Calibration at ZAMG

Alexander Kann, Benedikt Bica, Yong Wang

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Outline

- Work on calibration of ALADIN-LAEF (~5 -6 yrs ago)
- Comparison calibration ECMWF-EPS vs. LAEF
- Deterministic calibration: A multi-model post-processing method vs. single model calibration approach
- Open issues to be discussed …

Bias correction method (1st moment correction)

Bias Assessment: adaptive (Kalman filter type) algorithm

Implementation of decaying averaging for the first moment bias (from Bo Cui, NCEP, 2006), flow-dependant

decaying averaging mean error = (1-w) * prior t.m.e + w * (f – a)

a) Prior estimate to startup procedure: choose T0 as current date (00Z), calculate the time mean errors between T-46 and T-16 day.

b) Update: the prior estimate of the average state is multiplied by a factor 1-w (<1). Then, most recent verification error (f - a) is added to the decaying average for each lead time with a weight of w (operational: w = 2%).

c) Cycling: repeat step (b) every day.

d) Carry out steps (a) to (c) for each variable, for each lead time and on every grid point.

e) The bias correction is applied on each ensemble member.

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration method (2nd moment correction)

- Gneiting et. al., 2005: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Wea. Rev*., **133**, 1098-1118.
- Hagedorn, R., Hamill, T., Whitaker, J., 2007: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: 2-meter Temperatures. *Mon. Wea. Rev*., **136**, 2608-2619.

- Idea of NGR (Non-homogenous Gaussian regression):

  Based on multiple linear regression, addresses forecast bias and underdispersion. NGR yields probabilistic forecasts with Gaussian PDF‘s for continuous weather variables. The predictive mean is a bias-corrected average of the ensemble member forecasts. The predictive variance is a linear function of the ensemble variance. Fitting the regression coefficients, the method of minimum CRPS estimation is used (optimizing CRPS for the training data).

$$PDF = N(a + bX, c + dS^2)$$

a, b: bias & general performance of ensemble mean

c = 0; d = 1: large spread-skill relationship

d = 0: small spread-skill relationship

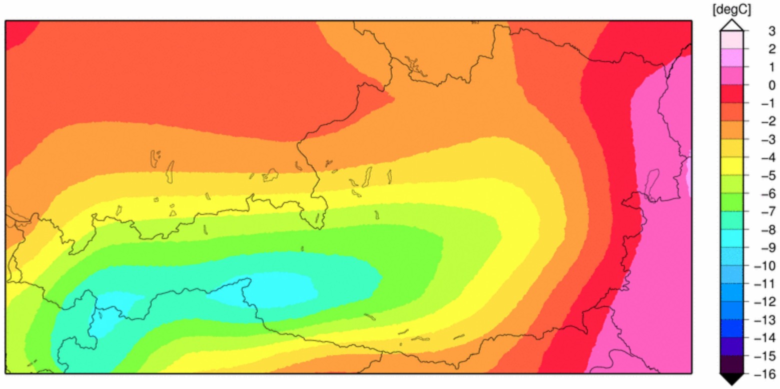
The calibration is done with INCA analyses on a 1 km*1 km horizontal grid covering Austria, a sliding 50 day training period is used. The experiment is carried out for one month (December 2007). Verification is done using station observation only.

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Forecast fields: Ensemble Mean

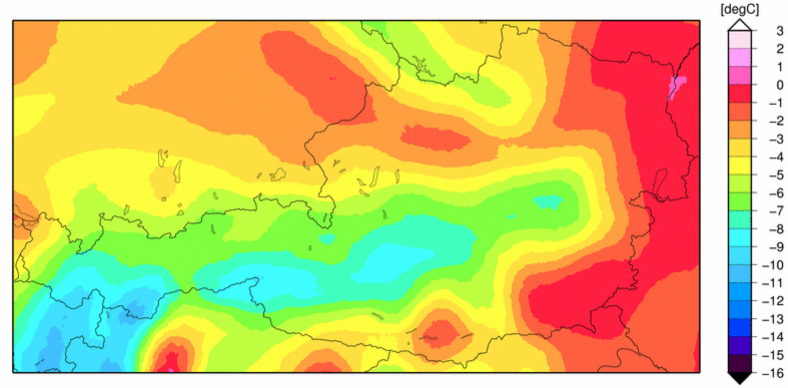
ECMWF: Uncalibrated 2m Temperature, Ensemble Mean

Forecast from: 20071216 00 UTC + 36h



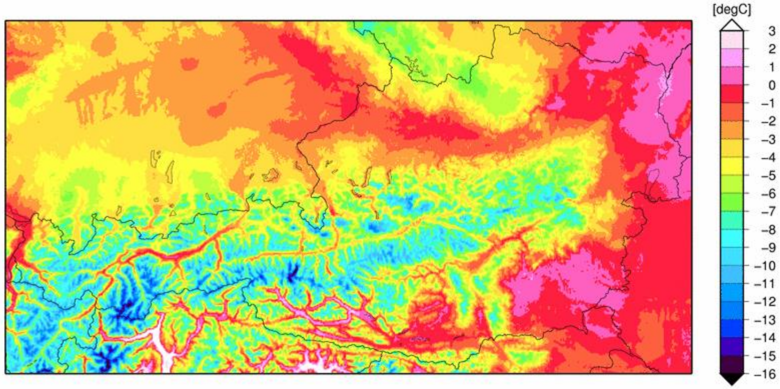
LAEF: Uncalibrated 2m Temperature, Ensemble Mean

Forecast from: 20071216, 00 UTC + 36h



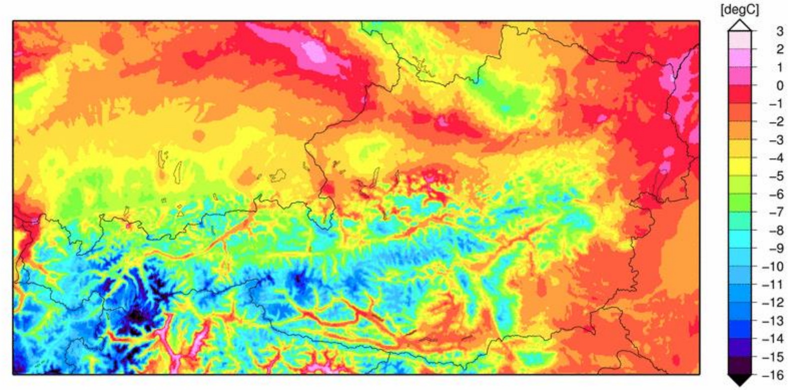
LAEF: Calibrated 2m Temperature, Ensemble Mean

