Machine learning methods for extremes

Sebastian Engelke
www.sengelke.com

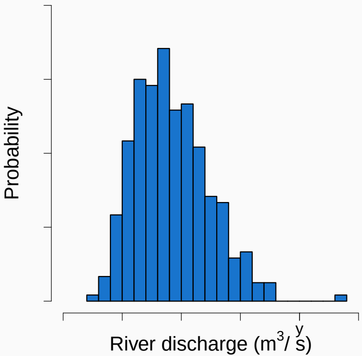
Jointly with
Jasper Velthoen, Clément Dombry, Juan-Juan Cai, Edossa Merga Terefe, Nicola Gnecco, Younes Boulaguiem, Jakob Zscheischler, Edoardo Vignotto, Karin van der Wiel.

Weather Extremes and Climate Change, Paris
January 19, 2022

UNIVERSITÉ DE GENÈVE
FNSNF
Swiss National Science Foundation

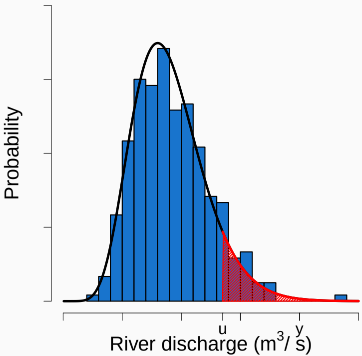
Prediction of extreme conditional quantiles

Generalized Pareto distribution



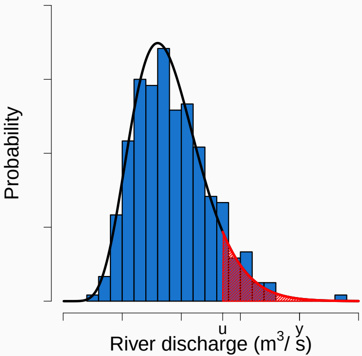
$\mathbb{P}(Y > y)$

Generalized Pareto distribution



$$\mathbb{P}(Y > y) = \mathbb{P}(Y > u) \times \mathbb{P}(Y > y \mid Y > u)$$

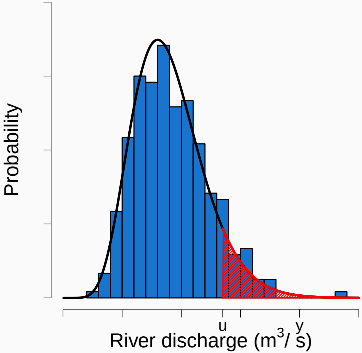
Generalized Pareto distribution



$$\mathbb{P}(Y > y) = \mathbb{P}(Y > u) \times \mathbb{P}(Y > y \mid Y > u)$$
$$\approx \mathbb{P}(Y > u) \times (1 - H_{\sigma,\gamma}(y - u))$$

where $H_{\sigma,\gamma}$ is the cdf of the GPD with scale and shape $\sigma > 0$ and $\gamma \in \mathbb{R}$.

Generalized Pareto distribution



$$\mathbb{P}(Y > y) = \mathbb{P}(Y > u) \times \mathbb{P}(Y > y \mid Y > u)$$
$$\approx \mathbb{P}(Y > u) \times (1 - H_{\sigma,\gamma}(y - u))$$
$$= \mathbb{P}(Y > u) \times \left(1 + \gamma \frac{y - u}{\sigma}\right)^{-1/\gamma}$$

where $H_{\sigma,\gamma}$ is the cdf of the GPD with scale and shape $\sigma > 0$ and $\gamma \in \mathbb{R}$.

Estimation

- Consider i.i.d. data $Y_1, \ldots, Y_n$ and estimate empirically the quantile $u = \hat{Q}(\tau_0)$ for an intermediate quantile level $\tau_0 < 1$.

- Define the exceedances above the threshold as

$$Z_i = \left(Y_i - \hat{Q}(\tau_0) \right)_+ .$$

Estimation

- Consider i.i.d. data $Y_1, \dots, Y_n$ and estimate empirically the quantile $u = \hat{Q}(\tau_0)$ for an intermediate quantile level $\tau_0 < 1$.
- Define the exceedances above the threshold as

$$Z_i = \left(Y_i - \hat{Q}(\tau_0)\right)_+ .$$

- The likelihood of the GPD model with parameters $\theta = (\sigma, \gamma)$ is

$$\ell_{Z_i}(\theta) = -\left[(1 + 1/\gamma)\log\left(1 + \gamma\frac{Z_i}{\sigma}\right) + \log\sigma\right]\mathbb{I}_{Z_i > 0}.$$

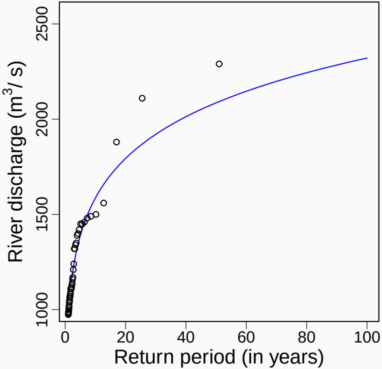
Estimate the parameters by maximum likelihood

$$\hat{\theta} = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \ell_{Z_i}(\theta).$$

