Short range ensemble forecast

Edit Hágel

Hungarian Meteorological Service

Supervisor: Jean Nicolau

Toulouse, 2 October 2004 - 27 November 2004

1 Introduction

The ensemble technique is based on the fact that small errors in the initial condition of any numerical weather prediction model (or errors in the model itself) can cause big errors in the forecast. When making an ensemble forecast the model is integrated not only once (starting from the original initial condition), but forecasts are also made using little bit different (perturbed) initial conditions. This ensemble of the initial conditions consists of equally likely analyses of the atmospheric initial state and, in an ideal case, encompasses the unknown 'true' state of the atmosphere. This technique is capable to predict rare or extreme events and has the advantage of predicting also the probability of future weather events or conditions. Despite its success, at the moment the ensemble method is mainly used for medium range forecasting and on global scales, but nowadays the emphasis is more and more moving towards the short ranges and smaller scales. However, methods used in the medium range can not be directly applied to short range forecasting. Research has already been done in this field and there are some operational or quasi-oparational short range ensemble systems (e.g. at Météo-France, at NCEP, the COSMO-LEPS, or the SRNWP-PEPS project at DWD). We also wish to develop a short range ensemble system for the Central European area, with the main goal being the better understanding and prediction of local extreme events like heavy precipitation, wind storms, big temperature-anomalies and also to have a high resolution probabilistic forecast for 2 meter temperature, 10 meter wind and precipitation in the 12-48 h time range.

For making an ensemble forecast lots of methods can be used (e.g. multi-model, multi-analysis, perturbation of observations, singular vector method, breeding etc.). It is not known yet (especially at mesoscale) which method would provide the best forecasts. It was decided to start our experiments with the downscaling of the global (ARPEGE based) ensemble. This work can be divided into two parts:

- Direct downscaling of the ARPEGE/PEACE[1] members
- Investigation of the impact of target domain and target time window and downscaling the ARPEGE ensemble members

From previous studies performed at the Hungarian Meteorological Service (HMS), we found that by simply downscaling the PEACE members the spread obtained is not big enough in the area of our interest (Central Europe). This seems reasonable if we consider that the PEACE system was calibrated in order to get enough spread over Western Europe between 24 and 72 h steps, for wind speed, 500 hPa geopotential and mean sea level pressure. The aim of the PEACE system is to detect strong storms. This raises some questions:

- Are the PEACE provided initial and boundary conditions convenient for the local EPS run, for a Central European application?
- What is the impact of different target domains and target times?

During this two month stay - as a continuation of the work started at HMS - we tried to investigate and better understand the effect of different target domains and target times. In our experiments an ARPEGE ensemble system was used, based on a previous version of the PEACE system:

- For the generation of perturbations the singular vector method was used

1 Prévision d'Ensemble A Courte Echéance; a short-range ensemble system operational at Météo-France, with 10+1 members, based on the model ARPEGE

- The singular vector computations were performed on T63 resolution
- 10 ensemble members were computed + the control run
- The integrations were performed on T199 resolution
- The forecast length was 54 h (we use 54 h because of the verification of precipitation, since the daily precipitation amount is observed at 06 UTC, so a 48 h forecast started from the 00 UTC analysis would only cover one 24 h period like this, while a 54 h forecast covers both)

The main difference is that the target domain and the target time was not fixed. For the target time 12 h and 24 h were used, and different target domains were defined.

2 Experiments

In previous studies performed at the Hungarian Meteorological Service, we tried to investigate the effect of different target domains. Four domains were defined:

- Domain 1: Atlantic Ocean and Western Europe (70N/260W/30S/20E; the same as used earlier in PEACE)
- Domain 2: Europe and some of the Atlantic (70N/330W/30S/35E)
- Domain 3: covering nearly whole Europe (60N/1W/30S/35E)
- Domain 4: slightly bigger than Hungary (49N/15W/45S/24E)

With the use of these domains case studies were performed. We concluded, that the use of domain 2 provides better results compared with domain 1, and also seems to be more rational than the use of domain 4. Domain 2 and domain 3 provided quite similar results in most of the cases, so next to domain 1 we chose domain 2 for a 10 day experiment (the target time in these experiments was 12 h).

The chosen period for the ten day experiment was 10-19 July 2004. The period was chosen randomly, the meteorological situation was not particularly interesting. At the beginning of the period there was some frontal activity at the area of interest, but in the second half of the period the weather situation was determined by an anticyclone over Central Europe.

The results of the 10 day experiment showed that by using domain 2 for singular vector computation we can obtain a bigger spread, and better scores, also.

2.1 Experiments with different target times

During this stay in Toulouse our first aim was to repeat these case studies and the 10 day experiment, but only with the use of domain 2, and with 24 h as target time instead of using 12 h.

2.1.1 Case studies

It is expected that in different meteorological situations the use of different target domains would provide the better results and a compromise should be found to choose the best domain. So far three different meteorological situations were examined (and a forth, with a situation that a cyclone was coming from South-East, has already started). One of them was a convective event in 2002 (18 July 2002). In this situation large quantity of precipitation (40-70 mm during 24 h) was measured at some places along the river Danube and all the models (ALADIN, ARPEGE, ECMWF) failed to forecast the event. The second case (22 June 2001) was a situation with a fast moving cold front coming from the west. This time the models overestimated the precipitation. The third situation (22 February 2004) was one with a quite big temperature overestimation. This error in the forecast of temperature caused a big problem: the models predicted rain, but in reality it was sleet.

Every time the ARPEGE ensemble runs were performed with singular vector target domain 2, and target time 24 h. The average standard deviation over Hungary was computed (for 850 hPa temperature, 10 meter wind speed, mean sea level pressure and 500 hPa geopotential) and we also looked at different meteorological parameters. The results were compared with those obtained from the previous experiments (performed at HMS).

