Verification of ECMWF products at the Deutscher Wetterdienst (DWD)

DWD – Martin Göber

1. Summary of major highlights

The usage of a combined GME-MOS and ECMWF-MOS continues to lead to a further increase in forecast accuracy. ECMWF high resolution forecasts in conjunction with GME forecasts are now being used for the production of a probabilistic warning guidance based on the MOS technology.

2. Use and application of products

2.1 Post-processing of model output

2.1.1 Statistical adaptation

The high resolution ECMWF model and DWD's model GME are statistically interpreted up to 7 days in terms of near surface weather elements by means of a perfect prog scheme (AFREG) as well as by MOS and subsequent averaging of the two interpretations to form „AFREG/MIX" and "MOS/MIX".

Since 2004 a MOS interpretation of the ECMW model (ECMOS) has been used operationally in addition to the traditional MOS of DWD's global model GME (GMOS). A weighted average of the two MOS' forms MOS/MIX - the best available guidance for the production of local short and medium range forecasts. The introduction of MOS/MIX continues to lead to a further increase in forecast accuracy.

Since 2008 ECMWF high resolution forecasts in conjunction with GME forecasts have been used for the production of a probabilistic warning guidance based on the MOS technology.

Some EPS surface variables have been refined by Kalman filtering.

2.1.2 Physical adaptation

2.1.3 Derived fields

2.2 Use of products

The high resolution ECMWF model forms together with DWD's model GME the general operational data base. ECMWF's high resolution model is always used together with other models in short- and medium-range forecasting. For medium range forecasting the EPS is used additionally; in the short range the LEPS (Local model nested into EPS clusters) provides ensemble information. EPS products are used intensively in order to create a daily simple confidence number and describe alternative solutions. Furthermore, they are used to estimate the prospect for extreme weather events. Here, extensive use of the Extreme Forecast Index (EFI) is made.

3. Verification of products

3.1 Objective verification

3.1.1 Direct ECMWF model output (both deterministic and EPS)

3.1.2 ECMWF model output compared to other NWP models

Upper air forecasts from ECMWF continued to exhibit smaller errors than DWD-GME forecasts (Fig. 1). The RMSE of the ECMWF model for 500hPa geopotential height has decreased by 10 % (1 gpm) in the short range from 2006 to 2008 and was stagnant for the GME. ECMWF MSLP error growth with forecast range is about half a day better than for DWD-GME in the short range (fig. 2). The RMSE of the GME model for MSLP has hardly improved from 2006 to 2008 and only slightly for ECMWF (about 0,1 hPa in the short range and by 0,3 hPa in the medium range).

 GERMANY

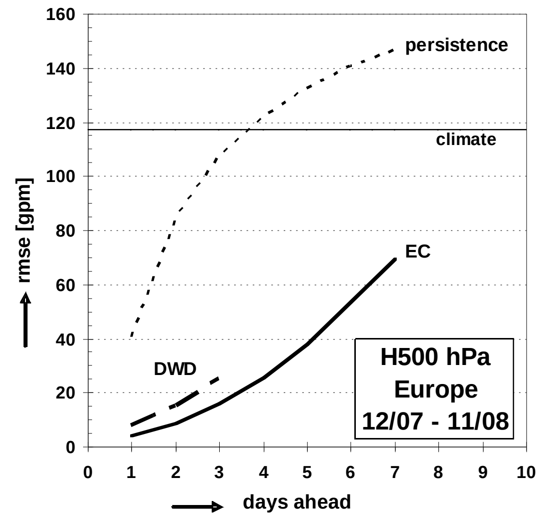

Fig. 1 RMSE 500hPa geopotential over Europe. DWD (Numerical Weather Prediction model GME), EC (high resolution ECMWF model), persistence (analysis from the initial state is used as a forecast for all following days), climate (long term mean of the predictand (H500, MSLP) serves as a constant forecast).

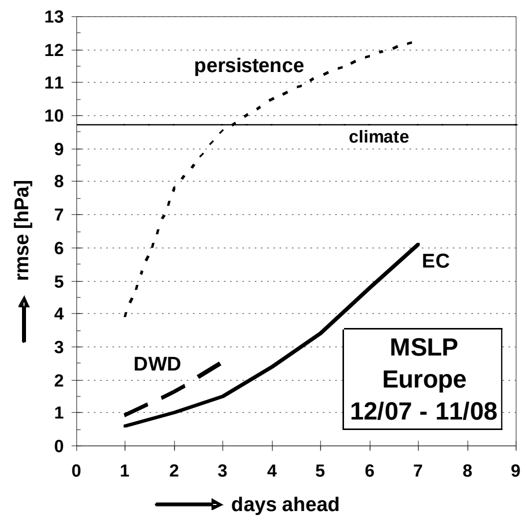

Fig. 2 Same as fig. 1, but for RMSE of mean sea level pressure.