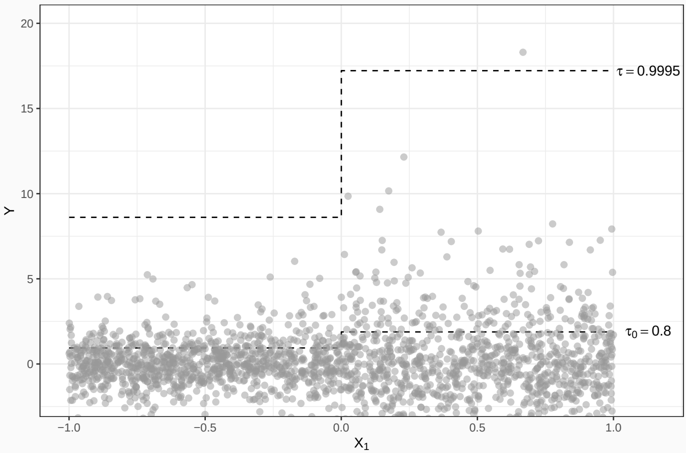
Extreme quantile estimation

- Inverting the cdf $H_{\hat{\sigma},\hat{\gamma}}$ of the GPD provides an approximation of the quantile for probability level $\tau > \tau_0$ by

$$\hat{Q}(\tau) = \hat{Q}(\tau_0) + \hat{\sigma}\,\frac{\left(\frac{1-\tau}{1-\tau_0}\right)^{-\hat{\gamma}} - 1}{\hat{\gamma}}.$$



Extreme quantile regression



Scale shift example with Student's t-distribution (dof=4) for Y; $n = 2000$.

- For i.i.d. data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- There are different scenarios depending on the quantile level $\tau = \tau_n$:
    - $\tau_n \equiv \tau_0 < 1$ (classical case)
    - $\tau_n \to 1$, and $n(1 - \tau_n) \to \infty$ (intermediate case)
    - $\tau_n \to 1$, and $n(1 - \tau_n) \to 0$ (extreme case)

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- There are different scenarios depending on the quantile level $\tau = \tau_n$:
  - $\tau_n \equiv \tau_0 < 1$ (classical case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to \infty$ (intermediate case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to 0$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$.

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- There are different scenarios depending on the quantile level $\tau = \tau_n$:
  - $\tau_n \equiv \tau_0 < 1$ (classical case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to \infty$ (intermediate case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to 0$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$.
- Methods from extreme value theory are not flexible enough (Chernozhukov [2005], Chavez-Demoulin & Davison [2005]) or do not generalize well into higher dimensions (Daouia, Gardes & Girard [2013]).

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- There are different scenarios depending on the quantile level $\tau = \tau_n$:
  - $\tau_n \equiv \tau_0 < 1$ (classical case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to \infty$ (intermediate case)
  - $\tau_n \to 1$, and $n(1 - \tau_n) \to 0$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$.
- Methods from extreme value theory are not flexible enough (Chernozhukov [2005], Chavez-Demoulin & Davison [2005]) or do not generalize well into higher dimensions (Daouia, Gardes & Girard [2013]).
- Goal: Develop a new method for extreme quantile regression that works well with high-dimensional and complex data.

Quantile regression

- Consider independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$.

Quantile regression

- Consider independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$.
- The goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

Quantile regression

- Consider independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$.
- The goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- Most approaches for quantile estimation rely on the property

$$Q_{\mathbf{x}}(\tau) = \operatorname*{argmin}_q \mathbb{E}\left[\rho_\tau(Y - q) \mid \mathbf{X} = \mathbf{x}\right],$$

where $\rho_\tau(u) = (\tau - \mathbb{I}\{u < 0\})u$ is the quantile check function.

Quantile regression

- Consider independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$.
- The goal is to predict the conditional quantile at level $\tau \in (0, 1)$

$$Q_{\mathbf{x}}(\tau) = F_Y^{-1}(\tau \mid \mathbf{X} = \mathbf{x}).$$

- Most approaches for quantile estimation rely on the property

$$Q_{\mathbf{x}}(\tau) = \operatorname*{argmin}_q \mathbb{E}\left[\rho_\tau(Y - q) \mid \mathbf{X} = \mathbf{x}\right],$$

where $\rho_\tau(u) = (\tau - \mathbb{I}\{u < 0\})u$ is the quantile check function.

Without parametric assumption, a pointwise estimator is

$$\hat{Q}_{\mathbf{x}}(\tau) = \operatorname*{argmin}_q \sum_{i=1}^{n} w_i(\mathbf{x}) \rho_\tau(Y_i - q).$$

where $w_1, \ldots, w_n$ is a sequence of localizing weight functions.

Extreme quantile regression

- Assume the GPD model

$$(Y - u \quad | \; Y > u \qquad) \sim H_{\sigma \quad , \gamma}$$

Extreme quantile regression

- Assume the conditional GPD model

$$(Y - u(\mathbf{x}) \mid Y > u(\mathbf{x}), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

Extreme quantile regression

- Assume the conditional GPD model

$$(Y - u(\mathbf{x}) \mid Y > u(\mathbf{x}), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

- Let τ_0 be an intermediate quantile level, and $u(\mathbf{x}) = \hat{Q}_{\mathbf{x}}(\tau_0)$ be an estimate of the conditional τ_0 quantile of $Y \mid \mathbf{X} = \mathbf{x}$; it can be estimated with classical methods, e.g., a quantile random forest.

