



HAL
open science

Intégration des pluies prévues du système AROME-PI dans le système AIGA

J. Demargne, P. Javelle

► **To cite this version:**

J. Demargne, P. Javelle. Intégration des pluies prévues du système AROME-PI dans le système AIGA. [Rapport de recherche] irstea. 2015, pp.38. hal-02606077

HAL Id: hal-02606077

<https://hal.inrae.fr/hal-02606077>

Submitted on 16 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.



PROGRAMME MEDDE (DGPR / SRNH) - IRSTEA 2015
Connaissance et prévention des risques naturels et hydrauliques

*Ces travaux bénéficient du soutien du Ministère chargé de l'Ecologie
par décision de subvention n°2101527675*

Action « prévision des inondations »

AXE III : « Avertissement crues soudaines en site non jaugé »

**A. Recherches et développements : Amélioration du produit
« Crues Soudaines »**

Intégration des pluies prévues

du système AROME-PI

dans le système AIGA

Auteur : Julie Demargne (HYDRIS-Hydrologie)

Suivi : Pierre Javelle (IRSTEA)

Janvier 2016

HYDRIS
hydrologie

Table des Matières

Résumé.....	3
1. Objectifs.....	6
a) Développements du système d'anticipation AIGA.....	6
b) Prévisions immédiates de pluie AROME-PI.....	9
c) Objectifs de l'étude	9
2. Données et Modélisation Hydrologique	10
a) Bassins versants	10
b) Modélisation hydrologique et données d'observations.....	11
c) Sélection des évènements	13
3. Méthodologies	14
a) Prévisions de pluie AROME-PI et ensembles timelag	14
b) Prévisions de débit et avertissements AIGA	16
c) Evaluation de la qualité des avertissements.....	17
d) Vérification des prévisions d'ensemble	19
4. Résultats.....	21
a) Qualité des prévisions d'ensembles de pluie de bassin	21
b) Qualité des prévisions d'ensembles de débit.....	25
c) Qualité des avertissements probabilisés.....	29
5. Conclusions et Perspectives.....	32
6. Références	35
Appendice : Liste des Figures.....	37

Résumé

Cette étude a été réalisée pour le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) pour analyser l'amélioration du système d'avertissement sur la possibilité de crues rapides AIGA en intégrant des prévisions de pluie future à courte échéance et fine résolution spatio-temporelle et en prenant en compte leurs incertitudes. Le SCHAPI développe actuellement un service d'avertissement automatisé valable sur l'ensemble de la France, qui sera complété par un service expertisé d'avertissement sur des bassins dits prioritaires. Ce service expertisé concerne les bassins avec des temps de réaction courts compris entre 2h et 6h et avec des enjeux particuliers.

Pour améliorer la caractérisation et l'anticipation des crues soudaines, l'objectif est d'intégrer des prévisions immédiates (appelées nowcasts) de courte échéance et à haute résolution spatio-temporelle, notamment les prévisions du modèle AROME de Météo-France. Les prévisions de pluie mises à disposition par Météo-France sont les prévisions immédiates du modèle AROME, dites AROME-PI, de type déterministe, au pas de temps de 15 minutes pour une échéance de 6 heures, à la résolution de 2,5 km, réactualisées toutes les heures (Auger et al. 2015). La fine résolution d'AROME permet de modéliser les processus convectifs, qui sont essentiels pour les alertes aux crues rapides. Les prévisions déterministes successives sont utilisées pour produire des prévisions d'ensemble au pas de temps horaire par décalage temporel, appelées AROMEPI-timelag, avec 6 membres (les prévisions d'AROME-PI plus anciennes étant complétées par la pluie nulle pour les plus longues échéances non couvertes par AROME-PI). L'utilisation de prévisions d'ensemble de pluie et débit a permis de définir des prévisions probabilistes en fonction du nombre de membres dépassant le seuil considéré.

L'analyse de l'apport des prévisions de pluie immédiate a porté sur 3 événements de crue de septembre 2014, mi-novembre 2014 et fin novembre 2014. Les simulations et prévisions de débit sont issues du modèle hydrologique semi-distribué dit opérationnel pour le futur système AIGA, avec les paramètres calibrés et régionalisés de la fin 2014 (Organde 2014). L'analyse a concerné les 185 bassins versants jaugés de surface entre 10 et 992 km² dans la zone Grand Sud, zone touchée par les 3 événements étudiés. Pour analyser l'apport des prévisions déterministes AROME-PI et prévisions d'ensemble AROMEPI-timelag, différentes prévisions déterministes de référence sont également évaluées : la prévision de pluie persistante (en maintenant sur toute l'échéance de prévision, la dernière pluie observée au moment de l'initialisation de la prévision), la prévision de pluie « parfaite » (i.e. pluie observée) pour montrer l'impact des erreurs et des incertitudes des prévisions d'ensemble de pluie, et la prévision de pluie nulle.

Comme pour l'étude portant sur la mise à jour du paramétrage de la modélisation hydrologique d'Organde en 2015, l'évaluation de la qualité des alertes s'est basée sur le dépassement d'un seuil de débit relatif correspondant à une période de retour de l'ordre de 2 ans. Cette approche permet de « débiaiser » les sorties du modèle hydrologique. L'évaluation des avertissements a inclus différents critères de contingence pour décrire la fiabilité, la précision et le biais en fréquence des dépassements prévus de seuil de débit, ainsi que l'anticipation des dépassements. Par ailleurs, les prévisions de pluie et de débit ont été évaluées avec le système de vérification EVS (Brown et al. 2010) en sélectionnant des critères pour décrire le biais relatif et la corrélation des moyennes des ensembles avec les observations, ainsi que le gain en termes d'erreur en probabilité et de discrimination pour les dépassements de quantiles. L'évaluation des prévisions a concerné les 56 bassins les plus touchés par les 3 événements. Comme la chaîne d'avertissement testée dans cette étude est

basée sur le dépassement d'un seuil relatif de débit, les prévisions de débit ont été comparées aux débits simulés.

Les résultats de l'évaluation des prévisions de pluie de bassin ont permis d'identifier le biais conditionnel des moyennes des ensembles de pluie, qui ont tendance à surestimer les faibles pluies et à significativement sous-estimer les fortes pluies. Mais les ensembles montrent une bonne capacité à discriminer les dépassements des seuils de quantiles observés 80% (issus de la chronique de pluie observée sur l'année 2014), notamment pour les 4 premières heures d'échéance. Pour la comparaison des prévisions de débit avec les débits simulés, l'évaluation a montré un biais négatif des ensembles AROMEPI-timelag pour les échéances les plus longues dû aux sous-estimations des pluies AROME-PI et à l'intégration des pluies nulles pour certains des membres de l'ensemble. Cependant, on note également des erreurs de surestimation pour les débits moyens et certains débits forts. En comparaison avec la prévision probabiliste climatologique (pour toutes les échéances) et la prévision persistante (pour les échéances de plus 2 heures), les ensembles ont de meilleures performances pour la détection des dépassements du quantile de débit simulé de 80%. Les résultats sont relativement similaires entre la prévision persistante et la prévision AROMEPI-timelag pour la première heure d'échéance de prévision. Du fait de l'échantillon réduit des prévisions pour les 3 événements étudiés, l'analyse des dépassements prévus concerne les quantiles simulés de débit de 80% (estimés à partir des simulations de l'année 2014), qui ont des valeurs inférieures aux valeurs de seuil relatif de débit utilisées pour générer les avertissements dans le système AIGA.

L'évaluation des avertissements prévus avec AROME-PI en comparaison au système AIGA actuel (qui n'intègre pas de prévision de pluie) a inclus 524 dépassements et non-dépassements sur les 185 bassins versants étudiés et les 3 événements. La prévision déterministe AROME-PI et la prévision d'ensemble AROMEPI-timelag n'utilisant qu'un seul membre pour déclencher l'avertissement de dépassement obtiennent des scores de contingence très similaires. Elles permettent une amélioration des avertissements en termes de détection (gain de l'ordre de 20%) mais avec une légère dégradation du taux de fausses alertes et donc une perte de fiabilité (perte de l'ordre de 13%). L'utilisation des prévisions d'ensemble AROMEPI-timelag prenant en compte 3 ou 6 membres de l'ensemble pour déclencher les avertissements permet de contrebalancer cette tendance à générer des fausses alertes plutôt que des alertes manquées ; l'utilisation des prévisions AROME-PI et AROMEPI-timelag sur une échéance de 3 heures permet également d'obtenir une légère amélioration. Le taux de détection obtenu est supérieur à 60%, mais la fiabilité reste légèrement inférieure à celle du système AIGA actuel (perte de l'ordre de 5%). En termes d'anticipation, les gains avec la prévision déterministe AROME-PI et la prévision d'ensemble AROMEPI-timelag pour l'échéance de 6 heures sont très significatifs : gain moyen en anticipation de 5 heures pour les 64 dépassements simulés par AIGA et de 6 heures pour les 37 dépassements observés.

En conclusion, cette étude montre l'intérêt d'inclure, dans la modélisation hydrologique du système AIGA, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag, pour améliorer significativement l'anticipation des détections de dépassement de seuil vis-à-vis des détections du système actuel sans pluie future. L'approche pragmatique de construction de prévisions d'ensemble en utilisant les prévisions déterministes successives a permis de définir des avertissements avec différents niveaux de probabilité afin de moduler le niveau de probabilité déclenchant l'avertissement. Cependant, il est nécessaire d'analyser le taux de détection et la fiabilité des dépassements prévus en fonction de l'échéance de prévision et du niveau de probabilité pour définir les dépassements avec les prévisions d'ensemble, afin de gagner en anticipation tout en conservant une fiabilité

comparable au système AIGA actuel. Ce type d'optimisation de la qualité des alertes, pour une meilleure détection des dépassements avec une anticipation améliorée et une fiabilité acceptable, doit également prendre en compte les enjeux pour l'ensemble de la chaîne d'avertissement et les besoins des utilisateurs.

Des pistes d'amélioration pour la combinaison des prévisions immédiates avec des prévisions à plus longue échéance, la prise en compte des incertitudes de source hydrologique, l'amélioration des conditions initiales du modèle hydrologique, et l'utilisation des informations post-crise de dégâts notamment pour les bassins non jaugés, sont également discutées en conclusion.

Le présent rapport constitue le livrable 00050 de la commande n°4500261231 du 15/04/2015 du marché « Mise en œuvre de la méthode AIGA » n°2014-01-16, conclu entre IRSTEA et Hydris Hydrologie.

1. Objectifs

Cette étude a pour objectif d'étudier l'apport des prévisions immédiates de pluie du système AROME pour le système d'avertissement sur la possibilité de crues soudaines AIGA. Ce travail est réalisé pour le SCHAPI qui développe actuellement un nouveau service d'avertissements expertisés sur des bassins dits prioritaires, dont le temps de réaction varie entre 2h et 6h, et pour lesquels des enjeux ont été identifiés concernant le risque de crue soudaine.

L'étude réalisée pour le SCHAPI en 2013 a montré l'apport des prévisions de pluie à haute résolution issues du modèle AROME (Application of Research to Operations at Mesoscale) de Météo-France pour améliorer la détection et l'anticipation des dépassements du débit biennal pour des bassins de Meuse-Moselle, Rhône-Cévennes, et Sud-Est (Demargne 2014). En 2014, l'évaluation a montré le bénéfice d'intégrer les prévisions d'ensemble de pluie à haute résolution issues du modèle COSMO-DE-EPS de l'agence Deutscher Wetterdienst (DWD) pour les bassins de Meuse-Moselle (Demargne 2015). Pour mieux caractériser les pluies futures des courtes échéances qui sont essentielles pour les crues soudaines, l'objectif est d'intégrer les prévisions immédiates AROME-PI de Météo-France, de type déterministe, disponibles toutes les heures (Auger et al. 2015). Ces prévisions ont été utilisées pour construire des prévisions d'ensemble par décalage temporel de prévisions déterministes successives (prévisions dites d'ensembles timelag). Cette approche pragmatique permet de prendre en compte les incertitudes sur les prévisions de pluie à une échelle spatio-temporelle fine (2,5 km), qui peuvent avoir un impact significatif pour des événements de pluie intense.

a) Développements du système d'anticipation AIGA

Pour les avertissements aux crues soudaines en milieu non jaugé, Irstea et Météo-France ont développé le système d'anticipation AIGA (Javelle et al. 2010, 2014), qui fonctionne en temps réel sur le pourtour méditerranéen depuis 2005. Le système AIGA a fait l'objet de nombreux développements concernant les améliorations du modèle hydrologique, des données d'observation et de prévisions, et de l'évaluation de la qualité.

Dans la version initiale du système tel qu'implémenté dans le Sud et sur la plateforme du projet RHYTMME (Risques HYdrométéorologiques en Territoires de Montagnes et MEditerranéens, rhytmme.irstea.fr), un modèle hydrologique distribué intègre les pluies horaires produites à la résolution de 1km^2 par Météo-France, à partir des radars météorologiques et des stations pluviométriques, pour estimer le débit en tout point des cours d'eau. Ces débits sont alors comparés en temps réel aux quantiles de débit préalablement estimés par la méthode régionalisée SHYREG (Arnaud et Lavabre 2002, Arnaud et al. 2008). Cette méthode a permis de fournir une base de quantiles de débit de pointe et de débits pour les durées de 1h à 72h pour les périodes de retour de 2 ans à 100 ans sur la France métropolitaine pour tout bassin versant de superficie comprise entre 10 et 2000 km^2 . Pour caractériser la gravité des événements en cours, le système AIGA affiche, toutes les 15 minutes, la carte des périodes de retour estimées le long des cours d'eau pour les catégories suivantes : entre 2 et 10 ans, entre 10 et 50 ans, et supérieur à 50 ans. Ces cartes sont actuellement disponibles en temps réel via la plateforme RHYTMME développée pour la région Provence-Alpes-Côte d'Azur, pour permettre aux services concernés de mieux gérer les événements de crue sur leurs territoires..

La calibration et la régionalisation du modèle pluie-débit ont été récemment améliorées par Organde (2013, 2014, 2015) pour l'implémentation d'ici 2016 par le SCHAPI d'un système national d'avertissement. En particulier, les travaux d'Organde de 2013 ont permis de tester

différentes options de modélisation hydrologique, événementielle ou continue, distribuée, semi-distribuée ou globale, ainsi que différents critères de calibration et différentes méthodes de régionalisation, afin de déterminer la ou les modélisations hydrologiques les plus pertinentes pour un système d'avertissement national. Les procédures de calibration et d'optimisation de la régionalisation pour la première mise en place du système national ont été finalisées dans l'étude d'Organde (2014).

Il est aussi proposé dans l'étude d'Organde (2015) de considérer des seuils d'avertissement de débit qui soient relatifs, i.e., basés sur l'estimation des quantiles de débit à partir des chroniques d'observations et des simulations de la modélisation hydrologique sur la même période (et non plus sur des quantiles de débit estimés par la méthode SHYREG). Cette approche permet de prendre en compte les biais du modèle hydrologique utilisé pour le système AIGA en considérant des quantiles de débit estimés à partir des débits simulés. Ainsi, dans le cas où le modèle hydrologique a tendance à surestimer les forts débits pour un bassin, le système AIGA aurait tendance à produire trop de fausses alertes en considérant des seuils de débit issus des débits observés ; en utilisant des seuils de débit plus forts issus des simulations du modèle, l'objectif est de réduire le nombre de fausses alertes. Ce débiaisement du modèle hydrologique en utilisant des seuils de débit relatifs fonctionne également si le modèle a tendance à sous-estimer les débits forts d'un bassin ; dans ce cas, les seuils issus des simulations de débits seront plus faibles pour permettre de réduire le nombre d'alertes manquées.

Les études de Demargne de 2013, 2014 et 2015 ont également montré les améliorations du système d'avertissement AIGA en termes de détection et d'anticipation des crues soudaines grâce à l'intégration des prévisions de pluie dans la modélisation hydrologique. Les développements récents de Météo-France et des autres agences de prévision météorologique pour générer des prévisions de pluie à plus fine échelle spatio-temporelle et issues d'une meilleure modélisation des phénomènes de pluie intense par les modèles atmosphériques numériques permettent d'intégrer ces prévisions en cohérence avec l'échelle de modélisation hydrologique.

Cependant, tout système de prévision est impacté par de nombreuses sources d'incertitude qui sont issues de la variabilité (naturelle et causée par l'homme) des phénomènes modélisés et des variables, des modélisations imparfaites de ces phénomènes, ainsi que des limitations en termes de mesures et données disponibles. Comme l'a souligné notamment le rapport du NRC (2006), ces incertitudes de prévisions doivent être prises en compte et quantifiées pour mettre à disposition des utilisateurs des prévisions avec leurs incertitudes associées et permettre ainsi leur utilisation optimale.

