

# Agrégation séquentielle d'experts probabilistes application à la prévision d'ensemble du vent

Michaël Zamo, Météo-France

with

Liliane Bel, AgroParisTech  
Olivier Mestre, Météo-France

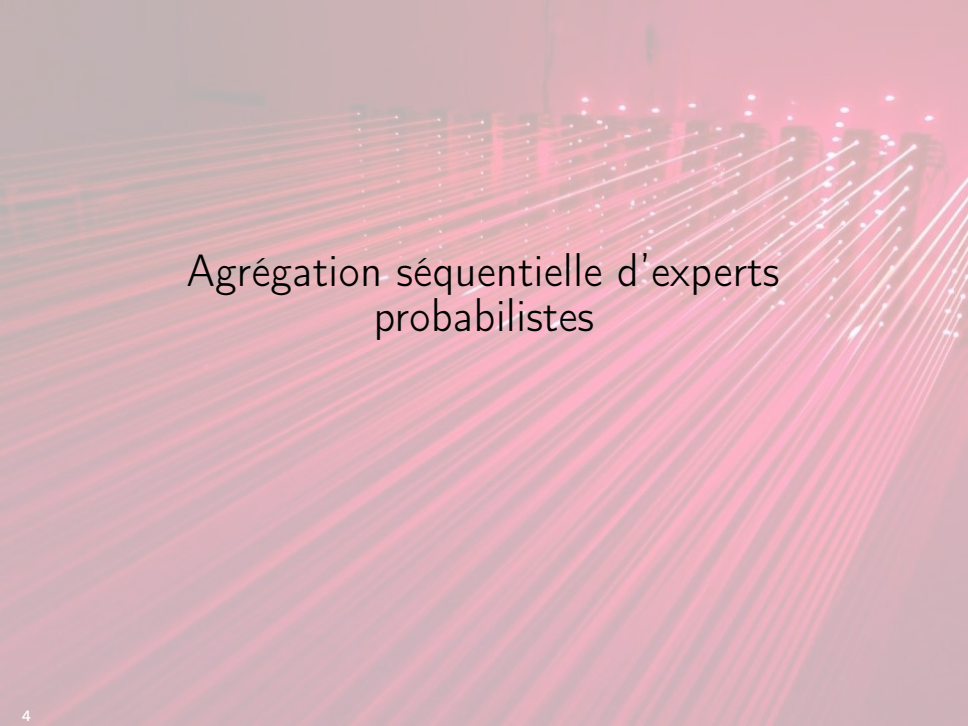
Combiner plusieurs prévisions permet d'exploiter au mieux plusieurs ensembles (post-traités ou pas)

Pourquoi l'agrégation séquentielle d'experts

- ▶ poids adaptatifs
- ▶ existence de garanties théoriques de performance
- ▶ gérer les changements des codes numériques des ensembles

Agrégation séquentielle d'experts probabilistes

Application aux prévisions d'ensemble de vent



# Agrégation séquentielle d'experts probabilistes

Des informations arrivent séquentiellement à des instants  
 $t \in 1, \dots, T$ ,

- ▶  $E$  “experts” (prévisions)  $\hat{y}_{e;t}$
- ▶ une observation  $y_t \in \mathbb{R}$ .

On produit une nouvelle prévision par combinaison des experts

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

Agrégation séquentielle : utilisation de l'information disponible à  $t$  :  
seulement  $\hat{y}_{e;t'} \forall t' \leq t$  et  $y_{t'} \forall t' < t$ .

---

1. Cesa-Bianchi et al. [2006]

Soit une fonction de perte  $\ell(\hat{y}, y) \in \mathbb{R}$ .

Si  $\ell$  est convexe en son premier argument, certains algorithmes<sup>2</sup> de calcul des poids  $\omega_{e;t}$  induisent une borne de la perte cumulée par rapport à celle d'un oracle  $\bar{y}$  (regret)

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

où l'oracle  $\bar{y}_t$  est calculé avec toute l'information sur la période complète.

