Bayesian Processor of Ensemble Members: combining the Bayesian Processor of Output with Bayesian Model Averaging for reliable ensemble forecasting

R. Marty[1,2] V. Fortin[2] H. Kuswanto[3] A.-C. Favre[4] E. Parent[5]

[1]Université Laval, Québec, Canada [2]Environnement Canada, Dorval, Canada [3]Institute Technology of Sepuluh Nopember, Indonesia
[4]Grenoble INP/LTHE, Grenoble, France [5]AgroParisTech/INRA, Paris, France

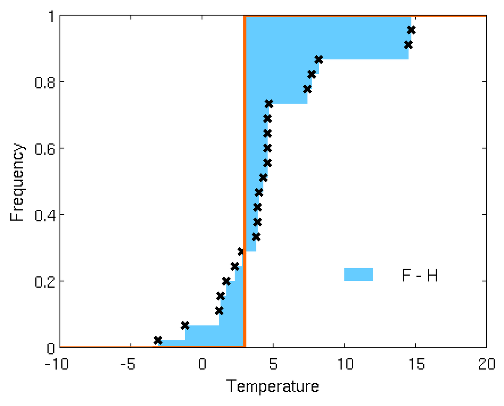
Why Calibrate Ensemble Predictions ?

- Ensemble forecasts: a mean to assess the uncertainty in meteorological forecasts
- ➡ Unfortunately uncertainty underestimated by current ensemble prediction systems (EPS)
- ➡ Unfortunately ensemble often provided at unsuitable spatial/temporal scales (e.g. for hydrological predictions)
- ➡ Statistical post-processing required to obtain reliable ensemble forecasts at appropriate scales

Verification Scores

Continuous Ranked Probability Score

- Comparison of cumulative distribution functions from ensemble forecasts **F**, with observation y through the Heaviside function
- Negatively oriented (smaller is better)
- See details in Hersbach (2000)



$$\mathrm{CRPS}\{F,y\} = \int_{-\infty}^{\infty} \{F(u) - H(u-y)\}^2 du$$

Reliability

- Statistical consistency between a priori predicted probabilities and a posteriori observed frequencies of the occurrence
- Reliability measured by the Reliability Component of the CRPS decomposition (Hersbach, 2000)

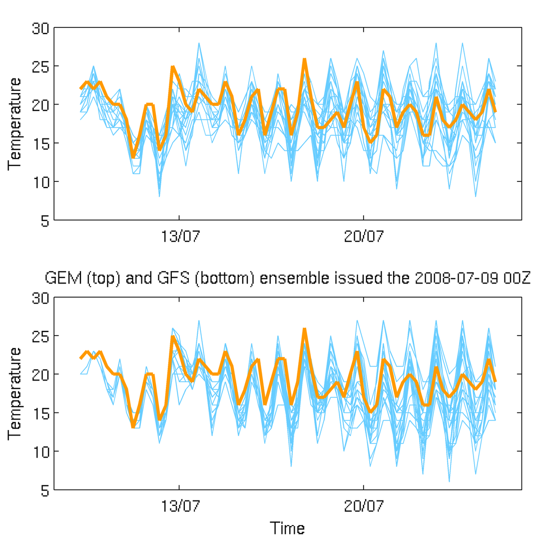
Observation and Forecast Datasets

Canadian Daily Climate Data

- Temperature observation from meteorological station in Jean Lesage Intl. Airport (YQB) in Quebec City over the period 1978-2007
- Temperature observed at 00Z
- `http://www.climat.meteo.gc.ca`

North American Ensemble Forecasting System

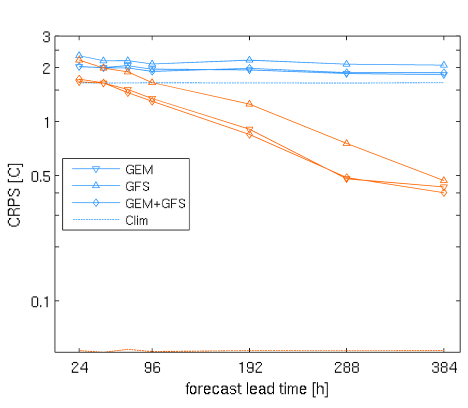
- 21 members from the Global Ensemble Prediction System of CMC based on the Global Environmental Multiscale (GEM) model
- 21 members from the ensemble configuration of the Global Forecast System (GFS) of NCEP
- Forecasts provided on a 1-degree grid
- Runs 00Z, lead times +24h to +384h by 24h
- from April 2008 to March 2009



observations vs ensemble forecast (Quebec City)

