# Machine learning for extreme values

Sebastian Engelke
www.sengelke.com

Jointly with
Nicola Gnecco, Edossa M. Terefe, Olivier Pasche and Gloria Buriticá

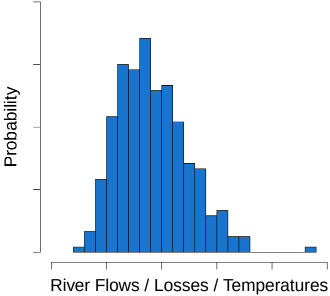
COSMOSTATS, Asiago
September 14, 2023

# Extreme value theory and statistics

- Analysis of rare phenomena with small probabilities
- Impact on various risks (health, environment, economy,...)

# Extreme value theory and statistics

- Analysis of rare phenomena with small probabilities
- Impact on various risks (health, environment, economy,...)

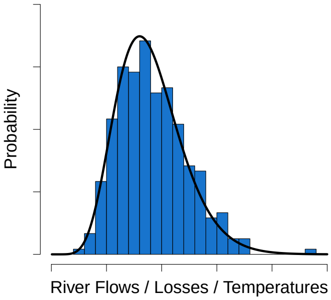





# Extreme value theory and statistics

– Analysis of rare phenomena with small probabilities

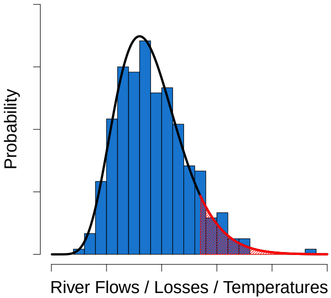
– Impact on various risks (health, environment, economy,...)






# Extreme value theory and statistics

- Analysis of rare phenomena with small probabilities
- Impact on various risks (health, environment, economy,...)






# My research interests

- Graphical models: sparsity and efficient computations [1] [2]
- Causal inference for extremes [3] [4]
- Machine learning: prediction beyond the data range

[1] S. Engelke and A. S. Hitz. "Graphical models for extremes (with discussion)". In: *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 82.4 (2020), pp. 871–932.

[2] S. Engelke and J. Ivanovs. "Sparse structures for multivariate extremes". In: *Annu. Rev. Stat. Appl.* 8 (2021), pp. 241–270.

[3] D. Deuber, J. Li, S. Engelke, and M. H. Maathuis. "Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions". In: *J. Amer. Statist. Assoc.* (2023). to appear.

[4] N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke. "Causal discovery in heavy-tailed models". In: *Ann. Statist.* 49.3 (2021), pp. 1755–1778.

# My research interests

- Graphical models: sparsity and efficient computations [1] [2]
- Causal inference for extremes [3] [4]
- Machine learning: prediction beyond the data range

[1] S. Engelke and A. S. Hitz. "Graphical models for extremes (with discussion)". In: *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 82.4 (2020), pp. 871–932.
[2] S. Engelke and J. Ivanovs. "Sparse structures for multivariate extremes". In: *Annu. Rev. Stat. Appl.* 8 (2021), pp. 241–270.
[3] D. Deuber, J. Li, S. Engelke, and M. H. Maathuis. "Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions". In: *J. Amer. Statist. Assoc.* (2023). to appear.
[4] N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke. "Causal discovery in heavy-tailed models". In: *Ann. Statist.* 49.3 (2021), pp. 1755–1778.

# My research interests

- Graphical models: sparsity and efficient computations [1] [2]
- Causal inference for extremes [3] [4]
- Machine learning: prediction beyond the data range

---

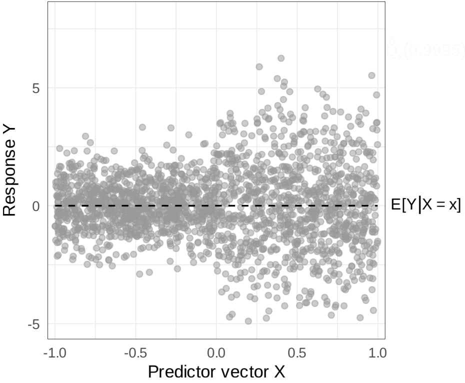
[1] S. Engelke and A. S. Hitz. "Graphical models for extremes (with discussion)". In: *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 82.4 (2020), pp. 871–932.

[2] S. Engelke and J. Ivanovs. "Sparse structures for multivariate extremes". In: *Annu. Rev. Stat. Appl.* 8 (2021), pp. 241–270.

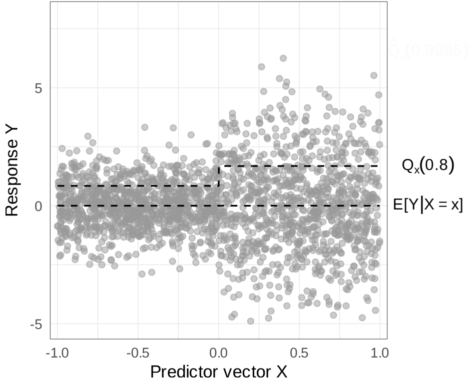
[3] D. Deuber, J. Li, S. Engelke, and M. H. Maathuis. "Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions". In: *J. Amer. Statist. Assoc.* (2023). to appear.

[4] N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke. "Causal discovery in heavy-tailed models". In: *Ann. Statist.* 49.3 (2021), pp. 1755–1778.

# Beyond data range: extreme quantile regression



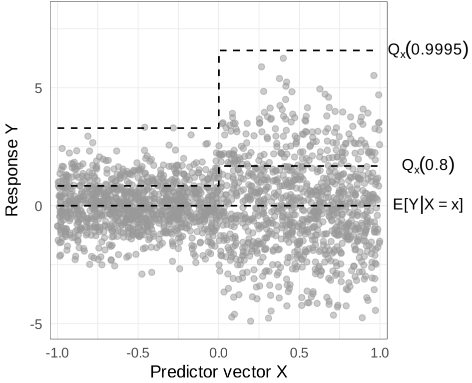
# Beyond data range: extreme quantile regression




– Prediction of conditional quantile at level $\tau \in (0, 1)$:

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

# Beyond data range: extreme quantile regression

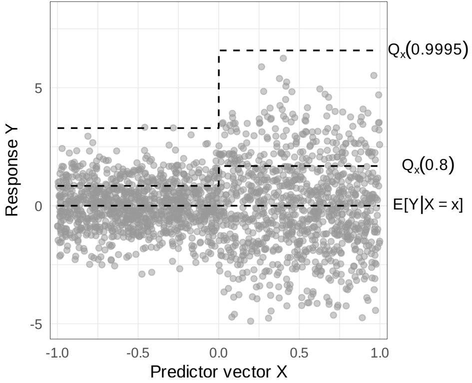



– Prediction of conditional quantile at level $\tau \in (0, 1)$:

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

– If $\tau$ is close to 1, we speak of extreme quantile regression

# Beyond data range: extreme quantile regression



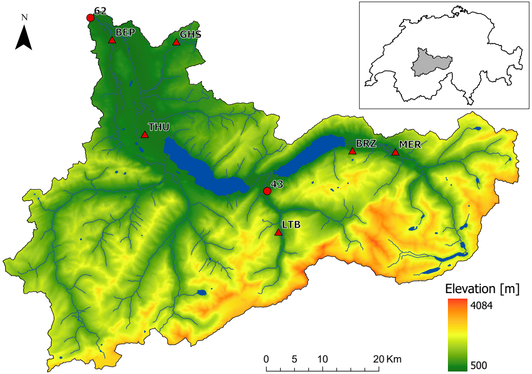

– Prediction of conditional quantile at level $\tau \in (0, 1)$:

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

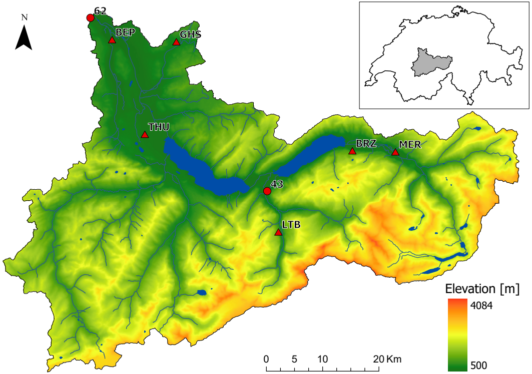
– If $\tau$ is close to 1, we speak of extreme quantile regression

– Again, classical machine learning methods perform poorly

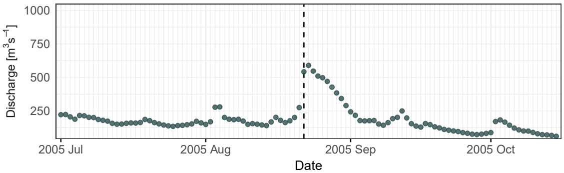
# Extreme quantile regression: Aare river in Bern



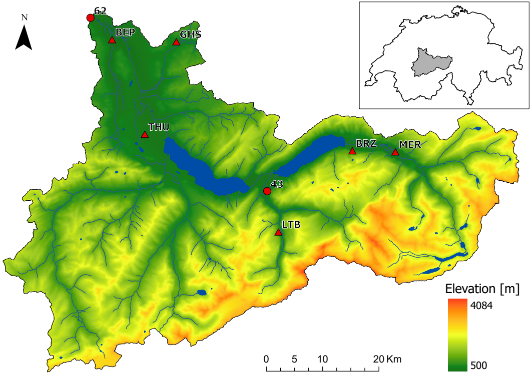
# Extreme quantile regression: Aare river in Bern



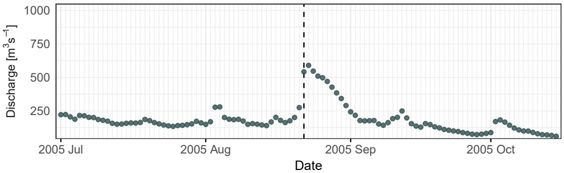

– Late warnings during flood in August 2005



# Extreme quantile regression: Aare river in Bern




- Late warnings during flood in August 2005




- $Y$ is daily discharge at Bern station
- $\mathbf{X}$ contains discharge and precipitation from previous days, etc.

# Extreme quantile regression

– For independent copies $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

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

[5] V. Chernozhukov. "Extremal quantile regression". In: *The Annals of Statistics* 33.2 (2005), pp. 806–839.

[6] V. Chavez-Demoulin and A. C. Davison. "Generalized additive modelling of sample extremes". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.1 (2005), pp. 207–222.

# Extreme quantile regression

- For independent copies $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ of $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, 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 $n(1 - \tau_n) \to c \in [0, \infty)$ (extreme case)

[5] V. Chernozhukov. "Extremal quantile regression". In: *The Annals of Statistics* 33.2 (2005), pp. 806–839.
[6] V. Chavez-Demoulin and A. C. Davison. "Generalized additive modelling of sample extremes". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.1 (2005), pp. 207–222.

# Extreme quantile regression

- For independent copies $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, 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 $n(1-\tau_n) \to c \in [0, \infty)$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$

[5] V. Chernozhukov. "Extremal quantile regression". In: *The Annals of Statistics* 33.2 (2005), pp. 806–839.
[6] V. Chavez-Demoulin and A. C. Davison. "Generalized additive modelling of sample extremes". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.1 (2005), pp. 207–222.

# Extreme quantile regression

- For independent copies $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, 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 $n(1-\tau_n) \to c \in [0, \infty)$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$
- Existing methods from extreme value theory are not flexible enough or do not generalize well to higher dimensions [5] [6]

---

[5] V. Chernozhukov. "Extremal quantile regression". In: *The Annals of Statistics* 33.2 (2005), pp. 806–839.
[6] V. Chavez-Demoulin and A. C. Davison. "Generalized additive modelling of sample extremes". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.1 (2005), pp. 207–222.

# Extreme quantile regression

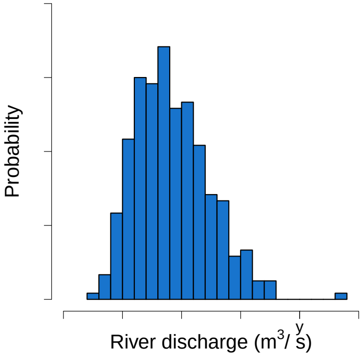
- For independent copies $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$, the goal is to predict the conditional quantile at level $\tau \in (0, 1)$

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

- 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 $n(1 - \tau_n) \to c \in [0, \infty)$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$
- Existing methods from extreme value theory are not flexible enough or do not generalize well to higher dimensions [5] [6]
- Goal: Develop a new method for extreme quantile regression that works well with high-dimensional and complex data

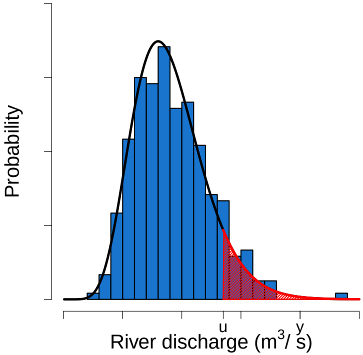
[5] V. Chernozhukov. "Extremal quantile regression". In: *The Annals of Statistics* 33.2 (2005), pp. 806–839.
[6] V. Chavez-Demoulin and A. C. Davison. "Generalized additive modelling of sample extremes". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.1 (2005), pp. 207–222.

