# Seasonal streamflow forecasting for reservoir management

Louise Crochemore

# Doctorat AgroParisTech

École doctorale *Géosciences, Ressources Naturelles et Environnement*

## Thèse

pour obtenir le grade de docteur délivré par

## l'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)

**Spécialité : Hydrologie**

présentée et soutenue publiquement par

**Louise Crochemore**

le 29 Avril 2016

# Seasonal streamflow forecasting for reservoir management

Co-directeur de thèse : **Vazken Andréassian**

Co-directrice de thèse : **Maria-Helena Ramos**

**Jury**

| | | |
|---|---|---|
| **Mme Sandrine ANQUETIN** | CNRS-LTHE, Grenoble, France | Rapporteure |
| **Mme Marie-Amélie BOUCHER** | UQAC, Chicoutimi, QC, Canada | Rapporteure |
| **Mme Flavie CERNESSON** | TETIS-AgroParisTech, Montpellier, France | Examinatrice |
| **Mme Florence HABETS** | METIS-UPMC, Paris, France | Examinatrice |
| **M. Micha WERNER** | UNESCO-IHE, Delft, Pays-Bas | Examinateur |
| **M. Aldo PENASSO** | IAV, La Roche Bernard, France | Invité |

**AgroParisTech**

**Irstea, Unité de recherche Hydrosystèmes et Bioprocédés**

1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France

# Remerciements

Ce manuscrit et ces trois années de thèse n'auraient pas été les mêmes sans un certain nombre de personnes que je souhaite remercier ici.

Avant tout, je souhaite remercier Maria-Helena Ramos qui a supervisé ces travaux de thèse et qui m'a beaucoup conseillée tout au long du chemin. Merci de m'avoir formée au monde de la recherche comme tu l'as fait, merci de m'avoir poussée à dépasser mes limites, et merci de m'avoir ouvert autant de portes. Je souhaite aussi remercier Vazken Andréassian, en tant que co-directeur de thèse, et Charles Perrin, en tant qu'instigateur de la collaboration avec l'IAV. Merci à tous les deux de m'avoir fait découvrir l'hydrologie, de m'avoir donné mes premières chances dans l'équipe et merci pour tout ce que vous m'avez appris tout au long de ces belles années.

Deux personnes extérieures à Irstea ont aussi été très importantes pour cette thèse. Je pense d'abord à Aldo Penasso de l'IAV qui a pris le temps de partager son expérience du barrage d'Arzal et qui m'a reçue à plusieurs reprises au sein de l'IAV. J'en profite pour remercier les membres de l'IAV que j'ai rencontrés à ces occasions et avec qui j'ai pu discuter, notamment Jean-Pierre Arrondeau. Je pense ensuite à Florian Pappenberger qui a joué un rôle important dans l'orientation scientifique de la thèse et qui a toujours été encourageant. Florian nous a notamment permis d'utiliser les prévisions saisonnières du CEPMMT et m'a donné l'opportunité de passer deux semaines au sein du CEPMMT. Un grand merci à l'équipe à ECMWF pour son accueil à chaque fois chaleureux, notamment Fredrik, Louise, Ervin, Rebecca, Francesca et Callum.

Je souhaite remercier les membres du jury d'avoir accepté d'évaluer ce travail de thèse, ainsi que pour les questions et discussions le jour de la soutenance. Je remercie également les membres du comité de pilotage: Éric Sauquet, David Dorchies et Matthieu Le Lay pour leurs conseils sur l'orientation de la thèse.

Les séjours à l'IAV n'auraient pas été les mêmes sans nos partenaires du WVER dans le cadre du projet DROP: Herbert, Christoff, Antje et Evelyn, merci pour votre accueil dans l'Eifel-Rur et ses "inspection galleries". Merci aussi aux autres partenaires du projet DROP.

À Irstea, j'aimerais remercier les équipes administrative et technique qui nous aident au quotidien, en particulier Nathalie et Roger, sans oublier Sylvain, Laurence et Elisabeth. Je souhaite remercier l'équipe Hydro d'Irstea Antony dans son ensemble. Je n'aurais pas pu imaginer un meilleur cadre pour faire cette thèse. C'est grâce à chacun d'entre vous, que vous y soyez actuellement, que vous en soyez partis ou que vous n'y soyez restés que le temps d'un été. Un grand merci à Carina, Guillaume, Pierre, Olivier et Alban pour leurs conseils et encouragements, pour les trajets dans ce merveilleux RER C, les soirées brassantes, et la mousse au chocolat (chacun se reconnaîtra). Laure, fantastique co-bureau, dans le calme comme dans la tempête, merci pour Tout. Julie, formidable colocataire et collègue, ton entrain et ton interception de balle de tennis sont sans pareil. Andrea et Carine, ces temps sont difficiles pour vous, merci à tous les deux d'avoir été présents dans cette période si étrange qu'est la rédaction. Sur cette note, j'en profite pour souhaiter le meilleur aux suivants: Angélica, Philippe, Cédric, Léonard, Sylvia, Morgane et Manon. Merci aux "anciens", comme le dirait si bien Andrea, qui ont montré la voie: Ioanna, Bahar, Laurent, François, Florent, Damien, Pierre B., Claire, Marine, Mathilde, Annie, Pierre-Yves et Raji et Dimitri. Merci aussi à l'ensemble de l'hydrologie sociale et à l'équipe de Mamane. Et merci à ceux sans qui le bâtiment Lavoisier et les pauses sportives n'auraient pas été aussi conviviales: Violaine, Nastassia, Laetitia, Carolina, Mathilde, Marine, Hajer Douidoui, Cécile, Olivier C., Sylvain, Simon S., Thomas, Simon P. , sans oublier Captain Valou.

Enfin, à ceux qui ont suivi du début jusqu'à la fin mon parcours sinueux et tumultueux au long de ces trois années, à ma famille, à Marilyn, mille mercis.

# Résumé

Les prévisions saisonnières de débits peuvent favoriser la gestion des risques dans de nombreux secteurs, tels que l'approvisionnement en eau potable, la production hydroélectrique ou la gestion de réservoirs multi-usage. Leur usage pour la prise de décision en contexte de risque nécessite de quantifier les incertitudes à longue échéance. Celles-ci peuvent par exemple être communiquées à l'utilisateur via des outils d'évaluation des risques. Cependant, l'implémentation des prévisions saisonnières rencontre encore des obstacles comme la qualité actuelle des prévisions, ou la difficulté de traduire l'information saisonnière pour les besoins opérationnels.

L'objectif de cette thèse est de faire progresser les connaissances sur la prévision saisonnière pour la gestion de réservoirs multi-usage. Dans un premier volet, les travaux ont évalué la qualité des prévisions saisonnières de pluies et de débits dans seize bassins français. De nouvelles méthodes de prévision des débits ont été proposées et testées dans ces bassins, notamment pour la prévision des étiages. Un deuxième volet est consacré aux prévisions saisonnières de débits pour la gestion de réservoirs. Un outil d'évaluation des risques de pénurie d'eau dans le réservoir d'Arzal, en Bretagne, a été développé et le rôle des prévisions saisonnières dans la gestion de réservoirs en contexte de risque a été analysé.

Dans le premier volet, nous avons montré que la correction des biais mensuels des prévisions saisonnières du CEPMMT permet d'améliorer la fiabilité des prévisions de débits, et d'harmoniser les performances obtenues dans les seize bassins. Plusieurs méthodes de prévision des débits, basées sur les prévisions saisonnières du CEPMMT et sur les données historiques de précipitation et de débit, ont ensuite été comparées. Un conditionnement des données historiques à partir des prévisions de précipitations a permis de combiner la fiabilité des données historiques et la finesse des prévisions météorologiques. Ces méthodes permettent de prévoir les étiages de manière fiable et peuvent aider à prévoir des évènements extrêmes de sécheresse.

Le deuxième volet a permis le développement d'un outil d'évaluation des risques en contexte de basses eaux pour le cas du réservoir d'Arzal. Le risque de ne pouvoir garantir la ressource en eau pour tous les usages et de devoir changer de stratégie de gestion pendant la saison estivale est quantifié à partir de prévisions saisonnières d'apports et d'un modèle de bilan du réservoir. Enfin, un jeu de rôle a été développé pour mieux comprendre comment l'information saisonnière peut influer la prise de décision en contexte de gestion de réservoirs. Ce jeu a mis en évidence l'enjeu mais aussi la difficulté d'incorporer l'information à longue échéance dans le processus de décision.

# Abstract

Seasonal forecasts can enhance risk assessment in a variety of applications ranging from multi-purpose reservoir management, drinking water supply and preparedness to droughts. Risk assessment tools, for instance, can benefit from seasonal probabilistic forecasting to support risk-based decision-making. However, the implementation of seasonal forecasts still faces impediments, including the quality of seasonal forecasts and the difficulty to tailor seasonal products to end-users' needs.

This thesis investigates seasonal streamflow forecasting for multi-purpose reservoir management. We first assess the quality of seasonal precipitation and streamflow forecasts in sixteen French catchments. Streamflow forecasting systems are proposed and tested in these catchments, and their potential is illustrated in low-flow and drought risk forecasting. Secondly, seasonal streamflow forecasts are applied in reservoir management. A risk assessment tool is developed to forecast risks of water shortages in the Arzal reservoir, in Brittany, France, and the role of seasonal forecasts for risk-based decision-making is assessed.

First, we showed that a bias correction of monthly biases in seasonal precipitation forecasts could increase the reliability of streamflow forecasts and harmonize their performances in the sixteen catchments. We then compared several streamflow forecasting systems, based either on ECMWF seasonal forecasts or on historical streamflows and precipitations. A conditioning of historical data based on precipitation forecasts allowed to take advantage of the reliability of historical data and of the sharpness of meteorological forecasts. The proposed methods provided reliable low-flow forecasts and showed a good ability to forecast drought events.

Secondly, a low-flow risk assessment tool was developed for the case of the Arzal reservoir. The seasonal streamflow forecasts used as input to a water balance model of the reservoir allowed us to quantify the risks of water shortages in summer. Lastly, a role-playing game was developed to better understand the role of long-term probabilistic information for decision-making in reservoir management.

# Résumé substantiel

## Introduction

Les prévisions saisonnières de débits peuvent favoriser la gestion des risques dans de nombreux secteurs, tels que l'approvisionnement en eau potable, la production hydroélectrique ou la gestion de réservoirs multi-usage. Leur usage pour la prise de décision en contexte de risque nécessite de quantifier les incertitudes à longue échéance. Celles-ci peuvent par exemple être communiquées à l'utilisateur via des outils d'évaluation des risques. Cependant, l'implémentation des prévisions saisonnières rencontre encore des obstacles comme la qualité actuelle des prévisions, ou la difficulté de traduire l'information saisonnière pour les besoins opérationnels.

La prévisibilité peut être définie comme la limite de notre capacité à prévoir un état futur à partir de la connaissance des conditions actuelles (Lorenz, 1984; Kirtman *et al.*, 2013). La prévisibilité dépend du processus naturel et de la localisation géographique, mais aussi de notre capacité à représenter le système physique (Blöschl and Zehe, 2005). Alors que les phénomènes météorologiques sont difficilement prévisibles au-delà de deux semaines à cause de la nature chaotique de l'atmosphère, les processus climatiques à plus grande échelle peuvent aider à prévoir jusqu'à des horizons de plusieurs années. Les informations provenant de ces systèmes peuvent ensuite améliorer les performances des prévisions hydrologiques, d'autant que l'inertie du bassin versant filtre les erreurs à hautes fréquences.

Les prévisions hydrologiques peuvent aider la prise de décision, notamment pour la gestion de réservoirs d'eau. La gestion de barrage est au cœur des stratégies d'optimisation de la ressource en eau. Le cas particulièrement complexe des barrages multi-usage requiert à la fois une bonne gestion et des outils sophistiqués d'aide à la décision (Castelletti *et al.*, 2008). Le processus de décision et l'évaluation du risque peuvent notamment bénéficier de prévisions d'apports aux réservoirs. En contexte de basses eaux ou sécheresse, lorsque la gestion vise à préserver la ressource en eau et garantir ses différents usages, des prévisions à longue échéance peuvent aider à optimiser la gestion et prendre des mesures préventives (Georgakakos and Graham, 2008; Regonda *et al.*, 2011).

La thèse s'est déroulée dans le cadre du projet DROP Interreg IVB qui visait à aider les régions du Nord-Ouest de l'Europe à mieux anticiper les périodes de sécheresses et de manque d'eau.Ce travail de thèse a été motivé par les besoins des gestionnaires du barrage d'Arzal, un des sites pilotes du projet, en outils de prévision des étiages pour venir en appui à la gestion. Un objectif de la thèse est d'étudier la qualité des prévisions saisonnières et leur apport en contexte de basses eaux et sécheresses pour la gestion de réservoir. La recherche aborde quatre questions: (1) Quel est l'impact de corriger le biais des prévisions de précipitations issues de modèles climatiques sur les prévisions de débits ? (2) Comment utiliser les sorties de modèles climatiques pour améliorer les méthodes traditionnelles de prévision des débits saisonniers basées sur les données historiques ? (3) Quel système de gestion des étiages pour le réservoir d'Arzal ? (4) Comment les décisionnaires exploitent l'information à long terme pour la gestion séquentielle de réservoir multi-usage ?

Après une présentation des bassins versants, des données hydrométéorologiques, du modèle hydrologique et de son évaluation, nous évaluons la qualité des prévisions saisonnières de précipitations du CEPMMT (Centre Européen pour les Prévisions Météorologiques à Moyen Terme) et de débits associées. Huit méthodes de correction des biais sont comparées et huit conditionnements des données historiques à partir d'indices de précipitations sont évalués. Ensuite, un outil d'évaluation des risques de pénurie d'eau est développé pour le réservoir d'Arzal, et le rôle des prévisions saisonnières dans la prise de décision pour la gestion de réservoirs est analysé. Enfin, les conclusions de la thèse sont présentées.

## 1. Bassins versants et jeux de données

Dans la thèse, un premier jeu de seize bassins versants a été utilisé pour évaluer les performances des prévisions saisonnières de précipitations pour la prévision des débits en France. Ces bassins présentent une variété de conditions hydrologiques et climatiques et ne sont que peu influencés par les activités humaines et par la neige. Ces bassins disposent de longues séries de débits observés (36 à 52 années) provenant de la Banque HYDRO. Les aire de ces bassins varient de 377 km$^2$ à 4320 km$^2$. Fig. R1 présente la localisation de ces bassins.

Le bassin versant de la Vilaine, situé en Bretagne, a ensuite permis de se concentrer sur les apports au réservoir d'Arzal situé en aval (cf. Fig. R1). Le bassin draine une surface de 10 000 km$^2$, majoritairement occupée par des terres agricoles. À son embouchure, la Vilaine se jette dans le réservoir d'Arzal puis dans

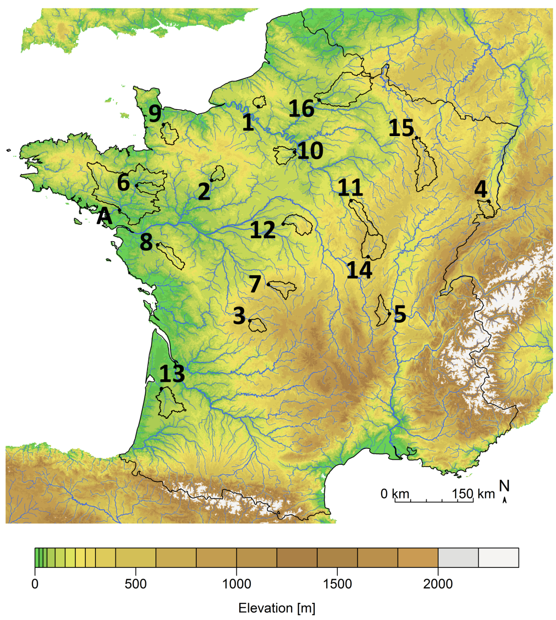



Figure R1: Localisation des 17 bassins versants étudiés. Les seize bassins versants sont numérotés du plus petit bassin au plus étendu. Le bassin de la Vilaine est indiqué par la lettre A.

l'Océan Atlantique. Le barrage d'Arzal est géré par l'IAV (Institution d'Aménagement de la Vilaine). Initialement construit pour arrêter les marées qui contribuaient aux inondations du bassin aval de la Vilaine, le barrage crée un réservoir d'eau douce permettant d'approvisionner la région en eau potable. Les apports au réservoir sont estimés à partir des débits observés à la station de Rieux, au Pont de Cran. Le débit moyen journalier de la Vilaine est de 60 $m^3/s$.

Pour l'ensemble de ces bassins, les données journalières de précipitations et températures proviennent de la réanalyse SAFRAN de Météo-France (Quintana-Seguí *et al.*, 2008; Vidal *et al.*, 2010). Ces données sont disponibles depuis 1958 à une résolution de 8x8 km. L'évapotranspiration potentielle est calculée à partir des températures et de la formule d'Oudin (Oudin *et al.*, 2005). Des mesures locales de précipitations pour le réservoir d'Arzal ont aussi été collectées auprès des gestionnaires du barrage.

Les prévisions saisonnières de précipitations journalières du Système 4 du CEPMMT ont été utilisées tout au long de la thèse. Ces prévisions émises le 1$^{er}$ de chaque mois comprennent 15 ou 51 membres et couvrent la période de 1981 à 2010 avec une résolution de 0.7° (Molteni *et al.*, 2011). Dans le cadre de la thèse, les prévisions sont agrégées à l'échelle du bassin versant et seuls les 90 premiers jours de la prévision sont utilisés.

## 2. Modélisation et prévision hydrologique

Le modèle hydrologique utilisé dans la thèse est le modèle conceptuel journalier GR6J développé à Irstea pour les étiages (Pushpalatha *et al.*, 2011). Les entrées du modèle sont les précipitations et l'évapotranspiration potentielle sur la surface du bassin versant, et sa sortie est le débit à l'exutoire. Le modèle est composé d'un réservoir de production, de deux hydrogrammes unitaires et de deux réservoirs de routage (Fig. R2).

Les six paramètres du modèle sont calés sur l'ensemble de la chronique d'observations disponibles, à laquelle on soustrait une année utilisée pour valider le jeu de paramètres. Cette procédure est répétée pour chaque année de la chronique. Trente jeux de paramètres sont obtenus pour la période de 1981 à 2010 dans chancun des seize bassins. Dix sont obtenus pour la période de 2003 à 2012 dans le bassin de la Vilaine.

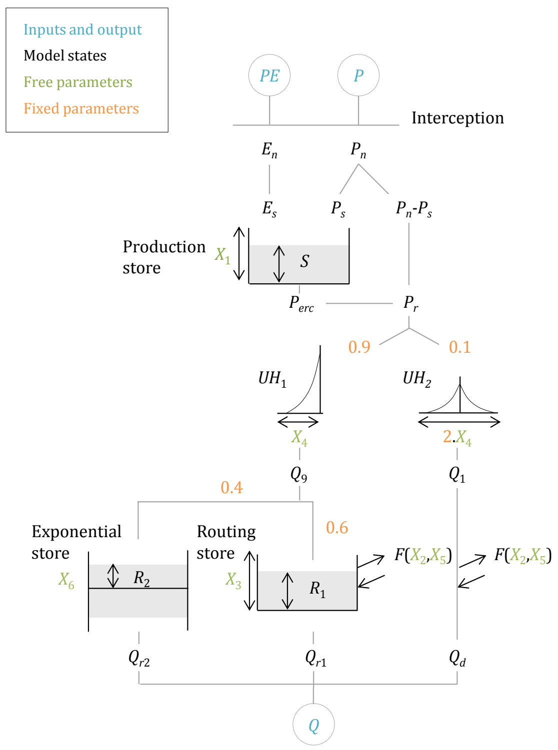



Figure R2: Structure du modèle GR6J (modifié de Pushpalatha *et al.*, 2011).

Le modèle est calé à l'aide: (1) du KGE (*Kling-Gupta efficiency*; Gupta *et al.*, 2009) appliqué à la racine des débits ($KGE_{rQ}$) pour mettre l'accent sur les débits médians et forts, et (2) du KGE appliqué à l'inverse des débits ($KGE_{iQ}$) pour mettre l'accent sur les bas débits. La valeur optimale de ces critères est 1. Les performances en calage sont globalement satisfaisantes avec des valeurs de $KGE_{rQ}$ entre 0.88 et 0.97, et des valeurs de $KGE_{iQ}$ entre 0.46 et 0.94.

Le modèle est ensuite validé à l'aide des mêmes critères sur les basses eaux et les débits médians à forts. Les performances du modèle calé avec $KGE_{rQ}$ sont excellentes sur les débits médians. En revanche, de bonnes performances ne sont pas assurées dans tous les bassins sur les bas débits. Un calage sur les bas débits permet d'atteindre des performances satisfaisantes entre 0.41 et 0.94 en basses eaux dans tous les bassins. De plus, le calage sur les bas débits assure des performances supérieures à 0.54 sur les débits médians à forts. Le biais des simulations en terme de volumes mensuels est entre -0.02 et 0.04 avec un calage sur les débits médians, et entre -0.14 et 0.12 à l'exception du bassin 8 où un biais de -0.94 est obtenu avec un calage sur les bas débits. Dans le bassin de la Vilaine, le modèle calé sur les bas débits atteint en validation un $KGE_{rQ}$ de 0.90, un $KGE_{iQ}$ de 0.76 et un biais de -0.04. Fig. R3 corrobore ces valeurs satisfaisantes avec les récessions simulées de Mai à Octobre de 2005 à 2012.

En prévision, les états du modèle sont d'abord initialisés sur l'année précédant la date de prévision. Les états des réservoirs de routage sont ensuite ajustés à partir du dernier débit observé. Enfin le modèle est forcé avec des scénarios de précipitations prévues et d'évapotranspiration. Ici, le scénario d'évapotranspiration est systématiquement l'évapotranspiration potentielle moyenne interannuelle. Les scénarios de débits ainsi obtenus forment une prévision d'ensemble offrant une représentation des incertitudes.

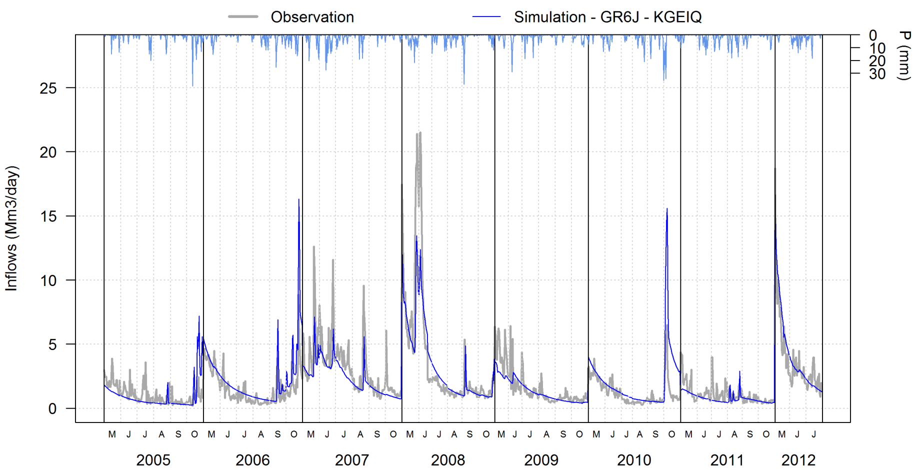



Figure R3: Débits observés et simulés avec GR6J et précipitations observées dans le bassin de la Vilaine à Rieux, de Mai à Octobre et de 2005 à 2012.

## 3. Évaluation des prévisions

Dans cette thèse, les performances des systèmes de prévision sont évaluées à l'aide de plusieurs critères numériques testant la précision, la fiabilité, la finesse, les performances globales et la capacité à détecter des évènements binaires des systèmes.

La précision d'une prévision correspond à sa distance à l'observation. Pour cela nous avons utilisé la prévision moyenne, et l'erreur moyenne absolue (MAE) ou l'erreur quadratique moyenne (RMSE).

La fiabilité représente la cohérence statistique entre les fréquences d'observations et les probabilités prévues. Elle est représentée à l'aide du diagramme de PIT (*Probability Integral Transform*; Fig. R4). Une prévision parfaitement fiable a un diagramme confondu avec la diagonale 1:1. Le diagramme de PIT peut donc être synthétisé par l'aire entre la diagonale 1:1 et la courbe que l'on cherche à minimiser (Renard *et al.*, 2010).

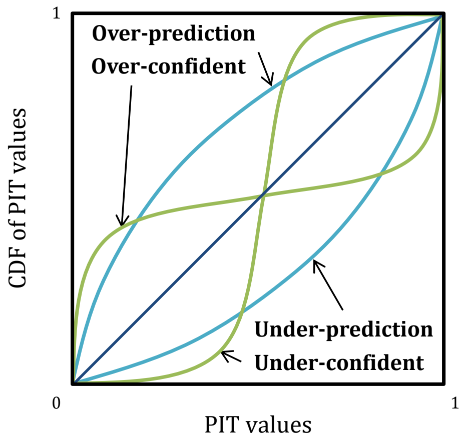



Figure R4: Interprétation du diagramme de PIT (Laio and Tamea, 2007).

La finesse correspond à la concentration de la distribution prédictive d'une prévision. Elle est évaluée par l'écart interquantile à 90 %. Suivant Gneiting *et al.* (2007), la finesse n'est considérée comme critère de performance qu'une fois la fiabilité assurée, i.e. entre deux systèmes fiables, le plus fin est préféré.

L'erreur globale des prévisions est calculée à l'aide du CRPS qui évalue l'erreur de la distribution prédictive (*Continuous Rank Probability Score*; Hersbach, 2000). Enfin, le diagramme de ROC (*Relative Operating Characteristics*; Mason and Graham, 1999) évalue la capacité d'un système à détecter un évènement défini par un seuil. L'aire sous la courbe synthétise le diagramme en une valeur numérique que l'on cherche à maximiser.

Les performances des systèmes de prévision évaluées par le MAE, l'aire du PIT, l'IQR et le CRPS sont ensuite comparées à des systèmes de référence dont les performances sont considérées "standard". Un système de référence commun pour les précipitations est la climatologie des précipitations. Dans le cas des débits, l'ESP, i.e. la climatologie des pluies utilisée en entrée du modèle hydrologique, ou la climatologie des débits sont souvent utilisés. Enfin, on peut aussi utiliser une version antérieure, avant modification, du système de prévision pour mesurer les apports d'une méthode. Dans ce dernier cas, on regarde l'horizon de prévision maximal pour lequel les performances du système sont améliorées (*Useful Forecasting Lead time*; Nicolle *et al.*, 2014).

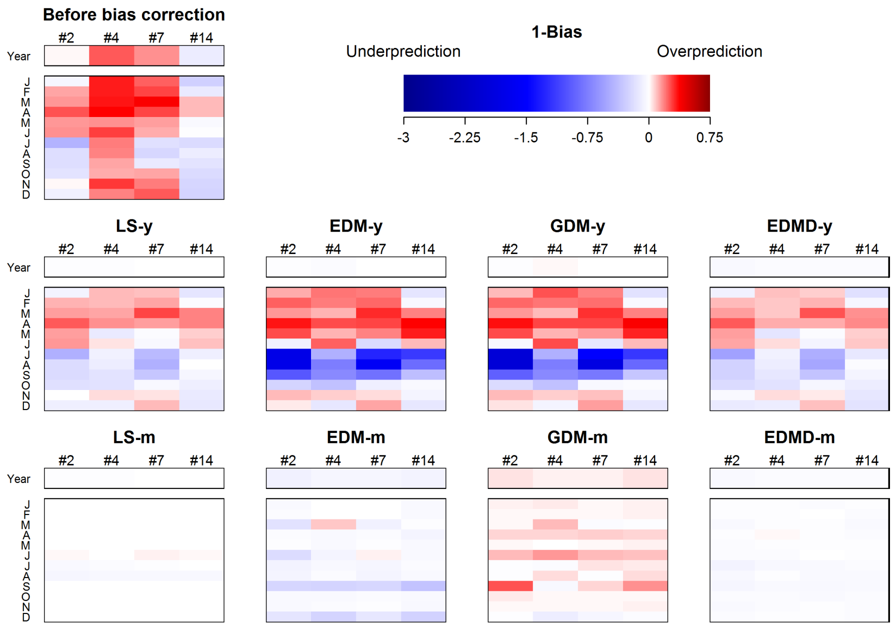



Figure R5: Biais des précipiations dans les bassins 2, 4, 7 et 14, sur la période 1981-2010, sur l'année entière et pour chaque mois de l'année, et pour les horizons 31 à 60 jours. Le graphe en haut à gauche représente le biais avant correction. Les autres représentent les biais restants après correction avec chacune des huit méthodes.

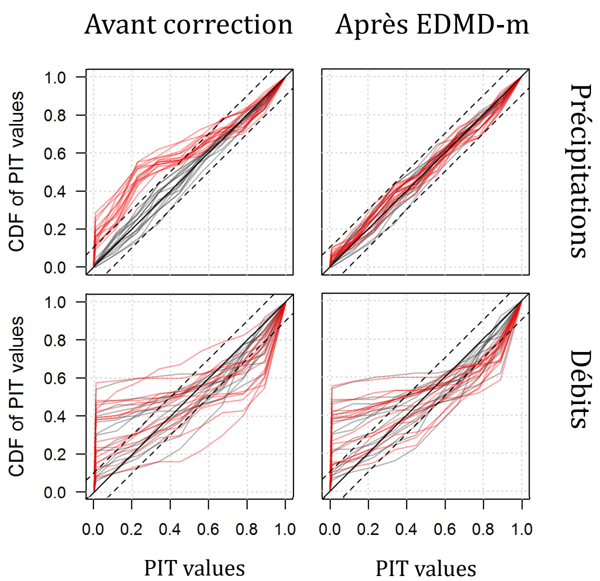



Figure R6: Diagrammes de PIT des précipitations et débits pour la saison Juin-Juillet-Août et un horizon de 30 jours. Chaque ligne correspond à un bassin. Les lignes rouges correspondent aux prévisions du Système 4, les lignes grises correspondent à la climatologie des pluies ou à l'ESP.

## 4. Correction du biais des prévisions de précipitations pour améliorer les performances des prévisions saisonnières de débits

Dans ce chapitre, nous estimons la qualité de prévisions saisonnières de précipitations et de débits en France et explorons l'impact de la correction du biais des prévisions de précipitations sur la qualité des prévisions de débits.

Dans un premier temps, les prévisions saisonnières de précipitations brutes (i.e., sans correction du biais) du CEPMMT sont évaluées pour prévoir les débits des seize bassins versants français de 1981 à 2010. Dans un second temps, le biais de ces prévisions de précipitations est évalué et corrigé à l'aide de huit méthodes de correction du biais présentées dans le Tableau R1. Les prévisions de précipitations et de débits obtenues à l'aide des huit méthodes sont, à leur tour, évaluées. La qualité des prévisions est caractérisée en terme de fiabilité, finesse, précision et performance globale. L'ensemble de référence utilisé pour l'évaluation des précipitations est basé sur la climatologie des pluies dans chaque bassin. L'ensemble de référence pour évaluer les débits est l'ESP qui utilise la climatologie des précipitations en entrée d'un modèle hydrologique.

Table R1: Abbréviation, calibration et description des huit corrections du biais.

| Abbré-viation | Calibré pour | Description |
|---|---|---|
| LS-y | l'année | Correction linéaire |
| LS-m | chaque mois | des valeurs mensuelles |
| EDM-y | l'année | Quantile-quantile de la distribution |
| EDM-m | chaque mois | empirique des données mensuelles |
| GDM-y | l'année | Quantile-quantile de la distribution |
| GDM-m | chaque mois | gamma des données mensuelles |
| EDMD-y | l'année | Quantile-quantile de la distribution |
| EDMD-m | chaque mois | empirique des données journalières |

Dans l'ensemble des bassins versants, les prévisions obtenues à partir des précipitations brutes du CEPMMT ont tendance à être plus fines mais moins fiables que celles obtenues avec la méthode ESP. La qualité des prévisions dépend fortement de la saison et du bassin considéré. Les biais observés en précipitation varient aussi de bassin en bassin, avec des biais mensuels importants qui ont tendance à se compenser sur l'année (Fig. R5). Par conséquent, seules les méthodes de correction du biais calées séparément pour chaque mois de l'année parviennent à corriger simultanément les biais annuels et mensuels. Plus particulièrement, la correction linéaire simple et la correction quantile-quantile des précipitations journalières minimisent le biais et apportent les meilleurs gains en performance.

La correction quantile-quantile des valeurs journalières est nettement supérieure pour améliorer la fiabilité des prévisions de précipitations et de débits. De plus, cette correction harmonise les performances entre bassins et saisons. Enfin, les prévisions corrigées deviennent aussi fiables que les prévisions de la méthode ESP tout en restant plus fines. Fig. R6 illustre l'impact de la correction quantile-quantile des valeurs journalières sur la fiabilité des précipitations et débits de la saison Juin-Juillet-Août. Les prévisions de précipitations corrigées pour la saison sont fiables. Les prévisions de débits obtenues après correction du biais ont gagné en fiabilité mais restent peu fiables pour la saison (une meilleure fiabilité est obtenue pour les autres saisons). Ceci suggère qu'un post-traitement des prévisions de débits est nécessaire pour atteindre des prévisions de débits fiables en été.

Enfin, nous avons regardé l'impact relatif de la correction du biais sur les précipitations et les débits. Les gains sont moyennés sur l'ensemble des horizons de prévisions, de 10 à 90 jours. L'idée est d'examiner comment une amélioration des prévisions de précipitations impacte les performances des prévisions de débits. Le gain en fiabilité est globalement supérieur en précipitations qu'en débits (Fig. R7). Enfin, un gain en performance globale des précipitations, même faible, permet d'obtenir un gain plus important en débit. Ceci met en évidence l'importance d'améliorer les performances des forçages météorologiques afin d'améliorer les performances des prévisions saisonnières de débits.

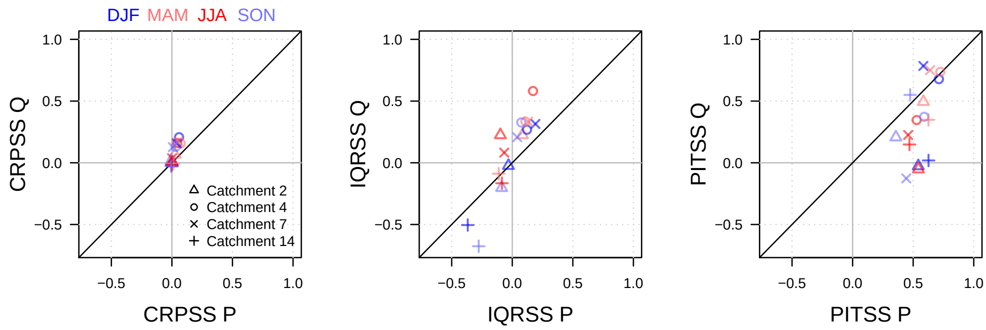



Figure R7: Gains en performance obtenus en débits grâce à la correction du biais versus gains en performance obtenus en précipitations, dans les bassins 2, 4, 7 et 14, pour chaque critère et saison.

## 5. Prévision saisonnière des débits en conditionnant les données historiques avec des indices saisonniers de précipitations

Dans le chapitre précédent, nous avons montré que la correction du biais des prévisions saisonnières de précipitations du CEPMMT peut améliorer la qualité des prévisions de débits, produites à l'aide d'un modèle hydrologique. Les prévisions de débits obtenues à partir des précipitations corrigées sont fines, mais restent peu fiables, en été par exemple. Dans ce chapitre, nous considérons d'autres méthodes dynamiques et statistiques de prévision des débits à l'échéance saisonnière exploitant les données historiques de précipitations ou de débits. Ces méthodes ont l'avantage d'être moins coûteuses en ressources informatiques et de produire des ensembles fiables, mais ne bénéficient pas des informations spécifiques au jour de la prévision. L'objectif de ce chapitre est de conditionner ces méthodes de prévision des débits basées sur les données historiques à partir des prévisions corrigées dans le chapitre précédent, afin d'améliorer la finesse des prévisions basées sur la climatologie.

Pour cela, des statistiques à long-terme des prévisions de précipitations du CEPMMT sont utilisées pour sélectionner des années parmi les précipitations et débits disponibles. Les statistiques utilisées sont basées sur le cumul de précipitations et le SPI (*Standardized Precipitation Index*) qui indique l'anomalie par rapport à la moyenne à long terme des précipitations ainsi que sa fréquence. Ces statistiques sont calculées sur l'ensemble de l'horizon de trois mois (e.g. SPI3) ou pour chacun des mois (e.g. SPI1-1, SPI1-2, SPI1-3). Les ensembles basés sur les données historiques de précipitations et de débits sont conditionnés par rapport à leur proximité aux statistiques prévues par Système 4. L'intérêt de la méthode est illustré en Fig. R8. Les statistiques prévues par Système 4 sont fiables, offrent des erreurs globales similaires à la climatologie des précipitations, et sont plus fines que la climatologie. Il est donc attendu que le conditionnement affine les scénarios prévus pour l'horizon de prévision.

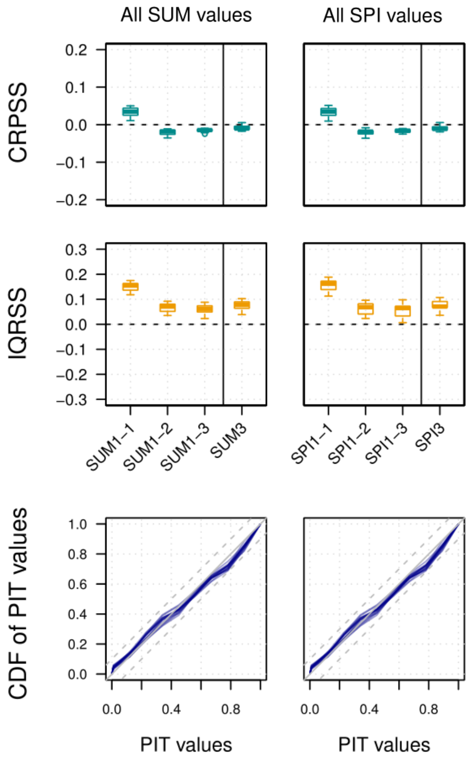



Figure R8: CRPSS, IQRSS et diagramme de PIT des statistiques de précipitations prévues par Système 4. L'ensemble de référence est la climatologie des précipitations.

La finesse, la fiabilité et les performances globales des ensembles produits sont évaluées dans le jeu de seize bassins versants de 1981 à 2010. Ces ensembles sont ensuite comparés en fonction de leur capacité à prévoir des évènements en basses eaux, ainsi que des variables d'intérêt en étiages telles que le nombre de jours ou le volume déficitaire sous un seuil de basses eaux. Enfin, les prévisions sont comparées à l'aide d'un graphique d'évaluation des risques afin d'illustrer ces différences de performances dans le cas de la sécheresse de 2003 dans un bassin français.

Les résultats de l'étude montrent que les sélections basées sur le SPI, et en particulier le SPI calculé sur trois mois (SPI3), produisent des ensembles aux performances plus homogènes entre bassins que les autres méthodes de sélection. Fig. R9 montre simultanément les gains en fiabilité et en finesse des débits obtenus en conditionnant les précipitations historiques avec le SPI3 (ESP_SPI3). Les gains sont définis par rapport aux débits obtenus avec les précipitations de Système 4 débiaisées, et par rapport à l'ESP. Les ensembles conditionnés sont majoritairement plus fins que l'ESP, comme attendu. À 10 jours, ces ensembles sont plus fiables mais moins fins que les débits dérivés de Système 4. Pour des horizons plus lointains, les performances des ensembles conditionnés se rapprochent de celles des prévisions dérivées de Système 4. Par ailleurs, les prévisions construites sont fiables.

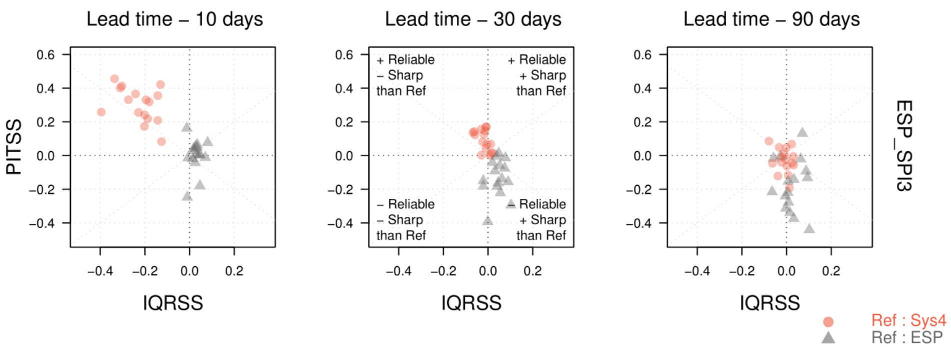



Figure R9: Gains en fiabilité et en finesse des débits obtenus à partir des précipitations conditionnées avec le SPI3, par rapport aux débits obtenus avec les précipitations de Système 4 débiaisées (rouge), et à l'ESP (gris), pour des horizons de 10, 30 et 90 jours.

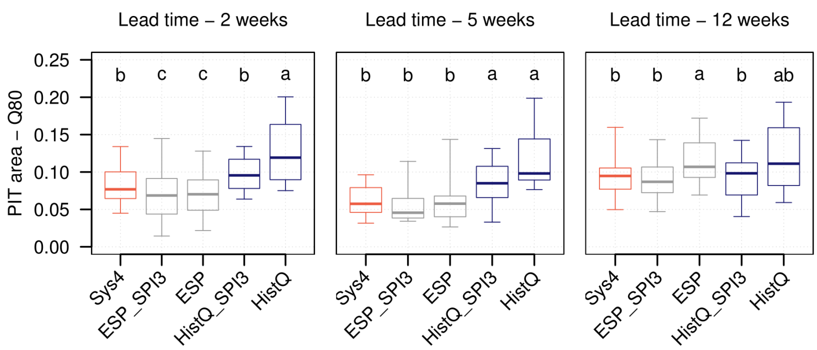



Figure R10: Aires du PIT des prévisions de volume déficitaires sous le quantile 80% obtenues avec cinq méthodes sélectionnées, pour trois horizons.

L'évaluation de la discrimination des ensembles de prévision montre de bons résultats pour tous les ensembles issus du modèle hydrologique (approche dynamique). De plus, les prévisions conditionnés permettent de prévoir le volume déficitaire sous le quantile à 80% (débit excédé par 80% des débits de la chronique disponible) de manière fiable (cf. Fig. R10). Enfin, une application au cas de la sécheresse de 2003 montre que les prévisions sélectionnées peuvent aider à prévoir des évènements extrêmes de manière plus précise en terme de durée et de volume déficitaire.

## 6. Outil d'évaluation des risques en période d'étiages pour le réservoir d'Arzal

Dans les chapitres précédents, nous avons évalué la qualité des prévisions saisonnières de débits en France et illustré leur potentiel pour la prévision des étiages. En effet, les outils d'évaluation des risques peuvent tirer profit des prévisions à long-terme afin d'aider la gestion de la ressource en eau ou la prise de décision des lâchers de réservoirs.

Le barrage d'Arzal se situe à l'embouchure de la Vilaine, juste avant qu'elle ne se déverse

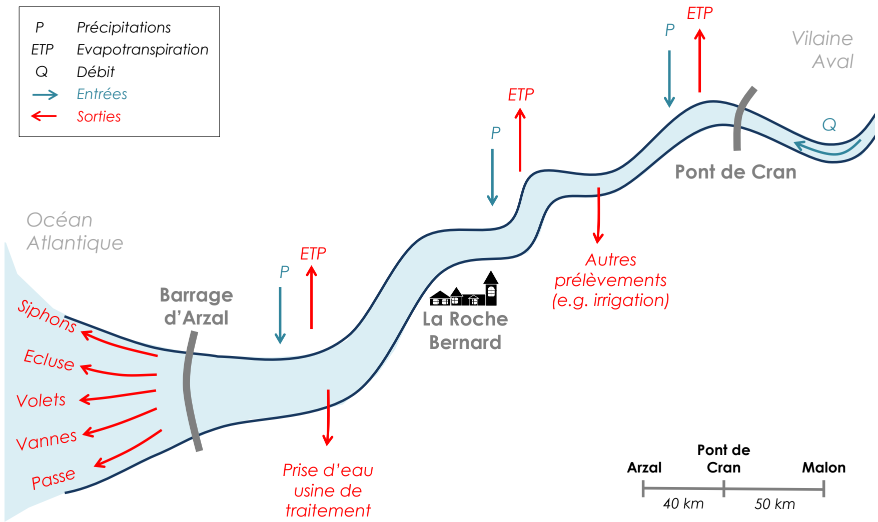



Figure R11: Schéma de la partie avale du réservoir du réservoir d'Arzal, entre le Pont de Cran et le barrage.

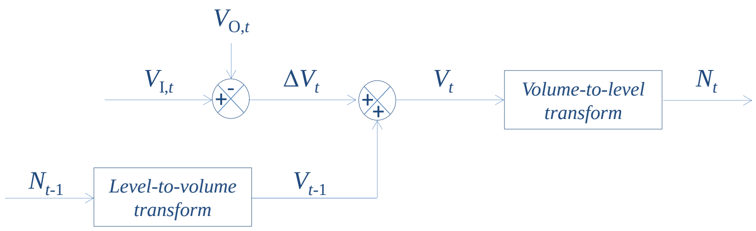



Figure R12: Schema of the structure of the reservoir balance model at a given time step.

dans l'Océan Atlantique (Fig. R11). Sa situation en fait une barrière entre l'eau salée de l'océan et l'eau douce de la Vilaine, créant une retenue d'eau douce de 50 Mm$^3$. Cette retenue est utilisée pour l'approvisionnement en eau potable et l'irrigation. De plus, le barrage permet la navigation et la migration des poissons entre l'océan et la rivière. Le réservoir a donc un rôle clé dans la gestion de la ressource en eau, notamment dans le bassin aval de la Vilaine. L'objectif de ce chapitre est de proposer un outil d'évaluation des risques en contexte de basses eaux pour le réservoir d'Arzal.

Grâce aux données des différentes entrées et sorties du réservoir d'Arzal (cf. Fig. R11) disponibles de 2005 à 2011, nous avons mis en place un modèle simple de bilan d'eau du réservoir. Les étapes du bilan à partir des données de niveaux observés sont schématisées dans la Fig. R12. Les entrées du réservoir sont les apports de la Vilaine et les précipitations. Les sorties comprennent l'eau évacuée via les vannes et volets pour la gestion du réservoir, via la passe à poissons, l'écluse, les siphons, ainsi que l'eau prélevée pour l'eau potable, et l'eau évaporée.

Une analyse préliminaire des données hydrométéorologiques et de gestion du réservoir nous permet de mieux comprendre les stratégies de gestion du barrage et l'importance relative des entrées et sorties du réservoir. Les lâchers des vannes sont étroitement liés aux apports (Fig. R13) et la gestion est très sensible aux apports de la Vilaine, même au pas de temps journalier. Jusqu'en 2014, en année sèche, les volets étaient préférés aux vannes (Fig. R13). La stratégie a été modifiée depuis, de manière à évacuer l'eau saline en fond de réservoir lors des lâchers de vanne.

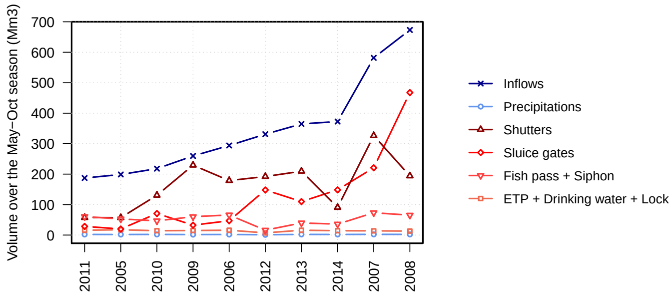



Figure R13: Volumes entrants et sortants cumulés sur la saison de Mai à Octobre, de 2005 à 2014. Les années sont classées de la plus sèche (2005) à la plus humide (2008) en terme d'apports.

Le modèle est utilisé pour estimer le risque de ne pas pouvoir maintenir un niveau minimum dans le réservoir entre Mai et Octobre. Pour évaluer ce risque, des prévisions saisonnières de débits (ESP), étudiées dans les chapitres précédents, sont utilisées en entrée du modèle de bilan. Les sorties des vannes et volets sont optimisées pour maintenir un niveau objectif dans le réservoir, tandis que les sorties liées aux autres usages sont supposées constantes et maximales. Les prévisions de niveaux de réservoir ainsi obtenues nous permettent d'identifier le risque de franchir le seuil minimal dans le réservoir et de déterminer le nombre de jours et le nombre de membres de la prévision d'ensemble atteignant ce seuil. Ces variables d'intérêt pour la gestion du risque sont réunies dans un graphique d'évaluation du risque (Fig. R14). Les graphiques ont été produits rétrospectivement pour les périodes de basses eaux de 2005 à 2010 afin d'évaluer leur potentiel en conditions opérationnelles.

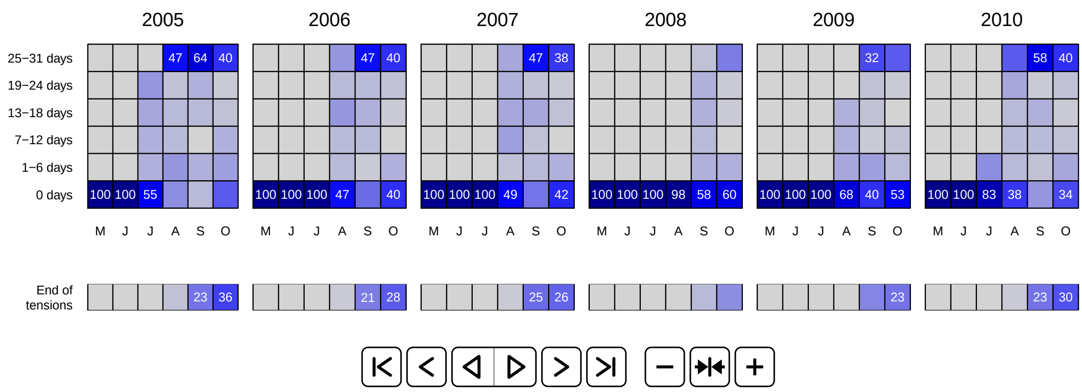



Figure R14: Graphes du risque d'atteindre le seuil minimal de gestion pour la saison à venir. La durée prévue sous le seuil est indiquée verticalement, et les mois de la saison horizontalement. Le graphe inférieur indique la probabilité de revenir à un niveau supérieur au niveau minimum au cours du mois. L'animation présente les prévisions issues du 1$^{er}$ Mai au 1$^{er}$ Septembre.

Enfin, une analyse de sensibilité est conduite afin de déterminer les paramètres du bilan qui influent sur le risque prévu (cf. Fig. R15). Ici, le risque est représenté par le nombre de jours sous le seuil minimal de gestion et est calculé pour l'été 2005. L'analyse montre clairement que le risque est très sensible au niveau maintenu dans le réservoir et très peu sensible aux prélèvements pour l'eau potable. De prime abord le risque est peu sensible aux ouvertures de l'écluse. En réalité, les ouvertures d'écluse sont étroitement liées au fonctionnement des siphons auxquels le risque est bien plus sensible. De même le nombre de jours prévus sous le seuil varie rapidement avec l'eau perdue via la passe à poissons. Cette analyse est en accord avec la gestion actuelle. En cas de sécheresse, la première action est de limiter les ouvertures d'écluses, puis de fermer la passe à poissons.

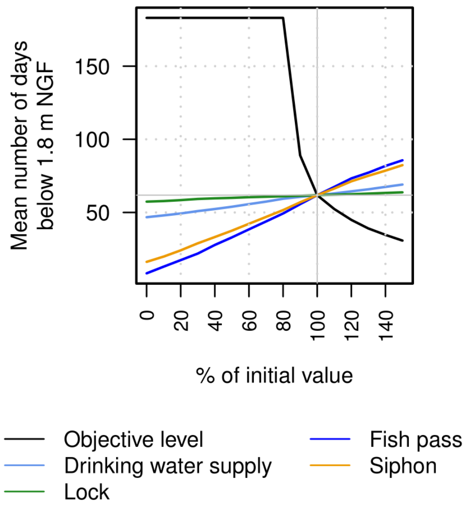



Figure R15: Sensibilité du nombre de jours sous le seuil de gestion aux usages et au niveau objectif, dans le cas de la sécheresse de 2005.

## 7. Une expérience de prise de décision pour la gestion de réservoir en contexte de risque

Les prévisions probabilistes ou d'ensemble, parce qu'elles prennent en compte les incertitudes de prévision, peuvent aider la prise de décision en contexte de risque. Des prévisions de débits aux échéances mensuelles ou saisonnières sont déjà utilisées opérationnellement en gestion de réservoirs, pour des objectifs tels que la répartition de la ressource, l'optimisation des lâchers d'eau ou l'anticipation du risque sécheresse. Dans le chapitre précédent, nous avons notamment proposé un outil d'évaluation des risques en période de basses eaux pour le barrage d'Arzal. Alors qu'il existe de nombreuses études cherchant à estimer l'apport des prévisions hydrométéorologiques d'ensemble pour de telles applications, peu étudient leur rôle pour la prise de décision. Les jeux de rôles peuvent s'avérer très utiles pour mieux comprendre le processus complexe de prise de décision en contexte de risque.

Ce chapitre propose une expérience, sous forme de jeu de rôle, pour mieux comprendre l'usage des prévisions probabilistes à longue échéance pour la décision de lâcher de réservoir. Durant le jeu, les participants endossaient le rôle de gestionnaire de réservoir. À partir d'une séquence de prévisions mensuelles d'apports au réservoir, et étant donnés un objectif de remplissage et des contraintes de lâchers, les participants décidaient séquentiellement des lâchers qu'ils allaient effectuer pour les mois à venir. À la fin de chaque mois, soit après chaque décision, les conséquences des décisions prises le mois précédent étaient évaluées en fonction des apports effectivement observés.

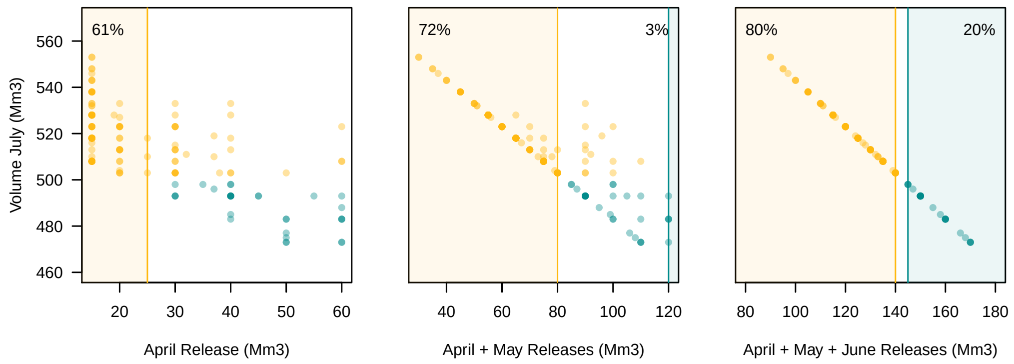



Figure R16: Volumes des 162 participants en fin de saison en fonction des lâchers décidés aux mois précédents. Le volume à ne pas dépasser est de 500 $Mm^3$. Les gagnants sont représentés en bleu, les perdants en jaune. A chaque étape, les pourcentages indiquent le nombre de participants qui sont d'ores et déjà sûrs d'être gagnant ou perdant, étant donnée leur séquence de décisions.

Pour cette étude, 162 feuilles de résultat collectées lors de huit évènements ont permis de mettre en évidence l'enjeu mais aussi la difficulté de prendre à la fois en compte l'information probabiliste et l'information à long terme. Pendant les séquences de jeu, un évènement de crue survenait au mois de juin. La stratégie permettant de finir le jeu sans faire déborder le réservoir consistait à vider progressivement le réservoir dans les mois précédant l'évènement pour pouvoir recevoir les apports de la crue de juin. L'analyse des feuilles de résultats a montré que, sur l'ensemble des séances, seulement 20 % des participants avaient réussi à finir la séquence de décisions sans faire déborder leur réservoir et en respectant les contraintes de lâchers. En réalité, la première décision de lâcher en Avril pouvait déjà déterminer le sort des participants quant à l'évènement de Juin. En effet, 61 % des participants, sans le savoir, avaient perdu dès le premier tour car ils n'avaient pas suffisamment vider le réservoir (cf. Fig. R16). Ceci souligne la nécessité d'anticiper les évènements extrêmes en gestion, et d'avoir les informations nécessaires pour.

Le manque d'anticipation de l'évènement ou la sous-estimation de l'évènement à venir en terme de volume sont des explications possibles au faible taux de réussite des participants. La manière de communiquer les prévisions et la quantité d'informations à intégrer dans le temps imparti pouvaient aussi présenter une difficulté pour les participants et influer la prise de décision. De manière plus générale, l'usage du jeu de rôle a permis de faciliter et créer un contexte favorable à la discussion des enjeux et limites à l'usage de l'information probabiliste à long-terme dans la prise de décision.

## Conclusion

Les prévisions saisonnières sont un sujet d'intérêt pour un grand nombre d'opérationnels dans des secteurs variés allant de la production hydroélectrique à la gestion de bassin. Cependant plusieurs barrières se dressent encore entre la production des prévisions et leur implémentation. Dans cette thèse nous avons abordé deux de ces barrières à l'implémentation en contexte opérationnel: le besoin d'évaluer la qualité des prévisions disponibles, et le besoin de traduire l'information saisonnière de manière à répondre aux besoins des utilisateurs. Nous résumons ici les principales conclusions de la thèse.

**Corriger le biais des prévisions saisonnières de précipitations peut améliorer la fiabilité des prévisions saisonnières de débits.** Dans le jeu de bassins français étudiés, nous avons vu que les précipitations issues du Système 4 du CEPMMT présentaient des biais mensuels qui avaient tendance à se compenser sur l'année. Parmi les huit méthodes testées, la correction quantile-quantile des précipitations journalières permettait d'améliorer la fiabilité des prévisions de précipitations et de débits. De plus, nous avons vu qu'une faible amélioration en performance des précipitations prévues pouvait amener à une amélioration plus conséquente des performances en débits.

**Conditionner les données historiques à partir d'indices de précipitations saisonniers peut améliorer la finesse des méthodes de prévision basées sur la climatologie.** Parmi les quatre indices testés, l'anomalie à 3 mois (SPI3) a permis d'améliorer la finesse des prévisions basées sur la climatologie tout en maintenant leur fiabilité et leurs performances globales. Cette étude a souligné le fait que la relation entre finesse et fiabilité est un compromis. Au-delà du travail de thèse, cette méthode pourrait être appliquée comme une descente d'échelle.

**Un outil pré-opérationnel d'évaluation du risque a été développé pour aider la gestion du barrage d'Arzal.** L'outil est le fruit d'un travail collaboratif avec les gestionnaires du barrage d'Arzal. Il indique, à partir des prévisions d'apports pour la saison, les risques de ne pouvoir maintenir un niveau minimum dans le réservoir en basses eaux. Une analyse de sensibilité pour le cas de la sécheresse de 2005 a montré que le niveau du réservoir est le paramètre le plus influant sur les risques de manque d'eau.

**Un jeu de rôle mettant en scène un réservoir théorique a permis d'analyser le processus de prise de décision basé sur l'information probabiliste saisonnière.** Le jeu a ouvert la discussion sur les enjeux et barrières à l'implémentation des prévision saisonnières. Il a aussi permis de montrer qu'en cas d'évènements extrêmes, la prévision saisonnière permettait une anticipation nécessaire. Cependant, l'expérience a aussi montré la difficulté à intégrer l'information saisonnière probabiliste dans la prise de décision.

# Contents

# Avant-Propos

During the past three years, I have had many opportunities to try and explain my PhD work to experts at project meetings or conferences and to non-experts at family dinners and parties with friends. First, there was the common yet awkward step of being unable to remember the title of my PhD as advertised by my supervisors. At times like that, my cinema references immediately came to my mind and I was often tempted to answer “Les chevaliers-paysans de l’an Mil au lac de Paladru” and walk away. Instead, I usually gathered myself and tried to explain my research with the simplest words I could find at the moment. I have never used a single combination of words nor a single reasoning process twice in that step, and I believe I have rarely been successful with non-experts. However, whenever I was successful, I encountered perplex and incredulous expressions and smirks often followed by questions or assertions that I found very unsettling at first: “but why are you working on this?”, “is someone really paying you to do this?”, “isn’t it a bit pointless?” I soon discovered that these questions would not leave me until the end of my PhD.

In the first months of my PhD, I wanted to avoid questions at all cost and thought the best strategy would be to remain as vague as possible. I tried a few times to simply say: “I work on water issues”, which, in French, translates more or less as “I work in the water”. But I quickly gave up on this approach because it triggered many more questions than it aimed to, including the now famous one “Do you work in a swimming pool?”

After realising I should put a little effort in explaining my work, I would say that I was working on forecasting droughts to help reservoir management. Then came the inevitable and legitimate question: “are you working on a specific dam?” I would respond by mentioning the Arzal dam in Brittany, France, which is the case study of my research and the main reason why the PhD subject was launched in the first place. It was also around that time that I created a short video[1] to explain my research. Therefore, when the questions on my PhD subject arose, I could simply refer the inquisitor to that video saying: “Haven’t you seen my video? Shame.” At this stage, I received various feedbacks. While some were very enthusiastic (mostly the ones who had seen the video), some remained puzzled. The most recurrent remark I got definitely was: “are there any droughts in Brittany?”