Pour un système de prévision hydrométéorologique, on différencie généralement (Demargne et al. 2014):

- les incertitudes météorologiques liées aux observations et prévisions des variables atmosphériques (ex. pluie et température) qui sont utilisées en entrée des modèles hydrologiques,
- les incertitudes hydrologiques, provenant des états initiaux des modèles hydrologiques, des paramètres de ces modèles, de la structure des modèles, et des régulations potentielles de débit.

Une des approches pour quantifier ces différentes sources d'incertitude est de produire des prévisions d'ensemble qui décrivent des scénarios équiprobables avec les différents membres de l'ensemble. L'approche ensembliste est mise en œuvre depuis plus de 20 ans dans les systèmes de prévisions météorologiques pour les prévisions à court, moyen et long terme, ce

b) Prévisions immédiates de pluie AROME-PI

Depuis décembre 2008, Météo-France produit des prévisions déterministes avec le modèle atmosphérique Application of Research to Operations at Mesoscale (AROME, Seity et al. 2011) à l'échelle de 2,5 km environ et au pas de temps horaire pour des échéances de 30h à 36h. La fine résolution spatiale d'AROME permet de modéliser les processus convectifs, qui sont essentiels pour les avertissements aux crues rapides. L'utilisation des prévisions AROME pour la prévision de crue fait également l'objet d'une expérimentation dans le cadre du projet CHROME mené par le SCHAPI avec différents modèles hydrologiques sur trois bassins versants cévenols. Météo-France développe actuellement des prévisions d'ensembles dites AROME-PE pour une mise en service opérationnelle prévue pour 2016.

Concernant la courte échéance, Météo-France a mis en place la production de prévision immédiate à partir du modèle AROME pour mieux caractériser les pluies des échéances de 1h à 6h. Ce système configuré pour la prévision immédiate est appelé AROME-NWC ou AROME-NOWCASTING (Auger et al. 2015). Il diffère notamment du système AROME par sa réactualisation horaire permettant d'utiliser les observations les plus récentes. Dans cette étude, nous proposons d'intégrer ces prévisions immédiates AROME-PI de type déterministe au système AIGA, ainsi que de construire des prévisions d'ensemble par décalage temporel de prévisions déterministes successives, prévisions dites d'ensembles timelag.

Les caractéristiques des prévisions AROME-PI sont les suivantes :

- prévisions issues du modèle AROME qui inclut une modélisation physique des processus convectifs ;
- des prévisions à la résolution spatiale d'environ 2,5 km, avec des cumuls incrémentés toutes les 15 minutes, pour une échéance maximale de 6 heures ;
- des prévisions disponibles toutes les heures.

Le délai de mise à disposition des prévisions par Météo-France n'a pas été considéré dans cette étude, mais il se limite à quelques minutes, ce qui ne devrait pas avoir un impact significatif sur l'utilisation de ces prévisions au sein d'AIGA.

c) Objectifs de l'étude

Comme décrit dans le schéma de la Figure 2, le travail proposé répond aux trois objectifs suivants :

- 1) évaluer l'intérêt d'intégrer la prévision déterministe AROME-PI pour mieux détecter et anticiper les dépassements des seuils de débit ;
- 2) prendre en compte l'incertitude des prévisions de pluie en intégrant des prévisions d'ensemble construites par décalage temporel des prévisions déterministes successives, pour générer des ensembles de débit, qui reflètent l'incertitude de pluie future, et des avertissements probabilisés de dépassement de seuil de débit ;
- 3) évaluer la qualité des prévisions d'ensemble de pluie et de débit, ainsi que la qualité des avertissements des dépassements vis-à-vis du système AIGA actuel (sans pluie future) et de prévisions déterministes de référence.

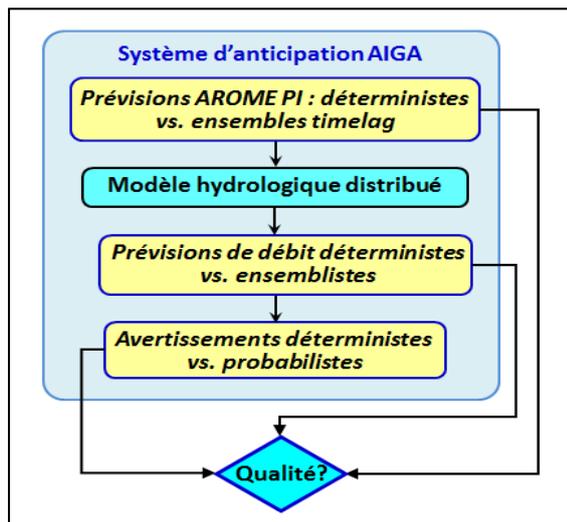


Figure 2 : Approche méthodologique pour évaluer l'apport des prévisions AROME-PI pour le système AIGA

2. Données et Modélisation Hydrologique

Le travail de cette étude utilise les données de débit et la modélisation hydrologique proposée dans l'étude d'Organde (2014) pour l'implémentation du système national AIGA, avec le jeu de paramètres calibrés et régionalisés de la fin d'année 2014. La zone d'étude correspond au Grand Sud qui a été impacté durant les 3 événements considérés.

a) Bassins versants

A partir de la base des bassins versants du SCHAPI, la sélection finale des bassins réalisée par Organde (2014) inclut 719 bassins versants sur le territoire national. Les critères de sélection de ces bassins retenus pour finaliser la calibration et la régionalisation de la modélisation hydrologique sont les suivants :

- les données de débit observées sont disponibles sur la période de calibration et régionalisation de la modélisation hydrologique 2001-2006 avec un taux de lacune inférieur jusqu'à 10% et dont la qualité en hautes eaux a été jugée acceptable par le réseau des SCP;
- les stations hydrométriques sont à une distance maximale de 100 km d'un radar opérationnel pour assurer une bonne qualité de la lame d'eau observée issue du signal radar ;
- les surfaces de bassin versant sont inférieures à 1 000 km² et les valeurs de superficie issues de la base de bassins versants du SCHAPI et de la banque HYDRO sont similaires ;
- les bassins versant sont « peu influencés » par les ouvrages hydrauliques ; la sélection est basée sur l'indice de confiance de Caruso (2012), estimé pour la méthode SHYREG en tout pixel de 50 m x 50 m et défini comme le ratio entre la superficie drainée par le(s) barrage(s) situé(s) sur le bassin versant et la surface du bassin versant à l'exutoire ; les bassins pour lesquelles l'indice de confiance est inférieur à 0,2 (influence dite faible selon Caruso (2012)) ont été retenus.

Pour centrer l'étude sur la zone de Grand Sud, on a sélectionné les 185 bassins versants pour lesquels la coordonnée Y en Lambert 2 étendu de l'exutoire du bassin est inférieure à 2050km. La Figure 3 donne la carte des 719 bassins versants incluant les 185 bassins retenus pour cette étude, ainsi que la distribution des surfaces de bassin, variant de 10 km² à 992 km², avec une valeur médiane de 161 km².

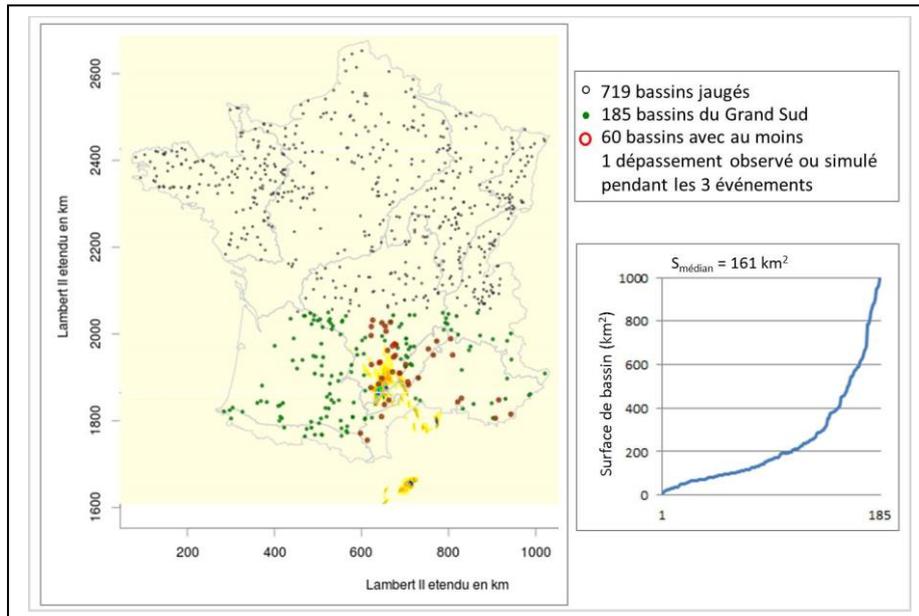


Figure 3 : Carte des 719 bassins versants calibrés et régionalisés en France, dont 185 bassins de la zone Grand Sud ont été sélectionnés pour cette étude, avec leur distribution de surface

b) Modélisation hydrologique et données d'observations

Comme le montre le schéma de la Figure 4, la modélisation hydrologique s'appuie sur le couplage de trois modèles (cf. Organde 2014 pour plus de détails) :

- le modèle événementiel horaire semi-distribué, noté GR-SD, basé sur une adaptation du modèle GR distribué du système AIGA actuel pour utiliser une production distribuée, calculée avec un réservoir de production en chaque pixel i du bassin, de capacité A_i , et un transfert global de bassin, estimé avec un unique réservoir de transfert, de capacité B .
- le modèle continu global journalier GR4J à quatre paramètres (Edijatno et al. 1999, Perrin 2002, et Perrin et al. 2003), qui fournit le débit journalier pour l'initialisation du niveau du réservoir de transfert du modèle horaire au début de l'événement ; le taux de remplissage initial du réservoir de transfert de GR-SD (R_0/B) est ajusté quotidiennement si la pluie journalière de bassin de la veille est inférieure à 10 mm afin de restituer le débit issu du modèle journalier global GR4J.
- le modèle continu distribuée journalier GR3J (Edijatno et Michel 1989) qui fournit le taux de remplissage du réservoir $S_i A_i(J)$ au pixel i du bassin; les valeurs de $S_i A_i(J)$ sont estimées à partir des grilles de pluies journalières et des grilles journalières d'ETP calculées à partir de la formule d'Oudin (Oudin et al., 2005) ; le taux de remplissage initial du réservoir de production du modèle GR-SD (S_{i0}/A_i) est ajusté quotidiennement si la pluie journalière de la veille au pixel est inférieure à 10 mm avec la relation suivante : $S_{i0}/A_i = a_i \cdot S_i A_i(J)$

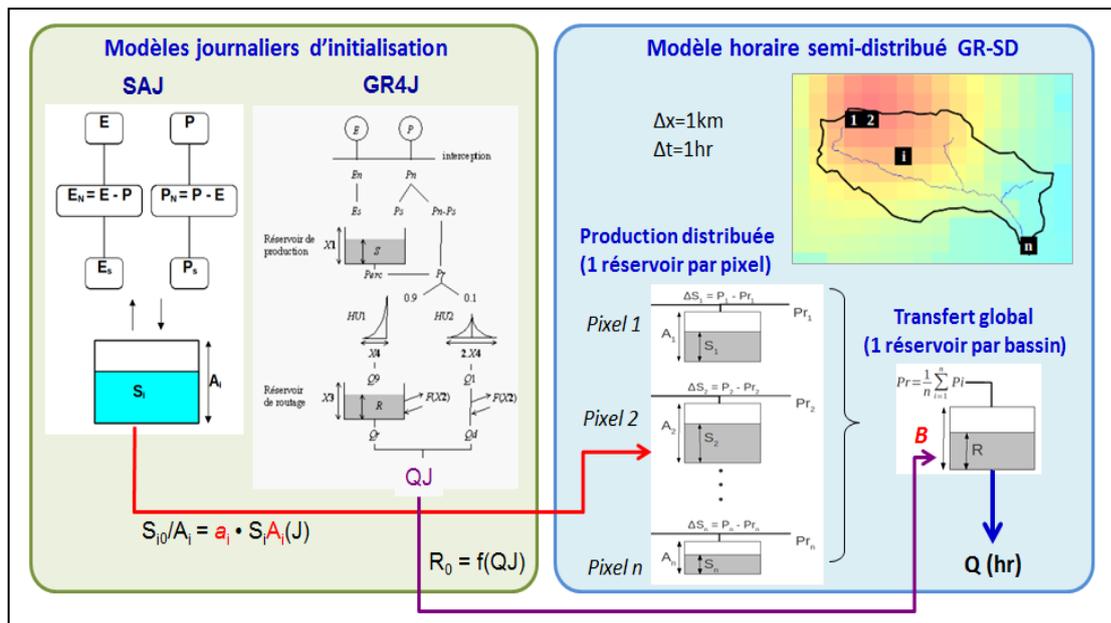


Figure 4 : Schéma du couplage des modèles journaliers continus SAJ et GR4J et du modèle horaire semi-distribué GR-SD pour le système AIGA national

Les paramètres des trois modèles GR-SD, SAJ et GR4J qui doivent être calibrés et régionalisés sont les suivants :

- capacité A_i du réservoir de production de GR-SD et du modèle SAJ, défini en chaque pixel i du territoire, qui a été fixé à 200 mm,
- paramètre a_i pour l'initialisation du modèle GR-SD, défini en chaque pixel i du territoire,
- capacité B du réservoir de transfert de GR-SD, défini pour chaque bassin versant,
- les 4 paramètres du modèle GR4J, définis en chaque bassin versant.

Pour améliorer la réactivité du modèle GRSD, la valeur minimale du niveau de remplissage initial du réservoir de production a été fixée : $S_{i0}/A_i = 0.2$.

Le calage des modèles GR-SD et GR4J a été réalisé à partir des données de pluie radar ré-analysées de Météo-France (COMEPHORE) sur la période 2002-2006. Les données hydrométriques proviennent de la banque HYDRO et de l'actualisation effectuée par le réseau de SPC à l'automne 2013.

Les paramètres a et B ont été calés et régionalisés sur la période 2002-2006 (Organde, 2013) et optimisés régionalement sur la période 2009-2012 avec les pluies opérationnelles (Organde, 2014). Les paramètres régionaux du modèle GR4J sont estimés comme les moyennes des paramètres calibrés des 3 voisins « donneurs » les plus proches (avec la distance euclidienne entre les barycentres des bassins versants). Les paramètres régionaux de GR-SD sont estimés par la méthode de régression en utilisant des critères morphométriques et climatiques. L'optimisation régionale a été réalisée pour les 9 hydro-écorégions en utilisant le critère de Heidke skill score (HSS) basé sur la détection des dépassements pour les seuils de quantiles de débit de période de retour de 2, 10 et 50 ans (en considérant le tirage aléatoire comme référence).

La Figure 5 montre les résultats en termes de Critical Success Index pour les dépassements du seuil de débit biennal (de la base de données SHYREG) pour la période 2009-2012 sur

622 bassins versants de l'étude d'Organde (2014). Ces cartes montrent que certains bassins versants de la zone Grand Sud, notamment dans le Sud du Massif Central, ont un CSI faible avec le modèle hydrologique opérationnel de 2014 utilisé dans cette étude.

