

**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-HYDROGEOLOGIE

Evaluation de la qualité des prévisions pour l'alerte aux crues

Rindra Annie RANDRIANASOLO

Encadrement : Maria-Helena Ramos



**CEMAGREF,
Parc de Tourvoie
92163 Antony**

Juin 2009



Abstract

At operational flood forecasting centres, forecasters usually have to deal with forecasts issued by different models and combine them to support their decisions and communicate flood alerts to end-users. The quality of the forecast is an important aspect of a forecasting system. In this report, a comparative analysis is conducted to assess the quality of streamflow forecasts issued by two different modelling conceptualizations of catchment response, both driven by the same weather ensemble prediction system. Weather forecasts come from the ensemble prediction system PEARP of Météo-France, which is based on the global spectral ARPEGE model zoomed over France. The model runs 11 perturbed members for a forecast range of 60 hours. The two hydrological modelling approaches used are: 1) the coupled physically-based hydro-meteorological model SAFRAN-ISBA-MODCOU (SIM) developed at Météo-France and based on a fully distributed catchment model, and 2) the GRPE forecasting system developed at Cemagref and based on a lumped soil-moisture-accounting type rainfall-runoff model. A tool is developed, comprising several scores usually used for forecast verification. The study is conducted on a set of 211 gauging stations representative of a wide range of upstream catchment areas and hydro-meteorological conditions in France. The discharges simulated by both systems are compared with observed discharges over a 17-month period (March 2005-July 2006). Skill scores are computed for the first two days of forecast range and the performance of both hydrologic ensemble forecasting systems is assessed.

This study shows that the PEARP-based ensemble streamflow forecasts have a good skill for short range prediction of discharges over the studied catchments. Better scores were obtained from the GRPE model with updating, while SIM results are closer to the results obtained with the GRPE model without updating. Although GRPE scores are better, we note that its hydrological ensemble has a lower spread. Ways for improvement remain open.

Keywords: ensemble forecast, forecast verification, updating, hydrological model

Remerciements

Un proverbe malgache dit "ny trano tsy vita ra tsy hifanakonana", ce qui signifie littéralement, une bâtisse ne peut être construite par une seule personne. Je crois que cette citation tient tout son propos lors de l'élaboration de ce rapport mais aussi lors de ce stage au Cemagref. En effet, il s'est construit autour de nombreux acteurs que je tiens ici à remercier.

D'abord merci à mon encadreur, Maria- Helena Ramos, pour le temps consacré à m'initier à la programmation, ce qui n'était pas gagné d'avance. Plus sérieusement, je lui adresse ma profonde reconnaissance pour toutes les connaissances acquises durant ce stage, pour sa disponibilité, sa méthodologie et sa rigueur.

Merci aussi à Vazken Andreassian et Charles Perrin pour leurs conseils avisés et leurs remarques constructives.

J'adresse également ma gratitude envers Météo-France, qui a fourni les données. Merci aussi à Guillaume Thirel pour son aide, ses explications et pour les échanges par e-mail qui m'ont fait avancer dans mon travail.

Je ne saurais oublier les autres membres de l'équipe Hydrologie, à Lionel Berthet pour sa formation sur R, à Audrey Valery pour les données de neige.

Finalement, un grand merci à toute l'équipe pour leur participation à la réunion de mi-parcours, ils m'ont aidé à progresser dans mon travail. Merci pour l'ambiance de travail qui règne au Cemagref.

Table des matières

AVANT-PROPOS	6
INTRODUCTION	8
CONTEXTE ET OBJECTIFS DU STAGE	8
PREMIERE PARTIE: SYNTHESE BIBLIOGRAPHIQUE	10
I-1 Généralités sur la modélisation hydrologique.....	10
I-2 La prévision d'ensemble. Qu'est-ce que c'est?	10
I-3 Vérification des prévisions d'ensemble	12
DEUXIEME PARTIE: ASPECTS METHODOLOGIQUES ET DONNEES.....	13
II-1 Méthodologie générale.....	13
II-2 Les données.....	13
II-2-1 Les 211 bassins versants étudiés	13
II-2-2 Les prévisions de pluie	14
II-2-3 Les prévisions de débits.....	15
II-3 Les scores de vérification.....	17
II-3-1 RMSE et ratio-RMSE.....	18
II-3-2 Ecart-type et ratio de l'écart-type.....	19
II-3-3 Le tableau de contingence	19
II-3-4 La persistance	20
II-3-5 Le Brier Score et le Brier Skill Score	21
II-3-6 Le Discrete Ranked Probability Score (DRPS) et le Skill Score associé (RPSS)	22
II-3-7 Le diagramme de Talagrand (ou histogramme de rang).....	23
II-4 L'outil de vérification développé	24
TROISIEME PARTIE : RESULTATS ET INTERPRETATIONS	25
III-1 Pluies et débits prévus	25
III-2 Erreurs sur les précipitations prévues par PEARP	29
III-3 Scores sur les débits prévus	31
III-3-1 Cartographie des erreurs	31
III-3-2 Dispersion : écart-type et ratio de l'écart-type.....	32
III-3-3 FAR, POD et BIAS : tableau de contingence	34
III-3-4 Brier score et Brier Skill score	37
III-3-5 Le RPSS	38
III-3-6 Diagramme de Talagrand.....	39
III-3-7 Synthèse des scores moyens.....	40
III-3-8 Les principaux résultats.....	42
CONCLUSION	44

Liste des figures

Figure 1 : Schéma général du système de prévision de débits GRPE.....	9
Figure 2 : Schéma général de l'approche traitée dans ce travail	9
Figure 3 : Carte de localisation des 211 bassins étudiés	14
Figure 4 : Représentation graphique des échéances de prévision	15
Figure 5 : Schéma du modèle hydrologique de prévision GR3P (Tangara, M. 2005).....	16
Figure 6 : Schéma du modèle SIM en prévision (<i>Rousset-Regimbeau et al.; 2008</i>)	17
Figure 7 : Données du bassin A1310310: L'Ill à Ensisheim, 1038 km ² pour les mois d'Avril et Octobre. Echéance 1.....	26
Figure 8 : Données du bassin M0500610: la Sarthe à Spay, 5285 km ² pour les mois d'Avril et d'Octobre. Echéance 1.....	27
Figure 9 : Hydrogrammes des débits prévus et observés du bassin K2981910: la Dore à Dorat, 1523 km ² pour les mois de Mars et Avril	28
Figure 10 : Evolution du débit avec un module neige sur GR.....	29
Figure 11 : Cartes des RMSE (haut) et des ratio-RMSE (bas) des pluies sur les deux échéances .	30
Figure 12 : Boxplots des ratio-RMSE des pluies (gauche) et des ratio-écart-types des pluies (droite).....	31
Figure 13 : Valeurs des ratio-RMSE des débits des trois modèles en échéance 1 (haut) et échéance 2 (bas).....	31
Figure 14 : Relation entre les Ratio RMSE P et Ratio RMSE Q.....	32
Figure 15 : Ecart-type des débits en échéance 1 et 2 (haut) et ratio-écart-type (bas) à gauche; mêmes scores pour le calcul en grille avec GR (colonne de droite).....	33
Figure 16 : POD, FAR BIAIS en fonction de la surface du bassin versant pour le seuil de référence de débit Q50 et différents pourcentages de membres de la prévision d'ensemble dépassant ce seuil : p1=20% (gauche), p2=50% (centre) et p2=80% (droite). Echéance jour 1 et Modèle GR.....	34
Figure 17 : POD, FAR BIAIS en fonction de la surface du bassin versant pour différents seuils de référence de débit : Q10 (gauche), Q50 (centre) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant (au non-dépassant pour Q10) le seuil de 50%. Echéance jour 1 et Modèle GR.	35
Figure 18 : Comparaison entre les modèles : POD, FAR BIAIS en fonction de la surface du bassin versant pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres 80% de la prévision d'ensemble dépassant le seuil. Echéance jour 1 et Modèle GR (haut), Modèle GR sans mise-à-jour (bas), Modèle SIM (centre).	35
Figure 19 : Comparaison entre les modèles : boxplots des valeurs de BIAIS pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant le seuil de 50%. Echéance jour 1.	36
Figure 20 : Comparaison entre les modèles : boxplots des valeurs de FAR pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant le seuil de 50%. Echéance jour 1.	36
Figure 21 : Brier score pour les seuils Q50 (haut) et Q90 (bas) sur les deux échéances	37
Figure 22 : Brier Skill Score de référence climatologique pour les seuils Q50 (haut) et Q90 (bas) pour les deux échéances	38
Figure 23 : RPSS pour les deux échéances et les trois modèles par rapport à la référence climatologique.....	39
Figure 24 : Diagrammes de Talagrand des bassins M0243010: l'Orne Saosnoise à Montbizot, A1080330: l'Ill à Didenheim, à partir des débits prévus par GR.	40

Avant-propos

Ce rapport est le fruit d'un stage de recherche de six mois effectué au Cemagref Antony. Ce stage s'est déroulé au sein de l'unité de recherche Hydrosystèmes et Bioprocédés, dans l'équipe Hydrologie des Bassins Versants. Cette équipe s'intéresse aux relations bassin-pluie-débit et le stage s'inscrit dans cette optique, sur les événements extrêmes et leurs prévisions. En prévision hydrologique, l'évaluation de la qualité des prévisions est l'un des principaux défis à relever et une question d'actualité dans les services de prévisions des crues et au sein du Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations (SCHAPI), avec lesquels le Cemagref collabore depuis plusieurs années.

Présentation du CEMAGREF

Issu du regroupement d'équipes scientifiques et techniques constituées par l'État, dans les années 1960-1970, en appui aux politiques nationales d'aménagement et d'équipement de l'espace rural, le Cemagref a été créé en 1981 et a acquis le statut d'établissement public à caractère scientifique et technologique (EPST) en 1986. Depuis lors, la tutelle d'État est confiée conjointement au ministre chargé de la recherche et au ministre chargé de l'agriculture.

Le Cemagref est un organisme public de recherche finalisée. Il a pour mission de répondre à des questions concrètes de société dans le domaine de la gestion durable des eaux et des territoires en produisant des connaissances nouvelles et des innovations techniques utiles aux gestionnaires, aux décideurs et aux entreprises. Il centre ses recherches sur les ressources en eau de surface, les systèmes écologiques aquatiques et terrestres, les espaces à dominante rurale, les technologies pour l'eau, les agro-systèmes et la sûreté des aliments dans la perspective de la gestion durable des eaux et des territoires. Le Cemagref compte 9 centres en France métropolitaine et un centre en Martinique. Il englobe également 21 unités de recherche propres, 6 unités mixtes de recherche (UMR) et une équipe de recherche technologique (ERT). La cible fixée à l'établissement par l'État est de devenir la référence en sciences pour l'ingénierie de la gestion durable des eaux et des territoires. Son domaine de recherche est l'environnement aux échelles intermédiaires, décliné à travers le triptyque "eaux continentales, écotecnologies, territoires".

L'équipe Hydrologie du centre d'Antony

Le centre d'Antony possède 3 unités de recherche dans les domaines de compétences de l'eau (unité de recherche HBAN), du froid (GPAN) et des agroéquipements (TSAN). Les travaux de recherche de l'unité « Hydrosystèmes et bioprocédés » (HBAN) sont appliqués à des domaines aux enjeux environnementaux forts. Ils portent sur le fonctionnement hydrologique des bassins versants, sous l'angle des relations bassins versants-pluies-débits, ainsi que sur l'impact des aménagements et des pratiques agricoles.

L'équipe Hydrologie des Bassins Versants s'occupe de l'étude des transferts d'eau au sein d'un bassin versant. Elle mène depuis plusieurs décennies des travaux de recherche axés sur la modélisation hydrologique pluie-débit et ses applications. Entre autres, les travaux de l'équipe ont abouti au développement d'une famille de modèles conceptuels, les modèles du Génie Rural GR (<http://www.cemagref.fr/webgr>). Sur la base de la modélisation GR à 4 paramètres (GR4J, Perrin, 2000 ; Perrin *et al.*, 2001, 2003), des travaux de recherche menés par C. Michel ont pu mettre en évidence une structure également parcimonieuse, adaptée à la prévision de crues, le modèle GR3P (Tangara, 2005), qui équipe aujourd'hui, en phase expérimentale, plusieurs services de prévision des crues (SPC) en France. Le modèle GR3P a été récemment adapté pour utiliser en entrée plusieurs scénarios de la prévision d'ensemble de pluies. Une première étude a été effectuée sur le

cas de la crue de la Seine en mars 2001 (Ramos *et al.*, 2008). Le stage ici rapporté contribue au développement de cette thématique au sein de l'équipe.

Le SCHAPI (source: MEEDM)

Le SCHAPI, Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations, a été 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 Mer (MEEDM), le SCHAPI est implanté à Toulouse pour favoriser les synergies avec Météo-France et les équipes scientifiques qui y sont rassemblées. Il réunit des experts en météorologie et en hydrologie. Ses principales missions consistent en l'appui aux services de prévision des crues au niveau national, ainsi qu'en une veille hydrométéorologique 24 heures sur 24 localisée sur les bassins rapides et des tronçons de vigilance. Le SHAPI met alors à disposition sur le web une carte de vigilance crues ou "vigicrues". Cette carte se présente sous une forme nationale ou plus locale sur une zone concernée. Elle est accompagnée de bulletins d'informations en temps réel sur la chronologie et l'évolution des crues. Chaque cours d'eau est visible sur cette carte, divisé en plusieurs tronçons. Chaque tronçon est affecté d'une couleur allant du vert (pas de vigilance) au rouge (risque de crue majeure), selon le niveau de vigilance nécessaire pour faire face aux inondations susceptibles de se produire dans les 24h à venir. L'expertise météorologique de Météo-France, contenue dans la vigilance météorologique, et l'expertise hydrologique de la Direction de l'Eau, contenue dans la vigilance crues, permettent de prendre en compte l'apparition et les conséquences de l'événement "pluie-inondation" pour une meilleure qualification du risque¹.

Les SPCs²

Les services de prévision de crues sont issus de la réorganisation de l'annonce de crues, décrite dans la circulaire du 1er Octobre 2002 du Ministère de l'écologie et du développement durable. Ce sont les premiers interlocuteurs des collectivités dans la prévision des crues afin de mieux anticiper les phénomènes. Le SPC les conseille et veille à la cohérence des outils et des méthodes employés. 22 services de prévision des crues (SPC) ont ainsi été créés pour assurer une mission de prévision des crues sur les cours d'eau gérés par l'Etat. Chaque SPC est rattaché à un territoire d'intervention ayant une logique de bassin versant. La surveillance des cours d'eau et la prévision sont les principales missions des SPC. Chaque SPC effectue la capitalisation des informations sur les crues des cours d'eau surveillés de la zone qu'il couvre. En liaison avec le SCHAPI, les SPC mettent à disposition et diffusent, en continu, des messages d'information individualisés par tronçons de cours d'eau et incluant notamment des prévisions d'évolution de la situation.

¹<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000229756&dateTexte>

² Source: www.ecologie.gouv.fr/IMG/pdf/44SPC.pdf

Introduction

L'inondation est une submersion rapide ou lente, d'une zone pouvant être habitée, et fait souvent référence au débordement d'un cours d'eau lors d'une crue. En France, les inondations, de plus en plus fréquentes et dévastatrices, sont devenues un véritable fléau ces dernières années pour les biens matériels et économiques, mais surtout pour la population touchée, en provoquant, dans certains cas, la perte de vie humaine. Ils sont de loin le premier risque naturel, avec 54 672 événements répertoriés sur 24 269 communes depuis 20 ans³. Dans le monde, les inondations causent plus de 50% des catastrophes naturelles avec, en moyenne, 20 000 morts/an (*IFEN, 2008*). Dans ce contexte, la prévision des crues et l'alerte précoce sont des solutions possibles pour diminuer les pertes et se prémunir de ces risques naturels. Une bonne qualité des prévisions devient alors primordiale. Les questions suivantes se posent : Quels sont les avantages de la prévision d'ensemble de pluies (plusieurs scénarios futurs) pour la prévision hydrologique ? Qu'apporte cette approche probabiliste de plus que la prévision déterministe (1 unique scénario futur) ? Comment vérifier la qualité des prévisions ? Comment les noter de manière objective et quels critères (ou scores) utiliser ?

Ce rapport est divisé en trois parties. Initialement, le contexte et les objectifs du stage sont décrits. Ensuite, une synthèse bibliographique essaye de balayer l'état des lieux des connaissances actuelles sur le sujet. Elle contient également quelques notions de base sur le sujet traité. Dans la deuxième partie, les données exploitées, ainsi que les méthodes employées durant le stage sont décrites et développées pour permettre une meilleure compréhension des calculs à venir. Chaque score programmé pour apprécier la qualité des prévisions est explicité. Les intérêts des calculs effectués et l'originalité des informations apportées par chaque score y sont aussi étayés. La mise en place de l'outil d'évaluation de la qualité des prévisions (également appelé ici, outil de vérification) y est aussi décrite. Finalement, dans la troisième partie, les résultats majeurs sont présentés, illustrés et interprétés. De là, les conclusions et les perspectives sont énoncées.

Contexte et objectifs du stage

Le stage s'inscrit dans la thématique des travaux menés au Cemagref dans le cadre du projet AVI (Aide à la Vigilance) de mise en place d'un outil d'aide à la vigilance pour le SCHAPI, afin d'améliorer l'anticipation aux événements de crues. En particulier, il est traité ici le volet portant sur l'utilisation des données de prévision météorologique d'ensemble dans la modélisation hydrologique des crues. Il s'agit d'un travail de coopération scientifique et d'échange d'expertise entre Météo-France, le SCHAPI et le Cemagref d'Antony autour de recherches en prévision hydrologique d'ensemble, avec une valorisation opérationnelle pour la prévision des crues. Au Cemagref, un système de prévision d'ensemble GRPE (GR pour le modèle pluie-débit utilisé et PE pour la prévision d'ensemble, Fig.1) est en développement afin de mieux exploiter les avantages de la prévision d'ensemble. Le stage traite uniquement la construction et le test du module de vérification de la branche "prévision en mode rétrospectif" du système GRPE. La prévision en temps réel fait l'objet d'une autre étude en cours au Cemagref, en partenariat avec le SCHAPI.

Concrètement, l'objectif du stage est d'évaluer la qualité des prévisions d'ensemble de débits basées sur la prévision d'ensemble météorologique pour l'alerte aux crues. Pour cela, nous cherchons à mettre en place un outil de vérification des prévisions d'ensemble grâce à la programmation de différents scores existants dans la littérature et au traitement d'une archive de

³ Source: <http://www.catnat.net/donneesstats/bd-catnat>

prévisions. Il devient ainsi possible de comparer les prévisions et les observations *a posteriori*, dans le but de pouvoir évaluer progressivement les modèles et finalement aboutir à des prévisions toujours plus fiables dans la prévision des crues.

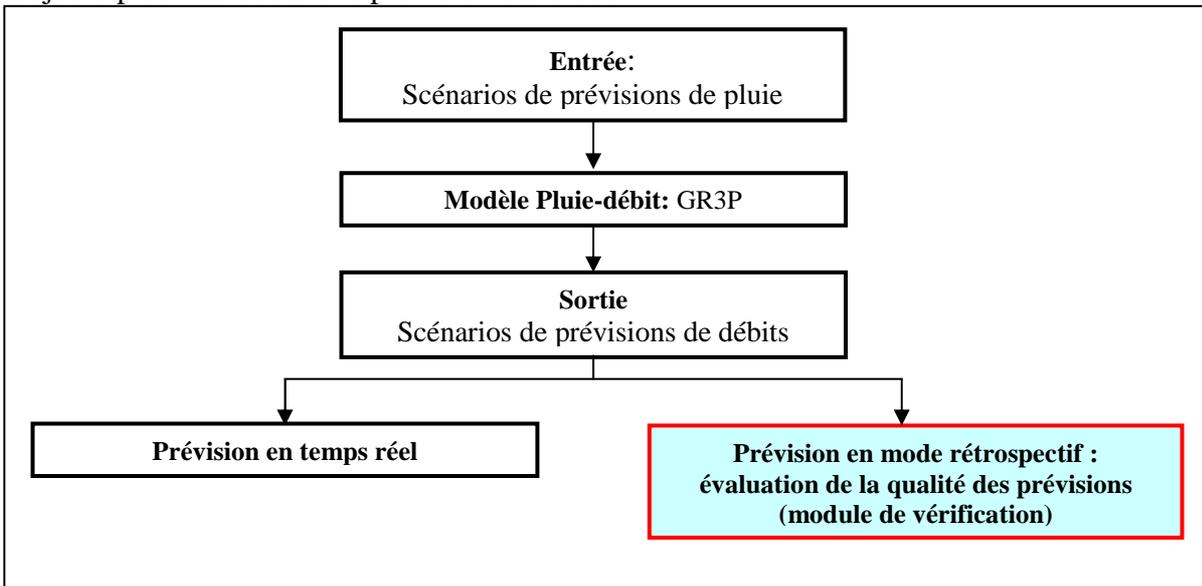


Figure 1 : Schéma général du système de prévision de débits GRPE

Pour construire l'outil de vérification, une archive de prévision d'ensemble de Météo-France (PEARP) a été exploitée. Les sorties de deux modèles hydrologiques (GR3P du Cemagref et SIM SAFRAN-ISBA-MODCOU de Météo-France), appliqués sur un grand nombre de bassins versants en France, constituent également la base de cette construction (Figure 2). Ce test comparatif de la performance entre ces deux modèles structurellement différents a pour objectif d'évaluer le comportement des scores de vérification sur les différentes conditions de la transformation de la pluie en débit. Puisque l'outil doit servir plus tard à évaluer les prévisions issues d'autres modèles, et sur la base d'autres prévisions météorologiques d'ensemble, on cherche à développer un module souple et facile à opérer.

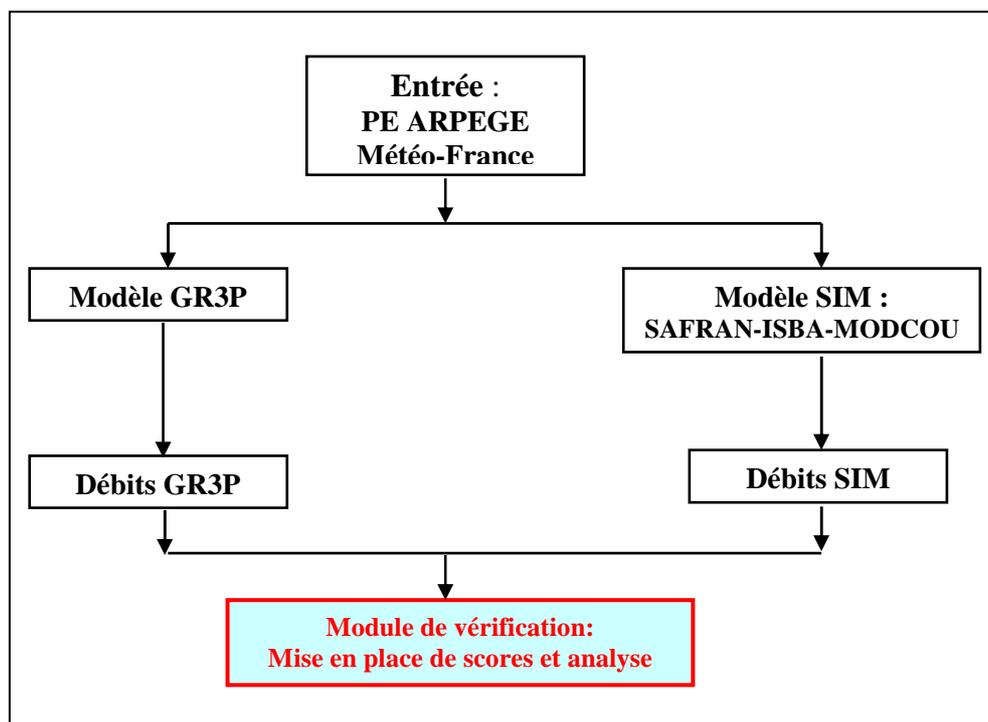


Figure 2 : Schéma général de l'approche traitée dans ce travail

Première partie: Synthèse bibliographique

I-1 Généralités sur la modélisation hydrologique

Un modèle hydrologique est un moyen de raisonner sur les phénomènes hydrologiques ; il représente une image de la réalité perçue par le modélisateur. Les processus hydrologiques sont étudiés à l'échelle du bassin versant. La modélisation hydrologique permet de répondre aux différentes problématiques de l'hydrologie opérationnelle, de simuler et de prévoir le fonctionnement d'un bassin. Elle permet de connaître les débits à l'exutoire, dont la mesure est parfois inexistante ou manquante. On peut classer les modèles hydrologiques :

- selon les données en entrée : de pluie, de température ou d'ETP, entrées soit de manière continue, soit de manière événementielle.
- selon la discrétisation du bassin dans l'espace : globale (l'unité est le bassin versant), semi-distribuée (l'unité est un sous bassin versant) ou distribuée (le bassin versant est discrétisé par plusieurs mailles).
- selon les fondements théoriques : mathématique (modèle boîte noire), conceptuel ou empirique (modèles à réservoirs), théorique (fondé sur la physique).
- selon le pas de temps de calcul : horaire, mensuel, annuel, interannuel.

