

UNE CONSTRUCTION STATISTIQUE ÉCHANGEABLE POUR LE POST-TRAITEMENT DES ENSEMBLES DE SÉRIES MÉTÉOROLOGIQUES

Éric Parent ¹, Marie Courbariaux ¹ & Pierre Barbillon ¹

¹ *INRA/AgroParisTech, UMR518 MIA, 75231 Paris, France*

Résumé. Les techniques de prévisions probabilistes visent à produire une distribution prédictive de la quantité d'intérêt au lieu d'une seule meilleure estimation ponctuelle. Pour les prévisions de débits en rivière qui intéressent les producteurs d'hydroélectricité tels EDF ou Hydro-Québec une des principales sources d'incertitude est due à la méconnaissance des pluies et températures futures.

Cette incertitude météorologique est aujourd'hui décrite grâce à la génération de plusieurs scénarios de l'évolution à plus ou moins long terme de variables météorologiques, telles la pluie et la température, sur le bassin-versant. Dans le jargon météorologique, ces trajectoires hypothétiques potentielles sont appelées membres d'une prévision d'ensemble. L'incertitude météorologique semblerait donc pouvoir être prise en utilisant tour à tour chaque membre d'ensemble comme un intrant pour le modèle pluie-débit. Cependant, les membres du système de prévision d'ensemble constituent un échantillon souvent biaisé et sous-dispersé d'une prévision probabiliste digne de ce nom.

Pour recalibrer les prévisions d'ensemble, nous développons des modèles fondés sur l'hypothèse d'échangeabilité, une propriété essentielle de tout système de prévision météorologique à base d'ensemble. Ces modèles ont l'avantage de facilement s'adapter aux situations dans lesquelles nous disposons d'ensembles provenant de plusieurs modèles climatiques. Nous appliquons nos modèles à des prévisions d'ensemble de températures et d'occurrence de précipitations.

Notre travail porte sur des séries de prévisions régulièrement émises par deux producteurs d'hydroélectricité en France et au Québec. Nous comparons les résultats de nos constructions statistiques à leurs systèmes actuels de prévision opérationnelle.

Mots-clés. Échangeabilité, prévisions probabilistes, post-traitement, membres d'ensemble météorologique .

Abstract. Probabilistic forecasting aims at producing a predictive distribution of the quantity of interest instead of a single best guess estimate. With regard to river flow forecasts, one of the main sources of uncertainty is due to the unknown future rainfalls and temperatures.

This meteorological uncertainty is nowadays taken into account using members of a meteorological ensemble, i.e. scenarios mimicking the possible multiple trajectories of

the meteorological system on a short to medium range. Such forecasts are considered as inputs to the rainfall-runoff model. However ensemble forecasts are often a biased and under-dispersed sample of a probabilistic prevision.

To recalibrate the forecasts, we develop models based on the hypothesis of exchangeability, a key property when dealing with any ensemble-based forecasting system. Those models have the advantage of being easily adapted to include several sources of ensemble forecasts. We apply them to ensemble forecasts of temperature and precipitation occurrence.

Our work focuses on series of forecasts routinely issued by two hydro-electricity producers in France and in Québec. We finally compare the results of our statistical elaborations to their present operational forecasting systems.

Keywords. Exchangeability, probabilistic forecasts, post-processing, ensemble weather forecasts.

1 Contexte : la prévision hydrologique probabiliste

Afin de produire des prévisions probabilistes de débits en rivière, les gestionnaires d'ouvrages hydrauliques disposent en général de prévisions météorologiques. Une première étape de leur système de prévision consiste à faire le lien entre les prévisions météorologiques dont ils disposent et les grandeurs météorologiques en intrant de leur modèle pluie-débit. Ces grandeurs sont en général des séries de températures et de précipitations (soit absentes, soit en eau ou en neige) moyennées sur le bassin versant d'intérêt.

Les prévisions météorologiques, elles, sont typiquement sous la forme d'un ensemble de K scénarios de une à deux semaines ($K = 50$ membres pour le Centre Européen de Prévision, $K = 20$ membres pour le modèle météorologique régional canadien) que le praticien met au même format que les grandeurs météorologiques à prévoir.

On pourrait tenter de modéliser des séries multivariées (par exemple de températures et de précipitations) mais il serait difficile de représenter leur structure complexe de dépendance. Nous avons donc choisi de procéder par étapes, avec dans un premier temps un modèle pour l'événement binaire "occurrence de précipitations" et un modèle pour la température, à chaque échéance de prévision indépendamment. Il restera à traiter le difficile cas des précipitations (zéro-inflatées) et à tenir compte de la dimension temporelle.

Par la suite, Y désignera la variable à prévoir : soit la température (minimale, maximale ou moyenne) du bassin versant, soit l'occurrence de précipitations ; X désignera le ou les ensemble(s) de prévisions météorologiques associés.

