{TURB}(h) = \max([V_{TURB}(h) - V_{\min} + S_{obs}(h+1)],0) \quad \text{Eq. 70}$$

Le stock à l'heure *h* est recalculé à partir de ce nouveau volume turbiné.

- si à l'heure *h*, la retenue se remplit trop ($S_{obs}(h+1) > V_{max}$), la centrale doit turbiner le nécessaire pour ne pas déverser ou pour déverser le moins possible, dans la limite de al capacité maximale de turbinage de l'usine :

$$V'_{TURB}(h) = \min([S_{obs}(h) - V_{\min}] + QT_{\max} * 0.0036], V_{TURB}(h-1)) \quad \text{Eq. 71}$$

Le stock à l'heure *h* est recalculé à partir de ce nouveau volume turbiné, ainsi que, si c'est le cas, le volume déversé :

$$V'_{DEVERSE}(h) = \max([S_{obs}(h+1) - V_{\max}],0) \quad \text{Eq. 72}$$

A la fin de l'étape de simulation au jour *J*, le stock $S_{obs}(h=24)$ au jour J est transféré comme stock initial pour la gestion du jour suivant, *J*+1 .

7.2.5 Etape d'évaluation

Nous considérons ici trois critères pour évaluer la valeur d'une prévision sur l'ensemble des jours de la période d'évaluation *T* :

– *le gain économique de la gestion* (en €) :

Ce gain est évalué en deux temps :

– Si nous considérons le rendement de l'usine E_{us}, les volumes effectivement turbinés et les prix de l'énergie à l'heure où ces volumes ont été turbinés, le gain en énergie produite, GainENERGIE s'écrit :

$$Gain_{ENERGIE} = \sum_{T}\sum_{h=1}^{24} PRIX_{seuil}(h) * \frac{V'_{TURB}(h)}{0.0036} * E_{us} \quad \text{Eq. 73}$$

, où $PRIX_{seuil}(h)$: le prix seuil à utiliser (en €/MWh) ;

$V'_{TURB}(h)$: le volume turbiné au pas de temps horaire (en Mm^3) ;

E_{us} : le rendement de l'usine égal à 1 MWh par m^3/s turbiné pendant 1 heure ;

- Dans le but d'illustrer plus clairement la performance de chaque prévision (pré/post-traitée), l'introduction d'un *score de gain économique de gestion,* S_{GAIN} s'impose. Le principe sur lequel ce score se bâtit est le simple ratio entre prévisions traitées et prévisions brutes. Pour cette raison le $Gain_{GESTION}$ est utilisé. Si les prévisions traitées apportent un gain supérieur à celui des prévisions brutes *PB* le score prend une valeur supérieure à 1. Si les prévisions corrigées *PT* sont moins performantes que les prévisions brutes le score devient inférieur à 1.

$$S_{Gain} = \frac{Gain_{PT}}{Gain_{PB}} \quad \text{Eq. 74}$$

- *les totaux de volumes turbinés* VT_t *et déversés* VT_d (en Mm^3) :

$$VT_t = \sum_{T}\sum_{h=1}^{24} V'_{TURB}(h) \quad \text{Eq. 75}$$

$$VT_d = \sum_{T}\sum_{h=1}^{24} V'_{DEVERSE}(h) \quad \text{Eq. 76}$$

, où $V'_{TURB}(h)$: le volume turbiné au pas de temps horaire (en Mm^3) ;

$V'_{DEVERSE}(h)$: le volume déversé au pas de temps horaire (en Mm^3) ;

- *l'écart* ΔS_T *entre stock de fin et de début de la période d'évaluation T* (en Mm^3) :

$$\Delta S_T = S_{obs}(h=24)_{j=T} - S_{obs}(h=24)_{j=1} \quad \text{Eq. 77}$$

, où $S_{obs}(h=24)_{j=T}$: le stock d'eau du réservoir (en Mm^3) à la dernière heure de la période d'évaluation T ;

$S_{obs}(h=24)_{j=0}$: le stock d'eau du réservoir (en Mm^3) à la première heure de la période d'évaluation T ;

- Le *gain en gestion*, $Gain_{GESTION}$, qui prend en compte non seulement l'énergie produite, mais aussi le gain potentiel correspondant au turbinage de l'écart ΔS_T entre stock de fin et de début de la période d'évaluation T. Etant donné que l'objectif de la gestion est d'avoir le même niveau du stock au début et à la fin de la période d'évaluation, l'écart entre le stock de fin et de début correspond à un gain ou une perte potentielle à prendre en compte :

$$Gain_{GESTION} = \sum_{T} \sum_{h=1}^{24} PRIX_{seuil}(h) * \frac{V'_{TURB}(h)}{0{,}0036} * E_{us} + \overline{PRIX}_{seuil} * \frac{\Delta S_T}{0{,}0036} * E_{us} \quad \text{Eq. 78}$$

, où $PRIX_{seuil}(h)$: le prix seuil à utiliser (en €/MWh) ;

$V'_{TURB}(h)$: le volume turbiné au pas de temps horaire (en Mm^3) ;

E_{us} : le rendement de l'usine égal à 1 MWh par m^3/s turbiné pendant 1 heure ;

$\overline{PRIX}_{seuil}$: la moyenne de tous les prix seuil horaires utilisés sur la dernière semaine d'évaluation (en €/MWh) ;

ΔS_T : l'écart ΔS_T entre stock de fin et de début de la période d'évaluation T.

7.2.6 Prise en compte des prévisions

Nous cherchons à évaluer les gains de gestion lorsque nous considérons les prévisions suivantes :

- Prévision déterministe de la climatologie : le débit prévu est égal au débit journalier moyen du jour en cours, estimé sur l'ensemble de la période historique d'observation disponible pour le basin versant,
- Prévision déterministe " parfaite " : les débits prévus sont identiques aux débits observés,
- Prévision d'ensemble brute : les débits prévus sont les membres de la prévision hydrologique d'ensemble brute (sans application d'aucun traitement statistique de correction de biais),
- Moyenne de la prévision d'ensemble brute : le débit prévu est la moyenne des débits prévus par la prévision hydrologique d'ensemble brute,
- Prévision déterministe : les débits prévus sont fournis par le modèle hydrologique avec pour entrées les prévisions déterministes de pluie et température,

- Prévision d'ensemble corrigée : les débits prévus sont les membres de la prévision hydrologique d'ensemble corrigée (après application des traitements statistiques de correction de biais présentés dans le Chapitre 6).

Les étapes du modèle de gestion décrites précédemment s'appliquent à un débit d'apport moyen sur les sept jours à venir. Dans l'intention d'appliquer ces équations dans le cadre des prévisions probabilistes nous avons examiné deux options possibles :

- dans le cas des prévisions à scénario unique, l'application est immédiate.
- dans le cas des prévisions d'ensemble, à *n* scénarios, la procédure adoptée pour chaque jour de prévision a été de faire la gestion des apports pour chacune des prévisions de l'ensemble et ensuite prendre la moyenne des *n* vecteurs de prix-seuil obtenus. C'est sur ces prix-seuil moyens que l'étape de simulation est ensuite appliquée et le gain du jour J calculé. Afin d'avoir une idée de la variabilité du gain basé sur les prix-seuil moyens, nous avons également considéré les vecteurs des prix-seuil maximum et minimum donnés par l'ensemble.

7.3 Résultats

Les résultats des différentes applications du modèle de gestion développé sont présentés selon quatre axes principaux :

i. Illustration d'un cas de gestion durant un événement de crue : nous illustrons la gestion d'un aménagement fictif, à capacité utile et coefficient de turbinage fixés, face aux apports en eau de la prévision d'ensemble brute du bassin versant de la Durance à Serre Ponçon pendant l'événement de crue qui a eu lieu du 1 au 9 septembre 2008.
ii. Analyse de sensibilité aux caractéristiques de l'aménagement : nous évaluons les gains apportés par la gestion de différents aménagements fictifs, à capacité utile et coefficient de turbinage variés, face aux apports en eau de la prévision d'ensemble brute du bassin versant de la Durance à Serre Ponçon pendant quatre années de prévision (2005 à 2008).
iii. Intérêt des prévisions d'ensemble :
    - nous analysons l'intérêt des prévisions d'ensemble en comparant, leurs gains de gestion, aux gains apportés par les prévisions données par la moyenne de la prévision d'ensemble, la prévision climatologique et la prévision parfaite. Dans un premier temps, le bassin versant de la Durance à Serre Ponçon et les quatre années de période d'évaluation (2005 à 2008) sont considérés ayant une capacité utile et un coefficient de turbinage constants.

iv. Impact de la qualité des prévisions sur les gains de gestion : nous évaluons l'impact de la correction de biais apportée aux prévisions d'ensemble brutes et corrigées.
   - Dans un premier temps, les données du bassin versant de la Durance à Serre Ponçon sur les quatre années d'évaluation (2005 à 2008) sont considérées ayant une capacité utile variable et un coefficient de turbinage constant.
   - Ensuite, nous évaluons l'impact de la qualité de prévision sur le gain de gestion face aux apports en eau de la prévision d'ensemble brute de l'ensemble de 11 bassins versants. Dans ce cas les caractéristiques d'aménagement sont considérées constantes.

7.3.1 Illustration d'un cas de gestion durant un événement de crue

Nous illustrons les résultats du modèle de gestion développé face aux apports en eau prévus et observés lors de l'événement du 1er au 9 septembre 2008 dans le bassin versant de la Durance à Serre Ponçon (3580 km²). Cet événement de crue a été caractérisé par des fortes précipitations qui sont tombées dans le Sud-Est de la France. Ces précipitations étaient bien captées par les prévisions d'ensemble du centre européen (Figure 143).

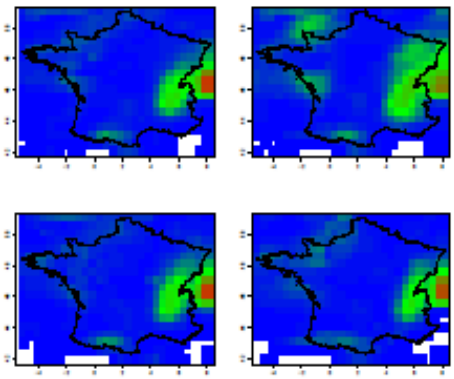


Figure 143 : Carte de prévisions sur la France pour la période de 1 à 9 septembre 2008. Ici les prévisions affichées sont extraites de 4 membres de l'ensemble à un horizon de 2 jours.

La Figure 144 (a) présente la pluie observée sur le bassin versant de la Durance à Serre Ponçon, ainsi que les boîtes à moustaches représentant les prévisions d'ensemble de pluie. La Figure 144 (b) présente l'hydrogramme de l'événement étudié, avec un débit journalier observé maximal de 131m^3/s enregistré le 6 septembre.

Pour la même période, nous présentons l'évolution des prix horaires extraits du marché européen d'énergie EPEX SPOT (Figure 145). Le prix journalier moyen sur la période est égal à 83,83 €.

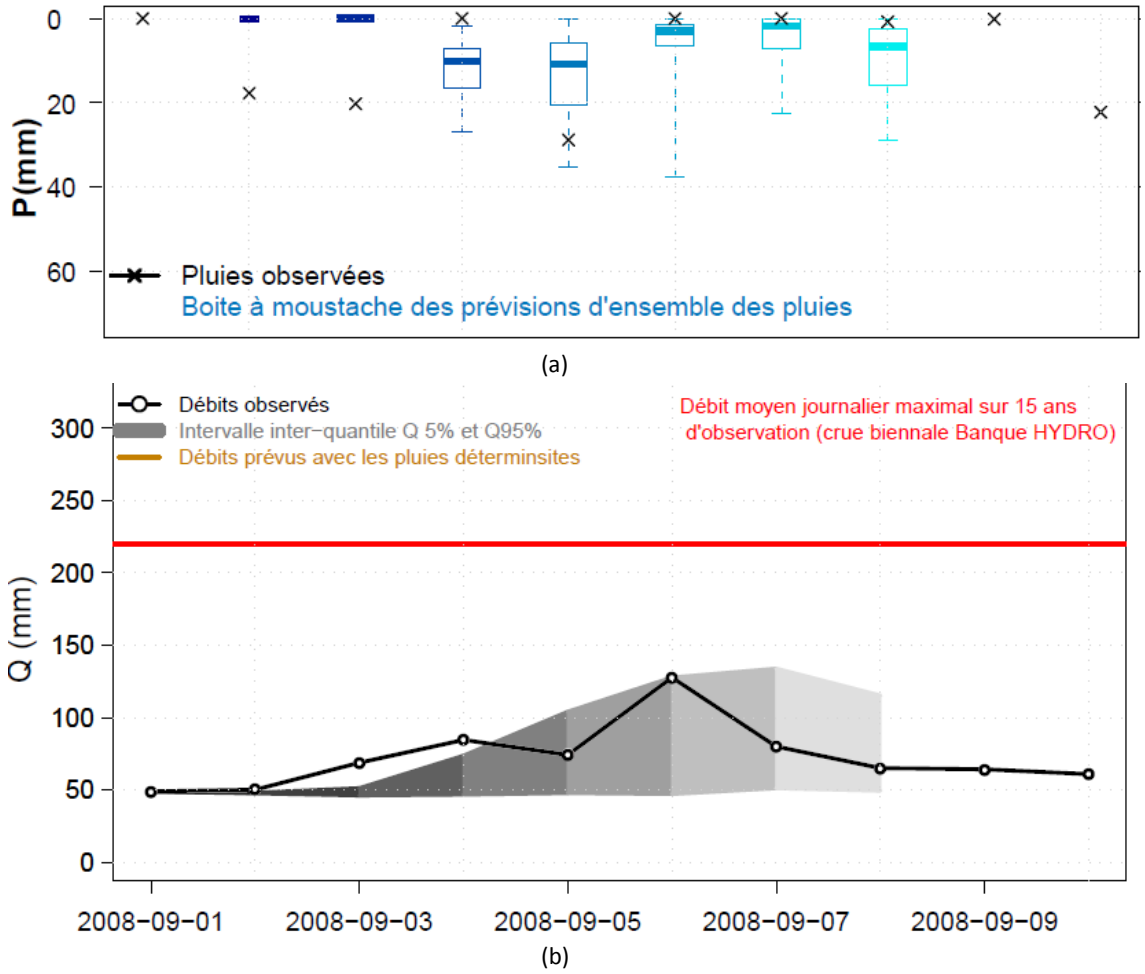


Figure 144 : (a) Boîtes à moustache pour les prévisions d'ensemble de précipitation pour la période de 1 à 9 septembre 2008, (b) l'hydrogramme de débits modélisés par le modèle MORDOR pour la période de 1 à 9 septembre 2008.

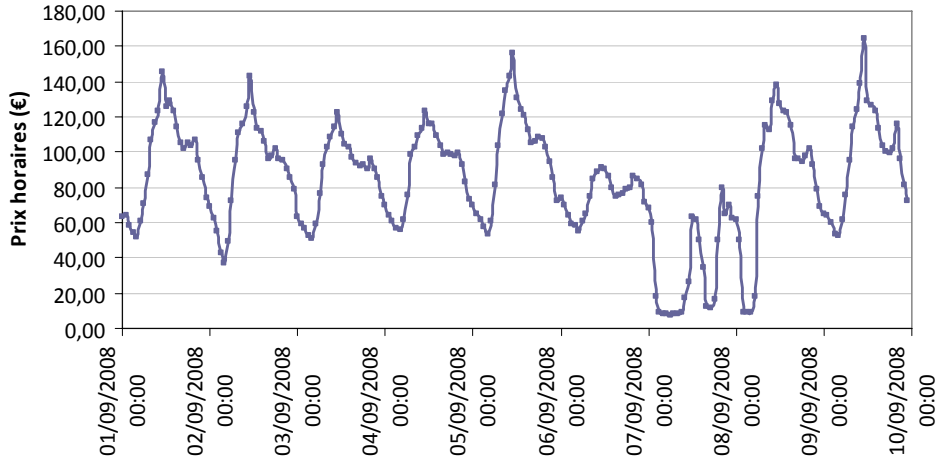


Figure 145 : Prix horaires extraits du marché européen d'énergie EPEX SPOT pour la même période.

Rappelons que le code heuristique vise à optimiser la gestion dans un horizon d'une semaine en respectant les contraintes de stock. Le but est de turbiner à pleine puissance le débit moyen prévu entrant dans le réservoir sur 7 jours. Le principe général se résume à programmer le turbinage des apports aux heures de prix plus élevés (afin de maximiser le gain de la gestion il est essentiel de distribuer les heures de turbinage sur les heures où les prix sont les plus élevés), tout en évitant les débordements et le non-respect des contraintes du réservoir.

La Figure 146 présente un exemple tiré de la gestion hebdomadaire obtenue quand nous considérons un aménagement hydroélectrique de coefficient de turbinage égal à 3 fois le débit moyen de la Durance à Serre Ponçon et une retenue de coefficient de capacité égal à 4 jours de stockage de ce débit moyen. La prévision parfaite (débits prévus identiques aux débits observés) a été considérée et le modèle hydrologique utilisé est *MORDOR*. L'algorithme de gestion a été executé en continu sur la période du 1er au 10 septembre 2008, avec stock initial au 1er septembre fixé à la moitié du stock maximal de la retenue.

Le résultat obtenu pour le 4 septembre 2008 est montré dans la Figure 146. La gestion a indiqué un prix seuil au-dessus duquel la centrale doit turbiner l'eau de 103,47 €/ MWh. Dans le graphique indiquant le programme de gestion de l'usine (Figure 146 en haut), la ligne bleue ciel pointillée représente la variation hebdomadaire des prix horaires de l'énergie (en €/ MWh, axe des ordonnées à gauche). La ligne bleue continue indique les instants où l'eau est turbinée : les heures où la ligne bleu continue suit la courbe des prix horaires sont celles où la centrale doit turbiner l'eau, tandis que la ligne bleue continue à l'horizontale sur le prix-seuil indique les instants où la turbine est en arrêt. Le niveau du stock de la retenue (en Mm^3) est aussi représenté (Figure 146 en bas).

Nous observons que pour cette gestion, les contraintes sont respectées sur l'ensemble de la période et aucune rectification n'a dû être faite. En effet, lorsque le niveau du stock approche le minimum accepté (contrainte Smin = 0), peu avant l'heure 48, l'usine arrête de turbiner et la retenue se remplit. Ceci correspond aux instants de niveau bas de prix, quand pour la gestion optimale il est préférable de réserver l'eau et la turbiner plus tard aux instants de prix plus élevés.

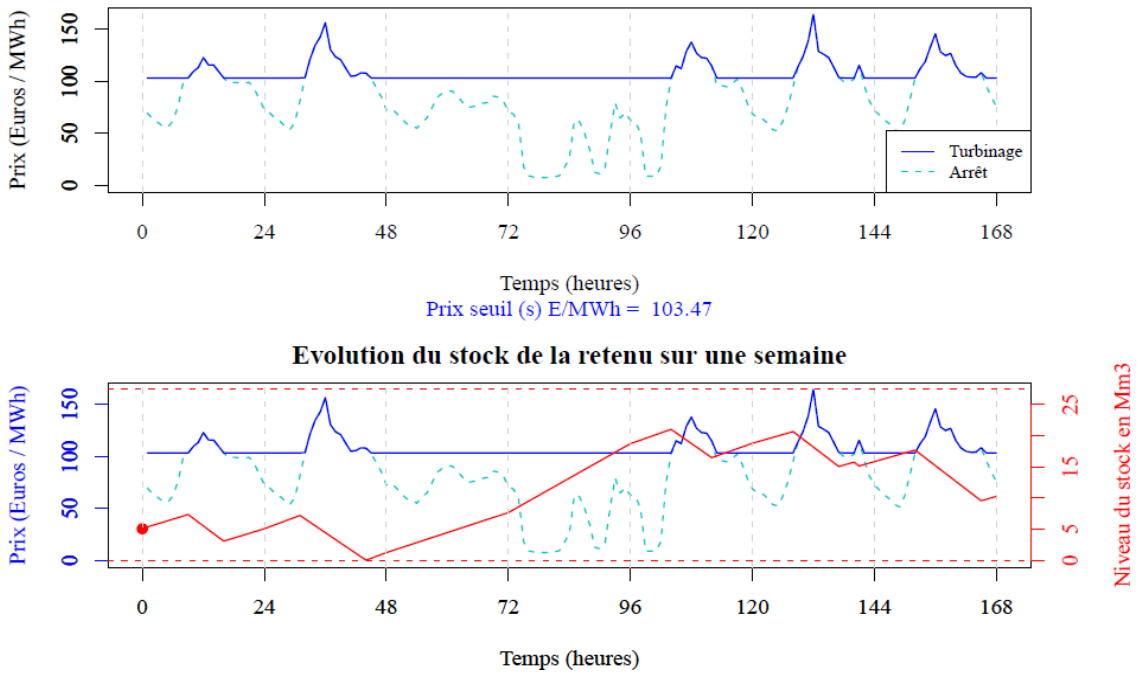


Figure 146: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 103,47 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. Aménagement hydroélectrique de coefficient de turbinage égal à 3 fois le débit moyen de la Durance à Serre Ponçon et coefficient de capacité égal à 4 jours de stockage du débit moyen.

L'équilibre entre les heures de turbinage, les heures où les prix sont les plus élevés et les contraintes du réservoir n'est cependant pas toujours facile à obtenir. En modifiant les coefficients de turbinage et de capacité, le code heuristique a besoin de plusieurs itérations (rectifications) avant de trouver la solution optimale de gestion. Les Figure 147 à Figure 149 présentent le même cas de gestion hebdomadaire pour le 4 septembre 2008, obtenu cette fois-ci à partir d'un aménagement à capacités de turbinage et de stockage réduites : coefficients de turbinage égal à 2 et capacité utile égale à 3 jours de débit moyen. Dans cette configuration, trois rectifications ont été nécessaires pour obtenir une gestion de l'usine avec respect des contraintes du réservoir.

Lors de la première itération (Figure 147 (en bas)), le prix-seuil obtenu est de 83,09 €/MWh (pour une capacité de turbinage moins élevé, le prix-seuil est aussi moins élevé). Nous observons que, avec ce prix-seuil, le stock est trop turbiné pendant les premières 48 heures, ce qui occasionne le non-respect de la contrainte de stock minimal. Une première rectification est alors nécessaire (Figure 148 (en bas)), où le prix-seuil fixé est plus élevé (94,19 €/MWh), ce qui aboutit à une diminution d'heures de turbinage. Néanmoins, même avec ce prix-seuil plus élevé, c'est à nouveau la contrainte de stock minimal qui est maintenant non-respectée : le stock obtient la valeur minimale admissible peu avant la 48ème heure. Une deuxième rectification est nécessaire (Figure 149 (en bas)). Cette fois-ci, deux prix-seuil sont définis pour la gestion de l'usine : 98,20 €/MWh jusque peu avant la 48ème heure et 96,02 €/MWh au-delà.

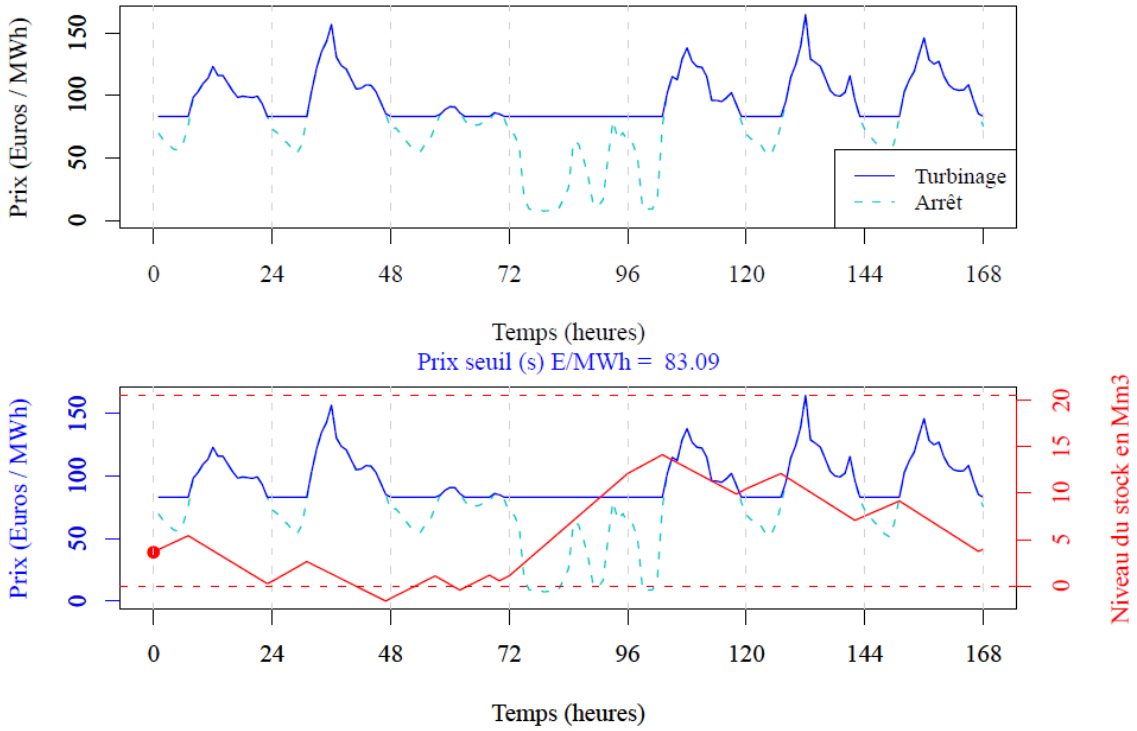


Figure 147: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 83,09 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. Aménagement hydroélectrique de coefficient de turbinage égal à 2 fois le débit moyen de la Durance à Serre Ponçon et coefficient de capacité égal à 3 jours de stockage du débit moyen.

