Probabilistic Seasonal Precipitation Forecasting over Africa and Europe using Quantile Regression

Petra Friederichs and Andreas Hense

Meteorological Institute
University of Bonn

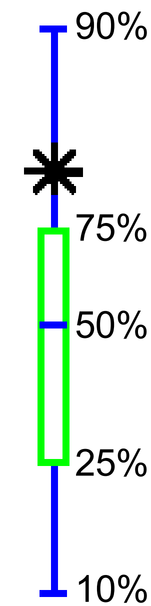
Potsdam, 13. Dezember 2005

Outline

- **Motivation**
- **Model and Data**
- **Probabilistic MOS**
  - **Quantile Regression**
  - **Proper Scoring Rule**
  - **Multivariate MOS – Data Reduction**
  - **Probabilistic MOS**
- **Conclusions**

Motivation

AIM: provide probabilistic forecasts in terms of conditional quantiles



Why:

- probabilistic forecasts for decision maker
- in terms of quantiles more intuitive for end user
- forecast of „extremal“ quantiles

What:

- quantile regression, introduced by Koenker and Bassett (1978)
- example in meteorology is Bremnes (Mon. Wea. Rev., 2004)
- classical regression derives conditional exspectation value
- conditional quantile of a response variable in linear model context

Wherefor:

- multiple regression used for Model Output Statistics (MOS)
- multiple quantile regression do derive probabilistic MOS

Model and Data

Model:

- Hamburg **ECHAM4-T42** (2.8° x 2.8°) spectral model
  from **1903-2002**
- **15 simulations** started from different initial states
- forced with **observed SST** and sea-ice boundary,
  5 with changing GHG concentration
  5 with additional GHG and solar/volcanic forcing

Data:

- **Precipitation** from Climatic Research Unit (**CRU TS 2.0**)
  monthly data from 1901-2002
  on regular **0.5° x 0.5° grid over land**

Quantile Regression

The **conditional τ - quantile** of variate Y given multivariate covariate $\boldsymbol{X}$ is modeled by

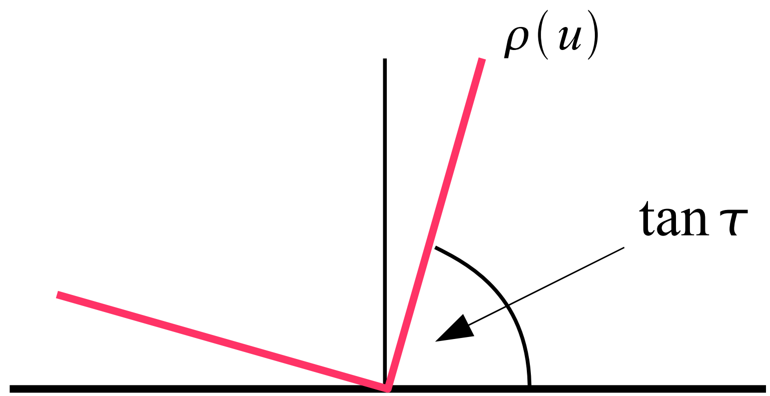
$$Q_Y(\tau|\boldsymbol{X}) = \beta_0^\tau + \sum_i \beta_i^\tau X_i + u = \boldsymbol{\beta}_\tau^T \boldsymbol{X} + u, \qquad u \in IID$$

The coefficients $\boldsymbol{\beta}_\tau$ are estimated by **minimization problem**

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

with the **check function**

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



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 $\hat{\boldsymbol{\beta}}_\tau$

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

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

(training period)

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

$$\hat{y}_n^\tau = \hat{\boldsymbol{\beta}}_\tau^T \boldsymbol{x}_n$$

Proper Scoring Rule

Forecast verification using cost function of QR

$$CQS(\tau, \hat{\boldsymbol{y}}, \boldsymbol{y}) = \sum_n \tau (y_n - \hat{y}_n^\tau) H_{y_n > \hat{y}_n^\tau} + (\tau - 1)(y_n - \hat{y}_n^\tau) H_{y_n \leqslant \hat{y}_n^\tau}$$

$H_{y_n > \hat{y}_n^\tau}$ is heavy-side function

This is a **proper scoring rule** (Gneiting and Raftery, 2004).

