

**Université Pierre et Marie Curie, École des Mines de Paris
& École Nationale du Génie Rural des Eaux et des Forêts**

**Master 2 Sciences de l'Univers, Environnement, Ecologie
Parcours Hydrologie-Hydrogéologie**

**Construction d'un faisceau de prévisions hydrométéorologiques
à partir d'une prévision à scénario-unique de précipitation**

Hajer DHOUIOUI

Encadrée par : Maria-Helena RAMOS



Cemagref, UR-HBAN

**1 rue Pierre Gilles de Gennes
92761 Antony Cedex, France**

Septembre 2011



Au terme de ce travail, je tiens à remercier Maria-Helena RAMOS, mon encadrante de stage, pour les conseils qu'elle n'a cessé de me prodiguer ainsi que pour son indulgence et la confiance qu'elle m'a toujours accordée. J'ai eu la chance d'apprécier ses qualités humaines et scientifiques et je lui suis très reconnaissante pour l'aide qu'elle m'a apporté et d'avoir su se rendre disponible malgré ses activités débordantes...

J'adresse par la même occasion tous mes remerciements à Benjamin RENARD du Cemagref à Lyon pour ses précieux conseils et suggestions en analyse statistique.

Je tiens également à exprimer mon profond respect et ma reconnaissance à Vazken ANDREASSIAN et Charles PERRIN, pour leurs commentaires constructifs et pour l'accueil qu'ils m'ont réservé, ainsi qu'à toute l'équipe HYDRO du Cemagref (Abderrahmen, Annie Damien, Florent, François, Gianluca, Ioanna, Julien, Pierre, Pierre-Yves, Louise et Raji) pour l'ambiance de travail qui était vraiment très enrichissante professionnellement et personnellement.

Un grand merci à Vazken ANDREASSIAN et à Maria Helena RAMOS pour les cours d'initialisation aux langages de programmation Fortran et R qui m'ont fait gagner beaucoup de temps.

Merci à Annie RANDRIANASOLO pour les données hydrométéorologiques et à Ioanna ZALACHORI pour nos discussions plus qu'utiles !

Merci à mes collègues du bureau Louise et Cécile pour l'ambiance chaleureuse et pour leurs bonne humeur. Merci aussi à tous les stagiaires, les thésards et le personnel de l'unité HBAN avec lesquels j'ai pu partager des moments agréables...

J'aimerais aussi remercier tous mes professeurs, pour la qualité de la formation dans ce master.

Mes remerciements s'adressent également au président et à tous les membres du jury pour avoir accepté d'évaluer ce travail.

Hajer,

Hydrometeorological forecasting is not an exact science and it should question about the magnitude of the uncertainties associated with the prediction announced. Even if it is sometimes accompanied by qualitative indications of the level of uncertainty, a deterministic forecast gives an illusion of certainty and can lead to inappropriate decisions by managers. In the operational forecast centers, managers are often required to make decisions despite the blur due to persistent use of such forecasting. However, working with ensemble forecasts allows to manage much more explicitly and accurately the potential risks and to deduce the probability of any future outcome. Thus, these techniques provide a decision support for managers.

In this study, we developed and tested a statistical perturbation method of a deterministic scenario forecast to generate hydrometeorological ensemble forecasts. Uncertainties due to the forecasted amount of precipitation, known to be the most important hydrological forecasting system, were quantified and used to create different scenarios of precipitation forecasts.

The procedure is based on statistical modelling of past errors of a single-valued forecast, by a multivariate Gaussian distribution using the Box-Cox transformation, through the different lead-times. For this, two statistical models were checked, a general model (processing all peer observation-prediction) and an alternative model (processing only peer-observation forecasting non-zero).

The method implemented in this work was applied using an archive of deterministic forecasts of the European Centre for Medium-Range Weather Forecasting (ECMWF) for 1301 days, aggregated on 74 watersheds in France. Ensemble forecasts generated were used on the input of hydrological model GRP of the Cemagref to conclude on the contributions of this method in a hydrological forecasting system (i.e. the quality of the ensemble flow forecasts resulting).

The results showed that, overall, the ensemble forecasts generated by the proposed procedures are better than the original deterministic forecast in terms of detecting high thresholds while being very similar to the ECMWF ensemble forecasts.

Keywords: precipitation forecast, statistical analysis, forecast errors, model error, Box-Cox transformation, multivariate normal distribution, forecast scenarios, discharge forecast, ensemble prediction.

Table de matières

Abstract	3
Chapitre 1: Introduction générale	8
1.1. Contexte et motivations de l'étude	10
1.2. Objectifs	11
1.3. Organisation du rapport.....	12
Chapitre 2 : Comment construire et évaluer des faisceaux de prévisions ?	13
2.1. Approches de modélisation : vers une prévision d'ensemble hydro-météorologique	13
L'approche multi-modèle	13
2.2. Evaluation des prévisions : comment reconnaître une bonne prévision ?	16
Chapitre 3 : Choix méthodologiques	19
3.1. Construire un modèle des erreurs passées de la prévision déterministe.....	19
3.1.1. Quel type de modèle ?	19
3.1.2. Quelles échéances de prévision pour le modèle d'erreur ?	22
3.1.3. Comment construire le modèle multivarié des résidus ?	24
3.2. Génération des faisceaux des prévisions des pluies	25
3.2.1. Tirage aléatoire d'une loi de probabilité multivariée	26
3.2.2. Construction des scénarios de prévision.....	27
3.3. Contrôle du modèle d'erreur	28
3.3.1. Les critères quadratiques	28
3.3.2. Les critères liés à des tables de contingence.....	29
3.3.3. Les critères graphiques	30
Chapitre 4 : Cas d'application Données, bassins versants et modèle	32
4.1. L'archive de prévisions météorologiques	32
4.2. Le modèle de prévision hydrologique GRP	33
4.3. Bassins versants étudiés et données hydrométéorologiques	34
Chapitre 5: Résultats & interprétations	36
5.1. Analyse exploratoire des données en période de calage	36
5.2. Calage du modèle statistique des erreurs de prévision.....	38
5.2.1. Calage des paramètres de la transformation Box-Cox (λ_1, λ_2)	38
5.2.2. Calage des paramètres du modèle gaussien multivarié (μ, Σ).....	39
5.3. Contrôle des modèles statistiques d'erreurs.....	45
5.3.4. Un évènement marqué	53
5.4. Synthèse des principaux résultats.....	56
Chapitre 6: Conclusions et perspectives	57
Références bibliographiques	59
Annexes : Liste des bassins versants étudiés	62

Liste des figures

Figure 1. Exemple de bulletin de prévision des précipitations tel que fourni aux SCPs et au SCHAPI en 2008.....	11
Figure 2. Schéma générale de la prévision hydrologique avec indication des parties du système qui sont concernées par cette étude (cases colorées).....	12
Figure 3. Illustration simplifiée du tirage aléatoire d'une loi binormale (si $e=2$).	27
Figure 4. Principe de la table de contingence : cas possibles pour un évènement donné en prévision. Chaque case rassemble le nombre d'occurrences du cas correspondant et la somme des quatre cases est égale au nombre de prévisions émises.	30
Figure 5. Exemple de 24 membres de la prévision d'ensemble du CEPMMT (EPS) émise le 28-02-2007 à 4 jours d'échéance sur la France.	33
Figure 6. Structure du modèle hydrologique de prévision GRP : à gauche un schéma simplifié du fonctionnement des deux réservoirs (production et routage) et à droite une légende récapitulative des principaux paramètres et variables d'état du modèle.....	34
Figure 7. Localisation des 74 bassins versants étudiés en France (contours noirs).....	35
Figure 8. Les cas envisageables lors d'un exercice d'évaluation d'une prévision.....	37
Figure 9. Diagnostic de l'échantillon des erreurs en période de calage du modèle d'erreurs et avant le traitement des données (a) Saisonnalité (b) Homoscédasticité (c) Normalité et (d) auto-corrélation.....	38
Figure 10. Optimisation du paramètre λ_1 de la transformation de Box-Cox pour les échéances 1, 5 et 9 jours de la prévision déterministe du CEPMMT. Exemple du bassin « La Meuse à Commercy » (2327,9 km ²).	39
Figure 11. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage pour l'exemple du bassin «La Sarthe à Neuville-sur-Sarthe» à l'échéance 2 de la prévision. ..	41
Figure 12. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage du modèle (B) pour le bassin L'Armançon à Aisy sur Armançon présentant la p-valeur maximale (échéance 9).....	42
Figure 13. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage du modèle d'erreur (B) pour le bassin La Midouze à Campagne présentant la p-valeur minimale (échéance 3).....	43
Figure 14. Cartes des résultats des tests statistiques de normalité.....	44
Figure 15. Ecart en erreurs quadratiques par rapport aux prévisions DET et EU sur les 74 bassins versants étudiés pour les échéances de prévision $j+1$, $j+5$ et $j+9$	46
Figure 16. Dispersion des scores BS de la prévision PDH du modèle B par rapport aux EU du CEPMMT (en haut) et QDH en fonction des prévisions de débits à partir des EU du CEPMMT (en bas) pour les échéances de prévision $j+1$, $j+5$, $j+9$ et le dépassement des seuils des quantiles Q95%, Q75%, Q50% et Q25%.....	47

Figure 17. Box-plots des indices de succès critiques (en haut), des ratios de fausses alertes (au milieu) et des taux de succès (en bas) pour les échéances de 1, 5 et 9 jours de la prévision déterministe des précipitations (DET) et pour au moins 80% des membres de la prévision d'ensemble PDH (modèles A et B) sur les 74 bassins considérés et pour les dépassements des quantiles Q95% et Q75%.....	49
Figure 18. Idem que la figure 17 pour les prévisions d'ensemble des débits.	50
Figure 19. Box-plots des scores AUC sur les 47 bassins versants étudiés pour les prévisions des pluies (PDH et EU) en haut et des débits (QDH et EU) en bas.	51
Figure 20. Courbes de ROC pour les prévisions d'ensemble des pluies pour le dépassement du quantile 95% (échéance j+1, j+5 et j+9)	52
Figure 21. Idem que la figure 13 pour les prévisions d'ensemble des débits.	53
Figure 22. Prévisions réalisées par le modèle GRP lors de l'évènement du 04-03-2010 dans le bassin «La Meuse à Commercy» (2327,9km ²) sur 9j d'échéances.....	55

Liste des tableaux

Tableau 1. Quelques exemples d'études de cas récentes concernant les prévisions d'ensemble hydrologiques.....	18
Tableau 2. Effets du paramètre λ_1 de la transformation Box-Cox (d'après Osborne, 2010 et modifié) ..	22
Tableau 3. Principales caractéristiques des bassins versants étudiés pour la période d'étude (sur la période 1958-2009).	35

Liste des acronymes

ACF	Auto-Correlation Function
AP	Alerte Prévision
AUC	Area Under the Curve
BGM	Breeding of Growing Modes
BS	Brier Score
CEPMMT	Centre Européen de Prévision Météorologique à Moyen Terme
CMC	Centre Météorologique Canadien
CMC	Centre Météorologique Canadien
COST731	Propagation of Uncertainty in Advanced Meteo-Hydrological Forecast Systems, Cooperation Science and Technology
DET	prévision déterministe
DGPR	Direction Générale de la Prévention des Risques
DH	Déterministe Habillée
DH	Déterministe Habillée
EDF	Electricité De France
EFAS	European Flood Alert System
EnKF	Ensemble Kalman Filter
EU	prévision d'ensemble du CEPMMT
FAR	False Alarm Ratio
FDP	Fonction de Distribution de Probabilité
GLUE	Generalized Likelihood Uncertainty Estimation
HEPEX	Hydrological Ensemble Prediction EXperiment
MAP-DPHASE	Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region

MEEDM	Ministère de l'Écologie, de l'Énergie, du Développement durable et de la Mer
NCEP	National Center for Environmental Prediction
OMM	Organisation Mondiale de la Météorologie
PDH	Prévision d'ensemble des précipitations DH
PE	Prévision d'Ensemble
PEARP	Prévision d'Ensemble ARpege
PHE	Prévision Hydrologique d'Ensemble
PME	Prévision Météorologique d'Ensemble
PNT	Prévision Numérique du Temps
POD	Probability Of Detection
POFD	Probability Of False Detection
PQP	Prévision Quantitative des Précipitations
QDH	Prévision d'ensemble des débits DH
RMSE	Root Mean Square Error
ROC	Relative Operating Curve
SCHAPI	Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations
SPC	Service de Prévision des Crues
SPE	Système de Prévision d'Ensemble
TIGGE	THORPEX Interactive Grand Global Ensemble

Notations

δ	résidu ou erreur de prévision
Y	pluie observée
X	pluie prévue
δ_t^i	résidu associée à la prévision émise à l'instant t et pour l'échéance i ,
Y_t^i	pluie observée à la même date de l'échéance i de la prévision émise à l'instant t
X_t^i	pluie prévue à l'instant t et pour l'échéance i
Θ	jeu de paramètres quelconque
Z	variable sur laquelle une transformation Box-Cox sera appliquée
$BC(Z)$	la transformée Box-Cox d'une variable Z
λ_1	paramètre qui définit le type de la transformation Box-Cox
λ_2	constante choisie arbitrairement pour rendre la variable Z non-nulle pour une transformation Box-Cox
ε_t^i	résidu transformé (après transformation Box-Cox des variables), associé à la prévision émise à l'instant t et pour l'échéance i
e	nombre total d'échéances de la prévision
N_t	nombre total de jours de prévision
M	nombre total de membres d'une prévision d'ensemble
μ	vecteur de moyenne sur tous les jours de calage pour chaque échéance; $\mu = (\mu_1, \mu_2, \dots, \mu_e)$
Σ	matrice de covariance des erreurs
C_e	matrice de corrélation des erreurs de prévisions
σ_i	écart-type du vecteur ε^i
σ_{ij}	covariance des vecteurs ε^i et ε^j
$z_{\alpha/2}$	une constante qui dépend du degré de confiance α souhaité dans un intervalle de confiance

Chapitre 1: Introduction générale

“Mieux vaut prévenir que guérir !”

La prévision hydrologique est devenue un atout économique et social important pour la maîtrise des risques liés à l'eau. En effet, les intempéries, les crues et les inondations, ainsi que les étiages sévères et les sécheresses, représentent, depuis longtemps, des pertes importantes pour les communautés humaines à travers le monde entier. D'après le bilan décennal des catastrophes naturelles en France et dans les DOM, les inondations en particulier ont représenté la première cause de catastrophe naturelle, avec 20% des événements recensés entre 2001 et 2010¹. Il est estimé qu'une crue de la Seine comme celle de 1910 coûterait à la collectivité nationale environ 10 milliards d'euros. D'où l'intérêt non seulement écologique, mais aussi sécuritaire et financier pour prédire, de la manière la plus fiable possible, les précipitations et les débits qui en découlent.

Néanmoins, de par le caractère incertain de tout phénomène naturel et de l'imperfection des modèles, la prévision des précipitations et des crues n'est, et ne sera jamais, fiable à cent pour cent (Lorenz, 1969 cité dans Cloke et Pappenberger, 2009 ; Hingray *et al.*, 2009). En hydrologie, il y a différents facteurs qui affectent la qualité de la prévision des débits, tels que (Krzysztofowicz, 2001b) :

- les incertitudes opérationnelles dues aux événements imprévisibles au cours du processus de prévision, telles que les ruptures de barrages. Ce genre d'incertitude ne peut pas être quantifiée à l'avance et donc reste inexprimée ;
- les incertitudes de la modélisation hydrologique, qui sont liées au calage des paramètres des modèles, à l'incertitude quant à la structure des modèles, à l'incertitude des courbes de tarage et données de débits, aux incertitudes sur l'état initial des bassins versants modélisés, etc. (voir, par exemple, Yang *et al.*, 2007) ;
- les incertitudes liées aux forçages atmosphériques des modèles hydrologiques, tels que les précipitations observées et prévues. Le membre le plus important de ce groupe reste la prévision quantitative des précipitations (PQP), qui représente une source majeure d'incertitude des systèmes de prévisions hydrologiques (Krzysztofowicz, 1998 ; Seo *et al.*, 2000 ; Ramos *et al.*, 2010 ; Vincendon, 2011 ; Wu *et al.*, 2011).

Les différentes sources d'incertitude affecteront de façon plus ou moins importante les prévisions de débits selon l'échéance de prévision à laquelle on s'intéresse. En effet, la prévision se distingue de la simulation par, entre autres, la prise en compte d'une dimension temporelle supplémentaire à travers l'échéance de prévision (ou horizon de validité d'une prévision).

Les prévisions hydrologiques réalisées à l'aide seulement des observations des pluies enregistrées au sein du bassin versant, ou bien celles basées sur une approche purement hydraulique (les données d'entrée aux modèles sont uniquement les débits mesurés en amont de l'exutoire où on souhaite prévoir les débits), sont souvent capables de prévoir à des horizons plus courts, quelques heures ou quelques jours, en fonction du temps de réponse du bassin versant. Elles seront ainsi limitées par le temps de transit de l'eau dans le bassin versant. Ces prévisions seront principalement affectées par les incertitudes des observations, des conditions

¹ http://www.catnat.net/Bilan_France.pdf

initiales ou des paramètres des modèles, notamment les paramètres liés à la fonction de transfert hydrologique.

Une prévision à courte échéance ne permet cependant pas de fournir toujours des prévisions en temps opportun pour les alertes et les différentes préparations avant un événement de crue. Par conséquent, afin de pouvoir étendre les échéances de prévision à des délais supérieurs (plusieurs jours), il est souvent nécessaire de faire appel à des prévisions quantitatives de précipitations (PQP), utilisées en entrée aux modèles pluie-débit (Maidment, 1993). Or, les PQP sont entachées de fortes incertitudes, ce qui aura une influence sur la qualité des prévisions hydrologiques.

L'exercice de prévision nécessite alors d'être accompagné de la prise en compte des principales sources d'incertitude en jeu et de leur quantification. Ce sujet suscite l'intérêt de nombreux chercheurs, notamment dans le but de répondre aux besoins de la prévision opérationnelle en termes d'évaluation du degré d'incertitude d'une prévision (Schaeffli *et al.*, 2007; Yang *et al.*, 2007; Joly et Descamps, 2009; Mascaro *et al.*, 2010). Le développement de systèmes de prévisions capables de fournir plusieurs scénarios futurs pour le même endroit de prévision et pour chaque échéance d'intérêt compte parmi les voies les plus explorées de la recherche scientifique, qu'il s'agisse de fournir l'ensemble de scénarios considérés possibles, une probabilité associée à un événement prévu ou des intervalles de confiance sur les prévisions émises.

Les systèmes de prévision d'ensemble existent en météorologie depuis les années 1990 (Palmer *et al.*, 2002). Ils génèrent plusieurs scénarios possibles prévus des variables météorologiques (températures, vents, précipitations, etc.), à partir des états initiaux légèrement perturbés de l'atmosphère, afin de simuler l'incertitude des conditions initiales et fournir à l'utilisateur une prévision probabiliste (Palmer et Buizza, 2007, cité dans Cloke et Pappenberger, 2009). En hydrologie, la prévision d'ensemble à partir de l'utilisation de la prévision d'ensemble météorologique en entrée à un modèle hydrologique devient une pratique de plus en plus courante (voir la synthèse proposée par Cloke et Pappenberger, 2009). Elle peut être combinée ou non avec l'utilisation de plusieurs modèles hydrologiques (jeux de structures ou de paramètres) pour la prise en compte également des incertitudes de la modélisation hydrologique (Velázquez *et al.*, 2011).

Contrairement aux techniques plus traditionnelles de prévision qui produisent une prévision unique, dite déterministe, la prévision d'ensemble donne une information quantitative sur l'incertitude des prévisions, en proposant un ensemble de scénarios alternatifs considérés équiprobables (Atger, 2001). Par conséquent, le prévisionniste pourra se baser sur les différents scénarios dont il dispose pour émettre un indice de confiance sur chaque événement prévu, ce qui l'aidera à la prise de décision concernant la gestion des risques : une forte probabilité d'occurrence d'un événement d'intensité donnée imposera une alerte et/ou une action protectrice contre des pertes (économiques et/ou humaines) possibles.

Du fait qu'une prévision déterministe ne soit pas accompagnée d'indications quantitatives ou même qualitatives concernant le niveau d'incertitude, ce type de prévision devient insuffisant et peut conduire à une mauvaise, voire dangereuse, gestion du risque associé. C'est la raison pour laquelle il est essentiel de prendre en compte les incertitudes dans les prévisions, notamment quand on s'intéresse à améliorer la qualité et la valeur de l'information émise.

1.1. Contexte et motivations de l'étude

Ce travail s'intéresse aux incertitudes des précipitations prévues, affectant les prévisions de débits. Il s'est déroulé au Cemagref² à Antony, dans l'unité de recherche « Hydrosystèmes et bioprocédés » (HBAN). C'est une étude qui a contribué à l'équipe « Hydrologie des bassins versants » (HYDRO) dans le cadre des collaborations menées entre le Cemagref, les 22 Services de Prévision des Crues en France (SPC), qui assurent la surveillance, la prévision et la transmission de l'information hydrologique en cas de crue pour l'anticipation des événements extrêmes, et le Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations (SCHAPI, créé en juin 2003). Rattaché à la Direction Générale de la Prévention des Risques (DGPR) du Ministère de l'Écologie, de l'Énergie, du Développement durable et de la Mer³ (MEEDM), le SCHAPI constitue un appui aux SPCs au niveau national et assure une veille hydrométéorologique continue (24h/24h) sur des tronçons de vigilance aux crues.

Actuellement, l'information sur la « vigilance crues » mise à disposition du public par le SCHAPI se résume à l'affichage des hauteurs d'eau (ou des débits) provenant des réseaux d'hydrométrie en temps réel et, en cas de vigilance particulière requise, d'informations sur le risque associé durant les prochaines 24 heures⁴. Des prévisions déterministes (un seul scénario) sont calculées en cours d'événements uniquement pour les besoins d'organisation de la sécurité civile. Cependant, il y a une véritable attente à ce que les prévisions et leurs incertitudes soient également affichées (Berthet *et al.*, 2011). Pour cela, plusieurs expérimentations, avec les prévisions d'ensemble (du Centre Européen de Prévision Météorologique à Moyen Terme dans la chaîne de modélisation SIM de Météo-France, du modèle Arpège de Météo-France, du système européen EFAS, entre autres) sont menées depuis quelques années par le SCHAPI.

En prévision opérationnelle, cependant, les prévisionnistes sont aujourd'hui amenés à évaluer leurs prévisions des débits et à fournir une indication concernant les risques hydrologiques à partir d'une information limitée sur les incertitudes autour des précipitations prévues, comme le montre l'exemple du « bulletin des précipitations » présenté dans la figure 1. Il s'agit d'une communication des hauteurs de précipitations observées et prévues, pour un horizon maximal de trois jours, fournies en moyenne spatiale sur des zones dites AP (Alerte Prévision), selon une fourchette min/max, ou sous la forme d'une hauteur ponctuelle maximale à l'intérieur de la zone. Sur la base de cette information, le prévisionniste doit constituer le scénario de forçage à considérer sur le(s) bassin(s) versant(s) surveillé(s), sachant qu'un bassin versant peut être partiellement ou totalement à l'intérieur d'une ou plusieurs zones AP.

De manière générale, la prévision déterministe utilisée actuellement correspond au « meilleur » jugement du prévisionniste, auquel il rajoute une marge d'incertitude « expertisée » qui peut aussi bien converger que diverger de la meilleure estimation (qui serait l'observation), ce qui peut être gênant dans une optique de prise de décision. Le développement d'une meilleure estimation de l'incertitude, basée sur une approche statistiquement cohérente, semble alors être pertinent et essentiel.

² <http://www.cemagref.fr/>

³ <http://www.developpement-durable.gouv.fr/>

⁴ <http://www.vigicrues.gouv.fr/>

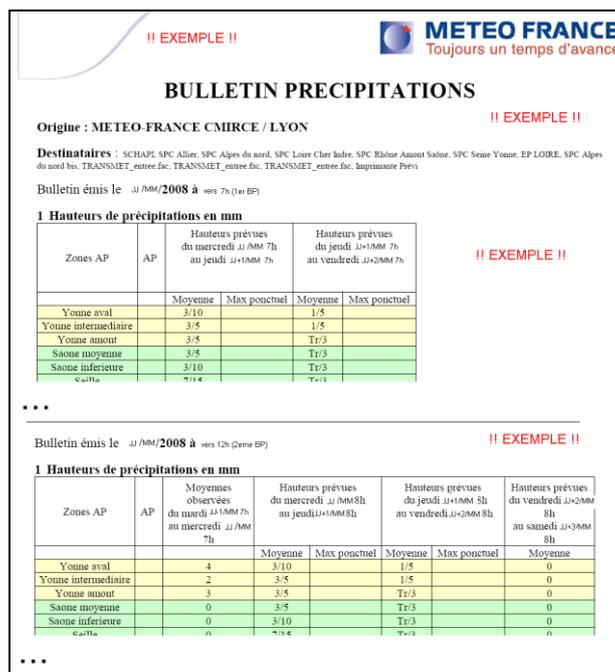


Figure 1. Exemple de bulletin de prévision des précipitations tel que fourni aux SCPs et au SCHAPI en 2008.

1.2. Objectifs

Cette étude vise à développer un modèle statistique permettant de construire des scénarios de prévisions de précipitations qui puissent aider les prévisionnistes et les gestionnaires en hydrologie à raffiner leurs niveaux de décisions selon leurs besoins (évaluation des dépassements des seuils d’alerte, analyse des scénarios temporels futurs, etc.), et leur permettre d’aller ainsi bien au-delà du binôme oui/non découlant d’un seul scénario de PQP.

Plus particulièrement, l’objectif est de développer et évaluer une procédure permettant de produire des faisceaux de prévisions (prévisions d’ensemble) à partir des prévisions à scénario unique (prévisions déterministes) pour rendre disponible un jeu (d’une cinquantaine) de prévisions des précipitations à chaque pas de temps. Ces prévisions seront utilisées en forçage à un modèle hydrologique dans le but de pouvoir améliorer la prévision hydrologique et apprécier l’incertitude de la prévision émise, due aux incertitudes des précipitations prévues.

La figure 2 montre un schéma général d’un système de prévision hydrologique avec indication des parties du système qui sont concernées par cette étude. Le travail est basé sur les étapes suivantes :

- développer une méthode de perturbation (habillage ou *Ensemble dressing* en anglais) des prévisions météorologiques de précipitations à scénario unique pour construire un ensemble de scénarios probables, en tenant compte des erreurs passées : construction d’un modèle statistique des erreurs passées de la prévision déterministe ;
- utiliser les différents scénarios construits en entrée d’un modèle hydrologique afin de générer une prévision de débits à plusieurs scénarios : construction d’une prévision hydrologique d’ensemble ;
- évaluer et quantifier l’apport de la méthode développée à l’aide de scores d’évaluation de la qualité des prévisions : évaluation de la performance des prévisions pour une meilleure détection des situations de crues.

