

*Communications in
Applied
Mathematics and
Computational
Science*

**ON INFERENCE OF STATISTICAL REGRESSION
MODELS
FOR EXTREME EVENTS BASED ON
INCOMPLETE OBSERVATION DATA**

OLGA KAISER AND ILLIA HORENKO

vol. 9 no. 1 2014

ON INFERENCE OF STATISTICAL REGRESSION MODELS FOR EXTREME EVENTS BASED ON INCOMPLETE OBSERVATION DATA

OLGA KAISER AND ILLIA HORENKO

We present a computationally efficient, semiparametric, nonstationary framework for statistical regression analysis of extremes with systematically missing covariates based on the generalized extreme value (GEV) distribution. It is shown that the involved regression model becomes nonstationary if some of the relevant model covariates are systematically missing. The resulting nonstationarity and the ill-posedness of the inverse problem are resolved by deploying the recently introduced finite-element time-series analysis methodology with bounded variation of model parameters (FEM-BV). The proposed FEM-BV-GEV approach allows a well-posed problem formulation and goes beyond probabilistic a priori assumptions of methods for analysis of extremes based on, e.g., nonstationary Bayesian mixture models, smoothing kernel methods or neural networks. FEM-BV-GEV determines the significant resolved covariates, reveals directly their influence on the trend behavior in probabilities of extremes and reflects the implicit impact of missing covariates. We compare the FEM-BV-GEV approach to the state-of-the-art GEV-CDN methodology (based on artificial neural networks) on test cases and real data according to four criteria: (1) information content of the models, (2) robustness with respect to the systematically missing information, (3) computational complexity and (4) interpretability of the models.

1. Introduction

Modeling of extreme events plays a crucial role in different areas of science (e.g., in weather/climate research, economics, biology/medicine) Simulation and prediction of such events is challenging since by definition they are rare and occur irregularly. To approach these challenges, statistical modeling of extreme events is widely accepted (as an alternative to deterministic physical/mathematical modeling based on “first principles”). Extreme value analysis (EVA) is a standard tool in statistics for description of probability distributions of extremes; its foundations were laid by

MSC2010: primary 62G05, 62G32, 65R32; secondary 65C50, 62F03.

Keywords: generalized extreme-value distribution, systematically missing information, nonstationary time-series analysis, nonparametric statistics, finite-element method.

E. Gumbel, W. Weibull and M. Fréchet already in the first half of the 20th century; for more details, see [13; 17; 19]. EVA distinguishes between two types of extremes: (a) exceedances over a higher or lower threshold and (b) block maxima or minima, e.g., annual extremes. In this work, we focus on block maxima. Consider a sample of identical and independently distributed (i.i.d.) variables Y_1, \dots, Y_n with common distribution $F(\cdot)$. Analogously to the limit distribution of the partial sums of the sample, described by the central limit theorem, there exists a limit distribution for the sample extremes $X_n = \max\{Y_1, \dots, Y_n\}$ (or $X_n = \min\{Y_1, \dots, Y_n\}$) as $n \rightarrow \infty$: the limit distribution of $\mathbb{P}[X_n \leq x] = F^n(x)$, as $n \rightarrow \infty$, is the generalized extreme value (GEV) distribution introduced by Fisher and Tippet (1928) and Gnedenko (1943), which has the form (see [17, Theorem 1.1.13])

$$G(x; \mu, \sigma, \xi) = \exp\left(-\left[1 + \xi \frac{x - \mu}{\sigma}\right]^{-1/\xi}\right) \quad (1)$$

with location, scale and shape parameters $\mu, \sigma, \xi \in \mathbb{R}$, respectively, and subject to $[1 + \xi(x - \mu)/\sigma] > 0$ and $\sigma > 0$. By fitting model (1) to a series of block maxima, we imply an unchanging behavior of the underlying dynamics (stationarity). This is obviously not always the case; e.g., in the context of climatology/meteorology, the monthly maxima of temperature or precipitation should be affected by the periodic seasonal effects. The most general way to release the stationarity assumption is to include external influence (denoted as covariates, modes or factors) by constructing the GEV parameters as functions of covariates, i.e., as regression models. And thus, the aim of data-based analysis of block maxima will be to infer the values of the GEV regression parameters from observed data. Standard state-of-the-art methods applicable to this task can be roughly divided into two groups: parametric and nonparametric regression approaches. In parametric approaches, the GEV parameters can be expressed as linear combinations of some explicitly known and given functions (e.g., sine/cosine functions to model the seasonal trends in meteorology). The nonlinearity of explicit covariate combinations is achieved deploying the standard tools from machine learning, e.g., artificial neural networks (ANNs) [8] and support vector machines (SVMs) [35]. Combination of GEV with a special form of ANN called conditional density-estimation network (CDN) has recently led to a creation of the GEV-CDN [8], a robust and flexible approach to the nonstationary and nonlinear extension of (1). However, GEV-CDN as well as all other parametric nonstationary extensions of (1) rely on the explicit availability of all of the relevant covariates and some strong probabilistic assumptions about the systematically missing/unresolved covariates, e.g., i.i.d. assumption for unresolved covariates. As a result, these methods implicitly assume time independence of regression coefficients (e.g., of the hidden neurons weights and biases in the case of GEV-CDN). But due to the multiscale nature of most of the realistic applications

(e.g., in climate research, economics or biology/medicine), one would never be able to guarantee that the set of information collected about the analyzed system is complete. One would also not be able to guarantee that all of the necessary probabilistic assumptions are fulfilled a priori for the analyzed system. As demonstrated in this paper, the violation of the i.i.d. condition for the systematically missing covariates leads to the conceptual and practical problems of the standard parametric approaches and may result in the wrong/biased analysis of the statistics of extremes.

Nonparametric approaches for regression analysis of block maxima exploit local likelihood smoothing [16] or Bayesian techniques [15; 42]. The limitations of these methods are their locality (e.g., a local stationarity assumption) and a priori parametric assumptions about the distributions of the GEV parameters. Another strategy is to involve mixture models and hidden Markov models (HMM) [38; 3; 2]. Such approaches require a priori knowledge about the probabilistic model for the time-dependent GEV parameters, e.g., stationarity and Markov assumptions for the hidden parameter switching process. Additionally, all of the above-mentioned state-of-the-art methods may lead to the ill-posed parameter identification problems (in sense of Hadamard [22]), resulting in the over-fitting of the available data. Most of the above approaches are realized as optimization algorithms for some nonlinear, nonconvex and often nondifferentiable quality functionals. That is, the obtained results are not unique and depend strongly on the initial values and other tuning parameters of the respective computational algorithms. One of the most frequently used methods of transforming an ill-posed problem to a well-posed one is called regularization. This approach is based on imposing some additional (reasonable) assumptions on the original problem formulation, e.g., that the solution should be “small” in an appropriately chosen norm [45]. In statistics and different areas of applied data analysis, approaches like Tikhonov and LASSO regularization are widely used in the context of, e.g., parametric regression and spline-interpolation problems [48; 44; 24], support vector machines [47], compressed sensing and matrix-completion methods [7].

Here we exploit a regularized and nonparametric strategy for general parameter identification in nonstationary problems [26; 27; 41]. It is based on the finite-element discretization (FEM) of the resulting inverse problem subject to bounded variation (BV) of the nonstationary model parameters in time. The FEM-BV framework allows computationally very efficient and highly scalable numerical implementation, either based on the adaptive FEM solvers (usually deployed for the adaptive numerical solution of partial differential equations) [25] or based on adaptive Markov-chain Monte Carlo (MCMC) schemes [18]. Resulting framework was demonstrated to be appropriate for a wide range of nonstationary inverse problems and applications, ranging from climate/weather research [26; 27], molecular dynamics [29] and DNA sequence analysis [41] to sociology [28] and economics [41].

In this paper, we present a unified approach for joint solving of all of the above discussed difficulties (i.e., the problem of systematically missing covariates, numerical complexity and the ill-posedness/over-fitting problem) in data-based analysis of block maxima. To address the issue of missing covariates, we exploit the central limit theorem for independent variables and express each GEV parameter by a fully non-stationary regression model, based only on resolved covariates, with a nonstationary additive noise. The resulting nonstationarity of (1) is interpolated by a linear convex combination of $K \geq 1$ local stationary models and a nonstationary switching process between them. The corresponding inverse problem is regularized by employing FEM-BV methodology [25; 41]. The resulting quality functional is optimized by adjusting the adaptive MCMC methodology (originally proposed in [18]) to a numerical solution of the resulting FEM-BV-GEV problem in optimization formulation.

The FEM-BV-GEV approach described in the current manuscript avoids a priori assumptions on stationarity and trend behavior of the GEV parameters. The proposed method allows an explicit data-driven recovery of the implicit impacts of unresolved modes in the situations when these unresolved modes are neither i.i.d. nor available in the measurement. The local linearity of the GEV parameter functions enables direct interpretation of the influence of covariates on the underlying dynamics of block maxima but can lead to the biased results in cases when the dynamics is locally highly nonlinear (i.e., in the scenarios where parametric GEV-CDN methodology based on neural networks is supposed to produce better estimates). We show that under appropriate assumptions FEM-BV-GEV includes/extends standard techniques based on HMM or kernel smoothing and, therefore, consider the nonlinear GEV-CDN approach as a main competitor for the FEM-BV-GEV in a series of numerical studies. This work shows that the resulting numerical framework, despite the local linearity, addresses the above-mentioned difficulties of the standard methods and demonstrates high robustness with respect to systematically missing covariates and is computationally highly efficient. These issues make the proposed methodology an adequate tool for analysis of extremes in very large realistic applications.

This manuscript is organized as follows. In Section 2, we derive in detail the FEM-BV-GEV approach. We compare FEM-BV-GEV to the state-of-the-art methods from conceptual and applied viewpoints in Section 3. The iterative numerical algorithm of FEM-BV-GEV is described in Section 4. In Section 5, we exemplify the application of FEM-BV-GEV and compare its performance with GEV-CDN [8] on test-cases and regression analysis of temperature extremes (30-day maxima for the period between 1950 and 2011) for Lugano and Berlin. Comparison is performed according to the four criteria: (1) information content of the models (jointly measuring complexity and quality of the model fit), (2) robustness with respect to the systematically missing information, (3) computational complexity and (4) understandability/interpretability of the models.

2. FEM-BV-GEV

In this work, we focus on the fully time-dependent GEV distribution defined by its probability density function (pdf)

$$f(x; \mu(t), \sigma(t), \xi(t)) = c(t) \exp\left(-\left[1 + \xi(t) \frac{x - \mu(t)}{\sigma(t)}\right]^{-1/\xi(t)}\right), \quad (2)$$

where t denotes the time variable and $c(t)$ the normalization constant

$$c(t) = \frac{1}{\sigma(t)} \left[1 + \xi(t) \frac{x - \mu(t)}{\sigma(t)}\right]^{-1/\xi(t)-1} \quad (3)$$

and the model parameters have to fulfill the constraints

$$\left[1 + \xi(t) \frac{x - \mu(t)}{\sigma(t)}\right] > 0 \quad \text{and} \quad \sigma(t) > 0 \quad \forall t. \quad (4)$$

In order to address the time-dependence, we intend to express each GEV parameter as a function dependent on covariates as a linear regression model. However, in real applications, one is usually confronted with the problem that some (or most) potentially relevant covariates are missing in the measurements.

One possible source for the systematically missing covariates is the multiscale dynamics of the underlying process; e.g., processes in climate or molecular dynamics may involve multiple time and length scales [39; 40; 11]. That is, only observing modes on a slow time scale (resolved modes), we neglect modes on the faster scale (unresolved modes). An additional reason for the missing information/measurements is that, even on just one single time scale, one cannot resolve all covariates because one is interested in regression models with a finite number of degrees of freedom. In particular, this is true for regression analysis of extremes because of the relatively small statistics. Thus, we have to select a set of resolved covariates and to account for the influence coming from the systematically unresolved/missing information.