If $CQS(\tau, \boldsymbol{r}, \boldsymbol{y})$ is a scoring rule, $\boldsymbol{s}$ best forecast, $\boldsymbol{r}$ any forecast, then

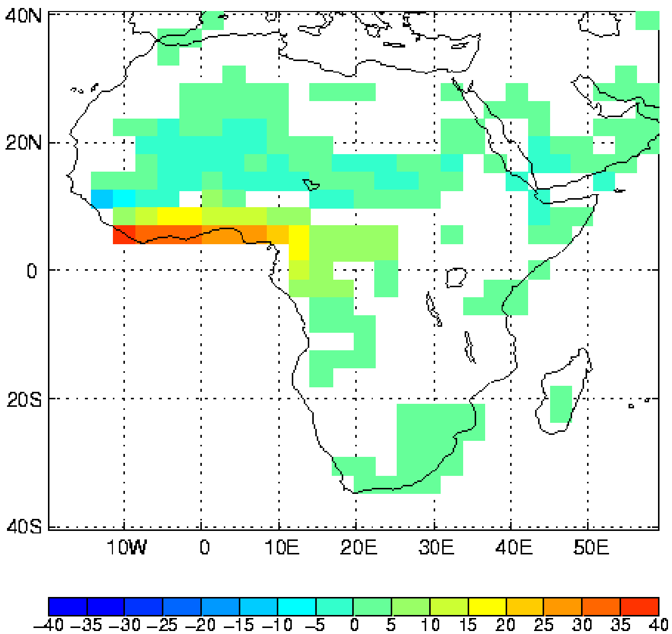
$$CQS(\tau, \boldsymbol{s}, \boldsymbol{y}) \leqslant CQS(\tau, r, \boldsymbol{y}) \quad \forall \boldsymbol{s} \neq \boldsymbol{r}$$

Define **Skill Score**

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

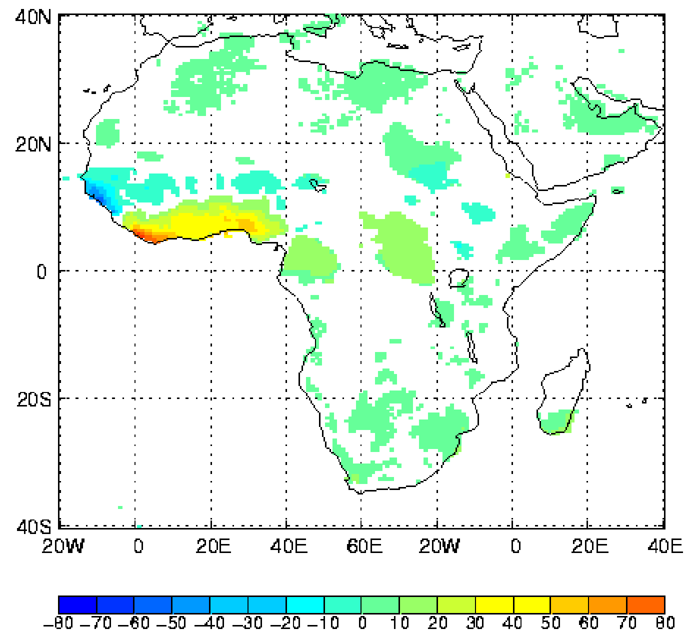
Multivariate MOS – Data Reduction

ECHAM4 precipitation (predictor)
ensemble mean seasonal anomalies over Africa
> data reduction needed



$\rightarrow \boldsymbol{p}_k$

CRU precipitation (predictand)
> gridded CRU precipitation anomalies
> univariate transforms
(e.g. area mean precipitation)



$\rightarrow \boldsymbol{q}_k$

MOS re-calibration using CCA
> CCA (Canonical Correlation Analysis): finds linear transforms between predictor and predictand that maximizes correlation

