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

Leonardo Olivetti¹, Gabriele Messori¹, and Shaobo Jin¹

¹Uppsala University

April 16, 2023

Abstract

We present an application of quantile generalised additive models (QGAMs) to study spatially compounding climate extremes, namely extremes that occur (near-) simultaneously in geographically remote regions. We take as an example wintertime cold spells in North America and co-occurring wet or windy extremes in Western Europe, which we collectively term Pan-Atlantic compound extremes. QGAMS are largely novel in climate science applications and present a number of key advantages over conventional statistical models of weather extremes. Specifically, they remove the need for a direct identification and parametrisation of the extremes themselves, since they model all quantiles of the distributions of interest. They thus make use of all information available, and not only of a small number of extreme values. Moreover, they do not require any a priori knowledge of the functional relationship between the predictors and the dependent variable. Here, we use QGAMs to both characterise the co-occurrence statistics and investigate the role of possible dynamical drivers of the Pan-Atlantic compound extremes. We find that cold spells in North America are a useful predictor of subsequent wet or windy extremes in Western Europe, and that QGAMs can predict those extremes more accurately than conventional peak-over-threshold models.

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}

5	¹ Department of Earth Sciences, Uppsala University
6	² Centre of Natural Hazards and Disaster Science, Uppsala University
7	³ Department of Meteorology, Stockholm University
8	⁴ Bolin Centre for Climate Research, Stockholm University
9	⁵ Department of Statistics, Uppsala University
10	⁶ Department of Mathematics, Uppsala University

Key Points:

1

2

3

4

11

12	•	Quantile general additive models (QGAMs) can model the relationship between
13		compound climate extremes flexibly and robustly.
14	•	North American cold spells are a useful predictor of subsequent wet or windy ex-
15		tremes in Western Europe.
16	•	QGAMs can predict those extremes more accurately than conventional peak-over-
17		threshold models.

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

18 Abstract

We present an application of quantile generalised additive models (QGAMs) to study 19 spatially compounding climate extremes, namely extremes that occur (near-) simulta-20 neously in geographically remote regions. We take as an example wintertime cold spells 21 in North America and co-occurring wet or windy extremes in Western Europe, which we 22 collectively term Pan-Atlantic compound extremes. QGAMS are largely novel in climate 23 science applications and present a number of key advantages over conventional statis-24 tical models of weather extremes. Specifically, they remove the need for a direct iden-25 tification and parametrisation of the extremes themselves, since they model all quantiles 26 of the distributions of interest. They thus make use of all information available, and not 27 only of a small number of extreme values. Moreover, they do not require any a priori knowl-28 edge of the functional relationship between the predictors and the dependent variable. 29 Here, we use QGAMs to both characterise the co-occurrence statistics and investigate 30 the role of possible dynamical drivers of the Pan-Atlantic compound extremes. We find 31 that cold spells in North America are a useful predictor of subsequent wet or windy ex-32 tremes in Western Europe, and that QGAMs can predict those extremes more accurately 33 than conventional peak-over-threshold models. 34

35 Plain Language Summary

In this paper we propose a new data-driven method to study climate extremes oc-36 37 curring simultaneously in multiple, possibly remote, locations. Such extremes can pose a greater threat to human societies than single, isolated extremes, as their effects may 38 exacerbate each other and lead to correlated losses. The method we suggest requires fewer 30 assumptions than conventional extreme value statistical techniques, and can help us to 40 identify previously unknown relationships between the extremes themselves and their 41 possible drivers. We exemplify its use by studying the co-occurrence of periods of un-42 usually cold weather in North America and uncommonly strong wind and abundant pre-43 cipitation in Western Europe. We find that the new method has better predictive power 44 for the European extremes than conventional statistical approaches. Furthermore, we 45 confirm the results of previous studies suggesting an association between the wintertime 46 extremes in North America and Western Europe. 47

48 1 Introduction

The statistical properties of climate extremes have been extensively studied using 49 parametric approaches to extreme value theory (EVT; e.g. Fisher & Tippett, 1928; Gum-50 bel, 1941; Davison & Smith, 1990; Coles, 2001; Mares et al., 2009; Elvidge & Angling, 51 2018; French et al., 2019), and since its inception, parametric EVT regression has been 52 successfully used to investigate extreme event drivers (Gumbel, 1958; Pickands, 1975; 53 Coles, 2001; Mares et al., 2009; do Nascimento et al., 2021). Parametric EVT aims to 54 identify and characterise extreme observations by identifying their underlying distribu-55 tion which, in turn, provides valuable information on the expected frequency and inten-56 sity of the extremes. Two fundamental approaches lie at the core of classic parametric 57 EVT: the block maxima approach (BM) (Fisher & Tippett, 1928; Gumbel, 1958) and 58 the peak over threshold (POT) approach (Pickands, 1975; Smith, 1984; Davison & Smith, 59 1990). 60

The BM approach defines k time periods (blocks) of equal length and extracts the n largest independent observations from each block (block maxima). According to the Fisher–Tippett–Gnedenko theorem (Fisher & Tippett, 1928), properly normalised block maxima converge to a distribution belonging to the generalised extreme value family of distributions (GEV). A challenge of the BM approach lies in finding a suitable size for the blocks (Ferreira & Haan, 2015). Small blocks may lead to the identification of some spurious extremes, whereas large blocks may ignore some extremes and slow down the convergence to the GEV family of distributions. Furthermore, given that the Fisher-Tippett-Gnedenko
 theorem is an asymptotic theorem, there is no guarantee that the appropriately normalised
 block maxima belong to the GEV family of distributions when the sample size is lim-

71 ited.

The POT approach addresses these challenges by selecting a high threshold u, and defining as extremes observations above that threshold (the maxima, μ). The excesses follow approximately a generalised Pareto distribution (GPD), in accordance with Pickands' theorem (Pickands, 1975). The challenge in this case is to define an appropriate threshold: sufficiently high to isolate extremes yet sufficiently low to ensure an appropriate sample size. Technical details of the BM and POT approaches are provided in the Supporting Information text S1.

