Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 642–651
Printed in Singapore.
TELLUS

Latest developments around the ALADIN operational short-range ensemble prediction system in Hungary

By ANDRÁS HORÁNYI*, MÁTÉ MILE and MIHÁLY SZŰCS, *Hungarian Meteorological Service, 1525 Budapest, P. O. Box 38, Hungary*

(Manuscript received 23 April 2010; in final form 1 February 2011)

ABSTRACT

Ensemble prediction systems (EPSs) are an essential part of numerical weather prediction for the provision of probabilistic forecast guidance. The Hungarian Meteorological Service has implemented a limited area EPS (called ALADIN HUNEPS) based on the ALADIN mesoscale limited area model coupled to the French global ARPEGE EPS (PEARP). The dynamical downscaling method is assessed in terms of ensemble verification scores taking also into account the recent upgrade of the PEARP global system. The verification results point towards some weaknesses of the ALADIN HUNEPS, mainly with respect to the near-surface parameters. Therefore, some improvements are needed so as to provide better and more reliable ensemble predictions for the Carpathian Basin. The application of near-surface perturbations into the surface data assimilation scheme is implemented and tested. The first results show that the surface perturbation method slightly improves the ALADIN HUNEPS, however, further experiments should be made to find out the optimal settings of the method for definite robust improvements of ALADIN HUNEPS.

1. Introduction

The main purpose of weather forecasting is to predict the future states of the atmosphere using the atmospheric hydrothermodynamics system of equations. The governing equations that determine the present and future atmospheric states are well known: conservation of momentum (Newton's second law), energy (first law of thermodynamics), mass and moisture (continuity equations) and finally the ideal gas law. This set of nonlinear partial differential equations has only numerical solution. NWP (numerical weather prediction) models make use of several approximations, notably about these equations, the model formulation and the parametrization schemes. The initial condition is decidedly needed to solve the governing equations. As similar systems, the atmosphere, described by non-linear equations, presents a chaotic nature and is, per force, very sensitive to the initial conditions to the extent that a slight modification leads to large forecast errors. The small number of representative observations is another major source of forecast uncertainties, as the initial condition cannot exactly represent the present state of the atmosphere.

Because of the abovementioned reasons, there will always be some uncertainties in the forecasts of NWP models. The only reasonable solution is to run a set, or, as it is commonly called, an ensemble of forecasts. These runs can differ for instance in the applied model (multimodel), in the used parametrization schemes (multiparametrization) and the fluxes of the parametrized physical processes can be perturbed by random numbers (stochastic physics), but usually the ensemble members are integrated from slightly different initial conditions which are also possible realizations of the 'true' atmospheric state. A forecast made by such an ensemble of runs is called an ensemble prediction system (EPS) and has a clear advantage against the regular 'deterministic' forecasts, namely that it predicts not only one probable future state, but that the members define the probability density functions (PDFs) of the different forecasted meteorological parameters. These PDFs contain useful information about the forecast uncertainty (which depends on the weather situation as well).

Since it was first operationally implemented in 1992 (Buizza et al., 1993; Toth and Kalnay, 1997) ensemble forecasting has been widely used by more and more forecasting centres in global medium-range time scale. In the last couple of years, intensive research has been carried out on the ensemble method linked to limited area models for short-range forecasts, because it seems a plausible way to predict the probability of severe and extreme weather events which are often related to small-scale forcing or geographical impacts (which can be properly described only by high-resolution numerical models).

The first real-time, operational regional EPS was implemented at NCEP (National Centers for Environmental

*Corresponding author.
e-mail: horanyi.a@met.hu
DOI: 10.1111/j.1600-0870.2011.00518.x

Prediction) in 2001. The initial condition perturbations were generated with the so-called 'breeding method' (Toth and Kalnay, 1997) similar to the global ensemble system of NCEP. At AEMET (Spanish Meteorological Service) initial and lateral boundary conditions (LBCs) are downscaled from four different global models and they are coupled with five different limited area models, which result in a multimodel multi-analysis multi-boundary system of 20 members. The SRNWP-PEPS (Short Range Numerical Weather Prediction-Poor man's Ensemble Prediction System) is based on the idea of bringing together all available operational regional high-resolution numerical forecasts (Heizenreder et al., 2006). Within the COSMO (Consortium for Small-Scale Modeling) framework the COSMO-LEPS (Limited area Ensemble Prediction System) was implemented to run an ensemble system with a non-hydrostatic limited area model based on the downscaled representative members of ECMWF (European Centre for Medium-Range Weather Forecasts) EPS (Montani et al., 2003; Marsigli et al., 2005). These members are selected using a clustering technique. It was found that COSMO-LEPS performed better than the lower resolution ECMWF EPS in forecasting (severe) precipitation events. The LAMEPS (Limited Area Model Ensemble Prediction System) of the Norwegian Meteorological Service uses TEPS (global targeted EPS) to provide boundary conditions to HIRLAM (High Resolution Limited Area Model, Frogner and Iversen, 2001, 2002; Jensen et al., 2006). TEPS is similar to the operational ECMWF EPS, but uses specific singular vectors (SVs) to generate initial condition perturbations, which are targeted to have their greatest impact in Northern Europe, which is the area of main concern. It was found that NORLAMEPS (Norwegian LAMEPS), which is the combination of TEPS (20 + 1 members) and their downscaled members (20 + 1 members), can give better results than ECMWF EPS for short-range forecasts. ZAMG (Austrian Meteorological Service), in collaboration with the international LACE (Limited Area Modeling for Central Europe) team, developed its own ALADIN-LAEF (ALADIN Limited Area Ensemble Forecasting) that is an ensemble of 16 perturbed members. In this system, perturbations of the initial condition are generated by a breeding–blending cycle. Forecasts are realized from several model setups (multi-physics) and LBCs are retrieved by the dynamical downscaling of the first 16 ECMWF EPS members (Wang et al., 2009).

The abovementioned results motivated the experimentation at the Hungarian Meteorological Service (HMS) whose final aim was to establish an operational LAMEPS for Central Europe (ALADIN HUNEPS). The dynamical downscaling of both the ARPEGE (Météo-France global spectral model) ensemble system (PEARP) and the ECMWF EPS system were first considered and assessed (Szintai and Ihász, 2006). PEARP was chosen and used operationally since February 2008 at HMS (Hágel and Horányi, 2006). In December 2009, the global PEARP system was upgraded which resulted in a strong impact on the lateral boundaries in ALADIN HUNEPS. A short description of the ALADIN HUNEPS system and the effect of these upgrades are presented in Section 2 illustrated by some standard EPS verification diagnostics.

