CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks

Younes Boulaguiem[1,]* , Jakob Zscheischler[2,3,4] , Edoardo Vignotto[1], Karin van der Wiel[5] and Sebastian Engelke[1]

[1]Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland
[2]Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
[3]Climate and Environmental Physics, University of Bern, Bern, Switzerland
[4]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[5]R&D weather and climate modelling, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
*Corresponding author. E-mail: younes.boulaguiem@unige.ch

Abstract

Modeling dependencies between climate extremes is important for climate risk assessment, for instance when allocating emergency management funds. In statistics, multivariate extreme value theory is often used to model spatial extremes. However, most commonly used approaches require strong assumptions and are either too simplistic or over-parameterized. From a machine learning perspective, generative adversarial networks (GANs) are a powerful tool to model dependencies in high-dimensional spaces. Yet in the standard setting, GANs do not well represent dependencies in the extremes. Here we combine GANs with extreme value theory (evtGAN) to model spatial dependencies in summer maxima of temperature and winter maxima in precipitation over a large part of western Europe. We use data from a stationary 2000-year climate model simulation to validate the approach and explore its sensitivity to small sample sizes. Our results show that evtGAN outperforms classical GANs and standard statistical approaches to model spatial extremes. Already with about 50 years of data, which corresponds to commonly available climate records, we obtain reasonably good performance. In general, dependencies between temperature extremes are better captured than dependencies between precipitation extremes due to the high spatial coherence in temperature fields. Our approach can be applied to other climate variables and can be used to emulate climate models when running very long simulations to determine dependencies in the extremes is deemed infeasible.

Impact Statement

Spatially co-occurring climate extremes such as heavy precipitation events or temperature extremes can have devastating impacts on human and natural systems. Modeling complex spatial dependencies between climate extremes in different locations are notoriously difficult and traditional approaches from the field of extreme value theory are relatively inflexible. We show that combining extreme value theory with a deep learning model (generative adversarial networks) can well represent complex spatial dependencies between extremes. Hence, instead of running expensive climate models, the approach can be used to sample many instances of spatially cooccurring extremes with realistic dependence structure, which may be used for climate risk modeling and stress testing of climate-sensitive systems.

---

1. Introduction

Understanding and modeling climate extremes such as floods, heatwaves, and heavy precipitation, are of paramount importance because they often lead to severe impacts on socioeconomic systems (Field et al., 2012). Many impacts are associated with compounding drivers, for instance, multiple hazards occurring at the same time at the same or different locations affecting the same system (Leonard et al., 2014; Zscheischler et al., 2018). Ignoring potential dependencies between multiple hazards can lead to severe misspecification of the associated risk (Zscheischler and Seneviratne, 2017; Hillier and Dixon, 2020). However, estimating dependencies between extremes is a challenging task, requiring large amounts of data and/or suitable approaches to model the phenomena of interest in a computationally feasible time.

A particularly challenging class of events is spatially compounding events (Zscheischler et al., 2020). Spatially compounding events occur when multiple connected locations are affected by the same or different hazards within a limited time window, thereby causing an impact. The compounding is established via a system capable of spatial integration, which accumulates hazard impacts in spatially distant locations. For instance, the global food systems are vulnerable to multiple co-occurring droughts and heatwaves in key crop-producing regions of the world (Anderson et al., 2019; Mehrabi and Ramankutty, 2019). Similarly, spatially extensive floods (Jongman et al., 2014) or sequences of cyclones (Raymond et al., 2020) can deplete emergency response funds. Climate extremes can be correlated over very large distances due to teleconnections (Boers et al., 2019) but modeling such dependencies is challenging.

One approach to tackle the challenge of spatially correlated extremes is to create a large amount of data by running very long simulations with state-of-the-art climate models, which have the physical spatial dependencies built into them. For instance, over the recent years for a number of climate models, large ensembles have been created (Deser et al., 2020). However, these simulations typically have rather coarse resolution, are usually not stationary in time and are very expensive to run.

Extreme value theory provides mathematically justified models for the tail region of a multivariate distribution $(X_1, \ldots, X_d)$, $d \geq 2$. This enables the extrapolation beyond the range of the data and the accurate estimation of the small probabilities of rare (multivariate) events. Statistical methods building on this theory are popular in a broad range of domains such as meteorology (Le et al., 2018), climate science (Naveau et al., 2020), and finance (Poon et al., 2004). Applications have so far been limited to low-dimensional settings for several reasons. On the one hand, even for moderately large dimensions d, the fitting and simulation of parametric models are computationally intensive because it requires computing complex likelihoods (Dombry et al., 2017). On the other hand, the extremal dependence structures in applications are difficult to model and the existing approaches are either simplistic or over-parameterized. In spatial applications, for instance, Euclidean distance is used to parametrize the isotropic dependence of stationary max-stable random fields (Blanchet and Davison, 2011). Most real-world datasets that cover larger spatial domains do, however, feature nonstationarities in space that cannot be captured by the stationarity assumed in current geostatistical models; see Engelke and Ivanovs (2021) for a review on recent extreme value methods in higher dimensions.

Methods from machine learning are well-suited for complex and nonstationary data sets. Their loss functions are, however, typically designed with the purpose to predict well in the bulk of the distribution. It is therefore a difficult problem to construct approaches with an accurate performance outside the range of the training data. In prediction problems, one possibility is to adapt the loss function to make the algorithm more sensitive to extreme values, as for instance done in quantile regression forests (Meinshausen, 2006) or extreme gradient boosting (Velthoen et al., 2021) for the prediction of high conditional quantiles. Jalalzai et al. (2018) discuss this problem for classification in extreme regions and propose a new notion of asymptotic risk that helps to define classifiers with good generalization capacity beyond the observations' range in the predictor space.

Rather than predicting an unknown response, in this work, we are interested in a generative model that is able to learn a high-dimensional distribution $(X_1, \ldots, X_d)$ both in the bulk and in the extremes. We concentrate on generative adversarial networks (GANs), which are known to be competitive models for

multivariate density estimation. While classical applications of GANs are often in the field of image analysis (Zhang et al., 2017; Choi et al., 2018; Karras et al., 2018), they have been used much more broadly in domains such as finance (Efimov et al., 2020), fraud detection (Zheng et al., 2019), speech recognition (Sahu et al., 2020), and medicine (Schlegl et al., 2017).

There are two different aspects that characterize the extremal properties of multivariate data: the univariate tail behavior of each margin and the extremal dependence between the largest observations. For GANs, it has recently been shown (Wiese et al., 2019; Huster et al., 2021) that the marginal distributions of the generated samples are either bounded or light-tailed if the input noise is uniform or Gaussian, respectively. Huster et al. (2021) propose to use heavy-tailed input to overcome this issue. Bhatia et al. (2021) propose the ExGAN algorithm that uses conditional GANs to perform importance sampling of extreme scenarios. The main difference to our approach is that ExGAN simulates single extreme events, while our model will rely on block maxima and therefore does not suffer from serial correlation and seasonality issues. Moreover, ExGAN does not model the marginal distributions with extreme value theory and, since the input noise of ExGAN is Gaussian, this may be problematic especially for heavy-tailed data sets in view of the results of Wiese et al. (2019) and Huster et al. (2021) mentioned above.

In this article, we concentrate on modeling the extremal properties of both the marginal distributions and the dependence structure in a realistic way even for high-dimensional and spatially nonstationary data. Our model, called evtGAN, combines the asymptotic theory of extremes with the flexibility of GANs to overcome the limitations of classical statistical approaches to low dimensional and stationary data. By using extrapolation for the marginals from extreme value theory, our method can be applied to datasets with arbitrary combinations of light-tailed and heavy-tailed distributions. We use a stationary 2000-year climate simulation to test and validate our approach.

In sum, in this article, we demonstrate how our method combines the best of extreme value theory and GANs in order to efficiently and accurately learn complex multidimensional distributions with nontrivial extremal dependence structures. Furthermore, one can easily sample from these learned distributions, which facilitates the study of complex data such as spatially distributed climate extremes with spatially highly heterogeneous dependence structures. The remainder of the article is structured as follows: in Section 2, we briefly lay out the theoretical background of extreme value theory and GANs, and present the main algorithm of our methodology. Section 3 contains a description of the data and the model's architecture and hyperparameters used for its training, and a presentation of the obtained results, which we then discuss in Section 4. Finally, concluding remarks are laid out in Section 5.