Extreme quantile regression

- Assume the conditional GPD model

$$(Y - u(\mathbf{x}) \mid Y > u(\mathbf{x}), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

- Let τ_0 be an intermediate quantile level, and $u(\mathbf{x}) = \hat{Q}_{\mathbf{x}}(\tau_0)$ be an estimate of the conditional τ_0 quantile of $Y \mid \mathbf{X} = \mathbf{x}$; it can be estimated with classical methods, e.g., a quantile random forest.
- For a possibly extreme level $\tau > \tau_0$ we can estimate

$$\hat{Q}_{\mathbf{x}}(\tau) = \hat{Q}_{\mathbf{x}}(\tau_0) + \hat{\sigma}(\mathbf{x}) \frac{\left(\frac{1-\tau}{1-\tau_0}\right)^{-\hat{\gamma}(\mathbf{x})} - 1}{\hat{\gamma}(\mathbf{x})},$$

where $\hat{\theta}(\mathbf{x}) = (\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$ is an estimate of the conditional GPD parameters.

Extreme quantile regression

- Assume the conditional GPD model

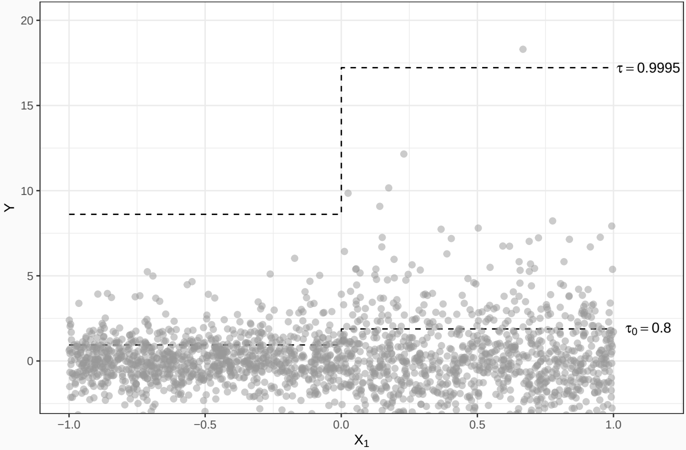
$$(Y - u(\mathbf{x}) \mid Y > u(\mathbf{x}), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

- Let τ_0 be an intermediate quantile level, and $u(\mathbf{x}) = \hat{Q}_{\mathbf{x}}(\tau_0)$ be an estimate of the conditional τ_0 quantile of $Y \mid \mathbf{X} = \mathbf{x}$; it can be estimated with classical methods, e.g., a quantile random forest.
- For a possibly extreme level $\tau > \tau_0$ we can estimate

$$\hat{Q}_{\mathbf{x}}(\tau) = \hat{Q}_{\mathbf{x}}(\tau_0) + \hat{\sigma}(\mathbf{x}) \frac{\left(\frac{1-\tau}{1-\tau_0}\right)^{-\hat{\gamma}(\mathbf{x})} - 1}{\hat{\gamma}(\mathbf{x})},$$

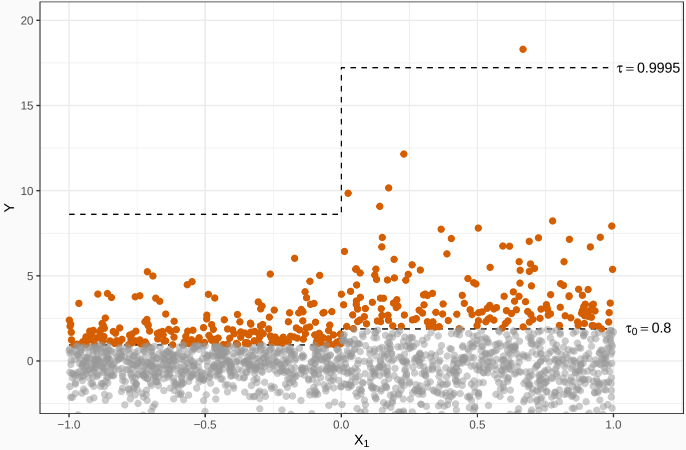
  where $\hat{\theta}(\mathbf{x}) = (\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$ is an estimate of the conditional GPD parameters.
- The triple $(\hat{Q}_{\mathbf{x}}(\tau_0), \hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$ provides a model for the tail of $Y \mid \mathbf{X} = \mathbf{x}$.

Extreme quantile regression



Scale shift example with Student's t-distribution (dof=4) for Y; $n = 2000$.

Extreme quantile regression



Scale shift example with Student's t-distribution (dof=4) for Y; $n = 2000$.

Extreme quantile regression

Two methods to estimate the GPD parameters $\hat{\theta}(\mathbf{x}) = (\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$, both maximize a localized likelihood:

$$\hat{\theta}(\mathbf{x}) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} w_i(\mathbf{x}) \ell_{Z_i}(\theta),$$

where Z_i are the conditional exceedances.

Extreme quantile regression

Two methods to estimate the GPD parameters $\hat{\theta}(\mathbf{x}) = (\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$, both maximize a localized likelihood:

$$\hat{\theta}(\mathbf{x}) = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} w_i(\mathbf{x}) \ell_{Z_i}(\theta),$$

where Z_i are the conditional exceedances.

- **Extremal gradient boosting (GBEX)**: The weights $w_i(\mathbf{x})$ are obtained through gradient boosting.

Velthoen, J., Dombry, C., Cai, J.-J., and Engelke, S. (2021). **Gradient boosting for extreme quantile regression.** https://arxiv.org/abs/2103.00808

Extreme quantile regression

Two methods to estimate the GPD parameters $\hat{\theta}(\mathbf{x}) = (\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x}))$, both maximize a localized likelihood:

$$\hat{\theta}(\mathbf{x}) = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} w_i(\mathbf{x}) \ell_{Z_i}(\theta),$$

where Z_i are the conditional exceedances.