Altogether the following runs were performed:

- Convective case
  - ✔ Integration started from 00 UTC, 17 July 2002, from 12 UTC, 17 July 2002, and from 00 UTC, 18 July 2002
    - target domain 1, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 24 h

- Situation with a fast moving cold front
  - ✔ Integration started from 00 UTC, 21 June 2001, from 12 UTC, 21 June 2001 and 00 UTC, 22 June 2001
    - target domain 1, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 24 h

- Temperature overestimation
  - ✔ Integration started from 00 UTC, 21 February 2004 and from 00 UTC, 22 February 2004
    - target domain 1, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 12 h (performed earlier at HMS)
    - target domain 2, target time 24 h

2.1.2 Results of case studies - Standard deviation

In nearly every situation it was found that with the use of singular vector target domain 1 and target time 12 h the average standard deviation was quite small in the beginning of the forecast and it increased quite slowly. Around the end of the forecast range it usually reached the values obtained by the use of domain 2, but we do not want to concentrate only on the last few hours of the forecast. Instead we would like to find an optimal target domain for the singular vector computation which guarantees sufficient spread in the 12-48 h time range.

When target domain 2 was used the (average) standard deviation was bigger. The second case (fast moving cold front) was the only one when standard deviations were nearly the same with the use of domain 1 and 2. The reason of this might be that in this case the examined phenomenon was a large scale one.

The use of 24 h as target time also (on average) increased the standard deviation (fig. 1, fig. 2, fig. 3).

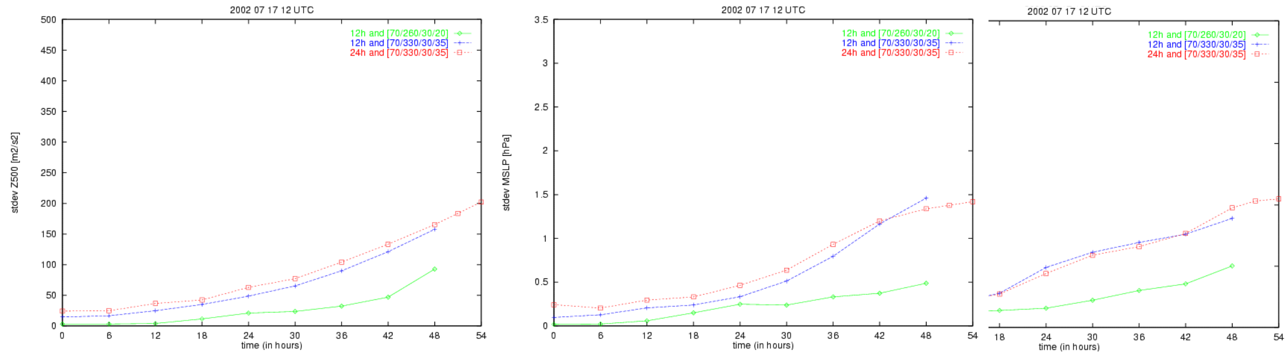


Figure 1. Average standard deviation over Hungary for the 17 July 2002 12 UTC run, for Z500, T850 (top row), MSLP and V10 (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 1 and target time 12h, the blue one is for target domain 2 and target time 12h, and the red is for target domain 2 and target time 24h.

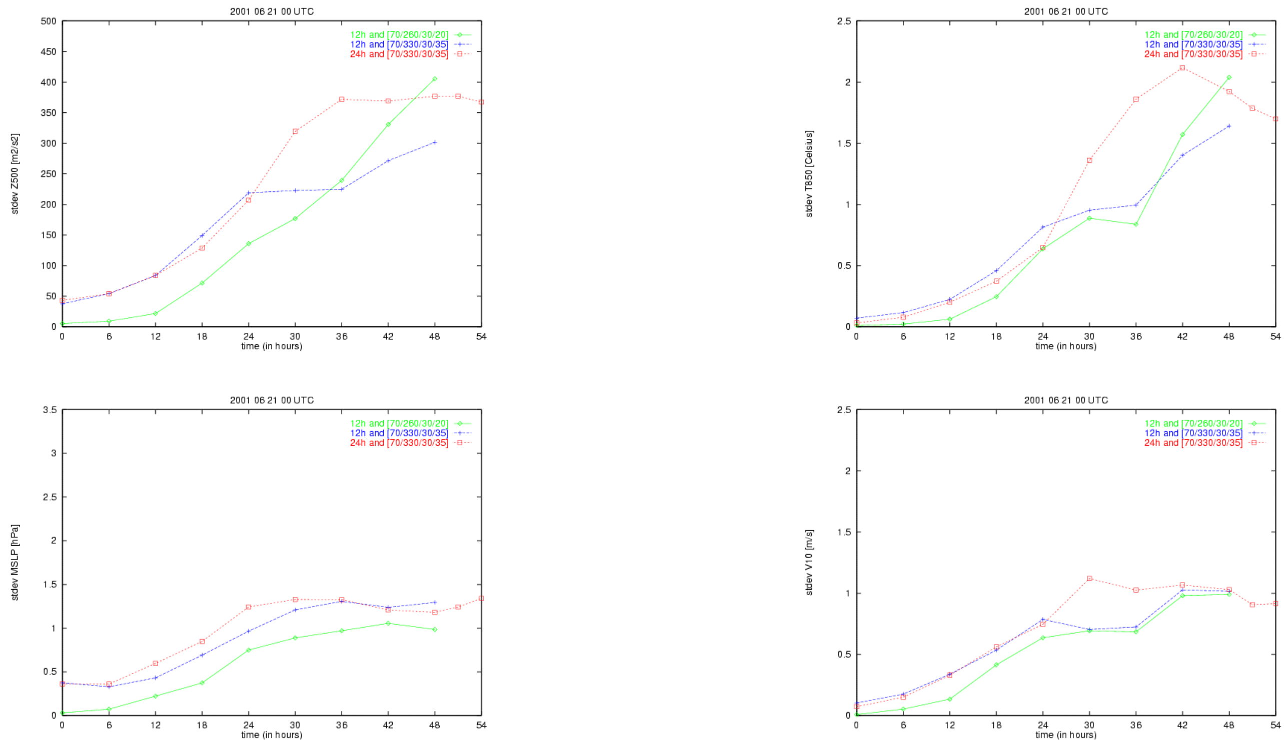
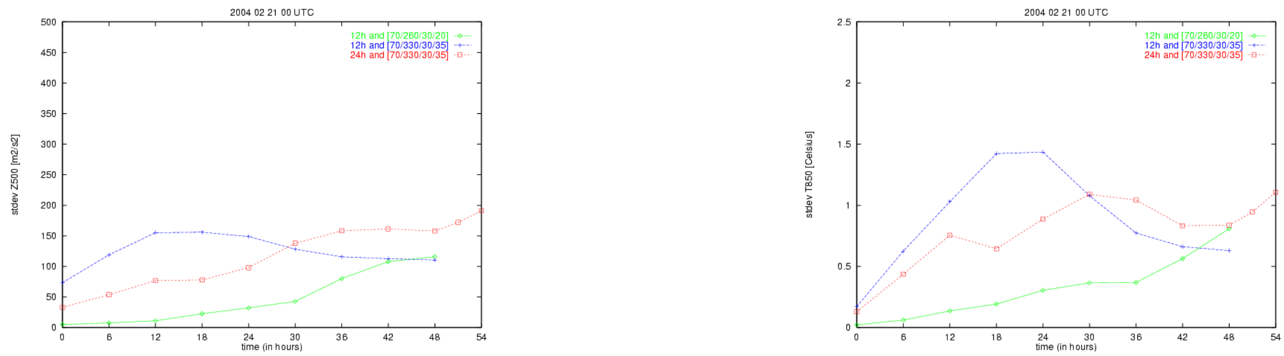


