



HAL
open science

EDITORIAL: SPECIAL ISSUE ON THE EXTREME VALUE ANALYSIS CONFERENCE CHALLENGE "PREDICTION OF EXTREMAL PRECIPITATION"

Olivier Wintenberger

► **To cite this version:**

Olivier Wintenberger. EDITORIAL: SPECIAL ISSUE ON THE EXTREME VALUE ANALYSIS CONFERENCE CHALLENGE "PREDICTION OF EXTREMAL PRECIPITATION". *Extremes*, 2018, 21, 10.1007/s10687-018-0327-7 . hal-01962884

HAL Id: hal-01962884

<https://hal.science/hal-01962884>

Submitted on 10 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**EDITORIAL: SPECIAL ISSUE ON THE EXTREME VALUE
ANALYSIS CONFERENCE CHALLENGE "PREDICTION OF
EXTREMAL PRECIPITATION"**

OLIVIER WINTENBERGER

At the Extreme Value Analysis conference in Delft in June 2017 a challenge for predicting spatio-temporal extremes was proposed. The aim of the challenge was to estimate high quantiles of daily rainfall and to extrapolate them in time and space. Eight teams competed in the challenge. A data set from the training period was given to each team. Based on the data from the training period each team predicted the corresponding high quantiles for the adjacent test period. The goal was to score those teams that achieved the best predictive power.



FIGURE 1. Training and test samples from different periods of observation.

1. THE DATA

Daily (24 hour) accumulations of precipitation $P_{j,t}$, $j = 1, \dots, 40$ (unit inches) have been recorded at 40 stations in the Netherlands during the 44 year period from $t=12/31/1972$ to $t=12/31/2016$.

The training sample corresponds to the 24 year period from $t=12/31/1972$ to $t=12/31/1995$; see Figure 1. The aim was to predict, from the training sample, a quantile of level corresponding to the extreme monthly precipitation over the next 20 years (the test period from $t=01/01/1996$ to $t=12/31/2016$) station by station. On the daily level, this event corresponds to the 0.998-quantile, i.e.,

$$0.998 = 1 - 0.002 \approx 1 - \frac{1}{20 * 30 \text{ days}}.$$

Financial support by the ANR network AMERISKA ANR 14 CE20 0006 01 is gratefully acknowledged by Olivier Wintenberger.

Under the assumption of strict stationarity, the quantile $q_{j,k}$ satisfies

$$\mathbb{P}(P_{j,t} > q_{j,k}) = 0.002,$$

for any day t of the test period in month $k \in \{1, \dots, 12\}$ and any station $j \in \{1, \dots, 40\}$.

The data set has some missing values (less than 2% of the active days) and plenty of days without rain (52%). The distributions of precipitation at the distinct stations on rainy days are quite similar. This may be due to the low altitude of the Netherlands.

The training and test periods have similar properties:

- They both contain some days with extreme values.
- Their tail indices take values around 2 as indicated by their Hill estimates; see Figure 2.

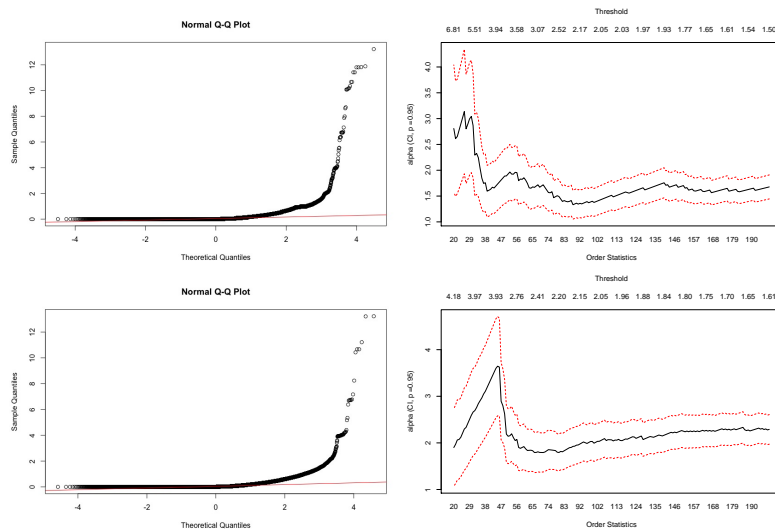


FIGURE 2. QQ-plots against the standard normal distribution (left) and Hill plots of the tail index kappa in the model $\mathbb{P}(X > x) = Cx^{-\kappa}$ (right) for the daily non-null precipitation in the training (top) and test (bottom) periods.

As regards the monthly 99.8%-quantiles, one observes similar seasonal behavior both in the training and test periods. Summer months appear to have heavier rainfall than other months. This may be due to convective storms which are more present in summer; see Figure 3.

