Verification of ECMWF products at the Deutscher Wetterdienst (DWD)

DWD – Martin Göber

1. Summary of major highlights

The high resolution ECMWF model forms in conjunction with DWD's model GME the general operational data base. For medium range forecasting the EPS is used additionally; in the short range the LEPS (Local model nested into EPS clusters) provides ensemble information. Extensive use is made of the Extreme Forecast Index (EFI) to estimate the likelihood of extreme events.

The introduction of a combined GME-MOS and ECMWF-MOS in 2004 continues to lead to a further increase in forecast accuracy.

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 PPM 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.

ECMWF high resolution forecasts will be used in the near future for the production of a probabilistic warning guidance based on the MOS technology.

Some EPS surface variables are 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 in conjunction 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

3.2

3.1.1 Direct ECMWF model output

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 7% (0,7 gpm) in the short range from 2005 to 2006 and by about 12% for the GME. ECMWF MSLP error growth with forecast range is about half a day better than for DWD-GME in the short range and one day in the medium range (fig. 2). The RMSE's of the GME model for MSLP have decreased in 2006 by about 0,1 hPa in the short range and by the same amount for the ECMWF model in the medium range.

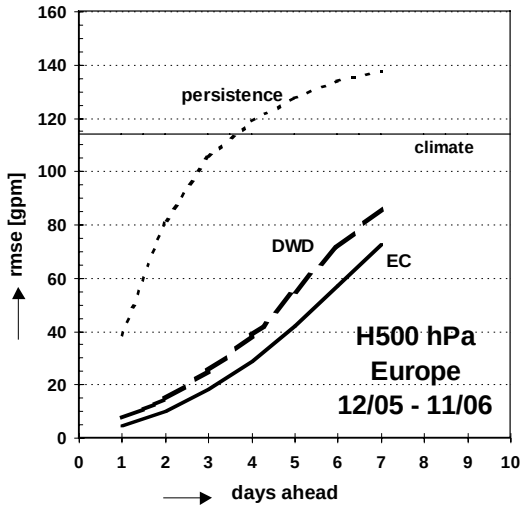

Fig. 1 RMSE 500hPa geopotential over Europe. DWD (Numerical Weather Prediction model GME), EC (high resolution ECMWF model), persistence (analysis from the initial state is used as a forecast for all following days), climate (long term mean of the predictand (H500, MSLP) serves as a constant forecast).

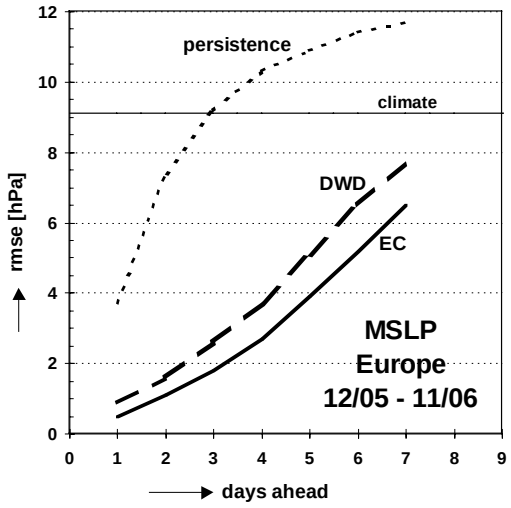

Fig. 2 Same as fig. 1, but for RMSE of mean sea level pressure.

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). PPP is generated for several areas of the whole Germany, but verified against point observations at 6 stations. MOS/MIX = PPP, 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*)_, 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	2,50	2,60	2,75	2,92	3,20	3,49	3,81	4,36	59
	MOS/MIX	1,62	1,94	2,25	2,63	2,97				71
MAX	AFREG/MIX	2,45	2,60	2,77	3,11	3,53	3,90	4,16	4,82	63
	MOS/MIX	1,85	2,11	2,45	2,88	3,29				72
SD	AFREG/MIX	24,9	25,4	26,1	27,3	28,6	29,6	29,8	31,7	27
dd[1)]	AFREG/MIX	42,0	46,0	52,0	57,1	67,3	73,5	76,9	87,0	63
	MOS/MIX	34,3	40,0	46,9	54,5	62,5				70
ff	AFREG/MIX	1,64	1,78	1,84	2,01	2,14	2,24	2,23	2,19	26
	MOS/MIX	1,48	1,66	1,81	1,94	2,04				33
PoP	AFREG/MIX	36,3	37,2	38,7	40,9	42,6	44,1	45,2	47,5	35
	MOS/MIX	34,9	36,6	39,0	41,0					36
PET	AFREG/MIX	0,704	0,726	0,743	0,764	0,804	0,838	0,837	0,964	37
HSS%										$\overline{HSS}$
RR	AFREG/MIX	58	55	50	44	37	33	26	0	52
	MOS/MIX	62	58	51	46					54

