Assessing hydrological forecasts at Hydro-Québec

Luc Perreault

Jocelyn Gaudet
James Merleau
Mylène Teasdale

CSHS Workshop
6-7 october 2011

Outline

> Hydrological forecasts at Hydro-Québec

> Assessing probabilistic forecasts

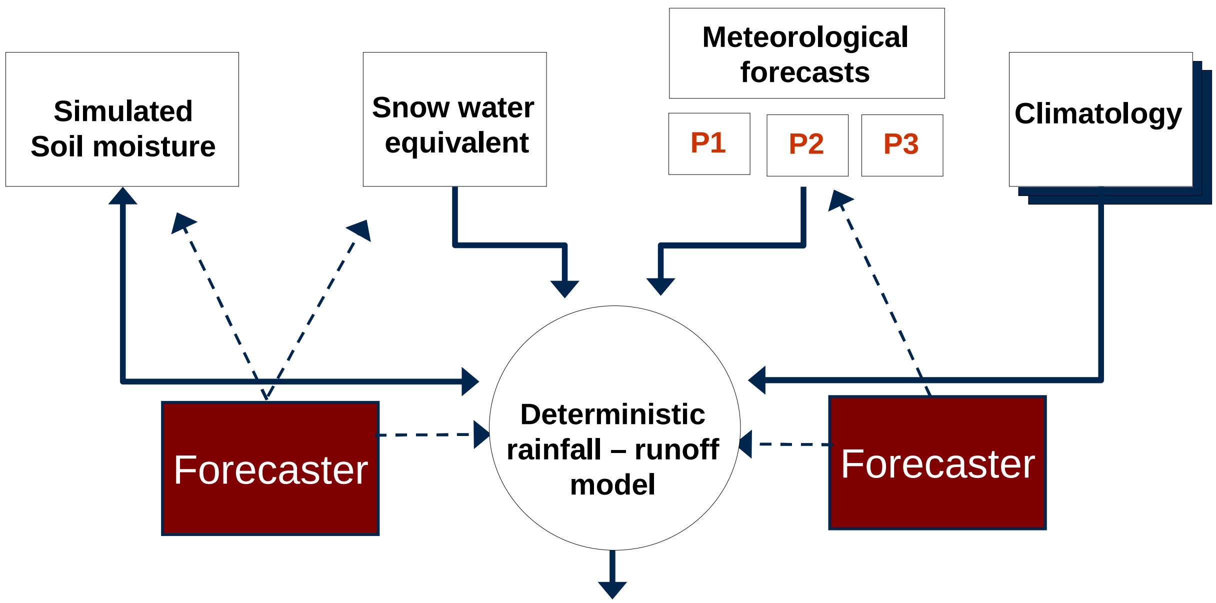
> Applications

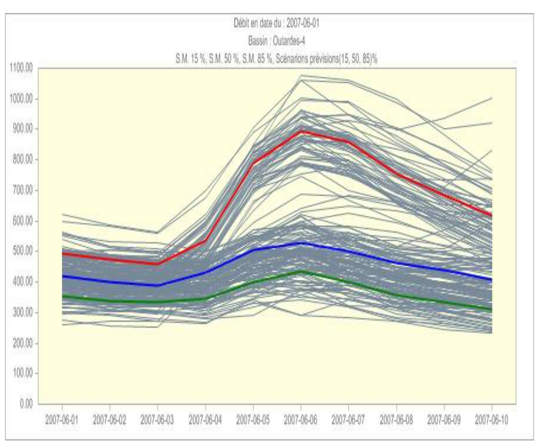
> Some thoughts and operational issues

> Conclusion

Hydrological forecasts at Hydro-Québec

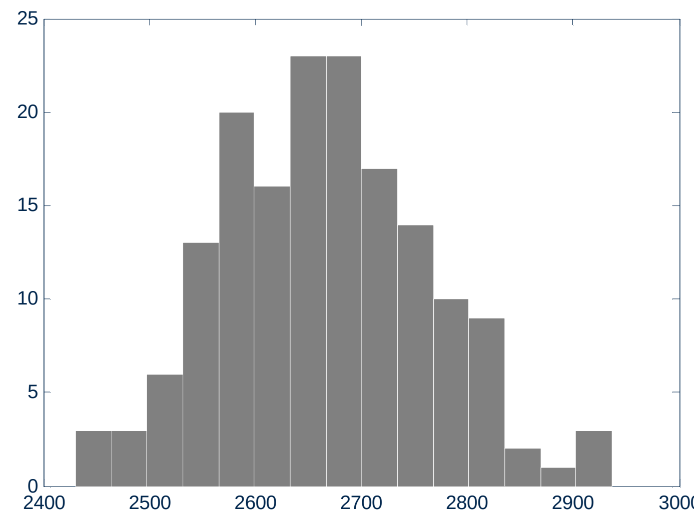
Short term forecasting : precipitation-runoff models





Extended Streamflow Prediction (ESP)

