



HAL
open science

Contributions statistiques aux prévisions hydrométéorologiques par méthodes d'ensemble

Marie Courbariaux

► **To cite this version:**

Marie Courbariaux. Contributions statistiques aux prévisions hydrométéorologiques par méthodes d'ensemble. Applications [stat.AP]. Université Paris Saclay (COMUE), 2017. Français. NNT : 2017SACLA003 . tel-01566195

HAL Id: tel-01566195

<https://theses.hal.science/tel-01566195>

Submitted on 20 Jul 2017

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

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

NNT : 2017SACLA003

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À AGROPARISTECH

ÉCOLE DOCTORALE N°581 :
Agriculture, alimentation, biologie, environnement et santé (ABIES)

Spécialité de doctorat : Statistiques appliquées

par

Marie COURBARIAUX

Contributions statistiques aux prévisions hydrométéorologiques
par méthodes d'ensemble

Thèse soutenue le 27 janvier 2017

Composition du jury :

Jean-Michel POGGI	Professeur (Université Paris Descartes)	Président du jury, Rapporteur
François ANCTIL	Professeur (Université Laval)	Rapporteur
Maria-Helena RAMOS	Chargée de recherche (IRSTEA)	Examinatrice
Olivier MESTRE	Chargé de recherche (Météo-France)	Examinateur
Éric PARENT	IGPEF (Professeur à AgroParisTech)	Directeur de thèse
Anne-Catherine FAVRE	Professeur (G-INP/LTHE/ENSE ³)	Co-directrice de thèse
Pierre BARBILLON	Maître de conférences (AgroParisTech)	Co-encadrant de thèse
Luc PERREAULT	Chercheur (IREQ)	Co-encadrant de thèse, invité
Rémy GARÇON	Expert (EDF-DTG)	Invité

Ce document a été rédigé sous LaTeX (logiciel libre, LaTeX *Project Public License*).
Les calculs et graphiques ont été réalisés sous R (logiciel libre, *GNU General Public License*).

Le template LaTeX a été réalisé par Loïc Schwaller.

UMR MIA-Paris AgroParisTech/INRA
Université Paris-Saclay
16, rue Claude Bernard
75005 Paris

Remerciements

Ce manuscrit ne pourrait être abouti sans ces remerciements aux personnes qui ont contribué à ce projet de doctorat.

Mes remerciements vont tout d’abord à l’équipe du projet, qui l’a porté jusque dans ses derniers jours, au fil des conférences, des rapports et articles, des réunions franco-québécoises, sur Skype, depuis Paris, Montréal ou Grenoble. C’était une joie de travailler avec des personnes aussi formidables au cours de ces trois années !

Merci en particulier à mon directeur de thèse, Éric Parent, grâce à qui j’ai beaucoup appris. Merci pour ta bienveillance et ton implication, considérables. Avec toi, les divergences de points de vue se muiaient en débats passionnés que j’espère nous continuerons à avoir.

Merci à Pierre Barbillon, vers lequel mes pas se dirigeaient à chaque interrogation avec l’assurance de toujours y trouver une réponse et des conseils avisés.

Merci à Anne-Catherine Favre, en particulier pour l’enthousiasme dont elle a entouré ce projet.

Merci à Joël Gailhard et à Luc Perreault, qui ont respectivement porté ce projet au sein d’EDF et d’Hydro-Québec et n’ont pas hésité à mettre la main à la pâte et à impliquer leurs équipes, notamment pour concevoir des jeux de données sur mesure et pour identifier les points pouvant faire l’objet de critiques dans nos travaux. J’en profite pour adresser un grand merci aux équipes d’EDF-DTG, d’Hydro-Québec et de l’IREQ qui m’ont accueillie au cours de chacun de mes séjours. Merci en particulier à ceux qui ont travaillé sur ma thèse, notamment Rémy Garçon, Fabien Rinaldi, Matthieu Le-Lay, Federico Garavaglia, et David Penot côté EDF-DTG, Catherine Guay, Isabelle Chartier, Jocelyn Gaudet et Marie Minville côté IREQ, Fabian Tito Arandia Martinez qui a conçu les jeux de données TIGGE du projet, Éric Crobeddu ou encore Marie-Claude Simard côté Hydro-Québec Production. Merci aussi à ceux qui ont pris de leur temps pour m’expliquer le fonctionnement des systèmes de prévision d’EDF et d’Hydro-Québec.

Merci aux personnes qui m’ont conseillée tout au long du projet, en conférence ou lors des comités de thèse. Je pense en particulier à Vincent Fortin, dont les idées ont eu beaucoup d’influence sur mon travail et m’ont plus d’une fois fait douter, à Rémy Garçon et à Maria-Helena Ramos, ainsi qu’à l’équipe “Hydrologie des bassins

versants” de l’IRSTEA d’Antony.

Merci à EDF et à Hydro-Québec, partenaires et financeurs de ma thèse.

Merci aux rapporteurs de ma thèse, Jean-Michel Poggi et François Anctil, et aux examinateurs de ma thèse, Maria-Helena Ramos et Olivier Mestre. Merci à tous les quatre d’avoir accepté de vous pencher sur mon travail et de m’avoir aidée, par vos rapports et/ou vos questions, à le mettre en perspective.

Merci à chacun des membres l’unité Mathématiques et Informatique Appliquées d’AgroParisTech. Grâce à vous, venir travailler était un bonheur. Il va m’être très difficile de résister à la tentation de repasser vous voir régulièrement, que ce soit pour les quizzes, pour les Vieux Chêne/Margen’s/... et autres moments riches en jeux de mots et en moqueries, ou plus simplement pour revoir des personnes que j’apprécie énormément.

Un pensée particulière pour le BDDPG. Loïc Schwaller Schwaller et Anna Boney M, cette thèse vous doit énormément, à commencer par un template LaTeX (merci Loïc !), mais aussi et surtout un soutien en toute circonstance et un milieu de travail des plus heureux, dont quelques éléments du bureau peuvent témoigner. Doctorant PiGlo, je tiens à te faire part de mon admiration pour tes qualités de chef d’orchestre¹. Timothée et Mathieu, je sais qu’entre vos mains le BDDPG prendra la bonne voile. Aux Orcéens : c’était une joie de vous trouver de temps en temps dans le (/à mon) bureau. Merci à nos surveillants et complices du bureau voisin ainsi qu’aux squatteurs qui animaient la pause thé. Merci aussi aux précédents occupants de ce qui fut le BDDDH : à ceux qui m’y ont accueillie, au créateur d’Hervé.

Merci à ceux avec qui j’ai partagé des moments heureux loin du bureau. Je pense à des buveurs de grenadine K-rément géniaux, à des gens tip et top, à un chat plein de sagesse, à des joueurs d’ultimate d’un club versaillais, à des airsoft gamers, et à bien d’autres personnes sans qui la vie serait moins belle.

Merci à mes parents, dont le dévouement et la confiance ont plus d’une fois recoloré ma vie en rose. Merci aussi à ma fratrie, dont le plus gros défaut est d’avoir traversé l’Atlantique.

Mes remerciements vont enfin à mon compagnon, qui a vécu cette thèse presque aussi intensément que moi, me redonnant confiance dans les moments les plus sombres, partageant ma joie dans les moments les plus lumineux. Un immense merci !

1. c’est à la fin du bal...

Table des matières

0	Introduction	1
0.1	Enjeux des prévisions hydrologiques à moyen-terme pour les producteurs d'hydro-électricité	2
0.2	Plus-value des prévisions assorties d'un ordre de grandeur de l'incertitude qui leur est associée	3
0.3	Les systèmes de prévision hydrologique actuels	4
0.4	Aide statistique à la représentation des incertitudes	7
0.5	Organisation du mémoire	8
1	Ressources pour les prévisions hydrométéorologiques et critères d'évaluation	11
1.1	Notations	12
1.2	Sources d'informations pour les prévisions hydrologiques à moyen-terme	12
1.3	Critères d'évaluation des prévisions hydrométéorologiques probabilistes	27
1.4	Définition avancée des objectifs du travail doctoral	35
2	État de l'art des systèmes de prévision hydrométéorologique	37
2.1	Prédominance d'une architecture séparant les sources d'incertitudes météorologique et hydrologique	39
2.2	Méthodes statistiques pour le post-traitement des prévisions	41
2.3	Vue d'ensemble des méthodes de post-traitement météorologique	47
2.4	Vue d'ensemble des méthodes de prise en compte de l'incertitude hydrologique	55
3	Un post-traitement des prévisions météorologiques reposant sur une hypothèse d'échangeabilité	65
3.1	Un modèle à variables latentes pour le post-traitement des prévisions de températures	67
3.2	Extension du modèle à variables latentes pour de post-traitement des prévisions de précipitations	84
3.3	Post-traitement des prévisions de précipitations avec la loi des fuites	97
3.4	Discussion et perspectives	99

4	Un post-traitement des prévisions hydrologiques tirant parti d'informations auxiliaires	101
4.1	Résumé de l'article soumis pour publication	102
4.2	Article soumis pour publication	105
5	Conclusion, discussion et perspectives	139
5.1	Contributions au problème météorologique et au problème hydrologique	140
5.2	Résultats du système complet proposé	142
5.3	Confrontation entre idéal statistique et réalité opérationnelle	146

O

Introduction

0.1	Enjeux des prévisions hydrologiques à moyen-terme pour les producteurs d'hydro-électricité	2
0.2	Plus-value des prévisions assorties d'un ordre de grandeur de l'incertitude qui leur est associée	3
0.3	Les systèmes de prévision hydrologique actuels	4
0.4	Aide statistique à la représentation des incertitudes	7
0.5	Organisation du mémoire	8

L'origine de ce projet de doctorat s'ancre dans la volonté de deux producteurs d'hydro-électricité, Électricité de France (EDF) et Hydro-Québec, de collaborer afin de (mieux) intégrer les prévisions météorologiques d'ensemble dans leurs systèmes de prévision hydrologique. Ce travail doctoral s'inscrit notamment dans le projet P66.01 de l'Institut de Recherche d'Hydro-Québec (IREQ) intitulé "*Prévisions hydrologiques probabilistes*" (Perreault, 2014, 2015) visant à refondre le système de prévision hydrologique d'Hydro-Québec notamment en l'adaptant pour l'utilisation des prévisions météorologiques d'ensemble.

Plutôt qu'un seul scénario (*prévision déterministe*), les services météorologiques délivrent en effet aujourd'hui, à l'usage des professionnels, un *ensemble* de scénarios, les *membres*, dont le faisceau illustre l'incertitude portant sur le futur à prévoir. Les hydro-électriciens utilisent cet ensemble de scénarios en entrée de leurs systèmes de prévision hydrologique à *moyen-terme*¹. Ces *systèmes de prévision hydrologique* visent à prévoir le débit à l'exutoire de bassins versants d'intérêt. Les prévisions

1. à un horizon de prévision de quelques jours à quelques semaines dans ce contexte.

qui en sont issues (dites *hydrologiques*) sont utilisées pour la gestion des ouvrages hydrauliques en avenir incertain.

Le raisonnement statistique devrait être inhérent à la conception et à la mise en opération de ces chaînes de prévision : le calcul des probabilités constitue l'outil approprié pour la prise en compte des incertitudes de façon cohérente, pour quantifier celles portant sur la météorologie future par exemple.

Dans ce chapitre introductif, nous expliquons le contexte de ce travail doctoral et dessinons ses objectifs. Nous commençons par présenter comment les prévisions hydrologiques à moyen-terme sont utilisées chez les producteurs d'hydro-électricité afin d'en cerner les enjeux (sections 0.1 et 0.2). Une brève description des systèmes de prévision actuels et notamment de leur façon d'intégrer les prévisions météorologiques nous permet ensuite de mettre en lumière les apports potentiels du raisonnement statistique pour les prévisions hydrologiques à moyen-terme (sections 0.3 et 0.4). Nous expliquons enfin l'organisation du présent mémoire (section 0.5).

0.1 Enjeux des prévisions hydrologiques à moyen-terme pour les producteurs d'hydro-électricité

ANTICIPATION DES CRUES ET DES ÉTIAGES Les prévisions hydrologiques à moyen-terme sont d'une grande importance pour l'anticipation des crues et des étiages (basses eaux). Elles sont utilisées à cette fin tant par les organismes de protection de la population contre les inondations (tels que les services qui interviennent dans le réseau VIGICRUES² en France) que par les gestionnaires d'ouvrages hydrauliques (ou en bord de cours d'eau) que sont EDF³ et Hydro-Québec. Chez EDF, passer au-dessous d'un seuil d'étiage peut, par exemple, impliquer un changement dans la politique de gestion des effluents des centrales thermiques riveraines (dont le refroidissement nécessite de l'eau).

GESTION DES OUVRAGES HYDRO-ÉLECTRIQUES ET OPTIMISATION DE LA PRODUCTION ÉLECTRIQUE Chez EDF et chez Hydro-Québec, à ce premier enjeu lié à l'anticipation des crues et des étiages s'ajoutent les enjeux de gestion des ouvrages hydro-électriques : les prévisions hydrologiques sont les principaux intrants des outils d'optimisation de la production.

Cet enjeu est particulièrement important pour Hydro-Québec, dont 99% de l'énergie électrique produite est d'origine hydro-électrique. Chez Hydro-Québec, l'objectif de gestion est plus précisément de maximiser le rendement de production tout en respectant la stratégie de production à plus long-terme (qui comprend, par exemple, le retrait périodique de groupes de production pour maintenance ou les engagements contractuels à long-terme de fourniture d'électricité aux États-Unis d'Amérique), en

2. <http://www.developpement-durable.gouv.fr/Le-service-central-d.html>

3. En 2014, EDF surveillait environ 130 seuils de crue ou d'étiage.

minimisant les variations de débit infra-journalières et en respectant des cotes minimales et maximales aux réservoirs. Ces dernières contraintes de cotes maximales ou minimales, dont certaines sont imposées par la réglementation en vigueur, visent par exemple à éviter tout risque d'inondation, à garantir le niveau d'eau nécessaire aux activités de loisir sur les plans d'eau ou à garantir les débits réservés permettant d'assurer en permanence le maintien des habitats des poissons.

En France, seuls 12% de l'énergie électrique produite était d'origine hydraulique en 2015⁴, mais la production des centrales hydro-électriques constitue, en association avec les centrales thermiques à flamme, le moyen de production privilégié pour ajuster en permanence la production électrique à la consommation, du fait de la souplesse d'utilisation d'une grande partie des aménagements.

En effet, la production des centrales hydro-électriques peut être contrôlée finement et dans des délais courts, contrairement à celle des centrales nucléaires, qui représentait 76% de l'énergie électrique produite en 2015, et à celle d'origine solaire ou éolienne, qui représentait 4% de l'énergie électrique produite en 2015. L'énergie électrique d'origine solaire ou éolienne est, de plus, intermittente et son abondance n'est pas contrôlable et difficilement prévisible sans pour autant que l'on dispose de moyens efficaces de stocker cette énergie.

Les prévisions à moyen-terme donnent enfin la possibilité aux hydro-électriciens d'optimiser la programmation des opérations de maintenance de leurs groupes de production (en prévoyant ces opérations lorsque la sollicitation des aménagements est probablement inutile), dans les limites fixées par les contraintes de ces interventions

0.2 Plus-value des prévisions assorties d'un ordre de grandeur de l'incertitude qui leur est associée

Comme le constate Gneiting (2008), le passage de prévisions déterministes à des *prévisions probabilistes*, c'est-à-dire sous la forme de distributions de probabilités de la quantité à prévoir, d'abord limité à la prévision d'événements binaires, concerne à présent d'autres catégories d'événements plus complexes, et ce dans de nombreux domaines tels que l'économie, la finance, l'épidémiologie et l'hydrométéorologie dont nous traitons ici. Le calcul des probabilités constitue l'outil approprié pour la prise en compte des incertitudes de façon cohérente (Dawid, 1984).

La représentation et la quantification des incertitudes portant sur l'hydrologie future sont d'une grande aide tant pour l'anticipation des crues et des étiages que pour la gestion des ouvrages hydro-électriques. Cette observation est à l'origine de la communauté *Hydrological Ensemble Prediction Experiment* (HEPEX)⁵ (Schaake et al., 2007). Cette communauté réunit des producteurs et des utilisateurs des prévisions

4. http://www.lemonde.fr/les-decodeurs/article/2016/04/17/transition-energetique-ou-en-est-la-france-en-matiere-d-energies-renouvelables_4903791_4355770.html .

5. <https://hepex.irstea.fr/about-hepex/>.

hydrologiques ainsi que des scientifiques dans le but de promouvoir l'utilisation et d'aider à la production des *prévisions hydrologiques d'ensemble*, c'est-à-dire de prévisions hydrologiques issues de l'utilisation des prévisions météorologiques d'ensemble (et donc assorties d'une quantification de l'incertitude qui leur est associée).

La valeur ajoutée des prévisions hydrologiques probabilistes par rapport à leur pendant déterministe⁶ a été montrée (notamment en matière de valeur économique) dans le cas de l'anticipation et de la gestion des inondations (Richardson, 2000; Buizza, 2008; Verkade & Werner, 2011; Pappenberger et al., 2015) et dans le cas de la gestion des retenues hydro-électriques (Zalachori, 2013).

EDF et Hydro-Québec ont conscience de cette valeur ajoutée depuis déjà plusieurs années. Chez EDF-DTG⁷, les prévisionnistes sont formés depuis plus de 10 ans au langage des probabilités et à la communication des incertitudes sous la forme de distributions de probabilités (Houdant, 2004).

Les prévisionnistes d'EDF-DTG produisent d'ailleurs à ce jour des prévisions probabilistes de dépassement des seuils surveillés et de débits à l'exutoire des bassins versants d'intérêt pour le groupe. Ils communiquent ces prévisions aux gestionnaires d'ouvrages sous une forme résumée :

- si la probabilité de dépassement des seuils surveillés selon la prévision est supérieure à $\frac{1}{2}$, un “*fort risque*” est communiqué, si elle est comprise entre $\frac{1}{10}$ et $\frac{1}{2}$, un “*risque*” est communiqué et, si elle est inférieure à $\frac{1}{10}$, un “*faible risque*” est communiqué,
- une fois par semaine à une fois par jour (en fonction des enjeux du moment pour chaque site), 50 scénarios équiprobables de débits à l'exutoire du bassin versant sont transmis, assortis du mode et des quantiles à 10% et à 90% de la distribution prédictive pour chaque jour de prévision.

L'incertitude portant sur les prévisions est aussi communiquée chez Hydro-Québec : les prévisions hydrologiques sont transmises quotidiennement aux gestionnaires sous la forme d'un ensemble de plus de 150 scénarios hydrologiques illustrant l'étendue des situations hydrologiques possibles.

0.3 Les systèmes de prévision hydrologique actuels

Dans cette thèse, nous nous concentrons sur la façon dont les prévisions hydrologiques sont produites, et l'incertitude qui leur est associée est quantifiée, à partir des informations météorologiques à disposition, et plus particulièrement à partir des prévisions météorologiques d'ensemble.

6. c'est-à-dire sous la forme d'une estimation ponctuelle pouvant correspondre au mode de la distribution prédictive, à sa médiane ou bien encore à une simulation selon cette distribution.

7. Division Technique Générale d'EDF, division chargée (entre autres) d'émettre des prévisions hydrométéorologiques pour le groupe EDF. EDF-DTG est partenaire et financeur du présent travail de recherche.

PRÉVISIONS MÉTÉOROLOGIQUES D'ENSEMBLE Produites par des centres de prévision météorologiques, les prévisions météorologiques d'ensemble constituent une tentative pour décrire l'incertitude relative au futur état de l'atmosphère en passant d'une prédiction unique et déterministe de cet état à un ensemble de scénarios illustrant ses réalisations possibles, les membres de l'ensemble. Ces membres sont obtenus en perturbant les conditions initiales (mal connues) ou la paramétrisation du modèle atmosphérique, ou en utilisant plusieurs modèles atmosphériques.

Le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT), par exemple, génère les 50 membres du CEPMMT-EPS (Ensemble Prediction System) en effectuant 50 perturbations des conditions initiales de son simulateur atmosphérique. Celles-ci sont effectuées selon les directions de plus grande déviation du simulateur (identifiées sur une version simplifiée de ce simulateur). Chaque perturbation est assortie de son opposée. L'ampleur des perturbations est choisie de façon à ce que la dispersion résultante corresponde aux erreurs attendues (Buizza et al., 2008).

Intrants

Les systèmes de prévision hydrologique prennent généralement en intrants des scénarios météorologiques transmis par des centres de prévision météorologique.

Les prévisions utilisées par défaut dans le système d'EDF sont les prévisions météorologiques d'ensemble du CEPMMT.

Le système de prévision d'Hydro-Québec prend en intrants, à ce jour, les prévisions déterministes du Centre Météorologique Canadien (CMC). Hydro-Québec s'est toutefois récemment tourné vers les prévisions météorologiques d'ensemble suite à des études menées dans le cadre du projet P66.01 (Perreault, 2014, 2015).

Simulateurs pluie-débit

La plupart des systèmes de prévision hydrologique actuels, dont ceux d'EDF et d'Hydro-Québec, font intervenir un simulateur pluie-débit (SPQ, ou modèle pluie-débit⁸).

Un SPQ est un code de calcul intégrant des opérateurs mathématiques pour mimer le processus hydrologique, fortement non linéaire, transformant la météorologie observée sur un bassin versant (généralement les précipitations et les températures⁹) en débits à son exutoire.

Prise en compte des incertitudes

Les incertitudes qui interviennent dans les systèmes de prévision hydrologique à moyen-terme rencontrés dans la littérature ou dont nous avons observé le fonc-

8. Dans ce manuscrit, le mot "modèle" est généralement employé pour "modèle statistique" tandis que le mot "simulateur" est employé pour les modèles physiques ou déterministes.

9. La dénomination *pluie-débit* est donc abusivement restrictive vis-à-vis de l'opération réalisée par les SPQ.

tionnement *in situ* proviennent de deux sources principales (Krzysztofowicz, 1999). La première est relative à la météorologie future. La seconde émane de l'imparfaite représentation de la transformation pluie-débit par les SPQ. Ces incertitudes sont successivement prises en compte dans la chaîne de prévision.

PRISE EN COMPTE DE L'INCERTITUDE MÉTÉOROLOGIQUE La première étape de la chaîne de prévision consiste généralement à prendre en compte l'incertitude portant sur la météorologie. Ce *post-traitement météorologique* vise à produire, à partir des prévisions météorologiques issues des centres de prévision, une prévision météorologique probabiliste *fiable*, c'est-à-dire telle que les observations météorologiques puissent être supposées être issues de la distribution de probabilité prédite. Ce n'est pas le cas lorsque l'on considère la distribution empirique obtenue à partir des membres des prévisions météorologiques d'ensemble (Hamill & Colucci, 1997; Bougeault et al., 2010; Perreault, 2014, 2015). Il est par conséquent nécessaire d'appliquer un post-traitement.

Les scénarios météorologiques simulés selon la loi de probabilité issue du post-traitement sont supposés être fiables. Ils sont donc supposés avoir les mêmes propriétés statistiques que les séries météorologiques observées. Ils peuvent par conséquent être utilisés en intrant d'un SPQ dont les paramètres ont été préalablement estimés à partir d'une série hydrométéorologique observée.

À cette étape, le SPQ intervient pour passer du domaine météorologique au domaine hydrologique. Chaque scénario météorologique simulé est utilisé en intrant du SPQ pour produire une simulation hydrologique.

PRISE EN COMPTE DE L'INCERTITUDE HYDROLOGIQUE Les simulations hydrologiques doivent aussi être post-traitées afin de tenir compte de l'imperfection des SPQ et afin d'intégrer l'information issue des derniers débits observés, non prise en compte par les SPQ.

Dans la plupart des systèmes de prévision hydrologique opérationnels que nous avons étudiés, un *post-traitement hydrologique*, dernier maillon de la chaîne de prévision, est effectué en plusieurs étapes successives :

- La première étape vise à assimiler les dernières observations de débits.
- La deuxième étape permet de tenir compte de l'incertitude issue de l'imperfection de la transformation opérée par le SPQ. On la réalise soit en perturbant le SPQ de plusieurs façons (chez Hydro-Québec par exemple), soit en effectuant des simulations selon une distribution prédictive (chez EDF par exemple).
- Une dernière étape vise enfin à recréer la cohérence spatio-temporelle souvent perdue lors des deux premières étapes.

Expertise

Chez EDF et chez Hydro-Québec, des experts interviennent tout au long de la chaîne de prévision. Ils veillent à la cohérence physique des prévisions émises, cor-

rigent les éventuels biais systématiques du système de prévision et intègrent des informations auxiliaires telles que des prévisions météorologiques alternatives émises par les centres de prévision.

De plus, certaines parties des systèmes de prévision décrites ci-dessus sont réalisées manuellement par les prévisionnistes. Chez Hydro-Québec, ils guident par exemple l'assimilation des dernières observations de débits lors de la première étape du post-traitement hydrologique et choisissent l'ampleur des perturbations effectuées au cours de la deuxième étape. Toujours chez Hydro-Québec, le post-traitement météorologique comprend une étape préalable de correction des prévisions du CMC qui est réalisée par un prévisionniste.

0.4 Aide statistique à la représentation des incertitudes

Ce travail doctoral avait pour but d'apporter une aide statistique pour l'amélioration des systèmes de prévision hydrologique (probabilistes) prenant en intrants des prévisions météorologiques d'ensemble.

Les principaux critères de performance des prévisions probabilistes sont leur fiabilité et leur précision. La section 1.3 définit ces critères et présente les scores qui permettent de les mesurer.

Nous nous sommes appuyés pour ce travail sur le raisonnement statistique. Ce raisonnement, s'il est bien conduit, permet de produire des prévisions probabilistes fiables en modélisant la relation entre les variables à prévoir et les informations disponibles pour les prévoir. Plusieurs choix d'inférence sont alors possibles pour les paramètres des modèles, nous utilisons pour notre part des méthodes d'inférence maximisant la vraisemblance.

Notre travail devait de plus contribuer au développement de méthodes pouvant s'intégrer dans les systèmes de prévision actuels d'EDF et d'Hydro-Québec.

CONTRIBUTION AU POST-TRAITEMENT MÉTÉOROLOGIQUE En ce qui concerne le post-traitement météorologique, une demande, commune à EDF et Hydro-Québec, était de permettre l'intégration automatique, dans leurs systèmes, de prévisions météorologiques alternatives, actuellement prises en compte manuellement. Une telle intégration devait résulter en un gain en précision par rapport aux systèmes automatiques actuels et en un gain de temps pour les prévisionnistes. Il nous fallait de plus proposer une méthode permettant l'intégration de prévisions expertisées produites en amont de certaines chaînes de prévision telles que celle d'Hydro-Québec.

Nous voulions de plus proposer une méthode parcimonieuse à même d'utiliser l'information issue de l'historique des mesures météorologiques, de tenir compte de la dépendance temporelle des phénomènes météorologiques et de prendre en compte la dépendance entre les variables météorologiques.

Les méthodes de post-traitement actuellement proposées dans la littérature ne possèdent pas toutes ces qualités à la fois. Nous nous sommes donc appliqués, dans ce travail, à proposer une structure ayant le potentiel de traiter l'ensemble de ces

aspects. Cette structure est fondée sur la propriété statistique d'échangeabilité. Nous présentons cette propriété dans le chapitre 3.

CONTRIBUTION AU POST-TRAITEMENT HYDROLOGIQUE Dans les systèmes de prévision considérés dans ce travail de thèse, le post-traitement hydrologique nécessite plusieurs étapes successives. L'intervention systématique de prévisionnistes y est de plus parfois nécessaire.

Il nous paraissait pertinent de développer un modèle permettant de réaliser l'ensemble de ces étapes de façon efficace et intégrée. De tels modèles sont déjà proposés dans la littérature mais ils ne donnent pas entière satisfaction car plusieurs régimes d'erreurs sont possibles et pas toujours représentés.

Dans une contribution présentée dans le chapitre 4, nous proposons un modèle intégré faisant intervenir deux régimes d'erreur pour le post-traitement hydrologique. L'appartenance à un régime est, dans ce modèle, guidée par des variables internes aux SPQ.

0.5 Organisation du mémoire

Ce mémoire décrit comment nous avons formalisé de façon probabiliste les incertitudes météorologiques et hydrologiques. Il est organisé comme suit :

- Dans le chapitre 1, nous décrivons, dans une première section, les ressources mises à notre disposition par les partenaires industriels, illustratives de celles dont les services opérationnels disposent pour émettre leurs prévisions (section 1.2) : les SPQ d'EDF et d'Hydro-Québec, des prévisions météorologiques et des historiques hydrométéorologiques relatifs à 15 bassins versants français et à 8 bassins versants québécois. Nous décrivons ensuite, dans la section 1.3, les outils scientifiques d'évaluation de la performance des systèmes de prévision. Il s'agit notamment d'outils diagnostiques de la fiabilité et de scores propres, qui évaluent à la fois la fiabilité et la précision des prévisions. Il résulte des sections 1.2 et 1.3 une définition plus avancée des objectifs des systèmes opérationnels et, par extension, de ce travail de doctorat, présentée dans la section 1.4.
- Dans le chapitre 2, nous présentons un état de l'art des systèmes de prévision hydrologique à moyen-terme. Un article de revue soumis pour publication (en Annexe) reprend en partie ce chapitre. Dans la section 2.1, nous expliquons la structure séparant les sources d'incertitude météorologique et hydrologique, commune à beaucoup de systèmes de prévision opérationnels et déjà décrite dans la section 0.3. Dans la section 2.2, nous discutons des outils de la statistique couramment utilisés dans ces systèmes, notamment des méthodes de copules et des méthodes permettant de retrouver un cadre gaussien. Dans les sections 2.3 et 2.4, nous présentons un état de l'art des méthodes de prise en compte des incertitudes respectivement météorologiques et hydrologiques.

Ce chapitre motive les contributions méthodologiques de ce travail de thèse, présentées dans les chapitres 3 et 4.

- Le chapitre 3 présente la méthode de post-traitement des prévisions météorologiques d'ensemble que nous proposons. La section 3.1 décrit la version de cette méthode adaptée au traitement de variables pouvant être supposées gaussiennes, telles que les températures. La section 3.2 étend notre méthode au cas des ensembles de précipitations. La section 3.3 présente enfin une méthode alternative que nous avons développée pour le post-traitement d'un ensemble de précipitations.
- Le chapitre 4 présente la méthode de post-traitement hydrologique que nous proposons. Ce chapitre correspond à un article soumis pour publication (en cours de révision).
- Dans un dernier chapitre, nous faisons le bilan de nos contributions méthodologiques dans une première section (section 5.1), nous les combinons en un système complet sur un exemple (section 5.2), puis discutons de leur implémentation en situation opérationnelle, où les hypothèses sur lesquelles s'appuient nos modèles peuvent se révéler fragiles (section 5.3). Nous dressons de plus les perspectives qui se dessinent à la fin de notre travail de recherche, dans un contexte où les systèmes de prévisions s'automatisent.

1

Ressources pour les prévisions hydrométéorologiques et critères d'évaluation

1.1	Notations	12
1.2	Sources d'informations pour les prévisions hydrologiques à moyen-terme	12
1.2.1	De longues séries d'observations hydrométéorologiques	15
1.2.2	Des prévisions météorologiques déterministes et d'ensemble	17
1.2.3	Les simulateurs pluie-débit d'EDF et d'Hydro-Québec	24
1.3	Critères d'évaluation des prévisions hydrométéorologiques probabilistes	27
1.3.1	La fiabilité, première attente vis-à-vis des prévisions probabilistes	27
1.3.2	La précision, second critère de performance des prévisions probabilistes	31
1.3.3	Des scores justes pour évaluer fiabilité et précision des prévisions probabilistes	32
1.3.4	Évaluation des prévisions hydrologiques chez les producteurs d'hydro-électricité	34
1.4	Définition avancée des objectifs du travail doctoral	35

1.1 Notations

Commençons par fixer les notations utilisées dans ce manuscrit. Nous utilisons ici l'écriture entre crochets pour la distribution de probabilité choisie pour représenter les phénomènes étudiés, comme proposé par Gelfand & Smith (1990) : soient X et Y des variables aléatoires, $[Y = y|X = x]$ désigne la densité de probabilité de Y en y conditionnellement à $\{X = x\}$ si Y est une variable aléatoire continue, la probabilité de l'événement $\{Y = y\}$ conditionnellement à l'événement $\{X = x\}$ si Y est une variable discrète. Les autres notations utilisées dans ce manuscrit sont définies dans le tableau 1.1. En situation de prévision, X et Y jouent généralement des rôles dissymétriques : Y désigne la quantité à prévoir tandis que X désigne une quantité informative connue.

Y	notation générique pour les quantités à prévoir
G	(hypothétique) fonction de répartition de la quantité à prévoir
F	fonction de répartition prédite pour la quantité à prévoir
f	densité associée à F (si elle existe)
y_t	réalisation de la variable aléatoire Y au temps t
X	notation générique pour les intrants de la prévision
Q	variable désignant les débits (journaliers)
A	variable désignant les prévisions météorologiques
P	variable désignant la météorologie réelle
$\mathcal{M} : x \mapsto \mathcal{M}(x)$	transformation opérée par un simulateur pluie-débit (SPQ)
$\mathcal{M}(P)$	sorties du SPQ prenant en intrant la météorologie observée

TABLEAU 1.1 – Notations du manuscrit

1.2 Sources d'informations pour les prévisions hydrologiques à moyen-terme

Nous décrivons dans cette partie les jeux de données que nous ont transmis EDF et Hydro-Québec, commanditaires de ce projet. Ces jeux de données sont représentatifs des informations à disposition, ou bien anticipées comme étant prochainement à disposition, des prévisionnistes en situation opérationnelle.

Il s'agit d'informations de plusieurs natures :

- un historique d'observations des variables météorologiques et hydrologiques d'intérêt (décrit dans la section 1.2.1),
- des prévisions météorologiques émises par des centres de prévision météorologique à des horizons de quelques jours (décrites dans la section 1.2.2),
- des simulateurs mimant le phénomène physique de la transformation pluie-débit (décrits dans la section 1.2.3).

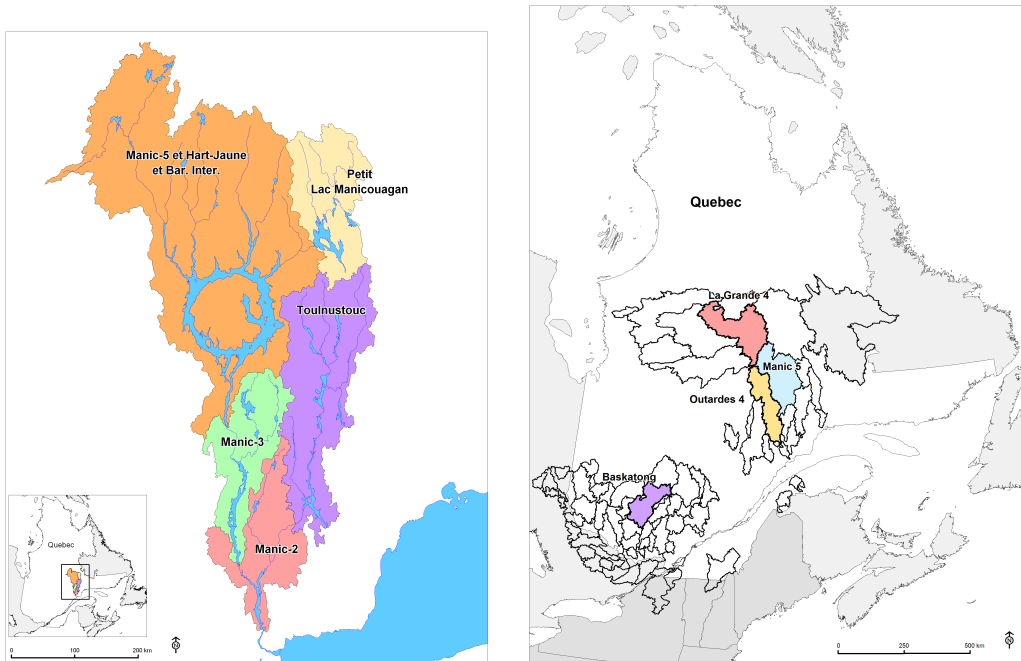


FIGURE 1.1 – Localisation et géographie des bassins versants québécois du projet pris en exemple dans le chapitre 3 (figure de gauche) et dans le chapitre 4 (figure de droite) (source : Frédéric Guay, IREQ).

EDF-DTG et Hydro-Québec nous ont respectivement fourni ces données pour 15 et 8 bassins versants, pour la plupart incluant une retenue gérée par l'hydro-électricien. La superficie moyenne des 15 bassins versants français est de $1\,266\text{ km}^2$, avec un minimum de 63 km^2 et un maximum de $3\,727\text{ km}^2$. Dans le cas québécois, la superficie des bassins versants considérés est beaucoup plus grande ($13\,031\text{ km}^2$ en moyenne, $4\,085\text{ km}^2$ au minimum et $28\,443\text{ km}^2$ au maximum) avec souvent des surfaces de retenues non négligeables (7% de la surface dans le cas de Manic 5, un bassin versant de $24\,622\text{ km}^2$, par exemple).

La figure 1.1 montre, à gauche, la localisation et la géographie de 5 des 8 bassins versants québécois étudiés, les bassins versants du complexe hydro-électrique Manicouagan, pris en exemple dans le chapitre 3. Parmi eux, le bassin versant de Manic 5 est aussi pris en exemple dans le chapitre 4. La partie droite de la figure 1.1 montre la localisation et la géographie des 4 bassins versants québécois pris en exemple dans le chapitre 4.

La figure 1.2 montre la localisation, la géographie et l'altitude des 15 bassins versants français pour lesquels EDF-DTG nous a transmis des données.

Les tableaux 1.2 et 1.3 listent les bassins versants (respectivement français et québécois) considérés dans ce travail et en donnent la superficie.

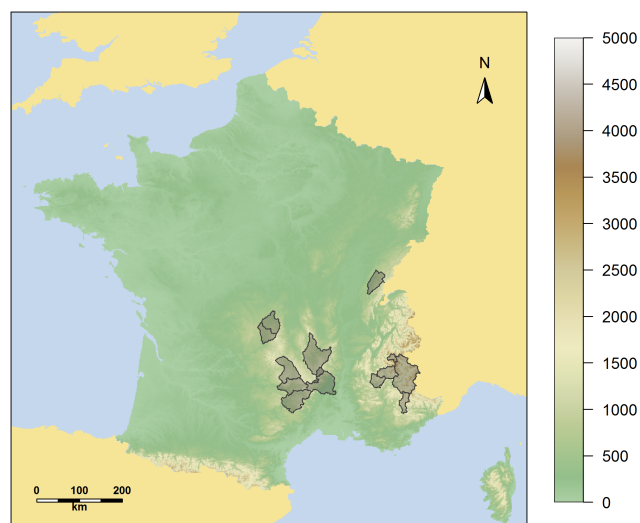


FIGURE 1.2 – Localisation, géographie et altitude (donnée en m par l'échelle à droite de la carte) des bassins versants français du projet (source : David Penot, EDF-DTG).

TABLEAU 1.2 – Bassins versants d'EDF considérés dans le projet et leur superficie.

Nom du bassin	Superficie (km^2)
L'Ain à Vouglans	1 125
L'Ardèche à Sauze	2 240
La Borne à Pont de Nicoulaud	63
Le Buech aux Chambons	723
La Dordogne à Bort	1 010
Le Drac au Sautet	1 018
La Durance à Serreponçon	3 727
La Loire à Bas-en-Basset	3 005
La Loire à Pont-la-Borie	229
Le Lot à Castelnau	1 650
La Romanche au Chambon	322
La Sioule à Fades	1 300
Le Tarn à Millau	2 170
La Truyère à Grandval	1 930
Le Verdon à Castillon	655

TABLEAU 1.3 – Bassins versants d'Hydro-Québec considérés dans le projet et leur superficie.

Nom du bassin	Superficie (km^2)
Baskatong	13 057
La Grande 4	28 443
Outardes 4	17 109
Manic 5	24 608
Manic 2	4 085
Manic 3	4 264
Petit Lac Manicouagan	4 553
Toulnostouc	8 117

1.2.1 De longues séries d'observations hydrométéorologiques

Une première source d'information à disposition des prévisionnistes est un historique de variables hydrométéorologiques (températures, précipitations et débits) observées sur les bassins versants d'intérêt au pas de temps journalier jusqu'au jour auquel sont émises les prévisions. Pour notre projet, nous disposons des historiques hydrométéorologiques correspondant aux 15 bassins versants français et 8 bassins versants québécois à l'étude (issus des bases de données opérationnelles). La plage de données commune aux variables hydrologiques et météorologiques s'étend de 1969 à 2008 pour les bassins versants français et de 1966 à 2013 pour les bassins versants québécois. La plage de données relative aux variables météorologiques est toutefois souvent beaucoup plus longue que celle relative à l'hydrologie. Le tableau 1.4 synthétise les données à notre disposition.

TABLEAU 1.4 – Historiques d'observations hydrométéorologiques journalières fournis par EDF-DTG et Hydro-Québec.

Source des données :	EDF-DTG	Hydro-Québec
Nature des données :	Température moyenne, cumul des précipitations totales, débit à l'exutoire	Températures minimale et maximale journalières, cumul des précipitations en pluie et en neige, débit à l'exutoire
Nombre de bassins versants :	15	8
Plage de données commune :	1969-2008*	1966-2013

*jusqu'à fin 2015 pour 3 des 15 bassins versants

Les variables météorologiques considérées chez Hydro-Québec et chez EDF correspondent aux intrants de leurs SPQ respectifs, décrits dans la section 1.2.3. Ces intrants météorologiques des SPQ sont obtenus en agrégeant, dans le temps (par jour) et dans l'espace (par bassin versant), des mesures effectuées en des stations météorologiques réparties sur et autour des bassins versants pour obtenir des *grandeurs de bassin*.

Chez EDF, par exemple, l'agrégation spatiale sur le bassin versant se fait en tenant compte de la courbe hypsométrique du bassin et du gradient de température associé dans le cas des températures, et peut se faire à l'aide de polygones de Thiessen dans le cas des précipitations.

Chez d'Hydro-Québec, les variables météorologiques prises en intrants du SPQ comprennent la température minimale et la température maximale journalières, le cumul des précipitations journalier sous forme de neige et sous forme de pluie. Chez EDF, ces informations sont davantage résumées : on considère la température moyenne journalière et le cumul des précipitations journalier.

Le phénomène de précipitation étant très hétérogène spatialement, il est difficile

d'évaluer la lame d'eau (précipitation totale) sur l'ensemble d'un bassin versant à partir de mesures aux stations météorologiques. Ceci est particulièrement vrai pour les bassins versants du Nord du Québec, sur lesquels la densité de pluviomètres est faible.

Les données de débits à l'exutoire sont, pour leur part, soit obtenues à partir d'une mesure de la hauteur d'eau dans la rivière (par le biais d'une courbe de tarage), lorsque c'est possible, soit calculées à partir de la hauteur d'eau au réservoir et du débit turbiné (par bilan hydrique). Dans ce dernier cas, le débit réel est très incertain (par exemple en raison de l'influence du vent sur la hauteur de l'eau, en particulier dans le cas des grands réservoirs québécois). Les débits peuvent de plus subir l'influence d'interventions humaines (turbinage d'un barrage amont par exemple). Les données de débits originales font donc l'objet d'un lissage, parfois revu *a posteriori* suite à de nouvelles mesures. En situation opérationnelle, le débit qui servira de référence (par exemple pour évaluer les prévisions) peut donc n'être connu que plusieurs jours après son occurrence, ce qui limite l'exploitation d'une persistance temporelle pour la prévision des débits.

Dans ce travail, nous ne tenons pas compte des incertitudes relatives aux observations hydrométéorologiques : nous travaillons avec les températures et précipitations journalières moyennées sur les bassins versants et avec les débits lissés de la base de données de référence des hydro-électriciens.

L'analyse exploratoire des données montre une forte saisonnalité des températures, comme l'illustre la figure 1.3 avec comme exemple les températures maximales journalières observées sur le bassin versant québécois de Baskatong par jour de l'année de 1950 à 2010. Dans le cas des précipitations et des débits, la saisonnalité est nettement plus marquée au Québec, où l'hiver et l'été sont plus contrastés qu'en France. La figure 1.4 l'illustre dans le cas des débits. Elle montre les débits journaliers des bassins versants de Baskatong (au Québec) et de l'Ain à Vouglans (en France) par jour de l'année de 1953 à 2008. Cette saisonnalité constitue une source d'information importante pour la prévision. En l'absence d'informations supplémentaires, des prévisions, dites *climatologiques*, peuvent être émises en se fondant sur les fluctuations météorologiques et hydrologiques du passé.

La possibilité d'utiliser un long historique pour représenter les fluctuations possibles du futur par un modèle probabiliste de la climatologie est toutefois limitée : les méthodes d'inférence de telles structures probabilistes reposent sur une hypothèse de stationnarité. Or, les données considérées dans ce travail de thèse présentent des manquements à cette hypothèse sous plusieurs formes. Certaines instationnarités sont dues à un changement climatique. De telles instationnarités sont particulièrement visibles sur les observations de températures comme l'illustre la figure 1.5 : on observe, entre 1950 et 2008, un accroissement des températures moyennes annuelles de l'ordre d'un degré Celsius sur les 11 bassins versants français de l'échantillon représenté. Dans le cas des débits, Perreault et al. (2000) montrent une soudaine diminution des débits moyens annuels au début des années quatre-vingt pour des

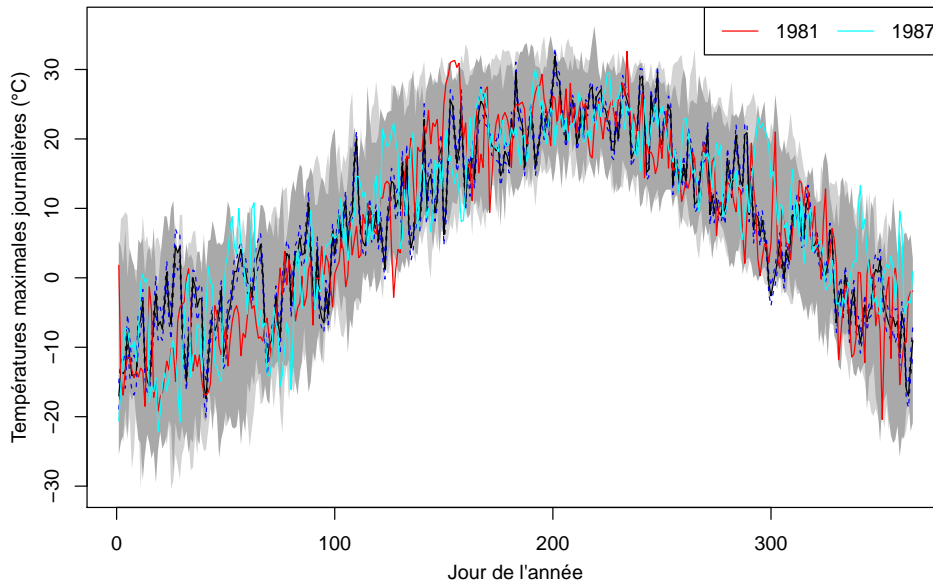


FIGURE 1.3 – Bagplot fonctionnel (Shang & Hyndman, 2016) des températures maximales journalières sur le bassin versant de Baskatong (province du Québec) de 1950 à 2010 par jour de l'année. La courbe noire représente la trajectoire médiane et les courbes en pointillés bleus l'intervalle de confiance à 95% qui lui est associé. La zone gris foncé regroupe 50% des trajectoires (les plus centrales) et la zone gris clair 99% des trajectoires (les plus centrales). Les années représentées sont les années identifiées comme particulières, non comprises dans la zone gris clair.

rivières situées au Nord-Est de la péninsule Québec-Labrador. Cette diminution est probablement liée à une variation climatique. D'autres instationnarités (appelées non-homogénéités) peuvent être causées par un changement dans la méthode d'acquisition des données (Peterson et al., 1998) ou bien par une modification dans la morphologie du bassin versant.

1.2.2 Des prévisions météorologiques déterministes et d'ensemble

En situation de prévision, en plus de l'historique météorologique, des prévisions météorologiques issues de simulateurs de la dynamique atmosphérique sont disponibles à un horizon de quelques jours.

Les jeux de données de prévisions météorologiques qui nous ont été fournis (après un formatage par les équipes opérationnelles) sont listés dans le tableau 1.5. Le tableau 1.6 décrit les prévisions produites par les centres de prévision météorologique correspondantes. Les prévisions, transmises par les centres climatiques à leurs clients sous la forme de grilles quadrillant l'espace, nous sont transmises agrégées par bassin versant et par jour, sous la forme de *prévisions de bassin*.

Chez EDF, les prévisions de bassin sont les prévisions du CEPMMT émises au point de grille auquel elles sont les plus corrélées avec les grandeurs de bassin observées. Chez Hydro-Québec, on obtient les prévisions de bassin en moyennant les prévisions issues par les centres climatiques aux points de grilles situés sur ou à proximité des bassins versants.

TABLEAU 1.5 – Jeux de données de prévisions météorologiques fournis par EDF-DTG (“EDF-DTG 1” et “EDF-DTG 2”) et par Hydro-Québec.

Jeu de données	EDF-DTG 1	EDF-DTG 2	Hydro-Québec
Prévisions	CEPMMT-EPS ANA*	CEPMMT-EPS CEPMMT-ctl CEPMMT-det et Exp	CEPMMT-EPS CMC-EPS NCEP-GEFS
Plage de données	2005-2008	2011-2015	2010-2014
Horizon de prévisions maximal (en jours)	7	9	10

* uniquement pour les précipitations

Certaines de ces prévisions sont des prévisions d’ensemble : les prévisions CEPMMT-EPS, ANA, CMC-EPS et NCEP-GEFS (voir le tableau 1.6). Leur production étant coûteuse (elle nécessite de faire tourner à plusieurs reprises le simulateur de la dynamique atmosphérique), les prévisions d’ensemble sont généralement produites à une résolution spatiale moins fine que les prévisions déterministes correspondantes.

EDF-DTG nous a successivement fourni deux jeux de données, intitulés “EDF-DTG 1” et “EDF-DTG 2” dans le tableau 1.5. Les prévisions du jeu de données “EDF-DTG 1” correspondent à celles actuellement utilisées opérationnellement par EDF (mais, aux opérations, l’horizon de prévision actuellement accessible est de 15 jours). Parmi elles, les prévisions dites Analogues (ANA), produites depuis longtemps par EDF-DTG (Obled et al., 2002), sont issues de l’idée que si, à deux dates, les champs de pression atmosphérique sont similaires, alors les précipitations de surface le seront aussi. Pour les produire, on compare chacun des 50 scénarios de champs de pressions produits par le CEPMMT à 700 hPa et à 1000 hPa avec un historique de champs de pression. 50 dates “analogues” sont ainsi identifiées pour chaque scénario, d’où 2500 scénarios de précipitations. On échantillonne ensuite 50 scénarios parmi les 2500 de sorte à ce qu’ils préservent des caractéristiques bien choisies des 50 membres des prévisions CEPMMT-EPS. Les prévisionnistes d’EDF-DTG ont aussi accès à d’autres sources de prévisions non présentes dans le jeu de données. Il s’agit par exemple des prévisions PEARP, AROME (avec un horizon maximal de 30h) et ARPEGE (avec un horizon maximal de 102h) de Météo-France.

Les prévisions fournies par Hydro-Québec proviennent de la base de données *THORPEX Interactive Grand Global Ensemble* (TIGGE). Il s’agit d’une base de

TABLEAU 1.6 – Caractéristiques de prévisions météorologiques produites par des centres de prévision météorologique et intervenant dans le jeu de données fourni par EDF-DTG ou par Hydro-Québec.

Prévisions	Nombre de membres	Horizon de prévision (en jours)	Résolution spatiale
CEPMMT-EPS ^a : prévisions d'ensemble du CEPMMT ^b	50	15	0.5°
CEPMMT-ctl : membre contrôle correspondant	1	15	0.5°
CEPMMT-det : prévision à haute résolution du CEPMMT	1		0.125°
ANA : prévisions analogues d'EDF-DTG	50	7	-
Exp : prévisions des experts d'EDF-DTG	1	7	-
CMC-EPS : prévisions d'ensemble du CMC ^c	20	16	1°
NCEP-GEFS ^d : prévisions d'ensemble du NCEP ^e	20	15	1°

a. Ensemble Prediction System.

b. Centre Européen pour les Prévisions Météorologiques à Moyen Terme.

c. Centre Météorologique Canadien.

d. Global Ensemble Forecast System.

e. National Centres for Environmental Prediction.

données constituée dans le cadre du programme *The Observing System Research and Predictability Experiment* (THORPEX) de l'Organisation Météorologique Mondiale visant à améliorer la précision des prévisions météorologiques à moyen-terme. Cette base de données comprend des prévisions d'ensemble émises depuis 10 ans par 10 centres de prévision météorologique globaux. Aux opérations, les prévisions d'Hydro-Québec se fondent surtout sur les prévisions déterministes produites par Environnement Canada à horizon 9 jours bien que d'autres prévisions soient aussi prises en compte par une expertise humaine (les prévisions d'ensemble du CMC et du NCEP par exemple). L'utilisation de la base de données TIGGE dans le cadre de ce projet vise à anticiper l'intégration des prévisions ensemblistes dans le système de prévision d'Hydro-Québec.

La figure 1.6 montre deux situations où l'on cherche à prévoir des températures de bassin (ligne tiretée noire), l'une sur le bassin versant de l'Ain à Vouglans (en France), et l'autre sur le bassin versant de Manic 2 (au Québec). On dispose pour

cela respectivement des prévisions de bassin des jeux de données “Hydro-Québec” et “EDF-DTG 2” (décrits dans le tableau 1.5).

Comme dans le cas des historiques hydrométéorologiques, les prévisions météorologiques présentent des instationnarités. Elles sont en particulier dues à des changements dans les simulateurs de la dynamique atmosphérique, auquel cas elles se présentent sous la forme de ruptures. À titre d'exemple, dans notre historique de prévisions, la méthode de production des prévisions d'ensemble du CEPMMT a fait l'objet de 6 changements entre 2006 et 2008, tandis que celle du CMC a fait l'objet de changements majeurs en août 2011 puis en février 2013¹. Ces changements constituent un obstacle à l'apprentissage de la relation entre les prévisions issues des simulateurs de la dynamique atmosphérique et les observations météorologiques. Pour pallier ce problème de rupture de stationnarité, certains centres climatiques (tels qu'Environnement Canada) effectuent des *reprévisions*, c'est-à-dire refont de prévisions avec la méthode de production des prévisions la plus récente (Hamill et al., 2006). Ces reprévisions ne sont toutefois pas toujours disponibles pour la version des simulateurs de la dynamique atmosphérique utilisée opérationnellement.

1. <https://software.ecmwf.int/wiki/display/TIGGE/Model+upgrades>.

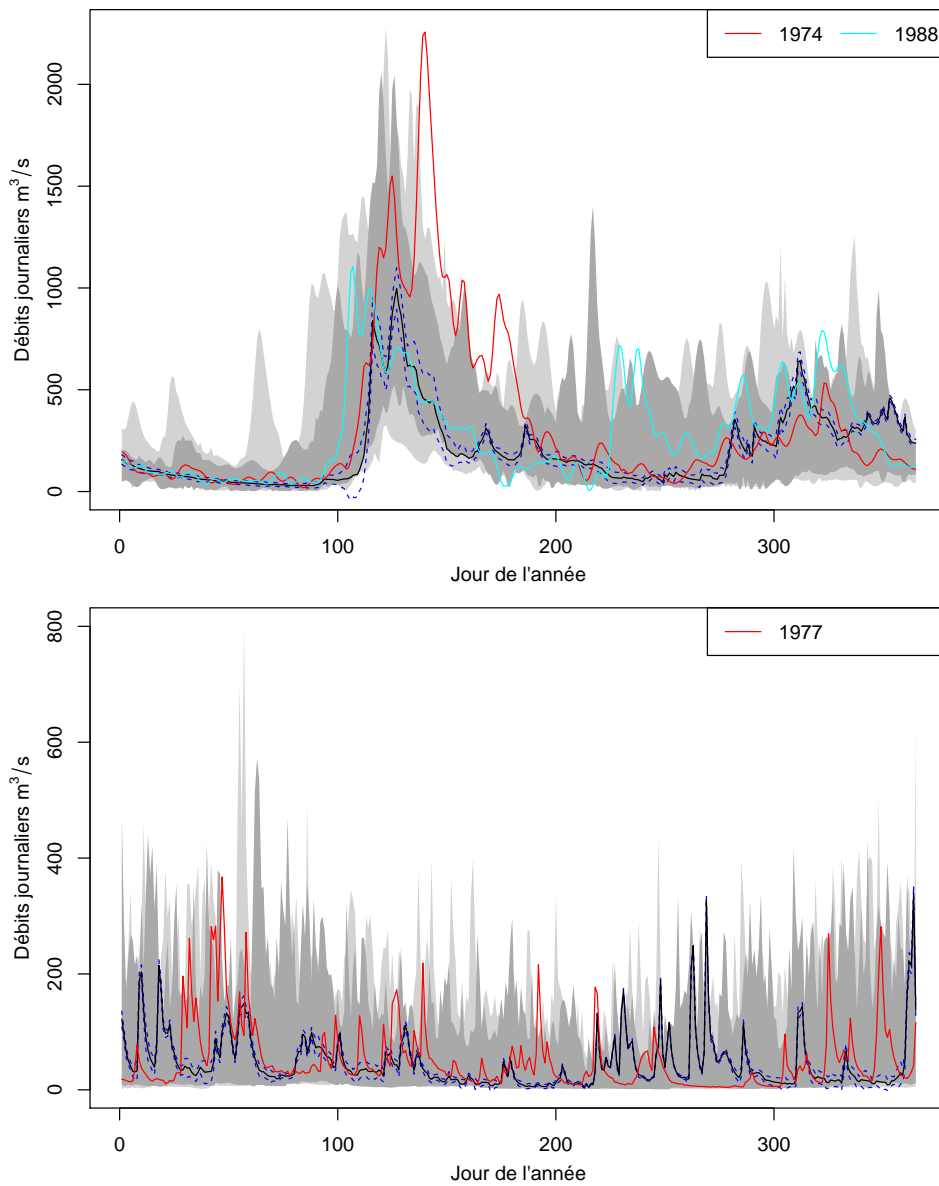


FIGURE 1.4 – Bagplots fonctionnels (Shang & Hyndman, 2016) des débits journaliers à l'exutoire du bassin versant de Baskatong (province du Québec, figure du haut) et à l'exutoire du bassin versant de l'Ain à Vouglans (France, figure du bas) de 1953 à 2008 par jour de l'année. Sur chaque figure, la courbe noire représente la trajectoire médiane et les courbes en pointillés bleus l'intervalle de confiance à 95% qui lui est associé. La zone gris foncé regroupe 50% des trajectoires (les plus centrales) et la zone gris clair 99% des trajectoires (les plus centrales). Les années représentées sont les années identifiées comme particulières, non comprises dans la zone gris clair.

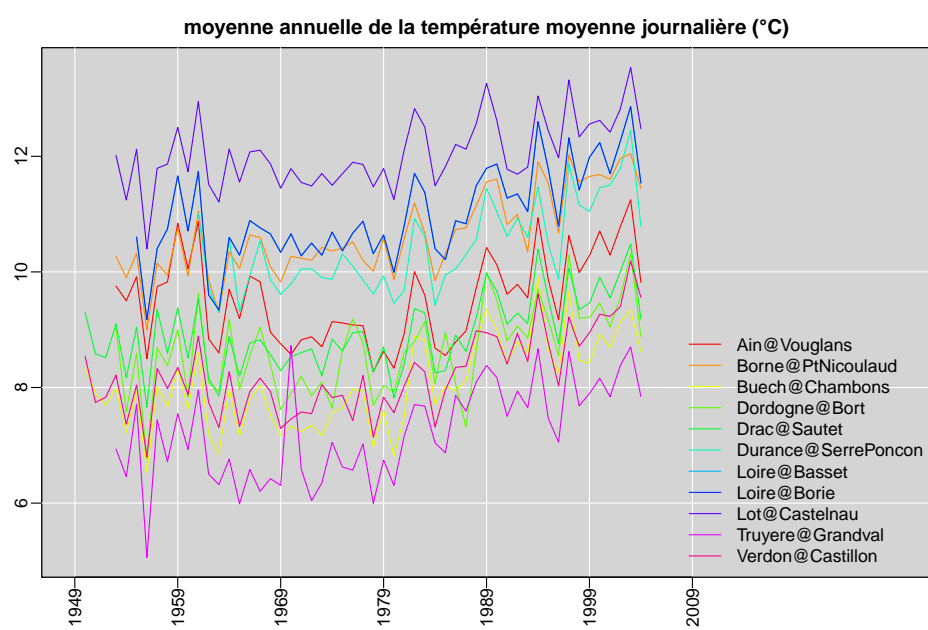


FIGURE 1.5 – Séries des températures moyennes annuelles de 1950 à 2008 sur les 11 bassins versants français considérés dans le chapitre 4.

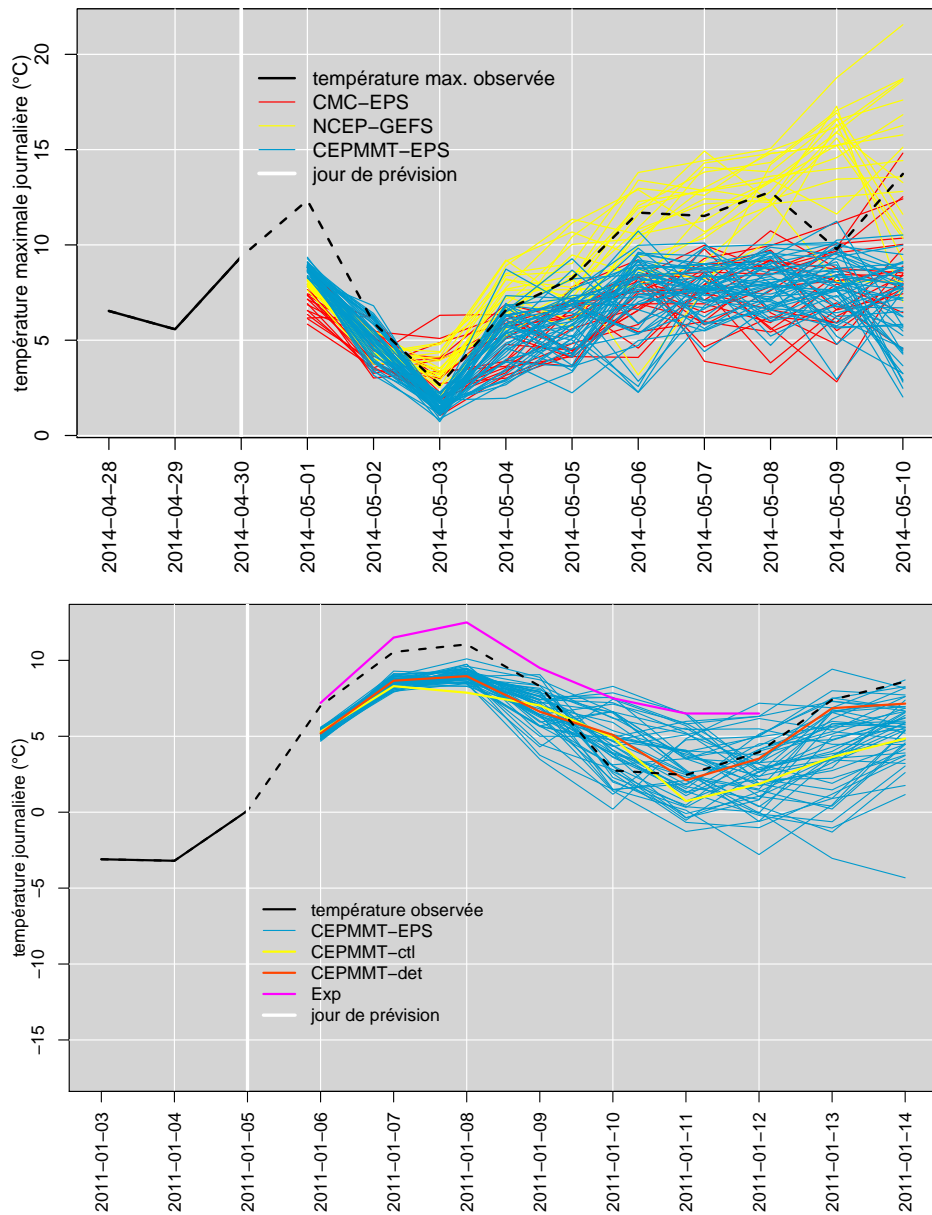


FIGURE 1.6 – Exemple de situation de prévision de températures maximales journalières sur le bassin versant québécois de Manic 2 (haut) et de températures moyennes journalières sur le bassin versant français de l’Ain à Vouglans (jeu de données “EDF-DTG 2”, bas). La barre verticale blanche désigne le jour où les prévisions sont émises. La ligne tiretée noire correspond aux températures à prévoir (qui ne seront observées qu’*a posteriori*).

1.2.3 Les simulateurs pluie-débit d'EDF et d'Hydro-Québec

Les précipitations et les températures de bassin peuvent être utilisées pour anticiper les futurs débits à l'exutoire des bassins versants. Pour ce faire, des simulateurs mimant le cheminement de l'eau dans un bassin versant ont été conçus : les simulateurs pluie-débit. Dans cette section, nous présentons les SPQ d'EDF (MORDOR, Garçon (1996)) et d'Hydro-Québec (HSAMI, Fortin (2000)).

MORDOR et HSAMI sont des transformations déterministes de type “modèle conceptuel global à réservoirs” à respectivement 22 et 23 paramètres. Comme les modèles à base physique mais contrairement aux modèles empiriques (tels que les réseaux de neurones, Coulibaly et al. (1999)), ce type de modèle cherche à mimer le processus réel grâce à une séquence d'opérateurs mathématiques. Mais contrairement aux modèles à base physique, il ne cherche pas à le représenter précisément à fine échelle en s'appuyant sur les équations de la physique des écoulements. Ainsi, il comprend des variables internes qui donnent une indication de l'état physique global du bassin versant sans pour autant chercher à parfaitement l'évaluer. Ces variables, dites d'état, correspondent essentiellement au niveau de remplissage de réservoirs conceptuels à interprétation physique (stock de neige, niveau d'une nappe souterraine, taux de saturation des sols,...). Les intrants de MORDOR et de HSAMI sont les grandeurs météorologiques de bassin précédemment décrites dans la section 1.2.1 : température moyenne et cumul de précipitations journaliers dans le cas de MORDOR ; température maximale, température minimale, pluie et neige journalières dans le cas de HSAMI. Leurs sorties sont des débits à l'exutoire des bassins versants. Les figures 1.7 et 1.8 illustrent schématiquement respectivement les SPQ MORDOR et HSAMI. Notons que le modèle HSAMI est actuellement en cours de révision majeure (Guay, 2014; Minville, 2014), l'illustration de la figure 1.8 pourrait donc bientôt être obsolète.

Nous disposons, dans le cadre de cette thèse, des paramètres des SPQ correspondant aux 15 bassins versants français et à 4 des 8 bassins versants québécois.

Le calage des paramètres des SPQ est effectué pour chaque bassin versant de telle sorte que les débits issus du SPQ à météorologie connue soient les plus proches possibles des débits observés selon des distances de type carré des écarts. Chez EDF (simulateur MORDOR), on maximise, selon les cas, le critère classique NSE (Nash-Sutcliffe Efficiency, Nash & Sutcliffe (1970)), le critère KGE (Kling-Gupta Efficiency, proposé par Gupta et al. (2009)), ces critères calculés uniquement sur certaines classes de débits, ou encore une combinaison de ces critères. Dans le cas de HSAMI, la méthode de calibration est en cours de révision majeure (Guay & Chartier, 2016; Minville, 2016) ; on maximise à ce jour un critère de type NSE.