1.3. Organisation du rapport

Le chapitre suivant (chapitre 2) présente une revue de la littérature traitant du sujet au cœur de cette étude : la construction des prévisions d'ensemble, la modélisation des erreurs de prévision et leurs rôles pour évaluer l'incertitude prédictive.

Le chapitre 3 décrit les étapes méthodologiques adoptées pour la génération des faisceaux de prévisions des pluies à partir d'un seul scénario et leur utilisation pour la prévision de débits. Les principes d'évaluation de la qualité des prévisions y sont également présentés. Ensuite, une description des données utilisées est fournie dans le chapitre 4.

Les résultats de la construction et du calage du modèle d'erreur des prévisions de pluies sont présentés et discutés dans le chapitre 5, ainsi que ceux de l'évaluation des prévisions générées (en pluies et en débits).

Des conclusions sont tirées et des recommandations sont données dans le chapitre 6.

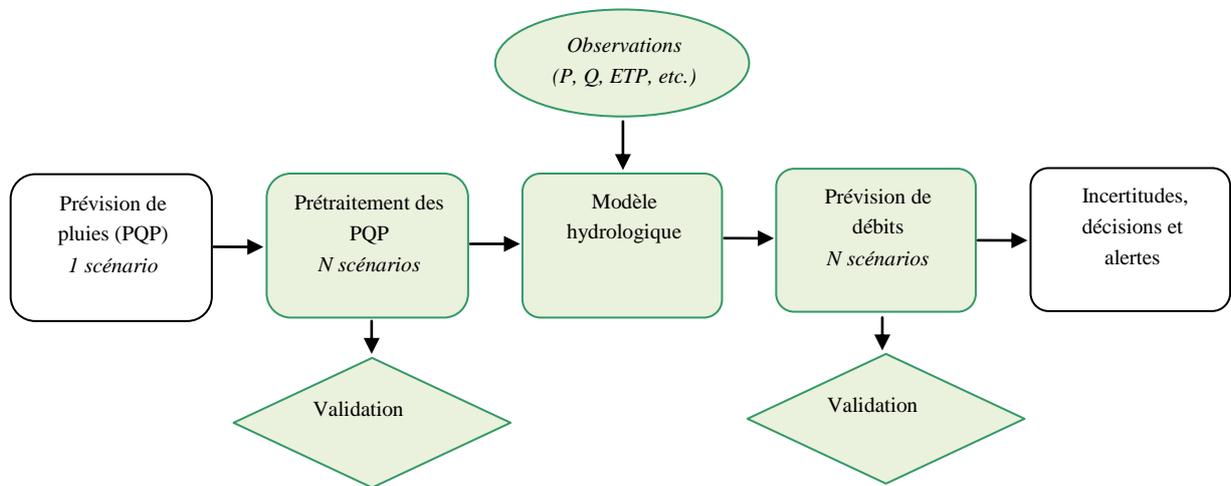


Figure 2. Schéma générale de la prévision hydrologique avec indication des parties du système qui sont concernées par cette étude (cases colorées).

Chapitre 2 : Comment construire et évaluer des faisceaux de prévisions ?

*“A forecast is just the set of probabilities attached to a set of future events...”
Baruch Fischhoff (psychologue américain)*

Ce chapitre offre une synthèse bibliographique sur les différentes techniques envisageables pour la génération d'un faisceau de prévisions ainsi que les critères d'analyse et de jugement de leur qualité. Les aspects méthodologiques et les conclusions des principales références citées dans ce qui suit sont brièvement synthétisés dans le tableau 1.

2.1. Approches de modélisation : vers une prévision d'ensemble hydro-météorologique

L'objectif d'une prévision d'ensemble consiste à prendre en compte les principales sources d'incertitudes liées à la prévision afin de fournir un faisceau de situations futures possibles et pouvoir évaluer un risque éventuel. Il existe plusieurs approches, dans la littérature, permettant la construction des prévisions hydrologiques d'ensemble. Les plus connues d'entre elles sont :

L'approche multi-modèle

L'idée est de se focaliser sur les incertitudes dues à la structure des modèles de prévision et aux différentes méthodes d'assimilation des données (mise à jour du modèle). En effet, des modèles différents utilisent rarement les mêmes méthodes d'assimilation de données et encore moins les mêmes stratégies de modélisation, ce qui explique la dissemblance des scénarios émis. L'approche multi-modèle consiste donc à utiliser plusieurs modèles de prévision au même temps pour construire un ensemble de prévisions déterministes provenant des sorties de différents modèles.

Appelée aussi la prévision d'ensemble « du pauvre » (*Poor man's ensemble* en anglais), les inconvénients de cette méthode sont : le nombre souvent très limité de modèles à disposition pour une interprétation probabiliste, l'impossibilité d'estimer la capacité de prévision de l'ensemble et les erreurs des modèles, ainsi que le fait que tous les modèles peuvent avoir des « mauvaises » prévisions au même temps s'ils partent du même état initial (Thirel, 2009).

La méthode des analogues

C'est le principe du « déjà vu » partant de l'hypothèse que deux situations météorologiques similaires doivent engendrer des effets locaux semblables. Autrement dit, les situations synoptiques à prévoir se sont déjà produites dans le passé sous une forme plus ou moins voisine. De ce fait, si le prévisionniste dispose de longues archives météorologiques, la recherche des situations analogues (rechercher dans le passé des pluies observées en présence de champs de pression semblables ; Garçon *et al.*, 2008) peut former un ensemble de scénarios probables à utiliser comme forçage d'un modèle hydrologique. Le nombre d'analogues sélectionnés détermine la taille de cet ensemble.

Le principal inconvénient de cette méthode est la perte de la cohérence de la dynamique temporelle des événements lors de la construction de scénarios couvrant différentes échéances, puisque les analogues sélectionnés sont indépendants dans le temps (Ben Daoud, 2010).

Couplage avec un modèle de prévision d'ensemble météorologique

Afin de pallier à l'inexactitude des prévisions et aux effets du comportement chaotique de l'atmosphère, notamment pour les échéances élevées (au-delà de 3 jours), les communautés météorologiques ont eu recours aux techniques de prévision à plusieurs scénarios ou encore « les Systèmes de Prévision d'Ensemble » (SPE). Les sorties de ces modèles « ensemblistes » ne peuvent ni s'interpréter comme un ensemble de prévisions déterministes, en isolant chaque scénario (membre ou « *run* » de la prévision), ni de la même manière qu'une prévision déterministe. L'ensemble doit être considéré dans sa totalité. Un prévisionniste qui dispose d'une prévision d'ensemble de précipitation peut l'appliquer directement en entrée à un modèle hydrologique de prévision pour construire ses scénarios de débits. Cette approche consiste à simuler la réponse hydrologique du système considéré aux différents scénarios d'une prévision d'ensemble issue d'un ou plusieurs modèles de Prévision Numérique du Temps (PNT).

Aujourd'hui, les prévisions météorologiques d'ensemble sont largement répandues auprès de nombreux centres de prévisions opérationnelles : par exemple, des prévisions d'ensemble météorologiques quotidiennes, assorties d'un indice de confiance et de produits probabilistes associés, sont émises depuis 1992 au CEPMMT et au NCEP (*National Center for Environmental Prediction*, Etats-Unis), depuis 1998 à Météo-France et depuis 2005 au CMC (Centre Météorologique Canadien). En ne considérant que l'initialisation des modèles comme source d'incertitude, différentes approches sont utilisées pour créer les variantes des états initiaux en entrées aux modèles PNT, à savoir :

- la méthode des Vecteurs Singuliers utilisée par le système de prévision d'ensemble du CEPMMT (EPS, pour *Ensemble Prediction System* en anglais) et par Météo-France (PEARP, Prévision d'Ensemble ARPege),
- la méthode des erreurs croissantes (BGM pour *Breeding of Growing Modes* en anglais), choisie par le NCEP pour la génération des ensembles,
- la méthode du filtre de Kalman (EnKF pour *Ensemble Kalman Filter* en anglais) utilisée par le CMC.

Le lecteur pourra se référer au résumé proposé par Thirel (2009) pour plus de détails.

Les sorties des modèles météorologiques d'ensemble doivent le plus souvent subir des prétraitements avant leurs utilisations dans un système de prévision hydrologique ; des désagrégations et/ou des réductions d'échelle sont généralement nécessaires. Plusieurs systèmes de prévision hydrologiques intégrant les prévisions météorologiques d'ensemble existent dans la littérature (voir la synthèse proposée par Cloke et Pappenberger, 2009). Ils font l'objet d'une variété d'études concernant, entre autres, l'évaluation de la qualité des prévisions, le couplage à d'autres sources d'incertitudes hydrologiques et la communication des prévisions probabilistes, à travers plusieurs initiatives et projets internationaux comme HEPEX (*Hydrological Ensemble Prediction EXperiment*) (Schaaque *et al.*, 2007), COST731 (*Propagation of Uncertainty in Advanced Meteo-Hydrological Forecast Systems, COoperation in Science an Technology*⁵), MAP-DPHASE (*Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region*⁶) ou EFAS (*European Flood Alert System*⁷).

5 <http://www.cost.esf.org/>

6 <http://www.map.meteoswiss.ch/d-phase>

7 <http://efas.jrc.it/>

Traitement statistique des prévisions déterministes

Dans le cas où le prévisionniste ne dispose que d'une prévision déterministe de précipitation et d'un seul modèle hydrologique, il peut faire recours à des techniques de prétraitement statistique permettant d'inclure une information sur l'incertitude à ce scénario unique (Shrestha *et al.*, 2011). Plusieurs méthodes ont été développées, y compris, les techniques de bruitage (ou encore habillage ou *dressing* en anglais). A l'aide d'une analyse critique de la prévision déterministe, sur la base de ses performances passées, en comparant les prévisions aux valeurs réellement observées sur le bassin ou bien à leurs répartitions dans le temps et leurs positionnements dans l'espace, on peut quantifier les incertitudes du système en question et attacher ces incertitudes aux prévisions.

L'habillage d'une prévision déterministe consiste à perturber la prévision actuelle avec les erreurs des prévisions passées, à partir de leur fonction de distribution de probabilité (FDP). Ainsi, un ensemble de prévision de débits peut être généré en forçant le modèle hydrologique avec une collection de scénarios probables des précipitations futures corrigés de leurs erreurs passées ou encore en attachant des erreurs passées à une prévision déterministe de débits.

Récemment, plusieurs études ont été menées dans le but de quantifier les erreurs des modèles de prévision hydrométéorologique (par exemple, Schaake *et al.*, 2007 ; England *et al.*, 2010 ; Vincendon *et al.*, 2010 ; Wu *et al.*, 2011) et de simulation hydrologique (par exemple, Schaeffli *et al.*, 2007 ; Yang *et al.*, 2007). La plupart de ces méthodes partagent une approche générale pour traiter les prévisions et les observations : des transformations des données sont généralement nécessaires afin de s'assurer que les erreurs étudiées peuvent être décrites à l'aide d'une distribution statistique donnée (par exemple, gaussienne, gaussienne mixte, etc.).

Dans cette étude, nous cherchons à construire et tester la validité d'un modèle statistique des erreurs des prévisions déterministes de précipitation pour générer une meilleure prévision hydrologique d'ensemble. Cette approche présente l'avantage de traiter, simultanément, toutes les sources d'incertitudes d'un système de PQP et d'en tenir compte pour corriger les biais des prévisions déterministes selon plusieurs scénarios (Wu *et al.*, 2011). Néanmoins, il est nécessaire de vérifier plusieurs hypothèses lors du traitement des données (cf. chapitre 3), ce qui n'est pas toujours évident et peut entraîner, dans le cas de faiblesse d'une hypothèse, des résultats biaisés et difficiles à interpréter (England *et al.*, 2010).

Les approches mixtes

Il est également possible de construire des ensembles de prévision en combinant deux ou plusieurs méthodes citées ci-dessus. Par exemple, le centre de prévision opérationnelle de l'Electricité de France (EDF) à Grenoble développe des approches mixtes de prévisions d'ensemble basées sur l'analyse des sorties du modèle ARPEGE de Météo-France, ainsi que celles du modèle statistique ANALOG, qui fournit les hauteurs de précipitations observées, dans le passé, par des situations météorologiques proches aux situations prévues, et celles du système EPS du CEPMMT (Garçon *et al.*, 2008).

Une autre expérience est celle du projet TIGGE (pour *THORPEX*⁸ *Interactive Grand Global Ensemble* en anglais) de l'Organisation Mondiale de la Météorologie (OMM) rassemblant les prévisions d'ensemble de plusieurs centres de prévision météorologique à travers le monde, dont le centre américain, britannique, européen, canadien, japonais, etc. Ce projet fait appel aux

⁸ THORPEX (THE Observing system Research and PRedictability EXperiment) est un programme de recherche de l'OMM pour accélérer l'amélioration de la précision des prévisions météorologiques (de 1 jour à 2 semaines) à fort impact pour le bénéfice de l'humanité (http://www.wmo.int/pages/prog/arep/wwrp/new/thorpex_new.html)

différents modèles de prévision d'ensemble pour former le « Grand Ensemble » de prévision météorologique. Utilisant des prévisions en provenance de différents centres, la base de données TIGGE permet de capturer les incertitudes dues à l'initialisation, la paramétrisation, la structure et la méthode d'assimilation de chaque modèle (cf. la discussion de Hagedorn *et al.*, 2010 pour plus de détails).

2.2. Evaluation des prévisions : comment reconnaître une bonne prévision ?

Dans l'évaluation des performances d'un système de prévision, deux points de vue différents apparaissent : celui de l'utilisateur et celui du prévisionniste.

L'utilisateur de la prévision se soucie généralement de sa valeur. Reliée à une décision basée sur cette prévision, la valeur dépend donc de son utilité en termes du coût économique, c'est-à-dire des gains et des pertes occasionnés par cette décision (Weijs *et al.*, 2010). Autrement dit, l'utilisateur se pose la question : combien la prévision rapporte effectivement ?

En tant que prévisionniste, c'est plutôt la qualité des prévisions émises qui se pose comme question dans l'évaluation d'un système de prévision. En effet, une fois que l'on dispose de scénarios *a priori* possibles pour un événement donné (un ou plusieurs scénarios), la qualité de la prévision dépend de son adéquation avec la réalité (l'observation *a posteriori*) (Hingray *et al.*, 2009). Dans le cadre de notre travail, c'est cet aspect qui sera considéré lors de l'évaluation des prévisions.

Une prévision déterministe ayant un seul scénario est souvent plus facile à évaluer qu'une prévision d'ensemble : soit l'événement prévu a été observé et la prévision est correcte, soit il ne l'a pas été et la prévision est erronée. En revanche, une prévision d'ensemble est sensée donner en plus une information sur le niveau d'incertitude et le risque de se tromper (par exemple sous la forme d'une probabilité : il y a 40% de chances de pleuvoir demain) ce qui rend son évaluation beaucoup plus complexe (par exemple, s'il pleut, la prévision ci-dessus est correcte et s'il ne pleut pas, aussi, car il y avait 60% de chances de ne pas pleuvoir !). L'évaluation des prévisions probabilistes requiert alors l'utilisation de critères statistiques appropriés et, par conséquent, une longue série de probabilités de prévision émises à comparer aux fréquences d'observation.

Des approches statistiques pour la mesure de la qualité des prévisions sont largement répandues aussi bien en météorologie^{9,10} (voir par exemple Rousseau, 2001 ; Atger, 2001 ; Yuan *et al.*, 2005) qu'en hydrologie (par exemple Weijs *et al.*, 2010 ; Randrianasolo *et al.*, 2010 ; Wu *et al.*, 2011) et elles suscitent actuellement beaucoup d'intérêt. Ces techniques sont basées sur le calcul des scores (ou critères) numériques d'évaluation appropriés pour la mesure de chaque « attribut » de la prévision (Wilks, 1995). Les définitions des attributs diffèrent souvent d'un auteur à l'autre. Ici, nous adopterons celles présentées par le centre européen⁹ :

- la fiabilité (*reliability* en anglais) indique dans quelle mesure l'observation est une réalisation de la distribution de probabilité des prévisions d'ensemble, autrement dit, jusqu'à quel point peut-on se fier à un ensemble qui prévoit une probabilité de précipitations de 40 %, par exemple ? En examinant ce qui s'est produit dans le passé avec toutes les prévisions d'ensemble qui indiquaient une probabilité de précipitations de 40 % on peut établir une mesure de cette fiabilité ;

⁹ Le CEPMMT http://www.ecmwf.int/products/forecasts/guide/The_verification_of_ECMWF_forecasts.html

¹⁰ Le Centre Australien <http://www.cawcr.gov.au/projects/verification/>

- la résolution (*resolution* en anglais) mesure, pour une fiabilité donnée, la dispersion des membres de la prévision indépendamment de l'observation. Autrement dit, c'est une mesure de la largeur du faisceau de la prévision d'ensemble (finesse ou *sharpness* en anglais). En effet, si l'ensemble des résultats divergent beaucoup (un très large faisceau de prévision), la confiance dans la prévision sera faible, puisque tous les scénarios sont possibles. Inversement, si toutes les prévisions convergent vers la même solution, la probabilité de réalisation est forte et, par conséquent, la prévision a une finesse parfaite (cas d'une prévision déterministe) ;
- la justesse (*accuracy* en anglais) mesure le degré moyen de correspondance entre la prévision et l'observation. En effet, l'observation *a posteriori* représente toujours la meilleure référence à laquelle on peut comparer une prévision. Or, dans le cas des prévisions d'ensemble il ne s'agit plus d'apprécier la correspondance entre un binôme prévu-observé mais plutôt plusieurs scénarios de prévisions possibles. Par conséquent, il faut chercher une valeur caractéristique pouvant représenter au mieux l'ensemble de prévision, par exemple la médiane (ou la moyenne) des membres. Dans ce cas, la justesse d'une prévision d'ensemble augmente si l'écart entre la valeur observé et la médiane (ou la moyenne) des membres diminue (Bontron, 2004).

La qualité des prévisions d'ensemble peut être aussi jugée par rapport au dépassement ou du non-dépassement d'un seuil critique, en évaluant les décisions binaires (oui/non, cf. §3.3.2) qui en découlent.

Enfin, la performance d'un système de prévision peut aussi être évaluée en le comparant à un système de prévision de référence, d'où la notion d'habileté ou de performance des prévisions (*skill* en anglais). Une prévision est habile si elle présente une valeur ajoutée en termes de succès et de capacité de prévision par rapport à une prévision non habile (tels que la climatologie ou la persistance).

Tableau 1. Quelques exemples d'études de cas récentes concernant les prévisions d'ensemble hydrologiques

<i>Objet de l'étude</i>	<i>Auteur</i>	<i>Méthode</i>	<i>Bassin et modèle</i>	<i>Principaux résultats</i>
Couplage avec les modèles de prévision d'ensemble météorologique	Randrianasolo <i>et al.</i> , 2010	- Evaluation des prévisions d'ensemble de débits de deux modèles hydrologiques différents forcés par des prévisions d'ensemble météorologiques et comparaison de leurs réponses	- PEARP (prévision d'ensemble, Météo-France) - Les modèles hydrologiques SIM et GRP - 211 bassins versants en France	Une bonne performance des deux modèles hydrologiques Amélioration des résultats de prévision du modèle GRP en rajoutant un module d'assimilation des données pour ensemble de prévisions à court terme
Prétraitement des prévisions météorologiques	Vincendon <i>et al.</i> , 2010 Wu <i>et al.</i> , 2011	- Une approche de vérification orientée objet pour le calcul des erreurs de Structure/ Amplitude/Localisation (mesure des ratios SAL) - Génération de l'ensemble via la multiplication des facteurs tirés aléatoirement des FDP des erreurs Habillage d'une prévision déterministe via un tirage aléatoire de la distribution statistique bivariée (modélisée via une transformation de données) "de la prévision sachant l'observation"	- AROME (prévision déterministe, Météo-France) - Prévision d'ensemble AROME-PEARP - Le modèle hydrologique ISBA-TOPMODEL - 3 bassins versants en France - NCEP HPC (prévision déterministe) - 3 bassins en Californie, en Pennsylvanie et au Arkansas et le bassin du Missouri	La vérification sur deux épisodes de crues éclair a montré que l'ensemble généré fonctionne mieux que la prévision déterministe. La performance de la méthode de perturbation a également été vérifiée par rapport à un de prévision d'ensemble météorologique et a montré des résultats beaucoup plus intéressants en termes de prévision des événements de crues éclairées Amélioration de la résolution des prévisions générées par rapport à la prévision déterministe
Approches mixtes	Pappenberger <i>et al.</i> , 2008 Velázquez <i>et al.</i> , 2011	Evaluation de la performance du système d'alerte aux crues européen EFAS, forcé par la base de données multi-modèle TIGGE Comparaison des ensembles formés via : - La réponse de différents modèles hydrologiques à la même prévision déterministe de pluies - La réponse de différents modèles hydrologiques à une prévision d'ensemble de pluies	-Prévision d'ensemble de la base de données TIGGE (7 centres de prévision), - le modèle hydrologique du système EFAS - 1 bassin en Roumanie -Prévisions déterministes et d'ensemble du CEPMMT - 16 modèles hydrologiques - 29 bassins en France	L'évaluation de la méthode pour le cas d'une inondation en Roumanie (10/2007) a montré que cet événement aurait pu être prévu 8 jours avant son occurrence en utilisant la base de données TIGGE, en offrant une marge d'anticipation des risques possibles Les prévisions d'ensemble produites par une combinaison de plusieurs structures de modèles hydrologiques et ensembles météorologiques sont plus fiables et ont plus d'habileté par rapport aux prévisions déduites uniquement par un simple modèle hydrologique alimenté par des prévisions d'ensemble météorologiques et par rapport à l'utilisation de différents modèles hydrologiques forcés par des prévisions météorologiques déterministes

Chapitre 3 : Choix méthodologiques

“Prévoir consiste à projeter dans l’avenir ce qu’on a perçu dans le passé...”
Henri Bergson (1859-1941)

Pour un système de prévision hydrométéorologique, il est souvent constaté que l’incertitude de la prévision des précipitations est une source principale d’erreur des prévisions des débits. Dans ce travail, un modèle statistique des erreurs passées de la prévision déterministe des précipitations a été développé. Ce modèle permet de capturer cette incertitude et de l’utiliser pour « corriger » un scénario de prévision selon plusieurs réalisations possibles. Nos choix méthodologiques, le principe de ce modèle, ainsi que les méthodes d’évaluation de l’approche proposée sont présentés dans ce qui suit.

3.1. Construire un modèle des erreurs passées de la prévision déterministe

Par analogie à une démarche de modélisation classique, ce travail a été développé en suivant les quatre étapes suivantes :

- sélection d’un type de modèle pertinent à la problématique via une analyse causale des erreurs de prévision, appliquée sur l’échantillon de données ;
- construction du modèle en formalisant la structure du problème mathématique et numérique (coder la formulation numérique du modèle) ;
- estimation d’une série de paramètres optimaux par calage ;
- contrôle du modèle par rapport à ses capacités à produire des scénarios de prévisions de bonne qualité, en pluies et en débits.

3.1.1. Quel type de modèle ?

Dans cette étude, nous cherchons à prendre en compte les dysfonctionnements d’un modèle de prévision météorologique, pour ce qui est de ces prévisions des précipitations, sans expliquer particulièrement les sources de ses incertitudes. Le modèle d’erreur ici développé permet donc de prendre en compte les biais systématiques (ou quasi-systématiques) des prévisions des précipitations (sur ou sous-estimation de l’observation).

Pour évaluer les erreurs de prévision, nous disposons d’une archive constituée à la fois de séries chronologiques de prévisions et d’observations sur une période donnée (cf. chapitre 4 : cas d’application). Les erreurs de prévision seront évaluées par l’écart entre les valeurs prévues et observées. Chaque résidu est donné par :

$$\delta = X - Y, \quad \forall (X, Y) \quad \text{Équation 1}$$

Où,

δ : est l’erreur de prévision, X : est la pluie prévue et Y : est la pluie observée.

Il est à noter que les résidus δ contiennent aussi une part d’erreur liée à l’observation, que nous avons supposé négligeable dans ce travail.

Le principe de cette première étape, consistant à choisir quel type de modèle d’erreur adopter, est de structurer le problème mathématique et numérique. Autrement dit, il s’agit de chercher la fonction f décrivant l’échantillon de données dont on dispose en fonction des variables prévues et observées X , Y , du pas de temps Δt , d’un ou plusieurs jeux de paramètres Θ (équation 2), et, ensuite, de développer le code numérique correspondant.

$$\delta = f(X, Y, \Delta t, \Theta \dots)$$

Équation 2

Afin d'éviter une faiblesse de l'un des maillons de la modélisation, et par la suite l'invalidité du modèle pour cause d'hypothèses manquantes ou « violées », notre choix a été précédé par des analyses descriptives et des traitements de l'échantillon de données, complétés par l'examen d'un certain nombre d'hypothèses indispensables pour l'acceptation du modèle. Ces hypothèses ont été vérifiées à travers de tests graphiques et/ou statistiques.

3.1.1.1. Nature des données et hypothèses de modélisation

Le choix du modèle repose sur la nature des données dont on dispose. Nous avons commencé par un diagnostic de l'échantillon, en analysant quatre principales caractéristiques des résidus de prévision (Maidment, 1993) : la saisonnalité, la stabilité de la variance (homoscédasticité), l'auto-corrélation des résidus et la loi de répartition des erreurs.

Il s'agit de répondre aux différentes questions suivantes :

- hypothèse de saisonnalité - les résidus ont-ils un caractère saisonnier ?

Lorsque l'on observe des phénomènes qui se reproduisent dans des périodes régulières (jour, mois, saison, etc.), on parle de « phénomène saisonnier ». Le modèle dépendra de ces fluctuations et il faut en tenir compte (Haan, 1977) ;

- hypothèse d'homoscédasticité - les résidus sont-ils indépendants des prévisions ?

