The German Practice in Medium-Range Forecasting (MRF) and Remarks on Developments

Konrad Balzer and Peter Emmrich

Deutscher Wetterdienst, Offenbach

General Remarks

Numerical weather forecasts are characterized by intrinsic uncertainties which grow with increasing forecast time. This is caused through both errors in the initial state as well as errors in the formulation of the numerical prediction model. It was and still is of high importance for the practice of weather forecasting and, in general, for economic reasons to quantify this uncertainty. This is what we call skill prediction and, ensemble prediction (EP) clearly aims at this. Unfortunately, the success of EPS in skill prediction is far behind the expectation. Today, we know that the improvement of skill by averaging the ensemble is the most useful and successful effect of EP.

The EPS approach of the ECMWF is based on the assumption that the uncertainty in the initial analysis is the main (only) source of uncertainty which means that the model used would be perfect. This is basically a disadvantage because the latter is not true. On the other side, the EPS now consists of 51 members. This is a remarkable progress and indeed an advantage.

A cost-effective alternative has been practized in the DWD for several years. It is the averaging of the two operational models German GM(T106) and ECMWF(T213). This is to be considered as a very poor mens 2-member-ensemble with two thoroughly important characteristics:

- It consists of two high developed models (with respect to MRF) and indeed,
- both models start from different analyses.

The advantage is that it reflects both sources of intrinsic uncertainty. But the minimum of ensemble size is to be considered as a decisive disadvantage. This simple approach works comparably well. On this background it was evident to test an expansion of the ensemble size to 4 members (models).

In the following a skill comparison between the operational poor mens 2-member-ensemble and the EPS will be presented and in addition, preliminary results regarding the expanded poor mens 4-model-ensemble are mentioned.

Basic guidance in use for practical MRF

The operational models GM and ECMWF are to be considered as the first and most important guidance in MRF.

On average both models show the same level of skill. Nevertheless, normally they diverge remarkably in medium-range from day to day and, nobody knows a priori which model performs better. Therefore, in order to make strictly use of the information provided by both models a simple linear combination of the two different forecasts has to be applied. It is a known fact that this is a quite successful approach in MRF - the simplest and most cost-effective EP approach. Indeed, the success is caused by the averaging effect which reduces both the error itself and the remarkably large and unwanted error variance (variance in time and among models).

Based on this findings a simple statistical 2-component PPM approach (called AFREG) is applied to both operational models in order to derive the real near surface weather parameter needed. The interpretation scheme uses the 1000-hPa- and 500-hPa-topography as basic predictors taken from a few grid points around Germany. The parameters interpreted are

Minimum and maximum temperature,
Relative sunshine duration,
Amount of daily precipitation and
Probability of precipitation yes/no and >5 mm resp.,
Wind direction and speed,
Probability of thunderstorm and
Probability of fog.

These parameters are determined for several areas of Germany which are roughly defined according to their climatological characteristics. Finally, both model interpretations are averaged in order to reduce the error and the error variance. The final result is called >AFREG MIX< (identical with MIX2 in this paper). It scores remarkably better than each single model interpretation and up to now still slightly better than EPS DMO.

Use of EPS guidance

A lot of information provided by the EPS is available in the daily routine work:

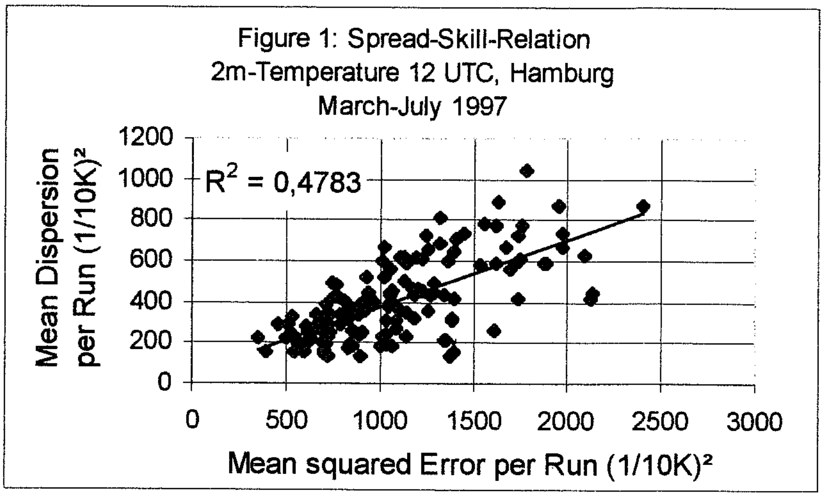
The cluster mean fields for domains G and C together with the corresponding fields of standard deviation,
6 plume diagrams and in addition
any weather parameter files provided by the Centre can be easely made available graphically at PC.

