Extreme weather, probabilistic forecast approaches and statistical downscaling of extremes

Petra Friederichs
Sabrina Bentzien, Andreas Hense

Meteorological Institute, University of Bonn

ESF-COST High-Level Research Conference on Extreme Environmental Events
Cambride, UK, 13 – 17 December 2010

Outline

Extreme weather and probabilistic prediction

- Atmospheric scales and mesoscale atmospheric dynamics
- Numerical weather prediction
- Forecast uncertainty and ensemble prediction systems

Statistical downscaling of extremes

- Downscaling and post-processing: Extract and calibrate information
- Verification

Results

What are extemes?

- **Mathematically:**

  Defined as block maxima or excedances of large thresholds.

  Events that lie in the tails of a distribution

- **Perseption:**

  Rare, exceptional, "large" and **high impact**

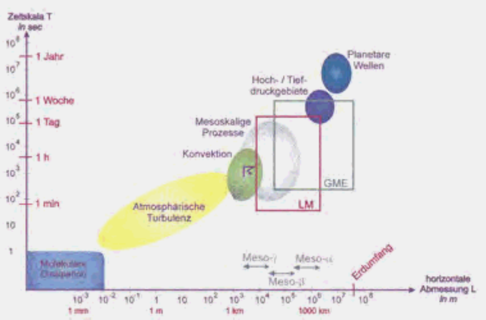
- **Problems:**
  - 95% quantile of daily precipitation: $\approx 10 - 15mm/d$
  - $\approx$ 2-5 years of data – only few extremes events for verification

Mesoscale Weather Prediction

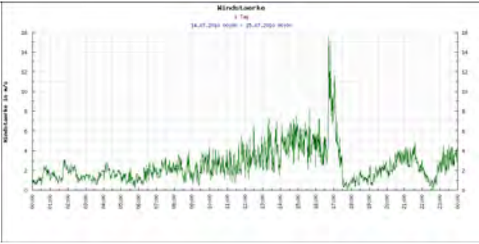
- Strong and disastrous impact of many weather extremes calls for reliable forecasts
- "Although forecasters have traditionally viewed weather prediction as deterministic, a cultural change towards probabilistic forecasting is in progress." (T N Palmer, 2002)
- Weather extremes do not come **"Out of the Blue"**
- Numerical weather forecast models provide reliable forecasts of the atmospheric circulation prone to generate extremes
- Combination of dynamical and statistical analysis methods

Atmospheric scales and mesoscale dynamics

- Different scales exhibit different dominant force balances, different wave dynamics
- Mesoscale on horizontal scales 2km – 2000km
- Complex force balances

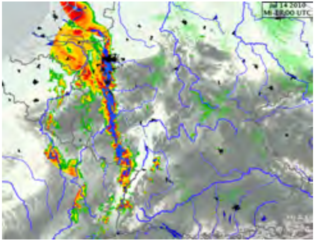


Steinhorst, Promet 35, 2010



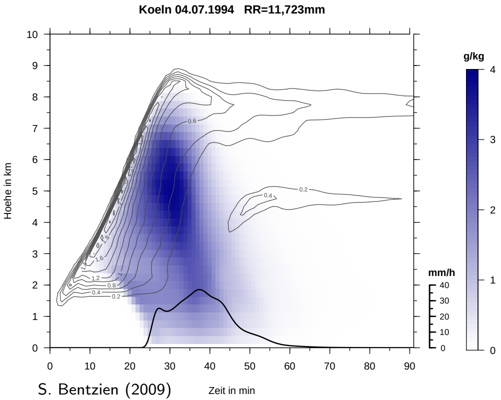
Mesoscale weather extremes

- Heavy thunderstorms on July 14, 2010
- Strong horizontal gradients
- Strong vertical mixing
- Embedded in larger scale squall line – embedded in synoptic situation



Connection of Extremes on Different Scales

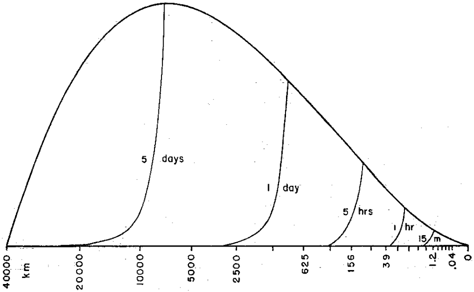
- Large vertical gradients of entropy
- Convective instability
- Deep convection lead to extremal vertical velocities
- Heavy precipitation and hailstones grow within this vertical circulation