Forecast from: 20071216, 00 UTC + 36h



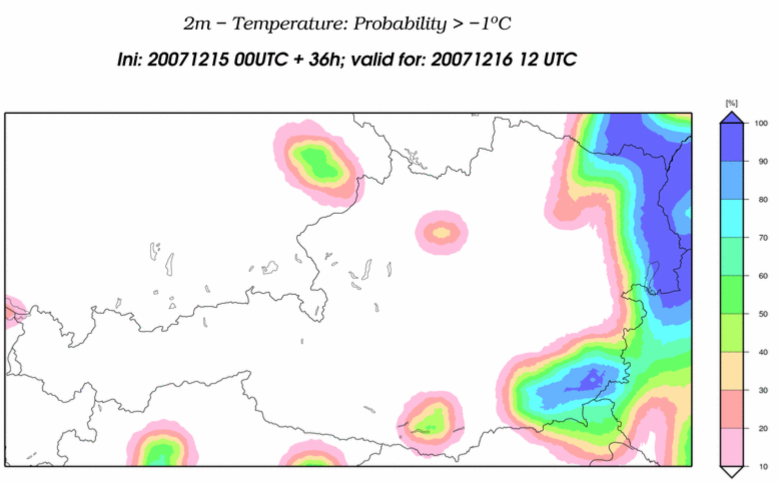
INCA: 2m Temperature

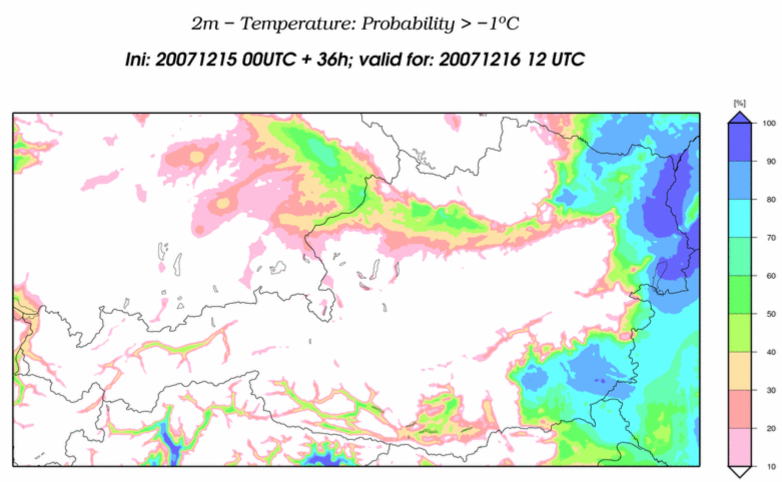
Analysis for: 20071217, 1200 UTC

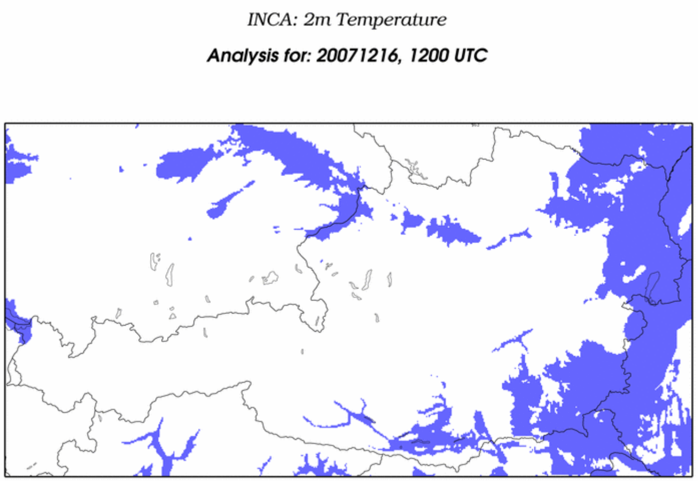


Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Forecast fields: Probability Charts



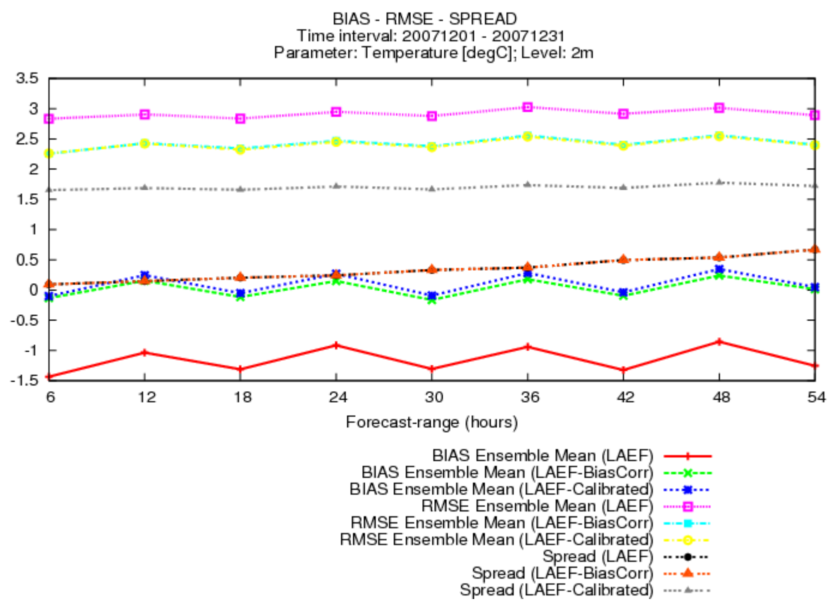


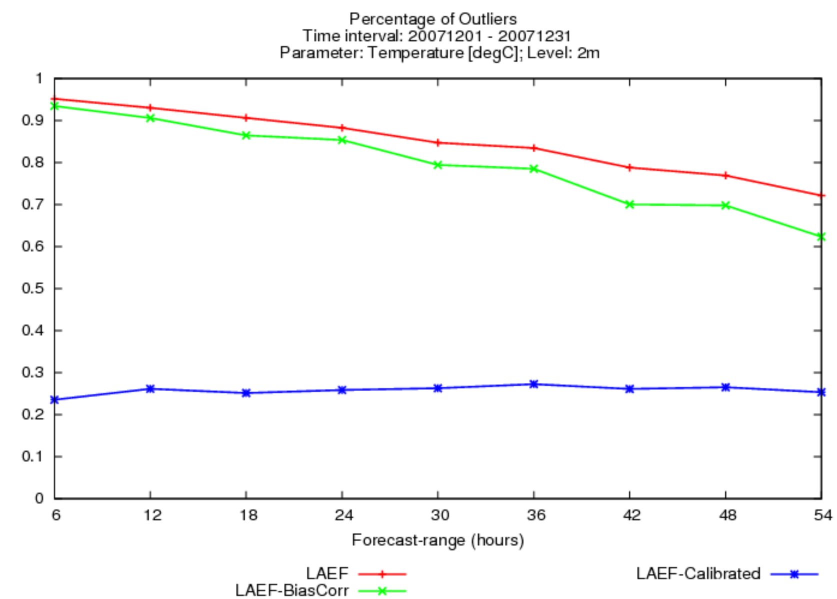


Probability plot for T2M>-1°C, forecast from 20071215, 00UTC + 36h. Raw LAEF (top left), calibrated LAEF (top right) and INCA analysis showing areas exceeding -1°C in blue (bottom left). Although LAEF roughly covers the areas, the calibration is able to add information particularly on local scale.

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Verification (Bias, RMSE, Spread; Outliers)



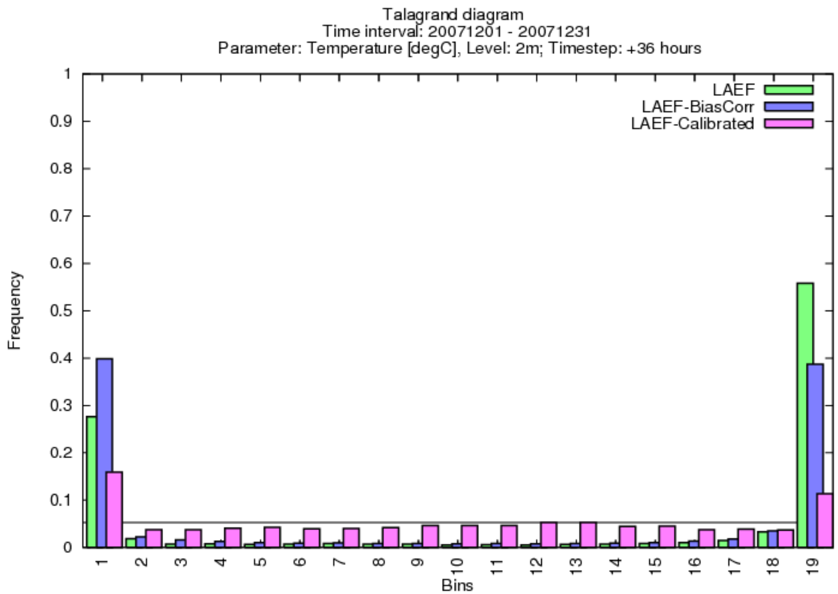


Bias, RMSE of Ensemble Mean and Ensemble Spread as a function of lead time for 2m temperature of raw LAEF, bias-corrected EPS and calibrated EPS. Bias correction leads to reduction of RMSE (from ~3K to ~2,4K), the spread is increased up to 1,5K by calibration.

