Incertitudes météorologique et hydrologique : investigation bayésienne sur la prévision probabiliste des débits (pour la production hydro-électrique)

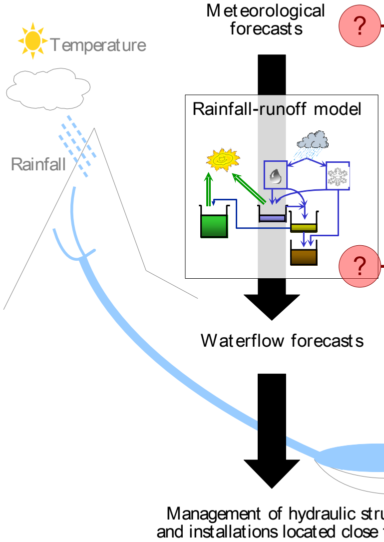
Éric Parent & Marie Courbariaux

INRA/AgroParisTech, UMR518 MIA, 75231 Paris, France

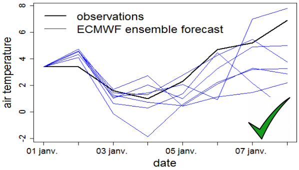
Partenaires :P. Barbillon (AgroParisTech), A.-C. Favre (LTHE), L. Perreault (IREQ), J. Gailhard et R. Garçon(EDF-DTG)

AppliBUGS, 28 novembre 2014

Les prévisions de débits : deux sources d'incertitudes



Uncertainty about future temperature and rainfall → ensemble forecasts

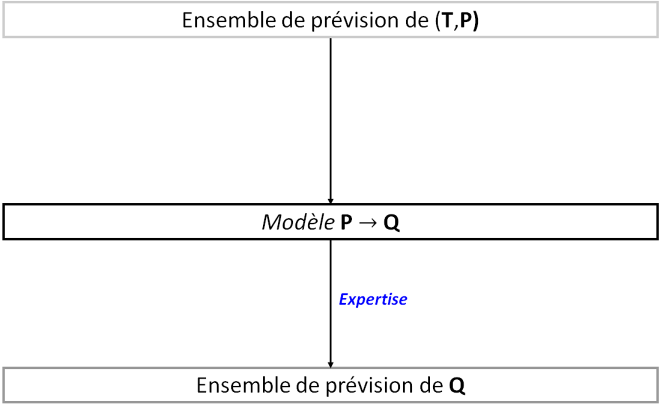


Limited capacity of the model to mimic physical processes

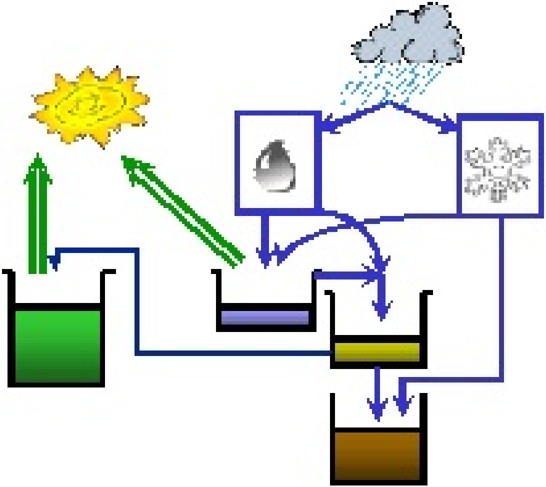
→ how should we take into account the resulting uncertainty ?

Images source : Matthieu Le Lay, EDF-DTG

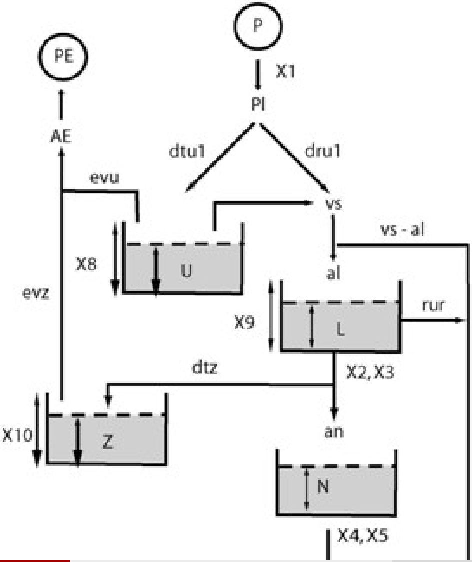
Structure Idéale...