Figure 2. Average standard deviation over Hungary for the 21 June 2001 00 UTC run, for Z500, T850 (top row), MSLP and V10 (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 1 and target time 12h, the blue one is for target domain 2 and target time 12h, and the red is for target domain 2 and target time 24h.



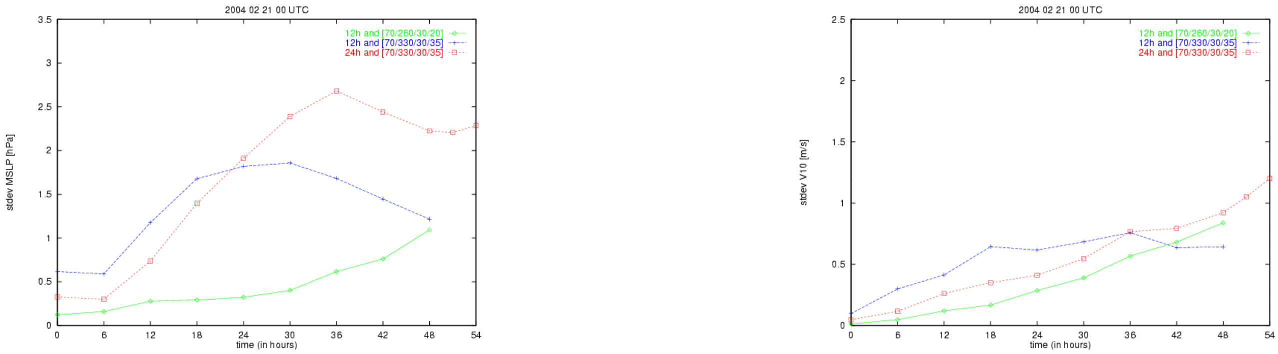


Figure 3. Average standard deviation over Hungary for the 21 February 2004 00 UTC run, for Z500, T850 (top row), MSLP and V10 (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 1 and target time 12 h, the blue one is for target domain 2 and target time 12 h, and the red is for target domain 2 and target time 24 h.

2.1.3 Results of case studies - Meteorological parameters

Not only the standard deviation was examined but also we looked at different meteorological parameters each time. In the first case (convective case) we got nearly no precipitation at all when we used target domain 1 and target time 12 h in the global singular vector computation. Using target domain 2 and target time 12 h gave slightly better results. Some members of the ensemble forecast started from 12 UTC, 17 July 2002 indicated bigger amount of precipitation, but the location and the quantity was not perfect. By changing target time from 12 h to 24 h (and using target domain 2), the best results were obtained from the integration started from 00 UTC, 17 July 2002 (fig. 4). Some members again predicted significant precipitation near the area where it occured in reality. The results obtained from the two other runs were worse. No significant amount of precipitation was forecasted along the river Danube (fig. 5, fig. 6).

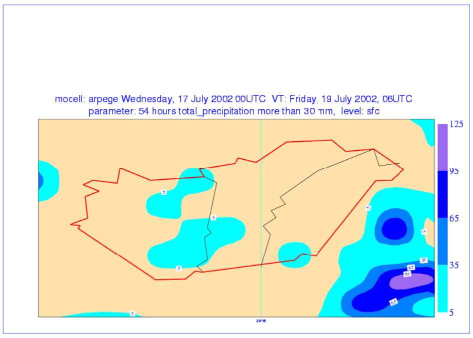


Figure 4. Probability of total precipitation more than 30 mm /54 h, forecast started from 00 UTC, 17 July 2002.

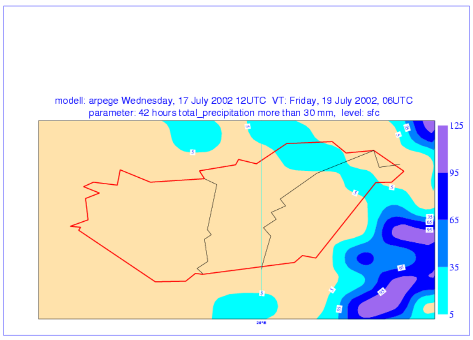


Figure 5. Probability of total precipitation more than 30 mm /42 h, forecast started from 12 UTC, 17 July 2002.

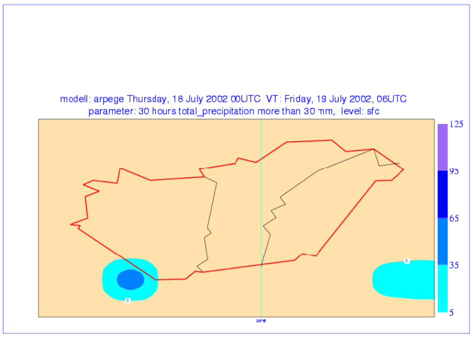


Figure 6. Probability of total precipitation more than 30 mm /30 h, forecast started from 00 UTC, 18 July 2002.

The second case was the only one where standard deviations were nearly the same with the use of domain 1 and 2, and also the predicted amount of precipitation was quite similar. The results of the forecasts showed that some members predicted too big amount of precipitation in the eastern part of Hungary along the river Tisza (which was also the problem with the operational forecast for that day, since the front in the model was not moving so fast than in reality), but there were also a significant number of members predicting much less amount of precipitation (fig. 7, fig. 8. fig. 9).