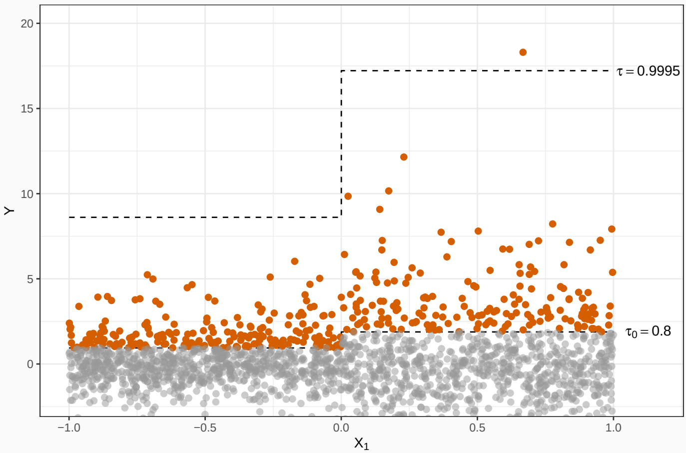
- **Extremal gradient boosting (GBEX)**: The weights $w_i(\mathbf{x})$ are obtained through gradient boosting.

  Velthoen, J., Dombry, C., Cai, J.-J., and Engelke, S. (2021). **Gradient boosting for extreme quantile regression.** https://arxiv.org/abs/2103.00808

- **Extremal random forest (ERF)**: The weights $w_i(\mathbf{x})$ are obtained through a generalized random forest Athey, Tibshirani & Wager [2019].

  Gnecco, N., Terefe, E.M., and Engelke, S. (2022+). **Extremal Random Forests.** *Preprint.*

Extreme quantile regression



Scale shift example with Student's t-distribution (dof=4) for Y; $n = 2000$.

Simulation setup

Setup from [Athey, Tibshirani & Wager, 2019]:

- Different dimensions d and sample size $n = 2000$.
- $\mathbf{X} = (X_1, \ldots, X_d)$ uniform distributed on $[-1, 1]^d$.
- Y follows a Student's t-distribution with dof $= 4$ and scale

  $$\mathrm{scale}(\mathbf{x}) = 1 + \mathbb{I}(x_1 > 0),$$

  that is $\gamma(\mathbf{x}) \equiv 1/4$ and $\sigma(\mathbf{x}) = \sigma(\tau_0)(1 + \mathbb{I}(x_1 > 0))$.

Simulation setup

Setup from [Athey, Tibshirani & Wager, 2019]:

- Different dimensions d and sample size $n = 2000$.
- $\mathbf{X} = (X_1, \ldots, X_d)$ uniform distributed on $[-1, 1]^d$.
- Y follows a Student's t-distribution with dof $= 4$ and scale

  $$\mathrm{scale}(\mathbf{x}) = 1 + \mathbb{I}(x_1 > 0),$$

  that is $\gamma(\mathbf{x}) \equiv 1/4$ and $\sigma(\mathbf{x}) = \sigma(\tau_0)(1 + \mathbb{I}(x_1 > 0))$.
- Intermediate threshold $\tau_0 = 0.8$ and extreme probability levels of interest $\tau = 0.99, 0.995, 0.999$ and 0.9995.

Simulation setup

Setup from [Athey, Tibshirani & Wager, 2019]:

- Different dimensions d and sample size $n = 2000$.
- $\mathbf{X} = (X_1, \ldots, X_d)$ uniform distributed on $[-1, 1]^d$.
- Y follows a Student's t-distribution with dof $= 4$ and scale

  $$\mathrm{scale}(\mathbf{x}) = 1 + \mathbb{I}(x_1 > 0),$$

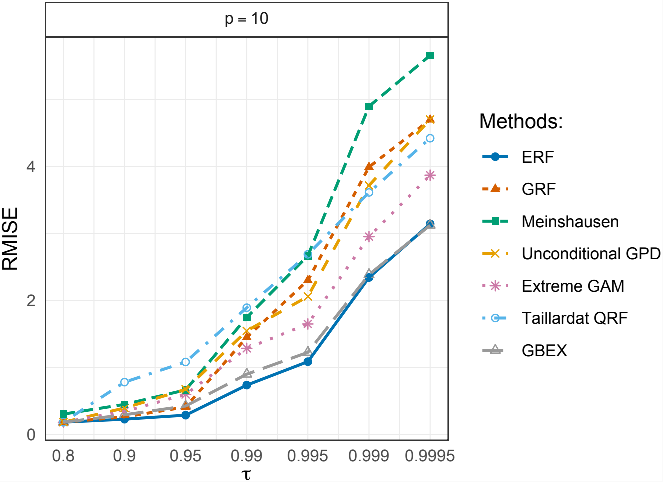
  that is $\gamma(\mathbf{x}) \equiv 1/4$ and $\sigma(\mathbf{x}) = \sigma(\tau_0)(1 + \mathbb{I}(x_1 > 0))$.
- Intermediate threshold $\tau_0 = 0.8$ and extreme probability levels of interest $\tau = 0.99, 0.995, 0.999$ and 0.9995.
- Evaluation metric is the integrated squared error (ISE) for level τ

  $$\mathrm{ISE} = \int_{[-1,1]^d} \left(\hat{Q}_{\mathbf{x}}(\tau) - Q_{\mathbf{x}}(\tau) \right)^2 d\mathbf{x}.$$

- Repeat experiment many times and compute the average of the ISE:

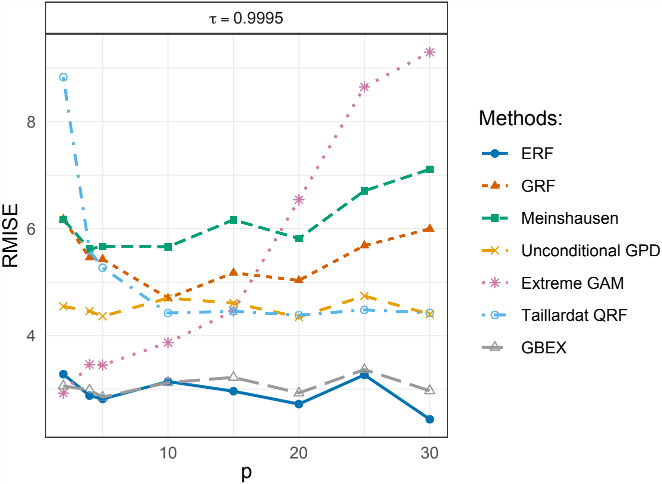
  $$\mathrm{MISE} = \mathbb{E}(\mathrm{ISE}).$$

Results



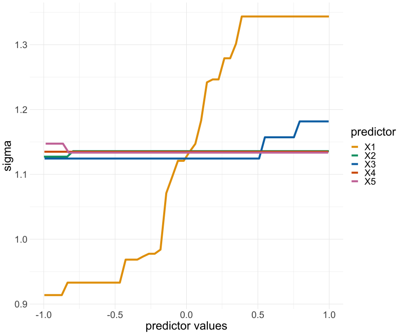
p = 10
RMISE
0
2
4
0.8
0.9
0.95
0.99
0.995
0.999
0.9995
τ
Methods:
ERF
GRF
Meinshausen
Unconditional GPD
Extreme GAM
Taillardat QRF
GBEX

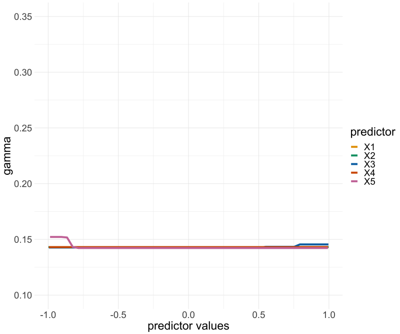
Results



τ = 0.9995
RMISE
8
6
4
10
20
30
p
Methods:
ERF
GRF
Meinshausen
Unconditional GPD
Extreme GAM
Taillardat QRF
GBEX

GBEX: variable importance





Data

- Consider ensembles of numerical weather predictions for precipitation forecasts from the European Centre for Medium-Range Weather Forecasts.
- We have $M = 51$ ensembles and consider a lead time of 36 hours.
- Statistical post-processing is a second step to correct these forecasts for systematic biases and over- or underdispersion.

Data

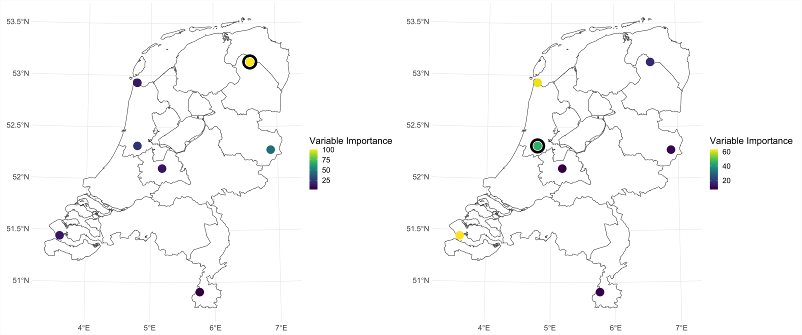
- Consider ensembles of numerical weather predictions for precipitation forecasts from the European Centre for Medium-Range Weather Forecasts.
- We have $M = 51$ ensembles and consider a lead time of 36 hours.
- Statistical post-processing is a second step to correct these forecasts for systematic biases and over- or underdispersion.
- We have 9 years of daily data at 7 locations in the Netherlands.

Data

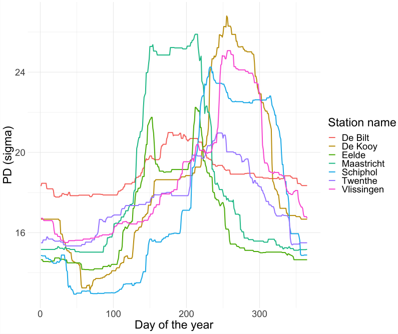
- Consider ensembles of numerical weather predictions for precipitation forecasts from the European Centre for Medium-Range Weather Forecasts.
- We have $M = 51$ ensembles and consider a lead time of 36 hours.
- Statistical post-processing is a second step to correct these forecasts for systematic biases and over- or underdispersion.
- We have 9 years of daily data at 7 locations in the Netherlands.
- We apply our gbex at each location with predictors from all locations:

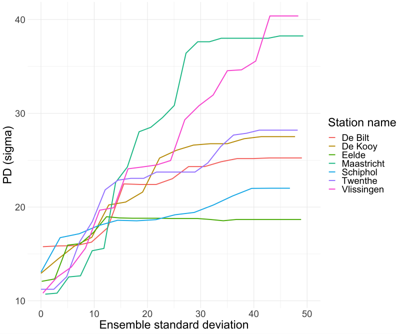
$$X_1, \ldots, X_7 = \texttt{ensemble mean from all locations}$$
$$X_8, \ldots, X_{14} = \texttt{ensemble st. dev. from all locations}$$
$$X_{15}, \ldots, X_{21} = \texttt{ensemble upper order statistics}$$
$$X_{22} = \sin(2\pi \text{day}/365)$$
$$X_{23} = \cos(2\pi \text{day}/365)$$

