Development and operational application of a short-range ensemble prediction system based on the ALADIN limited area model

Edit HÁGEL

Ph.D. Thesis

Eötvös Loránd University, Faculty of Science

Ph.D. School for Earth Sciences, Geography-Meteorology Program

School leader: Prof. Dr. Miklós Monostori, doctor of HAS

Program leader: Prof. Dr. Gyula Gábris, doctor of HAS

Thesis advisors:

Dr. András Horányi

Assoc. Prof. Dr. István Matyasovszky

Hungarian Meteorological Service

Numerical Modelling and Climate Dynamics Division

2009

Szüleimnek,
akik tanulmányaim során mindenkor segítettek és támogattak.

To my parents,
who have always helped and supported me throughout my studies.

Contents

List of abbreviations

3D-Var	3-Dimensional VARiational data assimilation
4D-Var	4-Dimensional VARiational data assimilation
AEMET	Agencia Estatal de METeorología
ALADIN	Aire Limitée Adaptation dynamique Développement International
ARPEGE	Action de Recherche Petite Echelle Grande Echelle (i.e. Research Project on Small and Large Scales)
CAPE	Convective Available Potential Energy
CFL	Courant-Friedrichs-Lewy
COSMO	COnsortium for Small-scale MOdeling
COSMO-LEPS	COnsortium for Small-scale MOdeling-Limited-area Ensemble Prediction System
ECMWF	European Centre for Medium-range Weather Forecasts
ENIAC	Electronic Numerical Integrator And Computer
EPS	Ensemble Prediction System
FA	Fichier Arpege (i.e. Arpege File)
GLAMEPS	Grand Limited Area Model Ensemble Prediction System

HAWK	Hungarian advanced WorKstation
HIRLAM	HIgh Resolution Limited Area Model
HMS	Hungarian Meteorological Service
IC	Initial Condition
LACE	Limited Area modelling in Central Europe
LAEF	Limited Area Ensemble Forecasting
LAM	Limited Area Model
LAMEPS	Limited Area Model Ensemble Prediction System
LBC	Lateral Boundary Condition
NCEP	National Centers for Environmental Prediction
NetCDF	Network Common Data Form
NWP	Numerical Weather Prediction
PEACE	Prévision d'Ensemble A Courte Échéance
PEARP	Prévision d'Ensemble ARPege
RMSE	Root Mean Squared Error
ROC	Relative Operating Characteristics
SRNWP-PEPS	Short Range Numerical Weather Prediction-Poor man's Ensemble Prediction System
SV	Singular Vector
TEPS	Targeted Ensemble Prediction System
UTC	Universal Time, Coordinated

Introduction

Making a weather forecast requires the use of mathematical models in order to predict the future state of the atmosphere. Forecasts are made by solving a set of partial differential equations, the so-called primitive equations. These equations are nonlinear and are impossible to solve analitically. Because of the nonlinear nature of the equations, the solution is highly dependent on the accuracy of the initial conditions. The problem is that the true state of the atmosphere cannot be known exactly. The reason of this is that the number of the observations is limited (smaller than the degrees of freedom in the models), their spread is uneven around the globe, there are inevitable observation errors, and also errors in the data assimilation techniques and in the models themselves. As a result, there will always be some uncertainty in the initial conditions of the numerical weather prediction models.

One possible solution of the above mentioned problem is to run a set, or as usually called, an *ensemble* of forecasts, each starting from a slightly different initial condition, thus they are equally likely realizations of the "true" atmospheric state. The advantage of this method is clear: the spread of the ensemble members can provide useful information on the predictability of the atmospheric state, and a probability value can be assigned to different weather events. Since its first operational application in 1992, ensemble forecasting has become a widely used technique by many meteorological services around the world. Despite its obvious benefits, it was used only on global scales and in the medium-range for a long time. In the last couple of years intensive research has started to apply the ensemble method in short-range limited area foracasting as well. The first real-time, operational regional ensemble prediction system was implemented at NCEP in 2001. To gen-

erate initial condition perturbations the breeding method was used, just like in the global ensemble system of NCEP ([44]). Research is also active in Europe with the application of different methods at different centres. At the Spanish Meteorological Service (AEMET) the multi-model multi-analysis multi-boundary method is used ([9]). The multi-model method is also used for the SRNWP-PEPS system ([20]). The COSMO consortium is running an ensemble system based on the downscaling of the representative members of ECMWF EPS ([34]). These representative members are selected by using a clustering technique. It was found that COSMO-LEPS is more skilful (in terms of Brier skill score and ROC area) in correctly forecasting high precipitation values over a larger area than the lower resolution ECMWF ensemble system. Yet another technique is used at the Norwegian Meteorological Service. They use TEPS to provide boundary conditions for the limited area model runs. TEPS is based on ECMWF EPS but singular vectors are targeted to have the greatest impact in the Northern European region. The system has 20+1 members in contrast to the 50+1 members of the original EPS ([24]). TEPS members are downscaled with the HIRLAM limited area model to form an ensemble system. It was found that combining TEPS and HIRLAM EPS into NORLAMEPS the new system is more skilful in predicting precipitation than the two individual ones, and also better than ECMWF EPS.

Motivated by the results of these experiments, research started at the Hungarian Meteorological Service with the final aim to establish an operational LAMEPS system for the Central European area and to see how it can improve the predictions of the existing global systems. It was decided to start the experiments with the dynamical downscaling of global ensemble forecasts. Two possible choices were considered: the downscaling of ARPEGE ensemble forecasts and the downscaling of ECMWF EPS members (for further information on ECMWF EPS downscaling see [42] about the work of Szintai and Ihász).

The aim of the PhD work is the development, investigation, and finally the operational application of a short-range limited area ensemble prediction system based on the ALADIN model, using ARPEGE EPS forecasts as initial and lateral boundary conditions. As a first step, sensitivity experiments were performed in order to

investigate whether or not it is possible to optimize the existing ARPEGE based global ensemble prediction system (PEARP) for Central Europe by changing the optimization area and optimization time used for the global singular vector computations. With this purpose several different optimization areas and times were defined and tested through case studies and longer test periods.

As a second step of the work, research has started with the computation of ALADIN singular vectors with the aim of perturbing the initial conditions of the LAMEPS locally. It is believed that by applying local perturbations, the initial uncertainties can be better addressed. As the scale of the perturbations will be more similar to the scale of those weather phenomena that are the most important in short-range limited area forecasting, it is expected that the skill of the system will improve.

Meanwhile, in order to gain experience on a day-to-day, real-time basis, a short-range limited area ensemble system - based on the ALADIN model - was put into operations at HMS. At present, the only operationally feasible solution is the direct downscaling of the PEARP members, therefore this method is used. The system has been running on a daily basis in quasi-operational status since February, 2008 and it is going to be developed and improved continuously, using results of the ongoing researches.

The thesis is organized as follows. In Chapter 1 a short introduction is given about numerical weather prediction in general and ensemble forecasting in particular. The singular vector method is discussed in more detail. In Chapter 2 the applied NWP models, ARPEGE and ALADIN are presented together with the ARPEGE ensemble system PEARP. Chapter 3 is dedicated to the results of the sensitivity studies with global singular vectors. The process of finding the most optimal optimization area and time is described. Verification results of case studies and longer test periods are analysed and the performance of the global and the limited area systems are compared to each other. In Chapter 4 the quasi-operational short-range limited area ensemble prediction system of HMS is presented. In Section 4.1 the characteristics of the system are briefly described, followed by a case study (Section 4.2) and verification results for a longer period (Section 4.3). Chapter 5 is

dedicated to the preliminary results of the ALADIN singular vector experiments. Finally, in the Appendix the most common verification and visualisation methods of ensemble forecasts are discussed and a detailed description is given about the quasi-operational LAMEPS of HMS.

Chapter 1

Numerical weather prediction

These days making a weather forecast would not be conceivable without the use of NWP models. All the experiments and results presented in the thesis were performed and achieved using NWP models. For this reason it is important to give a short introduction about numerical weather prediction. In Section 1.1 a brief overview is given about the history of NWP, followed by the description of present day models and the forecast process in Sections 1.2 and 1.3, respectively. As it will be shown later, NWP is based on the solution of a set of nonlinear partial differential equations, which is highly dependent on the accuracy of the initial conditions. Section 1.4 is dedicated to the topic of sensitivity to initial conditions. In Section 1.5 the ensemble technique is introduced and some of the possible methods to create an ensemble system are presented.

1.1 Historical background

Present day meteorological models - such as the ones used for the experiments presented in the thesis - are the results of more than 50 years of continuous development in the field of numerical weather prediction. The basic idea behind NWP is to solve the fundamental equations of hydrodynamics and thermodynamics (numerically), subject to the observed data, in order to describe the future state of the atmosphere. This idea was first suggested in 1904 by a Norwegian scientist, Vilhelm Bjerknes. He maintained that the basic equations governing the atmosphere

were known, but he was aware that a lot of preparatory work (both theoretical and practical) was needed to reach the final goal.

First attempt to perform numerical weather prediction was made by British scientist Lewis Fry Richardson in the 1910's ([39]). Richardson calculated the changes in pressure and the wind at two points and made all the calculations by hand. Unfortunately his forecast failed dramatically: he calculated a change of pressure of 145 hPa in 6 hours. His failure had two main causes: (i) the delicate dynamic balance that exists in the atmosphere between the pressure and the wind fields was not reflected in the initial conditions he used ([32]), and (ii) the applied time step (3 hours) did not fulfil the CFL criterion[1].

For a long time no one followed Richardson's footsteps. After the second world war, two important technological developments made NWP forecasts possible: the establishment of a hemispheric network of upper-air measurements and the development of the first electronic computers. Scientists of that time realized that successful predictions could only be achieved using a simplified set of equations, the so-called filtered equations. These filtered equations did not describe all possible forms of atmospheric motions, only the large scale dynamics of the atmosphere. The first successful numerical weather prediction was achieved in 1950 on the ENIAC[2] computer, using a model based on the filtered equations. Only by the development of more powerful computers could meteorologists return to the so-called primitive equations. The first global model based on these equations was put into operations in 1966. It had a 300 km grid and 6 vertical levels. This model had great similarities with the one used by Richardson almost 50 years earlier.

[1]The CFL criterion - discovered by Courant, Friedrichs and Lewy in 1928 - is a stability criterion which sais that the time step (Δt) to be applied in a model has to be smaller or equal than the ratio of the spatial resolution (Δx) and the phase speed (c) of the fastest propagating wave in the model:

$$\Delta t \leq \frac{\Delta x}{c} \tag{1.1}$$

Thus, if the model resolution increases, the time step must decrease, increasing the computational demand.

[2]ENIAC: An electronic, digital, general-purpose, programmable computer, completed in 1946.

1.2 Present day models

Present day models are based on the primitive equations, i.e. on the mathematical formulations of the conservation laws. These are:

- the conservation of the three dimensional momentum (equations of motion),
- the conservation of mass for dry air (continuity equation),
- the conservation of energy (first law of thermodynamics),
- the conservation of moisture.

These are called prognostic equations as they describe the dynamic change of the variables over a short time interval. Another equation, a so-called diagnostic equation is also needed. This is the equation of state for perfect gases (the gas law) that gives the relation between pressure, density and temperature. In several models the hydrostatic approximation is also used[3]. This approximation is valid only for horizontal scales finer than about 10 km.

The above mentioned equations form a set of nonlinear partial differential equations which does not have an analitical solution and can only be solved numerically. When solving it, different numerical approximations are used, the continuous equations are discretized in space (both horizontally and vertically) and time. Computation of the derivatives can be done by using finite differences or by applying the spectral technique ([38]).

1.2.1 Global and limited area models

Depending on the area covered, two types of models can be distinguished:

- global models that provide forecasts for the whole Earth,
- limited area models providing forecasts for only part of the Earth's surface.

[3]In the hydrostatic approximation the vertical acceleration is neglected in the vertical momentum equation, leading to the so-called hydrostatic equation.

As regards weather forecasting (and not talking about seasonal or climate predictions) global models are mainly used for making medium-range or even longer, i.e. monthly forecasts. Their forecast range goes typically up to 15 or even 30 days.

Limited area models are covering only part of the Earth's surface therefore they require boundary conditions not only at the upper and lower boundaries but also at the lateral boundaries of the limited area domain. These lateral boundary conditions are provided by a global model or another limited area model covering a larger domain. In NWP, limited area models are mainly used for short-range forecasting, typically up to day 3. The main reason why the lead time is shorter for limited area models than for global models is that after a certain time (depending on e.g. the integration domain of the LAM) the effect of the boundary conditions becomes dominant and determines the quality of the forecast. As limited area models cover a significantly smaller domain than the global models, given the same computer resources, they can be run on higher resolution than their global counterparts.

In recent years intensive effort was made to run models (mainly limited area ones) on very high, few kilometers resolution. In these scales the hydrostatic approximation does not hold any more, therefore these models have to be non-hydrostatic ([28]). As the resolution of these models is very high, their integration domain is usually very small and they are mainly used in ultra short-range forecasting (typically up to 1 day, as it would be computationally too demanding to run them on a larger domain and/or for longer lead time).

1.3 The forecast process

The different steps of making a numerical weather forecast are the following:

- data assimilation to define the initial conditions,
- initialization to restore the balance between the wind and the pressure fields in the initial conditions,
- integration of the model,

- post-processing of the raw model outputs.

In the following the different steps are going to be described.

1.3.1 Data assimilation

The aim of data assimilation is to prepare the best possible estimate of the present atmospheric state to be used as initial condition for the model integration. One of the problems one has to face when performing data assimilation is that the number of available observations (SYNOP and TEMP data, aircraft and satellite observations, etc.) is less than the degrees of freedom of the model (the number of the model variables $\times$ number of levels $\times$ number of grid points). In addition the distribution of the observations is not uniform in space and time (Fig. 1.1, Fig. 1.2 and Fig. 1.3.). For these reasons additional information is needed. This additional information is a short-range forecast called *first guess* or *background*. The advantage of using information from a forecast is that it contains the same number of data as the degrees of freedom of the model and this data is already on the model grid. The disadvantage is that it cannot be fully exact as it is subject to different types of errors, such as forecast error, analysis error, etc. (Certainly, the observations have their own errors as well.) The aim is to bring the first guess as close as possible to the real atmospheric state with the help of the observed data. However, the real atmospheric state can never be known exactly, therefore the purpose is to determine a balanced initial condition (see Section 1.3.2) from which the best possible predictions can be made. To reach this goal different methods can be applied, e.g. optimal interpolation or variational assimilation ([26]).

It is important to note that the quality of the initial conditions is very important as the set of partial differential equations to be solved during the model integration is very sensitive to the accuracy of the initial conditions (to be discussed later in Section 1.4).

1.3.2 Initialization

One of the main reasons of the failure of Richardson's attempt to perform numerical weather prediction was the missing balance between the initial wind and

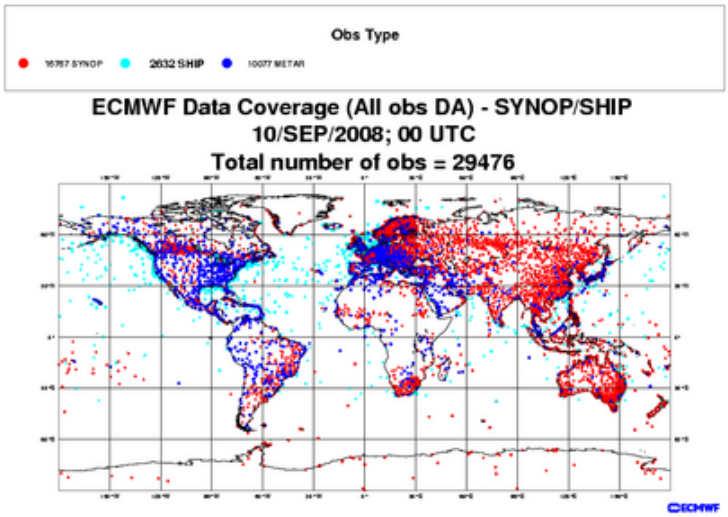

Figure 1.1: Data coverage - SYNOP/SHIP. 10 September 2008, 00 UTC. Total number of observations for the whole Earth: 29476. (Source of figure: ww.ecmwf.int)

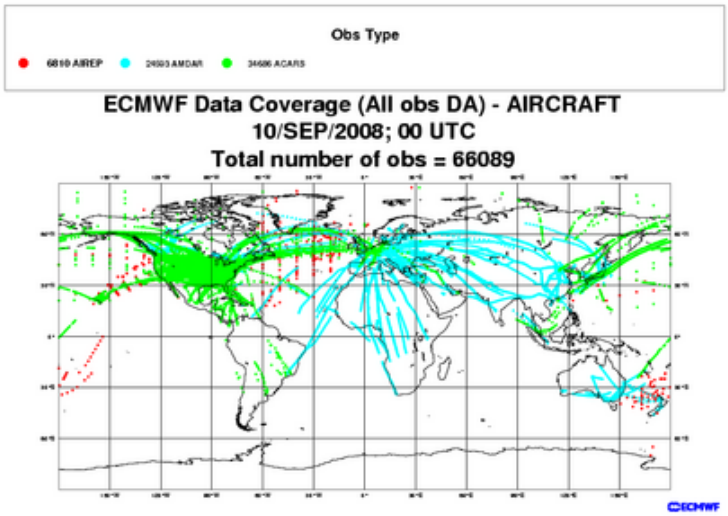

Figure 1.2: Data coverage - Aircraft. 10 September 2008, 00 UTC. Total number of observations for the whole Earth: 66089. (Source of figure: ww.ecmwf.int)

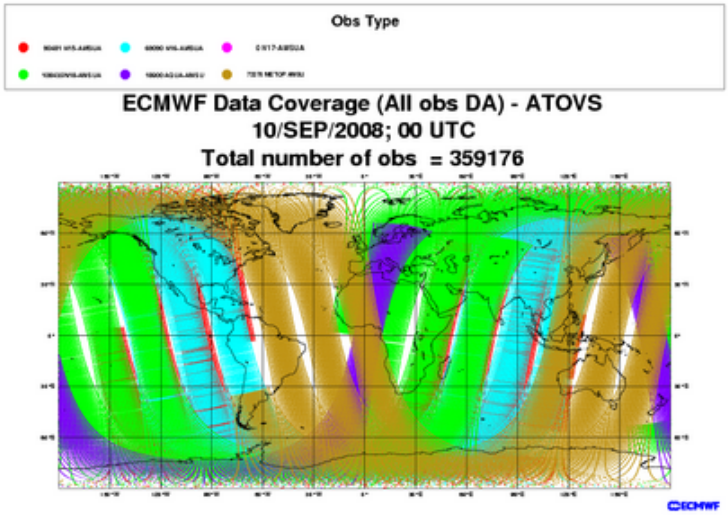

Figure 1.3: Data coverage - Satellite data (ATOVS - Advanced TIROS Operational Vertical Sounder). 10 September 2008, 00 UTC. Total number of observations for the whole Earth: 359176. (Source of figure: ww.ecmwf.int)

pressure fields. Several decades later, Peter Lynch repeated Richardson's experiment using the so-called *digital filter initialization* in order to restore this missing balance in the initial conditions. Applying initialization and using a time step that fulfiled the CFL criterion he got realistic results ([32]).

The primitive equations used in the NWP models have different type of solutions: (i) meteorologically important low frequency motions (with phase speeds of the order of ten m/s) and (ii) high frequency gravity waves (with phase speeds of hundreds of m/s). In the atmosphere a delicate balance exists between the wind and the pressure fields, ensuring that the gravity waves have much smaller amplitude than the low frequency, meteorologically important component. However, due to the imbalances in the initial data, during the integration of the model the amplitude of the gravity waves can be unrealistically large. This can lead to various problems, like e.g. the unrealistic results of Richardson's experiment. The problem can be handled through a process known as *initialization*, the aim of which is to define the initial conditions in such a way that the amplitude of the gravity waves remains small throughout the forecast ([31]).

1.3.3 Integration

Integration of a numerical weather prediction model involves solving the primitive equations in every time step throughout the whole forecast interval. However, the prognostic equations used in the NWP models do not describe all type of processes in the atmosphere: (i) there are subgrid-scale processes (e.g. convection), or (ii) processes that would be far too difficult to describe in an exact way (e.g. microphysics, radiation). These processes must be parameterized. Physical parameterization is also called in every time step during the integration.

1.3.4 Post-processing

A post-processing is performed on the raw model outputs in order to support the application of the model results by the forecasters, or by the end-users. Post-processing can consist of various steps. Special, meteorologically important diagnostic parameters can be derived from the model variables. Raw model outputs might also be transformed from one representation to the other. This may include horizontal and/or vertical transformations (e.g. change of projection, change of vertical coordinate system). Calibration of raw model outputs is also a possible part of post-processing.

1.4 Sensitivity to initial conditions

As it was mentioned before NWP is based on the solution of a set of nonlinear partial differential equations. Because of nonlinearity, this solution is highly dependent on the accuracy of the initial conditions. Before moving to the discussion of such complex systems as the atmosphere, let us first demonstrate the effect of nonlinearity in simple systems.

1.4.1 The effect of nonlinearity in simple systems

The effect of nonlinearity can be demonstrated through a very simple example. Let x_0 denote the initial value chosen from the [-2,2] interval. Take the square of x_0 and subtract 2 to get x_1 (i.e. the value in the next step). Continue this to get

a longer series of the x_i values. The process can be described with the following formula[4]:

$$x_{n+1} = x_n^2 - 2 \tag{1.2}$$

The system is deterministic, i.e. by chosing exactly the same x_0 we get the same result every time.

Let us consider two cases, where the initial conditions differ only slightly from each other. In the first case $x_0 = 0.4000$, while in the second case $x_0 = 0.4001$. Fig. 1.4 shows the effect of this rather small difference (0.0001). In the first ten steps the difference is small, the two curves are going close to each other, but after this initial phase, they start to differ significantly. The reason of this is the nonlinearity of the equation.

