



**HAL**  
open science

# Évaluation de l'apport des pluies prévues par le système opérationnel AROME-PI dans le système Vigicrues-Flash

J. Demargne, P. Javelle

► **To cite this version:**

J. Demargne, P. Javelle. Évaluation de l'apport des pluies prévues par le système opérationnel AROME-PI dans le système Vigicrues-Flash. [Rapport de recherche] irstea. 2016, pp.59. hal-02606079

**HAL Id: hal-02606079**

**<https://hal.inrae.fr/hal-02606079>**

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 2016**  
**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 »**

**Evaluation de l'apport des pluies prévues  
par le système opérationnel AROME-PI  
dans le système Vigicrues-Flash**

**Auteur : Julie Demargne (HYDRIS-Hydrologie)**

**Suivi : Pierre Javelle (IRSTEA)**

**Décembre 2016**

**HYDRIS**  
*hydrologie*

## Table des Matières

Résumé.....	3
1. Introduction .....	6
a) Contexte .....	6
b) Travaux antérieurs pour le développement de Vigicrues-Flash.....	7
c) Prévisions immédiates de pluie AROME-PI .....	10
d) Objectifs de l'étude .....	12
2. Données et Modélisation Hydrologique .....	12
a) Sélection des évènements .....	13
b) Bassins versants.....	13
c) Modélisations hydrologiques .....	15
d) Données hydrométéorologiques .....	18
3. Méthodologies .....	18
a) Prévisions de pluie AROME-PI et ensembles timelag .....	18
b) Prévisions de débit et avertissements .....	20
c) Evaluation de la qualité des avertissements.....	23
4. Résultats d'anticipation des dépassements .....	27
a) Echantillons des dépassements de seuil sans prévision de pluie non nulle .....	28
b) Degré d'anticipation avec des prévisions de pluie déterministes.....	30
c) Degré d'anticipation avec des prévisions de pluie déterministes et ensemblistes .....	34
5. Scores de contingence des avertissements.....	39
a) Scores de contingence globale des avertissements .....	39
b) Scores de contingence horaire des avertissements .....	48
6. Conclusions et Perspectives.....	51
7. Références .....	55
8. Liste des Figures et Tableaux .....	57

## 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 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, appelé Vigicrues-Flash, qui devrait être complété par un service expertisé d'avertissement sur des bassins dits prioritaires, 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. La fine résolution d'AROME permet de modéliser les processus convectifs, qui sont essentiels pour les alertes aux crues rapides. 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, réactualisées toutes les heures (Auger et al. 2015). Ces prévisions sont produites actuellement à la résolution d'environ 1,3 km (0.01 degré lat/lon), et précédemment (jusqu'en avril 2015), à la résolution d'environ 2,5 km (0.025 degré lat/lon). Elles sont reprojctées sur une grille régulière de 1km de résolution dans la projection Lambert 2 étendu pour leur intégration dans le système d'avertissement. Les prévisions déterministes successives sont utilisées pour produire des prévisions d'ensemble au pas de temps horaire par décalage temporel, appelées AROMEPI-timelag, avec un maximum de 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 8 événements de crue de septembre 2014 jusqu'à février 2016. Les simulations et prévisions de débit sont issues du modèle hydrologique semi-distribué actuellement utilisée dans le système Vigicrues-Flash, avec les paramètres calibrés et régionalisés de 2015 (Organde 2014 et 2015), ainsi que la modélisation hydrologique distribuée continue développée dans l'étude d'Organde de 2016. L'analyse a concerné 781 bassins versants jaugés de surface comprise entre 5 km<sup>2</sup> et 1548 km<sup>2</sup>. 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 nulle, la prévision de pluie persistante (en maintenant sur les échéances de prévision de +1h à +6h, la dernière pluie observée au moment de l'initialisation de la prévision), et la prévision de pluie « parfaite » (i.e. pluie observée) pour montrer l'impact des erreurs et des incertitudes des prévisions d'ensemble de pluie.

L'évaluation de la qualité des alertes s'est basée sur le dépassement d'un seuil de débit relatif correspondant aux périodes de retour de 2 ans et 5 ans (l'échantillon étant trop réduit pour l'analyse de la période de retour de 10 ans). L'utilisation de seuils relatifs de débit 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.

Les scores de contingence sont estimés par 2 approches complémentaires :

- 1) sur chaque événement dans son ensemble (i.e. pour toutes les dates d'initialisation de prévision de l'évènement) pour la contingence dite globale, qui correspond à une

évaluation a posteriori du système d'avertissement, lorsque, pour un événement donné, on peut comparer le premier dépassement de seuil qui a été observé, au premier dépassement de seuil qui a été prévu ;

- 2) pour chaque pas de temps horaire des différents événements, pour la contingence dite horaire, correspondant à une évaluation plus stricte du système d'avertissement qui peut être réalisée en temps réel, dès que les observations de débit sont disponibles.

Pour la contingence globale, les scores de contingence prennent en compte l'anticipation des dépassements prévus : si le dépassement est prévu en retard vis-à-vis du dépassement observé, il est considéré comme une alerte manquée. On définit également une valeur maximale d'anticipation pour les dépassements simulés et prévus afin de comparer, de façon plus cohérente, un dépassement simulé avec le dépassement observé lui correspondant sans pénaliser l'anticipation du modèle hydrologique.

Par ailleurs, afin de prendre en compte les incertitudes sur les valeurs de seuils de débit, les scores de contingence ont également été estimés en intégrant une tolérance de 10% sur ces valeurs de seuil. Dans ce cas, on considère comme des alertes correctes, d'une part, les cas de dépassements observés et manqués de peu par la prévision (le débit prévu dépassant les 90% de la valeur du seuil mais restant inférieur au seuil), et, d'autre part, les cas de fausses alertes presque observées (pour des dépassements prévus et non observés, le débit observé dépassant les 90% de la valeur du seuil). Cette approche (proposée par le Centre d'Ecologie et d'Hydrologie de Wallingford en Grande-Bretagne) permet de juger de l'amélioration potentielle des performances dans une approche expertisée du système d'avertissement, pour laquelle un prévisionniste ou un utilisateur peut ajuster la production aux avertissements de crue ou la réaction à ces avertissements en abaissant la valeur du seuil.

L'évaluation a permis de comparer les avertissements basés sur les prévisions AROME-PI et AROMEPI-timelag avec les avertissements du système Vigicrues-Flash actuel (utilisant le modèle GRSD sans prévision de pluie) et les avertissements produits avec les différentes pluies futures. Les principaux résultats obtenus sont les suivants :

- les gains en anticipation avec la prévision déterministe AROME-PI et la prévision d'ensemble AROMEPI-timelag sont très significatifs pour les 2 modèles, même si le modèle GRSD permet une meilleure anticipation que le modèle GRDc (pour le seuil de 2 ans, anticipation médiane de +7h avec le modèle GRSD et +4h avec le modèle GRDc en intégrant AROME-PI déterministe). Egalement, la proportion de dépassements prévus en retard est significativement diminuée. L'anticipation est davantage améliorée lorsque les prévisions AROME-PI des 3 dernières heures sont intégrées dans les ensembles timelag, en complément de la prévision AROME-PI la plus récente.
- en contingence globale, les prévisions déterministes AROME-PI améliorent significativement la détection des dépassements mais avec, généralement, une augmentation du nombre de fausses alertes et donc, une perte de fiabilité (pour le seuil de 2 ans, la Probabilité de Détection s'améliore de 51% à 80% avec GRSD en comparaison au système actuel, alors que le taux de succès se dégrade de 48% à 34%). Les prévisions d'ensemble, avec les différents niveaux de probabilité (définis par le nombre de membres de l'ensemble dépassant le seuil de débit), permettent de moduler le gain en détection et la perte concomitante de fiabilité, les meilleurs scores de détection étant obtenus avec le niveau de probabilité le plus faible (i.e. un seul membre de l'ensemble dépassant le seuil de débit).

- pour la contingence horaire, plus stricte (car elle concerne tout l'hydrogramme de chacun des 8 événements), les scores de fiabilité sont globalement meilleurs qu'en contingence globale, alors que les scores de détection sont globalement plus faibles, en particulier pour le modèle GRSD. En comparaison à la prévision de pluie nulle, on obtient par exemple, pour le seuil de 2 ans et le modèle GRDc, une détection de 60% pour une fiabilité de 59% avec la prévision d'ensemble à 6 membres et le niveau de probabilité le plus faible (dépassement prévu par au moins un membre). Pour toutes les prévisions de pluie non nulles, et notamment pour les prévisions AROME-PI déterministes et ensemblistes, les scores de détection et de fiabilité sont meilleurs avec le modèle GRDc qu'avec GRSD, avec des écarts entre les 2 modèles plus importants en fiabilité qu'en détection.
- les résultats obtenus en intégrant, d'une part, les pluies AROME-PI et, d'autre part, les pluies observées, sur toute l'échéance de +6h ont montré l'impact des erreurs des prévisions de pluie AROME-PI sur les prévisions de débit et les dépassements de seuil, avec une tendance plus marquée à surestimer les plus fortes pluies (mais le biais n'est pas systématique). Ce biais entraîne un nombre plus important de fausses alertes et donc, une perte de fiabilité, mais permet également d'améliorer la détection et l'anticipation des dépassements observés.

En conclusion, cette étude montre l'intérêt d'inclure, dans la modélisation hydrologique du système Vigicrues-Flash, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag. Ces prévisions permettent d'améliorer significativement l'anticipation des dépassements de seuil vis-à-vis de ceux du système actuel sans pluie future, ainsi que le taux de détection, même si elles dégradent partiellement la fiabilité. 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 (défini par le nombre minimum de membres de l'ensemble dépassant le seuil). Il est nécessaire d'analyser le taux de détection et la fiabilité des dépassements prévus en fonction du niveau de probabilité pour définir les dépassements avec les prévisions d'ensemble, afin de gagner en anticipation tout en conservant une fiabilité acceptable. 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 sont également discutées en conclusion, notamment : la combinaison des prévisions immédiates avec des prévisions de pluie à plus longue échéance, la prise en compte des incertitudes de source hydrologique (en utilisant des prévisions d'ensemble multi-modèles et des jeux de paramètres équiprobables), l'amélioration des conditions initiales du modèle hydrologique, en particulier avec l'assimilation des débits observés, et l'utilisation des informations post-crue de dégâts pour une évaluation du système d'avertissement sur des bassins non jaugés.

Le présent rapport constitue le livrable 00050 de la commande n°4500271150 du 21/03/2016 du marché « Mise en œuvre de la méthode AIGA » n°2014-01-16, conclu entre IRSTEA UR RHAX et Hydris Hydrologie.

# 1. Introduction

## a) Contexte

Cette étude a pour objectif d'étudier l'apport des prévisions immédiates de pluie du système opérationnel AROME-PI de Météo-France pour le système d'avertissement des crues soudaines. Ce travail est réalisé pour le SCHAPI qui développe actuellement un nouveau service d'avertissement dit Vigicrues-Flash, permettant de compléter les alertes du système national de vigilance, Vigicrues, qui concerne uniquement quelques 22000 km de réseau hydrographique. Le système d'avertissement Vigicrues-Flash, basé sur la méthode AIGA, a pour objectif de fournir des avertissements sur la possibilité de crues soudaines pour des bassins non réglementaires, dont le temps de réaction est supérieur à 1h30, et des bassins pour lesquels des enjeux ont été identifiés concernant le risque de crue soudaine (de Saint Aubin et al. 2016). Le service d'avertissement Vigicrues-Flash est actuellement en phase de test avant son implémentation opérationnelle nationale en 2017 pour quelques 11 000 communes concernées.

Grâce aux développements récents de Météo-France et des autres agences de prévision météorologique, des prévisions de pluie à fine échelle spatio-temporelle et issues d'une meilleure modélisation des phénomènes de pluie intense par les modèles atmosphériques numériques peuvent être intégrées dans les systèmes d'alertes au crue, en cohérence avec l'échelle de modélisation hydrologique. Le système d'avertissement Vigicrues-Flash en cours d'implémentation intègre uniquement les dernières observations de pluie. Pour préparer l'intégration des prévisions de pluie dans une future version de Vigicrues-Flash, plusieurs études ont été réalisées pour démontrer les bénéfices potentiels en termes de détection et d'anticipation des dépassements de seuils de débit. L'étude de Demargne (2013) a évalué l'apport des prévisions de pluie à haute résolution issues du modèle opérationnel AROME (Application of Research to Operations at Mesoscale, Seity et al. 2011) de Météo-France pour améliorer les dépassements du seuil de débit biennal pour des bassins de Meuse-Moselle, Rhône-Cévennes, et Sud-Est. Ces prévisions de pluie de type déterministe sont générées par le modèle atmosphérique AROME, à l'échelle de 2,5 km environ jusqu'en avril 2015, puis de 1,3 km environ par la suite, au pas de temps horaire pour des échéances de 30h à 36h. Pour évaluer l'apport de prévision de pluie de type ensembliste, l'étude de Demargne (2014) a concerné les prévisions d'ensemble de pluie à haute résolution issues du modèle COSMO-DE-EPS de l'agence Deutscher Wetterdienst (DWD, dont les prévisions sont disponibles sur le Nord-Est de la France) pour montrer leur apport pour les bassins de Meuse-Moselle. Cette étude a également intégré 5 jeux jugés équiprobables des paramètres du modèle hydrologique semi-distribué en complément des prévisions d'ensemble de pluie pour générer un plus grand nombre de prévisions d'ensemble de débit prenant en compte l'incertitude de pluie future et l'incertitude du paramétrage du modèle hydrologique, sans amélioration notable des performances pour les avertissements pour le seuil de débit biennal.

En 2015, Météo-France a fourni au SCHAPI, pour 3 événements de l'automne 2014, les prévisions immédiates du système AROME-PI alors en cours d'implémentation opérationnelle. Le système configuré pour la prévision immédiate, appelé AROME-PI et en anglais, AROME-NWC ou AROME-NOWCASTING (Auger et al. 2015), a pour objectif de mieux caractériser les pluies des échéances de +1h à +6h. Il diffère notamment du système AROME par sa réactualisation horaire permettant d'utiliser les observations les plus récentes. Le système AROME-PI est devenu opérationnel en mars 2016. Il est donc important pour le SCHAPI d'évaluer l'apport des prévisions déterministes AROME-PI pour le système Vigicrues-Flash puisque ce sont les prévisions opérationnelles de Météo-France a priori les

mieux adaptées pour les courtes échéances. Dans l'étude de Demargne (2015) et dans l'étude présentée ici, les prévisions déterministes AROME-PI sont également 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, qui peuvent avoir un impact significatif pour des événements de pluie intense.

L'étude présentée ici concerne 8 événements sur la période de septembre 2014 à février 2016 et l'ensemble des 781 bassins versants jaugés utilisés pour tester la modélisation hydrologique proposée pour le système national d'avertissement Vigicrues-Flash. L'objectif principal est d'évaluer comment l'intégration des prévisions de pluie AROME-PI permet d'améliorer la détection et l'anticipation des dépassements de seuil de débit utilisés pour définir les avertissements aux crues soudaines.

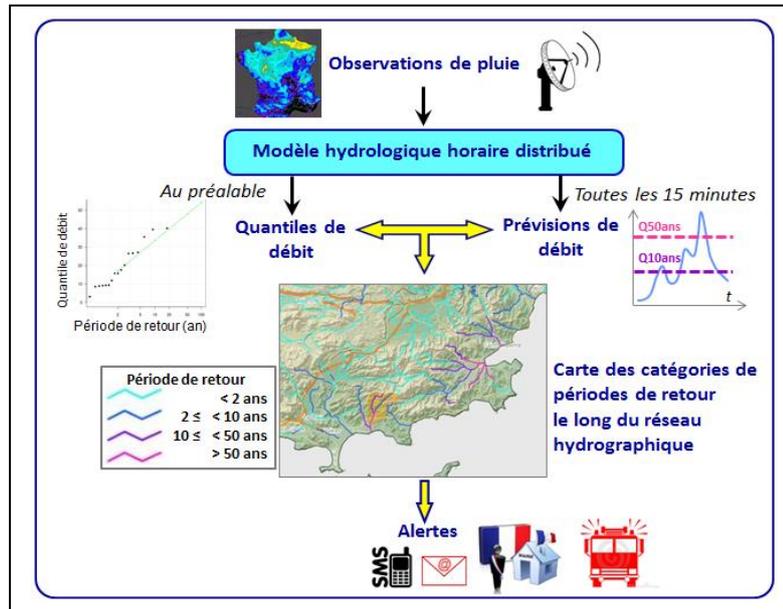
### ***b) Travaux antérieurs pour le développement de Vigicrues-Flash***

De nombreux travaux ont été réalisés dans le cadre du développement du système d'avertissement Vigicrues-Flash actuellement en cours d'implémentation par le SCHAPI grâce aux collaborations entre les équipes d'Irstea Aix, d'Hydris et du SCHAPI.

En ce qui concerne son principe de fonctionnement décrit en Figure 1 (cf. de Saint Aubin et al. 2016 pour plus de détails), le système d'avertissement Vigicrues-Flash s'appuie sur la méthode AIGA (Javelle et al. 2010, 2014, 2016). La méthode AIGA, développé par Irstea, fonctionne en temps réel depuis 2005 sur le pourtour méditerranéen et est utilisée actuellement sur la plateforme du projet RHYTMME (Risques HYdrométéorologiques en Territoires de Montagnes et MEditerranéens, [rhythme.irstea.fr](http://rhythme.irstea.fr)) pour la région Provence-Alpes-Côte d'Azur.

Dans la version testée actuellement par le SCHAPI pour Vigicrues-Flash, le modèle hydrologique semi-distribué et événementiel (Organde 2015) intègre les pluies horaires produites à la résolution de  $1\text{km}^2$  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 à partir des débits simulés sur une période de 18 ans. Pour caractériser la gravité des événements en cours, le système d'avertissement produit, toutes les 15 minutes (grâce à l'utilisation de 4 chaînes d'avertissement fonctionnant au pas de temps horaire), la carte des périodes de retour estimées le long des cours d'eau pour des catégories définies au préalable.

Dans les tests en cours du service Vigicrues-Flash, trois catégories de période de retour ont été fixées pour définir des avertissements de crue forte et de crue très forte (par exemple, période de retour inférieure à 5 ans, entre 5 et 20 ans, ou supérieure à 20 ans). Le service Vigicrues-Flash permet l'envoi d'un avertissement de crue forte ou crue très forte aux communes abonnées, ainsi que l'affichage des communes en avertissement et des cours d'eau en avertissement. Pour la plateforme RHYTMME actuelle, les catégories se basent sur les périodes de retour de 2 ans, 10 ans et 50 ans.



**Figure 1 : Principe de fonctionnement de la chaîne d’avertissement du système Vigicrues-Flash, se basant sur la méthode AIGA (les catégories de période de retour correspondent aux informations de la plateforme RHYTMME)**

Les développements relatifs à la méthode AIGA concernent les améliorations de la modélisation pluie-débit, ainsi que la prise en compte de prévisions de pluie et des incertitudes des prévisions de pluie et de débit.

### **Améliorations de la modélisation pluie-débit**

La calibration et la régionalisation du modèle pluie-débit ont été améliorées par Organde (2013, 2014, 2015) pour l’implémentation nationale du système d’avertissement par le SCHAPI. En particulier, ces études ont permis de valider une modélisation hydrologique semi-distribuée, comprenant un modèle journalier continu d’initialisation et un modèle horaire distribué événementiel, ainsi que la procédure de régionalisation des paramètres des 2 modèles hydrologiques. De plus, on considère 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 de simulations générées par la modélisation hydrologique sur la même période (qui est de 18 ans environ actuellement). Cette approche permet de prendre en compte les biais du modèle hydrologique utilisé pour le système d’avertissement 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 d’avertissement 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.

## **Prise en compte des prévisions et de leurs incertitudes**

Les études de Demargne (2013, 2014 et 2015a) ont montré les améliorations du système d'avertissement en termes de détection et d'anticipation des crues soudaines grâce à l'intégration des prévisions de pluie dans la 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 qui a conduit plus récemment au développement de systèmes de prévision d'ensemble hydrométéorologique. Les articles de Cloke et Pappenberger (2009), Zappa et al. (2010) et Pappenberger et al. (2016) donnent de nombreux exemples de systèmes opérationnels ou pré-opérationnels de prévision d'ensemble pour l'hydrologie.

Cette approche ensembliste a été mise en œuvre avec le système d'avertissement pour considérer les incertitudes des prévisions de pluie dans les études de Demargne (2013, 2014 et 2015a), ainsi que l'incertitude des paramètres du modèle hydrologique horaire semi-distribué dans l'étude de 2014. La Figure 2 décrit les principales sources d'incertitude du système d'avertissement Vigicrues-Flash en intégrant des prévisions d'ensemble de pluie lors de l'estimation des prévisions de débit et des dépassements de seuils de quantiles de crue. Avec la modélisation hydrologique actuellement utilisée, les incertitudes météorologiques sont liées aux observations de pluie, issues des données des radars et des pluviomètres, et aux prévisions de pluie. Pour le modèle hydrologique, les incertitudes concernent la paramétrisation du modèle, son initialisation, ainsi que la structure du modèle. Lorsque des prévisions d'ensemble de pluie sont utilisées avec un ou, potentiellement, plusieurs jeux équiprobables de pluie observée, de conditions initiales et de paramètres hydrologiques, le système produit des prévisions d'ensemble de débit qui sont également équiprobables. Les seuils de quantiles de crue, qu'ils soient issus de la base de données SHYREG (Arnaud et al. 2014) ou estimés à partir des chroniques de débit simulé, sont également issus d'une modélisation hydrométéorologique et donc entachés d'incertitude. La comparaison des prévisions d'ensemble de débit équiprobables avec les quantiles de crue (dont les incertitudes sont idéalement caractérisées) conduit à estimer la probabilité de dépasser le seuil de crue. Ainsi, le prévisionniste peut donner un avertissement de crue associé à un niveau de probabilité. En donnant de manière objective des niveaux de confiance aux avertissements,

les prévisionnistes et utilisateurs peuvent considérer des niveaux de tolérance acceptable vis-à-vis du risque de crue, qui peuvent varier selon les applications, et leur prise de décision peut donc être optimisée.

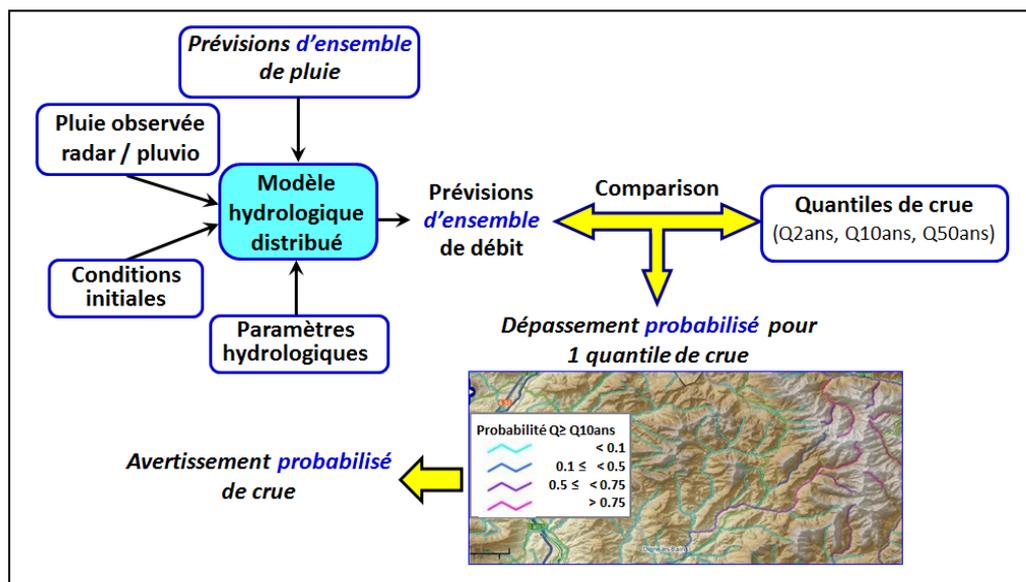


Figure 2 : Principales sources d'incertitude dans la chaîne d'avertissement du système Vigicrues-Flash

L'étude présentée ici considère l'impact de l'intégration de prévisions immédiates de pluie dans le système Vigicrues-Flash. On cherche également à quantifier les incertitudes des pluies futures grâce à l'intégration de prévisions d'ensemble de pluie, qui sont générées à partir des prévisions déterministes du système AROME-PI.

### c) 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) au pas de temps horaire pour des échéances de 30h à 36h. Les prévisions sont produites sur des grilles en latitude/longitude à l'échelle de 2,5 km environ (0,025 degrés de résolution) et, depuis le 13 avril 2015, à l'échelle de 1,3 km environ (0,01 degrés de résolution) (Météo-France, 2015). 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.

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.