S. Bentzien (2009)

Predictability

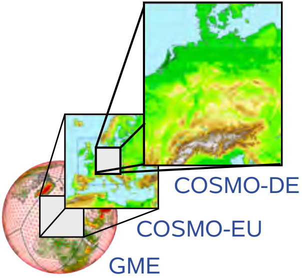
- Inherent limit of predictability
- Fastes error growth at smallest scales
- Predictability strongly depends on flow regime
- Moist convection is primary source of forecast-error growth
- Mesoscale forecasts are issued for ≤18h-24h

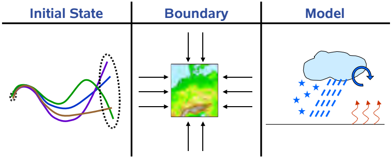


E N Lorenz, Tellus 21, 19 (1969)

COSMO-DE Ensemble Prediction System (EPS)

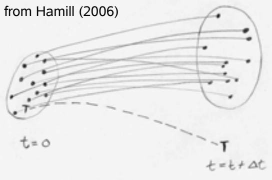
- COSMO-DE: 2.8 km grid spacing, convection resolving NWP model
- Operational forecasts 0-21 hours – high-impact weather by DWD
- EPS with 20 (40) members
- Uncertainty due to initial conditions, boundary conditions, and model parameterisation errors
- First EPS with convection resolving limited area NWP model





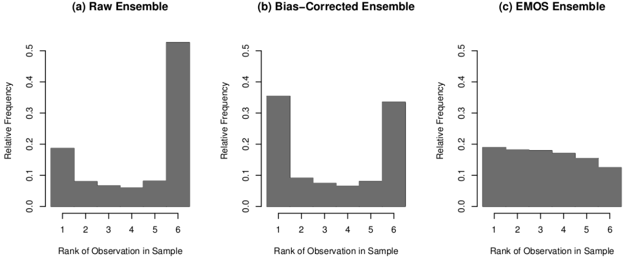
S. Theis, DWD (2010)

Probabilistic forecasting: Maximize sharpness of the predictive distribution subject to calibration



Calibration:

Raw ensemble data need adjustmend: biased and underdispersive

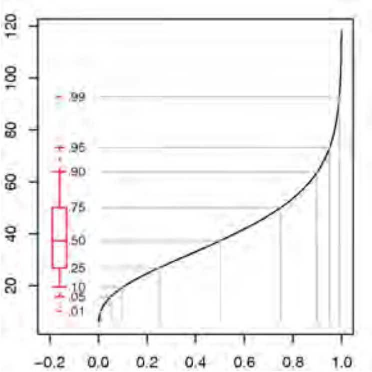


Gneiting et al. (2005)

Statistical downscaling for Extremes

- Global models do not resolve dynamics of many extremes
- More confidence in large scale flow patterns
- Find connection between local extreme and large scale flow
- Climate and weather prediction (early warning)
- Combination of dynamical and statistical analysis methods

Conditional quantile function



Semi-parametric

- A-priori probability τ, estimate conditional quantile $F^{-1}_{Y|\mathbf{X}}(\tau|\mathbf{x}) = \boldsymbol{\beta}_\tau^T \mathbf{x}$ via (linear) quantile regression

Parametric

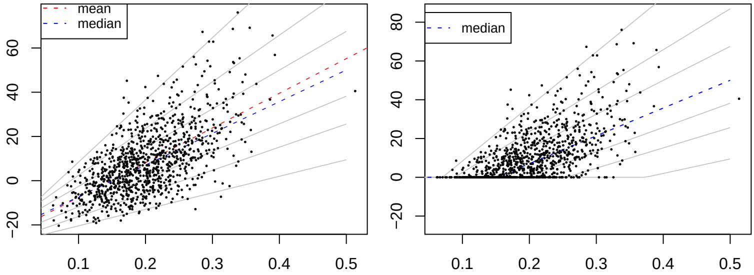
- A-priori assumption about parametric distribtion $F_{Y|\mathbf{X}}(y|\mathbf{x}) = G(y; \Theta(\mathbf{x}))$
  Estimate parameter function $\Theta(\mathbf{x})$

