Extreme weather and probabilistic forecast approaches

Petra Friederichs
Sabrina Bentzien, Andreas Hense

Meteorological Institute, University of Bonn

Statistical Methods for Meteorology and Climate Change
Montreal, 12 – 14 January 2011

Outline

Extreme weather and probabilistic prediction

- Atmospheric scales and mesoscale atmospheric dynamics
- High-impact weather
- Ensemble prediction systems

Ensemble post-processing

- 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:**

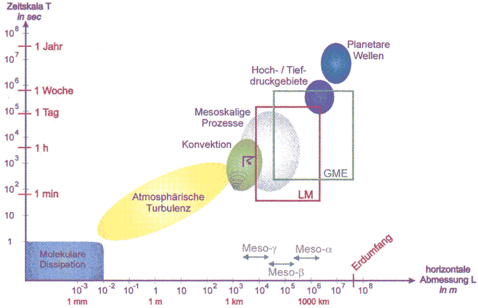
  Rare, exceptional, "large" and **high impact**
- **Problems:**
  - 95% quantile of daily precipitation: $\approx 10 - 15 mm/d$
  - $\approx$ 2 years of data – only few extremes events for verification

Mesoscale Weather Prediction

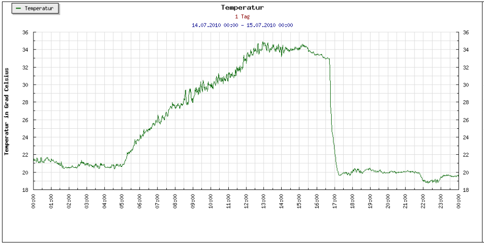
- Strong and disastrous impact of many weather extremes calls for reliable forecasts
- "Although forecasters have traditionally viewed weather prediction as deterministic, a cultural change towards probabilistic forecasting is in progress." (T N Palmer, 2002)
- Weather extremes do not come **"Out of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

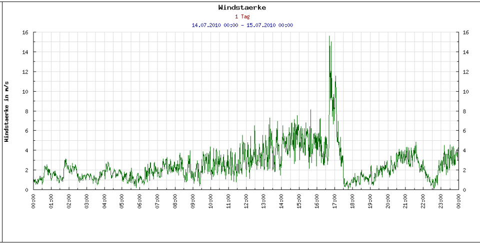
Atmospheric scales and mesoscale dynamics

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



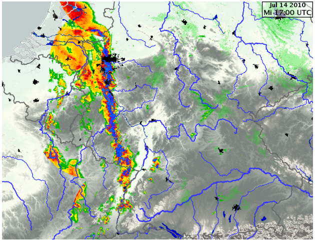
Steinhorst, Promet 35, 2010





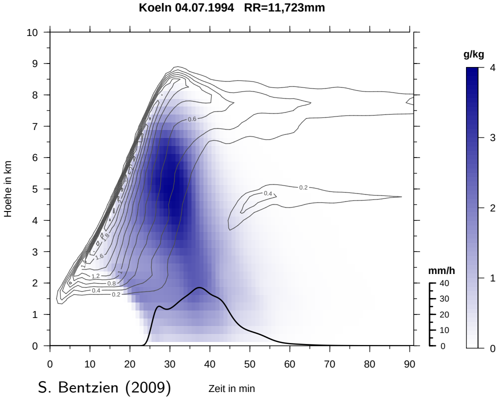
Mesoscale weather extremes

- 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

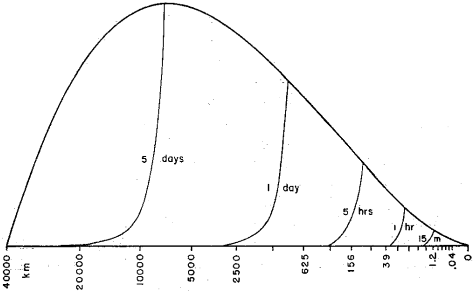
- Large vertical gradients of entropy
- Convective instability
- Deep convection lead to extremal vertical velocities
- Heavy precipitation and hailstones grow within this vertical circulation



S. Bentzien (2009)

Predictability

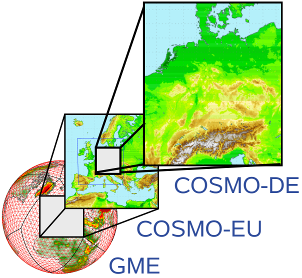
- Inherent limit of predictability
- Fastes error growth at smallest scales
- Predictability strongly depends on flow regime
- Moist convection is primary source of forecast-error growth
- Mesoscale forecasts are issued for $\leq$18h-24h