Les applications des modèles sont nombreuses, dont notamment :

- **La simulation** : où l'on génère une série chronologique de débit sans référence à une date précise. C'est également la reconstitution de données manquantes de débit, qui, elles, sont datées.
- **La prédétermination** : où l'on cherche à estimer la probabilité d'apparition d'un événement (souvent les valeurs extrêmes de crue ou d'étiage), ou la valeur de cet événement pour une probabilité d'apparition donnée. On se situe dans le domaine fréquentiel.
- **La prévision**, application sur laquelle on se focalise pendant ce stage, où l'on cherche à prédire et à anticiper les évolutions des débits futurs à partir d'un instant donné.

Pour utiliser un modèle hydrologique dans le but de faire de la prévision, une méthode de mise à jour permettant d'exploiter l'information disponible sur les débits observés est généralement employée. On utilise également un scénario de conditions de précipitation future. Selon les enjeux et les objectifs, ceci peut-être un scénario de précipitation future nulle, une prévision déterministe ou une prévision d'ensemble, ceux-ci généralement issues des modèles météorologiques.

I-2 La prévision d'ensemble. Qu'est-ce que c'est?

En amont de la prévision de débits, les modèles météorologiques⁴ jouent un rôle important. La capacité de ces modèles à prévoir des précipitations intenses reste faible et les incertitudes sont importantes. De plus, la fiabilité des prévisions de pluies diminue à mesure que l'échéance de prévision augmente. La limite de prévisibilité⁵ d'un modèle varie selon le type du modèle et le phénomène météorologique à prévoir. Actuellement, de manière générale, elle est inférieure à 3 jours pour une prévision fiable. Pour pouvoir aller au-delà de cette limite, les services météorologiques font appel aux systèmes de prévision d'ensemble.

⁴ Les modèles utilisés sont des modèles numériques qui prennent en compte les paramètres tels la température, la pression atmosphérique, l'humidité, etc. dans des systèmes d'équations mathématiques.

⁵ La prévisibilité est un délai, le plus souvent exprimé en jours, au-delà duquel on estime qu'un modèle sera trop éloigné de la réalité pour être fiable.

(Source: <http://comprendre.meteofrance.com>)

L'une des sources d'erreur de prévision qui limite la prévisibilité est l'inexactitude des données d'observation qui définissent l'état initial de l'atmosphère, sur lequel on se fonde pour procéder au départ des calculs. Afin de pallier ces inexactitudes, on réalise alors plusieurs états initiaux, obtenus chacun en imposant aux données observées de petites variations, plus petites que les erreurs normales de mesure ou d'interpolation. Puis, à partir de chacun de ces états initiaux, on lance le calcul des états futurs de l'atmosphère : pour une échéance donnée, on obtient donc autant d'états futurs que l'on a réalisés de variantes de l'état initial. Les états finaux ainsi obtenus peuvent être particulièrement divergents, ceci s'explique par l'état chaotique⁶ de l'atmosphère : deux états initiaux très proches peuvent aboutir à des prévisions très différentes. Chaque état initial aboutissant à une prévision est appelé "membre" ou "run". Les prévisions capturent alors la probabilité des événements météorologiques et la zone d'incertitude relative à chaque situation prévue pour des échéances allant jusqu'à 10-15 jours (prévisibilité supérieure à la prévision traditionnelle). On aboutit ainsi à un système de prévision d'ensemble. Plusieurs valeurs de prévisions sont obtenues avec des scénarios considérés équiprobables de l'évolution de l'atmosphère : chaque membre a une probabilité de "1/nombre total de scénarios" de se produire. Une autre illustration serait de prendre l'exemple de 5 scénarios sur 10 dépassant un seuil d'alerte : *a priori*, cela représenterait donc 50% de chance de dépasser ce seuil.

Ce caractère probabiliste de la prévision d'ensemble rend la vérification des prévisions moins immédiate car on ne peut pas le faire sur un événement. Une longue série d'événements prévus est nécessaire afin de vérifier, statistiquement, que les probabilités énoncées par les prévisions sont fiables et n'ont pas de biais. Les outils généraux utilisés en prévision simple peuvent être exploités : chaque membre peut être considéré comme une prévision déterministe. Mais, il est important d'exploiter les avantages d'un tel système de prévision d'ensemble et de vérifier la fiabilité de la distribution de probabilité d'événements possibles.

L'utilisation en hydrologie de la prévision d'ensemble est encore peu fréquente. Malgré tout, on note les initiatives européenne EFAS et internationale HEPEX. Le développement du Système européen d'alerte aux crues (EFAS⁷, pour *European Flood Alert System*) a été lancé en 2003 par la Commission européenne. EFAS a pour but d'augmenter le temps de préparation aux inondations dans les bassins fluviaux transnationaux à partir de la prévision de crues à moyenne échéance (3-10 jours) et la pré-alerte. Le modèle hydrologique utilisé est le LISFLOOD (modèle distribué) et quatre seuils critiques sont définis pour la pré-alerte. Le *Hydrologic Ensemble Prediction Experiment* (HEPEX)⁸ est une initiative internationale qui rassemble les communautés météorologiques et hydrologiques à travers le monde pour mener des recherches axées sur la promotion de techniques de prévisions hydrologiques probabilistes. Cela permet, d'une part, d'étudier le comportement des prévisions d'ensemble dans des conditions climatiques différentes et, d'autre part, d'accroître plus aisément la communauté scientifique travaillant autour de la thématique.

Au niveau national, Météo-France a développé un système couplé hydrométéorologique basé sur la chaîne de modèles SIM (SAFRAN-ISBA-MODCOU). Deux sources de prévisions météorologiques sont actuellement exploitées à titre expérimental : a) les 51 membres du système de prévision d'ensemble du Centre européen de prévision (CEPMMT) à 10 jours d'échéance, et b) les 11 scénarios de prévisions d'ensemble issues du modèle ARPEGE à 2 jours d'échéance (Rousset-Regimbeau *et al.*, 2007 ; Thirel *et al.*, 2008).

⁶ Météo-France, sur leur site, parle de l'effet papillon, illustré par un modèle à trois variables, qui consiste en la divergence rapide de deux trajectoires initialement très proches (un battement d'aile de papillon).

⁷ <http://natural-hazards.jrc.ec.europa.eu/index.html>

⁸ <http://hydis8.eng.uci.edu/hepex/>

I-3 Vérification des prévisions d'ensemble

Les prévisions d'ensemble étant de nature probabiliste, leur vérification et interprétation doivent être faites avec des scores adaptés (Martin *et al.*, 2008). La communauté météorologique utilise les prévisions d'ensemble depuis plusieurs années et a, de ce fait, vu le développement de nombreuses méthodes permettant l'évaluation de la performance de ces prévisions. Assez générales, certaines peuvent être appliquées à des données hydrologiques (Petit, 2008 ; Thirel *et al.* 2008). Quelques recommandations concernant des applications hydrologiques sont listées par Pappenberger *et al.* (2008), entre autres : la vérification doit être faite avec des valeurs réellement observés et à l'échelle du bassin versant.

Plusieurs scores ont été retrouvés dans la littérature⁹. La méthode de vérification la plus simple est la méthode visuelle dite "eyeball". C'est une simple représentation graphique, ou sur une carte, des observations et des prévisions. Cependant, ce n'est pas une méthode quantitative (seul un évènement est visualisé) et elle est très sujette aux particuliers, aux préjugés de l'interprétation subjective. De plus, dans le cadre de la prévision d'ensemble, la lecture est difficile car les scénarios peuvent varier beaucoup entre eux.

Dans la vérification de la prévision d'ensemble, il existe deux grands types de scores : la première catégorie mesure la dispersion et la deuxième, regroupe les autres scores qui intègrent la qualité de la prévision sur une série d'évènements prévu-observé. La dispersion de l'ensemble (écart-type des différents membres) caractérise la façon dont les différents membres couvrent l'espace des possibles. L'écart entre la moyenne des prévisions et chaque membre est calculé. Les autres scores évaluent surtout la qualité des prévisions : en quelques mots, ils vérifient comment la probabilité des évènements prévus dans l'ensemble correspond à la fréquence à laquelle les évènements sont observés. Le but des prévisions d'ensemble est de donner au prévisionniste une image précise de la distribution de probabilité d'évènements possibles. Il faut alors vérifier cette distribution et ses qualités statistiques.

Petit (2008) dans son étude évalue les prévisions d'ensemble des débits du bassin versant de la Lièvre, émises par Environnement Canada. Il utilise comme référence pour la vérification une prévision déterministe. Pour cela, des scores ont été programmés : le MAE (*Mean Absolute Error*), le RMSE (*Root Mean Square Error*), le RPS (*Ranked Probability Score*) et le diagramme de Talagrand. Dans la plupart des cas, la prévision d'ensemble donne de bons scores, mais présente une sous-dispersion. En France, Thirel *et al.* (2008) estiment les qualités des prévisions EPS (*Ensemble Prediction System*) du CEPMMT¹⁰ et de PEARP (Prévision d'Ensemble ARPege), utilisées dans un même modèle hydrologique. Ils comparent ces deux prévisions avec des débits simulés à partir des pluies observées. Les scores programmés dans leur étude sont : le RMSE, le BS et le BSS (*Brier Score et le Brier Skill Score*), le RPS et le RPSS (*Ranked Probability Skill Score*). Les résultats ont montré une faible dispersion des ensembles de pluie. Les débits semblent donner des résultats plus complexes : pour les pluies, les scores du Centre Européen sont moins bons que ceux de PEARP, mais cette tendance s'inverse quant il s'agit des débits, surtout, pour des bassins de taille importante.

⁹ Voir notamment la synthèse présentée dans http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html (site internet du WWRP World Weather Research Program et du WCRP World Climate Research Program)

¹⁰ CEPMMT est l'acronyme français de ECMWF (European Centre for Medium-Range Weather Forecasts) <http://www.ecmwf.int/>

Deuxième partie: Aspects méthodologiques et données

II-1 Méthodologie générale

Les démarches méthodologiques entreprises visaient à :

- choisir des scores de vérification représentatifs et adaptés, qui permettraient d'évaluer la qualité des prévisions. Pour cela, le choix a été dicté par les scores fréquemment utilisés par les météorologues et les hydrologues dans la littérature pour mesurer la qualité des prévisions d'ensemble.
- appliquer ces scores à une longue série de couplets prévision-observation. Les prévisions météorologiques à disposition s'étendaient sur 19 mois (10 Mars 2005 au 30 Septembre 2006). Cependant, les observations disponibles s'arrêtaient le 31 Juillet 2006 et l'ensemble des calculs était contraint de s'arrêter à cette date. Le travail a donc été effectué sur 17 mois.
- utiliser les séries de débits prévus obtenus à partir de deux modèles hydrologiques qui ont une structure différente : (i) SIM de Météo-France est un modèle distribué, fonctionnant avec des mailles et (ii) GR est un modèle global à l'échelle du bassin versant.
- appliquer l'étude non pas à un seul bassin, mais à un échantillon réparti sur l'ensemble du territoire français et représentatif de surfaces différentes. Les résultats obtenus sont alors plus robustes et valables dans des conditions climatiques et géographiques variables en France.
- choisir un langage de programmation souple pour construire l'outil de vérification, capable de traiter un grand nombre de données et permettant une visualisation rapide, synthétique ou au détail des résultats obtenus. C'est le logiciel R qui a donc été choisi. Plusieurs sont les raisons qui expliquent ceci : (i) R est un logiciel de calcul statistique libre et gratuit, les extensions (packages) sont aussi gratuits et facilement téléchargeables ; (ii) il existe un forum actualisé par tous les utilisateurs disponibles sur internet ; (iii) R traite les données selon une structure en matrice, le nombre élevé de valeurs ne l'handicape pas, au contraire, il les classe de façon optimale ; (iv) R est aussi un logiciel graphique, les résultats sont directement visualisés permettant une manipulation et une exploitation rapide de ces derniers. De plus, une formation interne sur ce logiciel a été organisée au Cemagref au début du stage ; il était donc important de mettre en valeur ces nouvelles connaissances.

II-2 Les données

II-2-1 Les 211 bassins versants étudiés

La sélection des bassins s'est faite en croisant la base de données du Cemagref et celle de Météo-France : 243 bassins ont alors été retenus au préalable. Un premier dépouillement des données a cependant montré qu'une partie de ces bassins présentaient beaucoup de valeurs manquantes durant la fenêtre d'étude, ce qui rendait alors le calcul des scores peu robustes. Au final, l'étude a été menée avec 211 bassins versants (Figure 3). La liste des bassins étudiés est présentée en annexe.

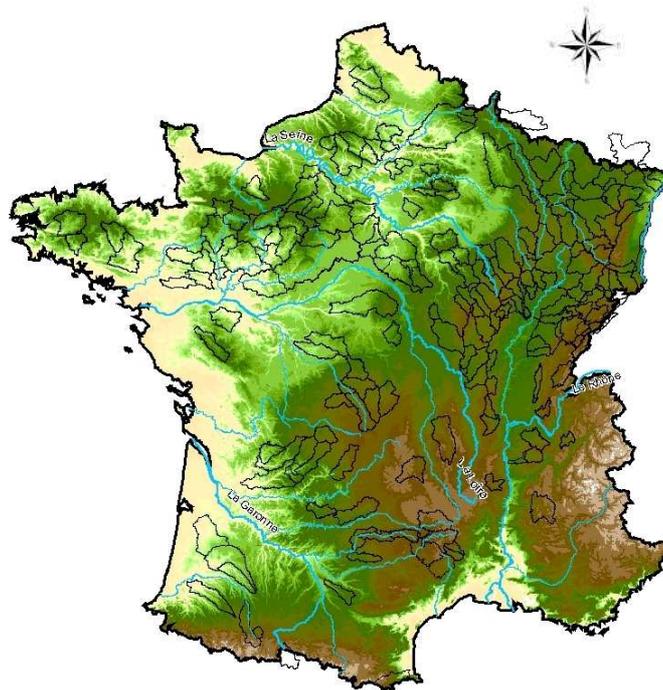


Figure 3 : Carte de localisation des 211 bassins étudiés

II-2-2 Les prévisions de pluie

La prévision de précipitations à Météo-France concerne des échelles spatio-temporelles extrêmement variées. Les modèles de prévision numérique sont les sources principales de ces prévisions de précipitations, mais l'expertise humaine et les observations jouent un rôle extrêmement important. Ainsi, lorsque les météorologues s'intéressent à des échéances de prévision très courtes (inférieures à 2 ou 3 heures), ils utilisent des techniques de suivi des systèmes précipitants à partir d'images satellite et d'images radar. Elles permettent d'évaluer leur nature et leur comportement à une échéance relativement brève ; on parle alors de « prévision immédiate ». Lorsque l'échéance de prévision augmente, l'utilisation de modèles de prévision numérique est incontournable. Météo-France dispose actuellement de 3 gammes de modèles qui se différencient par la taille (décroissante) du maillage géographique et par l'échéance, et se complètent en « s'emboîtant » : Arpège, Aladin, Arome. Les données utilisées dans ce travail viennent du modèle Arpège. Sa maille sur la France est d'environ 25 km sur l'horizontale. C'est un modèle dit « global » car il calcule l'évolution de l'atmosphère sur tout le globe terrestre. Il couvre 102 heures d'échéance, c'est-à-dire environ 4 jours (Vincendon et Carrière, 2008).

Le système de prévision d'ensemble de Météo-France est nommé PEARP. Les données ont été fournies dans une fourchette de prévisions de 60 heures, avec 11 membres, dont 1 dit membre de contrôle (le membre de contrôle part de la condition la plus favorable, proche d'une prévision déterministe). Dans la suite de notre étude, les précipitations prises en compte sont la somme des composantes "pluie" et "neige" de la PEARP.

Les données des 9892 mailles de la PEARP (mailles de 64 km²) qui recouvrent toute la France sont lues par un programme Fortran. Pour chaque bassin de cette étude, un traitement au préalable à l'aide du logiciel ArcGIS a permis de connaître le pourcentage de chaque maille se trouvant sur le bassin. Les pluies sur chaque bassin versant sont alors calculées par la moyenne spatiale pondérée par la superficie de la maille se trouvant sur le bassin.

Les données de précipitations prévues sont disponibles au pas de temps horaire. Les prévisions sont lancées à **18h** le jour **j** d'émission de la prévision. Les 6 premières heures et les 6 dernières heures ne sont pas utilisées dans cette étude. Les données restantes sont cumulées pour aboutir à des prévisions journalières entre la 7^{ème} et la 31^{ème} heure pour l'échéance 1 : **j + 1** ou **Day 1** (en d'autre terme, le lendemain) et entre la 32^{ème} et la 56^{ème} pour l'échéance 2 : **j + 2** ou **Day 2** (Figure 4). Les pluies prévues sont ainsi traitées dans une fourchette journalière (soit entre 0h et 23h59). Ceci permet de les comparer avec les pluies observées, qui sont également présentes dans la base de données dans la même fenêtre temporelle.

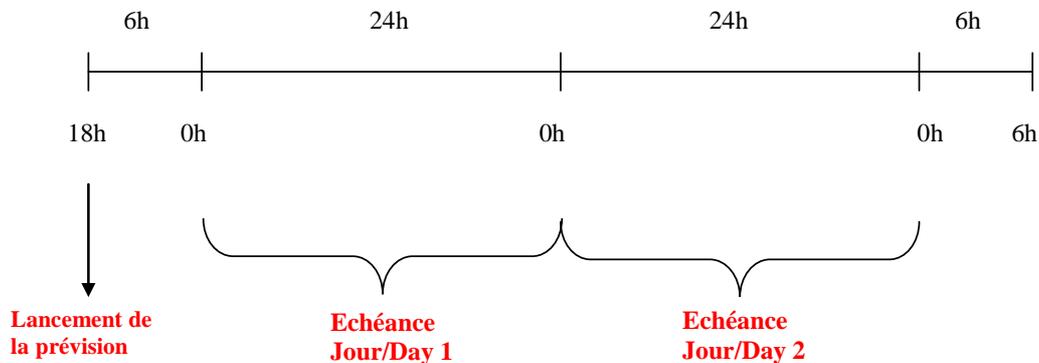


Figure 4 : Représentation graphique des échéances de prévision

II-2-3 Les prévisions de débits

Les prévisions de pluies ont été utilisées en entrée à deux modèles hydrologiques de la transformation pluie-débit, qui ont généré les séries de débits prévus évaluées dans cette étude. Les principes de base des modèles hydrologiques sont résumés ci-après.

a- Le modèle GR3P du Cemagref

(source : <http://www.Cemagref.fr/webgr/index.htm> ; Tangara, 2005)

Le modèle GR3P a été spécifiquement adapté pour la prévision des crues par Tangara (2005) et dérive du modèle de simulation pluie-débit GR4J, également développé par le Cemagref (Perrin *et al.*, 2001, 2003). Pour cette étude, le modèle a été utilisé au pas de temps journalier. Il a été préalablement calé pour chaque bassin et pour ses 3 paramètres (Figure 5) : (i) X1 est le coefficient d'échange, (ii) X2 est la capacité à un jour de son réservoir de stockage, (iii) X3 est la caractéristique de son hydrogramme unitaire HU dérivable en son sommet. Les pluies de bassin (P) et l'ETP constituent les données d'entrée au modèle. Les observations de pluie, ETP moyenne et des débits sur une longue série d'années sont utilisées dans le calage.

Quant au fonctionnement du modèle, un premier réservoir appelé réservoir de "production" gère, en fonction de son taux de remplissage, la part de pluie 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.

Le modèle permet des opérations en continu. Les calculs ne sont plus limités à une crue, mais intègrent une période plus longue, permettant de simuler l'état du bassin comme la résultante des conditions climatiques antérieures (mémoire des conditions antérieures). L'intérêt de cette approche est qu'elle permet de se soustraire à la détermination des conditions initiales.

La méthode de mise à jour (ou *updating*) 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.

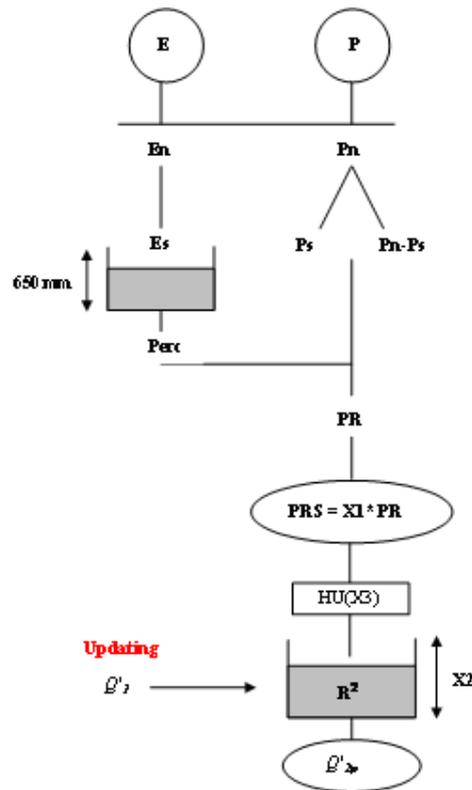


Figure 5 : Schéma du modèle hydrologique de prévision GR3P (Tangara, 2005)

b- Le modèle hydrologique SIM (SAFRAN-ISBA-MODCOU) de Météo-France

Le modèle **SIM** est une suite de trois modèles indépendants Martin *et al.*, (2008) :

- **SAFRAN** utilise des observations de surface, combinées à des données d'analyse de modèles météorologiques pour produire les paramètres (température, humidité, vent, précipitations solides et liquides, rayonnement solaire et infrarouge incident) nécessaires au fonctionnement d'ISBA. Ces paramètres sont analysés et sont ensuite interpolés sur une grille de calcul régulière (8x8 km²). Les deux premières étapes de l'analyse sont consacrées, d'une part, à l'analyse de la température, de l'humidité, du vent et de la nébulosité toutes les 6 heures et, d'autre part, à l'analyse des quantités quotidiennes de précipitation. La troisième étape est l'interpolation horaire de ces paramètres, ainsi que le calcul des termes de rayonnement.

- **ISBA** (Interaction sol-biosphère-atmosphère) simule les échanges d'eau et d'énergie entre le sol et l'atmosphère. La version utilisée dans la chaîne SIM utilise 3 couches de sol (surface, zone racinaire, zone profonde) et deux températures (température de surface globale du continuum sol-végétation et température profonde). ISBA simule l'ensemble des flux d'eau avec l'atmosphère (interception, évaporation, transpiration) et avec le sol (ruissellement des précipitations et drainage dans le sol).

- **MODCOU** est un modèle hydrologique distribué. Il utilise en entrée le ruissellement et le drainage d'ISBA pour calculer l'évolution des nappes (actuellement seules les nappes de la Seine et du Rhône sont représentées) et le débit des rivières (900 points simulés actuellement).

L'ensemble de la chaîne a été validée sur une période de 10 ans, par comparaison avec des hauteurs de nappe, hauteur de neige et débits des rivières. Il apparaît que globalement SIM restitue bien la dynamique du bilan hydrologique et des débits. En moyenne l'efficacité (critère de Nash) augmente avec la taille des bassins (Martin *et al.*, 2008). Les limitations principales du modèle

sont liées à la non prise en compte de certains systèmes importants, comme les nappes qui influencent fortement la temporalité des débits (par exemple celle de la Somme) et les barrages.

La première application de SIM en prévision d'ensemble s'est focalisée sur la prévision à moyen terme, permettant d'anticiper les crues dans les bassins les plus grands ou fournir des informations intéressantes sur l'évolution des débits les plus faibles, pour lesquels les réponses hydrographiques dépendent des niveaux de nappe.

