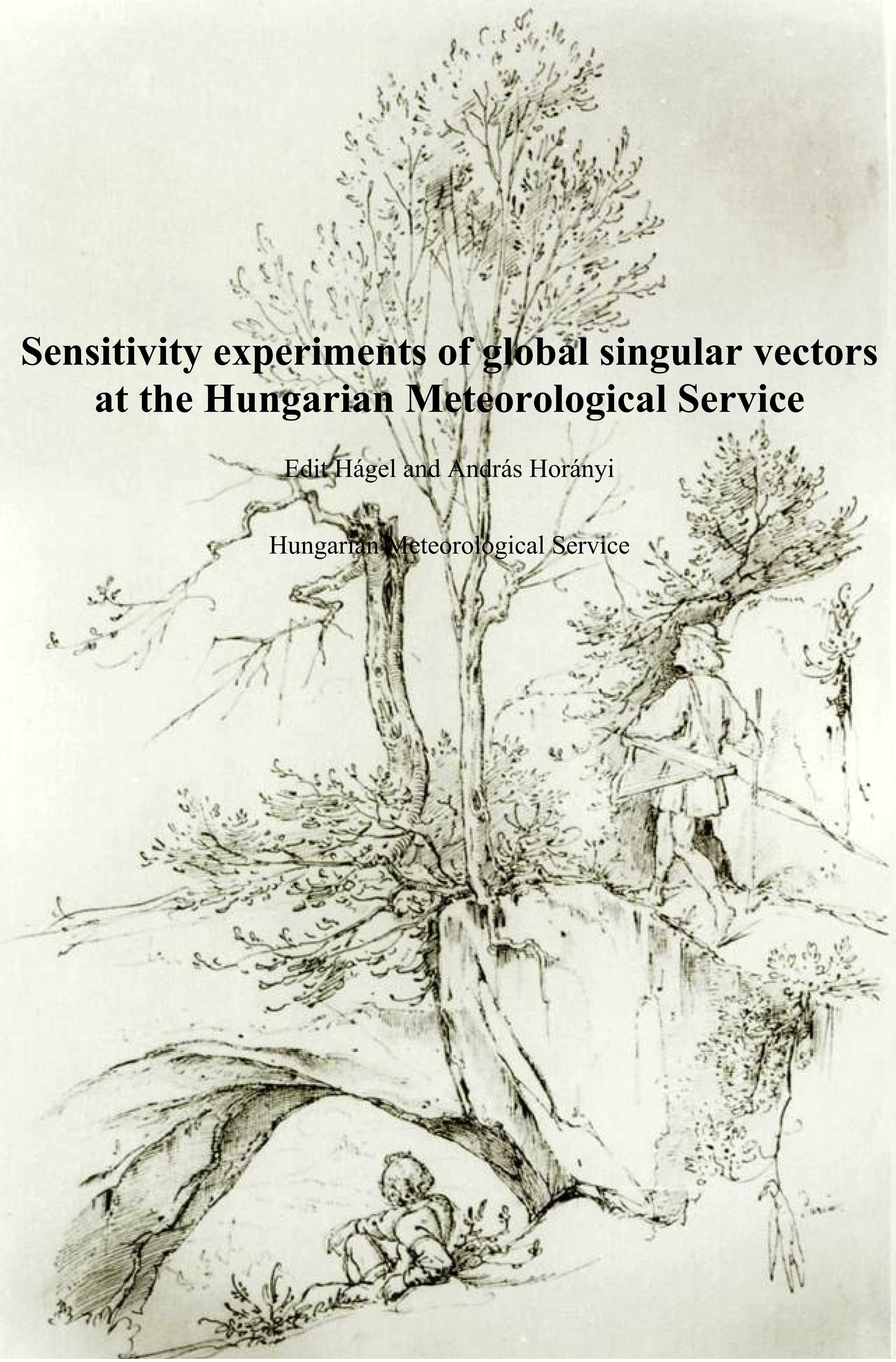
Sensitivity experiments of global singular vectors at the Hungarian Meteorological Service

Edit Hágel and András Horányi

Hungarian Meteorological Service



1. Introduction

In the last couple of years intensive research has started to develop short-range global and limited area ensemble prediction systems (LAMEPS) for the mesoscale. Most of the studies show the benefits of limited area ensemble forecasting, but it is not yet clear, which is the best method for the short-range mesoscale application. Motivated by these results research started on this field at the Hungarian Meteorological Service (HMS) too. It was decided to start with the direct downscaling of global ensemble members. The so-called PEACE[1] system was used to provide initial and lateral boundary conditions for the limited area experiments. In PEACE, targeted singular vectors are used to generate the initial perturbations.