Verification of raw ensemble forecasts

- GFS: lowest skill for all lead times
- GEM+GFS: skill closed to GEM
  - ➡ No additional benefit in forecasting by using 42 members
- Ensemble predictions less skillful than climatology
- Ensemble predictions are not reliable
  - ➡ Calibration of ensemble predictions from GEM



Skill (CRPS)
Reliability (CRPS Rel)
☞ (y) log scale

Notations and Hypotheses

Notations

- t valid date
- S ensemble size
- h forecast's lead time
- y_t predictand (quantity to predict)
- $\mathbf{X}_t^{(h)}$ ensemble forecasts
- $\mathbf{X}_t^{(h)} = \{X_{t,s}^{(h)}, s = 1, 2, ..., S\}$
- temporal independance: h is omitted

Assumptions

- Ensemble members generated in the same way
  - ➡ exchangeability
- Numerical model well suited for predicting an unobserved (latent) variable ξ_t (e.g. gridded temperature)
- Latent variable ξ_t exchangeable with all ensemble members $X_{t,s}$
- ξ_t contains all the information required to predict y_t
  - ➡ $\mathbf{X}_t$ and y_t are conditionally independant

BMA Component (Raftery et al., 2005)

- Law of total probability

$$p(y_t|\mathbf{X}_t) = \int p(y_t|\xi_t, \mathbf{X}_t)p(\xi_t|\mathbf{X}_t)d\xi_t$$

- As $\mathbf{X}_t$ and y_t are conditionally independant

$$p(y_t|\mathbf{X}_t) = \int p(y_t|\xi_t)p(\xi_t|\mathbf{X}_t)d\xi_t$$

- Since the latent variable is exchangeable with ensemble members

$$p(\xi_t|\mathbf{X}_t) \approx \frac{1}{S}\sum_{s=1}^{S}\delta(\xi_t - X_{t,s})$$

➡ BMA framework: Non-parametric approximation of the predictive distribution

$$p(y_t|\mathbf{X}_t) \approx \frac{1}{S}\sum_{s=1}^{S} p(y_t|X_{t,s})$$

BPO Component (Krzysztofowicz, 2004)

- Bayes' rule

$$\underbrace{p(y_t|\xi_t)}_{posterior} \propto \underbrace{p(\xi_t|y_t)}_{likelihood}\underbrace{p(y_t)}_{prior}$$

- Prior distribution of the predictand = climatology
  - ▷ Temperature: Gaussian distribution
  - ▷ Estimated from the past 30 years (1978–2007) with a moving window of 5 days around the valid date

$$p(y_t) = \mathcal{N}\left(y_t; \mu, \sigma^2\right)$$

- Likelihood function
  - ▷ Assuming linear model with Gaussian residuals between the predictand and the latent variable

$$\xi_t = \alpha + \beta y_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2_{\epsilon|y}\mathbb{1})$$

  - ▷ But ξ_t not observable

$$\bar{X}_t = \alpha + \beta y_t + \eta_t \qquad \eta_t \sim \mathcal{N}(0, \sigma^2_{\bar{X}|y}\mathbb{1})$$

  - ▷ Bayesian specification of the linear regression with informative conjugate priors defining 4 hyperparameters
  - ▷ Bayesian point estimators α, β and $\sigma^2_{\bar{X}|y}$: a compromise between a priori and least squares estimates
  - ▷ Less information in $\bar{X}_t$ than in ξ_t: $\sigma^2_{\bar{X}|y}$ is an upper bound of $\sigma^2_{\epsilon|y}$
    - ➡ Optimal variance $\sigma^2_{\epsilon|y}$ estimated by minimizing the CRPS