2. Methodology

In the climate community, there is a great need for methods able to efficiently provide an empirical description of the climate system, including extreme events, starting from as few ensembles runs as possible (Castruccio et al., 2019; Deser et al., 2020). Emulation techniques that tackle this challenge have been proposed in the last few years, usually focusing on the spatial correlation properties of the phenomena of interest (McKinnon et al., 2017; McKinnon and Deser, 2018; Link et al., 2019; Brunner et al., 2021). In this work, our aim is to build an emulator specifically designed for extreme events that is able to reproduce the spatial tail dependencies of the data and extrapolate outside the range of the training samples. Our method does not require to discard any simulations and is therefore highly efficient. In this section, we first recall some background on extreme value theory and GANs, and then describe our algorithm called evtGAN. Similar to a copula approach (e.g., Brunner et al., 2021), it relies on the idea to disentangle the modeling of the marginal distributions and the extremal dependence structure in order to use extreme value theory for the former and GANs for the latter.

2.1. Extreme value theory

Extreme value theory provides mathematically justified methods to analyze the tail of a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$. This branch of statistical theory has been widely applied in climate

science, for example, to study droughts (Burke et al., 2010), floods (Asadi et al., 2018), heatwaves (Tanarhte et al., 2015), and extreme rainfalls (Le et al., 2018).

For multivariate data, there are two different aspects that determine the quality of an extrapolation method: the univariate tail behavior of each margin and the extremal dependence between the largest observations. We explain how to measure and model those based on componentwise maxima. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, k$, be a sample of $\mathbf{X}$ and let $\mathbf{M}_k = (M_{k1}, \ldots, M_{kd})$, where $M_{kj} = \max(X_{1j}, \ldots, X_{kj})$ is the maximum over the jth margin.

2.1.1. Univariate theory

The Fisher–Tippett–Gnedenko theorem (e.g., Coles et al., 2001) describes the limit behavior of the univariate maximum M_{kj}. It states that if there exist sequences $a_{kj} \in \mathbb{R}$, $b_{kj} > 0$ such that

$$\lim_{k \to \infty} \mathbb{P}\left(\frac{M_{kj} - a_{kj}}{b_{kj}} \le z \right) = G_j(z), \quad n \to \infty, \tag{1}$$

then if G_j is the distribution function of a nondegenerate random variable Z_j, it is in the class of generalized extreme value (GEV) distributions with

$$G_j(z) = \exp\left[-\left\{ 1 + \xi_j \left(\frac{z - \mu_j}{\sigma_j} \right) \right\}_+^{-1/\xi_j} \right], \quad z \in \mathbb{R},$$

where x_+ denotes the positive part of a real number x, and $\xi_j \in \mathbb{R}$, $\mu_j \in \mathbb{R}$, and $\sigma_j > 0$ are the shape, location, and scale parameters, respectively. The shape parameter is the most important parameter since it indicates whether the jth margin is heavy-tailed ($\xi_j > 0$), light-tailed ($\xi = 0$) or whether it has a finite upper end-point ($\xi < 0$).

It can be shown that under mild conditions on the distribution of margin X_j, appropriate sequences exist for (1) to hold. The above theorem, therefore, suggests to fit a generalized extreme value distribution $\widehat{G}_j$ to maxima taken over blocks of the same length, which is common practice when modeling yearly maxima. The shape, location, and scale parameters can then be estimated in different ways, including moment-based estimators Hosking (1985), Bayesian methods (Yoon et al., 2010), and maximum likelihood estimation (Hosking et al., 1985). We use the latter in this work. This allows extrapolation in the direction of the jth marginal since the size of a T-year event can be approximated by the $(1 - 1/T)$-quantile of the distribution $\widehat{G}_j$, even if T is larger than the length of the data record.

2.1.2. Multivariate theory

For multivariate data, the correct extrapolation of the tail not only depends on correct models for the marginal extremes, but also on whether the dependence between marginally large values is well captured. For two components X_i and X_j with limits Z_i and Z_j of the corresponding maxima in (1), this dependence is summarized by the extremal correlation $\chi_{ij} \in [0, 1]$ (Schlather and Tawn, 2003)

$$\chi_{ij} = \lim_{q \to 1} \mathbb{P}\left(F_i(X_i) > q \mid F_j(X_j) > q \right) \in [0, 1].$$

The larger χ_{ij}, the stronger the dependence in the extremes between the two components, that is, between Z_i and Z_j. If $\chi_{ij} > 0$ we speak of asymptotic dependence, and otherwise of asymptotic independence. While the extremal correlation is a useful summary statistic, it does not reflect the complete extremal dependence structure between X_i and X_j. Under asymptotic dependence, the latter can be characterized by the so-called spectral distribution (de Haan and Resnick, 1977). Let $\widetilde{X}_j = -1/\log F_j(X_j), j = 1, \ldots, d$, be the data normalized to standard Fréchet margins. The spectral distribution describes the extremal angle of (X_i, X_j), given that the radius exceeds a high threshold, that is,

$$H(x) = \lim_{u \to \infty} \mathbb{P}\left(\frac{\widetilde{X}_i}{\widetilde{X}_i + \widetilde{X}_j} \le x \mid \widetilde{X}_i + \widetilde{X}_j > u \right), \quad x \in [0, 1].$$

Under strong extremal dependence, the spectral distribution centers around 1/2; under weak extremal independence, it has mass close to the boundary points 0 and 1.

In multivariate extreme value theory, a popular way to ensure a correct extrapolation of tail dependence under asymptotic dependence is by modeling the joint distribution $\mathbf{Z} = (Z_1, \ldots, Z_d)$ of the limits in (1) as a max-stable distribution with multivariate distribution function:

$$\mathbb{P}(G_1(Z_1) \leq z_1, \ldots, G_d(Z_d) \leq z_d) = \exp\{-\ell(z_1, \ldots, z_d)\}, \quad z_1, \ldots, z_d \in [0,1], \tag{2}$$

where ℓ is the so-called stable tail dependence function (e.g., de Haan and Ferreira, 2007). This is a copula approach in the sense that the right-hand side is independent of the original marginal distributions G_j and only describes the dependence structure. In practice, the G_j are replaced by estimates $\widehat{G}_j$ that are fitted to the data first. The right-hand side of (2) is also called an extreme value copula (e.g., Gudendorf and Segers, 2010).

A natural extension of max-stable distributions to the spatial setting is given by max-stable processes $\{Z(t) : t \in \mathbb{R}^d\}$ (e.g., Davison et al., 2012), where the maxima $Z_i = Z(t_i)$ are observed at spatial locations $t_i, i = 1, \ldots, d$. These types of models have been widely applied in many different areas such as flood risk (Asadi et al., 2015), heat waves (Engelke et al., 2019), and extreme precipitation events (Buishand et al., 2008). A popular parametric model for a max-stable process Z is the Brown–Resnick process (Kabluchko et al., 2009; Engelke et al., 2015). This model class is parameterized by a variogram function γ on $\mathbb{R}^d$. In this case, each bivariate distribution is in the so-called Hüsler–Reiss family (Hüsler and Reiss, 1989), and the corresponding dependence parameter is the variogram function evaluated at the distance between the two locations,

$$\lambda_{ij} = \gamma(\|t_i - t_j\|), \quad i, j = 1, \ldots, d.$$

One can show that the extremal correlation coefficient for this model is given by $\chi_{ij} = 2 - 2\Phi(\sqrt{\lambda_{ij}}/2)$, where Φ is the standard normal distribution function. A popular parametric family is the class of fractal variograms that can be written as $\gamma_{\alpha,s}(h) = h^\alpha/s$, $\alpha \in (0,2]$, $s > 0$. While these methods enjoy many interesting properties, they are often quite complex to fit (Dombry et al., 2017) and impose strong assumptions on the spatial phenomena of interest, such as spatial stationarity, isotropic behavior, and asymptotic dependence between all pairs of locations.

A simple way of fitting a Brown–Resnick process to data is to compute the empirical versions $\widehat{\chi}_{ij}$ of the extremal correlation coefficients and then numerically find the parameter values (e.g., α and s for the fractal variogram family) whose implied model extremal coefficients minimize the squared error compared to the empirical ones.

2.2. Generative adversarial network

GANs (Goodfellow et al., 2014) are attracting great interest thanks to their ability of learning complex multivariate distributions that are possibly supported on lower-dimensional manifolds. These methods are generative since they allow the generation of new realistic samples from the learned distribution that can be very different from the training samples.

Given observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$ from some d-dimensional unknown target distribution p_{data}, GANs can be seen as a game opposing two agents, a discriminator D and a generator G. The generator G takes as input a random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)$ from a p-dimensional latent space and transforms it to a new sample $\mathbf{U}^* = G(\mathbf{Y})$. The components $Y_j, j = 1, \ldots, p$, are independent and typically have a standard uniform or Gaussian distribution.

The discriminator D has to decide whether a new sample $\mathbf{U}^*$ generated by G is fake and following the distribution p_G, or coming from the real observations with distribution p_{data}. The discriminator expresses its guesses with a value between 0 and 1 corresponding to its predicted probability of the sample coming from the real data distribution p_{data}. Both the generator and the discriminator become better during the game, in which D is trained to maximize the probability of correctly labeling samples from both sources, and G is trained to minimize D's performance and thus learns to generate more realistic samples.