La variance est une mesure de la dispersion d'un échantillon ; le fait que la variance des résidus ne soit pas constante entraîne une baisse de précision dans le processus (la série est croissante et les fluctuations ont plus d'ampleur en fonction de la valeur prévue), ou encore l'hétéroscédasticité des données. Ayant des variances différentes, les résidus peuvent être vus comme provenant de différentes distributions. Deux alternatives peuvent être envisageables : soit appliquer une transformation des données pour stabiliser la variance (voir par exemple Box et Cox, 1964 ; Sakia, 1992) ou bien trouver un modèle approprié pour décrire la variance des résidus. Un tracé des résidus en fonction des valeurs prévues en forme d'entonnoir témoigne de la présence ou non d'une hétéroscédasticité du processus ;

- hypothèse d'auto-corrélation - les résidus sont-ils dépendants les uns des autres ?

L'auto-corrélation signifie qu'il existe des liaisons temporelles entre les résidus. Une auto-corrélation est dite positive lorsque des valeurs successives ont tendance à être proches les unes des autres au lieu de varier de manière aléatoire en fonction du temps. Par conséquent, les coefficients de corrélation vont être statistiquement significatifs (Haan, 1977) et vont décrire le lien entre les erreurs à différents pas de temps successifs. Pour conclure de l'auto-corrélation, des tracés dénommés « auto-corrélogrammes » sont utilisés (tracés de l'auto-corrélation à différents « lags » ou « retards ») ;

- hypothèse de normalité - les résidus suivent-ils une loi normale ?

La loi normale est très utilisée dans le cas du traitement des erreurs de modélisation. En effet, elle permet de se ramener à des distributions symétriques et de se débarrasser des valeurs extrêmes ou aberrantes pouvant avoir une très grande importance sur le résultat (cf. §3.2). En revanche, la normalité est difficilement tenable dans le cas des séries temporelles en hydrométéorologie. Pour cela, plusieurs traitements sont disponibles dans la littérature pour améliorer la normalité des données tels que les transformations des données par la racine carrée ou le logarithme, la transformation NQT (*Normal Quantile Transformation* en anglais) ou bien les familles de transformations de puissance (Osborne, 2010).

3.1.1.2. Choix d'une transformation adéquate des données

Dans ce contexte de traitement statistique des données, les transformations « normalisatrices » les plus communément répandues, sont la NQT et la transformation de Box-Cox. Engeland *et al.* (2010) ont testé l'effet de ces deux transformations pour modéliser, suivant une loi normale, les erreurs des prévisions des débits dans le bassin Langvatn au nord de la Norvège et ils ont montré que les deux méthodes présentent des résultats comparables.

L'avantage de la transformation de Box-Cox (Box et Cox, 1964) est qu'elle est plus générique, car paramétrable. Elle offre un degré de liberté pour se rapprocher de la distribution normale et traiter les problèmes d'hétéroscédasticité de la distribution des résidus à travers l'optimisation de son paramètre λ_1 (équation 3).

En effet, pour les $\lambda_1 < 1$, la transformation a pour effet de ramener à une amplitude constante des variations qui croissent avec le niveau moyen de la tendance générale. Plus λ_1 est proche de 0, plus cet effet est important. Pour les $\lambda_1 > 1$, l'effet est opposé et la transformation est alors plus adaptée aux séries dont les fluctuations croissent quand la série elle-même décroît (cf. tableau 2).

Cette transformation a été aussi utilisée par Thyer *et al.* (2002) et Yang *et al.* (2007) et a montré ses capacités à homogénéiser la variance et améliorer la normalité des données.

Toutes ces raisons font que nous avons décidé de faire recours à la transformation de Box-Cox pour traiter l'échantillon de données dont on dispose. Cette transformation s'écrit :

$$\left\{ \begin{array}{l} BC(Z) = \frac{(Z + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, \text{ si } \lambda_1 \neq 0 \\ BC(Z) = \log(Z + \lambda_2), \text{ si } \lambda_1 = 0 \end{array} \right. \quad \text{Équation 3}$$

Où,

Z : est la variable originale à transformer (dans notre cas, Z sera la pluie prévue ou observée),

$BC(Z)$: est la transformée Box-Cox de la variable Z ,

λ_1 : est le paramètre qui définit le type de la transformation Box-Cox (voir tableau 2)

λ_2 : est une constante choisie arbitrairement de telle sorte que $Z + \lambda_2 > 0$ (λ_2 vise essentiellement à rendre la variable non-nulle pour une « transformation puissance »).

On note par la suite ε les résidus évalués sur les données transformées. Par conséquent, l'équation 1 et l'équation 3 et donnent :

$$\varepsilon = BC(X) - BC(Y) \quad \text{Équation 4}$$

3.1.1.3. La transformation des données est-elle une garantie de la normalité et de l'homoscédasticité ?

L'hypothèse est que, parmi toutes les transformations de Box-Cox, avec différentes valeurs de λ_1 , les données transformées ont la plus forte probabilité d'être normalement distribuées en maximisant la fonction de vraisemblance (ou encore en maximisant le coefficient de corrélation entre la loi théorique et l'échantillon). Mais ceci n'est pas une garantie de normalité. Par conséquent, dans cette étude, la normalité des données a été vérifiée en utilisant un tracé de probabilité et des tests statistiques. De même pour l'hypothèse d'homoscédasticité, l'effet de la transformation a été examiné en fonction de la stabilité de la variance de l'échantillon des résidus (contrôle visuel sur le graphique).

Tableau 2. Effets du paramètre λ_1 de la transformation Box-Cox (d'après Osborne, 2010 et modifié)

<i>Effet de la transformation</i>	λ_1	$T=BC(Z)$
$\lambda_1 < 1$: stabiliser une variance qui croit avec les valeurs prévues	-2	$T^{-2} = \frac{1}{T^2}$
	-1	$T^{-1} = \frac{1}{T^1}$
	-0.5	$T^{-0.5} = \frac{1}{\sqrt{T}}$
	0	Log (T)
	0.5	$T^{0.5} = \sqrt{T}$
$\lambda_1=1$: pas d'effet sur la variable	1	$T^1 = T$
$\lambda_1 > 1$: stabiliser une variance qui décroît avec les valeurs prévues	2	T^2

3.1.1.4. Tester la normalité

Un test statistique de normalité est un outil pratique d'aide à la décision qui permet de trancher entre deux hypothèses dont une seule est vraie. La première est dite « nulle » (notée H_0 : l'échantillon suit une loi normale), et l'autre « alternative » (notée H_1 : l'échantillon ne suit pas une loi normale). La décision conduit à retenir l'une d'entre elles suivant un risque d'erreur associé ; un risque de 1^{ère} espèce noté α et appelé niveau ou seuil de signification du test (refuser H_0 lorsque H_0 est vraie) et un risque de 2^{ème} espèce noté β (accepter H_0 lorsque H_1 est vraie).

Le résultat du test est essentiellement un nombre appelé « p-valeur » ; c'est la probabilité de commettre une erreur de 1^{ère} espèce en rejetant l'hypothèse nulle H_0 . Plus précisément, si on veut un risque d'erreur de α % (par exemple, 5%), on ne rejette pas l'hypothèse nulle H_0 si la p-valeur $> \alpha/100$ (par exemple, p-valeur > 0.05) et on la rejette sinon (Haan, 1977).

Dans cette étude, deux tests ont été considérés : le test de Kolmogorov-Smirnov et le test de Shapiro-Wilk. Le premier est un test d'ajustement non paramétrique qui repose sur la mesure de la plus grande distance entre la fonction de distribution théorique (celle de la loi normale) et la distribution de l'échantillon (Massey, 1951). Le second est un test de normalité qui mesure le coefficient de détermination (le carré du coefficient de corrélation) entre la série des quantiles de la loi normale et les quantiles empiriques obtenus à partir de l'échantillon de données (Shapiro et Wilk, 1965).

En outre, la normalité a été également vérifiée en examinant des graphes dits « quantile-quantile » (ou encore les droites de Henry) permettant une appréciation de l'ajustement d'une distribution de l'échantillon de données au modèle théorique (ici, la loi normale).

3.1.2. Quelles échéances de prévision pour le modèle d'erreur ?

Les modèles de prévision météorologique fournissent des prévisions qui s'étalent sur plusieurs échéances (plusieurs heures ou jours de validité de la prévision). Ces prévisions sont d'autant moins précises que l'échéance est tardive, ce qui ajoute une dimension supplémentaire au

modèle d'erreur. En effet, à chaque instant de prévision t (date à laquelle une prévision est émise) on a :

$$X_t = (X_t^1, X_t^2, \dots, X_t^i, \dots, X_t^e), \quad \forall (i, t) \quad \text{Équation 5}$$

Où,

X_t^i : est la pluie prévue à l'instant t , pour l'échéance i et e : est le nombre total d'échéances de la prévision émise à l'instant t .

Pour tout pas de temps t pour lequel on dispose d'une prévision émise et pour toute échéance i pour laquelle la prévision est valable, l'équation 1 devient :

$$\delta_t^i = X_t^i - Y_t^i, \quad \forall (i, t) \quad \text{Équation 6}$$

Où,

δ_t^i : est l'erreur associée à la prévision émise à l'instant t et pour l'échéance i , X_t^i : est la pluie prévue à l'instant t et pour l'échéance i et Y_t^i : est la pluie observée à la même date de l'échéance i de la prévision émise à l'instant t .

En variables transformées (équation 3 et équation 4), les résidus de prévision (ici appelés, « résidus sur la variable transformée ») seront donnés par :

$$\varepsilon_t = (\varepsilon_t^1, \varepsilon_t^2, \dots, \varepsilon_t^i, \dots, \varepsilon_t^e), \quad \forall t \quad \text{Équation 7}$$

Où,

ε_t^i : est le résidu transformé associé à la prévision émise à l'instant t et pour l'échéance i et e : est le nombre total d'échéances de la prévision.

Les erreurs de prévisions sont alors une matrice de $N_t \times e$ dimensions de l'équation 8 :

$$\varepsilon = \begin{pmatrix} \varepsilon_1^1 & \varepsilon_1^2 & \varepsilon_1^3 & \dots & \varepsilon_1^e \\ \varepsilon_2^1 & \varepsilon_2^2 & \varepsilon_2^3 & \dots & \varepsilon_2^e \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \varepsilon_{N_t}^1 & \varepsilon_{N_t}^2 & \varepsilon_{N_t}^3 & \dots & \varepsilon_{N_t}^e \end{pmatrix}_{(N_t \times e)} \quad \text{Équation 8}$$

Où,

N_t : est le nombre total de prévisions émises (ici, il correspond au nombre de pas de temps de notre archive journalier, c'est-à-dire le nombre de jours de prévision), e : est nombre d'échéances de prévision et ε_t^i : est l'erreur de prévision au pas de temps t et à l'échéance i .

Dans la suite, on notera ε^i le vecteur des résidus sur les variables transformées de prévision des N_t pas de temps à l'échéance i :

$$\varepsilon^i = (\varepsilon_1^i, \varepsilon_2^i, \dots, \varepsilon_t^i, \dots, \varepsilon_{N_t}^i) \quad \text{Équation 9}$$

Généralement, dans le traitement des erreurs des modèles de prévision (pluies ou débits), les auteurs s'intéressent à la première échéance (voir par exemple Engeland *et al.*, 2010).

Dans cette étude, nous avons cherché à construire un modèle performant à toutes les échéances de prévision. Pour ce faire, deux approches sont envisageables :

- traiter chaque échéance à part, en ne considérant donc que l'aspect « univarié » du problème de telle sorte que le modèle dépendra de l'échéance de prévision (des paramètres différents pour chaque échéance),
- considérer l'aspect conjoint ou encore « multivarié », en conservant un pourcentage significatif des relations entre les pluies prévues à différentes échéances. Cette approche consiste à évaluer leurs sens de variations et ainsi qualifier la dépendance de ces variables, deux à deux, à l'aide de la mesure de la « covariance ». Le problème sera donc caractérisé par une matrice de variance-covariance suivante :

$$\Sigma = \text{cov}(\varepsilon) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1e} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2e} \\ \dots & \dots & \dots & \dots \\ \sigma_{e1} & \sigma_{e2} & \dots & \sigma_{ee} \end{pmatrix}_{(e \times e)} \quad \text{Équation 10}$$

Où,

$\sigma_{ij} = \text{cov}(\varepsilon^i, \varepsilon^j) = E[(\varepsilon^i - E[\varepsilon^i])(\varepsilon^j - E[\varepsilon^j])]$ est la covariance de ε^i et ε^j (les vecteurs des résidus de prévision aux échéances i et j , respectivement) avec $E[\]$, l'espérance estimée par la moyenne arithmétique du vecteur dans le cas de la loi normale, et $\sigma_{ii} = \text{cov}(\varepsilon^i, \varepsilon^i) = \text{var}(\varepsilon^i) = E[(\varepsilon^i)^2] - (E[\varepsilon^i])^2$, revient à la variance du vecteur des résidus à une échéance i .

Puisque nous nous intéressons particulièrement à la constitution de faisceaux de prévisions aussi cohérentes que possibles dans le temps, il nous a paru plus pertinent de choisir la deuxième approche : une approche de modélisation statistique multivariée. En effet, d'un point de vue de prévision hydrométéorologique, on s'attend à ce qu'il y ait une certaine cohérence temporelle dans les scénarios des pluies prévues, et à ce que deux échéances successives soient probablement corrélées.

3.1.3. Comment construire le modèle multivarié des résidus ?

Notre archive de prévisions déterministes des précipitations est une matrice de taille $(N_t \times e) =$ (nombre de pas de temps/jours de l'archive \times nombre d'échéances de prévision). A l'issu de l'étape de calage du modèle d'erreur, les résidus ε^i (après transformation Box-Cox des données) suivent une loi normale ayant comme paramètres la moyenne et la variance des résidus. Ces paramètres sont calculés sur les N_t pas de temps et sont notés (μ_i, σ_i^2) .

Par conséquent, les e vecteurs de prévisions suivent une loi jointe dite loi normale multivariée¹¹, définie par un vecteur de moyenne $\mu = (\mu_1, \mu_2, \dots, \mu_e)$, caractérisant la tendance centrale de l'échantillon, et une matrice de covariance Σ (celle donnée par l'équation 10), mesurant la dispersion des données.

¹¹ X est de loi normale k-variée si et seulement si les X_i ($i = 1, \dots, k$) suivent chacun une loi normale

La fonction de densité des résidus est donnée par :

$$f(\varepsilon^i) = 2\pi^{-\frac{N_i}{2}} [\det(\Sigma^{-1})]^{-1/2} \exp\left[-\frac{1}{2}(\varepsilon^i - \mu_i)\Sigma^{-1}(\varepsilon^i - \mu_i)'\right] \quad \text{Équation 11}$$

Où,

ε est l'erreur transformée de la prévision déterministe qui suit une loi normale multivariée :

$\varepsilon \sim N_e(\mu, \Sigma)$ et N_i est la dimension du vecteur ε^i et $\det(\)$ est le déterminant.

L'inconvénient de l'utilisation de la covariance dans la modélisation multivariée est lié au fait qu'elle dépend fortement de l'unité des variables (Haan, 1977). Par conséquent, il paraît souvent pertinent de faire appel à des grandeurs standardisées, permettant de prendre en compte, d'une façon relativement uniforme, tous les ordres de grandeur. Une distribution multivariée centrée réduite est caractérisée par une moyenne nulle et une matrice de covariance standardisée, ou encore une matrice de corrélation (équation 12) tel que $\frac{\varepsilon - \mu}{\sigma} \sim N_e(0, C_e)$.

$$C_e = \text{corr}(\varepsilon) = \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_1\sigma_2} & \dots & \frac{\sigma_{1e}}{\sigma_1\sigma_e} \\ \frac{\sigma_{21}}{\sigma_1\sigma_2} & 1 & \dots & \frac{\sigma_{2e}}{\sigma_2\sigma_e} \\ \dots & \dots & 1 & \dots \\ \frac{\sigma_{e1}}{\sigma_1\sigma_e} & \frac{\sigma_{e2}}{\sigma_2\sigma_e} & \dots & 1 \end{pmatrix}_{(e \times e)} \quad \text{Équation 12}$$

Où,

σ_i est l'écart-type du vecteur ε^i et σ_{ij} est la covariance des vecteurs ε^i et ε^j

Dans le modèle multivarié des résidus, ainsi définit, la valeur simulée ε prendra en compte, simultanément, l'association entre les différentes échéances. Les liens entre chacune d'entre-elles et ε sont estimés par les coefficients de corrélation de la matrice de l'équation 12.

Pour la construction du modèle d'erreur (en gros, l'estimation de la matrice de corrélation), nous avons adopté une procédure classique utilisée en modélisation hydrologique, la validation croisée (*Slipt sample test* en anglais ; Klêmes, 1986 cité dans Hingray *et al.*, 2009). Il s'agit d'utiliser un premier échantillon des données pour estimer les valeurs des paramètres du modèle qui permettent de rendre compte au mieux des observations, appelé échantillon de calage du modèle, et ensuite d'appliquer les paramètres ainsi calés sur un nouvel échantillon de données, indépendantes des premières, appelé échantillon de contrôle du modèle.

Dans notre étude, les performances en calage dépendent de l'ajustement de la loi normale à chaque échéance de prévision, ainsi que de la vérification des autres hypothèses de modélisation (cf. §3.1.1.1). En contrôle, le modèle a été évalué par rapport à ses qualités prédictives, c'est-à-dire par rapport à sa capacité à produire des scénarios de pluies « meilleurs » que la prévision déterministe et ainsi améliorer les prévisions des débits.

3.2. Génération des faisceaux des prévisions des pluies

A partir des simulations du modèle des erreurs passées de la prévision déterministe, décrit ci-dessus, nous avons pu « habiller » la prévision déterministe par ses erreurs afin de construire

une prévision d'ensemble de précipitations résultant des perturbations (ou encore des « corrections ») de ce scénario.

Pour chaque bassin versant étudié, et pour chaque pas de temps de la période de contrôle, nous avons :

- échantillonné aléatoirement de la loi de probabilité (FDP) multivariée calée, M vecteurs avec des valeurs d'erreurs pour l'ensemble des échéances de prévision traitées ;
- corrigé le scénario de prévision déterministe avec l'addition des M valeurs tirées.

3.2.1. Tirage aléatoire d'une loi de probabilité multivariée

Le principe est de simuler un échantillon d'erreurs tel que les valeurs pourront représenter un ensemble de scénarios probables pour corriger la prévision déterministe. Cet échantillon peut être probabiliste, si obtenu par un tirage aléatoire d'un nombre suffisamment grand d'éléments dans une population, ou non-probabiliste, si le tirage est effectué sur un sous-ensemble de la population.

La notion de précision ou de fiabilité d'échantillonnage probabiliste est matérialisée par un seuil de confiance (noté $1-\alpha$) et une marge d'erreur (notée α). En effet, on sait que pour une loi normale univariée (une gaussienne) de paramètres μ et σ^2 , une variable normale a $(1-\alpha).100\%$ de chance, ou de probabilité, d'être située dans l'intervalle :

$$\left[\mu - \frac{z_{\alpha/2} \times \sigma}{\sqrt{n}}, \mu + \frac{z_{\alpha/2} \times \sigma}{\sqrt{n}} \right] \quad \text{Équation 13}$$

Où,

n représente la taille de l'échantillon de données, et $z_{\alpha/2}$ est une constante qui dépend du degré de confiance α souhaité.

Ce segment, dit «intervalle de confiance à $(1-\alpha).100\%$ », recouvre la valeur moyenne μ , en considérant essentiellement les valeurs proches de cette moyenne ($\pm z_{\alpha/2} \times \sigma / \sqrt{n}$) et en écartant les valeurs extrêmes des queues de la distribution (Maidment, 1993).

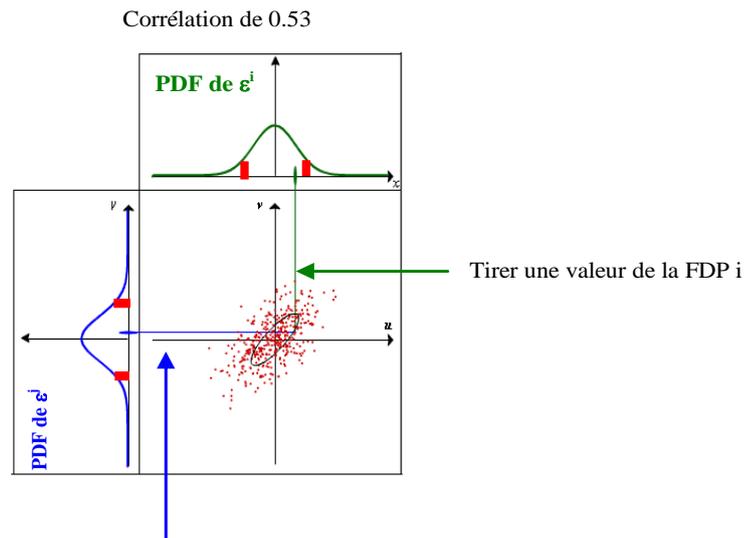
Par analogie, pour une distribution bivariée (si on considère deux échéances, $e=2$), la FDP de la loi jointe est une ellipse dont la forme dépend du coefficient de corrélation (cf. figure 3) et l'intervalle de confiance sera plutôt un carré, puisqu'on étudie simultanément l'action des deux échéances, en tenant compte de l'effet d'une sur l'autre. La loi multivariée est une généralisation du cas bivariée avec un polyèdre de confiance de 2^e-2 faces.

Dans ce travail, nous avons adopté une méthode de tirage dite « non probabiliste » : il s'agit d'effectuer le tirage aléatoire multivarié dans un intervalle de confiance choisi (on s'intéresse ici à l'intervalle à 80%). Une illustration schématique simplifiée du principe de ce tirage est donnée dans la figure 3: on échantillonne sur l'ensemble de la distribution, mais on ne garde que les valeurs qui se trouvent dans l'intervalle de confiance choisi (ceci pour toutes les variables considérées dans la loi multivariée).

Cette notion de tirage dite « non probabiliste » a été appliquée suite aux premiers tests de tirage qui ont montré qu'on peut aboutir parfois à des valeurs trop fortes des précipitations, au delà des limites acceptables des précipitations observées dans les bassins versants de l'étude, puisque nous adoptons une transformation des données de type puissance (transformation Box-Cox). Ceci posait un problème puisque nous nous intéressons à faire passer les scénarios de précipitations dans un modèle hydrologique. L'échantillonnage non-probabiliste nous a permis de se débarrasser de ce problème.

Pour chaque jour de la période de contrôle, le déroulement du tirage aléatoire est alors le suivant :

1. on tire un vecteur de e valeurs de la PDF multivariée, avec $e =$ nombre d'échéances,
2. pour chaque échéance, on vérifie que la valeur tirée est dans l'intervalle de confiance à 80% de la PDF univarié de cette échéance,
3. si ceci est vrai pour toutes les échéances, on garde le tirage et on passe à un deuxième tirage (retour à 1)
4. on répète la procédure M fois, M étant le nombre de scénarios que l'on souhaite tirer.



Tirer une valeur de la FDP j qui soit corrélée de 53% à celle tirée de la FDP i et dans un intervalle de confiance

Figure 3. Illustration simplifiée du tirage aléatoire d'une loi binormale (si $e=2$).

L'ellipse noire représente la zone la plus probable, les points rouges représentent les valeurs tirées en respectant la corrélation entre les deux échantillons. Les bornes rouges sur les FDP délimitent l'intervalle de confiance considéré (seules les valeurs qui se trouvent à l'intérieur de ces intervalles, pour toutes variables, sont prises en compte) (d'après <http://www.aiaccess.net/> et modifiée).

3.2.2. Construction des scénarios de prévision

Le modèle statistique des erreurs passées de la prévision déterministe est un modèle additif normal multivariée, construit via la transformation Box-Cox des données. Le principe de calage a été décrit dans les paragraphes précédents. Il a été mené de telle sorte que les résidus de prévision (évalués sur des données transformées) suivent une loi normale multivariée entre les échéances de prévision.

A partir de ce modèle, nous avons pu construire, à chaque instant t , M membres (ou scénarios) de prévision, à e échéances, et nous avons obtenu ce que nous appellerons ici DH , une prévision «Déterministe Habillée».

En effet, dans la période de contrôle, à chaque instant t , les différents membres de la prévision DH ont été générés par addition, à la prévision déterministe X du jour correspondant à l'instant t , des erreurs tirées aléatoirement de la FDP multivariée (vecteur d'erreurs de taille correspondant aux e échéances considérées) (équation 14).

$$BC(DH_t^i) = BC(X_t^i) + \begin{pmatrix} \varepsilon_{t,1}^i(\Theta) \\ \varepsilon_{t,1}^1(\Theta) \\ \dots \\ \varepsilon_{t,M}^i(\Theta) \end{pmatrix} = \begin{pmatrix} DH_{t,1}^i(\Theta) \\ DH_{t,2}^i(\Theta) \\ \dots \\ DH_{t,M}^i(\Theta) \end{pmatrix} \quad \text{Équation 14}$$

Afin d'obtenir des prévisions avec la même propriété du domaine original (non transformé), il a fallu que les prévisions l'équation 14 subissent une transformation inverse Box-Cox (BC^{-1}) :

$$DH_{t,m}^i(\Theta) = BC^{-1}(BC(Y_t^i) + \varepsilon_{t,m}^i); \forall m = 1 : M, \forall i = 1 : e \quad \text{Équation 15}$$

3.3. Contrôle du modèle d'erreur

Le contrôle du modèle d'erreur revient à vérifier les scénarios de prévision construits pour conclure du « succès » de l'approche mise en œuvre. Il s'agit essentiellement de :

- évaluer la qualité des faisceaux des prévisions : quelle est la précision et l'habileté des prévisions ? Quel est la capacité des prévisions à détecter les dépassements des seuils critiques ? Est-ce que le faisceau enveloppe l'observation ? Avec quelle fréquence ? (mesure de la fiabilité). Est-ce que la dispersion des faisceaux est étroite ? (mesure de la finesse). Quel est l'écart moyen entre les valeurs moyennes des prévisions et les valeurs observées sur l'ensemble de la période de contrôle ? (mesure de la justesse) ;
- comparer la qualité des prévisions des pluies générées (prévisions DH) avec d'autres systèmes de prévision : dans quelle mesure cette approche donne des meilleures prévisions qu'un autre un système de prévision (déterministe ou ensembliste) ?
- conclure de l'apport de cette méthode pour un système de prévision hydrologique : les prévisions DH générées, quand utilisées en entrée à un modèle de prévision de débits, présentent-elles une aide aux gestionnaires dans une optique de prise de décision face aux situations d'alerte aux crues ? Quelle est la qualité des prévisions de débits basées sur les prévisions DH ?

Dans cette partie, sont décrits les principaux critères (ou scores) utilisés pour l'évaluation de la qualité des prévisions : des critères quadratiques, des critères liés à des tables de contingence (capacité à prévoir des événements définis) et des critères graphiques. Nous présentons ceux que nous avons jugés suffisamment démonstratifs pour en tirer des conclusions sur la qualité des prévisions.