➡ BPO framework:

$$p(y_t|\xi_t) = \mathcal{N}(y_t; \mu_{y|\xi}, \sigma^2_{y|\xi})$$

$$\mu_{y|\xi} = \frac{\sigma^2\beta^2(\frac{\xi_t-\alpha}{\beta}) + \sigma^2_{\epsilon|y}\mu}{\sigma^2\beta^2 + \sigma^2_{\epsilon|y}} \qquad \sigma^2_{y|\xi} = \frac{\sigma^2\sigma^2_{\epsilon|y}}{\sigma^2\beta^2 + \sigma^2_{\epsilon|y}}$$

Bayesian Processor of Ensemble Members

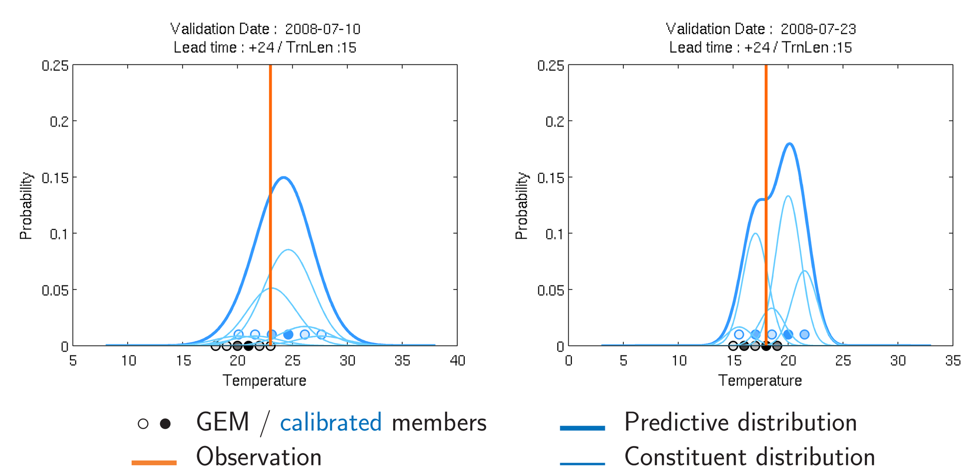
Combining the BMA and BPO frameworks

$$p(y_t|\mathbf{X}_t) \approx \frac{1}{S}\sum_{s=1}^{S}\mathcal{N}(y_t; \mu_{y|X_{t,s}}, \sigma^2_{y|X_{t,s}})$$

$$\mu_{y|X_{t,s}} = \frac{\sigma^2\beta^2(\frac{X_{t,s}-\alpha}{\beta}) + \sigma^2_{\epsilon|y}\mu}{\sigma^2\beta^2 + \sigma^2_{\epsilon|y}} \qquad \sigma^2_{y|X_{t,s}} = \frac{\sigma^2\sigma^2_{\epsilon|y}}{\sigma^2\beta^2 + \sigma^2_{\epsilon|y}}$$

- $\mu_{y|X_{t,s}}$: weighted average of bias-corrected member $\frac{X_{t,s}-\alpha}{\beta}$ and of the prior mean from climatology μ
- $\sigma^2_{y|X_{t,s}}$: weighted mixture of residuals variance of the linear model $\sigma^2_{\epsilon|y}$ and of the climatological variance σ^2
- The predictive distributions's shape depends on the empirical distribution of the ensemble members

➡ the predictive distribution is not necessary Gaussian



Identification of the Optimum Training Length

- training period: 10-days to 50-days joint samples
- No significant gain in forecasts's skill beyond 15 days
- Calibrated forecasts less skillful than climatology-based forecasts for longer-range lead times
- Optimal training length: 15 days
- Short training length: limit effects of seasonality and frequent changes to operational forecasting systems