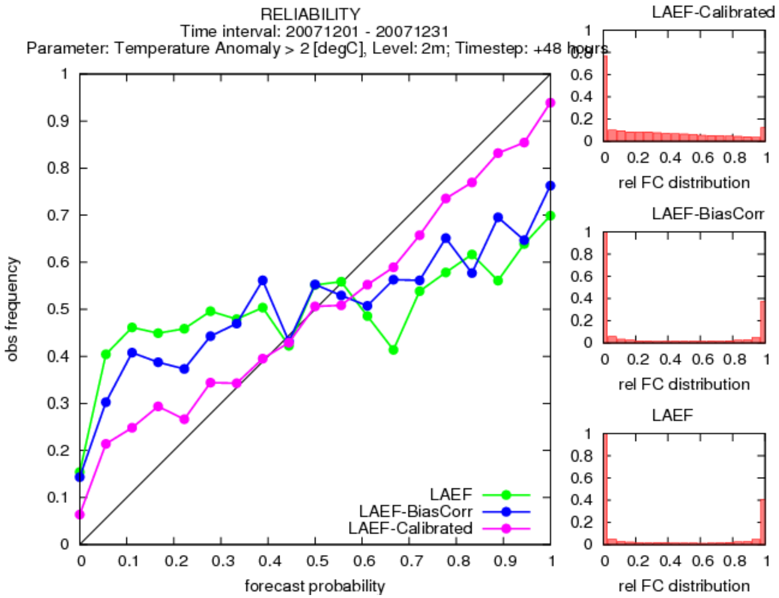
Percentage of outliers as a function of lead time for raw LAEF (red), bias-corrected LAEF (green) and calibrated LAEF (blue). The bias correction shifts the PDF and therefore slightly reduces the number of outliers, but with full calibration the percentage of outliers decreases to about 30% -35%.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Verification (Talagramm; Reliability)



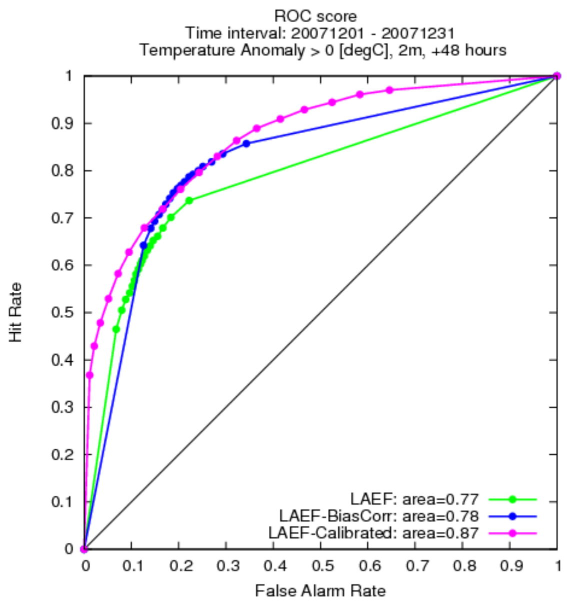
Talagrand diagram for 2m temperature, lead time +36hours. LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). The distribution becomes much flatter by calibrating, although it still remains slightly underdispersive.



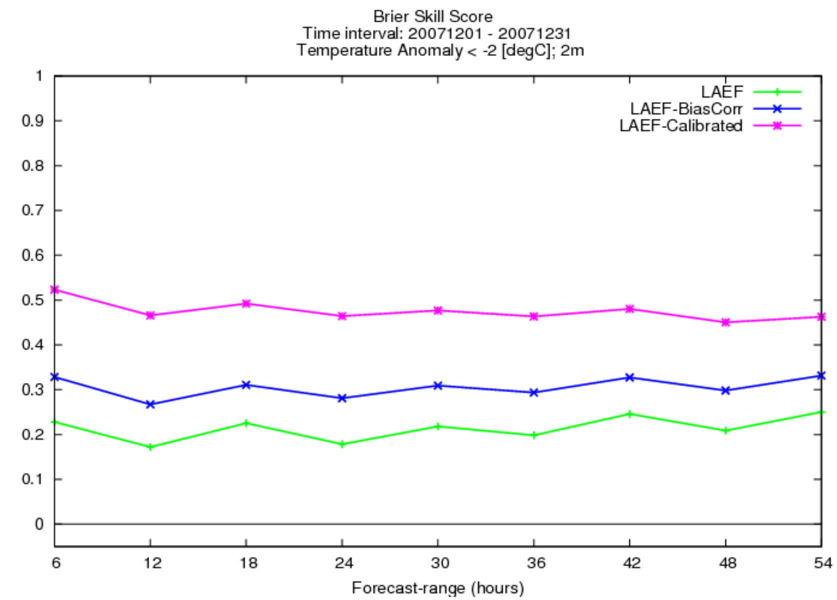
Reliability diagram for raw LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). The calibrated ensemble performs best, although they all tend to overforecast high probabilities (low probabilities are underforecast).

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Verification (ROC; Brier Skill Score)



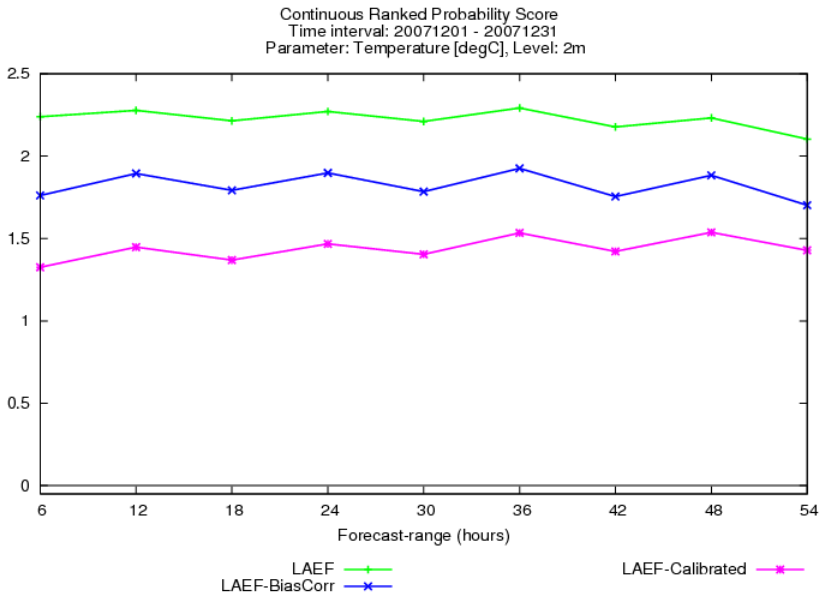
ROC curve and area under the ROC curve for 2m temperature anomaly > 0°C, lead time +48hours. LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). The area under the ROC curve for calibrated ensemble is about 10% higher than for raw or bias-corrected ensemble.

