

**PRÉVISIONS D'ENSEMBLE HYDROLOGIQUES AU SEIN DU RÉSEAU
NATIONAL DE PRÉVISION DES CRUES EN FRANCE « VIGICRUES » :
EXPÉRIMENTATIONS AVEC LE POST-TRAITEMENT DES
PRÉVISIONS DANS LE BASSIN DE L'ODET**

*Hydrologic ensemble forecasts at the French National Service for Flood
Forecasting « Vigicrues »: experiments with forecast post-processing in
the Odet catchment*

Auteurs : TIBERI-WADIER Anne-Laure^{1*}, BELLEUDY Anne², LE PAPE Etienne³, RAMOS Maria-Helena⁴, RICCI Sophie⁵, GOUTAL Nicole⁶

* *auteur correspondant*

¹Cerema Risques, Eaux et Mer, 155 rue Pierre Bouguer, 29280 PLOUZANE, France, anne-laure.tiberi-wadier@cerema.fr

²SCHAPI (MTES/DGPR/SRNH), 42 avenue Gaspard Coriolis, 31057 TOULOUSE, France, anne.belleudy@developpement-durable.gouv.fr

³SCHAPI (MTES/DGPR/SRNH), 42 avenue Gaspard Coriolis, 31057 TOULOUSE, France, Etienne.LEPAPE@developpement-durable.gouv.fr

⁴Université Paris-Saclay, INRAE, UR HYCAR, 92160 ANTONY, France, maria-helena.ramos@inrae.fr

⁵CECI, CERFACS/CNRS, 42 avenue Gaspard Coriolis, 31057 TOULOUSE, France, ricci@cerfacs.fr

⁶EDF/LNHE/Laboratoire d'Hydraulique Saint-Venant, 6 quai Watier, 78401 CHATOU, France, nicole.goutal@edf.fr

Résumé : L'étude des incertitudes de prévision est aujourd'hui largement développée pour les modèles météorologiques et hydrologiques. La réalisation de prévisions probabilistes fiables permet d'intégrer et de quantifier les incertitudes de prévision. L'étude présentée porte sur la mise en place et l'évaluation de prévisions d'ensemble hydrologiques (une prévision avec plusieurs membres qui décrivent les évolutions futures possibles de la variable à prévoir) sur trois sous-bassins du bassin versant de l'Odet situé dans le Finistère, en s'appuyant sur la méthode des forêts aléatoires « *Quantile Regression Forests* » (QRF). Le modèle hydrologique utilisé est le modèle non distribué GRP développé pour la prévision des crues par INRAE en collaboration avec le SCHAPI. Des prévisions d'ensemble brutes sont d'abord générées par perturbation des paramètres de calage du modèle GRP, puis un post-traitement statistique leur est appliqué avec la méthode QRF. L'article décrit la méthodologie mise en place ainsi que les résultats obtenus pour six configurations d'expériences, qui diffèrent par le type de pluie utilisée pour l'apprentissage et la validation (mesure ou prévision) et par la configuration du code hydrologique GRP (avec et sans assimilation de donnée). Il est mis en évidence que la mise en place du post-traitement permet d'utiliser le code GRP sans son module d'assimilation de donnée. Des tests sont également réalisés dans un objectif de passage vers l'opérationnel. Ils montrent que l'apprentissage de la méthode de post-traitement peut être réalisé avec la pluie mesurée et appliqué à la prévision de pluie (*transfer learning*), qu'il est possible de limiter le nombre maximal d'arbres à 30 pour gagner en temps de calcul et qu'il est possible d'appliquer la méthode QRF avec une prévision déterministe plutôt qu'une prévision d'ensemble.

Mots-clefs : prévisions d’ensemble hydrologiques, Quantile Regression Forests, modèle GRP, calibration statistique, bassin versant de l’Odet

Abstract : Forecast uncertainties are often provided together with meteorological or hydrologic forecasts. The production of reliable probabilistic forecasts allows the forecasters to quantify forecast uncertainties. This study investigates the implementation and evaluation of hydrologic ensemble forecasts (a forecast with several members that describe the possible future evolution of the variable to be forecasted) on three sub-basins of the Odet catchment located in Finistère (Brittany, France), based on the Quantile Regression Forests (QRF) method. The hydrologic model used is the lumped GRP model, developed by INRAE for operational flood forecasting in collaboration with the national service SCHAPI. Raw ensemble forecasts are first generated by perturbing the calibration parameters of the GRP model. A statistical post-processing is then applied with the QRF method. The paper describes the methodology and the results obtained for six experimental configurations, which differ in the type of rainfall data used for training and validation (measurement or forecast) and in the configuration of the GRP model (with and without data assimilation). The results highlight that the GRP model can be used without data assimilation when implementing the QRF method. Tests were also carried out with the aim to move towards the operational use of the post-processing method. They show that learning can be carried out with measured rainfall and applied to forecast rainfall (transfer learning), that it is possible to limit the maximum number of trees to 30 to limit the computing time and that it is possible to apply the QRF method with a deterministic forecast rather than an ensemble forecast.