# Generalized Pareto distribution



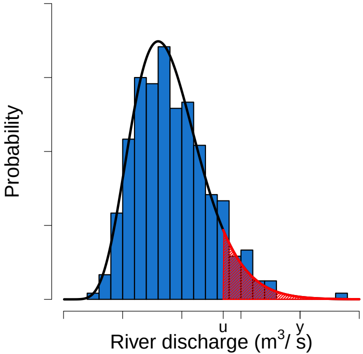

$\mathbb{P}(Y > y)$

# Generalized Pareto distribution




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

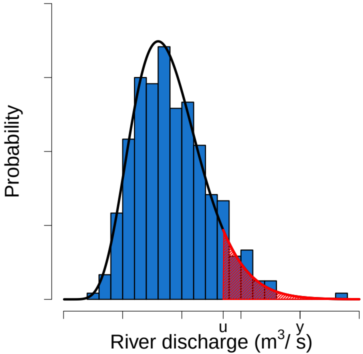
# Generalized Pareto distribution




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

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

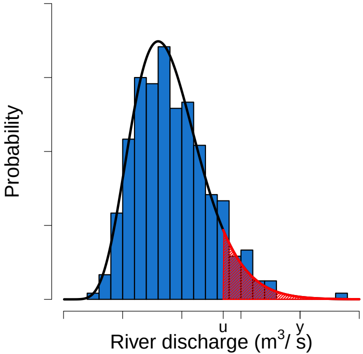
# Generalized Pareto distribution




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

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

# Generalized Pareto distribution




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

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

The sign of $\gamma$ characterizes the domain of attraction (DOA) of $Y$

- $\gamma > 0$: $Y$ has heavy tails (Fréchet DOA)
- $\gamma = 0$: $Y$ has light tails (Gumbel DOA)
- $\gamma < 0$: $Y$ has finite upper endpoint (Weibull DOA)

# 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

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

  follow approximately a GPD $H_{\sigma,\gamma}$,

# 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

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

follow approximately a GPD $H_{\sigma,\gamma}$, with log-likelihood of the parameters $\theta = (\sigma, \gamma)$

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

# 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

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

  follow approximately a GPD $H_{\sigma,\gamma}$, with log-likelihood of the parameters $\theta = (\sigma, \gamma)$

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

Estimate the parameters by maximum likelihood

$$\hat{\theta} = \underset{\theta=(\sigma,\gamma)}{\arg\max} \sum_{i=1}^{n} \ell_{Z_i}(\theta).$$

# Extreme quantile estimation

– Approximation of quantile $Q(\tau) = F_Y^{-1}(\tau)$ for probability level $\tau > \tau_0$ by inverting the cdf $H_{\hat{\sigma},\hat{\gamma}}$ of the GPD

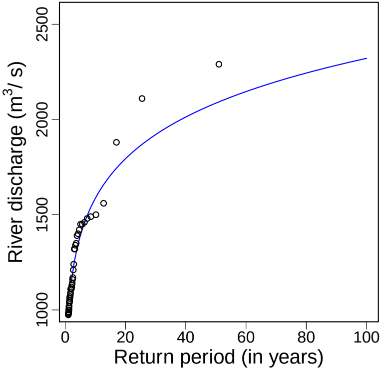
$$\hat{Q}(\tau) = \hat{Q}(\tau_0) + \hat{\sigma}\,\frac{\left(\frac{1-\tau}{1-\tau_0}\right)^{-\hat{\gamma}} - 1}{\hat{\gamma}}.$$

# Extreme quantile estimation

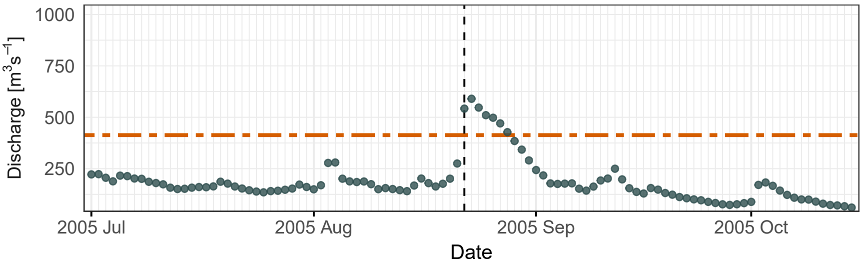
- Approximation of quantile $Q(\tau) = F_Y^{-1}(\tau)$ for probability level $\tau > \tau_0$ by inverting the cdf $H_{\hat{\sigma},\hat{\gamma}}$ of the GPD

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

- Compute a $T$-year event as $\hat{Q}(1 - 1/(n_Y T))$, where $n_Y$ is the number of observations per year.



# During the 2005 flood in Bern




– Daily observations during the 2005 flood in Bern together with 100-year return level estimate.

# Extreme quantile regression

– Assume the GPD model

$$(Y - \hat{Q}\ (\tau_0) \mid Y > \hat{Q}\ (\tau_0), \qquad ) \sim H_{\sigma \quad ,\gamma}$$

where $\tau_0$ is an intermediate quantile level

# Extreme quantile regression

– Assume the conditional GPD model

$$(Y - \hat{Q}_{\mathbf{x}}(\tau_0) \mid Y > \hat{Q}_{\mathbf{x}}(\tau_0), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

where $\tau_0$ is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional $\tau_0$ quantile of $Y \mid \mathbf{X} = \mathbf{x}$

# Extreme quantile regression

- Assume the conditional GPD model

$$(Y - \hat{Q}_{\mathbf{x}}(\tau_0) \mid Y > \hat{Q}_{\mathbf{x}}(\tau_0), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

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

# Extreme quantile regression

- Assume the conditional GPD model

$$(Y - \hat{Q}_{\mathbf{x}}(\tau_0) \mid Y > \hat{Q}_{\mathbf{x}}(\tau_0), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

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

$$(Y - \hat{Q}_{\mathsf{x}}(\tau_0) \mid Y > \hat{Q}_{\mathsf{x}}(\tau_0), \mathbf{X} = \mathbf{x}) \sim H_{\sigma(\mathbf{x}),\gamma(\mathbf{x})}$$

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

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

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

# Generalized random forest

– Generalized random forests (GRF) [7] are ensemble methods based on trees that can be seen as adaptive nearest neighbors that optimize

$$\hat{\theta}(\mathbf{x}) = \arg\min_{\theta} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i) L(\theta, Y_i),$$

where $L$ is a loss function and $w_n(\mathbf{x}, \mathbf{X}_i)$ are localizing weights from the GRF

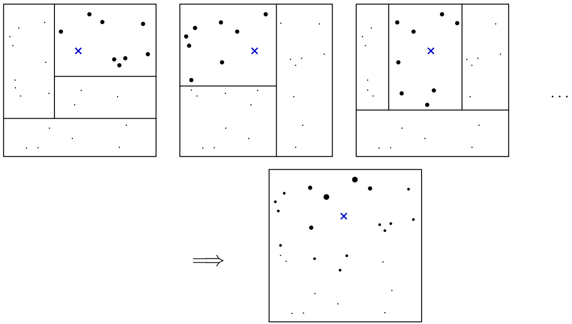
[7] S. Athey, J. Tibshirani, and S. Wager. "Generalized random forests". In: *Ann. Statist.* 47.2 (2019), pp. 1148–1178.