Several disciplines cover the issue of missing information; e.g., in statistical regression analysis, the issue of unresolved information is often addressed under the theme “unobserved heterogeneity” [6]. Thereby, the unobserved covariates are included via a stationary error term into the regression model and the posterior model depends on the a priori assumption about the distribution of this error term. However, there is often no closed expression of the posterior.

In this work, we reduce the involved linear regression model by splitting it into two linear parts, corresponding to resolved and unresolved modes, and incorporate the influence of unresolved modes as a nonstationary additive noise. In the following, we consider all possible modes dependent on time t and split them into resolved $U(t) \in \mathbb{R}^S$ and unresolved $U^{\text{un}}(t) \in \mathbb{R}^Q$ factors, further on denoted as $U_t = (u_1(t), \dots, u_S(t))$ and $U_t^{\text{un}} = (u_1^{\text{un}}(t), \dots, u_Q^{\text{un}}(t))$. Then we normalize the

latter and obtain

$$\mu(U_t, U_t^{\text{un}}) = \mu_0 + \sum_{s=1}^S \mu_s u_s(t) + \frac{1}{Q} \sum_{q=1}^Q v_q u_q^{\text{un}}(t), \quad (5)$$

where $\mu_s, s = 1, \dots, S$, and $v_q, q = 1, \dots, Q$, are the regression coefficients. Under the assumption that the unresolved modes are i.i.d. for all t , application of the central limit theorem reduces the unresolved modes to the additive noise

$$\mu(U_t) = \mu_0 + \sum_{s=1}^S \mu_s u_s(t) + \epsilon \quad \text{with } \epsilon \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}). \quad (6)$$

In real applications, the i.i.d. assumption may be too strong. Instead, we can apply the central limit theorem for independent variables in a formulation that requires a much weaker Lindeberg condition [36]. And in case the modes are not independent, the Karhunen–Loève transformation can be used to decorrelate the processes [37; 33]. Thus, under the assumption that the Lindeberg condition holds, we rewrite (5):

$$\begin{aligned} \mu(U_t, U_t^{\text{un}}) = \mu_0 + \sum_{s=1}^S \mu_s u_s(t) \\ + \underbrace{\frac{1}{Q} \sum_{q=1}^Q v_q (u_q^{\text{un}}(t) - \mathbb{E}[u_q^{\text{un}}(t)])}_{\rightarrow \epsilon(t)} + \frac{1}{Q} \sum_{q=1}^Q v_q \mathbb{E}[u_q^{\text{un}}(t)]. \end{aligned} \quad (7)$$

By inserting $\mu_0(t) = \mu_0 + (1/Q) \sum_{q=1}^Q v_q \mathbb{E}[u_q^{\text{un}}(t)]$ and $\epsilon(t) \sim \mathcal{N}(0, \hat{\sigma}(t))$ into (7), we obtain the reduced, nonstationary regression model:

$$\mu(t, U_t) = \mu_0(t) + \sum_{s=1}^S \mu_s u_s(t) + \epsilon(t). \quad (8)$$

Note that in the regression formulation (8) parameter $\mu_0(t)$ is a time-dependent function and not a constant number as in the case of parametric statistics. That is, application of parametric approaches would produce biased results, and thus, nonparametric statistical methods would be required in such situations. Moreover, without loss of consistency, we generalize (8) by releasing the stationarity assumption of the coefficients μ_s for $s = 1, \dots, S$:

$$\mu(t, U_t) = \mu_0(t) + \sum_{s=1}^S \mu_s(t) u_s(t) + \epsilon(t). \quad (9)$$

Analogously to (9), we express the scale and the shape parameters:

$$\sigma(t, U_t) = \sigma_0(t) + \sum_{s=1}^S \sigma_s(t) u_s(t) + \tilde{\epsilon}(t), \quad \tilde{\epsilon}(t) \sim \mathcal{N}(0, \tilde{\sigma}(t)), \quad (10)$$

$$\xi(t, U_t) = \xi_0(t) + \sum_{s=1}^S \xi_s(t) u_s(t) + \bar{\epsilon}(t), \quad \bar{\epsilon}(t) \sim \mathcal{N}(0, \bar{\sigma}(t)). \quad (11)$$

The regression models in (9)–(11), which are reduced to resolved covariates only, become stochastic. Thereby, each GEV model parameter has, e.g., a normal distribution as a prior in Bayesian inference context [13]. Since there is no closed formulation for the resulting posterior, MCMC-based algorithms can be used to obtain the distribution of the posterior [13]. In the current manuscript, we focus on the mean behavior of parameters and thus omit the normal distributed noise terms in (9)–(11). Please note, by considering the mean behavior, we obtain deterministic model parameters, which still account for the unresolved information through the nonstationary bias/off-set terms $\mu_0(t)$, $\sigma_0(t)$ and $\xi_0(t)$. The consideration of the complete stochastic regression model with explicit error terms remains for future study. Finally, the nonstationary GEV distribution (2) is parametrized by

$$\Theta(t) = (\mu_0(t), \dots, \mu_S(t), \sigma_0(t), \dots, \sigma_S(t), \xi_0(t), \dots, \xi_S(t)). \quad (12)$$

In this work, we aim to avoid a priori probabilistic or deterministic assumptions on $\Theta(t)$. Instead, we approximate the nonstationary distribution of block maxima by $K \geq 1$ local stationary GEV distributions and a hidden/latent switching process. Thereby, we consider a nonparametric and nonstationary hidden switching process in order to avoid a priori assumptions like stationarity or Gaussian or Markovian behavior (necessary for mixture models and HMMs [38; 3; 2]). Elimination of a priori assumptions implies ill-posedness of the optimization problem in the sense of [22]: in each time step, there will be infinitely more unknown variables than observations. To regularize the problem, we apply the FEM-BV methodology for time-series analysis introduced in [25; 26; 27]. FEM-BV formulates the inverse problem for nonstationary dynamical systems as a regularized variational problem by discretizing the hidden switching process with finite elements and restricting its bounded variation. In the following, we formulate the FEM-BV-GEV approach in two steps: (a) interpolation and (b) regularization.

2.1. FEM-BV interpolation. The FEM-BV approach assumes that the model parameter $\Theta(t)$ changes slower than the observed series $X(t)$ (in the following denoted by X_t with $t = 1, \dots, T$). Then the underlying dynamics can be approximated by a set of $K \geq 1$ local stationary models and a nonstationary switching process $\Gamma(t) = (\gamma_1(t), \dots, \gamma_K(t))$. For that, FEM-BV interpolates the model distance function $g(X_t, \Theta(t))$ (describing the error of the nonstationary model with parameters

$\Theta(t)$ at time t in reproducing the observation X_t) by a linear convex combination of K local stationary model distance functions.

The FEM-BV approach can be ranged into the class of clustering-based methods, where the K local stationary models correspond to K clusters and the switching process $\Gamma(t)$ is the affiliation of the data to one of the clusters. Most popular standard methods of data clustering (such as K -means, Gaussian mixture model, hidden Markov models, etc.) can be obtained as unregularized special cases of this more general FEM-BV-framework. These standard methods can be obtained in FEM-BV-framework for some specific choices of the model error function $g(X_t, \Theta(t))$; e.g., the choice of the l^2 -distance between X_t and $\Theta(t)$ as $g(X_t, \Theta(t))$ with no further regularization of $\Gamma(t)$ results in the standard K -means clustering [41]. The connection to classical mixture models and hidden Markov models is discussed in Section 3. We apply the FEM-BV interpolation to our problem by considering for each local GEV model the parametrization

$$\mu_i(U_t) = \mu_{i0} + \sum_{s=1}^S \mu_{is} u_s(t), \quad i = 1, \dots, K, \quad (13)$$

and analogous expressions for $\sigma_i(U_t)$ and $\xi_i(U_t)$ and defining the local model distance function as the local negative log-likelihood function with $\theta_i = (\mu_{i0}, \dots, \mu_{iS}, \sigma_{i0}, \dots, \sigma_{iS}, \xi_{i0}, \dots, \xi_{iS})$, $i = 1, \dots, K$,

$$g_{\text{GEV}}(X_t, \theta_i) = \log(\sigma_i(U_t)) + \left(1 + \xi_i(U_t) \frac{X_t - \mu_i(U_t)}{\sigma_i(U_t)}\right)^{-1/\xi_i(U_t)} + \left(1 + \frac{1}{\xi_i(U_t)}\right) \log\left(1 + \xi_i(U_t) \frac{X_t - \mu_i(U_t)}{\sigma_i(U_t)}\right). \quad (14)$$

Then for $\Theta = (\theta_1, \dots, \theta_K)$, the averaged (interpolated) model distance functional is defined by

$$\mathcal{L}(\Gamma(t), \Theta) = \sum_{t=1}^T \sum_{i=1}^K \gamma_i(t) g_{\text{GEV}}(X_t, \theta_i) \quad (15)$$

with constraints on model parameters

$$\left[1 + \xi_i(U_t) \frac{X_t - \mu_i(U_t)}{\sigma_i(U_t)}\right] > 0 \quad \text{and} \quad \sigma_i(U_t) > 0 \quad \text{for } t = 1, \dots, T \text{ and } i = 1, \dots, K \quad (16)$$

and with convexity constraints on $\Gamma(t) = (\gamma_1(t), \dots, \gamma_K(t))$

$$\sum_{i=1}^K \gamma_i(t) = 1, \quad t = 1, \dots, T, \quad (17)$$

$$\gamma_i(t) \geq 0, \quad t = 1, \dots, T, \quad i = 1, \dots, K. \quad (18)$$

2.2. FEM-BV regularization on $\Gamma(t)$. Minimization of (15) with constraints (16)–(18) is ill-posed. FEM-BV regularization exploits the observation that many real processes feature persistent behavior. Persistence can be imposed through the bounded variation of the switching process $\Gamma(t)$ by

$$|\gamma_i|_{\text{BV}(0,T)} = \sum_{t=1}^{T-1} |\gamma_i(t+1) - \gamma_i(t)| \leq C_i, \quad i = 1, \dots, K, \quad (19)$$

where C_i denotes the maximal number of allowed transitions between the model i and all the other models in the time interval $[1, T]$; further on, we will refer to $C = \max\{C_1, \dots, C_K\}$. Please notice that, since the natural boundary of C is given by T (the number of observed time steps), involving constraint (19) into the optimization problem does not confine the solution space. Now the minimization of (15) with constraints (16)–(19) is well-posed according to $\Gamma(t)$. In the following, we denote the minimization problem

$$(\Gamma^*(t), \Theta^*) = \underset{\Gamma(t), \Theta}{\operatorname{argmin}} \mathcal{L}(\Gamma(t), \Theta) \quad \text{with respect to constraints (16)–(19)} \quad (20)$$

as the FEM-BV-GEV approach and the optimal FEM-BV-GEV parameters as $(\Gamma^*(t), \Theta^*)$.

2.3. Model selection. In this section, we discuss how to choose the optimal FEM-BV-GEV parameters K and C . Moreover, we aim to detect the most significant combination of resolved covariates $U_t \in \mathbb{R}^S$ (S is the number of resolved covariates); in the following, we denote each combination by u_{comb} . Thus, for different K , C and u_{comb} , we apply the FEM-BV-GEV approach and obtain a candidate model M . Denoting the number of all possible K as N_K and the number of all possible C as N_C , we obtain in total

$$N_K N_C \sum_{s=1}^S \frac{S!}{(S-s)! s!} \quad (21)$$

different models and choose the optimal one according to model selection criteria, e.g., the second-order Akaike information criteria (AIC_c) [30]

$$\text{AIC}_c = 2L + 2|M| + \frac{2|M|(|M| + 1)}{T - |M| - 1}, \quad (22)$$

where L is the negative log-likelihood function for the estimated model M , $|M|$ denotes the number of parameters in model M and T is the length of the data. In FEM-BV-GEV formulation, the averaged model distance functional (15) corresponds to the averaged negative log-likelihood (NLL): $-L = \mathcal{L}(\Gamma(t), \Theta)$. The number of parameters depends on K , C and the dimension of u_{comb} so that

$|M| = |M(K, C, u_{\text{comb}})|$. AIC_c is a valid estimate for information content of data with finite length [5]. We compute AIC_c for each model M and choose the best model, denoted by M^* , with respect to $\min(\text{AIC}_c)$. Thus, incorporating the model selection criteria, the “complete FEM-BV-GEV formulation” is given by (20) and (22). In case S is big, the number of models M in (21) increases very fast and running over all possible combinations of U_t becomes computationally expensive. Instead, we can incorporate the LASSO shrinkage technique [44] on model parameter Θ :

$$|\theta_i|_{L_1} \leq C_L, \quad i = 1, \dots, K. \quad (23)$$

This constraint provides not only the most significant u_{comb} by setting the coefficients of insignificant covariates to zero but also improves the prediction accuracy of the regression by shrinking the coefficients [48; 24]. Also here we have to find the optimal C_L . Thus, with N_L as the number of all possible C_L , the total number of all models is $N_K N_C N_L$. The “LASSO FEM-BV-GEV formulation” is described by (20), (23) and (22), where now $|M| = |M(K, C, S)|$.