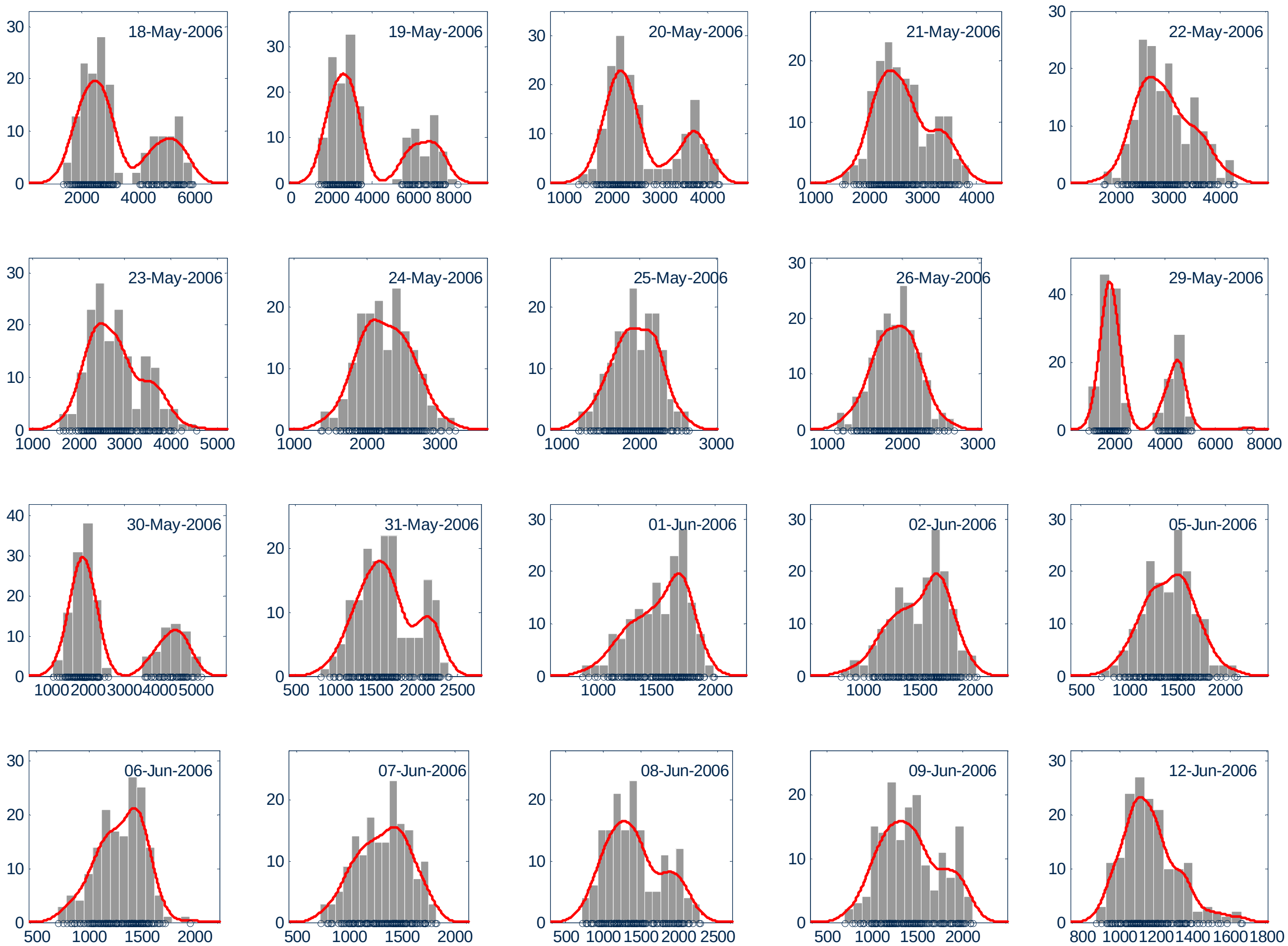
Predictive distribution for the volume



Hydrological forecasts at Hydro-Québec

Short term forecasting : precipitation-runoff models

7-days ahead volume



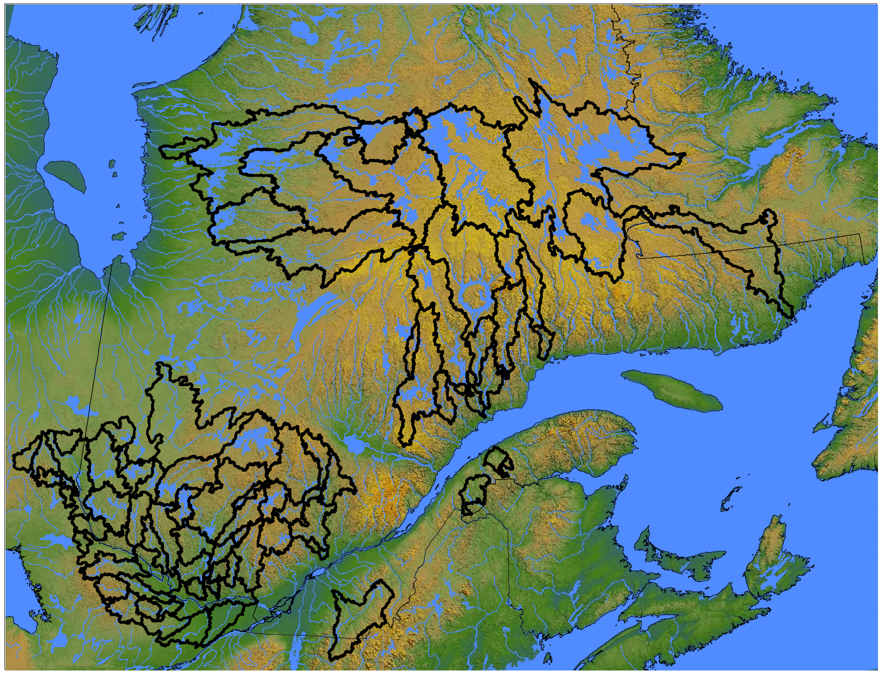
Hydro Québec

Hydrological forecasts at Hydro-Québec

Our mandate

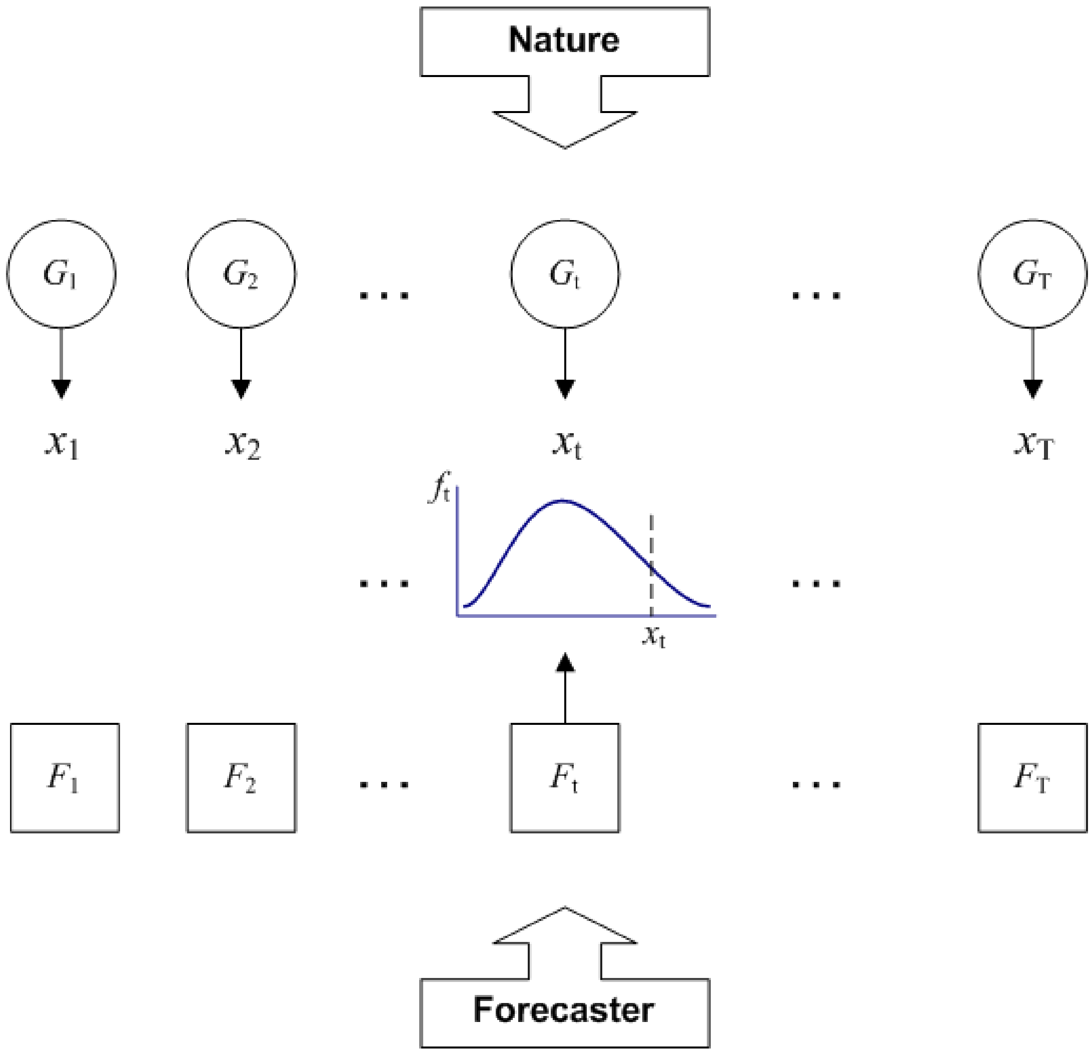
> Propose an **automatic** quality control system **suited** for the **probabilistic forecasts** produced by Hydro-Québec :

- to compare hydrological models
- to study seasonal performance (forecasting skill in critical conditions)
- to ensure the uncertainty is well calibrated
- to help forecasters ajust their interventions in the light of past performances
- …



Assessing probabilistic forecasts

Statistical framework



Assessing probabilistic forecasts

Statistical framework : a good forecast

Gneiting and Raftery (2006)

« Gneiting, Raftery, Balabdaoui and Westveld (2003) and Gneiting, Balabdaoui and Raftery (2005) contend that the goal of probabilistic forecasting is to *maximize the sharpness of the predictive distributions subject to calibration*. »

> Sharpness refers to the concentration of the predictive distribution

> Calibration refers to the statistical consistency between the predictive distribution and the observations

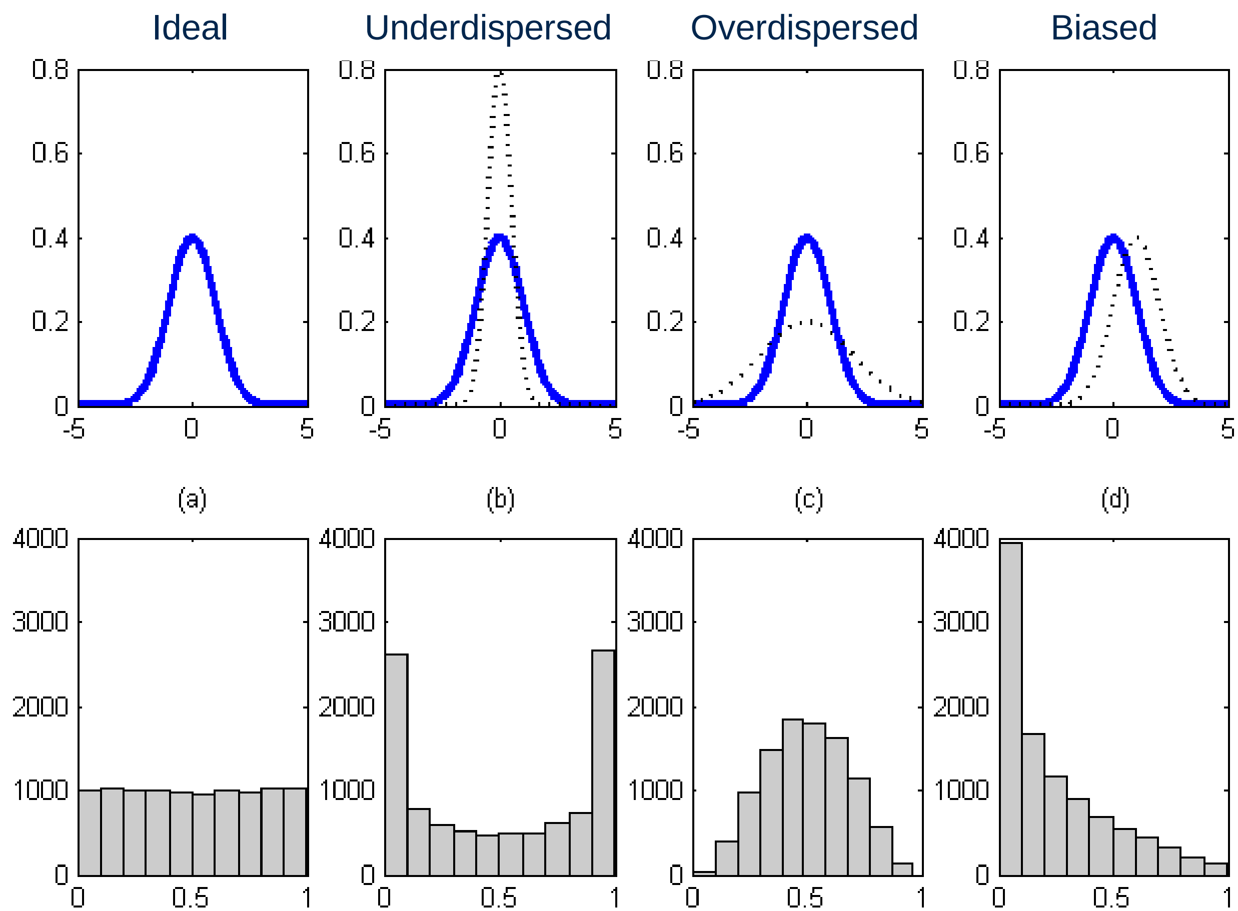
Assessing probabilistic forecasts

Graphical tools : PIT histogram

Dawid (1984)
Diebolt *et al.* (1998)

> Probability Integral Transform (PIT) values

$$p_t = F_t\left(x_t\right) \in \left[0,1\right]$$



Hydro Québec

Assessing probabilistic forecasts

Graphical tools : limitation of PIT histograms

> Nature

$$G_t(x) = N\left(x \middle| \mu_t, 1\right) \quad \text{with} \quad p(\mu_t) = N\left(\mu_t \middle| 0,1\right)$$

