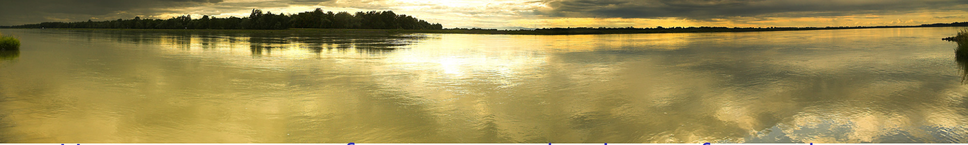


How to use new information technologies for prediction: Ensemble flow forecasting, verification and postprocessing

Maik Renner
Micha Werner*, Albrecht Weerts*, Paolo Reggiani*

TU Dresden / Germany – Institute of Hydrology and Meteorology
* Deltares, Delft The Netherlands

PUB 2011 workshop - Putting PUB into practise
Canmore, Canada - 14th May 2011

Are ensembles a luxury or a prerequistite?

Are ensembles a luxury or a prerequistite?

1. The future is uncertain!

Are ensembles a luxury or a prerequistite?

1. The future is uncertain!
2. Models are not reality!

Are ensembles a luxury or a prerequistite?

1. The future is uncertain!
2. Models are not reality!
3. But:
   - Our models may be close by.

Are ensembles a luxury or a prerequistite?

1. The future is uncertain!
2. Models are not reality!
3. But:
    - Our models may be close by.
    - We can try to express our certitude of a prediction with a probability.

Are ensembles a luxury or a prerequistite?

1. The future is uncertain!
2. Models are not reality!
3. But:
   - Our models may be close by.
   - We can try to express our certitude of a prediction with a probability.

   → decision can be made by a decision maker instead of the forecaster (Weerts et al. 2011)

Construction of an ensemble forecast:

- (ac)knowlegde uncertainties
  - input data, model simulation, model structure, future conditions
- simulate from different initial states – meteorological ensemble forecasts
- sample probability distributions: e.g. parameter distributions of a hydrologic model, sample inputs to the model

Expectations

- expect that the ensemble is a good representation of the predictive uncertainty

$\rightarrow$ the ensemble is drawn from the same distribution as the uncertainties

Research questions

1. Relevant sources of uncertainty for the given objective?

Research questions

1. Relevant sources of uncertainty for the given objective?
    - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)

Research questions

1. Relevant sources of uncertainty for the given objective?
   - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?

Research questions

1. Relevant sources of uncertainty for the given objective?
   - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?
3. How to compare forecasts?

Research questions

1. Relevant sources of uncertainty for the given objective?
    - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?
3. How to compare forecasts?
4. How to improve ensemble forecasts given observed data?

Research questions

1. Relevant sources of uncertainty for the given objective?
   - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?
3. How to compare forecasts?
4. How to improve ensemble forecasts given observed data?
   - Postprocessing methods

Research questions

1. Relevant sources of uncertainty for the given objective?
    - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?
3. How to compare forecasts?
4. How to improve ensemble forecasts given observed data?
    - Postprocessing methods
    - Data assimiliation

Research questions

1. Relevant sources of uncertainty for the given objective?
   - such objectives are e.g. : the forecast horizon and the respective basin (travel times, etc...)
2. Is an ensemble prediction accurate?
3. How to compare forecasts?
4. How to improve ensemble forecasts given observed data?
   - Postprocessing methods
   - Data assimiliation
5. Transferability of forecast accuracy towards ungauged basins?

