

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, les principales sources d'incertitude sont dues (a) à la méconnaissance des pluies et températures futures (incertitude météorologique), (b) aux erreurs de représentation de la transformation pluie-débit (incertitude hydrologique). Il faut d'abord modéliser séparément ces sources d'incertitudes avant de les intégrer pour obtenir une fonction prédictive de densité de probabilité. Cette communication focalise sur la modélisation de l'incertitude météorologique. (L' incertitude hydrologique est décrite dans une autre communication des mêmes auteurs à ces journées.)

L'incertitude météorologique est aujourd'hui décrite grâce à la génération de plusieurs scénarios de l'évolution à plus ou moins long terme des variables météorologiques généralement localisées, telles la pluie et la température sur un bassin-versant. Dans le jargon météorologique, ces trajectoires hypothétiques potentielles sont appelés 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 le post-traitement visant à recalibrer la prévision d'ensemble, nous développons un modèle fondé sur l'hypothèse d'échangeabilité, une propriété essentielle de tout système de prévision météorologique à base d'ensemble.

Notre travail porte sur des séries de prévisions de flux de la rivière régulièrement émises par deux producteurs d'hydroélectricité en France et au Québec. Nous comparons les résultats de nos constructions statistiques à leurs systèmes actuels empiriques 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 point-wise estimate.

With regard to river flow forecasts, the main sources of uncertainty are due (a) to the unknown future rainfalls and temperatures (input uncertainty), (b) to the inadequacy of

the deterministic model mimicking the rainfall-runoff transformation (hydrological uncertainty). We model them separately and then integrate the input uncertainty with the hydrologic uncertainty into the total uncertainty, which is quantified in terms of a predictive density function. This communication will focus on meteorological uncertainty. (A companion paper deals with hydrological uncertainty.)

Input 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 underdispersed sample of a probabilistic prevision. To postprocess ensemble members in order to recalibrate the forecasts, we develop a model based on the hypothesis of exchangeability, a key property when dealing with any ensemble based forecasting system.

Our work focuses on series of river flow forecasts routinely issued by two hydroelectricity 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, meteorological ensemble members.

1 La prévision probabiliste des débits

1.1 Structure du problème

Pour mettre en place un système de prévisions probabiliste de débits en rivière, l'analyste dispose de bases de données sur les divers bassins-versants qui intéressent les producteurs d'hydroélectricité tels EDF ou Hydro-Québec. Par bassin-versant, ces données sont typiquement les enregistrements de la réalisation de trois grandeurs aléatoires Q, P, A :

- Q désigne le débit de la rivière au site d'intérêt, pour une période de temps fixée.
- P représente les grandeurs météorologiques (pluie, température, évapotranspiration, etc) correspondantes, à un éventuel décalage temporel près (temps de concentration du bassin versant). Elles sont "génératrices" du débit de la rivière. Les services de recherche-développement de ces sociétés utilisent une transformation (pluie/débit) déterministe $\mathcal{M}(\cdot)$ permettant de passer de la météorologie P à l'hydrologie Q ; ne remettant pas en cause la pertinence de la transformation pluie/débit \mathcal{M} adoptée, on admet que cette transformation est *efficace* , c'est-à-dire que $Q|P = Q|\mathcal{M}(P)$
- A représente les anticipations météorologiques. Elles sont en général données sous la forme d'un ensemble de scénarios météorologiques (membres de prévisions) : on peut obtenir cet ensemble de membres avant que P , et donc Q , ne soient connus.

1.2 Deux sources principales d'incertitudes à intégrer

En situation opérationnelle, on ne dispose que des anticipations météorologiques A pour prévoir le débit Q sur un horizon fixé. Construire une prévision probabiliste, c'est donc fournir la loi prédictive $[Q|A]$, la notation crochets $[.]$ de Gelfand et Smith [1] désignant ici la distribution conditionnelle de probabilité. En suivant les règles du calcul des probabilités pour associer les principales sources d'incertitudes :

$$\begin{aligned} [Q|A] &= \int_P [Q, P|A] dP \\ [Q|A] &= \int_P [Q|P, A] \times [P|A] dP \end{aligned}$$

On fait l'hypothèse additionnelle réaliste que $[Q|P, A] = [Q|P]$, c'est à dire que P (la météo réelle) est *suffisante* pour prévoir Q .

Il s'agit alors d'effectuer la composition des incertitudes hydrologique $Q|\mathcal{M}(P)$ et météorologique $P|A$, déjà décrite par Krzysztofowicz [2] :

$$[Q|A] = \int_P [Q|\mathcal{M}(P)] \times [P|A] dP \quad (1)$$

Cette communication focalise sur la modélisation de l'incertitude météorologique. L'incertitude hydrologique est traitée dans l'autre communication des auteurs à ces journées.