Le critère NSE s'exprime de la façon suivante :

$$\text{NSE} = 1 - \frac{\sum_t (Q_t - \mathcal{M}(P)_t)^2}{\sum_t (Q_t - \bar{Q})^2},$$

où $\mathcal{M}(P)_t$ désigne la sortie du SPQ au temps t , Q_t désigne le débit au temps t

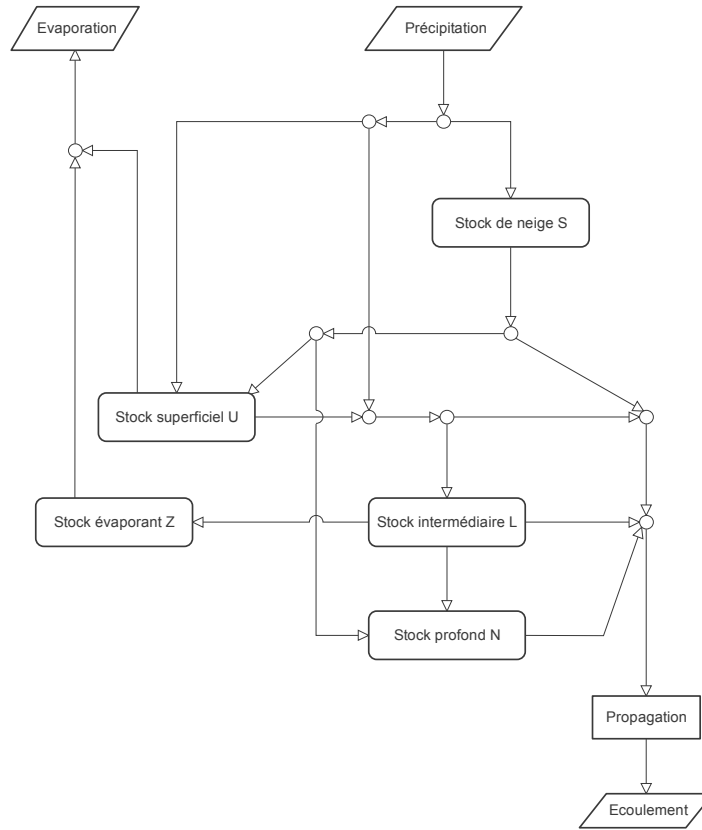


FIGURE 1.7 – Diagramme du simulateur pluie-débit d'EDF, MORDOR (Garçon, 1996).

et \bar{Q} désigne le débit moyen pendant la période de calage. Un NSE négatif signifie donc une prévision du SPQ moins performante, en somme des carrés des écarts, que celle obtenue en prévoyant systématiquement la moyenne des débits observés sur la période de calage.

Le KGE s'exprime de la façon suivante :

$$\text{KGE} = 1 - \sqrt{(1 - \alpha)^2 + (1 - r)^2 + (1 - \beta)^2}$$

$$\alpha = \frac{\hat{\sigma}_{\mathcal{M}(P)}}{\hat{\sigma}_Q} \quad \beta = \frac{\hat{\mu}_{\mathcal{M}(P)}}{\hat{\mu}_Q},$$

où $\hat{\mu}_X$ et $\hat{\sigma}_X$ désignent respectivement la moyenne et l'écart-type empiriques de la

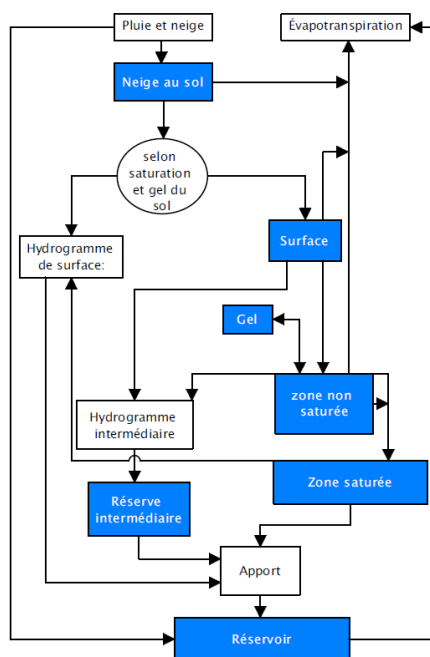


FIGURE 1.8 – Diagramme du simulateur pluie-débit d'Hydro-Québec, HSAMI (Fortin, 2000).

variable aléatoire X et r désigne le coefficient de corrélation linéaire empirique de Q et $\mathcal{M}(P)$. En maximisant le KGE, on cherche donc à réaliser une optimisation multi-critères dont les critères sont : (1) une corrélation maximale entre les sorties du SPQ et les débits observés (r doit être le plus proche possible de 1), (2) un biais multiplicatif minimal (β doit être le plus proche possible de 1), et (3) une dispersion des sorties du SPQ similaire à celle des débits observés (α doit être le plus proche possible de 1). À titre d'exemple, sur le bassin versant de l'Ain à Vouglans, avec le SPQ MORDOR, pour la période comprise entre 1974 et 2004, on obtient un KGE de 0.9443, avec $r = 0.9463$, $\beta = 1.0020$ et $\alpha = 1.0146$.

L'optimisation des critères KGE ou NSE est effectuée sous des contraintes visant à maintenir l'interprétabilité physique des simulateurs (le taux de fonte de la neige le jour doit être supérieur au taux de fonte de la neige la nuit par exemple).

Ces modèles fonctionnent par itérations successives. Au jour j , pour simuler des débits fictifs à horizon de h jours (pour les jours $j + 1$ à $j + h$), on effectue les étapes suivantes :

1. Le SPQ prend tout d'abord en intrants les dernières observations météorologiques (celles du jour j). Chacune des variables d'état internes (taux de remplissage des réservoirs, quantité de neige stockée,...) est mise à jour. On obtient ainsi les variables d'état en début de prévision (celles du jour j).

2. Le SPQ prend ensuite en intrant la prévision météorologique émise pour le jour $j + 1$. Les variables d'état internes sont mises à jour. Un débit de rivière est simulé pour le jour $j + 1$.
 3. Le SPQ prend ensuite en intrant la prévision météorologique du jour suivant ($j + 2$), met à jours les variables d'état et simule un débit pour le jour $j + 2$.
- On suit la même procédure pour chacun des jours de la prévision qui suivent ($j + 3$ à $j + h$).

Nous avons donné, dans cette première partie de chapitre, un aperçu de l'ensemble des informations à disposition des prévisionnistes opérationnels pour émettre des prévisions hydrologiques. Afin de comprendre la logique des systèmes de prévision, il nous paraît important de bien saisir, de plus, les objectifs visés, c'est-à-dire les critères de performance des prévisions (hydrométéorologiques) probabilistes.

1.3 Critères d'évaluation des prévisions hydrométéorologiques probabilistes

Une prévision probabiliste représente les fluctuations attendues d'une quantité d'intérêt par une loi de probabilité. Comment juger alors d'un "bon" système de prévision probabiliste ?

Suivant Gneiting & Raftery (2007), un grand nombre de praticiens considèrent que l'objectif de tels systèmes est de maximiser la précision des prévisions émises sous contrainte d'une garantie de fiabilité. Nous nous intéressons au concept de fiabilité et de précision respectivement dans la section 1.3.1. et dans la section 1.3.2. Nous nous intéressons ensuite aux scores strictement justes, des fonctions permettant d'évaluer conjointement fiabilité et précision, dans la section 1.3.3.

1.3.1 La fiabilité, première attente vis-à-vis des prévisions probabilistes

Gneiting et al. (2007) définissent la *fiabilité* (ou *calibration*) d'une prévision probabiliste comme la cohérence statistique entre la prévision probabiliste et la quantité d'intérêt observée. Lorsque l'on évalue si un système de prévision est fiable, on répond aux questions suivantes : "Peut-on supposer que les observations sont issues des distributions prédites ?", "Les observations sont-elles indiscernables de simulations selon les distributions prédites ?".

Plaçons-nous pour commencer dans la situation où la prévision porte sur un événement binaire, tel que l'occurrence de précipitations. Pour T prévisions émises, soit $\rho_T(p)$ la proportion de jours où l'événement à prévoir s'est produit parmi les jours où le système de prévision avait annoncé une probabilité p pour son occurrence. Dawid (1982), repris par DeGroot & Fienberg (1983), définit la bonne calibration (ou parfaite fiabilité) d'un prévisionniste (ou système de prévision) comme la propriété

suivante : $\lim_{T \rightarrow \infty} \rho_T(p) = p$ quel que soit p de mesure non nulle. Ainsi, si l'on met de côté le fait que $\{p = 80\%\}$ puisse être de mesure nulle, l'assertion "Il pleut 80% des jours où l'on a annoncé une probabilité de 80% pour l'événement "pluie" " signifie que le système de prévision est parfaitement fiable au niveau $p = 80\%$.

Il est conceptuellement aisé de passer du cas binaire au cas où la grandeur à prévoir est catégorielle à $r \geq 2$ modalités : la distribution F se présente alors sous la forme d'un point à r coordonnées, $p = (p_1, p_2, \dots, p_r)$, sur le simplexe $\{p_i \geq 0, \sum_{i=1}^r p_i = 1\}$. Dans ces contextes discrets, DeGroot & Fienberg (1981), repris par Bröcker (2009), ont souligné qu'un prévisionniste était bien calibré si chacune de ses prévisions probabilistes pouvait s'interpréter comme une loi conditionnelle de la quantité d'intérêt.

Gneiting et al. (2007) ont avancé plusieurs extensions possibles de la définition de la bonne calibration aux cas où la grandeur à prévoir est à valeurs réelles à une dimension. Leur définition de la *calibration probabiliste* de la séquence des prévisions probabilistes $(F_t)_{t=1,2,\dots}$ vis à vis de (l'hypothétique) séquence $(G_t)_{t=1,2,\dots}$ de (véritables) fonctions de répartition des quantités à prévoir est la suivante :

$$\forall p \in [0, 1], \quad \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T G_t(F_t^{-1}(p)) = p. \quad (1.1)$$

Cette définition étend directement celle adoptée pour la calibration dans le cas binaire.

Une des propriétés de la collection $(G_t(y_t))_{t=1,2,\dots}$ est qu'elle suit une distribution uniforme sur l'intervalle $[0, 1]$. Gneiting et al. (2007) prouvent que la propriété de calibration probabiliste est équivalente au caractère uniforme de la collection $(F_t(y_t))_{t=1,2,\dots}$.

Un bon moyen statistique pour visualiser la fiabilité (au sens de calibration probabiliste) des systèmes de prévision probabilistes consiste donc à visualiser l'histogramme de $(F_t(y_t))_{t=1,2,\dots}$ (Probability Intergral Transform Histogram, PITH). La p-valeur d'un test d'ajustement de $(F_t(y_t))_{t=1,2,\dots}$ à la loi $\mathcal{U}[0, 1]$ permet alors de quantifier la fiabilité du système de prévision ayant émis $(F_t(y_t))_{t=1,2,\dots}$. Un biais dans la prévision se traduira par une pente dans l'histogramme, une sous-dispersion par une forme en U, une sur-dispersion par une forme en cloche. La figure 1.9 montre le PITH pour 1826 simulations, dans le cas où $Y \sim \mathcal{N}(0, 1)$ et F désigne la fonction de répartition de la loi (1) $\mathcal{N}(0, 1)$ (prévisions fiables), (2) $\mathcal{N}(0, 0.85)$ (prévisions sous-dispersives), (3) $\mathcal{N}(0, 1.15)$ (prévisions sur-dispersives), et (4) $\mathcal{N}(0.1, 1)$ (prévisions biaisées positivement). On remarque qu'un seuil de 0.01 sur la p-valeur du test de Kolmogorov-Smirnov permet dans ce cas le rejet de l'hypothèse de fiabilité de façon appropriée.

Les prévisions d'ensemble décrites dans la section 1.2.2 (CEPMMT-EPS, ANA, CMC-EPS et NCEP-GEFS) peuvent être considérées comme étant, à chaque temps de prévision, des simulations selon une distribution prédictive. De même, il est possible de supposer les scénarios annuels passés comme représentatifs du futur, auquel cas ils forment un ensemble de prévisions dans lequel chaque membre est un scénario passé pour former ce que l'on nomme ici les prévisions climatologiques.

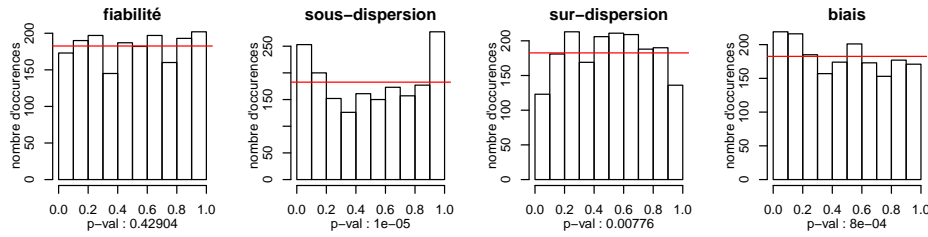


FIGURE 1.9 – PITH et p-valeurs du test d’uniformité de Kolmogorov-Smirnov associé pour $Y \sim \mathcal{N}(0, 1)$ et des prévisions (de gauche à droite) fiables, sous-dispersives $\mathcal{N}(0, 0.85)$, sur-dispersives $\mathcal{N}(0, 1.15)$ et biaisées $\mathcal{N}(0.1, 1)$, pour 1826 simulations.

Dans ces cas, on peut approcher le PITH par l’histogramme du rang de la quantité à prévoir parmi les membres de l’ensemble (*Rank Histogram*, RH, ou *diagramme de Talagrand*) (Anderson, 1996; Talagrand et al., 1997; Hamill & Colucci, 1997). Les deux premières colonnes de la figure 1.10 montrent des exemples de RH obtenus à partir des prévisions d’ensemble (CEPMMT-EPS) de températures maximales et à partir des prévisions climatologiques correspondantes, de 2010 à 2014, pour les échéances de prévision 1 et 9 jours, sur le bassin versant québécois de Manic 2. Pour un jour de l’année donné, les prévisions climatologiques sont obtenues en considérant les observations en ce même jour de 1960 à 2009 comme un ensemble de scénarios possibles. On note une forte sous-dispersion des prévisions d’ensemble, plus marquée à la première échéance de prévision qu’à la dernière. Cette observation est valable quel que soit l’ensemble considéré et quelle que soit la variable considérée. On note de plus une faible sous-dispersion et un faible biais négatif des prévisions climatologiques associées. Ils sont probablement dus au changement climatique illustré sur la figure 1.5. Les p-valeurs associées sont toutes inférieures au seuil classique de 0.01.

Dans les situations de prévisions ici décrites, l’aspect multidimensionnel des prévisions complique encore l’évaluation de la fiabilité. En effet, les prévisions ne portent pas seulement sur un événement mais sur un scénario spatial et temporel : il est attendu des scénarios des prévisions d’ensemble qu’ils soient à la fois probabilistiquement cohérents dans le temps et dans l’espace. La cohérence temporelle est particulièrement importante dans le cas des précipitations : c’est le cumul sur plusieurs jours de ces précipitations qui sera informatif pour la prévision hydrologique. Une première façon de tenir compte de cet aspect multidimensionnel est de procéder aux évaluations unidimensionnelles usuelles (telles que le PITH ou le RH), mais sur les cumuls des quantités à prévoir dans l’espace et dans le temps. La quatrième colonne de la figure 1.10 montre les RH obtenus de cette façon.

Il est également possible de généraliser les tests de calibration dans les cas multivariés. Plusieurs extensions du RH pour les prévisions multidimensionnelles ont été proposées à cet effet. Nous en présentons deux, le *Multivariate RH* (MRH) (Gneiting et al., 2008) et l’*Average RH* (ARH) (Thorarinsdottir et al., 2014).

Notons les membres X_1, \dots, X_K , où K désigne le nombre de membres et notons

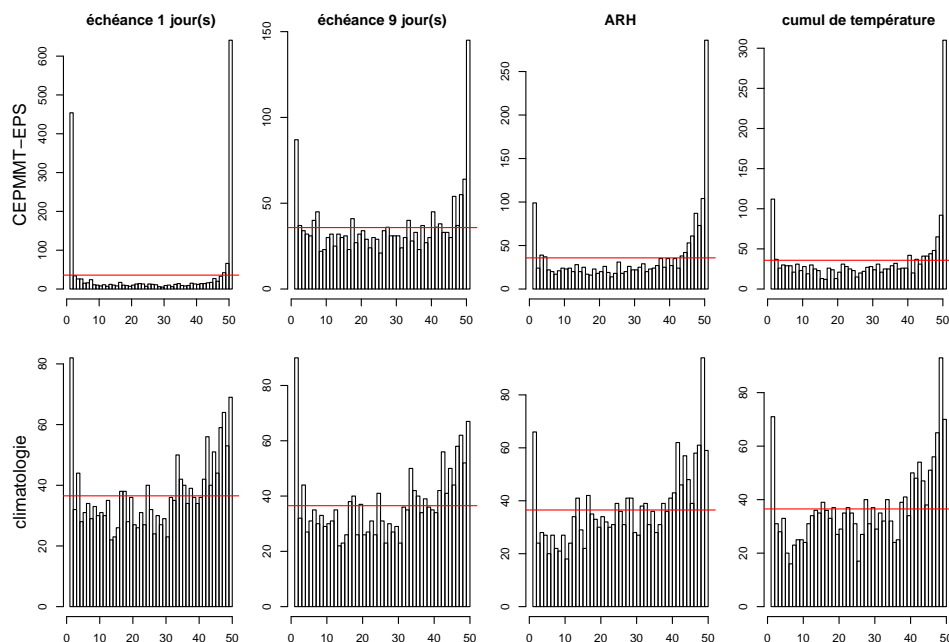


FIGURE 1.10 – RH des prévisions CEPMMT-EPS et des prévisions climatologiques de températures maximales journalières sur le bassin versant de Manic 2, de 2010 à 2014. De gauche à droite : à l'échéance 1, à l'échéance 9, ARH sur les 9 échéances et RH sur les prévisions de températures maximales cumulées au cours des 9 échéances de prévision. La ligne rouge correspond à l'histogramme théorique de prévisions calibrées.

$X_0 = Y$ l'observation. En dimension M , chaque X_k , $k \in (0, \dots, K)$ est un vecteur de dimension M : $X_k = (X_{k,1}, \dots, X_{k,M})^T$. Dans le MRH (Gneiting et al., 2008), on attribue à la quantité à prévoir, X_0 , le rang de son pré-rang. Le *pré-rang* ρ_k du membre k correspond au nombre de membres (observation incluse) dont l'ensemble des composantes sont inférieures ou égales aux composantes correspondantes du membre considéré : $\rho_k = \sum_{k'=0}^{k=K} \mathbb{I}_{\{\forall m \in 1, \dots, M, x_{k,m} \geq x_{k',m}\}}$. Dans le cas bi-dimensionnel, si l'on représente les membres selon deux axes correspondant à leurs deux composantes, le pré-rang d'un membre correspond au nombre de membres dans le carré inférieur gauche formé par la projection de ce membre sur les deux axes (en l'incluant). En dimension plus grande, le nombre de pré-rangs *ex-aequo* devient très grand, d'où une insensibilité du MRH à un éventuel manque de fiabilité. L'ARH ne présente pas cet inconvénient.

Dans l'ARH (Thorarinsdottir et al., 2014), le pré-rang de l'observation est donné par la moyenne de ses rangs unidimensionnels : $\rho_k = \frac{1}{M} \sum_{m=1}^{m=M} \text{rang}_k(x_{\cdot,m})$. Son interprétation est la même que celle des RH. La troisième colonne de la figure 1.10 montre les ARH obtenus avec les prévisions de températures maximales au cours des 9 échéances de prévision. Notons que ces histogrammes sont très similaires aux

RH obtenus pour les cumuls de températures maximales au cours de ces 9 échéances de prévision (illustrés en quatrième colonne de la figure).

En situation stationnaire, les prévisions climatologiques sont fiables par construction, puisqu'elles respectent la loi historique (et donc marginale) de la quantité d'intérêt. On conçoit cependant aisément qu'elles portent une information plus pauvre que les prévisions d'ensemble des centres climatiques pour la prévision des grandeurs météorologiques, même si ces dernières sont parfois peu fiables. La qualité d'une prévision probabiliste ne peut donc pas être évaluée par le seul critère de fiabilité. D'où un critère de performance complémentaire pour les prévisions probabilistes : leur précision.

1.3.2 La précision, second critère de performance des prévisions probabilistes

La *précision* (ou *finesse*) d'une prévision probabiliste correspond à sa dispersion. La moyenne temporelle des variances des prévisions probabilistes en donne par exemple une idée. Selon ce critère, les prévisions météorologiques d'ensemble, bien que peu fiables aux premières échéances, sont beaucoup plus fines que les prévisions climatologiques.

Perreault (2013) suggèrent, pour évaluer la précision d'une prévision, de s'appuyer sur la largeur des intervalles nécessaires à procurer un taux de couverture fixé.

Ce rapport reprend l'exemple introductif de Gneiting et al. (2007) qui considèrent quatre prévisionnistes, lesquels exploitent très différemment la même information X (supposée suivre une loi $\mathcal{N}(0, 1)$) portant sur une quantité à prévoir, Y . Pour se fixer les idées, x_t peut par exemple représenter une prévision déterministe de la température émise par un centre de prévision au niveau d'un village pour le jour t . La quantité cible peut par exemple être la température qui sera observée à une station météorologique proche du village au jour t , y_t . Dans cet exemple, pour chaque t , y_t est générée suivant une distribution gaussienne centrée sur la prévision x_t et de variance 1 : $\mathcal{N}(x_t, 1)$. Quand on consultera les fluctuations temporelles de la collection y_t , elle s'ajustera donc à une loi gaussienne $\mathcal{N}(0, 2)$ (la climatologie ou loi marginale de la température à la station). Nous considérons ici trois des quatre prévisionnistes pris en exemple par Gneiting et al. (2007) puis repris et discutés par Perreault (2013) :

- Le premier prévisionniste, dit “parfait”, émet la prévision probabiliste $F_t^1 = G_t$, où G_t est la fonction de répartition de la loi $\mathcal{N}(x_t, 1)$.
- Le deuxième prévisionniste utilise uniquement la climatologie et annonce $F_t^2(y) = [Y_t \leq y]$, la fonction de répartition de la loi marginale, $\mathcal{N}(0, 2)$.
- Le troisième prévisionniste, dit “hésitant”, prévoit un mélange à 50% entre une loi $\mathcal{N}(x_t, 1)$ et une loi $\mathcal{N}(x_t + \mu_t, 1)$, où μ_t prend la valeur -1 dans 50% des cas et 1 le reste du temps. On peut imaginer que ce prévisionniste manque de confiance vis-à-vis de l'information x_t et considère chaque jour t qu'elle a une probabilité de 50% d'être biaisée de μ_t et une probabilité de 50% de ne pas être biaisée.

Ces trois prévisionnistes se caractérisent par un même PITH uniforme, mais l'ordre $F^1 > F^3 > F^2$ est intuitivement le plus représentatif de la bonne exploitation de l'information. Il s'agit d'ailleurs aussi de celui donné par la moyenne intertemporelle des variances : $1 < 1.5 < 2$.

Cet ordre est aussi celui que l'on trouve en traçant les *diagrammes de fiabilité des intervalles de crédibilité*, un outil diagnostique de la performance des prévisions proposé par Perreault (2013). Ces courbes donnent le taux de couverture (de y) des intervalles de prévision en fonction de leur largeur pour différents niveaux de crédibilité. Ces taux de couverture empiriques sont donnés sur la figure 1.11, adaptée de Perreault (2013), pour les trois exemples de prévisionnistes décrits ci-dessus ainsi que pour un prévisionniste supplémentaire omniscient (qualifié d'“idéal” sur la figure). Ces taux sont obtenus sur une base de 3000 prévisions. Sur cette figure, les trois pré-

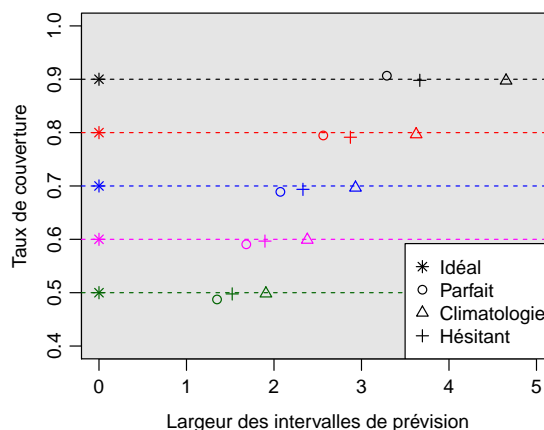


FIGURE 1.11 – Diagrammes de fiabilité des intervalles de crédibilité aux niveaux 50%, 60%, 70%, 80% et 90% pour trois prévisionnistes décrits dans Gneiting et al. (2007) et pour un prévisionniste “idéal” omniscient, sur une base de 3000 prévisions.

visionnistes décrits ci-haut apparaissent comme calibrés (les taux de couverture de leurs intervalles de prévision sont adéquats). En revanche, leur finesse, évaluée par la largeur des intervalles de prévision, dépend de la façon dont ils utilisent l'information portée par x .

Nous pouvons interpréter ces observations de la façon suivante : conditionnellement à la fiabilité, maximiser la finesse revient à trouver et exploiter au mieux l'information à disposition.

1.3.3 Des scores justes pour évaluer fiabilité et précision des prévisions probabilistes

Il faut, pour comparer la qualité de prévisions probabilistes, à la fois évaluer leur fiabilité et leur précision. Nous avons présenté dans la section précédente un moyen

de le faire de façon qualitative (en visualisant les diagrammes de fiabilité des intervalles de crédibilité). Une façon d'évaluer cette performance de façon quantitative est d'avoir recours aux scores justes.

Soit \mathcal{P} une classe convexe de probabilités sur (Ω, \mathcal{A}) , où Ω est un ensemble et \mathcal{A} est une tribu sur Ω . Un *score* sur \mathcal{P} est une fonction s qui attribue au couple (F, y) , où $F \in \mathcal{P}$ et $y \in \Omega$, une valeur dans \mathbb{R} . Notons $S(F, G) = \mathbb{E}_{Y \sim G} [s(F, Y)] = \int_{y \in \Omega} s(F, y) dG(y)$.

Un score *orienté négativement* (c'est-à-dire à minimiser) est *juste* par rapport à \mathcal{P} si et seulement si :

$$\forall (F, G) \in \mathcal{P}^2, \quad S(G, G) \leq S(F, G).$$

Il est *strictement juste* si l'égalité n'est vraie que pour $F = G$. Un prévisionniste évalué avec un tel score est incité à dévoiler ce qu'il pense être la véritable fonction de répartition de Y (ce qu'il pense être G) (Savage, 1971).

Le lien entre scores strictement justes, fiabilité et précision est bien établi : dans le cas où la variable à prévoir, Y , peut prendre un nombre fini de valeurs, on montre que tout score strictement juste peut être décomposé en un terme de fiabilité et un terme de précision (Bröcker, 2009).

En pratique, on évalue $S(F, G)$ avec le *Score moyen empirique* :

$$\hat{S}(F, G) = \frac{1}{\#(testset)} \sum_{t \in testset} s(F_t, y_t),$$

où *testset* désigne la période d'évaluation des prévisions et $\#(testset)$ est le nombre d'éléments de *testset*.

Le tableau 1.7 donne quelques exemples de scores. Un score s est dit *local* si $s(F, y)$ ne dépend que de $f(y)$. Il est dit *global* si il est aussi sensible à la probabilité annoncée pour les valeurs qui n'ont pas été observées.

TABLEAU 1.7 – Exemples de fonctions scores

Score	$S(F, y)$	Propriétés
Score linéaire	$-f(y)$	local, non juste
IS	$-\log(f(y))$	local, strictement juste*
CRPS	$\int_{-\infty}^{+\infty} (F(z) - \mathbb{I}_{\{z \geq y\}})^2 dz$	global, juste**

* par rapport à la classe des probabilités absolument continues par rapport à une σ -mesure finie de référence et ayant un premier moment fini.

** par rapport à la classe des probabilités boréliennes, \mathcal{P} , et strictement juste par rapport à la sous-classe de \mathcal{P} ayant un premier moment fini.

Tous les scores locaux et strictement justes sont nécessairement une fonction affine du score ignorance (IS) (Bernardo, 1979). Notons que les méthodes d'inférence statistique fondées sur la maximisation de la vraisemblance minimisent par la même occasion le score ignorance. Ce score présente l'inconvénient de ne pas être borné. Une prévision ponctuelle particulièrement mauvaise peut donc conduire à un score éliminatoire et la fiabilité a un poids très fort dans l'évaluation des systèmes de prévision par ce score.

Outre être global et juste, le Continuous Rank Probability Score (CRPS) a pour avantage de correspondre à une généralisation de l'erreur en valeur absolue (il s'exprime donc dans l'unité de la grandeur à prévoir) et de permettre ainsi la comparaison de prévisions déterministes et probabilistes. Il peut de plus être décomposé en un terme de fiabilité, un terme de précision et un terme d'incertitude (le terme d'incertitude correspondant au score atteint lorsque la prévision est parfaite) (Hersbach, 2000). Ces éléments en font le score le plus usité dans le domaines de la météorologie. En pratique, nous l'estimons dans nos travaux en approchant numériquement l'intégrale donnée dans le tableau 1.7 (à l'aide de la fonction `density` de R). Chatelain et al. (2015) repris par Perreault et al. (2017) ont montré sur plusieurs exemples que cette méthode d'estimation était généralement adéquate et qu'elle donnait des résultats similaires (en matière d'adéquation de l'estimation) à l'estimation fondée sur les *U-statistics* proposée par Naveau et al. (2014).

Pour le cas multidimensionnel, une généralisation du CRPS a été proposée par Gneiting & Raftery (2007) : l'*energy score* (ES). Ce score est défini par :

$$ES(F, y) = \frac{1}{2} \mathbb{E}_F \|X - X'\|^\beta - \mathbb{E}_F \|X - y\|^\beta, \quad (1.2)$$

où $\|\cdot\|$ désigne la norme euclidienne, X et X' ont pour fonction de répartition F et $\beta \in]0, 2[$. Il s'agit d'un score juste et, pour $\beta = 1$ et en dimension 1, on retrouve la définition du score CRPS (Gneiting & Raftery, 2007). Pinson & Tastu (2013) montrent que, dans le cas gaussien multidimensionnel, ce score permet bien d'évaluer le biais, mais que sa sensibilité à la sous-dispersion, à la sur-dispersion et à des défauts de modélisation de la dépendance reste limitée.

1.3.4 Évaluation des prévisions hydrologiques chez les producteurs d'hydro-électricité

La capacité des scores strictement justes à évaluer conjointement les deux facteurs de qualité reconnus des prévisions probabilistes (leur fiabilité et leur précision) en fait un moyen d'évaluation couramment utilisé par les producteurs de prévisions et leurs destinataires. Le score CRPS, par exemple, est utilisé pour l'évaluation des prévisions chez Hydro-Québec (Perreault, 2013) et chez EDF-DTG.

Chez les producteurs d'hydro-électricité, les prévisions hydrologiques sont par ailleurs évaluées de façon qualitative par les gestionnaires, ou bien de façon quantitative par la valeur économique des prévisions émises (Perreault et al., 2017).

1.4 Définition avancée des objectifs du travail doctoral

Dans la suite de ce mémoire, nous cherchons à apporter des contributions statistiques aux systèmes de prévision hydrologiques à moyen-terme, en nous focalisant sur ceux d'EDF-DTG et d'Hydro-Québec.

Soit H l'horizon de prévision auquel on travaille. Au jour t_0 , la cible des systèmes de prévision ici étudiés est une prévision probabiliste performante pour les débits qui seront observés au cours des H prochains jours, $Q_{t_0+(1,\dots,H)}$.

Les services de prévision disposent aujourd'hui d'une variété d'informations pour émettre ces prévisions. La section 1.2 en a donné un aperçu. Parmi elles, on trouve des prévisions particulièrement riches d'informations, transmises par les centres de prévision météorologiques sous la forme de prévisions pour les H jours à venir, $A_{t_0}^{1,\dots,H}$. On trouve aussi, parmi les informations à disposition, des observations hydrométéorologiques passées, notées Q_{t_0-} pour les débits et P_{t_0-} pour la météorologie, ainsi que des informations plus auxiliaires, telles que les variables d'état des SPQ.

Dans la section 1.3, nous avons étudié les propriétés faisant des systèmes de prévision probabilistes des systèmes performants : la fiabilité (ou calibration) et la précision (ou finesse).

Émettre des prévisions probabilistes fiables demande un travail de modélisation important afin de proposer des distributions prédictives pouvant s'interpréter comme celles de la quantité à prévoir conditionnellement aux quantités connues.

En l'occurrence, on cherche à construire des modèles statistiques dont on pourra dériver une distribution pouvant être interprétée comme la distribution suivante (ou, du moins, un moyen de l'approcher) :

$$\left[Q_{t_0+(1:H)} \mid A_{t_0}^{1:H}, P_{t_0-}, Q_{t_0-} \right]. \quad (1.3)$$

Afin de répondre aux objectifs de finesse, il nous faudra intégrer au mieux les informations disponibles pour la prévision, en particulier les informations portées par les membres des ensembles météorologiques, celles portées par les prévisions alternatives émises par les centres de prévision et les autres informations additionnelles utilisées par les prévisionnistes.

Une fois des modèles statistiques adaptés construits, l'inférence pourra être faite de sorte à minimiser un score strictement juste, donc de façon à en maximiser la performance. Nous utiliserons pour notre part des méthodes d'inférence de la statistique classique qui visent à estimer les paramètres du modèle statistique de telle sorte qu'ils maximisent la vraisemblance, ce qui revient à minimiser le score ignorance (strictement juste).

2

État de l'art des systèmes de prévision hydrométéorologique

2.1	Prédominance d'une architecture séparant les sources d'incertitudes météorologique et hydrologique	39
2.2	Méthodes statistiques pour le post-traitement des prévisions	41
2.2.1	Réduction de dimension avec les copules	41
2.2.2	Retrouver un cadre gaussien	44
2.2.3	Choix méthodologiques et inférence	45
2.3	Vue d'ensemble des méthodes de post-traitement météorologique	47
2.4	Vue d'ensemble des méthodes de prise en compte de l'incertitude hydrologique	55

Ce chapitre présente un état de l'art des systèmes de prévision hydrologique à moyen-terme. Un article de revue soumis pour publication (en Annexe) reprend et développe les sections 2.1 et 2.2 de ce chapitre. Notre réflexion s'appuie sur des exemples de systèmes de prévision opérationnels ou bien développés avec une visée opérationnelle (résumés dans le tableau 2.1), notamment ceux d'EDF et d'Hydro-Québec dont nous avons pu observer le fonctionnement *in situ*.

Systèmes de prévision hydrologique (à moyen-terme)		SCHEF, BoM (Australie) (en développement) [Bennett et al., 2014]	HEFS, NOAA (EU) (en développement) [Demargne et al., 2014]	EDF-DTG (France) (opérationnel) [Mathevet et al., 2009] [Gailhard, 2011]	Hydro-Québec (Canada) (opérationnel)	CEMIG (Brésil) (en développement) [Schwanenberg et al., 2015] [Fan et al., 2014]	
Intrants : prévisions météorologiques A		Prévisions déterministes de précipitations <i>global ACCESS-G</i> du <i>Numerical Weather Prediction</i> (NWP)	Prévisions météorologiques déterministes jusqu'à l'horizon 7 jours, puis moyenne des prévisions d'ensemble de plusieurs modèles du NWP	Prévisions d'ensemble CEPMMT-EPS	Prévisions déterministes d'Environnement Canada	Prévisions d'ensemble CEPMMT-EPS	
Post-traitement météorologique \sim [P]A	Normalisation	Transformation <i>Log-sinh</i> [Wang et al., 2012]	<i>Meteorological Ensemble Forecast Processor</i> (MEFP) [Wu et al., 2011, Schaake et al., 2007]	<i>Normal Quantile Transform</i> (NQT) [Krzysztofowicz, 1997]	∅	∅	
	Prise en compte de l'incertitude	Modèle Normal bidimensionnel		Modèle Normal bidimensionnel de type mixte pour les précipitations [Herr, 2005], autre modèle joint pour les températures	Méthode à noyaux inspirée de [Obled et al., 2002]	Noyau Normal pour les températures et Pearson type III pour les précipitations totales	Correction déterministe
	Structure multidimensionnelle	Copule empirique <i>Schaake shuffle</i> [Clark et al., 2004]		Copule empirique <i>Schaake shuffle</i> [Clark et al., 2004]	Copule empirique similaire à l' <i>Ensemble Copula Coupling</i> (ECC) de [Scheffzik et al., 2013]	Scénarios-quantiles	
P							
Simulateur pluie-débit $\mathcal{M}(P)$		GR4H [Perrin et al., 2003]	Snow-17 [Anderson, 1973], SAC-SMA [Burnash, 1995] et Unit HG	MORDOR [Garçon, 1996]	HSAMI [Fortin, 2000]	MGB-IPH [Collischonn et al. 2005] et <i>mini-basins</i> [Paiva et al., 2013]	
Post-traitement hydrologique \sim [Q]M(P)	Prise en compte de derniers débits observés	Correction déterministe : version modifiée de <i>dual-pass error correction model</i> de [Pagano et al., 2011]	Modification manuelle des paramètres et des variables d'état du SPQ	[Gailhard, 2014] Correction déterministe	Assimilation de données sur les variables d'état du SPQ	Assimilation de données sur deux variables d'état du SPQ [Paz et al., 2007]	
	Normalisation	∅	NQT	Transformation logarithmique [Mathevet, 2010]	∅	∅	
	Prise en compte de l'incertitude	∅	Modèle auto-régressif d'ordre 1 à variable auxiliaire : <i>Hydrologic Ensemble Postprocessor</i> (EnsPost) [Seo et al., 2006]	Modèle log-normal à variance hétéroscédastique s'inspirant de [Montanari and Brath, 2004][Schaefli et al., 2007]	Méthode d'ensemble : perturbation des variables d'état du SPQ	∅	
Structure multidimensionnelle	∅	Copule empirique similaire à l'ECC de [Scheffzik et al., 2013]		∅	∅		
Q							

TABLEAU 2.1 – Exemples de systèmes de prévision hydrologique à moyen-terme à visée opérationnelle. De haut en bas : des intrants aux sorties du système opérationnel. De gauche à droite : systèmes SCHEF (System for Continuous Hydrological Ensemble Forecasting, en développement au Bureau of Meteorology, en Australie), HEFS (Hydrologic Ensemble Forecast Service, en développement ou bien opérationnel depuis peu à la National Oceanic and Atmospheric Administration, Etats-Unis) et les systèmes opérationnels d'EDF, d'Hydro-Québec et de la CEMIG (Companhia Energética de Minas Gerais).

2.1 Prédominance d'une architecture séparant les sources d'incertitudes météorologique et hydrologique

Il existe une organisation commune remarquable des cinq principaux systèmes de prévision hydrologique pris en exemples dans cet état de l'art, listés dans les cinq colonnes du tableau 2.1 : l'incertitude émanant de la connaissance imparfaite de la météorologie future et celle émanant de la connaissance imparfaite des phénomènes hydrologiques sont traitées séparément. Cette organisation commune découle de deux points communs à l'ensemble de ces systèmes :

- Ils prennent en intrants des anticipations météorologiques, A (pour Anticipations), visant à prévoir la météorologie (généralement des précipitations et des températures) qui sera observée, P (pour Précipitations) ;
- Ils font intervenir un simulateur pluie-débit (SPQ) opérant une transformation, \mathcal{M} , calée de telle sorte qu'en prenant en intrant la météorologie observée sur une longue période, ses sorties, $\mathcal{M}(P)$, soient les plus proches possible des débits observés correspondants, Q .

Les prévisions météorologiques et le SPQ étant imparfaits, des opérations de post-traitements, l'un météorologique et l'autre hydrologique, y interviennent pour tenir compte de l'incertitude résiduelle respectivement vis-à-vis de P (lorsque l'on connaît A) et de Q (lorsque l'on connaît $\mathcal{M}(P)$).

Revenons à la distribution cible des systèmes de prévision au temps de prévision t_0 , établie dans le chapitre 1 à l'équation (1.3) :

$$\left[Q_{t_0+(1:H)} | A_{t_0}^{1:H}, P_{t_0}^-, Q_{t_0}^- \right].$$

Par souci de simplicité, plaçons-nous au temps t_0 et notons $\mathbf{Q} = Q_{t_0+(1:H)}$, $\mathbf{P} = P_{t_0+(1:H)}$ et $\mathbf{A} = A_{t_0}^{1:H}$, et omettons l'écriture de $Q_{t_0}^-$ et $P_{t_0}^-$ (qui interviennent dans tous les conditionnements).

Comme mis en évidence par Krzysztofowicz (1999), la chaîne de prévision ici décrite correspond à l'intégration de deux sources d'incertitudes, respectivement météorologique et hydrologique. On peut l'écrire formellement :

$$[\mathbf{Q}|\mathbf{A}] = \int [\mathbf{Q}|\mathcal{M}(\mathbf{P})] [\mathbf{P}|\mathbf{A}] d\mathbf{P}. \quad (2.1)$$

Parmi les hypothèses qui interviennent dans cette écriture, les deux principales sont les suivantes :

- Hypothèse d'exhaustivité de la météorologie observée (connaissant la météorologie, les anticipations ne sont plus informatives vis-à-vis de l'hydrologie) : $[\mathbf{Q}|\mathbf{A}, \mathbf{P}] = [\mathbf{Q}|\mathbf{P}]$;
- Hypothèse d'efficacité de la transformation pluie-débit (les SPQ produisent un résumé efficace des informations contenues dans la météorologie utiles à la connaissance des débits) : $[\mathbf{Q}|\mathbf{P}] = [\mathbf{Q}|\mathcal{M}(\mathbf{P})]$. En pratique, cette hypothèse peut être relâchée en utilisant des sorties auxiliaires du SPQ.

Ainsi, pour compléter la chaîne de prévision, deux modèles doivent être associés :

- Un modèle dont on dérivera $[\mathbf{P}|\mathbf{A}]$: le *post-traitement météorologique* évoqué ci-dessus ;
- Un modèle dont on dérivera $[\mathbf{Q}|\mathcal{M}(\mathbf{P})]$: le *post-traitement hydrologique* évoqué ci-dessus.

Afin d'obtenir la distribution prédictive cible, $[\mathbf{Q}|\mathbf{A}]$, il faut calculer l'intégrale de l'équation (2.1) sur l'espace des grandeurs météorologiques P , généralement de grande dimension (il s'agit de séries temporelles multidimensionnelles).

Plutôt que d'en faire le calcul analytique, on peut en approcher le résultat d'aussi près que l'on veut au travers de simulations. Il suffit d'utiliser des simulations issues de la distribution météorologique prédictive en intrant du SPQ. Chaque sortie du SPQ résultante est ensuite traitée comme si elle avait été obtenue avec un véritable scénario météorologique. On trouve cette procédure dans le système HEFS comme dans celui d'EDF. En pseudo-code, cela revient à opérer selon l'algorithme suivant :

Algorithme 1 : Algorithme (en simili-code) permettant d'approcher l'intégrale de l'équation (2.1).

Pour obtenir $\{q_1, \dots, q_K\} \sim [\mathbf{Q}|\mathbf{A}]$:

(1) Simuler $\{p_1, \dots, p_K\}$ selon $[\mathbf{P}|\mathbf{A}]$.

(2) Utiliser chaque p_k en intrant du SPQ pour obtenir

$m_1 = \mathcal{M}(p_1), \dots, m_K = \mathcal{M}(p_K)$.

(3) Pour chaque $m_k, k \in \{1, \dots, K\}$, simuler q_k selon $[\mathbf{Q}|\mathcal{M}(\mathbf{P}) = m_k]$.

Une alternative à cette structure en deux étapes des systèmes de prévision consiste à écrire un modèle liant Q à $\mathcal{M}(A)$ (sans chercher à passer par P) (Zalachori et al., 2012; Roulin & Vannitsem, 2015; Boucher et al., 2015).

Toutefois, la chaîne de prévision en deux étapes décrite ci-dessus présente des avantages qui expliquent l'ubiquité de cette façon de procéder :

- On dispose d'observations météorologiques (c'est-à-dire de P). Une chaîne de prévision en deux étapes permet de mieux les mettre à contribution pour estimer les paramètres des modèles que si on négligeait cette information.
- Un tel système fournit une distribution prédictive pour la météorologie future conditionnellement aux prévisions émises, $[\mathbf{P}|\mathbf{A}]$. Ce sous-produit est potentiellement utile à d'autres fins telles que la prévision probabiliste de la consommation électrique.
- Dans ces systèmes, il est possible aux prévisionnistes météorologues d'intervenir sur la prévision météorologique et de la rectifier en y intégrant des informations supplémentaires et sans avoir à corriger les biais systématiques des prévisions des simulateurs atmosphériques (car ils sont corrigés au préalable par le post-traitement météorologique).
- Les paramètres des SPQ sont généralement estimés de telle sorte que $\mathcal{M}(P)$ soit le plus proche possible de Q (selon des critères de type moindres carrés

- voir la section 1.2.3). Pour écrire un modèle liant les températures et les précipitations, réunies dans P , et les débits, Q , il est donc naturel de passer par la transformation \mathcal{M} . Ce n'est pas le cas quand il s'agit de lier les anticipations météorologiques A et les débits Q : utiliser A en intrant du SPQ risquerait d'en dégrader fortement les performances de façon difficilement prévisible. Pensons par exemple au cas où les prévisions de températures comporteraient un biais positif. Dans une situation où la température réelle sera légèrement inférieure zéro, le SPQ pourrait mimer un écoulement dû à la fonte d'un éventuel stock de neige. Ce type de répercussion non linéaire et avec changement de phase est difficile à prendre en compte dans un modèle statistique visant à anticiper les erreurs du SPQ.

Dans les sections qui suivent, nous donnons un aperçu des méthodes de prise en compte des incertitudes de la littérature ou opérationnelles. Nous commençons par mettre en relief les points communs à beaucoup d'entre elles dans la section 2.2.

2.2 Méthodes statistiques pour le post-traitement des prévisions ..

Un point commun à la plupart des méthodes de post-traitement des prévisions (météorologiques comme hydrologiques) intervenant dans les systèmes de prévision opérationnels est que l'on n'y modélise pas directement des variables de grande dimension. Cela devrait pourtant être le cas puisque l'on cherche à prévoir des phénomènes spatiaux (concernant plusieurs bassins versants voisins), temporels (pour plusieurs jours qui se suivent) et multi-variables (précipitations et températures dans la partie météorologique du problème).

La procédure statistique la plus employée pour construire un modèle multivarié simplement est d'employer des copules (Genest & Favre, 2007). Nous donnons une idée de cette construction dans la section 2.2.1. Une fois la dimension du problème réduite, on remarque que le modèle gaussien est omniprésent dans les systèmes de prévision ici considérés. Des méthodes utilisées dans le domaine pour se ramener à un cadre gaussien sont exposées dans la section 2.2.2. Enfin, l'approche de modélisation et la méthode d'inférence à utiliser font parfois débat. Quelques choix rencontrés dans la littérature ou *in situ* sont discutés dans la section 2.2.3.

2.2.1 Réduction de dimension avec les copules

D'après le théorème de représentation de Sklar (Sklar, 1959), si G est une fonction de répartition multivariée de dimension K , de marginales G_1, \dots, G_K , alors il existe (sous des conditions de régularité non restrictives) une fonction $C : [0, 1]^K \rightarrow [0, 1]$, appelée K -copule, telle que $G(y_1, \dots, y_K) = C(G_1(y_1), \dots, G_K(y_K))$ quel que soit le K -uplé $(y_1, \dots, y_K) \in \mathbb{R}^K$.

On peut ainsi, par exemple, supposer les variables à modéliser conjointement comme marginalement gaussiennes sans pour autant supposer une dépendance gaussienne entre ces variables, sortant ainsi du trop strict contexte gaussien où les es-

pérances conditionnelles sont nécessairement des fonctions linéaires des grandeurs conditionnantes et les variances conditionnelles n'en sont pas fonction.

En pratique, dans les systèmes de prévision hydrométéorologiques, les copules sont utilisées pour reconstituer des distributions prédictives multidimensionnelles (dans l'espace, dans le temps et entre variables) à partir de distributions prédictives unidimensionnelles ou de faibles dimensions (en un point donné de l'espace, pour un jour donné et pour une variable donnée par exemple).

Soient Y la variable à prévoir (le débit Q en un jour donné par exemple) et X le prédicteur (les sorties du SPQ, $\mathcal{M}(P)$, par exemple), avec Y et X de dimension K (où K correspond au nombre de bassins versants du domaine considéré par exemple). Dans les systèmes de prévision étudiés, on écrit typiquement un modèle liant chacun des Y_k , $k \in \{1, \dots, K\}$, au X_k correspondant et on émet à partir de ce modèle K distributions prédictives : $[Y_k|X_k]$, $k \in \{1, \dots, K\}$. À l'aide d'une copule appropriée, la distribution multidimensionnelle cible ($[Y_1, \dots, Y_K|X_1, \dots, X_K]$) peut être obtenue à partir des distributions prédictives unidimensionnelles déjà produites (en supposant que, conditionnellement à X_k , $X_{k'}$, $k' \neq k$ n'est pas informatif vis-à-vis de Y_k).

Möller et al. (2013), par exemple, utilisent un modèle de copule gaussienne pour proposer une distribution jointe de cinq variables météorologiques à partir de leurs distributions marginales obtenues en post-traitant des prévisions météorologiques. La copule gaussienne est un cas particulier de modèle de copule particulièrement simple à manipuler. Il existe un grand nombre d'autres modèles de copules, pour la plupart conçus pour les dimensions faibles. Pour une introduction à la modélisation à l'aide de copules, se référer à Nelsen (2007) et à Genest & Favre (2007).

Dans le domaine des prévisions hydrométéorologiques, il est le plus souvent fait appel à des méthodes de copules dites *empiriques*, non paramétriques.

Le *Schaake shuffle* (Clark et al., 2004), par exemple, consiste à transposer la structure de rang de scénarios passés (observés dans des situations climatiques similaires à la situation de prévision en météorologie) à un échantillon de simulations indépendantes selon les distributions prédictives unidimensionnelles. Cela revient à ordonner ces simulations de la même façon que les grandeurs étaient ordonnées dans les scénarios passés. Le *Schaake shuffle* est utilisé dans les systèmes SCHEF et HEFS par exemple.

Dans le cas où l'on cherche à émettre des prévisions météorologiques probabilistes à partir de prévisions d'ensemble, on peut, de façon analogue, transposer la structure de rang des membres de l'ensemble à des simulations selon les distributions marginales. Il s'agit alors de l'*Ensemble Copula Coupling* (ECC), une méthode dont Bremnes (2007) et Krzysztofowicz & Toth (2008) ont donné l'idée et formalisée par Schefzik et al. (2013). La méthode de copule empirique mise en place par EDF est similaire à l'ECC.

Schefzik (2015) proposent enfin d'identifier les situations passées similaires à la situation de prévision, dont on copiera la structure de rang, en comparant les prévisions émises par le passé aux prévisions du jour de prévision. Il s'agit alors de la méthode du *Sim-Schaake Shuffle*. Dans le cas où l'on se fonde sur des prévisions

météorologiques d'ensemble, on peut toutefois s'attendre à ce que l'historique de prévisions disponible soit trop limité pour obtenir un nombre intéressant de situations analogues par cette méthode.

Afin d'illustrer le principe des méthodes de copules empiriques, prenons par exemple une situation de prévision à laquelle nous avons trouvé les 3 situations analogues (sous la forme de scénarios passés ou bien de membres d'un ensemble de prévisions) illustrées sur le graphique du haut de la figure 2.1. Les numéros sur la figure sont le rang des scénarios à chaque pas de temps. Afin d'utiliser une copule empirique, nous simulons un échantillon de taille 3 (c'est-à-dire comprenant autant de simulations que nous avons identifié de situations analogues) selon nos lois marginales prédictives et obtenons le graphique en bas à gauche de la figure 2.1. L'application d'une copule empirique consiste à copier la structure de rang des scénarios analogues (la façon dont sont reliés les points de la figure du haut) pour obtenir des scénarios à partir des points, les 3 scénarios illustrés sur le graphique en bas à droite de la figure 2.1.

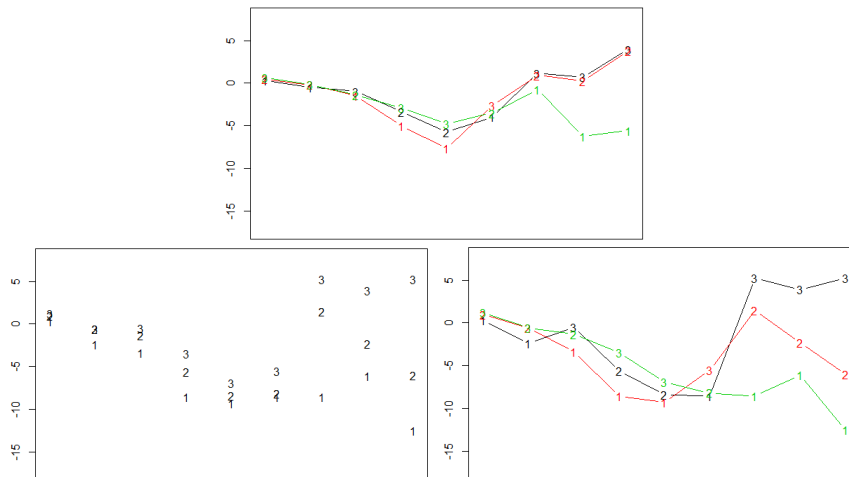


FIGURE 2.1 – En haut : situation de prévision dont on souhaite copier la structure. En bas à gauche : simulations selon les lois marginales à chaque pas de temps. En bas à droite : scénarios résultant de l'application d'une méthode de copule empirique aux lois marginales illustrées sur le graphique de gauche. Les numéros correspondent au rang des points à chaque pas de temps.

Un inconvénient de ces méthodes est que le nombre de simulations possibles est limité au nombre de scénarios analogues identifiés.

Afin de minimiser les conséquences de cette limitation, [Bremnes \(2007\)](#) et [Schefzik et al. \(2013\)](#) proposent d'appliquer une copule empirique à des quantiles équirépartis de la distribution prédictive, les quantiles $\frac{1}{K+1}, \frac{2}{K+1}, \dots, \frac{K}{K+1}$, plutôt qu'à K simulations selon cette distribution (où K désigne le nombre de situations analogues identifiées ou de membres de l'ensemble). Cette façon de procéder conserve l'uni-

formité des histogrammes de rang (Bröcker, 2012) tout en permettant de mieux garantir l'exploration du domaine des valeurs pouvant être prises par la variable à prévoir. Dans le cas où la méthode de copule empirique considérée est l'ECC, on parle alors d'ECC-Q (Scheffzik et al., 2013).

Il est aussi possible d'extraire, plutôt que des quantiles équirépartis, les quantiles correspondant aux situations analogues utilisées pour la copule, comme le proposent Krzysztofowicz & Toth (2008) et Scheffzik et al. (2013). Il faut pour cela proposer, pour chaque unité (par exemple pour chaque pas de temps de la prévision), une fonction de répartition, \hat{F} , telle que l'on puisse supposer les membres de l'ensemble analogue issus de la distribution de probabilité correspondante. L'échantillon auquel sera appliqué la copule empirique est obtenu en appliquant la fonction $F^{-1} \circ \hat{F}(\cdot)$ aux membres de l'ensemble d'analogues, où F^{-1} désigne l'inverse de la fonction de répartition prédictive. Dans le cas où la méthode de copule empirique considérée est l'ECC, on parle alors d'ECC-T (Scheffzik et al., 2013).

2.2.2 Retrouver un cadre gaussien

Comme on le remarque à la lecture du tableau 2.1, le modèle gaussien est omniprésent dans les systèmes de prévision considérés. Autant les variables relatives à la température de l'air peuvent éventuellement être supposées gaussiennes, autant ce n'est pas le cas des variables de type précipitations ou débits, comme l'illustre la figure 2.2 dans le cas des débits. Cela explique la présence d'une étape de normalisation parmi les chaînons des systèmes de prévision.

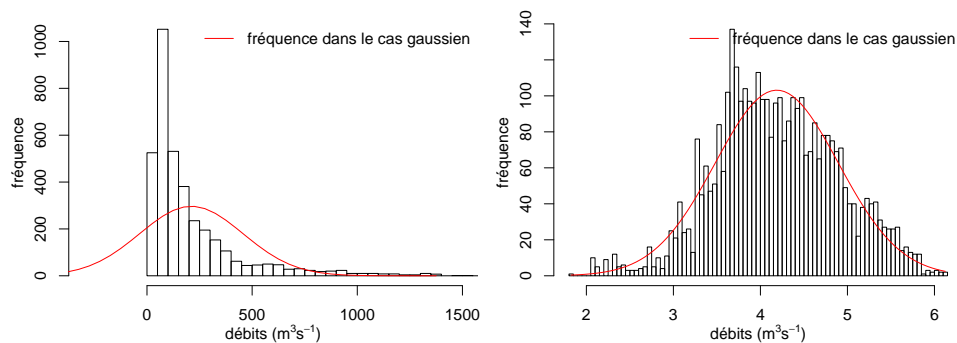


FIGURE 2.2 – Histogrammes des débits journaliers à l'exutoire du bassin versant de Baskatong (Québec) entre 1950 et 2013 avant (à gauche) et après (à droite) une transformation de Box-Cox. Les courbes rouges correspondent à l'histogramme de la loi gaussienne de même moyenne et variance.

Soit Y la variable à normaliser, G sa fonction de répartition marginale, $f_{\mathcal{N}}$ la transformation normalisante considérée, et Z la variable Y normalisée : $Z = f_{\mathcal{N}}(Y)$.

La transformation *Normal Quantile Transform* (NQT) est sans doute la plus générale. Elle consiste à appliquer à Y sa propre fonction de répartition marginale,

G , puis à y appliquer la fonction quantile de la loi normale centrée réduite, Φ^{-1} : $f_{\mathcal{N}}(Y) = \Phi^{-1}(G(Y))$. En pratique, on ne connaît pas G , on l'estime donc et on obtient \hat{G} . Si G est bien estimée par \hat{G} , $\hat{G}(Y)$ peut être supposé suivre une loi Uniforme sur l'intervalle $[0, 1]$. $\Phi^{-1}(\hat{G}(Y))$ suit alors avec une bonne approximation une loi normale centrée réduite. Cette transformation est par exemple utilisée pour normaliser des variables de type précipitations par Krzysztofowicz (2002) et dans HEFS, et de type débit à l'exutoire dans HEFS et par Krzysztofowicz (2002); Todini (2008) et Engeland et al. (2010).

Une transformation alternative est la transformation log-sinus hyperbolique (log-sinh), proposée par Wang et al. (2012). Elle s'écrit comme suit : $f_{\mathcal{N}}(Y) = \frac{\log\{\sinh(\alpha+\beta Y)\}}{\beta}$ où α et β sont des paramètres et $\sinh : x \mapsto \frac{e^x - e^{-x}}{2}$. Cette normalisation est utilisée pour les variables de type précipitations dans le système SCHEF.

Une normalisation courante pour les données de débit est la transformation logarithmique : $f_{\mathcal{N}}(Y) = \log(Y)$. Elle est par exemple utilisée dans le système d'EDF et par Collet et al. (2009) et Morawietz et al. (2011)). Il s'agit d'un cas particulier de la transformation de Box & Cox (1964), recommandée et appliquée par de nombreux auteurs dont Sorooshian & Dracup (1980); Kuczera (1983); Bates & Campbell (2001); Duan et al. (2007); Vrugt & Robinson (2007); Engeland et al. (2010); Hemri et al. (2013) et Hemri et al. (2015). Cette transformation s'exprime comme suit :

$$\begin{cases} f_{\mathcal{N}}(Y) = \log(Y) & \text{si } \lambda = 0 \\ f_{\mathcal{N}}(Y) = \frac{Y^\lambda - 1}{\lambda} & \text{sinon,} \end{cases}$$

où λ est un paramètre à estimer. Le résultat de l'application de la transformation de Box-Cox pour la normalisation de données de débit est illustré sur la figure 2.2 sur l'exemple des débits journaliers à l'exutoire du bassin versant de Baskatong (Québec) entre 1950 et 2013.

Pour une meilleure inférence, il faudrait estimer les paramètres de la normalisation de façon jointe avec ceux du modèle statistique de prévision. On le fait par exemple dans le système SCHEF (qui s'inspire des travaux de Wang et al. (2009)).

2.2.3 Choix méthodologiques et inférence

Plaçons-nous dans la situation où l'on cherche à écrire un modèle dont on puisse dériver $[Y|X]$, la loi de la quantité d'intérêt sachant le prédicteur.

La façon la plus directe de procéder consiste à écrire et ajuster un modèle donnant uniquement la loi conditionnelle $(Y|X)$. Les méthodes de régression procèdent de cette façon. L'inférence classique de tels modèles ne fait intervenir que la partie du jeu de données pour lequel on dispose conjointement d'observations de X et de Y , ce qui est regrettable si l'on dispose d'un jeu de données plus conséquent pour Y . Ce cas de figure intervient par exemple dans les problèmes de post-traitement des prévisions météorologiques : la plage de données pour laquelle on dispose de prévisions pour les prévisions d'ensemble est souvent très courte par rapport au jeu de

données des observations météorologiques climatologiques. Dans le cas des prévisions météorologiques d'ensemble du CMC par exemple, la dernière modification majeure de la méthode de production des prévisions météorologiques d'ensemble a eu lieu en 2013. Seules les prévisions émises après cette période sont donc utilisables pour un apprentissage statistique.

Afin d'utiliser les données relatives à Y non comprises dans l'échantillon joint, la solution mise en valeur par Krzysztofowicz (1999) consiste à écrire, plutôt qu'un unique modèle (conditionnel) pour $(Y|X)$, un modèle (marginal) pour Y et un modèle (conditionnel) pour $(X|Y)$. On obtient ensuite la distribution recherchée, $[Y|X]$, par la formule de Bayes : $[Y|X] = \frac{[X|Y][Y]}{[X]} \propto [X|Y][Y]$. Le jeu de données climatologique intervient dans l'inférence du modèle marginal dont on dérive $[Y]$.

Les raisons de procéder de la sorte ne se limitent pas à cela : l'écriture de deux modèles distincts, l'un dont on dérivera une proposition pour $[Y]$ et l'autre dont on dérivera une proposition pour $[X|Y]$, donne plus de flexibilité de modélisation comme l'illustrent sur un exemple Maranzano & Krzysztofowicz (2008). On peut par exemple imaginer faire des hypothèses de stationnarité différentes et faire intervenir des distributions de probabilité de différente nature pour l'un et pour l'autre.

Une distribution pour Y est de plus proposée en tant que sous-produit (et donc utilisable dans le cas éventuel où le prédicteur ne serait pas disponible au moment d'émettre la prévision par exemple).

Enfin, Krzysztofowicz (1999) met en avant que procéder ainsi apporte des garanties de calibration supplémentaires par rapport à une modélisation directe de $(Y|X)$. En effet, la loi marginale ($[Y]$), qui sert de référence pour évaluer la fiabilité des prévisions, intervient alors directement dans la construction du système de prévision. Toutefois, si l'inférence est conduite proprement et le jeu de données relatif à Y inclus dans le jeu de données conjoint de (X, Y) , on s'attend à retrouver les mêmes garanties.

Dans le contexte de cette thèse, la pertinence de cette modélisation en deux étapes dépend de la structure du jeu de données disponible :

- Dans le problème météorologique décrit dans le chapitre 3, où le jeu de données climatologique est beaucoup plus long que le jeu de données joint, cela paraît particulièrement pertinent. On cherche donc à s'en inspirer.
- Dans le problème hydrologique décrit dans le chapitre 4, où le jeu de données conjoint inclut le jeu de données climatologique et où nous nous plaçons dans un cadre intégralement gaussien, cette approche paraît moins avantageuse.

Le choix du critère à optimiser lors de l'inférence peut aussi faire l'objet d'une discussion. Un des critères d'inférence statistique les plus répandus, dans le contexte de l'approche classique, est celui du maximum de vraisemblance. Ce critère correspond au score ignorance décrit dans le chapitre 1. Il permet donc, si le jeu de données de validation est issu du même modèle que le jeu de données utilisé pour l'estimation, d'obtenir des prévisions probabilistes performantes conditionnellement

au modèle choisi. Nous avons vu que d'autres scores étaient couramment utilisés pour l'évaluation des prévisions probabilistes. Le CRPS, par exemple, présente des propriétés intéressantes (il est global, toujours fini et juste). Une alternative à la maximisation de la vraisemblance est donc la minimisation du CRPS. Ce choix de critère d'optimisation est fait dans la méthode Ensemble Model Output Statistics (EMOS) (Gneiting et al., 2005) par exemple. Dans cette thèse, nous préférons utiliser le critère du maximum de vraisemblance pour l'inférence de nos modèles afin d'avoir la possibilité d'utiliser des méthodes d'inférence bien établies en statistique, telles que l'algorithme Expectation-Maximisation (Dempster et al., 1977), utilisé dans les chapitres 3 et 4. Nous évaluons toutefois les prévisions ainsi obtenues à l'aide du CRPS afin de pallier le problème de scores ignorance infinis grâce au caractère global du CRPS.

Nous avons vu dans cette section les fondements communs aux méthodes de post-traitement statistique des prévisions météorologiques et hydrologiques. Dans les sections qui suivent, nous donnons un aperçu des méthodes proposées dans la littérature ou d'inspiration opérationnelle pour chacun de ces traitements.

2.3 Vue d'ensemble des méthodes de post-traitement météorologique

La distribution cible des post-traitements météorologiques, écrite $[\mathbf{P}|\mathbf{A}]$ sous sa forme simplifiée, est la suivante :

$$\left[P_{t_0+(1:H)} | A_{t_0}^{1:H}, P_{t_0}^- \right],$$

où t_0 désigne le jour d'émission des prévisions, H l'horizon de prévision et P et A désignent respectivement l'ensemble des variables météorologiques (de type températures et précipitations) et leurs prévisions sur l'ensemble des bassins versants.

Dans la grande majorité des méthodes de post-traitement des prévisions météorologiques existantes, les modèles statistiques proposés font intervenir une quantité à prévoir de dimension unitaire et un prédicteur de dimension K , où K désigne le nombre de membres de la prévision ($K = 1$ si la méthode vise à post-traiter des prévisions déterministe, $K > 1$ si la méthode vise à post-traiter des prévisions d'ensemble).

Dans les systèmes de prévision opérationnels, la variable à prévoir, $P_{t_0+(1:H)}$, et les prédicteurs, $A_{t_0}^{1:H}$ et $P_{t_0}^-$, sont donc de dimensions bien plus grandes (dimensions spatiale, temporelle et multi-variables) que dans les méthodes de post-traitement météorologique proposées dans la littérature. Comme déjà souligné dans la section 2.2.1, les systèmes opérationnels ont donc recours à ces méthodes pour chaque pas de temps, bassin versant et variable pris séparément. Ils font ensuite appel à des copules (empiriques dans HEFS, SCHEF et chez EDF) pour reconstituer des scénarios multidimensionnels pouvant être supposés simulés selon $[P_{t_0+(1:H)} | A_{t_0}^{1:H}]$. Notons que l'information provenant de la météorologie passée, $P_{t_0}^-$ reste toutefois

négligée, même après un recours aux copules empiriques. Il s'agit d'un point faible des méthodes de post-traitement météorologique décrites dans cette section.