Outline

Ensemble (probabilistic) forecast verification

- How well agree ensemble forecast with observed data?
  - water levels
  - discharge
  - economic losses
- Identify problems and improve your forecast!
- active research and application in meteorology
  - Wilks (2006), WMO (2010), mailing list `vx-discuss@rap.ucar.edu`

An ensemble flow forecast

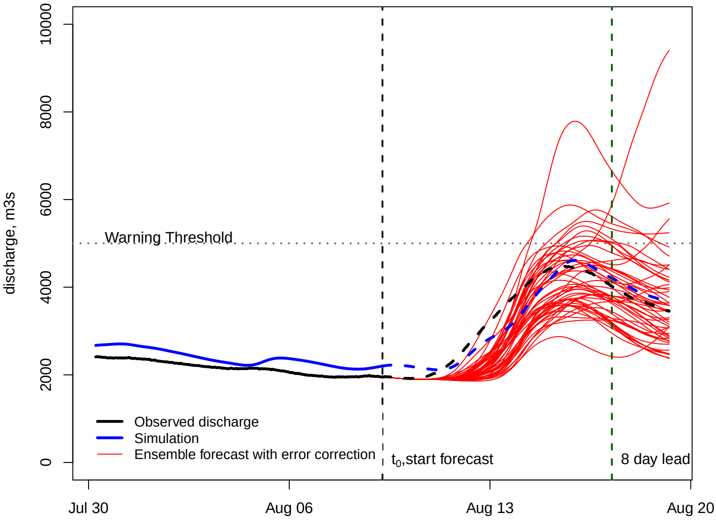


Figure: One ensemble forecast, with observations and threshold

Scalar Accuracy Measures

mean absolute error MAE of a set of m forecasts and observations:

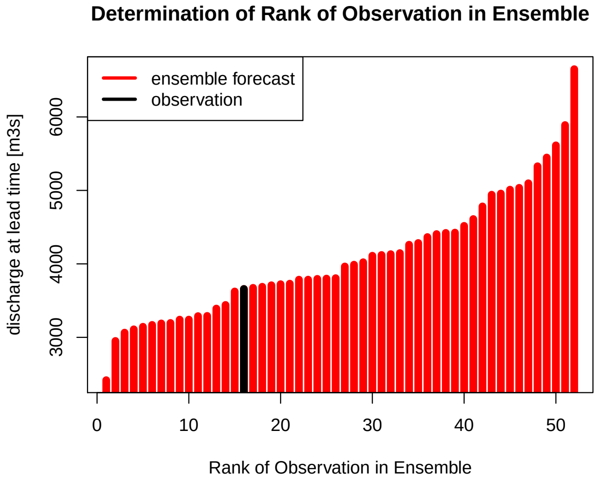
$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - o_i| \tag{1}$$

y_i is the deterministic forecast for issued at day i, for ensembles the ensemble mean and sometimes the median is used, probabilistic forecast the median is used

does not consider the ensemble spread!

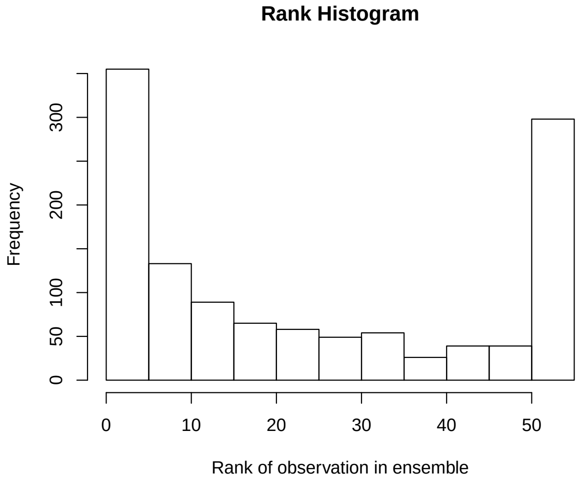
Reliability of an ensemble forecast - Rank Histogram

Considers the question if the ensemble is drawn from the same distribution as the predictive uncertainty

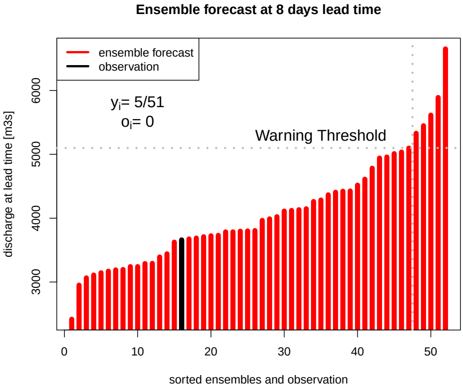


Reliability of an ensemble forecast - Rank Histogram

Considers the question if the ensemble is drawn from the same distribution as the predictive uncertainty