Keywords: hydrologic ensemble forecast, Quantile Regression Forests, GRP, statistical calibration, Odet catchment

1. INTRODUCTION

L’étude des incertitudes de prévision est aujourd’hui largement développée pour les modèles météorologiques et hydrologiques. Les principales incertitudes des prévisions hydrologiques ont deux origines : (i) les incertitudes sur le forçage météorologique et (ii) les incertitudes hydrologiques, qui regroupent les conditions initiales, la structure du modèle et ses paramètres. La réalisation de prévisions probabilistes fiables permet d’intégrer et de quantifier les incertitudes de prévision. La construction de prévisions d’ensemble (une prévision avec plusieurs membres qui décrivent les évolutions futures possibles de la variable à prévoir) est une manière parmi d’autres de générer des prévisions probabilistes. L’étude présentée porte sur la mise en place et l’évaluation de prévisions d’ensemble hydrologiques sur trois sous-bassins du bassin versant de l’Odet situé dans le Finistère, dont les exutoires se situent à Tréodet, Kerjean et Ty-Planche (Tiberi-Wadier, 2021 ; Tiberi-Wadier et al, 2021) (voir figure 1), en s’appuyant sur la méthode de calibration statistique « *Quantile Regression Forests* » (Meinshausen, 2006). Dans un souci de concision, seuls les résultats obtenus à la station de Tréodet sont présentés ici. L’objectif général est de contribuer aux travaux actuels ou passés visant à mieux comprendre et mettre en place des prévisions d’ensemble hydrologiques au sein du réseau national de prévision de crues « Vigicrues ».

La section 2 présente le principe de la vérification (évaluation de la qualité) des prévisions d’ensemble et les scores utilisés pour évaluer les performances des ensembles construits. La section 3 présente les caractéristiques du modèle hydrologique GRP utilisé dans l’étude. La section 4 s’attache à la description

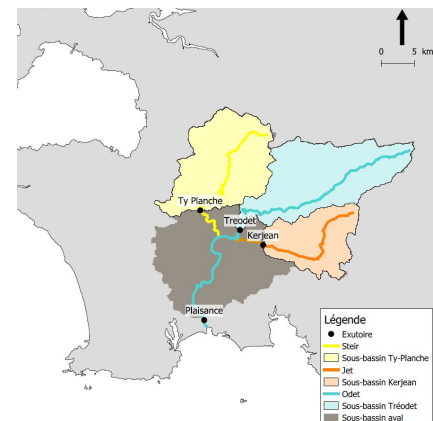


Figure 1 : Bassin versant de l’Odet

de la méthode de calibration statistique utilisée (*Quantile Regression Forests*). La section 5 présente les six configurations d’expériences mises en œuvre. La section 6 décrit les performances des ensembles hydrologiques bruts, et la section 7 présente celles des ensembles calibrés. Quelques recommandations pour la mise en place opérationnelle des méthodes développées sont également proposées dans cette section. Enfin, la section 8 présente des perspectives suite à cette étude.

2. LA VÉRIFICATION DES ENSEMBLES

Les deux caractéristiques souvent considérées comme étant les plus importantes d’une prévision probabiliste sont sa fiabilité et sa résolution (Jolliffe and Stephenson, 2003). Plusieurs critères numériques (ou scores) existent dans la littérature pour mesurer ces attributs de qualité. Dans cette étude, la qualité des prévisions d’ensemble est évaluée par l’utilisation des scores suivants (Tiberi-Wadier, 2021 ; Tiberi-Wadier et al, 2018) (voir figure 2) :

- Le CRPS (*Continuous Ranked Probability Score*) (Gneiting and Raftery, 2007) permet de juger conjointement de la fiabilité et de la résolution d’une prévision d’ensemble. C’est un score positif que l’on cherche à minimiser.
- Le calcul de la norme des histogrammes de rang (Talagrand et al., 1997) permet d’évaluer la fiabilité des ensembles. La valeur de la norme doit être la plus petite possible.
- La courbe ROC (*Relative Operating Characteristic*) permet d’analyser la capacité d’un système à discriminer entre l’occurrence ou la non-occurrence d’un événement donné. Ici, l’événement est le dépassement du quantile 90 de débit. Ce seuil permet de se focaliser sur les périodes de pluie. Il reste cependant suffisamment bas pour que le calcul des scores statistiques soit pertinent. L’aire sous la courbe ROC (AUC) doit être proche de 1 pour une bonne prévision.