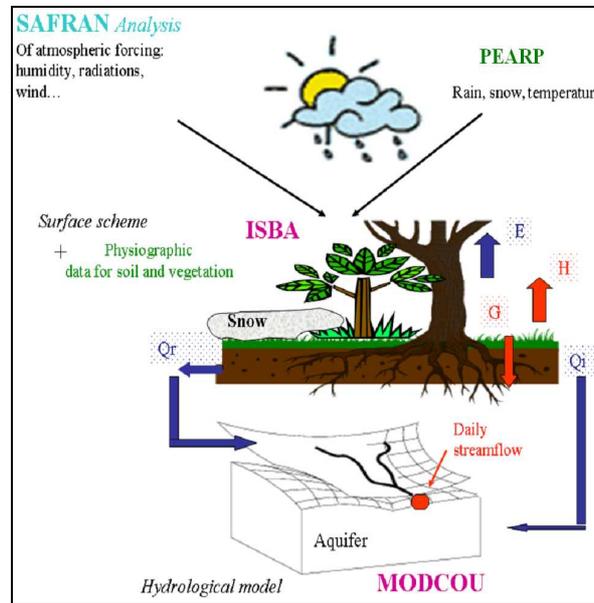


Figure 6 : Schéma du modèle SIM en prévision (adapté du poster EGU 2008 présenté par Rousset-Regimbeau et al. : *Medium-range ensemble streamflow forecast over France*)

Comme décrit ci-dessus, contrairement à GR3P (simplifié par GR dans la suite du rapport), le modèle SIM n'est pas mis à jour et n'assimile donc pas le dernier débit observé (des travaux de thèse en cours de G. Thirel étudient l'assimilation de débits dans SIM). Il était donc intéressant de considérer GR3P sans mis à jour et comparer ses performances avec SIM. Finalement, l'étude comparera alors trois modèles : GR avec et sans mise à jour et SIM (dans ce rapport *GR*, *GR no updating*, *SIM*).

II-2-4 Les observations

Pour évaluer les prévisions, des références de vérification sont nécessaires. Dans ce travail, ce sont les observations de débits qui ont été considérées. C'est une démarche originale car les études antérieures (cités précédemment) n'ont comme référence que d'autres prévisions (par exemple, la prévision déterministe) ou des débits simulés à partir des données météorologiques observées. Les débits observés viennent de la banque HYDRO. Les observations de pluies sont celles de la base SAFRAN de Météo-France. Ces données sont déjà sous format et disponibles dans la base de données du CEMAGREF.

Les débits de sortie de SIM et de GR sont comparés aux débits journaliers observés pour la même période d'étude (10 Mars 2005 au 31 Juillet 2006) et pour les deux échéances équivalentes à celles des pluies.

II-3 Les scores de vérification

Les scores ici présentés sont détaillés dans les références :

- Forecast Verification - Issues, Methods and FAQ
http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html/
- Jolliffe et Stephenson (2004), Forecast verification. A Practitioner's Guide in Atmospheric Science, John Wiley & Sons.
- "La prévision d'ensemble expliquée" www.meted.ucar.edu/nwp/pcul/ensemble_fr/

II-3-1 RMSE et ratio-RMSE

La racine carrée de la moyenne de l'erreur au carré ou le RMSE (*Root Mean Square Error*) mesure la magnitude moyenne des erreurs commises par la prévision. Dans la formule, les erreurs sont élevées au carré avant qu'elles ne soient moyennées : les RMSE donnent alors un poids relativement important aux différences élevées. Cela signifie que le RMSE est plus utile lorsque les grandes erreurs sont particulièrement indésirables.

Pour chaque échéance de prévision, les RMSE relatifs aux prévisions de débit et de pluie ont été calculés pour chaque bassin versant et sur l'ensemble de la période d'étude selon la formule :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - o_i)^2} \text{ (mm)}$$

Équation 1

Où,

o_i : est la donnée observée pour le jour i ,

m_i : est la moyenne des prévisions des 11 membres de l'ensemble pour le jour i ,

N : est le nombre de jours comptés pour calculer le score

Le RMSE mesure donc la différence moyenne entre l'observation et la moyenne des membres de la prévision d'ensemble. On ne considère que le nombre de jours pendant lesquels les données de précipitation et de débits observés ne sont pas manquantes.

Pour s'affranchir de l'effet de taille des bassins versants et avoir ainsi des valeurs adimensionnelles pour pouvoir comparer les résultats des différents bassins entre eux, les ratios des RMSE, par rapport aux valeurs moyennes observées, ont été calculés :

$$ratio_RMSE_P = \frac{RMSE}{\bar{P}}; \quad ratio_RMSE_Q = \frac{RMSE}{\bar{Q}}$$

Équation 2

Avec :

\bar{P} : la moyenne des pluies observées pendant la période d'étude,

\bar{Q} : la moyenne des débits observés pendant la même période.

II-3-2 Ecart-type et ratio de l'écart-type

Le calcul de l'écart-type permet d'avoir une mesure de la dispersion des membres : en effet, la différence entre la valeur prévue par un membre et la moyenne de l'ensemble est calculée pour chaque jour de prévision. La formule est définie ainsi :

$$\sigma = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{n} \sum_{k=1}^n (x_{k,i} - \bar{x}_i)^2} \quad (mm) \quad \text{Équation 3}$$

Où,

n : est le nombre de membres de la prévision (11 dans notre cas),

N : est le nombre de jours comptés pour le calcul du score,

\bar{x}_i : est la moyenne des prévisions de l'ensemble pour le jour i ,

$x_{k,i}$: est la valeur de la prévision d'un membre pour le même jour i .

De même, pour la mesure de la dispersion des membres, un ratio adimensionnel a été calculé :

$$\text{Ratio}_{\sigma_P} = \frac{\sigma}{PE}; \quad \text{Ratio}_{\sigma_Q} = \frac{\sigma}{QE}$$

Equation 4

Avec:

\overline{PE} : la moyenne des précipitations prévues par tous les membres et pour tous les jours,

\overline{QE} : la moyenne des débits prévus par tous les membres et pour tous les jours.

II-3-3 Le tableau de contingence

Les tableaux de contingence permettent d'étudier l'association de deux variables et d'en déduire des probabilités conjointes et conditionnelles. Ils indiquent le nombre d'observations de chaque combinaison possible des variables analysées. En prévision, le tableau de contingence compte le nombre d'éléments suivants (Tableau 1) :

Succès (ou hit) - évènement prévu et observé

Évènement manqué (ou miss) - évènement non prévu mais observé

Fausse alerte (ou false alarm) - évènement prévu mais pas observé

Rejets corrects (ou correct negative) - évènement non prévu et non observé

Ce tableau permet de visualiser quel type d'erreur a été commise. Une bonne prévision devrait logiquement ne contenir que des *succès* et des *rejets corrects*. Le tableau se présente comme suit :

		Evènement observé		Total
		OUI	NON	
Evènement prévu	OUI	Succès	Fausse alerte	Prévu
	NON	Manqué	Rejets Correct	Non prévu
Total		Observé	Non observé	Total

Tableau 1 : Présentation d'un tableau de contingence

Pour construire un tableau de contingence, il faut définir ce qu'est un évènement observé et un évènement prévu. Dans cette étude, trois seuils ont été choisis pour la définition des évènements observés et calculés sur la période d'étude (période de vérification) :

- Q_{ref1} = le quantile 10% des débits observés (Q10),
- Q_{ref2} = le quantile 50% des débits observés (Q50),
- Q_{ref3} = le quantile 90% des débits observés (Q90).

Pour le Q10, il permet de déterminer si l'on prévoit de manière correcte ou non les bas débits. Dans ce cas, on compte les événements inférieurs à ce seuil. Pour les deux autres quantiles, dans le contexte de la prévision des crues, ce sont les événements ayant dépassé les seuils qui intéressent.

Pour la définition des événements prévus, trois autres seuils ont été pris en compte :

Si un pourcentage (p%) des membres atteignent les seuils de débits, alors l'évènement est considéré comme un "évènement prévu". Sinon, il est "non prévu". Les pourcentages choisis sont :

- p1 = 20%
- p2 = 50%
- p3 = 80 %

Exemple pour p1 = 20% et Q_{ref}2=Q50 :

Si "évènement observé", c'est-à-dire que l'observation est supérieure à Q50, on compte le nombre X de membres pour lesquels les débits sont supérieurs à Q50.

*Si X > 20% * nombre de membres => on compte 1 "succès" dans le tableau de contingence.*

*Si X < 20% * nombre de membres => on compte 1 "manqué" dans le tableau de contingence.*

Si "évènement non observé", c'est-à-dire que l'observation est inférieure au débit seuil Q50, on compte le nombre Y de membres pour lesquels les débits sont supérieurs à Q50.

*Si Y > 20% * nombre de membres => on compte 1 "fausse alerte" dans le tableau de contingence*

*Si Y < 20% * nombre de membres => on compte 1 "rejet correct" dans le tableau de contingence*

Les probabilités suivantes peuvent être alors définies :

$$\text{Biais ou BIAS} = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} \quad \text{Équation 5}$$

Le BIAS mesure le ratio entre la fréquence des événements prévus et des événements observés. Il indique si le système de prévision a une tendance à sous-prévoir (BIAS < 1) ou à sur-prévoir (BIAS > 1). Il ne mesure pas comment les prévisions correspondent aux observations, mais calcule uniquement les fréquences relatives.

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}} \quad \text{Équation 6}$$

Le POD ou probabilité de détection est sensible aux "succès", mais ignore les "fausses alertes". Score parfait pour le BIAS et le POD = 1

$$\text{FAR} = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}} \quad \text{Équation 7}$$

Le FAR ou ratio des fausses alertes mesure la fraction des prévus "oui" qui ne se sont pas produits. Le score parfait est de 0.

II-3-4 La persistance

C'est un score couramment utilisé au Cemagref pour évaluer les performances de prévision. Les résultats du modèle sont alors comparés à un modèle naïf, qui consiste à considérer que le débit

prévu pour le jour i est le même que le dernier débit mesuré. La persistance est donnée par la formule de l'équation 8.

$$Pers = 100 \times \left[1 - \frac{\sum_{i=1}^N (O_i - F_i)^2}{\sum_{i=1}^N (O_i - O_{i-1})^2} \right] \quad -\infty < Pers < 100 \quad \text{Équation 8}$$

Où,

O_i : est le débit observé pour le jour i ,

O_{i-1} : est le débit observé pour le jour $i-1$ (1 étant l'échéance de la prévision). Ce débit est le dernier débit observé le jour de l'émission de la prévision.

F_i : est la moyenne des prévisions d'ensemble pour le jour i ($= \bar{x}_i$ dans le calcul de l'écart-type).

II-3-5 Le Brier Score et le Brier Skill Score

Le score de Brier et son "skill score" associé sont très largement utilisés pour évaluer les prévisions d'ensemble. Ils qualifient les performances d'un système probabiliste par rapport à un événement binaire (dépassement de seuil). Le skill score associé (BSS) varie entre $-\infty$ et 1 et compare le score de Brier à un score de référence pour lequel le BSS vaut 0 (Martin *et al.*, 2008).

Le Brier Score est donné par la formule:

$$BS = \frac{1}{N} \sum_{j=1}^n (p_j - x_j)^2 \quad \text{Équation 9}$$

Où,

N : est le nombre de réalisations (prévisions) sur la période de vérification, et pour chaque réalisation j ,

p_j : est la probabilité de prévision de l'évènement,

x_j : a une valeur égale à 1 si l'évènement s'est produit et 0, sinon.

Comme dans le cas du tableau de contingence, des seuils de débits, calculés sur la période d'étude, ont été fixés pour définir les événements :

- $Q_{ref1} = Q10$
- $Q_{ref2} = Q50$
- $Q_{ref3} = Q90$

Pour définir les probabilités de prévision p_j d'un événement, le nombre de membres dépassant le seuil est compté et divisé ensuite par le nombre total de membres dans le cas des quantiles Q50 et Q90. Pour les débits faibles, c'est le nombre de membres qui n'ont pas dépassé le débit de référence (Q10) qui est comptabilisé.

Le Brier minimum 0 est obtenu pour un système parfait (un système déterministe) où $p_j = x_j$ pour tous les j . Un tel système a une probabilité de prévision de 1 (0) à chaque fois que l'évènement est (n'est pas) observé. Parce qu'un tel système n'utilise pas des probabilités entre 0 et 1, il ne présente pas d'incertitude et peut être considéré comme un système de prévision déterministe binaire.

Au contraire, le Brier score prend la valeur maximale de 1 pour une erreur systématique de la prévision déterministe, qui prévoit avec certitude le mauvais évènement à chaque réalisation, c'est-à-dire $p_j = 1 - x_j$

Pour comparer un système de prévision à un autre système de référence, il est pratique de comparer le Brier score à un Brier de référence. On définit ainsi le Brier Skill Score, dont la formule est présentée dans l'équation 10.

$$BSS = \frac{BS - BS_{référence}}{0 - BS_{référence}} = 1 - \frac{BS}{BS_{référence}} \quad \text{Équation 10}$$

Le calcul d'un $BS_{référence}$ est alors nécessaire et il faut définir les p_j de l'équation 9 pour la référence. On peut utiliser, par exemple, la climatologie, c'est-à-dire comparer chaque observation à une prévision qui serait égale à la valeur moyenne climatologique, ou la persistance, c'est-à-dire comparer chaque observation à une prévision qui serait celle d'un modèle naïf où la prévision sur les échéances considérées est égale au dernier débit observé.

Dans cette étude, nous avons utilisé la climatologie et la persistance comme référence. Pour la climatologie, pour chaque bassin versant, les débits moyens journaliers ont été évalués à partir des données historiques disponibles dans la base de données observées (données issues de la Banque Hydro entre 1970 et 2006). Pour chaque jour, un débit est ainsi défini, correspondant à un débit moyen journalier pour le jour j . Prenons, par exemple, une date quelconque : le 01 Janvier. Le débit "climatologique" pour cette date est calculé en faisant la moyenne de tous les débits du 01 janvier répertoriés dans la base de données.

Dans le $BS_{référence}$, la probabilité p_j est alors soit égale à 1, si le débit climatologique est supérieur (inférieur) au seuil de débit de référence Q50 et Q90 (Q10), soit égal à 0 si ce débit est inférieur (supérieur). Pour le calcul de $BS_{référence\ persistance}$, on utilise, à la place du débit climatologique, le débit issu du critère de persistance.

Contrairement au Brier score, le BSS est positivement orienté (c'est-à-dire que des valeurs élevées indiquent meilleures performances). $BSS = 1$ pour un système parfait de prévision et $BSS = 0$ (ou négatif) pour un système dont la performance est égale (ou inférieure) au système de référence.

II-3-6 Le Discrete Ranked Probability Score (DRPS) et le Skill Score associé (RPSS)

Le score de probabilité ordonnée (DRPS ou RPS) est analogue au score de Brier (BS) mais s'applique à des catégories multiples (valeurs prévues discrètes). Il est une mesure de "distance" entre les distributions de probabilité cumulative des prévisions et des vérifications (ici, les observations). Il est alors sensible à la distance, c'est-à-dire, il donne un plus mauvais score aux prévisions les plus éloignées de l'observation (Rousset Regimbeau, 2007).

Pour le calcul du RPS, $J = 10$ classes ont été définies :

- débits < Q10
- Q10 < débits < Q20,
- Q20 < débits < Q30,
- ...
- Q80 < débits < Q90, et
- débits > Q90.

Pour chaque prévision d'ensemble, une probabilité de prévision (y_j) est associée à chaque catégorie (rapport entre le nombre de membres dans la catégorie et le nombre total de membres). Un vecteur pour l'observation (o_j) est également construit : on attribue la valeur de 1 à la catégorie dans laquelle l'évènement est observé et 0 aux autres catégories. Les sommes cumulées pour la prévision et l'observation sont ensuite calculées :

$$Y_m = \sum_{j=1}^m y_j$$

$$O_m = \sum_{j=1}^m o_j$$

$m = 1, \dots, J$

Où,

Y_m : est la somme cumulée des probabilités de prévision pour les catégories $j \leq m$,

O_m : est la somme cumulée des probabilités d'observation pour les catégories $j \leq m$.

Le RPS est alors défini comme la différence des écarts quadratiques des différentes composantes de prévision cumulée et d'observation cumulée :

$$RPS = \frac{1}{J-1} \sum_{m=1}^J (Y_m - O_m)^2$$

Équation 11

L'équation ci-dessus ne donne que le RPS d'un jour de prévision. Pour un nombre N de jours, le RPS global est simplement la moyenne des RPS individuels :

$$\overline{RPS} = \frac{1}{N} \sum_{k=1}^N RPS_k$$

Équation 12

De même que pour le score de Brier, il est possible de définir un Skill Score associé au RPS à partir d'un RPS de référence :

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{ref}}$$

Équation 13

$$-\infty \leq RPSS \leq 1$$

Le RPSS est orienté positivement, c'est-à-dire que les valeurs positives représentent un bon score. 1 est le score parfait. Si le RPSS = 0, le modèle est équivalent à la référence. Dans cette étude, la référence utilisée est la climatologie.

II-3-7 Le diagramme de Talagrand (ou histogramme de rang)

Le diagramme de Talagrand permet de situer l'analyse de vérification par rapport aux données de la prévision d'ensemble. C'est un excellent moyen pour détecter les défauts systématiques d'un système de prévision d'ensemble. Plus particulièrement, il permet de savoir comment la dispersion de l'ensemble de prévisions représente la variabilité des observations.

La construction du diagramme se fait pour chaque bassin versant. Pour chaque jour, on classe les prévisions (donc les 11 membres) et l'observation. On note alors le numéro du rang de l'observation. Pour tous les jours de prévision, on effectue le même procédé et on obtient à la fin un diagramme donnant le rang des observations en abscisse et le nombre de cas (ou la fréquence) en ordonnée. On obtient ainsi un histogramme de rang.

Les diagrammes de Talagrand peuvent permettre de diagnostiquer plusieurs problèmes communs dans la dispersion des membres. Par exemple, si l'écart dans la prévision d'ensemble est trop faible, la valeur observée tombera souvent en dehors de la gamme de valeurs prévues. La première et la dernière cases de rang seront alors surchargées en fréquence et le diagramme aura la forme d'un « U »; cette forme peut également être causée par des biais conditionnels dans les prévisions. En revanche, si l'écart de la prévision d'ensemble est trop important, la valeur analysée ne sera presque jamais en dehors de la gamme des valeurs prévues. La première et la dernière cases de rang contiendront alors très peu de valeurs et, par conséquent, on obtiendra un diagramme en

forme de « n » (plus haut au milieu). Si le diagramme est asymétrique, le modèle comporte un biais d'un côté. Pour remédier à ce problème, il peut être nécessaire d'effectuer un changement en profondeur de la procédure d'assimilation des données ou du modèle de prévision d'ensemble. On notera également qu'un diagramme plat n'indique pas nécessairement une bonne prévision ; il mesure uniquement si la distribution de probabilité des observations est bien représentée par l'ensemble.

II-4 L'outil de vérification développé

L'outil de vérification a été programmé sur R¹¹. Le script développé permet, sur les premières lignes, d'entrer d'une manière souple tous les chemins des fichiers qui seront lus : les dates de début et de fin, les paramètres nécessaires (le nombre de jours, le nombre de membres et le nombre d'échéances, etc.). Le script lit ensuite tous les fichiers et place les données dans des matrices dont les tailles sont définies au préalable. Suivent les calculs des scores, après une succession de boucles qui s'emboîtent. Enfin, la programmation des sorties permet de visualiser les résultats sous une forme graphique.

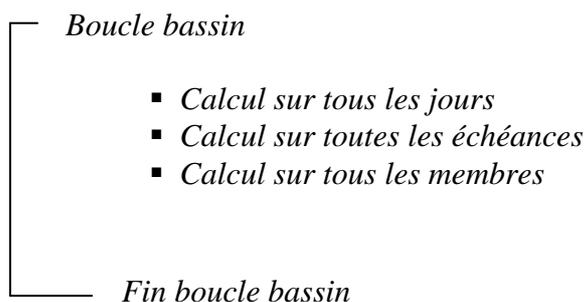
1. Entrées des paramètres :

Chemin des prévisions de SIM, de GR avec mise à jour et GR sans mise à jour

- *Nombre de bassins versants*
- *Date de début des prévisions*
- *Date de fin des prévisions*
- *Nombre de jours*
- *Nombre de membres*
- *Nombre d'échéances*
- *Etc.*

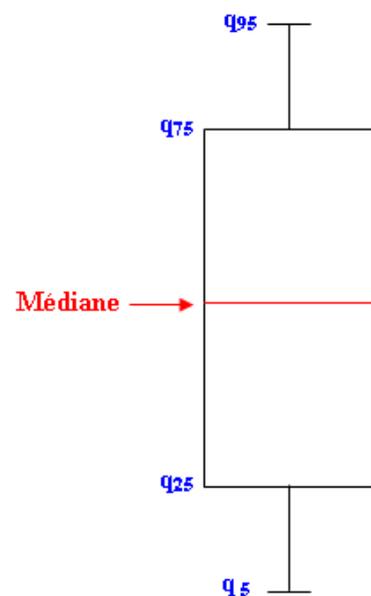
2. Lecture des fichiers

3. Calcul dans des boucles



4. Visualisation des résultats

- *soit par des graphes avec des axes (x, y),*
- *des cartes de France où chaque bassin est localisé,*
- *des boîtes à moustache ou boxplot (les quantiles définis dans les boîtes à moustache sont : la médiane, les quantiles 5%, 25%, 75%, 95%)*



¹¹ www.r-project.org

Troisième partie : Résultats et interprétations

Les résultats présentés ci-après sont issus de l'outil de vérification développé. Nous ne présentons pas tous les tests et les visualisations produits, mais un échantillon suffisamment démonstratif pour en tirer des conclusions sur la qualité des prévisions. Ci-après seul le score de persistance n'est pas présenté. En effet, lors de nos calculs de ce score, il s'est avéré que, malgré le fait que l'on obtenait des scores relativement bons pour GR (moyenne sur l'ensemble de bassins versants = 52,35 et 50,29 pour les échéances jour 1 et 2), les scores obtenus avec les débits de GR sans mise à jour et SIM ne sont pas exploitables puisqu'ils sont largement inférieures à -100.

III-1 Pluies et débits prévus

Avant de passer aux résultats des scores, une phase importante du travail consiste à avoir un aperçu général sur les prévisions de pluies et de débits. Cette partie préliminaire permet d'une part, de vérifier si la lecture de toutes les valeurs se fait correctement et d'autre part, d'effectuer un premier dépouillement visuel afin de rechercher une source d'erreur potentielle.

Les bassins représentés ici ont été choisis pour essayer de représenter au mieux la base de données, aussi bien sur la localisation en France, que sur la taille des bassins versants. Les deux mois d'Avril et d'Octobre sont censés refléter les variations saisonnières.

La figure 7 présente le bassin A1310310 : L'Ill à Ensisheim (affluent du Rhin) de 1038 km² et la figure 8, le bassin M0500610 : la Sarthe à Spay (affluent de la Loire) de 5285 km². Les graphiques comparent les pluies et les débits prévus aux valeurs observées. Ils permettent déjà dans un premier temps de regarder quelques tendances sur les prévisions.

Pour le mois d'Avril 2005 dans la figure 7, on discerne un cas où GR a tendance à prévoir plus par rapport aux débits observés. Au mois d'Octobre 2005 pour le même bassin, c'est le modèle SIM qui prévoit plutôt des débits supérieurs à ceux observés. Sur ces mêmes graphiques, la superposition des symboles d'un même modèle et pour un même débit observé indique une faible dispersion des membres de l'ensemble.

Le bassin M0500610 représenté dans la figure 8 montre un décalage entre les deux modèles GR et SIM au mois d'Avril. SIM ne se trouve pas au niveau de la première bissectrice : il prévoit moins que le débit observé. Les graphes de pluie (en haut des figures 7 et 8) montrent une dispersion des pluies prévues. En effet, pour chaque observation, on peut distinguer plus facilement que pour les débits, les 11 membres qui prévoient chacun une pluie différente.

La visualisation des débits prévus superposés aux hydrogrammes observés permet de discerner d'éventuels problèmes de prévision sur certains bassins ou certaines périodes. Le cas du bassin La Dore à Dorat (Puy-de-Dôme 63, Auvergne, 1523 km²), par exemple, est apparu intéressant. Tel qu'illustré par les hydrogrammes de la figure 9, au mois de Mars, malgré l'augmentation des débits observés, GR ne réagit pas et ne prévoit aucune augmentation de débit, contrairement à SIM. Au mois d'Avril, GR se rattrape et prévoit nettement mieux. Cette crue, en regardant de près la précipitation observée, est due à la pluie, tandis que celle de Mars semble être due à la fonte de la neige.