Threshold scores - The Brier score



- Transform ensemble and observations into probabilities (of an event)
- Goes back to Brier (1950)
- Analogue to the mean squared error for probabilites

$$BS = \frac{1}{N}\sum_{i=1}^{m}(y_i - o_i)^2$$

Ranked Probability Score

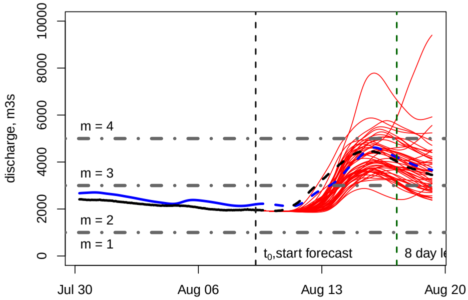


Figure: Definition of 4 prediction categories.

- extension to the Brier Score for several categories
- considers the distance of the forecast to the observations
- general accuracy of a probabilistic forecast
- negatively orientated, 0 best

$$RPS = \frac{1}{M-1} \sum_{m=1}^{M} \left[\left(\sum_{i=1}^{m} y_i \right) - \left(\sum_{i=1}^{m} o_i \right) \right]$$

Skill scores

- skill compared to a reference forecast e.g. climatology, persistence

$$\mathbf{Skillscore} = \mathbf{1} - \frac{\boldsymbol{Score}}{\boldsymbol{Score_{reference}}} \qquad (3)$$

- $\mathbf{1}$ ... perfect skill
- $\mathbf{0}$ ... no skill, i.e. equivalent
- $< \mathbf{0}$ ... less skill than reference

Flow forecasting at River Rhine

Aim: Medium range flow forecasting $\rightarrow$ 2 to 10 days lead time

river navigation

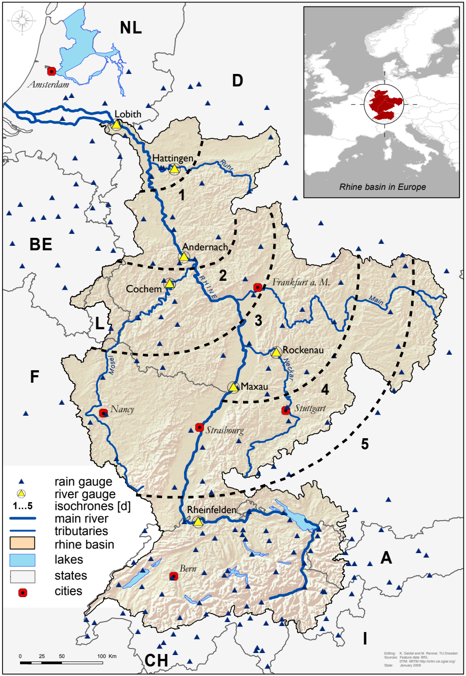
Length 1,233 km (766 mi)

Basin 170,000 km² (65,637 sq mi)

Discharge - average 2,000 m3/s (70,629 cu ft/s)

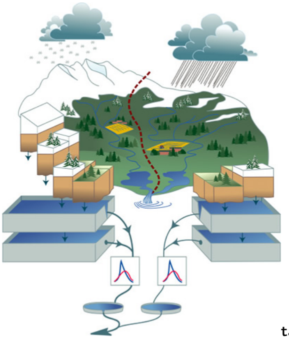
Details may be found in:

Renner, M., Werner, M., Rademacher, S. and Sprokkereef, E.: 2009, Verification of ensemble flow forecasts for the River Rhine, *Journal of Hydrology* **376**(3-4), 463–475.