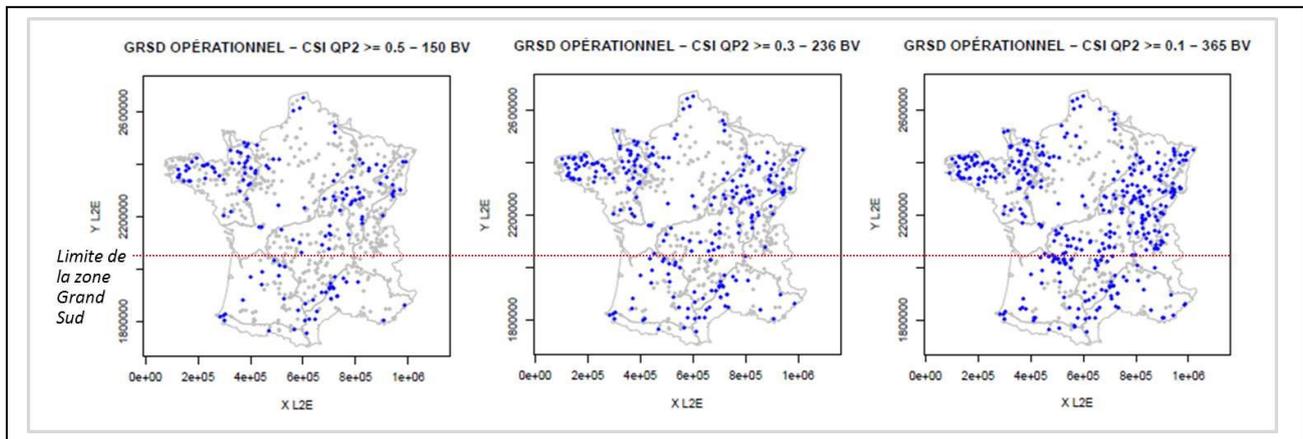


Figure 5 : Résultats du calage et de la régionalisation du modèle horaire semi-distribué GR-SD

Concernant les données d'observations de pluie, la chaîne d'avertissement en temps réel doit fonctionner avec les grilles de pluies opérationnelles produites par Météo-France à la résolution spatiale de 1 km^2 à partir des systèmes radars et des stations pluviométriques (Tabary, 2007; Tabary et al., 2007):

- les données « Antilope », qui permettent d'estimer les pluies journalières pour les 2 modèles journaliers SAJ et GR4J ;
- les données « Panthère » au pas de temps de 15 minutes, qui sont utilisées pour le modèle horaire GRSD.

On rappelle qu'il semble plus pertinent d'utiliser les pluies journalières Antilope pour les modèles journaliers (et non pas les pluies Panthère) car ces pluies Antilope bénéficient des informations issues des postes pluviométriques pour corriger les pluies issues de radar. Les données Panthère accordent un poids supérieur aux lames d'eau radar mais sont les seules données disponibles en temps réel toutes les 5 minutes. Pour Antilope, il y a une différenciation de traitement suivant le type de pluie : les lames d'eau radar sont utilisées telles quelles pour la partie convective des précipitations. Les données de pluviomètres sont spatialisées par krigeage pour la partie stratiforme. A partir d'août 2009, une correction des quantités de pluie issues des lames d'eau radar est introduite : une spatialisation en $1/r^2$ des résidus pluvio-radar au niveau des cumuls convectifs est faite. Le résultat de cette spatialisation est ensuite ajouté à la partie convective radar.

Pour produire les simulations et prévisions de débit sur les 3 événements étudiés, l'étude a utilisé les données horaires Panthère et les données journalières Antilope telles que produites rétrospectivement (dites données Antilope V2) pour mieux correspondre aux données produites et utilisées par AIGA en temps réel.

c) Sélection des événements

La sélection des événements a été réalisée en collaboration avec Météo-France pour le rejeu du modèle AROME avec la configuration spécifique pour les prévisions AROME-PI (le système n'étant pas encore opérationnel) (Auger et al. 2015).

Les prévisions AROME-PI ont été produites pour les 3 événements suivants :

- du 15 au 20 septembre 2014
- du 12 au 16 novembre 2014
- du 23 au 30 novembre 2014.

La Figure 6 présente des cartes de cumul des prévisions AROME-PI pour les 6 heures d'échéance pour une des dates avec de fortes pluies pour chacun des 3 événements étudiés.

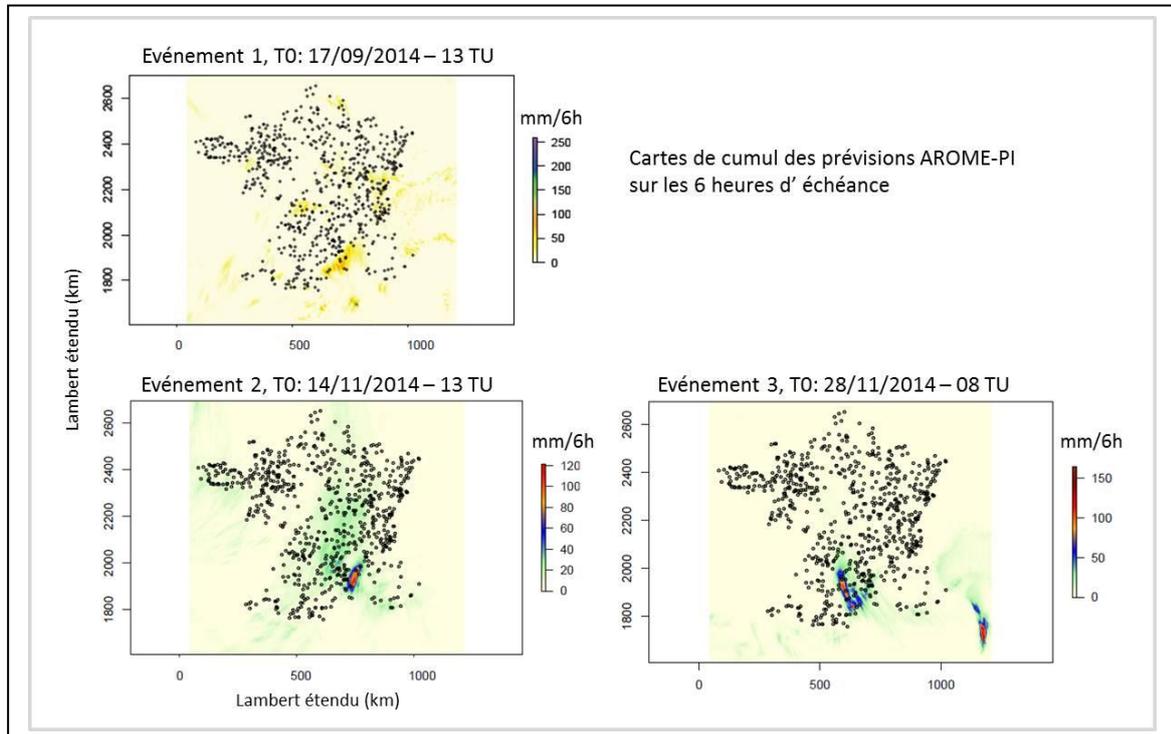


Figure 6 : Exemples de cartes de cumul des prévisions AROME-PI au cours des 3 événements étudiés

L'évaluation de la chaîne d'avertissement AIGA porte sur les dépassements du seuil relatif de débit de période de retour de l'ordre de 2 ans, ces seuils ayant été estimés dans l'étude d'Organde (2015). Pour les 3 événements retenus, l'analyse a inclus 524 dépassements et non-dépassements sur les 185 bassins versants étudiés de la zone Grand Sud.

3. Méthodologies

On décrit ici les étapes de la méthodologie pour construire des prévisions de pluie de type déterministe pouvant être intégrées dans AIGA, générer des prévisions d'ensemble à partir des prévisions de pluie AROME-PI déterministes, produire des avertissements de dépassement de seuil de débit, et pour évaluer la qualité des avertissements et des prévisions déterministes et ensemblistes.

a) *Prévisions de pluie AROME-PI et ensembles timelag*

Pour les trois événements sélectionnés, Météo-France nous a fourni les prévisions déterministes AROME-PI (Auger et al. 2015) avec les caractéristiques suivantes :

- paramètre PRECIP niveau SOL (quantité totale de précipitations en kg.m-2),
- cumuls avec une incrémentation de 15 minutes et une échéance maximale de 6h,

- réactualisation horaire
- grilles en coordonnées géographiques (latitude, longitude) de 0.0175 degrés de résolution, correspondant à une résolution de 2,5 km environ.

La première étape du travail a consisté à générer les grilles de prévisions de pluie horaire sur le domaine d'étude dans le système de projection et à la résolution spatiale utilisés dans le système AIGA. Ainsi, les grilles de prévision de cumul de pluie incrémentées au pas de temps de 15 minutes sont tout d'abord cumulées pour obtenir des cumuls horaires. Ensuite, les grilles de cumul horaire dans le système de projection cartographique latitude-longitude sont projetées dans le système Lambert 2 étendu. Finalement, les grilles sont interpolées sur une grille régulière à la résolution de 1 km² en utilisant la méthode du plus proche voisin. Les traitements sont réalisés sous le langage R avec le package raster, pour obtenir des grilles au format raster Rdata du langage R.

La construction d'ensembles par décalage temporel est une approche classique en prévision atmosphérique, du fait de sa relative simplicité de mise en œuvre lorsqu'une archive de prévisions déterministes est disponible (Ben Bouallègue et al. 2013). Elle permet de prendre en compte les incertitudes des prévisions météorologiques en termes d'intensité et de localisation dans l'espace et dans le temps grâce à la dispersion des prévisions déterministes successives (Mittermaier 2007) et de tester leur apport potentiel pour l'amélioration du système d'avertissement.

Dans cette étude, nous avons choisi de construire des prévisions d'ensembles AROME-PI timelag en considérant 6 membres, i.e. la dernière prévision AROME-PI ainsi que les 5 prévisions AROME-PI précédentes. Pour couvrir les 6 heures d'échéance de prévision, les valeurs des prévisions AROME-PI sont complétées par des grilles de pluie nulle. Ainsi pour toutes les heures d'initialisation, la prévision AROME-PI timelag inclut 6 membres pour les échéances de 1h à 6h. La Figure 7 présente le schéma de construction des prévisions d'ensembles AROME-PI timelag pour le jour J pour la date d'initialisation des prévisions T0 de 7h.

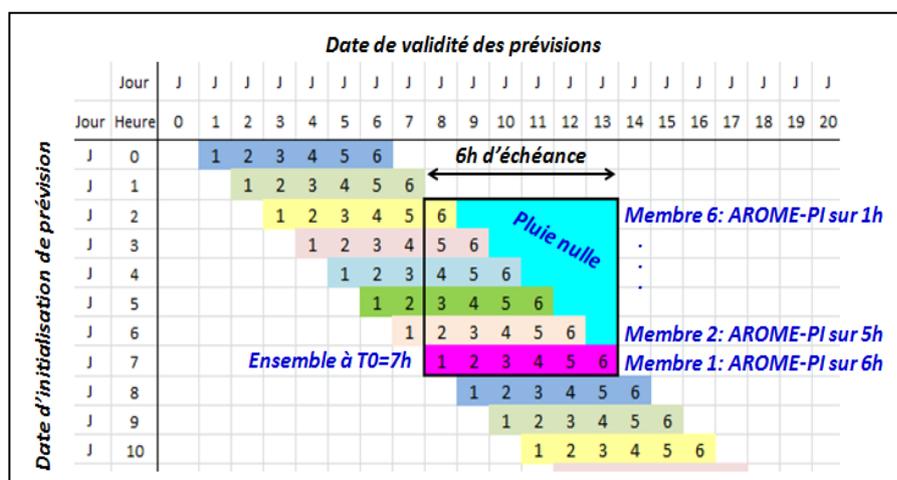


Figure 7 : Schéma de construction des ensembles timelag de pluie à partir des prévisions déterministes AROME-PI successives

Les 6 membres des prévisions d'ensemble AROME-PI timelag sont considérés comme des scénarios équiprobables. Ils sont utilisés pour générer les prévisions d'ensemble de débit sur les 6 heures d'échéance.

b) Prévisions de débit et avertissements AIGA

La seule source d'incertitude considérée dans cette étude correspond à celle des prévisions de pluie, en utilisant une approche simple et pragmatique utilisant les prévisions déterministes AROME-PI successives.

La prévision de débit est obtenue en intégrant la pluie observée jusqu'à la date d'initialisation de prévision T0 et ensuite la prévision de pluie dans le modèle hydrologique. Les prévisions d'ensemble de débit sont obtenues à partir des 6 membres de la prévision d'ensemble AROME-PI timelag, chaque ensemble de pluie générant un ensemble de débit. On obtient ainsi une prévision d'ensemble de débit avec 6 membres décrivant l'impact de l'incertitude de pluie future sur les prévisions de débit.

Pour montrer l'apport de la prévision d'ensemble de débit, nous utilisons les prévisions de référence suivantes :

- le débit simulé, sans pluie future, qui produit les avertissements du système AIGA actuel ; cette « prévision sans échéance » (notée Qsim) correspond au débit simulé obtenu à la date d'initialisation T0,
- le débit simulé avec pluie future nulle, qui est une des options de fonctionnement du système AIGA pré-opérationnel (notée Qpnul_ech6h) ;
- le débit simulé obtenu avec la pluie future « parfaite » (en négligeant les erreurs de la pluie observée), connue *a posteriori* ; la prévision de débit issue de la pluie observée est produite pour la même échéance que la prévision d'ensemble (prévision notée Qpobs_ech6h) ;
- la prévision de débit issue de la pluie persistante, définie en maintenant la valeur de pluie observée de la dernière heure avant l'initialisation des prévisions (prévision notée Qpers_ech6), qui est également une des options de fonctionnement du système AIGA pré-opérationnel.

Toutes les prévisions de débit sont produites toutes les heures pour les 3 événements de l'étude.

Comme pour l'étude portant sur la mise à jour du paramétrage de la modélisation hydrologique d'Organde en 2015, l'évaluation de la qualité des avertissements s'est basée sur le dépassement d'un seuil de débit relatif correspondant à une période de retour de l'ordre de 2 ans. Ainsi, 2 valeurs de seuil de débit sont définies : une valeur de seuil QOBS_2ans, estimé à partir des observations de débit sur la période 2002-2015, et une valeur de seuil QSIM_2ans, estimé à partir des simulations de débit sur cette même période 2002-2015. Cette approche permet de « débiaiser » les sorties du modèle hydrologique.

La Figure 8 présente un exemple des hydrogrammes de débit issus des 6 ensembles de pluie AROMEPI-timelag, du débit simulé et du débit observé avec une échéance de prévision de 6 heures. Les ensembles de pluie sont représentés par la moyenne et le maximum des 6 ensembles. Les 6 ensembles de débit correspondants sont représentés par leur enveloppe et la moyenne en calculant, pour chacune des 6 heures d'échéance, la valeur minimale, la valeur maximale et la valeur moyenne des 6 membres. Dans cet exemple, on peut noter l'intérêt d'utiliser 2 seuils de débit pour les avertissements, ce qui permet de prendre en compte la tendance à la surestimation du débit observé par le débit simulé pour ce bassin.

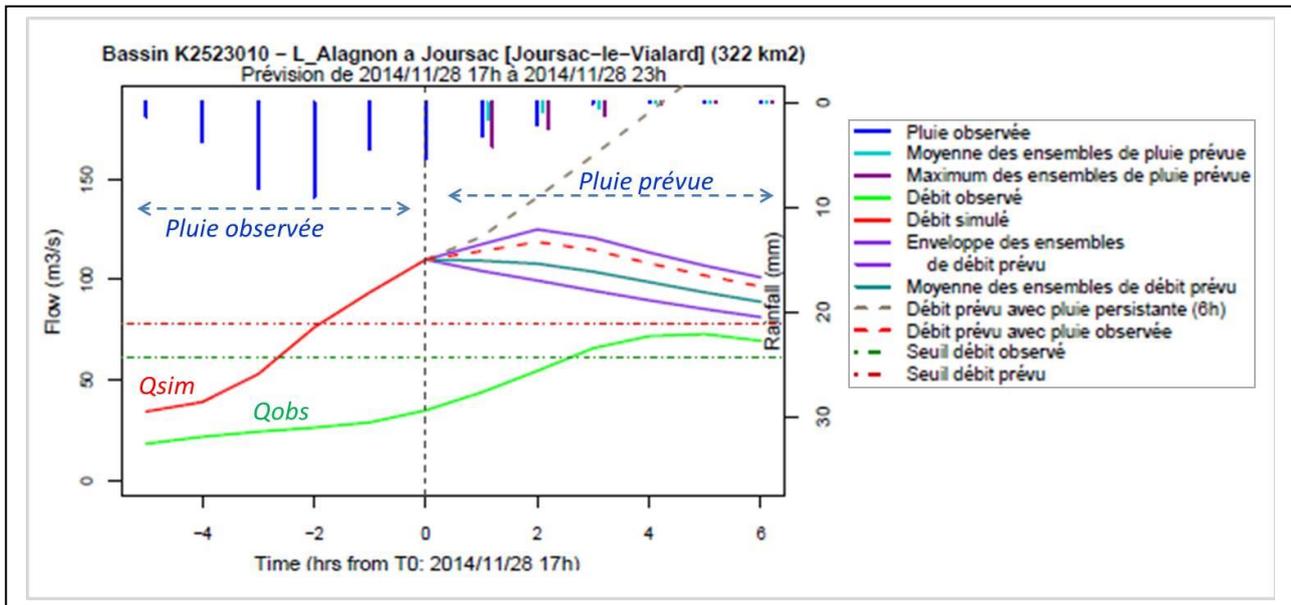


Figure 8 : Exemple d'hydrogrammes avec le débit simulé, l'enveloppe et la moyenne des 6 ensembles de débit, la prévision de débit issu de la pluie persistante pour la date d'initialisation du 28 novembre 2014 à 17h (TU) en comparaison avec l'hydrogramme de débit observé pour le bassin de l'Alagnon à Joursac

Pour les avertissements de dépassement d'un seuil de débit, les prévisions d'ensemble de débit conduisent à estimer une probabilité de dépassement du seuil en fonction du nombre de membres dépassant le seuil. Les membres de l'ensemble sont considérés comme tous équiprobables. La distribution empirique de probabilité pour la prévision d'ensemble de débit de dépasser un seuil de débit Q_{seuil} est estimée avec la formule suivante :

$$P[Q \geq Q_{seuil}] = i_{ens} / (nb_{ens} + 1)$$

avec i_{ens} le nombre de membres de l'ensemble dépassant le seuil de débit
et nb_{ens} le nombre total des membres de l'ensemble

Dans cette étude, les résultats pour les dépassements de seuil de débit sont présentés en considérant la probabilité de dépassement d'1 membre sur 6 (ce qui correspond au dépassement issu des valeurs maximales des 6 membres de l'ensemble), de 3 membres sur 6 et des 6 membres de l'ensemble.