3.3.1. Les critères quadratiques

C'est la famille de critères la plus utilisée en hydrologie. Ces critères donnent un poids relativement plus important aux forts écarts et est notamment utile lorsque ceux-ci sont particulièrement indésirables. Dans cette étude, nous avons considéré les critères quadratiques suivants :

L'Ecart Quadratique Moyen (RMSE)

L'Ecart Quadratique Moyen (RMSE pour *Root Mean Square Error* en anglais) représente la moyenne des carrés des erreurs de la prévision par rapport à une référence, en général l'observation. Pour chaque échéance de prévision, les RMSE des prévisions ont été calculés sur

la période de contrôle du modèle en considérant la moyenne des membres de la prévision d'ensemble :

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (m_j - o_j)^2}, \text{ en mm} \quad \text{Équation 16}$$

Où,

o_j est la donnée observée pour le jour j , m_j est la moyenne des prévisions des membres de l'ensemble pour le jour j et N est le nombre de pas de temps utilisés pour le calcul du score.

Le score optimal ($RMSE=0$) est atteint pour une prévision toujours égale à l'observation.

Afin de s'affranchir de l'effet de taille des bassins versants et avoir ainsi des valeurs adimensionnelles permettant de comparer sur les différents bassins entre eux, nous avons calculé le ratio entre le score RMSE et la moyenne journalière des observations (débit observé/pluie observé) de chaque bassin, ou encore le ratio-RMSE donné par:

$$\text{ratio_RMSE} = \frac{RMSE}{\bar{O}} \quad \text{Équation 17}$$

Où,

$RMSE$: est la valeur donnée par l'équation 16 et \bar{O} : est la moyenne des débits et/ou des pluies observés pendant la même période de calcul de la RMSE.

Le Score de Brier (BS)

Le score de Brier (BS pour *Brier Score* en anglais ; Brier, 1950) est une mesure intéressante pour quantifier les performances d'un système de prévision probabiliste, dont les prévisions d'ensemble. Le score BS est similaire au score RMSE, mais dans un espace de probabilités : il mesure la différence entre la probabilité de prévision d'un événement p et son occurrence o , exprimé par la valeur 1 ou 0 selon que l'événement a eu lieu ou non. Il est calculé par rapport à un événement donné, souvent défini par le dépassement d'un seuil critique (dans notre cas les seuils des quantiles 95%, 75%, 50%, 25% de l'observation), et évalue la somme des différences au carrées entre les probabilités de prévision et l'occurrence de l'observation :

$$BS = \frac{1}{N} \sum_{j=1}^N (p_j - o_j)^2 \quad \text{Équation 18}$$

Où,

p_j est la probabilité de prévision de l'évènement (pourcentage des membres de la prévision qui dépassent le seuil critique considéré) pour le jour j , o_j vaut 1 si l'évènement s'est produit au jour j et 0, sinon et N est le nombre de pas de temps utilisés pour calculer le score.

Pour améliorer le score de Brier, il faut augmenter la probabilité de prévision quand l'évènement se produit et la diminuer quand l'évènement ne se produit pas.

Comme le score RMSE, le BS est négativement orienté : plus petite sa valeur, meilleure la performance du système évalué. Le score optimal est nul ($BS=0$). Il est atteint pour une prévision déterministe toujours parfaite, qui associerait à l'évènement prévu une probabilité 0 et 1 toujours égale à l'occurrence de l'observation (aussi 0 ou 1).

3.3.2. Les critères liés à des tables de contingence

La table de contingence est un outil largement utilisé par les gestionnaires dans le cadre des décisions binaires (oui/non). En effet, pour un évènement donné (par exemple le dépassement

d'un seuil critique), quatre cas de figures peuvent se présenter selon que l'évènement est observé ou non et selon qu'il a été prévu ou non. Ces situations sont rassemblées dans la table de contingence de la manière suivante :

	Observé	Non observé
Prévu	Succès (H)	Fausse Alerte (FA)
Non prévu	Manqué (M)	Rejet correct (CR)

Figure 4. Principe de la table de contingence : cas possibles pour un évènement donné en prévision. Chaque case rassemble le nombre d'occurrences du cas correspondant et la somme des quatre cases est égale au nombre de prévisions émises.

A partir de cette table, plusieurs critères peuvent être calculés, à savoir :

- la probabilité de détection (ou taux de succès) $POD = \frac{H}{H + M}$
- le ratio de fausses alertes $FAR = \frac{FA}{FA + H}$
- la probabilité de fausse détection (ou taux de fausse alerte) $POFD = \frac{FA}{FA + CR}$
- l'indice de succès critique $CSI = \frac{H}{H + FA + M}$

3.3.3. Les critères graphiques

Les différents critères numériques d'évaluation peuvent être complétés par une analyse visuelle ou graphique des prévisions, permettant d'étudier un cas donné (un événement de pluie ou un hydrogramme de crue, par exemple) et de voir rapidement le comportement des prévisions par rapport à un comportement espéré (détection d'anomalies ou valeurs aberrantes) ou une référence (qui peut-être l'observation ou bien autres types de prévision telles la climatologie ou la persistance). Dans cette étude, nous avons considéré l'analyse graphique des courbes ROC et la visualisation de quelques épisodes pluvieux marqués, décrits ci-après.

Les courbes ROC

Le tracé du taux de fausses alarmes ($POFD$, en abscisse) en fonction de la probabilité de bonne détection (POD , en ordonnée) dans un graphique appelé courbe de ROC (pour *Relative Operating Curve* en anglais; Mason, 1982 cité dans Atger *et al.*, 2001) permet d'évaluer des décisions binaires (oui/non) en fonction d'un ensemble croissant de probabilités de prévision.

En effet, quand on calcule les cases d'un tableau de contingence à partir d'une prévision d'ensemble (avec M membres ou scénarios de prévision) pour un événement donnée (le dépassement d'un seuil critique), ceci peut être prévu avec différentes probabilités de prévision ; la probabilité de dépasser le seuil critique par au moins $i = 1, 2, 3 \dots$ ou M membres de la prévision d'ensemble. Pour chaque probabilité de dépassement du seuil i , on calcule un tableau de contingence pour déduire les scores $POFD_i$ et POD_i .

Par conséquent, si on a M membres de prévision, l'ensemble des M couples ($POFD$, POD) constitue la courbe ROC. Cette courbe donne une mesure de la capacité du système à détecter les événements observés et à ne pas faire des fausses alertes.

Plusieurs systèmes (ensembliste et/ou déterministe) peuvent être graphiquement comparés par rapport à leur position dans le graphique ROC (Atger, 2001 ; Wu *et al.*, 2011 ; Yuan *et al.*, 2005). Une prévision parfaite donnerait un taux de fausses alarmes nul, alors que le taux de bonnes prévisions serait de 100% ($POD=1$). Ainsi, plus la courbe sera proche du couple (0, 1), c'est-à-dire proche du coin supérieur gauche du graphe, meilleures seront les prévisions.

L'aire sous la courbe ROC est aussi fréquemment utilisée comme un score d'évaluation des prévisions, appelé *AUC* (pour *Area Under the Curve* en anglais). Un $AUC=0.5$ correspond à une courbe ROC confondu avec la 1^{ère} bissectrice et indique que la prévision n'a pas de performance. 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).

Chronologie des épisodes pluvieux marqués

A travers la sélection de quelques épisodes pluvieux les plus marquants de la période de contrôle, les chroniques des pluies et débits prévus ont été qualitativement évalués en comparant les hydrogrammes observés aux prévisions émises à différents instants et pour différentes échéances. Le but est de s'assurer que l'incertitude appliquée aux différents membres est correctement dispersée et que le système de prévision parvient à prévoir des événements (dépassements de seuils) que le scénario unique de la prévision déterministe n'a pas prévus.

Chapitre 4 : Cas d'application **Données, bassins versants et modèle**

“Make things as simple as possible, but not simpler”
Albert Einstein (1879 -1955)

Ce chapitre est une présentation des différentes caractéristiques des données et des outils dont nous disposons pour appliquer la démarche méthodologique de cette étude.

4.1. L'archive de prévisions météorologiques

La méthode d'habillage ici développée et testée repose sur les prévisions météorologiques de précipitations du Centre Européen de Prévision Météorologique à Moyen Terme (CEPMMT) sur la période allant du 11-03-2005 au 01-10-2006.

Il s'agit un système de prévision opérationnel avec un modèle à maille large et variable sur l'ensemble du globe. Les données ici utilisées ont été fournies à l'origine sur une grille de $0.5^{\circ} \times 0.5^{\circ}$ de résolution sur la France (environ 45 km²). Les prévisions sont émises à 12:00 UTC et s'étendent sur 240 heures (10 jours d'échéance).

L'archive utilisée dans ce travail est constituée des quantités de pluies cumulées au pas de temps de 24 h, en commençant à 0h, ce qui a réduit le nombre d'échéances à neuf jours. L'archive contient des précipitations agrégées spatialement pour constituer des pluies de bassin pour la modélisation hydrologique.

Pour la validation du modèle d'erreurs, en plus des observations, on dispose aussi des prévisions d'ensemble (EPS) du modèle IFS du CEPMMT pour dix jours d'échéances (51 membres à chaque jour de prévision). La figure 5 ci-dessous illustre un exemple des scénarios de prévisions émises par les EPS du CEPMMT le 28-02-2007.

Les ensembles du CEPMMT sont basés sur un système d'assimilation¹² de données à 4 dimensions variationnelle (4D-Var) et applique une technique de perturbation de la décomposition des vecteurs singuliers (Persson et Grazzini, 2007). En partant du principe qu'un faible écart dans les états initiaux de l'atmosphère peut générer d'importantes différences dans la prévision, la méthode des vecteurs singuliers recherche les perturbations de l'état initial qui auront le plus d'impact sur la prévision à partir d'une échéance donnée (2 jours pour le CEPMMT). Cette méthode implique une faible dispersion des membres de l'ensemble aux premières échéances du fait qu'elle est appliquée pour être optimale pour la prévision à moyen terme.

Les prévisions du CEPMMT ont été utilisées, dans plusieurs études, comme données de forçage aux modèles hydrologiques. Des exemples de références bibliographiques sont cités dans Velázquez *et al.* (2011).

¹² L'assimilation des données est une mise à jour du modèle via un réajustement de sa trajectoire en utilisant les observations, pour produire une nouvelle trajectoire corrigée.

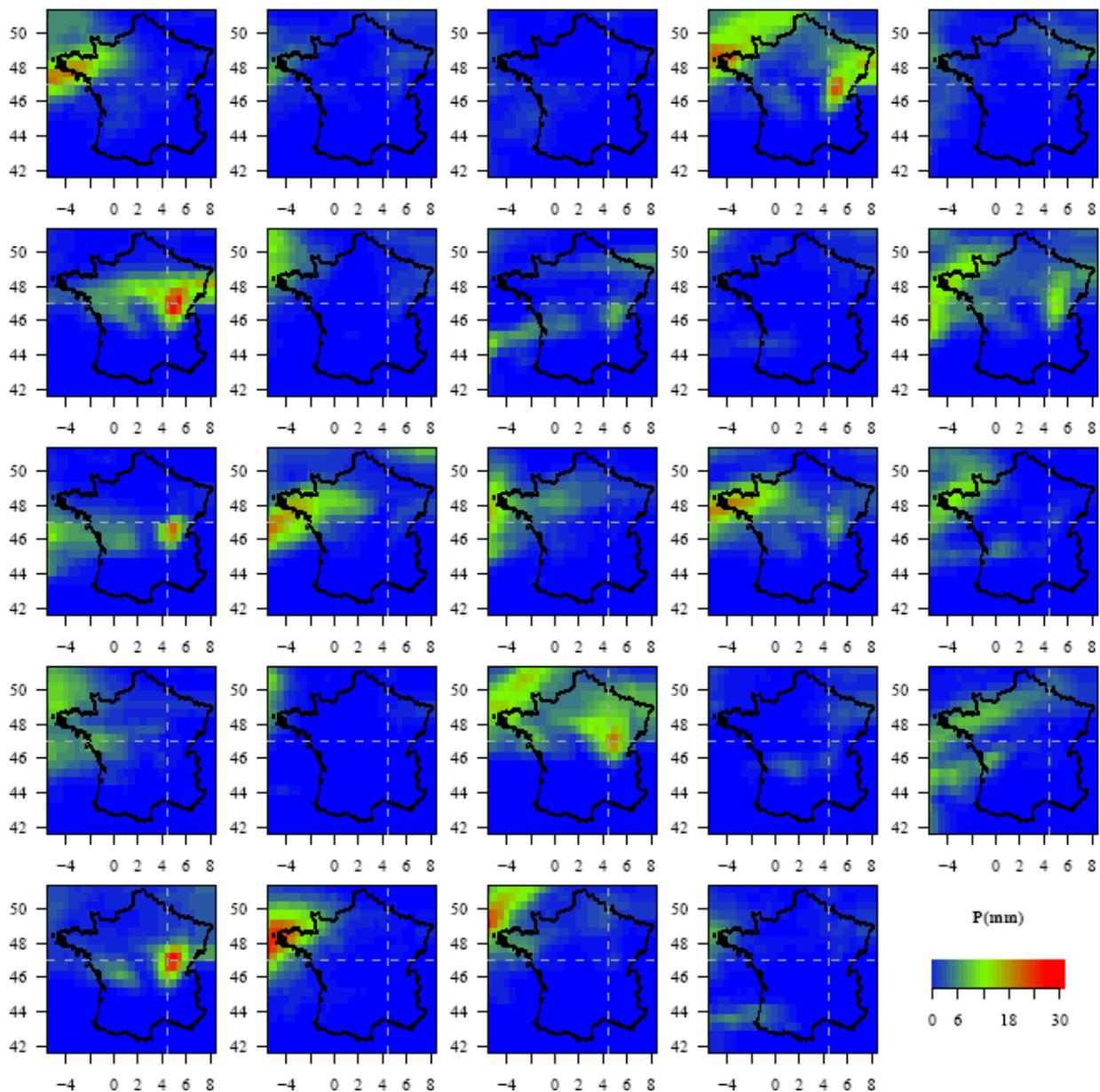


Figure 5. Exemple de 24 membres de la prévision d'ensemble du CEPMMT (EPS) émise le 28-02-2007 à 4 jours d'échéance sur la France.

4.2. Le modèle de prévision hydrologique GRP

Pour la prévision d'ensemble des débits à partir des ensembles générés dans ce travail, le modèle de prévision GRP¹³ développé au Cemagref a été utilisé (figure 6). Ce modèle dérive du modèle de simulation pluie-débit GR4J (Perrin *et al.*, 2001, 2003) et a été spécifiquement adapté pour la prévision des crues en continu et au pas de temps journalier par Tangara (2005). Le modèle a été préalablement calé avec les données climatiques d'environ 52 années¹⁴ d'observations de la base de données du Cemagref (cf. §4.3), pour chaque bassin versant étudié

¹³ http://www.cemagref.fr/webgr/Modeles/GRP/fonctionnement_GRP.htm

¹⁴ Les calculs doivent intégrer une période suffisamment longue, avant toute période d'évaluation, pour pouvoir bien simuler l'état initial du bassin versant comme la résultante des conditions climatiques antérieures.

et pour ses trois paramètres. Les précipitations prévues et observées et l'évapotranspiration potentielle (ETP), de chaque bassin, constituent les données d'entrée au modèle.

Le principe général du fonctionnement de GRP repose sur un premier réservoir dit de « production » qui gère, en fonction de son taux de remplissage, la part de pluie totale qui se transforme en pluie nette. Un second réservoir dit de « routage », associé à un hydrogramme unitaire, permet d'étaler le temps de cette pluie nette pour obtenir le débit de sortie. La version du modèle ici utilisé ne prend pas en compte le stockage et la fonte de la neige et ne simule pas les barrages, les digues ou autres ouvrages hydrauliques susceptibles d'être présents dans le bassin versant.

Une procédure de mise à jour intervient dans le fonctionnement du modèle dans le but de caractériser au mieux l'état du système avant de démarrer une prévision. Cette procédure est basée sur l'assimilation à l'aide de l'observation de débit du jour qui précède la prévision et intervient au niveau du réservoir de routage ; le niveau du réservoir est calculé directement à partir du dernier débit observé à l'instant de la prévision et modifié pour le mettre à jour (c'est-à-dire pour obtenir le meilleur accord des simulations aux observations).

L'évaluation statistique des séries temporelles des prévisions hydrologiques de GRP basées sur les prévisions des précipitations ont fait l'objet de plusieurs recherches menées dans le but d'examiner différents aspects de la prévision hydrologique d'ensemble (par exemple, Randrianasolo *et al.*, 2010, 2011 ; Velázquez *et al.*, 2010 ; Nicolle *et al.*, 2011 ; Van Pham, 2011).

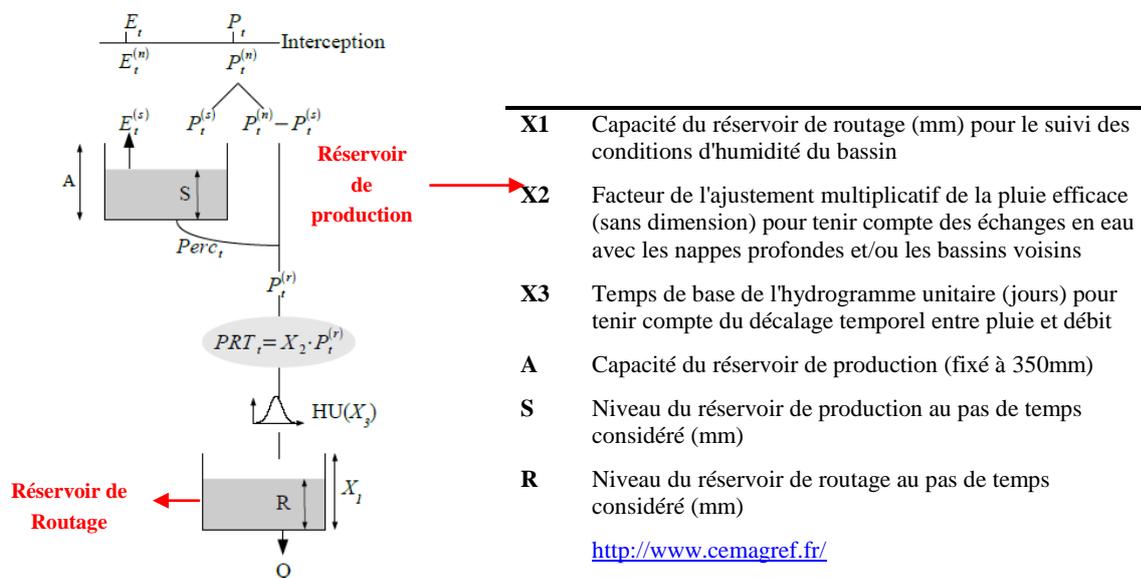


Figure 6. Structure du modèle hydrologique de prévision GRP : à gauche un schéma simplifié du fonctionnement des deux réservoirs (production et routage) et à droite une légende récapitulative des principaux paramètres et variables d'état du modèle

4.3. Bassins versants étudiés et données hydrométéorologiques

Les caractéristiques du modèle météorologique du CEPMMT (un maillage large) contraignent la taille des bassins versants que nous avons sélectionnés. En effet, les rapports d'échelle hydrologique / météorologique doivent être pris en considération. L'étude a été donc menée sur

74 bassins versants français ayant des grandes superficies, allant de 1080 km² à 43824 km², non ou peu influencés par la neige (un pourcentage maximal de neige par an de 35%) (tableau 3).

Les bassins sont répartis dans toute la France et représentent une grande variété de conditions physiques en termes de taille, de topographie, de géologie, des types et des occupations des sols et du climat (figure 7). Ce choix nous a permis de valider notre méthode sur différents types de bassins versants.

Pour ces différents bassins, le Cemagref dispose d'une ample base de données climatologique de Météo-France (pluies, températures, ETP, etc.) et hydrologique (débits) de la banque HYDRO entre 01-08-1958 et 31-07-2009 (avec un maximum de pourcentage de valeurs manquantes de 23% de toute la période). Ces données ont été utilisées pour le calage et la validation du modèle GRP, pour chaque bassin. Dans notre étude, elles ont été aussi utilisées pour le calage et la validation du modèle des erreurs des prévisions déterministes (données sur la période du 11-03-2005 au 01-10-2006, indépendante de la période de calage du modèle GRP).

Tableau 3. Principales caractéristiques des bassins versants étudiés pour la période d'étude (sur la période 1958-2009).

	Min	Max	Médiane
Superficie (km²)	1080	43824	2253
Pluie moyenne annuelle (mm)	624	1287	890
Débit moyen journalier (m³/s)	3	310	26
% Maximal de neige par an	2	35	9

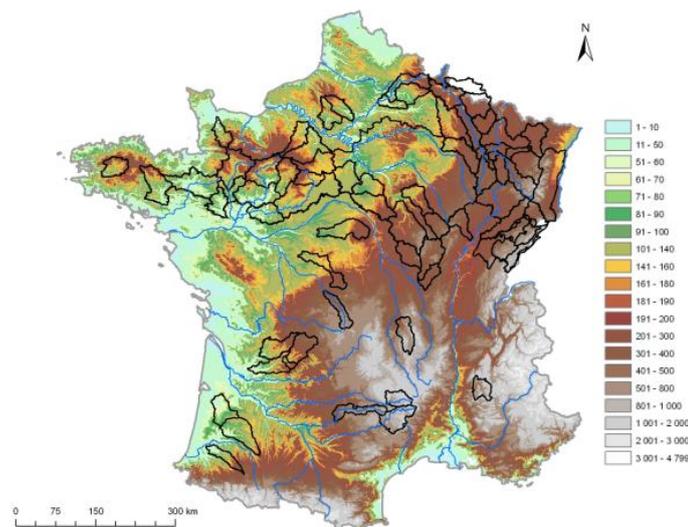


Figure 7. Localisation des 74 bassins versants étudiés en France (contours noirs).

Chapitre 5: Résultats & interprétations

*“We cannot solve our problems with the same thinking we used when we created them.”
Albert Einstein (1879 -1955)*

Le modèle des erreurs passées de la prévision déterministe développé au cours de ce stage a été programmé à l'aide du logiciel statistique R, qui a aussi été utilisé pour les représentations graphiques. Les calculs des scores d'évaluation des prévisions (pour les pluies et les débits) ont été programmés en Fortran.

Comme nous l'avons expliqué dans le chapitre 3, la période de contrôle du modèle d'erreurs (i.e. vérification des prévisions générées) est indépendante de sa période de calage. Nous avons fixé une période de 570j (du 11-03-2005 au 01-10-2006) pour la construction et le calage du modèle et 731 jours pour le contrôle (du 02-10-2006 au 01-10-2008). Ce chapitre présente les principaux résultats obtenus pour le calage et le contrôle du modèle d'erreurs, effectué pour chaque bassin versant étudié, ainsi qu'une évaluation des apports de la méthode pour améliorer la prévision des débits. Quelques bassins versants sont parfois choisis pour illustrer les graphiques et les résultats présentés, qui cependant ont été obtenus et analysés pour l'ensemble des 74 bassins versants étudiés.

5.1. Analyse exploratoire des données en période de calage

En première étape de ce travail, des analyses descriptives de l'échantillon de données ont été effectuées dans la période de calage, pour les 74 bassins versants étudiés, afin d'avoir un aperçu général sur les prévisions et de conclure du type de modèle approprié à notre problématique.

Lors d'un exercice de prévision classique, quatre cas de figure peuvent se présenter quand on compare les valeurs des prévisions aux observations : les deux valeurs sont (ou ne sont pas) nulles, ou encore une valeur est nulle quand l'autre ne l'est pas. Deux cas amènent à des erreurs nulles : si la prévision est strictement égale à l'observation ou si les deux sont nulles. Le premier cas d'erreur nulle n'a été pratiquement jamais observé dans notre échantillon (un pourcentage $\leq 0.35\%$). Pour les autres cas de figure, les diagrammes de fréquences sont présentés dans la figure 8 (valeurs minimales, médianes et maximales sur les 74 bassins versants et pour les 9 échéances de prévision). On observe que les résidus de la prévision déterministe du CEPMMT sont généralement positifs (pluies prévues supérieures aux observations (figure 8 en bas). La fréquence médiane dans les bassins versants étudiés varie entre les différentes échéances de 54% à 60% des jours de prévision en calage pour les erreurs positives, contre 10% au maximum pour les erreurs négatives, témoignant d'une tendance du système de prévision à surestimer les précipitations observées. Les erreurs sont plus importantes (différentes de zéro) pour les échéances élevées, comme le montre le diagramme de la figure 8 à gauche en bas (pourcentage médian de 20% des cas correspondent à des résidus > 0 pour l'échéance de 1 jour et 30% pour l'échéance 9 jours). Observons aussi que le pourcentage médian des résidus nuls (figure 8 en haut à gauche) décroît de 20% des jours de calage pour la 1^{ère} échéance à 10% pour la 9^{ième} échéance. Notons que les erreurs nulles sont dues principalement au « beau temps », ou encore les journées sèches durant lesquelles il n'y a ni observation ni prévision de pluies.

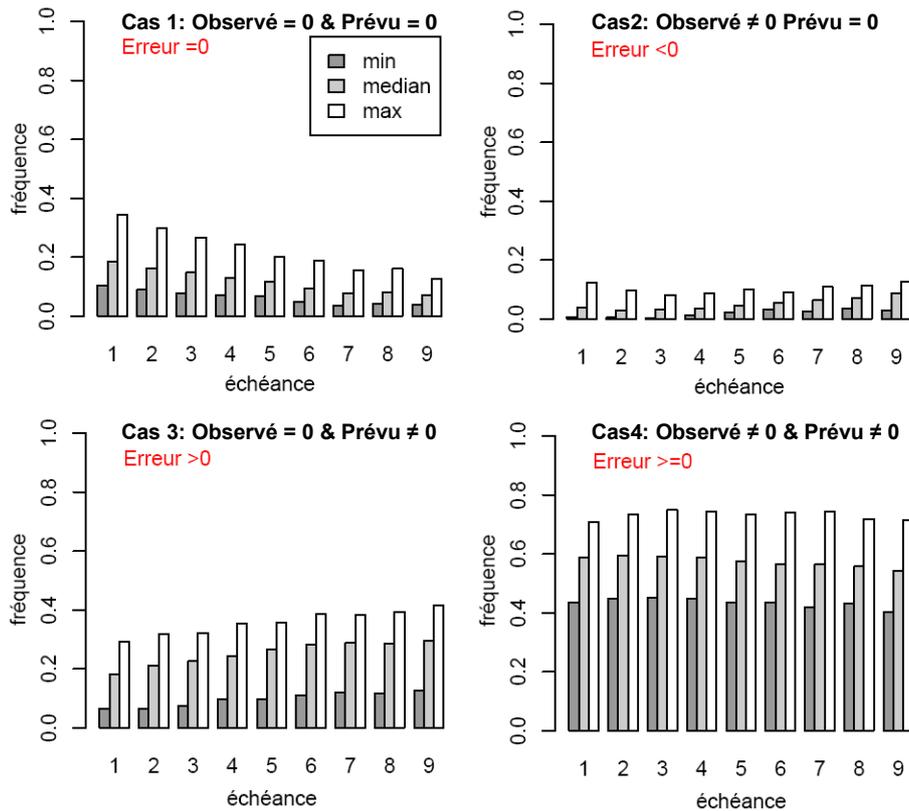


Figure 8. Les cas envisageables lors d'un exercice d'évaluation d'une prévision.