Table 1 Verification of operational medium range forecasts for 6 stations in Germany (Hamburg, Potsdam, Düsseldorf, Leipzig, Frankfurt/M., München), 01/06- 12/06. Day of issue = day +0 = today at noon. [1)] Here, persistence is used as a 'reference forecast'.

The skill (RV) of the forecasts in 2006 was better than 2005 by 1-5% for all variables (table 1). MOS forecasts improved more than in AFREG. 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 a perfect prog technology leads to a very reliable probability of YES/NO precipitation (PoP) forecast, whereas MOS/MIX exhibits a sizeable underestimation of the PoP.

Figs. 4-5a,b show two things: i) the MOS technology performs better than a perfect prog technology (AFREG) ; ii) mixing PP from both models leads to a moderate improvement of the forecast, especially in the medium range.

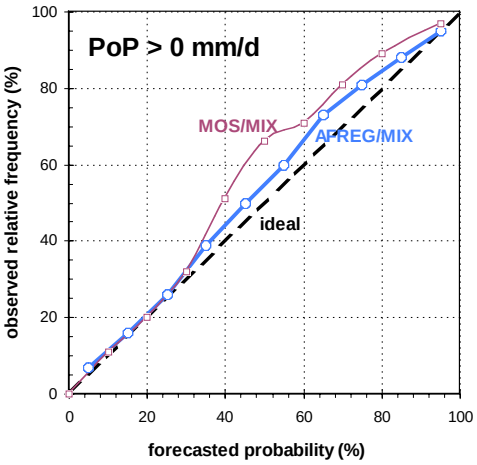

Fig. 3 Reliability diagram (6 stations, year 2006, day+2 ... day+7; only up to day+5 for MOS(MIX))

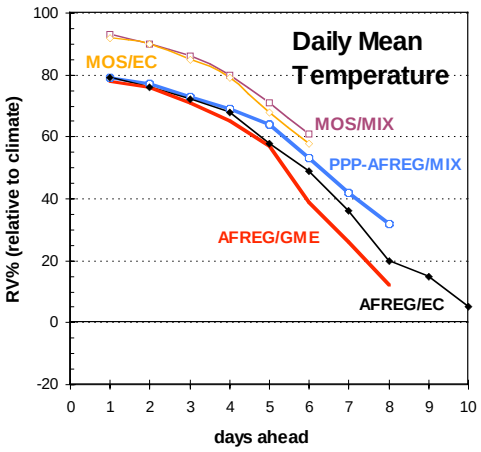

Fig. 4 Forecast skill RV for Daily Mean Temperature (DWD, 6 stations, 2006)

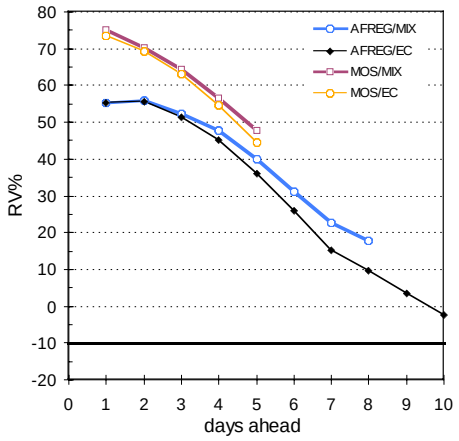

Fig. 5a Forecast skill RV as a function of range, averaged for all predictands taken in table 1 (without PET and RR)

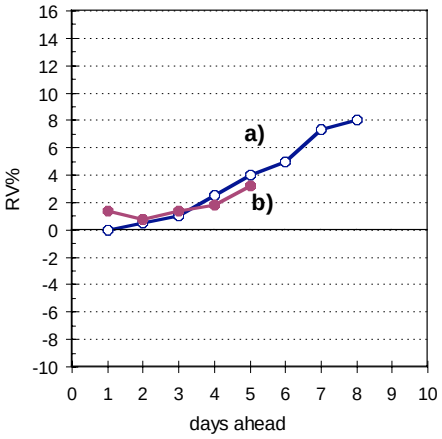

Fig. 5b Follows from fig. 5a:
a) Blue line:
RV (AFREG/MIX) – RV(AFREG/EC)
b) Claret red line:
RV(MOS/MIX) – RV(MOS/EC)

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 maximum temperature (Fig. 6).