Il faut noter que, pour la prévision sans pluie future et la prévision issue de la pluie observée, elles sont toutes les deux issues des valeurs de débit simulé. Mais la prévision issue de la pluie observée ayant la même échéance que les prévisions AROME-PI, elle inclut, pour la date d'initialisation T_0 , 6 valeurs de débit simulé de T_0+1hr jusqu'à T_0+6hr . La prévision sans pluie future est la valeur du débit simulé à T_0 .

c) Evaluation de la qualité des avertissements

L'analyse de la qualité des avertissements nécessite d'intégrer plusieurs critères de vérification pour quantifier les différents aspects de leur qualité, en termes notamment de fiabilité, justesse, et anticipation (Jolliffe et Stephenson 2003, Wilks 2011, WMO 2013). La qualité des avertissements utilise les scores de contingence des dépassements des seuils par le débit observé et par les différentes prévisions de débit. Les scores sont calculés pour chaque type de prévision pour les 3 événements étudiés et les 185 bassins du Grand Sud.

L'évaluation concerne également le degré d'anticipation des dépassements prévus vis-à-vis des dépassements de seuil observés.

Scores de contingence des dépassements de seuil de débit

Le tableau de contingence se définit pour un évènement donné, c'est-à-dire pour un seuil relatif fixé qui définit si oui ou non l'observation ou la prévision a dépassé le seuil relatif considéré. A noter que la prévision dépasse le seuil si, pour une des échéances, il y a dépassement. De plus, on comptabilise les dépassements du débit observé et du débit prévu pour chaque évènement dans son ensemble (i.e. pour toutes les dates d'initialisation de prévision de l'évènement). Sur les 3 évènements étudiés, on comptabilise les dépassements de seuil issus des observations et des prévisions de la façon suivante :

Evènement : dépassement d'un seuil relatif donné		Evènement observé	
		Oui	Non
Evènement prévu	Oui	a (succès/hit)	b (fausse alerte/false alarm)
	Non	c (manqué/miss)	d (rejet correct/ true negative)

A partir du tableau de contingence, on définit les scores d'évaluation suivants (WMO 2013):

- **Biais en fréquence** ou **Frequency Bias** (FB), qui décrit si le modèle a tendance à produire trop de fausses alertes ou trop d'alertes manquées : $FB = (a+b)/(a+c)$; score optimal $FB=1$.
- **Taux de bonne détection** ou **Probability Of Detection** (POD) (encore appelé Hit Rate), qui est conditionné par les dépassements observés : $POD = a/(a+c)$; le score varie entre 0 (toutes les alertes observées sont manquées) et 1 (toutes les alertes observées sont détectées), avec un score optimal égal à 1.
- **Taux d'alerte correcte** ou **Success Ratio** (SR), qui est conditionné par les alertes prévues pour mesurer la fiabilité des prévisions : $SR=a/(a+b)$; le score varie entre 0 (aucune alerte prévue qui soit correcte) et 1 (aucune fausse alerte), avec un score optimal égal à 1.
- **Critical Success Index** (CSI) qui permet de considérer les alertes manquées et les fausses alertes (mais sans différencier ces erreurs); $CSI=a/(a+b+c)$; le score varie entre 0 et 1, avec un score optimal égal à 1.
- **Taux de fausses alertes** ou **False Alarm Ratio** (FAR), qui est conditionné sur les dépassements prévus : $FAR = b/(a+b)$; le score varie entre 0 (aucune fausse alerte) et 1 (tous les événements prévus sont des fausses alertes), avec un score optimal égal à 0.

Prise en compte de l'anticipation des dépassements

L'analyse de la qualité des avertissements se base également sur l'anticipation des dépassements issus des prévisions. L'anticipation est définie vis-à-vis des alertes observées, i.e. par rapport à la date à laquelle le débit observé dépasse le seuil de débit biennal.

L'anticipation est positive lorsque la prévision à une date d'initialisation T_0 produit un dépassement du seuil de débit avant la date de dépassement effectif par le débit observé. Pour cette analyse relative à l'ensemble d'un évènement, l'anticipation est définie par la valeur la plus grande relative à la première date de dépassement par la prévision de débit.

L'anticipation est négative lorsque la première date du dépassement prévu T_0 intervient après

le dépassement effectif du débit observé. On peut donc comparer sur un même jeu de bassins l'anticipation des dépassements issus des différents types de prévisions.

Comme on cherche à optimiser l'anticipation des dépassements vis-à-vis des dépassements observés, les critères de contingence sont calculés en prenant en compte l'anticipation du dépassement de seuil par la prévision. Ainsi, un dépassement de seuil est prévu avec retard si la première date T_0 de dépassement prévu est plus tardive que la date du dépassement effectif du débit observé ; il est alors considéré comme une alerte manquée. Le dépassement de seuil est comptabilisé comme une bonne alerte si le dépassement prévu a une anticipation égale ou supérieure à 0.

d) Vérification des prévisions d'ensemble

Pour compléter l'analyse de la qualité des alertes, les prévisions d'ensemble de pluie de bassin et de débit sont évaluées pour quantifier la similarité entre les observations et les prévisions. Cependant, pour se limiter aux débits les plus forts pour les 3 événements étudiés, cette évaluation n'a concerné que les 56 bassins de la zone Grand Sud avec des dépassements observés ou simulés.

L'évaluation de la qualité des prévisions d'ensemble est réalisée avec le système de vérification Ensemble Verification System (EVS) (Brown et al. 2010, téléchargeable sur le site www.nws.noaa.gov/oh/evs.html). L'outil EVS permet d'évaluer les variables hydrométéorologiques (pluie, température, débit) pour différentes échéances et à de multiples échelles temporelles (ex. horaire, journalier). La qualité d'une prévision peut être analysée en fonction de différents facteurs tels que l'échéance de prévision, la saisonnalité, l'agrégation temporelle, la magnitude de l'évènement (dépassement d'un seuil), ou la valeur d'une autre variable (par exemple, qualité de la prévision de débit en fonction de la quantité de pluie observée ou prévue). EVS inclut un grand nombre de métriques de vérification pour analyser les différents aspects de fiabilité, justesse, discrimination, et gain vis-à-vis d'une prévision de référence. Ces différentes métriques permettent d'analyser la qualité des prévisions grâce à des scores et des graphiques plus ou moins détaillés. Il est également possible d'agréger les prévisions à de multiples échelles temporelles (ex. du pas de temps horaire au pas de temps tri-horaire). On peut également agréger les performances sur un jeu de bassins, soit en calculant les scores de vérification à partir de toutes les paires observations-prévisions de l'ensemble des bassins, soit en calculant une moyenne pondérée des scores de vérification obtenus pour chacun des bassins.

Concernant le choix des scores de vérification, les paires observations-prévisions des 56 bassins versants pour les 3 événements de cette étude constituent un échantillon de taille modeste. C'est pourquoi on a sélectionné des critères d'évaluation globaux, qui ne demandent pas un sous-échantillonnage des paires observations-prévisions (comme par exemple avec le diagramme de fiabilité). De plus, comme les bassins versants ont des surfaces très variables, on vérifie les débits spécifiques (i.e. débit divisé par la surface totale du bassin versant en km^2) et on privilégie des mesures relatives de vérification ou des scores de gain de performance par rapport à une référence. On peut alors agréger les résultats sur l'ensemble des bassins pour obtenir des résultats plus robustes. Pour analyser l'apport des prévisions d'ensemble vis-à-vis des prévisions déterministes de référence (la prévision de débit issue de la pluie observée et de la prévision persistante de pluie), les scores choisis peuvent être estimés pour des prévisions de type ensembliste ou déterministe.

Pour évaluer les différents aspects de la qualité des prévisions avec un nombre limité de scores, on a sélectionné les critères suivants :

- **l'erreur relative moyenne** (estimée pour la prévision déterministe ou pour la moyenne des ensembles), qui permet de quantifier le biais inconditionnel ou biais relatif des prévisions en mesurant la différence moyenne entre les prévisions et les observations, exprimée en termes de fraction de la moyenne des observations ; plus le critère est proche de 0, moins la prévision est biaisée ; une erreur relative moyenne négative / positive indique une sous-estimation / surestimation respectivement de la prévision déterministe ou de la moyenne des ensembles ;
- **le coefficient de corrélation de Pearson**, estimée pour la prévision déterministe ou pour la moyenne des ensembles, qui décrit la relation linéaire entre les prévisions et les observations ; plus le critère est proche de 1, meilleure est la relation linéaire entre les prévisions et les observations (cependant les prévisions peuvent être biaisées vis-à-vis des observations) ;
- **le Brier Skill Score**, score de gain en performance pour le critère de Brier Score (BS) ; le Brier Score permet de mesurer l'erreur quadratique moyenne en probabilité de la prévision pour un événement binaire donné (ex. dépassement d'une valeur de seuil) ; la probabilité associée à l'observation est de 1 en cas d'occurrence de l'événement et de 0 sinon (la probabilité associée à une prévision d'ensemble varie entre 0 et 1) ; le Brier Score a une valeur de 0 pour un système déterministe parfait et une valeur de 1 lorsque le système prévoit systématiquement l'inverse des occurrences observées de l'événement ; le Brier Score est moyenné sur l'ensemble des paires observations-prévisions de l'événement binaire considéré ; le score de performance BSS est alors calculé en comparant le Brier Score de la prévision avec le Brier Score d'une référence avec la formule suivante : $BSS = 1 - [BS_{\text{prévi}} / BS_{\text{ref}}]$; une valeur négative / positive de BSS indique que la prévision a une erreur en probabilité moins / plus forte que la référence respectivement ; dans cette étude, on utilise la référence de prévision probabiliste climatologique, définie par la distribution de probabilité empirique des observations de l'année 2014 ; cette référence permet de quantifier la performance des prévisions d'ensemble pour les situations les moins « communes » de l'échantillon ;
- **le Relative Operating Characteristic Score (ROCS)**, qui quantifie la capacité de discrimination des prévisions pour un événement binaire donné ; ce critère est estimé à partir de la courbe ROC de la probabilité de détection (POD) en fonction de la probabilité de fausse détection (POFD), qui est représentée en Figure 12 pour un événement donné (ici, le dépassement d'un seuil) ; pour une prévision ensembliste, on doit définir différents niveaux de probabilité pour obtenir une prévision binaire (dépassement prévu ou non) qui est alors comparée à l'observation (dépassement observé ou non) ; une prévision parfaite est définie par $POD=1$ et $POFD=0$; la prévision doit avoir une courbe ROC au-dessus de la diagonale (pour laquelle $POD=POFD$) pour pouvoir correctement prévoir les dépassements en évitant les fausses alarmes pour les non-dépassements ; le critère ROCS correspond à 2 fois l'aire entre la courbe ROC et la diagonale ; plus le ROCS est proche de 1, plus la prévision différencie les dépassements et les non-dépassements.

Pour définir des seuils de dépassements cohérents vis-à-vis des échantillons de prévisions-observations sur les différents bassins, on définit les seuils par les quantiles

des observations à partir de la distribution empirique sur l'année 2014. Dans cette étude, on s'intéresse uniquement aux quantiles observés 80% pour la pluie de bassin et le débit. A noter que l'utilisation du seuil de débit biennal aurait nécessité un échantillon de paires prévisions-observations plus large pour obtenir des résultats statistiques robustes.

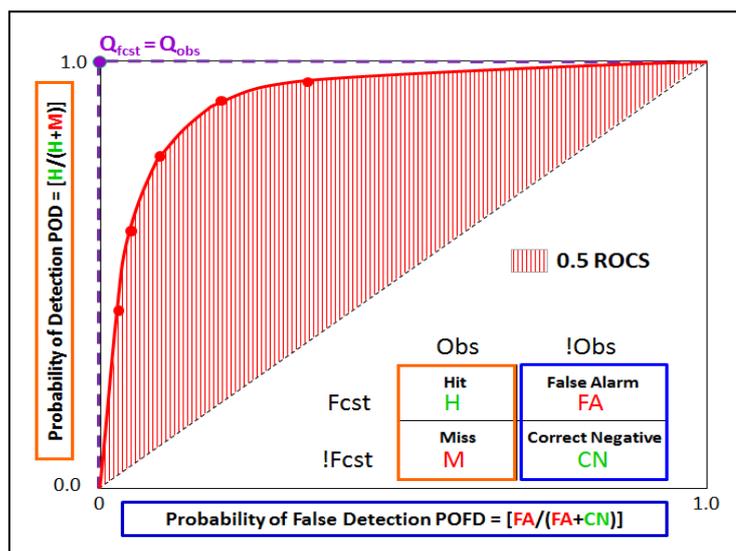


Figure 9 : Estimation du Relative Operating Curve Score (ROCS) pour une prévision probabiliste

Les quantiles utilisés pour le calcul du score de performance de Brier et du score de discrimination permettent également de présenter les scores d'erreur relative moyenne et de coefficient de corrélation pour les sous-échantillons des paires prévisions-observations dont les observations sont supérieures aux quantiles. Ceci permet de mieux analyser la qualité des prévisions pour les valeurs observées les plus fortes, ce qui est le focus du système d'avertissement aux crues.

4. Résultats

Les résultats de l'évaluation sont présentés tout d'abord pour la qualité des prévisions d'ensemble de pluie de bassin et de débit et, ensuite, pour la qualité des avertissements de dépassement du seuil relatif de débit biennal. Les résultats d'évaluation des prévisions d'ensemble ont été obtenus pour les 3 événements et pour les 56 bassins potentiellement les plus impactés. En faisant l'hypothèse que les scores d'évaluation ont des caractéristiques homogènes pour tous les bassins de la zone d'étude, les statistiques sont présentées après agrégation des résultats en considérant la moyenne des statistiques sur tous les bassins (un même poids étant donné à chaque bassin). Ces résultats mériteraient d'inclure un échantillon plus large de bassins et d'événements pour améliorer leur robustesse mais ils sont indicatifs de l'apport potentiel des prévisions AROME-PI pour les avertissements aux crues soudaines dans la zone Grand Sud.

a) Qualité des prévisions d'ensembles de pluie de bassin

Les prévisions d'ensembles AROMEPI-timelag construites à partir des prévisions déterministes successives permettent de produire des grilles de prévisions de pluie pour chaque bassin versant. Les prévisions de pluie sont vérifiées en pluie de bassin pour les échéances de 1hr à 6hr (mais en excluant les valeurs de pluie nulle, qui sont ajoutées

ultérieurement pour couvrir toutes les échéances de prévision de pluie lors de la production des prévisions d'ensemble de débit, tel que décrit par la Figure 7) et pour leurs cumuls sur 1hr, 3hr et 6hr. Il faut noter que, pour les prévisions de pluie, le cumul sur les 6 heures d'échéance correspond à la prévision déterministe AROME-PI la plus récente et donc au premier membre de la prévision d'ensemble. Les observations de pluie sont issues des données Panthère qui sont utilisées en entrée du modèle hydrologique horaire semi-distribué. Les résultats présentés incluent les histogrammes des erreurs des prévisions d'ensemble en fonction de la pluie observée et les scores de vérification. Les scores sont calculés pour chaque bassin pour toutes les paires observations-prévisions, ainsi que pour les paires dont les observations sont supérieures aux quantiles observés 80% ; les scores sont ensuite moyennés sur les 56 bassins étudiés. Les échantillons pour chaque bassin sont en moyenne de 357 paires au total et 72 paires pour le dépassement du quantile observé 80%.