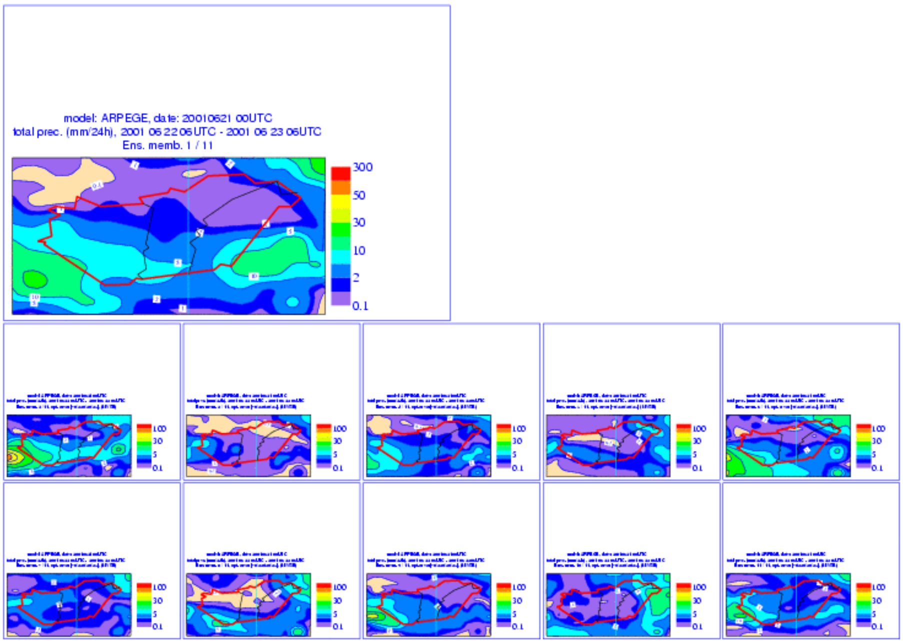


Figure7. Total precipitation from 22 June 2001, 06 UTC to 23 June 2001, 06 UTC for Hungary. Control forecast is in the top left panel. Forecast started from 00 UTC, 21 June 2001.

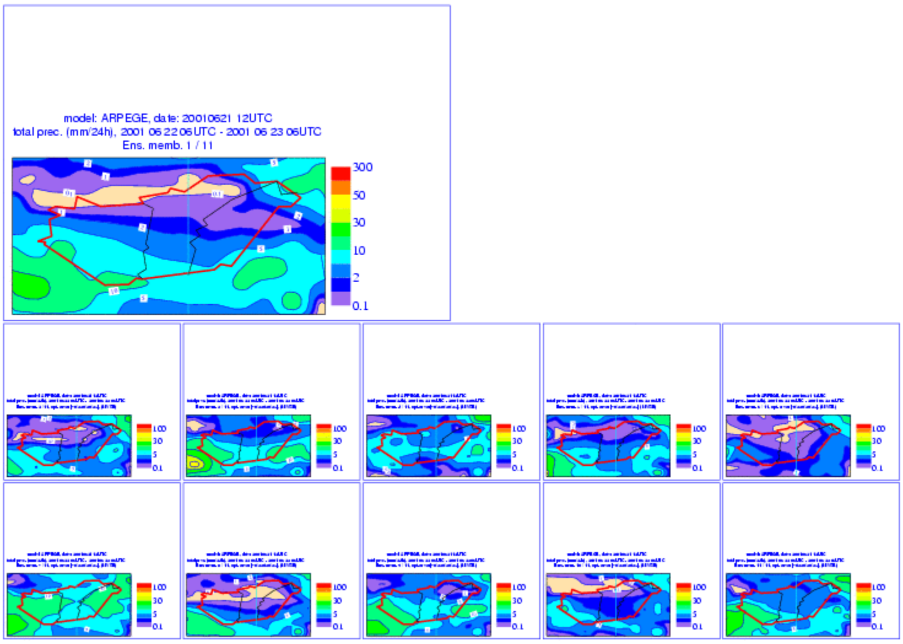


Figure 8. Total precipitation from 22 June 2001, 06 UTC to 23 June 2001, 06 UTC for Hungary. Control forecast is in the top left panel. Forecast started from 12 UTC, 21 June 2001.

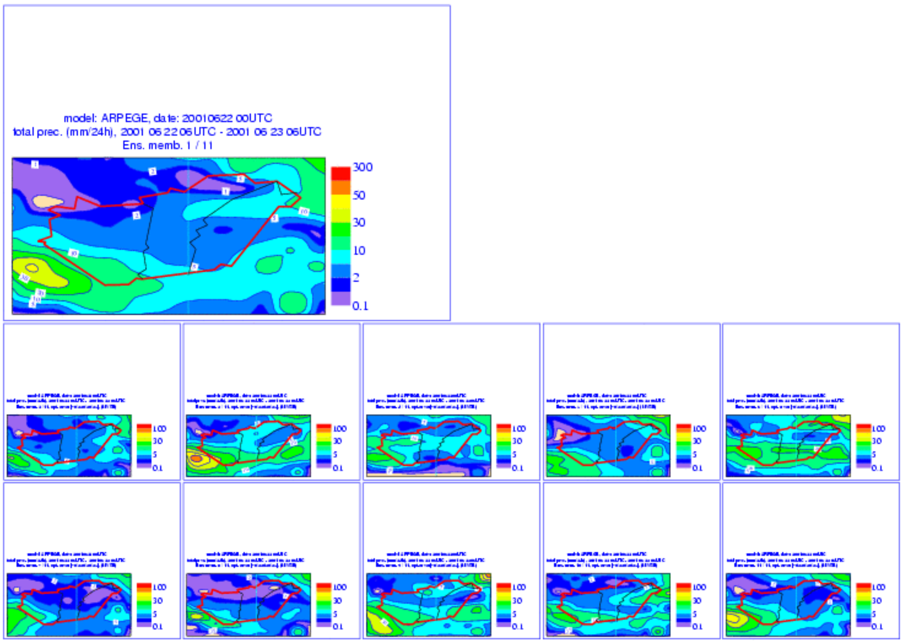


Figure 9. Total precipitation from 22 June 2001, 06 UTC to 23 June 2001, 06 UTC for Hungary. Control forecast is in the top left panel. Forecast started from 00 UTC, 22 June 2001.

The result obtained in the third case (temperature overestimation) was not so good. In reality the temperature was around or below zero celsius all day, but the models predicted much more. A sufficient spread was obtained when domain 2 was used, but still the values for the temperature were very high. At least some of the members were colder than the control one, but they were not cold enough (fig. 10, fig. 11, fig. 12, fig. 13).