Dans la suite de cette section, par souci de simplification, nous noterons A_{t_0} la prévision émise au temps t_0 pour l'échéance de prévision considérée, h , pour le bassin versant considéré, s , et pour la grandeur météorologique considérée, v , et P_{t_0} la variable météorologique correspondante. Autrement dit, $P_{t_0} = P_{t_0+h,s,v}$ et $A_{t_0} = A_{t_0,s,v}^h$.

La ligne 1 du tableau 2.1 donne un aperçu des intrants météorologiques possibles dans les systèmes de prévision opérationnels. Les lignes 2 à 4 résument les post-traitements météorologiques associés.

Parmi les systèmes de prévision pris en exemple dans le tableau 2.1, certains prennent en intrants des prévisions météorologiques déterministes : SCHEF, Hydro-Québec et, jusqu'à l'horizon 7 jours, HEFS. Le post-traitement météorologique consiste alors à proposer un modèle pour $(P|A)$ où P et A sont l'un et l'autre de dimension unitaire (P désigne par exemple la précipitation sur le bassin versant de Baskatong au cours de la journée suivant la date d'émission de la prévision et A la prévision déterministe d'Environnement Canada associée, à l'horizon un jour). Cela peut s'avérer relativement simple s'il s'agit de prévisions de températures moyennes ou maximales journalières, supposées gaussiennes, plus compliqué s'il s'agit de précipitations (typiquement non gaussiennes).

Dans le système SCHEF, après une transformation log-sinh dans le cas des précipitations, on suppose que la variable à prévoir et le prédicteur associés suivent conjointement une loi normale bidimensionnelle suivant la méthode du *Bayesian Joint Processor* adaptée de Wang et al. (2009) et Wang & Robertson (2011) par Robertson et al. (2013).

Les systèmes de prévision hydrologique prennent toutefois de plus en plus souvent en intrants des prévisions météorologiques d'ensemble. Dans ce cas, on pourrait s'attendre à ce que les membres de la prévision se comportent comme un échantillon issu d'une distribution pouvant être celle de la variable météorologique à prévoir. Ce choix de modélisation est par exemple sous-jacent au système de prévision de la CEMIG. Comme nous l'avons vu dans le chapitre 1 et comme souligné par Hamill & Colucci (1997); Bougeault et al. (2010); Perreault (2014) et Perreault (2015) (entre autres), cette hypothèse n'est malheureusement pas acceptable pour la plupart des prévisions d'ensemble considérées dans ce travail. Il faut donc également procéder à un post-traitement afin de simuler un nouvel échantillon, cette fois fiable.

Nous commençons par donner un aperçu des méthodes de post-traitement des prévisions météorologiques d'ensemble adaptées aux variables gaussiennes. Nous présentons ensuite des méthodes de post-traitement adaptées à d'autres variables, en nous concentrant sur les variables de type précipitations.

De nombreuses méthodes adaptées au post-traitement des prévisions d'ensemble de variables gaussiennes

Dans le cas où l'on a affaire à des prévisions d'ensemble, il est possible de faire appel, après une éventuelle transformation normalisante de \mathbf{P} et \mathbf{A} en X et Y , à la loi normale multidimensionnelle pour modéliser (Y, X) (cette fois-ci de dimension $K + 1$, où K désigne le nombre de membres et donc la dimension de X). Le Bayesian Processor of Output (BPO) (Krzysztofowicz & Maranzano, 2006) se fonde sur un tel modèle. Sans hypothèse supplémentaire sur la forme de la covariance entre les membres, ce modèle manque de parcimonie, ce qui est d'autant plus pénalisant que l'historique de prévisions disponible pour apprendre la loi de \mathbf{A} est en général très court, faute de reprévisions.

Une première possibilité pour gagner en parcimonie consiste à supposer que la moyenne de l'ensemble est la statistique exhaustive pour une modélisation paramétrique de l'ensemble. On peut alors imaginer traiter la moyenne d'ensemble comme une prévision déterministe et donc, par exemple, considérer un modèle gaussien bi-dimensionnel pour relier ce résumé et la quantité à prévoir. HEFS se fonde sur de tels modèles pour les échéances au-delà de 7 jours.

La méthode EMOS est plus parcimonieuse que le BPO et suppose que la dispersion des membres informe sur l'incertitude relative à la météorologie future. Le modèle conditionnel proposé est le suivant :

$$(Y|X) = a + \mathbf{b}^T X + \sqrt{c + dS_X^2} \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1),$$

où $a, c > 0$ et $d > 0$ sont des paramètres à estimer, \mathbf{b} un vecteur de paramètres à estimer (de taille K) et S_X désigne l'écart-type de l'ensemble X . Un inconvénient de cette méthode est qu'elle ne permet pas, contrairement au BPO par exemple, d'utiliser l'ensemble de l'information climatologique disponible sur Y pour l'inférence du modèle (et donc pour la prévision) : le modèle propose directement une loi conditionnelle sans passer par la loi jointe de X et de Y .

Une autre façon de rendre les modèles de post-traitement plus parcimonieux consiste à supposer que les membres de l'ensemble sont échangeables. Les membres d'un ensemble sont dits *échangeables* si leur distribution jointe est invariante par toute permutation de leurs indices. L'hypothèse d'échangeabilité peut être acceptée dans le cas des prévisions d'ensemble CEPMMT-EPS et NCEP-GEFS mais elle n'est pas réaliste dans le cas du CMC-EPS par exemple (voir le chapitre 1 pour une description de ces prévisions). En effet, chacun de ses membres est toujours produit avec sa propre combinaison de sous-modèles atmosphériques. Ses membres pairs, par exemple, proviennent du modèle de convection OldKuo tandis que ses membres impairs proviennent du modèle Kain-Fritsch¹.

En supposant l'échangeabilité des membres, le modèle EMOS s'écrit de la façon suivante :

$$(Y|X) = a + b\bar{X} + \sqrt{c + dS_X^2} \varepsilon \quad \varepsilon \sim \mathcal{N}(0, 1),$$

1. http://collaboration.cmc.ec.gc.ca/cmc/ensemble/doc/info_geps_e.pdf.

où \bar{X} désigne la moyenne de l'ensemble X et b est réduit à un scalaire. Les statistiques exhaustives de l'ensemble de prévisions sont, dans ce modèle, sa moyenne et sa variance.

Une limite aux modèles gaussiens décrits ci-dessus est qu'il en résulte systématiquement une distribution unimodale. Or, il est parfois pertinent d'émettre une distribution prédictive multimodale. Prenons par exemple une situation météorologique dans laquelle deux catégories de scénarios sont envisageables. Dans l'un, un orage éclate au-dessus du bassin versant considéré et des précipitations tombent en grande quantité. Dans l'autre, il reste bloqué sur le bassin versant voisin et le temps reste sec. Ce type de situation peut être illustré par les membres : la moitié d'entre eux peut être témoin de la première catégorie de scénario et l'autre de la seconde.

La figure 2.3 illustre la prévision d'ensemble CEPMMT-EPS émise le 18 octobre 2015 au niveau du bassin versant de l'Ain à Vouglans dont les membres présentent ce type de comportement : leur distribution empirique est multimodale à plusieurs échéances de prévision. Ce comportement reste toutefois très rare.

Dans le cas où il est souhaitable de pouvoir émettre des distributions prédictives multimodales, les méthodes de post-traitement *par habillage* (aussi dites à *noyaux*) peuvent être particulièrement pertinentes. Ces méthodes reposent sur l'hypothèse que, à chaque temps de prévision, un des membres de l'ensemble est le plus proche de la vérité. Le modèle sous-jacent à ces méthodes est le suivant :

$$\begin{aligned} (Y|X, k^*) &\sim \mathcal{G}(X_{k^*}, \boldsymbol{\theta}_{k^*}) \\ [k^* = k] &= \pi_k \end{aligned} \quad (2.2)$$

où $\sum_{k=1}^K \pi_k = 1$, \mathcal{G} est une loi à choisir, $\boldsymbol{\theta}_k$ et π_k , pour $k \in \{1, \dots, K\}$, sont des paramètres à estimer et k^* est l'indice du meilleur membre. Dans le cas gaussien considéré jusqu'alors, on prend pour \mathcal{G} une loi gaussienne. On peut par exemple prendre :

$$\mathcal{G}(X_k, \boldsymbol{\theta}_k) = \mathcal{N}(a_k X_k + b_k, \sigma_k^2),$$

avec $\boldsymbol{\theta}_k = (a_k, b_k, \sigma_k)$ pour $k \in \{1, \dots, K\}$. La distribution prédictive résultante est dans ce cas un mélange de lois gaussiennes. Dans le cas de membres échangeables, $\pi_k = \frac{1}{K} \quad \forall k \in \{1, \dots, K\}$ et il existe $\boldsymbol{\theta}$ tel que $\boldsymbol{\theta}_k = \boldsymbol{\theta} \quad \forall k \in \{1, \dots, K\}$.

Un autre avantage à utiliser des méthodes par habillage réside dans le fait que leur extension au cas multi-ensemble est directe : les membres de chaque ensemble ou sous-ensemble échangeable e partageront simplement les mêmes paramètres π_e et $\boldsymbol{\theta}_e$.

La *méthode du meilleur membre* de Roulston & Smith (2003) est, à notre connaissance, la première méthode procédant par habillage proposée pour le post-traitement des prévisions d'ensemble. Il s'agit d'une méthode non paramétrique. La probabilité que le membre k soit le meilleur, π_k , est donnée par la proportion de jours de la période d'apprentissage où le membre k était (parmi les K membres) le plus proche de l'observation selon une distance de type norme euclidienne. Roulston & Smith (2003) approchent la loi de $(Y|X, k^*)$ par l'échantillon d'erreurs du meilleur

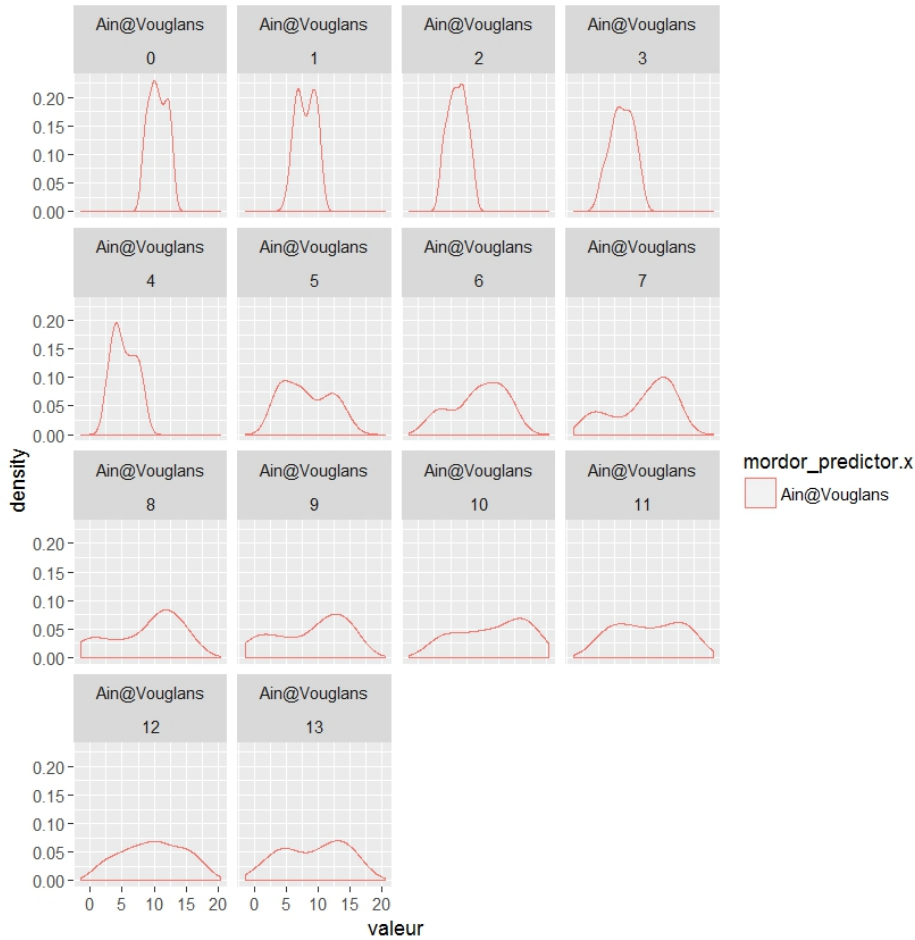


FIGURE 2.3 – Distributions de probabilités empiriques des prévisions d’ensemble de températures CEPMMT-EPS émises le 18 octobre 2015 pour les 13 jours qui suivent sur le bassin versant de l’Ain à Vouglans (France).

membre, $(Y - X_k^*)$, de la période d’apprentissage. Cette façon de faire n’offre pas de garantie de fiabilité et la taille de l’échantillon d’apprentissage disponible est fortement limitante, ce qui a conduit Wang & Bishop (2005) à proposer une version paramétrique de la méthode du meilleur membre pour le cas gaussien. Toutefois, cette nouvelle méthode conduit à une surestimation des valeurs extrêmes, le kurtosis ainsi estimé étant trop grand (Fortin et al., 2006). De plus, elle n’est pas adaptée au post-traitement des prévisions d’ensemble sur-dispersives car elle conduit alors des écart-types estimés négatifs.

L’adaptation de la méthode du *Bayesian Model Averaging* (BMA) proposée par Raftery et al. (2005) pour le post-traitement des prévisions d’ensemble peut aussi être vue comme une méthode de post-traitement par habillage. Le modèle proposé est celui donné par l’équation (2.2) et une inférence par algorithme EM est proposée dans

le cas gaussien, conduisant ainsi à des paramètres maximisant la vraisemblance et donc minimisant le score ignorance. Cette méthode présente comme les précédentes l'inconvénient de possiblement mal s'adapter au cas où l'ensemble à post-traiter serait sur-dispersif. Dans une expérience sur des données simulées menée par [Raftery et al. \(2005\)](#), cette méthode conduit, dans le cas où les prévisions sont sur-dispersives et où la moyenne et la variance des observations et des prévisions change d'un pas de temps au suivant, à des prévisions légèrement sur-dispersives (avec un taux de couverture des intervalles de prévisions à 66,7% et à 90% de respectivement 72% et 91.9%).

Une solution à ce dernier problème consiste à indiquer les membres par leur rang. Leur indice est alors porteur d'information. Cette idée fait partie des idées sous-jacentes à la *méthode des membres pondérés* de [Fortin et al. \(2006\)](#). On perd alors toutefois en parcimonie dans le cas échangeable.

Quelques méthodes adaptées au post-traitement des variables mixtes de type précipitations

Les variables de type précipitations posent un problème de modélisation particulier du fait de leur nature mixte : continue (pour leur intensité) et discrète (pour leur occurrence). Une approche classique pour leur modélisation consiste à séparer la modélisation de la variable "occurrence de précipitations", binaire, de la variable "quantité de précipitations conditionnellement à l'occurrence", continue. Dans le système de prévision d'Hydro-Québec par exemple, la quantité de précipitations conditionnellement à l'occurrence est supposée suivre une loi de Pearson type III.

Une des méthodes de post-traitement les plus proches du cadre gaussien et procédant à cette modélisation en deux parties est celle proposée par [Herr & Krzysztofowicz \(2005\)](#), appliquée par exemple dans le système HEFS. On procède dans cette méthode à une normalisation mais elle ne concerne que les parties positives des précipitations observées et prévues. On obtient ainsi Y et X à partir de P et A . Le modèle fait intervenir un mélange de lois gaussiennes (pour $(X, Y|A > 0, P > 0)$ par exemple) et de lois binomiales (pour $(A > 0, P = 0)$ par exemple). À ce jour, cette méthode n'a pas été généralisée au cas où les prévisions sont des prévisions d'ensemble.

Dans le cas des prévisions d'ensemble de précipitations, une méthode faisant l'hypothèse que les variables de type précipitations suivent une loi généralisée des extrêmes censurée a été proposée dans la continuité d'EMOS par [Scheuerer \(2014\)](#). Le modèle est le suivant :

$$\begin{cases} (Y|A) & \sim GEV(\mu_A, \gamma + \kappa MD_A, \xi) \\ \mu_A \text{ tel que } \mathbb{E}(Y|A) & = \alpha + \beta \bar{A} + \frac{1}{K} \sum_{k=1}^{k=K} \mathbb{I}_{A_k=0} \\ MD_A & = \frac{1}{K^2} \sum_{k,k'} |A_k - A_{k'}| \\ (P|Y) & = Y \mathbb{I}_{Y \geq 0} \end{cases}$$

où α , β , γ , κ et ξ sont des paramètres à estimer, A_k désigne le k^e membre de la pré-

vision météorologique d'ensemble, K désigne le nombre de membres (et donc la dimension) de \mathbf{A} , $MD_{\mathbf{A}}$ est une mesure de la dispersion de l'ensemble \mathbf{A} , et $GEV(\mu, \sigma, \xi)$ est la loi d'extremum généralisée de paramètre de position μ , de paramètre de dispersion $\sigma > 0$ et de paramètre de forme ξ . Les avantages et inconvénients de cette méthode sont les mêmes que ceux de la méthode EMOS.

Dans l'esprit des méthodes de post-traitement par habillage, une extension de l'approche BMA aux variables de précipitations dans le cas échangeable a été proposée par Slougher et al. (2007) :

$$\left\{ \begin{array}{ll} \text{logit} \left[\mathbb{P}^{\frac{1}{3}} = 0 | \mathbf{A}, k^* = k \right] & = d_0 + d_1 A_k^{\frac{1}{3}} + d_2 \mathbb{I}_{(A_k=0)} \\ \left(\mathbb{P}^{\frac{1}{3}} | \mathbf{A}, k^* = k, \mathbb{P} > 0 \right) = \mathbb{P}' & \sim \Gamma(\alpha(A_k), \beta(A_k)) \\ \alpha(A_k) \text{ et } \beta(A_k) & \text{tels que } \mathbb{E}(\mathbb{P}') = b_0 + b_1 A_k^{\frac{1}{3}} \\ & \text{et } \mathbb{V}(\mathbb{P}') = c_0 + c_1 A_k \\ [k^* = k] & = \frac{1}{K} \forall k \in \{1, \dots, K\} \end{array} \right.$$

où \mathbf{d} , \mathbf{b} et \mathbf{c} sont des vecteurs de paramètres à estimer et $\Gamma(\alpha, \beta)$ la loi gamma de paramètres α et β . On s'attend ici à ce que d_1 prenne une valeur négative : la probabilité de temps sec est ainsi d'autant plus faible que le meilleur membre prédit une faible quantité de précipitations. Les avantages et inconvénients de cette méthode sont ceux des méthodes par habillage.

La méthode choisie par EDF pour le post-traitement des prévisions de précipitations peut être interprétée comme une méthode par habillage. Des situations analogues passées (telles que celles utilisées pour l'application des copules empiriques par exemple) y sont utilisées afin de proposer une loi pour $(\mathbb{P} | \mathbf{A}, Z, k^* = k)$, où Z est la prévision de champs de pression du CEPMMT-EPS. On peut aussi rapprocher cette méthode des méthodes non paramétriques faisant appel à des arbres de décision : ces dernières impliquent aussi l'identification de situations analogues (telles que celles de la feuille atteinte en suivant un arbre de décision).

Des méthodes s'approchant davantage des méthodes de régression quantile ont aussi été proposées. Wilks (2009) a recours à une méthode de type régression logistique. Le modèle est le suivant :

$$\text{logit} [\mathbb{P} < p | \mathbf{A}] = \alpha + \boldsymbol{\beta}^T \mathbf{A} + \gamma g(p), \quad (2.3)$$

où α et γ sont des paramètres à estimer, $\boldsymbol{\beta}$ est un vecteur de paramètres à estimer et g est une fonction croissante à choisir. Wilks (2009) propose de prendre pour g la fonction racine carrée : $g(p) = \sqrt{p}$ et constate une légère amélioration des résultats en remplaçant $\boldsymbol{\beta}^T \mathbf{A}$ par $\beta \sqrt{\mathbf{A}}$ dans l'équation (2.3), avec β un paramètre à estimer.

Afin d'utiliser l'information contenue dans la dispersion des membres, une version hétéroscédastique de ce modèle est proposée par Messner et al. (2014). Le modèle

sous-jacent est le suivant :

$$\text{logit} [P < p|A] = \frac{\alpha + \beta^T A + \gamma g(p)}{\exp(\kappa S_A)},$$

où S_A désigne l'écart-type empirique de l'ensemble A , α , γ et κ sont des paramètres à estimer, β est un vecteur de paramètres à estimer et g une fonction croissante à choisir, par exemple la fonction racine carrée.

Ben Bouallègue (2013) propose par ailleurs d'ajouter un terme d'interaction dans le modèle de Wilks (2009). L'équation (2.3) devient alors :

$$\text{logit} [P < p|A] = \alpha + \beta^T A + \gamma g(p) + g(p)\beta'^T A,$$

où β' est un vecteur de paramètres à estimer.

Notons que, dans ces modèles, sous l'hypothèse d'échangeabilité, β_k aura la même valeur quel que soit l'indice $k \in \{1, \dots, K\}$.

Comme les méthodes de régression quantiles, ces méthodes ne produisent pas directement une distribution prédictive mais permettent d'en estimer les quantiles.

MÉTHODES NON PARAMÉTRIQUES Dans une toute autre démarche, des méthodes statistiques non paramétriques ont été proposées pour le post-traitement des prévisions météorologiques d'ensemble.

Taillardat et al. (2016), par exemple, ont utilisé une méthode combinant régression quantile et forêts aléatoires, développée par Meinshausen (2006), pour le post-traitement de prévisions d'ensemble de températures et de vitesses de vent.

Les forêts aléatoires sont obtenues en construisant un ensemble d'arbres de décision. La construction d'un *arbre de décision* (Breiman et al., 1984) consiste à diviser de façon itérative le jeu de données d'apprentissage en deux groupes au sein desquels les observations sont les plus homogènes possibles. Cette division en deux est opérée, à chaque itération, par le choix d'un seuil portant sur la valeur d'un des prédicteurs. Les arbres ainsi obtenus étant très dépendants du jeu de données d'apprentissage, Breiman (1996) propose de construire de cette façon plusieurs arbres après avoir échantillonné avec remise parmi les répétitions du jeu de données d'apprentissage. Breiman (2001) propose, de plus, d'échantillonner à plusieurs reprises quelques prédicteurs parmi les prédicteurs disponibles. L'ensemble des arbres de décision issus de ces deux échantillonnages est appelé *forêt aléatoire*. En situation de prévision, on obtiendra donc, en suivant chacun des arbres de décision ainsi construits, à chaque fois un groupe d'observations de la variable à prévoir, appelé *feuille* de l'arbre. Meinshausen (2006) propose une méthode permettant d'utiliser l'ensemble de ces groupes pour approcher une fonction de répartition prédictive.

Cette approche a l'avantage de ne pas spécifier d'hypothèse sur la forme de la distribution prédite (et donc notamment de permettre une multimodalité de la prévision) et de pouvoir intégrer des informations auxiliaires de natures variées. Il est de plus *a priori* aisé de l'adapter pour le post-traitement de variables de type mixte

telles que les précipitations. En contre-partie, elle ne produit pas directement une distribution prédictive, est coûteuse en temps de calcul et nécessite un jeu de données d'entraînement plus conséquent que les précédentes méthodes. Pour plus de détails concernant cette méthode, se référer à [Taillardat et al. \(2016\)](#), pages 2376 à 2378.

MÉTHODES D'AGRÉGATION SÉQUENTIELLE DE PRÉDICTEURS Il est enfin possible de considérer le problème de post-traitement des prévisions d'ensemble comme un problème d'agrégation séquentielle de prédicteurs ([Stoltz, 2010](#)), où les prédicteurs (ou experts) sont les membres de l'ensemble. Ces méthodes consistent à attribuer à chaque prédicteur (ou membre) un poids qui sera mis à jour à chaque nouvelle observation de sorte à ce que la prévision déterministe résultant de la combinaison des prédicteurs soit la plus performante possible (selon un critère appelé regret, à minimiser).

Les prévisions issues des méthodes d'agrégation séquentielle de prédicteurs ont pour avantage, du fait de leur caractère séquentiel, de s'adapter naturellement à d'éventuelles instationnarités. De plus, elles ne reposent pas sur des hypothèses vis-à-vis du processus stochastique dont est issue la variable à prévoir. Toutefois, elles sont historiquement conçues dans le but d'émettre une prévision déterministe. De plus, dans le cas échangeable, leur intérêt est limité puisque la meilleure combinaison consiste alors à attribuer le même poids à tous les prédicteurs.

2.4 Vue d'ensemble des méthodes de prise en compte de l'incertitude hydrologique

L'utilisation de SPQ le plus souvent conceptuels, peu coûteuse en temps de calcul, donne le champ libre à une multitude d'approches pour la prise en compte de l'incertitude hydrologique. Elles consistent à mettre en place une méthode permettant de produire des simulations se réclamant d'un modèle conditionnel prédictif donnant :

$$[\mathbf{Q}|\mathcal{M}(\mathbf{P})] = \left[Q_{t_0+(1:H)} | \mathcal{M} \left(P_{t_0}^-, P_{t_0+(1:H)} \right), Q_{t_0}^- \right] \quad (2.4)$$

pour tout t_0 , où t_0 est le jour d'émission de la prévision, H est l'horizon de prévision et $\mathcal{M}(\cdot)$ désigne la transformation opérée par le SPQ. Pour ce faire, les difficultés résident essentiellement dans l'intégration de l'information en provenance des derniers débits observés.

Dans les systèmes de prévision opérationnels comme dans les méthodes proposées dans la littérature, on considère le problème unité géographique par unité géographique (le plus souvent un bassin versant). Selon les cas, on fait parfois appel à une copule empirique pour retrouver la cohérence spatio-temporelle dans un second temps. Dans cette section, on considère le problème de modélisation pour un bassin versant donné. La variable à prévoir, $Q_{t_0+(1:H)}$, est donc de dimension H .

Dans la suite de cette section, on note $M = \mathcal{M}(P)$, $f_{\mathcal{N}}$ est la notation générique pour les transformations normalisantes, et on note $X = f_{\mathcal{N}}(M)$ et $Y = f_{\mathcal{N}}(Q)$.

Les lignes 6 à 9 du tableau 2.1 donnent un aperçu de l'approche choisie pour la prise en compte de l'incertitude hydrologique dans les cinq principaux systèmes de prévision pris en exemple dans ce manuscrit. Parmi les méthodes de prise en compte de l'incertitude hydrologique ici décrites, on distingue celles la prenant en compte (en partie ou totalement) là où elle intervient (les méthodes d'assimilation séquentielle de données et les méthodes bayésiennes) de celles la prenant en compte en totalité après-coup (les méthodes de post-traitement).

ASSIMILATION SÉQUENTIELLE DE DONNÉES Une première façon d'intégrer les dernières observations de débits consiste à procéder à leur assimilation séquentielle. L'assimilation de données consiste, pour chaque temps de prévision, à modifier les variables d'état et/ou les paramètres du SPQ afin de tenir compte des dernières observations de débits.

Le filtre de Kalman (Schweppe, 1973) en est la méthode de référence. Kitanidis & Bras (1980), par exemple, ont utilisé un filtre de Kalman pour la mise à jour séquentielle des variables d'état d'un SPQ. L'utilisation de cette méthode nécessite toutefois une simplification du SPQ, non linéaire, pour être applicable. Moradkhani et al. (2005b) proposent d'utiliser à la place un filtre de Kalman d'ensemble (EnKF, Evensen (1994)), une méthode couplant le filtre de Kalman avec une méthode de Monte-Carlo et dont les hypothèses sous-jacentes sont moins restrictives. Avec cette méthode, Moradkhani et al. (2005b) sont à même de mettre à jour les paramètres du SPQ en même temps que ses variables d'état.

Dans le cas où les hypothèses de l'EnKF ne seraient pas acceptables, Moradkhani et al. (2005a); Salamon & Feyen (2009) et DeChant & Moradkhani (2012) proposent d'utiliser un filtre particulaire (Gordon et al., 1993).

Bien que ces méthodes soient connues pour produire des prévisions déterministes très performantes, l'EnKF comme le filtre particulaire résultent généralement en des prévisions probabilistes sous-dispersives (voir par exemple Moradkhani et al. (2005a); Salamon & Feyen (2009)), car ils ne permettent pas de prendre en compte la part d'incertitude hydrologique relative à la structure du SPQ. Une solution partielle à ce problème, proposée et mise en œuvre par Thiboult et al. (2016), consiste à utiliser plusieurs SPQ à la fois.

ASSIMILATION DE DONNÉES COMBINÉE À UN MODÈLE D'ERREUR Une autre solution pour obtenir des prévisions probabilistes fiables, moins coûteuse en temps de calcul, consiste à coupler une méthode d'assimilation de données à un modèle d'erreur.

Dans un premier temps, la méthode d'assimilation de données est utilisée de sorte à ce que les sorties du SPQ ainsi modifié, M'_{DA} , soient plus proches des observations correspondantes, Q_{DA} , que ne l'étaient ses sorties originales, M_{DA} , où DA désigne la période utilisée pour l'assimilation de données.

Le système du CEMIG applique une telle méthode d'assimilation de données (Da Paz et al., 2007) tandis qu'une assimilation de données manuelle est opérée dans le système d'Hydro-Québec.

Une alternative consiste à utiliser une correction déterministe pour passer de M à M' . SCHEF applique par exemple une méthode appelée *dual-pass error correction technique* (Pagano et al., 2011). Cette correction est la suivante :

$$M'_{t_0+h} = \max \left(0, M_{t_0+h} + \beta \exp \left\{ -\frac{h-1}{\alpha} (Q_{t_0} - M_{t_0}) \right\} \right),$$

où α et β sont des paramètres à estimer.

EDF utilise une correction déterministe similaire faisant de plus intervenir des variables d'état du SPQ pour arbitrer entre deux modèles de dépendance temporelle des erreurs.

Enfin, dans la méthode *Error Reduction and Representation in Stages* (ERRIS), récemment proposée par Wang et al. (2016), après une normalisation des débits par la transformation log-sinh, une première étape fait intervenir une régression linéaire simple visant à corriger de façon déterministe un éventuel biais systématique du SPQ. Une deuxième étape consiste ensuite à utiliser le modèle auto-régressif proposé par Li et al. (2015) sous la forme d'une correction déterministe afin d'assimiler les dernières observations de débits. On obtient ainsi X' (l'équivalent de M' obtenu avec les précédentes méthodes mais dans un espace normalisé). Cette méthode est pour l'instant uniquement décrite pour le cas de la prévision à une seule échéance de prévisions, h .

Après l'assimilation de données ou bien la correction déterministe visant à tenir compte des derniers débits observés, un modèle doit être proposé afin de produire une distribution se réclamant de $[Q_{t_0+h}|M'_{t_0+h}]$ (ou bien $[Y_{t_0+h}|X'_{t_0+h}]$ après normalisation) pour chaque échéance h .

Le système d'EDF fait intervenir dans ce but un modèle gaussien hétéroscédastique après normalisation des variables de débits. Dans ce modèle :

$$(Y_{t_0+h}|X_{t_0+h} = x) = b_h(x) + x + \sigma_h(x) \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1), \quad (2.5)$$

où b_h et σ_h sont des fonctions tabulées de x (évaluées par gamme de valeurs de x) et la normalisation opérée préalablement est $f_{\mathcal{N}} = \log$.

ERRIS fait pour sa part intervenir un mélange de deux lois gaussiennes centrées sur X'_{t_0+h} :

$$\begin{aligned} (Y_{t_0+h}|X'_{t_0+h} = x, S = s) &\sim \mathcal{N}(x, \sigma_{s,h}^2) \\ [S = 1] &= \omega \quad [S = 0] = 1 - \omega, \end{aligned}$$

où S est une variable latente guidant le mélange entre les deux lois gaussiennes et ω et $(\sigma_{s,h})_{s \in \{0,1\}}$ sont des paramètres à estimer. Les paramètres sont tous estimés de sorte à maximiser la vraisemblance.

Lorsque l'on dispose d'un SPQ peu coûteux en temps, une autre solution, choisie par Hydro-Québec, consiste à créer des prévisions d'ensemble hydrologiques en perturbant ses variables d'état.

Enfin, dans le SCHEF et dans le système de la CEMIG, on suppose simplement que l'erreur de modélisation hydrologique résiduelle est négligeable vis-à-vis des autres sources d'incertitude : $Q_{t_0+(1:H)} = M'_{t_0+(1:H)}$.

APPROCHE BAYÉSIENNE ET DÉCOMPOSITION DE L'INCERTITUDE Une toute autre approche pour la prise en compte de l'incertitude hydrologique consiste à considérer le SPQ comme une partie intégrante du modèle de prévision. Les paramètres du SPQ font alors l'objet d'une inférence jointe avec ceux du modèle d'erreur.

Les premières méthodes de ce type proposées dans la littérature sont des méthodes bayésiennes. Ces méthodes permettent d'identifier et de tenir compte formellement de la part d'incertitude portant sur les paramètres du SPQ. L'estimation de la loi a posteriori des paramètres du SPQ, notés θ , nécessite de choisir une loi a priori sur ces paramètres, donnant $[\theta]$, un modèle d'erreur pour le SPQ, donnant $[\mathbf{Q}|\theta] = [\mathbf{Q}|\mathbf{M}_\theta]$, où \mathbf{M}_θ désigne les sorties du SPQ paramétrisé par θ , et enfin d'approcher $[\theta|\mathbf{Q}] \propto [\mathbf{Q}|\theta][\theta]$.

Kuczera (1983) est, à notre connaissance, le premier à avoir proposé une méthode bayésienne pour la prise en compte de l'incertitude hydrologique. Il effectue pour cela une approximation du premier ordre de la transformation opérée par le SPQ et suppose que les erreurs de paramétrisation sont additives.

Les hypothèses sous-jacentes à la méthode de Kuczera (1983) n'étant souvent pas valides en hydrologie, Beven & Binley (1992) proposent alternativement d'intégrer l'incertitude de paramétrisation par une méthode Monte-Carlo (ils effectuent des simulations selon des lois a priori sur les paramètres, $[\theta]$, puis en sélectionnent certaines). Ils ne spécifient toutefois pas explicitement un modèle d'erreur (donnant $[\mathbf{Q}|\theta]$) et la méthode d'échantillonnage utilisée est peu efficace (elle revient à effectuer de l'importance sampling avec une fonction auxiliaire uniforme).

Kuczera & Parent (1998) proposent en solution d'échantillonner selon la loi a posteriori des paramètres du SPQ, dont on sait, d'après la formule de Bayes, qu'elle est proportionnelle au produit de la loi a priori et de la vraisemblance : $[\theta|\mathbf{Q}] \propto [\mathbf{Q}|\theta][\theta]$, avec l'algorithme de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970). Il s'agit aussi de l'approche choisie dans la méthode *Differential Evolution Adaptive Metropolis* (DREAM) développée par Vrugt et al. (2009). La méthode DREAM a notamment été utilisée pour la prévision de l'incertitude hydrologique par Schoups & Vrugt (2010).

Kavetski et al. (2002) proposent une méthode, *Bayesian Total Error Analysis* (BATEA), permettant, de plus, de tenir compte explicitement de l'incertitude provenant de la méconnaissance de la lame d'eau réellement tombée sur le bassin versant, souvent approchée de façon très approximative à partir de mesures en quelques pluviomètres répartis sur ou bien à proximité du bassin versant. Elle a été utilisée assortie d'un modèle d'erreur par, entre autres, Kuczera et al. (2006); Thyer et al.

(2009) et Renard et al. (2011). BATEA permet donc de décomposer et d'intégrer l'incertitude hydrologique selon trois sources : les intrants du SPQ, la paramétrisation du SPQ et la structure du SPQ.

Evin et al. (2014) notent que l'inférence jointe des paramètres du SPQ et des paramètres du modèle d'erreur peut poser des problèmes de robustesse, notamment en raison d'interactions fortes entre leurs paramètres. Ainsi, dans leur étude comparative, un modèle complet résultant de l'inférence séparée des paramètres du SPQ et de ceux du modèle d'erreur produit des prévisions probabilistes plus performantes que ce même modèle dont les paramètres ont été estimés par une inférence jointe.

Il est donc parfois préférable de procéder à une inférence séparée des deux jeux de paramètres. On parle alors de *post-traitement*.

POST-TRAITEMENT Sorooshian & Dracup (1980) sont des précurseurs de cette approche. Elle a par la suite été développée et appliquée par de nombreux auteurs dont Krzysztofowicz (1999); Todini (2008); Engeland et al. (2010); Coccia & Todini (2011) et Pianosi & Raso (2012). Des méthodes de post-traitement non paramétriques ont aussi été proposées. Solomatine & Shrestha (2009) et Weerts et al. (2011) utilisent par exemple des méthodes de régression quantile.

Des approches de post-traitement permettent enfin de traiter conjointement les sorties de plusieurs SPQ. Duan et al. (2007) et Vrugt & Robinson (2007), par exemple, utilisent pour cela la méthode du BMA décrite dans la section 2.3. Les informations provenant des dernières observations de débits ne sont toutefois pas utilisées dans ces méthodes. Hemri et al. (2013) utilisent en solution une extension multivariée du BMA, originellement proposée par Berrocal et al. (2007) pour la prise en compte de la dépendance spatiale.

On considère alors que l'incertitude de paramétrisation du SPQ et l'incertitude portant sur ses intrants sont négligeables devant l'incertitude résiduelle et les autres incertitudes intervenant dans le système de prévision (notamment l'incertitude météorologique).

Nous avons choisi cette approche pour la méthode que nous développons dans le chapitre 4.

MODÈLE D'ERREUR POUR LES MÉTHODES BAYÉSIENNES ET LES MÉTHODES DE POST-TRAITEMENT

Les modèles d'erreur, dans les méthodes bayésiennes comme dans les post-traitements, ont pour cible la relation entre Q et M . La plupart des modèles proposés à cet effet dans la littérature visent plus spécifiquement la différence entre ces deux grandeurs, $W = Q - M$.

Deux difficultés majeures interviennent pour l'écriture de ces modèles : l'hétéroscédasticité des erreurs, d'une part, et leur dépendance temporelle, d'autre part (Sorooshian & Dracup, 1980).

La dépendance temporelle est typiquement prise en compte en formulant un modèle auto-régressif. Kuczera (1983) propose par exemple un modèle auto-régressif à moyenne mobile (ARMA, Box & Jenkins (1976)) pour $W = Q - M$ tandis que Bates

& Campbell (2001) et Pianosi & Raso (2012) proposent un modèle auto-régressif simple et beaucoup d'auteurs ont recours à un modèle auto-régressif d'ordre 1 (Sorooshian & Dracup, 1980; Schaeffi et al., 2007; Schoups & Vrugt, 2010; Engeland et al., 2010; Evin et al., 2013, 2014). Le modèle d'erreur prend dans ce dernier cas la forme suivante :

$$W_t = \rho W_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad (2.6)$$

où ρ et σ sont des paramètres à estimer.

L'hétéroscédasticité des erreurs, elle, est typiquement traitée en ayant recours à une transformation normalisante, $f_{\mathcal{N}}$, telle que celles décrites dans la section 2.2.2. Le modèle de l'équation (2.6) porte alors sur $W' = f_{\mathcal{N}}(W)$. Il s'agit par exemple de la solution choisie par Sorooshian & Dracup (1980); Kuczera (1983); Bates & Campbell (2001); Krzysztofowicz (2002); Krzysztofowicz & Maranzano (2004); Seo et al. (2006); Todini (2008) et Engeland et al. (2010).

Une autre façon de procéder consiste à tenir compte explicitement de cette hétéroscédasticité dans le modèle, par exemple en supposant que l'écart-type dépend linéairement du débit (Thyer et al., 2009; Schoups & Vrugt, 2010; Pianosi & Raso, 2012; Evin et al., 2013, 2014). σ est alors remplacé, dans l'équation (2.6), par $\sigma_t = \alpha_0 + \alpha_1 Q_t$, où α_0 et α_1 sont des paramètres à estimer. En pratique, l'inférence d'un tel modèle s'effectue en réduisant préalablement les données. Le modèle porte alors sur $W' = \frac{W}{\alpha_0 + \alpha_1 Q}$ (les SPQ étant généralement sans biais et, par conséquent, l'espérance de W nulle).

Enfin, d'autres modèles se détachent du cadre purement gaussien. Par exemple, Schaeffi et al. (2007) supposent que W se comporte comme un mélange de deux lois gaussiennes, l'une correspondant aux forts débits et l'autre aux faibles débits. Schoups & Vrugt (2010) font pour leur part intervenir une loi normale généralisée asymétrique (Azzalini, 1986).

Li et al. (2015) soulignent que, dans ce contexte, une transformation normalisante n'est pas toujours suffisante pour une prise en compte adéquate de l'hétéroscédasticité des erreurs de prévision. Ils montrent notamment que les modèles auto-régressifs usuels peuvent être à l'origine de sur-corrrections détériorant les prévisions originales du SPQ. Des solutions impliquant des contraintes sous la forme de seuils y sont proposées.

Ces sur-corrrections pourraient être dues à l'existence de plusieurs régimes d'erreur pour le SPQ, par exemple un comportement auto-régressif différent en montée et en descente de crue. Cette observation a par exemple été faite dans le cas du SPQ d'EDF : sur certains bassins versants, il est possible, en descente de crue, que l'erreur du SPQ se reporte simplement d'un pas de temps à l'autre tandis qu'en montée de crue l'erreur ait tendance à s'amenuiser avec les impulsions de débits simulées par le SPQ.

Dans le chapitre 4, nous proposons un post-traitement hydrologique impliquant deux régimes de dépendance entre les débits à prévoir et les sorties du SPQ.

Des méthodes impliquant plusieurs régimes de dépendance ont déjà été proposées

pour la prévision de débits à l'exutoire de bassins versants. Selon les approches, la modélisation porte soit directement sur la probabilité d'appartenance à un régime, soit sur les probabilités de transition d'un régime à l'autre.

Dans le modèle de [Schaeffi et al. \(2007\)](#), par exemple, les erreurs du SPQ peuvent se comporter selon deux modèles auto-régressifs différents selon que le débit est inférieur ou supérieur à une valeur seuil.

Dans la méthode proposée par [Furrer et al. \(2006\)](#), quatre régimes hydrologiques interviennent pour modéliser les débits en rivière (sans passer par l'intermédiaire d'un SPQ). Ces quatre régimes correspondent au régime hivernal, de crue printanière, de décrue et estival ou automnal des rivières québécoises. Les probabilités de transition d'un régime à l'autre y sont exprimées en fonction de variables auxiliaires par le biais d'un modèle de régression binaire de type Probit, selon l'approche développée par [Lu & Berliner \(1999\)](#).

De même, [Perreault et al. \(2007\)](#) utilisent, pour la prévision à long terme des apports en eau aux barrages, des approches développées par [Chib \(1996\)](#) et par [Albert & Chib \(1993\)](#) faisant intervenir deux régimes possibles. Dans l'approche s'appuyant sur [Chib \(1996\)](#), les probabilités de transition d'un régime à l'autre sont exprimées en fonction de variables météorologiques par le biais d'un modèle Probit. Dans la méthode s'appuyant sur [Albert & Chib \(1993\)](#), c'est la probabilité d'appartenance à l'un ou à l'autre des régimes qui est exprimée en fonction de ces variables météorologiques au travers d'un modèle Probit. Enfin, la correction déterministe du post-traitement hydrologique d'EDF fait intervenir des variables d'état du SPQ pour arbitrer entre deux modèles de dépendance temporelle des erreurs du SPQ.

Notre modèle de post-traitement s'approche de celui de [Albert & Chib \(1993\)](#) et s'inspire de celui d'EDF : la probabilité d'appartenance à l'un ou l'autre des régimes s'y exprime en fonction des variables d'état internes du SPQ au travers d'un modèle Probit. Nous réalisons au préalable une transformation normalisante des débits (en l'occurrence logarithmique). Au sein de chaque régime, la dépendance entre les sorties du SPQ transformées et les débits transformés est supposée linéaire. Nous écrivons pour cela un modèle ARX (*Auto Regressive model with eXternal inputs*) dont les variables externes sont les sorties du SPQ passées et futures. Une méthode de sélection de variables portant sur l'ordre du modèle auto-régressif et sur les variables d'état est mise en place.

Bien que nous ayons conscience de l'intérêt de la mise à jour bayésienne d'une loi marginale climatologique, préconisée par [Krzysztofowicz \(1999\)](#) et décrite dans la section 2.2.3, nous ne procédons pas selon cette approche. Nous nous plaçons en effet dans un cadre gaussien (où donc toutes les lois conditionnelles s'expriment sous forme de régressions) et le jeu de données joint dont nous disposons inclut le jeu de données climatologiques. Notre approche, plus directe que celle de [Krzysztofowicz \(1999\)](#), amène donc ici certainement des prévisions équivalentes.

Conclusion et introduction des contributions au problème météorologique et au problème hydrologique

Dans ce chapitre, nous avons donné un aperçu des systèmes de prévision hydrologique opérationnels ou développés pour des applications opérationnelles.

Nous avons montré qu'il existait une structure commune sous-jacente à ces systèmes justifiée à la fois d'un point de vue pratique et d'un point de vue plus théorique. Cette structure fait apparaître deux lois conditionnelles qui peuvent être dérivées à partir de méthodes respectivement nommées "post-traitement météorologique" et "post-traitement hydrologique".

Deux approches se distinguent parmi les méthodes de post-traitement statistique des prévisions météorologiques (de températures et de précipitations) d'ensemble :

- Dans l'une, chaque membre de l'ensemble est considéré comme un scénario possible pour la météorologie future. Les membres de l'ensemble sont en compétition entre eux. Les méthodes de post-traitement reposant sur cette approche consistent à sélectionner le meilleur membre. Un modèle donnant la loi de la variable à prévoir conditionnellement à la valeur prise par le meilleur membre est proposé. Il s'agit des méthodes dites par habillage ou bien à noyaux. La méthode la plus emblématique de cette approche est le BMA (Raftery et al., 2005).
- Dans l'autre, c'est l'ensemble des membres et sa structure qui sont considérés comme informatifs. Les membres collaborent pour apporter de l'information sur la météorologie future. Les méthodes de post-traitement correspondant à cette approche consistent à extraire l'information issue de la structure globale de l'ensemble (sa moyenne et sa variance empiriques dans le cas gaussien). Un modèle donnant la loi de la variable à prévoir conditionnellement à ces informations résumées est ensuite proposé. La méthode la plus emblématique de cette approche est EMOS (Gneiting et al., 2005).

Un inconvénient des méthodes par habillage est qu'elles sont conçues dans l'optique de corriger la sous-dispersion des prévisions et ne sont donc pas toujours bien adaptées au post-traitement d'éventuels ensembles sur-dispersés. Une adaptation de ces méthodes l'est (Fortin et al., 2006), mais elle nécessite de renoncer à une part importante de parcimonie, une propriété pourtant attendue, les jeux de données d'entraînement étant souvent très courts. Dans ce travail de doctorat, nous avons choisi de suivre la seconde approche.

Nous proposons, dans le chapitre 3, un cadre et des modèles permettant le post-traitement de prévisions météorologiques (de précipitations et de températures) issues de plusieurs sources à la fois (de plusieurs centres de prévision météorologique et d'experts météorologues par exemple). Nous nous sommes pour cela appuyés sur la propriété statistique d'échangeabilité. À notre connaissance, de tels modèles n'ont pas encore été proposés de façon formelle dans l'approche résumant les informations portées par l'ensemble.

Le cadre choisi devrait de plus permettre, à terme, une modélisation tenant compte de la structure multi-dimensionnelle (notamment temporelle et entre variables) de la météorologie.

Les post-traitements hydrologiques visent à assortir les sorties d'un SPQ (donnant le terme central de la prévision probabiliste à produire) d'une quantification de l'incertitude qui leur est associée pour prévoir les débits (la dispersion de la distribution prédictive). Ces post-traitements doivent de plus utiliser l'information provenant des dernières observations de débits.

On trouve dans la littérature et dans les systèmes de prévision observés *in situ* deux principales façons de faire :

- La première fait intervenir une étape d'assimilation de données puis des modèles statistiques. L'emploi de copules est ensuite généralement nécessaire pour la prise en compte de la dépendance temporelle.
- La seconde consiste à proposer un modèle statistique pour l'ensemble du post-traitement.

De ces deux approches, la seconde nous paraît plus pertinente. Il apparaît toutefois que cette modélisation plus directe par un processus homogène ne permet pas de bien représenter les erreurs du SPQ. Dans le chapitre 4 nous étudions une modélisation faisant intervenir les variables d'état du SPQ pour arbitrer entre deux régimes interprétables et permettre ainsi de représenter une incertitude hydrologique à variance hétérogène.

3

Un post-traitement des prévisions météorologiques reposant sur une hypothèse d'échangeabilité

3.1	Un modèle à variables latentes pour le post-traitement des prévisions de températures	67
3.1.1	Modèle	67
3.1.2	Inférence : algorithme Expectation-Maximization (EM)	68
3.1.3	Prévisions	72
3.1.4	Expérience sur des données simulées	73
3.1.5	Application aux données d'EDF-DTG et d'Hydro-Québec	74
3.2	Extension du modèle à variables latentes pour de post-traitement des prévisions de précipitations	84
3.2.1	Modèle	84
3.2.2	Inférence	86
3.2.3	Prévisions	87
3.2.4	Expériences sur des données simulées	87
3.2.5	Application aux données d'EDF-DTG et d'Hydro-Québec	88
3.3	Post-traitement des prévisions de précipitations avec la loi des fuites	97
3.3.1	Modèle	97
3.3.2	Inférence	98
3.3.3	Application aux données d'EDF-DTG et d'Hydro-Québec	98
3.4	Discussion et perspectives	99

Dans le présent chapitre, nous posons un cadre ayant le potentiel de réunir les qualités souhaitables pour les méthodes de post-traitement météorologique, qualités identifiées dans la section 2.3 : une parcimonie de la modélisation, la possibilité d'intégrer des informations de sources diverses et la modélisation multidimensionnelle des variables météorologiques (en prenant notamment en compte leur dimension temporelle).

Le cadre proposé s'appuie sur la propriété statistique d'échangeabilité, définie section 2.3. L'idée sous-jacente à notre approche est d'utiliser le théorème de représentation de de Finetti-Savage-Hewitt (de Finetti, 1931; Hewitt & Savage, 1955). Ce théorème montre que, sous des conditions techniques de régularité, si, pour tout $K \in \mathbb{N}^*$, X_1, \dots, X_K forment un échantillon de K variables échangeables, alors il existe une quantité conditionnante aléatoire Z telle que :

$$[X_{1:K}] = \int_z [Z = z] \prod_{k=1}^K [X_k | Z = z] dz.$$

Conditionnellement à cet effet aléatoire, Z , les X_k , $k \in \{1, \dots, K\}$, sont indépendantes et identiquement distribuées.

Nous avons vu dans la section 2.1 du chapitre 2, que la distribution cible des post-traitements météorologiques au jour de prévision t_0 peut s'écrire sous la forme suivante :

$$\left[P_{t_0+(1:H)} | A_{t_0}^{1:H}, P_{t_0}^- \right],$$

où H est l'horizon de prévision et P et A désignent respectivement les variables météorologiques (températures et précipitations) et leurs prévisions (sur un ensemble de bassins versants).

Comme dans la plupart des méthodes de post-traitement météorologique, nous commençons par considérer le problème de modélisation pour une unité géographique (bassin versant) donnée, s , pour une variable météorologique (de type température ou de type précipitations) donnée, v , et pour une échéance de prévision donnée, h . Autrement dit, nous proposons des modèles dont on pourra dériver $[P_{t_0+h,s,v} | A_{t_0,s,v}^h]$ pour tout t_0 , pour tout v , pour tout s et pour $h \in \{1, \dots, H\}$, où t_0 est le temps d'émission de la prévision. Comme dans la plupart des méthodes de post-traitement météorologique, nous retrouvons ensuite $[P_{t_0+(1:H)} | A_{t_0}^{1:H}, P_{t_0}^-]$ à l'aide de copules empiriques. Notons que ce choix de modélisation ne permet pas d'utiliser les dernières observations météorologiques $P_{t_0}^-$.

3.1 Un modèle à variables latentes pour le post-traitement des prévisions de températures

Dans cette section, après une éventuelle transformation normalisante de $P_{t_0+h,s}$ et $A_{t_0,s}^h$ en respectivement Y et \mathbf{X} , nous nous plaçons dans le cas où l'on peut supposer que la variable à prévoir, Y , et les prévisions correspondantes, \mathbf{X} , sont conjointement gaussiennes.

3.1.1 Modèle

Comme énoncé précédemment, nous proposons un modèle permettant d'intégrer plusieurs sources de prévisions. Soit E le nombre de sources de prévisions considérées (ensembles) et K_e le nombre de membres de la prévision e . Dans le cas où la prévision e est déterministe, on aura simplement $K_e = 1$. Afin de simplifier l'écriture des équations, $e = 0$ désignera désormais la variable à prévoir : $Y = X_0$, avec $K_0 = 1$.

Nous proposons le modèle suivant :

$$\begin{cases} (X_{e,k}|Z) &= a_e + b_e Z + c_e \varepsilon_{e,k} \forall e, k \\ (\varepsilon_{e,k}|\omega^{-2}) &\stackrel{iid}{\sim} \mathcal{N}(0, \omega^2) \\ (Z|\omega^{-2}) &\sim \mathcal{N}(0, \lambda \omega^2) \\ \omega^{-2} &\sim \Gamma(\alpha, \beta) \end{cases},$$

où $X_{e,k}$ désigne le k^e membre de la prévision e , Z et ω^2 sont des variables latentes conditionnellement auxquelles les membres des ensembles sont indépendants, $\Gamma(\alpha, \beta)$ désigne la loi gamma de paramètres α et β (à estimer), et λ et $\{a_e, b_e, c_e\}_{e \in \{0, \dots, E\}}$ sont des paramètres à estimer. Pour des questions d'identifiabilité, on impose $b_0 = c_0 = 1$. La figure 3.1 montre le graphe orienté acyclique (DAG) correspondant à ce modèle.

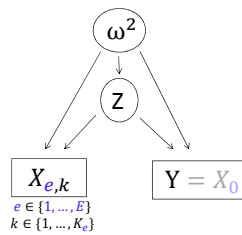


FIGURE 3.1 – Graphe orienté acyclique du modèle proposé. $X_{e,k}$ désigne le k^e membre de la prévision e et K_e le nombre de membres de la prévision e , $X_{0,K_0=1}$ désigne la variable à prévoir, et Z et ω^2 sont des variables latentes.

On peut supposer que des membres très dispersés indiquent une incertitude importante sur la variable latente Z , que l'on interprète comme "l'état sous-jacent de l'atmosphère". La deuxième variable latente, ω^2 , nous permet de tenir compte

de cette information. Précisons que la dépendance entre Z et ω^2 spécifiée dans le modèle facilite notablement l'inférence (en permettant d'utiliser la conjugaison gamma-normale). Notons par ailleurs que l'on fait ici l'hypothèse que la quantité conditionnante aléatoire émanant de la propriété d'échangeabilité, (Z, ω^{-2}) , est la même pour tous les ensembles.

Voici une interprétation naturelle des paramètres du modèle proposé :

- La différence entre le paramètre a associé aux prédicteurs météorologiques (a_e pour $e > 0$) et le paramètre a_0 (associé à la variable à prévoir), $a_e - a_0$, $e > 0$, peut être interprétée comme un biais additif des prédicteurs.
- Le rapport entre le paramètre b associé aux prédicteurs météorologiques (b_e pour $e > 0$) et le paramètre b_0 (associé à la variable à prévoir), $\frac{b_e}{b_0}$, $e > 0$, peut être interprété comme un biais multiplicatif des prédicteurs. Une absence de biais se traduit par un paramètre valant 1 (le paramètre b_0 étant fixé à 1). Les biais additifs et multiplicatifs peuvent se compenser en partie.
- Le rapport entre le paramètre c associé aux prédicteurs météorologiques (c_e pour $e > 0$) et le paramètre c_0 (associé à la variable à prévoir), $\frac{c_e}{c_0}$, $e > 0$, peut être interprété comme un biais de dispersion des prédicteurs. Un rapport supérieur à 1 s'interprète comme une sur-dispersion et un rapport inférieur à 1 comme une sous-dispersion de l'ensemble prédicteur (le paramètre c_0 étant fixé à 1).
- Le rapport $\frac{\beta}{\alpha-1}$ correspond à l'espérance de ω^2 et règle la variabilité de la quantité à prévoir Y autour de la variable latente Z . On s'attend donc à ce qu'il croisse avec l'échéance de prévision, les prévisions étant de moins en moins informatives à mesure que l'horizon de prévision est lointain.

Notons que le modèle proposé permet, en plus d'intégrer plusieurs sources de prévisions, de satisfaire nos objectifs de parcimonie : chaque source de prévisions supplémentaire n'implique l'ajout que de 3 paramètres (soit autant qu'avec la méthode du BMA décrite dans la section 2.3 par exemple).

3.1.2 Inférence : algorithme Expectation-Maximization (EM)

En supposant que les paramètres restent identiques sur une période d'apprentissage proche de la période de prédiction ou bien homogène à celle-ci, l'algorithme EM (Dempster et al., 1977) est un outil efficace pour conduire l'estimation des paramètres de ce modèle normal multivarié à effets aléatoires.

LOIS a posteriori DES VARIABLES LATENTES On s'intéresse tout d'abord au calcul des lois conditionnelles $[Z|\omega^{-2}, \mathbf{X}, Y]$ et $[\omega^{-2}|\mathbf{X}, Y]$, où \mathbf{X} désigne toujours l'ensemble des prédicteurs et n'englobe pas Y , dans l'idée d'effectuer l'étape "Expectation" de l'algorithme EM, qui consiste à calculer l'espérance de la log-vraisemblance complète par rapport à $(Z, \omega^{-2}|\mathbf{X}, Y)$.

Écrivons la loi jointe :

$$\begin{aligned} [Z, \omega^2, \mathbf{X}, Y] &= [\mathbf{X}, Y | Z, \omega^{-2}] [\omega^{-2}, Z] \\ &= [\mathbf{X} | Z, \omega^{-2}] [Y | Z, \omega^{-2}] [\omega^{-2}, Z] \end{aligned}$$

en l'interprétant comme une fonction de (ω^{-2}, Z) , cherchons à y reconnaître la loi de $(Z, \omega^{-2} | \mathbf{X}, Y)$ (à une constante multiplicative près puisque $[Z, \omega^{-2} | \mathbf{X}, Y] = [\omega^{-2}, Z, \mathbf{X}, Y] \times \left(\frac{1}{[\mathbf{X}, Y]}\right)$).

La déviance complète (moins deux fois la log-vraisemblance complète) s'écrit, aux constantes connues près :

$$\begin{aligned} &\sum_{e=0}^E \left\{ \sum_{k=1}^{K_e} (X_{e,k} - b_e Z - a_e)^2 \omega^{-2} c_e^{-2} - K_e \log(\omega^{-2}) - K_e \log(c_e^{-2}) \right\} \\ &+ Z^2 \lambda^{-1} \omega^{-2} - \log(\lambda^{-1}) - \log(\omega^{-2}) + 2\beta \omega^{-2} - 2(\alpha - 1) \log(\omega^{-2}) - 2 \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right). \end{aligned} \quad (3.1)$$

La loi conditionnelle recherchée se décompose sous la forme :

$$[Z, \omega^{-2} | \mathbf{X}, Y] = [Z | \omega^{-2}, \mathbf{X}, Y] [\omega^{-2} | \mathbf{X}, Y].$$

Supposons que nous sommes ici dans une situation de conjugaison, c'est à dire que, sachant (\mathbf{X}, Y) , nous avons encore une loi normale pour $(Z | \omega^{-2}, \mathbf{X}, Y)$ de la forme $\mathcal{N}(m', \lambda' \omega^2)$ et une loi gamma pour $(\omega^{-2} | \mathbf{X}, Y)$, $\Gamma(\alpha', \beta')$. La forme générale de la déviance de $[Z, \omega^{-2} | \mathbf{X}, Y]$ est alors :

$$(Z - m')^2 \lambda'^{-1} \omega^{-2} - \log(\omega^{-2}) - \log(\lambda'^{-1}) - 2(\alpha' - 1) \log(\omega^{-2}) + 2\beta' (\omega^{-2}).$$

Cherchons à identifier les paramètres $\alpha', \beta', \lambda', m'$ de l'équation ci-dessus à partir de la déviance de la loi conjointe donnée à l'équation (3.1).

En identifiant dans les deux expressions :

— les termes en facteur devant $\log(\omega^{-2})$, on obtient :

$$-2(\alpha' - 1) - 1 = - \sum_{e=0}^E K_e - 1 - 2(\alpha - 1)$$

$$\text{d'où : } \alpha' = \alpha + \frac{\sum_{e=0}^E K_e}{2}.$$

— les termes en facteur devant $Z^2 \omega^{-2}$, on obtient :

$$\lambda'^{-1} = \sum_{e=0}^E K_e b_e^2 c_e^{-2} + \lambda^{-1}.$$

— les termes en facteur devant $Z\omega^{-2}$, on obtient :

$$-2m'\lambda'^{-1} = -2 \sum_{e=0}^E \sum_{k=1}^{K_e} (X_{e,k} - a_e) b_e c_e^{-2}$$

$$m' = \lambda' \cdot \sum_{e=0}^E c_e^{-2} b_e \left(\sum_{k=1}^{K_e} X_{e,k} - K_e a_e \right)$$

d'où, en notant $\bar{X}_e = \frac{1}{K_e} \sum_{k=1}^{K_e} X_{e,k}$, $m' = \lambda' \cdot \sum_{e=0}^E c_e^{-2} b_e K_e (\bar{X}_e - a_e)$.

— les termes en facteur devant ω^{-2} , on obtient :

$$2\beta'^2 + m'^2 \lambda'^{-1} = 2\beta + \sum_{e=0}^E \sum_{k=1}^{K_e} c_e^{-2} (X_{e,k} - a_e)^2$$

d'où $\beta' = \beta + \frac{1}{2} \left\{ \sum_{e=0}^E \sum_{k=1}^{K_e} c_e^{-2} (X_{e,k} - a_e)^2 - m'^2 \lambda'^{-1} \right\}$.

Ainsi, nous sommes bien dans une situation de conjugaison : nous avons retrouvé l'expression de la loi conditionnelle $[Z, \omega^{-2} | \mathbf{X}, Y]$ sous la même forme gamma-normale que la loi a priori $[Z, \omega^{-2}]$.

Si on appelle ϕ la dérivée de la fonction $\log \{\Gamma(\cdot)\}$, on a alors :

$$\mathbb{E} \{ \log(\omega^{-2}) | \mathbf{X}, Y \} = -\log(\beta') + \phi(\alpha'),$$

$$\mathbb{E}(\omega^{-2} | \mathbf{X}, Y) = \frac{\alpha'}{\beta'},$$

$$\mathbb{E}(Z^2 \omega^{-2} | \mathbf{X}, Y) = \lambda' + m'^2 \frac{\alpha'}{\beta'},$$

$$\mathbb{E}(Z \omega^{-2} | \mathbf{X}, Y) = m' \frac{\alpha'}{\beta'}.$$

Nous avons donc tous les éléments nécessaires pour l'étape "Expectation" de l'algorithme EM.

MISE À JOUR DES PARAMÈTRES À CHAQUE ITÉRATION La déviance complète s'écrit, en ne gardant que les composantes concernées par les paramètres, en notant n le nombre de répétitions dans le jeu de données et en indiquant par t ces répétitions :

$$D(\alpha, \beta, \lambda, a, b, c) = \sum_{t=1}^n \left\{ Z_t^2 \lambda^{-1} \omega_t^{-2} - \log(\lambda^{-1}) - 2\alpha \log(\omega_t^{-2}) + 2\beta \omega_t^{-2} - 2\alpha \log(\beta) \right. \\ \left. + 2 \log \{ \Gamma(\alpha) \} \right. \\ \left. + \sum_{e=0}^E \left\{ \sum_{k=1}^{K_e} (X_{t,e,k} - a_e - b_e Z_t)^2 c_e^{-2} \omega_t^{-2} - \log(c_e^{-2}) \right\} \right\}.$$

En prenant l'espérance de $D(\alpha, \beta, \lambda, a, b, c)$ par rapport à $(Z, \omega^{-2} | \mathbf{X}, Y)$ puis en dérivant le résultat par rapport à chacun des paramètres, on obtient la mise à jour $(\alpha, \beta, \lambda, a, b, c) \rightarrow (\alpha_{new}, \beta_{new}, \lambda_{new}, a_{new}, b_{new}, c_{new})$:

— En dérivant par rapport à λ^{-1} , on obtient :

$$\lambda_{new} = \lambda' + \frac{\alpha'}{n} \sum_{t=1}^n \frac{m_t'^2}{\beta_t'}.$$

— En dérivant par rapport à β , on obtient :

$$\frac{\alpha_{new}}{\beta_{new}} = \frac{\sum_{t=1}^n \mathbb{E}(\omega_t^{-2} | \mathbf{X}, Y)}{n} = \frac{\alpha'}{n} \sum_{t=1}^n \frac{1}{\beta_t'}$$

$$\beta_{new} = \frac{n\alpha_{new}}{\alpha' \sum_{t=1}^n \frac{1}{\beta_t'}}.$$

— En dérivant par rapport à α , on obtient :

$$n \{ \log(\beta_{new}) - \phi(\alpha_{new}) \} = - \sum_{t=1}^n \mathbb{E} \{ \log(\omega_t^{-2} | \mathbf{X}, Y) \} = \sum_{t=1}^n \log(\beta_t') - n\phi(\alpha')$$

$$\text{d'où } \log \left(\frac{n\alpha_{new}}{\alpha' \sum_{t=1}^n \frac{1}{\beta_t'}} \right) - \phi(\alpha_{new}) = \frac{1}{n} \sum_{t=1}^n \log(\beta_t') - \phi(\alpha').$$

— En dérivant par rapport à c_e^2 , $e \in \{1, \dots, E\}$, on obtient :

$$nc_{e,new}^2 = \mathbb{E}_{(Z, \omega^{-2} | \mathbf{X}, Y)} \left\{ \sum_{t=1}^n \sum_{k=1}^{K_e} (X_{t,e,k} - a_{e,new} - b_{e,new} Z_t)^2 \omega_t^{-2} \right\}$$

$$= \sum_{t=1}^n \sum_{k=1}^{K_e} \left\{ b_{e,new}^2 \lambda' + \frac{\alpha'}{\beta_t'} \left\{ b_{e,new}^2 m_t'^2 - 2b_{e,new} m_t' (X_{t,e,k} - a_{e,new}) \right. \right.$$

$$\left. \left. + (X_{t,e,k} - a_{e,new})^2 \right\} \right\}$$

$$\text{d'où : } c_{e,new}^2 = K_e b_{e,new}^2 \lambda' + \frac{1}{n} \sum_{t=1}^n \frac{\alpha'}{\beta_t'} \sum_{k=1}^{K_e} (X_{t,e,k} - a_{e,new} - b_{e,new} m_t')^2.$$

— En dérivant par rapport à b_e , $e \in \{1, \dots, E\}$, on obtient :

$$\sum_{t=1}^n \sum_{k=1}^{K_e} \{ b_{e,new} \mathbb{E}(Z_t^2 \omega_t^{-2} | \mathbf{X}, Y) + (a_{e,new} - X_{t,e,k}) \mathbb{E}(Z_t \omega_t^{-2} | \mathbf{X}, Y) \} = 0$$

$$\sum_{t=1}^n \frac{(\bar{X}_{t,e} - b_{e,new} m_t') m_t'}{\beta_t'} - n \frac{\lambda'}{\alpha'} b_{e,new} = a_{e,new} \sum_{t=1}^n \frac{m_t'}{\beta_t'}$$

$$\sum_{t=1}^n \frac{m_t' \bar{X}_{t,e}}{\beta_t'} - b_{e,new} \left(\sum_{t=1}^n \frac{m_t'^2}{\beta_t'} - n \frac{\lambda'}{\alpha'} \right) = a_{e,new} \sum_{t=1}^n \frac{m_t'}{\beta_t'}. \quad (3.2)$$

— En dérivant par rapport à a_e , $e \in \{1, \dots, E\}$, on obtient :

$$\sum_{t=1}^n \frac{\bar{X}_{t,e}}{\beta'_t} - b_{e,new} \sum_{t=1}^n \frac{m'_t}{\beta'_t} = a_{e,new} \sum_{t=1}^n \frac{1}{\beta'_t}$$

et en rapprochant cette équation de l'équation (3.2) on obtient :

$$b_{e,new} = \frac{\frac{D_e}{B} - \frac{C_e}{G}}{\frac{G}{B} - \frac{H}{G} - \frac{n\lambda'}{G\alpha'}}$$

$$a_{e,new} = \frac{D_e}{B} - b_{e,new} \frac{G}{B}$$

où l'on note : $G = \sum_{t=1}^n \frac{m'_t}{\beta'_t}$, $B = \sum_{t=1}^n \frac{1}{\beta'_t}$, $C_e = \sum_{t=1}^n \frac{m'_t \bar{X}_{t,e}}{\beta'_t}$, $D_e = \sum_{t=1}^n \frac{\bar{X}_{t,e}}{\beta'_t}$
et $H = \sum_{t=1}^n \frac{m_t'^2}{\beta_t'}$.