Un modèle conceptuel pluie-débit : MORDOR



Modèle conceptuel pluie-débit $P \mapsto \mathcal{M}(P)$



Convolution des 2 sources d'incertitude météo et hydro

En suivant les règles du calcul des probabilités :

$$[Q|A] = \int_P [Q, P|A] dP$$

$$[Q|A] = \int_P [Q|P, A] \times [P|A] dP$$

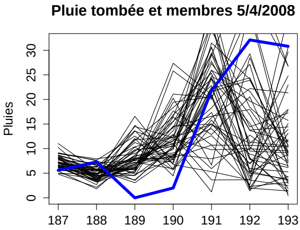
Hypothèses additionnelles

- $[Q|P, A] = [Q|P]$: P (la météo réelle) est *suffisante* pour prévoir Q (sans avoir besoin de A, l'ensemble des scénarios météos prévus)
- $[Q|P] = [Q|\mathcal{M}(P)]$: Le modèle de transformation pluie débit $P \mapsto \mathcal{M}(P)$ est *efficace*

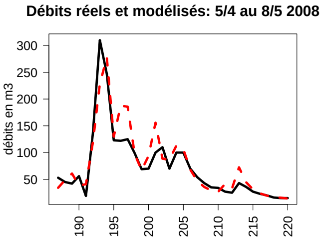
Composition des incertitudes hydrologiques $Q|\mathcal{M}(P)$ et météorologiques $P|A$:

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

Composition des 2 sources d'incertitude

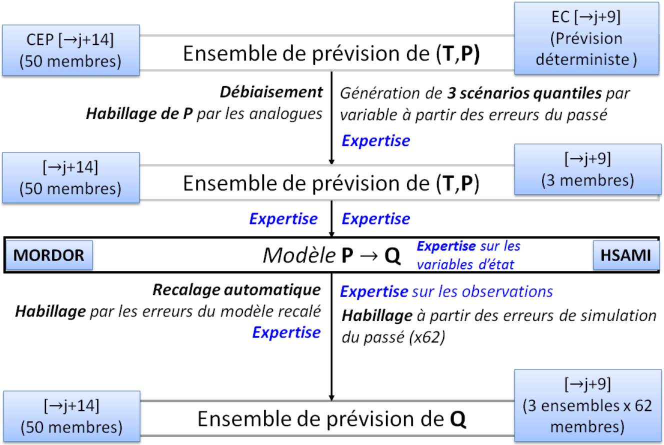


Incertitude Météo $[P|A]$

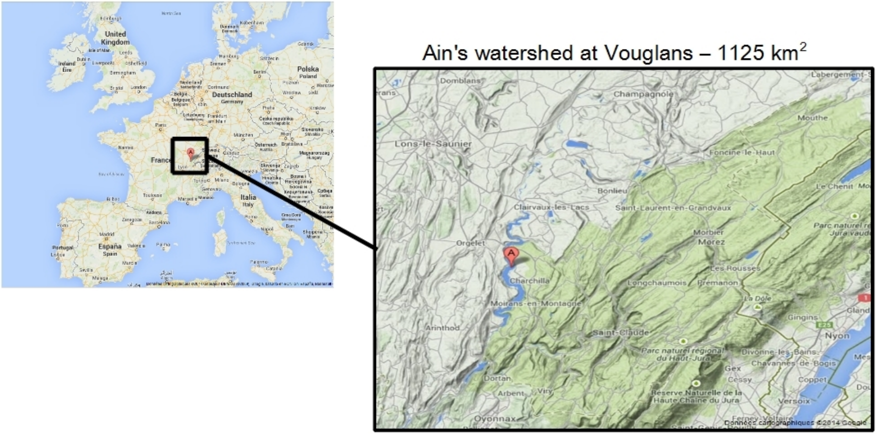


Incertitude Hydro $[Q|\mathcal{M}(P)]$

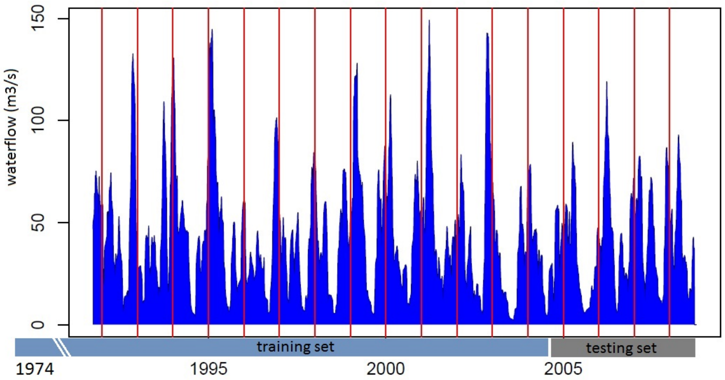
Le systeme mis en place à EDF-DTG & HQ



Bassin de l'Ain à Vouglans



Données disponibles



- Débits observés
- Débits modélisés et variables d'état correspondantes

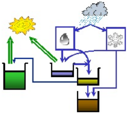
On teste les méthodes sur ces données en effectuant des prévisions à 7 jours.