---

2. Mallet et al. [2007]

3. Cesa-Bianchi et al. [2006]

Chaque expert  $e$  est une fonction de répartition (CDF) en escalier avec  $M_e$  marches de hauteur  $p_e^{m_e}$  en  $x_{e;t}^{m_e}$

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

La fonction de perte utilisée est le CRPS (continuous ranked probability score,<sup>4</sup>)

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

Le CRPS est convexe en son premier argument, ce qui assure l'existence des bornes pour certains algorithmes d'agrégation.

---

4. Matheson and Winkler [1976]

5. Zamo [2017]

Les poids d'agrégation doivent vérifier  $\omega_{e;t} \geq 0 \forall t, e$  et

$$\sum_{e=1}^E \omega_{e;t} = 1 \forall t$$

La prévision agrégée est une CDF en escalier est une fonction de répartition (CDF) en escalier avec  $\sum_{e=1}^E M_e$  marches de hauteur  $\omega_{e;t} p_e^{m_e}$  en  $x_{e;t}^{m_e}$

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

Cela généralise Baudin [2015], Thorey et al. [2015] qu'on retrouve en posant  $M_e = 1 \quad \forall e \in 1, \dots, E$ .



ou *exponentially weighted average forecaster* (EWA).

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


avec  $\eta \in \mathbb{R}^+$  le taux d'apprentissage et  $CRPS_{e;t}(W)$  la perte cumulée de l'expert  $e$  sur les  $W$  dernières prévisions.

Si  $W$  couvre toute la période avant  $t$  ( $W = t - 1$ )

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

où  $B$  est la borne supérieure (pratique) de la perte.

L'oracle est le meilleur des  $E$  experts (en terme de CRPS moyen) sur la période  $\{1, \dots, T\}$ .



# Application aux prévisions d'ensemble de vent

France

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

4 années, grille de  $0.5^\circ$ , 267 points de grille.

Pour chaque ensemble,  $\hat{y}_{e;t}$  est la CDF empirique.

---

6. Swinbank et al. [2016]

Six versions post-traitées de chaque ensemble (cf. présentation de M. Taillardat)

- ▶ Forêts de régression quantile (quantregForest)<sup>7</sup>, ajustées par validation croisée (3 ans d'apprentissage, un an de test).
- ▶ Régression non homogène (NR)<sup>8</sup> : apprentissage avec fenêtres glissantes (5 versions : 7, 30, 90, 365, et  $t - 1$  jours précédents). Discrétisée pour avoir des CDFs en escalier.

---

7. Meinshausen [2006]

8. Gneiting [2014]

9. Taillardat et al. [2016]

Analyse horaire combinant prévisions d'AROME et observations en points stations, sur la France.

Vent moyen à dix mètres, rafales horaires<sup>10</sup>.

---

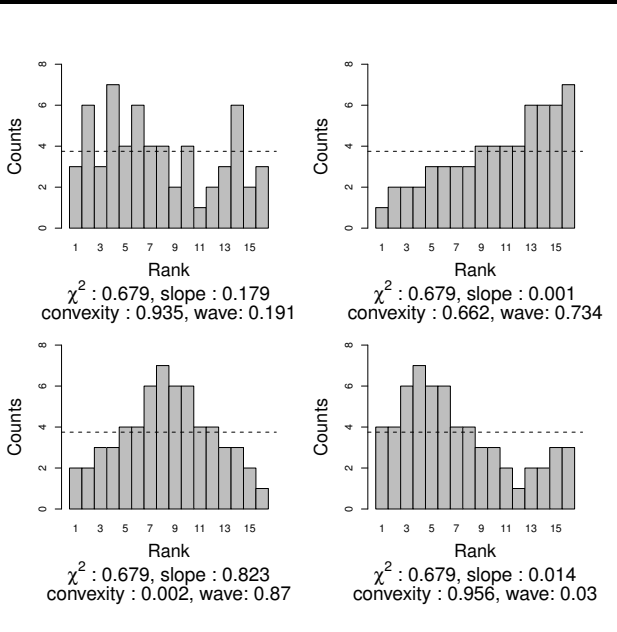
10. Zamo et al. [2016]

- ▶ CRPS : mesure globale de performance<sup>11</sup> ;
- ▶ Platitude de l'histogramme des rangs : tests de Jolliffe-Primo<sup>12</sup>.