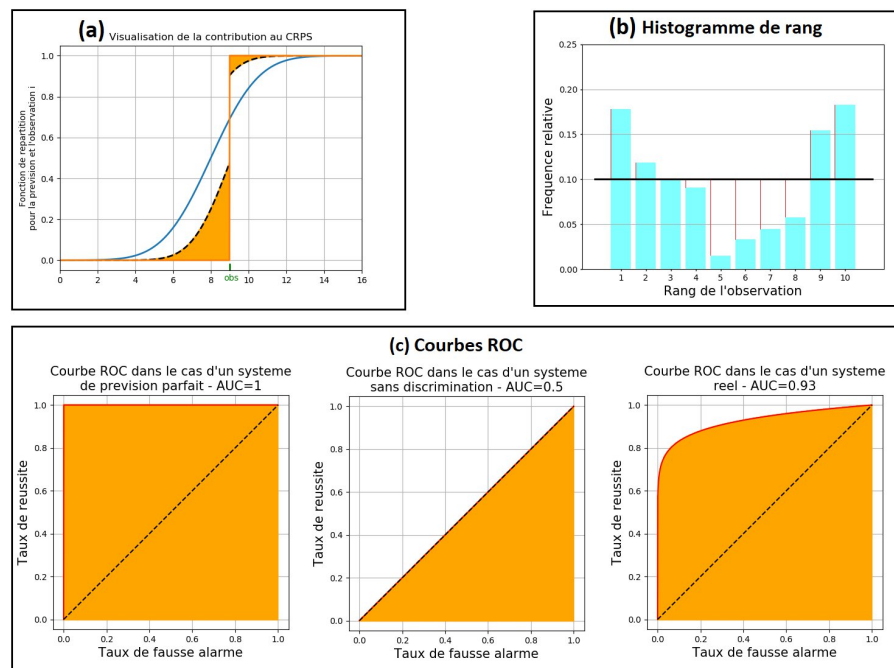


Figure 2 : Illustration des trois scores utilisés (a) CRPS : visualisation graphique de la contribution d’un couple (prévision d’ensemble, observation) au calcul du CRPS, (b) : Histogramme de rang et sa norme : les traits rouges correspondent à la contribution de chaque rang à la norme de l’histogramme, (c) Courbes ROC : illustration des courbes ROC pour différents systèmes de prévision : discrimination parfaite ($AUC=1$, gauche), aucune discrimination ($AUC=0.5$, milieu) et cas réaliste ($AUC=0.93$, droite).

3. MODÈLE HYDROLOGIQUE GRP ET DONNÉES UTILISÉES

Le modèle hydrologique utilisé est le modèle GRP (Viatgé et al., 2019a ; Viatgé et al., 2019b), largement utilisé par les services de prévision des crues au sein du réseau national de vigilance « Vigicrues ». Ce modèle est forcé par la pluie horaire et par l'évapotranspiration potentielle, estimée à partir de la température par la formule d'Oudin (Oudin, 2004 ; Oudin et al., 2005). Il est composé d'un module de production et d'un module de routage (figure 3). Il a trois paramètres libres à caler : $Corr$ (sans unité), T_B (heures) et $Rout$ (mm). Dans le module de production, la quantité totale d'eau allant alimenter l'écoulement est corrigée par le facteur $Corr$. Le module de routage permet de transférer la pluie efficace par l'intermédiaire de l'hydrogramme unitaire de temps de base T_B et d'un réservoir de routage de capacité $Rout$.

Le modèle opérationnel inclut une étape d'assimilation du dernier débit observé (Berthet, 2010) qui corrige le contenu en eau du réservoir de transfert, ainsi qu'une exploitation de l'erreur du modèle (Tangara, 2005). Ces deux étapes sont résumées sous le nom d'« assimilation de données » par la suite. Dans l'étude conduite ici, le modèle GRP est utilisé selon deux configurations : avec cette assimilation de données (ici nommé PRÉVISION) et sans (ici nommé SIMULATION).

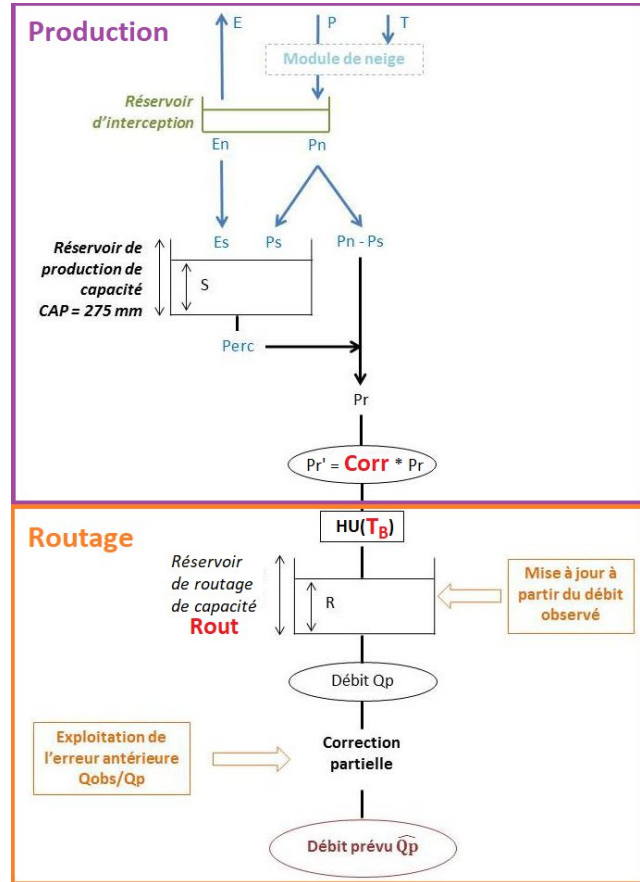


Figure 3 : Schéma fonctionnel du modèle hydrologique GRP. D'après INRAE [2020]

Le modèle a été préalablement calé sur les bassins de l'Odette (Tiberi-Wadier, 2021) sur des sous-périodes de la période 2007-2020 (voir paragraphe 6). Les données utilisées pour l'étude sont de natures diverses : (i) la pluie ANTILOPE J+1 : lame d'eau au pas de temps horaire qui combine la pluie radar et les relevés des pluviomètres, (ii) les prévisions de pluie RR3 : prévisions déterministes expertisées (Symposium 2) au pas de temps tri-horaire et (iii) les débits observés issus de la banque de données Hydroportail aux stations Tréodet, Kerjean et Ty-Planche au pas de temps horaire.