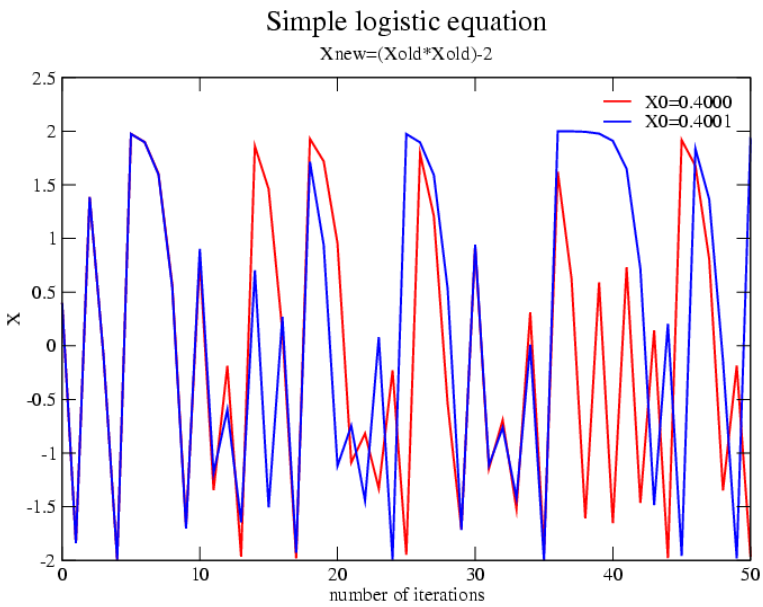

Figure 1.4: Simple example to demonstrate the sensitivity of nonlinear systems to the initial conditions. The difference of the two initial conditions is only 0.0001, however, after a few steps the two curves start to behave significantly different. The reason of this is the nonlinearity of the equation.

[4]The equation used in this example is the simplified form of the logistic equation used in the modelling of population growth.

The Lorenz model

Another well known example of the sensitivity to initial conditions is based on the Lorenz model[5] ([29]). The variables X, Y and Z are determined by the following equations:

$$\dot{X} = -\sigma X + \sigma Y \tag{1.3}$$

$$\dot{Y} = -XZ + rX - Y \tag{1.4}$$

$$\dot{Z} = XY - bZ \tag{1.5}$$

where σ, r and b are constants. In the experiments described in the paper of Lorenz ([29]) the following values were used: $\sigma = 10$, $r = 28$ and $b = 8/3$.

Fig. 1.5 shows the so-called Lorenz attractor (also called butterfly attractor). The initial circles represent initial conditions that differ only slightly from each other, while the arrows show the evolution of these points in time. If we are in a predictable state (top panel of Fig. 1.5), small differences (or errors) in the initial condition will not affect the forecast and the points remain close to each other. If we are in a less predictable state (bottom left panel of Fig. 1.5) the points only stay together for a limited time, while in the worst case, in the unpredictable state (bottom right panel of Fig. 1.5) they start to diverge almost immediately.

This example clearly demonstrates that nonlinear systems show large sensitivity to the initial conditions and this sensitivity depends on the initial state itself (see the three different cases in Fig. 1.5)

Over the years the Lorenz model and in particular the so-called butterfly attractor (Fig. 1.5) have become the symbols of chaos.

1.4.2 The chaotic nature of the atmosphere

Strictly speaking chaos is defined as the complicated behaviour of simple systems, i.e. of systems with only a few degrees of freedom ([10]) just like the second example in the previous section, the Lorenz model. The atmosphere is not by far a simple system, it is a complex system with very large degrees of freedom.

[5]The Lorenz model describes the convection motion of a fluid in a small, idealized Rayleigh-Bénard cell.

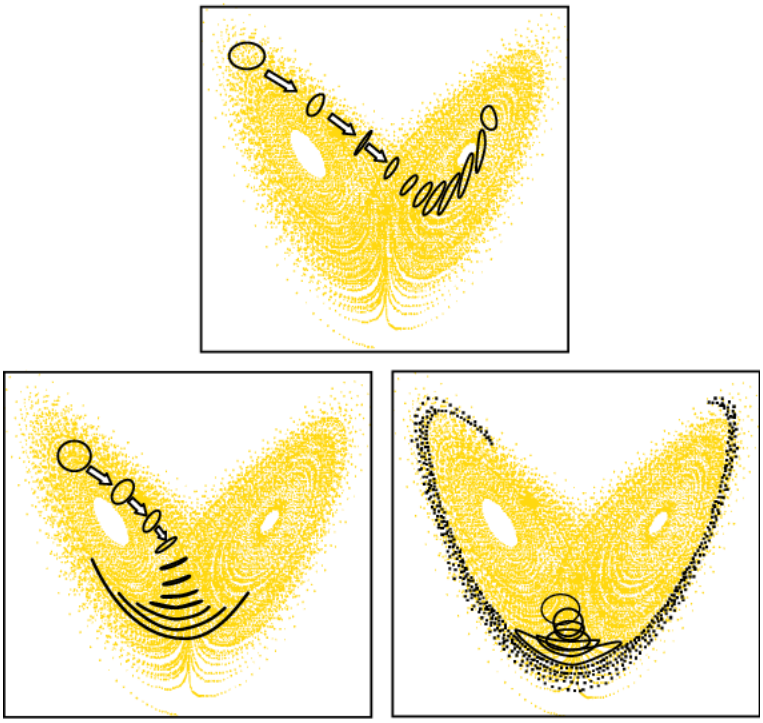

Figure 1.5: The Lorenz attractor (also called butterfly attractor) in the phase space of the system. Top panel shows the highly predictable state, when small differences (or errors) in the initial condition do not have a significant effect on the forecast. Bottom left panel shows a state with limited predictability, when the errors in the initial conditions lead to very different states after a certain time. Bottom right panel shows the unpredictable case, when the points start to diverge almost immediately. (Source of figure: www.ecmwf.int)

However, there is a strong link between chaos and meteorology. Chaos was "discovered" by Edward Lorenz, a meteorologist, while performing computer examinations of a relatively simple model of weather. Over the years the so-called butterfly attractor (Fig. 1.5) has become the symbol of chaos. The question "Does the flap of a butterfly's wing in Brazil set off a tornado in Texas?"[6] has become well known and the atmosphere is often mentioned as a chaotic system. Is it a proper statement? Can the atmosphere really be regarded as chaotic?

In his book "The Essence of CHAOS" ([30]) Lorenz defines chaos as *"the*

[6]Title of a presentation by Lorenz from 1972. In the presentation he avoided answering the question and noted that if a single flap could lead to a tornado, it could equally well prevent one ([30]).

property that characterizes a dynamical system in which most orbits exhibit sensitive dependence". In the same book Chapter 3 bears the title: *"Our Chaotic Weather"*.

Although the atmosphere is not a simple system, its basic large scale dynamical features can be described using simple models with degrees of freedom around 10. Also, there are studies showing that locally the atmosphere can behave as a low dimensional system ([43]). Because of these, and because of the fact that the atmosphere shows the two main features of chaotic systems - irregularity and sensitivity to initial conditions - it can be studied with the methods and theories developed for chaotic systems and it can be said that the atmosphere exhibits a chaotic behaviour.

The simple examples in the previous section, especially the example of the Lorenz model provided an insight into the effect of nonlinearity and into the effect of the sensitivity to initial conditions. The models used for NWP purposes are far more complex and in case of the atmosphere the exact determination of the initial condition can never be possible. Therefore it is not enough to forecast the future state of the atmosphere, one also has to predict the uncertainty related to this forecast. This finding led to the introduction of *probabilistic forecasts* and the *ensemble technique*.

1.5 Ensemble prediction

The original idea of the ensemble method - as it was put into operations in 1992 at NCEP ([44]) and ECMWF ([5], [37]) - can be described as follows: one may choose to integrate the NWP model not only once, but starting from several - slightly different - initial conditions. The difference between these initial conditions should have the same order of magnitude as the overall errors in the data assimilation process (analysis errors). It is considered that the ensemble of initial conditions would comprise the true state of the atmosphere. The model is then integrated from these different initial conditions. The advantage of the ensemble method is clear: it provides useful information on the predictability of the atmospheric state (the larger the spread of the ensemble members, the smaller the predictability, and vice versa) and also on the probability of the occurrence of

different weather events. Since its first operational application in 1992 ensemble forecasting has become a widely used technique by many meteorological services around the world. Despite its obvious benefits the ensemble technique was used only on global scales and in the medium-range for a long time.

In the last couple of years intensive research has started to apply the ensemble method in short-range limited area forecasting as well (see e.g. [24], [9], [34], [20]). The first real-time, operational regional ensemble prediction system was implemented at NCEP in 2001. In most of the references about limited area ensemble prediction system experiments one can find that in general it is rather difficult to achieve improvements with the limited area system with respect to the global one (which is providing the lateral boundary conditions). On the other hand, there are promising aspects as well. Several authors have found that in case of extreme events, improvements can be achieved.

While initially the ensemble method was used to represent initial condition errors, nowadays it is used to simulate other sources of errors as well. In the following we present some of the possible techniques for creating an ensemble system, depending on what type of uncertainty we wish to account for.

1.5.1 Methods for ensembles

Several types of errors can be distinguished that affect the quality of numerical weather forecasts. Most often these are separated into two different categories: initial condition errors and model errors. In case of limited area models there is an additional source: the errors related to the lateral boundary conditions. Initial condition errors and model errors are linked to one another: model error at a given time will lead to initial condition error for the next model run (Fig. 1.6). Since the initial conditions of the models always differ from the true state of the atmosphere, errors caused by these imperfections would appear even if the models were perfect. Because of the nonlinearity of the equations, even small errors in the initial conditions can lead to large forecast errors.

Depending on what type of uncertainty we wish to account for, different methods can be used for making ensemble forecasts. In the following these methods will be briefly presented.

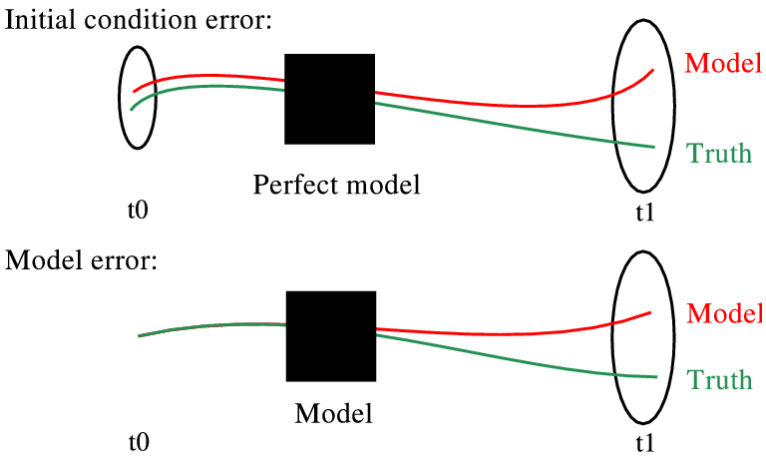

Figure 1.6: Schematic view of possible forecast errors. Initial condition errors and model errors are linked to one another: model error at time t_0 will lead to initial condition error at time t_1. Since the initial conditions always differ from the true atmospheric state, errors would appear even if the models were perfect.

Initial condition perturbations

The ensemble approach was first applied at ECMWF and NCEP. In both cases perturbations were added to and subtracted from the analysis in order to take into account the errors in the initial conditions. At ECMWF the singular vector technique was applied, while at NCEP the breeding method was used.

At ECMWF the aim of the singular vector approach is to find the fastest growing perturbations to a given initial state that have the maximal growth over a 48 hours period ([5], to be discussed in more details in Section 1.5.3).

At NCEP a different approach was used when operational ensemble forecasting started, the so-called breeding method. The main idea behind breeding is the following. As a first step initial conditions are randomly perturbed, then forecast is made from all these perturbed initial conditions. With a certain frequency (typically 6 hours) perturbations are re-scaled to their original size (Fig. 1.7), the actual (unperturbed) analysis is modified with these perturbations and the process continues. After 4-5 days this method leads to the selection (breeding) of fastest growing perturbations ([44]).

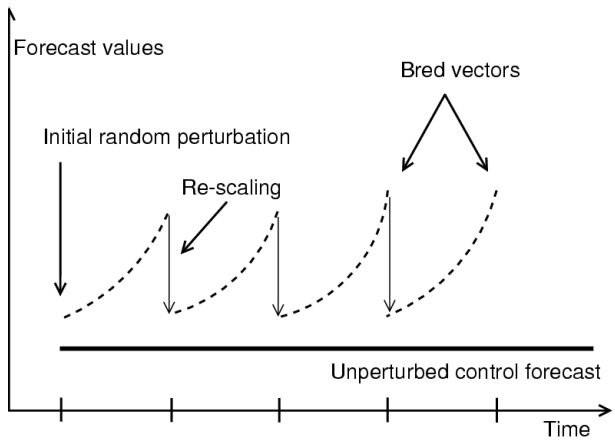

Figure 1.7: Schematic of a breeding cycle. As a first step initial conditions are randomly perturbed, then forecast is made from all these perturbed initial conditions. Perturbations are re-scaled in regular intervals (typically 6 hours). After 4-5 days the method leads to the selection (breeding) of the most unstable perturbations, the so-called breeding vectors.

Multi-analysis, multi-model method

Another way to represent the errors in the initial conditions is the use of several different analyses. When applying the multi-analysis method the ensemble is generated using different analyses made by e.g. different data assimilation techniques, or more often, made by different numerical weather prediction models. This technique is often combined with the multi-model and (in case of limited area models) multi-boundary approach.

In case of a multi-model system an ensemble of forecasts is generated with the use of different numerical weather prediction models. The different models may use different numerical schemes, different initial and/or boundary conditions, different parameterization packages, etc. which makes it possible to account for several different types of errors successfully.

As an example of multi-model, multi-analysis, multi-boundary system the short-range limited area ensemble prediction system of AEMET (the Spanish Meteorological Service) can be mentioned ([9]). The system is made up of five different limited area models and the multi-model technique is combined with the multi-analysis, multi-boundary approach: ICs and LBCs are coming from four

different global models, then each limited area model is coupled with all global models resulting a system of 20 members.

Multi-parameter, multi-parameterization method

To account for the uncertainties in the physical parameterizations one can apply the multi-parameter method. In this case the ensemble is built from forecasts integrated with the same model but using different parameters for a given parameterization scheme. This approach is mainly used in combination with other methods.

The multi-parameterization method is also a way to account for the uncertainties in the physical parameterizations. Instead of using the same parameterization package, each member of the ensemble is integrated using different parameterization schemes. As an example the ALADIN-LAEF system can be mentioned ([45]).

Stochastic physics

Another way to account for the errors caused by the physical parameterization schemes is the use of stochastic physics. In this case a random element is included in the model integration, therefore the system becomes non-deterministic. A stochastic physics scheme is used operationally at ECMWF in the ensemble prediction system ([4]). In the current implementation the tendencies of parameterized physical processes are perturbed using random numbers. The development of a more sophisticated stochastic physics scheme is underway at ECMWF.

1.5.2 Limited area ensemble systems

In case of limited area models lateral boundary conditions are required. It can be shown that during the initial part of the forecast the effect of the initial condition is dominant, while in the latter part the effect of lateral boundary conditions takes over the dominance. In case of a limited area ensemble prediction system based on a single model it is important to use not only different initial conditions, but also different lateral boundary conditions for each ensemble member. Without

perturbing the lateral boundary conditions, the members would become very similar (after an initial time) due to the dominance of the boundary conditions and the system would not have sufficient spread throughout the whole forecast range (see [8], or Section 3.1).

1.5.3 The singular vector technique

Our experiments were based on the computation of singular vectors therefore this technique is discussed in more detail. The aim of the singular vector approach is to find the fastest growing perturbations to a given initial state that have the maximal growth over a given $[t_0, t]$ time interval (the optimization time) and a given (geographical) domain (the optimization area). In order to understand better this technique, the singular vectors will be discussed in more detail. First the theoretical background is presented, followed by some details of the practical implementation.

Theoretical background

Let X be the state vector in the phase space of the system. The system of nonlinear primitive equations describing the evolution in time can be written formally in the following way:

$$\frac{dX}{dt} = A(X) \tag{1.6}$$

where A is the nonlinear model operator. Let x be a perturbation of the state vector X. Then the following can be written:

$$\frac{d(X+x)}{dt} = A(X+x) \tag{1.7}$$

If x is small, the Taylor expansion of $A(X+x)$ can be applied in the vicinity of X. Therefore one can write

$$A(X+x) \approx A(X) + A_l(x) \tag{1.8}$$

where $A_l = \frac{\partial A}{\partial X}|_{X(t)}$ is the linear operator that corresponds to the model operator A. From Eq. 1.7 and Eq. 1.8 one can write the linearized evolution equation for the

perturbation x:

$$\frac{dx}{dt} = A_l(x) \tag{1.9}$$

Integrating Eq. 1.9 from t_0 to t yields

$$x(t) = Lx(t_0) \tag{1.10}$$

where L stands for the tangent linear operator integrated from t_0 to t. Let <.,.> be an inner product and let us define the associated norm:

$$\|x\|^2 =< x,x > \tag{1.11}$$

Then the norm of the perturbation at time t_0 is

$$\|x(t_0)\| = \sqrt{< x(t_0),x(t_0) >} \tag{1.12}$$

and at time t the norm is

$$\|x(t)\| = \sqrt{< Lx(t_0),Lx(t_0) >} \tag{1.13}$$

Therefore the amplification of the perturbation from t_0 to t is

$$\frac{\|x(t)\|}{\|x(t_0)\|} = \sqrt{\frac{< Lx(t_0),Lx(t_0) >}{< x(t_0),x(t_0) >}} \tag{1.14}$$

and the vectors (perturbations) we are looking for are those that maximize this amplification in Eq. 1.14.

Let L^* denote the adjoint of the operator L with respect to the inner product <.,.>. By definition the following can be written:

$$< Ly,z >=< y,L^*z > \; and \; < y,Lz >=< L^*y,z > \tag{1.15}$$

Using Eq. 1.14 and Eq. 1.15 the amplification of the perturbation from t_0 to t can be written

$$\frac{\|x(t)\|}{\|x(t_0)\|} = \sqrt{\frac{< L^*Lx(t_0),x(t_0) >}{< x(t_0),x(t_0) >}} \tag{1.16}$$

By definition the square roots of the eigenvalues of L^*L are called the singular values of L, while the eigenvectors of L^*L are called the singular vectors of L. In our case L is real and the singular vectors are orthogonal ([5]). If the values σ_i^2 are the eigenvalues of L^*L and the vectors v_i are the eigenvectors of L^*L then the following holds:

$$L^*L = V\Sigma V^T \tag{1.17}$$

where Σ is a diagonal matrix with the values σ_i^2 on the diagonal and V is a matrix with columns defined by the eigenvectors of L^*L. Then the norm of a singular vector v_i at time t is given as

$$\|v_i(t)\|^2 =< L^*Lv_i(t_0), v_i(t_0) >= \sigma_i^2 \|v_i(t_0)\|^2 \tag{1.18}$$

and the amplification from t_0 to t is

$$\sqrt{\frac{< L^*Lv_i(t_0), v_i(t_0) >}{< v_i(t_0), v_i(t_0) >}} = \sqrt{\frac{< \sigma^2 v_i(t_0), v_i(t_0) >}{< v_i(t_0), v_i(t_0) >}} = |\sigma_i| \tag{1.19}$$

This amplification is maximal for the leading eigenvalue σ_1^2 (the eigenvalues are sorted, with σ_1^2 being the largest), associated to the singular vector v_1:

$$max\left\{ \frac{< L^*Lx(t_0), x(t_0) >}{< x(t_0), x(t_0) >} \;\middle|\; x(t_0) \neq 0 \right\} = \frac{< L^*Lv_1(t_0), v_1(t_0) >}{< v_1(t_0), v_1(t_0) >} = \sigma_1^2 \tag{1.20}$$

Thus the computation of the perturbations with the fastest growth is reduced to an eigenvalue problem which is solved using an iterative method. To solve this problem several assumptions and choices are needed.

Practical implementation

In this section the applied assumptions and the possible choices are highlighted.

- **Question of norms:** During the computation of the singular vectors it is possible to use different norms at initial and final time. The most commonly used norm is the total energy norm, but other norms e.g. the kinetic energy

norm, or the CAPE norm ([41]) can also be used. The structure of the singular vectors can be significantly different depending on the norm used.

- **Optimization area:** One may wish to compute the singular vectors targeted to the region(s) of interest, i.e. to produce the maximum spread amongst the ensemble members over a given geographical domain. This area is called optimization area and such vectors are called targeted singular vectors.

- **Optimization time:** Singular vectors are computed for a given $[t_0, t]$ time interval, the so-called optimization time. The time evolution of the perturbations is computed using the tangent linear model. As the assumption of linearity holds up to approx. 48 hours, the optimization time cannot be longer than that.

- **Generation of perturbations:** Once the singular vectors are computed perturbations can be generated. Most often a perturbation is made up of the linear combination of several singular vectors. The amplitude of the perturbations needs to be re-scaled to have the same order of magnitude as the analysis error. Evolved singular vectors (i.e. singular vectors computed for a previous model run and evolved to the time of the current analysis) can also be used to generate the initial condition perturbations. The horizontal resolution used for the singular vector computation, the number of singular vectors used for the generation of the perturbations and the number of iterations necessary for obtaining those singular vectors are also important issues.

- **Vertical optimization:** Singular vectors can be targeted not only horizontally (over a geographical domain), but vertically as well. They can be optimized for all model levels (from the highest level to the lowest), but also for a sub-layer of the atmosphere of the model.

The choices used for the singular vector computations in our experiments will be presented in Chapters 2 and 3.

Chapter 2

The applied models

The final goal of our research was to develop a short-range limited area ensemble prediction system to be used operationally at the Hungarian Meteorological Service. To achieve this goal experiments were made using the ARPEGE global and the ALADIN limited area models. The reason of choosing the ARPEGE/ALADIN model family was that its limited area member (ALADIN) is used operationally at the Hungarian Meteorological Service.

As it was mentioned before, limited area models require not only upper and lower, but lateral boundary conditions as well. In the experiments presented in the thesis global ensemble forecasts were made by running ARPEGE based ensemble systems. These global forecasts provided initial and lateral boundary conditions for the ALADIN limited area model. In this chapter the applied NWP models, ARPEGE and ALADIN are going to be presented together with the ARPEGE ensemble system, PEARP.

