



**HAL**  
open science

# Evaluation de la qualité des prévisions hydrologiques d'ensemble construites à partir des prévisions météorologiques d'ensemble de Météo-France (PEARP)

Annie Randrianasolo, Maria-Helena Ramos

► **To cite this version:**

Annie Randrianasolo, Maria-Helena Ramos. Evaluation de la qualité des prévisions hydrologiques d'ensemble construites à partir des prévisions météorologiques d'ensemble de Météo-France (PEARP). [Rapport de recherche] irstea. 2011, pp.31. hal-02595518


**HAL Id: hal-02595518**

**<https://hal.inrae.fr/hal-02595518v1>**

Submitted on 15 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# **Evaluation de la qualité des prévisions hydrologiques d'ensemble construites à partir des prévisions météorologiques d'ensemble de Météo-France (PEARP)**

*Rapport de Synthèse*

*Convention Cemagref-SCHAPI*

*Août 2011*

**Annie Randrianasolo**

**Maria-Helena Ramos**

Equipe Hydrologie

Unité de Recherche HBAN - Antony

Contact : [maria-helena.ramos@cemagref.fr](mailto:maria-helena.ramos@cemagref.fr)

1, rue Pierre-Gilles de Gennes  
92761 Antony Cedex FRANCE



**Cemagref**  
Sciences, eaux & territoires

# ***Evaluation de la qualité des prévisions hydrologiques d'ensemble construites à partir des prévisions météorologiques d'ensemble de Météo-France (PEARP)***

## ***Rapport de Synthèse***

### **Introduction**

Ce rapport s'inscrit dans l'action "Prévision des inondations", menée dans le cadre de la convention SCHAPI-Cemagref. Il présente les travaux menés au Cemagref sous le volet 2 de la convention 2010, "Développement du système d'aide à la vigilance hydrologique (AVI)" et correspond au livrable de l'action A2 portant sur l'évaluation de la qualité des prévisions hydrologiques construites à partir de l'archive étendue PEARP fournie par Météo-France.

Il s'agit de contribuer aux développements concernant la mise en place d'un outil d'aide à la vigilance pour le SCHAPI. Le but général est d'évaluer l'apport d'une prévision météorologique à scénarios multiples (prévision d'ensemble) à la prévision hydrologique de débits, pour mieux anticiper les événements de crues. Cette étude s'inscrit dans la continuité des travaux initiés dans le cadre du protocole d'expérimentation sur les prévisions d'ensemble de débits à partir des prévisions météorologiques d'ensemble ARPEGE (PEARP), mis en place en 2009 dans le cadre d'une coopération scientifique entre le SCHAPI, le Cemagref et Météo-France.

Les premiers résultats des études menées au Cemagref ont été présentés dans Randrianasolo (2009) et Randrianasolo *et al.* (2010). L'évaluation des sorties en débits du modèle hydrologique GRP (Cemagref) et de la chaîne de modélisation SIM (Météo-France ; débits générés dans le cadre de la thèse de G. Thirel, 2009), alimentés par une même entrée de précipitation prévue, la PEARP, a permis d'étudier la performance (qualité) des prévisions hydrologiques d'ensemble sur un large échantillon de bassins versants en France, ainsi que de mieux comprendre les approches et critères utilisés couramment pour évaluer les prévisions probabilistes. Suite à ces travaux, il a paru intéressant de poursuivre les démarches entreprises et explorer également une archive étendue de la PEARP, contenant des nouvelles années de précipitations prévues par le système, afin de consolider les statistiques d'évaluation des prévisions.

Ce rapport présente une synthèse des études concernant l'évaluation de la qualité des prévisions hydrologiques d'ensemble construites à partir des prévisions PEARP (archive initiale et nouvelle archive étendue), et fournit une vision intégrée des travaux réalisés en utilisant les prévisions météorologiques de la PEARP en hydrologie.

Il est organisé en deux parties. La première partie s'intéresse aux aspects méthodologiques (modèles et critères d'évaluation) et aux données utilisées (prévisions, observations et bassins versants). Dans la deuxième partie, les différents résultats obtenus sont présentés et discutés. En fin de rapport, des conclusions sont émises.

# Table de Matières

<b>INTRODUCTION .....</b>	<b>2</b>
<b>PARTIE I : ASPECTS METHODOLOGIQUES ET DONNEES.....</b>	<b>4</b>
I-1. METHODOLOGIE GENERALE .....	4
I-2. MODELES UTILISES .....	4
I-2.1 <i>Modèle de prévision hydrologique GRP du Cemagref</i> .....	4
I-2.2 <i>Chaîne de modélisation SIM de Météo-France</i> .....	6
I-3. DONNEES UTILISEES ET BASSINS VERSANTS.....	7
I-3.1 <i>Prévisions d'ensemble de pluie ARPEGE : PEARP</i> .....	7
I-3.2 <i>Bassins versants étudiés, débits prévus et observations</i> .....	9
I-4. SCORES D'EVALUATION DES PREVISIONS .....	10
I-4.1 <i>RMSE et ratio-RMSE</i> .....	10
I-4.2 <i>Tableau de contingence et CSI</i> .....	11
I-4.3 <i>Score de Brier (BS) et score de performance BSS</i> .....	12
I-4.4 <i>Score de probabilité ordonné (RPS) et RPSS</i> .....	13
I-4.5 <i>Référence pour les scores de performance et gain</i> .....	14
I-5. VISUALISATION DES RESULTATS .....	14
<b>PARTIE 2 : RESULTATS .....</b>	<b>15</b>
II-1. EVALUATION DES PREVISIONS HYDROLOGIQUES SIM ET GRP POUR LA PERIODE DE MARS 2005 A JUILLET 2006 .....	15
II-1.1 <i>Performances et impact de la mise à jour sur les prévisions</i> .....	15
II-1.2 <i>Evaluation des prévisions selon les surfaces des bassins</i> .....	18
II-1.3 <i>Evaluation des prévisions selon les durées caractéristiques de crue</i> .....	18
II-1.4 <i>Evaluation des prévisions selon les régimes hydrologiques</i> .....	21
II-2. EVALUATION DES PREVISIONS HYDROLOGIQUES GRP POUR LA PERIODE DE MARS 2005 A JUILLET 2009 (ARCHIVE ETENDUE) .....	23
II-2.1 <i>Caractéristiques descriptives des deux périodes d'évaluation</i> .....	23
II-2.2 <i>Performance comparée</i> .....	24
<b>CONCLUSIONS .....</b>	<b>27</b>
<b>REFERENCES BIBLIOGRAPHIQUES.....</b>	<b>29</b>

# Partie I : Aspects méthodologiques et données

## I-1. Méthodologie générale

Les démarches méthodologiques se sont structurées autour de :

- la mise en forme des séries de débits prévus, obtenues à partir des deux modèles hydrologiques étudiés : (1) le modèle global GRP fonctionnant à l'échelle du bassin versant et (2) la chaîne de modélisation distribuée SIM de Météo-France. Ces modèles ont différentes structures et approches de mise à jour des prévisions journalières des débits (§I-2), et ont été alimentés par une même entrée de précipitation prévue, la prévision météorologique d'ensemble PEARP de Météo-France (§I-3.1) ;
- le choix des critères d'évaluation de la qualité des prévisions hydrologiques (ou *scores de vérification*) (§I-4) ;
- l'application des critères d'évaluation à une longue série de couplets prévision-observation : d'abord, sur une période de 17 mois, entre mars 2005 et juillet 2006 (§II-1.1), et ensuite sur une archive étendue de 53 mois, allant jusqu'à juillet 2009 (§II-2) ;
- l'utilisation d'un large échantillon de bassins versants, répartis sur l'ensemble du territoire français et représentatifs de différentes surfaces, durées caractéristiques de crue et régimes hydrologiques (§I-3.2) ;
- l'étude de l'impact de la mise à jour employée dans la modélisation des débits sur la qualité des prévisions émises (§II-1.1), ainsi que de l'évolution des critères selon différentes caractéristiques des bassins versants (§II-1.2, §II-1.3 et §II-1.4) et selon la longueur des séries de prévisions disponibles pour l'évaluation (§II-2).

## I-2. Modèles utilisés

Les prévisions de débits évaluées dans cette étude proviennent de deux modèles hydrologiques : le modèle global GRP du Cemagref et la chaîne de modélisation distribuée SIM développée à Météo-France.

### I-2.1 Modèle de prévision hydrologique GRP du Cemagref

Le modèle GRP<sup>1</sup> a été spécifiquement adapté pour la prévision de crues par Tangara (2005) et dérive du modèle de simulation pluie-débit GR4J, également développé par le Cemagref (Perrin *et al.*, 2001, 2003)<sup>2</sup>. Les développements récents autour du modèle GRP ont notamment concerné : le fonctionnement au pas de temps horaire, avec l'évaluation de différentes stratégies de mise à jour (Berthet, 2010), l'adaptation du modèle journalier pour la prévision hydrologique d'ensemble (prise en compte de plusieurs scénarios de prévisions météorologiques en entrée) et son évaluation (Ramos *et al.*, 2008 ; Randrianasolo, 2009), ainsi que la mise en place d'une routine de simulation du stockage et de la fonte de la neige (Valéry, 2010) et son évaluation (Nicolle, 2010 ; Nicolle *et al.*, 2011).

---

<sup>1</sup> Ou "GR3P" pour indication de ses trois paramètres de calage.

<sup>2</sup> Pour plus d'information sur les modèles GR développés au Cemagref : <http://www.Cemagref.fr/webgr/>

Le modèle GRP ici utilisé est présenté dans la Figure 1. Il fonctionne au pas de temps journalier et a été préalablement calé pour chaque bassin versant étudié et pour ses trois paramètres : X1, la capacité à un jour de son réservoir de stockage, X2, le coefficient de correction, et X3, la durée caractéristique de son hydrogramme unitaire. Le modèle fonctionne en continu. Les calculs doivent ainsi intégrer une période suffisamment longue, avant toute période d'évaluation, pour pouvoir bien simuler l'état initial du bassin versant comme la résultante des conditions climatiques antérieures. La précipitation et l'évapotranspiration potentiel (ETP) de bassin constituent les données d'entrée du modèle.

Quant au fonctionnement général du modèle, un premier réservoir de "production" gère, en fonction de son taux de remplissage, la part de pluie totale qui se transforme en pluie nette. Un second réservoir (de routage), associé à un hydrogramme unitaire, permet d'étaler le temps de cette pluie nette pour obtenir le débit de sortie. La version du modèle ici utilisé ne prend pas en compte le stockage et la fonte de la neige (les précipitations en entrée sont considérées comme étant des précipitations liquides). De plus, le modèle ne simule pas les barrages, digues ou autres ouvrages présents à l'intérieur du bassin versant.

La première application du modèle GRP en prévision hydrologique d'ensemble s'est appuyée sur les scénarios de la prévision météorologique d'ensemble produite par le CEPMMT<sup>3</sup> et a concerné un événement de crue dans le bassin de la Seine (Ramos *et al.*, 2008). L'évaluation statistique des séries temporelles des prévisions hydrologiques basées sur les prévisions des précipitations du CEPMMT (EPS) et de Météo-France (PEARP) ont ensuite fait l'objet de plusieurs recherches menées dans le but d'examiner différents aspects de la prévision hydrologique d'ensemble (Randrianasolo *et al.*, 2010, 2011 ; Velázquez *et al.*, 2010 ; Nicolle *et al.*, 2011 ; Werner, 2010 ; Van Pham, 2011).

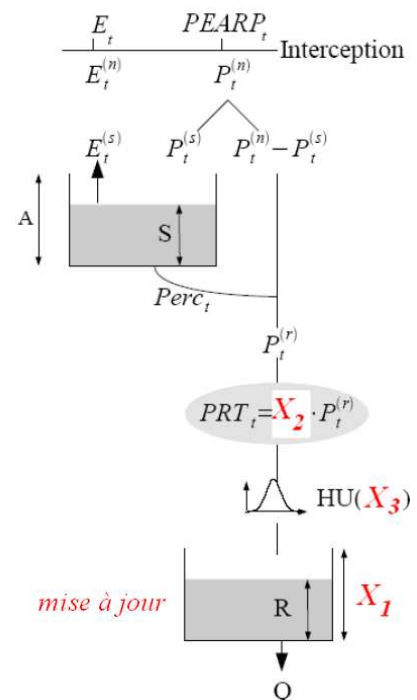


Figure 1 : Schéma du modèle hydrologique de prévision GRP

De manière générale, une procédure de mise à jour intervient souvent dans le fonctionnement d'un modèle de prévision hydrologique dans le but de caractériser au mieux l'état du système (le bassin versant) avant de démarrer une prévision. Cette procédure est basée sur l'assimilation des observations passées et présentes. Elle permet de modifier une ou plusieurs composantes du système (paramètres, états internes, sorties) pour obtenir un meilleur accord des simulations aux observations.

Dans le modèle GRP ici utilisé, l'assimilation des données se fait à l'aide de l'observation de débit du jour qui précède la prévision et intervient au niveau du réservoir de routage : le niveau du réservoir est calculé directement à partir du dernier débit observé à l'instant de la prévision et modifié pour le mettre à jour.

