Verification of ECMWF products at the Deutscher Wetterdienst (DWD)

DWD – Dr. Marcus Paulat, Dr. Ulrich Damrath, Gabriele Schweigert

1. Summary of major highlights

The usage of a combined GME-MOS and ECMWF-MOS at DWD continues to form the basis for the production of local short and medium range forecasts. It has been augmented in the short range by forecasts from the regional model COSMO-EU in the best available guidance called Objectively Optimised Guidance (OOG, available up to +168 h). 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. The use of ECMWF Ensemble data within DWD's visualisation software NinJo continues, too.

The high resolution ECMWF model is one of four driving models for the high resolution COSMO-DE-EPS which is operational since May 2012 and well accepted in forecasting deep convection fairly good especially compared to deterministic information.

2. Use and application of products

2.1 Post-processing of model output

2.1.1 Statistical adaptation

The high resolution ECMWF model (both 12 and 00 UTC run) and DWD's model GME are statistically interpreted up to 10 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". Because of the change from GRIB1 to GRIB 2 format in DWD mid of 2014 and the good skills from MOS technique AFREG/MIX is replaced by "MOSGEB" (forecasts produced for certain areas) in operations since October 2012. Because of this we show results for both the old AFREG data and the new MOSGEB data in this report once.

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 which will form the basis to produce automated warn status proposals.

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 COSMO-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, use of the Extreme Forecast Index (EFI) is made. There is high usage of the products as presented on the ECWMF website. To make some of these products more easily usable in the context of DWD environment (layer technique for comparison to other meteorological data), ECMWF-EPS, LEPS and COSMO-DE-EPS products are displayed within NinJo since two years now. As an example, figure 1 illustrates the visualisation of ECMWF ensemble data in a meteogram format for a certain time and location (available at each model grid point). The data can be viewed in 6 hourly time steps up to 240 hours for (top to bottom) total cloud cover, precipitation, CAPE, wind gusts and 2m temperature (Tmax in red, Tmin in blue). We also added deterministic information from ECMWF's high resolution model as symbols (total cloud cover), columns (precipitation) and lines (other parameters). The very helpful visualisation of wind direction ("pie slices") shown in ECMWF ensemble meteograms is still in planning. Note, that for winter season the CAPE chart is exchanged by snowfall.