Mathematically, the optimization problem is a two-player minimax game with cross-entropy objective function:

$$\min_G \max_D \mathbb{E}_{\mathbf{U} \sim p_{\text{data}}}\left[\log\left(D(\mathbf{U})\right)\right] + \mathbb{E}_{\mathbf{Y} \sim p_\mathbf{Y}}\left[\log\left(1 - D(G(\mathbf{Y}))\right)\right]. \tag{3}$$

In equilibrium, the optimal generator satisfies $p_G = p_{\text{data}}$ (Goodfellow et al., 2014).

In practice, the discriminator and the generator are modeled through feed-forward neural networks. While the equilibrium with $p_G = p_{\text{data}}$ is only guaranteed thanks to the convex–concave property in the nonparametric formulation (3), the results remain very good as long as suitable adjustments to the losses, training algorithm, and overall architecture are made to improve on the stability and convergence of training. For instance, a standard adjustment to the generator's loss function was already proposed in the GANs original paper (Goodfellow et al., 2014), and consists of training the generator to minimize $-\log\left(D(G(\mathbf{Y}))\right)$ rather than $\log\left(1 - D(G(\mathbf{Y}))\right)$. This new loss function is called the nonsaturating heuristic loss, and was suggested to mitigate close to null gradients that occur in early training. Indeed, as the generator is still poorly trained, the discriminator can easily detect its samples and therefore associates values close to zero to them. This results in a very slow improvement of the generator during backpropagation as the gradients associated with values around $D(G(\mathbf{Y})) = 0$ are too small, and thus its loss $L(G) = \log\left(1 - D(G(\mathbf{Y}))\right)$ saturates. Training the generator to minimize $-\log\left(D(G(\mathbf{Y}))\right)$ rather than $\log\left(1 - D(G(\mathbf{Y}))\right)$ provides stronger gradients in early training without compromising on the dynamics of the two agents.

The architecture of the generator G and the discriminator D can be specifically designed for the structure of the data under consideration. In this direction, Radford et al. (2016) introduce the use of the convolutional neural network as an add-on feature to the GAN yielding a deep convolutional generative adversarial network model (DCGAN). It is considered the preferred standard for dealing with image data and more generally any high dimensional vector of observations describing a complex underlying dependence structure, showing superior performance in the representation and recognition fields.

2.3. evtGAN

Here we give an overview of evtGAN. Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ be independent observations of dimensionality d, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$ and $m = k \cdot n$. Now let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent observations obtained as block maxima taken sequentially over the span of k observations, that is, $\mathbf{Z}_1 = \max(\mathbf{X}_1, \ldots, \mathbf{X}_k)$, $\mathbf{Z}_2 = \max(\mathbf{X}_{k+1}, \ldots, \mathbf{X}_{2k})$ and so on (see also Section 2.1), where the maxima are taken in each component. The evtGAN algorithm (Algorithm 1) below takes these empirical block maxima as input and follows a copula approach where marginal distributions and the dependence structure are treated separately. In particular, this allows us to impose different amounts of parametric assumptions on the margins and the dependence.

It is known that classical GANs that are trained with bounded or light-tailed noise input distributions in the latent space will also generate bounded or light-tailed samples, respectively (see Wiese et al., 2019; Huster et al., 2021). That means that they are not well-suited for extrapolating in a realistic way outside the range of the data. For this reason, we suppose that the approximation of the margins by generalized extreme value distributions G_j as in (1) holds, which allows for extrapolation beyond the data range. On the other hand, we do not make explicit use of the assumption of a multivariate max-stable distribution as in (2). This has two reasons. First, while the univariate limit (1) holds under very weak assumptions, the multivariate max-stability as in (2) requires much stronger assumptions (see Resnick, 2008) and may be too restrictive for cases with asymptotic independence (e.g., Wadsworth and Tawn, 2012). Second, even if the data follow a multivariate max-stable distribution (2), the probabilistic structure is difficult to impose on a flexible machine learning model such as a GAN.

Algorithm 1: evtGAN

Input: Observations $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})$, $i = 1, \ldots, n$.

1. For $j = 1, \ldots, d$, fit a GEV distribution $\widehat{G}_j$ to the data $Z_{1j}, \ldots, Z_{nj}$ with estimated parameters $\left(\widehat{\mu}_j, \widehat{\sigma}_j, \widehat{\xi}_j\right)$.
2. Normalize all margins empirically to a standard uniform distribution to obtain pseudo-observations:

$$\mathbf{U}_i = \left(\widehat{F}_1(Z_{i1}), \ldots, \widehat{F}_d(Z_{id})\right), \quad i = 1, \ldots, n,$$

where $\widehat{F}_j$ is the empirical distribution function of the $Z_{1j}, \ldots, Z_{nj}$.

3. Train a DCGAN G on the normalized data $\mathbf{U}_1, \ldots, \mathbf{U}_n$ based on the loss in Equation (3).
4. Generate n^* new data points $\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*$ from G with uniform margins, $i = 1, \ldots, n^*$.
5. Normalize back to the scale of the original observations:

$$\mathbf{Z}_i^* = \left(\widehat{G}_1^{-1}(U_{i1}^*), \ldots, \widehat{G}_d^{-1}(U_{id}^*)\right), \quad i = 1, \ldots, n^*.$$

Output: Set of new generated observations $\mathbf{Z}_i^* = (Z_{i1}, \ldots, Z_{id}^*)$, $i = 1, \ldots, n^*$.

The margins are first fitted by a parametric GEV distribution (line 1 in Algorithm 1). They are then normalized to (approximate) pseudo-observations by applying the empirical distribution functions to each margin (line 2). This standardization to uniform margins stabilizes the fitting of the GAN. Alternatively, we could use the fitted GEV distributions for normalization but this seems to give slightly worse results in practice. The pseudo-observations contain the extremal dependence structure of the original observations. Since this dependence structure is very complex in general and our focus is to reproduce gridded spatial fields, we do not rely on restrictive parametric assumptions but rather learn it by a highly flexible DCGAN (line 3). From the fitted model we can efficiently simulate any number of new pseudo-observations that have the correct extremal dependence (line 4). Finally, we transform the generated pseudo-observations back to the original scale to have realistic properties of the new samples also in terms of the margins (line 5).

3. Application

We assess the ability of the proposed method to correctly model the spatial tail dependence structure between climate extremes. Because observational records are relatively short and contain temporal trends, here we rely on 2000 years of climate model output that is representative of present-day weather over western Europe. In particular, we apply our approach to summer temperature and winter precipitation maxima, yielding 2000 and 1999 maxima for temperature and precipitations, respectively. We then compare the performance of our algorithm to the Brown–Resnick model, a state-of-the-art statistical model for spatial extremes. We use the fractal variogram family and the fitting described in Section 2.1.2. To retrieve an unbiased estimate of the error that each of the methods makes, we divide the data into a training set of 50 observations, and the rest is taken as a test set. A sensitivity analysis with other sample sizes is also carried out. Both methods are fitted using only the training data set and the evaluation is made considering the information contained in the test set as ground truth. When not stated otherwise, the results from the Brown–Resnick model are analytical, while the results from evtGAN are obtained simulating $10'000$ data points.

3.1. Data

The model experiment, which uses large ensemble simulations with the EC-Earth global climate model (v2.3, Hazeleger et al., 2012) has been widely used in climate impact studies (Van der Wiel et al., 2019a,

2021; Tschumi et al., 2021; Van Kempen et al., 2021; Vogel et al., 2021) and was originally designed to investigate the influence of natural variability on climate extremes. A detailed description of these climate simulations is provided in Van der Wiel et al. (2019b), here we only provide a short overview. The large ensemble contains 2000 years of daily weather data, representative of the present-day climate. Present-day was defined by the observed value of global mean surface temperature (GMST) over the period 2011–2015 (HadCRUT4 data, Morice et al., 2012); the 5-year time slice in long transient model simulations (forced by historical and Representative Concentration Pathway (RCP) 8.5 scenarios) with the same GMST was simulated repeatedly. To create the large ensemble, at the start of the time slice, 25 ensemble members were branched off from the 16 transient runs. Each ensemble member was integrated for the 5-year time slice period. Differences between ensemble members were forced by choosing different seeds in the atmospheric stochastic perturbations (Buizza et al., 1999). This resulted in a total of $16 \times 25 \times 5 = 2000$ years of meteorological data, at T159 horizontal resolution (approximately 1°), among which we selected temperature (Kelvin) and precipitations (meter/day) for this article.

We choose an area such that it is big enough to be relevant for climate application while being small enough to ensure a fair comparison of our approach and the classical approach in statistics to model spatial tail dependence, the Brown–Resnick process. Our analysis thus focuses on a total of 18×22 grid points covering most of western Europe. For that area, we compute for each grid point the summer temperature maxima and winter precipitation maxima.