2 Expression de l'incertitude météorologique

2.1 Passer dans le domaine des débits

Comment construire un modèle conjoint $[P, A]$ pour nourrir l'équation 1? Ce que recouvre le terme générique P est une variable complexe, constituée de la série multivariée des variables météorologiques servant de point d'entrée à la transformation pluie/débit $\mathcal{M}(\cdot)$: en particulier, la température et la pluie (soit absente, soit en eau ou en neige). Typiquement A est un ensemble de scénarios (50 membres pour le Centre Européen de Prévision, 21 pour le modèle météorologique régional canadien) de même structure que P . On pourrait tenter de modéliser des séries multivariées (par exemple température et pluie) mais il sera très difficile de représenter leur structure complexe de dépendance et les occurrences zéro-inflatées dues aux jours sans pluie. Nous avons donc choisi de passer dans un domaine ayant une structure analogue à une série temporelle hydrologique où l'on espère pouvoir travailler avec des modèles probabilistes plus simples et parcimonieux, du genre loi normale multivariée. En appelant $\mathcal{M}(A)$ l'ensemble de scénarios de débits générés terme à terme par chacun des scénarios météorologiques de A , on s'appuie sur l'idée que

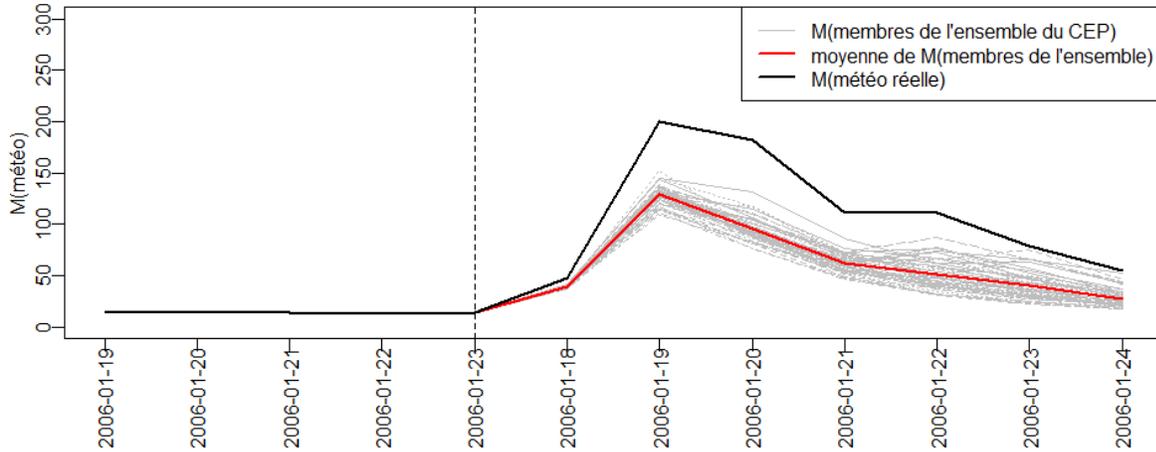


Figure 1: Station de l’Ain Vouglans : 50 membres d’un ensemble météorologique (faisceau en gris) pour représenter les incertitudes portant sur la prévision de la série des débits modélisés par la transformation pluie-débit \mathcal{M} (ligne continue noire).

la transformation $\mathcal{M}(\cdot)$ étant déterministe, chaque événement aléatoire ($P = p, A = a$) est en correspondance immédiate avec l’événement ($\mathcal{M}(P) = \mathcal{M}(p), \mathcal{M}(A) = \mathcal{M}(a)$). La figure 1 représente la série $\mathcal{M}(A)$ (faisceau de 50 membres en gris) pour la station de l’Ain Vouglans. Il s’agit d’anticipations des débits journaliers sur un horizon de 7 jours suivant le 23 janvier 2006. La série $\mathcal{M}(P)$ reportée en ligne continue noire, ne pourra bien sûr n’être calculée qu’à la fin de cette période, puisqu’elle est obtenue en nourrissant la transformation pluie-débit \mathcal{M} avec les entrées météo P effectives de la période du 12 au 18 janvier à prédire.

2.2 Notations

L’image de P par la transformation pluie/débit est une série temporelle $\{\mathcal{S}_t^*\}_{t=1:h} = \mathcal{M}(P)$, tandis que l’image de A est un ensemble de $K = 50$ séries temporelles $\{S_{k,t}\}_{t=1:h;k=1:K} = \mathcal{M}(A)$ de la même dimension, avec l’index k pour désigner son numéro de membre et l’index t pour désigner le jour.

2.3 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 indis-

cernables. 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 qui le composent.

Un modèle à effets aléatoire du type $\{S_{k,1:h}\} = \prod_{k=1}^K [S_{k,1:h}|L_t][L_t]$ respecte naturellement cette hypothèse constructive. Réciproquement, le théorème de représentation de DeFinetti-Savage-Hewitt [3], montre sous des conditions techniques de régularité, l'existence de ce type de décomposition et de la quantité conditionnante aléatoire L quand on postule l'échangeabilité quelle que soit la taille de l'ensemble généré.