<sup>3</sup> Centre européen de prévision météorologique à moyen terme (ECMWF, en anglais) : <http://www.ecmwf.int/>

## I-2.2 Chaîne de modélisation SIM de Météo-France

La chaîne de modélisation distribuée SIM (SAFRAN-ISBA-MODCOU) est une suite de trois modèles indépendants (Figure 2 ; 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 km<sup>2</sup>). 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 six heures et, d'autre part, à l'analyse des quantités quotidiennes de précipitation. La troisième étape est constituée de l'interpolation horaire de ces paramètres, ainsi que du 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 est basée sur trois couches de sol (surface, zone racinaire et 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 le ruissellement et le drainage d'ISBA en entrée pour calculer l'évolution des nappes (dans les bassins où celles-ci sont représentées) et les débits des rivières.

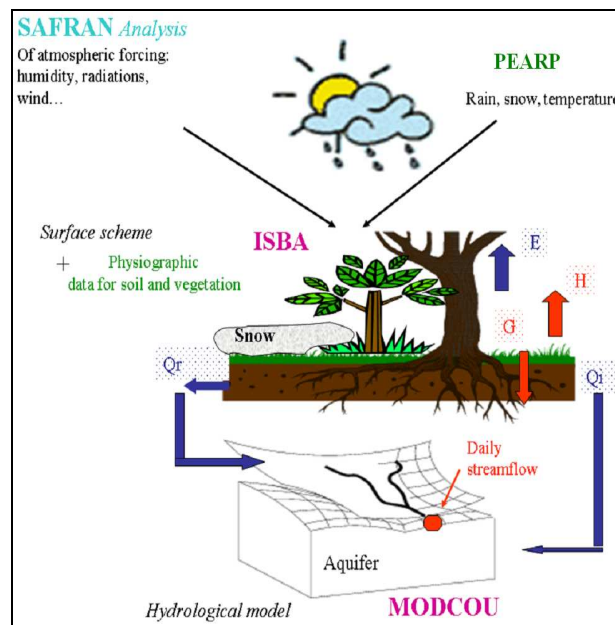


Figure 2 : Schéma de la chaîne de modélisation SIM  
(adapté du poster présenté par Rousset-Regimbeau *et al.* à la conférence EGU 2008:  
"Medium-range ensemble streamflow forecast over France")

La chaîne SIM a été validée sur une période de dix ans, par comparaison avec des hauteurs de nappe, hauteurs 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 (voir Martin *et al.*, 2008 et références citées dans l'article). Les principales limitations du modèle identifiées sont liées à des éléments qui ne sont pas pris en compte dans la modélisation, comme certaines nappes et les barrages.

Les premières applications de la chaîne de modélisation SIM en prévision hydrologique d'ensemble se sont focalisées sur la prévision à moyen terme, dans le but de pouvoir anticiper les crues dans les grands bassins ou fournir des informations sur l'évolution des débits les plus faibles, pour lesquels les réponses hydrographiques dépendent des niveaux des nappes. Cela a permis de mettre en place une chaîne de prévision à moyenne échéance (10 jours) basée sur les 51 membres (ou scénarios) de la prévision météorologique d'ensemble produite par le CEPMMT (Rousset Regimbeau, 2007 ; Rousset Regimbeau *et al.*, 2007). Cette chaîne tourne en temps réel en mode expérimental depuis 2004 et ses résultats sont mis à disposition du SCHAPI pour une sélection de stations hydrologiques<sup>4</sup>. Des études ont ensuite été menées en utilisant en entrée les prévisions d'ensemble météorologiques ARPEGE de Météo-France à courte échéance (PEARP ; 11 membres et 60 heures d'échéance) (Thirel, 2009 ; Thirel *et al.*, 2008).

Une procédure de mise à jour a été récemment développée et testée dans la chaîne de modélisation SIM, avec pour but d'améliorer les simulations de débits en ajustant l'humidité du sol. La mise à jour dans la chaîne de modélisation SIM se fait ainsi au niveau du modèle ISBA. La procédure est décrite en détail dans Thirel (2009) ; Thirel *et al.* (2010a, b). De manière générale, il s'agit de modifier les humidités des deux couches de sol (couche racinaire et couche profonde) pour rapprocher la trajectoire de la simulation le plus possible de l'ensemble des observations de débits disponibles.

Enfin, on notera que les débits prévus ici évalués, autant pour le modèle GRP que pour la chaîne SIM, ont été considérés avec et sans prise en compte de la procédure d'assimilation et mise à jour de chaque modèle. Il a été alors possible d'étudier l'impact de l'assimilation de données sur la qualité des prévisions de débits (voir §II-1.1).

### **I-3. Données utilisées et bassins versants**

#### **I-3.1 Prévisions d'ensemble de pluie ARPEGE : PEARP**

La prévision d'ensemble ARPEGE de Météo-France (PEARP) est basée sur la méthode des vecteurs singuliers pour la recherche des perturbations de l'état initial qui auront le plus d'impact sur la prévision à une échéance de 12 heures et sur une zone couvrant le proche Atlantique (pour plus de détail, voir Nicolau, 2002 et Thirel, 2009). Les prévisions PEARP ici utilisées sont constituées de dix perturbations de l'état initial, conduisant à dix prévisions de l'évolution de l'atmosphère, et une prévision de contrôle, obtenue avec un état initial non-perturbé, fourni par l'analyse initiale (résultat de l'assimilation des observations). A chaque instant de prévision, 11 membres de la PEARP sont ainsi disponibles pour une échéance maximale de prévision de 60 heures. Les membres de la prévision d'ensemble sont considérés équiprobables entre eux.

---

<sup>4</sup> Pour plus de détail, voir la présentation disponible sur : [http://www.cemagref.fr/webgr/Download/afq/04\\_AtelierFrance-Quebec\\_Regimbeau.pdf](http://www.cemagref.fr/webgr/Download/afq/04_AtelierFrance-Quebec_Regimbeau.pdf)



Les données PEARP mises à disposition du Cemagref couvrent l'ensemble de la France, avec un maillage de 8x8 km<sup>2</sup> (9892 mailles au total). Lancées une fois par jour (à 18h), les prévisions ont été fournies au pas de temps tri-horaire et sous la forme d'archives mises à disposition en deux temps : d'abord, une archive comprenant la période du 10/03/2005 au 30/09/2006, et, ensuite, une archive étendue allant jusqu'au 30/09/2009.

La mise en forme de ces prévisions a été effectuée pour pouvoir constituer les fichiers nécessaires pour le forçage du modèle hydrologique GR. Pour cela, les précipitations (somme des composantes "pluie" et "neige" de la PEARP) ont été agrégées au pas de temps journalier (cumul de 0h à 23h59). Au total, deux échéances de prévision ont été considérées : j+1 ("le lendemain" de la date d'émission de la prévision) et j+2 ("le surlendemain"). Ensuite, pour chaque bassin versant étudié, les précipitations prévues de bassin (moyennes surfaciques) ont été calculées en considérant le pourcentage de superficie de chaque maille se trouvant sur le bassin.

Les Figures 3 et 4 montrent deux exemples des 11 scénarios fournis par la prévision d'ensemble PEARP, à 2 jours d'échéance, pour le 16 avril 2005 et le 31 octobre 2005, respectivement.

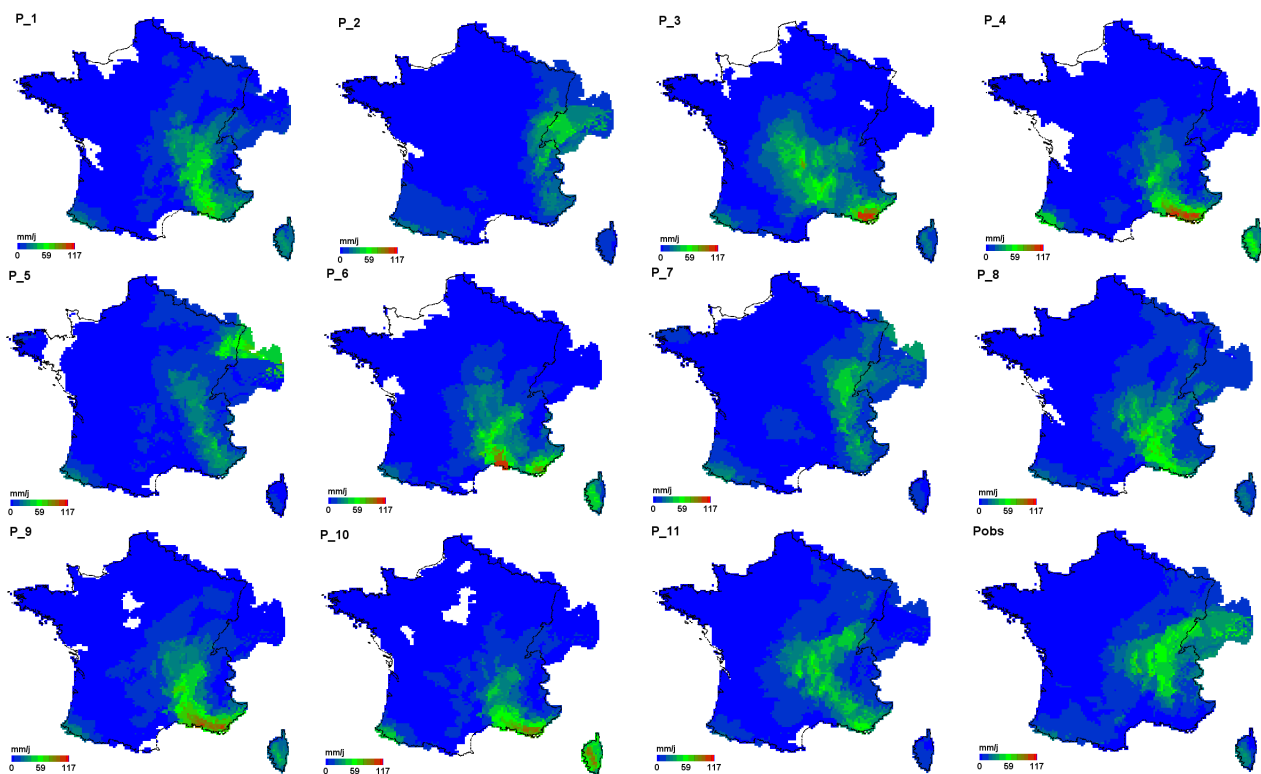


Figure 3 : Pluies journalières d'ensemble prévues par la PEARP sur la France pour le 16 avril 2005 à deux jours d'échéance (11 scénarios : P1 à P11). Les précipitations de la ré-analyse SAFRAN sont également présentées (Pobs)

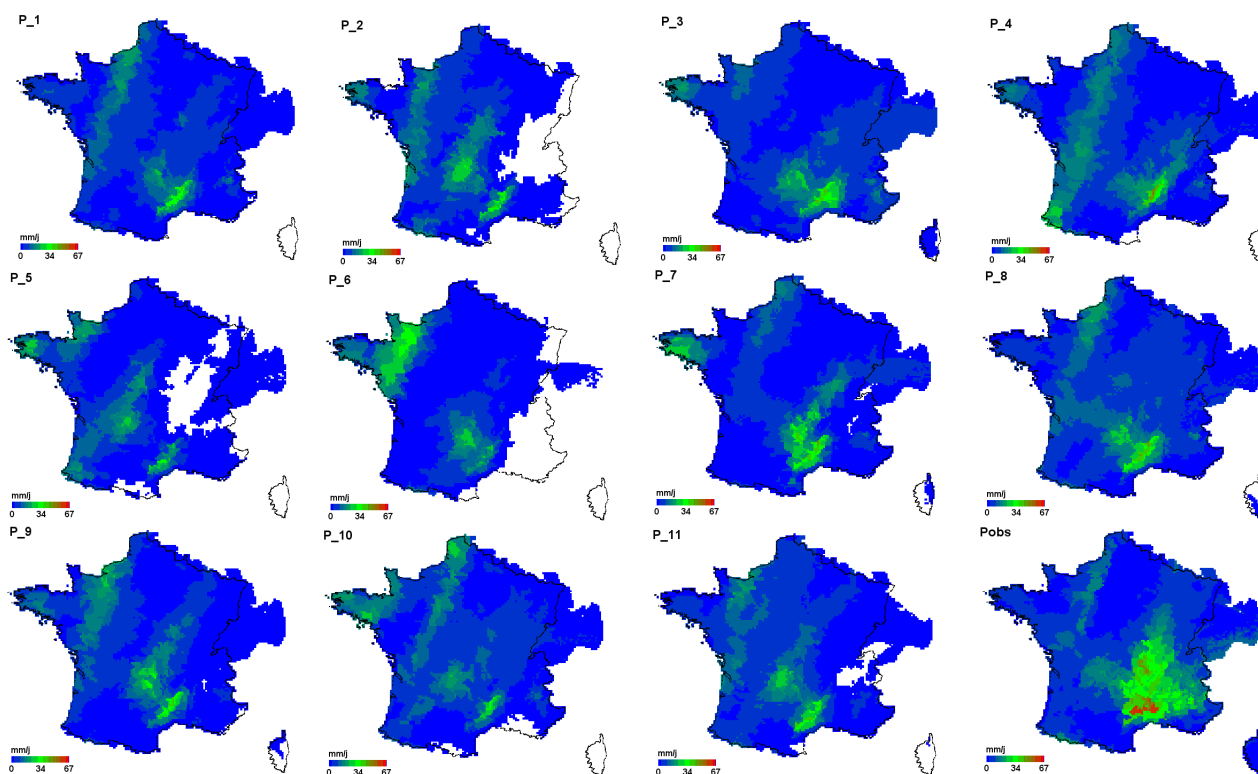


Figure 4 : Pluies journalières d'ensemble prévues par la PEARP sur la France pour le 31 octobre 2005 à deux jours d'échéance (11 scénarios : P1 à P11). Les précipitations de la ré-analyse SAFRAN sont également présentées (Pobs)

### I-3.2 Bassins versants étudiés, débits prévus et observations

Pour évaluer les prévisions, une référence de vérification est nécessaire. Dans ce travail, ce sont les observations de débits de la banque HYDRO qui ont été considérées. Ces données sont disponibles au Cemagref jusqu'au 31 juillet 2009.