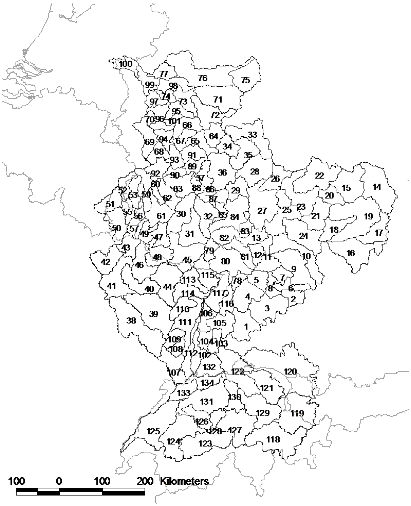


Model scheme

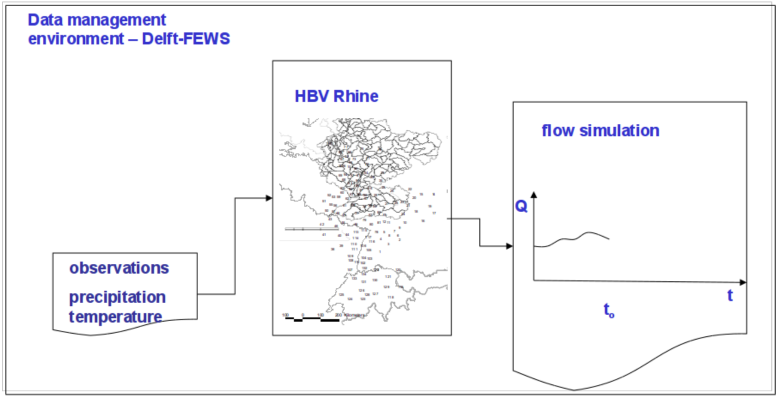
- use conceptual hydrological HBV-96 model
- simple routing, hydrodynamic also possible
- calibrated for 134 sub basins



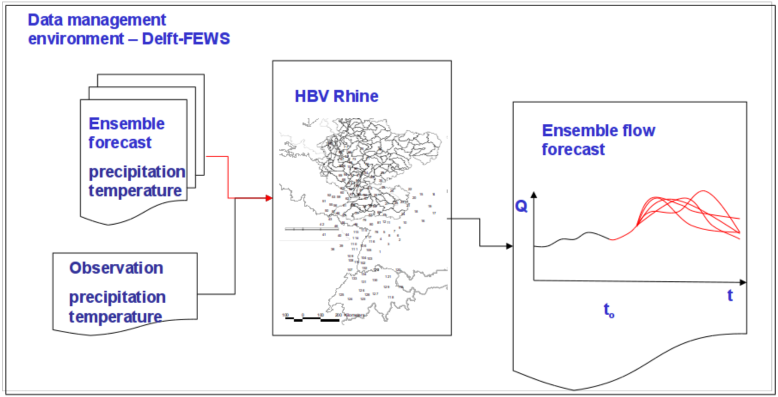
taken from http://www.smhi.se/foretag/m/hbv_demo/html/end.html



Hydrological Ensemble generation



Hydrological Ensemble generation



Hindcast set up

- baseline simulation (HBV) with observed meteorological inputs (1-6 hourly rain and temperature)
- daily hindcasts, HBV forced with meteorological forecasts
- **ECMWF-EPS** a global circulation model with a resolution of approx. 50 km and 51 ensemble members
- verification period 6/2004 - 10/2007
- **COSMO-LEPS** dynamic downscaling approach nested in ECMWF-EPS over Europe; 10 km resolution, 16 ensemble members,
- verification period 1/2007 - 10/2007
- statistical error correction with autoregressive (AR) based on 2-10 days of observed data, then predicting the error (Broersen and Weerts 2005)

Skill of ECMWF-EPS precipitation forecasts

- Precipitation forecasts on sub basin scale show to have Ranked Probability Skill Score ($\boldsymbol{RPSS}$) between 0.1 and 0.3 compared against climatology, the skill deteriorates with lead time and there is no skill after 5 to 7 days

