

Sequential aggregation of probabilistic experts – Application to ensemble wind speed forecasts

Michaël Zamo

Météo-France/CNRM
Liliane Bel, AgroParisTech
Olivier Mestre, Météo-France/CNRM

michael.zamo@meteo.fr

April 23, 2021

Overview

Sequential probabilistic expert aggregation

Application to ensemble wind speed forecasts

Framework ²

Framework

At times $t \in 1, \dots, T$ E experts $\hat{y}_{e;t}$ are used to forecast an observation $y_t \in \mathbb{R}$.

¹Allard et al. [2012]

²Cesa-Bianchi et al. [2006]

Framework ²

Framework

At times $t \in 1, \dots, T$ E experts $\hat{y}_{e;t}$ are used to forecast an observation $y_t \in \mathbb{R}$.

Combining several forecasts (“experts”) may allow to improve over each forecasts individually ¹. The aggregated forecast is a **linear combination** of the experts

$$\hat{y}_t = \sum_{e=1}^E \omega_{e;t} \hat{y}_{e;t}$$

¹Allard et al. [2012]

²Cesa-Bianchi et al. [2006]

Framework ²

Framework

At times $t \in 1, \dots, T$ E experts $\hat{y}_{e;t}$ are used to forecast an observation $y_t \in \mathbb{R}$.

Combining several forecasts (“experts”) may allow to improve over each forecasts individually ¹. The aggregated forecast is a **linear combination** of the experts

$$\hat{y}_t = \sum_{e=1}^E \omega_{e;t} \hat{y}_{e;t}$$

Sequential expert aggregation:

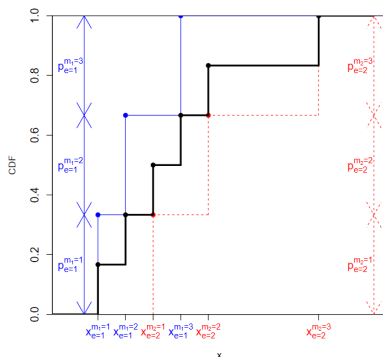
- adaptive weightings: uses only information available at time t :
 $\hat{y}_{e;t'} \ \forall t' \leq t$ and $y_{t'} \ \forall t' < t$
- theoretical performance theorems

¹Allard et al. [2012]

²Cesa-Bianchi et al. [2006]

Application to probabilistic experts³

Each expert e is a **step-wise cumulative distribution function** (CDF) with M_e steps of height $p_e^{m_e}$ at $x_{e;t}^{m_e}$



CRPS

The loss function is the Continuous Ranked Probability Score (CRPS⁴)

$$\ell(\hat{F}, y) = \int_{x \in \mathbb{R}} \left(\hat{F}(x) - H(x - y) \right)^2 dx$$

The lower the better.

The CRPS is **convex with its first argument**, so theoretical bounds may apply.

⁴Matheson and Winkler [1976]

Aggregated forecast

The weights must be **convex** $\omega_{e;t} \geq 0 \ \forall t, e$ and $\sum_{e=1}^E \omega_{e;t} = 1 \ \forall t$

The aggregated forecast is a valid stepwise CDF with $\sum_{e=1}^E M_e$ steps of height $\omega_{e;t} p_e^{m_e}$ at $x_{e;t}^{m_e}$

$$\hat{y}_t = \hat{F}_t(x) = \sum_{e=1}^E \sum_{m_e=1}^{M_e} \omega_{e;t} p_e^{m_e} H(x - x_{e;t}^{m_e})$$

This generalises Baudin [2015] and Thorey et al. [2015] (where $M_e = 1 \ \forall e \in 1, \dots, E$).

28 experts: 4 raw ensembles⁵

- Canada (CMC)
- Europe (ECMWF)
- France (MF, PEARP)
- USA (NCEP).

4 years, 0.5° grid, 267 grid points over France.

For each raw ensemble, $\hat{y}_{e;t}$ is the empirical CDF.

⁵Swinbank et al. [2016]

28 experts: 24 post-processed ensembles⁸

Each ensemble is post-processed with 6 machine learning methods (or version)

- Quantile regression forest (quantregForest)⁶, trained over three years (cross-validation).
- Non-homogeneous regression (NR)⁷: trained over a sliding window (5 versions: 7, 30, 90, 365, and $t - 1$ previous days). Parametric distribution, transformed in a step-wise CDF.