Dans un premier temps, les débits de sortie de la chaîne SIM et du modèle GRP sont comparés aux débits journaliers observés de la banque HYDRO sur la période du 10 mars 2005 au 31 juillet 2006, et ceci pour les deux échéances équivalentes à celles des précipitations prévues PEARP. Ensuite, les prévisions issues du modèle GRP sont évaluées sur une période plus longue (jusqu'au 31 juillet 2009). Seules les prévisions GRP sont étudiées sur cette période étendue car les débits prévus par SIM n'étaient pas disponibles au-delà du 30 septembre 2006.

Les débits prévus par la chaîne SIM sont issus de la thèse de G. Thirel (2009) et ont été fournis par Météo-France. Pour les débits prévus par le modèle GRP, nous avons d'abord calé le modèle en utilisant les données historiques disponibles sur chaque bassin d'étude (75% des bassins étudiés ont plus de 27 années de données journalières) et ensuite effectué les prévisions en utilisant les scénarios de la prévision PEARP en entrée du modèle. Autant pour le calage que pour la simulation en continu des débits, le modèle GRP utilise les données de pluies issues de la ré-analyse SAFRAN de Météo-France (Quintana-Seguí *et al.*, 2008). Les données d'évapotranspiration potentielle ont été obtenues à partir de l'équation proposée par Oudin *et al.*, (2005), qui prend en compte la courbe de régime des températures et la latitude du bassin pour estimer le rayonnement extraterrestre. Ces données font partie de la base de données disponible au Cemagref à Antony sur plus de 3000 bassins versants en France.

En fonction de la disponibilité concomitante des séries de débits prévus par SIM et GRP, un total de 86 bassins versants français a été sélectionné (Fig. 5).

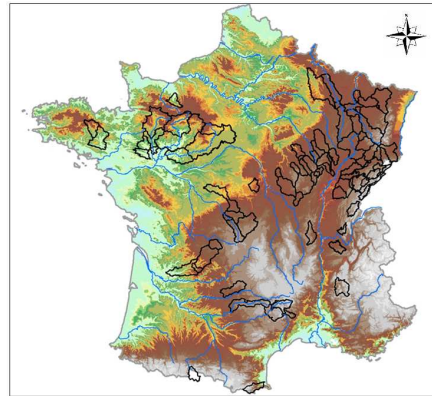


Figure 5 : Carte de localisation des 86 bassins versants étudiés en France

#### I-4. Scores d'évaluation des prévisions

Les scores (ou critères) utilisés pour évaluer la qualité des prévisions de débits sont présentés succinctement ci-après. Ils ont été calculés pour chaque échéance de prévision traitée ( $j+1$  et  $j+2$ ). Ils sont détaillés dans :

- *Forecast Verification - Issues, Methods and FAQ*. Disponible sur : <http://www.cawcr.gov.au/projects/verification/>
- "La prévision d'ensemble expliquée". Disponible sur : [http://www.meted.ucar.edu/nwp/pcu1/ensemble\\_fr/print.htm](http://www.meted.ucar.edu/nwp/pcu1/ensemble_fr/print.htm)
- Jolliffe, I.T. & Stephenson, D.B. (2003) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley & Sons, Chichester, 2003.

##### I-4.1 RMSE et ratio-RMSE

La RMSE (pour *Root Mean Square Error*, en anglais) mesure la magnitude moyenne des erreurs commises par la prévision (Eq. 1). Dans sa formulation, les erreurs sont élevées au carré avant qu'elles ne soient moyennées : les RMSE donnent alors un poids relativement plus important aux forts écarts. Cela signifie que la RMSE est un critère particulièrement utile lorsque les fortes erreurs sont particulièrement indésirables.

Pour chaque échéance de prévision, les RMSE des prévisions de débits ont été calculés pour chaque bassin versant et sur l'ensemble de chaque période d'évaluation en considérant la moyenne des membres de la prévision d'ensemble :

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

où,

$o_j$  : est la donnée observée pour le jour  $j$ ,

$m_j$  : est la moyenne des prévisions des 11 membres de l'ensemble pour le jour  $j$ ,

$N$  : est le nombre de jours utilisés pour calculer le score.

Pour s'affranchir de l'effet de taille des bassins versants et avoir ainsi des valeurs adimensionnelles comparables sur les différents bassins entre eux, le ratio entre la RMSE et le débit journalier moyen (débit observé) de chaque bassin a été calculé pour chaque jour de prévision (Eq. 2) :

$$ratio\_RMSE = \frac{RMSE}{\bar{Q}} \quad \text{Equation 2}$$

où,

$RMSE$  : est la valeur donnée par l'équation 1,

$\bar{Q}$  : est la moyenne des débits observés pendant la même période de calcul de la RMSE.

#### I-4.2 Tableau de contingence et CSI

Les tableaux de contingence permettent d'étudier l'association de deux variables et d'en déduire des probabilités conjointes et conditionnelles (Tab. 1). Ils indiquent le nombre de succès (événements prévu et observés), événements manqués (événements non prévus mais observés), fausses alertes (événements prévus mais pas observés) et rejets corrects (événements non prévus et non observés).

Pour construire un tableau de contingence, il faut définir un "événement observé" et un "événement prévu". Dans cette étude, un événement observé correspond au dépassement d'un seuil critique. Deux seuils ont été choisis : les quantiles 50% et 90% des débits observés, respectivement Q50 et Q90<sup>5</sup>. Ces quantiles sont calculés sur chaque période d'évaluation. Ainsi, un événement est compté "observé" (c'est-à-dire, colonne "OUI" dans le Tableau 1), si le débit observé dépasse le quantile considéré. Sinon, il est compté comme "non-observé" (dans la colonne "NON" du Tableau 1).

Pour la définition d'un événement prévu, trois seuils ont été considérés. Si un pourcentage (p%) des membres (scénarios) de débits prévus atteignent les seuils critiques de débits (Q50 et Q90 définis précédemment), alors l'évènement est considéré comme un "événement prévu" (ligne "OUI" dans Tab. 1). Sinon, il est "non prévu" (ligne "NON"). Les pourcentages étudiés ont été 20%, 50% et 80% des membres. Puisque nos premiers calculs ont montré peu de différence dans les résultats quand on considère ces différentes valeurs (Randrianasolo, 2009), nous montrerons les résultats pour ce score en utilisant un seul pourcentage.

Plusieurs scores peuvent être déduits d'un tableau de contingence. Nous avons choisi d'utiliser le CSI (*Critical Success Index*, en anglais), qui est un score qui fait la synthèse entre le taux de fausses alertes et le taux de succès (Eq. 3). Le CSI prend la valeur de 1 pour des "prévisions parfaites".

$$CSI = \frac{\text{succès}}{\text{succès} + \text{manqué} + \text{fausse alerte}} \quad \text{Équation 3}$$

		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 : Tableau de contingence

<sup>5</sup> Le quantile Q% est le plus petit élément q des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins Q% des données soient inférieures ou égales à q.

### I-4.3 Score de Brier (BS) et score de performance BSS

Le score de Brier (BS) et son score de performance, ou score d'habileté, (BSS, pour *Brier Skill Score*, en anglais) sont très largement utilisés pour évaluer les prévisions probabilistes, dont les prévisions d'ensemble. Ils qualifient les performances d'un système probabiliste par rapport à un événement particulier, ex., dépassement (ou non-dépassement) d'un seuil.

Le score BS est similaire à la RMSE, mais dans l'espace de probabilités : il compare les probabilités prévues et observées. Le BS mesure ainsi l'erreur quadratique moyenne de la probabilité prévue d'un événement (Eq. 4). Il est estimé en évaluant la somme des différences au carrées entre les probabilités de prévision et l'occurrence de l'observation (1 si l'événement a été observé et 0, sinon). Pour améliorer ce score, il faut augmenter la probabilité de prévision quand l'événement se produit et la diminuer quand l'événement ne se produit pas.

Comme la RMSE, le BS est négativement orienté : plus petite sa valeur, meilleure la performance du système évalué. Le score optimal ( $BS=0$ ) est atteint pour une prévision déterministe toujours correcte.

$$BS = \frac{1}{N} \sum_{j=1}^N (p_j - o_j)^2 \quad \text{Équation 4}$$

où,

$p_j$ : est la probabilité de prévision de l'évènement (pourcentage des membres des débits prévus qui dépassent le seuil critique) pour le jour  $j$ ,

$o_j$ : vaut 1 si l'évènement s'est produit au jour  $j$  et 0, sinon,

$N$ : est le nombre de jours utilisés pour calculer le score.

Le score de performance associé (BSS) varie entre  $-\infty$  et 1 et compare le score de Brier d'un système de prévision au score de Brier d'un système de référence (souvent la climatologie). Le BSS est positivement orienté (c'est-à-dire, les valeurs élevées indiquent meilleures performances). Un BSS de valeur zéro indique que le système de prévision évalué est équivalent au système de référence ; un BSS inférieur (ou supérieur) à zéro indique un système de performance inférieur (ou supérieur) au système de référence (Eq. 5).

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad \text{Équation 5}$$

où,

$BS$  : est le score de Brier du système évalué (Eq. 4),

$BS_{ref}$ : est le score de Brier du système utilisé comme référence.

De même que pour le calcul du score CSI, deux seuils de débits, calculés sur la période d'évaluation, ont été fixés pour définir les événements à évaluer avec le score de Brier : les quantiles Q50 et Q90 des débits observés.

#### I-4.4 Score de probabilité ordonné (RPS) et RPSS

Le principe du score de probabilité ordonnée (RPS, pour *Rank Probability Score*, en anglais) consiste à comparer non pas les probabilités prévues et observées, comme le fait le score BS pour un événement donné, mais à comparer les probabilités au non-dépassement prévues et observées sur un ensemble de catégories. Le score RPS est ainsi utilisé pour évaluer la précision d'un ensemble de prévisions de probabilités pour des variables ordonnées et s'applique à des multiples catégories : par exemple, les  $K+1 = 10$  catégories peuvent être les intervalles définis par les  $K$  quantiles Q10, Q20...Q90%. Dans cette étude, le calcul du RPS a été effectué en utilisant ces quantiles (discrétisation de la fonction de répartition prévue ici adoptée).

Chaque jour de prévision produit une séquence de probabilités de prévision :  $P_{jk}$ , avec  $P_j = (P_{j1} \leq P_{j2} \leq \dots \leq P_{jK})$ .  $P_{jk}$  est la probabilité cumulée prévue au jour  $j$  : probabilité prévue au non-dépassement du quantile  $k$ . Elle est donnée par le nombre de membres de la prévision d'ensemble qui ne dépassent pas le quantile  $k$  sur le nombre total de membres de l'ensemble. A chaque jour, un vecteur pour l'observation ( $O_j$ ) est également construit avec les  $k$  composantes :  $O_{jk} = 1$  ou  $0$ , selon que la vérification (l'observation) est respectivement plus petite ou plus grande que le quantile  $k$ . Le RPS dépend de la somme moyenne des différences au carré entre les distributions de probabilité cumulative observées et prévues pour l'ensemble de catégories (Eq. 6 ; Fig. 6) :

$$RPS_j = \frac{1}{K} \sum_{k=1}^K (P_{jk} - O_{jk})^2 \quad \text{Équation 6}$$

Pour un nombre  $N$  de jours de prévision :

$$\overline{RPS} = \frac{1}{N} \sum_{j=1}^N RPS_j \quad \text{Équation 7}$$

La valeur du RPS dépend du découpage en catégories adopté. La plage de valeurs admissibles se situe entre sa valeur minimale  $RPS^{min} = 0$  (cas parfait) est sa valeur maximale, qui dépend du nombre de catégories considérées (in extremis,  $RPS^{max} = K-1$ ). Ce score est en quelque sorte une mesure de "distance" entre les distributions de probabilité cumulative des prévisions et les observations (Fig. 6). Il est alors sensible à cette "distance", c'est-à-dire, il donne un plus mauvais score aux ensembles de prévision dont le centre de gravité (classe modale) est plus éloigné de la valeur observée.

