Machine learning beyond the data range: extreme quantile regression

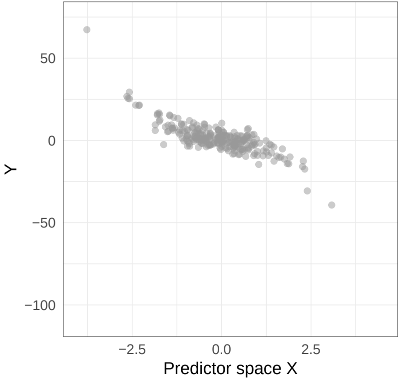
Sebastian Engelke (University of Geneva)
www.sengelke.com

Seminar in Mathematics, Physics & Machine Learning
Lisbon, January 12, 2023

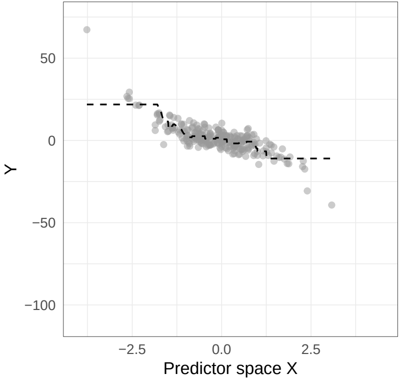
UNIVERSITÉ DE GENÈVE

FNSNF
Swiss National Science Foundation

Out-of-sample prediction: domain generalization

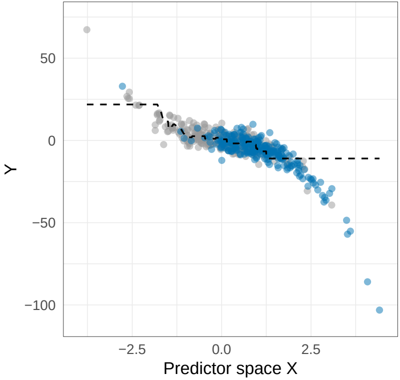


Out-of-sample prediction: domain generalization



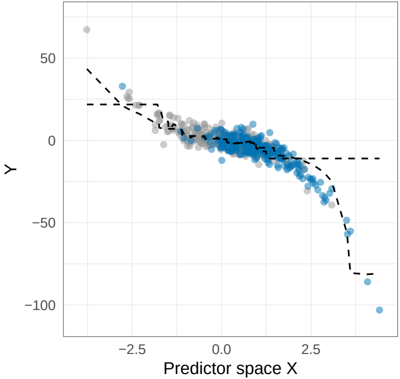
- Prediction of condition mean $\mathbb{E}[Y \mid X = x]$
- Machine learning methods typically perform poorly outside of the range of training distribution

Out-of-sample prediction: domain generalization



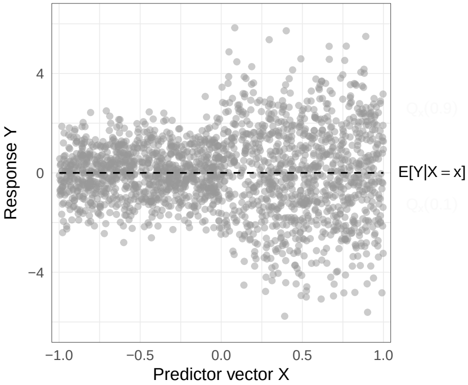
- Prediction of condition mean $\mathbb{E}[Y \mid X = x]$
- Machine learning methods typically perform poorly outside of the range of training distribution
- If test distribution is different, domain generalization is needed

Out-of-sample prediction: domain generalization

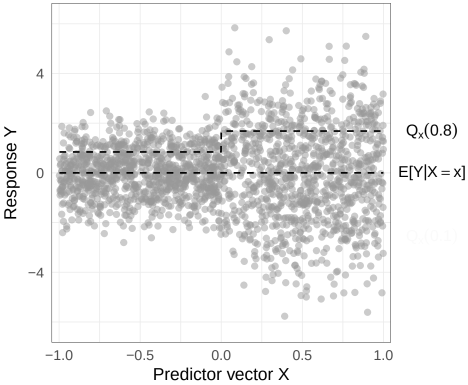


- Prediction of condition mean $\mathbb{E}[Y \mid X = x]$
- Machine learning methods typically perform poorly outside of the range of training distribution
- If test distribution is different, domain generalization is needed
- We work on methods based on extreme value theory

Out-of-sample prediction: extreme quantile regression

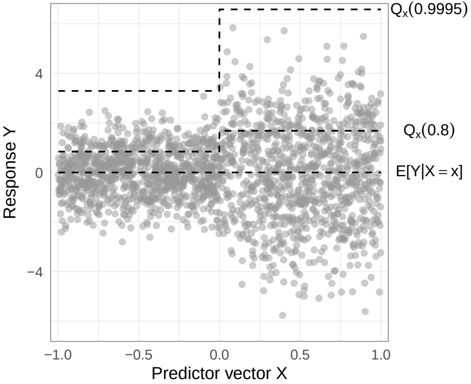


Out-of-sample prediction: extreme quantile regression



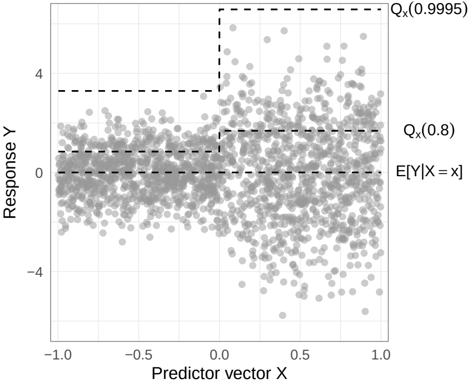
- Prediction of quantile $Q_x(\tau)$ at level $\tau \in (0,1)$