> Four predictive distributions

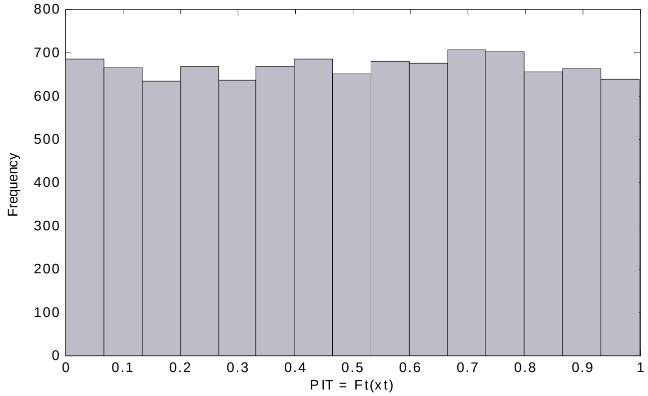
$$F_t(x) = G_t(x) = N\left(x \middle| \mu_t, 1\right)$$

$$F_t(x) = N\left(x \middle| 0, 2\right)$$

$$F_t(x) = \frac{1}{2} N\left(x \middle| \mu_t, 1\right) + \frac{1}{2} N\left(x \middle| \mu_t + \tau_t, 1\right) \quad \text{where} \quad \tau_t = \pm 1$$

$$F_t(x) = N\left(x \middle| \mu_t + \delta_t, \sigma_t^2\right) \quad \text{where} \quad \left(\delta_t, \sigma_t^2\right) = \left(\pm\frac{1}{2}, 1\right) \text{ or } \left(0, \frac{13}{10}\right)$$

All PIT histograms are uniform !!



Gneiting et al. (2007)

Assessing probabilistic forecasts

Scoring rules

- A scoring rule is a function *S*(*F*, *x*) that assigns a numerical value to each pair (*F*, *x*) , where *F* is a predictive distribution and *x* is the observation.

- We aim to minimize the average score

$$\boxed{S_T = \frac{1}{T}\sum_{t=1}^{T} S\left(F_t, x_t\right)}$$

- Important property : a proper score [Brocker and Smith (2006)]

We need a form of *S*(*F*, *x*) which guarantees that the expected score is minimized if and only if the predictive distribution is *G*.

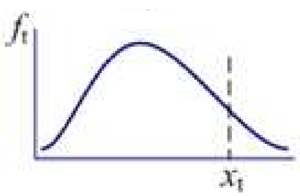
Forecasted skill of *F* Forecasted skill of *G*

$$E_G\left[S\left(F, X\right)\right] \geq E_G\left[S\left(G, X\right)\right]$$

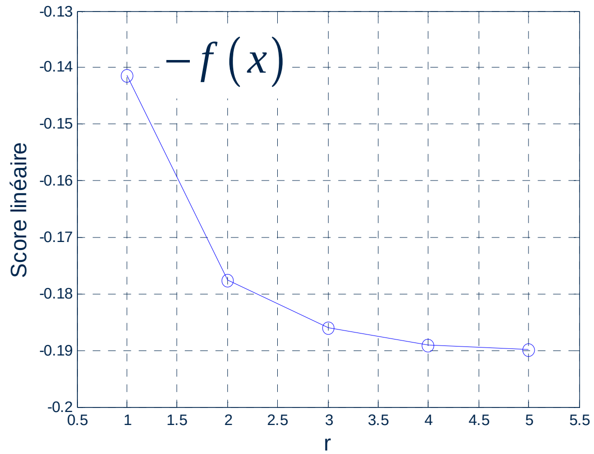
Assessing probabilistic forecasts

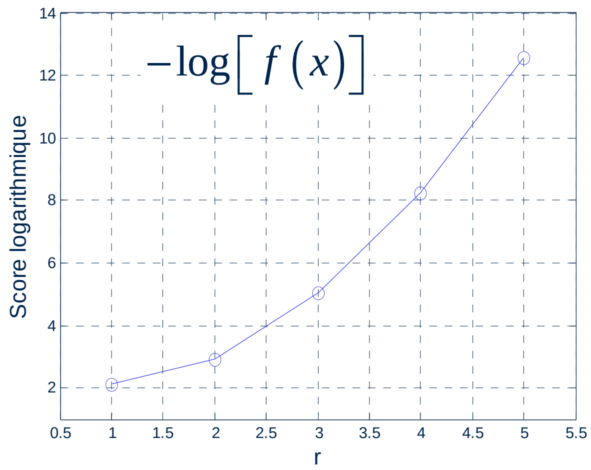
Proper scoring rules

- Nature G_t is a normal distribution N(0, 2)
- Predictive F_t is a normal distribution N(0, 2/r)
  - Unbiased but underdispersed if r > 1
  - $\sigma(F_t) = \sigma(G_t) / r$
- Improper scoring rule $f_t(x_t)$



Underdispersion

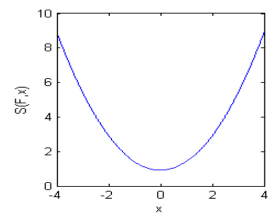




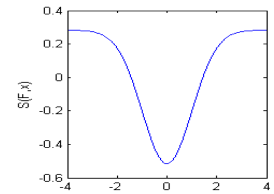
Assessing probabilistic forecasts

Proper scoring rules

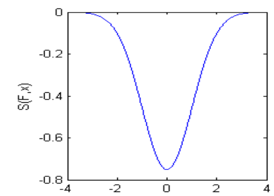
Logarithmic $$-\log\left[f\left(x\right)\right]$$



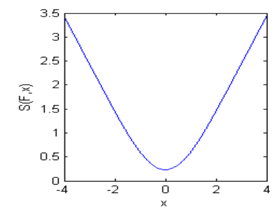
Quadratic $$E_F\left[f\left(X\right)\right]-2f\left(x\right)$$



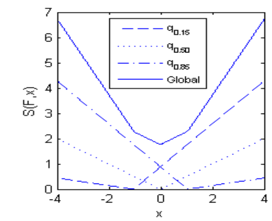
Spherical $$-f\left(x\right)/E_F^{1/2}\left[f\left(X\right)\right]$$



CRPS $$\int\left[F\left(y\right)-\mathbf{1}\left(y\geq x\right)\right]^2 dy$$



Quantile $$S\left(q,x\right)=\left(x-q\right)\left(\mathbf{1}\left\{x\geq q\right\}-\alpha\right)$$