Until now, this information is used only in order to try to judge the performance of the operational forecasts. There are two main problems that have to be taken into account in this context:

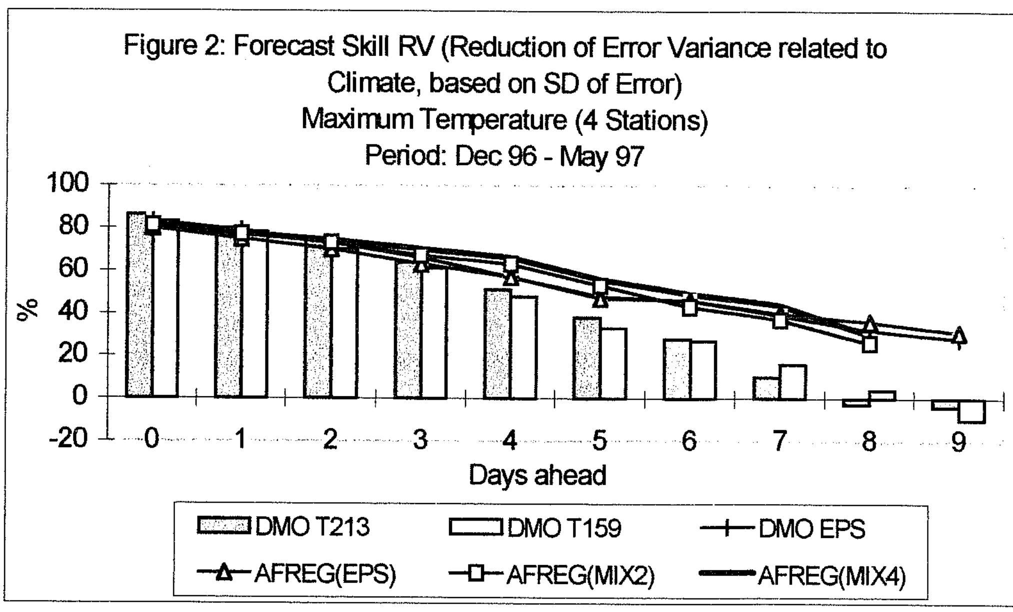
- the skill-spread-relationship produced by the EPS does not work as expected and
- the skill of EPS is still less than that of AFREG MIX. As seen below, especially the EPS temperature forecast is too much affected by a negative bias.

Remarks on the EPS spread-skill-relation

It is quite clear that the error and the ensemble dispersion are highly positively correlated. This is true with respect to the forecast error of each single ensemble member as well as to the error of the EMean or cluster mean. Unfortunately, this is not of prognostic value because small errors are always connected to small dispersion as a characteristics of short-range forecasts while in medium-range any combination can occur. On the other side, as seen in Figure 1, it is interesting that the ensemble as a



whole reacts quite reasonable on the general atmospheric predictibility at starting time expressing the general predictability in terms of mean dispersion per run. A large mean dispersion per run is statistically positive correlated with a large mean error per run. In this sence, one can show that the generally varying underlying atmospheric predictability affects the forecast skill. But, as shown, only up to 40-50% of the error variance can be explained in this way.



Performance characteristics of all the guidance available

As seen in **Figure 2,** the skill of the single forecasts T213 and T159 drops most rapidly to zero. Compared with the RV obtained by averaging (ensembling) forecasts (MIX2, EPS) the difference is at least 30% at day 8. This simply means, it does not make sense to base MRF only on the deterministic single model approach. AFREG MIX = MIX2 scores better than EPS for days 3 to 5. Afterwards the EPS Mean becomes slightly better. It can be seen that the experimental 4-model-mixture works most successfully in

medium-range. This may be to consider as a quite important signal, perhaps with some influence on a future design of EP.
More details are given in Figures 3 and 4. In Figures 3a,b and 4a the first three pairs of columns on the left represent the RV of the simple DMO, while the remaining three pairs on the right show the RV of AFREG applied to the other basic guidances EPS Mean, MIX2 and the experimental MIX4. As mentioned, MIX2 means the average of AFREG applied to German GM and ECMWF and the experimental MIX4 expresses the average of AFREG
applied to the 4 models German GM, ECMWF, UKMO GM and NCEP GM (poor mans

Figure 3a: Reduction of Error Variance related to Climate
Maximum Temperature (4 Stations)
Average over days +0...+8, May - July 1997