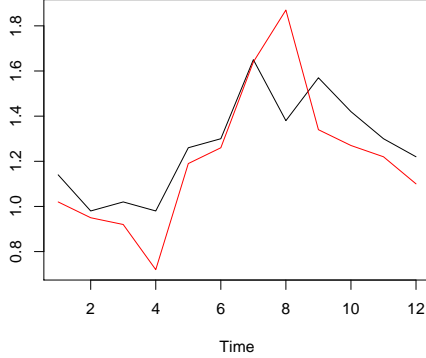


FIGURE 3. Monthly empirical 99.8% quantiles during the training (dark) and the test (red) periods

2. EVALUATION

The performance of the quantile predictions $\hat{q}_{j,k}$ for $j = 1, \dots, 40$ and $k = 1, \dots, 12$, is evaluated with the quantile loss function

$$\ell(x, y) = \alpha(x - y)1_{x > y} + (1 - \alpha)(y - x)1_{y \geq x}, \quad x, y \in \mathbb{R},$$

at the level $\alpha = 0.998$. The quantile predictions $\hat{q}_{j,k}$ are compared with the daily maxima $P_{j,t}$ at each station $j = 1, \dots, 40$, and month $k = 1, \dots, 12$:

$$S_{j,k}(\hat{q}_{j,k}) = \sum_{\text{days } t \text{ of the test period in month } k} \ell(P_{j,t}, \hat{q}_{j,k}).$$

Notice that, under strict stationarity, the risk $q \rightarrow \mathbb{E}[S_{j,k}(q)]$ is minimized in $q_{j,k}$; see [2]. The final scores of the predictive algorithm $\hat{q} = (\hat{q}_{j,k})$ are the sums of the quantile losses over the stations and months

$$S_i(\hat{q}) = \sum_{j \in C_i} \sum_{k=1}^{12} S_{j,k}(\hat{q}_{j,k}), \quad i = 1, 2.$$

The teams were informed that this score function will be used for the final ranking.

There are two scores based on two different sets C_i , $i = 1, 2$, of test stations:

- (1) The final score is the sum over C_1 , the set of the 29 stations of the training sample that were still active after the training period,



FIGURE 4. Locations of the 40 stations used in the challenge. The red circles show the 35 stations active during the training period, the dark crosses represent the 34 stations active during the test period.

- (2) The final score is the sum over C_2 , the set of the 34 stations active during the test period; see Figure 4.

The scores are compared with those of the benchmark, predicting the quantiles by the monthly maxima per station for the first score, extrapolating with the average of those monthly maxima for the second score.

The final score board ranks the relative difference of the scores S_i , $i = 1, 2$, of each team to those of the benchmark: the score of the benchmark based on the dataset C_i is denoted $S_{i,b}$ for $i = 1, 2$. Note that the ranking is the same for the two scores. Each team has been invited to present their algorithm in this special issue of the journal *Extremes*.

team's and members names	$\frac{S_1}{S_{1,b}} - 1$	$\frac{S_2}{S_{2,b}} - 1$
Melbs Saunders, Stephenson, Tafakori	60.6%	58.5%
STOR-i Lancaster Simpson Barlow Sharkey Shooter Rohrbeck	59.9%	57.7%
INLA goes extreme Huser Opitz Rue	59.7%	57.5%
SNUMultiscale Park Kwon Kim Oh	57.4%	55.5%
SUTD Bernhard Lai Nguyen	57.2%	55.1%
UNC Glotzer He Pipiras	56.2%	54.0%
FraC Mefeh Ho Biard Dombry	43.6%	42.0%
LUCY Liu Uchiumi Cheng Yumoto	40.7%	38.7%

3. DISCUSSION

The scores are impressive: many teams have improved the score of the benchmark by more than 50%. In an uncertain environment and for a difficult problem as predicting a high-level quantile, it is encouraging to see that modern algorithms can significantly improve the naive approach of the benchmark. A challenge is also a good opportunity to compare the predictive power of new algorithms.

One may discuss the choice of the score. The expectation of the quantile loss is minimal for a high-level quantile. A proper score, so that its expectation is minimal for a distribution that has the correct tail property, seems much more accurate; see e.g. [1]. Prediction challenges should use such scores.

It is also unclear whether, in such prediction challenge, the conditions are met so that a marginal quantity (quantiles or tail distribution) is actually inferred. One should take into account that the scores, as average losses, approximate the expectation *conditionally* on the information from the training sample. If so, a conditional quantity (quantiles or tail distribution) is inferred. The prediction problem is more complicated than the marginal one; for instance, it prevents the use of many classical and powerful cross-validation techniques.

4. CONCLUSION

We would like to thank all participants to rise to this challenge. The solutions proposed were imaginative and spanned a wide range of techniques from classical EVT to machine learning approaches. The teams were given the opportunity to submit their work in this journal and we hope that the attentive reader will enjoy, not only reading each paper, but also comparing them. Finally, this particular challenge clearly shows that quite different paths can lead to similar skill scores in predicting extreme rainfall quantiles. Such a conclusion reinforces the identified need of welcoming new ideas and communities at conferences like Extreme Value Analysis.

REFERENCES

- [1] GNEITING, T., AND RAFTERY, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *JASA* 102(477), 359-378.
- [2] KOENKER, R. (2005) *Quantile regression*. Cambridge University Press, Cambridge.

SORBONNE UNIVERSITY, UPMC UNIVERSITÉ PARIS 06, F-75005, PARIS, FRANCE
E-mail address: `olivier.wintenberger@upmc.fr`