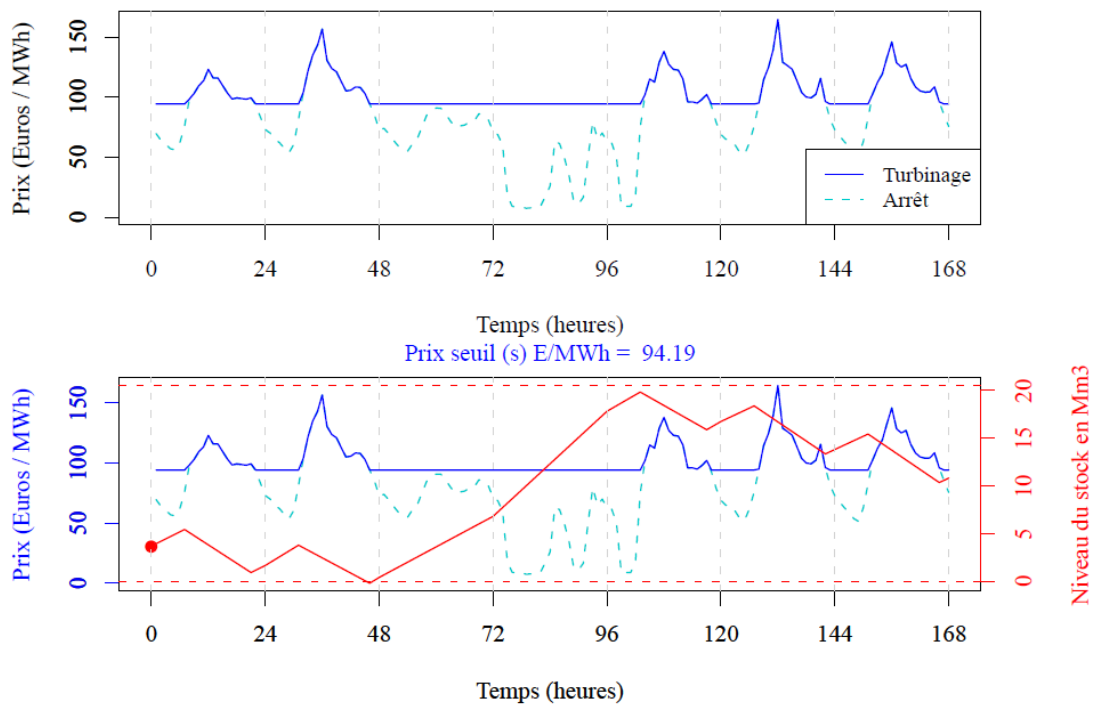


Figure 148: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 94,19 €/MWh). En bas : Niveau du stock de la retenue sur la semaine.

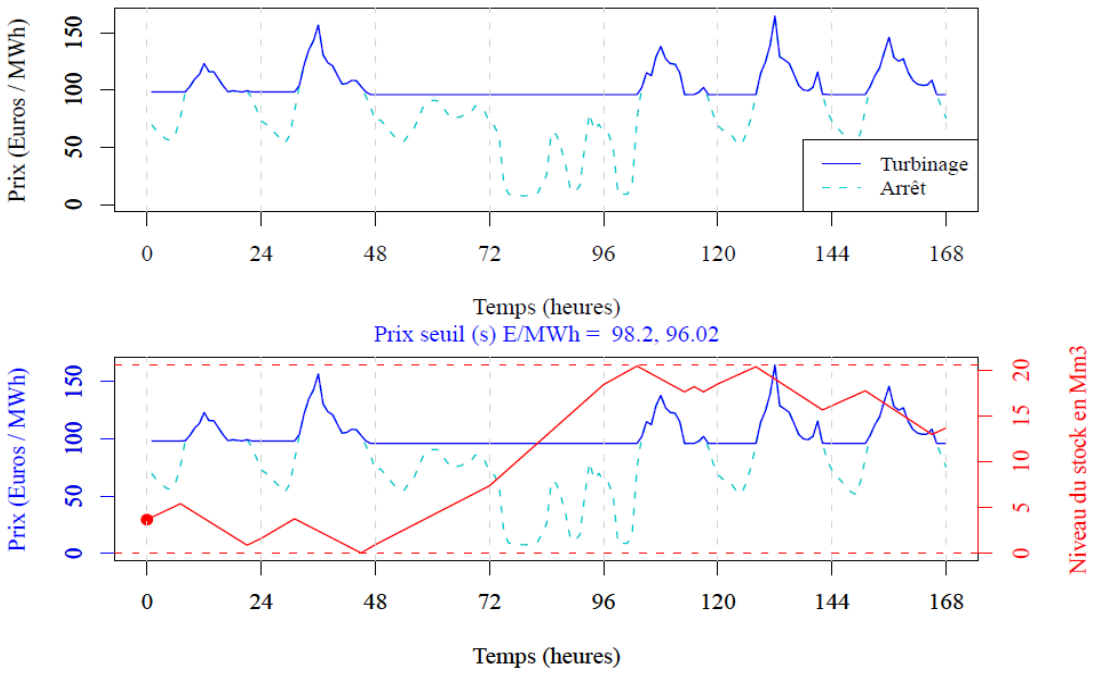


Figure 149: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 98,2 et 96,02 €/MWh). En bas : Niveau du stock de la retenue sur la semaine.

L'exemple montré a permis d'illustrer comment l'algorithme développé effectue l'étape de gestion et d'avoir un premier aperçu de l'impact des caractéristiques de l'aménagement sur la définition de la règle optimale de gestion de l'usine. Pour aller plus loin, nous présentons par la suite l'application de l'algorithme sur une période de quatre années (2005-2008) et nous étudions l'impact des caractéristiques de l'aménagement sur le gain économique évalué sur cette longue période.

7.3.2 Analyse de sensibilité aux caractéristiques de l'aménagement

L'illustration de l'événement ci-dessus est intéressante afin de démontrer la fonctionnalité du modèle de gestion, pourtant cela reste un exemple particulier qui manque de généralité. Dans ce paragraphe, nous présentons les résultats de la gestion sur l'ensemble de la période d'évaluation (1er janvier 2005 au 23 décembre 2008, 1453 jours). Les données sont toujours celles du bassin versant de la Durance à Serre Ponçon. L'hydrogramme de débits de cette période est présenté dans la Figure 150.

Le bassin versant de la Durance à Serre Ponçon est un bassin versant montagneux avec des fortes influences neigeuses. Par conséquent c'est souvent pendant la période de fonte que des événements importants de crue apparaissent. En mai 2008, par exemple, un événement de crue très important, avec un débit maximal de l'ordre de 800 m^3/s (Figure 150), a été observé le 30 mai sur le bassin versant. Lors de cet événement, la retenue de Serre Ponçon a déversé des volumes significatifs pour la deuxième fois en 50 ans d'existence.

Le modèle de gestion a été appliqué sur l'ensemble de la période d'évaluation. A chaque jour, les étapes de gestion et de simulation ont été exécutées. A la fin de la période, l'étape d'évaluation a permis d'estimer les gains de gestion. Pour étudier la sensibilité des résultats aux caractéristiques de l'aménagement fictif introduit à l'exutoire du bassin versant, nous avons fait varier les coefficients de turbinage et la capacité utile du réservoir.

Une première évaluation qualitative des résultats du modèle de gestion est présentée dans la Figure 151. Le modèle a été appliqué pour le cas des prévisions parfaites et pour un aménagement de coefficients de turbinage et de capacité égaux respectivement à 3 et 4. Les volumes déversés sur l'ensemble des jours de la période sont indiqués. Nous observons l'important déversement lors de l'événement de mai 2008. En effet, en dehors de cet événement le modèle a réussi à gérer les stocks dans les contraintes de gestion.

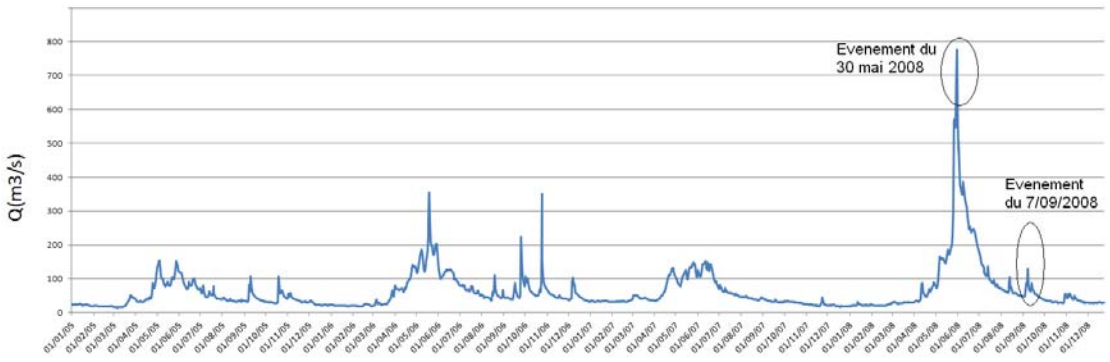


Figure 150: Hydrogramme des débits observés à la station de la Durance à Serre Ponçon sur la période d'évaluation (01 /01/2005 au 23/12/2008).

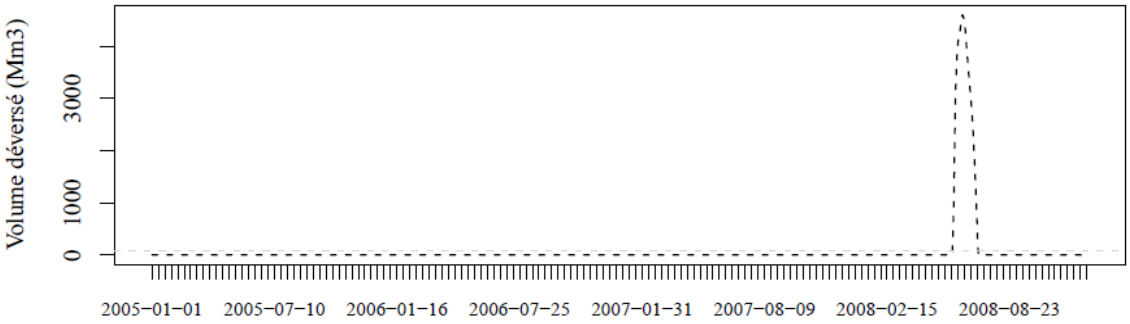


Figure 151: Volumes déversés d'après le modèle de gestion pour un aménagement fictif de coefficients de turbinage et de capacité égaux respectivement à 3 et 4, dans la Durance à Serre Ponçon sur la période d'évaluation (01 /01/2005 au 23/12/2008).

Dans l'exemple de l'événement du 1er au 10 septembre 2008, nous avons observé que les coefficients de capacité et de turbinage influencent la gestion de la retenue (Figure 146 à Figure 149). L'analyse de sensibilité sur l'ensemble de la période d'évaluation (01/01/2005 au 23/12/2008) est ici présentée en considérant les variations suivantes :

- capacité utile k_{max} pris entre 3 et 10 jours,
- coefficient de turbinage C_T pris entre 3 et 5.

Rappelons que ces coefficients (présentés dans la Figure 137) varient selon les caractéristiques de l'aménagement. Par conséquent pour les aménagements type éclusées examinés ici, les ordres de grandeur sont ceux décrits ci-dessus. La Figure 152 présente l'évolution du gain de la gestion (Eq. 78) en fonction des différentes valeurs du coefficient de capacité utile du réservoir (axe des abscisses) et du coefficient de turbinage (courbes représentées). Le cas étudié est celui des prévisions parfaites, afin de démontrer la sensibilité du modèle de gestion aux caractéristiques de l'aménagement.

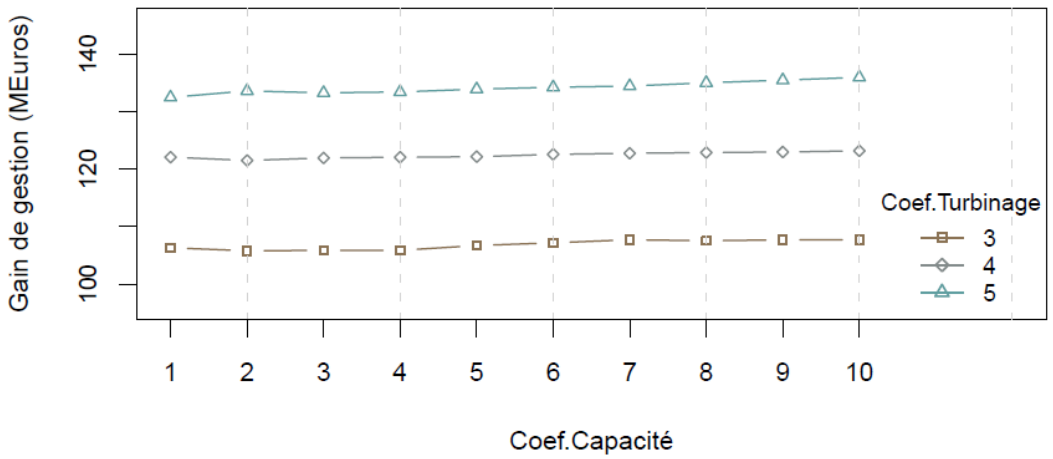


Figure 152: Gain de gestion en MEuros pour les prévisions parfaites : chaque ligne représente un différent coefficient de turbinage.

Selon ce graphique, plus le coefficient de turbinage est élevé, plus nous obtenons de gain avec la gestion des apports, ce qui est normal puisqu'une puissance plus importante permet de concentrer la production sur les heures les plus " chères ". En effet, les courbes correspondant aux trois coefficients de turbinage à valeurs croissantes sont pratiquement parallèles, car le gain de gestion est proportionnel au volume turbiné. De plus, nous observons que le gain de gestion croît légèrement en fonction de la capacité de la retenue, et ce jusqu'à une capacité donnée du débit moyen. A partir de cette valeur, le gain se stabilise. Cependant, le point de stabilisation varie en fonction du coefficient de turbinage, par exemple il est égal à 8 pour le coefficient de turbinage 3 et 4. Par contre pour le coefficient de turbinage 5, le gain accroît même au delà capacité 10.

La Figure 153 présente le graphique des volumes turbinés (Eq. 75) et déversés (Eq. 76) en fonction de la capacité utile de la retenue et de la puissance de turbinage. En accord avec les graphiques précédents, le cas étudié est celui des prévisions parfaites, afin d'examiner la sensibilité du modèle de gestion aux caractéristiques de l'aménagement. En ce qui concerne le volume turbiné, la Figure 153 montre qu'il dépend principalement de la capacité utile de retenue et que, pour l'ordre de grandeur de coefficient de turbinage testé ici, il est peu influencé par le coefficient de turbinage. Par conséquent, l'évolution du volume turbiné est linéaire pour tous les coefficients de turbinage étudiés.

Même si le réservoir n'a déversé qu'une seule fois – en Mai 2008 - pendant cette période, la Figure 153 montre que les volumes déversés sont plus importants pour le coefficient de turbinage égal à 3. Plus le coefficient de turbinage est fort, moins importants sont les volumes déversés. Cette remarque confirme le principe que, plus la puissance de turbinage est importante, moins d'énergie est perdue par le déversement.

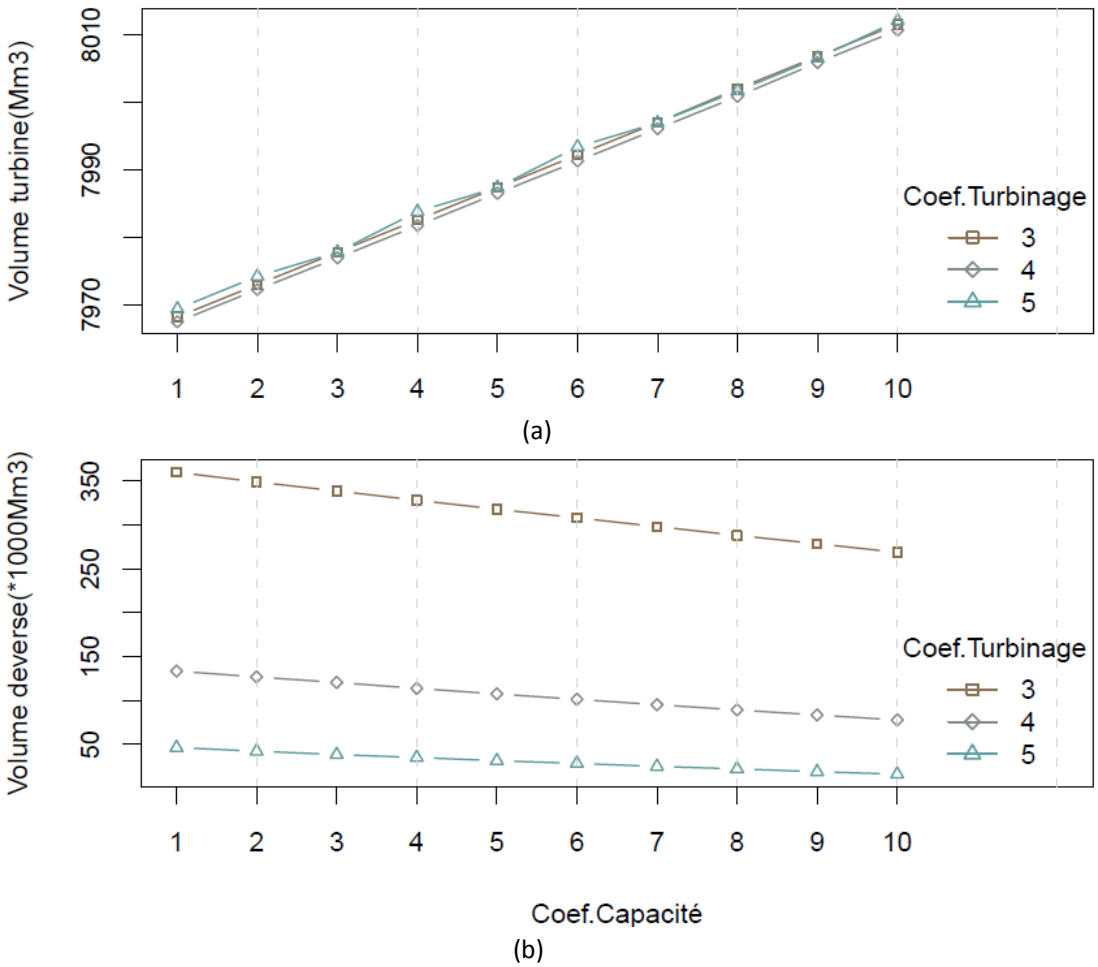


Figure 153: Evolution du (a) volume turbiné et (b) volume déversé pour les prévisions d'ensemble brutes (tous les membres sont considérés) en fonction du coefficient de capacité. Chaque ligne représente un différent coefficient de turbinage.

Toutes les conclusions établies ici s'appuient sur l'ensemble de la période d'évaluation et elles démontrent la sensibilité de la gestion d'un aménagement du type éclusée à ses caractéristiques : capacité et turbinage. Les graphiques présentés ci-dessus sont générés à partir de prévisions parfaites, afin de se concentrer uniquement sur la sensibilité du modèle aux caractéristiques de l'éclusée et non pas à la sensibilité vis-à-vis de la qualité des prévisions. Nous analysons cette dernière problématique dans les paragraphes suivants.

7.3.3 Intérêt des prévisions d'ensemble

La question qui se pose ici est d'évaluer l'intérêt (en valeur économique de gestion) de la prévision d'ensemble par rapport à une prévision naïve (climatologie), une prévision parfaite (observation) ou une prévision déterministe (ici, la prévision déterministe et la moyenne de l'ensemble). Dans un premier temps, la période d'évaluation est considérée dans son intégralité pour le cas du bassin versant de la Durance à Serre Ponçon.

La Figure 154 (a) à (c) présente le gain de gestion en MEuros (Eq. 78) pour la prévision d'ensemble, la climatologie, la prévision parfaite, la prévision déterministe et la moyenne d'ensemble, en fonction des différentes valeurs de capacité utile du réservoir et de coefficient de turbinage étudiés.

La Figure 154 (a) à (c) illustre le gain de gestion pour les coefficients de turbinage 3 à 5. D'un côté, nous observons que la prévision climatologique est celle qui donne le plus petit gain, indépendamment des caractéristiques considérées pour l'aménagement. D'un autre côté, les gains de gestion obtenus à partir des prévisions parfaites sont les plus importants dans tous les cas considérés. De plus, nous remarquons que l'information tirée d'une prévision (déterministe ou probabiliste) contribue à des gains supérieurs à ceux d'une prévision naïve (climatologie). Ainsi l'apport d'une prévision est démontré très nettement.

Plus particulièrement, selon ce graphique les prévisions d'ensemble apportent des gains très élevés et comparables à ceux de la prévision parfaite pour presque toutes les capacités de la retenue examinée. Une explication possible est liée au fait que très souvent, et surtout dans l'horizon d'une semaine, les prévisions d'ensemble sont fiables (l'observation tombe dans le faisceau des ensembles). En résultat, elles prévoient un volume d'apport proche du volume observé et elles produisent ainsi une gestion proche de celle produite par les prévisions parfaites.

De plus, la Figure 154 révèle aussi que les prévisions d'ensemble apportent un gain supérieur à celui de la moyenne des prévisions d'ensemble pour tous les coefficients de capacité testés. Les résultats sont similaires pour les prévisions déterministes. Toutefois, cette différence en termes de gain s'accentue avec la capacité de turbinage : plus le turbinage est fort, plus la valeur ajoutée de prévisions d'ensemble devient évidente. Ce résultat prouve qu'une prévision uni-valeur apportant une information réduite, aboutit en conséquence à un gain de gestion moins important. En ce qui concerne les prévisions déterministes, nous remarquons que leur performance est améliorée quand la capacité du réservoir est limitée (coefficient de capacité 1 ou 2), par contre elle se dégrade pour les coefficients de capacité supérieurs à 2.

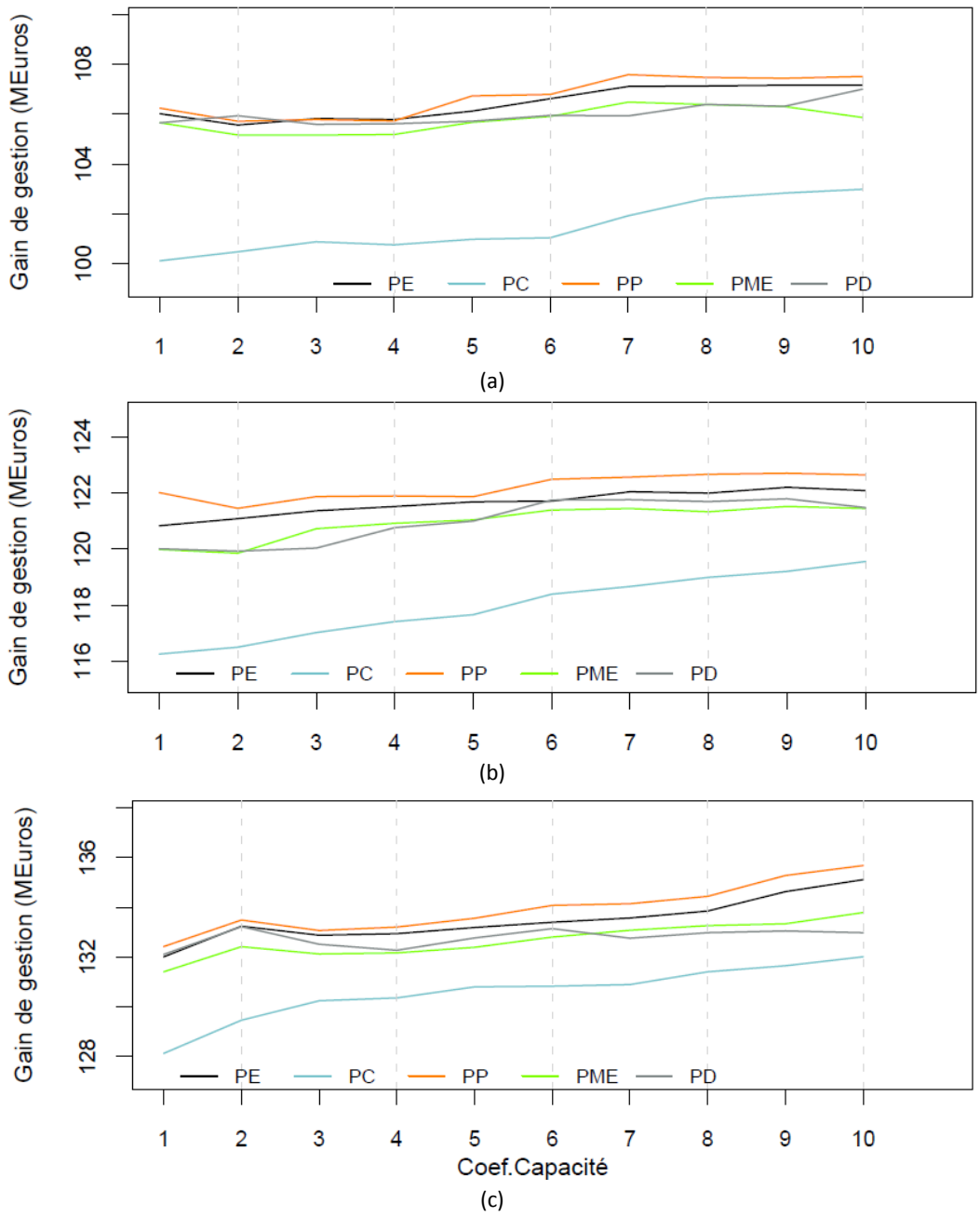


Figure 154 : Le gain de gestion en fonction du coefficient de capacité pour des coefficients : (a) 3, (b) 4 et (5) de turbinage et pour des prévisions : d'ensemble brutes (PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD).

