Verification of extremes using proper scoring rules and extreme value theory

Maxime Taillardat[1,2,3]

A-L. Fougères[3],
P. Naveau[2] and
O. Mestre[1]

[1] CNRM/Météo-France [2] LSCE [3] ICJ

May 8, 2017

Plan

Verification & extremes : a challenging issue

- **Verification habits**
  - Set of observed events and associated forecasts
  - Standard verification methods applied on the set
- **But for extremes**
  - Small number of observed events
  - Standard verification methods degenerate
  - Models (even ensemble forecasts) are usually quite bad
- Misguided inferences/assessments : *The forecaster's dilemma* (see Sebastian's talk)

Plan

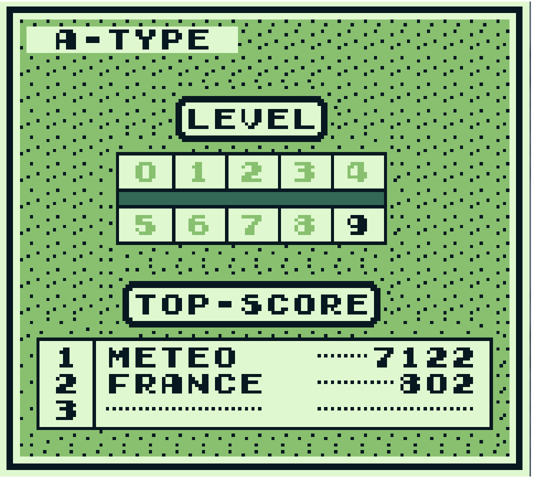
1 Extremes : difficult to forecast... and to verify

2 Weighted CRPS for extremes

3 Extreme Value Theory and CRPS distribution

4 A relevant case study

Proper scoring rules



Proper scoring rules

- Y : observation with CDF G (unknown...)
- X forecast with CDF F
- $s(.,.)$ function of $\mathcal{F} \times \mathbb{R}$ in $\mathbb{R}$

s is a proper scoring rule (Murphy 1968 ; Gneiting 2007)

$$\mathbb{E}_Y(s(G,Y)) \leq \mathbb{E}_Y(s(F,Y)) \quad (1)$$

Proper scoring rules

- Y : observation with CDF G (unknown...)
- X forecast with CDF F
- $s(.,.)$ function of $\mathcal{F} \times \mathbb{R}$ in $\mathbb{R}$

s is a proper scoring rule (Murphy 1968 ; Gneiting 2007)

$$\mathbb{E}_Y(s(G, Y)) \leq \mathbb{E}_Y(s(F, Y)) \tag{1}$$

The CRPS...

- A widely used proper score : the CRPS (Murphy 1969 ; Gneiting and Raftery 2007 ; Naveau et al. 2015 ; Taillardat et al. 2016)

$$
\begin{aligned}
CRPS(F, y) &= \int_{-\infty}^{\infty} (F(x) - \mathbf{1}\{x \geq y\})^2 \, \mathrm{d}x \\
&= \mathbb{E}_F|X - y| - \frac{1}{2}\mathbb{E}_F|X - X'| \\
&= y + 2\left[\overline{F}(y)\mathbb{E}_F(X - y|X > y) - \mathbb{E}_F(XF(X))\right] \\
&= \mathbb{E}_F|X - y| + \mathbb{E}_F(X) - 2\mathbb{E}_F(XF(X))
\end{aligned}
$$

... And its weighted derivation

- A weighted score : the wCRPS (Gneiting and Ranjan 2012)

$$
\begin{aligned}
wCRPS(F,y) &= \int_{-\infty}^{\infty} w(x)(F(x) - \mathbf{1}\{x \geq y\})^2 \, \mathrm{d}x \\
&= \mathbb{E}_F|W(X) - W(y)| - \frac{1}{2}\mathbb{E}_F|W(X) - W(X')| \\
&= W(y) + 2\left[\overline{F}(y)\mathbb{E}_F(W(X) - W(y)|X > y) - \mathbb{E}_F(W(X)F(X))\right] \\
&= \mathbb{E}_F|W(X) - W(y)| + \mathbb{E}_F(W(X)) - 2\mathbb{E}_F(W(X)F(X))
\end{aligned}
$$

where $W = \int w$ and $\int wf < \infty$

- The weight function cannot depend on the observation : it leads to improper scores.