The first sensitivity studies at HMS were undertaken to see whether forecast skill was improved when replacing the operational PEARP perturbations by differently targeted global SV-based perturbations. Different SV optimization domains were examined and it was found that an optimal choice for Central Europe can increase the spread and improve the skill of the ensemble forecasts (Hágel and Horányi, 2007).

More research was done to improve ALADIN HUNEPS after its operational implementation in Hungary. Some local (mesoscale) perturbations are to be introduced in the system to see whether they would improve the main weaknesses of the operational ensemble forecasts (especially near the surface, since surface uncertainties are relatively poorly described by the original ALADIN HUNEPS). The generation of near-surface perturbations through the surface (optimal interpolation) data assimilation cycle was implemented and tested. The scientific and technical details of this ongoing research about this near-surface perturbation scheme, and preliminary results are presented in Section 3. The results are summarized and discussed in Section 4 and some conclusions will be presented in Section 5.

2. The operational ALADIN HUNEPS and its evolution

The operational LAMEPS at the HMS (ALADIN HUNEPS) is based on the ALADIN limited area model. The latest configuration of ALADIN HUNEPS is fashioned after a simple downscaling method, with PEARP providing the initial and LBCs of the ALADIN integrations. At present the ALADIN ensemble system is operating without any additional local perturbations (and data assimilation), therefore the initial behaviour of the limited area ensemble is mostly determined by the perturbations of the global (PEARP) system. The impact of the LBCs grows in importance during the integration, which is another indication that the quality of the global EPS is an essential factor for the performance of the limited area system.

The 1.5 version of PEARP (hereafter referred as PEARP1.5) was based on the current version of the ARPEGE ensemble system when the LAMEPS experiments started in Hungary. The ARPEGE spectral model (Courtier et al., 1991) with spectral truncation of TL358 and a stretching coefficient of 2.4 was applied to the global ensemble system. The initial perturbations were generated with the combination of initial and evolved SVs. Initial SVs were computed with a 12-h optimization time to four different optimization areas. 16 SVs (on TL95 resolution) targeted on Europe and the North Atlantic region, 10 SVs (TL44) targeted on the Northern Hemisphere and the tropical band ($\pm 30^{\circ}$) and 20 SVs (TL44) targeted on the Southern Hemisphere. Besides that, 24 h evolved perturbations were used from the previous run of PEARP1.5 valid at the time of the current analysis.

Following this procedure, five perturbations were first calculated and then rescaled, based on the computed flow-dependent background errors from the ARPEGE 4D-Var (four-dimensional variational) data assimilation system. These perturbations were added to and subtracted from the initial unperturbed analysis (the 'control'). Thus PEARP1.5 contained 10 additional ensemble members to the control forecast, it ran once a day at 18 UTC with a 108 h integration time.

On 8 December 2009, France implemented a new PEARP version, called PEARP2.0 (Descamps et al., 2009). The main difference between the old and the new versions is an increase in the number of ensemble members (34 perturbed ones with one unperturbed control) and the use of a different perturbation generation method. Initial perturbations are now constructed by the combination of ensemble data assimilation (EDA) perturbations and targeted SVs. The main innovation of PEARP2.0 is a six-member ensemble assimilation suite (which was originally based on 3D-FGAT: 3D-Var with First-Guess at Appropriate Time and later replaced by the 4D-Var scheme; Berre et al., 2009). The EDA system produces six perturbed analyses. An additional 56 differently targeted SVs are computed and then combined into the 34 SV perturbations (by a Gaussian sampling procedure). 30 out of the 34 SV perturbations are then combined with the EDA-based perturbations (where each perturbed EDA analysis is used five times), the remaining four SV perturbations are linked to the control analysis (where no EDA-related perturbations are considered). Another new feature is the inclusion of a multi-physics approach to account for the model errors. The control forecast runs with the operational physical parametrization package of the 'deterministic' ARPEGE, whereas the perturbed members use an additional seven different physics packages in a random distribution.

The ALADIN HUNEPS is just a simple downscaling of the PEARP system. It is running on a domain covering almost the entire continental part of Europe (Fig. 1) with a 12 km horizontal resolution and with 46 levels in the vertical. The integration of the ALADIN ensemble members last up to 60 h (based on the 18 UTC run). The ensemble set of the ALADIN HUNEPS consists of 11 members which are the first 11 ensemble members of the upgraded PEARP (PEARP2.0).

2.1. Verification

The ALADIN HUNEPS was verified against ECMWF analysis data for pressure level fields and against surface observations for screen level parameters. ECMWF analysis as validation references were chosen since it is an independent good quality analysis system. The resolution of the ECMWF analysis as an appropriate one for validation was assessed. For this purpose, three resolutions were used (T399, T799 and T1279) and then verification scores were compared. It ensured that experiments are not sensitive to the ECMWF analysis resolution to the extent that even the poorest resolution can be used to quantify the relative merits of the different EPS systems. Verification measures were computed for the period of 1 January 2010 to 28 February 2010 for the old and new ALADIN HUNEPS systems (based on the PEARP1.5 and PEARP2.0 global systems, respectively) and also for the experimental ALADIN HUNEPS configurations. The verified parameters were temperature and geopotential at various pressure levels (500, 700, 850 and 1000 hPa) and at screen levels (e.g. at 2 m for temperature). The 1000 hPa diagnostics was discarded as much as possible as this pressure level may be suboptimal in mountainous regions (being below the surface), however, comparing results from the two different systems may give some invaluable information about their respective merits. With regards to screen level (2 m temperature for instance) verification measures it must be noted here that although no correction was made from the model height to the observation altitude, it is most probable that the intercomparison results are not effected by this fact (which is also the case when the global and limited area ensembles are compared, since the two systems are compared on the same grid, so consequently, possible verification deficiencies are affecting the two systems equally).

