Statistical downscaling of precipitation using extreme value theory

Petra Friederichs

Meteorological Institute
University of Bonn

E2C2-GIACS Summer School, Romania, September 3-12, 2007

Outline

Statistical downscaling – general overview

- Aim of statistical downscaling
- Approaches and Methods
- with regard to extremes

Our Problem

- Data
- Probabilistic Downscaling
- Extreme value theory
- Censored quantile regression
- Verification

Results

- Extreme precipitation events
- Skill of downscaling

Conclusions

Weather Forecasting

DYNAMICS:

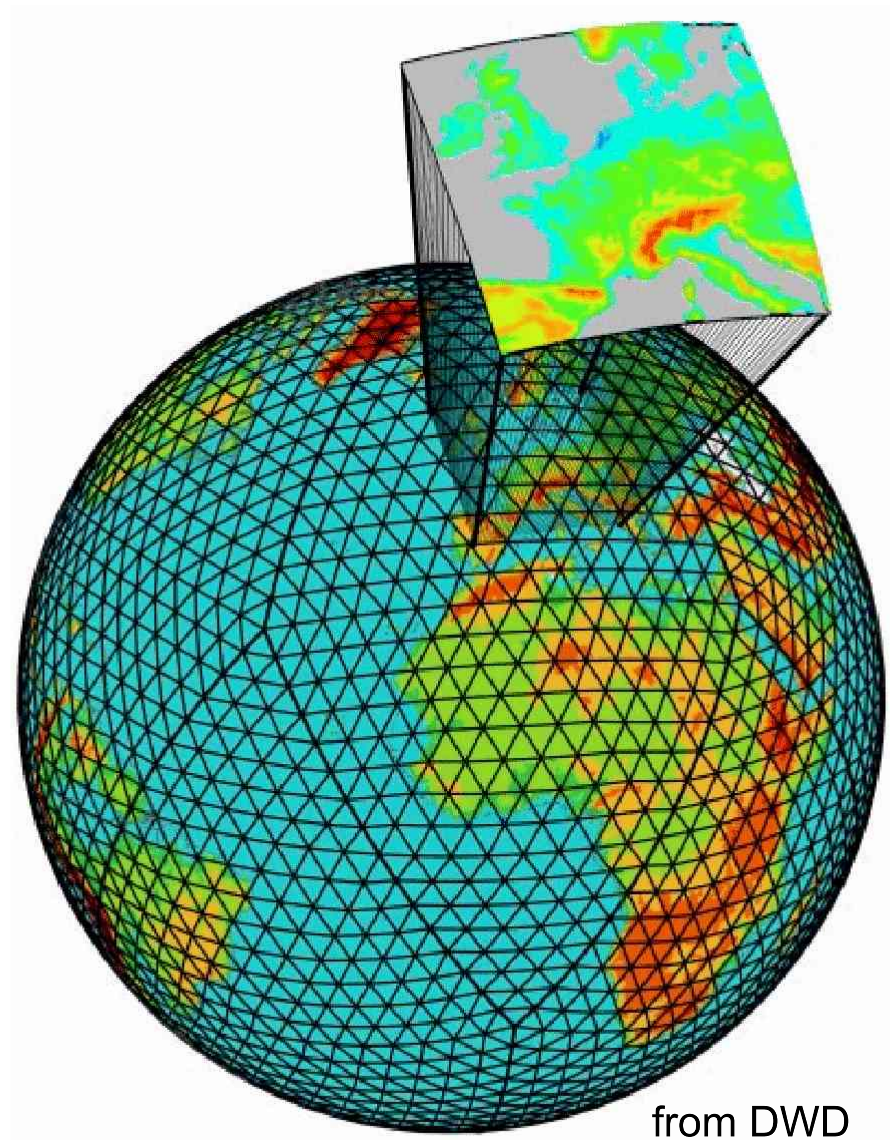
In numerical weather forecasting the **Navier Stokes equations** on a rotation sphere and the 1st principle of thermodynamics are numerically intergrated in time

PHYSICS:

sub-scale processes are parametrized
- cloud micro-physics
- radiative transfer
- exchange processes with land surface...
- turbulence
- **convection** ...

Extreme weather phenomena are (mostly) on small scales

–> post-processing

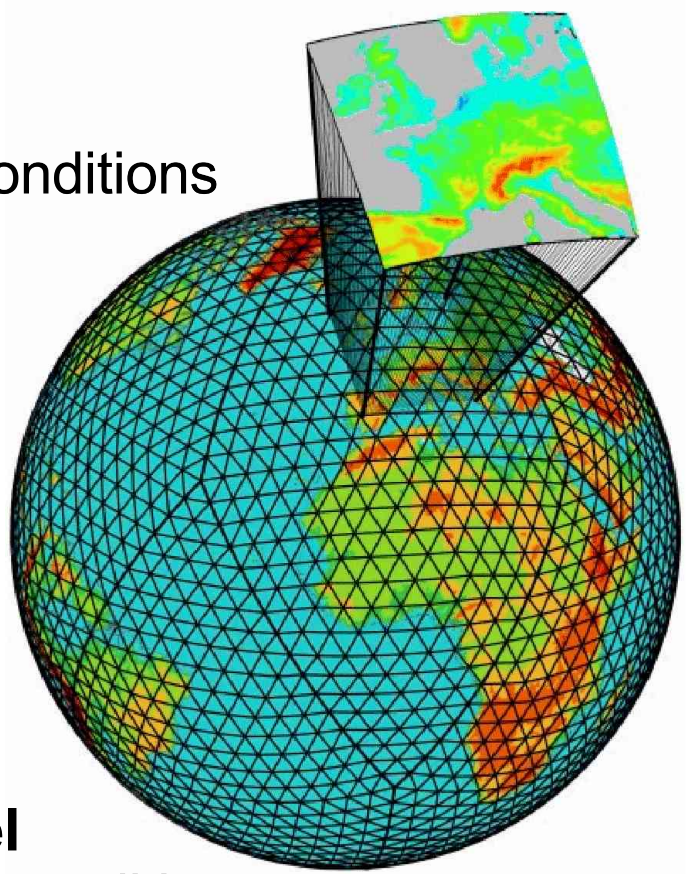


from DWD

Downscaling

Aim of downscaling:

Represent and forecast local weather or climate conditions on scales not resolved by the model.
Here: downscaling in space!



Dynamical and **statistical** downscaling

Dynamical downscaling uses a **regional, high resolution model**, where the boundary conditions are determined by the global model.

Statistical downscaling derives a **statistical model** between the large scale model state and the local conditions.

Downscaling

Applications:

Short range weather prediction
• regional models to account for local conditions (orography) and small scale processes (convection)
• statistical downscaling (MOS) for local (weather station) forecasts
• assessing probabilities e.g. of extreme events

Climate change projections
• estimate possible man induced changes on the local scale

Weather generator
• generate realistic precipitation patterns for hydrological modeling

Methods of Stat. Downscaling

Linear transfer functions

- multiple linear regression
- canonical correlation analysis

Non-linear transfer functions

- neural networks

Stochastic models

- multivariate autoregressive models
- conditional weather generator

Weather typing methods

- conditional resampling
- analogue-based methods

Probabilistic downscaling

- estimate conditional distribution function

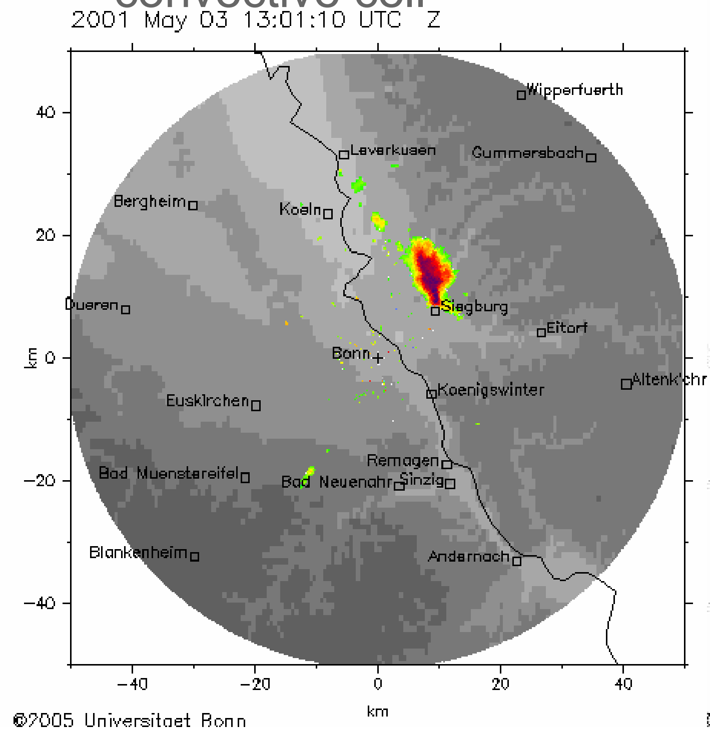
Meteorological Institute
University Bonn

Extreme Weather Phenomena

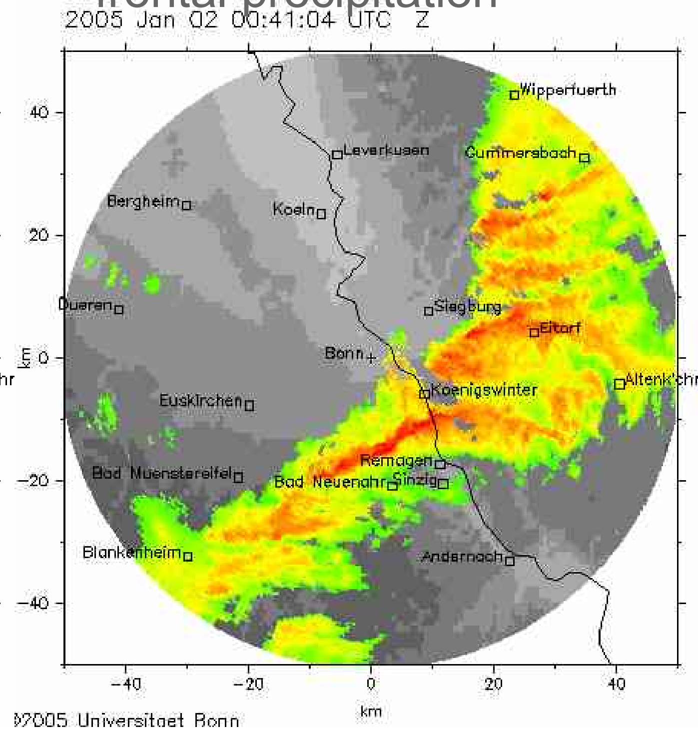
universität bonn

heavy precipitation events:
reasonable observational network
spacial dependence – unknown, depends on weather

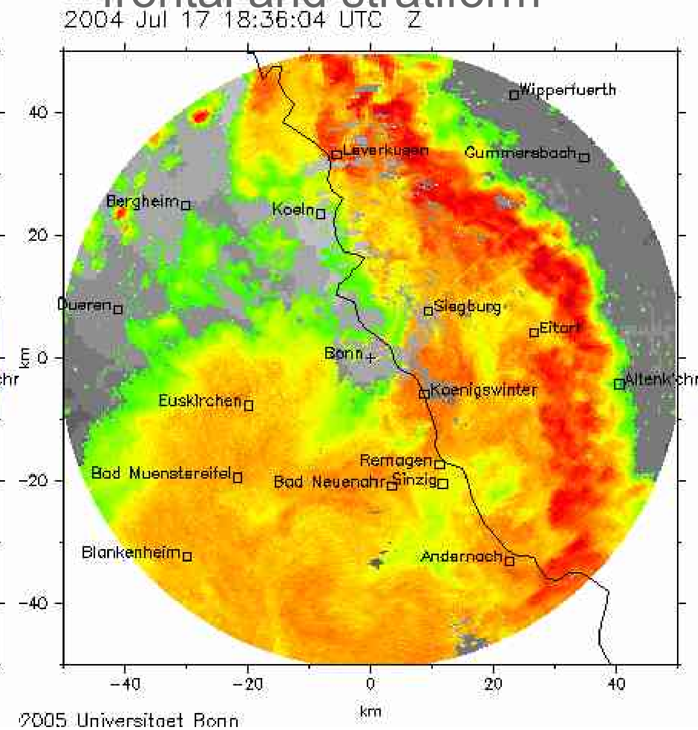
2001 May 03
convective cell



2005 Jan 02
frontal precipitation



2004 Jul 17
frontal and stratiform



wind gusts:
must less observations
processes not well known
spacial dependence - unknown

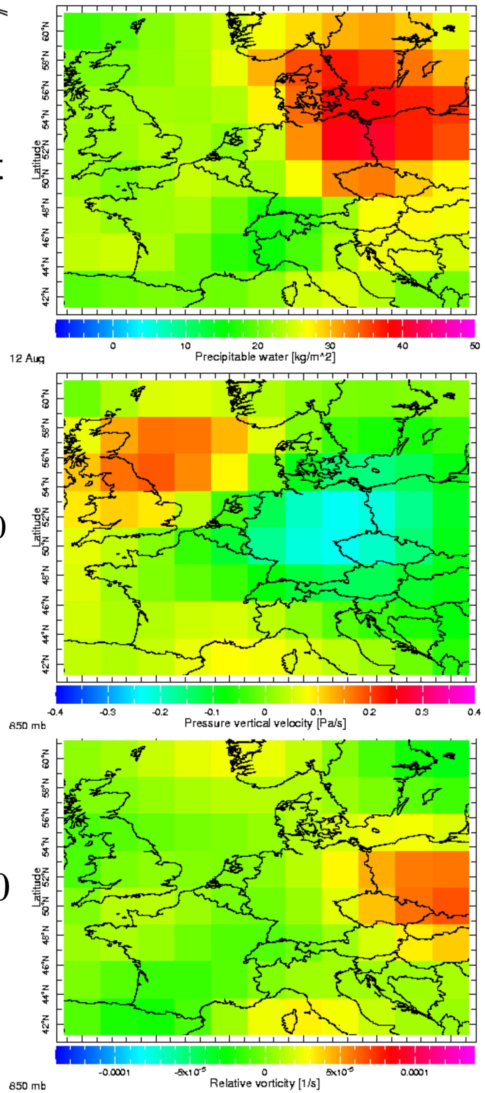
heat waves:
good observations
but long memory