Censored quantile regression

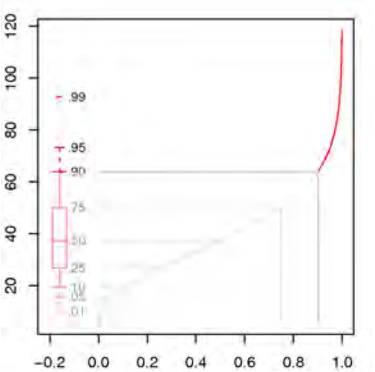
$$Q_{Z_{QR}}(\tau|\mathbf{X}) = \max(0, \boldsymbol{\beta}_\tau^T \mathbf{X}), \qquad \boldsymbol{\beta}_\tau^T = (\beta_0, \ldots, \beta_K)$$

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - \max(0, \boldsymbol{\beta}_\tau^T \mathbf{x}_i) \right)$$

with $\rho_\tau(u) = \tau u$ for $u \geq 0$ and $\rho_\tau(u) = (\tau - 1)u$ for $u < 0$



Conditional quantile function



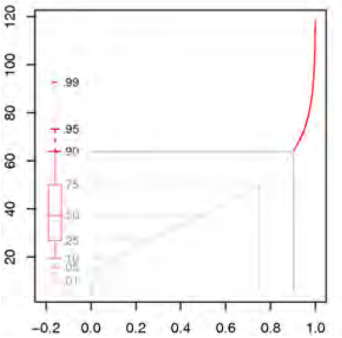
Semi-parametric

- A-priori probability τ, estimate conditional quantile $F^{-1}_{Y|\mathbf{X}}(\tau|\mathbf{x}) = \boldsymbol{\beta}^T_\tau \mathbf{x}$ via (linear) quantile regression

Parametric: Extreme Value Theory

- Parametric distribtion $F_{Y|\mathbf{X}}(y|\mathbf{x}) = G(y; \Theta(\mathbf{x}))$ is of the famililly of max-stable distributions.

Extreme value theory "Going beyond the range of the data"



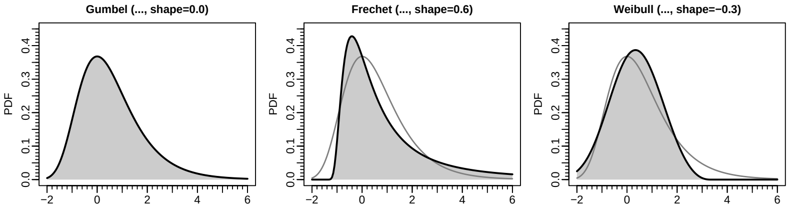
- **Limit theorem** for sample maxima
  → asymptotic distribution for extremes
- Condition of **max-stability** (de Haan, 1984)
  → maxima follow a generalized extreme value distribution
- Garantees **universal behavior of extremes**
  → enables extrapolation!

In praxis: often not enough data to reach asymptotic limit

Extreme value distribution

Generalized extreme value distribution (GEV)

$$G_\xi(y) = \begin{cases} \exp(-(1+\xi\frac{y-\mu}{\sigma})^{-1/\xi})_+, & \xi \neq 0 \\ \exp(-\exp(-\frac{y-\mu}{\sigma})), & \xi = 0 \end{cases},$$



Poisson point process model

For sufficiently large threshold u, $Z_i > u$ is Poisson point process on region $[0,1] \times (u, \infty)$ with intensity

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}.$$

for $A = [t_1, t_2] \times (u, z)$

μ, σ and ξ are parameters of corresponding GEV distribution.

Non-stationary Poisson point process model

Intensity of Poisson point process depends on multivariate covariate $\mathbf{X}$

$\mathbf{X}$ contains information from model (ensemble) forecast

$$\mu \rightarrow \boldsymbol{\mu}^T\mathbf{X} \qquad \sigma \rightarrow \boldsymbol{\sigma}^T\mathbf{X} \qquad \xi \rightarrow \boldsymbol{\xi}^T\mathbf{X}$$

The hyperparameter $\boldsymbol{\mu}^T = (\mu_0, \ldots, \mu_K)$, $\boldsymbol{\sigma}^T = (\sigma_0, \ldots, \sigma_K)$, $\boldsymbol{\xi}^T = (\xi_0, \ldots, \xi_K)$ are estimated by maximum likelihood method.