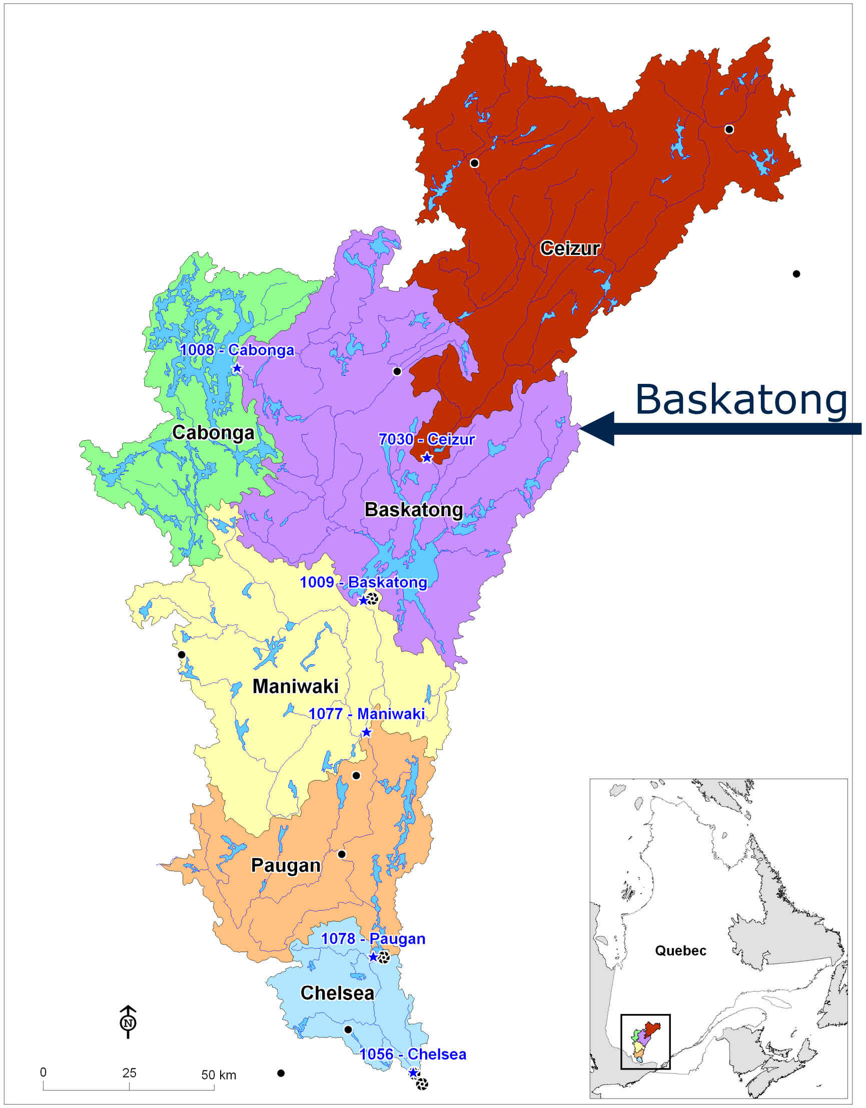


Hydro Québec

Applications

« Short term » forecasting

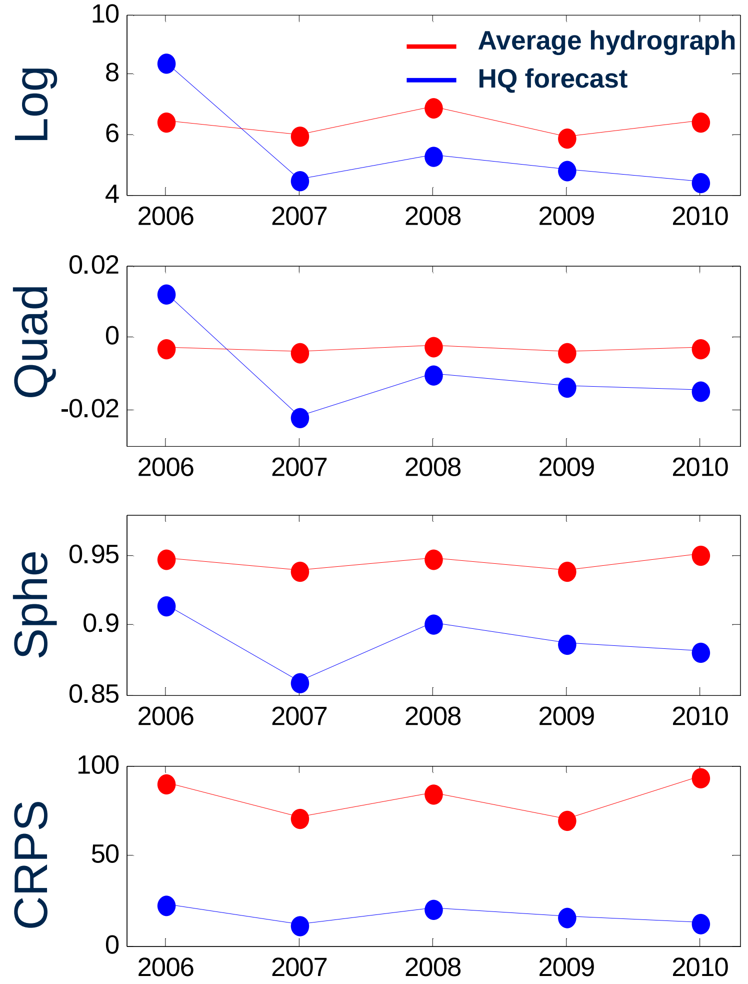
- Baskatong basin
- Archives of forecasts produced daily from 2006 to 2010
- For this illustration, only the 1-day ahead forecasts are considered
- Reference model : the average hydrograph (1950 up to the year of forecasting)



Applications

« Short term » forecasting : 1-day ahead

Global score



CRPS skill score

	Winter	Spring	S./Fall	**Global**
2006	0,35	0,86	0,76	0,75
2007	0,82	0,79	0,90	0,84
2008	0,81	0,77	0,73	0,76
2009	0,71	0,77	0,81	0,77
2010	0,73	0,91	0,84	0,86
Global	0,67	0,82	0,79	0,78

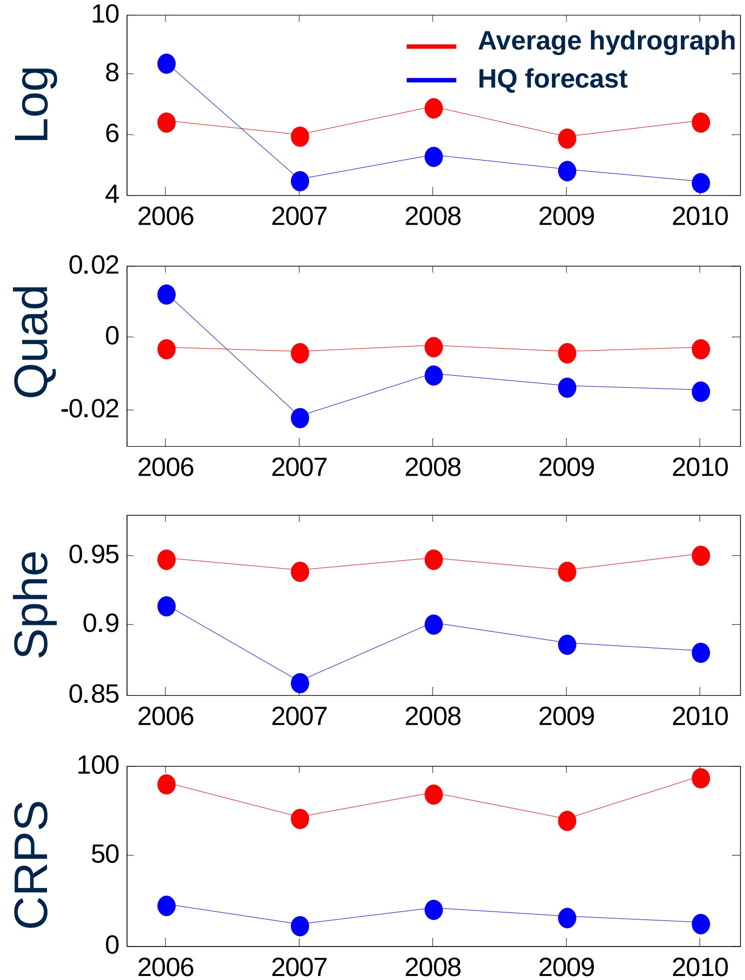
$$1-\frac{CRPS(\mathrm{HQ})}{CRPS(\mathrm{Ref})}$$

Hydro Québec

Applications

« Short term » forecasting : 1-day ahead

Global score



Log difference

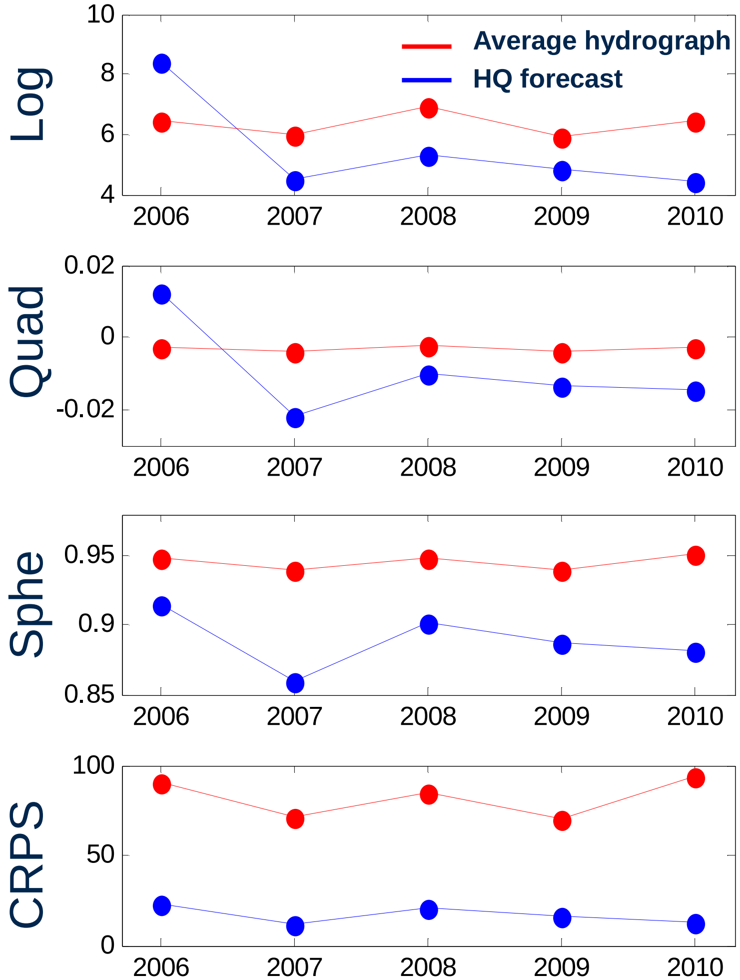
	Winter	Spring	S./Fall	**Global**
2006	-9,59	1,73	1,38	-1,97
2007	1,77	1,32	1,45	1,49
2008	2,06	0,95	1,59	1,57
2009	0,70	1,27	-1,35	1,10
2010	1,43	2,99	1,87	2,00
Global	-0,96	1,53	1,31	0,63

$$\log(\mathrm{Ref}) - \log(\mathrm{HQ})$$

Applications

« Short term » forecasting : 1-day ahead

Global score



Log difference

	Winter	Spring	S./Fall	**Global**
2006	-9,59	1,73	1,38	-1,97
2007	1,77	1,32	1,45	1,49
2008	2,06	0,95	1,59	1,57
2009	0,70	1,27	-1,35	1,10
2010	1,43	2,99	1,87	2,00
Global	-0,96	1,53	1,31	0,63

$$\log(\mathrm{Ref}) - \log(\mathrm{HQ})$$

Some thoughts and operational issues

Experiment 1 : Tails of the predictive distribution

> Monte-Carlo experiment

T = 10000, N_{ens} = 300

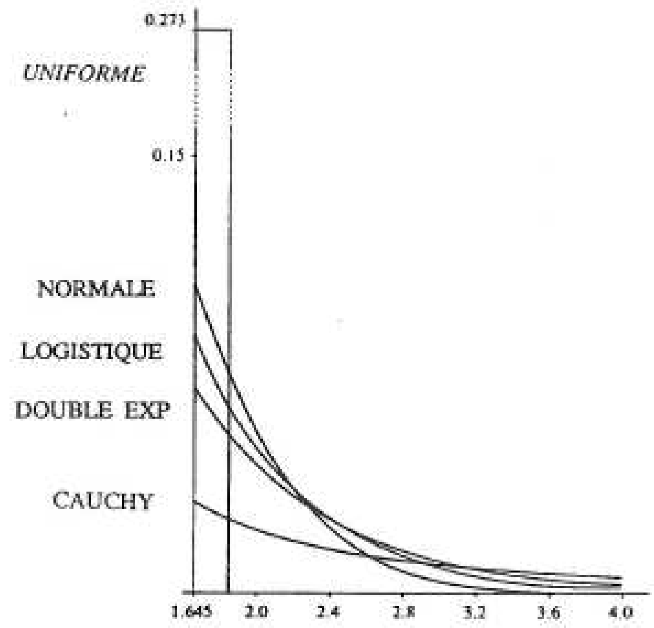
> Nature

$$G_t = N(\mu_t, 1) \quad \text{where} \quad p(\mu_t) = N(0,1)$$

> Four predictive distributions with

$E(X) = m_t$ and $V(X) = 1$

but with different tail behavior



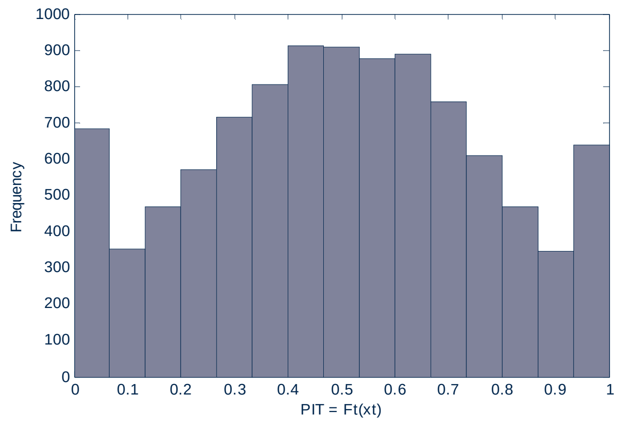
Uniform $<_s$ Normal (nature) $<_s$ Logistic $<_s$ Double exponential