Le deuxième critère d'évaluation est le volume turbiné (Figure 155). Dans ce graphique nous traçons le volume turbiné en fonction de la capacité de l'aménagement pour chaque coefficient de turbinage. Ici nous considérons le coefficient de turbinage égal à 3 (les conclusions obtenues sont similaires pour l'ensemble des coefficients de turbinage testés).

Sur ce graphique, nous constatons qu'en général le volume turbiné produit par la prévision parfaite est le plus important, suivi par le volume produit par la prévision d'ensemble. La prévision d'ensemble, la moyenne d'ensemble et la prévision deterministe démontrent des volumes turbinés qui s'alignent -pour ce coefficient de capacité en particulier- et qu'ils sont comparables en général. Cependant leurs valeurs restent inférieures à celles des prévisions parfaites. En ce qui concerne la climatologie, le volume turbiné est légèrement inferieur aux prévisions d'ensemble ; cependant il n'est pas turbiné aux moments où les prix sont les plus élevés (gestion non optimale). Toutefois, les volumes turbinés produits par les prévisions parfaites, d'ensemble, la climatologie, la deterministe et la moyenne d'ensemble évoluent presqu'en parallèle. Les résultats sont similaires pour tous les coefficients de turbinage examinés.

Quant aux volumes déversés, les différents types de prévisions ont donné des résultats comparables. Le volume déversé est principalement influencé par la puissance du turbinage et moins par les types de prévisions. De ce fait, pour un même coefficient de turbinage, les prévisions d'ensemble, les prévisions parfaites, la climatologie, la prévision déterministe et la moyenne d'ensemble présentent des courbes de volume déversé parallèles. Cependant, le volume déversé issu d'une gestion à partir de la climatologie et la prévision déterministe reste significativement plus élevé que le volume déversé calculé par les prévisions d'ensemble. La Figure 156 illustre le cas pour le coefficient de turbinage égal à 3.

Le gain de gestion est en partie composé par le volume turbiné et en partie par la valeur potentielle du stock en fin de période de gestion (Eq. 78). La Figure 157 présente l'évolution de la différence du stock en fonction de la capacité de la retenue pour un coefficient de turbinage égal à 3. La différence du stock produite par les prévisions parfaites est la plus élevée, suivie par la prévision d'ensemble, la moyenne d'ensemble et la prévision deterministe – similairement à la Figure 155. La même tendance générale est observée pour les autres coefficients de turbinage testés.

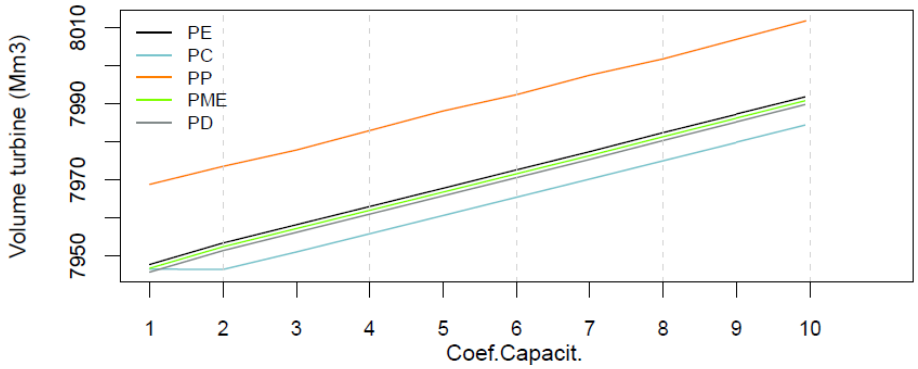


Figure 155 : Le volume turbiné en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes (PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD).

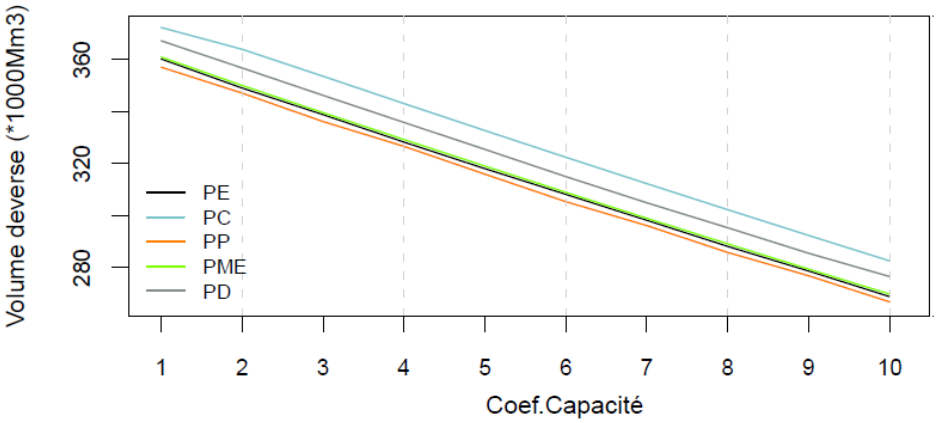


Figure 156: Le volume déversé en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes(PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD).

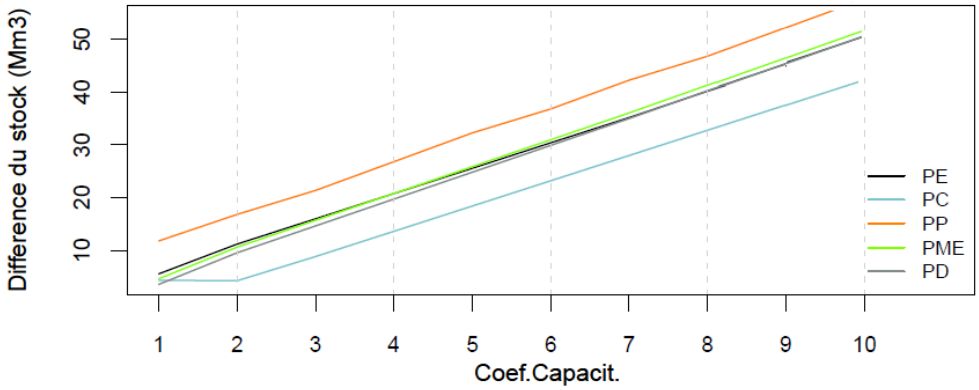


Figure 157: La différence du stock en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes(PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD).

7.3.4 Impact de la qualité des prévisions sur les gains de gestion

Dans ce paragraphe, nous évaluons les gains de la gestion des apports en fonction de différentes méthodes de pré- et post-traitement appliquées aux prévisions d'ensemble brutes. Les approches testées sont celles présentées et mises en place dans le Chapitre 7. Au total, nous examinons huit apports en prévision différents :

1. la prévision d'ensemble brute (sans traitement statistique),
2. la prévision d'ensemble avec l'habillage à l'aide des analogues,
3. la prévision d'ensemble avec l'habillage à l'aide de l'erreur du modèle hydrologique,
4. la prévision d'ensemble avec la combinaison des deux habillages précédents : analogues et erreur du modèle hydrologique,
5. la prévision d'ensemble avec l'habillage par la méthode du meilleur membre – acronyme *BMM 2003* (Roulston et Smith, 2003),
6. la prévision d'ensemble avec l'habillage par la méthode du meilleur membre modifiée – acronyme *BMM 2005* (Wang et Bishop, 2005),
7. la prévision d'ensemble avec l'habillage par la méthode du meilleur membre modifiée en prenant en compte la covariance des échéances,
8. la prévision d'ensemble avec l'habillage à partir des erreurs de 30% des meilleurs membres de l'ensemble,
9. la prévision d'ensemble avec l'habillage par la méthode du meilleur membre modifiée – acronyme *BMM 2006* (Fortin *et al.*, 2006).

7.3.4.1 Evaluation sur un bassin versant

Nous avons appliqué le modèle de gestion aux prévisions de la Durance à Serre Ponçon, que nous avons pré– et/ou post-traité avec les approches statistiques citées ci-dessus. Différents coefficients de turbinage et de capacité utile du réservoir ont également été testés.

La Figure 158 présente les gains de gestion obtenus en fonction de la méthode de correction statistique utilisée (axe des abscisses) et des caractéristiques de l'aménagement. Les 3 groupes colorés correspondent à différents coefficients de turbinage, tandis que les boîtes à moustaches montrent la variabilité des gains par rapport aux coefficients de capacité testés (Ck = 3 à 10 jours de capacité utile). Comme espéré, le gain augmente quand le turbinage devient plus puissant.

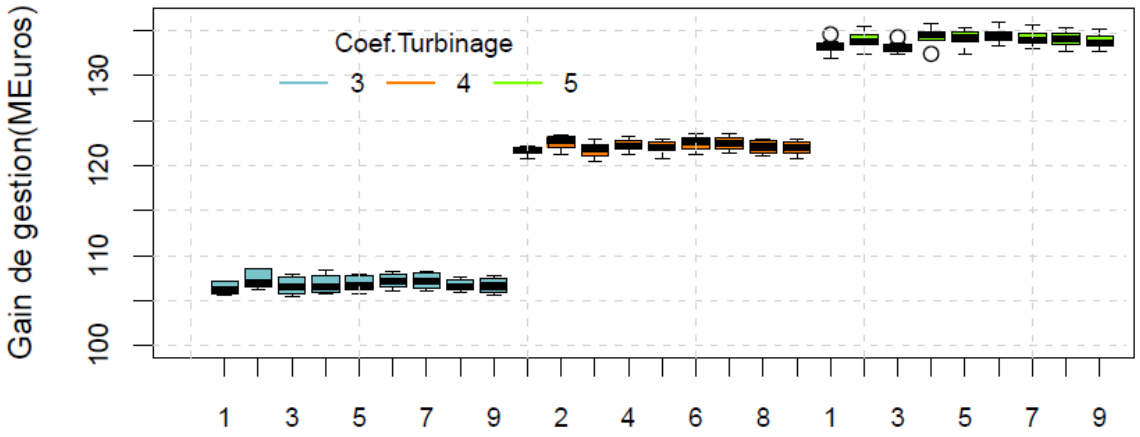


Figure 158: Le gain de gestion en MEuros pour différents coefficients de capacité (boîtes à moustaches) en fonction du coefficient de turbinage et par type de prévision étudié : (1) brutes, (2) habillage avec les Analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les Analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) habillées avec l'erreur de 30% de meilleurs membres et (9) habilées avec la BMM(2006).

Afin d'avoir une idée plus claire sur la performance de chaque approche de traitement statistique, nous introduisons le score de gain économique (Eq. 74) de gestion pour les différents coefficients de capacité dans la Figure 159. Les boîtes à moustache affichées correspondent aux différentes approches de traitement statistique. Les résultats sont présentés pour les trois coefficients de turbinage examinés. Pour tous ces coefficients de turbinage, presque toutes les prévisions d'ensemble (traitées et non-traitées) apportent un gain par rapport aux prévisions deterministes.

Plus particulièrement, pour un turbinage de puissance 3 et 4 (Figure 159 (a) et (b)), le score de gain est supérieur à 1 (qui est équivalent à un gain supérieur à celui produit par les prévisions d'ensemble brutes) pour presque toutes les approches testées. Plus particulièrement l'habillage avec les analogues - sans ou avec l'erreur du modèle - (2 et 4), ainsi que les méthodes du meilleur membre modifiées et ses variants (6 à 9) apportent le gain le plus élevé par rapport aux prévisions brutes pour tous les coefficients de capacité examinés.

De même, quand le turbinage atteint une puissance maximale égale à 5 (Figure 159 (c)), la plupart des approches de traitement statistique, à l'exception d'habillage avec l'erreur du modèle (3), apportent un gain supplémentaire par rapport aux prévisions d'ensemble brutes. L'habillage avec l'erreur du modèle (3) devient l'approche la moins performante par rapport aux prévisions brutes (scores de gain inférieurs à 1). Les résultats montrent aussi que les gains les plus significatifs sont apportés par les prévisions habillées avec les analogues et l'erreur du modèle hydrologique conjointement, ainsi que par les prévisions corrigées avec la méthode du meilleur membre modifiée.

Ce paragraphe a démontré en termes de gain économique de gestion l'intérêt d'appliquer les prévisions d'ensemble à la gestion des aménagements hydrauliques. Plus particulièrement, pour l'ensemble des coefficients de turbinage, presque toutes les prévisions d'ensemble (traitées et non-traitées) apportent un gain par rapport aux prévisions déterministes. De plus, ci-dessus nous avons étudié le gain économique apporté par différentes approches de traitement statistique de prévision. En général la plupart des approches statistiques testées arrivent à améliorer le gain de gestion. La performance des approches diffère selon le coefficient de turbinage, néanmoins l'habillage avec les analogues (couplé ou pas avec l'erreur du modèle) ; ainsi que les modifications de la méthode du meilleur membre (6 à 9) sont les plus performantes pour tous les coefficients de turbinage testés.

Cependant, la performance de l'habillage avec l'erreur du modèle hydrologique varie considerablement avec le coefficient de turbinage – et surtout au coefficient le plus élévé. Une explication possible est liée au fait que la correction avec l'erreur du modèle hydrologique a tendance à habiller la prévision d'ensemble brute avec une dispersion importante. Ceci donné à une fiabilité considérablement ameliorée, et une capacité à discriminer entre les événements pratiquement intouchée (voir : §6.5.1). Ce résultat en combinaison avec ceux tirés de la Figure 159 nous conduit à la conclusion qu'une dispersion importante ne conduit pas forcement à une gestion optimale : le gestionnaire recherche alors une stratégie " passe-partout " qui peut perdre les avantages d'une stratégie spécifique ciblant la partie centrale de la distribution des scénarios.

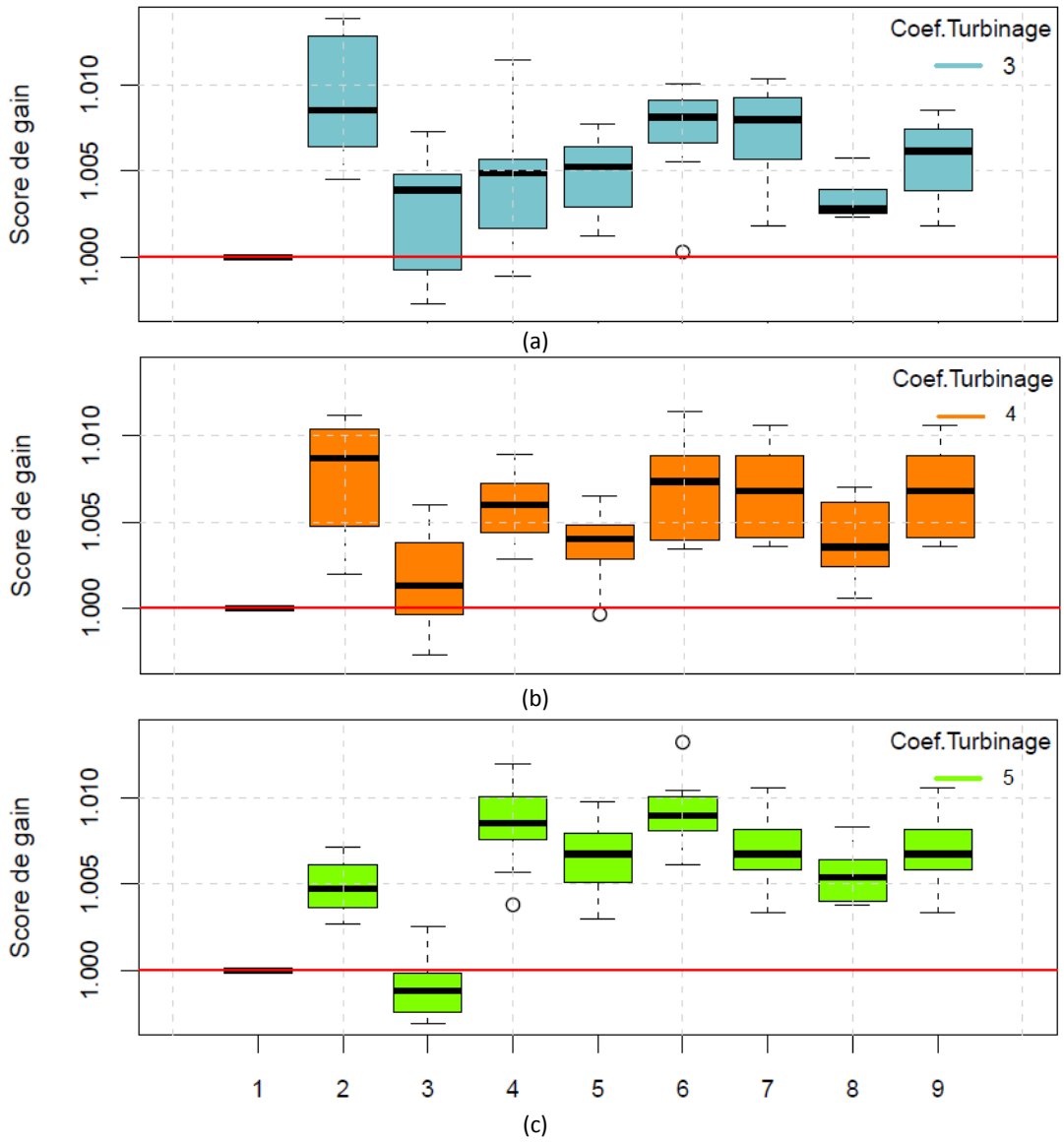


Figure 159: Gain de gestion de turbinage pour les coefficients de turbinage 3(a), 4(b) et 5(c) par type de prévisions : (1) brutes, (2) habillage avec les analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) habillées avec l'erreur de 30% de meilleurs membres et (9) habillées avec la BMM (2006).

7.3.4.2 Evaluation sur l'échantillon de 11 bassins versants

Précédemment nous avons montré que les méthodes de traitement statistique apportent un gain de gestion supplemantaire par rapport aux prévisions brutes. Afin de tester la robustesse de ces conclusions, il faut utiliser un échantillon de bassins versants plus large. L'échantillon utilisé ici est celui des 11 bassins versants à fort enjeu économique pour EDF (Chapitre2, §2.3).

La Figure 160 présente le score de gain économique de gestion lié aux différentes approches de traitement statistique par rapport aux prévisions brutes. Le graphique présenté correspond à des coefficients de capacité et de turbinage égaux respectivement à 4 et 3. Les boîtes à moustache correspondent aux scores de 11 bassins versants. La ligne noire épaisse montre la médiane, les bords bas et hauts de la boite correspondent aux quantiles 25% et 75% et les extrémités basses et hautes les valeurs minimales et maximales. La ligne rouge horizontale est le seuil au dessus duquel les prévisions corrigées apportent un gain vis-à-vis des prévisions d'ensemble. Les bassins versants présentant la meilleure performance en termes de gain économique sont présentés dans la Figure 161.

D'abord nous observons que la plupart des approches arrivent à améliorer les prévisions d'ensemble par rapport aux prévisions brutes pour un nombre important de bassins versants. Nous remarquons que la méthode la plus performante pour le plus grand nombre de bassins versants est la méthode du meilleur membre modifiée par Wang et Bishop (*6*). Les résultats de la méthode du meilleur membre modifiée avec la covariance des échéances (7), de l'habillage avec l'erreur de 30% des meilleurs membres (8) ; ainsi que celles de la méthode modifiée par Fortin *et al.* (9) démontrent également qu'elles apportent un gain amélioré - similaire à celui de la méthode du meilleur membre modifiée (6) - par rapport aux prévisions brutes pour la plupart de bassins testés. Cependant, la méthode du meilleur membre (*5*) arrive à améliorer le gain de gestion des prévisions brutes que sur un nombre limité des bassins versants. L'habillage avec l'erreur du modèle hydrologique – avec ou sans l'information des analogues - améliore considérablement la performance des prévisions brutes pour plus de la moitié des bassins versants examinés.

Pourtant, nous remarquons que la performance des approches testées varie selon le bassin versant. Afin d'approfondir la sensibilité des méthodes de traitement statistique à la localisation du bassin versant, nous avons affiché sur la Figure 161 un diagramme circulaire avec les deux méthodes les plus performantes en termes de score de gain par bassin versant. Chaque couleur correspond à une approche statistique. Ce graphique correspond aussi à des coefficients de capacité et de turbinage égaux respectivement à 4 et 3.

Sur les bassins versants des Alpes, la méthode du meilleur membre modifiée (*6*) et sa variante (7) apportent les gains les plus élevés. Sur le reste des bassins versants, et surtout les bassins versants du Massif Central et des Cévennes, l'habillage avec l'erreur du modèle (sans ou avec l'information des analogues) démontre la meilleure performance par rapport à toutes les méthodes testées.

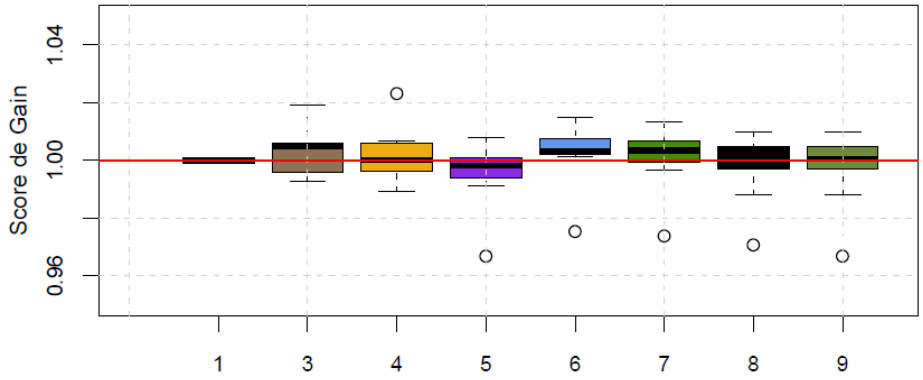


Figure 160: Boîtes à moustaches du score de gain par méthode de post-traitement appliquée. Le graphique présenté correspond à des coefficients de capacité et de turbinage égal à 4 et 3 respectivement.

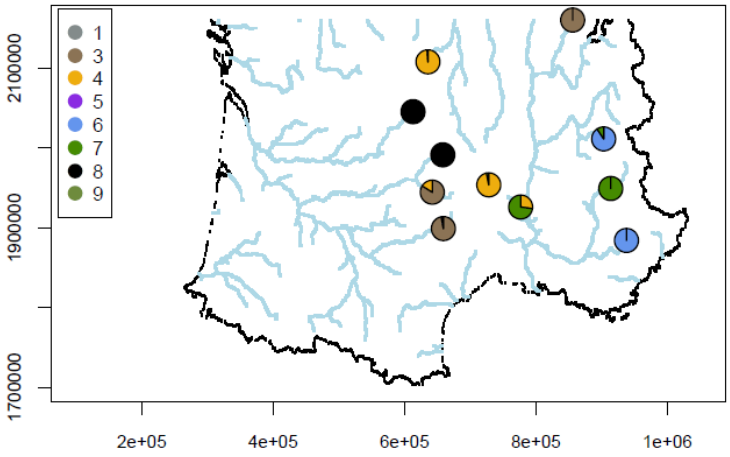


Figure 161: Score de gain en fonction de la localisation du bassin versant par type de prévisions : (1) brutes, (2) habillage avec les analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) avec l'erreur de 30% des meilleurs membres et (9) avec la BMM (2006). Le graphique présenté correspond à des coefficients de capacité et de turbinage égaux à 4 et 3 respectivement.

7.4 Synthèse

Dans ce chapitre nous avons essayé de valoriser les prévisions d'ensemble en termes économiques. Pour pallier l'absence d'un outil d'optimisation économique adapté aux prévisions d'ensemble nous avons créé un modèle heuristique d'optimisation. La fonctionnalité de ce modèle, la robustesse de ses résultats, ainsi que la sensibilité de la gestion en fonction des paramètres sont examinées dans ce chapitre.