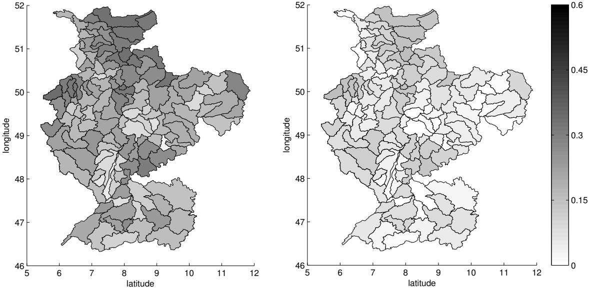


Figure: RPSS of precipitation forecasts against "observed" subbasin precip. Left one day, right panel 5 days lead time. A value of one indicates a perfect forecast, while 0 indicates no skill.

Skill of ECMWF-EPS driven discharge ensemble

- Skill of discharge forecasts at main river basin scale (from 4124 to 160,800 km^2) against climatology
- positive skill up to 9 days $\rightarrow$ long travel times
- tendency of increasing skill with catchment area

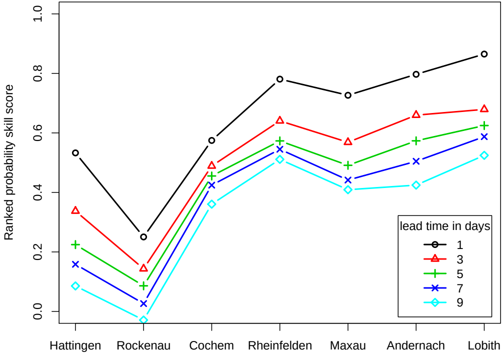
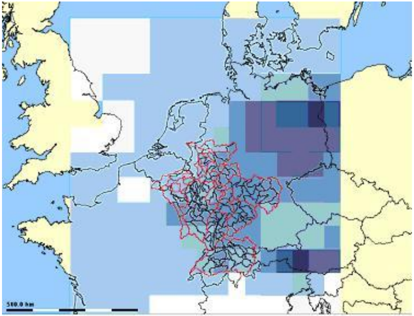
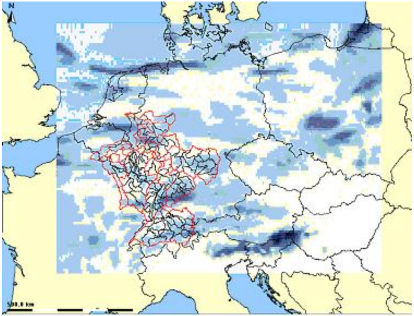


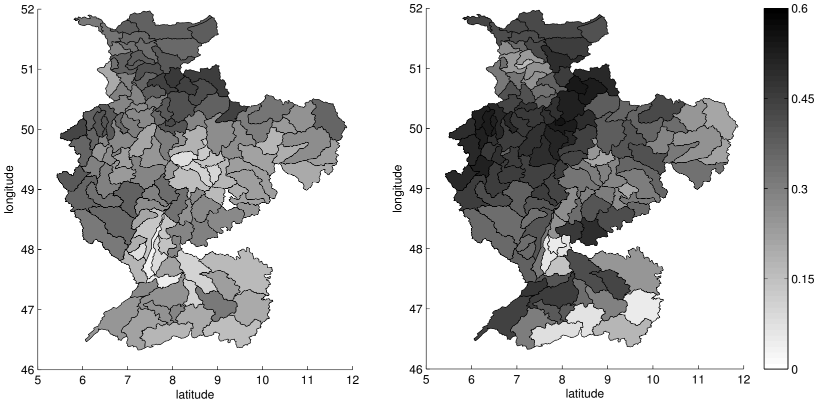
Figure: RPSS of error corrected flow forecasts at main river gauges, sorted by catchment size.