2.1 The ARPEGE global model

The ARPEGE global model ([47]) was developed in collaboration between Météo-France and ECMWF. ARPEGE is a spectral model which means that the prognostic variables are represented as sums of a finite set of smooth orthogonal functions ([38]). The big advantage of the spectral method is that the horizontal derivatives can be calculated analytically. The accuracy is only determined by the wavenum-

ber at which the truncation is applied. However, even in spectral models a significant part of the computations is done in gridpoint space. In ARPEGE a Gaussian grid is used for gridpoint computations.

In the vertical a hybrid coordinate system is used. This means that in the lower part of the atmosphere the vertical levels follow the Earth's surface, while in the upper part they are levels of constant pressure. The transition between the two types of levels is smooth (Fig. 2.1).

The ARPEGE model has a special feature, the so-called stretching ([40]) which enables the application of a variable resolution grid, i.e. higher horizontal resolution in the area of interest and lower resolution on the other side of the globe (Fig. 2.2).

The tangent linear and the adjoint versions of the model are coded and can be used for several purposes including data assimilation or the computation of singular vectors. Concerning data assimilation the 4 dimensional variational assimilation (4D-Var) is applied.

2.2 PEARP, the ARPEGE ensemble system

Based on the ARPEGE model an 11-member global short-range ensemble prediction system was built ([35]) which is called PEARP (formerly PEACE). From the eleven members ten are started from perturbed initial conditions and one - the so-called control member - is started from the unperturbed initial condition. The system has been running operationally at Météo-France once a day at 18 UTC up to 60 hours since June, 2004. In January, 2008 a major upgrade took place and a new PEARP version became operational. As part of the experiments were performed before this date it is important to describe not only the new, but the old configuration of PEARP as well.

2.2.1 The old version of the PEARP system

In the old version of the PEARP system perturbations were generated using only one set of targeted singular vectors. Targeted singular vectors were computed using an optimization time of 12 hours with a low resolution, without stretching

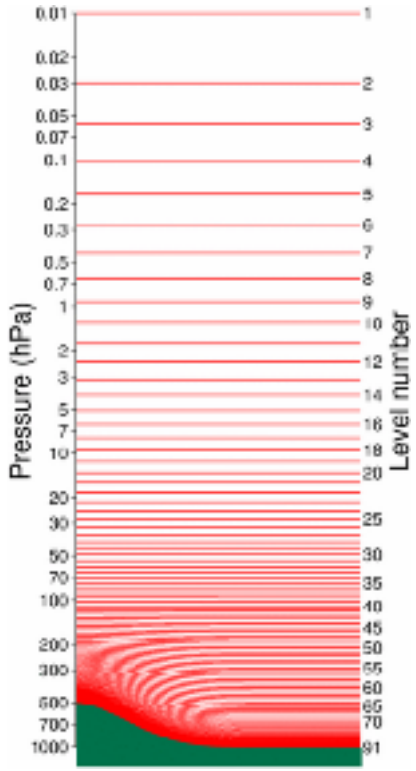

Figure 2.1: Example of the hybrid vertical coordinate system used in the models. In the lower part of the atmosphere the vertical levels follow the Earth's surface, while in the upper part they are levels of constant pressure. The transition between the two types of levels is smooth. (The 91 level system shown here is currently operational at ECMWF. Source of figure: www.ecmwf.int)

(TL95, approx. 180 km) over a limited area including Western Europe and the northern part of the Atlantic Ocean (see later on Fig. 3.2). By targeting, perturbations have their greatest impact over the area of interest (i.e. Western Europe, particularly France in this case). The orthogonal perturbations were computed by the linear combination of the first 16 targeted singular vectors. From these 16 singular vectors five perturbations were generated which were added to and subtracted from the unperturbed analysis leading to a total number of 10 perturbed initial conditions. The perturbations were scaled with an average analysis error estimate. For the integration of the 10+1 members the operational version of the ARPEGE model was used. At that time it was running with a spectral truncation

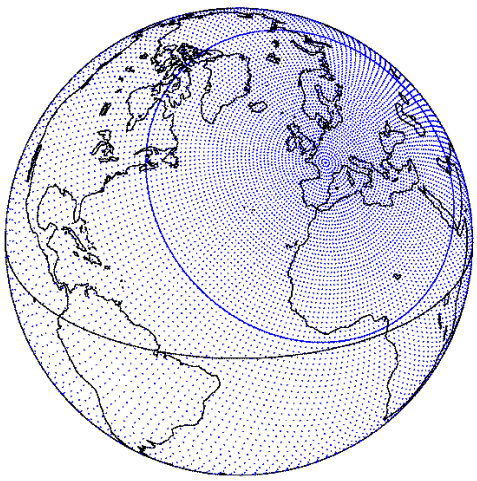

Figure 2.2: The variable resolution grid of the ARPEGE model. Higher horizontal resolution in the area of interest (that is Europe in the figure) and lower resolution on the other side of the globe.

of TL358 and a stretching coefficient of 2.4, i.e. with 23 km horizontal resolution over Europe (Fig. 2.3) and 100 km over New Zealand.

An advantage of this version was that its members, perturbed and unperturbed, had the same (high) resolution as the operational deterministic ARPEGE run of that time (23 km over Europe). The disadvantage is also clear: the high resolution limited the number of perturbed members to be used for the model integration. Nevertheless, ten members might be enough to have a reliable estimation of the probability density function.

2.2.2 The new version of the PEARP system

In January, 2008 a major upgrade took place and a new PEARP version became operational at Météo-France. The characteristics of the new version are the following: the number of members (10+1) remained unchanged, just like the horizontal resolution (TL358c2.4) while the operational deterministic ARPEGE run has now a better resolution (TL538c2.4). The number of vertical levels used for the model integration is now 55. The generation of the initial perturbations was revised: singular vectors are computed for four different areas (not only for Eu-

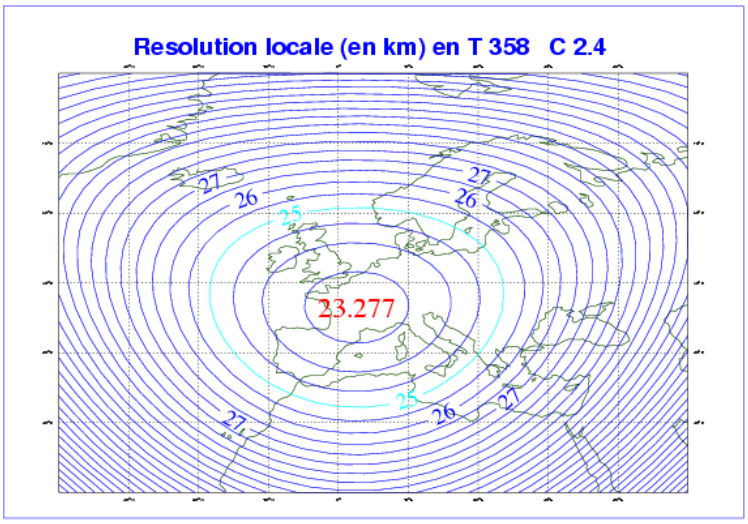

Figure 2.3: The variable resolution grid of the ARPEGE model as used in the PEARP system. Higher horizontal resolution in the area of interest (Europe, around 23 km) and lower resolution on the other side of the globe (not shown). Source of figure: Météo-France.

rope and the Northern Atlantic region) and they are all combined to create the perturbations. The different areas and the number of singular vectors computed are as follows:

- 16 singular vectors targeted for Europe and the Northern Atlantic region (as before, in the old version)
- 10 singular vectors targeted over the Northern Hemisphere
- 10 singular vectors targeted over the Tropical band (+/-30 deg.)
- 20 singular vectors targeted over the Southern Hemisphere

The first set of singular vectors (Europe and the Northern Atlantic region) is computed with a truncation of TL95 and 55 vertical levels. The other three sets are computed with a lower resolution (TL44) but using the same number of vertical levels. The optimization time remained 12 hours, but additionally 24 hours evolved perturbations are used (i.e. perturbations from the previous run of

PEARP, evolved to the time of the current analysis) to generate the initial perturbations. The perturbations are scaled by a flow-dependent background error estimate. These changes yielded a marked improvement in the spread of the ensemble members over the whole globe ([3]).

2.2.3 Some general characteristics of PEARP

As described before, in the PEARP system (both in the old and the new versions) the perturbations are added to and subtracted from the unperturbed analysis. Creating the initial conditions such a way means that they are going to be symmetric around the unperturbed analysis, which yields that the ensemble mean and the control member (the forecast started from the unperturbed initial condition) are almost identical in the initial part of the forecast (i.e. in the part when the evolution of the perturbations can be regarded as linear).

Another interesting issue is the quality of the perturbations. It was beyond the scope of the experiments (to be presented in Chapter 3), however, it might be of interest to mention this topic here. At initial time, when generating the perturbations from the singular vectors, the perturbations are rescaled with respect to a flow dependent background error estimate (average analysis error estimate in the old version). Certainly this rescaling only affects the magnitude of the perturbations and not their horizontal scale. The problem is that the scale of the perturbations is much larger than the resolution used for the model integration and also much larger than the scale of the data assimilation. Therefore the perturbations have a larger horizontal scale than the typical errors they try to simulate. This is true not only for the PEARP system but for several other ensemble systems as well (Anders Persson, personal communication).

2.3 The ALADIN limited area model

The limited area member of the ARPEGE/ALADIN model family is the ALADIN model ([22] and [23]). The ALADIN model has been developed by an international team with French leadership using as much as possible of the existing code of the ARPEGE global model. ALADIN is used operationally at the Hungarian

Meteorological Service.

Just like ARPEGE, ALADIN is a spectral model. The horizontal fields are represented with 2 dimensional Fourier functions. This representation requires the periodicity of the horizontal fields in both directions. In the ALADIN model an extension zone is used to make the horizontal fields periodic (Fig. 2.4). In this zone the values are artificial without any physical meaning.

ALADIN is a limited area model, thus lateral boundary conditions are required for the model integration. The ALADIN model typically takes the boundary conditions from ARPEGE, as it was done in case of the experiments presented in the thesis. Boundary conditions are provided only with a given frequency (e.g. every 3 hours), therefore an interpolation is performed by the limited area model for its intermediate integration steps. In operational context the almost universal approach is to overspecify the boundaries (i.e. specify the values in every point of the boundary) and damp the resultant noise with a relaxation scheme introduced by Davies ([6]). This is done in the so-called coupling zone (Fig. 2.4). On the inner border of the coupling zone values are only taken from the limited area model, while on the outer boundary only the global model is taken into account. Inside there is a smooth transition.

With the ALADIN model it is possible to run 3 dimensional variational assimilation (3D-Var) to create the initial conditions. If one decides to run the ALADIN model without data assimilation, initial conditions have to be taken from the global model (or from another limited area model covering a larger domain).

Like ARPEGE, ALADIN is using a hybrid coordinate system in the vertical (Fig. 2.1), i.e. in the lower part of the atmosphere the vertical levels follow the Earth's surface, while in the upper part they are levels of constant pressure.

In the experiments presented in Chapter 3, the ALADIN model was used with a horizontal resolution of 12 km. The time step used for the integration of the model was 450 sec. Integrations were performed over a domain covering large part of Continental Europe (Fig. 2.5). At the early stage of the work 37 levels were used in the vertical, this was later changed to 46. Initial and lateral boundary conditions were provided by ARPEGE EPS systems (PEARP or experimental setups) and a simple downscaling was performed (no data assimilation or computation of perturbations with the ALADIN model). Each member of the ARPEGE

ensemble system provided initial and lateral boundary conditions for one member of the ALADIN EPS, thus, like its global counterpart, the ALADIN ensemble system also had 11 members.

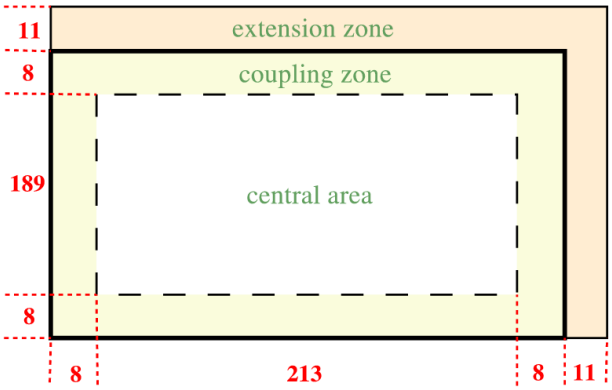

Figure 2.4: C+I+E zones used in the ALADIN model for our experiments. In the middle is the central area (in our experiments 213 gridpoints in the x direction and 189 in the y direction), where only values computed by the limited area model are taken into account. Arround the central area is the coupling zone (8 gridpoints in each direction), which is used for relaxation between values of the limited area model and the global model. In the extension zone (11 gridpoints) the values have no physical meaning and they are used only to make the horizontal fields periodic.

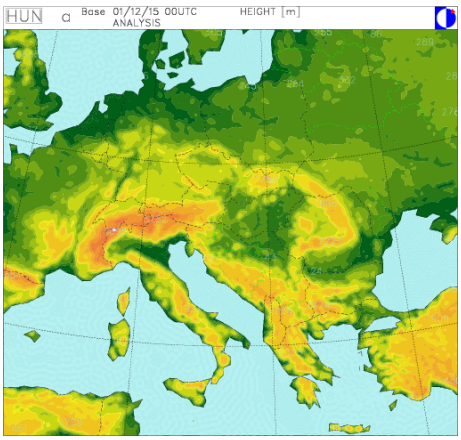

Figure 2.5: The integration domain and the orography of the ALADIN limited area model.

Chapter 3

Experiments and results

The studies that are going to be presented in this chapter were considered as a first step towards the establishment of a short-range limited area ensemble prediction system for the Central European area. The basis of the work was the downscaling of ARPEGE based global ensemble prediction systems (PEARP and experimental setups).

First issue to discuss is the question of lateral boundary conditions. To run a limited area model lateral boundary conditions are needed. This raises the following question immediately. In case of a LAMEPS, is it necessary to perturb the lateral boundary conditions as well (i.e. to run each member of the LAMEPS with different LBCs), or the same set of LBCs can be used for all members and only the ICs should be different? This question is discussed in more detail in Section 3.1.

After the first tests with the downscaling of the operational ARPEGE EPS it became clear that the system is not optimal for our area of interest. Sensitivity studies were performed in order to explore whether or not it was possible to optimize the existing operational ARPEGE EPS system (PEARP, formerly PEACE) for Central Europe by changing the optimization area and time used for the global singular vector computations. Results were analysed with the help of standard EPS verification measures (Section 3.2).

Finally, the verification results of the global and the limited area systems were compared to each other (Section 3.2.4) to see whether the limited area model could

improve the predictions of the global model or not.

3.1 The question of the lateral boundary conditions

When running a limited area model lateral boundary conditions are required. At the beginning of the forecast the initial conditions are dominant, while later the influence of the lateral boundary conditions becomes more important.

In the literature one can find that if all members of a limited area ensemble prediction system are running with the same set of LBCs and only the ICs are different, then the members will be very similar after a given time and the system will not have sufficient spread throughout the whole forecast range (for an example see e.g. [8]).

In order to check the above mentioned phenomenon in our system an experiment was performed using different initial conditions, but the same set of LBCs (the control member of PEARP) for each member of the ALADIN ensemble system. The experiment was covering one month between 01 May 2008 and 31 May 2008. Percentage of outliers diagrams[1] were plotted for 500 hPa geopotential (Fig. 3.1/a) and 850 hPa temperature (Fig. 3.1/b) in order to analyse the impact of using the same set of LBCs. It can be concluded for both parameters that the percentage of outliers was not decreasing but significantly increasing with time after the first 6-18 hours. This means that there was less and less spread in the ensemble system. This can be explained with the dominance of the LBCs.

Based on these experiments we can confirm that in case of limited area ensemble systems it is important to use not only different initial conditions but also different lateral boundary conditions for each ensemble member in order to maintain the difference between the members and to have sufficient spread, which - on average - is not decreasing but increasing with time.

[1] For a description of this verification method see B.5 in the Appendix.

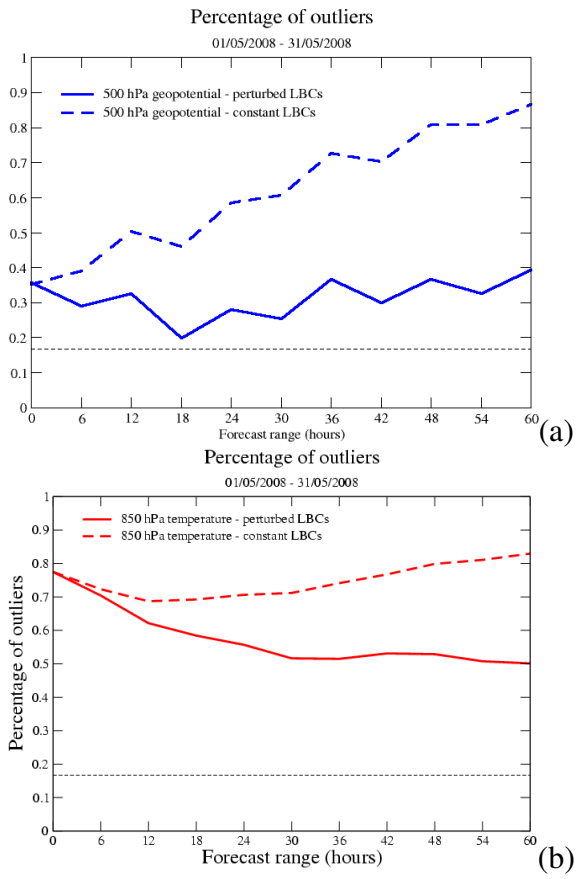

Figure 3.1: Percentage of outliers diagram for (a) 500 hPa geopotential and (b) 850 hPa temperature. Solid curve is the control experiment (with different LBCs for each ensemble member), dashed curve is the experimental run using the same set of LBCs for each ensemble member. The thin horizontal line is the expected value. Verification was performed against ECMWF analysis for the period 01/05/2008-31/05/2008.

3.2 Sensitivity studies with global singular vectors

As a first step the simple dynamical downscaling of the PEARP system was tested using the ALADIN limited area model. At the time of these experiments the optimization area used for the global singular vector computations was domain 1 in Fig. 3.2 and the system was meant to be efficient and skilful for Western Europe, particularly France. It was suspected and then proved that this system was not fully optimal for a Central European application, especially because the spread of the ensemble members was not satisfactory (not shown). Based on

these results and motivated by the findings of Frogner and Iversen ([7], [8]) and Hersbach et al. ([21]) further experiments were started to explore whether or not it is possible to optimize the existing global system for Central Europe by changing the optimization area and time used for the global singular vector computations.

To determine the singular vector optimization area and optimization time one can rely on the theories developed in the late 1940s by Rossby and Charney. According to these theories on the Northern Hemisphere mid-latitudes the large-scale dynamical influence spreads with an average speed of 30 longitude/day. In practice it means that e.g. a one day forecast for Europe is mostly determined by the initial conditions over the Eastern basin of the North Atlantic region ([36]). These studies can be used as a basis to determine the singular vector optimization area from theoretical point of view.

The experiments were concentrating on the sensitivity of global singular vectors with respect to their optimization area and optimization time. Because of the heavy computational costs the number of optimization areas used for the sensitivity experiments had to be restricted to a reasonable amount. Five different domains (covering entirely or partially the Euro-Atlantic region, Fig. 3.2) and two different optimization times were defined and tested through case studies and longer (10 and 32 days) test periods ([19], [12], [16], [15], [17]). The different optimization areas were as follows:

- Domain 1: This domain was used in the early version of the PEARP system, at the time when our experiments started. Used as a reference in the experiments.

- Domain 2: This domain was used in a newer version of the PEARP system. Used as a reference in the experiments.

- Domain 3: Smaller optimization domain, covering Europe and the Eastern basin of the North Atlantic region. According to the theories of Rossby and Charney, this domain could be more suitable for Central Europe.

- Domain 4: This domain was chosen in order to test the effect of targeting only over the integration area of the ALADIN model.

- Domain 5: Small domain, covering only Hungary.

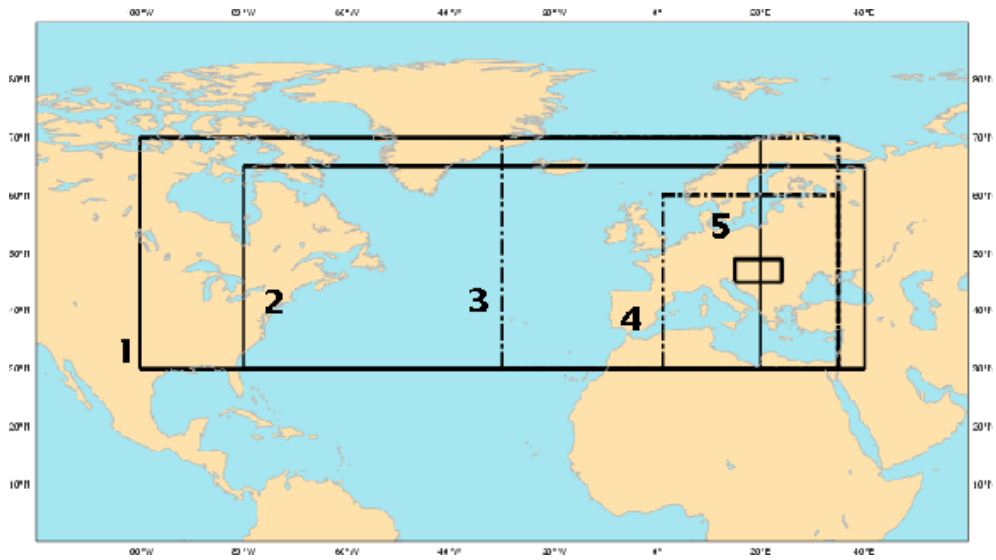

Figure 3.2: The different optimization areas used for the experiments. Domain 1 was used in the early version of the PEARP system (at the time when our experiments started). Later this was changed to domain 2. Domain 3 was used for the case studies and the experimental sets in the 10 and 32 day experiments. Domain 4, which is almost the same as the integration domain of the ALADIN model (see Fig. 2.5), and domain 5 (covering only Hungary) were used only for case studies.

For optimization time, 12 and 24 hours were defined. In the PEARP system 12 hours is applied, therefore this optimization time was used as a reference. A longer, 24 hours optimization time was also used. Based on the above mentioned theories, it might be more appropriate than 12 hours, given a Central European application.

