Verification of ECMWF products at the Deutscher Wetterdienst (DWD)

DWD – Martin Göber, Cristina Primo

1. Summary of major highlights

The usage of a combined GME-MOS and ECMWF-MOS continues to form the basis for the production of local short and medium range forecasts. It has now been augmented in the short range by forecasts from the regional model COSMO-EU in the best available guidance called Objectively Optimised Guidance (OOG). ECMWF high resolution forecasts in conjunction with GME forecasts are also 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 (now both 12 and 00 UTC run) 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 weighted averaging of the two interpretations to form „AFREG/MIX" and "MOS/MIX". 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. Furthermore, the ECMWF model is one of four driving models for the pre-operational high resolution COSMO-DE-EPS.

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. There is growing usage of the new products as presented on the ECWMF website.

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 not improved in the short range from 2009 to 2010, but it improved in the medium range by 2–4 %. GME short range forecast errors decreased by 3%. ECMWF MSLP error growth with forecast range is about one day better than for DWD-GME in the short range (Fig. 2). The RMSE of the GME model for MSLP has improved from 2009 to 2010 by 0,1 hPa (4%), yet for the ECMWF model the error increased by 0,1 hPa.

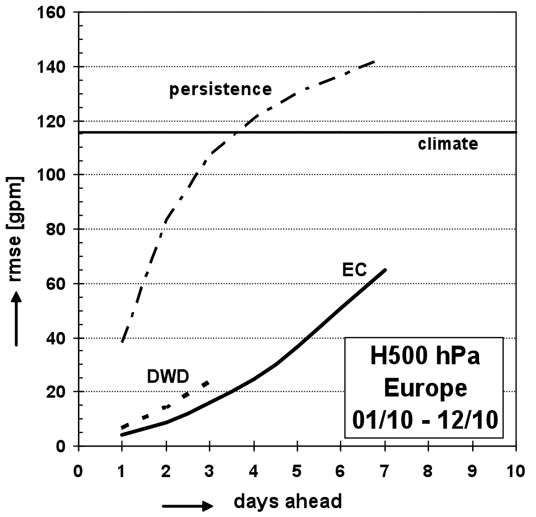

Figure 1 RMSE 500hPa geopotential over Europe. DWD (Numerical Weather Prediction model GME), EC (high resolution ECMWF model), persistence (analysis from the initial state is used as a forecast for all following days), climate (long term mean of the predictand (H500, MSLP) serves as a constant forecast).

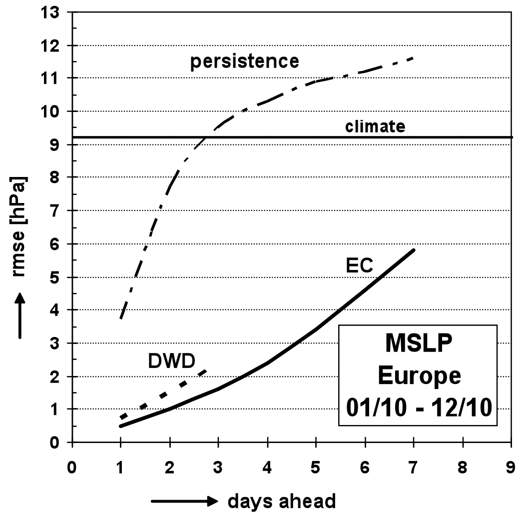

Figure 2 Same as Fig. 1, but for RMSE of mean sea level pressure.

3.1.3 Post-processed products

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

Predictands

MIN = daily minimum 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
$\overline{\textbf{HSS}}$ = mean value for day 2 ... 7

rmse		day +2	+3	+4	+5	+6	+7	+8	rmse* (climate)	RV [%]
MIN	AFREG/MIX	2,42	2,53	2,61	2,81	3,11	3,42	3,69	4,10	57
	MOS/MIX	1,77	2,03	2,28	2,51	2,87				68
MAX	AFREG/MIX	2,53	2,66	2,85	3,18	3,57	3,98	4,28	4,92	63
	MOS/MIX	1,80	2,19	2,58	3,07	3,50				70
SD	AFREG/MIX	24,3	25,1	25,8	26,7	27,8	29,0	29,4	29,4	19
dd[1)]	AFREG/MIX	45,8	48,2	53,5	60,1	66,9	72,8	80,2	94,4	66
	MOS/MIX	34,0	40,4	49,4	58,5	67,4				72
ff	AFREG/MIX	1,61	1,72	1,80	1,93	2,06	2,15	2,21	2,11	25
	MOS/MIX	1,45	1,63	1,76	1,90	2,01				30
PoP	AFREG/MIX	38,3	39,5	40,9	41,7	43,3	45,0	46,0	45,1	21
	MOS/MIX	36,6	37,9	39,8	41,6					23
PET	AFREG/MIX	0,737	0,753	0,792	0,817	0,843	0,903	0,917	0,941	26
HSS%										$\overline{\text{HSS}}$
RR	AFREG/MIX	47	42	39	34	27	22	19		41
	MOS/MIX	53	47	41	31					43