3.1.3 Post-processed products

Here, various statistically post-processed model forecasts are compared for the following:

Predictands

MIN = daily minumum temperature (°C)
MAX = daily maximum temperature (°C)
SD = daily relative sunshine duration (%)
dd = surface wind direction (°) 12 UTC. Only verified, if ff(obs) ≥ 3 m/s
ff = surface wind speed (m/s) 12 UTC
PoP = Probability of Precipitation > 0 mm/d
PET = potential evapotranspiration (mm/d)
RR = a binary predictand: precipitation amount > 0 mm/d: Yes/No;

Forecast Types

AFREG/MIX = Perfect prog product AFREG(MIX) = AFREG(EC)+AFREG(DWD)/2
EC = high res. ECMWF model, DWD = operational DWD Global Model "GME" (initial time: 00 UTC). AFREG is generated for several *areas* of the whole Germany, but verified against *point* observations at 6 stations.

MOS/MIX = post processed product, a weighted average of Model Output Statistics of MOS/GME and MOS/EC

and **Verification measures**

rmse is used for both categorical and probabilistic forecasts (equals square root of the Brier Score)
RV = Reduction of Variance against reference, $1-(rmse/rmse^*)^2$, here: mean value for day 2 ... 7
rmse* = smoothed climate as the best reference forecast to evaluate forecast skill
HSS = Heidke Skill Score, only for binary predictands
HSS = mean value for day 2 ... 7

rmse		day +2	+3	+4	+5	+6	+7	+8	rmse* (climate)	RV [%]
MIN	AFREG/MIX	2,28	2,28	2,37	2,49	2,70	3,00	3,23	3,95	62
	MOS/MIX	1,53	1,73	2,03	2,34	2,64				72
MAX	AFREG/MIX	2,38	2,49	2,69	2,85	3,10	3,38	3,67	4,36	61
	MOS/MIX	1,76	2,04	2,41	2,69	2,97				69
SD	AFREG/MIX	24,8	25,5	26,1	26,7	27,7	28,7	29,2	29,6	19
dd[1)]	AFREG/MIX	39,2	42,3	46,0	51,9	57,0	65,4	70,6	83,7	68
	MOS/MIX	30,6	37,3	43,2	50,6	59,1				72
ff	AFREG/MIX	1,67	1,76	1,90	2,02	2,16	2,23	2,34	2,32	32
	MOS/MIX	1,55	1,70	1,89	2,00	2,14				35
PoP	AFREG/MIX	39,2	40,0	41,4	41,9	43,8	45,4	46,5	46,9	25
	MOS/MIX	35,9	38,2	41,1	42,0					30
PET	AFREG/MIX	0,709	0,729	0,759	0,778	0,810	0,852	0,864	0,906	27
HSS%										HSS
RR	AFREG/MIX	45	43	39	37	30	26	20	0	41
	MOS/MIX	58	50	40	35					46

Table 1 Verification of operational medium range forecasts for 6 stations in Germany (Hamburg, Potsdam, Düsseldorf, Leipzig, Frankfurt/M., München); 12/2007- 11/2008; rmse and HSS, respectively. Day of issue = day +0 = today at noon. [1)] Here, persistence is used as a 'reference forecast'.

The skill (RV) of minimum temperature and wind (speed and direction) forecasts has improved significantly since 2006. However, the skill of "water related" PPM's (RR, PoP, PET, SD) has decreased considerably, especially in the medium range. MOS/MIX forecasts have substantially smaller errors than AFREG/MIX, which is only partly due to the lower (and thus less realistic) variability of MOS forecasts. The lower variability of MOS, especially in the medium range, is an obstacle for the use of it for forecasts of more severe weather. Here, the more variable solutions of the EPS serve as an important additional guidance.

The application of post-processing lead to largely reliable probability of YES/NO precipitation (PoP) forecasts (fig. 3), yet with a slight underconfidence in situations of low PoP.

Figs. 4-5a,b show two things: i) the MOS technology performs better than a perfect prog technology (AFREG) ; ii) mixing post-processed products from both models leads to a moderate improvement of the forecast, especially in the medium range, where the gain in skill is about half a day.