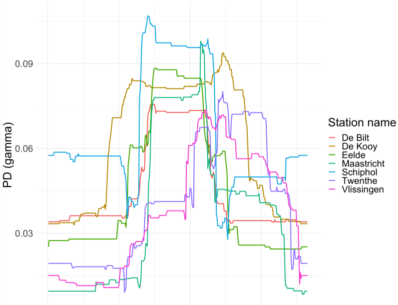
Variable importance

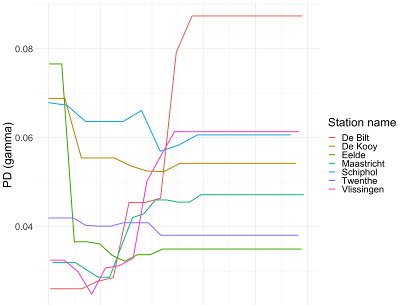


Partial dependence plots









evtGAN: combining EVT and GANs

Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model (v2.3, Hazeleger et al., 2012).
- Present-day climate conditions, stationary in time; see van der Wiel et al. (2019) for details.
- Data at $d = 18 \times 22$ grid points over western Europe.
- We consider annual maxima of precipitation and temperature, giving us $n = 2000$ observations of a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$.

Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model (v2.3, Hazeleger et al., 2012).
- Present-day climate conditions, stationary in time; see van der Wiel et al. (2019) for details.
- Data at $d = 18 \times 22$ grid points over western Europe.
- We consider annual maxima of precipitation and temperature, giving us $n = 2000$ observations of a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$.

Goal: Realistic simulations of $\boldsymbol{Z}$ for stress testing and extreme event simulation.

Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model (v2.3, Hazeleger et al., 2012).
- Present-day climate conditions, stationary in time; see van der Wiel et al. (2019) for details.
- Data at $d = 18 \times 22$ grid points over western Europe.
- We consider annual maxima of precipitation and temperature, giving us $n = 2000$ observations of a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$.

Goal: Realistic simulations of $\boldsymbol{Z}$ for stress testing and extreme event simulation.

Here we use $n_{\text{train}} = 50$ years for training, and $n_{\text{test}} = 1950$ for evaluation.

Marginal GEV parameters

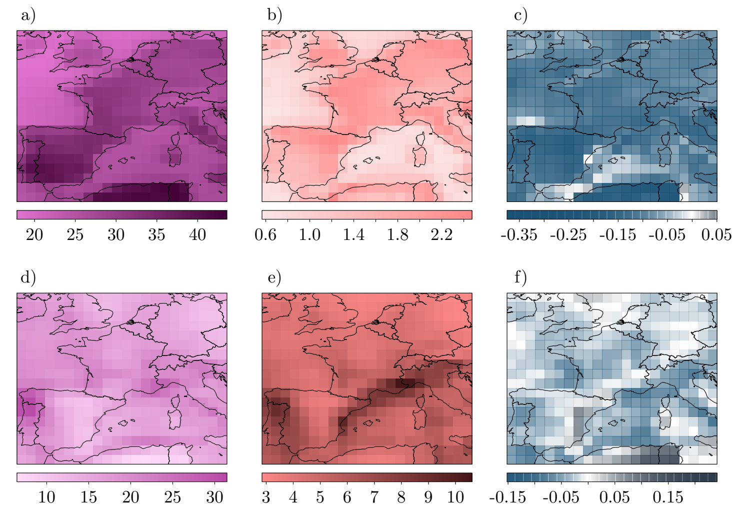


Figure 1: GEV parameters for temperature (a-c) and precipitation (d-f): mean parameter μ (left), scale parameter σ (center) and shape parameter ξ (right).

Two approaches

Classical EVT approach

- Spatial extreme value theory provides statistical models for $\boldsymbol{Z}$, e.g., a spatial max-stable model [Davison and Gholamrezaee, 2012].
- If the region is very large and heterogeneous, such models may not be flexible enough, because of:
  - spatial non-stationarities;
  - asymptotic independence between some stations;
  - etc.

Two approaches

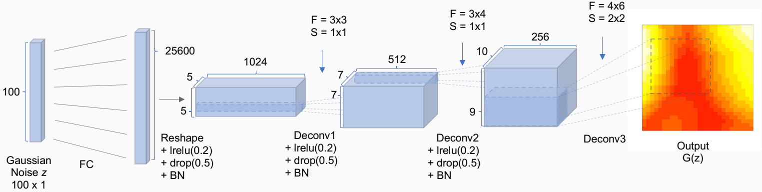
Classical EVT approach

- Spatial extreme value theory provides statistical models for $\boldsymbol{Z}$, e.g., a spatial max-stable model [Davison and Gholamrezaee, 2012].
- If the region is very large and heterogeneous, such models may not be flexible enough, because of:
  - spatial non-stationarities;
  - asymptotic independence between some stations;
  - etc.