Before presenting the experiments and their results let us spend some time on the question of verification. One of the essential issues when talking about verification is the definition of "truth". Most often SYNOP (surface synoptic observations) and TEMP (upper level temperature, humidity, wind) data are used to describe the true state of the atmosphere. However, the number of observations is variable in time and space, which means that over specific areas, such as oceans and/or in specific synoptic times (e.g. at 06 and 18 UTC) our knowledge about the atmosphere might be insufficient. Therefore analyses coming from numerical models or even short-range forecasts are often used for verification purposes.

To analyse the results of the experiments surface and upper air parameters such as 10 meter wind speed, 2 meter temperature, 850 hPa temperature and 500 hPa geopotential height were verified. The applied verification methods included several types of scores and diagrams, such as (i) BIAS and RMSE computations for the ensemble mean and the control forecast, (ii) ranked histograms and percentage of outliers diagrams, (iii) ROC and ROC area diagrams and (iv) reliability diagrams. (For a detailed description of the verification methods the reader is referred to the Appendix.) For the ten day experiment (Section 3.2.2) scores were computed against observations. ROC and reliability diagrams could not be used due to the shortness of the period and the resulting poor sampling size, only ranked histograms and percentage of outliers diagrams could be plotted. For the 32 day experiment (Section 3.2.3) verification was performed against ECMWF analysis in order to avoid the unsufficient sampling size. In this way all of the above mentioned scores could be computed and analysed.

3.2.1 Experiments with the optimization domain - Case studies

It was expected that the optimal setting of the two parameters (i.e. the optimization domain and time) would depend on the meteorological situation. In case of a large-scale phenomenon a larger domain could be more suitable, while in case of a small-scale one a smaller domain - targeted to the area of interest - could be better. Therefore a compromise solution had to be found to select the most optimal overall choice. To understand this consideration in detail, significantly different meteorological situations were selected. These situations differ in their scale, and also in the direction from which they had arrived.

- Convective event (18 July 2002). In this situation 40-70 mm/24 hours precipitation was measured at some places along the Danube and both the operational deterministic ARPEGE and ALADIN models failed to forecast the event.

- Fast moving cold front arriving to Hungary from Western Europe (22 June 2001). Both deterministic models (ARPEGE and ALADIN) overestimated the precipitation.

- Significant temperature overestimation (22 February 2004). Models predicted rain, but in reality, because of the much lower temperatures, it was sleet.
- Cyclone reaching the country from South East Europe (8 November 2004) and causing more than 30 mm/24 hours precipitation mainly in the middle and western part of Hungary.

First the emphasis was on the selection of the optimization domain, i.e. to restrict the possible choices as far as domain size and geographical location is concerned. Because of the heavy computational costs the number of singular vector optimization areas had to be restricted to a reasonable amount. For the case studies four[2] different optimization areas were used (domains 1, 3 , 4 and 5 on Fig. 3.2). Global ensemble forecasts were made with ARPEGE using these optimization domains for the global singular vector computation. The optimization time was set to 12 hours (as used in the PEARP system).

In order to analyse the results, the spread of the ensemble members was computed (around the ensemble mean) over Hungary for different meteorological parameters (such as 10-meter wind speed, 850 hPa temperature, 500 hPa geopotential, and mean sea level pressure) and the objective scores were completed with subjective verification.

Ensemble spread

Fig. 3.3 shows the standard deviation for two of the case studies, for 850 hPa temperature, as an example. It can be concluded that using the largest singular vector optimization domain (domain 1 on Fig. 3.2), the average standard deviation over Hungary (for all examined parameters) remained rather small during the entire forecast range. The reason of this behaviour is that using this optimization domain, the singular vectors are mainly located over the western basin of the North Atlantic region and they do not influence significantly the forecast over Hungary. Using optimization domain 3, the average standard deviation became significantly larger for all parameters, which can be explained by the theories of

[2]Domain 2 was not yet used in the PEARP system at that time.

Rossby and Charney. The perturbations in this case can have a larger effect on the forecast over Hungary as in the initial time they are located over the eastern basin of the North Atlantic region and during the model integration they reach the Central European area. The use of optimization domain 4 did not result in significantly different outcomes for all cases. Moreover, domain 3 proved to be more suitable in some of the examined cases. For the smallest singular vector optimization domain (domain 5 on Fig. 3.2) the initial standard deviation was rather large. However, it started to decrease with the forecast range, which can be explained by the fact that this optimization domain was very small. Therefore significant part of the initial perturbations propagated out of the optimization area after a short period of time.

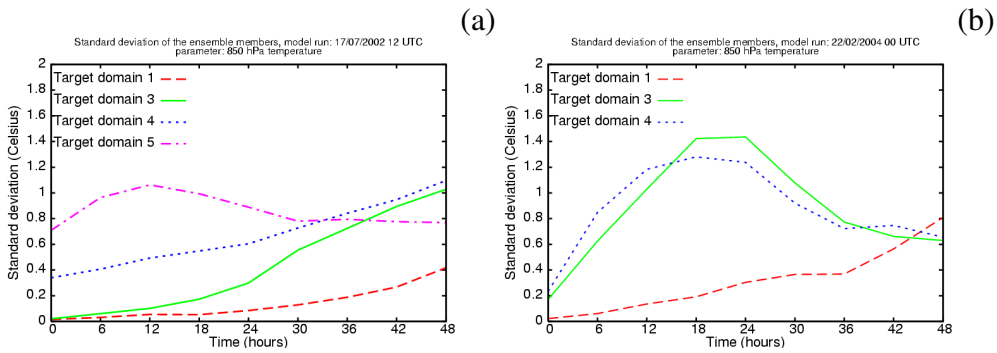

Figure 3.3: Standard deviation diagrams of the operational and experimental ARPEGE ensemble systems for two different model runs and for 850 hPa temperature. Standard deviation was computed over Hungary. a) Model run: 17 July 2002, 12 UTC (the convective event). Dashed line is standard deviation with the use of optimization domain 1, solid line is standard deviation with the use of optimization domain 3, dotted line is standard deviation with the use of optimization domain 4, chained line is standard deviation with the use of optimization domain 5. b) Model run: 22 February 2004, 00 UTC (the situation with significant temperature overestimation). Dashed line is standard deviation with the use of optimization domain 1, solid line is standard deviation with the use of optimization domain 3, dotted line is standard deviation with the use of optimization domain 4.

Subjective verification

From the subjective verification (which concentrated mainly on 2 meter temperature and precipitation) no clear conclusion can be drawn. As expected a priori,

in different meteorological situations different singular vector optimization domains proved to be the good choice in order to obtain the best ensemble forecasts. In some cases reducing the size of the optimization domain could increase the spread without improving the quality of the forecasts. On the contrary, there were cases (e.g. the convective event) when using a smaller singular vector optimization domain, the forecast became significantly better (not shown).

Results of the case studies (in accordance with the theories) suggested that domain 3 is the most optimal choice among the domains defined a priori, therefore it was selected for further experiments.

3.2.2 Experiments with the optimization domain and time - Summer period of 10 days

To confirm the preliminary conclusions drawn from the case studies, experiments for a longer, ten day period were performed. Not only the choice of the "optimal" optimization domain was tested, the additional goal of these experiments was to test the effect of different optimization times as well. It was suspected that using a longer optimization time the perturbations would have a larger effect over the area of our interest (Central Europe and particularly Hungary). Therefore, we have examined in detail two optimization domains and two optimization times, resulting altogether in four sets of experiments:

- Optimization domain 1 and optimization time 12 hours, as in the PEARP system at that time. Used as a reference in the experiment.
- Optimization domain 3 and optimization time 12 hours.
- Optimization domain 1 and optimization time 24 hours.
- Optimization domain 3 and optimization time 24 hours.

The randomly selected period was from 10 to 19 July, 2004. The first part of the period was characterized by frontal activity in the area of interest, and in the second half the weather situation over Central Europe was determined by an anticyclone.

The average ensemble spread for different meteorological parameters over Hungary was computed, and objective verification (using ranked histograms and percentage of outliers diagrams) was performed as well.

Ensemble spread

The results of the experiment showed that on average, the use of optimization domain 1 and optimization time 12 hours provided the smallest standard deviation over Hungary for all examined parameters (500 hPa geopotential height, 850 hPa temperature, mean sea level pressure, 10 meter wind speed). This can be explained with the large size of this optimization domain and the short optimization time. The perturbations - created from the singular vectors optimized to this area - typically have their maximum amplitude over the north part of the Atlantic Ocean, therefore, they do not influence significantly the Central European area in the course of the short-range forecast.

Using optimization domain 3, the spread (on average) was increased and even further improvement was obtained with 24 hour optimization time. Fig. 3.4 shows the values of standard deviation for 850 hPa temperature as an example. On average, this configuration (optimization domain 3 and optimization time 24 hours) provided the largest values in terms of standard deviation computed over Hungary.

Ranked histograms and percentage of outliers diagrams

Analysing the ranked histograms and the percentage of outliers diagrams, best results were obtained when optimization domain 3 together with optimization time 24 hours was used for the global singular vector computation. Nevertheless, for surface parameters the two outermost intervals of the ranked histogram (not shown) were still dominating, and the percentage of outliers remained much larger than the expected value which is about 0.2 in case of our ensemble system. Fig. 3.5 shows the comparison between the runs with different optimization areas (with 12 hours as optimization time).

Changing the singular vector optimization domain yielded clear improvements (especially on the higher atmospheric levels) over the verification area in terms of

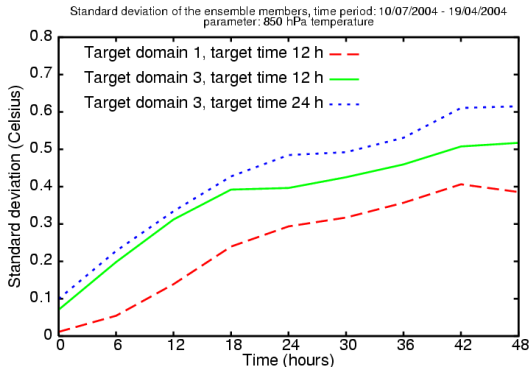

Figure 3.4: Standard deviation diagrams for the ARPEGE ensemble system for the period 10/07/2004-19/07/2004, for 850 hPa temperature. Standard deviation was computed over Hungary. Dashed line is standard deviation with the use of optimization domain 1 and optimization time 12 hours, solid line is standard deviation with the use of optimization domain 3 and optimization time 12 hours, dotted line is standard deviation with the use of optimization domain 3 and optimization time 24 hours.

spread and outliers. It is important to keep in mind that improvement in the spread does not necessarily result in more skilful ensemble forecasts, however, because of the shortness of the period, no other verification measures (e.g. ROC or reliability diagrams) could be computed and analysed.

Conclusion of the results

Results of the case studies and the ten day experiment showed that the use of optimization domain 3 provided better ensemble spread than the use of the original settings (domain 1). For optimization time, on average the runs with 24 hours had better results. However, it was also realised that a period of ten days was not long enough to draw reliable conclusions, therefore extended tests were made for a longer period in order to finalize the optimization area and optimization time.

3.2.3 Finalizing the optimization domain and time - Winter period of 32 days

The period for the extended tests - from 15 January 2005 to 15 February 2005 - was chosen randomly, and was characterized by an unusually cold weather. Verifi-

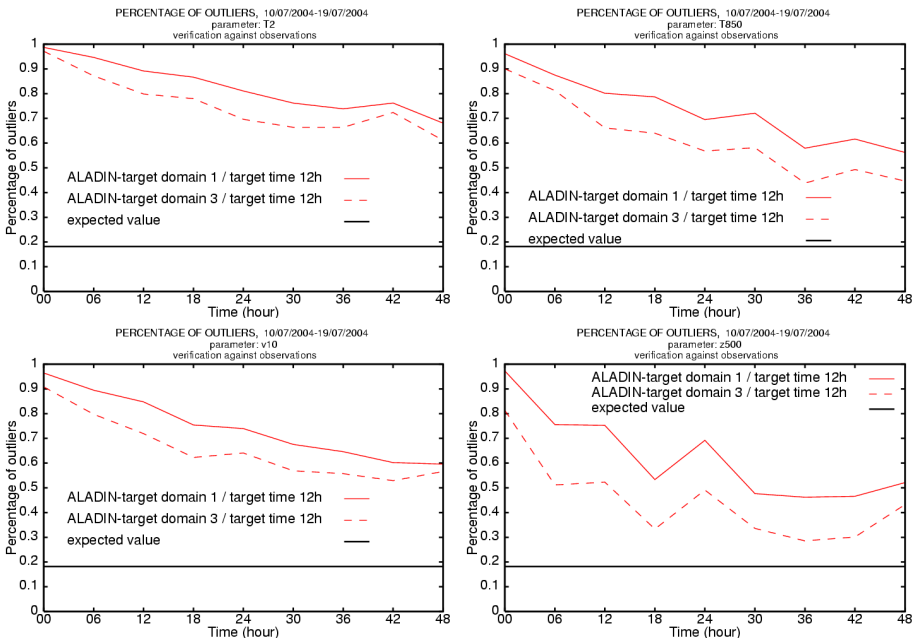

Figure 3.5: Percentage of outliers diagrams for the ALADIN ensemble system for the period 10/07/2004-19/07/2004. (a) 2 meter temperature, (b) 850 hPa temperature, (c) 10 meter wind speed, (d) 500 hPa geopotential height. Solid line is ALADIN coupled with ARPEGE ensemble members using optimization domain 1 and optimization time 12 hours for SV computation, dashed line is ALADIN coupled with ARPEGE ensemble members using optimization domain 3 and optimization time 12 hours for SV computation. Verification was performed against SYNOP and TEMP observations on the whole integration domain. The expected value is approx. 0.2 (see the thin horizontal lines).

cation was performed on the entire integration domain, against ECMWF analyses. The following configurations were tested for the 32 day period:

- ARPEGE-OPER: the operational PEARP system. At the time of this experiment the optimization domain used in the PEARP system was not domain 1 any more, it was changed to domain 2 (Fig. 3.2). The optimization time was 12 hours.
- ARPEGE-EXP: the experimental ARPEGE ensemble system where the optimization area and optimization time were changed with respect to PEARP. Smaller optimization domain - domain 3 instead of domain 2 in Fig. 3.2 - and 24 hours instead of 12 hours as optimization time.
- ALADIN-OPER: downscaling of the experiment ARPEGE-OPER (i.e. downscaling of the PEARP system) with the ALADIN model. Both initial and lateral boundary conditions were coming from ARPEGE-OPER.
- ALADIN-EXP: downscaling of the experiment ARPEGE-EXP with the ALADIN model. Both initial and lateral boundary conditions were coming from ARPEGE-EXP.

Comparing the error of the ensemble mean and the control member

As a first step, the relationship between the error of the control member (i.e. the forecast started from the unperturbed initial condition) and the error of the ensemble mean was analysed. Since the perturbations in the examined ensemble systems are symmetric around the unperturbed initial condition and have a small amplitude at initial time, the ensemble mean and the control forecast are almost identical in the early forecast ranges. This means that their RMSE is also very similar. However, after the initial linear phase it is expected that the ensemble mean has lower RMSE values than the control forecast since the averaging has the effect of filtering out the less predictable features and leaving only the more predictable ones that show agreement among the ensemble members. It was found that for certain parameters (e.g. 500 hPa geopotential height, Fig. 3.6/a) the ensemble mean was almost identical to the control member in terms of RMSE until 42-48 hours, while

for other parameters (e.g. 10 meter wind speed, Fig. 3.6/b) the ensemble mean had lower RMSE values already after 6 hours. This was true for all of the four studied systems (operational and experimental, global and limited area forecasts). From the results it seems that in case of geopotential, the ensemble members remain centered around the control member during the first 42-48 hours of the forecast, hence the ensemble mean and the control member are almost identical. For wind speed nonlinearity has a much stronger effect, therefore the ensemble mean and the control forecast start to differ from the early forecast ranges.

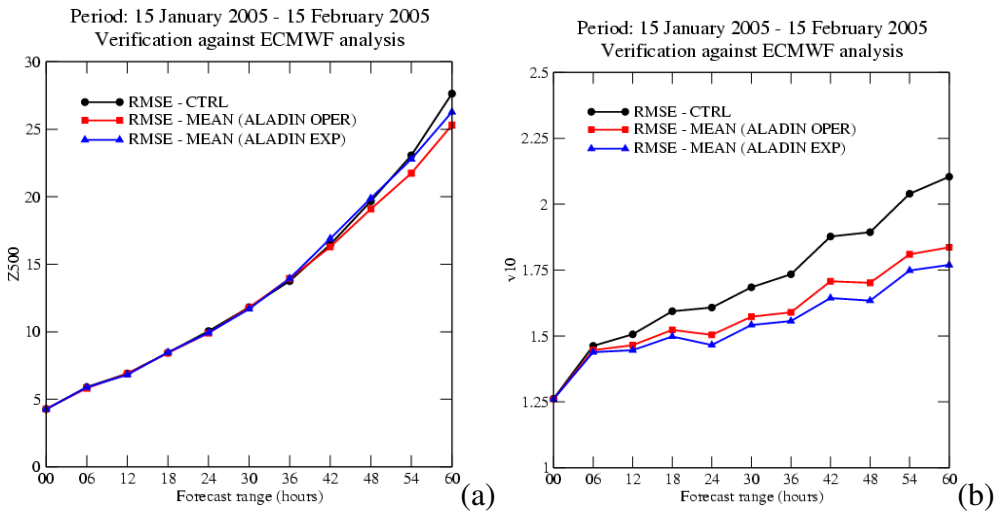

Figure 3.6: RMSE of the control member (black), RMSE of the ensemble mean for ALADIN-OPER (red) and RMSE of the ensemble mean for ALADIN-EXP (blue). a) 500 hPa geopotential height, b) 10 meter wind speed. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005.

Spread-skill relationship

Another important feature of an ensemble system is the spread-skill correspondence. The spread of the ensemble system (computed around the ensemble mean) should be in good agreement with the forecast error (e.g. RMSE of the ensemble mean). In case of large error, large spread is expected as a sign of high unpredictability. On the other hand, if the spread is small, it is expected that the situatuion has good predictability, therefore the error should be small as well. If the

spread is larger (smaller) than the error, then the system is said to be over- (under-) dispersive. Both ARPEGE-OPER (not shown) and ALADIN-OPER (Fig. 3.7) were found to be underdispersive for all of the verified parameters, especially for the surface ones (10 meter wind speed, 2 meter temperature). As expected a priori, the change in the optimization domain and optimization time resulted in a better spread-skill correspondence for ARPEGE-EXP (not shown) and ALADIN-EXP (Fig. 3.7). Moreover, for the 500 hPa geopotential height (Fig. 3.7/b), both of the experimental systems became slightly over-dispersive in several forecast steps (only ALADIN is shown). For the explanation of these results (over-dispersion at the higher altitudes and under-dispersion near to the surface, especially in case of 2 meter temperature) some speculative explanations can be given. It is believed that any of these aspects or the combination of them can contribute to this behaviour:

- Regarding surface variables only surface pressure is perturbed in the global ARPEGE ensemble system therefore the uncertainties around the surface might be treated insufficiently, resulting in too small spread.
- The main energy of the singular vectors is usually located near to the steering level (around 700hPa) and during the forecast it propagates upwards rather than downwards in the atmosphere ([26]).
- The uncertainties related to the physical parameterizations are not addressed at all, only vertical diffusion is used in the course of tangent linear and adjoint integrations, which might penalize the surface more than the upper air fields.

Ranked histograms, percentage of outliers diagrams

Ranked histograms and percentage of outliers are useful tools to analyse different characteristics of the ensemble system. In an ideal case the distribution of the ranked histogram should be flat. Different shapes of the distribution indicate different behaviours like BIAS, too small or too large spread. For both ARPEGE-OPER and ALADIN-OPER, the ranked histograms were far from being ideal in

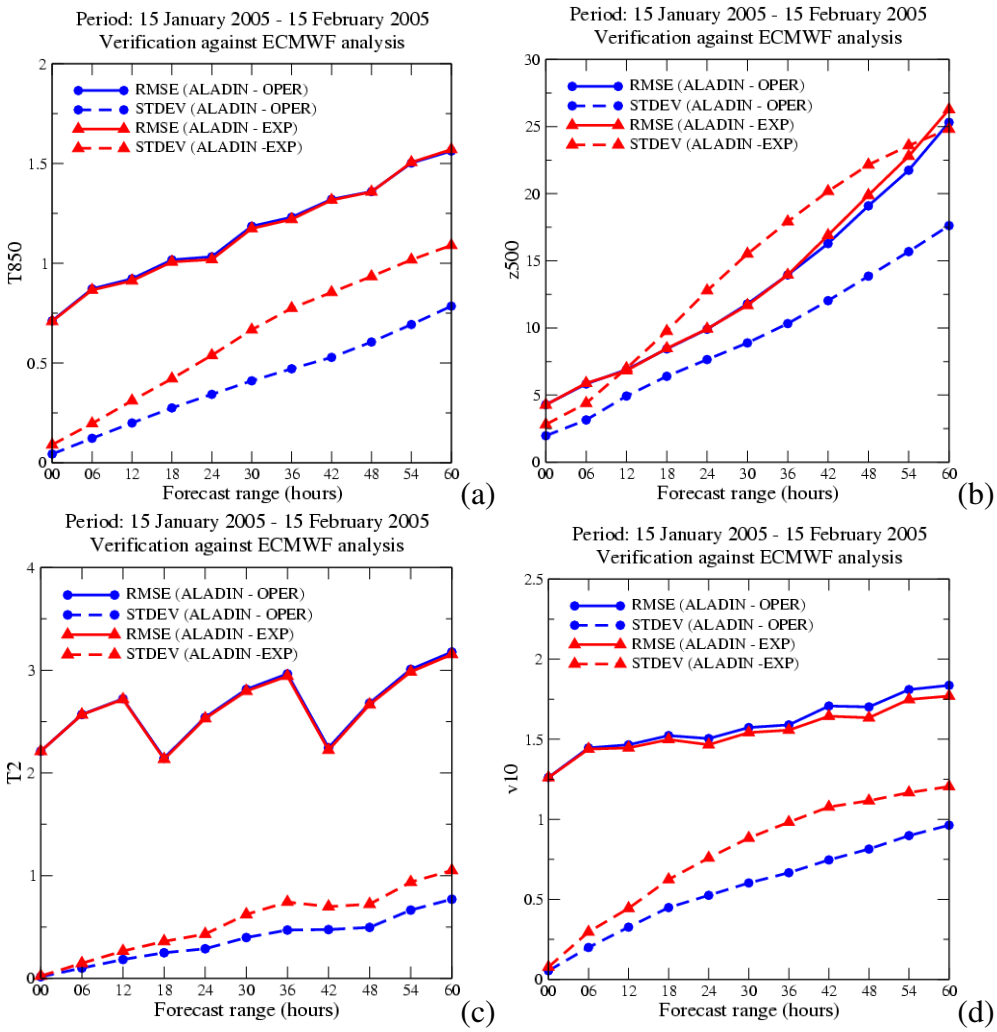

Figure 3.7: RMSE of the ensemble mean (solid line) and spread of the ensemble (dashed line) for the experiments ALADIN-OPER (circle symbols) and ALADIN-EXP (triangle symbols). a) 850 hPa temperature, b) 500 hPa geopotential height, c) 2 meter temperature and d) 10 meter wind speed. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005. In case of (a) and (c) the RMSE values of ALADIN-OPER and ALADIN-EXP are almost identical, thus they cannot be distinguished from each other in the figures.

every forecast step. This is especially true for 2 meter temperature (see Fig. 3.8/a for ALADIN-OPER, ARPEGE is not shown). For 500 hPa geopotential height the results are much better (Fig. 3.8/c), although the diagrams still show an "U" shape, indicating the lack of sufficient ensemble spread. The same was found through the spread-skill diagrams (Fig. 3.7). In case of the experimental sets (ARPEGE-EXP and ALADIN-EXP) the distributions moved towards the ideal one. For 2 meter temperature the improvement was not too significant (Fig. 3.8/b), but for 500 hPa geopotential height (Fig. 3.8/d) the ranked histograms became significantly flatter. Again, this is in good agreement with the results found through the spread-skill diagrams. The improvement is even more visible on the percentage of outliers (i.e. the sum of the two outermost bins of the ranked histogram, Fig. 3.9).