Precipitation ECMWF-EPS vs. COSMO-LEPS

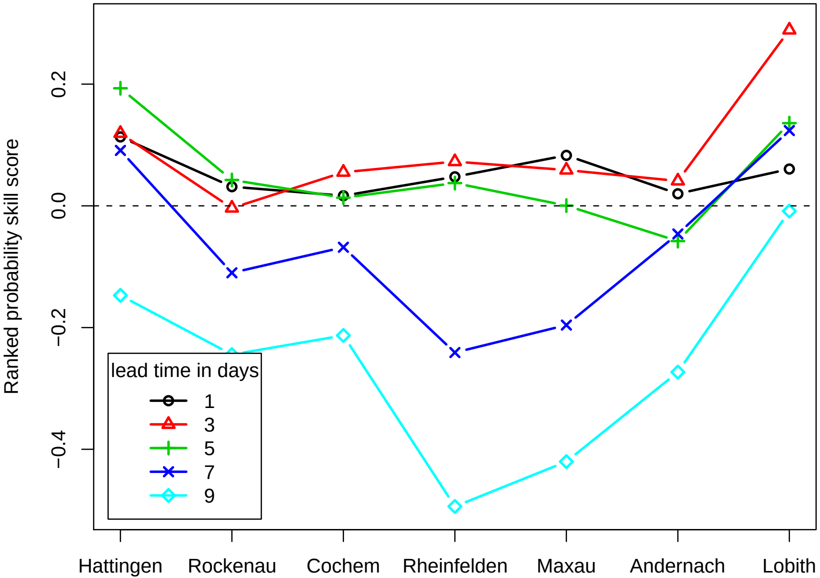




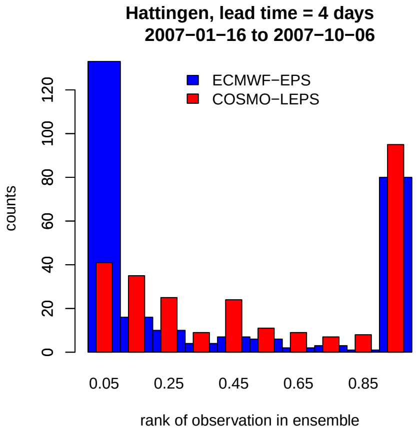
ECMWF-EPS and COSMO-LEPS precip skill at one day ahead

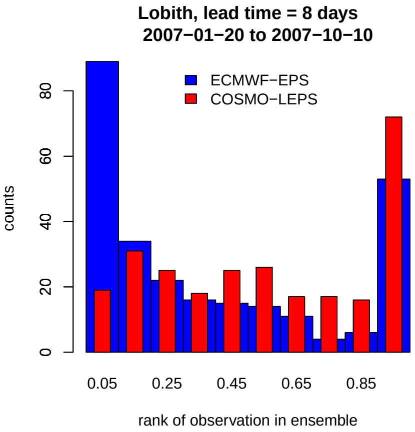


Discharge skill of COSMO-LEPS compared with ECMWF-EPS



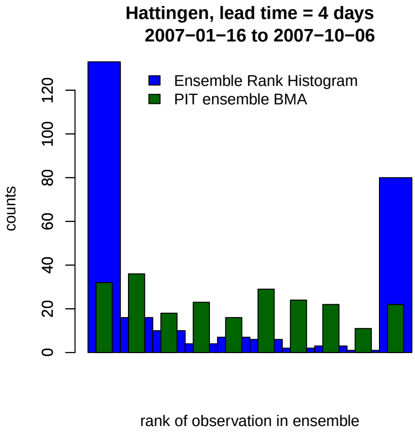
Reliability of flow forecasts forced with ECMWF-EPS vs. COSMO-LEPS

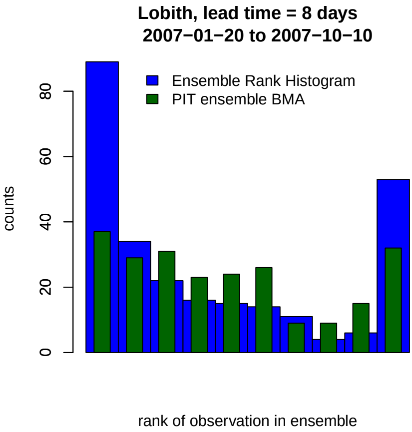




Improving forecasts with its own error: Ensemble postprocessing with Bayesian Model Averaging

