PROGRESS IN MEDIUM-RANGE FORECASTING BY MEANS OF ENSEMBLE FORECASTING

P.Emmrich, K.Balzer
Deutscher Wetterdienst
Offenbach, Germany

Summary: This is a brief discussion of the first german verification results concerning the first ensemble prediction experiment that has been running at the ECMWF since December 1992. The objects of evaluation were

- one plume diagram of temperature forecast 850 hPa and
- the cluster mean pattern .

The verification deals with two essential questions:

a) Is ensemble prediction a successful approach to provide an a priori forecast of the forecast skill ?

b) What is the benefit of ensemble prediction compared with the operational T213 forecast?

1. INTRODUCTION

From the very first beginning, the medium-range forecast suffered from the unwanted large day to day variability of skill. This is also the main reason why the internal consistency of one single prediction model is no skillful predictor for judging its confidence.

Every forecast aims at the minimization of knowledge uncertainty. The information provided by a forecast is more valuable the better its confidence is determined. The ensemble prediction strategy (EPS) is the strategy that provides an a priori information about the forecast's confidence. The EPS is the only appropriate strategy for forecasting forecast skill, deriving probabilistic forecast information and providing alternative large-scale circulation scenarios; all needed in medium-range forecasting. In addition, the atmosphere behaves neither deterministic nor pure chaotic but in a stochastic manner. It is the ensemble prediction that reflects this behaviour best.

The EPS experiment is given a lot of attention because ensemble prediction will be the future forecasting strategy in medium-range.

2. DISCUSSION OF EPS VERIFICATION RESULTS

The objects of evaluation are

- the plume diagram at 54N/12E (Rostock) and
- the following cluster mean pattern for the area of Europe
  - geopotential height 1000 hPa,
  - temperature forecast 850 hPa
  - temperature and geopotential height forecast 500 hPa.

2.1 Verification of the plume diagram

For each run and each forecast day (days 0 to 10) the following variables were determined:

obs:	temperature 850 hPa at radiosonde station Greifswald
max:	maximum temperature ensemble forecast
min:	minimum temperature ensemble forecast
EMode:	most probable estimate of ensemble forecast
T63:	control forecast
T213:	routine forecast
Sample size:	66 ensemble forecast runs from December 12, 1992 to August 18, 1993 (726 single day forecasts)

Error measures used:

$$RMSE = \sqrt{\left(\frac{1}{11}\left(\sum_{i=1}^{11} (obs_i - EMode_i)^2\right)\right)}$$

$$RMSE(d) = \sqrt{\left(\frac{1}{11}\left(\sum_{i=1}^{11} (max_i - min_i)^2\right)\right)}$$

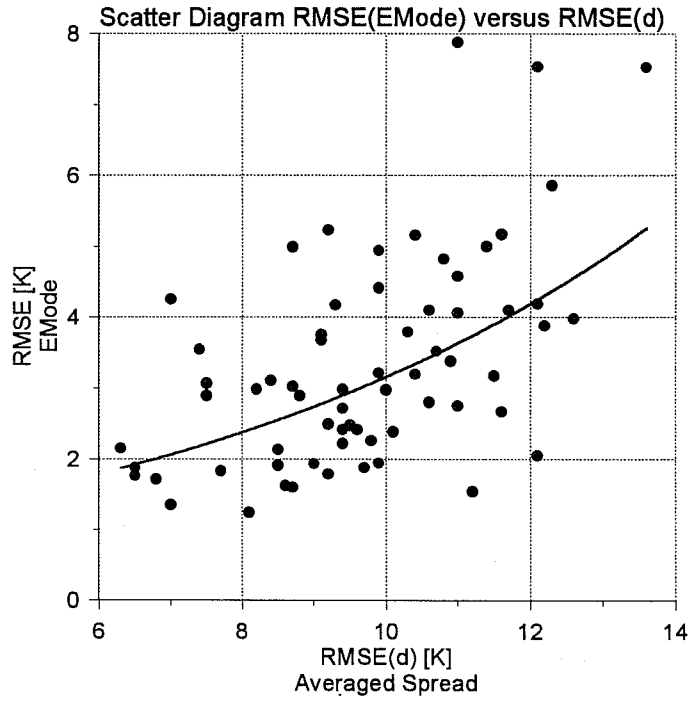
i = 1 forecast day 0

i = 11 forecast day 10

RMSE(d) is a measure of the "averaged ensemble forecast spread" per run. It is assumed that this measure is mainly a function of the starting conditions.

We only concentrate on the verification of the EMode because it may be - besides the ensemble spread - the most important parameter.