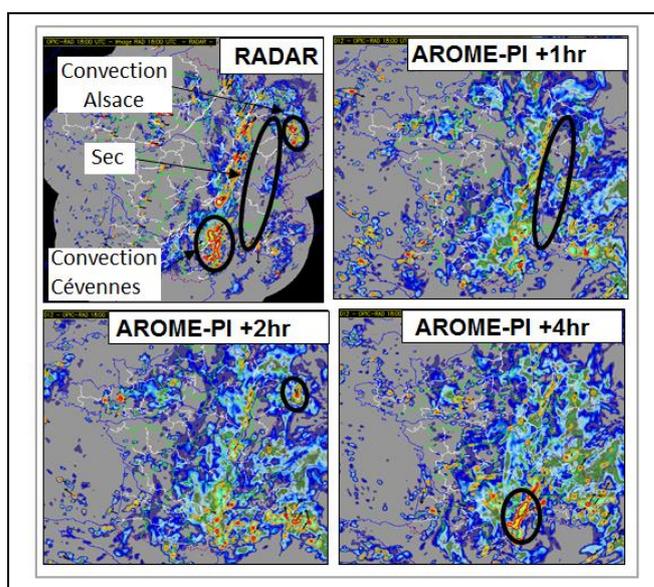
Les caractéristiques des prévisions AROME-PI utilisées dans cette étude sont les suivantes (Météo-France, 2015):

- prévisions issues du modèle AROME (alimenté par le modèle à plus grande échelle ARPEGE) qui inclut une modélisation physique des processus convectifs ;

- prévisions produites sur la grille FRANGP0025 sur le domaine 53N 38N, 8W 12E à la résolution spatiale d'environ 2,5 km (0.025dg lat/lon) jusqu'au 13 avril 2015 ; par la suite, prévisions produites sur la grille EURW1S100 sur le domaine 55.4N 37.5N, 12W 16E à la résolution spatiale d'environ 1,3 km (0.01dg lat/lon) ;
- des prévisions avec des cumuls incrémentés toutes les 15 minutes, pour une échéance maximale de 6 heures ;
- des prévisions réactualisées 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 de Vigicrues-Flash.

Dans cette étude, nous proposons d'intégrer ces prévisions immédiates AROME-PI de type déterministe au système Vigicrues-Flash, ainsi que de construire des prévisions d'ensemble par décalage temporel de prévisions déterministes successives, prévisions dites d'ensembles timelag. Il est en effet intéressant d'utiliser les prévisions déterministes AROME-PI plus anciennes car, du fait des incertitudes de ces sorties de modèle météorologique, ces prévisions plus anciennes contiennent en partie des scénarios de pluie future pertinents vis-à-vis des pluies observées. La Figure 3 (issue d'une présentation au SCHAPI de Joël Hoffman) en donne une illustration en comparant la carte de pluie observée issue du système radar avec les prévisions AROME-PI successives pour les horizons de +1h (la prévision la plus récente), +2h et +4h. On observe trois caractéristiques sur la carte de pluie observée : un temps sec sur l'Est avec, en parallèle plus à l'Ouest, des pluies intenses, ainsi que deux événements convectifs en Alsace et sur les Cévennes. Alors que la caractéristique de temps sec en parallèle du temps pluvieux est bien décrite par la prévision AROME-PI la plus récente, les 2 événements convectifs (qui sont critiques pour les crues soudaines) sont les mieux décrits dans les prévisions moins récentes. Il semble donc essentiel d'intégrer non seulement les prévisions AROME-PI les plus récentes, mais également les prévisions moins récentes.



**Figure 3 : Exemple de comparaison de la grille de pluie observée issue des radars et des grilles de prévisions AROME-PI successives pour les horizons de +1h, +2h et +4h (source : Joël Hoffman)**

#### d) Objectifs de l'étude

Comme décrit dans le schéma de la Figure 4, le travail proposé répond aux trois objectifs suivants :

- 1) évaluer l'intérêt d'intégrer la prévision déterministe AROME-PI pour générer des prévisions déterministes de débit et, potentiellement, 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 de pluie 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 avertissements des dépassements vis-à-vis du système Vigicrues-Flash actuel, i.e. sans pluie future et avec la modélisation hydrologique semi-distribuée, avec les avertissements des prévisions issues de la modélisation hydrologique continue développée par Organde en 2016, ainsi que les prévisions déterministes de débit de référence, correspondant aux débits issus d'une pluie future nulle, d'une pluie future persistante et de la pluie future « parfaite » (i.e. la pluie observée).

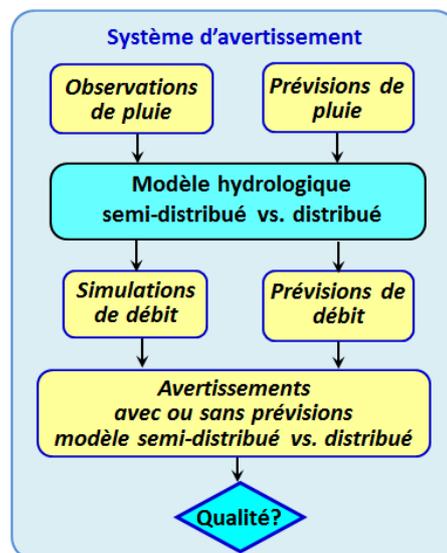


Figure 4 : Approche méthodologique pour évaluer l'apport des prévisions AROME-PI pour le système d'avertissement Vigicrues-Flash en utilisant la modélisation hydrologique semi-distribuée du système actuel et la modélisation hydrologique distribuée continue de l'étude Organde (2016)

## 2. Données et Modélisation Hydrologique

Les grilles de prévisions AROME-PI ont été mises à disposition du SCHAPI par Météo-France pour 8 événements de septembre 2014 à février 2016. Les pluies observées sur cette période sont issues du système opérationnel de Météo-France dit « Antilope KED » (en utilisant les stations pluviographiques et les radars avec krigeage à dérivée externe, une réanalyse ayant été produite par Météo-France pour obtenir des chroniques de pluie avant 2015 cohérentes avec les données opérationnelles).

Le travail de cette étude utilise les données de débit et deux des modélisations hydrologiques de l'étude d'Organde (2016) : 1) la modélisation hydrologique semi-distribuée (avec un modèle horaire événementiel) qui est actuellement intégrée dans le système national

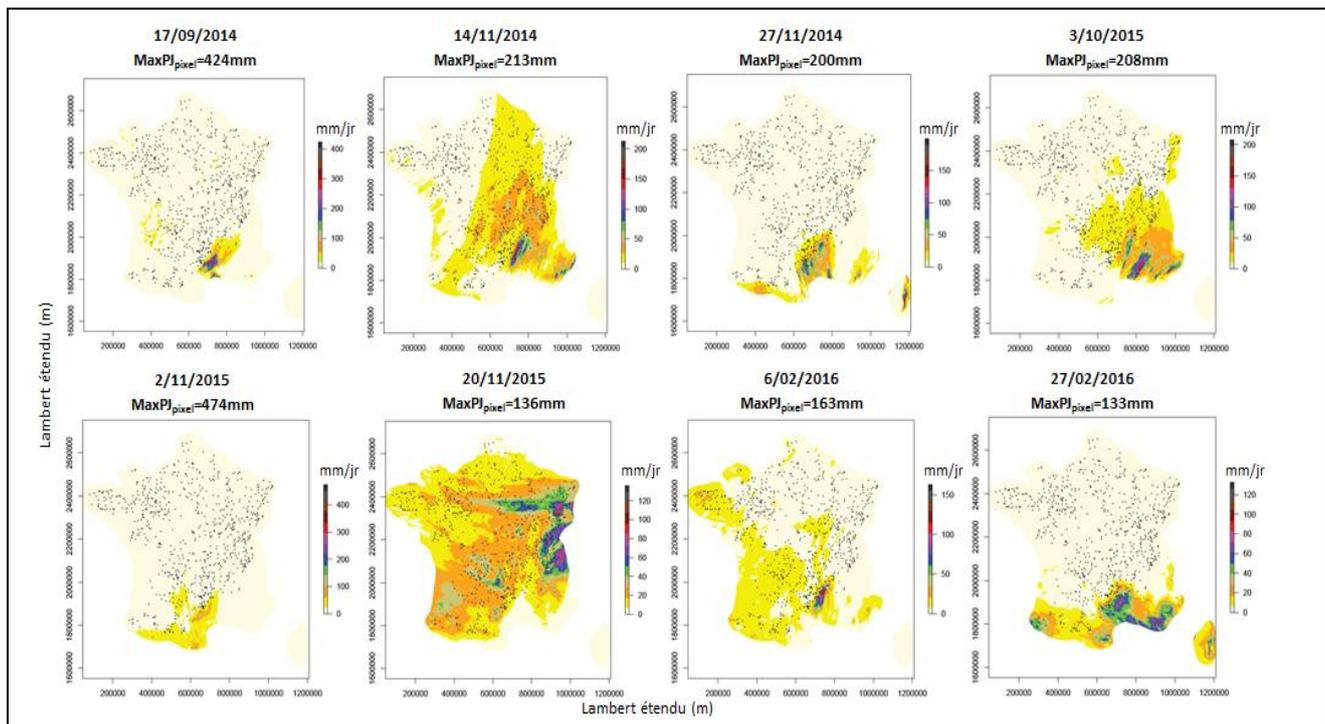
Vigicrues-Flash, avec le jeu de paramètres calibrés et régionalisés de 2015 ; 2) la modélisation hydrologique distribuée continue avec le jeu de paramètres calibrés et régionalisés de 2016.

### a) Sélection des événements

La sélection des événements a été réalisée en fonction de la disponibilité des prévisions AROME-PI de Météo-France (prévisions issues de rejeu avant mars 2016 et prévisions opérationnelles produites en temps réel par la suite) et pour une sélection des événements les plus importants en termes de pluie journalière. Les prévisions AROME-PI sont donc disponibles pour les 8 événements suivants :

- en 2014, du 15 au 20 septembre, du 12 au 16 novembre, du 23 au 30 novembre,
- en 2015, du 2 au 4 octobre, du 1<sup>er</sup> au 4 novembre, du 19 au 26 novembre,
- en 2016, du 5 au 9 février et du 27 au 29 février.

La Figure 5 présente les cartes des pluies observées journalières maximales (au pixel) pour chacun des 8 événements. Les événements concernent majoritairement la zone Sud-Est, à l'exception de l'événement de fin novembre 2015.



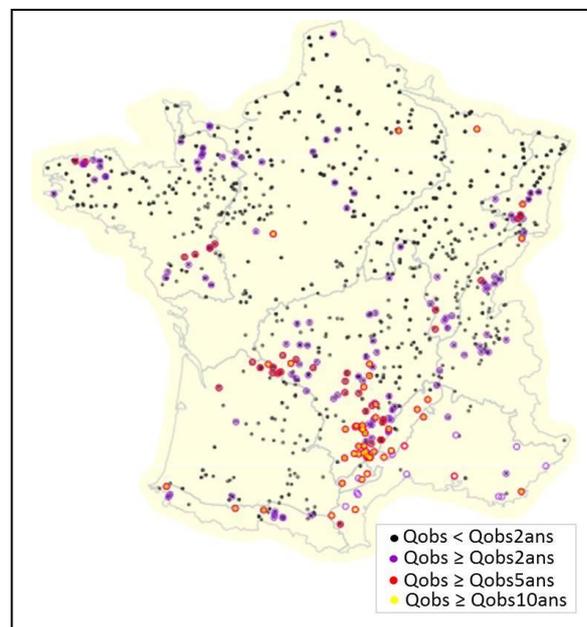
**Figure 5 : Cartes des pluies observées journalières maximales au pixel pour chacun des 8 événements (avec la valeur de pluie maximale observée au pixel), basées sur les pluies Antilope KED et la localisation des exutoires des 781 bassins versants étudiés**

### b) Bassins versants

A partir de la base des bassins versants du SCHAPI, la sélection des bassins réalisée par Organde (2016) inclut 781 bassins versants sur le territoire national. Les critères de sélection de ces bassins retenus pour la calibration et la régionalisation de la modélisation hydrologique sont les suivants (voir Organde 2016 pour plus de détails):

- les données de débit observées sont disponibles sur la période de calibration et régionalisation de la modélisation hydrologique de 2002-2016 pour un minimum de 10 ans et la qualité en hautes eaux a été jugée acceptable par le réseau des SPC;
- les surfaces de bassin versant sont inférieures à 1 500 km<sup>2</sup> 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 ;
- les bassins versants sont jugés suffisamment rapides pour leur utilisation pour le service Vigicrues-Flash.

La Figure 6 présente la carte des 781 bassins versants de cette étude avec la catégorie maximale de période de retour des dépassements observés sur les 8 événements étudiés. Les surfaces de bassin varient de 5 km<sup>2</sup> à 1548 km<sup>2</sup>, avec une valeur médiane de 143 km<sup>2</sup> (25% des bassins ont une surface inférieure à 71 km<sup>2</sup> et 75% des bassins ont une surface inférieure à 299 km<sup>2</sup>).



**Figure 6 : Carte des exutoires des 781 bassins versants avec la catégorie maximale de période de retour des dépassements observés sur les 8 événements de l'étude (pour les seuils de quantiles observés de 2, 5 et 10 ans)**

Sur la carte de période de retour maximale des dépassements observés, on observe que les dépassements des seuils de 5 et 10 ans concernent en majorité la moitié Sud de la France (en cohérence avec les cartes de pluies journalières maximales de la Figure 5).

La répartition par événement du nombre de dépassements observés et simulés avec la modélisation hydrologique semi-distribuée pour le premier seuil de période de retour de 2 ans est présentée en Figure 7. Les 242 dépassements observés et les 273 dépassements simulés

correspondants concernent principalement les événements de novembre 2014, novembre 2015, ainsi que l'événement du début de février 2016.

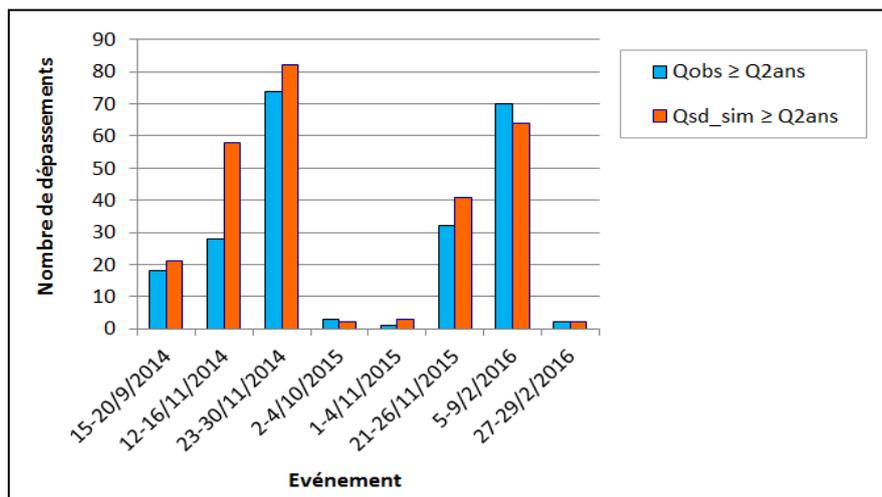


Figure 7 : Répartition par événement du nombre de dépassements observés (Qobs) et simulés avec la modélisation hydrologique semi-distribuée (Qsd\_sim) pour le seuil relatif de période de retour de 2 ans et les 781 bassins versants

### c) Modélisations hydrologiques

Dans cette étude, deux modélisations hydrologiques ont été comparées :

- la modélisation semi-distribuée finalisée en 2015 et actuellement en cours de tests pour le système Vigicrues-Flash (Organde 2015), qui sera notée GRSD par la suite ;
- la modélisation distribuée continue développée en 2016 (Organde 2016), notée GRDc.

#### Modélisation hydrologique semi-distribuée GRSD de 2015

Comme le montre le schéma de la Figure 8, la modélisation hydrologique semi-distribuée du système Vigicrues-Flash actuel s'appuie sur le couplage de trois modèles (cf. Organde 2014 pour plus de détails) :

- le modèle événementiel horaire semi-distribué, noté ici GRSDévé, qui inclut 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 GRSDévé ( $R_0/B$ ) 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 GRSDévé ( $S_{i0}/A_i$ ) est ajusté quotidiennement si la pluie journalière de la veille au pixel est inférieure à 10 mm avec la relation suivante :  $S_{i0}/A_i = a_i \cdot S_i A_i(J)$

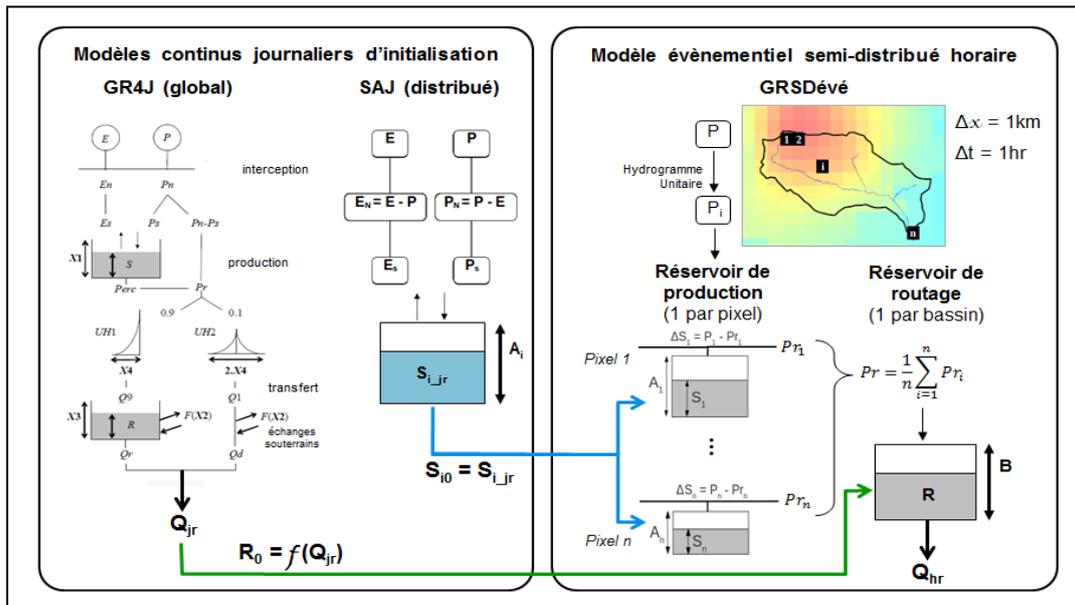


Figure 8 : Schéma du couplage des modèles journaliers continus SAJ et GR4J et du modèle horaire semi-distribué GRSDévé pour le système actuel Vigicrues-Flash

Les paramètres des trois modèles GRSDévé, SAJ et GR4J qui doivent être estimés sont les suivants :

- capacité  $A_i$  du réservoir de production de GRSDévé et du modèle SAJ, définie en chaque pixel  $i$  du territoire, qui a été fixée à 200 mm,
- paramètre  $a_i$  pour l'initialisation du modèle GRSDévé, défini en chaque pixel  $i$  du territoire, qui a été fixé à 1,
- capacité  $B$  (globale) du réservoir de transfert de GR-SD, définie pour chaque bassin versant, qui a été fixée à 50 mm, une valeur faible pour permettre de générer une rapide augmentation des débits,
- les 4 paramètres du modèle GR4J, définis en chaque bassin versant.

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

Le calage du modèle 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 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).

### Modélisation hydrologique distribuée et continue GRDc de 2016

Comme la modélisation hydrologique semi-distribuée est relativement complexe en associant trois modèles différents, une modélisation hydrologique plus simple, qui est distribuée,

continue et fonctionnant au pas de temps horaire, a été proposée dans l'étude d'Organde (2016), et dont le schéma est présenté en Figure 9.

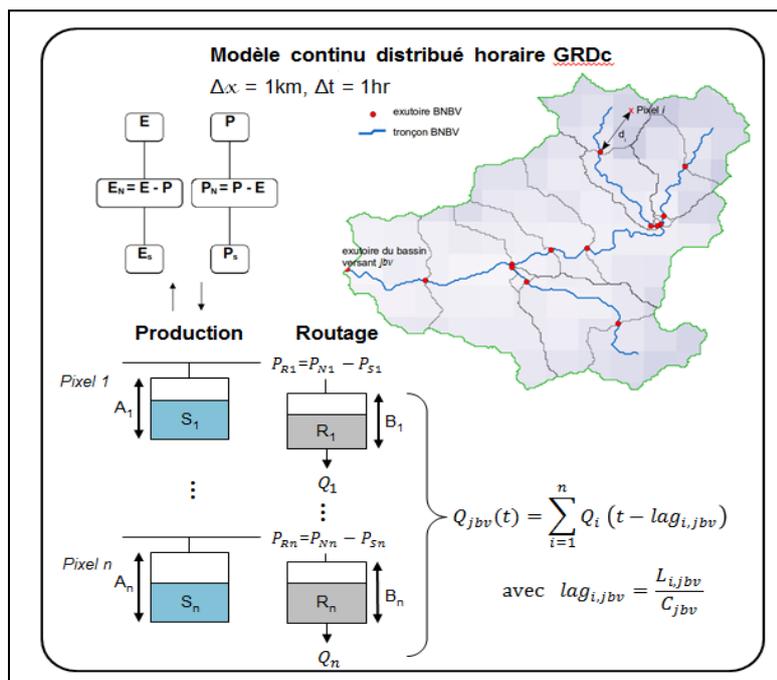


Figure 9 : Schéma du modèle distribué continu GRDc

La neutralisation de la pluie par l'évapotranspiration potentielle permet d'estimer une pluie nette ( $P_N$ ) qui est répartie entre le réservoir de production et le réservoir de transfert, qui sont tous deux définis au pixel. La proportion de pluie qui « ruisselle » ( $P_R$ ) est proportionnelle au carré du niveau  $S_i$  du réservoir de production  $A_i$ . Le réservoir  $B_i$  assure le transfert non linéaire de la pluie, au sein du pixel, en fonction de son niveau de remplissage  $R_i$  et de sa capacité  $B_i$ .

Les débits obtenus ainsi en chaque pixel du bassin versant sont alors sommés en incluant un décalage temporel afin de reproduire au mieux les hydrogrammes observés. Le décalage temporel est le rapport de la distance hydraulique entre le pixel et l'exutoire, et de la célérité entre le pixel et l'exutoire, supposée uniforme sur le bassin versant. La distance hydraulique est basée sur la distance euclidienne entre le centre du pixel et le plus proche exutoire de la base nationale de bassins versants vers lequel il s'écoule ; à cette distance, est ajoutée la longueur du parcours hydraulique vers l'exutoire du bassin versant, en se basant sur la longueur hydraulique des tronçons de la base nationale de bassins versants (cf. Organde 2016 pour plus de détails).

La calibration et la régionalisation du modèle GRDc ont été réalisées sur la période de 2002 à 2016. Pour la version du modèle régionalisé de cette étude, le paramètre  $A_i$  a été fixé à 200mm, le paramètre  $B_i$  a été optimisé pour obtenir une valeur par hydro-écorégion, et la célérité a été régionalisée pour chaque bassin versant avec la méthode des voisins.

Avec l'intégration d'un décalage temporel des débits au pixel, ce modèle a tendance à amortir les hydrogrammes et donc à perdre en anticipation vis-à-vis des simulations produites par le modèle semi-distribué. Par contre, même en incluant une prévision de pluie nulle, le débit à l'exutoire du bassin peut croître par l'apport des débits au pixel précédant l'instant  $t_0$  de prévision si les décalages temporels au sein de bassin sont supérieurs à 1h. Ce modèle doit donc être évalué, non pas en mode simulation (sans aucune pluie future) comme pour le

modèle actuel du système Vigicrues-Flash, mais en intégrant une prévision de pluie nulle afin de prendre en compte tous les éventuels dépassements de seuil.

#### **d) Données hydrométéorologiques**

Concernant les données d'observations de pluie, il a été décidé d'utiliser les meilleures pluies produites en temps réel par Météo-France à la résolution spatiale de 1 km<sup>2</sup> à partir des systèmes radars et des stations pluviométriques. C'est pourquoi ce sont les pluies Antilope avec krigeage avec dérive externe (appelées Antilope KED ou Antilope V2) qui sont utilisées pour les pluies horaires et pour le calcul des pluies journalières. Ces données Antilope KED ont été produites rétrospectivement par Météo-France pour les dates précédant septembre 2015, pour mieux correspondre aux données produites actuellement en temps réel.

Les pluies Antilope KED sont des produits composites basées sur les données pluviométriques et les lames d'eau radar. Cette fusion radar-stations permet d'améliorer significativement la qualité des champs de pluie en comparaison aux produits radar Panthère, pour réduire notamment le biais de sous-estimation des pluies pour des événements de forte pluie (Pauthier et al. 2016). Pour Antilope KED, les observations des stations pluviométriques sont tout d'abord corrigées grâce aux lames d'eau radar et à un produit de « classification nuageuse » issu du satellite MSG (pour éliminer les cas de fausse pluie). Les lames d'eau radar sont issues de la mosaïque Panthère au pas de temps de 5 minutes, qui permet une bonne représentation spatiale des champs de pluie. La fusion des données pluviométriques et radar se base sur une méthode de krigeage à dérivée externe. En sortie, on obtient des pluies avec une extension spatiale proche de celle du radar mais avec des valeurs ajustées en fonction des cumuls pluviométriques observés ; pour les pixels correspondants aux stations pluviométriques, on retrouve le cumul mesuré par le pluviomètre. Cette méthode de fusion nécessite l'utilisation d'un variogramme, qui est calculé à chaque heure à l'aide de l'image radar ; des variogrammes distincts sont calculés pour les pluies stratiformes et pour les pluies convectives.