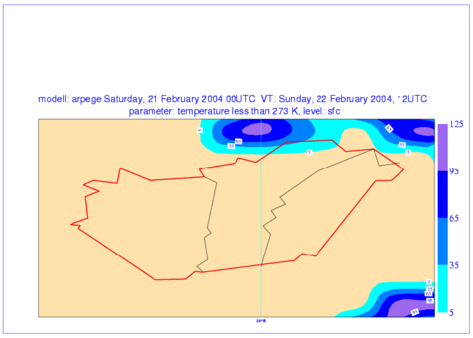


Figure 10. Probability of temperature less than 0 Celsius, forecast started from 00 UTC, 21 February 2004.

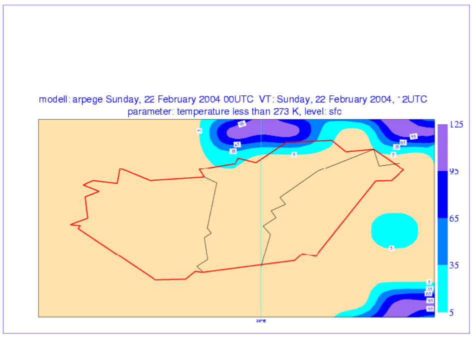


Figure 11. Probability of temperature less than 0 Celsius, forecast started from 00 UTC, 22 February 2004.

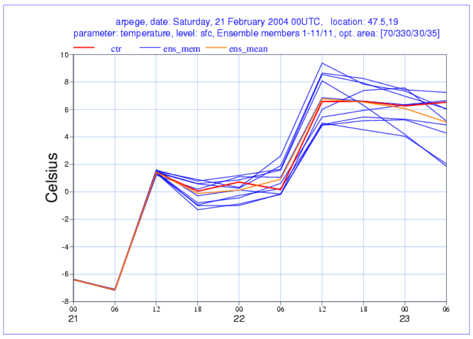


Figure 12. Plume diagram for Budapest for 2 meter temperature, forecast started from 00 UTC, 21 February 2004.

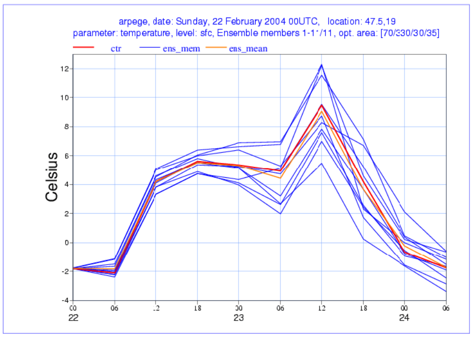


Figure 13. Plume diagram for Budapest for 2 meter temperature, forecast started from 00 UTC, 22 February 2004.

2.1.4 Ten day experiment

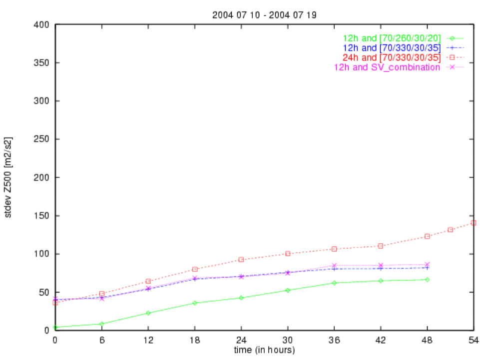
We repeated the 10 day experiment with the use of target domain 2 and with target time 24 h instead of 12 h.

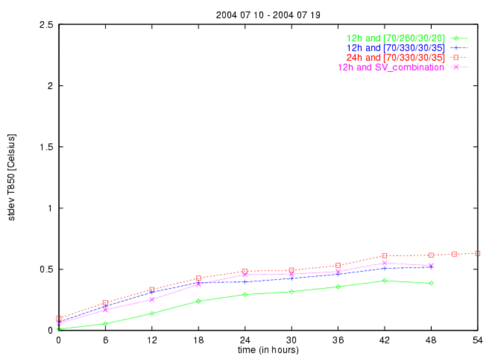
2.1.5 Ten day experiment results - Standard deviation

The results of the 10 day experiment show that on average, the use of configuration target domain 1 and target time 12 h would provide the smallest standard deviation values for all examined parameters (500 hPa geopotential, 850 hPa temperature, mean sea level pressure, 10 meter wind speed). This can be explained by the fact, that this domain is covering not only Western Europe but also the North Atlantic region and some part of the North American continent. The perturbations created usually have their maximum amplitude in the North Atlantic region and during a 54 hour forecast they not always have a significant effect on the forecast over the Central European area.

With the use of target domain 2 the standard deviation (on average) can be increased and further improvement can be obtained with the use of 24 h as target time. On average this configuration (target domain 2 and target time 24 h) provides the biggest values in terms of standard deviation computed over Hungary (fig. 14).

Looking at the forecasts one by one, instead of the ten day average, we can find that the spread was bigger in the first few days of the period in case of every target domain and target time. This is reasonable if we concider the fact that there was some frontal activity at that time in the area, and in the second half of the period an anticyclon was determining the weather situation.





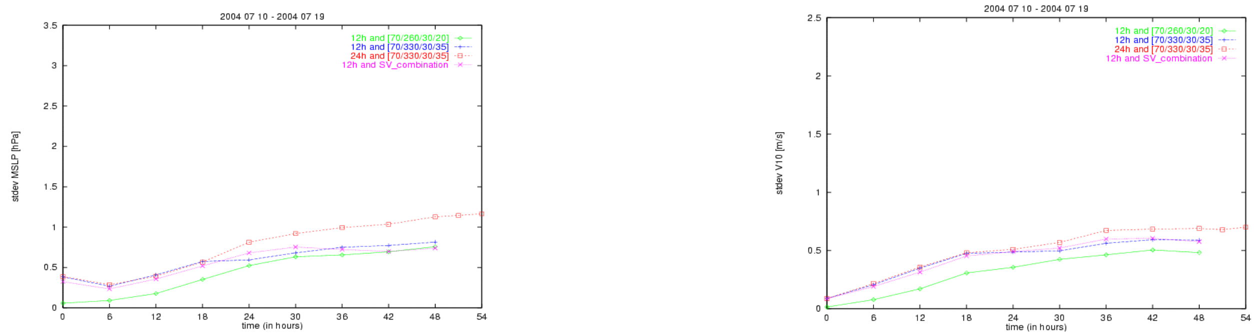


Figure 14. Average standard deviation over Hungary for the period 10 July 2004 - 19 July 2004, for Z500, T850 (top row), MSLP and V10 (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 1 and target time 12 h, the blue one is for target domain 2 and target time 12 h, red is for target domain 2 and target time 24 h, and magenta is for the experiment with two sets of singular vectors and target time 12 h.