Focus sur l'incertitude Hydro $[Q|\mathcal{M}(P)]$

Nous souhaitons produire la fonction de répartition prédictive (multivariée) pour :

$$Q_{1\to 7} \;\Big|\; \underbrace{\mathcal{M}(P_{1\to 7}) \quad , \quad (\mathcal{M}(P), Q)_{hist} \quad , \quad VE_{hist\to 7}}_{\mathcal{I}}$$

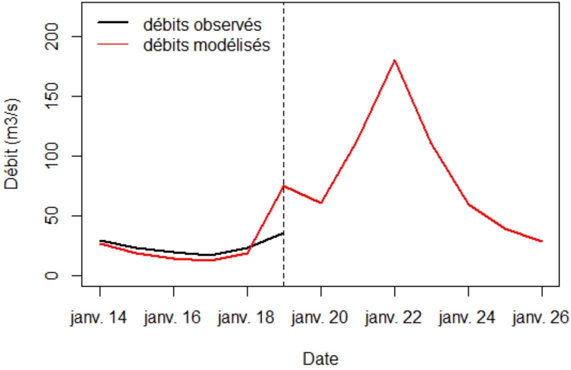
$Q_{1\to 7}$	$\mathcal{M}(P_{1\to 7})$	$(\mathcal{M}(P), Q)_{hist}$	$VE_{hist\to 7}$
Débit des 7 prochains jours	Débit modélisé pour les 7 prochains jours	Débits prévus et modélisés dans le passé	Covariables du modèle pluie-débit



- quantité d'eau stockée dans des réservoirs fictifs
- quantité de neige
- proportion des apports venant de la fonte de la neige
- ...

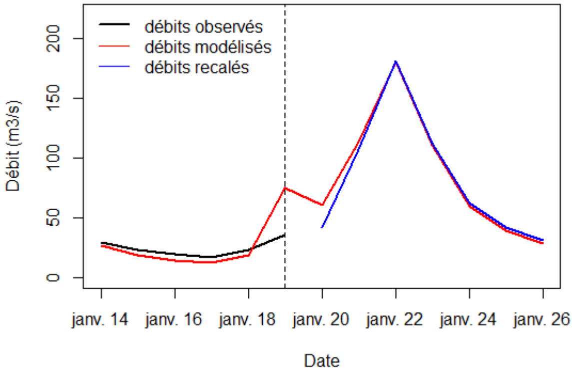
La méthode opérationnelle chez EDF

Etape 0 : trajectoire de débit prévue sans post-traitement (prévision déterministe)



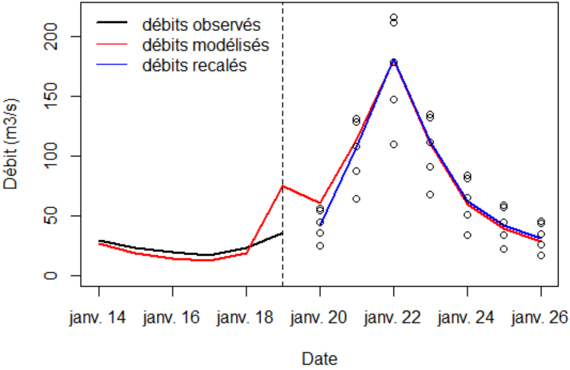
La méthode opérationnelle chez EDF

Etape 1 : trajectoire de débit prévue après recalage (assimilation de données) (prévision déterministe)