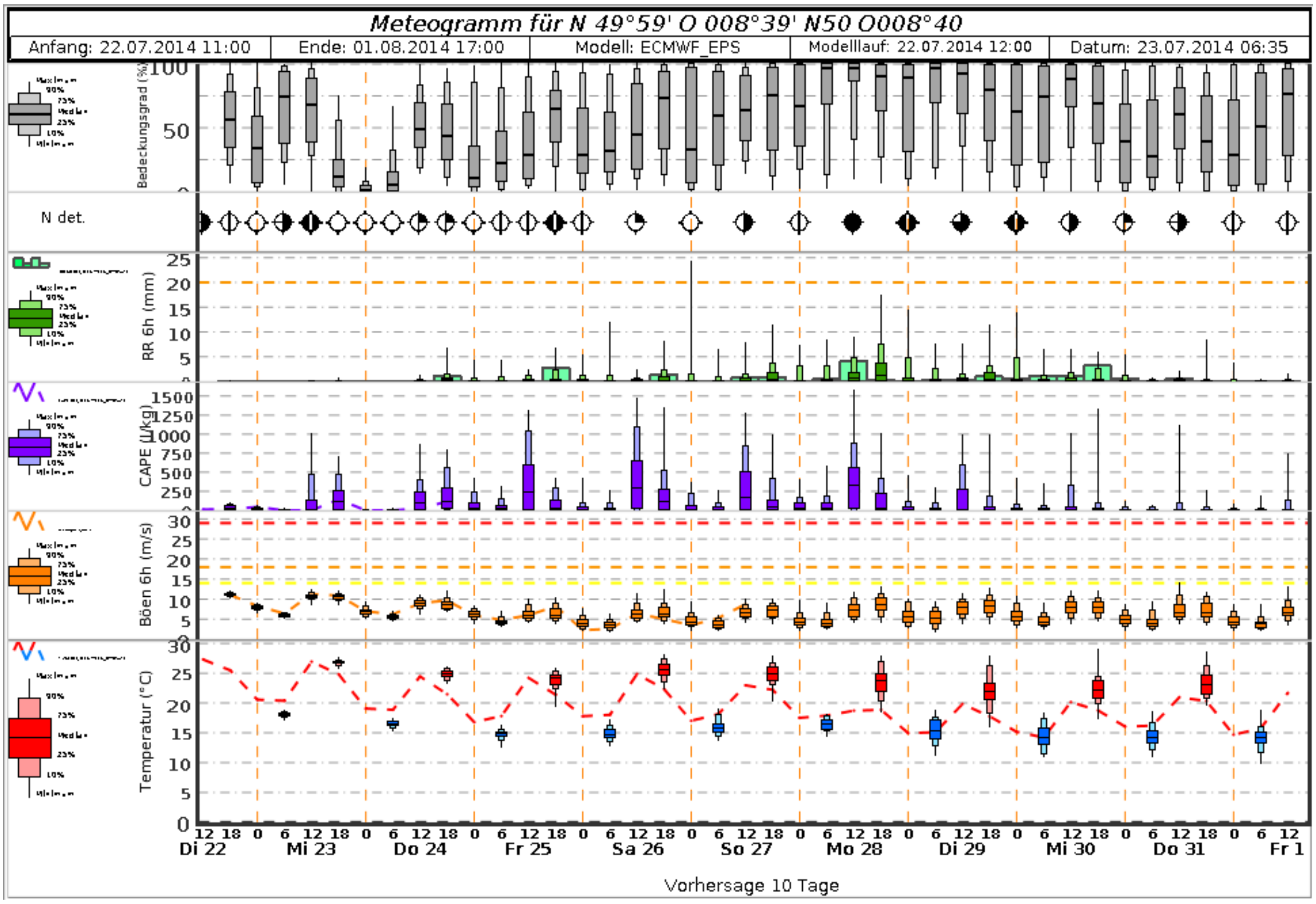

Fig. 1: Screenshot of a NinJo ensemble meteogram showing ECMWF-EPS data, location close to DWD centre at Offenbach/Main, 12 UTC run of 22nd of July 2014. Also added are data from the corresponding deterministic run. Horizontal coloured and dashed lines show warning threshold at DWD.

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

Following the results of former editions, upper air forecasts from ECMWF continued to exhibit smaller errors than DWD-GME forecasts (Fig. 2). The RMSE of the ECMWF model for 500hPa geopotential height has not significantly improved in the short range from 2009 to 2013. ECMWF MSLP error growth with forecast range remains about one day better than for DWD-GME in the short range (fig. 3) and has been improved at ECMWF especially for forecasts 4-7 days ahead.

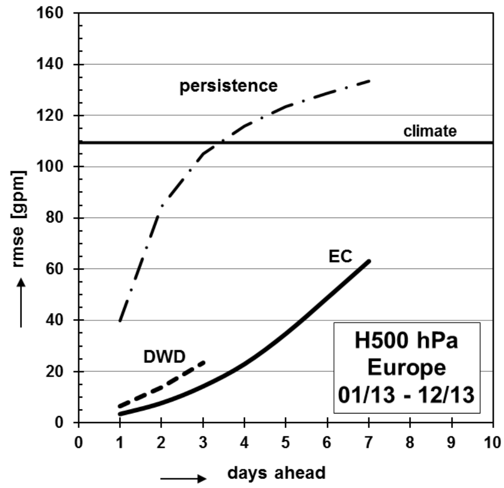

Fig. 2: RMSE 500hPa geopotential over Europe. DWD (Numerical Weather Prediction model GME), EC (high resolution ECMWF model), persistence (analysis from the initial state is used as a forecast for all following days), climate (long term mean of the predictand (H500, MSLP) serves as a constant forecast.