En ce qui concerne la sensibilité de la gestion en fonction des paramètres, il faut noter que les principaux paramètres de notre modèle sont le coefficient de capacité utile du réservoir et le coefficient de turbinage. Nous avons constaté que la gestion est particulièrement sensible au coefficient de turbinage. En augmentant la puissance de turbinage, le gain de gestion, le volume turbiné et la différence du stock deviennent plus importants alors que le volume déversé diminue. De plus la sensibilité de ces vecteurs au coefficient de capacité dépend aussi de la puissance de turbinage. La capacité de la retenue est plus sensible quand le coefficient de turbinage est, soit très petit, soit très grand. A ce sujet il faut remarquer que lors de notre étude nous avons examiné différents coefficients de turbinage (2 à 6). Néanmoins, nous avons choisi de ne présenter ici que les coefficients de turbinage de 3 à 5, car ils correspondent le mieux à un aménagement du type " éclusée ".

Par conséquent l'analyse présentée ici démontre que, selon la capacité de la retenue et les objectifs de gestion, un gestionnaire pourra appliquer notre modèle de gestion à n'importe quel aménagement de type éclusée en choisissant le coefficient de turbinage le plus adapté. Cela prouve que notre modèle heuristique peut servir de base à outil générique et facilement applicable.

De plus, nous avons démontré que les prévisions d'ensemble possèdent – sur l'ensemble de la période d'évaluation – une valeur supplémentaire par rapport à une prévision naïve (climatologie) et qu'elles sont également les plus performantes au regard de la prévision déterministe et la moyenne d'ensemble. De ce fait nous avons clairement demontré que la prévision d'ensemble apporte une valeur ajoutée pour la décision, pas seulement dans les cas des événements rares pour lesquelles elle est souvent consultée par les operationnels, mais aussi dans les cas des problèmatiques de la décision en "quotidien".

Lors de notre étude nous avons montré également qu'il est essentiel de prendre en compte l'ensemble de l'information incluse dans les prévisions probabilistes car le gain de gestion apporté est plus important.

Ces conclusions sont en désaccord avec McCollor et al. (2008) qui ont montré sur la base d'un bassin versant et de 2 ans de prévisions (2003-2005) que la moyenne d'ensemble est légèrement plus performante que les prévisions d'ensemble sur un horizon de prévision de 1 à 2 jours. Les modèles de gestion mis en place par McCollor et al. (2008) étaient un modèle coût - perte statique et un modèle de Théorie de décision. Ce désaccord peut être attribué à la longueur de la période d'évaluation (notre période d'évaluation est significativement plus longue : 2005-2008). De plus, il faut noter que depuis 2005 la qualité des prévisions d'ensemble fournies par les centres météorologiques s'est considérablement améliorée.

Un autre sujet que nous avons abordé lors de ce chapitre est la performance de diverses méthodes de traitement statistique en termes de valeur économique. Les conclusions montrent que la qualité des prévisions influence fortement la performance de la gestion et le gain produit. Dans cette partie de la thèse nous avons clairement demontré qu'une prévision d'ensemble (traitée ou non-traitée), qui est performante pour tous les attributs de qualité, produira un gain de gestion elevé. Nous avons montré qu'en général l'application d'un traitement statistique arrive à améliorer le gain de gestion produit par rapport à celui des prévisions d'ensemble brutes pour tous les bassins versants examinés. Cependant, certaines approches statistiques sont plus adaptées à certains bassins versants et régimes hydrologiques que d'autres.

En comparant les conclusions du paragraphe §7.3.4 avec celles tirées du Chapitre 6 (§6.5.4), nous constatons que, en ce qui concerne les performances de différentes approches de traitement statistique testées, la tendance principale est commune : l'habillage avec l'erreur du modèle (avec ou sans l'aide des analogues) se montre le plus performant en termes de qualité et d'utilité dans les bassins du Massif Central et les Cévennes ; ainsi que la méthode du meilleur membre et ses variantes sont plus performantes dans les Alpes. Cette conclusion nous permet de constater que les critères de qualité peuvent donner également une première estimation qualitative du gain économique potentiel de gestion. Bien évidemment, ils ne remplacent pas les outils de simulation de gestion, mais si un opérationnel n'a pas à sa disposition un outil de simulation de gestion, il peut examiner la performance des prévisions en termes d'attributs de qualité afin d'obtenir une estimation qualitative du gain potentiel de gestion.

“Sailors…say the weather is a great bluffer. I guess the same is true for our human society – things can look dark, then a break shows in the clouds, and all is changed.”

Elwyn Brooks White

8 Conclusions et perspectives

Dans ce chapitre, nous présentons les principales conclusions de cette thèse, en mettant en évidence les apports de notre recherche à l’étude des liens entre la qualité et l’utilité des prévisions hydrologiques d’ensemble. Nous proposons une " feuille de route " pour guider le choix parmi les différents critères d’évaluation existants dans la littérature et illustrés dans les chapitres précédents. Enfin, nous évoquons quelques perspectives de recherche qui apparaissent à l’issue de cette thèse.

8.1 Les objectifs, données et outils de cette thèse

Au début de cette thèse, nous avions une double intention derrière nos objectifs de recherche : a) éclairer les défis principaux des prévisions hydrologiques d'ensemble et b) examiner différents chemins possibles pour fournir des prévisions *bonnes* et *utiles* aux utilisateurs-décideurs.

Pour cette raison, notre travail s'est appuyé sur l'évaluation des principaux aspects d'une prévision : sa qualité et son utilité. Toute la démarche de cette thèse a été basée sur ces questions à répondre :

- *comment définir la qualité d'une prévision et quels sont les principaux attributs à rechercher dans un système de prévision hydrologique d'ensemble ?*
- *est-ce que nous arrivons à améliorer la qualité des prévisions hydrologiques d'ensemble en utilisant des approches de correction statistique de biais simples et faciles à implémenter opérationnellement ?*
- *est-ce que les prévisions hydrologiques d'ensemble apportent un gain économique supplémentaire par rapport aux prévisions déterministes et climatologiques dans un contexte de production hydroélectrique ?*

Afin de pouvoir apporter des éléments de réponse à ces questions, nous avons conceptualisé une chaîne de prévision qui s'étend de la sortie du modèle météorologique jusqu'à la valorisation des prévisions par l'utilisateur final (Figure 3). Dans cette chaîne, le rôle clé de l'évaluation de la qualité des prévisions, du traitement statistique de leurs biais et de l'utilité des prévisions a été mis en évidence, avec en parallèle leurs principaux défis actuels (Chapitre 1, §1.4).

Ces défis nous ont incités à étudier les prévisions hydrologiques d'ensemble à l'aide de plusieurs jeux de données et outils d'investigation. Ce travail a ainsi été basé sur :

- 2 jeux de bassins versants (avec 74 et 11 bassins français) : l'un avec 74 grands bassins versants et l'autre avec 11 bassins de fort intérêt pour la production hydroélectrique. Ces bassins sont caractérisés par de tailles et régimes hydrologiques différentes (Chapitre 2, §2.3),
- une base de données historique d'observations en pluie, température et débits sur plus de 40 années sur chaque bassin versant étudié (Chapitre 2, §2.3),

- 2 modèles hydrologiques globaux de représentation de la transformation pluie-débit, ici calés sur chaque bassin versant étudié : le modèle MORDOR à 10 et 11 paramètres et le modèle GRP, à 3 et 5 paramètres, avec et sans module neige, respectivement (Chapitre 2, §2.4),

- une archive des prévisions météorologiques d'ensemble (précipitation et température) d'une longueur variable entre 2 et 4 années mise en place à partir des prévisions émises par 8 centres météorologiques (base de prévisions TIGGE) et des prévisions issues de l'application d'une méthode de recherche de situations analogues, pour un horizon maximale de prévision de 15 jours (Chapitre 2, §2.2),

- 7 critères d'évaluation de la qualité des prévisions hydrométéorologiques d'ensemble, sélectionnés dans la littérature pour évaluer les attributs de fiabilité, précision, justesse, dispersion, discrimination et performance globale des prévisions (Chapitre 3, §3.2),

- 2 stratégies pricipales de combinaison de prévisions multi-modèles, pour évaluer l'apport de l'utilisation de prévisions émises par différents centres météorologiques à la prévision hydrologique d'ensemble (Chapitre 5, §5.3),

- 9 stratégies de prétraitement (correction statistique des précipitations prévues) et post-traitement (correction statistique des débits prévus) mises en place dans le but d'améliorer la qualité des prévisions hydrologiques d'ensemble.

  Les approches empiriques ont été choisies grâce à leur simplicité et à leur facilité d'application en opérationnel. Les approches utilisées appartiennent principalement à la famille de correction par habillage en utilisant l'information des analogues, une archive des erreurs du meilleur membre ou celles du modèle hydrologique. Les stratégies conçues ont ciblé le traitement à la fois de l'incertitude totale des prévisions ou d'une source particulière d'incertitude, ici l'erreur du modèle hydrologique. Elles ont été appliquées soit uniquement aux précipitations ou aux débits, soit aux deux (Chapitre 3 et Chapitre 6)

- Un modèle heuristique de gestion des apports en débit aux aménagements hydroélectriques, développé dans cette thèse pour évaluer le gain en termes d'impact économique apporté par les prévisions à moyen-terme, plus particulièrement par les prévisions probabilistes (d'ensemble), sur la gestion de la production d'énergie d'une retenue à capacités de réservoir et turbinage variables.

Ce modèle nous a permis d'estimer le gain économique supplémentaire apporté par les prévisions d'ensemble en le comparant avec celui des prévisions déterministes ou des prévisions basées la climatologie. De plus, à l'aide du modèle développé, nous avons également examiné l'influence de la qualité des prévisions sur le gain économique apporté, en comparant le gain des prévisions corrigées à ce des prévisions brutes (sans aucun pré-ou post-traitement) (Chapitre 4 et Chapitre 7).

Les principales conclusions de cette thèse sont résumées ci-après.

8.2 Conclusions

- Sur évaluation des prévisions météorologiques et hydrologiques d'ensemble

Dans cette thèse, l'évaluation de la qualité des prévisions a été effectuée soit sur la base de prévisions météorologiques TIGGE avec des prévisions fournies par 8 centres météorologiques, sur le jeu de 74 bassins versants, soit sur les prévisions du centre européen CEPMMT sur le jeu de 11 bassins versants. Pour la transformation de la pluie en débit, nous avons utilisé les modèles hydrologiques *MORDOR* et *GRP*.

Nous avons remarqué que les prévisions d'ensemble de précipitation (pluie moyenne de bassin) sont plutôt fiables pour la majorité des bassins versants étudiés et notamment - aux échéances plus longues de prévision (au-delà de 5 jours). Toutefois nous avons constaté un effet important de diminution de la dispersion des membres de l'ensemble et de perte de fiabilité dans les prévisions de débit obtenues à la sortie du modèle hydrologique, qui pourrait être lié au rôle intégrateur du bassin versant ou à l'influence du modèle hydrologique.

Ce fait est surtout observé dans le cas du modèle *GRP* et il est moins intense pour le cas du modèle *MORDOR*. Une explication possible pourrait être liée à la mise à jour spécifique du modèle *GRP*, qui impose que toute prévision démarre à partir du dernier débit observé. Nous retiendrons néanmoins que les prévisions hydrologiques des deux modèles affichent un défaut de qualité lié à la sous-dispersion des prévisions d'ensemble, ce qui relève un défis au traitement statistique des prévisions.

Si d'un côté les prévisions hydrologiques brutes montrent un manque de fiabilité, d'un autre côté, nous remarquons que le modèle hydrologique contribue à améliorer la précision par la capacité des prévisions à discriminer des événements binaires (par exemple, le dépassement ou le non-dépassement d'un seuil critique). En général, les prévisions hydrologiques affichent une meilleure performance selon cet attribut que les prévisions d'ensemble de précipitations.

Une dépendance forte entre la performance des prévisions hydrométéorologiques d'ensemble et la surface du bassin versant n'a pas pu être identifié sur le jeu de 74 grands bassins versants (surfaces variant de 1000 km² à 10000 km²). Néanmoins, sur l'échantillon de 11 bassins versants (surfaces variant de 60 km² à 3600 km²), les critères de performance globale des prévisions ont montré une légère amélioration de la qualité des prévisions pour les bassins versants avec une surface supérieure à environ 1000 km².

Parmi les différents centres météorologiques de la base TIGGE ici évalués, nous avons remarqué que certains fournissent des prévisions de précipitation de meilleure qualité que d'autres pour une majorité de bassins versants. Il s'agit du centre européen, du centre japonais et du centre canadien. Toutefois, il n'y a pas un centre qui est le meilleur pour tous les bassins versants. Nous avons également constaté que la performance des prévisions des différents centres dépend de la localisation du bassin versant, ainsi que de l'échéance de prévision.

- *Sur le traitement statistique de biais de prévision d'ensemble*

Dans un premier temps, nous avons examiné la correction des défauts de fiabilité des prévisions hydrologiques à partir de la combinaison des prévisions obtenues à partir de différents centres météorologiques fournisseurs des prévisions d'ensemble de précipitation, autrement dit par une approche " multi-modèle ".

La combinaison des prévisions d'ensemble provenant de différents centres, et par conséquence, modèles météorologiques n'est pas toujours une démarche simple. Dans cette thèse, nous avons proposé deux approches différentes de combinaison des prévisions : la première est basée sur l'équiprobabilité des centres et des membres des ensembles, et la deuxième, sur une pondération des prévisions d'ensemble des différents centres en fonction de leur performance par rapport à un critère numérique de qualité des prévisions, ici le score de probabilité ordonnée *RPS*.

Nous avons montré que le grand ensemble multi-modèle améliore considérablement la qualité globale des prévisions, ainsi que leur fiabilité et discrimination. La performance du grand ensemble pondéré en fonction de la performance des centres est légèrement meilleure que celle de l'approche basée sur l'équiprobabilité, surtout aux échéances plus lointaines (au-delà de 5 jours).

Une deuxième manière de correction des prévisions vient du traitement statistique des biais, qu'il soit appliqué en amont (prétraitement) ou en aval (post-traitement) du modèle hydrologique. Dans cette thèse, nous avons appliquée plusieurs approches empiriques existantes dans la littérature. Nous avons également proposé quelques modifications aux approches utilisées. Au total, neuf stratégies de traitement statistique ont été examinées (voir Chapitre 6, Tableau 8).

Nous avons conclue que presque tous les traitements statistiques appliqués améliorent soit la qualité globale des prévisions, soit une ou plusieurs de ses attributs. Le Tableau 10 résume l'influence de chaque approche sur les divers attributs de la qualité des prévisions

Nous avons également constaté qu'il n'y a pas une "meilleure" méthode qui permet d'améliorer la qualité des prévisions hydrologiques sur tous les bassins versants étudiées et pour les deux modèles hydrologiques utilisés. Toutefois, il y a des approches qui ressortent plus fréquemment que d'autres comme étant plus performantes : c'est le cas de l'habillage avec l'erreur du modèle hydrologique (sans ou avec, au préalable, le prétraitement des précipitations à l'aide de l'information des situations analogues), ainsi que de la méthode du meilleur membre modifiée par Wang et Bishop (2006) et ses variantes.

Nous avons observé que la performance des approches de traitement statistique peut être influencée par la longueur de la période de calage des paramètres. C'est notamment le cas pour l'habillage avec l'erreur du modèle et pour la méthode du meilleur membre modifiée par Fortin et al. (2006). Il s'est avéré qu'un nombre plus important d'années de données sont nécessaires pour caler les paramètres de ces méthodes.

Nous avons également examiné l'intérêt d'appliquer une approche de traitement statistique soit seulement sur les entrées du modèle hydrologique, soit seulement sur les sorties, soit sur les deux. Nous avons conclue que le passage par le modèle hydrologique peut limiter les effets positifs d'un traitement statistique appliqué aux prévisions de précipitation. Par conséquent, l'application d'une correction statistique des biais à la sortie du modèle hydrologique est presque obligée.

- *Sur la valorisation des prévisions hydrologiques d'ensemble*

Le modèle de gestion des réservoirs de production hydroélectrique que nous avons développé dans cette thèse est basé sur deux paramètres principaux que l'on peut faire varier pour mieux s'adapter à une situation donnée : le coefficient de turbinage de l'usine et la capacité de la retenue. Ces deux paramètres libres garantissent ainsi la facilité d'application du modèle à n'importe quel aménagement hydroélectrique.

Pour vérifier le caractère générique du modèle dans le cas des éclusées, nous l'avons appliqué aux apports en débit sur 7 jours et avons examinés plusieurs valeurs de ces paramètres. Le gain économique des prévisions en gestion de production électrique a été évalué sur la base du prix de marché de l'énergie (en €/MWh), supposé traduire la valeur de l'énergie.

Nous avons constaté que, pour le type d'aménagement étudié, la gestion est surtout sensible au coefficient de turbinage. Le gain économique supplémentaire qu'apportent les prévisions hydrologiques d'ensemble à l'utilisateur a été numériquement démontré. Nous avons montré que ce gain est supérieur à celui apporté par une prévision naïve (la climatologie) ou même à celui apporté par la prévision déterministe et par la moyenne de la prévision d'ensemble. Cette conclusion est importante car elle montre tout l'intérêt de l'utilisation de la prévision hydrologique d'ensemble sur une base quotidienne, pour des apports en eaux quotidiens, et non pas seulement lors des événements extrêmes, comme c'est souvent le cas à nos jours dans les services opérationnels.

Une autre problématique abordée dans cette thèse a été celle de la relation entre la qualité des prévisions et leur utilité (ici exprimée par le gain économique en gestion de la production hydroélectrique). Nous avons montré que le gain économique de gestion est fortement dépendant de la qualité des prévisions : plus la qualité des prévisions s'améliore, plus élevé devient le gain apporté par les prévisions

Nous avons ainsi observé qu'en utilisant des prévisions d'ensemble qui avaient été débiaisées à l'aide d'un traitement statistique appliqué à la sortie du modèle hydrologique (post-traitement), le gain économique de gestion apporté par les prévisions était plus élevé par rapport au gain obtenu à partir des prévisions brutes (sans aucun traitement statistique). Plus particulièrement, nous constatons que les approches statistiques, qui arrivent à améliorer la qualité globale des prévisions, mais aussi la fiabilité et la discrimination, conduisent à un gain économique plus important. Néanmoins, nous n'avons pas pu dégager une " meilleure " méthode à appliquer pour avoir le meilleur gain partout, car les performances des méthodes dépendent du bassin versant considéré.

En faisant la liaison entre les aspects de qualité et de l'utilité, nous pouvons déclarer qu'il suffit souvent d'examiner la performance d'une prévision en termes de qualité globale, fiabilité et discrimination, afin de constater si elle conduira à un gain économique potentiel. De cette façon, en cas de manque d'un outil adapté de gestion, une estimation qualitative du gain économique supplémentaire peut se faire par l'évaluation de la qualité de prévision.

Dans les cas de différentes approches de traitement statistique testées, cette conclusion est confirmée par l'examination des cartes indiquent, en parallèle et par bassin versant, les deux meilleures méthodes de traitement statistique qui donnent les meilleures performances en termes de qualité globale de prévision (critère *RPS* ; Figure 162, en haut) et de score de gain économique (Figure 162, en bas). Nous remarquons que 7 sur 11 bassins versants étudiés associent les mêmes meilleures méthodes de traitement en qualité et en utilité. De plus, la tendance qui ressort associe l'habillage avec l'erreur du modèle (sans ou avec l'aide des analogues– 3 et 4) aux bassins du Massif Central et des Cévennes et la méthode du meilleur membre et ses variantes (5 à 7) aux bassins des Alpes.

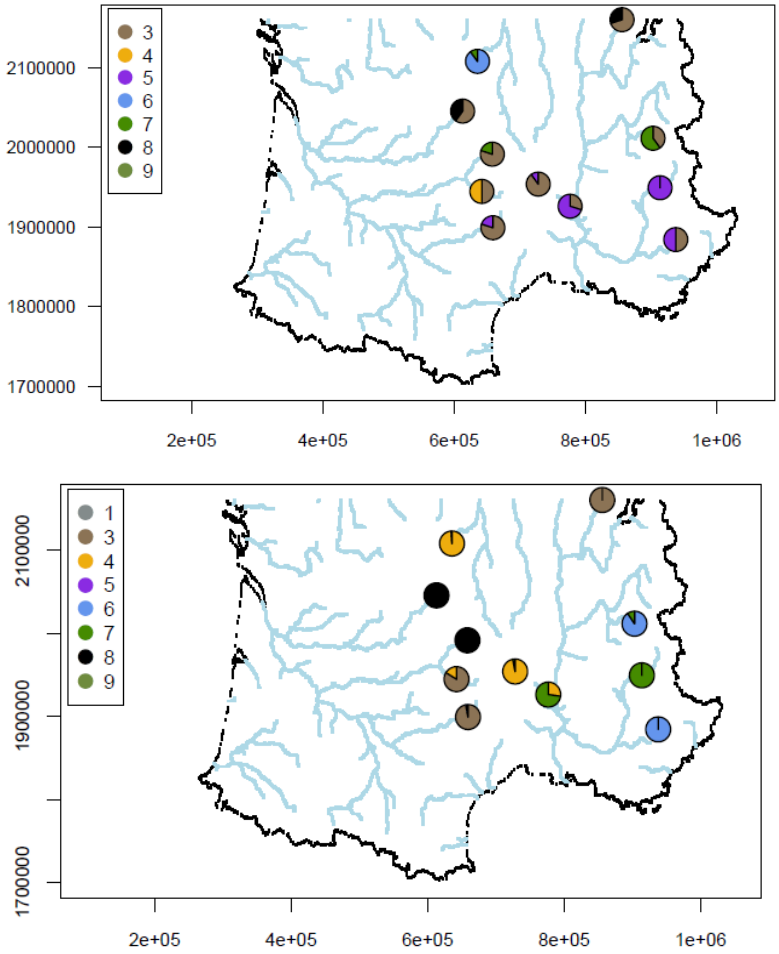


Figure 162: Lien entre qualité et utilité des prévisions : carte avec les deux meilleures méthodes de traitement statistique en termes de qualité globale de prévision selon le critère RPS (en haut) et de meilleur score de gain économique pour les coefficients de capacité et de turbinage égaux à 4 et 3, respectivement (en bas). Prévisions hydrologiques du modèle MORDOR à l'échéance 6 jours (pour les méthodes de traitement statistique, voir Tableau 8).

8.3 Perspectives

Une thèse n'a qu'une durée limitée, ce qui fait qu'il reste toujours plein de voies potentielles pour améliorer et poursuivre les travaux de recherche. De notre côté nous avons identifié trois principales questions qui méritent d'investigation supplémentaire.

- *Est-ce qu'une méthode de traitement statistique est partout et toujours la meilleure ?*

Dans cette thèse, nous avons essayé de traiter cette question. Néanmoins, la réponse n'est ni simple ni facile. En ce qui concerne le traitement statistique des prévisions, nous avons favorisé dans cette thèse les approches d'habillage empirique. Grâce à leur simplicité méthodologique et à leur facilité d'application, elles sont des méthodes attirantes dans un contexte opérationnel.

Toutefois, l'application d'autres méthodes (paramétriques et non paramétriques) plus élaborées, déjà développées dans le milieu de la recherche, offrira une opportunité enrichissante de comparaison. L'exploitation de telles approches répondra potentiellement à la question initiale que nous nous sommes posés (" *est-ce qu'il existe une meilleur méthode de traitement statistique partout et toujours ?* ").

La réponse doit être basée à l'application de plusieurs méthodes de traitement statistique sur un échantillon large de bassins versants. De plus, les recherches futures doivent également considérer les influences saisonnières sur la performance des différentes approches, ainsi qu'à leur performance en temps de crues ou d'étiage.

- *Comment aller plus loin avec une approche multi-modèle en utilisant une archive plus longue et actualisée ?*

Dans cette thèse nous avons proposé deux approches principales pour la combinaison des prévisions émises par les différents centres météorologiques de notre base de données. Etant donné la longueur limitée de l'archive à notre disposition (242 jours en commun à tous les centres), même si les combinaisons proposées offrent une alternative efficace, la robustesse des résultats mériterait d'être testée avec une archive plus longue. Actuellement, la base TIGGE possède plus de données et par conséquent nous considérons comme une perspective complémentaire aux travaux de cette thèse la poursuite des études de combinaison des différents centres en utilisant une archive actualisée.

De plus, il serait également intéressant de comparer les approches de pondération des centres proposées dans cette thèse avec d'autres approches de combinaison existantes dans la littérature, mais pas encore utilisées dans le cadre de la base TIGGE, telle que l'approche *BMA* (" *Bayesian Model Averaging* ") ou une approche de la famille des approches par régression.

- Est-ce que le modèle gestion proposé dans cette thèse est robuste ?

L'utilité de l'application d'un modèle heuristique de gestion pour évaluer le gain économique des prévisions a été clairement démontrée. Dans le contexte étudié (production hydroélectrique), les avantages de ce modèle par rapport à des modèles du type Coût - Pertes sont incontestables. Néanmoins, il serait intéressant de comparer la performance du modèle de gestion heuristique ici développé à celle d'un modèle plus élaboré, basé, par exemple, sur la programmation stochastique dynamique. Cette comparaison pourrait démontrer la robustesse de nos conclusions, ainsi que contribuer à mieux identifier les forces et faiblesses de l'approche heuristique.