2.1.6 Ten day experiment results - Scores

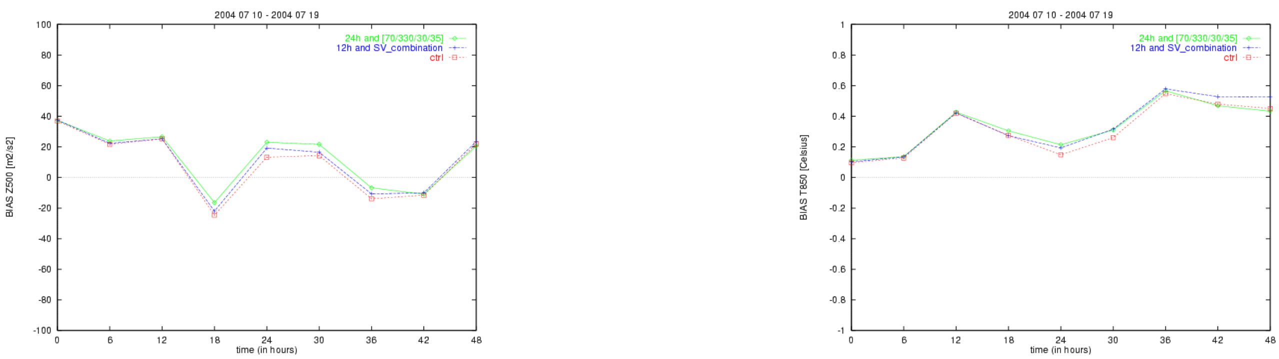
Root mean-square error (RMSE) and the systematic error (BIAS) were computed both for ensemble mean and for the control forecast.

Both RMSE and BIAS was computed for 500 hPa geopotential, 850 hPa temperature and mean sea level pressure for the ten day period over Hungary. Instead of observation, the analysis was used to compute these scores.

The BIAS of the ensemble mean and the control run on average seems to be quite similar, especially until +18 h. Between +18 h and +48 h the difference becomes bigger. In some time steps the control forecast performed better, in other cases the ensemble mean. If we look at the BIAS of the individual forecasts and not the ten day average, we can find cases when the BIAS of the ensemble mean and the control run is nearly identical (mainly in the second half of the period when an anticyclon was determining the weather situation) and also cases when one of them performed much better than the other (fig. 15).

For the 850 hPa temperature the control run and the ensemble mean performed nearly identically in terms of RMSE values. In the case of mean sea level pressure between +18 h and +48 h the control run was slightly better. For 500 hPa geopotential ensemble mean was better between +18 h and +30 h and the control run was better from +30 h. Looking at the forecasts one by one cases can be found when the ensemble mean outperformed the control run and vice versa. However, there are also cases (mainly in the second half of the period) when the RMSE of the control run and the ensemble mean was nearly equal (fig. 16).

The evaluation of the 10 day experiment will be continued (at HMS), by computing various kinds of probability scores such as ROC diagrams, Talagrand diagrams, Brier score and Brier skill score for several meteorological parameters and several thresholds.



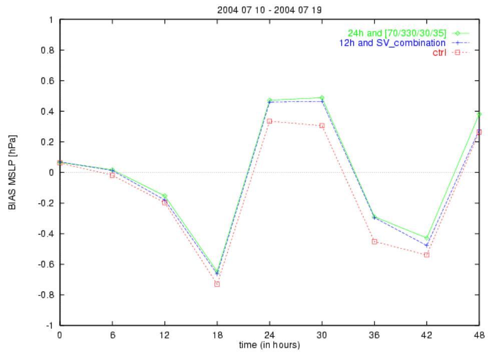


Figure 15. BIAS of the ensemble mean and the control forecast over Hungary for the period 10 July 2004 - 19 July 2004, for Z500, T850 (top row) and MSLP (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 2 and target time 24 h, the blue one is for the experiment with two sets of singular vectors and target time 12 h, and the red curve is for the control forecast.