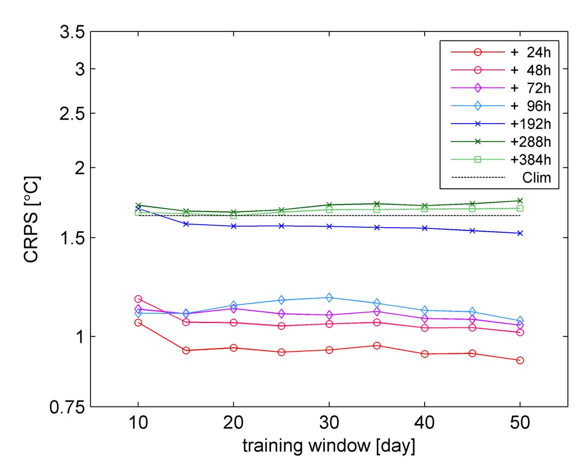
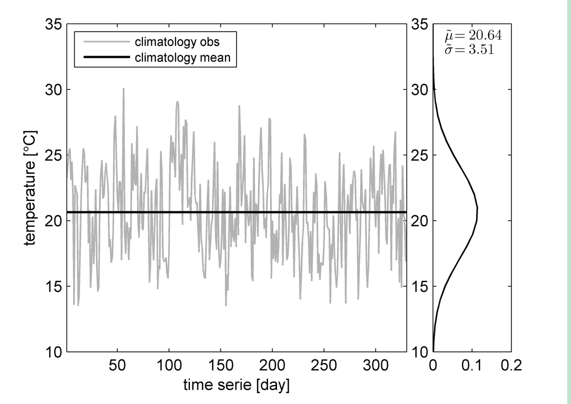


Illustration of the BPEM Method

24-hour ahead forecast valid the 10th of July 2008

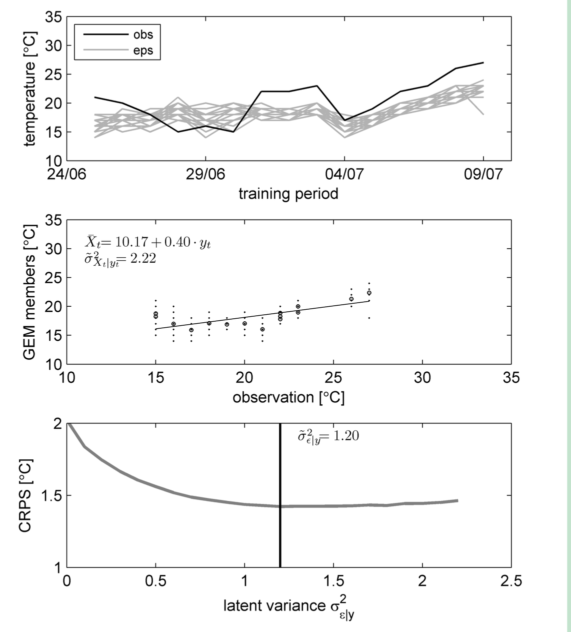
❶ Prior distribution

- Climatology: observations from the past 30 years (1978–2007) inside a moving window of 5 days around the valid date
- No seasonality effect nor climatological trend
- Estimated parameters: $\tilde{\mu} = 20.64$ and $\tilde{\sigma} = 3.51$



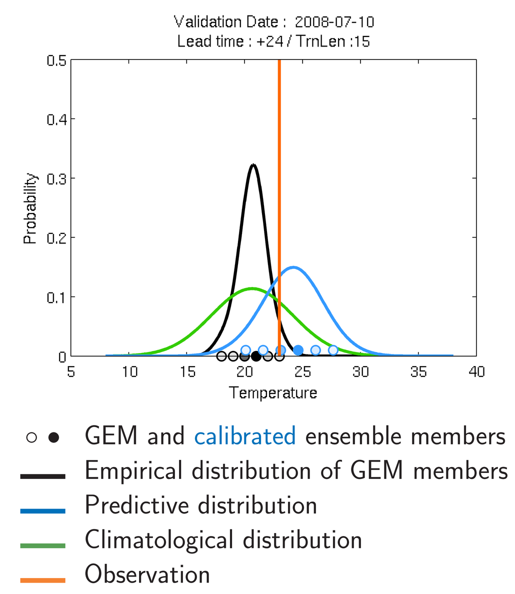
❷ Likelihood function

- Bayesian Linear Regression estimated from the joint sample (ensemble means and observations) over the training window
- Hyperparameters estimated from least squares linear regressions, in the same season, using joint samples of closed lead time with the same training length
- Estimated parameters: $\tilde{\alpha} = 10.17$, $\tilde{\beta} = 0.40$ and $\tilde{\sigma}^2_{\bar{X}|y} = 2.22$