While being a widely-used and versatile tool, implementing parametric EVT for 79 studying climate extremes presents some challenges. The first is finding a suitable em-80 pirical definition for the extremes, as discussed above and also highlighted by Passow and 81 Donner (2019). Second, using parametric EVT in a regression context to study extreme 82 event drivers requires some a priori knowledge or assumption of the functional relation-83 ship between the drivers and the extremes. Defining this relationship a priori is often 84 challenging and may result in simplistic or unrealistic assumptions about the relation-85 ship between the two. Third, parametric EVT does not make optimal use of all the sta-86 tistical information available, as it ignores all that is not an extreme (see also the dis-87 cussion in Passow & Donner, 2019). 88

The literature offers extensions to the classic parametric EVT theory (Lucarini et 89 al., 2016) as well as a number of non-parametric alternatives (Koenker & Hallock, 2001; 90 Yee, 2015; Fasiolo et al., 2021). The main appeal of non-parametric techniques, such as 91 quantile-based methods, is that they do not require an empirical definition of the extremes 92 in the same way as parametric EVT does. Furthermore, they make use of all statisti-93 cal information available, rather than just the extreme data. This is particularly relevant when the interest lies in extreme events with a relatively short return period (e.g. 95 windstorms recurring every 6 months or 1 year at a given location), for which the con-96 vergence of the GPD parameters may first be observed at quantiles larger than the one 97 of interest. 98

Here, we apply quantile generalised additive models (QGAMs, Fasiolo et al., 2021; 99 Koenker, 2011) to the study of spatially compounding climate extremes and their drivers, 100 namely extremes that occur (near-)simultaneously in geographically remote regions. The 101 synchronised - or compound - occurrence of remote extremes is often associated with 102 greater impacts than those of the corresponding individual extremes, for example expos-103 ing actors with international coverage to correlated losses (Mills, 2005) and imperilling 104 global food security (Kornhuber et al., 2020). QGAMs are a non-parametric approach 105 which is largely novel in the context of climate science. They are a recent extension to 106 generalised additive models, bearing the promise of addressing the aforementioned three 107 key limitations of parametric EVT. We benchmark QGAMs against other parametric 108 and non-parametric approaches. 109

To illustrate the methodology, we consider the repeated occurrence of wintertime 110 cold spells in Eastern North America and wet or windy extremes in Western Europe, which 111 we collectively term Pan-Atlantic compound extremes. The repeated occurrence of these 112 extremes in recent winters (e.g. Coumou & Rahmstorf, 2012; Lee et al., 2015; Trenary 113 et al., 2015; Dodet et al., 2019; Wild et al., 2015) has led to hypothesise a connection 114 between the two sets of extremes (Messori et al., 2016; De Luca et al., 2020; Leeding et 115 al., 2023; Messori & Faranda, 2023). Nonetheless, a systematic statistical characterisa-116 tion of such connection has largely been limited to simple co-occurrence statistics. The 117 aim here is to provide a proof-of-concept for the use of QGAMs in the study of compound 118 climate extremes. We therefore seek to evaluate the performance of QGAMs relative to 119 alternative statistical models by applying them to a previously studied set of spatially 120 compounding climate extremes, as opposed to investigating novel extreme occurrences 121 and related large-scale atmospheric drivers. We further highlight that we do not aim to 122

use the regression models we present as self-standing forecasting tools. In other words,
we do not aim to forecast extreme event occurrences using only information from several days before the extremes – as one would do in a conventional forecasting exercise.
Rather, we see these models as useful tools to robustly quantify the statistical connections between geographically remote extremes and investigate the roles of different potential dynamical drivers.

The remainder of this paper is structured as follows: Section 2 provides a short in-129 troduction to parametric EVT regression, and introduces quantile-based non-parametric 130 approaches as an alternative, including QGAMs. Section 3 defines the scope of this pa-131 per, and discusses practical concerns related to variable selection and model formulation. 132 Section 4 compares the methods presented in previous sections, by studying the Pan-133 Atlantic compound extremes and their possible dynamical drivers, considering specif-134 ically the North Atlantic jet stream and the North Atlantic Oscillation (NAO). Section 135 5 concludes the paper by providing a short summary of the findings and discussing the 136 strengths and limitations of QGAM applications to study climate extremes. 137

¹³⁸ 2 Extreme Value Statistical Models

This section presents some parametric and non-parametric approaches to EVT regression and discusses how their performance can be compared in practice.

¹⁴¹ 2.1 Parametric EVT regression

The relationship between the extremes and their likely precursors can be described parametrically through a generalised linear model, where the expected value of a BM or POT extreme at a time t, $E(M_t)$ is a function of previous values of itself, M_{t-k} and other factors likely to affect its strength, X_{t-k} . Then:

¹⁴⁶
$$E(M_t|M_{t-k}, X_{t-k}) = M_{t-k}\phi + X_{t-k}\beta.$$
 (1)

Since the extremes are selected through the BM or the POT approach, paramet ric EVT regression largely shares the same strengths and limitations of these approaches.

149

157

2.2 Non-parametric EVT regression

¹⁵⁰ Non-parametric models are a broad class of methods which do not rely on a pre-¹⁵¹ determined functional relationship between the outcome and the predictor, but estimate ¹⁵² it empirically (Härdle, 1990). Quantile-based models are a subset of non-parametric mod-¹⁵³ els which are particularly suitable to the analysis of extremes, as they can be used to es-¹⁵⁴ timate values in the tails of the distribution. A quantile $Q(\tau)$ is defined as the inverse ¹⁵⁵ of the cumulative distribution function, uniquely identifying the value of the cumulative ¹⁵⁶ distribution function corresponding to probability τ .

2.2.1 Quantile Regression

The linear relationship between a conditional quantile of an output and a predictor can be estimated non-parametrically through quantile regression (Koenker & Hallock, 2001; Koenker & Bassett, 1978). Quantile regression aims to estimate the conditional quantile of the dependent variable $Q_{Y|X}(\tau)$ as a function of the regressors X, so that:

$$Q_{Y|X}(\tau) = X\beta_{\tau} \tag{2}$$

$$\min E\{\rho_{\tau}(Y - X^*\beta)\},\tag{3}$$

where $\rho_{\tau} = (\tau - 1)zI(z < 0) + \tau zI(z \ge 0)$ is the so-called "pinball loss", punishing predictions which are further away from the quantile of interest. Here, z is the residual and τ the quantile of interest.

By estimating the effect of the regressors on an ensemble of extreme quantiles, it is possible to draw some conclusions on how the regressors affect the intensity of the extremes. This is done without relying on the large sample asymptotics employed by parametric EVT, thus without the need of reducing sample size.

The main limitation of quantile regression is that it is a linear model, ignoring all possible non-linear effects of the regressors on the quantile of interest.