Comme pour le BSS, il est également possible d'évaluer le score de performance RPSS par rapport à un système de référence (Eq. 8) :

$$RPSS = 1 - \frac{RPS}{RPS_{référence}} \quad \text{Équation 8}$$

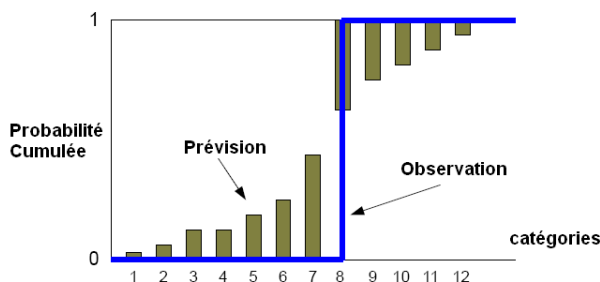


Figure 6 : Schéma représentant le calcul du RPS : pour les catégories inférieures à la catégorie où se place l'observation (ici, 8), la probabilité cumulée de prévision (au carré) de chaque catégorie participe à la somme de l'équation 6, tandis que pour les catégories supérieures, c'est la différence (au carré) entre chaque probabilité cumulée de prévision et la valeur 1 qui est prise en compte.

#### I-4.5 Référence pour les scores de performance et gain

Dans cette étude, nous avons utilisé la climatologie comme référence pour calculer les scores de performance BSS et RPSS. Pour la définir pour chaque bassin versant, les débits journaliers ont été obtenus grâce à des données historiques disponibles dans la base de données observées (données issues de la Banque Hydro entre 1970 et 2006). Pour chaque jour, une climatologie de débits est définie à partir des débits journaliers pour le même jour  $j$  des  $N$  années passées. Prenons, par exemple, une date quelconque : le 01 janvier. Les débits "climatologiques" pour cette date correspondent aux débits mesurés les 01 janvier des  $N$  dernières années. Dans cette étude, nous avons pris  $N = 20$  années.

De plus, pour comparer les performances des modèles avec et sans prise en compte d'une technique de mise à jour (assimilation d'une donnée observée avant le lancement de la prévision), un calcul de "gain" a été mise en place pour confronter les prévisions avec mise à jour aux prévisions effectuées sans mise à jour (Eq. 9). Les gains positifs (négatifs) sont obtenus quand la modélisation avec mise à jour fournit des prévisions de meilleure (moins bonne) performance par rapport à une modélisation sans mise à jour.

$$\begin{aligned} \text{Gain Ratio}_{RMSE} &= 1 - \frac{\text{RatioRMSE}_{\text{avec mise à jour}}}{\text{RatioRMSE}_{\text{sans mise à jour}}} \times 100 & \text{Gain CSI} &= 1 - \frac{\text{CSI}_{\text{sans mise à jour}}}{\text{CSI}_{\text{avec mise à jour}}} \times 100 \\ \text{Gain BS} &= 1 - \frac{\text{BS}_{\text{avec mise à jour}}}{\text{BS}_{\text{sans mise à jour}}} & \text{Gain RPS} &= 1 - \frac{\text{RPS}_{\text{avec mise à jour}}}{\text{RPS}_{\text{sans mise à jour}}} \end{aligned}$$

Équation 9

#### I-5. Visualisation des résultats

Pour permettre une visualisation globale des différents scores, nous avons choisi une représentation en boîte à moustaches.

Ce genre de graphique permet de représenter schématiquement la distribution d'une variable. Il est constitué de deux axes : un axe en ordonnée avec une échelle numérique qui correspond au quantile évalué selon le score représenté, et un axe en abscisse, sans aucune échelle, qui correspond aux différents systèmes évalués (GRP et SIM, avec et sans mise à jour).

Les quantiles représentés sont illustrés dans la Figure 7. Ils sont évalués en prenant en compte les scores des bassins versants étudiés. Selon cette représentation, la "boîte" contient alors 50% des valeurs du score représenté, tandis que 90% des valeurs se trouvent à l'intérieur des limites des moustaches.

Des cartes de France représentant le score moyen pour chaque bassin versant étudié sont également présentées. Elles offrent un aperçu de la performance individuelle de chaque bassin versant.

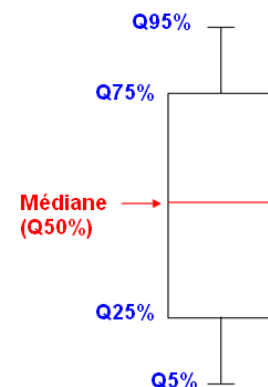


Figure 7 : Visualisation des résultats : quantiles des boîtes à moustaches



## Partie 2 : Résultats

### ***II-1. Evaluation des prévisions hydrologiques SIM et GRP pour la période de mars 2005 à juillet 2006***

Nous présentons ici les scores évalués sur les débits prévus de la chaîne SIM et du modèle GRP (modélisations avec et sans mise à jour) pour l'évaluation de la première archive PEARP, du 10/03/2005 au 31/07/2006. Une partie de cette étude avait déjà été menée dans le cadre du protocole d'expérimentation PEARP signé entre le SCHAPI, le Cemagref et Météo-France en 2009 (résultats présentés dans Randrianasolo, 2009 et Randrianasolo *et al.*, 2010). Ici, nous apportons deux éléments nouveaux : (i) l'étude de l'impact de la mise à jour sur les résultats de la modélisation SIM et (ii) l'évaluation de l'évolution des performances selon les surfaces, les durées caractéristiques de crue et les régimes hydrologiques des bassins versants.

Notons enfin que plusieurs scores ont été calculés et ici, dans un souci de synthèse, nous présentons seulement un choix représentatif des résultats. De plus, puisqu'il a été observé que les différences entre les valeurs des scores des deux échéances traitées étaient globalement faibles, notamment pour les débits issus de la chaîne SIM, dans la suite de ce rapport nous ne présentons que les scores obtenus pour l'échéance j+2.

#### **II-1.1 Performances et impact de la mise à jour sur les prévisions**

Les graphes de la Figure 8 illustrent le fait que, pour les deux modèles étudiés, les performances générales augmentent quand le modèle est utilisé avec sa technique de mise à jour en prévision : tous les scores médians sont meilleurs en situation de mise à jour.

Les scores probabilistes (BSS et RPSS) montrent que les prévisions sont, en général, bien meilleures qu'une prévision climatologique qui n'utiliserait pas les prévisions météorologiques PEARP (scores > 0), et ceci surtout quand les modèles sont considérés avec leurs procédures de mise à jour. Pour le RPSS, en particulier, toutes les situations donnent des prévisions supérieures à la climatologie, sauf pour quelques bassins versants modélisés par SIM sans technique de mise à jour.

Les gains en performance avec la mise à jour sont présentés sous la forme de cartes de France pour le ratio-RMSE (Fig. 9) et de boîtes à moustaches (Fig. 10) pour les autres scores. Les gains positifs (ou en bleus sur les cartes) indiquent une meilleure performance avec la mise à jour.

De manière générale, la majorité des bassins montre un gain positif avec la mise à jour, surtout pour le modèle GRP. Pour les débits prévus par le modèle SIM, quelques bassins ont montré plus de performance quand le modèle est utilisé sans la mise à jour (points rouges dans la Fig. 9 et scores inférieurs à zéro dans la Fig. 10). Pour les valeurs de RMSE de la Figure 9, les gains les plus importants sont observés pour les bassins situés dans le Nord-Ouest de la France (notamment stations situées dans le bassin de la Loire).



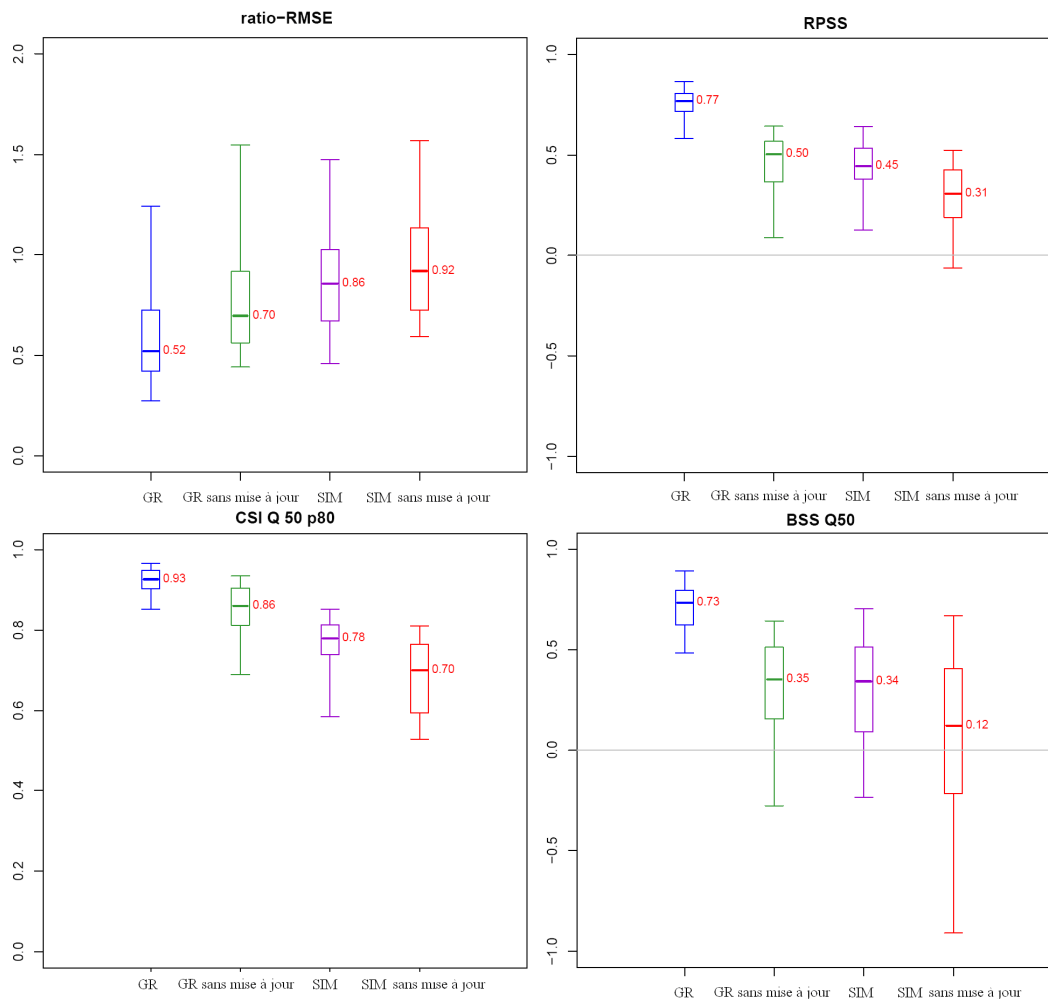


Figure 8 : Comparaison entre les modèles sans et avec mise à jour pour les débits prévus par le modèle GRP et la chaîne SIM : ratio-RMSE (haut à gauche), RPSS (haut à droite), CSI pour 80 % des membres dépassant le quantile de débit Q50 (bas à gauche) et BSS pour le dépassement du quantile Q50 des débits observés (bas à droite). Statistiques sur 86 bassins versants. Prévisions évaluées à deux jours d'échéance.

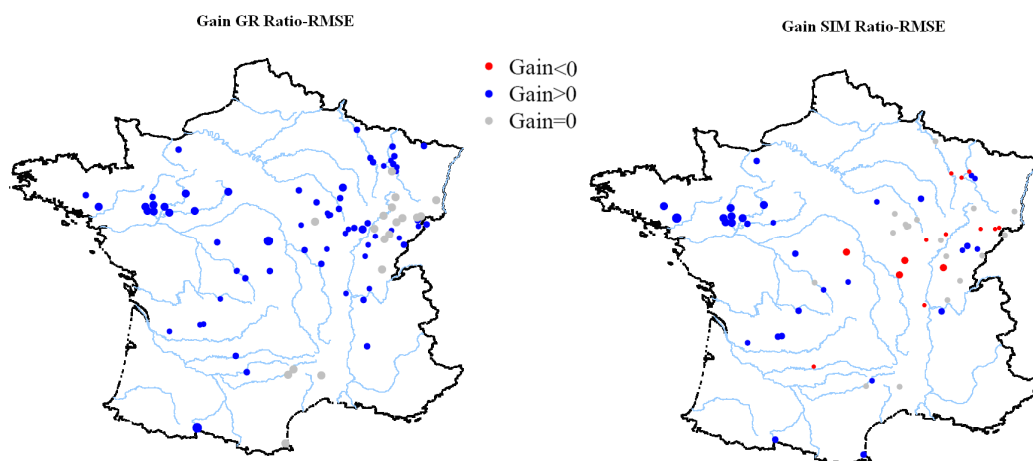


Figure 9 : Carte de gain avec la mise à jour du modèle pour chaque bassin versant étudié et selon le score ratio\_RMSE : modèle GRP (à gauche) et chaîne SIM (à droite). La taille des points indique un gain (positif ou négatif) plus ou moins grand. Prévisions évaluées à deux jours d'échéance.