Out-of-sample prediction: extreme quantile regression



- Prediction of quantile $Q_x(\tau)$ at level $\tau \in (0, 1)$
- If τ is close to 1, we speak of extreme quantile regression

Out-of-sample prediction: extreme quantile regression



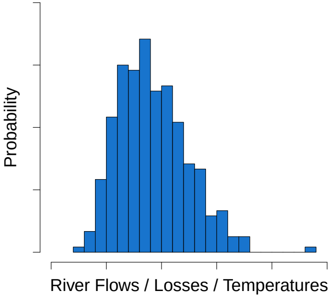
- Prediction of quantile $Q_x(\tau)$ at level $\tau \in (0, 1)$
- If τ is close to 1, we speak of extreme quantile regression
- Again, classical machine learning methods perform poorly

Extreme Value Theory and Statistics

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

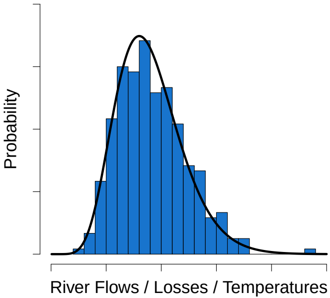
Extreme Value Theory and Statistics

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

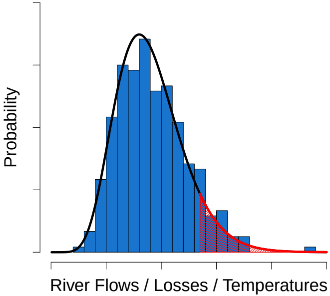


Extreme Value Theory and Statistics

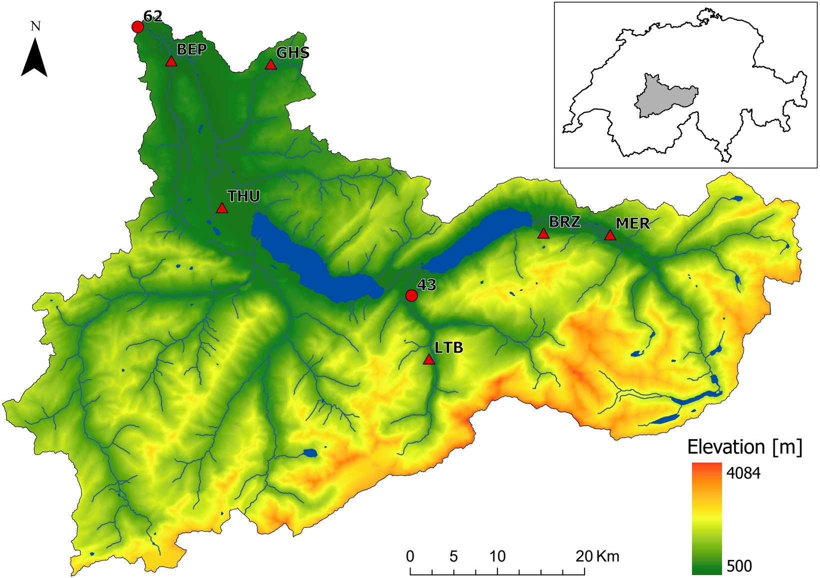
- Analysis of rare phenomena with small probabilities
- Impact on various risks (health, environment, economy,...)



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



Aare River in Bern: Catchment



Data

- Daily observations of average discharge at 2 stations and total precipitation at 6 stations.
- Training and validation data: 1930–1958 ($10,349$ obs.).
- Test data: 1958–2014 ($20,697$ obs.).

Aare River in Bern: Data

Data

- Daily observations of average discharge at 2 stations and total precipitation at 6 stations.
- Training and validation data: 1930–1958 ($10,349$ obs.).
- Test data: 1958–2014 ($20,697$ obs.).

2005 Flood

- Hydrological models underestimated the event, which led to late warnings.
- 3 billion CHF in damages, 6 deaths and countless injuries were reported.

Aare River in Bern: Data

Data

- Daily observations of average discharge at 2 stations and total precipitation at 6 stations.
- Training and validation data: 1930–1958 ($10,349$ obs.).
- Test data: 1958–2014 ($20,697$ obs.).

2005 Flood

- Hydrological models underestimated the event, which led to late warnings.
- 3 billion CHF in damages, 6 deaths and countless injuries were reported.

As regression problem

- Y is daily discharge at Bern station.
- $\mathbf{X}$ can contain: discharge from previous days at same and other stations; precipitation from close-by stations; other climatological variables.

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^p$ 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 $\mathbf{X}_i \in \mathbb{R}^p$ 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 $n(1 - \tau_n) \to c \in [0, \infty)$ (extreme case)

Extreme quantile regression

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

Extreme quantile regression

- For i.i.d. data $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ where $\mathbf{X}_i \in \mathbb{R}^p$ 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 $n(1-\tau_n) \to c \in [0, \infty)$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$.
- Methods from extreme value theory are not flexible enough [Chernozhukov, 2005, Chavez-Demoulin and Davison, 2005] or do not generalize well into higher dimensions [Daouia et al., 2011].