Les données hydrométriques sont issues de la base de données BD Hydro et de la critique des données de débit par les SPC. L'évaluation de la chaîne d'avertissement porte sur les dépassements de valeurs de seuil relatif pour différentes périodes de retour : 2 ans, 5 ans et 10 ans. Les quantiles issus des débits observés et des débits simulés sur la période 2002-2016 ont été estimés dans l'étude d'Organde (2016).

### **3. Méthodologies**

On décrit ici les étapes de la méthodologie pour construire des prévisions de pluie de type déterministe pouvant être intégrées dans Vigicrues-Flash, 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 issus des prévisions déterministes et ensemblistes.

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

Pour les huit événements sélectionnés, Météo-France 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 résolution d'environ 2,5 km (0.025 degrés) pour les 3 événements de 2014, et d'environ 1,3 km (0.01 degrés) pour les 5 autres événements de 2015 et 2016.

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 Vigicrues-Flash. 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 km<sup>2</sup> 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 un nombre de membres de 2 à 6 (i.e. pour un ensemble avec 6 membres, la prévision AROME-PI la plus récente 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 (à noter que l'on pourrait également considérer de maintenir la dernière valeur de pluie prévue, pour potentiellement remédier au problème de sous-estimation des pluies prévues, mais ceci risquerait de générer des surestimations, et donc des fausses alertes de dépassements). Ainsi pour toutes les heures d'initialisation, les prévisions AROME-PI timelag incluent de 2 à 6 membres pour toutes les échéances de 1h à 6h.

La Figure 10 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. Pour la prévision d'ensemble timelag avec 2 membres, le premier membre correspond à la prévision AROME-PI la plus récente issue à 7h, et le second membre à la prévision AROME-PI issue à 6h, avec les valeurs prévues pour les échéances de 2h à 6h et une pluie nulle pour le dernier pas de temps. Pour la prévision d'ensemble timelag avec 6 membres, on ajoute à ces 2 premiers membres, les membres 3 à 6 provenant des valeurs des prévisions AROME-PI générées entre 2h et 5h, complétées avec de la pluie nulle pour les échéances les plus lointaines (i.e. les 5 derniers pas de temps pour le membre 6, les 2 derniers pas de temps pour le membre 3).

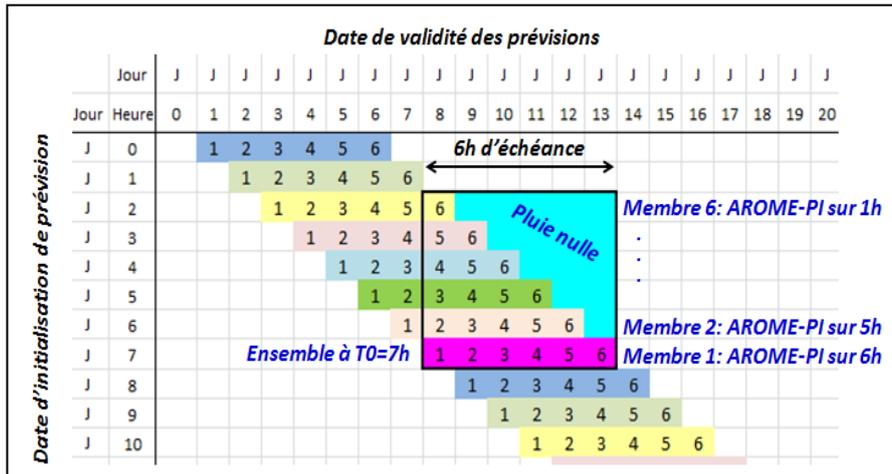


Figure 10 : Schéma de construction des ensembles timelag de pluie à partir des prévisions déterministes AROME-PI successives

Les membres des prévisions d'ensemble AROME-PI timelag sont considérés comme des scénarios équiprobables pour générer les prévisions d'ensemble de débit correspondantes sur les 6 heures d'échéance, chaque prévision d'ensemble de débit comprenant le même nombre de membres que la prévision d'ensemble de pluie. Les avertissements de dépassement de seuil sont donc issus des 6 prévisions de débit suivantes, couvrant toutes les 6 heures d'échéance : la prévision déterministe AROME-PI, ainsi que les 5 prévisions d'ensembles timelag comprenant de 2 membres à 6 membres.

### b) Prévisions de débit et avertissements

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 (Antilope KED) jusqu'à la date d'initialisation de prévision  $T_0$  et ensuite la prévision de pluie dans le modèle hydrologique. Chaque prévision déterministe AROME-PI génère une prévision de débit déterministe notée  $Q[pdet]$ . Comme décrit dans la section précédente, les prévisions d'ensemble de débit sont obtenues à partir des membres de la prévision d'ensemble AROME-PI timelag sur toute l'échéance de 6h, chaque ensemble de pluie générant un ensemble de débit. On obtient ainsi 5 prévisions d'ensemble de débit avec un nombre de membres compris entre 2 et 6, couvrant les 6h d'échéance, afin de décrire l'impact de l'incertitude de pluie future sur les prévisions de débit.

Pour montrer l'apport des prévisions AROME-PI déterministes et ensemblistes, nous utilisons les prévisions de référence suivantes :

- le débit simulé, sans pluie future, qui produit les avertissements du système Vigicrues-Flash actuel ; cette « prévision sans échéance » (notée  $Q_{sim}$ ) correspond au débit simulé obtenu à la date d'initialisation  $T_0$ ,
- le débit simulé avec une pluie future nulle sur les 6h d'échéance (noté  $Q[pnul]$ ) ;

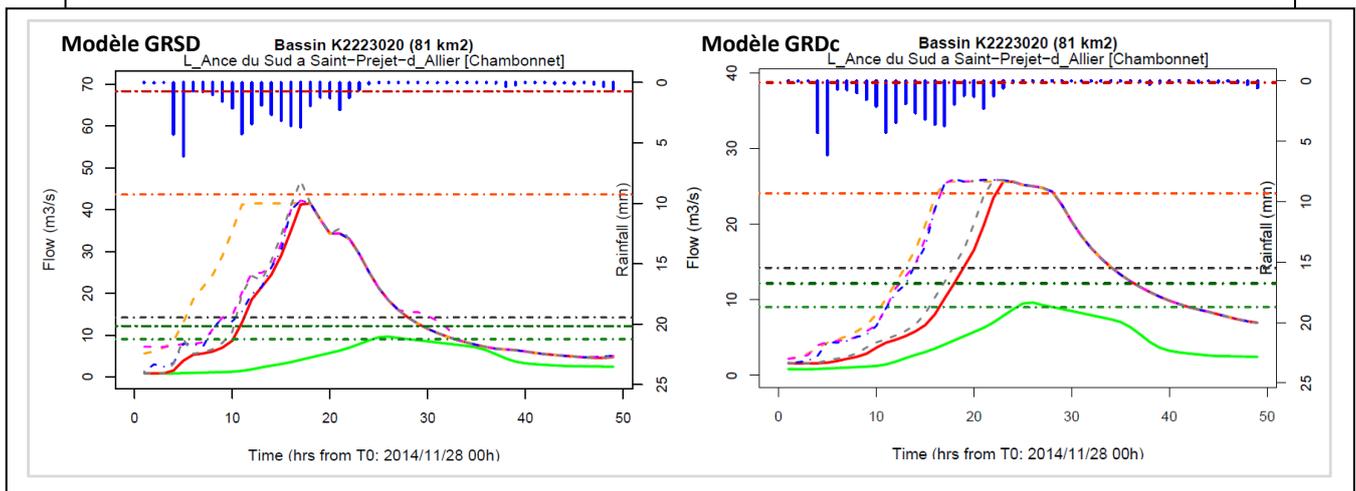
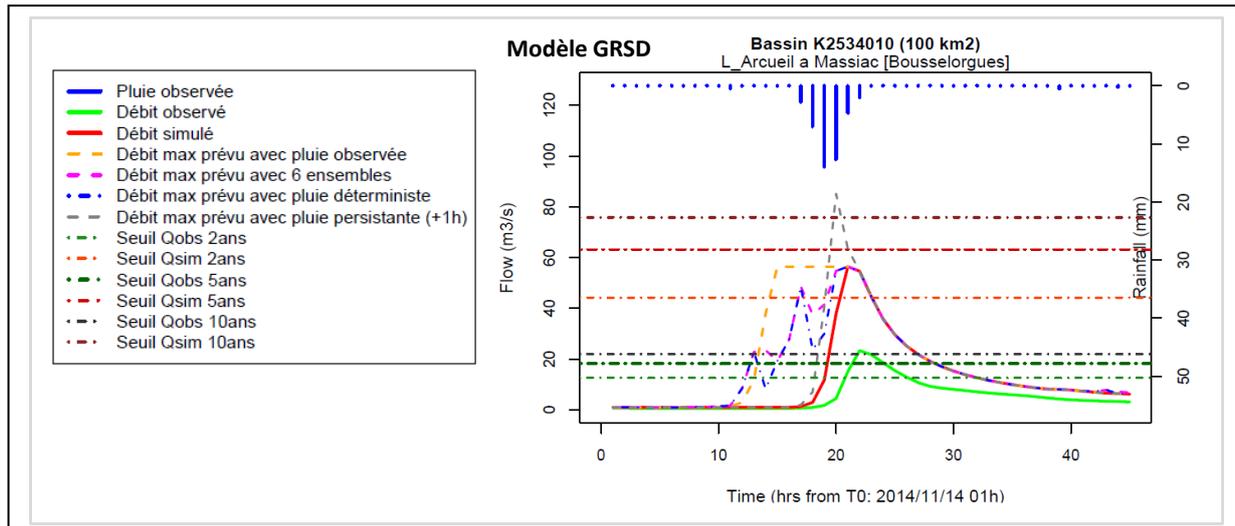
- le débit simulé obtenu avec la pluie future « parfaite » (en négligeant les erreurs de la pluie observée) connue *a posteriori*, sur les 6h d'échéance (prévision notée  $Q[\text{pobs}]$ );
- la prévision de débit issue de la pluie persistante, définie en maintenant la valeur de pluie observée de la dernière heure avant l'initialisation des prévisions sur une échéance  $n$  de +1h à +6h (prévision notée  $Q[\text{pers}_{nh}]$ ).

Toutes les prévisions de débit sont produites pour tous les pas de temps horaires pour lesquels les prévisions AROME-PI sont disponibles pour les 8 événements de l'étude.

L'évaluation de la qualité des avertissements s'est basée sur le dépassement de seuil de débit relatif correspondant à des périodes de retour de 2 ans, 5 ans et 10 ans. Ainsi, pour chaque période de retour  $T$ , deux valeurs distinctes de seuil de débit sont définies : une valeur de seuil  $Q_{\text{obs}T\text{ans}}$ , estimée à partir des observations de débit sur la période 2002-2016, et une valeur de seuil  $Q_{\text{sim}T\text{ans}}$ , estimée à partir des simulations de débit sur cette même période 2002-2016. Cette approche permet de « débiaiser » les sorties du modèle hydrologique.

La Figure 11 présente des exemples des hydrogrammes de débit observé et de débit maximal sur l'échéance de 6 heures (qui permet de déterminer les dépassements de seuil), qui ont été produits par le modèle GRSD et le modèle GRDc avec les différentes prévisions de pluie. On peut noter de légères différences entre le débit maximal issu de la prévision déterministe AROME-PI et le débit maximal issu des 6 membres des ensembles AROMEPI-timelag, ce qui confirme un apport des autres membres de l'ensemble, même si cet apport est limité (par construction des ensembles timelag, une pluie nulle étant utilisée pour les échéances sans prévision déterministe AROME-PI). Concernant la prévision issue de la pluie persistante, dans ces exemples, seul le débit maximal issu de la pluie persistante maintenue sur 1h d'échéance est représenté. On peut noter l'intérêt d'utiliser des seuils de débit relatifs pour les avertissements : pour le premier bassin par exemple, cette approche permet de prendre en compte la tendance à la surestimation du débit observé par le débit simulé issu de GRSD.

Dans le premier exemple de la Figure 11, le dépassement observé du seuil de 2 ans a été simulé par GRSD sans anticipation ; avec la prévision déterministe AROME-PI, le dépassement est prévu avec une anticipation de +4h. Par contre, le dépassement observé du seuil de 5 ans n'est prévu qu'avec la prévision de pluie persistante sur une échéance d'1 heure. Dans les exemples 2 et 3 pour un autre bassin avec un dépassement observé du seuil de 2 ans, le modèle GRSD a un pic de débit prévu qui reste légèrement inférieur au seuil de 2 ans (sauf avec la pluie persistante), alors que le pic de débit prévu avec le modèle GRDc dépasse la valeur du seuil. Le modèle GRDc produit un dépassement simulé avec 2 heures d'anticipation (comme attendu, le modèle GRSD est plus réactif, mais, dans ce cas, ne produit pas de dépassement du seuil), la prévision déterministe AROME-PI permet d'obtenir 8 heures d'anticipation.



**Figure 11 : Exemples d'hydrogrammes obtenus par les modèles GRSD et GRDc avec les différentes prévisions de pluie en comparaison avec l'hydrogramme de débit observé**

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

Ainsi, pour une prévision d'ensemble ayant  $K$  membres ( $K$  variant entre 2 et 6), on obtient des dépassements potentiellement différents en considérant la probabilité de dépassement d'1 seul membre sur  $K$ , de 2 membres sur  $K$ , ..., ou de  $K$  membres sur  $K$ . Ces prévisions probabilistes sont notées  $Q_{ensI/K}$  quand le dépassement est prévu lorsque  $I$  membres sur  $K$  membres dépassent le seuil de débit ( $K$  variant entre 2 et 6,  $I$  variant de 1 jusqu'à  $K$ ).

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

pluie future comprend uniquement la valeur du débit simulé à T0 (qui produit les avertissements du système Vigicrues-Flash actuel).

### 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 8 événements étudiés et les 781 bassins versants. 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 événement 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. Ainsi, 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	<b>a</b> (succès/hit)	<b>b</b> (fausse alerte/false alarm)
	Non	<b>c</b> (manqué/miss)	<b>d</b> (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, avec un score optimal égal à 1 :  $FB = (a+b) / (a+c) = POD / SR$
- **Taux de bonne détection** ou **Probability Of Detection** (POD) (également appelé Hit Rate), qui est conditionné par les dépassements observés ; 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:  $POD = a / (a+c)$
- **Probability Of False Detection** (POFD) (ou False Alarm Rate), qui est conditionné par les non-dépassements observés ; le score varie entre 0 (aucune fausse alerte) et 1 (des alertes prévues pour tous les non-dépassements observés), avec un score optimal égal à 0 :  $POFD = b / (b+d)$
- **Taux d'alerte correcte** ou **Success Ratio** (SR), qui est conditionné par l'ensemble des alertes prévues pour mesurer la fiabilité des prévisions ; le score varie entre 0 (aucune alerte prévue qui soit correcte) et 1 (aucune fausse alerte), avec un score optimal égal à 1 :  $SR = a / (a+b)$

- **Critical Success Index (CSI)** qui permet de considérer les alertes manquées et les fausses alertes (mais sans différencier ces erreurs); le score varie entre 0 et 1, avec un score optimal égal à 1 :

$$CSI = a / (a+b+c) = 1 / ( 1/SR + 1/POD - 1 )$$

- **Gilbert Skill Score (GSS, également appelé Equitable Threat Score)** qui adapte la formulation du CSI pour prendre en compte les alertes correctes obtenues par chance ; le score varie entre -1/3 et 1, avec un score optimal égal à 1;

$$GSS = [a - hits_{chance}] / [(a+b+c) - hits_{chance}] \text{ avec}$$

$$hits_{chance} = [(a+c)(a+b)] / [a+b+c+d]$$

### Contingence globale vs. Contingence horaire

Le tableau des contingences des dépassements observés et prévus peut être généré en considérant :

- l'évènement dans son ensemble (i.e. pour toutes les dates d'initialisation de prévision de l'évènement), qui est appelée **contingence globale** dans cette étude ; la contingence globale correspond à une évaluation *a posteriori* du système d'avertissement, lorsque, pour un évènement donné, on peut comparer le premier dépassement de seuil qui a été observé, au premier dépassement de seuil qui a été prévu ;
- chaque date d'initialisation de prévision de l'évènement, que l'on appelle ici **contingence horaire** ; la contingence horaire correspond à l'évaluation du système d'avertissement en temps réel, lorsque, pour un évènement en cours, on peut comparer les dépassements obtenus avec le débit observé et le débit prévu pour chaque pas de temps horaire, dès que les observations de débit sont disponibles.

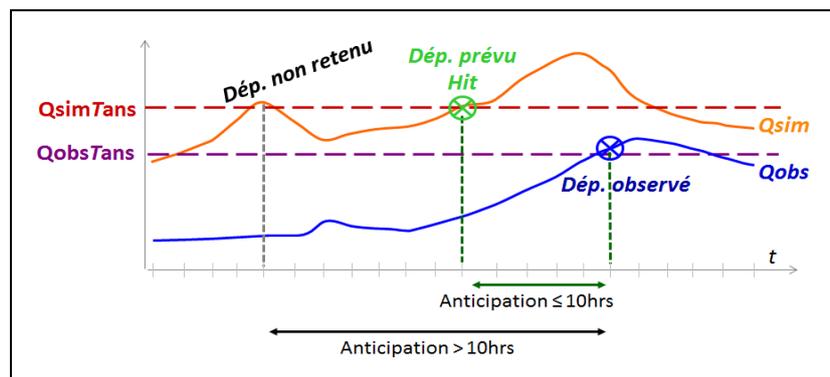
Dans cette étude, les résultats sont présentés principalement pour la contingence globale. Dans un second temps, quelques résultats obtenus avec la contingence horaire sont inclus, pour juger des performances des différentes prévisions lorsque l'évaluation porte sur chaque pas de temps horaire et donc, est plus stricte. Par exemple, si la prévision ou l'observation inclut plusieurs dépassements de seuil au cours de l'évènement, chaque dépassement prévu et/ou observé est évalué par la contingence horaire, alors que la contingence globale n'évalue que le premier dépassement observé au cours de l'évènement.

### Contingence globale et anticipation des dépassements

L'analyse de la qualité des avertissements en contingence globale se base également sur l'anticipation des dépassements issus des prévisions. L'anticipation se définit uniquement pour les dépassements de seuil observés, par la durée écoulée entre l'instant  $T_0$  où une prévision émise détecte un dépassement sur son échéance et l'instant  $T_{obs}$  où ce dépassement est effectivement constaté avec les observations :  $Anticipation = T_{obs} - T_0$  (h).

L'anticipation est donc comptée positivement lorsque  $T_0$  est antérieur à l'instant où le dépassement est observé ( $T_0 \leq T_{obs}$ ) (respectivement négativement lorsque  $T_0$  est postérieur, i.e.  $T_0 > T_{obs}$ ). Pour cette analyse relative à l'ensemble d'un évènement, l'anticipation est définie par la valeur maximale des anticipations calculées à chaque instant  $T_0$  (ce qui correspond à la première date de dépassement par la prévision de débit). On peut donc comparer, sur un même jeu de bassins, l'anticipation des dépassements issus des différents types de prévisions.

Par ailleurs, comme les dates de début et de fin des 8 événements ont été fixées pour tous les bassins de l'étude (pour obtenir auprès de Météo-France les prévisions AROME-PI), certains événements incluent différents pics de débit observé et/ou différents pics de débit prévu. Ainsi, pour certains bassins, le premier dépassement prévu peut se produire bien avant le premier dépassement observé car il correspond au premier pic de débit n'ayant pas conduit à un dépassement observé. Afin de mieux comparer les dépassements prévus et observés sur l'ensemble de l'événement, on fixe une valeur maximale d'anticipation de 10 heures pour le dépassement simulé (la valeur de 10 heures ayant été discutée avec le SCHAPI). C'est pourquoi, si le dépassement simulé se produit plus de 10 heures avant le dépassement observé, il n'est pas retenu dans l'analyse. Lorsque le dépassement est issu des prévisions de débit avec une échéance de 6 heures, la valeur maximale d'anticipation de dépassement prévu est alors de 16 heures (par exemple, dans le cas où le débit simulé à T0 anticipe un dépassement observé de 10h et que la prévision émise à T0-6h produit un dépassement prévu sur la 6eme heure de l'échéance de la prévision). Le schéma de la Figure 12 illustre ce principe de contingence globale avec une valeur maximale d'anticipation, qui permet d'associer de manière plus cohérente les dépassements observés avec les dépassements simulés ou prévus.



**Figure 12 : Principe de fonctionnement de la contingence globale (sur l'ensemble d'un événement) avec une valeur maximale d'anticipation pour associer un dépassement simulé avec un dépassement observé**

Comme l'objectif du système d'avertissement est d'anticiper au mieux les crues, les critères de contingence sont présentés ici en prenant en compte l'anticipation du dépassement de seuil par la prévision (à l'exception de la Figure 26 qui compare les valeurs du critère CSI estimé avec et sans prise en compte l'anticipation). Ainsi, un dépassement de seuil prévu avec une anticipation négative (le dépassement prévu est en retard vis-à-vis du dépassement effectif du seuil) est alors considéré comme une alerte manquée. Le dépassement de seuil prévu est comptabilisé comme une bonne alerte si son anticipation est égale ou supérieure à 0.

### **Contingence globale avec une tolérance sur les seuils de débit**

Nous savons que les valeurs de quantiles observés et simulés qui sont utilisées pour produire et évaluer les avertissements sont entachées d'incertitude. Les prévisionnistes et utilisateurs d'un système d'avertissement peuvent parfois réagir avec des débits prévus légèrement inférieurs aux quantiles simulés qui définissent les seuils si ces débits prévus sont suffisamment proches des quantiles et si les enjeux liés au risque de crue sont importants (cf. deuxième exemple donné en Figure 11 avec le modèle GRSD); dans ce cas, un avertissement pourrait être émis par le prévisionniste et devrait être compté comme correct si le débit observé dépasse le quantile observé. Par ailleurs, l'impact sur le terrain pour un débit observé

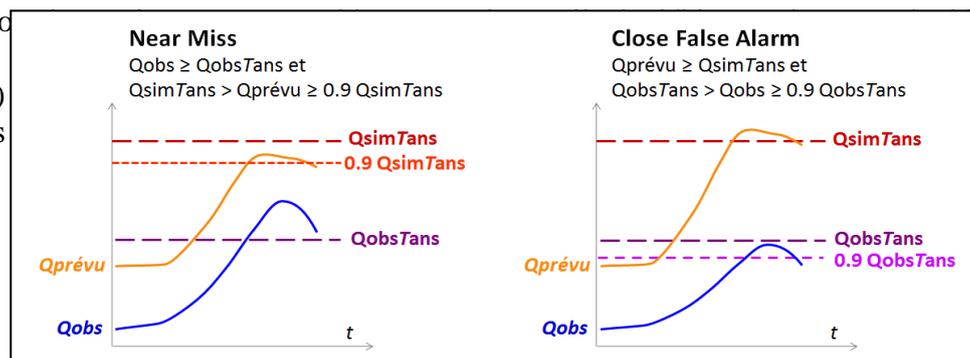
légèrement inférieur au quantile observé qui définit le seuil peut être suffisamment important pour préférer l'émission d'un avertissement, qui pourrait être comptabilisé comme une alerte correcte (même si le débit observé ne dépasse pas le seuil).

C'est pourquoi il nous paraît intéressant d'évaluer également les performances du système d'alerte lorsque l'on considère une tolérance sur la magnitude des seuils de débit pour définir des valeurs de seuil légèrement inférieures aux valeurs des quantiles de débit et comptabiliser ainsi comme alertes correctes, des alertes « quasi manquées » et de « quasi fausses » alertes. Ceci permet de juger de l'amélioration potentielle des performances dans une approche expertisée du système d'avertissement, pour laquelle un prévisionniste peut ajuster la production et l'évaluation des avertissements de crue. Cette approche, qui complète la description des performances sans tolérance sur les seuils de débit, a été retenue par le Centre d'Ecologie et d'Hydrologie de Wallingford en Grande-Bretagne pour leur évaluation nationale du système de prévision de crue, le National Flood Forecasting System (Wells et al. 2016).

Dans cette étude, on a retenu une tolérance de 10% sur les valeurs des quantiles de débit. On définit donc deux nouveaux cas d'avertissement jugé correct pour la contingence, au vu des incertitudes sur les quantiles et les dégâts potentiels sur le terrain (cf. Figure 13) :

- une alerte manquée de peu, dite « alerte quasi manquée » (« Near Miss ») lorsque le dépassement a été observé et le débit simulé ou prévu a seulement dépassé les 90% de la valeur du quantile simulé (on considère que le prévisionniste pourrait décider d'émettre un avertissement),
- une fausse alerte mais proche d'être observée, dite « quasi fausse alerte » (« Close False Alarm ») lorsque le dépassement a été prévu et le débit observé a seulement dépassé les 90% de la valeur du quantile observé (l'avertissement prévu pourrait être comptabilisé comme correct).

Ainsi, pour observer un dépassement (observé) d'alertes



assement he d'être e nombre ond cas).