Later on in my research, I had the chance to put my hands on the seasonal forecasting

[1] https://www.youtube.com/watch?v=Ap2bfDoy8Wc, winner of the EGU 2014 “Communicate your science” video competition.

products provided by the European Center for Medium-range Weather Forecasts. The description I would give of my research when asked thus evolved to: "forecasting streamflow at the seasonal range to help reservoir management". At this point, I received many more comments from people in the scientific community. Even though many were convinced of the importance of long-range forecasting for many applications, such as low flows and droughts for reservoir management or hydroelectricity and drinking water production, typical comments and questions now were: "does it work?", "do they have any predictability?", "no... not in France!", "... nor in Europe!", "we have tested it, it is pointless, at least in the current state of knowledge."

By the end of my PhD work, I realised that these questions give a meaning to our work and put words to the challenges of this PhD research rather than question its validity. It is precisely the objective of any research to answer such questions and challenge what we assume to be true. My objective in this report is to guide you one last time through the motivations of my PhD research. I will present the main results we reached and explain how we tried to contribute and "add our stone" to existing knowledge on seasonal forecasting in a context of low flows and droughts, with an application to a reservoir management case: the Arzal dam in France (and yes... in Brittany!).

# Introduction

Risk management is defined as "the systematic approach and practice of managing uncertainty to minimize potential harm and loss" (UNISDR, 2009). It aims at enhancing preparedness in the anticipation of an event, as opposed to crisis management, where organizations have to deal with a sudden emergency situation. The type of actions deployed in crisis management and in risk management can be very distinct, but should be coordinated to efficiently reduce and address disaster impacts. In the case of droughts, the transition from crisis management to risk management has long been advocated for and steps have been proposed in the last decades to enhance risk management practices (Wilhite *et al.*, 2000; FAO, 2014). A first step consists in understanding the concept of risk. The risk associated with a natural disaster integrates both the probability of the natural hazard and the likeliness of subsequent negative consequences, also called vulnerability (UNISDR, 2009; Turner *et al.*, 2003). The actions leading to a preventive risk management approach must address both components. These actions can be grouped in four main categories: monitoring, vulnerability assessment, prediction and early warning, and risk reduction (Figure 1).

A central feature of risk management is the prediction system, which assesses the probability of future hazards and informs decisions towards risk reduction. Prediction systems can be combined with warning systems to issue timely warnings to communities at risk or with decision support systems to optimize water resources management in river basins. The development of a prediction system requires a close collaboration between the end-users or decision-makers and the developers of the system (Palmer and Holmes, 1988; Simonović and Bender, 1996). Both have to work from their experience of the natural process at stake, which sets the common framework for the development of the system. The developers of the system will explore the predictability of the natural phenomenon and state what is technically possible in terms of modelling tools. End-users will set their needs and operational expectations as well as provide guidance on practical choices, often based on the particular characteristics of the system subject to decision-making.

The focus of this thesis is to investigate the potential of a low-flow and drought prediction system at the seasonal scale in the context of water reservoir management.

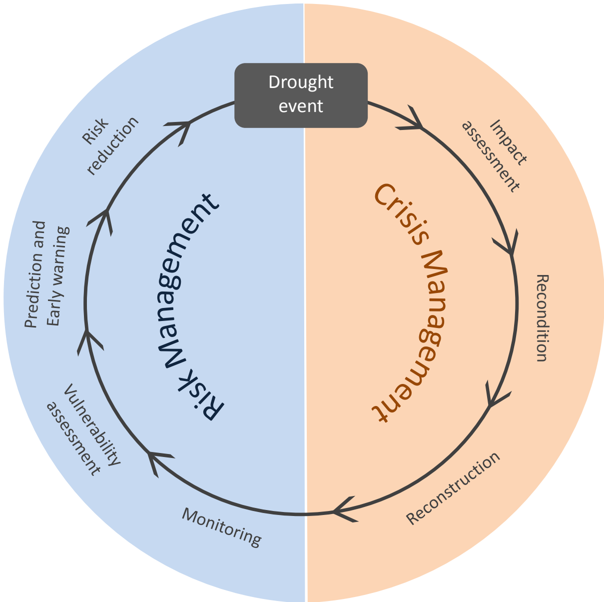



Figure 1: The cycle of risk and crisis management in the case of a drought (adapted from Wilhite *et al.*, 2000; FAO, 2014)

## Low flows and droughts

Low-flow periods are a natural part of a river's flow regime. In the mid-latitudes of the northern hemisphere, a distinction is made between summer and winter low flows. The former is caused by a lack of rainfall combined with high evaporation rates, while the latter occurs when precipitation is retained as snow and does not immediately contribute to catchment runoff (WMO, 2008).

Droughts are occasional phenomena that are defined relatively to conditions of rainfall and evapotranspiration perceived as "normal" in a specific area. Wilhite and Glantz (1985) defined four categories of droughts:

- a meteorological drought consists in a lack of precipitation over an extended period of time;
- an agricultural drought is a combination of weather, soil moisture, groundwater and surface water conditions that impact agricultural practices and production;
- a hydrological drought is a water deficiency in water bodies: surface water, groundwater or reservoirs. This type of drought is a consequence of a meteorological drought but, unlike agricultural droughts, can occur several months after precipitation deficiency;
- a socio-economic drought corresponds to a period during which hydrometeorological conditions cause a disequilibrium between water supply and demand, impacting economic goods or services, e.g. crops, drinking water distribution or hydroelectric power production.

Droughts can last much longer than low flows, with time scales ranging from one month to several years. Europe has known major drought events during the 20th and 21st centuries, the latest one being the drought of 2015. Each of these drought events has contributed to shaping the way countries prepare for and manage future events. Table 1 illustrates five European drought events and their impacts, with a focus on their contribution to changes in the French national policies.

Table 1: Time and spatial extent of the European droughts of 1976, 1989, 2003, 2011 and their main impacts (modified from La Jeunesse *et al.*, 2016[1]).

| Year | Spatial extent | Time period | Main impacts in France (EDC, 2013) |
|---|---|---|---|
| 1976 | Europe, more specifically northern Europe | Winter 1975 to summer 1976 | Human casualties in France<br>Severe loss of spring crops, 1.7 MT of straw transferred (Brochet, 1977)<br>Wildfires<br>Death of fish and livestock<br>Restriction on drinking water supply<br>High concentration of pollutants in rivers<br>Reduction of hydropower and nuclear power production<br>**Drought tax** to FFRANCS 2.2 billion |
| 1989 | Most European countries | Summer to autumn 1989, with impacts lasting until end of summer 1990 | Loss of crops, wildfires<br>High concentration of pollutants in rivers<br>Death of fish<br>Reduction of hydropower and nuclear power production<br>**Temporary water ban** through revision of the Water Law (passed in 1992) |
| 2003 | Europe, more specifically west Europe | Summer 2003 | 14,800 human casualties in France (Pirard *et al.*, 2005; Robine *et al.*, 2007)<br>Important loss of crops (UNEP, 2004), wildfires and insect invasions<br>Death of fish and livestock<br>Decrease in nuclear power production (UNEP, 2004)<br>Water use restrictions in 75% of the French departments<br>Natural Disaster Declaration for 4,400 municipalities<br>Drought-induced soil subsidence (Corti *et al.*, 2009)<br>Economic costs estimated to EUR 1.1 Billion<br>**Drought Action Plan and National Drought Committee** |
| 2011 | Most European countries | Spring 2011 | Loss of crops, wildfires<br>Difficulties with cattle feeding, farmers having to sell their animals<br>Water use restrictions in 75% of the French departments<br>Historic deficit of hydropower production<br>**National guarantee funds through agricultural disaster**<br>**Implementation of PROPLUVIA[2] to manage water bans** |
| 2015 | Most European countries | June and July 2015 | Civil and industrial water use restrictions (EDO-JRC, 2015)<br>Loss of crops, wildfires and insect invasions<br>Decrease in nuclear and hydropower production |

The increased severity of drought consequences in the last decades as well as the expected increase in drought occurrences in the coming years due to climate change have amplified

[1]A modified version of this table corresponds to my contribution to the DROP handbook chapter: La Jeunesse, I., Larrue, C., Furusho, C., Ramos, M.-H., Browne, A., de Boer, C., Vidaurre, R., **Crochemore, L.**, Penasso, A., Arrondeau, J.-P., In press. Chapter 6: The governance context of drought policy and pilot measures in the Arzal dam and reservoir, Vilaine catchment, Brittany, France., in: Governance for Drought Resilience. H. Bressers, N. Bressers, C. Larrue (Eds.).

[2]http://propluvia.developpement-durable.gouv.fr/

the need for a better understanding of drought mechanisms and for improved drought predictions (Mishra and Singh, 2010, 2011, and references therein). The prediction of droughts and low flows still raises many challenges such as identifying the predictability of drought characteristics or better understanding the local processes involved in drought mechanisms (Wood *et al.*, 2015). It is the topic of many recent research projects, e.g. the DEWFORA project in Africa, the DROUGHT R-SPI project in Europe (Gudmundsson *et al.*, 2014) or the PREMHYCE project in France (Nicolle *et al.*, 2014). Improving forecast skill and linking forecasts to user needs are often highlighted.

> This thesis focuses on the prediction of low flows at the seasonal time scale to anticipate hydrological droughts.

## Seasonal predictability and forecasting

Predictability can be interpreted as the extent to which errors in our knowledge of current states impact our capacity to predict future states (Lorenz, 1984; Kirtman *et al.*, 2013). The theoretically attainable predictability mainly depends on the natural process at stake and the geographical location where such process occurs. Identifying local sources of predictability can shed light on where to reduce or better represent uncertainties in order to enhance the skill of a prediction system (Shukla *et al.*, 2013; Yossef *et al.*, 2013; Wood *et al.*, 2016). Limits in hydrological predictability come from the limits in our capacity to represent the physical entity that is modelled in the system, i.e., the hydrological catchment, but also from the uncertainties in initial conditions in the first forecast horizons and in boundary forcings at longer forecast horizons (Blöschl and Zehe, 2005). The relative importance of initial conditions and boundary forcings will vary with the catchment inertia and its long-term memory of initial conditions.

Advances in meteorological and climate predictability can contribute to improving the skill of streamflow forecast systems at sub-seasonal to seasonal time scales. Meteorological processes occur daily in the atmosphere and define our local weather. These processes are hardly predictable up to more than two weeks due to the chaotic nature of the atmosphere and its short residence time for moisture (Chow *et al.*, 1988; Stocker *et al.*, 2001). By contrast, climate processes, as represented by coupled ocean-atmosphere global circulation models, are predictable for lead times up to several years ahead (Yuan *et al.*, 2015). These large-scale climate processes describe the expected long-term signals of meteorological variables, such as differences from average conditions. Coarse long-term meteorological signals can then be used to achieve additional predictability in long-term hydrological forecasting. Indeed, hydrological catchments and processes therein play the role of low-pass filters, preserving the long-term tendencies and filtering random high frequency errors.

Seasonal forecasting has been the topic of many projects launched in the past years, such as the EUPORIAS project (Hewitt *et al.*, 2013) or the S2S project (Robertson *et al.*, 2015). Both projects promote the uptake of products from climate and weather services by decision makers from different socio-economic sectors.

In this thesis, we aim to assess the available skill of seasonal streamflow forecasts, and extend the predictability of low flows and droughts, with a focus on water resources and reservoir management.

## Reservoir management

User needs and operational expectations are dictated by the characteristics and the vulnerability of the system subject to decision-making as well as the economic stakes of the sector impacted by the decisions taken. In this thesis, the targeted user of the investigated hydrological forecasts is the manager of a water reservoir.

Reservoir management is a central part of optimization strategies that aim at increasing resources efficiency. Indeed, dams and reservoirs are key elements for regulating water flows and harnessing water resources. As opposed to reservoir management during high flows, which focuses on limiting the consequences of flooding and the damages to hydraulic structures, the management of a reservoir during low flows or droughts focuses on preserving the water as a resource and on guaranteeing its priority uses. Multi-purpose water reservoirs may be subject to strong constraints which require high managing skills and, usually, sophisticated decision support systems (Castelletti *et al.*, 2008).

Decision-making is a complex process that mainly relies on past experience (Cannon-Bowers and Bell, 1997). Inflow prediction systems can help inform decision makers by quantifying risks that otherwise would not have been foreseen. During dry periods, reservoirs are strongly dependent on inflows from upstream catchments. Tools based on inflow forecasts can directly benefit from the long-range predictability of inflows to help dam managers anticipate and take preventive actions (Georgakakos and Graham, 2008; Regonda *et al.*, 2011). In addition, reservoir management models can be valuable to analyse several management scenarios and guide towards an optimal management for the running low-flow season (Yeh, 1985; Kelman *et al.*, 1990).

The work in this PhD thesis was prompted by the needs of the managers of the Arzal reservoir (Brittany, France) in terms of inflow forecasting and drought risk assessment. It was carried out within the framework of the DROP Interreg IVB project.

## The DROP project

The DROP project[3] (*Benefit of Governance in DROught AdaPtation*), aimed at helping Northwest Europe regions better anticipate and prepare for periods of drought and water scarcity. It was an Interreg IVB project that ran from 2013 to 2015. Six regional water authorities and five research institutes from the United Kingdom, Belgium, the Netherlands, Germany and France collaborated on drought governance and management topics. The project had six pilot sites, which were twinned to investigate drought issues under three perspectives: freshwater, agriculture and nature. One of the main outputs of the project is a drought governance assessment tool that was developed to help characterize local drought governance contexts and serve as a basis for recommendations on how to improve local governance (Bressers *et al.*, 2016).

The multi-purpose Arzal reservoir located in Brittany, France, was one of the freshwater pilot sites of the DROP project. The project showed that, in the freshwater pilots, drought governance is still based on crisis management. The transition towards risk-oriented practices is mainly lagged by the lack of recurrent severe drought events. Nevertheless, awareness of water scarcity exists and it is common knowledge that freshwater resources are limited in a context of increasing demand (Furusho *et al.*, 2016). In the specific case of the Arzal reservoir, the DROP project also showed that even though freshwater is recognized as the priority use by all stakeholders, restrictions on other uses can be sources of tension during summer (La Jeunesse *et al.*, 2016).

This PhD thesis was partly funded by the DROP project, with the aim to develop a pre-operational integrated low-flow forecasting and water management system to support reservoir management and drought adaptation for the Arzal reservoir. The project also supported the development of a game designed to better understand how probabilistic forecasts can be used for decision-making in reservoir management (Crochemore *et al.*, 2015b[4]). The game was played at the French pilot site and was later adapted by the stakeholders of the German pilot site of the DROP project (Wasserverband Eifel-Rur, WVER).

## Aims of the research

This thesis focuses on the seasonal time scale in hydrometeorology for the prediction of low flows and droughts, with the overall goal of enhancing decision-making in water reservoir management. It relies on the idea that better hydrological prediction tools can help bridge climate services and water managers needs, and lead to a faster adoption of novel products and developments.

The aims of this research can be summarized through four questions: the first two address seasonal streamflow forecasting and the remaining two focus on reservoir management.

[3] http://www.dropproject.eu

[4] The game is presented in Chapter 7 of this thesis and can be freely downloaded from http://hepex.irstea.fr/resources/gamestraining.

1 How much can we improve streamflow forecasting by bias correcting meteorological forecasts derived from climate models?

2 How can climate model outputs be used to improve traditional streamflow forecasting approaches based on historical records, and better capture low-flow duration and severity?

3 How should a low-flow forecasting system be tailored to respond to the needs of the managers of the Arzal reservoir during summer seasons?

4 What lessons can we learn from the way decision makers use probabilistic long-term information to manage a theoretical multi-objective reservoir?

## Structure of the thesis

This thesis is composed of seven chapters organized into three main parts.

Part I presents the main tools used in this research. Chapter 1 introduces the catchments as well as the hydrometeorological data available to this research. Chapter 2 presents the hydrological model used and details the framework adopted for its calibration. The preliminary step of validation of the hydrological model over the studied catchments is presented, as well as the ensemble forecasting framework. In Chapter 3, the different numerical criteria, either deterministic or probabilistic, used to assess the performance of forecasts are presented.

Part II focuses on the evaluation of seasonal forecasts and addresses the first two key questions of this thesis. In Chapter 4, the skill of precipitation forecasts from ECMWF (European Centre for Medium-Range Weather Forecasts) System 4 climate model to forecast streamflow over the studied catchments is investigated. Eight bias correction methods are applied and compared based on their impact on both precipitation and streamflow forecasts. Chapter 5 presents a conditioning of historical streamflows and precipitations based on System 4 precipitations developed to extend the range of predictability of low flows. It also introduces a risk visualisation tool developed to assess the risk of low reservoir inflows along the dry season.

Part III tackles reservoir management and risk-based decision-making in order to address the last two key questions of this thesis. Chapter 6 focuses on the case study of the Arzal dam in Brittany, France. The management data available to this study are presented and a quantitative model of the reservoir is developed. A risk assessment tool designed through interactions with end-users and thus adapted to user needs is proposed. Risk visualisations benefit from seasonal forecasts and can be used as early warnings for low-flow management at the Arzal dam. In Chapter 7, we present the set-up and results of a game experiment, which aims to improve our overall understanding of how long-term probabilistic forecast information can be used in sequential decision-making in a theoretical multi-objective reservoir.

Lastly, the main conclusions of this thesis are discussed and new perspectives are outlined.

# *I*

# Hydrometeorological data, hydrological model and forecast evaluation

# 1

# Data collection and control

## 1.1 Catchment sets and hydrological data

Two different catchment sets are used in this thesis. The first one is a countrywide set that includes 16 catchments with long time series of observed meteorological and streamflow data in France. This catchment set is used to investigate the skill of seasonal precipitation forecasts for streamflow forecasting in France (Chapters 4 and 5 of this thesis). The second catchment set focuses on the Vilaine river basin in Brittany, France. In this catchment, we are more specifically interested in the river inflows to the Arzal reservoir, located at the outlet of the catchment. Data from the Vilaine river are used to set up a water balance model of the Arzal reservoir and investigate the added-value of a low-flow forecasting and management system for reservoir operation (Chapter 6). The following sections describe the catchment sets and their data.

### 1.1.1 Countrywide catchment set

The catchment countrywide set was selected from the catchment set used in Nicolle *et al.* (2014), who compared several hydrological models for low-flow simulation and forecasting in France. Their database was built based on three constraints:

- to gather catchments that are representative of a variety of climatic and hydrologic conditions found in French catchments;
- to ensure that the selected catchments had long time series of meteorological and streamflow data for robust model calibration and validation;
- to select catchments that are not strongly influenced by human activities, while selecting enough catchments to draw general conclusions (e.g. Andréassian *et al.*, 2009).

In this thesis, an additional constraint was considered. Catchments were selected provided that their solid fraction of precipitation over the basin was below 10 % over the available record period. In other words, we only considered catchments that were not, or marginally, influenced by snow. This resulted in a set of 16 catchments. For each catchment, daily streamflow observations were available from the French HYDRO database (www.hydro.eaufrance.fr), which is a national database administered by the National flood forecasting centre (SCHAPI - *Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*). Discharges over French hydrological stations are provided by local public institutions, research institutes or private companies, who are also responsible for data quality.

Figure 1.1 presents the locations and the hydrological regimes of the studied catchments. Their main characteristics are presented in Table 1.1. Catchments are ranked from the smallest (377 km$^2$) to the largest (4320 km$^2$). We can see that streamflow data availability for these catchments varies from 36 to 52 years.

The graphs in Figure 1.1 show the mean and the 80 % confidence interval of monthly flows computed over the streamflow records available for each catchment. They also show the 80$^{\text{th}}$ exceedance percentile computed over all available data (i.e., the streamflow value that is exceeded by 80 % of the available data). The catchment set is mainly characterized by a pluvial regime and a strong intra-annual variability of streamflows. High flows are typically observed during the rainy season, running approximately from November to May. Low flows are typically observed between May and October, and very low flows, between July and September. It is also during these months of very low flows that the 80$^{\text{th}}$ exceedance percentile tends to be reached. As expected, the 80 % confidence interval is much wider during the months of high flows. Catchment 1 (the smallest in the set) shows a slightly different regime pattern, with a large 80 % confidence interval during low flows as well. We observe that the inter-annual variability in this catchment is of the same order of magnitude as the intra-annual variability.

In Figure 1.2, the Turc-Budyko representation (Andréassian and Perrin, 2012) was used to analyse the balance between the mean annual precipitation, the mean annual evapotranspiration and the mean annual flow of the catchments. These are represented in blue in the figure. More than 2000 catchments spread over France (represented in grey) provide an overview of the range of conditions found in French catchments. The studied catchments are located within the cloud of French catchments. This indicates that our countrywide catchment set is well within the range of conditions observed in French catchments and does not include any outliers, as far as the annual water balance is concerned. From this figure, we can also see that the selected catchments have no unexplained water gain or loss.

### 1.1.2 The Vilaine river basin

The Vilaine river basin has an upstream area of about 10,000 km$^2$ and is located in the southwest of the Brittany region, in France (Figure 1.3). The source of the Vilaine river is located in the hills of Juvigné, at an altitude of only 150 meters (the maximum altitude in the Vilaine catchment is 340 meters). About 220 km later, the Vilaine river flows into the reservoir of the Arzal dam, just before the Atlantic Ocean.

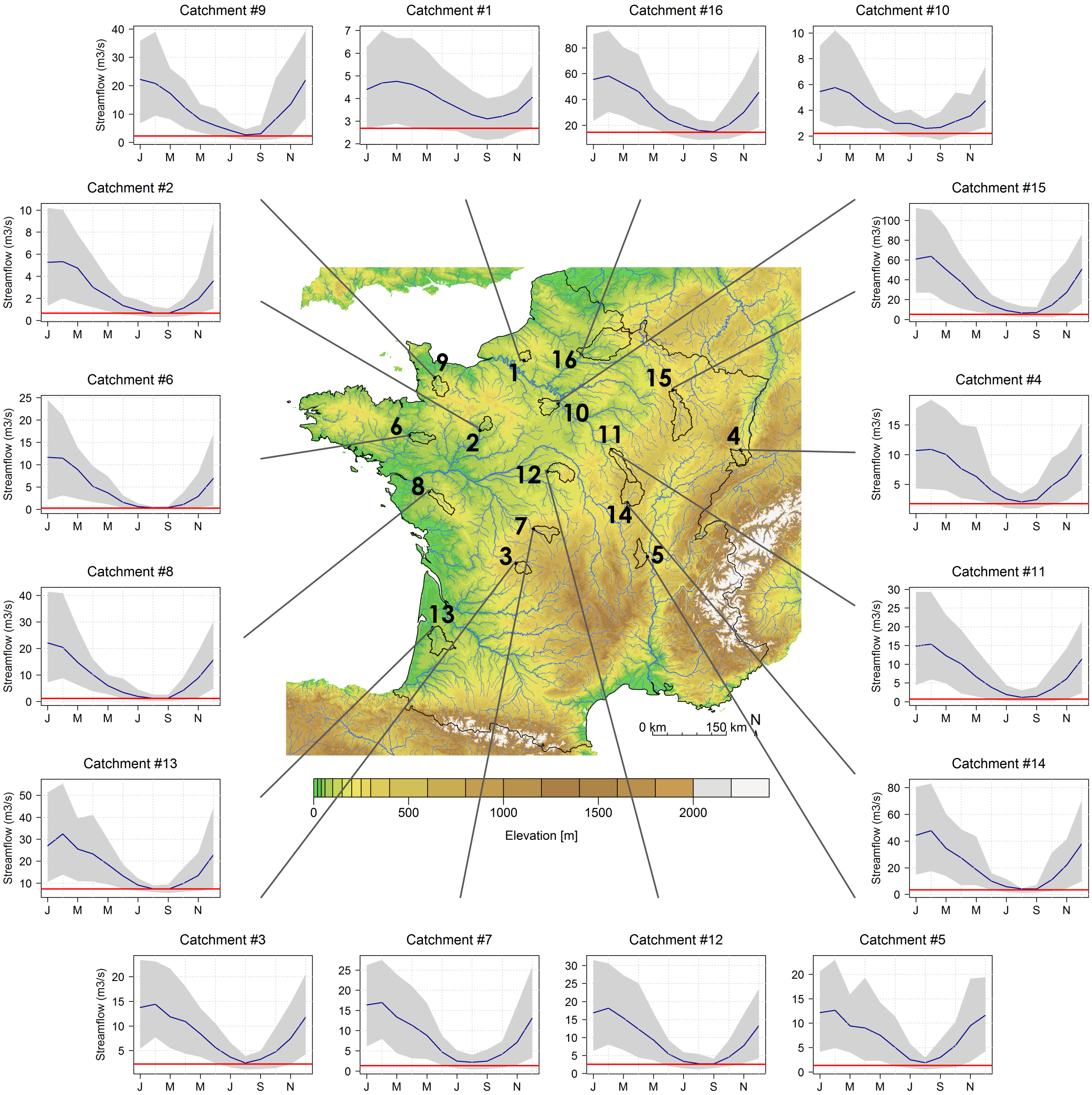



Figure 1.1: Location in France and hydrological regime of the 16 catchments studied in Chapters 4 and 5 of this thesis. Blues lines represent mean interannual monthly flows. Grey-shaded areas represent the 10$^{th}$ and 90$^{th}$ percentiles of interannual monthly flows. Red lines represent the 80$^{th}$ exceedance percentile (i.e. the daily flow exceeded by 80 % of the data). The catchments are numbered from the smallest to the largest. Statistics are computed over the streamflow record available for each catchment, i.e. 36 to 52 years (cf. Table 1.1).

Table 1.1: River and gauging station, catchment ID, period with available streamflow observations, area and main hydroclimatic characteristics of the 16 catchments in the countrywide set (ranked from the smallest to the largest).

| # | River | Gauging station | Catchment ID | Streamflow availability[1] | Area km$^2$ | Mean annual precipitation mm/yr | Mean annual potential evapotranspiration mm/yr | Mean annual streamflow mm/yr |
|---|---|---|---|---|---|---|---|---|
| 1 | Andelle | Vascoeuil | H8212010 | 01/01/1973-27/02/2010 | 377 | 952 | 628 | 332 |
| 2 | Orne Saosnoise | Montbizot [Moulin Neuf Cidrerie] | M0243010 | 01/12/1967-04/03/2010 | 501 | 735 | 696 | 163 |
| 3 | Briance | Condat-sur-Vienne [Chambon Veyrinas] | L0563010 | 01/01/1966-28/03/2010 | 605 | 1100 | 706 | 427 |
| 4 | Ill | Didenheim | A1080330 | 01/11/1973-02/03/2010 | 668 | 956 | 664 | 309 |
| 5 | Azergues | Lozanne | U4644010 | 01/01/1965-28/03/2010 | 798 | 931 | 689 | 296 |
| 6 | Seiche | Bruz [Carcé] | J7483010 | 01/12/1967-11/03/2010 | 809 | 732 | 696 | 181 |
| 7 | Petite Creuse | Fresselines [Puy Rageaud] | L4411710 | 01/08/1958-28/03/2010 | 853 | 899 | 680 | 316 |
| 8 | Sèvre Nantaise | Tiffauges [la Moulinette] | M7112410 | 01/11/1967-04/03/2010 | 872 | 898 | 712 | 331 |
| 9 | Vire | Saint-Lô [Moulin des Rondelles] | I5221010 | 01/01/1971-03/02/2010 | 882 | 958 | 629 | 448 |
| 10 | Orge | Morsang-sur-Orge | H4252010 | 01/10/1967-07/03/2010 | 934 | 658 | 680 | 131 |
| 11 | Serein | Chablis | H2342020 | 01/08/1958-03/03/2010 | 1119 | 842 | 675 | 220 |
| 12 | Sauldres | Salbris [Valaudran] | K6402520 | 01/01/1971-28/03/2010 | 1220 | 803 | 684 | 240 |
| 13 | Eyre | Salle | S2242510 | 01/01/1967-19/03/2010 | 1678 | 1025 | 785 | 323 |
| 14 | Arroux | Etang-sur-Arroux [Pont du Tacot] | K1321810 | 01/11/1971-27/03/2010 | 1792 | 981 | 655 | 390 |
| 15 | Meuse | Saint-Mihiel | B2220010 | 01/07/1968-03/01/2010 | 2543 | 948 | 639 | 372 |
| 16 | Oise | Sempigny | H7401010 | 01/08/1958-02/03/2010 | 4320 | 805 | 639 | 250 |

[1] Meteorological data were available from August 1958 to July 2010 in all catchments.

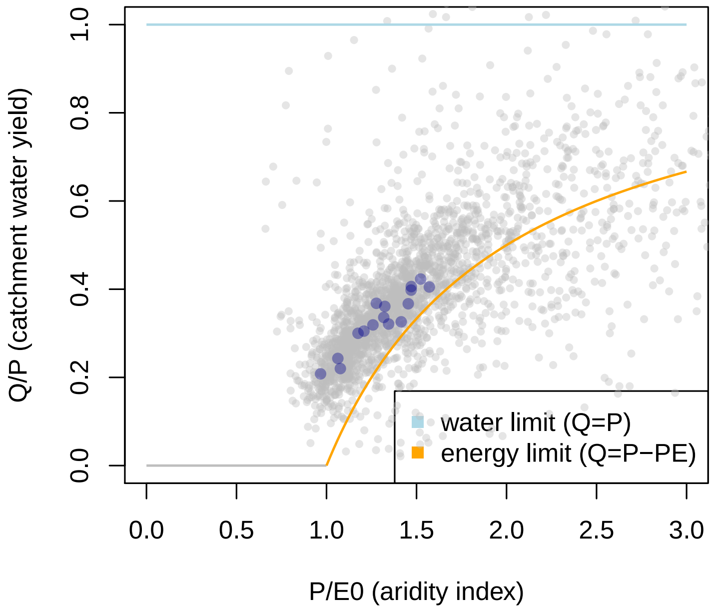



Figure 1.2: The Turc-Budyko representation of the 16 catchments from the countrywide catchment set (blue dots). Grey dots represent a set of more than 2000 catchments spread over France. Catchments should ideally be located below the blue line and above the orange curve. Catchments located above the blue line tend to gain water, while catchments located below the orange curve tend to lose water or leak.

The major part of the catchment area (62 %) is occupied by agricultural lands, producing mainly maize and wheat, and lands dedicated to livestock farming (SAGE, 2015a). Industries in the catchment are concentrated in its upper areas, with food production being the sector best represented. In 2012, 1,260,000 inhabitants lived in the Vilaine river catchment, including 700,000 inhabitants of the Rennes agglomeration. Industries, agriculture, and drinking water supply for local inhabitants and tourists induce important water withdrawals, which divide as follows: 68.5 $Mm^3$ (79 %) for drinking water supply, 10.4 $Mm^3$ (12 %) for industries and 8.1 $Mm^3$ (9 %) for irrigation (SAGE, 2015b). Other water uses include energy production, with 19 hydroelectric power plants located in the catchment, and tourism, with activities such as sailing and fishing (Figure 1.4).

The Arzal dam is an estuarian dam located at the outlet of the Vilaine river basin, just before the Atlantic Ocean. Its primary goal is to block the ocean tides, which used to contribute to the flooding in the downstream part of the catchment. It also creates a freshwater reservoir that contributes to meeting the water demand in the region. The inflows to the water reservoir can be estimated from the streamflow recorded at the Rieux gauging station located at Pont de Cran (the last station on the Vilaine river before the Atlantic Ocean). The streamflow recorded at this station comprises the flows from the Vilaine river and all its tributaries (Figure 1.3). In the upper part of the Vilaine, these

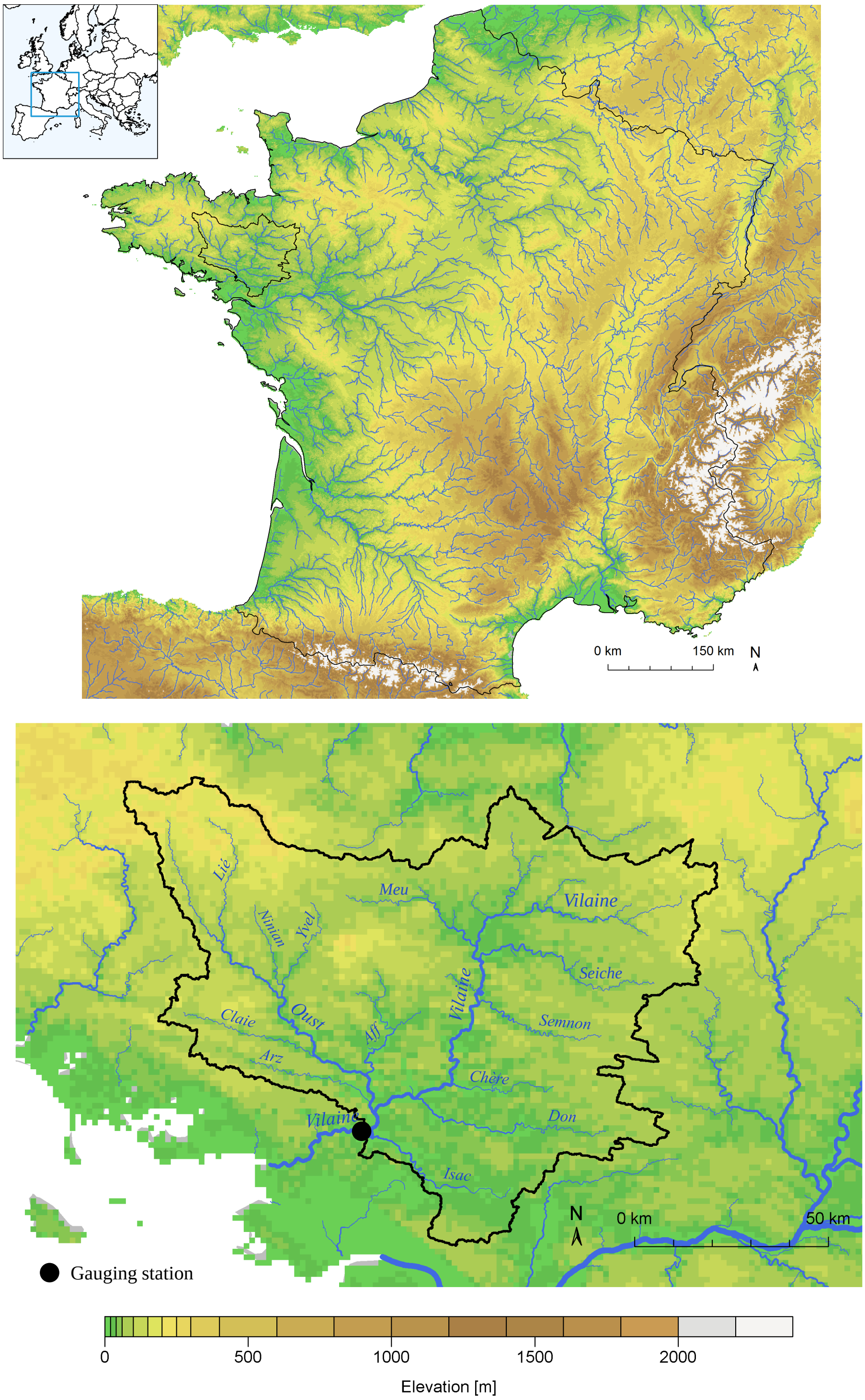



Figure 1.3: Location, topography, tributaries of the Vilaine river at Rieux, and gauging station.

Figure 1.4: View of the Vilaine river flowing towards the ocean at dusk, seen from the sailing harbour at La Roche Bernard.

tributaries include the Meu, the Seiche and the Semnon rivers, and later on, the Chère and Don rivers. Yet, the main tributary of the Vilaine river is the Oust which meets the Vilaine in its downstream segment. The Oust is fed by the Lié, the Ninian, the Yvel, the Claie, the Aff and the Arz rivers. The Isac is the last river to reach the Vilaine river before the Atlantic Ocean and before the Pont de Cran.

Flow records for the Vilaine river were directly collected from the IAV (*Institution d'Aménagement de la Vilaine*), which is the public institution responsible for the management of the Arzal dam. Among other specifications, WMO (2008) recommend using a minimum flow record of 5 years, ideally without discontinuities or interruptions, for low-flow analysis. For the purpose of this thesis, streamflow data recorded every 10 minutes at the Rieux gauging station (at Pont de Cran) were collected for the period running from April 2003 to July 2012. These observations are based on ultrasonic measurements of the flow velocity, as well as bathymetric data. According to the dam managers, these measures may not be accurate during very low flows, due to the calibration of the ultrasonic sensors. Furthermore, flow velocity can be influenced by dam operations when large water volumes are evacuated through the gates, leading to artificially higher velocities.

We prepared the streamflow data based on the specifications given by WMO (2008) for low-flow analysis. Prior to aggregating data at the daily time step, we removed negative values and linearly interpolated them at the 10-minute time step. Mean daily flows were then computed for days with at least half of the measures available and at least 25 % of the measures available in both halves of the day. To minimize the impact of remaining missing values, daily values were interpolated when gaps did not exceed three consecutive days. The characteristics of the raw and quality-controlled data are presented in Table 1.2.

Table 1.2: Main characteristics of the raw and quality-controlled streamflow data for the Vilaine river at Rieux.

| Characteristics | Raw | After quality control |
|---|---|---|
| First available data | 01/04/2003 17:10 | 02/04/2003 |
| Last available data | 21/03/2013 00:00 | 21/03/2013 |
| Frequency | Every 10 minutes | Every day |
| Negative values | 2.20 % | - |
| Missing data | 4.45 % | 4.12 % |
| Maximum consecutive negative values | 48 (0.3 day) | - |
| Maximum consecutive missing values | 853 ($\sim$ 6 days) | 6 days |
| Correct data | 93.35 % | 95.88 % |

The rainfall regime of the Vilaine catchment is governed by an oceanic climate, with a wet season observed from October to January and a dry season from June to August. On average, the catchment receives approximately 790 mm/yr of precipitation and has a potential evapotranspiration of about 680 mm/yr, with mean annual temperatures around 12°C. The discharge per unit area of the Vilaine river at Rieux is of 220 mm/yr, and the average daily flow is 60 $m^3/s$. Over the available data, extreme low flows reach flow values as low as 2 $m^3/s$ (October 2005), and extreme high flows reach values as high as 1000 $m^3/s$ (February 2014). Figure 1.5 presents the hydrological regime of the Vilaine at Rieux. As in Figure 1.1, grey-shaded areas represent the 80 % confidence interval of monthly streamflows, the blue line represents the mean interannual monthly streamflow and the red line represents the 80$^{th}$ exceedance percentile. We observe a wide intra-annual variability in flows as well as a wide interannual variability in high flows (represented by a wide interquantile range between December and March). The low-flow period can be identified as running from June to October. In August and September, the mean monthly flow is very close to the 80$^{th}$ exceedance percentile.

Figure 1.6 and Table 1.3 present some of the characteristics of the Vilaine river at Rieux in low flows. Figure 1.6 presents the cumulative distribution of observed flows, along with the 70$^{th}$ , 80$^{th}$ , 90$^{th}$ and 95$^{th}$ exceedance percentiles (named Q70, Q80, Q90 and Q95, respectively). The exceedance percentile $Q_N$ is defined as the flow exceeded N % of the

Table 1.3: Mean annual minimum flow (MAM) for flows averaged over 3, 7, 10, 30 and 90 days for the Vilaine river at Rieux and based on data from 2003 to 2012.

| Duration (days) | 1 | 3 | 7 | 10 | 30 | 90 |
|---|---|---|---|---|---|---|
| MAM ($m^3/s$) | 5.5 | 6.7 | 8.0 | 8.6 | 10.4 | 14.6 |

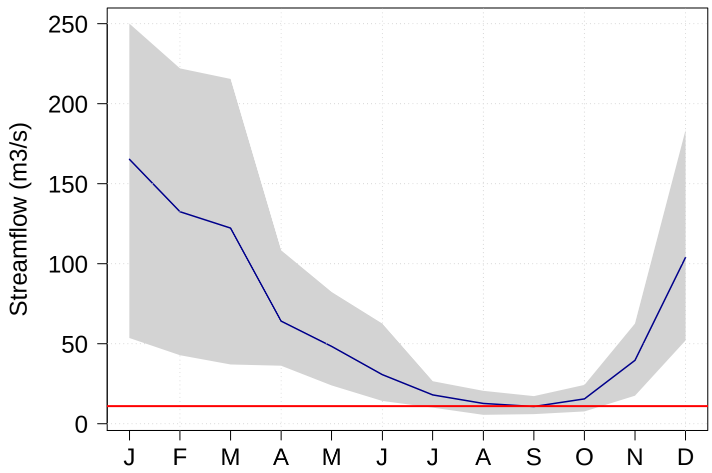



Figure 1.5: Hydrological regime for the Vilaine river at Rieux. The mean interannual monthly flows (blue line), the $80^{\text{th}}$ annual exceedance percentile (red line) and the $10^{\text{th}}$ and $90^{\text{th}}$ quantiles of monthly flows (grey-shaded area) are shown based on data from 2003 to 2012.

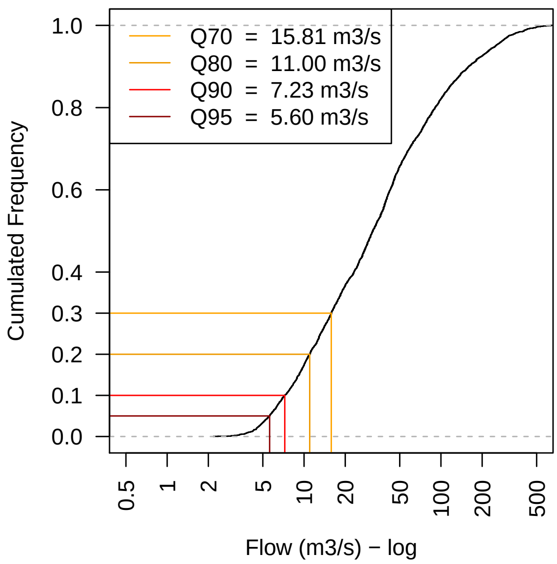



Figure 1.6: Flow duration curve and low-flow exceedance percentiles Q70, Q80, Q90 and Q95 for the Vilaine river at Rieux based on data from 2003 to 2012.

time in the data period. In Table 1.3, daily flows were averaged over time steps ranging from 3 to 90 days. For each aggregation, the minimum flow values observed each year were averaged so as to obtain a mean annual minimum. This mean of annual minima was only averaged over years with less than 5 % of missing data. We learn from this table that having three consecutive months with a mean streamflow of 15 m$^3$/s or a month with a mean streamflow of 10 m$^3$/s are average situations in the Vilaine at Rieux catchment.

## 1.2 Meteorological data

### 1.2.1 Observed meteorological data

The observed time series of daily precipitation and potential evapotranspiration used in this thesis come from the SAFRAN reanalysis of Météo-France (Quintana-Seguí *et al.*, 2008; Vidal *et al.*, 2010). Daily values are available for an 8 x 8 km grid covering France. Potential evapotranspiration was computed based on temperature data and extra-terrestrial radiation (Oudin *et al.*, 2005). These data were aggregated at the catchment scale to obtain time series of areal precipitation and potential evapotranspiration for each of the 17 catchments described in Section 1.1.

In addition to the daily meteorological data described above, we collected local precipitation data from a rain gauge located at the Arzal dam, downstream the Vilaine river. These data were collected directly from the dam managers (IAV). Precipitation values at the hourly time step were collected for the period running from January 1990 to February 2009, and averaged at the daily time step. Precipitation values already aggregated at the daily time step were collected for the period running from March 2009 to May 2015. In total, 25 complete years of data were available. These local precipitation data are used in the water balance model of the Arzal reservoir (Chapter 6).

### 1.2.2 Meteorological forecasts

Daily seasonal precipitation forecasts from ECMWF System 4 were used. System 4 is an ocean-atmosphere general circulation model that couples the NEMO ocean model and the IFS cycle 36r4 atmosphere model. System 4 provides a 51-member forecast ensemble for the next seven months at a TL255 (about 0.7°) spatial resolution (Molteni *et al.*, 2011). ECMWF retrospectively produced forecasts for the period running from 1981 to 2010. These are composed of 51 ensemble members for February, May, August and November, and 15 members for the other months. The first member of the ensemble is issued with unperturbed (control) initial conditions, which are obtained from reanalysis of atmosphere and ocean observations and from climatology. The other forecast members are obtained from as many perturbed oceanic and atmospheric initial conditions as necessary, but also from stochastic physics. Even though the role of stochastic physics prevails in the spread of the forecasts at the seasonal time scale, initial conditions remain important to well

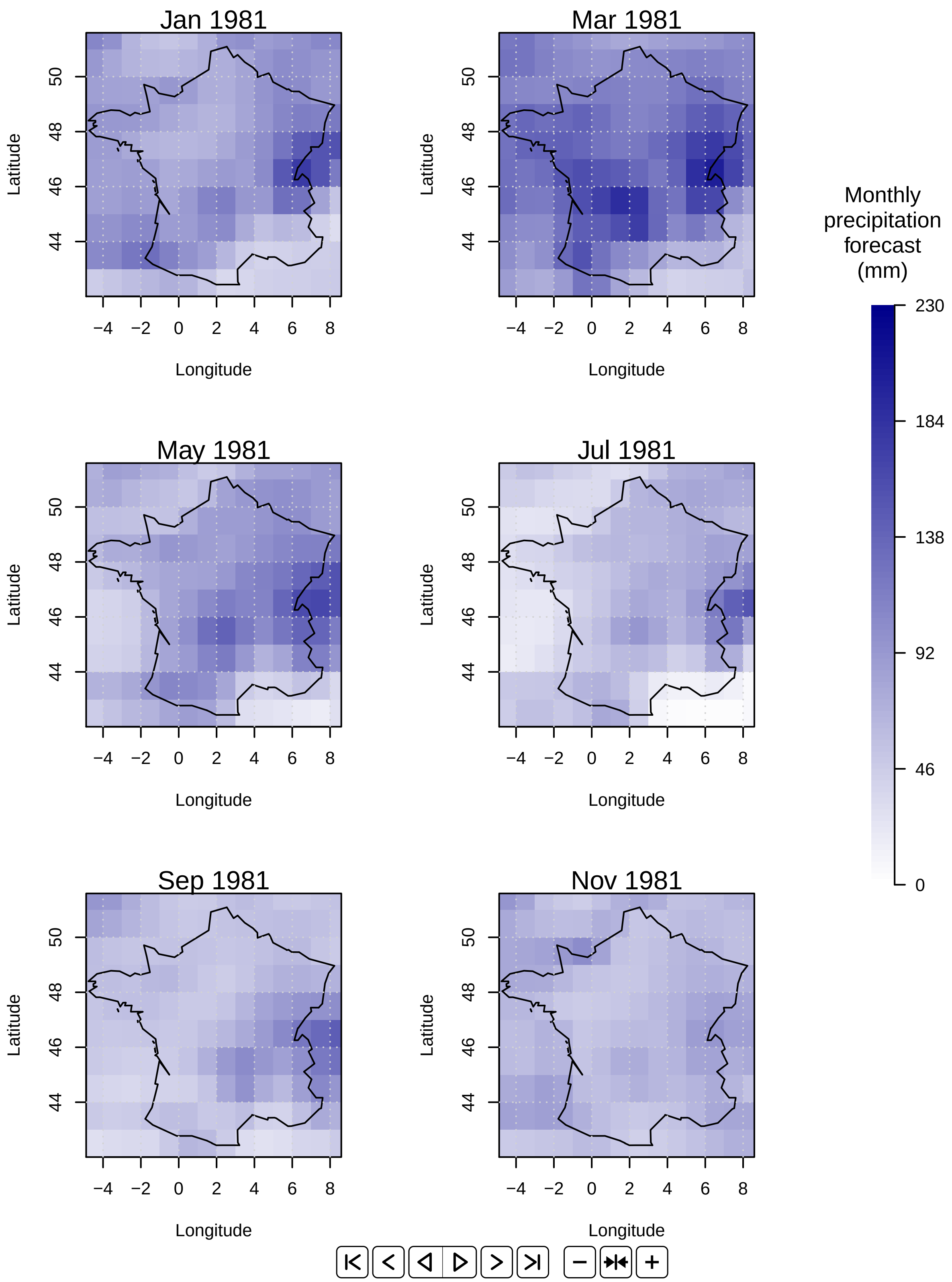



Figure 1.7: Monthly precipitation forecasts at a one-month lead from the ensemble mean of ECMWF System 4, from 1981 to 2010. Forecasts are presented for the months of January, March, May, July, September and November. Note that this figure is an animation that displays in the PDF version of this thesis. The printed version only shows the forecasts of 1981.

reproduce the dynamics in the first lead times.

Figure 1.7 gives an overview of the monthly precipitation forecasts issued by System 4 for the month ahead and over France, from 1981 to 2010. In this figure, the precipitation forecast ensemble mean is summed over the month following the forecast date. For the purpose of this PhD work, the 1981-2010 forecasts were aggregated at the catchment scale. Only the first 90 days of the forecast horizon were considered.

Weisheimer and Palmer (2014) evaluated the reliability of the precipitation forecasts issued by ECMWF System 4 on a scale ranging from “dangerous” to “perfect”. Over the world, precipitation forecasts often fell within the “marginally useful” category. In France, they were ranked as “marginally useful” during wet winters and summers, “not useful” in dry winters, and “dangerous” in dry summers. Kim *et al.* (2012) also evaluated the skill of System 4 precipitation and temperature at the global scale. Despite good overall performances, they identified systematic biases, e.g. a warm bias in the North Atlantic. Several studies have proposed to bias correct ECMWF System 4 precipitation forecasts in different contexts. Di Giuseppe *et al.* (2013) applied a spatially-based precipitation bias correction to improve malaria forecasts. Trambauer *et al.* (2015) applied a linear scaling method to forecast hydrological droughts in Southern Africa. In the same context, Wetterhall *et al.* (2015) applied a quantile mapping method to daily precipitation values, and showed that bias correction was able to improve the skill of the system to forecast dry spells. Bias correction of System 4 precipitation seasonal forecasts will be the topic of Chapter 4 of this thesis.

## 1.3 Conclusion

In this chapter, we first presented the catchments and hydrological data used in this thesis. A set of 16 catchments spread over France and used to draw general conclusions on the quality of seasonal streamflow forecasting in France (see Chapters 4 and 5), was introduced. In these catchments, 36 to 52 years of hydrological data were available from the French national HYDRO database. We have seen that the selected catchments are dominated by a pluvial regime, with high flows typically observed during the rainy season between November and May, and low flows observed during the dry season, between May and October. We have also seen that these catchments are well within the range of French catchments in terms of water balance, and that no atypical behaviours, such as unexplained water gain or water leakage, were observed.

The catchment of the Vilaine river at Rieux (Brittany, France) was also presented. This catchment is used in Chapter 6 of the thesis to study the inflows to the reservoir of the Arzal dam, located at the outlet of the catchment. The hydrological data for this catchment come directly from the managers of the Arzal dam. This catchment presents a strong intra-annual streamflow variability, and its lowest flows are observed in the months of August and September.

The observed and forecast meteorological data used in this thesis are presented. Most meteorological observations originate from the SAFRAN reanalysis provided by Météo-France. Seasonal precipitation forecasts derive from the ECMWF System 4, which is a coupled ocean-atmosphere model.

The hydrometeorological data presented in this chapter are used to simulate and forecast streamflow in the studied catchments, and evaluate model performances. In the following chapter, we present the hydrological model used throughout this thesis, as well as the procedures adopted to calibrate and validate the model, and forecast streamflows.

# 2

# Hydrological modelling and forecasting

## 2.1 Introduction

Hydrological models reproduce the rainfall-runoff transformation based on a set of equations. Chow *et al.* (1988) classify hydrological models based on their accounting for randomness in output variables and their accounting for variations in space and time. Following the first criterion, we distinguish between models with which similar inputs will always yield the same outputs (deterministic models), and models whose outputs result, at least partially, from a random process (stochastic models). Deterministic models are then subdivided depending on whether their outputs are uniform (lumped) or vary in space (distributed), and stochastic models are divided depending on whether they are space-correlated or space-independent. Similarly, deterministic models are subdivided into models whose outputs are constant and models whose outputs vary with time. Stochastic models are subdivided into models that are time-correlated and time-independent.

Just like the response of a catchment depends on its physical characteristics, the response of a hydrological model, either deterministic or stochastic, is determined by a parameter set. The parameters can be either fixed, if they are space- and time-independent, or left free. Prior to streamflow simulation or forecasting, it is necessary to set the free model parameters via a calibration step. The objective of model calibration is to identify a parameter set that not only minimizes an objective function (i.e., a chosen distance between observed streamflows and simulated streamflows obtained from observed inputs) but also is hydrologically realistic (Andréassian *et al.*, 2012). Once a parameter set is identified, its validity must be assessed based on an evaluation distance metric, that can either be the objective function, or another metric that focuses on other simulation attributes. This assessment is carried out over an evaluation period that is ideally independent from the calibration period (Klemes, 1986; Rykiel Jr., 1996; Refsgaard and Henriksen, 2004). Numerous optimisation and evaluation techniques, either based on numerical distance metrics or visual inspection, exist to calibrate and validate hydrological models (ASCE, 1993). The choice of the calibration and validation strategies should thus be guided by the modelling objectives.

Once a hydrological model has been calibrated and validated, it can be used to either simulate or forecast streamflow. In hydrology, simulation consists in using past meteorological observations to reproduce past streamflows by means of a hydrological model. It is a retrospective approach used, for example, to reconstruct streamflow time series or gain insight into the physical phenomena in action in a catchment. By contrast, forecasting consists in making informed assumptions on flows that have not yet been observed. In this configuration, meteorological and hydrometric observations are only available up to the time of the forecast, and predictions are made for future lead times.

Many studies advocate for probabilistic forecasting and have shown the value of these forecasts for decision-making and operational purposes (see e.g. Ramos *et al.*, 2010, 2013). Probabilistic forecasting proposes to issue forecasts that account for all sources of uncertainties and provide, for each lead time, a forecast probability density function rather than a single value (Krzysztofowicz, 2001). Sources of uncertainty in hydrological forecasting include meteorological forcings, initial hydrological conditions, the structure of the hydrological model and its parameters. Taking into account all these sources of uncertainties in streamflow forecasting can be fastidious and computationally costly. In seasonal hydrological forecasting, several studies have investigated the relative role of initial hydrological conditions and meteorological forcings to better understand how the different sources of uncertainty influence the quality of the final predictions (Shukla *et al.*, 2013; Yossef *et al.*, 2013; Wood *et al.*, 2016). They showed that the relative importance of these sources of uncertainty will depend on several factors such as the lead time, the study area or the season. Shukla *et al.* (2013) showed that, in the Northern hemisphere, initial conditions can prevail over meteorological forcings in snow-dominated regions. In France, Singla *et al.* (2012) also found a predominance of initial conditions in snow-dominated catchments, as opposed to catchments located in plains, in which the hydrologic predictability mainly depended on meteorological forcings.

The catchments used in this thesis (see Chapter 1) are not influenced by snow. Therefore, we focus on improving the assessment of uncertainties in meteorological forcings for seasonal streamflow forecasting. These meteorological forcings were then used as input to the GR6J hydrological model. In the next section, we present the model, the results of its calibration and validation as well as the forecasting framework adopted in this thesis.

## 2.2 The GR6J model

The GR (Génie Rural) models are lumped, conceptual rainfall-runoff models developed at Irstea. These models represent the rainfall-runoff processes in a simplified way by means of a series of storages and a delay function. Their inputs are precipitation and potential evapotranspiration time series, averaged over the catchment area. Their output is the corresponding streamflow time series at the catchment outlet.

There are several GR models at Irstea. The structure of each model varies with its purpose and its time step of application, which ranges from the hourly time step to the yearly time step. They all are characterized by a low number of parameters, ranging from one to six parameters, to be calibrated against observed streamflows. This is considered an asset for operational applications and research studies based on large catchment datasets. Examples of some of the GR models include:

- the GRP model (Tangara, 2005; Berthet, 2010) that runs at the hourly time step for forecasting purposes. It has three free parameters and is currently used operationally in 14 out of 19 French flood forecasting centres (*Services de Prévision des Crues*);
- the GR4J model (Perrin *et al.*, 2003) that runs at the daily time step and has four free parameters. This model has been used by several studies in France and worldwide;
- the GR5J model (Le Moine, 2008), which is based on the GR4J model but also accounts for catchment leakages (exchanges with groundwater).

The daily GR6J model was developed with the objective to further improve the GR models efficiency on low flows. Pushpalatha *et al.* (2011) tested 60 modified versions of the GR5J model on 1000 unregulated French catchments. The conclusion of the study led to the GR6J model: a version of GR5J with an exponential store added in parallel to the routing store to represent the contribution of pluriannual aquifers. Improvement in performance was observed in both high and low flows.

The structure of GR6J is presented in Figure 2.1. In the model, input rainfall ($P$) is first neutralized by input potential evapotranspiration ($PE$). Part of the resulting net rainfall ($P_n$) or net evapotranspiration ($E_n$), depending on whether rainfall is greater than evapotranspiration, are used to update the production store. Water then percolates from this store ($P_{erc}$) and is added to the part of rainfall that directly transforms into surface runoff ($P_n - P_s$). The resulting amount of water ($P_r$) is then routed through two unit hydrographs. Part of the delayed flow directly contributes to the catchment outlet ($Q_1$), but most of it ($Q_9$) feeds the routing and exponential stores which govern the propagation of the flow to the catchment outlet ($Q_{r1}$ and $Q_{r2}$). Before the flow reaches the catchment outlet, the intercatchment groundwater flow acts on the water in the routing store and the water that transits directly from the unit hydrograph to the catchment outlet. The resulting streamflow ($Q$) is the sum of all partial outputs.

The fixed parameters of the GR6J model intervene in the division of the water flow between the two unit hydrographs, and, later, in the allocation of water between the exponential and the routing stores. A fixed parameter also governs the relative base time of the unit hydrographs. Six other parameters are left free:

- $X_1$ ($mm$): the capacity of the production store,
- $X_2$ ($mm/day$): the maximum intercatchment groundwater exchange,
- $X_3$ ($mm$): the capacity of the routing store,
- $X_4$ ($day$): the base time of the unit hydrographs,

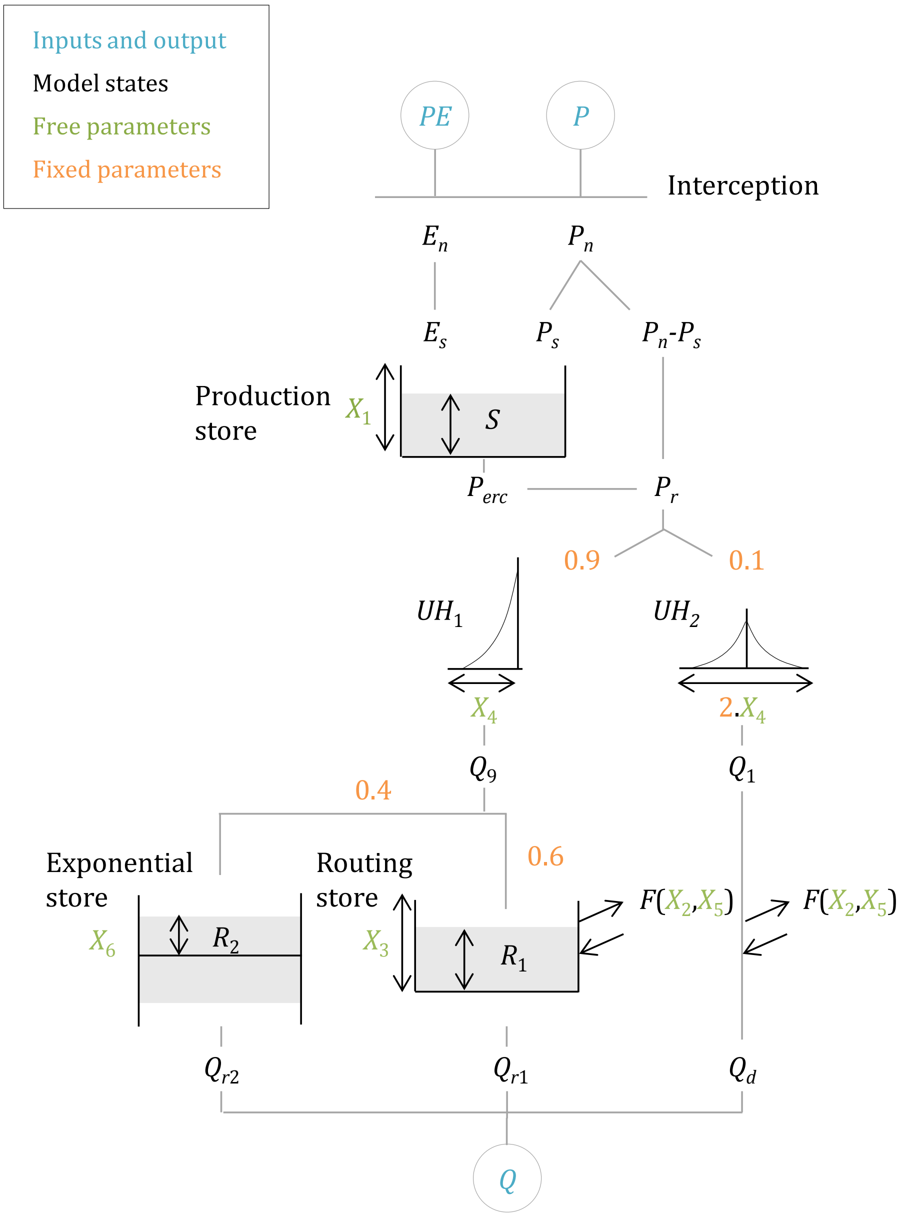



Figure 2.1: Structure of the GR6J model (modified from Pushpalatha *et al.*, 2011). The model is composed of three reservoirs represented by the grey-shaded areas and two delay functions represented by the curves. Model states are in black, free parameters are in green, fixed parameters are in orange and inputs and outputs are in blue.