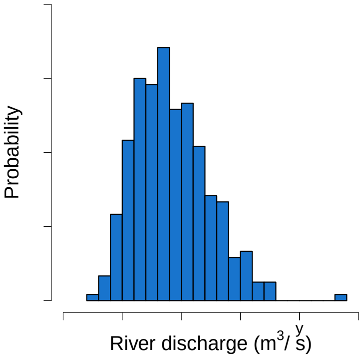
Extreme quantile regression

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

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

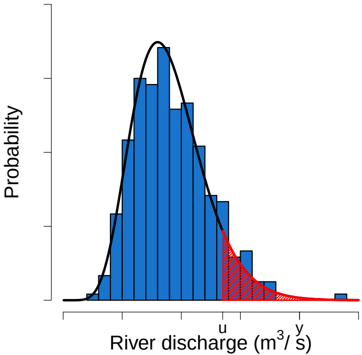
- There are different scenarios depending on the quantile level $\tau = \tau_n$:
  - $\tau_n \equiv \tau_0 < 1$ (classical case)
  - $\tau_n \to 1$, and $n(1-\tau_n) \to \infty$ (intermediate case)
  - $\tau_n \to 1$, and $n(1-\tau_n) \to c \in [0, \infty)$ (extreme case)
- Classical methods for quantile regression only work well in the case of fixed $\tau_n \equiv \tau_0 < 1$.
- Methods from extreme value theory are not flexible enough [Chernozhukov, 2005, Chavez-Demoulin and Davison, 2005] or do not generalize well into higher dimensions [Daouia et al., 2011].
- Goal: Develop a new method for extreme quantile regression that works well with high-dimensional and complex data.

Generalized Pareto distribution



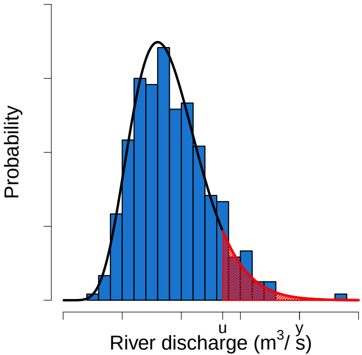
$\mathbb{P}(Y > y)$

Generalized Pareto distribution



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

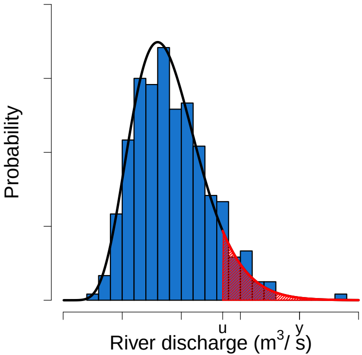
Generalized Pareto distribution



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

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

Generalized Pareto distribution



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

where $H_{\sigma,\gamma}$ is the cdf of the GPD with scale 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

$$\hat{\theta} = \arg\max_{\theta}\sum_{i=1}^{n}\ell_{Z_i}(\theta).$$

Extreme quantile estimation

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

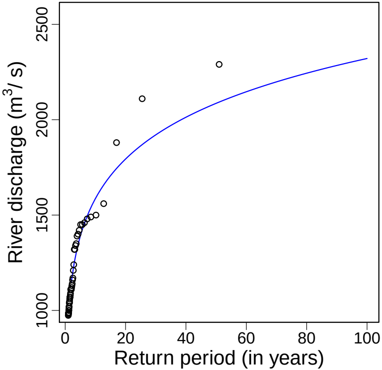
$$\hat{Q}(\tau) = \hat{Q}(\tau_0) + \hat{\sigma}\frac{\left(\frac{1-\tau}{1-\tau_0}\right)^{-\hat{\gamma}} - 1}{\hat{\gamma}}.$$

Extreme quantile estimation

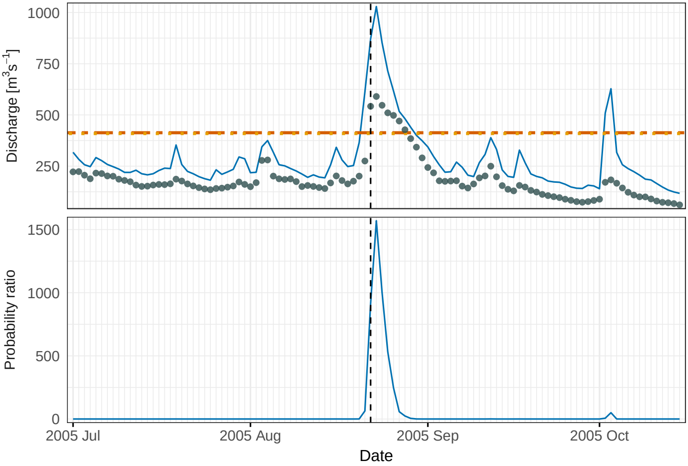
- Inverting the cdf $H_{\hat{\sigma},\hat{\gamma}}$ of the GPD provides an approximation of the quantile $Q(\tau) = F_Y^{-1}(\tau)$ for probability level $\tau > \tau_0$ by

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

- For independent data, we can compute a T-year event by $\hat{Q}(1 - 1/(n_Y T))$, where n_Y is the number of observations per year.



During the 2005 flood



- Top: Daily observations during the 2005 flood in Bern together with 100-year return level estimate. Vertical dashed line is the first exceedance of this level.

Forest-based quantile regression

- For $\tau \in (0,1)$ and $x \in [-1,1]^p$, the quantile regression function is defined as

  $$Q_{\boldsymbol{x}}(\tau) := \arg\min_q \mathbb{E}[\rho_\tau(Y-q) \mid \boldsymbol{X} = \boldsymbol{x}],$$

  where $\rho_\tau(c) = c(\tau - \mathbb{1}\{c<0\})$ is the quantile loss function.

Forest-based quantile regression

- For $\tau \in (0,1)$ and $x \in [-1,1]^p$, the quantile regression function is defined as

$$Q_{\boldsymbol{x}}(\tau) := \arg\min_q \mathbb{E}[\rho_\tau(Y - q) \mid \boldsymbol{X} = \boldsymbol{x}],$$

  where $\rho_\tau(c) = c(\tau - \mathbb{1}\{c < 0\})$ is the quantile loss function.
- Recently, [Meinshausen, 2006] and [Athey et al., 2019] proposed to estimate $Q_{\boldsymbol{x}}(\tau)$ by

$$\hat{Q}_{\boldsymbol{x}}(\tau) = \arg\min_q \sum_{i=1}^n w_n(\boldsymbol{x}, X_i)\rho_\tau(Y_i - q),$$

  where $(x, y) \mapsto w_n(\boldsymbol{x}, y)$ is a localizing weight function learned with a random forest.