4. LA CALIBRATION STATISTIQUE « QUANTILE REGRESSION FORESTS »

4.1. La méthode QRF

Les approches de prévision d'ensemble ont pour objectif de représenter les incertitudes de prévision. Cependant, celles-ci sont nombreuses et de ce fait rarement complètement couvertes par les modèles et techniques mises en place pour générer une prévision d'ensemble. Le post-traitement des ensembles générés (ensembles « bruts »), appelé encore « calibration statistique », est donc souvent indispensable avant une utilisation opérationnelle d'une prévision d'ensemble.

Dans cette étude, le post-traitement des prévisions est réalisé grâce à la méthode *Quantile Regression Forests*, appelée encore « méthode QRF ». C’est une méthode d’apprentissage statistique. Elle est détaillée dans Taillardat et al. [2019], Zamo et al. [2014]. Le principe de la méthode QRF est l’agrégation des observations selon leur prévision, cette dernière étant résumée par différents prédicteurs. Elle consiste en la construction d’une forêt aléatoire constituée d’un ensemble d’arbres de décisions binaires, chacun de ces arbres étant construit à partir d’un ensemble de prédicteurs. Les prédicteurs peuvent être naturellement des statistiques relatives à l’ensemble brut à calibrer (ici, les débits horaires), mais peuvent également être indépendants des caractéristiques de l’ensemble brut. Dans cette étude, les prédicteurs utilisés sont la moyenne et l’écart-type de l’ensemble brut, le mois et le moment de la journée de la prévision à calibrer, le débit et le gradient du débit mesurés à la date de référence de la prévision $Date_{Ref}^1$, la pluie mesurée sur le bassin versant avant $Date_{Ref}$ (période $[-(24\text{-échéeance calibrée}) \text{ heures}, Date_{Ref}]$) et la pluie prévue (période $[Date_{Ref} + 1 \text{ heure}, \text{échéeance}]$).

Pour la construction d’un arbre, on scinde de manière itérative l’échantillon d’apprentissage en deux groupes. Le partage est réalisé selon la valeur d’un seuil sur un des prédicteurs pour les prédicteurs quantitatifs, ou selon leur modalité pour les prédicteurs qualitatifs. Le prédicteur et le seuil sont choisis de manière à maximiser l’homogénéité des valeurs correspondantes de l’observation dans les deux sous-groupes créés. Chaque sous-groupe est alors redivisé en deux, jusqu’à l’atteinte d’un critère d’arrêt : nombre minimal de données ou décroissance insignifiante de la variance dans les sous-groupes résultants. Chaque sous-groupe final est appelé « feuille » : une feuille contient un ensemble d’observations. La figure 4 illustre cette étape avec deux prédicteurs $p1$ et $p2$ dont les valeurs sont comprises entre 0 et 1. À chaque point est associée une observation de débit, les prédicteurs associés à cette observation étant $p1$ et $p2$. L’arbre construit dans cet exemple contient trois feuilles constituées des valeurs des observations correspondantes.

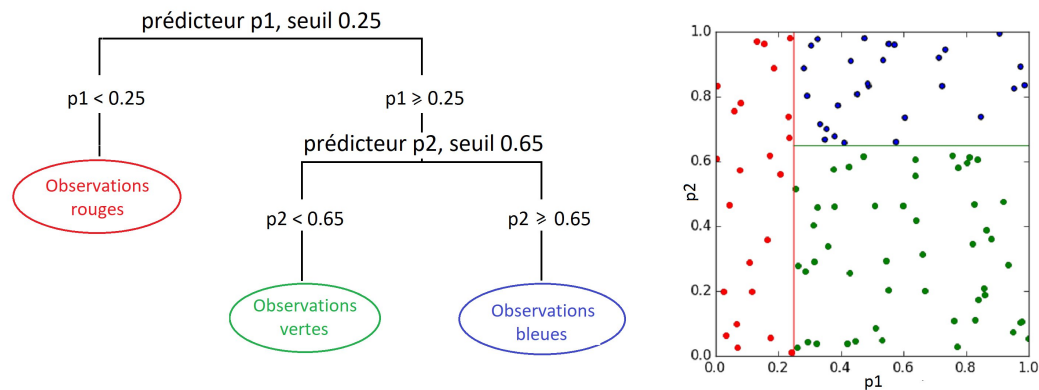


Figure 4 : Illustration d’un arbre de décision binaire construit avec deux prédicteurs $p1$ et $p2$

4.2. Périodes d’apprentissage et de validation

La période sur laquelle est optimisée la forêt aléatoire est appelée période d’apprentissage. Il s’agit de la première étape de l’étude. Dans un second temps, il est nécessaire d’évaluer la performance de cette forêt. On travaille alors sur une période de validation indépendante de la période d’apprentissage. Dans l’étude, les périodes d’apprentissage et de validation sont définies comme suit :

- Les mois d’été (juillet, août et septembre) sont exclus, cela est dû à la faible occurrence de pluies/crues pendant cette période sur le bassin de l’Odet.
- L’apprentissage est réalisé sur une période de 3 ans, du 1^{er} octobre 2011 au 30 juin 2014.

¹ date du temps présent au moment de la production de la prévision, notée par la suite $Date_{Ref}$

- La validation est réalisée sur une période de durée totale de 7 ans, composée de deux sous-périodes, du 1er janvier 2010 au 30 juin 2011 et du 1er octobre 2014 au 31 décembre 2019.