2.2.2 Quantile Generalised Additive Models

Generalised additive models (GAMs, Wood, 2017; Hastie & Tibshirani, 1986) are a broad family of non-parametric models describing the dependent variable as an additive function of unknown smooths of the regressors. A smooth function is a function which is derivable up to certain order at each point throughout its domain. GAMs can be described as follows:

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_i(X_i),$$
(4)

where E(Y) is the expected value of the outcome, g() is the link function, β_0 is an intercept, and $f_1(X_1) + f_2(X_2) + ... + f_i(X_i)$ are smooth functions of the predictors.

GAMs do not require determining the functional relationship between the outcome 184 and the predictors a priori. This is instead determined empirically, through a data-driven 185 process testing a large number of possible combinations. First, a set of bases is chosen 186 for the predictors, so that the original covariates $X_1, X_2...X_i$ are embedded into a larger 187 feature space X^* including higher order terms of the original covariates. Then, the best 188 model is chosen out of the expanded feature space by minimising a loss function of choice, 189 $L(X^*)$, while penalising for excessive complexity. In the absence of a link function, a nat-190 ural choice of loss function is the quadratic loss, so the best model is chosen according 191 to the following minimisation: 192

165

175

181

$$\min E\{(Y - X^*\beta)^2 + \lambda J\},\tag{5}$$

where $(Y - X^*\beta)^2$ is the sum of squared residuals, λ is a smoothing parameter and 194 J is a penalty term. A common penalty term is $J = \int f''(x)^2 dx$, with other forms of penalty 195 also being possible (James et al., 2022). J is larger when the function becomes wigglier, 196 punishing excessive functional complexity. Choosing larger values of the smoothing pa-197 rameter λ pushes the model towards a simpler functional form, so that $\lambda \to \infty$ makes 198 GAM equivalent to linear regression. The value of λ is usually determined empirically, 199 through generalised cross-validation or restricted maximum likelihood. For technical de-200 tails, the reader is referred to Wood (2017). 201

Classic GAMs as expressed in equation 4 model the expected value of the outcome 202 and not of the maxima, making them unsuitable for extreme value analysis. One approach 203 could be to model the expected value of a series of independent and identically distributed 204 maxima $(M_1, M_2, ..., M_N)$, by selecting the maxima through a BM or a POT approach; 205 however, this reintroduces into the model the limitations related to those approaches. 206 An alternative approach is to model a conditional quantile of the outcome as a function 207 of the predictors, as in quantile regression: this approach goes under the name of quan-208 tile generalised additive model (QGAM). 209

QGAMs are a recent extension of GAMs and quantile regression, which model a chosen conditional quantile of interest as a sum of unknown smooth functions of the regressors (Fasiolo et al., 2021). This builds on earlier theoretical work from Koenker (2011). QGAMs can be expressed as follows:

$$Q_{Y|X}(\tau) = f_1(X_1) + f_2(X_2) + \dots + f_i(X_i), \tag{6}$$

where $Q_{Y|X}(\tau)$ is a conditional quantile of choice of the dependent variable. They aim to minimise a loss function similar to the one described by equation 3,

$$\min E\{\rho_{\tau}^*(Y - X^*\beta)\},\tag{7}$$

where ρ_{τ}^* is defined as:

214

217

218

219

234

$$\rho_{\tau}^* = (\tau - 1)\frac{z}{\sigma}I(z < 0) + \lambda log(1 + e^{\frac{z}{\lambda\sigma}})$$
(8)

 $\begin{array}{ll} & \rho_{\tau}^{*} \text{ is the extended log-f loss, which, similarly to the pinball loss, punishes predic$ $tions which are further away from the quantile of interest. <math>\sigma > 0$ is a scale parameter and $\lambda > 0$ is a penalty term, meant to prevent excessive functional complexity. As λ approaches 0, ρ_{τ}^{*} becomes equivalent to ρ_{τ} , the pinball loss used in quantile regression (Fasiolo et al., 2021).

Similarly to quantile regression, QGAMs do not make any assumptions on the distribution of the extremes, and only require defining a set of quantiles of interest to study the effect of the regressors on the output. Furthermore, similarly to GAMs, they model the relationship between the output and the regressors empirically, without requiring any previous knowledge of the functional relationship between the two.

A possible limitation of GAMs and QGAMs is that they are additive in nature, and even though interactions between terms may be modelled, every interaction has a large effect on the computational burden of the model, thus limiting de facto the number of regressors which may be added to the model.

2.3 Benchmarking EVT models

In order to compare the performance of the EVT models presented so far, we ap-235 ply them to full-complexity climate data. We specifically consider the association between 236 cold spells in North America and wet or windy extremes in Western Europe. Following 237 previous literature looking at these extremes (Messori et al., 2016; De Luca et al., 2020; 238 Leeding et al., 2023; Messori & Faranda, 2023), we consider surface extremes occurring 239 one to a few times per year (e.g. 1% - 5% extreme quantiles), rather than extreme events 240 with multiannual return times. A useful statistical model should be able to verify whether 241 any relation between surface temperature in North America and surface extremes in West-242 ern Europe is present, and, if this is the case, make use of this and additional informa-243 tion on the state of the North Atlantic atmospheric circulation to improve the predic-244 tion of said extremes. In particular, we identify three key characteristics of a useful EVT 245 model in this context: 246

1. It should provide a spatially resolved prediction of extremes in Western Europe,
 which is more accurate than competing models, given that similar information is pro vided.

250
2. It should provide a consistent estimate of the spatially resolved return levels of
the extremes in Western Europe, where consistency is the defined as the property of an
estimator whose probability of being arbitrarily close to the true value tends to one for
increasing sample size.

3. It should improve its performance whenever relevant information is added to the
 model.

The following steps are implemented to assess the performance of the models:

256

259

268

- For objective 1, we compute and compare the pseudo R² quantile regression goodness of fit introduced by Koenker and Machado (1999). This is defined as:

$$R_{pseudo}^2 = 1 - \frac{L_{complete}}{L_{baseline}},\tag{9}$$

where $L_{complete}$ is the pinball loss of the model of interest, and $L_{baseline}$ is the pinball loss of a model using the unconditional quantile $Q_Y(\tau)$ of the anomalies at a given grid point as a fixed prediction for all observations at that same grid point.