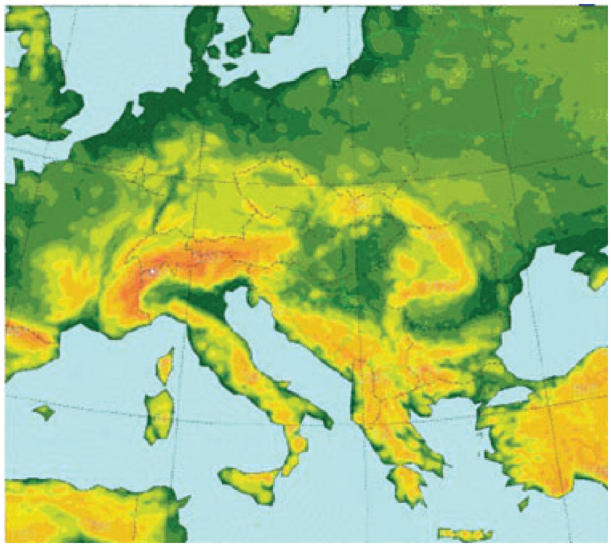

Fig. 1. The domain and the orography of the ALADIN-HUNEPS.

The spread–skill relationship can be used as a quality indicator of the ensemble system, that is, how the forecast error is represented by the ensemble. The main goal of this diagnostics is to determine whether the system is under- or overdispersive (the error is then under- or overestimated by the ensemble), and information about the vertical characteristics of the perturbations can be got from a diagnostic of the different levels.

In Fig. 2, the RMSE of the ensemble mean and of the ensemble spread are plotted for 500 hPa and 2 m temperature. The diagrams show comparative aspects of the ALADIN HUNEPS driven by PEARP1.5 and PEARP2.0, respectively. With the downscaled PEARP1.5, the results show a strong

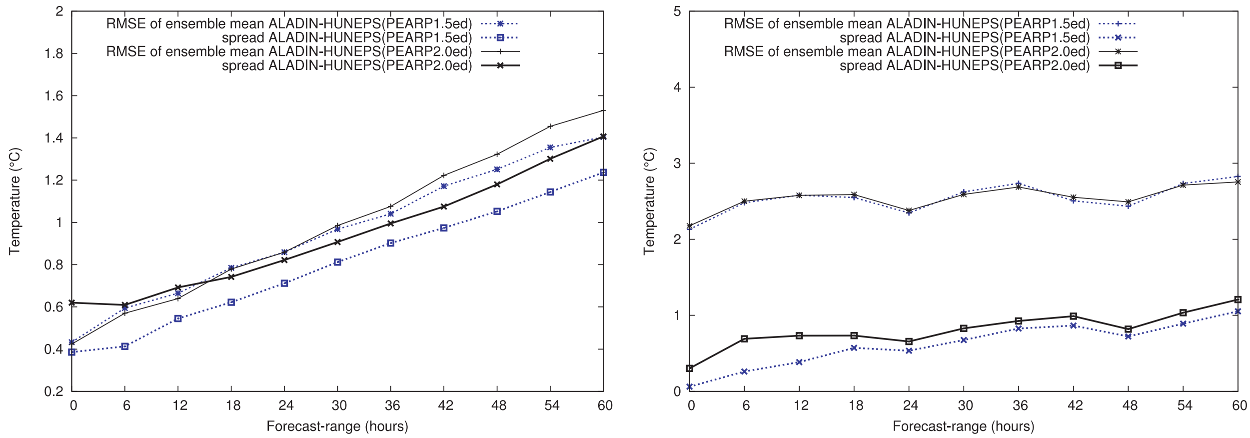

Fig. 2. RMSE of the ensemble mean (thin lines) and spread of the ensemble (bold lines) for temperature at 500 hPa (left figure) and 2 m height (right figure). Solid curves show the ALADIN-HUNEPS driven by PEARP2.0 and dashed curves show the ALADIN-HUNEPS driven by PEARP1.5. Verification period is 1 January 2010 to 28 February 2010. Verification was performed against ECMWF analysis (500 hPa) and Hungarian observations (2 m).

underdispersive behaviour of the ALADIN system for temperature mainly at the lower atmospheric levels (Fig. 2, right panel). With an upgraded PEARP, the skill of the ALADIN HUNEPS becomes significantly better for the higher atmosphere, but remains poor for lower levels and for the surface (though slightly better than with the previous system). It is important to note that this poor near-surface relationship is strongest for temperature (in case of geopotential - not shown - it is not so meaningful). Additionally, for the 500 hPa diagnostics, the new (ALADIN HUNEPS based on PEARP2.0) system has a higher RMSE of the ensemble mean after 18 h of integration than for ALADIN HUNEPS based on PEARP1.5. This somehow surprising feature (the new system is worse in that forecast ranges) might be related to the fact that PEARP2.0 uses a multi-physics option to account for the model uncertainties and the different physics packages used by the global EPS members might not be consistent with the single physics package used for the ALADIN integrations (the fact that the impact is seen only after 18 h indicates that the influence is coming from the LBCs).

Percentage of outliers and Talagrand diagrams are also commonly used ensemble verification characteristics for the representation of the spread and of the quality of the ensemble system (Anderson, 1996; Talagrand et al., 1998).

Figure 3 shows the percentage of outliers for 500 hPa and 2 m temperature. The difference between the old (based on PEARP1.5) and the new ALADIN HUNEPS (based on PEARP2.0) is significant. As it is the case for the spread–skill relationship, the improvement of the new ALADIN HUNEPS on the upper-air variables is quite manifest, which is not the case for the near-surface fields, although some slight improvements

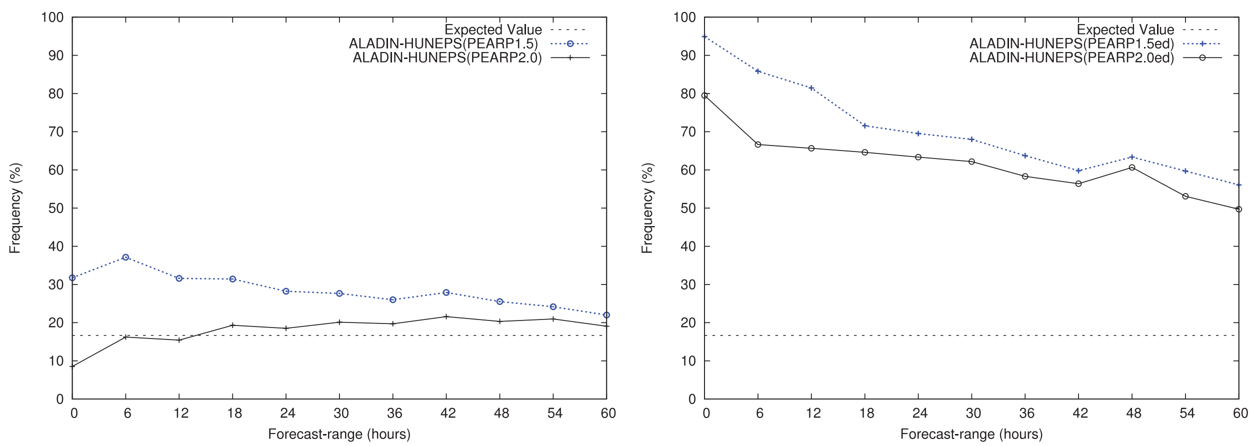

Fig. 3. Percentage of outliers for temperature at 500 hPa (left-hand side) and for 2 m temperature (right-hand side). Solid curve for ALADIN-HUNEPS (PEARP2.0) and dashed curve for ALADIN-HUNEPS (PEARP1.5). Verification period is 1 January 2010 to 28 February 2010. Verification was performed against ECMWF analysis (500 hPa) and Hungarian observations (2 m).