3.2. Architecture and hyper-parameters

In our application, we consider the nonsaturating heuristic loss (Goodfellow et al., 2014) for the generator, and the standard empirical cross-entropy loss for the discriminator. Training is done iteratively, where the discriminator is allowed to train two times longer than the generator. A bigger number of discriminator training iterations per generator training iteration was initially considered but it did not achieve better results for the considered sample sizes and input dimensions but only resulted in longer training. We incorporate an exponential moving average scheme (Gidel et al., 2019) to the training algorithm as it has demonstrated more stability, a far better convergence, and improved results.

Since we focus on reproducing gridded spatial fields, we make use of the DCGAN model (Radford et al., 2016) as an architectural design to take advantage of convolution (Fukushima, 1980), and for a more accurate analysis of the input data, we apply zero-padding of a single layer to the train set observations before they are fed to the discriminator (Albawi et al., 2017), increasing the dimension of its input to 20×24. We use classical regularization tools such as drop-out (Srivastava et al., 2014), batch-normal-ization (Ioffe and Szegedy, 2015), and the Adam optimizer (Kingma and Ba, 2015) with a learning rate of 2×10^{-4} and batch size of 50 to train the neural network, and a total of 30,000 training epochs. The value of the exponential moving average parameter in the training algorithm is set to $\alpha = 0.9$. The values of the hyperparameters are the result of an extensive heuristic tuning, and are only valid for the dimension of the input and the considered sample sizes involved in training the network. That being said, we are fairly confident that a minimal intervention in tuning the hyperparameters/architecture would be required if for instance higher resolution inputs were to be used, as these models are fairly robust to such changes. Indeed, models based on the general GANs paradigm are traditionally trained on much higher dimensional objects such as images, and they learn well using an architecture that is very close to the one we have adopted. One may argue that the underlying distributions defining real-world images are "simpler," and therefore easier to learn, which is why we propose the evtGAN model to address the increased complexity of the underlying distribution of the data at hand: initially, we started with a coarser grid (lower resolution input), as we considered France only, then decided to cover a larger part of Europe as the results were very good. In this change, the model's architecture needed to be adjusted only by adding a few layers to the generator. For a given input resolution, increasing the complexity of the model (using a deeper one) just increased training time and pushed convergence further down the line. Figures 1 and 2 illustrate the final architectures of the discriminator and the generator, as well as the values of some of the hyperparameters used for training.
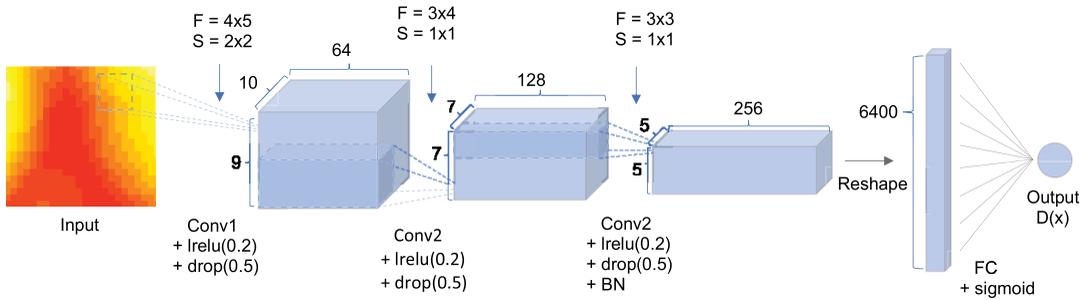
Figure 1. *Discriminator's architecture and values of hyperparameters used for training. lrelu stands for the leaky relu function and its argument corresponds to the value of the slope when x < 0; drop stands for drop-out (see Section 3.2) and its argument corresponds to the drop-out probability; F stands for the kernel's dimension, and S for the value of the stride; FC stands for fully connected, and sigmoid for the sigmoid activation function.*
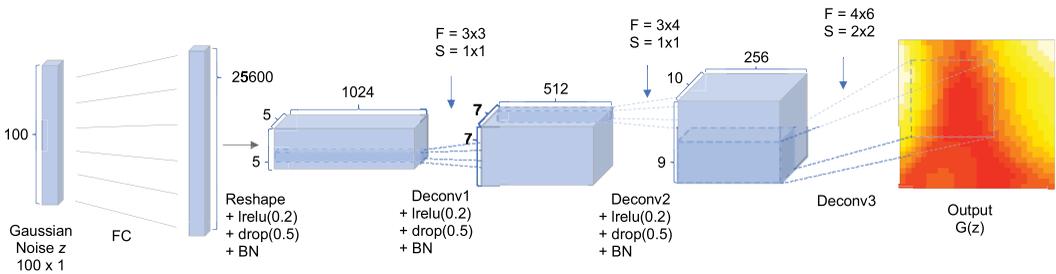
Figure 2. *Generator's architecture and values of hyper-parameters used for training. lrelu stands for the leaky relu function and its argument corresponds to the value of the slope when x < 0; drop stands for drop-out (see Section 3.2) and its argument corresponds to the drop-out probability; F stands for the kernel's dimension, and S for the value of the stride; BN stands for batch-normalization (see Section 3.2); FC stands for fully connected.*

3.3. Results

A particularity of evtGAN is that it decomposes the full high-dimensional distribution of the data into its marginal distributions and its dependence structure and processes them separately. We thus first report estimated key parameters of the marginal distributions. For temperature, the location parameter μ is typically higher over land than over oceans (Figure 3a). Furthermore, there is a trend toward lower values of μ for more northern regions, illustrating the well-known latitudinal temperature gradient from south to north. The scale parameter σ, representing the width of the distribution, is also higher over land than over the ocean, with a fairly homogeneous distribution over land (Figure 3b). The shape parameter ξ is well below 0 for most regions except some areas in the Atlantic north of Spain and in the Mediterranean at the North African coast, indicating a bounded tail (Figure 3c).

For precipitation, the location parameter μ is more similar between land and ocean compared to temperature but spatially more heterogeneous (Figure 3d), illustrating orographic effects on precipitation extremes. For the scale parameter σ, there is also no clear land-sea contrast but areas with relatively much higher values (corresponding to higher interannual variability in extremes) occur over the Western part of the Iberian peninsula and along a coastal band at the northern Mediterranean coast (Figure 3e). The shape parameter ξ is spatially quite heterogeneous, with values ranging from -0.15 to 0.25 (Figure 3f). Overall, the parameters of the extreme value distributions of precipitation extremes shows much higher spatial
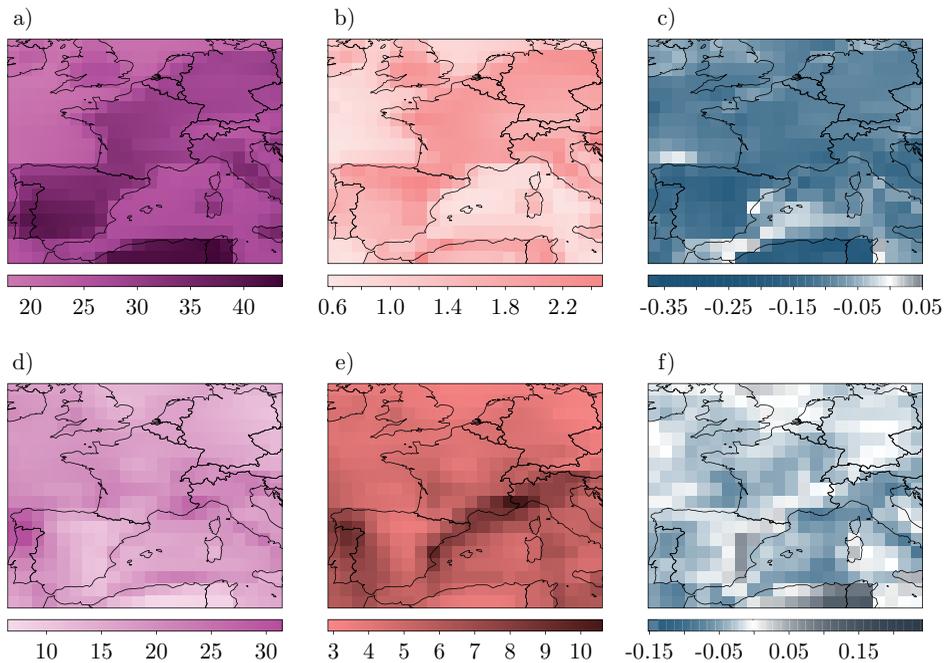
Figure 3. *The generalized extreme value distribution parameters estimated for each grid point for temperature (a–c) and precipitation (d–f) extremes. On the left (a,d) the mean parameter μ, in the middle (b,e) the scale parameter σ, and on the right (c,f) the shape parameter ξ.*

heterogeneity than the ones for temperature extremes, suggesting that it might be more difficult to learn a model that represents well all tail dependencies for the entire region.