Erreurs des ensembles de pluie en fonction de la pluie observée

La Figure 10 présente les histogrammes des erreurs des prévisions d'ensemble en fonction de la pluie observée pour identifier les éventuels biais ou valeurs aberrantes. Pour chaque paire d'observation-prévision on calcule les erreurs des différents membres de la prévision (ex. 6 membres pour l'échéance de 1hr et 4 membres pour l'échéance de 3hrs) ; la distribution de ces erreurs est alors représentée par une boîte à moustache avec les valeurs des quartiles, le minimum et le maximum. Toutes ces distributions d'erreur sont représentées pour chaque paire d'observation-prévision en fonction de la valeur observée. Les graphes d'erreur ci-dessous incluent les résultats des 56 bassins. Idéalement, les erreurs des prévisions doivent être faibles et capturer la valeur nulle. Les erreurs sont présentées pour les ensembles AROMEPI-timelag pour la première échéance pour les cumuls de pluie sur 1hr et 3hrs.

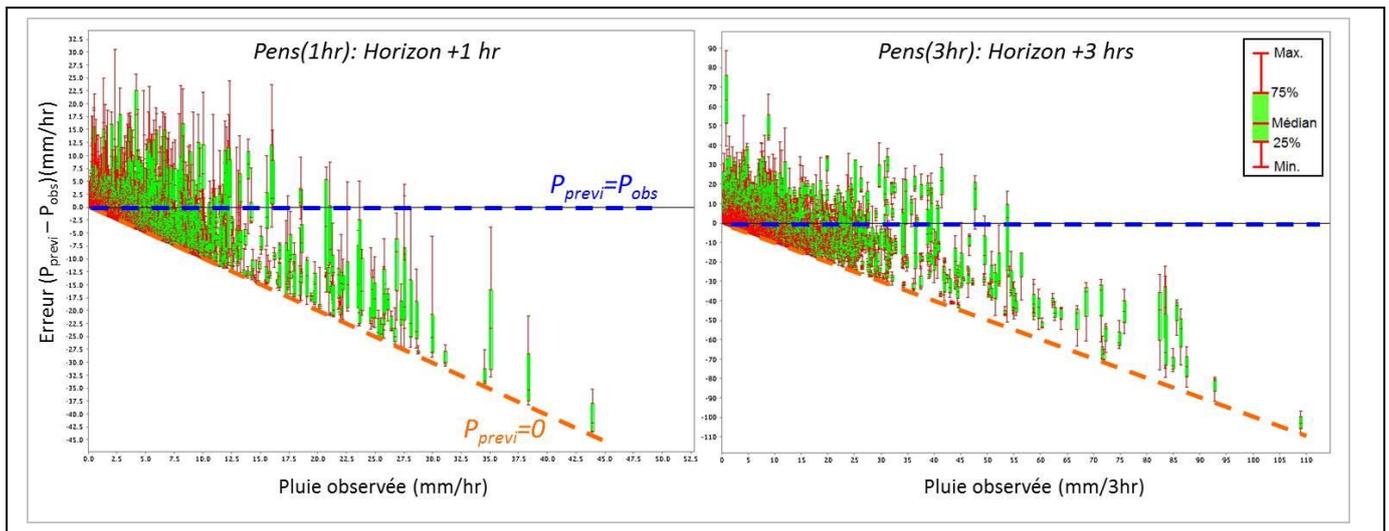


Figure 10 : Histogrammes des erreurs des prévisions d'ensemble de pluie de bassin en fonction de la pluie observée pour les cumuls sur 1hr et 3hrs

Le premier graphe de la Figure 10, qui concerne la pluie horaire pour la première échéance, montre une sous-estimation des fortes valeurs observées et une surestimation des faibles précipitations, ce qui démontre un biais conditionnel des prévisions de pluie (conditionnel car dépendant de la valeur observée). Pour les valeurs relativement fortes, les ensembles montrent une large dispersion mais, dans certains cas, la valeur observée n'est pas incluse dans l'intervalle [minimum, maximum] des 6 membres du fait du biais. Pour le cumul sur les 3 premières heures, on peut observer une tendance à la dégradation des erreurs avec un manque de dispersion des membres.

Impact de l'échéance sur la qualité des ensembles de pluie de bassin

Les scores d'évaluation sont représentés en fonction de l'échéance de prévision dans les graphes ci-dessous. On calcule également les scores pour les sous-échantillons pour lesquels les observations sont supérieures aux quantiles observés 80%. Ces quantiles observés sont les valeurs de seuil pour définir les dépassements pour le score de Brier et le score de discrimination ROCS. Les scores obtenus pour les prévisions AROMEPI-timelag sont comparés à ceux obtenus avec la pluie persistante.

L'erreur relative moyenne en Figure 11 de la prévision horaire montre une légère surestimation des pluies par la moyenne des ensembles (de l'ordre de 10%), qui peut être en partie due au choix de la méthode d'interpolation de plus proche voisin des grilles AROME-PI. Cependant la tendance est de sous-estimer les pluies les plus fortes (biais de 30% environ pour toutes les échéances), ce qui pourrait amener une sous-estimation des débits de pointe. La pluie persistante ne montre pas de biais général mais a une tendance marquée à sous-estimer les pluies les plus fortes, le biais se renforçant rapidement avec l'échéance ; ce biais négatif est supérieur à celui des prévisions AROMEPI-timelag à partir de l'échéance de 3 heures. Le coefficient de corrélation pour la moyenne des ensembles se dégrade également avec l'échéance et pour les pluies les plus fortes, mais sont meilleurs que les scores de la pluie persistante dès la deuxième heure d'échéance. Les scores de Brier en référence à la prévision probabiliste climatologique montrent une meilleure performance des prévisions d'ensemble des dépassements jusqu'à l'échéance de 4 heures ; le gain est meilleur avec la pluie persistante pour la première heure d'échéance uniquement. Le score de discrimination ROCS montre une bonne discrimination par les ensembles des dépassements du seuil, le score étant supérieur à celui de la pluie persistante pour toutes les échéances.

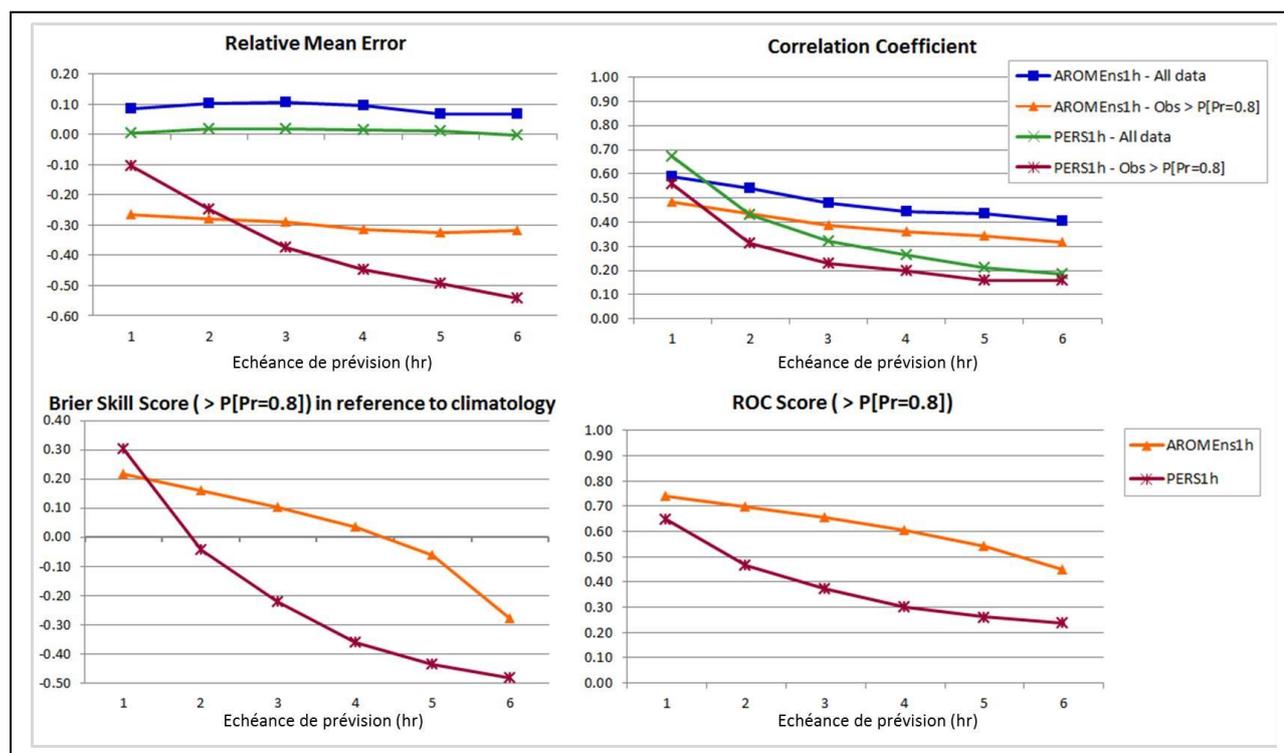


Figure 11 : Impact de l'échéance sur les scores de vérification des prévisions d'ensemble de pluie de bassin et de pluie persistante

Impact de la période d'agrégation sur la qualité des ensembles de pluie de bassin

La Figure 12 compare les scores d'évaluation en fonction de la période de cumul de la pluie de bassin (1hr, 3hr et 6hr) pour les observations supérieures au quantile observé 80%. On rappelle que le cumul sur les 6 heures d'échéance correspond à la prévision déterministe AROME-PI la plus récente et au premier membre de la prévision d'ensemble. Il faut noter que, généralement, l'agrégation améliore les scores d'évaluation par effet de lissage (les erreurs pouvant se compenser sur les périodes plus longues).

Les erreurs moyennes relatives montrent que la sous-estimation demeure pour les cumuls des ensembles de pluie sur 3hr et 6hr mais avec une réduction du biais (biais de l'ordre de 20% alors qu'il est de 30% pour la prévision horaire). Le biais de la pluie persistante est lui-aussi plus limité pour atteindre 20% environ pour le cumul sur 6hr. Le coefficient de corrélation de la moyenne des ensembles reste stable pour tous les cumuls (de l'ordre de 50%), alors qu'il se dégrade significativement pour la pluie persistante. Le gain en score de Brier des cumuls des ensembles sur 3hr et 6hr s'améliore vis-à-vis des cumuls horaires, mais il n'est positif que pour le cumul de 3hr (BSS de 17%). La discrimination pour les dépassements est améliorée avec les cumuls des pluies sur 3hr et 6hr, même si elle se dégrade pour 6hr, et le score reste supérieur à celui de la pluie persistante.

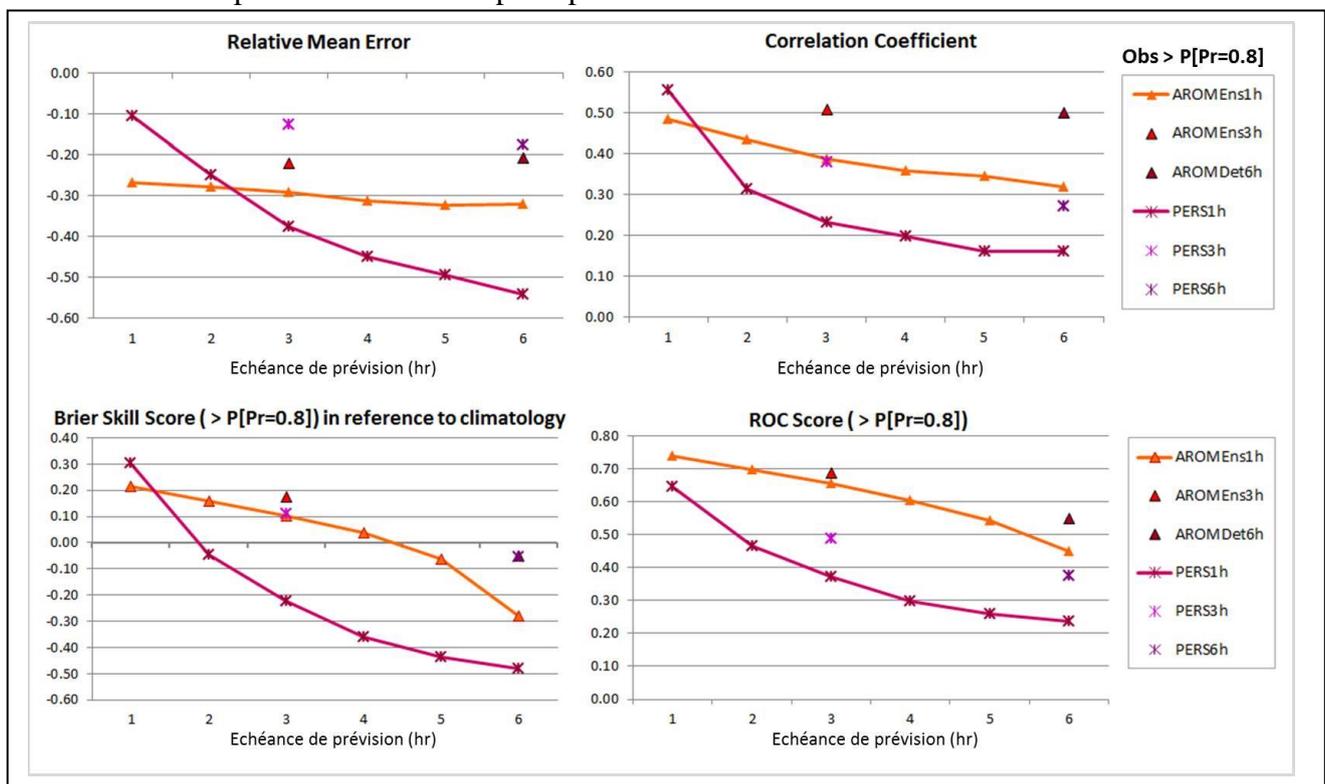


Figure 12 : Impact de l'agrégation sur les scores de vérification des prévisions d'ensemble et de pluie persistante pour les cumuls de pluie de bassin sur 1hr, 3hr et 6hr

Les résultats de l'évaluation des prévisions de pluie de bassin ont permis d'identifier le biais conditionnel des moyennes des ensembles de pluie, qui ont tendance à surestimer les faibles pluies et à significativement sous-estimer les fortes pluies. Mais les ensembles montrent une bonne capacité à discriminer les dépassements des seuils de quantiles observés 80% et de meilleurs scores de Brier en comparaison à la prévision climatologique jusqu'à l'échéance de 4 heures. La prévision de pluie persistante obtient de meilleurs scores que les ensembles de pluie pour la première échéance uniquement, à l'exception du score de discrimination ROCS qui est meilleur pour la prévision AROMEPI-timelag.

b) Qualité des prévisions d'ensembles de débit

Comme la chaîne d'avertissement testée dans cette étude est basée sur le dépassement d'un seuil relatif de débit, les prévisions horaires de débit sont vérifiées pour la variable de débit maximal sur les périodes de 1hr, 3hr et 6hr en comparaison avec les débits simulés. Ceci permet de mesurer l'impact des erreurs des prévisions de pluie sur les prévisions de débit. On rappelle également que les ensembles de pluie AROMEPI-timelag sont complétés avec de la pluie nulle pour les échéances non couvertes par les prévisions AROME-PI plus anciennes qui sont utilisées pour construire les prévisions AROMEPI-timelag (cf. Figure 7). De plus, comme les 56 bassins versants ont des surfaces très variables, on utilise la variable de débit spécifique (défini par le débit divisé par la surface du bassin) pour pouvoir agréger les scores sur tous les bassins.

Comme pour les pluies de bassin, les scores de vérification sont calculés pour chaque bassin pour toutes les paires observations-prévisions, ainsi que pour les paires dont les valeurs de débit simulé (utilisées comme « observations » pour cette analyse) sont supérieures aux quantiles de débit simulés 80% ; les scores sont ensuite moyennés sur les 56 bassins. Les échantillons sont en moyenne pour chaque bassin de 432 paires au total, et de 89 paires pour les dépassements du quantile simulé 80%. Il faut noter que le quantile simulé 80% défini à partir des simulations de débit de 2014 a une valeur inférieure au seuil relatif de débit biennal, qui est le seuil utilisé pour l'analyse de la qualité des avertissements. Mais la taille réduite de l'échantillon pour ce quantile simulé 80% n'a pas permis d'analyser les résultats pour un quantile plus élevé.