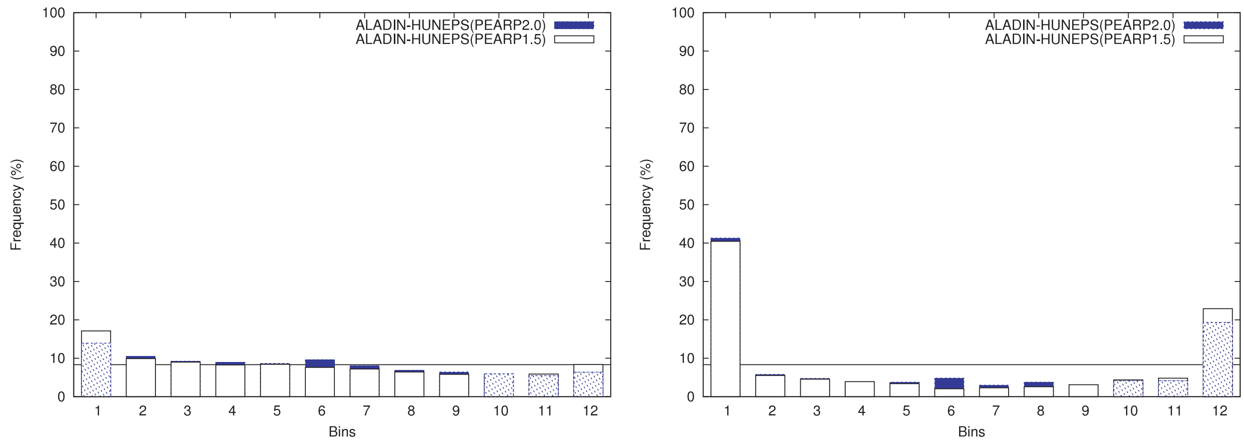

Fig. 4. Talagrand diagrams for 48 h temperature forecasts at 500 hPa (left-hand side) and 2 m height (right-hand side). White boxes for ALADIN-HUNEPS (PEARP1.5) and dark boxes the increments for ALADIN-HUNEPS (PEARP2.0). Verification period is 1 January 2010 to 28 February 2010. Verification was performed against ECMWF analysis (500 hPa) and Hungarian observations (2 m).

can be noticed, the skill of the system falls short of expectations. The new PEARP system now provides a flawless skill at 500 hPa, which is a desired feature of a global EPS. All these aspects are confirmed by the Talagrand diagrams (Fig. 4). The U-shape of Talagrand diagrams decreases for PEARP2.0 in the mid-atmosphere, but at 2 m and for the surface (not shown), the verifying analysis is still quite often outside the interval defined by the sorted ensemble members, hence the spread of the LAMEPS is insufficient near to the surface. The other important conclusion of the verification is that although the new set of global perturbations in principle (since the global EDA system-based perturbations are originated from the perturbation of different near-surface observations as well) improves the treatment of surface uncertainties, however in practise it does not simultaneously improve the surface spread for ALADIN HUNEPS, giving the impression that the surface behaviour of the limited area model is somehow inherently constrained in terms of further improvements during the downscaling procedure. Furthermore, improvements for the near-surface should be sought in the framework of the limited area model rather than of the global one.

A short comparative verification of the driving (global) and the limited area (LAM) systems was also performed in order to see whether the LAM (ALADIN HUNEPS) can provide an added value to the global EPS. The LBCs of the first 11 members of PEARP2.0 were post-processed at the same latitude–longitude grid, where the operational forecast products are considered. To be noted here that in terms of quality, no real difference was found between the PEARP system represented by its first 11 members and the full 35 members (not shown), consequently the conclusions below are also valid in the context of the lower resolution full (35 members) global ensemble (PEARP) and the higher resolution, but lower ensemble sized (11 members) limited area system (ALADIN HUNEPS). Figure 5 shows the spread–skill relationship of the global EPS and the ALADIN HUNEPS for 500 hPa and 2 m temperature, respectively. At 500 hPa, a slight overdispersive behaviour of the global PEARP2.0 can be seen where the spread of the global system is a bit larger than the RMSE of the ensemble mean, however the spread–skill relationship is rather satisfactory at this level. For the same level and for ALADIN-HUNEPS, a rather slight underdispersive behaviour can be noticed with a very moderate degradation of the spread–skill relationship with respect to the global system. At 2 m (as shown in previous figures) the skill of both ensemble systems is worse than for the upper air, however, the ALADIN HUNEPS limited area ensemble significantly improves in quality with decreasing values of RMSE (this being due to the higher resolution limited area model and with a better description of the corresponding surface). Therefore, the limited area ensemble is able to increase the quality of the global EPS near the surface, nevertheless, the spread–skill relationship is still far from being optimal, which means that the ALADIN HUNEPS system needs to be improved (to be described in the next section).

3. Computation of near-surface perturbations

The main objective of our efforts is to improve the skill of the near-surface fields in the ALADIN HUNEPS. The operational ALADIN HUNEPS verification (see Section 2.1 for more details), pointed out that one of the weakest aspect of the system was its performance near the surface, therefore it seemed a matter-of-course to start to deal with the generation of near-surface perturbations on the top of the dynamical downscaling of the PEARP2.0 system.