- $X_5$ (-): the threshold of inversion of groundwater exchanges,
- $X_6$ ($mm$): the output coefficient of the exponential store.

These six parameters are subject to calibration and validation prior to running the model.

## 2.3 Calibration and validation of the GR6J model

The six free parameters of the GR6J model were calibrated and validated with the one-year-leave-out method (Figure 2.2). This method allows to take full advantage of the available observations and provides a robust calibration. With this method, a parameter set is identified for each year of the record period by optimizing a target score on all other available years. The parameter set is then transferred to run the model for the target application year. The procedure is repeated to cover all years in the validation. In each catchment of the countrywide catchment set, 30 parameter sets were obtained for the period running from 1981 to 2010. In the Vilaine river catchment, 10 parameter sets were obtained for the period running from 2003 to 2012.

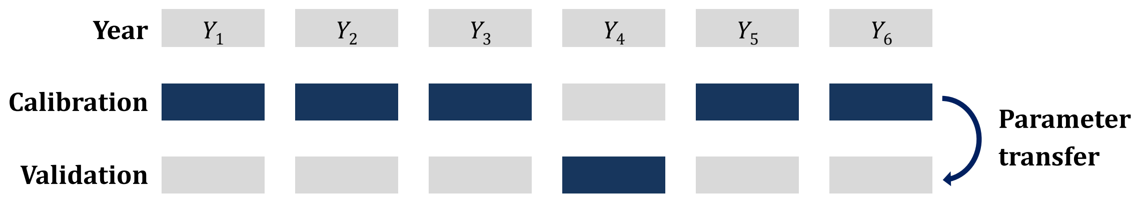



Figure 2.2: Schema of the one-year-leave-out method for calibration and validation. If the model is calibrated to simulate or forecast streamflows for year $Y_4$, all other years, i.e. $Y_1$, $Y_2$, $Y_3$, $Y_5$ and $Y_6$, are used in the calibration procedure. The calibrated parameter set is then transferred to run the model for $Y_4$ in validation.

Parameters were calibrated for each studied catchment by forcing GR6J with precipitation and evapotranspiration from the SAFRAN meteorological reanalysis (see Chapter 1). The metrics used to minimize the distance between simulated and observed streamflows derive from the Kling-Gupta efficiency (KGE; Gupta *et al.*, 2009), which evaluates simultaneously the bias, the error in standard deviation and the correlation between the observations and the simulation. It is defined by:

$$\mathrm{KGE} = 1 - \sqrt{(\frac{\sigma_{\hat{Q}}}{\sigma_Q} - 1)^2 + (\frac{1}{\mathrm{Bias}} - 1)^2 + (r-1)^2} \tag{2.1}$$

where $\sigma_Q$ is the standard deviation of the observed streamflows, $\sigma_{\hat{Q}}$ is the standard deviation of the simulated streamflows, $r$ is the Pearson correlation coefficient between the observation and the simulation and Bias represents the error in mean volume. The values taken by the

KGE criterion fall within the $(-\infty, 1]$ interval, with 1 corresponding to a perfect simulation. In the formulation of the KGE and throughout the thesis, the Bias is formulated as:

$$\text{Bias} = \frac{\sum_{i=1}^{n} Q_i}{\sum_{i=1}^{n} \hat{Q}_i} \tag{2.2}$$

where $Q$ and $\hat{Q}$ are the observed and simulated streamflow, and $n$ is the number of time steps used for the evaluation. A simulation that is not biased will score 1. A Bias higher (lower) than 1 indicates that the mean observation is underestimated (overestimated).

In this thesis, two versions of the KGE derived from different flow transforms were used as objective function to calibrate the GR6J model: the KGE applied to inverse flows ($\text{KGE}_{iQ}$; i.e. the KGE criterion computed for the time series of $\frac{1}{Q}$ instead of the time series of $Q$) and the KGE applied to root-squared flows ($\text{KGE}_{rQ}$). The inverse flow transform puts the focus on the lowest flows of the hydrographs and is thus preferred to study low flows, whereas the squared flow transform focuses on the medium to high flows of the hydrographs (Pushpalatha *et al.*, 2012).

In the validation step, we evaluate the simulations obtained for each year of the available record period and each catchment by juxtaposing them and evaluating a single simulation over the record period. The simulations were evaluated with the $\text{KGE}_{rQ}$ and $\text{KGE}_{iQ}$ criteria, but also with a normalized Nash-Sutcliffe efficiency and the Bias (Equation 2.2) to evaluate the simulations with criteria that are independent from the calibration criteria. The normalized Nash-Sutcliffe efficiency ($\text{C2M}_Q$; Mathevet *et al.*, 2006) evaluates the quadratic error between the simulation and the observations, and compares the evaluated model with a simple benchmark: the constant, long-term mean of the observed streamflows. It is formulated as follows:

$$\text{C2M}_Q = \frac{\sum_{i=1}^{n} (Q_i - \bar{Q})^2 - \sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \bar{Q})^2 + \sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2} \tag{2.3}$$

where $Q$ is the observed streamflow, $\hat{Q}$ is the simulated streamflow, $\bar{Q}$ is the mean observed streamflow and $n$ is the number of time steps used for the evaluation. $\text{C2M}_Q$ values fall within the $[-1, 1]$ interval.

Figure 2.3 presents the ranges of parameter values obtained with the one-year-leave-out calibration procedure. Each graph corresponds to a parameter of the GR6J model. Catchments are represented on the x-axis and variations in parameter values are represented by boxplots on the y-axis (in blue when $\text{KGE}_{rQ}$ is used for calibration and in red when $\text{KGE}_{iQ}$ is used). We observe that the parameter sets are usually consistent over the years, and variations between catchments are greater than variations between application years. However, one catchment stands out with a large variability of its calibrated parameters: catchment 13, which is the southernmost catchment (Figure 1.1). The smallest catchment, i.e. catchment 1, also stands out with large capacities of its production and routing stores. Apart from these two catchments, the parameter values are similar in all catchments, including in the Vilaine at Rieux catchment.

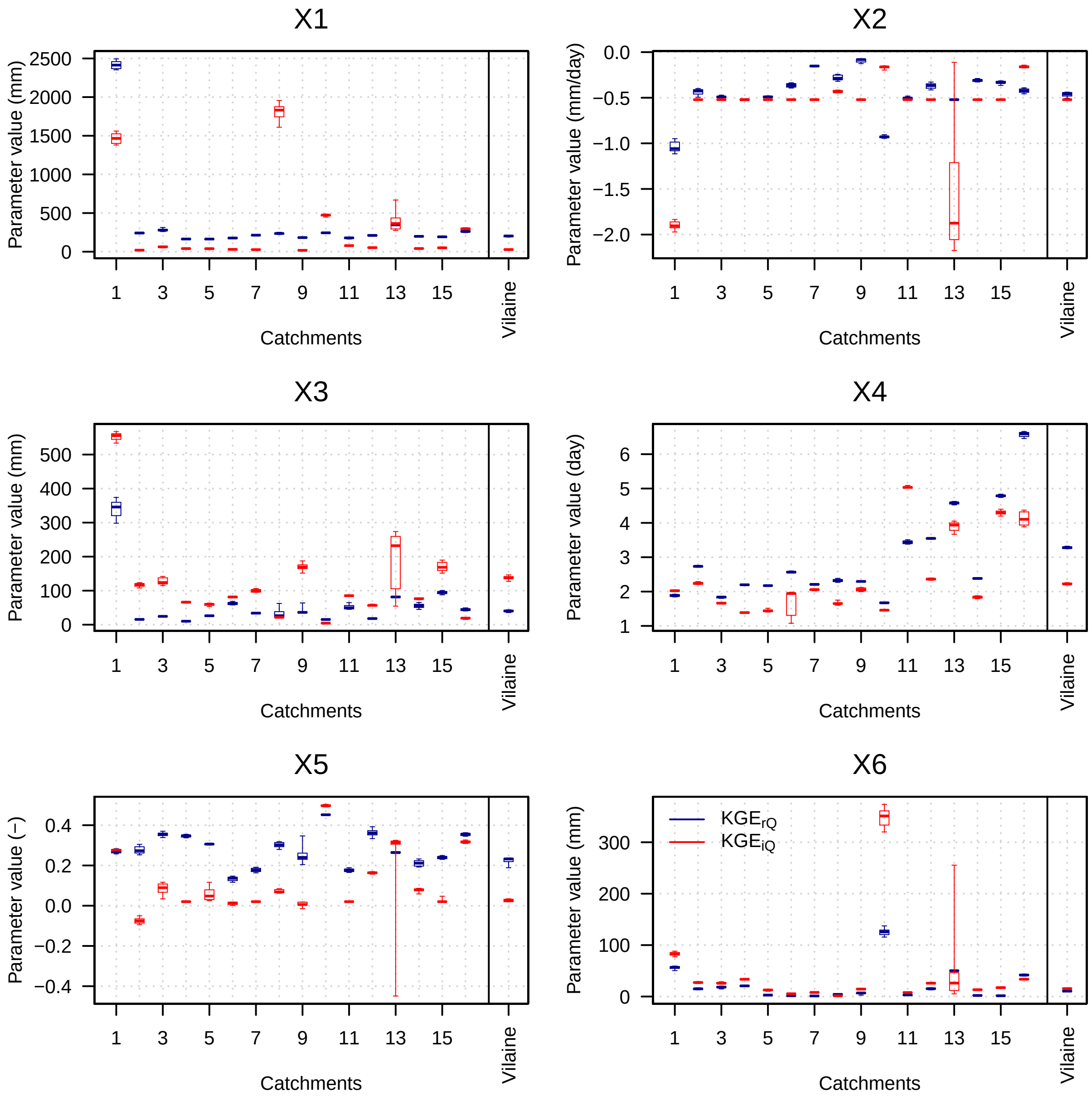



Figure 2.3: Ranges of parameter values obtained after calibration with $\mathrm{KGE}_{rQ}$ (in blue) and $\mathrm{KGE}_{iQ}$ (in red) with a one-year-leave-out procedure. Each graph corresponds to a parameter of the GR6J model. Catchments are represented on the x-axis, with catchments from the countrywide set numbered from 1 to 16 (from the smallest to the largest) and the Vilaine at Rieux catchment represented on the far right (see Figure 1.1 for catchment location in France). Variations in parameter values are represented on the y-axis by boxplots ($10^{\mathrm{th}}$ , $25^{\mathrm{th}}$ , $50^{\mathrm{th}}$ , $75^{\mathrm{th}}$ and $90^{\mathrm{th}}$ percentiles). For the countrywide catchment set, the boxplots are composed of 30 parameter values obtained for the period running from 1981 to 2010. For the Vilaine at Rieux, they are composed of 10 parameter values obtained over the period running from 2003 to 2012.

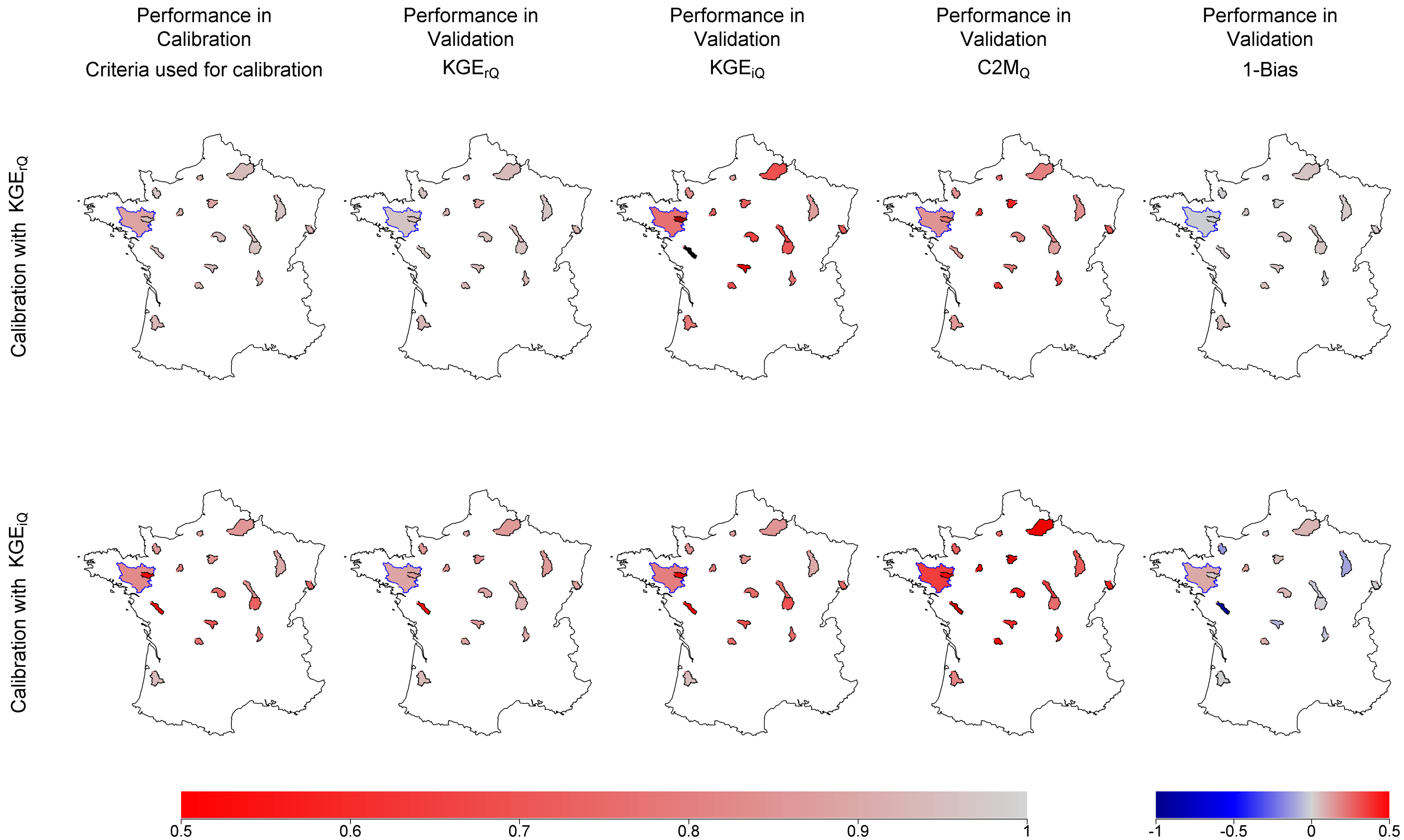



Figure 2.4: Maps of performance of GR6J in calibration and validation in the 17 studied catchments. Rows correspond to performance obtained with KGE$_{rQ}$ (top) and KGE$_{iQ}$ (bottom). The first column presents the average performance in calibration. The other four columns present the performance in validation as evaluated by KGE$_{rQ}$ (second column), KGE$_{iQ}$ (third column), C2M$_Q$ (fourth column) and the Bias (last column). Grey colours indicate good performance and red (or blue) colours indicate poorer performance. In the case of the Bias, blue (red) colours indicate an underestimation (overestimation) of streamflow.

Figure 2.4 shows the performance of GR6J in calibration and in validation in the 17 studied catchments. The performance in calibration is the mean calibration score optimized for each target application year. The performance in validation is the score obtained when evaluating the series of simulations obtained in validation. The first and second rows respectively correspond to the performances obtained after calibrating GR6J with $\mathrm{KGE}_{rQ}$ and $\mathrm{KGE}_{iQ}$. The first column shows the performances obtained in calibration as evaluated by the objective function. The other columns show the performances in validation as evaluated by $\mathrm{KGE}_{rQ}$, $\mathrm{KGE}_{iQ}$, $\mathrm{C2M}_{Q}$ and the Bias.

We observe good to excellent performances in calibration, with objective scores ranging between 0.88 and 0.97 when calibrating with $\mathrm{KGE}_{rQ}$, and between 0.46 and 0.94 when calibrating with $\mathrm{KGE}_{iQ}$. Interestingly, very good performances were obtained in catchments 1 and 13 even though they stood out in terms of variability of parameter values (cf. Figure 2.3). Performances obtained in validation with the criterion used in calibration (i.e., simulations calibrated with $\mathrm{KGE}_{rQ}$ and evaluated with $\mathrm{KGE}_{rQ}$, and simulations calibrated with $\mathrm{KGE}_{iQ}$ and evaluated with $\mathrm{KGE}_{iQ}$) are also very satisfying with performances very similar to the ones obtained in calibration. As expected, simulations calibrated with $\mathrm{KGE}_{iQ}$ perform better than simulations calibrated with $\mathrm{KGE}_{rQ}$ to simulate low flows. While the calibration with $\mathrm{KGE}_{iQ}$ yields $\mathrm{KGE}_{iQ}$ values between 0.41 and 0.94, the calibration with $\mathrm{KGE}_{rQ}$ yields $\mathrm{KGE}_{iQ}$ values in validation between 0.4 and 0.89 except in two catchments where scores of 0.11 and -0.02 are reached (in catchments 6 and 8). The calibration with $\mathrm{KGE}_{iQ}$ thus yields significant improvement to simulate low flows in these catchments. Calibrating GR6J on low flows (i.e. with $\mathrm{KGE}_{iQ}$ as objective function) still provides good performances in medium flows with $\mathrm{KGE}_{rQ}$ values always above 0.54 in all catchments in validation. Also, performances in $\mathrm{C2M}_{Q}$, which focuses on medium to high flows, are good to very good with values always superior to 0.58 when calibration is performed with $\mathrm{KGE}_{rQ}$, and between 0.28 and 0.81 when calibration is performed with $\mathrm{KGE}_{iQ}$. Lastly, the Bias in validation falls between -0.02 and 0.04 when GR6J is calibrated with $\mathrm{KGE}_{rQ}$, which is close to no bias. When the model is calibrated with $\mathrm{KGE}_{iQ}$, Bias values fall between -0.14 and 0.12, except in catchment 8, where a value of -0.94 is reached.

Figure 2.5 shows the observed and simulated hydrographs for the Vilaine at Rieux (the largest catchment in Figure 2.4, cf. Figure 1.1) when GR6J is calibrated with $\mathrm{KGE}_{iQ}$. The hydrographs were limited to the May to October period and allow for a visual validation of GR6J during low flows and recessions in the Vilaine river. In this catchment, GR6J gave very good numerical results, with a $\mathrm{KGE}_{rQ}$ of 0.90, a $\mathrm{KGE}_{iQ}$ of 0.76, a $\mathrm{C2M}_{Q}$ of 0.63 and a Bias of -0.04 in validation. We observe that the hydrological model succeeds in reproducing the general trend of the streamflow, including recessions and peak flows. We also observe that observations are much noisier than simulations during low flows. Two explanations are possible: either the frequent observed variations are actual signals and thus the hydrological model is not sufficiently reactive, or the variations are caused by noise in the observations, which the model does not simulate. Discussions with the managers of the Arzal dam, who are also in charge of the measurement station, supported the second hypothesis. Indeed, measuring devices tend to fail during low flows, which causes unreliable variations in the observed streamflow data (observation noises).

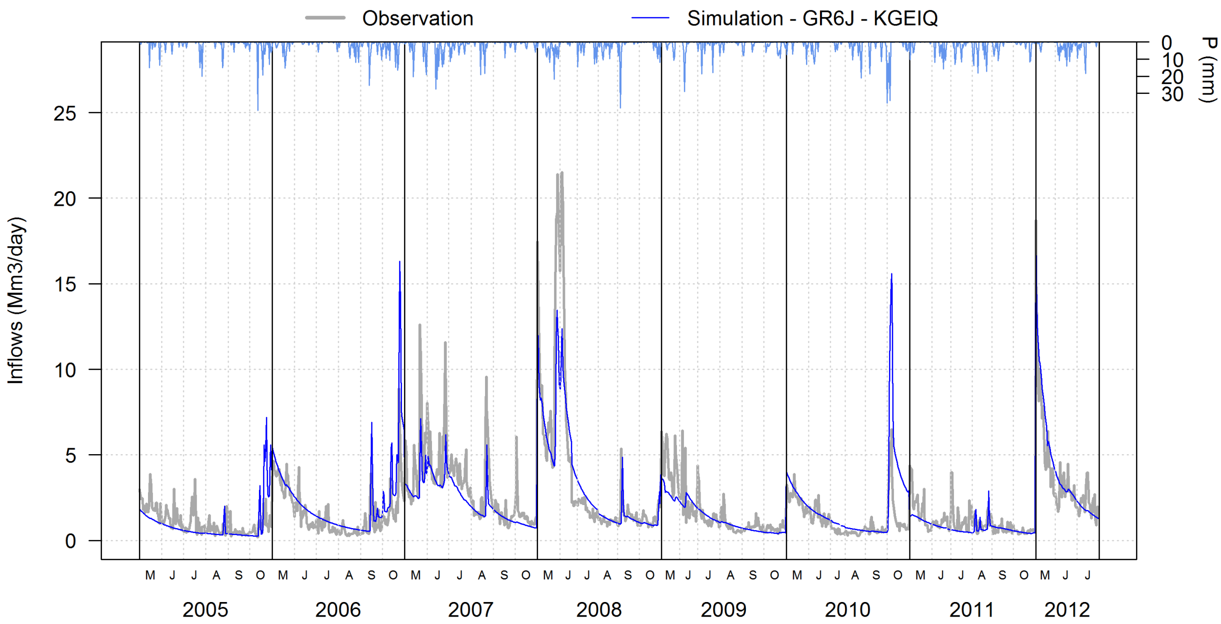



Figure 2.5: Observed and simulated streamflow and observed precipitations for the Vilaine river at Rieux, between May and October, from 2005 to 2012. Observed streamflow is represented in grey, simulated streamflow in dark blue, and precipitation in light blue.

## 2.4 Using the GR6J model for streamflow forecasting

In this thesis, the GR6J model is used to forecast streamflow in the catchments presented in Chapter 1. The forecasting procedure can be divided into two steps: the initialisation of the model states and the input of meteorological forecasts to run the model in forecasting mode. Figure 2.6 summarizes how a streamflow forecast is issued with the GR6J model, from a precipitation forecast scenario.

The initialisation of the model consists in estimating adequate values for the states of the GR6J model at the time of the forecast. In this thesis, these states are initialised by running the model with observed inputs for a year up to the time of the forecast. To allow for a better representation of the initial conditions at the time of the forecast, the model is then updated based on the last observed streamflow value. To that effect, the levels in the routing and exponential stores are corrected based on the inverted reservoir balance equations.

Once the model states are initialized and updated, the GR6J model is fed with me-

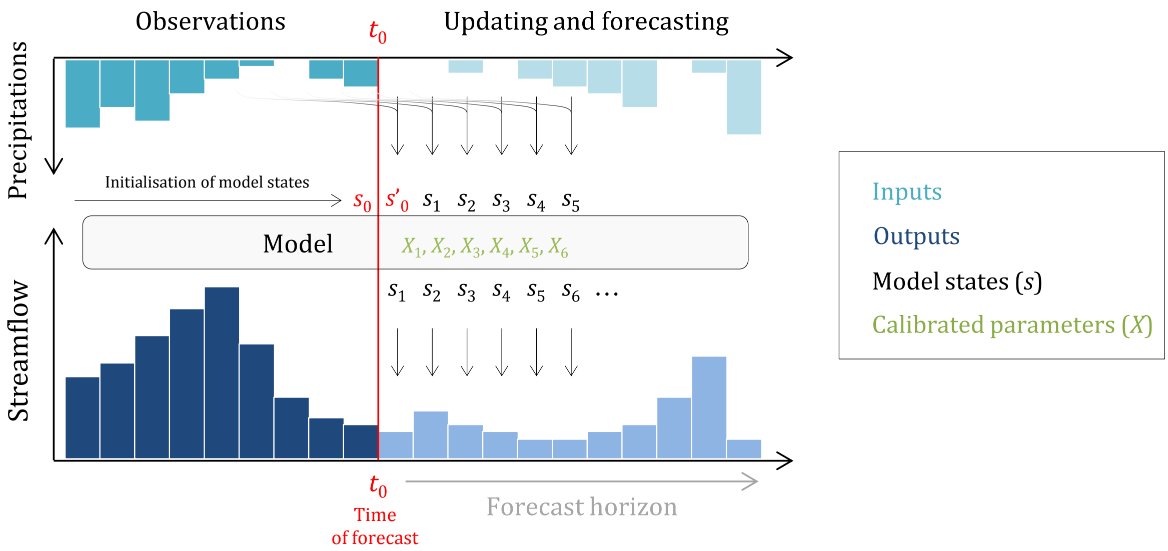



Figure 2.6: Schema of the method used to forecast streamflows with the GR6J model. Model states are initialized using previous observations of model inputs and calibrated parameters. The model is run up to the time of forecast $t_0$ and we obtain the model states $s_0$. Before issuing a forecast, the last observed streamflow is used to update model states ($s'_0$). The precipitation forecast is then used as input to the hydrological model. The model is run with updated states and previously calibrated parameters. The forecast streamflow is obtained for each lead time of the forecast horizon. The procedure is repeated for each input forecast. In this study, GR6J updating only concerns the updating of the states of the routing and the exponential stores.

teorological inputs (precipitation and potential evapotranspiration) and run in forecasting mode. We then obtain streamflow forecasts and the procedure is repeated for each forecast inputs. In this thesis, the focus is placed on the influence of precipitation inputs on streamflow forecasting. The model is fed with interannual potential evapotranspiration to have a standard reference scenario for potential evapotranspiration. Since the selected catchments are not influenced by snow, this set up is not expected to have a major impact on water volumes. However, it may underestimate water losses due to evapotranspiration in years with temperatures well above the climatological average.

The uncertainty in precipitation forcings is represented with ensemble forecasts. Here, we use precipitation ensembles composed of a finite number of scenarios (members), each member corresponding to a possible outcome. Furthermore, all members are assumed to be equiprobable, meaning that they are equally likely to be the observed outcome. Each member of the precipitation ensemble is used as input to the GR6J hydrological model to produce a possible streamflow outcome. The resulting streamflow scenarios constitute an ensemble streamflow forecast composed of as many members as the input precipitation ensemble. Because not all sources of uncertainty are taken into account in the resulting streamflow forecasts, we refer to the method as ensemble forecasting rather than probabilistic forecasting (Krzysztofowicz, 2001).

## 2.5 Conclusion

In this chapter, we presented the GR6J hydrological model that is used throughout this thesis for streamflow forecasting. The structure of the model and its parameters were first detailed. The model was then calibrated and validated in the studied catchments (see Chapter 1) by means of the one-year-leave-out procedure and with two variants of the Kling-Gupta efficiency (KGE): the KGE applied to root-square flows and the KGE applied to inverse flows. Parameter sets obtained with this procedure were consistent over the calibration period in most catchments. The performance of the model in validation were overall satisfactory. We have seen that calibrating GR6J with the KGE applied to inverse flows rather than the KGE applied to root-square flows provided better performances in low flows in some of the catchments and shall be preferred to model low flows in our catchment set. In addition, the calibration with $\mathrm{KGE}_{iQ}$ also provides good performances in medium flows as evaluated by $\mathrm{KGE}_{rQ}$. Lastly, we have described how we use the GR6J model to issue streamflow forecasts from initial hydrological conditions and meteorological forecast inputs.

In the following chapter, we present the evaluation criteria used in this thesis to assess the quality of the streamflow forecasts obtained with the GR6J model with the procedure described in this chapter, as well as the quality of the meteorological forecast inputs.

# 3

# Forecast evaluation

## 3.1 Introduction

Once a forecasting system is set up, it is essential to evaluate its performance and limitations. Indeed, forecast evaluation fosters a critical approach to the forecast system outputs, but also provides a diagnostic of the deficiencies of a forecast system to guide further improvement. Forecast quality comprises many forecast attributes. Jolliffe and Stephenson (2003) distinguishes between evaluation approaches that assess the capacity to:

- forecast binary events (also named yes/no events). In this case, the range of possible outcomes is divided in two exclusive and exhaustive categories corresponding to the occurrence of the event or its non-occurrence, as is the case of events defined by a threshold;
- forecast categorical events, which are an extension of binary events. In this case, the range of possible outcomes is divided in a finite number of categories, superior to two;
- forecast continuous variables.

Regardless of the approach, a thorough forecast evaluation is done retrospectively over an extended time period. For instance, meteorological centres usually run newly-developed models over extended past records to evaluate the new system and compare it with previous versions. These retrospective forecasts, often named hindcasts, are thus essential to evaluate the performance of the new forecast system.

In general, an evaluation framework should assess forecast quality in regard to several attributes, and, ideally, combine visual and numerical evaluations (Chiew and McMahon, 1993; Krause *et al.*, 2005; Crochemore *et al.*, 2015a[1]). A wide range of evaluation metrics exist to evaluate hydrological models and numerical predictions. Since each metric focuses

[1]This paper shows the results of my Master work whose publication was finalized during my PhD thesis: Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S.P., Grimaldi, S., Gupta, H., Paturel, J.-E., 2015. Comparing expert judgement and numerical criteria for hydrograph evaluation. Hydrological Sciences Journal 60, 402–423.

on different attributes of model outputs (Jachner *et al.*, 2007), the objective of the evaluation should guide the choice of a set of metrics (Krause *et al.*, 2005; Pushpalatha *et al.*, 2012). Once a set of metrics is chosen, the performance of a forecast system can be assessed independently from any reference, i.e. in absolute terms, or comparatively to another forecast system, i.e. in relative terms. In the latter case, the choice of the benchmark is crucial in order not to inflate the performance of a system by using a too naïve reference or to deflate its performance by choosing a too demanding reference (Perrin *et al.*, 2006; Pappenberger *et al.*, 2015).

In this thesis, we evaluated the performance of forecasting systems considering continuous variables and binary events. The quality of the forecasts in regard to continuous variables was assessed by looking at the accuracy, the reliability, the sharpness and the overall performance of the forecast systems. The quality of the forecasts in regard to binary events was assessed by looking at the discrimination attribute. Each of these forecast attributes was evaluated by means of different evaluation scores. The performance of the forecast systems was also assessed relatively to chosen reference forecasts. In this chapter, the forecast attributes used in this thesis, as well as the evaluation criteria used to assess these attributes are first presented. We then detail the skill scores used to compare forecast systems.

## 3.2 Evaluation scores

### 3.2.1 Accuracy

The accuracy of a forecast (or simulation) corresponds to its distance to the observation. In the case of a probabilistic or an ensemble forecasting system, the forecast mean, or sometimes the forecast median, is used in place of the deterministic simulation. In this thesis, we assess the accuracy with the Mean Absolute Error (MAE) and the Root-Mean-Square Error (RMSE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Q_i - \hat{Q}_i| \tag{3.1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2} \tag{3.2}$$

where $Q$ is the observed streamflow (or precipitation), $\hat{Q}$ is the simulated streamflow (or precipitation) and $n$ is the number of time steps used for the evaluation. These metrics vary within $[0, \infty)$. They can be considered as distances: a value of 0 means that the simulation and the observation are identical.

### 3.2.2 Reliability

Reliability is a forecast attribute that refers to the statistical consistency between observed frequencies and forecast probabilities. It indicates whether a system can be trusted, i.e., given a forecast, where is my observation most likely to fall within the forecast range. In this thesis, reliability is evaluated with the Probability Integral Transform (PIT) diagram (Gneiting *et al.*, 2007; Laio and Tamea, 2007). The PIT diagram is the cumulative distribution of the positions of the observation within the cumulative forecast distribution (or PIT values). The interpretation of the PIT diagram is illustrated in Figure 3.1. A reliable forecast has a PIT diagram superposed with the 1:1 diagonal (the dark blue line in Figure 3.1). If the PIT diagram shows a curve systematically above (below) the diagonal (the blue lines in Figure 3.1), the observed values are too frequently located in the lower (upper) parts of the forecast distribution, suggesting a systematic bias of the forecasts towards overprediction (underprediction). If the points in the diagram are too concentrated in the vicinity of the end points (0 and 1), forecasts are over-confident and observations fall more frequently than expected on the tails of the forecast distribution. On the contrary, too many points concentrated in the midrange indicate a forecast distribution that is under-confident. In order to numerically compare results among catchments, we also computed the area between the curve of the PIT diagram and the 1:1 diagonal, as proposed by Renard *et al.* (2010). The PIT area thus ranges between 0 and 0.5. The smaller this area, the more reliable the ensemble is.

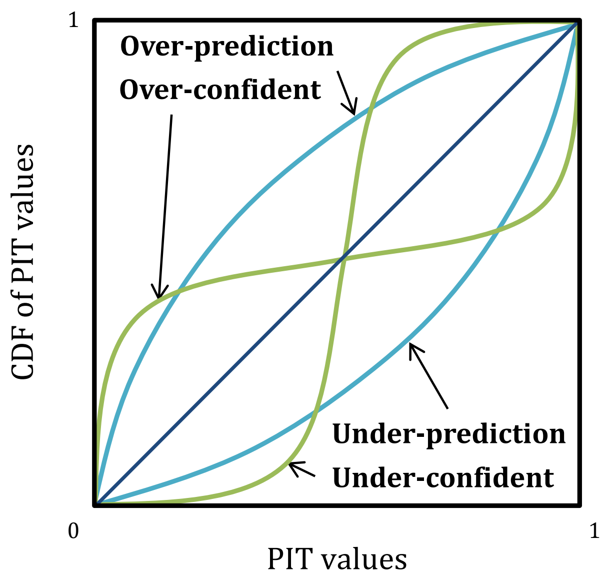



Figure 3.1: Illustration of how symptomatic shapes of the PIT diagram are interpreted (modified from Laio and Tamea, 2007; Bourgin, 2014).

### 3.2.3 Sharpness

Rather than a performance metric, sharpness is a property of the forecast system. It refers to the concentration of the predictive distribution and, in the case of an ensemble forecast, indicates how spread the members of the ensemble are. In this thesis, sharpness was evaluated with the 90 % interquartile range (IQR; Gneiting *et al.*, 2007), i.e., the difference between the 95$^{\text{th}}$ and the 5$^{\text{th}}$ percentiles of the forecast distribution averaged over the evaluation period. The narrower the IQR, the sharper the ensemble is. Gneiting *et al.* (2007) introduced the paradigm of "maximizing the sharpness of the predictive distributions subject to calibration". This paradigm relates sharpness to reliability (here referred to as "calibration"), i.e., sharpness can only become an evaluation criterion once reliability has been achieved. Given two reliable forecast systems, the sharpest one is preferred.

### 3.2.4 Overall performance

The Continuous Rank Probability Score (CRPS; Hersbach, 2000) assesses the overall performance of a forecast system. It evaluates the difference between the forecast distribution and that of the observation. It is defined by:

$$\text{CRPS} = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(\mathrm{F}_i^{\hat{Q}}(x) - \mathrm{F}_i^{Q}(x))^2 dx \tag{3.3}$$

where $\mathrm{F}^{\hat{Q}}$ is the cumulative distribution of the forecast, $\mathrm{F}^{Q}$ is the Heaviside step function corresponding to the observation and $n$ is the number of time steps used in the evaluation. The lower the CRPS is, the better the overall performance of the forecasts is. The green area in Figure 3.2 illustrates a term of the CRPS for a given forecast and corresponding observation. Note that the CRPS of a deterministic forecast corresponds to its MAE. Hersbach (2000) showed that the CRPS can be broken down into a reliability term, a resolution term and an uncertainty term. We do not apply this decomposition in this thesis, but we evaluate the CRPS as a whole to describe the overall performance of forecasting systems.

### 3.2.5 Discrimination

The Relative Operating Characteristics diagram (ROC; Mason and Graham, 1999) is often used to assess the capacity of a forecast system to discriminate between situations when an event is observed or not. The ROC diagram is plotted for a given threshold that is used to compute the Probability of Detection (POD) and the False Alarm Ratio (FAR). In order to build the diagram for an ensemble forecasting system, the proportion of ensemble members below the threshold (in the case of a low-flow event) necessary to trigger an alert varies from none to all ensemble members. For each proportion, the POD is plotted against the FAR. The ROC diagram is plotted for a given threshold, catchment and forecast lead time. The Area under the Curve (AUC) summarizes the ROC diagram into one numerical

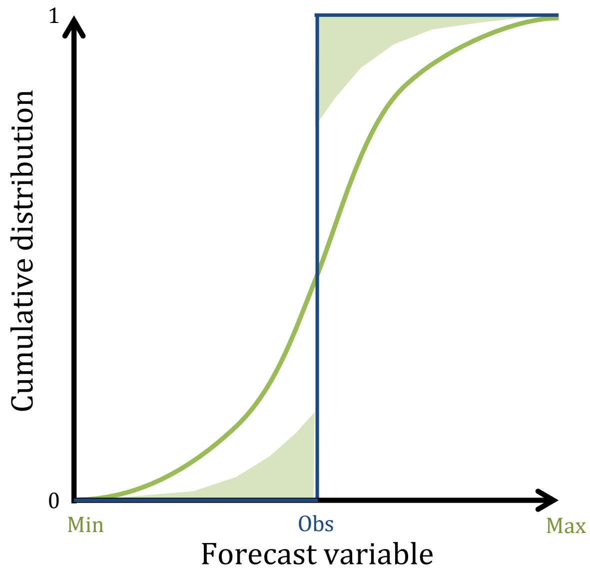



Figure 3.2: Illustration of the CRPS. The blue line represents the Heaviside step function of the observation. The green line represents the cumulative distribution of the forecast. The green area represents the area used in the computation of the CRPS (see Equation 3.3), for the displayed forecast and observation.

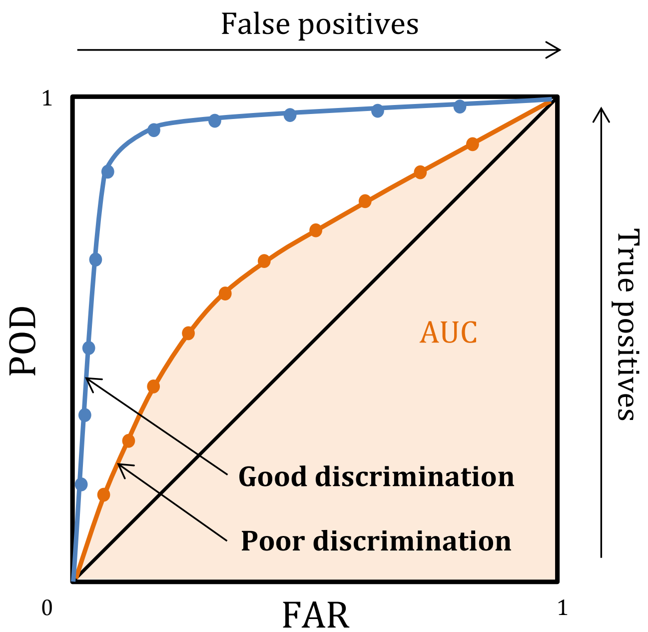



Figure 3.3: Illustration of the ROC diagram and AUC. The blue curve represents the ROC curve of a forecasting system with good discrimination. The orange curve presents the ROC curve of a forecasting system with poor discrimination. The AUC is illustrated for the orange curve and corresponds to the area under this curve.

value, which allows for easier comparisons between thresholds, catchments, lead times and forecast systems. The closer the AUC is to 1, the better the discrimination of the system is. An AUC value close to 0.5 shows no discrimination. Figure 3.3 illustrates the ROC diagram and the AUC.

## 3.3 Skill scores

Forecast skill is evaluated by comparing the performance of a given forecast system with the performance of a reference forecast. In this thesis, the skill scores were computed for some of the scores presented in the previous section, namely the MAE, the PIT area, the IQR and the CRPS. The skill score $SS$ is computed as follows:

$$SS_i = \frac{S_i^{\text{REF}} - S_i^{\text{SYS}}}{S_i^{\text{REF}}} \tag{3.4}$$

where $S$ corresponds to the chosen evaluation score, $S^{\text{REF}}$ is the score computed for the reference forecast, $S^{\text{SYS}}$ is the score computed for the forecast system and $i$ is the considered lead time. The skill score ranges within $(-\infty, 1]$. When it is superior to zero, the forecast system has skill with respect to the reference forecast. When the skill score is equal to zero, the system and the reference have equivalent performances with regard to the chosen evaluation score. A normalized version of the skill score can also be used:

$$S_i = \frac{S_i^{\text{REF}} - S_i^{\text{SYS}}}{S_i^{\text{REF}} + S_i^{\text{SYS}}} \tag{3.5}$$

This normalized formulation ranges within $]-1, 1]$ and is useful to visualize and compare the skill of a system in several catchments when performances strongly vary among catchments.

### 3.3.1 Using a reference forecast

The reference used in the computation of the skill score defines the scale used for evaluation. For instance, choosing a reference with poor performances will inflate the skill of the evaluated forecast system. Pappenberger *et al.* (2015) list a number of references used in streamflow forecasting. The authors investigated the impact of different references on the assessed performance of a flood forecasting system. Their study highlighted the need to clearly define the reference chosen for evaluation and to orientate this choice based on the application.

A first option is to compare the forecast system with a forecast that is commonly used and whose performance can serve as a standard performance. A common reference used to evaluate precipitation forecasts over an area is based on past observations and is representative of the climatology in this area: for a given day and year, it is the ensemble of precipitation values observed on that same Gregorian day in past years of the observation period. A reference that is often used to evaluate seasonal streamflow forecasts is the

Extended Streamflow Prediction (ESP), which corresponds to using the precipitation climatology (as described above) as input to a hydrological model. Another common reference used to evaluate streamflow forecasts is the ensemble based on past streamflow observations (on the same day as the given forecast day). This reference does not use any precipitation forecasts or hydrological model. Lastly, the forecast system can also be compared with a former version of this forecast system. In this case, the skill quantifies the improvement achieved between the two versions of the system.

### 3.3.2 Ensemble size

Studies have shown that the size of an ensemble forecasting system influences the computation of some evaluation scores, such as the Brier score or the rank probability score (see e.g. Buizza and Palmer, 1998). An increase in the ensemble size of a forecast system leads to an increase in performance as evaluated by these scores, though to different extents. This has a direct impact when comparing the performances of forecasting systems of different ensemble sizes.

Several studies have thus proposed methods to remove the bias induced when comparing ensembles of different sizes. Ferro *et al.* (2008) provided a synthesis of existing previous studies, and detailed the bias correction methods that can be applied in the case of the Brier score and other probability scores. The authors proposed an approach to remove the bias in the computation of the CRPS. This approach was followed in this thesis to compute the CRPS skill score with systems and references of different ensemble sizes. Given a forecast system of ensemble size $M$ and a reference forecast of size $m$, the correction computes the score of the reference ensemble as though it had $M$ members. The corrected CRPS of the reference ensemble of size $m$ is defined by:

$$\mathrm{CRPS}_M = \mathrm{CRPS}_m - D \tag{3.6}$$

where $D$ is the correction factor expressed as:

$$D = \frac{M-m}{2Mm}\frac{1}{m(m-1)}\sum_{i=1}^{n}\sum_{k\neq l}(\hat{Q}_{i,k} - \hat{Q}_{i,l}) \tag{3.7}$$

where $\hat{Q}$ is the simulated streamflow (or precipitation), $n$ is the number of time steps used in the evaluation, and $k$ and $l$ refer to members of the forecast ensemble $\hat{Q}$. When the ensemble size varies with the month, as is the case for ECMWF seasonal forecasts, we chose to use the ensemble size averaged over one year.

### 3.3.3 Useful Forecasting Lead Time

In this thesis, we also used skill scores to estimate the gain in performance brought by applying bias correction methods. To that effect, we use the raw (uncorrected) forecasts as reference in the computation of the skill scores. An indicator of forecast performance can be derived from the evolution of these skill scores: the lead time up to which bias corrected seasonal forecasts have more skill than the raw forecasts. Buizza and Leutbecher (2015) defined the forecast skill horizon, with respect to the climatology, as “the forecast time when the average CRPS of the [forecast system] ceases to be statistically significantly lower, at the 99$^{th}$-percentile level, than the CRPS of the climatological ensemble [...]”. In another study, Nicolle *et al.* (2014) defined an indicator named UFL (Useful Forecasting Lead time) as “the lead time beyond which model performance is not at least 20 % better than benchmark performance”. Here, we considered the lead time beyond which the seven-day moving average of the skill score of the bias corrected forecast system, with regard to the raw forecast system, becomes negative. UFL values were then grouped in four categories: (1) None: no improvement over the forecast reference, (2) $<$30: gain up to 30 days, (3) $<$60: gain greater than 30 days and up to 60 days and (4) $>$60: gain greater than 60 days. Figure 3.4 illustrates the determination of the UFL and the UFL category in six examples.

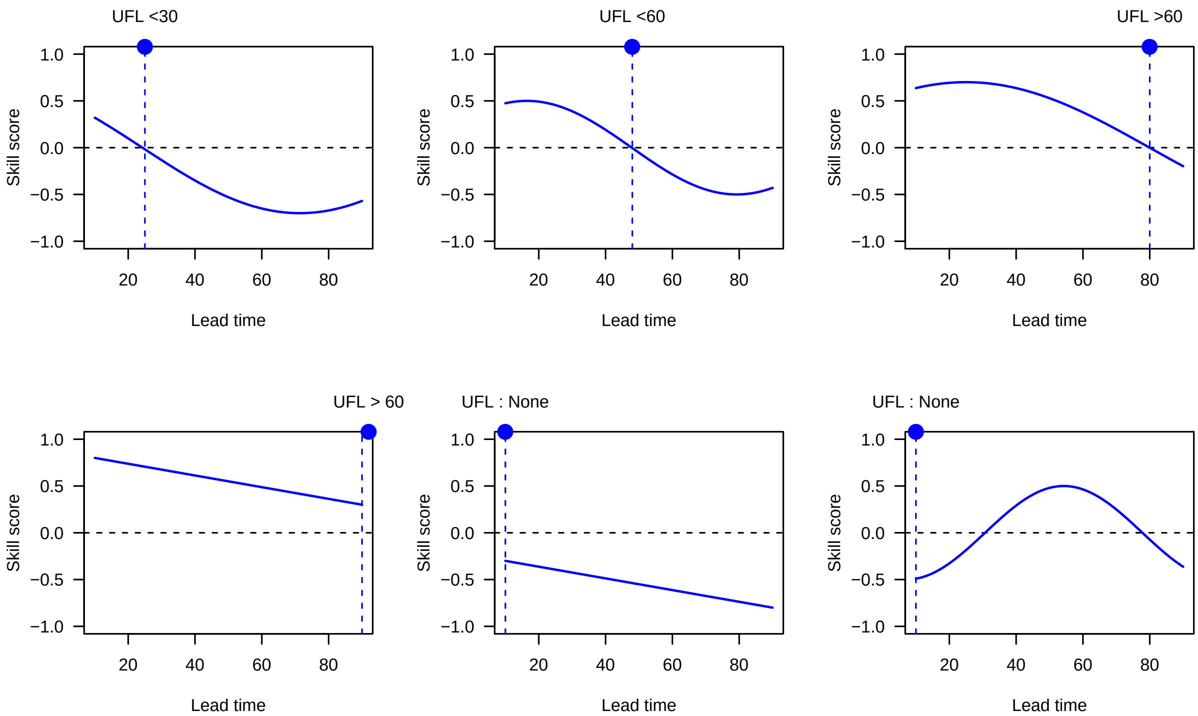



Figure 3.4: Illustration of how UFL values are determined and associated with a UFL category in six different cases of evolution of the skill score with lead time.

## 3.4 Conclusion

In this chapter we presented the forecast attributes evaluated in this thesis. The forecast systems investigated in Chapters 4, 5 and 6 were evaluated based on their accuracy, reliability, sharpness, overall performance and discrimination. The accuracy is assessed with the Mean Absolute Error (MAE) and the Root-Mean-Square Error (RMSE), the reliability is assessed with the PIT diagram and the PIT area, the sharpness is assessed with the Interquantile Range (IQR), the overall performance is assessed with the Continuous Rank Probability Score (CRPS) and the discrimination is assessed with the ROC diagram and the AUC. Skill scores are used in Part II of this thesis. In each case, the evaluation criterion and the reference used in the computation of the skill score are specified. The Useful Forecasting Lead time (UFL) is also used to evaluate forecasts in Chapter 4.

# *II*

# Seasonal Forecasting

# 4

# Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts

**This chapter is based on a paper under review in *Hydrology and Earth System Sciences***: Crochemore L., Ramos M.-H., Pappenberger F., 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrology and Earth System Sciences Discussions, doi:10.5194/hess-2016-78.

**Résumé**

De nombreux secteurs, tels que l'approvisionnement en eau potable, la gestion du risque sécheresse ou la production hydroélectrique, peuvent bénéficier de prévisions saisonnières de débits afin d'anticiper et de planifier leur gestion. De telles prévisions peuvent notamment être obtenues à l'aide de modèles pluie-débit. On parle alors de l'approche dynamique. Dans ce cas, les prévisions saisonnières provenant de modèles de circulation peuvent être utilisées en entrée des modèles hydrologiques pour bénéficier des avancées des centres météorologiques en terme de prévisions saisonnières. Néanmoins, ces produits étant souvent générés à des résolutions spatiales plus larges que celle du bassin versant, leur utilisation pour la prévision des débits nécessite un travail préliminaire de correction des biais. Dans ce chapitre, nous estimons la qualité de prévisions saisonnières de précipitations et de débits en France et explorons l'impact de la correction du biais des prévisions de précipitations sur la qualité des prévisions de débits.

Dans un premier temps, les prévisions saisonnières de précipitations brutes (i.e., sans correction du biais) du CEPMMT sont évaluées pour prévoir les débits de seize bassins versants français de 1981 à 2010. Dans un second temps, le biais de ces prévisions de précipitations est évalué et corrigé à l'aide de huit méthodes de correction du biais, basées sur une régression linéaire simple ou sur la correction des distributions prévues. Les prévisions de précipitations et de débits obtenues à l'aide de ces huit méthodes sont, à leur tour, évaluées. La qualité des prévisions est caractérisée en terme de fiabilité, finesse, précision et performance globale. L'ensemble de référence utilisé pour l'évaluation des pluies est basé sur la climatologie des pluies dans chaque bassin. L'ensemble de référence pour évaluer les débits est la méthode communément appelée Extended Streamflow Prediction (ESP) qui utilise la climatologie des précipitations en entrée d'un modèle hydrologique.

Dans l'ensemble des bassins versants, les prévisions obtenues à partir des précipitations brutes du CEPMMT sont plus fines mais moins fiables que celles obtenues avec la méthode ESP. La qualité des prévisions dépend fortement de la saison et du bassin considéré. Les biais observés en précipitation varient aussi de bassin en bassin, avec des biais mensuels importants qui ont tendance à se compenser si on ne regarde que le biais annuel. La correction linéaire simple et la correction des distributions des précipitations journalières apportent les meilleurs gains en performance. La correction linéaire a tendance à d'avantage améliorer la finesse des ensembles alors que la correction des distributions des valeurs journalières est nettement supérieure pour améliorer la fiabilité. La correction des distributions des précipitations journalières harmonise les performances entre bassins et saisons. Enfin, les prévisions corrigées deviennent aussi fiables que les prévisions de la méthode ESP tout en restant plus fines.

**Abstract**

Meteorological centres make sustained efforts to provide seasonal forecasts that are increasingly skilful, which has the potential to benefit streamflow forecasting. Seasonal streamflow forecasts can help to take anticipatory measures for a range of applications, such as water supply or hydropower reservoir operation and drought risk management.

This study assesses the skill of seasonal precipitation and streamflow forecasts in France to provide insights into the way bias correcting precipitation forecasts can improve the skill of streamflow forecasts at extended lead times. We apply eight variants of bias correction approaches to the precipitation forecasts prior to generating the streamflow forecasts. The approaches are based on the linear scaling and the distribution mapping methods. A daily hydrological model is applied at the catchment scale to transform precipitation into streamflow. We then evaluate the skill of raw (without bias correction) and bias corrected precipitation and streamflow ensemble forecasts in sixteen catchments in France. The skill of the ensemble forecasts is assessed in reliability, sharpness, accuracy, and overall performance. A reference prediction system, based on historical observed precipitation and catchment initial conditions at the time of forecast (i.e., ESP method), is used as benchmark in the computation of the skill.

The results show that, in most catchments, raw seasonal precipitation and streamflow forecasts are often more skilful than the conventional ESP method in terms of sharpness. However, they are not significantly better in terms of reliability. Forecast skill is generally improved when applying bias correction. Two bias correction methods show the best performance for the studied catchments, each method being more successful in improving specific attributes of the forecasts: the simple linear scaling of monthly values contribute mainly to increasing forecast sharpness and accuracy, while the empirical distribution mapping of daily values is successful in improving forecast reliability.

## 4.1 Introduction

Numerous activities with economic, environmental and political stakes benefit from knowing and anticipating future streamflow conditions at different lead times. While flood forecasting requires forecasts up to several hours or days ahead, other areas such as water supply reservoir operations or drought risk management need forecasts for the months or season ahead. Regardless of the considered lead time, streamflow forecasting systems are frequently updated to take the latest useful information into account (e.g. last observed discharge, soil moisture or snow cover) and can use numerical weather model outputs to extend the range of skilful predictions.

Seasonal forecasts can contribute to a proactive risk management, for example, for drought management (e.g. Wilhite *et al.*, 2000; Dutra *et al.*, 2014; Mwangi *et al.*, 2014; Wetterhall *et al.*, 2015). Extended-range forecasting systems can be valuable tools to help decision-makers in planning long-term strategies for water storage (Crochemore *et al.*, 2015b) and to support adaptation to climate change (Winsemius *et al.*, 2014). Nevertheless, several users still remain doubtful whether seasonal forecasts can be trustworthy or skilful enough to enhance decision-making in an operational context (Rayner *et al.*, 2005). Lemos *et al.* (2002) list the performance of seasonal forecasts, the misuse of seasonal forecasts by end-users and the lack of consideration of end-users' needs in the development of products as major obstacles to the widespread of seasonal forecasting in North-East Brazil. It is therefore crucial to assess the potential of available seasonal forecasting products and communicate on the assets and shortcomings of the different approaches that can benefit the water sector (Hartmann *et al.*, 2002).

Seasonal forecasting methods in hydrology can be broadly divided into two categories: statistical methods which use a statistical relationship between a predictor and a predictand, and dynamical methods which use seasonal meteorological forecasts as input to a hydrological model. More recently, mixed approaches have been investigated in the attempt to take advantage of initial land surface conditions, seasonal predictions of atmospheric variables and the predictability information contained in large-scale climate features (see Robertson *et al.*, 2013; Yuan *et al.*, 2015, and references therein). Extended Streamflow Prediction (ESP; Day, 1985) is a dynamical method widely used to forecast low flows and reservoir inflows at long lead times (Faber and Stedinger, 2001; Nicolle *et al.*, 2014; Demirel *et al.*, 2015). It consists in using historical weather data as input to a hydrological model whose states were initialized for the time of the forecast. The ESP method is also used along with the Reverse-ESP method to determine the relative impacts of meteorological forcings and hydrological initial conditions on the skill of streamflow predictions (Wood and Lettenmaier, 2008; Shukla *et al.*, 2013; Yossef *et al.*, 2013). An alternative dynamical method consists in using seasonal forecasts from regional climate models (RCMs) (Wood *et al.*, 2005). This approach yields better results when seasonal predictability is enhanced by meteorological forcings rather than by initial conditions. Climate model outputs may also be more suitable to capture the specific climate conditions at the time of the forecast, whereas ESP-based methods will be limited to the range of past observations and challenged by climate non-stationarity.

The use of climate model outputs in hydrology has however some methodological implications. Outputs are produced for grid scales that are usually too coarse for streamflow forecasting at the catchment scale. This can lead to errors in capturing forecast uncertainty and introduce significant biases. Post-processing (including bias correction techniques and downscaling procedures) is usually a necessary first step prior to using climate model outputs to model streamflow. A range of methods has been proposed in the literature and the best method usually depends on the modelling chain being investigated and the studied area, with levels of performance that may vary with the forecast horizon or the targeted application (Christensen *et al.*, 2008; Gudmundsson *et al.*, 2012).

Bias correction is usually an integral part of post-processing techniques applied to forecasting systems. Weather forecasting has performed bias correction of numerical model outputs through model output statistics (MOS) for decades. In hydrologic ensemble prediction systems, post-processing has become more and more popular in the last decade, particularly for medium-range ensemble forecasting (e.g. Weerts *et al.*, 2011; Zalachori *et al.*, 2012; Verkade *et al.*, 2013; Madadgar *et al.*, 2014; Roulin and Vannitsem, 2015). In seasonal forecasting, two popular bias correction methods are linear scaling and distribution mapping. Linear scaling corrects the mean of the forecasts based on the difference between observed and forecast means, whereas distribution mapping matches the statistical distribution of forecasts to the distribution of observations. These approaches, which can also be applied to improve the performance of ESP forecasts (Wood and Schaake, 2008), focus on increasing forecast skill and reliability, by reducing errors in the forecast mean and improving forecast spread.

Studies comparing different bias correction methods in seasonal hydrological forecasting are still rare in the literature. However, we can find studies reviewing and comparing methods to bias correct RCM outputs and quantify climate change impacts, although their efficiency in this context is still a topic of discussion (Ehret *et al.*, 2012; Muerth *et al.*, 2013; Teutschbein and Seibert, 2013). Teutschbein and Seibert (2012) compared six methods, among which linear scaling and parametric distribution mapping, to bias correct RCM simulations of precipitation and temperature in Sweden. The authors recommended using the distribution mapping method for current climate conditions. They also highlighted the need to assume that bias correction procedures are stationary to correct future climate projections and evaluate changes in flow regimes. In Norway, Gudmundsson *et al.* (2012) proposed a comparison of eleven methods to bias correct RCM precipitation. The methods derived from distribution transformations (e.g. distribution mapping based on fitted theoretical distributions), parametric transformations such as linear scaling, and nonparametric transformations such as distribution mapping based on empirical distributions. Their study highlighted the differences between the bias corrections and the necessity to test methods prior to their application. The authors recommended using nonparametric methods since these methods were the most effective to reduce the bias and did not require any approximations of the empirical distributions.

This chapter aims to further investigate the potential of bias correcting the seasonal forecasts produced by ECMWF System 4 forecasts from GCM simulations. The utlimate objective is to improve streamflow forecasting at extended lead times. By comparing several variants of linear scaling and distribution mapping methods, the study provides insights into the way bias correcting seasonal precipitation forecasts can contribute to the skill of seasonal streamflow predictions. Forecasts are evaluated over the 1981-2010 period in 16 catchments in France. Section 4.2 presents the catchment set, the forecast and observed data, as well as the hydrological model used. Section 4.3 presents the bias correction methods investigated, as well as the calibration and evaluation frameworks adopted. Results are presented in Sections 4.4 to 4.6 for the quality of the raw (uncorrected) and the bias corrected forecasts. In Section 4.7, conclusions and limitations are discussed.

## 4.2 Data and hydrological model

### 4.2.1 Seasonal forecasts and observed data

This study is based on daily seasonal precipitation forecasts from ECMWF System 4. For the purpose of this study, the 1981-2010 forecasts were aggregated at the catchment scale and only the first 90 days of the forecast horizon were considered. Observed precipitation data used for the calibration and evaluation of the bias correction methods come from the SAFRAN reanalysis of Météo-France. Daily streamflow data at the outlet of each catchment come from the French national archive. Chapter 1 provides a detailed description of observed and forecast hydrometeorological data.

### 4.2.2 Studied catchments and hydrological model

The catchment set used in this study includes 16 catchments spread over France with a dominant pluvial regime. The main characteristics and locations of the catchments are presented in Chapter 1. The conceptual, reservoir-based GR6J hydrological model (Pushpalatha *et al.*, 2011) was run at the daily time step with daily precipitation and potential evapotranspiration inputs at the catchment scale. The model output is the daily streamflow at the catchment outlet. Interannual potential evapotranspiration was used to focus solely on the influence of precipitation inputs on streamflow forecasts. The model was calibrated in each catchment with the Kling-Gupta Efficiency (Gupta *et al.*, 2009) applied to root-square flows ($\mathrm{KGE}_{rQ}$) and applied within the forecasting framework described in Chapter 2.

## 4.3 Methods

### 4.3.1 Overview of the calibration-evaluation approach

Bias correction methods were calibrated and evaluated in each catchment over the 1981-2010 period. The one-year-leave-out cross-validation method was applied to calibrate and evaluate the methods over independent periods. This method is further detailed in Chapter 2.

In the calibration step, we considered two approaches. The simplest calibration uses all days of the years within the calibration dataset. An alternative approach consists in calibrating the bias correction methods for each calendar month. Additionally, since we are dealing with forecasts issued up to 90 days ahead, and since forecast performance varies with lead time, calibration also takes the lead time into account. In this study, lead times were grouped from 1 to 30 days, 31 to 60 days and 61 to 90 days ahead. The calibrated bias correction factors are then applied to the daily values of the ensemble precipitation forecasts in the target application year. The hydrological model is forced by precipitation forecasts and streamflow ensemble forecasts are obtained. The modelling chain is applied to raw and bias corrected precipitation forecasts. Precipitation and streamflow forecasts are then evaluated with deterministic and probabilistic scores commonly used in ensemble forecasting.

### 4.3.2 Bias correction methods

We applied the linear scaling (LS) and the distribution mapping (DM) methods to the raw System 4 precipitation forecasts. The distribution mapping method was applied following three variants: considering the empirical distribution of monthly values (EDM), a fitted gamma distribution of monthly values (GDM), and the empirical distribution of daily values (EDMD). Each method was applied on a monthly (-m) or a yearly (-y) basis (Table 4.1).