Erreurs des ensembles de débit en fonction du débit simulé

La Figure 13 présente les histogrammes des erreurs des prévisions d'ensemble en fonction du débit simulé pour identifier les éventuels biais ou valeurs aberrantes des prévisions. Les erreurs sont calculées pour chacun des 6 membres de la prévision et sont représentées avec les valeurs des quartiles, le minimum et le maximum. Les graphes d'erreur ci-dessous incluent les résultats des 56 bassins.

Le graphe concernant la première échéance montre des erreurs limitées des ensembles, qui s'expliquent par la persistance du débit à cette très courte échéance, mais également des sous-estimations pour les débits simulés moyens et forts, dus aux erreurs de sous-estimation des ensembles de pluie (cf. les résultats d'erreur relative moyenne en Figure 11). Les graphes pour les échéances de 3hr et 6hr montrent une sous-estimation marquée des valeurs les plus fortes, sous-estimation qui se dégrade avec l'échéance, et également une surestimation de certains débits moyens. Cette dégradation marquée pour la plus longue échéance vient notamment des valeurs de pluie nulle utilisées pour les plus longues échéances pour 5 des 6 membres de l'ensemble et du biais conditionnel des pluies (cf. Figures 10 et 11). Il faut remarquer que le dernier graphe relatif au débit maximal sur les 6 heures d'échéance montre les erreurs de surestimation et sous-estimation des débits moyens et forts.

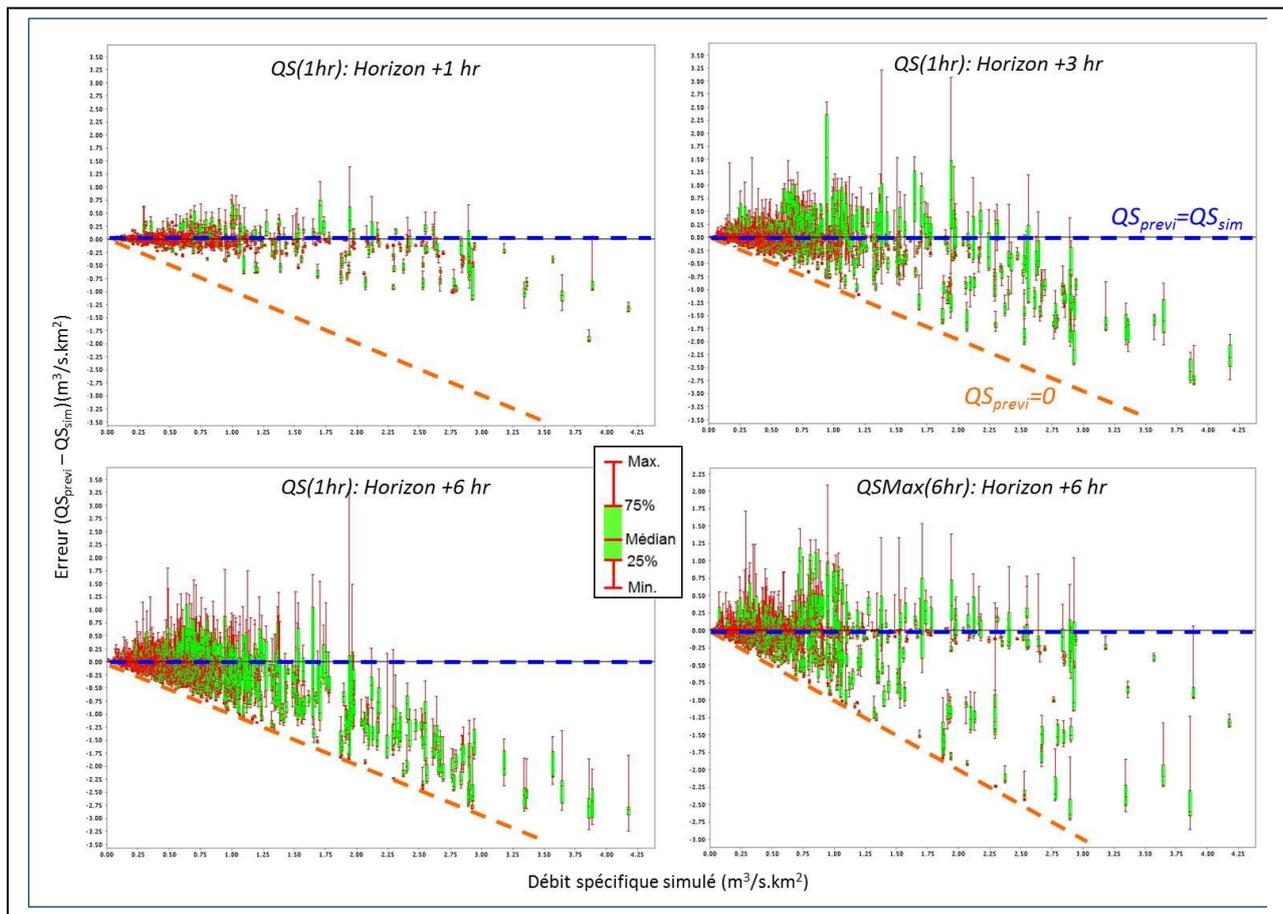


Figure 13 : Histogrammes des erreurs des ensembles de débit spécifique en fonction du débit spécifique simulé pour les horizons de 1hr, 3hr et 6hr pour les prévisions horaires et pour les prévisions de débit maximal sur 6 heures

Impact de l'échéance sur la qualité des ensembles de débit

La Figure 14 présente les scores d'évaluation en fonction de l'échéance de prévision. On calcule également les scores relatifs aux quantiles simulés 80%.

L'erreur relative moyenne en Figure 14 de la moyenne des ensembles de débit horaire montre un biais positif très faible sur les 2 premières échéances (on rappelle que les débits prévus sont vérifiés avec les débits simulés), mais qui devient négatif et se dégrade avec l'échéance : la sous-estimation est de 12% pour l'échéance de 6hr en ce qui concerne les débits simulés les plus forts. Ces biais sont significativement meilleurs que ceux relatifs au débit issu de la pluie persistante, atteignant jusqu'à -27%. Les coefficients de corrélation, proches de 1 pour la première échéance pour les 2 prévisions (ensembles AROMEPI-timelag et persistance), se dégradent avec l'échéance, avec de meilleurs scores pour les ensembles AROMEPI-timelag que pour la persistance. Les scores de Brier et les scores de discrimination ROCS montrent des gains très significatifs sur toutes les échéances, même s'ils se dégradent, et ces gains restent supérieurs à ceux obtenus pour la persistance.

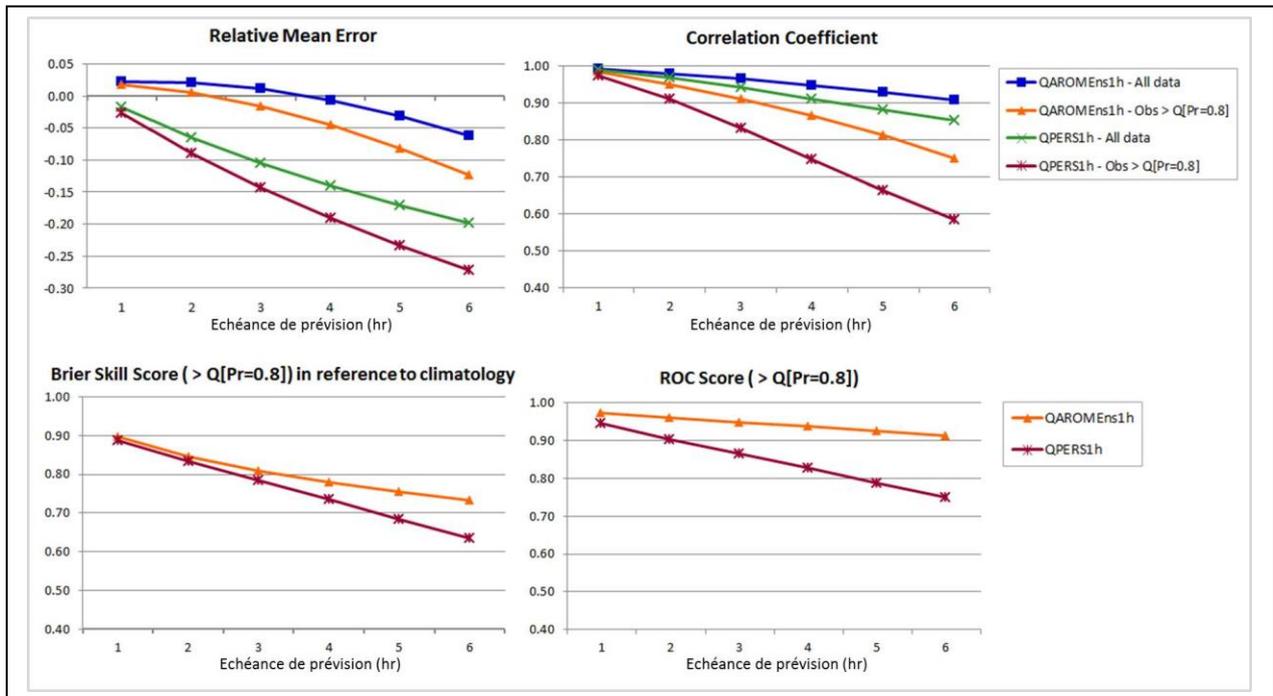


Figure 14 : Impact de l'échéance sur les scores de vérification des prévisions de débit issues des ensembles AROMEPI-timelag et de la pluie persistante

Impact de la période d'agrégation sur la qualité des ensembles de débit

En Figure 15, les scores d'évaluation sont comparés pour les différentes périodes d'estimation du pic de débit de 3hrs et 6 hrs pour les débits simulés dépassant le quantile 80%. On inclut également les scores pour le pic de la prévision déterministe de débit issu de la prévision immédiate de pluie sur les 6 heures d'échéance (qui correspond au premier membre de l'ensemble). Cette comparaison des scores permet de juger l'apport potentiel de l'intégration des prévisions déterministes plus anciennes pour décrire de manière pragmatique l'incertitude des pluies.

L'erreur relative moyenne des moyennes des ensembles de débit maximal montre un biais nul pour l'échéance de 3hr mais un biais négatif de 8% pour l'échéance de 6hr, qui montre l'impact significatif de l'intégration des valeurs de prévision nulle pour 5 des 6 membres de l'ensemble sur les échéances les plus longues. En comparaison, la prévision déterministe de débit issu de la prévision AROME-PI a un biais nul pour l'échéance de 6hr, qui s'explique sans doute par la compensation entre les erreurs de surestimation des débits moyens et les erreurs de sous-estimation des débits plus forts. L'agrégation permet de dégrader plus faiblement les coefficients de corrélation, le score étant similaire pour les ensembles AROMEPI-timelag et la prévision déterministe pour l'échéance de 6hr, et meilleur que celui de persistance. Pour le gain en score de Brier et le score de discrimination ROCS, les résultats sont meilleurs pour les ensembles AROMEPI-timelag que pour la prévision déterministe, ce qui indique que les ensembles AROMEPI-timelag peuvent permettre d'améliorer la détection des dépassements de seuil.

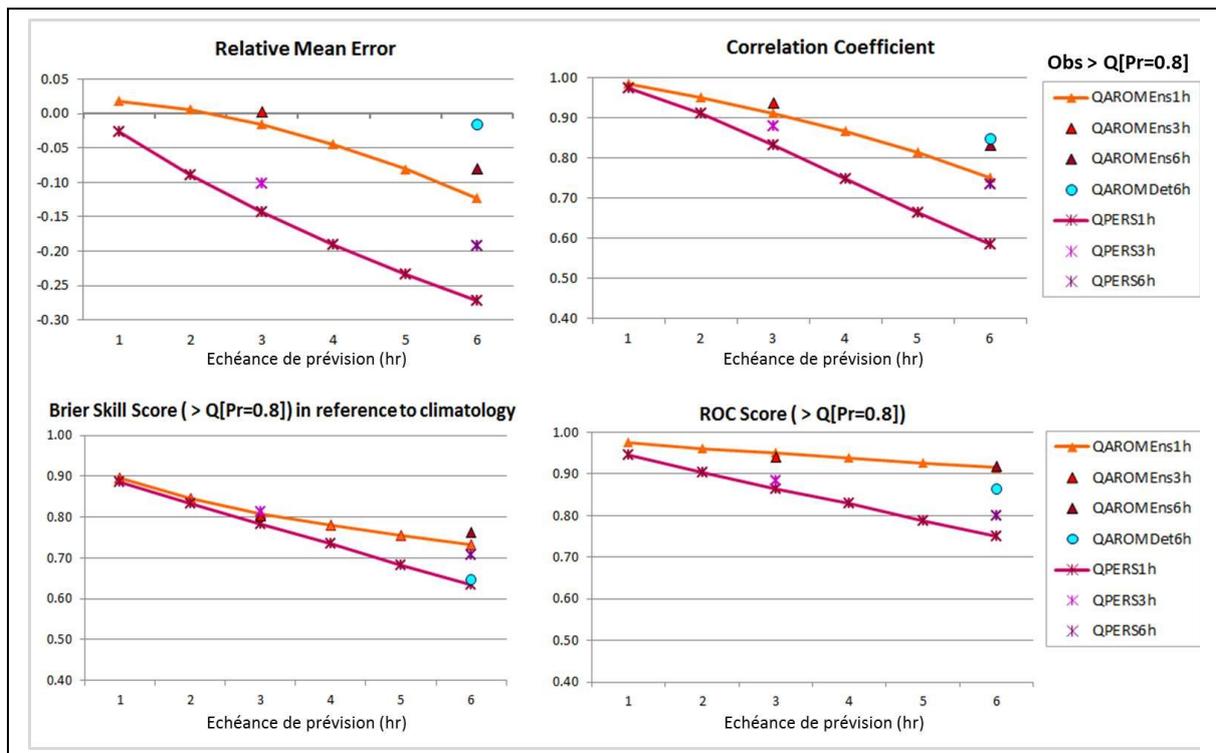


Figure 15 : Impact de l'agrégation sur les scores de vérification des prévisions de débit issues des ensembles AROMEPI-timelag et de la pluie persistante

En conclusion, l'évaluation des prévisions d'ensemble a montré un biais négatif des ensembles AROMEPI-timelag de débit pour les échéances les plus longues dû aux sous-estimations des pluies les plus fortes par le système AROME-PI et à l'intégration des pluies nulles pour certains des membres de l'ensemble. Mais on note également des compensations entre sous-estimation des débits les plus forts et surestimation des débits moyens. La corrélation des moyennes des ensembles se dégrade légèrement avec l'échéance pour les pluies et les débits mais reste supérieure à celle des prévisions de persistance. En comparaison avec la prévision probabiliste climatologique (pour toutes les échéances) et la prévision persistante (pour les échéances de plus 2 heures), les ensembles ont de meilleures performances pour la détection des dépassements de seuil, ce qui semble prometteur pour la qualité des avertissements basées sur les prévisions AROMEPI-timelag de débit. Les résultats sont relativement similaires entre la prévision persistante et la prévision AROMEPI-timelag pour la première heure d'échéance de prévision. Cependant, ces résultats sont basés sur les 56 bassins les plus touchés par les 3 événements et sur les dépassements correspondants aux quantiles simulés de débit de 80%, dont les valeurs sont largement inférieures aux valeurs de seuil relatif de débit utilisées pour générer les avertissements dans le système AIGA.

c) Qualité des avertissements probabilisés

Cette section présente les résultats de l'analyse de la qualité des avertissements pour le dépassement du seuil relatif de débit biennal pour les 185 bassins de la zone Grand Sud, en se basant sur les scores de contingence et l'anticipation des dépassements.

Scores de contingence pour les avertissements

La qualité des avertissements est analysée avec les prévisions suivantes de débit sur les échéances de N heures ($N=6$ en Figure 16 et $N=3$ en Figure 17) :

- les prévisions d'ensemble pour 3 niveaux de probabilité différentes, notées Qens1/6_echNh (en considérant qu'il y a dépassement si 1 seul des 6 membres est supérieur au seuil de débit), Qens3/6_echNh (3 membres sur 6 indiquant le dépassement ou non du seuil de débit), et Qens6/6_echNh (le dépassement est prévu lorsque les 6 membres sont supérieurs au seuil de débit),
- la prévision déterministe AROME-PI (notée Qdet_echNh),
- la prévision issue de la pluie future nulle (Qpnul_echNh) (l'option est disponible dans l'outil AIGA depuis 2014),
- la prévision de débit sans pluie future (Qsim), correspondant au système AIGA actuel du Sud de la France, mais en utilisant un seuil relatif de débit,
- les 2 prévisions de référence, la prévision persistante (notée Qpers_echNh, l'option étant également disponible dans l'outil AIGA depuis 2014) et la prévision avec la pluie observée ou pluie « future parfaite » (Qpobs_echNh).