⁶Meinshausen [2006]

⁷Gneiting [2014]

⁸Taillardat et al. [2016]

Observation

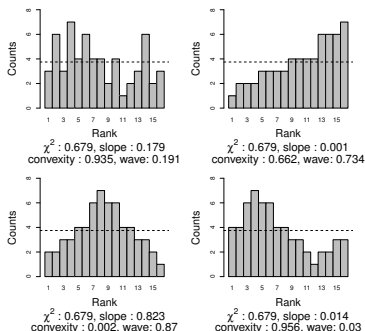
Hourly analysis combining AROME analysis and observations at station locations.

10-m height wind speed and gust.⁹.

⁹Zamo et al. [2016]

Verification tools

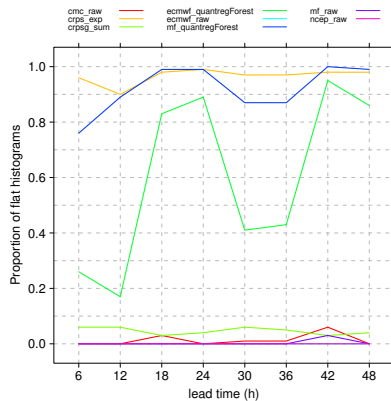
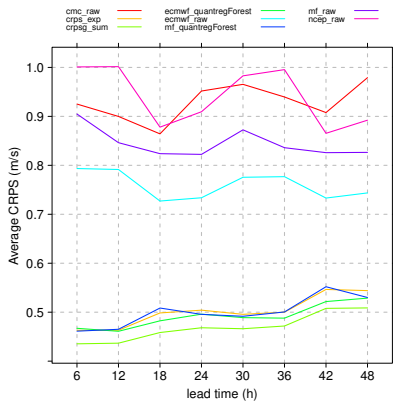
- CRPS: global measure of performance¹⁰ == most skillful forecast
- Flatness of the rank histogram: Jolliffe-Primo tests¹¹ == most "reliable" forecast.



¹⁰Hersbach [2000]

¹¹Jolliffe and Primo [2008]

Improvement of the CRPS and the flatness



Conclusion

- adaptation of the sequential aggregation framework to probabilistic forecasts
- illustration on ensemble wind speed forecasts
- verification methodology: added-value to used several verification tools (CRPS and Jolliffe-Primo tests)

More details in: Zamo, Michaël, Liliane Bel, and Olivier Mestre.
"Sequential aggregation of probabilistic forecasts—application to wind speed ensemble forecasts." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 70.1 (2021): 202-225.
<https://arxiv.org/pdf/2005.03540.pdf>

Thanks! Merci! Danke Schön! Grazie mille! etc.

References I

- D. Allard, A. Comunian, and P. Renard. Probability aggregation methods in geoscience. *Mathematical Geosciences*, 44(5): 545–581, 2012.
- Paul Baudin. *Prévision séquentielle par agrégation d'ensemble: application à des prévisions météorologiques assorties d'incertitudes*. PhD thesis, Université Paris-Saclay, 2015.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.
- N. Cesa-Bianchi, G. Lugosi, et al. *Prediction, learning, and games*, volume 1. Cambridge University Press Cambridge, 2006.

References II

- T. Gneiting, F. Balabdaoui, and A.E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2): 243–268, 2007.
- Tilmann Gneiting. Calibration of medium-range weather forecasts. In *Technical Memoranda 719*. ECMWF, 2014.
- H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.
- Ian T Jolliffe and Cristina Primo. Evaluating rank histograms using decompositions of the chi-square test statistic. *Monthly Weather Review*, 136(6):2133–2139, 2008.

References III

Vivien Mallet, Boris Mauricette, and Gilles Stoltz. Description of Sequential Aggregation Methods and their Performances for Ozone Ensemble Forecasting. *Technical report, École normale supérieure, DMA and CEREa*, 2007.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management science*, 22 (10):1087–1096, 1976.

N. Meinshausen. Quantile regression forests. *The Journal of Machine Learning Research*, 7:983–999, 2006.

Tim Palmer and David Richardson. *Decisions, decisions...!*, pages 12–14. Number 141 in ECMWF Newsletter. European Center for Medium Range Weather Forecasts, Summer 2014.

References IV

- Richard Swinbank, Masayuki Kyouda, Piers Buchanan, Lizzie Froude, Thomas M Hamill, Tim D Hewson, Julia H Keller, Mio Matsueda, John Methven, Florian Pappenberger, et al. The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97(1):49–67, 2016.
- Maxime Taillardat, Olivier Mestre, Michaël Zamo, and Philippe Naveau. Calibrated Ensemble Forecasts using Quantile Regression Forests and Ensemble Model Output Statistics. *Monthly Weather Review*, 144(6):2375–2393, 2016.
- Jean Thorey, Vivien Mallet, Christophe Chaussin, Laurent Descamps, and Philippe Blanc. Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database. *Solar Energy*, 120:232–243, 2015.