**Figure 13 : Principe de fonctionnement de la contingence avec une tolérance de 10% sur les seuils pour comptabiliser les dépassements manqués de peu (« Near Miss ») et les quasi fausses alertes (« Close False Alarm »)**

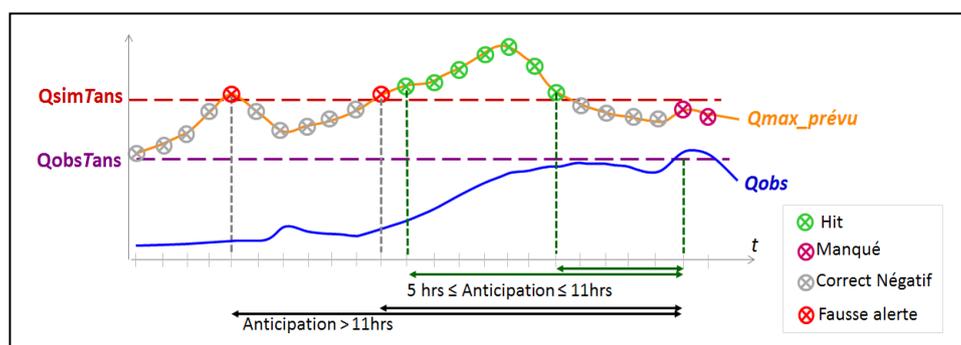
La contingence avec tolérance sur les seuils de débit est donc modifiée : le nombre d'alertes manquées de peu et le nombre de fausses alertes presque observées sont, tous deux, ajoutés au nombre des alertes correctes et soustraits, respectivement, du nombre des alertes manquées et du nombre des fausses alertes. Les scores de contingence vont donc être nécessairement améliorés mais cette amélioration permettra de quantifier la similitude entre les avertissements prévus et les avertissements observés dans une approche expertisée du

système d'avertissement, pour laquelle un prévisionniste peut ajuster la production et l'évaluation des avertissements de crue.

Les résultats de contingence sont présentés ci-dessous, tout d'abord sans tolérance sur les seuils de débit, et ensuite, avec une tolérance en magnitude de 10% sur ces seuils.

### Contingence horaire

L'analyse en contingence horaire, sur chaque pas de temps horaire au cours de chacun des 8 événements, conduit à une évaluation plus stricte qu'en contingence globale : à chaque pas de temps, on établit s'il y a ou non dépassement observé et dépassement prévu. Pour les différentes prévisions de débit, on prend en compte le maximum du débit prévu sur toute l'échéance de +6h. Afin de ne pas pénaliser trop fortement l'anticipation potentielle des dépassements prévus, on prend en compte une valeur maximale d'anticipation (valeur de +11 heures qui permet une anticipation de +5h du débit simulé sur le débit observé, à laquelle on ajoute les 6 heures de l'échéance des débits prévus). Si le dépassement prévu se produit plus de 11 heures en avance du dépassement observé, il est comptabilisé comme une fausse alerte. Le schéma de la Figure 14 illustre ce principe de contingence horaire avec une valeur maximale d'anticipation de +11 heures.



**Figure 14 : Principe de fonctionnement de la contingence horaire, pour chaque pas de temps de l'événement, avec une valeur maximale d'anticipation pour associer un dépassement prévu avec un dépassement observé**

Comme la contingence horaire évalue tout l'hydrogramme de débit prévu, cette évaluation est plus contraignante que l'évaluation de contingence globale, qui ne compare qu'un seul pic de débit prévu avec le pic de débit observé. Afin d'évaluer sur une même fenêtre d'échéance, des éventuels dépassements produits par les différentes prévisions, le système d'avertissement de référence pour les 2 modèles hydrologiques est basé sur la prévision de pluie future nulle.

## **4. Résultats d'anticipation des dépassements**

L'analyse des performances du système d'avertissement et de l'apport des prévisions AROME-PI est présentée pour les 2 modélisations hydrologiques, notées GRSD pour la modélisation semi-distribuée, et GRDC pour la modélisation distribuée continue.

Tout d'abord, on présente les échantillons des dépassements observés et des dépassements prévus sans intégrer de prévision de pluie non nulle pour les 8 événements étudiés et pour les 3 quantiles de période de retour de 2, 5 et 10 ans. Les échantillons incluent d'une part, tous les dépassements prévus en avance et en retard et, d'autre part, uniquement les dépassements

prévus avec une anticipation positive. Ceci nous permet de déterminer les seuils de dépassement ayant des échantillons de taille suffisante pour obtenir des scores d'évaluation robustes.

Ensuite, les résultats de l'anticipation des dépassements observés et prévus avec les différentes prévisions et les 2 modélisations hydrologiques sont analysés.

### a) Echantillons des dépassements de seuil sans prévision de pluie non nulle

Le Tableau 1 et la Figure 15 permettent de comparer le nombre des dépassements observés et prévus sans intégrer de prévision de pluie non nulle pour les 781 bassins et les 8 événements étudiés (ce qui représente environ 6200 bassins-événements). On distingue les dépassements générés par chacun des deux modèles, GRSD pour le semi-distribué et GRDc pour le distribué continu, pour les 3 seuils relatifs de débit de période de retour de 2, 5 et 10 ans, en incluant, d'une part, tous les dépassements prévus en avance et en retard (statistiques notées « sans anticipation ») et, d'autre part, uniquement les dépassements prévus avec une anticipation positive (statistiques notées « avec anticipation »). On rappelle que les dépassements prévus en retard vis-à-vis des dépassements observés sont comptabilisés comme des alertes manquées pour les scores de contingence.

Dépassements	Sans anticipation			Avec anticipation		
	≥ Q2ans	≥ Q5ans	≥ Q10ans	≥ Q2ans	≥ Q5ans	≥ Q10ans
<b>Qobs</b>	228	81	45	228	81	45
<b>QGRSDsim</b>	273	120	66	240	106	60
<b>QGRSD[pnul]</b>	273	120	66	242	107	60
<b>QGRDcsim</b>	218	79	50	134	48	28
<b>QGRDc[pnul]</b>	238	84	54	195	66	39

Tableau 1 : Nombre de dépassements observés, simulés et prévus avec une pluie future nulle avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils relatifs de quantiles de période de retour de 2, 5 et 10 ans

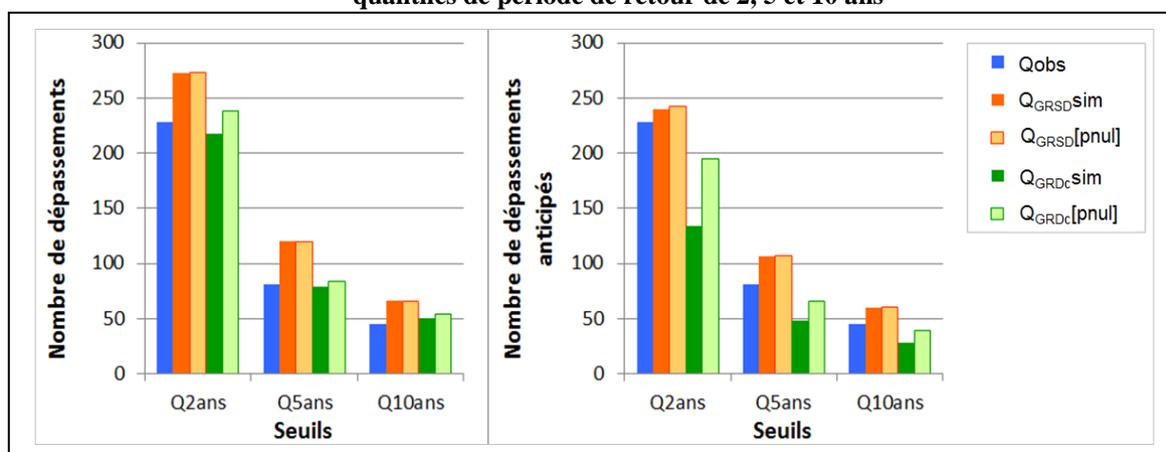
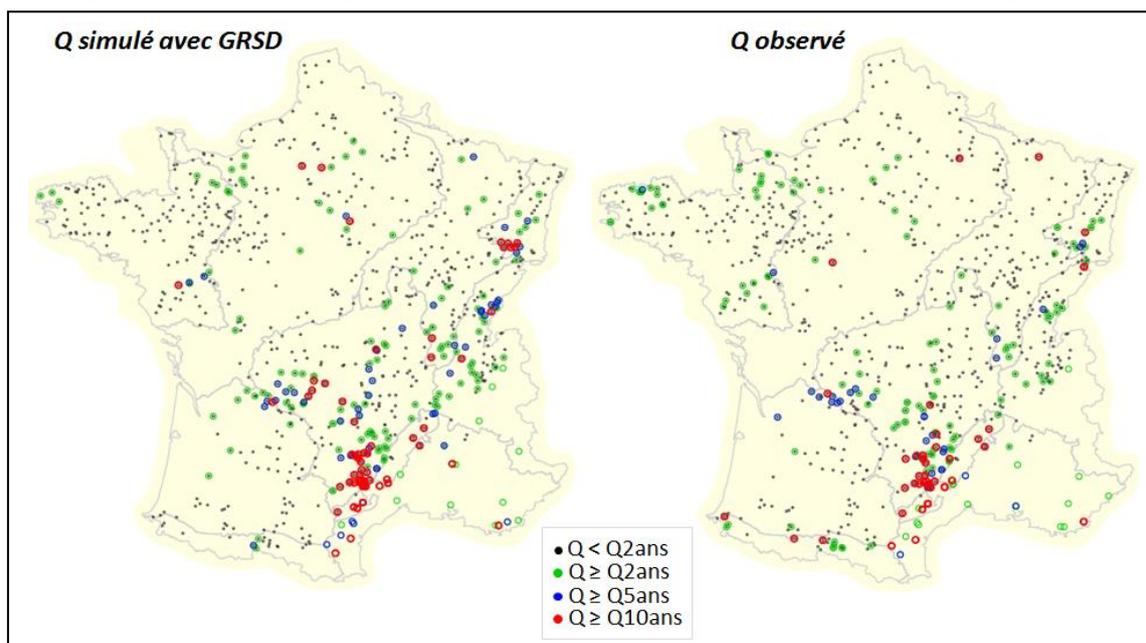


Figure 15 : Histogrammes des dépassements observés, simulés et prévus avec une pluie future nulle avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils relatifs de quantiles de période de retour de 2, 5 et 10 ans, en incluant, à gauche, tous les dépassements prévus en avance et en retard et, à droite, uniquement les dépassements prévus avec une anticipation positive

On rappelle que les seuils relatifs ont des valeurs différentes pour les débits issus des observations et ceux issus de chacun des modèles hydrologiques. Par construction, sur la période d'estimation de ces quantiles de 2002-2016, le nombre de dépassements observés et simulés se correspondent. Par contre, pour les 8 événements étudiés de septembre 2014 jusqu'à février 2016, on note que le modèle GRSD a tendance à surestimer le nombre de dépassements pour les 3 seuils, alors que, pour le modèle GRDc et sans considérer l'anticipation des dépassements (Figure 15 à gauche), le nombre des dépassements simulés et prévus est très similaire au nombre des dépassements observés. On note également l'impact d'intégrer une prévision de pluie nulle pour le modèle GRDc (qui prend en compte un temps de parcours au sein du bassin lors de la sommation des débits au pixel). L'intégration de la pluie future nulle permet de produire un nombre plus élevé de dépassements prévus (issus des valeurs de débit sur les pas de temps de  $T_0 + 1h$  jusqu'à  $T_0 + 6h$ ) en comparaison aux dépassements issus des débits simulés (« prévisions sans échéance »), qui n'incluent que la valeur du débit à  $T_0$ . Cependant, lorsque l'on prend en compte l'anticipation des dépassements, le modèle GRDc a tendance à sous-estimer le nombre de dépassements. En effet, le modèle GRDc a été développé pour mieux reproduire l'hydrogramme observé, sans chercher à anticiper la montée des débits. Il semble que cette première version du modèle GRDc avec les prévisions de pluie nulle génère des crues qui sont généralement trop en retard.

Pour compléter cette comparaison des échantillons des dépassements observés et des dépassements simulés et prévus avec une pluie future nulle, la Figure 16 présente une comparaison des cartes des dépassements maximaux observés et simulés avec GRSD sur l'ensemble des 8 événements étudiés. Même si globalement, on observe une bonne cohérence spatiale des dépassements observés et simulés, on peut noter des différences, qui vont se traduire par des alertes manquées et des fausses alertes dans les scores de contingence.



**Figure 16 : Comparaison des cartes de dépassements observés et des dépassements simulés issus du modèle semi-distribué avec la catégorie maximale de période de retour sur les 8 événements de l'étude**

Etant donné de la taille réduite des échantillons des dépassements observés et simulés avec les 2 modèles GRSD et GRDc pour la période de retour de 10 ans (cf. Tableau1), on présente ci-dessous les résultats sur le degré d'anticipation et les scores de contingence pour les périodes de 2 ans et 5 ans uniquement.

## ***b) Degré d'anticipation avec des prévisions de pluie déterministes***

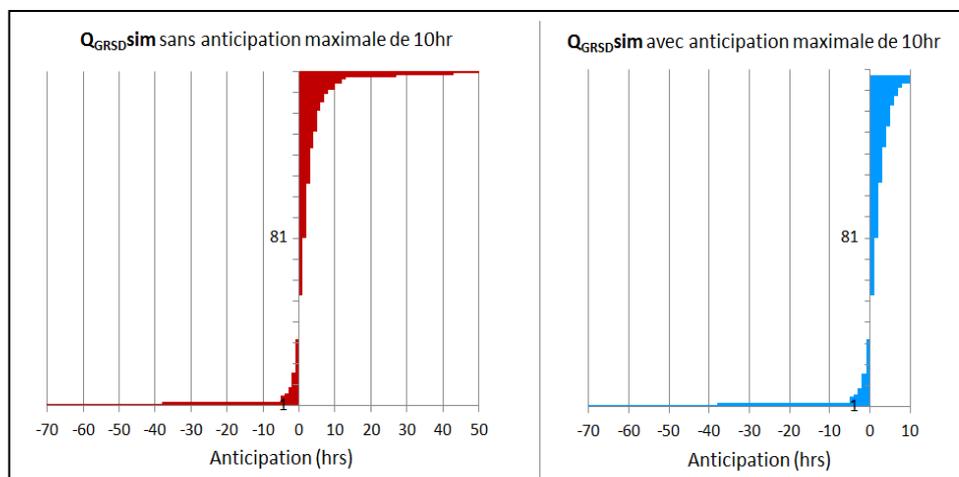
L'objectif est d'évaluer l'apport des prévisions de pluie déterministes AROME-PI sur l'anticipation des dépassements observés, en comparaison avec les avertissements actuels issus des débits simulés (« prévisions sans échéance »), et les avertissements produits en intégrant une prévision de pluie nulle et une prévision de pluie persistante. L'évaluation de l'anticipation nécessite de comparer le premier dépassement simulé ou prévu correspondant à un dépassement observé. On rappelle que, pour le choix du premier dépassement simulé ou prévu correspondant à un dépassement observé et pour l'estimation du degré d'anticipation, une valeur de tolérance maximale de 10 heures a été fixée.

### **Impact de la tolérance maximale**

La Figure 17 illustre l'impact du choix d'une valeur maximale d'anticipation sur les échantillons de dépassements simulés et observés, ainsi que sur la distribution du degré d'anticipation. La Figure 17 se base sur les degrés d'anticipation obtenus avec le modèle GRSD (actuellement intégré au système Vigicrues-Flash) pour les dépassements simulés et observés pour le seuil de 2 ans, en incluant tous les pas de temps horaires des 8 événements.

Sur la gauche en Figure 17, la distribution est relative aux 160 dépassements obtenus sans fixer une anticipation maximale de 10 heures : seuls 6 dépassements simulés ont un degré d'anticipation supérieur à 10 heures, incluant 3 anticipations de 27h, 43h et 50h, et le degré d'anticipation médian est de +1h. Comme décrit dans la section méthodologique, les très fortes valeurs d'anticipation ne peuvent être considérées comme des dépassements prévus corrects correspondants aux dépassements observés.

Lorsque l'on fixe l'anticipation maximale à 10 heures, on obtient la distribution du degré d'anticipation des 157 dépassements simulés et observés sur la droite en Figure 17. On peut noter que 20% des dépassements sont en retard et 13% des dépassements ont une anticipation nulle ; le degré d'anticipation médian est de +1h, avec 90% des valeurs d'anticipation comprises entre -3h et +7h.



**Figure 17 : Distributions du degré d'anticipation des dépassements simulés avec le modèle semi-distribué GRSD pour le seuil de débit biennal, sans inclure d'anticipation maximale de 10h (à gauche) et en intégrant une anticipation maximale de 10h (à droite)**

Tous les résultats d'anticipation et de contingence globale présentés par la suite sont basés sur la valeur maximale d'anticipation de 10h.

### **Apport des AROME-PI déterministes**

Afin d'évaluer l'apport des prévisions AROME-PI, on inclut uniquement dans les résultats suivants les débits simulés et prévus pour les pas de temps horaires pour lesquels les prévisions de pluie AROME-PI sont disponibles. Sur 5 des 8 événements de l'étude (événements de fin novembre 2014, début et fin novembre 2015, et début février 2016), les prévisions AROME-PI manquent pour la construction des ensembles timelag pour un total de 75 pas de temps horaires. En excluant ces pas de temps horaires de l'analyse, on observe une légère diminution du nombre des dépassements simulés et observés, par exemple de 160 à 147 pour la période de retour de 2 ans avec le modèle GRSD.

Comme l'apport des prévisions AROME-PI dépend du modèle hydrologique employé, les résultats sont présentés tout d'abord avec le modèle hydrologique semi-distribué GRSD (employé dans Vigicrues-Flash), et ensuite avec le modèle hydrologique GRDc. L'analyse de la distribution du degré d'anticipation pour le seuil de période de retour de 2 ans pour les dépassements simulés et dépassements prévus issus de la pluie future nulle, la pluie observée et la prévision déterministe AROME-PI est présentée pour le modèle GRSD en Figure 18 et sur le Tableau 2, et pour le modèle GRDc en Figure 19 et sur le Tableau 3.

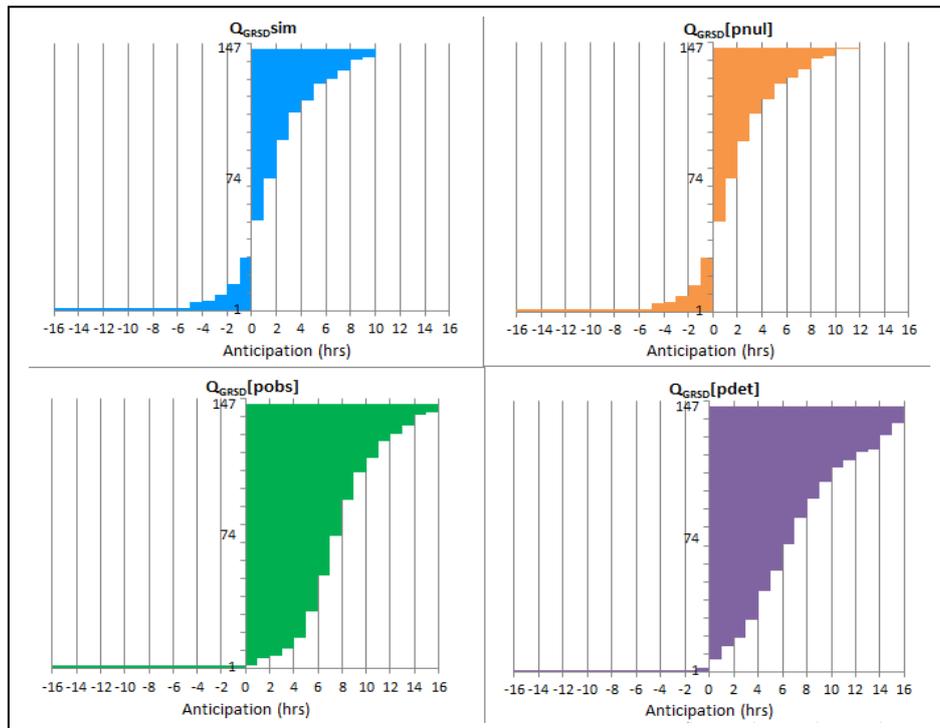


Figure 18 : Distributions du degré d'anticipation des 147 dépassements simulés ( $Q_{GRSDsim}$ ) et prévus issus de la pluie future nulle ( $Q_{GRSD[pnul]}$ ), la pluie observée ( $Q_{GRSD[pobs]}$ ) et la prévision déterministe AROME-PI ( $Q_{GRSD[pdet]}$ ) avec le modèle semi-distribué pour le seuil de période de retour de 2 ans

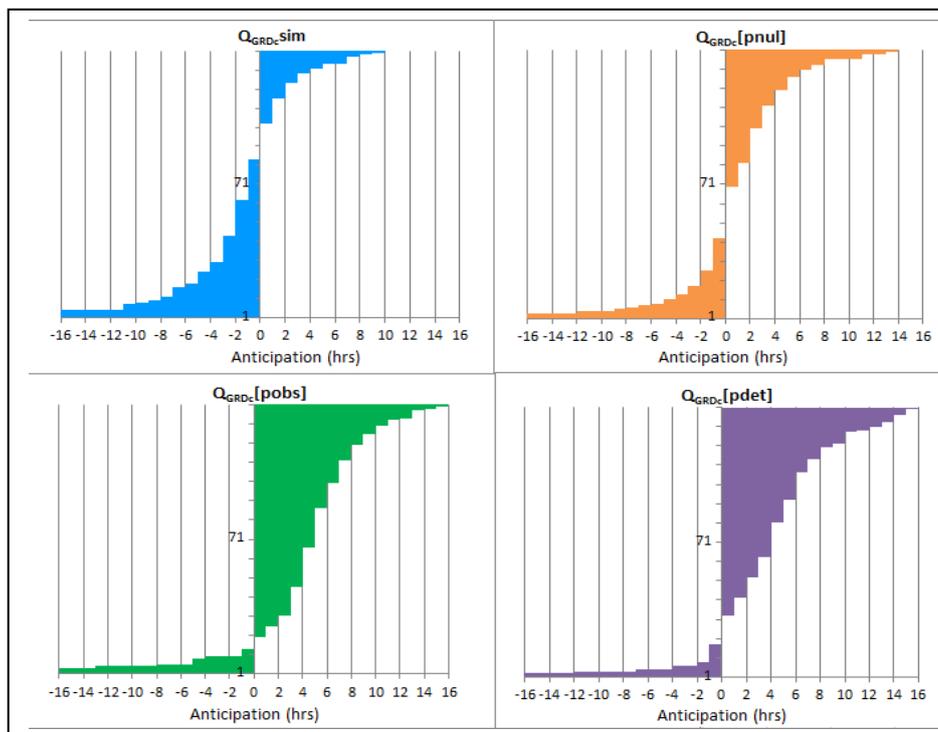


Figure 19 : Distributions du degré d'anticipation des 140 dépassements simulés ( $Q_{GRDc-sim}$ ) et prévus issus de la pluie future nulle ( $Q_{GRDc[pnul]}$ ), la pluie observée ( $Q_{GRDc[pobs]}$ ) et la prévision déterministe AROME-PI ( $Q_{GRDc[pdet]}$ ) avec le modèle distribué continu pour le seuil de période de retour de 2 ans

<b>147 dépassements de Q2ans</b>	<b>Q<sub>GRSDsim</sub></b>	<b>Q<sub>GRSD[pnul]</sub></b>	<b>Q<sub>GRSD[pobs]</sub></b>	<b>Q<sub>GRSD[pdet]</sub></b>
<b>en retard</b>	20,4%	20,4%	1,4%	1,4%
<b>anticipation nulle</b>	14,3%	13,6%	0%	3,4%
<b>90% des anticipations</b>	[-3h, +8h]	[-3h, +8h]	[+3h, +14h]	[+1h, +16h]

**Tableau 2 : Statistiques de la distribution du degré d'anticipation des 147 dépassements simulés (Q<sub>GRSDsim</sub>) et prévus avec la pluie future nulle (Q<sub>GRSD[pnul]</sub>), la pluie observée (Q<sub>GRSD[pobs]</sub>) et la prévision déterministe (Q<sub>GRSD[pdet]</sub>) issu de la pluie déterministe AROME-PI avec le modèle semi-distribué pour le seuil de période de retour de 2 ans**

<b>140 dépassements de Q2ans</b>	<b>Q<sub>GRDc sim</sub></b>	<b>Q<sub>GRDc [pnul]</sub></b>	<b>Q<sub>GRDc [pobs]</sub></b>	<b>Q<sub>GRDc [pdet]</sub></b>
<b>en retard</b>	59,3%	30%	9,3%	12,1%
<b>anticipation nulle</b>	13,6%	19,3%	4,3%	10,7%
<b>90% des anticipations</b>	[-10h, +5h]	[-6h, +8h]	[-5h, +12h]	[-2h, +14h]

**Tableau 3 : Statistiques de la distribution du degré d'anticipation des 140 dépassements simulés (Q<sub>GRDc sim</sub>) et prévus avec la pluie future nulle (Q<sub>GRDc [pnul]</sub>), la pluie observée (Q<sub>GRDc [pobs]</sub>) et la prévision déterministe (Q<sub>GRDc [pdet]</sub>) issu de la pluie déterministe AROME-PI avec le modèle distribué continu**

Pour le modèle GRSD, l'anticipation des dépassements est significativement améliorée avec la prévision AROME-PI : la proportion de dépassements prévus en retard qui est de 20% avec le débit simulé est inférieure à 2% avec AROME-PI. Les erreurs des pluies AROME-PI ne permettent pas d'obtenir une distribution du degré d'anticipation similaire aux dépassements issus de la pluie future parfaite mais 90% des degrés d'anticipation des 147 dépassements issus des prévisions AROME-PI sont supérieures à +1h.