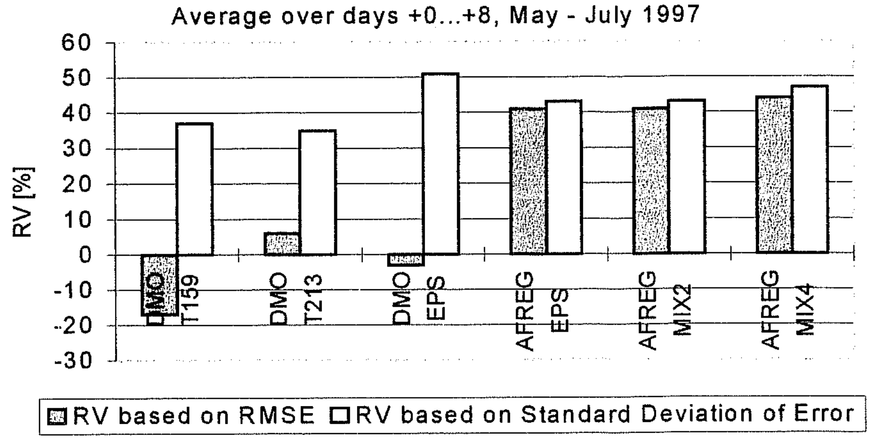
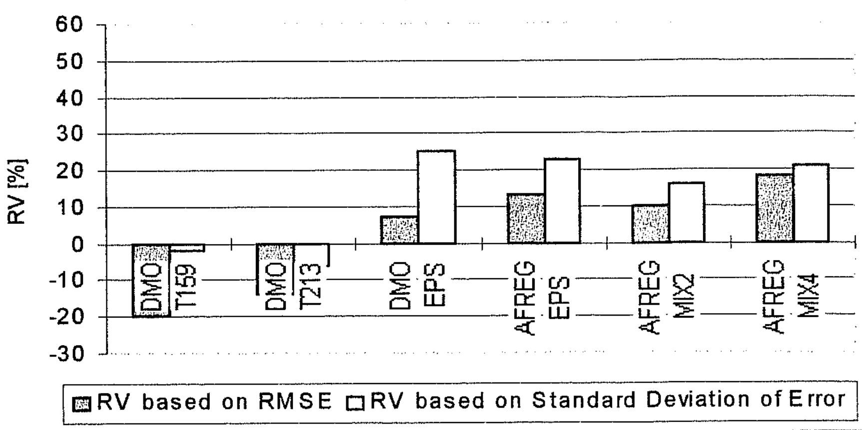


Figure 3b: as in Figure 3a
Wind Speed 12 UTC



4-model-ensemble). Notice, the results shown are very stratifying because they represent the average of RV over forecast days 0 to 8 and in so far, they are of general meaning. The dark bars show the real skill (RV based on RSME) while the light bars stand for the hypothetical skill based on the standard deviation of the error (error free of bias $SD = \sqrt{(rmse^2 - bias^2)}$).

As regards the maximum temperature (Figure 3a), the DMO is more or less worthless. But in the case of a forecast free of bias the DMO of EPS would be the best guidance

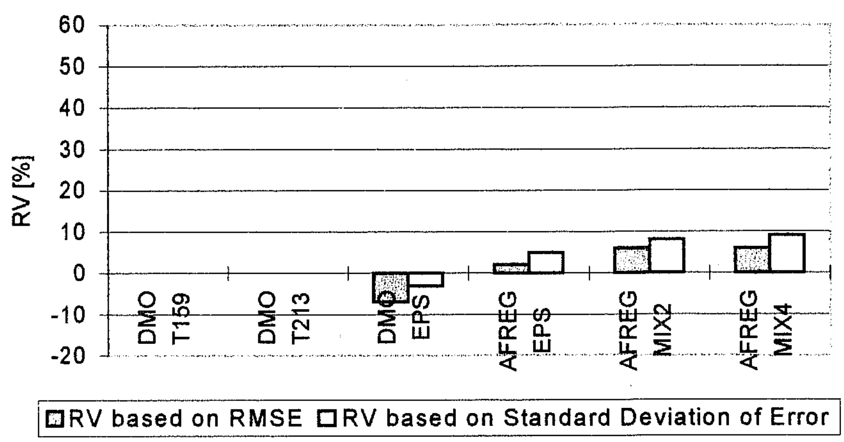


Figure 4a: As in Figure 3a
PoP > 5 mm/day

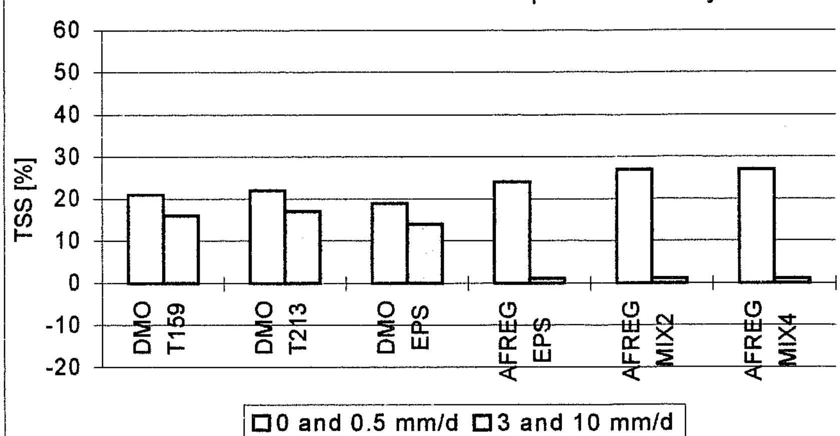


Figure 4b: As in Figure 3a
True Skill Statistics TSS for Precipitation mm/day

which means that the DMO of EPS suffers from a large (negative) bias. Nearly the same characteristics are shown according to wind speed (Figure 3b), but the general level of skill is much lower than for maximum temperature. The AFREG scheme (shown on the right) works nearly without bias and, the differences in skill are very small among EPS, MIX2 and MIX4.