Next, we look at examples of bivariate scatterplots of simulated temperature extremes from the different approaches for three cases with varying tail dependence (Figure 4). The three rows correspond to three pairs of grid points with weak, mild, and strong tail dependence, respectively. The train sets (Figure 4a,f,k) illustrate the distributions from which the models are tasked to learn while the test sets (Figure 4b,g,l) are used as ground truth. As can be clearly seen, extremes are much less likely to co-occur between locations that are characterized by weak tail dependence (Figure 4b) compared to locations that are characterized by strong tail dependence (Figure 4l). Purely based on visual judgment, it seems that evtGAN is able to characterize the different tail dependencies relatively well and can simulate samples outside of the range of the train set (Figure 4c,h,m) whereas DCGAN only simulates distributions bounded to the range of the train set (Figure 4d,i,n) and Brown–Resnick tends to overestimate tail dependence in cases where tail dependence is weak (Figure 4e,j).

The corresponding figure for precipitation extremes is shown in Figure 5. Conclusions are similar to the ones derived based on temperature extremes. The evtGAN simulates distributions with different tail dependencies and samples that are far outside the range of the train set (Figure 5c,h,m). DCGAN simulates distributions that are bounded to the train set range (Figure 5d,i,n). On the other hand, Brown–Resnick overestimates tail dependence in the case of mild tail dependence (Figure 5j), and underestimates tail dependence when tail dependence is strong (Figure 5o).

A scatterplot of bivariate extremal correlations between 100 randomly selected locations estimated from the train set, evtGAN, and Brown–Resnick, respectively, against estimates based on the test set (1950 samples) is shown in Figure 6. The estimates derived directly from the train sets (Figure 6a,d) are the benchmark, and by design better performance is not possible. Clearly, pairs of locations with stronger tail dependence are much more likely for temperature (Figure 6a) than for precipitation (Figure 6d),
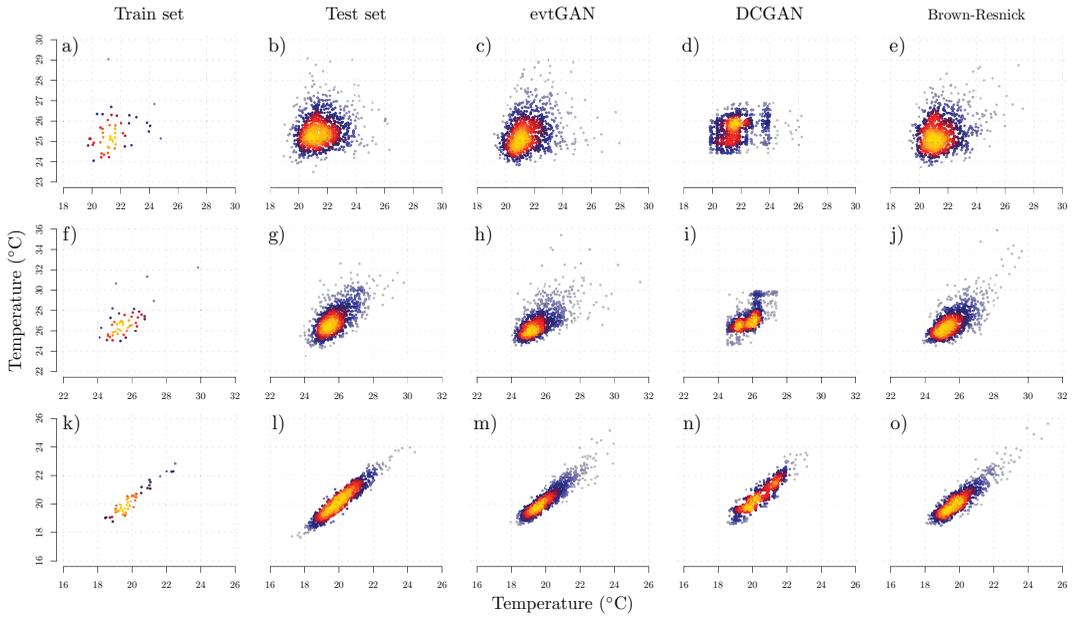
Figure 4. *Bivariate plots of temperature extremes based on train and test sets and simulated through the different presented methods. Shown are selected pairs of locations with varying tail dependence. Columns from left to right: train, test, evtGAN, DCGAN and Brown–Resnick. From top to bottom: weak tail dependence (a–e), mild tail dependence (f–j), and strong tail dependence (k–o). The colors represent the empirically estimated density: the spectrum goes from yellow to gray, reflecting a decrease in density.*
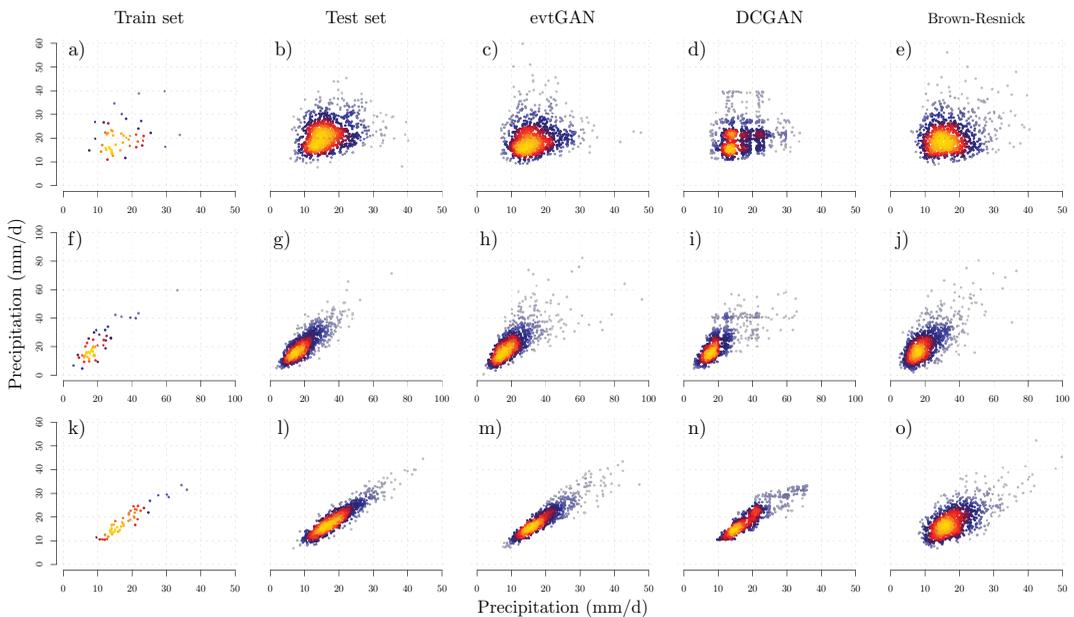
Figure 5. *Bivariate plots of precipitation extremes based on train and test sets and simulated through the different presented methods. Shown are selected pairs of locations with varying tail dependence. Columns from left to right: train, test, evtGAN, DCGAN, and Brown–Resnick. From top to bottom: weak tail dependence (a–e), mild tail dependence (f–j), and strong tail dependence (k–o). The colors represent the empirically estimated density: the spectrum goes from yellow to gray, reflecting a decrease in density.*
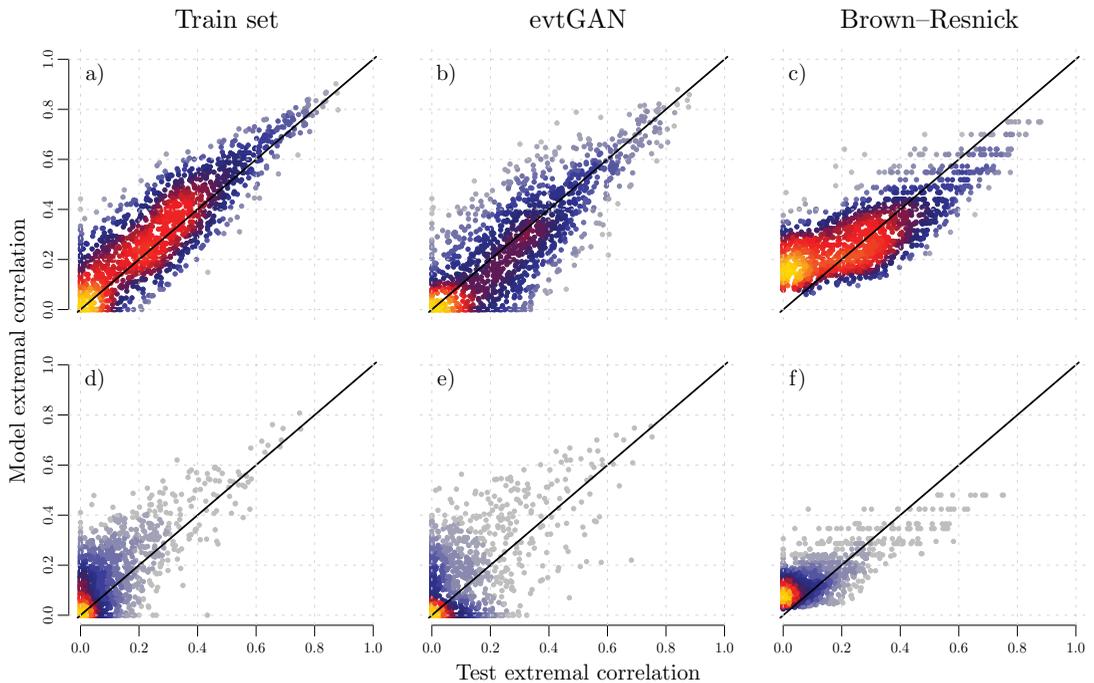
Figure 6. *Scatter plots of the extremal correlations for temperature extremes (a–c) and precipitation extremes (d–f) between* 100 *randomly selected locations. The x-axes always show the estimates based on the test set. Estimates on the y-axes are based on the train set (a,d), on the output of evtGAN (b,d), and on the Brown–Resnick model. The colors represent the empirically estimated density: the spectrum goes from yellow to gray, reflecting a decrease in density.*