3.1.3 Prévisions

Pour un nouveau temps, t' , on a $\mathbf{X} = \mathbf{x}_{t'}$ et on cherche la loi de $(Y_{t'} | \mathbf{X} = \mathbf{x}_{t'})$. De même que dans la première étape de l'inférence EM (premier paragraphe de la section 3.1.2), sachant $\mathbf{X} = \mathbf{x}_{t'}$, nous aurons encore une loi normale pour $(Z | \omega^{-2}, \mathbf{X} = \mathbf{x}_{t'})$, de la forme $\mathcal{N}(m_{t'}'', \lambda'' \omega^2)$, et une loi gamma pour $(\omega^{-2} | \mathbf{X} = \mathbf{x}_{t'})$, $\Gamma(\alpha'', \beta_{t'}'')$. On identifie terme à terme les paramètres $m_{t'}''$, λ'' , α'' et $\beta_{t'}''$ dans les expressions de la déviance :

$$\alpha'' = \alpha + \frac{\sum_{e=1}^E K_e}{2}$$

$$\lambda''^{-1} = \sum_{e=1}^E K_e b_e^2 c_e^{-2} + \lambda^{-1}$$

$$m_{t'}'' = \lambda'' \cdot \sum_{e=1}^E c_e^{-2} b_e K_e (\bar{x}_{e,t'} - a_e)$$

$$\beta_{t'}'' = \beta + \frac{1}{2} \left\{ \sum_{e=1}^E \sum_{k=1}^{K_e} c_e^{-2} (x_{e,k,t'} - a_e)^2 - m_{t'}''^2 \lambda''^{-1} \right\}.$$

$(Y | \mathbf{X} = \mathbf{x}_{t'})$ suit alors une loi de Student de paramètre d'échelle $\sqrt{\frac{(\lambda''+1)\beta_{t'}''}{\alpha''}}$, de paramètre de localisation $a_0 + m_{t'}''$ et à $2\alpha''$ degrés de liberté. En effet, on a :

$$(Y|Z) = a_0 + Z + \varepsilon$$

$$(Z|\omega^2, \mathbf{X} = \mathbf{x}_{t'}) \sim \mathcal{N}(m_{t'}'', \lambda'' \omega^2)$$

$$(\varepsilon|\omega^2) \sim \mathcal{N}(0, \omega^2)$$

donc $\left(\frac{Y}{\sqrt{\lambda''+1}} \mid \omega^2, \mathbf{X} = \mathbf{x}_{t'}\right) \sim \mathcal{N}\left(\frac{a_0+m_{t'}}{\sqrt{\lambda''+1}}, \omega^2\right)$.

De plus, $(\omega^{-2} \mid \mathbf{X} = \mathbf{x}_{t'}) \sim \Gamma(\alpha'', \beta_{t'})$. D'où le résultat annoncé (d'après les formules de conjugaison dans le cas d'un modèle normal à moyenne fixe et à variance de loi inverse-gamma).

Notons que l'espérance prédite pour $(Y \mid \mathbf{X} = \mathbf{x}_{t'})$ est donnée par :

$$\mathbb{E}(Y \mid \mathbf{X} = \mathbf{x}_{t'}) = a_0 + \lambda'' \sum_{e=1}^E \frac{b_e}{c_e^2} K_e (\bar{x}_{e,t'} - a_e). \quad (3.3)$$

et la variance prédite pour $(Y \mid \mathbf{X} = \mathbf{x}_{t'})$ est telle que :

$$\mathbb{V}(Y \mid \mathbf{X} = \mathbf{x}_{t'}) \propto \beta + \frac{1}{2} \left\{ \sum_{e=1}^E c_e^{-2} \sum_{k=1}^{K_e} (x_{e,k,t'} - a_e)^2 - \lambda'' \left\{ \sum_{e=1}^E \frac{b_e}{c_e^2} K_e (\bar{x}_{e,t'} - a_e) \right\}^2 \right\}.$$

On retrouve ainsi une distribution prédictive similaire à celle issue de la méthode EMOS, décrite dans la section 2.3, où l'espérance prédictive s'exprime linéairement en fonction de la moyenne de l'ensemble et la variance prédictive s'exprime en fonction de la variance de l'ensemble.

Dans l'équation (3.3), il apparaît qu'un membre de la prévision e a d'autant plus d'impact sur la prévision que le rapport $\frac{b_e}{c_e^2}$ est grand. Nous définissons donc la *contribution* d'une source de prévision e à la prévision finale par :

$$contrib_e = \frac{\frac{b_e}{c_e^2}}{\sum_{e'=1}^{e'=E} K_{e'} \frac{b_{e'}}{c_{e'}^2}}.$$

3.1.4 Expérience sur des données simulées

Nous vérifions dans cette section que nous serons à même d'estimer correctement les paramètres de notre modèle avec l'inférence par algorithme EM décrite. Nous simulons pour cela des données selon le modèle décrit dans la section précédente avec des paramètres réalistes pour nos jeux de données de températures (c'est-à-dire proches de ceux que l'on apprendrait sur ces jeux de données) :

$$\begin{aligned} K &= (1, 10, 35, 1) \\ a &= (0, 1, 0.3, -0.1) \\ b &= (1, 1.1, 0.8, 0.9) \\ c &= (1, 0.8, 0.9, 1.1) \\ \alpha &= 2.5, \quad \beta = 3, \quad \lambda = 20 \end{aligned}$$

où le premier élément de chaque vecteur est celui indicé par 0, relatif aux observations. Dans cet exemple, la prévision correspondant à l'indice 3 est déterministe

($K_3 = 1$). Nous simulons selon ce modèle 100 jeux de données de 500 éléments chacun (car le nombre de jours disponibles pour l'inférence en pratique est de cet ordre de grandeur) (nous effectuons donc 50000 simulations selon le modèle). L'algorithme EM précédemment décrit est ensuite utilisé sur chacun des 100 jeux de données pour l'inférence du modèle. La valeur initiale des paramètres est choisie par une méthode des moments. Notons que les paramètres b_0 et c_0 sont contraints à valoir 1 et ne sont donc pas estimés. L'algorithme effectue des itérations jusqu'à ce que la différence relative entre un paramètre à une itération et ce même paramètre à l'itération suivante soit au plus de $5 \cdot 10^{-4}$. Les paramètres estimés ainsi obtenus sur les 100 jeux de données sont illustrés sous la forme de boîtes à moustaches sur la figure 3.2. Ces résultats valident notre méthode d'inférence pour le type de données dont nous disposons.

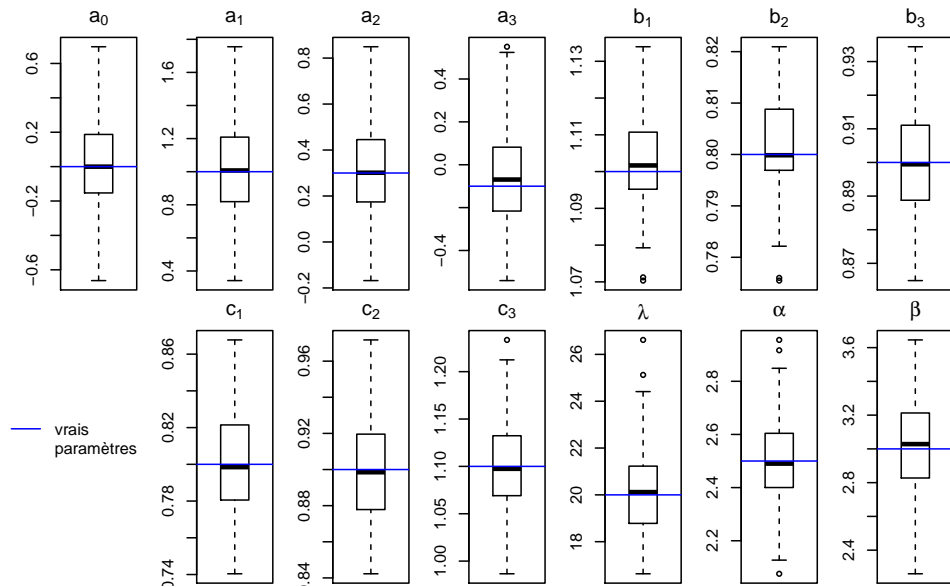


FIGURE 3.2 – Paramètres obtenus avec l'algorithme EM pour 100 simulations de 500 répétitions selon le modèle. Les paramètres utilisés pour la simulation sont figurés par les lignes bleues.

3.1.5 Application aux données d'EDF-DTG et d'Hydro-Québec

Nous utilisons le modèle décrit ci-dessus pour le post-traitement des prévisions de températures des jeux de données de prévisions météorologiques “EDF-DTG 2” et “Hydro-Québec” décrits dans la section 1.2.2. Ils nous sont fournis pour respectivement 3 des 15 bassins versants français et pour 5 des 8 bassins versants québécois à l'étude dans cette thèse.

Nous n'utiliserons pas le jeu de données “EDF-DTG 1” car, comme exposé dans le chapitre 1 : (1) Il ne comprend qu'une source de prévision dans le cas des tempéra-

tures (les prévisions provenant du CEPMMT-EPS). En effet, les prévisions analogues (ANA) ne sont disponibles que dans le cas des précipitations. (2) La méthode de production des prévisions d'ensemble du CEPMMT a fait l'objet de 6 changements sur la période 2006-2008.

En raison de changements majeurs survenus dans la façon de produire les prévisions CMC-EPS en 2011 et début 2013, nous n'utiliserons pas non plus les données de prévisions antérieures à début 2013 du jeu de données québécois. Les prévisions CEPMMT-EPS fournies pour le Québec comportant, de plus, des anomalies à l'échéance de prévision 10 jours, nous étudions ici les 9 premières échéances seulement.

Par ailleurs, les prévisions expertisées du jeu de données "EDF-DTG 2" n'étant fournies que pour quelques jours par semaine, à un horizon de 7 jours et n'étant pas produites de façon homogène, nous ne les utilisons pas ici.

Enfin, la méthode de post-traitement ici étudiée faisant l'hypothèse que la variable à prévoir est gaussienne, nous choisissons de travailler avec les données relatives aux températures (température moyenne journalière dans le cas des données françaises, température maximale journalière dans le cas des données québécoises).

Une première étape consiste à vérifier l'hypothèse d'échangeabilité des membres des ensembles. Comme énoncé dans la section 2.3, les membres des prévisions CEPMMT-EPS et NCEP-GEFS peuvent *a priori* être supposés échangeables, mais ce n'est pas le cas des membres de l'ensemble CMC-EPS. Dans ce cas, une solution consiste à identifier des sous-groupes de membres au sein desquels l'échangeabilité est une hypothèse acceptable, puis à traiter chacun de ces sous-groupes comme un ensemble à part entière dans notre méthode de post-traitement. Dans le cas de l'ensemble CMC-EPS, une analyse rapide de l'inertie montre que, dans le cas des précipitations, l'hypothèse d'échangeabilité est acceptable au sein des membres pairs et au sein des membres impairs. En effet, la variance cumulée des membres de ses sous-ensembles n'est pas significativement augmentée par des permutations aléatoires de leurs membres d'un pas de temps à l'autre. Dans le cas des prévisions de températures, par contre, nous n'avons pas pu identifier de partition de ce type. Nous avons tout de même recours à cette hypothèse pour travailler sur les données de températures afin de préserver une parcimonie de la modélisation, mais notons que, si une réanalyse conséquente de prévisions d'ensemble du CMC-EPS était disponible, il serait intéressant d'étudier de façon plus approfondie l'échangeabilité, totale ou partielle, au sein de cet ensemble.

Illustration sur le bassin versant de Manic 2

Prenons pour commencer l'exemple du bassin versant québécois de Manic 2. Nous disposons, sur ce bassin versant, des prévisions d'ensemble CMC-EPS, NCEP-GEFS et CEPMMT-EPS et souhaitons utiliser notre méthode pour les post-traiter.

Nous apprenons les paramètres du modèle de prévision à partir du jeu de données de l'année 2013, dans l'idée d'ensuite utiliser le jeu de données de l'année 2014 pour

évaluer les prévisions qui en seront issues.

Les paramètres obtenus sont représentés sur la figure 3.3 en fonction de l'échéance de prévision (le problème étant traité échéance de prévision par échéance de prévision, on procède à l'inférence pour chaque échéance).

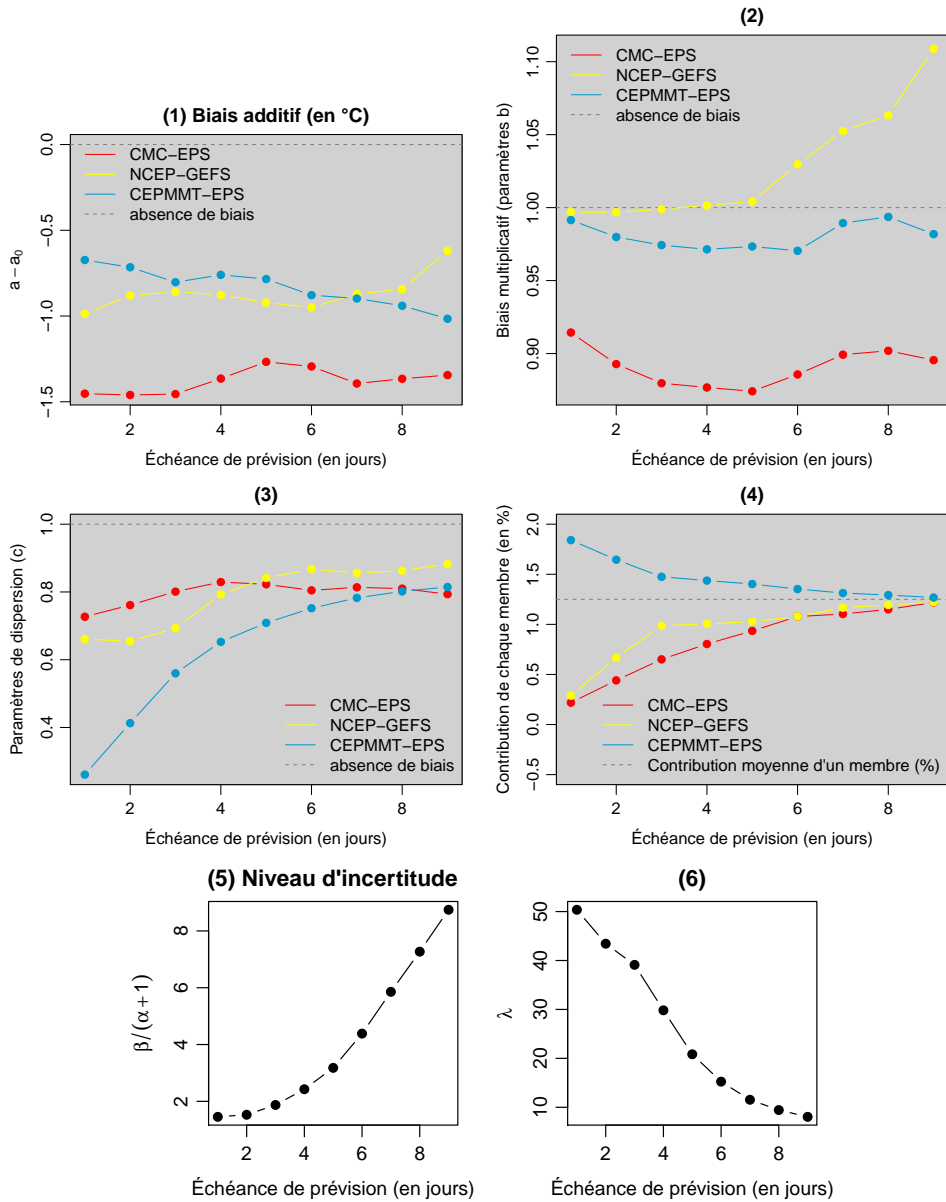


FIGURE 3.3 – Paramètres obtenus avec l’algorithme EM en fonction de l’échéance de prévision pour le bassin versant de Manic 2 avec $E = 3$ sources de prévisions (les ensembles CMC-EPS, NCEP-GEFS et CEPMMT-EPS), avec un apprentissage sur l’année 2013.

Le graphique (1) de la figure illustre le biais additif de la prévision, donné par la différence entre les paramètres a_1, a_2 et a_3 , relatifs aux 3 ensembles, et le paramètre a_0 , correspondant aux observations. On remarque que tous les ensembles, pour le bassin versant de Manic 2, sont biaisés négativement pendant l'année 2013 (dans ce jeu de données), et ce quelle que soit l'échéance de prévision.

Le graphique (2) illustre les résultats d'inférence pour les paramètres b . Ces paramètres peuvent être interprétés comme les paramètres relatifs aux biais multiplicatifs des prévisions. Le biais multiplicatif des prévisions est négatif dans le cas des ensembles CMC-EPS et CEPMMT-EPS, amplifiant encore le diagnostique issu du premier graphique. En revanche, cette figure fait apparaître, à partir de l'échéance 6 jours, un biais multiplicatif positif de l'ensemble NCEP-GEFS pouvant potentiellement compenser le biais additif négatif préalablement observé.

Le graphique (3) montre les résultats pour les paramètres c réglant la dispersion inter-membres de l'ensemble relativement à celle de l'observation. On remarque que tous les paramètres c croissent avec l'échéance de prévision et prennent des valeurs inférieures à 1. Les ensembles sont donc sous-dispersifs, et ce d'autant plus que l'échéance de prévision considérée est courte.

De même que les paramètres c , le rapport $\frac{\beta}{\alpha-1}$, illustré sur le graphique (5), croît avec l'échéance de prévision. Ce rapport correspond à l'espérance de ω^2 et règle donc la variabilité de la quantité à prévoir Y autour de la variable latente Z . L'incertitude portant sur Y croît donc avec l'échéance de prévision. Nous nous attendions à ce résultat, les prévisions étant de moins en moins informatives à mesure que l'horizon de prévision est lointain.

Le graphique (4) montre la contribution, $contrib_e$, de chaque membre d'une source de prévision e à la prévision finale. Elle est donnée en pourcentage de la contribution totale des prévisions d'ensemble. On remarque que la contribution des membres du CEPMMT-EPS est beaucoup plus importante que celle des membres des autres ensembles et que, plus l'horizon est lointain, plus les prévisions deviennent équivalentes en matière de contributions. L'ensemble CEPMMT comprenant de plus 50 membres tandis que les autres en comprennent chacun 20, la majeure partie de l'information provient de l'ensemble CEPMMT. Cette observation est aussi illustrée par la figure 3.7, qui montre la contribution relative de chaque source de prévision moyennée sur les 9 échéances de prévision, donnée pour la prévision e (et pour une échéance de prévision) par $K_e \times contrib_e$. Si l'information portée par chaque ensemble était la même, la contribution relative des prévisions CMC-EPS et NCEP-GEFS réunies devrait être d'environ 0.44. Elle se situe autour de 0.35 sur les 5 bassins versants québécois, les membres des ensembles CMC-EPS et NCEP-GEFS sont donc sous-pondérés au profit des membres du CEPMMT-EPS.

Les paramètres moyennés sur les échéances obtenus pour Manic 2 sont mis en évidence en rouge sur la figure 3.6.

Nous testons ensuite notre méthode de post-traitement sur l'année 2014. La figure 3.4 illustre un exemple de prévisions des températures maximales journalières

sur le bassin versant de Manic 2. Les prévisions représentées sont issues du post-

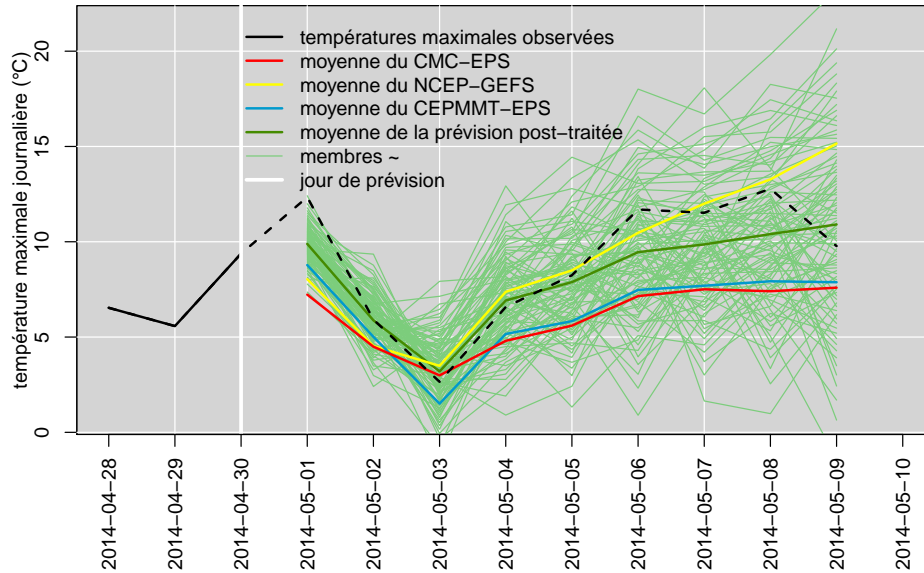


FIGURE 3.4 – Exemple de prévision émise pour les températures maximales journalières du bassin versant de Manic 2 en utilisant notre méthode de post-traitement prenant en intrants les ensembles NCEP-GEFS, CMC-EPS et CEPMMT-EPS. Les scénarios prédictifs sont obtenus à partir des prévisions par échéance de prévision à l’aide de l’ECC-Q. La température maximale journalière à prévoir est figurée par la ligne tiretée noire.

traitement que nous proposons et émises le 30 avril 2014 pour les 9 jours qui suivent. Cette situation correspond à celle illustrée sur la figure 1.6. Les températures maximales sont représentées en noir, celles à prévoir par une ligne tiretée. Figurent également sur le graphique les moyennes respectives des ensembles intervenant dans la prévision (en bleu, en jaune et en rouge). La moyenne de la prévision issue du post-traitement, qui utilise les 3 sources de prévisions disponibles, est en vert. Il s’agit manifestement d’un compromis entre les prévisions CEPMMT-EPS et CMC-EPS, d’une part, et NCEP-GEFS, d’autre part.

Les (90) scénarios (“membres”) de la prévision illustrés par des traits fins verts sont obtenus à partir des prévisions produites échéance de prévision par échéance de prévision en ayant recours à une copule empirique (l’*Ensemble Copula Coupling* ECC-Q, décrite page 42). Cette méthode consiste ici à copier la structure de rang du grand ensemble, un ensemble constitué des ensembles CMC-EPS, NCEP-GEFS et CEPMMT-EPS réunis. Nous obtenons ainsi des scénarios spatiaux, temporels (et multi-variables) et sommes donc à même, par exemple, d’émettre des prévisions de températures cumulées dans le temps (en estimant une distribution à partir des températures cumulées de chaque scénario).

D’après les histogrammes de rang, les prévisions obtenues (sur l’année 2014, pour

le bassin versant de Manic 2) peuvent être considérées comme fiables quelle que soit l'échéance de prévision (non illustré).

Bien que nous ayons choisi une méthode d'inférence minimisant le score ignorance, nous évaluons les prévisions avec le CRPS afin de pallier le problème de scores ignorance parfois infinis.

L'évaluation des prévisions (sur l'année 2014) avec le CRPS (score décrit dans la section 1.3.3) est illustrée par la figure 3.5 pour le bassin versant de Manic 2 en fonction de l'échéance de prévision. Sont comparées sur cette figure :

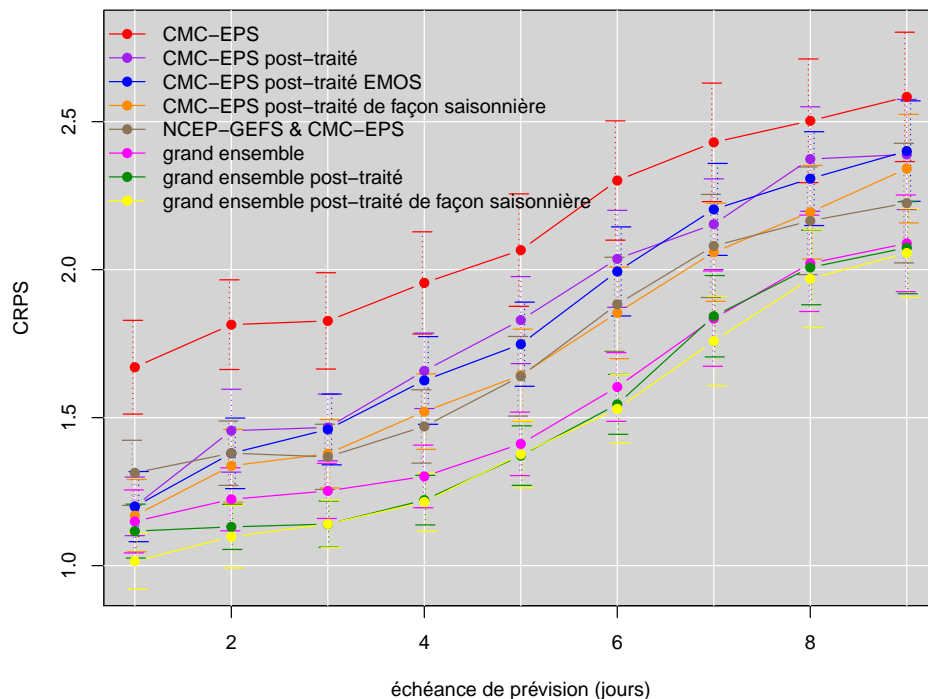


FIGURE 3.5 – CRPS (en °C) relatifs aux prévisions de températures maximales pour le bassin versant de Manic 2 en fonction de l'échéance de prévision et intervalles bootstrap associés, pour un apprentissage sur l'année 2013 suivi d'un test sur l'année 2014. Un CRPS est d'autant meilleur qu'il est petit.

- les prévisions obtenues en considérant les membres de l'ensemble CMC-EPS comme des simulations selon une distribution prédictive, **CMC-EPS**,
- les prévisions obtenues en post-traitant le CMC-EPS avec notre méthode en apprenant les paramètres du modèle sur l'année 2013 (dans ce cas, $E = 1$), **CMC-EPS post-traité**,
- les prévisions obtenues en post-traitant le CMC-EPS avec notre méthode en apprenant les paramètres du modèle par une inférence saisonnière et glissante (pour émettre une prévision le 30 avril 2014, par exemple, nous effectuons l'inférence par algorithme EM sur la période comprise entre le 30 mars 2013

- et le 30 mai 2013 et entre le 30 mars 2014 et le 30 avril 2014), **CMC-EPS post-traité de façon saisonnière**,
- les prévisions obtenues en post-traitant le CMC-EPS avec cette fois la méthode EMOS, décrite dans le chapitre 2 (mais avec une paramétrisation issue de la maximisation de la vraisemblance, et non du CRPS comme préconisé dans la méthode EMOS originale), **CMC-EPS post-traité EMOS**,
 - les prévisions obtenues en considérant les membres des ensembles CMC-EPS et NCEP-GEFS réunis comme des simulations selon une distribution prédictive, **CMC-EPS & NCEP-GEFS**,
 - les prévisions obtenues en considérant les membres des ensembles CMC-EPS, NCEP-GEFS et CEPMMT-EPS réunis en un grand ensemble comme des simulations selon une distribution prédictive, **grand ensemble**,
 - les prévisions obtenues en post-traitant les ensembles CMC-EPS, NCEP-GEFS et CEPMMT-EPS considérés comme 3 sources de prévisions distinctes ($E = 3$) selon la méthode de post-traitement que nous proposons, selon l'exemple traité ci-dessus, dont les paramètres correspondants sont illustrés sur la figure 3.3, **grand ensemble post-traité**,
 - les résultats de ce même post-traitement effectué de façon saisonnière, **grand ensemble post-traité de façon saisonnière**.

On constate que les prévisions CMC-EPS sont grandement améliorées (selon le CRPS) par le post-traitement proposé, et encore davantage si le post-traitement est saisonnier. Cela corrobore l'intuition selon laquelle la relation entre les prévisions d'ensemble et la météorologie est saisonnière. Ce post-traitement (uni-ensemble) du CMC-EPS permet d'obtenir des prévisions à peu près aussi bonnes qu'en réunissant les ensembles CMC-EPS et NCEP-GEFS. Sa performance, telle qu'évaluée par le CRPS, est du même ordre de grandeur que celle de la méthode EMOS, ce qui était attendu, ces méthodes résultant en des prévisions très proches. Enfin, le post-traitement multi-ensemble permet d'améliorer les prévisions du grand ensemble, en particulier sur les premières échéances de prévision.

Les résultats des prévisions de températures cumulées à l'horizon 9 jours, évaluées par le CRPS, sont illustrés sur la figure 3.8. Ceux relatifs au bassin versant de Manic 2 y figurent dans la première colonne. On remarque que les prévisions post-traitées ressortent avantagées de ce passage de l'unidimensionnel au temporel : leurs scores en sont améliorés relativement à ceux des autres prévisions. ce qui est étonnant puisqu'on ne fait ici que mimer la structure temporelle des ensembles.

Extension des résultats à l'ensemble des bassins versants

La figure 3.6 illustre les paramètres obtenus pour les 5 bassins versants québécois et les 3 bassins versants français puis moyennés sur les 9 échéances de prévision. On remarque que les paramètres de biais multiplicatifs et les paramètres dispersifs des sources de prévisions sont ordonnés de façon similaires quel que soit le bassin

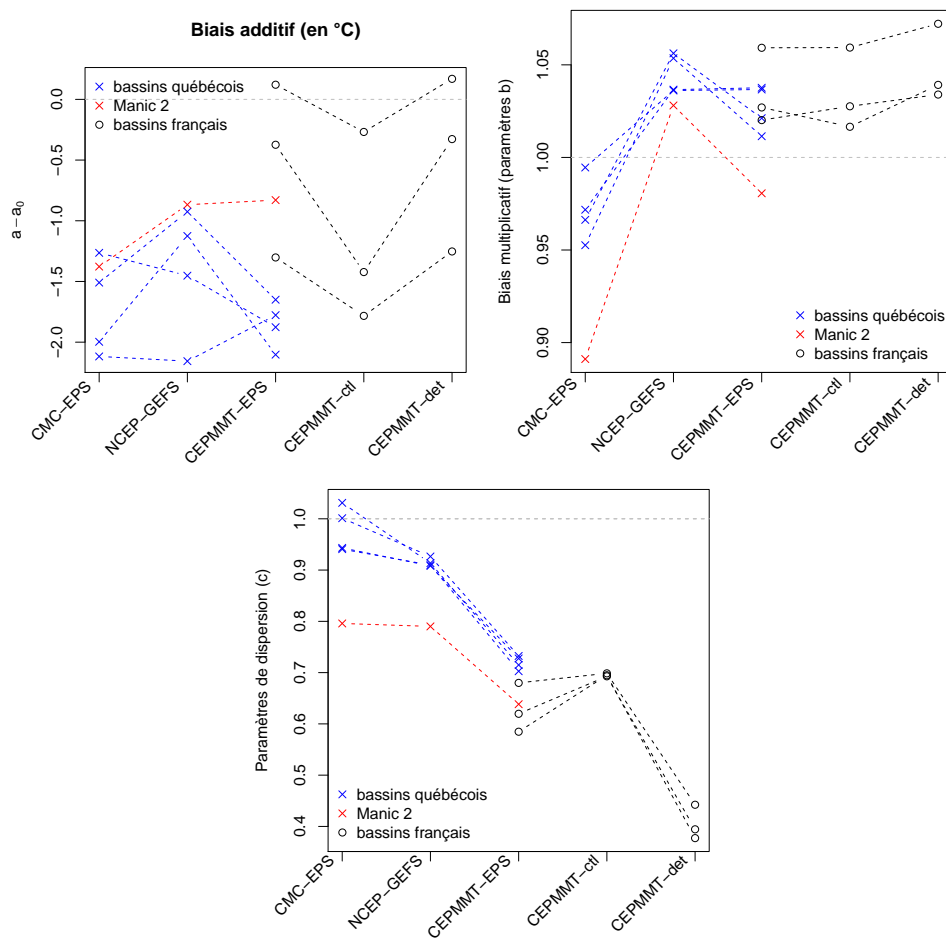


FIGURE 3.6 – Paramètres obtenus avec l’algorithme EM moyenné sur les échéances de prévision et avec un apprentissage sur l’année 2013 pour les 5 bassins versants québécois et sur les années 2011 à 2013 pour les 3 bassins versants français de l’étude.

versant. Les paramètres λ , α et β sont du même ordre de grandeur que dans le cas de Manic 2 sur les autres bassins versants (non montré). Le post-traitement opérera donc de façon similaire sur les huit bassins versants ici considérés.

La figure 3.7 montre que les observations faites sur les contributions relatives des sources de prévisions pour Manic 2 s’étendent aux autres bassins versants québécois. Sur les bassins versants français, on note que les prévisions déterministes du CEPMMT (CEPMMT-det) sont particulièrement informatives comparées aux membres de l’ensemble CEPMMT-EPS, y compris à son membre contrôle (CEPMMT-ctl) : la contribution relative de CEPMMT-det est de l’ordre de 5% alors qu’il ne représente que 2% des membres intervenant dans la prévision post-traitée.

D’après les histogrammes de rang, les prévisions obtenues (sur l’année 2014)

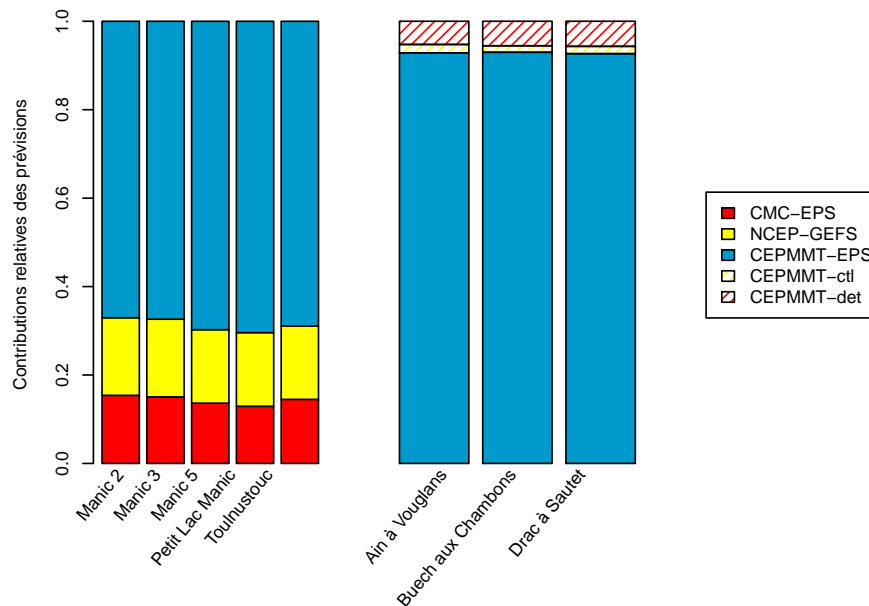


FIGURE 3.7 – Contributions relatives des sources de prévision selon le modèle à deux variables latentes, moyennées sur les 9 échéances de prévisions. À gauche : pour l'année 2013 et sur les bassins versants québécois. À droite : pour les années 2011 à 2013 et sur les bassins versants français.

peuvent être considérées comme fiables quel que soit le bassin versant considéré (non illustré).

La figure 3.8 étend les résultats en matière de performance des prévisions de cumuls de températures aux autres bassins versants de l'étude. Comme attendu, les prévisions sont, selon le CRPS, toujours meilleures après post-traitement qu'avant post-traitement et tenir compte de la saisonnalité améliore les résultats (ce qui est illustré sur les bassins versants québécois). De plus, les scores obtenus après un post-traitement saisonnier des prévisions CMC-EPS (sur les bassins versants québécois) sont généralement meilleurs que ceux des prévisions issues du grand ensemble non post-traité. Quelques résultats surprenants apparaissent toutefois : il arrive que les prévisions du grand ensemble post-traité soient moins bonnes que celles du CMC-EPS post-traité (nous le constatons sur les bassins versants de Toulousteouc et de Manic 3) et, sur le bassin versant du Buech aux Chambons, les prévisions du CEPMMT-EPS seul post-traité sont meilleures que les 3 prévisions du CEPMMT réunies et post-traitées. Toutefois, dans le cas des bassins versants québécois, cette observation n'est valable que sur les prévisions de températures cumulées : échéance de prévision par échéance de prévision, les prévisions issues du grand ensemble post-traité sont meilleures que celles issues de CMC-EPS seul post-traité. Cette dégradation est donc due à une mauvaise prise en compte de la dépendance temporelle.

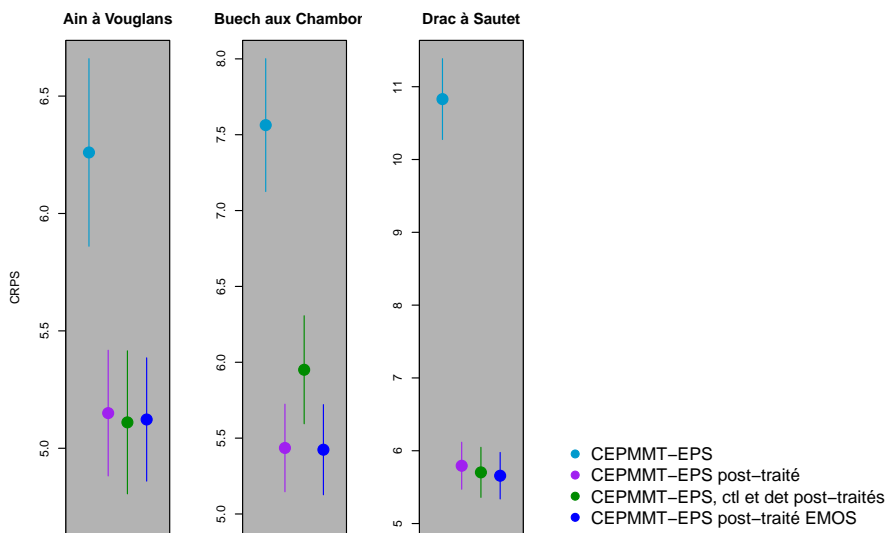
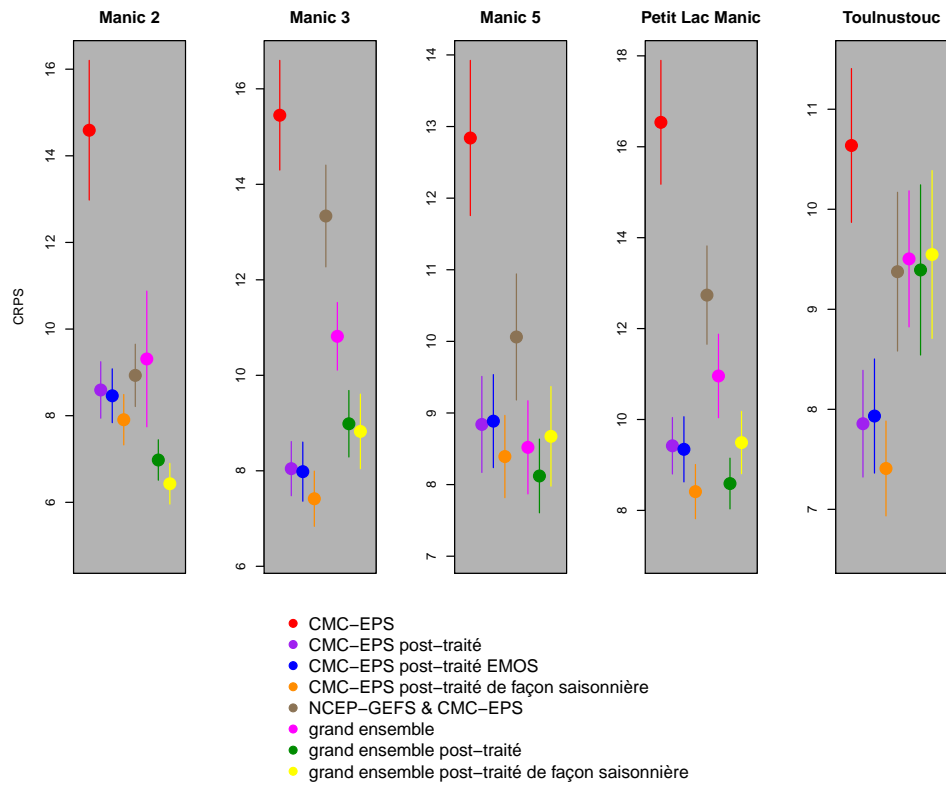


FIGURE 3.8 – CRPS (en °C) relatifs aux prévisions de températures cumulées dans le temps et intervalles bootstrap associés, pour un apprentissage sur l’année 2013 suivi d’un test sur l’année 2014 pour les bassins versants québécois (haut), et un apprentissage sur les années 2011 à 2013 suivi d’un test sur les années 2014 et 2015 pour les bassins versants français (bas).

3.2 Extension du modèle à variables latentes pour de post-traitement des prévisions de précipitations

Les investigations de cette section visent à proposer une adaptation de la méthode de post-traitement fondée sur l'hypothèse d'échangeabilité pour les variables de type précipitations. Ces variables ne peuvent pas être supposées issues d'une loi gaussienne : elles sont mixtes avec une composante discrète en zéro et une composante continue positive. Nous souhaitons, à terme, être à même de post-traiter conjointement des variables de type température et des variables de type précipitations, c'est une des raisons pour lesquelles nous cherchons à rester dans le cadre commode de la famille gaussienne.

3.2.1 Modèle

L'idée sous-jacente au modèle que nous proposons est que les précipitations que nous observons, des variables avec une composante discrète (en zéro) et une composante continue (supérieure à zéro), seraient une censure à gauche d'une variable continue. Des travaux reposent déjà sur cette idée pour mettre au point une méthode de post-traitement des prévisions de précipitations (Schultz et al., 2008). Dans ces travaux, la variable continue associée aux précipitations a une interprétation physique et est appelée *pseudo-précipitations*.

Dans notre modèle, nous laissons de côté toute interprétation physique et supposons que ces pseudo-précipitations sont gaussiennes à une transformation près, $f_{\mathcal{N}}$. Le modèle que nous proposons pour les pseudo-précipitations normalisées est le même que celui proposé pour les variables de type températures. Soient Y' et \mathbf{X}' respectivement la variable précipitations (à prévoir) et ses prévisions. Soient Y et \mathbf{X} les pseudo-précipitations normalisées correspondantes. Le modèle que nous proposons est le suivant : $\forall e, k$,

$$\begin{cases} (X_{e,k}|Z) & = a_e + b_e Z + c_e \varepsilon_{e,k} \\ (X'_{e,k}|X_{e,k}) & = \mathbb{I}_{X_{e,k} > \nu} f_{\mathcal{N}}^{-1}(X_{e,k}) \\ (\varepsilon_{e,k}|\omega^{-2}) & \stackrel{iid}{\sim}_{e,k} \mathcal{N}(0, \omega^2) \\ (Z|\omega^{-2}) & \sim \mathcal{N}(0, \lambda\omega^2) \\ \omega^{-2} & \sim \Gamma(\alpha, \beta) \end{cases}$$

où Z et ω^2 sont des variables latentes conditionnellement auxquelles les membres de ensembles sont indépendants, et $\alpha, \beta, \lambda, \nu$ et $\{a_e, b_e, c_e\}_{e \in \{0, \dots, E\}}$ sont des paramètres à estimer. Pour des questions d'identifiabilité, on impose à nouveau $b_0 = c_0 = 1$ et $a_0 = 0$. $(X_{e,k})_{e \in \{0, \dots, E\} k \in \{1, \dots, K_e\}}$ sont alors des variables latentes (toutefois observées quand elles sont supérieures à ν). La figure 3.9 montre le graphe orienté acyclique (DAG) correspondant à ce modèle.

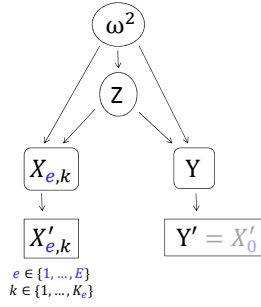


FIGURE 3.9 – Graphe orienté acyclique du modèle proposé. $X'_{e,k}$ désigne le k^e membre de la prévision e , K_e désigne le nombre de membres de la prévision e , $X'_{0,K_0=1}$ désigne la variable à prévoir et \mathbf{X} , Y , Z et ω^2 sont des variables latentes (partiellement observées dans le cas de \mathbf{X} et Y).

TRANSFORMATION NORMALISANTE La question de la transformation normalisante, $f_{\mathcal{N}}$, est importante. Notons qu'il faut veiller à ce que la transformation inverse, $f_{\mathcal{N}}^{-1}$, soit à valeurs positives sur $]\nu, +\infty[$.

Une première possibilité consiste à choisir la fonction identité, auquel cas ν correspond au plus petit cumul de précipitations journalier non nul observé jusqu'ici (très proche de zéro dans nos données) et on suppose que la variables précipitation se comporte suivant une loi normale censurée à gauche en zéro, ce qui est peu crédible.

Une autre possibilité consiste à utiliser la transformation proposée par Allard (2012) pour les variables de type précipitations. Nous en considérons une forme simplifiée donnée par :

$$f_{\mathcal{N}}^{-1}(x) = \psi_{\iota, \kappa, \nu, \gamma}(x) = \iota + \kappa \left(e^{\gamma(x-\nu)} - 1 \right),$$

où ι , κ et γ sont des paramètres à estimer, ν est le même paramètres que précédemment et x est la réalisation d'une variable normale centrée et réduite, X . $f_{\mathcal{N}}^{-1}(X)$ est alors supposée se comporter comme une variable pouvant représenter les précipitations. Toutefois, dans nos expériences sur les données réelles, les simulations selon la distribution prédictive résultante comprennent souvent des précipitations extrêmes irréalistes. Nous ne développons donc pas davantage cette idée dans ce manuscrit.

Une dernière possibilité de transformation, celle retenue pour notre travail, consiste à simplement appliquer une puissance à la variable précipitations : $f_{\mathcal{N}}(x') = x'^{\gamma}$, où γ est un paramètre à estimer. Effectuer cette transformation revient essentiellement à effectuer la transformation Box-Cox décrite dans la section 2.2.2.

Nous choisissons d'utiliser le même paramètre de transformation quel que soit l'ensemble de précipitations considéré, celui estimé pour la variable à prévoir : $f_{\mathcal{N}}$ ne dépend pas de e dans notre modèle. Il est ainsi possible d'utiliser le large historique d'observations de précipitations dont nous disposons pour l'inférence du ou des paramètre(s) de transformation (en l'occurrence un unique paramètre).

3.2.2 Inférence

Paramètres de normalisation

Nous faisons dans ce travail une approximation en fixant $\nu = 0$ (cela correspond à une sensibilité très fine du pluviomètre, avec des valeurs non nulles de précipitations observées très proches de 0).

Nous apprenons le paramètre γ au préalable à partir de l'historique météorologique. Il s'agit de trouver une valeur de γ telle que $(Y'^\gamma | Y' > 0)$ puisse être supposée suivre une loi normale tronquée à gauche en zéro.

La log-vraisemblance, selon ce modèle, s'écrit, en supposant l'indépendance temporelle du phénomène de précipitations :

$$\begin{aligned} \log [Y' = y'] &= \sum_{t \text{ tel que } y'_t > 0} \log [Y' = y'_t] \\ &= \sum_{t \text{ tel que } y'_t > 0} \log [Y'^\gamma = y_t'^\gamma] + (\gamma - 1) \log (y'_t) + \log (\gamma) \\ &= \sum_{t \text{ tel que } y'_t > 0} \log \mathcal{N}_{>0} (y_t'^\gamma; \mu, \sigma^2) + (\gamma - 1) \log (y'_t) + \log (\gamma), \end{aligned}$$

où μ et σ sont (aussi) des paramètres à estimer et $\mathcal{N}_{>0} (x; \mu, \sigma^2)$ est la densité en x d'une loi gaussienne d'espérance μ et de variance σ^2 tronquée à gauche en zéro.

En pratique, nous estimons les paramètres μ, σ^2 et γ maximisant cette vraisemblance par une méthode d'optimisation numérique (celle de Nelder-Mead implémentée sur R) puis utilisons le paramètre γ ainsi obtenu par la suite.

Autres paramètres : algorithme EM stochastique

Pour réaliser l'étape "Expectation" de l'algorithme EM, il nous faut calculer les lois des variables latentes conditionnellement à (\mathbf{X}', Y') . Le calcul ne conduit pas à une expression analytique de ces lois.

En revanche, on connaît la loi de $(\mathbf{X}, Y | Z, \omega^{-2}, \mathbf{X}', Y')$ (selon notre modèle) :

$$[X_{e,k} | Z, \omega^{-2}, \mathbf{X}', Y'] \begin{cases} = \mathbb{I}_{\{X_{e,k} = f_{\mathcal{N}}(X'_{e,k})\}} \text{ si } X'_{e,k} > 0 \\ = \mathcal{N}_{<\nu} (X_{e,k}; a_e + b_e Z, c_e^2 \omega^2) \text{ si } X'_{e,k} = 0 \end{cases}$$

où \mathbb{I} désigne la fonction indicatrice et $\mathcal{N}_{<\nu} (x; \mu, \sigma^2)$ désigne la densité en x de la loi normale d'espérance μ et de variance σ^2 tronquée à droite en ν .

Nous connaissons aussi la loi de $(Z, \omega^{-2} | \mathbf{X}, Y, \mathbf{X}', Y')$: il s'agit de la même loi que dans le cas gaussien car $[Z, \omega^{-2} | \mathbf{X}, Y, \mathbf{X}', Y'] = [Z, \omega^{-2} | \mathbf{X}, Y]$.

Nous pouvons donc, plutôt que calculer l'espérance par rapport aux variables latentes conditionnellement à (\mathbf{X}', Y') , réaliser des simulations de $(\mathbf{X}, Y, Z, \omega^{-2} | \mathbf{X}', Y')$ selon un algorithme de Gibbs (prenant pour point de départ les lois marginales de

Z et ω^2) puis considérer la première simulation de (\mathbf{X}, Y) obtenue après la période de chauffe comme s'il s'agissait d'une observation.

L'étape Maximisation de l'algorithme EM n'est donc pas impactée par le passage du modèle de la section 3.1 à celui-ci.

Il s'agit alors d'un EM (partiellement) stochastique (SEM, Broniatowski et al. (1983); Celeux & Diebolt (1985)), mais seulement en partie car on conserve l'étape "Expectation" par rapport à $(Z, \omega^{-2} | \mathbf{X}', Y')$.

Autrement dit, à l'itération $(i + 1)$ de l'algorithme EM :

1. On simule w^{-2} selon $\Gamma(\alpha^{(i)}, \beta^{(i)})$ puis z selon $\mathcal{N}(0, \lambda^{(i)} w^2)$.
2. Jusqu'à convergence de l'algorithme de Gibbs :
 - (a) on simule (\mathbf{x}, y) selon $[\mathbf{X}, Y | Z = z, \omega^{-2} = w^{-2}, \mathbf{X}', Y']$ (lorsque l'observation de \mathbf{X}' ou de Y' correspondante est nulle),
 - (b) on simule z et w^{-2} selon $[Z, \omega^{-2} | \mathbf{X} = \mathbf{x}, Y = y, \mathbf{X}', Y'] = [Z, \omega^{-2} | \mathbf{X} = \mathbf{x}, Y = y]$.
3. On simule une dernière fois selon $[\mathbf{X}, Y | Z = z, \omega^{-2} = w^{-2}, \mathbf{X}', Y']$, on obtient $(\mathbf{x}, y)^{(i)}$, et on fait comme s'il s'agissait d'observations de (\mathbf{X}, Y) (on remplace $\mathbb{E}\{\mathbf{X}, Y | \mathbf{X}', Y'\}$ par $(\mathbf{x}, y)^{(i)}$ dans les équations de l'étape maximisation).

3.2.3 Prévisions

Afin de simuler y' selon la loi de $(Y' | \mathbf{X}')$, nous procédons comme auparavant à un échantillonnage de Gibbs afin de simuler selon la loi de $(Z, \omega^{-2}, \mathbf{X} | \mathbf{X}')$. Nous simulons ensuite un y pour chaque couple (z, w^{-2}) obtenu après la période de chauffe, puis prenons $y' = f_{\mathcal{N}}^{-1}(y) \mathbb{I}_{y > \nu}$.

3.2.4 Expérience sur des données simulées

Nous vérifions dans cette section que, dans un cadre réaliste (c'est-à-dire pour des paramètres proches de ceux que l'on apprendrait par une inférence conduite sur des jeux de données réels), nous sommes à même d'estimer correctement les paramètres de notre modèle avec l'inférence par algorithme SEM précédemment décrite. Le paramètre γ de la transformation normalisante étant appris séparément, on se place, pour cette expérience numérique, directement dans l'espace normalisé. Nous simulons pour cela des données selon le modèle décrit dans la section 3.2.1 avec les paramètres suivants :

$$\begin{aligned} K &= (1, 10, 35, 1) \\ a &= (0, 1, 0.3, -0.1) \\ b &= (1, 1.1, 0.8, 0.9) \\ c &= (1, 0.8, 1.3, 1.1) \\ \alpha &= 2.5, \quad \beta = 3, \quad \lambda = 0.5, \end{aligned}$$

où le premier élément de chaque vecteur est celui indicé par 0, relatif aux observations. Nous simulons selon ce modèle 100 jeux de données de 100 éléments chacun (nous effectuons donc 10000 simulations selon le modèle). L'algorithme SEM précédemment décrit est ensuite utilisé sur chacun des 100 jeux de données pour l'inférence des paramètres du modèle. La valeur initiale des paramètres est choisie par une méthode des moments. Notons que les paramètres b_0 et c_0 sont contraints à valoir 1 et ne sont donc pas estimés. L'algorithme SEM effectue 1000 itérations et, au sein de chaque itération, l'algorithme de Gibbs en effectue 4, ce qui apparaît comme suffisant dans le cas ici considéré au vue des résultats d'estimation. Les estimations des paramètres ainsi obtenues sur les 100 jeux de données sont illustrées sous la forme de boîtes à moustaches sur la figure 3.10. L'estimation résultante ne fait apparaître aucun biais quel que soit le paramètre considéré. Ces paramètres sont estimés de façon moins précise que sur la figure 3.2, mais il faut noter que le nombre de répétitions ici utilisé pour l'apprentissage (100) est 5 fois inférieur à celui utilisé dans le cas gaussien (500).

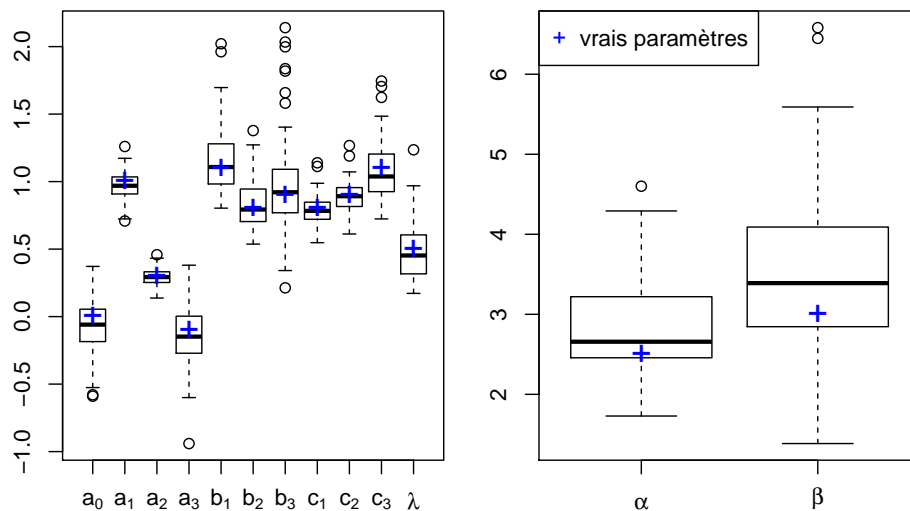


FIGURE 3.10 – Paramètres obtenus pour 100 simulations de 100 répétitions selon le modèle écrit pour le cas des précipitations. Les paramètres utilisés pour la simulation sont représentés par les croix bleues.

3.2.5 Application aux données d'EDF-DTG et d'Hydro-Québec

Nous testons ensuite notre méthode de post-traitement des précipitations sur les données opérationnelles. Une première étape consiste à vérifier l'hypothèse d'échangeabilité des membres des ensembles. Comme relevé page 49, les membres des prévisions CEPMMT-EPS et NCEP-GEFS peuvent être supposés échangeables tandis que, dans le cas de l'ensemble CMC-EPS, dans le cas des précipitations, l'hypo-

thèse d'échangeabilité paraît plus acceptable parmi les membres impairs et parmi les membres pairs. Nous traitons donc désormais ces deux groupes de membres comme deux ensembles distincts et les notons respectivement CMC-EPS-1 et CMC-EPS-2.

Comme dans la section 3.1.5, nous commençons par traiter l'exemple du bassin versant québécois de Manic 2.

INFÉRENCE Nous apprenons les paramètres du modèle de prévision à partir de jeu de données de l'année 2013. Les paramètres obtenus sont représentés sur la figure 3.11 en fonction de l'échéance de prévision. Le paramètre de transformation, γ , est préalablement estimé, la valeur obtenue est 0.43.

On remarque que :

- Les paramètres obtenus pour les sous-ensembles CMC-EPS-1 et CMC-EPS-2 sont très différents (en particulier les paramètres de biais additif et de biais multiplicatif, illustrés sur les graphiques (1) et (2)), cette subdivision de l'ensemble CMC-EPS est donc pertinente. Toutefois, excepté à la 9^e échéance, les biais additifs et multiplicatifs diffèrent de façon opposée, ils se compensent donc et la différence entre ces deux ensembles est moins importante qu'il n'y paraît au premier abord.
- Contrairement aux ensembles de températures, les ensembles de précipitations montrent, aux échéances de prévisions médianes (4 à 6 jours) en particulier, un comportement sur-dispersif ($c > 1$). Il est toutefois nécessaire de faire preuve de prudence dans l'interprétation de ce comportement : les paramètres ici décrits sont relatifs à des variables latentes qu'il faudra (transformer et) tronquer pour obtenir les variables relatives aux ensembles que nous observons.
- Conformément à notre intuition, quand l'échéance de prévision augmente, le mode de ω^2 (donné par $\frac{\beta}{\alpha+1}$ et illustré sur le graphique (5)) augmente, l'incertitude relative à la météorologie future augmente donc. Cette augmentation est toutefois partiellement compensée par la diminution de la valeur de λ (illustrée sur le graphique (6)).
- Comme illustré sur le graphique (4), l'ensemble européen contribue davantage à la prévision au cours de l'année 2013 que les autres ensembles, et ce même en ajustant le calcul de sa contribution par le nombre de membres qu'il comprend. Quand l'échéance de prévision augmente, les contributions des membres s'équilibrent exceptée celle de l'ensemble constitué par les membres impairs de l'ensemble canadien, qui reste très inférieure aux autres.
- Quel que soit l'ensemble considéré, le biais additif augmente au cours des échéances de prévision, mais cette augmentation est compensé par une diminution simultanée du biais multiplicatif. En tenant compte de ces deux biais, les informations provenant des ensembles européen et américain telles que résumées dans notre jeu de données sont biaisées positivement à toutes les échéances.

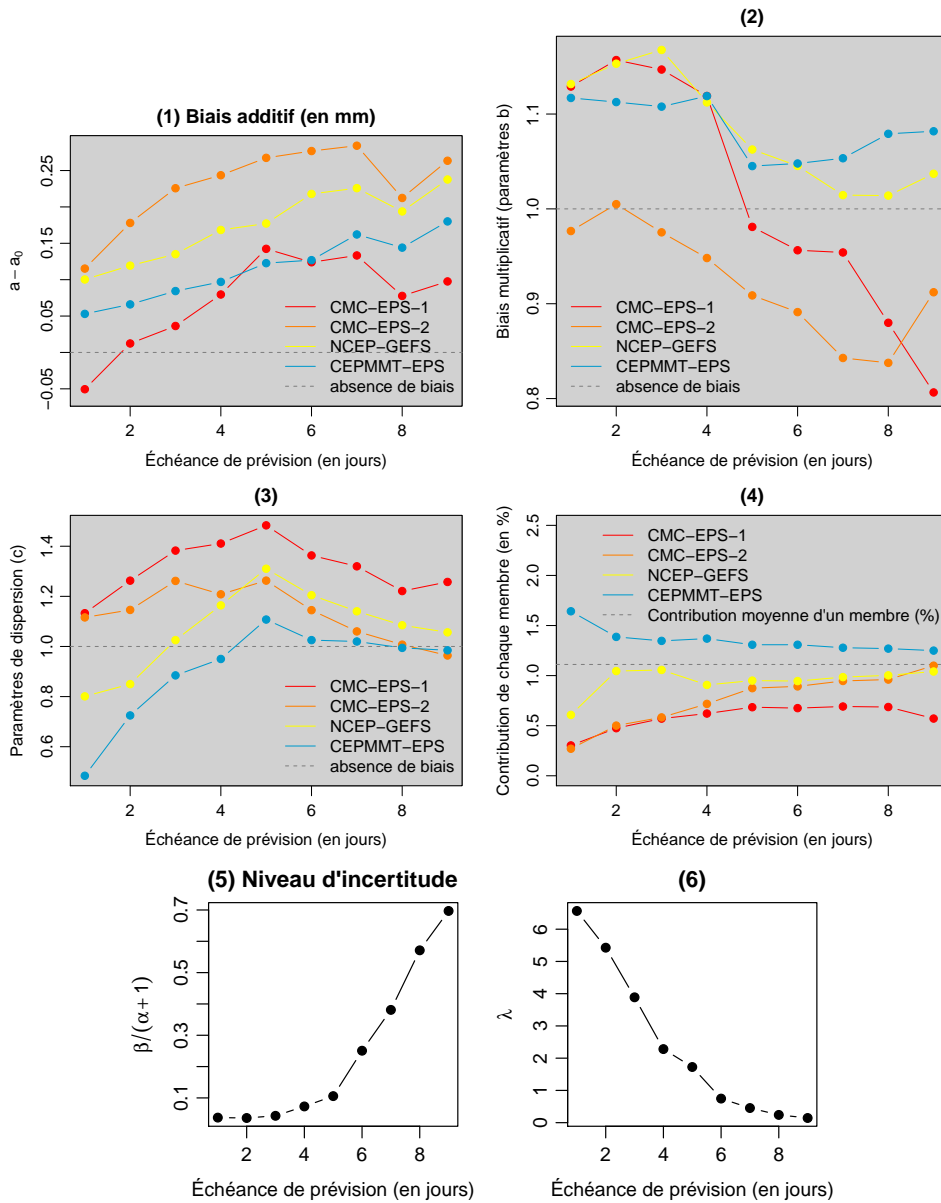


FIGURE 3.11 – Paramètres obtenus en fonction de l'échéance de prévision pour le bassin versant de Manic 2 avec $E = 4$ sources de prévisions pour le modèle de post-traitement des précipitations avec un paramètres de transformation $\gamma = 0.43$ (les membres impairs de l'ensemble CMC-EPS (CMC-EPS-1), les membres pairs de l'ensemble CMC-EPS (CMC-EPS-2) et les ensembles NCEP-GEFS et CEPMMT-EPS), avec un apprentissage sur l'année 2013.

PRÉVISIONS Poursuivant notre étude prenant en exemple le bassin versant québécois de "Manic 2", nous testons notre méthode de post-traitement sur l'année 2014.

Le post-traitement réalisé conduit, comme attendu, à un gain en fiabilité par rapport aux prévisions d'ensemble d'origine. Ce gain est particulièrement marqué aux premières échéances de prévision. La figure 3.12 l'illustre par les histogrammes du rang des observations parmi les membres des ensembles (graphiques (1) à (4)), ou bien parmi des simulations selon les distributions prédictives issues du post-traitement (graphique (6)), sur l'exemple de l'échéance de prévision 3 jours (cette façon d'évaluer la fiabilité des prévisions est décrite dans la section 1.3.1). L'histogramme de rangs correspondant aux prévisions post-traitées est en effet plus plat que ceux associés aux prévisions d'ensemble, qui montrent, dans le cas des ensembles NCEP-GEFS et CEPMMT-EPS, une sous-dispersion et un biais positif, et, dans le cas de l'ensemble CMC-EPS et du grand ensemble obtenu en agrégeant les trois ensembles précédents, un biais positif. Cette sous-dispersion et ce biais sont davantage marqués aux échéances de prévision 1 et 2 jours et s'atténuent à mesure que l'échéance de prévision augmente. Ils montrent de plus, aux échéances médianes, une légère sur-dispersion. Ces éléments corroborent l'interprétation des paramètres de dispersion (c) et de biais (a et b) proposée dans la section 3.2.5.

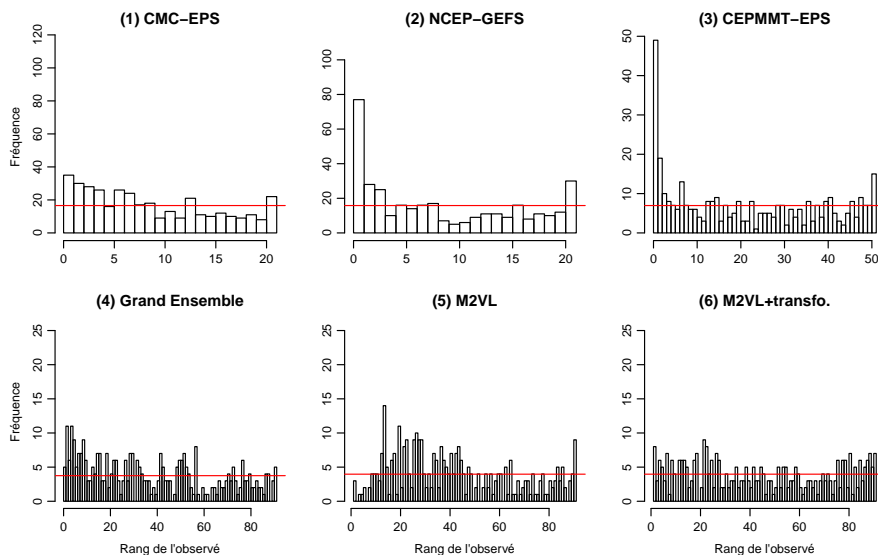


FIGURE 3.12 – Histogrammes du rang des observations : (1-3) parmi les membres des ensembles (1) CMC-EPS (2) NCEP-GEFS et (3) CEPMMT-EPS, (4) parmi les membres du grand ensemble constitué des membres des 3 ensembles précédents, (5-6) parmi des simulations selon le modèle de post-traitement des précipitations appris sur l'année 2013 (5) sans transformation (6) avec transformation normalisante. On considère ici l'échéance de prévision 3 jours, le bassin versant de Manic 2 et l'année 2014. La ligne rouge indique l'objectif visé pour une fiabilité des prévisions.