Figure 1: Plume Diagram ROSTOCK
Temperature Forecast 850 hPa



The first question to be answered is:

Is the ensemble spread (max-min) an useful a priori information of the skill of the EMode (expressed in terms of RMSE)**?** The hypothesis may be: The larger the ensemble forecast spread the more unreliable the EMode forecast.

Figure 1 shows for each forecast day the RMSE(EMode) as function of the ensemble forecast spread. The relationship shown is rather weak. Nevertheless, it is statistically significant (correlation coefficient CC = 0.55, sample size N = 726) and it is best expressed by

$$RMSE_{EMode} = EXP(0.14*RMSE(d)) + 0.77.$$

But in fact, only 30% of the error variance are explained by the ensemble spread. This relationship is too weak to make use of it prognostically.

Furthermore, there is a weak indication that large forecast errors are more unlikely in the case of a small spread. Unfortunately, this is typical for short-range forecasts up to days 2-4 only. It seems that broad plumes are not a strong signal for bad EMode forecasts. On average the spread increases by approximately 1.5K per forecast day. This growing rate is in most cases remarkably consistent. Therefore, the day to day variability of RMSE(d) (as function of the starting conditions) is to be considered as too weak. The prognostic information provided by this "quasi consistent" spread is too less. Consequently, **at this time the spread of the ensemble forecast is still not a properly working parameter.**

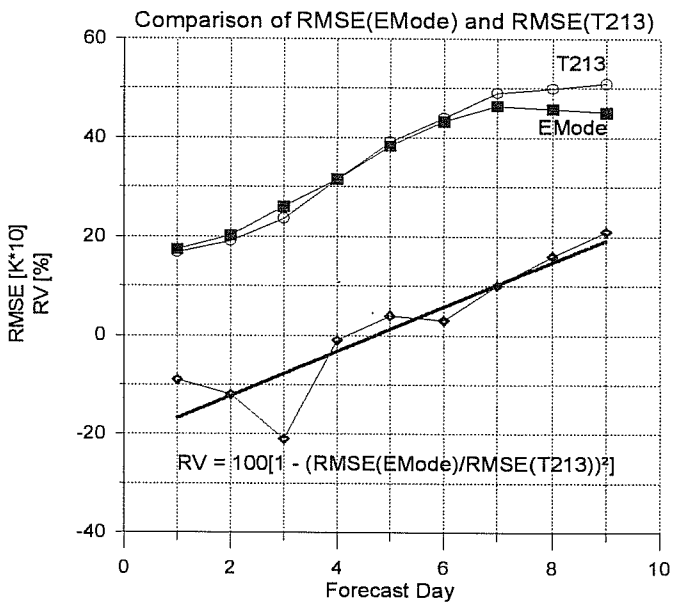
The second question of importance is:

What are the benefits of the EMode forecast compared with the operational T213 forecast?

The following table contains the intercomparison of the success of the T213, T63 and EMode forecast in terms of RMSE[K] and RV[%] (= reduction of error variance related to T213). The sample size of the data set verified is again N=726 single day forecasts.

	EMode [K]	T213 [K]	T63 [K]
averaged run	3.61	3.8	3.88
RV [%]	10		-4

Figure 2: Plume Diagram ROSTOCK
Temperature Forecast 850 hPa



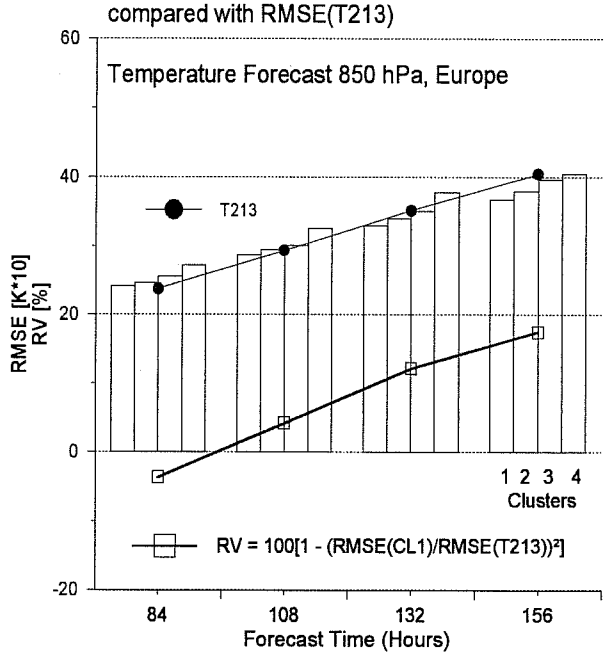
In general the RMSE differences among these three forecast strategies are rather small. It is first of importance that the EMode forecast (mean over all forecast days of a run) is on average better than the corresponding T213 forecast (RV = 10%). Secondly, there is a surprisingly small (statistically nonsignificant) difference between T213 and T63. This may be an indication that the T63 model version is more skillful in medium-range forecasting than initially expected when the experiment was started. (The latest results including all ensemble forecast runs till December 6, 1993 (N=103 runs) suggest a much higher superiority of both EMode (RV = 22%) and T63(RV = +10!)).