Forecast verification by means of scores

- Cost functions or distance between forecast and data
- Utility measure in a Bayesian context

A score is proper iff

$$\mathrm{E}_{y\sim Q}\left[S(P,y)\right] \geq \mathrm{E}_{y\sim Q}\left[S(Q,y)\right] \quad \forall \quad P \neq Q$$

$S(P,y)$: score function

Q forecasters best guess

$\mathrm{E}_{y\sim Q}\left[S(.,y)\right]$ expectation of $S(.,y)$ over $y \sim Q$

Verification: Goodness-of-fit criterion

$$\mathrm{QVS}(\tau) = \min_{\{\boldsymbol{\beta}\in\mathbb{R}^q\}} \sum_i \rho_\tau(y_i - \boldsymbol{\beta}^T \mathbf{x}_i) \qquad \mathrm{QVS}_{ref}(\tau) = \min_{\{\beta_0\in\mathbb{R}\}} \sum_i \rho_\tau(y_i - \beta_0)$$

Quantile verification skill score

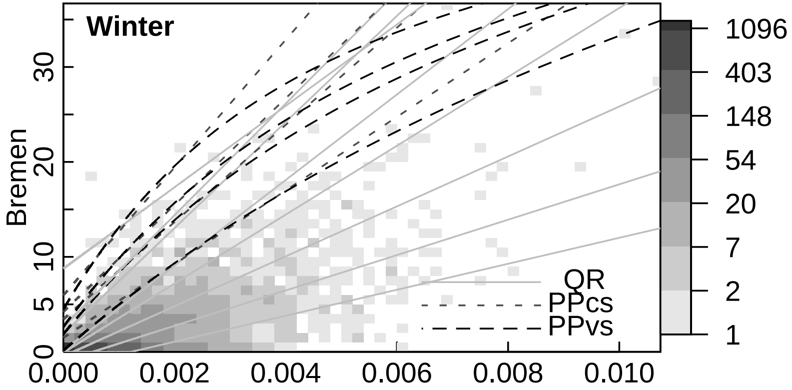
$$\mathrm{QVSS}(\tau) = 1 - \frac{\mathrm{QVS}(\tau)}{\mathrm{QVS}_{ref}(\tau)}$$

Log-likelihood ratio test: asymmetric Laplacian regression

$$f_\tau(u) = \frac{\tau(1-\tau)}{\sigma_L} \exp(-\rho_\tau(u)/\sigma_L).$$

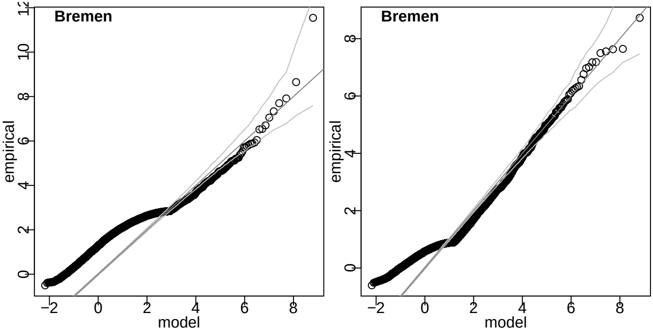
proportional to $\log(\mathrm{QVS}(\tau)/\mathrm{QVS}_{ref}(\tau))$