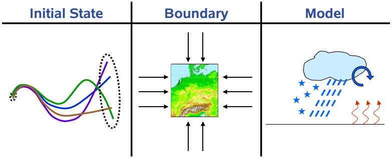


E N Lorenz, Tellus 21, 19 (1969)

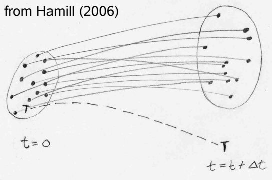
COSMO-DE Ensemble Prediction System (EPS)

- 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





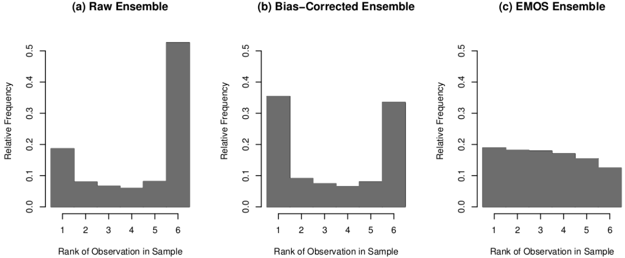
S. Theis, DWD (2010)



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

Calibration:

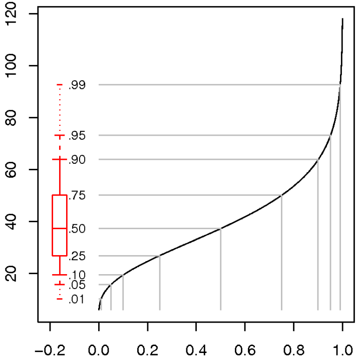
Raw ensemble data need adjustmend: biased and underdispersive



Gneiting et al. (2005)

Sharpness: Information of forecasts

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

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

Quantile regression

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

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

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 $u < 0$

Censoring

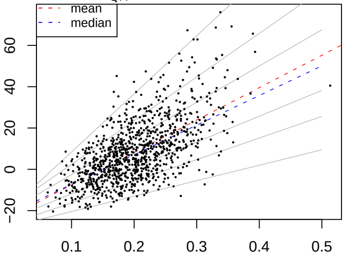
Equivariance with respect to non-decreasing function $h(\cdot)$

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

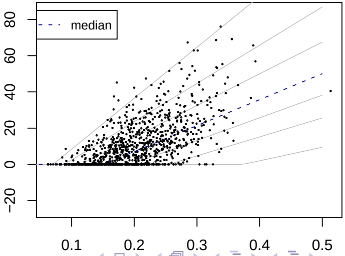
Hidden process Y^* observed through censored variable Y

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

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



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



Generalized linear model – Mixture model

- Probability of precipitation ($Y \geq 0.1mm$) – Logistic regression

$$Pr(Y \geq 0.1 \mid \mathbf{x}) = \pi(\mathbf{x})$$

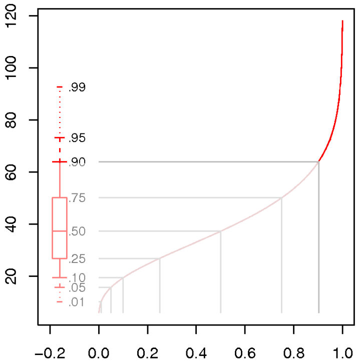
- Distribution of precipitation – Gamma GLM

$$F(Y \mid \mathbf{x}, Y \geq 0.1) = G_\Gamma(Y; \Theta(\mathbf{x}))$$

Conditional mixture model for precipitation

$$F_{Y_{mix}}(y \mid \mathbf{x}) = (1 - \pi) + \pi \, G_\Gamma(y; \Theta(\mathbf{x})) \, \mathrm{I}_{y \geq 0.1}$$

Conditional mixture model



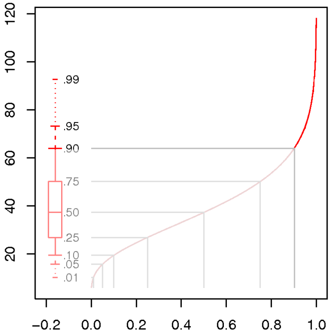
Parametric – 'normal'

- A-priori assumption about parametric distribtion for 'normal' part $F_{Y|\mathbf{x}}(y|\mathbf{x}) = G(y; \Theta(\mathbf{x}))$
  Estimate parameter function $\Theta(\mathbf{x})$
  Generalized linear model (GLM)

Parametric: 'extreme'

- Above threshold/quantile: 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"