Data

PWat

ω_{850}

ζ_{850}



NCEP reanalysis

Daily mean fields of ω_{850}, ζ_{850} and PWat

Period from:

1948 to 2004

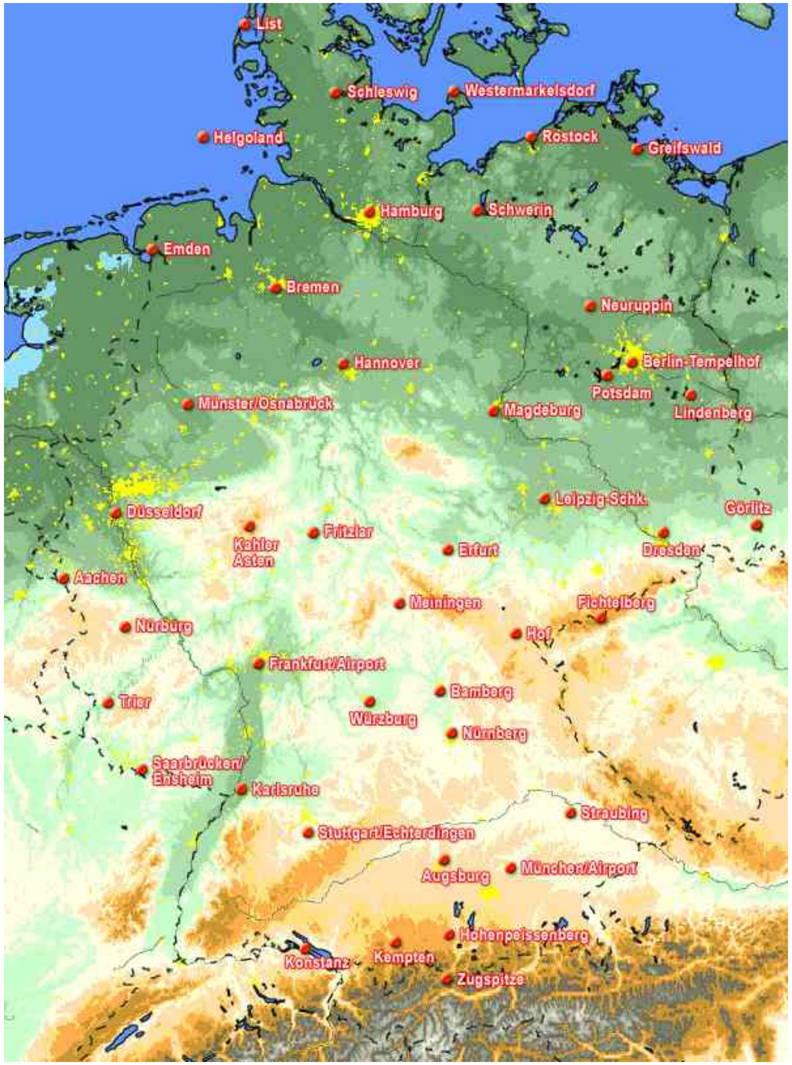
Separate:

cold (NDJFM) and

warm seasons (MJJAS)

About:

8600 daily totals



DWD (German Weather Service)

~ 2000 stations in Germany

Daily precipitation totals

Downscaling

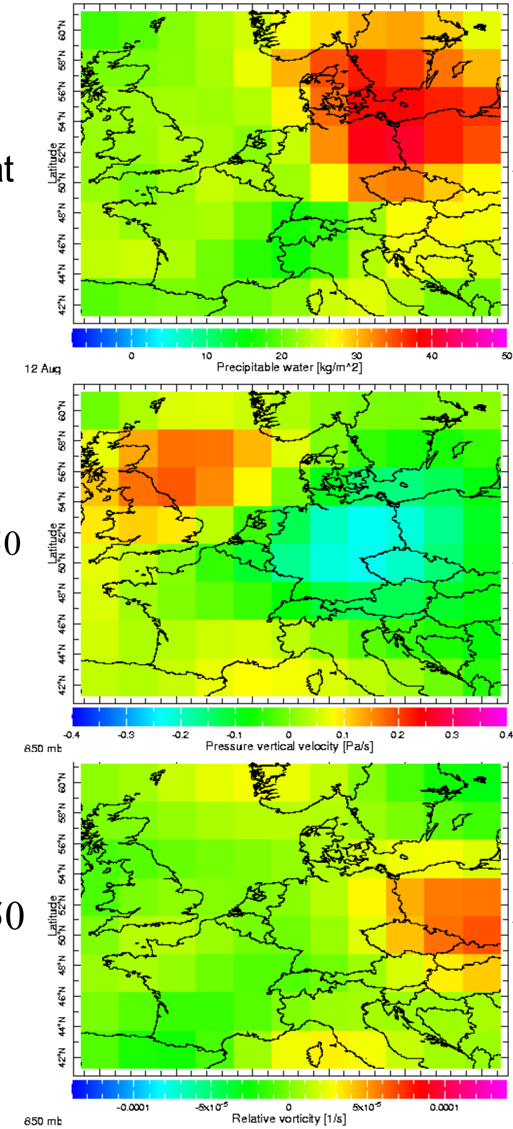
Downscaling seeks for:
a **statistical model** between the **covariate X** and a **variable** ***y***

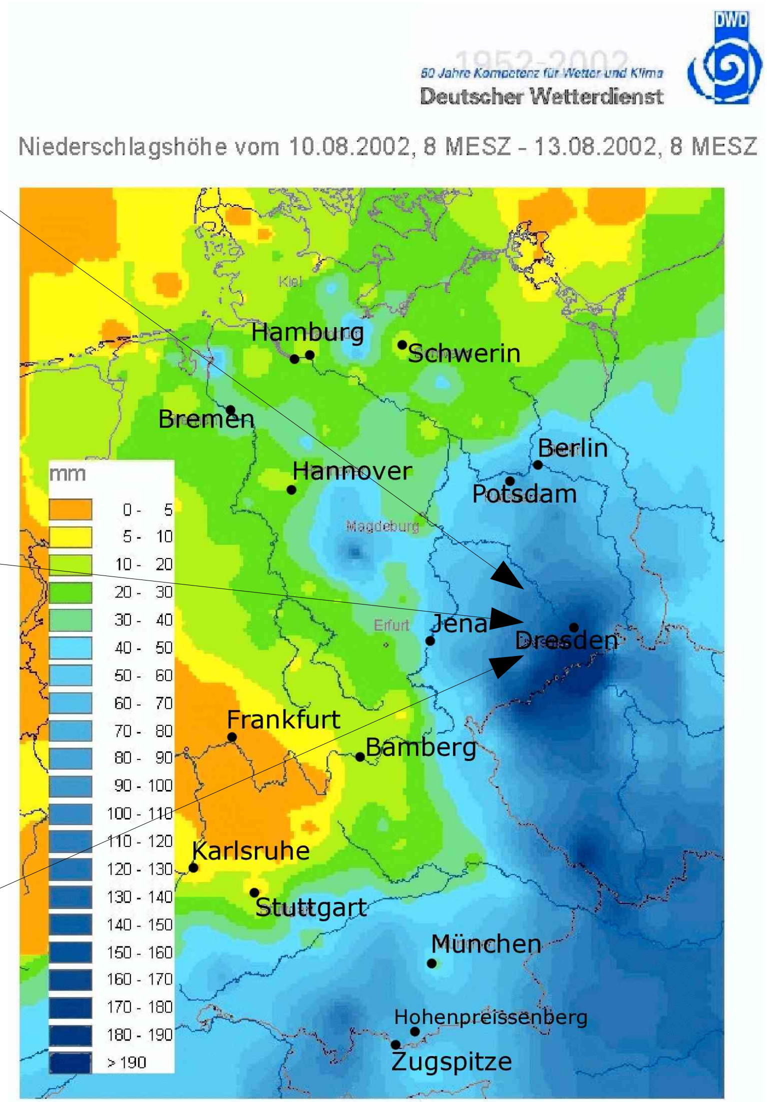
7 EOF

PWat

ω_{850}

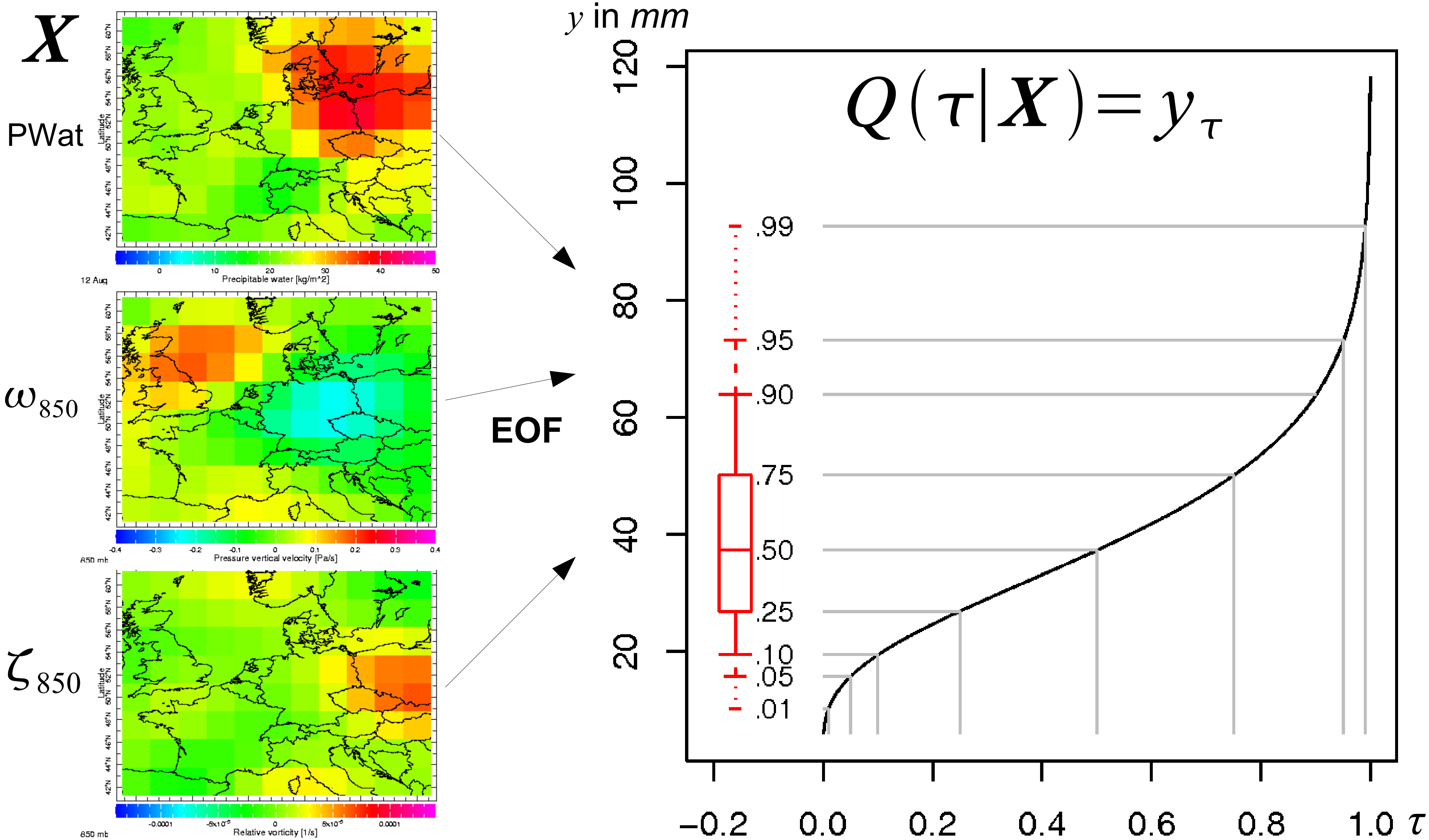
ζ_{850}





Probabilistic Downscaling...