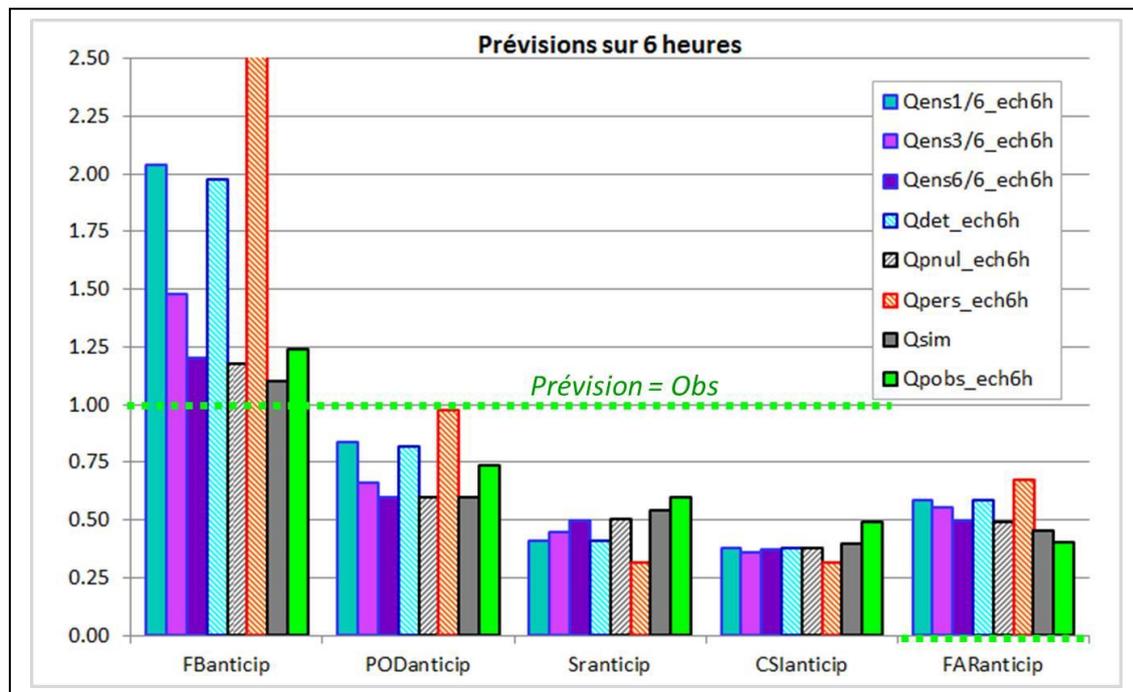


Figure 16 : Scores de contingence des avertissements de dépassement du débit biennal avec les prévisions de débit sur une échéance de 6 heures, issues des ensembles de pluie AROMEPI-timelag, de la prévision déterministe AROME-PI, de la pluie future nulle, de la pluie persistante, de la prévision de débit sans pluie future et de la pluie observée

La Figure 16 présente les scores de contingence (en prenant en compte l'anticipation) obtenus pour les 3 événements étudiés sur les 185 bassins du Grand Sud. Si l'on considère la

prévision de référence issue de la pluie observée (Q_{pobs_ech6h}), le système a une légère tendance à générer plus de fausses alertes que d'alertes manquées ; le score de détection $POD_{anticip}$ des dépassements avec une anticipation non nulle est de 74%, le score de $CSI_{anticip}$ de 49% et le taux de fausses alertes de 40%. Pour le système AIGA actuel, les scores de contingence de prévision de débit sans pluie future (Q_{sim}) sont dégradés par un plus fort taux de fausses alertes ($FAR_{anticip}=45\%$), avec un score de détection de 60% et un score de $CSI_{anticip}$ de 40%. La prévision avec pluie future nulle obtient des résultats très similaires à la prévision de débit sans pluie future mais en dégradant légèrement le taux de fausses alertes, avec un score de détection de 60% et un score de $CSI_{anticip}$ de 38%. La prévision de pluie persistante sur 6hr permet d'obtenir un score de détection proche de 1 mais au détriment d'un très fort taux de fausses alertes ($FAR_{anticip}=68\%$) et donc une dégradation du $CSI_{anticip}$ pour la valeur de 32%.

La prévision déterministe AROME-PI et la prévision ensembliste AROMEPI-timelag n'utilisant qu'un seul membre pour déclencher un avertissement, obtiennent des scores de contingence très similaires. Elles ont tendance à générer trop de fausses alertes en comparaison des alertes manquées (biais en fréquence proche de 2), ce qui améliore le score de détection $POD_{anticip}$ (supérieur à 82%) mais au détriment de la fiabilité ($SR_{anticip}=41\%$ contre 55% pour le système AIGA actuel) et du score $CSI_{anticip}$ (d'une valeur de 38%). Les deux autres prévisions probabilistes, prenant en compte 3 ou 6 membres de l'ensemble pour déclencher les avertissements, permettent d'améliorer les scores en réduisant le taux de fausses alertes avec un taux de détection supérieur à 60%. Mais, vis-à-vis du système AIGA actuel, elles dégradent la fiabilité ($SR_{anticip}$ inférieur à 50%) et le score de $CSI_{anticip}$ (inférieur à 38%).

La Figure 17 permet de comparer uniquement les scores de probabilité de détection et de fiabilité (avec le taux de succès) en utilisant les prévisions pour les échéances de 6hrs (graphe de gauche) et de 6hrs et 3hrs (graphe de droite). La prévision la plus favorable doit avoir des valeurs proches de 1 pour ces 2 scores.

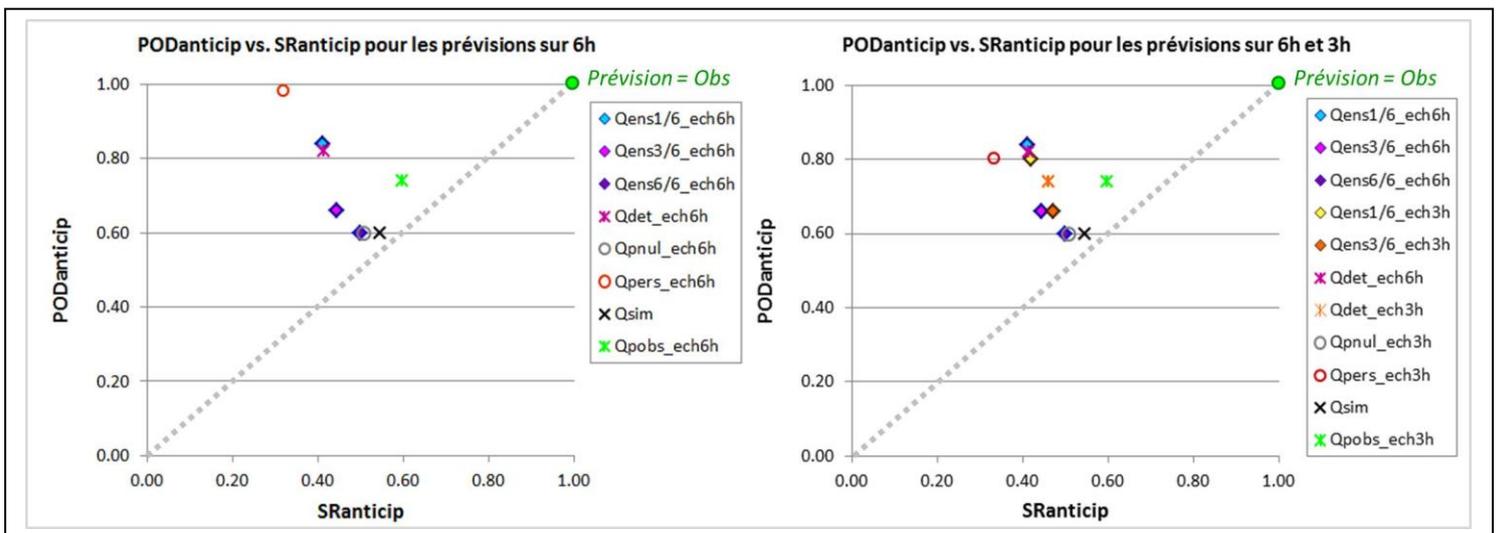


Figure 17 : Comparaison de la Probabilité de Détection et du Taux de Succès pour les avertissements de dépassement du débit biennal avec les prévisions de débit sur les échéances de 6 et 3 heures, issues des ensembles de pluie AROMEPI-timelag, de la prévision déterministe AROME-PI, de la pluie future nulle, de la pluie persistante, de la prévision de débit sans pluie future et de la pluie observée

Ces 2 graphes montrent clairement que les prévisions AROME-PI et AROMEPI-timelag permettent un gain en taux de détection mais pour une légère perte de fiabilité vis-à-vis de la

prévision de débit sans pluie future. Les scores pour les prévisions AROME-PI et AROMEPI-timelag sont légèrement améliorés en termes de fiabilité lorsque l'on considère une échéance de 3hr au lieu de 6hr.

Cette première évaluation serait à compléter par l'analyse d'autres événements sur un plus grand nombre de bassins pour obtenir des statistiques plus robustes. Elle est complétée ci-dessous par l'analyse de l'amélioration potentielle du degré d'anticipation des dépassements observés et simulés.

Degré d'anticipation des dépassements

L'analyse du degré d'anticipation des dépassements prévus est réalisée à partir de l'échantillon des 64 dépassements simulés pour les 3 événements étudiés et pour les 37 dépassements observés et simulés. Pour calculer le gain en anticipation pour une prévision vis-à-vis du système AIGA actuel, seuls les dépassements prévus par la prévision sans pluie future sont inclus. La Figure 18 représente les statistiques de minimum, maximum et moyenne des gains en anticipation vis-à-vis du système AIGA actuel pour les différentes prévisions pour l'échéance de 6hrs.

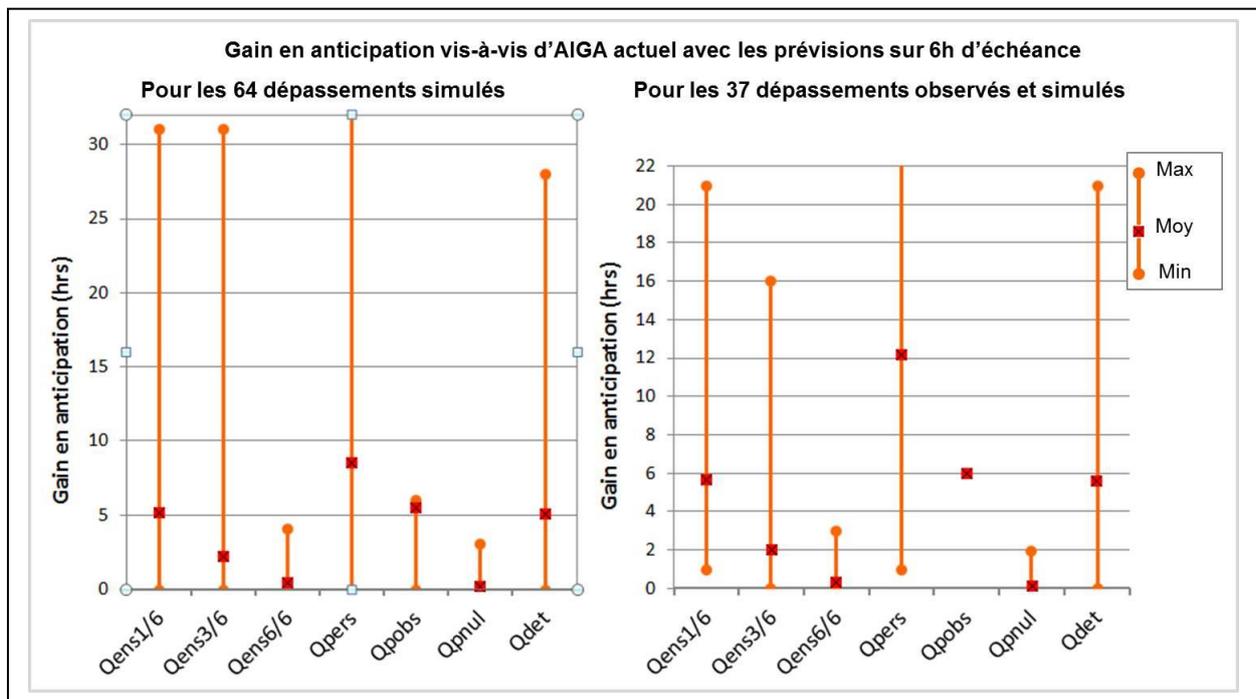


Figure 18 : Statistiques des gains en anticipation des dépassements du débit biennal avec les prévisions de débit sur l'échéance de 6 heures, issues des ensembles de pluie AROMEPI-timelag, de la pluie persistante, de la pluie observée, de la pluie future nulle et de la prévision déterministe AROMEPI

La moyenne du gain en anticipation avec la prévision AROMEPI-timelag considérant un seul des 6 membres pour les dépassements et avec la prévision déterministe AROME-PI sont du même ordre que le gain moyen pour la prévision issue de la pluie parfaite : un gain en anticipation de 5 heures pour les 64 dépassements simulés, et de 6 heures pour les 37 dépassements observés et simulés. Ce gain moyen est réduit à 2 heures environ pour la prévision AROMEPI-timelag considérant 3 des 6 membres pour les dépassements, et pour moins d'1 heure lorsque les 6 membres de l'ensemble définissent les dépassements. Il faut noter que les très fortes valeurs d'anticipation pour les prévisions AROMEPI-timelag et AROME-PI, ainsi que pour les prévisions de persistance, sont liées à une surestimation des

pluies pour certains bassins, qui peut également être aggravée par un modèle hydrologique trop réactif.

En conclusion, cette analyse sur l'amélioration du degré d'anticipation montre l'intérêt d'inclure, dans la modélisation hydrologique du système AIGA, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag pour améliorer significativement l'anticipation des dépassements de seuil vis-à-vis des ceux du système actuel sans pluie future. Cependant l'analyse des scores de contingence a mis en évidence une tendance à générer plus de fausses alertes que d'alertes manquées avec les prévisions AROME-PI et AROMEPI-timelag, ce qui entraîne une meilleure détection des dépassements au détriment de la fiabilité. L'utilisation des prévisions d'ensemble AROMEPI-timelag prenant en compte 3 ou 6 membres de l'ensemble pour déclencher les avertissements permet de contrebalancer cette tendance à générer des fausses alertes plutôt que des alertes manquées. Dans une moindre mesure, la réduction de l'échéance de prévision à 3 heures permet également d'obtenir une légère amélioration. Le taux de détection obtenu est supérieur à 60%, mais la fiabilité reste légèrement inférieure à celle du système AIGA actuel (perte de l'ordre de 5%). Plus généralement, il est nécessaire d'analyser le taux de détection et la fiabilité des dépassements prévus en fonction de l'échéance de prévision et du niveau de probabilité pour définir les dépassements avec les prévisions d'ensemble, afin de gagner en anticipation et en détection, tout en conservant une fiabilité comparable au système AIGA actuel.

5. Conclusions et Perspectives

L'analyse de l'apport des prévisions de pluie immédiate a porté sur 3 événements de crue de septembre 2014, mi-novembre 2014 et fin novembre 2014. Les simulations et prévisions de débit sont issues du modèle hydrologique semi-distribué du futur système d'avertissement AIGA, avec les paramètres calibrés et régionalisés de la fin 2014 (Organde 2014). L'analyse a concerné les 185 bassins versants jaugés dans la zone Grand Sud, zone touchée par les 3 événements étudiés. Pour analyser l'apport des prévisions déterministes AROME-PI et prévisions d'ensemble AROMEPI-timelag, différentes prévisions déterministes de référence sont également évaluées : la prévision de pluie persistante, la prévision de pluie « parfaite » (i.e. pluie observée) et la prévision de pluie nulle.