When applying the singular vector method to generate initial perturbations for ensemble forecasting one has to keep in mind the importance of the singular vector target domain and target time (*Frogner* and *Iversen,* 2001, 2002*; Hersbach et al.,* 2000). These characteristics should be chosen such that they yield perturbations optimized to the area of interest (i.e. Central Europe and particularly Hungary in our case) and to the given forecast length (typically 48 hours). In the PEACE system the SV target domain is a rather large area covering Europe, the northern part of the Atlantic Ocean and even a small part of the North American continent (Figure 2). The SV target time is fixed to 12 hours. Altogether the system was calibrated in order to get enough ensemble spread over Western Europe for wind speed, 500 hPa geopotential height and mean sea level pressure. This raises some important questions as far as the design of a similar system for Central Europe is concerned:

- Are the initial and lateral boundary conditions directly provided by PEACE convenient for a Central European LAMEPS application?
- Is there a large sensitivity with respect to target domain and target time used in the global singular vector computation? If so, what is the optimal configuration for our purposes (i.e. LAMEPS for Central Europe)?

To answer these questions several experiments have been performed. From the beginning this work was divided into two parts. On the one hand the direct downscaling of the PEACE members was examined. On the other hand sensitivity experiments were carried out to investigate the impact of different target domains and target times during the global SV computation. Results of the direct downscaling and the sensitivity experiments were compared to one another and they are going to be presented in this article.

2. Methodology

2.1. The applied models

For the experiments the ARPEGE/ALADIN modeling system was used. The singular vector computations and the global integrations were performed with the ARPEGE model, while the limited area experiments were carried out with the ALADIN model. On the one hand the direct downscaling of the PEACE members was examined. On the other hand sensitivity experiments were carried out to investigate the impact of different target domains and target times during the global SV computation. Therefore a global ARPEGE ensemble system was set up for the experiments based on the PEACE system. The only difference was in the choice of target domain and target time used for the global singular vector computations. For the limited area experiments the ALADIN model was used on 12 km horizontal resolution with 37 vertical levels. The integration domain is shown on Figure 1.

The initial and lateral boundary conditions were provided by the global ensemble systems described above.

[1]PEACE: **P**revision d'**E**nsemble **A** **C**ourte **E**chéance, an ARPEGE based global short-range ensemble system which runs operationally at Météo-France.

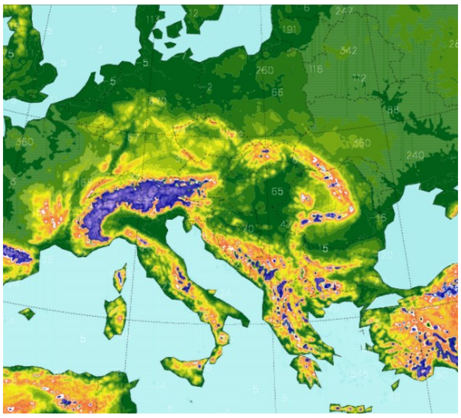


Figure 1. The integration domain and orography of the ALADIN model.

2.2. Description of the experiments

Motivated by some earlier results in the field of short-range limited area ensemble forecasting (*Frogner* and *Iversen,* 2001, 2002*; Hersbach et al.,* 2000) it was decided to investigate the sensitivity of the global singular vector computation in terms of target domain and target time with the main goal to find an optimal configuration for a Central European application. First, case studies were investigated for significantly different meteorological situations in order to see whether the change of the target domain and target time for the global singular vector computations can have a significant effect on the quality of the forecasts valid for the Central European area. Target domains were chosen with different size and location as follows (Figure 2):

- Domain 1: covering the Atlantic Ocean and Western Europe (as used in a former PEACE version, when experiments were started at HMS),
- Domain 2: covering Europe and some of the Atlantic Ocean,
- Domain 3: covering nearly whole Europe,
- Domain 4: covering a slightly larger area than Hungary.