Quantile Estimates

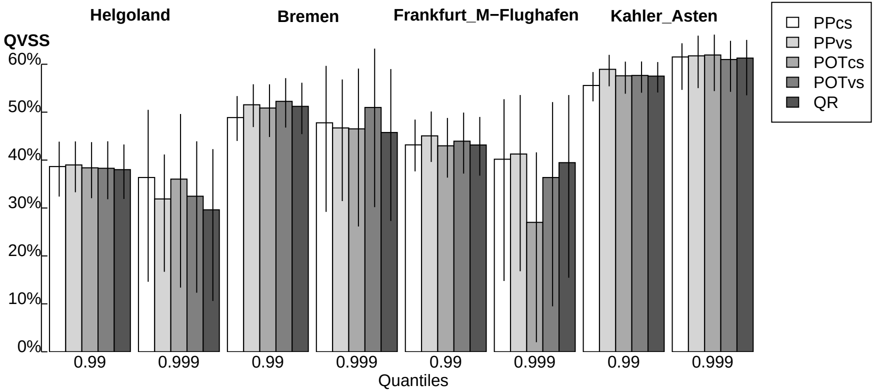


Residual quantile plots

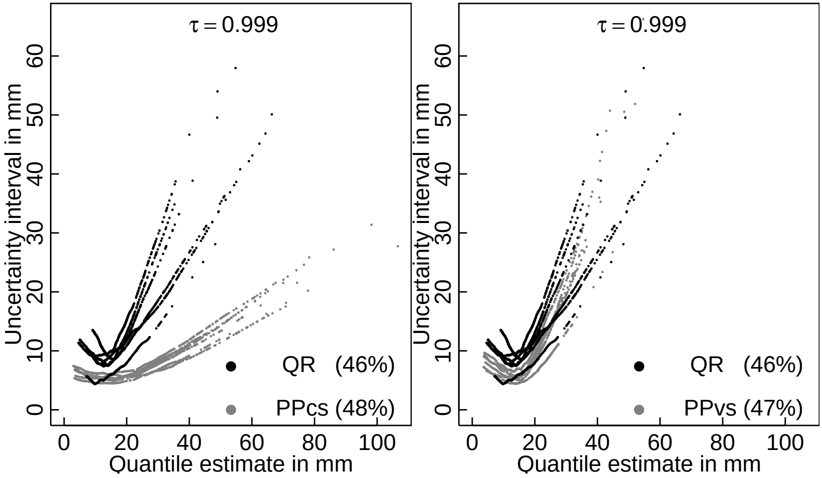
$$\left\{\left(-\log(-\log(\frac{i}{n+1})), -\log((1+\xi_{(i)}\frac{z_{(i)}-\mu_{(i)}}{\sigma_{(i)}})^{-1/\xi_{(i)}})\right), i=1,\ldots,n\right\}$$



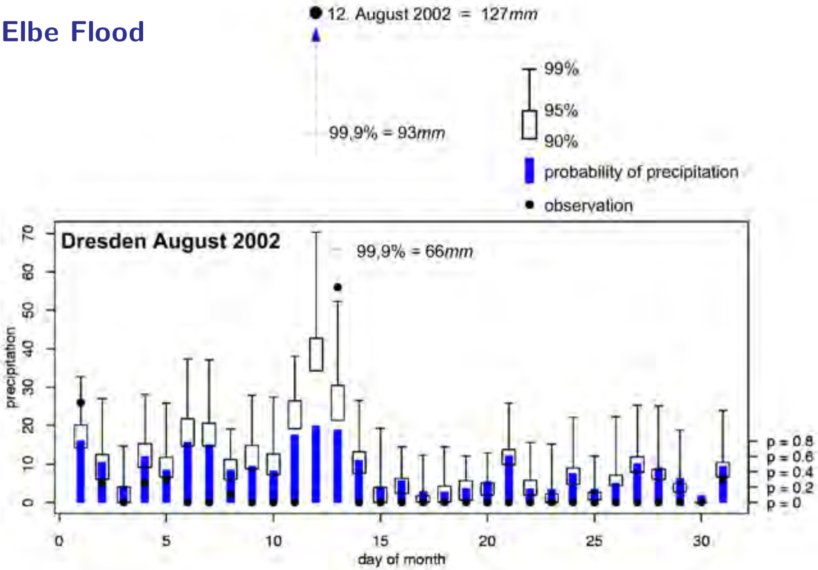
Quantile Verification Score



Uncertainty of Quantile Estimates



Elbe Flood


12. August 2002 = 127mm
99,9% = 93mm
99%
95%
90%
probability of precipitation
observation
Dresden August 2002
99,9% = 66mm
precipitation
0
10
20
30
40
50
60
70
day of month
0
5
10
15
20
25
30
p = 0.8
p = 0.6
p = 0.4
p = 0.2
p = 0

Conclusions

- Weather forcasts provide information that conditions occurence of extremes
- Linear (non-linear) statistical modeling extracts information
- Extreme value theory provides distributions tailored for extremes
- Parametric method less uncertain than non-parametric method and non-linear dependecy (shape parameter) is not parsimony
- High-impact weather: insufficient data available for training and for validation

Challenges

- Improve physical understanding of generation processes of extremes
- Application to multi-variable and spatio-temporal predictionswhich
  - Combine spatial statistics with model post-processing (Berrocal et al., 2007)
  - Develop methods for multivariate post-processing
  - Develop ensemble methods tailored to extremes
- Verification tailored to extremes
- Verification for probabilistic multivariate and spatial forecasts

Reference

- Friederichs, P., 2010: Statistical downscaling of extreme precipitation using extreme value theory. Extremes 13, 109-132.