---

11. Hersbach [2000]

12. Jolliffe and Primo [2008]





Soient  $k$  rangs peuplés avec  $n_k$  observations.

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

est le vecteur des écarts à la platitude. Alors sous l'hypothèse  $H_0$  que l'histogramme des rangs est issu d'une prévision fiable

$$\|\boldsymbol{\delta}\|^2 \sim \chi_k^2.$$

Soit  $\mathbf{u}_{i=1,\dots,k}$  une base orthonormée de  $\mathbb{R}^k$ , alors sous l'hypothèse  $H_0$

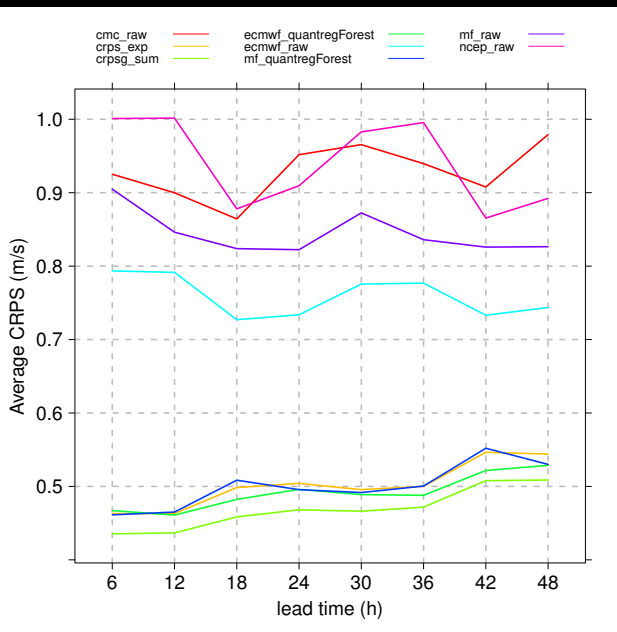
$$\boldsymbol{\delta} \cdot \mathbf{u}_i \sim \chi_1^2$$

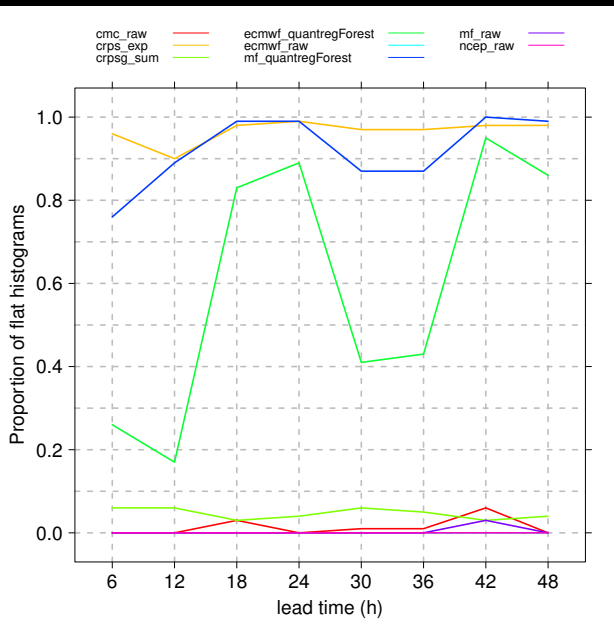
pour  $i \in \{1, \dots, k-1\}$ .

