Construction bayésienne de prévisions probabilistes à partir des sorties d'un modèle déterministe pluie-débit

Marie Courbariaux[‡] , Éric Parent[‡] & Pierre Barbillon[‡]

[‡] *INRA/AgroParisTech, UMR518 MIA, 75231 Paris, France*
{marie.courbariaux,eric.parent,pierre.barbillon}@agroparistech.fr

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 hydrologique (l' incertitude météorologique est décrite dans une autre communication des mêmes auteurs à ces journées).

Pour quantifier l'incertitude hydrologique, un modèle conjoint de la série des débits modélisés à partir de la pluie et des débits observés est développé. Nous nous appuyons sur une construction bayésienne : après modélisation (normale sur variables transformées) du comportement *a priori* du régime naturel des débits, la prévision est mise à jour en tenant compte de l'information véhiculée par les sorties de la transformation pluie-débit. On cherche ensuite à améliorer la modélisation autorégressive des erreurs de la transformation pluie-débit en introduisant deux régimes de fonctionnement commandés par des variables explicatives selon un modèle Probit, estimé par l'algorithme EM.

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. algorithme EM, erreur de modèle, hydrologie, modélisation bayésienne, prévisions probabilistes, modèle pluie-débit.

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 hydrological uncertainty. (A companion paper deals with meteorological uncertainty.)

To quantify hydrological uncertainty, a joint model for the time series of rainfall-runoff errors along the whole trajectory has to be developped. We rely on a Bayesian approach, first considering the prior behavior of the natural regime of river flows and then updating their predictions by taking into account the likelihood of the information conveyed through the outputs of the rainfall-runoff transformation. We then try to improve the autoregressive model of errors for the rainfall-runoff transformation by introducing two regimes controlled by explanatory variables through a Probit model to be estimated by the EM algorithm.

Our work focuses on series of river flow 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. Bayesian modelling, EM algorithm, hydrology, model uncertainty, probabilistic forecasts, rainfall-runoff model.

1 La prévision probabiliste des débits

Structure du problème

Pour mettre en place un système de prévisions probabiliste de débits en rivière, l'analyste dispose d'une base de données sur les divers bassins-versants qui intéressent les producteurs d'hydroélectricité tels EDF ou Hydro-Québec. Pour un bassin-versant donné, 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 $\mathcal{M}(.)$ 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.

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 & Smith (1990) désignant ici la distribution conditionnelle de probabilité. En suivant les règles du calcul des probabilités pour associer les principales sources d'incertitudes :

$$[Q|A] = \int\limits_{P} [Q, P|A] dP$$

$$[Q|A] = \int_{P} [Q|P, A] \times [P|A] dP$$

On fait l'hypothèse additionnelle assez 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 :

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

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

2 Expression de l'incertitude hydrologique

Comme dans le Bayesian system for probabilistic river stage forecasting de Krzysztofowicz (2002), nous nous appuyons sur une construction bayésienne.

Celle-ci consiste à modéliser dans un premier temps le comportement *a priori* du régime naturel de la rivière puis à mettre à jour la prévision résultante avec les informations véhiculées par les sorties du modèle pluie-débit.

Concrètement, on obtient la loi recherchée, $[Q|\mathcal{M}(P)]$, en mettant à jour la loi *a priori* des débits futurs,$[Q]$, par la vraisemblance des sorties du modèle pluie-débit, $[\mathcal{M}(P)|Q]$, via la formule de Bayes :

$$[Q|\mathcal{M}(P)] \propto [Q] [\mathcal{M}(P)|Q]$$

Cette approche en deux temps (par rapport à une approche donnant directement $[Q|\mathcal{M}(P)]$) nous donne une plus grande flexibilité de modélisation. Nous exploitons de plus la totalité de l'historique climatologique des débits, et non uniquement la période pour laquelle on dispose d'observations conjointes de $\mathcal{M}(P)$ et de Q. Enfin, en partant de la loi marginale de Q, nous offrons une garantie de calibration de par la façon même dont elle se mesure.