Table 1 Verification of operational medium range forecasts for 6 stations in Germany (Hamburg, Potsdam, Düsseldorf, Leipzig, Frankfurt/M., München); 01/2010 – 12/2010; 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 forecasts was significantly lower in 2010 than in many of the preceding years. Particularly **minimum temperature forecasts were much too low** during a prolonged very cold period in January and February. GME exhibited the same problem. It was not possible to decide whether the change in ECMWF resolution contributed to this problem, since both the ECMWF and GME resolution changed at almost the same time and the weather itself was also unusual during that time.

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). 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 very moderate improvement of the forecast. However, in the medium range the gain in skill due to mixing is about half a day.

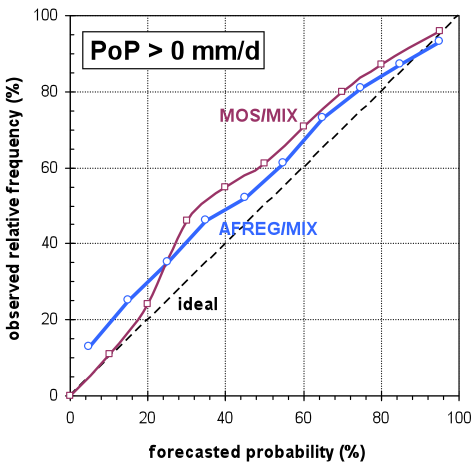

Figure 3 Reliability diagram (6 stations, 01/10 – 12/10, day+2 ... day+7; only up to day+5 for MOS(MIX))

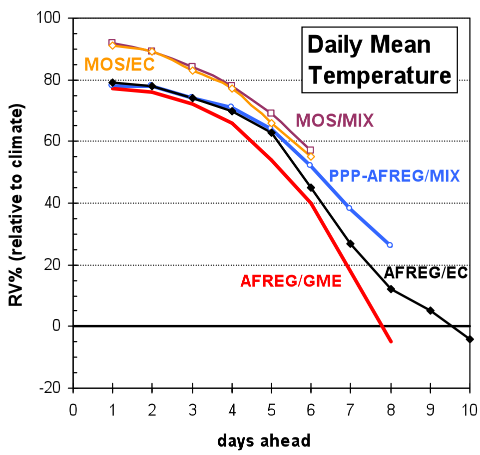

Figure 4 Forecast skill RV for Daily Mean Temperature (DWD, 6 stations, 01/10 – 12/10

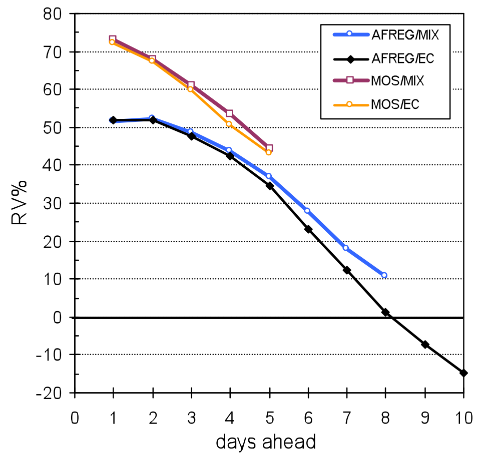

Figure 5a Forecast skill RV as a function of range, averaged for all predictands taken in table 1 (without PET and RR)