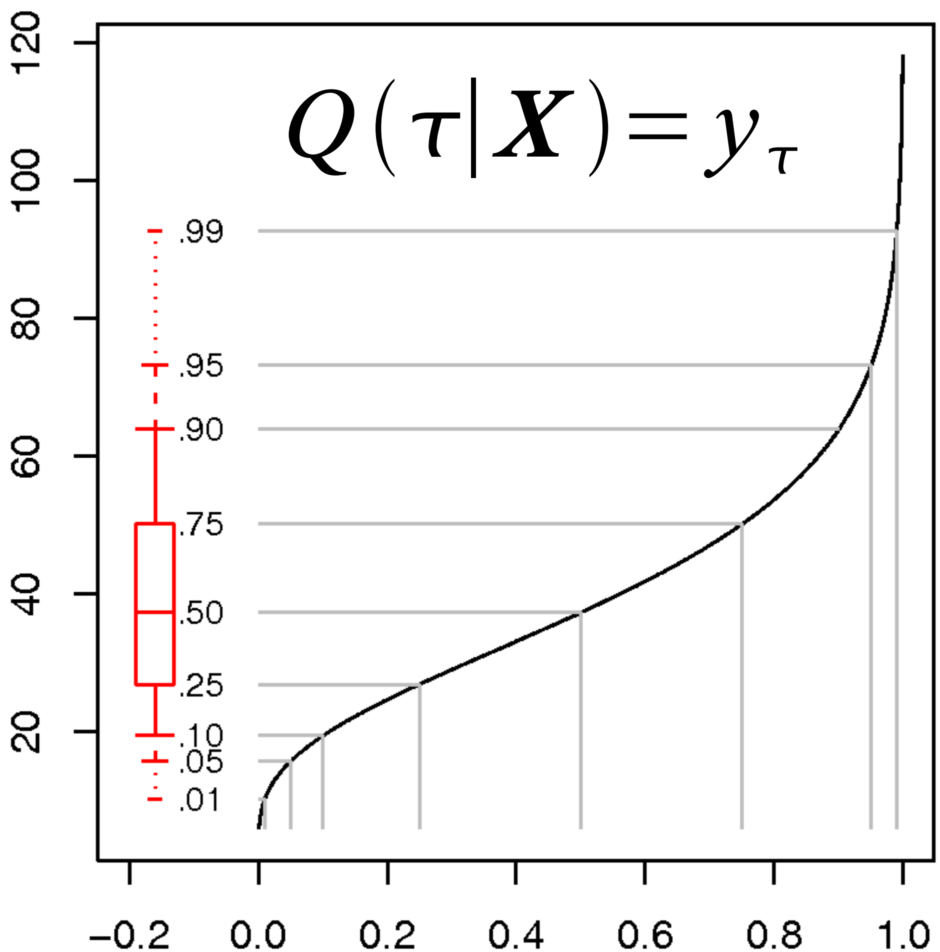
... seeks for a **statistical model** between the **covariate X** and a **probabilistic measure of** *y* - with special emphasis on



Probabilistic Downscaling

How to estimate a probabilistic measure of *y* given **X**?

- Distribution function: $F(y|\mathbf{X}) = p$
- Quantile function: $Q(\tau|\mathbf{X}) = y_\tau$



Semi-parametric:

Fix threshold *y*,
estimate 1-*p* probability of threshold exceedance using logistic regression

**Fix probability τ=*p*,
estimate y quantile using quantile regression**

Parametric:

Distributional assumption:
estimate parameter vector given **X**

Poisson point – GPD process
estimate parameter μ, σ, ξ given **X**

POT approach with u = u(**X**)
with parameters σ_u, ξ given **X**

Censored Quantile Regression

The **censored** linear model

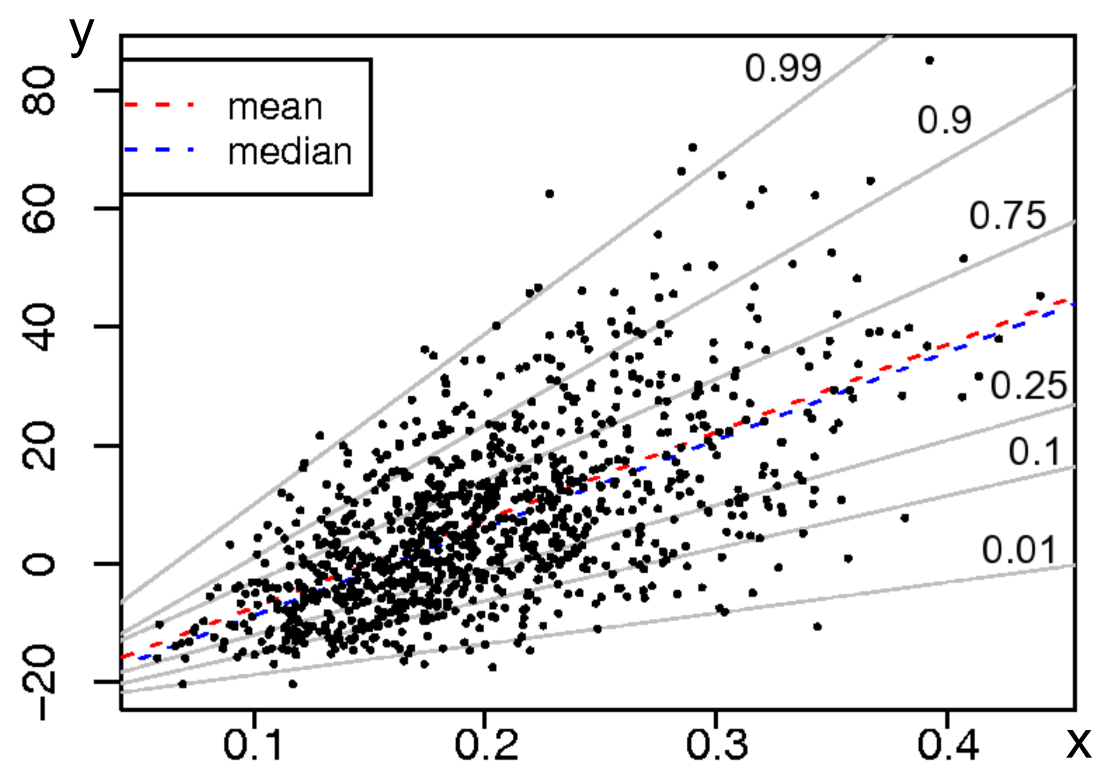
$$Y|X = max(0, \boldsymbol{\beta}^T X + \gamma^T X u) \qquad u \sim i.i.d.$$

the conditional quantile function

$$Q_y(\tau|X) = max(0, \hat{\boldsymbol{\beta}}_{\tau}^T X)$$

Example
no censoring:

$$Q_{\tilde{y}}(\tau|X) = \hat{\boldsymbol{\beta}}_{\tau}^T X$$



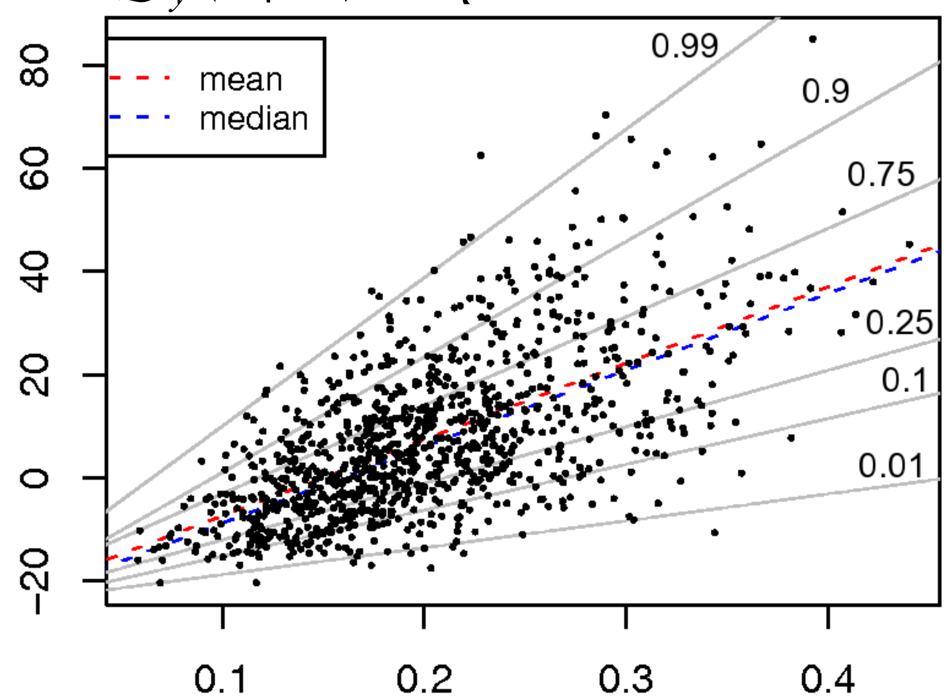
Censoring

Conditional quantiles are **equivariant to monotone transformations** $h(.)$

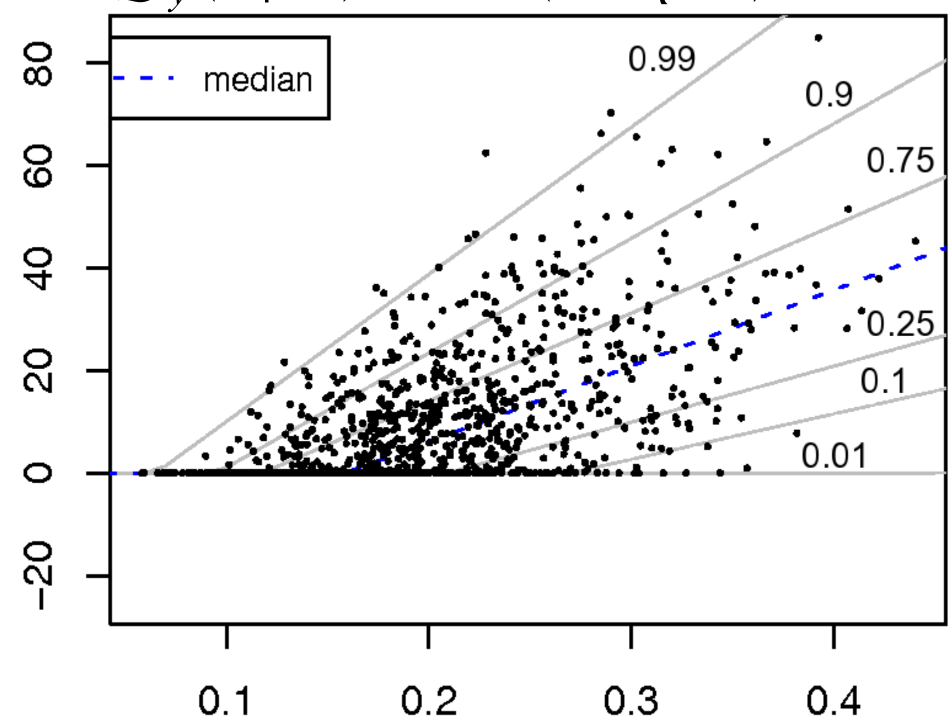
$$Q_{h(\tilde{Y})}(\tau) = h(Q_{\tilde{Y}}(\tau))$$

in out case $y = h(\tilde{y}) = max(0, y)$

$Q_{\tilde{y}}(\tau|X) = \hat{\beta}_\tau^T X$



$Q_y(\tau|X) = max(0, \hat{\beta}_\tau^T X)$



Censored Quantile Regression

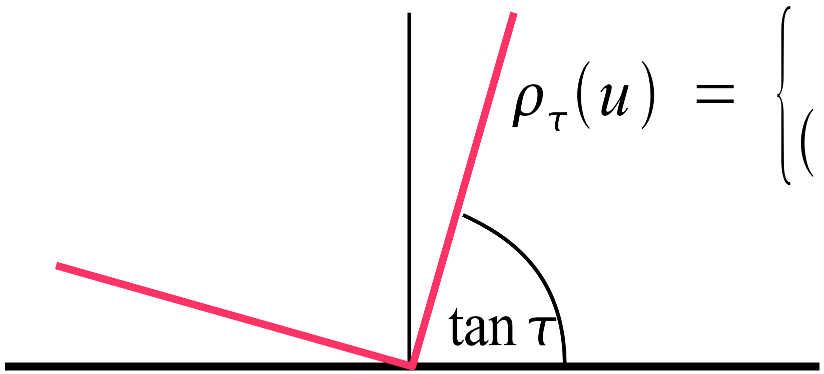
The **censored** linear model

$$Y|\boldsymbol{X} = max(0, \boldsymbol{\beta}^T \boldsymbol{X} + \gamma^T \boldsymbol{X} u) \qquad u \sim i.i.d.$$

the conditional quantile function $\hat{y}_\tau = Q_y(\tau|\boldsymbol{X}) = max(0, \hat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}^{\boldsymbol{T}} \boldsymbol{X})$

Censored quantile regression minimizes the **absolute deviation function**

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} = arg\, min_{\boldsymbol{\beta}_\tau} \sum_n \rho_\tau [y_n - max(0, \boldsymbol{\beta}_\tau^T \boldsymbol{x}_n)]$$



$$\rho_\tau(u) = \begin{cases} \tau u & if\ u \geq 0 \\ (\tau - 1) u & if\ u < 0 \end{cases}$$