30 days training period, gaussian error model (Fraley et al. 2010)



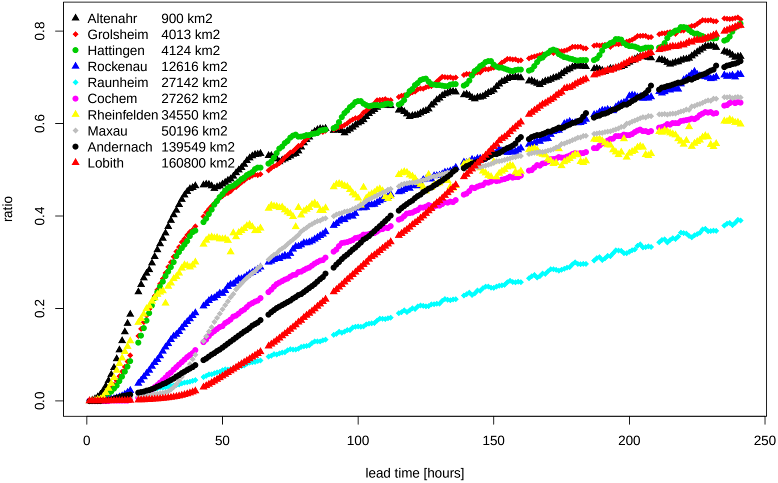


What we learnt from verification

1. medium range ensemble flow forecasts provide some skill
2. downscaling of precipitation forecasts improves skill in flow forecasts at all basin sizes
3. calibration of ensemble forecasts increase reliability (and often the spread)
4. for interesting warning threshold verification long hindcasts are necessary

Dominant sources of uncertainty

use separation of error (MAE) by the ratio of: $\frac{forecast\ vs.\ simulation}{forecast\ vs.\ observation}$



Recommendations to improve flow forecasts

Future meteorological conditions are dominant:

1. force hydrological models with meteo. ensembles
2. downscaling further improves hydrologic forecasts

Short term forecasts

data rich:

1. Error correction
   - statistical
   - model state updating with Ensemble Kalman Filter (Weerts and El Serafy 2006, Vrugt and Robinson 2007)
2. Postprocessing, e.g. BMA, Model Output Statistics (MOS), ...
3. Hydrological Ensembles (Multimodel (Georgakakos et al. 2004, Velázquez et al. 2011), Sampling Input Uncertainty

Short term forecasts

data poor:

1. Hydrological Ensembles
2. Model state updating using other data, e.g. Remote sensed soil moisture (Komma et al. 2008)
3. any PUB ideas?

some data:

- Geostatistical interpolation of forecast accuracy or postprocessing
- hydrodynamic modelling with updated states (Weerts et al. 2010)

Thank you for listening!

Questions?

References

Brier, G.: 1950, Verification of forecasts expressed in terms of probability, *Monthly weather review* **78**(1), 1–3.

Broersen, P. and Weerts, A.: 2005, Automatic Error Correction of Rainfall-Runoff models in Flood Forecasting Systems, *Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE*, Vol. 2, pp. 963– 968.

Fraley, C., Raftery, A. E., Sloughter, J. M., Gneiting, T., of Washington With contributions from Bobby Yuen, U. and Polokowski, M.: 2010, *ensembleBMA: Probabilistic Forecasting using Ensembles and Bayesian Model Averaging*. R package version 4.5.
URL: *http://CRAN.R-project.org/package=ensembleBMA*

Georgakakos, K., Seo, D., Gupta, H., Schaake, J. and Butts, M.: 2004, Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *Journal of hydrology* **298**(1-4), 222–241.

Komma, J., Bloschl, G. and Reszler, C.: 2008, Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, *Journal of Hydrology* **357**(3-4), 228–242.

Velázquez, J., Anctil, F., Ramos, M. and Perrin, C.: 2011, Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, *Advances in Geosciences* **29**, 33–42.

Vrugt, J. and Robinson, B.: 2007, Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res* **43**, W01411.