En effet, après analyse de la localisation du bassin versant, de l'évolution de la température et du stock de neige, une crue nivale a été mise en évidence pour l'évènement de Mars 2005. La figure 10 montre une hausse de température entraînant une diminution du stock de neige. Une pointe de débit arrive vers le 15 Mars due à la fonte de la neige et provoque la pointe de crue que GR ne

prévoit pas (figure 9, gauche). Au mois d'Avril, on observe que les pics de débit sont les réponses à des événements de pluies. Dans ce cas, les prévisions GR sont meilleures (figure 9, droite).

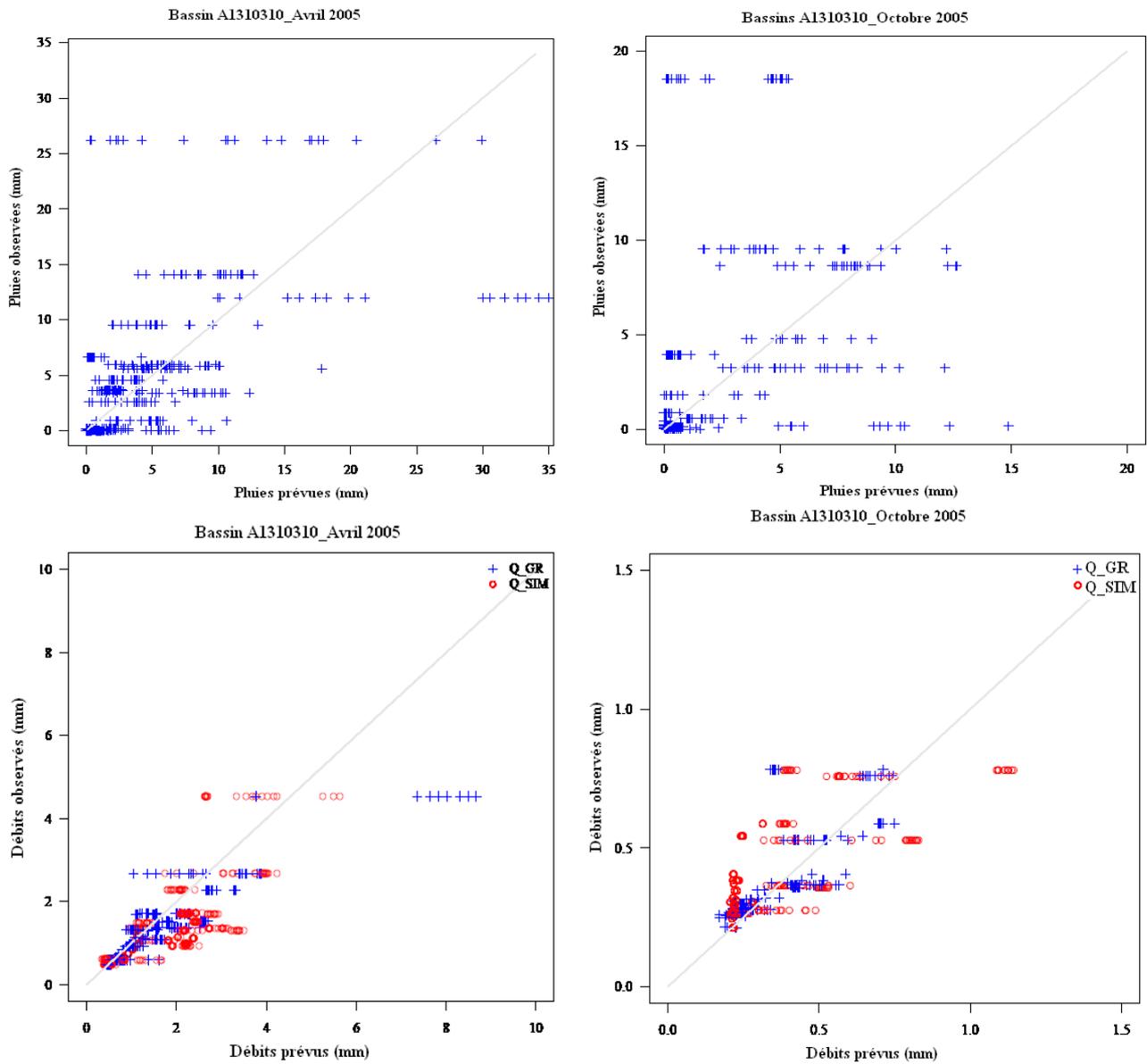


Figure 7 : Données du bassin A1310310: L'III à Ensisheim, 1038 km² pour les mois d'Avril et Octobre. Echéance 1

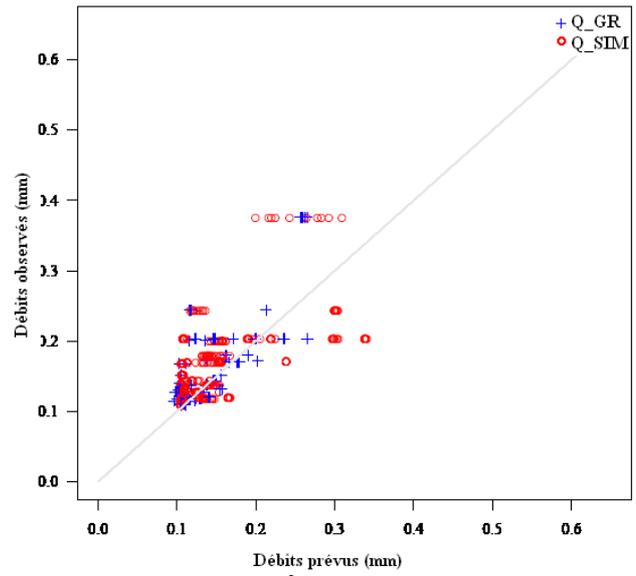
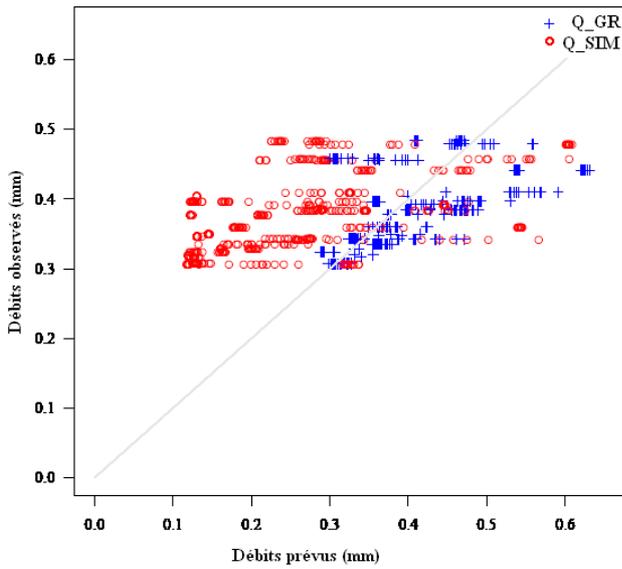
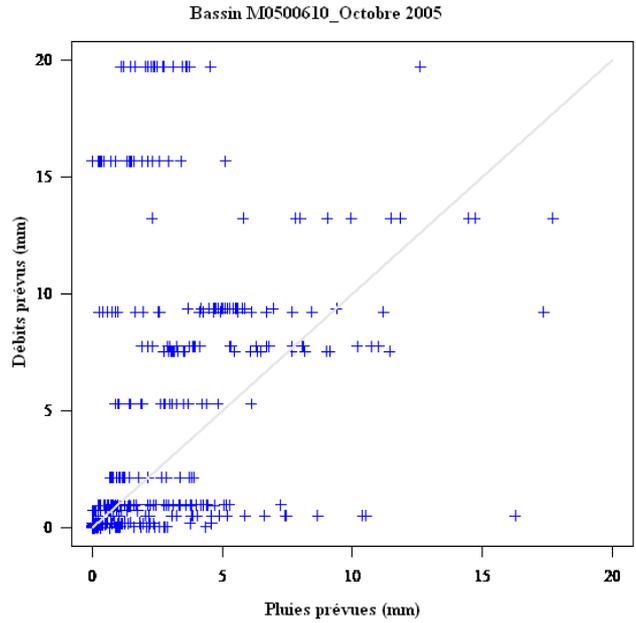
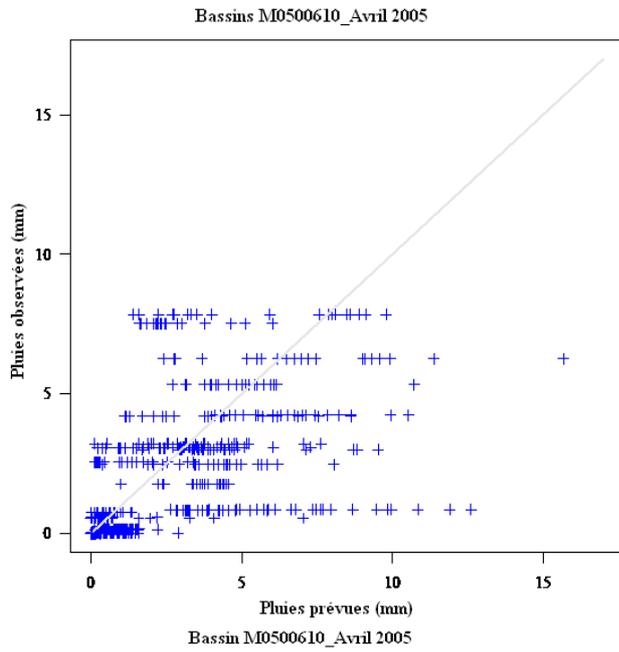


Figure 8 : Données du bassin M0500610: la Sarthe à Spay, 5285 km² pour les mois d'Avril et d'Octobre. Echance 1

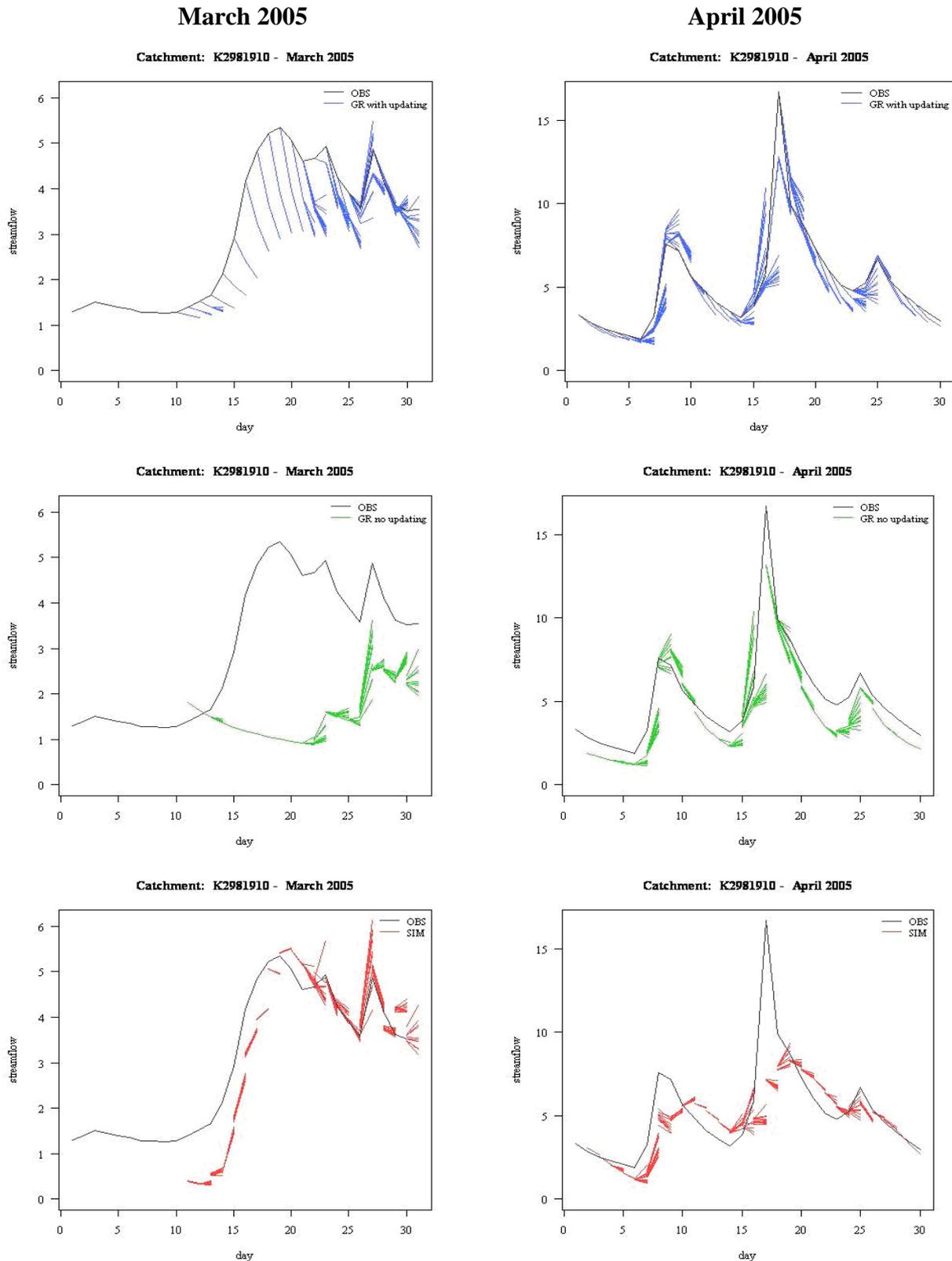


Figure 9 : Hydrogrammes des débits prévus et observés du bassin K2981910: la Dore à Dorat, 1523 km² pour les mois de Mars et Avril

Ce cas illustre bien l'importance de prendre en compte la neige dans les modèles hydrologiques et la nécessité d'un module neige sur GR, qui permettrait de prendre en compte les variations de température et la fonte de la neige au printemps, facteurs causant des crues non négligeables. Ce qu'illustre la figure 10 : vers début mars, la température augmente, causant une diminution du

stock de neige et une hausse de débit. GR, dans sa version avec module neige (en développement), arrive à prévoir une augmentation de débit (léger pic) et réussit donc à mieux reproduire la montée des débits, par rapport au cas de modélisation sans module neige. Pour le mois d'Avril, les crues sont mieux reproduites par le modèle car elles sont essentiellement dues à la pluie. On peut espérer alors que l'introduction d'un module neige dans la version GR pour la prévision des crues (GR3P) pourra augmenter la qualité de ses prévisions.

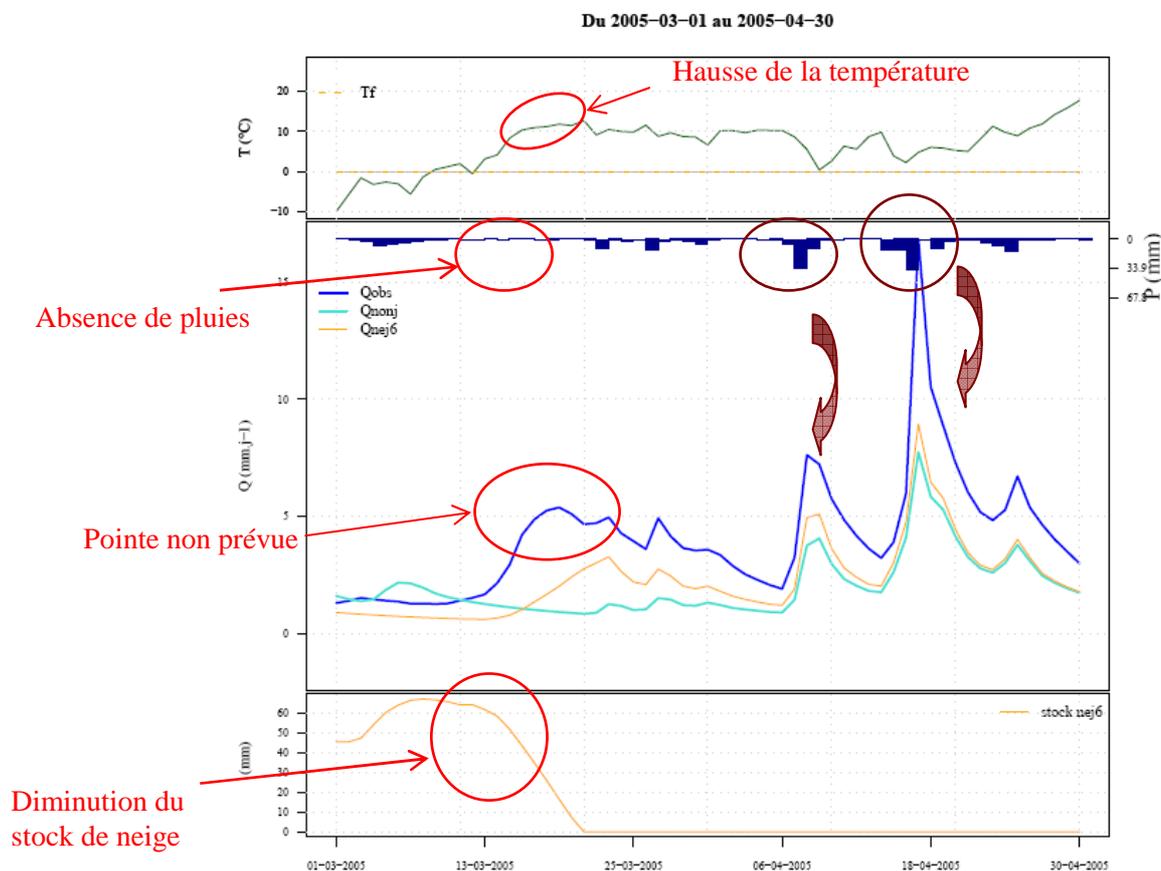


Figure 10 : Evolution du débit avec un module neige sur GR

Source: Extrait des calculs de Valéry Audrey (Cemagref, thèse en cours)

III-2 Erreurs sur les précipitations prévues par PEARP

Le but du stage n'est pas d'évaluer directement les pluies PEARP de Météo-France, mais il est important d'avoir un aperçu de la partie de l'erreur dans les débits qui pourrait être due à la prévision de pluie. La figure 11 montre les cartes des RMSE et ratio-RMSE calculés sur les pluies, tandis que la figure 12 montre les box-plots (quantiles sur les 211 bassins versants) des ratio-RMSE et des ratio-écart-type.

Les RMSE sont supérieurs à 2 mm dans tous les bassins. Les bassins où les erreurs sont les plus faibles sont ceux de la Seine et du quart nord-est de la France. Cette disparité n'apparaît plus dans les résultats des Ratio-RMSE, où la seule tendance visible reste les faibles valeurs dans le nord-est. Les scores de quelques bassins dégradent légèrement avec l'échéance. En regardant les ratios, nous nous rendons compte que dans quelques cas les erreurs dans la prévision de pluie représentent deux fois la pluie moyenne sur la période d'étude. Les erreurs les plus élevées sont dans les bassins situés dans le sud de la France.

Thirel *et al.*, (2008) a calculé ces scores sur ces mêmes données de pluies, mais sur l'ensemble des grilles de la PEARP (9892 grilles). Il trouve les résultats : RMSE du jour 1 = 3.76 mm/jour et RMSE du jour 2 = 3.98 mm/jour.

Durant cette même période, il est indiqué que la pluie moyenne observée est de 2.33 mm/jour. Le rapport entre le RMSE moyen et la pluie observée est alors supérieur à 1.5. Ce résultat montre que les erreurs relativement élevées constatées dans notre étude ne viennent pas de la lecture des données, car elles sont retrouvées en même ordre de grandeur dans le calcul en passant par toutes les mailles d'ARPEGE. Des valeurs relativement faibles sont aussi retrouvées pour les valeurs de dispersion des pluies d'ensemble (écart-type). Le tableau 2 montre que les résultats calculés par grille et par bassin versant se recourent.

	Valeurs de Thirel <i>et al.</i> , (2008), moyenne sur chaque point de grille en mm/j	Valeurs moyennes calculées sur tous les bassins versants en mm/j
RMSE échéance 1	3.76	3.45
RMSE échéance 2	3.98	3.60
Ecart-type échéance 1	0.80	0.58
Ecart-type échéance 2	1.24	1.07

Tableau 2 : Valeurs moyennes des écart-types et des RMSE des pluies PEARP

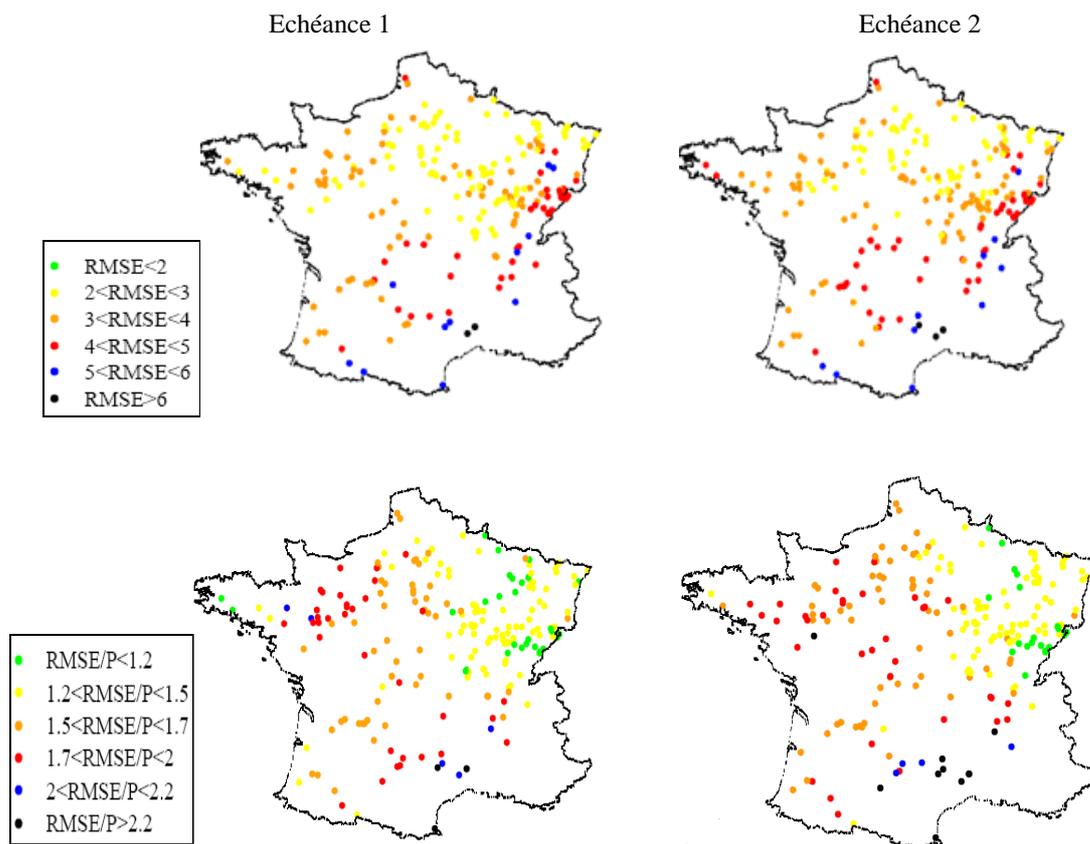


Figure 11 : Cartes des RMSE (haut) et des ratio-RMSE (bas) des pluies sur les deux échéances

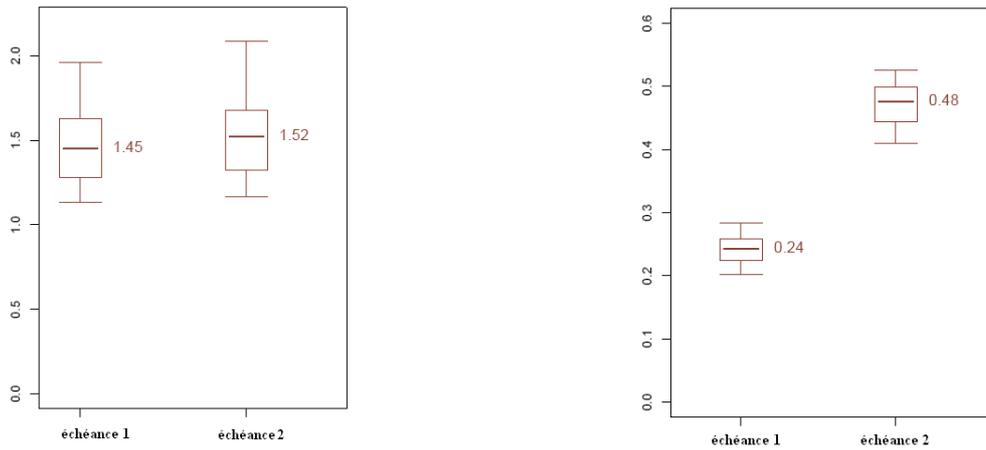


Figure 12 : Boxplots des ratio-RMSE des pluies (gauche) et des ratio-écart-type des pluies (droite)

III-3 Scores sur les débits prévus

III-3-1 Cartographie des erreurs

La figure 13 montre les ratio-RMSE pour les trois modèles pris en compte. Il est à noter que, pour la définition de ces classes, les intervalles pour les petites valeurs sont plus resserrés. Ceci permet de mieux voir la variabilité des RMSE.