Van Zwet (1964)
Capéràa (1986)

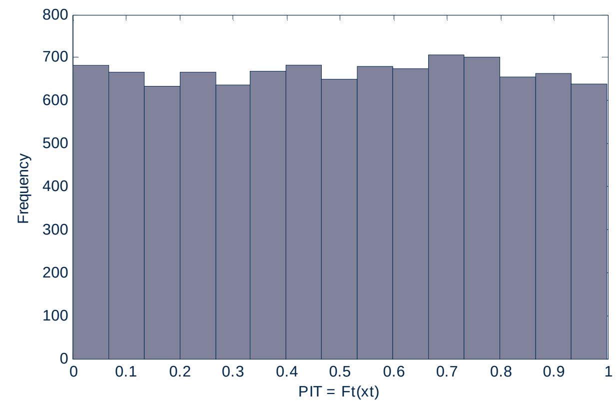
Some thoughts and operational issues

Experiment 1 : PIT histograms

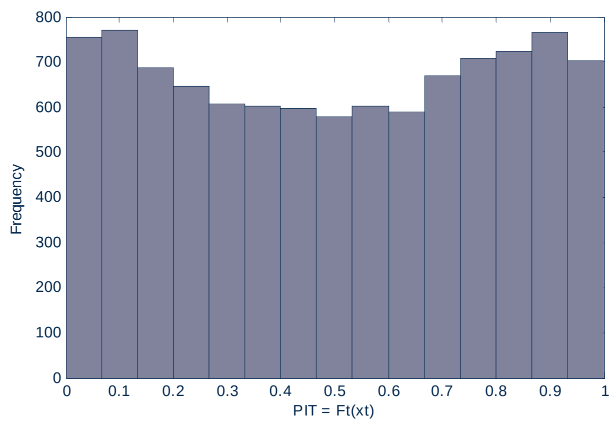
Uniform



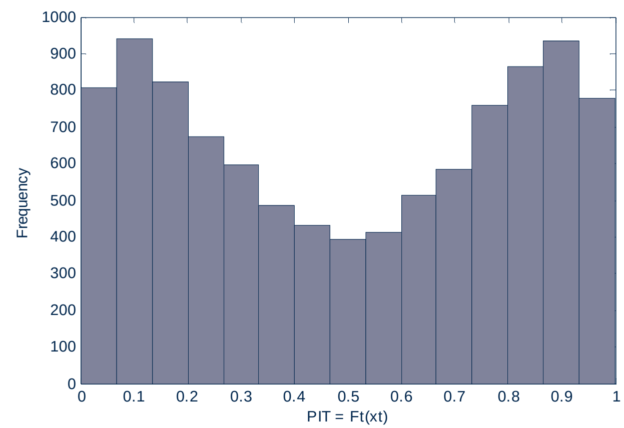
Ideal : normal



Logistic



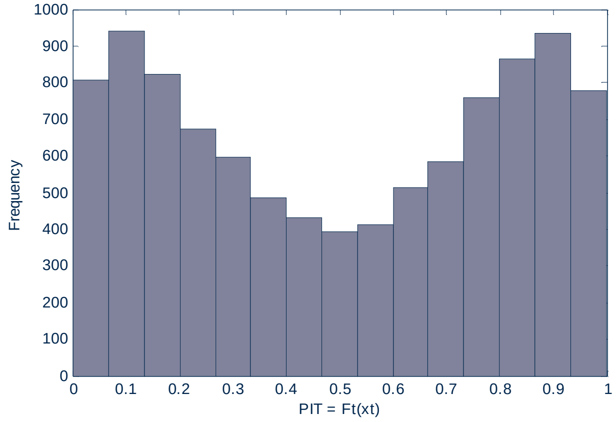
Double exp.



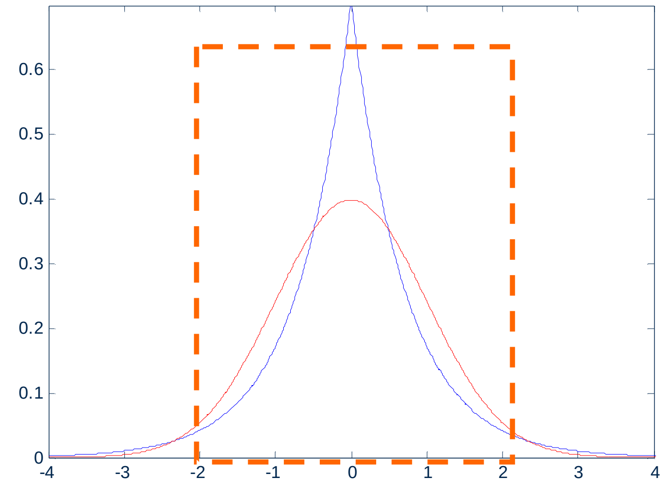
Some thoughts and operational issues

Experiment 1 : PIT histograms

Double exp.



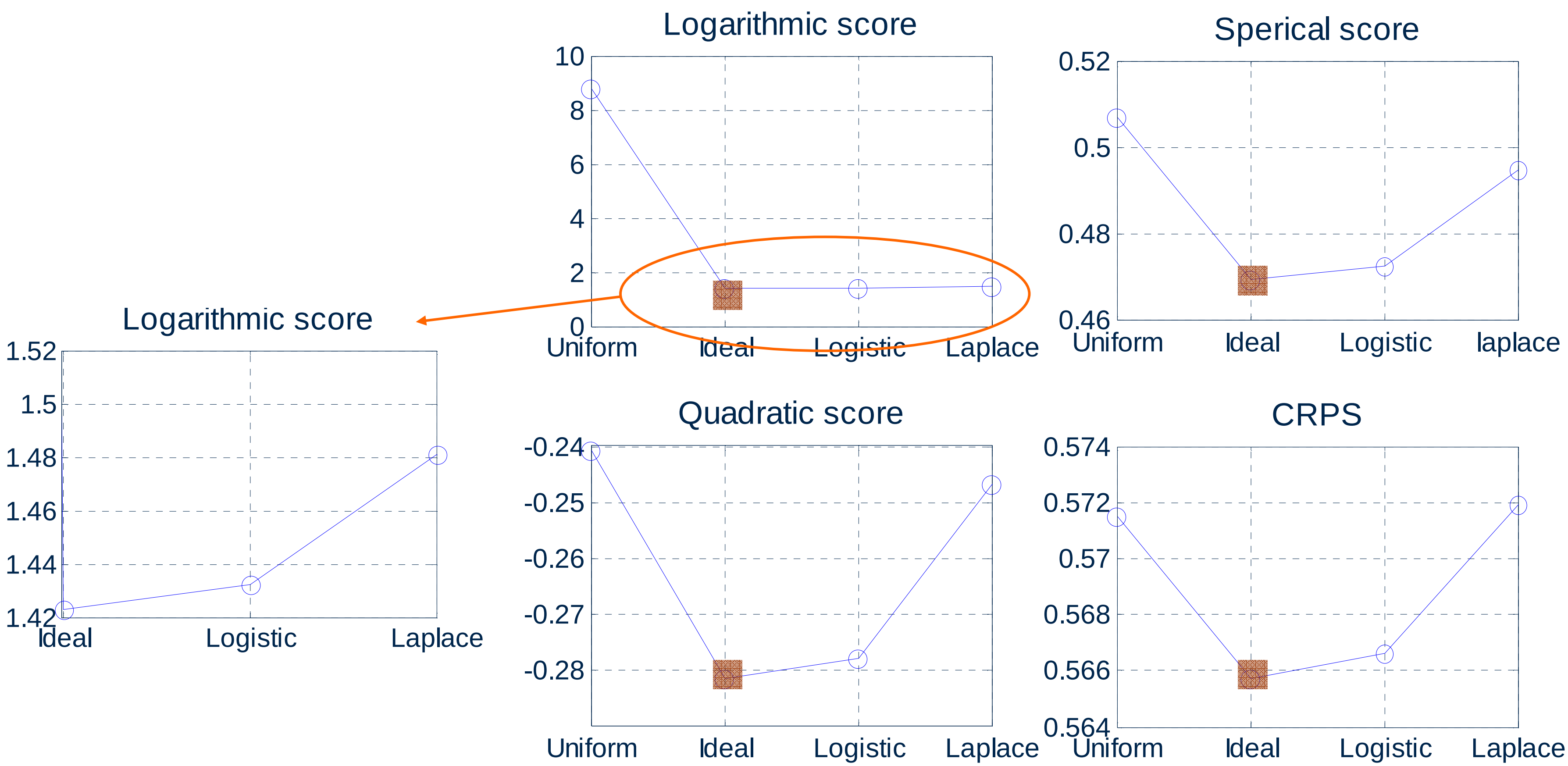
Normal vs Double exp.