2 Des modèles fondés sur l’hypothèse d’échangeabilité pour traiter l’incertitude météorologique

2.1 Echangeabilité

En perturbant les conditions initiales pour décrire leur incertitude à l’aide de trajectoires possibles générées par leur modèle de la dynamique météorologique du système terre, les fournisseurs d’ensemble météorologique visent à satisfaire une propriété naturellement désirable pour les membres d’un ensemble, à savoir qu’ils soient statistiquement indiscernables.

Du point de vue probabiliste, cette hypothèse d’échangeabilité signifie que l’ensemble généré est tel que sa loi conjointe soit invariante par toute permutation des numéros des membres ($k \in \{1, \dots, K\}$) qui le composent. Un modèle à effets aléatoire du type $[X_t] = \prod_{k=1}^K [X_{t,k}|Z_t] [Z_t]$, où Z est une variable latente et t l’indice du temps, respecte naturellement cette hypothèse constructive.

Réciproquement, le théorème de représentation de DeFinetti-Savage-Hewitt [De Finetti, 1937], montre sous des conditions techniques de régularité, l’existence de ce type de décomposition et de la quantité conditionnante aléatoire Z quand on postule l’échangeabilité quelle que soit la taille de l’ensemble généré.

2.2 Cas binaire : prévisions de l’occurrences de précipitations

Dans le cas où l’événement à prévoir pour le temps t est l’occurrence de précipitations, la variable cible, Y_t , est binaire (occurrence de précipitations / pas de précipitations le jour t sur le bassin versant).

Nous proposons le modèle suivant, fondé sur l’hypothèse d’échangeabilité :

$$\begin{aligned} (X_{k,t}|Z_t) &\sim B(Z_t) \\ Y_t &\sim B(p), p \in [0, 1] \\ (Z_t|Y_t = y) &\sim \beta(\alpha_y, \beta_y) \end{aligned}$$

Alors :

$$\begin{aligned} [Y_t = 1|X_t] &\simeq [Y_t = 1|Z_t = \bar{X}_t], \bar{X}_t = \frac{1}{K} \sum_k X_{k,t} \\ &= \frac{[Z_t = \bar{X}_t|Y_t = 1] [Y_t = 1]}{\sum_{y \in \{0,1\}} [Z_t = \bar{X}_t|Y_t = y] [Y_t = y]} \end{aligned}$$

En pratique, nous étendons ce modèle pour exploiter la dépendance temporelle des événements de précipitations (en supposant que Y est Markovienne).

Dans le cas où l'on dispose de $E > 1$ ensembles de prévisions météorologiques (cas illustré Figure 1 pour $E = 3$), nous indiquons les variables latentes Z_t par l'ensemble considéré, e , et supposons qu'après une transformation appropriée (en l'occurrence une Normal Quantile Transform [Krzysztofowicz, 1997]) les Z_t des E ensembles suivent conjointement une loi normale multivariée.

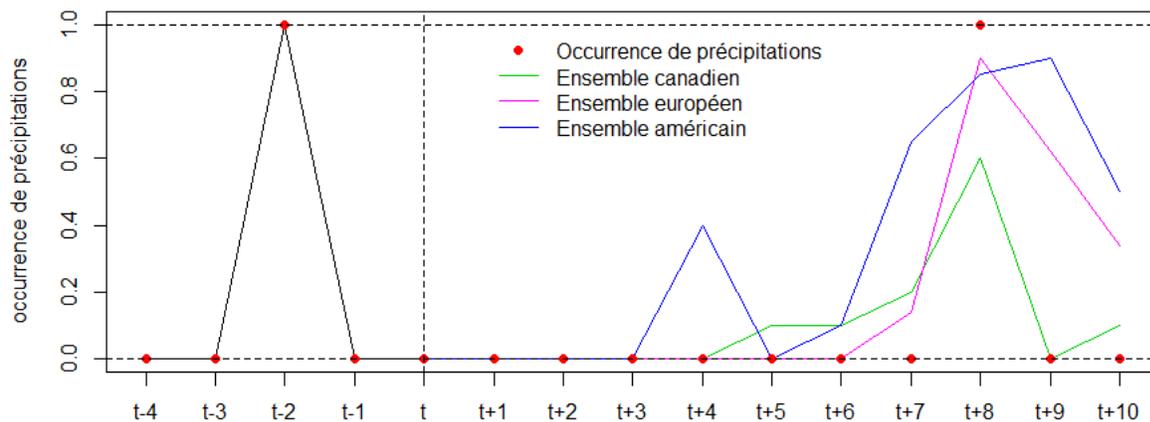


Figure 1: Situation de prévision au 24 février 2012 ($=t$) pour le bassin versant de Baskatong au Québec. Les trajectoires correspondent au nombre de membres non nuls dans les prévisions météorologiques d'ensemble. Le point rouge donne la réalisation de l'événement "occurrence de précipitations", à prévoir.

Les prévisions obtenues avec ce modèle statistique, évaluées par le score de Brier, sont satisfaisantes tant sur les bassins versants québécois que sur les bassins versants français.