3.1. Methodology

In order to use all available information to determine the initial state of the atmosphere, modern data assimilation algorithms are used to merge several kind of observations and model

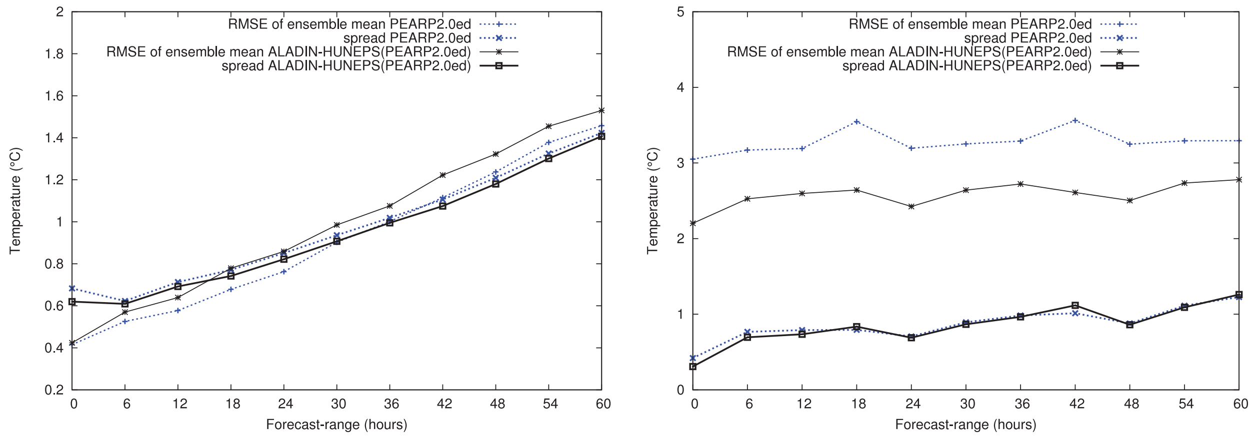

Fig. 5. RMSE of the ensemble mean (thin lines) and spread of the ensemble (bold lines) for temperature at 500 hPa (left-hand side) and 2 m (right-hand side). Dashed curves show the PEARP2.0 and solid curves show the ALADIN HUNEPS driven by PEARP2.0. Verification period is 1 January 2010 to 28 February 2010. Verification was performed against ECMWF analysis (500 hPa) and Hungarian observations (2 m).

information (background) to the analysis. At the HMS, the initial conditions of the ALADIN model are operationally produced by the surface optimal interpolation (OI) and by the upper-air 3D-Var (three-dimensional variational data assimilation) techniques (Bouttier and Courtier, 1999; Bölöni, 2006; Horányi et al., 2006).

For the ensemble system, at this stage of the experiment, the OI configuration of the ALADIN model is applied to inject near-surface perturbations through a surface assimilation cycle. SYNOP (surface synoptic) and TEMP (upper-air radiosonde) observations for the 2 m temperature and humidity are used in the surface assimilation in order to update the surface (soil) temperature and the moisture (Bouttier et al., 1993; Giard and Bazile, 2000). The perturbation of near-surface observations (2 m temperature and humidity) implies a perturbation in the initial soil temperature and moisture, which impacts the background forecasts as well. This results in a combined EPS, composed of an ensemble of surface assimilation cycles for the representation of the surface initial uncertainties and of the downscaled PEARP2.0 system. The core of the near-surface perturbation procedure is a pseudo-random-number generator which defines an initial random number from a normal distribution. The algorithm provides perturbations, which also take into consideration the observation errors, therefore, every observation is multiplied by a random value, which is computed based on a normal distribution, which is characterized by zero mean and standard deviation equal to the observation error. It was furthermore ensured and checked that the analysis quality control did not filter out the perturbed observations. 11 sets of near-surface perturbations were created and then analysed with a 6 h frequency, providing 11 different assimilation cycles. Every day at 18 UTC, integrations were realised for 60 h (as it is the case for the operational ALADIN HUNEPS). Regarding the LBCs the PEARP2.0 forecasts are used to run data assimilation cycles (for each member of the surface assimilation cycle ensemble, different boundary conditions were used) as well as in production. Figure 6 shows the schematic figure of the 'perturbed' assimilation cycle.

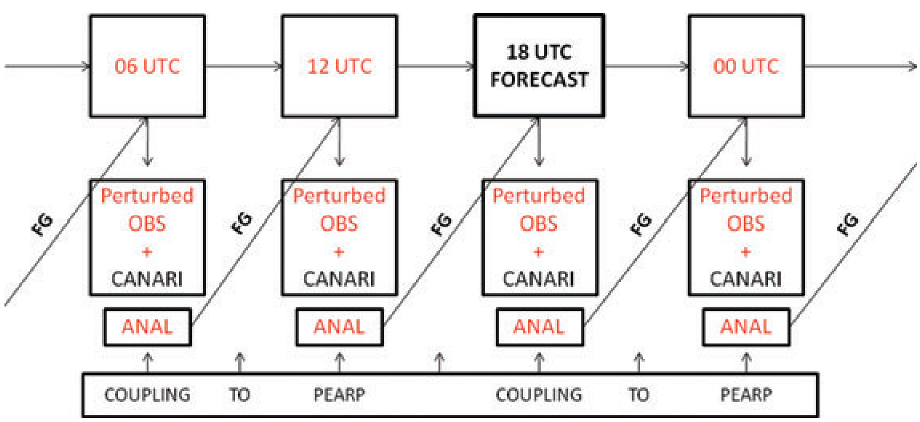

Fig. 6. The schematic figure of the perturbed assimilation cycle.