confirming the impression obtained from the spatial homogeneity of the parameters in the extreme value distributions (Figure 3). Furthermore, evtGAN seems to better capture the observed relationship. Brown–Resnick has difficulties in particular with pairs that have weak or no tail dependence (extremal correlation equals 0, lower left in the figures), which is consistent with Figures 4 and 5.

To further explore the behavior of evtGAN and Brown–Resnick in dealing with different levels of dependence, Figure 7 shows a comparison of the estimated spectral distribution for the same pairs of chosen grid points shown in Figures 4 and 5, characterized by weak (a,d), mild (b,e) and strong tail dependence (c,f), respectively. The magenta bars show the estimates based on the 1950 samples in the test set. The estimates of evtGAN are very close to the ground truth for all cases, that is, weak, mild, and strong dependence (red lines in Figure 7), except for some bias in the case of mild dependence in precipitation extremes (Figure 7e). In contrast, the performance of the Brown–Resnick model (black lines in Figure 7) is much more variable. It captures relatively well the two pairs with weak tail dependence (Figure 7d) and does a decent job for strong dependence in temperature extremes (Figure 7c) but fails completely for the remaining three cases (Figure 7b,e,f). Furthermore, in contrast to Brown–Resnick, evtGAN is able to represent asymmetric dependence structures, as is evident from Figure 7.

We finally present a sensitivity analysis of the performance of evtGAN for different realistic sample sizes, namely 30, 50, and 100, while learning for 30,000 epochs (Figure 8). Overall the error tends to decrease the longer the models learn and it does not seem to matter whether we evaluate the performance on the train set or the test set (difference between black and red lines in Figure 8 is small). For small sample sizes ($n = 30$) the best performance might be reached for epochs smaller than 30,000 (Figure 8a,b) and the l^2 norm between the extremal coefficients of evtGAN and the train set could be used as a stopping criterion. An improvement in performance with increasing sample size is clearly visible (compare Figure 8a,b with Figure 8e,f). Furthermore, the model errors for temperature extremes are smaller than the ones for precipitation extremes (left vs. right column in Figure 8).
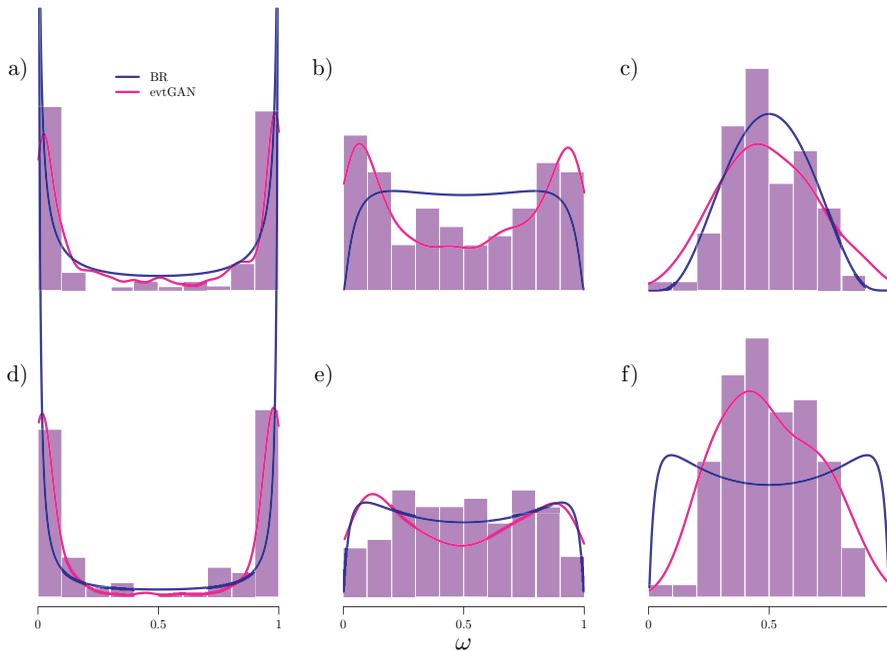
Figure 7. *Spectral distributions for a threshold of 0.95 for selected pairs of locations with varying tail dependence for temperature (a–c) and precipitation (d–f). (a,d) weak tail dependence, (b,e) mild tail dependence, (c,f) strong tail dependence. Shown are kernel density estimates of the evtGAN (red), the Brown–Resnick model (blue) and bars for the ground truth (magenta).*

4. Discussion

The use of evtGAN combines the best of two worlds: correct extrapolation based on extreme value theory on the one hand, and flexible dependence modeling through GANs on the other hand. GANs are an excellent tool to model complex dependencies in high-dimensional spaces. However, they are typically not tailored to model extremes in the marginals well. Indeed, for a standard DCGAN implementation where the marginals are not estimated by GEV distributions but empirically transformed, Figures 4d,i,n and 5d,i,n show that the generated samples are bounded by the range of the training sample. For an accurate extrapolation that resembles the marginal distributions of the test set, extreme value theory is required (panels c,e,h,j,m, and o of Figures 4 and 5).

On the other hand, classical methods of spatial extreme value theory such as the Brown–Resnick process have accurate extrapolation properties for the marginal distributions. However, for an application to a spatially heterogeneous data set on a large domain (Figure 3), their parametric assumption may be too restrictive. Indeed, Figure 6c,f shows a clear bias of the Brown–Resnick model in terms of bivariate extremal correlations, which is particularly visible for pairs with weak extremal dependence. Another indication for this bias can be seen in Figure 5j,o where the distributions of the Brown–Resnick samples differ strongly from the test set distributions. The evtGAN does not make prior assumptions on spatial stationarity or isotropy and therefore it does not exhibit a bias (points in Figure 6b,e are centered around the diagonal). This is particularly noteworthy since modeling complex nonstationarities for extremes is very difficult with parametric models (Huser and Genton, 2016; Engelke and Hitz, 2020). Considering the fitted bivariate distributions of evtGAN and Brown–Resnick underlines this point.