Most interesting results are the ones concerning PoP (Figure 4a) and precipitation amount (the latter expressed in terms TSS in Figure 4b) because there were some early indications that the EPS could be successful in providing useful signals for the more rare events of large precipitation. As shown in Figures 4a and 4b, this no longer holds true. The results are disappointing. The maximum RV that can be achived is less than 10%. Looking at Figure 4b containing TSS, one can state that the DMO of EPS is not as good as the single T213 and T159 forecasts for both small and large precipitation amount. The AFREG scheme applied to EPS mean, MIX2 and MIX4 works with

considerable success in predicting small amount of precipitation but there is no real skill according to the forecast of large amount of precipitation.
Therefore, presently we have to state that there is no strong indication for prefering the EPS. On average the stratifying concept of mixing different high sophisticated models by means of the AFREG scheme (ore comparable ones) is more successful.
Notice, this may change considerably when the bias of the EPS near surface weather parameters will be minimized (Kalman filtering). In order to achive that the DWD is now developing a Kalman filtering of the relevant near surface weather elements derived from EPS.

The EPS (single-model ensemble) versus the poor mens 4-model ensemble approach
As known the philosophy of the ECMWF's single-model EPS is based on the assumption that the forecast uncertainty is mainly caused by the uncertainty in the initial analysis. This implies that the prediction model is perfect. Unfurtunately, this is not the case. In reality the different operational models known always produce different non-perfect forecasts.
This natural disagreement of models can (should) also be considered as a kind of spread measure comparable to the spread of the single-model EPS. As mentioned in the introduction and, in contrast to the single-model EPS the multi-model EP approach includes both different initial states as well as different model formulations.
Based on this consideration on one hand and the remarkable success of the DWD's AFREG MIX approach on the other hand, an experiment was started in 1996 which aims at testing the 4-model-mixture using the operational models GM(T106), ECMWF(T213), UKMO GM and NCEP GM. So, indeed there are 4 different models involved starting from 4 different analyses.
At first the investigation concentrated on the skill of forecasts of near surface local weather parameters derived from the 4 models by means of the AFREG interpretation scheme. Preliminary results indicate that (1) the skill of MIX4 exceeds that of MIX2 and EPS Mean (see Figures 1, 3 and 4) (see also: K.Balzer and P.Emmrich: Proceedings EPS Expert Meeting 1996: >Some Remarks on the Assessment of EPS<).
The direct comparison of MIX4 and EPS concerning the spread-skill-relation on the basis of a simple field verification of the 500 hPa topography in the area of Europe (winter period December 96 – March 97) shows (2) significant higher correlations for the EPS than for MIX4. Nevertheless, the maximum correlation is still less than 50% and, this is unacceptable for practical prognostic usage. In general it is true that - considering a perfect model - a small spread is connected to a small error. But in reality it can also happen that the error is small in the case of a large spread when the realization is randomly close to the control or the ensemble mean.
The project will be continued in order to become more familiar with the fitness of the poor mans MIX4 with special emphasis on the comparison with the EPS and, most important, to examine the general fitness of the two systems to produce a practical useful estimate of forecasting forecast skill.

Kalman-Filtering of the DMO of EPS

The Kalman filtering system now under development is of the type:

$$KAL(E1) = Const(t) + DMO\_Factor(t) \times DMO(E1)$$

with continously adapted coefficients.

Advantages are (1) easy to handle, (2) quick reaction on model changes

Disadvantages are (1) no a priori safety indication of the adaptation time (velocity) of the filter, (2) filtering parameters depend on weather situation and season.

At first we aimed at optimizing the velocity of adaptation empirically by means of a historical data set of DMO and the corresponding observations.

For each combination of 5 × 5 different velocities of adaptation (from no adaptation up to very fast adaptation) the filter were quasi-operational calculated (5 for both the filtering constants and the coefficients).

On the basis of the a posteriori verification results the filter showing the optimal velocity of adaptation was then selected for operational application. This was done individually for each element, station and forecast distance.

The results allow in a first approximation a statement on the performance of the filter and, so far as they are available up to now, they are encouraging at least for elements like temperature, wind and cloud coverage.