# Generalized random forest

– Generalized random forests (GRF) [7] are ensemble methods based on trees that can be seen as adaptive nearest neighbors that optimize

$$\hat{\theta}(\mathbf{x}) = \arg\min_{\theta} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i) L(\theta, Y_i),$$

where $L$ is a loss function and $w_n(\mathbf{x}, \mathbf{X}_i)$ are localizing weights from the GRF



[7] S. Athey, J. Tibshirani, and S. Wager. "Generalized random forests". In: *Ann. Statist.* 47.2 (2019), pp. 1148–1178.

# Generalized random forest

- Generalized random forests (GRF) [7] are ensemble methods based on trees that can be seen as adaptive nearest neighbors that optimize

$$\hat{\theta}(\mathbf{x}) = \arg\min_{\theta} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i) L(\theta, Y_i),$$

  where $L$ is a loss function and $w_n(\mathbf{x}, \mathbf{X}_i)$ are localizing weights from the GRF

- Least squares regression: loss $L(\theta, Y) = (Y - \theta)^2$ and

$$\hat{\theta}(\mathbf{x}) \approx \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$$

[7] S. Athey, J. Tibshirani, and S. Wager. "Generalized random forests". In: *Ann. Statist.* 47.2 (2019), pp. 1148–1178.

# Generalized random forest

- Generalized random forests (GRF) [7] are ensemble methods based on trees that can be seen as adaptive nearest neighbors that optimize

$$\hat{\theta}(\mathbf{x}) = \arg\min_{\theta} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i) L(\theta, Y_i),$$

  where $L$ is a loss function and $w_n(\mathbf{x}, \mathbf{X}_i)$ are localizing weights from the GRF

- Least squares regression: loss $L(\theta, Y) = (Y - \theta)^2$ and

$$\hat{\theta}(\mathbf{x}) \approx \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$$

- Quantile regression at $\tau \in (0, 1)$: $L(\theta, Y) = \rho_\tau(\theta, Y)$ is quantile loss and

$$\hat{\theta}(\mathbf{x}) \approx Q_{\mathbf{x}}(\tau)$$

[7] S. Athey, J. Tibshirani, and S. Wager. "Generalized random forests". In: *Ann. Statist.* 47.2 (2019), pp. 1148–1178.

# Generalized random forest

- Generalized random forests (GRF) [7] are ensemble methods based on trees that can be seen as adaptive nearest neighbors that optimize

$$\hat{\theta}(\mathbf{x}) = \arg\min_{\theta} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i) L(\theta, Y_i),$$

  where $L$ is a loss function and $w_n(\mathbf{x}, \mathbf{X}_i)$ are localizing weights from the GRF

- Least squares regression: loss $L(\theta, Y) = (Y - \theta)^2$ and

$$\hat{\theta}(\mathbf{x}) \approx \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$$

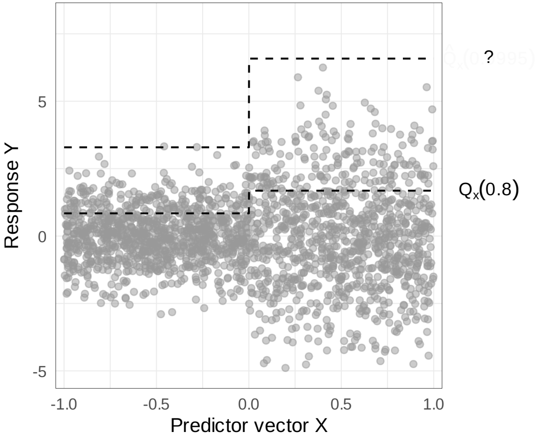
- Quantile regression at $\tau \in (0, 1)$: $L(\theta, Y) = \rho_\tau(\theta, Y)$ is quantile loss and

$$\hat{\theta}(\mathbf{x}) \approx Q_{\mathbf{x}}(\tau)$$

- For extreme quantile levels $\tau \approx 1$, this does not work well

[7] S. Athey, J. Tibshirani, and S. Wager. "Generalized random forests". In: *Ann. Statist.* 47.2 (2019), pp. 1148–1178.

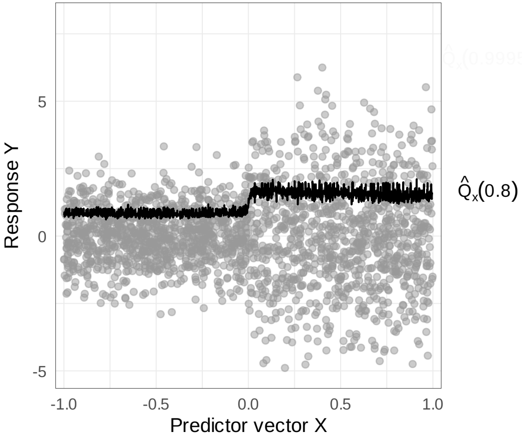
# Extremal Random Forest (ERF)




[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

# Extremal Random Forest (ERF)

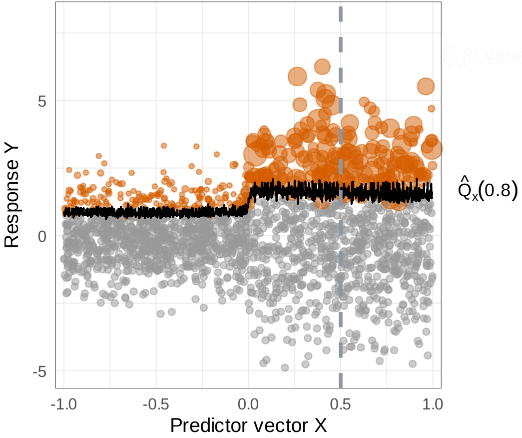
– Estimate intermediate quantile $\hat{Q}_x(\tau_0)$