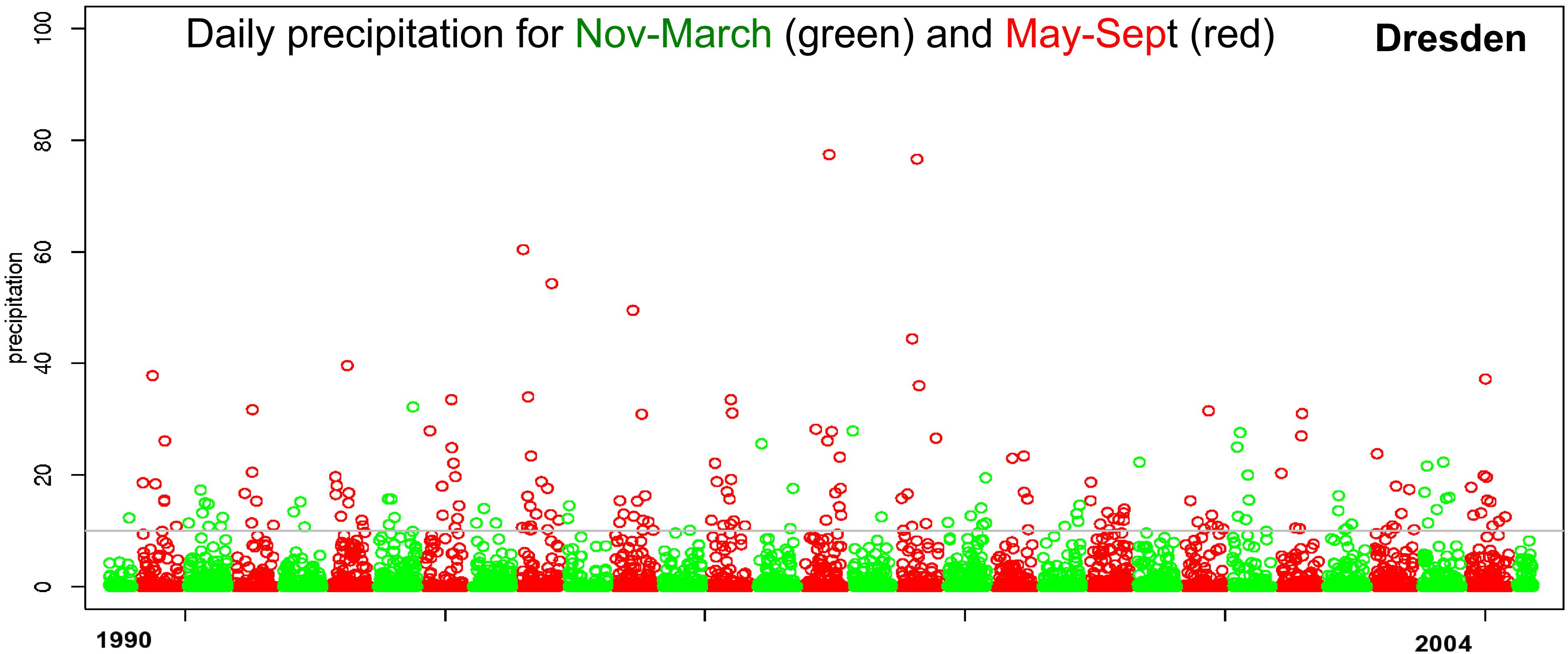
Peak Over Threshold

Peak over Threshold (POT)

$$\{ Y_i - u | Y_i > u \}$$

very large threshold u

follow a **Generalized Pareto Distribution (GPD)**



Peak Over Threshold

Peak over Threshold (POT)

The distribution of $Z_i := Y_i - u | Y_i > u$ exceedances over large threshold u are asymptotically distributed following a

Generalized Pareto Distribution (GPD)

$$H(z|Y_i > u) = \begin{cases} 1 - \left(1 + \xi \dfrac{z}{\sigma_u}\right)^{-1/\xi} & 1 + \xi \dfrac{z}{\sigma_u} > 0 \quad \xi > 0 \\ 1 - \exp\left(\dfrac{-z}{\sigma_u}\right) & z > 0 \quad \xi = 0 \end{cases}$$

two parameters

- σ_u scale parameter
- ξ shape parameter

POT with variable threshold

For the model we have 3 parameters

threshold u is definded as conditional τ-quantile

$$\hat{u}_{\tau} = Q_y(\tau_u | \boldsymbol{X}) = \hat{\boldsymbol{\beta}}^T \boldsymbol{X}$$

scale and **shape parameter** of GPD

$$\hat{\sigma} = \boldsymbol{\sigma}^T \boldsymbol{x}_n \qquad \hat{\xi} = \boldsymbol{\xi}^T \boldsymbol{x}_n$$

Estimation of threshold uses censored quantile regression, conditional GPD parameters of the exceedances are estimated using maximum likelihood method.

POT with variable threshold

The conditional distribution of precipitation is

$$F(y|\boldsymbol{X}) = H(y-u_x|y>u_x,\boldsymbol{X})\,Prob(y>u_x|\boldsymbol{X}) + Prob(y\leqslant u_x|\boldsymbol{X})$$

$$= H(u-u_x|y>u_x,\boldsymbol{X})(1-\tau_u)+\tau_u \;=\; \tau>\tau_u$$

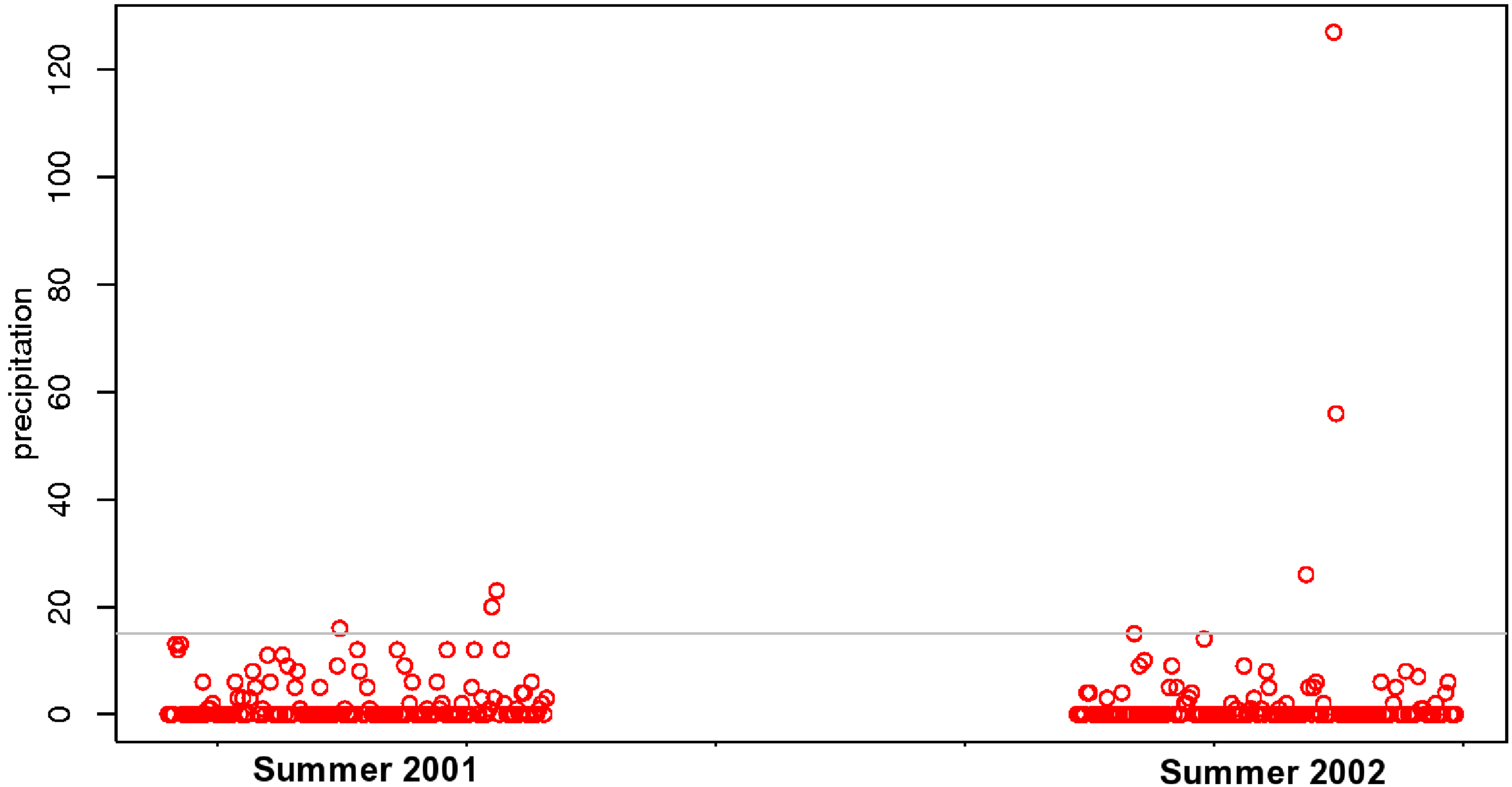
The respective τ-is estimated as

$$\hat{y}_\tau = H^{-1}(\tau|\boldsymbol{X}) = \begin{cases} u_x + \dfrac{\hat{\sigma}_u}{\hat{\xi}}\left[(1-\tilde{\tau})^{-\hat{\xi}}\right] & \xi>0 \\ u_x + \hat{\sigma}_u \log(1-\tilde{\tau}) & \hat{\xi}=0 \end{cases}$$

Poisson Point - GPD

Poisson point – GPD process with intensity

$$\Lambda(A) = (t_2 - t_1)\left[1 + \xi\left(\frac{y-u}{\sigma_u}\right)\right]^{-1/\xi} \quad \text{on} \quad A = (t_1, t_2) \times (y, \infty)$$



Poisson Point - GPD

Poisson point – GPD process with intensity

$$\Lambda(A)=(t_2-t_1)[1+\xi(\frac{y-u}{\sigma_u})]^{-1/\xi} \quad \text{on} \quad A=(t_{1,}t_2)\times(y,\infty)$$

Non-stationarity:
assume parameters to depend linearly on covariate **X**

$$\hat{\mu}_n=\boldsymbol{\mu}^T\boldsymbol{x}_n \qquad \hat{\sigma}_n=\boldsymbol{\sigma}^T\boldsymbol{x}_n \qquad (\hat{\xi}_n=\boldsymbol{\xi}^T\boldsymbol{x}_n)$$

Estimate parameters using **maximum likelihood**

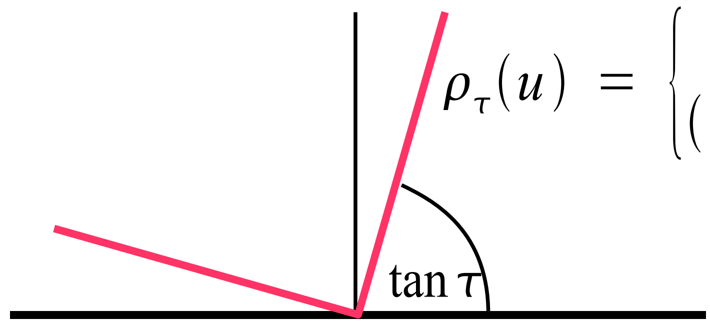
Get the **conditional extremal quantiles**

$$\hat{y}_\tau=u-\frac{(\hat{\sigma}+\hat{\xi}(u-\hat{\mu}))}{\hat{\xi}}(1-(\frac{-\log\tau}{\hat{\lambda}})^{-\hat{\xi}})$$

Quantile verification score

Quantile regression minimizes
the **absolute deviation function**

$$QVS = \sum_n \rho_\tau \left[y_n - \hat{y}_{n,\tau} \right]$$



$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geqslant 0 \\ (\tau - 1) u & \text{if } u < 0 \end{cases}$$

proper scoring rule (Gneiting and Raftery 2005)

censored quantile verification skill score

$$CQVSS(\tau) = 1 - \frac{CQVS(\tau, \hat{\boldsymbol{y}})}{CQVS(\tau, \boldsymbol{y}_{ref})}$$

Forecast verification

A **Scoring rule** $S(P,y)_{y\sim Q}$

measures the quality or utility of a probabilistic forecast

here in terms of a cost function

Let $$E[S(P,y)]_Q := S(P,Q)$$

Let Q be the forecasters best judgement
S is a proper scoring rule if for all P and Q

$$S(P,Q) \geqslant S(Q,Q)$$

strictly proper if

$$S(P,Q) = S(Q,Q) \quad iff \quad P = Q$$

Gneiting T, A E Raftery (2007): Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association 102,359-378

Forecast verification

Compare QR, POT and PP approach
extend censoring to threshold u

$$S_\tau(P_{QR}, \boldsymbol{y}) = \sum_n \rho_\tau[y_n - max(0, \hat{\boldsymbol{\beta}}_\tau^T \boldsymbol{x}_n)]$$

$$S_\tau(P_{POT/PP}, \boldsymbol{y}) = \sum_n \rho_\tau[y_n - \hat{y}_{n,\tau}]$$

$$S_\tau(P_{ref}, \boldsymbol{y}) = \sum_n \rho_\tau[y_n - Q_y(\tau)]$$

define a quantile verification skill score

$$QVSS_\tau = 1 - \frac{S_\tau(P_{QR/POT/PP}, \boldsymbol{y})}{S_\tau(P_{ref}, \boldsymbol{y})}$$

cross-validation

Cross validation

Derive forecast using **cross-validation**
through separation in **training** period and **target** season

To obtain set of forecasts $\{\hat{y}_n\}$ (target season)

we derive estimates of coefficients $\hat{\boldsymbol{\beta}}_\tau, \quad \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\xi}}, \quad u_x, \hat{\boldsymbol{\sigma}}_u, \hat{\boldsymbol{\xi}}$

from set of observations $\{\ldots, y_{n'\neq n}, \ldots\}$

(training period)