Extreme quantile regression

Given: Independent data $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ of response $Y \in \mathbb{R}$ and covariates/predictors vector $\mathbf{X} \in \mathbb{R}^d$.

Goal:

- Predict extreme conditional quantiles of Y given $\mathbf{X} = \mathbf{x}$:

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

Extreme quantile regression

Given: Independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of response $Y \in \mathbb{R}$ and covariates/predictors vector $\mathbf{X} \in \mathbb{R}^d$.

Goal:

- Predict extreme conditional quantiles of Y given $\mathbf{X} = \mathbf{x}$:

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

- Predict exceedance probability of Y over high threshold Q (e.g., 100-year level) given $\mathbf{X} = \mathbf{x}$:

$$\mathbb{P}(Y > Q \mid \mathbf{X} = \mathbf{x}).$$

Extreme quantile regression

Given: Independent data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ of response $Y \in \mathbb{R}$ and covariates/predictors vector $\mathbf{X} \in \mathbb{R}^d$.

Goal:

- Predict extreme conditional quantiles of Y given $\mathbf{X} = \mathbf{x}$:

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

- Predict exceedance probability of Y over high threshold Q (e.g., 100-year level) given $\mathbf{X} = \mathbf{x}$:

$$\mathbb{P}(Y > Q \mid \mathbf{X} = \mathbf{x}).$$

Examples:

- Bern river discharge:
  - Y is daily discharge at Bern station.
  - $\mathbf{X}$ can contain: discharge from previous days at same and other stations; precipitation from close-by stations; other climatological variables.
- Risk of heat waves:
  - Y daily temperature measurement at some location.
  - $\mathbf{X}$ can contain: altitude; day of the year; other variables on land use, climate, etc.

Extreme quantile regression

- Assume the GPD model

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

where τ_0 is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional τ_0 quantile of $Y \mid \mathbf{X} = \mathbf{x}$

Extreme quantile regression

- Assume the conditional GPD model

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

where τ_0 is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional τ_0 quantile of $Y \mid \mathbf{X} = \mathbf{x}$

Extreme quantile regression

- Assume the conditional GPD model

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

  where τ_0 is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional τ_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

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

  where τ_0 is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional τ_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

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

  where τ_0 is an intermediate quantile level, and $\hat{Q}_{\mathbf{x}}(\tau_0)$ is an estimate of the conditional τ_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.
- 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

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

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i)$$

$\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with exceedances $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$.

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:

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i)$$

$\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with exceedances $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$.

- **Extremal gradient boosting (GBEX)**: The weights $w_n(\mathbf{x}, X_i)$ 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:

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i)$$

$\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with exceedances $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$.

- **Extremal gradient boosting (GBEX)**: The weights $w_n(\mathbf{x}, X_i)$ 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 $w_n(\mathbf{x}, X_i)$ are obtained through a GRF

  Gnecco, N., Terefe, E.M., and Engelke, S. (2022).
  Extremal Random Forests.
  `https://arxiv.org/abs/2201.12865`

Extremal Random Forest (ERF)

- $(\hat{\sigma}(\boldsymbol{x}), \hat{\gamma}(\boldsymbol{x}))$ is the maximizer of the weighted log-likelihood

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i).$$

Extremal Random Forest (ERF)

- $(\hat{\sigma}(\boldsymbol{x}), \hat{\gamma}(\boldsymbol{x}))$ is the maximizer of the weighted log-likelihood

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i).$$

- $w_n(\boldsymbol{x}, X_i)$ are localizing weights from a GRF.

Extremal Random Forest (ERF)

- $(\hat{\sigma}(\boldsymbol{x}), \hat{\gamma}(\boldsymbol{x}))$ is the maximizer of the weighted log-likelihood

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i).$$

- $w_n(\boldsymbol{x}, X_i)$ are localizing weights from a GRF.
- $\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$, $i = 1, \ldots, n$.

Extremal Random Forest (ERF)

- $(\hat{\sigma}(\boldsymbol{x}), \hat{\gamma}(\boldsymbol{x}))$ is the maximizer of the weighted log-likelihood

$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_{(\sigma,\gamma)}(Z_i).$$

- $w_n(\boldsymbol{x}, X_i)$ are localizing weights from a GRF.
- $\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$, $i = 1, \ldots, n$.

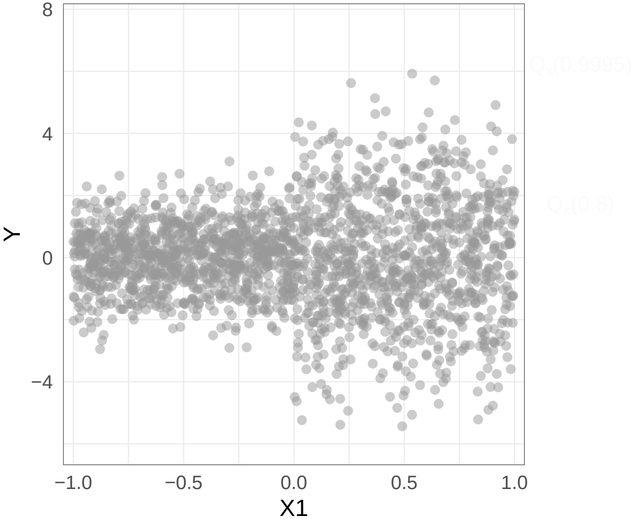
Extremal Random Forest (ERF)