Une autre piste possible est liée à l'utilisation d'autres bassins versants et de prévisions d'autres modèles météorologiques et/ou hydrologiques dans le modèle de gestion. Ici nous n'avons utilisé que les prévisions du centre européen *CEPMMT* et le modèle hydrologique *MORDOR* sur un échantillon de 11 bassins versants français. Il serait intéressant de tester la sensibilité de la gestion aux prévisions des différents centres météorologiques de la base TIGGE, y compris du grand ensemble, et du modèle *GRP* (voire d'autres modèles hydrologiques), sur un échantillon plus important de bassins versants en France ou ailleurs.

- Comment communiquer efficacement aux utilisateurs l'information des prévisions hydrologiques d'ensemble ?

Une autre problématique qui émerge de l'utilisation opérationnelle des prévisions d'ensemble est celle liée à leur communication. Les décideurs finaux, habitués à recevoir des prévisions déterministes, sont souvent réticents à l'utilisation d'une prévision probabiliste. Quelques études faites sur ce sujet (par exemple, Houdant, 2004 ; Ramos *et al.*, 2010 ; Norbert *et al.*, 2010 ; Pappenberger *et al.*, 2013) ont montré l'intérêt à trouver une plateforme pour familiariser les décideurs avec les prévisions d'ensemble. Ce sujet était hors des objectives de cette thèse, mais il émerge comme une des problématiques intéressantes à creuser par des études futures qui s'intéresseraient à l'utilité des prévisions hydrologiques d'ensemble.

Cette liste non exhaustive de perspectives montre quelques champs ouverts qui restent à explorer dans le domaine relativement nouveau de la prévision hydrologique d'ensemble. Il s'agit d'un domaine avec plusieurs acteurs fortement intéressés par les bénéfices des prévisions de bonne qualité et utiles. Plein de recherches fructueuses sont surement à venir.

Bibliographie

Addor, N., Jaun, S., Fundel, F., Zappa, M. (2011). An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scénarios. *Hydrol. Earth Syst. Sci.*, 15(7), 2327-2347.

Ahrens, B., Jaun, S. (2007). On evaluation of ensemble precipitation forecasts with observation-based ensembles. *Advances in Geosciences,* 10, 139–144.

Alfieri, L., Thielen, J. (2012). A European precipitation index for extreme rain-storm and flash flood early warning. *Met. Apps*. Online Version, doi: 10.1002/met.1328.

Andreatis, K. M., Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model, *Adv. Water Resour*., 29, 872–886.

Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*. Yrjo Jahnssonin Saatio : Helsinki, 65 pages.

Atger, F. (1999). The skill of ensemble prediction systems, *Monthly Weather Review*, 127, 1947-1953.

Atger, F. (2001). Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes Geophys*., 8, 401–417

Bao, L., Gneiting, T., Grimit, E. P., Guttorp, P. and Raftery, A. E. (2010). Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction, *Monthly Weather Review*, 138, 1811– 1821.

Bao, H.-J., Zhao, L.-N., He, Y., Li, Z.-J., Wetterhall, F., Cloke, H.L., Pappenberger, F. and Manful, D., (2011). Coupling ensemble weather predictions based on TIGGE database with Grid-Xinanjiang model for flood forecast, *Advances in Geosciences*, 29, 61–67.

Bao, H.J., Zhao, LN (2012). Development and Application of an Atmospheric-Hydrologic-Hydraulic Flood Forecasting Model Driven by TIGGE Ensemble Forecasts, *Acta Meteorologica Sinica*, 26 (1), 93-102.

Bartholmes, J.C. and Todini, E. (2005). Coupling meteorological and hydrological models for flood forecasting, *Hydrol. Earth Syst. Sci.*, 9, 333-346.

Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S. (2009). The European Flood Alert System EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.*, 13 (2), 141–153.

Bell, M., Martin, M., Nichols, N. (2004). Assimilation of data into an ocean model with systematic errors near the equator, *Q. J. Roy. Meteor. Soc.*, 130, 873–894.

Ben Daoud, A., Sauquet, E., Lang, M., Obled, C., Bontron, G. (2009). La prévision des précipitations par recherche d'analogues : état de l'art et perspectives, *L'Huille Blanche*, 6, 60-65.

Berthet, L. (2010). *Prevision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de debit dans un modele hydrologique*. These de doctorat. AgroParisTech-Cemagref, 603 pages.

Beven, K.J., Binley, A., (1992). The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes*, 6, 279-298.

Bishop, C.H., Etherton, B.J., Majumdar, S.J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Monthly Weather Review,* 129, 420–436.

Bogner, K., Pappenberger, F., and Cloke, H. L. (2011). Technical Note: The Normal Quantile Transformation and its application in a flood forecasting system, *Hydrol. Earth Syst. Sci. Discuss*., 8, 9275-9297.

Bogner, K., Pappenberger, F., (2011). Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research*, 47 (7), W07524.

Bogner, K., Kalas, M. (2008). Error-correction methods and evaluation of an ensemble based hydrologic forecasting system for the Upper Danube catchment*, Atmos. Sci. Lett*., 9(2), 95–102.

Bontron, G. (2004). *Prévision quantitative des precipitations : adaptation probabiliste par recherche d'analogues*. Thèse de doctorat, INPG, Grenoble, 257pages.

Boucher M.-A., D. Tremblay, L. Delorme, L. Perreault, F. Anctil (2012). Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology,* 416–417, 133–144.

Boucher, M.-A., Anctil, F., Perreault, L., and Tremblay, D. (2011). A comparison between ensemble and deterministic hydrological forecasts in an operational context, *Advances in Geosciences*, 29, 85-94.

Bougeault, P., Toth, Z., Bishop, C., Brown B., Burridge,D., Chen D. H., Ebert, B., Fuentes, M., Hamill, T.M., Mylne,K., Nicolau,J., Paccagnella,T., Park,Y.Y., Parsons, D. Raoult,B. Schuster,D., Silva P.D., Swinbank,R., Takeuchi,Y., Tennant, W., Wilson, L. and Worley, S., (2010). The Thorpex Interactive Grand Global Ensemble (TIGGE), *Bulletin of the American Meteorological Society*, 91, 1059–1072.

Bradley, A., Schwartz, S.S., Hashino, T. (2004). Distributions-Oriented Verification of Ensemble Streamflow Predictions. *J. Hydrometeorology*, 5, 532–545.

Brier, G. W. (1950). Verification forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1-3.

Brown, J. D., Demargne, J., Seo, D.J., Liu, Y. (2010). The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modeling & Software*, 25, 854-872.

Brown J.D., Seo, D.-J. (2010). A nonparametric post-processor for bias-correction of hydrometeorological and hydrologic ensemble forecasts. *Journal of Hydrometeorology*, 11(3), 642-665.

Brown J.D., Seo, D.-J. (2012). Evaluation of a nonparametric post-processor for bias-correction and uncertainty estimation of hydrologic predictions. *Hydrological Processes (Special issue on Hydrologic Ensemble Prediction System)*, 27(1), 83-105.

Buizza, R., Palmer, T. N. (1995). The singular-vector structure of the atmospheric global circulation, *J. Atmos. Sci*., 52, 1434–1456.

Buizza, R., Miller, M., Palmer, T.N. (1999). Stochastic representation of model uncertainties in the CEPMMT ensemble prediction system. *Q. J. Roy. Meteorol. Soc.,* 125 (560), 2887–2908.

Buizza, R. (2001). Accuracy and Potential Economic Value of Categorical and Probabilistic Forecasts of Discrete Events. *Monthly Weather Review*, 129(9), 2329-2345.

Buizza, R., Houtekamer, P.L., Pellerin, G., Toth, Z., Zhu, Y., Wei, M., (2005). A comparison of the CEPMMT, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133 (5), 1076–1097.

Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., Vitart, F. (2007a). The new CEPMMT VAREPS. *Q. J. Roy. Meteorol. Soc*., 133, 681-695.

Buizza, R., Asensio, H., Balint, G., Bartholmes, J., Bliefernicht, J., Bogner, K., Chavaux, F., de Roo, A., Donnadille, J., Ducrocq, V., Edlund, C., Kotroni, V., Krahe, P., Kunz, M., Lacire, K., Lelay, M., Marsigli, C., Milelli, M., Montani, A., Pappenberger, F., Rabufetti, D., Ramos, M.H., Ritter, B., W. Schipper, J., Steiner, P., Thielen-Del Pozzo, J., Vincendon, B. (2007b) EURORISK/PREVIEW report on the technical quality, functional quality and forecast value of meteorological and hydrological forecasts. *CEPMMT Technical Memorandum* 516, Reading, England, February 2007.

Buizza, R., (2008). The value of probabilistic prediction. *Atmos. Sci. Lett*., 9, 36–42.

Buizza, R., (2010). The Value of a Variable Resolution Approach to Numerical Weather Prediction. *Monthly Weather Review*, 138 (4), 1026-1042.

Candille, G. (2003). *Validation des systèmes de prévisions météorologiques probabilistes*. Thèse en Mathematiques Appliquées, Université Pierre et Marie Curie. Thèse de doctorat: 146 pages.

Casati, B., Wilson, L.J., Stephenson, D.B (2008) Forecast verification: current status and future directions, *Meteorological Applications*, 15 (1), 3–18.

Cloke, H., Pappenberger, F. (2009a). Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures, *Meteorological Applications*, 15, 181–197.

Cloke, H., Pappenberger, F. (2009b). Ensemble flood forecasting: a review. *Journal of Hydrology*, 375 (3–4), 613–626.

Coccia, G., Todini, E. (2011). Recent developments in predictive uncertainty assessment based on the model conditional processor approach, *Hydrol. Earth Syst. Sci.*, 15, 3253-3274.

Davolio, S., Davolio, S., Miglietta, M. M., Diomede, T., Marsigli, C., Morgillo, A., Moscatello, A. (2008). A meteo-hydrological prediction system based on a multimodel approach for precipitation forecasting. *Natural Hazards Earth System Sciences*, 8, 143–159.

Day, G.N. (1985). "Extended Streamflow Forecasting Using NWSRFS*". Journal of Water Resources Planning and Management*, 111(2), 157-170.

De Roo, A., Gouweleeuw, B., Thielen, J., Bates, P., Hollingsworth, A., et al. (2003). Development of a European Flood Forecasting System, *International Journal of River Basin Management*, 1(1), 49–59.

Demargne J, Mullusky M, Werner K, Adams T, Lindsey S, Schwein N, Marosi W, Welles E. (2009). Application of forecast verification science to operational river forecasting in the US national weather service. *Bulletin of the American Meteorological Society,* 90, 779–784.

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., Ramos, M.H. (2007). Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environmental Hazards*, 7, 115-127.

Doblas-Reyes FJ, Hagedorn R, Palmer RN. (2005). The rationale behind the success of multimodel ensembles in seasonal forecasting - II. Calibration and combination. *Tellus*, 57A, 234–252.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S. (2007). Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging, *Advances in Water Resources*, 30, 1371–1386.

Duband, D. (1970) *Reconnaissance dynamique de la forme des situations météorologiques. Application à la prévision quantitative des précipitations*. Thèse de 3ème cycle, Faculté des Sciences de Paris, 400 pages.

Duband, D. (1980) Dynamic selection of analogue flow patterns to produce quantitative precipitation forecasts, *WMO symposium on probabilistic and statistical methods in weather forecasting*, Nice, France, September 1980, 487-492.

Ehrendorfer, M., Tribbia, J. J. (1997). Optimal prediction of forecast error covariances through singular vectors*. J. Atmos. Sci*., 54, 286-313.

Epstein, E. S., Murphy A. H. (1965). A note on the attributes of probabilistic prediction and the probability score. *Journal of Appl. Meteorology*, 4, 297-299.

Epstein, E. S. (1969). A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Appl. Meteorology*, 8, 985-987.

European Commision (2007). Directive 2007/60/EC on the assessment and management of flood risks (OJ L288, 6.11.2007), 27 pages.

Ferro, C. A. T. (2007). A Probability Model for Verifying Deterministic Forecasts of Extreme Events. *Weather and Forecasting*, 22, 1089-1100.

Fortin, V., Favre, A.C., Said, M. (2006). Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member, *Q. J. Roy. Meteorol. Soc*., 132, 1349–1369.

François, B., Hingray, B., Hendricks, F. Creutin, J.D. (2012). The value of storage water in reservoir systems: a climatological signature for global change impact studies. In preparation for *Water Resources Research*.

Garçon, R. (1996). Prévision opérationnelle des appons de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. *La Houille Blanche,* 5, 71-76.

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.

Garçon, R., Houdant, B., Garavaglia, F., Mathevet, T., Paquet, E., Gailhard, J. (2009), Expertise humaine des prévisions hydrométéorologiques et communication de leurs incertitudes dans un contexte décisionnel, *La Houille Blanche,* 5, 71-80.

Geleyn, J.-F., Bazile, E., Bougeault, P., Déqué, M., lvanovici, V., Joly, A., Labbé, L., Piédelièvre, A.D., Piriou, J.-M., Royer, J.-F. (1994) Atmospheric parametrization schemes in Météo-France's Arpège NWP model, *ECMWF seminar proceedings on Parametrization of Sub-grid Scale Physical Processes*, Reading, UK, September 1994, 385-402.

Glahn, H. R., Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting., *Journal of Appl. Meteorology*., 11, 1203–1211.

Glahn, B., Peroutka, M., Wiedenfeld, J., Wagner, J., Zylstra, G., Schuknecht, B., Jackson, B. (2009), MOS Uncertainty Estimates in an Ensemble Framework. *Monthly Weather Review*, 137, 246–268.

Gneiting, T., Raftery, A.E., Westveld, A.H., Goldman, T., (2005). Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, 133, 1098-1118.

Gneiting, T. ,Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal *of the American Statistical Association*, 102 (477), 359-378.

Goldstein, M., Rougier, J. (2009). Reified Bayesian modelling and inference for physical systems, *J. Stat. Plan. Infer.*, 139, 1221–1239.

Goswani, M., O'Connor, K.M., Bhattarai, K.P., Shamsedlin, A.Y. (2005). Assessing the performance of eight real-time updating models and procedures for the brosna river. *Hydrol. Earth Syst. Sci.*, 9(4), 394–411.

Goswami, M., O'Connor, K. M., Bhattarai, K. P. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment, *Journal of Hydrology*, 333 (2– 4), 517– 531.

Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80-91.

Haan, C.T. (1977). *Statistical methods in hydrology*. The Iowa State University Press/Ames, 378 pages.

Hagedorn, R., Doblas-Reyes, F.J., Palmer, T. N. (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A*, 57, 219–233.

Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., Palmer, T. N. (2012), Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Q.J.Roy. Meteorol. Soc.*, 138, 1814–1827.

Hashino, T., A. A. Bradley, Schwartz, S. S. (2006), Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol. Earth Syst. Sci. Discuss*., 3, 561–594.

Hamill, T. M. (1997). Reliability Diagrams for Multicategory Probabilistic Forecasts. *Weather and Forecasting*, 12, 736-741.

Hamill, T. M. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, 129, 550-560.

Hamill, T. M., Whitaker, J.S., Wei, X. (2004). Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review*, 132, 1434-1447.

Hamill, T.M., Hagedorn, R., Whitaker, J.S. (2008). Probabilistic Forecast Calibration Using CEPMMT and GFS Ensemble Reforecasts. Part II: Precipitation. *Monthly Weather Review,* 136(7), 2620-2632.

Hamill, T. M. (2012). Verification of TIGGE Multimodel and Ecmwf Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States. *Monthly Weather Review*, 140, 2232–2252.

He Y., Wetterhall F., Cloke H. L., Pappenberger F., Wilson M., Freer J., McGregor G. ,(2009). Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, *Meteorological Applications*, 16, 91–101.

He Y., Wetterhall F., Bao H., Cloke H. L., Li Z., Pappenberger F., Hu Y., Manful D., Huang Y, (2010). Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study, *Atmos. Sci. Let*., 11 (2), 132-138.

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15, 559-570.

Houtekamer, P.L., Lefaivre, L. (1997). Using ensemble forecasts for model validation. *Monthly Weather Review,* 125, 2416–2426.

Houdant, B. (2004). *Contribution à l'amélioration de la prévision hydrométéorologique opérationnelle pour l'usage des probabilités dans la communication entre acteurs*, Thèse de doctorat, ENGREF - EDF, 209 pages.

Jasper, K., Gurtz, J., Lang, H. (2002). Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. *Journal of Hydrology,* 267 (1–2), 40–52.

Jaun, S., Ahrens, B., Walser, A., Ewen, T., Schär, C. (2008). A probabilistic view on the August 2005 floods in the upper Rhine catchment, *Nat. Hazards Earth Syst. Sci*., 8, 281-291.

Jaun, S., Ahrens, B. (2009). Evaluation of a probabilistic hydrometeorological forecast system. *Hydrol. Earth Syst. Sci.*, 13, 1031–1043.

Johnson, C., Swinbank, R. (2009). Medium-range multimodel ensemble combination and calibra tion*, Q. J. Roy. Meteor. Soc*., 135, 777–794.

Jolliffe, I. T., Stephenson D. B., (2003). *Forecast Verification: A practitioner's guide in Atmospheric Science*, Wiley & Sons, Ltd, 273 pages.

Joly, A. (2008). De l'incertitude de prévision à la prévision de l'incertitude, *Colloque SHF*, Lyon, France, Novermbre 2008, 18-19.

Johnson, C. and Swinbank, R. (2009). Medium-range multimodel ensemble combination and calibration. Q*. J. Roy. Meteor. Soc*., 135 (640), 777-794.

Ilich, N. (2001). The benefits of replacing LP solvers in basin allocation models with a generalized nonlinear evolutionary network flow solver, *Proc. World Water and Environmental Resources Congress*, Orlando, Florida,Mai 20-24, 1-20.

Kalas, M., Ramos, M.H., Thielen, J., Babiakova, G. (2008) Evaluation of the medium-range European flood forecasts for the March-April 2006 flood in the Morava River*. Journal of Hydrology and Hydromechanics*, 56 (2), 116-132.

Kalnay, E., (2003). *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 341 pages.

Kang, T.-H., Kim, Y.-O., Hong, I.-P. (2010). Comparison of pre- and post-processors for ensemble streamfow prediction, *Atmos. Sci. Let*., 11, 153-159.

Katz, R.W., Murphy A.H., (2000). *Economic Value of Weather and Climate Forecasts*, Cambridge University Press, 222 pages.

Keller, J.H., Jones, S.C., Evans, J.L., Harr, P.A. (2011). Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition, *Geophysical Research Letters*, 38 (12), art. no. L12802.

Kearns, M., Ron, D. (1999) Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural Computation*, 11, 1427 1453.

Kharin V, Zwiers F. (2002). Climate predictions with multimodel ensembles. *J. Climate,* 15, 793–799.

Kim, Y.O., Eum, H.I., Lee, E.G., Ko, I.H., (2007). Optimizing operational policies of a korean multireservoir system using sampling stochastic dynamic programming with ensemble streamflow prediction. *J. Water Resour. Plan. Manage*., 133, 4-14.

Kitanidis, P., Bras, R. (1980). Real-time forecasting wih a conceptual hydrologic model Part I : analysis of uncertainty. *Water Resources Research*, 16(6), 1025_1033.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13-24.

Komma, J., Reszler, C., Bloschl, G., Haiden, T. (2007). Ensemble prediction of floods – catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.,* 7, 431–444.

Koutsoyiannis, D. (2010). HESS Opinions "A random walk on water", *Hydrol. Earth Syst. Sci*., 14, 585–601.

Krishnamurti T, Kishtawal C, LaRow T, Bachiiiochi DR, Williford C, Gadgil S, Surendran S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science,* 13, 4196–4216.

Krzysztofowicz, R., Kelly, K. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research,* 36(11), 3265-3278.

Krzysztofowicz, R., (2002). Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268 (1–4), 16–40.

Krzysztofowicz, R., and Maranzano, C.J., (2004). Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology*, 293 (1-4), 57-73.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm*, Journal of Hydrology*, 211, 69–85.

Labadie, J.W., (2004). Optimal Operation of Multireservoir Systems: State-of-the-Art Review *J. Water Resour. Plan. Manage*, 130 (2), 93-111.

Laio, F., Tamea S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrol. Earth Syst. Sci.,* 11**,** 1267–1277.

Le Moine, N. (2008), *Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?*, Thèse de doctorat, Université Pierre et Marie Curie - Cemagref, 348 pages.

Leutbecher, M., Parlmer, T.N. (2007). Ensemble forecasting. *Journal of Computational Physics*. 227, 3515–3539.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., Restrepo, P., (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.*, 16, 3863-3887.

Lorenz, E.N. (1969). Atmospheric predictability as revealed by naturally occurring analogues*. J.Atmos. Sci*, 26, 636–646.

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus,* 34, 505–513.

Madadgar, S., Moradkhani, H., Garen, D., (2012). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, Online version, doi: 10.1002/hyp.9562.

Marsigli, C., Montani, A., Paccagnella, T. (2008). A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorological Applications*, 15, 125–143.

Marty R., Zin I., Obled Ch. (2012). Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of PQPF: flash flood case-studies in the Cévennes-Vivarais region (Southern France). *Hydrological Processes*, 27 (1), 33–44.

Mason, I. (1982). A model for assessment of weather forecasts. *Aust. Meteor. Mag*., 30, 291–303.

Mason, S.J., Stephenson, D.B. (2008). *How do we know whether seasonal climate forecasts are any good? Seasonal Climate: Forecasting and Managing Risk*. Edition: A. Troccoli, M. Harrison, D.L.T. Anderson and S.J. Mason, Dordrecht: Springer, 259-289.

Matheson, J. E., Wikler, R.L. (1976). Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10), 1087-1096.

Mathevet, T. (2005). *Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants*, Ecole Nationale du Génie Rural, des Eaux et Forêts, Paris, 463 pages.

Mathevet, T. (2010). *Erreur empirique de modèle*. Note technique internet EDF-DTG (ref : D4165/NT/2010-00395-A), 11 pages.

McCollor, D., Stull, R., (2008). Hydrometeorological short-range ensemble forecasts in complex terrain. Part II: Economic evaluation. *Weather and Forecasting,* 23, 557– 574.

Molteni, F., Palmer, T.N. (1993). Predictability and finite time instability of the northern winter circulation. *Q. J. Roy. Meteor. Soc*., 119, 269-298.

Molteni, F., Buizza, R., Palmer, T. N., Petroliagis, T. (1996), The CEPMMT Ensemble Prediction System: Methodology and validation. Q*. J. Roy. Meteor. Soc*., 122, 73–119.

Muller, W. A. A., Doblas-Reyes, F.J., Liniger, M.A. (2004). A Debiased Ranked Probability Skill Score to Evaluate Probabilistic Ensemble Forecasts with Small Ensemble Sizes. *Journal of Climate,* 18, 1513-1523.

Muluye, G.Y., (2011). Implications of medium-range numerical weather model output in hydrologic applications: Assessment of skill and economic value. *Journal of Hydrology,* 400, 448–464

Murphy, A. H., Epstein, E.S. (1967). Verification of probability predictions: a brief review. *Journal of Appl. Meteorology*, 6, 748-755.

Murphy, A. H. (1971). A note on the Ranked Probability Score. *Journal of Appl. Meteorology,* 10, 155-156.

Murphy, A. H. (1972a). Scalar and Vector Partitions of the Probability Score : Part I. Two State Situation. *Journal of Appl. Meteorology,* 11, 273-282.

Murphy, A. H. (1972b). Scalar and Vector Partitions of the Probability Score : Part II. N-State Situation. *Journal of Appl. Meteorology,* 11, 1183-1192.

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Appl. Meteorology*, 12, 595-600.

Murphy, A. H., Daan, H. (1985). *Forecast evaluation. Probability, Statistics, and Decision Making in the Atmospheric Sciences*, Editors : Murphy et Katz, Boulder, CO, Westview Press, 379 pages.

Murphy, A. H., Winkler., R.L. (1987). A general framework for forecast verification. *Monthly Weather Review*, 115, 1330-1338.

Murphy, J., Sexton, D., Barnett, D., Jones, G.,Webb, M., Collins, M., Stainforth, D. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 20, 430, 768–772.

Nash, J. E., J.V. Sutcliffe (1970). River flow forecasting through conceptual models Part I - A discussion of principles. *Journal of Hydrology*, 27 (3), 282-290.

Nicolle, P. (2010). *Prise en compte de la neige dans la prévision hydrologique : quel apport pour mieux prévoir les crues nivales ?,* Maitrise de Master, UPMC - ENGREF, 53pages.

Nicolle, P., Valéry, A., Ramos, M.H., Perrin, C. Andréassian, V. (2012). Mieux prévoir les crues nivales : évaluation de prévisions probabilistes de débit sur des bassins versants de montagne français. *La Houille Blanche*, 2, 26-33.

Nobert, S., Demeritt, D., Cloke, H. L. (2010). Using Ensemble Predictions for operational flood forecasting: Lessons from Sweden, *J. Flood Risk Manage*., 3, 72–79.