2.4 Modèle et Inférence

Un modèle échangeable simple s'obtient par tirage indépendant de chaque scénario k selon une loi normale multivariée à h dimensions conditionnellement à une même grandeur aléatoire latente normale L préalablement générée, par exemple elle aussi normale.

$$\begin{aligned} L_{1:h} &\sim N_h(\mu_{1:h}, \Sigma_{h \times h}) \\ S_{k,1:h} &\underset{iid}{\sim} N_h(L_{1:h}, \Omega_{h \times h}) \end{aligned} \quad (2)$$

Un modèle parcimonieux reliant la série à prédire $S_{1:h}^*$ et les membres $\{S_{k,t}\}_{t=1:h; k=1:K}$ s'appuiera sur une indépendance conditionnelle grâce à la variable latente mère L à travers une relation structurée de même type:

$$S_{1:h}^* \sim N_h(A_{h \times h}^* L_{1:h} + B_{1:h}^*, \Omega_{h \times h}^*) \quad (3)$$

Un modèle encore plus simple fait les hypothèses additionnelles $A_{h \times h}^*$ et $\Sigma_{h \times h}$ diagonales (la série pivot latente $L_{1:h}$ n'est pas autocorrélée, l'espérance conditionnelle de S^* s'exprime marginalement).

En supposant que les paramètres restent identiques sur une période d'apprentissage proche de la période de prédiction, un algorithme EM effectue l'estimation de ce modèle normal multivarié à effet aléatoire.

Pour la prédiction (cf. figure 2), on calculera facilement la loi normale multivariée $[S^*|S] = \int_L [S^*|L][L|S]dL$ grâce à l'indépendance conditionnelle $[S^*|S, L] = [S^*|L]$.

2.5 Discussion et conclusion

D'autres modèles sont possibles : par exemple supposer une même structure de corrélation pour $\Omega_{h \times h}$ et $\Omega_{h \times h}^*$, ou imaginer que la corrélation conditionnelle entre $S_{k,t}|L$ et $S_{k,t+\Delta t}|L$ diminue avec l'écart Δt mais qu'*a contrario* la variance conditionnelle de $S_{k,t}|L$ augmente avec le temps t . Rappelons néanmoins que l'objectif de cette modélisation est de

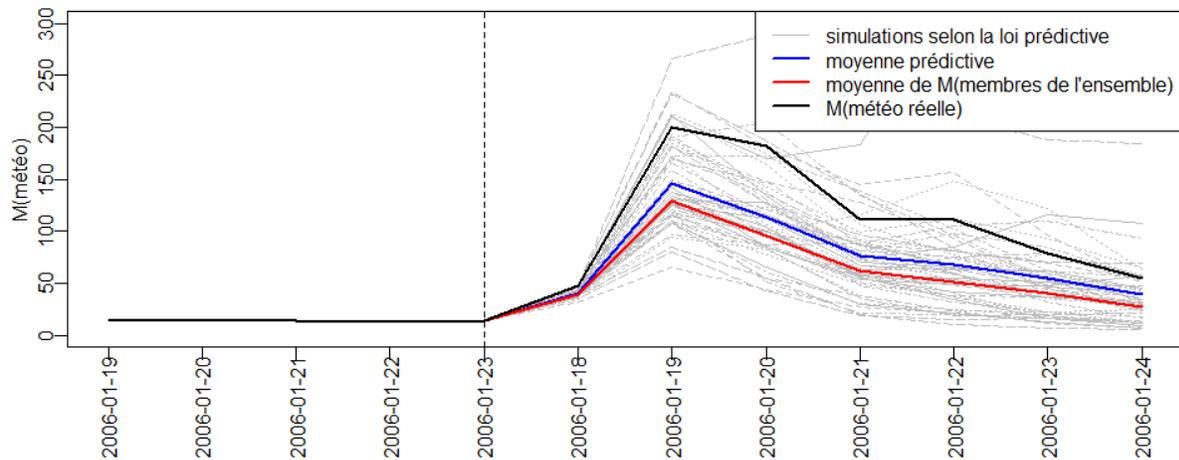


Figure 2: Prédiction probabiliste du débit modélisé S^* (ligne continue noire) grâce un échantillon de la loi conditionnelle sachant l' ensemble météorologique S . La moyenne prédictive est en bleue, la moyenne des membres en rouge

fournir la brique conceptuelle représentant l'incertitude météorologique pour l'évaluation de l'incertitude totale via l'équation 1 : un modèle n'est jamais "vrai", même si on s'appuie sur une propriété-clé (ici l'échangeabilité cf eq. 2) pour le construire, tout au plus peut-il être utile...

2.6 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.

Bibliographie

- [1] Gelfand, A.E. and Smith, A.F.M. (1990), Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, 398–409.
- [2] Krzysztofowicz, R. (2001), The case for probabilistic forecasting in hydrology, *Journal of Hydrology*, 249, 2.
- [3] Hewitt, E. and Savage, L. J. (1955), Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80, 470–501.