- $(\hat{\sigma}(\boldsymbol{x}), \hat{\gamma}(\boldsymbol{x}))$ is the maximizer of the weighted log-likelihood

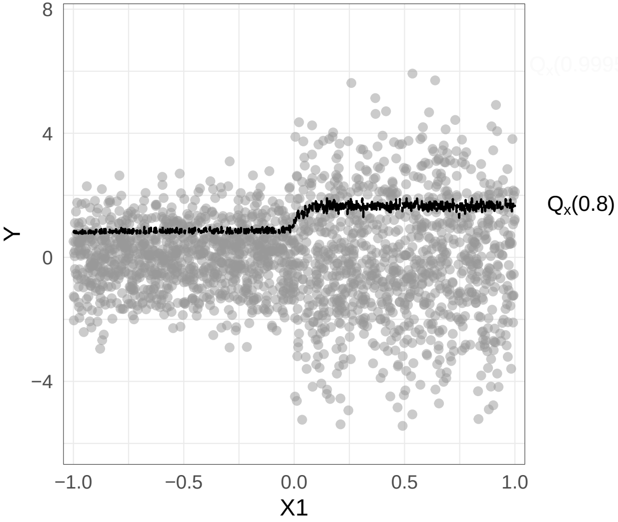
$$\sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i) \ell_{(\sigma,\gamma)}(Z_i).$$

- $w_n(\boldsymbol{x}, X_i)$ are localizing weights from a GRF.
- $\ell_{(\sigma,\gamma)}(Z_i)$ is the GPD log-likelihood, with $Z_i = (Y_i - \hat{Q}_{\boldsymbol{x}}(\tau_0))_+$, $i = 1, \ldots, n$.

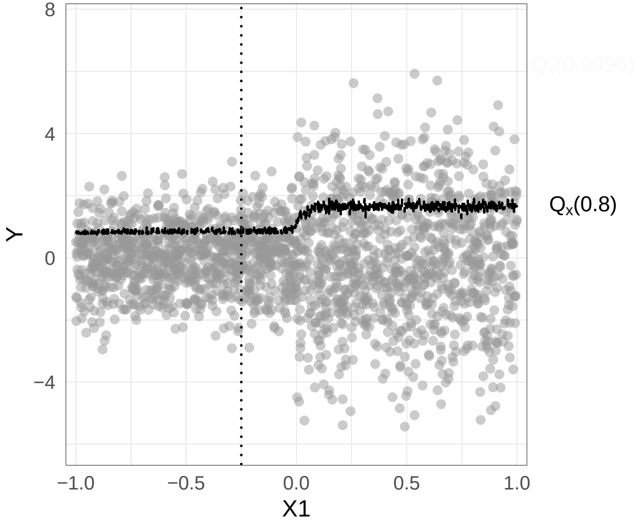
Extremal Random Forest (ERF)



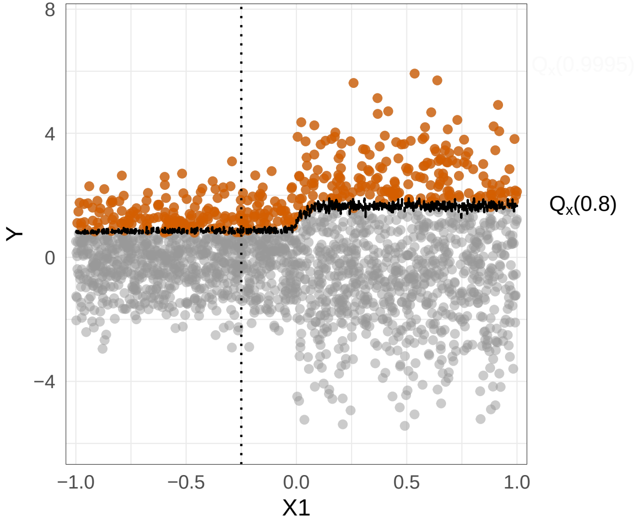
Extremal Random Forest (ERF)



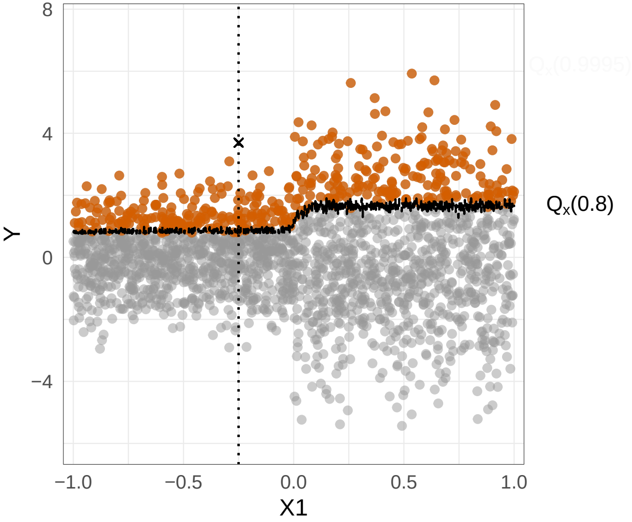
Extremal Random Forest (ERF)



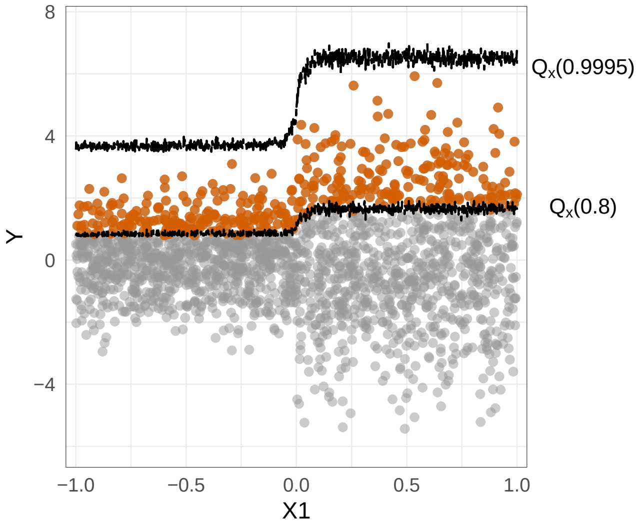
Extremal Random Forest (ERF)