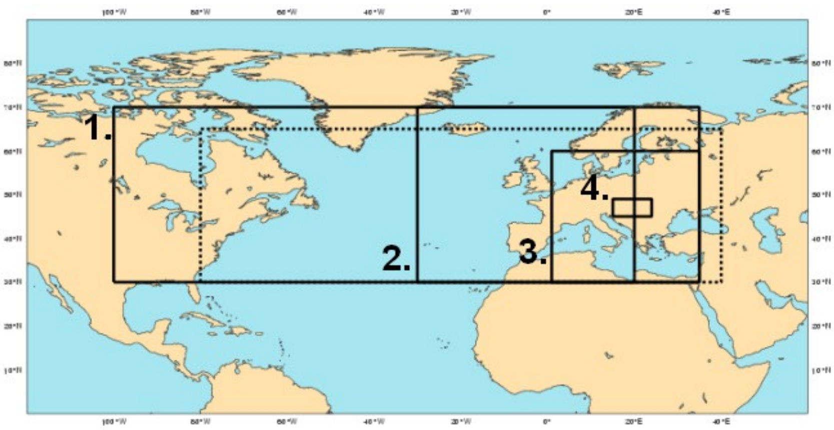


Figure 2. The location of the four different target domains used for the experiments and the target domain used in the present PEACE system (dotted line).

As far as target time is concerned, 12 hours (as used in the PEACE system) and 24 hours were chosen. Due to the linearity assumptions within the theory of SV computations the maximum length of the target time is about 48 hours. However the primary aim is to provide short-range forecasts, therefore a target time considerably less than 48 hours should be chosen for ensuring the desired impact of the perturbations during the forecast range. This argumentation justifies the choice having 12 hours and 24 hours as target times for the experiments.

Based on the results of the case studies (*Hágel* and *Szépszó*, 2004; *Hágel*, 2005), further experiments were carried out for a 10 day summer period. Then the following four configurations were examined in detail:

- SV target domain 1, target time 12 hours (as used in a former PEACE version, when experiments were started at HMS)
- SV target domain 1, target time 24 hours
- SV target domain 2, target time 12 hours
- SV target domain 2, target time 24 hours

Based on the result of the 10 day summer period (*Hágel*, 2005) and inspired by the fact that in between important changes took place in the PEACE system, it was decided to examine the following two configurations for an additional 32 day winter period:

- target domain and target time as used in the present PEACE system (dotted rectangle on Figure 2 as target domain and 12 hours as target time)
- target domain 2 and target time 24 hours

2.3. Verification methods

Results of the case studies and the experiments covering longer periods were examined in detail. Both subjective and objective verification were performed. For subjective verification the ensemble members were visualized in the form of probability maps, "stamp" and "plume" diagrams. For the objective verification, different scores were computed and several types of diagrams were derived such as Talagrand diagram, Percentage of outliers, ROC and Reliability diagrams (*Toth et al.,* 2003*; Persson and Grazzini,* 2005). The performance of the ensemble mean and the control forecast was compared to one another. The objective verification was performed against SYNOP (surface) and TEMP (upper air) data. Additionally for the winter period, verification was also carried out with respect to the ECMWF 4d-var analysis. The verification area was the entire integration domain of the ALADIN model (Figure 1).

3. Results

The experimentation was concentrating on the sensitivity of global singular vectors with respect to their target domain and target time (altogether 5 target domains and 2 target times were considered). Case studies for some significantly different meteorological situations and investigations for longer periods (10 days during summer and 32 days during winter) were analyzed to understand the impact of these important characteristics of the singular vector calculations.

Results of the case studies and the 10 day period were already described in previous articles in the ALADIN Newsletter (*Hágel* and *Szépszó*, 2004; *Hágel*, 2005). Hereafter the results of the 32 day winter period and the overall conclusions of the sensitivity experiments will be presented.