3. Conceptual comparison with standard methods

The FEM-BV-GEV is a semiparametric approach as a combination of the parametric GEV and the nonparametric FEM-BV description of the hidden switching process. The influence of unresolved factors, expressed as the nonstationarity of model parameters (9)–(11), is reflected by $\Gamma(t)$. The key issue that makes the FEM-BV-GEV problem well-posed is the fact that decreasing the value of C in (19) results in shrinking of the parameter space for $\Gamma(t)$, limiting the number of the local minima for $\mathcal{L}(\Gamma(t), \Theta)$ in (20). The current realization of the FEM-BV-GEV approach assigns $\gamma_i(t) \in \{0, 1\}$, $i = 1, \dots, K$, for all t . That is, according to the result in [41], interpolation of the model distance function provides the direct interpolation of the nonstationarity of the model parameter $\Theta(t)$:

$$\Theta(t) \approx \sum_{i=1}^K \gamma_i(t) \theta_i. \quad (24)$$

Moreover, the FEM-BV-GEV approach includes some state-of-the-art approaches as special cases: in case the whole information is provided for the regression analysis of extremes, FEM-BV-GEV with $C = 0$ (no transitions between the models and thus $K = 1$) corresponds to stationary parametric regression models and results in a well-posed inverse problem. For $K \geq 2$, FEM-BV-GEV provides a nonlocal extension of the nonparametric kernel smoothing approach: equivalently to adaptive multimodal optimization, the nonstationary switching process $\Gamma(t)$ enables us to consider all observations that underlie similar dynamics as one ensemble (in contrast

to, e.g., methods based on a moving window [16; 10] where the kernel is a priori chosen as some fixed local parametric function, e.g., a Gaussian of a certain width).

Furthermore, under an additional a priori assumption that $\Gamma(t)$ is a homogenous Markov process, FEM-BV-GEV is restrained to the family of hidden Markov models (HMMs) as applied to GEV, e.g., becoming equivalent to the standard HMM-based methods of extreme value analysis [2; 1]. However, the difference between the FEM-BV-GEV and Bayesian techniques for analysis of block maxima, e.g., based on mixture models or HMMs, is in the modeling of the hidden process or the hidden variable, respectively. In more detail, the main conceptual advantage of FEM-BV in its general form over the Bayesian mixture models, e.g., HMM, is that FEM-BV is a nonparametric approach. HMM is a purely parametric approach with strong a priori assumptions. In the HMM context, the hidden process is parametrized by a probabilistic model, e.g., homogenous Markov, and requires an initial hidden probability. In contrast, FEM-BV avoids a priori parametrization and is also applicable beyond these assumptions. The only two assumptions involved in FEM-BV are (1) $\Gamma(t)$ is considered as a function in some (very broad) function space, e.g., BV-space and (2) the smoothness of $\Gamma(t)$ is bounded in the corresponding function space (compare also discussions in [41]). The estimation of $\Gamma(t)$ results in a linear minimization problem [25] or can be carried out using MCMC techniques [18]. Both approaches result in very efficient computational numerical schemes scaling well for very large problems unlike the expectation maximization algorithm (the most prominent and widely used algorithm for Bayesian mixture models).

Thus, exploiting the advantages of FEM-BV and involving stochastic mode reduction for linear regression analysis, the proposed FEM-BV-GEV provides an unbiased estimator for GEV parameters in context of systematically missing information. However, the linearity assumption for the GEV parameters may impose a disadvantage as soon as the influence of covariates on the dynamics of extremes is nonlinear. Considering a set of numerical examples, we will compare the proposed FEM-BV-GEV to the intrinsically nonlinear GEV-CDN methodology, which exploits a conditional density network (CDN) for nonlinear regression analysis based on time-dependent covariates with time-independent (i.e., stationary) neuron weights and biases [8].

4. Implementation

In the following, we discuss the algorithmic implementation of the ‘‘FEM-BV-GEV framework’’. The FEM-BV-GEV approach was integrated in the existing FEM-BV MATLAB toolbox and can be provided by the authors on email request. The main steps of the general FEM-BV formulation are sketched in Algorithm 1: (1) for different K , C and u_{comb} , a candidate model M is estimated (see Algorithm 1, line 4),

input : Observed series X_t , list of u_{comb} , list of K , list of C
output : Optimal model M^* (u_{comb}^* and $(\Gamma^*(t), \Theta^*)$)

```

1 for  $u_{\text{comb}}$  do
2   for  $K$  list do
3     for  $C$  list do
4       Step 1  $(\Gamma^*(t), \Theta^*) = \text{getOptimalParameterSet}(u_{\text{comb}}, K, C)$ . For fixed  $u_{\text{comb}}$ ,
        $K$  and  $C$  estimate the global optimal parameter set  $(\Gamma^*(t), \Theta^*)$  (compare
       Algorithm 2).
5       Step 2  $M^* = \text{updateOptimalModel}(\Gamma^*(t), \Theta^*, u_{\text{comb}}^*)$ . Estimate the IC value
       according to (22) for every model  $M$ . If the current IC value is smaller than the
       previous one, assign  $M^* = M$ .

```

Algorithm 1: The general FEM-BV algorithm.

and (2) the optimal model M^* , i.e., optimal K^* , C^* and u_{comb}^* , is chosen according to AIC_c in (22) (see Algorithm 1, line 5). Thereby, for a fixed set $\{K, C, u_{\text{comb}}\}$, a model M is obtained by solving (20). The minimization is implemented as a subspace iteration (see Algorithm 2): starting with a randomly initialized $\Gamma(t)$, in an alternating order, we estimate Θ for a fixed $\Gamma(t)$ and then $\Gamma(t)$ for a fixed Θ , thereby obviously reducing in each step the value of (15). The subspace iteration converges to a local optimum. The convergence is achieved if the decrease of the averaged model distance functional (15) is less than a predefined minimization threshold Tol . To obtain the global optimum FEM-BV framework involves an annealing-like strategy: in each annealing step, $\Gamma(t)$ is initialized randomly (for more details on the general FEM-BV annealing-like algorithmic strategy sketched above, see [41]).

The two steps of the subspace iteration are carried out as follows. For a fixed parameter Θ , $\Gamma(t)$ is discretized by the finite element method and estimation of $\Gamma_{\text{opt}}(t)$ results in a linear constrained minimization problem that can be solved using standard numerical tools, e.g., the simplex method [26; 41]. For a fixed $\Gamma(t)$, Θ_{opt} is obtained by minimizing the averaged negative log-likelihood functional (15) with respect to the constraints (16). For minimization, we take advantage of the fact that the averaged model distance functional (15) is uncoupled for different $i = 1, \dots, K$. Thus, Θ_{opt} can be estimated by solving

$$\min_{\theta_i} \sum_{t=1}^T \gamma_i(t) g_{\text{GEV}}(X_t, \theta_i), \quad (25)$$

with respect to constraints (16) for $i = 1, \dots, K$, K times with standard likelihood maximization techniques [13; 19]. Note that the corresponding function in (25) is strongly nonlinear and nonconvex. Additionally, in practical applications, it

input : Observed series X_t , fixed $\{K, C, u_{\text{comb}}\}$, minimization threshold value Tol, number of annealing steps annealing, maximal number of subspace iterations maxSubspace

output : Global optimal parameter set $(\Gamma^*(t), \Theta^*)$

```

1  $\mathcal{L}(\Gamma^*(t), \Theta^*) = \inf$ 
2 for  $a = 1$  : annealing do
3      $\Gamma_{\text{old}}(t)$  generate randomly with respect to constraints (17)–(19)
4      $\Theta_{\text{old}} = \operatorname{argmin}_{\Theta} \mathcal{L}(\Gamma_{\text{old}}(t), \Theta)$ 
5     while  $|\mathcal{L}(\Gamma_{\text{opt}}(t), \Theta_{\text{opt}}) - \mathcal{L}(\Gamma_{\text{old}}(t), \Theta_{\text{old}})| > \text{Tol}$  or maxSubspace do
6         Step 1  $\Gamma_{\text{opt}}(t) = \operatorname{argmin} \mathcal{L}(\Gamma(t), \Theta_{\text{old}})$ . The constrained minimization with respect
            to  $\Gamma(t)$  results for BV-regularization in a linear problem; standard methods, e.g.,
            simplex method, can be applied.
7         Step 2  $\Theta_{\text{opt}} = \operatorname{argmin} \mathcal{L}(\Gamma_{\text{opt}}(t), \Theta)$ . The required numerical optimization method
            with respect to  $\Theta$  depends on the model distance function  $g(\cdot)$ . In FEM-BV-GEV,
             $g(\cdot)$  is the negative log-likelihood and the minimization is carried out by applying the
            MCMC method (compare Algorithm 3).
8     if  $\mathcal{L}(\Gamma^*(t), \Theta^*) > \mathcal{L}(\Gamma_{\text{opt}}(t), \Theta_{\text{opt}})$  then
9          $\Theta^* = \Theta_{\text{opt}}$ 
10         $\Gamma^*(t) = \Gamma_{\text{opt}}(t)$ 
    
```

Algorithm 2: getOptimalParameterSet: annealing and subspace iteration.

may be nondifferentiable (or may exhibit very large values of the first derivative). Because of these reasons, minimization using standard gradient-based methods like Newton’s method and gradient-descent approaches would strongly depend on the initial value and on the boundedness of the first derivatives (e.g., as in the case of the Levenberg–Marquardt optimization algorithm deployed in GEV-CDN [8]). To avoid this difficulty, we consider a gradient-free optimization technique based on the Metropolis algorithm, which is a Markov-chain Monte Carlo (MCMC) method. In particular, we employ the adaptive MCMC methodology proposed in [18], where the adaptive MCMC optimization method considers the Boltzmann distribution as the target density:

$$\pi(\cdot) = \frac{1}{z} \exp(-\beta h(\cdot)) \quad (26)$$

with normalization constant z , inverse temperature parameter β and some energy function $h(\cdot)$. For $\beta \rightarrow \infty$, Boltzmann-distributed samples converge towards the minimal energy of $h(\cdot)$. The adaptivity of the MCMC in [18] comes from adjusting the noise, used for proposing the next sample, and from increasing β (i.e., from annealing). Thus, this approach can be used as an optimization method to get Θ_{opt} for fixed $\Gamma(t)$. For that, we set $h(\Theta) = \mathcal{L}(\Gamma(t), \Theta)$ and modify the MCMC in [18] by adjusting the initialization and the proposed next step (taking into account the

constraints (16) and the dimensionality of Θ). The main steps of the deployed adaptive MCMC are sketched in Algorithm 3 in Appendix A.

We would like to emphasize that in each run of the MCMC algorithm it is sufficient to sample a parameter Θ_{new} that provides a smaller value of $\mathcal{L}(\Gamma(t), \Theta_{\text{new}})$ instead of sampling the whole distribution (refer to Algorithm 3, lines 3 and 9). The subspace iteration deployed by FEM-BV improves in each step the parameters in the sense of minimizing $\mathcal{L}(\Gamma(t), \Theta)$ and provides the optimal parameter set $(\Theta_{\text{opt}}, \Gamma_{\text{opt}})$ for each annealing step. Moreover, MCMC optimization techniques do not depend on the initial start values: the MCMC algorithm also allows us to accept parameters with higher value of the functional (15); thus, there is a chance to obtain the global minima starting from a bad initial value. As will be demonstrated on the numerical examples below in Section 5, the deployed MCMC optimization technique is efficient in terms of computational time.