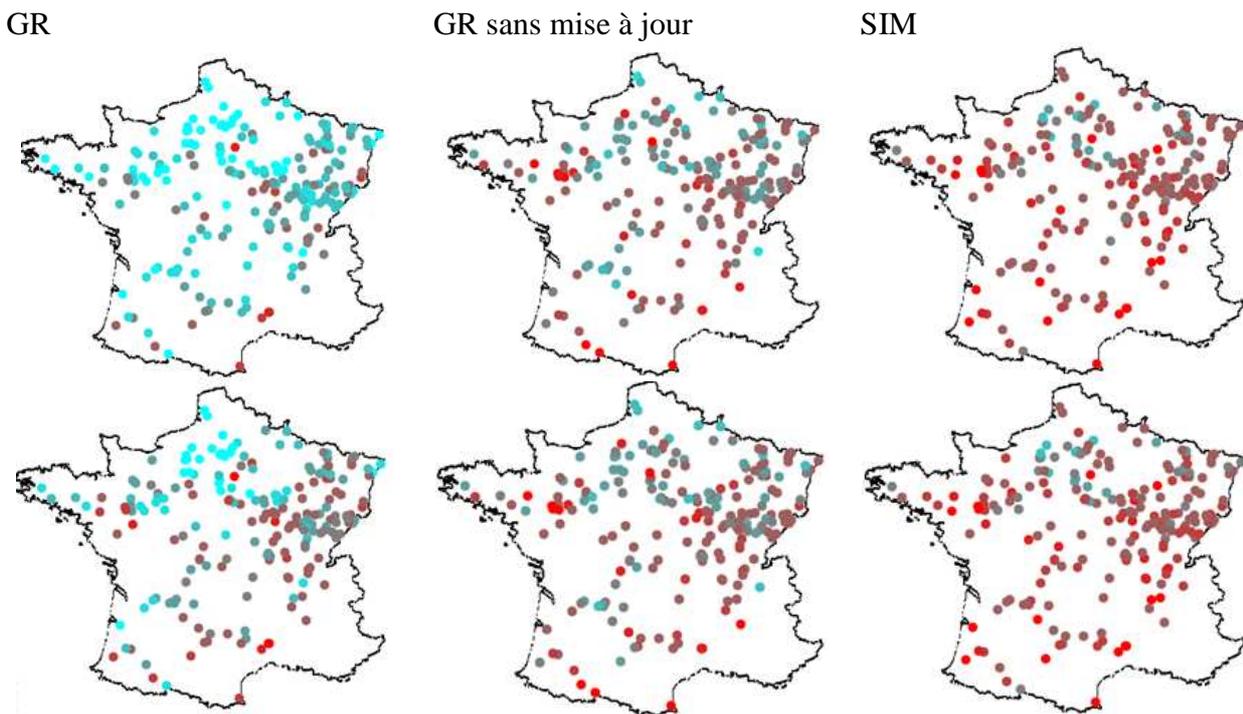
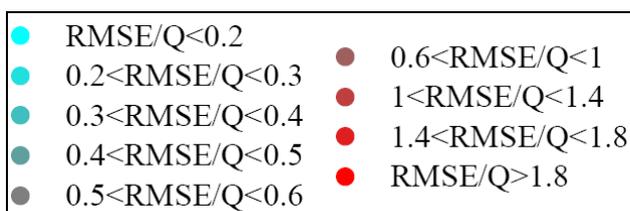


Figure 13 : Valeurs des ratio-RMSE des débits des trois modèles en échéance 1 (haut) et échéance 2 (bas)



Les scores deviennent moins bons avec l'augmentation de l'échéance. GR réalise généralement une bonne performance au niveau des bassins de la Seine. Pour GR sans mise à jour et SIM, il est difficile, d'après ces graphes, de délimiter des zones où les ratio-RMSE sont moins élevés. Dans l'ensemble, les scores de GR sans mis à jour et de SIM ne sont pas éloignés. On remarque que la tendance pour un même bassin est la même : si GR sans mise à jour est bon pour un bassin alors SIM a également un ratio-RMSE plus petit pour celui-ci.

Il était intéressant de voir s'il existait une relation entre les erreurs de débit et de pluie. Les graphes de la figure 14 montrent alors les relations entre les erreurs (ratio-RMSE) de pluie et de débits (GR).

On observe sur ces graphes qu'une corrélation entre les deux erreurs n'est pas immédiate. Néanmoins, quelques valeurs de RMSE de pluie plus fortes sont associées aux RMSE des débits élevés et donc une partie des erreurs sur les débits peut venir des erreurs des pluies. Il est mis en évidence ici également que les erreurs de pluie sont déjà décalés d'une unité (Ratio RMSE P > 1).

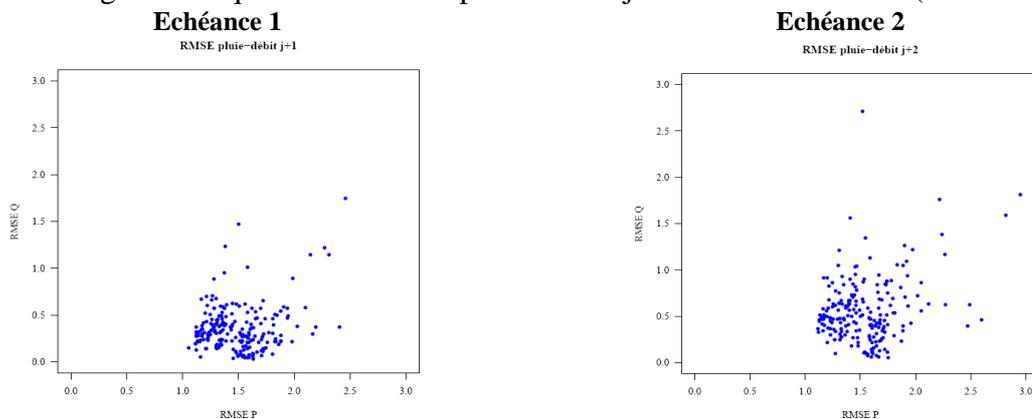


Figure 14 : Relation entre les Ratio RMSE P et Ratio RMSE Q

III-3-2 Dispersion : écart-type et ratio de l'écart-type

Le calcul des écarts-types et des ratio-écarts-types des modèles sont présentés dans la figure 15. La principale remarque que l'on peut faire est le peu de dispersion des débits prévus, notamment pour la première échéance et pour GR avec ou sans mise à jour. La dispersion du modèle GR évolue rapidement en échéance 2, sans toutefois rattraper celle de SIM. Le fait que les membres de GR soient peu dispersés par rapport à ceux de SIM pourrait s'expliquer par son caractère global qui, en moyennant la pluie sur la surface du bassin versant, réduit la variabilité des membres de prévision.

Pour tester un moyen d'introduire de la dispersion dans les sorties de GR, sans toucher au modèle, le caractère distribué des pluies en entrée a été exploité. Pour cela, la pluie prévue de chaque maille a été prise comme une pluie globale du bassin versant. Ainsi, pour chaque maille d'un bassin, nous avons une donnée de pluie avec chacune 11 membres. Des nouvelles "prévisions d'ensemble" sont construites : chaque bassin a alors une prévision avec un nombre de membres égal à 11 fois le nombre de mailles recouvrant le bassin.

Les écarts-types des points de grille sont présentés dans la figure 16 (à droite). Nous avons gardé les mêmes échelles pour faciliter la comparaison. Les écart-types ont donc largement augmenté et sont même devenus trop importants. On note que les valeurs des médianes sont proches de celles de SIM, mais les intervalles interquantiles sont beaucoup plus grands. Le but n'étant pas d'augmenter à tout prix cette dispersion, il faut voir si les résultats restent rationnels. Dans ce cas-ci, certes la dispersion a évolué, mais de manière exagérée sur une bonne partie des bassins versants ; la notion d'"incertitude" justifiant les dispersions peut-être surreprésentée. Des études plus approfondies seraient nécessaires pour mieux examiner cet aspect.

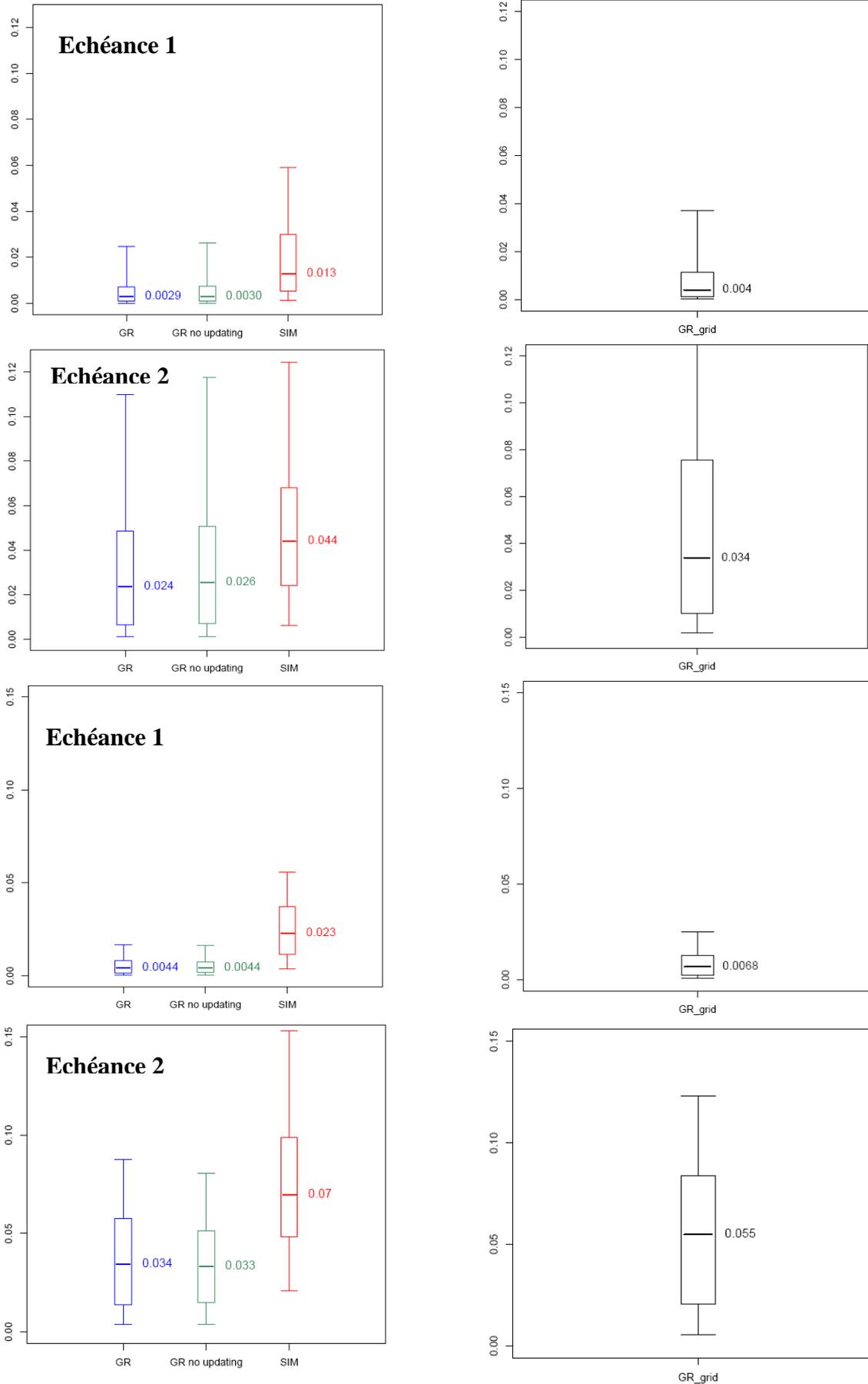


Figure 15 : Ecart-type des débits en échéance 1 et 2 (haut) et ratio-écart-type (bas) à gauche; mêmes scores pour le calcul par mailles avec GR (colonne de droite)

III-3-3 FAR, POD et BIAS : tableau de contingence

Tous les seuils de débits ont été programmés (Q10, Q50, Q90) ainsi que tous les pourcentages de membres de la prévision d'ensemble pour chaque seuil de référence. Cependant, puisqu'on s'intéresse plus dans cette étude aux cas des crues, la majorité des graphes proposés dans cette partie ne montrent que les résultats avec Q50 et Q90. Pour l'échéance, seul le jour 1 est présenté car on distingue très peu de variations entre les résultats des deux échéances.

La figure 16 illustre comment évoluent les scores en fonction de la surface du bassin et pour un nombre de membres croissant dépassant les quantiles de référence. De gauche à droite, **20% membres > Q_{ref}**, **50% membres > Q_{ref}**, **90% membres > Q_{ref}**. On remarque qu'il y a pas de sensibilité au nombre de membres: peu de changement dans les résultats. Les graphes ne changent que légèrement. Ceci est probablement expliqué par la faible dispersion des membres (résultat démontré précédemment).

L'effet de l'augmentation des seuils des débits a été testé en fixant $p_2=50\%$ (Figure 17). Les graphes montrent qu'en augmentant le quantile de référence, les scores deviennent moins bons : les POD diminuent et beaucoup de valeurs sont en dessous de la ligne unitaire. Les FAR s'éloignent de 0. Les BIAS se regroupent toujours autour de l'unité mais avec une variabilité plus importante. Tous ces résultats présentés ont été calculés à partir des débits de sortie du modèle GR dans son fonctionnement avec mise à jour.

La comparaison avec les performances des débits de sortie de SIM et de GR sans mise à jour est présentée dans les graphes de la figure 18. Ils illustrent les scores des modèles pour les seuils de référence Q50 et Q90 et pour un pourcentage fixe de 80 % des membres de prévision d'ensemble dépassant ces seuils de débits. On remarque que les scores sont moins bons pour les modèles qui ne font pas appel à la mise à jour.

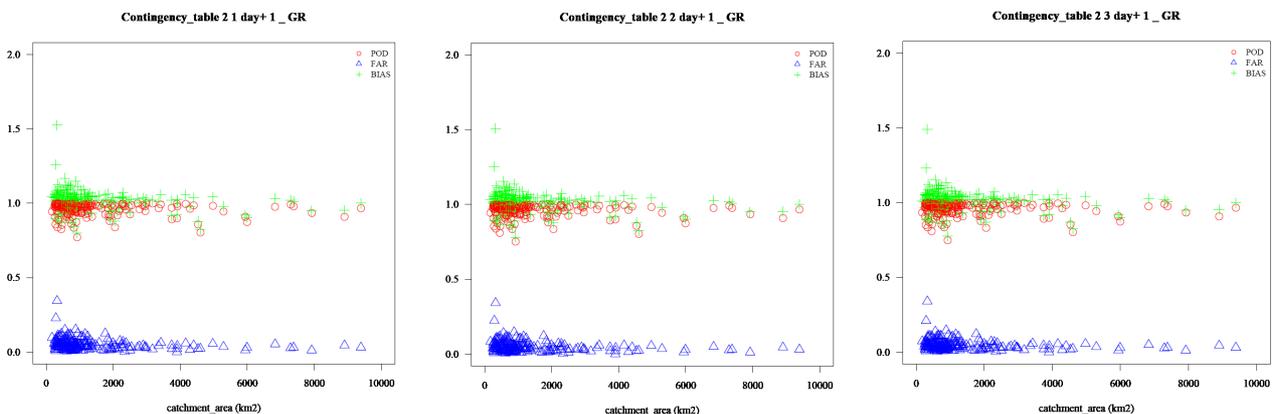


Figure 16 : POD, FAR BIAIS en fonction de la surface du bassin versant pour le seuil de référence de débit Q50 et différents pourcentages de membres de la prévision d'ensemble dépassant ce seuil : p1=20% (gauche), p2=50% (centre) et p2=80% (droite). Echéance jour 1 et Modèle GR.

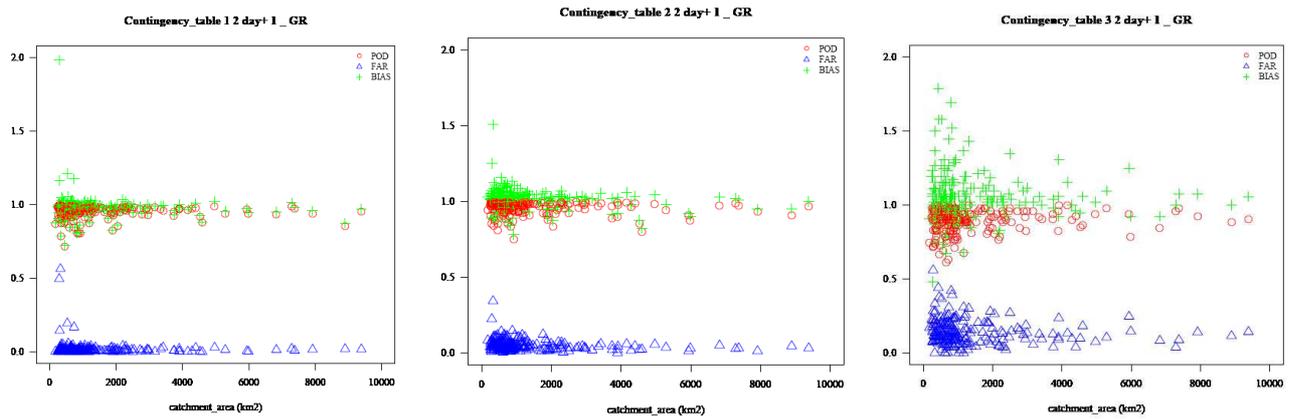


Figure 17 : POD, FAR BIAIS en fonction de la surface du bassin versant pour différents seuils de référence de débit : Q10 (gauche), Q50 (centre) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant (au non-dépassant pour Q10) le seuil de 50%. Echéance jour 1 et Modèle GR.

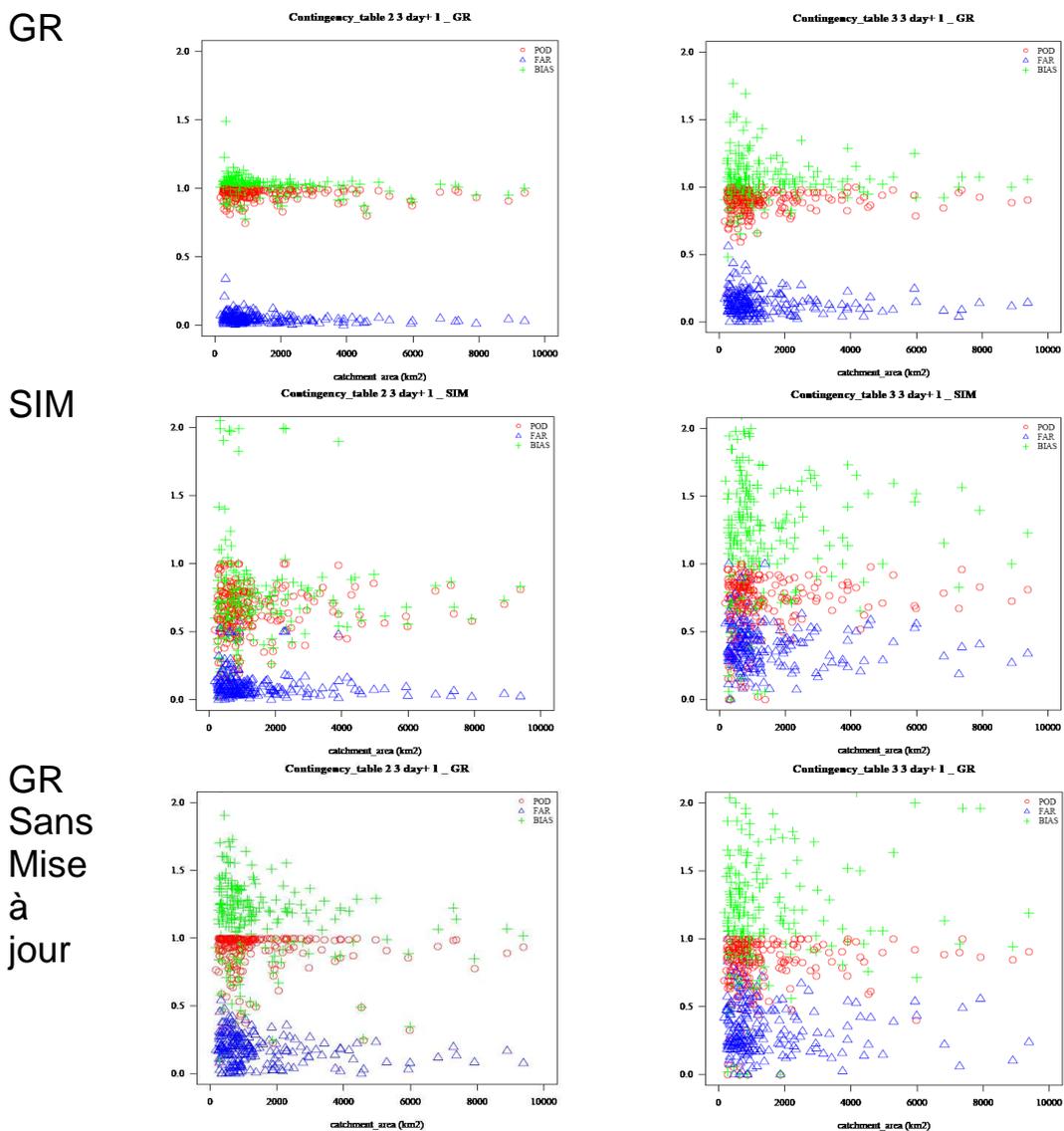


Figure 18 : Comparaison entre les modèles : POD, FAR BIAIS en fonction de la surface du bassin versant pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres 80% de la prévision d'ensemble dépassant le seuil. Echéance jour 1 et Modèle GR (haut), Modèle GR sans mise-à-jour (bas), Modèle SIM (centre).

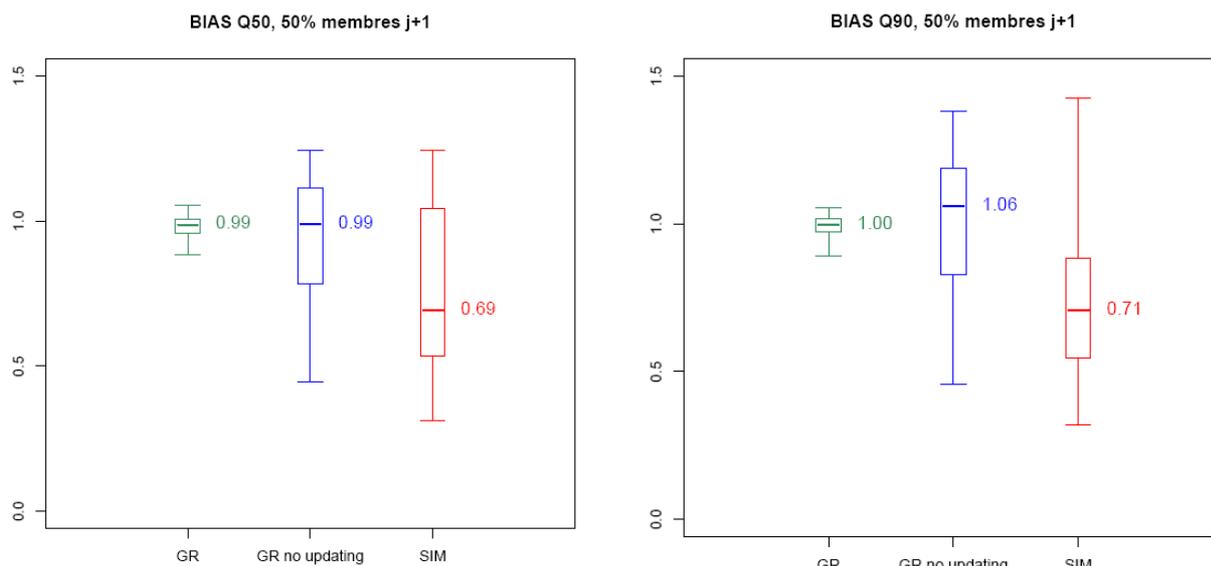


Figure 19 : Comparaison entre les modèles : boxplots des valeurs de BIAIS pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant le seuil de 50%. Echéance jour 1.