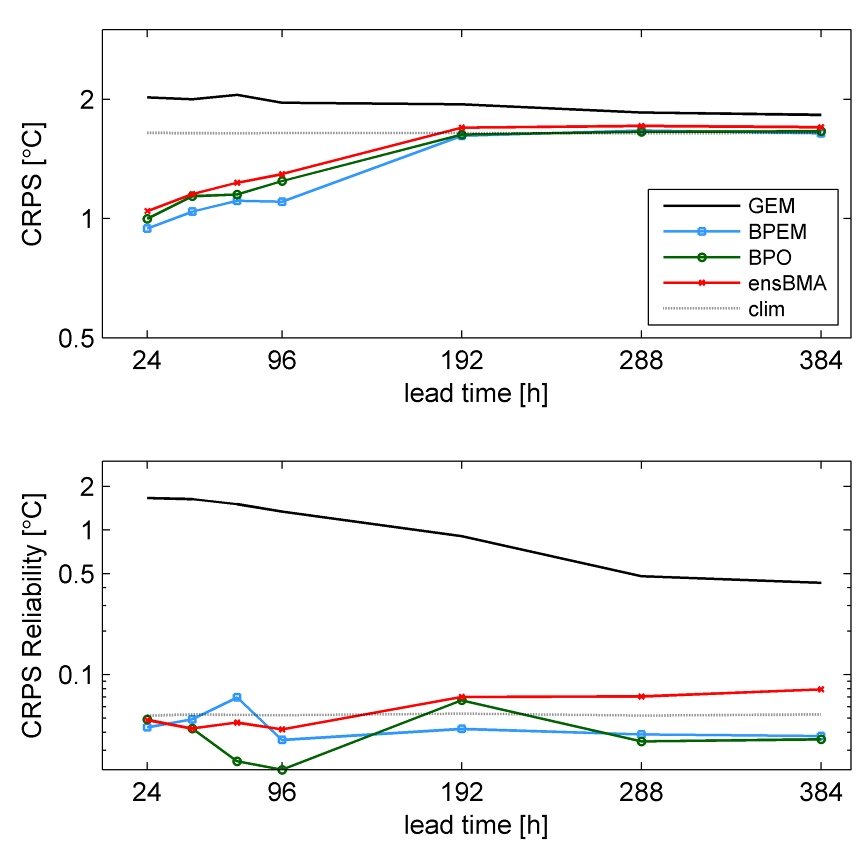


❸ Optimal error variance of the latent variable

- Standard error of the latent variable determined by grid search
- Minimisation of CRPS within the interval [0.0, 2.22]
- Optimal error variance: $\tilde{\sigma}^2_{\epsilon|y} = 1.20$



Verification of Calibrated Ensemble Forecasts



- Verification performed on the summer season (July and August 2008)
- Optimal training length for each calibration method
- Forecasts' skill improved by calibration (up to +192h) with similar pattern for each calibration method
- Significant improvment of forecasts' reliability

Conclusions on BPEM

... a new approach to calibrated ensemble forecasts

- Based on BMA and BPO frameworks
- Capable to generate reliable forecasts
- Outperforms slightly both the BMA and BPO approaches as well as a climatology
- Short optimal training length: avoid negative impacts of seasonality and of frequent changes to operational forecasting systems
- Successfully applied to 7 other stations accross the Quebec

Some References

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for ensemble prediction systems. *Weather and Forecasting 15*(5), 559–570.

Krzysztofowicz, R. (2004). Bayesian Processor of Output: a new technique for probabilistic weather forecasting. In *17th Conference on Probability and Statistics in the Atmospheric Sciences*, Volume 4.2, Seattle, Washington, USA. American Meteorological Society.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005). Using Bayesian Model Averaging to calibrate forecast ensembles. *Monthly Weather Review 133*(5), 1155–1174.