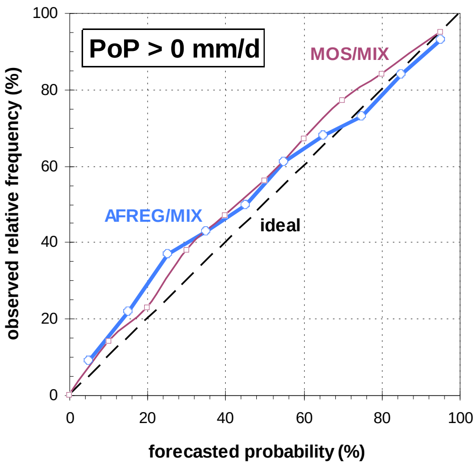

Fig. 3 Reliability diagram (6 stations, 12/07 - 11/08, day+2 ... day+7; only up to day+5 for MOS(MIX))

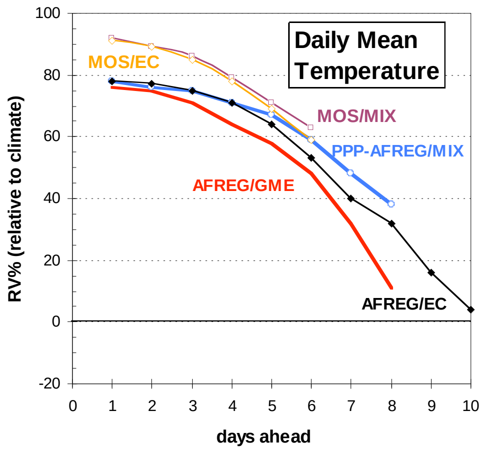

Fig. 4 Forecast skill RV for Daily Mean Temperature (DWD, 6 stations, 12/07 - 11/08

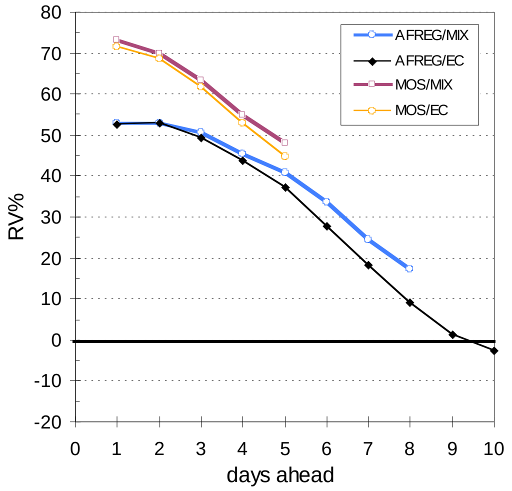

Fig. 5a Forecast skill RV as a function of range, averaged for all predictands taken in table 1 (without PET and RR)

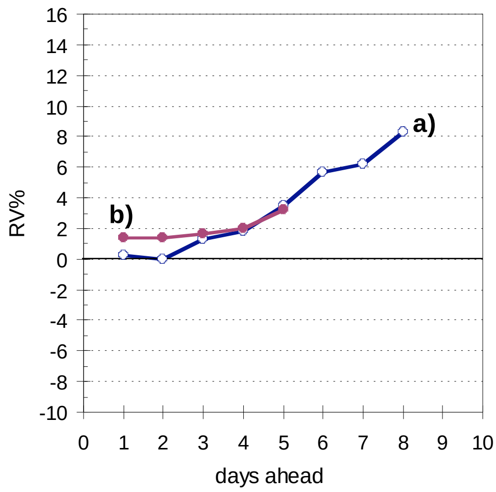

Fig. 5b follows from fig. 5a:
a) Blue line:
RV(AFREG/MIX) - RV(AFREG/EC)
b) Claret red line:
RV(MOS/MIX) - RV(MOS/EC)

EPS products are only verified in a PP form as a Kalman filtered mean of the ensemble for continuous variables and as a relative frequency for probability forecasts, respectively. The verification is done against point observations from Synop's.

Up to 3 days ahead, MOS/MIX presents by far the best guidance. In the medium range, AFREG/MIX is of similar quality compared to MOS/MIX for PoP, whereas the Kalman-filtered EPS is most suitable as an additional guidance for wind speed, cloud cover forecasts and latterly minimum and maximum temperature (Fig. 6). All guidances outperform climate forecasts at least up to day 7.

Probability of YES/NO precipitation forecasts continued to slightly underestimate the PoP, with MOS/MIX exhibiting the best resolution followed by AFREG/MIX (Fig. 7). Stronger events (>5mm/d, Fig. 7) were hardly ever forecasted from the EPS with a sizeable probability, which is only partly attributable to the mismatch between areal precipitation forecasts and point observations. On the other hand, MOS/MIX achieved a good calibration for this rather rare event.