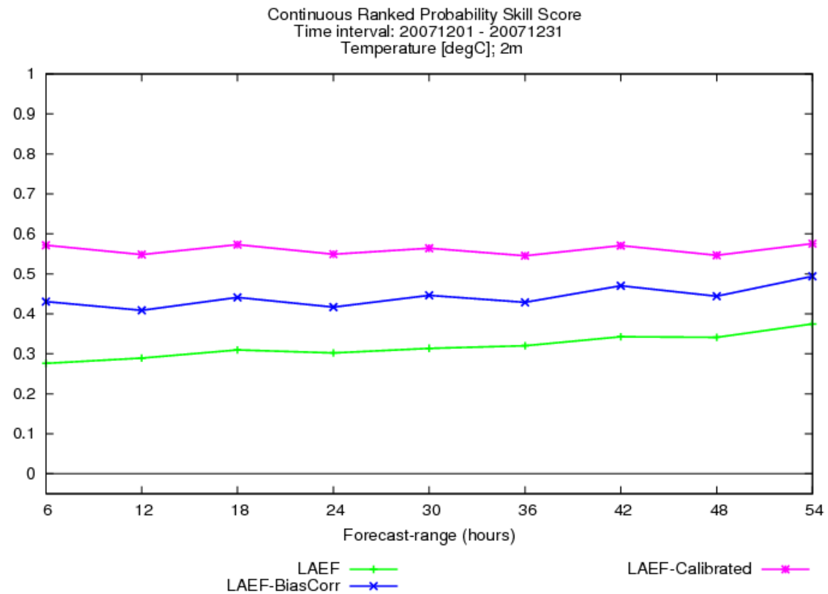


Brier Skill Score for 2m temperature anomaly < 0°C; raw LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). The calibrated ensemble performs much better, 30% - 40% of the improvement is achieved by the bias-correction.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Calibration results: Verification (CRPS; CRPSS)



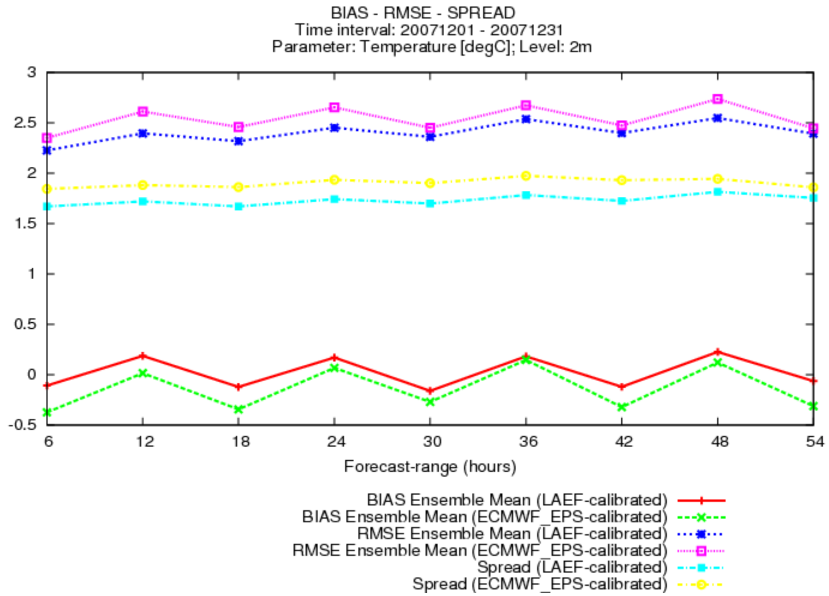


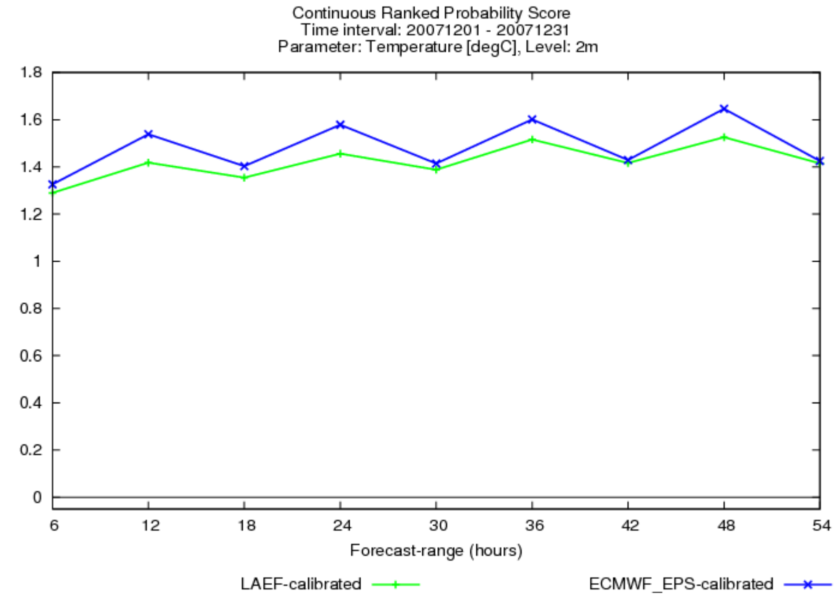
CRPS for T2M as a function of lead time. LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). About 50% of the calibration improvement is achieved by bias correction, the CRPS is decreased from ~2,3K to ~1,5K !!

CRPSS (with deterministic Aladin-Austria as a reference) for T2M as a function of lead time. LAEF (green), bias-corrected LAEF (blue) and calibrated LAEF (purple). Again, about 50% of the total calibration improvement is obtained by bias correction, the CRPSS is approx. doubled from ~0,3 to almost 0,6 !