3.1. Experiments for a winter period of 32 days

According to previous experiments (case studies, 10 day summer period) it was concluded that great sensitivity (at least in terms of spread) could be found with respect to the target domain and target time used in the global singular vector computation. It was additionally realized that a period of ten days is not sufficiently long for drawing reliable conclusions therefore larger sample is desirable. However it could be concluded that the target domain 2 with target time 24 hours seems to be a better choice for a Central European application than target domain 1 complemented with target time 12 hours (as used in the PEACE system at that time). In addition and simultaneously to these preliminary conclusions, important changes (and operational introduction) had been encountered at Météo-France PEACE system. The following characteristics were changed:

- the resolution used for the SV computation was changed from T63 to T95,
- the target domain became smaller and was shifted towards east,
- the resolution used for the integration was changed from T199 to T358.

Therefore extended experiments were made for another (longer) period (the choice of this period was again arbitrary) covering 32 days in January and February, 2005. It is important to note that this period was characterized by an unusually cold weather.

Altogether two different configurations were examined: the operational PEACE configuration and target domain 2 together with target time 24 hours to be used for the global SV computations. For the objective evaluation Talagrand, ROC and reliability diagrams were drawn, bias and RMSE of the ensemble mean and the control forecast were computed for ARPEGE and ALADIN respectively.

3.1.1. Ensemble mean vs. control forecast

The first, basic validation of an ensemble system is the comparison of the performance of the ensemble mean and the control forecast (the minimum requirement is that the ensemble mean should provide better results than the control run). For every examined parameter (10 meter wind, 2 meter temperature, 500 hPa geopotential height, 850 hPa temperature) the values of the ensemble mean and the control forecast were relatively close to one another with a slight advantage to the ensemble mean (not shown). The improvement of the ensemble mean is more pronounced near the surface. All this only means that the ensemble system meets the above-mentioned (basic) criterion and further evaluations can be performed.

3.1.2. Spread vs. RMSE

It is expected that the behavior of the ensemble spread and the error is similar (i.e. if the error is small, then the spread should be small as well and vice versa). For the examined parameters it was found that the spread is usually smaller than the error, however the use of the smaller SV target domain (domain 2) and the 24 hours target time reduced the difference between them. Moreover for 500 hPa geopotential the spread became even larger than the RMSE of the ensemble mean (not shown). It can be concluded that there is a discrepancy between the error and the spread, however with the correct choice of SV target domain and target time this can be reduced (especially at higher levels).

3.1.3. Talagrand diagrams and percentage of outliers

It was found that the change of the target domain and target time during the global SV computation could improve the system's ability to comprise the true state of the atmosphere. For all parameters the Talagrand diagrams became flatter, the distribution moved towards the ideal one (not shown). Looking at the percentage of outliers, clear improvement can be seen especially for upper level parameters, but also to some extent for the surface ones (see Figure 3). It is also interesting to notice that on the surface the improvement for the wind speed is more emphasized than that of the temperature. Moreover the 2 meter temperature is one of the worst parameters in that characteristics (it is expected that the surface wind is a rather good parameter of the dynamical adaptation due to the fine scale surface description, however the erroneous behavior of the temperature is a rather puzzling feature).