5. Numerical examples

In this section, we illustrate the proposed FEM-BV-GEV methodology on two test cases and real data. The two test cases are used to investigate the robustness with respect to the systematically missing covariates, the approximation of nonstationary behavior and the computational performance of the framework (with respect to accuracy and computational time). In the real-data example, we analyze a series of block maxima surface temperatures for locations Lugano, Switzerland and Berlin, Germany. In each application, the performance of the FEM-BV-GEV framework is compared to the GEV-CDN approach. GEV-CDN exploits a conditional density network (CDN) for nonlinear regression analysis based on time-dependent covariates with constant weights and biases [8]. The GEV-CDN analysis is performed using the package GEV-CDN provided in the statistical toolbox R [8; 9]. The main tuning parameters of GEV-CDN are the number of hidden neurons (further on denoted by N_H) in the network, the hidden layer transfer function (identity or logistic function) and the number of trials (to avoid the local optima). In all of the numerical examples considered below, an optimal configuration of GEV-CDN with respect to these tuning parameters was determined according to the AIC_c criterion in the way as described in [8].

5.1. Stationary test case. The first example aims to verify the regression analysis of block maxima based only on resolved covariates. We would like to roughly mimic the true underlying dynamics of block maxima in real meteorological applications. Therefore, as covariates, we consider a linear trend, a periodic function with a one-year period and daily averaged measurements of the total solar intensity (TSI) [20; 21].¹ In general, the TSI factor describes the total amount of the solar radiative

¹Data were retrieved from <http://www.pmodwrc.ch/pmod.php?topic=tsi/composite/SolarConstant>.

	Optimal models for stationary test case			
	Settings	NLL	$ M $	AIC_c
FEM-BV-GEV	$K = 3, C = 4$	1.7173×10^3	38	3.5144×10^3
GEV-CDN	$N_H = 12$	2.1116×10^3	75	4.3703×10^3

Table 1. Optimal results for FEM-BV-GEV and GEV-CDN for a stationary test case. The exact negative log-likelihood for X_t using the original parameters is $NLL_{\text{exact}} = 1.7042 \times 10^3$. As described below, smaller values of NLL indicate the models with a better fit, whereas smaller values of AIC_c indicate more informative models.

energy that is hitting the earth’s upper atmosphere [21]. However, for this example, we consider only a segment of the TSI measurements (staring from the year 1950) of length $T = 800$, and thus, this factor is only responsible for more fluctuation in the generated block maxima. Now, with covariates $\widehat{U}_t = (u_1(t), u_2(t), u_3(t))$ defined by

$$u_1(t) = \frac{1}{400}t, \quad u_2(t) = \sin\left(\frac{\pi}{2} + \frac{1}{365}2\pi t\right), \quad u_3(t) = \text{TSI}, \quad (27)$$

we generate an artificial series of block maxima using the following parametrization of the GEV model (2):

$$\mu(\widehat{U}_t) = +1 - 5u_1(t) + 2u_2(t) + 1u_3(t), \quad (28)$$

$$\sigma(\widehat{U}_t) = +2.1018 - 0.7132u_1(t) - 0.8203u_2(t) + 0.1356u_3(t), \quad (29)$$

$$\xi(\widehat{U}_t) = -0.0627 - 0.4051u_1(t) + 0.0022u_2(t) - 0.0026u_3(t). \quad (30)$$

By assigning a relatively high coefficient to the factor $u_1(t)$ in (28), we stress the linear-trend behavior in the dynamics of block maxima. The coefficients in (29)–(30) were generated randomly. We use MATLAB function `gevrnd` for sampling:

$$X_t \sim \text{GEV}(\mu(\widehat{U}_t), \sigma(\widehat{U}_t), \xi(\widehat{U}_t)) \quad \text{for } t = 1, \dots, 800. \quad (31)$$

In the next step, we split the covariates \widehat{U}_t into resolved and unresolved subsets $U_t = (u_2(t), u_3(t))$ and $U_t^{\text{un}} = u_1(t)$, respectively, and apply the FEM-BV-GEV and GEV-CDN methods for solving the inverse problem. For given X_t and U_t , we fit the model parameters to describe the distribution of X_t . We want to emphasize that by deliberately missing the most relevant covariate, the linear trend, we would expect both methods to react to this issue by exploiting the intrinsic nonlinearity in the case of GEV-CDN and the nonstationarity in the case of FEM-BV-GEV.

FEM-BV-GEV is supplied with $K_{\text{list}} = \{1, 2, 3\}$, $C_{\text{list}} = \{2 : 1 : 6\}$ and the following configurations: the number of annealing steps is set to 100, the maximal number of the subspace iterations to 150 and the minimization threshold to $\text{Tol} = 5.0 \times 10^{-5}$. The GEV-CDN approach is configured with $N_H = \{1, 2 : 2 : 18\}$, the hidden transfer function is the logistic function and the number of trials is 100. The results are summarized in Table 1, featuring the minimal AIC_c values achieved by the respective

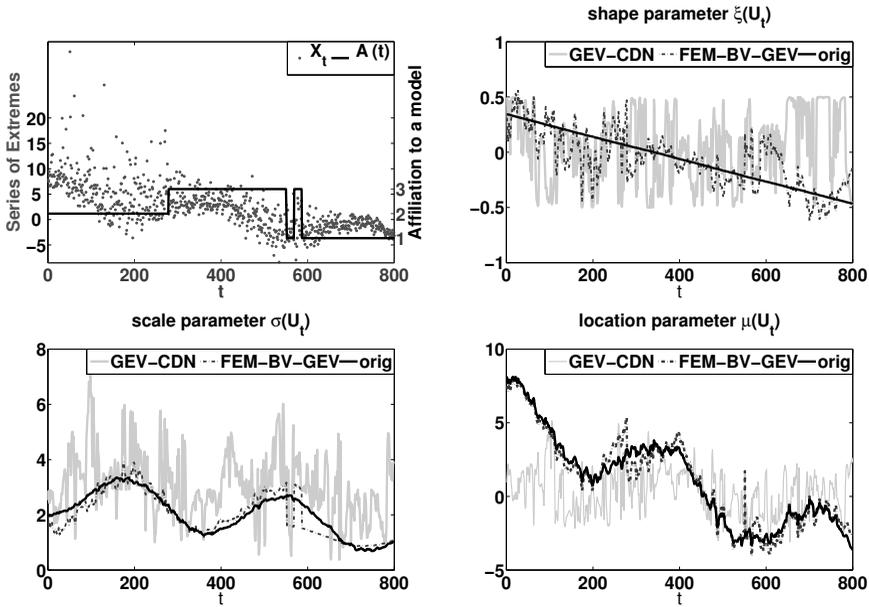


Figure 1. Stationary test case: this figure shows the results for the application of FEM-BV-GEV and GEV-CDN to (31). The upper left figure shows the artificially generated series of extremes X_t versus the optimal switching process $\Gamma^*(t)$, expressed by the affiliation vector $A(t)$. The remaining panels represent the evaluation of the shape, scale and location parameters according to the original (black solid line), optimal FEM-BV-GEV (dashed line) and GEV-CDN (gray solid line) parameters.

methods. Resulting optimal models are $K = 3$ and $C = 4$ for FEM-BV-GEV and $N_H = 12$ for CDN-GEV. The regression analysis of X_t based on resolved covariates was performed better by the FEM-BV-GEV than by the GEV-CDN approach (with a smaller NLL and a lower total number of model parameters). As seen from Figure 1, top left, the optimal switching process $\Gamma^*(t)$, expressed by the affiliation $A(t) \in \mathbb{R}$ (with $A(t) = \{i : i = \operatorname{argmax} \gamma_i^*(t) \text{ over } i = 1, \dots, K\}$), assigns X_t to three different models. Thereby, it explicitly resolves the implicit linear trend in the systematically missing covariate U^{un} via a switching process that subsequently goes through three local parameter regimes. We cannot compare the original and the resulting coefficients for the regression models explicitly. Instead, we evaluate the approximated $\mu^*(U_t)$, $\xi^*(U_t)$ and $\sigma^*(U_t)$ according to the FEM-BV-GEV and the GEV-CDN models and compare them with the original evaluations according to (28)–(30). The comparison is shown in Figure 1. The top right, bottom left and bottom right panels represent the shape, the scale and the location parameters, respectively. The parameters obtained from FEM-BV-GEV resolve the underlying trend very reliably. In contrast, due to the intrinsic assumption that the neuron weights and biases are constant, GEV-CDN is not able to recover the impact of this missing covariate.

	Optimal models for nonstationary test case			
	Settings	NLL	$ M $	AIC_c
FEM-BV-GEV	$K = 2, C = 12$	1.2041×10^3	37	2.4859×10^3
GEV-CDN	$N_H = 7$	1.2545×10^3	52	2.6203×10^3

Table 2. Optimal results for FEM-BV-GEV ($K = 2$ and $C = 12$) and GEV-CDN ($N_H = 7$) for the nonstationary test case. The exact negative log-likelihood for X_t using the original parameters is $NLL_{\text{exact}} = 1.2289 \times 10^3$.

5.2. Nonstationary test case. Now we consider a nonstationary test case and use it to verify the accuracy and the performance of the FEM-BV-GEV. We generate X_t according to a mixture model with a nonstationary switching process

$$X_t \sim \gamma_1(t) \text{GEV}_1 + \gamma_2(t) \text{GEV}_2, \tag{32}$$

where GEV_1 is parametrized according to (28)–(30) and GEV_2 according to

$$\mu_2(\widehat{U}_t) = -0.5 - 3u_1(t) + 0.5u_2(t) + 0.5u_3(t), \tag{33}$$

$$\sigma_2(\widehat{U}_t) = +0.6729 + 0.0183u_1(t) - 0.4131u_2(t) + 0.1378u_3(t), \tag{34}$$

$$\xi_2(\widehat{U}_t) = -0.0780 - 0.1398u_1(t) - 0.1608u_2(t) + 0.0266u_3(t). \tag{35}$$

Here we consider the same covariates \widehat{U}_t as in the stationary case. The nonstationary switching process $\Gamma(t) = (\gamma_1(t), \gamma_2(t))$ is generated artificially with $C = 6$ switches. Now for given X_t and $U_t = (u_1(t), u_2(t), u_3(t))$, we apply FEM-BV-GEV and the GEV-CDN approach to capture the nonstationarity of (32). FEM-BV-GEV is supplied with $K_{\text{list}} = \{1, 2, 3\}$ and $C_{\text{list}} = \{2 : 1 : 14\}$; remaining configurations are the same as for the stationary test case. Also the configurations of the GEV-CDN approach do not change. Because we provide the full information, $U_t = \widehat{U}_t$, to both methods, they both perform well; compare Table 2 and Figure 2. FEM-BV-GEV approximates the dynamics of X_t with less parameters and a smaller NLL. The inconsistency of the number of switches in $\Gamma^*(t)$ with $C = 12$ (Figure 2 upper left panel) and the original $\Gamma(t)$ with $C = 6$ can be neglected due to the relatively large confidence intervals for $\Gamma^*(t)$ and Θ^* (compare Appendix B, Figure 4 and Table 5).

Also GEV-CDN captures the underlying trend in parameters; compare Figure 2. The computational performance of FEM-BV-GEV and GEV-CDN is compared by considering the CPU time for one annealing step dependent on the increasing number of parameters (configurations do not change). The results are shown in Figure 3. The plots contain the average CPU time over 100 runs. FEM-BV-GEV obviously outperforms the GEV-CDN approach with respect to the computational performance for the growing number of parameters (e.g., corresponding to the larger number of involved covariates or hidden neurons).