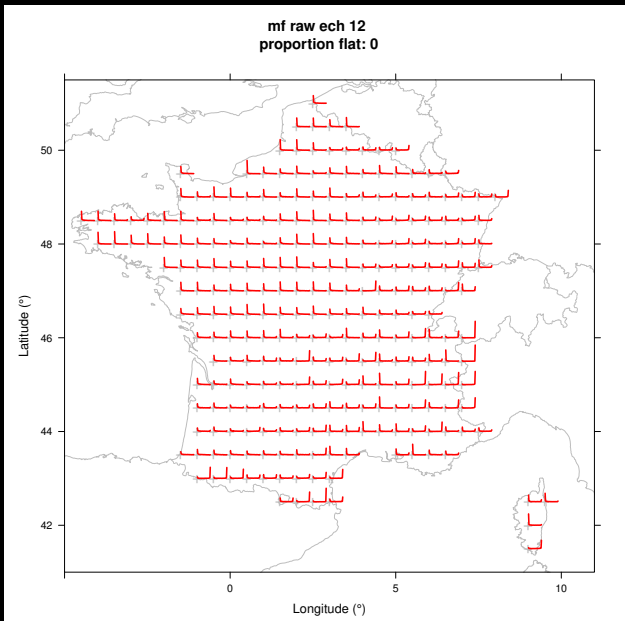
Prise en compte des tests multiples par le procédure de Benjamini-Hochberg<sup>13</sup>.

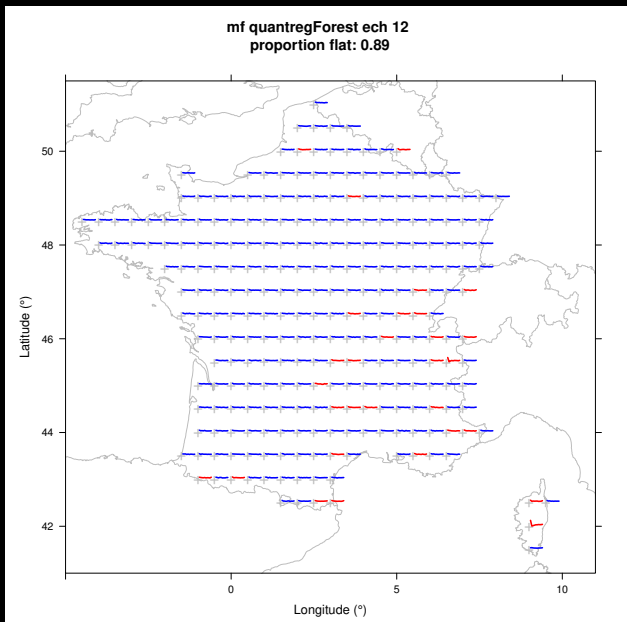
---

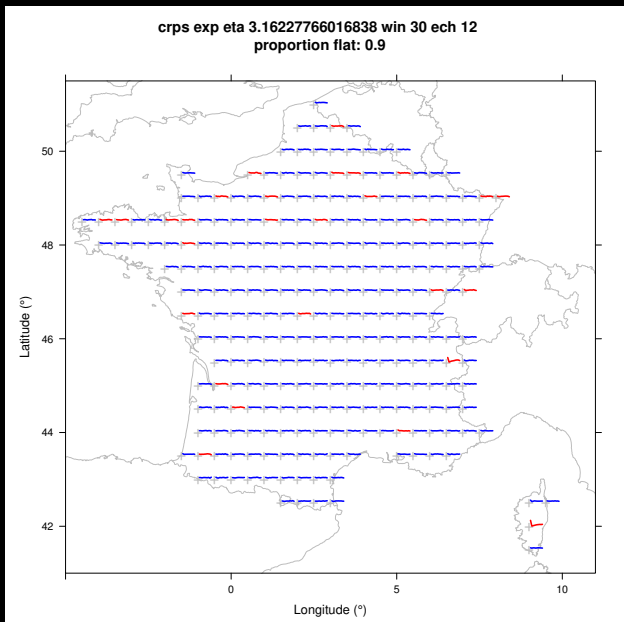
13. Benjamini and Hochberg [1995]