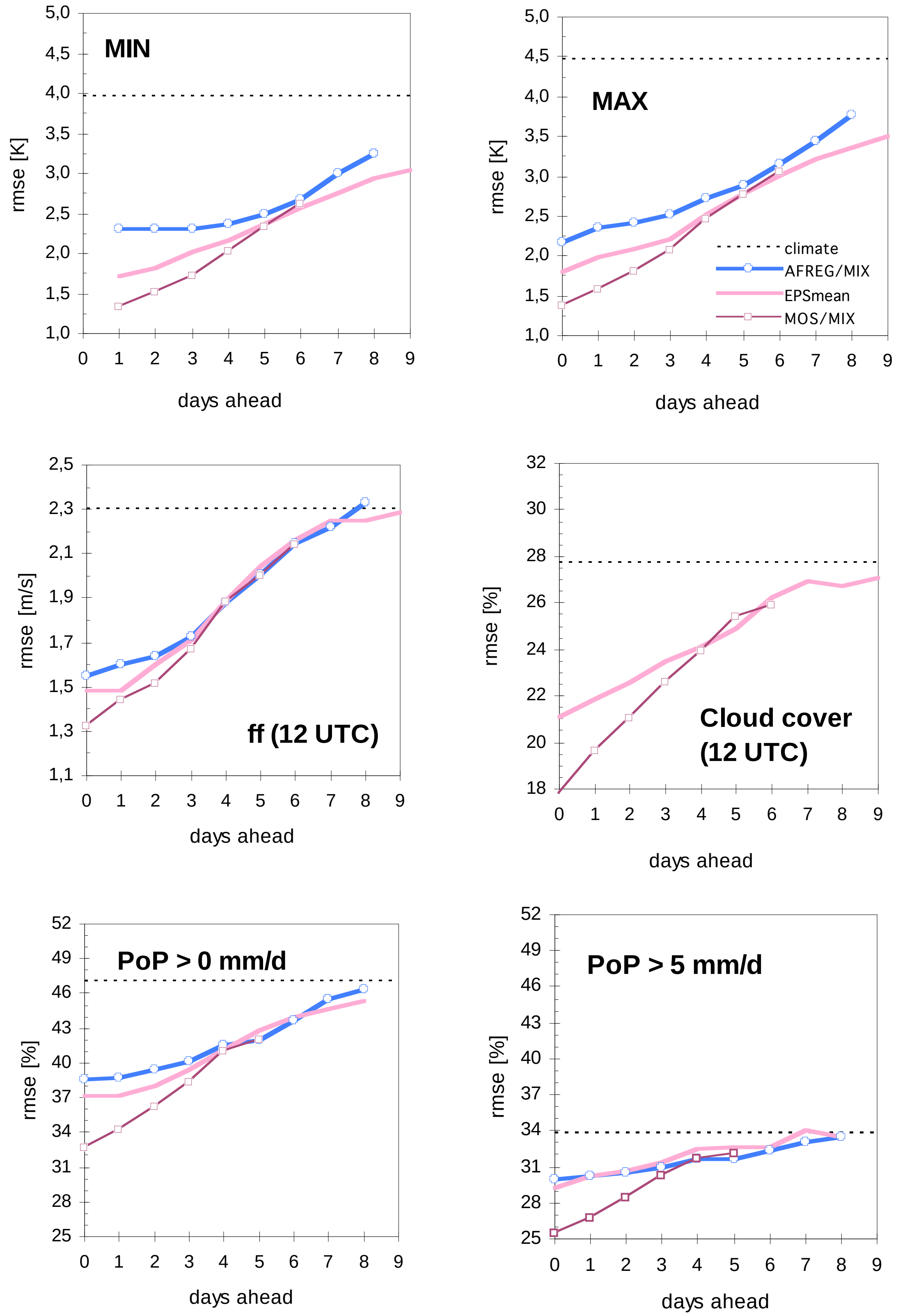

Fig. 6 DMO(EPSmean)+KAL (pink) versus AFREG-MIX (blue) and MOS/MIX (magenta), dotted line = rmse(climate).
Sample: 12/07 - 11/08, DWD (5 stations).The EPS forecast for cloud cover, wind speed ff and maximum temperature MAX is the arithmetical mean of all 51 ensemble members. PoP forecast is the relative frequency of the "yes-event forecast". Notice, rmse is identical to SQRT(BS), BS = Brier Score.