En comparaison aux résultats avec le modèle semi-distribué GRSD, les distributions du degré d'anticipation du modèle distribué continu GRDc montrent des degrés d'anticipation plus faibles pour toutes les prévisions déterministes de débit, même en incluant la pluie future parfaite (i.e. la pluie observée) : environ 9% des dépassements prévus avec Q<sub>GRDc [pobs]</sub> sont en retard vis-à-vis des dépassements observés. Cependant, la prévision déterministe AROME-PI permet d'améliorer significativement l'anticipation des dépassements vis-à-vis des prévisions basées sur une pluie future nulle : par exemple, la proportion de dépassements prévus en retard passe de 30% avec Q<sub>GRDc [pnul]</sub> à 12% avec AROME-PI.

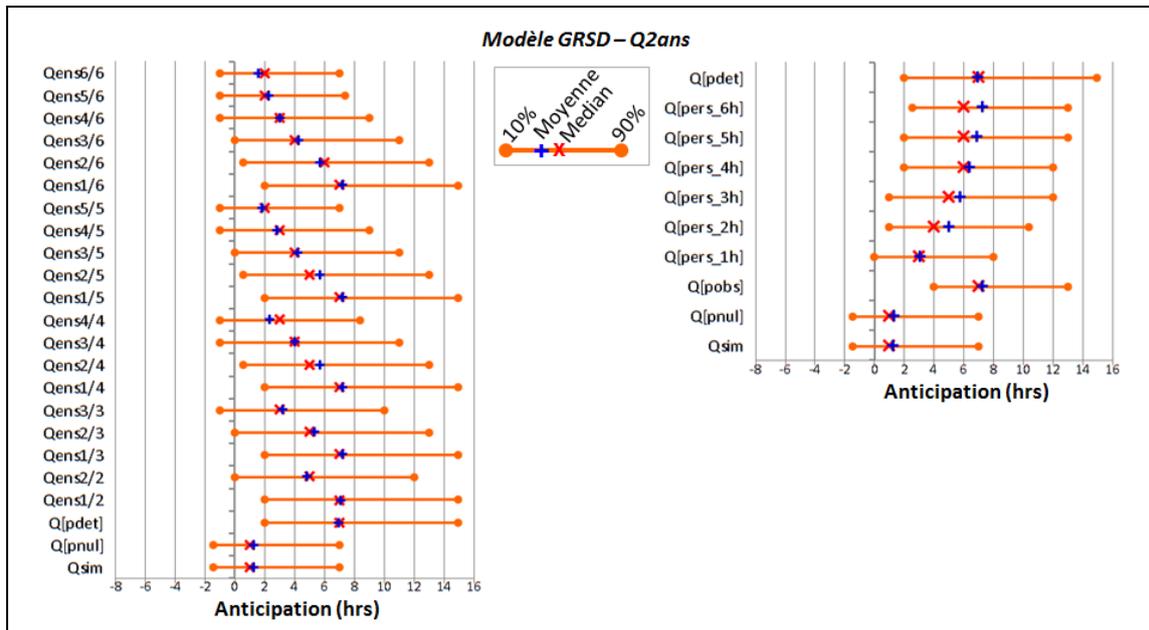
Il faut noter que les très fortes valeurs d'anticipation pour les prévisions issues de la pluie future parfaite (i.e. pluie observée), notamment pour le modèle GRSD, sont dues au modèle hydrologique très réactif. Avec les prévisions AROME-PI, de fortes valeurs d'anticipation peuvent aussi être dues à une surestimation des pluies observées.

**c) Degré d'anticipation avec des prévisions de pluie déterministes et ensemblistes**

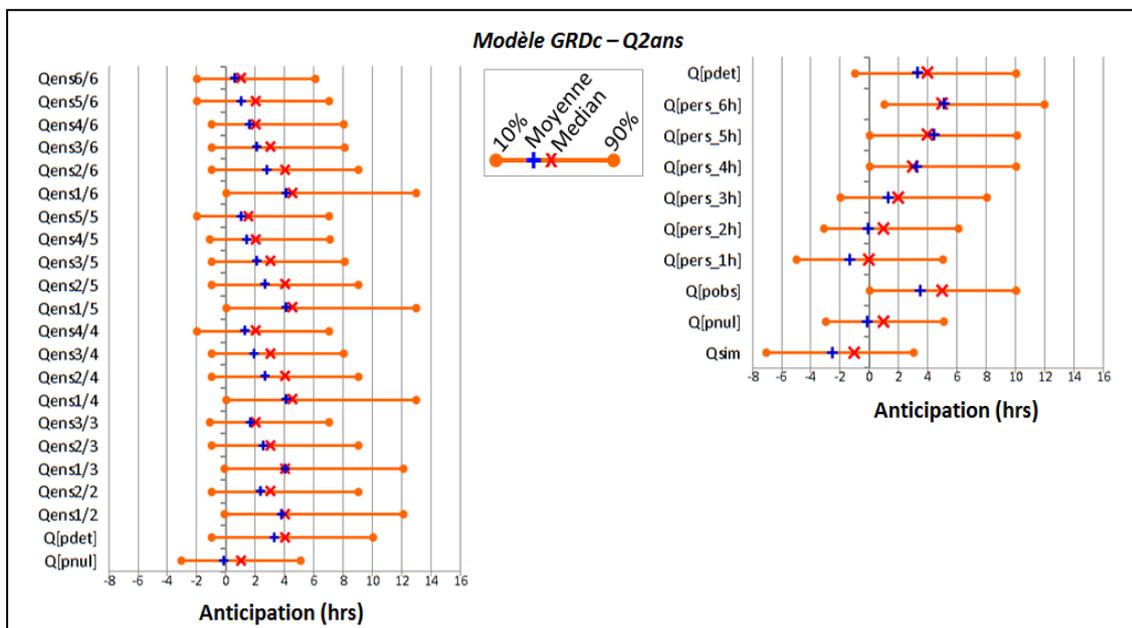
Pour analyser l'apport des prévisions déterministes et ensemblistes AROME-PI, les Figures 20 et 21 pour le seuil de 2 ans, et les Figures 24 et 25 pour le seuil de 5 ans, permettent de comparer la distribution des degrés d'anticipation issus des différentes prévisions en représentant l'intervalle incluant 80% des anticipations, la valeur médiane et la moyenne.

Sur les graphes de gauche, les distributions issues des prévisions ensemblistes, pour tous les niveaux de probabilité et pour un nombre  $K$  de membres variant de 2 à 6, sont comparées aux distributions d'AROME-PI déterministes et les prévisions de référence (débit simulé ou issu de la pluie future nulle). Les avertissements issus de la prévision d'ensemble avec  $K$  membres lorsque  $I$  membres sur  $K$  membres dépassent le seuil de débit sont notés  $Q_{ensI/K}$  (avec  $K$  variant de 2 à 6,  $I$  variant de 1 à  $K$ ).

Sur les graphes de droite, on compare les distributions pour toutes les prévisions déterministes, notamment les prévisions de pluie persistante pour les échéances variant entre +1h et +6h. Les Figures 22 et 23 représentent la distribution des gains en anticipation (pour chaque dépassement) vis-à-vis de  $Q_{GRSDsim}$  et de  $Q_{GRDC[pnul]}$ .



**Figure 20 : Distributions du degré d'anticipation des 147 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle semi-distribué pour le seuil de période de retour de 2 ans**



**Figure 21 : Distributions du degré d'anticipation des 140 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle distribué continu pour le seuil de période de retour de 2 ans**

Pour le modèle GRSD (Figure 20), la prévision déterministe AROME-PI est supérieure en termes d'anticipation à toutes les prévisions de pluie persistante, avec une anticipation médiane de l'ordre de +7h. Avec les prévisions d'ensembles timelag, lorsque tous les membres de l'ensemble définissent les dépassements prévus (niveau de probabilité le plus élevé), les résultats d'anticipation sont légèrement meilleurs que ceux du système actuel (avec le débit simulé  $Q_{GRSDsim}$ ) avec une anticipation médiane entre +2h (pour Qens6/6 par exemple) et +5h (pour Qens2/2). Mais, lorsqu'un seul membre de l'ensemble est considéré pour définir les dépassements (niveau de probabilité le plus faible), l'anticipation est

significativement améliorée vis-à-vis du débit simulé : valeur médiane de +7h et valeur du percentile 10% améliorée de +3h. L'anticipation ne semble pas significativement améliorée en comparaison avec la prévision déterministe AROME-PI (le degré moyen d'anticipation est très légèrement supérieur). Il semble que l'intégration des membres 5 et 6 dans la prévision d'ensemble n'apporte aucune amélioration notable du degré d'anticipation avec les différents niveaux de probabilité. Il y a donc une information pertinente dans les prévisions AROME-PI des 3 dernières heures lorsqu'elles sont intégrées dans les ensembles timelag, en complément de la prévision AROME-PI la plus récente.

Pour le modèle GRDc (Figure 21), malgré des degrés d'anticipation plus réduits, les prévisions déterministes et ensemblistes AROME-PI permettent d'améliorer significativement les résultats. La prévision déterministe AROME-PI est supérieure en termes d'anticipation aux prévisions de pluie persistante jusqu'à +3h d'échéance, avec une anticipation médiane de l'ordre de +4h (contre une anticipation médiane de +1h pour  $Q_{GRDC}[pnul]$ ). On observe des résultats similaires avec les prévisions d'ensembles timelag : une faible amélioration pour le niveau de probabilité le plus élevé, mais pour le niveau de probabilité le plus faible (i.e. 1 seul membre dépassant le seuil), l'anticipation est significativement améliorée vis-à-vis de  $Q_{GRDC}[pnul]$  : valeur médiane passant de +1h à +4h, valeur du percentile 10% de -3h à 0h. L'anticipation est aussi légèrement améliorée en comparaison avec la prévision déterministe AROME-PI lorsque les prévisions AROME-PI des 3 dernières heures sont intégrées dans les ensembles timelag, en complément de la prévision AROME-PI la plus récente.

Afin de comparer, pour chaque dépassement, le gain en anticipation avec les prévisions AROME-PI, les Figures 22 et 23 représentent la distribution du gain en anticipation, pour le modèle GRSD en comparaison avec le système actuel utilisant le débit simulé ( $Q_{GRSDsim}$ ), et pour le modèle GRDc, vis-à-vis des prévisions issues de la pluie future nulle ( $Q_{GRDC}[pnul]$ ). Pour le modèle GRSD (Figure 22) et la prévision déterministe AROME-PI, 80% des gains en anticipation sont compris entre +1h et +11h, avec un gain médian de +5h. Les prévisions d'ensemble avec le niveau de probabilité le plus faible (i.e. 1 seul membre dépassant le seuil) améliorent très légèrement le gain moyen, jusqu'à +6h, à partir de l'intégration d'au moins 3 membres dans la prévision d'ensemble. Pour le modèle GRDc (Figure 23), les gains en anticipation sont plus réduits, de l'ordre de +2h pour la prévision déterministe AROME-PI et de +3h pour les prévisions d'ensemble avec le niveau de probabilité le plus faible.

On peut également comparer les gains en anticipation obtenus avec AROME-PI avec ceux de la prévision de la pluie parfaite,  $Q[pobs]$ . Pour les 2 modèles, l'intervalle de 80% des gains en anticipation avec AROME-PI est plus large que l'intervalle du gain avec  $Q[pobs]$  : percentile 10% égal à +1h et percentile 90% égal à +11h pour GRSD, percentiles 10% 90% de +0h à +7h pour GRDc (contre +0h à +5h for  $Q[pobs]$ ), les gains médians et moyens étant légèrement plus faibles que ceux de  $Q[pobs]$ . Ceci indique que les pluies AROME-PI ont des erreurs de sous-estimation et surestimation, entraînant, dans le premier cas, des gains d'anticipation nuls ou très faibles, et dans le second cas, de larges gains d'anticipation. Comme le montrent les résultats des scores de contingence, ces erreurs de sous-estimation et surestimation des pluies futures avec AROME-PI entraînent également des avertissements manqués et des fausses alertes, qui vont dégrader les scores de contingence. Cependant, on peut noter que ces résultats se basent sur les prévisions de débit maximal sur toute l'échéance de 6 heures des prévisions AROME-PI. En intégrant les prévisions AROME-PI sur des échéances plus courtes, on pourrait potentiellement réduire les biais de surestimation et sous-estimation (les prévisions à plus courte échéance étant généralement de meilleure qualité). Mais cela réduirait aussi le degré d'anticipation des dépassements.

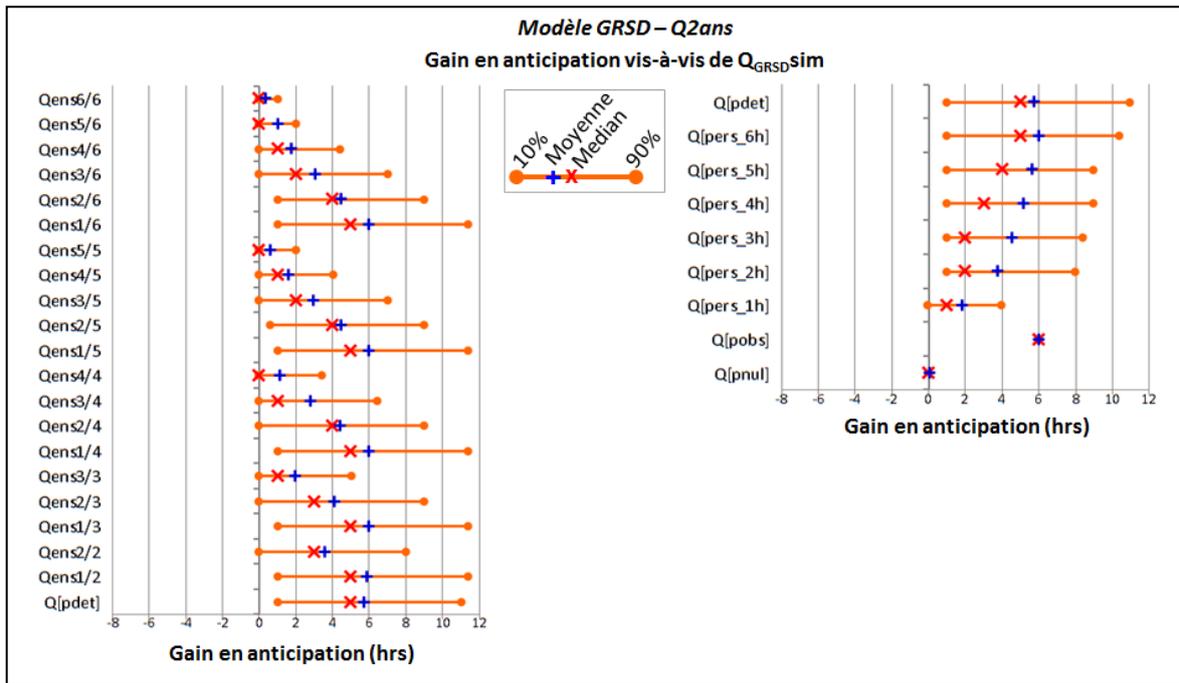


Figure 22 : Distributions du gain en anticipation des 147 dépassements observés et prévus avec le modèle semi-distribué vis-à-vis des dépassements simulés ( $Q_{GRSDsim}$ ) avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) pour le seuil de période de retour de 2 ans

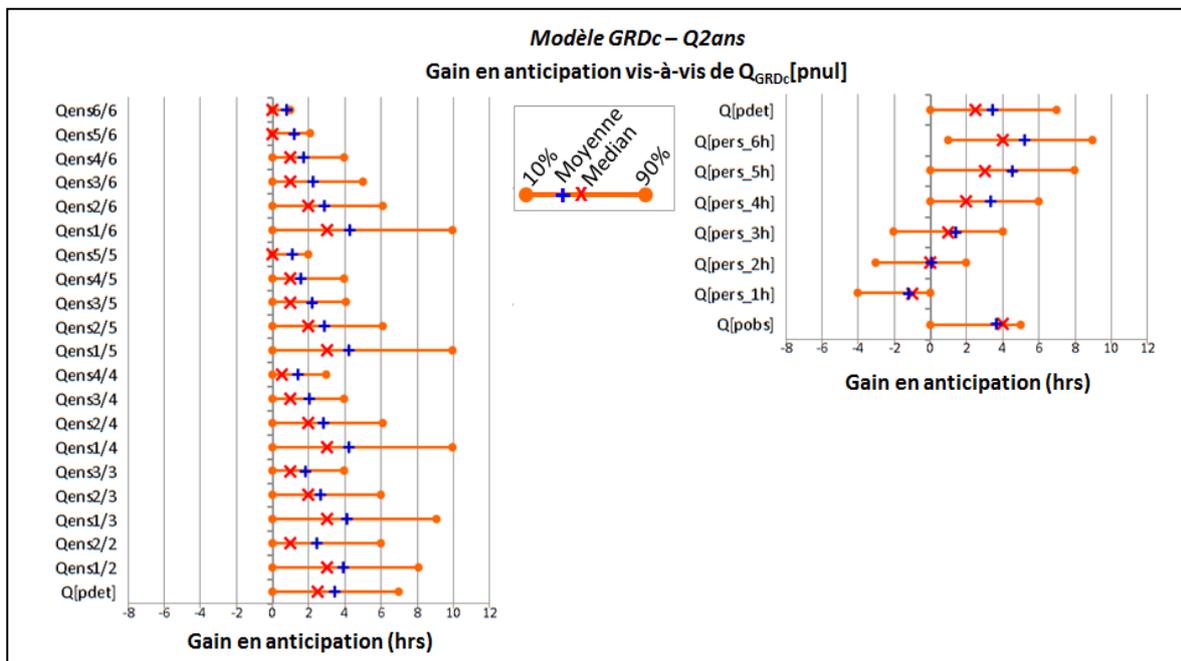
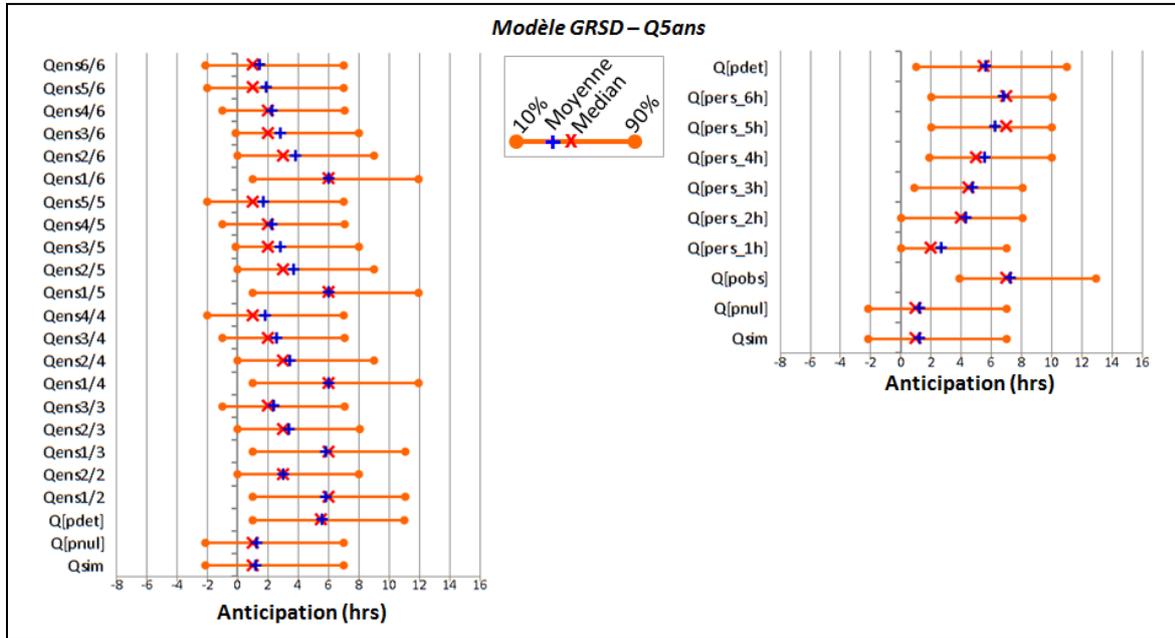


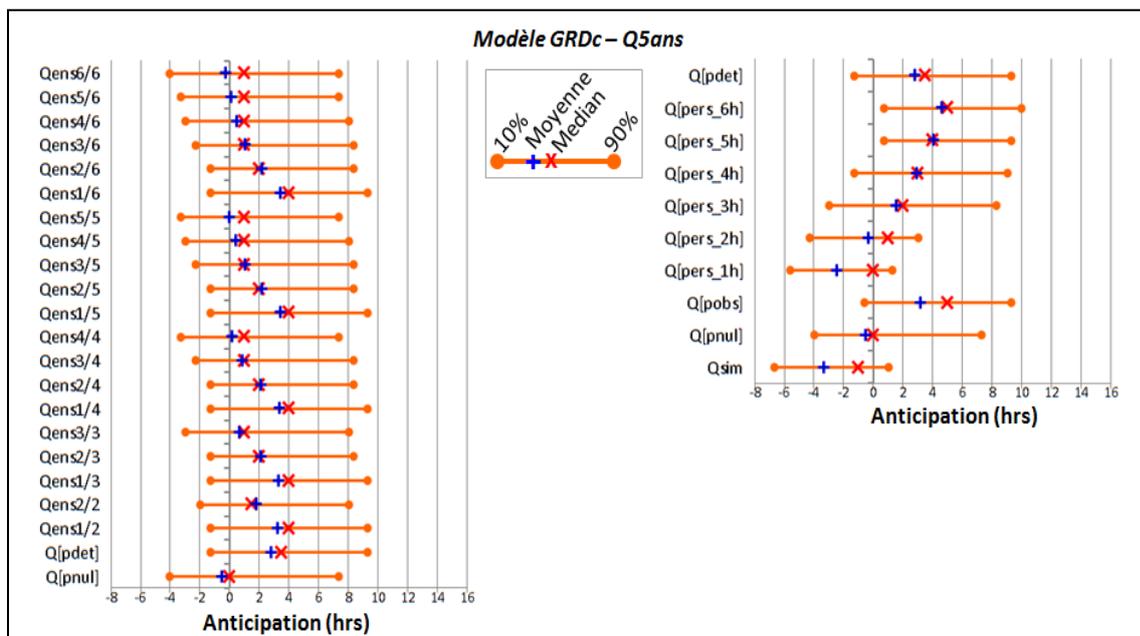
Figure 23 : Distributions du gain en anticipation des 140 dépassements observés et prévus avec le modèle distribué continu vis-à-vis des dépassements prévus avec une pluie future nulle ( $Q_{GRDc[pnul]}$ ) avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) pour le seuil de période de retour de 2 ans

Les Figures 24 et 25 présentent les distributions des degrés d'anticipation pour les modèles GRSD et GRDc et la période de retour de 5 ans. Les degrés d'anticipation sont réduits en comparaison de ceux obtenus pour le seuil de 2 ans : par exemple, le degré médian d'anticipation avec AROME-PI est de +5h (au lieu de +7h pour le seuil de 2 ans) pour le

modèle GRSD et d'environ +3h pour le modèle GRDc (au lieu de +4h pour le seuil de 2ans). Pour les prévisions AROME-PI déterministes et ensemblistes avec le niveau de probabilité le plus faible (Qens1/4, Qens1/5 et Qens1/6), le gain en anticipation reste significatif : environ +3h en gain médian pour GRSD (par rapport au débit simulé) et +2h pour GRDc (par rapport à  $Q_{GRDc}[pnul]$ ).



**Figure 24 : Distributions du degré d'anticipation des 50 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle semi-distribué pour le seuil de période de retour de 5 ans**



**Figure 25 : Distributions du degré d'anticipation des 48 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle distribué continu pour le seuil de période de retour de 5 ans**

En conclusion, cette analyse du degré d'anticipation montre l'intérêt d'inclure, dans la modélisation hydrologique, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag pour améliorer significativement l'anticipation des dépassements de seuil. Avec le modèle semi-distribué du système Vigicrues-Flash actuel, le gain médian est de +5h pour le seuil de 2 ans et de +3h pour le seuil de 5 ans. L'anticipation est légèrement améliorée pour les prévisions d'ensembles timelag avec 4 membres et lorsque l'on définit le dépassement avec le niveau de probabilité le plus faible (1 seul membre dépassant le seuil). Les membres 5 et 6 (pour lesquels les valeurs des prévisions de pluie AROME-PI ne concernent que les 2 premières heures d'échéance et la première heure respectivement) ne permettent pas d'améliorer le degré d'anticipation. Avec le modèle distribué continu, les degrés d'anticipation sont légèrement plus faibles, mais les gains sont significatifs vis-à-vis de des prévisions issues de la pluie future nulle : +3h pour le seuil de 2 ans et de +2h pour le seuil de 5 ans.