Obled, C., Bontron, G., Garçon, R. (2002). Quantitative precipitation forecasts: a statistical adaptation of model outputs though an analogues sorting approach. *Atmospheric Research,* 63, 303-324.

Olsson, J., Lindström, G (2008). Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *Journal of Hydrology*, 350, 14–24.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil, F., Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, 303, 290-306.

Pagano, T. C., Shrestha, D. L., Wang, Q. J., Robertson, D., Hapuarachchi, P. (2012), Ensemble dressing for hydrological applications. *Hydrological Processes,* doi: 10.1002/hyp.9313.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., Tribbia, J. (1993). Ensemble Prediction. *Proc. of the CEPMMT Seminar on Validation of Models over Europe*, 1, 21-66.

Palmer, T., Barkmeijer, J., Buizza, R., Klinker, E., Richardson, D. (2002). L'avenir de la prevision d'ensemble. *La Météorologie*, 36, 22-30.

Palmer, T., Buizza, R., Hagedorn, R., Lawrence, A., Leutbecher, M., Smith, L. (2005) Ensemble prediction; a pedagogical perspective. *CEPMMT Newsletter,* 106, 10-17.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., De Roo, A.P.J., (2005). Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions with the European Flood Forecasting System (EFFS). *Hydrol. Earth Syst. Sci.*, 9, 381–393.

Pappenberger F., Bartholmes J., Thielen J., Cloke H. L., R. Buizza, de Roo A., (2008). New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophysical Research Letters*, 35, 1-7.

Pappenberger, F., Stephens, E., Thielen, J., Salamon, P., Demeritt, D., van Andel, S.J., Wetterhall, F., Alfieri, L. (2013). Visualising probabilistic flood forecast information: expert preferences and

perceptions of best practice in uncertainty communication. *Hydrological Processes*, 27(1), 132-146.

Paquet, E. (2004). Modèle MORDOR à 10 paramètres, *MOPEX Workshop n°4*. Paris.

Park, Y.-Y., Buizza, R., Leutbecher, M., (2008). TIGGE: preliminary results on comparing and combining ensembles. *Q. J. Roy. Meteorol. Soc*., 134, 2029-2050.

Parker, D., Fordham, M. (1996). Evaluation of flood forecasting, warning and response systems in the European Union, *Water Resource Management*, 10, 279–302.

Pena M, van den Dool H. (2008). Consolidation of multi model forecasts by ridge regression: application to Pacific sea surface temperature. *J. Climate,* 21, 6521–6538.

Perrin, C. (2000). *Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative*. Thèse de doctorat, INPG, Grenoble, Paris, 530 pages.

Pierce, C., Bowler, N., Seed, A., Jones, D., Moore, R. (2005). Towards stochastic fluvial flood forecasting: quantification of uncertainty in very short range QPFs and its propagation through hydrological and decision making models, *ACTIF 2nd Workshop, Quantification, Reduction and Dissemination of Uncertainty in Flood Forecasting*, Delft, 3594-3603.

Pratt, J.W. (1964). Risk aversion in the small and in the large. *Econometrica*. 32, 98-107.

Raftery, A. E., Gneiting, T., Balabdaoui, F., Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174.

Ramos, M.H., Thielen J., Pappenberger F., (2008). Utilisation de la prévision météorologique d'ensemble pour la prévision hydrologique opérationnelle et l'alerte aux crues. *Proc. Colloque SHF Prévisions hydrométéorologiques*, Lyon, 18-19 Novermber 2008, 191-199.

Ramos, M.H., Mathevet Th., Thielen J., Pappenberger F., (2010) Communicating uncertainty in hydro-meteorological forecasts: mission impossible?, *Meteorological Applications*, 17 (2), 223-265.

Ramos, M.H., van Andel, S.J., Pappenberger, F. (2012) Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, 9, 13569-13607.

Randrianasolo, A., Ramos, M.H., Thirel, G., Andreassian, V., Martin, E. (2010) Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmos. Sci. Let*., 11(2), 100-107.

Reggiani, P., Weerts, A.H. (2008). A Bayesian Approach to Decision-Making Under Uncertainty: An Application to Real-Time Forecasting in the River Rhine, *Journal of Hydrology*, 356(1-2), 56-69.

Reggiani, P., Renner, M., Weerts, A. H., van Gelder, P. A. H. J. M. (2009). Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resources Research*, 45, W02428.

Richardson, D.S., (2000). Skill and economic value of the CEPMMT Ensemble Prediction System. *Q. J. Roy. Meteorol. Soc.* 126, 649–668.

Richardson, D., (2005). The THORPEX interactive grand global ensemble (TIGGE). *First Workshop on Thorpex Interactive Grand Global Ensemble*, Reading, UK, 1-3 March 2005, 1-39.

Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments, *Water Resources Research*, 48 (5), W05520.

Rochas, M. (2000). La prévision numérique du temps. *Atmosphériques*, 21-22.

Roche, P.-A., Miquel, J., Gaume, E., (2012). *Hydrologie Quantitative : Processus, modèles et aide à la décision*. Springer, 620 pages.

Rodwell, M. (2005) Comparing and combining deterministic and ensemble forecasts: how to predict rainfall occurrence better. *CEPMMT Newsletter,* 106, 17-23.

Roulin, E., Vannitsem, S. (2005). Skill of medium-range hydrological ensemble predictions. *J. Hydrometeorology*, 6, 729–744.

Roulin, E. (2007). Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrol. Earth Syst. Sci.,* 11, 725–737.

Roulin, E., Vannitsem, S. (2012). Postprocessing of Ensemble Precipitation Predictions with Extended Logistic Regression Based on Hindcasts. *Monthly Weather Review*, 140, 874–888.

Rousset-Regimbeau, F., Habets, F., Martin, E., Noilhan, J. (2007). Ensemble streamflow forecasts over France*, CEPMMT Newsletter*, 111, 21–27.

Rousset-Regimbeau, F., Noilhan, J., Thirel, G., Martin, E., Habets, F., (2008). Mediumrange ensemble streamflow forecast over France. *Geophysical Research Abstracts*, 11, A-03111.

Roulston, M. R., Smith, L. A. (2003). Evaluating Probabilistic Forecasts Using Information Theory. *Monthly Weather Review,* 130, 1653-1660.

Quintana-Segui P., Le Moigne P., Durand Y., Martin E., Habets F., Baillon M., Canellas C., Franchisteguy L., Morel S., (2008), Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47 (1), 92-107.

Sauquet E., 2006, *Cartographie des écoulements moyens mensuels en France*, Note de synthèse Cemagref, 92 pages.

Schaake, J., Franz, K., Bradley, A., Buizza, R. (2005). The Hydrologic Ensemble Prediction EXperiment (HEPEX), *Hydrol. Earth Syst. Sci. Discuss.*, 3, 3321–3332.

Schaake, J.C., Hamill, T.H., Buizza, R., Clark, M., 2007. HEPEX – the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society,* 88(10), 1541-1547.

Seo, D.-J., Koren, V., Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydro meteorological data into operational hydrologic forecasting. *J. Hydrometeorology*, 4, 627–641.

Seo, D.-J., Herr, H. D., Schaake, J. C. (2006). A statistical postprocessor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, *Hydrol. Earth Syst. Sci. Discuss*., 3, 1987–2035.

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, C., Wergen, W. (1989). The ECMWF medium-range prediction models. Development of the numerical formulations and the impact of increased resolution. *Mel. Atmos. Phys*., 40, 28-60.

Smith, J. A., Day, G. N., Kane, M. D. (1992). Nonparametric framework for long-range streamflow forecasting, *J. Water Resour. Plan. Manage.*, 118, 82–91.

Smith, L. (2007). *Chaos : A very short introduction*. Oxford University Press, UK, 180 pages.

Stefanova, L., Krishnamurti, (2002) Interpretation of seasonal climate forecast using Brier score, the Florida State University superensemble, and the AMIP-I dataset. *J. Climate*, 15, 537– 544.

Stensrud, J. S. W., Wandishin, M.S. (2000). The Correspondence Ratio in Forecast Evaluation. *Weather and Forecasting,* 15, 593-601.

Stephenson, D. B. (2007). Two Extra Components in the Brier Score Decomposition. *Weather and Forecating,* 23(4), 752-757.

Sloughter, J. M., Raftery, A. E., Gneiting, T., Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135, 3209–3220.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychol. Bull.,* 99, 100-117.

Talagrand, O., Vautard, R., Strauss, B. (1997). Evaluation of probabilistic prediction systems. *Proceedings of workshop on predictability, European Centre for Medium-Range Weather Forecasts*, Reading, UK, 0ctober 1997, 1-25.

Tangara, M. (2005). *Amélioration d'une méthode de prévision de crue utilisant un modèle global pluie-débit*. Thèse de Doctorat, EPHE, Paris, 374pages.

Teweless, S., Wobus,B. (1954). Verfication of prognostic charts. *Bulletin American Meteorological Society*, 35, 455-463.

Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., (2009). The European flood alert system – Part 1: concept and development. *Hydrol. Earth Syst. Sci.,* 13(2), 125–140.

Thirel, G., Rousset-Regimbeau, F., Martin, E., Habets, F. (2008). On the impact of short-range meteorological forecasts for ensemble stream flow predictions, *J. Hydrometeorology*, 9(6), 1301–1317.

Thompson, J. C. (1952) . On the operational deficiencies in categorical weather forecasts. *Bulletin of Americal Meteorological Society*, 33, 223-226.

Toth, Z., Kalnay, E. (1993). Ensemble forecasting at NMC: the generation of initial perturbations. *Bulletin of Americal Meteorological Society*, 74, 2317-2330.

Toth, Z., Kalnay, E. (1997). Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125, 3297-3319.

Tsikalakis, A. G., Hatziargyriou, N. D., Katsigiannis, Y. A., Georgilakis, P. S. (2009), Impact of wind power forecasting error bias on the economic operation of autonomous power systems. *Wind Energ.*, 12, 315–331.

US Army Corps of Engineers. (1956). *Snow Hydrology*. USACE North Pacific Division, Portland,Oregon, 437 pages.

Valéry, A. (2010). *Modélisation précipitations-débit sous influence nivale, élaboration d'un module neige et évaluation sur 380 bassins versants*, thèse de doctorat, ENGREF-Cemagref, Paris, 405pages.

Van den Bergh, J., Roulin, E. (2010), Hydrological ensemble prediction and verification for the Meuse and Scheldt basins. *Atmos. Sci. Let*., 11, 64–71.

Velazquez, J. A., Petit, T., Lavoie, A., Boucher, M.-A., Turcotte, R., Fortin, V., Anctil, F. (2009). An evaluation of the Canadian global meteorological ensemble prediction system for shortterm hydrological forecasting, *Hydrol. Earth Syst. Sci*., 13, 2221–2231.

Verkade J. S., Werner, M. G. F. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *Hydrol. Earth Syst. Sci.,* 15, 3751–3765.

Verkade, J., Brown, J., Weerts, A., Regianni, P. (2012). Streamow prediction in River Rhine: Exploring combinations of bias-correcting forcing ensembles and bias-correcting streamow ensembles, *HEPEX meeting*, Beijing Normal University, October 2012 (oral presentation).

Vidal J.-P., Martin, É., Franchistéguy, L., Baillon, M., Soubeyroux, J.-M., (2009). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30(11), 1627-1644.

Von Neumann, J., Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, 625 pages.

Vrugt, J. A., Gupta, H., Bouten, W., Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, 39(8), art. No. 1201.

Vrugt, J.A., Dirks, C.G., Clark, M. (2008). Ensemble Bayesian model averaging using Markov Chain Monte Carlo Sampling. *Environmental Fluid Mechanics*, 8, 579-595.

Yamaguchi, M., T. Nakazawa, Hoshino, S. (2012). On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific, *Q.J.Roy. Meteorol. Soc.*, 138, 2019–2029.

Yang, C., Yan, Z., Shao, Y. (2012). Probabilistic precipitation forecasting based on ensemble output using generalized additive models and bayesian model averaging. *Acta Meteorologica Sinica*, 26 (1), 1-12.

Younis, J., Anquetin, S., Thielen, J. (2008a). The benefit of high-resolution operational weather forecasts for flash flood warning, *Hydrol. Earth Syst. Sci*., 12, 1039-1051.

Younis, J., Ramos, M.H., Thielen, J. (2008b) EFAS forecasts for the March-April 2006 flood in the Czech part of the Elbe River Basin - a case study. *Atmos. Sci. Let*., 9(2), 88-94.

Wang, X., Bishop, C. H. (2005). Improvement of ensemble reliability with a new dressing kernel. *Q. J. Roy. Meteorol. Soc*., 131, 965–986.

Weerts, A. H., Ghada, Y., Serafy, El., Hummel, S., Dhondia, J., Gerritsen, H. (2010). Application of generic data assimilation tools (DATools) for flood forecasting purposes. *Comput. Geosci,* 36(4), 453-463.

Weerts, A. H., Winsemius, H. C., Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), *Hydrol. Earth Syst. Sci.*, 15, 255–265.

Weigel, A. P. (2006). The discrete Brier and Ranked Probability Skill Scores. *Monthly Weather Review*, 135, 118-124.

Weijs, S., Van Nooijen, R., Van de Giesen, N. (2010). Kullback-Leibler divergence as a forecast skill score with classical reliability-resolution-uncertainty-decomposition, *Monthly Weather Review*, 138, 3387–3399.

Weijs, S. (2011). *Information Theory for Risk-based Water System Operation*, PhD Thesis published by VSSD, Delft, The Netherlands, 210 pages.

Whitaker, J.S., Loughe, A.F. (2003). The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review*, 126, 3292-3302.

Wilks, D.S., Hamill, T.M. (1995). Potential Economic Value of Ensemble-Based Surface Weather Forecasts. *Monthly Weather Review*, 123, 3565–3575.

Wilks, D. S. (2009). Extending logistic regression to provide full-probabilitydistribution MOS forecasts. *Meteorological Applications*, 16, 361–368.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences : An Introduction*, Academic Press, 676 pages.

Wilson, L. J., Burrows, W.R., Lanzinger, A. (1999). A Strategy for Verification of Weather Element Forecasts from an Ensemble Prediction System. *Monthly Weather Review*, 127, 956-970.

Wilson LJ, Valléee M. (2002). The Canadian Updateable Model Output Statistics (UMOS) system: design and development tests. *Weather and Forecasting*, 17, 206–222.

Winkler, R. L., (1972). *An introduction to Bayesian inference and decision*. Editors : Winkler Holt, Rinehart and Winston, New York, 563 pages.

Wood, A.W., Maurer, E.P., Kumar, A., Lettenmaier, D.P., (2002). Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research: Atmospheres*, 107 (D20), 4429.

Wood, A., Lettenmaier, D., (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters,* 35, L14401.

Wood, A. W., Schaake, J. C. (2008). Correcting Errors in Streamflow Forecast Ensemble Mean and Spread, *J. Hydrometeorology*, 9(1), 132–148.

Woodcock, F., Engel, C. (2005). Operational consensus forecasts. *Weather and Forecasting*, 20, 101–111.

Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research*, 8, 135-141.

Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J., Wunram, C., (2008). MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Let*., 9, 80–87.

Zhao, L., Duan, Q., Schaake, J., Ye, A., and Xia, J. (2011). A hydrologic post-processor for ensemble streamflow predictions*, Adv. Geosci*., 29, 51-59.

Zhi, X., Qi, H., Bai, Y., Lin, C. (2012). A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data, *Acta Meteorologica Sinica*, 26 (1), pp. 41-51.

Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability, *Advances in Atmospheric Sciences*, 22(6), 781-788.

Zorita, E., von Storch H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods*. J. Climate*, 12, 2474-2489.

Zupanski, D., Zupanski, M. (2006). Model error estimation employing an ensemble data assimilation approach, *Monthly Weather Review*, 134, 3597, 3599.

Liste des illustrations

Figure 1 : Schéma général d'une chaîne de prévision hydrologique. 18

Figure 2: Schéma des prévisions de l'évolution de l'atmosphère pour le cas d'une prévision déterministe et des prévisions d'ensemble et des échéances allant jusque J+4. 22

Figure 3: Aspects clés de la chaîne de prévision hydrologique d'ensemble 26

Figure 4 : Carte des régions où : (a) le modèle GRP est utilisé en opérationnel/ pré-opérationnel ou des tests préliminaires sont effectués par les services de prévision de crue SPC (source : SCHAPI) et (b) le modèle MORDOR est utilisé en opérationnel (source : EDF-DTG). 35

Figure 5 : Cartes de France illustrant les précipitations prévues par les 51 membres de la prévision d'ensemble du centre météorologique européen (CEPMMT) pour le 30 mai 2008 et un horizon de prévision de 7 jours (prévisions émises le 23 mai 2008)...... 37

Figure 6 : Centres météorologiques contribuant à la base de prévisions de précipitation TIGGE utilisée dans cette étude, avec indication de la date de début de participation de chaque centre...... 38

Figure 7 : Description schématique de l'approche de prévisions par analogie (source : Bontron, 2004). 41

Figure 8 : Localisation du jeu de 74 larges bassins versants français...... 43

Figure 9 : Localisation du jeu de 11 bassins versants de fort intérêt en hydroélectricité de cette étude. 45

Figure 10 : Schématique de la structure du modèle GRP sans module neige (Tangara, 2005)..... 49

Figure 11 : Schématique de l'intégration du module neige Cemaneige dans la structure du modèle GRP (Valéry, 2010 ; Nicolle et al., 2012)...... 50

Figure 12: Schématique du fonctionnement du modèle hydrologique MORDOR (source : Paquet, 2004)...... 51

Figure 13 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) le module Cemaneige selon le critère de NASH et pour l'échantillon de : (a) 74 et (b) 11 bassins versants.. 55

Figure 14 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) selon le critère de C2MP et pour l'échantillon de : (a) 74 et (b) 11 bassins versants. 55

Figure 15 : Performance au calage du modèle MORDOR sur un échantillon de 11 bassins versants selon le critère de: (a) Nash-Sutcliffe (NASH), (b) Kling-Gupta (KGE)...... 56

Figure 16 : Représentation schématique de la justesse d'une prévision. 64

Figure 17 : Illustration de : (a) la justesse et (b) la précision d'une prévision pour attendre sa cible, l'observation. 66

Figure 18 : Interprétation du critère '' PIT '', a) prévision fiable, b) prévision sous-dispersée, c) prévision sur-dispersée, d) prévision biaisée. Diagramme de PIT en haut et histogramme de '' PIT '' en bas. 70

Figure 19 : Tableau de contingence pour une prévision. 72

Figure 20 : Schéma représentant la construction d'une courbe ROC (H=a, M=c, FA=b, CR=d du Tableau de Contingence). 73

Figure 21 : La chaîne de la prévision du point de vue de l'utilisateur à l'arrivée d'un événement (d'après Parker et Fordham, 1996). 98

Figure 22: Illustration d'un arbre de décision. 102

Figure 23: Modèle Coût-Pertes de décision couplé à un tableau de contingence des événements prévus/observés. 104

Figure 24: Courbes de valeur économique relative pour une prévision probabiliste (tiré de http://www.cawcr.gov.au/projects/verification/) 106

Figure 25 : Prévisions de précipitation valables pour le 25 mai 2008 sur la France des 8 centres météorologiques de la base TIGGE de cette étude et à 3 jours d'échéance : répresentation de 4 membres de chaque prévision d'ensemble. 117

Figure 26 : Prévisions de précipitation valables pour le 25 mai 2008 sur la France des 8 centres météorologiques de la base TIGGE de cette étude et à 1 jour d'échéance : répresentation de 4 membres de chaque prévision d'ensemble. 118

Figure 27 : Boîtes à moustaches pour les prévisions d'ensemble de précipitation émises le 23 mai 2008 sur le bassin versant de la Drôme à Saillant pour les 5 jours à venir et selon les 8 centres météorologiques de la base TIGGE de cette étude. Les précipitations observées sont indiquées à l'aide du symbole "x". 119

Figure 28 : Histogrammes de PIT pour les prévisions de pluie sur le bassin de la Drôme à Saillant à un horizon de 7 jours des centres météorologiques TIGGE : (1) ECMF, (2) EGRR, (3) RJTD), (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée. 125

Figure 29 : Histogrammes de PIT pour les prévisions de pluie sur le bassin de l'Oise à Sempigny à un horizon de 7 jours des centres météorologiques TIGGE : (1) ECMF, (2) EGRR, (3) RJTD), (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée. 126

Figure 30 : Les courbes ROC pour le bassin de la Drôme à Saillant et pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – (en parenthèse). La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée. 127

Figure 31 : Les courbes ROC pour le bassin de l'Oise à Sempigny et pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – (en parenthèse). La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée. 128

Figure 32 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion de l'ensemble pour le bassin de la Drôme à Saillant pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des diagrammes est faite sur la période d'évaluation commune disponible (242 jours) pour chaque centre. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée. 129

Figure 33 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion de l'ensemble pour le bassin de l'Oise à Sempigny pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. La construction des diagrammes est faite sur la période d'évaluation commune disponible (242 jours) pour chaque centre. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée. 130

Figure 34 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision pour le bassin de la Drôme à Saillant. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF et (2) SBSJ. La construction des boîtes à moustaches est faite sur la période d'évaluation commune disponible (242 jours). 131

Figure 35 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision pour le bassin de l'Oise à Sempigny. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF et (8) SBSJ. La construction des boîtes à moustaches est faite sur la période d'évaluation commune disponible (242 jours). 131

Figure 36 : Boîtes à moustache pour le score de probabilité ordonnée RPS sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les boîtes à moustache sont calculées à partir des prévisions d'ensemble de pluie tirées des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. 133

Figure 37 : Boîtes à moustache pour le score de divergence sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions d'ensemble de pluie sont extraits des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. 135

Figure 38 : Graphique du score de probabilité ordonnée RPS en fonction de la surface du bassin versant pour les 74 bassins versants étudiés et pour les échéances 3 (a) et 7 (b) jours. Les prévisions d'ensemble de pluie sont extraites des centres : ECMF (vert clair), EGRR (vert foncé), RJTD (orange), KWBC (rouge), BABJ (marron), AMMC (mauve), CWAO (ciel) et SBSJ (bleu). 136

Figure 39 : Carte de localisation des centres qui donnent la meilleure performance selon le critère RPS pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d). 140

Figure 40 : Carte de localisation des centres qui donnent la meilleure performance selon le critère BSQ90 pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d)...... 142

Figure 41 : Prévisions d'ensemble de débits émises le 28 mai 2008 par le modèle GRP sur le bassin versant de la Drôme à Saillant pour les 5 jours à venir et selon les 8 centres météorologiques de la base TIGGE de cette étude (intervalle inter-quantile Q20%-Q80%). En trait plein, les débits observés...... 144

Figure 42 : Histogrammes de PIT pour le bassin de la Drôme à Saillant à partir des prévisions de débit à un horizon de 7 jours. Notons que les prévisions de pluie sont issues des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée...... 148

Figure 43 : Histogrammes de PIT pour le bassin de l'Oise à Sempigny à partir des prévisions de débit à un horizon de 7 jours. Notons que les prévisions de pluie sont issues des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempigny sur la carte de France est également présentée...... 149

Figure 44 : Les courbes ROC pour les dé bits du bassin de la Drôme à Saillant pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée. 150

Figure 45 : Les courbes ROC pour les débits du bassin de l'Oise à Sempigny pour un seuil de non-dépassement égal au quantile 70% et pour l'échéance 7 jours. Les prévisions d'ensemble de pluie sont celles des centres : (1) ECMF (730 jours), (2) EGRR (730 jours), (3) RJTD (730 jours), (4) KWBC (561 jours), (5) BABJ (504 jours), (6) AMMC (393 jours), (7) CWAO (363 jours) et (8) SBSJ (242 jours). La construction des histogrammes est faite sur la période entière d'évaluation disponible pour chaque centre – en parenthèse. La localisation du bassin versant de l'Oise à Sempigny sur la carte de France est également présentée. 151

Figure 46 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion des prévisions d'ensemble de débit pour le bassin de la Drôme à Saillant pour l'échéance 7 jours. Les prévisions de pluies sont issues des centres : (1) européen ECMF, (2) britannique EGRR, (3) japonais RJTD, (4) américain KWBC, (5) chinois BABJ, (6) australien AMMC, (7) canadien CWAO et (8) brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. La localisation du bassin versant de la Drôme à Saillant sur la carte de France est également présentée. 152

Figure 47 : La racine carrée de l'erreur quadratique moyenne en fonction de la dispersion des prévisions d'ensemble de débit pour le bassin de l'Oise à Sempigny pour l'échéance 7 jours. Les prévisions de pluie sont issues des centres : (1) européen ECMF, (2) britannique EGRR, (3) japonais RJTD, (4) américain KWBC, (5) chinois BABJ, (6) australien AMMC, (7) canadien CWAO et (8) brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. La localisation du bassin versant de l'Oise à Sempignysur la carte de France est également présentée. 153

Figure 48 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision sous forme de boîtes à moustache pour le bassin de la Drôme à Saillant (1) et de l'Oise à Sempigny (2). Les prévisions de pluie sont issues du centre européen ECMF. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. 154