- For objective 2, we define \hat{PO} as the percentage of overpredictions in the test sample. An unbiased EVT model should overpredict the value of the output a percentage of times corresponding to the probability τ of the target conditional quantile $Q_{Y|X}(\tau)$, so that $\hat{PO} = \tau$. An estimate of the bias of the model is then given by the absolute difference between the percentage of overpredictions \hat{PO} and τ , so that

$$\widehat{Bias} = |\hat{PO} - \tau| \tag{10}$$

- For objective 3, we first use the models to make spatially resolved predictions of extreme wind or precipitation events in Western Europe based on the latitude and the longitude of each location. Then, we add some information on the upstream large-scale atmospheric circulation and compare the results. A model fulfilling objective 3 is expected to progressively improve its performance, given that more information is provided and that this information has been identified as relevant to the extremes in previous studies (Messori et al., 2016; Leeding et al., 2023).

Holdout cross-validation is performed by splitting the dataset into three parts: a training set, a validation set and a test set, containing approximately 50%, 25% and 25% of the available observations, respectively. The random split uses seasonal data blocks (we focus on the winter season, see Sect. 3) to minimise information leakage between the different sets. The models are trained using the training set, with feature selection aided by the use of the validation set in addition to previous research. All metrics of model performance are computed based on the test set.

3 Defining Pan-Atlantic Compound Extremes and Prediction Models

Atmospheric data are taken from the ERA 5 global reanalysis (Hersbach et al., 2020) 285 with a daily time resolution and a 0.5° horizontal resolution. We consider the period Novem-286 ber 1959–January 2022, and the months of November, December, January and Febru-287 ary (NDJF). The daily NAO values, are taken from the NOAA online archive (NOAA/ 288 National Weather Service, 2022). Cold spells in North America are defined as days with 289 2m temperature (t2m) anomalies relative to the daily climatology, area-averaged over 290 30° - 45° N, 100° - 70° W, below the 5th quantile of the distribution. The daily clima-291 tology is obtained by applying a 7 days running mean and then averaging over each cal-292 endar day over the full time period. The domain of the North American cold spells fol-293 lows Messori et al. (2016) and is illustrated in Figure 1, together with the t2m anoma-294 lies associated with the selected cold spells. To decluster the extremes, we require that 295 at least five days elapse between separate cold spells. Whenever several days within a 296 five-day period meet the criteria for being classified as a cold spell, only the first day is 297 selected. The choice of selecting the first day rather than the coldest aims to prevent the 298 skill of the statistical models from being affected by the duration of the cold spells. 299



Figure 1. Domain of North American cold spells and mean t2m anomaly during cold spells. Darker shades are associated with larger negative t2m anomalies.

For Western Europe, we focus on daily mean 10m wind speed and daily precipitation over Iberia and Western France: regions that Messori et al. (2016) and Messori and Faranda (2023) highlighted as experiencing large anomalies following the North American cold spells. We test the statistical model predictions on the 95th and 99th quantiles of the local distributions of these two variables. These correspond to return periods of approximately six times and once per winter season, respectively.

Based on previous work, we consider the characteristics of the Polar – or eddy-driven 306 - jet stream over the North Atlantic region as a possible predictor of the European ex-307 tremes (Messori et al., 2016; Leeding et al., 2023). Jet stream speed is defined as the largest 308 zonally averaged zonal wind anomaly from a seven-day smoothed climatology at 250 hPa 309 over a North Atlantic domain spanning 30° - 75° N, 70° - 5° W. The location of the jet 310 is given by the latitude displaying the largest zonal mean zonal wind anomaly as defined 311 above. The choice of time lag of temperature in North America and zonal wind over the 312 North Atlantic to be included in the models has been aided by cross-validation. The lag 313 is relative to the prediction date for European extremes. Only one lag for each variable 314 is included in the models in order to avoid multicollinearity issues, as both t2m in North 315 America and the above jet indices display a high degree of autocorrelation. 316

We provide the statistical models with three levels of information for predicting t2m and daily mean 10m wind speed in Western Europe:

- Basic models, making predictions as a function of time, latitude, longitude, and a training-set based seasonal climatology, only.

- Cold spell models, where t2m in North America from two days prior to the prediction date is added to the regressors.

- Cold spell and jet stream models, where the NAO and strength and location of the Polar jet from one day prior are also added to the regressors. The NAO values are included here as a way to control for possible confounders affecting North American surface temperatures and surface weather in Western Europe, which might otherwise lead to biased estimates of the association between the two.

For each information level, we build three models: a QGAM, a linear quantile regression (QREG) and a POT model. Therefore, a total of nine models is estimated for each target variable in Western Europe. The models' performance on the test data is compared as outlined in Subsection 2.3. Technical details including the exact formulation of each model can be found in Supporting Information text S2.

4 Non-parametric EVT in practice: a QGAM approach to the study of Pan-Atlantic Compound Extremes

This section applies the models presented in Sections 2 and 3 to study Pan-Atlantic compound extremes and their drivers. It aims to support previous analyses of the connection between cold spells in North America and wet or windy extremes in Europe (Messori et al., 2016; De Luca et al., 2020; Leeding et al., 2023; Messori & Faranda, 2023) through robust statistical estimation, and verify whether QGAMs outperform alternative statistical models.

A first overview of the relation between cold spells in North America and wet or 341 windy extremes in Western Europe can be obtained by means of a composite analysis 342 (Fig. 2). There is a heightened frequency of positive anomalies in 10m wind speed and 343 precipitation in Western Europe in conjunction with cold spells in North America (Fig. 344 2a, b, c), with the largest effect in Western Europe being observed immediately after the 345 cold spell in North America. This is particularly true for middle-sized to large positive 346 anomalies, pointing to the possibility of near-simultaneous extremes in the two regions. 347 Fig. 2d, e, f corroborate this hypothesis, by showing how the mean, the 95th and the 99th 348 quantiles of daily mean 10m wind speed and daily precipitation in Western Europe are 349 significantly higher than usual in the aftermath of North American cold spells (Fig. 2d-350 351 f).



Figure 2. **a-c**: Kernel density estimate of daily mean 10m wind speed (left) and daily precipitation anomalies (right) in Western Europe: climatological distribution (blue) and distribution of events two days before (**a**), the same day (**b**), and two days after (**c**) a cold spell in North America (red). Mean (**d-e**), 95th quantile (**f-g**) and 99th quantile of (**h-i**) of daily mean 10m wind speed (left) and daily precipitation anomalies (right) in Western Europe 15 days before and after a cold spell in North America. The solid black lines are the overall mean/quantile, while the dashed lines mark approximate 95% significance levels, computed by means of a Monte Carlo permutation test with 20000 replications.