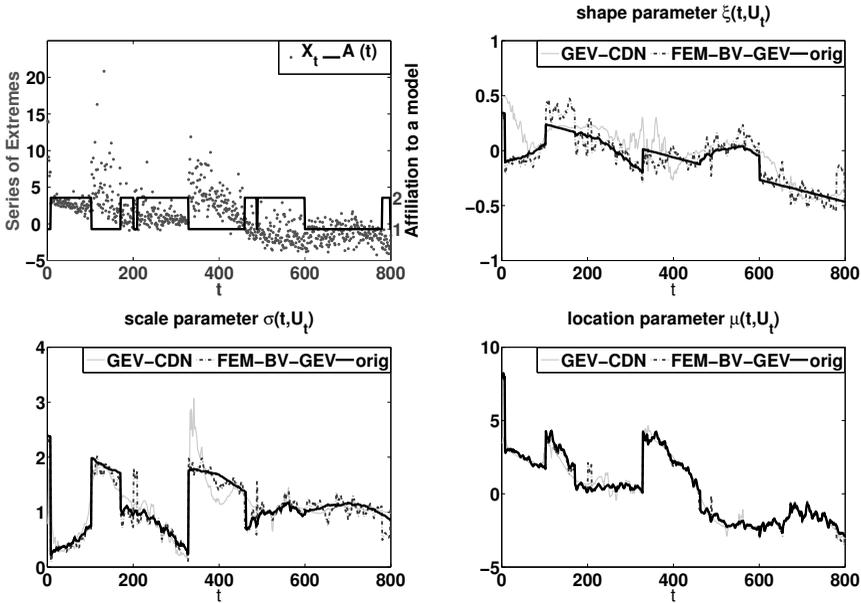


Figure 2. Nonstationary test case: this figure shows the results for the application of FEM-BV-GEV and GEV-CDN to (32). The upper left figure shows the artificial generated series of extremes X_t versus the optimal switching process $\Gamma^*(t)$, expressed by the affiliation vector $A(t)$. The remaining panels represent the evaluation of the shape, scale and location parameters according to original (black solid line), optimal FEM-BV-GEV (dashed line) and GEV-CDN (gray solid line) parameters.

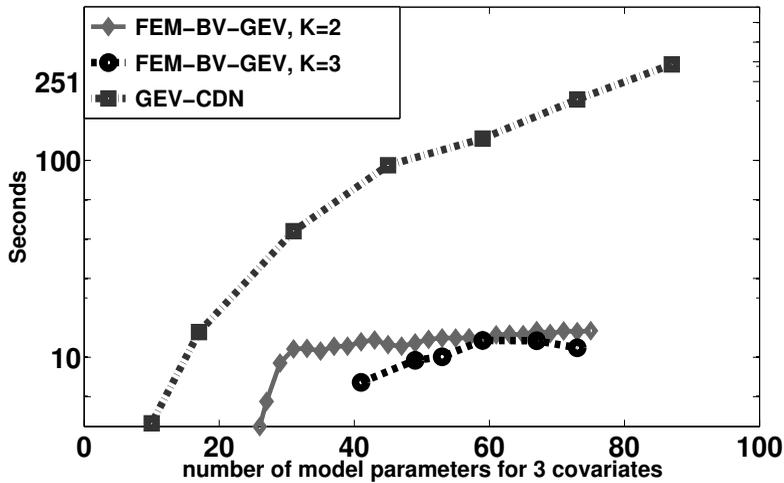


Figure 3. Nonstationary test case: this figure compares the computational time performance of FEM-BV-GEV (diamonds marker for $K = 2$ and circles for $K = 3$) and GEV-CDN (square markers) using a logarithmic time scale (seconds). The number of covariates is fixed; thus, the increase of the number of model parameters is due to increasing of C for FEM-BV-GEV and the number of hidden neurons for GEV-CDN.

5.3. Real-data application. In this section, we apply FEM-BV-GEV and GEV-CDN to real data, where we do not a priori have the knowledge about the underlying dynamics and we have to deal with unresolved modes. In the following, we consider historical daily records of temperature from January 1, 1950 until January 1, 2011 for Lugano, Switzerland (46° N, 8.9667° E) and Berlin, Germany (52.4649° N, 13.3017° E)² [32]. We restrict the data to this period because observations for some of the involved factors are only available starting from 1950. Before extracting 30-day block maxima, we deseasonalize the data. The dedicated series of block maxima for each location contains 742 maxima in the observed period. For the regression analysis, we consider the following set of covariates:

- (1) arctic oscillation (AO),²
- (2) North Atlantic oscillation (NAO),³
- (3) total solar irradiance (TSI), averaged over one day [20; 21],⁴
- (4) ENSO, represented through mean sea surface temperature anomalies in the Nino3.4 region [46],
- (5) $\log(\text{CO}_2)$, with logarithmic dependence according to [43],
- (6) seasonal periodical phase $\text{Per}_I = \sin\left(\frac{1}{365}2\pi t\right)$,
- (7) seasonal periodical phase $\text{Per}_{II} = \sin\left(\frac{3}{2.1}\pi + \frac{1}{365}2\pi t\right)$, and
- (8) Madden–Julian oscillation (MJO) containing the first two empirical orthogonal functions.⁵

The covariates $U_t \in \mathbb{R}^8$ are scaled, with $u_s(t) \in [-1, 1]$ for $s = 1, \dots, 8$, so we can interpret their relative influences on trends in model parameters. For the following GEV regression analysis, we consider the covariates at the same time steps when the maxima in each block are observed. First of all, we want to extract the most significant u_{comb} out of all possible, in total 255, combinations of U_t . For this task, we use the FEM-BV-GEV framework with the following configurations: $K_{\text{list}} = \{1, 2, 3\}$, $C_{\text{list}} = \{5 : 5 : 100\}$, the number of annealing steps is fixed to 100, the number of the subspace iterations is set to 250 and the minimization threshold is set to $\text{Tol} = 5.0 \times 10^{-5}$. Then, according to the minimal AIC_c , we obtain for each location the optimal model including the most significant combination, denoted by u_{comb}^* . For Lugano, u_{comb}^* is $[\text{NAO}, \log(\text{CO}_2), \text{Per}_I, \text{Per}_{II}]$, and for Berlin, $u_{\text{comb}}^* = [\text{AO}, \text{NAO}, \text{Per}_I]$. In the second step, we compare FEM-BV-GEV and GEV-CDN applied to two different settings: (a) we provide the complete set

²Data were retrieved from the NOAA's National Climatic Data Center web page.

³Data were retrieved from <ftp://ftp.cpc.ncep.noaa.gov/cwlinks/>.

⁴Data were retrieved from <http://pmodwrc.ch/pmod.php?topic=tsi/composite/SolarConstant>.

⁵Data were retrieved from <http://cawcr.gov.au/staff/mwheeler/maproom/RMM/>.

	Location Lugano with $u_{\text{comb}}^* = [\text{NAO}, \log(\text{CO}_2), \text{Per}_I, \text{Per}_{II}]$			Location Lugano with $\hat{u}_{\text{comb}}^* = [\text{NAO}, \text{Per}_I, \text{Per}_{II}]$		
	NLL	$ M $	AIC_c	NLL	$ M $	AIC_c
FEM-BV-GEV	1.5739×10^3	70	3.3026×10^3	1.6089×10^3	64	3.3580×10^3
GEV-CDN	1.4940×10^3	115	3.2606×10^3	1.6729×10^3	45	3.4416×10^9

Table 3. Comparison of FEM-BV-GEV and GEV-CDN according to AIC_c model selection criteria for Lugano according to the resolved and unresolved modes. The optimal models for resolved modes are FEM-BV-GEV with $K = 2$ and $C = 40$ and GEV-CDN with $N_H = 14$. The optimal models for unresolved modes are FEM-BV-GEV with $K = 2$ and $C = 40$ and GEV-CDN with $N_H = 6$.

	Location Berlin with $u_{\text{comb}}^* = [\text{AO}, \text{NAO}, \text{Per}_I]$			Location Berlin with $\hat{u}_{\text{comb}}^* = [\text{NAO}, \text{Per}_I]$		
	NLL	$ M $	AIC_c	NLL	$ M $	AIC_c
FEM-BV-GEV	1.6428×10^3	109	3.5415×10^3	1.6756×10^3	89	3.5538×10^3
GEV-CDN	1.7818×10^3	45	3.6595×10^3	1.7927×10^3	39	3.6678×10^3

Table 4. Comparison of FEM-BV-GEV and GEV-CDN according to AIC_c model selection criteria for Berlin according to the resolved and unresolved modes. The optimal models for resolved modes are FEM-BV-GEV with $K = 2$ and $C = 85$ and GEV-CDN with $N_H = 6$. The optimal models for unresolved modes are FEM-BV-GEV with $K = 2$ and $C = 70$ and GEV-CDN with $N_H = 6$.

of optimal covariates for the regression analysis u_{comb}^* , and (b) we provide an incomplete set $\hat{u}_{\text{comb}}^* = [\text{NAO}, \text{Per}_I, \text{Per}_{II}]$ and keep back $\log(\text{CO}_2)$ for Lugano and provide $\hat{u}_{\text{comb}}^* = [\text{NAO}, \text{Per}_I]$ and keep back AO for Berlin. Note that u_{comb}^* is significant according to the FEM-BV-GEV approach and one could argue that for the GEV-CDN approach another set of covariates could be more important.⁶ In return, in real application, we will never know a priori which covariates may be important, and in any case, we do not dispose of complete system measurements. Moreover, the influence of u_{comb}^* on the dynamics of block maxima can be interpreted physically; refer to Appendix C. The results for settings (a) and (b) are shown in Table 3 for Lugano and in Table 4 for Berlin. Thereby, the optimal GEV-CDN model is chosen from $N_H = \{2 : 2 : 16\}$. Additionally, the more interested reader can find a short postinference according to the optimal models in Appendix C: we compute the expectation value of block maxima with the corresponding quantiles for both locations and discuss its behavior.

⁶Application of GEV-CDN to identify the most significant combination of covariates is not feasible because of prohibitively high computational cost to get through all 255 covariates combinations (see Figure 3 for computational-cost comparisons of the two methods).

Comparing the optimal FEM-BV-GEV and GEV-CDN models, we can conclude that in the case when the set of covariates is “complete” the nonlinear GEV-CDN provides a better description of the block maxima for Lugano in terms of information theory (as measured by AIC_c), suggesting that the underlying dynamics is nonlinear rather than nonstationary. In contrast, FEM-BV-GEV provides a better description of block maxima for Berlin. Moreover, in case some information is “missing”, the nonstationary FEM-BV-GEV approach approximates the underlying dynamics better by reflecting the unresolved modes through the switching process for both considered cases (Berlin and Lugano).

6. Conclusion and outlook

In this work, we presented an extension of the GEV methodology for statistical regression analysis of block maxima with systematically missing covariates. We applied the strategy of stochastic covariate reduction and expressed the GEV parameters as fully nonstationary regression models based on resolved covariates only. The involved nonstationarity is interpolated by $K \geq 1$ local models and a nonstationary hidden switching process. The corresponding inverse problem was regularized using the nonparametric FEM-BV methodology by assuming persistence of the switching process (the number of switches between the local models is limited by a parameter C). The well-posed inverse problem is solved by deploying the gradient-free optimization methodology based on the Metropolis algorithm. The selection of optimal K and C and the significant subset of resolved covariates is carried out using the AIC_c information criteria. The proposed FEM-BV-GEV framework allows a computationally efficient, semiparametric and nonstationary analysis and goes beyond strong a priori probabilistic and deterministic assumptions typical for standard approaches deploying, e.g., mixture models, hidden Markov models, spline interpolation or neural networks. FEM-BV-GEV includes methods based on parametric regression, hidden Markov models and local kernel smoothing as special cases. However, the linear regression, which is involved in FEM-BV-GEV and provides an easily interpretable and understandable statistical model, becomes a weakness as soon as the influence of covariates is strongly nonlinear. For that reason, we considered the GEV-CDN approach as a main competitor. GEV-CDN is based on GEV and neural networks: each model parameter is described as a nonlinear function of covariates with constant coefficients exploiting the conditional density network.