[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

# Extremal Random Forest (ERF)

- Estimate intermediate quantile $\hat{Q}_{\mathbf{x}}(\tau_0)$
- Compute exceedances $Z_i = (Y_i - \hat{Q}_{\mathbf{x}}(\tau_0))_+$ and weights $w_n(\mathbf{x}, \mathbf{X}_i)$

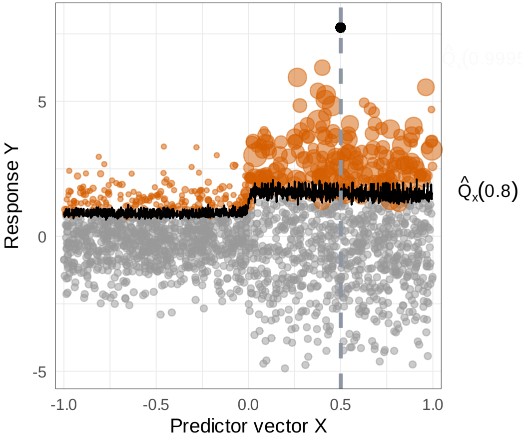



[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

# Extremal Random Forest (ERF)

- Estimate intermediate quantile $\hat{Q}_x(\tau_0)$
- Compute exceedances $Z_i = (Y_i - \hat{Q}_x(\tau_0))_+$ and weights $w_n(\mathbf{x}, \mathbf{X}_i)$
- Our extremal random forest uses negative GPD log-likelihood as loss $L(\theta, Z) = \ell_\theta(Z)$ and solves

$$(\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x})) = \arg\min_{\theta=(\sigma,\gamma)} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i)\ell_\theta(Z_i)$$

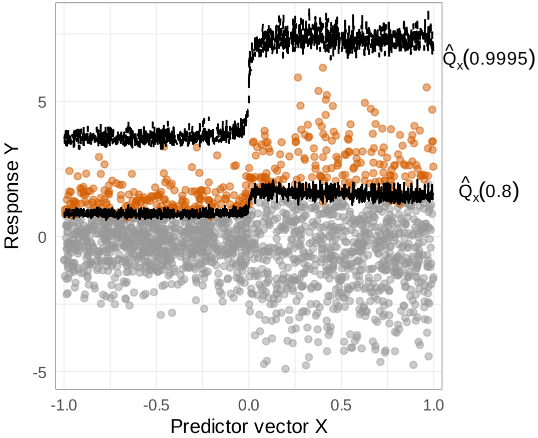



[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

# Extremal Random Forest (ERF)

- Estimate intermediate quantile $\hat{Q}_{\mathbf{x}}(\tau_0)$
- Compute exceedances $Z_i = (Y_i - \hat{Q}_{\mathbf{x}}(\tau_0))_+$ and weights $w_n(\mathbf{x}, \mathbf{X}_i)$
- Our extremal random forest uses negative GPD log-likelihood as loss $L(\theta, Z) = \ell_\theta(Z)$ and solves

$$(\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x})) = \underset{\theta=(\sigma,\gamma)}{\arg\min} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i)\ell_\theta(Z_i)$$




[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

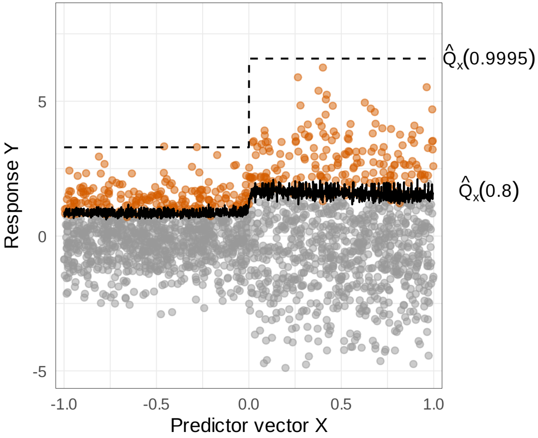
# Extremal Random Forest (ERF)

- Estimate intermediate quantile $\hat{Q}_{\mathbf{x}}(\tau_0)$
- Compute exceedances $Z_i = (Y_i - \hat{Q}_{\mathbf{x}}(\tau_0))_+$ and weights $w_n(\mathbf{x}, \mathbf{X}_i)$
- Our extremal random forest uses negative GPD log-likelihood as loss $L(\theta, Z) = \ell_\theta(Z)$ and solves

$$(\hat{\sigma}(\mathbf{x}), \hat{\gamma}(\mathbf{x})) = \underset{\theta=(\sigma,\gamma)}{\arg\min} \sum_{i=1}^{n} w_n(\mathbf{x}, \mathbf{X}_i)\ell_\theta(Z_i)$$

- Under some assumptions, ERF estimates are consistent

$$\hat{\theta}(\mathbf{x}) \xrightarrow{\mathbb{P}} \theta(\mathbf{x}), \quad \text{for all } \mathbf{x} \in [-1, 1]^p$$




[8] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

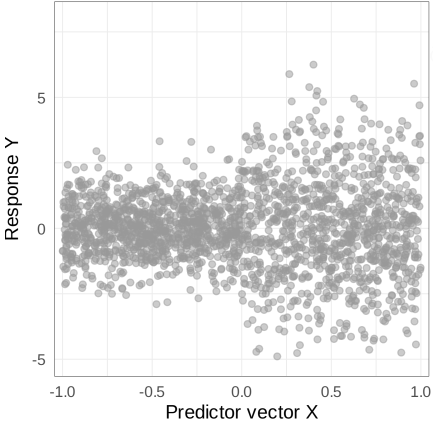
# Simulation study

- Sample $n = 2000$ iid copies of $(X, Y)$ from

$$\begin{cases} \mathbf{X} \sim U\left([-1, 1]^p\right), \\ (Y \mid \mathbf{X} = \mathbf{x}) \sim s(\mathbf{x})\, T_4, \end{cases}$$

  where $s(\mathbf{x}) = 1 + \mathbf{1}\{x_1 > 0\}$ and $\gamma(\mathbf{x}) = 1/4$.
- Compare ERF and GBEX with quantile regression and other extreme value methods
- On a test data set $\{\mathbf{x}_i\}_{i=1}^{n'}$, evaluate the integrated squared error (ISE)