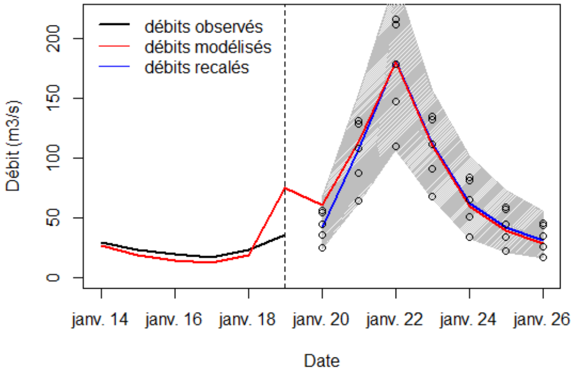
La méthode opérationnelle chez EDF

Etape 2 : habillage des prévisions (1 densité prédictive par échéance de prévision)



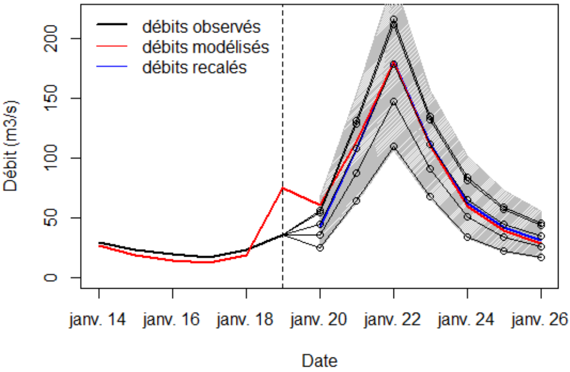
La méthode opérationnelle chez EDF

Etape 2 : habillage de la prévision (1 densité prédictive par échéance de prévision)



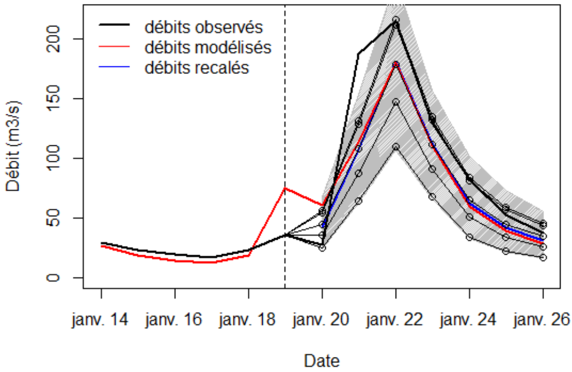
La méthode opérationnelle chez EDF

Etape 3 : construction de trajectoires à partir des points de l'habillage (densité prédictive multivariée)



La méthode opérationnelle chez EDF

Pb : Représentation de trajectoires temporellement cohérentes



Approche de série temporelle

- **Une seule étape de modélisation multivariée**
- Un modèle avec une classification inspirée par l'expertise d'EDF
- Modèle de base similaire à celui de l'Ensemble Streamflow Prediction System de la NWS

Décomposer la loi multivariée en produit de conditionnelles

On peut écrire :

$$[Q_{1\to 7}|\mathcal{I}] = \prod_{h=1}^{h=7}[Q_h|Q_{1\to h-1},\mathcal{I}]$$

Par conséquent, pour tirer des scénarios dans $[Q_{1\to 7}|\mathcal{I}]$:

(for h in 1, ..., 7){
tirer q_h selon $[Q_h|\mathcal{I}, q_1 \to q_{h-1}]$
}

...Nous n'avons ainsi que besoin d'un modèle conditionnel pour $(Q_h|Q_{1\to h-1},\mathcal{I})$

Un modèle ARX avec classification

Proposition de modélisation :

$$VE_{hist \to 7} \to classe_h$$

$$(Y_h | Y_{1 \to h-1}, \mathcal{I}, classe_{\mathrm{h}} = c) = \mu_c + \beta_c Y_{h-3 \to h-1} + \gamma_c X_{h-1 \to h+1} + \sigma_c \varepsilon_h$$

$\varepsilon_h \sim N(0,1), \quad \varepsilon_{t'} \perp \varepsilon_t$
$Y = ln(Q)$
$X = ln(Q_{mod})$

$\Rightarrow$ Gain en cohérence & simplicité par rapport à la méthode opérationnelle actuelle
$\Rightarrow$ Pb : Quid quand plus de conditionnement ? Calibration ? Raccord sur le comportement "climatique" ?

Modélisation directe ARX $[Q|P]$ et calibration

Le modèle probabiliste ARX est *en principe* une loi *conditionnelle* $[Q|P]$.
Vérifie-t-elle la bonne calibration ? c-a-d :

$$
\begin{aligned}
[Q] &= \int_P [Q, P] dP \\
&= \int_P [Q|P][P] dP \\
&\approx \frac{1}{G} \sum_{p_g \sim [P]} [Q|p_g]
\end{aligned}
$$

les G tirages g étant, en pratique, tout ou partie de l'échantillon disponible de P.