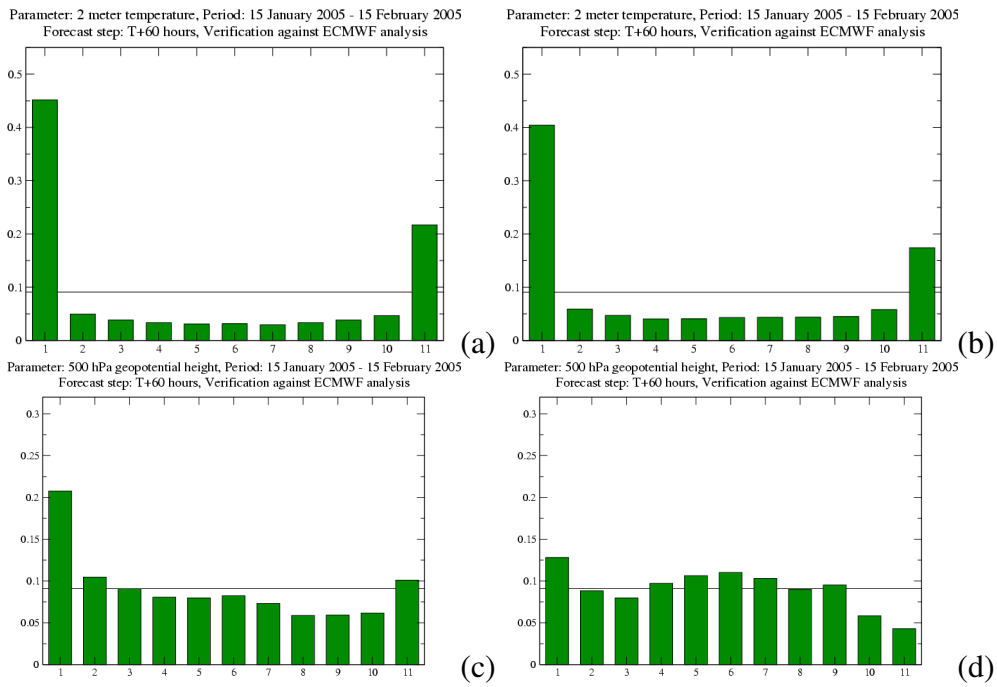

Figure 3.8: Ranked histograms at T+60 hours for a) 2 meter temperature, ALADIN - OPER, b) 2 meter temperature, ALADIN - EXP, c) 500 hPa geopotential height, ALADIN - OPER, d) 500 hPa geopotential height, ALADIN - EXP. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005. Horizontal line is the expected value, i.e. 1/(ensemble members+1).

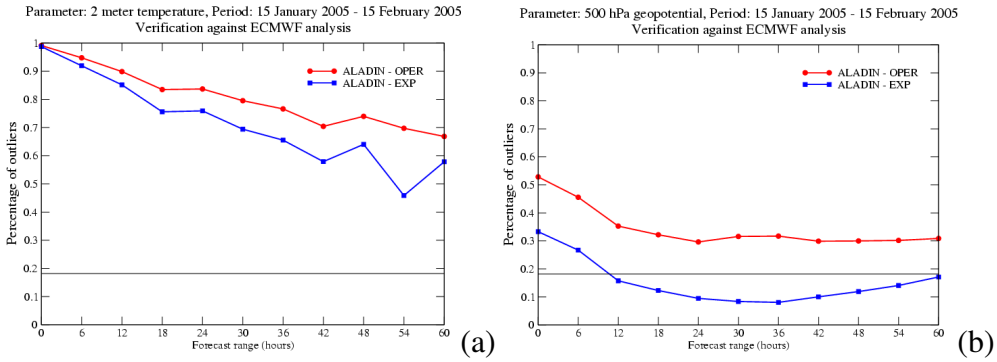

Figure 3.9: Percentage of outliers diagrams for a) 2 meter temperature, ALADIN-OPER (red), ALADIN-EXP (blue), b) 500 hPa geopotential height, ALADIN-OPER (red), ALADIN-EXP (blue). Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005. Horizontal line is the expected value, i.e. 2/(ensemble members+1).

ROC and ROC area diagrams

ROC and ROC area diagrams represent the skill of the ensemble system compared to the use of climatological statistics. A ROC area of 1 represents a perfect system, while an area less than 0.5 means the forecasts have no skill compared to climatological data. To verify our experiments ROC diagrams were plotted and the ROC area was calculated for 10 meter wind speed with thresholds 2, 5, 10 and 15 m/s and for 850 hPa temperature anomaly with thresholds $\pm 8°$C and $\pm 4°$C. For the 850 hPa temperature anomaly the scores were quite good already for ALADIN-OPER, with ROC areas significantly higher than 0.5 (i.e. around 0.85 0.95). The use of the experimental set (ALADIN-EXP) showed further improvement (Fig. 3.10/a). The 10 meter wind speed scores were also good, but the improvement was not significant. However, the change in optimization area and optimization time improved or at least did not degrade the quality of the forecasts (Fig. 3.10/b).

Reliability diagrams

Reliability diagrams are used to test the ability of the system to correctly forecast probabilities of a certain event. For that reason forecast probabilities are plot-

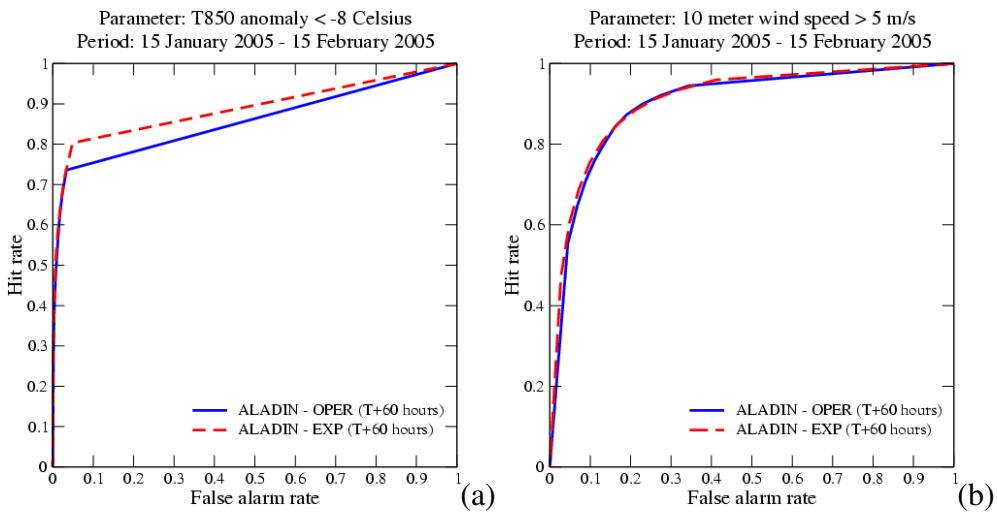

Figure 3.10: ROC diagrams at T+60 hours for the experiments ALADIN-OPER (solid line) and ALADIN-EXP (dashed line). a) 850 hPa temperature anomaly < -8 Celsius. b) 10 meter wind speed > 5 m/s. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005.

ted against conditional observed frequencies (with forecast probabilities on the x-axis and observed frequencies on the y-axis). For a perfect system the points lie along the diagonal. Reliability diagrams were plotted for 10 meter wind speed with thresholds 2, 5, 10 and 15 m/s and for 850 hPa temperature anomaly with thresholds $\pm 8°$C and $\pm 4°$C.

Fig. 3.11 shows some example of the results. For 850 hPa temperature anomaly below -4 Celsius (Fig. 3.11/a) both the operational and the experimental sets were underestimating the event. For 10 meter wind speed larger than 5 m/s (Fig. 3.11/b) the reliability was good for both sets for low probability values. For high probabilities both ALADIN-OPER and ALADIN-EXP show overestimation.

Comparing the operational and the experiamental sets, in this case we cannot come to a clear conclusion. In certain forecast steps and for certain forecast probabilities ALADIN-OPER performed better, in other cases ALADIN-EXP had better scores (Fig. 3.11). It can be concluded however, that on average the experimental set at least kept the forecast quality.

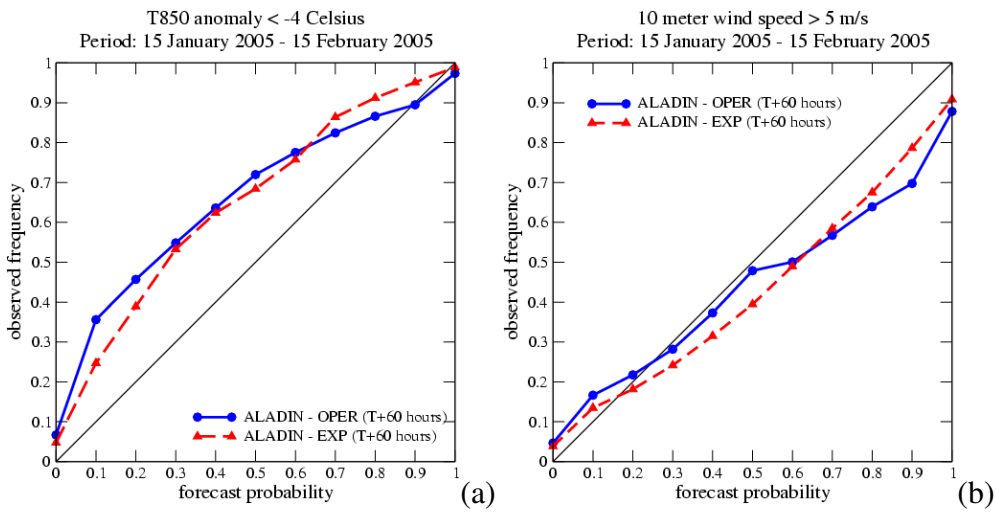

Figure 3.11: Reliability diagrams at T+60 hours for the experiments ALADIN-OPER (solid line, circle symbols) and ALADIN-EXP (dashed line, triangle symbols). a) 850 hPa temperature anomaly < -4 Celsius. b) 10 meter wind speed > 5 m/s. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005.

Conclusion of the results

As an overall conclusion it can be said that ALADIN-EXP (i.e. the use of a smaller optimization area targeted to Central Europe and a longer optimization time for the global singular vector computation) performs significantly better according to several verification measures (e.g. ranked histograms) while according to other tools (e.g. reliability diagrams) it keeps the quality of ALADIN-OPER. For a brief summary of the results see Tab. 3.1.

It should be noted that in spite of the improvements achieved by the use of optimization domain 3 and 24 hours as optimization time, the spread-skill relationship is still not satisfactory for the majority of the verified parameters, especially for the surface ones. Possible reasons of this behaviour were already mentioned before (see the part about the spread-skill relationship) and it seems that the lack of surface perturbations is one of the weak spots of the system.

Table 3.1: Summary of the results of the different experiments. EXP v OPER denotes the impact of using the experimental sets (ARPEGE-EXP and ALADIN-EXP) instead of the operational ones (ARPEGE-OPER and ALADIN-OPER). ALADIN v ARPEGE denotes the impact of using the high resolution limited area model (ALADIN) instead of the global model (ARPEGE).

Verification measure	**EXP v OPER**	**ALADIN v ARPEGE**
BIAS, RMSE, spread	BIAS and RMSE of the ensemble mean remained similar. Improvement for all parameters and all forecast steps in terms of spread. Improved spread-skill relationship.	Slight enhancement for some of the parameters, but no significant overall improvement.
Ranked histogram, percentage of outliers diagram	Improvement for all parameters and all forecast steps. Spread increased and the two outermost ranks of the ranked histograms moved towards the expected value.	Slight enhancement for some of the parameters, but no significant overall improvement.
ROC diagram, ROC area	For all parameters and all forecast steps at least slight improvement exists.	Slight enhancement for some of the parameters, but no significant overall improvement.
Reliability diagram	No clear improvement, but no degradation of forecast value, either.	Slight enhancement for some of the parameters, but no significant overall improvement.

3.2.4 Comparison of global and limited area ensemble systems

When running limited area forecasts it is always important to know whether the limited area model can improve the predictions of the global model or not. Therefore the verification scores of ARPEGE and ALADIN were compared to each other for the 32 day period from 15/01/2005 to 15/02/2005. It can be said that by simply downscaling the global ARPEGE ensemble forecasts using the higher

resolution ALADIN model it is very difficult to achieve significant - overall - improvements ([15], [17]). For some parameters and verification measures the limited area ensemble forecasts performed better (Fig. 3.12/a), in other cases the global forecasts were more skilful (Fig. 3.12/b). Also, in a couple of cases, the two models had nearly the same scores (see Tab. 3.1). Some aspects behind these results might be the relatively small resolution difference between the global and the limited area models (approx. 23 km and 12 km, respectively), or the too strong impact of the lateral boundary conditions.

As limited area models are running with a better horizontal resolution than their global counterparts, the representation of the orography is more realistic, which is very important in case of e.g. wind or precipitation forecasts. As a result of the increased resolution, limited area models can produce better defined mesoscale structures. Therefore, it is important to run limited area models even if there is no significant - overall - improvement with respect to the global model providing the lateral boundary conditions.

In addition, one should not forget, that it is a common phenomenon that high resolution models might perform worse (on average, not for all individual cases) than the low resolution ones when usual verification measures are applied. Although the increased resolution generally produces more realistic results, inevitable errors in timing and position can lead to larger RMSE values than for the smoother forecasts of the low resolution model. This is known as the *double penalty* problem. Therefore the results presented above (i.e. no significant overall improvement by the LAM) should be interpreted with care.

3.2.5 Conclusion of the sensitivity experiments

The basis of the experiments presented in the previous sections was the dynamical downscaling of global ARPEGE EPS systems (the operational PEARP and experimental setups) with the ALADIN limited area model. The skills and capabilities of both the global and the limited area ensemble systems were investigated in detail. Due to the fact that - especially the old version of - PEARP is meant to be efficient and skilful mainly for Western Europe (particularly for France) it was

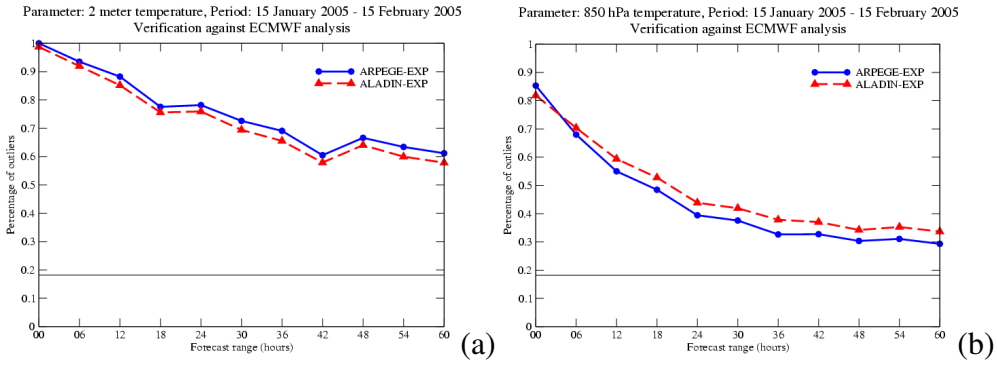

Figure 3.12: Percentage of outliers for the experiments ARPEGE-EXP (solid line, circle symbols) and ALADIN-EXP (dashed line, triangle symbols). a) 2 meter temperature, b) 850 hPa temperature. Verification was performed against ECMWF analysis for the period 15/01/2005-15/02/2005. Horizontal line is the expected value.

suspected and then proved that it is possible to optimize the system for Central Europe by changing the optimization area and the optimization time used for the global singular vector computation.

The verification results of our experiments confirmed that the proper choice of the singular vector optimization area and time can increase the spread and on average can improve the skill of the ensemble for the Central European area. Similar results were found by Frogner and Iversen in Norway. They found that using targeted singular vectors for the global SV computations, the skill of the LAMEPS can be improved for Northern Europe ([7], [8]). Therefore, instead of downscaling members of the ECMWF EPS, they use a system called TEPS to provide boundary conditions for their operational limited area system ([24]). TEPS is based on ECMWF EPS, but the singular vectors are targeted to have the greatest impact in the Northern European region.

Although at the moment we do not have the necessary conditions to run on a daily basis a targeted version of PEARP in order to obtain initial and lateral boundary conditions for our ALADIN ensemble system, the results presented in this chapter have a clear theoretical, and a potential practical value, and - just like in Norway - can be used in the operational practice some way in the future.

Analysing the verification results it can be concluded that for surface parameters, especially for temperature, the ensemble spread is not satisfactory. On the

one hand the system is underdispersive (i.e. the RMSE of the ensemble mean is larger than the enemble spread), on the other hand the percentage of outliers is much larger than the expected value, which means that the verifying analysis falls out of the interval defined by the (sorted) ensemble members too often. It is suspected that this behaviour is caused by the lack of surface perturbations. Therefore the issue of surface perturbations needs be investigated (and solved) in the future.

Chapter 4

The (quasi-) operational LAMEPS system of HMS

The studies presented in the previous chapter were considered as a first step towards the establishment of a short-range limited area ensemble prediction system for Central Europe (particularly for Hungary). As a second step, a quasi-operational system was built at HMS. At present, the only operationally feasible solution was the direct downscaling of the PEARP members, therefore this method is used. The system has been running on a daily basis since February, 2008. Having an operational system is very important, both for the forecasters and for the developers, in order to gain experience not only from case studies and longer test periods, but also on a day-to-day, real-time basis.

In this chapter this (quasi-) operational ensemble prediction system will be described. First the characteristics of the system are briefly presented in Section 4.1 (for a detailed description see part C of the Appendix), followed by a case study in Section 4.2 and finally the verification results are discussed in Section 4.3.

4.1 Characteristics of the system

The 11-member short-range limited area ensemble prediction system of HMS has been running every day, in quasi-operational status, since February, 2008 ([18]).

It is run with the ALADIN limited area model and it is driven by the members of the global PEARP system. At present no local data assimilation or generation of local perturbations are applied for the LAMEPS. Forecasts are made once a day starting from the 18 UTC data.

In order to be able to use the outputs of the global model as initial and lateral boundary conditions an interpolation is needed to the exact domain and resolution (approx. 12 km) which is used for the model integration. Once the initial and lateral boundary conditions are in the proper format (resolution, domain, etc.) the integration of the model can start. The ALADIN ensemble system is running on a domain covering large part of Continental Europe (Fig. 2.5) with a horizontal resolution of approx. 12 km. In the vertical 46 levels are used. Forecast length is 60 hours and the time step used for the integration is 450 seconds (7.5 minutes). The output frequency of forecasted fields is set to 3 hours (can be changed arbitrarily).

The next step is the post-processing of raw model outputs in order to support the application of the model results by forecasters, or by the end-users. After performing post-processing to a latitude-longitude grid, the outputs of the LAMEPS system are mainly visualized using HAWK ([1]). HAWK is used in the everyday work of the forecasters to visualize the outputs of several NWP models (both deterministic and probabilistic), observations, radar and satellite data, etc. The available products from our LAMEPS system are the ensemble mean, the ensemble spread (computed around the mean), individual ensemble members and probability fields[1] for several parameters. The individual members can be visualized in the form of spaghetti diagrams. In addition, plume diagrams are also plotted for several parameters and selected Hungarian locations.

4.2 Case study - heavy precipitation event

In this section we present an example of the use of LAMEPS products in case of an intense precipitation event that occured in February 2009. Fig. 4.1 displays the analysed weather situation at 00 UTC, 09 February 2009. The main feature of this weather pattern is the strong convergence zone present through the west part

[1]For a detailed description of the visualization methods the reader is referred to the Appendix.

of Hungary and Slovakia. This intense frontal system stayed almost at the same position for about 24-36 hours, providing ideal conditions for the occurrence of heavy precipitation (Fig. 4.2). 15-25 mm/24 hours of precipitation was observed in the west and north parts of Hungary, showing quite good agreement with the position of the stationary frontal zone.