We compared the two methods on test cases and real data according to the four criteria: (1) information content of the models, (2) ability to handle unresolved covariates, (3) computational complexity and (4) interpretability of the models. The results in Section 5.1 show that if some relevant information is missing then the nonstationary FEM-BV-GEV approach approximates the underlying dynamics

better by reflecting the unresolved modes through the switching process. In contrast, the GEV-CDN approach seems to average out the underlying trends and in these situations performs worse. The second example (compare Section 5.2) investigates the performance of the two methods applied to data, which is generated according to a switching process and different models. The FEM-BV-GEV approach performs better. GEV-CDN seems to capture the general dynamics but requires more model parameters to describe the underlying switching process (implicitly resolving the nonstationary switching process through the nonlinear stationary function). The third example (compare Section 5.3) demonstrates the performance of FEM-BV-GEV and GEV-CDN on real data analyzing 30-day block maxima surface temperatures for locations Lugano, Switzerland and Berlin, Germany. The FEM-BV-GEV approach allows a better description of block maxima for Berlin. Thereby, FEM-BV-GEV indicates two different models ($K = 2$) pointing to systematically missing covariates in the statistical regression analysis. GEV-CDN performs better applied to block maxima for Lugano. The better performance of the nonlinear GEV-CDN approach might lead to the suggestion that the dynamics of the block maxima at Lugano can be better explained by the stationary nonlinear regression than by the nonstationary linear one. Additionally, FEM-BV-GEV outperforms GEV-CDN in the interpretability and understandability of the models and in the far more favorable computational complexity and scalability. Consequently, we conclude that the FEM-BV-GEV approach should be used in cases where not all potentially significant covariates can be resolved explicitly and the observed data (series of extremes and the number of resolved covariates) is big; correspondingly, GEV-CDN should be applied in cases where the complete information is known and the provided statistics are rather small.

Outlook. A point of interest in data-based analysis of block maxima in the context of the FEM-BV-GEV approach is to understand the dynamics of the switching process, implicitly reflecting the dynamics of the most relevant covariates that are systematically missing in the analyzed data. One can either try to find a set of covariates to resolve the observed dynamics or parametrize the switching process. The latter can be done by considering the switching process as a discrete time series and study the dynamics with time-series analysis methods, e.g., a FEM-BV-Markov method [27]. Another extension of the presented FEM-BV-GEV methodology goes toward space-time modeling of block maxima. The spatial extension of FEM-BV-GEV requires appropriate regularization in space, e.g., based on distances between the locations where the measurements are taken. Besides that, by replacing GEV by the generalized Pareto distribution (GPD) methodology, we can straightforwardly derive the FEM-BV-GPD framework to study threshold exceedances. All these issues are matters of future research.

```

input :  $X_t$  series of extremes,  $u_{\text{comb}}$ ,  $\Gamma(t)$ ,  $\mathcal{L}(\Gamma, \Theta_{\text{old}})$ 
output :  $\Theta_{\text{opt}}$ 
1  $\Theta_{\text{new}} = \text{generateInitialValue}(\Gamma(t), X_t, u_{\text{comb}})$ 
2 if  $\mathcal{L}(\Gamma(t), \Theta_{\text{new}}) < \mathcal{L}(\Gamma(t), \Theta_{\text{old}})$  then
3      $\Theta_{\text{opt}} = \Theta_{\text{new}}$ 
4     return  $\Theta_{\text{opt}}$ 
5 initialize  $\delta, \beta, \Sigma$ , counterAccept = 0
6 for sampleStep = 1 : sampleSizeMCMC do
7      $\Theta_{\text{next}} = \text{proposeNext}(\Theta_{\text{new}}, \Gamma(t), X_t, u_{\text{comb}}, \Sigma, \text{noise}, \beta)$ 
8     if  $\mathcal{L}(\Gamma(t), \Theta_{\text{next}}) < \mathcal{L}(\Gamma(t), \Theta_{\text{old}})$  then
9          $\Theta_{\text{opt}} = \Theta_{\text{next}}$ 
10        return  $\Theta_{\text{opt}}$ 
11    else if  $\text{checkAcceptance}(\beta, \Theta_{\text{next}}, \Theta_{\text{new}})$  then
12         $\Theta_{\text{new}} = \Theta_{\text{next}}$ 
13        counterAccept += 1
14         $\text{updateCovMatrix}(\Theta_{\text{new}}, \Sigma)$ 
15    if sampleStep  $\geq 50$  then
16         $[\delta, \beta] = \text{adaptStep}(\delta, \beta, \text{counterAccept}, \text{sampleStep})$ 
17  $\Theta_{\text{opt}} = \Theta_{\text{old}}$ 
    
```

Algorithm 3: MCMC-based optimization algorithm for fixed $\Gamma(t)$.

Appendix A: Details of the adaptive MCMC algorithm

In the following, we point out the main steps of the deployed MCMC-based optimization; see Algorithm 3. The algorithm is based on [18] and differs mainly in two steps: lines 3 and 6 (explained in more details in the next two sections). Please note that the convergence condition for this algorithm is fulfilled if MCMC proposes a new parameter set that provides a smaller $\mathcal{L}(\Gamma, \Theta_{\text{new}})$ value for fixed $\Gamma(t)$. In practical applications, in the beginning of the FEM-BV subspace-minimization procedure, Algorithm 3 proposes a better parameter set already after few steps. However, the number of samplings is limited by the parameter sampleSizeMCMC; see Algorithm 3, line 5. In particular, it is recommended to limit the number of samples because as soon as we get into the area of the local optima it becomes hard to propose a better parameter set. And if the algorithm fails, meaning it does not provide a better set of parameters, it returns with $\Theta_{\text{opt}} = \Theta_{\text{old}}$; see Algorithm 3, line 16. For the applications presented in this work, this parameter was assigned to sampleSizeMCMC = 1000.

A.1. Generate an initial value. MCMC sampling starts with generating an initial value (we refer to Algorithm 3, line 3). The scale and the shape parameters have to

fulfill the constraints

$$0 < \sigma_i(U_t) = \sigma_i^{(0)} + \sum_{s=1}^S \sigma_i^{(s)} u_s(t) < \text{const}, \quad i = 1, \dots, K, \quad \forall t, \quad (1)$$

$$-0.5 < \xi_i(U_t) = \xi_i^{(0)} + \sum_{s=1}^S \xi_i^{(s)} u_s(t) < 0.5, \quad i = 1, \dots, K, \quad \forall t \quad (2)$$

(constraint (2) ensures a regular likelihood estimator [14]). Applying a simple uniform distribution would not necessarily provide an appropriate initial value. To hold the constraints, we reformulate them: since $\xi_i(U_t)$ and $\sigma_i(U_t)$ attain their unique maximum/minimum values in one of the corners of the convex hull defined by U_t , $t = 1, \dots, T$ [27], it is sufficient to fulfill the constraints (1)–(2) on all corners of the convex hull of U_t . Using a matrix $A \in \mathbb{R}^{(S+1) \times 2^S}$ that contains all combinations of maximal/minimal values of $U(t)$, $t = 1, \dots, T$, we can reformulate the constraint for $\xi_i = (\xi_i^{(0)}, \dots, \xi_i^{(S)})$:

$$\begin{aligned} -A\xi_i &< -lb_\xi, & lb_\xi &= -0.5 \cdot \mathbf{1} \in \mathbb{R}^{2^S}, \\ A\xi_i &< +ub_\xi, & ub_\xi &= 0.5 \cdot \mathbf{1} \in \mathbb{R}^{2^S}. \end{aligned}$$

The same applies for σ . Finally, if we slightly strengthen the constraints

$$\sigma_i(U_t) \in [\epsilon, \text{const}] \quad \text{and} \quad \xi_i(U_t) \in [-0.5 + \epsilon, 0.5 - \epsilon]$$

with $\epsilon > 0$ small and $\text{const} \in \mathbb{R}$ some high value, we can use some convex sampler to get random, uniformly distributed values within this convex hull. The same approach can be applied to sample $\mu_i = (\mu_i^{(0)}, \dots, \mu_i^{(S)})$ in a way such that the constraint (16) in Section 2 is fulfilled. Another way is to estimate the initial value for μ_i by applying ordinary least squares [14]. Note that this estimation is not considered as the trend estimate for the GEV distribution but as a procedure to generate an initial value that is adjusted within the MCMC and the subspace (Algorithm 2 in Section 4) procedure. Both possibilities are implemented in the FEM-BV-GEV framework. In this paper, the second one was deployed.

A.2. Propose next. The performance of the Metropolis algorithm can be improved with an appropriate proposal distribution [4; 34]. However, it is not obvious which proposal density should be chosen for the current target density. In this work, we refer to the discussions in [4] and deploy the adaptive Metropolis algorithm where the next proposal, denoted here by Y_{n+1} , is sampled according to a mixture distribution with respect to the information of all previous accepted samples, denoted here by X_0, \dots, X_n :

$$Y_{n+1} \sim (1 - \delta)\mathcal{N}\left(X_n, \frac{2.38^2}{d}\Sigma_n\right) + \delta\mathcal{N}(X_n, \Sigma_0), \quad (3)$$

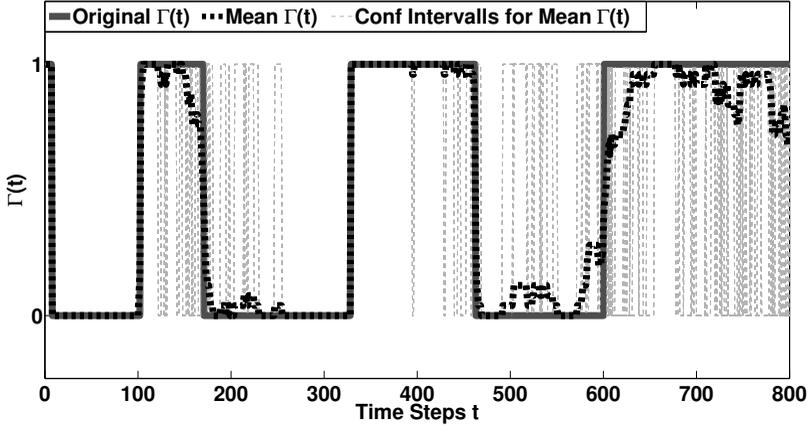


Figure 4. Nonstationary test case: the figure compares the averaged optimal $\hat{\Gamma}^*(t)$ (black dashed line) with its confidence intervals (gray dashed lines) versus the original switching process (gray solid line).

where d is the dimension of X_n and $\Sigma_n \in \mathbb{R}^{d \times d}$ corresponds to the empirical covariance matrix of X_0, \dots, X_n . The parameter $0 < \delta < 1$ controls the acceptance rate of the Metropolis algorithm; the acceptance rate is increasing for $\delta \rightarrow 1$ and decreasing for $\delta \rightarrow 0$. For more details on the adaption step, see [18].

Appendix B: Confidence intervals

In the following, we refer to the nonstationary test case in Section 5 and provide the confidence intervals for the optimal estimation of $(\Gamma^*(t), \Theta^*)$ to verify the accuracy of the proposed FEM-BV-GEV approach. We obtain the confidence intervals via bootstrapping procedure [12]; i.e., we resample $X_t, t = 1, \dots, T$, according to (32) in Section 5 N times and apply FEM-BV-GEV (with $K_{\text{list}} = \{1, 2, 3\}$ and $C_{\text{list}} = \{2 : 1 : 14\}$) each time. Then each optimal result $(\Gamma^*(t), \Theta^*)$ is stored, and we can estimate the averaged parameters as well as the confidence intervals. For this example, we consider $N = 150$. Figure 4 shows the original $\Gamma(t)$ (gray solid line), the averaged optimal $\hat{\Gamma}^*(t)$ (black dashed line) and all other realizations of $\Gamma^*(t)$ that can be considered as the confidence intervals (gray dashed lines). Table 5 contains the corresponding confidence intervals for the averaged optimal model parameters $\hat{\theta}_i^*$ for $i = 1, 2$.