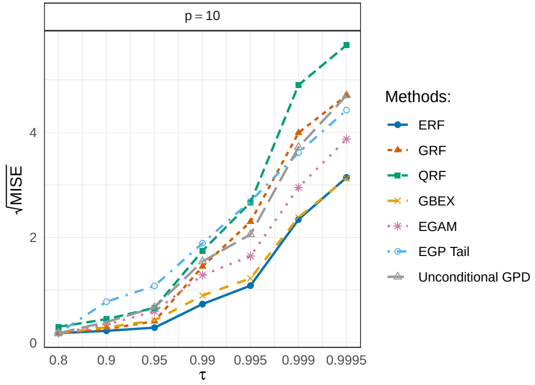
$$\text{ISE} = \frac{1}{n'} \sum_{i=1}^{n'} \left( \hat{Q}_{\mathbf{x}_i}(\tau) - Q_{\mathbf{x}_i}(\tau) \right)^2 .$$

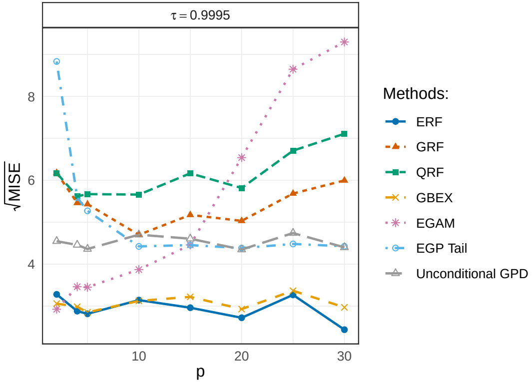



[9] N. Gnecco, E. M. Terefe, and S. Engelke. "Extremal random forests". In: *J. Amer. Statist. Assoc.* (2023). conditionally accepted.

[10] J. Velthoen, C. Dombry, J.-J. Cai, and S. Engelke. "Gradient boosting for extreme quantile regression". In: *Extremes* (2023). to appear.

# Simulation study results



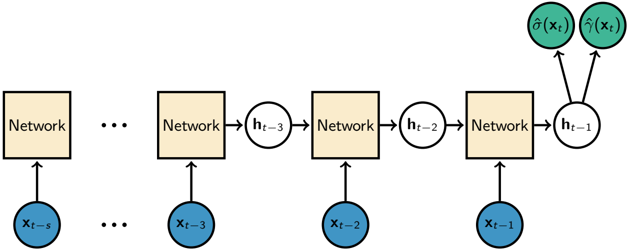



# Extreme quantile regression neural networks (EQRN)

- If there is sequential dependence as in time series, then this structure can be used in recurrent neural networks
- Predict quantiles of $Y_t$ (discharge at time $t$) using past observations

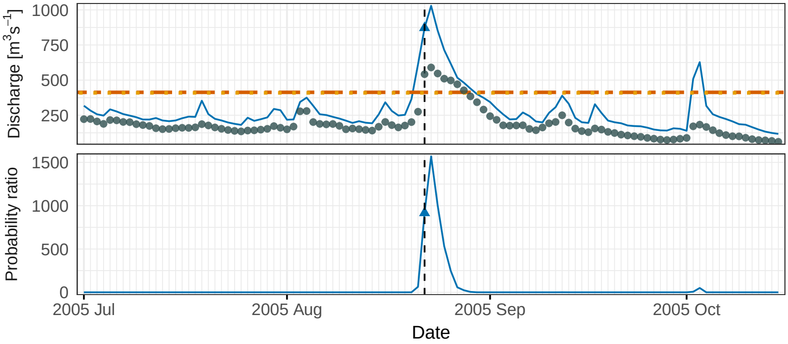
$$\mathbf{X} = (Y_{t-1}, Y_{t-2}, \dots, X^1_{t-1}, X^1_{t-2}, \dots)$$

from response on other covariates $X^1, X^2, \dots$ (e.g., precipitation at locations 1, 2, etc.)



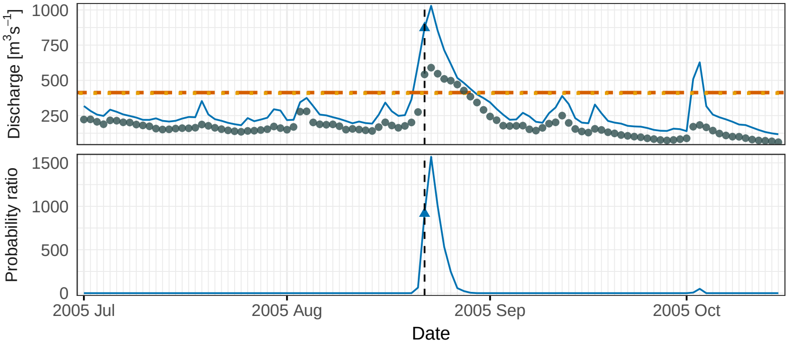

[11] O. C. Pasche and S. Engelke. *Neural networks for extreme quantile regression with an application to forecasting of flood risk*. Available on https://arxiv.org/abs/2208.07590. 2022.

# Results for the 2005 flood in Bern




– Top: One-day-ahead forecasted conditional 100-year return level $\hat{Q}_{\mathbf{x}}^{100}$ (blue line)

# Results for the 2005 flood in Bern




- Top: One-day-ahead forecasted conditional 100-year return level $\hat{Q}_{\mathbf{x}}^{100}$ (blue line)
- Bottom: Ratio of conditional exceedance probability compared to unconditional estimate

$$\frac{\hat{\mathbb{P}}(Y > \hat{Q}^{100} \mid \mathbf{X} = \mathbf{x})}{\hat{\mathbb{P}}(Y > \hat{Q}^{100})}$$

# evtGAN: combining EVT and GANs

# Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model [12].
- Present-day climate conditions, stationary in time; see [13] 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 $\mathbf{Z} = (Z_1, \ldots, Z_d)$.

[12] W. Hazeleger et al. "EC-Earth V2. 2: description and validation of a new seamless earth system prediction model". In: *Climate dynamics* 39.11 (2012), pp. 2611–2629.

[13] K. Van der Wiel, N. Wanders, F. Selten, and M. Bierkens. "Added value of large ensemble simulations for assessing extreme river discharge in a 2 C warmer world". In: *Geophysical Research Letters* 46.4 (2019), pp. 2093–2102.

[14] Y. Boulaguiem et al. "Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks". In: *Environmental Data Science* 1 (2022), e5.

# Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model [12].
- Present-day climate conditions, stationary in time; see [13] 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 $\mathbf{Z} = (Z_1, \ldots, Z_d)$.