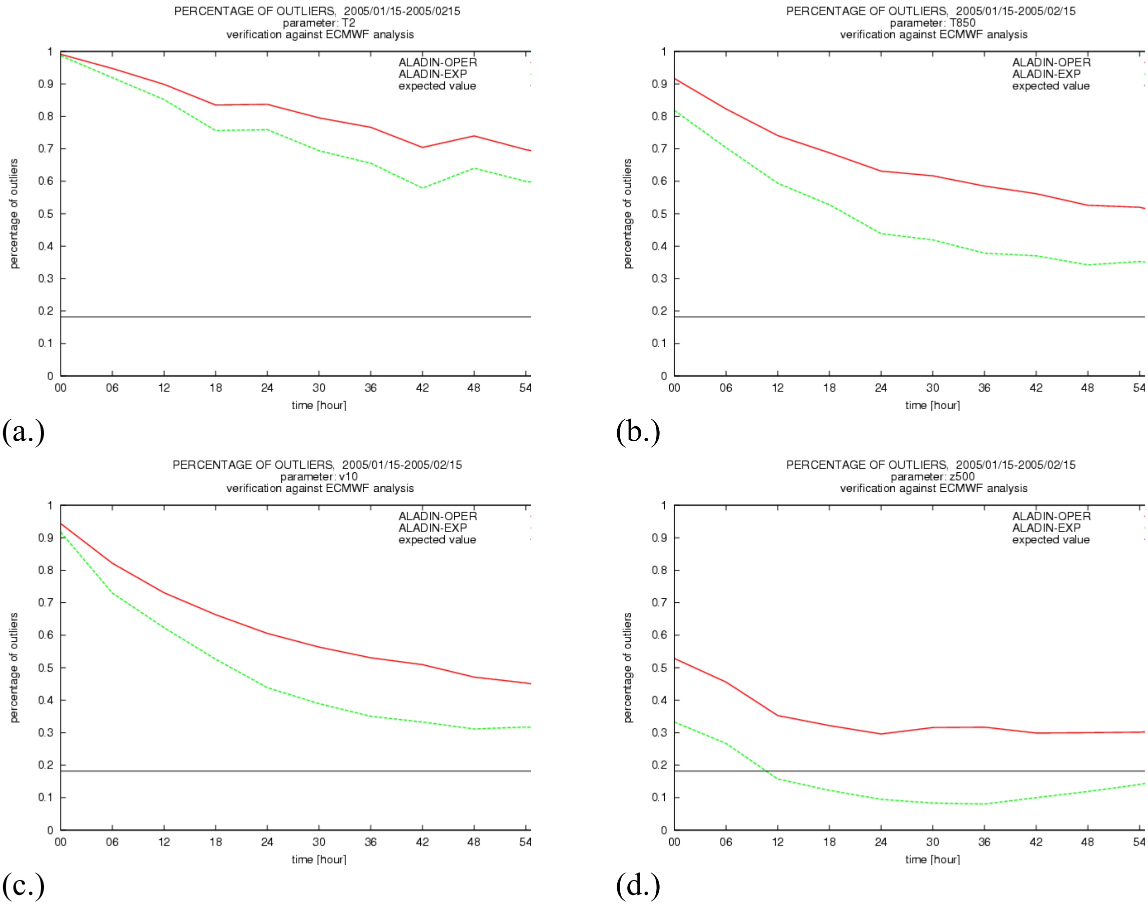


Figure 3. Percentage of outliers diagrams for the ALADIN ensemble system for the period 2005/01/15-2005/02/15. (a.) 2 meter temperature, (b.) 850 hPa temperature, (c.) 10 meter wind speed, (d.) 500 hPa geopotential height. Red line is ALADIN coupled with the operational PEACE forecasts, green line is ALADIN coupled with the experimental set. Verification was performed against ECMWF analysis. The expected value is ~ 0.2 (see the thin horizontal lines).

3.1.4. ROC area

As already mentioned before, changing the singular vector target domain and target time yields clear improvement in terms of spread. ROC diagrams were derived and examined in detail for 10 meter wind speed (with thresholds such as 2, 5, 10 and 15 m/s respectively) and 850 hPa temperature anomaly (with thresholds ±8 Celsius and ±4 Celsius). The integral area under the ROC curve was computed and results from the two configurations (operational and experimental) were compared.

For the 850 hPa temperature anomaly better results were obtained while using the experimental set (using modified target domain and target time for the global SV computation) of global ensemble forecasts as initial and lateral boundary conditions for the ALADIN model (Figure 4). The ROC area shows rather good scores for the –4 Celsius threshold (without loss of quality with the integration time), however the relative improvement is higher for the –8 Celsius threshold value.

For the 10 meter wind speed the improvement is less significant compared to the 850 hPa temperature anomaly. However, the change of the target domain and target time yields clear improvement for this parameter as well (see Figure 4). Maybe two additional features can be further mentioned for the 10 meter wind speed (based also on the figure for 10 m/s threshold, not shown): on the one hand the scores are getting better while using higher threshold values (the quality of the ensemble system increases for stronger wind values which is an encouraging result, especially if one would like to represent correctly extreme events). On the other hand there is a jump in quality for the bigger thresholds just after the analysis time (this might correspond with some spin-up effects).