5. SIX EXPÉRIENCES DIFFÉRENTES

Trois configurations pour l'apprentissage et la validation sont étudiées :

- **Configuration $A_{ANT}V_{ANT}$** : le modèle GRP est forcé par la pluie ANTILOPE sur les périodes d'apprentissage et de validation. Cette configuration, non opérationnelle, est testée pour évaluer quels sont les scores de référence avec une pluie « parfaite » (c'est-à-dire identique à l'observation historique) utilisée en prévision.
- **Configuration $A_{RR3}V_{RR3}$** : le modèle GRP est forcé par la pluie prévue RR3 sur les périodes d'apprentissage et de validation. C'est la meilleure configuration opérationnelle que l'on peut mettre en place car la nature de l'erreur reste la même entre apprentissage et validation.
- **Configuration $A_{ANT}V_{RR3}$** : le modèle GRP est forcé par la pluie ANTILOPE sur la période d'apprentissage, et par la pluie prévue RR3 sur la période de validation. Ici, la nature de l'erreur n'est pas la même sur la période d'apprentissage et sur la période de validation. L'intérêt est d'évaluer si cela induit une dégradation importante par rapport à la meilleure configuration opérationnelle ($A_{RR3}V_{RR3}$), plus complexe à mettre en œuvre.

Pour chacune de ces trois configurations, le code GRP est utilisé en PREVISION (avec assimilation de donnée) et en SIMULATION (sans assimilation de donnée) ce qui fait en tout six expériences différentes évaluées. Dans un objectif opérationnel, seuls les résultats des expériences $A_{RR3}V_{RR3}$ et $A_{ANT}V_{RR3}$ sont présentés dans la suite (rappel : l'expérience $A_{ANT}V_{ANT}$ n'est pas réalisable en opérationnel).

6. CARACTÉRISTIQUES DES ENSEMBLES HYDROLOGIQUES BRUTS

Les ensembles hydrologiques bruts sont générés par perturbation des trois paramètres de calage du modèle GRP. Les bornes de variation des paramètres sont définies par un ensemble de calages successifs du modèle sur des périodes de 2 et 3 ans entre 2007 et 2020. Sur chaque bassin versant de l'étude, l'ensemble brut est constitué de 99 membres. Quatre ensembles bruts ont été générés à chaque station :

- **raw-PRE-ANT** : code GRP en PREVISION, forçage ANTILOPE après $Date_{Ref}$;
- **raw-SIM-ANT** : code GRP en SIMULATION, forçage ANTILOPE après $Date_{Ref}$;
- **raw-PRE-RR3** : code GRP en PREVISION, forçage prévisions RR3 après $Date_{Ref}$;
- **raw-SIM-RR3** : code GRP en SIMULATION, forçage prévisions RR3 après $Date_{Ref}$;

On notera que seuls les ensembles **raw-PRE-RR3** et **raw-SIM-RR3** reproduisent des conditions qui peuvent être retrouvées en opérationnel. Les ensembles **raw-PRE-ANT** et **raw-SIM-ANT** servent de référence. La figure 5 présente les scores par échéance de prévision des quatre ensembles hydrologiques bruts générés à Tréodet sur la période 2010-2019.

L'analyse de ces graphiques met en évidence les éléments suivants :

- Pour chaque score étudié, deux groupes de prévisions d'ensemble présentent des scores d'évolution semblable avec les échéances selon que le code GRP est utilisé en PREVISION ou en SIMULATION (groupes (**raw-PRE-ANT**, **raw-PRE-RR3**) et (**raw-SIM-ANT**, **raw-SIM-RR3**)).
- Pour chacun de ces deux types d'utilisation du modèle GRP, les scores sont analogues aux courtes échéances de prévision (moins de 6-12 heures) que l'on utilise ANTILOPE ou RR3

comme forçage. Cela est probablement lié au temps de réponse du bassin versant. À partir de ces échéances, les scores sont dégradés avec RR3 par rapport à ANTILOPE. Ceci est normal puisque la pluie prévue RR3 est de moins bonne qualité que ANTILOPE.

- Aux courtes échéances, l'utilisation du modèle en PREVISION fournit les meilleurs scores CRPS et AUC. À longue échéance en revanche (au-delà de 12-15h), l'utilisation du modèle en SIMULATION est préférable pour ces scores, et ce notamment pour le cas où l'on utilise la pluie RR3.
- Pour toutes les échéances de prévision, la fiabilité est toujours significativement meilleure lorsqu'on utilise le modèle GRP en SIMULATION (ensembles **raw-SIM-ANT** et **raw-SIM-RR3**).
- Ces conclusions confirment l'intérêt d'appliquer la calibration QRF aux deux types d'utilisation du code GRP afin de mettre en évidence la meilleure stratégie globale à retenir.

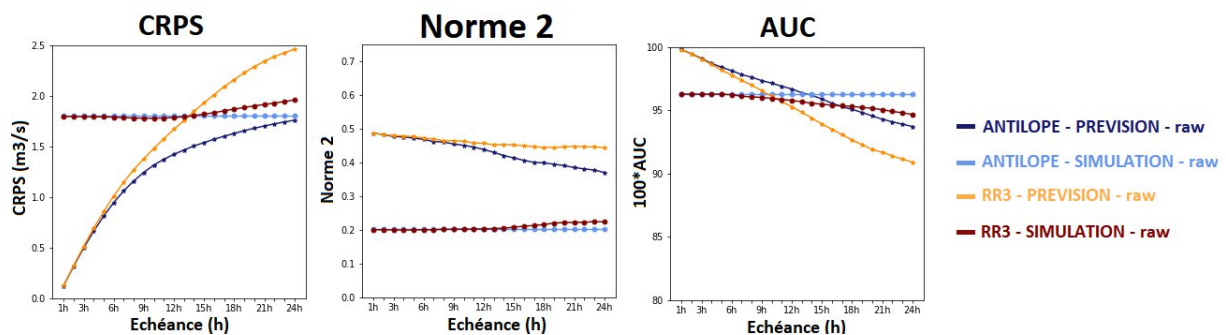


Figure 5 : CRPS (gauche), norme 2 de l'histogramme de rang (milieu) et aire AUC de la courbe ROC (droite) pour les quatre ensembles bruts générés à Tréodet. Échéances de prévision de 1h à 24h. Évaluation sur la période 2010-2019.