We next test whether QGAMs and the other models introduced in Section 2 can 352 leverage this association to make statistical forecasts of extreme events in Europe (95th 353 and 99th quantiles of daily mean 10m wind speed and daily precipitation at locations 354 in Iberia and Western France). We present the results for the three QGAM models de-355 scribed in Section 3 (basic, cold spell and cold spell and jet stream), and compare their 356 performance to the quantile regression and POT models. 357

QGAMs can identify the relation between t2m in North America and weather ex-358 tremes in Western Europe. Figure 3 shows the partial effect of t2m in North America 359 on near-surface weather in Western Europe, when holding all other variables included 360 in the model constant. Lower temperatures in North America are significantly associ-361 ated with higher values of the extreme quantiles and therefore more extreme weather events 362 in Western Europe. This effect is at its strongest for temperature anomalies of two stan-363 dard deviations below the mean, i.e. cold spells. In the model also including jet stream 364 and NAO information (Fig. 3e-h), the effect of North American temperatures is weaker 365 due to the fact that part of the effect is likely mediated by the jet and/or NAO. 366



Figure 3. Partial effect of t2m anomalies in North America at lag -2 days on the 95th and 99th quantiles of daily mean 10m wind speed (red) and daily precipitation (blue) in Western Europe generated through QGAMs (standardised variables). The effect is measured in terms of standard deviations of the dependent variable, when holding all other variables in the model fixed. a-d: Cold spell model, holds latitude, longitude and time-variable fixed. e-h: Cold spell and jet stream model, holds all of the above, NAO, and strength and location of the jet stream fixed. The p-values indicate the overall significance of the smooth term in the respective model.

374

Figure 4 shows the bias associated with the prediction of extreme quantiles of daily mean 10m wind speed in Western Europe through QGAMs. Ideally, the bias of our models should be held under 1- τ for most grid points, since a model with bias greater than 1- τ has a larger bias than a model providing a systematic underprediction. At the same time, zero bias is not an aim itself due to the variance-bias trade-off, and thus some bias is to be expected. Figure 4 suggests that QGAMs mostly have a bias lower than 1- τ when predicting daily mean 10m wind speed for most grid points. The models predicting the 95th quantile may appear to perform better than those predicting the 99th quantile, but

this is mostly a question of scale, since the figure colourscale is always capped at $1-\tau$,

which is lower for the 99th than for the 95th quantile.



Figure 4. Estimated bias of QGAMs in terms of absolute distance between the percentage of overpredictions (\hat{PO}) and the theoretical quantile (τ) . Estimation of 95th quantile of daily mean 10m wind. Basic model (a), model with information on t2m in North America at lag -2 days (b), model with same information as above plus jet stream and NAO information at lag -1 days (c). d-f: As a-c, but for the 99th quantile of daily mean 10m wind speed. In each row, the colourscale is capped at $1-\tau$.

Similar conclusions can be drawn for daily precipitation extremes (Figure 5). The bias of the models at most grid points is acceptable, as it is well under the $1-\tau$ threshold.



Figure 5. As Fig. 4 but for daily precipitation.

Figures 6 and 7 show the performance of QGAMs in terms of pseudo R^2 . The pseudo R² can take any values between one and minus one, where one represents a perfect model and minus one the worst possible model. Here, a pseudo R^2 over zero indicates that the model is better than the seasonal climatology of the quantile of interest at the given grid point, whereas values under zero indicate that it performs worse than the seasonal climatology. Stippling is added to grid points with a pseudo R^2 lower than zero.

All QGAMs predicting extreme quantiles of daily mean 10m wind speed appear 386 to gradually improve their performance as they are provided with more information on 387 the upstream large-scale atmospheric state. Sizeable gains can be observed already when 388 adding information on t2m to the models, so that cold spell models (Figures 6 b and e) 389 are better than the seasonal climatology alone at most grid points. However, the largest 390 gains occur at the last step, after the information on the state of the North Atlantic and 391 the Polar jet is added. Figures 6 c and f show that cold spell and jet stream models are 392 better than the seasonal climatology alone at all grid points apart from on the Pyrenees 393 and some mountainous areas in South-Eastern Spain, where the difference between the 394 two models is in any case small. 395



Figure 6. As Figure 4 but for pseudo \mathbb{R}^2 . Stippling indicates that the QGAM performs worse than the seasonal climatology of the quantile of interest.

The overall trend is similar for models predicting extreme quantiles of daily pre-396 cipitation. All the models in Figure 7 show gradual improvements as more information 397 is added, with the largest improvement occurring as above at the last step. This sug-398 gests that information on the state of the North Atlantic atmospheric circulation is key 399 to the predictability of surface extremes in Western Europe following North American 400 cold spells. The basic model is approximately equivalent to the seasonal climatology of 401 the quantile, whereas the cold spell and jet stream models systematically outperform the 402 seasonal quantile (and the base model) for both 10m wind speed and precipitation. Even 403 in the case of precipitation, QGAMs perform poorly in some mountainous areas, sug-404 gesting that they may have difficulties in accounting for the effect of local orographic fea-405 tures. 406



Figure 7. As Figure 5 but for pseudo R^2 . Stippling indicates that the QGAM performs worse than the seasonal climatology of the quantile of interest.

Figure 8 and 9 compare the performance of QGAMs to conventional POT models. The same set of variables is used for the two models, and a new Pseudo R^2 is computed using the POT models as baseline. A negative pseudo R^2 at a given grid point is to be interpreted as POT models performing better for that grid point, whereas a positive pseudo R^2 suggests that the QGAMs perform better.



Figure 8. As Figure 6, but using the linear POT model as baseline for the computation of the pseudo R^2 . Stippling indicates that the QGAM performs worse than the baseline model.

QGAMs become gradually better than POT models at most grid points as more 412 information is added to the models, both when predicting daily mean 10m wind speed 413 (Figure 8) and daily precipitation (Figure 9). This suggests that QGAMs are better than 414 POT models at modelling the non-linear effects of upstream atmospheric factors on the 415 surface extremes as a whole. However, some regional differences can be observed. In the 416 case wind, QGAMs are clearly superior for Western and Central Iberia, whereas the dif-417 ference is smaller in Eastern Iberia and Western France. POT models outperform QGAMs 418 in mountainous areas and in North-Western France. In the case of precipitation, QGAMS 419 outperform POT models almost everywhere, with the exception of North-Western Iberia 420 and the Pyrenees. The fact that the QGAMs struggle in mountainous areas is consis-421 tent with what found for the comparison with the seasonal climatology of the quantile 422 (Figures 6 and 7) 423



Figure 9. As Figure 7, but using the linear POT model as baseline for the computation of the pseudo R^2 . Stippling indicates that the QGAM performs worse than the baseline model.