$\Rightarrow$ Aucune raison *théorique* d'être calibré : la météo P est subie ! On ne peut que vérifier, après le choix de $[Q|P]$, si la discrépance entre $[Q]$ et $\frac{1}{G} \sum_{p_g \sim [P]} [Q|p_g]$ est acceptable ou non !

Idée

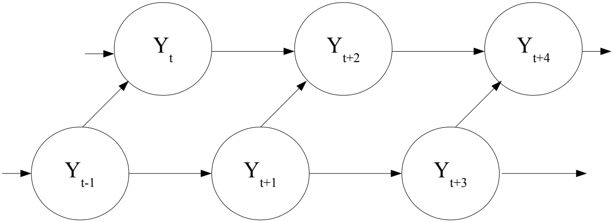
Formule de Bayes :

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

Interprétation phénoménologique :

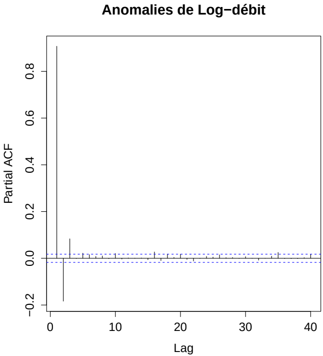
1. $[Q]$ = régime naturel de la rivière : *a priori*, comment les débits évoluent-ils d'un pas de temps au suivant ?
2. $\mathcal{M}(P)$, le débit imparfait fourni par le modèle pluies/débits est une *information* apportée sur l'état naturel du système, ici Q, le *vrai* débit. $[\mathcal{M}(P)|Q]$ =*vraisemblance* du modèle statistique ! La nature Q (inconnue) délivre une information $\mathcal{M}(P)$ (observable qui aidera à la caractériser).
3. Bayes= assimilation de l'information $\mathcal{M}(P)$ pour de produire la loi prédictive $[Q|P]$ en un temps où Q n'est pas connu ?

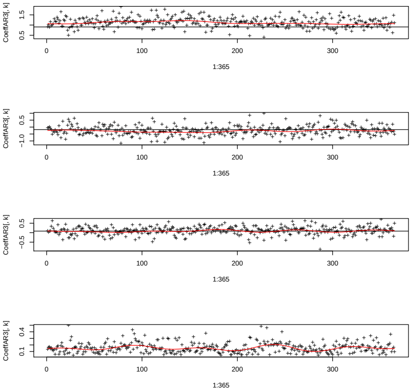
Un modèle marginal $[Q]$simple reliant les débits avec une mémoire à deux pas de temps



$$Y = log(Q)$$

L'Ain à Vouglans : graphe des autocorrelations partielles sur 34 années de débits transformés

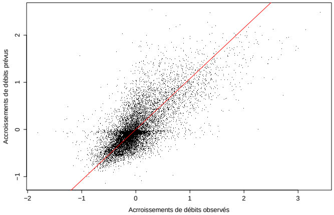


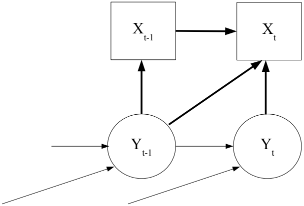


$$Y_{t+3} = \rho_{1,t} Y_{t+2} + \rho_{2,t} Y_{t+1} + \rho_{3,t} Y_t + \sigma_t \epsilon_t$$

$$Y = log(Q),\ \epsilon_t \sim N(0,1)$$

L'Ain à Vouglans : graphe des accroissements de débits transformés prévus *versus* observés

