Supporting Information for "A Quantile Generalised Additive Approach for Compound Climate Extremes: Pan-Atlantic Extremes as a Case Study"

Leonardo Olivetti[1,2], Gabriele Messori[1,2,3,4], Shaobo Jin[5,6]

[1]Department of Earth Sciences, Uppsala University

[2]Centre of Natural Hazards and Disaster Science, Uppsala University

[3]Department of Meteorology, Stockholm University

[4]Bolin Centre for Climate Research, Stockholm University

[5]Department of Statistics, Uppsala University

[6]Department of Mathematics, Uppsala University

Contents of this file

1. Text S1 to S2
2. Figures S1 to S8
3. Table S1

Introduction

In this supporting information, we provide a brief introduction to classic parametric extreme value theory (text S1), some details on the formulation of our models (text

Corresponding author: L.Olivetti, leonardo.olivetti@geo.uu.se

S2), and additional figures and tables where the analysis in the main paper is repeated accounting for the autocorrelation in the residuals, assuming an AR 1 process.

Text S1. Parametric EVT

Block Maxima approach

BM follow the GEV family of distributions, which has the following density function (Coles, 2001):

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1} e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}},$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter.

The BM approach aims to reliably estimate the shape parameter ξ. This is because ξ can be used to infer how often a certain value of the distribution is likely to occur (return level), and obtain information on the thickness of the tail of the distribution. Based on the value of ξ, the GEV family may be subdivided into three subfamilies of distributions:

- $\xi < 0$ corresponds to the Weibull family. Distributions within this family have a well-defined upper bound and light tails.
- $\xi = 0$ corresponds to the Gumbel family. Distributions within this family have no theoretical upper bound, but a fast decaying exponential tail, which makes extreme events above a certain threshold very unlikely to occur.
- $\xi > 0$ corresponds to the Fréchet family. Distributions within this family have heavy, slowly decaying tails and no upper bound, which makes it impossible to draw conclusions on the maximum magnitude of the extremes.

Peak-Over-Threshold approach

The excesses, defined as the difference between the selected maxima M_i and the threshold u, given independence, belong asymptotically to the generalised Pareto distribution (GPD), in accordance with Pickands' theorem (Pickands, 1975).

The probability density function of the GPD is defined as follows (Coles, 2001):

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1},$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. The shape parameter can be interpreted in relation to the tails of the distribution similar to the BM approach, where a larger value of ξ is associated with heavier tails. Given the same underlying distribution, the shape parameter estimated through the BM and the POT approach is approximately the same (Coles, 2001).

S2. Model Formulation

Below, we describe the statistical models used in the main text and Figures in the supporting materials. The same models are used for the prediction of daily mean 10m wind speed and daily precipitation.

QGAM models presented in the Main Text

Base model:

$$\begin{aligned} Q_{Y|X}(\tau) = {} & s(lat, lon, k=20, d=2) + s(time, k=5) \\ & + seaonsal\_quantile\_climatology + month, \end{aligned} \tag{1}$$

where k is the basis dimension, s is a smooth, and d the dimension of the product smooth. Month is treated as a categorical variable.

Cold spell model:

$$
\begin{aligned}
Q_{Y|X}(\tau) &= s(lat, lon, k = 20, d = 2) + s(time, k = 5) + month \\
&+seaonsal\_quantile\_climatology + s(US\_temp\_lag2, k = 5)
\end{aligned} \tag{2}
$$

cold spell and jet stream model:

$$
\begin{aligned}
Q_{Y|X}(\tau) &= s(lat, lon, k = 30, d = 2) + s(time, k = 5) + month \\
&+seaonsal\_quantile\_climatology + s(US\_temp\_lag2, k = 5) \\
&+s(Jet\_strength\_lag1, k = 10) + s(NAO\_lag1, k = 10) \\
&+s(lat\_jet\_lag1, k = 10) + s(jet\_proximity, k = 10),
\end{aligned} \tag{3}
$$

where jet proximity is given by the distance in degrees between the latitude of the maximum zonally averaged jet and the latitude of the grid point of interest.