The spectral distributions of the Brown–Resnick model are restricted to a parametric class of functions, which, for instance, are symmetric around $1/2$. The blue lines in Figure 7 show that this is too restrictive for our data since the strength of dependence is not correctly modeled (Figure 7b) or the asymmetry is not captured (Figure 7c,f). The evtGAN (red lines) on the other hand can model weak and strong
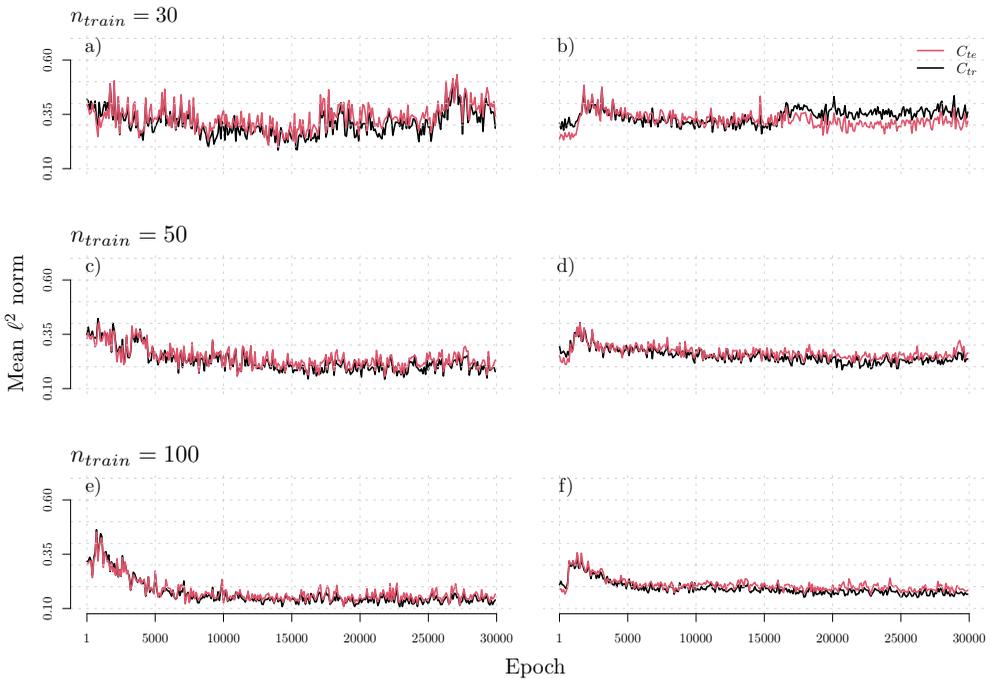
Figure 8. *Model performance versus training epochs for different sample sizes in evtGAN for temperature (a,c,e) and precipitation (b,d,f) extremes. A number of observations equal to $n_{\text{test}} = N - n_{\text{train}}$ were sampled from evtGAN, where $N = 2000$ for temperature, and $N = 1999$ for precipitation. The mean l^2 norms for train (black) and test set (red) are defined as $C_{\text{te}} = \|\chi_{\text{evtG}} - \chi_{\text{te}}\|_2$, $C_{\text{tr}} = \|\chi_{\text{evtG}} - \chi_{\text{tr}}\|_2$, where χ_{evtG}, χ_{te}, and χ_{tr} denote the vectors of extremal correlations calculated on the samples from evtGAN, the test set and the train set, respectively.*

dependences, and it even adapts to possible asymmetries. This is also evident from the scatterplots in Figures 4 and 5, where the shape of the Brown–Resnick samples is restricted to a parametric subclass of distributions, the so-called Hüsler–Reiss family (Hüsler and Reiss, 1989).

A further restriction of classical spatial max-stable processes is the fact that they are limited to modeling asymptotic dependence. For precipitation, it can be seen in Figure 6d–f that most of the test extremal correlations are close to zero, indicating asymptotic independence. While the evtGAN is able to capture this fairly well (Figure 6e), the Brown–Ressnick model always has positive extremal correlations, explaining the bias in the bottom left corner of Figure 6f. A spatial asymptotically independent model (e.g., Wadsworth and Tawn, 2012) would be a possible remedy for this, but such processes would still suffer from the limitation induced by nonstationarity and asymmetry described above.

Overall, evtGAN tends to perform better in capturing dependencies between temperature extremes than precipitation extremes (Figure 8). This is likely related to the fact that extremes in temperature are more spatially coherent (Keggenhoff et al., 2014; Figure 3).

5. Conclusions

Understanding and evaluating the risk associated with extreme events is of primal importance for society, as recently emphasized in the 6th Assessment Report of the Intergovernmental Panel on Climate Change (Seneviratne et al., 2021). Extreme event analysis and impact assessments are often limited by the available sample sizes. Furthermore, simulations with complex climate models are very expensive. Here we combine a machine learning approach with extreme value theory to model complex spatial

dependencies between extreme events in temperature and precipitation across Europe based on a limited sample size. We demonstrate that this hybrid approach outperforms the typically used approach in multivariate extreme value theory and can well represent the marginal distributions and extremal dependencies across spatially distributed climate extremes. The approach can be easily adapted to other types of extremes and used to create large sample sizes that are often required for climate risk assessments.

References

Albawi S, **Mohammed TA and Al-Zawi S** (2017) Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, Antalya, Turkey New York: IEEE, pp. 1–6.

Anderson WB, **Seager R**, **Baethgen W**, **Cane M and You L** (2019) Synchronous crop failures and climate-forced production variability. *Science Advances 5*(7), eaaw1976.

Asadi P, **Davison AC and Engelke S** (2015) Extremes on river networks. *The Annals of Applied Statistics 9*(4), 2023–2050.

Asadi P, **Engelke S and Davison AC** (2018) Optimal regionalization of extreme value distributions for flood estimation. *Journal of Hydrology 556*, 182–193.

Bhatia S, **Jain A and Hooi B** (2021) Exgan: Adversarial generation of extreme samples. *Proceedings of the AAAI Conference on Artificial Intelligence 35*(8), 6750–6758.

Blanchet J and Davison AC (2011) Spatial modeling of extreme snow depth. *The Annals of Applied Statistics 5*, 1699–1725.

Boers N, **Goswami B**, **Rheinwalt A**, **Bookhagen B**, **Hoskins B and Kurths J** (2019) Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature 566*(7744), 373–377.

Brunner MI, **Gilleland E and Wood AW** (2021) Space–time dependence of compound hot–dry events in the United States: Assessment using a multi-site multi-variable weather generator. *Earth System Dynamics 12*(2), 621–634.

Buishand T, **de Haan L and Zhou C** (2008) On spatial extremes: With application to a rainfall problem. *The Annals of Applied Statistics 2*(2), 624–642.

Buizza R, **Milleer M and Palmer TN** (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society 125*(560), 2887–2908.

Burke EJ, **Perry RH and Brown SJ** (2010) An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology 388*(1–2), 131–143.

Castruccio S, **Hu Z**, **Sanderson B**, **Karspeck A and Hammerling D** (2019) Reproducing internal variability with few ensemble runs. *Journal of Climate 32*(24), 8511–8522.

Choi Y, **Choi M**, **Kim M**, **Ha J-W**, **Kim S and Choo J** (2018) Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT New York: IEEE, pp. 8789–8797.

Coles S, **Bawa J**, **Trenner L and Dorazio P** (2001) *An Introduction to Statistical Modeling of Extreme Values*, vol. 208. London: Springer.

Davison AC, **Padoan SA and Ribatet M** (2012) Statistical modeling of spatial extremes. *Statistical Science 27*, 161–186.

de Haan L and Ferreira A (2007) *Extreme Value Theory: An Introduction*. New York, NY: Springer Science & Business Media.

de Haan L and Resnick SI (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 40*(4), 317–337.

Deser C, **Lehner F**, **Rodgers K**, **Ault T**, **Delworth T**, **DiNezio P**, **Fiore A**, **Frankignoul C**, **Fyfe J**, **Horton D**, **Kay JE**, **Knutti R**, **Lovenduski NS**, **Marotzke J**, **McKinnon KA**, **Minobe S**, **Randerson J**, **Screen JA**, **Simpson IR and Ting M** (2020) Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change 10*, 1–10.

Dombry C, **Engelke S and Oesting M** (2017) Bayesian inference for multivariate extreme value distributions. *Electronic Journal of Statistics 11*, 4813–4844.

Efimov D, **Xu D**, **Kong L**, **Nefedov A and Anandakrishnan A** (2020) Using generative adversarial networks to synthesize artificial financial datasets. *Preprint*, arXiv:2002.02271.

Engelke S, de Fondeville R and Oesting M (2019) Extremal behaviour of aggregated data with an application to downscaling. *Biometrika 106*, 127–144.

Engelke S and Hitz AS (2020) Graphical models for extremes. *Journal of the Royal Statistical Society, Series B 82*(4), 871–932.

Engelke S and Ivanovs J (2021) Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application 8*, 241–270.

Engelke S, **Malinowski A**, **Kabluchko Z and Schlather M** (2015) Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society, Series B 77*, 239–265.

Field CB, **Barros V and Stocker TF** (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics 36*, 193–202.

Gidel G, **Berard H**, **Vignoud G**, **Vincent P and Lacoste-Julien S** (2019) A variational inequality perspective on generative adversarial networks. In *International Conference on Learni ng Representations*.

Goodfellow I, **Pouget-Abadie J**, **Mirza M**, **Xu B**, **Warde-Farley D**, **Ozair S**, **Courville A and Bengio Y** (2014) Generative adversarial nets. In Ghahramani Z, Welling M, Cortes C, Lawrence N and Weinberger KQ (eds), *Advances in Neural Information Processing Systems*. New York: Curran Associates, pp. 2672–2680.

Gudendorf G and Segers J (2010) Extreme-value copulas. In Jaworski P, Durante F, Härdle W and Rychlik T (eds), *Copula Theory and Its Applications*. Berlin, Heidelberg: Springer, pp. 127–145.

Hazeleger W, **Wang X**, **Severijns C**, **Stefnescuă S**, **Bintanja R**, **Sterl A**, **Wyser K**, **Semmler T**, **Yang S**, **van den Hurk B**, **van der Linden E and van der Wiel K** (2012) EC-Earth v2. 2: Description and validation of a new seamless earth system prediction model. *Climate Dynamics 39*(11), 2611–2629.