Figure 2 shows the advantage and disadvantage of the EMode forecast compared with the operational T213 forecast distributed over forecast days 1 to 9 (running average over 3 days including day 0). There is no doubt, in short-range the deterministic T213 strategy is significantly better than the probabilistic strategy. But from forecast day 4 upwards the EMode forecast becomes better than the T213 forecast with a superiority of about 20% at forecast day 9!

2.2 Verification of the cluster mean pattern

The cluster mean fields of H1000, H500, T500 and T850 have been verified. Both the RMSE and the tendency correlation coefficient have been determined. The area of concern covers Europe. The data set contains all ensemble forecasts available from March 21 to December 29, 1993 (N =109). As per definitionem, cluster 1 represents the one that contains the most members, it is to consider as the mode of the distribution obtained from the cluster analysis. Therefore, the corresponding hypotheses is:

Figure 3a: RMSE of Cluster Mean Pattern compared with RMSE(T213)

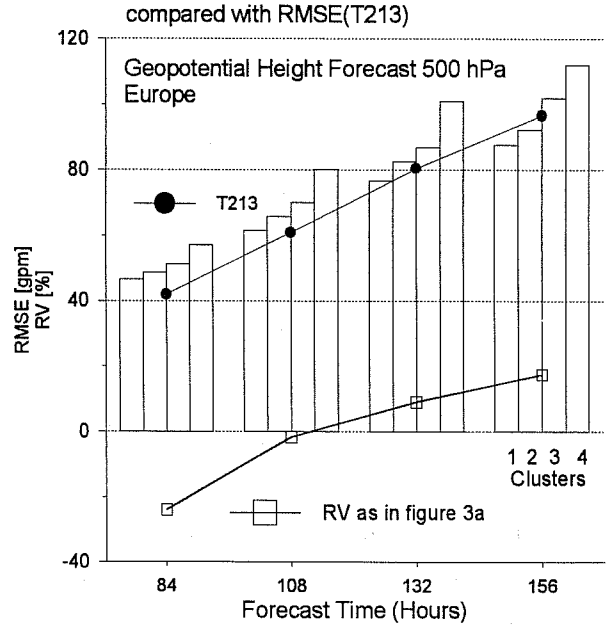


Cluster 1 mean pattern represents the most probable weather regime. In other words: **It is expected that the forecast error grows with the order of the clusters.**

As it is seen from figure 3a,b this hypothesis ist true without exception. The most important signal is that there is no advantage of the T213 forecast compared with Cluster 1 (except for geopotential height forecast 500 hPa at H+84). Cluster 1 achieves in relation to the T213 forecast a RV = 18% at H+156. This is in line with the result obtained from the verification of the plume diagram. Cluster 2 is partly also better than T213, at least for forecast times H+132 and H+156.

So, it is to state from the verification of the cluster mean pattern: Considering the forecast time range provided by the clustering the cluster 1

Figure 3b: RMSE of Cluster Mean Pattern compared with RMSE(T213)



mean pattern represents a more valuable forecast information than the operational T213 forecast does.

3. CONCLUSION

There are very similar and encouraging results for both the plume diagram verification and the cluster mean pattern verification. Judging these verification results in general one can conclude:
Up to forecast day 4 the deterministic T213 forecast strategy provides the better forecast. Beyond forecast day 4 there is a distinct advantage of the T63 ensemble forecast strategy. Therefore, it is no longer a need to follow the deterministic strategy beyond day 4. It is possible to improve the medium-range guidance by applying the ensemble forecast. This is what was realized at NMC Washington recently [1][2].

Not all relevant questions are answered up to now by the experiment. Therefore, the experiment has to be continued, at least in order

- to examine more extensively the ensemble spread as a function of the starting analysis
- to develope a more sufficient clustering scheme and
- to test the impact of models with higher performance.

References:

[1] Z.Toth and E.Kalnay, 1993: Ensemble Forecasting at NMC: The Generation of Perturbations. Bull. Amer. Meteor. Soc., submitted

[2] M.S.Tracton and E.Kalnay, 1993: Operational Ensemble Prediction at the National Meteorological Centre: Practical Aspects. Weather and Forecasting, Vol. 8, 379-398