(Weighted) CRPS embarassing properties

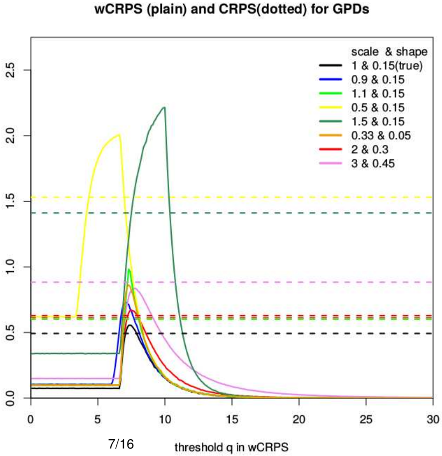
$$w_q(x) = \log(x)\mathbf{1}\{x \geq q\}$$

This weight function is closely linked to the Hill's tail-index estimator.

(Weighted) CRPS embarassing properties

$$w_q(x) = \log(x)\mathbf{1}\{x \geq q\}$$

This weight function appears suitable for extremes but...



(Weighted) CRPS embarassing properties II

- Tail equivalence

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c \in (0, \infty)$$

- For any given $\epsilon > 0$, it is always possible to construct a CDF F that is **not tail equivalent** to G and such that

$$|\mathbb{E}_Y(wCRPS(G, Y)) - \mathbb{E}_Y(wCRPS(F, Y))| \leq \epsilon$$

(Weighted) CRPS embarassing properties II

- Tail equivalence

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c \in (0, \infty)$$

- For any given $\epsilon > 0$, it is always possible to construct a CDF F that is **not tail equivalent** to G and such that

$$|\mathbb{E}_Y(wCRPS(G, Y)) - \mathbb{E}_Y(wCRPS(F, Y))| \leq \epsilon$$

Paradigm of verification for extremes ?

"The paradigm of maximizing the sharpness of the predictive distributions subject to calibration" (Gneiting et al. 2006)

"Extreme events are often the result of some extreme atmospheric conditions and combinations : Most of the time just few members in the ensemble leads to such events. We could just look at the information brought by the forecast. But how ?"

- **Consequence** : we do not care about reliability here ! (More in "detection" logic)
- An example : The ROC Curve
- **Different criterion** : Be skillful for extremes subject to a good overall performance.
- **Question** : How combining an extreme verification tool with the CRPS ?

Plan

How using extreme value theory with CRPS ?

The classical Continuous Ranked Probability Score (CRPS) can be written as :

$$CRPS(F,y) = \mathbb{E}|X-y| + \mathbb{E}(X) - 2\mathbb{E}(XF(X))$$

And for large y it is possible to show that :

$$CRPS(F,y) \approx y - 2\mathbb{E}(XF(X))$$

Pickands-Balkema-De Haan Theorem (1974-1975)

If the observed value is viewed as a random draw Y with CDF G, the survival distribution of $CRPS(F,Y)$ can be approximated by a GPD with parameters σ_G and ξ_G :

$$\mathbb{P}(CRPS(F,Y) > t + u \,|\, CRPS(F,Y) > u) \sim GP_t(\sigma_G, \xi_G)$$

How using extreme value theory with CRPS ?

The classical Continuous Ranked Probability Score (CRPS) can be written as :

$$CRPS(F, y) = \mathbb{E}|X - y| + \mathbb{E}(X) - 2\mathbb{E}(XF(X))$$

And for large y it is possible to show that :

$$CRPS(F, y) \approx y - 2\mathbb{E}(XF(X))$$

Pickands-Balkema-De Haan Theorem (1974-1975)

If the observed value is viewed as a random draw Y with CDF G, the survival distribution of $CRPS(F, Y)$ can be approximated by a GPD with parameters σ_G and ξ_G :

$$\mathbb{P}(CRPS(F, Y) > t + u \,|\, \mathbf{Y} > u) \sim GP_{t'}(\sigma_G, \xi_G)$$

Under assumptions on G (satisfied for extremes)

And so what ?