Cette analyse de l'anticipation est limitée aux dépassements qui sont prévus et observés. Les résultats de contingence ci-dessous permettent d'analyser la qualité de détection et la fiabilité de tous les dépassements prévus, en incluant les fausses alertes et les alertes manquées.

## **5. Scores de contingence des avertissements**

Les résultats de contingence des avertissements sont tout d'abord présentés pour la contingence globale (sur l'ensemble d'un événement, en comparant un seul pic de débit observé avec un seul pic de débit prévu), puis pour la contingence horaire (pour chaque pas de temps horaire de l'événement). Pour ne pas pénaliser la réactivité du modèle hydrologique, qui peut conduire à des dépassements prévus en avance sur les dépassements observés, une valeur maximale d'anticipation des dépassements prévus est prise en compte pour comptabiliser les bonnes alertes, les fausses alertes et les alertes manquées.

En faisant l'hypothèse que les scores d'évaluation ont des caractéristiques homogènes sur toute la France, les scores de contingence sont présentés après agrégation des résultats sur tous les bassins et sont indicatifs de l'apport potentiel des prévisions AROME-PI pour les avertissements aux crues soudaines sur l'ensemble du territoire.

### ***a) Scores de contingence globale des avertissements***

Cette section présente les résultats de la contingence globale pour les dépassements des seuils relatifs de période de retour de 2 et 5 ans. On rappelle que les scores de contingence sont relatifs aux 781 bassins et aux 8 événements. L'anticipation des dépassements prévus est prise en compte : si le dépassement est prévu en retard vis-à-vis du dépassement observé, il est considéré comme une alerte manquée. Les résultats sont tout d'abord relatifs aux seuils de débit sans tolérance sur les valeurs des seuils. Dans un deuxième temps, une tolérance de 10% sur les valeurs de seuil est prise en compte pour juger de l'amélioration potentielle des performances dans une approche expertisée du système d'avertissement, pour laquelle un prévisionniste ou un utilisateur peut ajuster la production ou la réaction aux avertissements de crue en abaissant la valeur du seuil.

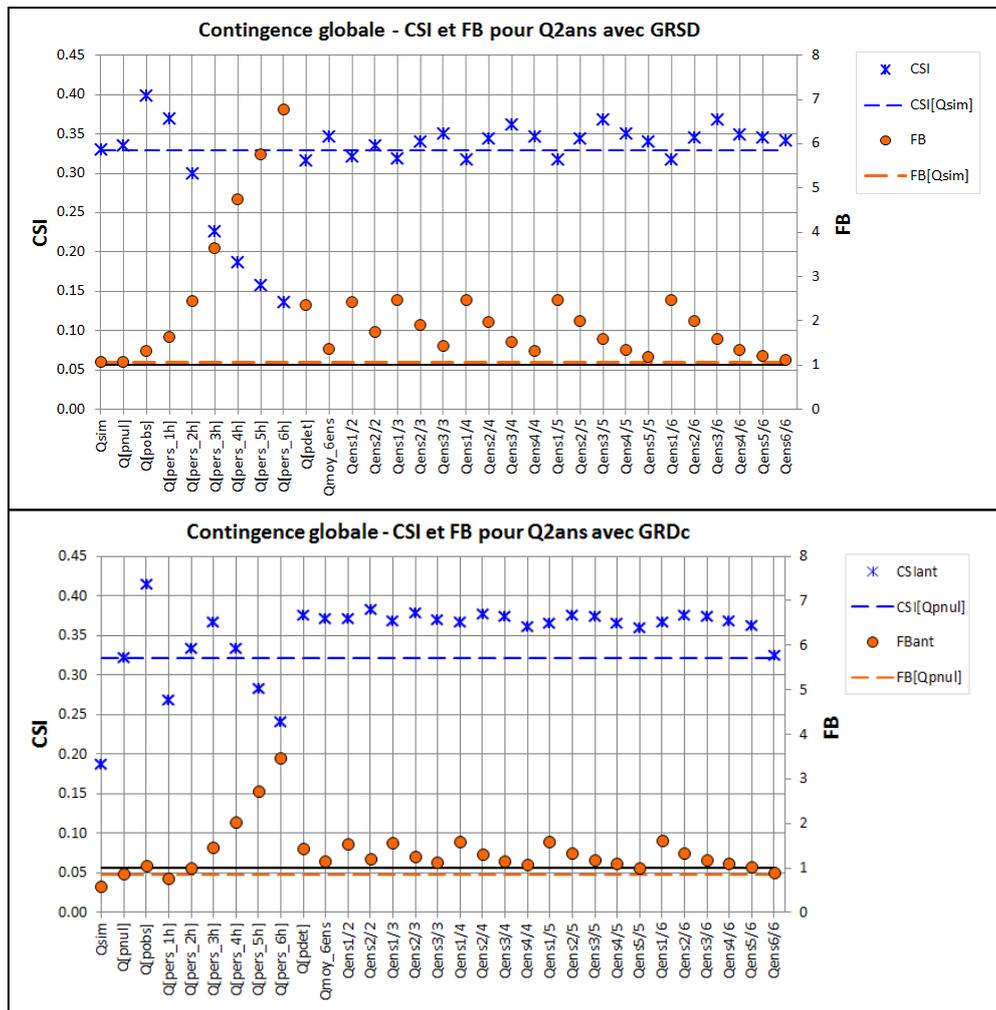
### **Scores de contingence sans tolérance sur les seuils**

Les deux graphes de la Figure 26 permettent de comparer le CSI (qui prend en compte le nombre d'alertes manquées et de fausses alertes) et le biais en fréquence (FB) obtenus pour le seuil de 2 ans avec les différentes prévisions et la référence : débit simulé pour le modèle

GRSD et  $Q_{GRDC}[pnul]$  pour le modèle distribué continu. Il faut noter que, d'après les discussions avec les prévisionnistes du SCHAPI, il est moins problématique d'avoir un système d'avertissement produisant un plus grand nombre de fausses alertes que d'alertes manquées si cela permet de limiter le nombre d'alertes manquées. Par contre, il est difficile de quantifier cette tolérance du prévisionniste aux fausses alertes, sachant qu'une trop grande proportion de fausses alertes décrédibilise un système d'avertissement auprès des utilisateurs.

Pour le modèle GRSD, le biais est légèrement positif pour le débit simulé (utilisé dans le système actuel) et le débit prévu avec la pluie future nulle : le biais en fréquence est de 105% et 106% respectivement pour le seuil de 2 ans. Par contre, pour toutes les autres prévisions, le biais en fréquence est supérieur à 111% : les prévisions ont tendance à produire plus de fausses alertes que d'alertes manquées. Par exemple,  $Q_{GRSD}[pobs]$  obtient un biais en fréquence de 131% pour le seuil de 2 ans, ce qui signifie que les dépassements ont été prévus 1,31 fois plus souvent qu'ils n'ont été observés. Concernant le CSI,  $Q_{GRSDsim}$  obtient un score de 33% ; pour  $Q_{GRSD}[pnul]$ , le CSI est de 34%. La prévision déterministe AROME-PI dégrade légèrement le CSI, avec un CSI de 32%. Pour les prévisions d'ensembles AROMEPI-timelag, ce sont les prévisions Qens3/5 et Qens3/6 qui améliore le plus significativement le CSI, avec un score de 37%. Les prévisions persistantes sur une échéance d'au moins +2h obtiennent des scores significativement dégradés. Il faut noter que la prévision de pluie observée obtient un CSI de 40%, ce qui signifie que, de tous les dépassements prévus et/ou observés, seulement 40% ont été correctement prévus par le modèle en intégrant une pluie future « parfaite » sur l'échéance de +6h. Les erreurs des prévisions de pluie AROME-PI dégradent donc légèrement le score de CSI (de 3% pour Qens3/5 et Qens3/6 par exemple). Pour le seuil de 5 ans, les scores sont dégradés, mais la hiérarchie des performances des différentes prévisions est similaire : CSI de 34% avec  $Q_{GRSD}[pobs]$ , 26% avec  $Q_{GRSDsim}$ , et 29% avec Qens5/6 (par exemple). On peut noter que le biais positif est plus important pour le seuil de 5 ans, y compris pour le débit simulé (utilisé dans le système actuel) et le débit prévu avec la pluie future nulle, avec des scores de 131% et 132% respectivement ;  $Q_{GRSD}[pobs]$  obtient un biais en fréquence de 152%.

Pour le modèle GRDC, la prévision de pluie observée conduit à un léger biais en fréquence (valeur de 104% avec un nombre de fausses alertes légèrement supérieur au nombre d'alertes manquées). Pour la prévision de pluie nulle, GRDC a tendance à produire plus d'alertes manquées que de fausses alertes (biais de 86%). (On peut également noter la nécessité d'inclure une prévision de pluie future nulle pour l'analyse des dépassements produits avec  $Q_{GRDC}[pnul]$ , les scores obtenus avec le débit simulé  $Q_{GRDCsim}$  étant très dégradés). Toutes les prévisions issues des AROME-PI (sauf Qens5/5 et Qens6/6) obtiennent des biais supérieurs à 1 et améliorent le CSI par rapport à  $Q_{GRDC}[pnul]$  : valeurs de CSI entre 36% et 38%, pour un CSI de 32% avec  $Q_{GRDC}[pnul]$ . Avec la prévision de pluie observée, le CSI est de 41%. Pour les prévisions issues de la pluie persistante, c'est l'échéance de +3h qui obtient le meilleur CSI, d'une valeur de 37%. Les scores sont dégradés pour le seuil de 5 ans, mais la hiérarchie des performances des différentes prévisions est similaire : CSI de 38% avec  $Q_{GRDC}[pobs]$ , 28% avec  $Q_{GRDC}[pnul]$ , et 29% avec  $Q_{GRDC}[pdet]$  et Qens3/5 (par exemple).

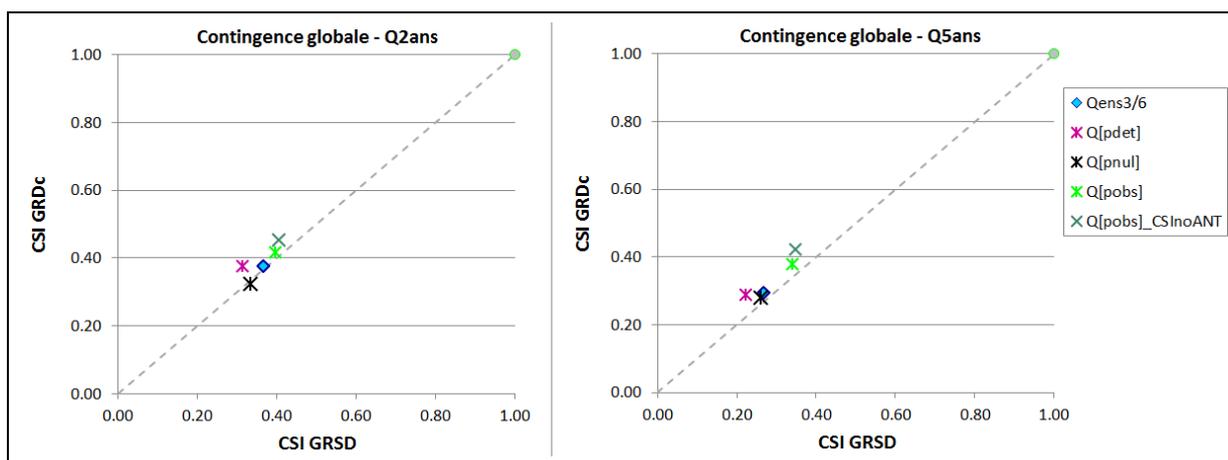


**Figure 26 : Scores de contingence CSI (échelle sur la gauche) et de Biases en Fréquence (FB, échelle sur la droite) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans**

La Figure 27 permet de comparer le score de CSI (prenant en compte l'anticipation) qui a été obtenu avec les 2 modèles hydrologiques pour les prévisions issues de la pluie future nulle, la pluie future parfaite, la prévision AROME-PI déterministe, ainsi que les prévisions d'ensemble à 6 membres pour lesquelles le dépassement est prévu quand, au minimum, 3 membres dépassent le seuil. On a également ajouté les valeurs de CSI avec la pluie future parfaite sans prise en compte de l'anticipation des dépassements prévus (noté Q[pobs]\_CSI<sub>noANT</sub>). Globalement, pour les seuils de 2 et 5 ans, les 2 modèles obtiennent des scores de CSI similaires, légèrement supérieurs pour le modèle GRDc (à l'exception de la prévision Q[pnul] pour laquelle GRSD est meilleur pour le seuil de 2 ans). Les prévisions AROME-PI déterministes (Q[pdet]) dégradent légèrement le CSI pour GRSD (car elles augmentent trop le nombre de fausses alertes, même si elles permettent aussi de diminuer le nombre d'alertes manquées), mais améliorent le CSI pour GRDc. Par contre, les dépassements issus des prévisions d'ensemble Qens3/6 améliorent le CSI, avec un même score de 37% pour les 2 modèles. Pour le seuil de 5 ans, les différences entre les 2 modèles et pour l'apport des AROME-PI sont plus faibles.

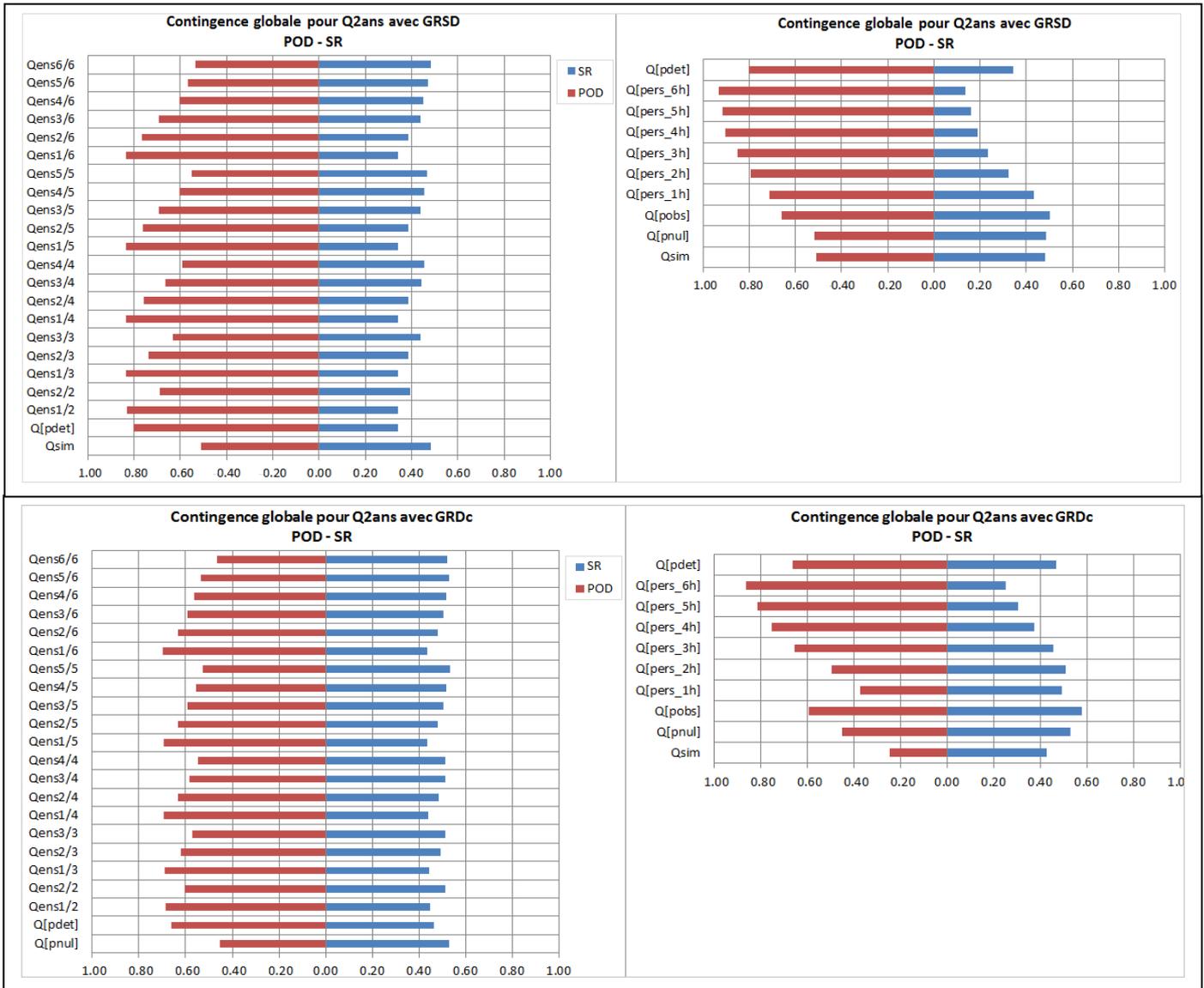
Par ailleurs, pour juger du potentiel des 2 modèles s'ils pouvaient intégrer des pluies futures parfaites sur une échéance de +6h et évaluer l'impact de la capacité des modèles à anticiper les dépassements, on compare les valeurs de CSI avec et sans prise en compte de

l'anticipation (notés  $Q[pobs]$  et  $Q[pobs]_{CSInoANT}$  respectivement). Les 2 scores sont très similaires pour GRSD, alors que la prise en compte de l'anticipation dégrade plus significativement le CSI pour le modèle GRDc. Le modèle GRDc apparaît comme légèrement meilleur en CSI que le modèle GRSD, notamment lorsque l'anticipation n'est pas considérée. Le modèle GRSD ayant été construit pour anticiper les dépassements avec le débit simulé, le score de CSI est globalement moins impacté par la prise en compte de l'anticipation. Par contre, en considérant l'anticipation, le modèle GRDc a une performance dégradée (de 45% à 41% pour le seuil de 2 ans) du fait du retard des dépassements prévus. On rappelle qu'une valeur de CSI de 41% signifie que, de tous les dépassements prévus et/ou observés, seulement 41% ont été correctement prévus. A noter que le Gilbert Skill Score (qui adapte la formulation du CSI pour prendre en compte les alertes correctes obtenues par chance) obtient une même hiérarchie entre les 2 modèles et les différentes prévisions (les valeurs de GSS sont globalement de 2% et 1% plus faibles que celles de CSI pour les seuils de 2 ans et 5 ans respectivement).



**Figure 27 : Comparaison des scores de CSI avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de période de retour de 2 et 5 ans**

Afin de différencier les erreurs d'alertes manquées et de fausses alertes, les Figures 28 et 29 comparent la Probabilité de Détection (POD, défini sur l'ensemble des dépassements observés, et qui diminue avec un nombre plus grand d'alertes manquées) et le Taux de Succès (SR, défini sur l'ensemble des dépassements prévus, et qui diminue avec un nombre plus grand de fausses alertes) pour les seuils de 2 et 5 ans. On cherche à obtenir des valeurs de Probabilité de Détection et de Taux de Succès les plus proches de 1, en minimisant le nombre d'alertes manquées (pour le POD) et le nombre de fausses alertes (pour le SR). Les graphes de gauche comparent les scores pour les dépassements issus des prévisions d'ensemble avec les différents niveaux de probabilité, les graphes de droite, ceux obtenus avec les prévisions déterministes, avec les scores de la prévision de référence ( $Q_{sim}$  pour le modèle GRSD et  $Q_{pnul}$  pour le modèle GRDc).



**Figure 28 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans**

Le modèle GRSD obtient, pour toutes les prévisions et le débit simulé, un meilleur score en POD qu'en SR, donc une meilleure détection des dépassements que de fiabilité de ces dépassements prévus. Pour  $Q_{GRSDsim}$ , le POD de 0,51 signifie que 51% des dépassements observés ont été correctement prévus, alors que le SR de 0,48 signifie que 48% des dépassements prévus ont actuellement été observés. La prévision déterministe AROME-PI  $Q_{GRSD}[pdet]$  permet d'améliorer significativement le POD de 29% (avec un score de 80%) mais en dégradant (dans une moindre mesure) le SR de 14% (score SR de 34%). La pluie persistante sur +1h obtient un score de détection de 71% avec une fiabilité de 43% ; les prévisions issues des pluies persistantes sur une échéance d'au moins +2h dégradent fortement la fiabilité (SR en dessous de 30%). Les prévisions d'ensemble avec les différents niveaux de probabilité permettent de moduler le gain en détection et la perte concomitante de fiabilité, avec des résultats légèrement améliorés vis-à-vis de la prévision déterministe. Pour une meilleure détection, il faut privilégier le niveau de probabilité le plus faible (i.e. un seul membre dépassant le seuil de débit). On obtient par exemple, en comparaison avec le débit

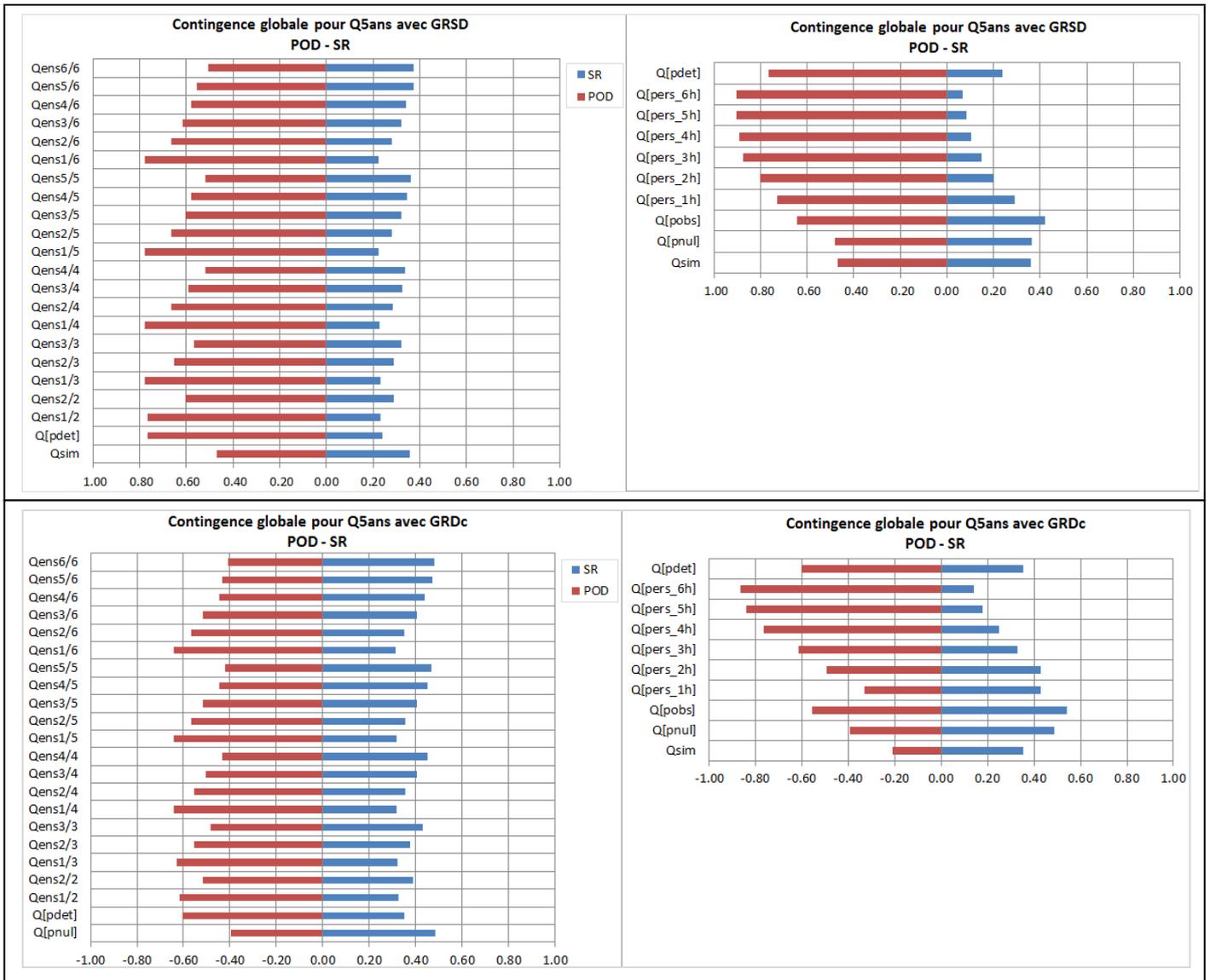
simulé, un gain en détection de 32% (score POD de 83%) et une perte de fiabilité de 14% (score SR de 34%) pour  $Q_{GRSDens1/6}$ , alors que  $Q_{GRSDens6/6}$  ne dégrade pas la fiabilité mais améliore la détection de 3% seulement (score POD de 54%).