Weerts, A. and El Serafy, G.: 2006, Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water resources research* **42**(9), W09403.

Weerts, A., El Serafy, G., Hummel, S., Dhondia, J. and Gerritsen, H.: 2010, Application of generic data assimilation tools (DATools) for flood forecasting purposes, *Computers & Geosciences* **36**(4), 453–463.

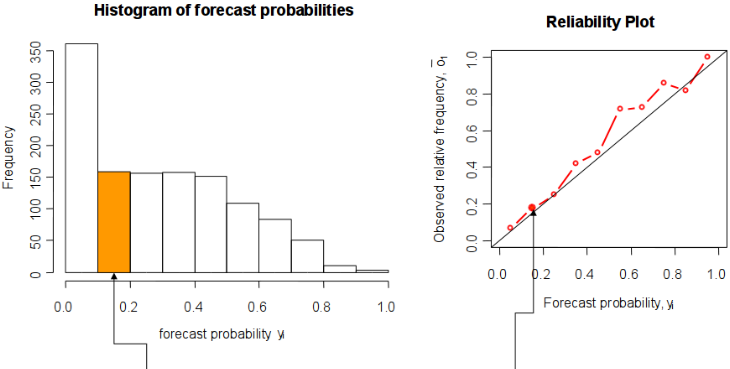
Weerts, A., Winsemius, H. and Verkade, J.: 2011, Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), *Hydrol. Earth Syst. Sci* **15**, 255–265.

Wilks, D. S.: 2006, *Statistical Methods in the Atmospheric Sciences*, 2nd edn, International Geophysics Series, Vol. 59, Academic Press.

WMO: 2010, Forecast verification - issues, methods and faq, WWRP/WGNE Joint Working Group on Verification.
URL: *http://www.cawcr.gov.au/projects/verification/*

Reliability at a given threshold

- specific for a threshold of the target variable and a period T
- given all cases, when a certain probability of exceedance of this threshold was issued by the forecast, how often this exceedance has been observed



$$\bar{o}_{1,k=2} = p\left(o_1 \middle| 0.1 < y_i \le 0.2\right) = \frac{1}{n}\sum o_{i,k=2}$$

Reliability of ECMWF-EPS driven discharge forecasts

- evaluate threshold exceeding $\mathbf{1460}\ m^3/s$ at lead time of $\mathbf{8}$ days at Maxau/Upper Rhine
- 80% quantile –> observed at 200 out of 1000 days (sample climatologic probability)

→ very long hindcast data set would be rquired to verifiy forecasts for warning thresholds!

- forecast probabilities histogram: "sharp" but mostly 0 → low sample size for probabilites in between
- reliability diagram: close to 1-1 line → is relatively reliable
- error corrected forecast vs. observed full uncertainties
- forecast vs. baseline simulation meteo. forecast uncertainty only
- deviations: low probability of exceedance (important for high risk decisions) are underpredicted due to meteo. forecasts!
- high exceedance probability predicted

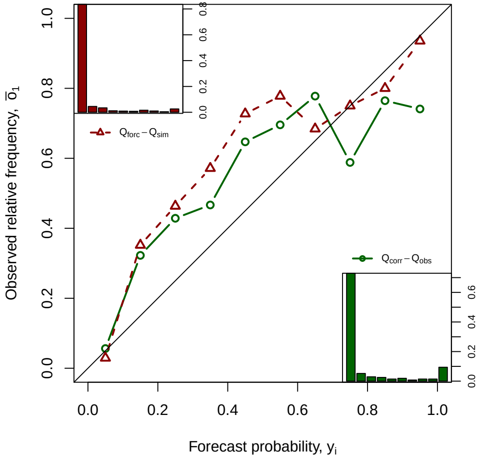


Figure: Reliability diagram for Maxau at a lead time of $\mathbf{8}$ days exceeding 1460 m^3/s. The dark green line marked with circles shows the reliability line for the verification of the error-corrected flow forecast against observations ($Q_{corr} - Q_{obs}$). The dark red dashed

1