Appendix C: Postinference

In this section, we discuss the postinference for Lugano and Berlin according to the optimal FEM-BV-GEV and GEV-CDN models. The locally linear FEM-BV-GEV model allows direct interpretation of the influence of covariates on the dynamics of GEV parameters; see Table 6 for Lugano and Table 7 for Berlin. For

	Model Parameters							
	μ_0	μ_1	μ_2	μ_3	σ_0	σ_1	σ_2	σ_3
					ξ_0	ξ_1	ξ_2	ξ_3
θ_1	1.0000	-5.0000	2.0000	-1.0000	1.4983	-0.6987	0.1937	0.0353
$\hat{\theta}_1^*$	1.0502	-4.9507	1.9786	-1.0385	1.4821	-0.6654	0.1888	0.0297
std±	0.1478	0.1690	0.2492	0.3418	0.1085	0.1652	0.0979	0.2109
θ_2	-0.5000	-3.0000	0.5000	0.5000	0.6729	0.0183	-0.4131	0.1378
$\hat{\theta}_2^*$	-0.4938	-2.9626	0.5342	0.5300	0.7187	0.1113	-0.3720	0.1944
std±	0.0715	0.1427	0.0897	0.1066	0.0573	0.1108	0.0645	0.0740
					0.0500	0.0550	0.0446	0.1051
					-0.0627	-0.4051	0.0022	-0.0026
					-0.0719	-0.4490	0.0421	-0.0355
					-0.0780	-0.1398	-0.1608	0.0266
					-0.0852	-0.1472	-0.1542	0.0122
					0.0684	0.1219	0.1175	0.1537

Table 5. The original parameters θ_1 and θ_2 according to (33)–(35) and (28)–(30) in Section 5, averaged optimal parameters $\hat{\theta}_1^*$ and $\hat{\theta}_2^*$ and the corresponding standard deviations (std±).

	Model Parameters for Lugano with $u_{\text{comb}}^* = [\text{NAO}, \log(\text{CO}_2), \text{PerI}, \text{PerII}]$									
	μ_0	μ_1	μ_2	μ_3	μ_4	σ_0	σ_1	σ_2	σ_3	σ_4
						ξ_0	ξ_1	ξ_2	ξ_3	ξ_4
θ_1^*	3.92	0.78	-2.12	1.70	-0.34	1.71	0.60	-0.17	0.39	-0.42
						-0.05	0.16	0.03	-0.19	-0.09
θ_2^*	4.29	-0.19	1.97	0.74	-1.39	1.99	-0.10	0.05	0.21	-0.61
						-0.37	0.39	0.40	-0.15	-0.09

Table 6. The table contains optimal parameters θ_1^* and θ_2^* (the values are rounded to two decimal places).

	Model Parameters for Berlin with $u_{\text{comb}}^* = [\text{AO}, \text{NAO}, \text{PerI}]$							
	μ_0	μ_1	μ_2	μ_3	σ_0	σ_1	σ_2	σ_3
					ξ_0	ξ_1	ξ_2	ξ_3
θ_1^*	4.73	1.89	0.1	0.38	2.59	-0.52	-0.03	0.6873
					-0.22	-0.27	-0.40	0.17
θ_2^*	8.43	2.00	-1.13	0.59	2.15	-0.21	0.69	-0.12
					-0.32	0.01	-0.10	0.00

Table 7. The table contains optimal parameters θ_1^* and θ_2^* (the values are rounded to two decimal places).

the neural-networks-based GEV-CDN approach, we obtain a matrix of weights and have to evaluate the parameters according to the transfer function (a logistic function in our particular case), making the fitted models not easy for interpretation and understanding. The identification of these factors is physically meaningful. Positive phase of AO causes dry and hot conditions in Mediterranean regions. AO has a direct influence on atmospheric circulation blocking events: it induces a ridge of high pressure in the mid-latitude jet streams that can cause persistently high temperatures (as well as cold conditions) [23]. Positive phases of NAO cause warm, wet winters in northern and dry winters in southern Europe. Due to the anthropogenic influence of CO₂ concentration, log(CO₂) holds a positive trend with oscillating dynamics (with maximum value in May and minimum in October) [31]. The relevance of Per_I and Per_{II} points to strong seasonal dependence of block maxima in both locations (this is obvious since we consider monthly maxima). In order to study the long-term trend in distribution of block maxima, we evaluate the nonstationary expectation value

$$\mathbb{E}_{K=2}[X_t, t] = \sum_{i=1}^2 \gamma_i(t) \left(\mu_i(U_t) + \sigma_i(U_t) \frac{\tilde{\Gamma}(1 - \xi_i(U_t)) - 1}{\xi_i(U_t)} \right), \tag{1}$$

$$\mathbb{E}_{\text{CDN}}[X_t, t] = \mu_{\text{CDN}}(U_t) + \sigma_{\text{CDN}}(U_t) \frac{\tilde{\Gamma}(1 - \xi_{\text{CDN}}(U_t)) - 1}{\xi_{\text{CDN}}(U_t)} \tag{2}$$

with $t = 1, \dots, 742$, where $K = 2$ corresponds to FEM-BV-GEV (with parametrization according to (13) in Section 2) and CDN to GEV-CDN models and $\tilde{\Gamma}$ denotes the gamma function. Figures 5 and 7 show the results according to FEM-BV-GEV and Figures 6 and 8 according to GEV-CDN. The 0.99- and 0.10-quantiles are the

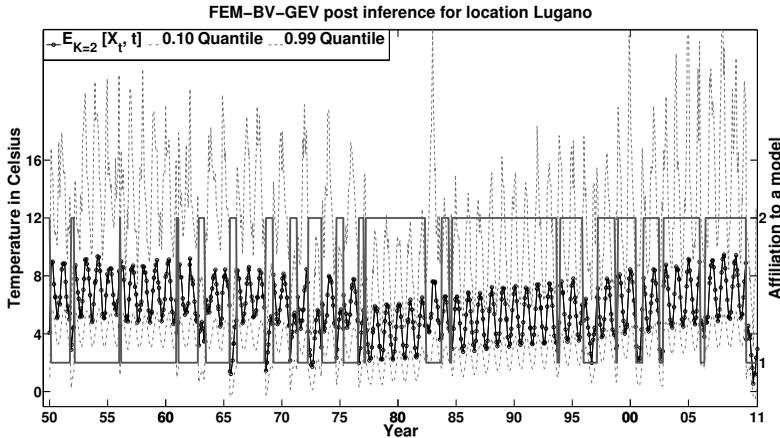


Figure 5. Location Lugano: the plot of the expectation value for the optimal FEM-BV-GEV model, $K = 2$ and $C = 40$.

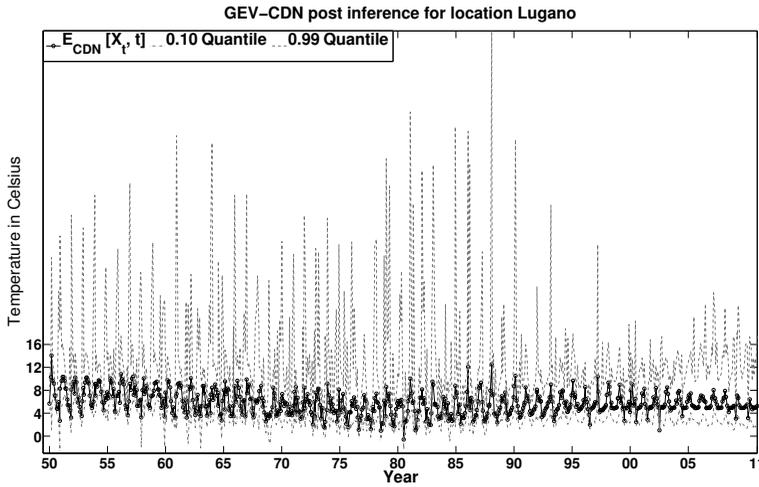


Figure 6. Location Lugano: the plot of the expectation value for the optimal GEV-CDN model with $N_H = 14$.

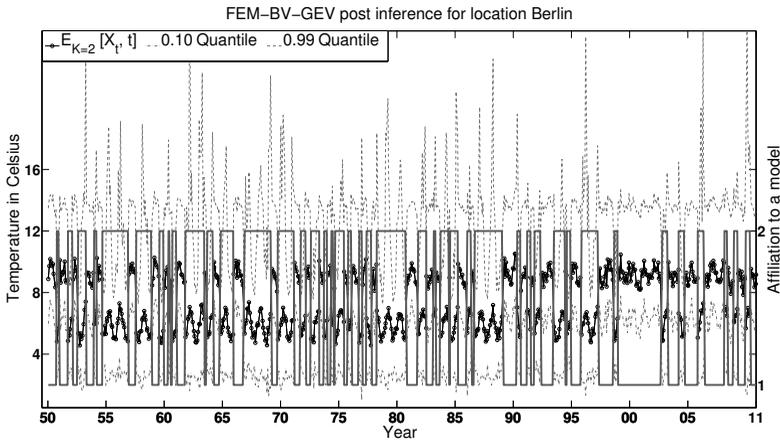


Figure 7. Location Berlin: the plot of the expectation value for the optimal FEM-BV-GEV model, $K = 2$ and $C = 85$.

confidence intervals containing 89% of the distribution. In particular, the 0.99-quantile corresponds to the 100-year return level. According to the FEM-BV-GEV results, the mean for Lugano shows a slightly negative trend in the first model, but after the 1980s, the second model dominates, where $\log(\text{CO}_2)$ has a positive influence and so the trend in block maxima becomes positive. In contrast, according to the GEV-CDN model, there is no obvious trend; however, the confidence intervals for the GEV distribution increase in the last ten years. For Berlin, the trend of the expectation value is separated according to two FEM-BV-GEV models, one model corresponds to higher block maxima. The GEV-CDN model averages these dynamics and provides a unchanging behavior with some outliers.

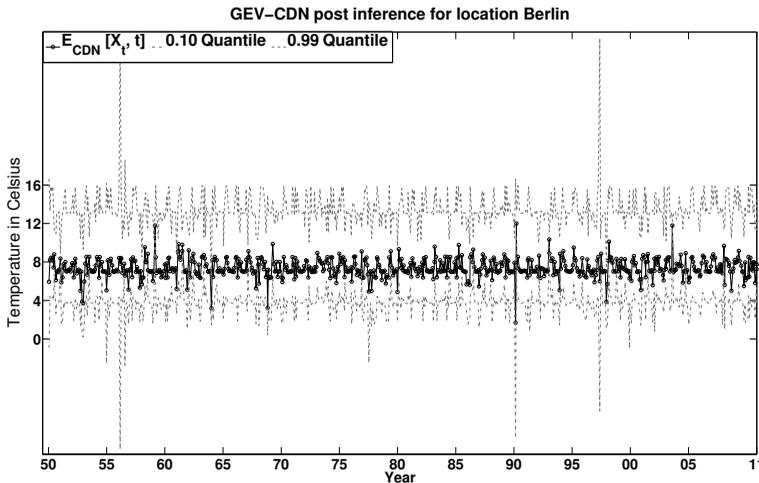


Figure 8. Location Berlin: the plots of the expectation value for the optimal GEV-CDN model, $N_H = 6$.

Acknowledgments

The authors would like to thank the staff and the organizers of the IPAM program “Model and Data Hierarchies for Simulating and Understanding Climate”, March 8–June 11, 2010. Important input regarding the understanding of standard methods to study the dynamics of extremes is due to other participants of the program (special thanks to Richard L. Smith for his talk that motivated the problem formulation considered in this manuscript; we would also like to thank E. Gerber for helpful hints regarding the $\log(\text{CO}_2)$ impact on temperature). We also thank Harald von Waldow and Olivia Romppainen-Martius for a helpful discussion.