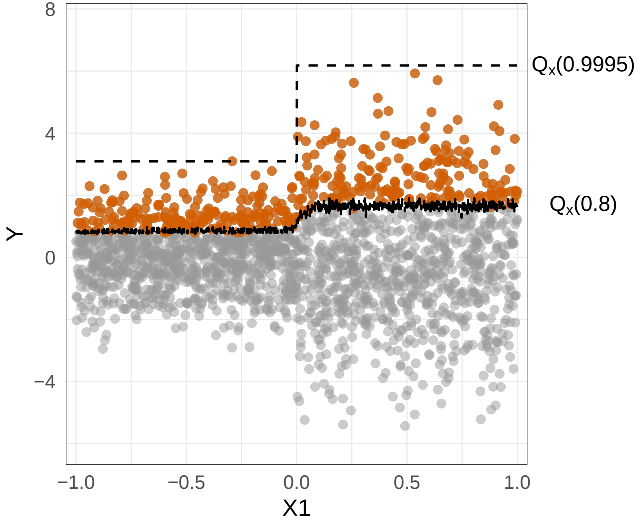
Extremal Random Forest (ERF)



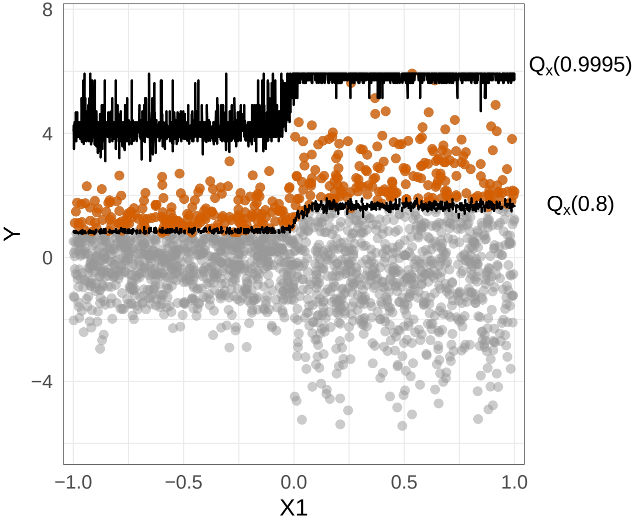
Extremal Random Forest (ERF)



Extremal Random Forest (ERF)



Extremal Random Forest (ERF)



Consistency

- Let $x \in [-1,1]^p$ and $\theta := (\sigma, \gamma)$. Want to show that

$$\hat{\theta}(x) \in \arg\max_{\theta} \sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_\theta(Z_i) \xrightarrow{\mathbb{P}} \theta(x).$$

Consistency

- Let $x \in [-1,1]^p$ and $\theta := (\sigma, \gamma)$. Want to show that

$$\hat{\theta}(x) \in \arg\max_{\theta} \sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_\theta(Z_i) \xrightarrow{\mathbb{P}} \theta(x).$$

- Regularity conditions from [Athey et al., 2019] do not hold in our setting.

Consistency

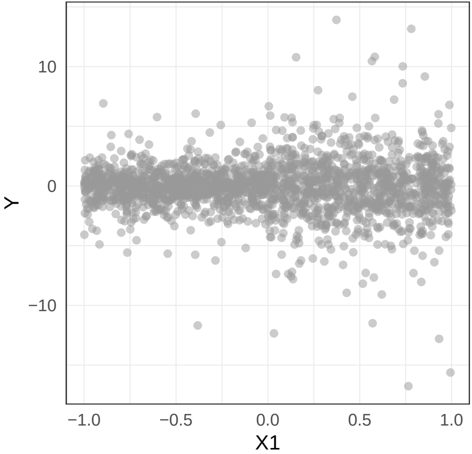
- Let $x \in [-1, 1]^p$ and $\theta := (\sigma, \gamma)$. Want to show that

$$\hat{\theta}(x) \in \arg\max_{\theta} \sum_{i=1}^{n} w_n(\boldsymbol{x}, X_i)\ell_\theta(Z_i) \xrightarrow{\mathbb{P}} \theta(x).$$

- Regularity conditions from [Athey et al., 2019] do not hold in our setting.
- Under some assumptions, ERF estimates are consistent

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

Simulation Study I



Simulation Study I

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

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

where $s(x) = 1 + \mathbb{1}\{x_1 > 0\}$ and $\gamma(x) = 1/4$.

Simulation Study I

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

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

  where $s(x) = 1 + \mathbb{1}\{x_1 > 0\}$ and $\gamma(x) = 1/4$.
- Compare ERF and GBEX with QRF [Meinshausen, 2006], GRF [Athey et al., 2019] Extreme GAM [Youngman, 2019], [Taillardat et al., 2019].