Figures 10 and 11 display a comparison between QGAMs and QREG models. The comparison is performed similarly to the previous case, where in this case the QREG models are used as baseline for the Pseudo R² computation.



Figure 10. As Figure 6, but using the quantile regression model as baseline for the computation of the pseudo R^2

The difference between the two models is overall smaller in this case, with QREG 427 models generally performing better for simpler models and QGAMs improving their per-428 formance relative to QREG models as more information is added (Figures 10 and 11). 429 A difference between the models predicting daily mean 10m wind speed and those pre-430 dicting daily precipitation is that in the first case QGAMs show significant improvements 431 compared to QREG models only when information on both t2m in North America and 432 the Polar jet stream and NAO are added to the model (Figure 10), whereas in the sec-433 ond case the trend is less clear, as QGAMs outperform QREG models at most locations 434 already when information on t2m is added to the model (Figure 11). In the cold spell 435 and jet stream models, QGAMs outperform QREG models in inland Iberia, perform sim-436 ilarly to QREG models on the coast in North-Western France, and are outperformed in 437 the Pyrenees and other mountainous regions (Figures 10 c,f and 11 c,f). This difference 438 appears to be consistent across output variable and quantile of choice. 439



Figure 11. As Figure 7, but using the quantile regression as baseline for the computation of the pseudo R^2

An overview of the overall performance of the cold spell and jet stream models in terms of Pseudo R^2 is given in Table 1. As suggested from previous figures, QGAMs perform overall better than conventional alternatives (corresponding to positive Pseudo R^2 values in the table 1) when both t2m and atmospheric circulation information is provided. This holds in all cases bar when estimating the 95th quantile of 10m wind speed, where QREG models perform approximately at the same level.

Table 1. 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 seasonal climatology	0.0476	0.0495
Daily mean 10m wind speed	POT	0.0316	0.0353
Daily mean 10m wind speed	QREG	-0.0036	0.0085
Daily precipitation	Quantile of seasonal climatology	0.0982	0.0830
Daily precipitation	POT	0.0713	0.0518
Daily precipitation	QREG	0.0125	0.0155

In the supporting information (Figures S1-S8), we repeat the same analysis per-446 formed in this section while adding the previous lag of the variable of interest to the mod-447 els. This is done to show how the comparison between models changes when we correct 448 for the autocorrelation in the extremes. This autocrorrelation is ignored here since the 449 models do not make use of any information on the outcome variable of interest in Eu-450 rope, except for the training set-based seasonal climatology. The trends observed in the 451 main analysis largely hold even in the supporting materials, with the key difference be-452 453 ing that QGAMs improve further in comparison to other models.

5 Discussion and Conclusions

This paper has introduced QGAMs (Fasiolo et al., 2021) as an alternative to con-455 ventional parametric methods for the analysis of spatially compounding climate extremes. 456 Through a case study on pan-Atlantic cold spells in North America and wet or windy 457 extremes in Europe, it has been shown that non-parametric quantile-based methods gen-458 erally forecast near-surface weather extremes with a short return period more accurately 459 than through a conventional POT approach (Tables 1 and S1). The difference between 460 QGAMs and QREG models is relatively small, with QGAMs generally being superior 461 when introducing information on the state of Polar jet and the North Atlantic atmosphere into the models (Figures 10 and 11 and Tables 1 and S1). This suggests that QGAMs 463 may be recommended over other techniques when a larger number of drivers is explored. 464 The advantage of QGAMs over alternative techniques widens when information on the 465 autocorrelation of the extremes is added to the models (Figures S7 and S8). 466

Despite the overall superiority of QGAMs, some interesting regional differences could 467 be observed, with QGAMs performing at their best in inland Iberia, and at their worst 468 in the Pyrenees and in other mountainous regions (Figures 6 and 7). The poor perfor-469 mance of QGAMs in those regions is probably to be ascribed to the lack of orographic 470 information in the models and the relatively coarse spatial resolution, which make it hard 471 for QGAMs to reconstruct realistic spatial patterns. The fact that other models seems 472 to suffer less from this may be due to the fact that simpler linear models give greater 473 relative importance to the seasonal climatology compared to spatial features. 474

Non-parametric quantile methods have additional applications for the analysis of 475 near-surface extremes, which we have touched upon in this paper. First, they may be 476 used to assess the impact of a particular driver on a given quantile of a downstream atmospheric variable of interest. This is particularly useful for spatially compounding ex-478 tremes, namely extremes that occur (near-)simultaneously in geographically remote re-479 gions. In this study, we used QGAMs to show that lower area-averaged 2m temperatures 480 in North America are significantly associated with higher values of the extreme quan-481 tiles of daily mean 10m wind speed and daily precipitation in Western Europe (Figures 482 2 and 3). Second, non-parametric quantile-based methods may be used to provide ranges 483 of uncertainty to deterministic numerical forecasts. In particular, Figures 4 and 5 show 484 that QGAMs overpredict the value of the variable of interest a percentage of times close 485 to τ , thus displaying a good empirical coverage of the upper boundary of uncertainty when 486 used for estimation of the range of uncertainty of the forecast. 487

This paper focused on a specific case study of previously studied spatially compound-488 ing extremes. We considered a limited range of conventional models for comparison to 489 QGAMs, and only tested a small number of possible large-scale dynamical drivers of the 490 extremes. Our work should be viewed as a proof-of-concept to show the potential of QGAMs 491 compared to conventional parametric models for the study and understanding of spa-492 tially compounding extremes, rather than an attempt to build a statistical forecast model 493 or to investigate novel extreme occurrences and the related large-scale atmospheric drivers. 494 We also note that, even though QGAMs perform better than linear POT models in fore-495 casting near-surface extremes with short return periods, it does not mean that they are 496 equally effective for extremes with longer return periods. More research may be needed 497 to verify the robustness of QGAMs when analysing extremes of this nature. 498

The statistical analysis in this paper should also be contextualised relative to pre-499 vious research on Pan-Atlantic compound extremes (Messori et al., 2016; De Luca et al., 500 2020; Leeding et al., 2023; Messori & Faranda, 2023). Our results strengthen the hypoth-501 esis of a connection between wintertime North American cold spells and wet or windy 502 extremes in Western Europe, by showing that introducing information on surface tem-503 504 perature in North America has a clear effect in the model and significantly improves the prediction of extreme quantiles of 10m wind and precipitation in Iberia and Western France 505 (Figures 3, 6, 7). The fact that the effect of temperature in North America is weakened 506 but still significant when adding information on the Polar jet and the state of the North 507 Atlantic atmosphere to the models, points to the presence of teleconnections which can-508

not be fully explained by a simple causal flow in which cold spells influence the behaviour
of the Polar jet, which in turn affects surface weather in Europe. This might suggest the
presence of more complex or multiple pathways through which Pan-Atlantic compound
extremes may be engendered.