Figure 49 : L'évolution de la dispersion de l'ensemble en fonction de l'horizon de prévision sous forme de boîtes à moustache pour le bassin de la Drôme à Saillant (1) et de l'Oise à Sempigny(2). Les prévisions d'ensemble de pluie sont issues du centre brésilien SBSJ. La construction des histogrammes est faite sur la période d'évaluation disponible (242 jours) pour chaque centre. 154

Figure 50 : Boîtes à moustache pour le score de probabilité ordonnée RPS sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions de pluie sont issues des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ.... 157

Figure 51 : Boîtes à moustache pour le score de divergence sur les 74 bassins versants étudiés et pour les échéances 3(a), 7(b) et 9(c) jours. Les prévisions de pluie sont issues des centres : (1) ECMF, (2) EGRR, (3) RJTD, (4) KWBC, (5) BABJ, (6) AMMC, (7) CWAO et (8) SBSJ. 158

Figure 52 : Graphique du score de probabilité ordonnée RPS, des prévisions d'ensemble de débit en fonction de la surface du bassin versant pour les 74 bassins versants étudiés et pour les échéances 3 (a) et 7 (b) jours. Les prévisions de pluie sont issues des centres : ECMF (vert clair), EGRR (vert foncé), RJTD (orange), KWBC (rouge), BABJ (marron), AMMC (mauve), CWAO(ciel) et SBSJ (bleu). 159

Figure 53 : Carte de localisation des centres qui donnent la meilleure performance selon le critère RPS pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d). 161

Figure 54 : Carte de localisation des centres qui donnent la meilleure performance selon le score de Brier pour le non-dépassement du quantile 90% pour les échéances 3 jours (a), 5 jours (b), 7 jours (c) et 9 jours (d). 163

Figure 55 : Histogrammes de PIT pour les prévisions de pluie pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées des centres européen et brésilien évaluées sur les 242 jours sont présentées dans la figure (d) & (e). 170

Figure 56 : Histogrammes de PIT pour les prévisions de pluie pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées des centres européen et brésilien évaluées sur les 242 jours sont présentées dans la figure (d) & (e). 170

Figure 57 : Courbes ROC pour les prévisions de pluie pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées du grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres selon le score RPS, (c) la pondération en fonction de leur performance en fonction du RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien (d) & (e). 171

Figure 58 : Courbes ROC pour les prévisions de pluie pour le bassin de l'Oise à Sempigny à un horizon de 7 jours. 171

Figure 59 : Boîtes à moustaches représentant l'évolution de la dispersion des membres de l'ensemble de pluie en fonction de l'échéance de prévision pour le bassin de la Drôme à Saillant. Les prévisions sont tirées : (a) du grand ensemble utilisant une pondération de la moyenne d'ensembles des différents centres en fonction du score RPS et (b) du centre européen. 172

Figure 60 : Boîtes à moustaches représentant la performance des prévisions de pluie en termes de scores de probabilité ordonnée RPS pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours tirées sur les prévisions du : (1) centre européen, (2) grand ensemble basé sur l'équiprobabilité de membres, (3) le grand ensemble construit en pondérant la moyenne d'ensemble des différents centres en fonction du score RPS, (4) le grand ensemble en pondérant les membres en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie sont évaluées sur 242 jours. 173

Figure 61 : Boîtes à moustaches représentant la performance des prévisions de pluie en termes de scores de divergence pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours tirées sur les prévisions du : (1) centre européen, (2) grand ensemble basé sur l'équiprobabilité de membres, (3) le grand ensemble construit en pondérant la moyenne d'ensembles des différents centres en fonction du score RPS, (4) le grand ensemble en pondérant les membres en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie sont évaluées sur 242 jours. 174

Figure 62 : Histogrammes de PIT pour les prévisions de débit pour le bassin de la Drôme à Saillant à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le du RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien évaluées sur 242 jours pour tous les centres sont présentées dans la figure (d) & (e). 176

Figure 63 : Histogrammes de PIT pour les prévisions de débit pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées d'un grand ensemble basé sur : (a) l'équiprobabilité des membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de pluie tirées du centre européen et brésilien évaluées sur 242 jours pour tous les centres sont présentées dans la figure (d) & (e). 177

Figure 64 : Boîtes à moustaches représentant l'évolution de la dispersion des membres de l'ensemble de débit en fonction de l'échéance de prévision pour le bassin de la Drôme à Saillant. Les prévisions sont tirées : (a) du grand ensemble utilisant une pondération en fonction du nombre de membres et (b) du centre européen. 177

Figure 65 : Courbes ROC pour les prévisions de débit pour le bassin de la Drôme à Saillant à un horizon de 7 jours. 178

Figure 66 : Courbes ROC pour les prévisions de débit pour le bassin de l'Oise à Sempigny à un horizon de 7 jours tirées du grand ensemble basé sur :(a) l'équiprobabilité de membres, (b) la pondération de la moyenne d'ensemble des différents centres en fonction du score RPS, (c) la pondération en fonction de leur performance selon le RPS. Les prévisions d'ensemble de débit tirées du centre européen et brésilien (d) & (e). 178

Figure 67 : Boîtes à moustaches représentant la performance des prévisions de débit en termes de scores de probabilité ordonnée RPS pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours pour les prévisions tirées du : (1) centre européen, (2) grand ensemble basé sur l'equifinalité de membres, (3) grand ensemble construit en pondérant la moyenne d'ensemble des différents centres en fonction du score RPS, (4) grand ensemble en pondérant en fonction de leur performance selon le RPS. 180

Figure 68 : Boîtes à moustaches représentant la performance des prévisions de débit en termes de scores de divergence pour l'échantillon de 74 bassins versants sur la période entière d'évaluation et à un horizon de 3 (a) et 7 (b) jours. 180

Figure 69 : Localisation des trois bassins versants utilisés en exemple. 188

Figure 70 : Histogrammes de PIT – avec le diagramme de probabilités cumulées en haut gauche - pour les prévisions d'ensemble de pluies du CEPMMT brutes pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite). 193

Figure 71 : Evolution de la dispersion des membres de la prévision d'ensemble du CEPMMT brutes des pluies en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 194

Figure 72 : Relation de la racine carrée de l'erreur quadratique moyen (RMSE) et la dispersion de la prévision d'ensemble des pluies du CEPMMT brutes pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours. 195

Figure 73 : Courbes ROC de la prévision d'ensemble des pluies de bassin brutes pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). 196

Figure 74 : CRPS moyens de la prévision d'ensemble des précipitations de bassin brutes en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. 197

Figure 75 : Histogrammes et diagrammes de PIT pour les prévisions d'ensemble de débit brutes du modèle MORDOR pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite). Prévisions d'ensemble de pluie du CEPMMT utilisées en entrée du modèle hydrologique. 201

Figure 76 : Histogrammes et diagrammes de PIT pour les prévisions d'ensemble de débit brutes du modèle GRP pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 jours (à gauche) et 7 jours (à droite). Prévisions d'ensemble de pluie du CEPMMT utilisées en entrée du modèle hydrologique. 202

Figure 77 : Evolution de la dispersion des membres de la prévision hydrologique d'ensemble fournies par les modèles MORDOR (1) et GRP (2) de bassin brutes en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noire dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 203

Figure 78 : La racine carrée de l'erreur quadratique moyen (RMSE) en fonction de la dispersion de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours. 204

Figure 79 : Erreur quadratique moyen (RMSE) en fonction de la dispersion de la prévision d'ensemble des débits brutes fournies par le modèle GRP pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances de 3 (à gauche) et 7 (à droite) jours. 205

Figure 80 : Courbes ROC de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). 206

Figure 81 : Courbes ROC de la prévision d'ensemble des débits brutes fournies par le modèle GRP pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). 207

Figure 82 : CRPS moyens de la prévision d'ensemble des débits brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. 208

Figure 83 : CRPS moyens de la prévision d'ensemble des débits brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. 209

Figure 84 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de pluie pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiques sont calculés à partir des pluies du centre européen CEPMMT brutes (1) et corrigées en utilisant un facteur multiplicatif (2A) et la méthode des analogues (2B). 212

Figure 85 : Evolution de la dispersion des membres de la prévision d'ensemble de précipitations : brute (1), biaisée avec la correction multiplicative (2A) et corrigée avec l'information des analogues (2B) en fonction de l'échéance de prévision pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 213

Figure 86 : Courbes ROC de la prévision d'ensemble des précipitations pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). Les courbes ROC sont calculées à partir des prévisions de pluies du centre européen brutes (1), corrigées en utilisant un facteur multiplicatif (2A) et l'information des analogues (2B). 214

Figure 87 : CRPS moyens de la prévision d'ensemble des précipitations de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions de pluies du centre européen brutes (1) et corrigées en utilisant un facteur multiplicatif (2A) et l'information des analogues (2B). 215

Figure 88 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de débit pour : (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiques sont calculés à partir des pluies du centre européen CEPMMT brutes (1) et corrigées en utilisant la méthode des analogues (2B). 217

Figure 89 : Evolution de la dispersion des membres de la prévision d'ensemble de débits : brutes (1) et à partir des prévisions de pluie corrigées avec l'information des analogues (2B) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 218

Figure 90 : Courbes ROC de la prévision hydrologiques d'ensemble pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances 3 jours (à gauche) et 7 jours (à droite). Les courbes ROC sont calculées à partir des prévisions de pluies du centre européen brutes (1) et l'information des analogues (2B). 219

Figure 91 : CRPS moyens de la prévision d'ensemble des débits fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculés à partir des prévisions de pluies du centre européen brutes (1) et corrigées en utilisant l'information des analogues (2B). 220

Figure 92 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS selon l'habillage avec l'information des analogues. Les graphiques sont calculés à partir des prévisions hydrologiques produites par le modèle MORDOR et à une échéance de prévisions égale à (a) 3 et (b) 7 jours. 221

Figure 93 : Diagrammes de probabilité cumulée pour les prévisions d'ensemble de débits pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les graphiquess sont calculés à partir des prévisions hydrologiques habillées avec l'erreur du modèle hydrologique (3) : MORDOR (1) et GRP (2). .. 225

Figure 94 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec l'erreur du modèle hydrologique (3) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 226

Figure 95 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec l'erreur du modèle hydrologique (3) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 227

Figure 96 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle (3). 228

Figure 97 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle hydrologique (3). 230

Figure 98 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions brutes (1) et corrigées avec l'erreur du modèle hydrologique (3). 230

Figure 99 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques du modèle MORDOR par bassin versant en termes de score de RPS par l'habillage avec l'erreur du modèle. Les graphiques sont calculés à partir des prévisions hydrologiques à une échéance de prévisions égale à : 3 (à gauche) et 7 (à droite) jours. 231

Figure 100 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques du modèle GRP par bassin versant en termes de score de RPS par l'habillage avec l'erreur du modèle. Les graphiques sont calculés à partir des prévisions hydrologiques à une échéance de prévisions égale à : 3 (à gauche) et 7 (à droite) jours. 231

Figure 101 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour les échéances : (1) 3 et (2) 7 jours. Les histogrammes sont calculés à partir des prévisions hydrologiques habillées avec l'erreur du modèle hydrologique MORDOR – sans et avec la correction à l'aide des analogues. 233

Figure 102 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), des prévisions corrigées avec l'aide des analogues (2), corrigées avec l'erreur du modèle (3) et l'approche combinée (4). 234

Figure 103 : Boîtes à moustaches du score de performance RPSS (avec en référence les prévisions brutes) pour les prévisions hydrologiques d'ensemble et les échéances 3 (à gauche), 5 (centre) et 7 jours (à droite). Les approches de traitement statistique utilisées sont les : (1) prévisions brutes, (2) l'habillage avec l'information des analogues, (3) l'habillage avec l'erreur du modèle et (4) une combinaison des traitements (2) et (3). Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 235

Figure 104 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre et sa modification par Wang et Bishop (2005) et extraites par le modèle hydrologique : (1) MORDOR et (2) GRP. 239

Figure 105 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 240

Figure 106 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 241

Figure 107 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6). 242

Figure 108 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6). 244

Figure 109 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre (5) et sa modification par Wang et Bishop (6). 244

Figure 110 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7), extraites par le modèle hydrologique : (1) MORDOR et (2) GRP. 247

Figure 111 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 248

Figure 112 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 249

Figure 113 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 3 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (2005) et la modification que nous proposons et extraites par le modèle hydrologique : (1) MORDOR et (2) GRP. 250

Figure 114 : Courbes ROC de la prévision hydrologiques d'ensemble fournies par le modèle : MORDOR (1) et GRP (2) pour un seuil de dépassement égal au quantile 70% des pluies observées pour le bassin (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance de 7 jours. Les courbes ROC sont calculées à partir des prévisions brutes (1), habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et la modification avec la covariance des échéances (7). 251

Figure 115 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et modifiée avec covariance des échéances (7). 253

Figure 116 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec la méthode du meilleur membre modifiée (6) et modifiée avec covariance des échéances (7). 253

Figure 117 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS la méthode du meilleur membre modifiée avec covariance des échéances. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours. . 254

Figure 118 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec l'erreur de 30% de meilleurs membres et fournies par le modèle hydrologique : (1) MORDOR et (2) GRP. 257

Figure 119 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et avec l'erreur de 30% de meilleurs membres (8) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 258

Figure 120 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et avec l'erreur de 30% de meilleurs membres (8) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 259

Figure 121 : Les courbes ROC pour un seuil de dépassement égal au quantile de 70% pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les courbes ROC sont calculées à partir des prévisions hydrologiques habillées avec l'erreur de tous les membres produites par le modèle hydrologique : (1) MORDOR et (2) GRP. 260

Figure 122 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et habillées avec l'erreur de 30% des meilleurs membres (8). 262

Figure 123 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée (6) et habillées avec l'erreur de 30% des meilleurs membres (8). 262

Figure 124 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS par la méthode prenant en compte l'erreur de 30% des meilleurs membres. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours. 263

Figure 125 : Diagrammes de probabilité cumulée pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les graphiques sont calculés à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifée par Fortin et al. (2006) et fournies par le modèle hydrologique : (1) MORDOR et (2) GRP. 266

Figure 126 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle MORDOR : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et Fortin et al. (9) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 267

Figure 127 : Evolution de la dispersion des membres de la prévision hydrologiques d'ensemble du modèle GRP : brutes (1) et habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9) en fonction de l'échéance de prévision pour (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons. Boîtes à moustaches : quantiles 25% et 75%, trait inférieur et supérieur de la boîte, respectivement, quantile 50%, trait horizontal noir dans la boîte, valeurs 5% et 95%, extrémités des moustaches. 268

Figure 128 : Les courbes ROC pour un seuil de dépassement égal au quantile de 70% pour le bassin de (a) la Durance à Serre Ponçon, (b) le Verdon à Castillon et (c) la Romanche à Chambons et pour l'échéance 7 jours. Les courbes ROC sont calculées à partir des prévisions hydrologiques habillées avec la méthode du meilleur membre modifiée par Fortin et al. (2006) produites par le modèle hydrologique : (1) MORDOR et (2) GRP. 269

Figure 129 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle MORDOR en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle MORDOR habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9). 270

Figure 130 : CRPS moyens de la prévision d'ensemble des débits de bassin brutes fournies par le modèle GRP en fonction de la surface du bassin pour les 11 bassins versants de l'étude et pour les échéances (a) 3 jours et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec la méthode du meilleur membre modifiée par Wang et Bishop (6) et par Fortin et al. (9). 270

Figure 131 : Carte avec l'amélioration apportée à la qualité de prévisions hydrologiques par bassin versant en termes de score de RPS par la méthode du meilleur membre modifiée par Fortin et al. (2006). Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle : (a) MORDOR et (b) GRP et à une échéance de prévisions égale à (1) 3 et (2) 7 jours. . 271

Figure 132 : Carte avec la méthode correspondant à la meilleur performance par bassin versant selon le critère RPS et pour une échéance de prévisions égale à (a) 3, (b) 5 et (c) 7 jours. Ici les prévisions hydrologiques sont fournies par les modèles MORDOR (1) et GRP (2). 273

Figure 133 : Valeurs moyennes de CRPS en fonction de la surface du bassin pour l'ensemble de 11 bassins versants testés et pour les échéances : (a) 3 et (b) 7 jours. Les graphiques sont calculées à partir des prévisions hydrologiques produites par le modèle GRP habillées avec l'erreur du modèle hydrologique avec un calage de longueur de 35 (cercle) et de 4 ans (triangle). 275

Figure 134 : Boîtes à moustache représentant l'incertitude de l'estimation de la moyenne des erreurs du meilleur membre pour les échéances : (a) 3 et (b) 7 jours. Les boîtes à moustaches sont calculées à partir des prévisions hydrologiques d'ensemble produite par le modèle GRP. 277

Figure 135 : Boîtes à moustache représentant l'incertitude de l'estimation de l'écart-type des erreurs du meilleur membre pour les échéances : (a) 3 et (b) 7 jours. Les boîtes à moustaches sont calculées à partir des prévisions hydrologiques d'ensemble produite par le modèle GRP. 277

Figure 136: Schématique simplifié d'une réserve hydroélectrique. 284

Figure 137: Classification générale des aménagements hydroélectriques en fonction du coefficient de turbinage et de la capacité utile (source des données : EDF). 287

Figure 138: Schématique de la procédure développée pour l'optimisation heuristique d'une réserve hydroélectrique pour un jour J. 289

Figure 139 : Prix horaire de l'énergie EPEX-SPOT pour les années 2003 (à gauche) et 2007 (à droite). Prix en €/MWh. 293

Figure 140 : Prix horaire de l'énergie EPEX-SPOT au long de la semaine du 8 au 14 septembre 2008. Prix en €/MWh. 294

Figure 141 : Prix horaire de l'énergie EPEX-SPOT au long de la journée du 8 septembre 2008 (lundi). Prix en €/MWh. 294

Figure 142 : Illustration des cas où les contraintes de gestion ne sont pas respectées et il faut faire une rectification de la règle de gestion : quand le stock dépasse la capacité maximale du réservoir (à gauche) et quand le stock devient inférieur au volume minimum du réservoir (à droite). 297

Figure 143 : Carte de prévisions sur la France pour la période de 1 à 9 septembre 2008. Ici les prévisions affichées sont extraites de 4 membres de l'ensemble à un horizon de 2 jours. 303

Figure 144 : (a) Boîtes à moustache pour les prévisions d'ensemble de précipitation pour la période de 1 à 9 septembre 2008, (b) l'hydrogramme de débits modélisés par le modèle MORDOR pour la période de 1 à 9 septembre 2008. 304

Figure 145 : Prix horaires extraits du marché européen d'énergie EPEX SPOT pour la même période. 305

Figure 146: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 103,47 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. Aménagement hydroélectrique de coefficient de turbinage égal à 3 fois le débit moyen de la Durance à Serre Ponçon et coefficient de capacité égal à 4 jours de stockage du débit moyen. 306

Figure 147: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 83,09 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. Aménagement hydroélectrique de coefficient de turbinage égal à 2 fois le débit moyen de la Durance à Serre Ponçon et coefficient de capacité égal à 3 jours de stockage du débit moyen. 307

Figure 148: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 94,19 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. 308

Figure 149: Résultat de la gestion au jour 4 septembre 2008. En haut : Programme de gestion de l'usine sur une semaine (prix-seuil = 98,2 et 96,02 €/MWh). En bas : Niveau du stock de la retenue sur la semaine. 308

Figure 150: Hydrogramme des débits observés à la station de la Durance à Serre Ponçon sur la période d'évaluation (01 /01/2005 au 23/12/2008). 310

Figure 151: Volumes déversés d'après le modèle de gestion pour un aménagement fictif de coefficients de turbinage et de capacité égaux respectivement à 3 et 4, dans la Durance à Serre Ponçon sur la période d'évaluation (01 /01/2005 au 23/12/2008). 310

Figure 152: Gain de gestion en MEuros pour les prévisions parfaites : chaque ligne représente un différent coefficient de turbinage. 311

Figure 153: Evolution du (a) volume turbiné et (b) volume déversé pour les prévisions d'ensemble brutes (tous les membres sont considérés) en fonction du coefficient de capacité. Chaque ligne représente un différent coefficient de turbinage. 312

Figure 154 : Le gain de gestion en fonction du coefficient de capacité pour des coefficients : (a) 3, (b) 4 et (5) de turbinage et pour des prévisions : d'ensemble brutes (PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD). 314

Figure 155 : Le volume turbiné en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes (PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD). 316

Figure 156: Le volume déversé en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes(PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD). 316

Figure 157: La différence du stock en fonction du coefficient de capacité pour un coefficient de turbinage égal à 3 et pour des prévisions : d'ensemble brutes(PE), climatologiques (PC), parfaites (PP), la moyenne d'ensemble (PME) et la deterministe (PD). 316

Figure 158: Le gain de gestion en MEuros pour différents coefficients de capacité (boîtes à moustaches) en fonction du coefficient de turbinage et par type de prévision étudié : (1) brutes, (2) habillage avec les Analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les Analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) habillées avec l'erreur de 30% de meilleurs membres et (9) habilées avec la BMM(2006). 318

Figure 159: Gain de gestion de turbinage pour les coefficients de turbinage 3(a), 4(b) et 5(c) par type de prévisions : (1) brutes, (2) habillage avec les analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) habillées avec l'erreur de 30% de meilleurs membres et (9) habillées avec la BMM (2006). 320

Figure 160: Boîtes à moustaches du score de gain par méthode de post-traitement appliquée. Le graphique présenté correspond à des coefficients de capacité et de turbinage égal à 4 et 3 respectivement. 322

Figure 161: Score de gain en fonction de la localisation du bassin versant par type de prévisions : (1) brutes, (2) habillage avec les analogues, (3) habillées avec l'erreur du modèle, (4) habillées avec les analogues et l'erreur du modèle, (5) corrigées avec la BMM (2003), (6) avec la BMM (2005), (7) avec la BMM (2005) modifiée avec la covariance, (8) avec l'erreur de 30% des meilleurs membres et (9) avec la BMM (2006). Le graphique présenté correspond à des coefficients de capacité et de turbinage égaux à 4 et 3 respectivement. 322

Figure 162: Lien entre qualité et utilité des prévisions : carte avec les deux meilleures méthodes de traitement statistique en termes de qualité globale de prévision selon le critère RPS (en haut) et de meilleur score de gain économique pour les coefficients de capacité et de turbinage égaux à 4 et 3, respectivement (en bas). Prévisions hydrologiques du modèle MORDOR à l'échéance 6 jours (pour les méthodes de traitement statistique, voir Tableau 8). 334

Tableau 1 : Caractéristiques principales des centres météorologiques qui forment la base de prévisions de précipitation TIGGE utilisée dans cette étude (d'après : http://tigge.CEPMMT.int/models.html). 39

Tableau 2 : Principales caractéristiques des 74 bassins versants étudiés en termes de valeurs moyennes (entre les crochets sont affichées les valeurs minimales et maximales). 44

Tableau 3 : Principales caractéristiques des 11 bassins versants étudiés en termes de valeurs moyennes (entre crochets sont affichées les valeurs minimales et maximales). 46

Tableau 4 : Principales caractéristiques des retenues liées aux 11 bassins versants de fort intérêt en hydroélectricité de cette étude..... 47

Tableau 5 : Périodes de calage et de validation pour chaque approche de traitement statistique testée. 93

Tableau 6 : Pourcentage de bassins versants pour lesquels chaque centre démontre les meilleures (1ère et 2ème) performances selon le critère RPS par échéance de prévision. 164

Tableau 7 : Pourcentage de bassins versants pour lesquels chaque centre demontre les meilleures (1ère et 2ème) performances selon le critère BSQ 90 par échéance de prévision. 164

Tableau 8 : Liste de scénarios de traitement statistique évalués. 189

Tableau 9 : Les sous-périodes de calage en utilisant le concept de fenêtres glissantes : leur longueur en mois (en haut) et le nombre de périodes dans chaque sous-échantillon (en bas). 276

Tableau 10 : Illustration comparative entre les différentes approches de traitement statistique utilisées en marquant les attributs de la qualité des prévisions hydrologiques que chaque approche améliore. 280

Annexes

ANNEXE A : Les modèles utilisés

A-1 : Le modèle GRP

"Le modèle GRP (modèle du Génie Rural pour la Prévision) est un modèle de prévision des crues conçu pour des utilisateurs souhaitant disposer de prévisions en temps réel sur des bassins jaugés. C'est un modèle :

- ***hydrologique** : il se sert des données de pluie disponibles sur le bassin versant pour calculer les débits à l'exutoire ;*
- ***continu** : il comporte une représentation du taux de saturation des sols pour tenir compte des conditions pluviométriques passées ;*
- ***global** : il représente le bassin versant de manière simplifiée, limitant ainsi les besoins en données ;*
- *fonctionnant au **pas de temps horaire** pour s'adapter à des dynamiques de crue variées."*

Le modèle utilisé dans cette thèse est une adaptation du modèle *GRP*, décrit par Berthet (2010), à la prévision d'ensemble au pas de temps journalier (Ramos, 2010). La Figure 10 représente schématiquement le modèle. Ci- après nous décrivons les principales fonctions du modèle utilisé.