References V

Michaël Zamo. *Statistical Post-processing of Deterministic and Ensemble Wind Speed Forecasts on a Grid*. PhD thesis, AgroParisTech, 2017. (in English).

Michaël Zamo, Liliane Bel, Olivier Mestre, and Joël Stein. Improved gridded windspeed forecasts by statistical post-processing of numerical models with block regression. *Weather and Forecasting*, 31(6):1929–1945, 2016.

The End

Framework: theoretical bounds¹³

Let $\ell(\hat{y}, y) \in \mathbb{R}$ be a loss function.

If ℓ is **convex** with its first argument (i.e. experts), some algorithms to compute the weights $\omega_{e;t}$ ¹² have an upper bound of the cumulative loss respectively to an oracle $\bar{y}$ (**regret**)

$$\sup_{y_{t=1}, \dots, T} \mathcal{R}(\bar{y}, \hat{y}) = \sup_{y_{t=1}, \dots, T} \left(\sum_{t=1}^T \ell(\hat{y}_t, y_t) - \sum_{t=1}^T \ell(\bar{y}_t, y_t) \right) \leq O(T)$$

The oracle $\bar{y}_t$ is computed with all the information on $t = \{1, \dots, T\}$.

The supremum is over **all possible sequences** of observations.

¹²Mallet et al. [2007]¹³Cesa-Bianchi et al. [2006]

Example: exponentially weighted average forecaster (EWA)

$$\omega_{e;t}^{EWA} = \frac{\exp\{-\eta CRPS_{e;t}(W)\}}{\sum_{e=1}^E \exp\{-\eta CRPS_{e;t}(W)\}},$$

with $\eta \in \mathbb{R}^+$ the learning rate

$CRPS_{e;t}(W)$ the loss of expert e cumulated over the last W forecasts.

Example: EWA

If $W = t - 1$

$$\sup_{y_{t=1,\dots,T}} \mathcal{R}(\bar{y}, \hat{y}) \leq \frac{\ln E}{\eta} + \frac{\eta T}{8} B^2$$

where B is the upper bound of the loss.

The oracle is the best of the E experts (in terms of mean CRPS over the whole period $\{1, \dots, T\}$).

Jolliffe-Primo flatness tests

Let K ranks with counts n_k .

$$\boldsymbol{\delta} = \left(\frac{n_1 - n_{th}}{\sqrt{n_{th}}}, \dots, \frac{n_K - n_{th}}{\sqrt{n_{th}}} \right)'$$

is the vector of normalized deviation from flatness. Under the hypothesis H_0 : *the ranks are uniformly distributed*

$$\|\boldsymbol{\delta}\|^2 \sim \chi_{K-1}^2.$$

Let $\mathbf{u}_{i=1,\dots,K}$ an orthonormal basis of $\mathbb{R}^K$, then under H_0

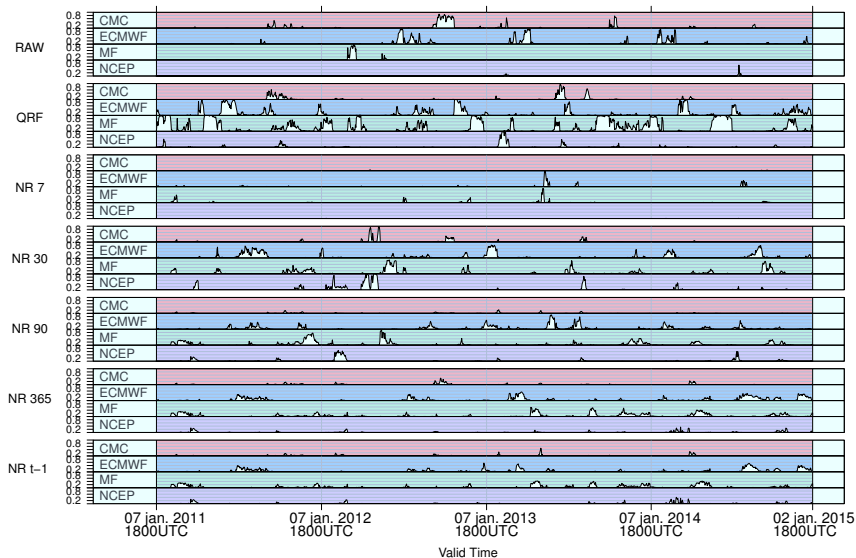
$$\boldsymbol{\delta} \cdot \mathbf{u}_i \sim \chi_1^2$$

for each $i \in \{1, \dots, K-1\}$.