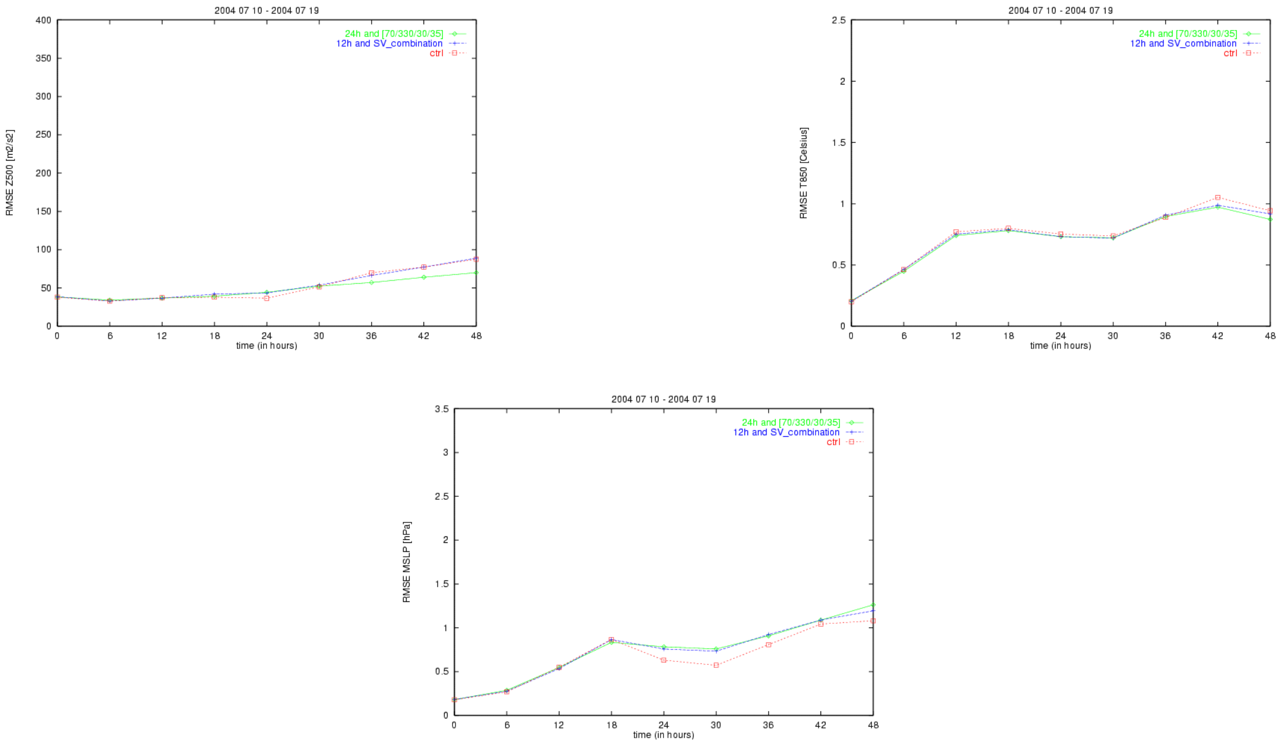


Figure 16. RMSE of the ensemble mean and the control forecast over Hungary for the period 10 July 2004 - 19 July 2004, for Z500, T850 (top row) and MSLP (bottom row). Computed from ARPEGE ensemble forecasts. The green curve is for target domain 2 and target time 24 h, the blue one is for the experiment with two sets of singular vectors and target time 12 h, and the red curve is for the control forecast.

2.2 Experiments with combining different sets of singular vectors

The results show that the spread in the ensemble system over Central Europe is - usually - not big enough with the use of the configuration target domain 1 and target time 12 h. Changing the target domain and also the target time seems to be a good way of increasing the spread over the area of our interest, but this method requires the rerun of the global ensemble system.

As our final goal is to develop an operational short range ensemble system, an alternative solution has to be found which does not require the local integration of a global ensemble system.

The most obvious solution would be to compute singular vectors in the framework of ALADIN. Preliminary works have already started at HMS, but up to now we have not been able to run this configuration. Until this problem is solved an other possible alternative solution can be the combination of different sets of singular vectors. The idea is the following:

Next to the singular vectors computed operationally every day at the PEACE system, a second set of singular vectors, using different target area and probably different target time, could be computed locally (at HMS). From this second set of singular vectors, perturbations can be built. The global ensemble run (PEACE) could provide the lateral boundary conditions for the limited area model (ALADIN) and the initial conditions could be produced by combining the initial conditions coming from PEACE and the perturbations generated from the second set of singular vectors (fig. 17).

Since this is a very complex system, first we concentrated only on a small part of it. We wanted to examine, whether the combination of two different sets of singular vectors can inprove the quality of the ensemble system in terms of spread. For the shake of simplicity as a start we did the combination in the framework of ARPEGE on the following way:

- Singular vectors with the use of target domain 1 and target time 12 h were computed
- Singular vectors with the use of a different target domain (one which is inside the LACE domain, 55N/2W/30S/40E) and target time 12 h were computed
- Independency check was performed to select singular vectors from the second set which are independent from the vectors in the first set (this was necessary, because we wanted to be sure, that the spread will not be reduced, the perturbations from the two sets of singular vectors will not weaken each other)
- After checking the independency, perturbed initial conditions were built from the vectors of the first set and the selected vectors of the second set
- Integration of the global ensemble system was performed for the ten day period (10-19 July, 2004)

2.2.1 Independency check

We performed the independency check the following way: scalar products were computed between the vectors of the two sets (16 vectors in each set). If the vectors are independent, their scalar product is zero. Of course we can not expect to have values exactly equal to zero, therefore we had to set a threshold; if the scalar product is below this value we consider the vectors to be independent. First we chose this threshold to be 0.1, but we found that there were cases when only one or two singular vector was selected from the second set with the use of this threshold. With a threshold 0.2 the situation was better (fig. 18).

So finally from the second set we used only the vectors which had scalar product less than 0.2 with all the vectors of the first set; from the first set all of the vectors were used.