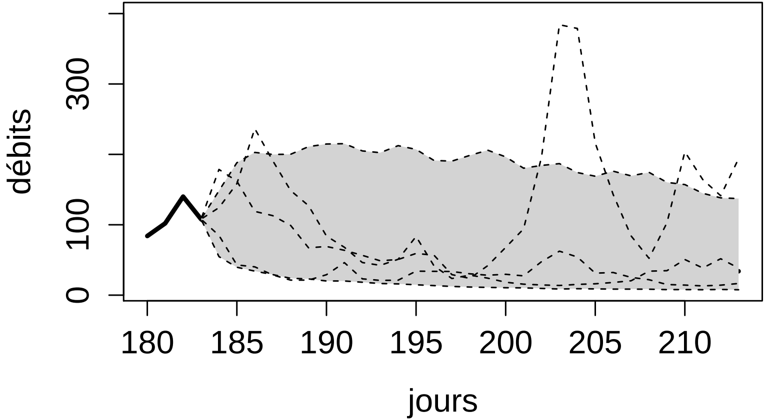




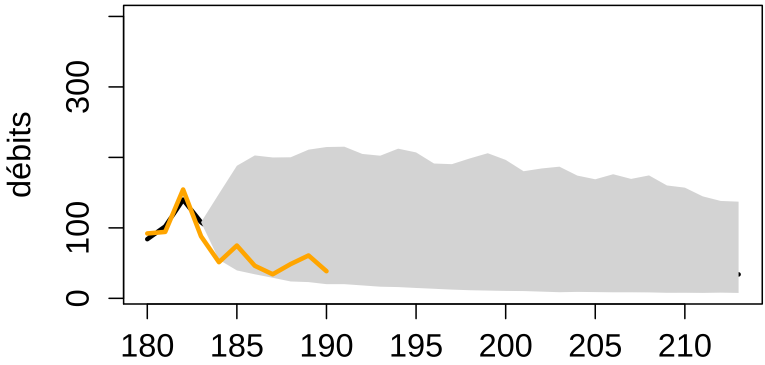
$$X_{t+1} = \alpha_t X_t + \beta_{1,t} Y_{t+1} + \beta_{2,t} Y_t + \tau_t \nu_t$$

$$\nu_t \sim N(0,1),\ Y = log(Q),\ X = log(\mathcal{M}(P))$$

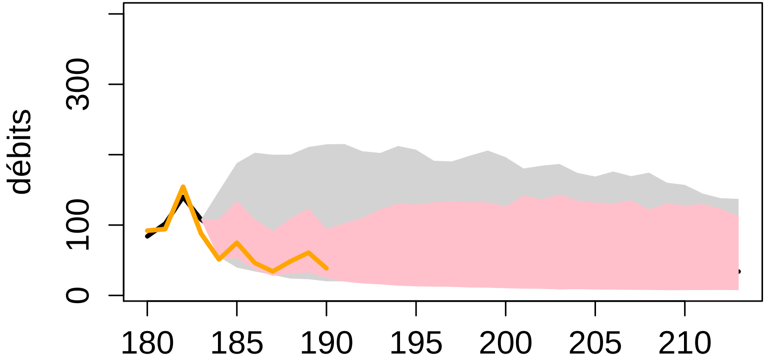
Prévision marginale d'un mois de débits à la station de Vouglans (Ain), en date du 1er avril 2008 (temps 183)



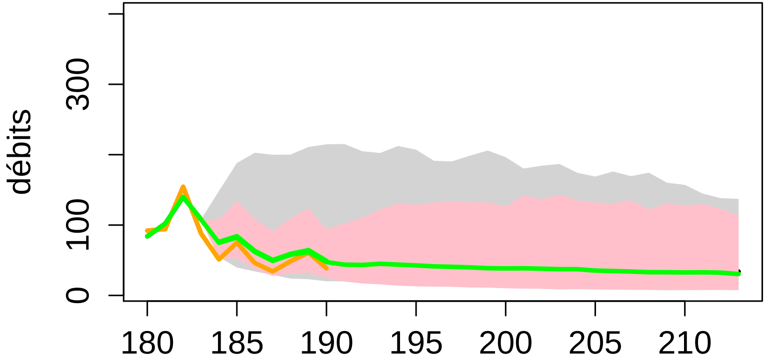
Prévision marginale d'un mois de débits à la station de Vouglans (Ain), en date du 1er avril 2008 (temps 183)+info $\mathcal{M}(P)$



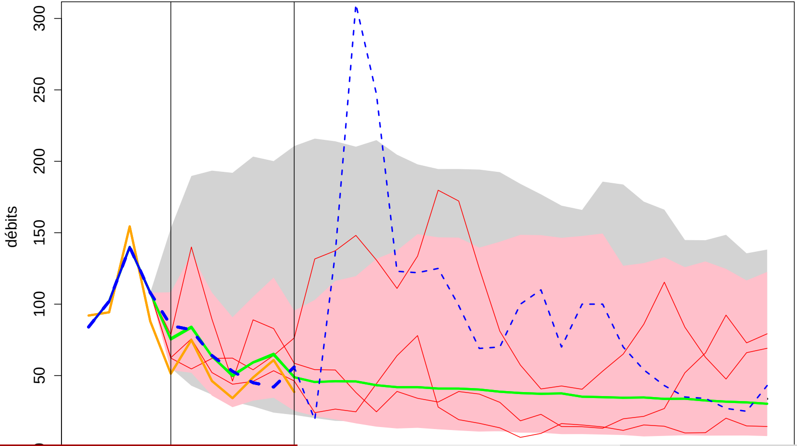
Prévision d'un mois de débits à la station de Vouglans (Ain), en date du 1er avril 2008 (temps 183) conditionnellement à $\mathcal{M}(P)$