Comme pour l'étude portant sur l'amélioration de la modélisation hydrologique d'Organde en 2015, l'évaluation de la qualité des avertissements s'est basée sur le dépassement d'un seuil de débit relatif correspondant à une période de retour de l'ordre de 2 ans. Cette approche permet de « débiaiser » les sorties du modèle hydrologique. L'évaluation a inclus différents critères de contingence pour décrire la fiabilité, la précision et le biais en fréquence des avertissements de dépassement de seuil de débit, ainsi que l'anticipation des dépassements. Par ailleurs, les prévisions de pluie et de débit ont été évaluées avec le système de vérification EVS (Brown et al. 2010) en sélectionnant des critères pour décrire le biais relatif et la corrélation des moyennes des ensembles avec les observations, ainsi que le gain en termes d'erreur en probabilité et de discrimination pour les dépassements de quantiles. L'évaluation a concerné les 56 bassins les plus touchés par les 3 événements. Comme la chaîne d'avertissement testée dans cette étude est basée sur le dépassement d'un seuil relatif de débit, les prévisions de débit ont été comparées aux débits simulés.

Les résultats de l'évaluation des prévisions de pluie de bassin ont permis d'identifier le biais conditionnel des moyennes des ensembles de pluie, qui ont tendance à surestimer les faibles pluies et à significativement sous-estimer les fortes pluies. Mais les ensembles montrent une

bonne capacité à discriminer les dépassements des seuils de quantiles observés 80% (issus de la chronique de pluie observée sur l'année 2014), notamment pour les 4 premières heures d'échéance. Pour la comparaison des prévisions de débit avec les débits simulés, l'évaluation a montré un biais négatif des ensembles AROMEPI-timelag pour les échéances les plus longues dû aux sous-estimations des pluies AROME-PI et à l'intégration des pluies nulles pour certains des membres de l'ensemble. Cependant, on note également des erreurs de surestimation pour les débits moyens et certains débits forts. En comparaison avec la prévision probabiliste climatologique (pour toutes les échéances) et la prévision persistante (pour les échéances de plus 2 heures), les ensembles ont de meilleures performances pour la détection des dépassements de seuil. Les résultats sont relativement similaires entre la prévision persistante et la prévision AROMEPI-timelag pour la première heure d'échéance de prévision. Du fait de la taille modeste de l'échantillon issu des 3 événements de crue, l'analyse des dépassements prévus a concerné le seuil des quantiles simulés de débit de 80% (estimés à partir des simulations de l'année 2014), qui ont des valeurs inférieures aux valeurs de seuil relatif de débit utilisées pour générer les avertissements dans le système AIGA (seuil relatif basé sur le quantile empirique de période de retour de l'ordre de 2 ans).

L'évaluation des avertissements basés sur les prévisions AROME-PI et AROMEPI-timelag en comparaison au système AIGA actuel (qui n'intègre pas de prévision de pluie) a montré une amélioration en termes de détection (gain de l'ordre de 20%) mais avec une légère dégradation du taux de fausses alertes et donc une légère perte de fiabilité (perte de l'ordre de 13%). L'utilisation des prévisions d'ensemble AROMEPI-timelag, prenant en compte 3 ou 6 membres de l'ensemble pour déclencher les avertissements, permet de contrebalancer cette tendance à générer des fausses alertes plutôt que des alertes manquées ; l'utilisation des prévisions AROME-PI et AROMEPI-timelag sur une échéance de 3 heures permet également, dans une moindre mesure, d'obtenir une légère amélioration. Le taux de détection obtenu est supérieur à 60%, mais la fiabilité reste légèrement inférieure à celle du système AIGA actuel (perte de l'ordre de 5%). En termes d'anticipation, les gains avec la prévision déterministe AROME-PI et la prévision d'ensemble AROMEPI-timelag pour l'échéance de 6 heures sont très significatifs : gain moyen en anticipation de 5 heures pour les 64 dépassements simulés par AIGA et de 6 heures pour les 37 dépassements observés.

En conclusion, cette étude montre l'intérêt d'inclure, dans la modélisation hydrologique du système AIGA, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag, pour améliorer significativement l'anticipation des dépassements de seuil vis-à-vis de ceux du système actuel sans pluie future. L'approche pragmatique de construction de prévisions d'ensemble en utilisant les prévisions déterministes successives a permis de définir des avertissements avec différents niveaux de probabilité afin de moduler le niveau de probabilité déclenchant l'avertissement. Cependant, il est nécessaire d'analyser le taux de détection et la fiabilité des dépassements prévus en fonction de l'échéance de prévision et du niveau de probabilité pour définir les dépassements avec les prévisions d'ensemble, afin de gagner en anticipation tout en conservant une fiabilité comparable au système AIGA actuel. Ce type d'optimisation de la qualité des avertissements, pour une meilleure détection des dépassements avec une anticipation améliorée et une fiabilité acceptable, doit également prendre en compte les enjeux pour l'ensemble de la chaîne d'avertissement et les besoins des utilisateurs.

Des pistes d'amélioration peuvent être proposées pour compléter cette analyse et pour les améliorations futures de la chaîne d'avertissement AIGA :

- il serait nécessaire de poursuivre l'évaluation des performances avec un plus grand nombre d'événements pour travailler avec des échantillons plus importants et obtenir

ainsi des scores plus robustes, et pour analyser les performances pour des bassins répartis sur l'ensemble du territoire ; cela permettrait également d'étudier les dépassements pour des seuils de périodes de retour plus grandes ; il faut noter qu'il est souhaitable que les observations de débit soient préalablement critiquées, par le réseau de SPC notamment, pour intégrer des informations de qualité satisfaisante pour les hautes eaux ; l'évaluation de la qualité des avertissements devrait aussi utiliser des retours post-crue sur les dégâts observés sur le terrain pour ne pas limiter l'analyse des résultats aux bassins jaugés ;

- les seuils de débit utilisés dans cette étude pourraient être modulés en fonction des enjeux, pour une prise en compte de la vulnérabilité et fournir ainsi des avertissements sur le risque (par croisement entre les informations d'aléa et de vulnérabilité) ;
- la chaîne d'avertissement AIGA devrait bénéficier de la combinaison des différentes prévisions de type déterministe (prévisions pouvant définir des ensembles timelag par décalage temporel), ensembliste ou probabiliste ; les prévisions immédiates pourraient être complétées par des prévisions produites par les modèles atmosphériques sur des échéances plus longues, notamment les prévisions d'ensemble AROME-PE ou COSMO-DE-EPS ; les incertitudes sur les pluies observées pourraient également être prises en compte (cf. le travail de thèse d'Angelica Caseri, Caseri et al. 2015), ce qui permettrait d'améliorer la dispersion (et donc la fiabilité) des ensembles de débit pour les toutes premières échéances ;
- les incertitudes hydrologiques ont un impact significatif sur les résultats, notamment du fait de l'application d'un modèle semi-distribué horaire en milieu non-jaugé après régionalisation des paramètres ; les améliorations du modèle pluie-débit concernent notamment l'initialisation du modèle, grâce à l'intégration des informations sur les derniers débits observés pour des bassins jaugés voisins et cette assimilation de données peut également permettre d'initialiser le modèle de manière ensembliste (Demargne 2016) ; la prise en compte des incertitudes de paramétrisation du modèle pourrait également être poursuivie pour travailler avec différents jeux de paramètres équiprobables ; plus généralement, toute amélioration de la modélisation hydrologique (par exemple, prise en compte des processus de neige, modélisation au pas de temps infra-horaire) et de la chaîne d'avertissement AIGA pourra potentiellement conduire à une meilleure prise en compte des incertitudes.

6. Références

- Arnaud P., Y. Aubert, D. Organde, P. Cantet, C. Fouchier, N. Folton, 2014. Estimation de l'aléa hydro-météorologique par une méthode par simulation: La méthode SHYREG : Présentation – Performances – Base de données, *La Houille Blanche* n°2, 20-26.
- Aubert Y., P. Arnaud, P. Ribstein et J. A. Fine, 2014. La méthode SHYREG débit, application sur 1605 bassins versants en France Métropolitaine. *Hydrological Sciences Journal*, 59 (5), 1-13.
- Auger L., Dupont O., Hagelin S., Brousseau P., Brovelli P., 2015. AROME–NWC: a new nowcasting tool based on an operational mesoscale forecasting system. *Quarterly Journal of the Royal Meteorological Society*, 141: 1603–1611, doi:10.1002/qj.2463
- Ben Bouallègue Z., Theis S. E., Gebhardt C. 2013. Enhancing COSMO-DE-EPS ensemble forecasts by inexpensive techniques. *Meteorologische Zeitschrift*, 22(1), 49-59.
- Brown J. D., Demargne J., Seo D.-J., et Y. Liu, 2010. The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling Software*, 25, 854-872.
- Casari A., Javelle P., Ramos M.H., Leblois E., 2015. Generating precipitation ensembles for flood alert and risk management. *Journal of Flood Risk Management* , DOI: 10.1111/jfr3.12203.
- Cloke, H.L., et F. Pappenberger, 2009. Ensemble flood forecasting: a review. *Journal of Hydrology*, 375, 613-626.
- Demargne J., 2013. Application de la méthode de Flash Flood Guidance avec AIGA pour la prévision des crues éclairs en milieu non jaugeé. Convention DGPR/SCHAPI – IRSTEA 2012, *Rapport d'étude HYDRIS Hydrologie*, 63 p.
- Demargne J. 2014. Amélioration de l'approche AIGA par l'intégration des pluies prévues. Convention DGPR/SCHAPI – IRSTEA 2013, *Rapport d'étude HYDRIS Hydrologie*, 71 p.
- Demargne J. 2015. Intégration des pluies prévues du système COSMO-DE-EPS dans le système AIGA. Convention DGPR/SCHAPI – IRSTEA 2014, *Rapport d'étude HYDRIS Hydrologie*, 56 p.
- Demargne J. 2016. Assimilation des débits dans le modèle hydrologique du système AIGA. Programme MEDDE (DGPR / SRNH) - IRSTEA 2015, *Rapport d'étude HYDRIS Hydrologie*, 37 p.
- Demargne, J., Wu, L., Regonda, S., Brown, J.D., Lee, H., He, M., Seo, D.-J., Hartman, R., Fresch, M., Schaake, J., et Zhu, Y. 2014. The science of NOAA's operational Hydrologic Ensemble Forecast Service. *Bulletin of American Meteorological Society*, 20 p. doi: 10.1175/BAMS-D-12-00081.1
- Edijatno et Michel C., 1989. Un modèle pluie-débit à trois paramètres. *La Houille Blanche*, 2, 113-121.
- Edijatno, Nascimento N.O., Yang X., Makhlof Z. et Michel C., 1999. GR3J : a daily watershed model with three free parameters. *Hydrological Sciences Journal*, 44(2), 263-278.

- Javelle P., Fouchier C., Arnaud P. et Lavabre J., 2010. Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *Journal of Hydrology*, 394(1-2), 267-274.
- Javelle P., Demargne J., Defrance D., Arnaud P., 2014. Evaluating flash flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system. *Hydrological Sciences Journal*. doi: 10.1080/02626667.2014.923970
- Jolliffe I.T., Stephenson, D.B., 2003. Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley and Sons, Chichester, England, 2003, 254 p.
- National Research Council of the National Academies (NRC), 2006. Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. The National Academies Press, 124 pp.
- Organde D., Arnaud P., Fine J.-A., Fouchier C., Folton N. et Lavabre J., 2013. Régionalisation d'une méthode de prédétermination de crue sur l'ensemble du territoire français : la méthode SHYREG. *Revue des Sciences de l'Eau*, 26 (1) : 65-78.
- Organde D., 2013. Extension nationale de la méthode AIGA, Évaluation de AIGA débit sur un échantillon de bassins versants du territoire national. Convention DGPR/SCHAPI – IRSTEA 2013, *Rapport d'étude HYDRIS Hydrologie*, 39 p.
- Organde D., 2014. Extension nationale de l'approche AIGA, Amélioration de la réactivité du modèle. Convention DGPR/SCHAPI – IRSTEA 2014, *Rapport d'étude HYDRIS Hydrologie*, 20 p.
- Organde D., 2015. Mise à jour de la paramétrisation de la méthode AIGA et calcul des quantiles de débit issus des chroniques simulées. Convention DGPR/SCHAPI – IRSTEA 2015, *Rapport d'étude HYDRIS Hydrologie*, 19 p.
- 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. *Journal of Hydrology*, 303(1-4), 290-306.
- Perrin C., 2002. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. *La Houille Blanche*, 6-7, 84-91.
- Perrin C., Michel C., et Andréassian V., 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275-289.
- Tabary P. 2007. The New French operational radar rainfall product. Part I: Methodology. *Weather and Forecasting*, 22(3), 393-408.
- Tabary P. et al. 2011. Evaluation of two integrated polarimetric Quantitative Precipitation Estimation (QPE) algorithms at C-band. *Journal of Hydrology*, 405(3-4), 248-260.
- Wilks D.S., 2006. Statistical Methods in Atmospheric Sciences, Academic Press, San Diego, California, 627 p.
- World Meteorological Organization (WMO), 2013. Forecast Verification – Issues, Methods and FAQ, WWRP/WGNE Joint Working Group on Verification, <http://www.cawcr.gov.au/projects/verification/>
- Zappa, M., et co-auteurs, 2010. Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. *Atmospheric Science Letters*, 11 (2), 83-91.

Appendice : Liste des Figures

Figure 1 : Principales sources d’incertitude dans la chaîne d’avertissement sur la possibilité de crues soudaines du système AIGA.....	8
Figure 2 : Approche méthodologique pour évaluer l’apport des prévisions AROME-PI pour le système AIGA	10
Figure 3 : Carte des 719 bassins versants calibrés et régionalisés en France, dont 185 bassins de la zone Grand Sud ont été sélectionnés pour cette étude, avec leur distribution de surface	11
Figure 4 : Schéma du couplage des modèles journaliers continus SAJ et GR4J et du modèle horaire semi-distribué GR-SD pour le système AIGA national.....	12
Figure 5 : Résultats du calage et de la régionalisation du modèle horaire semi-distribué GR-SD.....	13
Figure 6 : Exemples de cartes de cumul des prévisions AROME-PI au cours des 3 événements étudiés	14
Figure 7 : Schéma de construction des ensembles timelag de pluie à partir des prévisions déterministes AROME-PI successives.....	15
Figure 8 : Exemple d’hydrogrammes avec le débit simulé, l’enveloppe et la moyenne des 6 ensembles de débit, la prévision de débit issu de la pluie persistante pour la date d’initialisation du 28 novembre 2014 à 17h (TU) en comparaison avec l’hydrogramme de débit observé pour le bassin de l’Alagnon à Joursac.....	17
Figure 9 : Estimation du Relative Operating Curve Score (ROCS) pour une prévision probabiliste	21
Figure 10 : Histogrammes des erreurs des prévisions d’ensemble de pluie de bassin en fonction de la pluie observée pour les cumuls sur 1hr et 3hrs.....	22
Figure 11 : Impact de l’échéance sur les scores de vérification des prévisions d’ensemble de pluie de bassin et de pluie persistante	23
Figure 12 : Impact de l’agrégation sur les scores de vérification des prévisions d’ensemble et de pluie persistante pour les cumuls de pluie de bassin sur 1hr, 3hr et 6hr	24
Figure 13 : Histogrammes des erreurs des ensembles de débit spécifique en fonction du débit spécifique simulé pour les horizons de 1hr, 3hr et 6hr pour les prévisions horaires et pour les prévisions de débit maximal sur 6 heures	26
Figure 14 : Impact de l’échéance sur les scores de vérification des prévisions de débit issues des ensembles AROMEPI-timelag et de la pluie persistante	27

Figure 15 : Impact de l'agrégation sur les scores de vérification des prévisions de débit issues des ensembles AROMEPI-timelag et de la pluie persistante	28
Figure 16 : Scores de contingence des avertissements de dépassement du débit biennal avec les prévisions de débit sur une échéance de 6 heures, issues des ensembles de pluie AROMEPI-timelag, de la prévision déterministe AROME-PI, de la pluie future nulle, de la pluie persistante, de la prévision de débit sans pluie future et de la pluie observée.....	29
Figure 17 : Comparaison de la Probabilité de Détection et du Taux de Succès pour les avertissements de dépassement du débit biennal avec les prévisions de débit sur les échéances de 6 et 3 heures, issues des ensembles de pluie AROMEPI-timelag, de la prévision déterministe AROME-PI, de la pluie future nulle, de la pluie persistante, de la prévision de débit sans pluie future et de la pluie observée	30
Figure 18 : Statistiques des gains en anticipation des dépassements du débit biennal avec les prévisions de débit sur l'échéance de 6 heures, issues des ensembles de pluie AROMEPI-timelag, de la pluie persistante, de la pluie observée, de la pluie future nulle et de la prévision déterministe AROMEPI.....	31