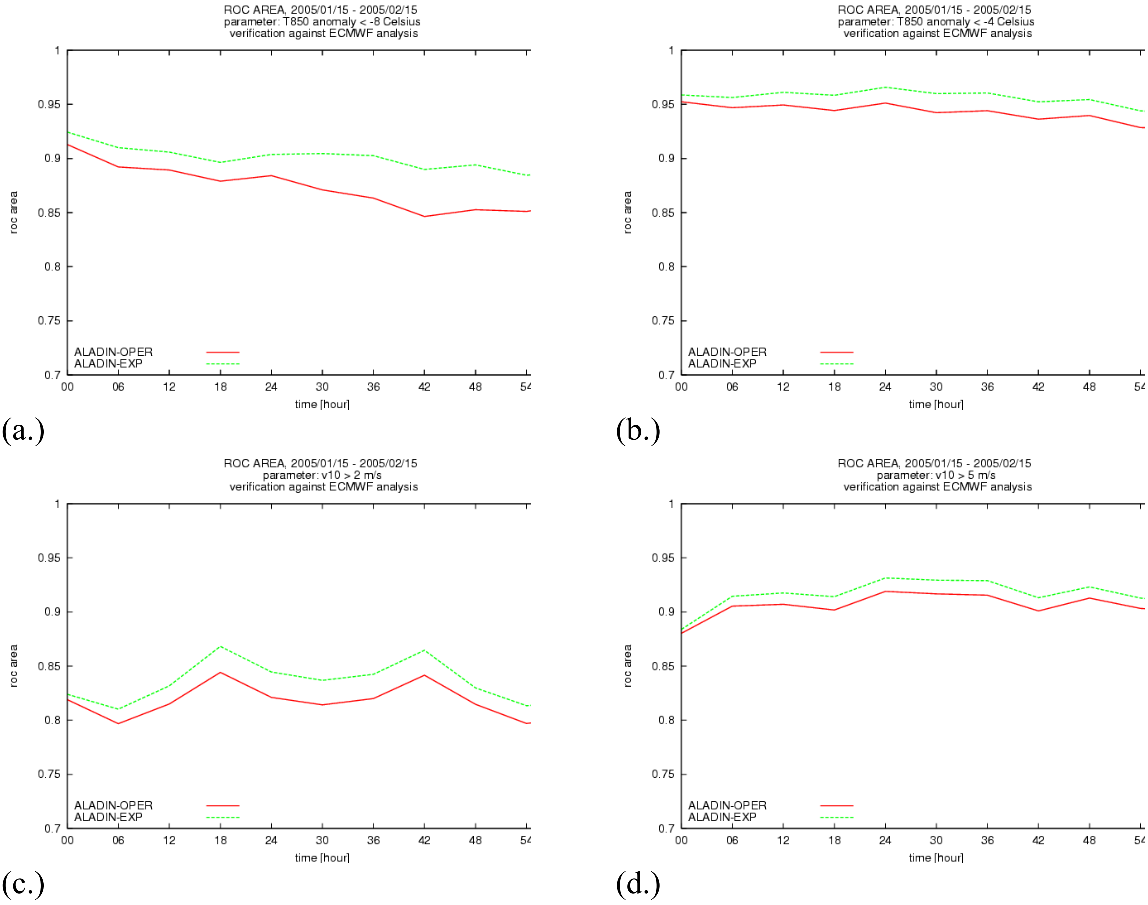


Figure 4. ROC area for the ALADIN ensemble system for the period 2005/01/15-2005/02/15. (a.) 850 hPa temperature anomaly less than –8 Celsius, (b.) 850 hPa temperature anomaly less then –4 Celsius, (c.) 10 meter wind speed greater than 2 m/s, (d.) 10 meter wind speed greater than 5 m/s. Red line is ALADIN coupled with the operational PEACE forecasts, green line is ALADIN coupled with the experimental set. Verification was performed against ECMWF analysis. (The ROC area of a perfect forecast is 1.)