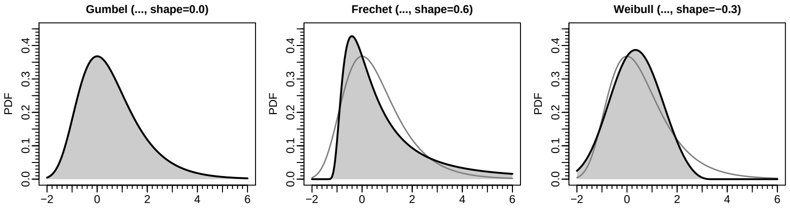
- **Limit theorem** for sample maxima
  → asymptotic distribution for extremes
- Condition of **max-stability** (de Haan, 1984)
  → maxima follow a generalized extreme value distribution
- Garantees **universal behavior of extremes**
  → enables extrapolation!

In praxis: often not enough data to reach asymptotic limit

Extreme value distribution

Generalized extreme value distribution (GEV)

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



Generalized Pareto distribution

Threshold excesses $Z = Y - u$ follow a GPD

$$Prob(Z \leq y - u) = \begin{cases} 1 - \left(1 + \xi \frac{y-u}{\sigma_u}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y-u}{\sigma_u}\right), & \xi = 0 \end{cases},$$

for $y - u > 0$ and $(1 + \xi \frac{y-u}{\sigma_u}) > 0$

Mixture GLM including Extremes

- Conditional mixture model for precipitation

$$F_{Y_{mix}}(y \mid \mathbf{x}) = (1-\pi) + \pi\, G_\Gamma(y;\Theta(\mathbf{x}))\ \mathrm{I}_{y\geq 0.1}$$

- GPD for 'extreme' precipitation – above $u_\tau = F_Y^{-1}(\tau \mid \mathbf{x})$

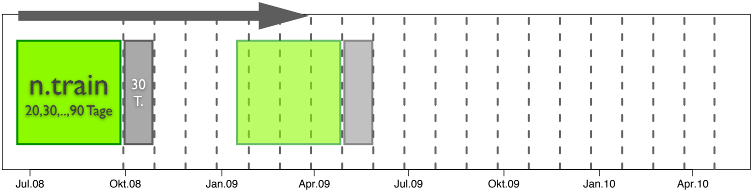
$$F(Y \mid \mathbf{x}, Y > u_\tau) = G_{GPD}(Y - u_\tau; \sigma(\mathbf{x}))$$

Conditional mixture model for precipitation with extremes

$$\begin{aligned} F_{Y_{GPD}}(y \mid \mathbf{x}) &= (1-\pi) + \pi\, G_\Gamma(y;\Theta(\mathbf{x}))\ \mathrm{I}_{Y\geq 0.1, Y\leq u_\tau} \\ &+ (1-\tau) G_{GPD}(Y - u_\tau; \sigma(\mathbf{x})\ \mathrm{I}_{Y>u_\tau} \end{aligned}$$

Prediction and Verification

- Model parameter training
- Verification on independent data



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

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

Quantile verification skill score

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

Log-likelihood ratio test: asymmetric Laplacian regression

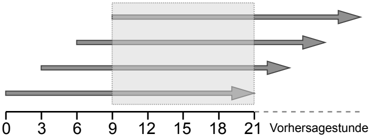
$$f_\tau(u) = \frac{\tau(1-\tau)}{\sigma_L} \exp(-\rho_\tau(u)/\sigma_L).$$

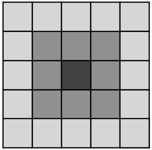
proportional to $\log(\text{QVS}(\tau)/\text{QVS}_{ref}(\tau))$

SKeleton EPS Interim Solution – Neighborhood method

- 'Time-lagged' ensemble of COSMO-DE
- Initialized every 3 h, forecasts for 21 h
- 1 July 2008 – 30 April 2010

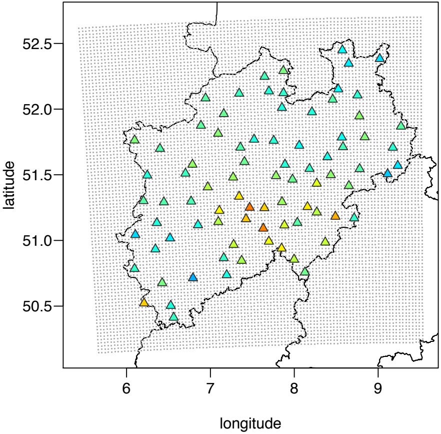
- First guess probability (fgp)
- First guess 0.9-quantile (fgq9)
- 4 COSMO-DE forecasts and 5 $\times$ 5 neighborhood