Les prévisions probabilistes issues du post-traitement proposé ne peuvent toutefois être considérées comme parfaitement fiables. L'histogramme de rang correspondant

(graphique (6)) montre en effet une légère sur-dispersion associée à un léger biais négatif. Une interprétation possible est que la distribution normale (censurée) élevée à une puissance ne permet pas d'obtenir des queues de distribution suffisamment lourdes pour mimer des variables de précipitations : malgré une sur-dispersion, de trop nombreuses observations occupent les rangs les plus élevés. La transformation normalisante opérée a toutefois permis un gain important en fiabilité : sans transformation préalable, ce patron serait beaucoup plus marqué, comme illustré sur le graphique (5) de la figure 3.12.

La figure 3.13 montre un exemple de situation de prévision des cumuls de précipitations journaliers du bassin versant de Manic 2. Dans cet exemple, les prévisions sont émises le 30 avril 2014 pour les 10 jours qui suivent. On remarque des membres négatifs dans l'ensemble CEPMMT-EPS au 10^e jours de prévision, il s'agit d'une des anomalies qui appuient notre choix de ne considérer que les 9 premières échéances de prévision.

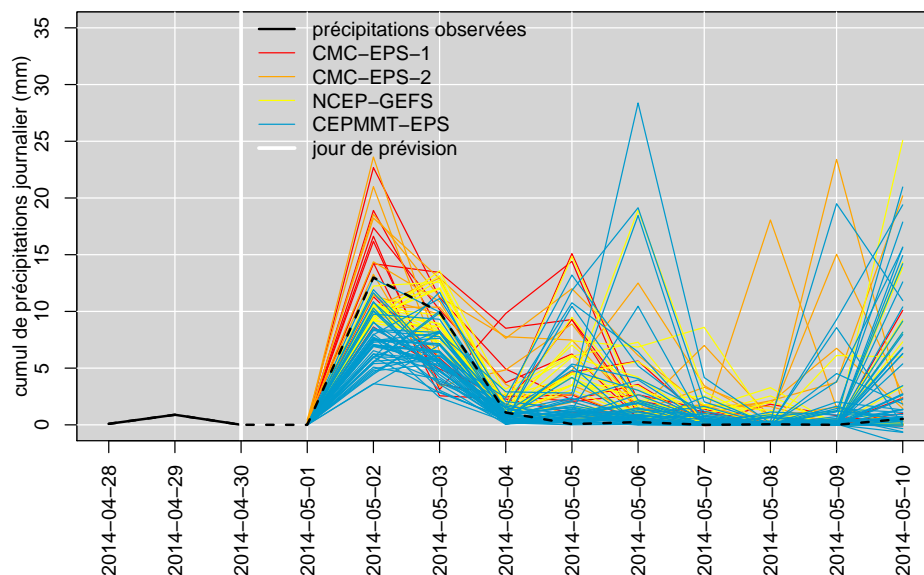


FIGURE 3.13 – Exemple de situation de prévision des précipitations journalières du bassin versant de Manic 2. Les informations à disposition pour émettre la prévision sont les ensembles NCEP-GEFS, CMC-EPS (membres impairs : 1 et pairs : 2) et CEPMMT-EPS. Les précipitations à prévoir sont en noir tireté.

La figure 3.14 illustre les prévisions émises à l'issue du post-traitement (en haut : avant la censure à gauche en zéro, en bas : après cette censure). Les (90) scénarios de prévisions post-traitées illustrés sont obtenus à partir des prévisions post-traitées échéance de prévision par échéance de prévision par une copule empirique (ici l'ECC). Le code couleur utilisé est le même que celui de la figure 3.4.

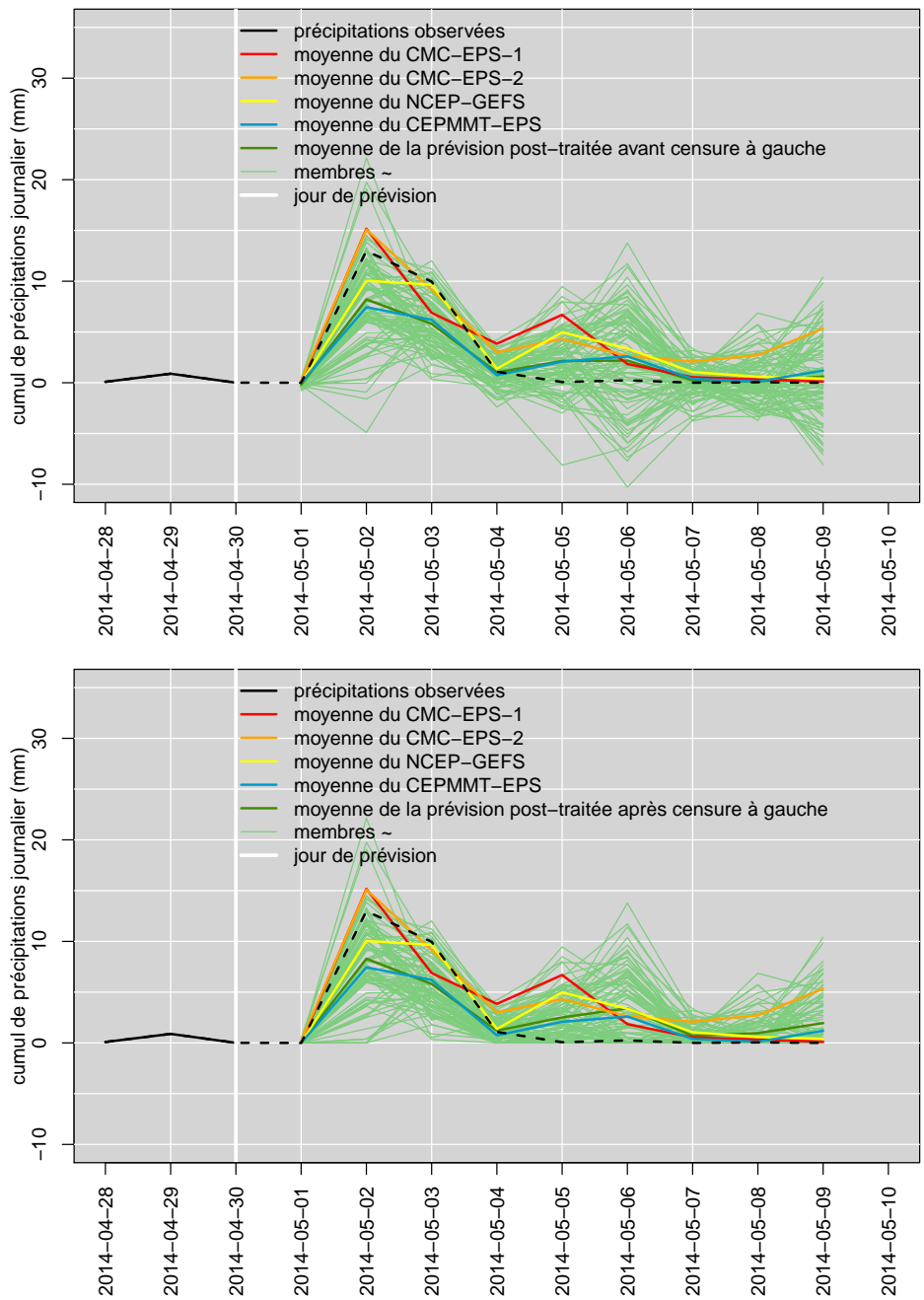


FIGURE 3.14 – Exemple de prévision émise pour les précipitations journalières du bassin versant de Manic 2 avec notre méthode de post-traitement appliquée aux ensembles NCEP-GEFS, CMC-EPS (impairs : 1 et pairs : 2) et CEPMMT-EPS. En haut : avant censure. En bas : après censure à gauche en zéro. Les précipitations à prévoir sont en noir tireté.

La figure 3.15 montre la performance des prévisions ainsi obtenues évaluée par le CRPS (cette façon d'évaluer la performance des prévisions est expliquée dans la section 1.3.3 du chapitre 1). Sur cette figure, **M2VL + transfo.** désigne la méthode de post-traitement décrite dans la présente section, **M2VL** cette même méthode mais sans application préalable d'une transformation normalisante et **fuites** est relatif à une méthode de post-traitement des précipitations alternative s'appuyant sur un raisonnement similaire à celui de EMOS décrite plus loin, dans la section 3.3.

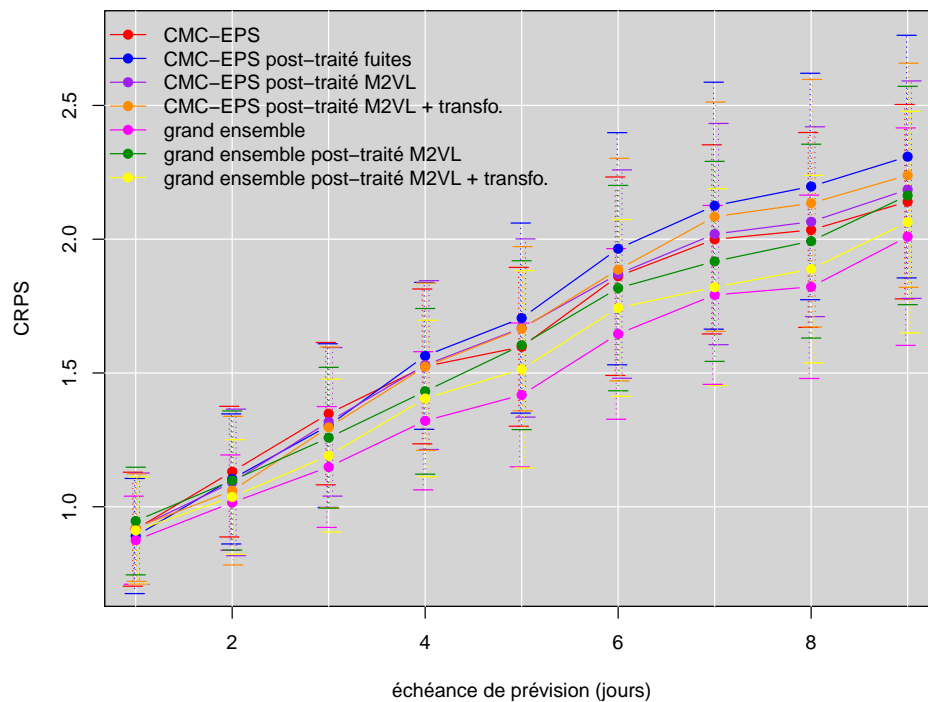


FIGURE 3.15 – CRPS (en mm) relatifs aux prévisions de précipitations pour le bassin versant de Manic 2 en fonction de l'échéance de prévision et intervalles bootstrap associés, pour un apprentissage sur l'année 2013 suivi d'un test sur l'année 2014.

Les résultats d'inférences montrés sur la figure 3.11 sont utilisés pour obtenir la prévision **grand ensemble post-traité M2VL + transfo.** Nous avons de plus appliqué notre méthode de post-traitement à l'ensemble CMC-EPS seul et ainsi obtenu la prévision **CMC-EPS post-traité M2VL + transfo.** La prévision CMC-EPS est obtenue en considérant la fonction de répartition empirique de l'ensemble comme une prévision probabiliste. Excepté dans les méthodes **fuites** et **CMC-EPS**, nous traitons les membres impairs et pairs du CMC-EPS comme deux ensembles distincts.

L'ajout des prévisions CEPMMT-EPS conduit comme attendu à une amélioration des prévisions et, dans le cas multi-ensemble, la transformation améliore comme attendu notablement les prévisions post-traitées.

Les performances des prévisions post-traitées (évaluées par le CRPS) sont toute-

fois décevantes sur ce bassin versant : les scores associés aux prévisions sans post-traitement sont meilleurs que ceux associés aux prévisions avec post-traitement.

Des non-homogénéités dans le jeu de données pourraient expliquer ces résultats : nous savons qu'un changement majeur du modèle atmosphérique du CMC a eu lieu en février 2013, soit au début de notre période d'apprentissage, avec un impact fort sur les précipitations.

Sur les autres bassins versants québécois, les résultats obtenus sont bien meilleurs : les prévisions sont en général légèrement améliorées par l'application de notre post-traitement, en particulier sur les premières échéances (jusqu'à l'horizon 4 jours) (non montré). Ceci apparaît davantage lorsque l'on regarde les résultats obtenus pour les prévisions de cumuls de précipitations, comme illustré sur la figure 3.16. Les prévisions de cumuls de précipitations du CMC-EPS sont améliorées par le post-traitement quel que soit le bassin versant considéré. Les prévisions de cumuls du grand ensemble ne sont en revanche améliorées que pour 3 des 5 bassins versants québécois, ce qui peut s'expliquer par une non représentativité des prévisions d'ensemble CEPMMT-EPS ou bien NCEP-GEFS émises l'année 2013 par rapport à l'année 2014 sur ces bassins versants. Enfin, les prévisions du grand ensemble prises telles quelles sont toujours meilleures que celles du CMC-EPS seul.

La figure 3.16 montre aussi les résultats obtenus sur les 3 bassins versants français étudiés (bas de la figure). Ils sont très contrastés : la performance des prévisions est améliorée par le post-traitement sur le bassin versant du Drac à Sautet, mais le post-traitement les détériore légèrement sur le bassin versant de l'Ain à Vouglans et très fortement sur le bassin versant du Buech aux Chambons. Dans le cas du bassin versant du Buech aux Chambons, la transformation ne permet pas d'obtenir une distribution s'approchant de celle d'une gaussienne tronquée, produisant ainsi des extrêmes irréalistes après transformation inverse (γ est estimé à 0.07 sur ce bassin versant, contre 0.6 sur les deux autres bassins versants français). Dans ce cas, il est nécessaire de trouver une transformation plus adéquate, éventuellement en levant la contrainte portant sur ν ($\nu = 0$).

Sur les bassins versants québécois, le paramètre de transformation estimé prend des valeurs plus homogènes; elles sont comprises entre 0.43 (pour le bassin versant de Manic 2 et pour le bassin versant de Toulnostouc) et 0.5 (pour le bassin versant du Petit Lac Manicouagan).

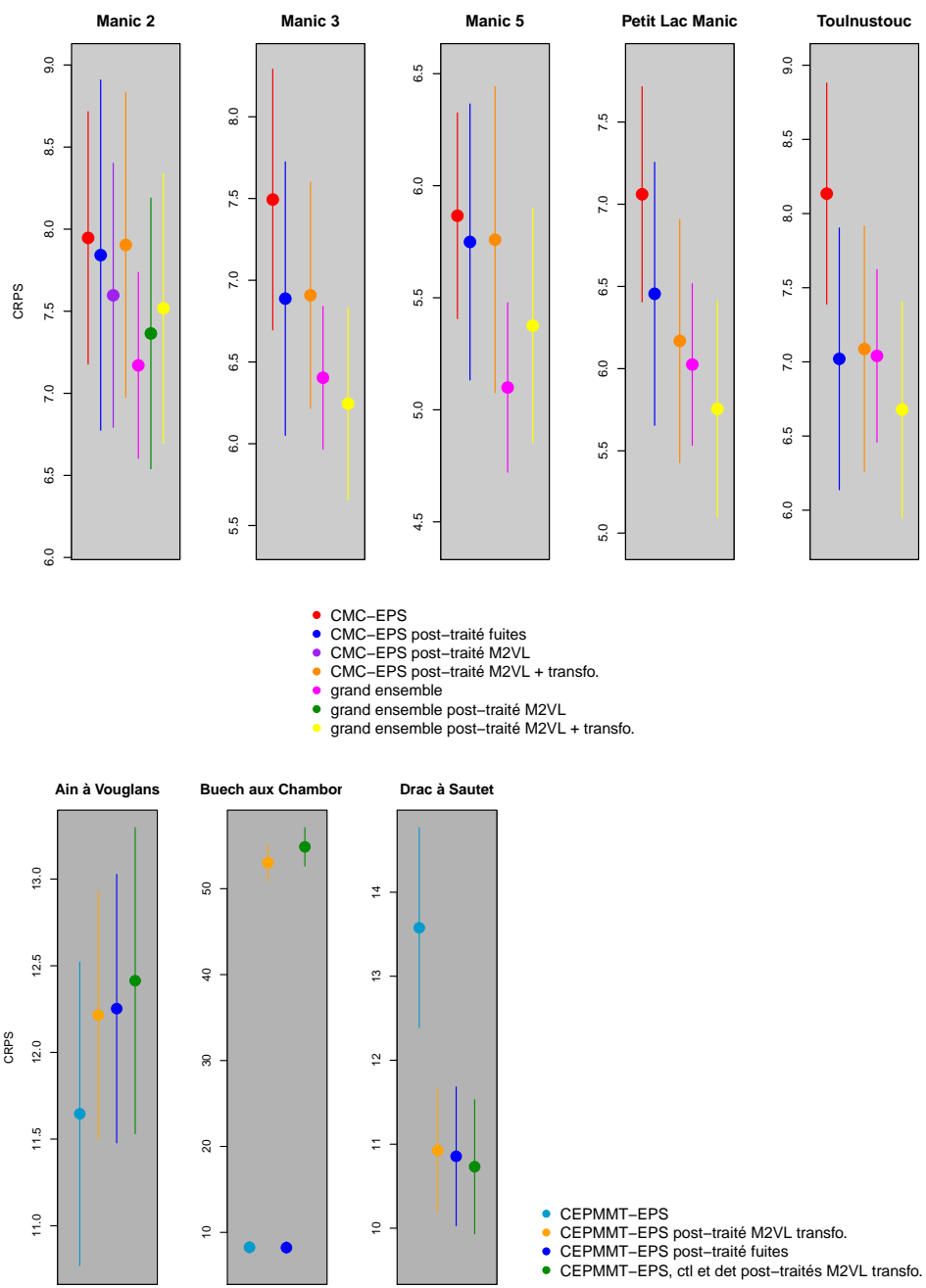


FIGURE 3.16 – CRPS (en mm) relatifs aux prévisions de précipitations cumulées dans le temps et intervalles bootstrap associés, pour un apprentissage des paramètres sur l’année 2013 suivi d’un test sur l’année 2014, dans le cas des bassins versants québécois, et un apprentissage sur les années 2011 à 2013 suivi d’un test sur les années 2014 et 2015 dans le cas des bassins versants français.

3.3 Post-traitement des prévisions de précipitations avec la loi des fuites

Ce travail est un travail joint avec et sur une idée de Luc Perreault (IREQ).

Toujours dans l'idée de ne pas séparer modélisation de l'occurrence de précipitations et modélisation de la quantité de précipitations conditionnellement à leur occurrence, nous avons exploré en parallèle une méthode de post-traitement des prévisions d'ensemble de précipitations fondée sur l'utilisation de la loi des fuites dans un modèle s'inscrivant dans la logique de EMOS.

La loi des fuites (Babusiaux, 1969; Bernier & Fandoux, 1970; Ballay et al., 1996) est la loi suivie par le cumul de quantités produites selon une loi exponentielle au cours d'événements dont l'occurrence est dictée par une loi de Poisson. Elle est similaire à la loi de Tweedie, proposée plus tard pour la modélisation des précipitations par Dunn (2004). Soit Y une variable aléatoire. Y suit une loi des fuites s'il existe β et λ tels que :

$$\begin{cases} (Y|\mathbf{Z}, \mathcal{I} = I) & = \sum_{i=1:I} Z_i \\ Z_i & \stackrel{iid}{\sim} Exp(\beta) \\ \mathcal{I} & \sim \mathcal{P}(\lambda) \end{cases},$$

où $Exp(\beta)$ désigne la loi Exponentielle de paramètre β .

La densité de la loi des fuites résultante, de paramètre d'échelle γ et de paramètre de forme λ , est donnée par :

$$f(y; \gamma, \lambda) = \begin{cases} e^{-\lambda} & \text{si } y = 0 \\ \frac{\lambda}{\gamma} e^{-\lambda - \frac{y}{\gamma}} I_1 \left(2\sqrt{\frac{\lambda y}{\gamma}} \right) \sqrt{\frac{\gamma}{\lambda y}} & \text{si } y > 0 \end{cases},$$

où I_1 est la fonction de Bessel modifiée d'ordre 1.

L'espérance et la variance associées à cette loi des fuites sont données par :

$$\begin{cases} \mathbb{E}(Y) & = \lambda\gamma \\ \mathbb{V}(Y) & = 2\lambda\gamma^2 \end{cases}.$$

Une interprétation possible de la modélisation des précipitations journalières selon une loi des fuites est la suivante : le cumul de précipitations quotidien est la somme des précipitations au cours de I événements pluvieux ayant lieu dans la journée dont le nombre d'occurrences est issu d'une loi de Poisson.

3.3.1 Modèle

Plaçons-nous dans le cas où nous disposerions d'une seule source de prévisions et ce sous la forme de prévisions d'ensemble. Soit Y la variable (de type précipitations)

à prévoir et $\mathbf{X} = X_1, \dots, X_K$ la prévision d'ensemble correspondante, avec K le nombre de membres de l'ensemble.

Suivant la logique de EMOS, nous supposons que la moyenne, \bar{X} , et l'écart-type, S_X , de l'ensemble portent toute l'information relative à Y contenue dans les membres de l'ensemble. Nous supposons de plus que l'espérance de $(Y|\mathbf{X})$ est une fonction linéaire de \bar{X} et que l'écart-type de $(Y|\mathbf{X})$ est une fonction linéaire de S_X .

Le modèle que nous proposons est donc le suivant :

$$\begin{cases} (Y|\mathbf{X} = x) & \sim \mathcal{Fuites}(\lambda(x), \gamma(x)) \\ \lambda(x) & = \frac{2(a+b\bar{x})^2}{(c+dS_x)^2} \\ \gamma(x) & = \frac{(c+dS_x)^2}{2(a+b\bar{x})} \end{cases}$$

où $\mathcal{Fuites}(\lambda, \gamma)$ désigne la loi des fuites de paramètre d'échelle γ et de paramètre de forme λ .

3.3.2 Inférence

Nous proposons de choisir pour les paramètres les estimateurs du maximum de vraisemblance en utilisant la méthode de Nelder-Mead (implémentée dans la fonction `optim` de R).

3.3.3 Application aux données d'EDF-DTG et d'Hydro-Québec

Les prévisions obtenues en post-traitant l'ensemble CMC-EPS (traité comme un unique ensemble) avec cette méthode présentent de façon surprenante le même problème de fiabilité que dans le cas de la modélisation par une loi normale censurée : les histogrammes de rang obtenus ressemblent à celui du graphique (5) la figure 3.12. Appliquer ce post-traitement aux cumuls successifs de précipitations (au cours du premier jour, puis des deux premiers jours,...) pourrait permettre de résoudre en partie ce problème de fiabilité : l'interprétation proposée pour le modèle serait alors plus conforme à la réalité. Il est en effet plus probable que plusieurs événements pluvieux aient lieu au cours d'une série de jours que durant une plage horaire limitée à 24h.

L'évaluation des prévisions issues de ce post-traitement par le CRPS (sur l'année 2014 après un apprentissage sur l'année 2013 pour les bassins versants québécois, sur les années 2014 et 2015 après un apprentissage sur les années comprises entre 2011 et 2013 pour les bassins versants français) montre globalement une bonne performance des prévisions comme l'illustre la figure 3.16 : les prévisions de cumuls de précipitations sont améliorées par le post-traitement sur la majorité des bassins versants considérés. Les prévisions de précipitations le sont en particulier pour les premières échéances de prévision (non montré).

Comme le montre la figure 3.15, ce post-traitement conduit toutefois, dans le cas du bassin versant de Manic 2, à une détérioration du CRPS par rapport à la

prévision sans post-traitement pour les dernières échéances de prévision. Ce résultat peut, là encore, s'expliquer par un manque de stationnarité du jeu de données.

3.4 Discussion et perspectives

En nous appuyant sur la propriété d'échangeabilité, nous avons décrit, dans les sections 3.1 et 3.2, un cadre pour le post-traitement des variables météorologiques permettant d'intégrer plusieurs sources de prévisions (qu'elles soient de type déterministe ou ensembliste et pouvant être issues d'une expertise) et d'écrire des modèles parcimonieux. Nous avons proposé, dans ce cadre, des modèles statistiques pour le post-traitement des prévisions de températures et pour le post-traitement des prévisions de précipitations, ainsi que des méthodes d'inférence par algorithme EM permettant d'en apprendre les paramètres maximisant la vraisemblance. Nous avons vérifié l'algorithme d'inférence sur des données simulées avant d'appliquer les méthodes de post-traitement résultantes sur des données réelles fournies par EDF-DTG et par Hydro-Québec.

Les résultats obtenus sont satisfaisants en ce qui concerne la partie du post-traitement relative aux températures : les CRPS sont presque systématiquement améliorés par les post-traitements opérés.

La partie de la méthode adaptée aux prévisions de précipitations n'est, elle, pas encore tout-à-fait satisfaisante : moyennant une transformation normalisante des précipitations, elle donne lieu à des prévisions relativement fiables mais ses performances sont décevantes sur la moitié des bassins versants considérés, probablement en raison d'un problème de non-homogénéité du jeu de données, et, sur un des bassins versants, la transformation ne permet pas la normalisation souhaitée, ce qui a pour conséquence la simulation de précipitations extrêmes irréalistes.

Il est possible, avec la méthode proposée, d'intégrer des prévisions expertisées en tant que source de prévision additionnelle de la même façon que nous intégrons des prévisions de nature déterministe, à condition de disposer de données expertisées homogènes sur la période d'apprentissage. C'est par exemple le cas chez Hydro-Québec. Cela donnerait aux prévisionnistes qui le souhaitent la possibilité de (continuer à) travailler avec une vision déterministe du phénomène météorologique.

Par ailleurs, nous nous sommes ici simplement appuyés sur des copules empiriques pour rétablir une cohérence dans l'espace, dans le temps et entre variables. Nous ne sommes pas à même, avec cette méthode, de garantir que les températures maximales journalières simulées selon notre modèle de prévision seront toujours supérieures aux températures minimales journalières correspondantes. Cette façon de faire induit de plus une limite dans le nombre de simulations productibles : celle du nombre de scénarios similaires dont on copie la structure multidimensionnelle (en l'occurrence le nombre de scénarios de prévisions d'ensemble à disposition).

Une première solution, naturelle ici puisque nous avons construit nos modèles sur

un socle gaussien, serait de considérer la loi jointe des variables latentes Z intervenant dans chacun de ces modèles, à la fois entre variables, pour plusieurs pas de temps et à plusieurs localisations. Cela permettrait de plus d'utiliser l'information provenant des dernières observations météorologiques.

Une autre solution consisterait à utiliser, à la place, des modèles de copules paramétriques tels que les modèles de copules gaussiennes. Afin de profiter des très nombreux modèles de copules conçus pour des dimensions faibles, il est aussi possible d'utiliser des modèles de copules bidimensionnelles imbriqués les uns dans les autres pour obtenir des structures de plus grandes dimensions comme le proposent Aas et al. (2009).

Nous pouvons par ailleurs envisager de dériver des modèles proposés les lois marginales des variables à prévoir afin d'intégrer des informations issues de la climatologie, par exemple en réalisant une inférence distincte des paramètres des lois marginales (ce qui résulterait en des contraintes sur les autres paramètres lors de l'inférence par algorithme EM). Toutefois, cette façon de faire ne découle pas naturellement de la modélisation. Nous explorons pour cette raison actuellement une autre voie de modélisation possible pour les variables de type températures, similaire à celle de ce manuscrit mais inversant les rôles de Y et de Z . Une telle façon de faire devrait permettre non seulement d'intégrer les données climatologiques avec plus de facilité, mais aussi d'aisément modéliser la structure temporelle des phénomènes.

Nous avons finalement exploré une voie alternative pour le post-traitement des prévisions de précipitations. Celle-ci repose sur l'idée de modéliser les précipitations selon une loi des fuites, *a priori* adaptée à ce type de données mais présentant des problèmes de fiabilité.

Les résultats du post-traitement qui en résulte sont bons sur 6 des 8 bassins versants considérés.

Travailler sur les cumuls successifs de précipitations au cours des jours de prévision pourrait améliorer la fiabilité des prévisions mais induirait de nouvelles difficultés : en utilisant des copules empiriques comme nous le faisons, il serait par exemple possible de simuler un cumul de précipitations sur 3 jours inférieur à celui des 2 jours précédents.

Une limitation, commune à l'ensemble des méthodes ici décrites, est qu'elles ne permettent pas d'émettre des distributions prédictives multimodales. Il n'est pas facile de mesurer à quel point cet inconvénient est pénalisant en pratique : sur nos jeux de données, les cas où les membres d'une même prévision d'ensemble forment une distribution empirique multimodale sont très rares. Nous pourrions imaginer des modèles de mélange pour permettre ce type de comportement, comme nous le faisons dans la modélisation du chapitre qui suit pour l'hydrologie.

4

Un post-traitement des prévisions hydrologiques tirant parti d'informations auxiliaires

4.1	Résumé de l'article soumis pour publication	102
4.1.1	Modèle et inférence	102
4.1.2	Application aux données d'EDF et d'Hydro-Québec et résultats	103
4.2	Article soumis pour publication	105

Nous avons conclu dans le chapitre 2 que, parmi les méthodes de post-traitement hydrologique existantes, celles consistant à proposer un modèle statistique pour l'ensemble du post-traitement paraissaient plus avantageuses et adaptées aux systèmes opérationnels que nous considérons, mais que l'usuelle modélisation par un processus homogène ne permettait pas de bien représenter les erreurs du SPQ.

Dans le présent chapitre, nous étudions une modélisation faisant intervenir les variables d'état du SPQ pour déclencher des régimes de fonctionnement différents et permettre ainsi de représenter une incertitude hydrologique à variance hétérogène.

La méthode opérationnelle d'EDF, succinctement décrite dans l'Annexe B de l'article, utilise déjà les variables d'état du SPQ (notamment la proportion du débit à l'exutoire en provenance d'un écoulement profond) pour distinguer les régimes d'erreur de crue et d'étiage. Elle procède toutefois en plusieurs étapes successives (prise en compte des dernières observations de débits puis de l'incertitude et enfin de la dépendance temporelle).

Nous avons appliqué la méthode que nous proposons aux jeux de données fournis par EDF-DTG et par Hydro-Québec. Ils comprenaient, au moment des travaux, les 15 bassins versants français et 4 des 8 bassins versants québécois décrits dans le

chapitre 1. Parmi les 15 bassins versants français, 4 ont été laissés de côté car ils présentaient des données manquantes ou bien des défauts de stationnarité flagrants.

Ce chapitre correspond à un article soumis pour publication (en cours de révision, situé à la section 4.2). Dans la section 4.1, nous en présentons le résumé en guise d'introduction.

4.1 Résumé de l'article soumis pour publication

4.1.1 Modèle et inférence

MODÈLE Soient Y et X les variables aléatoires obtenues en normalisant respectivement les variables relatives aux débits à l'exutoire d'un bassin versant (à prévoir) et aux sorties correspondantes d'un SPQ prenant en intrant les précipitations et les températures observées sur le bassin versant en question. En reprenant les notations de la section 2.4, $X = f_{\mathcal{N}}(M)$ et $Y = f_{\mathcal{N}}(Q)$. Dans ce travail, nous avons considéré la transformation $f_{\mathcal{N}}(\cdot) = \log(\cdot)$ utilisée chez EDF.

Nous proposons le modèle de mélange ARX suivant :

$$\begin{aligned} (Y_t | Y_{t^-}, X_{t^-}, X_t, S_t = k) &= a_k + \mathbf{b}_k^T X_{t-r, \dots, t} + \mathbf{c}_k^T Y_{t-s, \dots, t-1} + \sigma_k \varepsilon_t \\ \varepsilon_t &\stackrel{iid}{\sim} \mathcal{N}(0, 1) \\ S_t &= \mathbb{I}_{Z_t \leq 0} \\ Z_t &\sim \mathcal{N}(\mathbf{B}^T \mathbf{V}_t, 1) \end{aligned}$$

Où :

- t^- désigne les temps antérieurs au temps t ,
- s est l'ordre maximal du modèle auto-régressif,
- r est l'ordre maximal de la dépendance dans le temps de la variable à prévoir, Y , à son prédicteur, X ,
- S_t est le régime d'erreur au temps t , il peut prendre les valeurs 0 et 1,
- \mathbf{V} est un vecteur de variables d'état du SPQ, à sélectionner,
- Z est une variable latente guidant l'appartenance aux régimes d'erreur 0 et 1 et dont la loi dépend des variables d'état \mathbf{V} ,
- et $(a_k, \mathbf{b}_k, \mathbf{c}_k, \sigma_k)_{k \in \{0,1\}}$ et \mathbf{B} sont des paramètres ou des vecteurs paramètres à estimer.

Dans les exemples ici traités, $r = s = 6$ est un ordre suffisant pour représenter la mémoire du phénomène étudié. Notons que, dans la littérature, on prend très souvent une persistance d'un seul pas de temps ($r = s = 1$), comme exposé dans la section 2.4.

INFÉRENCE Le modèle proposé faisant intervenir des variables latentes, Z et S , l'algorithme EM est un outil efficace pour conduire l'inférence.

SÉLECTION DE VARIABLES EXPLICATIVES Nous dérivons de plus la matrice d'information de Fisher (FIM) associée à la précision des estimations. Nous nous en servons pour la sélection des variables d'état du SPQ et des termes auto-régressifs et relatifs aux sorties du SPQ intervenant dans le modèle. La méthode de sélection de variables que nous proposons utilise la FIM pour identifier successivement les variables les moins informatives et le Bayesian Information Criterion (BIC) pour valider chaque suppression.

4.1.2 Application aux données d'EDF et d'Hydro-Québec et résultats

RÉSULTATS D'INFÉRENCE ET DE SÉLECTION DE VARIABLE Les résultats d'inférence sur les bassins versants français et québécois font apparaître deux régimes d'erreur qui s'interprètent de façon similaire pour la majorité des bassins versants considérés :

- Dans le premier régime, par convention $S_t = 0$, seuls les accroissements de débits interviennent dans le modèle, la dépendance temporelle est courte (les termes antérieurs à X_{t-3} et Y_{t-3} ne sont généralement pas sélectionnés) et l'incertitude sur les débits futurs est limitée (σ_0 prend de faibles valeurs). Le modèle, dans ce régime, prend typiquement la forme suivante :

$$(Y_t - Y_{t-1}) = b_{1,0}(X_t - X_{t-1}) + \mathbf{b}_{(4,3),0}X_{t-3,t-2} + \mathbf{c}_{(3,2),0}Y_{t-3,t-2} + \sigma_0\varepsilon_t,$$

$$\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, 1),$$

où $b_{i+1,0}$ désigne le $(i+1)^e$ paramètre du vecteur \mathbf{b}_0 , relatif à X_{t-i} , et $c_{i,0}$ désigne le i^e paramètre du vecteur \mathbf{c}_0 , relatif à Y_{t-i} .

Ce régime “d’*anticipation des accroissements*” intervient typiquement lors des périodes d'étiage : il est d'autant plus probable que $S_t = 0$ que la proportion du débit à l'exutoire en provenance d'un écoulement profond simulée par le SPQ est forte.

- Dans le second régime, $S_t = 1$, ce sont plutôt les erreurs du SPQ ($Y - X$) qui interviennent dans le modèle, la dépendance temporelle est plus longue (les termes X_{t-4} et Y_{t-4} sont généralement sélectionnés) et l'incertitude sur les débits futurs est plus grande (σ_1 prend de fortes valeurs). Le modèle, dans ce régime “de *correction des erreurs*”, prend typiquement la forme suivante :

$$(Y_t - X_t) = c_{1,1}(Y_{t-1} - X_{t-1}) + \dots + c_{4,1}(Y_{t-4} - X_{t-4}) + \sigma_1\varepsilon_t,$$

$$\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, 1).$$

Ce régime intervient typiquement lors des périodes de crue : $S_t = 1$ est d'autant plus probable que le débit à l'exutoire simulé par le SPQ est fort.

PRÉVISIONS RÉSULTANTES Nous évaluons les performances des prévisions hydrologiques probabilistes (à météorologie future connue) obtenues avec notre modèle de post-traitement hydrologique par le CRPS.

Par rapport aux prévisions climatologiques, notre post-traitement permet de réduire le CRPS de l'ordre de 70% sur les bassins versants français et de l'ordre de 50% sur les bassins versants québécois.

Ce modèle de post-traitement se montre de plus, comme attendu, toujours plus performant que son homologue ne permettant pas plusieurs régimes d'erreur (un modèle ARX).

Sur les bassins versants français, les performances de notre post-traitement sont comparables à celles de la méthode opérationnelle automatique d'EDF : son CRPS est respectivement moins bon, équivalent et meilleur que celui de la méthode opérationnelle sur 4, 1 et 6 des 11 bassins versants français étudiés.

La méthode de post-traitement proposée s'adapte avec facilité à de nouveaux contextes : nous avons pu aisément transférer le post-traitement initialement proposé pour des situations de prévision d'EDF à celles d'Hydro-Québec, pourtant très différentes. Les bassins québécois présentent en effet une superficie bien supérieure à celle des bassins versants français (comme le montrent les tableaux 1.2 et 1.3) et des hydrogrammes beaucoup plus marqués par les saisons que ceux des bassins versants français (comme l'illustre la figure 1.4).

PERSPECTIVES On peut imaginer utiliser cette méthode dans tout système de prévision probabiliste et séquentielle reposant sur l'utilisation d'un simulateur (ou modèle déterministe). Si aucune variable d'état n'est disponible pour informer sur le régime, les sorties du simulateur elles-mêmes peuvent être utilisées comme telles.

On pourrait également chercher à étendre le modèle à plus de deux régimes et construire une méthode de sélection du nombre de composantes du modèle de mélange ainsi construit.

On peut aussi envisager d'exploiter des sorties de SPQ concurrents ou d'associer notre modèle de post-traitement à un modèle pluie-débit probabiliste et tenir compte de l'incertitude portant sur ses paramètres, à l'exemple de Kuczera (1983); Beven & Binley (1992); Kuczera & Parent (1998); Kavetski et al. (2002); Kuczera et al. (2006); Thyer et al. (2009); Renard et al. (2011); Schoups & Vrugt (2010); Evin et al. (2014) (entre autres). On rencontrerait alors toutefois les difficultés d'inférence déjà exposées par Evin et al. (2014), dues à des interactions fortes entre les paramètres du SPQ et ceux du modèle auto-régressifs.

4.2 Article soumis pour publication

Water flow probabilistic predictions based on a
rainfall-runoff simulator: a two-regime model with
variable selection

Marie Courbariaux
UMR MIA-Paris, AgroParisTech, INRA,
Université Paris-Saclay,
75005, Paris, France

Pierre Barbillon*
UMR MIA-Paris, AgroParisTech, INRA,
Université Paris-Saclay,
75005, Paris, France

Éric Parent
UMR MIA-Paris, AgroParisTech, INRA,
Université Paris-Saclay,
75005, Paris, France

February 6, 2017

*corresponding author: pierre.barbillon@agroparistech.fr

Abstract

Probabilistic forecasting aims at producing a predictive distribution of the quantity of interest instead of a single best guess point-wise estimate. With regard to water flow forecasts, the two main sources of uncertainty stem from unknown future rainfall and temperature (input error *i.e.* meteorological uncertainty) and from the inadequacy of the deterministic simulator mimicking the rainfall-runoff (RR) transformation (hydrological uncertainty or RR error). These two sources of uncertainty can be dealt with separately and only the latter will be considered here. Only hydrological uncertainty is at stake when recorded meteorological data (instead of meteorological forecasts) are used as inputs to feed the RR simulator (RRS) for probabilistic predictions. The predictive performance of the RRS may strongly depend on the hydrological regimes: rapid flood variations induce large errors of anticipation but a series of dry events will translate into a much more smoother sequence of river levels due to the easily predictable behavior of the soil reservoir emptying. Consequently, a model with several regimes adapted to different error structures appears as a solution to cope with the issue of unstationary predictive variance. The river regime is modeled as a latent variable, the distribution of which is based on additional outputs of the RRS to be selected. Inference is performed by the EM algorithm with both steps leading to explicit analytic expressions. Asymptotic confidence regions for the estimates are provided within the same EM framework. Model selection is also performed, including the length of the model memory as well as the choice of explanatory variables for the latent regimes. The model is applied to a series of water flow forecasts routinely issued by two hydroelectricity producers in France and in Québec and compared with their present operational forecasting methods

KEYWORDS: EM algorithm; Probit model; Model uncertainty; Probabilistic forecasts; Hydrology; Rainfall-runoff model.

1 Introduction

In his pioneering work framing Bayesian Forecasting Systems, Krzysztofowicz (2002) pointed out that the two main sources of uncertainty at stake when forecasting water flows are the uncertainty on the meteorological forecasts (meteorological uncertainty) and the uncertainty on the capability of representation of the output of the Rainfall-Runoff Simulator (RRS) given the true meteorological conditions (hydrological uncertainty). Krzysztofowicz (2002) also recalled out that those two uncertainty sources can be dealt with separately. For the treatment of the former source of uncertainty, see for example Gneiting et al. (2005) and Raftery et al. (2005). In this paper, we shall only deal with the latter source of uncertainty, *i.e.* we assume that future rainfall and temperature are known when undertaking probabilistic prediction of the time series of river flows. RRSs deterministically mimic the water

progression through a given watershed from its surface to its outflow (the quantity to be predicted here). RRSs take as input the weather forcing conditions and output water flow predictions, denoted by \mathbf{M} . No information about the accuracy of the prediction is provided by RRSs.

Statistical post-processing methods (post-processor) embed the RRS and actual water flow series, denoted by \mathbf{Q} , into a statistical model. The goal of this post-processor is to eventually obtain, for any forecasting time t , probabilistic forecasts of water flow from time t to time $t + H$ (H denotes the forecasting horizon) based on the outputs of the RRS until time $t + H$ (obtained by processing the true meteorological conditions until time $t + H$) and on the water flow recorded up to time t . The target at time t is then the conditional distribution of $(\mathbf{Q}_{t+1}^{t+H} | \mathbf{M}_{t+1}^{t+H}, \mathbf{Q}^t)$, where \mathbf{A}_{t+1}^{t+H} stands for A at all times prior to t (t included) and $\mathbf{A}_{t+1}^t = (A_{t+1}, \dots, A_t)^T$.

In an operational situation, such post-processor will be combined with a weather probabilistic forecast to produce hydrological probabilistic forecasts taking into account the two main uncertainty sources previously described. More specifically, each simulation from the future weather distribution will be inputted in the RRS as if it was the actual meteorology, and the corresponding output from the RRS will be treated by the post-processor as such. See Krzysztofowicz (2002) for more details about how to combine the two main uncertainty sources in an operational forecasting system. In what follows, we will only use the term *probabilistic predictions* instead of *forecasts* to emphasize that we do not consider the real-time forecasting framework but only predictions conditionally on the true meteorology.

The residual errors, *i.e.* the difference between \mathbf{Q} and \mathbf{M} denoted by \mathbf{W} , are known to be both auto-correlated and heteroscedastic (Sorooshian and Dracup, 1980). Heteroscedasticity can either be accounted for by transforming \mathbf{Q} and \mathbf{M} to respectively get \mathbf{Y} and \mathbf{X} (sometimes only the residual \mathbf{W} is transformed) or by explicitly expressing heteroscedasticity in the model as, for instance, done by Thyer et al. (2002); Schoups and Vrugt (2010); Pianosi and Raso (2012) and Evin et al. (2013). The transformation can be operated through a Box Cox transformation (Box and Cox, 1964) as in Sorooshian and Dracup (1980); Kuczera (1983); Bates and Campbell (2001); Vrugt and Robinson (2007); Engeland et al. (2010); Hemri et al. (2013, 2015), through a Normal Quantile Transformation as in Krzysztofowicz (2002); Todini (2008); Engeland et al. (2010) or through a log-sinh transformation (Wang et al., 2012) for instance. The logarithmic transformation, used for water flow by Collet et al. (2009), by EDF (Mathevet, 2010) and by Morawietz et al. (2011) (for example), is a special case of the Box-Cox transformation. Auto-correlation is usually accounted for, in post-processing methods, by considering an auto-regressive model (AR) (Sorooshian and Dracup (1980); Kuczera (1983); Bates and Campbell (2001); Schaeffli et al. (2007); Schoups and Vrugt (2010); Engeland et al. (2010); Pianosi and Raso (2012); Evin et al. (2013) and Evin et al. (2014)). An AR of order one (*i.e.* only taking into account the observation at the previous time step), denoted AR(1), is most often considered as sufficient (Sorooshian

and Dracup, 1980; Schaeffi et al., 2007; Schoups and Vrugt, 2010; Evin et al., 2013, 2014; Engeland et al., 2010). For instance, Sorooshian and Dracup (1980); Evin et al. (2014) and Engeland et al. (2010) consider the following AR(1) model on a transformation of the residuals \mathbf{W} , denoted by $\tilde{\mathbf{W}}$:

$$\tilde{W}_t = \rho \tilde{W}_{t-1} + \varepsilon_t, \quad \varepsilon_t \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2),$$

where ρ and σ are parameters to be estimated and $\mathcal{N}(\mu, \sigma^2)$ stands for the Gaussian distribution of mean μ and variance σ^2 . There is an abundant literature about such AR models in hydrology, see for instance Box and Jenkins (1970).

Li et al. (2015) stress out that relying on AR models can result in some over-correction of the RRS predictions. This may be due to changes in the assumed relation between \mathbf{X} and \mathbf{Y} along with changes in the behavior of the rivers. This relation may not only change in mean but, more importantly, in variance. For instance, under flood conditions, one would expect a larger residual variance than in the following water receding. We thus propose a post-processing method implying two possible dependence regimes between \mathbf{X} and \mathbf{Y} . In the literature, the methods implying several dependence regimes either model the probability of belonging to each regime (Albert and Chib, 1993; Schaeffi et al., 2007) or the probability to switch from one regime to another (Chib, 1996; Lu and Berliner, 1999; Furrer et al., 2006; Ailliot and Monbet, 2012). In the model proposed by Schaeffi et al. (2007) for instance, the residuals \mathbf{W} behave differently depending on whether the water flow exceeds or not a predetermined threshold. Furrer et al. (2006) model the water flow time-series \mathbf{Q} according to four ordered possible hydrological regimes and without resorting to any RRS. In their model, the transition probability from one regime to the following depends on auxiliary variables through a probit-type binary regression, following the approach developed by Lu and Berliner (1999). Similarly, Perreault et al. (2007) produce long-term forecasts of dams water intakes without the use of any RRS and with a model involving two possible hydrological regimes. To do so, they make use of two approaches respectively developed by Chib (1996) and by Albert and Chib (1993). In the latter approach, the probability of being in a regime or in the other depends on climatological conditions through a probit model.

The RRSs we consider in this paper do not only provide a predictor but also mimic the river behavior by a system of interconnected reservoirs. As a consequence, their internal state variables (such as snowmelt intensity, fraction of the superficial flow in the whole flow, etc.) may bring valuable information about the regimes of the river. We then propose a model in which RRS internal state variables arbitrate between two possible temporal dependence structures of $(\mathbf{Y}|\mathbf{X})$. The probability of being in a regime or in the other depends on these RRS internal state variables through a probit model which is similar to one of the approaches of Perreault et al. (2007). The use of internal state variables is inspired

by the operational method of the French national electricity provider (EDF), described in Appendix A. Conditionally on the regime, the transformed water flow \mathbf{Y} is assumed to behave according to a regime-dependent AR model with eXternal inputs (ARX). The external inputs are the transformed outputs from the RRS (\mathbf{X}).

We restrict to the most interesting case of a two-regime model which is realistic for many rivers with two contrasted modes (flood rise versus flood receding) as in European or North American countries. In addition, the extension to more than two components for the latent regime would have risen mathematical difficulties out of the scope of this paper: choosing between an ordered or a multivariate probit is not a straightforward modeling issue and leads to rather complicated inference procedures.

Inference is performed by an Expectation-Maximization (EM) algorithm (Dempster et al., 1977). Since all computations for the second order moments can be conducted explicitly within the EM framework, following the approach of Louis (1982), we also provide asymptotic confidence intervals for the EM estimates. The Bayesian Information Criterion (BIC) is used to select the best subset of explanatory state variables as well as the external terms and the auto-regressive terms of the ARX model. The BIC is a penalized likelihood criterion proposed by Schwarz (1978) in a Bayesian framework to select a model. The selected model can be then applied in predictive mode for a series of water flow predictions routinely issued by two hydroelectricity producers in France and in Québec, EDF and Hydro-Québec, and compared with the present operational predictive method (when inputs are actual meteorological conditions) in the case of watersheds managed by EDF.

Conversely to most forecasting systems in operation, the method presented here is fully probabilistic and formally model-based. In the hydrological forecasting system of Hydro-Québec, the recently observed water flows are taken into account beforehand, through data assimilation, by tuning the state variables of the RRS. Similarly, in the forecasting system of the French national electricity supplier (EDF), they are taken into account beforehand through a deterministic correction. In both cases, uncertainty on the prediction is re-introduced in a second step, through perturbations.

The paper is organized as follows. In Section 2, we describe the two RRSs considered in the present work: MORDOR (EDF) and HSAMI (Hydro-Québec). In Section 3.1, the model with latent regimes is mathematically stated and the EM inference algorithm is detailed. In Section 4, the performance of the method is first checked on artificial data and then applied to numerous real cases provided by EDF in France and by Hydro-Québec in Québec, Canada. A concluding section discusses possible extensions of the present work.

2 Global conceptual Rainfall Runoff Simulators

In this paper, we consider *global conceptual RRSs* with interconnected reservoirs (such as snow storage, surface water storage and groundwater water storage) mimicking the water pathway in the watershed. A *conceptual* RRS seeks to represent the main processes of the rainfall-flow relationship without using the physical laws governing the processes involved. In *global* RRSs, watersheds are supposed to be spatially homogeneous. We call the filling levels of the reservoirs (*i.e.* the quantity of water stored in each reservoir), increments and combinations of those filling levels, *state variables* of the simulator, and we denote them by \mathbf{V} .

Those RRSs behave as a dynamic system with deterministic transition equations: for each time t , a RRS processes (daily in the cases considered hereby) recorded weather data (*i.e.*, typically, precipitation and temperature over a watershed) by updating its state variables from the previous time, \mathbf{V}_{t-1} , notably its reservoirs filling levels, into \mathbf{V}_t . The RRS finally outputs a daily water flow prediction, M_t , which is to be compared with the actual water flow Q_t . The reservoirs filling levels are updated according to their theoretical response to weather events, to their mass balance equations and to the transferring delays within the various components of the modeled system. The relations between reservoirs are modeled through rather simple mathematical operators (with parameters previously fitted on a training database).

All global conceptual RRSs share the same basic principles. Diagrams of EDF's global conceptual RRS, MORDOR, and of HQ's global conceptual RRS, HSAMI, are given as twin examples in Figure 1. On those alike diagrams, only the precipitations inputs are explicitly represented but the temperature inputs intervene notably in the evaporation and snow melting phenomena. Table 1 describes the state variables of MORDOR and HSAMI used in the post-processing method we propose.

3 Accounting for RRS state variables: Probabilistic hydrological predictions need a two-regime model

In the subsequent sections, we will consider $\mathbf{X} = \log(\mathbf{M})$ and $\mathbf{Y} = \log(\mathbf{Q})$. In current practice, taking the logarithm of the water flow often proves a suitable transformation to achieve data normalization and model daily water flows in most cases (Collet et al., 2009; Mathevet, 2010; Morawietz et al., 2011). In Table 2, we recall notations previously introduced and give new notations.

Before stating the model, we motivate the necessity to propose a model which can adapt to different hydrological behaviors with an example.

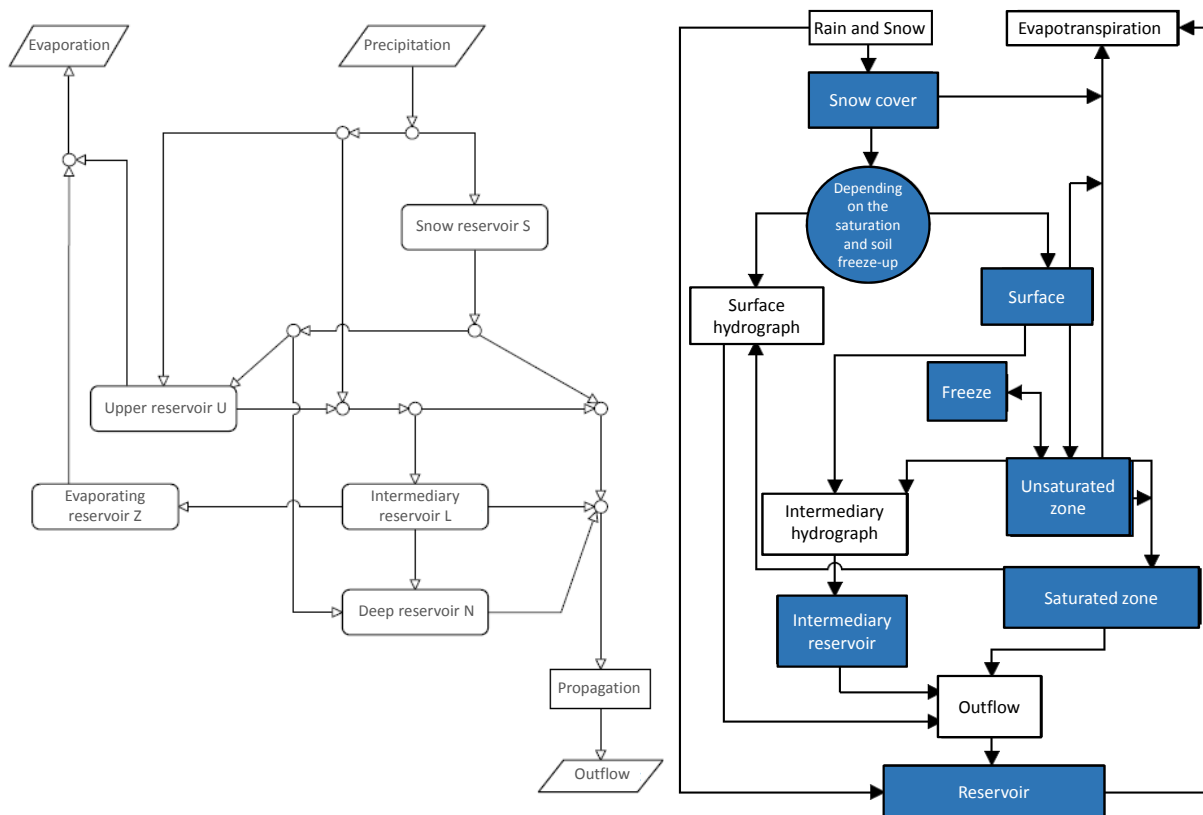


Figure 1: Diagram of EDF (left, adapted from Garçon (1996)) and HQ (right, adapted from Fortin (2000)) rainfall-run off simulators (resp. MORDOR and HSAMI)

3.1 An example of two different hydrological behaviors

Figure 2 illustrates, on an example relative to the (French) Dordogne watershed at Bort, some prediction situations commonly observed on the watersheds we consider. Such frequently observed situations suggest focusing on a model with two regimes: (i) in the common situation when the RRS ingests rainfall inputs at the time considered for the forecast, marked by a dark gray background on the figure, the link between the RRS output and the water-flow is subjected to a high level of uncertainty, (ii) in situations where there is no recent rain event to be inputted in the RRS, marked by a light gray background on the figure, the time increments of \mathbf{Y} and \mathbf{X} (daily increases of the logarithm of the water flows) are very close to each other as the state variables of the simulator are no longer updated and thus the actual error is reported identically from one time step to the next step. Consequently, the prediction uncertainty is much smaller.

Table 1: State variables of MORDOR and HSAMI used in the post-processing method (*i.e.*, at time t , elements of the vector \mathbf{V}_t). All the state variables are scaled (centered and divided by their standard deviation) before being inputted in the post-processor. The last columns indicate whether the state variable is used in the post-processor for each of the RRSs.

Name	State variable (at time t)	MORDOR	HSAMI
Const_t	1	×	×
M_t	M_t (water flow outputted by the RRS at time t)	×	
dM_t	$M_t - M_{t-1}$	×	×
$\frac{1}{M_t}$	$\frac{1}{M_t}$	×	×
$\frac{dM}{M}_t$	$\frac{M_t - M_{t-1}}{M_t}$	×	×
X_t	$X_t = \log(M_t)$	×	×
Precip_t	average precipitation over the watershed	×	
Temp_t	average temperature over the watershed	×	
Upper_t	filling level of the upper reservoir (MORDOR) or unsaturated zone (HSAMI)	×	×
$w\text{Upper}_t$	water flow originating from the upper reservoir	×	
$d\text{Upper}_t$	$\text{Upper}_t - \text{Upper}_{t-1}$	×	
Inter_t	filling level of the intermediary reservoir	×	×
$w\text{Inter}_t$	water flow originating from the intermediary reservoir	×	×
Evap_t	filling level of the evaporating reservoir	×	
PotEvap_t	potential evapotranspiration	×	×
TopEvap_t	total evapotranspiration	×	×
Deep_t	filling level of the deep reservoir (MORDOR) or saturated zone (HSAMI)	×	×
$w\text{Deep}_t$	water flow originating from the deep reservoir	×	
$\frac{w\text{Deep}}{M}_t$	$\frac{w\text{Deep}_t}{M_t}$	×	×
Snow_t	filling level of the snow reservoir	×	×
SnowMelt_t	$\mathbb{I}_{\{\text{Snow}_t < \text{Snow}_{t-1}\}}$	×	×
$\frac{w\text{Snow}}{M}_t$	portion of water flow originating from snow melt	×	×
SnowCov_t	portion of the watershed under a snow coverage	×	
MeltTemp_t	1 if the melting temperature is overpassed day t		×
Freeze_t	amount of freezed water		×

3.2 Model

We would like our model to properly identify different regimes such as the ones presented above and to predict which one is adequate for any prediction situation. We propose to make the conditional distribution of $(Y_t | \mathbf{X}_{\dots}^t, \mathbf{Y}^{t-1})$ depend on the sign of a latent variable denoted

t	time (day) index
M_t	water flow outputted by the RRS at time t
Q_t	actual water flow at time t
X_t	$\log(M_t)$: logarithm of the water flow outputted by the RRS at time t
Y_t	$\log(Q_t)$: logarithm of the actual water flow at time t
\mathbf{V}_t	vector of state variables in the RRS at time t
Z_t	latent variable indicating the regime of the river at time t
ϕ	probability density function of the standard Gaussian distribution
Φ	cumulative distribution function of the standard Gaussian distribution
\mathbf{A}_{\dots}^t	A at all times prior to t (t included)
$\mathbf{A}_{t'}^t$	$(A_{t'}, \dots, A_t)^T$
$[A = a B = b]$	conditional probability density function of A in a , given the event $\{B = b\}$, if A is continuous, its mass probability if A is discrete
$\mathbb{I}_{\{A=a\}}$	indicator function of the event $\{A = a\}$

Table 2: Notations

by Z_t indicating the regime of the river and thus the nature of the relationship between Y_t , \mathbf{X}_{\dots}^{t+1} and \mathbf{Y}_{\dots}^t . We assume that such a latent variable follows a Gaussian distribution, the mean of which is a function of the RRS state variables at the considered time, \mathbf{V}_t . Thus, this model exhibits two regimes that we name regime 0 and regime 1. The regime is given by the binary random variable $S_t = \mathbb{I}_{\{Z_t \leq 0\}}$ (where $\mathbb{I}_{\{A=a\}}$ is the indicator function of the event $\{A = a\}$) which says whether the RRS behaves according to regime 0 or regime 1 at time t . Let \mathbf{V}_t denote the vector of state variables of the RRS at time t and we assume that:

$$Z_t \underset{i.i.d.}{\sim} \mathcal{N}(\mathbf{B}^T \mathbf{V}_t, 1), \quad (1)$$

where the vector \mathbf{B} contains unknown parameters to be estimated. We emphasize that we only assume that the Z_t s (hence the S_t s) are independent conditionally to the state variables \mathbf{V}_t . Hence, the regimes at time t and $t + 1$ are still dependent since \mathbf{V}_t and \mathbf{V}_{t+1} are dependent.

We moreover assume that, conditionally on $(S_t = k)$, $(\mathbf{Y}|\mathbf{X})$ behaves according to an ARX model given by the following equation:

$$Y_t = a_k + \mathbf{b}_k^T \mathbf{X}_t^{t-r} + \mathbf{c}_k^T \mathbf{Y}_{t-1}^{t-s} + \sigma_k \varepsilon_t \quad \varepsilon_t \underset{i.i.d.}{\sim} \mathcal{N}(0, 1) \quad (2)$$

where $\mathbf{Y}_{t-1}^{t-s} = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-s})^T$, $\mathbf{X}_t^{t-r} = (X_t, X_{t-1}, \dots, X_{t-r})^T$ and $\boldsymbol{\theta} = (\theta_k)_{k \in \{0,1\}} = (a_k, \mathbf{b}_k^T, \mathbf{c}_k^T, \sigma_k)_{k \in \{0,1\}}$ stands for the set of unknown parameters $\{a_k, \sigma_k\}_{k \in \{0,1\}}$ and vectors of parameters $\{\mathbf{b}_k, \mathbf{c}_k\}_{k \in \{0,1\}}$ to be estimated (in addition to the vector \mathbf{B}). Whatever the regime k , \mathbf{b}_k is of size $r + 1$ and \mathbf{c}_k is of size s .

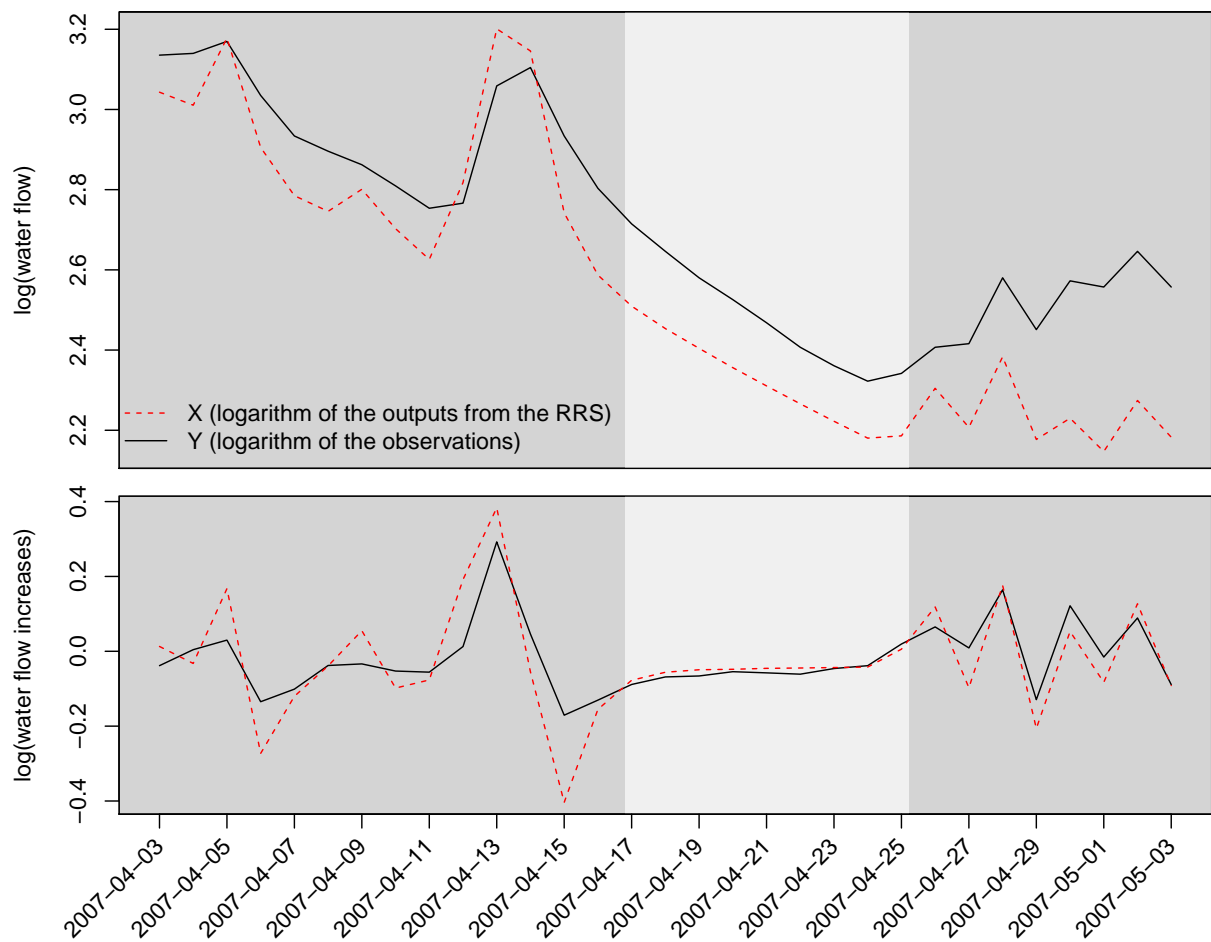


Figure 2: Illustration of two possible regimes for the water flow simulations of MORDOR on the Dordogne watershed at Bort in april 2007. The two regimes are represented by the dark and light gray backgrounds and are associated with, respectively, high and low uncertainty for the prediction of the water flows conditionally on past observed water flows and on simulations from the RRS.

Thus, the statistical modeling of the link between the outputs of the RRS, X_t, X_{t-1}, \dots , and the actual water flow, Y_t, Y_{t-1}, \dots , has two regimes ($S_t = 0$ or $S_t = 1$) which depend on the latent variable Z_t governed by the known state variables \mathbf{V}_t .

3.3 Inference

Inference is conducted conditionally on \mathbf{X} by the EM algorithm and this conditioning is omitted in the following to keep the notations simple. We adopt here the bracket notation for probability distribution as advocated in Gelfand and Smith (1990). Namely, once a probabilistic framework has been assumed and specified for the couple of random variables A and B , $[A = a|B = b]$ stands for the conditional probability density function of A in a , given the event $\{B = b\}$, if A is continuous, or its mass probability if A is discrete.

The complete likelihood associated to the ARX model given by Equation (2) and by the assumption in Equation (1), $[\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]$, is decomposable as:

$$[\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}] = \prod_{t>t_{min}} [Y_t|\mathbf{Y}_{t-1}^{t-s}, Z_t; \boldsymbol{\theta}][Z_t; \mathbf{B}]. \quad (3)$$

Let us recall that $[Y_t|\mathbf{Y}_{t-1}^{t-s}, Z_t; \boldsymbol{\theta}] = [Y_t|\mathbf{Y}_{t-1}^{t-s}, S_t; \boldsymbol{\theta}]$. Here, we assume that we are estimating the parameters conditionally on the first times of observation: $t_{min} = \max(s, r)$. The conditional distribution of $(Y_t|\mathbf{Y}_{t-1}^{t-s}, Z_t; \boldsymbol{\theta})$ is a Gaussian distribution. Then, the complete likelihood can be written as:

$$\log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]) = \sum_{t>t_{min}} \left\{ \log([Z_t; \mathbf{B}]) + \sum_k \mathbb{I}_{\{S_t=k\}} \log(\phi_{\theta_k}(Y_t)) \right\}, \quad (4)$$

where

$$\phi_{\theta_k}(Y_t) = \phi \left(\frac{Y_t - (a_k + \mathbf{b}_k^T \mathbf{X}_t^{t-r} + \mathbf{c}_k^T \mathbf{Y}_{t-1}^{t-s})}{\sigma_k} \right),$$

with ϕ the probability density function of the standard Gaussian distribution $\mathcal{N}(0, 1)$.