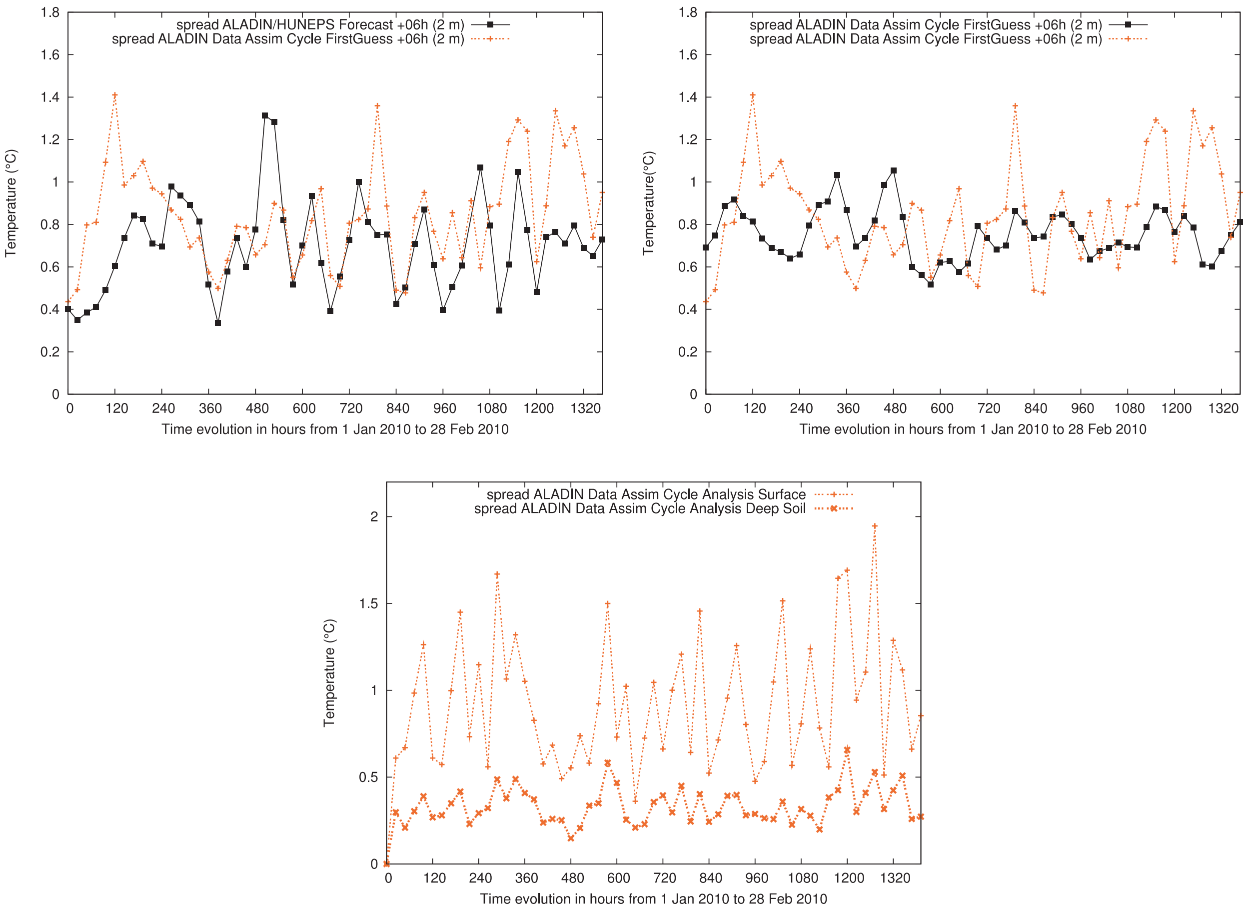

Fig. 7. The spread of the different temperature perturbations related to the near-surface perturbation scheme. Upper left panel: the spread of the first guess in the experimental data assimilation cycle and in the operational ALADIN HUNEPS. Dashed lines correspond with the spread of the first guesses at 2 m. Solid line marks the spread of the 6 h operational ALADIN HUNEPS forecasts at 2 m. Upper right panel: the same as the left panel except that the solid line corresponds with the spread of the first guess at 850 hPa level. Bottom panel: the spread of the surface temperature analyses at two different surface layers in the data assimilation cycle with near-surface perturbations. Dashed line correspond to the surface values and the thick dashed lines mark the values at soil level nearest to the surface (at around 1.5 m depth). Period: 1 January 2010 to 28 February 2010.

3.2. Results

The ensemble of surface assimilation cycles ran from the 1 January 2010 to the 28 February 2010.

The question of impact of the near-surface perturbations through the inclusion of random near-surface observational perturbations correctly affecting the surface assimilation cycle was first assessed. Several diagnostics were made (Fig. 7): the intercomparison of the spread of temperature in the data assimilation first guesses and in the 6 h forecasts of the reference ALADIN HUNEPS members; the comparison of spread at 2 m and at 850 hPa in the assimilation cycle first guesses; the spread of the surface perturbations themselves in the data assimilation analyses. The first diagnostics (Fig. 7, upper-left panel) gives an indication as to the verisimilitude of the amplitudes of the inserted perturbations with respect to the global PEARP2.0 ones. It seems that the new 2 m temperature perturbations, usually larger than the PEARP2.0 ones, are of a realistic size. This latter characteristic was expected since the PEARP2.0 system does not directly include surface perturbations, consequently, the downscaled PEARP2.0 (ALADIN HUNEPS) gets near-surface perturbations only through the atmospheric forcing of the combination of the SV- and the EDA-based perturbations. It is hoped that the larger perturbations implied by the inclusion of the direct surface initial perturbations might improve the quality of the ensemble near to the surface. Additionally, it is noted that no spin-up effects are visible, therefore it is not necessary to exclude a spin up period from the investigations. Furthermore, the near-surface perturbations have also an impact at the higher atmosphere (upper-right panel of Fig. 7), which is weaker in terms of amplitude of the spread. The generated near-surface perturbations are growing reasonably during the assimilation cycle. The last plot (lower panel of Fig. 7) indicates the impact of the near-surface perturbations to the surface and to the

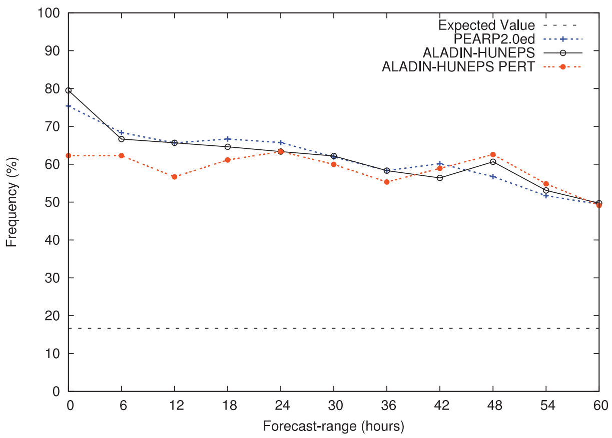

Fig. 8. Percentage of outliers of 2 m temperature for PEARP (dotted), for ALADIN HUNEPS (solid) and experimental with near-surface perturbations (dashed). Period: 1 January 2010 to 28 February 2010.

soil (at around 1.5 m under the surface) temperature fields. The values and the relative size of perturbations between the two levels seem also realistic. Therefore, the basic diagnostics of the surface perturbation scheme show that the surface perturbations have a true impact on the surface, the near-surface and even on the higher levels. The amplitudes and the relative sizes of these perturbations seem adequate and reasonable, therefore as the perturbation generation algorithm is behaving as expected, it seems worthy to probe deeper into the performance of the perturbed ensemble system.