3.1.5. Reliability diagrams

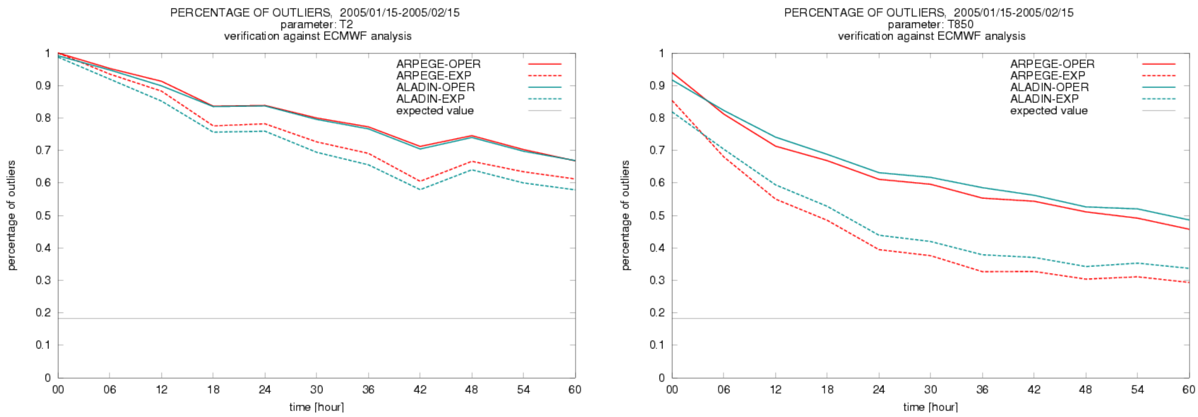
Reliability diagrams were drawn for the same parameters (10 meter wind speed and 850 hPa temperature anomaly) and thresholds as for the ROC diagram. In this case the use of target domain 2 and target time 24 hours did not result in significantly better forecasts, the diagrams of the two ALADIN configurations (ALADIN coupled with the PEACE members and ALADIN coupled with the experimental set) were rather similar (not shown). Nevertheless, it can be concluded that the use of target domain 2 and target time 24 hours kept the same quality of the forecasts in this particular measure.

As an overall conclusion for the 32 day experiment it can be said, that the change of target area from domain 1 to domain 2, together with the change of target time from 12 hours to 24 hours can increase the quality of the ensemble forecasts valid for the verification area. This improvement is true for both the ARPEGE and the ALADIN ensemble systems. For upper level parameters (e.g. 500 hPa geopotential) the improvement is more notable than for some surface parameters. Regarding the surface variables there are large differences between temperature and wind speed: the 2 meter temperature is a rather weak point of the system (seen from the percentage of outliers), while the 10 meter wind speed is proven to be a well-predictable parameter in ensemble sense as well (especially for the higher threshold values). This contradictory surface behavior might be explained by the fact that regarding the surface characteristics only pressure is perturbed in the global ensemble system.

3.2. Comparison of global and limited area ensemble systems

When making (ensemble) forecasts with a limited area model it is always a key aspect to consider whether the limited area model is producing more enhanced ensemble forecasts than the global one. Therefore during the objective verification both the ARPEGE (global) and the ALADIN (limited area) models were verified and then inter-compared.

Looking at the percentage of outliers one can conclude that the simple downscaling of the global ensemble system with the ALADIN model does not yield significant improvement. For some parameters the ALADIN forecasts have better scores, for others the ARPEGE ones. In Figure 5 one can see that for 2 meter temperature ALADIN coupled with the experimental set performs better, while for 850 hPa temperature the experimental ARPEGE ensemble system has the best results (for any case the differences are small). This result can be explained with the consideration that the higher resolution ALADIN forecasts are gaining advantage near the surface due to the more precise description of surface characteristics and processes.



(a.) (b.)

Figure 5. Percentage of outliers diagrams for ARPEGE and ALADIN ensemble systems for the period 2005/01/15-2005/02/15. (a.) 2 meter temperature, (b.) 850 hPa temperature. Solid red line is the operational PEACE forecasts (ARPEGE-OPER), solid green line is ALADIN coupled with the operational PEACE members (ALADIN-OPER), dashed red line is the experimental ARPEGE ensemble (ARPEGE-EXP), dashed green line is the ALADIN model coupled with the experimental set (ALADIN-EXP). Verification was performed against ECMWF analysis. The expected value is ~ 0.2 (see the thin horizontal lines).

When examining the ROC area diagrams, for both parameters (10 meter wind, 850 hPa temperature) it seems to be hard to tell whether ALADIN or ARPEGE performs better. For certain thresholds and parameters ALADIN had better scores, for other thresholds ARPEGE was more successful. There were also combinations (in terms of variables and thresholds) when the two models had nearly the same skill (not shown).

As far as the reliability diagrams are concerned (for 10 meter wind speed and 850 hPa temperature) no significant differences can be seen between the results of the global and the limited area ensemble systems (not shown).