QREG models presented in the Main Text

Base model:

$$
Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal\_quantile\_climatology + year + month \tag{4}
$$

Cold spell model:

$$Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal\_quantile\_climatology$$
$$+ year + month + US\_temp\_lag2 \tag{5}$$

Cold spell and jet stream model:

$$Q_{Y|X}(\tau) = Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal\_quantile\_climatology$$
$$+ year + month + US\_temp\_lag2 + Jet\_strength\_lag1 + NAO\_lag1$$
$$+ lat\_jet\_lag1 + jet\_proximity \tag{6}$$

POT models presented in the Main Text

Since information on previous lags of the outcome in Europe apart from seasonal quantile climatology is not available to the model, declustering is based on the lag -1 of the jet speed, which is the variable most closely correlated to the extremes in 10m wind speed and precipitation.

A threshold is set at the 90th quantile of the outcome of interest, and values above that are considered extreme. Then declustering is performed, by requiring that at least five days elapse between days with lag -1 of the jet speed over the 90th quantile of the seasonal climatology. Whenever several days within 5 days of each other meet the criteria for being classified as extreme, only the first day to meet the criteria is selected.

Otherwise, the same regressors as QREG models are used, but predicting the location parameter of the GPD. The shape parameter is held fixed. The estimated parameters are then used to generate the predictions, assuming that extremes follow a GPD.

QGAMs presented in the Supporting Information

As the QGAMs in the main paper, but adding $s(y_{ij}\_lag\_1, k = 5)$ as a regressor to all the models, where y_{ij} is the outcome variable of interest at the grid point of interest, i.e. lag -1 of 10m wind speed for the wind speed models, and lag -1 of precipitation for the precipitation models.

QREG models presented in the Supporting Information

As the QREG models in the main paper, but adding $y_{ij}\_lag\_1$ as a regressor to all the models.

POT models presented in the Supporting Information

Declustering is based on lag -1 of the outcome of interested instead of on the speed of the jet, following the same declustering technique described for the models in the main paper.

Otherwise, as the POT models in the main paper, but adding $y_{ij}\_lag\_1$ as a regressor to all the models.

Results of the autoregressive models

The figures below, differently from those in the main paper, make use of the last lag of the outcome variable of interest among the regressors (AR 1 models) in addition to all other information, and in the case of the POT models also use information of previous lags for declustering purposes (see above for details).

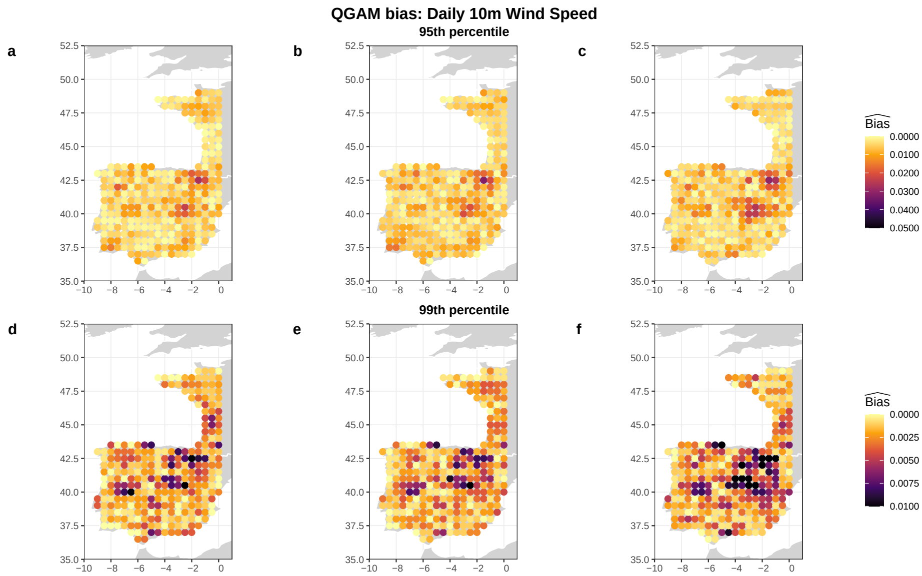

Figure S1. Estimated bias of QGAMs in terms of absolute distance between the percentage of overpredictions ($\hat{PO}$) and the theoretical quantile (τ). **a-c**: Estimation of 95th quantile of daily mean 10m wind speed. Basic model with no information on the state of the North Atlantic atmosphere (**a**), model with information on surface temperature in North America at lag -2 days (**b**), model with same information as above plus speed and location of the Polar jet stream at lag -1 days (**c**). **d-f**: As a-c, but for the 99th quantile of daily mean 10m wind speed.