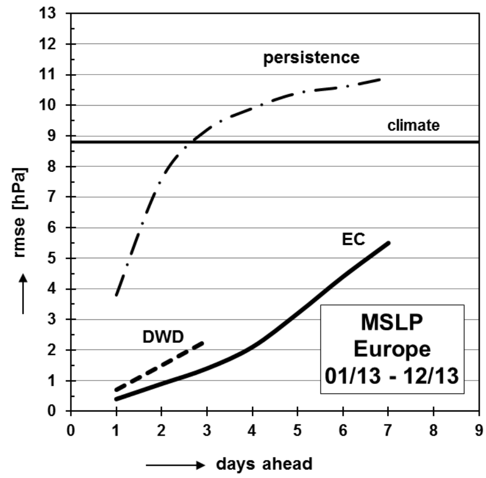

Fig. 3: Same as fig. 2, but for RMSE of mean sea level pressure.

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 is generated for several areas of 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, verified at 6 stations
MOSGEB = Area forecasts calculated by averaging selected point forecasts of MOS/MIX output, verified at 6 stations. AFREG was replaced by MOSGEB in operations since October 2012

Verification measures

RMSE is used for both categorical and probabilistic forecasts (equals square root of the Brier Score)

RV = Reduction of Variance against reference, $1-(\text{RMSE}/\text{RMSE}^*)^2$
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								RMSE*	
		+2	+3	+4	+5	+6	+7	+8	+9	(climate)	RV [%]
MIN	AFREG/MIX	2,43	2,49	2,64	2,76	3,08	3,38	3,59			57
	MOSGEB	1,72	1,90	2,14	2,40	2,76	3,00	3,29	3,49	4,10	71
	MOS/MIX	1,51	1,70	1,98	2,28	2,65					74
MAX	AFREG/MIX	2,44	2,53	2,78	3,07	3,47	3,73	4,06			64
	MOSGEB	1,83	2,03	2,35	2,77	3,23	3,63	4,06	4,26	4,78	73
	MOS/MIX	1,63	1,88	2,26	2,72	3,19					75
SD	AFREG/MIX	24,6	25,0	25,7	26,7	27,8	29,4	30,5		30,0	21
	MOSGEB	19,5	22,0	24,3	26,4	28,1	29,4	30,5			30
dd[1)]	AFREG/MIX	45,1	48,4	53,2	59,5	66,4	72,6	78,0			62
	MOSGEB	36,4	42,5	49,9	59,5	66,0	74,9	82,8	88,4	88,9	67
	MOS/MIX	33,1	39,0	47,9	56,7	64,7					70
ff	AFREG/MIX	1,63	1,68	1,79	1,94	2,06	2,13	2,17			26
	MOSGEB	1,42	1,55	1,73	1,86	1,99	2,06	2,09	2,10	2,13	34
	MOS/MIX	1,33	1,47	1,66	1,82	1,97					39
PoP	AFREG/MIX	38,9	39,3	40,2	42,6	44,0	46,5	48,3			25
	MOSGEB	37,1	39,5	41,7	44,7	46,4	48,0	48,8	48,9	46,4	23
	MOS/MIX	34,3	36,5	38,6	41,9						33
PET	AFREG/MIX	0,695	0,706	0,734	0,766	0,791	0,828	0,877		0,905	30
	MOSGEB	0,890	1,07	1,03	1,09	1,14	1,32	1,40			-32
HSS%											$\overline{\text{HSS}}$
RR	AFREG/MIX	48	47	44	38	30	18	15			44
	MOSGEB	60	54	48	37	29	22	16	10	0	50
	MOS/MIX	63	57	49	39						52

Table 1: Verification of operational medium range forecasts for 6 stations in Germany (Hamburg, Potsdam, Düsseldorf, Leipzig, Frankfurt/Main, München); 01/2013 - 12/2013; RMSE and HSS, respectively. Day of issue = day +0 = today at noon. [1)] Here, persistence is used as a 'reference forecast'.