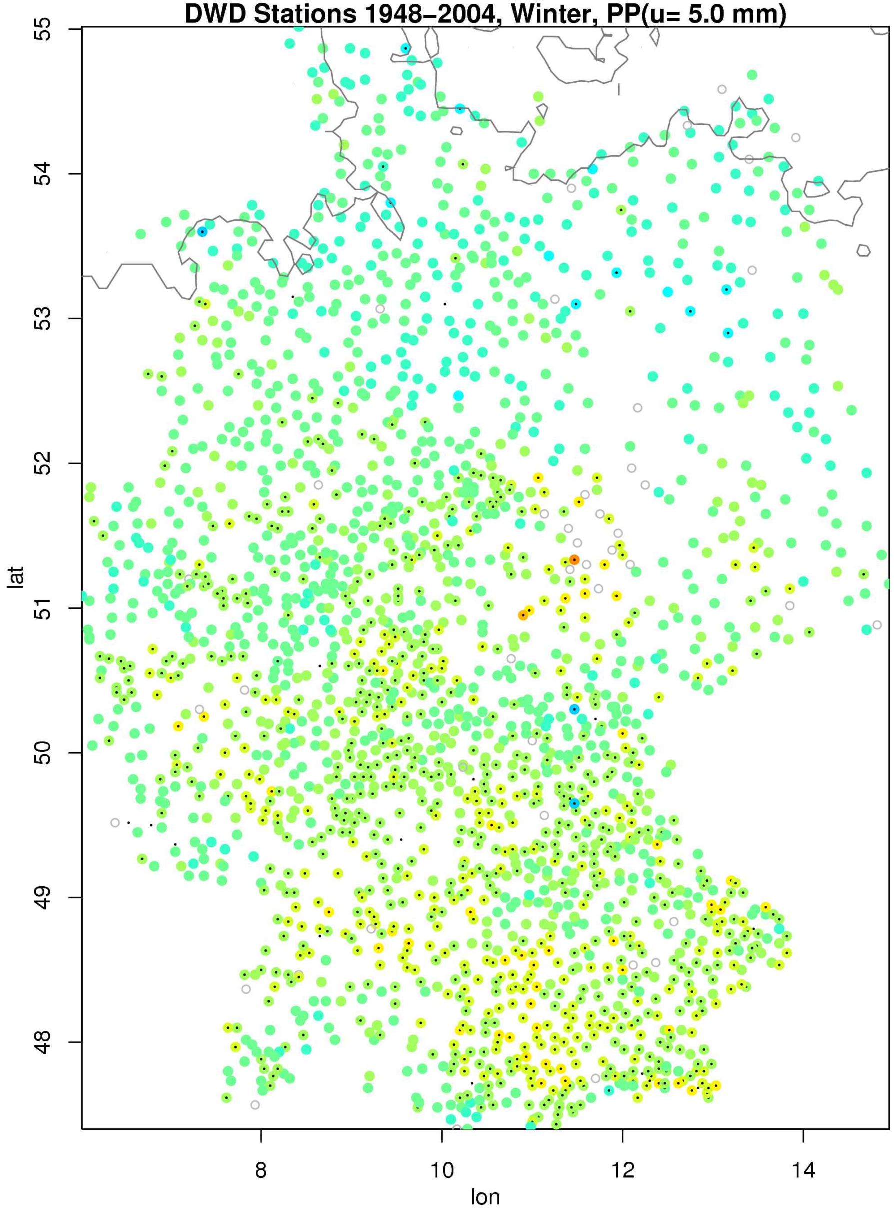
and covariates $\{\ldots, \boldsymbol{x}_{n'\neq n}, \ldots\}$

The forecast is derived for each target season in the time sequence

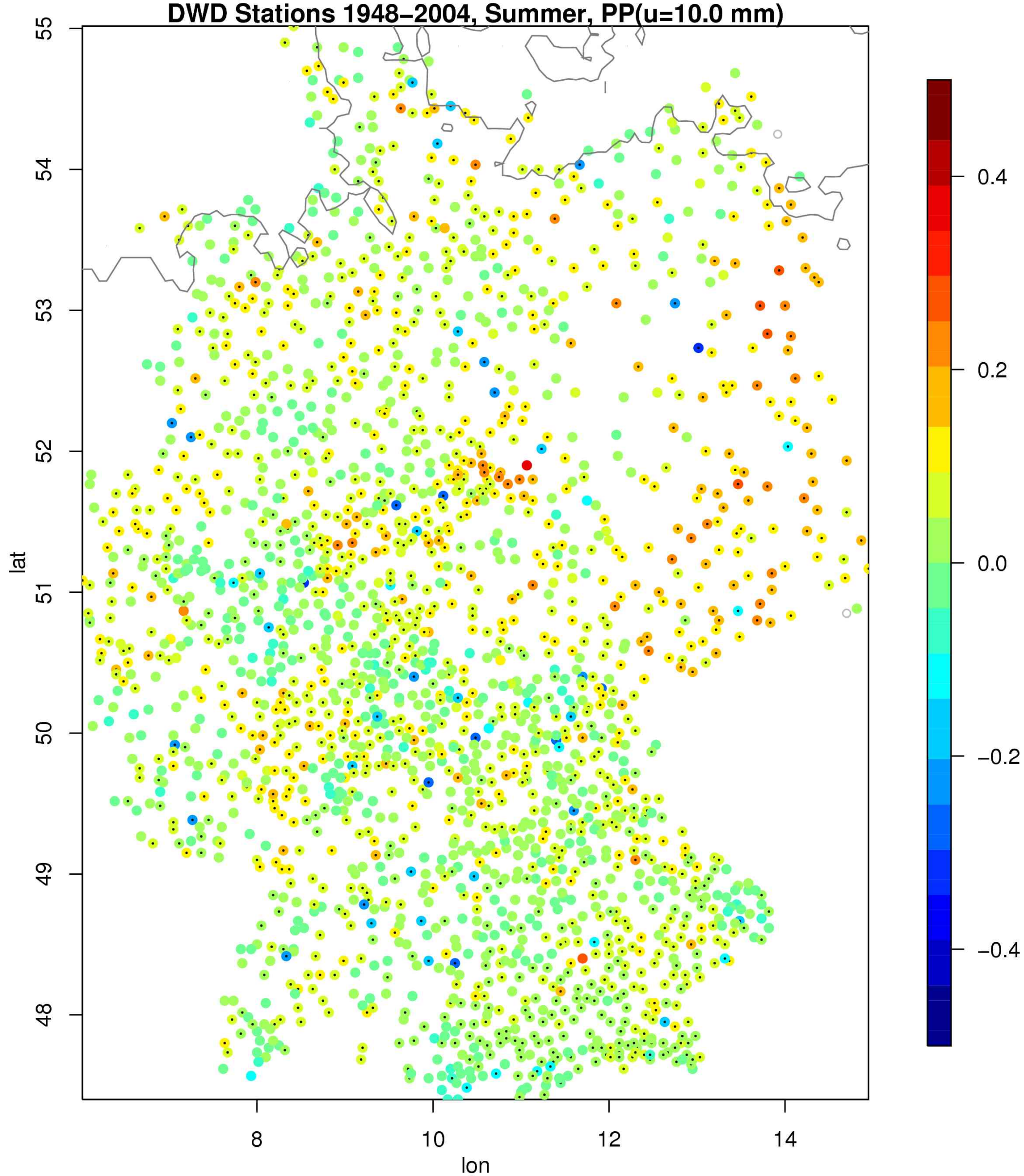
$$\hat{y}_{n,\tau} = max(0, \hat{\boldsymbol{\beta}}_\tau^T \boldsymbol{x}_n) \qquad \hat{y}_{n,\tau} = Q_{y,POT/PP}(\tau|\boldsymbol{x}_n)$$