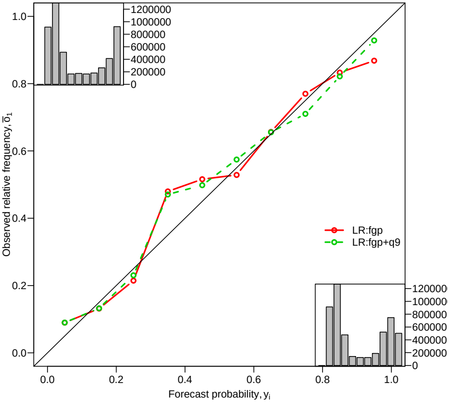


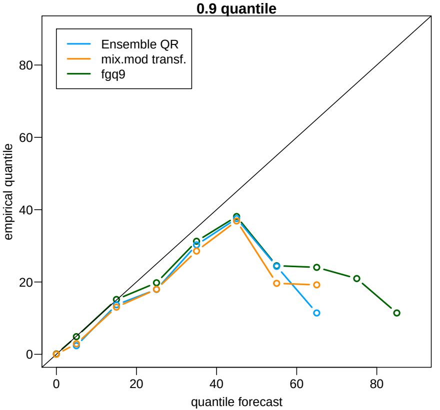
12h accumulated precipitation between 12 and 00 UTC