513 6 Open Research

The ERA 5 data used in this study is freely available from the Copernicus Climate 514 Change service at https://doi.org/10.24381/cds.adbb2d47 (Hersbach et al., 2020) 515 and https://doi.org/10.24381/cds.bd0915c6(Hersbach et al., 2020). The daily NAO 516 index is available through the NOAA online archive at https://www.cpc.ncep.noaa 517 .gov/products/precip/CWlink/pna/nao.shtml (NOAA/ National Weather Service, 518 2022). The models have been built and tested with the help of open-source R Statisti-519 cal Software (R Core Team, 2021). All software is freely available through CRAN at https:// 520 cran.r-project.org/. More details can be obtained by contacting the corresponding 521 author L.Olivetti, leonardo.olivetti@geo.uu.se. 522

523 Acknowledgments

L. Olivetti and G. Messori thankfully acknowledge the support of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project CENÆ: compound Climate Extremes in North America and Europe: from dynamics to predictability, Grant Agreement No. 948309). The authors also acknowledge valuable discussions with R. Leeding, J. Riboldi and A. Segalini.

⁵²⁹ 7 Conflict of interest

530

The authors declare no conflicts of interest relevant to this study.

531 **References**

- ⁵³² Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values.
- Coumou, D., & Rahmstorf, S. (2012, July). A decade of weather extremes. Na ture Climate Change, 2(7), 491–496. (Number: 7 Publisher: Nature Publishing
 Group) doi: 10.1038/nclimate1452
- Davison, A. C., & Smith, R. L. (1990). Models for Exceedances over High Thresh olds. Journal of the Royal Statistical Society. Series B (Methodological), 52(3),
 393-442. (Publisher: Royal Statistical Society, Wiley)
- ⁵³⁹ De Luca, P., Messori, G., Pons, F. M. E., & Faranda, D. (2020). Dynamical systems
 ⁵⁴⁰ theory sheds new light on compound climate extremes in Europe and East ⁵⁴¹ ern North America. *Quarterly Journal of the Royal Meteorological Society*,
- ⁵⁴² 146(729), 1636–1650. doi: 10.1002/qj.3757 ⁵⁴³ Dodet, G., Castelle, B., Masselink, G., Scott, T., Davidson, M., Floc'h, F., ...
- Dodet, G., Castelle, B., Masselink, G., Scott, T., Davidson, M., Floc'h, F., ...
 Suanez, S. (2019). Beach recovery from extreme storm activity during the
 2013–14 winter along the Atlantic coast of Europe. Earth Surface Processes
 and Landforms, 44(1), 393–401. doi: 10.1002/esp.4500
- do Nascimento, F. F., Azevedo, A. A., & Ferraz, V. R. d. S. (2021). Regression mod els to dependence for exceedance. *Journal of Applied Statistics*, 48(16), 3048–
 3059. doi: 10.1080/02664763.2020.1795088
- Elvidge, S., & Angling, M. (2018, April). Using Extreme Value Theory for Determining the Probability of Carrington-Like Solar Flares. Space Weather, 16, 417-421. doi: 10.1002/2017SW001727
- Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., & Goude, Y. (2021, July). Fast
 calibrated additive quantile regression. Journal of the American Statistical As-

555	sociation, 116(535), 1402–1412,
555	Ferreira A & Haan L d (2015 February) On the block maxima method
550	in extreme value theory: PWM estimators The Annals of Statistics
557	/3(1) 276–298 (Publisher: Institute of Mathematical Statistics) doi:
550	$\frac{10}{1214}$ $\frac{112}{14}$ $$
559	Fisher B \land by Tinnett L H C (1028 \land pril) Limiting forms of the frequency
560	distribution of the largest or smallest member of a sample Mathematical Pro-
501	ceedings of the Cambridge Philosophical Society 2/(2) 180–100 (Publisher:
502	Cambridge University Press) doi: 10.1017/S030500/100015681
505	French I Kokoszka P Stoov S & Hall I (2010 March) Quantifying the
564	righ of heat waves using extreme value theory and spatio temporal fune
565	tional data Computational Statistics & Data Analysis 121 176 103
500	10 1016/j csdp 2018 07 004
567	Cumbol F I (10/1) The Boturn Poriod of Flood Flows The Annals of Mathemat
568	<i>ical Statistice</i> 12(2) 163–100 (Publisher: Institute of Mathematical Statis
569	(1 ubisfield institute of Wathematical Statistics)
570	Cumbel F. I. (1058). Statistics of Extremes. Columbia University Proce. (Publics
571	tion Title: Statistics of Extremes) doi: 10.7312/gumb02058
572	Hastia T & Tibahirani P (1086 August) Conservatized Additive Models Statistic
573	nastie, 1., & Hoshirani, R. (1980, August). Generalized Additive Models. Stutisti-
574	Lowbach H. Doll P. Domisford D. Hindhand S. Honónyi A. Muñoz Sabatan J.
575	Thersbach, H., Dell, D., Dellisiond, F., Illianara, S., Horanyi, A., Mulloz-Babater, J., Thersbach, I. N. (2020) The FDA5 global reconclusion Orientaria Learned of
576	the Bouel Meteorological Society 1/6(730) 1000 2040 doi: 10.1002/cj.2803
577	Uindle W (1000) Applied Neprocessical Comparison Comparison University Process
578	hardle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.
579	James, G., Hastle, I., Witten, D., & Hosnirani, R. (2022). An introduction to sta-
580	tistical learning with applications in K (Second ed.). New York: Springer Sci-
581	ence. Veenken D. (2011 Nevershen) Additive medele for eventile remeasion. Medel selec
582	Koenker, R. (2011, November). Additive models for quantile regression: Model selec-
583	tion and confidence bandalds. <i>Brazilian Statistical Association</i>) doi: 10.1214/10. 05(2) 220.262 (Publisher: Prezilian Statistical Association) doi: 10.1214/10.
584	25(5), $259-202$. (Fublisher: Drazinali Statistical Association) doi: 10.1214/10 B IDS121
585	-DJI JIJI Kaankan D & Degestt C (1079) Degregation Quantiles Econometrics $l_{\ell}(1)$ 22
586	50 (Dublishow [Wiley Econometric Society]) doi: 10.2207/1012642
587	50. (Fublisher: [Whey, Econometric Society]) doi: 10.2507/1915045
588	Koenker, R., & Hallock, K. F. (2001, December). Quantile Regression. Journal of Economic Parameticae $15(4)$ 142 156 doi: 10.1257/iop.15.4.143
589	Koopkon P. & Machada I. A. E. (1000 December). Coodness of Fit and Polated
590	Informace Processor for Quantile Regression — Lowrnal of the American Statistic
591	cal Association 0/(448) 1206–1310 (Publisher: Taylor & Francis) doi: 10
592	$1080/01691450 \ 1000 \ 10473889$
593	Kornhuber K Courney D Vogel F Losk C Donges I F Lehmann I k
594	Horton R M (2020 January) Amplified Rossby waves enhance risk
595	of concurrent heatwayes in major breadbasket regions Nature Climate
590	Change 10(1) 48–53 (Number: 1 Publisher: Nature Publishing Group)
508	doi: 10.1038/s41558-019-0637-z
590	Lee M-Y Hong C-C & Hsu H-H (2015) Compounding effects of warm
599	sea surface temperature and reduced sea ice on the extreme circulation over
601	the extratropical North Pacific and North America during the 2013–2014
602	boreal winter Geophysical Research Letters 42(5) 1612–1618 doi:
603	10.1002/2014GL062956
604	Leeding B Biboldi J & Messori G (2023 March) On Pan-Atlantic cold wet
605	and windy compound extremes. Weather and Climate Extremes 39 100524
606	doi: 10.1016/i.wace.2022.100524
607	Lucarini, V., Faranda, D., Freitas, A. C. G. M. M. d. Freitas, J. M. M. d. Holland
608	M., Kuna, T., Vaienti, S. (2016). Extremes and Recurrence in Dynamical
609	Systems (1st edition ed.). Hoboken, New Jersev: Wilev.