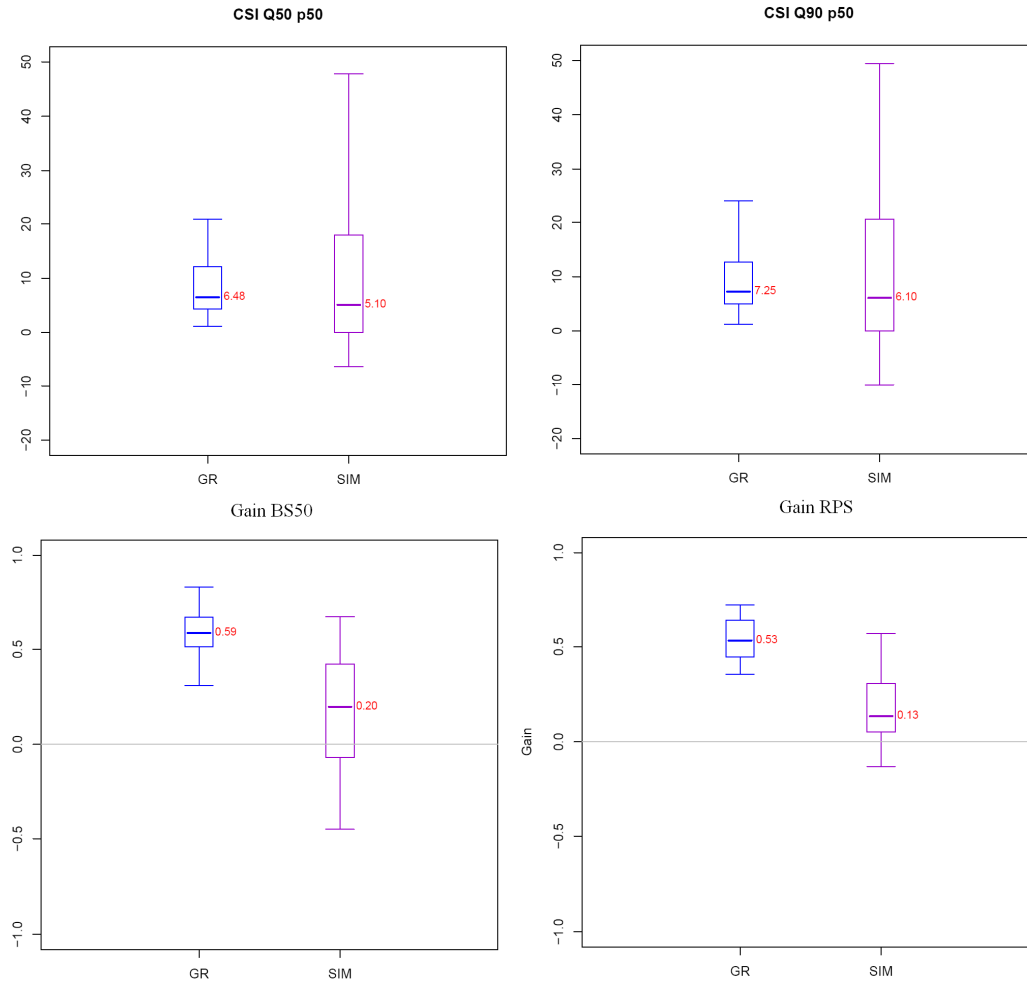


Figure 10 : Boîtes à moustaches des valeurs de gains avec la mise à jour pour les scores : CSI avec 50% des membres dépassant les seuils de débit Q50 (haut à gauche) et Q90 (haut à droite) ; BS pour le dépassement du quantile Q50 (bas à gauche) et RPS (bas à droite). Statistiques sur 86 bassins versants. Prévisions évaluées à deux jours d'échéance.

La Figure 10 montre que les gains les plus importants en matière de score CSI sont observés pour le modèle SIM, autant pour ce qui est du dépassement du quantile de débit Q50 que pour le quantile Q90. Les gains par rapport aux scores de Brier et RPS sont tous positifs pour les prévisions de GRP : les prévisions avec mise à jour sont toujours plus performantes. Les valeurs médianes pour les débits issus de la chaîne SIM sont également positifs, même si quelques bassins présentent des scores négatifs ou nuls. En comparant les deux modèles, on observe que les valeurs de gains de GRP pour les scores BS et RPS sont généralement plus fortes, c'est-à-dire, que l'influence de la mise à jour est plus forte sur ce modèle selon ces critères.

Les résultats ont montré que la modélisation avec mise à jour produit, dans les deux cas de modélisation SIM et GRP, des meilleures prévisions. Dans les analyses qui suivent nous n'évaluons donc que les débits issus des modélisations avec mise à jour.

### **II-1.2 Evaluation des prévisions selon les surfaces des bassins**

Les bassins versants de la base de données ont été divisés en trois groupes selon leur surface. La performance des prévisions a ensuite été calculée pour chaque groupe :

- *Petits bassins* :  $S < 800 \text{ km}^2$  ( $n = 29$  bassins)
- *Bassins moyens* :  $800 < S < 2000 \text{ km}^2$  ( $n = 32$  bassins)
- *Grands bassins* :  $S > 2000 \text{ km}^2$  ( $n = 25$  bassins)

La Figure 11 montre les résultats obtenus pour chaque groupe de surface et pour les quatre scores de performance étudiés (ratio\_RMSE, CSI avec Q50, BSS avec Q50 et RPSS). Globalement, on observe que les meilleures performances sont obtenues pour le groupe avec les plus grands bassins versants (surface supérieur à  $2000 \text{ km}^2$ ), quel que soit le score considéré et pour les deux cas de modélisation (GRP et SIM). De manière générale, les performances s'améliorent quand on passe du groupe des plus petits bassins (surface inférieurs à  $800 \text{ km}^2$ ) au groupe représentant les plus grands bassins de l'échantillon.

### **II-1.3 Evaluation des prévisions selon les durées caractéristiques de crue**

Les bassins versants étudiés ont été classés selon leurs durées caractéristiques de crue. Pour évaluer cette grandeur, nous avons utilisé l'approche par hydrogramme moyen de crue normalisé. La construction des formes des crues normalisées a suivi les étapes présentées dans Sauquet *et al.* (2008b). Pour chaque bassin versant, il s'agit de sélectionner les hydrogrammes de crue, les normaliser par le débit de pointe et ensuite évaluer un hydrogramme moyen, duquel nous récupérerons une durée caractéristique. Cette durée est donnée par la durée continue pendant laquelle le débit dépasse un débit-seuil fixé. Ici, le seuil pris a été la moitié du débit de pointe normalisé. Les bassins plus rapides ont ainsi une durée caractéristique de crue plus courte.

La Figure 12 présente l'évolution des quatre scores de performance étudiés (ratio\_RMSE, CSI avec Q50, BSS avec Q50 et RPSS) selon la durée caractéristique de crue des bassins versants de cette étude.

Globalement, on observe que les valeurs des scores s'améliorent avec l'augmentation des durées caractéristiques de crue : plus le bassin est lent, plus la performance des prévisions est meilleure, quel que soit le score pris en compte.

Cette tendance est néanmoins plus marquée pour le modèle GRP que pour le modèle SIM quand on considère, notamment, les scores probabilistes BSS et RPSS (les résultats de ces scores pour le modèle SIM sont peu sensibles à la durée caractéristique de crue des bassins). Pour les débits prévus par le modèle SIM, la durée caractéristique de crue semble, en effet, jouer un plus grand rôle quand on considère les scores ratio-RMSE et CSI.

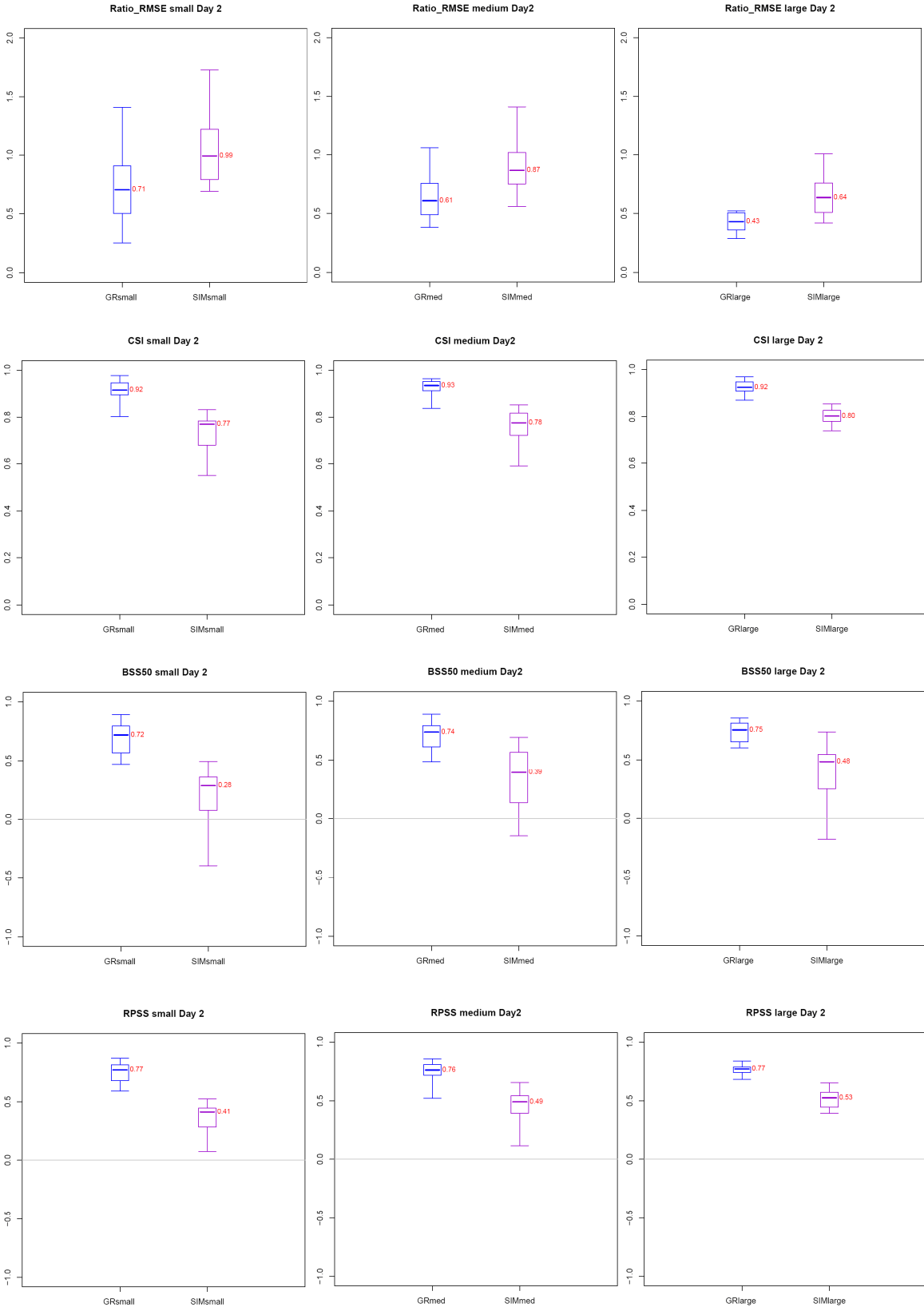


Figure 11 : Boîtes à moustaches des valeurs des scores selon chaque groupe de surface des bassins étudiés. Petits bassins (à gauche)  $S < 800 \text{ km}^2$  ( $n = 29$  bassins) ; Bassins moyens (au centre)  $800 \text{ km}^2 < S < 2000 \text{ km}^2$  ( $n = 32$  bassins) ; Grands bassins (à droite)  $S > 2000 \text{ km}^2$  ( $n = 25$  bassins). Prévisions évaluées à deux jours d'échéance.