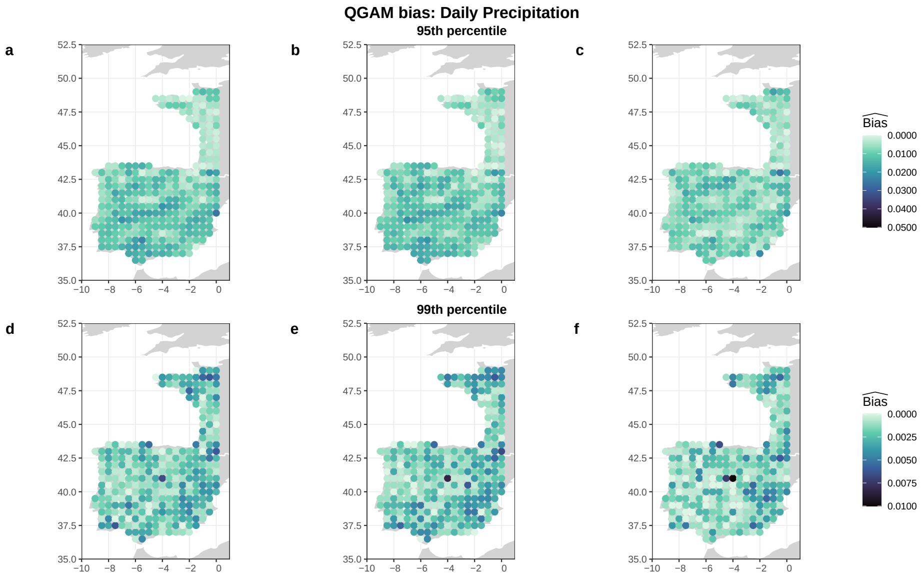

Figure S2. As Fig. S1 but for daily precipitation.