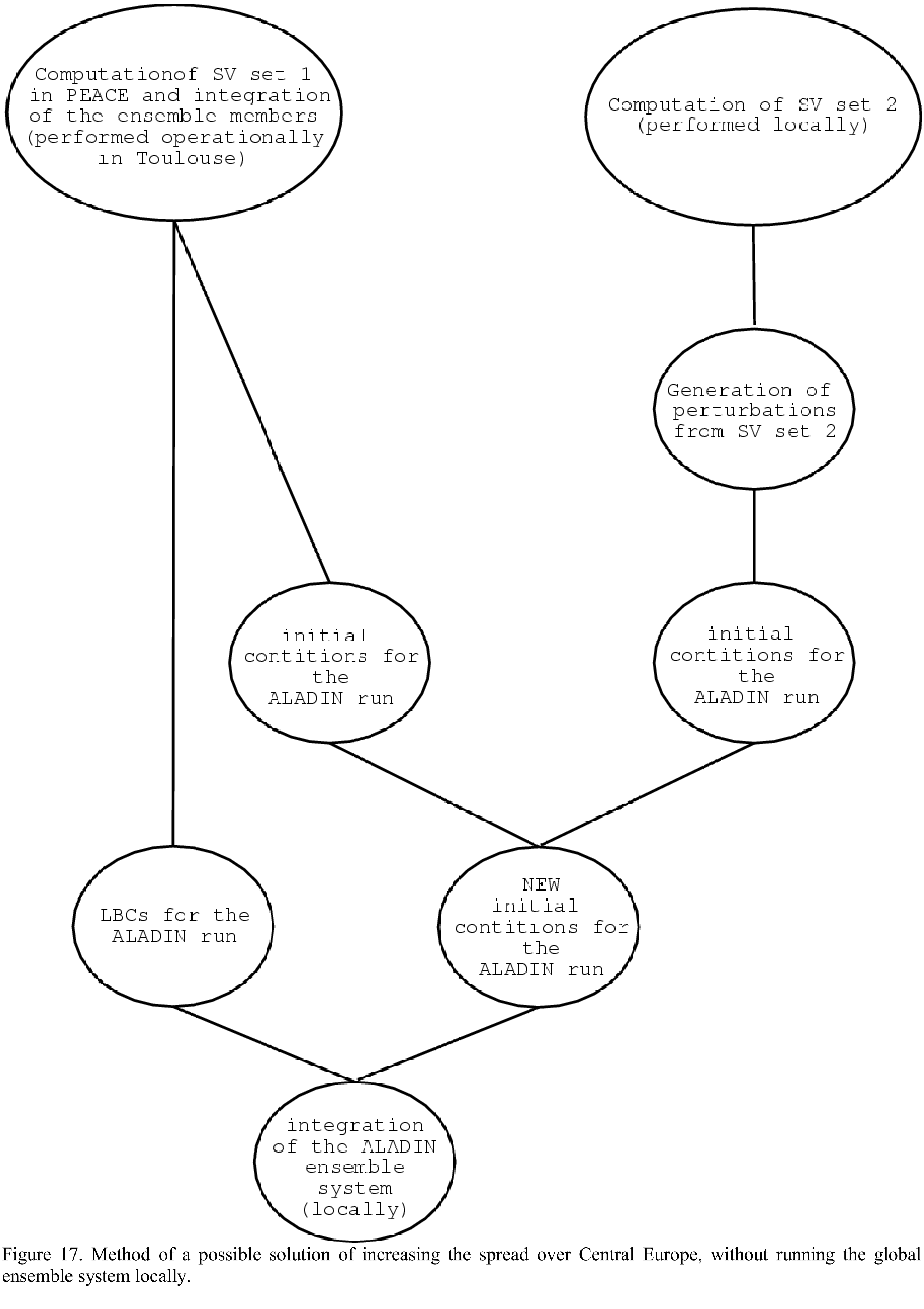


Figure 17. Method of a possible solution of increasing the spread over Central Europe, without running the global ensemble system locally.

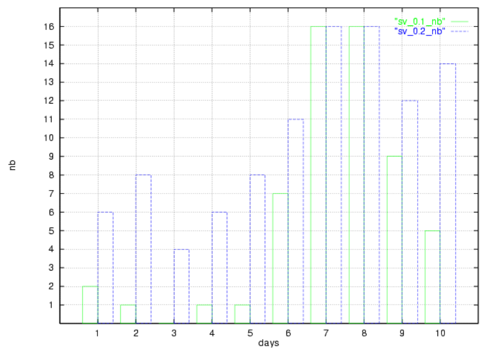


Figure 18. Number of singular vectors selected from the second set for each day of the ten day period. The green curve is representing the case, when vectors with scalar product less then 0.1 were selected, the blue curve is for the case when vectors with scalar product less then 0.2 were selected.

2.2.2 Ten day experiment results - Standard deviation

The results of this experiment show that by combining the two sets of singular vectors, the average standard deviation over Hungary is similar to results of the experiment using target domain 2 and target time 12 h. A clear improvement can be found compared with the spread obtained by the use of target domain 1 and target time 12 h, but still, the highest standard deviation values (over Hungary) are provided by the use of target domain 2 and target time 24 h (fig. 14). An experiment has started to repeat this one, but using 24 h as target time during the computation of the second set of vectors.

2.2.3 Ten day experiment results - Scores

In this case the same conclusions can be drawn than for the experiment with 24 h as optimization time, since the scores of the two experiment were very similar on average. Also it is true, that if we look the forecasts one by one and not the ten day average, bigger differences can be found between the performance of the ensemble mean in the two experiment, and also between the performance of the ensemble mean and the control run (fig. 15, fig. 16).

In this case also, the evaluation of the 10 day experiment will be continued (at HMS), by computing various kinds of probability scores such as ROC diagrams, Talagrand diagrams, Brier score and Brier skill score for several meteorological parameters and several thresholds.

3 Preliminary conclusions

From the case studies and the experiment with downscaling the PEACE members it seems that the PEACE provided initial and boundary conditions are not really optimal for the local ensemble run, for a Central European application. It can be understood if we consider that it was calibrated to Western Europe. Our aim is to find an optimal method, which fits our purposes.

Changing the target domain and possibly also the target time seems to be a good way of increasing the spread over the area of our interest, but this method requires the rerun of the global ensemble system.

An alternative method can be the combination of two different sets of singular vectors. Preliminary results seem to be promising, but still lots of work has to be done in this field.

4 Future plans

We would like to continue to further investigate the topic of combining two sets of singular vectors. The experiment should be continued with combining the two sets not in the framework of ARPEGE, but in the framework of ALADIN, on the way which is described in section 2.2. and on fig. 17, and check whether the the same improvement can be obtained like what we achieved in the case of ARPEGE.

Besides it is important to test the ensemble system on (much) more cases distributed in all four seasons (so far we ran experiments for four consecutive days from autumn 2003, ten consecutive days from summer 2004, and three case studies, two from the summer period and one from the winter, but the sample is not big enough so far), and to test it on independent cases instead of consecutive days.

Also it is planned to start the experiments with the other methods especially with ALADIN native SV perturbations, but there is still a lot of work to be done to be able to run this configuration.

The errors in the forecasts are not only caused by the errors in the analysis, but also by the errors in the model itself (e.g. from the parameterisation of physical processes). A possible approach of this problem could be to run the model with different parameterisation schemes and/or by changing the parameters that represent important assumptions in the parameterisation. Work on this field has already started at Météo-France, and it would be useful to investigate the efficiency of such an ensemble system.

As our final goal is to develop an operational short range ensemble system we also have to concider the problem of transfering the lateral boundary conditions. Since the ensemble system consists of 10+1 members, there is a significant amount of data which has to be transfered. To solve this problem different proposals can be made:

- To discriminate the information coming from the perturbed lateral boundary conditions and the perturbed initial conditions. If it is found that the information coming from the perturbed initial conditions is more important than the information coming from the perturbed lateral boundary conditions, a possible solution could be that e.g. in the first 24 h of the forecast the lateral boundary conditions for every member would be provided by the control run of the global ensemble system, and only after 24 h would we use the perturbed lateral boundary conditions. This would reduce the amount of data which has to be transfered by nearly 50%.
- The PEACE system runs every day starting from the 18 UTC analysis. By runing the LAMEPS starting at 00 UTC and using initial and lateral boundary conditions from the (previous) 18 UTC PEACE run, we could gain some time which could be used to download the lateral boundary conditions for the 10+1 members. This possibility could also be investigated in detail.

CONTENTS