EVD: Shape Parameter

Winter, u = 5mm



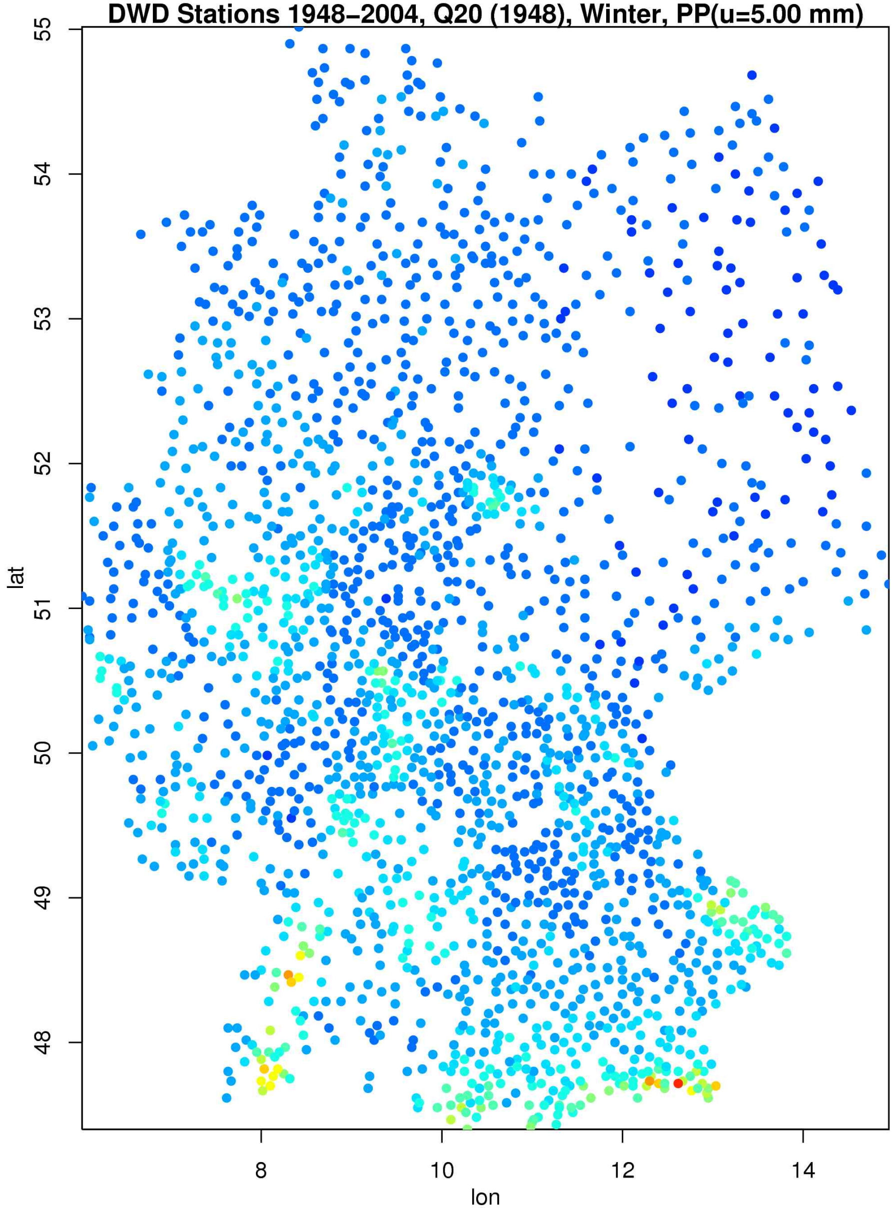
Summer, u = 10mm

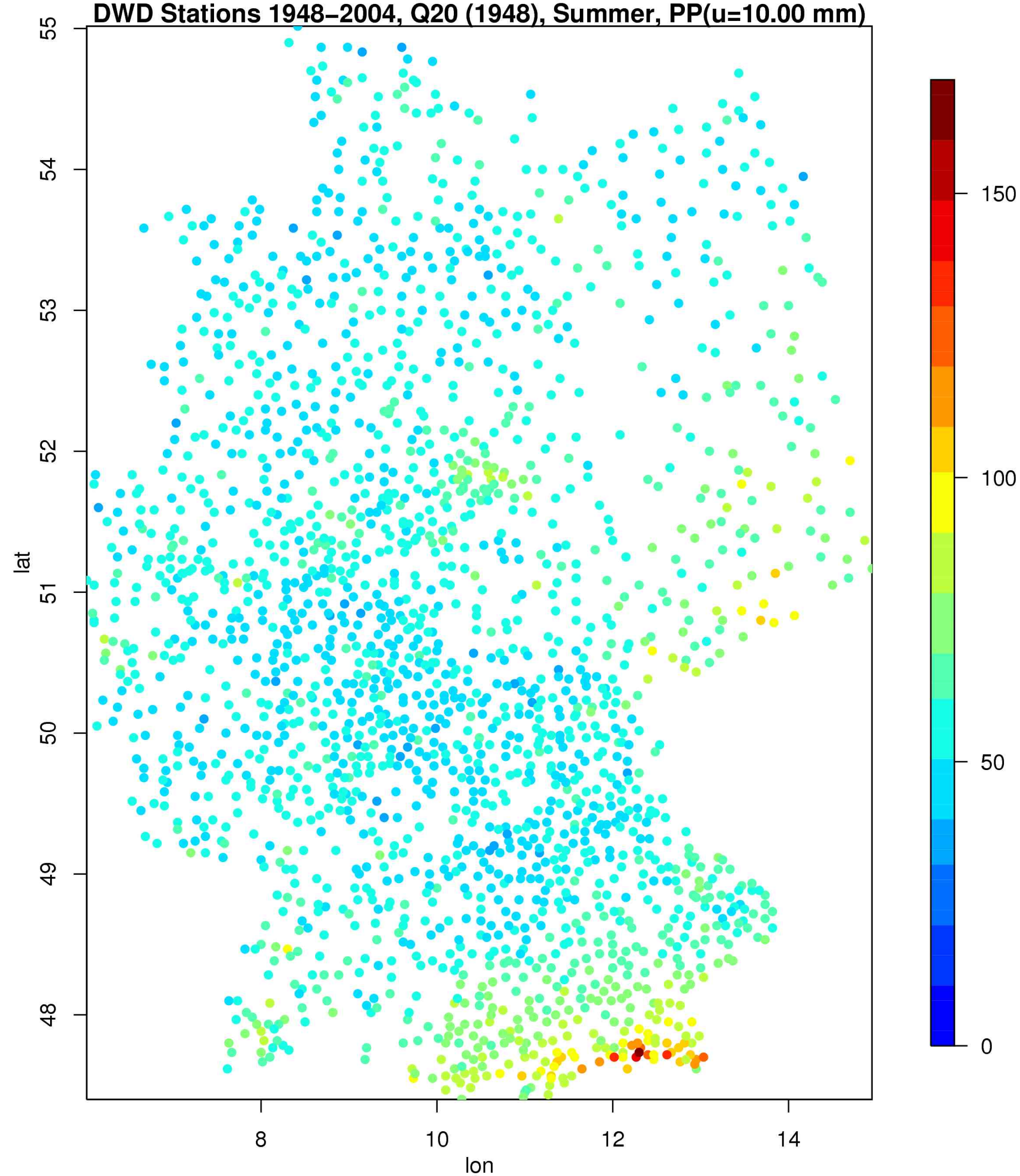


EVD: 20y return values

Winter, u = 5mm

Summer, u = 10mm

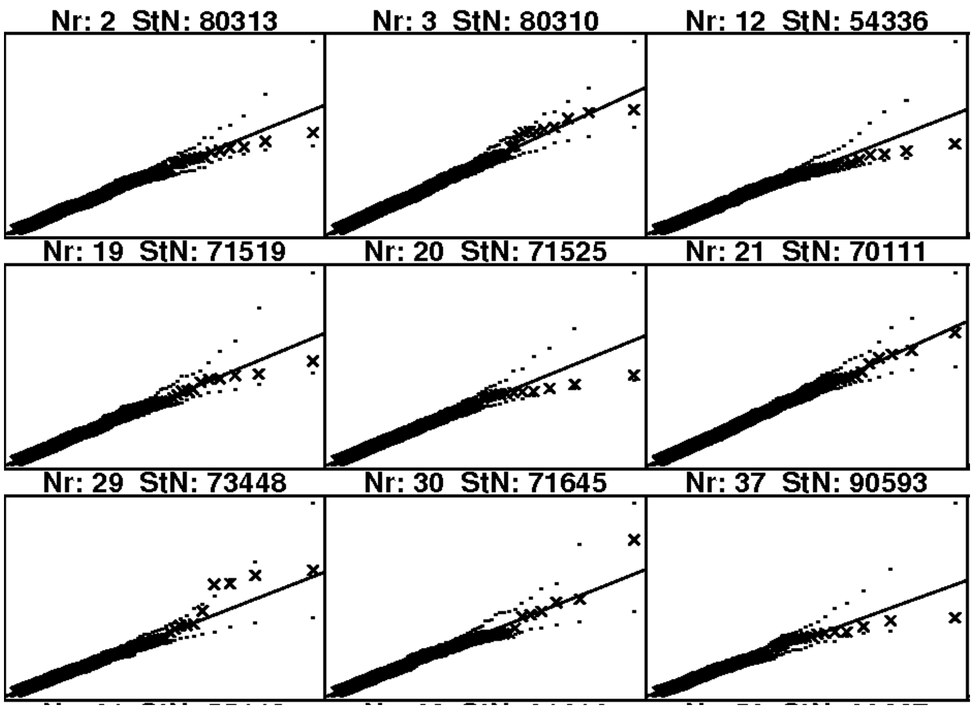




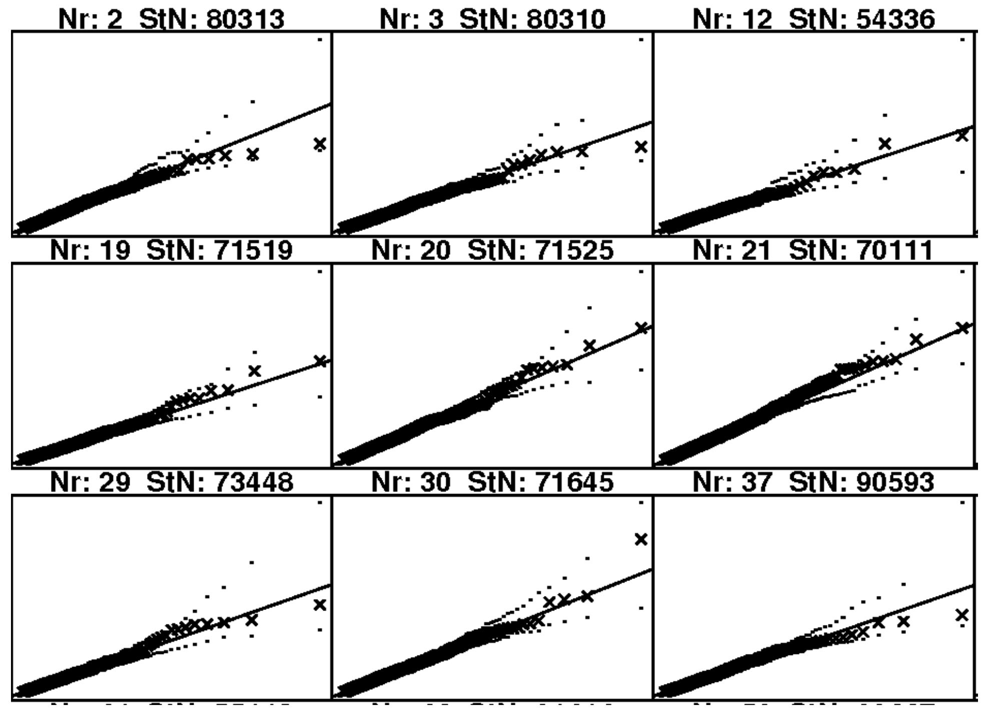
Goodness of Fit

u = 5mm

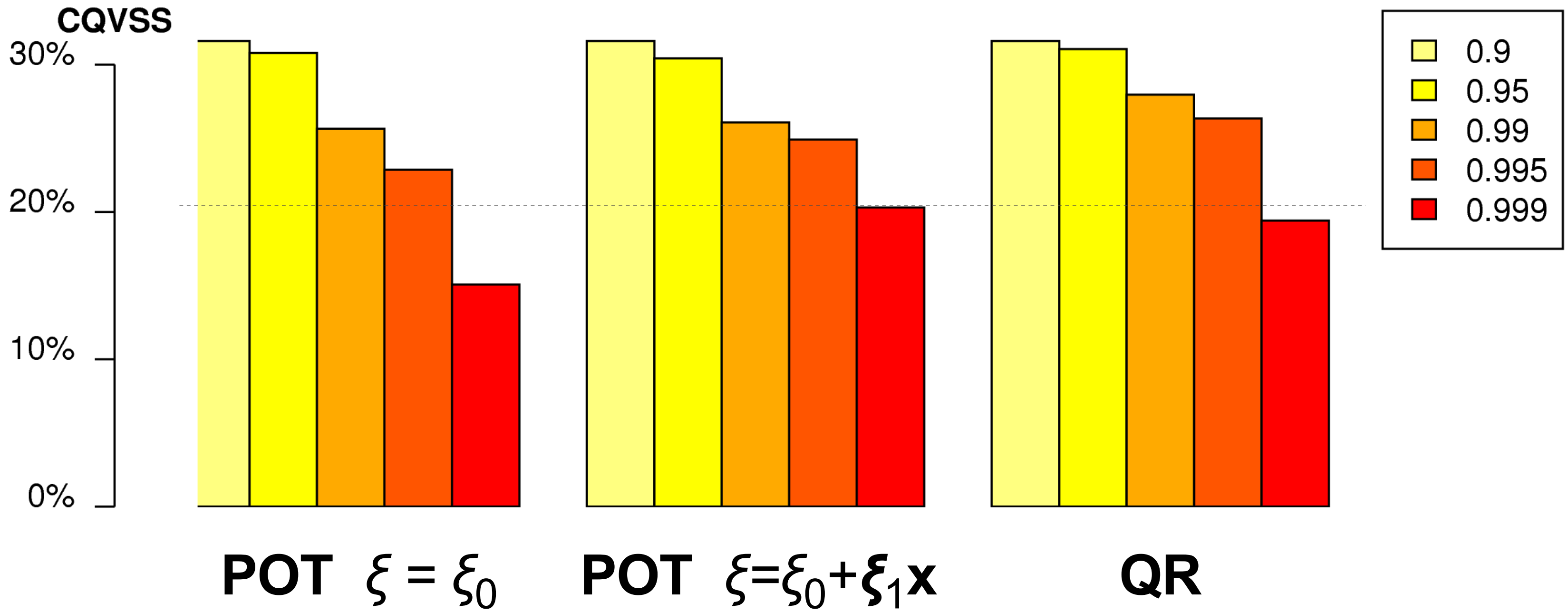
Winter



Summer



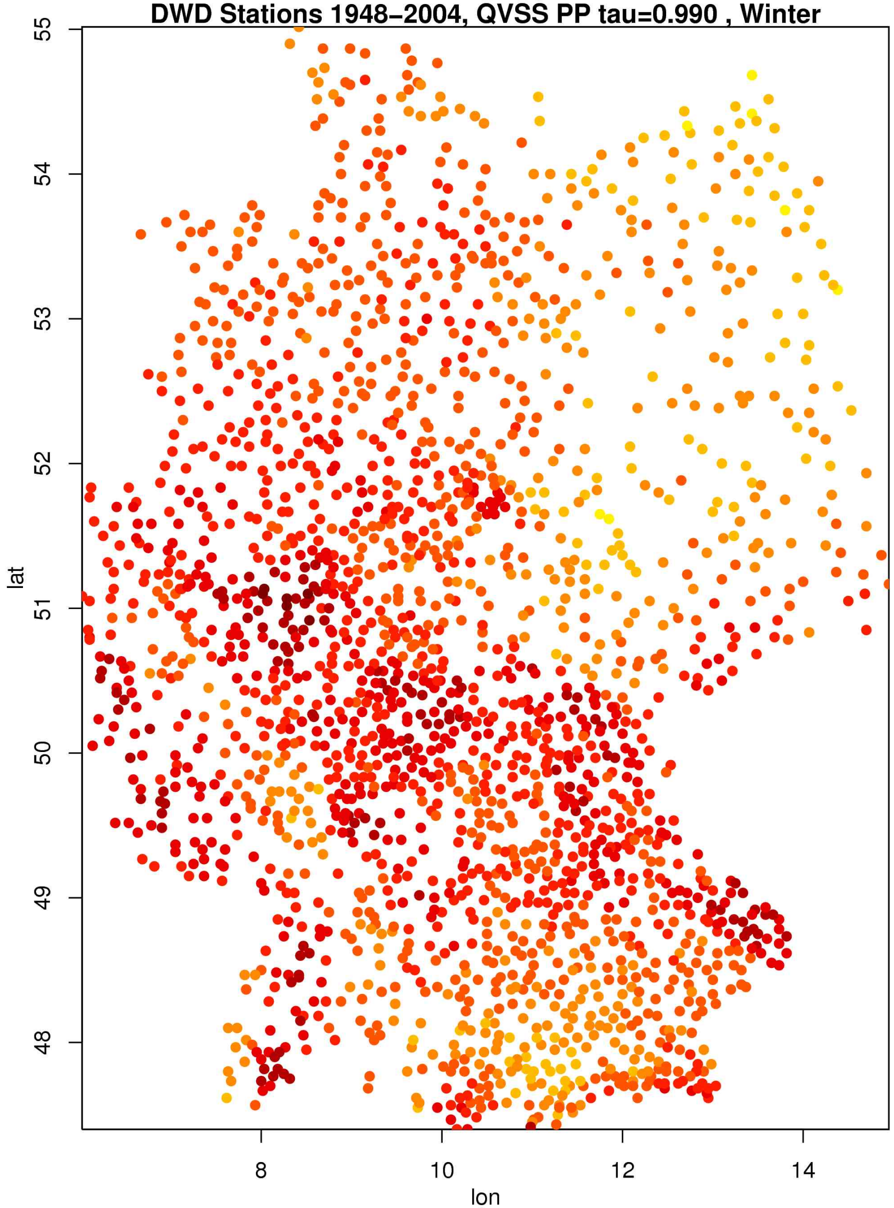
QVSS - Winter



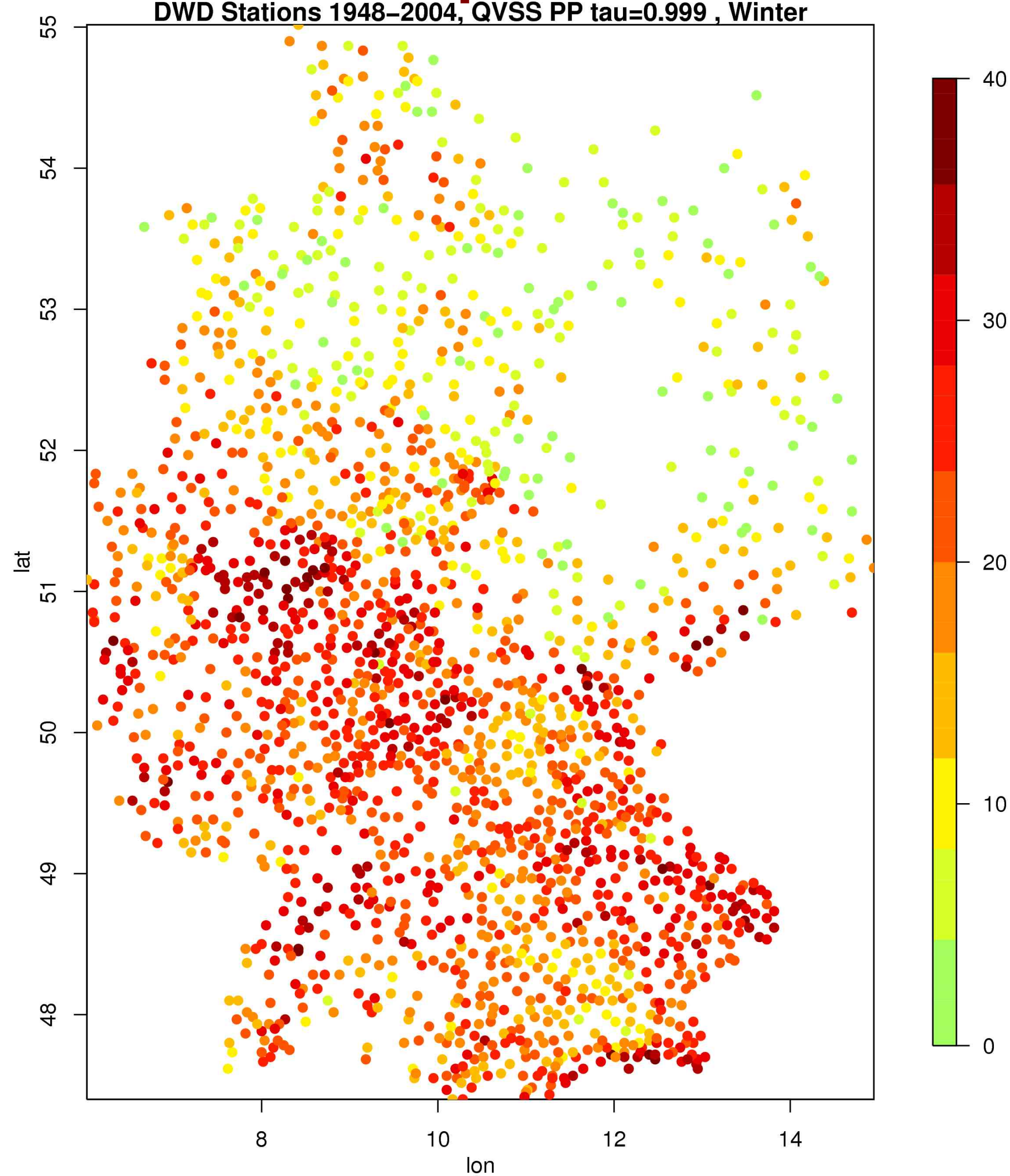
$$\tau_u = 0.9$$

QVSS - Winter all Stations

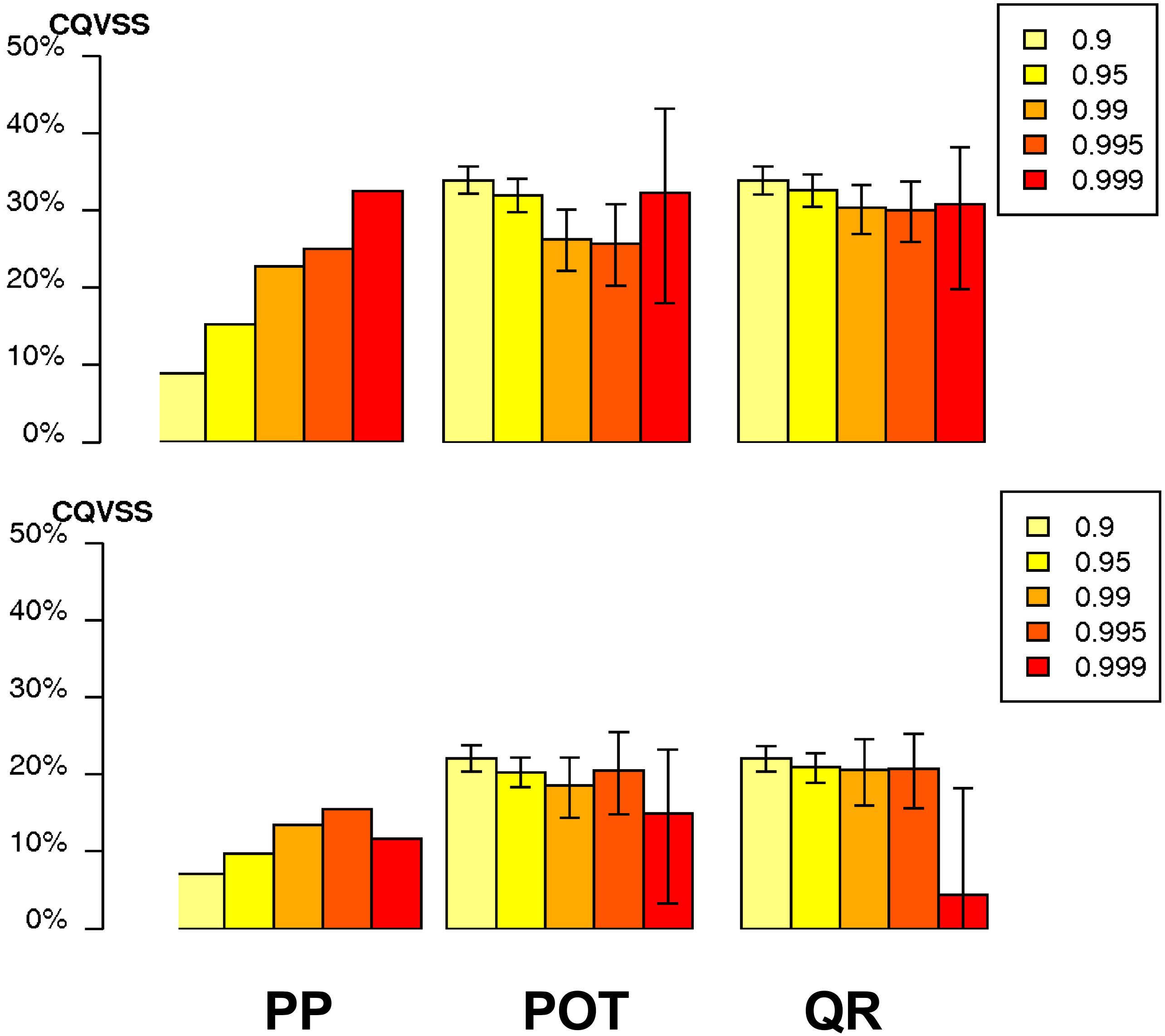
0.99 quantile



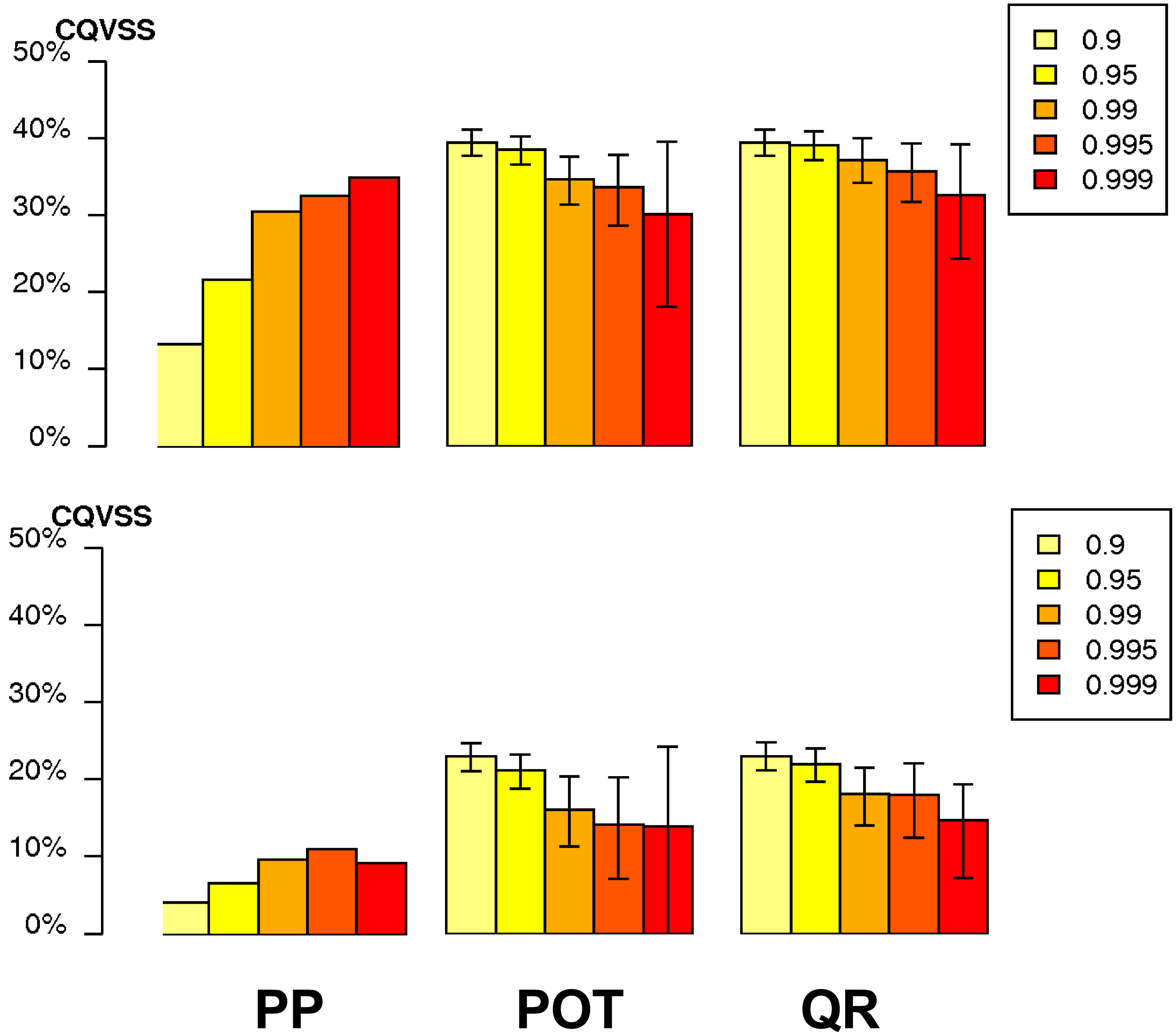
0.999 quantile



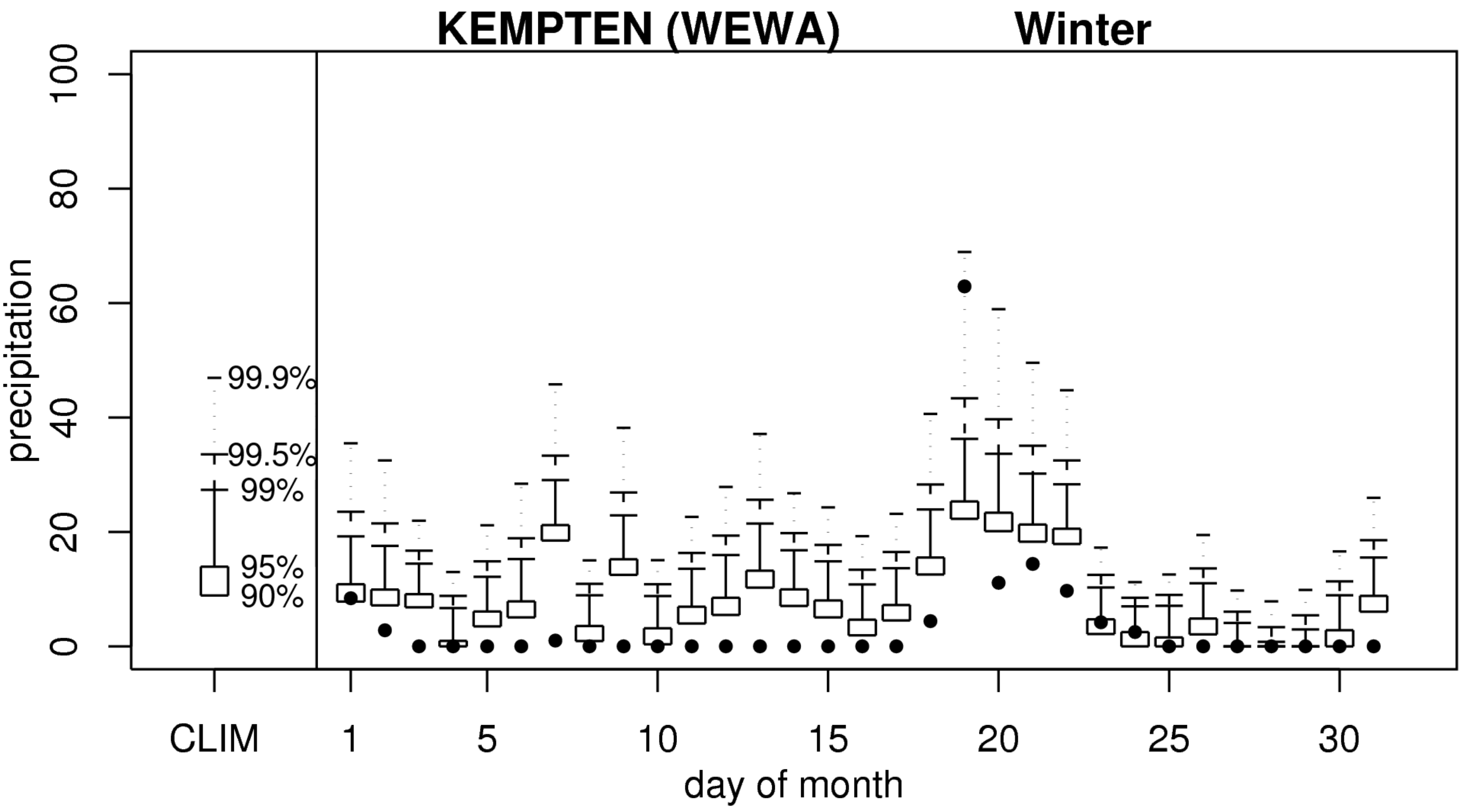
QVSS - Kempten



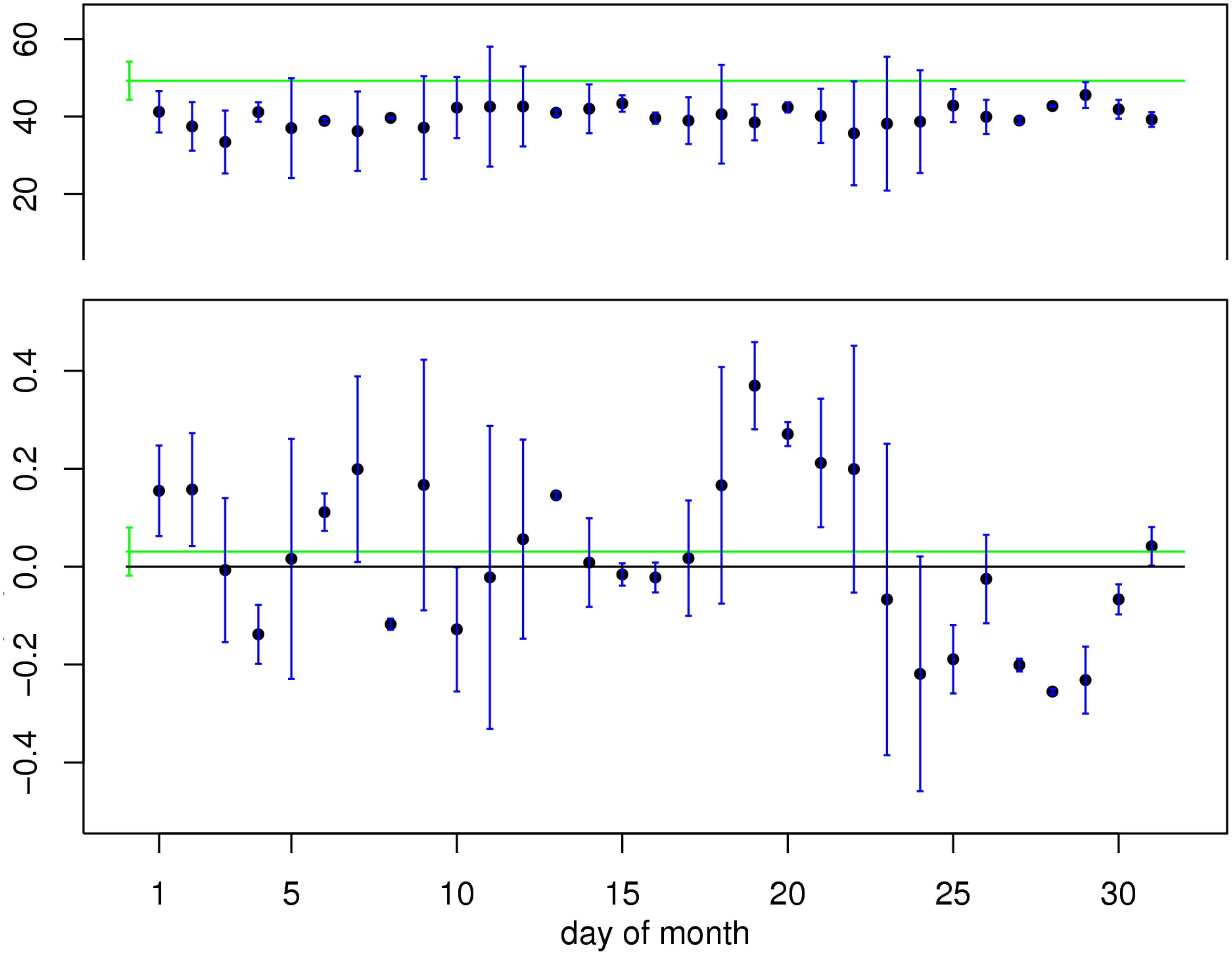
QVSS - Fichtelberg