Normalisation et stationnarisation des séries temporelles

Les modèles pour lesquels on dispose du plus d'outils d'analyse et d'inférence reposent sur des hypothèses de normalité non valables dans notre cas. Afin de nous y ramener, nous avons recours à une transformation Box-Cox de paramètre λ. Une analyse préliminaire montre que la valeur optimale de λ n'est pas significativement différente de zéro. Nous choisissons donc $\lambda \to 0$, ce qui revient à simplement effectuer un passage au logarithme.

Afin de stationnariser et de désaisonnaliser nos séries temporelles, nous estimons ensuite moyenne et variance des variables transformées pour chaque jour calendaire puis centrons et réduisons.

Dans la suite de ce document, Y_t et X_t désignent respectivement le débit normalisé-centré-réduit au jour t et la sortie du modèle pluie-débit transformée correspondante.

Soit H l'horizon de prévision. Au temps t, nous cherchons la loi de Y au cours des H jours qui suivent conditionnellement à la connaissance de X pour ces H jours et à la connaissance du passé :

$$[Y_{t+1:H}|X_{t+1:H}, X_{\to t}, Y_{\to t}] \propto [X_{t+1:H}|Y_{t+1:H}, X_{\to t}, Y_{\to t}]\,[Y_{t+1:H}|X_{\to t}, Y_{\to t}].$$

Des modèles auto-régressifs pour la climatologie des débits et pour l'erreur de transformation pluie-débit

Une analyse préliminaire de notre jeu de données montre qu'un modèle markovien d'ordre 3 pourrait convenir dans le cas de la loi marginale des débits. On considère donc le modèle suivant :

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \rho_3 Y_{t-3} + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0, \sigma^2),\ \varepsilon_t \perp \varepsilon_{t'}\ \forall t, t'$$

Dans le cas de l'erreur de transformation, nous considérons un modèle linéaire avec persistance de l'erreur :

$$X_t = aX_{t-1} + a_1 Y_t + a_2 Y_{t-1} + \eta_t,\ \eta_t \sim \mathcal{N}(0, v^2),\ \eta_t \perp \varepsilon_{t'} \text{ et } \eta_t \perp \eta_{t'}\ \forall t, t'$$

Inférence et prévisions

Les paramètres de nos modèles sont estimés par maximum de vraisemblance pour chaque jour calendaire.

On obtient facilement une expression analytique de $[Y_{t+1:H}|X_{\to t}, Y_{\to t}]$ en fonction de $Y_{\to t}$ et de $[X_{t+1:H}|Y_{t+1:H}, X_{\to t}, Y_{\to t}]$ en fonction de $Y_{t+1:H}, X_{\to t}$ et de $Y_{\to t}$. À l'aide des formules de conjugaison normale (ou, de façon équivalente, du filtre de Kalman), on en déduit une expression analytique de $[Y_{t+1:H}|X_{t+1:H}, X_{\to t}, Y_{\to t}]$.

La figure 1 montre un exemple d'application de la méthode décrite ci-dessus sur des débits, à l'exutoire du bassin versant français de l'Ain à Vouglans. Il s'agit des prévisions émises le 1er avril 2008 pour les 30 jours qui suivent, dont 14 pour lesquels on dispose de prévisions issues du modèles conceptuel.

La surface orangée correspond à l'intervalle prédictif à 90%.

Les trois trajectoires en traits fins et oranges sont des exemples de trajectoires conduisant aux quantiles 0.05, 0.5 et 0.95 des volumes prédictifs à horizon 30 jours.

La surface grisée correspond à l'intervalle de confiance à 90% dans le cas où les prévisions seraient émises à partir d'un simple modèle auto-régressif (sans utiliser les sorties du modèle conceptuel pluie-débit).