7. CARACTÉRISTIQUES DES ENSEMBLES CALIBRES ET RECOMMANDATIONS

7.1. Résultats généraux

La figure 6 présente les scores des ensembles bruts et calibrés à Tréodet pour les expériences d'apprentissage (A) et validation (V) $A_{RR3}V_{RR3}$ et $A_{ANT}V_{RR3}$. Les ensembles sont les suivants :

- **qrf-PRE-RR3** est obtenu par calibration statistique de **raw-PRE-RR3** ;
- **qrf-SIM-RR3** est obtenu par calibration statistique de **raw-SIM-RR3** ;
- **qrf-PRE-ANT/RR3** est construit par calibration statistique de **raw-PRE-RR3** (forçage RR3) par la forêt aléatoire optimisée pour l'ensemble brut **raw-PRE-ANT** (apprentissage ANTILOPE) ;
- **qrf-SIM-ANT/RR3** est construit par calibration statistique de **raw-SIM-RR3** (forçage RR3) par la forêt aléatoire optimisée pour l'ensemble brut **raw-SIM-ANT** (apprentissage ANTILOPE).

On rappelle ici que l'intérêt des expériences **qrf-PRE-ANT/RR3** et **qrf-SIM-ANT/RR3** est de tester la faisabilité opérationnelle de la méthodologie. En effet, la construction *a posteriori* des prévisions d'ensembles **raw-PRE-RR3** et **raw-SIM-RR3** n'est pas aisée puisqu'il faut reconstruire l'historique des prévisions RR3 à chaque date $Date_{Ref}$ de la période d'apprentissage. Réaliser l'apprentissage statistique à partir de **raw-PRE-ANT** et **raw-SIM-ANT** est beaucoup plus simple.

Les conclusions concernant l'expérience $A_{RR3}V_{RR3}$ sont les suivantes :

- la calibration statistique QRF améliore toujours les scores des ensembles bruts ;

- l'ensemble **qrf-SIM-RR3** présente les meilleures valeurs de CRPS après calibration parmi tous les ensembles calibrés ;
 - la fiabilité et la discrimination des ensembles **qrf-SIM-RR3** et **qrf-PRE-RR3** sont analogues.
- Ainsi, pour le bassin versant étudié ici, il est recommandé de construire **un ensemble brut avec le code GRP utilisé en SIMULATION**, afin de garantir une bonne performance sur l'ensemble des critères de qualité de la prévision (CRPS, fiabilité et discrimination).

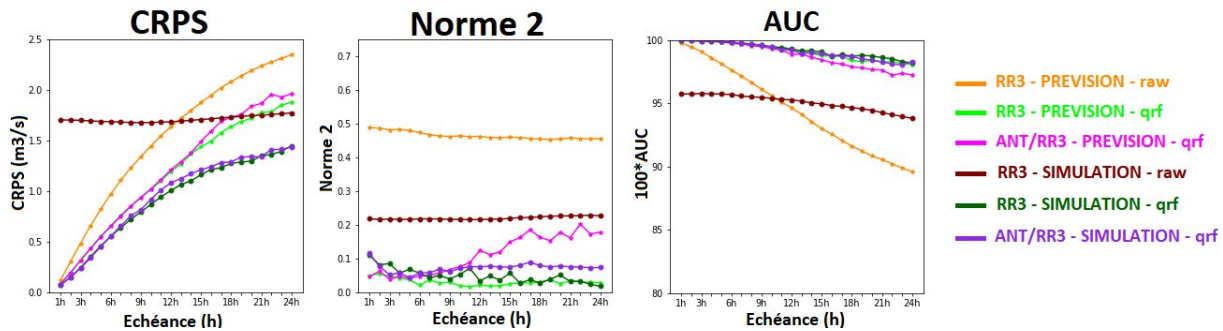


Figure 6 : CRPS (gauche), norme 2 de l'histogramme de rang (milieu) et aire AUC de la courbe ROC (droite) pour les quatre ensembles calibrés à Tréodet. Échéances de prévision de 1h à 24h. Évaluation sur la période de validation.

En ce qui concerne l'expérience $A_{ANT}V_{RR3}$, nous pouvons émettre les conclusions suivantes :

- les performances des ensembles calibrés **qrf-PRE-RR3** et **qrf-PRE-ANT/RR3** (respectivement des ensembles **qrf-SIM-RR3** et **qrf-SIM-ANT/RR3**) sont proches pour le CRPS et AUC ;
- la fiabilité est un peu dégradée dans l'expérience $A_{ANT}V_{RR3}$ (ensembles **qrf-PRE-ANT/RR3** et **qrf-SIM-ANT/RR3**) par rapport à l'expérience $A_{RR3}V_{RR3}$ (ensembles **qrf-PRE-RR3** et **qrf-SIM-RR3**).