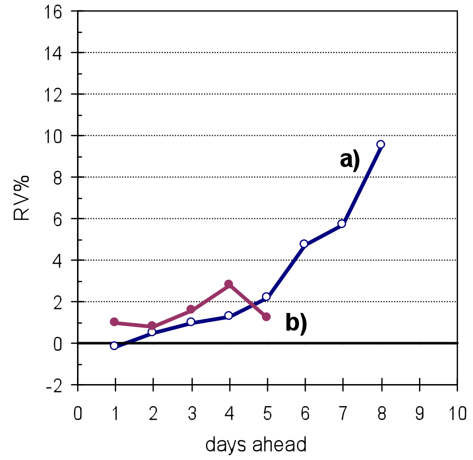

Figure 5b Follows from Fig. 5a: a) Blue line: RV(AFREG/MIX) – RV(AFREG/EC) b) Claret red line: RV(MOS/MIX) – RV(MOS/EC)

Finally, first results (Fig. 6) are presented of the verification of a probabilistic guidance for precipitation warning purposes (WarnMOS, a MOS for extreme events based on ECMWF and GME models). Note, that one of the goals of using the MOS technology is to produce calibrated forecasts, which is particularly difficult for rare event forecasts, because of an often unstable statistics.

The forecasts turn out to be quite reliable over the whole range of forecast values. Yet the forecast probability never exceeded 53% for this data set. Note that, whereas non calibrated, raw EPS extreme event forecasts sometimes show high confidence (high probability), this is mostly not justified, i.e. the forecasts are over-confident. On the contrary, the MOS technology "learns" that its knowledge is not sufficient to issue extreme event forecasts with very high probability.

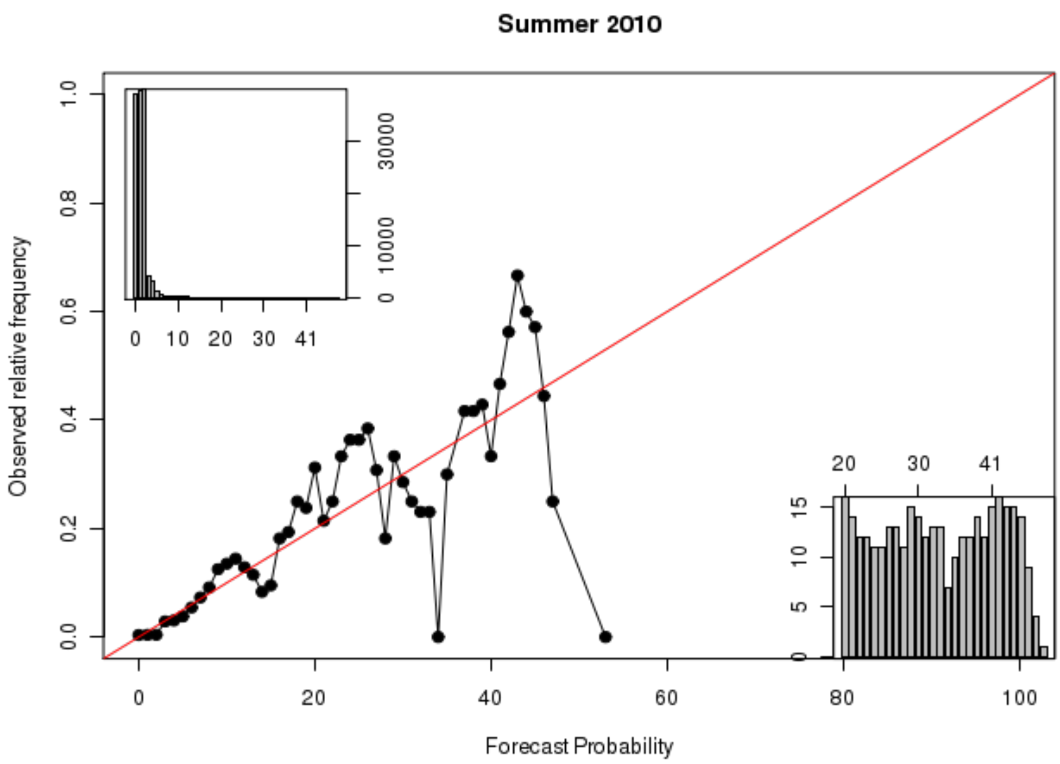

Figure 6 Reliability diagram of 18 hour forecasts of 6-hourly precipitation accumulation above the warning criterion of 20 mm/6h for the summer 2010 (April–September). The upper left inlet shows the frequency distribution of the forecasts, the lower right inlet is a zoom into the high forecast end of the complete distribution to highlight the small sample size. The diagonal signifies perfect reliability.