#### Linear scaling of precipitations

The linear scaling method consists in correcting the monthly mean values of the forecasts to match the monthly mean values of the observations. A scaling factor (or bias) is calculated considering the ratio between the observed and the forecast (ensemble mean) values. A scaling factor higher (lower) than 1 indicates that the mean ensemble forecast underpredicts (overpredicts) the mean observed value. A value of 1 indicates no bias in the forecasts. The scaling factor obtained through calibration is then applied as a multiplicative factor to correct raw daily precipitation forecasts.

### Distribution mapping of precipitations

The distribution mapping method consists in correcting the precipitation forecasts so that their statistical distribution matches that of the observations. There are several ways to match forecast and observed distributions or quantiles, and existing techniques mainly differ on how the forecast and observed cumulative distribution functions (CDF) are considered. In some techniques, a parametric distribution is fitted to the forecast and observed datasets, while in others the empirical distributions and linear interpolations between data points or estimated quantiles are considered. In any case, observed and forecast CDFs must be determined from long data series.

In this study, the calibration of the distribution mapping method was first carried out considering empirical (EDM) and gamma-fitted (GDM) distributions of observed and forecast (ensemble mean) precipitation values averaged monthly. A third variant considered directly the empirical distribution of the daily values of the ensemble members (EDMD). These variants are listed in Table 4.1 After calibration, bias correction is applied to the daily precipitation forecasts of each application period. In the case of EDM and GDM, all daily values are corrected based on the correction suited to their monthly average. In the case of EDMD, each daily precipitation value of each forecast member is corrected individually.

Table 4.1: Abbreviation, calibration period and description of tested bias correction methods.

| Abbreviation | Calibration based on | Description |
|---|---|---|
| LS-y | the whole year | Linear scaling of monthly values |
| LS-m | calendar months | |
| EDM-y | the whole year | Empirical distribution mapping of monthly values |
| EDM-m | calendar months | |
| GDM-y | the whole year | Gamma distribution mapping of monthly values |
| GDM-m | calendar months | |
| EDMD-y | the whole year | Empirical distribution mapping of daily values |
| EDMD-m | calendar months | |

## 4.3.3 Evaluation framework

For each catchment, daily forecasts are issued once every month, up to 90 days ahead, during the 1981-2010 period. The quality of the forecasts was evaluated at the weekly time step (i.e., daily forecasts and observations are averaged over the week). Scores were computed as a function of lead time and for the winter (December-January-February), the spring (March-April-May), the summer (June-July-August) and the autumn (September-October-November) seasons. The reliability, sharpness, accuracy and overall performance of the ensemble forecasts were assessed based on the PIT diagram and area, the Interquantile Range (IQR), the Mean Absolute Error (MAE) and the Continuous Rank Probability Score (CRPS).

The standard computation of the skill (i.e. the version that is not normalized) was applied in this study to compare forecast systems. The skill scores were computed for the PIT area, the IQR and the CRPS (noted PITSS, IQRSS and CRPSS hereafter). The forecast skill was first evaluated by comparing the performance of the studied forecast systems with the performance of common reference forecasts. The reference used to evaluate precipitation forecasts is based on past observations and is representative of the catchment climatology. The references used to evaluate streamflow forecasts are the Extended Streamflow Prediction (ESP) and an ensemble based on historical streamflows. The ESP allows applying the same hydrological modelling setup to both the precipitation forecasts and the reference precipitation ensemble. Therefore, differences in performance are mainly due to differences between the precipitation inputs to the model. One would expect that precipitation and streamflow forecasts perform better than precipitation climatology, ESP or historical precipitations, at least in the first lead times. At longer lead times, natural variability should end up being a sound forecast. When skill scores were computed to indicate the gain in performance brought by bias correction methods, the raw (uncorrected) forecasts were used as reference in the computation of the skill scores. The UFL (Useful Forecasting Lead time) was derived from the evolution of these skill scores to indicate the lead time up to which bias corrected seasonal forecasts have more skill than raw forecasts.

The evaluation criteria, the computation of the skill and the references used in the computation of the skill scores are described in Chapter 3.

## 4.4 Quality of the raw seasonal forecasts

### 4.4.1 Performance of raw precipitation forecasts

Figure 4.1 presents the evolution of IQRSS and CRPSS with lead time, for winter (DJF) and summer (JJA). Each line corresponds to a catchment. Skill in sharpness and overall performance is very similar in winter and in summer (as well as in spring and autumn, not shown). Precipitation forecasts are overall sharper than historical precipitations in the large majority of catchments and up to long lead times. Some exceptions appear for lead times longer than three weeks, and especially in winter (wetter season in the majority of catchments). In terms of overall performance, precipitation forecasts clearly have skill up to two to three weeks ahead for 7-day averaged areal precipitation. At longer lead times, they are equivalent or perform slightly worse than historical precipitations.

Figure 4.2 shows the PIT diagrams for lead times of 30 and 90 days, for winter and summer. Grey lines represent the reliability of historical precipitations and coloured

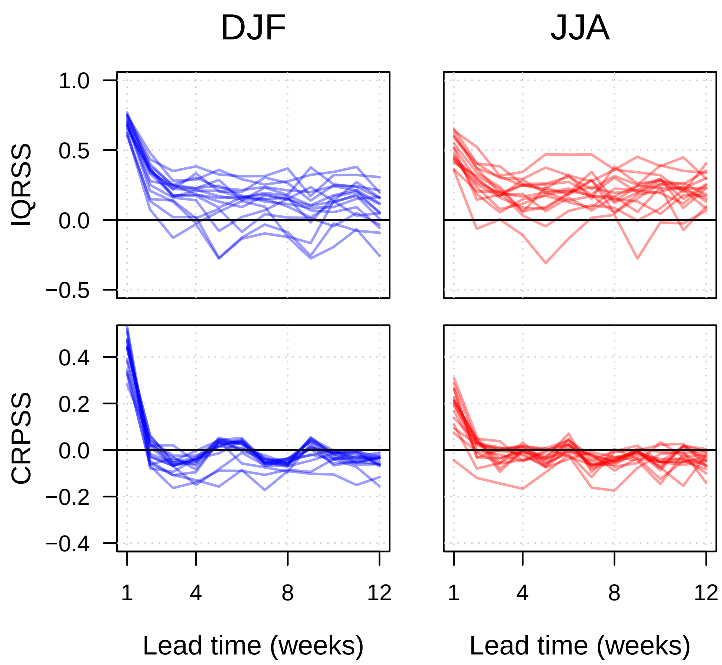



Figure 4.1: Skill of raw weekly precipitation forecasts as a function of the lead time for all catchments and all seasons. The skill is computed based on the IQR (top) and the CRPS (bottom) and the reference is historical precipitations. Each column corresponds to a target season. Each line represents the skill score in a catchment for forecast horizons within the target season.

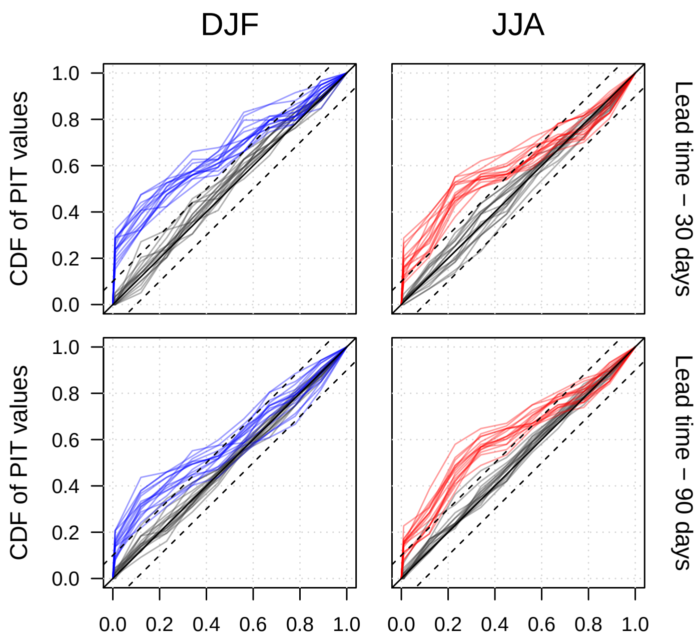



Figure 4.2: PIT diagram of raw precipitation forecasts (coloured lines) and historical precipitations (grey lines) for lead times of 30 days (top) and 90 days (bottom). Each column corresponds to a target season. Each line represents the PIT diagram in a catchment for forecast horizons within the target season.

lines represent the reliability of System 4 precipitation forecasts in each catchment. Dotted lines represent deviations of +0.1 and -0.1 from the bisector. The two seasons yield very similar results (also observed in spring and autumn, not shown). In all catchments and for both lead times, historical precipitations are reliable, as expected. Seasonal precipitation forecasts also show some reliability, but tend to overpredict precipitations in both seasons and at both lead times. The concentration of points in the zero end points in most of the curves of the System 4 forecasts shows that low values of the observations are too often falling in the lower tail of the forecast distribution. This effect tends to decrease with increasing lead time. This is an indication that forecasts are too narrow and overpredict the lowest observations. It can also translate a difficulty of the system to forecast null precipitation.

### 4.4.2 Performance of raw streamflow forecasts

Streamflow forecasts are generated by using raw seasonal precipitation forecasts as input to the hydrological model. Forecast skill is evaluated using the ESP method as reference (Figure 4.3). Differences in forecast skill between the winter and summer seasons are more noticeable when evaluating streamflow forecasts rather than precipitation forecasts. Streamflow forecasts generated from raw precipitation forecasts are sharper than ESP up to twelve weeks ahead in most catchments (IQRSS above zero in Figure 4.3). Approximately, only four catchments stand out in both seasons with lower skill than ESP (six in spring and one in autumn, not shown). However, even in these catchments, sharpness can be improved using seasonal precipitation forecasts for lead times up to three weeks in winter (as well as in spring and autumn, not shown). Concerning overall performance (CRPSS in Figure 4.3), skill can be observed for lead times up to four weeks in some catchments. In winter, as well as in spring and autumn (not shown), this is observed in the majority of catchments, while in summer, this concerns only a couple of catchments. At longer lead times, ESP and streamflow forecasts generated from raw precipitation forecasts are equivalent in most catchments for the winter season. In summer, as well as in spring and autumn (not shown), the difference in skill at longer lead times is more pronounced and most catchments have a clearly negative skill in terms of overall forecast performance.

PIT diagrams are shown for each catchment, for the winter and summer seasons, and for lead times of 30 and 90 days (Figure 4.4). In winter and spring (not shown), ESP forecasts and seasonal streamflow forecasts generated from raw precipitation forecasts show good reliability, although the curves above the diagonal indicate that forecasts are slightly overpredicting streamflow. Streamflow forecasts for the autumn season (not shown) also show good reliability, but with a tendency to underpredict streamflow. In summer (Figure 4.4, right), streamflow forecasts from both, ESP forecasts and forecasts generated from raw seasonal precipitation forecasts, show problems in forecast reliability. PIT curves clearly indicate a concentration of points at the end points of the

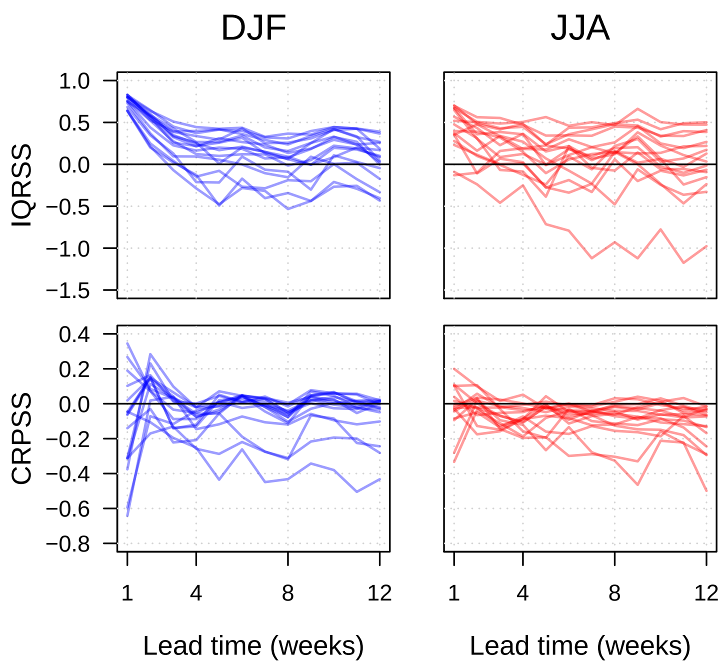



Figure 4.3: Skill of weekly streamflow forecasts from raw precipitation forecasts as a function of the lead time for all catchments and all seasons. The skill is computed based on the IQR (top) and the CRPS (bottom) and the reference is Extended Streamflow Prediction. Each column corresponds to a target season. Each line represents the skill score in a catchment for forecast horizons within the target season.

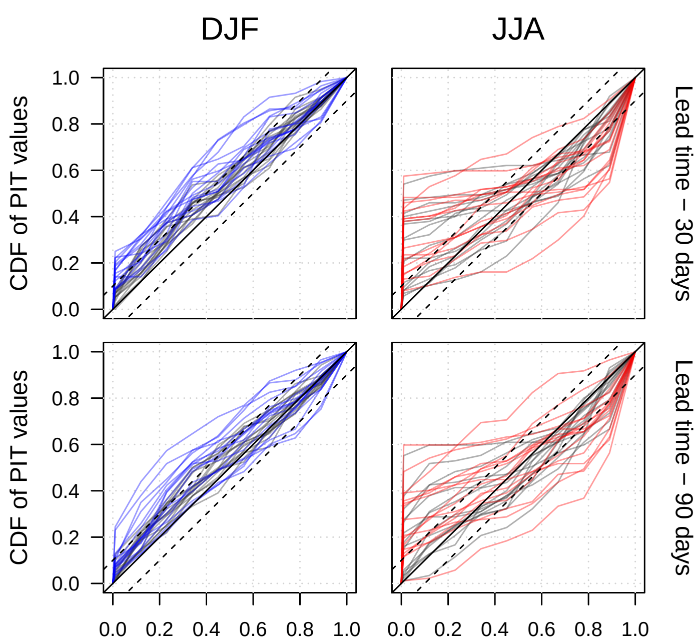



Figure 4.4: PIT diagram of streamflow forecasts from raw precipitation forecasts (coloured lines) and Extended Streamflow Prediction (grey lines) for lead times of 30 days (top) and 90 days (bottom). Each column corresponds to a target season. Each line represents the PIT diagram in a catchment for forecast horizons within the target season.

diagram and, consequently, narrow ensemble forecasts. In most catchments, 20 % to 60 % of observed values fall in the lowest interval of the forecast distribution or below it, i.e., outside the forecast range. Although reliability is slightly improved with lead time, streamflow ensemble forecasts remain under-dispersive at 90 days of lead time. This could be the result of at least two factors acting alone or jointly: a difficulty of the hydrological model to reach the lowest streamflow values in the simulations of the recession periods, and the influence of not considering uncertainties in the hydrological initial conditions at the time of forecasting.

### 4.4.3 Summary of the quality of raw seasonal forecasts

Skill in the overall performance of System 4 raw precipitation forecasts, at the catchment scale and over a reference forecast based on past observed precipitations, was observed up to two to three weeks in the studied catchments. When looking at streamflow forecasts generated from the input of raw seasonal forecasts to a hydrological model, skill over the traditional ESP method was observed up to four weeks, but only in few catchments. The asset of System 4 raw precipitation forecasts and related streamflow forecasts over historical precipitations and ESP, respectively, resides mainly in their sharpness. However, the evaluation of forecast quality shows also that forecasts are often too narrow and suffer from underprediction or overprediction. Improving forecast reliability, while maintaining forecast sharpness is clearly a challenge. In the following section, we investigate the presence of biases in System 4 precipitation forecasts and the impact of bias correction on seasonal precipitation and streamflow forecasts.

## 4.5 Bias correction of seasonal precipitation forecasts

### 4.5.1 Overview of the effectiveness of the bias correction methods

Forecast bias, i.e. the ratio between the mean observation and the average forecast ensemble mean, was computed for each catchment over the 1981-2010 period. The bias was computed for each calendar month, but also considering the whole year. Figure 4.5 shows the biases expressed as deviations from 1 (i.e., 1−Bias), before and after applying the bias correction methods. It illustrates the results obtained in four catchments at the 2-month lead time (i.e., considering the forecasts issued for day 31 to day 60 in the forecast range). The effectiveness of each bias correction method can be easily seen from the coloured charts: unbiased forecasts have a deviation equal to 0 (white colour); positive deviations (red colour) and negative deviations (blue colour) indicate overprediction and underprediction, respectively. A deviation equal to 0.75 (-3) can be interpreted as the mean forecast being four times larger (smaller) than the mean observation. Overall, when computing the deviations for all monthly lead times of the forecast range, we observed that the biases vary more with the calendar month of the

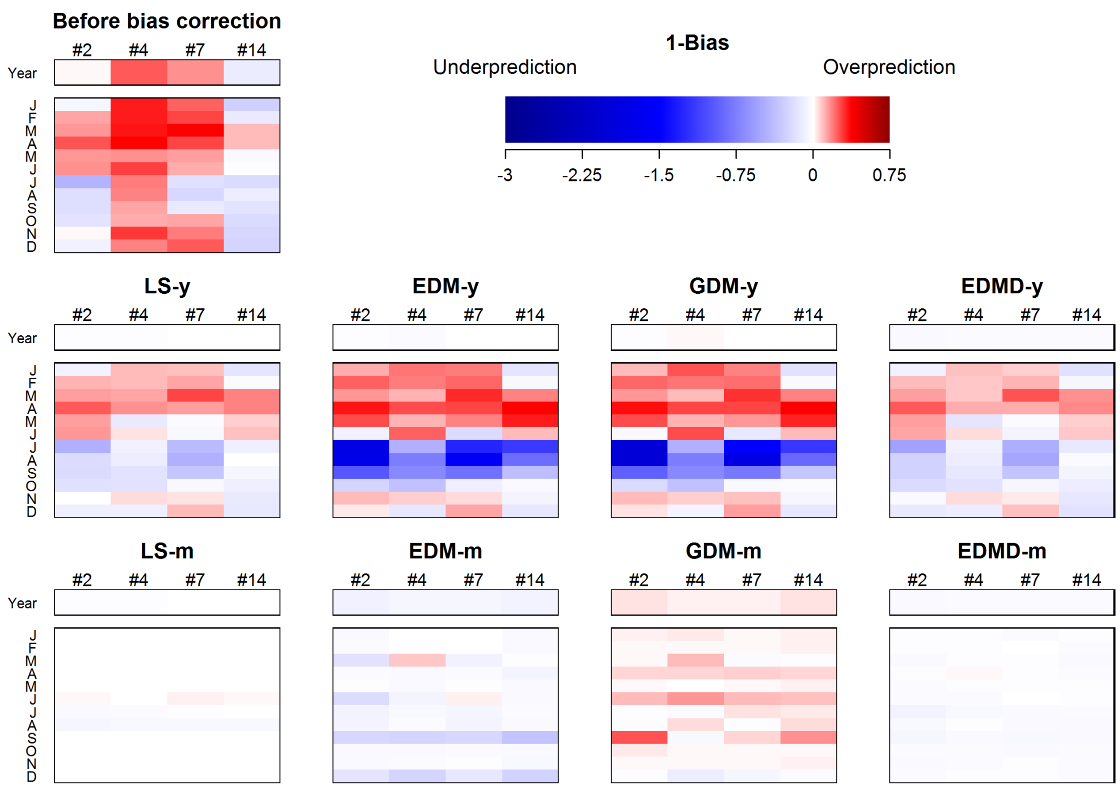



Figure 4.5: Deviation of the precipitation bias from 1, for catchments 2, 4, 7 and 14, over the 1981-2010 period. The deviation is shown for the whole year (top line) and for each calendar month. The bias is only shown for lead times between 31 and 60 days. Blue-shaded areas (negative values) represent a tendency of underpredicting precipitations and red-shaded areas (positive values) a tendency of overpredicting precipitations. The top left graph represents the bias of raw precipitation forecasts, and each of the other graphs represents the bias after applying one of the bias correction methods.

forecast horizon than with lead time. For this reason, we only show the 2-month lead time.

In general, seasonal forecasts tend to overpredict precipitations over the entire year in most catchments. Overprediction tends to occur near the end of the winter (rainy) season and throughout the spring season. Conversely, precipitations tend to be underpredicted from the end of the summer (dry) season and until the beginning, and sometimes throughout, the autumn season. The four selected catchments illustrate the variety of conditions we encountered in the bias correction analysis. In catchment 2, precipitations could be considered unbiased when carrying the analysis over the year. However, this result hides monthly underpredicting and overpredicting biases which compensate over the year. In this catchment, forecasts tend to overpredict from February to June and underpredict from July to October. The yearly result may also be a reflection of the lack of important biases in the months of December and January, which are, climatologically, the rainiest months in this catchment. This type of variation in bias was also observed in catchments 6, 11, 12 and 13. In catchment 4, precipitation forecasts are strongly overpredicting observations in all calendar months and thus over the year. This catchment stands out because in no other catchment do we observe a similarly strong and systematic bias. This catchment is the one located at the most eastern part of France. Its main river (l'Ill) is a tributary of the Rhine river. In catchment 7, precipitations are overpredicted over the year, with the strongest positive deviations concentrated during the rainy season, basically from November to April. The same behaviour is found in catchments 5, 10 and 15. Interestingly, catchments with a clear overprediction, i.e. catchments following the patterns depicted in Figure 4.5 for catchments 4 and 7, correspond to the catchments in which System 4 raw precipitation and streamflow forecasts showed low skill in sharpness and/or overall performance. Last, catchment 14 is representative of catchments 1, 3, 8, 9 and 16 in the database. Forecasts slightly underpredict precipitations over the year, with a tendency to underpredict precipitations in all seasons but the spring season, whose precipitations are slightly overpredicted.

Figure 4.5 also presents the remaining biases after the application of the eight bias correction methods to the raw precipitation forecasts. We present the results over the whole year and for each month. The same four selected catchments illustrate the results for the 2-month lead time. All correction methods are effective to correct biases of precipitation forecasts over the year. However, this is not observed in the bias correction for each calendar month. Results for the methods calibrated on a yearly basis (LS-y, EDM-y, GDM-y, EDMD-y) show that the absence of bias over the year is mainly achieved through an effect of compensation between over and underprediction among the calendar months. Particularly EDM-y and GDM-y methods show a tendency to increase monthly biases, towards overprediction of precipitations in winter and spring, and underprediction in summer and autumn.

By construction, monthly calibrated methods perform much better when looking at monthly biases. LS-m and EDMD-m are particularly effective in all catchments. Forecasts corrected with EDM-m tend to slightly underpredict precipitations, while forecasts corrected with GDM-m tend to overpredict precipitations. This may be an effect of the application of distribution mapping based on monthly values. Distribution mapping requires that the time structure of forecast and observed precipitation are coherent, so that upper forecast values are shifted towards upper observed values and conversely. However, raw monthly forecast means from System 4 do not always reproduce the time structure of monthly observations and often fail to reach extreme monthly values. Therefore, correction factors obtained with a distribution mapping based on monthly values show poorer performance, and the method can wrongly increase or decrease daily precipitation values.

### 4.5.2 Comparison of bias correction factors for LS and EDMD methods

The LS and EDMD methods showed more effectiveness in reducing bias in the precipitation forecasts. In order to better understand how the two methods compare, we plotted in Figure 4.6 their correction factors for catchment 7 over the 1981-2010 period for the 2-month lead time. Black lines represent correction factors from LS. Each day, one correction factor is applied to all members of the ensemble forecast at the 2-month lead time. Grey-shaded areas represent the range of correction factors applied with EDMD, and darker grey lines represent the median correction factor. For EDMD, each precipitation value has a specific correction factor depending on its probability of occurrence. Therefore, for a given day and lead time, the number of correction factors is equal to the number of ensemble members.

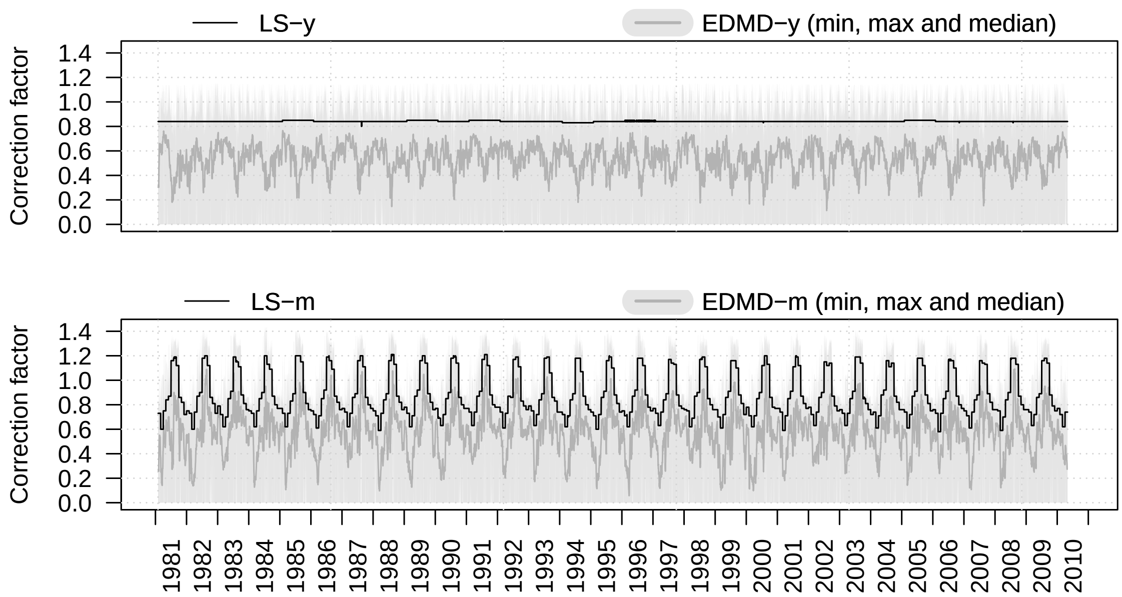



Figure 4.6: Bias correction factors applied to each day of the record period with the LS and EDMD methods. Correction factors are only shown in the case of catchment 7 and for the second month lead of the precipitation forecasts. The top graph presents correction factors obtained with LS and EDMD calibrated over the whole year, and the bottom graph presents correction factors obtained with LS and EDMD calibrated monthly.

LS-y provides relatively constant bias correction factors over the study period. Since, on average, precipitations in catchment 7 are overpredicted by System 4 forecasts, this correction factor is smaller than 1. The bias correction factors are obtained with the one-year-leave-out calibration framework. It is interesting to note that removing one year within the 30 years of the calibration period has little impact over the calibrated correction factors, even for an extreme dry year such as 1989 in this catchment. With EDMD-y, correction factors vary for each day of the study period. These factors remain smaller or close to 1. Their median values are smaller than the LS-y correction factors and the maximum values are slightly greater than the LS factors. When calibrated monthly, correction factors obtained with LS-m depict a variation, ranging from 0.6 to 1.2. They present a recurring pattern over the year, which follows what was shown in Figure 4.5, i.e., that precipitations in catchment 7 are, on average, overpredicted during the winter and spring seasons, leading to correction factors smaller than 1, and underpredicted from July to September, leading to bias correction factors greater than 1. This pattern in the factors indicates that the LS method might be further simplified to provide correction factors that would solely vary with the calendar month, regardless of the year, or in the case of LS-y, be constant over the target period. Correction factors computed with EDMD-m present a similar pattern to the one observed with LS-m, but their range is more variable, with values between 0 and 1.4. This method is particularly interesting because, as opposed to LS, it also corrects the frequency of precipitation days, given the null values of some correction factors.

### 4.5.3 Impact of bias correction on the useful forecasting lead time

The four criteria used to evaluate reliability, accuracy, sharpness and overall performance were applied to the precipitation forecasts bias corrected with each of the eight bias correction methods. They were also applied to the seasonal streamflow forecasts generated from inputting the different bias corrected precipitation forecasts to the hydrological model. Skill scores were computed with the raw seasonal precipitation forecasts as reference forecast for precipitation, and with the (raw) streamflow forecasts generated from raw precipitation forecasts as reference forecast for streamflow. For each variable (precipitation and streamflow), each criterion, each bias correction method, each catchment and each season, we obtained the corresponding UFL (Useful Forecasting Lead time). We then evaluated the proportion of catchments falling in each UFL group (as defined in Section 3.3.3). Results are shown in Figure 4.7 and Figure 4.8, for precipitation and streamflow forecasts, respectively.

In Figure 4.7, the two bias correction methods that stand out regarding overall performance (CRPS), in all seasons, are LS and EDMD. This is in accordance with our previous results on the efficiency of each method to correct biases. When looking more closely at improvements in the PIT criterion, as measured by the UFL, EDMD clearly stands out from the other methods. The proportion of catchments with skill improvement over raw forecasts is almost always 100 %, and skill is often extended up to 60 days and more. The other methods are quite equivalent to each other, although LS performs slightly better, with greater improvements in larger proportions of catchments, especially in winter and spring, for reliability (PIT), accuracy (MAE) and overall performance (CRPS). In terms of sharpness (IQR), the best performing method varies with the season. Precipitation forecasts in spring (MAM) are sharper when corrected with methods calibrated monthly, while forecasts in summer and autumn are sharper with methods calibrated yearly. To effectively address the tendency to overestimate spring precipitations, the multiplicative correction factor of a monthly calibrated bias correction for the spring season will be smaller than 1, and much smaller than the correction factor obtained with a yearly calibrated correction. Therefore, the spring interquartile range will be further reduced by the method calibrated monthly than by the method calibrated yearly. This reasoning only applies to LS, EDM and GDM since EDMD corrects each ensemble member independently.

Figure 4.8 shows the results for the streamflow forecasts. LS and EDMD methods are able to extend the lead time of bias corrected predictions further than other methods, and for a higher proportion of catchments in the large majority of seasons and criteria. Again, EDMD methods yield the best improvements in reliability. LS yields results slightly better than EDMD in sharpness and accuracy. EDM and GDM clearly have lower performance, except in some cases in sharpness and for spring and summer.

### 4.5.4 Summary of the comparison of bias correction methods

In general, LS and EDMD bias correction methods show good performance for precipitation forecasts, although in a distinct way. While EDMD clearly improves forecast reliability, LS shows better performance in improving sharpness. In terms of streamflow forecasts, LS and EDMD are the methods that offer the best performance. Again, EDMD may be preferred if focus is placed on forecast reliability, while LS may be preferred if sharpness and accuracy are the criteria one is looking to improve. Since streamflow forecasts generated from raw System 4 precipitation forecasts are already, in most of the studied catchments, sharper than the ESP reference, but lack reliability (as shown in Figure 4.3 and Figure 4.4), it seems appropriate to give priority to a correction method that improves reliability, while providing good overall performance. Therefore, in the following, we will only consider the monthly calibrated version of EDMD (EDMD-m) to further investigate the skill of bias corrected seasonal forecasts in the 16 selected French catchments. The monthly version is chosen to ensure that monthly biases are removed and that the correction will perform relatively equally in all seasons, while avoiding the "mis-estimation" of forecast skill (Hamill and Juras, 2006).

## 4.6 Skill scores of bias corrected seasonal forecasts

### 4.6.1 Performance of bias corrected precipitation forecasts

Figure 4.9 (for sharpness and overall performance) and Figure 4.10 (for reliability) present the skill of seasonal precipitation forecasts bias corrected with EDMD-m. Skill

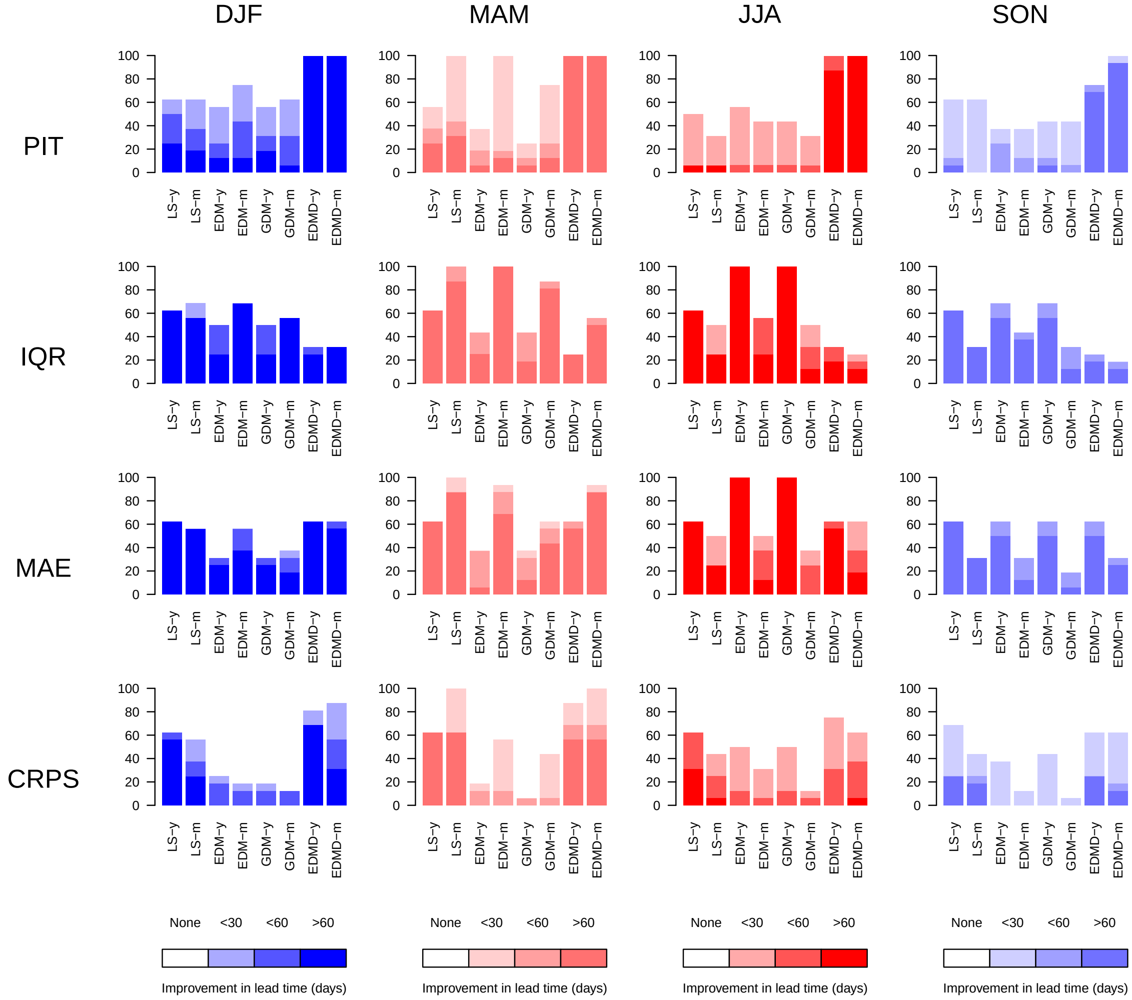



Figure 4.7: Number of catchments (%) in each UFL value category, i.e. number of catchments in which bias corrections increase the lead time up to which seasonal precipitation forecasts have skill in regards to raw seasonal precipitation forecasts. Each row corresponds to an evaluation criterion and each column corresponds to a season. Colour shades indicate the UFL category, i.e. the lead time up to which precipitation forecasts are improved.

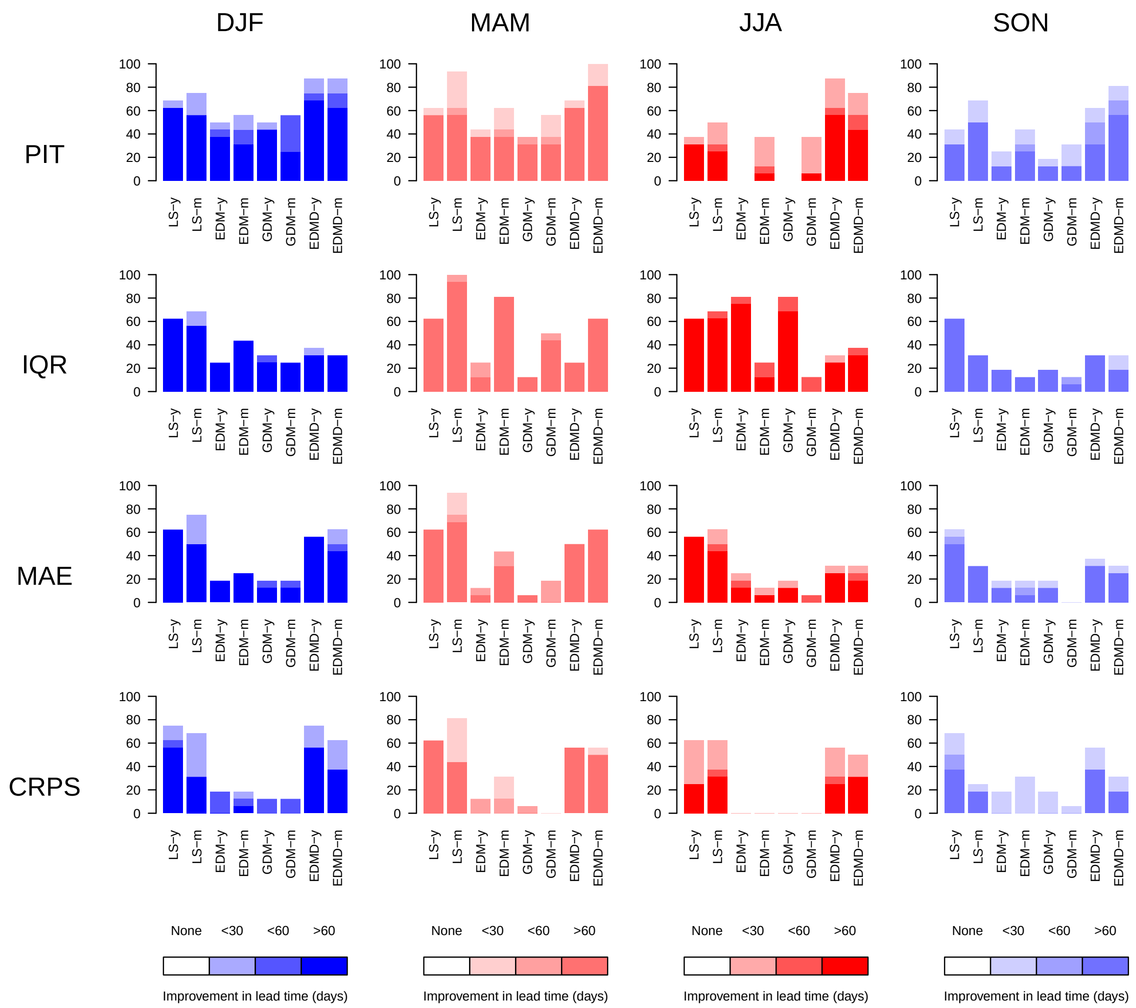



Figure 4.8: Number of catchments (%) in each UFL value category, i.e. number of catchments in which bias corrections increase the lead time up to which seasonal streamflow forecasts have skill in regards to seasonal streamflow forecasts generated from raw seasonal precipitation forecasts. Each row corresponds to an evaluation criterion and each column corresponds to a season. Colour shades indicate the UFL category, i.e. the lead time up to which streamflow forecasts are improved.

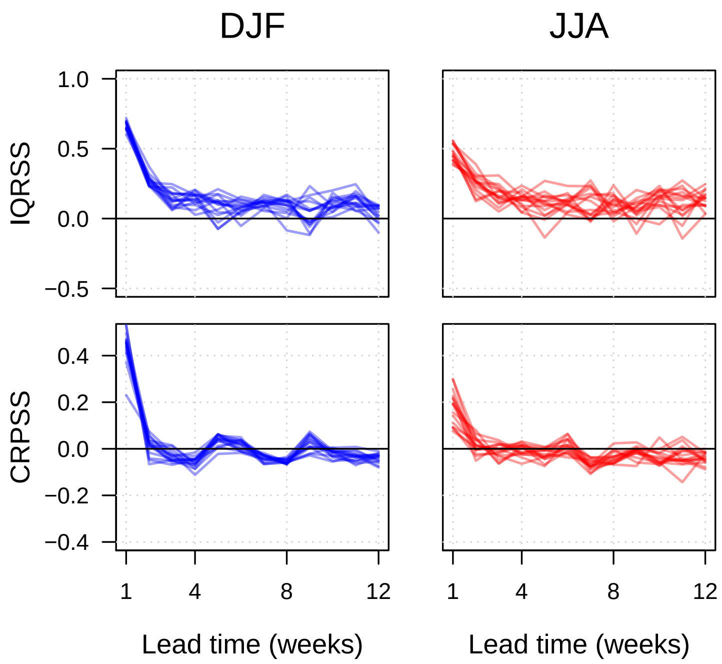



Figure 4.9: Skill of weekly precipitation forecasts corrected with EDMD-m as a function of the lead time for all catchments and all seasons. The skill is computed based on the IQR (top) and the CRPS (bottom) and the reference is historical precipitations. Each column corresponds to a target season. Each line represents the skill score in a catchment for forecast horizons within the target season.

scores are computed with historical precipitation as the reference. In order to better evaluate the impact of bias correction on forecast skill, the y-axes in Figure 4.9 are the same as in Figure 4.1. The comparison of these two figures shows that bias correcting the raw System 4 forecasts reduces the differences in skill between catchments. After bias correction, catchments present very similar evolutions of the skill with the lead time. We can also infer that, after bias correction, in some catchments, the values of IQR and CRPS are lower than before bias correction. Nevertheless, bias corrected forecasts remain sharper than the reference (i.e., skill scores are always greater than zero). In the catchments where the raw forecasts performed worse than historical precipitations (i.e., skill scores lower than zero in Figure 4.1), bias corrected forecasts become sharper and gain skill in regards to the reference. Forecast skill in overall performance (CRPSS) is observed up to two to three weeks ahead, after which forecasts attain skill equal to that of the reference forecast. Skill is improved in catchments that performed worse than the reference prior to bias correction (i.e., skill scores lower than zero in Figure 4.1). Figure 4.9 illustrates these findings for winter (DJF) and summer (JJA), but results are similar for spring and autumn (not shown).

Figure 4.10 shows that the most remarkable improvement in performance due to bias correction is achieved in reliability. While precipitation forecasts had a tendency to overpredict prior to bias correction, bias corrected precipitations are reliable in all

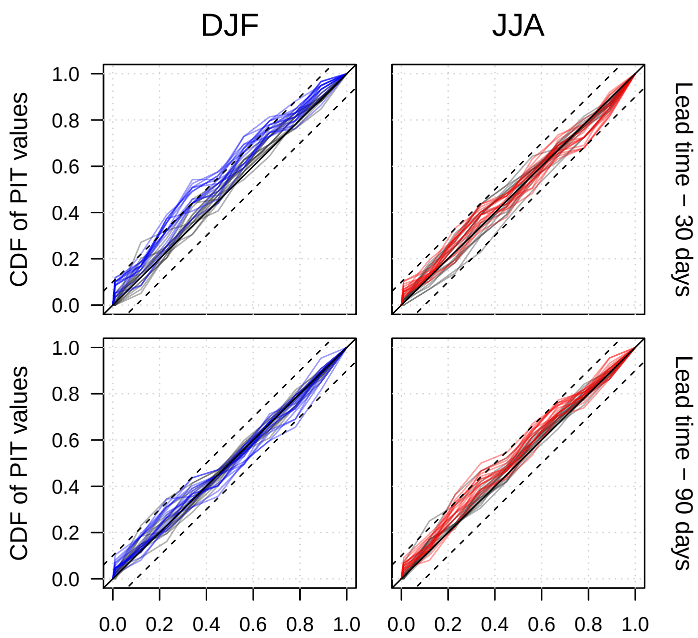



Figure 4.10: PIT diagram of precipitation forecasts corrected with EDMD-m (coloured lines) and historical precipitations (grey lines) for lead times of 30 days (top) and 90 days (bottom). Each column corresponds to a target season. Each line represents the PIT diagram in a catchment for forecast horizons within the target season.

catchments. Figure 4.10 shows the results for winter and summer, and for lead times of 30 and 90 days, but conclusions are similar in the other seasons and lead times (not shown). Even though a slight tendency to overpredict precipitations remains in winter for short lead times, the improvements are noticeable. The EDMD-m bias correction was able to address the concentration of points in the zero end point observed in Figure 4.2 for the raw forecasts.

### 4.6.2 Performance of bias corrected streamflow forecasts

The quality of the streamflow forecasts generated from the precipitation forecasts corrected with EDMD-m is investigated in Figure 4.11 and Figure 4.12 (IQRSS and CRPSS) and in Figure 4.13 (PIT diagrams). These figures can be compared to Figure 4.3 and Figure 4.4 which were obtained from the analysis of streamflow forecasts generated from raw precipitation forecasts. As seen with precipitation forecasts, bias correction also reduces the differences in streamflow forecast skill between catchments and seasons (Figure 4.11). Again, this translates into a loss in skill in catchments with the sharpest ensemble forecasts before bias correction, but also in a gain in skill in catchments where raw streamflow forecasts had negative skill. Overall, after bias correction, streamflow forecasts are sharper than ESP in all catchments and seasons (only the winter and summer seasons are shown but results are similar for the spring

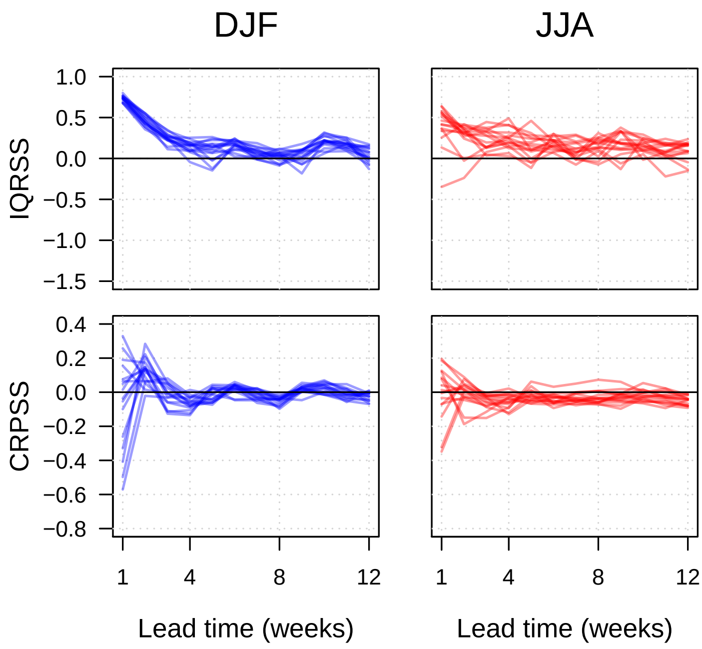



Figure 4.11: Skill of streamflow forecasts obtained from precipitation forecasts corrected with EDMD-m as a function of the lead time for all catchments and all seasons. The skill is computed based on the IQR (top) and the CRPS (bottom) and the reference is Extended Streamflow Prediction. Each column corresponds to a target season. Each line represents the skill score in a catchment for forecast horizons within the target season.

and autumn seasons). In terms of overall performance (CRPSS), the skill of streamflow forecasts was largely improved, especially in catchments that had very low skill prior to bias correction (i.e., CRPSS values well below zero in Figure 4.3). In winter, autumn and spring, skill over the ESP reference is observed up to four weeks ahead in several catchments (even up to five weeks ahead in spring and autumn), while in summer, it is observed up to two to three weeks. At longer lead times, streamflow forecasts show an overall performance equivalent or slightly lower than the performance of the ESP method.

Some studies use past streamflow observations (referred to as streamflow climatology) as the reference forecast to assess the skill of streamflow forecasts (e.g. Trambauer *et al.*, 2015; Wetterhall *et al.*, 2015). Figure 4.12 shows the skill in overall performance and sharpness when streamflow climatology is used as reference to calculate the skill of EDMD-m bias corrected forecasts. As expected, streamflow forecasts generated from bias corrected precipitation forecasts are sharper and present better overall performance than streamflow climatology, even for lead times of up to twelve weeks in some catchments. In one catchment (catchment 1), skill scores are systematically higher than the scores of the other catchments. In this catchment, streamflow climatology is very wide, with interannual variability of the same order of magnitude as interseasonal variability.

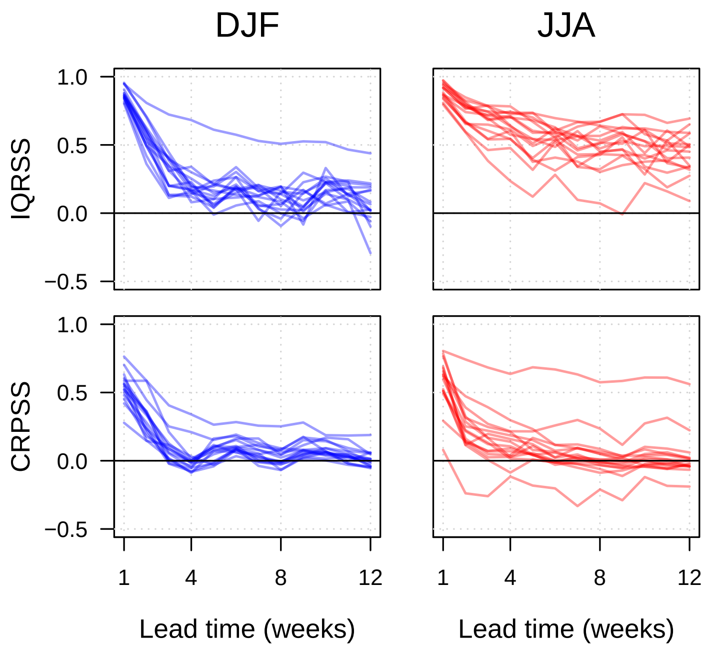



Figure 4.12: Skill of EDMD-m debiased streamflow forecasts as a function of the lead time for all catchments and all seasons. The skill is computed based on the IQR (top) and based on the CRPS (right) and the reference is historical streamflow. Each column corresponds to the target season of forecast lead times. Each plotted line represents the performance of a catchment.

The PIT diagrams in Figure 4.13 show that the reliability of streamflow forecasts is also improved after bias correcting precipitation forecasts. In winter (DJF) and spring (not shown), streamflow forecasts are now reliable and equivalent to ESP, although forecasts still show a slight tendency to overpredict streamflows. In autumn (not shown), streamflow forecasts are also reliable in most catchments, but with a tendency to underpredict streamflows. Summer (JJA) streamflow forecasts are also more reliable than they were prior to bias correction, but they still depict poor reliability and show that there is room for improvements. As shown by other studies in ensemble forecasting (Zalachori *et al.*, 2012; Verkade *et al.*, 2013; Roulin and Vannitsem, 2015), a simple bias correction of meteorological inputs is obviously not enough to achieve streamflow forecast reliability. In our case, the difficulties of the hydrological model in reaching lower streamflow values remain. This highlights the need for taking into account other sources of hydrological modelling uncertainties and including additional post-processing, targeting directly streamflow forecasts.

### 4.6.3 How improvements in precipitation forecasts propagate to streamflow forecasts?

We have seen that the use of reliable precipitation forecasts as input to a hydrological model does not automatically generate reliable streamflow forecasts. In order to further

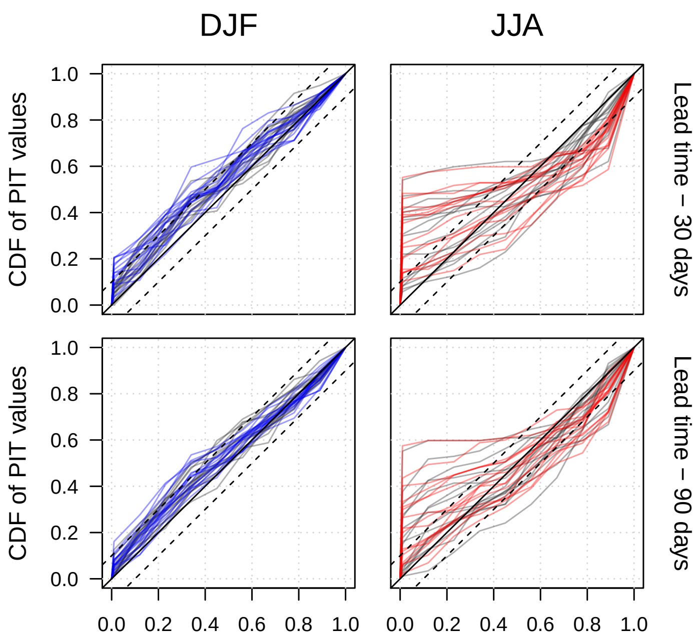



Figure 4.13: PIT diagram of streamflow forecasts obtained from precipitation forecasts bias corrected with EDMD-m (coloured lines) and Extended Streamflow Prediction (grey lines) for lead times of 30 days (top) and 90 days (bottom). Each column corresponds to a target season. Each line represents the PIT diagram in a catchment for forecast horizons within the target season.

understand how improvements in precipitation forecasts propagate to streamflow forecasts, we compared the skill scores of EDMD-m bias corrected precipitation forecasts with the skill scores of the streamflow forecasts generated from these bias corrected precipitations. We focused the analysis on the four catchments previously selected as representative of the database, i.e. catchments 2, 4, 7 and 14.

Figure 4.14 presents the results for the CRPSS, IQRSS and the PITSS (PIT area) in these four catchments. The reference forecast for the computation of the skill scores of the bias corrected forecasts is the raw forecast. The skill thus represents a measure of the improvement due to bias correction. Skill scores were averaged over lead times of 10 days to 90 days.

In overall performance (CRPSS), bias correcting precipitation forecasts either led to a gain in skill in both precipitation and streamflow forecasts, as in catchments 4 and 7 and in some seasons in catchment 2, or to a skill equivalent to the skill prior to bias correction, as in catchment 14. Since catchments 4 and 7 were the ones with the most biased forecasts (cf. Figure 4.5), there was more room for improvement in these catchments. Catchment 14 had the smallest bias of the four catchments. Bias correction had thus little impact on precipitation forecasts, and therefore also on streamflow forecasts. Interestingly, the improvement achieved in streamflow is always

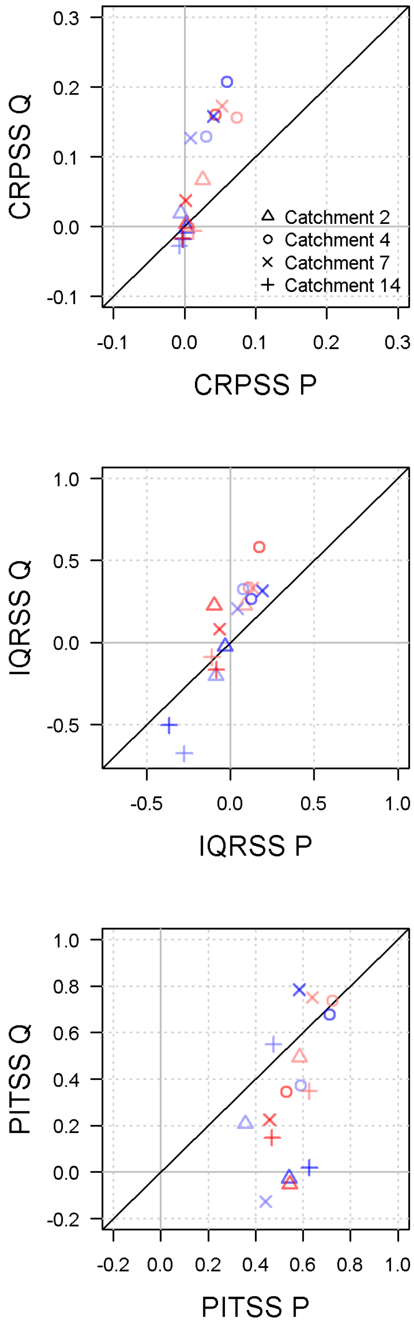



Figure 4.14: Skill scores of streamflow forecasts after correction with EDMD-m against skill scores of precipitation forecasts after correction with EDMD-m. The skill score of forecasts corrected with EDMD-m is computed in regards to raw forecasts. It is then averaged over lead times 10 to 90 days to obtain a single value. Results are shown for all four seasons in four selected catchments (Catchments 2, 4, 7 and 14). Skill scores were obtained based on the CRPS (top), the IQR (middle) and the PIT diagram area (bottom). The 1:1 diagonal corresponds to an equivalent performance increase in precipitation and streamflow.

superior to the improvement achieved in precipitation, or equivalent when there was no gain in skill. It seems therefore that a small improvement in the overall performance of precipitation inputs (as measured by the CRPS) can translate in a greater improvement in streamflow forecasts.

If we look at the skill in sharpness (IQRSS) and in reliability (PITSS) of the ensemble forecasts, we observe different behaviours. In sharpness, a loss in skill was observed in catchments 2 and 14, while a gain was observed in catchments 4 and 7. When a gain was achieved, the gain is superior in streamflow forecasts than in precipitation forecasts. If we look at reliability, skill was always improved by bias correcting the precipitation forecasts, with skill scores always superior to 0.3. The gain in streamflow is mainly positive, but not always, as in the case of precipitation forecasts. Although the majority of skill scores are superior to 0.1, some values are below the zero skill score line. The gain in reliability from the application of bias correction to raw precipitation forecasts is, in general, superior in precipitation forecasts than it is in streamflow forecasts.

Based on our results, we can say that in catchments with small biases, here represented by catchments 2 and 14, overall performance was mainly stable from precipitation to streamflow forecasts. However, in these catchments, a gain in reliability was generally associated with a loss in sharpness. In catchments with greater biases, here represented by catchments 4 and 7, overall performance, sharpness and reliability were improved for both precipitation and streamflow forecasts by simply bias correcting the precipitation forecasts.

### 4.6.4 Example of forecast hydrographs in a selected catchment

Figure 4.15 presents the hydrographs of the forecasts obtained from historical streamflow (HistQ), ESP, and seasonal forecasts bias corrected with LS-m and EDMD-m, from April 2004 to April 2007 in catchment 7. We show forecasts for lead times from 31 days to 60 days, i.e., forecasts issued in the previous month. Ensemble forecasts are represented by the median forecasts and two prediction intervals: the 25 % - 75 % interval containing 50 % of the ensemble members (dark grey zone), and the 5 % - 95 % interval with 90 % of the ensemble members (light grey zone). Observed streamflow is also shown. In this catchment, seasonal forecasts had a strong bias and bias correction methods performed well.

The hydrograph for historical streamflow represents the interannual variability in streamflow in the catchment, except that the forecast year is excluded for cross-validation. It relies on past observations of streamflow and does not include seasonal meteorological forecasts. We can see that the observations fall within the forecast ranges in most cases, which indicates, as expected with climatology, good forecast reliability. However, the forecast lacks sharpness during low-flow periods. Accuracy of the median forecast is, in general, good, although too high and low peak flows are not

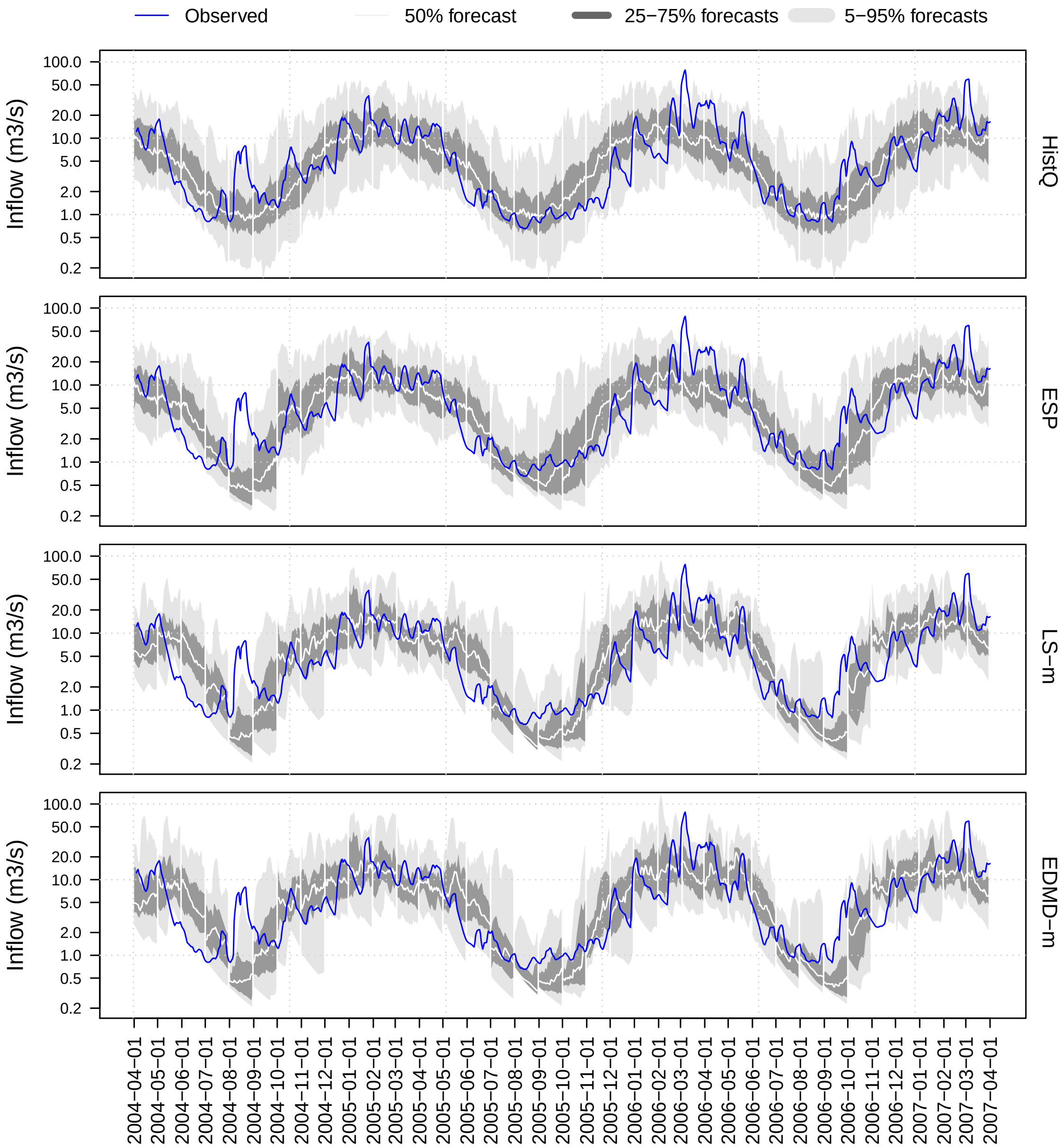



Figure 4.15: Hydrographs obtained with historical streamflow, ESP, seasonal forecasts corrected with LS-m and seasonal forecasts corrected with EDMD-m in catchment 7 from 1 April 2004 until 1 April 2007. The vertical axis is logarithmic. The blue line represents the observed streamflow. The grey shaded areas present the forecasts issued in the previous month, i.e. 31 to 60 days prior to the observations.

well reproduced.