Fonction de production

Cette fonction a pour objectif de calculer la pluie efficace. Elle se présente comme un module de suivi de l'humidité du sol (communément connu sous le sigle SMA, *"soil moisture accounting"* en anglais).

Interception

Dans la suite de cette description, les pluies en entrée du modèle sont symbolisées par P_k. La pluie nette P_n et l'évapotranspiration nette E_n sont calculées à l'aide des équations :

$$P_n = P_k - E \ et \ E_n = 0 \ \ si \ \ P_k \geq E \qquad \text{Eq. A- 1}$$

$$E_n = E - P_k \ et \ P_n = 0 \ \ si \ \ E \geq P_k \qquad \text{Eq. A- 2}$$

Réservoir de production

Le flux net interagit avec le réservoir de production. Si la pluie nette est positive, une partie P_s est dirigée vers le réservoir de production tandis que le reste P_n-P_s s'écoule vers la fonction de routage. La fraction de la pluie nette stockée P_s dans le réservoir dépend du niveau de remplissage de ce réservoir (du taux d'humidité du bassin). Pour une capacité de réservoir de production égale à *A* et son niveau au pas de temps *t* égal à S_t :

$$P_s = A\frac{(1-\frac{S_{t-1}}{A})^2 \tanh(\frac{P_n}{A})}{1+\frac{S_{t-1}}{A}\tanh(\frac{P_n}{A})} \quad \text{Eq. A- 3}$$

L'eau ne sort pas du réservoir par évapotranspiration si E_s est egal à 0. Au contraire, quand l'évapotranspiration est positive, une quantité E_s s'évapore du réservoir et est considérée comme perdue par le modèle.

$$E_s = \frac{S_{t-1}(2-\frac{S_{t-1}}{A})\tanh(\frac{P_n}{A})}{1+\frac{S_{t-1}}{A}\tanh(\frac{P_n}{A})} \quad \text{Eq. A- 4}$$

Le niveau du réservoir de production modifié, pour prendre en compte l'humidité du sol et de moduler l'écoulement direct :

$$S_t^+ = S_{t-1} - Es + Ps \quad \text{Eq. A- 5}$$

Percolation

Le réservoir perd de l'eau vers la fonction de production, suivant la loi de puissance :

$$dS = -kS^5 \quad \text{Eq. A- 6}$$

Cette fonction est appelée fonction de percolation. Cette loi intégrée sur le pas de temps devient alors :

$$Perc_t = S_t^+(1-(1+\frac{S_t^+}{K})^4)^{-1/4} \quad \text{Eq. A- 7}$$

La variable K est liée au pas de temps par la relation :

$$K = k(a-1)\Delta t)^{1/1-a} \quad \text{Eq. A- 8}$$

Et le niveau du réservoir de production devient :

$$S_t = S_t^{+} - Perc_t \quad \text{Eq. A- 9}$$

L'eau percolée rejoint l'écoulement direct, qui subit ensuite un ajustement multiplicatif par un facteur *X2* (un des paramètres libres du modèle) pour donner la pluie efficace entrant dans la fonction de routage :

$$PRT_t = X2*(P_n - P_s + -Perc_t) \quad \text{Eq. A- 10}$$

Evapotranspiration

L' equation suivante décrit mathématiquement la formule proposé par Oudin pour l'ETP journalière en fonction de la température de l'air T_a (en °C), du rayonnement extraterrestre *Re* (MJ*m^{-2}*jr^{-1}), de la chaleur latente de vaporisation de l'eau λ (MJ*kg^{-1}) et sa masse volumique ρ (kg*m^{-3}).

$$ETP(j) = \begin{cases} \frac{\mathrm{Re}(\phi,j)}{\lambda\rho}\left(\frac{T_a(j)+5}{100}\right) & si\ T_a(j) > -5 \\ 0 & si\ non \end{cases} \quad \text{Eq. A- 11}$$

Le rayonnement extraterrestre *Re* et la chaleur latente de vaporisation de l'eau λ sont exprimés par les équations suivantes :

$$\lambda = 2.501 - 0.002361*T_a \quad \text{Eq. A- 12}$$

$$\mathrm{Re} = \frac{24*60}{\pi} G_{sc} d_r \left[\omega_s \sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\sin(\omega_s)\right] \quad \text{Eq. A- 13}$$

où, G_{sc} : la constante solaire (égale à 0.0820 MJ*m^{-2}*min^{-1}),

d_r : la distance relative inverse Terre-Soleil,

ω_s : l'angle horaire au coucher de Soleil (rad),

ϕ : latitude (rad),

δ : déclinaison solaire (rad),

j : jour julien.

Fonction du routage

La fonction de routage retarde la sortie de la pluie efficace PRT_t sur les futurs pas de temps. Elle relie un routage linéaire (hydrogramme unitaire) et un routage non linéaire (par un réservoir de routage).

Routage linéaire

La pluie efficace PRT_t constitue l'entrée de l'hydrogramme unitaire symétrique, dont la fonction *UHC* (*hydrogramme unitaire cumulé*) est exprimée par :

$$UHC(i) = 0 \;\; si \;\; i \le 0$$
$$UHC(i) = \frac{i^{\alpha}}{i_{\alpha} + (X3 - i)^{a}} \;\; si \;\; i \le i \prec X3 \qquad \text{Eq. A- 14}$$
$$UHC = 1 \;\; si \;\; i \ge X3$$

Le temps de base de l'hydrogramme unitaire X3 est un des paramètres libres et l'exposant α est un paramètre fixe. La sortie de l'hydrogramme unitaire s'ecrit alors :

$$QUH_t = \sum_{i=1}^{X3} \gamma_i PRT_{t-1+1} \qquad \text{Eq. A- 15}$$

$$\gamma_i = UHC(i) - UHC(i-1) \qquad \text{Eq. A- 16}$$

La quantilé d'eau dan le réservoir de routage R_t^+ est egalé à la somme de la sortie de l'hydrogramme unitaire QUH_t et le volume de l'eau déjà présent dans le reservoir R_{t-1}.

Sortie du réservoir de routage

La sortie du réservoir suit une loi de vidage de type puissance :

$$\widehat{Q} = -dR = kR^{\beta} dt \qquad \text{Eq. A- 17}$$

Le réservoir est quadratique selon Tagara (2005) avec β=2, et la loi de vidage decrivant le débit de sortie devient alors :

$$\widehat{Q}_t = fX_{1,\beta=2}(R_t^+) = \frac{R_t^{+2}}{R_t^+ + X1} \qquad \text{Eq. A- 18}$$

, où X1 est un parmètre libre.

Le niveau du reservoir de production devient alors :

$$R_t = R_t^+ - \widehat{Q}_t \qquad \text{Eq. A- 19}$$

Mise à jour du réservoir de routage

Tagara (2005) propose une combinaison de deux mises à jour pour le modèle GRP. Le premier utilise la dernière observation des débits pour ajuster le niveau du réservoir de routage, tandis que le second utilise la dernière erreur observée pour corriger la prévision à venir en sortie du modèle. Seule la mise à jour du réservoir de routage est considérée dans cette étude.

La mise à jour du réservoir de routage est complète, elle calcule le niveau donnant le débit observé Q_t. Pour un reservoir quadratique, le niveau est donné par :

$$R_{t/t}^+ = f^{-1} X_{1,\beta=2}(Q_t) = \frac{\sqrt{Q_t^2 + 4X1 * Q_t} + Q_t}{2} \qquad \text{Eq. A- 20}$$

Le niveau du réservoir mis à jour après vidange s'ecrit alors :

$$R_{t/t} = f^{-1} X_{1,\beta=2}(Q_t) = \frac{\sqrt{Q_t^2 + 4X1 * Q_t} + Q_t}{2} \qquad \text{Eq. A- 21}$$

Valeurs des différents paramètres

Le *GRPE* a trois paramètres (X1, X2, X3) qui sont à caler pour chaque bassin. De plus, un certain nombre de paramètres fixes est defini à l'interieur du modèle de prévision. Au pas de temps journalier, ils sont donnés par Tangara (2005) et elles sont présentées dans le Tableau A-1.

Tableau A-1 : Les valeurs de paramètres fixes utilisées dans cette thèse.

Paramètres fixes	Symboles	Valeurs
Capacité du réservoir de production	A	350mm
Coefficient de la fonction de percolation	B	2,25
Exposant de l'hydrogramme unitaire	α	2,5
Exposant de la fonctionde vidqge du réservoir de routage	β	2,0

A-2 : Le modèle MORDOR

La version du modèle *MORDOR* utilisée dans cette thèse est une adaptation à 10 paramètres du modèle *MORDOR* original (Garçon, 1999 et Paquet, 2004). Le modèle est adapté à la prévision d'ensemble au pas de temps journalier. La Figure 12 représente schématiquement le modèle. Ci- après nous décrivons les principales fonctions du modèle utilisé.

A l'entrée du modèle hydrologique les précipitations totales sont corrigées par un facteur multiplicatif *X1.*

$$Pl = P * X1 \qquad \text{Eq. A- 22}$$

Par la suite la répartition de la pluie entre le réservoir superficiel U *dtr1* en fonction de la capacité maximale du réservoir *X8* (paramètre du modèle) et le réservoir intermédiaire L *dru1.*

$$dtr1 = \frac{Pl * U}{X8} \qquad \text{Eq. A- 23}$$

$$dtu1 = Pl - dtr1 \qquad \text{Eq. A- 24}$$

Par conséquent l'évolution *vs* du réservoir U :

$$vs = dtr1 + \max(0, U - X8) \qquad \text{Eq. A- 25}$$

Et le niveau réservoir U :

$$U = \min(U + dtu1, X8) \qquad \text{Eq. A- 26}$$

L'évapotranspiration *evu* par le réservoir superficiel U est égale à :

$$evu = \min(X8, \frac{E.U}{X8}) \qquad \text{Eq. A- 27}$$

Et le niveau final du réservoir U en prenant en compte l'évapotranspiration :

$$U' = U - evu \qquad \text{Eq. A- 28}$$

En ce qui concerne le réservoir intermédiaire L, l'évolution de son niveau *al* est influencé par sa capacité maximale *X9* ; ainsi que par l'évolution *vs* du niveau du réservoir U :

$$al = \min(X9 - L, vs(1 - \frac{L}{X9}))$$ Eq. A- 29

Par conséquent le niveau du réservoir intermédiaire L devient :

$$L' = L + al$$ Eq. A- 30

En ce qui concerne la vidange *vl1* et *vl2* du réservoir L, il dépend de son niveau L et des coefficients de vidange *X2, X3* (paramètre du modèle) :

$$vl1 = \frac{L'}{X2 * X3}$$ Eq. A- 31

$$vl2 = \frac{L'}{X3}$$ Eq. A- 32

Et la vidange totale est egale à :

$$vl = vl1 + vl2$$ Eq. A- 33

Si la vidange du réservoir *vl* est superieur à la capacité du réservoir L :

$$vl1 = \frac{vl1 * L}{vl}$$ Eq. A- 34

$$vl2 = \frac{vl2 * L}{vl}$$ Eq. A- 35

Et le niveau du réservoir L devient :

$$L'' = L' - vl$$ Eq. A- 36

Afin de calculer l'evolution du réservoir Z :

$$dtz = vl * (1 - Z / X10)$$ Eq. A- 37

$$rur = vl1 * (Z / X10)$$ Eq. A- 38

$$an = vl2 * (Z / X10)$$ Eq. A- 39

Alors le niveau du reservoir Z devient :

$$Z' = Z + dtz \quad \text{Eq. A- 40}$$

$$evz = \min(Z', (E - evu) * Z / X10) \quad \text{Eq. A- 41}$$

$$Z'' = \min(X10, Z' - evz) \quad \text{Eq. A- 42}$$

En ce qui concerne le reservoir N :

$$N' = N + an \quad \text{Eq. A- 43}$$

$$vn = \min(N', N' / X4)^{X5}) \quad \text{Eq. A- 44}$$

Le niveau du reservoir :

$$N'' = N' - vn \quad \text{Eq. A- 45}$$

La somme Q_t de differentes contributions :

$$Qt = vs - al + rur + vn \quad \text{Eq. A- 46}$$

Le routage de ces contributions est faite par une loi de Weibull :

$$Q = Weibull(X6, X7, Qt) \quad \text{Eq. A- 47}$$

A-3 : Module Neige du modèle *GRP*

Le module *CemaNeige* a été développé et testé pour la simulation hydrologique par Valéry (2010). Il a été ensuite adapté pour la prévision par Nicolle *et al.* (2011).

Il s'agit d'un module qui suit une approche empirique pour la modélisation des stocks et de la fonte de la neige. Le module *CemaNeige* est inséré en amont du modèle hydrologique. Les deux paramètres supplémentaires apportés par le module CemaNeige sont le facteur de fonte (*Kf*) et le coefficient de pondération de l'état thermique du manteau (CT_G). Les données de forçage sont les pluies (*P*), la température de l'air (*Tair*) et l'évapotranspiration potentielle (*ETP*), tandis que la mise à jour a besoin du dernier débit observé pour être effective.

Dans l'adaptation du module *CemaNeige* au modèle de prévision *GRP* (Nicolle *et al.*, 2011), la séparation de la pluie en liquide ou solide n'est pas basée sur les températures minimales et maximales, car souvent nous ne disposons pas ces données, mais seulement d'une température moyenne journalière. Ainsi, c'est cette variable qui est utilisée dans la fonction introduit par US Army Corps of Engineer (1956) et présenté dans la Figure A-1.

Pour le fonctionnement du module Neige :

- les bassins versants sont distribués en 5 zones d'altitude d'égale surface. Pour chaque zone, les valeurs de température et de pluie sont extrapolées, en étant corrigées par des facteurs altitudinaux Pz et Tz,
- Chacune de zones d'altitude z est caractérisée par 2 états internes pour chaque pas de temps j : le stock de neige (G(z,j)) et l'etat thermique du stock (eT(z,j)),
- la transition de pluie liquide en pluie solide est déterminée en combinant deux fonctions différentes répandues dans la littérature selon l'altitude moyenne du bassin versant (inférieure ou supérieure à 1500m, respectivement),
- il y a quatre paramètres fixes : l'intervalle de température dans lequel les fractions pluie liquide/solide sont estimées (ΔT_s), la température de fonte (T_f = 0°C), le seuil à partir duquel toute la zone d'altitude est considérée comme enneigée (G_{seuil}) et la % de K_f qui correspond à la vitesse de fond minimale,

Fonction $f([-1;+3°C])$, type USACE (US Army Corps of Engineers, 1956)	Fonction $f(T_{min}, T_{max})$, type Hydrotel (Turcotte et al., 2007)
Si $Tz < -1°C$, $FracNeige = 1$	Si $Tz\max \leq 0°C$, $FracNeige = 1$
Si $Tz > +3°C$, $FracNeige = 0$	Si $Tz\min \geq 0°C$, $FracNeige = 0$
Si $Tz \in [-1;3]°C$, $FracNeige = 1 - \frac{Tz - (-1)}{3 - (-1)}$	Sinon, $FracNeige = 1 - \frac{Tz\max}{Tz\max - Tz\min}$

Figure A-1 : Les fonctions de détermination de la nature des précipitations selon l'altitude moyenne du bassin versant (source: Valéry, 2010).

A-4 : Module Neige du modèle *MORDOR*

Aux dix paramètres initiales, il faut ajouter les paramètres libres du module neige : leur nombre varie de 4 à 11 (Garçon, 1999) selon les objectifs du calage. La plupart de modules neige existants dans la littérature se basent sur le principe que les précipitations sont très variables spatialement et selon l'altitude et les températures locales. Dans cette direction pour le traitement de la composante nivale nous trouvons deux approches dans la littérature : le traitement du bassin versant dans sa globalité ou divisé en bandes d'altitude sur lesquelles le module neige est appliqué indépendamment. La première approche est suivie dans le cas du modèle *MORDOR* global permettant la modélisation du comportement des manteaux neigeux en utilisant une courbe hypsométrique pour déterminer la quantité de pluie et de neige (mais pas en utilisant une distribution par bandes d'altitude). Néanmoins, le module neige considère un état interne où il calcule pour chaque pas de temps qui est le pourcentage de bassin enneigé.

Sommairement, la version du modèle utilisé lors de cette thèse présente 11 paramètres libres au total (les dix paramètres décrits à l'ANNEXE A-2, plus un paramètre qui dirige la lame de fonte de neige à différents endroits du modèle hydrologique) et quatre paramètres fixes. C'est à noter que la module neige du modèle MORDOR est intégrée dans sa structure. Dans le module neige une partie de la lame de fonte est directement dirigée vers le ruissellement tandis qu'une autre partie est ajoutée soit au réservoir de surface *U*, soit au réservoir profond *N*. La répartition se fait à l'aide d'un paramètre à optimiser lors de la phase de calage.

Les coefficients de la fraction liquide de la précipitation consistent les 4 paramètres fixes du module et sont calculé à partir de la courbe hypsométrique. Ces derniers sont déterminés en fonction des caractéristiques altitudinales du bassin versant et de celles du point de mesure de la température de l'air. L'hypsométrie du bassin versant est prise en compte par une fonction seuil empirique du type *FracNeige = f (Tair)*. Une représentation de cette fonction seuil empirique est illustrée dans la Figure A- 2. De manière analytique, le module neige a deux états internes pour la modélisation nivale : les indices thermiques du stock de neige au pars de temps *j* (*TG(j)*) et de fonte au pas de temps *j* (*Tfonte(j)*). La transition de pluie en neige (et vise versa) est déterminé sur un intervalle de *ΔTs* définit par le US Army Corps of Engineer (1956).

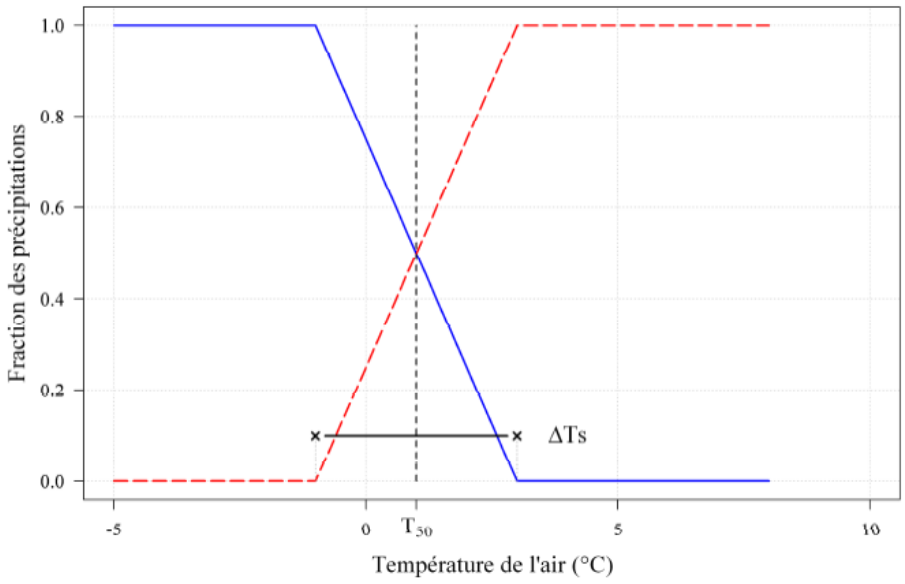


Figure A- 1 : Fonction seuil empirique de forme linéaire représentant les fractions solide (ligne continue) et liquide (ligne pointillée) en fonction de la température de l'air. Un mélange pluie / neige est possible sur l'intervalle ΔTs (tiré de Valéry, 2010).

Le paramètre libre du module neige à optimiser, qui représente la diversion de la lame de fonte de neige à différents endroits du modèle hydrologique, sera multiplié par le pourcentage de bassin enneigé. Ainsi, la fonte à la base n'est plus constante mais dépend de la superficie enneigée du bassin à un pas de temps donné *j* et pour une zone d'altitude *z*.

$$Fonte_{base}(z, j) = K_{f,base} * p_{neige}(z, j) \qquad \text{Eq. 79}$$

, ou $Fonte_{base}$: la quantité d'eau de fonte à la base du manteau neigeux en mm ;

$p_{neige}(z, j)$: le pourcentage de surface enneigée ;

$K_{f,base}$: la quantité d'eau fondue à la base du stock quand toute la surface considérée est recouverte de neige, en mm.

ANNEXE B : Extrait du modèle heuristique de gestion

Tableau B-1 : Extrait du code heuristique présentant le gain total, les volumes turbinés/ déversés ; ainsi que la différence de stock initial pour les différents coefficients de capacité et de turbinage.

Coef. Turbinage= 3	Gain Total(kEuros)	Volume turbiné total(Mm3)	Vol total déversé(Mm3)	Dif. stock initial et final(Mm3)
coef. capacité= 2	3412.42	98.3864	0	−5.84
coef. capacité= 3	3512.88	101.8086	0	−9.27
coef. capacité= 4	3597.41	105.2307	0	−12.69
coef. capacité= 5	3689.48	108.6529	0	−16.11
coef. capacité= 6	3773.77	112.075	0	−19.53
coef. capacité= 7	3829.55	115.4971	0	−22.95
coef. capacité= 8	3787.47	117.2082	0	−24.67
coef. capacité= 9	3788.43	118.0637	0	−25.52
coef. capacité= 10	3788.43	118.0637	0	−25.52

ANNEXE C : Conférences et ateliers

Conférences/ Ateliers:

- European Geophysics Union General Assembly (Vienna, 6-10 avril 2013) : "Hydrological ensemble predictions for reservoir inflow management." (présentation sous forme de poster).
- HEPEX workshop (Beijin, 22-24 octobre 2012) : "The value of hydrological ensemble predictions for reservoir inflow management" (présentation orale).
- European Geophysics Union General Assembly (Vienna, 22-27 avril 2012) : "Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of bias correction strategies." (présentation sous forme de poster).
- 11th European Meteorological Society Annual Meeting (Berlin, 12-16 september 2011) : "Post-processing approaches for hydrological ensemble prediction: a comparative study on two hydrological forecasting chains" (présentation orale).
- Journée de Doctorants d'EDF DTG- R&D (Grenoble, 20 juin 2011) : "Prévision probabiliste des crues à partir de la prévision d'ensemble de pluies : post-traitement des faisceaux de simulation et évaluation des incertitudes" (présentation orale)
- HEPEX workshop (Unesco, Delft, 7-9 juin 2011) : "Impact of sample size on forecast verification scores and post processing parameters of hydrological ensemble predictions"(présentation orale).

- Journée de Doctorants de l'Ecole Doctorante GRN (Ecole de Mines ParisTech, 10-10 mai 2011) : "Prévisions probabilistes des débits à partir des prévisions d'ensemble : faisceaux et traitement" (présentation orale).
- European Geophysics Union General Assembly (Vienna, 4-8 avril 2011) : "Ensemble hydrological forecasting based on weather predictions from the TIGGE database" (présentation sous forme de poster).
- European Geophysics Union General Assembly (Vienna, 4-8 avril 2011) : "How much the length of data samples influences forecast verification and post processing of hydrological ensemble predictions?" (présentation sous forme de poster).
- European Geophysics Union General Assembly (Vienna, 4-8 avril 2011) : "Operational use of ensemble hydrometeorological forecasts at EDF (French energy producer)" (présentation sous forme de poster).
- American Geophysics Union Fall Meeting (San Francisco, 13-17 december 2010) : "Using the TIGGE database for ensemble hydrological forecasting: a study on 74 catchments in France" (présentation sous forme de poster).

ANNEXE D : Article

Article :

- Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. Advances in Science and Research, 8: 135-141.

Adv. Sci. Res., 8, 135–141, 2012
www.adv-sci-res.net/8/135/2012/
doi:10.5194/asr-8-135-2012
© Author(s) 2012. CC Attribution 3.0 License.



Advances in Science & Research
Open Access Proceedings

Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies

I. Zalachori[1], M.-H. Ramos[1], R. Garçon[2], T. Mathevet[2], and J. Gailhard[2]

[1]IRSTEA, Hydrology Group, UR HBAN, 1 rue Pierre-Gilles de Gennes CS 10030, 92761 Antony, France
[2]EDF-DTG, Electricité de France, Direction Technique de Grenoble, France

Correspondence to: I. Zalachori (ioanna.zalachori@irstea.fr)

Received: 15 January 2012 – Revised: 16 May 2012 – Accepted: 7 July 2012 – Published: 25 July 2012

Abstract. The aim of this paper is to investigate the use of statistical correction techniques in hydrological ensemble prediction. Ensemble weather forecasts (precipitation and temperature) are used as forcing variables to a hydrologic forecasting model for the production of ensemble streamflow forecasts. The impact of different bias correction strategies on the quality of the forecasts is examined. The performance of the system is evaluated when statistical processing is applied: to precipitation and temperature forecasts only (*pre-processing* from the hydrological model point of view), to flow forecasts (*post-processing*) and to both. The pre-processing technique combines precipitation ensemble predictions with an analog forecasting approach, while the post-processing is based on past errors of the hydrological model when simulating streamflows. Forecasts from 11 catchments in France are evaluated. Results illustrate the importance of taking into account hydrological uncertainties to improve the quality of operational streamflow forecasts.