Multivariate MOS – Data Reduction

Training period

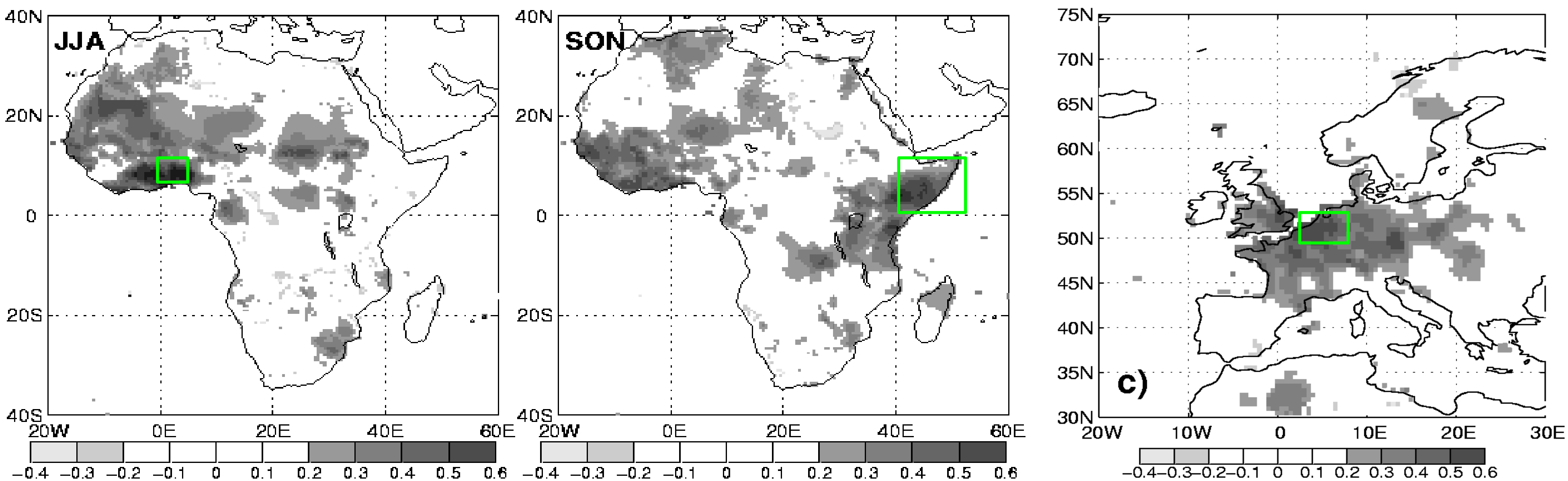
canonical patterns $\boldsymbol{p}_k, \boldsymbol{q}_k$

canonical correlations ρ_k

and covariance matrix $\boldsymbol{\Sigma}_{xx}^{-1}$ are calculated for the training period

Forecast for target season

$$\boldsymbol{Y}_{for} = \sum_k \rho_k \alpha_k \boldsymbol{q}_k \quad \text{with} \quad \alpha_k = \boldsymbol{p}_k^T \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{X}_{tar}$$