Hydro Québec

Some thoughts and operational issues

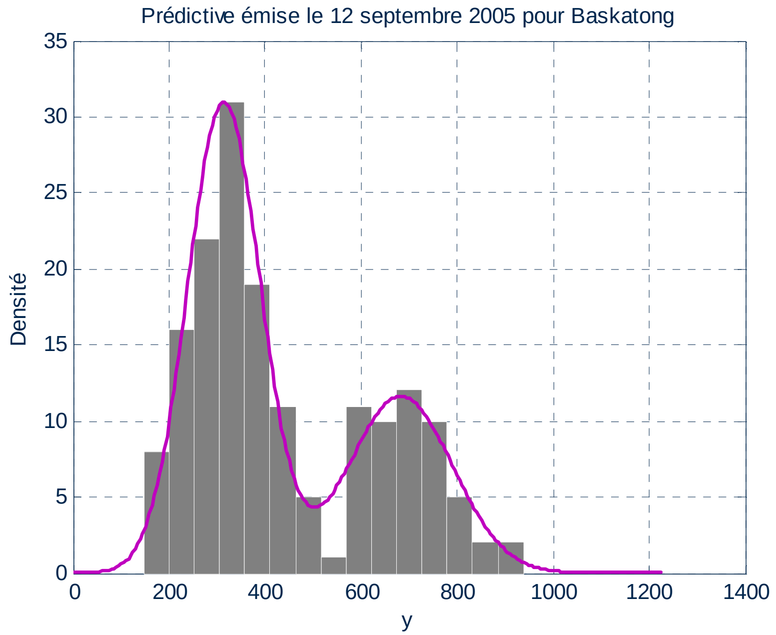
Experiment 1 : Proper scoring rules



Some thoughts and operational issues

Experiment 2 : « Distance sensitivity »

> Four predictive distributions
- Uniform
- Normal
- Gamma
- 2-component mixture of normal distributions

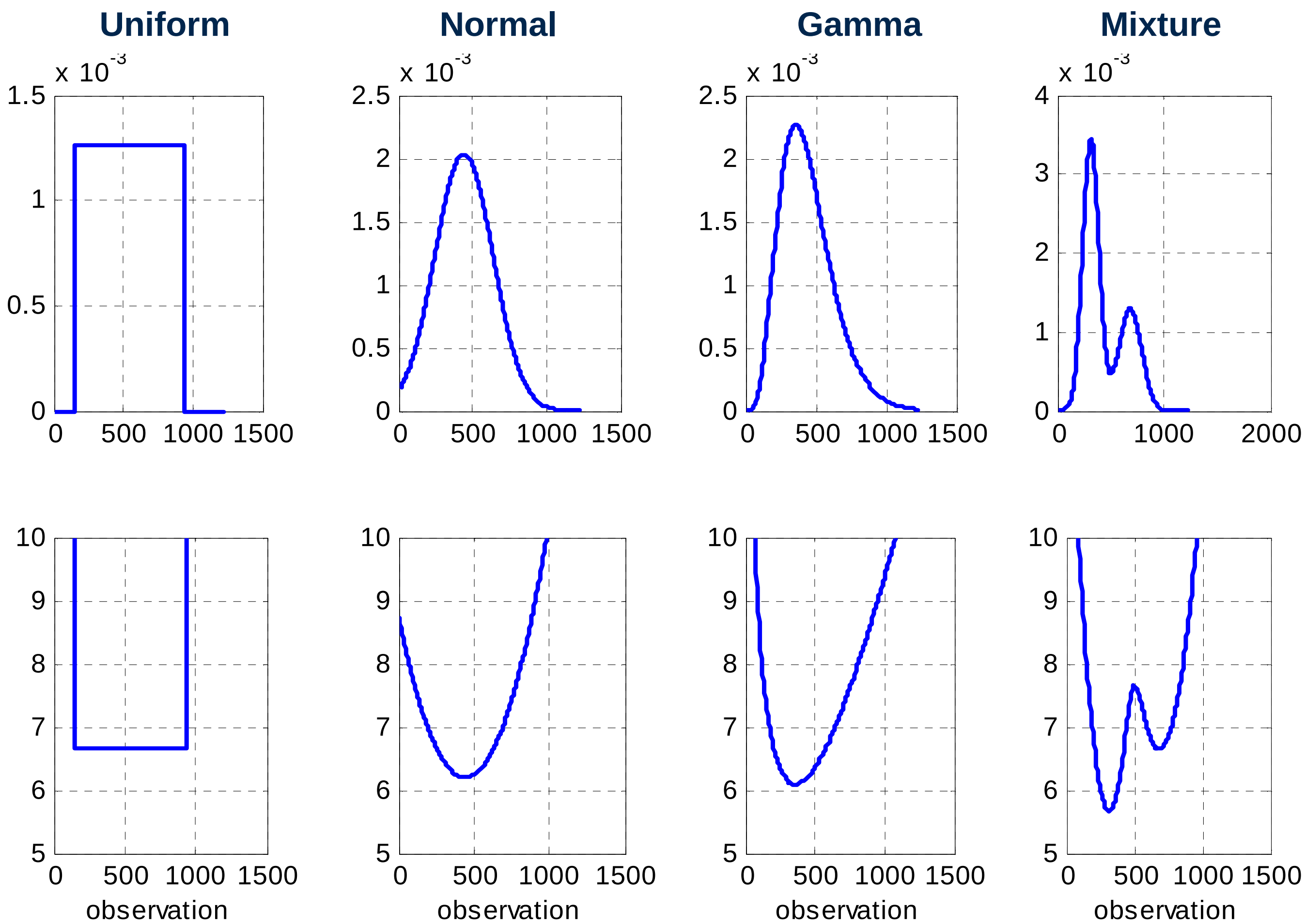


> Parameters estimated using the 7-day ahead forecast produced in September 2005 for Baskatong

Some thoughts and operational issues

Experiment 2 : « Distance sensitivity »

Log

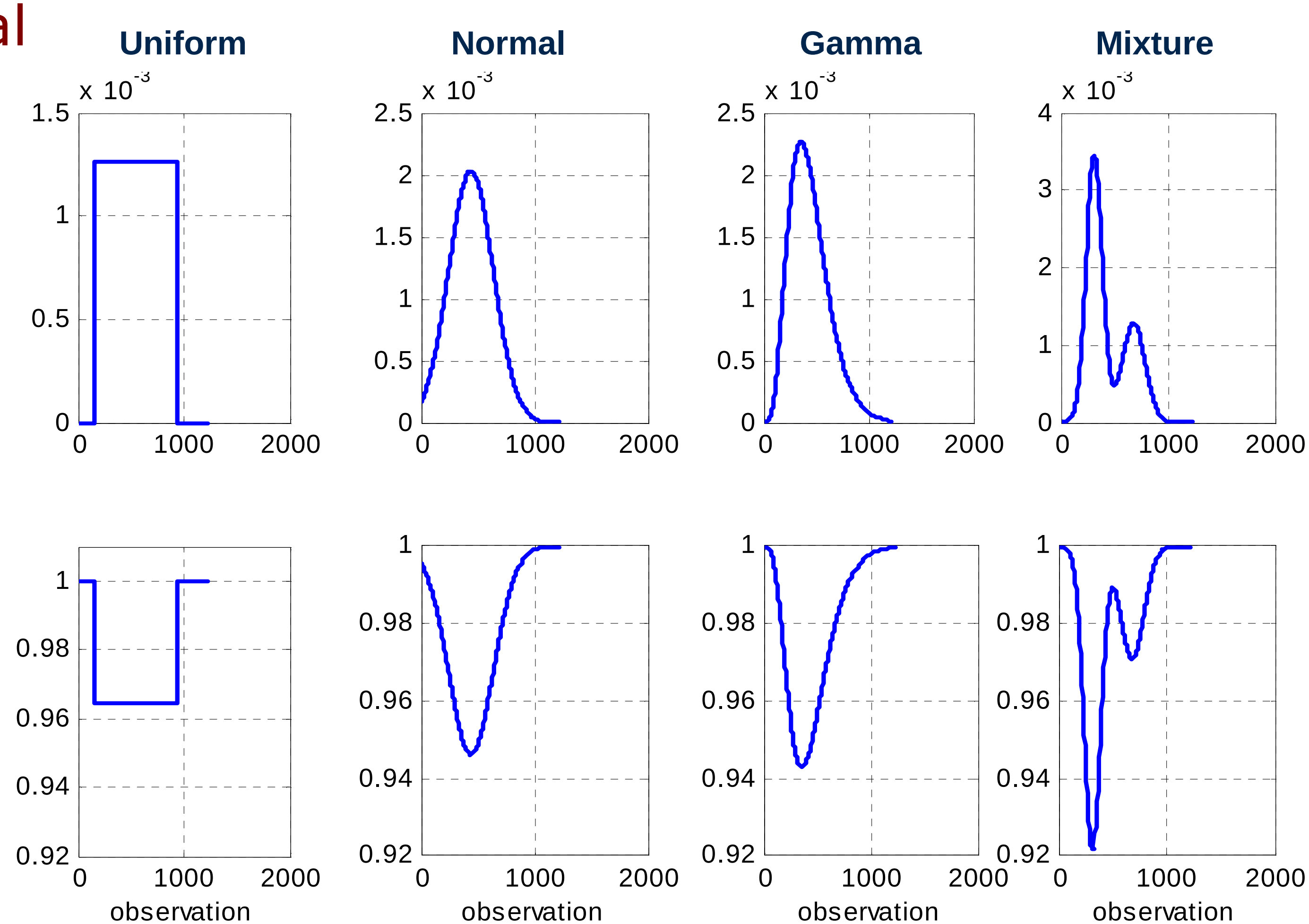


Hydro Québec

Some thoughts and operational issues

Experiment 2 : « Distance sensitivity »