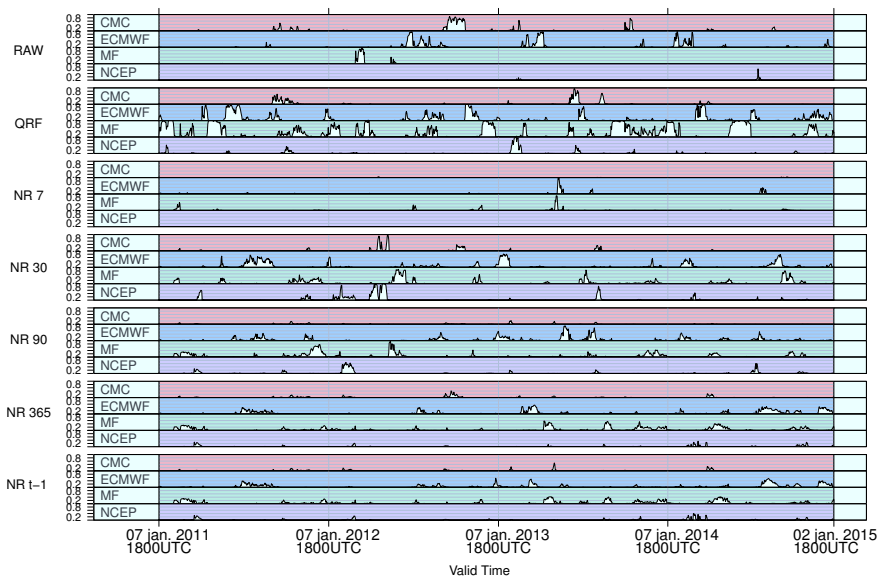




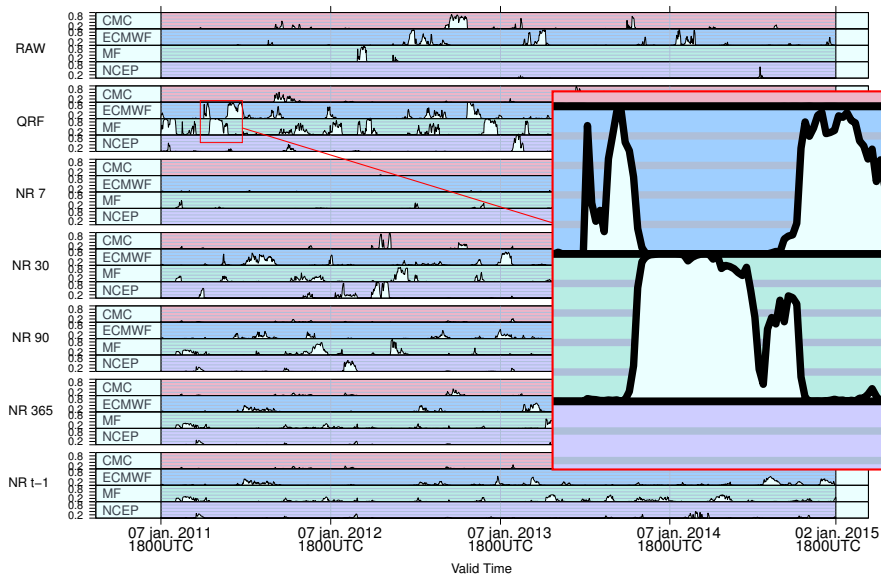




# Evolution des poids d'agrégation

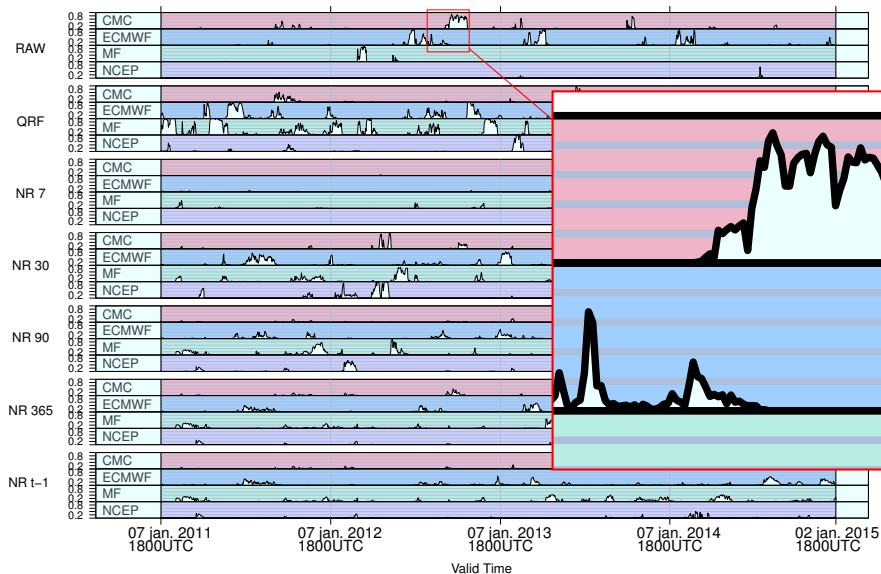


# Evolution des poids d'agrégation





# Evolution des poids d'agrégation



- ▶ adaptation du cadre de l'agrégation séquentielle d'experts aux prévisions probabilistes
- ▶ illustration sur des prévisions de vent
- ▶ nécessité d'employer plusieurs outils de vérification (CRPS et tests de Jolliffe-Primo)

A noter : article soumis à JRSS-C

Merci

- 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.
- 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.
- 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 CEREAs*, 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.

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.

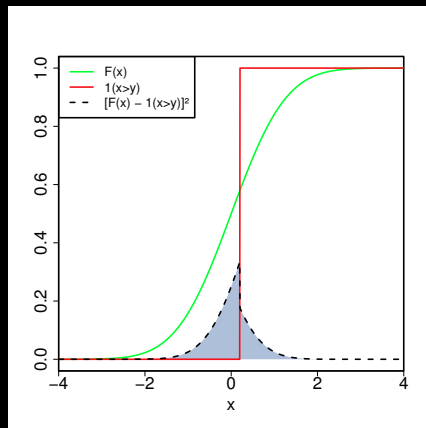