Valeur minimale, maximale et médiane des fréquences (en jours de prévision) calculées pour la prévision déterministe du CEPMMT sur les 74 bassins versants étudiés et pour la période de calage du modèle d'erreur.

La figure 9 récapitule un exemple des résultats des diagnostics de l'échantillon de calage du modèle d'erreurs sur $N_t=570$ jours et $e=9$ échéances de prévision (ici seulement l'échéance 4 est représentée), pour l'exemple du bassin versant «La Sarthe à Neuville-sur-Sarthe» (2713,19 km²). Cet exemple a été choisi pour représenter l'ensemble des bassins étudiés car il correspond à un bassin de taille proche à la surface médiane des bassins versants.

Un examen préliminaire des erreurs de la prévision déterministe en fonction du temps ne révèle aucun caractère saisonnier marqué, comme le montre la figure 9.a. Par contre, l'hétéroscédasticité est évidente : nous observons sur la figure 9.b une évolution en « entonnoir » ; les erreurs sont d'autant plus importantes que les valeurs prévues augmentent, témoignant d'une variance non constante.

Le graphe quantile-quantile (figure 9.c), traçant les quantiles de l'échantillon en fonction des quantiles théoriques de la loi normale, indique la nature non-gaussienne des erreurs : les points du graphique suivent une distribution en S alors qu'ils devraient se situer le long de la ligne d'égaux quantiles.

D'autre part, la fonction d'auto-corrélation des erreurs (ACF pour *Auto-Correlation Function* en anglais) dans la figure 9.d illustre l'évolution des coefficients d'auto-corrélation des erreurs à différents « lags » (observons que pour le lag zéro le coefficient d'auto-corrélation est égale à 1; une erreur est parfaitement corrélée avec elle-même). Les lignes bleues pointillées, sur ce graphique délimitent les seuils de significativité de $\pm 0,1$. Nous remarquons que les auto-corrélations ne sont pas significatives ; elles ne dépassent que très légèrement ces seuils pour les différents lags considérés.

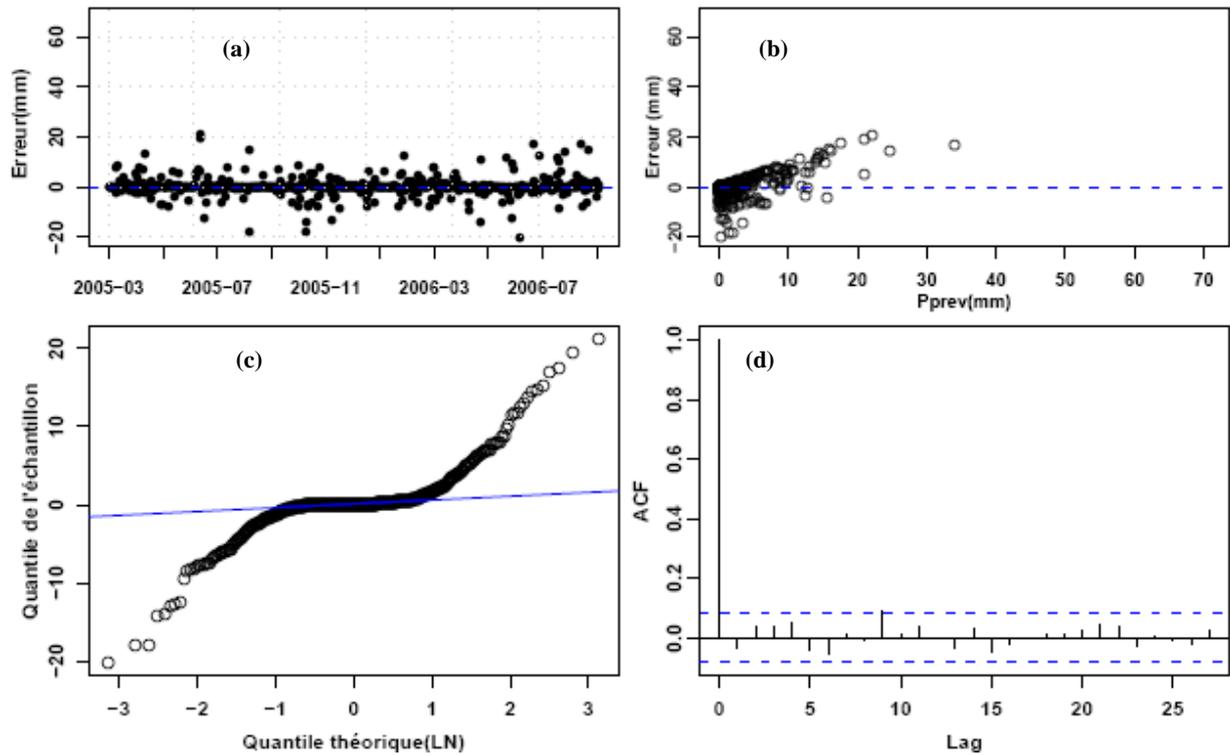


Figure 9. Diagnostic de l'échantillon des erreurs en période de calage du modèle d'erreurs et avant le traitement des données (a) Saisonnalité (b) Homoscédasticité (c) Normalité et (d) auto-corrélation. Exemple du bassin «La Sarthe à Neuville-sur-Sarthe» pour la 4^{ème} échéance de prévision.

Sur la base de ces analyses exploratoires des résidus sur la période de calage, nous avons pu voir que les erreurs de la prévision déterministe des précipitations du CEPMMT ne sont ni auto-corrélées, ni saisonnières. En revanche, une transformation de données s'avère nécessaire pour traiter les problèmes d'hétéroscédasticité et de non-normalité des erreurs.

5.2. Calage du modèle statistique des erreurs de prévision

Rappelons que le modèle statistique des erreurs passées de la prévision déterministe développé pour chaque bassin versant étudié est un modèle additif gaussien multivarié, construit via la transformation de Box-Cox. La formulation mathématique et la justification de la démarche adoptée ont été détaillées dans le chapitre 3. Afin de s'assurer du bon comportement du modèle, l'optimisation des paramètres qui définiront le modèle a été d'abord effectuée sur la période de calage.

5.2.1. Calage des paramètres de la transformation Box-Cox (λ_1, λ_2)

La transformation de Box-Cox (équation 3) a été appliquée sur les données de l'échantillon de calage du modèle ($N_t = 570$ jours de prévision, $e = 9$ échéances de prévisions) pour les 74 bassins versants étudiés. Nous avons choisi de fixer λ_2 à une valeur arbitraire (0.01) et nous ne manipulons qu'un seul paramètre, λ_1 par la suite.

Dans un premier temps, afin d'optimiser le paramètre λ_1 de la transformation de Box-Cox, pour chaque échéance de prévision, nous avons analysé les variations de l'écart-type¹⁵ des erreurs en fonction de différentes valeurs de λ_1 (figure 10). Les valeurs minimales des écart-types varient d'une échéance à autre et sont globalement atteintes pour des λ_1 compris entre ± 0.5 . Dans cet intervalle, on remarque que l'écart-type des erreurs varie très légèrement. Pour cette raison, et par souci de simplification, nous avons décidé de fixer le paramètre $\lambda_1 = 0.5$ pour le traitement des données dans ce qui suit. Ceci a été appliqué sur les 74 bassins versants étudiés, qui ont tous montré un comportement similaire à celui représenté dans la figure 10.

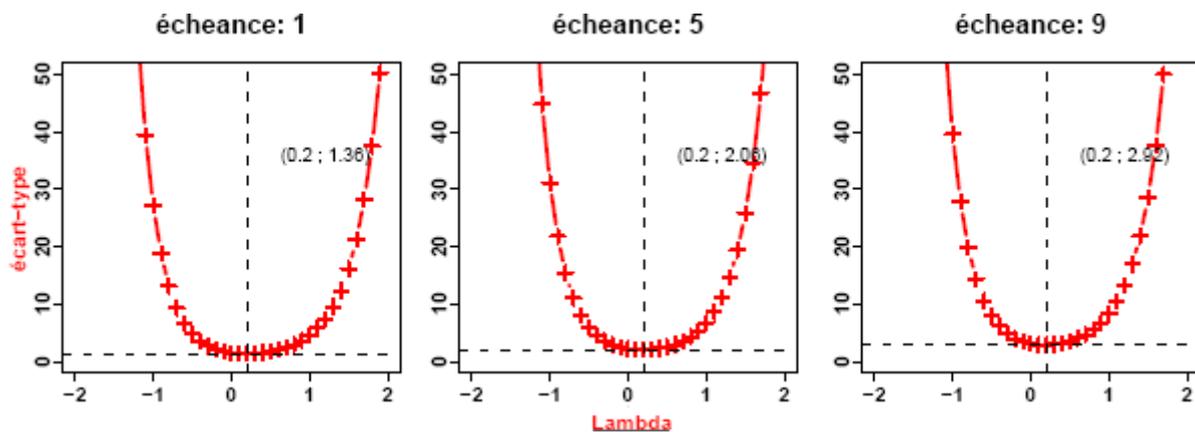


Figure 10. Optimisation du paramètre λ_1 de la transformation de Box-Cox pour les échéances 1, 5 et 9 jours de la prévision déterministe du CEPMMT. Exemple du bassin « La Meuse à Commercy » (2327,9 km²).

5.2.2. Calage des paramètres du modèle gaussien multivarié (μ , Σ)

Une fois définie la transformation Box-Cox à appliquer sur les séries des valeurs observées et prévues, les résidus de prévision ont été calculés et les paramètres de modèle gaussien multivarié ont été estimés (moyenne μ des résidus et matrice de covariance Σ ou de corrélation de résidus normalisés C_e ; voir chapitre 3). Pour cela, deux modèles ont été testés : 1) un modèle général construit avec l'ensemble de paires observations-prévisions de la série disponible pour chaque bassin versant et 2) un modèle de résidus non-nuls, construit sans la prise en compte des situations de « beau temps », où les pluies observées et prévues sont nulles.

5.2.2.1. Un premier calage : modèle général (modèle A)

La visualisation des résidus pour chaque bassin et chaque échéance de prévision nous a permis d'évaluer la capacité de la transformation de Box-Cox à corriger les problèmes que nous avons pu remarquer lors des premiers diagnostics (cf. figure 9). La figure 11 montre les résultats obtenus après transformation de données du modèle général pour l'exemple du bassin « La Sarthe à Neuville-sur-Sarthe » (2713,1 km²) à l'échéance 2 de la prévision.

La transformation de Box-Cox est apparue pertinente et a donné des résultats encourageants quant à l'hétéroscédasticité des résidus. Nous pouvons remarquer dans la figure 11.b que la dispersion des résidus est quasiment homogène sur tout le domaine des valeurs prévues.

¹⁵ Plusieurs auteurs minimisent l'écart-type des valeurs transformées afin d'optimiser la valeur de λ_1 de la transformation Box-Cox. Voir par exemple les explications données par: http://www.isixsigma.com/index.php?option=com_k2&view=item&id=1096:making-data-normal-using-box-cox-power-transformation&Itemid=208

En outre, il n'y a toujours pas de caractères significatifs remarquables pour l'auto-corrélation et la saisonnalité (figure 11.a et d).

En revanche, le comportement des erreurs transformées ($\lambda_1=0.5$; $\lambda_2=0.01$) a montré que les tests de statistiques utilisés (test de Shapiro-Wilk et test de Kolmogorov-Smirnov) ont rejeté l'hypothèse de normalité des données pour l'ensemble des bassins versants étudiés. Nous avons remarqué aussi que les p-valeurs de ces tests s'approchent du seuil de significativité, fixé à $\alpha=0.01$, au fur et à mesure que l'échéance de prévision augmente (i.e. un meilleur résultat est obtenu pour les échéances élevées).

La figure 11.c révèle l'influence des résidus situés dans les queues de distribution sur la qualité de calage de la loi normale (valeurs transformées extrêmes, ici $> \pm 4\text{mm}$), mais aussi l'effet des résidus nuls (marqués par un cercle rouge). Les écarts par rapport à la normalité semblent être fortement influencés par la présence de ces erreurs nulles, comme le montre la figure 11.e et la figure 11.f.

En effet, l'histogramme des erreurs transformées montre que les classes des erreurs nulles ont la probabilité la plus élevée d'être réalisées (cercle rouge sur la figure 11.e), ce qui n'a pu être modélisé ni par la densité de probabilité d'une loi normale de même moyenne et de même écart-type (gaussienne en courbe bleue pointillée sur le graphique) ni par une loi normale de moyenne nulle¹⁶ (gaussienne en courbe rouge), vu le décalage important entre les deux classes de probabilités successives. Ce pic de probabilité au quantile zéro semble être à l'origine des faibles p-valeur du test de Shapiro-Wilk (voir §3.1.1.4).

La figure 11.f confirme ces observations en comparant la fonction de répartition empirique des résidus transformés (courbe noire) à celle de la loi normale (courbe bleue) : une discontinuité au niveau des résidus nuls entraîne des écarts plus importants entre les deux fonctions et le rejet de l'hypothèse de normalité des résidus par les tests de Kolmogorov-Smirnov (voir §3.1.1.4).

Cependant, puisque visuellement nous avons constaté que l'écart à la normalité ne semble pas toujours si prononcé, même si l'hypothèse de normalité est rejetée par les tests statistiques, nous avons quand même décidé, par sa simplicité, de garder ce modèle des résidus pour la construction des faisceaux de prévision déterministes habillées (*DH*). Les paramètres de la loi gaussienne multivariée ont alors été estimés et appliqués en période de contrôle (la validation du modèle est présentée dans §5.3).

Les différentes constatations précédentes nous ont néanmoins incités à se poser des questions sur la nécessité de prendre en compte et de traiter le cas des erreurs nulles à part, même si ceci complexifie la démarche de construction du modèle d'erreur. C'est le cas du modèle alternatif présenté par la suite (modèle B).

¹⁶ Rappelons ici qu'une densité de probabilité d'une loi normale de moyenne nulle atteint sa valeur maximale à zéro. Ceci nous a amené à comparer la densité de probabilité de l'échantillon à celle d'une loi normale de moyenne nulle et de même écart-type.

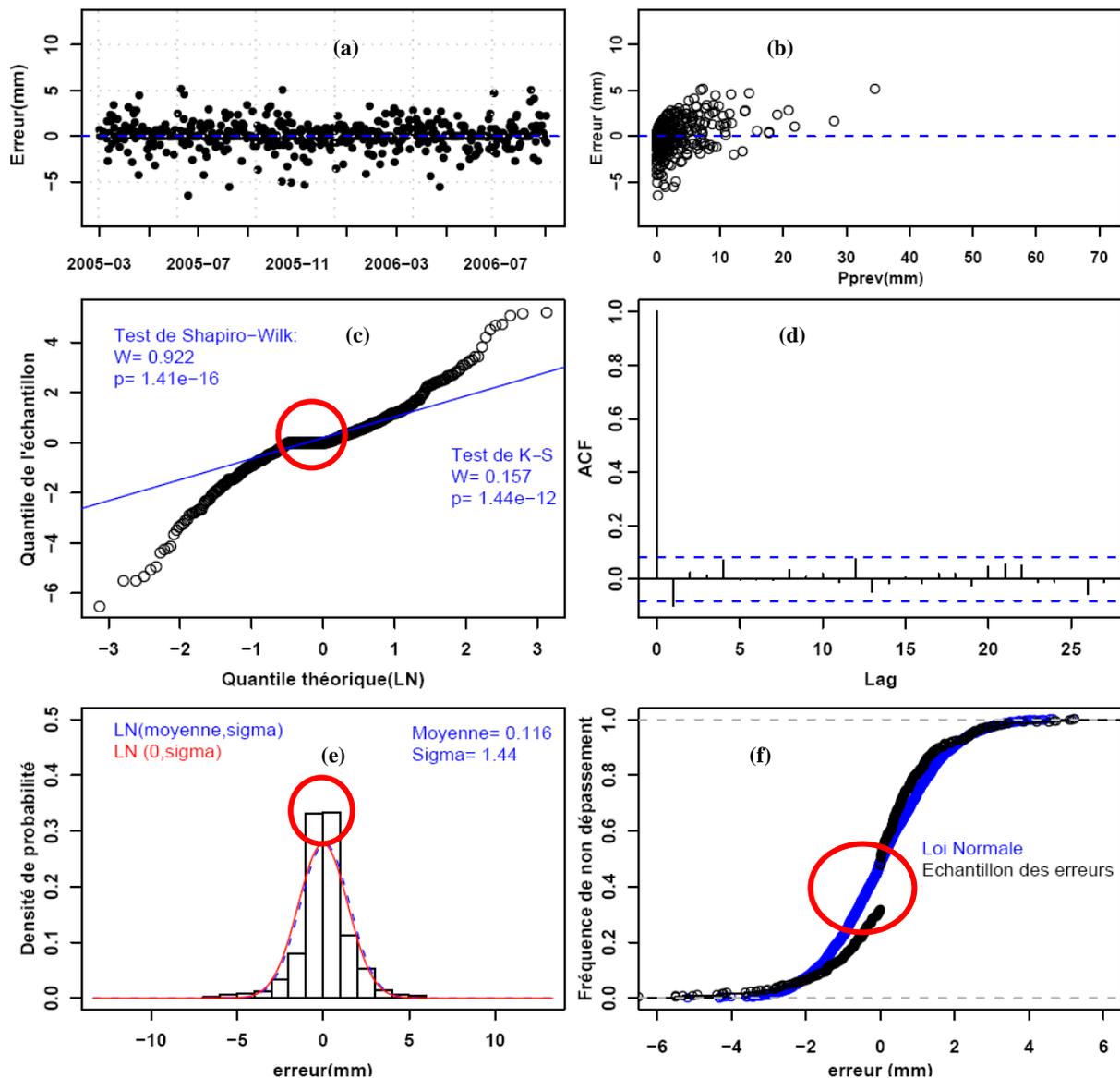


Figure 11. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage pour l'exemple du bassin «La Sarthe à Neuville-sur-Sarthe» à l'échéance 2 de la prévision.

(a) Saisonnalité (b) Homoscédasticité (c) Graphe quantile-quantile des résidus de prévision avec les résultats des tests de Shapiro-Wilk et Kolmogorov-Smirnov en légende (d) Fonction d'auto-corrélation des résidus de prévision (e) Histogramme des résidus de prévision et densité de probabilité d'une loi normale ayant les mêmes paramètres (moyenne, écart-type) (f) Fonctions de répartition empiriques d'une loi normale (courbe bleue) et de l'échantillon des résidus de prévision (courbe noire).

5.2.2.2. Un deuxième calage : modèle sans les pas de temps secs (modèle B)

Pour ce modèle, nous avons enlevé les pas de temps secs, c'est-à-dire les jours où les pluies observées et prévues sont nulles, de la procédure de transformation Box-Cox des données et donc de l'échantillon de calage du modèle d'erreur. La fréquence de « beau temps », évaluée pour chaque bassin versant et pour chaque échéance de prévision (voir la synthèse dans la figure 8 en haut à gauche), va constituer un paramètre supplémentaire du modèle d'erreur (f_b ; cf. expliqué ci-après).

En ne considérant que les cas 2, 3 et 4 de la figure 8, les résultats du diagnostic des résidus obtenus par la transformation Box-Cox des données montrent des meilleurs ajustements de la loi normale pour tous les bassins étudiés. La figure 12 et figure 13 présentent, respectivement,

le meilleur et le pire ajustement de la loi normale en fonction des résultats des tests de Kolmogorov-Smirnov et Shapiro-Wilk. Nous observons un bon ajustement de la fonction de répartition (figure 12.d et figure 13.d) ainsi qu'une adéquation assez approuvable de la FDP normale malgré que l'aplatissement de la loi n'est pas très prononcé dans le deuxième cas (la distribution normale ne s'élève pas assez haut dans la figure 13.c).

Les graphiques quantile-quantile montrent aussi des résultats plus intéressants : nous pouvons remarquer que, même dans le cas de la plus faible p-valeur, et mises à part les queues de la distribution, les quantiles des données situés aux voisinages de la moyenne sont bien alignés avec ceux de la loi normale (cercle rouge dans la figure 13.a). Etant donné que le modèle d'erreur que nous adoptons s'appliquera à un intervalle de confiance de 80% autour de la moyenne des résidus (voir chapitre 3), l'adéquation obtenue peut être considérée suffisante

D'autre part, apercevons sur la figure 12.b et la figure 13.b que l'hypothèse d'homoscédasticité reste toujours acceptable (la variance des erreurs n'est pas croissante avec la valeur prévue).

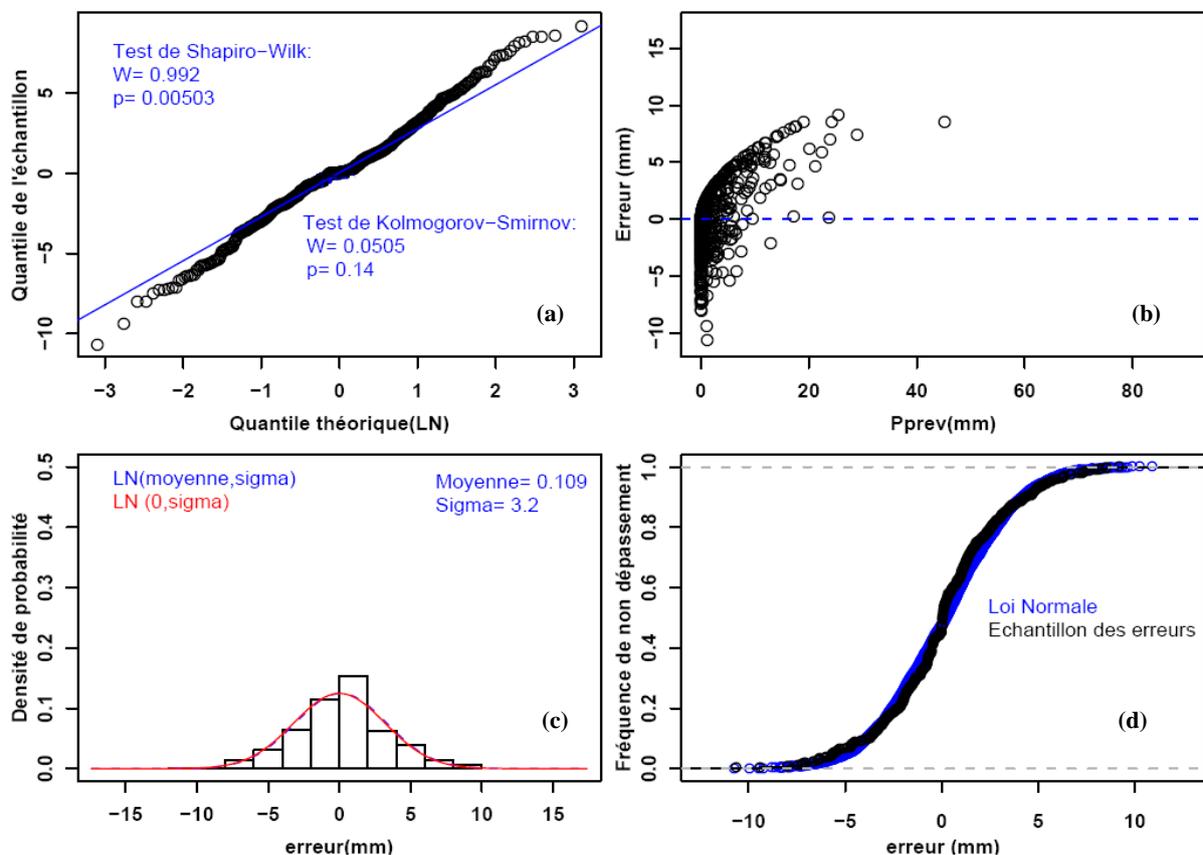


Figure 12. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage du modèle (B) pour le bassin L'Armançon à Aisy sur Armançon présentant la p-valeur maximale (échiance 9).