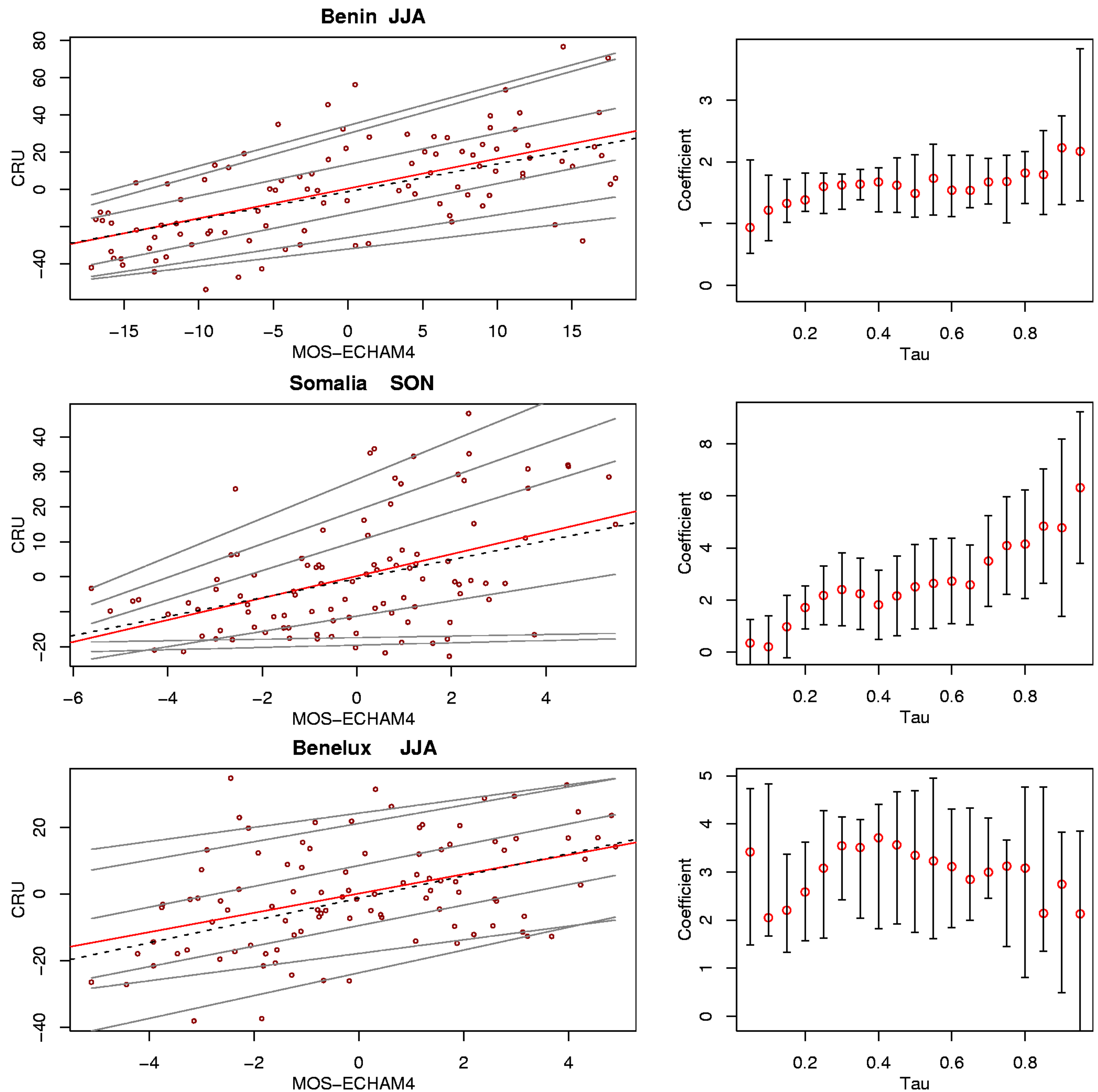


Probabilistic Forecast

$$\hat{y}_\tau = \hat{\beta}_0^\tau + \hat{\beta}_1^\tau y_{for}$$

Predictor: y_{for}
area mean
MOS re-calibrated
precipitation

Predictant: y
area mean
CRU-precipitation



Probabilistic MOS

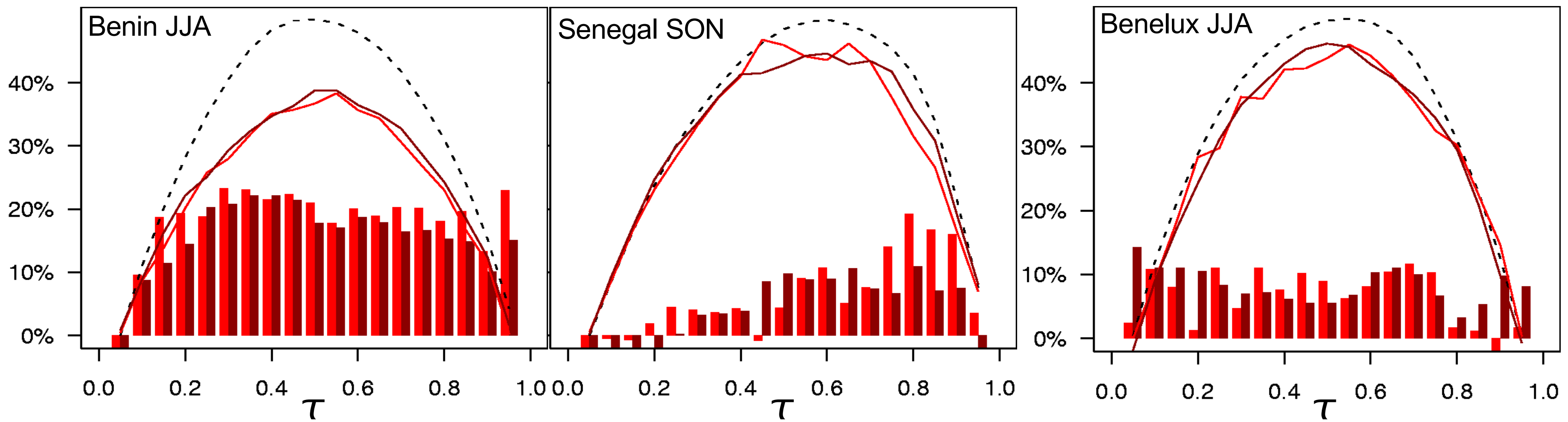
$$\hat{y}_\tau = \hat{\beta}_0^\tau + \sum_k \hat{\beta}_k^\tau \alpha_k \qquad \alpha_k = \boldsymbol{p}_k^T \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{X}_{tar}$$