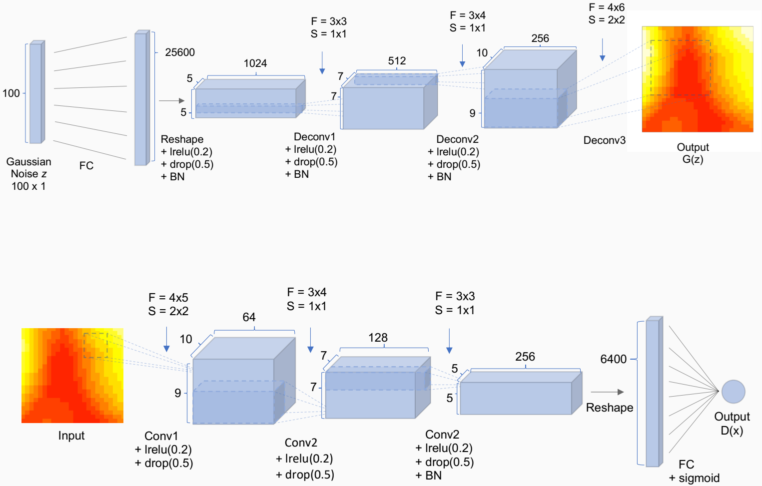
Our ML approach

- Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] are a flexible way of learning and sampling from a multivariate distribution $\boldsymbol{Z}$.
- They are usually used to sample from image data using convolutional neural networks.
- We can treat our spatial climatological data $\boldsymbol{Z}$ as an image.

GAN architecture: generator and discriminator



GAN architecture: generator and discriminator



GANs for extremes

- GANs are trained on the bulk of the distribution.
- There are two main challenges concerning extremes:
  - accurate extrapolation of the marginal distributions;
  - accurate modeling of the extremal dependence structure.

GANs for extremes

- GANs are trained on the bulk of the distribution.
- There are two main challenges concerning extremes:
  - accurate extrapolation of the marginal distributions;
  - accurate modeling of the extremal dependence structure.
- If the input noise is bounded/light-tailed, then the generator output is bounded/light-tailed [Wiese et al., 2019].
- Huster et al. (2021) and Allouche et al. (2021) develop GANs that can generate heavy-tailed output.
- Bhatia et al. (2020) propose a conditional GAN for importance sampling of extreme events.

GANs for extremes

- GANs are trained on the bulk of the distribution.
- There are two main challenges concerning extremes:
  - accurate extrapolation of the marginal distributions;
  - accurate modeling of the extremal dependence structure.
- If the input noise is bounded/light-tailed, then the generator output is bounded/light-tailed [Wiese et al., 2019].
- Huster et al. (2021) and Allouche et al. (2021) develop GANs that can generate heavy-tailed output.
- Bhatia et al. (2020) propose a conditional GAN for importance sampling of extreme events.

Our evtGAN is copula approach where marginals use EVT approximations and dependence the structure is generated by the GAN.

The evtGAN algorithm

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

The evtGAN algorithm

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

1. Fit a GEV distribution $\widehat{G}_j$ to the jth marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.

The evtGAN algorithm

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

1. Fit a GEV distribution $\widehat{G}_j$ to the jth marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.
2. Normalize empirically to a std uniform distribution to obtain pseudo observations

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

   where $\widehat{F}_j$ is the empirical distribution function of the $Z_{1j}, \dots, Z_{nj}$.
3. Train a GAN G on the normalized data $\mathbf{U}_1, \dots, \mathbf{U}_n$.

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

1. Fit a GEV distribution $\widehat{G}_j$ to the jth marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.
2. Normalize empirically to a std uniform distribution to obtain pseudo observations

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

   where $\widehat{F}_j$ is the empirical distribution function of the $Z_{1j}, \dots, Z_{nj}$.
3. Train a GAN G on the normalized data $\mathbf{U}_1, \dots, \mathbf{U}_n$.
4. Generate n^* new data points $\mathbf{U}_1^*, \dots, \mathbf{U}_{n^*}^*$ from G with uniform margins.
5. Normalize back to the scale of the original observations