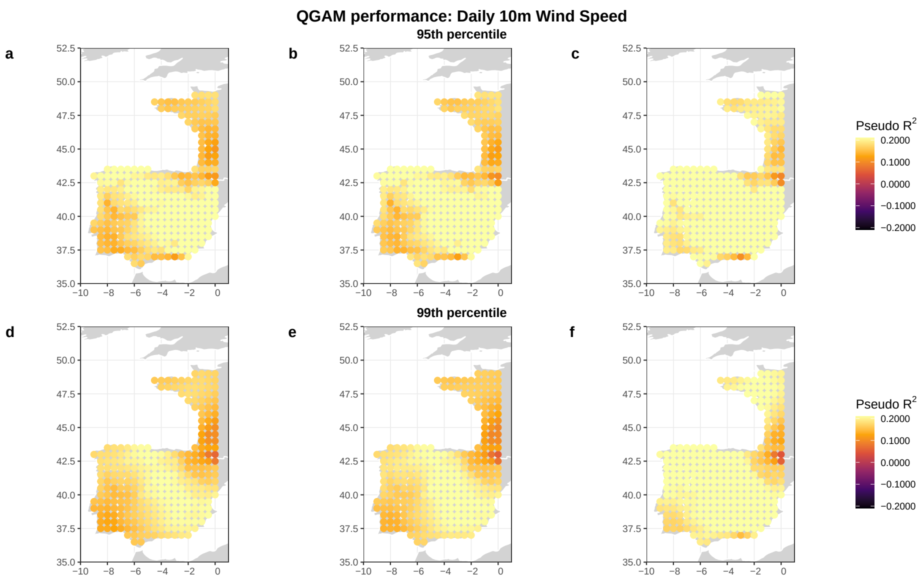

Figure S3. As Figure S1 but for pseudo R^2

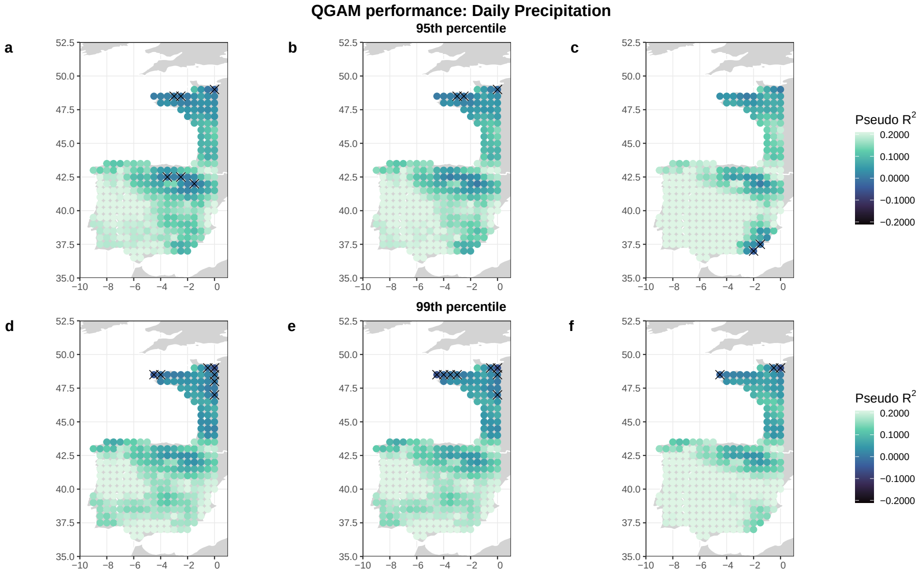

Figure S4. As Figure S2 but for pseudo R^2

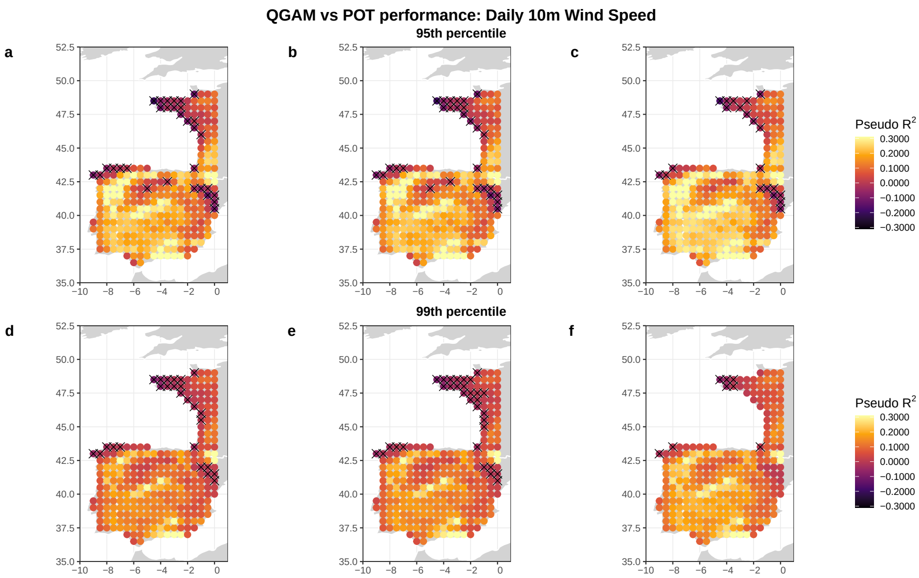

Figure S5. As Figure S3, but using the linear POT model as baseline for the computation of the pseudo R^2

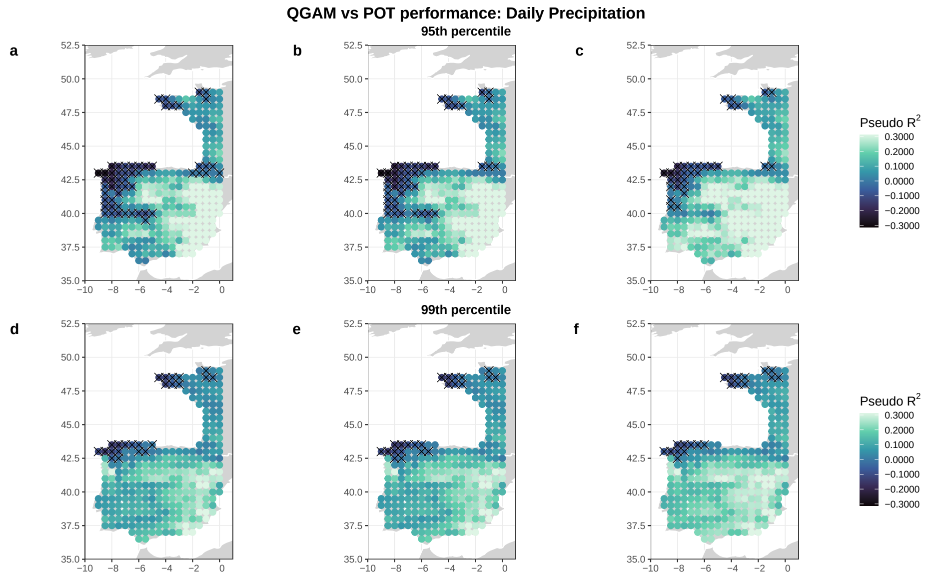

Figure S6. As Figure S4, but using the linear POT model as baseline for the computation of the pseudo R^2