ZAMG Zentralanstalt für Meteorologie und Geodynamik

ECMWF calibration vs. LAEF calibration



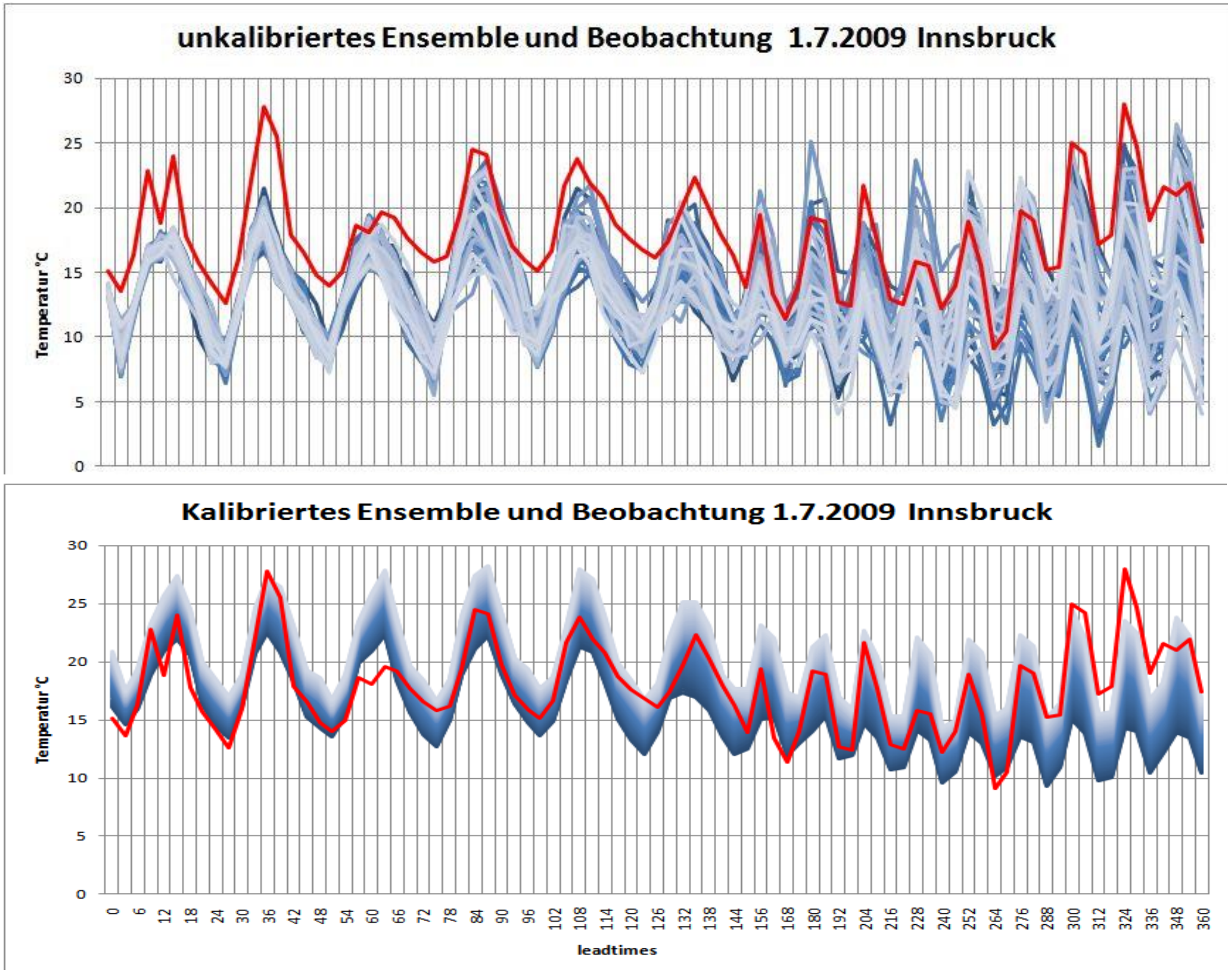


Bias, RMSE of Ensemble Mean and Ensemble Spread as a function of lead time for 2m temperature of calibrated LAEF and calibrated ECMWF-EPS. For ECMWF, a slight bias remains, both spread and RMSE are higher.

CRPS for T2M as a function of lead time. Calibrated LAEF (green), calibrated ECMWF-EPS (blue). LAEF performs slightly better, although differences are not very overwhelming.

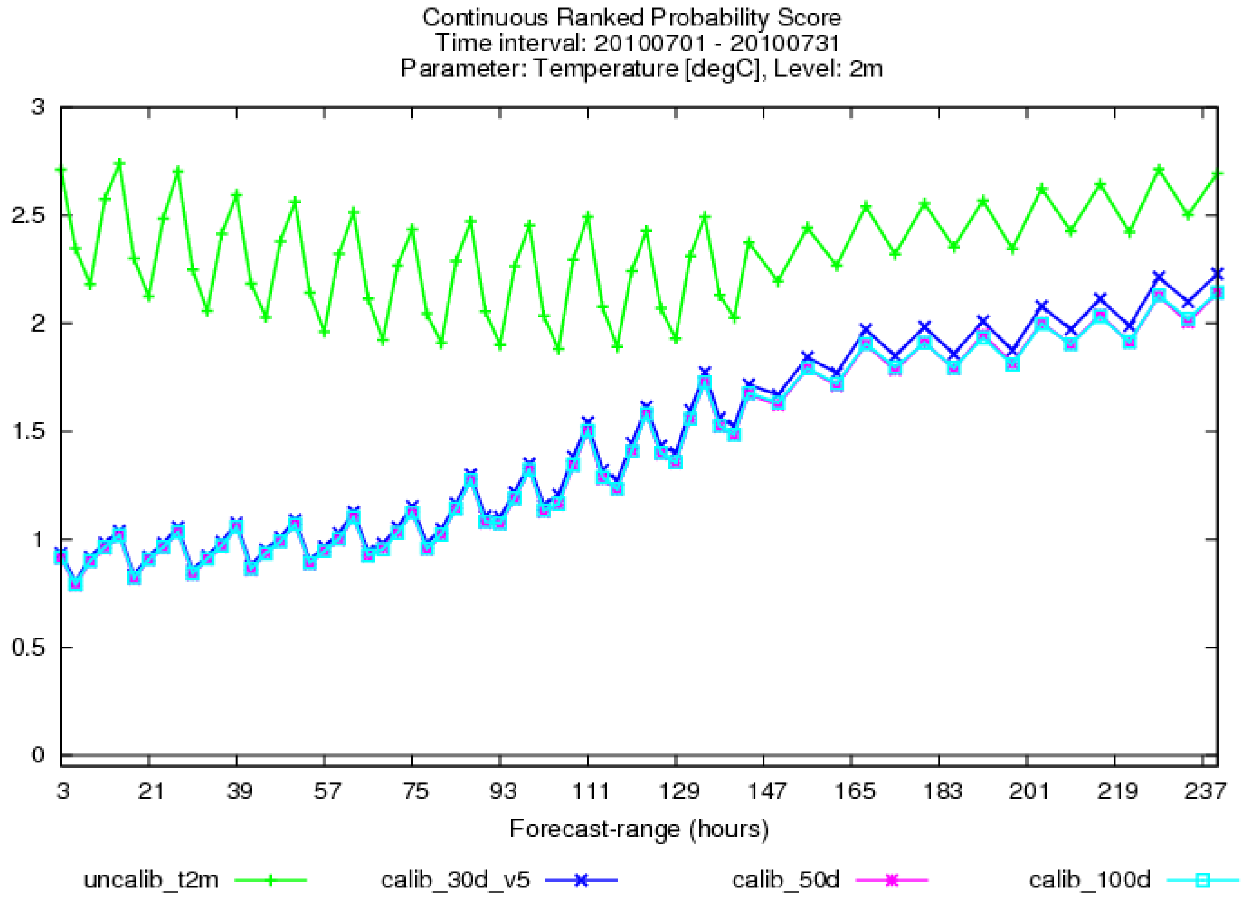
ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration of ECMWF 2-m temperature with NGR



ZAMG
Zentralanstalt für Meteorologie und Geodynamik

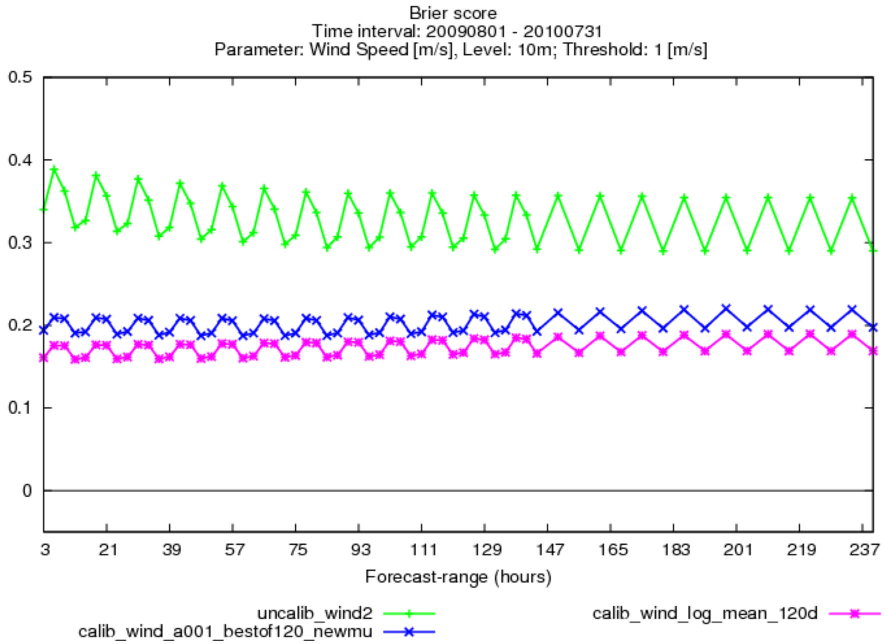
Validation raw vs. calibrated: 10 day forecasts