**Goal:** Realistic simulations of $\mathbf{Z}$ for stress testing and extreme event simulation. [14]

[12] W. Hazeleger et al. "EC-Earth V2. 2: description and validation of a new seamless earth system prediction model". In: *Climate dynamics* 39.11 (2012), pp. 2611–2629.
[13] K. Van der Wiel, N. Wanders, F. Selten, and M. Bierkens. "Added value of large ensemble simulations for assessing extreme river discharge in a 2 C warmer world". In: *Geophysical Research Letters* 46.4 (2019), pp. 2093–2102.
[14] Y. Boulaguiem et al. "Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks". In: *Environmental Data Science* 1 (2022), e5.

# Climate model data

- We use 2000 years of large ensemble simulations with the EC-Earth global climate model [12].
- Present-day climate conditions, stationary in time; see [13] 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 $\mathbf{Z} = (Z_1, \ldots, Z_d)$.

**Goal:** Realistic simulations of $\mathbf{Z}$ for stress testing and extreme event simulation. [14]
Here we use $n_{\text{train}} = 50$ years for training, and $n_{\text{test}} = 1950$ for evaluation.


[12] W. Hazeleger et al. "EC-Earth V2. 2: description and validation of a new seamless earth system prediction model". In: *Climate dynamics* 39.11 (2012), pp. 2611–2629.
[13] K. Van der Wiel, N. Wanders, F. Selten, and M. Bierkens. "Added value of large ensemble simulations for assessing extreme river discharge in a 2 C warmer world". In: *Geophysical Research Letters* 46.4 (2019), pp. 2093–2102.
[14] Y. Boulaguiem et al. "Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks". In: *Environmental Data Science* 1 (2022), e5.

# Marginal GEV parameters

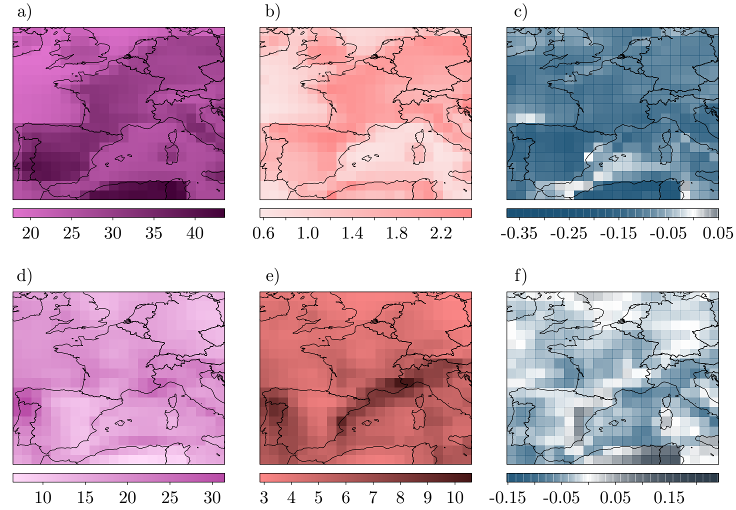



Figure: GEV parameters for temperature (a-c) and precipitation (d-f): mean parameter $\mu$ (left), scale parameter $\sigma$ (center) and shape parameter $\xi$ (right).

# Two approaches

**Classical EVT approach**

- Spatial extreme value theory provides statistical models for $\mathbf{Z}$, e.g., a spatial max-stable model [15].
- 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.

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

[16] I. Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

# Two approaches

**Classical EVT approach**

- Spatial extreme value theory provides statistical models for $\mathbf{Z}$, e.g., a spatial max-stable model [15].
- 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.

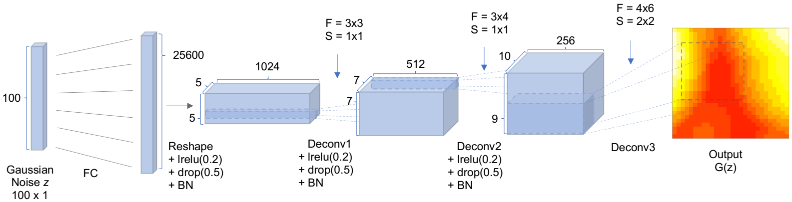
**Our ML approach**

- Generative Adversarial Networks (GANs) [16] are a flexible way of learning and sampling from a multivariate distribution $\mathbf{Z}$.
- They are usually used to sample from image data using convolutional neural networks.
- We can treat our spatial climatological data $\mathbf{Z}$ as an image.

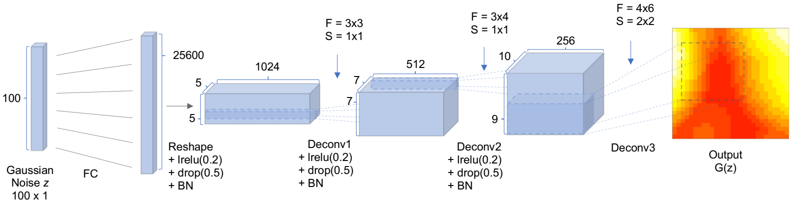
[15] A. C. Davison and M. M. Gholamrezaee. "Geostatistics of extremes". In: *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 468 (2012), pp. 581–608.

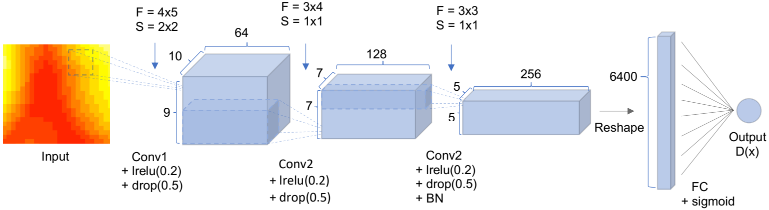
[16] I. Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

# 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.

[17] M. Wiese, R. Knobloch, and R. Korn. *Copula & Marginal Flows: Disentangling the Marginal from its Joint*. 2019.

[18] T. Huster et al. *Pareto GAN: Extending the Representational Power of GANs to Heavy-Tailed Distributions*. 2021.

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

[20] S. Bhatia, A. Jain, and B. Hooi. "ExGAN: Adversarial Generation of Extreme Samples". In: *arXiv preprint arXiv:2009.08454* (2020).

# 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 [17].
- [18] and [19] develop GANs that can generate heavy-tailed output.
- [20] propose a conditional GAN for importance sampling of extreme events.

[17] M. Wiese, R. Knobloch, and R. Korn. *Copula & Marginal Flows: Disentangling the Marginal from its Joint*. 2019.
[18] T. Huster et al. *Pareto GAN: Extending the Representational Power of GANs to Heavy-Tailed Distributions*. 2021.
[19] M. Allouche, S. Girard, and E. Gobet. "EV-GAN: Simulation of extreme events with ReLU neural networks". hal-03250663v2. 2021.
[20] S. Bhatia, A. Jain, and B. Hooi. "ExGAN: Adversarial Generation of Extreme Samples". In: *arXiv preprint arXiv:2009.08454* (2020).