At iteration i , we denote by $\theta_k^{(i)}$ the current value of the vector of parameters θ_k (to be inferred).

E-step At the iteration i of the EM algorithm, the E-step consists in computing

$$\mathbb{E} \left\{ \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]) | \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)} \right\}.$$

To perform the E-step at iteration i , we first need to compute

$$\tau_{kt}^{(i)} = \mathbb{E} \left(\mathbb{I}_{\{S_t=k\}} | \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)} \right), \quad (5)$$

which is equal to

$$\begin{aligned}
\tau_{kt}^{(i)} &= \left[S_t = k \mid \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)} \right] \\
&= \frac{[Y_t \mid \mathbf{Y}_{t-1}^{t-s}, S_t = k; \boldsymbol{\theta}_k^{(i-1)}][\mathbf{Y}_{t-1}^{t-s}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)}][S_t = k; \mathbf{B}^{(i-1)}]}{[Y_t \mid \mathbf{Y}_{t-1}^{t-s}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)}][\mathbf{Y}_{t-1}^{t-s}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)}]} \\
&= \frac{\phi_{\boldsymbol{\theta}_k^{(i-1)}}(Y_t) [S_t = k; \mathbf{B}^{(i-1)}]}{\sum_{k'} \phi_{\boldsymbol{\theta}_{k'}^{(i-1)}}(Y_t) [S_t = k'; \mathbf{B}^{(i-1)}]},
\end{aligned}$$

where $[S_t = 1; \mathbf{B}^{(i-1)}] = [Z_t \leq 0; \mathbf{B}^{(i-1)}] = \Phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t)$ and $[S_t = 0; \mathbf{B}^{(i-1)}] = 1 - \Phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t) = \Phi((\mathbf{B}^{(i-1)})^T \mathbf{V}_t)$.

To perform the E-step at iteration i , we moreover need to compute $\mathbb{E}(\log([Z_t; \mathbf{B}]) \mid \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)})$ which is equal to:

$$\begin{aligned}
&\mathbb{E}(\log([Z_t; \mathbf{B}]) \mid \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)}) \\
&= \sum_k \int_{z: S_t=k} \log([Z_t = z; \mathbf{B}]) \frac{\phi_{\boldsymbol{\theta}_k^{(i-1)}}(Y_t) \phi(z - (\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}{\sum_{k'} \phi_{\boldsymbol{\theta}_{k'}^{(i-1)}}(Y_t) [S_t = k'; \mathbf{B}^{(i-1)}]} dz \\
&= \sum_k \frac{\phi_{\boldsymbol{\theta}_k^{(i-1)}}(Y_t) [S_t = k; \mathbf{B}^{(i-1)}]}{\sum_{k'} \phi_{\boldsymbol{\theta}_{k'}^{(i-1)}}(Y_t) [S_t = k'; \mathbf{B}^{(i-1)}]} \int \log([Z_t = z; \mathbf{B}]) \frac{\phi(z - (\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}{[S_t = k; \mathbf{B}^{(i-1)}]} \mathbb{I}_{\{z: S_t=k\}} dz.
\end{aligned}$$

The set $\{z : S_t = k\}$ either corresponds to $z > 0$ (if $k = 0$) or to $z \leq 0$ (if $k = 1$), hence the integration is done with respect to the density of a truncated Gaussian distribution.

Then, by using well-known formulas to compute the mean of a truncated Gaussian distribution (see for instance Johnson et al. (1994)), we obtain up to a constant (independent of \mathbf{B}) denoted by C :

$$\mathbb{E}(\log([Z_t; \mathbf{B}]) \mid \mathbf{Y}; \boldsymbol{\theta}^{(i-1)}, \mathbf{B}^{(i-1)}) = C + \sum_k \tau_{kt}^{(i)} \left(-(\mathbf{B}^T \mathbf{V}_t)^2 / 2 + \mathbf{B}^T \mathbf{V}_t \cdot E_{kt}^{(i)} \right)$$

where

$$\begin{aligned}
E_{1t}^{(i)} &= (\mathbf{B}^{(i-1)})^T \mathbf{V}_t - \frac{\phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}{\Phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}, \\
E_{0t}^{(i)} &= (\mathbf{B}^{(i-1)})^T \mathbf{V}_t + \frac{\phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}{1 - \Phi(-(\mathbf{B}^{(i-1)})^T \mathbf{V}_t)}.
\end{aligned}$$

M-step The maximization with respect to \mathbf{B} is then provided by:

$$\mathbf{B}^{(i)} = \left(\sum_{t>t_{min}} \mathbf{V}_t \mathbf{V}_t^T \right)^{-1} \cdot \sum_{t>t_{min}} \sum_k \tau_{kt}^{(i)} E_{kt}^{(i)} \mathbf{V}_t.$$

The maximization with respect to the parameters θ_k corresponding to the Gaussian distribution is standard and given by:

$$\begin{aligned} \gamma_k^{(i)} &= \left(\sum_{t>t_{min}} \tau_{kt}^{(i)} \mathbf{U}_t \mathbf{U}_t^T \right)^{-1} \cdot \sum_{t>t_{min}} \tau_{kt}^{(i)} \mathbf{U}_t Y_t, \\ \sigma_k^{2(i)} &= \left(\sum_{t>t_{min}} \tau_{kt}^{(i)} \right)^{-1} \sum_{t>t_{min}} \tau_{kt}^{(i)} (Y_t - (\gamma_k^{(i)})^T \mathbf{U}_t)^2, \end{aligned}$$

with $\mathbf{U}_t^T = (1, (\mathbf{X}_t^{t-r})^T, (\mathbf{Y}_{t-1}^{t-s})^T)$ and $(\gamma_k^{(i)})^T = (a_k, (\mathbf{b}_k^{(i)})^T, (\mathbf{c}_k^{(i)})^T)$

Fisher information matrix The observed Fisher Information Matrix (FIM) will provide confidence intervals for the estimated parameters. By applying the formula given in Louis (1982), one can compute the Hessian matrix associated with the integrated likelihood from the complete likelihood. The observed FIM is given by:

$$I_{(\boldsymbol{\theta}, \mathbf{B})} = -\frac{\partial^2}{\partial(\boldsymbol{\theta}, \mathbf{B})^2} \log([\mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}]).$$

Louis's formula states that:

$$\begin{aligned} \frac{\partial^2}{\partial(\boldsymbol{\theta}, \mathbf{B})^2} \log([\mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}]) &= \mathbb{E} \left(\frac{\partial}{\partial(\boldsymbol{\theta}, \mathbf{B})} G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) | \mathbf{Y} \right) + \mathbb{E} (G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) \cdot G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z})^T | \mathbf{Y}) \\ &\quad - \mathbb{E} (G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) | \mathbf{Y}) \mathbb{E} (G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) | \mathbf{Y})^T, \end{aligned}$$

where the score G is the gradient of the complete log-likelihood: $G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) = \frac{\partial}{\partial(\boldsymbol{\theta}, \mathbf{B})} \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}])$.

If $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{B}}) = \arg \max_{(\boldsymbol{\theta}, \mathbf{B})} \log([\mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}])$, then the last term is zero. We have

$$G_{(\boldsymbol{\theta}, \mathbf{B})}(\mathbf{Y}, \mathbf{Z}) = \left((G_{\theta_k}(\mathbf{Y}, \mathbf{Z}) = \frac{\partial}{\partial \theta_k} \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]))_{k \in \{0,1\}}, G_{\mathbf{B}}(\mathbf{Y}, \mathbf{Z}) = \frac{\partial}{\partial \mathbf{B}} \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]) \right)^T,$$

with

$$G_{\theta_k}(\mathbf{Y}, \mathbf{Z}) = \left(G_{\gamma_k}(\mathbf{Y}, \mathbf{Z}) = \frac{\partial}{\partial \gamma_k} \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]), G_{\sigma_k^2}(\mathbf{Y}, \mathbf{Z}) = \frac{\partial}{\partial \sigma_k^2} \log([\mathbf{Y}, \mathbf{Z}; \boldsymbol{\theta}, \mathbf{B}]) \right)^T.$$

The detailed computations are provided in Appendix B.

Hereafter, the first index of a parameter refers to its position in the parameter vector and the second one refers to the value of k : $\mathbf{b}_k = (b_{1k}, \dots, b_{(r+1)k})^T$ and $\mathbf{c}_k = (c_{1k}, \dots, c_{sk})^T$.

3.4 Model selection

We propose a backward stepwise model selection procedure. This selection should help avoid overfitting. Beforehand, a re-parametrization of the parameters of $\{\gamma_k\}_{k \in \{0,1\}}$ is performed such that model testing results in testing whether a collection of parameters values can be set to zero. For instance, if the constraint “ $c_{10} = 1$ ” seems relevant (when looking at the inference results on several watersheds), we re-parametrize c_{10} into $c'_{10} = c_{10} - 1$ and then test, in our model selection procedure, whether $c'_{10} = 0$. Similarly, if the constraint “ $b_{20} = -b_{10}$ ” seems relevant, we re-parametrize b_{20} into $b'_{20} = b_{20} + b_{10}$ and then test, in our model selection procedure, whether $b'_{20} = 0$.

The model selection procedure we propose is described in Algorithm 1. It is based on the BIC and on the absolute significant index. The BIC is a penalized likelihood which is computed as the deviance of the model plus the number of free parameters associated with the AR relationships (θ) and the internal state variables (\mathbf{B}). A model with the lowest BIC is to be preferred. The *absolute significant index* associated to a parameter, used in the model selection procedure, is defined by $\left| \frac{\hat{\beta}}{\hat{\sigma}_\beta} \right|$, where $\hat{\beta}$ stands for the estimated value of the parameter and $\hat{\sigma}_\beta$ stands for the standard error (on the parameter value estimation) estimated in the FIM.

Algorithm 1: Model selection algorithm for the two-regime model.

Initialization:

Re-parametrize the model such that each constraint to be tested corresponds to a parameter value to be set to zero;

Set $BIC^{(0)} = +\infty$;

Set $j = 1$;

Proceed to the inference of the (unconstrained) model parameters;

Estimate the associated FIM and compute the associated BIC: $BIC^{(1)}$;

while $BIC^{(j)} \leq BIC^{(j-1)}$ **do**

Step 1 Compute the absolute significant index associated to each (not yet constrained to zero) parameter;

Step 2 Constrain the value of the parameter with the lowest absolute significant index to zero;

Step 3 Set $j = j + 1$;

Step 4 Proceed to the inference of the parameters of the model with the new constraint in addition to the previously selected constraints;

Step 5 Estimate the associated FIM and compute the associated BIC: $BIC^{(j)}$;

end

Take the parameters obtained at iteration $(j - 1)$.

Computational experiments (not reported here) show that proceeding to a BIC-only model selection would result in the similar final constraint selection as in our case. In what we call a BIC-only model selection procedure, Step 3 is kept, Steps 4 and 5 are removed and Steps 1, and 2 are replaced with:

Step 1 Proceed to the inference of the parameters of the models with every possible new constraint and compute associated BICs.

Step 2 Keep the additional constraint for which the lowest BIC is reached. Define then $BIC^{(j)}$ as the lowest BIC.

Even if the backward selection procedure relying only on the BIC remains feasible here, we prefer to take advantage of the absolute significant index that we can compute at no additional cost. The absolute significant index makes it possible to proceed to model selection for as many state variables (up to 25 state variables in the cases considered hereby) and as many auto-regressive terms ($r = 6$ and $s = 6$ in the cases considered hereby, where r is the order of the auto-regression and s is the number of \mathbf{X} elements taken into account for each prediction time) as required, to test our model on many watersheds and possibly in many different contexts, and eventually to proceed to statistical learning with a sliding window if needed.

3.5 Probabilistic prediction

Once inference has been conducted, the probabilistic prediction of future water flow given the outputs and the state variables of the RRS is obtained as a mixture of Gaussian distributions with weights computed from a probit link which depends on the state variables.

More specifically:

$$(S_{t+1}|\mathbf{B}) \sim \mathcal{B}(\Phi(-\mathbf{B}^T \mathbf{V}_{t+1}))$$

$$(Y_{t+1}|\mathbf{Y}^t, \mathbf{X}^{t+1}, S_{t+1} = k; \boldsymbol{\theta}) \sim \mathcal{N}(a_k + \mathbf{b}_k^T \mathbf{X}_{t+1}^{t-r+1} + \mathbf{c}_k^T \mathbf{Y}_t^{t-s+1}, \sigma_k^2),$$

where $\mathcal{B}(\pi)$ stands for the Bernoulli distribution of parameter π .

The common practice in operational systems is to use scenarios, also dubbed *ensemble forecasts*. To simulate a predictive sample of water flow at time t for the future times up to $t + H$, one can generate a series of regimes from the independent Bernoulli distributions of \mathbf{S}_{t+1}^{t+H} using the state variables of the RRS for those future times (\mathbf{V}_{t+1}^{t+H}). Then, given the simulated set of regimes, the corresponding and past outputs from the RRS (realizations of $\mathbf{M}_{t-r+1}^{t+H} = \exp(\mathbf{X}_{t-r+1}^{t+H})$), and the recently observed water flows (realizations of $\mathbf{Q}_t^{t-s+1} = \exp(\mathbf{Y}_t^{t-s+1})$), generate simulations from the multivariate Gaussian distribution of $(\mathbf{Y}_{t+1}^{t+H}|\mathbf{Y}_t^{t-s+1}, \mathbf{X}_{t+H}^{t-r+1}, \mathbf{S}_{t+1}^{t+H}; \boldsymbol{\theta})$ and take the exponential to get simulations from the predictive distribution of the water flows. This multivariate distribution can be derived from Equation (2) by moving to matrix notations.

4 Numerical illustrations

4.1 Artificial data

We first check our estimation methods on artificial data. The goals are to check whether the EM algorithm gives a good estimation of the parameters, whether the Fisher Information Matrix properly estimates the standard error of the inference procedure, and whether our model selection method chooses the appropriate model.

We simulate artificial data according to the models described in the previous sections and we keep as close as possible to the behavior of the time series of the (French) Dordogne watershed at Bort by choosing parameters close to the one obtained for this watershed from 1995 to 2004. The \mathbf{X} time series of the Dordogne watershed at Bort from 1995 to 2004 and three corresponding state variables (“Const”, “dUpper” and “SnowMelt”) are taken as such. See Table 1 for a description of those state variables. The conditional model ($\mathbf{Y}|\mathbf{X}$) with the following parameters is run to produce replicates of the artificial \mathbf{Y} time series:

$$\begin{aligned}\mathbf{B} &= (0.4, -0.9, 0.0)^T, \quad r = s = 6, \\ a_0 &= 0.0, \quad a_1 = 0.1, \quad \sigma_0 = 0.1, \quad \sigma_1 = 0.3, \\ \mathbf{b}_0 &= (0.4, -0.4, 0.0, 0.0, 0.0, 0.0, 0.0)^T, \quad \mathbf{b}_1 = (0.9, -0.5, 0.0, 0.0, 0.0, -0.2, 0.0)^T, \\ \mathbf{c}_0 &= (1.0, -0.1, 0.0, 0.0, 0.0, 0.0)^T, \quad \mathbf{c}_1 = (0.5, 0.1, 0.0, 0.0, 0.1, 0.0)^T.\end{aligned}$$

For each simulation, we launch the EM algorithm previously described. This experiment is repeated 1000 times. We initialize the estimation with starting values based on a simple regression model (*i.e.* the one-regime model corresponding to our two-regime model). To avoid label switching (*i.e.* switches between $k = 0$ and $k = 1$), the starting value of c_{10} is initially set higher than the corresponding regression estimate while c_{11} is initially set lower than it and we proceed similarly for the other parameters. The EM algorithm is then run until the maximal relative difference between the estimated parameters from one iteration to the next one is equal to or less than 1.10^{-4} and the log-likelihood rises by less than 0.5. In the considered cases (with training set of 3653 units), it takes a few seconds to run the inference on R: typically 200 iterations of the EM algorithm are required, which takes about 15 s.

The resulting estimated parameters are illustrated in Figure 3. The biases are notably close to 0. In this figure as in the rest of the article, by convention, the light gray and dark gray backgrounds respectively correspond to the low uncertainty regime (regime 0) and to the high uncertainty regime (regime 1). Note that if a parameter of the vector \mathbf{B} is negative, then the probability of belonging to the regime 1 is larger when the corresponding state variable takes a larger value (and conversely when the parameter of the vector \mathbf{B} is positive).

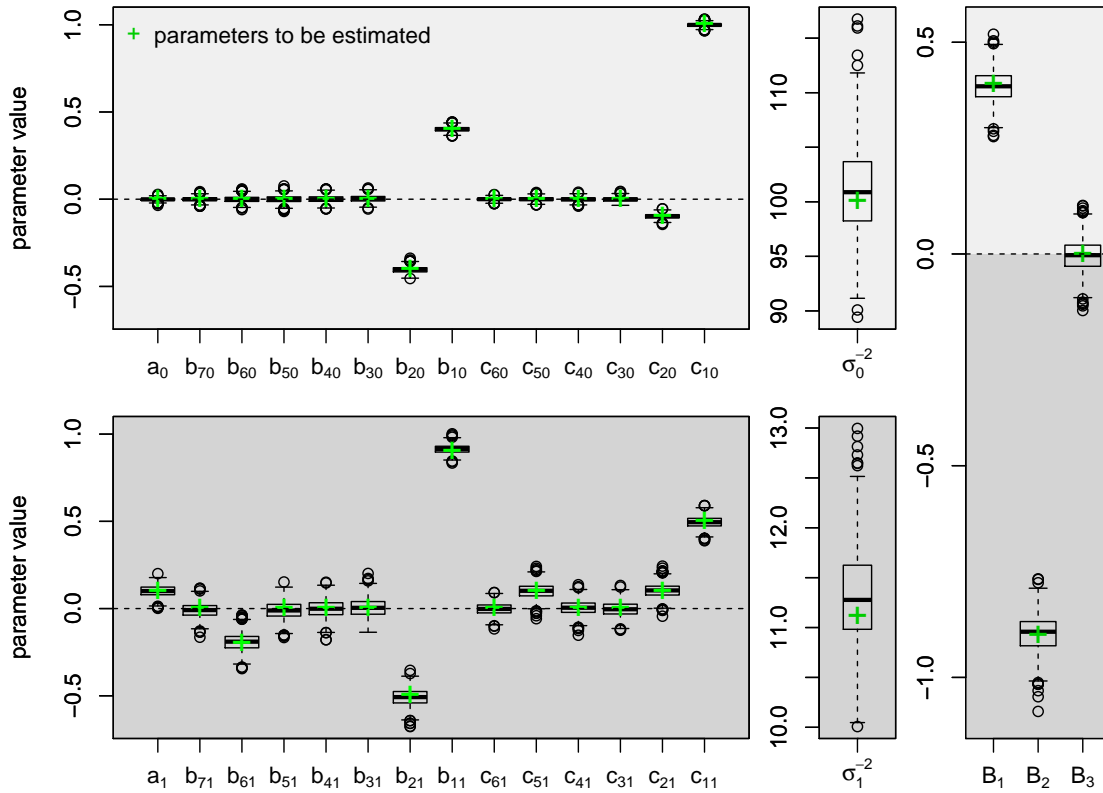


Figure 3: Estimated parameters over 1000 simulations and respective true value of parameters. The light gray and dark gray backgrounds respectively correspond to the regime 0 and to the regime 1.

Coverage rates based on Fisher information matrix 95% confidence intervals are given in Figure 4. We expect an empirical coverage rate of 95%. The resulting coverage rates are rather satisfying: the number of simulations (out of 1000) for which the Fisher information matrix 95% confidence interval contains the true parameter (“coverage event”) is about 950 for most of the parameters. Only the state variable parameters (\mathbf{B}) are not covered often enough (around 770 coverage events out of 1000 simulations). This may arise from correlation between the state variables.

Eventually, the backward stepwise model selection procedure we propose in Section 3.4 is performed. The results of the model selection procedure on artificial data is given in Figure 5. The real constraints are almost always selected (2% of false negatives: 431 out of 1000 simulations and 21 constraints to be selected). Inadequate constraint selections mainly affect the auto-regressive terms with the smallest values ($a_{11} = c_{21} = c_{51} = 0.1$, $b_{11} - 1 = -0.1$

and $b_{61} = -0.2$) and in the regime with the larger uncertainty ($\sigma_1 = 0.3$), which are probably mistaken for 0 by the algorithm.

Additional simulations on artificial data (not reported here) show that the performances of the inference algorithm remains acceptable (*i.e.* no significant bias is observed) for shorter training sets (down to 100 days) and for artificial data simulated accordingly to other watersheds.

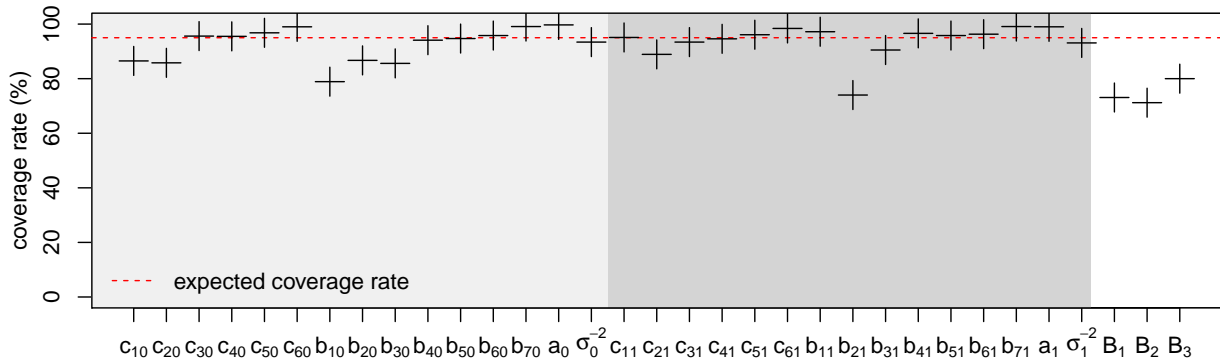


Figure 4: Coverage rates of FIM 95% confidence intervals (%) for 1000 simulations.

4.2 Application to water flow data from France and Québec

To assess the performance of the two-regime model in an operational context, EDF and Hydro-Québec provided us with past records of daily water flows and meteorological data for some of their watersheds of interest for hydroelectricity production (11 watersheds for EDF and 4 watersheds for Hydro-Québec, from back to 1948 and until 2013). They also supplied the RRSs of their operational systems. In addition, EDF gave its own probabilistic predictions processed with regard to hydrological uncertainty, from 2005 to 2008. The operational EDF method is described in Appendix A.

Inference and model selection results

Inference and model selection results for the 11 French watersheds and the 4 Québec watersheds are illustrated in Figure 6 and Figure 7. The parameters set to zero (after reparametrization and model selection) do not appear on those figures. The training set ranges from 1995 to 2004 (3653 days), we use the model selection procedure described Section 3.4 and the EM algorithm is run until convergence as in the artificial data experiment described in Section 4.1.

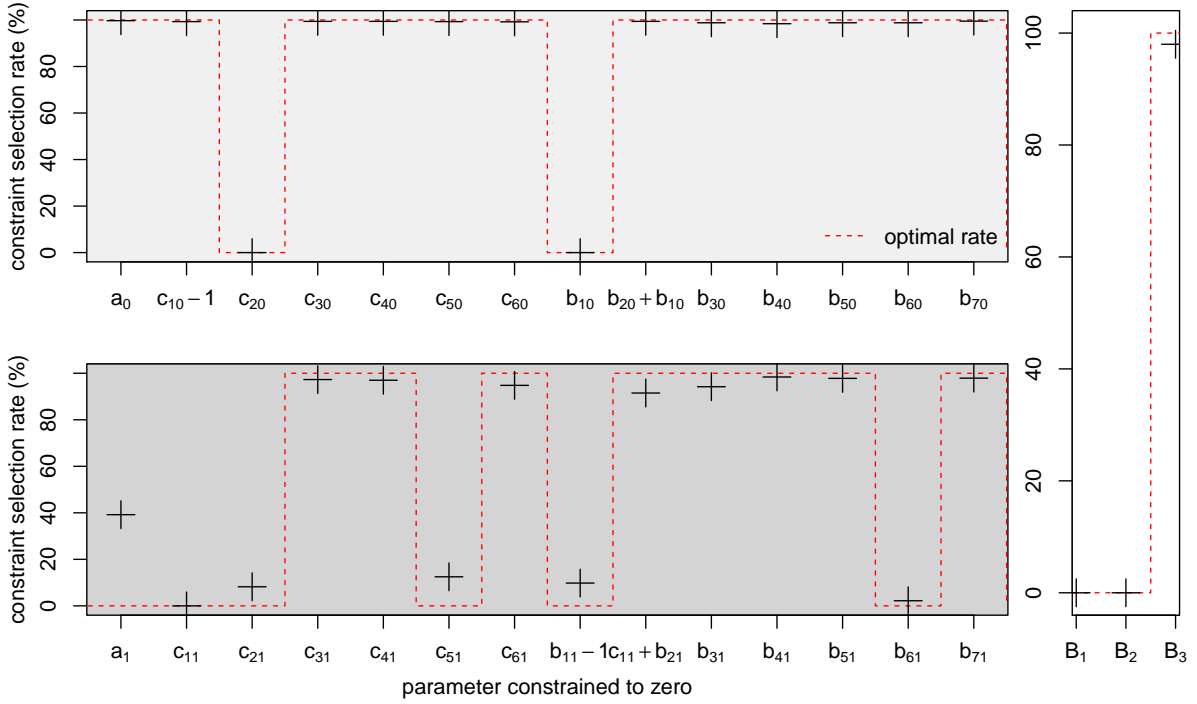


Figure 5: Constraint selection rates (%) for 1000 simulations.

As far as the BIC is concerned, we do not need to check beyond $r = 6$ and $s = 6$, the orders of the auto-regressive relation with respect to \mathbf{X} and \mathbf{Y} , for all considered watersheds: additional terms would generally be constrained to have a zero value by the model selection procedure (*i.e.* this constraint would contribute to a BIC reduction).

On Figures 6 and 7, the light gray background parts contain the regime 0 parameters while the dark gray background parts contain the regime 1 parameters. The size of the points is proportional to the absolute significant indexes of the parameters. On Figure 6, one can identify, for most watersheds:

- A regime (by convention $S = 0$) in which approximately only the time increments of \mathbf{X} and \mathbf{Y} (*i.e.* daily increases) are considered ($b_{10} \simeq -b_{20}$ and $c_{10} \simeq 1$),
- A regime (by convention $S = 1$) in which the predicted water flow depends more on the RRS outputs ($c_{11} < c_{10}$ and $b_{11} > b_{10}$), approximately only the errors ($\mathbf{Y} - \mathbf{X}$) matter ($b_{11} \simeq 1$ and $c_{11} \simeq -b_{21}$) and with a higher uncertainty level ($\sigma_1 > \sigma_0$).

Let us recall that if a parameter of the vector \mathbf{B} is negative, then the probability of belonging to the regime 1 is larger when the corresponding state variable takes a larger value (and reversely when the parameter of the vector \mathbf{B} is positive). According to the significance

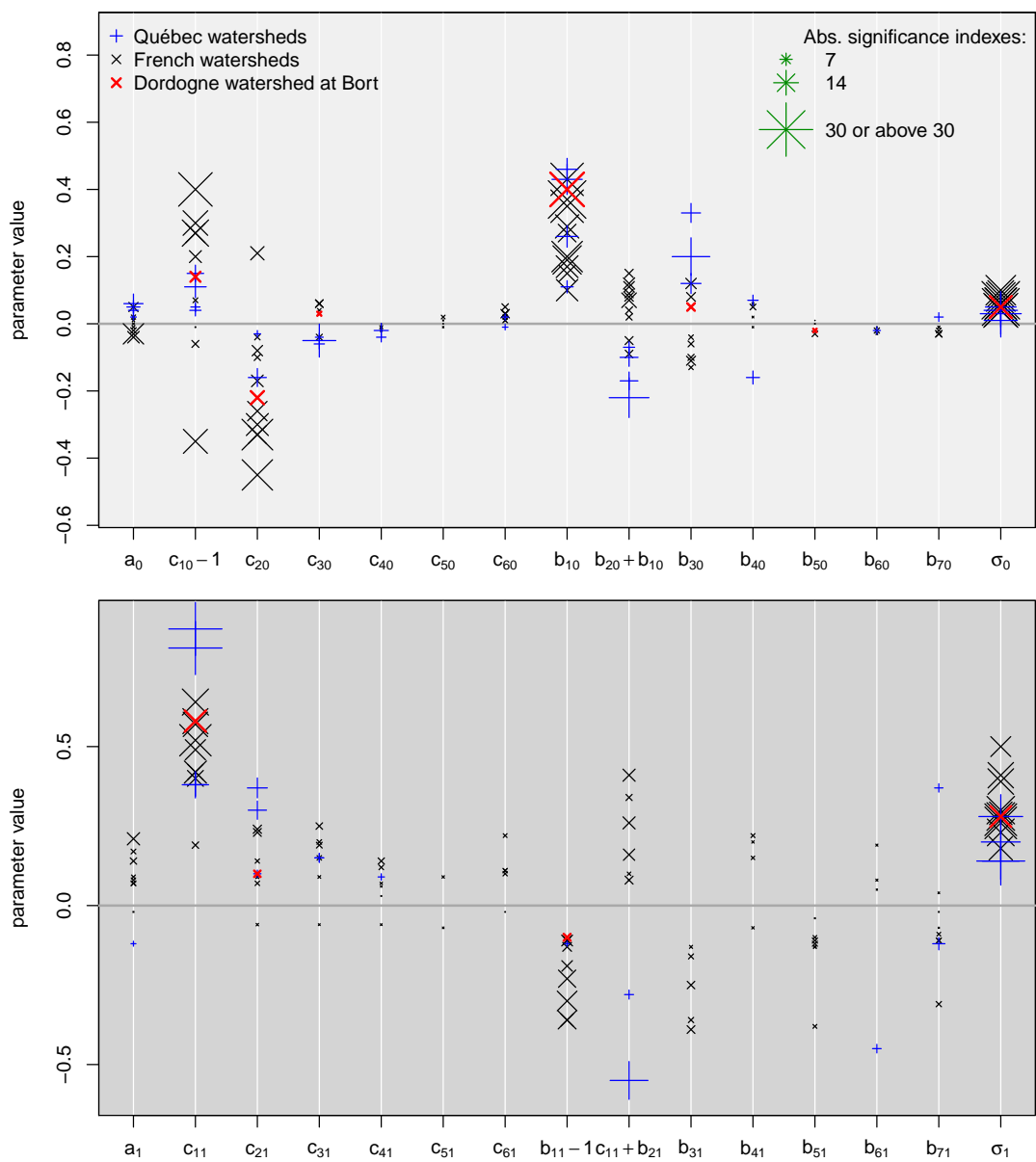


Figure 6: a_0, a_1, b_0, b_1, c_0 and c_1 parameters estimates for the 11 France watersheds and the 4 Québec watersheds when the training set ranges from 1995 to 2004. The size of the points is proportional to the value of the absolute significance indexes. The parameters set to zero do not appear. Dordogne watershed at Bort parameters are highlighted in red.

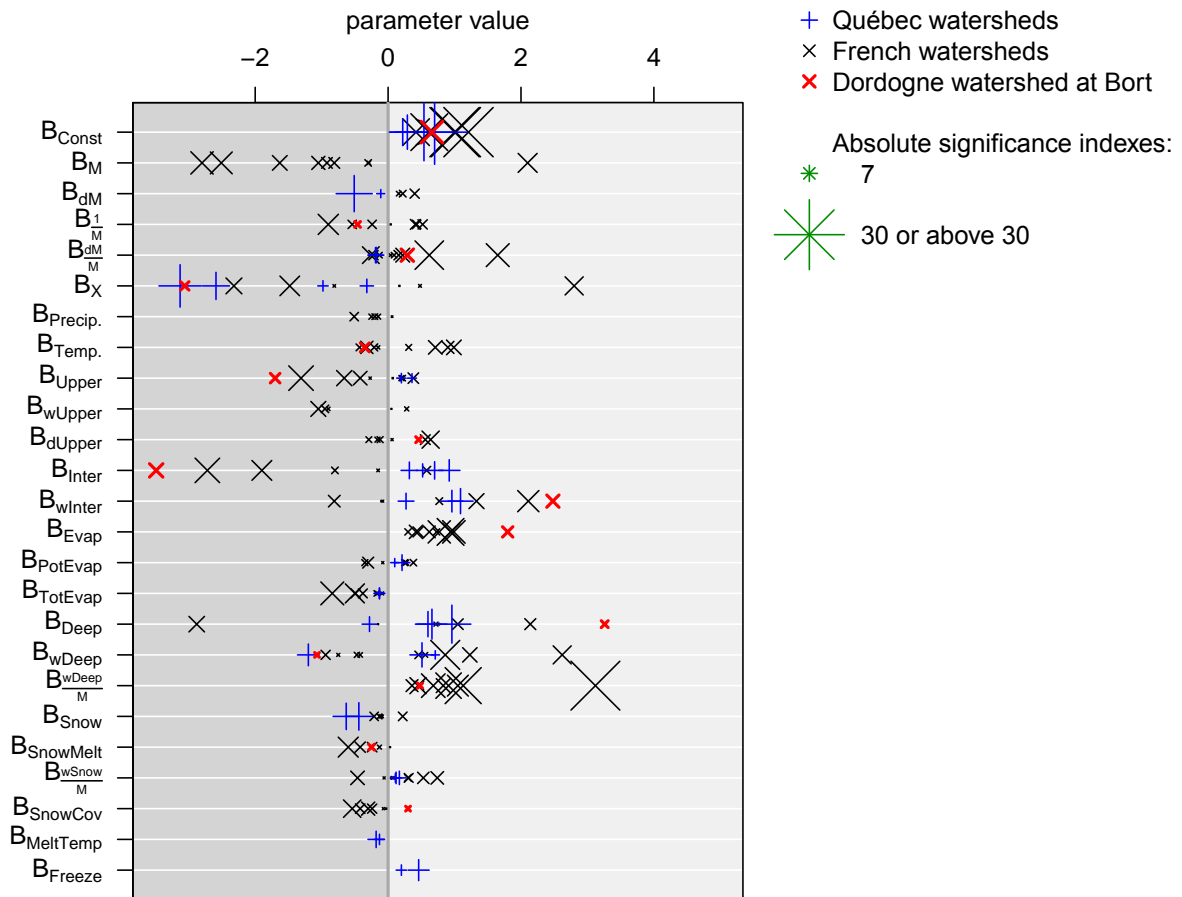


Figure 7: **B** parameters estimates for the 11 France watersheds and the 4 Québec watersheds when the training set ranges from 1995 to 2004. The size of the points is proportional to the value of the absolute significance indexes. The parameters set to zero do not appear. Dordogne watershed at Bort parameters are highlighted in red. Some state variables only exist in EDF's RRS (MORDOR) while others only exist in HQ's RRS (HSAMI).

index defined in Section 4 and to the value of the **B** parameters displayed on Figure 7, the state variables most significantly associated to the regime $S = 0$ (positive **B** terms, light gray zone) are:

- the proportion of the water flow originating from the deep reservoir ($\frac{w_{Deep}}{M}$),
- the filling level of the evaporating reservoir (“Evap”),
- the quantity of water flow originating from the intermediary reservoir (“wInter”),

- the filling level of the deep reservoir (“Deep”),
- the proportion of water originating from snow melt ($\frac{w_{\text{Snow}}}{M}$).

On the contrary, the state variables most significantly associated to the regime $S = 1$ (negative \mathbf{B} terms, dark gray zone) are:

- the value of the RRS water flow or its logarithm (“M” or “X”),
- for the French watersheds, the filling levels of the intermediary and of the upper reservoirs (“Inter” and “Upper”),
- for the Québec watersheds, the filling level of the snow reservoir (“Snow”).

The regime $S = 1$ can thus be dubbed *flood regime*, while the regime $S = 0$ can be designated as a *low-flow regime*.

Illustrative results on the Dordogne watershed at Bort

We first focus on the Dordogne watershed at Bort to analyze the prediction results. Inference and model selection results for this watershed are highlighted in red in Figure 6 and in Figure 7.

We test the prediction scheme on the years ranging from 2005 to 2008 for a forecasting horizon of 7 days (which more or less corresponds to the operational forecasting horizon for mid-term dam management). At each forecasting day, the target distribution is thus the multivariate distribution of the water flows for the 7 following days and the predictive distribution is a multivariate distribution (of dimension 7).

A prediction example is illustrated in Figure 8 in the case where the low flow mode dominates ($S = 0$) previously considered Figure 2 (month of April 2007). In the illustrated situation, the forecaster has to provide a water flow prediction the 11th of April 2007 (vertical line) for the 7 upcoming days. The available information includes the past water flow observations (black line) until that day, the RRS outputs until that day and for the next 7 days, as well as the associated state variables (not shown). The target value (unknown at the forecasting time) is the future water flow (black line) from the 12th of April to the 18th of April. When using a simple linear model inputting all the information except the state variables (one-regime model), the mean of the forecasting predictive distribution is the turquoise line. In that case, the model is a simple (ARX) regression model and we consider the ordinary least squares estimates of its parameters. When using the proposed two-regime model, the mean of the resulting predictive distribution is the deep blue line, and its associated 95% prediction intervals are illustrated in shaded blue. In the upper right box, for each forecasting day, the proportion of the dark gray filling corresponds to the predicted probability of belonging to the regime $S = 1$. Eventually, we compare our probabilistic predictions to the

ones from EDF's operational method. The mean of the operational predictive distribution is the purple line. Its associated 95% prediction intervals are illustrated in shaded purple. In

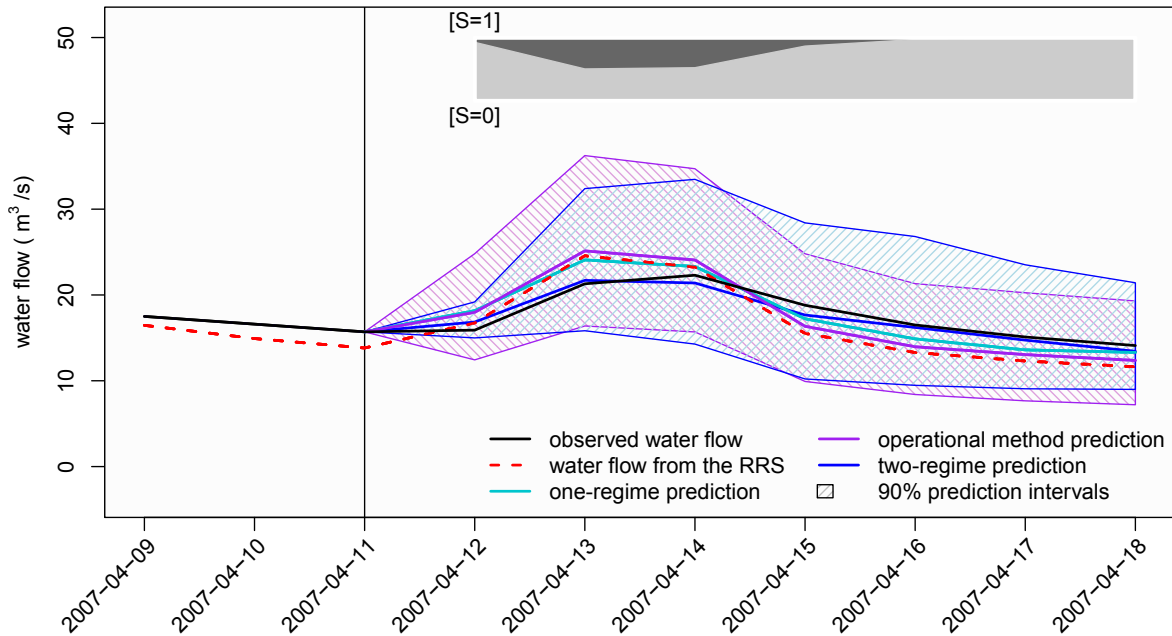


Figure 8: Water flow prediction on the Dordogne watershed at Bort. The proportion of the dark gray filling of the upper right box corresponds to the predicted probability of being in the high uncertainty (flood) regime ($S = 1$). The vertical line corresponds to the date until which one has access to the water flow observations (*i.e.* the forecasting day). The prediction targets the water flow observations after that day.

this example, the mean of the distribution resulting from the two-regime model is closer to the predictand than the means for the operational predictive method and for the one-regime prediction method. The two-regime model seems to achieve a significant improvement compared with the single regime model: as expected, the two-regime model is able to properly identify the low-flow (lower uncertainty) period.

Results for the whole testing period in terms of (lead time by lead time) Continuous Rank Probability Score (CRPS) (Matheson and Winkler, 1976; Hersbach, 2000) are illustrated in Figure 9. This score can be understood as a generalization of the absolute value error (Gneiting et al., 2005) and is estimated as follows:

$$S_F = \frac{1}{T} \sum_{t \in \text{test set}} \int (F_t(z) - \mathbb{I}_{\{z \geq y_t\}})^2 dz \quad (6)$$

where F_t is the predicted cumulative distribution function produced for the day t of the testing set of length T and y_t is the value actually taken by the variable to be predicted that day. A lower value of the CRPS indicates a better probabilistic prediction. Bootstrap 95% intervals, obtained by sampling with replacement in the testing set, are also pictured to better assess the achieved improvements.

The single regime model, the two-regime model and the operational method all significantly outperform the climatological prediction for which CRPS value is about 8.6 (not shown for readability). This *climatological* prediction serves as a reference. For a given day, it is obtained as the empirical distribution of past water flows (available from 1960 on some watersheds) which were observed around the same day of the year (20 days just before or just after the targeted day). Here, we only use the records from 1995 to 2004 to avoid problems arising from a lack of stationarity. For a given day of the year, this prediction is the same whatever the lead time: no use is made of the lastly observed water flows. According

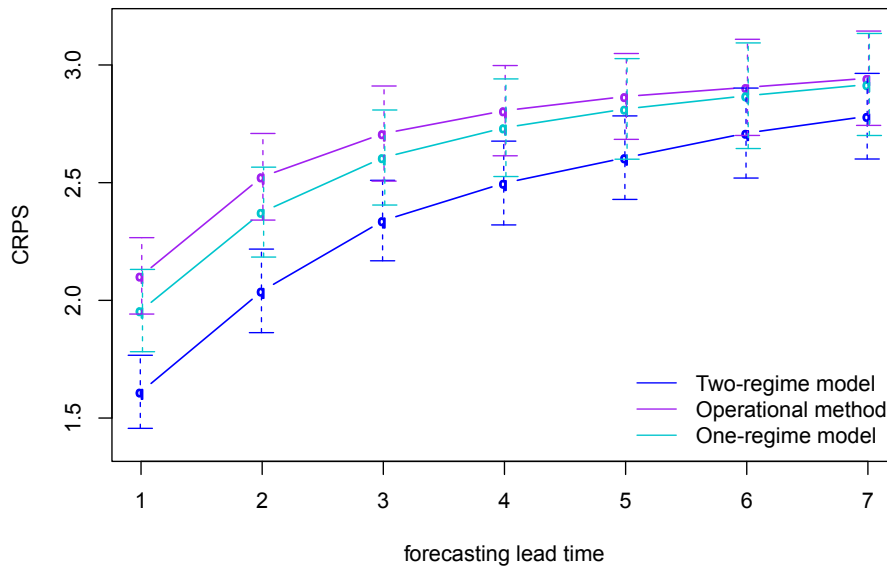


Figure 9: CRPS for the operational method, a method using a one regime ARX model and the proposed method (using a two-regime ARX model) for each forecasting lead time (in days), for the Dordogne watershed at Bort water flows on a testing set ranging from 2005 to 2008 with a training set ranging from 1995 to 2004. Bootstrap is performed to derive 95% confidence intervals for differences.

to the CRPS, the three post-processing methods achieve comparable performances and the addition of a second regime clearly improves the quality of the predictions at all lead times, especially at the first ones.

As stated by Gneiting et al. (2005), the CRPS both evaluates the reliability of the predictions and their sharpness, which are the two expected qualities from a probabilistic prediction. Reliability refers to the statistical consistency between the predictive distributions and the observations. Sharpness refers to the concentration of the predictive distributions (e.g. evaluated by the length of their prediction intervals in an univariate setting).

We expect the predictions resulting from the model we propose to be reliable. Since this model accounts for time correlation, we moreover expect the predictions to be consistent with the observations in time: the resulting predictive distributions of the accumulation of the water flows over the 7 forecasting days should also be reliable. To approximate this predictive distribution, we draw $K = 50$ simulations from the multivariate predictive distribution over the 7 forecasting days. For each simulation, we then cumulate the water flows predicted over the 7 days and obtain a value for each simulation (thus $K = 50$ values). The predictive distribution of the seven-day accumulated water flows is obtained by approximating an (empirical) distribution from the $K = 50$ values.

If the prediction is reliable, when drawing K simulations from the predictive distribution, the rank of the quantity to be predicted among the simulations should range from 1 to $K + 1$ with equal probability. The validity of this statement can be pictured by drawing the histogram of those ranks, the so-called *Rank Histogram* (RH) (Anderson, 1996; Talagrand et al., 1997). The RH for the prediction of the seven-day accumulated water flows (resulting from the 7 days ahead predictions) are illustrated in Figure 10. Those RH show some over-dispersion of both the operational method predictions, the one-regime model predictions and the two-regime model predictions: the extreme ranks are under-represented. This over-dispersion may result from the big length of the training set and is also observed at any lead time (not shown). However, the one-regime model predictions and the two-regime model predictions are slightly less over-dispersive than the operational method ones in which over-dispersion arises from the empirical copula used to recover the temporal structure, which assumes too much time correlation. We also observe such over-dispersive behaviors on the other watersheds considered hereby (not shown).

Results on all watersheds

The predicting scheme is trained on the years ranging from 1995 to 2004 and tested on the years ranging from 2005 to 2008, on the 10 other French watersheds and on the 4 Québec watersheds of the data set.

To assess the overall performance of our predicting scheme, we compute CRPS scores for the predictions of the (seven-day) accumulated water flows for all the considered watersheds. To directly compare those predictions to the ones of the climatological predictions, we furthermore compute the CRPS skill scores corresponding to those CRPS. Here, the skill score corresponds to the relative score enhancement reached with the predictive method when compared with the climatological prediction. Let us denote by S_C the CRPS for the

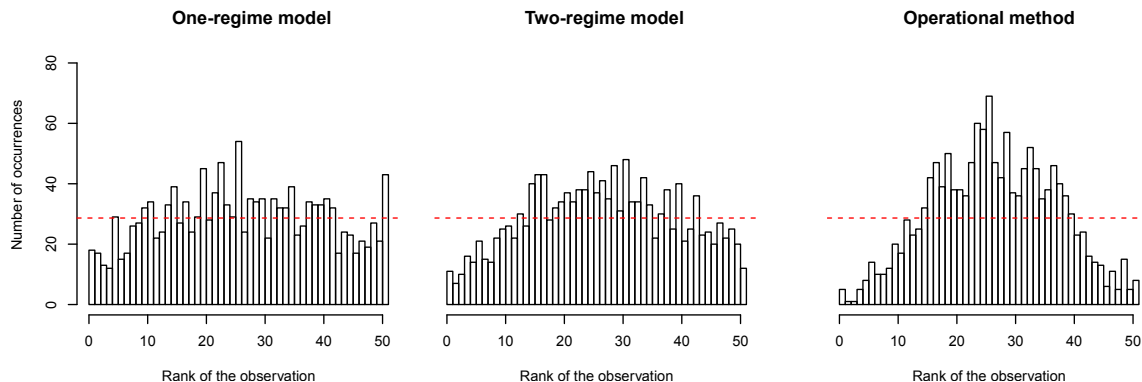


Figure 10: Rank histograms for a method using a one-regime ARX model, the proposed method (using a two-regime ARX model) and the operational method when predicting the water flow cumulated over the 7 days of prediction, for the Dordogne watershed at Bort on a testing set ranging from 2005 to 2008 with a training set ranging from 1995 to 2004. The red dotted line corresponds to the expected number of occurrences for a reliable forecast.

climatological prediction and by S_F the CRPS for the considered prediction. The CRPS skill score, $CRPSS$, is given by:

$$CRPSS = \frac{S_C - S_F}{S_C}. \quad (7)$$

The CRPSS results are illustrated in Figure 11. On the French watersheds, we can compare the probabilistic predictions based on our approach with the operational predictions. On the Québec watersheds, the operational method was not available for comparison. Results for the Dordogne watershed at Bort are pictured in the fourth column of Figure 11. The improvement of the CRPS with respect to the climatological prediction is not much affected by the move from lead time by lead time to accumulated water flow. The temporal consistency of the proposed predictive methods is thus satisfying.

According to the CRPSS, the two-regime method performs better than does a standard single regime ARX model in any case. More interestingly for hydropower generation purposes, the two-regime method shows improvement compared with the operational predictions (when known) for more than half of the watersheds and suggests the perspective of a 50% increase (with regard to CRPS) when departing from the climatology for the small subset of watersheds provided by HydroQuebec.

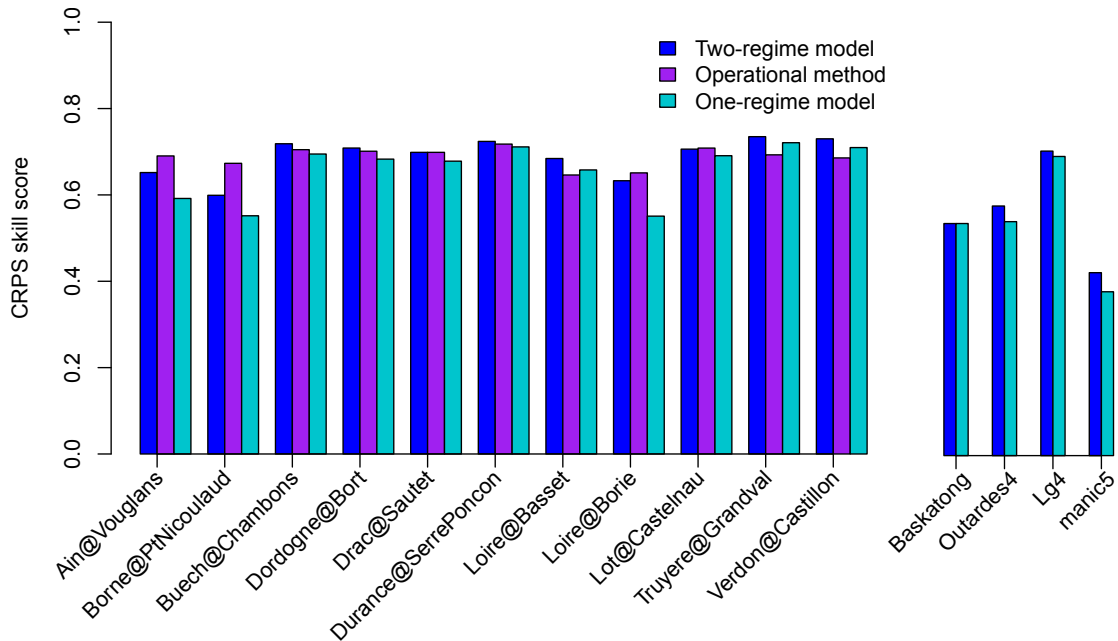


Figure 11: CRPS skill scores for the probabilistic predictions of accumulated water flows over 7 days, for the French watersheds (left) and for the Québec watersheds (right). The reference prediction is the climatological prediction. In the case of the French watersheds, the proposed method is compared with the operational method.

5 Conclusion and discussion

The goal of this work was to propose a post-processing method for RRSs allowing for two possible regimes of errors. The main idea is to use the RRS state variables as additional information indicating the type of relation (out of two) between the actual water flow and the outputs of the RRS. When applied to French and Quebec watersheds and associated RRSs, the same kinds of contrasted regimes are identified with a similar dependence on state variables: a flood regime when the time increments of the RRS errors are linearly linked and the level of uncertainty is high and a low-flow regime when only the time increments of the water flow and of the RRS are linearly linked and the uncertainty level is low.

A probit model is used to incorporate the state variables as an explanation for the choice between regimes. Inference is conducted with an exact EM algorithm and the Fisher information matrix is derived for building confidence intervals on parameters. A model selection procedure based on those confidence intervals and on the BIC removes uninformative state variables and simplifies the error model. A simulation study on artificial data shows a good behavior of both the inference and model selection procedures. Then, the probabilistic

predictions produced with our method is compared with a one-regime model and with the operational predictive method on the French watersheds. As expected, the two-regime model outperforms the single regime model but more surprisingly also outperforms the operational predictive method in more than half of the cases (according to the CRPSS). The operational predictive method often relies on several steps (e.g. data assimilation, deterministic correction, empirical error model) while our method is all-in-one.

We have applied this method to two simulators (EDF's and Hydro-Québec's) without any specific implementation trick, thus showing a nice adaptability of the method. One could imagine applying this method for any forecasting system relying on a simulator. Even if there are no intrinsic state variables, outputs of the simulator themselves, or increments of these outputs for example, can be considered as such.

A natural extension of the proposed method would be to consider additional regimes, this could be achieved by the use of an ordered probit model (in that case, there is still only one latent variable Z and there are several thresholds to determine the regime) or a multinomial probit model (in such a case, there are several latent variables and the regime is determined by the index of the largest one).

Acknowledgments: This work was supported by Électricité de France and by Hydro-Québec [research grant number 694R] through the thesis of M. Courbariaux. We would like to thank Joël Gailhard and Luc Perreault for their unfailing help and constructive comments on earlier drafts of the article. The forecasting and development teams at EDF-DTG and Hydro-Québec have provided the necessary material and case studies as well as many valuable advises ; we thank in particular Catherine Guay, Isabelle Chartier and Marie Minville from IREQ, Rémy Garçon, Matthieu Le-Lay and Federico Garavaglia from EDF-DTG. We also thank Joan Sobota for English proofreading. We finally thank the Associate Editor and the two reviewers for their comments and questions which help us to improve the paper.

References

- Ailliot, P. and Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30:92–101.
- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.
- Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7):1518–1530.

- Bates, B. C. and Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resources Research, 37(4):937–947.
- Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, Ca.
- Box, G. E. and Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), pages 211–252.
- Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. Journal of Econometrics, 75(1):79–97.
- Collet, J., Épiard, X., and Coudray, P. (2009). Simulating hydraulic inflows using PCA and ARMAX. The European Physical Journal-Special Topics, 174(1):125–134.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (methodological), pages 1–38.
- Engeland, K., Renard, B., Steinsland, I., and Kolberg, S. (2010). Evaluation of statistical models for forecast errors from the HBV model. Journal of Hydrology, 384(1):142–155.
- Evin, G., Kavetski, D., Thyer, M., and Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. Water Resources Research, 49(7):4518–4524.
- Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resources Research, 50(3):2350–2375.
- Fortin, V. (2000). Le modèle météo-apport HSAMI: historique, théorie et application. Institut de recherche d'Hydro-Québec, Varennes.
- Furrer, E. M., Jacques, C., and Favre, A.-C. (2006). Short term discharge prediction using a Markovian regime switching model. Technical report, INRS-ETE.
- Gailhard, J. (2014). Algorithme de recalage associé à MORDOR diagnostic et proposition d'améliorations. Note Technique Interne H-44200965-2014-000075, EDF-DTG.
- Garçon, R. (1996). Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. bilan de l'année 1994-1995. La Houille Blanche, (5):71–76.

- Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85(410):398–409.
- 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(5):1098–1118.
- Hemri, S., Fundel, F., and Zappa, M. (2013). Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. Water Resources Research, 49(10):6744–6755.
- Hemri, S., Lisniak, D., and Klein, B. (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. Water Resources Research, 51(9):7436–7451.
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15(5):559–570.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). Continuous Univariate Distributions, vol. 1-2. New York: John Wiley & Sons.
- Krzysztofowicz, R. (2002). Bayesian system for probabilistic river stage forecasting. Journal of Hydrology, 268(1):16–40.
- Kuczera, G. (1983). Improved parameter inference in catchment models: 1. evaluating parameter uncertainty. Water Resources Research, 19(5):1151–1162.
- Li, M., Wang, Q., Bennett, J., and Robertson, D. (2015). A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts. Hydrology and Earth System Sciences, 19(1):1–15.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. Series B, 44(2):226–233.
- Lu, Z.-Q. and Berliner, L. M. (1999). Markov switching time series models with application to a daily runoff series. Water Resources Research, 35(2):523–534.
- Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. Management Science, 22(10):1087–1096.
- Mathevet, T. (2010). Erreur empirique de modèle. Note Technique Interne D4165/NT/2010-00395-A, EDF-DTG.
- Morawietz, M., Xu, C.-Y., Gottschalk, L., and Tallaksen, L. M. (2011). Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. Journal of Hydrology, 407(1):58–72.

- Perreault, L., Garçon, R., and Gaudet, J. (2007). Analyse de séquences de variables aléatoires hydrologiques à l'aide de modèles de changement de régime exploitant des variables atmosphériques. La Houille Blanche, (6):111–123.
- Pianosi, F. and Raso, L. (2012). Dynamic modeling of predictive uncertainty by regression on absolute errors. Water Resources Research, 48(3).
- Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review, 133(5).
- Schaeffli, B., Talamba, D. B., and Musy, A. (2007). Quantifying hydrological modeling errors through a mixture of normal distributions. Journal of Hydrology, 332(3):303–315.
- Schoups, G. and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46(10).
- Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2):461–464.
- Sorooshian, S. and Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resources Research, 16(2):430–442.
- Talagrand, O., Vautard, R., and Strauss, B. (1997). Evaluation of probabilistic prediction systems. In Proc. ECMWF Workshop on Predictability, volume 1, page 25.
- Thyer, M., Kuczera, G., and Wang, Q. (2002). Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. Journal of Hydrology, 265(1):246–257.
- Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. International Journal of River Basin Management, 6(2):123–137.
- Vrugt, J. A. and Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1).
- Wang, Q., Shrestha, D. L., Robertson, D., and Pokhrel, P. (2012). A log-sinh transformation for data normalization and variance stabilization. Water Resources Research, 48(5).

A Operational predictive method

EDF's operational predictive method consists of 3 independent modules: a deterministic model, an error model and an empirical copula.

Deterministic model (Gailhard, 2014) The deterministic model in use at EDF is an auto-regressive model combined with exponential smoothing. The strength of the auto-correlation is supposed to increase with the rate of water flow coming from the deep reservoirs of the watershed.

Error model (Mathevet, 2010) The error model is a heteroscedastic conditional normal model derived for each forecasting lead time h (after normalization):

$$(Y_{t+h}|X_{t+h} = x) = b_h(x) + x + \sigma_h(x)\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1),$$

where b_h and σ_h are tabulated functions of x .

Empirical copula One finally resorts to an empirical copula to get samples of a space and time multivariate distribution from samples from the marginal (lead time by lead time) distributions.

B Fisher information matrix

$$\begin{aligned} G_{\gamma_k}(\mathbf{Y}, \mathbf{Z}) &= \sum_{t>t_{min}} \mathbb{I}_{\{S_t=k\}} (\mathbf{U}_t/\sigma_k^2 \cdot Y_t + 0 \cdot Y_t^2 - \mathbf{U}_t \mathbf{U}_t^T \gamma_k / \sigma_k^2) \\ &= \sum_{t>t_{min}} \mathbb{I}_{\{S_t=k\}} (\mathbf{U}_t/\sigma_k^2 \cdot Y_t - \mathbf{U}_t \mathbf{U}_t^T \gamma_k / \sigma_k^2), \\ G_{\sigma_k^2}(\mathbf{Y}, \mathbf{Z}) &= \sum_{t>t_{min}} \mathbb{I}_{\{S_t=k\}} \left(\frac{(Y_t - \gamma_k^T \mathbf{U}_t)^2}{2\sigma_k^4} - \frac{1}{2\sigma_k^2} \right), \\ G_{\mathbf{B}}(\mathbf{Y}, \mathbf{Z}) &= \sum_{t>t_{min}} (Z_t - \mathbf{B}^T \mathbf{V}_t) \mathbf{V}_t. \end{aligned}$$

For any $k, k' \neq k$,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\gamma}_k} G_{\boldsymbol{\gamma}_k}(\mathbf{Y}, \mathbf{Z}) &= - \sum_{t > t_{min}} \mathbb{I}_{\{S_t=k\}} \mathbf{U}_t \mathbf{U}_t^T / \sigma_k^2, \\ \frac{\partial}{\partial \sigma_k^2} G_{\sigma_k^2}(\mathbf{Y}, \mathbf{Z}) &= \sum_{t > t_{min}} \mathbb{I}_{\{S_t=k\}} \left(\frac{-(Y_t - \boldsymbol{\gamma}_k^T \mathbf{U}_t)^2}{\sigma_k^6} + \frac{1}{2\sigma_k^4} \right), \\ \frac{\partial}{\partial \boldsymbol{\gamma}_k} G_{\sigma_k^2}(\mathbf{Y}, \mathbf{Z}) &= \sum_{t > t_{min}} \mathbb{I}_{\{S_t=k\}} \left(\frac{-\mathbf{U}_t (Y_t - \boldsymbol{\gamma}_k^T \mathbf{U}_t)}{\sigma_k^4} \right), \\ \frac{\partial}{\partial \theta_{k'}} G_{\theta_k}(\mathbf{Y}, \mathbf{Z}) &= \frac{\partial}{\partial \theta_k} G_{\mathbf{B}}(\mathbf{Y}, \mathbf{Z}) = 0, \\ \frac{\partial}{\partial \mathbf{B}} G_{\mathbf{B}}(\mathbf{Y}, \mathbf{Z}) &= - \sum_{t > t_{min}} \mathbf{V}_t \mathbf{V}_t^T.\end{aligned}$$

Then, the first term in the Louis decomposition is computed since $\mathbb{E}(\mathbb{I}_{\{S_t=k\}} | \mathbf{Y}; (\boldsymbol{\theta}, \mathbf{B})) = \tau_{kt}$ is computed above.

For the second term in the Louis decomposition, we notice that:

$$\begin{aligned}\mathbb{E}(\mathbb{I}_{\{S_t=k\}}^2 | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \mathbb{E}(\mathbb{I}_{\{S_t=k\}} | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) = \tau_{kt}, \\ \mathbb{E}(\mathbb{I}_{\{S_t=k\}} \mathbb{I}_{\{S_t=k'\}} | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= 0 \text{ for } k \neq k', \\ \mathbb{E}(\mathbb{I}_{\{S_t=k\}} \mathbb{I}_{\{S_{t'}=k\}} | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \mathbb{E}(\mathbb{I}_{\{S_t=k\}} | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) \mathbb{E}(\mathbb{I}_{\{S_{t'}=k\}} | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) \text{ by independence.}\end{aligned}$$

We also need to compute:

$$\begin{aligned}\mathbb{E}(Z_t | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \sum_k \tau_{kt} E_{kt} \quad \text{and} \\ \mathbb{E}((Z_t - \mathbf{B}^T \mathbf{V}_t)^2 | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \text{var}(Z_t | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) = \sum_k \tau_{kt} s_{kt},\end{aligned}$$

where

$$\begin{aligned}s_{1t} &= 1 - \frac{\phi(-\mathbf{B}^T \mathbf{V}_t)}{\Phi(-\mathbf{B}^T \mathbf{V}_t)} \left(\frac{\phi(-\mathbf{B}^T \mathbf{V}_t)}{\Phi(-\mathbf{B}^T \mathbf{V}_t)} - \mathbf{B}^T \mathbf{V}_t \right), \\ s_{0t} &= 1 - \frac{\phi(-\mathbf{B}^T \mathbf{V}_t)}{1 - \Phi(-\mathbf{B}^T \mathbf{V}_t)} \left(\frac{\phi(-\mathbf{B}^T \mathbf{V}_t)}{1 - \Phi(-\mathbf{B}^T \mathbf{V}_t)} + \mathbf{B}^T \mathbf{V}_t \right).\end{aligned}$$

Again, we rely on the independence between the Z_t s.

The remaining terms are easily evaluated:

$$\begin{aligned}\mathbb{E}(\mathbb{I}_{\{S_{t'}=k\}} Z_t | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \tau_{kt'} \mathbb{E}(Z_t | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) \text{ by independence,} \\ \mathbb{E}(\mathbb{I}_{\{S_t=k\}} Z_t | \mathbf{Y}; \boldsymbol{\theta}, \mathbf{B}) &= \tau_{kt} E_{kt}.\end{aligned}$$

5

Conclusion, discussion et perspectives

Au cours de cette thèse, nous nous sommes intéressés à la quantification des incertitudes dans les systèmes de prévision hydrologique utilisant en intrant des prévisions météorologiques d'ensemble.

Dans le chapitre 1, nous avons défini la façon dont les méthodes proposées dans le cadre de ce travail de thèse seraient évaluées (à l'aide d'outils de diagnostic de la fiabilité et de scores évaluant à la fois la fiabilité et la précision des prévisions). Nous avons aussi présenté, dans ce chapitre, les ressources à notre disposition pour mener à bien ce travail : les SPQ d'EDF et d'Hydro-Québec et des jeux de données comprenant des prévisions météorologiques (déterministes et d'ensemble) et les historiques hydrométéorologiques (couvrant plus de 40 ans) relatifs à 15 bassins versants français et à 8 bassins versants québécois.

Dans le chapitre 2, nous avons proposé une présentation unifiée des méthodes actuelles de quantification des incertitudes dans les systèmes de prévision hydrologique. Cette revue nous a permis d'identifier les principaux axes d'amélioration de ces méthodes, notamment une meilleure utilisation des informations auxiliaires à disposition et une meilleure prise en compte des aspects multidimensionnels et de l'existence de plusieurs régimes d'erreurs. Nous avons développé de nouvelles méthodes de post-traitement, décrites dans les chapitres 3 et 4, levant partiellement ou en totalité les verrous identifiés. La section 5.1 en dresse le bilan et nous montrons, sur un exemple, les résultats du système complet issu de l'assemblage de ces méthodes dans la section 5.2

Enfin, des séjours au sein des équipes d'EDF-DTG, d'Hydro-Québec et de l'Institut de Recherche d'Hydro-Québec (IREQ) ont rythmé ce travail doctoral et permis de mettre en perspective nos travaux. Nous présentons des éléments de réflexion

issus de ces séjours dans la section 5.3.

5.1 Contributions au problème météorologique et au problème hydrologique

Des incertitudes issues de deux sources principales interviennent dans les systèmes de prévisions étudiés : l'incertitude sur la météorologie future, d'une part, et l'incertitude liée aux erreurs de modélisation de la transformation des précipitations en débit à l'exutoire des bassins versants, d'autre part. Nous avons proposé, pour chacune de ces sources d'incertitudes, une méthode de quantification s'appuyant sur un raisonnement probabiliste.

POST-TRAITEMENT MÉTÉOROLOGIQUE La méthode statistique proposée pour la prise en compte de l'incertitude sur la météorologie future est fondée sur l'hypothèse d'échangeabilité des prévisions d'ensemble. Cette hypothèse n'est pas toujours vérifiée, un compromis doit alors être trouvé entre réalisme, respect des hypothèses, parcimonie et commodité de mise en œuvre.

La structure qui en résulte est parcimonieuse, répondant ainsi aux problèmes de non-homogénéité des séries temporelles de prévisions météorologiques d'ensemble décrits dans le chapitre 1.

Cette structure permet par ailleurs d'intégrer plusieurs sources de prévisions conjointement. Chez EDF et chez Hydro-Québec, les sources de prévisions météorologiques alternatives sont actuellement prises en compte par des experts de façon manuelle. Une telle méthode de post-traitement pourrait faciliter ce travail. Il est de plus possible de considérer des prévisions issues de l'expertise comme des sources de prévisions alternatives, permettant ainsi aux experts qui le souhaitent de (continuer à) émettre des prévisions de nature déterministe.