2.3 Cas normal : prévisions de températures

Dans le cas où la variable à prévoir est une température sur le bassin versant, la variable cible peut en général être considérée comme gaussienne. Un modèle échangeable normal s'obtient par tirage indépendant de chaque scénario k selon une loi normale conditionnellement à une même grandeur aléatoire latente normale Z . En toute généralité, on peut écrire:

$$(X_{k,t}|Z_t) = a + bZ_t + c\varepsilon_{k,t}, \quad \varepsilon_{k,t} \underset{iid}{\sim} N(0, \omega^2)$$

$$Z_t \underset{iid}{\sim} N(0, \sigma^2)$$

Avec $(a, b, c) \in \mathbb{R}$. Pour relier la série à prédire Y_t et les membres $\{X_{k,t}\}_{k=1:K}$ on s'appuiera sur une indépendance conditionnelle grâce à la variable latente mère Z_t à

travers une relation structurée de même type :

$$Y_t = a_0 + Z_t + \varepsilon_{0,t}, \quad \varepsilon_{0,t} \underset{iid}{\sim} N(0, \omega^2)$$

où $\varepsilon_{0,t} \perp \varepsilon_{k,t} \perp \varepsilon_{k',t} \perp Z_t$, $(a_0 - a)$ est un biais systématique et b et c sont des facteurs de dilatation des variances entre X_t et Y_t qui traduisent les erreurs d'inadéquation des membres à la réalité qu'on cherche à prédire.

Modèle normal-gamma Dans le modèle précédent, la dispersion des membres n'a que peu d'influence sur la variance de la loi conditionnelle prédictive. Or on peut penser que des membres très dispersés indiquent une incertitude plus importante sur la variable à prévoir. Nous proposons donc l'extension suivante du modèle (avec, pour faciliter le calcul des lois conditionnelles, une loi hiérarchique sur les couples (ω_t, Z_t)) :

$$\begin{aligned} (\varepsilon_{k,t} | \omega_t^2) &\underset{iid}{\sim} \mathcal{N}(0, \omega_t^2) \\ (Z_t | \omega_t^2) &\underset{iid}{\sim} \mathcal{N}(0, \lambda \omega_t^2), \quad \lambda \in \mathbb{R}_+ \\ \omega_t^2 &\underset{iid}{\sim} \mathcal{IG}(\alpha, \beta) \end{aligned}$$

Nous proposons de plus une extension de ce modèle au cas où l'on disposerait de $E > 1$ ensembles en posant, pour chaque ensemble e , $(X_{e,k,t} | Z_t) = a_e + b_e Z_t + c_e \varepsilon_{e,k,t}$, avec $(a_e, b_e, c_e)_{e \in \{0, \dots, E\}} \in \mathbb{R}^3$.

L'inférence de tels modèles se fait classiquement avec un algorithme Expectation-Maximization.

Les résultats ainsi obtenus pour la prévision de la température sont évalués avec les scores ignorance et CRPS. Ils sont satisfaisants, tant sur les bassins français que sur les bassins québécois.

3 Conclusion et discussion

L'hypothèse d'échangeabilité nous a ainsi permis de proposer des modèles parcimonieux pour émettre des prévisions probabilistes à partir de prévisions d'ensemble. Nous avons de plus pu naturellement les étendre pour utiliser plusieurs sources de prévisions d'ensemble simultanément.

Ces modèles se montrent performants pour émettre des prévisions probabilistes dans un cas gaussien (prévisions de températures) et dans un cas binaire (prévision de l'occurrence de précipitations) pour un jour et un bassin versant cibles. Une des grandeurs météorologiques d'intérêt pour les praticiens est toutefois d'un type mixte : les précipitations totales sur le bassin versant. Le caractère mixte de cette quantité la rend particulièrement difficile à modéliser.

Par ailleurs, nous ne nous sommes ici intéressés qu'à un bassin versant, une grandeur météorologique et à une échéance de prévision à la fois, alors que nous souhaitons à terme émettre des prévisions multivariées. Une solution prisée dans le domaine des prévisions hydrologiques consiste à faire l'usage de copules empiriques ([Scheffzik et al., 2013, Clark et al., 2004]). De telles copules copient la structure de rang d'un ensemble de prévisions et la transposent à des simulations selon les distributions prédictives marginales. On obtient ainsi des scénarios (dans le temps, dans l'espace et multi-grandeurs) ayant une structure multivariée crédible mais n'exploitant pas à proprement parler les dépendances ainsi illustrées.

4 Remerciements

Ce travail fait l'objet du contrat de recherche Adeprina/AgroParisTech 694R financé par EDF et Hydro-Québec. Les réflexions méthodologiques de cette communication ont été alimentées par les échanges soutenus avec les cadres de leurs services de recherche-développement et de prévision, notamment Rémy Garçon, Luc Perreault et Joël Gailhard.

References

- [Clark et al., 2004] Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R. (2004). The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1):243–262.
- [De Finetti, 1937] De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68.
- [Krzysztofowicz, 1997] Krzysztofowicz, R. (1997). Transformation and normalization of variates with specified distributions. *Journal of Hydrology*, 197(1):286–292.
- [Scheffzik et al., 2013] Scheffzik, R., Thorarinsdottir, T. L., Gneiting, T., et al. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28(4):616–640.