# 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 [17].
- [18] and [19] develop GANs that can generate heavy-tailed output.
- [20] 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.

---


[17] M. Wiese, R. Knobloch, and R. Korn. *Copula & Marginal Flows: Disentangling the Marginal from its Joint*. 2019.
[18] T. Huster et al. *Pareto GAN: Extending the Representational Power of GANs to Heavy-Tailed Distributions*. 2021.
[19] M. Allouche, S. Girard, and E. Gobet. "EV-GAN: Simulation of extreme events with ReLU neural networks". hal-03250663v2. 2021.
[20] S. Bhatia, A. Jain, and B. Hooi. "ExGAN: Adversarial Generation of Extreme Samples". In: *arXiv preprint arXiv:2009.08454* (2020).

# 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 $j$th marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.

# 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 $j$th marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.
2. Normalize empirically to a std uniform distribution to obtain pseudo observations

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

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

# 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 $j$th marginal with parameters $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$.
2. Normalize empirically to a std uniform distribution to obtain pseudo observations

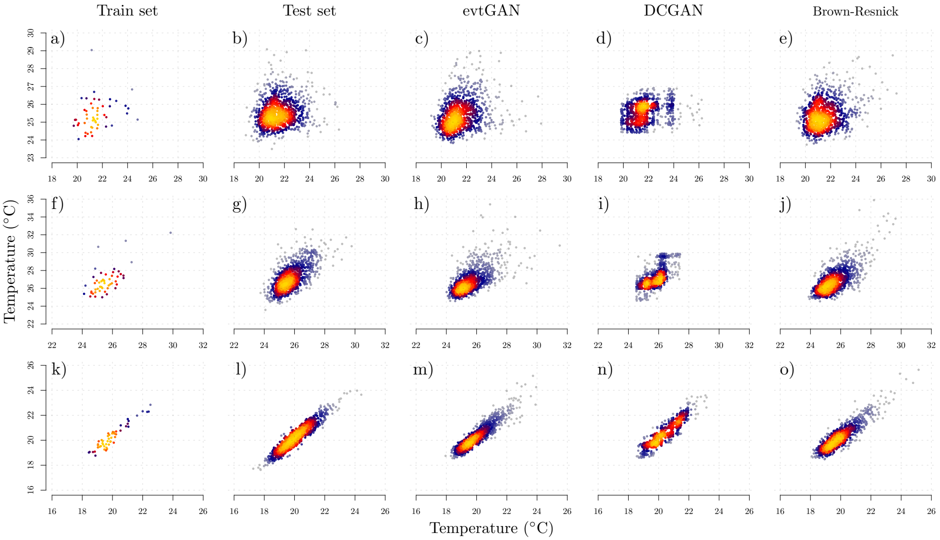
$$\mathbf{U}_i = (\widehat{F}_1(Z_{i1}), \dots, \widehat{F}_d(Z_{id})), \quad i = 1, \dots, n,$$

   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

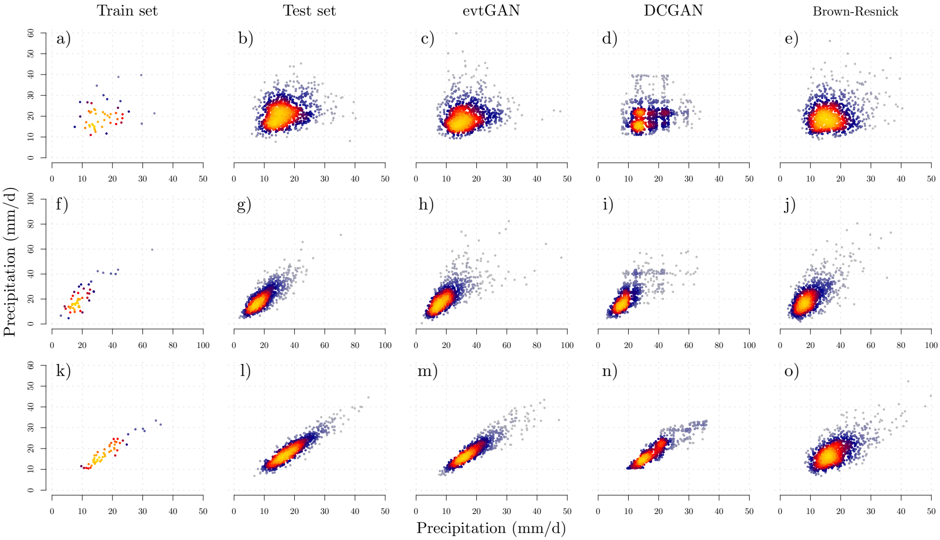
$$\mathbf{Z}_i^* = (\widehat{G}_1^{-1}(U_{i1}^*), \dots, \widehat{G}_d^{-1}(U_{id}^*)), \quad i = 1, \dots, n^*.$$

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

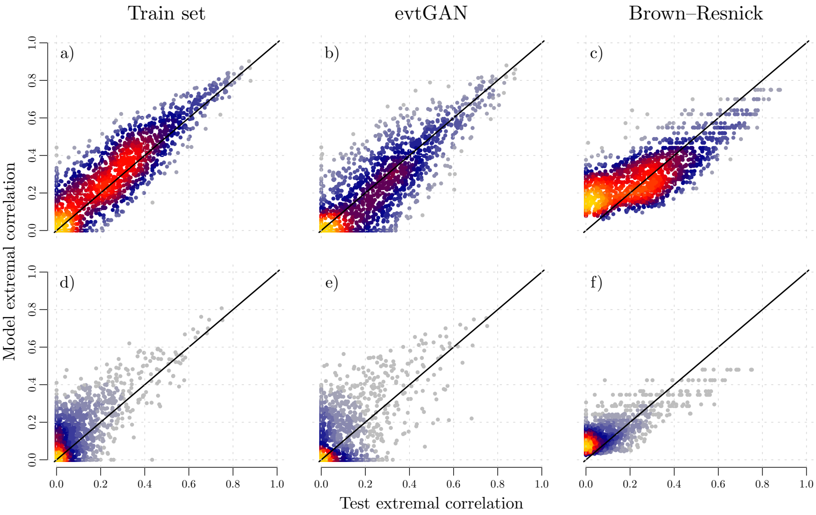
# Bivariate samples of temperature



# Bivariate samples of precipitation



# Extremal correlation plots



Thank you!