Ainsi, l'utilisation d'une forêt aléatoire construite avec un apprentissage (A) avec ANTILOPE (plus facile à mettre en œuvre qu'avec les RR3) dégrade surtout la fiabilité des ensembles calibrés, évalués en validation avec RR3. Les autres scores sont légèrement impactés mais pas de manière notable. **Il semble envisageable d'utiliser cette stratégie pour une application opérationnelle.**

7.2. Tests de simplification en vue d'une application opérationnelle

Limitation du nombre maximal d'arbres dans la forêt. Dans la pratique, la construction d'une forêt aléatoire nécessite la prescription de trois hyper-paramètres : (i) mtry : nombre de prédicteurs tirés aléatoirement parmi l'ensemble des prédicteurs à chaque nœud de chaque arbre, (ii) nodesize : nombre d'éléments dans les feuilles et (iii) ntree : nombre d'arbres dans la forêt. Ces trois hyper-paramètres doivent être optimisés pour chaque bassin, pour chaque échéance de prévision entre 1 heure et 24 heures et pour chaque ensemble brut généré (4 ensembles bruts, voir paragraphe 6). Il y a donc 288 (=24 heures * 4 ensembles * 3 sous-bassins) forêts à optimiser sur la période d'apprentissage.

Pour l'optimisation d'une forêt, on fait varier les valeurs du triplet d'hyper-paramètres entre des bornes définies, et on conserve celui pour lequel les performances de la forêt sont les meilleures. Les résultats présentés précédemment sont issus de forêts dans lesquelles on a autorisé jusqu'à 100 arbres (paramètre ntree) au moment de l'optimisation, ce qui nécessite d'importants moyens de calcul (1 semaine par bassin versant sur le calculateur Datarmor sur lequel il est autorisé de travailler avec 18 nœuds à 56 cœurs). Le temps de calcul des forêts est directement lié au nombre d'arbres à calculer. Il est donc pertinent d'étudier l'impact du nombre maximal d'arbres autorisé dans la forêt sur les performances des forêts construites. Le test réalisé a mis en évidence que les résultats obtenus avec un nombre maximal

d'arbres autorisé compris entre 30 et 100 sont quasiment identiques. En revanche, avec un nombre maximal d'arbres égal à 20, on commence à voir une légère dégradation des scores. Il n'est donc pas pertinent d'augmenter la durée de simulation pour l'optimisation en autorisant jusqu'à 100 arbres. La limitation à un nombre maximal de 30 arbres donne des résultats quasiment équivalents. Le gain de temps pour l'optimisation est environ d'un facteur 3.

Application de la calibration QRF à une prévision déterministe. Une autre limite de la méthode pour un passage vers l'opérationnel est qu'il est fastidieux de construire une prévision d'ensemble brute : il faut réaliser de multiples calages pour déterminer les bornes de variation des paramètres, puis lancer 99 membres sur une période d'apprentissage et sur une période de validation. Or, la calibration statistique réalisée repose uniquement sur 2 prédicteurs issus de l'ensemble brut : la moyenne de l'ensemble et son écart-type. Les autres prédicteurs sont indépendants des valeurs de l'ensemble. Afin de simplifier la méthodologie, le test suivant a été réalisé : on remplace le prédicteur « moyenne de l'ensemble » par la valeur d'une prévision déterministe réalisée avec RR3, et on supprime le prédicteur « écart-type ». Le test a été réalisé uniquement sur la station Tréodet, pour les échéances de prévision 6h, 12h, 18h et 24h. Il a montré qu'il semble *a priori* envisageable de mettre en place une calibration statistique basée sur une prévision déterministe uniquement. Cependant, il faut garder à l'esprit que cette méthodologie simplifiée ne permettrait pas de reconstruire des membres cohérents pour utiliser ces prévisions hydrologiques d'ensemble en forçage d'un modèle hydraulique.

8. CONCLUSIONS ET PERSPECTIVES

Dans cette étude, des prévisions hydrologiques d'ensemble ont été construites à partir du modèle hydrologique GRP (et des incertitudes de ses paramètres) et de la méthode de calibration statistique (post-traitement des sorties du modèle GRP) « *Quantile Regression Forests* » (QRF). Il a été montré que la calibration statistique d'une prévision d'ensemble améliore toujours sa performance. De plus, il est recommandé d'utiliser le code GRP sans l'étape d'assimilation de données avant d'appliquer le post-traitement QRF. Cela est probablement dû au fait que ce post-traitement contient déjà dans sa technique des prédicteurs liés aux variables utilisées dans l'assimilation de donnée de GRP (à savoir, le débit observé et le gradient du débit à l'instant de la prévision).

Dans une optique de passage vers l'opérationnel, plusieurs tests ont été réalisés pour faciliter le transfert méthodologique. Ils ont montré qu'il est possible de réaliser l'apprentissage du post-traitement QRF sur des prévisions hydrologiques réalisées avec en entrée au modèle GRP des pluies observées (dans notre cas, forçage ANTILOPE), bien qu'en prévision opérationnelle en temps réel ce soient des prévisions de pluie qui servent de forçage au modèle (concept de *transfer learning*). Par ailleurs, il est possible de réduire le temps de calcul pour l'optimisation des forêts aléatoires de la méthode QRF en limitant le nombre maximal d'arbres dans la forêt à environ 30 arbres. Enfin, il semble également possible de se passer de l'étape de construction d'une prévision d'ensemble brute, un travail souvent fastidieux quand on ne dispose pas de prévisions météorologiques d'ensemble pour forcer le modèle hydrologique, en ajustant les prédicteurs de la méthode QRF pour travailler uniquement avec une prévision déterministe, sans qu'il y ait une dégradation significative de performance du post-traitement.