La structure multidimensionnelle des phénomènes météorologiques est, dans nos exemples, prise en compte par l'emploi de copules empiriques. Cependant, le socle gaussien commun aux post-traitements proposés pour les prévisions de températures et de précipitations ouvre la voie à une modélisation jointe de ces deux types de variables ainsi qu'à une modélisation incluant leur structure spatiale et temporelle. On peut aussi envisager d'utiliser, plutôt que des copules empiriques, des copules paramétriques ne présentant pas l'inconvénient (des copules empiriques) de restreindre le nombre de simulations prédictives possibles au nombre de scénarios analogues à disposition.

Les résultats obtenus sur les données de températures sont tout-à-fait satisfaisants : les scores CRPS sont quasiment systématiquement améliorés par le post-traitement opéré. Les modèles proposés pour le post-traitement des précipitations le sont moins : ils donnent lieu à des prévisions relativement fiables mais leurs performances sont moins bonnes que celles des ensembles gardés tels quels sur la moitié des bassins versants considérés dans l'étude. Le caractère mixte de la variable précipitations ne permet en effet pas une modélisation simple s'appuyant sur les

propriétés de la famille gaussienne : des transformations *ad-hoc* de cette variable doivent être réalisées et compliquent fortement le traitement.

La parcimonie et la possibilité d'intégrer plusieurs sources d'informations sont des propriétés communes avec les méthodes de post-traitement par habillage proposées dans la littérature, telles que le BMA. Ces méthodes permettent de plus, contrairement aux nôtres, de produire des distributions prédictives multi-modales. Le post-traitement que nous proposons traite cependant naturellement la sous-dispersion comme la sur-dispersion des ensembles météorologiques, et la perspective de post-traiter conjointement précipitations et températures en renforce l'intérêt.

POST-TRAITEMENT HYDROLOGIQUE Nous avons proposé, pour la prise en compte de l'incertitude hydrologique, un post-traitement modélisant la structure temporelle des débits et leur dépendance aux sorties du SPQ. Ce post-traitement se distingue de ceux proposés dans la littérature en faisant intervenir les variables d'état du SPQ pour arbitrer entre deux régimes de dépendance contrastés et permettre ainsi de représenter une incertitude hydrologique à variance hétérogène.

La méthode qui en résulte est, d'une part, plus formelle et, d'autre part, d'une complexité moindre comparée à celles d'EDF et d'Hydro-Québec qui comprennent plusieurs étapes successives (dont certaines sont réalisées manuellement par des experts chez Hydro-Québec) résumées dans le tableau 2.1 (colonnes 3 et 4).

Dans tous les cas considérés, le post-traitement proposé se montre plus performant que son homologue ne permettant pas plusieurs régimes, mettant ainsi en évidence l'intérêt de notre proposition. Il présente de plus une performance, telle qu'évaluée par le score CRPS, comparable à celle de la méthode opérationnelle d'EDF : elle la surpasse pour 6 des 11 bassins versants français considérés.

Cette méthode s'adapte avec facilité à de nouveaux contextes : nous avons pu aisément transférer le post-traitement initialement proposé pour des situations de prévision d'EDF à celles d'Hydro-Québec, pourtant très différentes. On peut même imaginer utiliser cette méthode dans tout système de prévision probabiliste et séquentielle reposant sur l'utilisation d'un simulateur (ou d'un modèle déterministe).

Une extension naturelle serait l'ajout de régimes supplémentaires. L'exploitation des sorties de SPQ concurrents est aussi envisageable ainsi que la prise en compte des incertitudes portant sur les paramètres du modèle d'erreur, ceux du SPQ ou encore les intrants du SPQ, à l'exemple de Kuczera (1983); Beven & Binley (1992); Kuczera & Parent (1998); Kavetski et al. (2002); Kuczera et al. (2006); Thyer et al. (2009); Renard et al. (2011); Schoups & Vrugt (2010) et Evin et al. (2014) (entre autres).

5.2 Résultats du système complet proposé

Dans le chapitre 4, nous avons proposé une méthode permettant d'obtenir des prévisions probabilistes hydrologiques en avenir météorologique connu, soit une distribution se réclamant de $[\mathbf{Q}|\mathcal{M}(\mathbf{P})]$ (en reprenant les notations de la section 2.1). La météorologie future est toutefois sujette à des incertitudes. Les méthodes décrites dans le chapitre 3 ciblent ces incertitudes météorologiques. Elles visent à émettre des simulations selon une distribution se réclamant de $[\mathbf{P}|\mathbf{A}]$.

Un système de prévision hydrologique probabiliste complet s'obtient à partir de la méthode de post-traitement du chapitre 4 et de celles du chapitre 3 en approchant l'intégrale donnée par l'équation (2.1) :

$$[\mathbf{Q}|\mathbf{A}] = \int [\mathbf{Q}|\mathcal{M}(\mathbf{P})] [\mathbf{P}|\mathbf{A}] d\mathbf{P}.$$

L'algorithme 1, décrit dans la section 2.1, le permet.

Dans la présente section, nous illustrons sur un exemple l'application de l'algorithme 1 pour émettre des prévisions hydrologiques probabilistes en situation de prévision à partir des méthodes présentées dans les chapitres 3 et 4.

Nous prenons en exemple le bassin versant français du Drac au Sautet au cours de l'année 2014, sur lequel la méthode de post-traitement hydrologique a montré des performances comparables à celles de la méthode automatique d'EDF et sur lequel les méthodes de post-traitement météorologiques ont montré de bonnes performances.

Nous utilisons les simulations p_1, \dots, p_{50} supposées simulées selon $[\mathbf{P}|\mathbf{A}]$ issues du travail présenté dans le chapitre 3 :

- dans le cas des températures d'air, par la méthode à deux variables latentes (décrite dans la section 3.1) appliquée aux prévisions CEPMMT-EPS, CEPMMT-dét et CEPMMT-ctl,
- dans le cas des précipitations, par la méthode ayant recours à la loi des fuites (présentée dans la section 3.3) appliquée aux prévisions CEPMMT-EPS,

puis en utilisant la méthode de l'ECC afin d'obtenir des simulations de structure temporelle, spatiale et inter-variables cohérente avec celle des membres du CEPMMT-EPS (et donc, *a priori*, de la météorologie observée correspondante). Les paramètres des modèles sont appris à partir des jeux de données relatifs aux années 2011, 2012 et 2013.

Ces simulations météorologiques sont ensuite utilisées en intrants du modèle pluie-débit MORDOR. Nous obtenons ainsi m_1, \dots, m_{50} et les variables d'état associées à chacun de ces scénarios hydrologiques.

Pour chaque m_k , $k \in \{1, \dots, 50\}$, nous utilisons la méthode présentée dans le chapitre 4 pour obtenir un scénario q_k pouvant être supposé simulé selon $[\mathbf{Q}|\mathcal{M}(\mathbf{P}) = m_k]$. Les paramètres du modèle statistique sous-jacent sont appris à partir du jeu de données relatif aux années 2009 à 2013.

L'ensemble q_1, \dots, q_{50} résultant est supposé simulé selon $[Q|A]$.

La figure 5.1 illustre les prévisions qui auraient été émises avec ce système le 16 mai 2014. La part de l'incertitude d'origine respectivement météorologique et hydrologique dans ces prévisions peut être évaluée en observant la largeur des intervalles de prévision avant et après le post-traitement hydrologique. On constate que, dans cet exemple, les derniers débits observés sont très informatifs vis-à-vis des débits aux premières échéances de prévision.

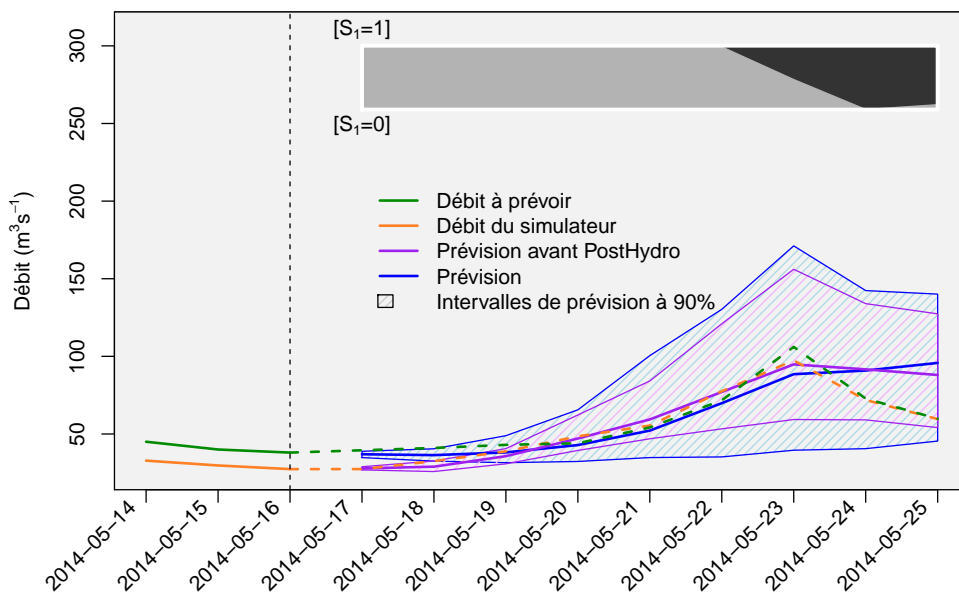


FIGURE 5.1 – En bleu, la prévision qui aurait été émise à partir de nos méthodes de post-traitement le 16 mai 2014 pour les débits journaliers à l'exutoire du bassin versant du Drac au Sautet les 9 jours suivants. Les débits à prévoir sont figurés par ligne tiretée verte. Les sorties du SPQ correspondantes en avenir météorologique connu sont figurées par une ligne tiretée orange. L'encadré grisé illustre la probabilité d'appartenance à chaque régime hydrologique en avenir météorologique connu. La prévision issue du post-traitement météorologique seul (portant sur les prévisions CEPMMT-EPS, CEPMMT-det et CEPMMT-ctl) et l'intervalle de prévision à 90% associé sont figurés en violet.

Nous comparons les prévisions produites par le système que nous proposons (prévisions CEP-PostMeteo--PostHydro) aux prévisions obtenues :

- sans appliquer de post-traitement météorologique (prévisions CEP--PostHydro),
- en appliquant le post-traitement hydrologique automatique d'EDF plutôt que celui que nous proposons (prévisions CEP-PostMeteo--PostHydroEDF).

Les histogrammes de rang des débits cumulés sur 9 jours au cours de l'année 2014

sont présentés dans la figure 5.2. Ils font apparaître que :

- Les prévisions issues du système proposé se montrent plutôt fiables mais présentent un léger biais positif et une légère sur-dispersion.
- Le post-traitement météorologique permet de corriger un biais négatif important (visible sur l'histogramme de rang de gauche).
- Le post-traitement hydrologique d'EDF induit une sur-dispersion des prévisions.

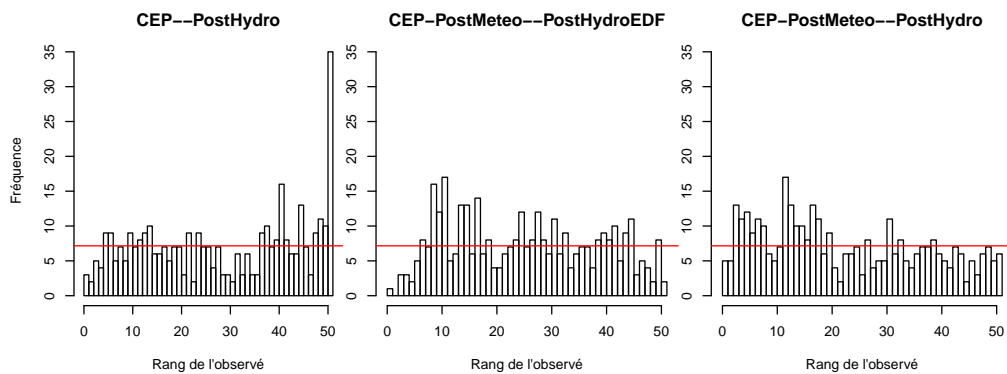


FIGURE 5.2 – Histogrammes de rang des débits cumulés sur 9 jours parmi des simulations selon les distribution prédictives correspondantes au cours de l'année 2014, à l'exutoire du bassin versant du Drac au Sautet.

Les CRPS relatifs aux prévisions de débits à chaque échéance de prévision et aux prévisions de débits cumulés sont illustrés sur la figure 5.3. Ils font apparaître que :

- L'application du post-traitement météorologique proposé conduit à un gain en performance (illustré par la différence entre les courbes bleue et violette).
- Le biais observé sans post-traitement météorologique, peu pénalisant échéance de prévision par échéance de prévision, l'est davantage lorsque l'on considère les débits cumulés, ce que montre l'inversion du classement des prévisions CEP--PostHydro et CEP-PostMeteo--PostHydroEDF de la figure de gauche à celle de droite.
- Le post-traitement hydrologique proposé permet, sur cet exemple, un gain en performance par rapport à la méthode automatique opérationnelle d'EDF jusqu'à un horizon de prévision de 7 jours.

Nous avons ainsi montré, sur un exemple, que les résultats du système de prévision proposé dans ce mémoire obtenus en pratique rejoignent les attentes théoriques formulées à son commencement.

CONCLUSION SUR LES PROPOSITIONS DE MÉTHODES DE POST-TRAITEMENT Les méthodes de post-traitement élaborées au cours de ce travail de thèse s'appuient sur des

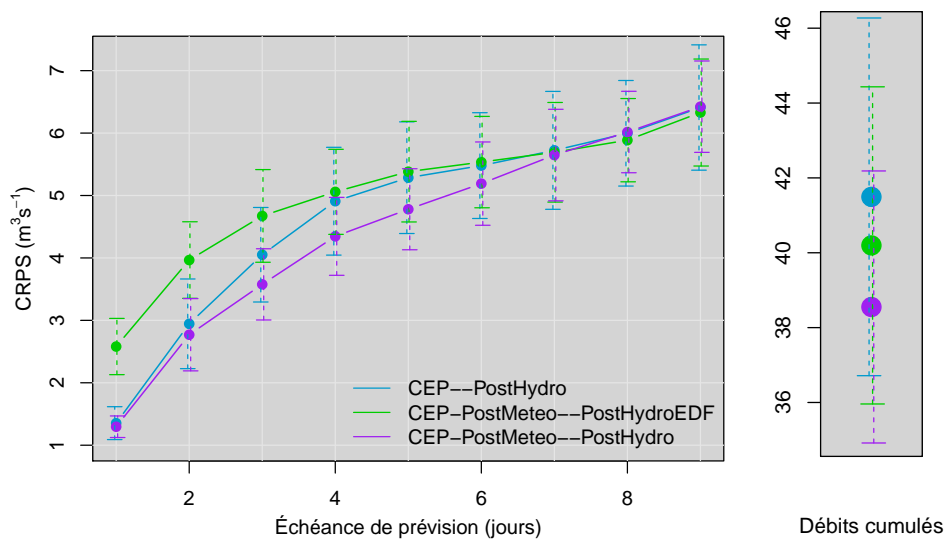


FIGURE 5.3 – CRPS relatifs aux prévisions de débits à chaque échéance de prévision et aux prévisions de débits cumulés sur 9 jours au cours de l'année 2014, à l'exutoire du bassin versant du Drac au Sautet.

constructions statistiques rigoureuses qui garantissent, si les hypothèses sur lesquelles elles reposent sont respectées, une fiabilité des prévisions qui en résultent.

Les post-traitements proposés ont de plus montré des performances intéressantes (selon le score CRPS) dans des situations de prévision variées proposées par EDF et par Hydro-Québec. Le score CRPS évaluant à la fois la fiabilité et la précision des prévisions, le gain en précision par l'intégration d'informations auxiliaires escompté doit avoir été obtenu.

Une corrélation significative entre le *Continuous Ranked Probability Skill Score* (CRPSS), défini dans le chapitre 4, et le gain net a été montrée par Perreault et al. (2017) dans le cas de l'exploitation de la centrale de Churchill Falls par Hydro-Québec. La figure 5.4 illustre ce résultat qui corrobore les observations de Zalachori (2013). Ainsi, le gain en CRPS obtenu avec les post-traitements que nous avons proposés pourra se traduire en un gain financier net pour l'entreprise s'ils sont utilisés pour la gestion des ouvrages hydro-électriques dans des contextes tels que ceux d'EDF et d'Hydro-Québec (décrits dans le chapitre introductif).

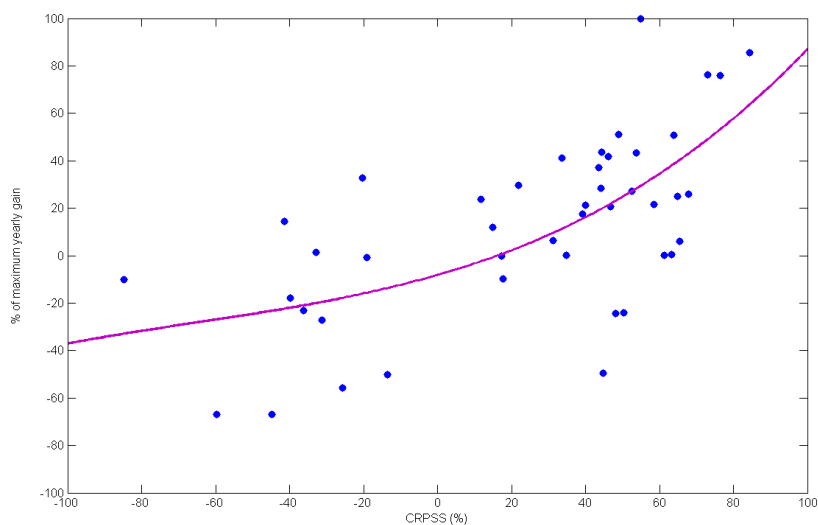


FIGURE 5.4 – Relation entre les gains financiers nets et les valeurs correspondantes du *Continuous Ranked Probability Skill Score* de prévisions probabilistes dans le cas de la centrale de Churchill Falls (Canada). Cette figure est adaptée de Perreault et al. (2017).

5.3 Confrontation entre idéal statistique et réalité opérationnelle ...

À Hydro-Québec, la méthode de post-traitement proposée pour les températures fait actuellement l'objet d'essais en vue d'une éventuelle mise en place opérationnelle. Des essais sur le post-traitement des précipitations suivront sous peu et la méthode de post-traitement hydrologique fera l'objet d'essais en 2017.

La confrontation entre idéal statistique et réalité opérationnelle peut toutefois faire apparaître des critères de performance non retranscrits par le CRPS, considéré comme critère de performance jusqu'ici.

FIABILITÉ Les hypothèses de modélisation nécessaires aux garanties de fiabilité des prévisions post-traitées ne sont pas toujours bien vérifiées, ce qui apparaît lors de confrontations aux réelles conditions de fonctionnement des systèmes de prévision. Pour que fiabilité du système de prévision il y ait, il faut qu'elle soit observable (avec des outils diagnostiques tel que l'histogramme de rang, présenté dans le chapitre 1) pour tout conditionnement de la variable à prévoir, comme souligné dans la section 1.3.1. Dans les méthodes que nous proposons, nous tâchons de la garantir, mais des conditionnements peuvent nous échapper.

Quand il s'agit de débits par exemple, la prévision des principaux événements de crue est d'une importance majeure. On vérifie donc, aux opérations, la fiabilité des prévisions produites en focalisant sur les événements de forts débits. Or, il apparaît

que des transformations utilisées pour normaliser les données de débits, telles que les transformations logarithmique ou Box-Cox, peuvent donner lieu, après transformation inverse, à des débits extrêmes irréalistes d'un point de vue hydrologique. En solution à ce problème, Hemri et al. (2015) proposent de supposer que la variable qui résulte de la transformation Box-Cox est gaussienne tronquée (à droite) plutôt que gaussienne. On ne trouve pas ce type d'aberration à l'issue de notre post-traitement hydrologique (sans doute parce qu'il tient compte de la dépendance temporelle des débits et des sorties du SPQ), mais une analyse saisonnière des débits extrêmes simulés, réalisée par EDF-DTG et illustrée sur la figure 5.5, montre un manque de fiabilité de nos prévisions. Une inférence saisonnière pourrait améliorer la performance de la méthode de post-traitement hydrologique que nous proposons. Dans le

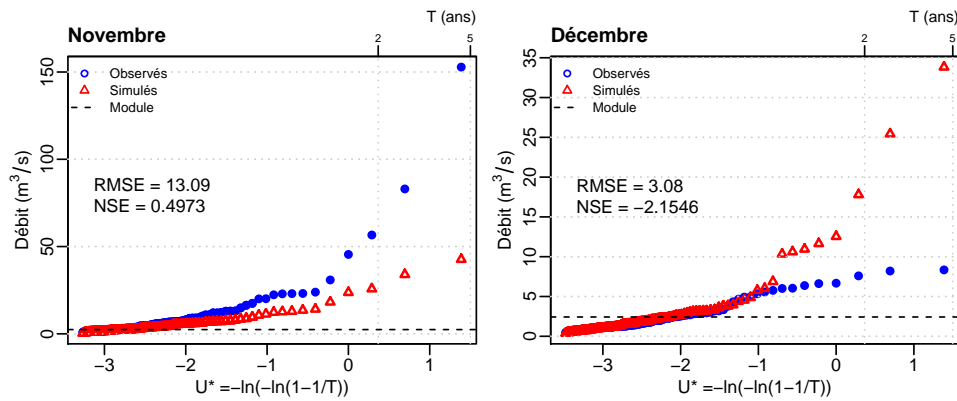


FIGURE 5.5 – Diagramme de Gumbel des débits extrêmes simulés à l'échéance 7 jours selon notre méthode de post-traitement et débits observés correspondants, pour les mois de novembre (à gauche) et de décembre (à droite) des années 2005 à 2008, sur le bassin versant de la Borne à Pont de Nicoulaud.

cas où seule la prévision des débits extrêmes serait visée, d'autres modèles (spécifiques aux extrêmes) doivent être développés. De tels modèles ont été développés pour le dimensionnement des évacuateurs de crue des barrages par EDF-DTG par exemple (Garavaglia, 2011).

L'exemple d'application considéré pour le post-traitement météorologique montre de plus que des instationnarités peuvent s'immiscer dans notre jeu de données et rendre inefficaces nos méthodes de post-traitement (par exemple un changement majeur de la méthode de production des prévisions d'ensemble météorologiques ou bien un hiver d'un froid exceptionnel). Cela met en évidence la pertinence de l'intervention d'une expertise humaine en aval de la chaîne de prévision, car elle est à même de détecter un tel changement et d'y adapter les prévisions malgré la rupture de l'hypothèse de stationnarité locale, indispensable à l'inférence statistique.

INCERTITUDES DE MESURE ET RÉOLUTION SPATIO-TEMPORELLE Comme souligné dans la section 1.2.1 du chapitre 1, les difficultés liées à l'incertitude de mesure aux stations et aux biais induits par les méthodes d'agrégation spatiale et temporelle ont été laissées de côté dans ce travail de thèse : nous avons choisi de chercher à prévoir les mesures (ou les agrégations) et non les "vraies" variables (qui, de fait, ne sont pas toujours observées).

Or, en pratique, ces incertitudes négligées dans notre approche posent parfois des problèmes majeurs :

- Comme l'illustre la figure 5.6, des débits négatifs (obtenus par bilan hydrique à partir de la hauteur d'eau au réservoir et des volumes turbinés) sont parfois à prévoir au Québec ; faut-il les ignorer ou bien intégrer cette possibilité de débits négatifs dans la modélisation ?
- Les défauts de résolution spatiale (des prévisions, des réseaux de mesures et des SPQ) posent particulièrement problème pour l'utilisation des prévisions de précipitations. Par exemple, les SPQ considérés dans ce travail ne prennent en intrants que les précipitations moyennées sur chaque bassin versant. Or, des précipitations localisées tout en amont d'un bassin versant ne se traduisent pas de la même façon en débit à l'exutoire que la même quantité de précipitations équirépartie sur ce même bassin versant.

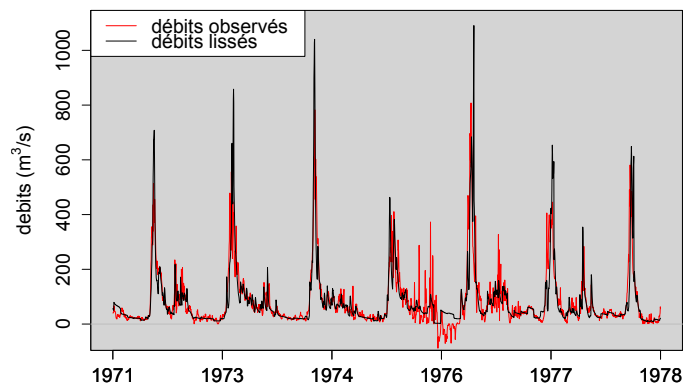


FIGURE 5.6 – Débits calculés par bilans hydriques ("observés") et débits utilisés dans cette thèse ("lissés") - exemple de l'historique hydrologique du bassin versant de Manic 2 de 1971 à 1978.

Il apparaît là encore que l'intervention des prévisionnistes s'avère parfois cruciale. Dans le dernier exemple donné ci-dessus par exemple, où des précipitations sont localisées tout en amont d'un bassin versant, il arrive que les prévisionnistes modifient les intrants du SPQ (en faisant partiellement passer les précipitations d'un jour pour celles du lendemain par exemple) afin de compenser le manque de résolution spatiale du SPQ.

Une façon d'améliorer les systèmes de prévision est donc d'améliorer la résolution (spatiale et temporelle) des SPQ. En poussant plus loin cette stratégie, on pourrait imaginer coupler simulateurs atmosphériques et SPQ afin de directement émettre des prévisions hydrologiques d'ensemble. Cela fait partie des objectifs de la Modélisation Environnementale Communautaire, développée par Environnement Canada depuis déjà une dizaine d'années (Pietroniro et al., 2007).

Par ailleurs, la disponibilité croissante de mesures de la couverture neigeuse des bassins versants par télédétection (par exemple par le biais du produit MODIS¹) ou bien de l'équivalent en eau de la neige laisse entrevoir d'autres sources de données pour l'affinement des prévisions hydrologiques.

EXPERTISE, COMMUNICATION ET AUTOMATISATION Il est d'autant plus important que les prévisionnistes s'impliquent dans la production des prévisions qu'ils ont souvent aussi pour rôle leur communication (et donc leur explication) aux preneurs de décisions. Communiquer les prévisions et leur caractère probabiliste de façon adéquate est en effet important pour qu'elles soient bien utilisées mais reste un défi (Ramos et al., 2010). L'automatisation des systèmes de prévision hydrologique doit donc être accompagnée d'une réflexion sur ses potentiels impacts sur le travail des experts et sur la façon dont les prévisions et leur caractère probabiliste seront communiqués (Pagano et al., 2016).

1. <http://modis-snow-ice.gsfc.nasa.gov/>

Bibliographie

- K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198, 2009.
- J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679, 1993.
- D. Allard. Modeling spatial and spatio-temporal non Gaussian processes. In *Advances and Challenges in Space-time Modelling of Natural Events*, pages 141–164. Springer, 2012.
- E. A. Anderson. *National Weather Service river forecast system: Snow accumulation and ablation model*, volume 17. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, 1973.
- J. L. Anderson. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7):1518–1530, 1996.
- V. Andreassian, S. Bergstrom, N. Chahinian, Q. Duan, Y. Gusev, I. Littlewood, T. Mathevet, C. Michel, A. Montanari, G. Moretti, et al. Catalogue of the models used in MOPEX 2004/2005. *IAHS publication*, 307:41, 2006.
- A. Azzalini. Further results on a class of distributions which includes the normal ones. *Statistica*, 46(2):199–208, 1986.
- C. Babusiaux. *Etude statistique de la loi des fuites*. PhD thesis, 1969.
- S. Ballay, L. Perreault, and B. Bobée. *Loi des fuites*. Number R479 in Rapport de recherche. INRS-Eau, Québec, Septembre 1996. URL <http://espace.inrs.ca/593/>.
- B. C. Bates and E. P. Campbell. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, 37(4):937–947, 2001.
- Z. Ben Bouallègue. Calibrated short-range ensemble precipitation forecasts using extended logistic regression with interaction terms. *Weather and Forecasting*, 28(2):515–524, 2013.

- J. M. Bernardo. Expected information as expected utility. *The Annals of Statistics*, pages 686–690, 1979.
- J. Bernier and D. Fandoux. Théorie du renouvellement. application à l'étude statistique des précipitations mensuelles. *Revue de statistique appliquée*, 18(2):75–87, 1970.
- V. J. Berrocal, A. E. Raftery, and T. Gneiting. Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Monthly Weather Review*, 135(4):1386–1402, 2007.
- K. Beven and A. Binley. The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6(3):279–298, 1992.
- M.-A. Boucher, L. Perreault, F. Anctil, and A.-C. Favre. Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. *Hydrological Processes*, 29(6):1141–1155, 2015.
- P. Bougeault, Z. Toth, C. Bishop, B. Brown, D. Burridge, D. H. Chen, B. Ebert, M. Fuentes, T. M. Hamill, K. Mylne, et al. The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, 91(8):1059–1072, 2010.
- G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, Ca, 1970.
- G. E. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.
- G. E. Box and G. M. Jenkins. *Time series analysis: forecasting and control, revised ed.* Holden-Day, 1976.
- L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.
- J. Bremnes. Improved calibration of precipitation forecasts using ensemble techniques. Part 2: Statistical calibration methods. met. Technical report, research report number 04, 2007.
- J. Bröcker. Reliability, sufficiency, and the decomposition of proper scores. *Quarterly Journal of the Royal Meteorological Society*, 135(643):1512–1519, 2009.
- J. Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Quarterly Journal of the Royal Meteorological Society*, 138(667):1611–1617, 2012.

- M. Broniatowski, G. Celeux, and J. Diebolt. Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste. *Data analysis and informatics*, 3: 359–373, 1983.
- R. Buizza. The value of probabilistic prediction. *Atmospheric Science Letters*, 9(2): 36–42, 2008.
- R. Buizza, M. Leutbecher, and L. Isaksen. Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 134(637):2051–2066, 2008.
- R. Burnash, V. Singh, et al. The NWS river forecast system-catchment modeling. *Computer models of watershed hydrology.*, pages 311–366, 1995.
- G. Celeux and J. Diebolt. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational statistics quarterly*, 2(1):73–82, 1985.
- S. Chatelain, L. Perreault, and S. Tardif. Vérification des prévisions hydrométéorologiques : expérience de simulation et application sur données réelles. Master's thesis, Université McGill, 2015.
- S. Chib. Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75(1):79–97, 1996.
- M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby. The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1):243–262, 2004.
- G. Coccia and E. Todini. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences*, 15(10):3253–3274, 2011.
- J. Collet, X. Épiard, and P. Coudray. Simulating hydraulic inflows using PCA and ARMAX. *The European Physical Journal-Special Topics*, 174(1):125–134, 2009.
- W. Collischonn, R. Haas, I. Andreolli, and C. E. M. Tucci. Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *Journal of Hydrology*, 305(1):87–98, 2005.
- P. Coulibaly, F. Anctil, and B. Bobée. Prévision hydrologique par réseaux de neurones artificiels: état de l'art. *Canadian Journal of civil engineering*, 26(3):293–304, 1999.
- A. R. Da Paz, W. Collischonn, C. Tucci, R. Clarke, D. Allasia, E. Boegh, H. Kunstmann, T. Wagener, A. Hall, L. Bastidas, et al. Data assimilation in a large-scale distributed hydrological model for mediumrange flow forecasts. *IAHS PUBLICATION*, 313:471, 2007.

- A. P. Dawid. The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77(379):605–610, 1982.
- A. P. Dawid. Present position and potential developments: Some personal views. statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, pages 278–292, 1984.
- B. de Finetti. Funzione caratteristica di un fenomeno aleatorio. 1931.
- R. C. D. de Paiva, D. C. Buarque, W. Collischonn, M.-P. Bonnet, F. Frappart, S. Calmant, and C. A. Bulhões Mendes. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research*, 49(3):1226–1243, 2013.
- C. M. DeChant and H. Moradkhani. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48(4), 2012.
- M. H. DeGroot and S. E. Fienberg. Assessing probability assessors: Calibration and refinement. Technical report, DTIC Document, 1981.
- M. H. DeGroot and S. E. Fienberg. The comparison and evaluation of forecasters. *The statistician*, 32(1/2):12–22, 1983.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- Q. Duan, N. K. Ajami, X. Gao, and S. Sorooshian. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5): 1371–1386, 2007.
- P. K. Dunn. Occurrence and quantity of precipitation can be modelled simultaneously. *International Journal of Climatology*, 24(10):1231–1239, 2004.
- K. Engeland, B. Renard, I. Steinsland, and S. Kolberg. Evaluation of statistical models for forecast errors from the HBV model. *Journal of hydrology*, 384(1): 142–155, 2010.
- G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- G. Evin, D. Kavetski, M. Thyer, and G. Kuczera. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research*, 49(7):4518–4524, 2013.

- G. Evin, M. Thyer, D. Kavetski, D. McInerney, and G. Kuczera. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3):2350–2375, 2014.
- F. M. Fan, W. Collischonn, A. Meller, and L. C. M. Botelho. Ensemble streamflow forecasting experiments in a tropical basin: The São Francisco river case study. *Journal of Hydrology*, 519:2906–2919, 2014.
- V. Fortin. Le modèle météo-apport HSAMI: historique, théorie et application. *Institut de recherche d'Hydro-Québec, Varennes*, 2000.
- V. Fortin, A.-c. Favre, and M. Said. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. *Quarterly Journal of the Royal Meteorological Society*, 132(617):1349–1369, 2006.
- E. M. Furrer, C. Jacques, and A.-C. Favre. Short term discharge prediction using a Markovian regime switching model. technical report, INRS-ETE, 2006.
- J. Gailhard. Outil PREDICTOR notice méthodologique. Note Technique Interne D4165/NT/2011-00121-A, EDF-DTG, 2011.
- J. Gailhard. Algorithme de recalage associé à MORDOR diagnostic et proposition d'améliorations. Note Technique Interne H-44200965-2014-000075, EDF-DTG, 2014.
- F. Garavaglia. *Méthode SCHADEX de prédétermination des crues extrêmes*. PhD thesis, Grenoble, 2011.
- R. Garçon. Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. bilan de l'année 1994-1995. *La Houille Blanche*, (5): 71–76, 1996.
- A. E. Gelfand and A. F. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.
- C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368, 2007.
- T. Gneiting. Editorial: probabilistic forecasting. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 319–321, 2008.
- T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

- T. Gneiting, A. E. Raftery, A. H. Westveld, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.
- T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.
- T. Gneiting, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, 17(2):211–235, 2008.
- N. J. Gordon, D. J. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F-Radar and Signal Processing*, volume 140, pages 107–113. IET, 1993.
- C. Guay. IREQ-2014 Modification de la modélisation des processus dans HSAMI : écoulements verticaux. Technical report, Institut de recherche d’Hydro-Québec, Varennes, QC, Canada, 2014.
- C. Guay and I. Chartier. IREQ-2016 Stratégies de calage pour la paramétrisation opérationnelle d’un modèle hydrologique : déploiement et analyse sur les bassins versants d’Hydro-Québec. Technical report, Institut de recherche d’Hydro-Québec, 2016.
- H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1):80–91, 2009.
- T. M. Hamill and S. J. Colucci. Verification of eta-rsm short-range ensemble forecasts. *Monthly Weather Review*, 125(6):1312–1327, 1997.
- T. M. Hamill, J. S. Whitaker, and S. L. Mullen. Reforecasts: An important dataset for improving weather predictions. *Bulletin of the American Meteorological Society*, 87(1):33, 2006.
- W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- S. Hemri, F. Fundel, and M. Zappa. Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. *Water Resources Research*, 49(10):6744–6755, 2013.
- S. Hemri, D. Lisniak, and B. Klein. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research*, 51(9):7436–7451, 2015.

- H. D. Herr and R. Krzysztofowicz. Generic probability distribution of rainfall in space: the bivariate model. *Journal of Hydrology*, 306(1):234–263, 2005.
- H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.
- E. Hewitt and L. J. Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955.
- D. Higdon, M. Kennedy, J. Cavendish, J. Cafeo, and R. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, 2004.
- B. Houdant. *Contribution à l'amélioration de la prévision hydrométéorologique opérationnelle. Pour l'usage des probabilités dans la communication entre acteurs*. PhD thesis, ENGREF (AgroParisTech), 2004.
- N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions, vol. 1-2*. New York: John Wiley & Sons, 1994.
- D. Kavetski, S. W. Franks, and G. Kuczera. Confronting input uncertainty in environmental modelling. *Calibration of watershed models*, pages 49–68, 2002.
- P. K. Kitanidis and R. L. Bras. Adaptive filtering through detection of isolated transient errors in rainfall-runoff models. *Water Resources Research*, 16(4):740–748, 1980.
- R. Krzysztofowicz. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9):2739–2750, 1999.
- R. Krzysztofowicz. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268(1):16–40, 2002.
- R. Krzysztofowicz and C. J. Maranzano. Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology*, 293(1):57–73, 2004.
- R. Krzysztofowicz and C. J. Maranzano. Bayesian processor of output for probabilistic quantitative precipitation forecasts. *Manuscript in review*, 2006.
- R. Krzysztofowicz and Z. Toth. Bayesian processor of ensemble (BPE): Concept and implementation. In *Slides presented at the 4th NCEP/NWS Ensemble User Workshop, Laurel, MD*, 2008.
- G. Kuczera. Improved parameter inference in catchment models: 1. evaluating parameter uncertainty. *Water Resources Research*, 19(5):1151–1162, 1983.
- G. Kuczera and E. Parent. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *Journal of Hydrology*, 211(1):69–85, 1998.

- G. Kuczera, D. Kavetski, S. Franks, and M. Thyer. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331(1):161–177, 2006.
- M. Li, Q. Wang, J. Bennett, and D. Robertson. A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts. *Hydrology and Earth System Sciences*, 19(1):1–15, 2015.
- T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 44(2):226–233, 1982.
- Z.-Q. Lu and L. M. Berliner. Markov switching time series models with application to a daily runoff series. *Water Resources Research*, 35(2):523–534, 1999.
- C. J. Maranzano and R. Krzysztofowicz. Bayesian reanalysis of the Challenger O-Ring data. *Risk analysis*, 28(4):1053–1067, 2008.
- T. Mathevet. Erreur empirique de modèle. Note Technique Interne D4165/NT/2010-00395-A, EDF-DTG, 2010.
- T. Mathevet, F. Garavaglia, R. Garçon, J. Gailhard, and E. Paquet. Operational hydrological ensemble forecasts in france. recent development of the french hydropower company (edf), taking into account rainfall and hydrological model uncertainties. In *EGU General Assembly Conference Abstracts*, volume 11, page 10248, 2009.
- N. Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999, 2006.
- J. W. Messner, G. J. Mayr, A. Zeileis, and D. S. Wilks. Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Monthly Weather Review*, 142(1):448–456, 2014.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- M. Minville. IREQ-2014-0104 Modification du modèle hydrologique HSAMI : Incorporation d’un modèle mixte degré-jour et bilan énergétique pour la modélisation du couvert nival. Technical report, Institut de recherche d’Hydro-Québec, Varennes, QC, Canada, 2014.
- M. Minville. IREQ-2106 Stratégies de calage pour la paramétrisation opérationnelle d’un modèle hydrologique : déploiement et analyse sur un échantillon de bassins versants. Technical report, Institut de recherche d’Hydro-Québec, 2016.

- A. Möller, A. Lenkoski, and T. L. Thorarinsdottir. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991, 2013. ISSN 1477-870X. doi: 10.1002/qj.2009. URL <http://dx.doi.org/10.1002/qj.2009>.
- A. Montanari and A. Brath. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 40(1), 2004.
- H. Moradkhani, K.-L. Hsu, H. Gupta, and S. Sorooshian. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5), 2005a.
- H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, 28(2):135–147, 2005b.
- M. Morawietz, C.-Y. Xu, L. Gottschalk, and L. M. Tallaksen. Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. *Journal of hydrology*, 407(1):58–72, 2011.
- J. Nash and J. Sutcliffe. River flow forecasting through conceptual models part I-A discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.
- P. Naveau, R. de Fondeville, D. Cooley, and H. Benveniste. Scores CRPS, inference and extremes. In *Séminaire Statistique au Sommet de Rochebrune*, 2014.
- R. B. Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.
- C. Obled, G. Bontron, and R. Garçon. Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. *Atmospheric research*, 63(3):303–324, 2002.
- T. Pagano, Q. Wang, P. Hapuarachchi, and D. Robertson. A dual-pass error-correction technique for forecasting streamflow. *Journal of hydrology*, 405(3): 367–381, 2011.
- T. C. Pagano, F. Pappenberger, A. W. Wood, M.-H. Ramos, A. Persson, and B. Anderson. Automation and human expertise in operational river forecasting. *Wiley Interdisciplinary Reviews: Water*, 3(5):692–705, 2016.
- F. Pappenberger, H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen. The monetary benefit of early flood warnings in europe. *Environmental Science & Policy*, 51:278–291, 2015.
- L. Perreault. Vérification de prévisions hydrologiques probabilistes - Version 2. Rapport scientifique IREQ-2013-0149, Institut de recherche d’Hydro-Québec, 2013.

- L. Perreault. Comparaison des PME d'Environnement Canada avec les prévisions expertisées de l'unité HOA. Présentation à la direction de la planification de la production. Technical report, Institut de recherche d'Hydro-Québec, 2014.
- L. Perreault. Évaluation des prévisions météorologiques d'ensemble : quelques expériences de vérification. Rapport scientifique IREQ-2015-XXXX, Institut de recherche d'Hydro-Québec, 2015.
- L. Perreault, E. Parent, J. Bernier, B. Bobee, and M. Slivitzky. Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment*, 14(4-5):243–261, 2000.
- L. Perreault, R. Garçon, and J. Gaudet. Analyse de séquences de variables aléatoires hydrologiques à l'aide de modèles de changement de régime exploitant des variables atmosphériques. *La Houille Blanche*, (6):111–123, 2007.
- L. Perreault, J. Gaudet, L. Delorme, and S. Chatelain. Verification of medium-to-long range hydrological probabilistic forecasts. In Qingyun Duan, Florian Pappenberger, Jutta Thielen, Andy Wood, Hannah L. Cloke, and John C. Schaake, editors, *Handbook of Hydrometeorological Ensemble Forecasting*. Springer-Verlag Berlin Heidelberg, 2017.
- T. C. Peterson, D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, et al. Homogeneity adjustments of in situ atmospheric climate data: a review. *International journal of climatology*, 18(13):1493–1517, 1998.
- F. Pianosi and L. Raso. Dynamic modeling of predictive uncertainty by regression on absolute errors. *Water Resources Research*, 48(3), 2012.
- A. Pietroniro, V. Fortin, N. Kouwen, C. Neal, R. Turcotte, B. Davison, D. Verseghy, E. Soulis, R. Caldwell, N. Evora, et al. Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*, 11(4):1279–1294, 2007.
- P. Pinson and J. Tastu. Discrimination ability of the Energy score. Technical report, Technical University of Denmark, 2013.
- A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5), 2005.
- M.-H. Ramos, T. Mathevet, J. Thielen, and F. Pappenberger. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications*, 17(2):223–235, 2010.

- B. Renard, D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47(11), 2011.
- D. S. Richardson. Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126(563): 649–667, 2000.
- D. Robertson, D. Shrestha, and Q. Wang. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences*, 17(9):3587–3603, 2013.
- E. Roulin and S. Vannitsem. Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. *Hydrological Processes*, 29(6):1434–1449, 2015.
- M. Roulston and L. A. Smith. Combining dynamical and statistical ensembles. *Tellus A*, 55(1):16–30, 2003.
- P. Salamon and L. Feyen. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, 376(3):428–442, 2009.
- L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.
- J. C. Schaake, T. M. Hamill, R. Buizza, and M. Clark. HEPEx: the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society*, 88(10):1541–1547, 2007.
- B. Schaefli, D. B. Talamba, and A. Musy. Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology*, 332(3):303–315, 2007.
- R. Schefzik. A similarity-based implementation of the schaaake shuffle. *arXiv preprint arXiv:1507.02079*, 2015.
- R. Schefzik, T. L. Thorarinsdottir, T. Gneiting, et al. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28(4):616–640, 2013.
- M. Scheuerer. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society*, 140(680):1086–1096, 2014.
- G. Schoups and J. A. Vrugt. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10), 2010.

- P. Schultz, H. Yuan, M. Charles, R. Krzysztofowicz, and Z. Toth. A continuous variable for the statistical processing of precipitation. 2008.
- D. Schwanenberg, F. M. Fan, S. Naumann, J. I. Kuwajima, R. A. Montero, and A. A. dos Reis. Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty. *Water Resources Management*, 29(5):1635–1651, 2015.
- G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.
- F. C. Schwappe. *Uncertain dynamic systems*. Prentice Hall, 1973.
- D.-J. Seo, H. Herr, and J. Schaake. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions*, 3(4):1987–2035, 2006.
- H. L. Shang and R. J. Hyndman. *rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data*, 2016. URL <https://CRAN.R-project.org/package=rainbow>. R package version 3.4.
- M. Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.
- J. M. L. Sloughter, A. E. Raftery, T. Gneiting, and C. Fraley. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135(9):3209–3220, 2007.
- D. P. Solomatine and D. L. Shrestha. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, 45(12), 2009.
- S. Sorooshian and J. A. Dracup. Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, 16(2):430–442, 1980.
- G. Stoltz. Agrégation séquentielle de prédicteurs: méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique. *Journal de la Société française de Statistique*, 151(2):66–106, 2010.
- M. Taillardat, O. Mestre, M. Zamo, and P. Naveau. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review*, (2016), 2016.
- O. Talagrand, R. Vautard, and B. Strauss. Evaluation of probabilistic prediction systems. In *Proc. ECMWF Workshop on Predictability*, volume 1, page 25, 1997.
- A. Thiboult, F. Anctil, and M.-A. Boucher. Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrology and Earth System Sciences*, 20(5):1809–1825, 2016.

- T. L. Thorarinsdottir, M. Scheuerer, and C. Heinz. Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *Journal of Computational and Graphical Statistics*, (just-accepted):00–00, 2014.
- M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45(12), 2009.
- E. Todini. A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, 6(2):123–137, 2008.
- J. Verkade and M. Werner. Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences Discussions*, 8, 2011, 2011.
- J. A. Vrugt and B. A. Robinson. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, 43(1), 2007.
- J. A. Vrugt, C. Ter Braak, C. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290, 2009.
- Q. Wang and D. Robertson. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, 47(2), 2011.
- Q. Wang, D. Robertson, and F. Chiew. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, 45(5), 2009.
- Q. Wang, D. L. Shrestha, D. Robertson, and P. Pokhrel. A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research*, 48(5), 2012.
- Q. Wang, J. C. Bennett, and D. E. Robertson. Error reduction and representation in stages (erris) in hydrological modelling for ensemble streamflow forecasting. *Hydrology and Earth System Sciences*, 20(9):3561, 2016.
- X. Wang and C. H. Bishop. Improvement of ensemble reliability with a new dressing kernel. *Quarterly Journal of the Royal Meteorological Society*, 131(607):965–986, 2005.
- A. Weerts, H. Winsemius, and J. Verkade. Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (england and wales). *Hydrology and Earth System Sciences*, 15,(1), 2011.

- D. S. Wilks. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, 16(3):361–368, 2009.
- I. Zalachori. *Prévisions hydrologiques d'ensemble: développements pour améliorer la qualité des prévisions et estimer leur utilité*. PhD thesis, AgroParisTech, 2013.
- I. Zalachori, M. Ramos, R. Garçon, T. Mathevet, and J. Gailhard. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science & Research*, 8:p–135, 2012.

Annexe

Article de revue soumis pour publication

A statistical outlook on hydrological forecasting systems with ensemble weather forecasts: gap or bridge between operational practices and probability theory?

Marie Courbariaux^{a,d}, Pierre Barbillon^a, Luc Perreault^b, Anne-Catherine Favre^c,
Éric Parent^a

^aUMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

^bInstitut de Recherche d'Hydro-Québec (IREQ)

1800 boul. Lionel Boulet, Varennes, Québec J3X 1S1

^cUniversité Grenoble Alpes (UGA), Grenoble INP, LTHE

CS 40700, 38058 Grenoble cedex 9

^dCorresponding author. Address: UMR MIA-Paris, AgroParisTech, 16 rue Claude Bernard,
75005, Paris, France; Email address: marie.courbariaux@agroparistech.fr

Abstract

This article proposes a general framework to understand how hydrological forecasting systems make probabilistic predictions, with emphasis on information and uncertainty conveyed by ensemble weather forecasts used as inputs to a rainfall/runoff simulator. Why standard statistical techniques do not satisfactorily solve this particular problem of hydrological probabilistic prediction? Keeping a statistical point of view, the discrepancies between the ideal world of statistical modeling and the challenges of the operational constraints are pointed out: (i) hydrologists and statisticians think differently about the use and misuse of a probabilistic model although both communities are concerned by the objectives of calibration and sharpness (ii) non-stationarity, lack of Gaussianity and high-dimensionality are more the rule than the exception (iii) meteorological and hydrological sources of uncertainties have to be distinguished. General purposes of the reviewed case studies are the management of an hydro-system, mostly for hydro-power generation and flood awareness systems. A unified view on how such operational systems re-employ, formally or less explicitly, the various items of the normal statistical toolbox to design their operational forecasting scheme is given. Finally, the benefits that would become visible through a better collaboration between forecasters and statisticians are demonstrated.

Keywords: statistical modeling, probabilistic forecasts, ensemble weather forecasts, hydrology, review, uncertainty propagation /rainfall-runoff model

1. Introduction

One of the practical purposes of hydrology is to make inflow forecasts for the future. Strategic decisions are taken daily on the basis of these forecasts, namely to prevent flooding damages and to avoid operating losses. Hydrological processes, models, and the variables associated with forecasting future hydrological states all contain a wide part of uncertainty. Therefore, hydrological forecasts should be probabilistic and expressed in the form of a probability distribution, or an ensemble of realizations from that predictive distribution, in order to properly represent the phenomenon and what we know about it. Besides, as illustrated in a large number of applications (see e.g. Katz and Murphy, 1997), and conceptually demonstrated by the mind game of Weather Roulette (Hagedorn and Smith, 2009), the use of probabilistic forecasts can save the user money. For such reasons, the forecasting teams of most hydro-power companies have enforced their move from deterministic to probabilistic hydro-meteorological forecasts. As an illustration of the impact of improving hydrological forecasts, Figure 1 shows a convincing argument for Hydro-Québec (HQ) to keep on enhancing their probabilistic forecasting system. Retrospectively, high positive financial returns would have been obtained in average if the Churchill Falls hydropower complex could have been managed during the last 43 years on the basis of a more evolved probabilistic seasonal forecasting system for inflows. This economic experiment was performed by considering the new hydrological seasonal forecasts as input to HQ's stochastic mid-term planning model designed for optimizing electricity production. This optimization model is designed to determine for each week of a given planning horizon (often, one to two years) an optimal set of decision rules for the use and storage of water in system reservoirs, while satisfying a known weekly demand in electricity and environmental constraints (Perreault et al. (2017)). The maximization of water value in reservoirs at the end of the planning horizon is the main component of the cost function to minimize.

Streamflow predictions benefit from the recent availability of ensemble weather forecasts (Sivillo et al., 1997). Although meteorological numerical codes remain computationally very demanding (Zwiefelhofer and Kreitz, 2001), the ever-increasing power of computer technology makes it now possible to launch earth model numerical codes (at lower resolutions when necessary) for different initial conditions and several combinations of physical sub-models or their regional variants (Buizza et al., 2005). Such runs give the end-user worthwhile information (Cloke and Pappenberger, 2009). The dispersion of ensemble members also provides an appraisal of the uncertainty limitations of the earth model's weather predictions, a major source of uncertainty which in turn affects the streamflow predictability for a water catchment.

What type of operations does a water management system forecaster perform to process the resulting possible scenarios of the future weather as seen by the computer simulator (the ensemble members) into targets of hydrological interest? Hydrological forecasting teams routinely issue a few days ahead best-guess estimates of the streamflows as well as the order of magnitude of the associated uncertainty: how do they consider the connection they make between water science and operational

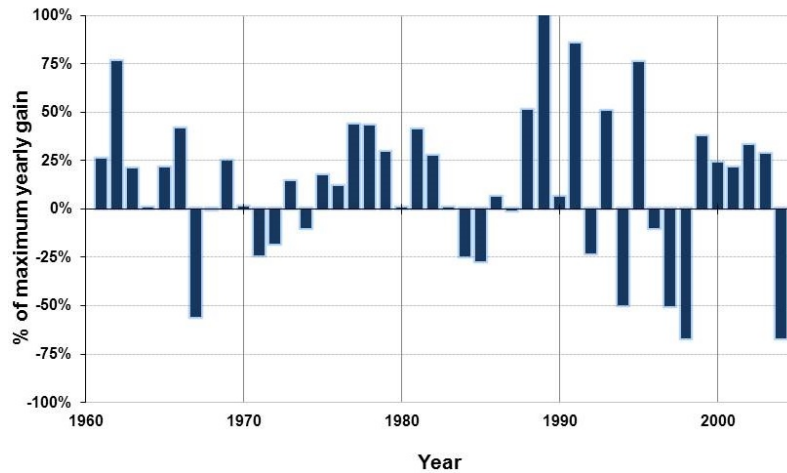


Figure 1: Net gain (in %) implied by the use of new seasonal inflow forecasting model based on atmospheric variables for Churchill Falls hydropower complex compared to the old mid-term forecasting approach based upon climatology (adapted from Perreault et al. (2017)).

prediction? Do they rely on some explicitly formalized model? Do they consider ensemble members as some sort of a finite sample from a non-formalized model? Do they simply work out an engineering recipe with good sense? These are intriguing questions for us, statisticians (this is the writing position that the authors will adopt throughout this article) since we are at the same time more inclined to be wary of the overoptimistic nature of forecasts issued by numerical codes (Buizza et al., 2005) but also fully appreciative of such an impressive task of processing very complex pieces of information.

In this paper, we analyze from a statistical point of view the challenges fulfilled by hydro-meteorological operational forecasting systems in order to provide probabilistic river flow forecasts, with emphasis on meteorological ensemble-based forecasts. Five hydrological forecasting systems serve as illustrative examples. They are detailed in Table 1. Three of those hydrological forecasting systems, namely Electricité de France’s (EDF), Hydro-Québec’s (HQ) and Companhia Energética de Minas Gerais’s (CEMIG) are designed for hydro-power production purposes. Information about those systems comes from the literature as well as from personal interactions with forecasting teams or engineering departments (at HQ and EDF). National Oceanic and Atmospheric Administration’s (NOAA) Hydrologic Ensemble Forecast Service (HEFS), Australian Government Bureau of Meteorology’s System for Continuous Hydrological Ensemble Forecasting (SCHEF) and CEMIG’s system are respectively described in Demargne et al. (2014); Zwiefelhofer and Kreitz (2001); Bennett et al. (2014). Our approach may thus be considered as a middle way between (i) a review of the statistical methods for streamflow forecasting such as Fortin et al. (1997) and (ii) a review of the operational streamflow forecasting systems such as Cloke and Pappenberger (2009) as well as (iii) a synthesis of daily echoes

in the many posts of the blog of the Hydrological Ensemble Prediction Experiment (HEPEX, Schaake et al., 2007b).

(short term) streamflow forecasting system	SCHEF, BoM (Australia) (in development) [Bennett et al., 2014]	HEFS, NOAA (USA) (in development) [Demargne et al., 2014]	EDF-DTG (operational) [Mathevet et al., 2009]	Hydro-Québec (operational)	CEMIG (Brazil) (in development) [Schwanenberg et al., 2015][Fan et al., 2014]	
Inputs	[deterministic forecasts] Deterministic precipitation forecasts from global ACCESS-G Numerical Weather Prediction (NWP) model	[deterministic/ensemble forecasts] Deterministic precipitation and temperature forecasts up to 7 days; ensemble mean forecasts from multiple NWP models for longer ranges	[ensemble forecasts] ECMWF-EPS precipitation and temperature forecasts	[deterministic forecasts] Deterministic precipitation and temperature forecasts from Environnement Canada	[ensemble forecasts] ECMWF-EPS precipitation forecasts	
Meteorological uncertainty processor	Normalization	Log-sinh transform [Wang et al., 2012]	Precipitation: Normal Quantile Transform (NQT) [Krzysztofowicz, 1997]	∅	∅	Deterministic bias correction
	Uncertainty processor	[Robertson et al., 2013] Modified version of the Bayesian joint probability (BJP) modelling approach by [Wang et al., 2009]	[bivariate normal model] Temperature: bivariate normal model; Precipitation: mixed-type bivariate meta-Gaussian model [Herr, 2005]	[truth+error model] Similar to kernel dressing methods and departing from [Obled et al., 2002]	[truth+error model] Normal kernel around temperature forecasts and Pearson III kernel around total precipitation forecasts	
	Multivariate structure	[empirical copula] Schaake shuffle [Clark et al., 2004]	[empirical copula] Schaake shuffle [Clark et al., 2004]	[empirical copula] Method similar to the ECC by [Scheffzik et al., 2013]	Quantile scenarios	
Rainfall-runoff simulator (RRS)	GR4H [Perrin et al., 2003]	Snow-17 [Anderson, 1973], SAC-SMA [Burnash, 1995], Unit HG	MORDOR [Garçon, 1996]	HSAMI [Fortin, 2000]	MGB-IPH [Collischonn et al. 2005] and mini-basins [Paiva et al., 2013]	
Hydrological uncertainty processor	Normalization	∅	NQT	log- transform	∅	∅
	Assimilation of last streamflows	[deterministic correction] Modified version of the dual-pass error correction model by [Pagano et al., 2011]	[manual data assimilation (DA)] Manual modifications of RRS state variables and parameters	[deterministic correction] Calimero: an error correction method specific to EDF-DTG	[automatic/manual DA] Manual DA on RRS state variables or automatic data assimilation	[automatic DA] Automatic DA on two RRS state variables [Paz et al., 2007]
	Uncertainty processor	Hydrological uncertainty is neglected	[ARX(1,1)] Hydrologic Ensemble Postprocessor (EnsPost) [Seo et al., 2006]	[normal model] Unformal log-normal linear model with heteroschedastic variance, inspired by [Montanari and Brath, 2004][Schaeffli et al., 2007]	[ensemble forecasts] Perturbations of RRS state variables	∅
	Multivariate structure	∅		[empirical copula] Method similar to the ECC by [Scheffzik et al., 2013]	∅	∅

Table 1: An outlook on miscellaneous hydrological forecasting systems. The main information sources are bracketed. Details about meteorological uncertainty processors (lines 2, 3 and 4 of the table) can be found Section 7.1. Details about hydrological uncertainty processors (lines 6 to 9 of the table) can be found Section 7.2. Some Rainfall-runoff simulators (line 5) are described Appendix A.

In Sections 2 and 3, we point out that statisticians and operational forecasters differ through their faith on probabilistic models and discuss why many practitioners might believe that a full statistical forecasting system remains out of reach:

essentially, non-stationarity and seasonality blur the ideal probabilistic model-based framework. Section 4 recalls the standard statistical toolbox of multivariate normal models to deal with the forecasting challenges and how to retrieve marginal normality while Section 5 focuses on how to take into account alternative multivariate structures. Section 6 distinguishes meteorological and hydrological sources of uncertainties: both statisticians and operational teams agree on treating them separately and reassembling them afterwards. In the case studies from various companies, practitioners do try to describe their main sources of uncertainty, rather explicitly for NOAA or less formally for EDF. We give a unified view on how they re-employ the various items of the previous statistical toolbox to design their operational forecasting scheme in Section 7. Sections 7.1 and 7.2 are respectively devoted to meteorological and hydrological uncertainties as met in operational forecasting systems. Section 8 discusses how they combine them into an operational probabilistic forecast. Section 9 provides a brief summary and some unexplored perspectives.

2. To bid or not to bid for a fully explicit statistical model \mathcal{H}

We define in Table 2 the notations used throughout the paper. We adopt here the bracket notation for probability distribution as advocated in Gelfand and Smith (1990). Namely, once a probabilistic framework has been assumed and specified for the couple of random variables X and Y , $[Y = y|X = x]$ stands for the conditional probability density function (pdf) of Y in y , given the event $\{X = x\}$, if Y is continuous, or its mass probability if Y is discrete.

Y	generic notation for the quantities of interest to be predicted
X	generic notation for the forecasting inputs
Q	variable referring to (daily) observed streamflow
A	variable referring to (daily) (ensemble) weather forecasts
P	variable referring to (daily) observed weather
$\mathcal{M} : x \mapsto \mathcal{M}(x)$	rainfall-runoff simulator (RRS) transformation
V	variable referring to (daily) watershed states in the RRS
$\mathcal{M}(P)$	rainfall-runoff simulator outputs from observed weather

Table 2: Notations used in this paper

From a statistician’s viewpoint, probabilistic forecasting consists merely in providing the probability distribution of the predictand Y (such as the streamflows Q) conditional onto predictors X (*e.g.* the weather anticipations A provided by the meteorological ensemble forecast members). Hence, the statistician’s graal is the *design* of a joint pdf $[Y, X|\mathcal{H}]$, with \mathcal{H} denoting the *statistical model*, *i.e.* all the necessary hypotheses to cast a coherent probabilistic structure onto the variables of interest. After inference on past records, the conditional pdf $[Y|X, \mathcal{H}]$ is derived following the rules of probability theory and then used for prediction.

The hydrologist’s quest is similar in essence but overlooks the mathematical existence of any statistical model \mathcal{H} . The practitioner’s operational behavior can be

understood as an engineering recipe to map any piece of information $X = x$ into a function $F^{(x)}(\cdot)$ for the predictand Y , this function having all the specific properties of a predictive cumulative distribution (monotonously increasing between 0 and 1). The mapping $X \mapsto F^{(X)}$ will be dubbed as a *probabilistic forecasting scheme* in the rest of the paper. For example, suppose the predictand Y is the temperature and the predictor X is the ensemble of members for two days ahead temperature forecasts; the following probabilistic forecasting scheme is a good example of a common sense engineering recipe: first, get rid of the members' largest and smallest values, then compute the mean and variance of the remaining members and finally adopt as a probabilistic forecasting scheme the normal distribution centered on the robustified mean with variance inflated by 10%.

As a simpler categorical example, consider the situation where the predictand Y is the binary variable "rain/ no rain" at a given location and the predictors X are 50 meteorological members of precipitation provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), h days ahead. The ratio π_h of members predicting rain h days ahead appears as a rather natural engineering recipe mapping predictors to a probability distribution for binary events (*i.e.* a number π between 0 and 1) ; by construction such a probabilistic forecasting scheme, $X \mapsto \pi$, takes values within $K = (50 + 1)$ categories . There are as many probabilistic forecasting schemes as forecasters: one could for instance pick $0 < a < b$ and makes predictions based on the adjusted ratio $\frac{50 \times \pi_h + a}{50 + b}$.

2.1. An actual conditional distribution would theoretically match the operational objectives of reliability and sharpness.

Once the mathematical apparatus \mathcal{H} is set (no matter its detailed formulation, delayed until Section 4) the main forecasting virtues will naturally stem *per se* from the probabilistic framework, *i.e.* in the ideal world where the quantities X and Y would comply to the statements assumed in \mathcal{H} . However, the desired operational properties of a probabilistic forecasting scheme are formulated by practitioners without reference to any probabilistic framework with requirements as follows:

- *reliability* (also termed *calibration*). Roughly speaking, reliability expresses the agreement between the distribution of a forecaster's predictions $F^{(X)}(\cdot)$ and the actual observed relative frequency of the predictand Y . For the previous probabilistic scheme based on the ratio π of members predicting rain $h = 4$ days ahead during summer on the Vouglan watershed, Fig 2 demonstrates that the probabilistic forecasting scheme we opted for is *not* reliable. The statistical agreement has been practically checked on a large dataset (the 2005-2008 summer periods): the ordinate axis of Fig 2 gives the long-run relative frequency of rain among the days when the forecast π belonged to category k which should –but does not at all– match the corresponding announced value π_k (in abscissa), even considering sampling variability (not represented here). Interestingly, a reliable forecasting scheme can always be written as a

conditional probability and any conditional probability of Y is a reliable (or calibrated) forecast (Bröcker, 2008).

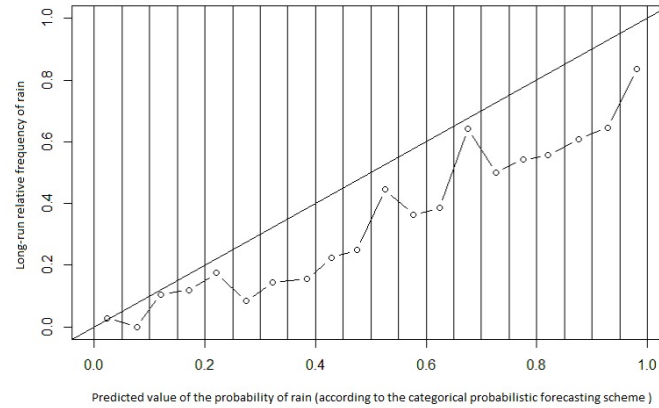


Figure 2: Long-run relative frequency of rain versus binned forecast π (ratio of 50 ECMWF members predicting rain) for $h = 4$ days ahead on the Ain watershed at Vouglans (France) during summer periods from 2005 to 2008. The first diagonal visualizes reliability agreement: the probabilistic forecasting system is not calibrated.

- *sharpness* (sometimes termed as *precision* or *refinement*). Sharpness refers to how spread out or how sharp a forecaster's predictions are. Fig 3 gives four examples ($h = 1, 3, 5, 7$) of forecasting systems predicting rain/no-rain based on the ratio π_h of "rainy" members on the ordinate axis versus time on the abscissa; the dots are rain or no rain events, filled dots correspond to observations not predicted by the majority of members. As the predicting horizon h moves away from the day of the event, sharpness keeps on degrading as shown by the different shapes of the various π_h 's: the behavior of π_h tends to deviate less from the mean value of Y , thus being less and less informative (with regard to a non risky mean predicting attitude) as the horizon step h increases. Given a probabilistic framework \mathcal{H} and a joint distribution $[X, Y|\mathcal{H}]$, statistical sufficiency proves that any deterministic transformation of X to predict Y (indeed any probabilistic forecasting scheme F^X) must be less refined than the conditional distribution $[Y|X, \mathcal{H}]$.
- *optimum scoring rewards*. A scoring function $S(F, y)$ is to be interpreted literally as how well a probabilistic forecasting scheme F scores when the quantity of interest Y takes the value y . Strictly proper scoring rules assess a quantified performance of a forecaster: they can be decomposed into additive components accounting respectively for reliability and sharpness (Gneiting and Raftery, 2007). Not surprisingly, *given* \mathcal{H} and the knowledge of X , $F^*(y) =$

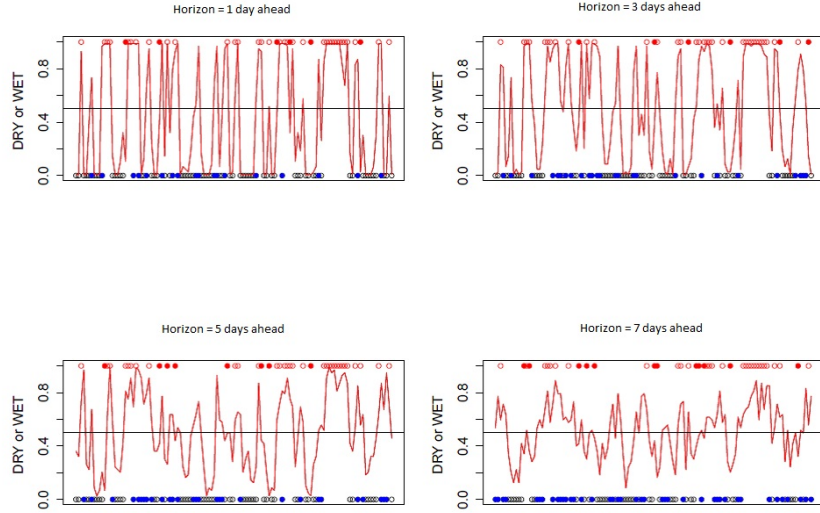


Figure 3: Observed events of rain/no rain versus daily forecasts π_h (ratio of 50 ECMWF members predicting rain for $h = 1, 3, 5, 7$ days ahead) on the Ain watershed at Vouglans (France) during summer days in 2005. The filled dots indicate false alarms and missed events (with regards to the 50% ratio corresponding to the majority rule)

$[Y = y|X, \mathcal{H}]$ automatically appears as the unique minimizer of the expectation with regards to Y of any *strictly proper scoring rule* $S(F, y)$ (this property served to De Finetti (1937) as a definite argument for probabilistic coherence and to Savage (1971) as a proof for honest probability elicitation).