Some usual ensemble verification scores were also computed in order to compare the reference (operational) and the experimental (with near-surface perturbation) ensemble systems. The percentage of outliers diagram was plotted and the near-surface characteristics were carefully compared (Fig. 8). Clearly, some slight improvements can be identified for the first 24 h of the integration (although the value of the curve is still far from the expected value). The comparison of the global PEARP2.0, the ALADIN-HUNEPS and the experimental systems shows small but substantial improvements and that the experimental systems based on the near-surface perturbations behave best and PEARP2.0 worst. Concurrently, at higher levels, the influence is slightly negative (not shown) indicating a too strong impact of the surface perturbation scheme. This result casts a shadow as the impact of near-surface perturbations is still too weak at the lower atmosphere, but already too strong at the mid-atmosphere. This might indicate that further improvements of the scheme at the surface should be made simultaneously with a proper tuning of the mesoscale downscaling model. The surface perturbation method might be efficiently combined with another, upper-air perturbation algorithm (for instance SV computations).

Improvements in terms of the spread–skill relationship are less clear. For the 2 m temperature (upper panels of Fig. 9) the spread and the RMSE of the experimental run are both slightly increased with respect to the operational one with a similar gap between the spread and the RMSE of the ensemble mean. Therefore, no detectable improvements can be identified on the base of this diagnostics. Looking at the 1000 hPa values (lower panels of Fig. 9) a telling improvement is seen up to the 24 h forecast range (which is in agreement with the results provided by the percentage of outliers). The main enhancements are coming from the reduced RMSE values of the ensemble mean at the beginning of the integration. For any case, the scheme, based on the presented ensemble verification scores of the near-surface perturbation method, can improve the quality of the operational ALADIN HUNEPS system.

4. Discussion

In this paper a limited area EPS for Hungary (called ALADIN HUNEPS) was examined. The backbone of this system rests on a dynamical downscaling of the global PEARP ARPEGE-based EPS. The downscaling was ensured by the mesoscale limited area model ALADIN. This version of the ALADIN HUNEPS is entirely determined by the initial and LBCs provided by the global system, therefore, the upgrade of PEARP (realised at the beginning of December 2009) had a major impact on the performance of the limited area ensemble system. The quality of ALADIN HUNEPS before and after the upgrade was carefully examined through standard EPS verification scores (spread–skill relationships, percentage of outliers, Talagrand diagrams). The impact of the new PEARP system is decidedly significant, with better EPS characteristics especially at the higher model levels. The comparison between the global (PEARP EPS) and the limited area (ALADIN HUNEPS) EPS systems revealed that some improvements near the surface could be gained, using the high-resolution ALADIN model, however, this headway fell short of expectations, making room for further upgrades. In order to address this deficiency, the computation of near-surface perturbations through perturbed surface assimilation cycles was validated and tested. This method is thought to be helpful to directly account for the weaknesses of the reference system near to the surface.

5. Conclusions

The first (preliminary) tests of the near-surface perturbations showed that although the implementation of the method brought some improvements, these were rather thin near the surface whereas some degradations was observed at higher levels. Results show that this method is potentially helpful, but it needs to be further and carefully tuned to boost improvements near the surface and prevent the negative impacts higher in the atmosphere. Among other leads, this method could be combined with other techniques (such as using SVs) to account for the upper-air uncertainties, together with a proper tuning of the ALADIN model (with special regard to the boundary layer mixing characteristics) itself in order to stave off the detrimental impact