The NWP models that are available for short range at the Hungarian Meteorological Service were all giving good indication of this significant precipitation event. Fig. 4.3 displays the 24 hour accumulated precipitation forecasts from two consecutive runs of the operational (deterministic) ALADIN model (started from 18 UTC, 07 February and from 00 UTC, 08 February 2009), the ensemble mean field from the quasi-operational LAMEPS run started from 18 UTC, 07 February and the ECMWF high resolution deterministic run started from 00 UTC, 08 February. All of the forecasts shown on Fig. 4.3 are verifying on the observational period from 06 UTC, 08 February to 06 UTC, 09 February. The different forecasts all seem to show correctly the main large scale pattern of this event, i.e. the larger precipitation amounts were situated mostly in the west and north parts of the country, where the stationary frontal system characterized the weather. However, the smaller scale features are quite different.

Both deterministic ALADIN forecasts indicated a band of precipitation along the hills of the north, northwest part of Hungary, having very extreme values of 30-50 mm. These forecast values were significantly above the actual observations, the difference was around 10-20 mm over the area of the forecast maximum. Since in this situation the precipitation was expected to fall mostly as snow in the hilly area, the potential reliability of these extreme predicted values was very important to the forecasters on duty.

Regarding the ECMWF forecast, the severe precipitation of the deterministic ALADIN runs was supported by this model as well. Although the ECMWF model was not directly showing values above 30 mm, the large size of the *more than 20 mm* area (in combination with the coarser grid used by the ECMWF model) was an alerting signal for a potentially much heavier precipitation on local scales.

The ensemble mean precipitation forecast from the LAMEPS, however, was showing a different and more realistic picture. The maximum forecast values were mostly found in the west, southwest part of Hungary with up to 20-25 mm,

in good agreement with the observations. More importatly the LAMEPS mean was predicting only 10-22 mm over the hilly area which was closer to reality, even if these values were a bit below the observations.

The probability maps based on the LAMEPS and the ECMWF EPS systems highlight the same difference (Fig. 4.4). The two probability maps look rather similar regarding the southwest part of Hungary, whereas they differ significantly over the hilly area north of the lake Balaton. The ECMWF EPS shows almost 100% probability in the northwest area, while the LAMEPS indicates only rather moderate chance for more than 20 mm over the same region, which is more realistic compared to the observations.

As an example of a possible interpretation of the LAMEPS forecasts Fig. 4.5 shows a plume diagram displaying all the 11 ensemble members for different parameters (2 meter temperature, total precipitation and 10 meter wind speed) for the location of Nagykanizsa in the southwest of Hungary, somewhere in the middle of the high probability area.

As a summary it can be concluded that in this event the LAMEPS forecast was delivering a very good extra guidance to the forecasters, complementing the operational deterministic ALADIN run. Although it slightly underestimated the precipitation over the hills west and north of Budapest, on the whole it was helping a lot the forecasters in the process of making the forecasts as accurate as possible in a difficult and potentially very severe situation.

4.3 Verification results

Verification of the quasi-operational LAMEPS system was performed for almost nine months from 10 March 2008 to 30 November 2008 against ECMWF analysis. Scores were computed for the whole period and also for the different seasons separately (only the results for the whole period are going to be shown, not the separate seasons). The verified parameters were temperature, geopotential and wind speed on several levels (500, 700, 850, 925 and 1000 hPa). The common

LACE[2] verification package was used. This verification package was developed in collaboration with colleagues from other LACE countries ([13], [33] and [27]). For a detailed description of the verification methods the reader is referred to the Appendix.

4.3.1 Comparing the error of the ensemble mean and the control member

As a first step the error of the control member was compared with the error of the ensemble mean. For geopotential the RMSE of the control member and the RMSE of the ensemble mean were almost identical during the whole period. (Fig. 4.6/a shows the results for 500 hPa.) For temperature (Fig. 4.6/b shows the results for 500 hPa) and wind speed (see Fig. 4.6/c for the scores on 1000 hPa) there was a difference between the ensemble mean and the control forecast already after the first 6 hours, with the ensemble mean having lower RMSE values. Similar behaviour was found when analysing the results of the 32 day experiment (Section 3.2.3). Results indicate that in case of geopotential, the ensemble members remain centered around the control member during almost the whole forecast, hence the RMSE of the ensemble mean and the control member are almost identical. For wind speed and temperature the ensemble mean and the control forecast start to differ from the early forecast ranges, which suggests that nonlinearity has a stronger effect on these parameters.

4.3.2 Spread-skill relationship, percentage of outliers

An important feature of an ensemble system is the spread-skill relationship. In Fig. 4.7 the spread-skill relationship is plotted for geopotential, temperature and wind speed for two levels: 500 and 850 hPa. In all cases the system was found to be underdispersive, i.e. the spread was smaller than the RMSE of the ensemble mean.

Another way of analysing the spread of an ensemble system is the use of per-

[2]LACE: A collaborative project with seven participating National (Hydro-) Meteorological Services (including Hungary) in the field of mesoscale numerical weather modelling.

centage of outliers diagrams. These diagrams tell us how often the verifying analysis lies out of the interval defined by the ensemble members. Fig. 4.8 shows the percentage of outliers for geopotential, temperature and wind speed for 500, 700, 850, 925 and 1000 hPa. For all levels and parameters the percentage of outliers was above the expected value. This means that the verifying analysis falls out of the interval defined be the (sorted) ensemble members too often, indicating that the spread of the ensemble is not sufficient.

Both in case of the spread-skill relationship and the percentage of outliers diagrams best results were obtained for 500 hPa and got worse as going closer to the surface. As already mentioned in Section 3.2.3, this behaviour might have several reasons, related to the perturbations used in the PEARP system:

- No perturbation of surface parameters, except for surface pressure.
- The maximum of the energy of the singular vectors is located around 700 hPa and during their evolution the energy propagates upwards, rather than downwards.
- The uncertainties related to the physical parameterizations are not addressed at all during the singular vector computation.

This behaviour (better results for higher levels) suggests that surface perturbations should also be included in the system. One can also conclude that even for 500 hPa, the spread of the ensemble is smaller than the RMSE of the ensemble mean, and the percentage of outliers is above the expected value. This indicates that local perturbations, targeted especially to the area of our interest would be needed to improve the quality of the system.

4.3.3 ROC and reliability diagrams

ROC and reliability diagrams were plotted for wind speed only (for technical reasons), with thresholds 1, 2, 5 and 10 m/s on 5 levels (500, 700, 850, 925 and 1000 hPa). Comparison was made between (i) different thresholds, (ii) different forecast ranges and (iii) different vertical levels.

The comparison between the different thresholds reveals that results are better for higher wind speeds. For reliability diagrams, in case of 1 m/s threshold the

system shows significant underestimation for low and middle probabilities and slight overestimation for high probabilities. For higher thresholds the underestimation for low probabilities and overestimation for high probabilities remains, but the curves move significantly closer to the diagonal. Fig. 4.10/a shows examples for 1000 hPa and T+60 hours. Similar results were obtained analysing the ROC diagrams (Fig. 4.9/a): the ROC area was well above 0.5 for all thresholds, with more skilful forecasts for wind speeds larger than e.g. 10 m/s than wind speeds exceeding 1 m/s.

The comparison between different forecast ranges shows that the system has similar skill throughout the whole forecast inteval, however results are somewhat worse in the early forecast ranges according to the reliability diagrams (this was more pronounced in autumn than during the spring and the summer). Fig. 4.9/b and Fig. 4.10/b show examples for 1000 hPa and 10 m/s as threshold.

As regards the different vertical levels, ROC and reliability curves were plotted for different vertical levels as well. In terms of ROC, results are quite similar for all levels, the only exception is 1000 hPa where the shape of the curves was slightly different, but the ROC area was similar to the other levels. Fig. 4.9/c shows examples for T+60 hours and 10 m/s as threshold. Reliability diagrams show that for low probabilities results are quite similar for all levels and show a very good reliability in case of higher thresholds. For high probabilities differences are more significant and the curves lie somewhat farther from the diagonal with better results for higher levels. This is in agreement with the conclusions drawn from other verification measures. Fig. 4.10/c shows examples for T+60 hours and 10 m/s as threshold.

4.3.4 Conclusion of the verification results

As an overall conclusion it can be said that the system behaved very similar in all three seasons. In spite of the problems with the spread-skill relationship (especially observed at lower levels and the surface) there are cases when the the ensemble members show large spread, indicating the high uncertainty of the situation. In such cases the LAMEPS is definitely a useful complement of the operational deterministic ALADIN run.

Results have shown that better scores are obtained for higher levels. The possible reasons of this behaviour were described in Section 4.3.2. However, it is important to have skilful prediction of surface parameters as well. It would also be desirable to compute perturbations that are targeted for the area of our interest (Central Europe, particularly Hungary), in order to improve the spread-skill relationship and the quality of the forecasts as well.

For these reasons it was decided to work on the computation of local perturbations. Experiments have started to compute singular vectors with the ALADIN model ([14] and Chapter 5). The aim of these experiments is to generate perturbations from the ALADIN singular vectors, and use them to perturb locally the initial conditions of the LAMEPS system. As lateral boundary conditions, PEARP or ECMWF EPS members will be used. Perturbation of surface fields is also an important issue to be solved. A possible solution could be the method applied in the LAEF system, see [45] for more details.

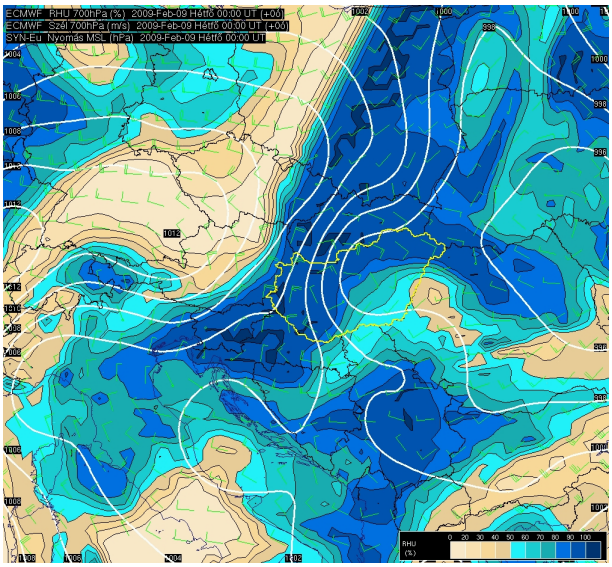

Figure 4.1: Analysed weather situation at 00 UTC, 09 February 2009. The white isolines display the mean sea level pressure (in hPa), the shaded areas show the 700 hPa relative humidity (in %), and the wind arrows indicate the wind conditions at 700 hPa.

Figure 4.2: Observed precipitation (in mm) at the stations of the Hungarian observing network accumulated over a 24 hour period from 06 UTC, 08 February to 06 UTC, 09 February 2009.

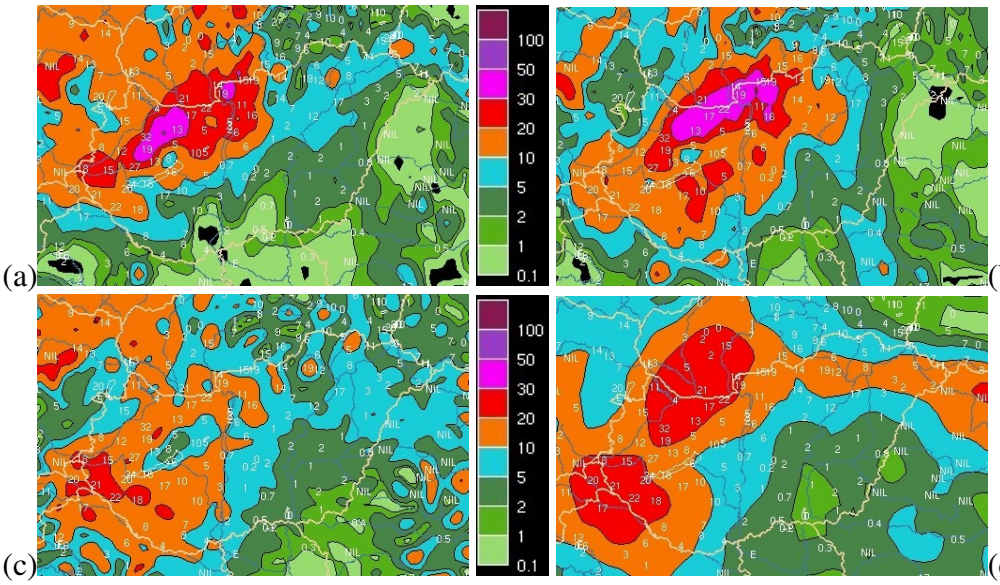

Figure 4.3: Precipitation forecasts of the different models available, with the corresponding observed values accumulated over a 24 hour period from 06 UTC, 08 February to 06 UTC, 09 February. (a) ALADIN deterministic run started from 18 UTC, 07 February, (b) ALADIN deterministic run started from 00 UTC, 08 February, (c) ensemble mean field from the LAMEPS run started from 18 UTC, February 07, (d) ECMWF deterministic run started from 00 UTC, 08 February.

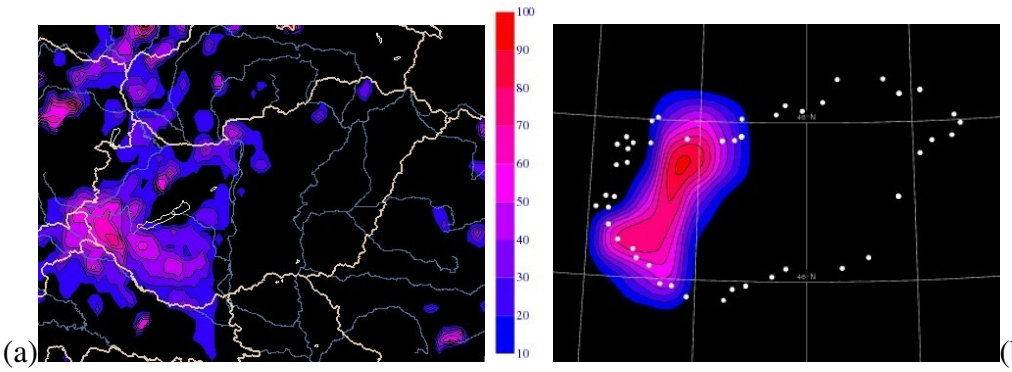

Figure 4.4: Probability for more than 20 mm precipitation in 24 hours based on (a) the LAMEPS system and (b) the ECMWF EPS system. The accumulation period is from 06 UTC, 08 February to 06 UTC, 09 February. (a) LAMEPS run started from 18 UTC, February 07, (b) ECMWF EPS run started from 00 UTC, 08 February.

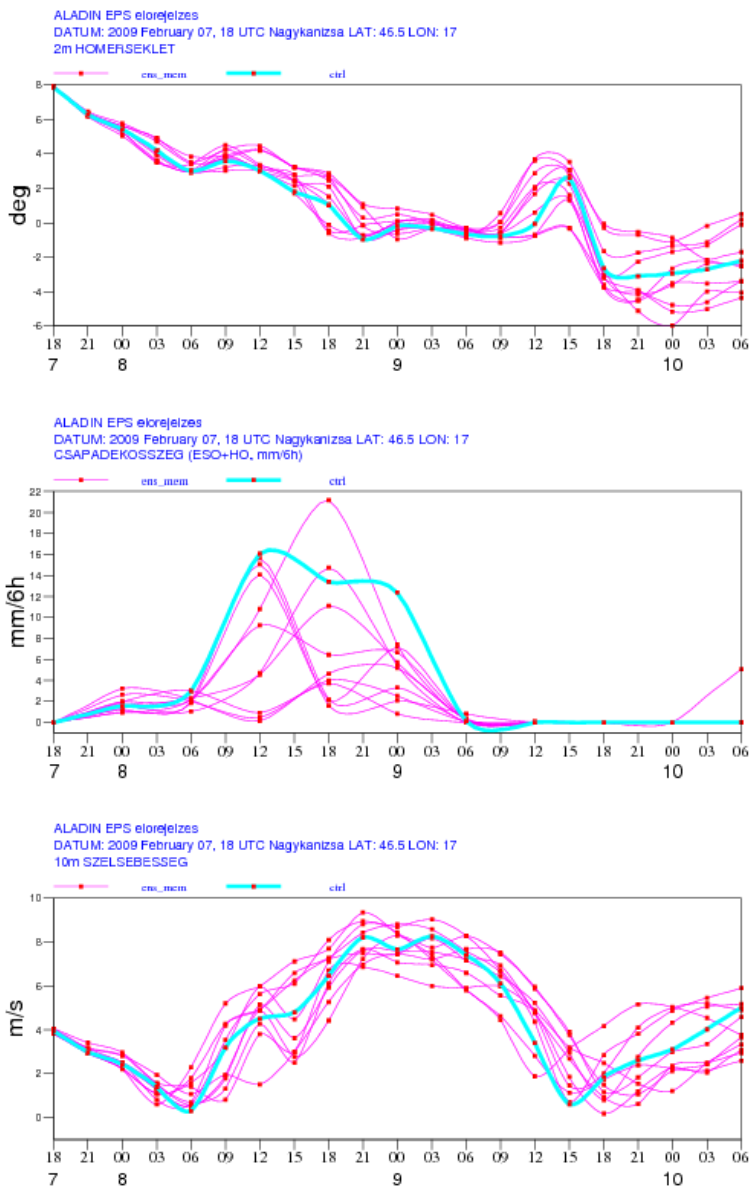

Figure 4.5: Plume diagram for Nagykanizsa based on the LAMEPS forecast started from 18 UTC, 07 February 2009. The diagram displays the time evolution of the distribution of 2 meter temperature (top), total precipitation (middle) and 10 meter wind speed (bottom). Values are plotted for all perturbed ensemble members (pink curves) and the control forecast (blue curve).

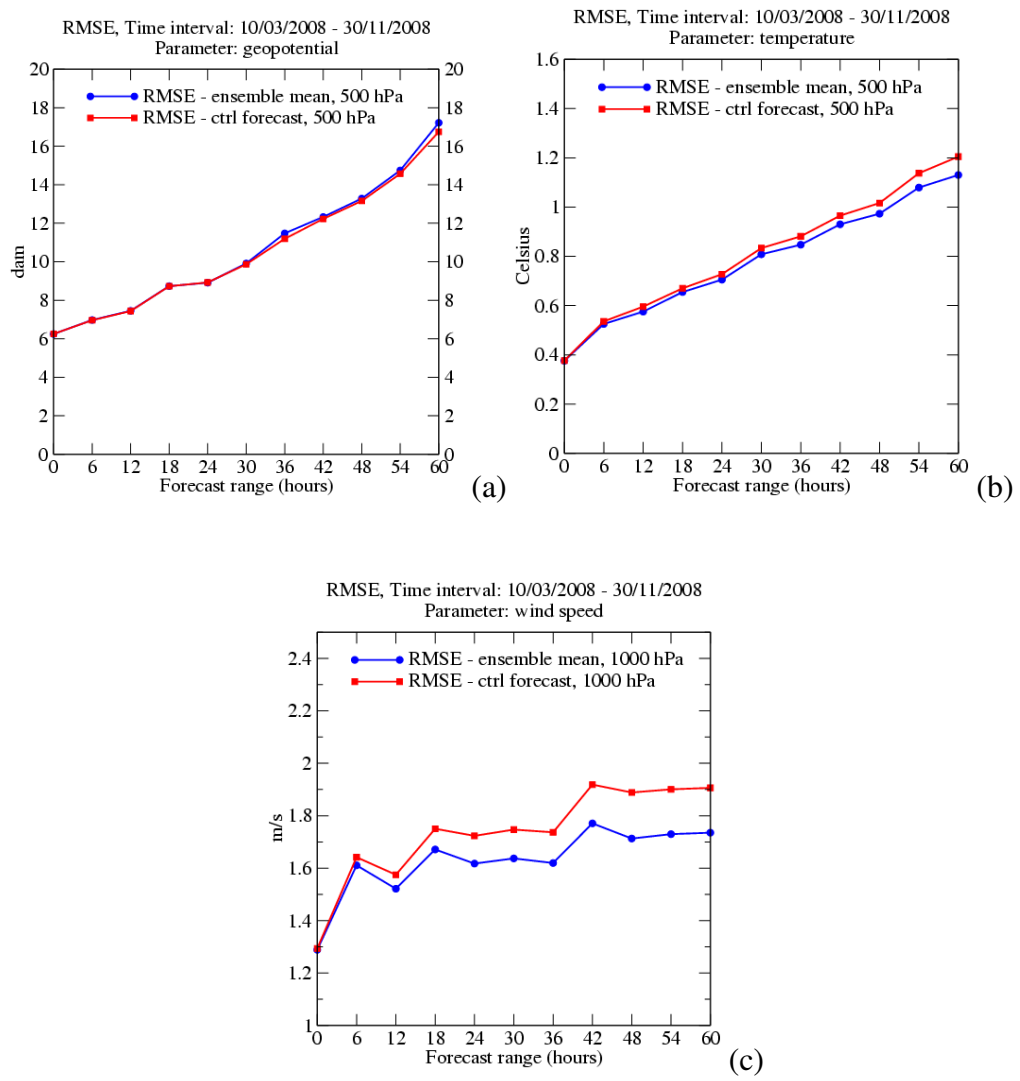

Figure 4.6: RMSE of the ensemble mean (blue lines, circle symbols) and RMSE of the control member (red lines, square symbols) for (a) geopotential at 500 hPa, (b) temperature at 500 hPa and (c) wind speed at 1000 hPa. Verification interval: 10/03/2008 - 30/11/2008. Verification was performed against ECMWF analysis.