Simulation Study I

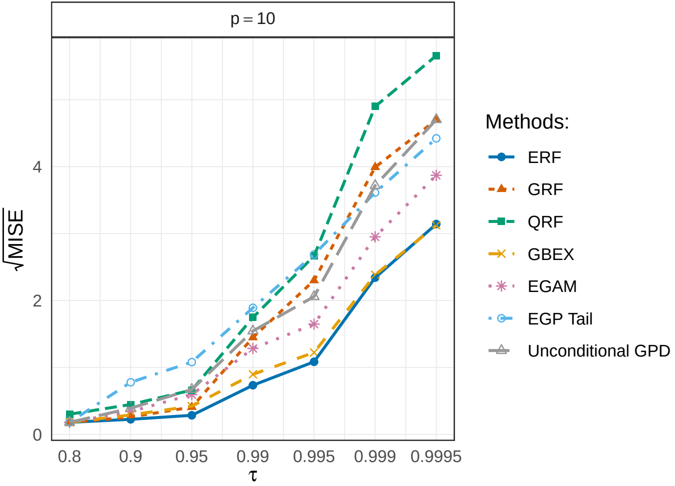
- Sample $n = 2000$ iid copies of (X, Y) from

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

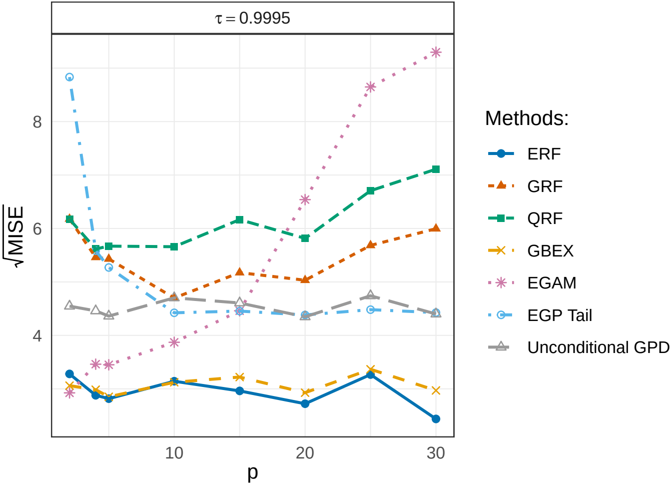
  where $s(x) = 1 + \mathbb{1}\{x_1 > 0\}$ and $\gamma(x) = 1/4$.
- Compare ERF and GBEX with QRF [Meinshausen, 2006], GRF [Athey et al., 2019] Extreme GAM [Youngman, 2019], [Taillardat et al., 2019].
- On a test data set $\{\boldsymbol{x}_i\}_{i=1}^{n'}$, evaluate the integrated squared error (ISE)

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

Simulation Study I



Simulation Study I



Simulation Study II

- Sample $n = 5000$ iid copies of (X, Y) from

$$\begin{cases} \boldsymbol{X} \sim U\left([-1,1]^p\right), \\ (Y \mid \boldsymbol{X} = \boldsymbol{x}) \sim s_j(\boldsymbol{x}) T_{\nu(\boldsymbol{x})}, \quad j = 1, 2, 3. \end{cases}$$

Simulation Study II

- Sample $n = 5000$ iid copies of (X, Y) from

$$\begin{cases} \boldsymbol{X} \sim U\left([-1,1]^p\right), \\ (Y \mid \boldsymbol{X} = \boldsymbol{x}) \sim s_j(\boldsymbol{x}) T_{\nu(\boldsymbol{x})}, \quad j = 1, 2, 3. \end{cases}$$

- $\nu(x) := 3 + 3[1 + \tanh(-2X_1)]$ where $\gamma(x) = 1/\nu(x)$.

Simulation Study II

- Sample $n = 5000$ iid copies of (X, Y) from

$$\begin{cases} \boldsymbol{X} \sim U\left([-1,1]^p\right), \\ (Y \mid \boldsymbol{X} = \boldsymbol{x}) \sim s_j(\boldsymbol{x}) T_{\nu(\boldsymbol{x})}, \quad j = 1, 2, 3. \end{cases}$$

- $\nu(x) := 3 + 3[1 + \tanh(-2X_1)]$ where $\gamma(x) = 1/\nu(x)$.
- $s_1(x) := (2 + \tanh(2X_1)) \cdot (1 + \frac{1}{2}X_2)$.

Simulation Study II

- Sample $n = 5000$ iid copies of (X, Y) from

$$\begin{cases} \boldsymbol{X} \sim U\left([-1,1]^p\right), \\ (Y \mid \boldsymbol{X} = \boldsymbol{x}) \sim s_j(\boldsymbol{x}) T_{\nu(\boldsymbol{x})}, \quad j = 1, 2, 3. \end{cases}$$

- $\nu(x) := 3 + 3[1 + \tanh(-2X_1)]$ where $\gamma(x) = 1/\nu(x)$.
- $s_1(x) := (2 + \tanh(2X_1)) \cdot (1 + \frac{1}{2}X_2)$.
- $s_2(x) := 4 - (X_1^2 + 2X_2^2)$.

Simulation Study II

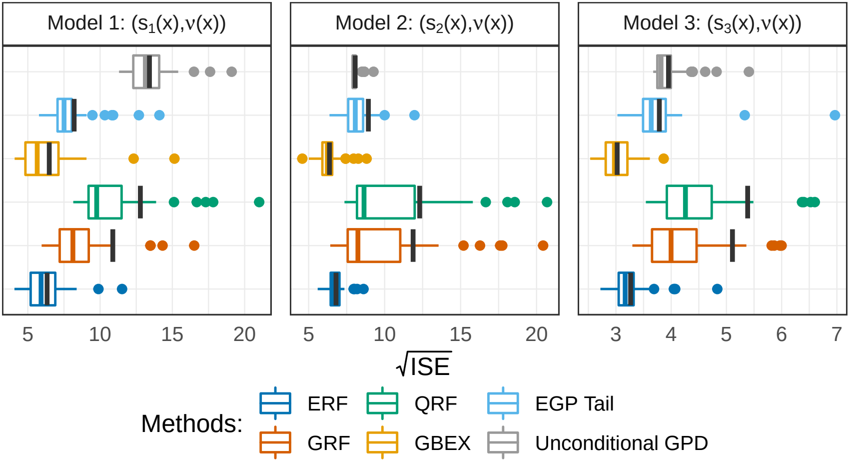
- Sample $n = 5000$ iid copies of (X, Y) from

$$\begin{cases} \boldsymbol{X} \sim U\left([-1,1]^p\right), \\ (Y \mid \boldsymbol{X} = \boldsymbol{x}) \sim s_j(\boldsymbol{x}) T_{\nu(\boldsymbol{x})}, \quad j = 1,2,3. \end{cases}$$