Enfin, plusieurs pistes se dégagent afin de poursuivre cette étude. Dans un premier temps, il serait intéressant d'étudier les résultats de l'application de cette méthodologie sur un plus grand nombre de bassins versants aux caractéristiques variées. Il serait intéressant de poursuivre les tests sur l'utilisation d'une prévision déterministe et sur la limitation du nombre d'arbres dans les forêts dans un cadre d'application à un large échantillon national de bassins. Enfin, les résultats obtenus avec la méthode QRF pourraient être comparés avec ceux que l'on peut obtenir avec la méthode automatisée de quantification des incertitudes prédictives OTAMIN (Bourgin, 2014 ; Viatgé et al., 2019a), utilisée

aujourd'hui en opérationnel par les Services de Prévision des Crues pour l'estimation des incertitudes de prévision (<https://webgr.inrae.fr/logiciels/otamin/>).

9. REMERCIEMENTS

Les auteurs remercient INRAE pour la mise à disposition du modèle GRP ainsi que Charles Perrin (INRAE) et Bruno Janet (SCHAPI) pour leurs conseils. Les auteurs remercient également le Pôle de Calcul et de Données Marines (PCDM) pour fournir DATARMOR pour les ressources de calcul et le stockage.

10. REFERENCES

- Berthet, L., 2010. Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique (Theses). AgroParisTech. <https://hal.inrae.fr/tel-02594604>
- Bourgin, F., 2014. Comment quantifier l'incertitude prédictive en modélisation hydrologique ? : Travail exploratoire sur un grand échantillon de bassins versants (PhD Thesis). AgroParisTech. <https://pastel.hal.science/tel-01130084v2>
- Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association 102, 359–378. <https://doi.org/10.1198/016214506000001437>
- INRAE, 2020. Calage et application opérationnelle du modèle de prévision de crue GRP - Manuel d'utilisation. INRAE.
- Jolliffe, I.T., Stephenson, D.B., 2003. Forecast verification: a practitioner's guide in atmospheric science. Wiley, Chichester.
- Meinshausen, N., 2006. Quantile Regression Forests. Journal of Machine Learning Research 7, 17. <https://www.jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf>
- Oudin, L., 2004. Recherche d'un modèle d'évapotranspiration potentielle pertinent comme entrée d'un modèle pluie-débit global (PhD Thesis). <https://theses.hal.science/pastel-00000931/>
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Journal of Hydrology 303, 290–306. <https://doi.org/10.1016/j.jhydrol.2004.08.026>
- Taillardat, M., Fougères, A.-L., Naveau, P., Mestre, O., 2019. Forest-Based and Semiparametric Methods for the Postprocessing of Rainfall Ensemble Forecasting. Weather and Forecasting 34, 617–634. <https://doi.org/10.1175/WAF-D-18-0149.1>
- Talagrand, O., Vautard, R., Strauss, B., 1997. Evaluation of probabilistic prediction systems, in: Workshop on Predictability, 20-22 October 1997. ECMWF, Shinfield Park, Reading, pp. 1–26. <https://www.ecmwf.int/en/elibrary/76596-evaluation-probabilistic-prediction-systems>
- Tangara, M., 2005. Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global (PhD Thesis). <https://webgr.inrae.fr/wp-content/uploads/2012/07/2005-TANGARA-THESE.pdf>
- Tiberi-Wadier, A.-L., 2021. Prévisions d'ensemble hydrologiques et hydrauliques pour la vigilance crues (PhD Thesis). <https://www.theses.fr/2021PESC2031>
- Tiberi-Wadier, A.-L., Goutal, N., Ricci, S., Sergent, P., Taillardat, M., Bouttier, F., Monteil, C., 2021. Strategies for hydrologic ensemble generation and calibration: on the merits of using model-based predictors. Journal of Hydrology 126233. <https://doi.org/10.1016/j.jhydrol.2021.126233>
- Tiberi-Wadier, A.-L., Taillardat, M., Goutal, N., Ricci, S., Sergent, P., Bouttier, F., Le Pape, E., 2018. Vers une approche ensembliste de la prévision des crues, in: De La Prévision Des Crues à La Gestion de Crise. SHF, Avignon, France. <https://hal.science/hal-03250002>
- Viatgé, J., Berthet, L., Marty, R., Bourgin, F., Piotte, O., Ramos, M.-H., Perrin, C., 2019a. Vers une production en temps réel d'intervalles prédictifs associés aux prévisions de crue dans Vigicrues en France. La Houille Blanche 105, 63–71. <https://doi.org/10.1051/lhb/2019016>
- Viatgé, J., Pinna, T., Ficchi, A., Perrin, C., Dorchie, D., Garandeau, L., Tilmant, F., 2019b. Vers une plus grande flexibilité temporelle du modèle opérationnel de prévision des crues GRP. La Houille Blanche 105, 72–80. <https://doi.org/10.1051/lhb/2019017>
- Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. Solar Energy 105, 804–816. <https://doi.org/10.1016/j.solener.2014.03.026>