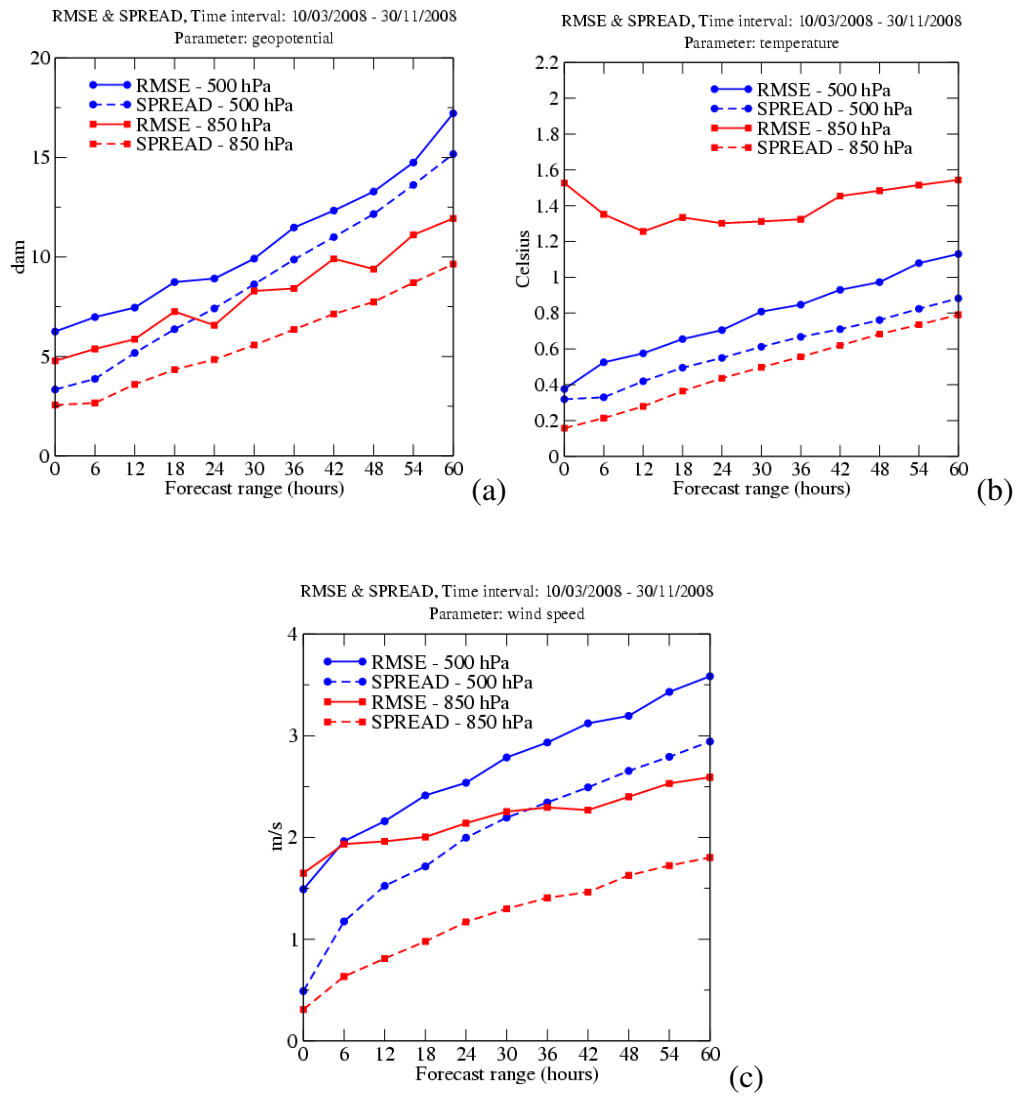

Figure 4.7: RMSE of the ensemble mean (solid lines) and spread of the ensemble (dashed lines) for (a) geopotential, (b) temperature and (c) wind speed at two levels, 500 hPa (blue curves, circle symbols) and 850 hPa (red curves, square symbols). Verification interval: 10/03/2008 - 30/11/2008. Verification was performed against ECMWF analysis.

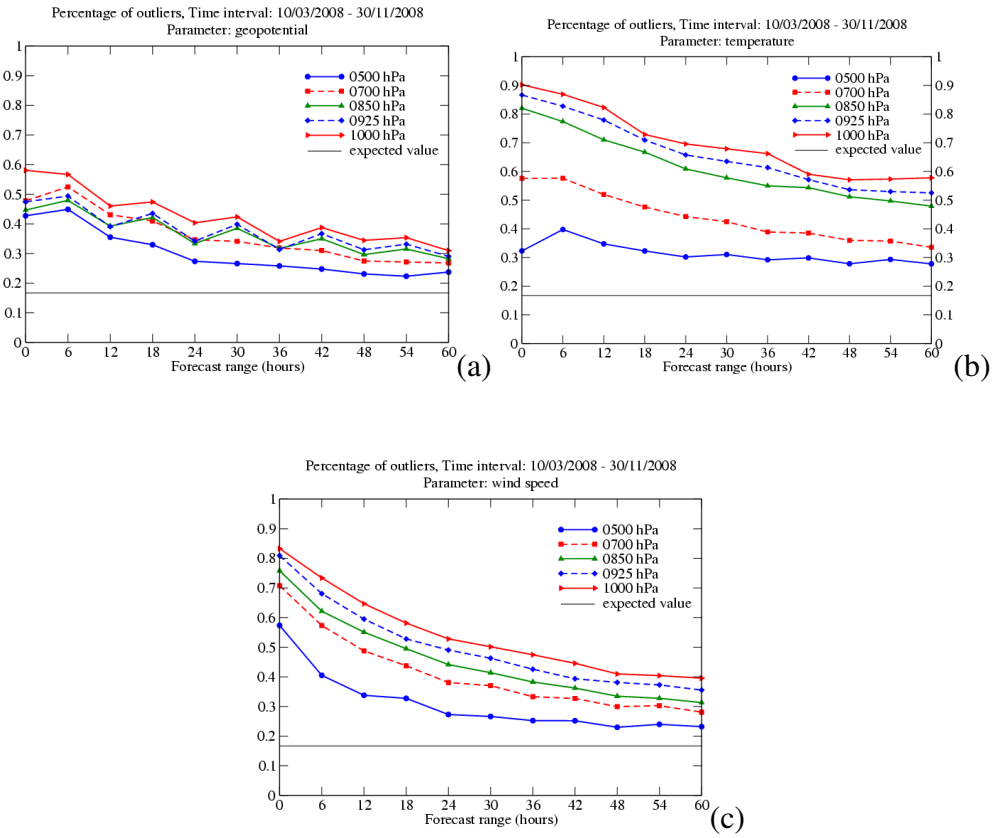

Figure 4.8: Percentage of outliers diagrams for (a) geopotential (b) temperature and (c) wind speed at 500 hPa (blue solid line with circle symbols), 700 hPa (red dashed line with square symbols), 850 hPa (green solid line with triangle symbols), 925 hPa (blue dashed line with diamond symbols) and 1000 hPa (red solid line with triangle symbols). Verification interval: 10/03/2008 - 30/11/2008. Verification was performed against ECMWF analysis. The thin horizontal line is the expected value, i.e. 2/(ensemble members+1).

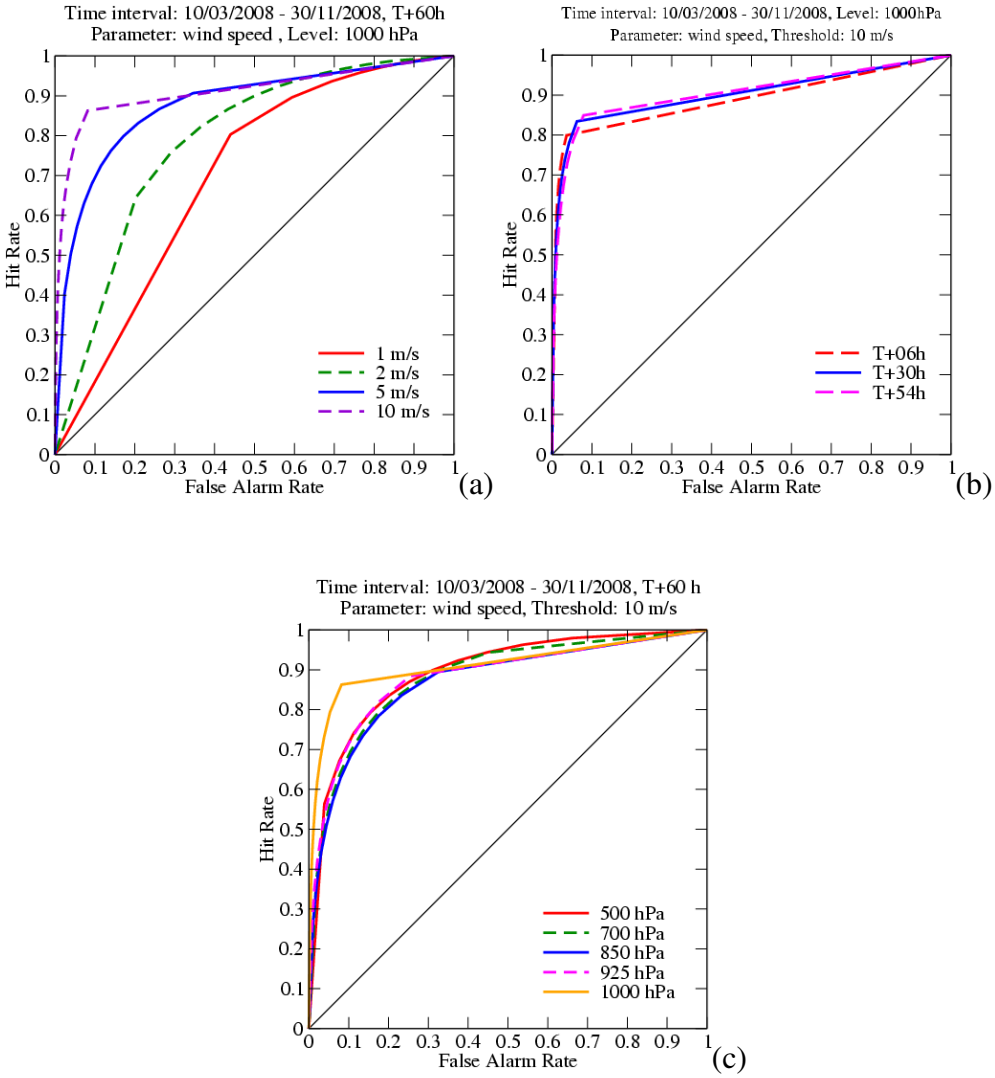

Figure 4.9: ROC diagrams for wind speed. (a) Diagrams for thresholds 1 m/s (solid red line), 2 m/s (dashed green line), 5 m/s (solid blue line), and 10 m/s (dashed purple line) at 1000 hPa and T+60 hours. (b) Diagrams for time range T+06 hours (red dashed line), T+30 hours (blue solid line) and T+54 hours (magenta dashed line) at 1000 hPa and 10 m/s threshold. (c) Diagrams for 500 hPa (red solid line), 700 hPa (green dashed line), 850 hPa (blue solid line), 925 hPa (magenta dashed line) and 1000 hPa (orange solid line) at T+60 hours and 10m/s threshold. Verification interval: 10/03/2008 - 30/11/2008. Verification was performed against ECMWF analysis.

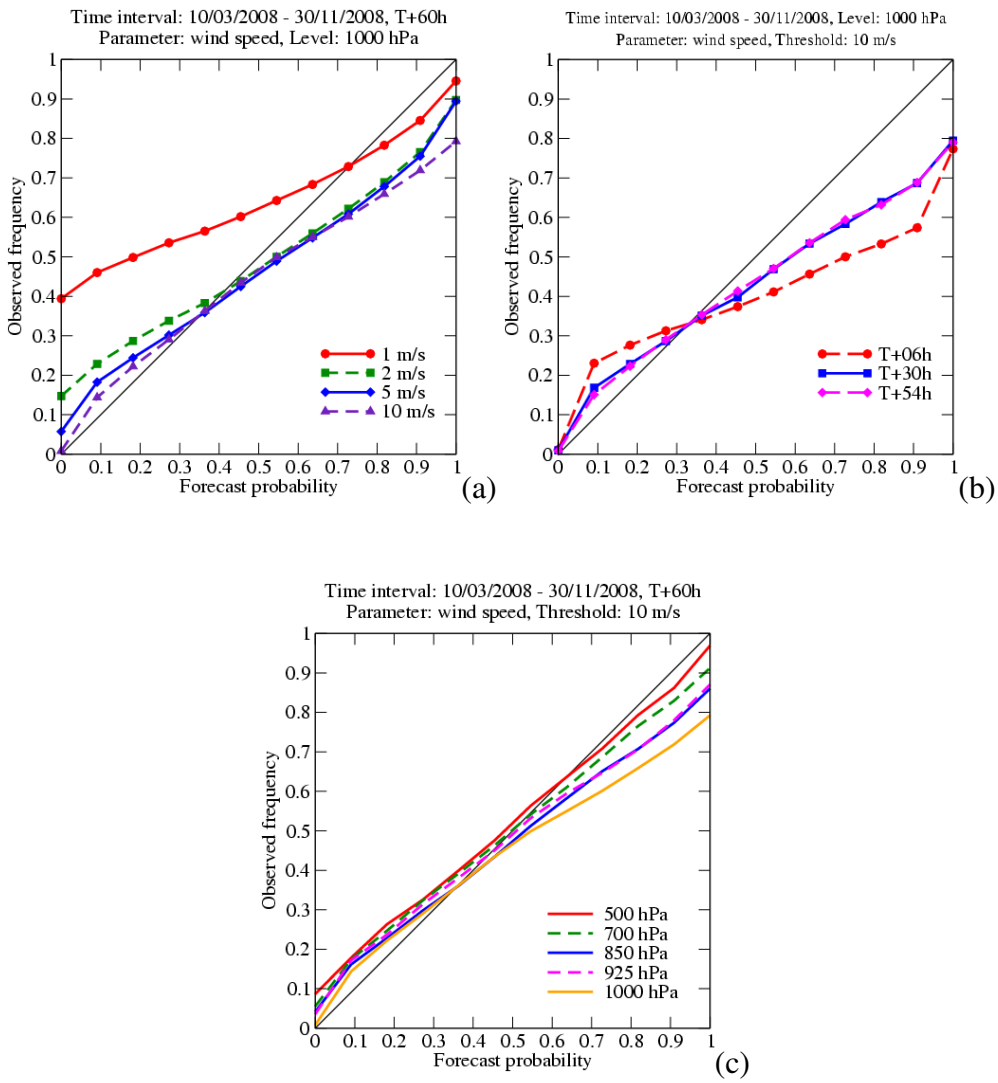

Figure 4.10: Reliability diagrams for wind speed. (a) Diagrams for thresholds 1 m/s (solid red line, circle symbols), 2 m/s (dashed green line, square symbols), 5 m/s (solid blue line, diamond symbols), and 10 m/s (dashed purple line, triangle symbols) at 1000 hPa and T+60 hours. (b) Diagrams for time range T+06 hours (red dashed line, circle symbols), T+30 hours (blue solid line, square symbols) and T+54 hours (magenta dashed line, diamond symbols) at 1000 hPa and 10 m/s threshold. (c) Diagrams for 500 hPa (red solid line), 700 hPa (green dashed line), 850 hPa (blue solid line), 925 hPa (magenta dashed line) and 1000 hPa (orange solid line) at T+60 hours and 10m/s threshold. Verification interval: 10/03/2008 - 30/11/2008. Verification was performed against ECMWF analysis.

Chapter 5

Experiments with ALADIN singular vectors

Verification results of the quasi-operational LAMEPS system (presented in Section 4.3) have indicated the two main lines of future research: (i) computation of local perturbations with the ALADIN model, and (ii) perturbation of surface fields locally. The detailed investigation of these topics were beyond the scope of the present PhD research. However, as it was mentioned before, experiments have already started to compute singular vectors with the ALADIN model. It is expected that using SVs computed with the ALADIN model, targeted to our area of interest, the spread-skill relationship will improve. Additionally, as the scale of the perturbations will be much smaller than in case of the global SVs, the perturbations will have a similar horizontal scale than those weather phenomena that are especially important in short-range forecasting, e.g. convection.

The final aim of this research is to generate perturbations from the ALADIN singular vectors and use them to perturb the initial conditions of the LAMEPS locally. In order to do this, first the important characteristics and the behaviour of the singular vectors have to be understood and investigated in detail. The preliminary results of these experiments are going to be presented in this chapter.

It is important to note, that the singular vector configuration (conf. 601) had not been used for a long time in ALADIN, the last (known) experiments before ours had been made in 2001 ([11]). Therefore technical tests were also needed in

order to check the different steps of the singular vector computation.

All the experiments presented in the following sections were made with cycle 30 of the ALADIN model on the supercomputer of Météo-France (which was first a Fujitsu VPP5000 machine, and later a NEC machine).

5.1 Technical tests

As it was mentioned already in Section 1.5.3, the computation of singular vectors requires the use of the tangent linear and the adjoint codes. Therefore these configurations were tested as a first step of our experiments (Section 5.1.1 and 5.1.2). The purpose of these tests was to check whether the tangent linear code and the adjoint code were working properly in a technical sense.

The number of iterations necessary to obtain a singular value with appropriate precision was also tested (Section 5.1.3).

5.1.1 Test of the tangent linear code

While the direct (nonlinear) model M computes the evolution of the state vector X, the tangent linear model L computes the evolution of a small perturbation x, assuming that the evolution of this perturbation is linear:

$$X(t) = MX(t_0) \tag{5.1}$$

$$x(t) = Lx(t_0) \tag{5.2}$$

In ALADIN, configuration 501 allows us to test the tangent linear code. The way to do that is to compute the following ratio:

$$\frac{M(X+\alpha x) - M(X)}{\alpha L(x)} \tag{5.3}$$

in order to check how well the solution of the tangent linear model (L) approximates the solution of the nonlinear model (M). The above mentioned ratio is computed for different α values, where $\alpha = 10^{-b}$ and b is an integer going from 0 to 10. As α is getting smaller and smaller, this ratio must converge to 1. In

practice, this sequence of numbers gets closest to 1 around b=5. If b is further increased (i.e. α is further decreased), the values start to diverge because of numerical errors ([47]).

5.1.2 Test of the adjoint code

The test of the adjoint code is done by configuration 401, which compares the following two scalar products S_1 and S_2 and checks if they are equal:

$$S_1 = \langle Lx, y \rangle \tag{5.4}$$

$$S_2 = \langle x, L^* y \rangle \tag{5.5}$$

In the equation Lx is obtained after a tangent linear integration from the initial perturbation $x(t_0)$, while L^*y is obtained after an adjoint integration from the perturbation $y(t)$ ([47]).

The results of both 401 and 501 confirmed that the adjoint and the tangent linear codes were working properly in the given cycle of the ALADIN model. Therefore work could continue with the computation of singular vectors (configuration 601).

5.1.3 The number of iterations

As it was mentioned in Section 1.5.3, singular vectors are computed using an iterative method ([5]). The number of iterations determines the accuracy of the computations. The more iterations made, the more singular vectors can be obtained. Tests were performed to determine the number of iterations needed in order to have an acceptable precision of the first few singular values. The characteristics of the experiments were the following. The domain is given by $NLON \times NLAT \times NLEV = 150 \times 135 \times 46$ (the number of points in horizontal and vertical directions, Fig. 2.5) with a horizontal resolution of 20 km. The SV optimization time was set to 12 hours with a timestep of 90 sec. The dry total energy norm was used both at initial and final time.

It was found that in general, one needs three times more iterations than the number of singular vectors desired (Fig. 5.1), e.g. for 20 singular vectors one

needs to perform about 60 iterations. It is in agreement with the results which can be found in the literature (see e.g. [5]).

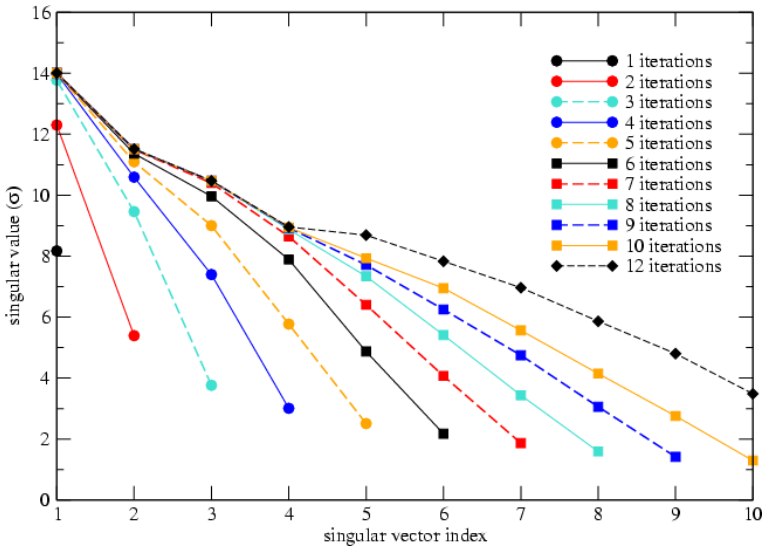

Figure 5.1: Singular values as a function of the number of iterations. Analysing the figure, one can see that for obtaining the first singular value with acceptable precision, 1 or 2 iterations are not enough as the values - denoted by the black and red circles - are different, at least 3 iterations are needed (from 3 iterations onwards, the points coincide). For the second singular value at least 6 iterations are needed. In general one needs three times more iterations than the number of singular values/singular vectors desired.

5.2 Case studies

After the technical tests, "real" experiments could start. Experiments have been performed for two different dates. First one was from 2006 (28 June 2006, starting from the 12UTC analysis) and the second from 2007 (27 August 2007, starting from the 00UTC analysis). In the first case the domain was covering large part of Continental Europe (Fig. 2.5) with 20 km resolution. Optimization of SVs was performed on the whole domain. In the second case the domain was larger (the so-called GLAMEPS[1] domain) with resolutions of 22 and 44 km. However (due

[1]GLAMEPS is a joint HIRLAM-ALADIN project for short-range limited area ensemble forecasting.

to the high computational costs), optimization of SVs was only performed on a subdomain (which was the domain used for the first case study).

5.2.1 Case study #1: 28 June 2006

The characteristics of the first case study were the following. The domain is given by $NLON \times NLAT \times NLEV = 150 \times 135 \times 46$ with a horizontal resolution of 20 km. The SV optimization time was set to 12 hours with a timestep of 90 sec. The optimization area was the whole domain (Fig. 2.5). The dry total energy norm was used both at initial and final time. Computations were started from the 12 UTC analysis on 28 June 2006. Lateral boundary conditions were obtained from ARPEGE every 3 hours. Singular values are plotted on Fig. 5.2. The leading singular value was around 14. As it was shown in Section 1.5.3, the singular value gives us the amplification of the perturbation from t_0 to t.