Hillier JK and Dixon RS (2020) Seasonal impact-based mapping of compound hazards. *Environmental Research Letters 15*(11), 114013.

Hosking J (1985) Algorithm as 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Journal of the Royal Statistical Society, Series C 34*(3), 301–310.

Hosking J, **Wallis JR and Wood EF** (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27*(3), 251–261.

Huser R and Genton MG (2016) Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics 21*(3), 470–491.

Hüsler J and Reiss R-D (1989) Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters 7*, 283–286.

Huster T, **Cohen JEJ**, **Lin Z**, **Chan K**, **Kamhoua C**, **Leslie N**, **Chiang C-YJ and Sekar V** (2021) Pareto gan: Extending the representational power of gans to heavy-tailed distributions. In Meila M and Zhang T (eds), *Proceedings of the 38th International Conference on Machine Learning*, Vol. *139*. *Proceedings of Machine Learning Research*, pp. 4523–4532.

Ioffe S and Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Bach F and Blei D (eds), *Proceedings of the 32nd International Conference on Machine Learning*, Vol. *37*. Lille, France: *Proceedings of Machine Learning Research*, pp. 448–456.

Jalalzai H, **Clémençon S and Sabourin A** (2018) On binary classification in extreme regions. In *NeurIPS*, pp. 3096–3104.

Jongman B, **Hochrainer-Stigler S**, **Feyen L**, **Aerts JCJH**, **Mechler R**, **Botzen WJW**, **Bouwer LM**, **Pflug G**, **Rojas R and Ward PJ** (2014) Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change 4*(4), 264–268.

Kabluchko Z, **Schlather M and de Haan L** (2009) Stationary max-stable fields associated to negative definite functions. *Annals of Probability 37*, 2042–2065.

Karras T, **Aila T**, **Laine S and Lehtinen J** (2018) Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations*.

Keggenhoff I, **Elizbarashvili M**, **Amiri-Farahani A and King L** (2014) Trends in daily temperature and precipitation extremes over Georgia, 1971–2010. *Weather and Climate Extremes 4*, 75–85.

Kingma DP and Ba J (2015) Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR) (Poster)*.

Le PD, **Davison AC**, **Engelke S**, **Leonard M and Westra S** (2018) Dependence properties of spatial rainfall extremes and areal reduction factors. *Journal of Hydrology 565*, 711–719.

Leonard M, **Westra S**, **Phatak A**, **Lambert M**, **van den Hurk B**, **McInnes K**, **Risbey J**, **Schuster S**, **Jakob D and Stafford-Smith M** (2014) A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change 5*(1), 113–128.

Link R, **Snyder A**, **Lynch C**, **Hartin C**, **Kravitz B and Bond-Lamberty B** (2019) Fldgen v1. 0: An emulator with internal variability and space–time correlation for earth system models. *Geoscientific Model Development 12*, 1477–1489.

McKinnon KA and Deser C (2018) Internal variability and regional climate trends in an observational large ensemble. *Journal of Climate 31*(17), 6783–6802.

McKinnon KA, **Poppick A**, **Dunn-Sigouin E and Deser C** (2017) An "observational large ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *Journal of Climate 30*(19), 7585–7598.

Mehrabi Z and Ramankutty N (2019) Synchronized failure of global crop production. *Nature Ecology and Evolution 3*(5), 780–786.

Meinshausen N (2006) Quantile regression forests. *Journal of Machine Learning Research 7*, 983–999.

Morice CP, **Kennedy JJ**, **Rayner NA and Jones PD** (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres 117* (D8), 1–22.

Naveau P, **Hannart A and Ribes A** (2020) Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application 7*, 89–110.

Poon S-H, **Rockinger M and Tawn J** (2004) Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies 17*, 581–610.

Radford A, **Metz L and Chintala S** (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434.

Raymond C, **Horton RM**, **Zscheischler J**, **Martius O**, **AghaKouchak A**, **Balch J**, **Bowen SG**, **Camargo SJ**, **Hess J**, **Kornhuber K**, **Oppenheimer M**, **Ruane AC**, **Wahl T and White K** (2020) Understanding and managing connected extreme events. *Nature Climate Change 10*(7), 611–621.

Resnick SI (2008) *Extreme Values, Regular Variation and Point Processes*. New York: Springer.

Sahu S, **Gupta R and Espy-Wilson C** (2020) Modeling feature representations for affective speech using generative adversarial networks. *IEEE Transactions on Affective Computing*, *1*.

Schlather M and Tawn JA (2003) A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika 90*, 139–156.

Schlegl T, **Seeböck P**, **Waldstein SM**, **Schmidt-Erfurth U and Langs G** (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*. Springer, pp. 146–157.

Seneviratne SI, **Zhang X**, **Adnan M**, **Badi W**, **Dereczynski C**, **Di Luca A**, **Ghosh S**, **Iskandar I**, **Kossin J**, **Lewis S**, **Otto F**, **Pinto I**, **Satoh M**, **Vicente-Serrano SM**, **Wehner M and Zhou B** (2021) Weather and climate extreme events in a changing climate. In Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R and Zhou B (eds), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Srivastava N, **Hinton G**, **Krizhevsky A**, **Sutskever I and Salakhutdinov R** (2014) Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research 15*, 1929–1958.

Tanarhte M, **Hadjinicolaou P and Lelieveld J** (2015) Heat wave characteristics in the eastern Mediterranean and Middle East using extreme value theory. *Climate Research 63*, 99–113.

Tschumi E, **Lienert S**, **van der Wiel K**, **Joos F and Zscheischler J** (2021) A climate database with varying drought-heat signatures for climate impact modelling. *Geoscience Data Journal*, 1–13. https://doi.org/10.1002/gdj3.129.

van der Wiel K, **Lenderink G and De Vries H** (2021) Physical storylines of future European drought events like 2018 based on ensemble climate modelling. *Weather and Climate Extremes 33*, 100350.

van der Wiel K, **Stoop LP**, **Van Zuijlen B**, **Blackport R**, **van den Broek M and Selten F** (2019a) Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall. *Renewable and Sustainable Energy Reviews 111*, 261–275.

van der Wiel K, **Wanders N**, **Selten F and Bierkens M** (2019b) Added value of large ensemble simulations for assessing extreme river discharge in a 2 C warmer world. *Geophysical Research Letters 46*(4), 2093–2102.

van Kempen G, **van der Wiel K and Melsen LA** (2021) The impact of hydrological model structure on the simulation of extreme runoff events. *Natural Hazards and Earth System Sciences 21*(3), 961–976.

Velthoen J, **Dombry C**, **Cai J-J and Engelke S** (2021) Gradient boosting for extreme quantile regression. *Preprint*, arXiv: 2103.00808.

Vogel J, **Rivoire P**, **Deidda C**, **Rahimi L**, **Sauter CA**, **Tschumi E**, **van der Wiel K**, **Zhang T and Zscheischler J** (2021) Identifying meteorological drivers of extreme impacts: An application to simulated crop yields. *Earth System Dynamics 12*(1), 151–172.

Wadsworth JL and Tawn JA (2012) Dependence modelling for spatial extremes. *Biometrika 99*(2), 253–272.

Wiese M, **Knobloch R and Korn R** (2019). Copula & marginal flows: Disentangling the marginal from its joint. *Preprint*, arXiv: 1907.03361.

Yoon S, **Cho W**, **Heo J-H and Kim CE** (2010) A full Bayesian approach to generalized maximum likelihood estimation of generalized extreme value distribution. *Stochastic Environmental Research and Risk Assessment 24*(5), 761–770.

Zhang H, **Xu T**, **Li H**, **Zhang S**, **Wang X**, **Huang X and Metaxas DN** (2017) Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, pp. 5907–5915.

Zheng P, **Yuan S**, **Wu X**, **Li J and Lu A** (2019) One-class adversarial nets for fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. *33*. California: Stanford University, pp. 1286–1293.

Zscheischler J, **Martius O**, **Westra S**, **Bevacqua E**, **Raymond C**, **Horton RM**, **van den Hurk B**, **AghaKouchak A**, **Jézéquel A**, **Mahecha MD**, **Maraun D**, **Ramos AM**, **Ridder N**, **Thiery W and Vignotto E** (2020) A typology of compound weather and climate events. *Nature Reviews Earth & Environment 1*, 333–347.

Zscheischler J and Seneviratne SI (2017) Dependence of drivers affects risks associated with compound events. *Science Advances* *3*(6), e1700263.

Zscheischler J, **Westra S**, **van den Hurk BJ**, **Seneviratne SI**, **Ward PJ**, **Pitman A**, **AghaKouchak A**, **Bresch DN**, **Leonard M**, **Wahl T and Zhang X** (2018) Future climate risk from compound events. *Nature Climate Change 8*(6), 469–477.