The forecasts obtained with the ESP method use past observations of precipitation as input to the hydrological model rather than seasonal meteorological forecasts. They show visible improvements in sharpness during low flow periods, while reliability seems preserved. Accuracy of the median forecasts seems equal or lower than observed with historical streamflow.

The hydrographs representing the streamflow forecasts obtained from bias corrected System 4 precipitation seasonal forecasts show forecasts that are sometimes even sharper than ESP forecasts, as seen, for instance, for the rising limb in 2005. Overall, the observed streamflow falls within the forecast ranges. In some situations, as in the peak event in August 2004, prediction intervals of bias corrected seasonal forecasts, particularly in the EDMD-m case, are closer to observations than ESP forecasts. In general, visual differences in quality between seasonal streamflow forecasts obtained from precipitation forecasts corrected with LS-m and EDMD-m are hardly noticeable. Although EDMD-m forecasts seem to present slightly larger prediction intervals, which could result in better reliability but lower sharpness comparatively to LS-m, the accuracy of their median forecasts is practically identical. The visual inspection of these graphs for all catchments indicates similar results. Although our analyses and evaluation criteria have indicated the EDMD-m as the preferred method for the studied catchments, LS-m also yields good improvements in precipitation and streamflow forecasts. Since this method is easier to implement, it can be an alternative to the application of EDMD-m in operational forecasting systems.

## 4.7 Discussion and conclusions

We assessed the quality of ECMWF System 4 precipitation forecasts for seasonal streamflow forecasting in 16 catchments in France. We evaluated areal precipitation forecasts over the catchments and streamflow forecasts generated from inputting precipitation forecasts to a lumped hydrological model. Results show that, in most catchments, raw (uncorrected) System 4 precipitation forecasts are sharper than precipitation climatology (i.e., ensemble forecasts built from climatological precipitations) in all seasons. However, raw precipitation forecasts show poor reliability and a tendency to overpredict precipitations. Likewise, streamflow forecasts generated from raw System 4 precipitations are sharper, but far less reliable than forecasts based on the ESP approach (i.e., ensemble forecasts obtained from running the hydrological model with current initial conditions and past observed precipitations). Yet, in overall performance, raw precipitation forecasts yield improvements up to two weeks in all catchments over precipitation climatology, and streamflow forecasts yield improvements up to three to four weeks over ESP in some catchments. In general, improving forecast reliability, while maintaining (or not diminishing too much) forecast sharpness, was clearly a challenge for bias correction methods.

An in-depth analysis of the biases of System 4 seasonal precipitation forecasts showed strong monthly biases sometimes hidden at the scale of the year, depending on the catchment. Bias correction methods calibrated over the whole year were therefore less efficient when evaluating forecasts over calendar months. In the majority of catchments, the empirical distribution mapping of daily values (EDMD) or the simple linear scaling method (LS) applied to raw System 4 precipitation forecasts showed more effectiveness in correcting the yearly but also the monthly biases. These methods also gave the highest increase in overall performance for streamflow forecasting. Empirical distribution mapping of daily values calibrated for each calendar month (EDMD-m) was particularly efficient to increase reliability of precipitation and streamflow forecasts, while linear scaling (LS-m) led to higher improvements in sharpness and accuracy.

The EDMD-m bias correction method was further investigated to better understand its impact on the skill of bias corrected seasonal forecasts in the studied catchments. Overall, the application of bias correction reduced the differences in forecast performance between seasons and catchments for precipitation and streamflow forecasts. Also, bias correction ensured that precipitation and streamflow forecasts were at least equivalent in performance to the historical precipitations and streamflow forecasts based on historical precipitations, respectively, up to three months ahead. In catchments with greater biases, overall performance, sharpness and reliability were improved for both precipitation and streamflow forecasts by simply bias correcting the precipitation forecasts. Overall performance was mainly stable in catchments with small biases. However, in these catchments, a gain in reliability was generally associated with a loss in sharpness. The evaluation of forecasts after bias correction, for the purposes of operational applications on water and risk management, may therefore involve a trade-off between sharpness and reliability. Furthermore, while precipitation forecast reliability is improved with bias correction, the evaluation of streamflow forecast reliability shows that there is still room for improvement. Notably, bias correction of precipitation inputs was not enough to achieve good reliability in summer streamflow forecasts. This highlighted the need for adding a step of streamflow post-processing to the forecasting system.

This study compared eight simple bias correction methods to correct precipitation seasonal forecasts and investigated how one of them impacts the skill of streamflow forecasts. The catchments studied were not influenced by snowmelt flows and thus only precipitation was considered in the bias correction procedures. In other contexts, it may be interesting to also include bias correction of temperature forecasts, with appropriate methods to consider space-time interdependencies of the meteorological variables. The explicit consideration of temperature forecasts could also benefit the skill of low flow forecasts in summer, when evapotranspiration can play a crucial role.

Several other approaches for post-processing and bias correction exist, for instance, based on MOS techniques, space-time disaggregation schemes or Bayesian Model Averaging (Gneiting *et al.*, 2005; Raftery *et al.*, 2005; Liu *et al.*, 2013; Hemri *et al.*, 2014). These could be investigated to contribute to the comprehensive comparison of options for bias correcting precipitation and temperature forecasts prior to seasonal streamflow forecasting.

Last, other forecasting methods selecting historical precipitations based on climate indicators have been investigated in the literature for seasonal hydrological forecasting in regions where strong correlations have been observed, e.g. in the United States or in Australia. In France, weak correlations have often shown that climate indicators may not be adapted to forecast precipitations at the seasonal scale. However, the use of indicators derived from seasonal forecasts could potentially improve the selection of past precipitation scenarios, which might enhance the skill of ESP methods to forecast streamflow.

# 5

# Seasonal streamflow forecasting by conditioning climatology with precipitation indices

**This chapter is based on a paper under review in *Hydrology and Earth System Sciences*:** Crochemore L., Ramos M.-H., Pappenberger F., Perrin C., 2016. Seasonal streamflow forecasting by conditioning climatology with precipitation indices. Hydrology and Earth System Sciences Discussions, doi:10.5194/hess-2016-285.

**Résumé**

Dans le chapitre précédent, nous avons montré que la correction du biais des prévisions saisonnières de précipitations du CEPMMT peut améliorer la qualité des prévisions de débits, produites à l'aide d'un modèle hydrologique. Les prévisions de débits obtenues à partir des précipitations corrigées sont fines, mais restent peu fiables, en été par exemple. Dans ce chapitre, nous considérons d'autres méthodes dynamiques et statistiques de prévision des débits à l'échéance saisonnière exploitant la climatologie des pluies ou les données historiques de débits. Ces méthodes ont l'avantage d'être moins coûteuses en ressources informatiques et de produire des ensembles fiables, mais ne bénéficient pas des informations spécifiques au jour de la prévision. L'objectif de ce chapitre est de conditionner ces méthodes de prévision des débits basées sur les données historiques à partir des prévisions présentées dans le chapitre précédent, afin de bénéficier à la fois de la fiabilité des prévisions basées sur la climatologie, et de la finesse des prévisions basées sur les prévisions du CEPMMT.

Pour cela, des statistiques à long-terme des prévisions de pluies du CEPMMT, i.e., le cumul de précipitations et le SPI (Standardized Precipitation Index), sont utilisées pour sélectionner des années parmi les années de précipitations et de débits disponibles. Les ensembles produits restent basés sur les données historiques de pluies et de débits, mais sont spécifiques à la période de prévision. La finesse, la fiabilité et les performances globales des ensembles produits sont évaluées dans seize bassins versants français de 1981 à 2010. Ces ensembles sont ensuite comparés en fonction de leur capacité à prévoir des évènements en basses eaux, ainsi que des variables d'intérêt en étiages telles que le nombre de jours ou le volume déficitaire sous un seuil de basses eaux. Enfin, les prévisions sont comparées à l'aide d'un graphique d'évaluation des risques afin d'illustrer ces différences de performances dans le cas de la sécheresse de 2003 dans un bassin français.

Les résultats de l'étude montrent que les sélections basées sur le SPI, et en particulier le SPI calculé sur trois mois (SPI3), produisent des ensembles aux performances plus homogènes entre bassins que les autres méthodes de sélection. Les ensembles produits avec cette sélection sont fiables et plus fins que les ensembles de départ basés sur la climatologie. Pour des horizons supérieurs à un mois, les performances des ensembles sélectionnés ressemblent aux performances des prévisions issues des précipitations du CEPMMT. L'évaluation de la discrimination des ensembles de prévision montre de bons résultats pour les ensembles issus du modèle hydrologique (approche dynamique). Enfin, une application au cas de la sécheresse de 2003 montre que les prévisions sélectionnées peuvent aider à prévoir des évènements extrêmes de manière plus précise.

## Abstract

Many fields such as drought risk assessment or reservoir management can benefit from long-range streamflow forecasts. Climatology (i.e. time series of climate conditions recorded over a long time period) has long been used in long-range streamflow forecasting. In the last decade, the use of coupled general circulation model (GCM) outputs as input to hydrological models has developed. While precipitation climatology and historical streamflows offer reliable ensembles, forecasts based on GCM outputs can offer sharper ensembles, partly due to the initialisation of GCMs and hydrological models.

This study proposes to condition historical data based on GCM precipitation forecasts to get the most out of both data sources and improve seasonal streamflow forecasting in France. Four conditioning statistics based on ECMWF System 4 forecasts of cumulative precipitation or of the Standardized Precipitation Index (SPI) were used to select traces within historical streamflows and historical precipitations. This resulted in eight conditioned ensemble forecast scenarios. These conditioned scenarios were compared to an ensemble based on historical streamflows, to the widespread Extended Streamflow Prediction (ESP) ensemble, and to System 4 precipitation forecasts used as input to the GR6J hydrological model. These ensembles were evaluated based on their sharpness, reliability and overall performance.

An overall comparison of forecast ensembles showed that conditioning past observations based on the three-month Standardized Precipitation Index (SPI3) improved the sharpness of ensembles based on historical data, while maintaining a good reliability. An evaluation of forecast ensembles in low-flow forecasting showed that the SPI3-conditioned ensembles provided reliable forecasts of low flow duration and deficit volume based on the 80$^{\text{th}}$ exceedance percentile. Last, drought risk forecasting is illustrated for the 2003 drought.

## 5.1 Introduction

Numerical prediction is valuable to proactively manage risks in areas such as hydropower, drinking water production and drought preparedness (Wilhite *et al.*, 2000). Regardless of the application, probabilistic forecasts are preferred over deterministic ones to convey uncertainties (Krzysztofowicz, 2001; Ramos *et al.*, 2013). The main sources of uncertainty that play a central role in informing decision-making depend on the variable being forecast, the forecast horizon, but also on the location. For instance, region-specific tools have been developed in the world to predict and anticipate drought events weeks, months or even years in advance (Anderson *et al.*, 2000; Sheffield *et al.*, 2013; Ceppi *et al.*, 2014; Hao *et al.*, 2014; Shukla *et al.*, 2014). Nevertheless, anticipating river runoff events at long lead times remains a challenge (Yuan *et al.*, 2015).

The predictability of streamflow at long lead times lies in the initial hydrological conditions and the meteorological forcing. Research has shown that the relative role of each source of predictability mainly depends on the studied basin, the forecast season and the forecast lead time (Shukla *et al.*, 2013; Wood and Lettenmaier, 2008; Yossef *et al.*, 2013). Yossef *et al.* (2013) showed that in Western Europe, from July to October, streamflow forecasts are more dependent on meteorological forcing than they are on initial conditions, even one month ahead. The conclusions of Shukla *et al.* (2013) are quite consistent with these findings. They found that the predictability of a forecast issued in July in France lies in the meteorological forcing for horizons longer than three months. However, their results at shorter lead times are more nuanced, with predictability being led either by initial conditions or meteorological forcing, depending on the geographical location in France.

In practice, two approaches are often used to forecast streamflow at the seasonal scale (Easey *et al.*, 2006). Statistical approaches rely on past observations and statistical relationships between a predictor and a predictand. Dynamical approaches rely on coupled general circulation model (CGCM) outputs or past observations to feed a hydrological rainfall-runoff model. The choice of one approach over the other will depend on the purpose of the forecast, the region of interest and on the available data. More importantly, some studies have shown that the two approaches can complement and benefit from each other (Block and Rajagopalan, 2009; Seibert and Trambauer, 2015).

Climatology (past observations) is considered a good indicator of the range of possible outcomes for a given time of the year. Day (1985) introduced the Extended Streamflow Prediction (ESP), which is an approach that uses precipitation climatology as input to a hydrological model previously initialised for the forecast date. This approach has been extensively used, for research purposes and operationally, in seasonal streamflow forecasting (Wang *et al.*, 2011) and reservoir operations (Faber and

Stedinger, 2001), among other fields. An alternative to climatology is the seasonal forecasts issued by CGCMs (Yuan *et al.*, 2015). While these are initialized and forced for a specific forecast day, precipitation climatology simply provides a range of what has been previously observed on the forecast day, regardless of the current atmospheric situation and latest observations.

More recently, research has focused on fine-tuning the traditional ESP method by selecting relevant years within the climatology. In that context, several studies have proposed to condition or weight past observations based on climate signals. The proposed approaches are commonly divided in pre-ESP (prior to hydrological modelling) and post-ESP approaches (after hydrological modelling). In Northern America, several studies have taken advantage of the influence of the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) to improve the skill of seasonal forecasts. Hamlet and Lettenmaier (1999) selected past precipitations based on categories of ENSO and PDO to feed a hydrological model for streamflow forecasting, and, later on, for reservoir operation (Hamlet *et al.*, 2002). Werner *et al.* (2004) selected and weighted traces based on the ENSO before and after hydrological modelling. The authors showed that the post-ESP method yielded greater improvements in forecast skill than the pre-ESP method. Their post-ESP method was recently applied by Trambauer *et al.* (2015) in Southern Africa. Gobena and Gan (2010) used the PDO in several pre- and post-ESP resampling, including a pre-ESP approach benefiting from monthly precipitation and temperature statistically derived from climate model outputs. Recent studies have investigated the use of multiple other climate indices in post-ESP techniques (Najafi *et al.*, 2012). At the scale of the globe, van Dijk *et al.* (2013) selected traces within precipitation climatology based on climate indicators that were proven influential for the region and time period. They showed that using climate information improved forecast skill in Southeast Asia and South America.

In Europe, teleconnections show complex patterns and strongly depend on the season (Ionita *et al.*, 2015). Bierkens and van Beek (2009) exploited the teleconnection found between winter precipitations and the Northern Atlantic Oscillation (NAO) to select traces within the precipitation climatology and forecast seasonal streamflows. In Czech Republic, Šípek and Daňhelka (2015) ran a hydrological model with synthetic series of precipitation and temperature generated from climate forecasts and historical meteorological series. In France, Sauquet *et al.* (2008) forecast low flows in the Rhine river by selecting past precipitation scenarios that were close to the forecast day in terms of previous amounts of precipitation. Other approaches have consisted in directly taking profit of the information offered by long streamflow records. For instance, Svensson (2016) selected analogues within historical streamflows based on the streamflow anomaly observed in the month prior to the forecast date. The author aimed to forecast mean streamflow over the coming month or the coming three months in the United Kingdom.

In California, Carpenter and Georgakakos (2001) and Yao and Georgakakos (2001) tested several streamflow forecasting methods to forecast the inflows to the Folsom Lake. Based on the hypothesis that "*It is not necessary [...] that low skill in reproducing regional precipitation is an index of the utility of GCM information for systems acting as low-pass filters, such as the hydrological and reservoir systems are.*" Carpenter and Georgakakos (2001) conditioned historical precipitations based on the precipitation anomaly forecast by a GCM. They found that this conditioning was particularly efficient to forecast the low 30-day inflows to the lake: "*Global climate model information from the Canadian coupled global climate model CGCM1 benefits the mean forecasts significantly mainly for low observed 30-day inflow volumes.*" Yao and Georgakakos (2001) compared this method with the ESP method, and with a forecast ensemble conditioned from historical streamflows based on the latest observed reservoir inflows. They found that the GCM-conditioned ensemble outperformed the ESP method, although the ensemble conditioned from historical streamflows, which was the most reliable, managed to completely eliminate flood damage and generate more energy than the other two ensembles.

This study proposes to investigate how selecting historical data based on forecast precipitation indices contributes to the skill of seasonal streamflow forecasts. Our approach selects traces of past observed precipitations and streamflows based on precipitation indices derived from the System 4 seasonal precipitation forecasts issued by the European Centre for Medium-range Weather Forecasts (ECMWF). The aim is to generate forecasts that benefit from the reliability of climatology-based ensembles and the sharpness of System 4 precipitation forecasts. In the previous chapter, we assessed the performance of System 4 precipitation forecasts for seasonal streamflow forecasting. Despite the good overall performance of the streamflow forecasts after bias correction, we still observed a lack of reliability of the forecasts generated with the hydrological model in summer. In accordance with the results from Carpenter and Georgakakos (2001), we evaluate the proposed methods in contexts of low flows and droughts.

Section 5.2 presents the data and the methodology used to build streamflow forecasts. In Section 5.3, we present the evaluation of the different studied scenarios. First, we analyse the impact of the conditioning on the overall performance, sharpness and reliability of seasonal streamflow forecasts over the whole year. Then, we investigate the discrimination and reliability of the ensemble prediction systems to forecast low-flow events. We also illustrate the performance of our approach in forecasting drought risks through the case of the 2003 severe drought in France. In Section 5.4, we discuss the main outcomes and perspectives of the study.

## 5.2 Data and methods

### 5.2.1 Observed and forecast hydrometeorological data

Mean areal observed precipitation and potential evapotranspiration data come from the SAFRAN reanalysis of Météo-France. Daily streamflow data at the outlet of each catchment come from the French HYDRO national archive. Seasonal precipitation forecasts are from ECMWF System 4. They were aggregated at the catchment scale and only the first 90 days of the forecast horizon were considered. Chapter 1 provides the detailed description of the observed and forecast hydrometeorological data. In Chapter 4, we compared several methods to bias correct System 4 precipitation forecasts for seasonal streamflow forecasting. We showed that the empirical distribution mapping of daily values improved the reliability of both precipitation and streamflow forecasts. Following these results, the System 4 precipitation forecasts used in this chapter were previously bias corrected with this method.

### 5.2.2 Catchments and hydrological model

The catchment set used in this chapter is the same as in Chapter 4. It includes 16 catchments spread over France, whose main characteristics and locations are presented in Chapter 1. In these catchments, low flows are observed between May and October. One of the major drought events in these catchments is the 2003 drought, which caused approximately 15,000 deaths and cost over a billion euros just in France (UNEP, 2004; Poumadère *et al.*, 2005). This drought event is used to illustrate the work on drought risk assessment in Section 5.3.3.

The conceptual, reservoir-based GR6J hydrological model (Pushpalatha *et al.*, 2011) was run at the daily time step with daily precipitation and potential evapotranspiration inputs at the catchment scale. The model output is the daily streamflow at the catchment outlet. Interannual potential evapotranspiration was used to focus solely on the influence of precipitation inputs on streamflow forecasts. The model was calibrated in each catchment with the Kling-Gupta Efficiency (Gupta *et al.*, 2009) applied to inverse flows ($\mathrm{KGE}_{iQ}$) to focus on the lowest flows of the hydrograph. Details on the hydrological model and on the forecasting framework are described in Chapter 2.

### 5.2.3 Evaluation framework

The quality of the forecasts was evaluated at the daily time step and up to 90 days ahead. We assessed the sharpness and the reliability of the ensemble forecast systems following the paradigm introduced by Gneiting *et al.* (2007), which is maximizing sharpness while guaranteeing reliability. These were assessed based on the Interquantile Range (IQR), and on the PIT diagram and its area. The overall performance and the discrimination of the forecasts were also evaluated through the Continuous Rank

Probability Score (CRPS) and the ROC diagram and its area (AUC), both computed based on the $80^{\text{th}}$ exceedance percentile. Details on the evaluation criteria and the normalized skill score used in this chapter are presented in Chapter 3.

### 5.2.4 Forecast scenario building method

Eleven ensemble forecast scenarios were compared based on their performance in forecasting streamflows. Three scenarios are based on methods commonly used in seasonal streamflow forecasting. These are named “base ensembles” in the following. The remaining eight scenarios are based on these base ensembles and specific conditioning statistics. Table 5.1 summarizes the different ensemble forecast scenarios compared in this study.

#### Description of base ensembles

The simplest ensemble forecast scenario uses the long-term statistical variability of historical streamflows. It is assumed that the streamflow at a given day of the year is likely to fall within the range of past streamflows observed in previous years, on that same day. Apart from the necessity to have a long time series of streamflow records, this ensemble is not computationally costly. It is named **HistQ** hereafter.

Another base ensemble is the traditional **ESP** method. It requires a hydrological model and a long time series of precipitation records. This ensemble is based on the assumption that the precipitation of a given day is likely to fall within the range of past precipitations observed in previous years, on that same day. For a given forecast day, a precipitation ensemble is thus built by using precipitations observed in previous years. The precipitation ensemble has as many members as the number of years available in the precipitation record. The states of the GR6J hydrological model are first initialized with a one year run up to the forecast date. The precipitation ensemble and interannual evapotranspiration are then used as input to the model.

The third base ensemble uses the ECMWF System 4 seasonal precipitation forecasts as input to the GR6J hydrological model. Both the System 4 GCM and the hydrological model are initialized for the forecast day. This ensemble can be considered the most costly in terms of implementation and computational needs. Hereafter, this ensemble is named **Sys4**.

#### Description of conditioned scenarios

From the base ensembles, we built eight other scenarios by selecting traces within the HistQ and the ESP ensembles. The conditioning was based on statistics derived from the System 4 precipitation forecasts. Four statistics were computed for each forecast date and each member of the seasonal forecasting system. Two are based on cumulative rainfalls, and two on the standardized precipitation index (SPI). The SPI transforms the distribution fitted to a long precipitation record into a normal distribution (McKee

Table 5.1: Summary of the methodology used to build the ensemble forecast scenarios.

| Name | Statistic on seasonal forecast used as condition | Additional condition | Size | Initial hydrological conditions | Hydrological model | Precipitation forecast |
|---|---|---|---|---|---|---|
| **HistQ** | No condition | - | Past available streamflows | no | no | no |
| **HistQ_Sum3** | Precipitation volume | previous streamflow | 15 or 51 | yes | no | no |
| **HistQ_Sum1** | Monthly precipitation volume | previous streamflow | 15 or 51 | yes | no | no |
| **HistQ_SPI3** | SPI3 | previous streamflow | 15 or 51 | yes | no | no |
| **HistQ_SPI1** | SPI1 | previous streamflow | 15 or 51 | yes | no | no |
| **ESP** | No condition | - | Past available precipitations | yes | yes | no |
| **ESP_Sum3** | Precipitation volume | - | 15 or 51 | yes | yes | no |
| **ESP_Sum1** | Monthly precipitation volume | - | 15 or 51 | yes | yes | no |
| **ESP_SPI3** | SPI3 | - | 15 or 51 | yes | yes | no |
| **ESP_SPI1** | SPI1 | - | 15 or 51 | yes | yes | no |
| **Sys4** | No condition | - | 15 or 51 | yes | yes | yes |

*et al.*, 1993; WMO, 2012). An SPI value of 0 corresponds to conditions close to the long-term average of precipitations. Negative (positive) SPI values correspond to drier (wetter) conditions. The four conditioning statistics are:

- the cumulative precipitation forecast over the first three months of lead time (**Sum3**);
- the series of cumulative precipitation forecast over the first, second and third months (i.e. one value per lead time, **Sum1**);
- the SPI over the first three months altogether (**SPI3**);
- the SPI over the first, second and third months separately (i.e. one value per lead time, **SPI1**).

The statistics (SPI or precipitation volume) derived from System 4 forecasts are then used to select traces within HistQ and ESP. For that purpose, statistics are also computed for sequences of historical precipitations. Here, we consider sequences that start within 15 days of the forecast date, observed in years different from the forecast year. For a given forecast member, the sequence that is the closest in terms of the Euclidian distance, and with regard to the considered statistics, is selected. Note that different forecast members can be associated with the same "closest" historical sequence.

Once the sequences are selected, two options can then lead to a streamflow forecast ensemble: (a) the selected precipitation sequences can be used as input to the hydrological model to generate a streamflow forecast ensemble (**ESP_Sum3**, **ESP_Sum1**, **ESP_SPI3**, **ESP_SPI1**), or (b) the historical streamflows corresponding to the selected sequences can be directly used as ensemble members to build a streamflow ensemble (**HistQ_Sum3**, **HistQ_Sum1**, **HistQ_SPI3**, **HistQ_SPI1**). In the latter case, conditioning streamflow sequences based on rainfall statistics may result in unrealistic forecasts due to initial conditions far from what is observed on the forecast date. Therefore, when directly selecting scenarios from past streamflow observations, the last observed streamflow is added as a conditioning criterion in the computation of the Euclidian distance.

Before evaluating the performance of the eleven ensemble forecast scenarios, we evaluated the skill of System 4 in forecasting the conditioning statistics (cumulative precipitations and SPI). Figure 5.1 shows the skill in overall performance (CRPSS) and in sharpness (IQRSS), and the reliability (PIT diagram). The reference forecast used to compute the skill scores is historical precipitations (i.e. climatology). Regardless of the considered statistic, System 4 performs as well as climatology while being sharper. In addition, SPI forecasts issued from System 4 are reliable overall and in standard precipitation conditions. In dry conditions (i.e. SPI values smaller than -1), however, forecasts tend to overestimate SPI values, while in wet conditions (i.e. SPI values greater than 1) forecasts tend to underestimate SPI values. Similar PIT diagrams are observed with SPI forecasts from historical precipitations (not shown). Dutra *et al.* (2014) did a similar comparison and showed that System 4 forecasts performed better than or similarly to historical precipitations to forecast SPI values in South Africa.

## 5.3 Performance of the streamflow forecasting systems

### 5.3.1 Statistical evaluation of accuracy and reliability

#### Influence of conditioning on streamflow forecasts performance

We evaluated the gain and loss in skill of daily streamflow forecasts due to the four types of conditioning applied to the HistQ base ensemble. Figure 5.2 shows the CRPSS, IQRSS and PITSS for lead times up to 90 days, and the PIT diagram for a lead time of 45 days. The reference for the computation of the skill is HistQ, i.e. historical streamflows with all available years. Each line corresponds to one of the 16 catchments.

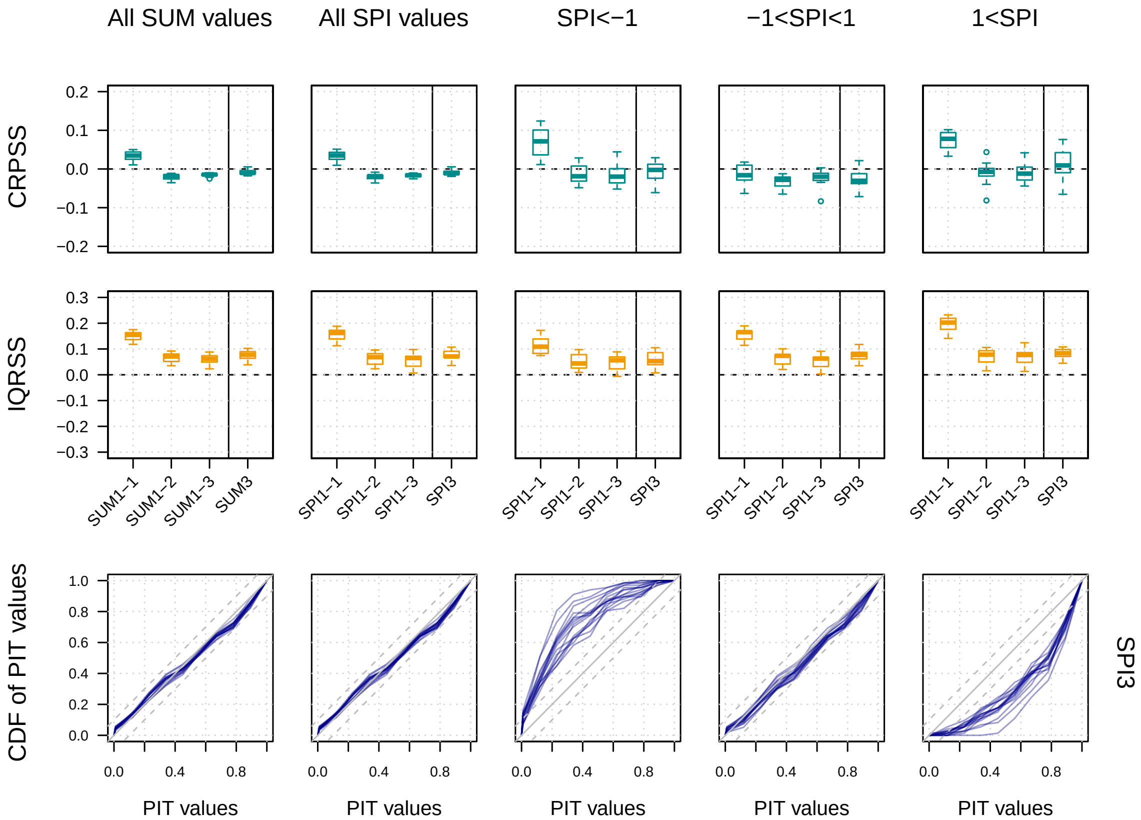



Figure 5.1: CRPSS, IQRSS and reliability of SPI forecasts and forecasts of cumulative precipitations produced from bias corrected System 4 precipitation forecasts. The reference for the skill scores is climatology. Skill scores are presented for statistics calculated over one month and three months altogether (Sum1 and Sum3; SPI1 and SPI3). PIT diagrams are presented for statistics calculated over the first three months altogether (Sum3 and SPI3). Columns correspond to scores computed for sums, SPI values, SPI values smaller than -1 (dry), SPI values within -1 and 1 (normal) and SPI values greater than 1 (wet).

The first conclusion from this figure is that all four conditionings lead to similar results. Their impact on forecasts reliability (PIT) and sharpness (IQR) is uniform over the lead times, while their impact on overall performance (CRPS) is greater at shorter lead times. Conditioning HistQ improves sharpness at most lead times (IQRSS above zero) and, for all conditioning statistics (Sum or SPI). However, as a direct result of narrower ensembles, there is a decrease in the PIT values (reliability) at most lead times (PITSS below zero). Nevertheless, the PIT diagrams at 45 days show that this decrease does not affect the overall reliability of the conditioned ensembles: they remain quite reliable (PIT values close to the diagonal line) for all conditioning statistics, especially when conditioning based on the SPI. Regarding overall performance (CRPS), the conditioning increases performance up to 15 to 30 days ahead in most catchments. Improvement is greater when traces are selected based on cumulative precipitations (Sum3 or Sum1) or SPI3 than when they are selected based on the series of SPI1 values. This improvement in overall performance in the first lead times can be attributed to the fact that the conditioning of historical streamflow takes into account the last observed streamflow. At longer lead times, the overall performance of conditioned scenarios is, in the majority of catchments, equivalent or slightly worse than that of HistQ. In one of the catchments, however, we observed improvements up to 90 days ahead. This catchment corresponds to catchment 1, in which interannual streamflow variability dominates over seasonality (cf. Chapter 1) due to a high base flow index.

We also examined the loss and gain in skill due to conditioning the ESP base ensemble. Figure 5.3 is similar to Figure 5.2. It plots the skill scores against lead time and the PIT diagram for a lead time of 45 days, but this time the reference used in the computation of the skill is ESP. Here again, the four conditionings seem to have a similar impact on performance. Conditioned streamflow forecasts appear to be as performant or slightly worse than ESP in terms of overall performance (CRPSS), for all lead times. This often translates in a gain in sharpness (IQRSS) associated with a loss in reliability (PITSS), as observed with the scenarios conditioned from the HistQ base ensemble. Some distinctions between the conditionings based on cumulative precipitations and the conditionings based on the SPI can be seen. First, conditionings based on the SPI provide more homogeneous results between catchments for all evaluation criteria. We also observe that the loss in overall performance is greater with the conditionings based on cumulative precipitations, while overall performance of the ensembles conditioned with the SPI tend to be equivalent to that of ESP. The PIT diagrams show that ensembles selected based on cumulative precipitations are not perfectly reliable, with observations too often falling below the forecast range in most catchments. Ensembles selected based on the SPI show a similar tendency, but in fewer catchments. In general, PIT values are closer to the diagonal when conditioning based on SPI values, especially with ESP_SPI3, which gives more reliable forecasts in most catchments.

Figure 5.2 and Figure 5.3 have shown that the conditionings provide quite consistent results, with a tendency to increase sharpness and maintain or just slightly decrease reliability. Overall performance is quite stable at all lead times, when conditioning the ESP base ensemble. When conditioning HistQ, the use of an additional condition on streamflows to select traces increases overall performance in the first lead times. Conditioning based on the SPI provides more consistent results between catchments and tends to produce more reliable forecasts than conditioning based on cumulative precipitations. More specifically, conditioning based on SPI3 minimizes the loss in reliability and in overall performance comparatively to the ESP ensemble. In the following paragraphs of our analysis, the quality of conditioned ensembles is further explored. The scenarios investigated were restrained to HistQ_SPI3 and ESP_SPI3.

#### Comparison of conditioned scenarios with the Sys4 base ensemble

In Figure 5.4, we compare the quality of ESP, ESP_SPI3, HistQ and HistQ_SPI3 comparatively to Sys4. Figure 5.4 is similar to Figures 5.2 and 5.3 in that it represents

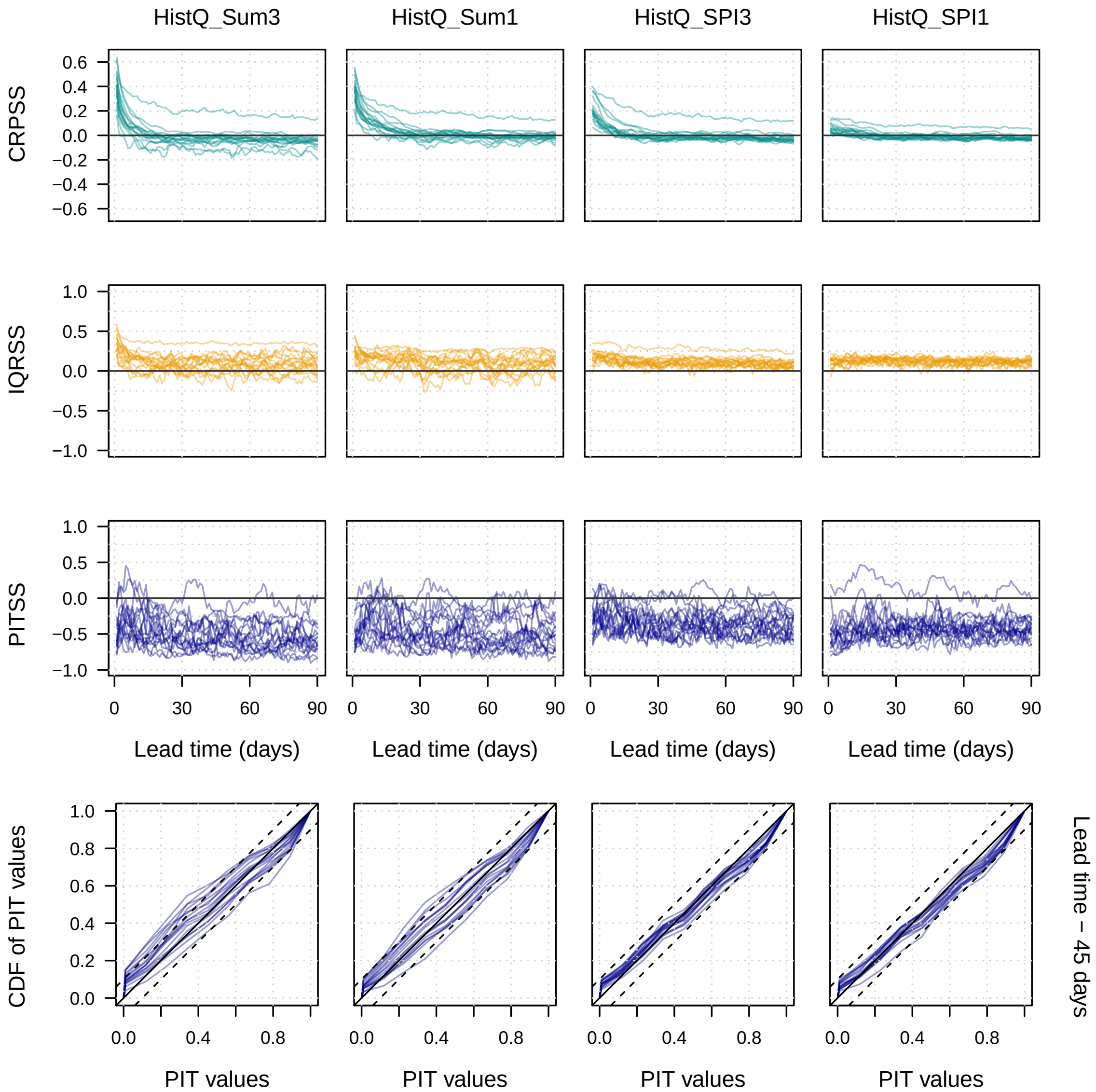



Figure 5.2: Skill scores (CRPSS, IQRSS, PITSS; first three rows) and PIT diagrams for a lead time of 45 days (last row) of the conditioned ensemble forecast scenarios: HistQ_Sum3, HistQ_Sum1, HistQ_SPI3 and HistQ_SPI1. In the skill scores, the reference forecast is the base ensemble HistQ. Each line represents one of the 16 catchments investigated.

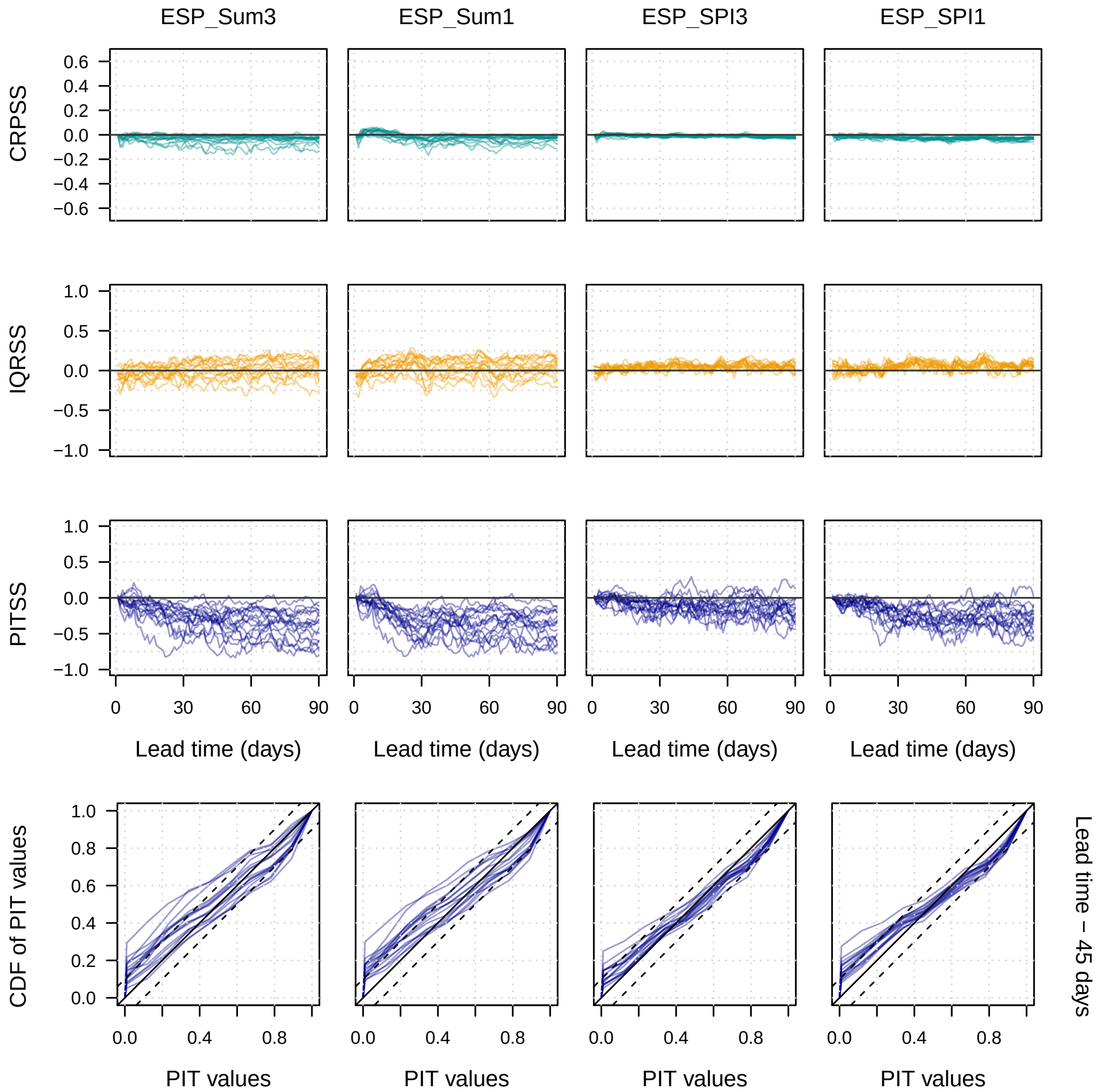



Figure 5.3: Same as Figure 5.2 but the forecast ensembles are ESP_Sum3, ESP_Sum1, ESP_SPI3 and ESP_SPI1 and the reference for the computation of the skill is ESP.

the skill in overall performance, reliability and sharpness as a function of lead time, as well as the PIT diagrams at 45 days lead.

The behaviour of ESP is very similar to that of ESP_SPI3 with respect to Sys4. Both have better overall performance than Sys4 for lead times shorter than 5 to 10 days, worse performance for lead times from 5 to 10 days and up to 20 days, and equivalent performance at longer lead times. In terms of reliability and sharpness, ESP and ESP_SPI3 are overall more reliable than Sys4 but not as sharp, though ESP_SPI3 becomes equivalent to Sys4 for lead times longer than 45 days. The PIT diagrams show that ESP and ESP_SPI3 are visually equivalent in terms of reliability, though the previously observed tendency of observations falling below the forecast range persists in a few catchments. This tendency may not be caused by precipitation inputs but by the hydrological model.

If we now look at ensembles based on historical streamflows, we observe that HistQ performs worse than Sys4, at least for lead times shorter than 50 days. Even though HistQ is more reliable than Sys4, it is not as sharp, especially for lead times shorter than 30 days. HistQ_SPI3 also has lower overall performance than Sys4 but the gap in performance is reduced for lead times shorter than 15 days. HistQ_SPI3, following HistQ characteristics, provides forecasts that are more reliable than Sys4, except at long lead times in some catchments. Contrary to HistQ, conditioning allows HistQ_SPI3 to be as sharp as Sys4 for horizons longer than 30 days. The reliability of HistQ and HistQ_SPI3 is confirmed by their PIT diagrams. These diagrams also show that ensembles based on historical streamflows (HistQ) are more reliable than ensembles based on precipitation climatology (ESP).

**Overall comparison of base and conditioned ensembles**

The objective now is to see whether we succeeded in benefiting from the reliability of climatology and the sharpness of Sys4 when conditioning ensemble forecast scenarios. Figure 5.5 proposes a simultaneous evaluation of the reliability (PIT area) and sharpness (IQR) of ESP_SPI3 and HistQ_SPI3. For a given catchment, lead time and reference, the skill in reliability is plotted against the skill in sharpness. Each point corresponds to a catchment, each column corresponds to a lead time and each row corresponds to a forecast ensemble. Two references are chosen for each ensemble: ESP_SPI3 is evaluated against ESP and Sys4, and HistQ_SPI3, against HistQ and Sys4. Each reference is identified by its colour and shape (cf. legend). If a point is located in the upper left part of the graph, the conditioned ensemble is more reliable but less sharp than the reference (indicated by the colour of the point) in the corresponding catchment. Reversely, if a point is located in the lower right part, the conditioned ensemble is sharper but less reliable than the reference. At best, both reliability and sharpness are improved, and points are located in the upper right part of the graph. At worst, both reliability and sharpness are deteriorated with respect to the reference,

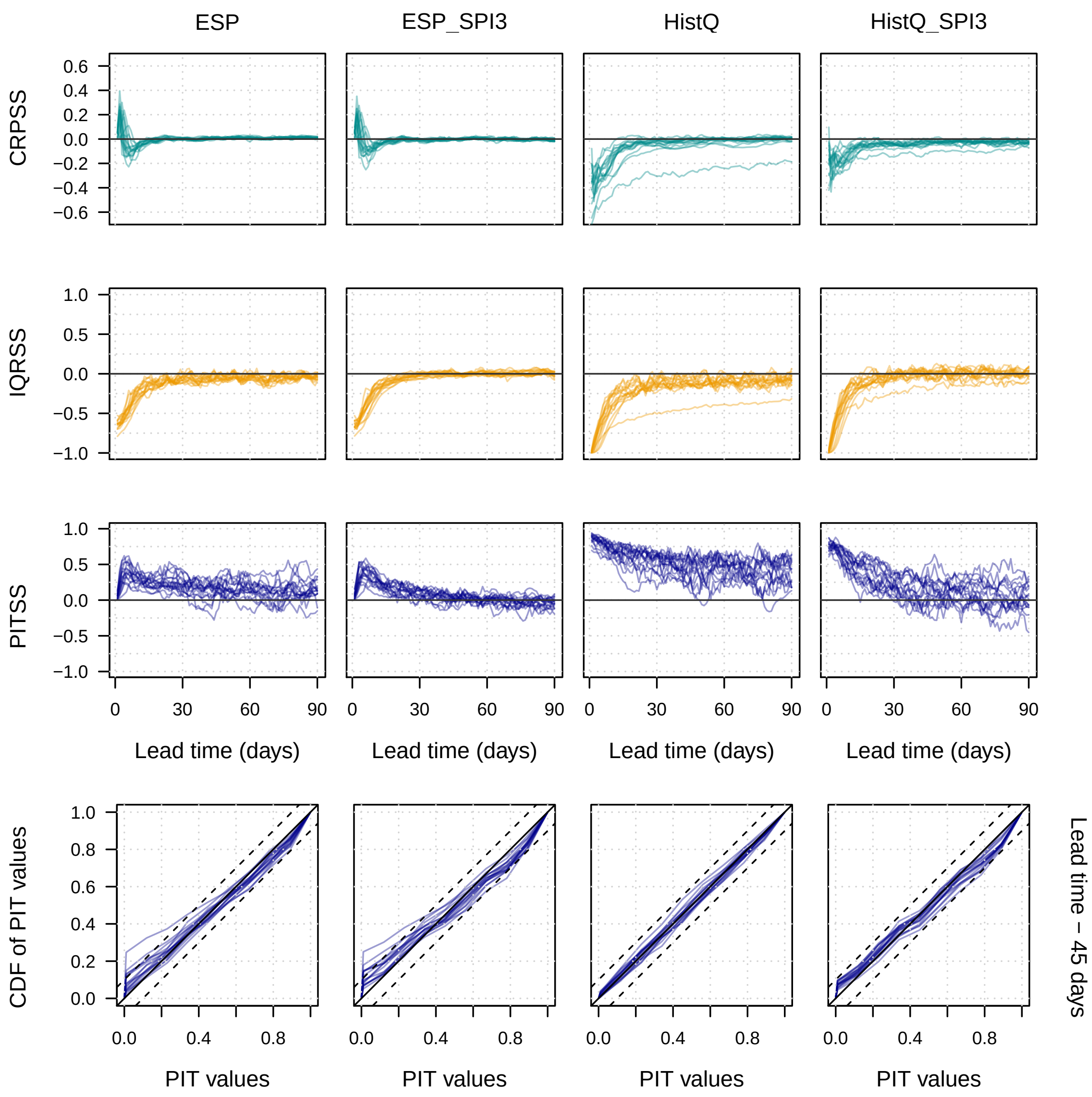



Figure 5.4: Same as Figure 5.2 but the forecast ensembles are ESP, ESP_SPI3, HistQ and HistQ_SPI3 and the reference for the computation of the skill is Sys4.

and points are located in the bottom left part of the graph.

Overall, we can observe that the conditioning tends to have more impact on reliability than on sharpness (y-axes extend further than x-axes). Also, conditioned ensembles are generally more reliable but less sharp than Sys4, and sharper but less reliable than the ensembles they are selected from. More specifically, we observe that:

- For a lead time of 10 days, ESP_SPI3 and HistQ_SPI3 can be more reliable and sharper than the ensembles they are selected from. This applies to most catchments with ESP_SPI3, and to at least two catchments with HistQ_SPI3;
- For a lead time of 30 days, fewer catchments benefit from a gain in both reliability and sharpness. The loss in sharpness and the gain in reliability with respect to Sys4 are less pronounced than for a lead time of 10 days. For instance, the maximum PITSS values for ESP_SPI3 move from 0.45 (for a lead time of 10 days) to 0.2 (for a lead time of 30 days) and those for HistQ_SPI3 move from 0.7 to 0.4. The gain in sharpness and the loss in reliability with regard to ESP and HistQ remain in the same ranges as observed for a lead time of 10 days;
- For a lead time of 90 days, the gain of ESP_SPI3 over Sys4 is further reduced and varies with the catchment. The same is observed to a lesser extent for HistQ_SPI3, even though a positive impact of the conditioning on the reliability can still be observed in several catchments. At this lead time, both ESP_SPI3 and HistQ_SPI3 provide forecasts that are still sharper, yet less reliable, than the climatology they are selected from.

Figure 5.5 can also be interpreted in terms of distance between approaches. Indeed, the (0,0) coordinate corresponds to the location of the references. From this perspective, we observe that ESP_SPI3 is closer to ESP than to Sys4 for a lead time of 10 days. But as the lead time increases, ESP_SPI3 becomes closer to Sys4 and further apart from ESP. The proximity between ESP_SPI3 and Sys4 at longer lead times can be attributed to the conditioning itself. The proximity between ESP_SPI3 and ESP and their distance to Sys4 at shorter lead times may be attributed to the repercussion of the initialization of the climate model on Sys4 streamflow forecasts. Indeed, since initial hydrological conditions are the same for the three forecast ensembles, differences are caused by meteorological forcings only. The main difference between System 4 precipitations and climatology at such lead times is the initialization of the GCM, which leads to sharper System 4 forecasts in the first lead times. Similarly, we observe that HistQ_SPI3 becomes closer to Sys4 as the lead time increases due to conditioning. However, its distance to HistQ remains the same at all lead times. This distance is probably due to the use of previous streamflow conditions as a conditioning criterion within HistQ. Therefore the three ensembles, HistQ, HistQ_SPI3 and Sys4 are equally distant in the first lead times.

Table 5.2 proposes a ranking of the different ensembles investigated based on overall performance, reliability and sharpness and for different lead time ranges: from 10 to

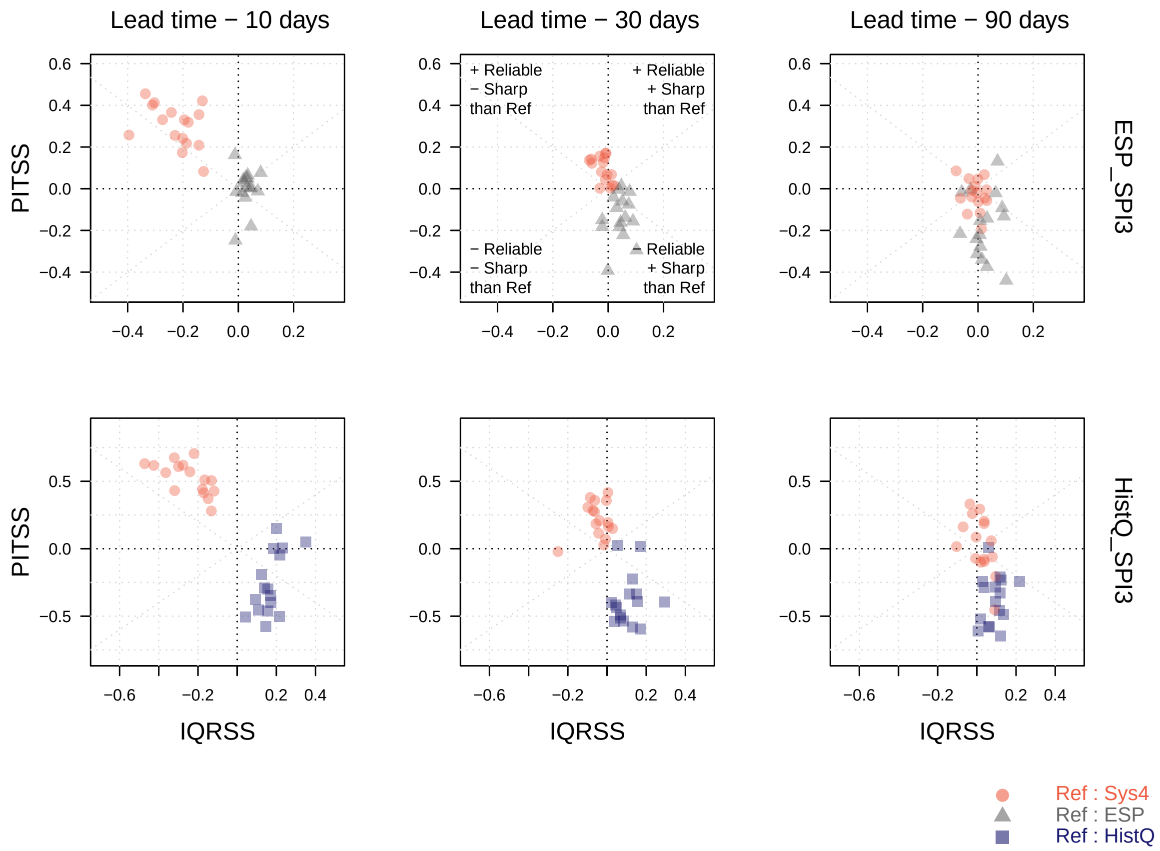



Figure 5.5: PITSS (reliability) versus IQRSS (sharpness) for ESP_SPI3 (upper row) and HistQ_SPI3 (lower row), and lead times of 10, 30 and 90 days (columns). ESP_SPI3 is compared to Sys4 (red) and ESP (grey), while HistQ_SPI3 is compared to Sys4 (red) and HistQ (blue). Each point represents one of the 16 catchments.

30 days, from 30 to 60 days and from 60 to 90 days. The rankings are based on the visual evaluation of Figure 5.4. The mean rank is calculated as the mean of the ranks obtained in the nine cells of the 3x3 table. Overall performance, reliability and sharpness are thus considered equivalent in this final ranking. Note that this may not be representative of operational expectations, since, in operational conditions, one could choose to emphasize one of the three characteristics over the others.

Based on Table 5.2, we can say that, if one seeks an overall performing ensemble with 10 to 30 days lead, one would use Sys4. For horizons longer than 30 days, ESP and ESP_SPI3 offer good alternatives. If one seeks, above all, a reliable ensemble, one could simply use HistQ, ESP, or even HistQ_SPI3 for lead times shorter than 30 days. However, for ensembles that are both sharp and reliable, and for horizons longer than 30 days, one could turn to the following ensembles: ESP_SPI3 for an emphasis on sharpness, or HistQ_SPI3 for an emphasis on reliability.

Table 5.2: Rankings of the Sys4, ESP_SPI3, ESP, HistQ_SPI3 and HistQ streamflow ensembles, as evaluated by three evaluation criteria (in rows) and three lead time ranges (columns). The mean rank is calculated for each ensemble and is a simple mean of the ranks obtained by this ensemble in the nine cells of the 3x3 table.

| | 10-30 days lead | 30-60 days lead | 60-90 days lead |
|---|---|---|---|
| CRPS | 1. Sys4 | 1. Sys4 | 1. Sys4 |
| | 2. ESP_SPI3 | 1. ESP_SPI3 | 1. ESP_SPI3 |
| | 2. ESP | 1. ESP | 1. ESP |
| | 4. HistQ_SPI3 | 4. HistQ_SPI3 | 4. HistQ_SPI3 |
| | 5. HistQ | 5. HistQ | 4. HistQ |
| IQR | 1. Sys4 | 1. Sys4 | 1. Sys4 |
| | 2. ESP_SPI3 | 2. ESP_SPI3 | 1. ESP_SPI3 |
| | 3. ESP | 3. HistQ_SPI3 | 1. HistQ_SPI3 |
| | 4. HistQ_SPI3 | 4. ESP | 4. ESP |
| | 5. HistQ | 5. HistQ | 5. HistQ |
| PIT area | 1. HistQ | 1. HistQ | 1. HistQ |
| | 2. HistQ_SPI3 | 2. ESP | 2. ESP |
| | 3. ESP | 3. HistQ_SPI3 | 3. HistQ_SPI3 |
| | 4. ESP_SPI3 | 4. ESP_SPI3 | 4. ESP_SPI3 |
| | 5. Sys4 | 5. Sys4 | 4. Sys4 |

| | |
|---|---|
| Mean Rank | 1. Sys4 : 2.22 |
| | 2. ESP_SPI3 : 2.33 |
| | 3. ESP : 2.44 |
| | 4. HistQ_SPI3 : 3.11 |
| | 5. HistQ : 3.55 |

### 5.3.2 Statistical evaluation of low flows

We assess the performance of the ensemble forecast scenarios to forecast summer low flows and drought risks. Many ways of characterizing severe low flows and droughts exist in the literature (Tallaksen *et al.*, 1997; Smakhtin, 2001; WMO, 2008; Mishra and Singh, 2010). In the following, the low-flow variables considered are the low-flow duration and deficit volume, both computed for the $80^{th}$ exceedance percentile. In this section, only forecast horizons falling within the May to October period are considered.

#### Capacity of the ensembles to forecast low-flow events

The capacity of the different systems to discriminate between low-flow events and non-events is assessed. Figure 5.6 presents the ranges of the Area Under the Curve (AUC) of the ROC diagram obtained from the five ensemble forecast scenarios, namely Sys4, ESP_SPI3, ESP, HistQ_SPI3 and HistQ. AUC values were assessed for the $80^{th}$ exceedance percentile and for lead times of 10 days, 30 days and 90 days. Each boxplot gathers the AUC values obtained in the 16 catchments. The letters below the boxplots result from the Friedman test (see for instance Lowry, 1999). This test consists in considering catchments as judges of the five methods. The test, which is based on rankings as evaluated by the catchments, assesses whether the methods are significantly different by assessing whether their rankings resemble a random shuffling. Based on this test, two boxplots sharing a letter at a given lead time are not significantly different.

Results show that all ensembles but HistQ are very close in terms of discrimination. As expected, their performance decreases as the lead time increases, except for HistQ, whose discrimination does not vary with the lead time. For all lead times, ESP significantly provides the best discrimination with most AUC values superior to 0.88. ESP_SPI3 and Sys4 are tied in terms of discrimination and appear as second best, with most AUC values greater than 0.82. HistQ_SPI3 is also very close to the performances of Sys4 and ESP_SPI3, but does not score as high as they do, especially for longer lead times. Nevertheless, HistQ_SPI3 mostly provides AUC values larger than 0.81. HistQ always provides AUC values between 0.8 and 0.9, except in Catchment 1, in which we have seen that this ensemble forecast has very low performances. Overall, ensembles based on hydrological modelling (Sys4, ESP and ESP_SPI3) provide the best skills in discrimination, at least for lead times shorter than 90 days, probably because they take into account initial hydrological conditions. We note that all these conclusions are also valid when the 60$^{\text{th}}$ exceedance percentile is used as threshold (not shown).

#### Capacity of the ensembles to forecast low-flow variables

We now compare the forecast systems based on variables of interest for water management during low flows, that is the weekly deficit duration and the weekly deficit volume. The weekly deficit duration corresponds to the number of days per week during which the daily streamflow is below a given threshold. The weekly deficit volume corresponds to the flow volume per week below this threshold. Figure 5.7 presents the PIT areas obtained with Sys4, ESP_SPI3, ESP, HistQ_SPI3 and HistQ when forecasting the weekly number of days below the 80$^{\text{th}}$ exceedance percentile. Boxplots represent the range of PIT areas obtained over the catchment set. Results are presented for lead times of two weeks, five weeks and twelve weeks (columns). Again, letters represent the results of the Friedman test. Two boxplots that share a letter are not significantly different. Figure 5.8 proposes the same evaluation for the weekly streamflow deficit volume below the 80$^{\text{th}}$ exceedance percentile.