To sum it up, both statisticians and practitioners do agree that reliability and sharpness form the cornerstones for judging of a forecasting system: at EDF, at HQ and in NOAA's Ensemble Verification Service (EVS), verification metrics include some reliability checks (Probability Integral Transform Histograms at HQ, Reliability diagrams at EDF and in the EVS for example) and strictly proper scoring rules such as the Continuous Ranked Probability Score (CRPS) (Jolliffe and Stephenson, 2012). Setting $Y = Q$ and $X = A$ for applications in hydrology, provided one has set a satisfactorily probabilistic model for $[Q, A|\mathcal{H}]$, the hydrological forecasting scheme obtained by the conditional distribution $[Q|A, \mathcal{H}]$, is:

- well calibrated,
- more refined than any other forecasting scheme,
- dominating the scoring performance of any other forecasting scheme, whatever the strictly proper scoring rule adopted.

Yet, although “*there is nothing more practical than a good theory*” as wrote by Lewin (1952, p.169), designing a *good* statistical model \mathcal{H} to correctly encompass the behaviors of Q and A , and working out its statistical inference do not quite appear as very appealing tasks to most practitioners: their reluctance becomes understandable when considering the challenging multidimensional nature of Q and A (both in the number of variables to deal with and their time-serial dimensions) . Conversely to statistical model-makers, forecasters often prefer to consistently improve their *ad hoc* probabilistic forecasting scheme $F^{(A)}$ to predict Q within a straightforward trial-and-error verification procedure in the operational context than bothering about ways to rebut, question or design the global underlying rationale \mathcal{H} upon which a reliable and sharp probabilistic forecasting scheme $[Q|A, \mathcal{H}]$ could be straightfully derived and (re)built.

2.2. In operational contexts, statistical models may provide additional advantages

In addition to sharpness and reliability theoretical warranties, explicit statistical modeling also comes with operational advantages:

- The underlying hypotheses are not swept under the carpet and the statements assumed in \mathcal{H} are honestly brought forward for discussion. Potential improvements can then be attempted by relaxing some of them that appear unrealistic. For example, in Section 6, one of the hypotheses allowing meteorological and hydrological sub-problems to be treated independently speculates that meteorological measurements should no longer be informative relatively to future streamflows once the RRS’s outputs are known. If one does not agree with such a conditional independence hypothesis, one can, for instance, increase the number of conditioning variables and develop alternative models based on RRS’s state variables to forecast streamflows (in addition to RRS’s outputs).
- A model based system can be adapted to a new forecasting situation with few cost. On the contrary, as EDF’s forecasting system is not presently based on some explicit statistical theory, each change in the system currently implies to manually calibrate parameters and formula, watershed by watershed.
- A model based system can provide on request a full probabilistic judgment for any quantity of interest. Most likely value, simulations, quantiles and probability of over-passing a threshold are intrinsically borne within a whole coherent probabilistic framework. Conversely, operational forecasts often restrictively provide experts and end-users with a sample of a limited number of scenarios illustrating the stochasticity of the predicted process. Such systems usually cannot easily extend the number of issued simulations. For instance, simulations of the future streamflows from the present EDF forecasting system are constrained by the size of the weather ensemble, while they are limited by the number of observation years for HQ’s forecasting system and HEFS. This is due to the use of empirical copulas [Ensemble Copula Coupling (ECC) at

EDF (Scheffzik et al., 2013) and Schaake Shuffle at HQ and in HEFS (Clark et al., 2004)]. Section 5 provides more details about those techniques.

3. Stationarity: the challenge for statistical learning.

However, the Beauty of probabilistic modeling should not hide the Beast of non-stationarity that impairs statistical learning. “*Loosely speaking, a stationary process is one whose statistical properties do not change over time*” (Nason, 2006). Statistical inference requires stationarity of the observed process: as a learning set, the past data must be statistically representative of the future behavior of the phenomenon to be predicted. Alas, in hydro-meteorological contexts, non-stationarity is more the rule than the exception. Non-stationarities can arise from:

- Changes induced by human actions. For instance, changes in the behavior of the catchment (due land-use development such as dike construction) or changes in the way ensemble weather forecasts are produced may have occurred. For instance the Canadian Meteorological Center (CMC) changes (listed on Environment Canada website http://collaboration.cmc.ec.gc.ca/cmc/cmci/product_guide/docs/changes_e.html) include three major upgrades of its Global Ensemble Prediction System (GEPS) over the last three years.
- Abrupt changes or slowly varying trends due to climate variability. For example, Perreault et al. (2000), Evin et al. (2011) and Jandhyala et al. (2009) note an overall abrupt decrease in average streamflow for rivers situated in the Northern Québec Labrador region in the early eighties, while Guay et al. (2015) anticipate a monotone increase in the mean annual streamflow at the 2050 horizon for the whole Québec province due to climate change.
- Periodic changes such as seasonality. Figure 4 (top panel) illustrates the daily streamflows in the Baskatong watershed (province of Québec) from 1950 to 2013, where the hydrological regime is characterized by annual snow melt floods. Figure 4 (bottom panel) shows the daily streamflow of the Ain watershed at Vouglans (France) from 1953 to 2008, where seasonality is less marked, with high flows depending more of rain than of snow. Seasonality thus requires specific *ad hoc* statistical treatments.

A rather sophisticated manner to recover stationarity is to identify and model the mechanism of non-stationarity (with Fourier series for seasonality for example) and to take it off from the data so as to work with stationarized residuals.

The simplest way to take into account non-stationarity might be to confine oneself to assuming some sort of local stationarity, i.e. to infer model parameters through a narrow sliding window or for each homogeneous temporal section, if known, or day of the year by day of the year (such as in NOAA’s Meteorological Ensemble Forecast Processor for example).

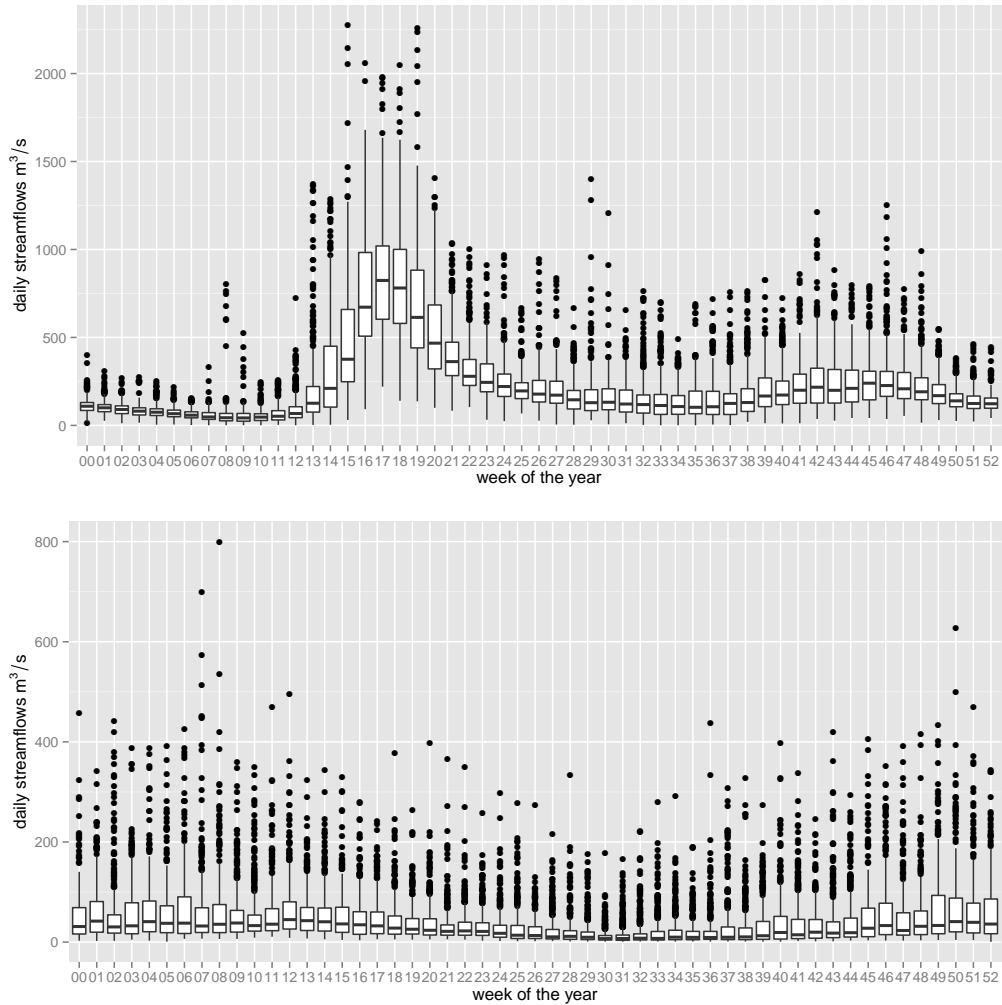


Figure 4: Boxplots of daily streamflows of the Baskatong watershed (province of Québec, top figure) and of the Ain watershed at Vouglans (France, bottom figure) from 1953 to 2008 by week of the year.

All solutions, and particularly sliding windows and homogeneous sections for parameter learning, require models with few parameters to stay robust when predicting. The overall Champion for parsimony in the statistical toolbox is the Normal model. Section 4 depicts how and why the multivariate normal model can be involved to achieve probabilistic forecasting.

4. The multivariate normal model: statistical modeling made easy?

The normal model is the reference model for statisticians. Once setting $\mathcal{H} = \{\text{all variables live in a normal world}\}$, their behavior is parsimoniously parameterized by a mean vector and a covariance matrix: a normal model of dimension d is only specified by $\frac{d(d+3)}{2}$ parameters. In addition, a multi-purpose wide range of sparse statistical structures have been developed based on the normal setting: Principal Component Analysis, regression, Anova and mixed linear models, ARIMA and PARMA time series, Gaussian Markov random fields and so on. The corresponding inference algorithms have been implemented within many **R** or **Matlab** functions and have been intensively tested on various case studies: elaborated routines dedicated to multivariate normal data are nowadays freely available to the hydrological community.

In forecasting problems, one is interested in a potentially multidimensional variable Y (the predictand, i.e. one-to-one transformation of the quantity of interest to predict such as the log of the streamflows on many sites or for many lead times) and is provided with some other variable X (termed predictors, generally issued from some meteorological source of information). Section 4.1 recalls the multivariate normal structure $[X, Y|\mathcal{H}]$ and points out the linear and homoscedastic features of its conditional pdf $[Y|X, \mathcal{H}]$ when achieving probabilistic forecasting.

Of course, it would be very remarkable if the real world would be adequately represented by the oversimple model $\mathcal{H} = \{\text{all variables live in a normal world}\}$. Thomson (1994) claimed that “*Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates*”. Temperature can sometimes be acceptably considered as normally distributed, but precipitation data are clearly not. In practice, forecasting systems make recourse to various working ad-hockeries to retrieve Gaussianity that are described in Section 4.2. When dealing with meteorological data, for instance the vector of temperatures and precipitations, the strength of their correlation does change with their levels. We review existing approaches to deal with this disrupting behavior for a brave normal model in Section 5.

4.1. Advantages and limits of the normal model

4.1.1. Modeling

Keeping in mind the block decomposition between X and Y , it is convenient to further decompose the mean and variance matrix tuning parameters into $\mu_X, \mu_Y, \Sigma_{XX}, \Sigma_{XY}$ and Σ_{YY} , so that:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^t & \Sigma_{YY} \end{pmatrix} \right) \quad (1)$$

If Y is the quantity of interest, μ_Y and Σ_{YY} characterize the *hydrological* (when Y stands for the streamflows) or the *climatological* (when Y stands for a meteorological variable) (normal) regime.

Deriving $[Y|X, \mathcal{H}]$ is straightforward:

$$(Y|X = x) \sim \mathcal{N}(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}) \quad (2)$$

Since components X and Y can stem from any subsetting of the model, the salient feature of the normal pdf is that any conditional pdf expressed in Eq. (2) is a *regression*:

1. the mean $\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X)$ is a *linear* form of the predictor X ,
2. the conditional variance $\Sigma_{Y|X} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$ does *not* depend on the predictor X (*homoscedasticity*),
3. as expected, $Y|X$ remains normally distributed with the conditional variance being smaller than the marginal (unconditional) variance Σ_{YY} . The strength of this reduction is ruled by the covariance between Y and X .

In an operational context, the first and third properties often seem acceptable, at least in a restrictive range of the variables around a working point. Specifying the hydrological regime or the climatology by a mean and a variance is also a current practice. The second property often questions the realism of the normal model: can one accept, at least as a first approximation, that the level x of (any) predictor X is irrelevant to the range of the possible additive error that will perturb the expected linear mean of (any) predictand Y given $X = x$?

4.1.2. Inference

Performing inference either directly on the conditional pdf (Eq. (2)) or on the joint model (Eq. (1)), and then derive the targeted model $[Y|X, \mathcal{H}]$ for probabilistic forecasting may give rather different results, depending on the structure of the available database for X and Y .

It is tempting to directly fit a probabilistic forecasting scheme $F^{(X)}$, adopting the same normal form as the conditional pdf (Eq. (2)) and write $F^{(X)}(\cdot)$ under the following regression like representation:

$$(Y|X = x) = \hat{\alpha} + \hat{\beta}x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \hat{\Sigma}_{Y|X}) \quad (3)$$

This direct and intuitive approach underlies the regression-like methods such as Ensemble Model Output Statistics (EMOS) (Gneiting et al., 2005), Extended Logistic Regression (ELR) (Wilks, 2009) and hydrological uncertainty modeling in EDF's system. Why in any case bother with Eq. (1)? The point is that only the joint records of X and Y are used when fitting Eq. (3). As a consequence, the estimates of $\hat{\mu}_Y = \hat{\alpha} + \hat{\beta}\hat{\mu}_X$ and $\hat{\Sigma}_{YY} = \hat{\Sigma}_{Y|X} + \hat{\beta}\hat{\Sigma}_{XX}\hat{\beta}'$ can be rather far away from their true values μ_Y and Σ_{YY} when obtained by simply running the fitted probabilistic forecasting scheme since its previous blunt fitting will not have taken into account all the available data of the predictand Y (namely what we will call the historical records of Y , *i.e.* the Y 's without X). This is particularly relevant when dealing with ensemble weather forecasts. In fact, new ensemble weather forecasts are rarely

accompanied by substantial re-forecasts. Thus the series of predictands data Y is often much longer than the series of their predictors X (the meteorological forecasts). In such a situation, a lot of historical data for the quantity of interest Y is unduly ignored. As an adverse consequence, the probabilistic forecasting scheme expressed in Eq. (3) can be poorly calibrated because it may not easily recover the full statistical features μ_Y and Σ_{YY} .

Following the seminal ideas of Krzysztofowicz (2002), the other avenue of thought begins with extracting valuable information from the historical records of Y , which ensures that the probabilistic forecasting scheme will be well calibrated. Two models are derived, a marginal model for Y and a conditional one for $(X|Y)$:

$$Y \sim \mathcal{N}(\mu_Y, \Sigma_{YY}) \quad (4)$$

$$(X|Y = y) = \psi + \gamma y + \eta, \quad \eta \sim \mathcal{N}(0, \Sigma_{X|Y}) \quad (5)$$

Bayes rule allows to obtain the targeted predictive distribution expressed in Eq. (2) from Eq. (4) and Eq. (5), where

$$\Sigma_{Y|X}^{-1} = \gamma' \Sigma_{X|Y}^{-1} \gamma + \Sigma_{YY}^{-1}$$

$$\mathbb{E}_{Y|X=x} = \Sigma_{Y|X} \left\{ \gamma' \Sigma_{X|Y}^{-1} (x - \psi) + \Sigma_{YY}^{-1} \mu_Y \right\}$$

The above normal conjugate formula are to be interpreted in terms of a Bayesian processor of information. The prior knowledge of Y (i.e. its overall behavior depicted in the historical records and summarized by $\hat{\mu}_Y$ and $\hat{\Sigma}_{YY}$) is updated through the likelihood (characterized by $\hat{\psi}$, $\hat{\gamma}$ and $\hat{\Sigma}_{X|Y}$) fed by the predictor X that gives information with regards to the present situation. Bayesian multivariate normal modeling underlies, for instance, HEFS's and SCHEF's meteorological uncertainty processors as well as Krzysztofowicz's Bayesian system for river stage forecasting (Krzysztofowicz, 2002).

Hydrological blunt "common sense" would surely favor working with structures in the vein of Eq. (3): the predictors X become available before the predictand Y does, they often "explain" it since there is certainly a causal link in the case when Y is the streamflow and X some weather supply of the rain. Such engineering good sense to interpret a regression is misleading: probability relies on correlation, not causality! Going one step further, it may even be easier to give rise to novel suggestions from Eqs. (4) and (5) than from Eq. (3): different hypotheses of partial stationarity to pick adapted training periods as well as different sparse stochastic models may be fruitfully chosen for marginal models and conditional ones. One could for instance consider

- i) an ARX conditional dependency between streamflows Y and weather measurements X fitted on a short training window to get $[X|Y]$ and
- ii) a long-term seasonal stationary PARMA model fitted on many years for streamflows Y , as a marginal behavior $[Y]$.

4.2. Retrieving marginal normality

In the hydro-meteorological context, scientists have to tackle obviously non-Gaussian variables, such as rainfalls and streamflows. Figure 5 (left) illustrates the deviation from normality of the Basketong daily streamflows from 1950 to 2013 while the right panel shows the normalized data.

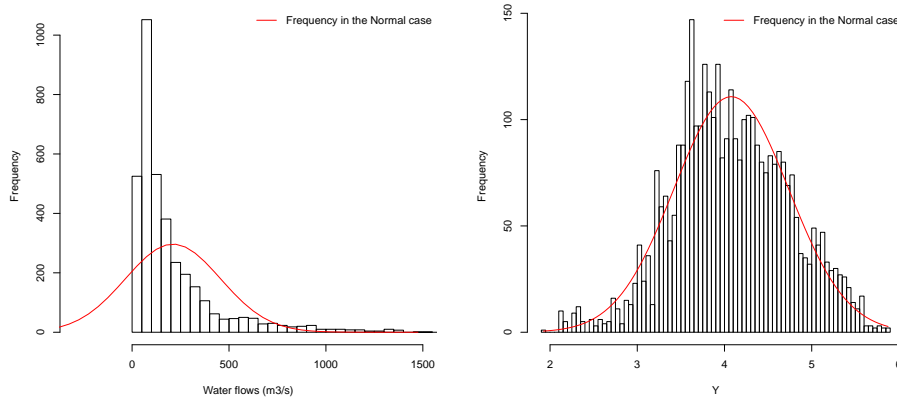


Figure 5: Histogram of the daily streamflows of the Basketong (province of Québec) from 1950 to 2013 (left) and histogram after normalization through a Box-Cox transformation (right). The lines indicate the normal cdf.

The Normal Quantile Transform (NQT) (Krzysztofowicz, 1997) is certainly the most direct technique to overcome non-normality. It works as follows: (i) Apply its own (empirical) cumulative distribution function G to the data x (ii) The resulting (uniformly distributed) data $G(x)$ is transformed by the inverse standard normal distribution function Φ^{-1} into standard normal-like data $\Phi^{-1}\{G(x)\}$. It is for instance in use in both HEFS's meteorological and hydrological processors (see also Krzysztofowicz (2002) and Todini (2008)).

Alternative normalization methods include the log-sinh transform (Wang et al., 2012) expressed as $x \rightarrow \frac{\log\{\sinh(a+bx)\}}{b}$, where a and b are parameters and $\sinh : x \mapsto \frac{e^x - e^{-x}}{2}$. It is for instance in use in SCHEF's meteorological uncertainty processor.

Another common normalization method resorts to the logarithmic transform: $x \mapsto \log(x)$ (see for instance EDF methods (Mathevet, 2010) and (Collet et al., 2009)). The logarithmic transform is a special case of the Box-Cox transform (Box and Cox, 1964), applied in Duan et al. (2007) and Hemri et al. (2015) and expressed as:

$$\begin{cases} x \mapsto \log(x) & \text{if } \lambda = 0 \\ x \mapsto \frac{x^\lambda - 1}{\lambda} & \text{otherwise} \end{cases}$$

The result of Box-Cox normalization on the same data set as Figure 5 (left) is illustrated in Figure 5 (right). Note that the transformed data is not exactly normal and may result into unlikely (extreme in this specific case) events after back-transformation, as also highlighted in Roulin and Vannitsem (2015).

To avoid miscalibration, normalization parameters should be estimated jointly with the other forecasting model parameters. This is for instance the case in SCHEF's meteorological uncertainty processor (inspired by Wang et al. (2009)).

Retrieving marginal normality through a one-to-one transform often brings a good behavior with regards to variance homogeneity (marginal and conditional ones), a necessary condition to catch up with the convenience of the normal multivariate framework. Some very specific methods however, dealing with zero inflated data of precipitations, avoid recourse to pre-normalization (see, as examples, Wilk's Extended Logistic Regression (Wilks, 2009); the censored GEV in Scheuerer's EMOS (Scheuerer, 2014); the gamma adapted BMA of Sloughter et al. (2007)).

Marginal normality (eventually artificially retrieved) of each variable is not the end of the story: what can be said about conditional dependence? Modelers should worry about troubles stemming from multidimensionality. To describe multivariate links, additional modeling efforts have to be undertaken for instance design a large multidimensional variance-covariance matrix if the pleasant but tight scheme of multivariate normal models is acceptable, or make recourse to a wider family of copulas.

5. Accounting for multivariate structures

Hydrological forecasting systems fed by meteorological ensembles forecast require a multidimensional representation:

- Time and space dependencies are inherent to the multiple locations and various forecast lead-times (up to 200 days ahead at HQ). For instance, EDF forecasting system includes about a hundred watersheds and requires coherent series of a dozen days ahead daily probabilistic forecasts on each of them.
- It is unavoidable to take into account several meteorological variables simultaneously: the interaction between precipitation and temperature has to be considered when one deals with hydrology (see the meteorological sub-problem identified in Section 6).
- Ensemble weather forecasts generate several members, for each variable, at each lead time and location.

It is hopeless to imagine that one will succeed in one step to cast all variables after some transformation into the multivariate normal framework: variables can still depend non linearly and may still have heterogeneous variances. A first solution is to break the problem into pieces that can eventually be approximated by a multivariate model around a working reference point that may change with seasons or river and meteorological regimes. Other solutions involve recourse to more sophisticated multivariate models than the normal structure. In this case copulas (such as in Möller et al. (2013)) appears as a tool of choice (Genest and Favre, 2007) as they are able to model the dependence independently of the marginal distributions.

For sake of simplicity current practice has focused on non parametric copula. The Schaake shuffle (Clark et al., 2004), for instance, copies the multivariate space-time rank structure of past events (for situations similar to the one to be predicted) to a simulated sample from independent marginal predictive distributions. This method has been chosen both in SCHEF (Bennett et al., 2014) and HEFS (Demargne et al., 2014). In the case of ensemble weather forecasts, one can similarly copy the ensemble rank structure. This type of method (similar to the one in use in EDF's forecasting system) is called Ensemble Copula Coupling (Scheffzik et al., 2013). It has been extended by the SimSchaake shuffle (Scheffzik, 2015) that makes recourse to both past and ensemble ranks.

Sections 3 and 4 have dealt with general modeling challenges and solutions that emerged from probabilistic forecasting problems. Which value can be awarded to these theoretical solutions when they are confronted with the operational context? How far do practitioners worry about the modeling challenges of probabilistic forecasting in the various case studies? How are practical solutions implemented in operational systems? How do they relate to, or can they be interpreted in terms of, statistical thinking?

6. Splitting the problem into meteorological and hydrological sub-problems

In the hereby considered operational contexts, a natural modeling scheme in association with a (supposedly efficient) RRS leads to the identification and the separate treatment of two main uncertainty sources: (i) *meteorological uncertainty* (ii) *hydrological uncertainty* i.e. uncertainty due to imperfect representation of the hydrological phenomenon by the RRS.

The natural formalization of our forecasting problem relates the predictand Q (streamflows) to the predictor A (meteorological forecasts) through P (meteorological measurements). Summarized as a probabilistic graphical model, it can be written as $A \rightarrow P \rightarrow \mathcal{M}(P) \rightarrow Q$. Thus:

- it is highly recommended to adequately exploit the (generally) provided actual precipitation record P , especially (as it is generally the case) when joint (Q, P) records are larger than (Q, A) 's (see Section 4.1.2 for detailed arguments in favor of indirect inference schemes). This amounts to making the following decomposition based on conditional distributions: $[Q|A] = \int [Q|P, A] [P|A] dP$. Arguments to proceed this way are similar to the ones advocating the Bayesian modeling scheme (see Section 4.1).
- $[Q|P, A]=[Q|P]$: the modeling framework encompasses a P -sufficiency assumption. Putting words on it, once the actual meteorology is known, the meteorological forecasts are no longer informative relatively to the streamflows to be predicted. Thus, $\int [Q|P, A] [P|A] dP = \int [Q|P] [P|A] dP$.

A model for $[Q|P]$ cannot be developed without further hydrological knowledge: the streamflow at time t does not only depend on meteorology at time t but also, and even more, on past meteorology over a period of several days or months. Hopefully, practitioners are usually provided with a RRS. A RRS is a highly non-linear function \mathcal{M} taking as inputs meteorological variables (such as precipitation and temperature) and outputting variables in the hydrological domain: $\mathcal{M} : P \rightarrow \mathcal{M}(P)$ (a series of streamflows). The lumped RRS model \mathcal{M} assembles mathematical operators to mimic the fate of a waterdrop after landing and its parameters are chosen such that $\mathcal{M}(P)$ is as close as possible (in general according to the Nash criterion, i.e. a slightly modified mean square error criterion) to the corresponding measured streamflows Q . Appendix A summarizes the internal mechanics of some RRS's in use in operational contexts.

Directly deriving a model for $[Q|\mathcal{M}(A)]$ in one step without the help of meteorological records P may be enticing. However, as underlined above, P records should be taken into account. Moreover, the apparent simplicity of such a modeling scheme can be deceptive: \mathcal{M} is far from being regular and rough behaviors of both the meteorological and the hydrological treatments (respectively described in Sections 7.1 and 7.2) interfere.

As noticed by Krzysztofowicz (1999), under some \mathcal{M} -efficiency, i.e. assuming

$[Q|\mathcal{M}(P), P] = [Q|\mathcal{M}(P)]$, we have:

$$[Q|A] = \int_P [Q|\mathcal{M}(P)] \times [P|A] dP \quad (6)$$

Then, two models can be derived and calibrated separately:

- i) a model to get $[P|A]$ (the *pre-processor*, to deal with meteorological uncertainty) and
- ii) a model to get $[Q|\mathcal{M}(P)]$ (a *post-processor*, to deal with hydrological uncertainty).

The resulting forecasting structure is illustrated in Figure 6.

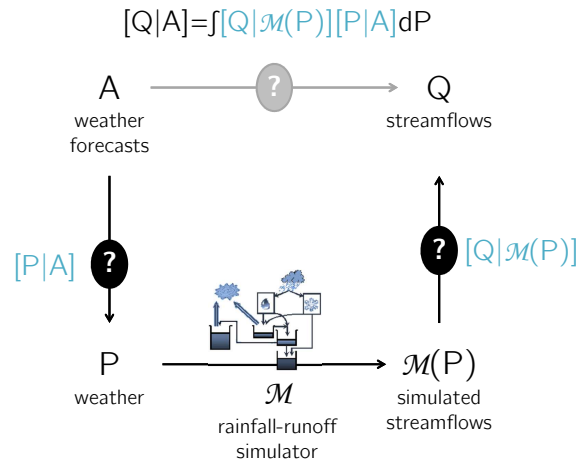


Figure 6: Forecasting requires the combination of the meteorological and hydrological sub-problems. This figure illustrates the forecasting structure resulting from the P -sufficiency and the \mathcal{M} -efficiency assumptions described Section 6.

Note that the \mathcal{M} -efficiency assumption may be partially relaxed under the form:

$$[Q|P] = [Q|\mathcal{M}(P), \mathcal{V}(P)]$$

although \mathcal{M} is generally fitted so that $\mathcal{M}(P)$ is as informative as possible about the streamflow Q , with regards to its expected value, some additional information $\mathcal{V}(P)$ about its conditional variance may be found elsewhere. For instance, it can be obtained by extending RRS outputs to include its state variables. Appendix A depicts in details some RRS structure and state variables.

The separate treatment of uncertainties is ubiquitous in the operational context (see for example Krzysztofowicz (2002), SCHEF and EDF's and NOAA's (Demargne et al., 2014) operational forecasting systems). This is not only because of a theoretical "Divide and Conquer" argument, but very practical reasons also come into play:

- Meteorological probabilistic forecasts $[P|A]$ are a valued by-product of the system: they can be of interest for other purposes such as power plant and drainage structures design and management or for energy supply, water temperature and ice layer forecasting.
- Human experts, either meteorologists or hydrologists, can play their specific role without perturbing the forecasting system as a whole: when one splits the problem as described above, weather experts can easily take action modifying the predictive distribution $[P|A]$ while experts in hydrology would focus on $[Q|P]$. They can integrate additional information relative to weather and to hydrology such as an alternative forecast of different nature. At EDF, for instance, it may be information from volunteering observers. Both at EDF and HQ, experts moreover have to get involved in the forecasting system in order to fully understand the forecast as well as being able to communicate it.

7. Getting back to the standard statistical toolbox in the meteorological and in the hydrological subproblems

As mentioned in Section 5, multidimensionality is a major modeling challenge, common to the meteorological and the hydrological sub-problems: time and space dependencies (as well as dependencies between several variables in the meteorological setting) have to be accounted for to issue a probability distribution of a multivariate predictand.

All the operational systems under study except HEFS's hydrological uncertainty processor dodge this challenge by considering the subproblems location by location, variable by variable and lead time by lead time. Afterwards, simulations may regain multidimensional coherence with the help of some empirical copulas (mentioned in Section 5) as in EDF's and in HEFS's meteorological and hydrological uncertainty processors and in SCHEF's meteorological uncertainty processor.

However, empirical copulas do not fully take advantage of the information provided by correlations. When considering temporal consistency, this drawback is very significant but may be reduced by using a data assimilation (DA) technique: such a technique takes into account the last measurements of the quantity of interest as well as the last forecasts errors. Climate simulators producing meteorological forecasts include a DA procedure, thus justifying the lead time by lead time meteorological uncertainty processing. Similarly, operational hydrological uncertainty processors, such as at EDF-DTG's, may proceed lead time by lead time when some sort of DA procedure is associated with their RRS.

From our theoretical position of statisticians, Y and X hereafter respectively denote the predictands (meteorological measurements P in the meteorological subproblem, streamflows Q in the hydrological one) and the predictors (meteorological forecasts A in the meteorological subproblem, outputs from the RRS $\mathcal{M}(P)$ in the hydrological one) at a specific location, and for a specific variable in the meteorological subproblem (the minimal temperature or the amount of rain for instance). Denoting by t the time when the forecast is issued (generally at daily rates), h the lead time with H the forecasting horizon and X_t^h the forecast issued at time t for time $t + h$, the target distribution is:

- $\left[Y_{t+1, \dots, t+H} | X_{1, \dots, t}^{1, \dots, H}, Y_{1, \dots, t} \right]$ in the general case,
- $\left[Y_{t+h} | X_{1, \dots, t}^{1, \dots, H}, Y_{1, \dots, t} \right]_{h \in (1:H)}$ when proceeding lead time by lead time, which is reduced to $\left[Y_{t+h} | X_t^h \right]_{h \in (1:H)}$ if assuming, as often in the meteorological setting, that the available information X_t^h at time t encompasses all necessary and sufficient information to predict Y at time $t + h$.

In the meteorological subproblem ($Y = P$, $X = A$), forecasting systems generally proceed lead time by lead time. The main difficulty is the multidimensionality of the ensemble A_t^h at any time t and lead time h . We deal with this problem in Section 7.1. In the hydrological subproblem ($Y = Q$, $X = \mathcal{M}(P)$), an important challenge is to take into account the heterogeneous behavior of the RRS transform $\mathcal{M}()$: the discrepancy between Q and $\mathcal{M}(P)$ strongly depends on the hydrological regime. Some forecasting systems derive a model accounting for temporal dependencies while others proceed in two steps: first dealing with temporal dependencies and then accounting for uncertainties. We describe existing operational methods in Section 7.2.

7.1. From weather forecasts to probabilistic weather forecasts: dealing with ensembles

In the meteorological subproblem, the target distribution is generally $\left[P_{t+h} | A_t^h \right]$, where P_{t+h} stands for a meteorological target at time $t + h$ and A_t^h stands for a meteorological forecast issued at time t for time $t + h$. A_t^h may take the form of a single deterministic weather forecast \hat{P}_{t+h} , an ensemble of K members $\left\{ \hat{P}_{k, t+h} \right\}_{k=1:K}$.

Line 1 of Table 1 gives an overview of possible meteorological inputs in operational hydrological forecasting systems. Lines 2 to 4 summarize the associated meteorological uncertainty processors.

Some companies take as input a deterministic weather forecast $A_t^h = \hat{P}_{t+h}$. This is for instance (up to now) the case of HEFS for short range forecasts (up to seven days), of HQ's system (up to nine days) and of SCHEF. In those cases, A_t^h is unidimensional. After a possible normalization procedure (which can be a difficult challenge - see Section 4.2), $\left[P_{t+h} | A_t^h \right]$ can be derived from a standard bivariate normal model following the modeling framework of Section 4. It is for instance the modeling choice in SCHEF's and HEFS's meteorological uncertainty processors

(Robertson et al., 2013), which make use of joint bivariate normal models, while HQ makes use of more complex conditional models in a non-homogeneous regression manner.

An ever increasing number of hydrological forecasting systems take as input ensemble weather forecasts $A_t^h = \left\{ \hat{P}_{k,t+h} \right\}_{k=1:K}$. This is for instance the case of EDF's and CEMIG's systems, which use ensemble weather forecasts from the ECMWF for short range forecasts, and of HEFS when it comes to long range forecasts.

Ensemble weather forecasts $[A_t^h = \left\{ \hat{P}_{k,t+h} \right\}_{k=1:K}]$ should ideally be considered to be drawn from the actual weather (P_{t+h}) distribution. It is to say that the observed weather trajectory should not be statistically distinguishable from the ensemble members trajectories. This is for instance the underlying modeling choice in CEMIG's forecasting system: no meteorological uncertainty processor is included. However, this assumption hardly ever holds (Bougeault et al., 2010) since the way ensemble members are generated only illustrates a part of the uncertainty due to the miss-representation of the phenomenon by climatic numerical models (essentially the uncertainty on initial conditions). Furthermore it does not rely on a statistical model. Therefore it does not provide any reliability warranty and requires a post-processing step modeling ($P|A$).

When inputting ensemble weather forecasts in the system, $A_t^h = \left\{ \hat{P}_{k,t+h} \right\}_{k=1:K}$ becomes by definition multidimensional. One may directly apply a multivariate normal model as in Section 4. This is for instance the modeling choice of Krzysztofowicz and Maranzano (2006) in their Bayesian Processor of outputs (BPO). However, operational systems traditionally use additional hypotheses, leading to advantageously parsimonious models:

- A rather common assumption is that most ensemble weather forecasts information is conveyed by the ensemble mean: $(P_{t+h}|A_t^h) = (P_{t+h}|\bar{A}_t^h)$. For instance, in HEFS, a bivariate normal model is derived for (P_{t+h}, \bar{A}_t^h) (Wu et al., 2011; Schaake et al., 2007a).
- The previous assumption can be relaxed assuming that ensemble forecasts information is mostly conveyed by their sample mean and standard deviation. For instance, a normal model for $(P_{t+h}|\bar{A}_t^h, s(A_t^h))$ is derived in the well-known Ensemble Model Output Statistics (EMOS) method (Gneiting et al., 2005).

Those hypotheses can be interpreted in terms of exchangeability assumptions for the ensemble members, with some random effect, such as the latent ensemble mean at time $(t+h)$ in the HEFS case.

Truth + errors models may also be relevant to model meteorological uncertainty with ensemble weather forecasts. One can for instance assume there exists, at each forecasting time, a best but unknown member A_t^{h*} among all members (when dealing with exchangeable ensembles, each member has the same probability to be the best one). Conditional upon member k being the best member ($A_t^{h*} = A_{t,k}^h$), P_{t+h} is

supposed to occur nearby A_t^{h*} , with a vicinity modeled via a kernel smoothing function. Those mixture models have the advantage to avoid assumption about the shape of the final distribution such as unimodality. The most popular of them is the Bayesian Model Averaging (BMA) method of Raftery et al. (2005). EDF's meteorological uncertainty processor, called MIX, can be considered as a variant of that modeling framework. In MIX, $(P_{t+h}|A_t^{h*} = A_{t,k}^h)$ is derived on the basis of past analogue atmospheric situations.

7.2. From rainfall-run off simulators' outputs to hydrological probabilistic forecasts: dealing with temporal dependency

For the hydrological subproblem, in the general case, the target distribution is

$$\left[Q_{t+1, \dots, t+h, \dots, t+H} | \mathcal{M}(P)_{1, \dots, t}^{1, \dots, H}, Q_{1, \dots, t} \right].$$

A major challenge here comes from the necessity to account for streamflow temporal consistency, since river memory may last over several months. This is for instance critical for HQ, where streamflow forecasts are mainly used for dam management.

Lines 6 to 9 of Table 1 give an overview of hydrological uncertainty processors in some operational hydrological forecasting systems.

7.2.1. Joint treatment of temporal dependency and hydrological uncertainty

A typical model for both accounting for the temporal correlation structure and taking into account external values (in our case outputs from a RRS) is the Auto Regressive model with eXternal inputs (ARX). This is for instance the modeling choice in HEFS hydrological uncertainty processor, which is described in Seo et al. (2006). Similarly, Krzysztofowicz and Maranzano (2004) suggest to make use of an ARX-like model but in the Bayesian setting described in Section 4.1.2. They first derive an AR model for Q , then a conditional normal model for $(\mathcal{M}(P)|Q)$ and finally get back to $[Q|\mathcal{M}(P)]$ through the Bayes' formula.

A difficulty is then to retrieve spatial consistency: it has to be retrieved separately from temporal consistency which is not envisaged in the usual empirical copula schemes. To our best knowledge, this problem has not been treated yet in the operational forecasting systems exemplified in this paper but is of due interest for operational use.

7.2.2. A separate treatment of temporal dependency and hydrological uncertainty

Another way of dealing with temporal consistency in the hydrological context is to proceed in two steps: first (i) resorting to some sort of DA and then (ii) accounting for hydrological uncertainty by treating the problem lead time by lead time as in the meteorological problem.

Conventionally, DA consists in a modification of the RRS's state variables such that last RRS's outputs (taking meteorological measurements as inputs) would have been identical to the corresponding streamflow measurements. For instance, CEMIG relies on an automatic DA by Da Paz et al. (2007), while HQ relies both upon automatic and manual DA.

Other operational systems alternatively use a deterministic correction to take into account last streamflow measurements. For example, the SCHEF system involves the "dual pass error-correction technique" (Pagano et al., 2011). With this technique, the target distribution becomes $\left[Q_{t+1,\dots,t+H}|\mathcal{M}(P)'_t^{1,\dots,H}\right]$, where:

$$\mathcal{M}(P)'_t^h = \max\left(0, \mathcal{M}(P)_t^h + \beta \exp\left\{-\frac{h-1}{f}\left(Q_t - \mathcal{M}(P)_t^0\right)\right\}\right)$$

and β and f are parameters to be estimated. EDF uses a similar deterministic correction which emanates from forecasters' experience.

After resorting to a DA or a deterministic correction technique to get $\mathcal{M}(P)'$ such that:

$$\left[Q_{t+1,\dots,t+H}|\mathcal{M}(P)_{1,\dots,t}^{1,\dots,H}, Q_{1,\dots,t}\right] \simeq \left[Q_{t+1,\dots,t+H}|\mathcal{M}(P)'_t^{1,\dots,H}\right],$$

it becomes possible to treat the hydrological subproblem similarly to the meteorological one, i.e. lead time by lead time. For instance, at EDF, a heteroscedastic conditional normal model is derived for each forecasting lead time (after log-normalization):

$$\left(\ln(Q_{t+h})|\mathcal{M}(P)'_t^h = m\right) = b_h(m) + \ln(m) + \sigma_h(m)\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1),$$

where b_h and σ_h are tabulated functions of m . Then, as classically in the meteorological subproblem, after a lead time by lead time forecast, one can have recourse to empirical copulas (as usually done in the meteorological setting). EDF makes use of a technique similar to ECC to this aim.

Alternatively, at HQ, some RRS state variables are perturbed to generate several hydrological ensemble members. Such choice demonstrates HQ's confidence to their RRS: it assumes hydrological uncertainty can be fully retrieved through the RRS itself. In SCHEF and in CEMIG's system, it is even assumed that hydrological uncertainty can be neglected for short forecasting lead times forecasts (up to 9 days in SCHEF): $\left(Q_{t+1,\dots,t+H}|\mathcal{M}(P)'_t^{1,\dots,H}\right) = \mathcal{M}(P)'_t^{1,\dots,H}$.

This assumption is questionable. In fact, for short forecasting lead times, meteorological forecasts have a poor impact on predicted streamflows (except when

considering very small watersheds), the uncertainty is thus mainly of hydrological nature. Neglecting hydrological uncertainty should thus result in too under-dispersed forecasts. This choice however comes with the advantage of preserving spatial and temporal structures originating from weather forecasts, thus avoiding the use of copulas.

8. Integrating meteorological uncertainty with hydrological uncertainty back into total uncertainty

When the system in use treats meteorological and hydrological uncertainties separately (as described in Section 6), two models are provided:

- i) a model to get a distribution forecast of meteorology from meteorological forecasts, $[P|A]$, and
- ii) a model to get a distribution forecast of streamflows from the outputs of a RRS (taking as input meteorology), $[Q|\mathcal{M}(P)]$.

To get the target-distribution $[Q|A = a_{new}]$ for a newly issued ensemble weather forecast a_{new} , one should thus perform the combination expressed in Eq. (6). One can proceed as EDF and HEFS by inputting simulations from the meteorological predictive distribution in the RRS. Each resulting output is then treated as if it was obtained from a meteorological measurement. In a code-like expression, we would have:

To get $\{q_1, \dots, q_K\} \sim [Q|A = a_{new}]$:

1. Simulate $\{p_1, \dots, p_K\}$ from $[P|A = a_{new}]$
2. Input each p_k in the RRS to get $m_1 = \mathcal{M}(p_1), \dots, m_K = \mathcal{M}(p_K)$
3. For each m_k , $k \in \{1, \dots, K\}$, simulate q_k from $[Q|\mathcal{M}(P) = m_k]$

The target distribution can then be approximated with a parametrized distribution to evaluate any subproduct of interest such as most likely value, simulations, quantiles, percentiles, probability of over-passing a threshold, etc. The most simple way to do so is to resort to a kernel method (as in the R `density` function for instance).

9. Conclusion and discussion

This article has proposed a general framework to read how operational hydrological forecasting systems make predictions. We placed emphasis on information as well as uncertainty conveyed by an ensemble of meteorological simulator runs used as inputs to a rainfall/runoff simulator also subjected to errors.

We do not advocate systematic recourse to probabilistic modeling where good engineering sense and human forecasters' judgments support efficient operational actions, but simply warn against considering probabilistic models as unrealistically over-simplistic: all statistical models simplify, but each simplification must be evaluated in terms of its dividends. The purpose of this article is to point out, based on miscellaneous case-studies of hydrological forecasting systems, that uncertainty is everywhere nested along the components which constitute the existing forecasting devices and to demonstrate the benefits that would accrue through a better collaboration between forecasters and statisticians. Not assessing uncertainty effectively, *i.e.* with reluctance to the probabilistic framework, leaves the door open to misinterpretations of system predictions and economic inefficiency, through unspecified hypotheses for instance.

However, in some examples of the forecasting systems we studied, probability theory and forecasting practice go along well:

- Calibration is a shared objective even if the ways to pertain this property differ. Statisticians would make sure their model is calibrated by recourse to inference with an appropriate likelihood maximization (see Section 2) while forecasters would regularly check their past forecasts by calibration diagnostics tools. For instance, many times a year, especially when the forecasting devices are updated, EDF scientists are in charge to run diagnostic tools including calibration checks. In HEFS and HQ, verification (calibration and sharpness) procedure is an important part of their system (Demargne et al., 2014; Perreault et al., 2017).
- In miscellaneous forecasting systems, components that take into account uncertainty either refer to a standard probabilistic model or can be interpreted in terms of a known statistical recipe. For example in SCHEF and HEFS, explicit recourse is made to bivariate normal models to depict meteorological uncertainty, in EDF a tabulated function of the inflow gives the effect of the hydrological uncertainty on the rainfall/run-off proposal, which can be interpreted as an unformalized log-transformed linear model with heteroschedastic variance.
- The division of the main sources of uncertainty between meteorological and hydrological ones, which appears natural and convenient for a probabilistic modeler, seems to be implemented in the effective forecasting operations. This practice also traduces, in the operational context, a will to benefit from meteorological forecasters' expertise. Indeed, many teams such as HQ involve different people with either a meteorological skill or an hydrological one.

Nevertheless, there remains a gap between a purely model-based view of the problem and the implementation that is achieved in practice:

- Even if the necessity of a conceptually rigorously defined framework is well understood in the statistical community, there are numerous reasons why probabilistic reasoning provides little intuitive insights for a general audience of hydrological forecasters.
 - i) While a standard training on probability would heavily focus on univariate probability distributions with trivial examples mostly idealized from the domain of a simple observable random mechanism, the type of uncertainty met in ensemble-based forecasts is very complicated to represent and to assess since multidimensional random variables, most of them often unobservable, are linked with complex black-box numerical codes.
 - ii) Part of misunderstanding can also be attributed to language traps and deeply embedded educational traits: an hydrologist is a physicist trained to understand a "model" as a set of equations as detailed as possible, deterministically depicting the behavior of a system, while a statistician will rather consider that a model is merely stochastic by nature; the former makes statements argued on causality while correlation is all that matters for the judgments made by the latter; practitioners would rather visualize randomness by the understandable mean of a (moderate) sample of scenarios while statisticians would rely on an abstract object, a (sometimes non analytical) probability distribution function. The technique of assimilation of last streamflows in EDF's FS and SCHEF (dual-pass error correction model) clearly illustrates this duality in the perception of the notion of model, as well as hydrologists' faith in their RRS. Hydrologists may dream that RRS's state variables would fit with the physical quantities they were meant to mimic when the RRS was conceived. For instance, HQ forecasters regularly update snow water equivalent quantities stored in HSAMI with field measurements. Statisticians do not have such expectations (except if parameters' training set includes such measurements).
- Forecasting systems have been implemented progressively and modularly with rather conservative upgrades when improvements appeared necessary. For instance, it is planned to add an automatic data assimilation procedure to HEFS (Demargne et al., 2014). EDF's hydrological processor is divided into three parts, first the last observed streamflow is taken into account through a deterministic correction, then a noisy part is added to take into account hydrological uncertainty and finally simulations are reordered to respect some sort of time memory through an empirical copula. Such sequences of interventions are made independently at different version steps of the forecasting device, with the successive implementation of separate sub-models that should not change too much the working routine of the forecasters, whilst statisticians

would rather plead to perform the three stages in one go...but of course this is only possible in a quiet office away from the pressure of operational demands.

- Statistical models attempt to try to represent "known unknowns" but statisticians are left unarmed when facing the "unknown unknowns" that the practitioners routinely meet. Surprising events occur for instance when what was considered as truth becomes random: due to wind and other perturbations, inflows, especially for large North American hydro-power reservoirs, may become precisely evaluated only a week later than the day prediction is needed for.
- Some operational constraints may be set aside by statisticians. For example, forecasters of both EDF and HQ forecasting systems have to be able to communicate previsions issued by those systems. This makes it even more necessary to involve them as actors of the forecasting system. Some further thinking about how to integrate expertise in the system efficiently and with no loss of reliability warranties is needed.
- While statisticians can theoretically be satisfied with any choice of strictly proper scoring rule, practitioners are concerned by choosing a "good" one, i.e. a scoring rule consistent with operational objectives. Figure 7 shows positive correlation between the performance of some probabilistic forecasts as evaluated by the CRPSS (Continuous Ranked Probability Skill Score) and the corresponding net gain from Churchill Falls complex management. The CRPS(S) is thus, to the aim of maximizing the net gain, a "good" score. One might wonder how to construct a score tailored to specific operational objectives, with a possible focus on some extreme events for instance. Other criteria of more technical nature may guide the practitioner toward a particular scoring rule. For instance, contrary to the CRPS, the ignorance score is *local* i.e. "no credit is given for assigning high probabilities to values near but not identical to the one materializing" (Gneiting et al., 2005).

One might imagine a future when the (hydro-)meteorological representation of the water cycle will be so detailed and accurate that quantifying their veracity (the job of statisticians) and assessing streamflows (the job of hydrologists) will no longer be useful nor relevant. This might already be ongoing, with the development of the MESH modeling system Pietroniro et al. (2007) and the development of ECMWF-IFS for instance. ECMWF-IFS integrates a land surface component (HTESSSEL), thus directly produces ensemble forecasts of hydrological nature (Alfieri et al., 2013) to which only a river-routing part has to be appended to get the target hydrological forecasts. However, statistics in hydrology will remain a "necessary evil" to properly communicate uncertainty in the most amenable quantified form. A final step towards an ideal forecasting system would rely on a joint inference of both physical and

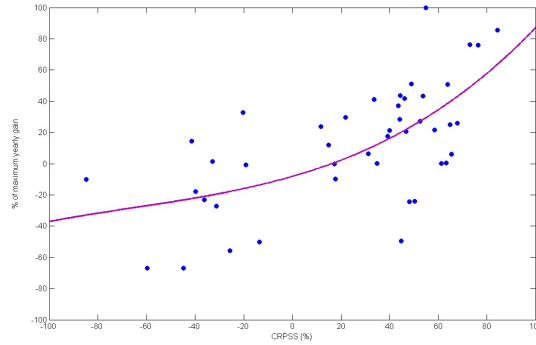


Figure 7: Relation between the net gains and the corresponding values of the CRPSS (Continuous Ranked Probability Skill Score) of some probabilistic forecasts. Figure adapted from Perreault et al. (2017)

statistical models, as attempted, although with identifiability problems, in Schoups and Vrugt (2010) and Evin et al. (2014).

Acknowledgements

This work was supported by Électricité de France and by Hydro-Québec [research grant number 694R] through the thesis of the first author.

We would like to thank Joël Gailhard, Rémy Garçon and the forecasting and development team at EDF-DTG, Vincent Fortin at Environment Canada, the forecasting team at Hydro-Québec, the research team at IREQ, Fabian Tito Arandia Martinez for supplying the processed data for our application and Jacques Bernier for his very helpful comments on previous versions of this article.

Appendix A. Rainfall-Runoff simulators, tools of choice for operational hydrology

RRSs deterministically mimic the water progression through a given watershed from the surface of the watershed to the outflow (which is the quantity to be predicted here). In the hereby considered cases, the RRSs are of global lumped type. In such simulators, the main stages of the water pathway in the basin are represented by conceptual interrelated reservoirs, such as a snow storage, a surface storage and a groundwater storage. We will hereafter call their filling ratios, or combinations of them, *state variables* of the simulator, V .

Global RRSs look all alike: diagrams of EDF's conceptual RRS, MORDOR, and of GR4J (Perrin et al., 2003), are given as twin examples in Figure A.8. The SCHEF's simulator structure is very similar to GR4J and Hydro-Québec also resorts to a RRS of conceptual type called HSAMI (Fortin, 2000).

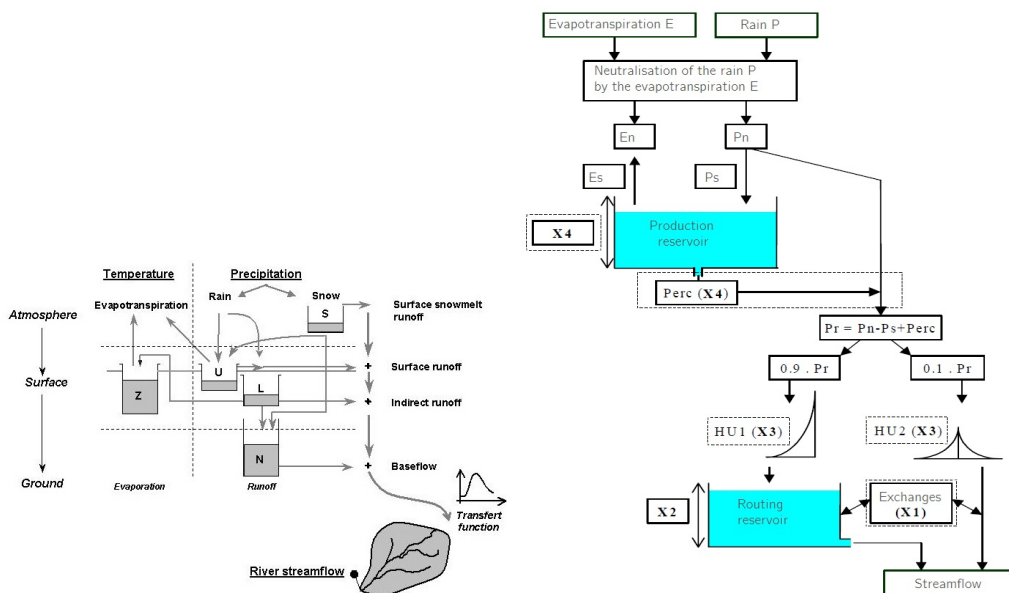


Figure A.8: Diagram of EDF-DTG's (left, from Andreassian et al. (2006)) and ISTEA's (right, translated from Perrin (2000)) rainfall-run off simulators (resp. MORDOR and GR4J)

For each forecast day, the simulator takes as inputs the daily watershed's meteorological variables (P) which can be (liquid and solid / total) precipitations and (minimal and maximal / mean) daily temperature for example. The simulator then updates the filling level of its fictitious reservoirs according to their theoretical response to weather events (such as a rainfall) and according to their relations, depicted by assembling simple mathematical operators (with parameters previously fitted on a training database). The simulator finally outputs a potential daily outflow $\mathcal{M}(P)$, and we denote by $\mathcal{M} : P \mapsto \mathcal{M}(P)$ the transformation operated by the simulator.

References

- Alferi, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS—global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci* 17, 1161–1175.
- Anderson, E.A., 1973. National Weather Service river forecast system: Snow accumulation and ablation model. volume 17. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.
- Andreassian, V., Bergstrom, S., Chahinian, N., Duan, Q., Gusev, Y., Littlewood, I., Mathevet, T., Michel, C., Montanari, A., Moretti, G., et al., 2006. Catalogue of the models used in MOPEX 2004/2005. IAHS publication 307, 41.
- Bennett, J.C., Robertson, D.E., Shrestha, D.L., Wang, Q., Enever, D., Hapuarachchi, P., Tuteja, N.K., 2014. A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days. *Journal of Hydrology* 519, 2832–2846.
- Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D.H., Ebert, B., Fuentes, M., Hamill, T.M., Mylne, K., et al., 2010. The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society* 91, 1059–1072.
- Box, G.E., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* , 211–252.
- Bröcker, J., 2008. On reliability analysis of multi-categorical forecasts. *Nonlinear Processes in Geophysics* 15, 661–673.
- Buizza, R., Houtekamer, P., Pellerin, G., Toth, Z., Zhu, Y., Wei, M., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review* 133, 1076–1097.
- Burnash, R., Singh, V., et al., 1995. The NWS river forecast system-catchment modeling. *Computer models of watershed hydrology.* , 311–366.
- Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology* 5, 243–262.
- Cloke, H., Pappenberger, F., 2009. Ensemble flood forecasting: a review. *Journal of Hydrology* 375, 613–626.
- Collet, J., Épiard, X., Coudray, P., 2009. Simulating hydraulic inflows using PCA and ARMAX. *The European Physical Journal-Special Topics* 174, 125–134.

- Collischonn, W., Haas, R., Andreolli, I., Tucci, C.E.M., 2005. Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *Journal of Hydrology* 305, 87–98.
- Da Paz, A.R., Collischonn, W., Tucci, C., Clarke, R., Allasia, D., Boegh, E., Kunstmann, H., Wagener, T., Hall, A., Bastidas, L., et al., 2007. Data assimilation in a large-scale distributed hydrological model for mediumrange flow forecasts. *IAHS PUBLICATION* 313, 471.
- De Finetti, B., 1937. La prévision : ses lois logiques, ses sources subjectives, in: *Annales de l'institut Henri Poincaré*, pp. 1–68.
- Demargne, J., Wu, L., Regonda, S.K., Brown, J.D., Lee, H., He, M., Seo, D.J., Hartman, R., Herr, H.D., Fresch, M., et al., 2014. The science of NOAA's operational hydrologic ensemble forecast service. *Bulletin of the American Meteorological Society* 95, 79–98.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources* 30, 1371–1386.
- Evin, G., Merleau, J., Perreault, L., 2011. Two-component mixtures of normal, gamma, and Gumbel distributions for hydrological applications. *Water Resources Research* 47.
- Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research* 50, 2350–2375.
- Fan, F.M., Collischonn, W., Meller, A., Botelho, L.C.M., 2014. Ensemble streamflow forecasting experiments in a tropical basin: The São Francisco river case study. *Journal of Hydrology* 519, 2906–2919.
- Fortin, V., 2000. Le modèle météo-apport HSAMI: historique, théorie et application. Institut de recherche d'Hydro-Québec, Varennes .
- Fortin, V., Ouarda, T., Rasmussen, P., Bobée, B., 1997. Revue bibliographique des méthodes de prévision des débits. *Revue des sciences de l'eau/Journal of Water Science* 10, 461–487.
- Garçon, R., 1996. Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. Bilan de l'année 1994-1995. *La Houille Blanche* , 71–76.
- Gelfand, A.E., Smith, A.F., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* 85, 398–409.

- Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering* 12, 347–368.
- Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359–378.
- Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review* 133, 1098–1118.
- Guay, C., Minville, M., Braun, M., 2015. A global portrait of hydrological changes at the 2050 horizon for the province of Québec. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* 40, 285–302.
- Hagedorn, R., Smith, L.A., 2009. Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Applications* 16, 143–155.
- Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research* .
- Herr, H.D., Krzysztofowicz, R., 2005. Generic probability distribution of rainfall in space: the bivariate model. *Journal of Hydrology* 306, 234–263.
- Jandhyala, V.K., Liu, P., Fotopoulos, S.B., 2009. River stream flows in the northern Québec Labrador region: A multivariate change point analysis via maximum likelihood. *Water resources research* 45.
- Jolliffe, I.T., Stephenson, D.B., 2012. *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.
- Katz, R.W., Murphy, A.H., 1997. *Economic value of weather and climate forecasts*. Cambridge University Press.
- Krzysztofowicz, R., 1997. Transformation and normalization of variates with specified distributions. *Journal of Hydrology* 197, 286–292.
- Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research* 35, 2739–2750.
- Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology* 268, 16–40.
- Krzysztofowicz, R., Maranzano, C.J., 2004. Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology* 299, 15–44.
- Krzysztofowicz, R., Maranzano, C.J., 2006. Bayesian processor of output for probabilistic quantitative precipitation forecasts. *Manuscript in review* .

- Lewin, K., 1952. *Field theory in social science: selected theoretical papers*. London: Tavistock.
- Mathevet, T., 2010. Erreur empirique de modèle. Note technique interne D4165/NT/2010-00395-A. EDF-DTG.
- Mathevet, T., Garavaglia, F., Garçon, R., Gailhard, J., Paquet, E., 2009. Operational hydrological ensemble forecasts in france. recent development of the french hydropower company (edf), taking into account rainfall and hydrological model uncertainties., in: EGU General Assembly Conference Abstracts, p. 10248.
- Möller, A., Lenkoski, A., Thorarinsdottir, T.L., 2013. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society* 139, 982–991. URL: <http://dx.doi.org/10.1002/qj.2009>, doi:10.1002/qj.2009.
- Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research* 40.
- Nason, G., 2006. Stationary and non-stationary times series. *Statistics in Volcanology*. Special Publications of IAVCEI 1, 000–000.
- Obled, C., Bontron, G., Garçon, R., 2002. Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. *Atmospheric research* 63, 303–324.
- Pagano, T., Wang, Q., Hapuarachchi, P., Robertson, D., 2011. A dual-pass error-correction technique for forecasting streamflow. *Journal of hydrology* 405, 367–381.
- de Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.P., Frappart, F., Calmant, S., Bulhões Mendes, C.A., 2013. Large-scale hydrologic and hydrodynamic modeling of the amazon river basin. *Water Resources Research* 49, 1226–1243.
- Perreault, L., Gaudet, J., Delorme, L., Chatelain, S., 2017. Verification of medium-to-long range hydrological probabilistic forecasts, in: *Handbook of Hydrometeorological Ensemble Forecasting*, Editors Qingyun Duan, Florian Pappenberger, Jutta Thielen, Andy Wood, Hannah L. Cloke and John C. Schaake, Springer-Verlag Berlin Heidelberg (submitted).
- Perreault, L., Parent, E., Bernier, J., Bobee, B., Slivitzky, M., 2000. Retrospective multivariate bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment* 14, 243–261.
- Perrin, C., 2000. *Vers une amélioration d'un modèle global pluie-débit*. Ph.D. thesis. Institut National Polytechnique de Grenoble-INPG.

- Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* 279, 275–289.
- Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Versegny, D., Soulis, E., Caldwell, R., Evora, N., et al., 2007. Development of the mesh modelling system for hydrological ensemble forecasting of the laurentian great lakes at the regional scale. *Hydrology and Earth System Sciences* 11, 1279–1294.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* 133.
- Robertson, D., Shrestha, D., Wang, Q., 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences* 17, 3587–3603.
- Roulin, E., Vannitsem, S., 2015. Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. *Hydrological Processes* 29, 1434–1449.
- Savage, L.J., 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66, 783–801.
- Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D., 2007a. Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrology and Earth System Sciences Discussions* 4, 655–717.
- Schaake, J.C., Hamill, T.M., Buizza, R., Clark, M., 2007b. Hepex: the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society* 88, 1541–1547.
- Schaefli, B., Talamba, D.B., Musy, A., 2007. Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology* 332, 303–315.
- Schefzik, R., 2015. Physically coherent probabilistic weather forecasts using multivariate discrete copula-based ensemble postprocessing methods. Master's thesis. Universität Heidelberg, Faculty of Mathematics and Computer Science.
- Schefzik, R., Thorarinsdottir, T.L., Gneiting, T., et al., 2013. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science* 28, 616–640.
- Scheuerer, M., 2014. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society* 140, 1086–1096.

- Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water Resources Research* 46.
- Schwanenberg, D., Fan, F.M., Naumann, S., Kuwajima, J.I., Montero, R.A., dos Reis, A.A., 2015. Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty. *Water Resources Management* 29, 1635–1651.
- Seo, D.J., Herr, H., Schaake, J., 2006. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions* 3, 1987–2035.
- Sivillo, J.K., Ahlquist, J.E., Toth, Z., 1997. An ensemble forecasting primer. *Weather and Forecasting* 12, 809–818.
- Sloughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Monthly Weather Review* 135, 3209–3220.
- Thomson, D.J., 1994. Jackknifing multiple-window spectra, in: *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on, IEEE*. pp. VI–73.
- Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* 6, 123–137.
- Wang, Q., Robertson, D., Chiew, F., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research* 45.
- Wang, Q., Shrestha, D.L., Robertson, D., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research* 48.
- Wilks, D.S., 2009. Extending logistic regression to provide full-probability-distribution mos forecasts. *Meteorological Applications* 16, 361–368.
- Wu, L., Seo, D.J., Demargne, J., Brown, J.D., Cong, S., Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. *Journal of hydrology* 399, 281–298.
- Zwiefelhofer, W., Kreitz, N., 2001. *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*. World Scientific.

Titre : Contributions statistiques aux prévisions hydrométéorologiques par méthodes d'ensemble.

Mots clés : prévisions d'ensemble, hydrométéorologie, post-traitement statistique, algorithme EM, prévisions probabilistes.

Résumé : Dans cette thèse, nous nous intéressons à la représentation et à la prise en compte des incertitudes dans les systèmes de prévision hydrologique probabilistes à moyen-terme. Ces incertitudes proviennent principalement de deux sources : (1) de l'imperfection des prévisions météorologiques (utilisées en intrant de ces systèmes) et (2) de l'imperfection de la représentation du processus hydrologique par le simulateur pluie-débit (SPQ) (au cœur de ces systèmes).

La performance d'un système de prévision probabiliste s'évalue par la précision de ses prévisions conditionnellement à sa fiabilité. L'approche statistique que nous suivons procure une garantie de fiabilité à condition que les hypothèses qu'elle implique soient réalistes. Nous cherchons de plus à gagner en précision en incorporant des informations auxiliaires.

Nous proposons, pour chacune des sources

d'incertitudes, une méthode permettant cette incorporation : (1) un post-traitement des prévisions météorologiques s'appuyant sur la propriété statistique d'échangeabilité et permettant la prise en compte de plusieurs sources de prévisions, ensemblistes ou déterministes ; (2) un post-traitement hydrologique utilisant les variables d'état des SPQ par le biais d'un modèle Probit arbitrant entre deux régimes hydrologiques interprétables et permettant ainsi de représenter une incertitude à variance hétérogène.

Ces deux méthodes montrent de bonnes capacités d'adaptation aux cas d'application variés fournis par EDF et Hydro-Québec, partenaires et financeurs du projet. Elles présentent de plus un gain en simplicité et en formalisme par rapport aux méthodes opérationnelles tout en montrant des performances similaires.

Title : Statistical contributions to hydrometeorological forecasting from ensemble methods.

Keywords : ensemble forecasts, hydrometeorology, statistical post-processing, EM algorithm, probabilistic forecasts.

Abstract : In this thesis, we are interested in representing and taking into account uncertainties in medium term probabilistic hydrological prediction systems.

These uncertainties mainly come from two sources: (1) from the imperfection of meteorological forecasts (used as inputs to these systems) and (2) from the imperfection of the representation of the hydrological process by the rainfall-runoff simulator (RRS) (at the heart of these systems).

The performance of a probabilistic forecasting system is assessed by the sharpness of its predictions conditional on its reliability. The statistical approach we follow provides a guarantee of reliability if the assumptions it implies are complied with. We are also seeking to incorporate auxiliary information to get sharper.

We propose, for each source of uncertainty, a method enabling this incorporation: (1) a meteorological post-processor based on the statistical property of exchangeability and enabling to take into account several (ensemble or deterministic) forecasts; (2) a hydrological post-processor using the RRS state variables through a Probit model arbitrating between two interpretable hydrological regimes and thus representing an uncertainty with heterogeneous variance.

These two methods demonstrate adaptability on the various application cases provided by EDF and Hydro-Québec, which are partners and funders of the project. Those methods are moreover simpler and more formal than the operational methods while demonstrating similar performances.