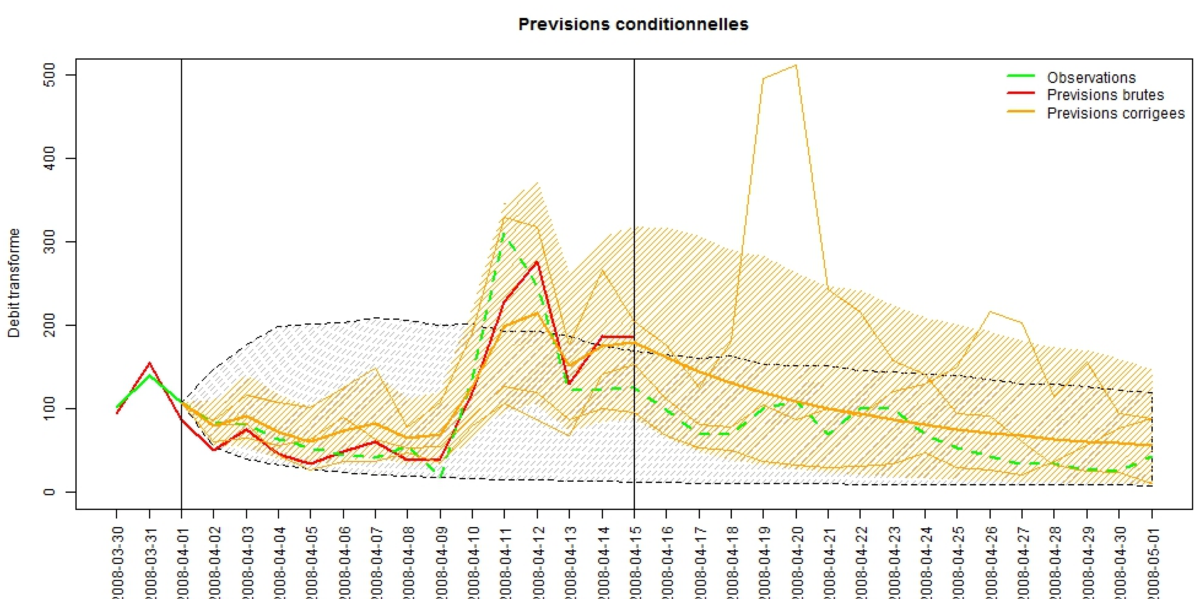

Figure 1 – Résultat des prévisions par filtre de Kalman - 1er avril 2008, Ain à Vouglans

3 Perspectives

Une façon de raffiner le modèle d'erreur consiste à identifier des régimes de fonctionnement du modèle pluie-débit pour lesquels la nature de l'erreur diffère.

En nous inspirant de Furrer et al. (2006), Perreault et al. (2007) et de Lu & Berliner (1999), nous proposons un modèle de mélange de type Probit dont la classification repose sur les variables d'état internes du modèle conceptuel pluie-débit (taux de remplissage des réservoirs internes et taux d'apport en eau provenant de la fonte par exemple) :

$$\begin{cases} (X_t|S_t = s) & = a_s X_{t-1} + a_{1,s} Y_t + a_{2,s} Y_{t-1} + \eta_{s,t},\ \eta_{s,t} \sim \mathcal{N}(0, {v_s}^2) \\ S_t & = \mathbf{1}_{Z_t<0} \\ Z_t & \sim \mathcal{N}\left(\beta^T E_t, 1\right) \end{cases}$$

Où $\eta_{s,t} \perp \varepsilon_{t'}$ et $\eta_{s,t} \perp \eta_{s',t'}$ $\forall t, t', s, s'$ et E_t est un vecteur de variables d'état.

L'inférence du modèle se fait avec l'algorithme Expectation Maximization. Un échantillonneur de Gibbs nous donne des simulations selon la loi prédictive.

Remerciements

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

Références

Furrer, E. M., Jacques, C. & Favre, A.-C. (2006), Short term discharge prediction using a markovian regime switching model, technical report, INRS-ETE.

Gelfand, A. E. & Smith, A. F. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American statistical association* **85**(410), 398–409.

Krzysztofowicz, R. (2002), 'Bayesian system for probabilistic river stage forecasting', *Journal of Hydrology* **268**(1), 16–40.

Lu, Z.-Q. & Berliner, L. M. (1999), 'Markov switching time series models with application to a daily runoff series', *Water Resources Research* **35**(2), 523–534.

Perreault, L., Garçon, R. & 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.