As a summary it can be said that generally speaking by the simple downscaling of the ARPEGE ensemble system with the higher resolution ALADIN model it is very difficult to achieve significant improvements. One explanation behind this result might be that on the one hand the resolution difference between the ARPEGE and ALADIN models is too small, on the other hand the influence coming from the lateral boundary conditions results in a rather strong forcing for the results of the limited area model. Additional explanation might come from the fact that the formulation and especially the physical parameterization package of the global (ARPEGE) and the limited area (ALADIN) models are rather similar. For the surface fields, where one would expect improvements (due to the more precise description of surface characteristics in the higher resolution model) maybe the benefits (which are reflected in the near surface wind fields, but not in the temperature field) are compensated by the fact that only the surface pressure as model prognostic variable is perturbed by the global system, therefore the initial uncertainties in the surface description are not properly addressed with the limited area ensemble system.

4. Summary, conclusions and future plans

Extended experiments were performed to investigate the sensitivity of global singular vector computations in terms of target domain and target time. Global (ARPEGE) ensemble members were downscaled with the limited area model ALADIN. The experimentation consisted of individual case studies, 10 days (in summer) and 32 days (in winter) continuous tests. Results show that the proper choice of the SV target domain and target time are important factors for the increase of the ensemble spread and on average for the improvement of the skill of the ensemble system (at least on average level). This conclusion is valid for ARPEGE global and ALADIN limited area forecasts as well. Thus, changing the target domain and target time can improve the system's ability to comprise the true state of the atmosphere. The improvements are clearly demonstrated for all parameters (especially at upper levels) by the percentage of outliers and ROC area diagrams.

A systematic comparison between ARPEGE and ALADIN ensemble systems was also carried out. From the results one can conclude that the simple downscaling of the ARPEGE ensemble members with the higher resolution ALADIN model does not improve significantly the forecast skill (even more for certain parameters the ARPEGE model performs better). The reason of this feature might be sought in the limited resolution difference between the global and the limited area models, the too strong impact of the lateral boundary conditions, the similarities between the model formulations and the lack of perturbations for the surface fields.

These conclusions indicate that the direct downscaling of the ARPEGE ensemble system is not sufficient to obtain a good, high resolution limited area ensemble system: there is a strong need of the development of methods, which are properly and directly accounting for the mesoscale uncertainties in the initial conditions of the ALADIN model. At the same time research should be pursued towards the consideration of other sources of uncertainties in the limited area models (for instance deficiencies in the description of the parameterized processes) as well.

Acknowledgements - This work was supported by the Hungarian National Research Fund (OTKA, grant Nº T/F 047295) and the Hungarian National Office for Research and Technology (NKFP, grant Nº 3A/051/2004 and JÁP, grant Nº 2/007/2005).

5. References

Frogner, I.-L. and Iversen, T., 2001: Targeted ensemble prediction for northen Europe and parts of the north Atlantic Ocean. *Tellus 53A*, 35-55

Frogner, I.-L. and Iversen, T., 2002: High-resolution limited-area ensemble predictions based on low-resolution targeted singular vectors. *Q. J. R. Meteorol. Soc.128*, 1321-1341

Hágel, E. and Szépszó, G., 2004: Preliminary results of LAMEPS experiments at the Hungarian Meteorological Service. *ALADIN Newsletter no 26*.

Hágel, E., 2005: Latest results of the LAMEPS experiments. *ALADIN Newsletter no 27*.

Hersbach, H., Mureau, R. Opsteegh, J. D. and Barkmeijer, J., 2000: A Short-Range to Early-Medium-Range ensemble Prediction System for the European Area. *Monthly Weather Review 128*, 3501-3519

Persson, A. and Grazzini, F., 2005: The verification of ECMWF forecasts. In: *User Giude to ECMWF Forecast Products.* 67-84. (http://www.ecmwf.int/products/forecasts/guide/user_guide.pdf)

Toth, Z., Talagrand, O., Candille, G. and *Zhu, Y.*, 2003: Probability and ensemble forecasts. In: *Forecast Verification: A practitioner's guide in Atmospheric Science.* (eds.: *I. T. Jolliffe* and *D. B. Stephenson.*) Wiley & Sons, Ltd, 137-163.

Contents