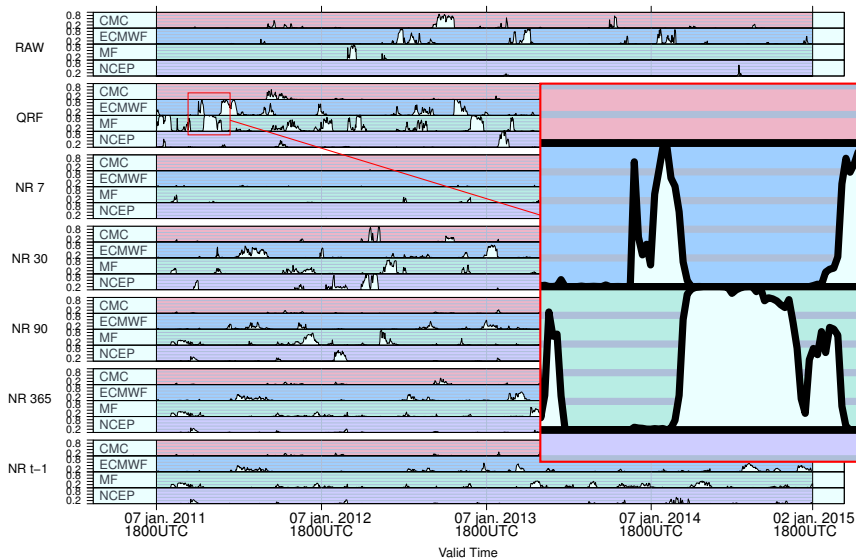
Multiple testing is taken into account thanks to the Benjamini-Hochberg procedure¹⁴.

¹⁴Benjamini and Hochberg [1995]