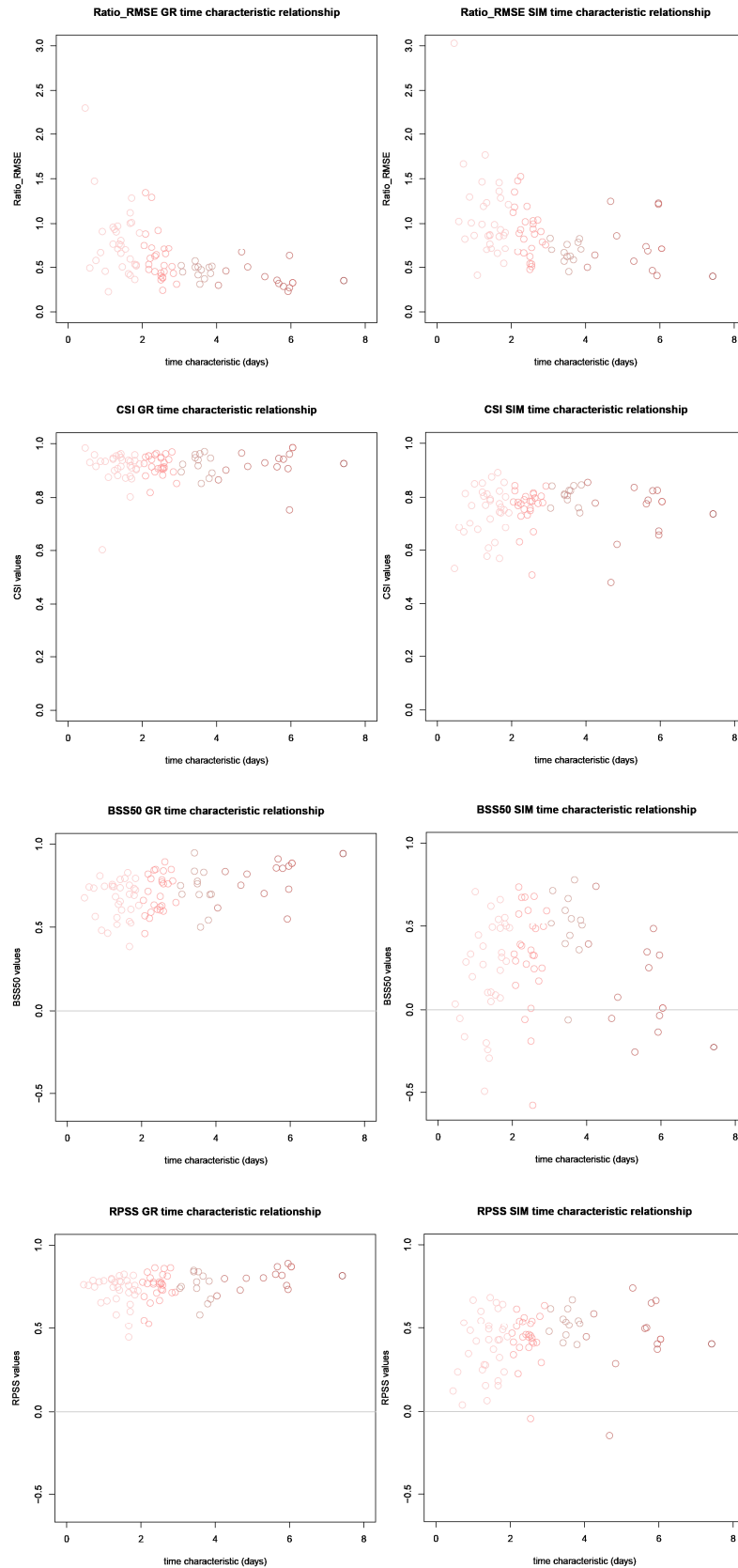


Figure 12 : Evolution des scores de performance selon les durées caractéristiques de crue des bassins versants étudiés pour les prévisions de débits du modèle GR (à gauche) et de la chaîne SIM (à droite). Prévisions évaluées à deux jours d'échéance.

## II-1.4 Evaluation des prévisions selon les régimes hydrologiques

Le régime hydrologique est le reflet de plusieurs facteurs influençant l'écoulement, tels que les précipitations, la température, le relief, la nature du bassin versant, etc. Il existe plusieurs types de régimes hydrologiques, fondés principalement sur le mode d'alimentation du cours d'eau (pluvial, nival, glaciaire), ainsi que plusieurs approches de classification. Ici, le groupement des bassins versants en trois grands régimes hydrologiques a été fait sur la base de la méthode présentée par Sauquet (2006) et Sauquet *et al.* (2008a), qui a proposée une classification de régimes hydrologiques en 12 groupes, à partir de l'analyse statistique des chroniques de débits mensuels de 872 stations de jaugeage de cours d'eau français sur la période 1981-2000. Cette classification repose sur le principe simple de similitude, selon une stratégie de regroupement qui s'appuie sur la méthode de Ward (minimisation des distances intra-classes, maximisation des distances inter-classes).

Sur la base des chroniques journalières de débits disponibles, le régime hydrologique de chacun des 86 bassins versants de cette étude a été évalué. Pour cela, nous avons transformé les débits journaliers en débits mensuels moyens, puis en pourcentage du débit annuel, pour pouvoir les comparer aux hydrogrammes de référence proposés par la méthode de Sauquet (2006). A fin de disposer d'un échantillon statistiquement représentatif sur chaque classe, un regroupement a été fait. Nous avons ainsi classé les bassins versants de l'étude en trois régimes hydrologiques (Fig. 13) :

- Régime 1 : régime à dominante pluviale, hautes eaux observées principalement en hiver ( $n = 45$  bassins)
- Régime 2 : régime de transition, hautes eaux observées au printemps et en hiver ( $n = 36$  bassins)
- Régime 3 : régime à dominante nivale ( $n = 5$  bassins)

Les résultats obtenus pour les scores de performance des débits prévus par les modèles SIM et GRP sont présentés dans la Figure 14. Ils montrent que, en général, les valeurs des scores étudiés diffèrent peu selon les régimes hydrologiques. Pour le modèle GRP, les performances sont légèrement meilleures pour les régimes à dominante pluviale et de transition (régimes 1 et 2). Les différences entre régimes sont plus importantes pour le modèle SIM, avec le régime 3, à dominante nivale, présentant les scores les plus mauvais (à noter cependant que le régime nivale est mal représenté dans notre échantillon car il ne contient que 5 bassins versants).

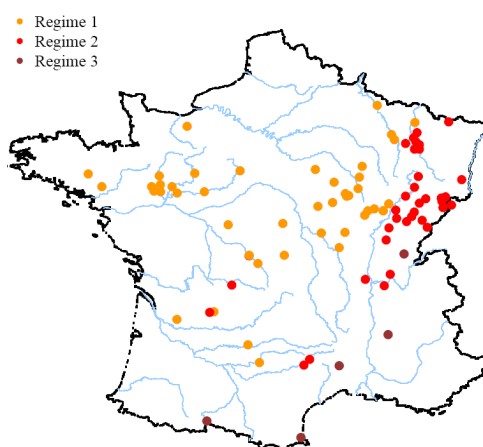


Figure 13 : Représentation des bassins d'étude selon leur régime hydrologique. Régime 1 à dominante pluviale ; Régime 2 de transition ; Régime 3 à dominante nivale.

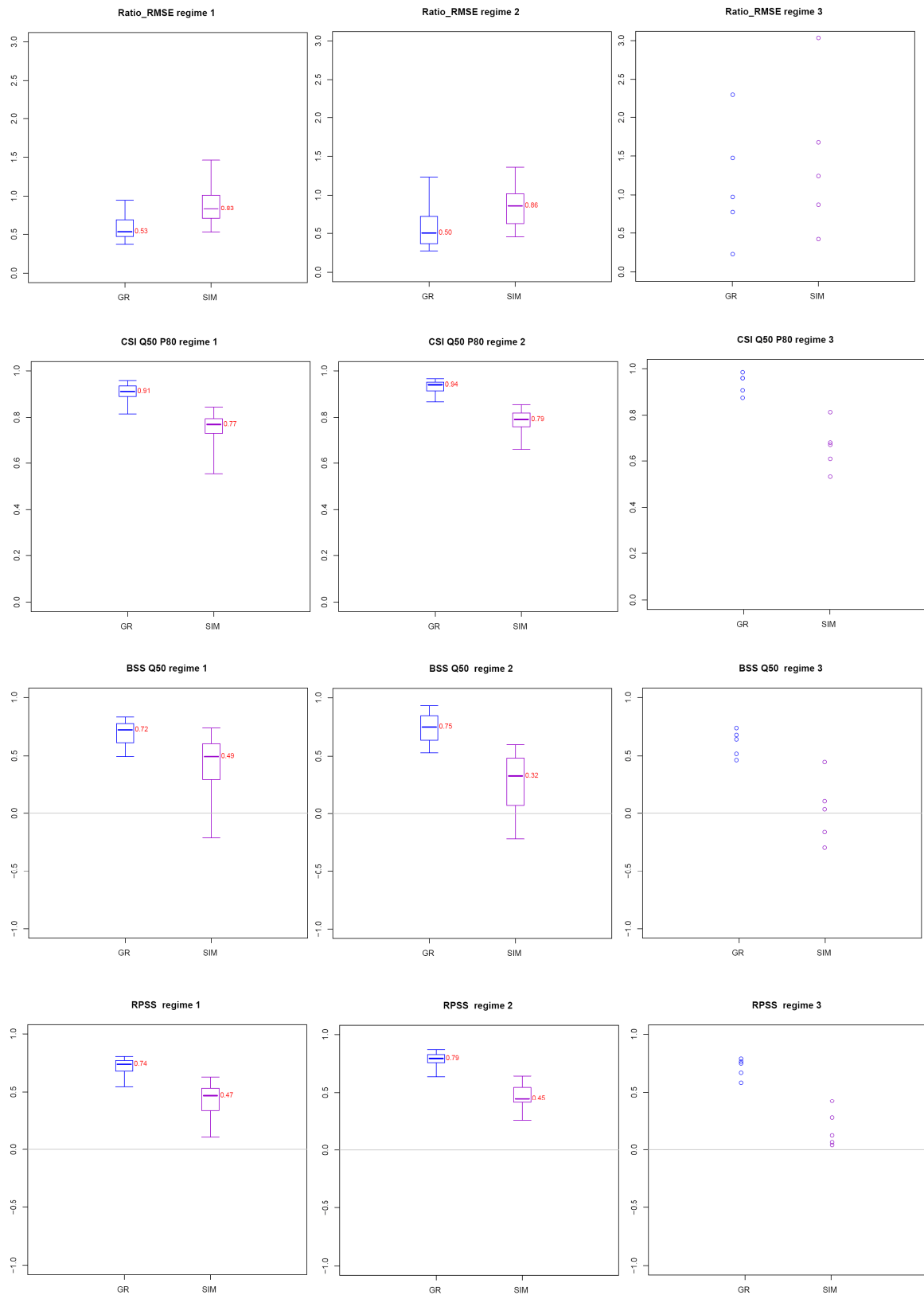


Figure 14 : Boîtes à moustaches des valeurs des scores selon les régimes hydrologiques des bassins versants étudiés. Régime 1, à dominante pluviale ( $n = 45$  bassins) ; Régime 2, de transition ( $n = 36$  bassins). Colonnes de gauche : scores pour les  $n = 5$  bassins du Régime 3, à dominante nivale. Prévisions évaluées à deux jours d'échéance.

## II-2. Evaluation des prévisions hydrologiques GRP pour la période de mars 2005 à juillet 2009 (archive étendue)

La première expérimentation avec les prévisions hydrologiques des modèles GRP et SIM, alimentés par les prévisions météorologiques d'ensemble PEARP de Météo-France, a encouragé la poursuite des études et une extension de l'archive PEARP a été mise à disposition du Cemagref. Cette nouvelle archive étendue prolongeait les données de prévision PEARP jusqu'à septembre 2009. Nous avons ensuite alimenté le modèle GRP journalier avec ces prévisions et le résultat de l'évaluation de la qualité des prévisions sur cette archive étendue est présenté dans ce chapitre.

Nous utilisons les mêmes critères statistiques décrits dans §1.4, mais ici ils ne sont appliqués qu'aux débits prévus par le modèle GRP (les débits prévus pas la chaîne SIM n'étant pas disponibles sur l'ensemble de cette période étendue). De plus, puisque la base de données de débits de la banque HYDRO disponible au Cemagref s'arrête au 31 juillet 2009, la période d'évaluation s'arrête également à cette date.

Nous présentons une comparaison entre les performances des débits prévus pendant la première période, plus courte, de mars 2005 à juillet 2006 (cf. résultats précédents) et les performances pendant la nouvelle période plus longue (mars 2005 à juillet 2009). L'effet de la taille de l'échantillon de vérification des prévisions peut ainsi être évalué.

### II-2.1 Caractéristiques descriptives des deux périodes d'évaluation

Le Tableau 2 compare quelques statistiques descriptives des débits et pluies observés pour chaque période d'évaluation. Nous considérons ici les valeurs moyennes sur l'ensemble de l'échantillon et les valeurs maximales et minimales observées parmi celles de chaque bassin de l'échantillon. On observe que les moyennes des pluies et débits journaliers moyens sur les bassins de l'échantillon ne diffèrent pas significativement, ainsi que les valeurs de pluies et débits journaliers moyens maximaux et minimaux et les statistiques en cumul de pluie annuel (janvier à décembre).

Cependant, quand on regarde en détail les dates des débits et pluies journaliers maximaux de chaque bassin, on observe qu'environ 60% des bassins ont enregistré leurs débits maximaux pendant la période étendue (c'est-à-dire, après le 31/07/2006) et plus de 90% ont enregistré une pluie journalière maximale pendant également la période étendue. Cette analyse sommaire semble indiquer que, même si la période étendue est comparable à la période courte en statistiques moyennes, elle comprend des épisodes plus forts, en pluies et débits, sur les bassins versants de l'échantillon. Les dates des valeurs maximales indiquent qu'environ 80% des événements responsables de ces maximaux en pluies et débits journaliers ont eu lieu pendant la période étendue.