Special thanks to:

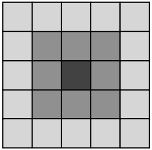
Susanne Theis, Deutscher Wetterdienst, Offenbach

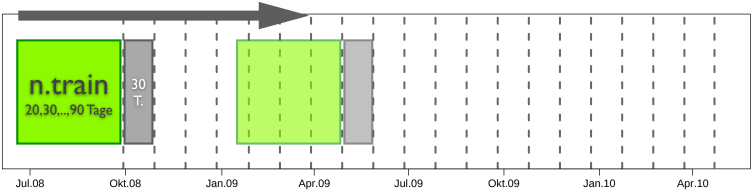
Thank You for Your Attention!

Ensemble Post-Processing

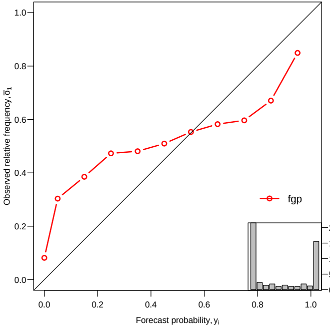
- COSMO-DE forecasts 1 July 2008 – 30 April 2010
- 12 h accumulated precipitation between 12 and 00 UTC

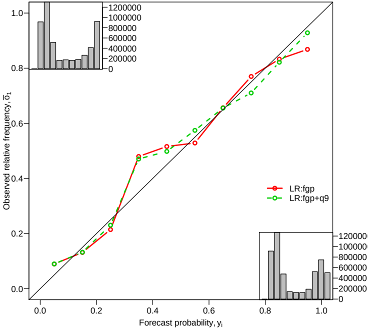
First guess probability (fgp) and 0.9-quantile (fgq9) from 5×5 neighborhood and 4 COSMO-DE forecasts



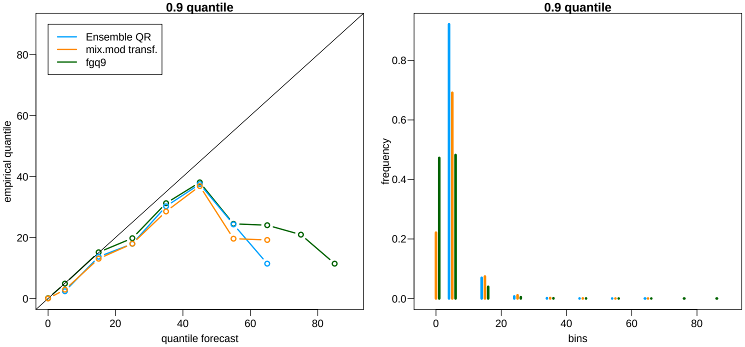


Logistic regression

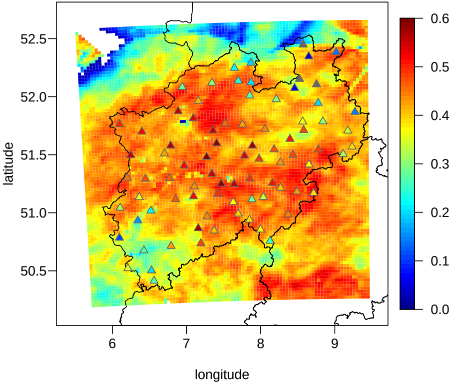




Quantile forecasts - reliability



Quantile forecasts - Scores



- 0.9-Quantil
- QVSS
- MIX: $fgp + \sqrt[3]{fgq9}$
- 60 days training
- Station data

Censoring

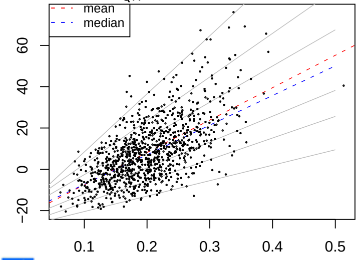
Equivariance with respect to non-decreasing function $h(\cdot)$

$$Q_{h(Y)}(\tau) = h(Q_Y(\tau))$$

Hidden process Y^* observed through censored variable Y

$$Y = h(Y^*) = \max[0, Y^*]$$

$$Q_{Y^*_{QR}}(\tau|\mathbf{X}) = \beta_\tau^T \mathbf{X}$$



$$Q_{Y_{QR}}(\tau|\mathbf{X}) = \max(0, \beta_\tau^T \mathbf{X})$$