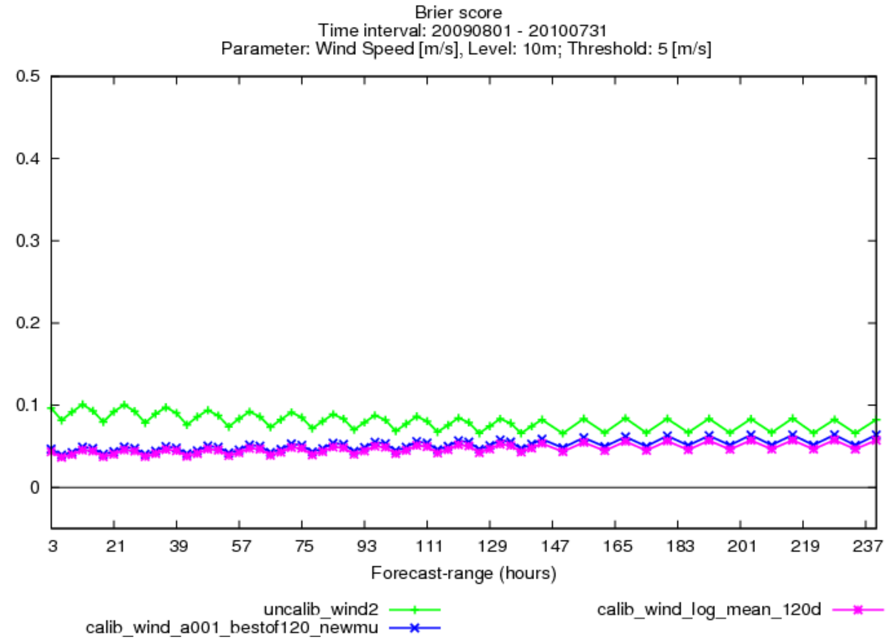
ZAMG Zentralanstalt für Meteorologie und Geodynamik

Calibration ECMWF 10-m Wind

Brier Score 1m/s



Brier Score 5m/s



- Raw: green; cut-off NGR: blue; logistic Regression: purple

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Operational ‚META‘ Forecasting System

- Based on : Woodcock, F. and C. Engel, 2005: Operational Consensus Forecasts. *Weath. Forcasting*, **20**, 101-111.
- META-Forecast for 00/06/12/18 UTC
  - Hourly until Day 3, 3-hourly until Day 5, afterwards 6-hourly until Day 10
- Parameter: T-2m, Td-2m, Wind, ff, Tmin, Tmax, Precip, Global Radiation, Cloudiness
- Site-specific forecasts interpolated from NWPs
- Bias correction based on previous 30/10 days (multi-stages)
- Weighting: MAE of latest 10 days and current error
- Weighting for model i depends on station k, on lead time t und on the parameter:

$$w_i^k(t) = a_i^k(t)^{-1} \left(\sum_i^n a_i^k(t)^{-1} \right)^{-1}$$

$$\text{Modell } i, \text{ Station } k, \text{ Zeit } t, \text{ Fehler (MAE) } a_i^k$$

- Number of available NWP models depends on the lead time

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

‚META‘ Forecast

1) NWP Forecasts (height-corrected)