(mm)	Période courte (17 mois)	Période étendue (53 mois)
$P_j$ moy [ $P_j$ max ; $P_j$ min]	2.75 [4.67 ; 1.62]	2.89 [4.96 ; 1.70]
$PN$ moy [ $PN$ max ; $PN$ min]	892 [1370 ; 563]	880 [1339 ; 549]
$Q_j$ moy [ $Q_j$ max ; $Q_j$ min]	1.08 [6.01 ; 0.12]	1.22 [6.13 ; 0.18]

Tableau 2 : Statistiques des observations sur les deux périodes d'évaluation (en mm) : moy = moyenne sur l'ensemble de bassins versants ; max et min = valeur maximale et minimale sur un bassin versant de l'échantillon ;  $Q_j$  et  $P_j$  pour débit et pluie journaliers et  $PN$  pour cumul annuel des pluies.



## II-2.2 Performance comparée

Pour des raisons de disponibilité de données au moment de l'étude de la période étendue, quatre bassins versants ont dû être écartés de l'analyse des prévisions. Les scores appliqués précédemment ont été calculés pour la nouvelle période et comparés avec les performances sur la période plus courte. La Figure 15 montre les résultats pour les scores ratio-RMSE et CSI (dépassement des quantiles Q50% et Q90%), tandis que la Figure 16 présente les résultats pour les scores probabilistes BS et RPS.

Le score ratio-RMSE montre, en général, des valeurs plus élevées (plus mauvaise performance) pour la plus longue période. Nous avons observé que les bassins à score RMSE élevé de la période courte sont restés mauvais et ont même empirés pour la période étendue (ceci sauf pour deux bassins en aval de la Loire, pour lesquels les ratio-RMSE ont diminué pour la plus longue période).

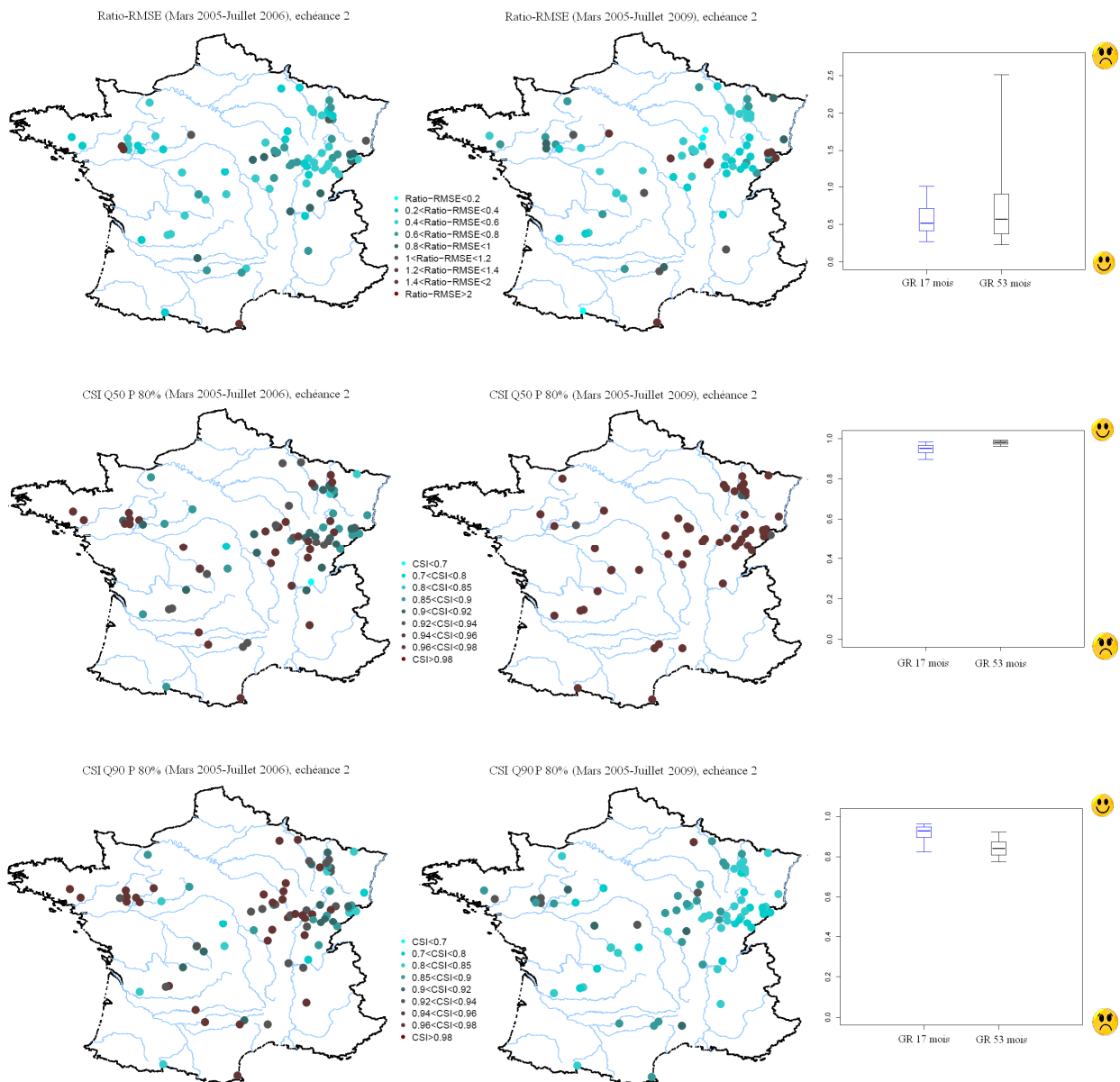


Figure 15 : Cartes et boîtes à moustaches comparant les scores ratio-RMSE et CSI des deux périodes d'évaluation des prévisions. A gauche, carte des scores de chaque bassin versant pour la période courte (03/2005-07/2006) ; au milieu, carte pour la période étendue (03/2005-07/2009) ; à droite, boîtes à moustaches. Prévisions évaluées à deux jours d'échéance.

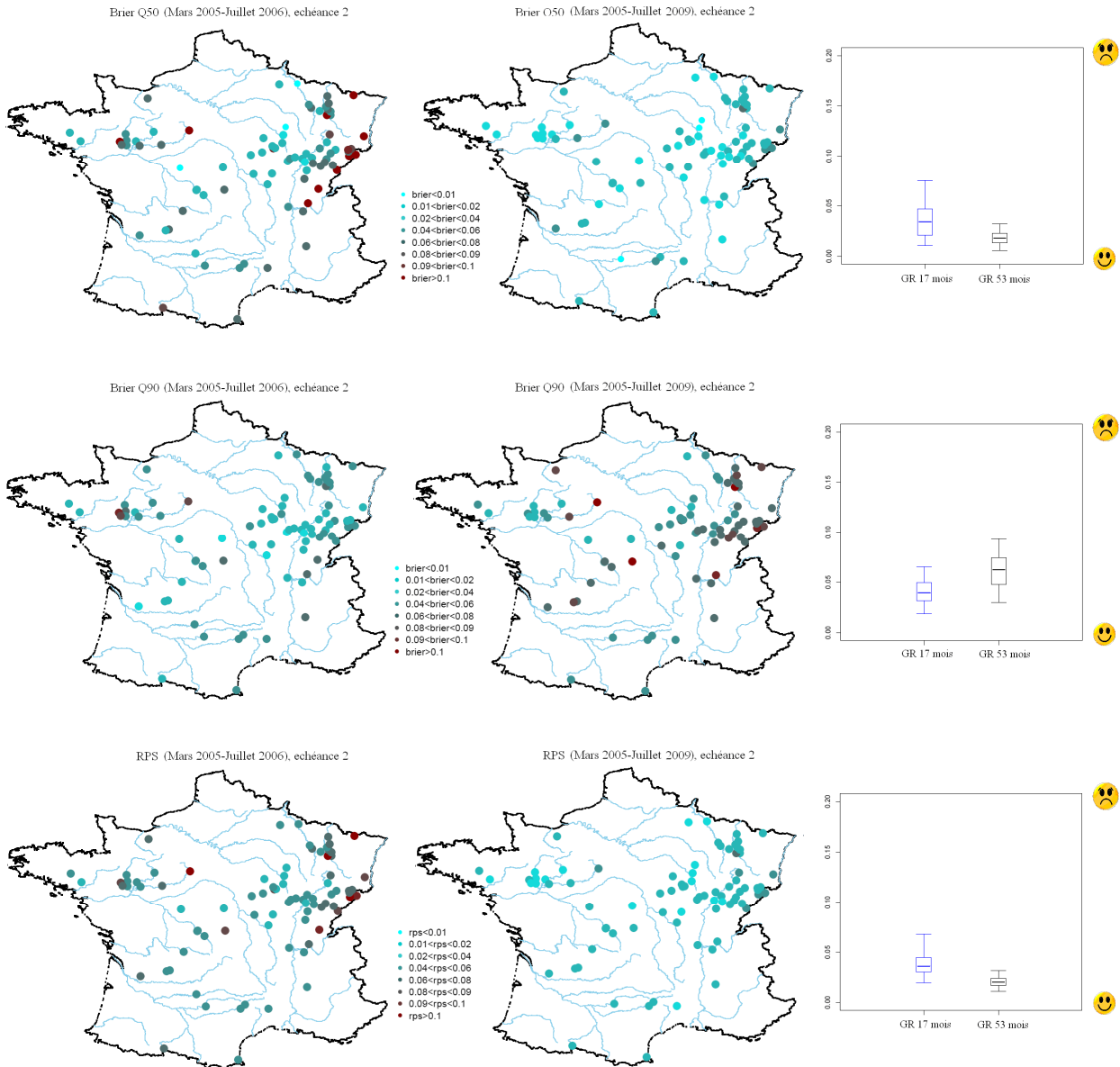
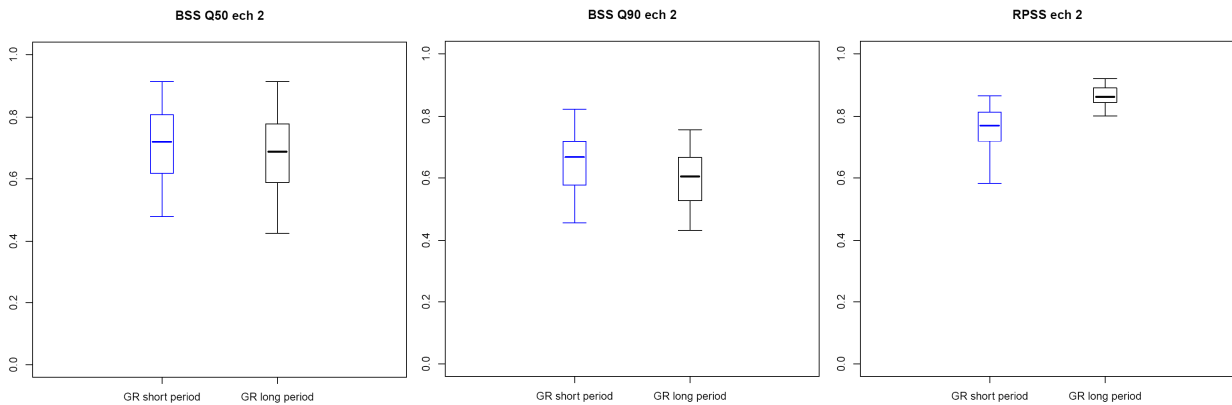


Figure 16 : Cartes et boîtes à moustaches comparant les scores BS et RPS des deux périodes d'évaluation des prévisions. A gauche, carte des scores de chaque bassin versant pour la période courte (03/2005-07/2006) ; au milieu, carte pour la période étendue (03/2005-07/2009) ; à droite, boîtes à moustaches. Prévisions évaluées à deux jours d'échéance.

Le score CSI (Fig. 15) montre deux tendances différentes que l'on considère le dépassement du quantile-seuil Q50% ou celui du Q90% (pour un même pourcentage de membres considérés) : la performance pour le dépassement du quantile Q50 (médiane des débits observés) s'améliore avec le rallongement de la période de vérification, tandis que les scores pour le dépassement du quantile Q90% deviennent moins bons.

Il en est de même pour le score de Brier (Fig. 16) : l'évaluation des probabilités de dépassement du quantile Q90 donne un score plus mauvais pour la longue période de 53 mois, alors que pour le seuil Q50, le score de Brier diminue (meilleure performance). Pour le score RPS, tous les bassins voient leur score diminuer : les performances sont donc meilleures quand on considère la plus longue période de vérification. Le fait que la période étendue semble contenir plus d'événements forts joue donc un rôle important sur les scores qui ciblent les prévisions de dépassements des quantiles plus élevés.

Au regard des scores probabilistes comparés avec la climatologie (BSS et RPSS) (Fig. 17), on observe une performance stable du BSS entre la courte et la longue période d'évaluation pour les dépassements du quantile Q50, et une légère baisse de la valeur du score pour le dépassement du quantile Q90 pendant la période étendue. Le score de probabilité ordonné comparé à la climatologie, RPSS, montre une meilleure performance quant on dispose d'une période plus longue d'évaluation des prévisions. Pour ces deux scores et pour tous les bassins, le système de prévision présente une meilleure performance que la climatologie (scores > 0).