Probability of YES/NO precipitation forecasts continued to 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, 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.

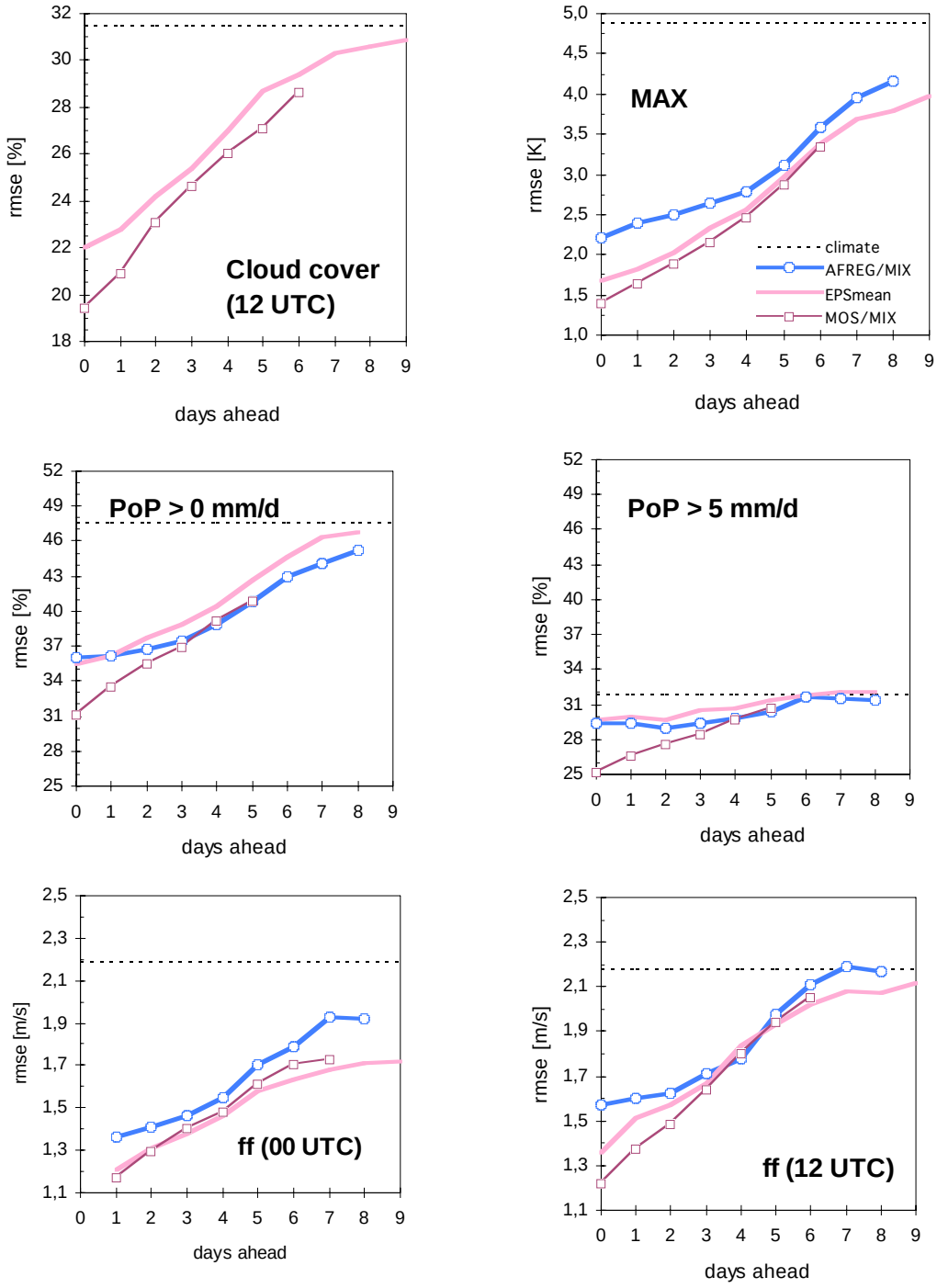

Fig. 6 DMO(EPSmean)+KAL (pink) versus AFREG-MIX (blue) and MOS/MIX (magenta), dotted line = rmse(climate).
Sample: 01/06 - 12/06, DWD (5 stations).The EPS forecast for cloud cover, wind speed ff and maximum temperature MAX is the arithmetical mean of all 51 ensemble members. PoP forecast is the relative frequency of the "yes-event forecast". Notice, rmse is identical to SQR(BS), BS = Brier Score.