Pickands-Balkema-De Haan Theorem (1974-1975)

If the observed value is viewed as a random draw Y with CDF G, the survival distribution of $CRPS(F, Y)$ can be approximated by a GPD with parameters σ_G and ξ_G :

$$\mathbb{P}(CRPS(F, Y) > t + u \,|\, \mathbf{Y} > u) \sim GP_{t'}(\sigma_G, \xi_G)$$

Under assumptions on G (satisfied for extremes)

- Are we trapped ? Parameters are the same whatever the forecast
- **Crucial (and unrealistic) assumption here** : F and G are independent
- In practice, the convergence to these parameters is driven by the skill of ensembles for extreme events

And so what ?

Pickands-Balkema-De Haan Theorem (1974-1975)

If the observed value is viewed as a random draw Y with CDF G, the survival distribution of $CRPS(F, Y)$ can be approximated by a GPD with parameters σ_G and ξ_G :

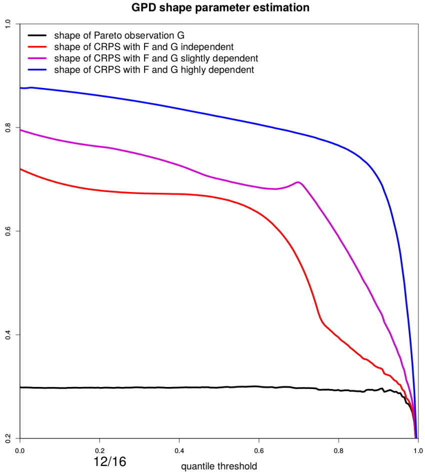
$$\mathbb{P}(CRPS(F, Y) > t + u \,|\, \mathbf{Y} > u) \sim GP_{t'}(\sigma_G, \xi_G)$$

Under assumptions on G (satisfied for extremes)

- Are we trapped ? Parameters are the same whatever the forecast
- **Crucial (and unrealistic) assumption here** : F and G are independent
- In practice, the convergence to these parameters is driven by the skill of ensembles for extreme events

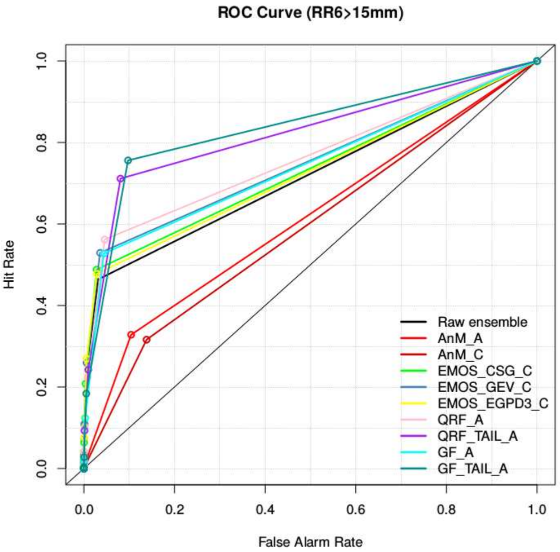
And so what ?