The operational replacement of AFREG by MOSGEB was mainly done because of technical reasons. Although MOSGEB is setup similar concerning parameters and areas there are some more differences in the calculation of these area forecasts. One can easily see that comparing results between both methods shown in table 1. Focussing on AFREG versus MOSGEB first there is a general clear improvement using MOSGEB except precipitation (no clear difference) and PET (AFREG remains clearly better). The nearly unrealistic results of PET forecasted by MOSGEB (especially looking to the RV value) are part of current investigation.

A comparison between MOS/MIX point forecasts of 2013 with results of 2012 show some slight improvement in 2013 except wind direction (slightly worse). Interesting to mention that RMSE for minimum temperatures (MIN) is smaller than 2012 while RV cannot reach the same level (MIN(2013) =74; MIN(2012)=76). In 2013 it was obviously harder to beat the mean climate forecast than in 2012.

As mentioned above the reliability diagram in figure 4 shows AFREG as probably best forecast data to answer the questions of precipitation YES/NO on daily sums. But point data of MOS/MIX are of similar quality. It will be part of further investigations why MOSGEB performs not that well. More general, all of these forecast methods show an "underforecasting" of probability of precipitation > 0 mm/day.

Figs. 5 and 6 contain two major results as it could be seen the last years:

i) the MOS technology performs better than a perfect prog technology (AFREG);

ii) mixing post-processed products from both models lead 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.

To summarize, the quality of mixed post-processed forecasts reaches a satisfying level. The differences between 2013 and 2012 are not significant.

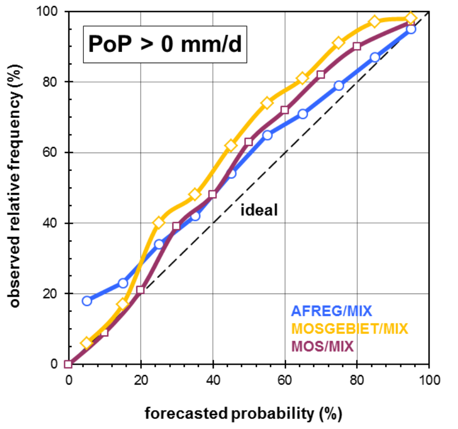

Fig. 4: Reliability diagram for precipitation > 0 mm/day (6 stations, 01/13 – 12/13, day+2 ... day+7; only up to day+5 for MOS/MIX)

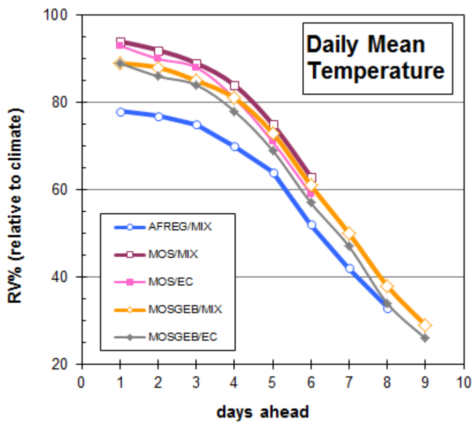

Fig. 5: Forecast skill RV for Daily Mean Temperature (DWD, 6 stations, 01/13 – 12/13)

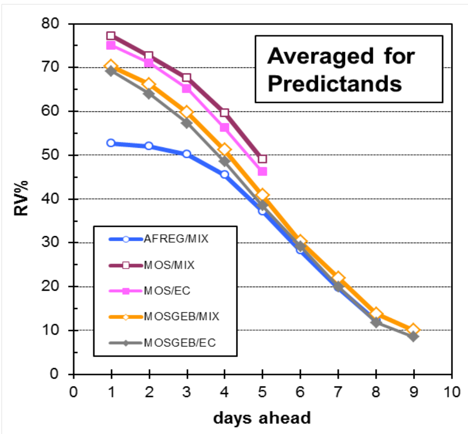

Fig. 6: Forecast skill RV as a function of range, averaged for all predictands taken in table 1 (without PET and RR)