- Mares, C., Mares, I., & Stanciu, A. (2009, March). Extreme value analysis in the
 Danube lower basin discharge time series in the twentieth century. *Theoretical* and Applied Climatology, 95(3), 223–233. doi: 10.1007/s00704-008-0001-0
- Messori, G., Caballero, R., & Gaetani, M. (2016). On cold spells in North America
 and storminess in western Europe. *Geophysical Research Letters*, 43(12), 6620–
 6628.
- 616Messori, G., & Faranda, D.(2023).On the Systematic Occurrence of Com-617pound Cold Spells in North America and Wet or Windy Extremes in618Europe.Geophysical Research Letters, 50(7), e2022GL101008.61910.1029/2022GL101008

620

621

622

632

633

634

- Mills, E. (2005, August). Insurance in a Climate of Change. Science, 309(5737), 1040–1044. (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1112121
- NOAA/ National Weather Service. (2022). Climate Prediction Center Teleconnections: North Atlantic Oscillation. Retrieved 2022-12-22, from https://www
 .cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml
- Passow, C., & Donner, R. V. (2019). A Rigorous Statistical Assessment of Recent
 Trends in Intensity of Heavy Precipitation Over Germany. Frontiers in Envi ronmental Science, 7.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. The An nals of Statistics, 3(1), 119–131. (Publisher: Institute of Mathematical Statis tics)
 - R Core Team. (2021). R: A Language and Environment for Statistical Computing [software]. Retrieved from https://www.R-project.org/ (Vienna, Austria. R Foundation for Statistical Computing)
- Smith, R. L. (1984). Threshold Methods for Sample Extremes. In J. T. de
 Oliveira (Ed.), *Statistical Extremes and Applications* (pp. 621–638). Dordrecht:
 Springer Netherlands. (doi: 10.1007/978-94-017-3069-3_48)
- Trenary, L., DelSole, T., Tippett, M. K., & Doty, B. (2015, December). Was the
 Cold Eastern Us Winter of 2014 Due to Increased Variability? Bulletin of the
 American Meteorological Society, 96(12), S15–S19. doi: 10.1175/BAMS-D-15
 -00138.1
- Wild, S., Befort, D. J., & Leckebusch, G. C. (2015, December). Was the Extreme
 Storm Season in Winter 2013/14 Over the North Atlantic and the United
 Kingdom Triggered by Changes in the West Pacific Warm Pool? Bulletin of
 the American Meteorological Society, 96(12), S29–S34. (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological
 Society) doi: 10.1175/BAMS-D-15-00118.1
- Wood, S. (2017). Generalized Additive Models: An Introduction With R (Vol. 66). Productivity Press.
- Yee, T. W. (2015). Vector Generalized Linear and Additive Models. Springer.

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}

 $^{1}\mathrm{Department}$ of Earth Sciences, Uppsala University

 $^2\mathrm{Centre}$ of Natural Hazards and Disaster Science, Uppsala University

³Department of Meteorology, Stockholm University

⁴Bolin Centre for Climate Research, Stockholm University

⁵Department of Statistics, Uppsala University

⁶Department 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:

$$Q_{Y|X}(\tau) = s(lat, lon, k = 20, d = 2) + s(time, k = 5)$$

$$+ seaonsal_quantile_climatology + month, \tag{1}$$

X - 4 OLIVETTI ET AL.: A QGAM APPROACH FOR COMPOUND EXTREMES

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:

$$Q_{Y|X}(\tau) = s(lat, lon, k = 20, d = 2) + s(time, k = 5) + month$$

+seaonsal_quantile_climatology + s(US_temp_lag2, k = 5) (2)

cold spell and jet stream model:

$$Q_{Y|X}(\tau) = s(lat, lon, k = 30, d = 2) + s(time, k = 5) + month$$

+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), (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:

$$Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal_quantile_climatology + year + month$$
(4)

Cold spell model:

 $Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal_quantile_climatology$

$$+year + month + US_temp_lag2$$
(5)

Cold spell and jet stream model:

$$Q_{Y|X}(\tau) = Q_{Y|X}(\tau) = lat + lon + lat \cdot lon + seaonsal_quantile_climatology$$
$$+year + month + US_temp_lag2 + Jet_strength_lag1 + NAO_lag1$$
$$+lat_jet_lag1 + jet_proximity$$
(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 \mathbb{R}^2



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



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



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



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

the pseudo \mathbf{R}^2

Table S1. Overall pseudo \mathbb{R}^2 of cold spell and jet stream QGAMs used to forecast the 95thand 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				