Comme noté précédemment, le modèle GRDc est globalement plus fiable que le modèle GRSD. Pour  $Q_{GRDc}[pnul]$ , le POD de 0,45 signifie que 45% des dépassements observés ont été correctement prévus, alors que le SR de 0,53 signifie que 53% des dépassements prévus ont actuellement été observés. La prévision déterministe  $Q_{GRDc}[pdet]$  issue des AROME-PI permet d'améliorer le POD de 21% (avec une valeur de POD de 66%) mais en dégradant le SR de 6% (valeur de SR de 46%). La prévision issue d'une pluie persistante sur une échéance de +3h obtient des scores similaires. Comme avec le modèle GRSD, les prévisions d'ensemble avec les différents niveaux de probabilité permettent de moduler le gain en détection et la perte concomitante de fiabilité, les meilleurs scores de détection étant obtenus avec le niveau de probabilité le plus faible (i.e. un seul membre dépassant le seuil de débit). C'est la prévision  $Q_{GRDc}ens1/6$  qui obtient la meilleure détection de 70% (gain de 25%) pour une fiabilité de 44% (perte de 9%), avec une légère amélioration de la détection vis-à-vis de ceux obtenus avec  $Q_{GRDc}[pdet]$ .

En ce qui concerne les erreurs des prévisions de pluie AROME-PI et leur impact sur les prévisions de débit et les dépassements de seuil, on peut comparer les scores obtenus avec  $Q[pdet]$  et  $Q[pobs]$  pour chacun des 2 modèles (on note une même tendance avec les scores de POD et SR qui ne prennent pas en compte l'anticipation des dépassements). La Probabilité de Détection est supérieure avec les pluies AROME-PI, conduisant à faire moins d'alertes manquées : les prévisions AROME-PI ont une tendance à surestimer les pluies fortes, ce qui permet de corriger la sous-estimation du débit futur par le modèle hydrologique avec les pluies observées. Dans le même temps, le Taux de Succès SR est dégradé avec les pluies AROME-PI car la surestimation des pluies fortes conduit dans certains cas à une augmentation du nombre de fausses alertes.

La Figure 29 présente les scores de détection et fiabilité pour le seuil de 5 ans : en comparaison aux scores pour le seuil de 2 ans, la détection et la fiabilité des dépassements sont plus faibles pour toutes les prévisions et le système de référence, la réduction du score de fiabilité étant généralement plus importante que celui de détection. Les gains en détection et les pertes en fiabilité grâce aux prévisions de pluie AROME-PI sont quasiment similaires à ceux obtenus pour le seuil de 2 ans.

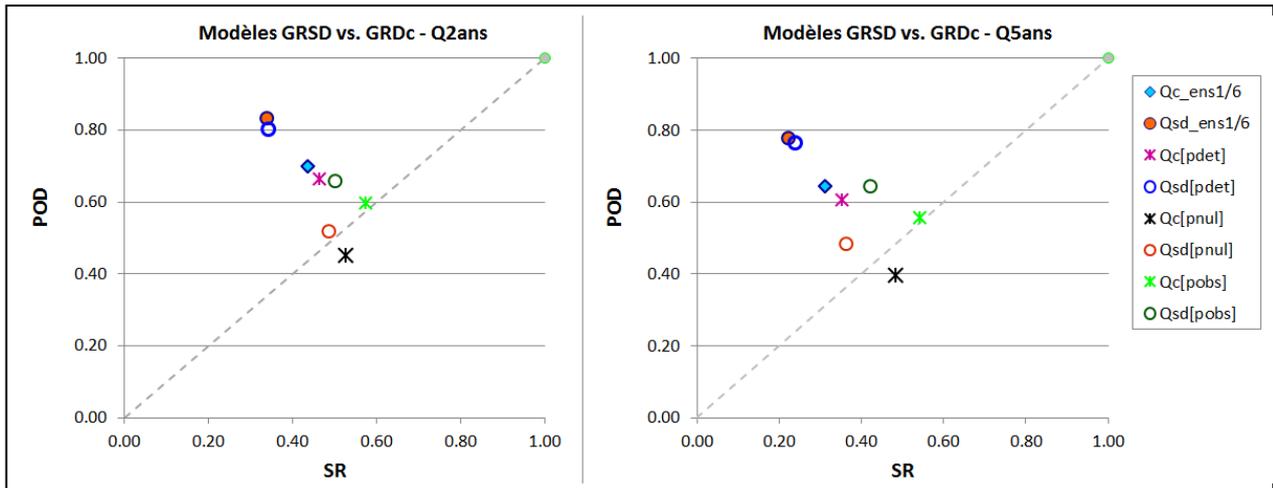
Par exemple, pour le modèle GRSD, la prévision déterministe  $Q_{GRSD}[pdet]$  permet d'améliorer le POD de 30% (avec une valeur de POD de 77%) mais en dégradant le SR de 12% (valeur de SR de 24%). La prévision d'ensemble  $Q_{GRSD}ens1/6$  améliore le POD de 31% (avec une valeur de POD de 78%) mais dégrade le SR de 13% (valeur de SR de 22%), alors que la prévision d'ensemble  $Q_{GRSD}ens6/6$  améliore le POD de 4% seulement (pour obtenir un POD de 51%) mais améliore légèrement le SR de 1% (valeur de SR de 37%). En ce qui concerne le modèle GRDc, la prévision déterministe  $Q_{GRDc}[pdet]$  permet d'améliorer le POD de 21% (avec une valeur de POD de 60%) mais en dégradant le SR de 13% (valeur de SR de 35%) ; la prévision d'ensemble  $Q_{GRDc}ens1/6$  améliore le POD de 25% (avec une valeur de POD de 64%) mais dégrade le SR de 17% (valeur de SR de 31%).



**Figure 29 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 5 ans**

La Figure 30 permet de comparer la détection et la fiabilité obtenue avec différentes prévisions par les 2 modèles hydrologiques et pour les seuils de 2 ans et 5 ans. Pour comparer de manière cohérente les scores obtenus par les 2 modèles, ce sont les scores des débits prévus avec la pluie future nulle,  $Qsd[pnul]$  et  $Qc[pnul]$  (et non le débit simulé) qui sont représentés. Cependant, pour le modèle GRSD, les scores obtenus avec  $Qsd[pnul]$  sont quasiment équivalents (globalement 1% plus forts) aux scores produits avec le débit simulé (correspondant au système d'avertissement actuel).

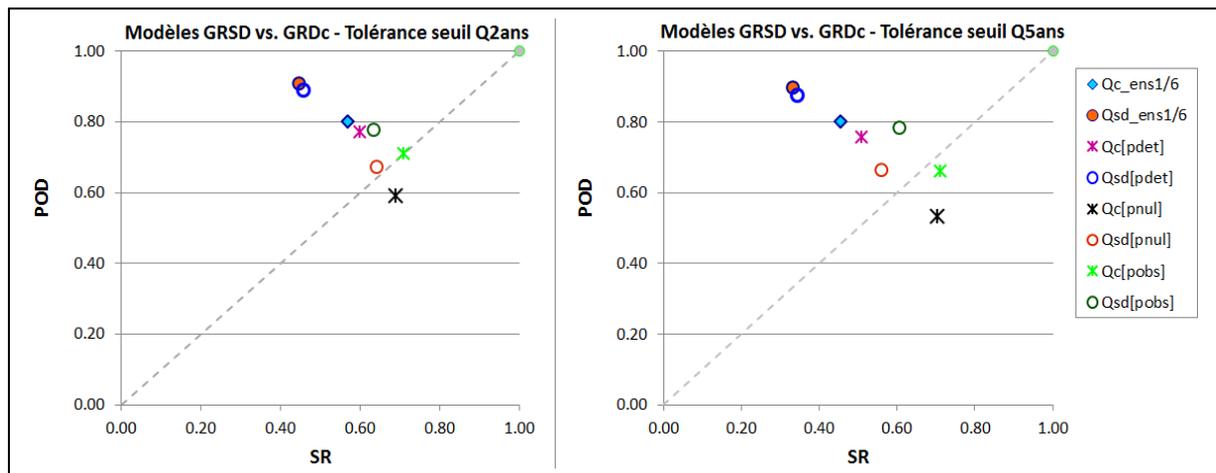
Comme noté précédemment, le modèle GRSD permet une meilleure détection, alors que le modèle GRDc est meilleur en fiabilité. La surestimation des pluies futures par AROME-PI permet d'augmenter significativement la détection vis-à-vis de la prévision basée sur la pluie future parfaite mais, en parallèle, dégrade la fiabilité des dépassements (cette perte de fiabilité étant plus forte pour le seuil de 5 ans). Du fait des performances complémentaires des modèles GRSD et GRDc, il pourrait être intéressant d'évaluer les performances des avertissements issus d'une prévision d'ensemble multi-modèle, en combinant les ensembles de débit issus du modèle GRSD avec ceux issus du modèle GRDc (tout en conservant des valeurs de seuils de dépassement propres à chaque modèle).



**Figure 30 : Comparaison des scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de 2 et 5 ans**

### Scores de contingence avec une tolérance de 10% sur les seuils

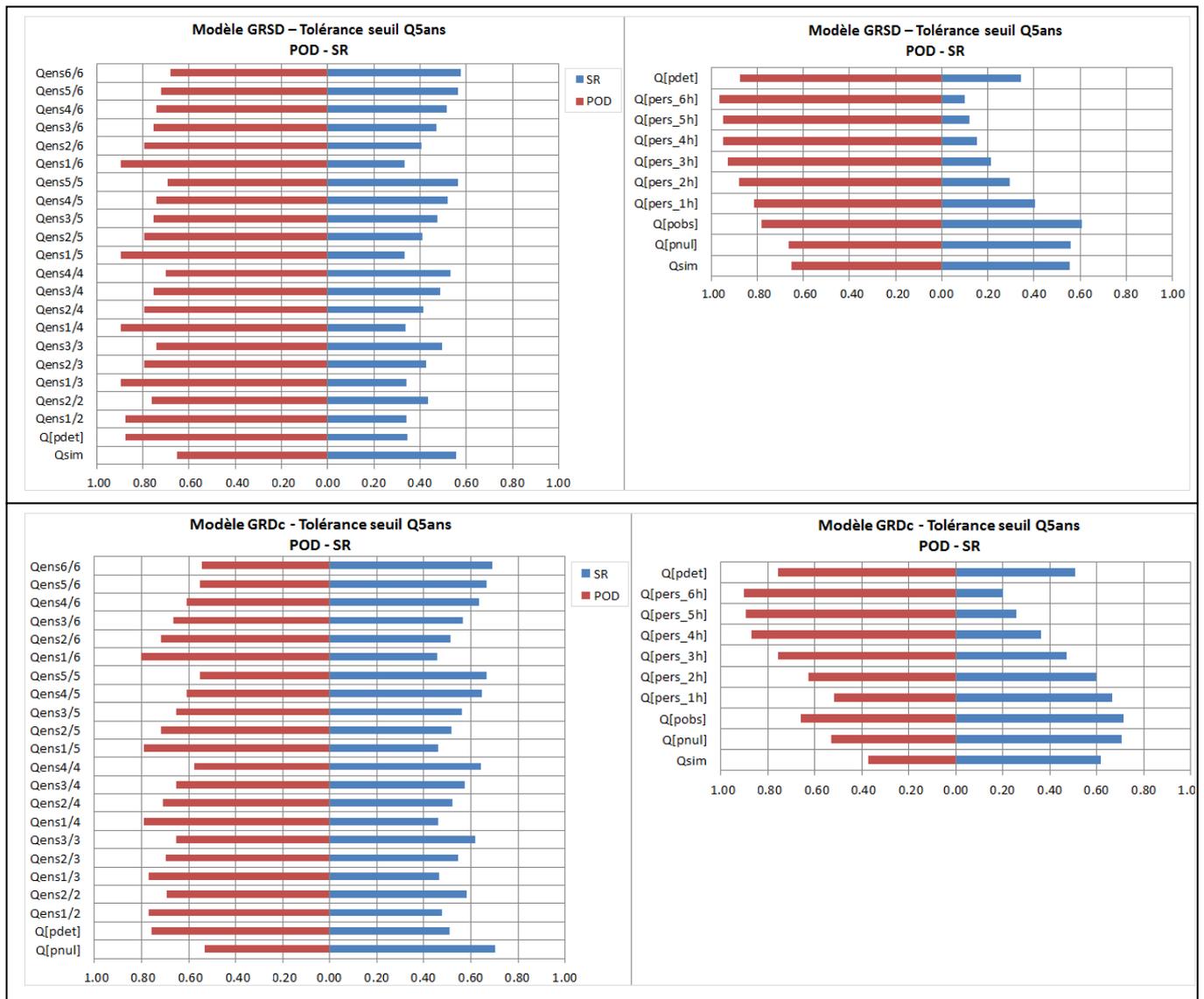
La Figure 31 présente les scores de détection et de fiabilité en intégrant une tolérance de 10% sur les seuils pour inclure, dans les alertes correctes, les alertes manquées de peu et les fausses alertes presque observées. La comparaison de la Figure 30 (sans tolérance) avec la Figure 31 (avec tolérance) permet de juger de l'amélioration des 2 scores, pour une hiérarchie similaire entre les différentes prévisions et les 2 modèles.



**Figure 31 : Comparaison des scores de Probabilité de Détection (POD) et Taux de Succès (SR) basés sur une tolérance de 10% sur les seuils de 2 et 5 ans avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc)**

Concernant l'amélioration des scores en intégrant une tolérance sur les seuils, pour  $Q_{GRSDsim}$  et le seuil de 2 ans, les gains en détection et fiabilité sont de 16% (pour un POD de 67% et un SR de 64%). Pour la prévision déterministe  $Q_{GRSD}[pdet]$ , le gain en détection est de 9% (pour un POD de 89%) et le gain en fiabilité de 12% (pour un SR de 46%). Pour le modèle distribué continu,  $Q_{GRDc}[pnul]$  présente un gain en détection de 14% (pour un POD de 69%) et un gain en fiabilité de 16% (pour un SR de 59%). Pour la prévision déterministe  $Q_{GRDc}[pdet]$ , le gain en détection est de 11% (pour un POD de 77%) et le gain en fiabilité de 14% (pour un SR de 60%). Les gains sont généralement supérieurs pour le seuil de 5 ans : par

exemple, la prévision déterministe  $Q_{GRSD}[pdet]$  présente un gain en détection de 11% et en fiabilité de 10% (pour un POD de 88% et un SR de 34%) et la prévision déterministe  $Q_{GRDc}[pdet]$  présente un gain en détection de 15% et en fiabilité de 14% (pour un POD de 76% et un SR de 51%).



**Figure 32 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) basés sur une tolérance de 10% sur les seuils de période de retour de 5 ans avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc)**

Pour le modèle GRSD, on peut noter que le débit simulé  $Q_{GRSDsim}$  et toutes les prévisions de débit ont des probabilités de détection supérieures à 67% (ce qui signifie que plus de 67% des dépassements observés ont été correctement prévus), avec un maximum de 91% pour  $Q_{GRSDens1/4}$ ,  $Q_{GRSDens1/5}$  et  $Q_{GRSDens1/6}$ . Sur le graphe de gauche, la fiabilité varie entre 45% pour les prévisions d'ensemble avec le plus faible niveau de probabilité ( $Q_{GRSDens1/2}$  à  $Q_{GRSDens1/6}$ ) et 64% pour  $Q_{GRSDsim}$  et  $Q_{GRSDens6/6}$  (une valeur de 64% signifiant que 64% des dépassements prévus ont actuellement été observés). Les prévisions issues de la pluie persistante pour des échéances de plus de 2 heures produisent une surestimation trop forte des débits prévus, ce qui dégrade fortement la fiabilité. Les prévisions AROME-PI  $Q_{GRSD}[pdet]$  permettent d'améliorer la probabilité de détection (gain de 24% avec un POD de 89%), mais

en dégradant la fiabilité (perte de 19% pour un SR de 45%) du fait de la surestimation des pluies prévues. En considérant les prévisions d'ensemble, on peut choisir un niveau de probabilité qui ne dégrade pas autant la fiabilité des dépassements, tout en améliorant la qualité des détections vis-à-vis du débit simulé: par exemple, la prévision  $Q_{GRSDens3/4}$  permet d'obtenir un POD de 81% pour un SR de 58%.

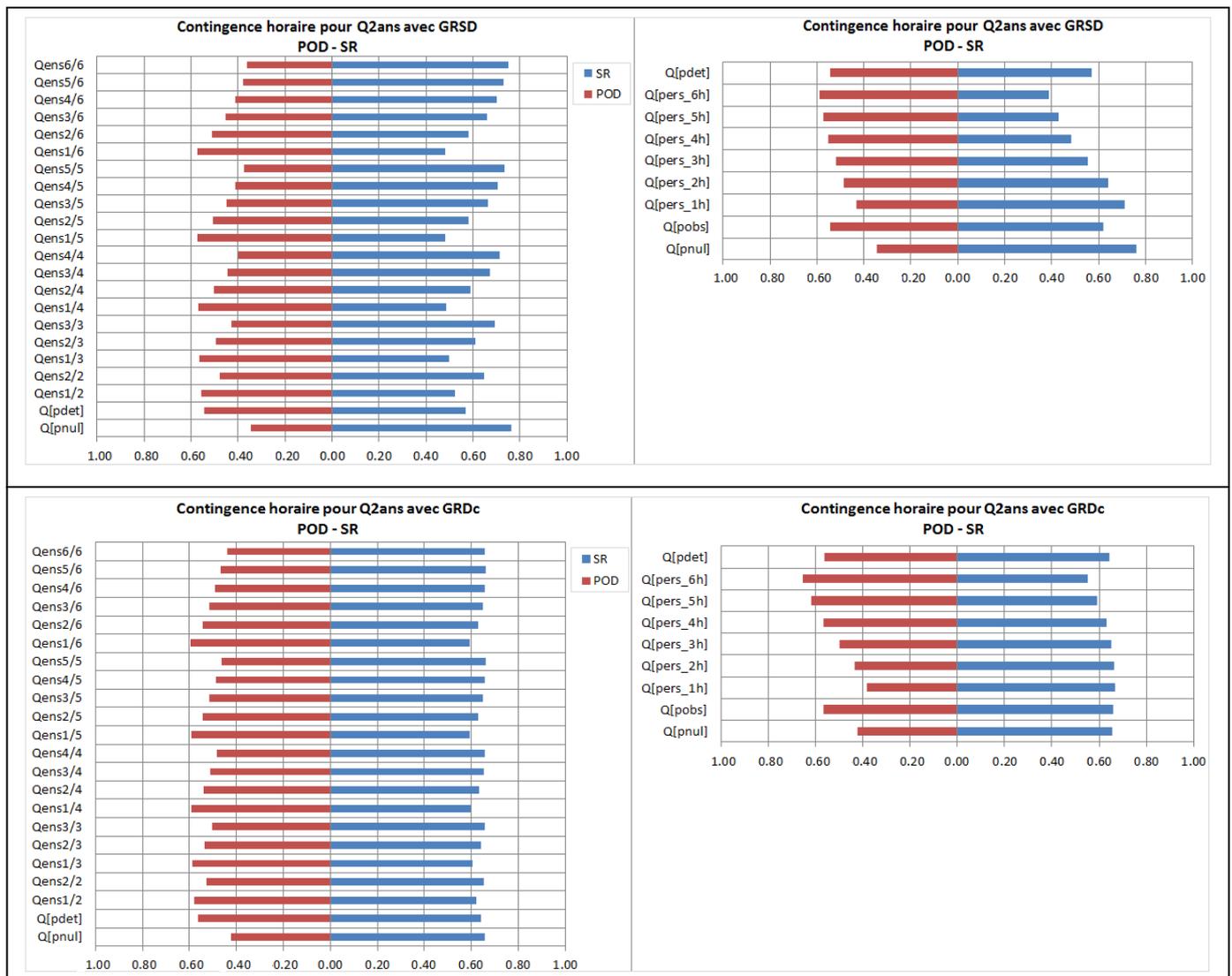
En ce qui concerne le modèle GRDc, les conclusions sont similaires, notamment sur l'apport des prévisions AROME-PI, même si les scores obtenus sont plus faibles en POD et plus forts en fiabilité vis-à-vis des scores avec le modèle GRSD. Par exemple, toutes les prévisions ont des probabilités de détection supérieures à 53%, avec une valeur maximale de 80% pour  $Q_{GRDcens1/6}$ . La fiabilité varie entre 46% pour  $Q_{GRDcens1/6}$  et 70% pour  $Q_{GRDc}[pnul]$ . Les prévisions AROME-PI  $Q_{GRDc}[pdet]$  permettent d'améliorer la probabilité de détection de 23% (avec un POD de 76%), mais en dégradant la fiabilité de 19% (pour un SR de 51%). Pour les prévisions d'ensemble, la prévision  $Q_{GRSDens3/4}$  permet d'obtenir un POD de 65% pour un SR de 57%.

Il faut noter qu'en ce qui concerne les prévisions d'ensemble avec les 2 modèles, les scores sont similaires pour les ensembles avec 4, 5 et 6 membres (pour des niveaux de probabilité comparables). Dans les prévisions d'ensemble timelag de pluie, les informations les plus pertinentes concernent les 4 prévisions AROME-PI les plus récentes, qui permettent d'améliorer la fiabilité des dépassements et, dans une moindre mesure, la qualité de détection vis-à-vis de la prévision déterministe (qui correspond à la plus récente).

### ***b) Scores de contingence horaire des avertissements***

On présente ici les résultats de contingence horaire, lorsque la contingence est établie à chaque pas de temps horaire au cours des 8 événements. Comme la contingence horaire évalue tout l'hydrogramme de débit prévu, et donc potentiellement les multiples dépassements de seuil prévus et/ou observés au cours d'un événement, cette évaluation est plus stricte que l'évaluation de contingence globale (qui ne compare qu'un seul pic de débit prévu avec le pic de débit observé). Pour juger de l'apport des prévisions AROME-PI et comparer les performances obtenues avec les 2 modèles hydrologiques sur une même fenêtre d'échéance, le système d'avertissement de référence se base sur les prévisions de débit issues de la pluie future nulle (et non sur le débit simulé). Ici, on ne présente que les résultats en considérant les valeurs maximales de débits prévus pour toute l'échéance de +6h. Il serait également intéressant d'évaluer les scores de contingence horaire pour les échéances plus courtes de débit (la qualité des prévisions de pluie ayant tendance à se dégrader avec l'échéance, les scores pourraient potentiellement s'améliorer).

La Figure 33 compare la Probabilité de Détection (POD) et le Taux de Succès (SR) estimés pour la contingence horaire et le seuil de 2 ans (les scores sont plus faibles pour le seuil de 5 ans mais avec des hiérarchies similaires entre les différentes prévisions).



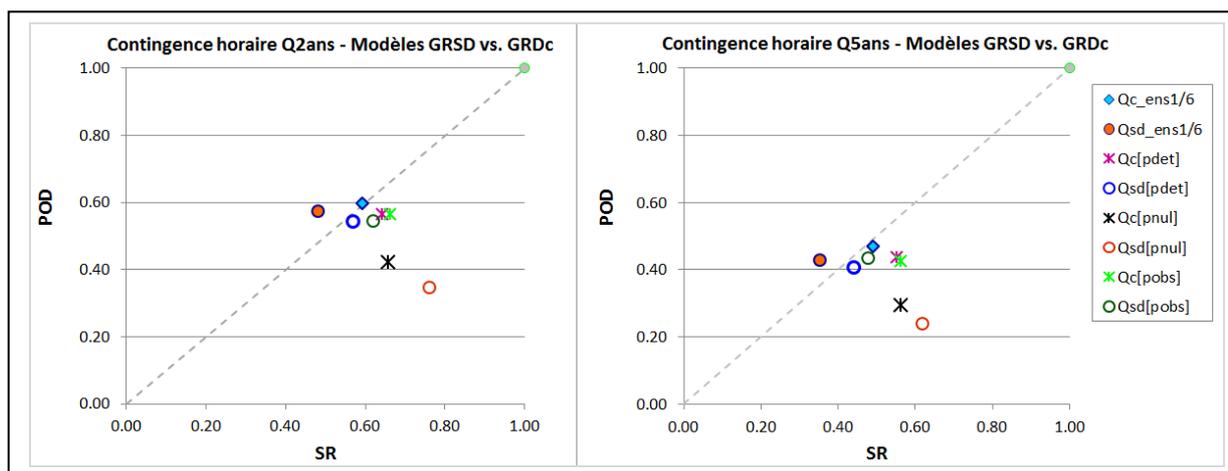
**Figure 33 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) de la contingence horaire avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans**

En comparant les Figures 28 et 33, on peut noter que, globalement, les scores de fiabilité sont meilleurs en contingence horaire qu'en contingence globale, alors que les scores de détection sont plus faibles. Les résultats avec le modèle GRDc (qui est défini pour mieux reproduire tout l'hydrogramme de débit observé) sont relativement stables avec les différentes prévisions AROME-PI déterministes et ensemblistes. En comparaison avec les scores de  $Q_{GRDc}[pnul]$ , la fiabilité est maintenue avec les ensembles pour les niveaux de probabilité les plus élevés alors que le POD est légèrement amélioré (par exemple, pour  $Q_{GRDc}ens3/4$ , gain en POD de 9% pour un score POD égal à 51% et un score SR maintenu à 65%). C'est la prévision  $Q_{GRDc}ens1/6$  qui obtient la meilleure détection de 60% (gain de 17%) pour une fiabilité de 59% (perte de 6%), alors que la prévision déterministe  $Q_{GRDc}[pdet]$  obtient un score de détection de 56% et un score de fiabilité de 64%.