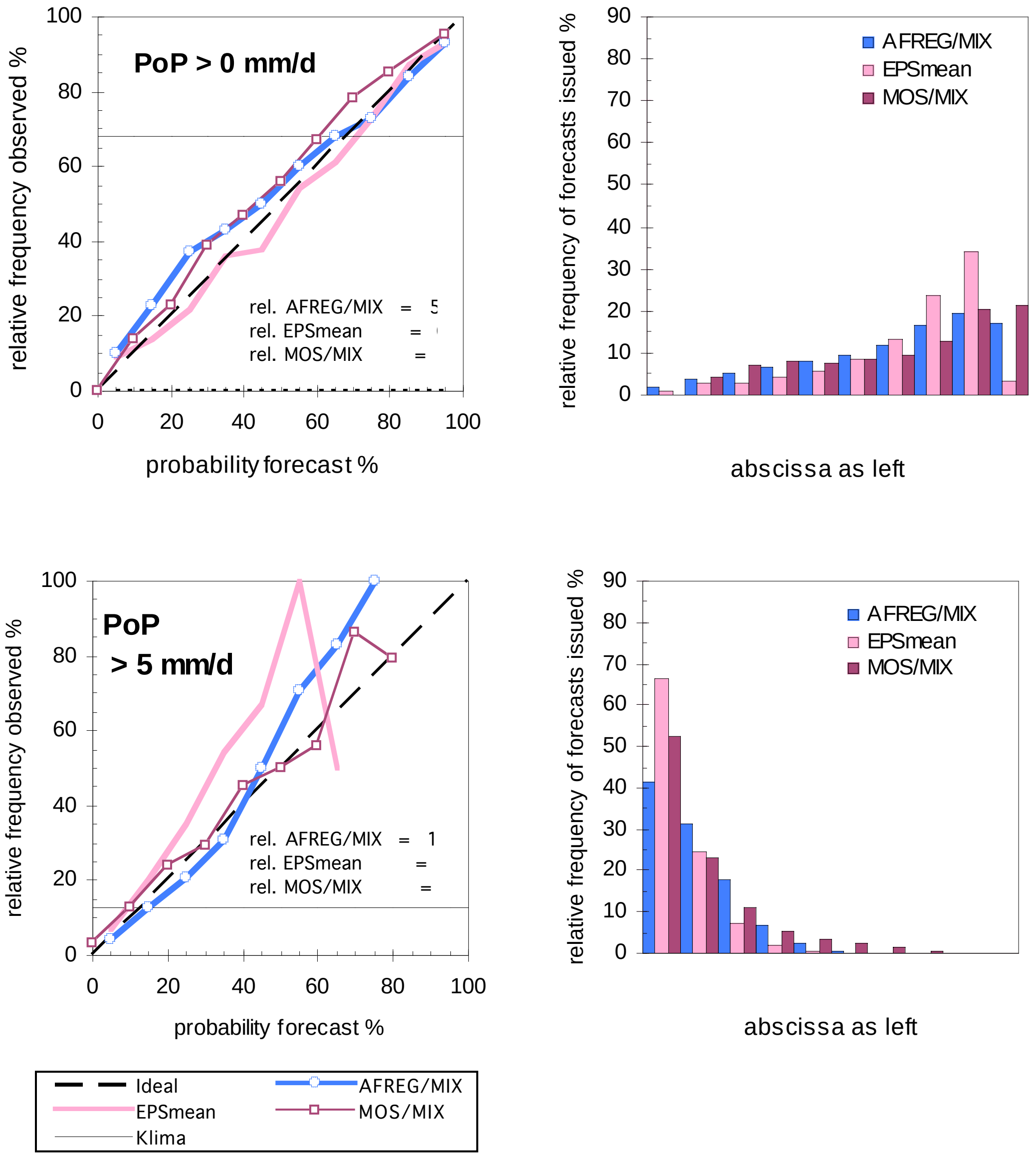

Fig. 7 Reliability (left) and resolution (right) of probability of precipitation (PoP) forecasts by EPS (pink), AFREG/MIX (blue) and MOS/MIX (magenta). 5 stations, day +2 ... +7, (MOS/MIX only up to day+5). No resolution forecasts of sample climatological probability are indicated by the horizontal line.

3.1.4 End products delivered to users

3.2 Subjective verification

3.2.1 Subjective scores

3.2.2 Synoptic studies

4. References to relevant publications

Göber, M., Zsoter, E. and **D.S. Richardson**, 2008: Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteorol. Appl.*, **15**: 359-365.