2) Bias-Correction

3) Evaluation ➔ Weights

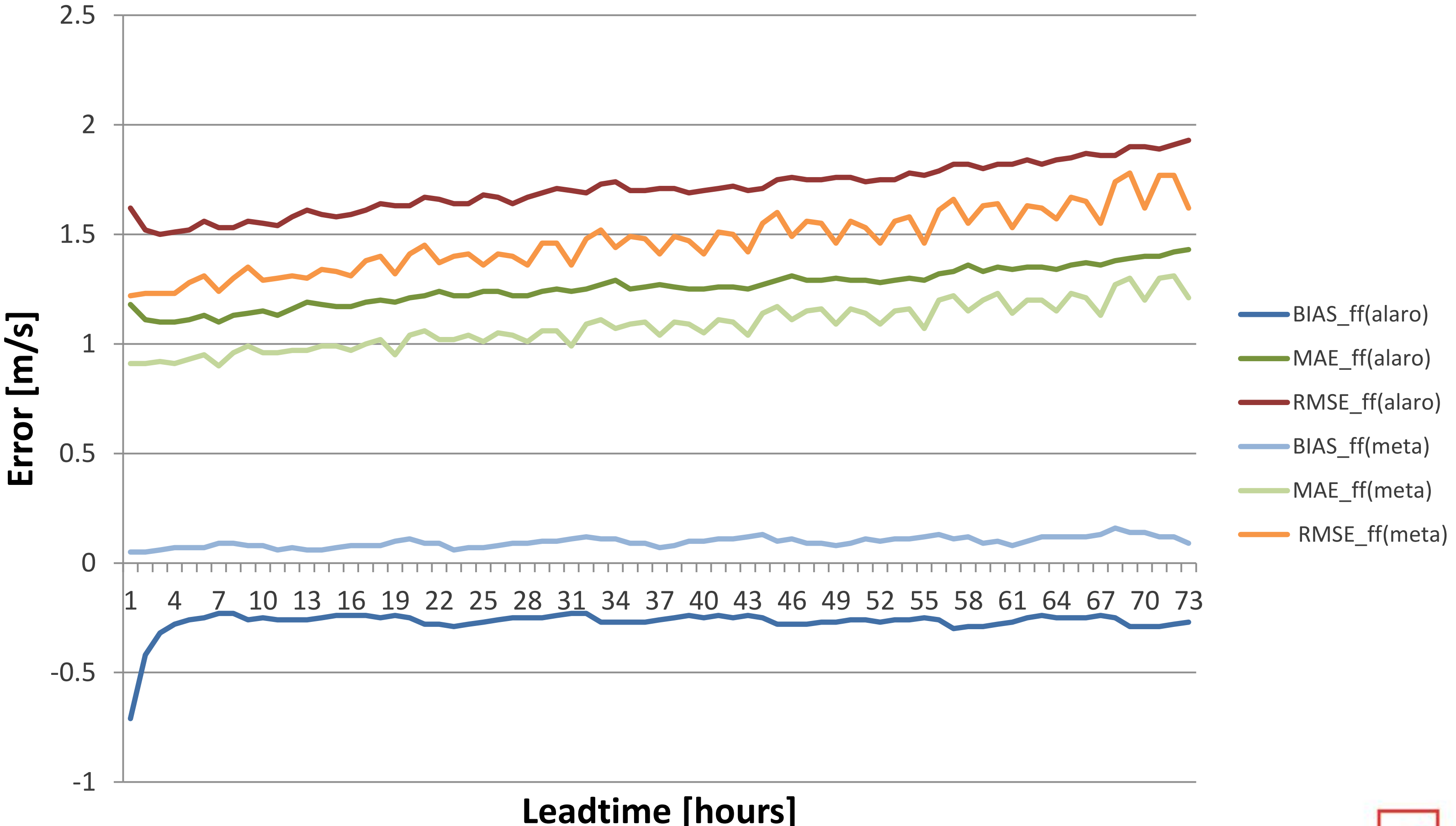
4) Weighted Mean of NWP forecasts

META - Forecast

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

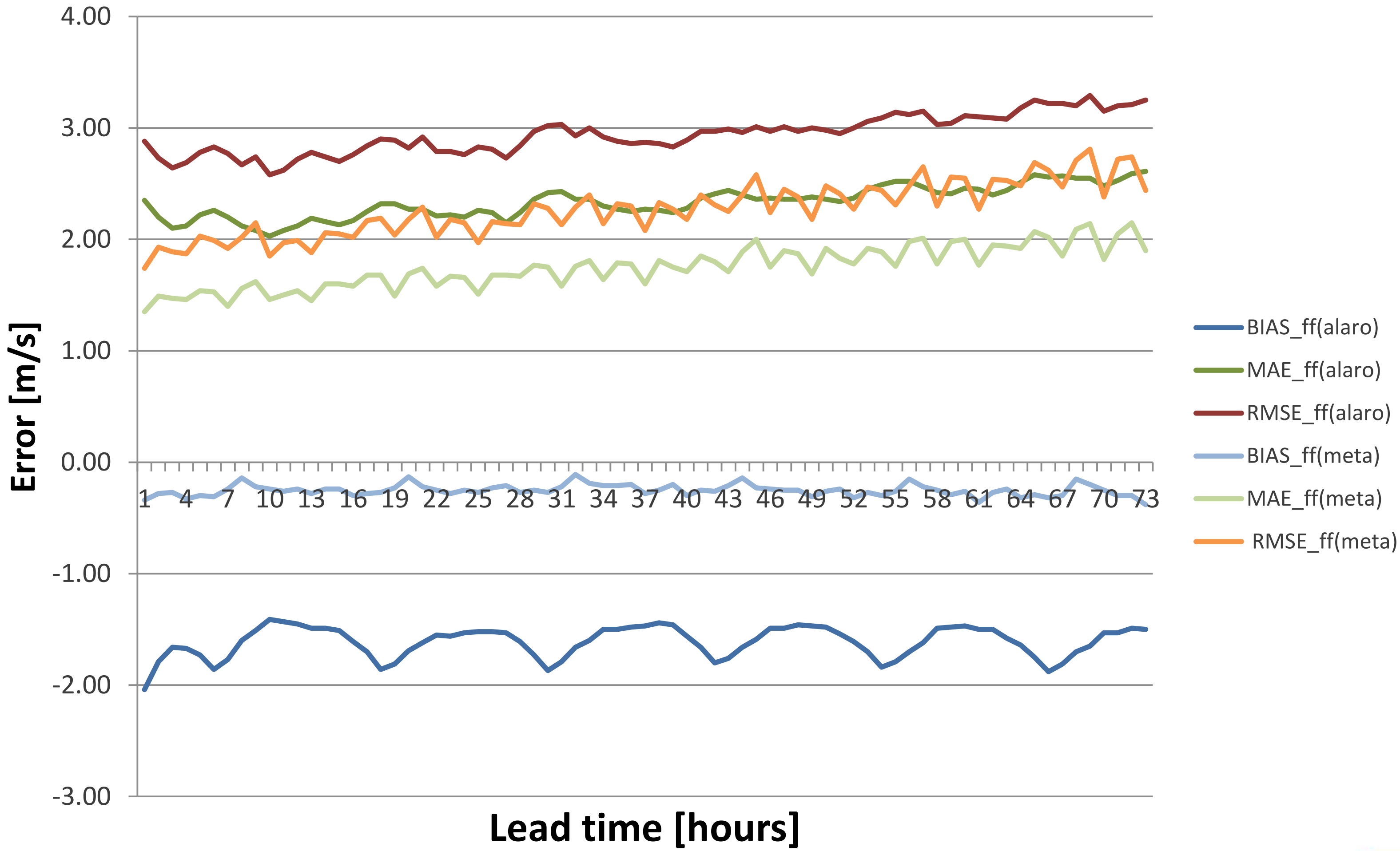
10-m wind speed: META vs. post-processed ALARO (lin Reg)

- Average over some flat stations (Zwerndorf, Allentsteig, Wolfsegg, St. Pölten, Wien-Unterlaa, Leiser Berge, Neusiedl, Andau)



10-m wind speed: META vs. post-processed ALARO (lin Reg)

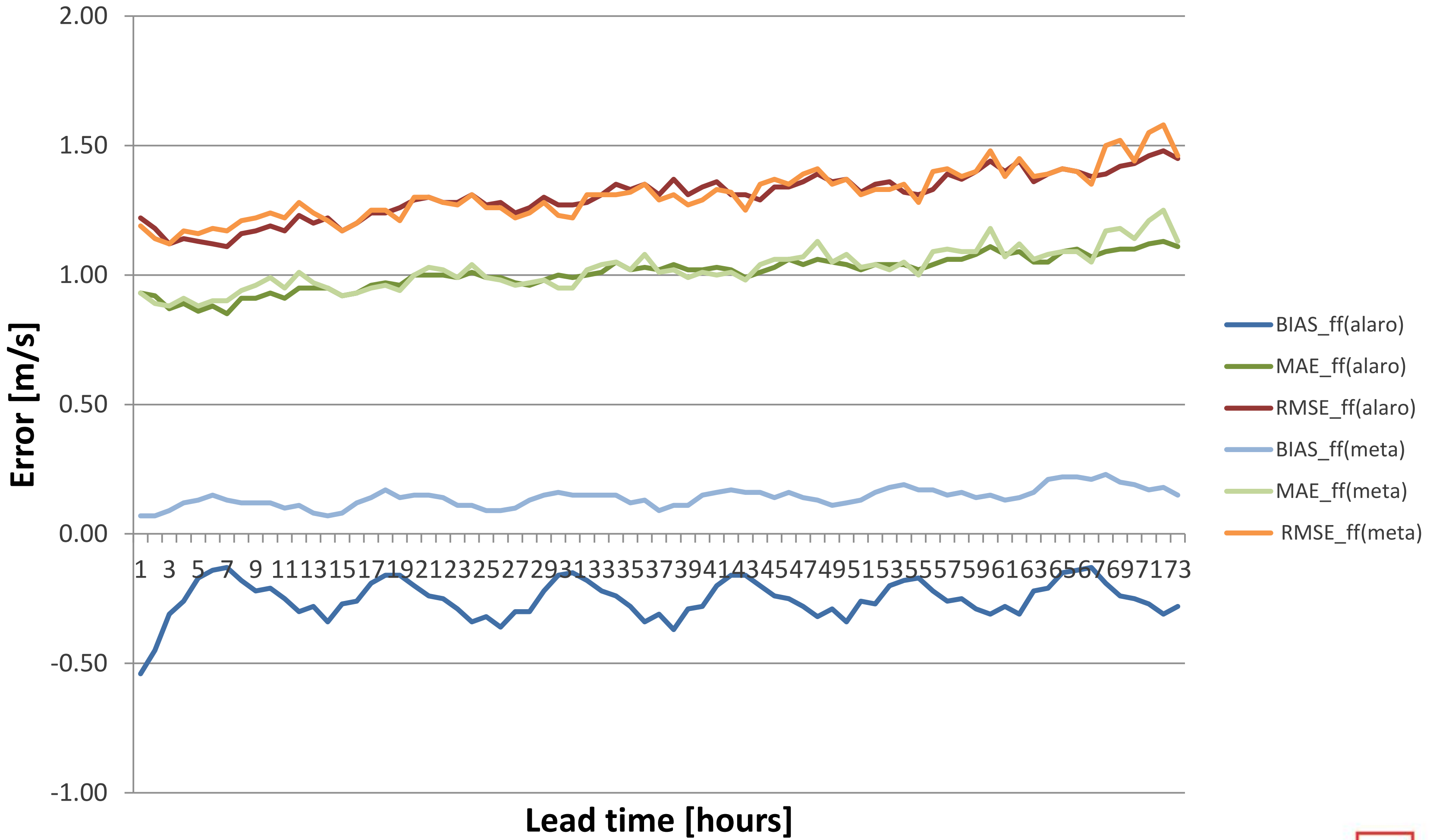
- Leiser Berge (hilly region)



ZAMG Zentralanstalt für Meteorologie und Geodynamik

10-m wind speed: META vs. post-processed ALARO (lin Reg)