Figure 5.7 shows that the difference between the five ensembles is very tenuous when forecasting the deficit duration. For instance, all lower and upper quartiles of Sys4, ESP_SPI3, ESP and HistQ_SPI3 are included in the $[0.01, 0.08]$ interval of PIT area values, regardless of the lead time. Overall, ESP, ESP_SPI3 or HistQ_SPI3 perform best to forecast the deficit duration. All ensembles but HistQ provide quite reliable forecasts (PIT area values close to zero). HistQ_SPI3 is significantly the

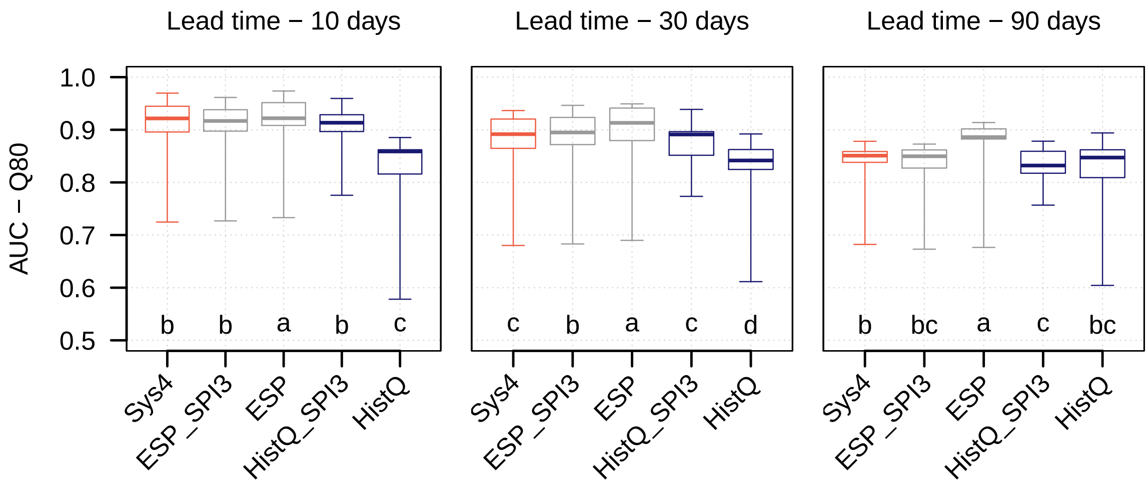



Figure 5.6: Ranges of the Area Under the Curve (AUC) of the ROC diagram based on the 80th exceedance percentile for each of the five selected ensemble forecasts (Sys4, ESP, HistQ, ESP_SPI3, HistQ_SPI3). Boxplots gather the AUC values for the 16 catchments. The boxes extend to the 25th and 75th percentiles and the whiskers, to the data extremes. Graphs are presented for 10-day, 30-days and 90-day lead times (columns). The letters below the boxplots result from the Friedman test. For a given lead time, two boxplots sharing a letter are not significantly different.

best performing ensemble for a lead time of two weeks. For a lead time of five or twelve weeks, both ESP and HistQ_SPI3 are the best options. The analysis of the corresponding PIT diagrams (not presented) showed that all ensembles are equivalently reliable, except for HistQ, which systematically overestimates the deficit duration.

The gap between ensembles widens when looking at the deficit volume (Figure 5.8). For lead times of two and five weeks, ESP and ESP_SPI3 provide consistently reliable ensembles, and lower PIT areas than the others. For a lead time of twelve weeks, ESP_SPI3, along with Sys4 and HistQ_SPI3, provide the most reliable ensembles. The corresponding PIT diagrams (not presented) showed that HistQ_SPI3 tends to underestimate durations at all lead times. Ensembles issued with hydrological modelling also slightly underestimate the deficit volume at long lead times. Overall, ESP_SPI3 systematically appears to be one of the best options to forecast deficit volumes.

### 5.3.3 Drought impact evaluation

Figure 5.9 illustrates the case of the 2003 drought with the streamflow forecasts issued on July 1st 2003 for the three months ahead, in catchment 5, the Azergues at Lozanne. Each column represents the graphs obtained with one of the five ensemble forecasts (Sys4, ESP_SPI3, ESP, HistQ_SPI3 and HistQ). The upper row presents the graphical

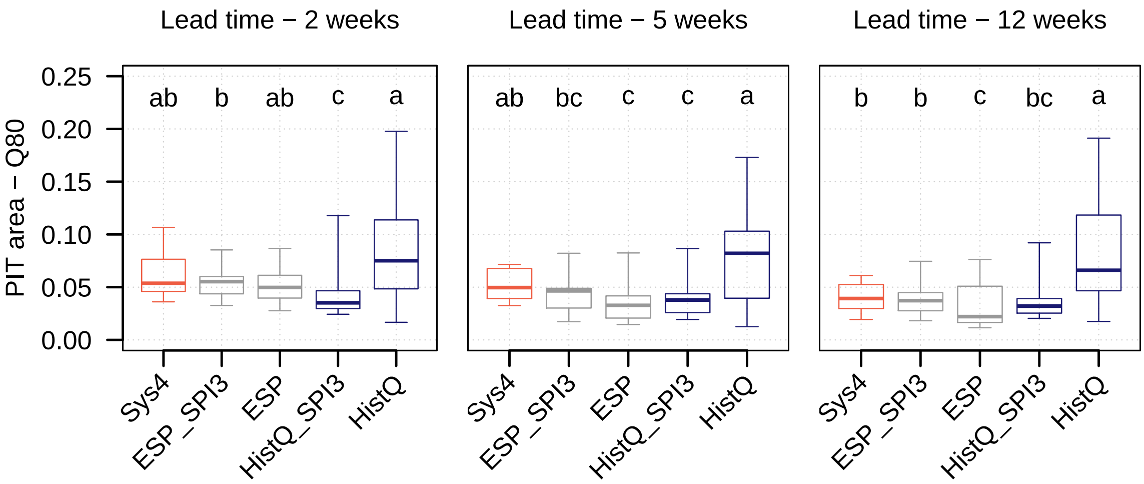



Figure 5.7: Same as Figure 5.6 but for PIT area ranges computed for deficit duration. Ranges are represented by boxplots which gather the PIT areas for the 16 catchments. Graphs are presented for lead times of two weeks, five weeks and twelve weeks (columns).

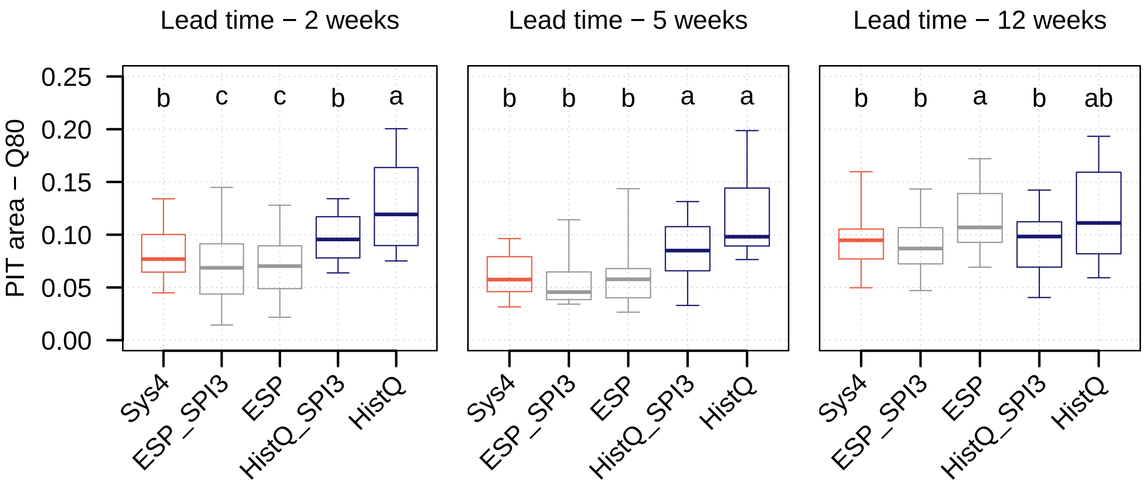



Figure 5.8: Same as Figure 5.7 for deficit volume.

representation we propose to assess drought risks based on the ensemble forecasts. The graphs represent the deficit duration against the deficit volume, both computed based on the 80$^{\text{th}}$ exceedance percentile. The graph is divided into 49 boxes corresponding to possible combinations and ranges of deficit volumes and durations. The colour within each of these boxes indicates the percentage of ensemble members that falls within each box. Coloured dots represent the observation and two references: the 1976 drought and the historical mean duration and deficit volume over the forecast period (climatology). The lower row presents the corresponding hydrographs over the forecast period. The black line represents the observed streamflow, the red line represents the 80$^{\text{th}}$ exceedance percentile and the blue lines represent the members of the ensemble forecast.

All ensembles produce similar patterns, but with different probabilities. The maximum probability is obtained with HistQ_SPI3 with 60 % of the ensemble members falling in the same cell. Ensembles based on hydrological modelling reach maximum probabilities of 20 to 30 %, and HistQ does not exceed a probability of 14 %. These colours translate in a way the sharpness of the ensemble forecasts. The objective with the graph is to have a maximum of darker cells close to the observation (represented by the black dot). We observe that the graph obtained with HistQ puts equivalent weights to a wide range of scenarios indicating no risk to high risks. This ensemble thus conveys little information to assess drought risks. HistQ_SPI3, as opposed to HistQ, offers a more confident risk assessment with the highest forecast probabilities and only three coloured cells. Eighty percent of the forecast members indicate a drought equivalent or more severe than that of 1976. The high probability may be explained by the fact that SPI forecast members were often best represented by the SPI observed in the same driest year (as suggested by the hydrographs), possibly 1976.

The ESP forecast provides a wider view of risks, with higher probabilities located in the upper right part of the graph, and small probabilities of having months with moderately dry conditions. ESP is able to forecast a more severe event than observed during the 1976 drought. This good performance can only be attributed to the initial hydrological conditions since ESP does not have any information on future precipitations apart from climatology. Conditioning ESP (ESP_SPI3) slightly reduces the number of coloured cells with slightly higher probabilities in some of the upper right cells. The difference between ESP and ESP_SPI3 is clear when looking at the hydrographs. With ESP_SPI3, the number of high-flow peaks is reduced.

Sys4 also provides a quite good risk assessment since only upper right cells are coloured. While ensembles based on hydrological modelling, i.e. ESP, ESP_SPI3 and Sys4, are limited by the capacity of the model to reproduce small low-flow variations and thus slightly underestimate the deficit volume, ensembles based on historical streamflows are limited within the range of past precipitation and streamflow scenarios. This highlights the fact that the studied methods, and here specifically Sys4, ESP_SPI3 and HistQ_SPI3, have different limitations, but also different assets. We have illustrated their performances to forecast a given drought event in France. We should however keep in mind that different contexts might penalize or favour different methods.

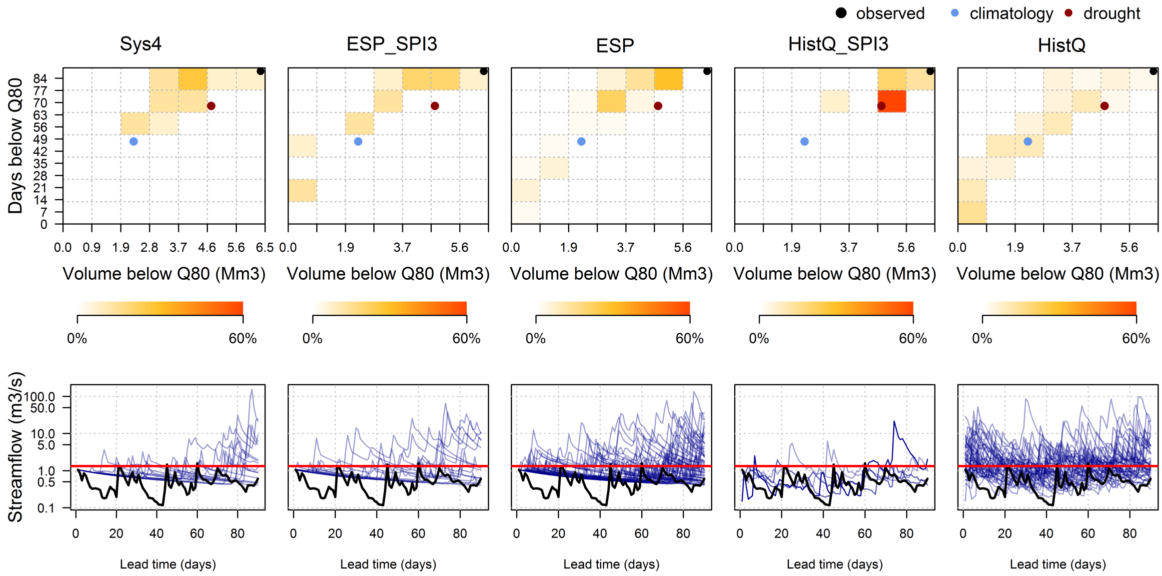



Figure 5.9: Risk graphs presenting the probabilities of deficit duration versus deficit volume based on the 80$^{th}$ exceedance percentile (upper row) and corresponding hydrographs (lower row). The maximum probability varies with the ensemble and the situation and is indicated in the colour scale. The black point corresponds to the observation, the dark red dot to the drought of 1976 and the blue dot to the mean duration and deficit volume observed in past streamflows. Each column corresponds to one of the five ensemble forecasts. Forecasts were issued for the Azergues at Lozanne (catchment 5) for the period running from July 1$^{st}$ to September 30$^{th}$ 2003.

## 5.4 Conclusion

We have investigated the potential of seasonal streamflow forecast ensembles built by conditioning precipitation climatology and historical streamflows based on precipitation indices derived from ECMWF System 4 (GCM) forecasts. In a first step, the performance of the conditioned ensembles was assessed in terms of overall performance, sharpness and reliability for lead times up to 90 days. Here are the main conclusions from this comparison:

- Selecting traces within precipitation climatology or historical streamflow generally improved sharpness and decreased reliability. Conditioning based on the SPI provided more consistent results between catchments and more reliable forecasts than conditioning based on cumulative precipitations. More specifically, conditioning based on SPI3 improved overall performance as compared to historical streamflow and maintained overall performance as compared to precipitation climatology used as input to a hydrological model, while providing reliable forecasts.

- A simultaneous evaluation of the sharpness and reliability of the conditioned ensembles showed that conditioning led to ensembles that were more reliable and less sharp than streamflow forecasts generated from System 4 precipitations, and less reliable and sharper than the ensembles they were selected from. Also, the conditioned ensembles benefit from the information of either precipitation climatology or historical streamflows at shorter lead times and from the information of GCM-based forecasts at longer lead times.
- Ensembles selected from precipitation climatology and historical streamflow offer a good compromise between sharpness and reliability, with an emphasis on sharpness with precipitation climatology, and an emphasis on reliability with historical streamflows.

The performance of the ensembles in forecasting low-flow events and low-flow variables was then evaluated, with an illustration on the 2003 drought in France. Their capacity to discriminate between low-flow events and non-events and their capacity to forecast streamflow deficit volume and duration, as defined by the $80^{\mathrm{th}}$ exceedance percentile, were assessed. The main conclusions from this second evaluation are:

- Forecast ensembles using hydrological modelling provided better discrimination than ensembles based on historical streamflows. Nevertheless, all forecast ensembles provided good performance, except for historical streamflows for lead times shorter than a month.
- Even though differences between ensembles are tenuous when forecasting low-flow duration, the gap widens when forecasting deficit volume. The ensemble selected within precipitation climatology systematically provides some of the most reliable deficit volume forecasts.
- Lastly, a graphic representation of the forecast drought risks was proposed. It was illustrated with the 2003 drought. We showed that, for this drought event, ensemble forecasts based on conditionings within climatologies (either precipitation or streamflow climatology) provided good drought risk assessment.

We investigated conditionings within climatology solely based on past precipitations and catchment conditions. SPI values were computed after an aggregation of System 4 precipitation forecasts at the catchment scale, therefore the conditioning and the spatial aggregation were independent. Further investigations could assess the potential of this method for spatial downscaling of System 4 precipitation forecasts.

One important parameter to forecast low flows and droughts is the temperature. A more advanced approach would consist in selecting past scenarios based on the SPEI (Standardized Precipitation-Evapotranspiration Index) calculated from seasonal precipitation and temperature forecasts.

Finally, other types of combinations can be found in the literature and could be investigated along with the proposed conditionings. As an example, Werner *et al.* (2005) or Shukla *et al.* (2012) have investigated the use of medium-range weather forecasts to improve long-range forecasting. These approaches are based on the fact that short-term events are well forecast by short-term to medium-term forecasts issued by GCMs and that the benefit from medium-range forecasts can be extended to longer lead times through the inertia of a catchment.

# III

# Reservoir Management

# 6

# Risk assessment tool for the Arzal reservoir during low-flow periods

**Résumé**

Dans les chapitres précédents, nous avons évalué la qualité des prévisions saisonnières de débits en France et illustré leur potentiel pour la prévision des étiages. En effet, les outils d'évaluation des risques peuvent tirer profit des prévisions à long-terme, afin d'aider la gestion de la ressource en eau ou la prise de décision des lâchers de réservoirs pour les mois à venir. Le barrage d'Arzal se situe à l'embouchure de la Vilaine, juste avant qu'elle ne se déverse dans l'Océan Atlantique. Sa situation en fait une barrière entre l'eau salée de l'océan et l'eau douce de la Vilaine, créant une retenue d'eau douce de 50 $Mm^3$. Cette retenue est utilisée pour l'approvisionnement en eau potable et l'irrigation. De plus, le barrage permet la navigation et la migration des poissons entre l'océan et la rivière. Le réservoir a donc un rôle clé dans la gestion de la ressource en eau, notamment dans le bassin aval de la Vilaine. L'objectif de ce chapitre est d'analyser le potentiel d'un outil d'évaluation des risques en contexte de basses eaux pour le réservoir d'Arzal.

Nous commençons par décrire plus en détail le réservoir multi-usage d'Arzal et sa gestion, et nous présentons les données de gestion disponibles ainsi que le modèle de bilan d'eau du réservoir que nous avons mis en place. Une analyse préliminaire des données hydrométéorologiques et de gestion du réservoir nous permet de mieux comprendre les stratégies de gestion du barrage et l'importance relative des entrées et sorties du réservoir. Cette analyse montre que le modèle de bilan est très réactif aux entrées d'eau provenant du bassin amont, ainsi qu'aux sorties des vannes et volets qui permettent aux gestionnaires de contrôler le niveau du réservoir. Le modèle est ensuite utilisé pour reproduire les niveaux de réservoir sur la période de Mai à Octobre, i.e., la période de basses eaux dans le bassin de la Vilaine.

Enfin, le modèle est utilisé pour estimer le risque de ne pas pouvoir maintenir un niveau minimum dans le réservoir entre Mai et Octobre. Pour évaluer ce risque, des prévisions saisonnières de débits (ESP), étudiées dans les chapitres précédents, sont utilisées en entrée du modèle de bilan. Les sorties des vannes et volets sont optimisées pour maintenir un niveau objectif dans le réservoir, tandis que les sorties liées aux autres usages sont supposées constantes et maximales. Les prévisions de niveaux de réservoir ainsi obtenues nous permettent d'identifier le risque de franchir le seuil minimal dans le réservoir et de déterminer le nombre de jours et le nombre de membres de la prévision d'ensemble atteignant ce seuil. Ces variables d'intérêt pour la gestion du risque sont réunies dans un graphique d'évaluation du risque proposé à la fin du chapitre. Les graphiques ont été produits retrospectivement pour les périodes de basses eaux de 2005 à 2010 afin d'évaluer leur potentiel en conditions opérationnelles.

**Abstract**

Risk assessment tools can help anticipate and corroborate decisions in reservoir management. The Arzal dam in Brittany, France, is located at the outlet of the Vilaine river basin, just before the Atlantic Ocean. Being a barrier to the salt water of the ocean, it creates a 50 $Mm^3$ freshwater reservoir that serves multiple water uses: drinking water, flood control, irrigation, sailing and fish by-passing. The reservoir thus plays an essential role in the regional water management system, and more specifically in the lower Vilaine river basin. In this chapter, we describe the multi-purpose Arzal reservoir, the elements of the Arzal dam, the available reservoir management data and a water balance model of the reservoir that we have developed.

A preliminary analysis investigates the management of the dam based on reservoir management data and hydrometeorological data, and the relative importance of inflows and outflows. This analysis shows that the Arzal reservoir model is very reactive to inflows from the upstream catchment and to outflows through the shutters and sluice gates of the dam. The reservoir water balance model is then used to assess the potential to reproduce reservoir levels between May and October, i.e. the low-flow period in the Vilaine river basin.

The model is also used to estimate risks in the Arzal reservoir. In the risk assessment framework, the model is run with seasonal ensemble inflow forecasts and constant maximum outflows for the different water uses. Outflows through the gates and shutters are optimized to maintain a constant objective level in the reservoir. The ensemble of reservoir level simulations thus obtained allow us to estimate the risk of not being able to maintain a constant level in the reservoir throughout the May to October period. This risk is quantified in terms of number of days spent below a minimum acceptable reservoir level and in terms of the number of ensemble members reaching this low threshold. These variables were summarized in a risk assessment graph proposed at the end of the chapter. Graphs are produced retrospectively for the May to October periods of 2005 to 2010 in order to evaluate the potential of these graphs in operational conditions.

## 6.1 Introduction

The Arzal dam (or Arzal-Camoël dam) is located at the mouth of the Vilaine river basin, just before the Atlantic Ocean (Figure 6.1). It was built in 1970 with the objective to reduce flood risks upstream the Vilaine estuary and has been managed by IAV since then (*Institution d'Aménagement de la Vilaine*). Being a barrier between the freshwater of the Vilaine River and the saline water of the Atlantic Ocean, the dam prevents water from the Atlantic tides from entering the river and thus limits the risks of flooding from the combination of high tides and high flows coming from the Vilaine river.

Figure 6.1: The Arzal dam seen from upstream (top), from downstream with the lock in the foreground (bottom left) and from downstream with a lateral view of the dam (bottom right).

The dam creates a 50 $Mm^3$ reservoir of freshwater that is used for drinking water supply in the region. A water treatment plant located at Férel with a maximum treatment capacity of 100,000 $m^3$ per day supplies 15 to 20 $Mm^3$ of freshwater annually. During the summer season, up to one million inhabitants are supplied with drinking water from this plant. The Arzal reservoir is the surface water mass that is the most impacted by water withdrawals in the Vilaine basin (SAGE, 2015a). Indeed, Arzal is the main withdrawal point with more than 15 $Mm^3$ of water withdrawn per year from surface water and up to 80,000 $m^3$ of water pumped per day for drinking water supply.

In addition to being a key element in flood risk reduction and drinking water supply, the dam is located in a key area for leisure boating. About 16,000 boats cross its lock each year to navigate from the Vilaine to the ocean, and vice versa, mainly during the summer season. The dam also allows for fish migration from one side of the dam to the other by means of a fish pass and a fish ladder, in operation since autumn 1995.

In summer, during low flows, the management of the dam focuses, first and foremost, on supplying drinking water to the local and tourist population, but also on enabling lock openings for leisure boat sailing. An adequate water quantity and a good water quality in the reservoir are required to make possible and minimize the costs of drinking water production. A minimum level in the reservoir is also required to facilitate boats getting in and out of the lock and to ensure the correct functioning of the pumps that send water from the reservoir to the water treatment plant.

However, as leisure boating and lock openings intensify, the intrusion of saline water in the reservoir becomes more frequent, degrading water quality. These intrusions combined with lower water inflows during dry periods in summer, can threaten both the water quantity and the water quality in the reservoir. As a result, during the summer season, the management objectives for the Arzal reservoir consist in ensuring adequate levels in the reservoir to preserve the estuarine environment and guarantee priority uses, while limiting and evacuating saline water intrusions so as to preserve the quality of the freshwater coming from the Vilaine river.

Table 6.1 presents the threshold reservoir levels that constrain the management of the reservoir and some consequences when these are not reached. Reservoir levels are expressed in meters NGF (*Nivellement Général de la France*). The NGF is based on a tide gauge located in Marseille and defines a common reference level over France. Levels expressed in meters NGF are computed relative to this reference level. When the management objectives cannot be met, measures such as limiting lock openings or preventively increasing the reservoir level can be taken. However, while the former may cause tensions with boat owners, the latter may cause flooding of protected marshlands upstream the reservoir and tensions with farmers of the lower part of the Vilaine due to losses of harvest during the haymaking season.

Table 6.1: Threshold reservoir levels in the Arzal reservoir and associated consequences or management constraints.

| Thresholds | Consequences or management constraints |
|---|---|
| $N > 2.10$ m NGF | Flooding of nearby fields and swamps |
| $N < 1.80$ m NGF | Difficulties in letting boats in and out of the lock |
| $N < 1.60$ m NGF | No navigation |
| $N < 0.80$ m NGF | Navigation is prohibited |
| $N < 0$ m NGF | Difficulties in pumping for drinking supply |

The general objective of the work presented in this chapter is to contribute to the development of a water management tool for the Arzal reservoir to support operations during low flows. The managers of the Arzal dam expressed the need for drought risk outlooks over the period from May to October, as well as for a simple reservoir model that could be used to contemplate several management scenarios and serve as a basis for discussion with stakeholders. They expect to be able to better inform users on future risks, limit tensions between stakeholders, and, when necessary, take preventive actions to reduce the risks.

In this chapter, we first present the management of the dam in more details and the different elements subject to decision-making during the May to October period. The management data (reservoir inflows and outflows) available to this study are introduced. A simple reservoir balance model is proposed and its application for the simulation and the forecasting of reservoir levels is presented. An in-depth analysis of the data and the simulations is carried out. This analysis then guides the design of a risk assessment tool presented at the end of the chapter. Lastly, we conclude on the results of the work and discuss the perspectives for further developments and operational use of the proposed tool.

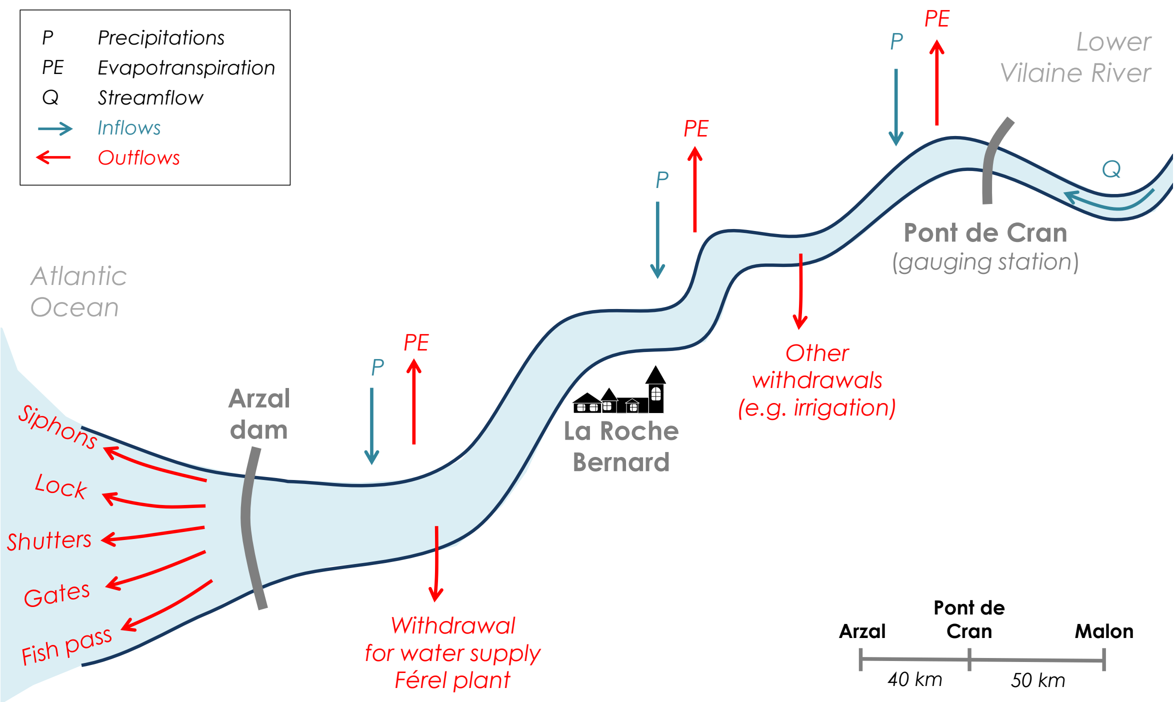



Figure 6.2: Schematic drawing of the downstream part of the Arzal reservoir between the Pont de Cran (bridge) and the Arzal dam. The reservoir extends upstream, up to the Malon lock. This lock is located 50 km upstream the Pont de Cran, and 90 km upstream the Arzal dam.

## 6.2 Management of the dam and data

To address the objectives of this chapter, it is necessary to fully understand the elements subject to decision-making in the management of the dam, how they relate to each other and how they are operated during low flows. In this section, we present these elements, the associated available data and the streamflow forecast data used to build and evaluate the reservoir balance model of the Arzal dam.

### 6.2.1 Elements of the dam

The dam is operated so as to reach the objectives listed in the previous section. Figure 6.2 provides an overview of the different elements of the dam, with a schematic drawing of the Arzal forebay and some key locations. These elements are described hereafter.

Five sluice gates and five shutters are operated to maintain a constant level in the reservoir. They let the water flow out of the reservoir and into the ocean (Figure 6.3). Shutters act as an overflow outlet for relatively small quantities of water. Sluice gates, however, open from the bottom of the reservoir and can release large volumes of water. They are thus preferred in case of flood risks, but also to evacuate the denser saline water at the bottom of the reservoir. The level of the Vilaine river is relatively constant, whereas the level of the ocean keeps changing due to the tides. At high tide, when the level of the ocean rises above the level in the reservoir, sluice gates are closed and shutters are positioned upward so as to prevent saline intrusions. At low tide, when the level of the ocean falls below the level in the reservoir, sluice gates or shutters can be opened to evacuate excess water.

Figure 6.3: Pictures of a shutter in operation (left) and a shutter and upper gate taken out of the dam for maintenance (right).

The lock enables boats to sail between the Vilaine river and the Atlantic Ocean (Figure 6.4). Lock openings are scheduled in advance based on tide coefficients. Each opening causes saline water intrusions in the reservoir. When inflows to the reservoir are low, restrictions on lock openings can be imposed to preserve the freshwater in the reservoir.

The siphon located at the bottom of the reservoir helps evacuate the denser salt water from the bottom of the reservoir to the ocean. This siphon is associated with a salinity sensor. When a salinity threshold is reached, and if the reservoir level is higher than the ocean level, the siphon automatically starts evacuating saline water from the

Figure 6.4: Lock operations just before the summer season from the ocean to the Vilaine (left) and from the Vilaine to the ocean (right).

reservoir. The use of the siphon is closely related to lock operations. During low flows, evacuating through the siphon can only be stopped provided that lock operations are interrupted for an extended period.

The fish ladder and the fish pass allow fish species, such as eels, mullets, lampreys or shads, to cross the dam and migrate from the salt water of the ocean into the freshwater of the river and vice versa (Figure 6.5). The fish ladder more specifically enables the counting of glass eels which is a protected species. The flow through the fish pass is around 28,000 $m^3$ per day. If the water in the reservoir is inadequate to meet the needs of the water supply treatment plant, either because its salinity is too high or the reservoir level is too low, and once restrictions on lock openings are already set, the fish pass may be closed.


Figure 6.5: Fish crossing the fish pass (left), the fish scale (top right) and glass eels waiting to be counted and transferred into the Vilaine River after their ascension of the fish scale (bottom right).

Table 6.2 lists several scenarios of hydrological conditions (from critical low flows to flood) and the associated management objectives and constraints on the different components of the dam. We observe that, as the risk of flood increases, the protection of populations and zones of economic interest is emphasized, and, as the risk of critical low flows increases, drinking water supply is emphasized. Also, as the streamflow increases, the objective level in the reservoir decreases, in order to make room for high inflows. During low flows, the reservoir level should be as high as possible to keep a maximum volume of water in the reservoir throughout the low inflow period. Sluice gates and shutters are mainly operated in conditions ranging from regular flows to high flows, and the fish pass and the lock are exclusively opened in non-critical conditions.

### 6.2.2 Reservoir management data

IAV measures reservoir levels at three stations: Pont de Cran, la Roche Bernard and Arzal, by means of level sensors. The bottom of the Arzal reservoir is located at -7.72 m NGF, while the water level in the reservoir varies around 2 m NGF.

The outflow through the shutters ($m^3$/day) is calculated at IAV based on the measured level in the reservoir and open channel hydraulics principles. Similarly, the outflow through the siphons and through the fish pass ($m^3$/day) are both estimated with laws of hydraulics based on level sensors located both in the reservoir and at sea.

Table 6.2: Hydrological conditions, associated dam management objectives, and associated constraints on reservoir levels and elements of the dam (the shutters and sluice gates, the fish pass, the siphon and the lock) (modified from PAGD, 2015).

| Hydrological conditions | Dam management objectives | Reservoir levels | Dam outflows | Siphon | Fish pass | Lock |
|---|---|---|---|---|---|---|
| Flood<br>*Crisis*<br>$Q > 250$ m$^3$/s | Protection of populations and zones of economic interest | Between 0 and 0.80 m NGF | Sluice gates | - | Closed | Boating prohibited |
| Small flood<br>*Vigilance*<br>$Q$ from 100 to 250 m$^3$/s | Protection of populations and zones of economic interest<br>Drinking water supply<br>Estuarine environment | Optimum between 0.80 and 1.20 m NGF | Sluice gates | - | Open if level $>$ 1.30 m NGF | Regular |
| Low flows<br>*Regular*<br>$Q$ from 10 to 100 m$^3$/s | Drinking water supply<br>Estuarine environment | Between 1.60 and 2.30 m NGF | Shutters and gates | Open | Open | Regular |
| Severe low flows<br>*Vigilance*<br>$Q$ from 2.5 to 10 m$^3$/s | Drinking water supply<br>Estuarine environment | As high as possible | None | Open | Restricted | Reduced |
| Critical low flows<br>*Crisis*<br>$Q < 2.5$ m$^3$/s | Drinking water supply | As high as possible | None | Closed | Closed | Closed |

The outflow through the lock (m$^3$/day) is estimated based on the difference between the level in the reservoir and the sea level, and the dimensions of the lock.

The outflow through the sluice gates (m$^3$/day) was computed at IAV based on the measures from the level sensors located both in the reservoir and at sea, and the equations for bottom valves. The equations used to compute the outflow were parametrized. Parameters were optimized to minimize the distance between the sluice gate outflows and the inflows observed at Rieux (Pont de Cran) for the period running from 2012 to 2014. Gate outflows were then calculated with the obtained set of parameters for the period running from 2003 to 2014. Lastly, the withdrawals for drinking water supply (m$^3$/day) are directly measured with a flow sensor at the pumping station. On average, about 52,000 m$^3$ are pumped each day.

Table 6.3 summarizes the periods for which the data provided by IAV, as well as hydrometeorological data were available. The time period for which we had available data for all variables extends from 2005 to 2011. Uncertainties in the different sensors as well as in the representations used in the hydraulics equations lead to data sets with different accuracies.

Table 6.3: Years of available hydrometeorological and management data. Three stars in the *Trust* column indicates that dam managers consider data as reliable, whereas one star indicates that dam managers consider data as not reliable.

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IAV observed streamflow | | | | x | x | x | x | x | x | x | x | x | x | x | x | / |
| IAV local precipitation | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | / |
| SAFRAN precipitation | x | x | x | x | x | x | x | x | x | x | x | x | / | | | |
| Interannual SAFRAN ETP | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Outflow – Sluice gates | | | | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Outflow – Shutters | | | x | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Outflow – Siphons | | | x | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Outflow – Fish pass | | | x | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Outflow – Lock | | | x | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Withdrawal – Water supply | | | | | | x | x | x | x | x | x | x | x | x | x | |
| Water levels – Arzal | | | | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Water levels – Roche Bernard | | | | x | x | x | x | x | x | x | x | x | x | x | x | / |
| Water levels – Pont de Cran | | | | x | x | x | x | x | x | x | x | x | x | x | x | / |

| | |
|---|---|
| x | Years with period May-October complete |
| / | Years with period May-October incomplete |

Table 6.4: Correspondence between reservoir levels and reservoir volumes.

| Level ($m\ NGF$) | Volume ($Mm^3$) |
|---|---|
| 0 | 25.52 |
| 0.5 | 29.31 |
| 1 | 33.41 |
| 1.5 | 37.72 |
| 2 | 42.18 |
| 2.1 | 43.09 |
| 2.2 | 44.01 |
| 2.3 | 44.99 |
| 2.4 | 46.07 |
| 2.5 | 47.31 |

The correspondence between levels in the reservoir and its volume were obtained at IAV through a GIS analysis. The bathymetric data used in this analysis mostly come from high-resolution surveys, but also from medium-resolution surveys and numerical model outputs (LIDAR data from the Litto 3D model). Table 6.4 presents the correspondence between reservoir levels and reservoir volumes as assessed by IAV with the GIS system.

### 6.2.3 Reservoir inflow forecasts

The GR6J model was set up to simulate and retrospectively forecast streamflow at the outlet of the Vilaine river basin (Rieux gauging station at Pont de Cran, see Chapter 1). In this section, the numerical criterion used to optimize the calibration process is the Kling Gupta Efficiency applied to inverse flows ($\text{KGE}_{iQ}$), which reinforces performances on low-flow periods. The calibration method and the results obtained in validation are presented in Chapter 2. Forecasts of inflows to the reservoir were obtained with the ESP method. The method consists in using precipitation time series from previous years as input to a hydrological model that has been initialised with information available up to the time of the forecast (see Chapter 5). These forecasts were produced up to July 2012 because SAFRAN precipitation data used to initialize the hydrological model states were available up to this date.

## 6.3 Reservoir balance model and evaluation framework

### 6.3.1 Formulation of the reservoir balance model

A reservoir balance model was set up for the Arzal reservoir. The reservoir volume $V$ at time $t$ can be calculated from the volume observed at time $t-1$ as:

$$V_t = V_{t-1} + \Delta V_t \tag{6.1}$$

with,

$$\Delta V_t = V_{\text{I},t} + V_{\text{O},t} \tag{6.2}$$

where $V_\text{I}$ comprises all reservoir inflows and $V_\text{O}$ comprises all reservoir outflows.

The reservoir inflows $V_\text{I}$ are defined as:

$$V_{\text{I},t} = V_{Q,t} + V_{P,t} \tag{6.3}$$

where $V_Q$ is the inflow volume from the upstream catchment, defined with an outlet at Rieux, and $V_P$ is the volume of precipitation over the area of the reservoir, represented by the volume of precipitation observed at the rain gauge located at the Arzal dam.

The reservoir outflows $V_\text{O}$ are defined as:

$$V_{\text{O},t} = V_{\text{s},t} + V_{\text{g},t} + V_{\text{w},t} + V_{\text{f},t} + V_{\text{l},t} + V_{\text{si},t} + V_{PE,t} \tag{6.4}$$

where $V_\mathrm{s}$ and $V_\mathrm{g}$ are the volumes evacuated by the dam managers through the shutters and sluice gates, respectively, $V_\mathrm{w}$ is the water volume pumped for drinking water supply, $V_\mathrm{f}$, $V_\mathrm{l}$ and $V_\mathrm{si}$ are the volumes lost to the ocean through the fish pass, the lock and the siphon, respectively, and $V_{PE}$ is the water volume evaporated over the area of the reservoir, approximated by the interannual potential evapotranspiration.

The volumes associated with precipitation and evapotranspiration over the area of the reservoir are calculated as follows:

$$V_{PE,t} = H_{PE,t}.S_{\mathrm{R},t} \tag{6.5}$$

$$V_{P,t} = H_{P,t}.S_{\mathrm{R},t} \tag{6.6}$$

where $H_{PE}$ is the height of water evaporating as evaluated by the interannual potential evapotranspiration, $H_P$ is the height of water due to precipitation and $S_\mathrm{R}$ is the surface of the reservoir. $S_\mathrm{R}$ was approximated by the water surface of the reservoir corresponding to a reservoir level of 2 m NGF, i.e. 6.2 km$^2$. Indeed, 2 m NGF corresponds to the average water level in the reservoir. Note that a preliminary sensitivity analysis had shown that varying the surface $S_\mathrm{R}$ between 5 and 50 km$^2$ had little impact on the performances of the water balance model. Reservoir evaporation and precipitation thus have a very small impact on the water balance.

The previous equations are used to simulate time series of volumes in the Arzal reservoir. However, the management of the dam is based on reservoir levels. Level time series can be deduced from the volume time series by applying a volume-to-level transform $f$ to the simulated reservoir volume. This volume-to-level transform is a linear function fitted to the reservoir level and volume values presented in Table 6.4. Figure 6.6 shows the computed points, the fitted linear volume-to-level transform $f$, as well as its coefficients. From this figure, we can see that an error of 10 Mm$^3$ in reservoir volume can lead to an error superior to one meter in the reservoir level.

The reservoir level time series can be written as:

$$N_t = f(V_t) = aV_t + b \tag{6.7}$$

where $N$ is the reservoir level, and $a$ and $b$ are the slope and the intercept of the volume-to-level transform (see Figure 6.6). Based on this equation, the water balance equation can be expressed in terms of reservoir levels:

$$N_t = N_{t-1} + a\Delta V_t \tag{6.8}$$

The global structure of the model is represented in Figure 6.7. We can identify the upper left branch that corresponds to the inflow and outflow data and the lower left branch that contributes with the previous reservoir volume.

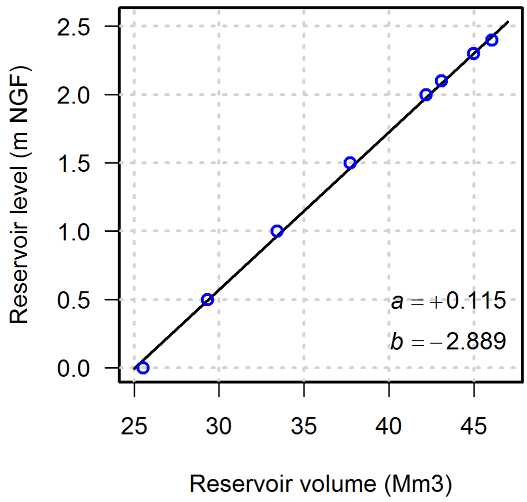



Figure 6.6: Reservoir levels versus reservoir volumes in the Arzal reservoir. Blue circles correspond to the values assessed by the GIS system and presented in Table 6.4. The black line represents the fitted linear transform, whose coefficients are displayed in the bottom right corner.

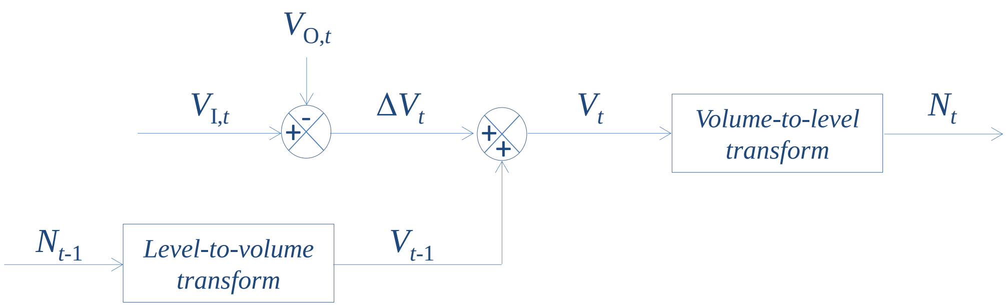



Figure 6.7: Schema of the structure of the reservoir balance model at a given time step.

### 6.3.2 Forecasting framework

The model of the reservoir balance is run at the daily time step based on Equation 6.8. A model run is defined by an initialization date and a forecast horizon. In the first time step, i.e. on the initialization day, the reservoir level observed on the previous day is used along with inflow and outflow data to calculate the updated reservoir level. In all following steps, we assume that future observed reservoir levels are still unknown. The reservoir level is thus calculated based on the previously simulated reservoir level and on inflow and outflow data. This is repeated up to the maximum forecast horizon. The equation of the reservoir in this configuration can be written as:

$$\hat{N}_t = \begin{cases} N_{t_0-1} + a\Delta V_{t_0}, & \text{if } t = t_0 \\ \hat{N}_{t-1} + a\Delta V_t, & \text{otherwise} \end{cases} \tag{6.9}$$

where $t_0$ is the initialization date, $N$ is the observed reservoir level, $\hat{N}$ is the simulated reservoir level, $a$ is the slope of the volume-to-level transform and $\Delta V$ is defined in Equation 6.2. Recursively, we can express the simulated reservoir level at time $t$ as a combination of the initial reservoir level $N_{t_0-1}$ and the inflow and outflow data:

$$\hat{N}_t = N_{t_0-1} + a \sum_{i=t_0}^{t} \Delta V_i \tag{6.10}$$

The performance of the system is strongly related to the reservoir level data, the inflow and outflow data, and the slope of the volume-to-level transform. If we consider that the errors in the reservoir level and the errors in the volume-to-level transform are negligible in regard to the errors in the inflow and outflow estimations, the error in the simulated reservoir levels solely resides in the inflow and outflow data. Since inflows and outflows are summed over time steps, the contribution of the error in the inflow and outflow data to the error in reservoir level is expected to increase with lead time.

### 6.3.3 Evaluation framework

The objective of the simulations obtained with the water balance model is to reproduce the observed levels in the reservoir. The quality of the simulated levels is assessed with the Root-Mean-Square Error (RMSE; see Chapter 3). Observed levels were collected from three different measurement stations: Arzal, la Roche Bernard and Pont de Cran (cf. Figure 6.2). The portion of the reservoir running from the Pont de Cran station to the Arzal station is 40 km long and a hydraulic gradient exists between these stations. Figure 6.8 shows the scatterplots of levels obtained from these three stations to allow a comparison of their records.

Strong differences in levels are observed from November to April between the Pont de Cran station and the other two downstream stations. These differences could, for instance, be explained by the impact of the wind or of dam operations on water levels. Indeed, winds from the ocean can push the water level up at Pont de Cran and subsequently cause levels at Arzal or at la Roche Bernard to decrease. Also, and as seen previously, the opening of the sluice gates at the dam can cause water levels to artificially decrease in the vicinity of the dam. Despite these differences in levels between November and April, levels at the three stations are very similar from May to October. Differences in level between Pont de Cran and Arzal are, on average, close to 2.5 cm and rarely exceed 6 cm. The level at la Roche Bernard would be a good reference for a median level in the reservoir. However, the acquisition of level data at this station changed between 2009 and 2010 and data prior to this period are overestimated. Therefore, we chose to evaluate our level simulations against the mean of the levels observed at the Pont de Cran and Arzal stations.

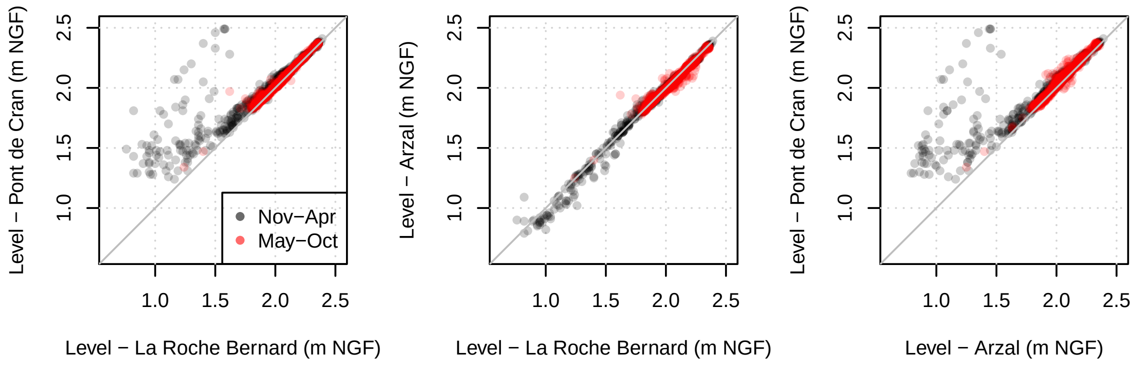



Figure 6.8: Scatterplots comparing reservoir levels observed at the Arzal, Pont de Cran and la Roche Bernard stations for data acquired from 2010 to 2014. Black dots correspond to levels measured between November and April (wet period), and red dots represent levels measured between May and October (dry period).

### 6.3.4 Setup of model runs

Three configurations of the model are used in the following sections. These configurations are presented in Table 6.5. Each configuration is characterized by the frequency of the initialization, which ranges from a daily initialization to a single initialization on the first of each month, and the target horizon, which ranges from one day ahead to one month ahead.

The first two configurations from Table 6.5 are used in Section 6.4 to better understand the behaviour of the water balance model and evaluate its performance in reproducing the reservoir levels over the years. In these configurations, the model is run with observed outflow data and both, observed and simulated (with GR6J), inflows. The evapotranspiration is represented by the interanual evapotranspiration. In the first configuration, the model is initialized each day of the available period and run only up to the day ahead. This configuration allows us to visualize the errors in the reservoir level simulations for a day lead. At each time step, these errors come from the model structure, the previous observed level and the inflow and outflow data at this time step. In the second configuration, the model is initialized every day of the available record period and run up to a month ahead. These simulations are used to numerically evaluate the model with the lead time.

The third configuration is used to develop the risk assessment tool presented in Section 6.5. In this configuration, the model is run with forecast inflows. Gate and shutter outflows are optimized to maintain a constant level in the reservoir. Outflows through the lock, the fish pass and the siphon, and the amount of water pumped for drinking water are assumed constant and equal to their respective maximum values. In this configuration as well, the evapotranspiration is represented by the interannual evapotranspiration.

Table 6.5: Characteristics of the configurations of the water balance model.

| # | Time step | Initialization | Horizon | Inflow data | Gates and shutters data | Other outflow data |
|---|---|---|---|---|---|---|
| 1 | daily | daily | one day | - observed<br>- simulated | observed | observed |
| 2 | daily | daily | one month | - observed<br>- simulated | observed | observed |
| 3 | daily | $1^{st}$ of each month | Oct $31^{st}$ | - ESP forecast<br>- Null $P$ forecast | optimized for constant level | constant, maximum value |

## 6.4 Preliminary data analysis and water balance simulations

### 6.4.1 Relative importance of the different inflows and outflows

In order to better understand the contribution of each data source to the simulation errors, we analysed the relative importance of each input data associated with the reservoir inflows and outflows. Figure 6.9 shows the relative importance of the different inflows and outflows to the reservoir, based on interannual volumes over the May to October and November to April periods. The relative importance of inflows is very similar between periods, with more than 99 % of the inflows coming from the upstream catchment. In both periods, total outflows are slightly superior to total inflows, which suggests that a perfect balance may not be achieved based on the available management and hydrometeorological data.

As expected in a season-dominated climate, we observe large differences in mean interannual values in total inflows and outflows between the two seasons. Observed inflows and outflows to the reservoir from November to April are almost five times greater than the volumes observed from May to October. Overall, all mean values increase from November to April as compared to May to October, except for the siphons, which are mainly used to evacuate salt water in summer, for the water pumped for drinking water, whose consumption increases in summer, for the lock, since the boating season is from May to October, and, finally, for the evapotranspiration, which is higher in summer. Another difference between periods is the relative importance of outflow sources. From May to October, the major outflows of the reservoir are outflows from shutters and sluice gates (80 % of total outflows). Other outflows are non-negligible (20 % of total outflows), and dominated by outflows through the siphons (10 %) and the fish pass (4 %). From November to April, sluice gates and shutters represent 97 % of the outflows from the reservoir, with almost 75 % of the outflows solely due to gate openings. Other outflows are negligible in this period of the year.

Figure 6.10 complements Figure 6.9 by presenting the relative importance of the different inflows and outflows on a year-to-year basis from 2005 to 2014 for the May to October period. Years were ranked based on their May to October inflow volumes, from the driest (187 $Mm^3$ in 2011) to the wettest year (673 $Mm^3$ in 2008), to highlight differences in management due to hydrometeorological conditions. From this figure, we observe that outflows over the season through the fish pass, the siphons, the lock and from the pumps for drinking water supply do not vary with the year, and thus do not depend on average inflows. The outflow through the shutters and sluice gates however do vary with inflows since shutters and sluice gates allow managers to adjust the reservoir level depending on the inflows. In the driest years of 2005 and 2011, outflows through the shutters are equivalent to outflows summed for the fish pass and the siphons, while they usually represent more than 40 % of total outflows. In average years, we observe that the management of the dam is done in priority through the shutters. The use of the gates increases with higher inflows and only surpasses the use of the shutters in 2008, the wettest year of the period, and in 2014. The drop in volume evacuated through the shutters observed in 2014 actually reflects a change in strategy by the managers of the Arzal dam. Since 2014, the managers evacuate water in priority through the gates, rather than through the shutters, during the May to October season. This new strategy was chosen to evacuate the salt water at the bottom of the water column without losing freshwater at the top of the water column.

### 6.4.2 Analysis of monthly management strategies

Figure 6.11 zooms in on inflows and outflows at the monthly time step for the years of 2005, 2008 and 2014. 2005 is one of the driest years in the available record period and is representative of the reservoir management during dry periods. 2008 is the wettest year of the record period and is more representative of the management during wetter years, but also during wetter months of the year. Lastly, 2014 is representative of the latest management strategies in the Arzal reservoir within the period analysed in this thesis.

In 2005, we observe that the reservoir is primarily managed with shutters, with small water volumes evacuated through the gates. In June, the volumes evacuated through the shutters and sluice gates strongly decrease and, from July to September, both gates and shutters are closed. Closing gates and shutters allows managers to retain all inflows in the reservoir and water levels from decreasing significantly. Meanwhile, 10 to 15 $Mm^3$ are still used monthly for drinking water supply, lock crossings and fish migration. In October, inflows start increasing and water is evacuated through the gates again. We note however that inflows are still far from an average monthly value for this period (the average monthly inflow from May to October is about 60 $Mm^3$).

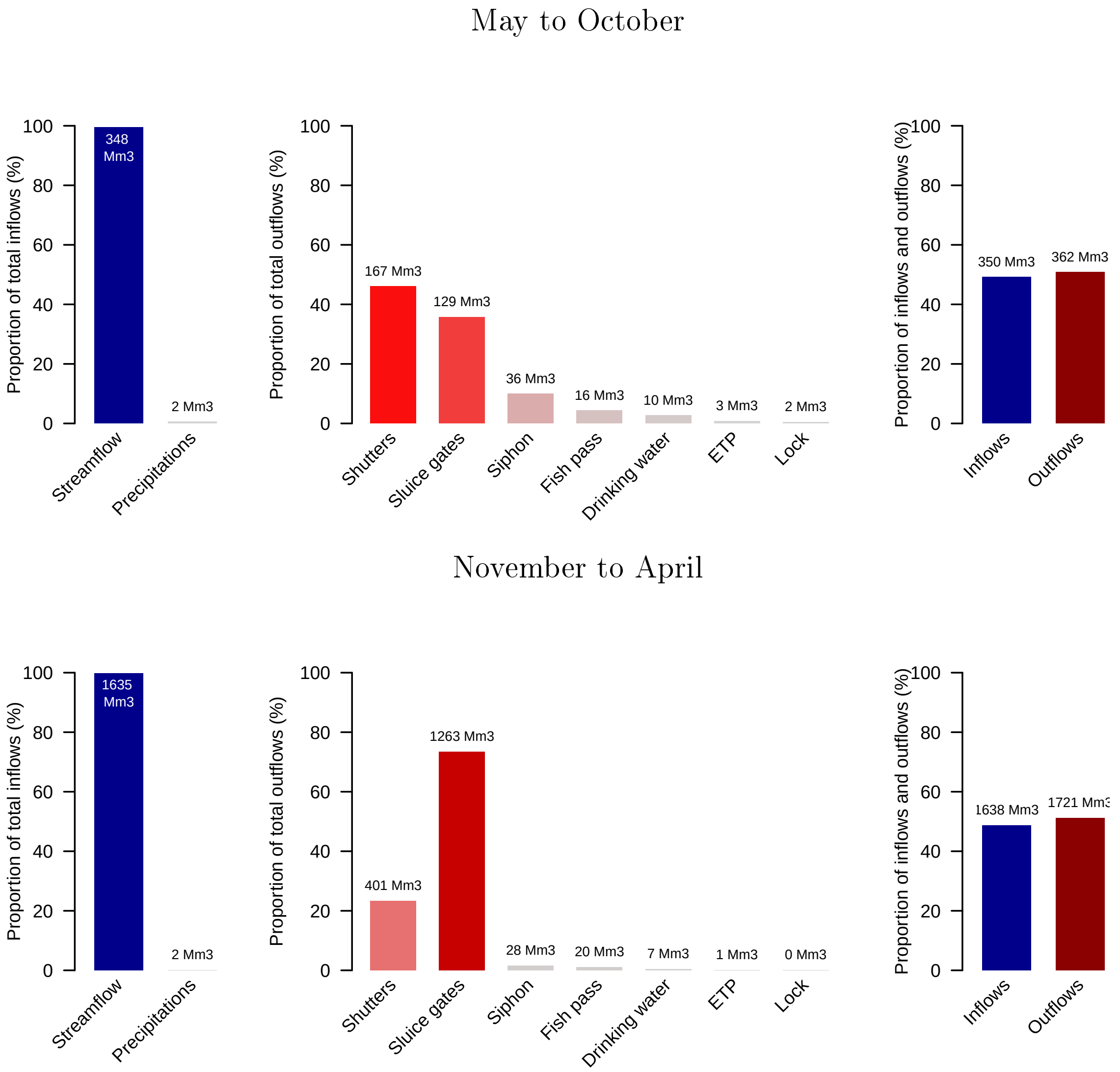



Figure 6.9: Proportion and mean seasonal value of each inflow and outflow component of the Arzal water balance model. Mean seasonal volumes are averaged over 10 years (2005-2014) from May to October and from November to April. The left graph compares inflow volumes, the middle graph compares outflow volumes, and the right graph compares inflow and outflow volumes.

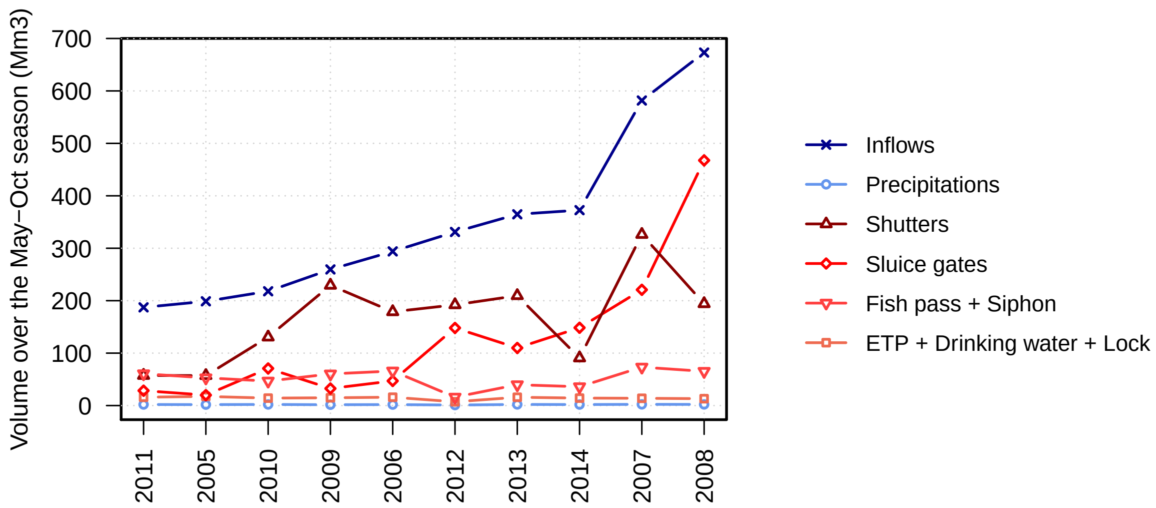



Figure 6.10: Cumulative inflows and outflows over the May to October period for 2005 to 2014. Years are ranked based on their May to October inflow volumes, from the driest (2011) to the wettest year (2008).

In 2008, the high inflows of May and June were mostly evacuated with gate openings. From July to October, inflows became equivalent to the inflows observed in May 2005, i.e. close to 50 $Mm^3$. From this month on, water was evacuated from the reservoir through the shutters, as was observed in May 2005, with small volumes of water still evacuated through the gates. In this year, which is characterized by large inflows during May and June, it was never necessary to completely close the shutters and gates in order to guarantee enough water for the May to October period.

In 2014, the inflows over the May to October period were slightly above average. In May, shutters and sluice gates were equally used to evacuate water volumes from the reservoir. From July to October, when inflows reached volumes equivalent to those observed in May 2005 or July 2008, gates were preferred over shutters to evacuate water, as opposed to the strategies of 2005 and 2008. In fact, this allows managers to evacuate an important quantity of salt water. By doing so, they may ultimately avoid closing the lock for boat sailing during the summer season.