Predictor: $\boldsymbol{\alpha}$
reduced ECHAM4
ensemble mean precipitation

Predictant: $\boldsymbol{y}$
area mean
CRU-precipitation

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

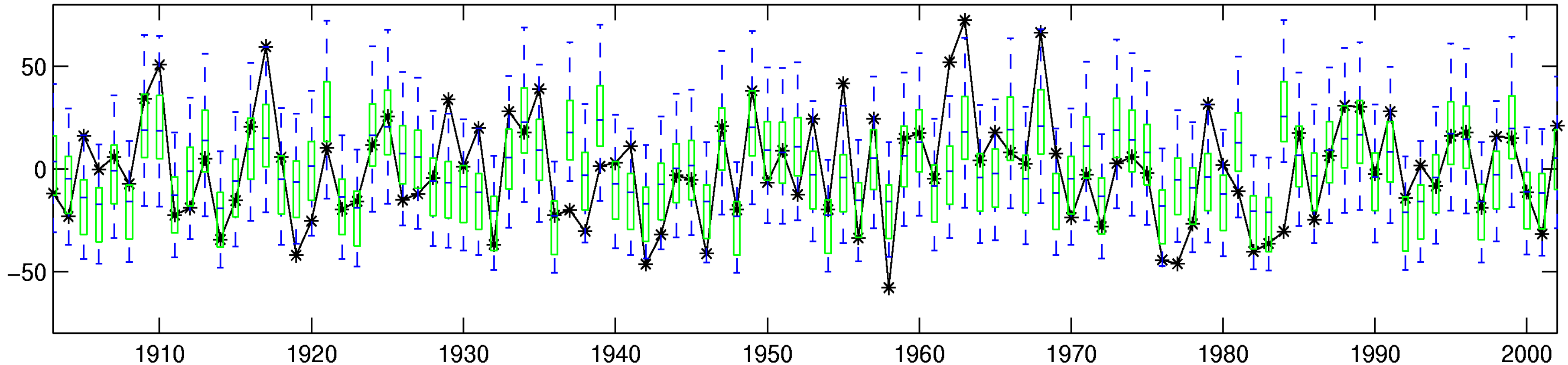
$\boldsymbol{y}_{ref}$ is climatological τ - quantile



Conclusions

Probabilistic seasonal precipitation forecasts using quantile regression > probabilistic MOS

- no strong assumptions
- for dependent time series:
  - cross-validation
  - resampling methods
- proper scoring rule to assess forecast skill
- forecast „extreme“ quantiles > 5% or 95% - quantile



References

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

(2) Gneiting, T. and Raftery, A.E., 2004: Strictly proper scoring rules, prediction, and estimation. University of Washington TechnicalReport, **463**, available at: http://www.stat.washington.edu/www/research/reports/2004/tr463.pdf.

(3) Koenker, R., 2005: Quantile Regression. Econometric Society Monographs. Cambridge, 349p.

Thank you!