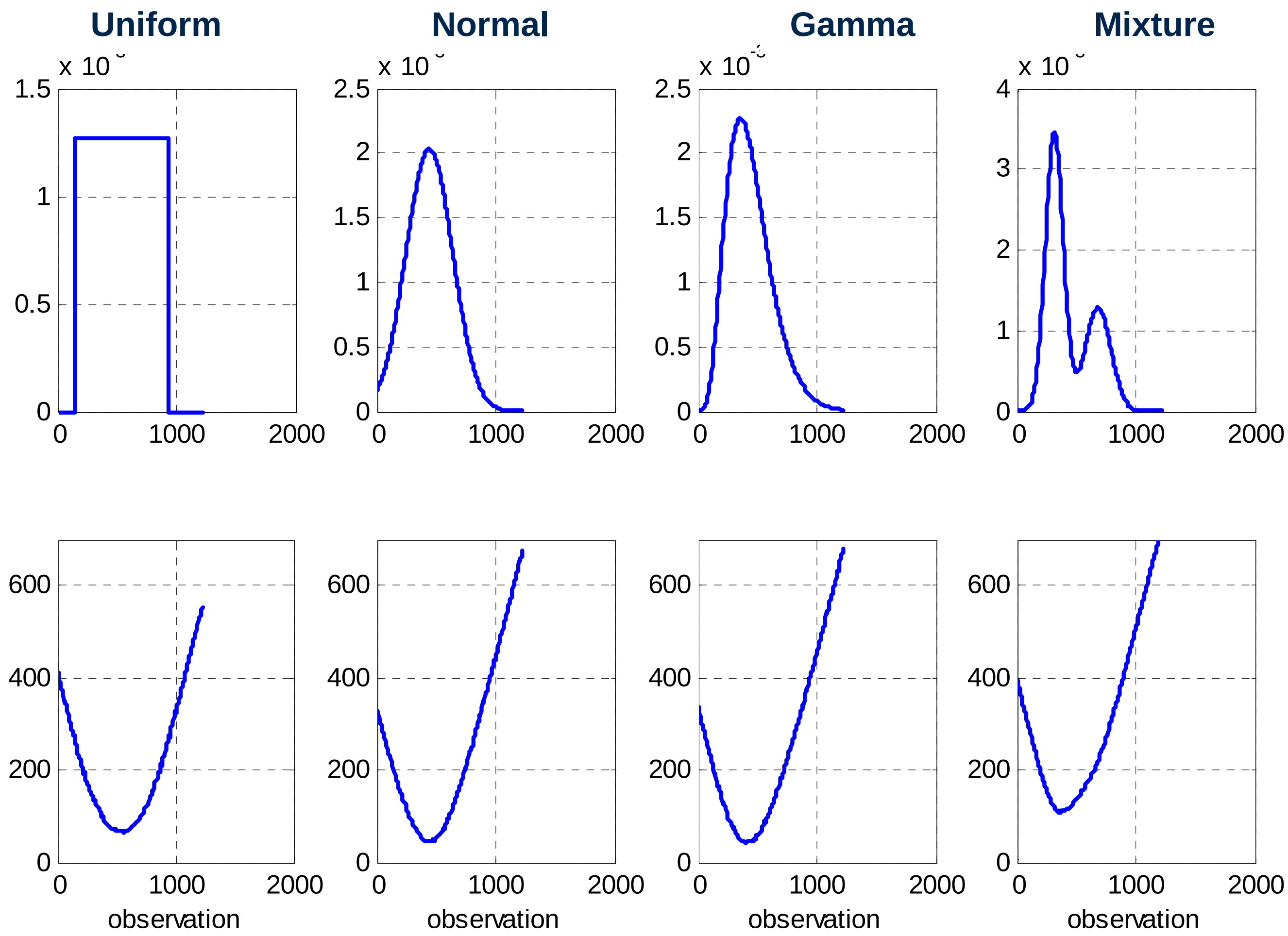
Spherical



Some thoughts and operational issues

Experiment 2 : « Distance sensitivity »

CRPS

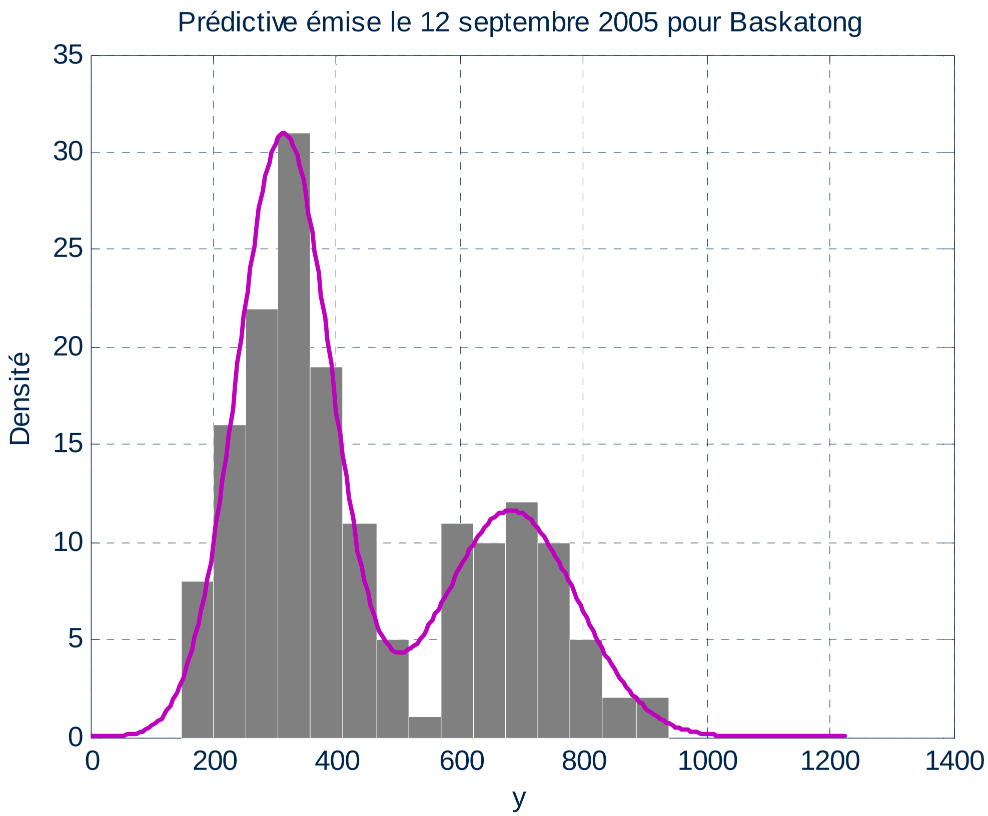


Conclusion

- A proper score guarantees that it can identify the “true” distribution which in practice is never known. Defining a sensible ‘scale’ criterion to measure the degree of improvement is not trivial.
- As simple experiments show, diagnostic tools which assess the quality of forecasts need to be interpreted carefully.
- PIT histograms monitor dispersion mainly where the bulk of the observations concentrate and are not necessarily sensitive to the tail behavior of predictive distributions.
- CRPS seems to be an adequate measure for assessing average behavior, but it doesn’t give precise information on the local behavior of a predictive distribution. For short term forecasting, a score which is more sensitive to the shape of the distribution should be used.

Hydrological forecasts at Hydro-Québec

Short term forecasting : why bimodal ?

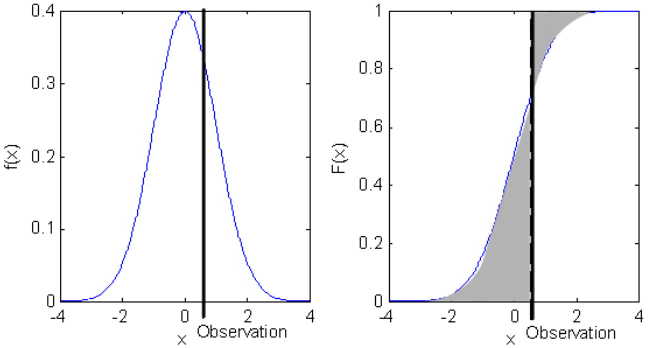


	P1	P2	P3
2005-09-12	0	0	6,4
2005-09-13	0	0	2,2
2005-09-14	3,7	6,2	30
2005-09-15	0	2,2	5

Assessing probabilistic forecasts

Proper scoring rules : CRPS

$$\mathrm{CRPS}(F,\ x) = \int \left[F(y) - \mathbf{1}(y \geq x) \right]^2 dy$$
$$= E_F\,|X - x| - \frac{1}{2} E_F\,|X - X'|$$



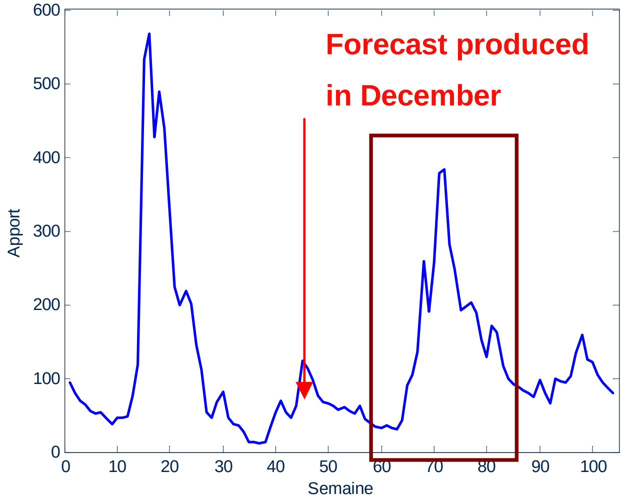
> Some interesting properties
- Possesses the same units as the observation
- Generalizes the absolute error (MAE)
- It is a « robust » scoring rule (M-estimator, Huber (1982))
- Can be evaluated using Monte Carlo integration