### 6.4.3 Daily inflows and outflows and reservoir reactivity

In general, decisions concerning water quantities in the Arzal reservoir are made at the daily time step. A finer time step is usually only necessary during high-flow periods, when urgent actions must be taken due to upcoming flood events. In addition, the tidal cycle at the sub-daily time step imposes certain constraints on the management of the shutters and sluice gates, the lock and, sometimes, the fish pass. The temporal distribution of releases at the sub-daily time step is outside the scope of this study since

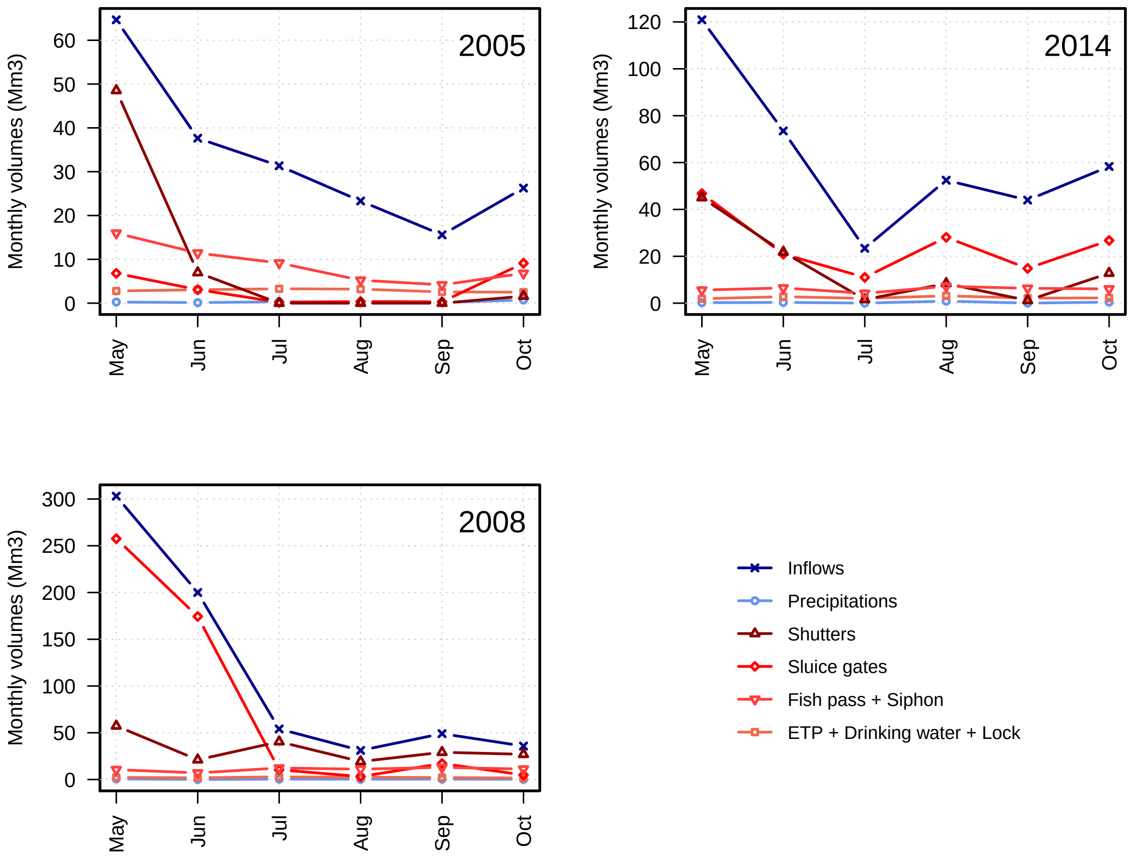
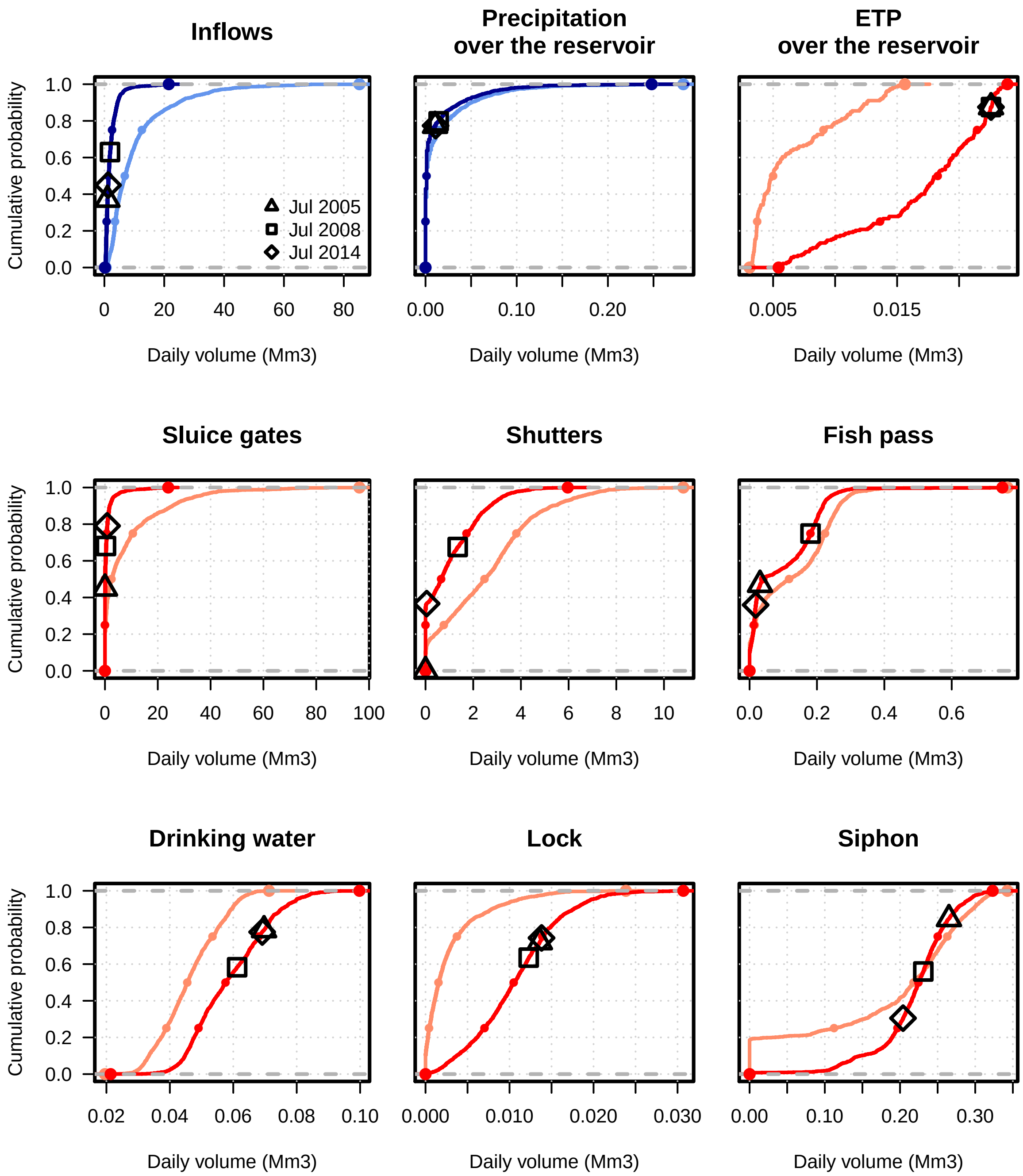



Figure 6.11: Evolution of monthly inflows and outflows over the May to October period in 2005 (dry year), 2008 (wet year) and 2014 (latest management strategy).

we focus on the low-flow season. However, the relative importance and the variability of the input and output variables at the daily time step are of interest since the decisions on reservoir volumes taken daily can influence (in the case of hydrometeorological variables) and reflect (in the case of management variables) the long-term reservoir volume planning.

Figure 6.12 shows the cumulative distributions of the daily volumes for each inflow and outflow of the dam, for the periods running from May to October (darker lines) and November to April (lighter lines). The mean daily values observed in July of 2005, 2008 and 2014 are also represented. Below the figure, a conversion table gives the equivalence between flows in $Mm^3/day$, the unit used in this chapter, and flows in $m^3/s$, the unit previously used in this thesis to describe hydrological conditions in the catchment (see e.g. Table 6.2).

The inflow-capacity ratio characterizes the relative importance of daily inflows as compared to the reservoir capacity. This ratio can help describe how dynamic a reservoir system is. For instance, the level of a reservoir with a high capacity and low




| Volume in Mm$^3$/day | 0.01 | 0.05 | 0.1 | 0.5 | 1 | 5 | 10 | 50 | 80 |
|---|---|---|---|---|---|---|---|---|---|
| Volume in m$^3$/s | 0.1 | 0.6 | 1.2 | 5.8 | 12 | 58 | 115 | 580 | 925 |

Figure 6.12: Empirical cumulative distributions of the daily volumes for each element of the dam and each hydrometeorological variable, for the periods running from May to October (darker lines) and from November to April (lighter lines). Large dots represent the minimum and the maximum of each distribution. Smaller dots represent the 25$^{\text{th}}$ , 50$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. We show a conversion table of flows from Mm$^3$/day to m$^3$/s. Mean daily volumes for July 2005, 2008 and 2014 are also indicated.

inflows (i.e. with a low inflow-capacity ratio) will vary slowly. Provided that management constraints for such a reservoir are not too strict, decisions may not need to be as frequent or fine-tuned as for a reservoir with a quickly varying volume (or a high inflow-capacity ratio). In the case of the Arzal reservoir, inflows and outflows can reach, and even exceed, the reservoir capacity, i.e. 50 $Mm^3$, during the high-flow season. During low flows, maximum inflows and outflows can reach 20 $Mm^3$, i.e. 40 % of the reservoir capacity. Therefore, the Arzal reservoir can be considered as a dynamic system. During high flows, the reservoir can be filled in a day or two. On the opposite, a release during low inflows may drastically lower the reservoir level for an extended period of time.

Figure 6.12 shows that the years of 2005, 2008 and 2014 differ in terms of inflows to the reservoir, but also in terms of management strategy with respect to some of the dam elements, as seen previously and as reflected in the month of July. Inflows of July 2005 are, on average, exceeded by 60 % of the daily observed inflows, while inflows of July 2008 are exceeded by less than 40 % of the observed inflows. July 2014 can be considered as an average month in terms of inflows, even though it is slightly drier than the median inflow. We observe at the daily time step what was also observed at the monthly time step. In 2005, daily shutter and gate outflows are null or almost null to preserve the freshwater in the reservoir. In 2008 and 2014, however, it was necessary to use shutters and gates to evacuate water from the reservoir. Gate outflows are superior in 2014, with only 20 % of observed outflows above the mean volume of 2014 and 30 % above the mean volume of 2008. Since July 2014 was an average month, drier than July 2008, the difference in gate outflows between these months can only be explained by the change in management strategy. The difference in the volumes evacuated through the shutters between July 2008 and July 2014 may be another consequence of this change in strategy. Since more water is evacuated through the gates in 2014, there is less need to operate the shutters.

Other differences between the months of July 2005, 2008 and 2014 are observed in water pumped for drinking water, and water lost through the lock and the siphon. In July 2008, the volume pumped for drinking water supply and the volume lost through the lock are inferior to the volumes of 2014 and 2005, possibly due to the preceding wet months. The volumes lost through the siphons, which evacuate salt water from the bottom of the reservoir and are thus closely related to lock openings, differ significantly between July 2005 and July 2014 . The average outflow through the siphons is close to 0.26 $Mm^3$ in 2005, a value exceeded only about 15 % of the time, and close to 0.20 $Mm^3$ in 2014, a value exceeded 70 % of the time. In both years however, the volumes lost through the lock and through the pumps for drinking water were similar. This difference may reflect the advantage of the new management strategy. Since gate openings evacuate more salt water than freshwater, less salt water needs to be evacuated through the siphons, even in average years such as 2014.

### 6.4.4 Analysis of errors in simulated reservoir levels

In this section, we assess the potential of the reservoir balance model to simulate reservoir levels. Errors in the simulations of the water levels in the Arzal reservoir were evaluated based on the first two configurations described in Table 6.5. Figure 6.13 presents the observed reservoir levels and the reservoir levels simulated with the first configuration, for the May to October periods of 2005 to 2012. Figure 6.14 complements this figure with the time series of outflows through the shutters and sluice gates, the observed inflows and the inflows simulated with GR6J.

In the first configuration, when the model is initialized every day and run for the day ahead, the simulated levels are very close to the observed levels. The simulation based on the observed inflow has an RMSE of 0.09 m over the May to October periods of 2005 to 2012. The simulation based on the GR6J simulated inflows has an RMSE of 0.18 m. This error can be related to the inflows simulated with GR6J, as seen in Figure 6.14. Indeed, we observe large errors in the inflows simulated with GR6J during high flows. This is fairly expected because the model was calibrated to minimize errors in low flows. The errors in simulated inflows lead to errors in the reservoir levels of up to a meter, as in 2007. Nevertheless, during low flows, the simulations of reservoir levels based on GR6J simulated inflows provide good results, as in 2005 and 2011. In these two cases, the reservoir level simulations obtained with simulated inflows show even better results than the level simulations obtained with observed inflows.

In the second configuration, the model is initialized every day and run for the month ahead. The RMSE values of these simulations were averaged for initialization dates within the May to October period and initialization dates within the November to April period. The water balance model was run at the daily time step, but also at the weekly time step (i.e. based on weekly-averaged inflows and outflows), and both the observed inflows and the GR6J simulated inflows were used as input to the water balance model. Figure 6.15 presents the evolution of the RMSE of these reservoir level simulations with the lead time. As expected, the error increases with the lead time. Errors are also much larger for flows simulated between November and April, probably due to the larger absolute errors in observed inflows and outflows directly translated in high reservoir level errors. Errors are also larger when using GR6J simulated inflows rather than observed inflows, especially during the November to April period. Figure 6.15 shows that errors are smaller when running the model with weekly inflows and outflows to simulate weekly reservoir levels. While errors at the daily time step can reach values up to 12 meters for a lead time of 30 days, errors at the weekly time step do not exceed 1.5 meters for a lead time of four weeks. It seems that the water balance model of the reservoir can provide more accurate level simulations for the month ahead when considering weekly-averaged levels rather than daily levels.

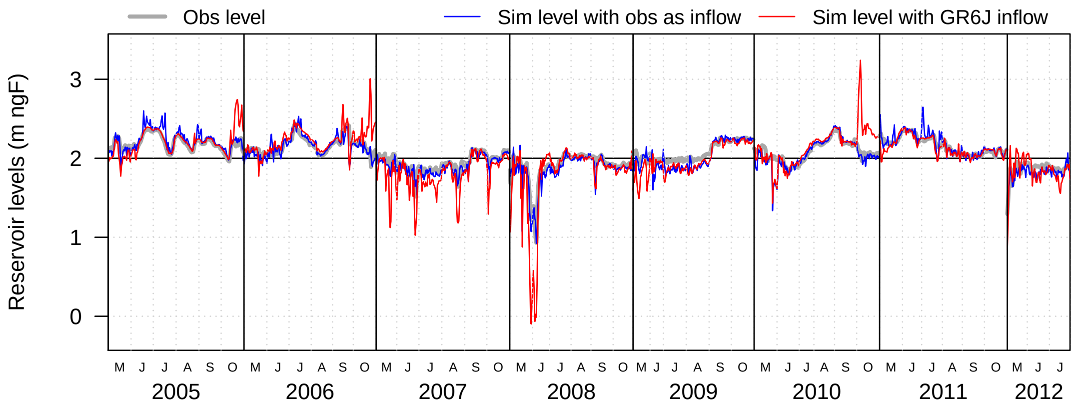



Figure 6.13: Observed and simulated levels in the Arzal reservoir obtained with the first configuration of Table 6.5 for the May to October periods of 2005 to 2012. Observed levels are represented in grey, levels simulated from observed inflows are in blue and levels simulated from inflows simulated with GR6J are in red.

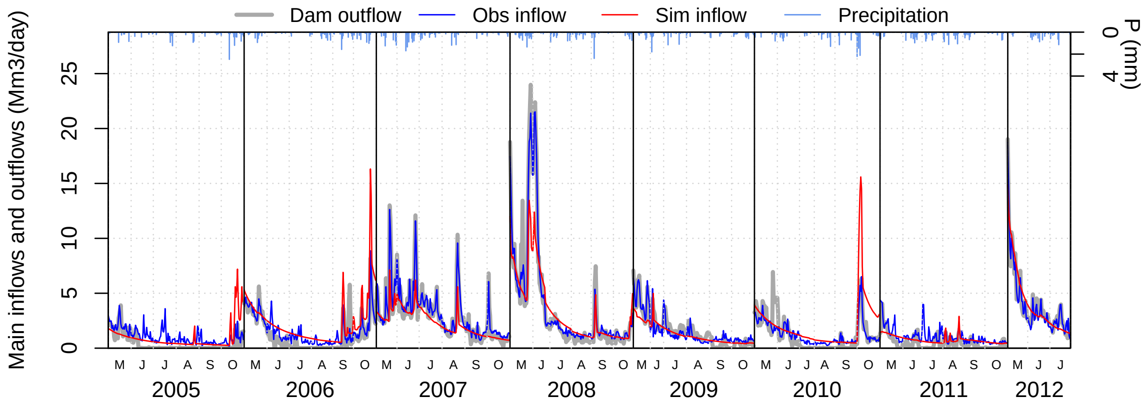



Figure 6.14: Major inflows and outflows of the Arzal reservoir for the May to October periods of 2005 to 2012. Inflows from the upstream catchment and outflows from shutters and sluice gates are also represented (bottom). Summed outflows for the shutters and sluice gates are represented in grey. Observed inflows (streamflow measured at Pont de Cran) are in blue and inflows simulated with GR6J are in red. Observed precipitations are represented in light blue.

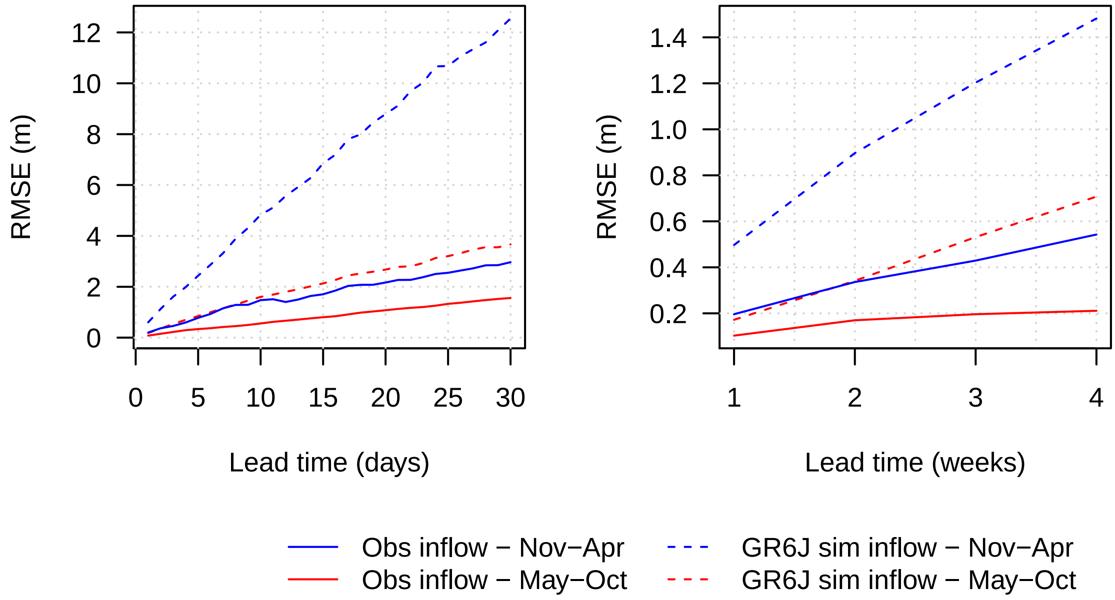



Figure 6.15: Evolution with lead time of the RMSE of reservoir level simulations obtained with the second configuration of Table 6.5. Simulations at the daily (left) and at the weekly (right) time step, for May to October (continuous line) and November to April (dashed line), and for simulations with observed inflows (blue) and GR6J simulated inflows (red).

## 6.5 The risk assessment tool

Several exchanges with IAV, the institution responsible for the management of the Arzal dam, based on the results of the preliminary analysis, led to the development of a risk assessment tool. The design of the tool was guided by their management objectives and needs.

### 6.5.1 Design of the risk assessment tool

The management of the Arzal dam is based on water levels and restrictions are formulated relatively to threshold levels (cf. Table 6.2). Therefore, a useful management tool should focus on reservoir levels. Another main request was to have a tool that would run at the daily time step and that could be updated every day to support daily decisions. Predictions would ideally cover the May to October period, and the maximum horizon targeted by the tool would be fixed to October 31$^{st}$, i.e. the end of the low-flow season in the Vilaine catchment. End-users expressed an interest in having a seasonal outlook of potential water shortages so as to agree on measures with stakeholders earlier on in the season and limit tensions.

Several discussions with the dam managers on the management objectives and the performances of the water balance model were necessary to decide on how hydrological modelling could best serve reservoir management. Several points were brought to light during these discussions. One main issue was that the objective of the dam managers would not be to use the final tool to optimize outflows from the dam. Indeed, the work carried out in this thesis only targets water quantities in the Arzal reservoir, while water quality (i.e., salt related issues) is in fact a major factor in the decision process. Therefore, it was not in the objectives of the risk assessment tool to provide quantitative guidance on optimal releases for the Arzal reservoir.

However, the proposed tool should provide a qualitative idea of the risk of severe low flows for the coming season and of the risk of having to modify the management strategy of the reservoir during the season. More specifically, the objective was formulated as follows: *what is the risk of not being able to ensure a minimum level in the reservoir at some point during the low-flow season ?* To answer this question, the tool was designed with the following inflow and outflow characteristics:

- The reservoir balance was fed with streamflow forecasts that reproduce the natural variability in streamflow in the upstream catchment. Streamflow forecasts were issued with the ESP method (as described in Chapters 4 and 5). These forecasts represent a wide range of possibilities, which are only constrained by catchment initial conditions. In addition to these, a forecast generated by running the GR6J hydrological model with zero rainfall as input was used. This forecast represents a worst-case scenario.
- Dam outflows were computed to maintain a constant objective level in the reservoir. This objective level was first set at 2 m NGF. Models of the releases of the Arzal dam based on inflows were tested but, given the reactivity of the reservoir and the different uncertainties in the observed streamflow and management data, these models were discarded. Dam releases optimised to maintain a constant reservoir level were the closest option to the way the Arzal dam is currently managed. They also guaranteed realistic reservoir levels and avoided setting unrealistic constraints on dam releases since these are not constrained in the current practice.
- Outflows from the dam other than releases through shutters and sluice gates were assumed constant between May and October. We have seen that these outflows have little impact on the reservoir balance model, even though their relative weight increases during the summer season. Outflow values were fixed at their maximum observed values: water pumped for drinking water supply was fixed at 0.1 $Mm^3$ per day, outflows due to lock openings were fixed at 0.03 $Mm^3$ per day, outflows through the fish pass were fixed at 0.4 $Mm^3$ per day and outflows through the siphon were fixed at 0.33 $Mm^3$ per day (cf. Figure 6.12). By fixing outflows at their maximum values, we consider water uses that are more demanding than they are in reality. However this guarantees all uses in their current volume at all times.
- Evaporation from the free water surface was assumed to be equal to the inter-

annual daily potential evapotranspiration; precipitation falling directly in the reservoir between the dam and Pont de Cran was assumed to be null.

### 6.5.2 Risk-oriented simulations

Figure 6.16 shows the simulations obtained with the above characteristics between 2005 and 2010. The model is run in the third configuration of Table 6.5, i.e., the model is initialized at the beginning of each month between May and September, and run up to October 31$^{st}$. Note that the figure is interactive (this feature requires the PDF file) and that each frame represents a different initialization date.

These simulations assess the periods during which managers will likely be able to maintain a constant objective level in the reservoir. They also assess the periods during which the inflows to the reservoir may not be adequate to maintain a minimum level in the reservoir, given no outflows through the shutters and sluice gates and constant values for other outflows. The minimum acceptable level was set at 1.80 m NGF by the dam managers. In Figure 6.16, the main difference between the simulations for each of the six years is the initial hydrological conditions used to run the hydrological model.

The years 2007 and 2008, initialized in May and in June, illustrate the impact of initial conditions on the assessed risk. The higher the initial streamflow and reservoir conditions, the lower the risk of not being able to maintain the objective reservoir level. When the simulations are initialized in May, the risk assessed for the end of the season strongly depends on the meteorological forcings used as input to the GR6J model. Since these forcings are very similar from one year to the next, the risk assessed as early as in May is very similar in the six years. However, differences in the risk assessed for the six years start appearing when the simulations are initialized in June. As early as in June, the risk assessed by the simulations seems higher in 2005, 2006, and lower in 2007 and 2008. This is coherent with the ranking from drier to wetter years shown in Figure 6.10, except for the year 2006 which was not as dry as 2009 and 2010.

### 6.5.3 Risk-oriented graphs

Risk-oriented graphs were developed to summarize the risk information provided by the simulations of the previous section. They are shown in Figure 6.17. Each graph summarizes the simulations by representing the number of days below the minimum acceptable reservoir level (here, 1.80 m NGF), for a given initialization date and for each month of the forecast horizon. The number of days below the threshold is divided in six categories: the threshold is never reached (0 days), the threshold is reached during one to six days, seven to 12 days, 13 to 18 days, 19 to 24 days, and 25 to 31 days. For each month of the forecast period, the percentage of scenarios falling within each of these categories is shown. The percentage is represented by the intensity of the colour in the corresponding cell, and values greater than twenty percent are written inside the cells.

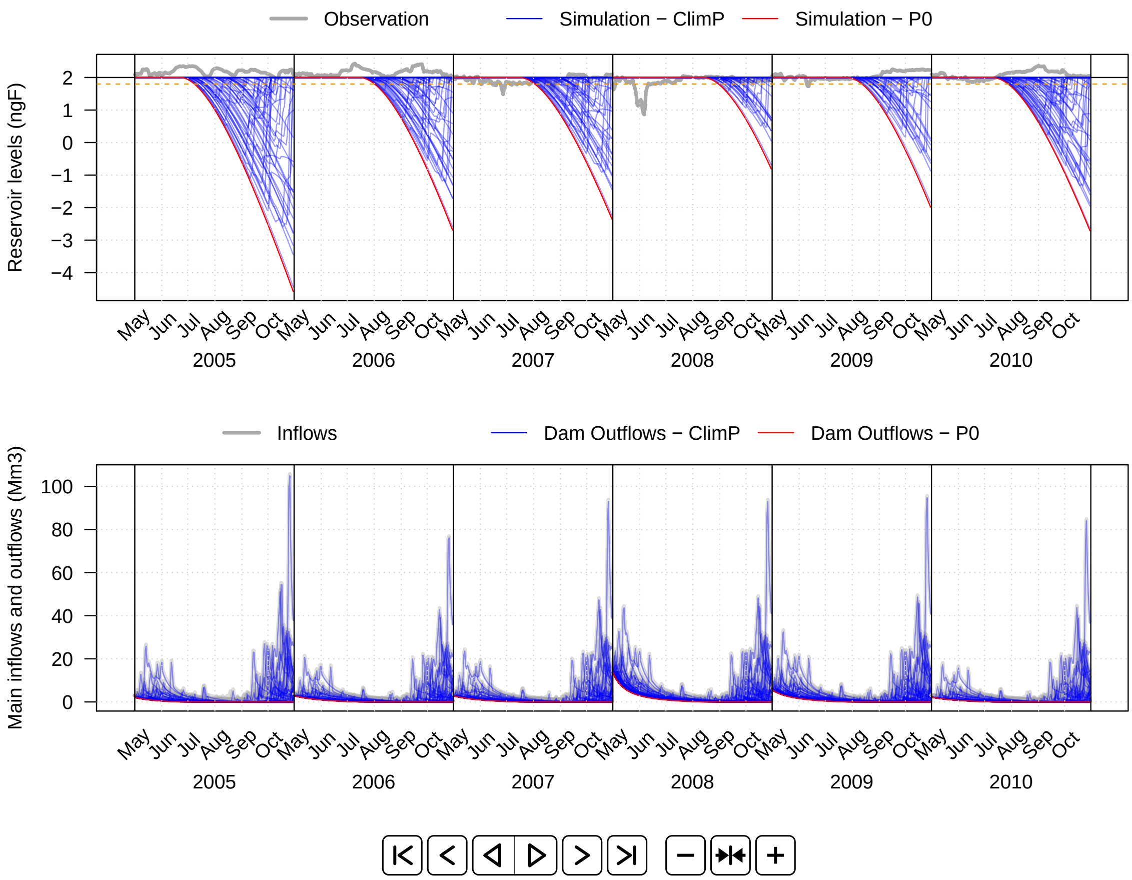



Figure 6.16: Level simulations (top), and inflow forecasts and corresponding shutter and gate outflows (bottom). The objective level is 2 m NGF. Each column corresponds to a different year, and each frame corresponds to a different initialization date, i.e. from May 1$^{st}$ to September 1$^{st}$. Blue lines correspond to reservoir levels and inflow forecasts generated with the ESP method and red lines correspond to the forecast generated with the GR6J model with null precipitation as input.

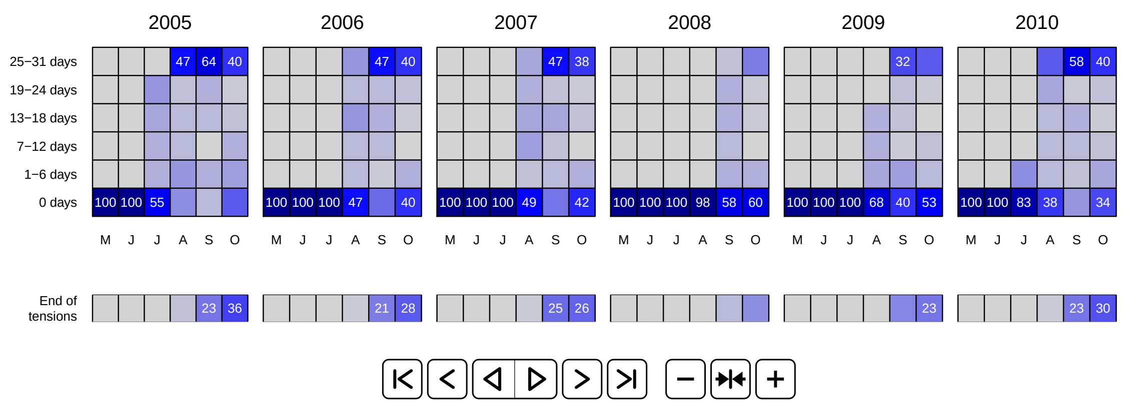



Figure 6.17: Risk graphs used to represent the probability of being below the minimum acceptable reservoir level of 1.80 m NGF for different durations and the probability of being able to return to a level superior to 1.80 m NGF. The objective level is 2 m NGF. Each column corresponds to a different year, and each frame corresponds to a different start date, i.e. from May 1$^{st}$ to September 1$^{st}$.

Below this graph, another graph indicates the percentage of scenarios returning to levels superior to 1.80 m NGF for each month. It is indicated as “End of tensions” because it represents the likelihood of going back to a situation where inflows are adequate to maintain the objective level and therefore to return to standard management rules.

Note that it was not possible to evaluate the proposed tool against observations, because the tool does not use actual reservoir releases, but releases that are optimized to maintain a constant level and that are chosen to always allow uses. Actual reservoir releases are decided based on a complex learning process from preceding and current situations that also includes water quality objectives. Therefore, the observed reservoir levels and the actual number of days below the 1.80 m NGF limit cannot be used as observations.

### 6.5.4 Sensitivity to simulation parameters

We investigated the impact of variations in simulation parameters on the risk as evaluated by the risk-oriented graph. Five parameters were looked at: the objective reservoir level, the constant water amount allocated to drinking water supply, the constant water amount that is assumed lost through the lock, the constant outflow through the fish pass and the constant outflow through the siphon. The risk as represented in Figure 6.17 was reduced to two indicators: (1) the number of days below the 1.80 m NGF threshold, averaged over all ensemble members, and (2) the number of members that predict at least one day below the 1.80 m NGF threshold. Lower values of these indicators correspond to lower risks of not being able to guarantee the minimum acceptable water level. In order to analyse the sensitivity of the assessed risk to the parameters, each parameter value was varied between 0 and 150 % of its initial value, while the other parameters were fixed at their initial value. Percentages of variation and corresponding parameter values are shown in Table 6.6. We looked at how the risk indicators were impacted by these variations.

In Figure 6.18, we show the analysis carried out for the year 2005, and for a simulation initialised on May 1$^{st}$. We can see that the parameter that has the most impact on the assessed risk is the objective level in the reservoir. Reducing this level drastically increases the risk since there is less water stored in the reservoir in anticipation of low inflows (note that an increase of 10 cm in objective level corresponds to an increase of 0.9 $Mm^3$ in reservoir volume). Increasing this level reduces the risk, but one should remember that the maximum level in the reservoir is constrained by the risks of flooding the upstream fields and marshes. We observe that variations in the water allocated to drinking water supply have very little impact on the assessed risk. In practice, the water allocated to drinking water supply is not considered as a variable of adjustment by IAV, since it is prioritized over all other outflows. Reducing its volume is thus out of the question for the dam managers. However, we see here that an increase in the amount of water pumped for drinking water supply would have little impact on the risk

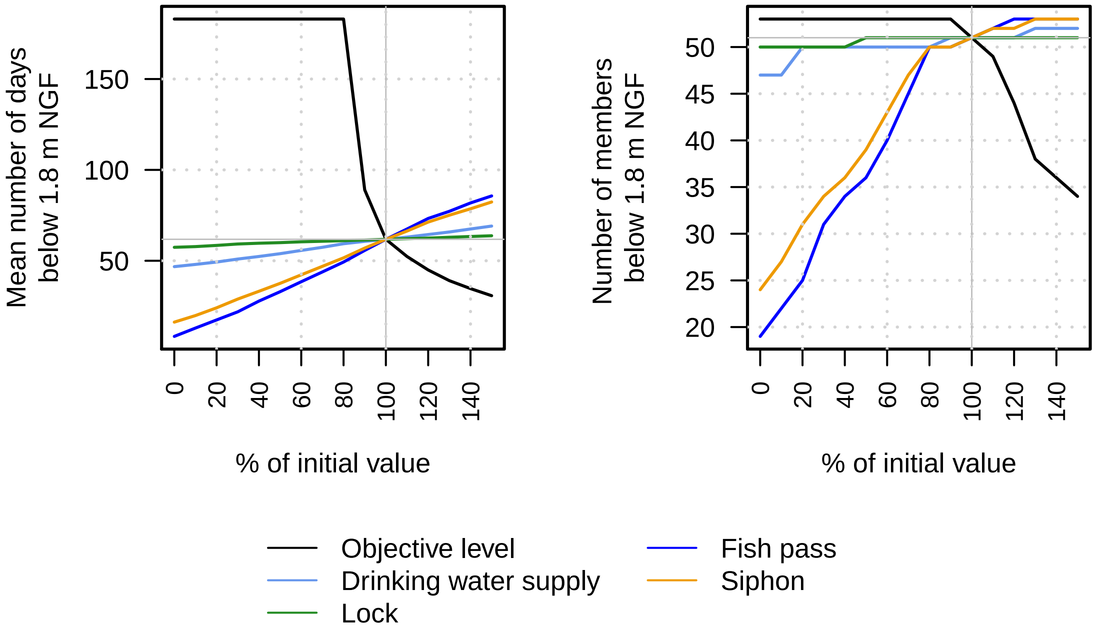



Figure 6.18: Sensitivity of the mean number of days and the number of members below the 1.80 m NGF threshold to changes in the objective reservoir level and outflows other than shutters and sluice gates for the year 2005. The mean number of days is averaged over all ensemble members. The number of members corresponds to the members that predict at least one day below the 1.80 m NGF threshold.

in terms of water quantity in the reservoir. The water lost through the lock also has very little impact on the risk variables. However, this volume cannot be dissociated from the volume evacuated through the siphon, which has more impact on the assessed risk. Reducing the volume through the siphon can almost neutralize the risk assessed in 2005. This partly explains why closing the lock can help in the management of the dam during dry seasons. When the lock is closed, the water quality in the reservoir is not deteriorated by salt intrusions and using the siphon is not necessary. This reduces one of the sources of water loss in the reservoir and strongly reduces the risk. Similarly to the volume evacuated through the siphon, the volume allocated to the fish pass has a strong impact on the assessed risk. Reducing the flow through the fish pass, or closing it, can substantially help reduce the risks of reaching below the 1.80 m NGF threshold. However, because preserving the estuarine environment is one of the management objectives for the dam, the outflow through the fish pass can only be reduced or closed once lock openings have been stopped.

## 6.6 Conclusions

This chapter focused on the case study of the Arzal dam in Brittany, France, and the management of its water reservoir during low-flow periods. The management objectives

Table 6.6: Correspondence between percentage of initial value and actual values of the management variables studied in the risk sensitivity analysis.

| % of initial value | Objective level (m NGF) | Fish pass ($Mm^3$) | Siphons ($Mm^3$) | Drinking water supply ($Mm^3$) | Lock ($Mm^3$) |
|---|---|---|---|---|---|
| 150 | 3.00 | 0.60 | 0.495 | 0.15 | 0.045 |
| 140 | 2.80 | 0.56 | 0.462 | 0.14 | 0.042 |
| 130 | 2.60 | 0.52 | 0.429 | 0.13 | 0.039 |
| 120 | 2.40 | 0.48 | 0.396 | 0.12 | 0.036 |
| 110 | 2.20 | 0.44 | 0.363 | 0.11 | 0.033 |
| **100** | **2.00** | **0.40** | **0.330** | **0.10** | **0.030** |
| 90 | 1.80 | 0.36 | 0.297 | 0.09 | 0.027 |
| 80 | 1.60 | 0.32 | 0.264 | 0.08 | 0.024 |
| 70 | 1.40 | 0.28 | 0.231 | 0.07 | 0.021 |
| 60 | 1.20 | 0.24 | 0.198 | 0.06 | 0.018 |
| 50 | 1.00 | 0.20 | 0.165 | 0.05 | 0.015 |
| 40 | 0.80 | 0.16 | 0.132 | 0.04 | 0.012 |
| 30 | 0.60 | 0.12 | 0.099 | 0.03 | 0.009 |
| 20 | 0.40 | 0.08 | 0.066 | 0.02 | 0.006 |
| 10 | 0.20 | 0.04 | 0.033 | 0.01 | 0.003 |
| 0 | 0.00 | 0.00 | 0.000 | 0.00 | 0.000 |

and the elements subject to decision in the dam were detailed. Available data for the management of the dam and hydrometeorological conditions were listed. A water balance model of the reservoir was proposed. Its use in forecasting conditions allowed us to evaluate the potential of a risk assessment tool proposed to detect possible situations of tensions during low-flow periods.

A first analysis of the available data highlighted the relative importance of the elements of the dam in terms of volumes. The streamflow from the upstream catchment is the main inflow to the reservoir, and the volumes evacuated through the shutters and sluice gates are the main outflows. In practice, shutters and sluice gates are the parameters of adjustment for dam managers to evacuate excess water from the reservoir. However, during the summer season, other outflows associated with drinking water supply, lock openings and the fish pass are non-negligible. This was first observed in the analysis of yearly inflows and outflows, and confirmed at the monthly scale. Monthly inflows and outflows were quite variable depending on the month of the May to October period and on hydrological conditions. They were therefore associated with different management strategies. The analysis of inflow and outflow values at the daily time step also showed that the reservoir is extremely reactive. This characteristic of the reservoir caused the reservoir balance model to be very sensitive to errors in input data. Errors in preliminary simulations seemed to originate from poor peak simulations or poor low-flow observations.

Based on these analyses, and in close collaboration with the dam managers, a risk assessment tool was developed. The aim was to investigate how such a tool could foresee the risks of not being able to maintain a minimum level in the reservoir. The tool was developed in a framework that minimized the impact of the reactivity of the reservoir, with the assumption that outflows through the shutters and sluice gates were optimized to maintain a constant objective level in the reservoir. An objective level was set and the minor outflows were assumed constant. Examples of simulations and graphs were provided for the May to October seasons of 2005 to 2010. The risk assessment tool was developed to summarize the information of these simulations in terms of risk of reaching a minimum acceptable reservoir level. The graphs predicted both the average number of days below this reservoir level and the number of ensemble members reaching this level. We showed that these risk graphs were sensitive to initial hydrological conditions. The sensitivity of the predicted number of days and members below the minimum reservoir level was also quantified. These two variables were sensitive to the volumes evacuated through the fish pass and the siphon. Risk parameters were also sensitive to the objective level to be maintained in the reservoir, even though, in practice, there is little possibility to change this objective level due to several management constraints in the reservoir.

In the following chapter, we investigate how risk-based information can be used in reservoir management. We developed a game experiment to better understand how decision-makers react to long-term information to decide on reservoir releases.

# 7

# An experiment on risk-based decision-making in water management using monthly probabilistic forecasts

**This chapter is based on a paper published in *Bulletin of American Meteorological Society*:** Crochemore L., Ramos M.-H., Pappenberger F., van Andel S. J., Wood A. W. (2016), An experiment on risk-based decision-making in water management using monthly probabilistic forecasts. doi:10.1175/BAMS-D-14-00270.1.

**Résumé**

Les prévisions probabilistes ou d'ensemble, parce qu'elles prennent en compte les incertitudes de prévision, peuvent aider la prise de décision en contexte de risque. Des prévisions de débits aux échéances mensuelles ou saisonnières sont déjà utilisées opérationnellement en gestion de réservoirs, pour des objectifs tels que la répartition de la ressource, l'optimisation des lâchers d'eau ou l'anticipation du risque sécheresse. Dans le chapitre précédent, nous avons notamment proposé un outil d'évaluation des risques en période de basses eaux pour le barrage d'Arzal. Alors qu'il existe de nombreuses études cherchant à estimer la valeur des prévisions hydrométéorologiques d'ensemble pour de telles applications, peu étudient leur rôle pour la prise de décision. Les jeux de rôles peuvent s'avérer très utiles pour mieux comprendre le processus complexe de prise de décision en contexte de risque.

Ce chapitre propose une expérience, sous forme de jeu de rôle, pour mieux comprendre l'usage des prévisions probabilistes à longue échéance pour la décision de lâcher de réservoir. Durant le jeu, les participants endossaient le rôle de gestionnaire de réservoir. À partir d'une séquence de prévisions mensuelles d'apports au réservoir, et étant donnés un objectif de remplissage et des contraintes de lâchers, les participants décidaient séquentiellement des lâchers qu'ils allaient effectuer pour les mois à venir. À la fin de chaque mois, soit après chaque décision, les conséquences des décisions prises le mois précédent étaient évaluées en fonction des apports effectivement observés.

Pour cette étude, 162 feuilles de résultat collectées lors de huit évènements ont permis de mettre en évidence l'enjeu mais aussi la difficulté de prendre à la fois en compte l'information probabiliste et l'information à long terme. Pendant les séquences de jeu, un évènement de crue survenait au mois de juin. La stratégie permettant de finir le jeu sans faire déborder le réservoir consistait à vider progressivement le réservoir dans les mois précédant l'évènement pour pouvoir recevoir les apports de la crue de juin. L'analyse des feuilles de résultats a montré que, sur l'ensemble des séances, seulement 20 % des participants avaient réussi à finir la séquence de décisions sans faire déborder leur réservoir et en respectant les contraintes de lâchers. Le manque d'anticipation de l'évènement ou la sous-estimation de l'évènement à venir en terme de volume sont des explications possibles à ce faible taux de réussite. La manière de communiquer les prévisions et la quantité d'informations à intégrer dans le temps imparti pouvaient aussi présenter une difficulté pour les participants et influer la prise de décision. De manière plus générale, l'usage du jeu de rôle a permis de faciliter et créer un contexte favorable à la discussion des enjeux et limites à l'usage de l'information probabiliste à long-terme dans la prise de décision séquentielle.

**Abstract**

The use of probabilistic forecasts is necessary to take into account uncertainties and allow for optimal risk-based decisions in streamflow forecasting at monthly to seasonal lead times. Such probabilistic forecasts have long been used by practitioners in the operation of water reservoirs, in water allocation and management, and more recently in drought preparedness activities. Various studies assert the potential value of hydrometeorological forecasting efforts, but few investigate how these forecasts are used in the decision-making process. Role-play games can help scientists, managers and decision-makers understand the extremely complex process behind risk-based decision.

In this chapter, we present an experiment focusing on the use of probabilistic forecasts to make decisions on reservoir outflows. The setup was a risk-based game, during which participants acted as water managers. Participants determined monthly reservoir releases based on a sequence of probabilistic inflow forecasts, reservoir volume objectives and release constraints. After each decision, consequences were evaluated based on the actual inflow.

The analysis of 162 game sheets collected after eight applications of the game illustrates the importance of leveraging not only the probabilistic information in the forecasts but also predictions for a range of lead times. Winning strategies tended to gradually empty the reservoir in the months before the peak inflow period to accommodate its volume and avoid overtopping. Twenty percent of the participants managed to do so and finished the management period without having exceeded the maximum reservoir capacity or violating downstream release constraints. The role-playing approach successfully created an open atmosphere to discuss the challenges of using probabilistic forecasts in sequential .

## 7.1 Introduction

Seasonal climate forecasts are used in a large number of water resources applications ranging from droughts (Anderson *et al.*, 2000), urban water (Chiew *et al.*, 2000), hydropower operations (Block, 2011), water supply (Bracken *et al.*, 2010), water allocation (Mushtaq *et al.*, 2012), agriculture (Ghile and Schulze, 2008) and ground water levels (Guo *et al.*, 2009). Without exception, seasonal forecasting systems are probabilistic. They incorporate uncertainties about the future state of the climate (Brown and Ward, 2013), which is especially useful for risk assessment.

Coelho and Costa (2010) defined several challenges for integrating seasonal climate forecasts in operational management, ranging from the production of the forecasts to the effective implementation into user applications (e.g. Hartmann *et al.*, 2002; Lemos *et al.*, 2002). In water resource applications, this often requires translating predictions of precipitation and temperature into predictions of streamflow or inflows to reservoirs (Regonda *et al.*, 2011). Practitioners can then make management decisions that are informed by the hydrologic forecasts.

Coelho and Costa (2010) particularly emphasise end-user as a key challenge when implementing seasonal forecasts for water management, even given the common existence of quantitative water decision support systems (e.g. Dutta *et al.*, 2013; Regonda *et al.*, 2011). The process and the policies for water management are extremely complex, for they have to satisfy a large range of possibly conflicting objectives, comprising technical, socio-economic and environmental issues, and also cover a considerable range of hydrologic conditions (Simonović and Marino, 1982; Welsh *et al.*, 2013). A number of studies connect forecast performance to (e.g. Golembesky *et al.*, 2009), but few recognize that a skilful forecast does not necessarily lead to the forecast actually used by decision makers (Chiew *et al.*, 2003; Kiem and Verdon-Kidd, 2011; Ritchie *et al.*, 2004).

In addition, almost all studies are unable to incorporate all operating rules or key decisions due to the complexity of the task. Technical complexities together with intricate governance settings contribute to barriers which lower the uptake of seasonal climate forecasts in water resource management (for a detailed review see Kirchhoff *et al.*, 2013; Lemos, 2008). Amongst others, these barriers include the challenge in incorporating information in the process, and the often insufficient human and institutional capacities. Kirchhoff *et al.* (2013) highlight individual water manager behaviour and risk perception as important areas which need to be addressed to realise the value of probabilistic seasonal forecasts (Block and Goddard, 2012).

These issues can be addressed through an improved process, using structured models, and training. It is however extremely complex to replicate a full reality, with all possible consequences. The use of role-play games can aid this process, whilst allowing the investigation of key research questions, such as how decision rules can be formulated or whether probabilistic forecasts lead to better decisions (Ramos *et al.*, 2013). The decision process includes a range of potential actions, a number of possible events, various consequences for each combination of action and event, and a set of probabilities for each combination (Faber and Stedinger, 2001; Sankarasubramanian *et al.*, 2009). These can be controlled in a game setting, whilst providing an experience which can be close to reality (Cannon-Bowers and Bell, 1997) and helpful to enhance understanding.

This chapter presents a game experiment that focuses on a realistic decision sequence for managing a reservoir — one based on actual hydrology, seasonal forecasts, and typical reservoir management objectives from a setting in the western United States (US). The game was played with different groups of students, researchers, operational hydrologists, forecasters, decision-makers and water managers during conferences and meetings. Players were asked to manage a reservoir used for flood control and water supply, for a four-month period (referred to as the management period). They were presented with a monthly series of probabilistic inflow forecasts for runoff in the coming 1 to 4 months: such forecasts are referred to as seasonal forecasts in practice due to their monthly to seasonal lead times. Participants were required to plan outflows at each decision step while complying with reservoir capacity and release constraints. The objectives of this chapter are to present the results obtained and to analyse the decision making process of the participants.

In the following sections, we present the game setup, the way it unfolded and how participants managed their reservoirs. The last section is dedicated to discussion and conclusions.

## 7.2 Material

### 7.2.1 Game setup

The setting of the game is a reservoir in a watershed with a pronounced annual runoff cycle defined by a winter-spring snow accumulation and spring-summer melt period, followed by a low runoff regime in the summer and fall. The reservoir management objectives are typical of many managed systems: water supply, minimum environmental releases and flood control. The game was adapted from training material for a course given at the 91$^{st}$ Annual Meeting of the American Meteorological Society. It was modified to be played in an auditorium, in 20-25 minutes, and to be accessible for audiences ranging from students, researchers, operational hydrologists, forecasters, decision-makers and water managers.

**Game setting** Each participant plays the role of a water manager for 'Lake Dual', which is a reservoir with a capacity of 500 Mm$^3$ that serves two primary functions: water supply for 'Swof Town' and flood control for 'Safe Town'. The residents of 'Swof Town' would like to see the reservoir full (500 Mm$^3$) on August 1$^{st}$ to ensure drinking water supply until the end of summer, while the residents of 'Safe Town' are interested in keeping monthly releases below 60 $Mm^3$ to prevent flood damage to their homes. Probabilistic forecasts of monthly inflows are available and updated on the first day of each month during the management period running from April to August. At the beginning of each month, participants have to decide on the monthly reservoir releases for the remaining months in this period.

**Management objectives** The goal of each participant is to have the reservoir volume as close as possible to 500 $Mm^3$ on August $1^{st}$ without ever exceeding this maximum capacity. Participants also have to maintain a minimum release of 15 $Mm^3$ for environmental flow and their maximum release cannot exceed 60 $Mm^3$. As a penalty, participants are fired from their management role in the game if they fail to meet the constraint of maximum capacity. At the end of the game, the 'winner' is the manager that has the highest reservoir volume on August $1^{st}$ without having exceeded its maximum capacity during the management period.

### 7.2.2 Playing the game

To collect the results of each participant, a worksheet was distributed at the beginning of the game. It allowed participants to record their releases and update their reservoir volume. It also provided information on the long-term flow climatology for each month (the median inflow in $Mm^3$ over the past 30 years). Before starting the game, an example was given for the month of March. Participants were guided on how to fill in the worksheet. Lastly, we asked for a volunteer to play the game in front of the group, bringing a more lively atmosphere to the experience. Since computations for the volunteer could not be executed in real-time, the volunteer was presented with three pre-defined options of releases. They were designed after several test-plays with small groups. For the volunteer, all possible combinations of sequential decisions were pre-calculated, but only the chosen sequence was displayed. Furthermore, the volunteer's choice and play did not interfere with the play of the other participants, because participants had to define their releases before the choices available for the volunteer were presented.

To illustrate the forecast-decision procedure, Figure 7.1 shows the slides presented for the first month. First the inflow forecasts are displayed with the help of boxplots showing the 5 % (min), 25 %, 50 %, 75 %, and the 95 % (max) percentiles (Figure 7.1a) and participants are given a few minutes to decide on their reservoir releases. On the next slide (Figure 7.1b), it is the volunteer's turn to make a decision, choosing among options A, B and C. The next slide (Figure 7.1c) displays the actual inflow of the month and participants update their reservoir levels accordingly, balancing the inflow with the release they had decided on for that month. In the last slide of the round (Figure 7.1d), we assess the volunteer's decision, by showing the calculation of reservoir volume at the end of the month for the release option chosen by the volunteer. It also indicates whether the volunteer still has a job. This marks the end of a forecast-decision-update sequence and the game moves on to the next decision (i.e., next month).

The game is a repetition of this sequence of steps for forecasts issued once a month from April 1st to July 1st until we reach August 1st. If participants are fired before the last round, they are encouraged to keep playing and try to recover their jobs by lowering the reservoir volume below its maximum capacity.

A major flow event was included to occur in June to test the participants' capacity to hedge for the possibility of high inflows to the reservoir. In addition, the probabilistic forecasts were designed such that their median values were below the actual inflow, to discriminate whether participants became sensitive to the risk represented by the upper tail of the forecast distribution. To help participants spot this pitfall, forecasts displayed both observed and forecast inflows for the past month.

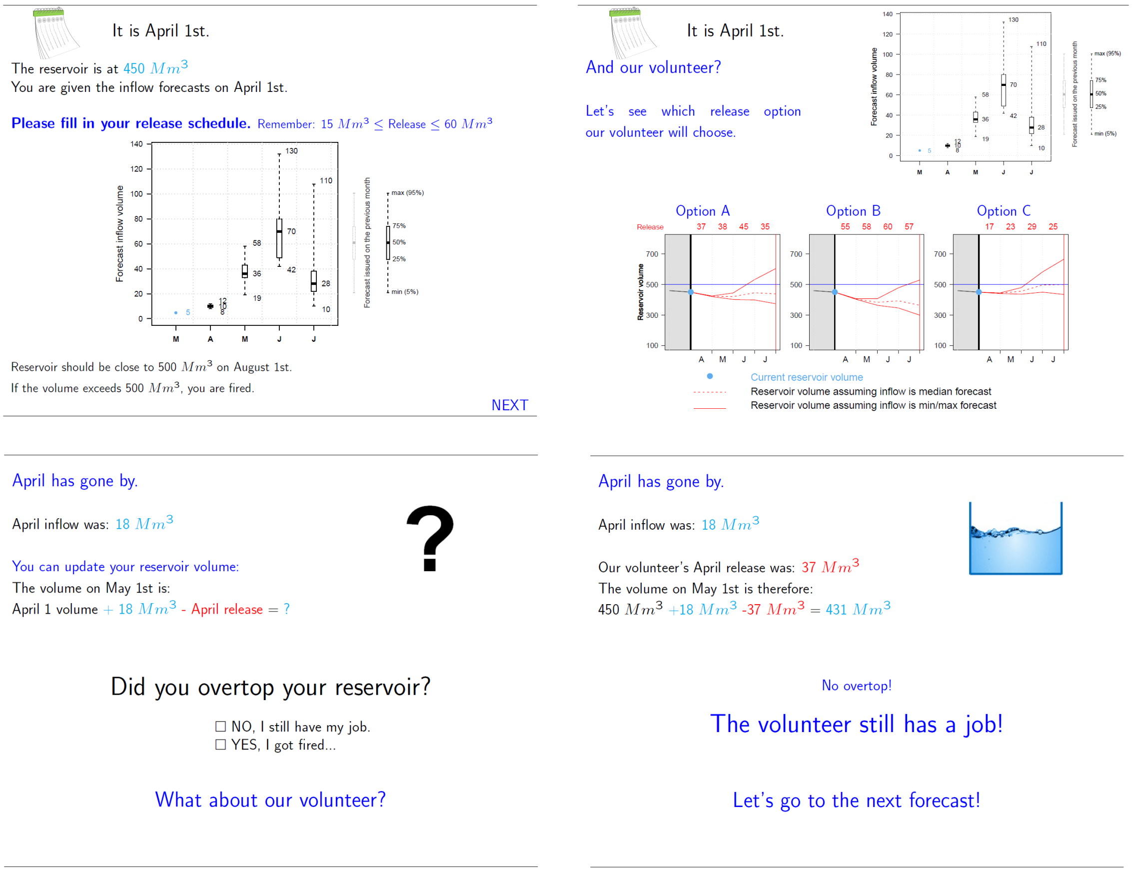



Figure 7.1: Example of a sequence of forecast-decision in the game: forecast issued on April 1st and results for a volunteer that chose option C as release schedule.

Table 7.1: Characteristics of the applications of the game.

| Context of the application (oral presentations) | Place | Number of work-sheets collected | Participants |
|---|---|---|---|
| European Geophysical Union General Assembly 2013 (Session *Ensemble hydro-meteorological forecasting for improved risk management: across scales and applications*) | Vienna, Austria | 85 | Students, researchers, operational hydrologists, forecasters, decision-makers, water managers |
| Users workshop of the European Flood Awareness System (EFAS) | Reading, UK | 23 | Operational forecasters from EU flood forecasting national services |
| Trans-national 'Drought team' meeting of the Interreg NEW IVB DROP project | Brittany, France | 8 | Operational hydrologists and water-supply reservoir operators from France and Germany |
| Seminar at Université du Québec á Chicoutimi (UQAC) | Chicoutimi, Canada | 11 | Undergraduate and post-graduate students, professors |
| Seminar at Centre d'Expertise Hydrique du Québec (CEHQ) | Québec city, Canada | 10 | Operational hydrologists and flood forecasters |
| Seminar at Hydro-Québec's research institute (IREQ) | Varennes, Canada | 11 | Researchers, operational forecasters, and decision-makers |
| Training course on Predictability, Diagnostics and Forecasting at ECMWF | Reading, UK | 20 | Undergraduate and post-graduate students |
| HEPEX $10^{th}$ Anniversary international workshop | Maryland, USA | 35 | Scientists, operational forecasters in meteorology and hydrology, decision-makers |
| TOTAL | | 203 | |

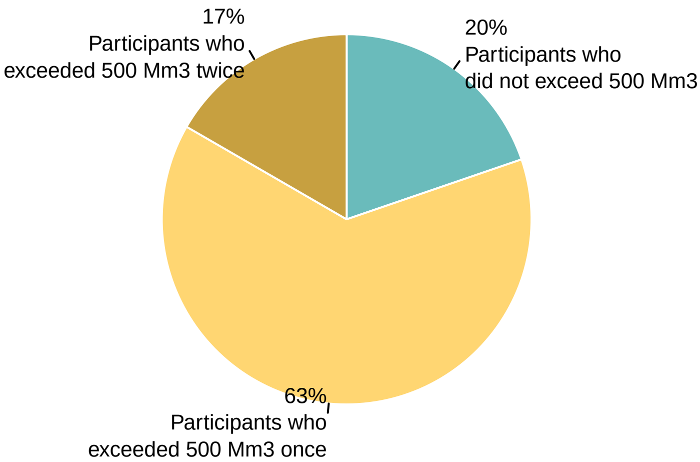



Figure 7.2: Results of participants in terms of the main reservoir constraint: not exceed the maximum reservoir capacity of 500 $Mm^3$.

## 7.3 Results

### 7.3.1 Worksheets collected

We collected a total of 203 worksheets through eight distinct presentations of the game (Table 7.1). Seventy-five percent presented release schedules filled in for each month (and not only for the first month). Thirty-two percent showed a miscomprehension of the release constraints and had releases greater than 60 $Mm^3$ or less than 15 $Mm^3$. Miscalculations were observed in 63 worksheets: most were small computational errors that did not impact the reasoning process and few were miscalculations of the reservoir inflow-outflow balance. We discarded the worksheets that did not respect the constraints in the releases for the first month of the schedule and those that presented miscalculations that could have led to erroneous decisions. In the end, the analysis considered 162 worksheets, 126 of which had complete release schedules. The volunteer play was not considered.

### 7.3.2 Decision-makers' behavior during the game: who won? who lost?

In terms of the main decision-makers' constraint, i.e., keep the reservoir level below 500 $Mm^3$, we observe (Figure 7.2) that twenty percent of the participants (32) never exceeded the reservoir capacity during the management period and kept their jobs until the end of the game. Eighty percent of the participants (130) exceeded the reservoir capacity at some point during the game. Among them, 103 exceeded the reservoir capacity once, in June, and lost their jobs, but then managed to recover (i.e., released enough in July to achieve a reservoir level back below 500 $Mm^3$ on August 1st), and 27 exceeded the reservoir capacity in June, but were unable to bring the volume below 500 $Mm^3$ at the end.

Figure 7.3 shows the monthly evolution of reservoir volumes (Figure 7.3a) and the distribution of releases at one-month lead (Figure 7.3b), for participants who exceeded the maximum reservoir capacity and for those who never did. All participants who won decreased their reservoir volumes by 10 to 50 $Mm^3$ in the first two months. These two months were crucial to make the difference between winners and losers. April and May inflows summed to 73 $Mm^3$, which, given the minimum compulsory release (15 $Mm^3$ for each month) and the initial reservoir level (450 $Mm^3$), forced participants to reach at most a reservoir volume of 493 $Mm^3$ at the end of May. This level left little flexibility to keep the volume lower than the maximum capacity in the subsequent months. In the group of participants who won, releases are higher than 30 $Mm^3$ in April, combined with, in most cases, additional relatively high releases in May (higher than 50 $Mm^3$). In contrast, more than half of the participants who lost opted to release the minimum allowed in April and consequently, saw their reservoir volumes increase by the end of the month since the April inflow (18 $Mm^3$) was greater than their releases. In May, 75 % of these participants had their reservoir volumes greater or equal to the initial volume of 450 $Mm^3$, with the highest value being the maximum possible, 493 $Mm^3$.

In June, the month for which the high runoff event was forecast, 80 % of all participants released the maximum 60 $Mm^3$. This represents 62 % of the participants who won and 84 % of the participants who lost. At the end of the month, the reservoir volumes of participants who lost were between 503 and 553 $Mm^3$, while the volumes of participants who won were between 473 and 498 $Mm^3$.