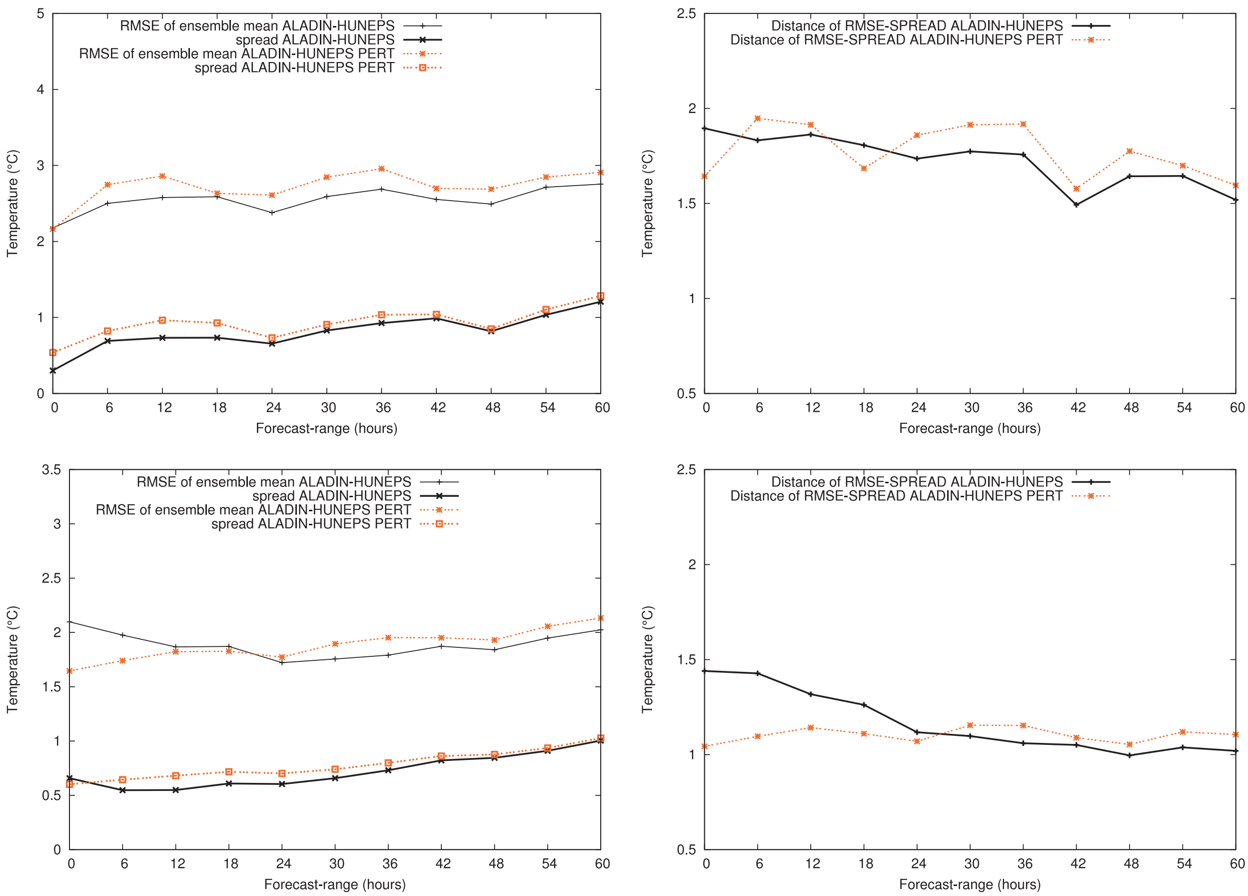

Fig. 9. Spread-skill relationship (left-hand side) and the difference between the root mean square error of the ensemble mean and the spread (right-hand side) of the operational (solid) and experimental (dashed) ensemble systems. In the left panels the lower (thick) lines indicate the spread and the upper (thin) ones the RMSE of the ensemble mean. The upper panels show the 2 m temperature scores and the lower ones the 1000 hPa temperature ones. Period: 1 January 2010 to 28 February 2010.

of the near-surface perturbations to the higher atmospheric levels.

Good results, using the initial perturbations obtained on the basis of the ALADIN SVs, should be brought forth in the near future, paving the way to explore the SV technique (such as to find the best optimization domain and time for mesoscale SVs, the possible inclusion of physical processes into the SV computations, the optimization of the scaling of SVs towards thorough studies of the impact of the singular-vector based perturbations into the ALADIN HUNEPS).

6. Acknowledgments

The authors are very grateful to the scientists of the international ALADIN limited area NWP Consortium for the development of the NWP tools (various configurations of the ALADIN model) used in the described investigations. We are particularly grateful to Laurent Descamps and Carole Labadie for their help and support on the PEARP (and its updates) matters. The reruns of the PEARP1.5 system would not be possible without the generous help of Carole and Laurent, therefore their help in this particular (and essential) aspect of our work is additionally emphasized. The work of Andrea Storto for the coding of the surface perturbation algorithm is highly acknowledged. The contribution of Edit Hágel for implementing the operational ALADIN HUNEPS in Hungary is also recognised and appreciated. The colleagues of the NWP team of the HMS are also thanked for their support in various practical aspects of the work. Among them, special thanks to Gergely Bölöni, who helped tremendously in the interpretation of the results with the ALADIN surface perturbation data assimilation cycles. The generous help of Jean Maziejewski improving the English of the manuscript is also warmly acknowledged. The comments of the two anonymous reviewers are also appreciated and acknowledged, which helped to significantly improve the manuscript.

References

Anderson, J. L. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Clim.* **9**, 1518–1530.

Berre, L., Desroziers, G., Raynaud, L., Montroty, R. and Gibier, F. 2009. Consistent operational variational assimilation. *Extended abstracts of the Fifth WMO International Symposium on Data Assimilation.* Melbourne, Australia, paper no. 196.

Bouttier, F. and Courtier, Ph. 1999. Data assimilation concepts and methods. Meteorological Training Course. Internal Report of ECMWF, 72.

Bouttier, F., Mahfouf, J-F. and Noilhan, J. 1993. Sequential assimilation of soil moisture from atmospheric low-level parameters, Part I: sensitivity and calibration studies. *J. Appl. Meteorol.* **32**, 1335–1351.

Bölöni, G. 2006. Development of a variational data assimilation system for a limited area model at the Hungarian Meteorological Service. *Időjárás* **110**, 309–327.

Buizza, R., Tribbia, J., Molteni, F. and Palmer, T. 1993. Computation of optimal unstable structures for a numerical weather prediction model. *Tellus* **45A**, 388–407.

Courtier, Ph., Freydier, C., Geleyn, J-F., Rabier, F. and Rochas, M. 1991. The ARPEGE Project at Météo-France. *ECMWF 1991 Seminar Proceedings: Numerical Methods in Atmospheric Models*, ECMWF Reading. United Kingdom. 193–231.

Descamps, L., Labadier, C., Joly, A. and Nicolau, J. 2009. Ensemble Prediction at Météo France (poster introduction by Olivier Riviere) 31st EWGLAM and 16th SRNWP meetings 28th September – 1st October, 2009. Available at: http://srnwp.met.hu/Annual_Meetings/2009/download/sept29/morning/posterpearp.pdf

Frogner, I-L. and Iversen, T. 2001. Targeted ensemble prediction for northern 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.

Giard, D. and Bazile, E. 2000. Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.* **128**, 997–1015.

Hágel, E. and Horányi, A. 2006. The development of a limited area ensemble prediction system at the Hungarian Meteorological Service: sensitivity experiments using global singular vectors, preliminary results. *Időjárás* **110**, 229–252.

Hágel, E. and Horányi, A. 2007. The ARPEGE/ALADIN limited area ensemble prediction system: the impact of global targeted singular vectors. *Meteorologische Zeitschrift* **16**(6), 653–663.

Heizenreder, D., Trepte, S. and Denhard, M. 2006. A regional multimodel ensemble in Europe. The European forecaster. *Newsletter of the WGCEF* **11**, 29–35.

Horányi, A., Kertész, S., Kullmann, L. and Radnóti, G. 2006. The ARPEGE/ALADIN mesoscale numerical modeling system and its application at the Hungarian Meteorological Service. *Időjárás* **110**, 203–227.

Jensen, M. H., Frogner, I-L., Iversen, T. and Vignes, O. 2006. Limited area ensemble forecasting in Norway using targeted EPS. *ECMWF Newsletter* **107**, 23–29.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T. 2005. The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Process. Geophysics* **12**, 527–536.

Montani, A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U., and co-authors. 2003. Operational limited area ensemble forecasts based on the 'Lokal Modell'. *ECMWF Newsletter* **98**, 2–7.

Szintai, B. and Ihász, I. 2006. The dynamical downscaling of ECMWF EPS products with the ALADIN mesoscale limited area model: preliminary evaluation. *Időjárás* **110**, 253–277.

Talagrand, O., Vautard, R. and Strauss, B. 1998. Evaluation of probabilistic prediction systems. *Proc. of ECMWF Workshop on Predictability.* 20–22 October 1997, 1–25, ECMWF, Shinfield Park, Reading, U.K.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Ivatek-Sahdan, S., and co-authors. 2009. The Central-European limited area ensemble forecasting system: ALADIN-LAEF. LACE Report. Available at: http://www.rclace.eu