Michaël Zamo. *Statistical Post-processing of Deterministic and Ensemble Wind Speed Forecasts on a Grid*. PhD thesis, AgroParisTech, 2017. (in English).

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.

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

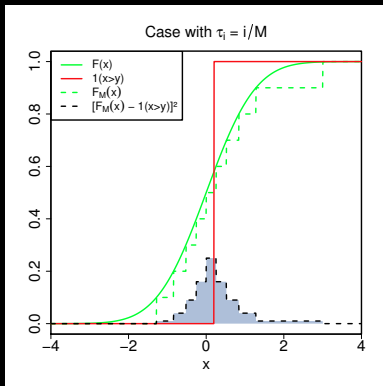
$$CRPS(F, y) = \int_{x \in \mathbb{R}} [F(x) - \mathbb{1}(x > y)]^2 dx \in [0; +\infty).$$

Its mean is linked to the overall economic value of the forecast <sup>14</sup>.



In quantile regression,  $F$  is known only through a set of  $M$  values  $x_i$  with order  $\tau_i$ , with  $i \in \{1, \dots, M\}$ .  $\text{CRPS}(F, y)$  is estimated with

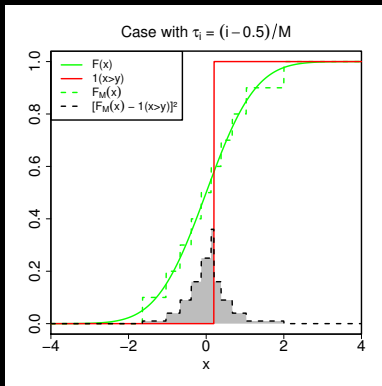
$$\widehat{\text{crps}}_{INT}(M, y) = \int_{\mathbb{R}} \left[ \frac{1}{M} \sum_{i=1}^M \mathbb{1}(x \geq x_i) - \mathbb{1}(x \geq y) \right]^2 dx.$$