All participants who won decreased their reservoir levels in the first two months, anticipating the peak runoff month, which was reflected in both the forecasts and the climatology. In the beginning of June, their reservoirs were then low enough to collect the water from the high inflow without overtopping. Participants who lost had a general tendency of increasing their reservoir volumes after the first decisions, reflecting emphasis on the goal of having the reservoir level as close to 500 $Mm^3$ as possible on August 1st, rather than on the risk of high inflow in June. Therefore, in June, they suffered the consequence of overtopping the reservoir. The maximum reservoir volume obtained by a participant reached 553 $Mm^3$ at the end of June, which corresponds to releasing the minimum 15 $Mm^3$ in April and in May, and the maximum 60 $Mm^3$ in June. This is the profile of a decision-maker that ignored or failed to comprehend the implications of the June inflow forecasts and climatology, and became trapped by release constraints in June.

To recover their role as reservoir managers, participants who had exceeded the reservoir capacity by the end of June had to release enough in July to decrease their reservoir volumes by the end of the month. Almost 75 % of this group released more

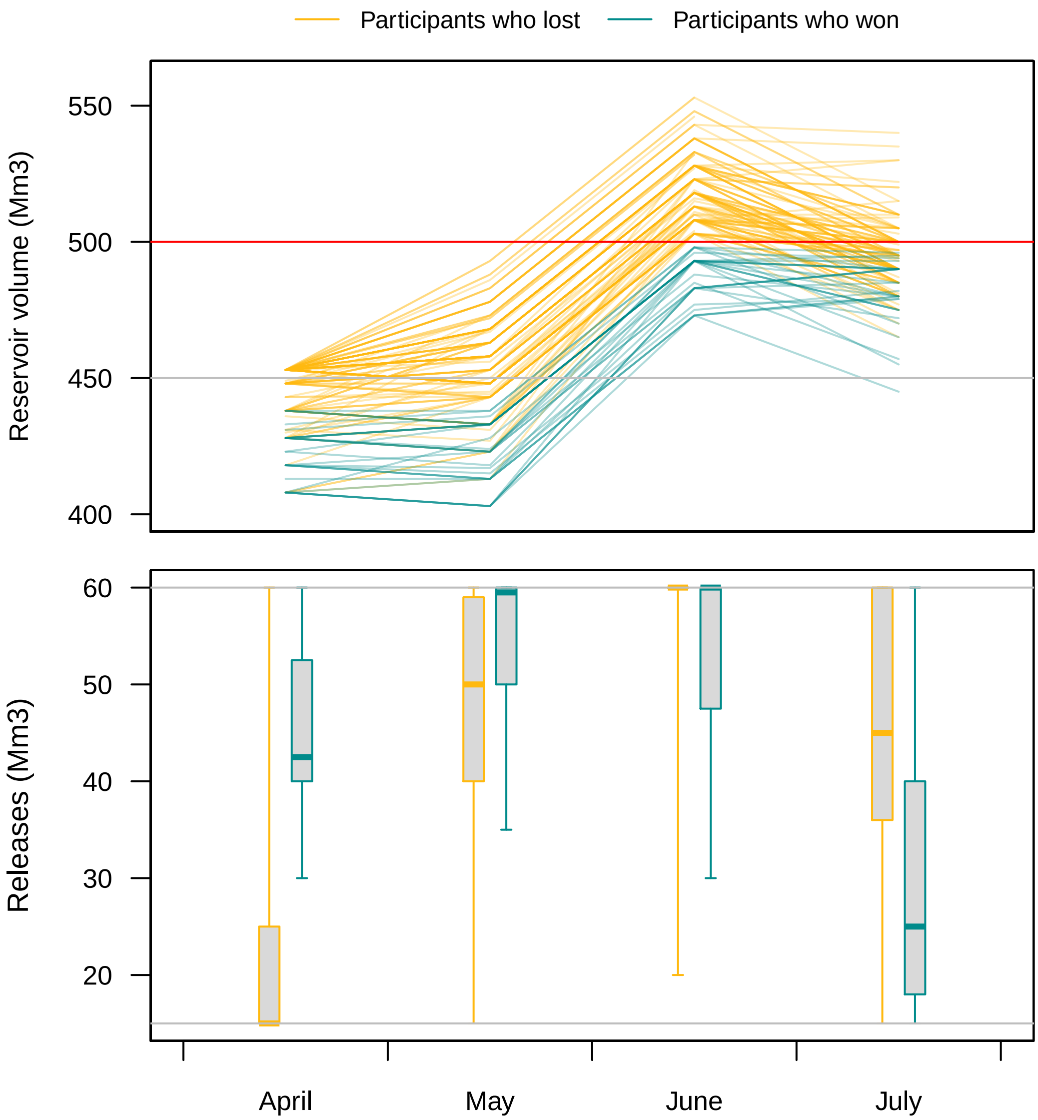



Figure 7.3: (a) Evolution of reservoir volumes at the end of each month and (b) distribution of releases for the coming month (one-month lead) for participants who exceeded the reservoir capacity (participants who lost, 130 worksheets, in yellow) and participants who did not (participants who won, 32 worksheets, in blue). In (a), the grey and red horizontal lines mark the reservoir volume at the beginning of the game (450 $Mm^3$) and the constraint of reservoir capacity (500 $Mm^3$), respectively. In (b), the grey horizontal lines mark the minimum (15 $Mm^3$) and the maximum (60 $Mm^3$) allowed for releases. Box plots display minimum value, percentiles 25 %, 50 %, 75 % and maximum value over 162 participants.

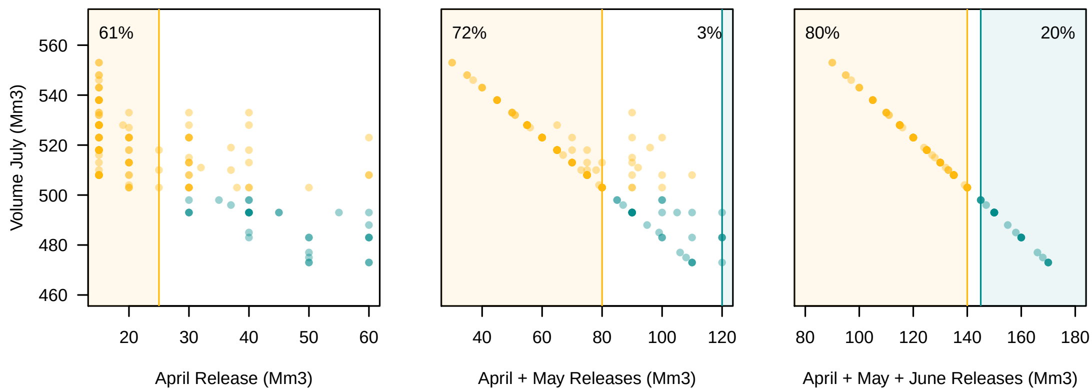



Figure 7.4: Reservoir volumes of 162 participants at the end of July as a function of the one-month lead releases for April (a), April plus May (b) and the sum of April, May and June (c). Participants who won (never exceeded the maximum reservoir capacity) are indicated in blue, and participants who lost, in yellow. The darker the dot, the more participants adopted the volume/release scenario. Shaded areas highlight the strict separation between winners and losers by indicating the domain occupied exclusively by participants who won, and by participants who lost, in the blue and yellow colors respectively. A percentage indicates the proportions of dots in these areas.

than 40 Mm$^3$ in July, while 75 % of the participants who never exceeded the maximum capacity released, in the same month, less than 40 Mm$^3$. The return to a reservoir volume below 500 Mm$^3$ was achieved by 79 % of the participants who had lost their jobs with the high inflow in June. Despite their efforts, 27 unfortunate players saw their reservoirs remain above the 'spill' volume.

The urgency to decrease the reservoir volume after June to avoid further overtopping, or maybe just to secure their jobs, led participants who had lost to release large quantities of water in July. Out of the 32 participants who had not exceeded 500 Mm$^3$ after the event in June and who had then more flexibility to adjust their reservoir volume so as to bring it as close as possible to the goal on August 1$^{st}$, 11 were able to increase their reservoir volumes in July by managing their releases.

### 7.3.3 And the winner is...: optimal one-month lead release schedule

Figure 7.4 presents the reservoir volumes of each participant at the end of July as a function of their one-month lead releases. We consider, progressively, their releases in April (Figure 7.4a), the summed releases for the months of April and May (Figure 7.4b) and the summed releases for all the months prior to the high inflow event, i.e., April, May and June (Figure 7.4c). This figure allows a retrospective analysis of which decisions led to overtopping the reservoir.

Figure 7.4a shows that participants who released 25 Mm$^3$ or less in April (61 % of all 162 participants) lost their jobs on July 1st. The same is observed for participants who released an accumulated volume of 80 $Mm^3$ or less by May (Figure 7.4b). By this time, they already represent 72% of all participants. By the end of June, participants who had released an accumulated volume of 140 $Mm^3$ or less (80 % of all participants; Figure 7.4c) had lost the game. On the opposite end of the release spectrum, all participants who released 120 $Mm^3$ by May (i.e., that released the maximum allowed in April and May) did not overtop their reservoirs during the game (Figure 7.4b). At the end of the first three sequential decisions, all participants who had released more than 145 $Mm^3$ had their reservoirs prepared for the high runoff event in June (Figure 7.4c).

In order not to overtop their reservoirs, participants had to lower the initial reservoir volume of 450 $Mm^3$ by at least 10 $Mm^3$ in the first two decisions. In terms of releases, this means that they had to release at least 83 $Mm^3$ over the first two decisions (given the 18 $Mm^3$ and 55 $Mm^3$ inflows of April and May, respectively). Given the constraint of maximum release (60 $Mm^3$), these two months were thus essential to adjust the reservoir volume prior to the high inflow event. From the worksheets, it emerges that the high-score winner applied the following sequence of releases - 35; 50; 60; 24 $Mm^3$ — to achieve the following volumes: 433; 438; 498; 496 $Mm^3$. This winner was among the audience of the HEPEX 10th Anniversary workshop, which was, most probably, the venue with the most specialized users of probabilistic forecasts in the audience.

As noted earlier, the game was designed such that the seasonal forecasts were under-predicting the coming high inflow event if judged on the 50 % percentile. In fact, the observed peak inflow of June was close to the 95 % percentiles of the first three (April, May, June) probabilistic forecasts for June. These forecasts foreshadowed that an upcoming major event was possible, albeit with a low forecast probability of occurring. The flow climatology information in the worksheets indicated that June was historically a month of high inflows and also provided a warning for participants to be cautious and prepare for the coming event with an appropriate release strategy.

### 7.3.4 Evolution of release schedules

The way participants were planning their releases months ahead, and how they changed their planning or not as the June high inflow approached, was investigated with the help of the 126 worksheets (out of 162) that had release schedules fully filled. In both groups, i.e., participants who lost (102 worksheets) and participants who won (24 worksheets), we had approximately the same proportion of players that fully filled in their release schedules (78 % and 75 %, respectively).

We observe that participants who won had basically planned their releases for the month of May and June already in the first decision on April 1st. When May 1st and later June 1st arrived, the majority confirmed their previous decisions or just increased their releases of approximately 5 to 10 $Mm^3$ more to accommodate the high inflows in the reservoir without overtopping.

On the other hand, participants who lost were, in general, planning very low releases on the first months and already, since the first decision in April, planning to release the maximum in June. What they had not anticipated was that this would not be enough to accommodate the high inflow event and that a better strategy would have been to gradually empty the reservoir already in April and May. On June 1st, players from this group may have been frustrated by the fact that they had their reservoir volumes too high but could not release more than the maximum allowed.

In general, the planned releases for July were progressively increased when moving from April to June in both groups. On July 1st, however, when the only release they had to plan was for the coming month and the highest inflow had already passed, the majority in the group of participants who won was able to decrease the values they had planned to release previously on June 1st and, therefore, better target the final goal of having the reservoir volume as close as possible to 500 $Mm^3$ on August 1st. On the other hand, in the group of participants who lost, half of the players had to decide on releasing more than what they had scheduled in the previous decisions to have a chance of getting their reservoir below 500 $Mm^3$ and, consequently, their jobs back.

### 7.3.5 How might participants have used the probabilistic forecasts and the flow climatology when making decisions?

It was left to participants to choose how they would take into account forecast and climatology information in their decisions. We could not, unfortunately, follow this process within each participant's mind. Nevertheless, using the worksheets only, we tried to identify which forecast quantile participants based their releases on. To do so, we calculated the August 1st reservoir volumes they would have obtained if the observed inflows had consistently matched one of the forecast quantiles, or, alternatively, the flow climatology. Volumes obtained with the observed inflows were also estimated. This was done for each decision step (month), so that we could also evaluate the release planning strategy. Figure 7.5 shows the results.

The players who took proactive actions as early as April and May to balance the game objective with the overtopping risk likely focused on upper quantiles of the forecasts (Q75 and Q95), with maybe also some support from the flow climatology. Indeed, most of these participants would not have overtopped their reservoirs had the Q75 and even the Q95 forecast quantiles been verified as observed inflows throughout the game, as early as in the first decision step. On the third and fourth decisions, they seemed to have set aside flow climatology and rather used an 'adjusted' upper forecast quantile (Q95) to evaluate the possible observed inflow and optimise their releases to the goals of the game. This might have been the result of having previously noted that the forecasts were, in general, under-predicting the observed inflows. The fact that part of the distribution for August volumes given actual observed flows is above the reservoir maximum during the first two decisions (April, May) means that participants tolerated or were not able to eliminate the overtopping risk early in the management period, but took steps to eliminate it when the potentially high inflow month was imminent.

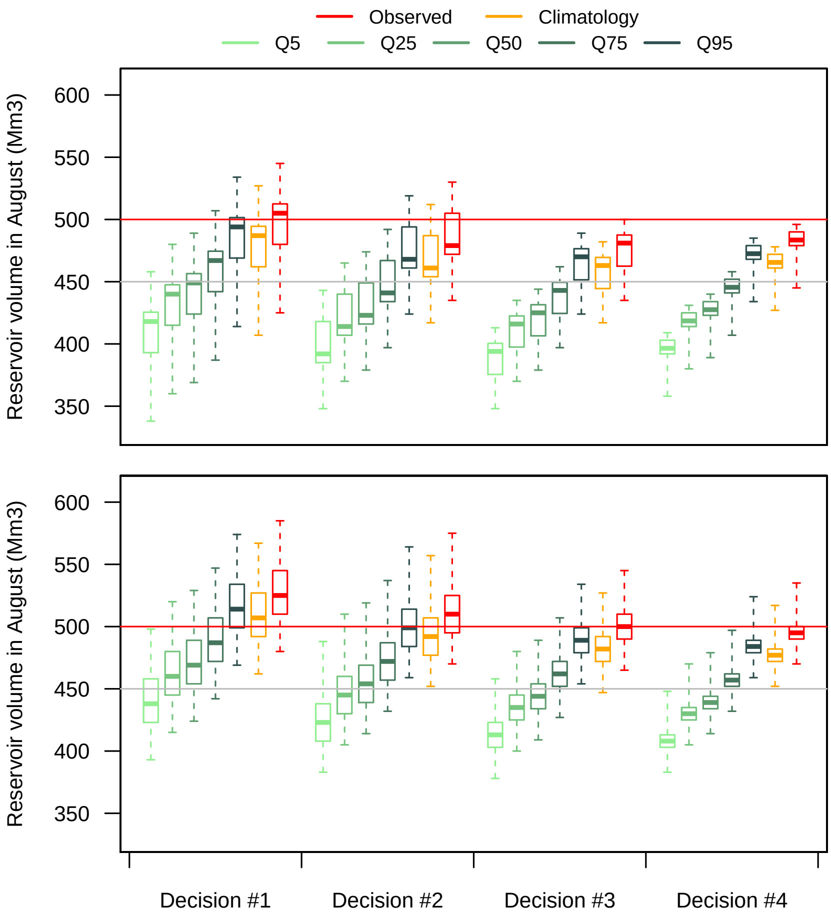



Figure 7.5: Volumes that participants would have in their reservoirs on August 1st by applying their release programs at each decision month (April-July) to the cases where inflows are equal to: the forecast quantiles Q5 to Q95 (shades of green), the flow climatology (orange) and the observed inflows (red). Boxplots represent statistics over 24 participants who won (top) and 102 participants who lost (bottom).

In the group of participants who lost, most participants might have been guided by the medium to lower quantiles (Q50 and Q25) in the first decisions. Figure 7.5 shows that if the upper quantile forecasts or even the flow climatology values had verified as observed inflows, these participants would have reached reservoir volumes much higher than the allowed 500 $Mm^3$ in August. In contrast, if lower quantiles had verified as observed inflows, most would have been safe from overtopping their reservoirs. Later on, on the third and fourth decisions, these participants might have acted based on the upper forecast quantiles (Q75 and Q90) and, eventually, on the flow climatology to plan their releases.

## 7.4 Discussion and conclusions

This chapter presented the results of a game experiment designed to mimic risk-based decision-making in water management using probabilistic forecasts of inflows to a reservoir. From the analysis of the worksheets collected during the application of the game in eight different contexts, we were able to illustrate key issues on the use of probabilistic forecasts in sequential decision making.

In the game setup, seasonal forecasts had to be viewed as a whole and not as independent monthly forecasts. Even though the forecast time evolution warned the participants about a high inflow event, monthly 50 % percentile forecast values were under-estimating flows. Given the reservoir constraints, as expressed in the rules of the game, and the goals of the management, it was necessary to look at forecasts months ahead before deciding on the reservoir releases. A winning strategy would be the one that would gradually empty the reservoir two months ahead of the expected high inflow to accommodate its potential volume without overtopping. Although the worksheets were designed to invite participants to adopt this long-term approach, notably by asking them to schedule their releases for the whole management period, still about 25 % of the participants filled their releases for the coming month only.

Approximately 20 % of the participants to the game were able to finish the management period without exceeding the maximum reservoir capacity at any time during the sequenced decisions, and approximately 17 % of the participants not only caused the reservoir to overtop, but also were unable to bring its volume back below the threshold. Winners were those who had programmed the releases in a way that they succeeded to adequately decrease their reservoir volumes in the first two months, anticipating the potential high inflows. Losers of the game were those who did not recognize early enough the significance of the high inflow risk, or did not comprehend its potential impact on the reservoir volume.

In this chapter, answers that showed signs of miscomprehension of the balance equation were not used in the interpretation. Therefore, it has been implicitly assumed in the analysis of the remaining results that the participants were able to take full advantage of the proposed information, once given the necessary tools to read the probabilistic graphs and understand the climatological values at the beginning of the game. The results based on the remaining answers may however still be influenced by the comprehension of the proposed tools and the way they were presented. It is possible that given the timed nature of the game (participants had to decide within a limited amount of time), some participants could understand the way forecasts were displayed, but were not fast enough in applying their understanding. Communicating forecasts and, more generally, probabilities is still a major challenge in hydrometeorology, and can be a barrier to the widespread use of ensemble forecasts in operational contexts.

The information collected from the worksheets shows which decisions were taken, but not why they were taken. Therefore, care has to be taken in the interpretation of the results of this game experiment. We can only speculate on what could have been the reasons for many participants to take a certain sequence of release decisions, and on what strategy winners had with respect to the forecast inflows. Therefore, even if the results indicate that the participants who won the game might have considered the upper quantiles of the forecasts, at least in their first decisions, information on why they followed this procedure is not available through the game setup. A possible solution to this limitation could be to expand the game by asking these relevant questions orally or directly in the worksheets. Recording whether participants had prior water resource management or forecasting experience could also help indicate the value of training toward improving the application of probabilistic information. Collecting specific information on the main occupation and background of each participant could also be helpful to evaluate how strategies may vary between different groups (e.g., students, managers or forecasters).

Finally, this game is a simplified representation of reality and does not intend to reproduce the full context of operational environments in reservoir management. Indeed, a participant who had real-life experience managing reservoirs noted during one game session that the winning strategy (described in Section 3.2) could have raised alarms in practice for having allowed too much flooding risk en route to achieving a near-perfect target level. Despite the simplifications, however, we received positive feedback after the different applications, even though howls of indignation were often heard in the rooms when the observed June inflow was revealed. The role-playing approach, and the penalty experienced by participants of being fired from their jobs, added a light touch to the experience and created a pleasant atmosphere to discuss the challenges of using probabilistic seasonal forecasts in sequential decision-making, where choices have delayed consequences. Notably, the game has been successfully used as support material during teaching and training activities.

# General conclusion

The quest for extending predictability in hydrometeorological forecasting is far from being over. Seasonal forecasts are a perfect example of this: many are curious about them and show an interest, but many remain sceptical, eyebrows often rise in their vicinity, and few trust them enough to implement them. This is a perfect demonstration of the stake of seasonal forecasting and of all that is left to prove and demonstrate. In streamflow forecasting, one can start by assessing the theoretical predictability of streamflows. How far ahead can we hope and should we expect skilful predictions? This approach to predictability is at the heart of recent publications and on-going research in several research teams around the world. This PhD work has tackled the matter by directly investigating the performance of seasonal forecasting systems. It has addressed two aspects that limit the widespread use of seasonal streamflow forecasts: the need to assess the quality of seasonal streamflow forecasts and the necessity to translate the seasonal information into risk information that answers user needs.

Throughout this PhD research, we have assessed the quality of seasonal streamflow forecasts in France and investigated their potential to enhance decision-making in reservoir management during periods of low flows and droughts. This research has more specifically addressed four aspects, listed in the Introduction of this thesis:

1. the bias correction of precipitation forecasts to improve seasonal streamflow forecasts;
2. the conditioning of historical data with seasonal precipitation forecasts to better capture low-flow characteristics;
3. the development of a risk assessment tool tailored to the needs of dam managers to help decision-making and alert against water shortages in summer;
4. the set up of a role-playing game to better understand role of seasonal probabilistic information in risk-based decision making.

We hereafter summarize the conclusions related to each of these questions, and list some perspectives to this work.

## Conclusions on seasonal streamflow forecasting in France

**1.** Bias correcting precipitation forecasts from climate services can improve the overall performance and the reliability of GCM-based seasonal streamflow forecasts.

We investigated the ability of precipitation forecasts issued by the System 4 GCM of ECMWF to forecast seasonal streamflows. The observed monthly biases of the System 4 precipitation in the sixteen French catchments often compensated over the year. This highlighted that efficient bias correction should address monthly biases rather than biases over the year. Eight bias correction methods were studied. The empirical distribution mapping of daily precipitations performed best: it allowed improving overall performance by improving forecast reliability, even though this was sometimes at the expense of forecast sharpness. The linear scaling method increased overall performance of streamflow forecasts by improving the sharpness of precipitation forecasts. Overall, a small improvement in the overall performance of precipitation forecasts led to greater improvements in the overall performance of streamflow forecasts. This result is an incentive to improving seasonal precipitation forecasts, either through improved climate model outputs or through post-treatment methods, to also improve seasonal streamflow forecasts.

**2.** Conditioning historical scenarios with GCM-based precipitation indices can improve the sharpness of climatology-based streamflow forecasts, and provide better forecasts of low-flow events.

The conditioning of historical data based on GCM precipitation forecasts allowed to take advantage of the reliability of historical data and the sharpness of meteorological forecasts. A conditioning based on forecasts of the anomaly in precipitations produced reliable forecasts, narrowed the spread of climatology-based forecasts and preserved their overall performance. This research has highlighted the trade-off between sharpness and reliability in the search for better forecast quality. Improving one may often be at the expense of the other. We can suggest that an universal optimal balance may not exist and that the best trade-off between sharpness and reliability is specific to the application. Lastly, an advantage of the proposed methods is that they may also serve as bias correction and downscaling of GCM outputs as highlighted by Carpenter and Georgakakos (2001). While the conditioning of historical precipitations can be seen as a bias correction prior to hydrological modelling, the conditioning of historical streamflows can directly serve as a bias correction of streamflow forecasts. Additionally, conditioning methods can help translate the large-scale seasonal GCM information into meteorological and hydrological scenarios observed in the catchment.

## Conclusions on seasonal forecasting for reservoir management

**3.** A risk assessment tool was developed to forecast the risk of not being able to maintain all water uses of the Arzal reservoir during the low-flow season.

An in-depth analysis of the quantitative management data of the Arzal reservoir, supported by several discussions with the managers of the IAV (the organization in charge of the management of the dam), highlighted the objectives and constraints of the dam. This analysis also helped understand how these constraints and objectives relate to hydrometeorological conditions. It has led to the development a risk assessment tool for the management of the reservoir. The seasonal streamflow forecasts used as input to a water balance model of the reservoir allowed us to quantitatively assess the risk of reaching a minimum acceptable reservoir level throughout the low-flow season. A pre-operational visualisation tool of this risk emerged from this preliminary work. A fully operational risk assessment tool would require further analysis after implementing the tool operationally, which would allow a complete evaluation of its strengths and a final tailoring to fulfil the user needs. Lastly, a sensitivity analysis of the risk assessed in the case of the 2005 drought, highlighted the importance of the objective management level in reducing the risk. Interestingly, the analysis also showed that increasing drinking water supply would have little impact on the quantified risk, while restricting the openings of the lock for boating can indirectly help reduce this risk.

**4.** A good communication and interpretation of seasonal products is key to taking full advantage of the seasonal information for decision-making in reservoir management.

A role-playing game experiment was conducted to investigate the process of sequential decision-making based on seasonal probabilistic forecasts in a theoretical reservoir. Overall, we observed that the game approach facilitated discussions on challenges and limits to using long-term information in decision-making. The experiment showed that seasonal probabilistic forecasts can be necessary in risk-based decision-making to anticipate extreme events months in advance. The experiment also highlighted the challenge of incorporating long-term probabilistic information in decision-making. Indeed, among the game participants, few were able to manage the theoretical reservoir throughout the season without overtopping it. Participants may not have sufficiently anticipated the hydrological event despite the seasonal forecasts. Also, the way the forecast information was communicated may have been a critical aspect impacting the decision-making.

## Collaboration with climate services and end-users was crucial

The conclusions of this PhD work are the fruit of active collaborations with climate services, represented by ECMWF, but also with end-users, here the dam managers of the Arzal reservoir. Rather than adopting a top-down approach from the climate services to the end-users, this work has involved all parties throughout the research process.

On the one hand, it was the end-users that initiated the collaboration on this PhD work within the Interreg IVB NWE DROP project. They were included in the scientific process and their experience was decisive in the development of adapted tools for reservoir management. An operational goal involving the work of this PhD thesis was to develop a pre-operational low-flow forecasting and water management system to support reservoir management and drought adaptation for the Arzal reservoir. The DROP project also allowed for exchanges with end-users from different horizons. We have been in close contact with the dam managers from WVER, in Germany, who had different experiences of similar issues, which gave an enlarged vision of this work. On the other hand, the scientists from ECMWF not only provided the seasonal precipitation forecasts used in this research, but also actively contributed to the application and evaluation of these forecasts for streamflow forecasting. They offered their contribution throughout the implementation of the precipitation forecasts within the streamflow forecasting framework. After the framework was implemented, they shared their expertise in the forecasts and participated in the interpretation of the results of this research.

## Perspectives

This PhD research builds on a forecasting system based on the precipitation forecasts of the ECMWF System 4 GCM, on the GR6J hydrological model and on a reservoir balance model developed for the purpose of this thesis. All the conclusions from the performance analyses carried out in this thesis thus rely on the performance of these models and may not be applicable to other forecasting systems. Future directions for this research are also highly related to this modelling framework.

A direct perspective is to include seasonal temperature forecasts in the streamflow forecasting framework. Throughout this thesis, the hydrological model was fed with the mean interannual evapotranspiration. However, studies have shown the importance of considering temperature when forecasting dry events, as in the case of the 2014 drought in California (Kollipara, 2015). Adding temperature forecasts to the analyses would, however, have implications on the bias corrections applied and on the selection of past scenarios to predict future events. Indeed, bias corrections would have to correct simultaneously precipitations and temperatures, and preserve their spatio-temporal consistency. A conditioning of past temperatures may also be challenging due to nonstationarities in long temperature records necessary to forecast streamflows based on historical meteorological data. A basic assumption in the way we built our conditioned forecasts is that past observations can represent current climate conditions, which may not be valid if trends are present in the observed time series.

In this thesis, we have investigated several ways to condition climatology in order to forecast streamflow. Other conditionings are extensively used in the literature, such as selections based on climate indices (e.g. the North Atlantic Oscillation in Europe). Yet, other approaches, such as the Bayesian Model Averaging (BMA) or the nonhomogeneous Gaussian regression (NGR), propose to optimally combine multiple forecasting systems by learning from their past performance. It would be interesting to compare the results from such combinations with the results of the conditioning presented in this thesis. Furthermore, the BMA and NGR methods could be interpreted as diagnostics of the performance of the different investigated forecasting ensembles.

Another direct perspective that would serve the future development of forecasting tools for the Arzal reservoir is the identification of the sources of predictability in the Vilaine catchment. Knowing the sources of predictability can indeed serve as a local diagnostic to improving the skill of streamflow forecasts (Wood *et al.*, 2016). In practice, this could help guide future efforts, either towards better data acquisition tools to improve initial conditions, or towards skilful monthly or seasonal forecasts from climate services to improve meteorological forcings.

Unfortunately, we could not implement the conditioned streamflow forecasts based on historical streamflow in the risk assessment tool of the Arzal reservoir. Indeed, this method requires long streamflow records that are not yet available for the Vilaine at Rieux. Nevertheless, when streamflow records allow for it, it would be interesting to use the method to forecast reservoir inflows, since it provided a good assessment of drought and low-flow risk at a low computational cost. In a similar line, there is still work to be done to bring the risk assessment tool to an operational level. Its pre-operational version could be implemented and tested to evaluate the actual potential of the tool in supporting decision-making. Comparing in real-time its risk assessment to observed situations could help validate its performance, and, later on, we could assess if and how the tool can be used in the management of the multi-purpose Arzal reservoir.

The role-playing game presented in this thesis, as well as other game experiments (Ramos *et al.*, 2013; Arnal *et al.*, 2016), have shown that games can open the discussion on touchy subjects, by creating a theoretical and therefore safer context. Speaking of touchy subjects, and in the line of reservoir management, the allocation of the water resource for different uses could be a perfect subject for a future game including reservoir volume forecasts.

Further along the line, the impact of climate change on streamflows might dictate the value of this work for future years. Climate change is omnipresent when discussing low flows and droughts. Even though studies on the impact of climate change on low flows and droughts do not always converge to similar conclusions (Hisdal *et al.*, 2001; Lehner *et al.*, 2006; Vidal *et al.*, 2012; Corzo Perez *et al.*, 2011), there is a chance that the occurrence or the severity of droughts will increase in the coming century. In France, within the Explore 2070 project, Chauveau *et al.* (2013) showed that more severe summer low flows are to be expected in large catchments. Taking into account climate change implies that the hypothesis of stationarity is no longer valid. This would require to take into account the bias between past and future hydrometeorological observations, for example in the calibration of the hydrological model (Coron, 2013), in the calibration of bias corrections or in the selection of likely scenarios within past observations.

# References

Anderson, M.L., Mierzwa, M.D. and Kavvas, M.L. (2000). Probabilistic seasonal forecasts of droughts with a simplified coupled hydrologic-atmospheric model for water resources planning. *Stochastic Environmental Research and Risk Assessment* **14**, 263–274. 84, 146

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.H. and Valéry, A. (2009). HESS Opinions "Crash tests for a standardized evaluation of hydrological models". *Hydrology and Earth System Sciences* **13**, 1757–1764. 11

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J. and Berthet, L. (2012). All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes* **26**, 2206–2210. 25

Andréassian, V. and Perrin, C. (2012). On the ambiguous interpretation of the Turc-Budyko nondimensional graph. *Water Resources Research* **48**, W10601. 12

Arnal, L., Ramos, M.H., Coughlan, E., Cloke, H.L., Stephens, E., Wetterhall, F., van Andel, S.J. and Pappenberger, F. (2016). Willingness-to-pay for a probabilistic flood forecast: a risk-based decision-making game. *Hydrology and Earth System Sciences Discussions* **2016**, 1–38. 165

ASCE (1993). *Criteria for evaluation of watershed models*, vol. 119. American Society of Civil Engineers, Reston, VA, ETATS-UNIS. 25

Berthet, L. (2010). *Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique*. Phd thesis, Cemagref (Antony), AgroParisTech (Paris). 27

Bierkens, M.F.P. and van Beek, L.P.H. (2009). Seasonal Predictability of European Discharge: NAO and Hydrological Response Time. *Journal of Hydrometeorology* **10**, 953–968. 85

Block, P. (2011). Tailoring seasonal climate forecasts for hydropower operations. *Hydrology and Earth System Sciences* **15**, 1355–1368. 146

Block, P. and Goddard, L. (2012). Statistical and Dynamical Climate Predictions to Guide Water Resources in Ethiopia. *Journal of Water Resources Planning and Management* **138**, 287–298. 146

Block, P. and Rajagopalan, B. (2009). Statistical–Dynamical Approach for Streamflow Modeling at Malakal, Sudan, on the White Nile River. *Journal of Hydrologic Engineering* **14**, 185–196. 84

Blöschl, G. and Zehe, E. (2005). On hydrological predictability. *Hydrological Processes* **19**, 3923–3929. vii, 4

Bourgin, F. (2014). *Comment quantifier l'incertitude prédictive en modélisation hydrologique? Travail exploratoire sur un grand échantillon de bassins versants.* PhD thesis, AgroParisTech. 39

Bracken, C., Rajagopalan, B. and Prairie, J. (2010). A multisite seasonal ensemble streamflow forecasting technique. *Water Resources Research* **46**, W03532. 146

Bressers, H., Bressers, N., Kuks, S. and Larrue, C. (2016). Chapter 3: The Governance Assessment Tool and its Use. *Governance for Drought Resilience*, H. Bressers, N. Bressers, C. Larrue (Eds.), springer edn. 6

Brochet, P. (1977). La sécheresse 1976 en France : aspects climatologiques et conséquences / The 1976 drought in France: climatological aspects and consequences. *Hydrological Sciences Bulletin* **22**, 393–411. 3

Brown, C. and Ward, M.N. (2013). Chapter 3: Climate variability and hydrologic predictability. *Managing Climate Risk in Water Supply Systems*, pp. 27–39, IWA Publishing, London. 146

Buizza, R. and Palmer, T.N. (1998). Impact of Ensemble Size on Ensemble Prediction. *Monthly Weather Review* **126**, 2503–2518. 43

Buizza, R. and Leutbecher, M. (2015). The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society* **141**, 3366–3382. 44

Cannon-Bowers, J.A. and Bell, H.R. (1997). Training decision makers for complex environments: Implications of the naturalistic decision making perspective. *Naturalistic decision making* (eds. C. Zsambok & G. Klein), pp. 99–110, C. Zsambok and G. Klein, Eds., Lawrence Erlbaum Associates, Hillsdale, NJ: LEA. 5, 147

Carpenter, T.M. and Georgakakos, K.P. (2001). Assessment of Folsom lake response to historical and potential future climate scenarios: 1. Forecasting. *Journal of Hydrology* **249**, 148 – 175. 86, 162

Castelletti, A., Pianosi, F. and Soncini-Sessa, R. (2008). Water reservoir control under economic, social and environmental constraints. *Automatica* **44**, 1595–1607. vii, 5

Ceppi, A., Ravazzani, G., Corbari, C., Salerno, R., Meucci, S. and Mancini, M. (2014). Real-time drought forecasting system for irrigation management. *Hydrology and Earth System Sciences* **18**, 3353–3366. 84

Chauveau, M., Chazot, S., Perrin, C., Bourgin, P.Y., Sauquet, E., Vidal, J.P., Rouchy, N., Martin, E., David, J., Norotte, T., Maugis, P. and De Lacaze, X. (2013). Quels impacts des changements climatiques sur les eaux de surface en France à l'horizon 2070 ? *La Houille Blanche* pp. 5–15. 166

Chiew, F.H.S. and McMahon, T.A. (1993). Assessing the Adequacy of Catchment Streamflow Yield Estimates. *Australian Journal of Soil Research* **31**, 665–680. 37

Chiew, F.H.S., McMahon, T.A., Zhou, S.L. and Piechota, T. (2000). Streamflow Variability, Seasonal Forecasting and Water Resources Systems. *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems* (eds. G.L. Hammer, N. Nicholls & C. Mitchell), vol. 21 of *Atmospheric and Oceanographic Sciences Library*, pp. 409–428, Springer Netherlands. 146

Chiew, F.H.S., Zhou, S.L. and McMahon, T.A. (2003). Use of seasonal streamflow forecasts in water resources management. *Journal of Hydrology* **270**, 135–144. 146

Chow, V.T., Maidment, D. and Mays, L. (1988). *Applied Hydrology*. Civil Engineering, McGraw-Hill. 4, 25

Christensen, J.H., Boberg, F., Christensen, O.B. and Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters* **35**, L20709. 53

Coelho, C.A.S. and Costa, S.M.S. (2010). Challenges for integrating seasonal climate forecasts in user applications. *Current Opinion in Environmental Sustainability* **2**, 317–325. 146

Coron, L. (2013). *Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution ?* Phd thesis, AgroParisTech (Paris), Irstea (Antony). 166

Corti, T., Muccione, V., Köllner-Heck, P., Bresch, D. and Seneviratne, S.I. (2009). Simulating past droughts and associated building damages in France. *Hydrology and Earth System Sciences* **13**, 1739–1747. 3

Corzo Perez, G.A., van Lanen, H.A.J., Bertrand, N., Chen, C., Clark, D., Folwell, S., Gosling, S.N., Hanasaki, N., Heinke, J. and Voß, F. (2011). Drought at the Global Scale in the 21st Century. Tech. Rep. 43, WATCH. 166

Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S.P., Grimaldi, S., Gupta, H. and Paturel, J.E. (2015a). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal* **60**, 402–423. 37

Crochemore, L., Ramos, M.H., Pappenberger, F., van Andel, S.J. and Wood, A.W. (2015b). An experiment on risk-based decision-making in water management using monthly probabilistic forecasts. *Bulletin of the American Meteorological Society* . 6, 52

Day, G. (1985). Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management* **111**, 157–170. 52, 84

Demirel, M.C., Booij, M.J. and Hoekstra, A.Y. (2015). The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models. *Hydrology and Earth System Sciences* **19**, 275–291. 52

Di Giuseppe, F., Molteni, F. and Tompkins, A.M. (2013). A rainfall calibration methodology for impacts modelling based on spatial mapping. *Quarterly Journal of the Royal Meteorological Society* **139**, 1389–1401. 22

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., Barbosa, P., Vogt, J. and Pappenberger, F. (2014). Global meteorological drought – Part 2: Seasonal forecasts. *Hydrology and Earth System Sciences* **18**, 2669–2678. 52, 90

Dutta, D., Wilson, K., Welsh, W.D., Nicholls, D., Kim, S. and Cetin, L. (2013). A new river system modelling tool for sustainable operational management of water resources. *Journal of Environmental Management* **121**, 13–28. 146

Easey, J., Prudhomme, C. and Hannah, D.M. (2006). Seasonal forecasting of river flows: a review of the state-of-the-art. *Proceedings of the fifth FRIEND World Conference*, vol. 308, pp. 158–162, IAHS Publ., Havana, Cuba. 84

EDC (2013). European Drought Impact Report Inventory (EDII) and European Drought Reference (EDR) database. 3

EDO-JRC (2015). Drought News August 2015. EDO Drought News, European Commission - Joint Research Centre - European Drought Observatory. 3

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J. (2012). HESS Opinions "Should we apply bias correction to global and regional climate model data?". *Hydrology and Earth System Sciences* **16**, 3391–3404. 53

Faber, B.A. and Stedinger, J.R. (2001). Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. *Journal of Hydrology* **249**, 113–133. 52, 84, 147

FAO (2014). Towards Risk-Based Drought Management in Europe and Central Asia. Agenda Item AU816, Food and Agriculture Organization of the United Nations, Bucharest, Romania. 1, 2

Ferro, C.A.T., Richardson, D.S. and Weigel, A.P. (2008). On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications* **15**, 19–24. 43

Furusho, C., Vidaurre, R., La Jeunesse, I. and Ramos, M.H. (2016). Chapter 11: Cross-cutting perspective freshwater. *Governance for Drought Resilience*, H. Bressers, N. Bressers, C. Larrue (Eds.), springer edn. 6

Georgakakos, K.P. and Graham, N.E. (2008). Potential Benefits of Seasonal Inflow Prediction Uncertainty for Reservoir Release Decisions. *Journal of Applied Meteorology and Climatology* **47**, 1297–1321. vii, 5

Ghile, Y.B. and Schulze, R.E. (2008). Development of a framework for an integrated time-varying agrohydrological forecast system for Southern Africa : Initial results for seasonal forecasts. *Water SA* **34**, 315–322. 146

Gneiting, T., Balabdaoui, F. and Raftery, A.E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 243–268. xi, 39, 40, 87

Gneiting, T., Raftery, A.E., Westveld, A.H. and Goldman, T. (2005). Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review* **133**, 1098–1118. 79

Gobena, A. and Gan, T. (2010). Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology* **385**, 336–352. 85

Golembesky, K., Sankarasubramanian, A. and Devineni, N. (2009). Improved Drought Management of Falls Lake Reservoir: Role of Multimodel Streamflow Forecasts in Setting up Restrictions. *J Water Res Plan Manage* **135**, 188–197. 146

Gudmundsson, L., Bremnes, J.B., Haugen, J.E. and Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences* **16**, 3383–3390. 53

Gudmundsson, L., Van Loon, A.F., Tallaksen, L.M., Seneviratne, S.I., Stagge, J.H., Stahl, K. and van Lanen, H.A.J. (2014). Guidelines for monitoring and early warning of drought in Europe. DROUGHT-R&SPI Technical Report 21, DROUGHT-R&SPI Project. 4

Guo, W., Zhao, J. and Wang, F. (2009). The seasonal forecast method of Sanjing Plain underground water level. *Journal of Northeast Agricultural University* **5**, 104–107. 146

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377**, 80–91. x, 29, 54, 87

Hamill, T.M. and Juras, J. (2006). Measuring forecast skill: is it real skill or is it the varying climatology? *Quarterly Journal of the Royal Meteorological Society* **132**, 2905–2923. 66

Hamlet, A.F., Huppert, D. and Lettenmaier, D.P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management* **128**, 91–101. 85

Hamlet, A.F. and Lettenmaier, D.P. (1999). Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals. *Journal of Water Resources Planning and Management* **125**, 333–341. 85

Hao, Z., AghaKouchak, A., Nakhjiri, N. and Farahmand, A. (2014). Global integrated drought monitoring and prediction system. *Scientific Data* **1**, 140001. 84

Hartmann, H.C., Pagano, T.C., Sorooshian, S. and Bales, R. (2002). Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives. *Bulletin of the American Meteorological Society* **83**, 683–698. 52, 146

Hemri, S., Scheuerer, M., Pappenberger, F., Bogner, K. and Haiden, T. (2014). Trends in the predictive performance of raw ensemble weather forecasts. *Geophysical Research Letters* **41**, 9197–9205. 79

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**, 559–570. xi, 40

Hewitt, C., Buontempo, C. and Newton, P. (2013). Using Climate Predictions to Better Serve Society's Needs. *Eos, Transactions American Geophysical Union* **94**, 105–107. 4

Hisdal, H., Stahl, K., Tallaksen, L.M. and Demuth, S. (2001). Have streamflow droughts in Europe become more severe or frequent? *International Journal of Climatology* **21**, 317–333. 166

Ionita, M., Boroneant, C. and Chelcea, S. (2015). Seasonal modes of dryness and wetness variability over Europe and their connections with large scale atmospheric circulation and global sea surface temperature. *Climate Dynamics* **45**, 2803–2829. 85

Jachner, S., van den Boogaart, K.G. and Petzoldt, T. (2007). Statistical methods for the qualitative assessment of dynamic models with time delay (R package qualV). *Journal of Statistical Software* **22**, 1–30. 38

Jolliffe, I.T. and Stephenson, D.B. (2003). *Forecast Verification: A Practicioner's Guide in Atmospheric Science*. John Wiley. 37

Kelman, J., Stedinger, J.R., Cooper, L.A., Hsu, E. and Yuan, S.Q. (1990). Sampling stochastic dynamic programming applied to reservoir operation. *Water Resources Research* **26**, 447–454. 5

Kiem, A.S. and Verdon-Kidd, D.C. (2011). Steps toward "useful" hydroclimatic scenarios for water resource management in the Murray-Darling Basin. *Water Resources Research* **47**, W00G06. 146

Kim, H.M., Webster, P.J. and Curry, J.A. (2012). Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dynamics* **39**, 2957–2973. 22

Kirchhoff, C.J., Lemos, M.C. and Engle, N.L. (2013). What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S. *Environmental Science & Policy* **26**, 6–18. 146

Kirtman, B., Power, S., Adedoyin, J., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G., Vecchi, G. and Wang, H. (2013). Near-term Climate Change: Projections and Predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. T. Stocker, D. Qin, G.K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. Midgley), pp. 953–1028, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. vii, 4

Klemes, V. (1986). Operational Testing of Hydrological Simulation-Models. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* **31**, 13–24. 25

Kollipara, P. (2015). When predicting drought risk, do not overlook temperature. *Eos* . 164

Krause, P., Boyle, D.P. and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**, 89–97. 37, 38

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology* **249**, 2 – 9. 26, 36, 84

La Jeunesse, I., Larrue, C., Furusho, C., Ramos, M.H., Browne, A., de Boer, C., Vidaurre, R., Crochemore, L., Penasso, A. and Arrondeau, J.P. (2016). Chapter 6: The governance context of drought policy and pilot measures in the Arzal dam and reservoir, Vilaine catchment, Brittany, France. *Governance for Drought Resilience*, H. Bressers, N. Bressers, C. Larrue (Eds.), springer edn. 3, 6

Laio, F. and Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences* **11**, 1267–1277. xi, 39

Le Moine, N. (2008). *Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?* Phd thesis, ENGREF (Paris), Cemagref (Antony). 27

Lehner, B., Döll, P., Alcamo, J., Henrichs, T. and Kaspar, F. (2006). Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis. *Climatic Change* **75**, 273–299. 166

Lemos, M.C. (2008). What Influences Innovation Adoption by Water Managers? Climate Information Use in Brazil and the United States. *JAWRA Journal of the American Water Resources Association* **44**, 1388–1396. 146

Lemos, M., Finan, T., Fox, R., Nelson, D. and Tucker, J. (2002). The Use of Seasonal Climate Forecasting in Policymaking: Lessons from Northeast Brazil. *Climatic Change* **55**, 479–507. 52, 146

Liu, Y., Duan, Q., Zhao, L., Ye, A., Tao, Y., Miao, C., Mu, X. and Schaake, J.C. (2013). Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin. *Hydrological Processes* **27**, 57–74. 79

Lorenz, E. (1984). Some Aspects of Atmospheric Predictability. *Problems and Prospects in Long and Medium Range Weather Forecasting* (eds. D. Burridge & E. Källén), Topics in Atmospheric and Oceanographic Sciences, pp. 1–20, Springer Berlin Heidelberg. vii, 4

Lowry, R. (1999). *Concepts and Applications of Inferential Statistics*. Vassar College. 99

Madadgar, S., Moradkhani, H. and Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes* **28**, 104–122. 53

Mason, S.J. and Graham, N.E. (1999). Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels. *Weather and Forecasting* **14**, 713–725. xi, 40

Mathevet, T., Michel, C., Andréassian, V. and Perrin, C. (2006). A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS Red Books Series*, vol. 307, pp. 211–219. 30

McKee, T., Doeskin, N. and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. pp. 179–184. 88

Mishra, A.K. and Singh, V.P. (2010). A review of drought concepts. *Journal of Hydrology* **391**, 202–216. 4, 99

Mishra, A.K. and Singh, V.P. (2011). Drought modeling – A review. *Journal of Hydrology* **403**, 157–175. 4

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. and Vitart, F. (2011). The new ECMWF seasonal forecast system (System 4). *ECMWF Tech. Memo.* **656**, 49 pp. ix, 20

Muerth, M.J., Gauvin St-Denis, B., Ricard, S., Velázquez, J.A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R. and Turcotte, R. (2013). On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. *Hydrology and Earth System Sciences* **17**, 1189–1204. 53

Mushtaq, S., Chen, C., Hafeez, M., Maroulis, J. and Gabriel, H. (2012). The economic value of improved agrometeorological information to irrigators amid climate variability. *International Journal of Climatology* **32**, 567–581. 146

Mwangi, E., Wetterhall, F., Dutra, E., Di Giuseppe, F. and Pappenberger, F. (2014). Forecasting droughts in East Africa. *Hydrology and Earth System Sciences* **18**, 611–620. 52

Najafi, M.R., Moradkhani, H. and Piechota, T.C. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *Journal of Hydrology* **442–443**, 105–116. 85

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J.M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B. and Morice, E. (2014). Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. *Hydrol. Earth Syst. Sci.* **18**, 2829–2857. xi, 4, 11, 44, 52

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. and Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model ? Part 2 — Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology* **303**, 290–306. ix, 20

PAGD (2015). Plan d'aménagement et de gestion durable. Tech. rep., Commission Locale de l'Eau. 118

Palmer, R. and Holmes, K. (1988). Operational Guidance During Droughts: Expert System Approach. *Journal of Water Resources Planning and Management* **114**, 647–666. 1

Pappenberger, F., Ramos, M.H., Cloke, H.L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A. and Salamon, P. (2015). How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology* **522**, 697 – 713. 38, 42

Perrin, C., Andreassian, V. and Michel, C. (2006). *Simple benchmark models as a basis for model efficiency criteria*. Schweizerbart, Stuttgart, ALLEMAGNE. 38

Perrin, C., Michel, C. and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* **279**, 275–289. 27

Pirard, P., Vandentorren, S. and Pascal, M. (2005). Summary of the mortality impact assessment of the 2003 heat wave in France. *Eurosurveillance: European Communicable Disease Journal* **10**. 3

Poumadère, M., Mays, C., Le Mer, S. and Blong, R. (2005). The 2003 Heat Wave in France: Dangerous Climate Change Here and Now. *Risk Analysis* **25**, 1483–1494. 87

Pushpalatha, R., Perrin, C., Mathevet, T. and Andreassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology* **411**, 66–76. ix, 27, 28, 54, 87

Pushpalatha, R., Perrin, C., Moine, N.L. and Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology* **420–421**, 171–182. 30, 38

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. and Morel, S. (2008). Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. *Journal of Applied Meteorology and Climatology* **47**, 92–107. ix, 20

Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review* **133**, 1155–1174. 79

Ramos, M.H., van Andel, S.J. and Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences* **17**, 2219–2232. 26, 84, 147, 165

Ramos, M.H., Mathevet, T., Thielen, J. and Pappenberger, F. (2010). Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications* **17**, 223–235. 26

Rayner, S., Lach, D. and Ingram, H. (2005). Weather Forecasts are for Wimps: Why Water Resource Managers Do Not Use Climate Forecasts. *Climatic Change* **69**, 197–227. 52

Refsgaard, J.C. and Henriksen, H.J. (2004). Modelling guidelines - terminology and guiding principles. *Advances in Water Resources* **27**, 71–82. 25

Regonda, S., Zagona, E. and Rajagopalan, B. (2011). Prototype Decision Support System for Operations on the Gunnison Basin with Improved Forecasts. *Journal of Water Resources Planning and Management* **137**, 428–438. vii, 5, 146

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S.W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research* **46**, W05521. xi, 39

Ritchie, J.W., Zammit, C. and Beal, D. (2004). Can seasonal climate forecasting assist in catchment water management decision-making?: A case study of the Border Rivers catchment in Australia. *Agriculture, Ecosystems & Environment* **104**, 553–565. 146

Robertson, A., Kumar, A., Peña, M. and Vitart, F. (2015). Improving and Promoting Subseasonal to Seasonal Prediction. *Bull. of the Amer. Meteor. Soc.* **96**, ES49–ES53. 4

Robertson, D.E., Pokhrel, P. and Wang, Q.J. (2013). Improving statistical forecasts of seasonal streamflows using hydrological model output. *Hydrology and Earth System Sciences* **17**, 579–593. 52

Robine, J.M., Cheung, S.L., Le Roy, S., van Oyen, H. and Hermann, F.R. (2007). Report on excess mortality in Europe during summer 2003. Tech. rep., EU Community Action Programme for Public Health. 3

Roulin, E. and Vannitsem, S. (2015). Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. *Hydrological Processes* **29**, 1434–1449. 53, 72

Rykiel Jr., E.J. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling* **90**, 229–244. 25

SAGE (2015a). Atlas - Révision du Schéma d'Aménagement et de Gestion des Eaux de la Vilaine. Tech. rep., Commission Locale de l'Eau. 15, 112

SAGE (2015b). Synthèse et état des lieux - Révision du Schéma d'Aménagement et de Gestion des Eaux de la Vilaine. Tech. rep., Commission Locale de l'Eau. 15

Sankarasubramanian, A., Lall, U., Souza Filho, F.A. and Sharma, A. (2009). Improved water allocation utilizing probabilistic climate forecasts: Short-term water contracts in a risk management framework. *Water Resources Research* **45**, W11409. 147

Sauquet, E., Lerat, J. and Prudhomme, C. (2008). La prévision hydro-météorologique à 3-6 mois. Etat des connaissances et applications. *La Houille Blanche* pp. 77–84. 85

Seibert, M. and Trambauer, P. (2015). Seasonal forecasts of hydrological drought in the Limpopo basin: Getting the most out of a bouquet of methods. *Drought: Research and Science-Policy Interfacing*, pp. 307–313, CRC Press. 84

Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S. and Ogallo, L. (2013). A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security. *Bulletin of the American Meteorological Society* **95**, 861–882. 84

Shukla, S., McNally, A., Husak, G. and Funk, C. (2014). A seasonal agricultural drought forecast system for food-insecure regions of East Africa. *Hydrology and Earth System Sciences* **18**, 3907–3921. 84

Shukla, S., Sheffield, J., Wood, E.F. and Lettenmaier, D.P. (2013). On the sources of global land surface hydrologic predictability. *Hydrology and Earth System Sciences* **17**, 2781–2796. 4, 26, 52, 84

Shukla, S., Voisin, N. and Lettenmaier, D.P. (2012). Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill. *Hydrology and Earth System Sciences* **16**, 2825–2838. 106

Simonović, S.P. and Bender, M.J. (1996). Collaborative planning-support system: an approach for determining evaluation criteria. *Journal of Hydrology* **177**, 237–251. 1

Simonović, S.P. and Marino, M.A. (1982). Reliability programing in reservoir management: 3. System of multipurpose reservoirs. *Water Resources Research* **18**, 735–743. 146

Singla, S., Céron, J.P., Martin, E., Regimbeau, F., Déqué, M., Habets, F. and Vidal, J.P. (2012). Predictability of soil moisture and river flows over France for the spring season. *Hydrology and Earth System Sciences* **16**, 201–216. 26

Šípek, V. and Daňhelka, J. (2015). Modification of input datasets for the Ensemble Streamflow Prediction based on large-scale climatic indices and weather generator. *Journal of Hydrology* **528**, 720 – 733. 85

Smakhtin, V.U. (2001). Low flow hydrology: a review. *Journal of Hydrology* **240**, 147 – 186. 99

Stocker, T., Clarke, G., Le Treut, H., Lindzen, R., Meleshko, V., Mugara, R., Palmer, T., Pierrehumbert, R., Sellers, P., Trenberth, K. and others (2001). Physical Climate Processes and Feedbacks. *In: IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change/Houghton, JT, Y. Ding, DJ Griggs, M. Noguer, PJ van der Linden, X. Dai, K. Maskell, CA Johnson (eds.).-Cambridge and New York: Cambridge University Press, 2001.-ISBN 0521 01495 6* . 4

Svensson, C. (2016). Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. *Hydrological Sciences Journal* **61**, 19–35. 85

Tallaksen, L.M., Madsen, H. and Clausen, B. (1997). On the definition and modelling of streamflow drought duration and deficit volume. *Hydrological Sciences Journal* **42**, 15–33. 99

Tangara, M. (2005). *Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global*. Phd thesis, EPHE (Paris), Cemagref (Antony). 27

Teutschbein, C. and Seibert, J. (2013). Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrology and Earth System Sciences* **17**, 5061–5077. 53

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology* **456-457**, 12–29. 53

Trambauer, P., Werner, M., Winsemius, H.C., Maskey, S., Dutra, E. and Uhlenbrook, S. (2015). Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa. *Hydrology and Earth System Sciences* **19**, 1695–1711. 22, 71, 85

Turner, B.L., Kasperson, R.E., Matson, P.A., McCarthy, J.J., Corell, R.W., Christensen, L., Eckley, N., Kasperson, J.X., Luers, A., Martello, M.L., Polsky, C., Pulsipher, A. and Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences* **100**, 8074–8079. 1

UNEP (2004). Impacts of summer 2003 heat wave in Europe. Environment Alert Bulletin 2, United Nations Environment Programme, Nairobi. 3, 87

UNISDR (2009). UNISDR Terminology on Disaster Risk Reduction. Glossary, UN International Srategy for Disaster Reduction, Geneva. 1

van Dijk, A.I.J.M., Peña-Arancibia, J.L., Wood, E.F., Sheffield, J. and Beck, H.E. (2013). Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resources Research* **49**, 2729–2746. 85

Verkade, J.S., Brown, J.D., Reggiani, P. and Weerts, A.H. (2013). Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology* **501**, 73–91. 53, 72

Vidal, J.P., Martin, E., Kitova, N., Najac, J. and Soubeyroux, J.M. (2012). Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios. *Hydrol. Earth Syst. Sci. Discuss.* **9**, 1619–1670. 166

Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M. and Soubeyroux, J.M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology* **30**, 1627–1644. ix, 20

Wang, E., Zhang, Y., Luo, J., Chiew, F. and Wang, Q. (2011). Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. *Water Resources Research* **47**. 84

Weerts, A.H., Winsemius, H.C. and Verkade, J.S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences* **15**, 255–265. 53

Weisheimer, A. and Palmer, T.N. (2014). On the reliability of seasonal climate forecasts. *Journal of The Royal Society Interface* **11**, 20131162. 22

Welsh, W.D., Vaze, J., Dutta, D., Rassam, D., Rahman, J.M., Jolly, I.D., Wallbrink, P., Podger, G.M., Bethune, M., Hardy, M.J., Teng, J. and Lerat, J. (2013). An integrated modelling framework for regulated river systems. *Environmental Modelling & Software* **39**, 81–102. 146

Werner, K., Brandon, D., Clark, M. and Gangopadhyay, S. (2005). Incorporating medium-range numerical weather model output into the Ensemble Streamflow Prediction system of the National Weather Service. *Journal of Hydrometeorology* **6**, 101–114. 106

Werner, K., Brandon, D., Clark, M. and Gangopadhyay, S. (2004). Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. *Journal of Hydrometeorology* **5**, 1076–1090. 85

Wetterhall, F., Winsemius, H.C., Dutra, E., Werner, M. and Pappenberger, E. (2015). Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin. *Hydrology and Earth System Sciences* **19**, 2577–2586. 22, 52, 71

Wilhite, D.A. and Glantz, M.H. (1985). Understanding: the Drought Phenomenon: The Role of Definitions. *Water International* **10**, 111–120. 2

Wilhite, D.A., Hayes, M.J., Knutson, C. and Smith, K.H. (2000). Planning for drought: Moving from crisis to risk management. *JAWRA Journal of the American Water Resources Association* **36**, 697–710. 1, 2, 52, 84

Winsemius, H.C., Dutra, E., Engelbrecht, F.A., Archer Van Garderen, E., Wetterhall, F., Pappenberger, F. and Werner, M.G.F. (2014). The potential value of seasonal forecasts in a changing climate in southern Africa. *Hydrology and Earth System Sciences* **18**, 1525–1538. 52

WMO (2008). *Manual on Low-flow Estimation and Prediction*. World Meteorological Organization. 2, 17, 99

WMO (2012). *Standardized Precipitation Index User Guide*. World Meteorological Organization. 89

Wood, A.W., Hopson, T., Newman, A., Brekke, L., Arnold, J. and Clark, M. (2016). Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. *Journal of Hydrometeorology* **17**, 651–668. 4, 26, 165

Wood, A.W., Kumar, A. and Lettenmaier, D.P. (2005). A retrospective assessment of National Centers for Environmental Prediction climate model–based ensemble hydrologic forecasting in the western United States. *Journal of Geophysical Research: Atmospheres* **110**, D04105. 52

Wood, A.W. and Lettenmaier, D.P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters* **35**, L14401. 52, 84

Wood, A.W. and Schaake, J.C. (2008). Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. *Journal of Hydrometeorology* **9**, 132–148. 53

Wood, E.F., Schubert, S.D., Wood, A.W., Peters-Lidard, C.D., Mo, K.C., Mariotti, A. and Pulwarty, R.S. (2015). Prospects for Advancing Drought Understanding, Monitoring and Prediction. *Journal of Hydrometeorology* . 4

Yao, H. and Georgakakos, A. (2001). Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management. *Journal of Hydrology* **249**, 176–196. 86

Yeh, W.W.G. (1985). Reservoir Management and Operations Models: A State-of-the-Art Review. *Water Resources Research* **21**, 1797–1818. 5

Yossef, N.C., Winsemius, H., Weerts, A., van Beek, R. and Bierkens, M.F.P. (2013). Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resources Research* **49**, 4687–4699. 4, 26, 52, 84

Yuan, X., Wood, E.F. and Ma, Z. (2015). A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *Wiley Interdisciplinary Reviews: Water* pp. 523–536. 4, 52, 84, 85

Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T. and Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research* **8**, 135–141. 53, 72