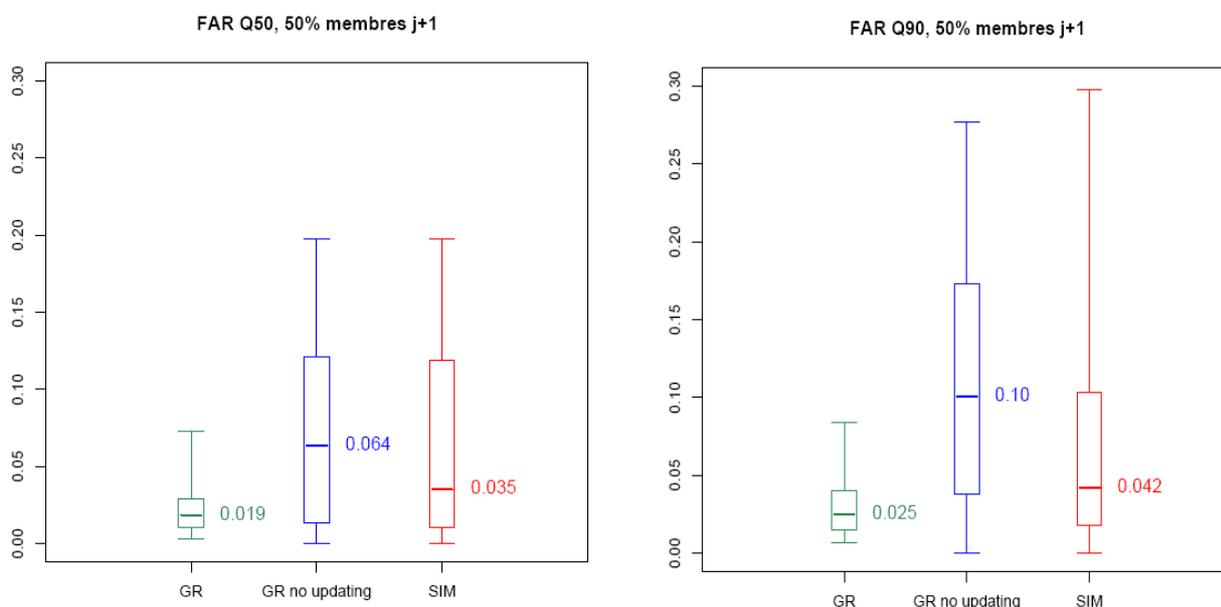


Figure 20 : Comparaison entre les modèles : boxplots des valeurs de FAR pour les seuils de référence de débit Q50 (gauche) et Q90 (droite) et pour un pourcentage de membres de la prévision d'ensemble dépassant le seuil de 50%. Echéance jour 1.

Pour GR, les BIAS sont resserrés autour de 1 (Figure 19) et les variabilités sont faibles, contrairement au cas de GR sans mise à jour et de SIM. GR sans mise à jour a une moyenne et une médiane légèrement supérieures à 1, tandis que SIM présente des valeurs inférieures à l'unité. Un $BIAS^{12} > 1$ représente une sur-prévision et $BIAS < 1$, une sous prévision.

Concernant les probabilités de détection (POD) et les taux de fausses alertes (FAR, figure 20), GR a des meilleurs scores. Cependant, SIM surpasse GR sans mise à jour sur ces deux scores.

En résumé, les principaux résultats déduits des tableaux de contingence sont :

¹² BIAS est le rapport entre le nombre d'évènements prévus sur le nombre d'évènements observés. Quand il est supérieur à 1, cela signifie que l'on a plus prévu un dépassement de seuil par rapport à ce qui a été réellement observé.

- POD, FAR et BIAS ne varient pas significativement avec le pourcentage de membres pris en compte dans le dépassement de seuils de débits (non dépassement pour le quantile 10).
- Ces scores deviennent moins bons en augmentant les quantiles de débit Q50 à Q90, utilisés pour le dépassement de seuils.
- Les scores ne varient pas beaucoup entre les deux échéances.
- Les scores de SIM et de GR sans mise à jour sont moins bons que ceux de GR avec mise à jour.
- GR sans mise à jour a tendance à sur-prévoir ($BIAS > 1$) et SIM à faire l'opposé ($BIAS < 1$).
- Les scores varient plus pour les petits BV. Ils s'améliorent et deviennent stables à partir de 5000 km².

III-3-4 Brier score et Brier Skill score

Tous les scores calculés précédemment sont des scores qui considèrent soit la moyenne de l'ensemble, soit chaque membre (ou *run*) comme une prévision déterministe. Dans la suite, les scores programmés prendront en compte l'aspect probabiliste des prévisions. Dans ce paragraphe, le BS et le BSS permettent de qualifier le comportement du système de prévision face à un évènement binaire de type dépassement de seuil. Les résultats obtenus pour les dépassements des seuils Q50 et Q90 sont présentés dans la figure 21 pour le BS et dans la figure 22 pour le BSS.

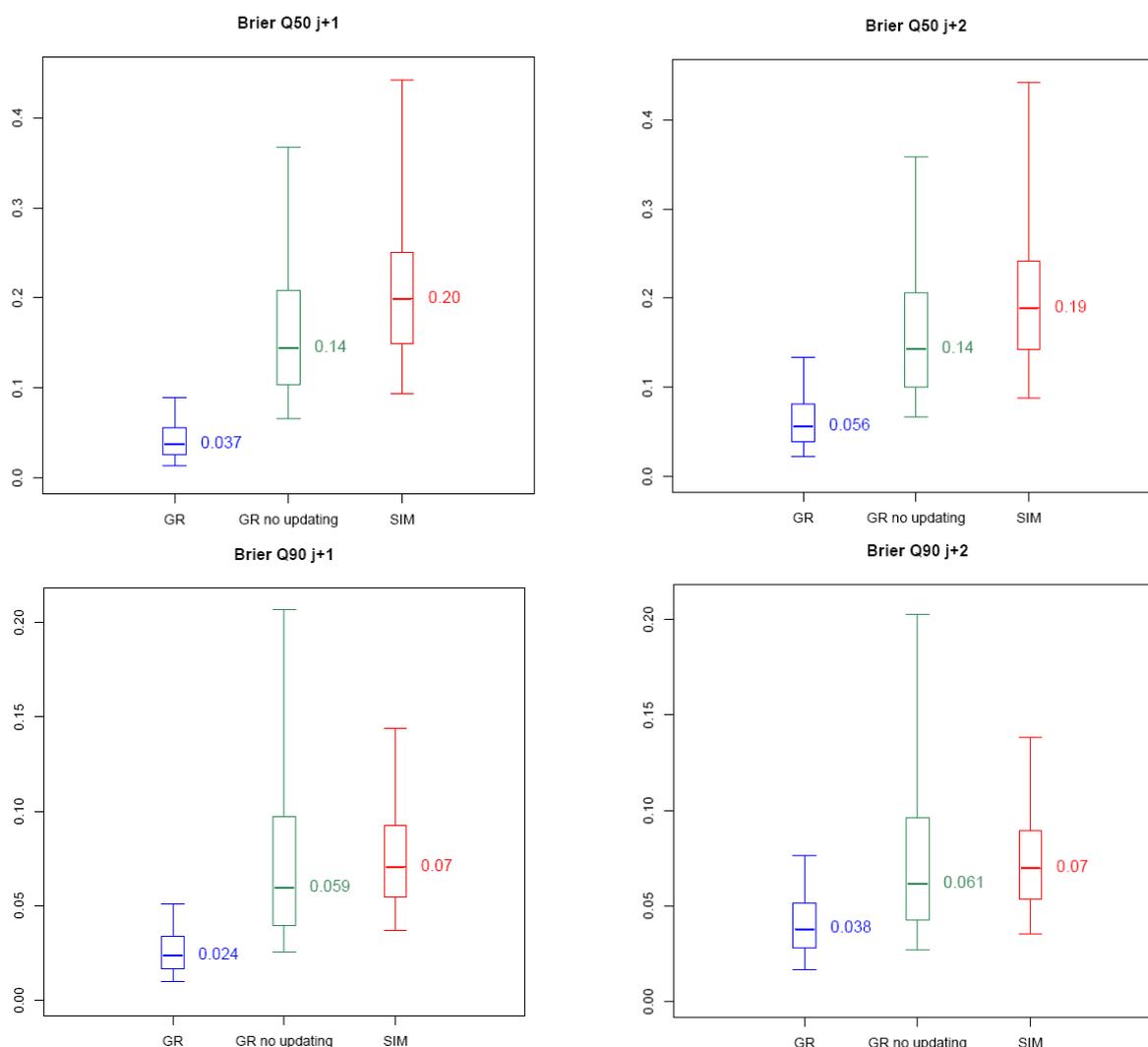


Figure 21 : Brier score pour les seuils Q50 (haut) et Q90 (bas) sur les deux échéances

Les valeurs de BS mettent en évidence une meilleure performance du modèle GR (score plus proche de 0). Entre les deux échéances, les valeurs de GR avec mise à jour empirent et les

intervalles interquantiles augmentent. Cependant, pour GR sans mise à jour et SIM, les différences entre les échéances ne sont pas significatives. Quand on compare les résultats pour les différents seuils (Q10, Q50 et Q90), la meilleure prévision est donc pour les débits plus élevés (scores de Q10 proche de Q50 et supérieures aux scores en Q90).

Pour les BSS (figure 22), il est constaté qu'en général les performances du modèle GR diminuent légèrement avec l'échéance (les valeurs plus élevées indiquent une meilleure performance), tandis que les différences sont moindres pour les modèles SIM et GR sans mise à jour. Les scores sont meilleurs en Q90. Les scores qui ont pour référence la climatologie sont plus élevés que les BSS calculés par rapport à la persistance : il est donc plus facile de prévoir mieux que la climatologie (c'est-à-dire que la prévision est plus proche de la dernière observation que du débit moyen journalier calculé sur une longue période).

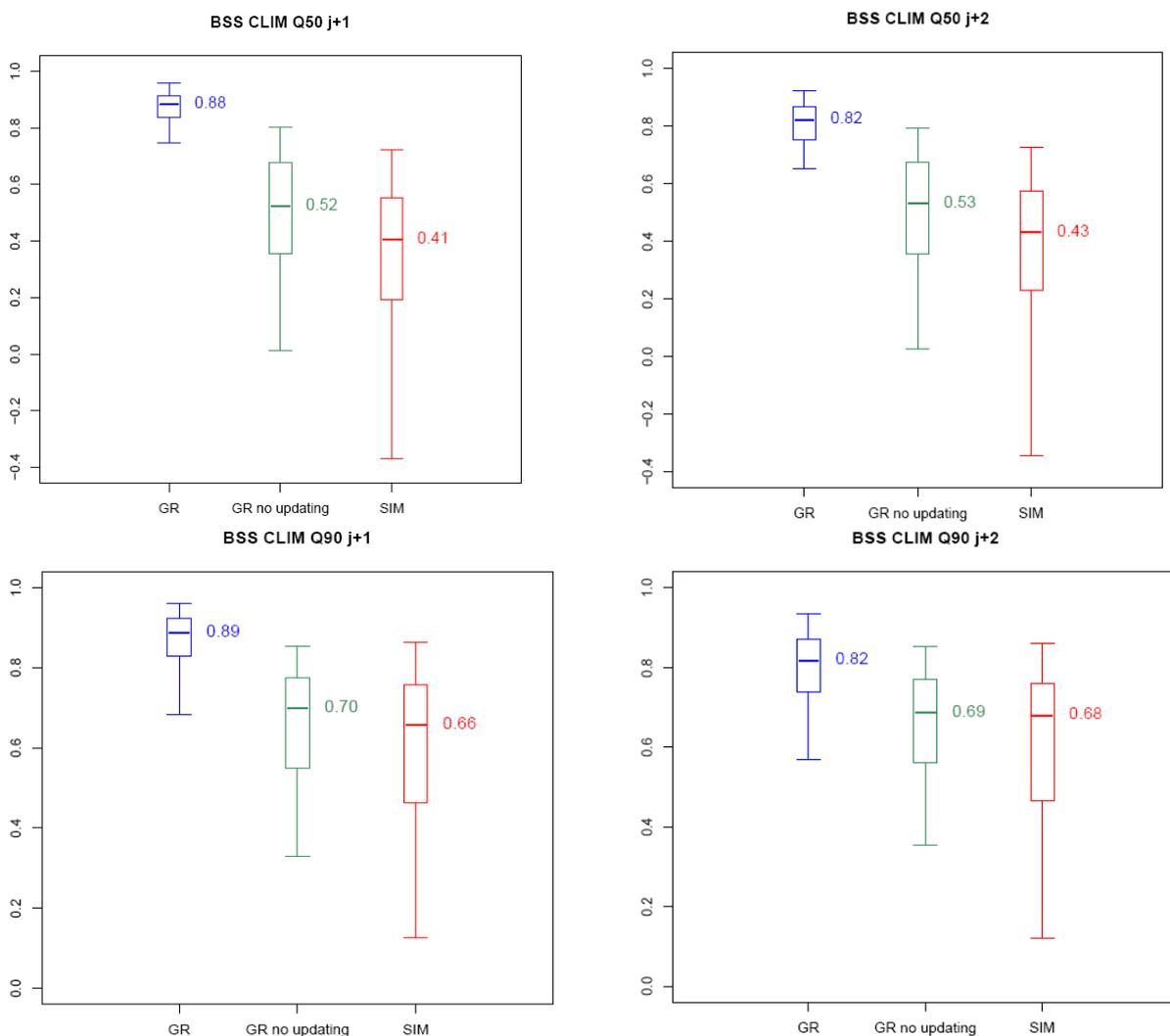


Figure 22 : Brier Skill Score de référence climatologique pour les seuils Q50 (haut) et Q90 (bas) pour les deux échéances

III-3-5 Le RPSS

La figure 23 présente les résultats des RPSS de référence climatologique pour les trois modèles. Les valeurs des RPSS montrent aussi une meilleure performance pour le modèle GR avec mise à jour (les valeurs des RPSS sont bonnes lorsqu'elles sont positives et s'approchent de 1). Les scores se détériorent légèrement en échéance 2. GR est le seul modèle à présenter des scores positifs ; il est donc ici le seul modèle dont les performances sont supérieures à celle de la climatologie. Les

valeurs médianes des deux autres modèles sont négatives, indiquant que ces derniers donnent des performances de prévision inférieures à la climatologie.

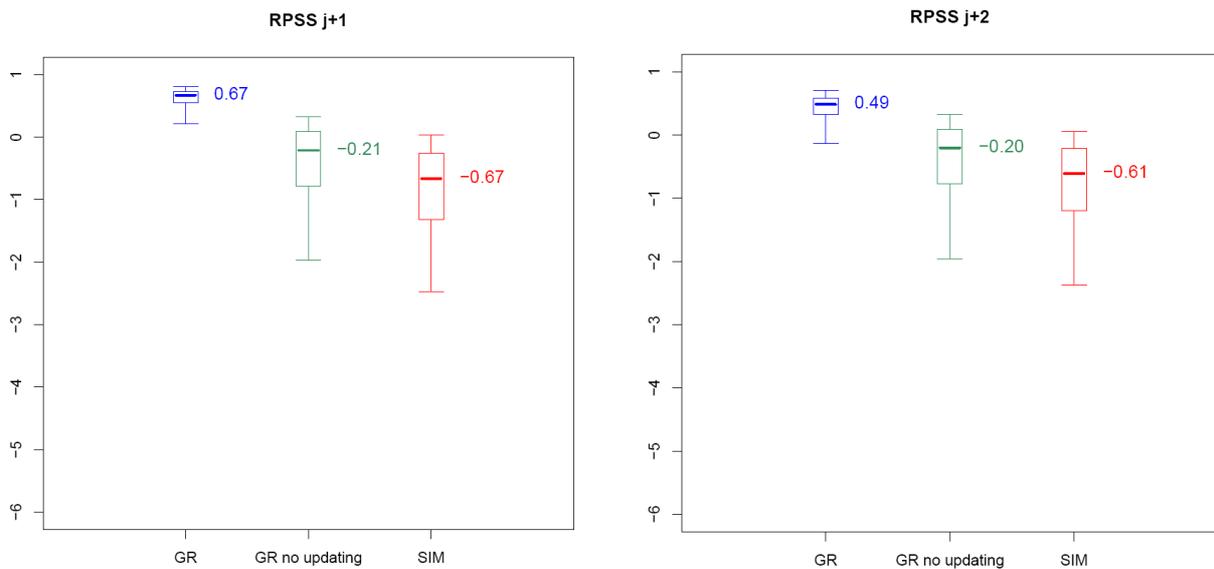


Figure 23 : RPSS pour les deux échéances et les trois modèles par rapport à la référence climatologique

III-3-6 Diagramme de Talagrand

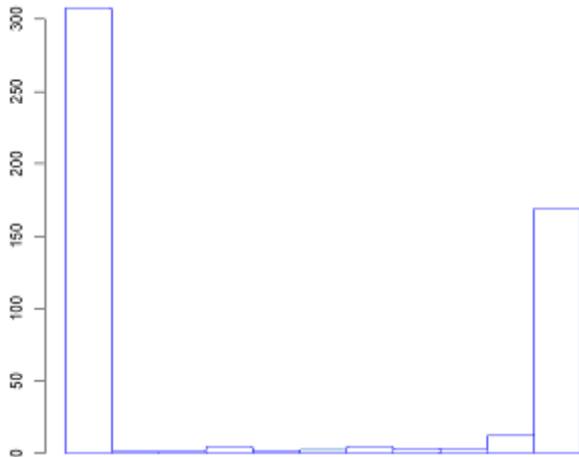
Idéalement, dans une prévision parfaite, les valeurs de l'ensemble et la vérification (ici, les débits observés) classées sont comme des échantillons aléatoires de la même distribution de probabilité. En comptant le rang de la vérification pour un certain nombre de cas indépendants, on doit donc obtenir un histogramme plat du nombre de cas pour chaque rang. Les diagrammes de Talagrand présentés dans la figure 24 ont été choisis car ils représentent assez bien les formes des résultats obtenus sur l'ensemble des 211 bassins versants.

Les bassins présentés sont :

- M0243010 : L'Orne Saosnoise à Montbizot, Sarthe (72), 510 km²
- A1080330 : L'Ill à Didenheim, Haut-Rhin (68), 656 km²

Dans tous les cas, les diagrammes tracés sont en forme de U, donc les observations sont soit majoritairement inférieures aux prévisions, soit supérieures à celles-ci. Ceci confirme alors la sous-dispersion des membres. Il est probable que toutes les incertitudes ne sont pas prises en compte dans la prévision d'ensemble (valable autant pour les précipitations que pour les débits).

M0243010: l'Orne Saosnoise à Montbizot



A1080330: l'Ille à Didenheim

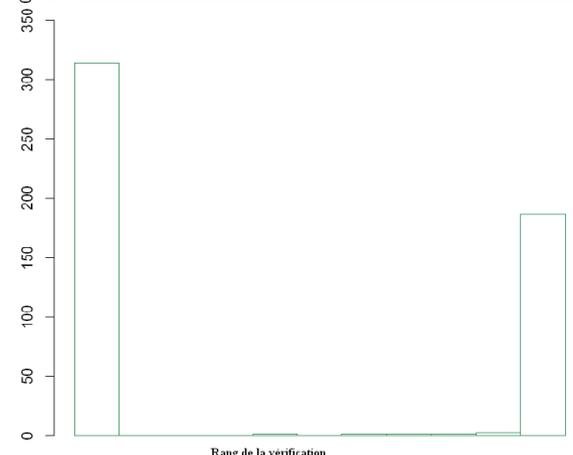
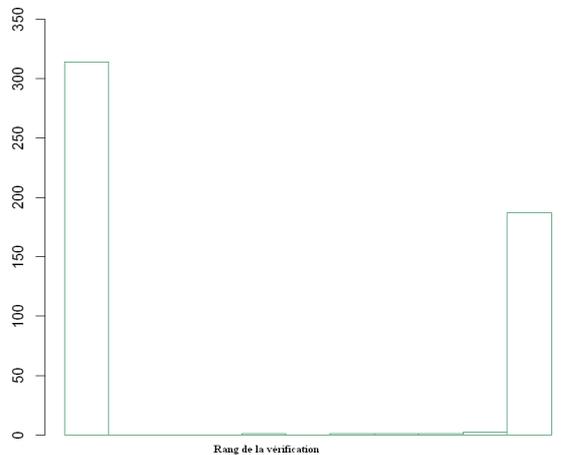
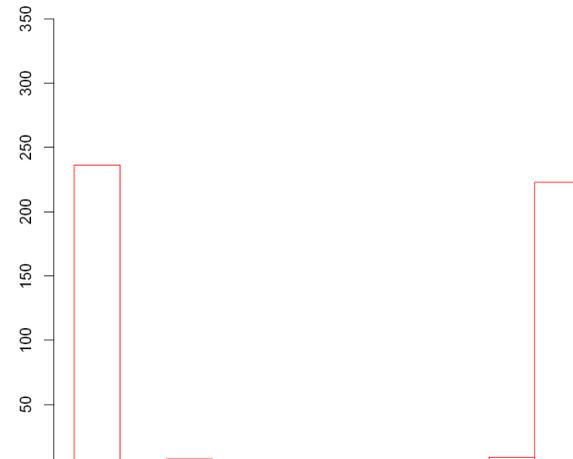
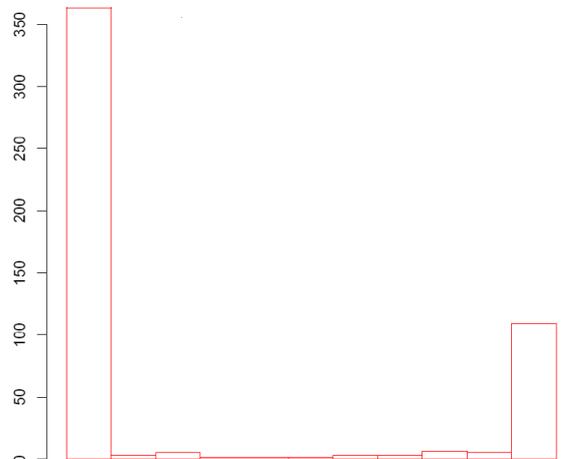
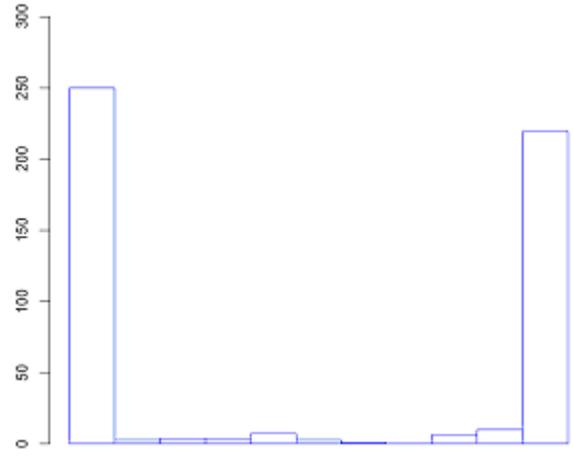


Figure 24 : Diagrammes de Talagrand des bassins M0243010: l'Orne Saosnoise à Montbizot (gauche), A1080330: l'Ille à Didenheim (droite), à partir des débits prévus par de GR (haut), de SIM (centre) et GR sans mise à jour (bas).

III-3-7 Synthèse des scores moyens

Etant donné la grande quantité de données analysées, tous les graphes et toutes les valeurs calculées n'ont pas pu être présentés dans cette étude. Le tableau suivant (tableau 3) présente une synthèse des scores moyens obtenus.