- $\nu(x) := 3 + 3[1 + \tanh(-2X_1)]$ where $\gamma(x) = 1/\nu(x)$.
- $s_1(x) := (2 + \tanh(2X_1)) \cdot (1 + \frac{1}{2}X_2)$.
- $s_2(x) := 4 - (X_1^2 + 2X_2^2)$.
- $s_3(x) := 1 + 2\pi\phi(2X_1, 2X_2)$, where $\phi(X_1, X_2)$ is a centered bivariate Gaussian with unit variance and correlation equal to $3/4$.

Simulation Study II

$n = 5000,\ p = 10,\ \tau = 0.9995$



Extreme quantile regression

- Suppose the data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ are NOT i.i.d., but have a time series structure
- Then we can use recurrent neural networks to model the GPD parameters $(\sigma(\boldsymbol{x}), \gamma(\boldsymbol{x}))$
- **Extreme quantile regression neural networks (EQRN)**:

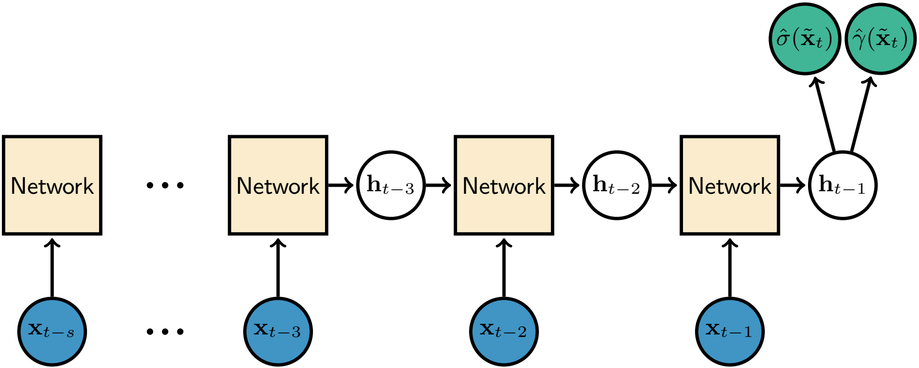
Pasche, O.C. and Engelke, S. (2022).
Neural Networks for Extreme Quantile Regression with an Application to Forecasting of Flood Risk.
https://arxiv.org/abs/2208.07590

Extreme quantile regression neural networks (EQRN)

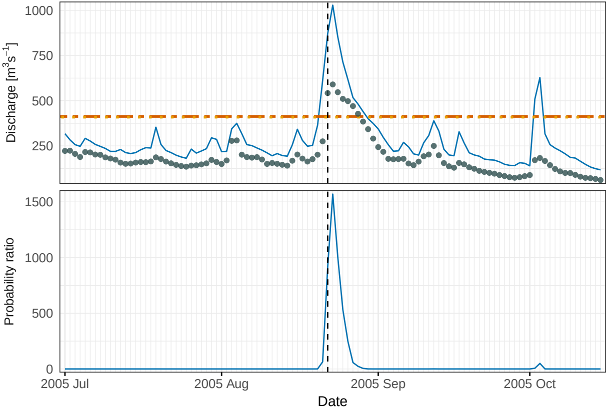
- If there is sequential dependence as in time series, then this structure can be used in recurrent neural networks.
- Predict quantiles of Y_t (discharge at time t) using past observations

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

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

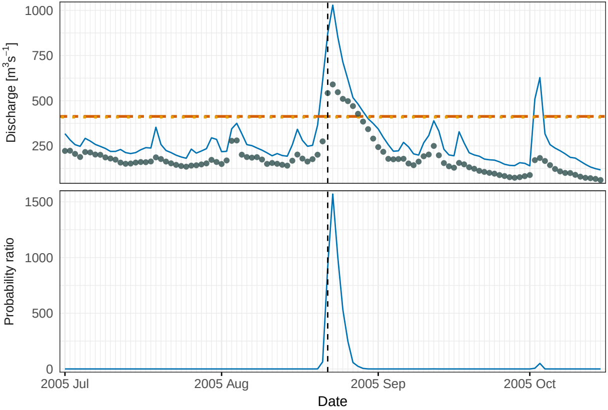


Results for the 2005 flood



- Top: Blue line is the one-day-ahead forecasted (conditional) 100-year return level $\hat{Q}_{\mathbf{x}}^{100}$.

Results for the 2005 flood



- Top: Blue line is the one-day-ahead forecasted (conditional) 100-year return level $\hat{Q}^{100}_{\mathbf{x}}$.
- Bottom: Blue line the the ratio of conditional exceedance probability compared to unconditional estimate

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

References I

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

Chavez-Demoulin, V. and Davison, A. C. (2005).
Generalized additive modelling of sample extremes.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(1):207–222.

Chernozhukov, V. (2005).
Extremal quantile regression.
The Annals of Statistics, 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.

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

References II

Taillardat, M., Fougères, A.-L., Naveau, P., and Mestre, O. (2019).
Forest-based and semiparametric methods for the postprocessing of rainfall ensemble forecasting.
Weather and Forecasting, 34(3):617 – 634.

Youngman, B. D. (2019).
Generalized additive models for exceedances of high thresholds with an application to return level estimation for u.s. wind gusts.
Journal of the American Statistical Association, 114(528):1865–1879.