- 83 Station in NRW
- Rain radar composite

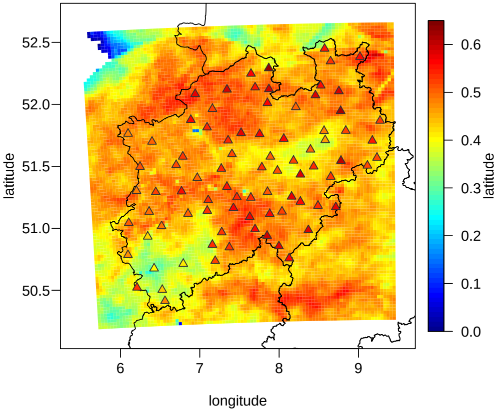


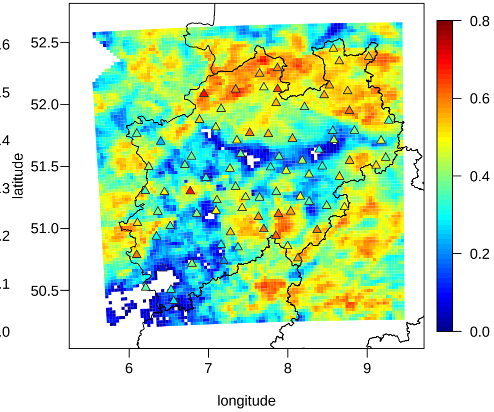
Reliability



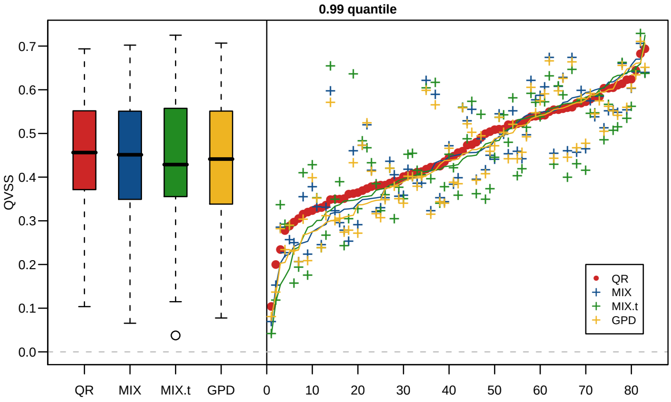


Quantile forecasts - Skill Score

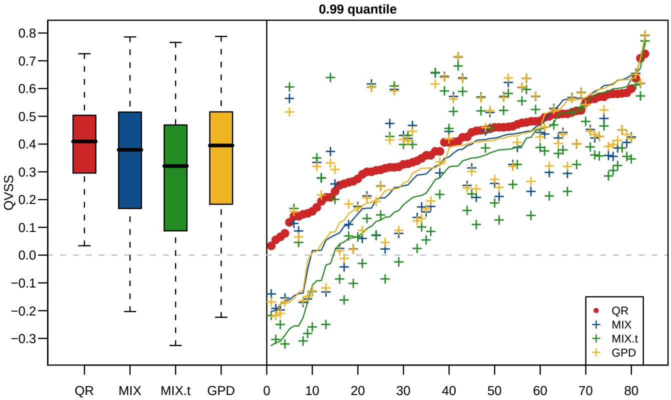


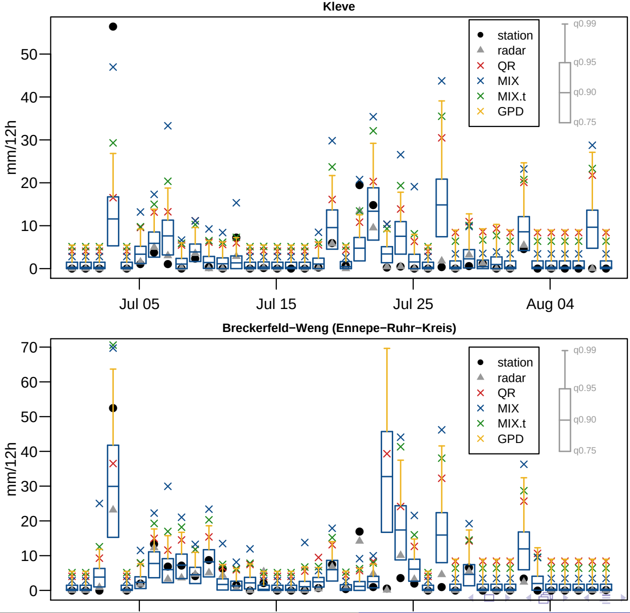


Quantile forecasts - Skill Score

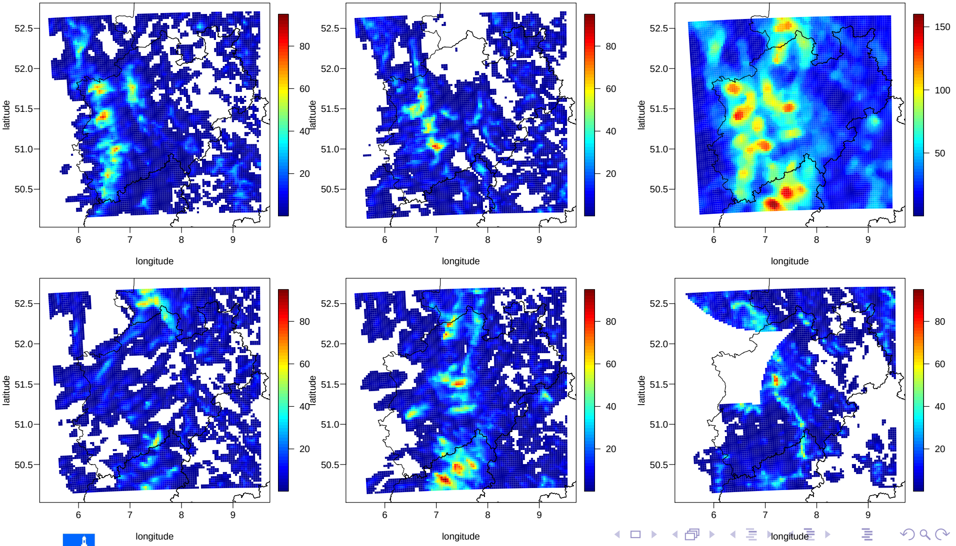


Quantile forecasts - Skill Score




Kleve
mm/12h
0
10
20
30
40
50
Jul 05
Jul 15
Jul 25
Aug 04
station
radar
QR
MIX
MIX.t
GPD
q0.99
q0.95
q0.90
q0.75
Breckerfeld−Weng (Ennepe−Ruhr−Kreis)
mm/12h
0
10
20
30
40
50
60
70
station
radar
QR
MIX
MIX.t
GPD
q0.99
q0.95
q0.90
q0.75

Lagged ensemble forecasts and radar measurements



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 predictions
  - Combine spatial statistics with model post-processing (Berrocal et al., 2007)
  - Develop methods for multivariate post-processing
  - Develop novel ensemble methods tailored to extremes (Bayesian model averaging)
- Verification tailored to extremes
- Verification for probabilistic multivariate and spatial forecasts

Reference

- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133, 1098-1118.
- C. Gebhardt, S.E. Theis, M. Paulat, and Z. Ben Bouallègue, 2010: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmos. Res. (in press).
- Friederichs, P., 2010: Statistical downscaling of extreme precipitation using extreme value theory. Extremes 13, 109-132.

Special thanks to:

Susanne Theis, Martin Göber, Deutscher Wetterdienst, Offenbach

Thank You for Your Attention!