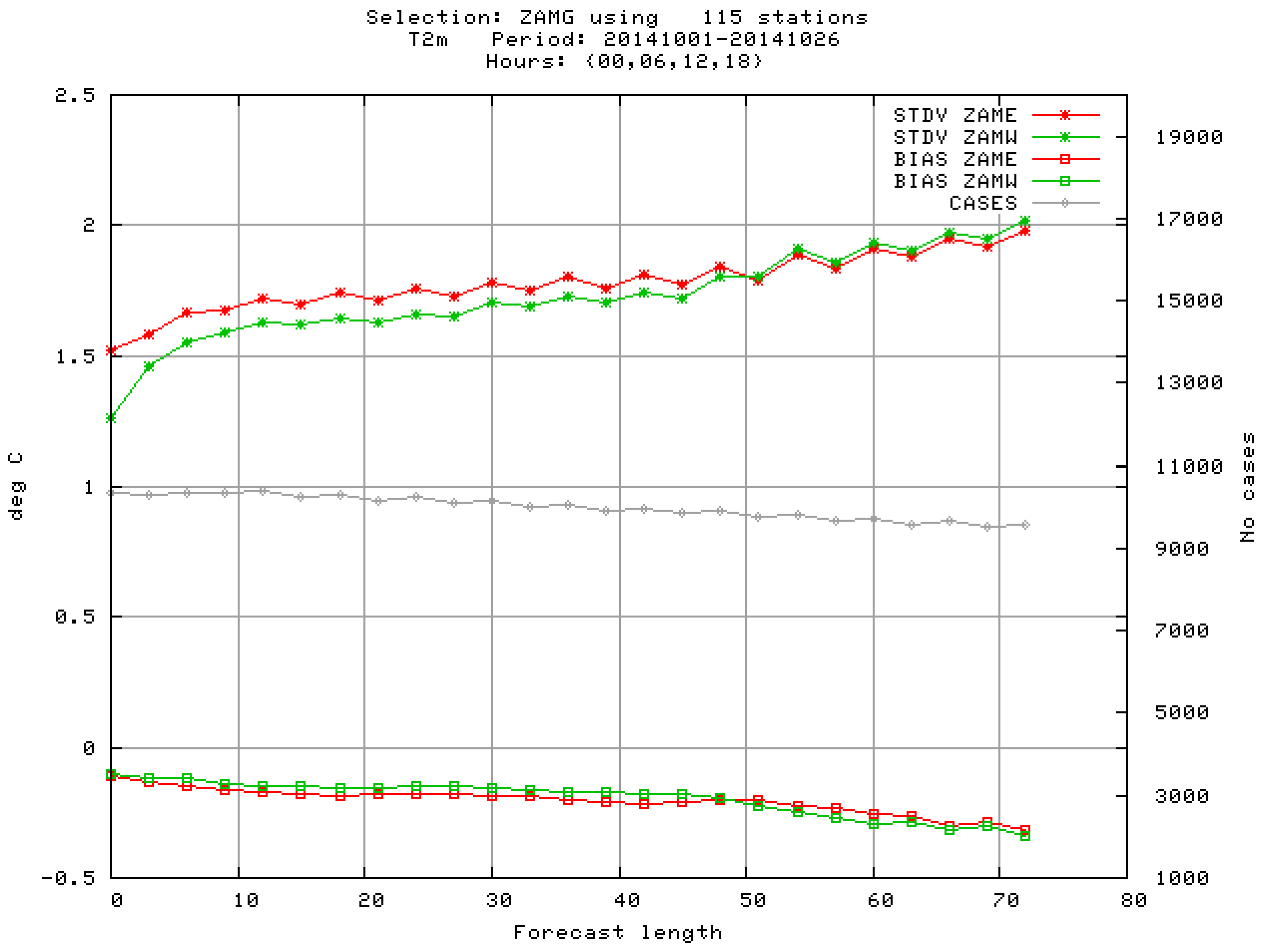
- Neusiedl (flatland)



ZAMG Zentralanstalt für Meteorologie und Geodynamik

New: regime-dependant META forecast: 2m-temperature

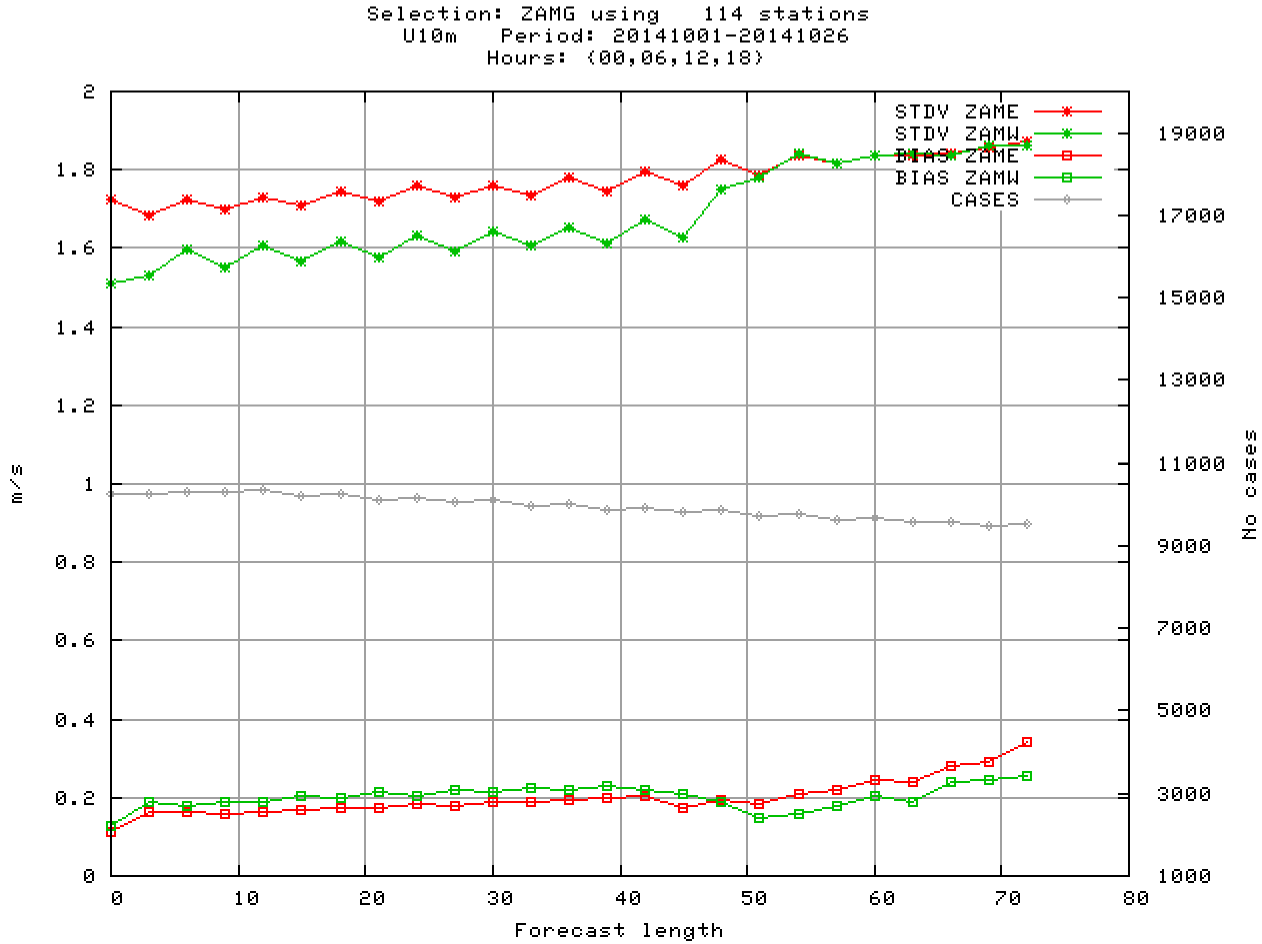
- Using automatic weather type classification (based on DWD‘s classification developed within COST 733)



ZAMG Zentralanstalt für Meteorologie und Geodynamik

New: regime-dependant META forecast: 10-m wind speed

- Using automatic weather type classification (based on DWD's classification developed within COST 733)



Issues to be discussed …

- Site-specific versus spatial calibration: Considering spatial-correlations to avoid artificial noise.

- Using analyses for gridded calibration is – in principle – preferable, but danger of introducing observational problems into calibrated forecast…

- Multi-variate calibration: applying methods to calibrate more parameter simultaneously in order to keep the temporal characteristics of the raw EPS.

- Do we need to work more on methods?

- Cooperation on methods?