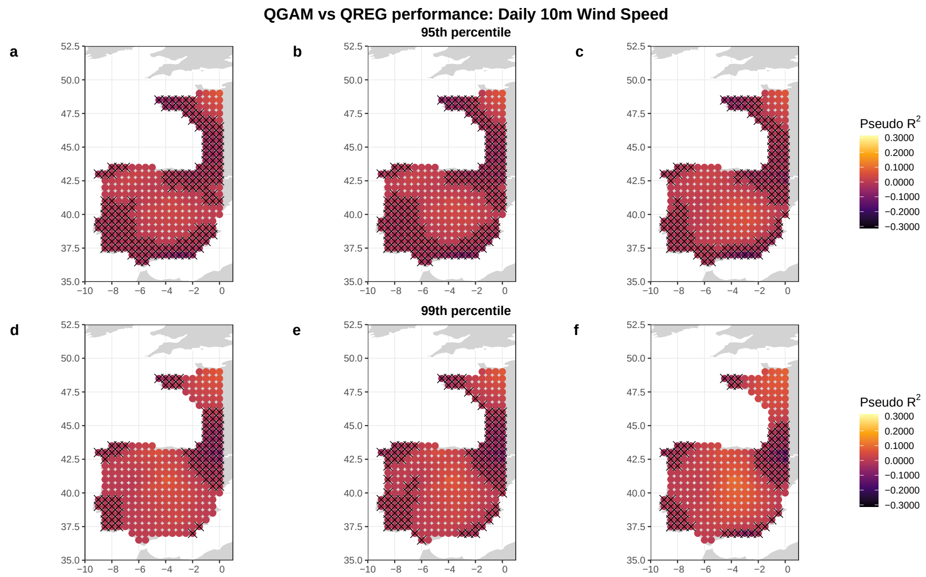

Figure S7. As Figure S3, but using the quantile regression model as baseline for the computation of the pseudo R^2

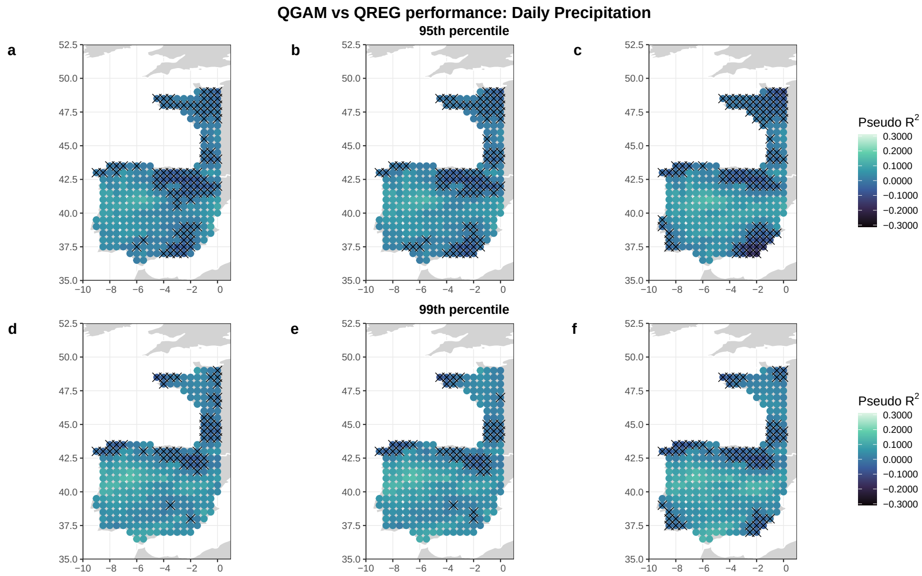

Figure S8. As Figure S4, but using the quantile regression as baseline for the computation of the pseudo R^2

Table S1. Overall pseudo R^2 of cold spell and jet stream QGAMs used to forecast the 95th and 99th quantiles of daily mean 10m wind speed and daily precipitation compared to different baseline models.

Variable	Baseline model	95th quantile	99th quantile
Daily mean 10m wind speed	Quantile of the climatology	0.2210	0.2118
Daily mean 10m wind speed	POT	0.1733	0.1302
Daily mean 10m wind speed	QREG	0.0047	0.0209
Daily precipitation	Quantile of the seasonal climatology	0.2171	0.2005
Daily precipitation	POT	0.1850	0.1574
Daily precipitation	QREG	0.0295	0.0443