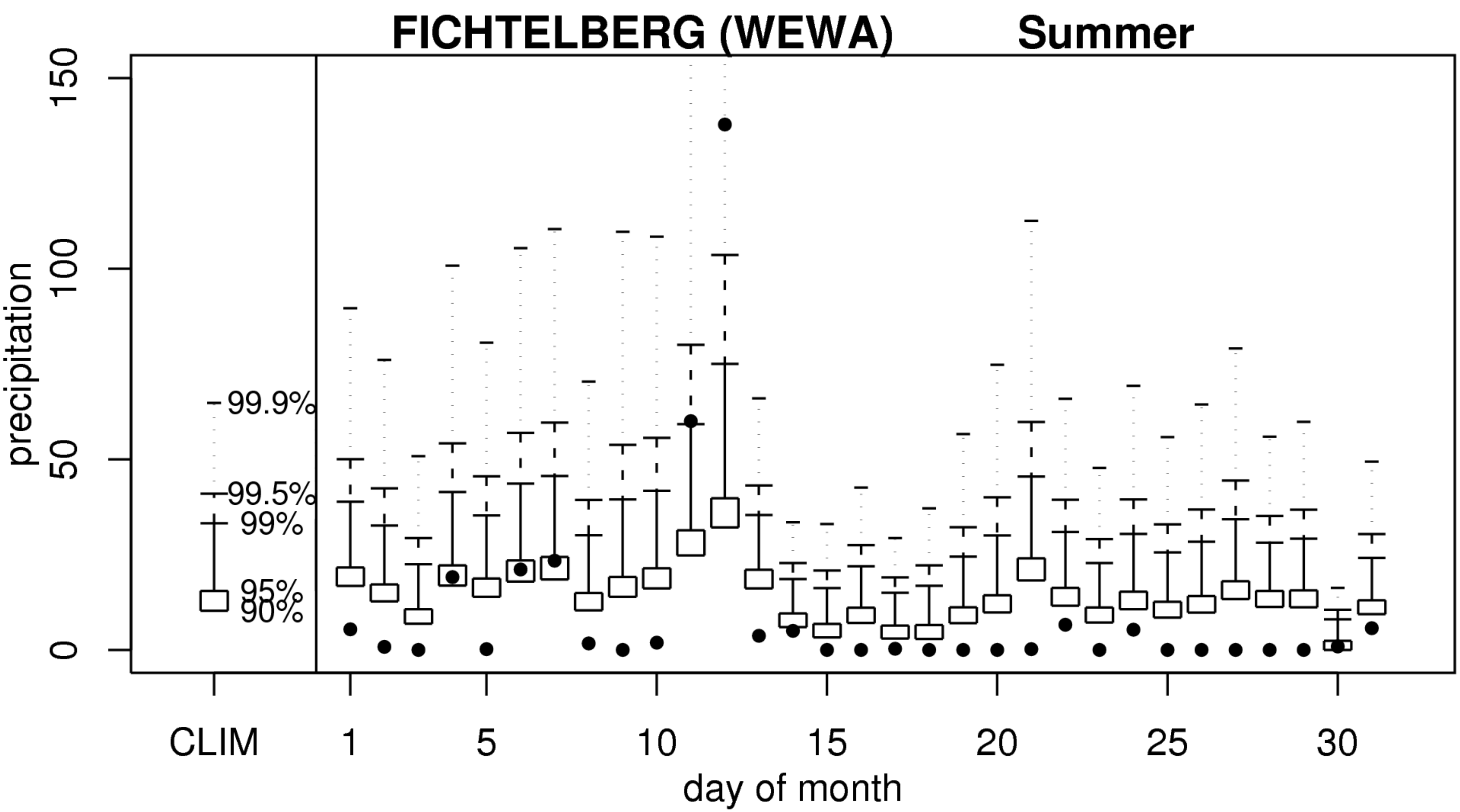
Kempten March 2002



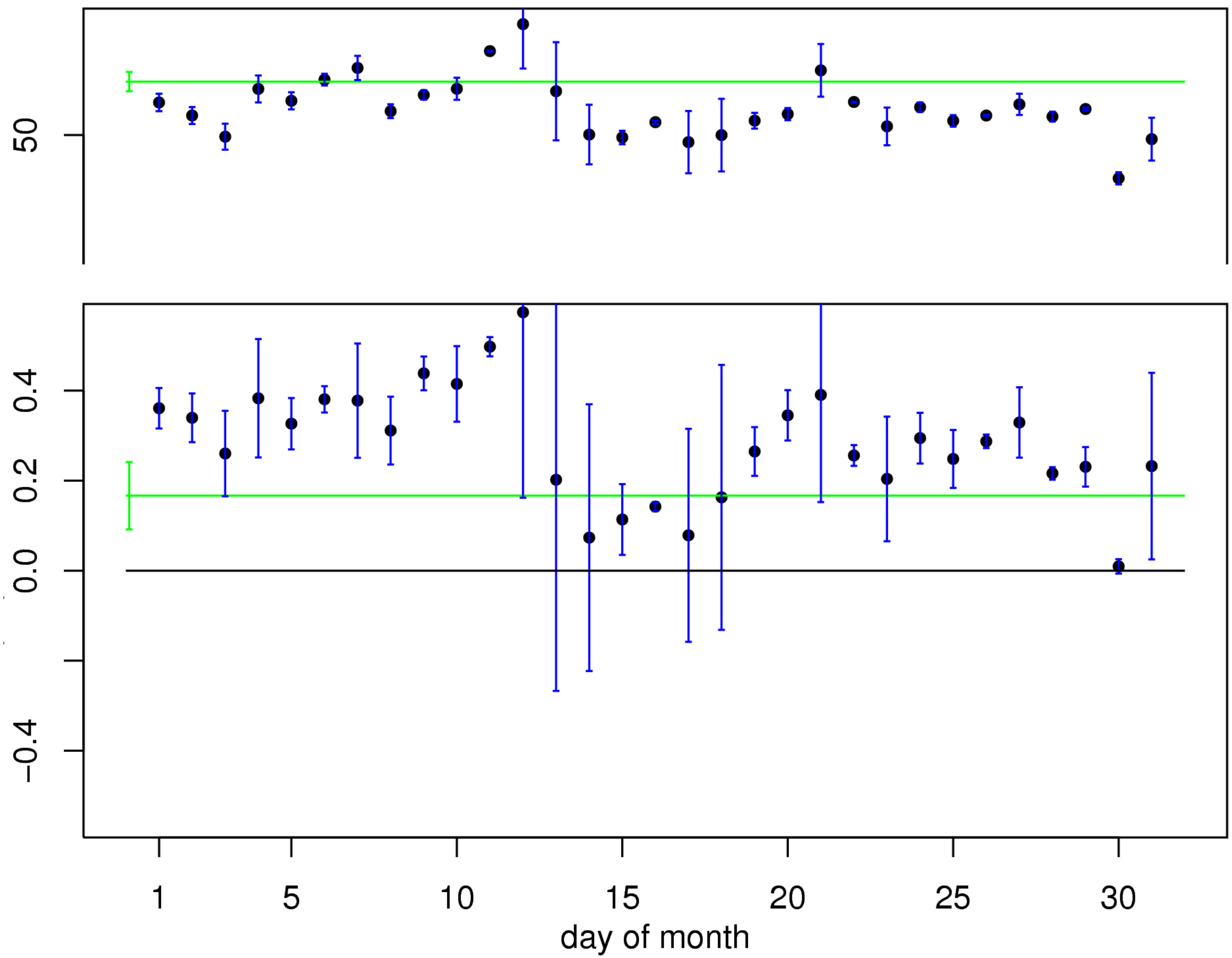
Kempten March 2002



Fichtelberg August 2002



Fichtelberg August 2002



Conclusions

Skill

more skill for winter than summer precipitation

(between20% to 40% in winter, and 10% to 30% in summer)

largest skill obtained for high quantiles

Threshold

largest skill obtained with threshold of about 2*mm*

EV parameter

shape parameter is positive in winter and summer

allowing regression for shape parameter induces large error in parameter estimates
-> statistical downscaling model becomes instable

regression for shape parameter increases skill in winter

Outlook

separate convective and stratriform precipitation

estimate quantiles taking into account spacial dependence

References

Beirlant, J. and co-authors, 2004: Statistics of Extremes. Wiley.

Bremnes, J.B., 2004: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Mon. Wea. Rev*., **132**, 338-347.

Gneiting, T. and Raftery, A.E., 2005: Strictly proper scoring rules, prediction, and estimation. University of Washington Technical Report, **463R**

Koenker, R., 2005: Quantile regression. Cambridge University Press. 349p

Friederichs, P. and A. Hense, 2007: Statistical downscaling of extreme precipitation events using censored quantile regression. *Mon. Wea. Rev.* (in press)

Friederichs, P., and A. Hense, 2007: A Probabilistic forecast approach for daily precipitation totals. *Weather and Forecasting* (submitted)