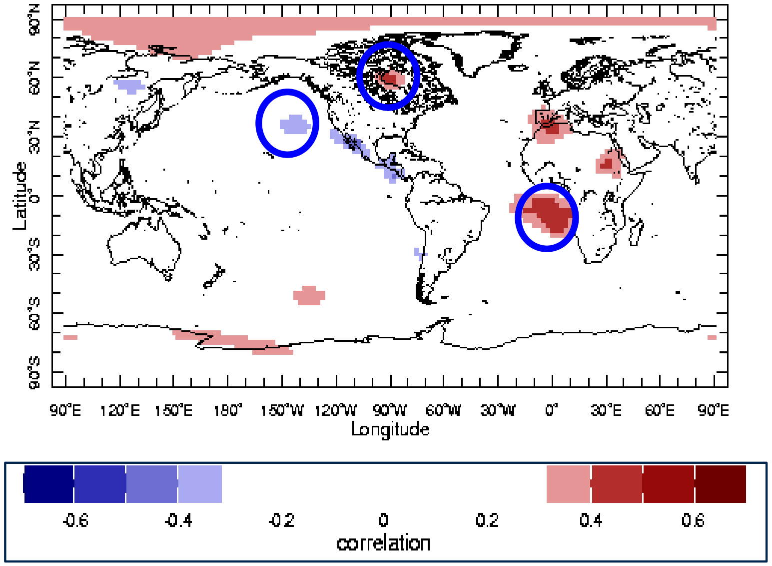
Gneiting et al. (2005) : « our favorite score »

Hydrological forecasts at Hydro-Québec

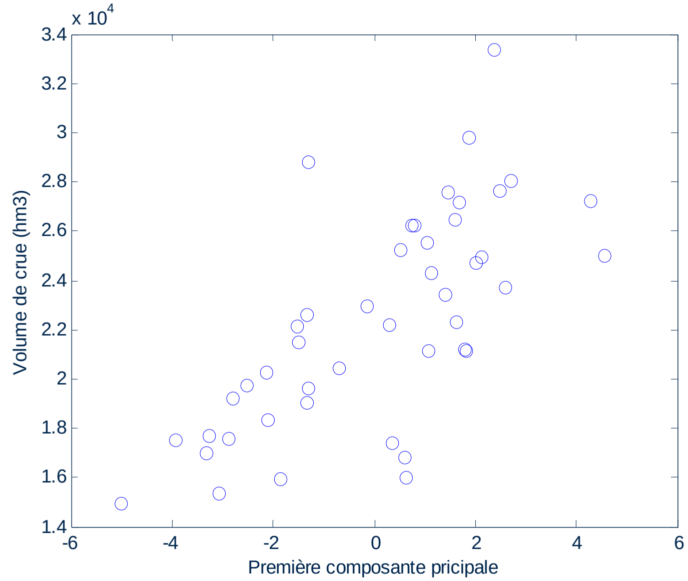
Long term forecasting : dynamic regression model based on measures of atmospheric circulation

Sveinsson *et al*. (2007)





$$y_j = \beta_0 + \beta_1 x_{j-1,1} + \beta_2 x_{j-1,2} + \varepsilon_j$$



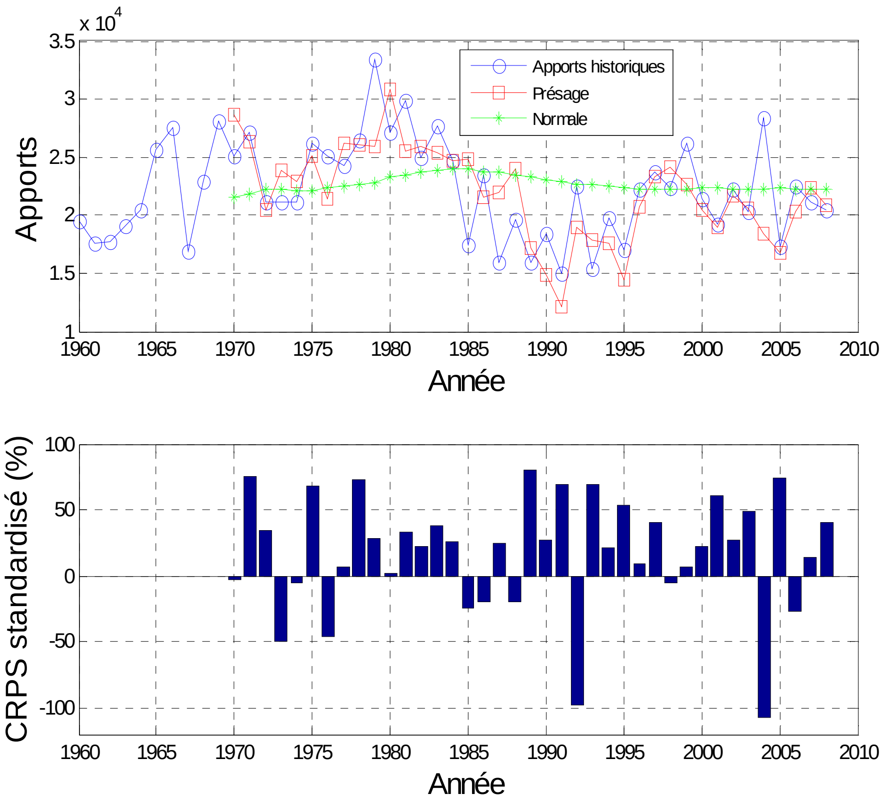
Hydro Québec

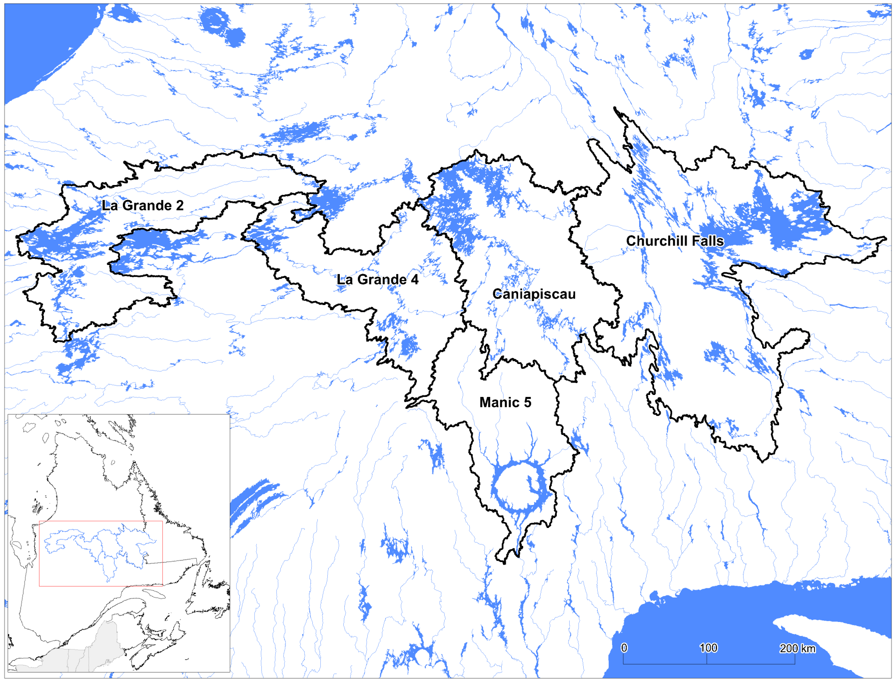
Applications

« Long term » forecasting : 5 to 6 months ahead

Sveinsson *et al*. (2007)

> Dynamic regression model based on measures of atmospheric circulation



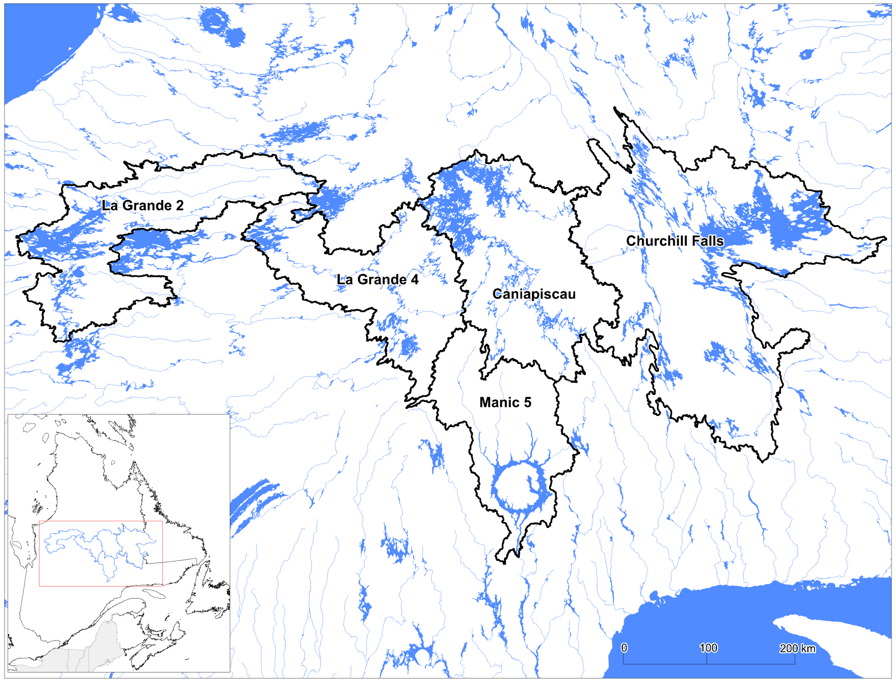


Applications

« Long term » forecasting : 5 to 6 months ahead

CRPS skill score

	ATM	AR1	AR2	AR3
Churchill Falls	26%	-12%	12%	4%
Manic 5	24%	-5%	-2%	2%
Caniapiscau	14%	-28%	-15%	-16%
La Grande 4	11%	-14%	-9%	-11%
La Grande 2	7%	-27%	-26%	13%



Hydro
Québec