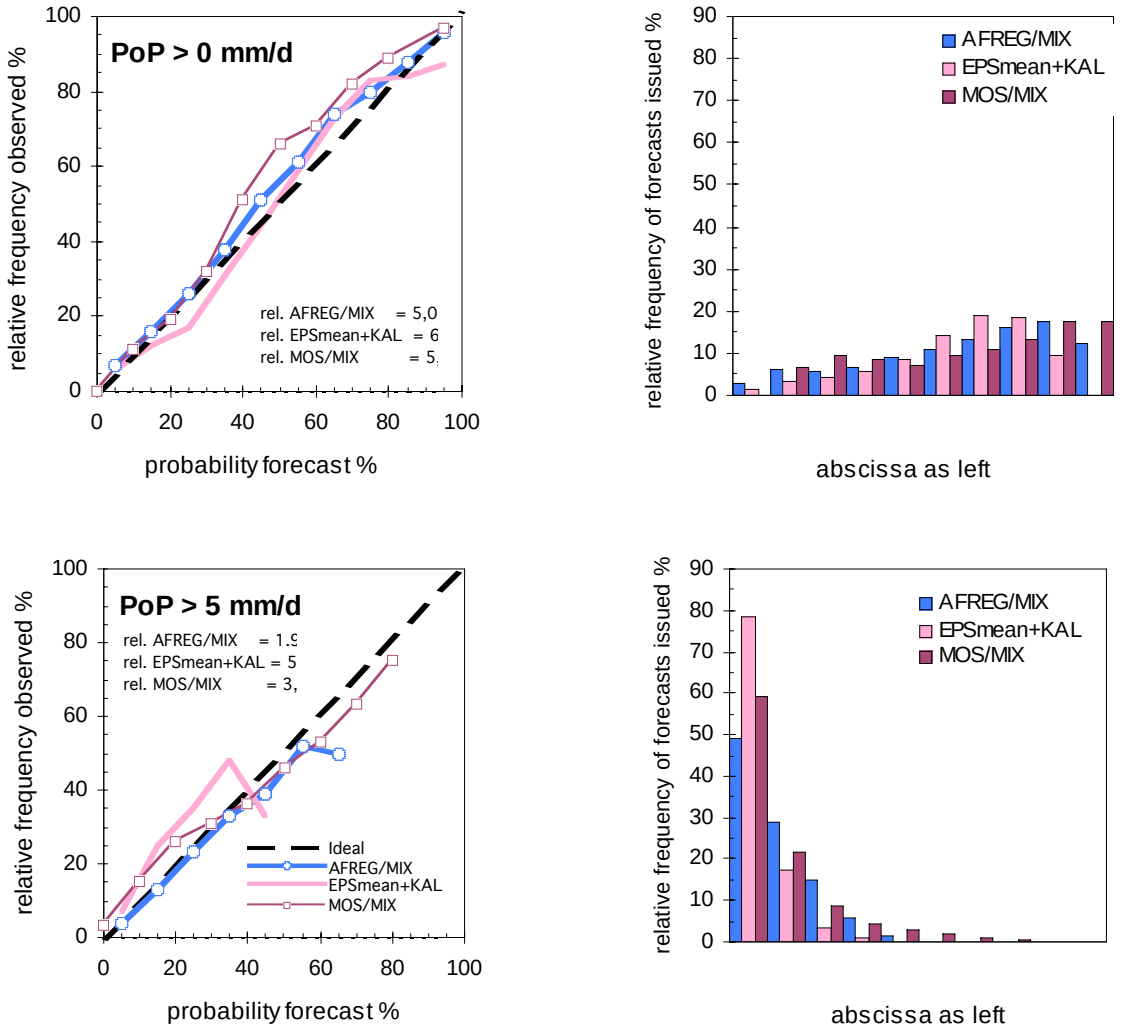

Fig. 7 Reliability of PoP forecasts by EPS (pink), AFREG/MIX (blue) and MOS/MIX (magenta) 5 stations, day +2 ... +7, (MOS/MIX only up to day+5)

3.1.3 End products delivered to users

3.2 Subjective verification

3.2.1 Subjective scores

3.2.2 Synoptic studies

4. References to relevant publications