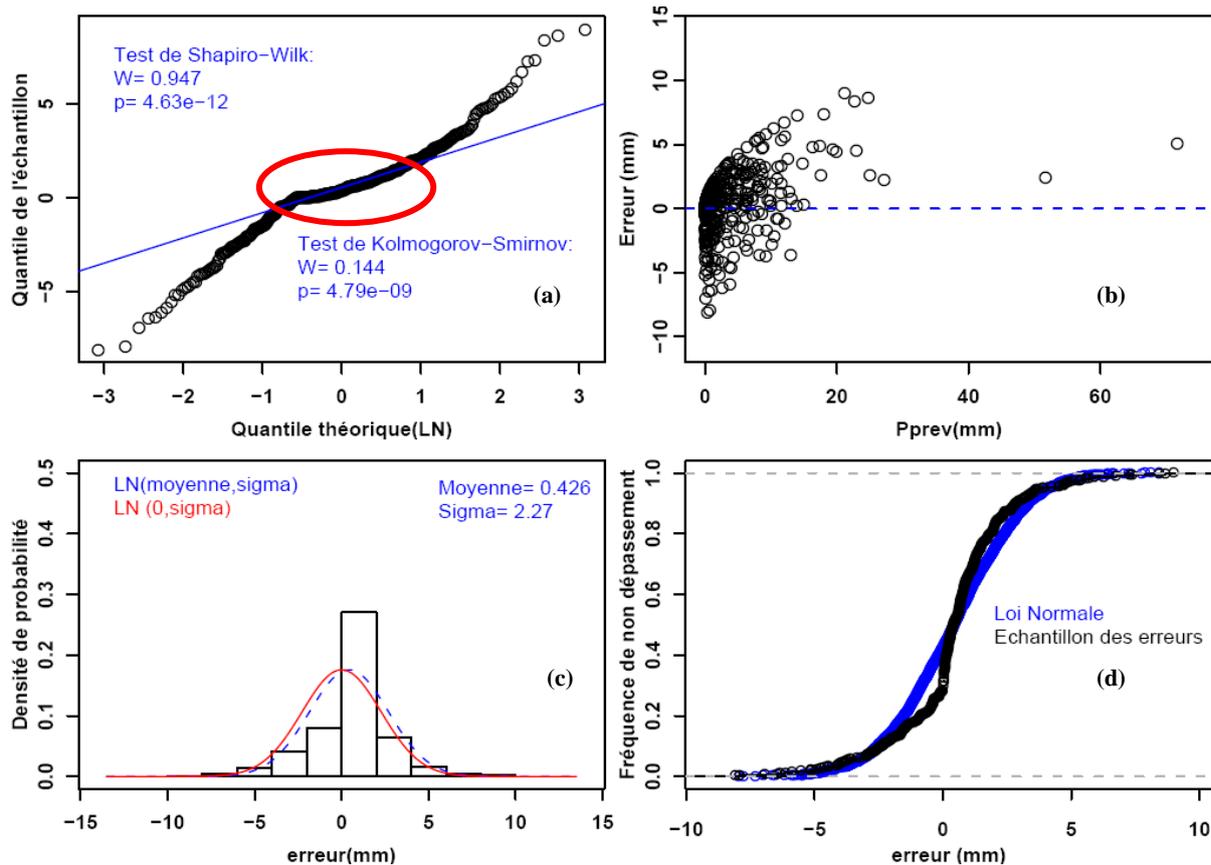


Figure 13. Diagnostic de l'échantillon des erreurs transformées (Box-Cox $\lambda_1=0.5$; $\lambda_2=0.01$) en période de calage du modèle d'erreur (B) pour le bassin La Midouze à Campagne présentant la p-valeur minimale (échéance 3).

L'analyse des résultats du diagnostic sur l'ensemble des bassins versants et échéances de prévision a montré que soit l'hypothèse de normalité ne peut pas être rejetée par les tests de normalité, soit, pour les bassins où l'hypothèse est rejetée par au moins l'un des deux tests considérés, une quasi-normalité (faible p-valeur) était néanmoins observée. Les cartes de la figure 14 indiquent que, globalement, pour entre 20% et 50% des bassins étudiés l'hypothèse de normalité n'est pas rejetée pour au moins un des tests appliqués, quand on considère les échéances allant de 7 à 9 jours, et ce notamment pour le cas des bassins situés plus au nord de la France.

En résumé, le modèle B, ici testé, présente un ajustement plus ou moins bien aux données, selon le bassin et selon l'échéance de prévision, tout en étant plutôt acceptable dans son ensemble. Par conséquent, nous l'appliquerons sur l'archive de contrôle dont nous disposons, tout en étant néanmoins conscients de ses limitations (hypothèse de normalité plus tenable pour les échéances élevées et pour les bassins du nord-est de la France).

Dans la période de contrôle du modèle et pour la construction des membres du faisceau de prévisions DH , nous avons dû reconsidérer la fréquence du « beau temps » (pluies prévues et observées nulles). Le modèle d'erreur a donc un autre paramètre f_b ; celui de la fréquence de l'erreur nulle. Nous traitons ce modèle de la manière suivante :

- à chaque pas de temps pour lequel nous souhaitons construire un faisceau de prévision, un tirage aléatoire d'une loi uniforme d'une fréquence entre 0 et 1 est faite ;

- si la prévision déterministe émise à la 1^{ère} échéance¹⁷ est égale à zéro et cette fréquence tirée aléatoirement est inférieure à f_b , l'erreur est considérée non nulle et les membres sont créés avec le modèle gaussien multivarié calé ;
- si la prévision déterministe émise à la 1^{ère} échéance est égale à zéro et cette fréquence tirée aléatoirement est supérieure à f_b , l'erreur est considérée nulle et la prévision déterministe n'est pas corrigée (tous les membres des faisceaux sont identiques).

Le choix de rajouter une étape supplémentaire au modèle d'erreur, traitant les erreurs nulles, est le fruit de plusieurs tests (non présentés dans ce rapport) qui ont montré que cela permettait effectivement de ne pas toujours rajouter des erreurs à la prévision déterministe (forte surestimation de la réalité notamment quand celle-ci était nulle) et d'améliorer la génération des scénarios.

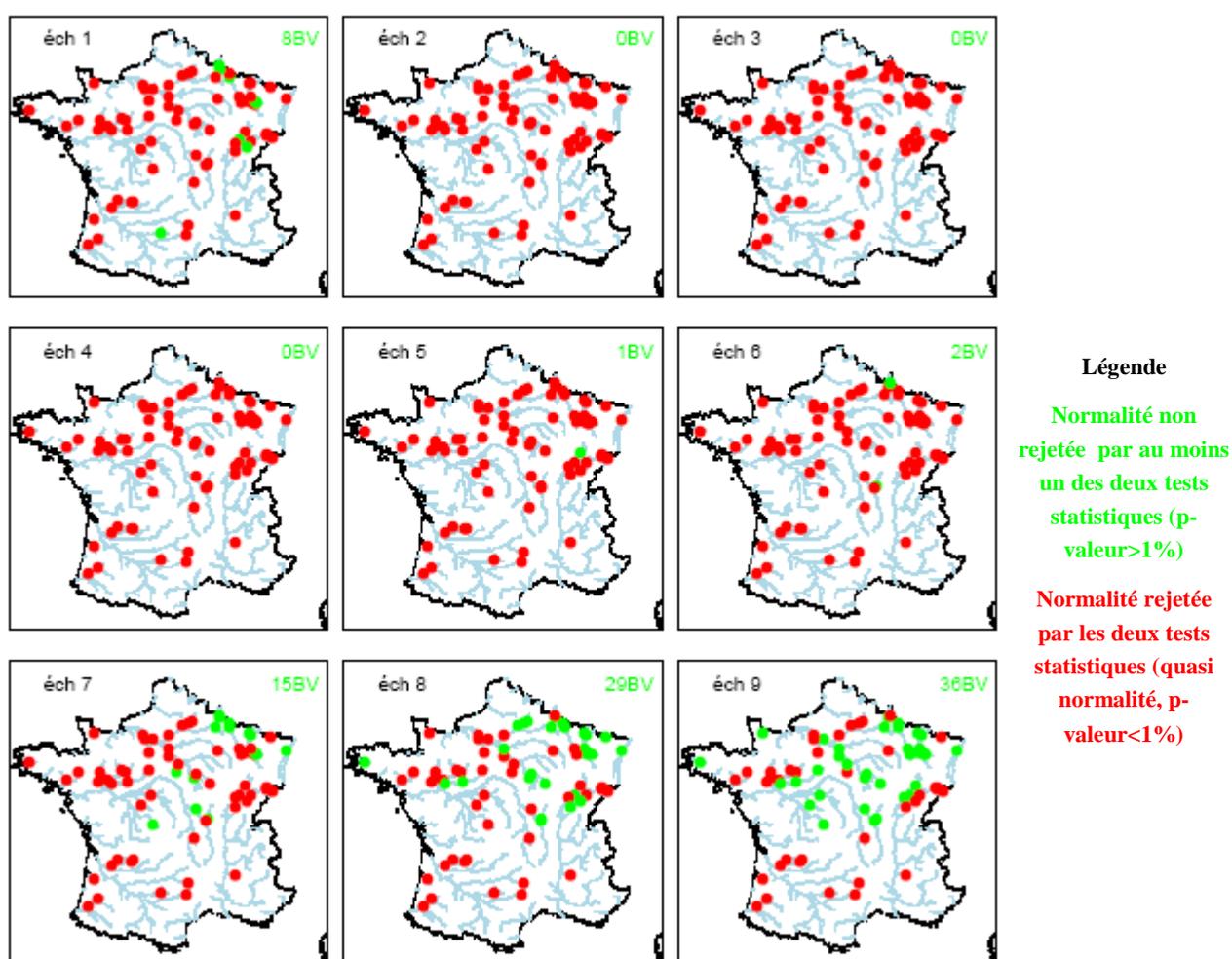


Figure 14. Cartes des résultats des tests statistiques de normalité.

¹⁷ Le test est effectué en fonction de la 1^{ère} échéance puisqu'elle est sensée être la plus proche de l'observation (pluie réellement observée).

5.3. Contrôle des modèles statistiques d'erreurs

A l'issue de la période de calage, les paramètres des deux modèles construits pour les 74 bassins versants ont été appliqués à l'échantillon de contrôle et ainsi les faisceaux des prévisions de précipitation (notées *PDH*) ont été construits (51 scénarios pour chaque jour de la période de calage et pour les neuf échéances journalières de prévision). Ces ensembles ont été utilisés en entrée du modèle hydrologique GRP pour générer les prévisions d'ensemble de débits (notée *QDH*).

Dans cette partie, nous évaluerons la qualité de la prévision des précipitations *PDH* en la comparant à la prévision déterministe du CEPMMT (notée *DET*) et aux différents scénarios du SPE du CEPMMT (notés *EU*). Les performances de la prévision *PDH* vont être aussi jugées par rapport à ses capacités à améliorer les prévisions hydrologiques (évaluation des *QDH*).

5.3.1. Les erreurs quadratiques des prévisions DH

Le score ratio-RMSE est souvent le score le plus simple qui représente la qualité d'une prévision d'ensemble en moyenne par rapport à l'observation sans être influencé par la taille du bassin versant étudié.

La figure 15, ci-dessus, présente les écarts¹⁸ entre les ratio-RMSE (équation 17) des modèles A et B par rapport aux ratio-RMSE de la prévision déterministe *DET* (points verts) et la prévision *EU* (points noirs) des pluies (haut) et des débits (bas), pour les 74 bassins versants étudiés. Sur cette figure, les valeurs négatives représentent donc un «gain» en erreurs quadratiques pour le modèle A et/ou B (i.e. moins d'erreur par rapport à *DET* et/ou *EU*) et les valeurs positives représente une «perte».

Nous pouvons constater que pour la 1^{ère} échéance de prévision, les *PDH* (figure 15 en haut à gauche) issues des modèles A et B sont très comparables (points situés sur la 1^{ère} bissectrice). Pour les différents bassins nous observons des écarts compris entre -0.01 et 0.25 témoignant d'une faible perte, en termes d'erreurs quadratiques, par rapport aux prévisions *EU* et *DET* du CEPMMT. Pour les échéances élevées, ce score est de plus en plus à l'avantage des prévisions *EU*; les écarts varient entre 0.25 et 1.4. Par rapport à la prévision *DET*, en moyenne les erreurs des membres DH restent proches de ceux de la prévision déterministe, nous observons des écarts plus faibles compris entre -0.01 et 0.5. Nous observons aussi un très faible décalage des points vers le bas du graphique à fur et à mesure que l'échéance augmente; c'est le modèle B qui présente légèrement plus d'erreurs quadratiques (généralement un décalage entre +0.1 et +0.3 par rapport au modèle A).

En ce qui concerne les prévisions des débits, nous remarquons sur la figure 15 en bas que, pour la 1^{ère} échéance, les erreurs des prévisions *QDH* issues des modèles A et B et les erreurs de la prévision *DET* et *EU* sont quasiment identiques (des écarts de l'ordre de 10^{-2}). Ces valeurs sont probablement dues à l'effet de la mise à jour du modèle GRP sur la 1^{ère} échéance de prévision: elle est plus influencée par le dernier débit observé (le débit de mise à jour) que par la précipitation prévue. En effet, le but de la mise à jour en elle-même est de réduire l'écart aux observations, et plus précisément le RMSE, les prévisions de GRP seront donc très proches pour les différents systèmes de prévisions testés.

En allant vers l'échéance 9 les écarts augmentent, tout en restant légèrement plus faible pour les *PDH* issues du modèle A (entre -0.25 et 1.8 pour le modèle B et -0.25 et 1.25 pour le

¹⁸ Différence entre Ratio-RMSE de la prévision A et Ratio-RMSE de la prévision DET en fonction de celle des prévisions B (points verts) et la même chose pour les prévisions EU (points noirs).

modèle A), ce qui s'explique par la dégradation de la qualité de la prévision en moyenne avec l'échéance.

Pour le calcul du score RMSE, la moyenne des membres a été considérée comme une prévision déterministe. Néanmoins, la moyenne ne peut vraiment pas refléter le comportement de tout l'ensemble. En effet, un ensemble peut contenir des membres parfaits (très proches de l'observation) et des membres aberrants qui auront un effet important sur la valeur moyenne.

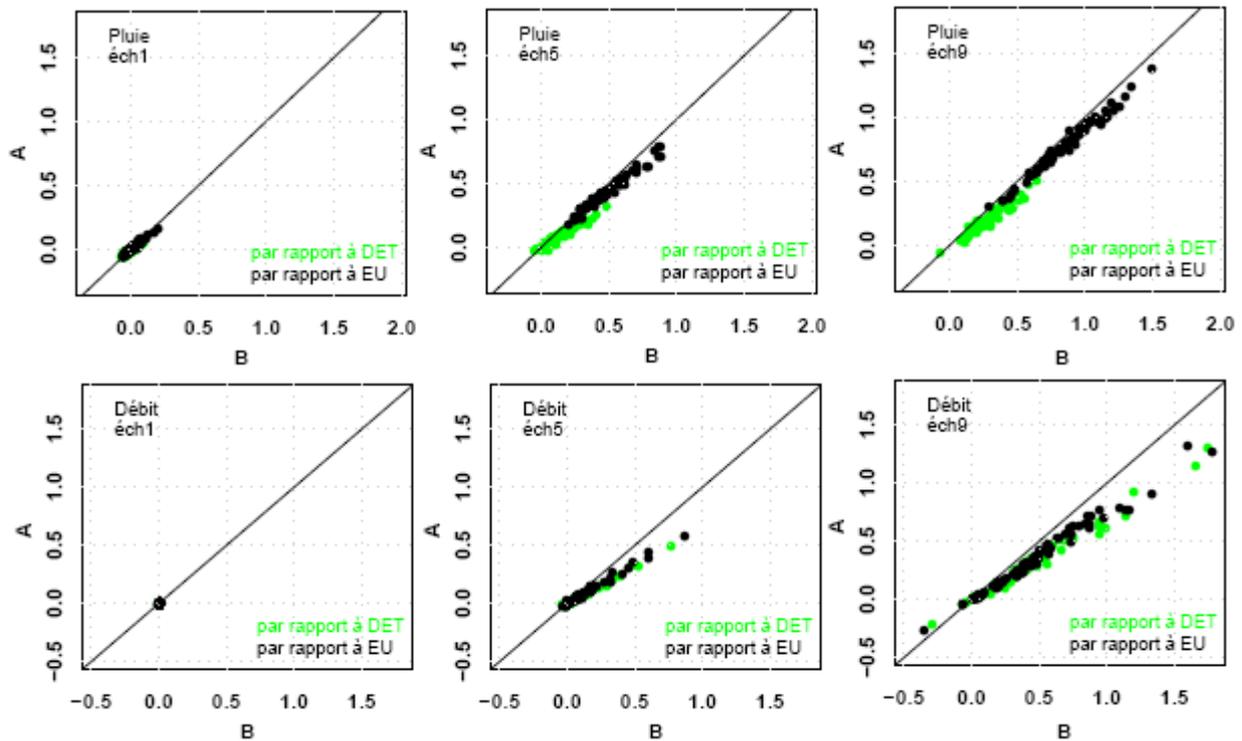


Figure 15. Ecarts en erreurs quadratiques par rapport aux prévisions DET et EU sur les 74 bassins versants étudiés pour les échéances de prévision $j+1$, $j+5$ et $j+9$.

En haut écart entre les ratio-RMSE de la moyenne des membres issus des modèles A en fonction des écarts des ratio-RMSE de la moyenne des membres issus des modèles B
En bas idem pour les QDH (modèle A et B).

Afin d'étudier davantage les erreurs des modèles en terme de probabilités de prévision des membres des prévisions *DH* (en pluies et débits) et *EU*, nous avons étudié le score BS pour chaque bassin versant et pour différents seuils (dépassement des quantiles Q95%, Q75%, Q50% et Q25%). Rappelons que ce score permet de qualifier le comportement des deux systèmes de prévision d'ensemble face à un évènement binaire (dépassement de chacun des seuils mentionnés) et que, comme pour le RMSE, les meilleures prévisions auront un score plus faible. La dispersion des scores BS des prévisions est montrée dans la figure 16. Pour chaque seuil de dépassement étudié, les 74 valeurs de BS obtenus pour chaque bassin versant de l'étude sont représentées. Ici nous représentons uniquement les résultats du modèle B, qui sont très proches de ceux du modèle A.

Le graphiques qui comparent les scores de la *PDH* à ceux des prévisions *EU* (figure 16 en haut) confirment bien les résultats précédents. Nous observons, à la 1^{ère} échéance, un léger décalage vers le bas des points du graphique (les BS des *PDH* sont plus faibles) notamment pour les

faibles seuils (Q25% et Q50%). Plus l'échéance augmente, plus les scores BS s'éloignent de la diagonale, au profit des prévisions *EU*. Notons que pour les seuils Q95% et Q75%, les BS restent faibles, respectivement, à 0.10 et 0.30 et que ces scores ne varient pas significativement d'un bassin à un autre.

Observons maintenant les prévisions d'ensemble des débits (figure 16 en bas). Nous pouvons remarquer, tout d'abord, que les BS deviennent plus dispersés entre les bassins versants. En outre, pour la 1^{ère} échéance les deux modèles de prévisions présentent des scores quasiment identiques (les points sont collés à la diagonale) et globalement inférieurs à 0.40.

Au fur et à mesure que l'échéance augmente, les scores commencent à se différencier, avec les scores de la prévision *DH* devenant supérieurs aux scores de la prévision *EU*, notamment pour les seuils supérieurs au Q50%. Cependant, on notera que les scores des prévisions habillées *DH* restent faibles et proches de ceux de la prévision hydrologique d'ensemble basée sur les prévisions météorologiques d'ensemble *EU*, notamment pour les échéances jusqu'à 6 jours.

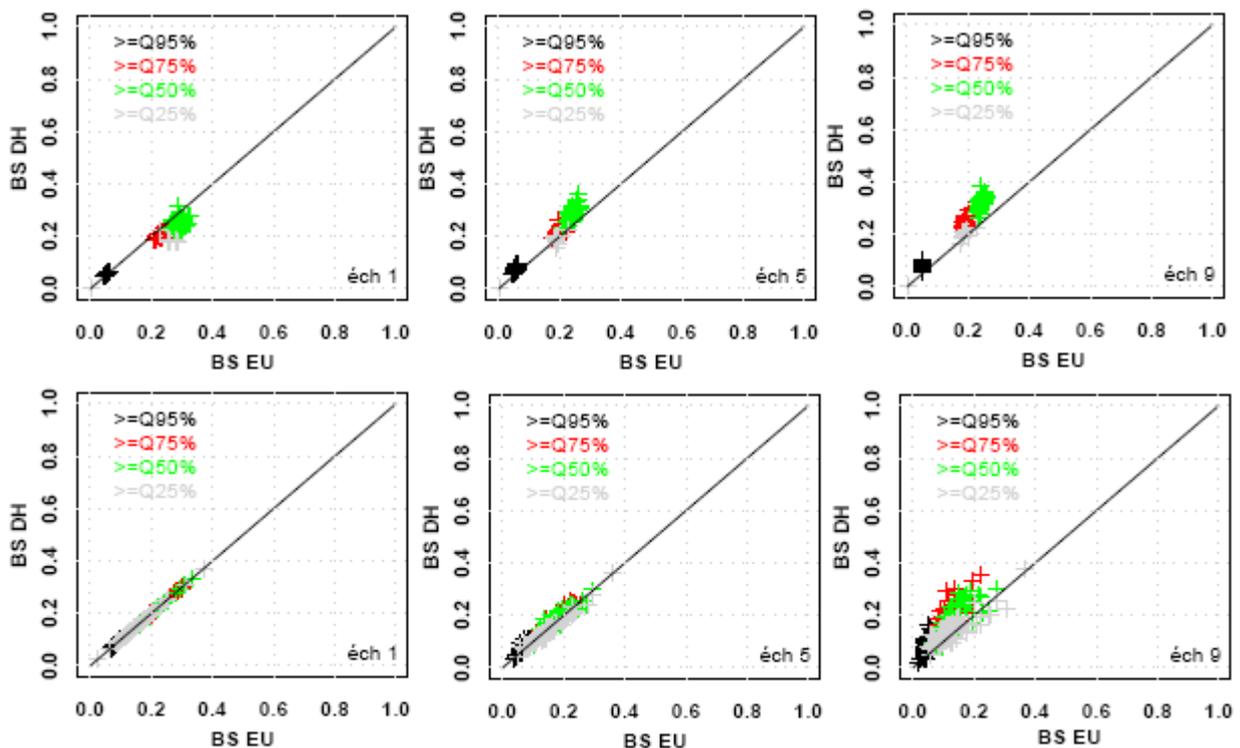


Figure 16. Dispersion des scores BS de la prévision PDH du modèle B par rapport aux EU du CEPMMT (en haut) et QDH en fonction des prévisions de débits à partir des EU du CEPMMT (en bas) pour les échéances de prévision $j+1, j+5, j+9$ et le dépassement des seuils des quantiles Q95%, Q75%, Q50% et Q25%.

5.3.2. Les critères issus des tables de contingence

Pour la construction des tables de contingence, nous avons considéré les dépassements des seuils élevés de débits par plusieurs pourcentages des membres des prévisions d'ensemble (au moins 10% des membres, au moins 20% des membres, etc.). Ici, nous montrerons les résultats pour les scores de la table de contingence en utilisant au moins 80% des membres (une forte probabilité de dépassement des seuils Q75% et Q95%).

La figure 17 et la figure 18 illustrent les scores de la table de contingence des trois systèmes de prévisions des pluies (*PDH* des modèles A et B et *DET*) pour les échéances 1, 5 et 9 jours et

pour le dépassement des quantiles Q95% et Q75%. Les valeurs sont représentées sous forme de boîtes à moustaches (box-plots) notamment pour les quantiles 95%, 75%, 50%, 25% et 5% des bassins (voir la légende du graphique). Rappelons que plus l'indice de succès critique (CSI) et le taux de succès (POD) sont élevés et plus le ratio de fausses alertes (FAR) est faible, plus le système de prévision est performant.

Pour les trois échéances représentées, nous observons sur la figure 17 en haut que les trois systèmes de prévisions montrent des scores CSI globalement inférieurs à 20% pour le dépassement du quantile Q95 % avec un très faible écart entre les bassins versants (un faible intervalle inter-quantile pour les box-plots). Cependant, ce score passe, de l'échéance de 1 jour à l'échéance de 9 jours, de l'intervalle [7%, 20%] à l'intervalle [17%, 40%] pour le dépassement du seuil du quantile Q75%. Pour ce même seuil, les prévisions *PDH* déduites des deux modèles A et B donnent des résultats très comparables, voire meilleurs que ceux proposés par la prévision *DET*, plus particulièrement pour les échéances élevées (un écart de 20% entre le CSI des *PDH* et *DET* pour l'échéance de 9 jours).

Les taux de fausses alertes (FAR, figure 17 au milieu) restent assez forts pour les trois systèmes de prévision évalués. Nous observons des valeurs globalement supérieures à 60% pour le seuil Q95%. Nous constatons aussi une amélioration de ces ratios pour les dépassements du seuil Q75%. Notamment pour ce seuil de dépassement, les modèles A et B construits présentent une meilleure performance par rapport à la prévision *DET* (plus faibles valeurs de FAR).

Pour ce qui est du score POD (probabilité de bonne détection, figure 17 en bas), nous pouvons voir, que les prévisions *PDH* améliorent aussi le taux de succès notamment pour les échéances élevées (à partir de l'échéance 5) et pour le seuil Q75%.

Finalement, il est à noter que les résultats des deux modèles A et B sont toujours très proches, avec quelques meilleures performance du modèle B pour les échéances moyennes (graphiques pour l'échéance à 5 jours) et le dépassement du quantile Q75%. Remarquons également que les scores des différents bassins varient plus dans le cas des modèles *DH* construits que pour les prévisions *DET* (étendu des boîtes et moustaches dans mes graphiques). Dans le cas du score CSI, par exemple, on observe un plus grand nombre de bassins avec un score plus élevé avec les prévisions de *DH* de notre modèle qu'avec la prévision déterministe *DET*.

Les mêmes scores (CSI, FAR, POD) ont été examinés pour les prévisions d'ensemble des débits (figure 18). L'effet de la mise à jour des débits est à nouveau observé sur la première échéance : les scores basés sur les modèles d'ensemble (A et B) construits dans notre étude sont tous très proches des scores basés sur les prévisions déterministes *DET* pour cette échéance. Pour les échéances plus lointaines, cependant, nous pouvons voir que les débits d'ensemble, quand on considère une probabilité de dépassement de seuil donné par au moins 80% des membres *QDH*, sont nettement meilleurs : en termes de taux de succès, par exemple, des valeurs de POD supérieures à 80% pour le dépassement du seuil Q75% sont observés, nettement supérieures aux valeurs de POD de la prévision déterministe (figure 18 en bas).

Enfin, remarquons également que les différences entre les modèles A et B de construction des faisceaux des prévisions des précipitations sont bien plus faibles quand on évalue les débits que lorsque l'on évalue les précipitations (figure 10).