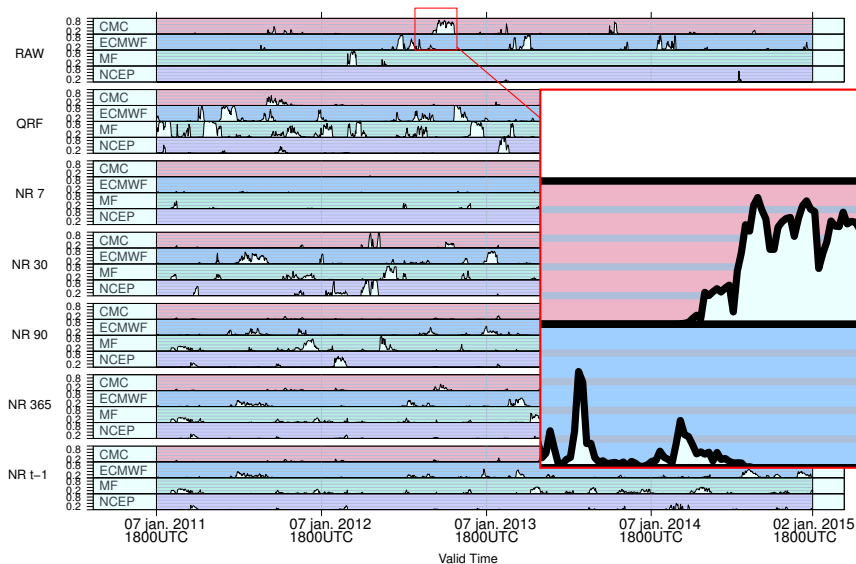
Evolution of the aggregation weights



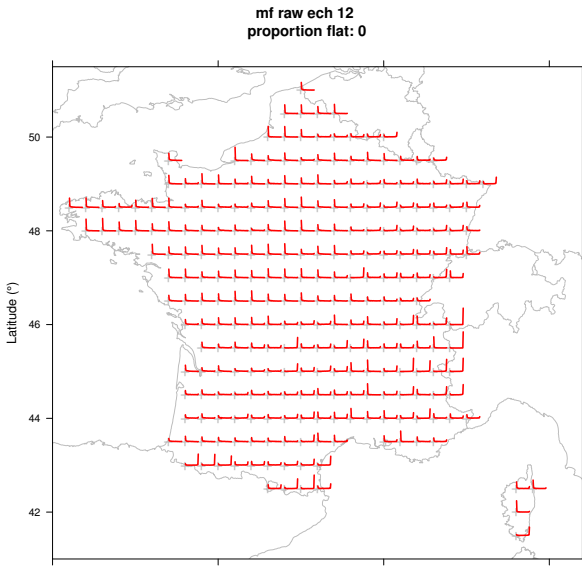
Evolution of the aggregation weights



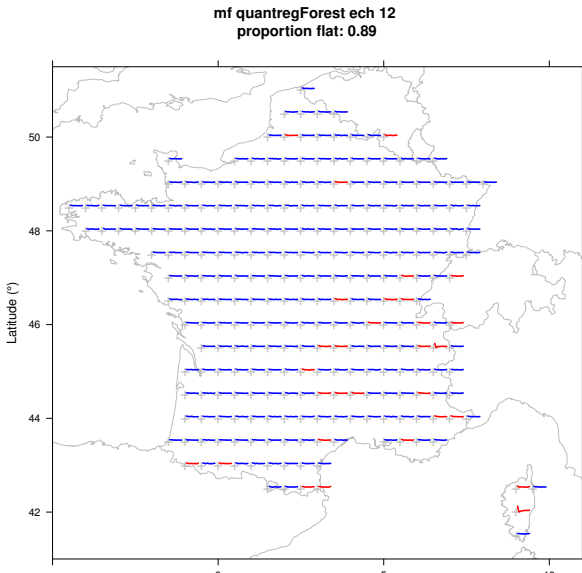
Evolution of the aggregation weights



Rank histograms: RAW MF ENSEMBLE

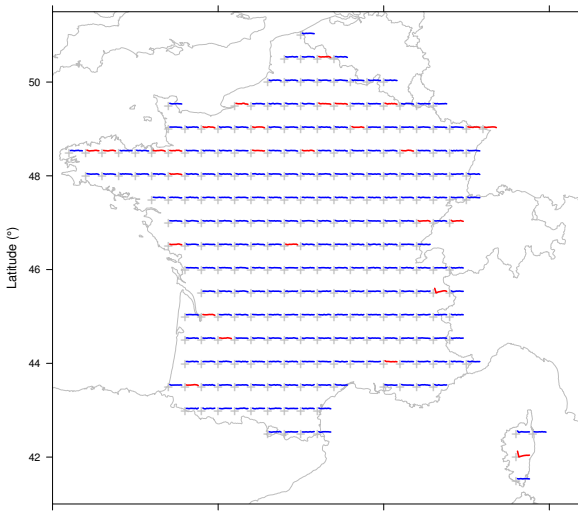


Rank histograms: MF quantregForest



Rank histograms: aggregation with EWA

crps exp eta 3.16227766016838 win 30 ech 12
proportion flat: 0.9



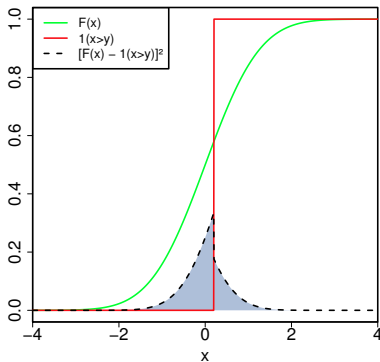
Probabilistic forecast verification paradigm: maximize sharpness under calibration [Gneiting et al., 2007]

Continuous ranked probability score (CRPS)

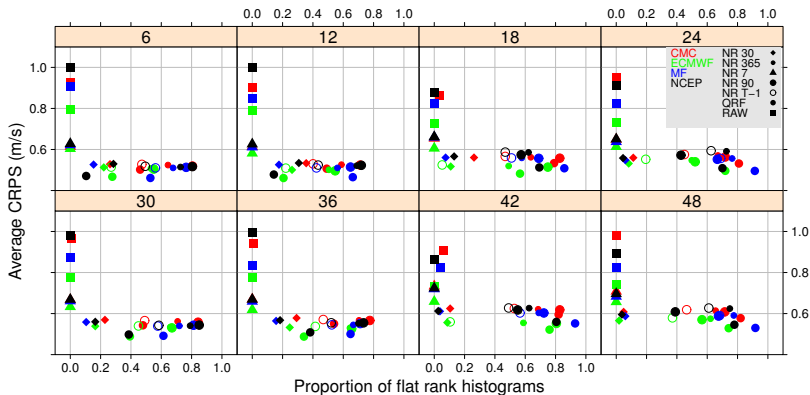
For a forecast cumulative distribution function F of an observation y ,

$$CRPS(F, y) = \int_{x \in \mathbb{R}} [F(x) - \mathbb{1}(x > y)]^2 dx \in [0; +\infty).$$

Its mean is linked to the overall economic value of the forecast¹⁵.



Relationship between CRPS and JP-test



Decomposition of the CRPS

$$\overline{CRPS(F, y)} = \textit{Reliability} - \textit{Resolution} + \textit{Uncertainty}$$

When $\textit{Reliability} \sim 0$, the CRPS is mainly a measure of resolution.

Relationship between *Reliability* and JP-test