Les scores pour le modèle GRSD sont plus variables en fonction des prévisions et inférieurs aux scores obtenus avec GRDc. En comparant aux scores obtenus avec  $Q_{GRSD}[pnul]$ , la prévision déterministe AROME-PI  $Q_{GRSD}[pdet]$  permet d'obtenir un POD de 54% (gain de 20%) et un SR de 57% (perte de 19%). Comme pour la contingence globale, les prévisions d'ensemble avec les différents niveaux de probabilité permettent de moduler le gain en

détection et la perte concomitante de fiabilité. C'est la prévision  $Q_{GRSDens1/4}$  qui obtient la meilleure détection de 57% (gain de 13%) pour une fiabilité de 49% (perte de 24%), mais avec des scores plus faibles que ceux obtenus avec les prévisions d'ensemble produites par GRDc.

La Figure 34 permet de comparer la détection et la fiabilité obtenues par les 2 modèles hydrologiques avec différentes prévisions pour les seuils de 2 ans et 5 ans. Pour toutes les prévisions de pluie non nulles, et notamment pour les prévisions AROME-PI déterministes et ensemblistes, les scores de détection et de fiabilité sont supérieurs avec le modèle GRDc, avec des écarts entre les 2 modèles plus importants en fiabilité qu'en détection. Pour la prévision de pluie nulle, c'est le modèle GRSD qui obtient un score de fiabilité plus fort (en produisant moins de fausses alertes que GRDc) mais pour une détection plus faible (en manquant plus de dépassements observés que GRDc). Ceci s'explique en partie par le fait que GRSD a tendance à produire des hydrogrammes plus « pointus » que GRDc, de par sa réactivité, et sans nécessairement reproduire la décrue observée.



**Figure 34 : Comparaison des scores obtenus en contingence horaire pour la Probabilité de Détection (POD) et le Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de 2 et 5 ans**

Comme pour la contingence globale, on peut comparer les scores obtenus avec  $Q[pdet]$  et  $Q[pobs]$  pour chacun des 2 modèles pour évaluer l'impact des erreurs des prévisions de pluie AROME-PI sur les prévisions de débit et les dépassements de seuil. Pour les 2 modèles, la Probabilité de Détection pour ces 2 prévisions est relativement similaire (sauf pour GRSD et le seuil de 5 ans, le POD se dégradant de 3%) ; le Taux de Succès est dégradé avec les pluies AROME-PI, avec un impact plus fort pour GRSD (perte de 5% et 4% pour GRSD et de 2% et 1% pour GRDc pour les seuils de 2 ans et 5 ans). Ainsi, le nombre de fausses alertes est globalement plus important avec les prévisions AROME-PI qu'avec la pluie observée. Ceci confirme que, sur tous les pas de temps de l'événement, lorsque l'on considère toute l'échéance de +6h, il y a une tendance générale des pluies déterministes AROME-PI à surestimer les plus fortes pluies, ce qui conduit à un nombre plus important de fausses alertes et donc, à une dégradation de la fiabilité. Cependant, ces erreurs de surestimation des fortes pluies pourraient potentiellement être améliorées en utilisant une échéance de prévision plus courte, la qualité des prévisions de pluie ayant tendance à se dégrader avec l'échéance. Mais cela impliquerait également une diminution de l'anticipation des dépassements observés.

## 6. Conclusions et Perspectives

L'analyse de l'apport des prévisions de pluie immédiate a porté sur 8 événements de crue entre septembre 2014 et février 2016. Les simulations et prévisions de débit sont issues du modèle hydrologique semi-distribué GRSD actuellement utilisé dans le système d'avertissement Vigicrues-Flash, avec les paramètres calibrés et régionalisés de la fin 2015 (Organde 2015), ainsi que du modèle distribué continu GRDC développé par Organde en 2016 (Organde 2016). L'analyse a concerné 781 bassins versants jaugés sur la France entière. 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 nulle, la prévision de pluie persistante et la prévision de pluie « parfaite » (i.e. pluie observée).

L'évaluation de la qualité des avertissements s'est basée sur le dépassement de seuils de débit relatif correspondant aux périodes de retour de 2 ans et 5 ans, pour lesquels les échantillons des dépassements prévus et observés étaient suffisamment larges. L'utilisation de seuils relatifs de débit (ayant des valeurs différentes pour les débits simulés ou prévus et les débits observés) 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.

Les scores de contingence sont estimés par 2 approches complémentaires :

- 1) sur chaque événement dans son ensemble (i.e. pour toutes les dates d'initialisation de prévision de l'évènement) pour la contingence dite globale, qui correspond à une évaluation a posteriori du système d'avertissement, lorsque, pour un événement donné, on peut comparer le premier dépassement de seuil qui a été observé, au premier dépassement de seuil qui a été prévu ;
- 2) pour chaque pas de temps horaire des différents évènements, pour la contingence dite horaire, correspondant à une évaluation plus stricte du système d'avertissement (e.g. évaluation de multiples dépassements prévus et/ou observés au cours de chacun des évènements), qui peut être réalisée en temps réel, dès que les observations de débit sont disponibles.

Pour la contingence globale, les scores de contingence prennent en compte l'anticipation des dépassements prévus : si le dépassement est prévu en retard vis-à-vis du dépassement observé, il est considéré comme une alerte manquée. On définit également une valeur maximale d'anticipation pour les dépassements simulés et prévus afin de comparer, de façon plus cohérente, un dépassement simulé avec le dépassement observé lui correspondant sans pénaliser l'anticipation du modèle hydrologique. Par ailleurs, afin de prendre en compte les incertitudes sur les valeurs de seuils de débit, les scores de contingence ont également été estimés en intégrant une tolérance de 10% sur ces valeurs de seuil. On considère comme des alertes correctes, les cas de dépassements observés et manqués de peu par la prévision, et les cas de fausses alertes presque observées pour des dépassements prévus et non observés. Cette approche (proposée par le Centre d'Ecologie et d'Hydrologie de Wallingford en Grande-Bretagne) permet de juger de l'amélioration potentielle des performances dans une approche expertisée du système d'avertissement, pour laquelle un prévisionniste ou un utilisateur peut ajuster la production ou la réaction aux avertissements de crue en abaissant la valeur du seuil.

Le degré d'anticipation des dépassements observés et les scores de contingence globale et de contingence horaire permettent de comparer la qualité des avertissements basés sur les prévisions de débit issus des prévisions déterministes de pluie AROME-PI et des prévisions d'ensemble AROMEPI-timelag avec les avertissements du système Vigicrues-Flash actuel (utilisant le modèle GRSD et sans prévision de pluie), et avec les avertissements issus d'une pluie future nulle pour les modèles GRSD et GRDc. Les principaux résultats obtenus sont les suivants :

- les gains en anticipation avec la prévision déterministe AROME-PI et la prévision d'ensemble AROMEPI-timelag sont très significatifs pour les 2 modèles, même si le modèle GRSD permet une meilleure anticipation que le modèle GRDc : avec AROME-PI déterministe et le seuil de 2 ans par exemple, anticipation médiane de l'ordre de +7h avec le modèle GRSD (gain médian de +5h vis-à-vis du système actuel) et de +4h avec le modèle distribué continu GRDc (gain médian de +3h vis-à-vis du débit issu d'une pluie future nulle). Egalement, la proportion de dépassements prévus en retard est significativement diminuée. Pour les ensembles timelag, l'anticipation est aussi légèrement améliorée en comparaison à AROME-PI déterministe lorsque les prévisions AROME-PI des 3 dernières heures sont intégrées dans les ensembles timelag, en complément de la prévision AROME-PI la plus récente. Par contre, les membres 5 et 6, provenant des prévisions de pluie AROME-PI les moins récentes et qui ne concernent que les 2 premières heures d'échéance, ne permettent pas d'améliorer l'anticipation.
- en contingence globale, les prévisions déterministes AROME-PI améliorent significativement la détection des dépassements mais avec, généralement, une augmentation du nombre de fausses alertes et donc, une perte de fiabilité. Pour le modèle GRSD, la Probabilité de Détection s'améliore de 51% à 80% avec AROME-PI en comparaison au système actuel, alors que le taux de succès se dégrade de 48% à 34%. Le modèle GRDc obtient une détection plus faible de 66% et un meilleur score de fiabilité de 46% en intégrant AROME-PI. Les prévisions d'ensemble, avec les différents niveaux de probabilité (définis par le nombre de membres de l'ensemble dépassant le seuil de débit), permettent de moduler le gain en détection et la perte concomitante de fiabilité, les meilleurs scores de détection étant obtenus avec le niveau de probabilité le plus faible (i.e. un seul membre de l'ensemble dépassant le seuil de débit).
- pour la contingence horaire, plus stricte (car elle concerne tout l'hydrogramme de chacun des 8 événements), les scores de fiabilité sont globalement meilleurs qu'en contingence globale, alors que les scores de détection sont globalement plus faibles, en particulier pour le modèle GRSD. En comparaison à la prévision issue de la pluie future nulle, on obtient par exemple, pour le modèle GRDc, une détection de 60% (gain de 17%) pour une fiabilité de 59% (perte de 6%) avec la prévision  $Q_{GRDc,ens1/6}$ . Pour le modèle GRSD, la prévision  $Q_{GRSD,ens1/4}$  obtient une probabilité de détection de 57% (gain de 13%) et une fiabilité de 49% (perte de 24%). Pour toutes les prévisions de pluie non nulles, et notamment pour les prévisions AROME-PI déterministes et ensemblistes, les scores de détection et de fiabilité sont meilleurs avec le modèle GRDc qu'avec GRSD, avec des écarts entre les 2 modèles plus importants en fiabilité qu'en détection.
- les résultats obtenus en intégrant, d'une part, les pluies AROME-PI et, d'autre part, les pluies observées, sur toute l'échéance de +6h ont montré l'impact des erreurs des

prévisions de pluie AROME-PI sur les prévisions de débit et les dépassements de seuil, avec une tendance plus marquée à surestimer les plus fortes pluies (mais le biais n'est pas systématique). Ce biais entraîne un nombre plus important de fausses alertes et donc, une perte de fiabilité, mais permet également d'améliorer la détection et l'anticipation des dépassements observés.

En conclusion, cette étude montre l'intérêt d'inclure, dans la modélisation hydrologique du système Vigicrues-Flash, les prévisions immédiates de pluie AROME-PI et les prévisions d'ensemble AROMEPI-timelag. Ces prévisions permettent d'améliorer significativement l'anticipation des dépassements de seuil vis-à-vis de ceux du système actuel sans pluie future, ainsi que le taux de détection, même si elles dégradent partiellement la fiabilité. 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 (défini par le nombre minimum de membres de l'ensemble dépassant le seuil). Il est nécessaire d'analyser le taux de détection et la fiabilité des dépassements prévus en fonction du niveau de probabilité pour définir les dépassements avec les prévisions d'ensemble, afin de gagner en anticipation tout en conservant une fiabilité acceptable. 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 de travail peuvent être proposées pour compléter cette analyse et améliorer la chaîne d'avertissement Vigicrues-Flash :

- du fait des performances complémentaires des 2 modèles hydrologiques (GRSD étant globalement meilleur en détection et GRDc permettant une meilleure fiabilité pour les scores de contingence globale), il serait intéressant d'évaluer les performances des avertissements issus de prévisions d'ensemble multi-modèles, en combinant les ensembles de débit issus du modèle GRSD avec ceux issus du modèle GRDc (tout en conservant des valeurs de seuils de dépassement propres à chaque modèle hydrologique);
- 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 ; 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 (cf. l'évaluation avec les dégâts observés pour les événements de crue de juin 2010 et d'octobre 2015 présentée dans Javelle et al. 2016);
- la qualité des avertissements issus des prévisions AROME-PI devrait également être évaluée en se limitant à des échéances plus courtes car les prévisions de pluie ont tendance à se dégrader sur les échéances les plus longues ; l'évaluation de qualité devrait aussi être conduite pour les prévisions de pluie elles-mêmes (en termes de pluie de bassin et de pluie au pixel) et pour les prévisions de débit afin de déterminer plus précisément des points forts et points faibles de ces prévisions, et ainsi optimiser leur utilisation dans le système d'avertissement ; par ailleurs, l'incertitude sur les

scores obtenus devrait être estimée (par exemple avec une méthode de bootstrap pour déterminer des intervalles de confiance) ;

- les seuils de débit utilisés dans cette étude pourraient également être modulés en fonction des enjeux, pour une prise en compte de la vulnérabilité et ainsi, permettre de fournir des avertissements sur le risque de crue lorsque les informations d'aléa et de vulnérabilité peuvent être croisées ; la vulnérabilité peut également être qualifiée grâce à un indice d'exposition des zones inondables, dépendant de l'occupation des sols et des enjeux, comme il a été proposé par Clotilde Saint-Martin (Saint-Martin et al. 2016) ;
- la chaîne d'avertissement Vigicrues-Flash devrait bénéficier de la combinaison des différentes prévisions de pluie de type déterministe (ces 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 (Demargne 2015b) ; l'assimilation de données peut également permettre d'initialiser le modèle en prenant en compte les incertitudes des états du modèle, par exemple avec l'assimilation variationnelle de l'algorithme 4DVAR testée par Jay-Allemand (2016) ; la prise en compte des incertitudes de paramétrisation du modèle (testée dans l'étude Demargne 2014) pourrait également être poursuivie pour travailler avec différents jeux de paramètres équiprobables ; une approche utilisant plusieurs modèles hydrologiques avec des structures différentes (par exemple les modèles GRSD et GRDc) permettrait de prendre en compte les incertitudes de structure de la modélisation hydrologique et pourrait potentiellement mieux caractériser les incertitudes hydrologiques des prévisions de débit ; 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 Vigicrues-Flash pourra potentiellement conduire à une meilleure prise en compte des incertitudes.

## 7. 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.
- 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.
- Caseri 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. 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. 2014. 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. 2015a. Intégration des pluies prévues du système AROME-PI dans le système AIGA. Programme MEDDE (DGPR / SRNH) - IRSTEA 2015, *Rapport d'étude HYDRIS Hydrologie*, 38 p.
- Demargne J. 2015b. Assimilation des débits pour améliorer l'initialisation du modèle hydrologique du système AIGA. Programme MEDDE (DGPR / SRNH) - IRSTEA 2015, *Rapport d'étude HYDRIS Hydrologie*, 29 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., Makhlouf 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
- Javelle P., Organde D., Demargne J., Saint-Martin C., de Saint-Aubin C., Garandeau L., Janet B., 2016. Setting up a French national flash flood warning system for ungauged

- catchments based on the AIGA method. Conférence FLOODrisk 2016 - 3rd European Conference on Flood Risk Management, doi: 10.1051/e3sconf/20160718010.
- Jay-Allemand M., Gejadze I., Javelle P., Organde D., Fine J.-A., 2016. Implémentation de l'algorithme 4DVAR dans un modèle hydrologique conceptuel, continu et distribué (GRDc). Application à la prévision des crues soudaines. Convention DGPR/SCHAPI – IRSTEA 2016, *Rapport d'étude HYDRIS Hydrologie*, 53 p.
- Jolliffe I.T., Stephenson, D.B., 2003. Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley and Sons, Chichester, England, 2003, 254 p.
- Meteo-France, 2015 : Caractéristiques du modèle Arome, documentation, 6p. ; disponible sur le lien suivant  
[https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=131&id\\_rubrique=51](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=131&id_rubrique=51)
- 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., 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. Convention DGPR/SCHAPI – IRSTEA 2015, *Rapport d'étude HYDRIS Hydrologie*, 19 p.
- Organde D., 2016. Améliorations de la modélisation hydrologique de la méthode AIGA. Convention DGPR/SCHAPI – IRSTEA 2016, *Rapport d'étude HYDRIS Hydrologie*, 49 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.
- Pappenberger, F., et co-auteurs, 2016: Hydrological Ensemble Prediction Systems around the globe. *Handbook of Hydrometeorological Ensemble Forecasting*, éditions Springer, (<http://refworks.springer.com/mrw/index.php?id=5946>), 36 p.
- Pauthier B., Bois B., Castel T., Thevenin D., Château Smith C., et Richard Y., 2016. Mesoscale and local scale evaluations of Quantitative Precipitation Estimates by weather radar products during a heavy rainfall event. *Advances in Meteorology*, 2016, 9p. <http://dx.doi.org/10.1155/2016/6089319>
- 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.
- de Saint-Aubin C., Garandeau L., Janet B., et Javelle P. 2016. A new French flash flood warning service. Conférence FLOODrisk 2016 - 3rd European Conference on Flood Risk Management, doi: 10.1051/e3sconf/20160718024.
- Saint-Martin C., Fouchier C., Javelle P., Douvinet J., et Vinet F. 2016. Assessing the exposure to floods to estimate the risk of flood-related damage in French

Mediterranean basins. Conférence FLOODrisk 2016 - 3rd European Conference on Flood Risk Management, doi: 10.1051/e3sconf/20160704013.

Wells S., Robson A., Moore R., Cole S. and Rudd A, 2016. Understanding the national performance of flood forecasting models to guide incident management and investment. Conférence FLOODrisk 2016 - 3rd European Conference on Flood Risk Management, doi: 10.1051/e3sconf/20160718007.

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.

## 8. Liste des Figures et Tableaux

### Liste des Figures

Figure 1 : Principe de fonctionnement de la chaîne d’avertissement du système Vigicrues-Flash, se basant sur la méthode AIGA (les catégories de période de retour correspondent aux informations de la plateforme RHYTMME) .....	8
Figure 2 : Principales sources d’incertitude dans la chaîne d’avertissement du système Vigicrues-Flash .....	10
Figure 3 : Exemple de comparaison de la grille de pluie observée issue des radars et des grilles de prévisions AROME-PI successives pour les horizons de +1h, +2h et +4h (source : Joël Hoffman) .....	11
Figure 4 : Approche méthodologique pour évaluer l’apport des prévisions AROME-PI pour le système d’avertissement Vigicrues-Flash en utilisant la modélisation hydrologique semi-distribuée du système actuel et la modélisation hydrologique distribuée continue de l’étude Organde (2016) .....	12
Figure 5 : Cartes des pluies observées journalières maximales au pixel pour chacun des 8 événements (avec la valeur de pluie maximale observée au pixel), basées sur les pluies Antilope KED et la localisation des exutoires des 781 bassins versants étudiés.....	13
Figure 6 : Carte des exutoires des 781 bassins versants avec la catégorie maximale de période de retour des dépassements observés sur les 8 événements de l’étude (pour les seuils de quantiles observés de 2, 5 et 10 ans).....	14
Figure 7 : Répartition par événement du nombre de dépassements observés (Qobs) et simulés avec la modélisation hydrologique semi-distribuée (Qsd_sim) pour le seuil relatif de période de retour de 2 ans et les 781 bassins versants .....	15
Figure 8 : Schéma du couplage des modèles journaliers continus SAJ et GR4J et du modèle horaire semi-distribué GRSDévé pour le système actuel Vigicrues-Flash.....	16
Figure 9 : Schéma du modèle distribué continu GRDc .....	17
Figure 10 : Schéma de construction des ensembles timelag de pluie à partir des prévisions déterministes AROME-PI successives .....	20

Figure 11 : Exemples d'hydrogrammes obtenus par les modèles GRSD et GRDc avec les différentes prévisions de pluie en comparaison avec l'hydrogramme de débit observé.....	22
Figure 12 : Principe de fonctionnement de la contingence globale (sur l'ensemble d'un événement) avec une valeur maximale d'anticipation pour associer un dépassement simulé avec un dépassement observé .....	25
Figure 13 : Principe de fonctionnement de la contingence avec une tolérance de 10% sur les seuils pour comptabiliser les dépassements manqués de peu (« Near Miss ») et les quasi fausses alertes (« Close False Alarm ») .....	26
Figure 14 : Principe de fonctionnement de la contingence horaire, pour chaque pas de temps de l'événement, avec une valeur maximale d'anticipation pour associer un dépassement prévu avec un dépassement observé .....	27
Figure 15 : Histogrammes des dépassements observés, simulés et prévus avec une pluie future nulle avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils relatifs de quantiles de période de retour de 2, 5 et 10 ans, en incluant, à gauche, tous les dépassements prévus en avance et en retard et, à droite, uniquement les dépassements prévus avec une anticipation positive .....	28
Figure 16 : Comparaison des cartes de dépassements observés et des dépassements simulés issus du modèle semi-distribué avec la catégorie maximale de période de retour sur les 8 événements de l'étude.....	29
Figure 17 : Distributions du degré d'anticipation des dépassements simulés avec le modèle semi-distribué GRSD pour le seuil de débit biennal, sans inclure d'anticipation maximale de 10h (à gauche) et en intégrant une anticipation maximale de 10h (à droite).....	30
Figure 18 : Distributions du degré d'anticipation des 147 dépassements simulés ( $Q_{GRSDsim}$ ) et prévus issus de la pluie future nulle ( $Q_{GRSD[pnul]}$ ), la pluie observée ( $Q_{GRSD[pobs]}$ ) et la prévision déterministe AROME-PI ( $Q_{GRSD[pdet]}$ ) avec le modèle semi-distribué pour le seuil de période de retour de 2 ans .....	32
Figure 19 : Distributions du degré d'anticipation des 140 dépassements simulés ( $Q_{GRDcsim}$ ) et prévus issus de la pluie future nulle ( $Q_{GRDc[pnul]}$ ), la pluie observée ( $Q_{GRDc[pobs]}$ ) et la prévision déterministe AROME-PI ( $Q_{GRDc[pdet]}$ ) avec le modèle distribué continu pour le seuil de période de retour de 2 ans.....	32
Figure 20 : Distributions du degré d'anticipation des 147 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle semi-distribué pour le seuil de période de retour de 2 ans.....	35
Figure 21 : Distributions du degré d'anticipation des 140 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle distribué continu pour le seuil de période de retour de 2 ans.....	35
Figure 22 : Distributions du gain en anticipation des 147 dépassements observés et prévus avec le modèle semi-distribué vis-à-vis des dépassements simulés ( $Q_{GRSDsim}$ ) avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) pour le seuil de période de retour de 2 ans .....	37
Figure 23 : Distributions du gain en anticipation des 140 dépassements observés et prévus avec le modèle distribué continu vis-à-vis des dépassements prévus avec une pluie future nulle ( $Q_{GRDc[pnul]}$ ) avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) pour le seuil de période de retour de 2 ans.....	37
Figure 24 : Distributions du degré d'anticipation des 50 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle semi-distribué pour le seuil de période de retour de 5 ans.....	38

Figure 25 : Distributions du degré d'anticipation des 48 dépassements observés, simulés et prévus avec les prévisions d'ensembles (à gauche) et les prévisions déterministes (à droite) avec le modèle distribué continu pour le seuil de période de retour de 5 ans.....	38
Figure 26 : Scores de contingence CSI (échelle sur la gauche) et de Biais en Fréquence (FB, échelle sur la droite) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans.....	41
Figure 27 : Comparaison des scores de CSI avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de période de retour de 2 et 5 ans.....	42
Figure 28 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans.....	43
Figure 29 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 5 ans.....	45
Figure 30 : Comparaison des scores de Probabilité de Détection (POD) et Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de 2 et 5 ans.....	46
Figure 31 : Comparaison des scores de Probabilité de Détection (POD) et Taux de Succès (SR) basés sur une tolérance de 10% sur les seuils de 2 et 5 ans avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc).....	46
Figure 32 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) basés sur une tolérance de 10% sur les seuils de période de retour de 5 ans avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc).....	47
Figure 33 : Scores de Probabilité de Détection (POD) et Taux de Succès (SR) de la contingence horaire avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour le seuil de période de retour de 2 ans.....	49
Figure 34 : Comparaison des scores obtenus en contingence horaire pour la Probabilité de Détection (POD) et le Taux de Succès (SR) avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils de 2 et 5 ans.....	50

### **Liste des Tableaux**

Tableau 1 : Nombre de dépassements observés, simulés et prévus avec une pluie future nulle avec le modèle semi-distribué (GRSD) et le modèle distribué continu (GRDc) pour les seuils relatifs de quantiles de période de retour de 2, 5 et 10 ans.....	28
Tableau 2 : Statistiques de la distribution du degré d'anticipation des 147 dépassements simulés ( $Q_{GRSDsim}$ ) et prévus avec la pluie future nulle ( $Q_{GRSD[pnul]}$ ), la pluie observée ( $Q_{GRSD[pobs]}$ ) et la prévision déterministe ( $Q_{GRSD[pdet]}$ ) issu de la pluie déterministe AROME-PI avec le modèle semi-distribué pour le seuil de période de retour de 2 ans.....	33
Tableau 3 : Statistiques de la distribution du degré d'anticipation des 140 dépassements simulés ( $Q_{GRDcsim}$ ) et prévus avec la pluie future nulle ( $Q_{GRDc[pnul]}$ ), la pluie observée ( $Q_{GRDc[pobs]}$ ) et la prévision déterministe ( $Q_{GRDc[pdet]}$ ) issu de la pluie déterministe AROME-PI avec le modèle distribué continu.....	33