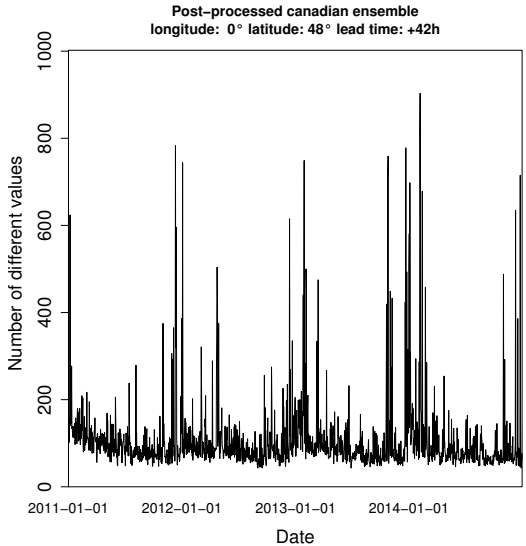


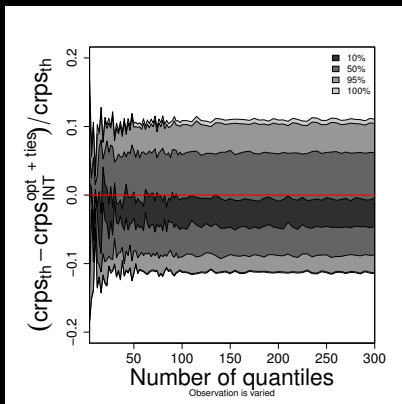
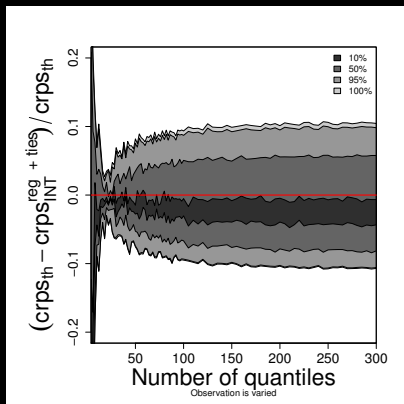


In quantile regression,  $F$  is known only through a set of  $M$  values  $x_i$  with order  $\tau_i$ , with  $i \in \{1, \dots, M\}$ .  $\text{CRPS}(F, y)$  is estimated with

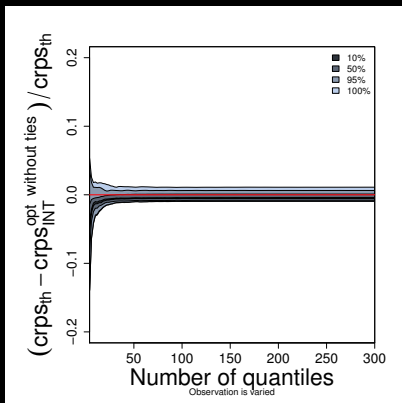
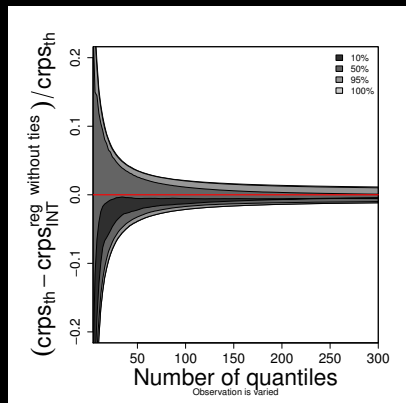
$$\widehat{\text{crps}}_{INT}(M, y) = \int_{\mathbb{R}} \left[ \frac{1}{M} \sum_{i=1}^M \mathbb{1}(x \geq x_i) - \mathbb{1}(x \geq y) \right]^2 dx.$$







# Case of quantile regression with ties removed



More details in .