- **Crucial (and unrealistic) assumption here** : F and G are independent
- In practice, the convergence to these parameters is driven by the skill of ensembles for extreme events

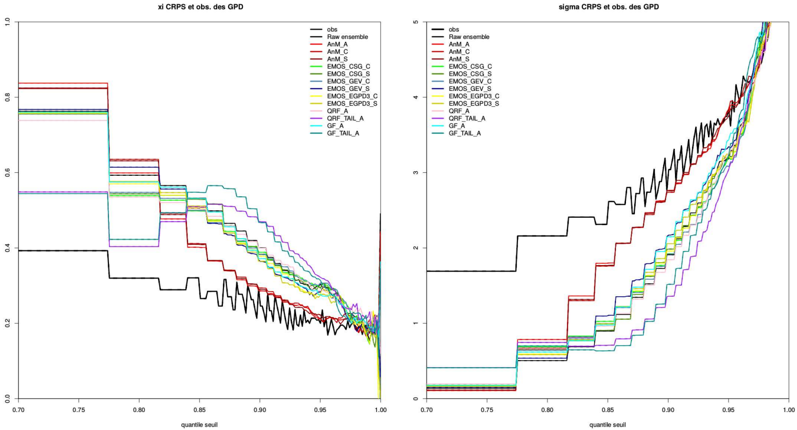


Plan

Post-processing of 6-h rainfall for extremes



Post-processing of 6-h rainfall for extremes



Estimations of GPD parameters are highly correlated

Conclusions

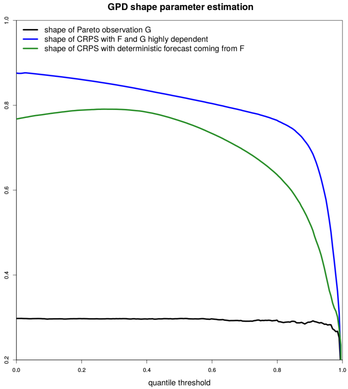
- Some properties of (w)CRPS are debated... And also used
- A different criterion for extreme verification is established
  Be skillful for extremes subject to a good overall performance
- A new way to verify ensemble (only ?) forecasts for extremes is shown
- This tool can be viewed as a summary of ROCs among thresholds.
- It seems to be consistent with simulations and real data

Conclusions

- Some properties of (w)CRPS are debated... And also used
- A different criterion for extreme verification is established
  Be skillful for extremes subject to a good overall performance
- A new way to verify ensemble (only ?) forecasts for extremes is shown
- This tool can be viewed as a summary of ROCs among thresholds.
- It seems to be consistent with simulations and real data

Conclusions

- Some properties of (w)CRPS are debated... And also used
- A different criterion for extreme verification is established
  Be skillful for extremes subject to a good overall performance
- A new way to verify ensemble (only ?) forecasts for extremes is shown



Conclusions

- Some properties of (w)CRPS are debated... And also used
- A different criterion for extreme verification is established
  Be skillful for extremes subject to a good overall performance
- A new way to verify ensemble (only ?) forecasts for extremes is shown

- This tool can be viewed as a summary of ROCs among thresholds.
- It seems to be consistent with simulations and real data

Conclusions

- Some properties of (w)CRPS are debated... And also used
- A different criterion for extreme verification is established
  Be skillful for extremes subject to a good overall performance
- A new way to verify ensemble (only ?) forecasts for extremes is shown

- This tool can be viewed as a summary of ROCs among thresholds.
- It seems to be consistent with simulations and real data

UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES
université PARIS-SACLAY

Lyon 1

cnrs

METEO FRANCE

Références I

Jochen Bröcker. Resolution and discrimination–two sides of the same coin. *Quarterly Journal of the Royal Meteorological Society*, 141 (689) :1277–1282, 2015.

Jochen Bröcker and Leonard A Smith. Scoring probabilistic forecasts : The importance of being proper. *Weather and Forecasting*, 22(2) :382–388, 2007.

Laurens De Haan and Ana Ferreira. *Extreme value theory : an introduction*. Springer Science & Business Media, 2007.

Petra Friederichs and Thordis L Thorarinsdottir. Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics*, 23(7) :579–594, 2012.

Références II

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69 (2) :243–268, 2007.

Sebastian Lerch, Thordis L Thorarinsdottir, Francesco Ravazzolo, Tilmann Gneiting, et al. Forecaster's dilemma : extreme events and forecast evaluation. *Statistical Science*, 32(1) :106–127, 2017.

David S Richardson. Skill and relative economic value of the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126(563) :649–667, 2000.