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

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

Marginal GEV parameters

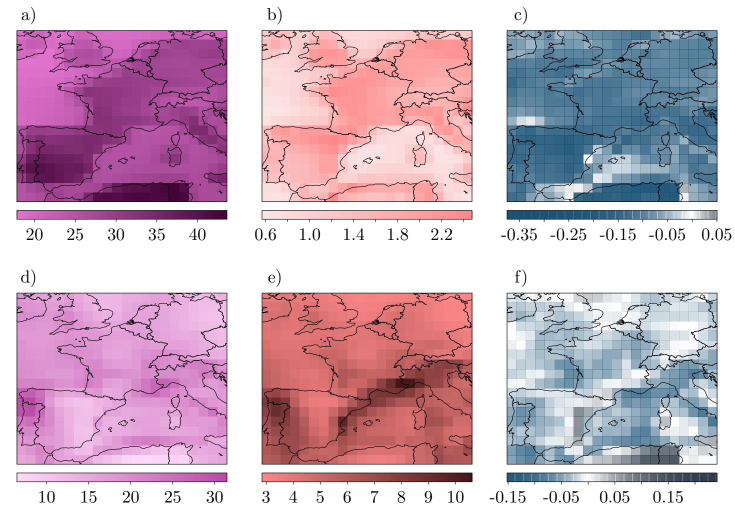
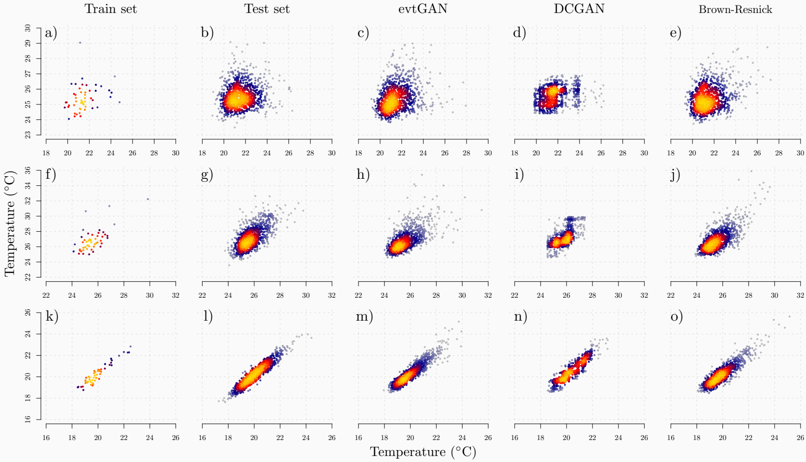
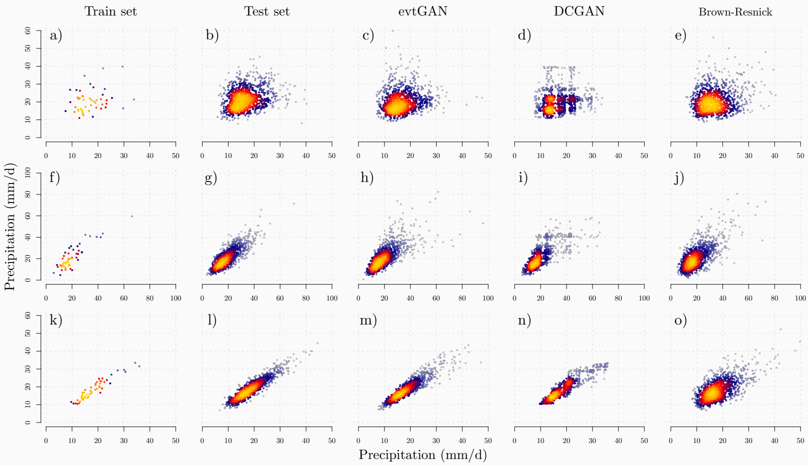


Figure 2: GEV parameters for temperature (a-c) and precipitation (d-f): mean parameter μ (left), scale parameter σ (center) and shape parameter ξ (right).

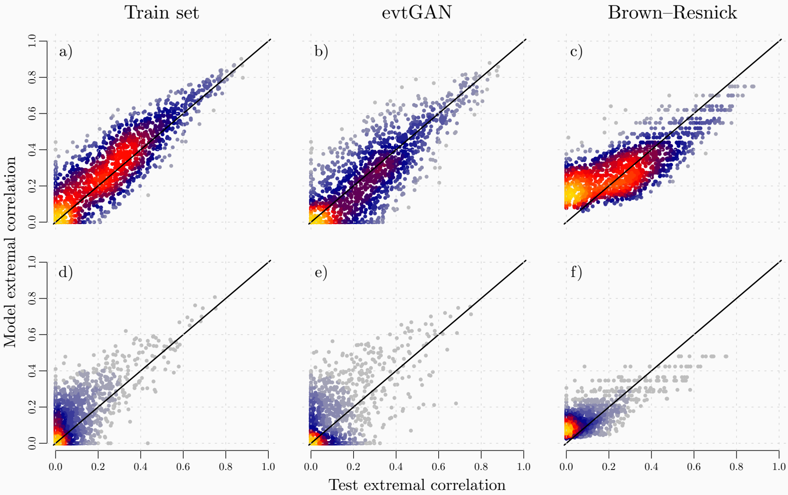
Bivariate samples of temperature



Bivariate samples of precipitation



Extremal correlation plots



Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K. and Engelke, S. (2021).
Modelling and simulating spatial extremes by combining extreme value theory with generative adversarial networks.
https://arxiv.org/abs/2111.00267

Thank You!

Allouche, M., Girard, S., and Gobet, E. (2021).
EV-GAN: Simulation of extreme events with ReLU neural networks.
hal-03250663v2.

Athey, S., Tibshirani, J., Wager, S., et al. (2019).
Generalized random forests.
The Annals of Statistics, 47(2):1148–1178.

Bhatia, S., Jain, A., and Hooi, B. (2020).
Exgan: Adversarial generation of extreme samples.
arXiv preprint arXiv:2009.08454.

Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K., and Engelke, S. (2021).
Modelling and simulating spatial extremes by combining extreme value theory with generative adversarial networks.
Available from `https://arxiv.org/abs/2111.00267`.

Chavez-Demoulin, V. and Davison, A. (2005).
Generalized additive modelling of sample extremes.
Journal of the Royal Statistical Society, series C, 54.

Chernozhukov, V. (2005).
Extremal quantile regression.
Ann. Statist., 33(2):806–839.

Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011).
Kernel estimators of extreme level curves.
Test, Spanish Society of Statistics and Operations Research/Springer, 20(2):311 – 333.

Davison, A. C. and Gholamrezaee, M. M. (2012).
Geostatistics of extremes.
Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 468:581–608.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative adversarial nets.
In *Advances in neural information processing systems*, pages 2672–2680.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al. (2012).
EC-Earth v2. 2: description and validation of a new seamless earth system prediction model.
Climate dynamics, 39(11):2611–2629.

Huster, T., Cohen, J. E. J., Lin, Z., Chan, K., Kamhoua, C., Leslie, N., Chiang, C.-Y. J., and Sekar, V. (2021).
Pareto gan: Extending the representational power of gans to heavy-tailed distributions.
Available from http://arxiv.org/abs/2101.09113.

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

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

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

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