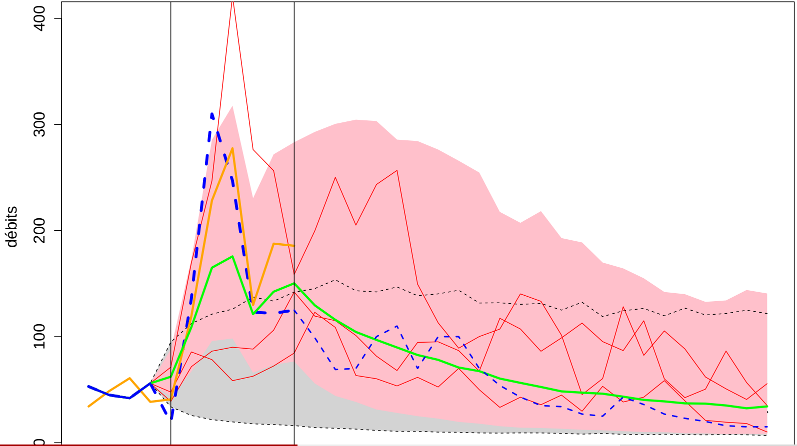
Prévision d'un mois de débits à la station de Vouglans (Ain), en date du 1er avril 2008 (temps 183) conditionnellement à $\mathcal{M}(P)$



Prévision d'un mois de débits à la station de Vouglans (Ain), en date du 1er avril 2008 (temps 183) conditionnellement à $\mathcal{M}(P)$



Prévision d'un mois de débits à la station de Vouglans (Ain), en date du 8er avril 2008 (temps 190) conditionnellement à $\mathcal{M}(P)$



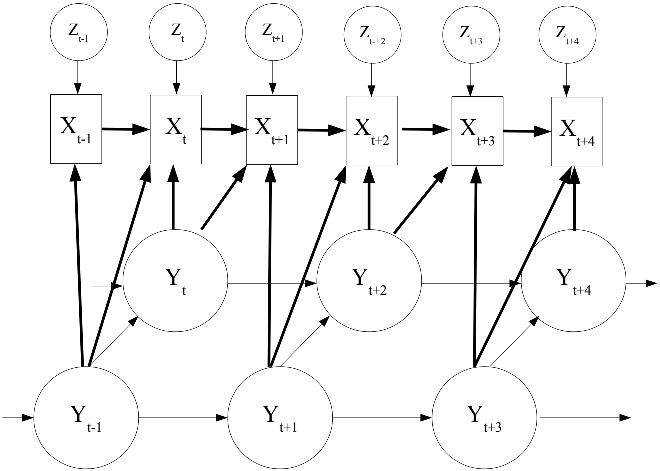
Résultats

- Plus grande liberté de modélisation $[Y]$ et aussi $[X|Y]$ avec $Y = log(Q)$ et $X = log(\mathcal{M}(P))$
- Approche plus naturelle (état de la nature statistique= état de la nature hydro !) conduisant à un raccord à la climatologie à mesure que les débits modélisés deviennent moins informatifs
- Auto-calibrée (modélisation jointe $[X, Y]$)
- A l'encontre des réflexes traditionnels en régression : réponse= f(var explicatives) ! ! !

Limites de la méthode proposée

- Des hypothèses de normalité conjointe injustifiées (en particulier homogénéité des résidus)
- Pertes de conjugaison normal-normal (Kalman) : vers les filtres particulaires
- Classification ? Introduire des modes de fonctionnement suivant type de régimes hydrologiques

Modèle avec déclenchement du mode d'occurrence des débits prévus $M(P)$ selon une variable cachée Z_t



Perpectives

- Adaptation de la méthode aux bassins d'EDF et d'Hydro-Québec
- Mettre en place une classification non supervisée
- Se comparer aux méthodes opérationnelles déjà implantées
- Exploiter la cohérence spatiale
- Modéliser l'incertitude météo (membres d'ensemble)
- Mesurer le gain du à l'intervention de l'expertise...