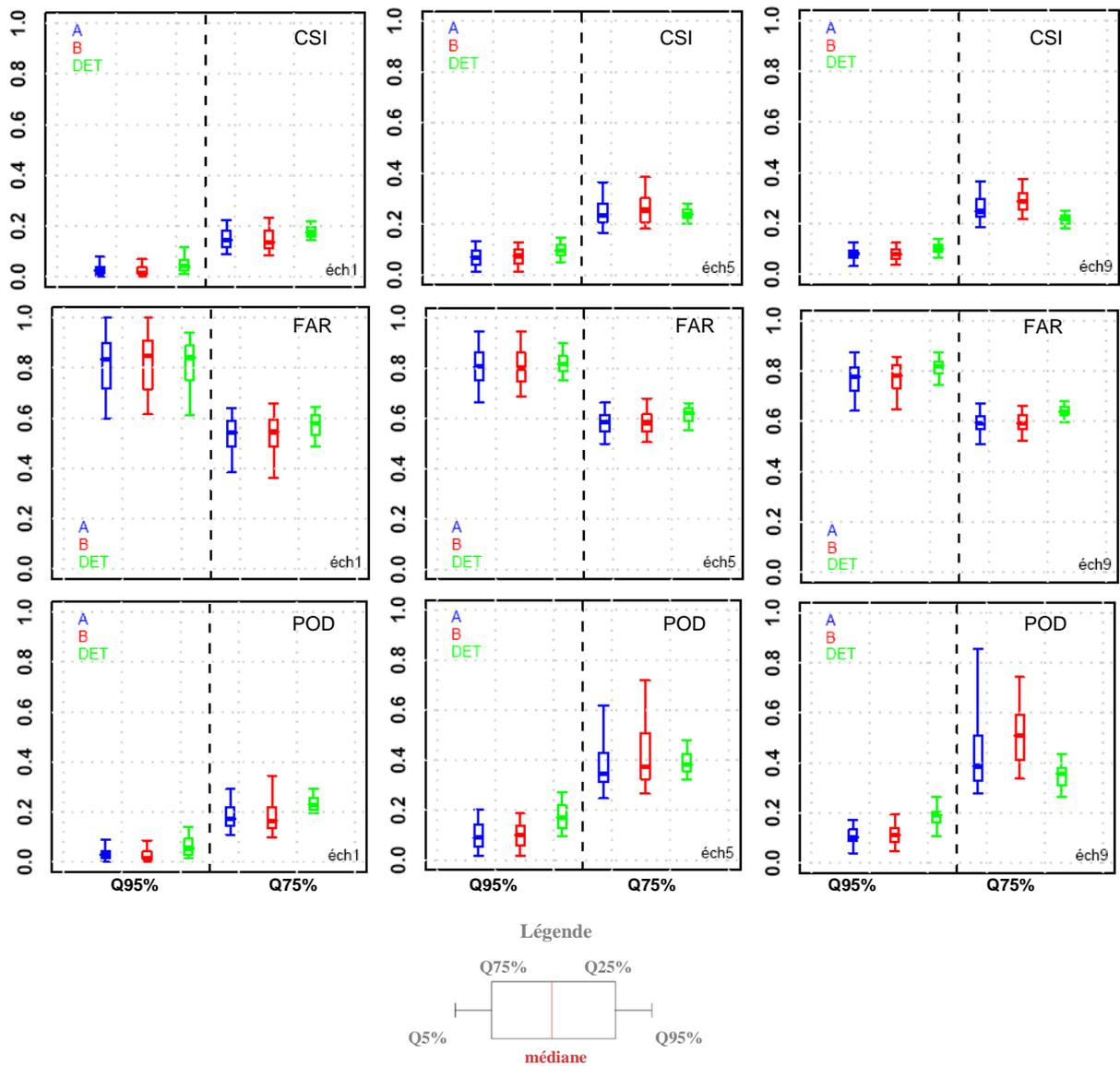


Figure 17. Box-plots des indices de succès critiques (en haut), des ratios de fausses alertes (au milieu) et des taux de succès (en bas) pour les échéances de 1, 5 et 9 jours de la prévision déterministe des précipitations (DET) et pour au moins 80% des membres de la prévision d'ensemble PDH (modèles A et B) sur les 74 bassins considérés et pour les dépassements des quantiles Q95% et Q75%.

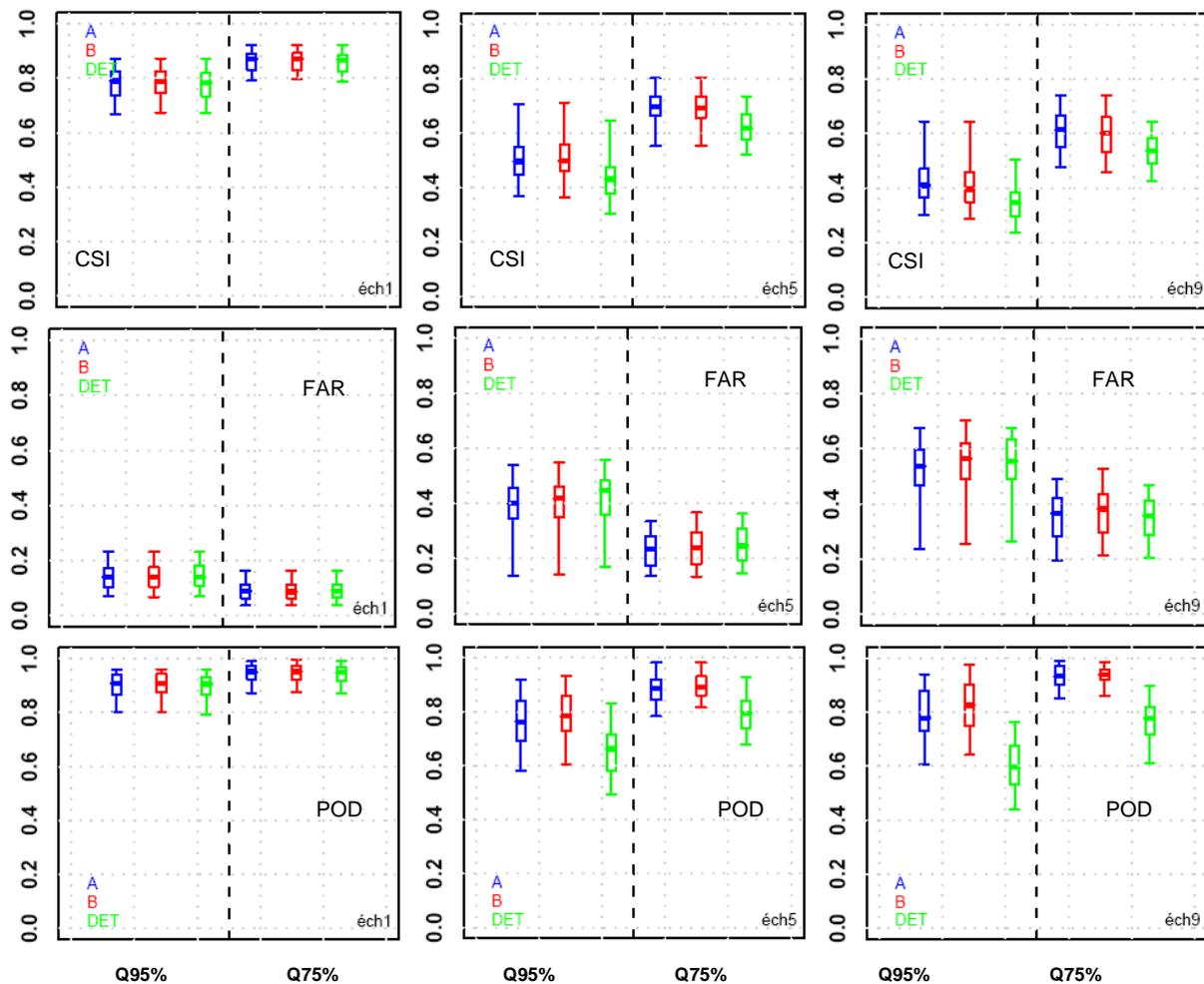


Figure 18. Idem que la figure 17 pour les prévisions d'ensemble des débits.

5.3.3. Les courbes de ROC

Nous avons commencé par étudier le comportement général de l'ensemble des bassins versants étudiés pour le score AUC. Rappelons que ce score est compris entre 0.5 (pour une prévision qui n'a pas de performance) et 1 (score parfait).

La figure 19 représente les boîtes à moustaches (box-plots) des scores AUC des 74 bassins versants étudiés. Les résultats de l'évaluation des prévisions des précipitations (faisceaux construits à l'aide des modèles d'erreurs A et B et prévision d'ensemble *EU*) donnent des résultats intéressants, plus particulièrement pour la 1^{ère} échéance ($j+1$): la meilleure performance des prévisions *PDH* des deux modèles A et B est évidente notamment pour le dépassement du quantile Q75% (un écart de 0.1 entre les *PDH* et les *EU*).

En ce qui concerne les prévisions d'ensemble des débits, les scores sont très élevés (scores médians supérieurs à 0.9 pour l'échéance 1 jour et à 0.8 pour l'échéance 9 jours), avec des comportements très similaires pour les trois systèmes de prévision évalués.

Les courbes de ROC ont été analysées pour chaque bassin versant. Les résultats sont ici présentés à titre d'exemple pour deux bassins versants.

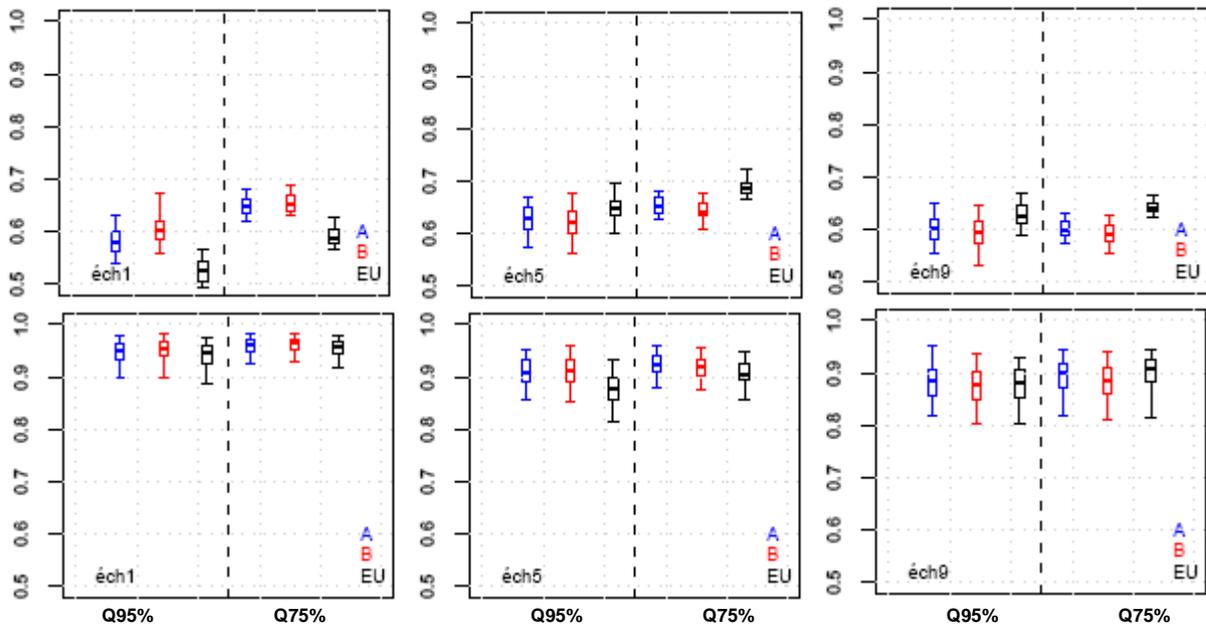


Figure 19. Box-plots des scores AUC sur les 47 bassins versants étudiés pour les prévisions des pluies (PDH et EU) en haut et des débits (QDH et EU) en bas. Les traits discontinus délimitent les seuils des quantiles Q95% et Q75% pour chaque score.

La figure 20 représente les courbes de ROC pour les trois systèmes de prévision des précipitations; *DH*, *EU* et *DET*, pour les échéances 1, 5 et 9 jours pour les bassins versants «L'Armançon à Aisy-sur-Armançon» (forte hypothèse de normalité des erreurs) et «La Midouze à Campagne» (faible hypothèse de normalité). Ces mêmes courbes pour les prévisions des débits sont données dans la figure 21. Nous illustrons avec ces courbes les taux de succès et de fausses alertes quand au moins 1, 2, ..., M membres (dans notre cas $M=51$) dépassent le seuil du quantile Q95%. Rappelons aussi que plus les courbes s'approchent du coin supérieur gauche, meilleure est la prévision (taux de fausses alarmes = 0 et taux de succès = 1).

De manière générale, nous observons une amélioration apportée par les prévisions *PDH*, même si les différences sont faibles. En effet, les courbes de ROC des prévisions *DH* sont légèrement situées au dessus de celles des prévisions *EU*. Remarquons aussi que la prévision déterministe est généralement proche du coin gauche en bas, montrant ainsi ses faibles capacités à prévoir les dépassements de ce seuil élevé. Enfin, même si les courbes de ROC sont moins bonnes pour le 1^{er} bassin (courbes proches de la 1^{ère} bissectrice), elles montrent bien que les performances de prévisions *DH* sont proches de celles des prévisions *EU*.

En visualisant les mêmes courbes pour les prévisions des débits, nous remarquons que, pour la 1^{ère} échéance, les trois systèmes de prévision (*DH*, *EU* et *DET*) sont fortement concentrés sur le coin haut-gauche du graphique, c'est-à-dire les taux de fausses alarmes et de réussite varient peu avec le pourcentage de membres considérés et sont proches entre les prévisions d'ensemble et la prévision déterministe. Ici aussi nous observons l'effet de la mise à jour du modèle hydrologique à partir du dernier débit observé. Les scores de la première échéance (1 jour) montrent des très faibles taux de fausses alertes et des taux de réussites élevés (>0,8). Cependant, ces bons scores peuvent très probablement être dus au très faible nombre de cas impliqués (si le nombre des évènements manqués est très faible, alors le taux de réussite sera élevé mais pas forcément avec un nombre de succès élevé). Les points de la courbe ROC, représentant les différents pourcentages de membres détectant les dépassements du quantile de

débit Q95%, sont beaucoup plus étalés lorsque les échéances augmentent. Globalement, les performances des systèmes de prévision, telles que mesurées par les courbes ROC, diminuent avec l'augmentation des échéances, et notamment pour le système déterministe *DET* et la prévision d'ensemble *EU*.

Les prévisions *QDH* améliorent le pourcentage de membres détectant le dépassement de ce seuil pour les longues échéances (points rouge au dessus des points noirs, i.e. une probabilité de détection plus forte). En revanche, remarquons que les prévisions *EU* présentent des ratios de fausses alertes (POFD) globalement plus faibles, d'où la légère diminution des scores AUC pour les prévisions *DH* à l'échéance 9.

Enfin, même si les courbes de ROC sont légèrement moins bonnes pour le 9^{ème} jour d'échéance, l'information apportée par le système *DH* reste intéressante et améliorée par rapport aux prévisions déterministes *DET* d'origine.

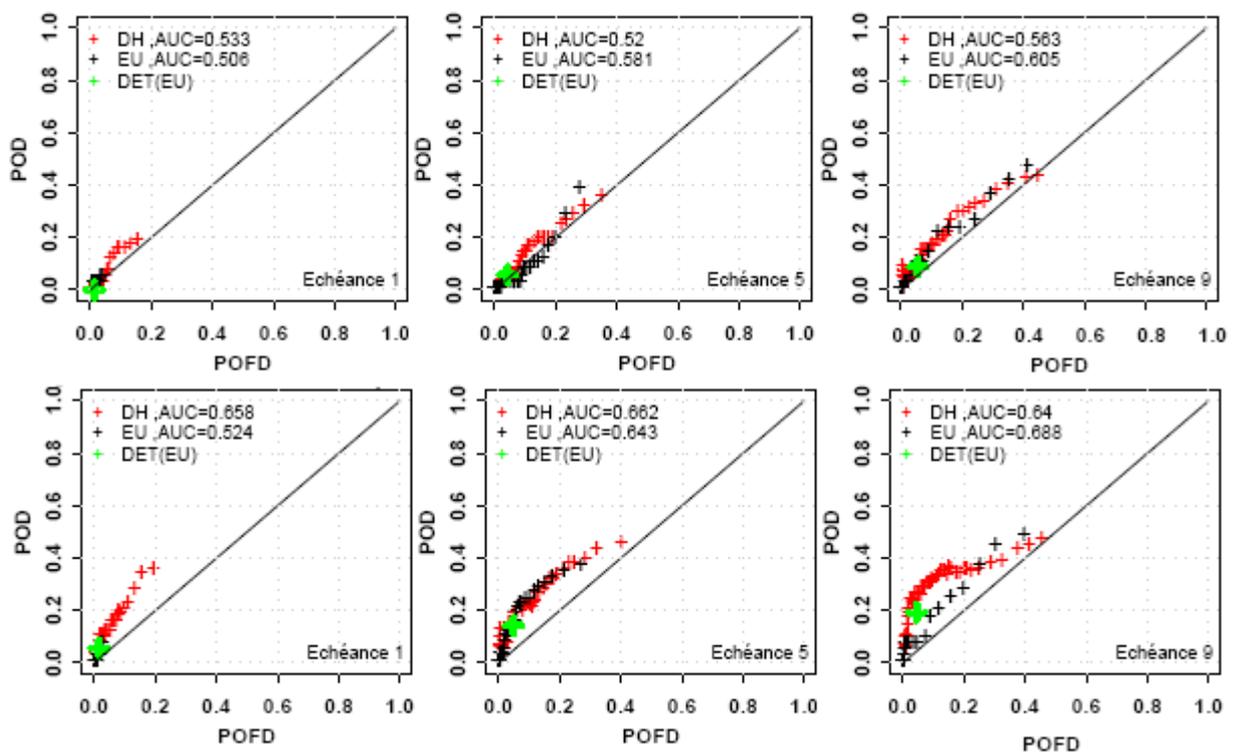


Figure 20. Courbes de ROC pour les prévisions d'ensemble des pluies pour le dépassement du quantile 95% (échéance $j+1$, $j+5$ et $j+9$)

En haut, les courbes de ROC des prévisions PDH (modèle B), EU et DET pour le bassin L'Armançon à Aisy-sur-Armançon et en bas idem pour le bassin La Midouze à Campagne.

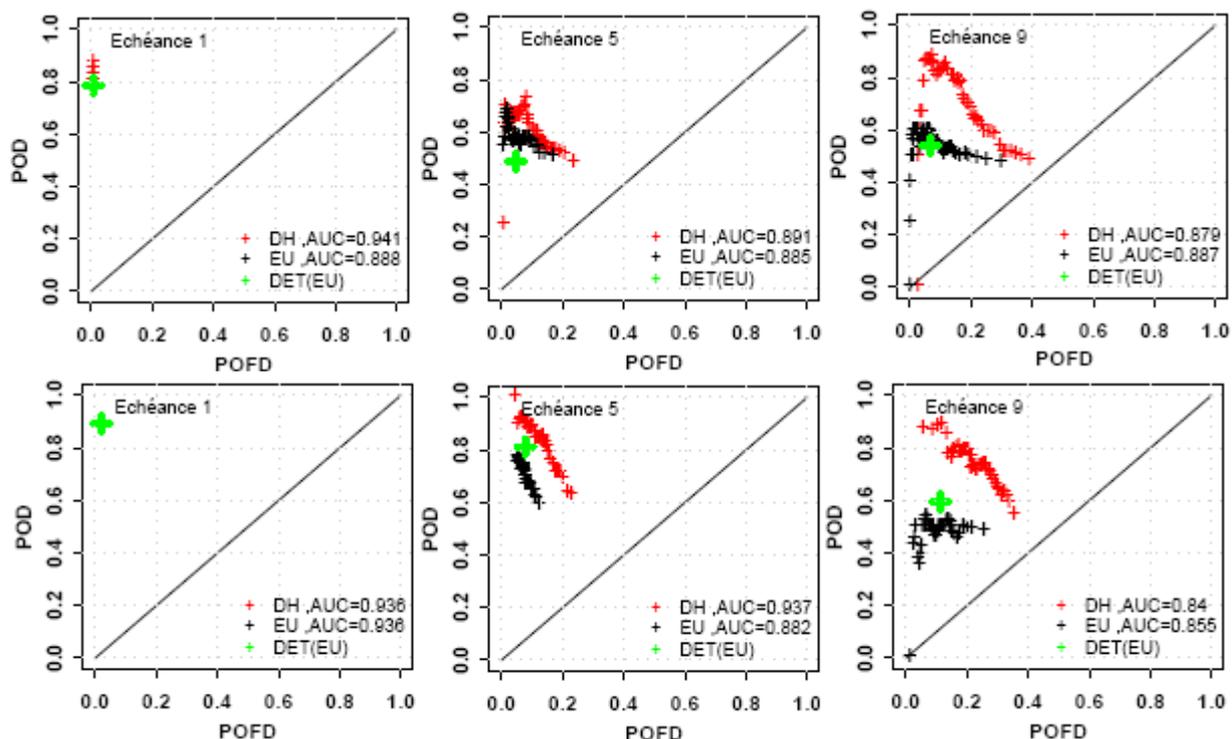


Figure 21. Idem que la figure 13 pour les prévisions d'ensemble des débits.

5.3.4. Un évènement marqué

Afin d'illustrer qualitativement les apports des scénarios de prévisions *QDH*, construits dans cette étude (sorties du modèle GRP forcé par les ensembles *PDH*), nous avons décidé de chercher les évènements marqués dans les chroniques des prévisions de chaque bassin versant. Cette analyse visuelle nous a permis d'apercevoir la capacité de nos ensembles de détecter nombreuses situations de crues.

Nous nous illustrons ici, à titre d'exemple, un évènement de crue perçu le 04 mars 2007 sur le bassin «La Meuse à Commercy» (2327,9 km²) pour le dépassement du seuil critique de débit (ici nous avons choisi un exemple de $Q=7\text{mm}$). En effet, observons sur la figure 22.a la sortie du modèle GRP forcé par la prévision déterministe du CEPMMT pour le 03 avril 2007 et valable jusqu'au 12-03-2007 (i.e. neuf échéances de prévision). A la 1^{ère} échéance ($j+1$) de la prévision, en explorant les différences entre les informations contenues dans les hydrogrammes journaliers, observé (courbe noire) et prévu (courbe marron), nous pouvons constater une cohérence appréciable de ces deux courbes. Le scénario de prévision déterministe s'approche de la pointe de crue, mais ne prévoit pas avec précision le franchissement du seuil critique (marqué par la ligne rouge horizontale sur la figure 22.a).

Examinons maintenant ce même évènement à une échéance de prévision plus lointaine, par exemple pour la prévision émise le 25-02-2007. Nous constatons sur la figure 22.b que le scénario déterministe est assez atténué par rapport à l'hydrogramme observé (courbe noire). Il n'est possible d'évaluer ni la pointe de crue ni le risque de dépassement du seuil à l'échéance 8j de prévision à l'aide d'un tel hydrogramme. Ceci est naturellement lié au caractère chaotique des prévisions déterministes qui trouvent souvent des difficultés à prévoir à longue échéance.

Sur la même figure est aussi représenté le «schéma spaghetti» des scénarios de débits prévus par chaque membre de la prévision d'ensemble *QDH* (courbes bleues). Nous voyons que plusieurs membres sont nettement rapprochés des débits observés, indiquant ainsi une assez forte probabilité de dépassement du seuil critique fixé.

Quoiqu'elles présentent un retard d'environ un jour (les pics sont un peu décalés par rapport à l'observation), probablement dû à l'agrégation temporelle des prévisions déterministes du CEPMMT (cf. chapitre 4), elles font preuve d'une capacité remarquable d'anticiper un tel évènement à 8j d'échéance.

D'autre part, il est important de remarquer que l'allure générale du faisceau de débits prévus *QDH* respecte une dynamique temporelle cohérente entre les échéances de prévision (observons sur la même figure, via les dégradations des couleurs, que les membres *DH* ne présente pas des fluctuations aberrantes d'une échéance à une autre, mais plutôt des signaux continus en fonction du temps). Ce résultat est la conséquence de l'analyse multivariée entre les échéances de prévision pour laquelle nous avons opté et qui représente un des plus importants apports de ce travail; il prend en compte la dynamique temporelle lors de la construction des scénarios contrairement à d'autre méthode (comme la méthode des analogues par exemple).

Sur la même figure en haut sont représentés les hiétogrammes des pluies observées au cours de cet évènement (en bleu clair) et ceux de la valeur médiane des membres de la prévision *PDH* (en bleu foncé). Nous pouvons remarquer que les deux hiétogrammes sont très comparables confirmant ainsi que les scénarios construits sont nettement rapprochés de l'observation sans avoir utilisé cette information (les pluies observées) lors de la construction de ces membres de prévision.

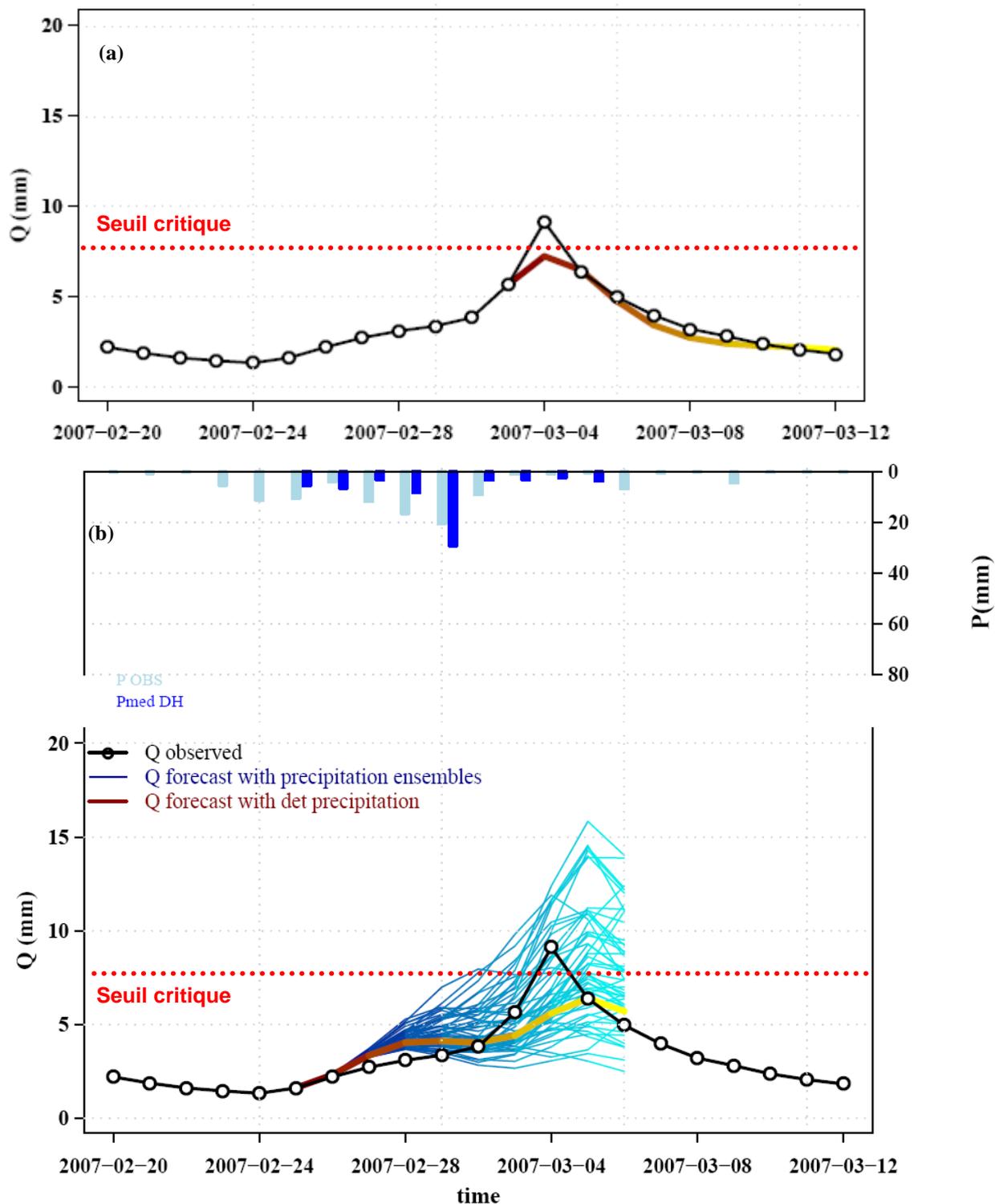


Figure 22. Prévisions réalisées par le modèle GRP lors de l'évènement du 04-03-2007 dans le bassin «La Meuse à Commercy» (2327,9km²) sur 9j d'échéances.