		GR (with updating)						GR (no updating)						SIM					
p = 20% members	Leadtime	Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90	
		Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
		0,9574	0,937	0,9639	0,9482	0,895	0,851	0,8416	0,8417	0,891	0,891	0,816	0,802	0,7214	0,7351	0,675	0,691	0,751	0,747
		0,0206	0,026	0,0512	0,0828	0,145	0,246	0,0443	0,0452	0,18	0,184	0,351	0,374	0,0464	0,0492	0,104	0,114	0,381	0,399
		0,984	0,97	1,0181	1,0387	1,059	1,16	0,8935	0,8946	1,121	1,126	1,424	1,445	0,7746	0,7907	0,786	0,813	1,344	1,387
p = 50% members	Leadtime	Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90	
		Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
		0,9567	0,9337	0,9629	0,9443	0,893	0,835	0,8411	0,8393	0,89	0,887	0,814	0,791	0,7144	0,7167	0,6669	0,6669	0,743	0,724
		0,02	0,0245	0,0497	0,0762	0,139	0,223	0,0442	0,0448	0,179	0,18	0,348	0,356	0,0453	0,0462	0,0994	0,0963	0,372	0,374
		0,9827	0,9643	1,0154	1,0264	1,048	1,1	0,8929	0,8917	1,119	1,116	1,414	1,388	0,7668	0,7701	0,7722	0,77	1,309	1,284
p = 80% members	Leadtime	Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90		Qref = Q10		Qref = Q50		Qref = Q90	
		Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
		0,9562	0,9309	0,9621	0,9399	0,89	0,816	0,8407	0,8371	0,89	0,884	0,812	0,777	0,7078	0,7001	0,6597	0,6446	0,737	0,7
		0,0197	0,0237	0,0486	0,0711	0,135	0,208	0,0441	0,0444	0,179	0,177	0,346	0,343	0,0445	0,0445	0,0957	0,0851	0,365	0,349
		0,9819	0,9607	1,0132	1,0156	1,039	1,051	0,8923	0,8891	1,117	1,108	1,405	1,339	0,7596	0,752	0,7609	0,7362	1,283	1,205
Streamflows		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2			
Ratio-rmse	0,360		0,565		0,745		0,803		1,040		1,087								
Ratio-σ	0,006		0,038		0,006		0,036		0,028		0,079								
Persistence	52,34		50,29		-488		-165		-1520		-626,9								
RMSE	0,314		0,482		0,569		0,621		0,781		0,826								
Ecart-type	0,006		0,035		0,006		0,037		0,020		0,054								
RPS	0,034		0,052		0,120		0,121		0,176		0,165								
RPSS	0,674		0,481		-0,234		-0,221		-0,671		-0,612								
Brier Score		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2							
BS10	0,046		0,070		0,180		0,181		0,291		0,282								
BS50	0,043		0,065		0,169		0,168		0,215		0,206								
BS90	0,026		0,041		0,080		0,080		0,082		0,081								
BSS10	0,555		0,328		-0,057		-0,041		-0,735		-0,730								
BSS50	0,870		0,804		0,481		0,480		0,479		0,483								
BSS90	0,864		0,790		0,450		0,641		0,647		0,644								

Tableau 3 : Tableau récapitulatif des moyennes des scores de vérification

III-3-8 Les principaux résultats

Concernant la précipitation :

- La carte des RMSE montre que les scores sont plus mauvais pour l'échéance 2 sur quelques bassins versants.
- Le calcul avec le ratio des RMSE, (c'est-à-dire RMSE normalisé par la pluie moyenne observée sur la période de vérification) indique que le RMSE médian est égal à 2-3 fois la pluie moyenne. Les erreurs sur les pluies sont donc déjà considérables.

Les valeurs de POD, FAR et BIAS :

- les scores ne diffèrent pas beaucoup pour les différents pourcentages de membres définis pour le dépassement des seuils de débits.
- FAR augmente et POD diminue en augmentant le seuil de débit. L'évolution des BIAS va dans le sens de la sur-prévision¹³ (c'est-à-dire que la majorité des valeurs des BIAS sont supérieures à 1).
- Les valeurs élevées de FAR et les faibles valeurs de POD sont observées pour les petits bassins versants. Pour ces derniers, le BIAS supérieur à 1 montre une tendance à sur-prévoir. Au-dessus de 5000 km², les scores s'améliorent (POD et BIAS sont proches de 1 et FAR de 0) et deviennent plus stables (les différences entre les bassins versants sont moins significatives). On note pourtant que la base de données ne présente pas beaucoup de bassins dont la taille est supérieure à 5000 km².
- Les différences des scores entre les deux échéances sont plus importantes pour les petits bassins ; le POD devient plus faible tandis que FAR augmente et les valeurs des BIAS ont tendance à s'éloigner de la ligne unitaire (=1).
- Concernant les valeurs médianes sur tous les bassins : POD, FAR et BIAS calculés à partir des sorties de SIM sont proches de celles de GR sans mise à jour. On remarque également que le BIAS de GR sans mis à jour est supérieur à 1, donc ce modèle a tendance à sur-prévoir. Ce qui n'est pas le cas de SIM, dont la médiane est inférieure à 1: SIM sous-prévoit. Les scores sont globalement mauvais par rapport au GR mis à jour. Les différences sont observées aussi bien pour les petits et les grands BV. En résumé, la mise à jour de GR améliore considérablement les scores du tableau de contingence.

Sur les débits :

- La carte des RMSE montre que les scores deviennent mauvais en échéance 2 (jour+2).
- On note également que les différences entre les deux échéances sont plus importantes pour le modèle GR, tandis que le modèle SIM semble être moins impacté par l'augmentation de l'échéance.
- Lorsque l'on compare les scores de SIM et de GR (avec mise à jour), on observe des valeurs proches sur les bassins versants de la Seine et de la Loire aval. Les RMSE de GR sont inférieurs à ceux de SIM pour les bassins du Rhône, de la Loire amont, de la Garonne. Ces différences tendent à diminuer lorsque l'on considère GR sans mise à jour.
- En considérant les ratios des RMSE, on observe toujours que les valeurs de GR sont inférieures à celles de SIM. On n'observe plus alors les mêmes zones définies précédemment. SIM est globalement moins performant que GR avec mise à jour (mais plus proche de GR sans mise à jour). Pour tous les autres scores, les différences entre GR et SIM sont moins importantes en considérant GR sans mise à jour.

¹³ Le score BIAS du tableau de contingence montre soit une sur-prévision, soit une sous-prévision: on dit qu'il y a sous-prévision (sur-prévision) quand le système prévoit moins (plus) fréquemment un dépassement de seuil qu'il en est observé.

L'indice de persistance est un score difficile à interpréter : des valeurs relativement élevées et positives signifient que le modèle est meilleur qu'un modèle naïf qui considère que la valeur de la prévision est égale à la dernière observation. Ce modèle naïf perd en augmentant les échéances, ce qui est aussi le cas du modèle basé sur les prévisions de pluie. Les résultats montrent que GR avec mise à jour présente de bons scores en persistance (les valeurs sont positives, à part quelques bassins). Ces valeurs sont meilleures que celles de GR sans mise à jour. Pour SIM, la persistance est majoritairement négative déjà pour l'échéance 1. A part GR avec mise à jour, pourtant, l'indice de persistance augmente considérablement à l'échéance 2.

L'écart-type permet de mesurer la dispersion des membres autour de la moyenne de l'ensemble :

- En général, l'écart-type (aussi bien de la précipitation que des débits) est plus élevé à l'échéance 2 ; les membres sont donc plus dispersés à jour+2.
- GR est moins dispersé que SIM. Les moyennes des valeurs sont à peu près égales pour GR avec et sans mise à jour, ce qui est explicable par le fait que l'approche de modélisation pluie-débit est la même et la comparaison ne tient pas en compte si les membres sont proches ou éloignés du débit observé. Au contraire, le score RMSE tient compte de cette différence, et les scores de GR avec mise à jour sont alors meilleurs.
- Pour les deux modèles GR, la dispersion augmente avec l'échéance, mais elle reste faible par rapport à la dispersion de SIM, même si GR augmente plus rapidement (dispersion au jour 2 = dispersion au jour 1*6) que SIM (3 fois plus).

Pour le BS et le BSS :

- La performance des prévisions de GR avec mise à jour est meilleure : le score BS est proche de 0.
- Les modèles sont plus performants quand on considère le seuil de référence Q90, ce qui montre que les modèles sont meilleurs en prévision de dépassement de ce seuil, donc pour les débits plus élevés.
- Pour les BSS, les scores qui ont pour référence la climatologie sont meilleurs par rapport à ceux qui utilisent comme référence la persistance. Les scores sont globalement supérieurs à 1, surtout pour le Q90. Cela veut dire que la prévision est meilleure que la climatologie, ce qui est normal car il est facile de prévoir mieux que la moyenne des débits journaliers dès qu'il y a une crue, par exemple.

Le RPS et le RPSS apportent plus que le Brier car ils donnent beaucoup plus de poids aux valeurs éloignées de l'observation. Du fait qu'il s'agisse de probabilités cumulées, les scores prennent en compte la distance entre la prévision et l'observation. Une prévision qui n'est pas dans la même catégorie que l'observation, mais dans une catégorie proche, aura un meilleur score que la prévision qui tombe dans une catégorie plus éloignée. Les résultats montrent que les RPSS (par rapport à la climatologie) sont proches de 1 et positifs pour GR avec mise à jour sur au moins 90% des stations : l'ensemble est une meilleure prévision que la climatologie. Pour SIM et GR sans mise à jour, les RPSS sont loin de 1 et positifs pour seulement 5% à 25% des stations. Pour ces modèles, la dégradation des scores avec l'échéance est moins prononcée.

Tous les diagrammes de Talagrand pour tous les modèles étudiés ont une forme en U, avec des valeurs élevées dans la première et la dernière classe. Cela confirme le manque de dispersion des ensembles hydrologiques, mais est aussi le signe d'un biais dans les prévisions (toutes les valeurs sont soit inférieures aux observations, soit supérieures).

Conclusion

La synthèse des résultats et leur interprétation constituent une étape difficile dans la prévision d'ensemble. En effet, les variables sont très nombreuses et les valeurs moyennes seules ne reflètent pas toute la richesse et la variabilité des résultats. Les scores, eux aussi, sont parfois difficiles à interpréter, non seulement à cause de leur diversité, mais également à cause des paramètres qui interviennent dans leurs calculs et de la relation probabiliste de chaque prévision. De plus, les scores n'indiquent pas toujours le même sens, c'est-à-dire que les conclusions obtenues par un score ne sont pas toujours compatibles avec un autre.

Dans cette étude, la principale conclusion reste la sous-dispersion des débits simulés : tous les résultats des scores vont d'ailleurs dans ce sens. Certes, il est intéressant dans la définition même de la prévision d'ensemble d'être dispersée, mais il est important que toute démarche entreprise pour prendre en compte les incertitudes de prévision reste réaliste et hydro-météorologiquement cohérente. Il est souligné également ici l'importance de la mise à jour, et donc de l'assimilation du dernier débit observé, pour une meilleure qualité des prévisions des crues.

Nous avons fait un effort pendant ce stage pour mener une analyse comparative en étudiant les sorties en débits de modèles hydrologiquement différents, mais qui ont utilisé une même entrée de précipitation prévue, la PEARP, et sur lesquelles on applique les mêmes méthodes, les mêmes références et les mêmes scores.

La comparaison de nos résultats avec d'autres études est complexe : les données, les bassins, les méthodes, la climatologie et les modèles sont différents. Citons, par exemple, le cas en France de l'étude de Thirel et *al.*, (2008). Il a calculé les scores sur une période de 569 jours, avec 2 échéances et en prenant en compte les valeurs de chaque point de grille dans les bassins pour le modèle SIM. Il a obtenu comme BSS en moyenne: $Q50 : 0.88 ; 0.86$ (échéance 1 et 2) et $Q90 : 0.91 ; 0.88$ (contre $Q50 : 0.47 ; 0.48$ et $Q90 : 0.65 ; 0.64$ dans cette étude) ; pour le RPSS moyen, Thirel (communication personnelle) obtient 0.05 pour le modèle SIM sans mise à jour et 0.35 avec assimilation (contre -0.6 dans ce travail pour SIM sans mise à jour). Ces scores sont meilleurs que ceux trouvés dans ce rapport. Cela peut-être expliqué par le fait que la vérification n'est pas faite avec les réelles observations, mais avec des débits simulés avec des observations de pluie en entrée. Par ailleurs, c'est là l'originalité de notre travail : il prend pour la validation les débits réellement observés.

La prévision d'ensemble permet d'avoir non pas une seule valeur prévue, mais plusieurs, qui sont équiprobables. Cet aspect de cette technique de prévision permet au prévisionniste et aux preneurs de décision de mieux prendre en compte toutes les incertitudes liées aux phénomènes naturels et à l'état de l'atmosphère. De ce fait, la prévision probabiliste est particulièrement favorable à la gestion des risques inondations. Il est cependant plus difficile de vérifier la qualité des prévisions d'ensemble, et donc la performance des modèles hydrologiques de prévision, car il est nécessaire d'avoir à disposition une longue série d'évènements prévus et observés. C'est pour cela que des scores adaptés doivent être mis en place pour pouvoir les interpréter et les traiter dans un but d'amélioration de la prévision et de l'émission d'alerte à des dépassements de seuils critiques. La vérification est donc une étape importante dans le développement d'un système de prévisions.

Enfin, nombreux sont encore les améliorations possibles à ce travail : par exemple, la programmation de nouveaux scores, l'augmentation de la période traitée et du nombre de bassins versants, l'amélioration de la dispersion, l'utilisation d'autres scénarios de pluie, l'application pour des utilisations opérationnelles de la prévision d'ensemble, etc.

Annexes



Impact of the use of two different hydrological models on scores of hydrological ensemble forecasts

A. Randrianasejo (1), M.H. Ramos (1), Q. Thirel (2), V. Andréassian (1), E. Martin (2)

(1) Hydrology Research Group, Cemagref HBAN, Antony, France (maria-iviana.ramos@cemagref.fr)
 (2) CNRM-GAME, Météo-France, CNRS, GMMEMOSAYC, Toulouse, France (gulliers.thirel@meteo.fr)



1 Objective

To assess the quality of ensemble streamflow forecasts issued by two different modelling conceptualizations of catchment response, both driven by the same weather ensemble prediction system

2 Data and methods

- 211 catchments in France (170 to 9390 km²) (Fig. 1)
- Weather forecasts from the PEARP ensemble prediction system of Météo-France (March 2005-July 2006):
 - 11 perturbed members for a forecast range of 60 hours (skill scores are computed for the first two days of forecast range)
- Time series of observed data: daily discharge, precipitation, temperature
- Two hydrological models:
 - the coupled physically-based hydro-meteorological model SAFRAN-ISBA-MODCOU developed at Météo-France, based on a distributed catchment model (Fig. 2a)
 - the lumped soil-moisture-accounting type rainfall-runoff model (GR3P) developed at Cemagref (Fig. 2b)



Fig. 1: Location of the catchments

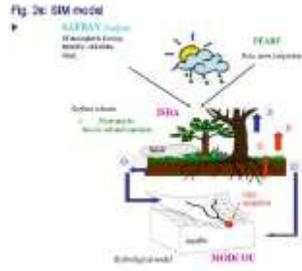


Fig. 2a: SIM model

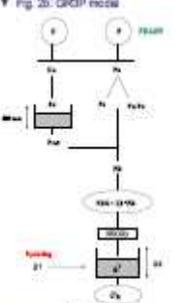


Fig. 2b: GR3P model

3 Skill scores

Contingency table

	Observed	True
Forecast	Hit	Miss
	False alarm	Correctly classified

$ACC = \frac{Hit}{Hit + Miss}$
 $POD = \frac{Hit}{Hit + Miss + False\ alarm}$
 $BIAS = \frac{Hit + False\ alarm}{Hit + Miss}$
 $RF = \frac{Hit}{Hit + Miss + False\ alarm + Correctly\ classified}$

- Two critical thresholds for observed events: Q_{eff1} = 50th percentile (Q50), and Q_{eff2} = 90th percentile (Q90)
 - Threshold for forecasted events: T_p = 50% of the number are greater than Q_{eff1} . The event is considered as a "forecasted event"

4 Results

Fig. 3: POD, FAR, BIAS for leadtime day 1

Fig. 4: Ratio-RMSE values

Fig. 5: Mean values over the 211 catchments

Metric	GR3P model with updating		SIM model		GR3P model without updating	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
Ratio-RMSE	0.56	0.52	1.24	1.07	1.40	0.83
Ratio-C	0.07	0.08	0.23	0.19	0.28	0.28
POD	0.92	0.89	0.89	0.94	0.99	0.94
FAR	0.04	0.07	0.08	0.09	0.03	0.07

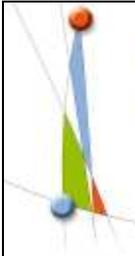
5 Conclusions

- PEARP-based ensemble streamflow forecasts predicted well discharges over the studied catchments
- Better scores are obtained from the GR3P model with updating, while SIM results are closer to the results from GR3P model without updating (for data assimilation in SIM model, see Thirel et al., 2009)
- Ensemble streamflow forecasts from the GR model show lower spread than forecasts from the SIM model
- Ongoing studies focus on the investigation of other probabilistic skill scores

www.cemagref.fr/web/guest

References:
 1. Thirel, Q., Andréassian, V., Lefebvre, F., Lefebvre, F., Adam, P. (2008) On the impact of hydrological models on ensemble streamflow prediction. *Hydrological Science Journal*, 53(1), 104-117.
 2. Legras, M. (2010) Nouvelle méthode de prévision de crues à court terme en mode prédictif. PhD Thesis, Cemagref, Antony, France, 201 p.
 3. Legras, M., Andréassian, V., Lefebvre, F., Adam, P. (2005) Ensemble streamflow prediction: a comparison of two hydrological models. *Hydrological Science Journal*, 50(1), 101-114.
 4. Thirel, Q., Adam, P., Andréassian, V., Lefebvre, F. (2010) An ensemble streamflow prediction system for hydrological forecasting over France. *Journal of Hydrology*, 380(3-4), 339-350.

La liste des bassins est ici présentée, ainsi que les posters qui ont été produits au long de ce stage. Le premier poster présente les premiers scores programmés et présentés à mi-parcours à la conférence EGU en Avril 2009 à Vienne. Le deuxième est celui présenté lors de la conférence HEPEX 2009 à Toulouse en Juin 2009.



Impact of the use of two different hydrological models on scores of hydrological ensemble forecasts

A. Randrianasolo (1), M.H. Ramos (1), G. Thirel (2), V. Andréassian (1), E. Martin (2)

(1) Hydrology Research Group, Cemagref HBAN, Antony, France (maria-helena.ramos@cemagref.fr)
(2) CNRM-GAME, Météo-France, CNRS, GMMEIMOSAYC, Toulouse, France (guillaume.thirel@meteo.fr)

1 Objective: to assess the quality of ensemble streamflow forecasts issued by two different modelling conceptualizations of catchment response, both driven by the same weather ensemble prediction system

2 Data and methods:

- 211 catchments in France (170 to 8390 km²) (Fig. 1)
- Weather forecasts from the PEARP ensemble prediction system of Météo-France (March 2005-July 2008):
 - 11 perturbed members for a forecast range of 60 h (skill scores computed for the first two days of forecast range)
- Time series of observed data: daily discharge, precipitation, temperature
- Two hydrological models:



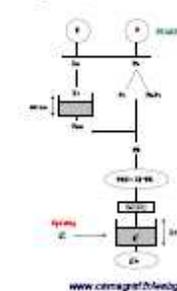
Fig. 1: Location of the catchments

- 1) the coupled physically-based hydro-meteorological model SAFRAN-ISBA-MODCOU developed at Météo-France, based on a distributed catchment model (Fig. 2a)
- 2) the lumped soil-moisture-accounting type rainfall-runoff model (GR3P) developed at Cemagref (Fig. 2b)

Fig. 2a: SIM model



Fig. 2b: GR3P model



3 Skill scores

Contingency table

	Observed	Not	Total
Forecast	Hit	False alarm	Forecast
Not	Miss	Correctly forecasted	Not forecast
Total	Observed	Not observed	Not

- Two critical thresholds for observed events:
Qref1 = 50th percentile (Q50)
Qref2 = 90th percentile (Q90)
- Threshold for forecasted events:
If p = 50% of the members are greater than Qref, the event is considered as a forecasted event

$$POD = \frac{Hit}{Hit + Miss} \quad FAR = \frac{False\ alarm}{Hit + False\ alarm} \quad BIAS = \frac{Hit + False\ alarm}{Hit + Miss}$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - F_i)^2}$$

O_i : observed data for the day
 F_i : mean of the ensemble forecasts for the day
 N : number of days used to compute the score

Standard deviation (or spread)

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - \bar{O})^2}$$

N : number of observed members
 \bar{O} : mean of the ensemble forecasts for the score
 N : number of days used to compute the score
 σ : spread of the member i for the day i

Brier Skill Score

$$BS = 1 - \frac{RMSE^2}{\sigma^2} \quad BSS = 1 - \frac{RMSE^2}{RMSE_{ref}^2}$$

N : number of days used to compute the score
 σ^2 : (the score is zero)
 $\sigma^2 = 1$ (the score does not exist)
 σ^2 : probability of the event to occur

Ratio-RMSE: RMSE / Mean of observed streamflow
Ratio- σ : Standard Deviation / Mean of forecasted streamflow

References

1. Thirel, G., Andréassian, V., Martin, E., Ramos, M.H. (2009) The impact of distributed hydrological models on ensemble streamflow predictions. *J. Hydroinformatics* 11(5), 1201-1217.
2. Thirel, G. (2009) Nouvelle méthode de prévision de crues basée sur un modèle pluie-débit global. PhD Thesis (PhD) Cemagref Paris, 174 p.
3. Thirel, G., B. Martin, J.P. Vazeux, S. Anquetin, B. Rivet, M.H. Ramos (2008) Ensemble streamflow predictions for ensemble observations forecast over France. *Hydrological Sciences Journal* 53(2), 208-220.
4. Randrianasolo, A. (2008) Evaluation de la qualité des prévisions pour l'écoulement des rivières. PhD Thesis (PhD) Cemagref (ongoing)

4 Results

Fig. 3: POD, FAR, BIAS (Leadtime = Day 1 - Qref2 = Q90)

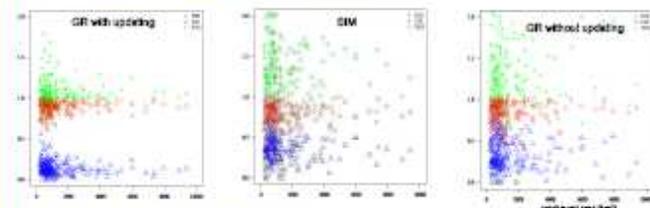


Fig. 4: Ratio-RMSE values (Leadtime = Day 2)



Fig. 5: Spread (Ratio- σ)