Merci pour votre attention !

[1] Eva Maria Furrer, Christiane Jacques, and Anne-Catherine Favre.
Short term discharge prediction using a markovian regime switching model.
technical report, INRS-ETE, 2006.

[2] Henry Herr, Edwin Welles, Mary Mullusky, Limin Wu, John Shaake, and Dongjun Seo.
Probabilistic hydrologic forecasting : An ensemble approach.
In *Second Federal Interagency Hydrologic Modelling Conference, Las Vegas, NV*, volume 28, 2002.

[3] Roman Krzysztofowicz.
Bayesian system for probabilistic river stage forecasting.
Journal of Hydrology, 268(1) :16–40, 2002.

[4] Roman Krzysztofowicz and Coire J Maranzano.
Bayesian system for probabilistic stage transition forecasting.
Journal of Hydrology, 299(1) :15–44, 2004.

[5] Zhan-Qian Lu and L Mark Berliner.
Markov switching time series models with application to a daily runoff series.
Water Resources Research, 35(2) :523–534, 1999.

[6] T. Mathevet.
Erreur empirique de modèle.
Note technique interne D4165/NT/2010-00395-A, EDF-DTG, 2010.

Operational situation

(Inspired from [3]) The distribution of interest is $[Q|M_{mod}, hist]$ (M_{mod} : meteorological ensemble forecast). We have :

- $[Q|Q_{mod}, VE, hist]$ (Q_{mod} : modelled waterflow when M is known),
- $[Q_{mod}, VE|M]$ (through the RRM)
- $[M|M_{mod}]$ (ensemble forecasts scenarios are supposed to be drawn from $[M]$)

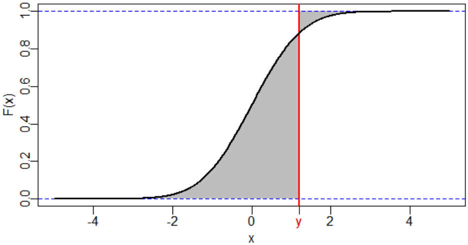
$$[Q|M_{mod}, hist] = \int [Q|Q_{mod}, VE, M_{mod}, hist]$$
$$[Q_{mod}, VE|M_{mod}, hist]d(Q_{mod}, VE)$$
$$[Q_{mod}, VE|M_{mod}, hist] = \int [Q_{mod}, VE|M, M_{mod}, hist][M|M_{mod}, hist]dM$$
$$[Q|M_{mod}, hist] = \int [Q|Q_{mod}, VE, hist]$$
$$\int [Q_{mod}, VE|M][M|M_{mod}]dMd(Q_{mod}, VE)$$

CRPS (Continuous Rank Probability Score)

$$CRPS(F_t, q_t) = \int \left(F_t(x) - 1_{q_t \leq x}\right)^2 dx$$

$$= \mathbb{E}_{F_t}|X - q_t| - \frac{1}{2}\mathbb{E}_{F_t}|X - X'|$$

(to be minimized)

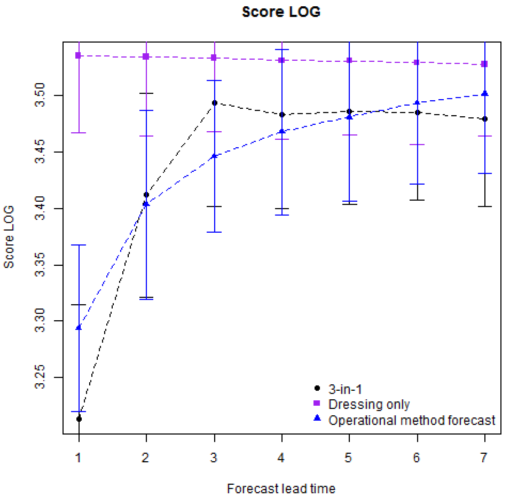


Property (*proper* scoring rule) : if G is the "true" distribution of Q, then
$\mathbb{E}_G CRPS(G, Q) \leq \mathbb{E}_G CRPS(F, Q) \quad \forall F$

We consider $\overline{CRPS} = \frac{1}{T}\sum_{t=1}^{t=T} CRPS(F_t, q_t)$

Logarithmic score results

(evaluation of marginal overall quality)



Reliability diagram results

Reliability diagram for total amount of water in 7 days (evaluation of joint overall quality)