*Figure 17 : Scores de performance BSS et RPSS évalués sur deux périodes de prévision : à gauche de chaque graphique, période courte (03/2005-07/2006) ; à droite, période étendue (03/2005-07/2009). Prévisions évaluées à deux jours d'échéance.*

# Conclusions

Ce rapport a présenté une synthèse des résultats obtenus par les études menées au Cemagref pour évaluer la qualité des prévisions hydrologiques émises par le modèle hydrologique GRP (Cemagref) et la chaîne de modélisation SIM (Météo-France) à partir d'une même entrée de précipitation prévue, la prévision météorologique d'ensemble PEARP de Météo-France (11 scénarios futurs et 60 heures d'échéance maximale de prévision). La qualité des prévisions hydrologiques d'ensemble journalières a été évaluée sur un large échantillon de bassins versants en France et sur une sélection de critères couramment utilisés dans la littérature.

Deux périodes d'évaluation ont été examinées : une période plus courte, allant du 10/03/2005 au 31/07/2006, et une période étendue, jusqu'au 31/07/2009. Nous avons examiné (i) l'impact de la mise à jour sur les résultats de chaque modèle et (ii) l'évolution des performances selon les surfaces, les durées caractéristiques de crue et les régimes hydrologiques des bassins versants.

Les principales conclusions sont :

## Impact de la mise à jour des modèles sur les prévisions de débits

Les performances des prévisions hydrologiques sont meilleures quand on considère la modélisation avec mise à jour, que ce soit pour la chaîne de modélisation SIM ou le modèle GRP. Au vue des scores de performances comparées (BSS et RPSS), il ressort que ces systèmes de prévision sont plus performants qu'une prévision basée sur la climatologie des débits passés (ceci sauf pour moins d'environ 15% des bassins versants modélisés avec la chaîne SIM). Ces résultats soulignent l'importance des prévisions météorologiques et de la mise à jour en prévision hydrologique.

## Evolution des scores

### ➤ avec l'échéance

Pour tous les scores et modèles, les performances diminuent avec l'échéance de prévision ; un résultat conforme à celui attendu. Les différences entre les deux échéances journalières évaluées sont plus importantes pour le modèle GRP, que pour le modèle SIM, qui semble être moins impacté par l'augmentation de l'échéance.

### ➤ avec les surfaces des bassins

De manière générale, les performances augmentent avec la taille des bassins versants pour les deux modèles utilisés avec mise à jour.

### ➤ avec les durées caractéristiques de crue

Les performances des débits prévus par le modèle GRP sont globalement meilleures pour les bassins les plus lents. Quant aux débits prévus par SIM, la même tendance est observée en examinant les scores déterministes, même si les scores probabilistes semblent être moins sensibles à l'augmentation de la durée caractéristique de crue.

➤ avec les régimes hydrologiques

En général, les scores sont légèrement meilleurs pour les régimes à dominante pluviale. Cependant, ces conclusions ne sont pas robustes car notre échantillon ne contenait que cinq bassins caractérisés par le régime à dominante nivale. Des études complémentaires sur un échantillon plus large de bassins versants sont alors nécessaires, notamment en utilisant le modèle GRP avec le module CemaNeige qui a été récemment développé au Cemagref.

➤ avec la longueur de la période d'évaluation

Le score RMSE et les scores liés au dépassement du quantile 90% des débits observés (CSI Q90 et BS Q90) ont montré une performance plus faible des systèmes de prévision lors qu'ils ont été évalués sur la période étendue (53 mois) par rapport aux scores obtenus sur la période plus courte (17 mois). Le fait que la période étendue contienne probablement plus d'événements forts semble alors impacter la valeur des scores. Notons que le caractère "moins arrosé" de la période plus courte avait déjà été signalé par les travaux de Thirel *et al.* (2008). Des études plus approfondies sont nécessaires pour mieux examiner cette question.

Enfin, il faut souligner que l'évaluation d'une prévision d'ensemble n'est souvent pas une tâche aisée. En effet, une prévision déterministe (scénario futur unique) est souvent plus facile à évaluer qu'une prévision d'ensemble (plusieurs scénarios et prévisions exprimées sous la forme de probabilité au (non-)dépassement de seuils critiques). Dans le cas déterministe, soit l'événement a été observé, soit il n'a pas été observé. Or, dans le cas d'une prévision sous la forme de probabilités (ex., 40% de chances de dépasser un seuil critique), il faut évaluer la fiabilité et la justesse d'une telle prévision en comparant les probabilités de prévision et les fréquences d'observation sur une longue série de prévisions émises et pour un ensemble de valeurs de débits.

Néanmoins, les prévisions probabilistes contiennent une information clé pour les prévisionnistes. L'approche de la prévision d'ensemble permet de disposer d'une indication de l'incertitude de prévision, ce qui s'avère plus adapté à la prise de décisions face aux risques hydrologiques. L'utilisation de ces prévisions n'est sans doute pas simple et la pratique sur des événements hydrologiques passés est essentielle pour mieux comprendre la valeur d'une prévision probabiliste, ses limites et avantages.

Enfin, les études que nous avons menées font ressortir plusieurs aspects concernant l'usage et l'évaluation de la qualité des prévisions hydrologiques d'ensemble, notamment :

- l'importance des prévisions météorologiques dans la chaîne de prévision hydrologique et de la mise à jour des modèles pour une meilleure prévision hydrologique ;
- l'importance de mener les études d'évaluation des prévisions de débits sur un large échantillon de bassins versants à caractéristiques variées et soumises à des conditions climatiques diverses ;
- l'importance de disposer d'archives longues de prévisions météorologiques d'ensemble pour pouvoir alimenter les modèles hydrologiques, évaluer les prévisions de débits de manière robuste et améliorer leur qualité (fiabilité et justesse), ainsi que pour pouvoir examiner plus en détail les événements hydrologiques forts observés dans le passé et à intérêt particulier dans la gestion des risques ;
- l'importance de la prévision probabiliste dans la quantification des incertitudes de prévision et dans l'aide à la prise de décision.

## Références bibliographiques

- Berthet L. (2010). *Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique*. Thèse de doctorat, AgroParisTech, Cemagref, 405 p.
- Jolliffe, I.T., Stephenson, D.B. (2003) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley and Sons, Chichester, England, 2003, 254 p.
- Martin, E., G. Thirel, F. Regimbeau, J. Noilhan, L. Franchistéguy, J-M. Soubeyroux, F. Habets (2008) Vers une prévision d'ensemble de débits à l'échelle des grands bassins français. *Colloque SHF Prévisions hydrométéorologiques*, Lyon, 18-19/11/2008, 8p.
- Nicolau J. (2002) Short-range ensemble forecasting. In *Proceedings WMO/CBS Technical Conference on Data Processing and Forecasting Systems*, Cairns, Australia, 2-3 Dec, 6p. Disponible sur: <http://www.wmo.int/pages/prog/www/DPS/TC-DPFS-2002/Papers-Posters/Topic1-Nicolau.pdf>
- Nicolle, P. (2010) *Prise en compte de la neige dans la prévision hydrologique : quel apport pour mieux prévoir les crues nivales ?* Mémoire de Master, UPMC, Cemagref, 53 p.
- Nicolle, P., Ramos, M.H., Andréassian, V., Valéry, A. (2011) Mieux prévoir les crues nivales : évaluation de prévisions probabilistes de débit sur des bassins versants de montagne français. *Proc. Colloque SHF "Eau en montagne"*, Lyon, 16-17 mars 2011, 163-170.
- Oudin L., Hervieu F., Michel C., Perrin C., Andréassian V., Anctil F. et Loumagne C. (2005) Which potential evapotranspiration input for a rainfall-runoff model? Part 2 - Towards a simple and efficient PE model for rainfall-runoff modelling. *J. of Hydrology*, 303(1-4): 290-306.
- Perrin, C., Michel, C., Andréassian, V. (2003) Improvement of a parsimonious model for streamflow simulation. *J. of Hydrology*, 279 (1-4): 275-289.
- Perrin, C., Michel, C., Andréassian, V. (2001) Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. of Hydrology*, 242: 275-301.
- Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., Morel, S. (2008) Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47 (1): 92-107.
- 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 Prévisions hydrométéorologiques*, Lyon, 18-19/11/2008, 191-199.
- Randrianasolo, A. (2009) *Evaluation de la qualité des prévisions pour l'alerte aux crues*. Mémoire de Master UPMC-AgroParisTech, Cemagref, 50 p.
- Randrianasolo, A., Ramos, M.H., Andreassian, V. (2011) Hydrological ensemble forecasting at ungauged basins: using neighbour catchments for model setup and updating. *Adv. in Geosciences*, 29: 1-11.
- Randrianasolo, A., Ramos, M-H, Thirel, G., Andréassian V., Martin, E. (2010) Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmospheric Science Letters* 11: 100-107.
- 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, Météo-France, 233 p.
- Rousset Regimbeau, F., Habets, F., Martin, E., Noilhan, J. (2007) Ensemble streamflow forecasts over France. *ECMWF Newsletter* 111, 21-27.
- Sauquet E., (2006) *Cartographie des écoulements moyens mensuels en France*, note de synthèse CEMAGREF, 92 p.
- Sauquet E., Gottschalk L., Krasovskaïa I. (2008) Estimating mean monthly runoff at ungauged locations: an application to France. *Hydrology Research*, 39 (5-6): 403-423.
- Sauquet, E., Ramos, M.H., Chapel, L., Bernardara, P. (2008) Streamflow scaling properties: investigating characteristic scales from different statistical approaches. *Hydrol. Processes*, 22 (17): 3462-3475.
- Tangara, M. (2005) *Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global*. Thèse de doctorat, EPHE, Cemagref, Paris, 374 p.
- Thirel, G., (2009) *Amélioration des prévisions d'ensemble des débits sur la France de SAFRAN-ISBA-MODCOU*. Thèse de doctorat, Université Toulouse III, CNRM/GAME Météo-France, 252 p.

- Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., Regimbeau, F., Habets, F. (2010a) A past discharges assimilation system for ensemble streamflow forecasts over France - Part 1: Description and validation of the assimilation system. *Hydrology and Earth System Sciences*, 14, 1623-1637.
- Thirel, G., Rousset-Regimbeau, F., Martin, E., Habets, F. (2008) On the impact of short-range meteorological forecasts for ensemble streamflow prediction. *J. of Hydrometeorology*, 9: 1301-1317.
- Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., Regimbeau, F., Habets, F. (2010b) A past discharge assimilation system for ensemble streamflow forecasts over France - Part 2: Impact on the ensemble streamflow forecasts. *Hydrology and Earth System Sciences*, 14, 1639-1653.
- Valéry A., (2010) *Modélisation précipitations-débit sous influence nivale, élaboration d'un module neige et évaluation sur 380 bassins versants*. Thèse de doctorat, AgroParisTech, Cemagref, 313 p.
- Van Pham, T. (2011) *Tracking the uncertainty in streamflow prediction through a hydrological forecasting system*. MSc Thesis, University of Twente, Cemagref, the Netherlands (en cours).
- Velázquez, J.A., Anctil, F., Ramos, M.H., Perrin, C. (2011) Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, 29: 33-42.
- Weeink, W. H. A. (2010) *Thresholds for flood forecasting and warning*. MSc Thesis, University of Twente, Cemagref, the Netherlands, 88p.

### **Remerciements**

*Nous remercions Météo-France pour la mise à disposition des prévisions d'ensemble PEARP et, en particulier G. Thirel et E. Martin, pour avoir accompagné nos travaux et nous avoir fourni les données des débits prévus de la chaîne SIM. Le SCHAPI est également remercié pour son soutien financier et pour l'intérêt porté à nos travaux en prévision hydrologique d'ensemble.*

## Résumé

Ce rapport s'inscrit dans l'action "Prévision des inondations", menée dans le cadre de la convention SCHAPI-Cemagref, sous le volet "Développement du système d'aide à la vigilance hydrologique (AVI)". Le but général des travaux présentés est d'évaluer l'apport d'une prévision météorologique à scénarios multiples (prévision d'ensemble) à la prévision hydrologique de débits, pour mieux anticiper les événements de crues. Le modèle hydrologique GRP (Cemagref) et la chaîne de modélisation SIM (Météo-France) ont été alimentés par une même entrée de précipitation prévue, la prévision d'ensemble PEARP de Météo-France. La qualité des prévisions hydrologiques d'ensemble a été évaluée sur un large échantillon de bassins versants en France. Ce rapport présente une synthèse des résultats obtenus et fournit une vision intégrée des travaux réalisés en utilisant les prévisions météorologiques de la PEARP en hydrologie. Plusieurs aspects concernant l'usage et l'évaluation de la qualité des prévisions hydrologiques d'ensemble sont abordés dans les conclusions de ce rapport.



Direction générale  
1, rue Pierre-Gilles de Gennes  
92761 Antony Cedex France  
Tél. 01 40 96 61 21  
Fax 01 40 96 62 25  
[www.cemagref.fr](http://www.cemagref.fr)