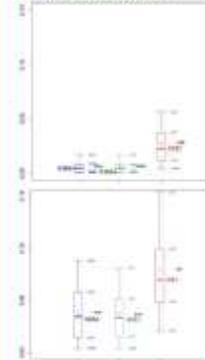
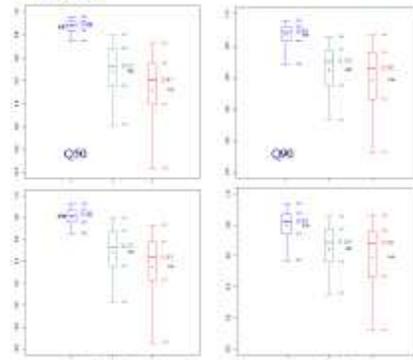
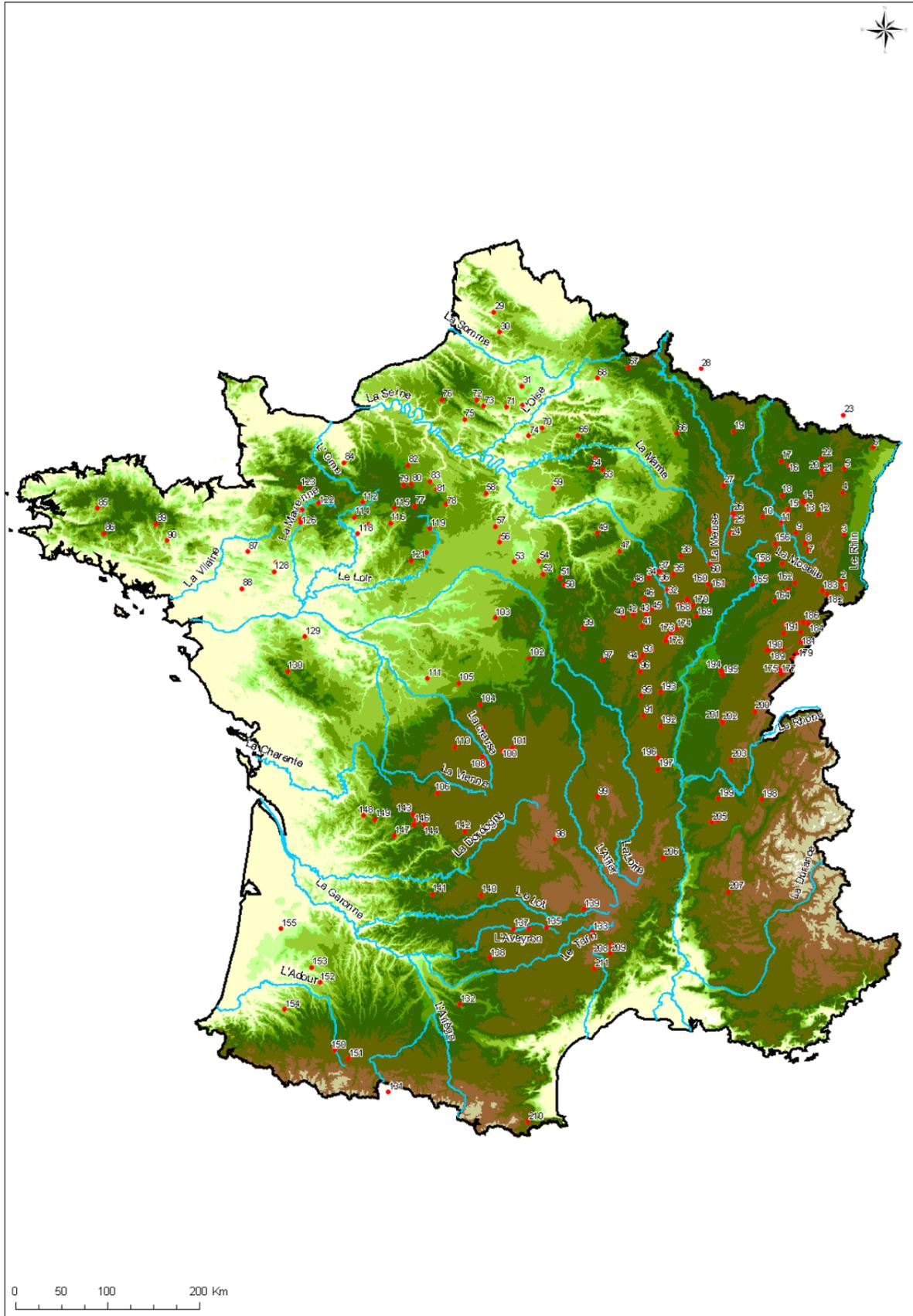


Fig. 6: BSS



5 Conclusions

- PEARP-based ensemble streamflow forecasts predicted well discharges over the studied catchments
- Better scores are obtained from the GR3P model with updating, while SIM results are closer to the results from GR3P model without updating (for data assimilation in SIM model, see Thirel et al., 2009)



Numéro	CODE HYDRO	Surface (km2)	Cours d'eau	Numéro	CODE HYDRO	Surface (km2)	Cours d'eau
1	A1080330	656	L' ILL A DIDENHEIM	944	L4210710		LA CREUSE A GLENIC
2	A1310310	1040	L' ILL A ENSISHEIM	108	L4220710	1240	LA GRANDE CREUSE A FRESSELINES
3	A2280350	4600	L' ILL A STRASBOURG	109	L4411710	850	LA PETITE CREUSE A FRESSELINES
4	A2860110	688	LA BRUCHE A HOLTZHEIM (2)	110	L5101810	570	LA GARTEMPE A BESSINES-SUR-GAR
5	A3472010	688	LA ZORN A WALTENHEIM-SUR-ZORN	111	L6202030	897	LA CLAUSE AU GRAND-PRESSIGNY (2)
6	A3792010	541	LA SAUER A BEINHEIM 1020.462	112	M0050620	908	LA SARTHE A ST-CENERI-LE-GEREI
7	A4200630	621	LA MOSELLE A ST-NABORD	113	M0243010	510	L' ORNE SAOSNOISE A MONTBIZOT
8	A4250640	1220	LA MOSELLE A EPINAL	114	M0250610	2720	LA SARTHE A NEUVILLE-SUR-SARTH
9	A5110610	1990	LA MOSELLE A TONNOY	115	M0361510	827	L' HUISNE A NOGENT-LE-ROUJOU
10	A5431010	940	LE MADON A PULLIGNY	116	M0421510	1890	L' HUISNE A MONTFORT-LE-GEISNOI
11	A5730610	3350	LA MOSELLE A TOUL	117	M0500610	5290	LA SARTHE A SPAY
12	A6221010	960	LA MEURTHE A AZERAILLES	118	M0680610	7380	LA SARTHE A ST-DENIS-D'ANJOU
13	A6921010	2780	LA MEURTHE A LANEUVILLE-DEVA	119	M1041610	1160	LE LOIR A ST-MAUR-SUR-LE-LOIR
14	A6941020	2960	LA MEURTHE A MALZEVILLE (2)	120	M1341610	5940	LE LOIR A FLEE (PORT-GAUTIER)
15	A7010610	6830	LA MOSELLE A CUSTINES	121	M1531610	7920	LE LOIR A DURTAL
16	A7821010	925	LA SEILLE A NOMENY	122	M3060910	828	LA MAYENNE A AMBRIERES-LES-VAL
17	A7881010	1280	LA SEILLE A METZ	123	M3133010	517	LA VARENNE A ST-FRAIMBAULT
18	A7930610	9390	LA MOSELLE A HAUCONCOURT	124	M3340910	2890	LA MAYENNE A L'HUISSERIE
19	A8431010	1230	L' ORNE A ROSSELANGE	125	M3600910	3910	LA MAYENNE A CHATEAU-GONTIER
20	A9091050	878	LA SARRE A KESKASTEL	126	M3630910	4160	LA MAYENNE A CHAMBELLAY
21	A9091060	721	LA SARRE A DIEDENDORF	127	M3771810	734	L' OUDON A CHATELAIS (MARCILLE
22	A9221010	1760	LA SARRE A SARREINSMING	128	M3851810	1310	L' OUDON A SEGRE (ECLUSE)
23	A9425050	1930	LA BLIES A BLIESBRUCK	129	M5222010	920	LE LAYON A ST-LAMBERT-DU-LATTA
24	B1150010	1030	LA MEUSE A DOMREMY LA PUCELLE	130	M7112410	814	LA SEVRE NANTAISE A TIFFAUGES
25	B2130010	2290	LA MEUSE A COMMERCY	131	O0010040	640	LA GARONNE A ST-BEAT
26	B2220010	2540	LA MEUSE A ST-MIHIEL	132	O2344010	526	LE GIROU A CEPET
27	B3150020	3900	LA MEUSE A STENAY (2)	133	O3141010	925	LE TARN A MOSTUEJOULS
28	B6111010	1340	LA SEMOY A HAILME	134	O3401010	2170	LE TARN A MILLAU (2)
29	E5400310	894	LA CANCHE A BRIMEUX	135	O5092520	556	L' AVEYRON A ONET-LE-CHATEAU
30	E5505720	784	L' AUTHIE A DOMPIERRE-SUR-AUTH	136	O5192520	1040	L' AVEYRON A VILLEFRANCHE-DE-R
31	E6406010	642	L' AVRE A MOREUIL	137	O5292510	1580	L' AVEYRON A LAGUEPIE (1)
32	H0100010	371	LA SEINE A NOD-SUR-SEINE	138	O5664010	503	LE CEROU A MILHARS
33	H0100020	704	LA SEINE A PLAINES-ST-LANGE	139	O7101510	1160	LE LOT A BANASSAC
34	H0203030	674	LA LAIGNES AUX RICEYS	140	O8133520	1190	LE CELE A ORNIAC
35	H0321030	548	L' OURCE A AUTRICOURT	141	P2484010	603	LE CEUO A ST-CYBRANET
36	H0400010	2340	LA SEINE A BAR-SUR-SEINE	142	P3922510	947	LA CORREZE A BRIVE-LA-GAILLARD
37	H0400020	2380	LA SEINE A COURTENOT	143	P6161510	800	L' ISLE A MAYAC
38	H1201010	1280	L' AUBE A BAR-SUR-AUBE	144	P6382510	884	L' AUIVEZERE AU CHANGE (AUBARED
39	H2062010	264	LE BEUVRON A OUAGE (CHAMPMORE)	145	P7001510	1860	L' ISLE A BASSILAC (CHARRIERAS
40	H2172320	366	LE COUSIN A AVALLON	146	P7041510	2120	L'ISLE A PERIGUEUX
41	H2322010	268	LE SEREIN A BIERRE-LES-SEMUR	147	F7261510	3750	L' ISLE A ABZAC
42	H2332020	643	LE SEREIN A DISSANGIS	148	P8284010	640	LA LIZONNE A ST-SEVERIN
43	H2412010	476	L' ARMANCON A QUINCY-LE-VICOMT	149	P8312520	1940	LA DRONNE A BONNES
44	H2442340	732	LA BRENNÉ A MONTBARD	150	Q0280030	906	L' ADOUR A ESTRICAC
45	H2452020	1350	L' ARMANCON A AISY-SUR-ARMONCON	151	Q0522520	173	L' ARROS A GOURGUE
46	H2462020	1970	L' ARMANCON A TRONCHOU	152	Q2192510	800	LE MIDOU A MONT-DE-MARSAN
47	H2473010	480	L' ARMANCON A CHESSY-LES-PRES	153	Q2593310	2500	LA MIDOUZE A CAMPAGNE
48	H2482010	2990	L' ARMANCON A BRIENON-SUR-ARMA	154	Q3464010	1150	LE LUY A ST-PANDELON
49	H2622010	866	LA VANNE A PONT-SUR-VANNE	155	S2242510	1650	L' EYRE A SALLES
50	H3122010	562	L' OUANNE A CHARNY	156	U0124010	317	LE CONEY A FONTENOY-LE-CHATEAU
51	H3122020	883	L' OUANNE A GY-LES-NOAINES	157	U0474010	1020	LA LANterne A FLEUREY-LES-FAVE
52	H3201010	2300	LE LOING A CHALETTE-SUR-LOING	158	U0610010	3740	LA SAONE A RAY-SUR-SAONE
53	H3322010	339	LA BEZONDE A PANNES	159	U0724010	390	LE SALON A DENEVRE
54	H3621010	3900	LE LOING A PISY	160	U0924010	417	LA VINGEANNE A ST-MAURICE-SUR-
55	H4022020	875	L' ESSONNE A GUIGNEVILLE-SUR-E	161	U0924020	623	LA VINGEANNE A OISILLY
56	H4022030	609	L' ESSONNE A BOULANCOURT	162	U1044010	866	L' OGNON A CHASSEY-LES-MONTBOZ
57	H4042010	1870	L' ESSONNE A BALLANCOURT-SUR-E	163	U1054010	1250	L' OGNON A BEAUMOTTE AUBERTANS
58	H4252010	922	L' ORGE A MORSANG-SUR-ORGE	164	U1084010	2040	L' OGNON A PESMES
59	H4322030	429	L' YERRES A COURTOMER (PARADIS	165	U1120010	8900	LA SAONE A AUXONNE
60	H5011020	360	LA MARNE A MARNAY-SUR-MARNE	166	U1215030	300	L' IGNON A VILLECOMTE
61	H5062010	614	LE ROGNON A DOULAIN-SAUCOURT	167	U1224010	700	LA TILLE A ARCEAU (ARCELOT)
62	H5302010	454	LE SURMELIN A ST-EUGENE	168	U1224020	744	LA TILLE A CESSEY-SUR-TILLE
63	H5412010	354	LE PETIT MORIN A MONTMIRAIL	169	U1235020	264	LA NORGES A GENLIS
64	H5412020	605	LE PETIT MORIN A JOUARRE	170	U1244040	1100	LA TILLE A CHAMPDOTE
65	H5522010	345	L' OURCQ A CHOUY	171	U1314010	304	L' OUCHE A LA BUSSIÈRE-SUR-OUC
66	H6221010	2940	L' AISNE A GIVRY	172	U1314020	442	L' OUCHE A STE-MARIE-SUR-OUCHE
67	H7033010	258	LE THON A ORIGNY-EN-THERACHE	173	U1324010	655	L' OUCHE A PLOMBIERES-LES-DIJO
68	H7401010	4290	L' OISE A SEMPIGNY	174	U1334010	887	L' OUCHE A TROUHANS
69	H7423710	284	L' ARONDE A CLAIROIX	175	U2022010	350	LE DOUBS A LA CLUSE-ET-MIOUX
70	H7513010	279	L' AUTOMNE A SAINTINES	176	U2022020	396	LE DOUBS A DOUBS
71	H7602010	468	LA BRECHE A NOGENT-SUR-OISE	177	U2102010	658	LE DOUBS A VILLE-DU-PONT
72	H7742010	747	LE THERAIN A BEAUVAIS	178	U2122010	1060	LE DOUBS A GOUMOIS
73	H7742020	1200	LE THERAIN A MAYSEL	179	U2142010	1240	LE DOUBS A GLERE (COURCLAVON)
74	H7813010	338	LA NONETTE A COURTEUIL	180	U2215020	560	LE DESSOUBRE A ST-HIPPOLYTE
75	H8042010	1370	L' EPTÉ A FOURGES	181	U2222010	2200	LE DOUBS A MATHAY
76	H8212010	371	L' ANDELLE A VASCOEUIL	182	U2334010	725	L' ALLAN A FESCHES-LE-CHATEL
77	H9021010	330	L' EURE A ST-LUPERCE	183	U2354010	1120	L' ALLAN A COURCELLES-LES-MONT
78	H9121010	2050	L' EURE A CHARPONT	184	U2402010	3420	LE DOUBS A VOUEAUCOURT
79	H9202010	490	L' AVRE A ACON	185	U2425260	358	LE CUSANCIEN A BAUME-LES-DAMES
80	H9222010	880	L' AVRE A MUZY	186	U2512010	4400	LE DOUBS A BESANCON
81	H9331010	4530	L' EURE A CAILLY-SUR-EURE	187	U2542010	4970	LE DOUBS A ROCHEFORT-SUR-NENON
82	H9402030	1050	L' ITON A NORMANVILLE	188	U2604030	326	LA LOUE A VULLAFANS
83	H9501010	5990	L' EURE A LOUVIERS	189	U2624010	1160	LA LOUE A CHENECEY-BUILLON
84	I2051040	638	LA DIVES AU MESSNIL-MAUGER	190	U2634010	1380	LA LOUE A CHAMPAGNE-SUR-LOUE
85	J3811810	1220	L' AULNE A CHATEAUNEUF-DU-FAOU	191	U2722010	7290	LE DOUBS A NEUBLANS-ABERGEMENT
86	J4742010	578	L' ELLE A ARZANO (PONT	192	U3214010	332	LA GROSNE A JALOGNY (CLUNY)
87	J7483010	820	LA SEICHE A BRUZ (CARCE)	193	U3225010	273	LA GUYE A SIGY-LE-CHATEL
88	J7963010	598	LE DON A GUEMENE-PENFAO (UZET	194	U3415030	508	LA BRENNÉ A SENS-SUR-SEILLE
89	J8202310	929	L' OUST A PLEUGRIFFET (LA	195	U3424010	790	LA SEILLE A ST-USUGE
90	J8502310	2470	L' OUST A ST-GRAVE (LE	196	U4624010	336	L' AZERGUES A CHATILLON
91	K1173210	599	L' ARCONCE A MONTCEAUX-L'ETOIL	197	U4644010	792	L' AZERGUES A LOZANNE
92	K1251810	776	L' ARROUX A DRACY-ST-LOUP	198	V1315020	280	LA LEYSSE A LA MOTTE-SERVOLEX
93	K1321810	1800	L' ARROUX A ETANG-SUR-ARROUX	199	V1774010	703	LA BOURBRE A TIGNIEU-JAMEYZIEU
94	K1341810	2280	L' ARROUX A RIGNY-SUR-ARROUX	200	V2444020	650	LA BIENNE A JEURRE
95	K1383010	819	LA BOURBINCE A VITRY-EN-CHAROL	201	V2814020	324	LE SURAN A NEUVILLE-SUR-AIN (L
96	K1391810	3170	L' ARROUX A DIGOUIN	202	V2814030	349	LE SURAN A PONT-D'AIN
97	K1773010	1470	L' ARON A VERNEUIL	203	V2934010	288	L' ALBARINE A ST-DENIS-EN-BUGE
98	K2593010	984	L' ALAGNON A LEMPEDES	204	V3424310	461	LE RIVAL A BEAUFORT
99	K2981910	1520	LA DORE A DORAT	205	V3434010	650	LES COLLIERES A ST-RAMBERT-D'A
100	K5183010	854	LA TARDES A EVAUX-LES-BAINS	206	V3724010	378	LE DOUX A COLOMBIER-LE-VIEUX
101	K5220910	1840	LE CHER A ST-VICTOR	207	V4264010	1150	LA DROME A SAILLANS
102	K5552300	531	L' YEVRE A SAVIGNY-EN-SEPTAINE	208	V7124010	240	LE GARDON DE MALET A GENERARGUES
103	K6492510	2250	LA SAULDRE A SELLES-SUR-CHER	209	V7135010	263	LE GARDON DE ST-JEAN A
104	K7202610	697	L' INDRE A ARDENTES	210	Y0284060	729	LE TECH A ARGELES-SUR-MER
105	K7312610	1710	L' INDRE A ST-CYRAN-DU-JAMBOT	211	Y2102010	760	L' HERAULT A LAROCQUE
106	L0563010	597	LA BRIANCE A CONDAT-SUR-VIENNE				

Bibliographie

- Bartholmes, J.C., Thielen, J., Ramos, M.H., S. Gentilini (2009), "The European flood alert system EFAS - Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts", *Hydrology and Earth System Sciences*, 13, 141-153.
- Bontron, G. (2004), "*Prévision quantitative des précipitations: Adaptation probabiliste par recherche d'analogues*". Thèse de doctorat, Institut national polytechnique de Grenoble, Grenoble, 150p.
- Brier, G. W. (1950), "Verification of forecasts expressed in term of probability", *Monthly Weather Review*, 78, 1-3.
- Candille, G. (2003), "*Validation des systèmes de prévisions météorologiques probabilistes*". Thèse de Doctorat Université Pierre et Marie Curie, 146 p.
- Candille, G., O. Talagrand (2005), "Evaluation of probabilistic prediction systems", *Quart. J. Roy. Meteor. Soc.*, 131, 2131–2150.
- Daoud, A. B., and et J. Plasse (2008), "Application d'une méthode de prévisions des pluies par analogie au bassin de la Seine à Paris", Rapport Cemagref, 80 p.
- Habets, F., Ducrocq, V. and Noilhan, J. (2004), "Prévisions hydrologiques et échelles spatiales : l'exemple des modèles opérationnels de Météo-France", *Géophysique externe, climat et environnement. C. R. Geoscience*, 337, 181–192.
- Hamill, M. T. (2009), "National Center for Atmospheric Research", *Boulder, Colorado*, 129, 11.
- Houdant, B. (2006), "*Contribution à l'amélioration de la prévision hydrométéorologique opérationnelle pour l'usage de probabilités dans la communication entre acteurs*", Thèse de Doctorat Université de Toulouse, Paris, 209 p.
- Houtekamer, P. (1998), "Rudiments de la prévision d'ensemble", in *Atelier de prévision météorologique numérique*, Canada, Centre météorologique canadien, 26p.
- Javelle, P., Berthet, L., Arnaud, P., Lavabre, J. and Perrin, C., (2008), "Comparaison de deux versions du modèle GR pour la prévision des crues sur un grand échantillon de bassins versants français", *Colloque SHF-191e CST-"Prévisions hydrométéorologiques"*, 18-19 Novembre, Lyon, 105-109.
- Jolliffe, I. T., D. B. Stephenson (2004), "Forecast verification. A Practitioner's Guide in Atmospheric Science", John Wiley, Chichester England, 240p.
- Mallet, V. (2008), "Prévision d'ensemble", INRIA, Université Paris Est.
- Martin E., Thirel G., Regimbeau, F. Noilhan J., Franchistéguy, L. Soubeyroux, J-M. (2008), "Vers une prévision d'ensemble des débits à l'échelle des grands bassins français". *Colloque SHF-191e CST-"Prévisions hydrométéorologiques"*, 18-19 Novembre, Lyon, 201-208.
- Murphy, A.H., (1993), "What is a good forecast? An essay on the nature of goodness in weather forecasting". *Weather Forecasting*, 8, 281-293.
- Palmer, T. N., J. Barkmeijer, R. Buizza, E. Klinker, D. Richardson (2002), "L'avenir de la prévision d'ensemble", *La Météorologie*, 36, 22-30.
- Pappenberger, F., Scipal, K., Buizza, R., (2008), "Hydrological aspects of meteorological verification", *Atmospheric Science Letters*, 9 (2), 43-52.
- Paradis, E. (2005), "R pour les débutants", Université de Montpellier II, Montpellier, 81p.

- Perrin, C. (2000), "*Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative*". Thèse de doctorat, INPG, Grenoble, Paris, 530 p.
- Perrin, C., Lerat, J., Oudin, L., Andréassian, V. (2005), " Inventaire et diagnostic des modèles simples de prévision des crues déjà existants sur le bassin de la Seine", Rapport Unité HBAN, RIE / TRANSFEAU, Cemagref Antony, 133 p.
- 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". *J. of Hydrology*, 242, 275-301.
- Perrin, C., Michel, C., Andréassian, V. (2003), "Improvement of a parsimonious model for streamflow simulation". *J. of Hydrology*, 279 (1-4), 275-289.
- Petit, T. (2008), "*Evaluation de la performance de prévisions hydrologiques d'ensemble issues de prévisions météorologiques d'ensemble*", Master report, Université de Laval, Québec, 81p.
- Ramos, M.-H., Thielen, J., Pappenberger, F. (2008), "Utilisation de la prévision météorologique d'ensemble pour la prévision opérationnelle et l'alerte aux crues", *Colloque SHF-191e CST- "Prévisions hydrométéorologiques"*, 18-19 Novembre, Lyon, 191-199.
- Rousset Regimbeau, F., Habets, F., Martin, E., Noilhan, J. (2007), Ensemble streamflow forecasts over France. *ECMWF Newsletter* 111, 21-27.
- Rousset-Regimbeau, F. (2007), "*Modélisation des bilans de surface et des débits sur la France, application à la prévision d'ensemble des débits*", Thèse de Doctorat, Université Toulouse III, Paul Sabatier, Toulouse, 216 p.
- 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., Rousset-Regimbeau, F., Martin, E., Habets, F. (2008), "On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Prediction". *Journal of hydrometeorology*, 9, 1301-1317.
- Vincendon, J.-C., J.-M. Carrière (2008), "Les prévisions opérationnelles des précipitations à Météo-France", *Colloque SHF-191e CST- "Prévisions hydrométéorologiques"*, 18-19 Novembre Lyon, 5-11.

Evaluation de la qualité des prévisions pour l'alerte aux crues

Rindra Annie RANDRIANASOLO

Mémoire de master soutenue le 30 Juin 2009 à Paris.

Encadrement : Maria-Helena Ramos

Résumé

Dans les centres opérationnels de prévision, les prévisionnistes doivent traiter des données issues de différents modèles et les combiner entre elles, afin de pouvoir orienter leurs décisions et communiquer les alertes pour les usagers. La qualité des prévisions devient alors primordiale.

Une analyse comparative a été réalisée entre deux sorties de débits issus de deux modèles hydrologiques structurellement distincts. Les entrées sont les mêmes : les prévisions d'ensemble de précipitations PEARP (de Météo-France) sur 211 bassins, pendant 17 mois (Mars 2005-Juillet 2006). Le modèle météorologique tourne avec 11 membres, et permet de prévoir sur deux échéances (jour $j+1$ et jour $j+2$). Les deux modèles hydrologiques sont : (i) le modèle GRPE, modèle global du Cemagref (GR3P), adapté à la prévision d'ensemble, et (ii) le modèle distribué SAFRAN-ISBA-MODCOU (SIM) développé à Météo-France. L'étude consistait donc à produire un outil de vérification permettant d'évaluer la qualité des prévisions d'ensemble issues de ces deux modèles.

Il est apparu alors à l'issue de ce travail, que GRPE a de meilleurs scores, i.e., il prévoit mieux que le modèle distribué de Météo-France. Sa performance est liée de près à une de ses caractéristiques : la mise à jour, c'est-à-dire l'assimilation dans le modèle de la dernière observation de débit. Malgré ces bons scores, il a été noté que la dispersion des membres du modèle GR est faible. Des voies d'amélioration restent encore ouvertes.

Mots-clés: prévision d'ensemble, vérification de la prévision, mis à jour, membre, modèle hydrologique