En haut, les hiétogrammes de pluies observées (bleu clair) et de la médiane des membres *PDH* (bleu foncé). En bas, les courbes bleues représentent les membres de l'ensemble *QDH*, la courbe marron illustre la prévision déterministe de GRP (forcé par la prévision *DET* du CEPMMT). Pour ces courbes les dégradations de couleurs s'étalent suivant l'évolution de l'échéance de prévision (couleur foncée à l'échéance 1 et plus clair à échéance 9). La courbe noire indique le débit observé et la ligne rouge horizontale en pointillé le seuil d'alerte fixé.

5.4. Synthèse des principaux résultats

Dans cette partie les performances des prévisions d'ensemble *PDH* et *QDH* ont été évalués en les comparant à un système de prévisions d'ensemble météorologique (du CEPMMT) et à la prévision déterministe du départ (sur laquelle a été appliquée la méthode d'habillage). Les principaux résultats sont :

- en comparant les modèles A et B :

Pour les différents scores calculés sur la période de contrôle des modèles des erreurs sur les 74 bassins versants étudiés (en pluie et en débits), nous avons globalement remarqué des comportements qui semblent être équivalents pour les deux modèles A et B, voire légèrement meilleur (notamment en POD, FAR, CSI et AUC) pour le modèle B. En plus, ce dernier présente une meilleure qualité en calage (hypothèse de normalité vérifiée dans plusieurs bassins versants).

- en les comparant à la prévision *DET* :

Nous avons remarqué que les prévisions *DET* présentent des ratio-RMSE légèrement plus faible que la moyenne des membres du modèle A et du modèle B. En revanche, les prévisions *PDH* et *QDH* ont montré leur capacité à améliorer le taux de succès, le CSI et à diminuer le taux de fausse alerte, même pour le pourcentage assez exigeant du nombre des membres (au moins 80% des membres qui dépassent le seuil) que nous avons illustré, notamment pour les échéances supérieures à 5j et pour le quantile 75%. En outre, nous avons montré que, pour le quantile 95%, la prévision *PDH* et plus particulièrement *QDH* sont plus performantes par rapport à la prévision *DET* (respectivement en pluie et en débit) en termes de capacité de détection des dépassements de ce seuil élevé (courbes ROC),

- en les comparant à la prévision *EU* :

En observant les probabilités de prévision des membres *DH* et *EU* en pluie et en débit, nous avons remarqué que les scores BS étaient légèrement au profit des prévisions *EU*. Nous avons aussi montré que les performances (ou encore l'habilité) des prévisions *DH* sont très comparables à celles des prévisions *EU* (les calculs des scores AUC),

- qualité de la distribution des membres *DH* :

Nous avons, également, pu évaluer la qualité des prévisions *DH* en sélectionnant des événements de crue sur les bassins versants étudiés et montrer que les scénarios générés sont des scénarios réalistes et suivants une dynamique temporelle cohérente qui permettent d'anticiper des dépassements des seuils critiques à plusieurs jours d'échéance. En revanche, un décalage temporel d'environ un jour a été aussi constaté.

En outre, nous avons remarqué, pour les différents scores précédents, l'effet de la mise à jour du modèle GRP sur la première échéance de prévision qui a montré généralement les mêmes résultats pour les systèmes de prévisions évalués.

Chapitre 6: Conclusions et perspectives

"As far as the laws of mathematics refer to reality, they are not certain; as far as they are certain, they do not refer to reality." Albert Einstein (1879 -1955)

Ce travail de stage portait sur les prévisions d'ensemble hydrométéorologiques, axées sur la moyenne échéance pour l'amélioration des prévisions des débits. Dans le présent rapport, nous nous sommes, particulièrement, intéressés à évaluer une méthode d'habillage, d'un seul scénario de prévision déterministe de précipitations, par ses erreurs passées en amplitude.

Pour cela, plusieurs tests et diagnostics ont été effectués afin d'atteindre notre principale objectif : fournir un outil statistique pratique qui améliore les prévisions de débit.

Le travail, ici présenté, s'est donc déroulé en trois grandes étapes : 1) formalisation mathématique et numérique du modèle envisageable, 2) traitement des données et calage du modèle et enfin 3) génération des scénarios de prévision et évaluation des apports de la méthode en fonction des résultats de différents scores ensemblistes (pour comparer deux prévisions d'ensembles) et déterministes (pour comparer une prévision d'ensemble à une prévision déterministe), en pluie et en débit, pour quantifier les apports de cette méthode.

Chacune de ces étapes est le fruit de plusieurs essais, comparaisons et études qui servaient à s'assurer de nos choix avant le passage aux parties suivantes.

Deux modèles statistiques d'erreurs ont été testés en utilisant une archive météorologique du CEPMMT pour la période entre 11-03-2006 et 01-10-2008 (1301j au total) sur 74 bassins versants en France: un modèle général gaussien multivarié via la transformation de Box-Cox, construit avec l'ensemble de paires observations-prévisions de la série disponible pour chaque bassin versant et un modèle alternatif traitant seulement les résidus non-nuls (les pluies observées et prévues sont nulles). Une comparaison statistique des prévisions aux observations sur la période de calage fixée à 570j, a permis de distinguer ces deux différents cas.

Les résultats dépendent de l'échéance de prévision (généralement les résultats s'améliorent pour les échéances élevées) et du bassin versant et ont montré globalement un meilleur calage aux données pour le modèle B aux échéances supérieures à 7j. En contrôle, les scores d'évaluation de la qualité des prévisions ont montré que les deux modèles sont très comparables. De ce fait, il semble être plus pertinent de choisir plutôt le modèle B.

L'importance de cette étude se manifeste, d'une part, du fait que nous avons pu modéliser «correctement» les résidus de prévision en passant par plusieurs diagnostics et transformations de données, d'autre part, du fait que nous avons réussi à générer des scénarios réalistes et cohérents en fonction de l'échéance de prévision, qui sont comparables à ceux du SPE du CEPMMT et assez meilleurs que la prévision déterministe de départ (i.e. sur laquelle la méthode d'habillage a été appliquée). Cette méthode ayant un coût numérique « faible », par rapport à d'autres méthodes de perturbation des prévisions, nous a permis d'améliorer la prévision des crues (dépassement des quantiles 95% et 75%) et d'anticiper des événements marqués à des échéances élevées (avant 7 à 8j de l'évènement).

En perspectives à ce travail, le modèle d'erreurs peut être testé pour différents événements de crue et/ou étiage sévère qui était assez restreint dans notre cas vu la taille limitée de l'archive de prévision dont nous disposons. En outre, l'évaluation de cette méthode peut être détaillée

davantage en utilisant des autres scores de mesure de la qualité ainsi que ceux évaluant la valeur économique des prévisions. D'autre part, ce modèle d'erreurs peut être complété par des modules complémentaires de corrections d'autre type d'erreurs, par exemple, de la répartition temporelle et/ou de localisation géographique et voire son adaptation à d'autres types de prévisions (horaires ou saisonnières par exemple) ou avec d'autres modèles hydrologiques.

Enfin, il reste à signaler qu'en manipulant des prévisions d'ensemble, les variables sont très nombreuses (le pas de temps, le bassin, l'échéance et les membres de la prévision, le seuil de dépassement ou de non dépassement) et la variabilité des résultats en fonction de toutes ces variables rend les tâches d'évaluation et de présentation des résultats beaucoup plus difficile, puisque un exemple ne traduit pas cette grande variabilité. Un autre problème est celui de la comparaison d'un ensemble à un seul scénario; en effet, la moyenne des membres ne reflète toujours pas la réalité et peut sous-estimer le résultat.

Références bibliographiques

“Writing an article a day keeps the Doctor away!”

- Atger F. (1999): The Skill of Ensemble Prediction Systems. *Monthly Weather Review*, 127, 1941-1953
- Atger F. (2001): Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes in Geophysics*, 8, 401-417
- Ben Daoud A. (2010): Améliorations et développements d'une méthode de prévision probabiliste des pluies par analogie. Application à la prévision hydrologique sur les grands bassins fluviaux de la Saône et de la Seine. *Thèse de doctorat de l'Université de Grenoble/ UR Hydrologie-Hydraulique, Cemagref, Lyon*, 340p
- Berthet, L., Janet B. et de Saint-Aubin C. (2011): La prévision opérationnelle des crues en France: organisation, pratiques et ambitions. Présentation orale lors de l'Atelier France-Québec sur les Prévisions Hydrologiques d'Ensemble, Grenoble, 11-13 avril 2011. Disponible sur: http://www.cemagref.fr/webgr/Download/afq/13_AtelierFrance-Quebec_Berthet.pdf
- Beven, K. et Binley, A. (1992): The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes*, 6, 279-298
- Bontron G. (2004): Prévision quantitative des précipitations : adaptation probabiliste par recherche d'analogie. Thèse de doctorat *LTHE/ CNRS-INPG-IRD-UJF*, 286p
- Box G. E. P. and Cox D. R (1964): An Analysis of Transformations. *Journal of the Royal Statistical Society*, B, 26, 211-252
- Brier G.W. (1950): Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1, 1-3
- Cloke H.L., Pappenberger F. (2009): Ensemble flood forecasting: A review. *Journal of Hydrology* 375, 613-626
- Demargne J., Brown J., Liu Y., Seo D.-J., Wu L., Toth Z., Zhu Y. (2010): Diagnostic verification of hydrometeorological and hydrologic ensembles. *Atmospheric Science Letters HEPEX special issue* 11 (2), 114-122
- Demeritta D., Cloke H., Pappenberger F., Thielenc J., Bartholmesc J., Ramos M.-H. (2007): Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environmental Hazards*, 7, 115-127
- Ehrendorfer M. (1997): Predicting the uncertainty of numerical weather forecasts: A review. *Meteorologische Zeitschrift*, Neue Folge 6, 147-183
- Engeland K., Renard B., Steinsland I., Kolberg S. (2010): Evaluation of statistical models for forecast errors from the HBV model. *Journal of Hydrology*, 384, 142-155
- Garçon R., Houdant B., Garavaglia F., Mathevet T., Paquet E., Gailhard J. (2008): Expertise humaine des prévisions hydrométéorologiques et communication de leurs incertitudes dans un contexte décisionnel. *La Houille Blanche*, 5, 71-80
- Haan C.-T. (1977): Statistical methods in hydrology. *The Iowa State University Press/Ames*, 378p
- Hagedorn R., Buizza R., Hamill T.-M., Leutbecher M., Palmer T.-N. (2010): Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. Soumis au *Monthly Weather Review*, le 12/07/2010, 53p, disponible sur: http://www.esrl.noaa.gov/psd/people/tom.hamill/tigge_refcst.pdf
- Hingray B., Picouet C., Musy A. (2009): Hydrologie 2, une science pour l'ingénieur. Chap.3: Principe de la modélisation hydrologique. *Presses Polytechniques et universitaires romandes*, 1^{ère} édition, ISBN 978-2-88074-798-5
- 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 ENREF/EDF-DTG*, 209p.
- Joly A., Descamps L., (2009): La prévision et l'utilisation des incertitudes en météorologie. *La Houille Blanche*, 6, 42-51

- Krzysztofowicz R. (2001b): Integrator of uncertainties for probabilistic river stage forecasting: precipitation dependent model. *Journal of Hydrology*, Amsterdam, the Netherlands, Elsevier, 249, 69-85
- Krzysztofowicz R. (1998): Probabilistic hydrometeorological forecasts: Toward a new era in operational forecasting. *Bulletin of the American Meteorological Society*, 79 (2), 243-251
- Leith C.-E. (1974): Theoretical skill of Monte-Carlo forecasts. *Monthly Weather Review*, 102, 405-418
- Lorenz E. (1969): The predictability of a flow which contains many scales of motion. *Tellus A*, 21, 289-307
- Maidment D. (1993): Handbook of Hydrology. USA, McGraw-Hill, Inc. Chap.17: Statistical analysis of hydrological data, Chap.26: Hydrological forecasting, ISBN 0-07-039732-5
- Mascaro G., Vivoni E.-R., Deidda R. (2010): Implication of ensemble quantitative forecast errors on distributed streamflow forecasting. *American Meteorological Society*, 11, 69-86
- Mason, I. (1982): A model for assessment of weather forecasts. *Australian Meteorological Magazine*, 30, 291-303
- Massey F.-J. (1951): The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46, 253, 68-78
- Nicolle P., Ramos, M.H., Andréassian, V., Valéry, A. (2011): Mieux prévoir les crues navales : évaluation de prévisions probabilistes de débit sur des bassins versants de montagne français. *Proc. Colloque SHF "Eau en montagne"*, Lyon, 16-17 mars 2011, 163-170
- Osborne J.-W. (2010): Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, Volume 15(12), ISSN 1531-7714
- Palmer T., Buizza R. (2007): Fifteenth anniversary of EPS. *ECMWF Newsletter 114*, 14
- Palmer T.-N., Barkmeijer J., Buizza R., Klinker E. et Richardson D. (2002): L'avenir de la prévision d'ensemble. *La Météorologie*, 36, 22-30
- Pappenberger F., Bartholmes J., Thielen J., Cloke H.-L., Buizza R., Roo A.-D (2008): New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophysical Research Letters*, 35, 1-7
- Perrin C., Michel, C., Andréassian, V. (2003): Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279 (1-4), 275-289
- Perrin C., Michel, C., Andréassian, V. (2001) Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242, 275-301
- Persson A., Grazzini F. (2007): User Guide to ECMWF forecast products. Version 4.0, *ECMWF*
- Ramos M-H, Mathevet T., Thielen J., Pappenberger F. (2010): Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Royal Meteorological Society. Meteorol. Appl.* 17, 223-235
- Randrianasolo A., Ramos M.H., Andreassian V. (2011): Hydrological ensemble forecasting at ungauged basins: using neighbour catchments for model setup and updating. *Adv. in Geosciences*, 29, 1-11
- Randrianasolo A., Ramos, M-H, Thirel, G., Andréassian V., Martin, E. (2010): Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmospheric Science Letters*, 11, 100-107
- Rousseau D. (2001): Un critère général d'évaluation des prévisions déterministes et probabilistes Application aux prévisions locales diffusées sur Internet. *La Météorologie*, 35, 48-60
- Sakia R. M. (1992): The Box-Cox transformation technique: a review. *The Statistician*, 41, 169-178
- Schaake J., T. Hamill R. Buizza and C. M. (2007): HEPEX: the Hydrometeorological Ensemble Prediction EXperiment. *Journal of HydroMeteorology*, 88, 1541-1547
- Schaeffli B., Talamba D.-B., Musy A. (2007): Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology*, 332, 303-315
- Seo D.-J., Perica S., Welles E., Schaake J.-C. (2000): Simulation of precipitation fields from probabilistic quantitative precipitation forecast. *Journal of Hydrology*, 239, 203-229
- Shapiro S.-S. and Wilk M.-B. (1965): An analysis of variance test for normality (complete samples). *Biometrika* 52, 591-611
- Shrestha D.-L., Pagano T., Wang Q., Robertson D. (2011): Application of Ensemble Dressing for Hydrological Applications. *Geophysical Research Abstracts*, résumé soumis au «European

- Geosciences Union General Assembly», Vienne (Autriche), 3-8 avril 2011, disponible sur: <http://meetingorganizer.copernicus.org/EGU2011/EGU2011-5397.pdf>
- Tangara M. (2005): Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global. *Thèse de doctorat EPHE/Cemagref, Paris*, 374 p
- Thirel G. (2009) : Amélioration des prévisions d'ensemble des débits sur la France de SAFRAN-ISBA-MODCOU, *thèse de doctorat de l'Université de Toulouse, Sciences de l'Univers, de l'Environnement et de l'Espace / CNRM/GAME Météo-France/CNRS*, 252p
- Thyera M., Kuczera G., Wang Q.J. (2002): Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. *Journal of Hydrology*, 265, 246-257
- Van Pham T. (2011): Tracking the uncertainty in streamflow prediction through a hydrological forecasting system. Mémoire de Master, *University of Twente, Cemagref Antony*, 74p
- Velázquez J. A., Anctil F., Ramos M. H., and Perrin C. (2011): Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, 29, 33-42
- Vincendon B., Ducrocq V., Nuissier O. and Vié B. (2011): Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting. *Natural Hazards Earth System Sciences*, 11, 1529-1544
- Wejís S.V., Schoups G., Van De Giesen (2010): Why hydrological predictions should be evaluated using information theory. *Hydrology and Earth System Science*, 14, 2545-2558
- Wilks D. S. (2006): Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorological Application*, 13, 243-256
- Wilks D.S. (1995): *Statistical Methods in the Atmospheric Sciences*. USA, *Academic Press*, chap.7: Forecast verification, ISBN 0-12-751965-3
- Wu L., Seo D.-J., Demargne J., Brown J.-D., Cong S. and Schaake J. (2011): Generating of ensemble precipitation forecast from single-valued quantitative precipitation for hydrological ensemble prediction. *Journal of Hydrology*, 399, 281-298
- Yang J., Reichert P., Abbaspour K.-C., Yang H. (2007): Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference. *Journal of Hydrology*, 340, 167-182
- Yuan H., Mullen S.-L., Gao X., Sorooshian S. (2005): Verification of Probabilistic Quantitative Precipitation Forecasts over the Southwest United States during Winter 2002/03 by the RSM Ensemble System. *American Meteorological Society*, 133, 279-294

Annexes : Liste des bassins versants étudiés

Tableau A.1. Liste et nomination des bassins versants étudiés

N°	BV code de la BANQUE HYDRO	Surface (Km ²)	Cours d'eau	N°	BV code de la BANQUE HYDRO	Surface (Km ²)	Cours d'eau
1	A2280350	5248	L'Ill à Strasbourg [Chasseur Froid, La Robertsau]	38	K1341810	2271,1	L'Arroux à Rigny-sur-Arroux
2	A5110610	1976,5	La Moselle à Tonnoy	39	K1391810	3181,9	L'Arroux à Digoïn
3	A5500610	3074,4	La Moselle à Pont-Saint-Vincent	40	K1773010	1467,5	L'Aron à Verneuil
4	A5730610	3345,7	La Moselle à Toul	41	K2981910	1522,4	La Dore à Dorat
5	A6761010	2293,6	La Meurthe à Damelevières	42	K6492510	2296,9	La Sauldre à Selles-sur-Cher
6	A6921010	2787	La Meurthe à Laneuveville-devant-Nancy	43	K7312610	1706,6	L'Indre à Saint-Cyran-du-Jambot
7	A7010610	6834,9	La Moselle à Custines	44	L4220710	1235,3	La Grande Creuse à Fresselines
8	A7930610	9384,4	La Moselle à Hauconcourt	45	M0250610	2713,1	La Sarthe à Neuville-sur-Sarthe [Montreuil]
9	A8431010	1241,3	L'Orne à Rosselange	46	M0421510	1936	L'Huisne à Montfort-le-Gesnois [la Pécardière]
10	A9221010	1760,3	La Sarre à Sarreinsming	47	M0680610	7523	La Sarthe à Saint-Denis-d'Anjou [Beffes]
11	B1340010	1728,4	La Meuse à Vaucouleurs [Chalaines]	48	M1041610	1080,5	Le Loir à Saint-Maur-sur-le-Loir
12	B2130010	2327,9	La Meuse à Commercy	49	M1341610	6150,1	Le Loir à Flée [Port-Gautier]
13	B3150020	3914,6	La Meuse à Stenay	50	M1531610	7918,2	Le Loir à Durtal
14	B4631010	1978,3	La Chiers à Carignan	51	M3340910	2907,9	La Mayenne à l' Huisserie [Bonne]
15	B5600010	7796,6	La Meuse à Montcy-Notre-Dame	52	M3600910	3934,8	La Mayenne à Château-Gontier
16	B6111010	1338,6	La Semoy à Haulmé	53	M3851810	1317,4	L'Oudon à Segré [Écluse de Maingué]
17	H2221010	3819,8	L'Yonne à Gurgy	54	O3401010	2142,9	Le Tarn à Millau [2]
18	H2452020	1349,5	L'Armançon à Aisy-sur-Armançon [aval]	55	O5292510	1603,5	L'Aveyron à Laguépie [1]
19	H2482010	2979,8	L'Armançon à Briennon-sur-Armançon	56	O7101510	1158,2	Le Lot à Banassac [La Mothe]
20	H3201010	2302	Le Loing à Châlette-sur-Loing	57	P7001510	1859,4	L'Isle à Bassilac [Charrieras]
21	H3621010	3916,7	Le Loing à Épisy	58	P7041510	2111	L'Isle à Périgueux
22	H4042010	1857,4	L'Essonne à Ballancourt-sur-Essonne	59	P7261510	3756,7	L'Isle à Abzac
23	H5172010	2109,1	La Saulx à Vitry-en-Perthois	60	P8312520	1911,6	La Dronne à Bonnes
24	H5920010	43824,7	La Seine à Paris [Austerlitz après création lacs]	61	Q2593310	2477,6	La Midouze à Campagne
25	H6221010	2887,6	L'Aisne à Givry	62	Q3464010	1144,4	Le Luy à Saint-Pandelon
26	H7162010	1637,4	La Serre à Nouvion-et-Catillon	63	S2242510	1678,1	L'Eyre à Salles
27	H7201010	3287,7	L'Oise à Condren	64	U0610010	3761,3	La Saône à Ray-sur-Saône

28	H7401010	4320,3	L'Oise à Sempigny	65	U1084010	2070,6	L'Ognon à Pesmes
29	H7742020	1210,2	Le Thérain à Maysel	66	U1244040	1263	La Tille à Champdôtre
30	H8042010	1385,6	L'Epte à Fourges	67	U2142010	1306,4	Le Doubs à Glère [Courclavon]
31	H9121010	2028	L'Eure à Charpont	68	U2222010	2235,6	Le Doubs à Mathay
32	H9331010	4560,6	L'Eure à Cailly-sur-Eure	69	U2354010	1109,1	L'Allan à Courcelles-lès-Montbéliard
33	H9501010	5891,3	L'Eure à Louviers	70	U2402010	3556,3	Le Doubs à Voujeaucourt
34	J3821810	1081,2	L'Aulne à Gouézec [Pont-Coblant]	71	U2512010	4658,1	Le Doubs à Besançon
35	I5231010	1397,9	La Vire à Montmartin-en-Graignes	72	U2634010	1264	La Loue à Champagne-sur-Loue
36	J7500610	3300,7	La Vilaine à Guichen [Le Boël]	73	U2722010	7346,2	Le Doubs à Neublans-Abergement
37	J8502310	2453,9	L'Oust à Saint-Gravé [Echuse Le Guélin]	74	V4264010	1127,8	La Drôme à Saillans

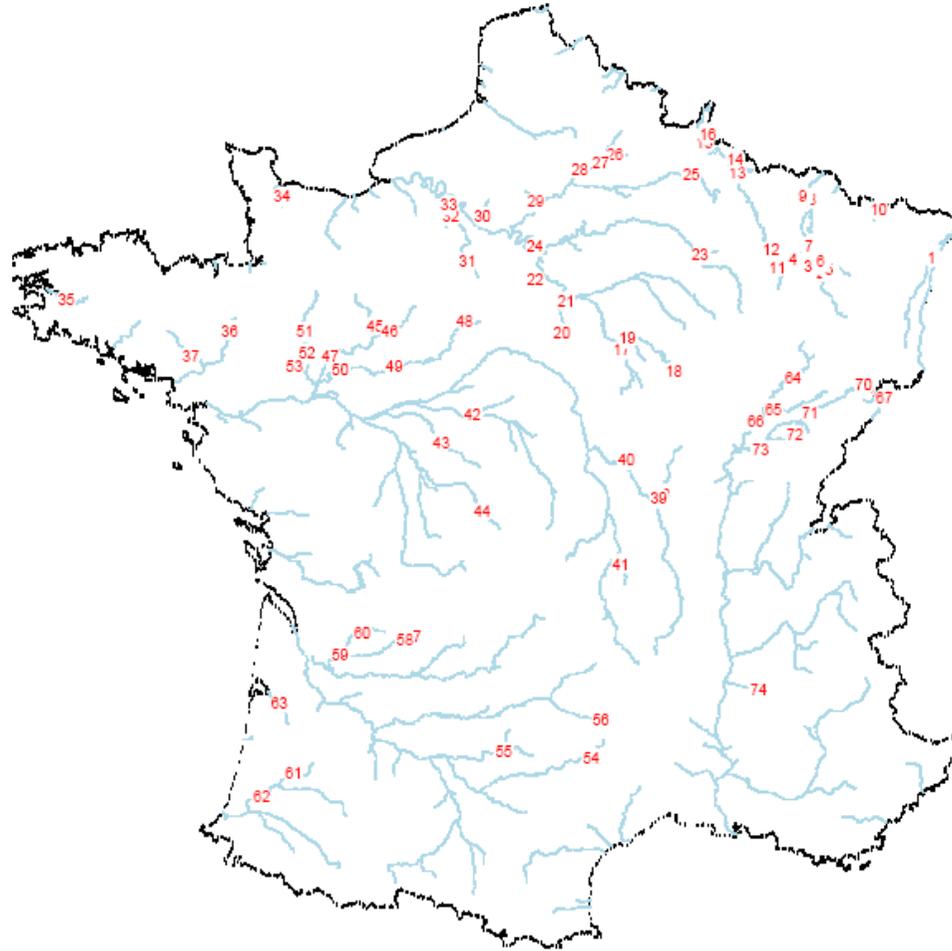


Figure A1. Carte de la localisation des différents bassins versants étudiés selon la numérotation du tableau ci-dessus.