References

- [1] B. Betrò, A. Bodini, and Q. A. Cossu, *Using a hidden Markov model to analyse extreme rainfall events in central-east Sardinia*, *Environmetrics* **19** (2008), no. 7, 702–713. MR 2530393
- [2] ———, *Regional-scale analysis of extreme rainfalls via HMM*, presented at the V International Workshop on Spatio-Temporal Modelling, 2010.
- [3] M.-O. Boldi and A. C. Davison, *A mixture model for multivariate extremes*, *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **69** (2007), no. 2, 217–229. MR 2325273 Zbl 1120.62030
- [4] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng (eds.), *Handbook of Markov chain Monte Carlo*, CRC Press, Boca Raton, FL, 2011. MR 2012g:62004 Zbl 1218.65001
- [5] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd ed., Springer, New York, 2002. MR 1919620 Zbl 1005.62007
- [6] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*, 2nd ed., Econometric Society Monographs, no. 53, Cambridge University Press, 2013. MR 3155491

- [7] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223. MR 2007f:94007 Zbl 1098.94009
- [8] A. J. Cannon, *A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology*, Hydrol. Process. **24** (2010), no. 24, 673–685.
- [9] ———, *GEVcdn: an R package for nonstationary extreme value analysis by generalized extreme value conditional density estimation network*, Comput. Geosci. **37** (2011), 1532–1533.
- [10] V. Chavez-Demoulin and A. C. Davison, *Generalized additive modelling of sample extremes*, J. R. Stat. Soc., Ser. C, Appl. Stat. **54** (2005), no. 1, 207–222. MR 2134607 Zbl 05188681
- [11] A. J. Chorin and O. H. Hald, *Stochastic tools in mathematics and science*, Surveys and Tutorials in the Applied Mathematical Sciences, no. 1, Springer, New York, 2006. MR 2006j:60001 Zbl 1086.60001
- [12] B. Clarke, E. Fokoué, and H. H. Zhang, *Principles and theory for data mining and machine learning*, Springer, Dordrecht, 2009. MR 2012i:62008 Zbl 1179.62004
- [13] S. G. Coles, *An introduction to statistical modeling of extreme values*, Springer, London, 2001. MR 2003h:62002 Zbl 0980.62043
- [14] S. G. Coles and M. J. Dixon, *Likelihood-based inference for extreme value models*, Extremes **2** (1999), no. 1, 5–23. Zbl 0938.62013
- [15] S. G. Coles and E. A. Powell, *Bayesian methods in extreme value modelling: a review and new developments*, Int. Stat. Rev. **64** (1996), no. 1, 119–136. Zbl 0853.62025
- [16] A. C. Davison and N. I. Ramesh, *Local likelihood smoothing of sample extremes*, J. R. Stat. Soc., Ser. B, Stat. Methodol. **62** (2000), no. 1, 191–208. MR 1747404 Zbl 0942.62058
- [17] L. de Haan and A. Ferreira, *Extreme value theory: an introduction*, Springer, New York, 2006. MR 2007g:62008 Zbl 1101.62002
- [18] J. de Wiljes, A. Majda, and I. Horenko, *An adaptive Markov chain Monte Carlo approach to time series clustering of processes with regime transition behavior*, Multiscale Model. Simul. **11** (2013), no. 2, 415–441. MR 3047436
- [19] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events for insurance and finance*, 8th ed., Applications of Mathematics, no. 33, Springer, Berlin, 1997. MR 98k:60080 Zbl 0873.62116
- [20] C. Fröhlich, *Observations of irradiance variations*, Space Sci. Rev. **94** (2000), no. 1–2, 15–24.
- [21] ———, *Solar irradiance variability since 1978: revision of the PMOD composite during solar cycle 21*, Space Sci. Rev. **125** (2006), no. 1–4, 53–65.
- [22] J. Hadamard, *Sur les problèmes aux dérivées partielles et leur signification physique*, Princeton University Bulletin **13** (1902), 49–52.
- [23] S. Häkkinen, P. B. Rhines, and D. L. Worthen, *Atmospheric blocking and Atlantic multidecadal ocean variability*, Science **334** (2011), no. 6056, 665–659.
- [24] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed., Springer, New York, 2009. MR 2012d:62081 Zbl 1273.62005
- [25] I. Horenko, *Finite element approach to clustering of multidimensional time series*, SIAM J. Sci. Comput. **32** (2010), no. 1, 62–83. MR 2011b:62009 Zbl 1206.62150
- [26] ———, *On identification of nonstationary factor models and their application to atmospheric data analysis*, J. Atmos. Sci. **67** (2010), no. 5, 1559–1574.
- [27] ———, *Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling*, J. Atmos. Sci. **68** (2011), no. 7, 1493–1506.

- [28] ———, *On analysis of nonstationary categorical data time series: dynamical dimension reduction, model selection, and applications to computational sociology*, *Multiscale Model. Simul.* **9** (2011), no. 4, 1700–1726. MR 2012j:60196 Zbl 1244.60070
- [29] I. Horenko and C. Schütte, *On metastable conformational analysis of nonequilibrium biomolecular time series*, *Multiscale Model. Simul.* **8** (2010), no. 2, 701–716. MR 2011a:62008 Zbl 05719784
- [30] C. M. Hurvich and C.-L. Tsai, *Regression and time series model selection in small samples*, *Biometrika* **76** (1989), no. 2, 297–307. MR 91c:62083 Zbl 0669.62085
- [31] R. P. Kane and E. R. de Paula, *Atmospheric CO₂ changes at Mahuna Loa, Hawaii*, *J. Atmos. Terr. Phys.* **58** (1996), no. 15, 1673–1681.
- [32] A. M. G. Klein Tank, J. B. Wijngaard, G. P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A. F. V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L. V. Alexander, and P. Petrovic, *Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment*, *Int. J. Climatol.* **22** (2002), no. 12, 1441–1453.
- [33] W. Kozek, *Optimally Karhunen–Loeve-like STFT expansion of nonstationary processes*, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, IEEE, Piscataway, NJ, 1993, pp. 428–431.
- [34] F. Liang, C. Liu, and R. J. Carroll, *Advanced Markov chain Monte Carlo methods: learning from past samples*, Wiley, Chichester, UK, 2010. MR 2012m:60172 Zbl 1209.62009
- [35] A. R. Lima, A. J. Cannon, and W. W. Hsieh, *Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy*, *Comput. Geosci.* **50** (2013), 136–144.
- [36] J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, *Math. Z.* **15** (1922), no. 1, 211–225. MR 1544569 JFM 48.0602.04
- [37] M. Loève, *Probability theory, II*, 4th ed., Graduate Texts in Mathematics, no. 46, Springer, New York, 1978. MR 0651018 Zbl 0385.60001
- [38] A. MacDonald, C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell, *A flexible extreme value mixture model*, *Comput. Statist. Data Anal.* **55** (2011), no. 6, 2137–2157. MR 2785120
- [39] A. J. Majda, R. V. Abramov, and M. J. Grote, *Information theory and stochastics for multiscale nonlinear systems*, CRM Monograph Series, no. 25, American Mathematical Society, Providence, RI, 2005. MR 2006k:76110 Zbl 1082.60002
- [40] E. Meerbach, E. Dittmer, I. Horenko, and C. Schütte, *Multiscale modelling in molecular dynamics: biomolecular conformations as metastable states*, *Computer simulations in condensed matter systems: from materials to chemical biology, I* (M. Ferrario, G. Ciccotti, and K. Binder, eds.), Lecture Notes in Physics, no. 703, Springer, Berlin, 2006, pp. 495–517.
- [41] P. Metzner, L. Putzig, and I. Horenko, *Analysis of persistent nonstationary time series and applications*, *Commun. Appl. Math. Comput. Sci.* **7** (2012), no. 2, 175–229. MR 3005737 Zbl 1275.62067
- [42] S. E. Neville, M. J. Palmer, and M. P. Wand, *Generalized extreme value additive model analysis via mean field variational Bayes*, *Aust. N. Z. J. Stat.* **53** (2011), no. 3, 305–330. MR 2897373
- [43] R. T. Pierrehumbert, *Energy balance models*, 2001 Program in Geophysical Fluid Dynamics (J.-L. Thiffeault, ed.), Woods Hole Oceanographic Institution, Woods Holes, MA, 2001, pp. 72–87.

- [44] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc., Ser. B, Stat. Methodol. **58** (1996), no. 1, 267–288. MR 96j:62134 Zbl 0850.62538
- [45] A. N. Tikhonov, *On the solution of ill-posed problems and the method of regularization*, Dokl. Akad. Nauk SSSR **151** (1963), 501–504, In Russian; translated in *Sov. Math., Dokl.* **4** (1963), 1035–1038. MR 28 #5576 Zbl 03227378
- [46] K. E. Trenberth, *The definition of El Niño*, B. Am. Meteorol. Soc. **78** (1997), 2771–2777.
- [47] V. N. Vapnik, *The nature of statistical learning theory*, Springer, New York, 1995. MR 98a:68159 Zbl 0833.62008
- [48] G. Wahba, *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, no. 59, SIAM, Philadelphia, PA, 1990. MR 91g:62028 Zbl 0813.62001

Received May 20, 2013. Revised November 28, 2013.

OLGA KAISER: olga.kaiser@usi.ch

Institute of Computational Science, Università della Svizzera italiana, Via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland

ILLIA HORENKO: illia.horenko@usi.ch

Institute of Computational Science, Università della Svizzera italiana, Via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland

Communications in Applied Mathematics and Computational Science

msp.org/camcos

EDITORS

MANAGING EDITOR

John B. Bell
Lawrence Berkeley National Laboratory, USA
jbbell@lbl.gov

BOARD OF EDITORS

Marsha Berger	New York University berger@cs.nyu.edu	Ahmed Ghoniem	Massachusetts Inst. of Technology, USA ghoniem@mit.edu
Alexandre Chorin	University of California, Berkeley, USA chorin@math.berkeley.edu	Raz Kupferman	The Hebrew University, Israel raz@math.huji.ac.il
Phil Colella	Lawrence Berkeley Nat. Lab., USA pcolella@lbl.gov	Randall J. LeVeque	University of Washington, USA rjl@amath.washington.edu
Peter Constantin	University of Chicago, USA const@cs.uchicago.edu	Mitchell Luskin	University of Minnesota, USA luskin@umn.edu
Maksymilian Dryja	Warsaw University, Poland maksymilian.dryja@acn.waw.pl	Yvon Maday	Université Pierre et Marie Curie, France maday@ann.jussieu.fr
M. Gregory Forest	University of North Carolina, USA forest@amath.unc.edu	James Sethian	University of California, Berkeley, USA sethian@math.berkeley.edu
Leslie Greengard	New York University, USA greengard@cims.nyu.edu	Juan Luis Vázquez	Universidad Autónoma de Madrid, Spain juanluis.vazquez@uam.es
Rupert Klein	Freie Universität Berlin, Germany rupert.klein@pik-potsdam.de	Alfio Quarteroni	Ecole Polytech. Féd. Lausanne, Switzerland alfio.quarteroni@epfl.ch
Nigel Goldenfeld	University of Illinois, USA nigel@uiuc.edu	Eitan Tadmor	University of Maryland, USA etadmor@cscamm.umd.edu
		Denis Talay	INRIA, France denis.talay@inria.fr

PRODUCTION

production@msp.org

Silvio Levy, Scientific Editor

See inside back cover or msp.org/camcos for submission instructions.

The subscription price for 2014 is US \$75/year for the electronic version, and \$105/year (+\$15, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Communications in Applied Mathematics and Computational Science (ISSN 2157-5452 electronic, 1559-3940 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

CAMCoS peer review and production are managed by EditFLOW[®] from Mathematical Sciences Publishers.

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2014 Mathematical Sciences Publishers

Communications in Applied Mathematics and Computational Science

vol. 9

no. 1

2014

-
- | | |
|---------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Discrete nonhomogeneous and nonstationary logistic and Markov regression models for spatiotemporal data with unresolved external influences | 1 |
| JANA DE WILJES, LARS PUTZIG and ILLIA HORENKO | |
| Low Mach number fluctuating hydrodynamics of diffusively mixing fluids | 47 |
| ALEKSANDAR DONEV, ANDY NONAKA, YIFEI SUN, THOMAS G. FAI,
ALEJANDRO L. GARCIA AND JOHN B. BELL | |
| High-order methods for computing distances to implicitly defined surfaces | 107 |
| ROBERT I. SAYE | |
| On inference of statistical regression models for extreme events based on incomplete observation data | 143 |
| OLGA KAISER AND ILLIA HORENKO | |



1559-3940(2014)9:1;1-0