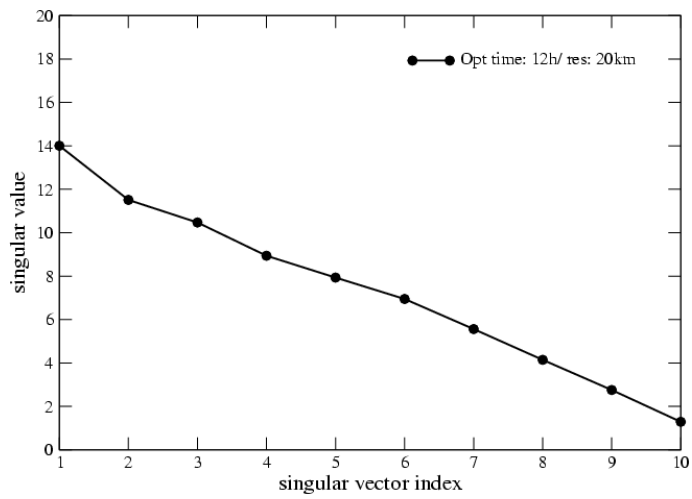

Figure 5.2: Singular values obtained in the experiment started on 28 June 2006, 12 UTC.

For comparison, SVs were computed with the global ARPEGE model as well. The optimization time and area was the same as for ALADIN, but the resolution was different, a truncation of T95 (approx. 210 km) was used.

The synoptic situation is shown in Fig. 5.3 at 00 UTC, 29 June 2006. This is the time the singular vectors were optimized to.

Temperature fields on model level 32 (about 727 hPa) are shown for both models on Fig. 5.4 and 5.5 (for ARPEGE only at initial time, for ALADIN both at initial and final time). One can realize that the location is quite similar at initial time for both models but the values and the area covered are different.

Comparing the leading ALADIN SV at initial and final time it can be realized that the structure is slightly moved to the northeast during the evolution of the SV and also the area covered became somewhat larger, but still very localized. The evolution of the SVs is important, because later - if a LAMEPS system is based on this method - this will determine the spread of the ensemble system. If the SVs (and the perturbations generated from them) are very localized in space, then the ensemble members are going to be very similar to each other and to the unperturbed control forecasts.

It can also be noted that the area of maximum amplitude at final time is located around Sardinia, at the same place where a low pressure zone can be observed on Fig. 5.3.

Energy distribution was plotted separately for the wind and the temperature component of the total energy (surface pressure part was not included in the computations). Fig. 5.6 reveals that at initial time the total energy is dominated by the temperature component, while at final time the wind component is more dominant. It can also be mentioned that the total energy propagates rather upwards than downwards. The maximum of the energy was around model level 30-32 (660-727 hPa). This seems to be in agreement with the values that can be found in the literature about the behaviour of global SVs (e.g. [26]).

5.2.2 Case study #2: 27 August 2007

For this experiment a larger integration domain was used with two different resolutions: 22 km and 44 km. The domains can be given by $NLON \times NLAT \times NLEV = 320 \times 300 \times 46$ for 22 km and $NLON \times NLAT \times NLEV = 160 \times 150 \times 46$ for 44 km. Two different optimization times were used: 12 hours and 24 hours, both with a timestep of 90 sec. The optimization area was not covering the whole domain, it was the same as in the previous experiment. The dry total energy norm was used both at initial and final time. Computations were started from 00 UTC analysis

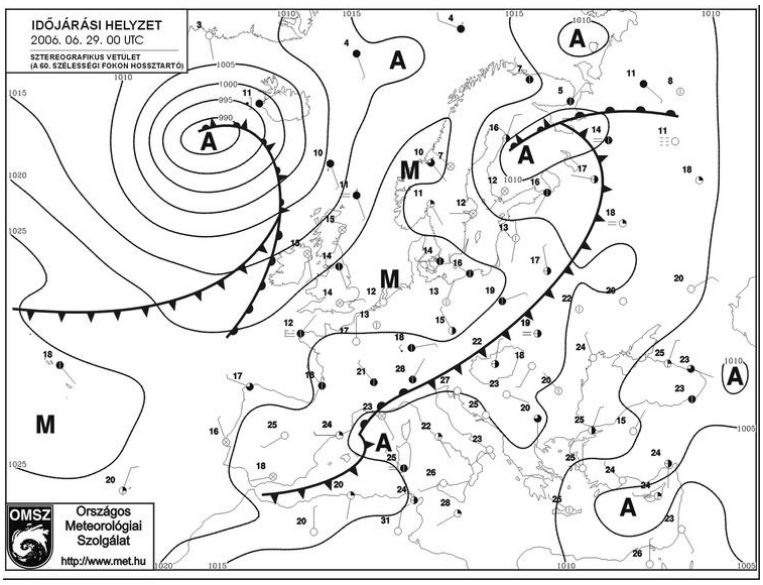

Figure 5.3: Synoptic situation at 00 UTC on 29 June 2006.

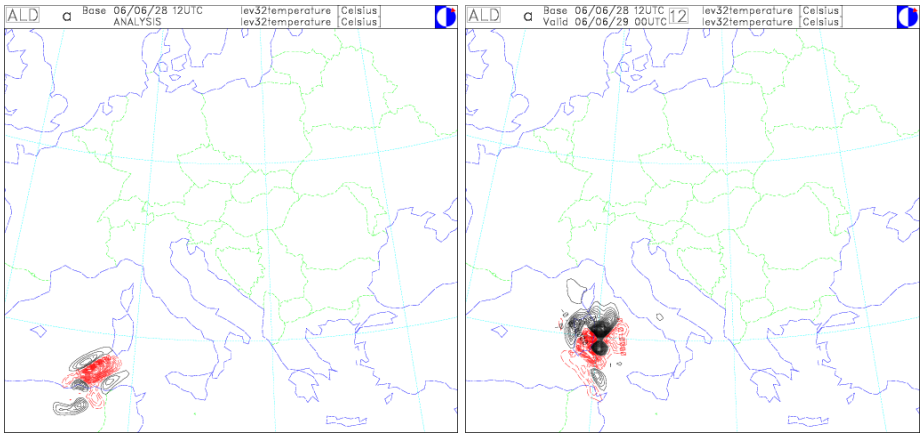

Figure 5.4: Temperature component (at model level 32) of the ALADIN leading singular vector at initial (left) and final (right) time started from 12 UTC, 28 June 2006. Contour interval: 0.01 Celsius. Resolution used for the computations was 20 km and the optimization time was 12 hours.

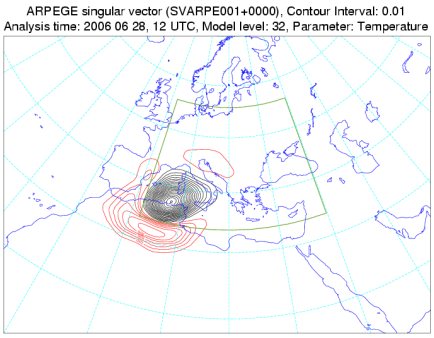

Figure 5.5: Temperature component (at model level 32) of the ARPEGE leading singular vector at initial time started from 12 UTC, 28 June 2006. Contour interval: 0.01 Celsius. Truncation used for the computations was T95 and the optimization time was 12 hours. Optimization area was the same as for the ALADIN model, shown in green in the figure.

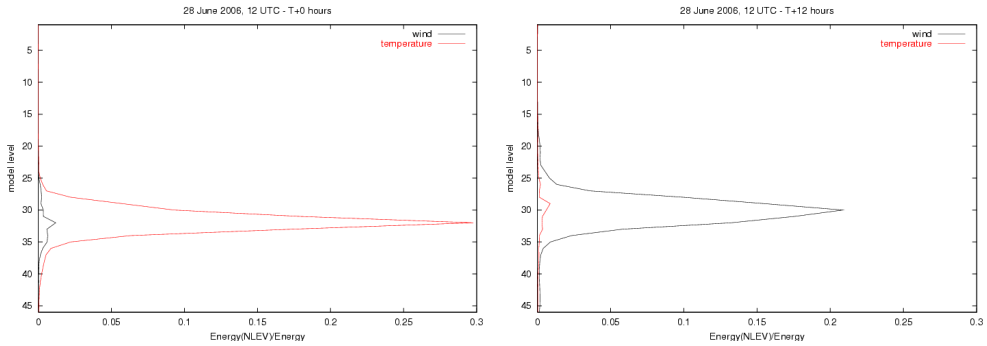

Figure 5.6: Vertical energy distribution of the leading SV started from 12 UTC, 28 June 2006. Wind (black) and temperature (red) component of the total energy is plotted at initial (left) and final (right) time. Energy of each model level is normalized with the total energy of all levels. The optimization time was 12 hours and the resolution was 20 km.

on 27 August 2007. Lateral boundary conditions were obtained from ARPEGE every 3 hours.

Energy distribution was plotted separately for the wind and the temperature component of the total energy (surface pressure part was not included in the computations). Figures reveal that at initial time the total energy is dominated by the temperature component, while at final time the wind component is more dominant (Fig. 5.8 and 5.9). It can also be mentioned that the total energy propagates rather upwards than downwards. Compared to the previous case study, it should be mentioned that the maximum of the energy was much higher in the second case, around model level 20. This suggests that the distribution of the energy depends very much on the synoptic situation. Further case studies will be made to investigate this issue in more detail.

Singular vectors were also plotted for the different experiments performed with the use of different resolution and optimization time (Fig. 5.10, 5.11, 5.12 and 5.13). Results suggest that the difference in the resolution on which the SVs were computed does not have a large effect on the structure of the SVs. On the other hand there is a difference in the singular values (Fig. 5.7). It was found that if the resolution is higher, the singular values are larger as well.

The difference in the optimization time has the effect of changing the location of the SVs. With 24 hours optimization time SVs are located more to the west at initial time compared to those computed with 12 hours optimization time. There is a difference at final time as well. 24 hours SVs cover a considerably larger area at final time than 12 hours SVs.

One can also note from the plots of the SVs - in agreement with the energy distributions - that at the initial time the temperature fields have larger values, while at final time the wind components are more dominant.

5.3 Further plans with the ALADIN SVs

As presented in this chapter, experiments have started with the computation of ALADIN singular vectors. The final goal is to generate initial condition perturbations from these singular vectors and use them to perturb the initial conditions of the LAMEPS locally. However, there are still a lot of issues to examine in detail,

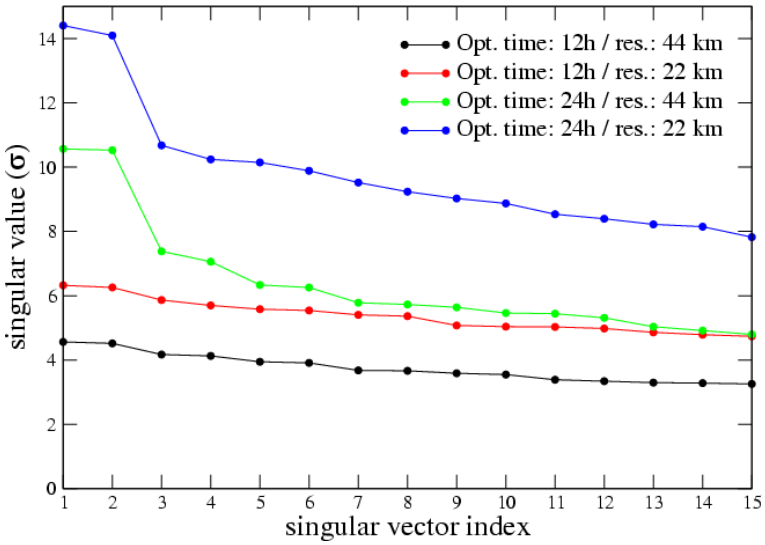

Figure 5.7: Singular values of the first 15 singular vectors started from 00 UTC, 27 August 2007, with different optimization times (12 and 24 hours) and different resolutions (22 and 44 km).

since from the two case studies only preliminary conclusions could be drawn. The most important questions to investigate are:

- The *optimization area and time* to be used for the singular vector computation. From the results presented in Chapter 3 it is clear, that the proper choice of the SV optimization area and time can significantly improve the skill of an ensemble system.

- The *resolution* to be used for the computation of singular vectors. The SV calculation is computationally very demanding, therefore it is important to find a compromise between the resolution and the CPU time available.

- What *norm* should be used for the SV computation (e.g. total energy norm, CAPE norm, etc)? This is an important issue as the structure of the SVs (both horizontally and vertically) can be significantly different depending on the applied norm. The most commonly used norm for global SV computations is the total energy norm, however, it should be investigated in detail what is the most appropriate choice for short-range limited area model perturbations.

- The *number of SVs* to be used to generate the IC perturbations and the *way to build the perturbations* from the SVs.

Work with ALADIN singular vectors is ongoing at present and the future experiments will focus on the above mentioned questions.

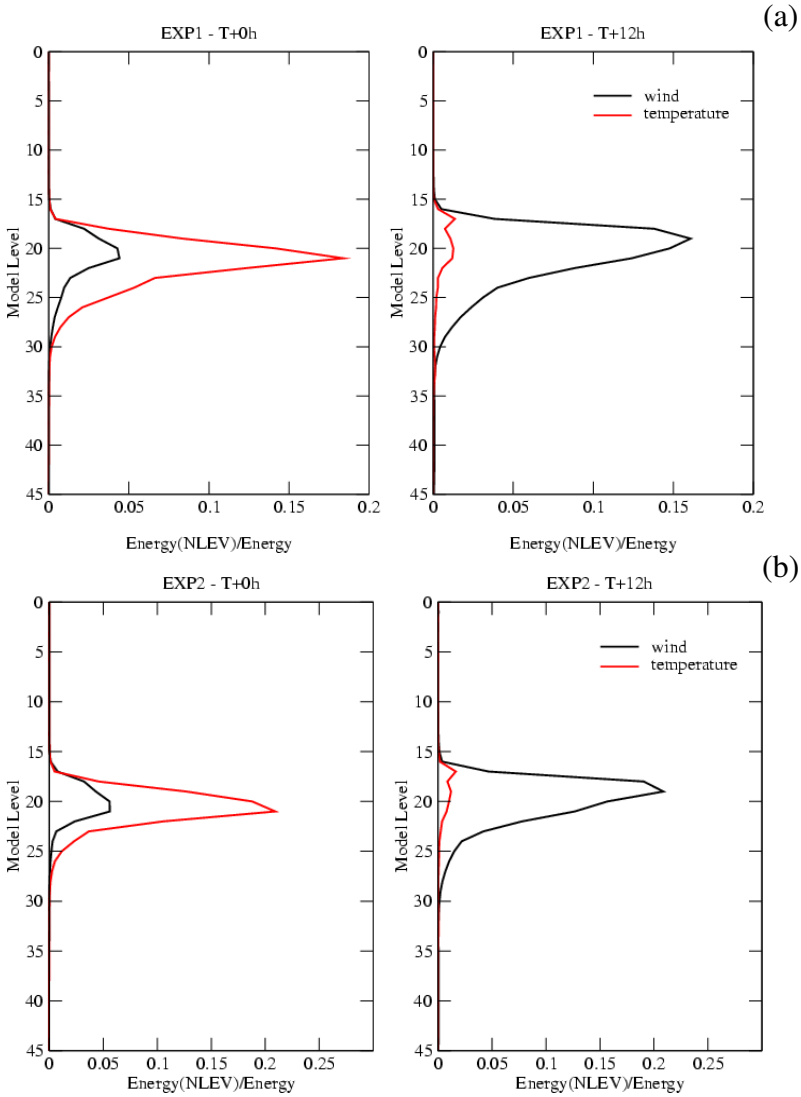

Figure 5.8: Vertical energy distribution of the leading SV started from 00 UTC, 27 August 2007. Wind (black) and temperature (red) component of the total energy are plotted separately both for initial (left panels) and final (right panels) time. Energy of each model level is normalized by the total energy of all levels. (a) Optimization time was 12 hours and resolution was 44 km. (b) Optimization time was 12 hours and resolution was 22 km.

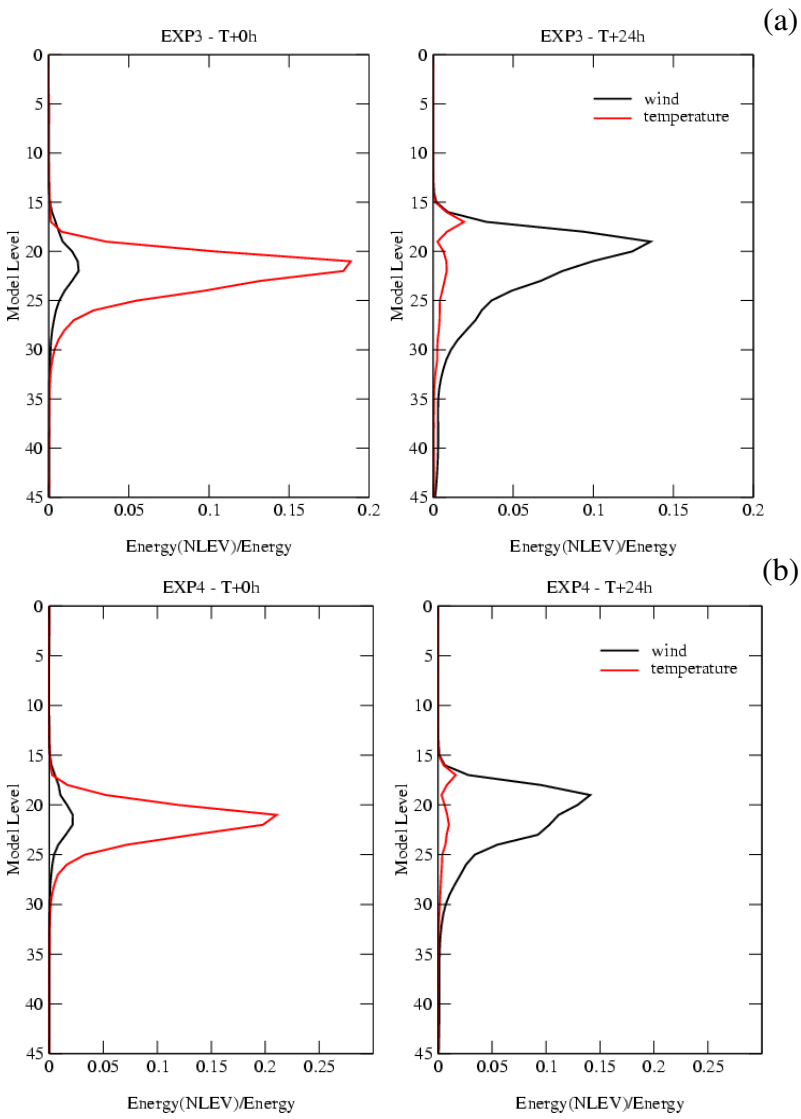

Figure 5.9: Vertical energy distribution of the leading SV started from 00 UTC, 27 August 2007. Wind (black) and temperature (red) component of the total energy are plotted separately both for initial (left panels) and final (right panels) time. Energy of each model level is normalized by the total energy of all levels. (a) Optimization time was 24 hours and resolution was 44 km. (b) Optimization time was 24 hours and resolution was 22 km.

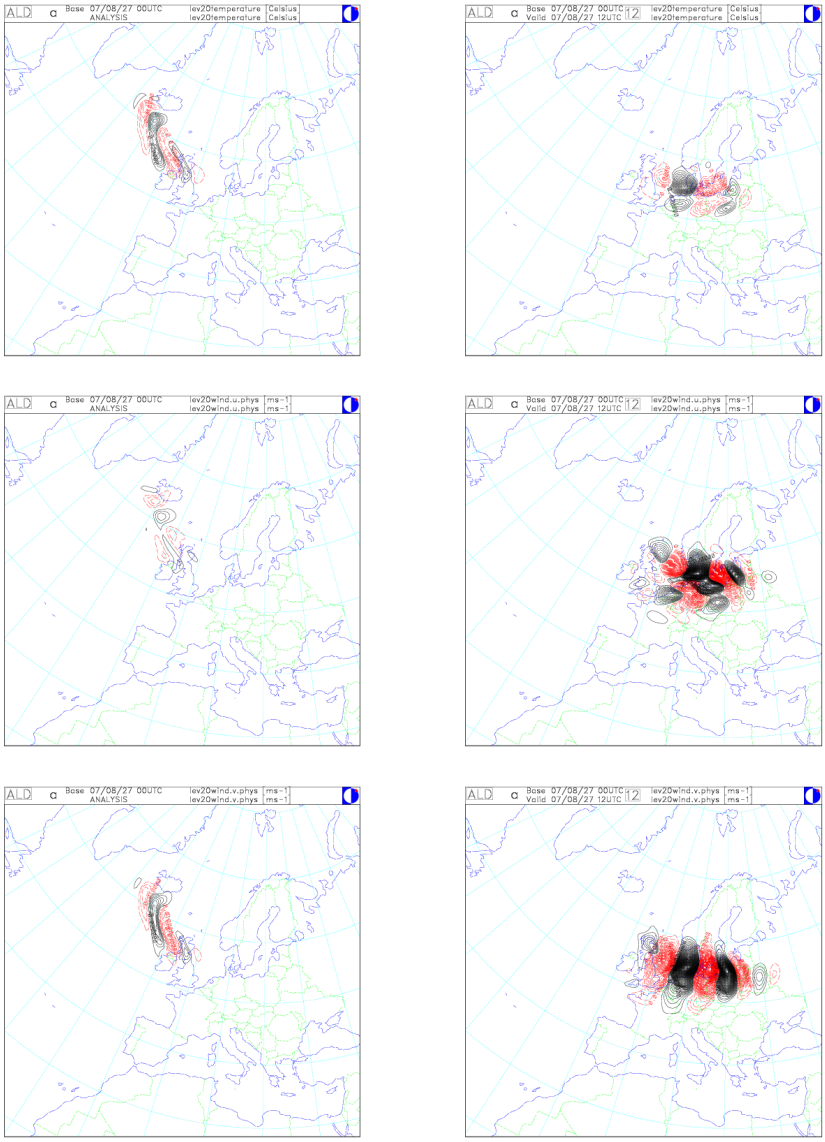

Figure 5.10: Temperature (top row), u (middle row) and v (bottom row) component of the leading singular vector (at model level 20) at initial (left column) and final (right column) time started from 00 UTC, 27 August 2007. Contour interval: 0.01 Celsius for temperature and 0.01 m/s for the wind components. Resolution used for computations was 44 km and the optimization time was 12 hours.

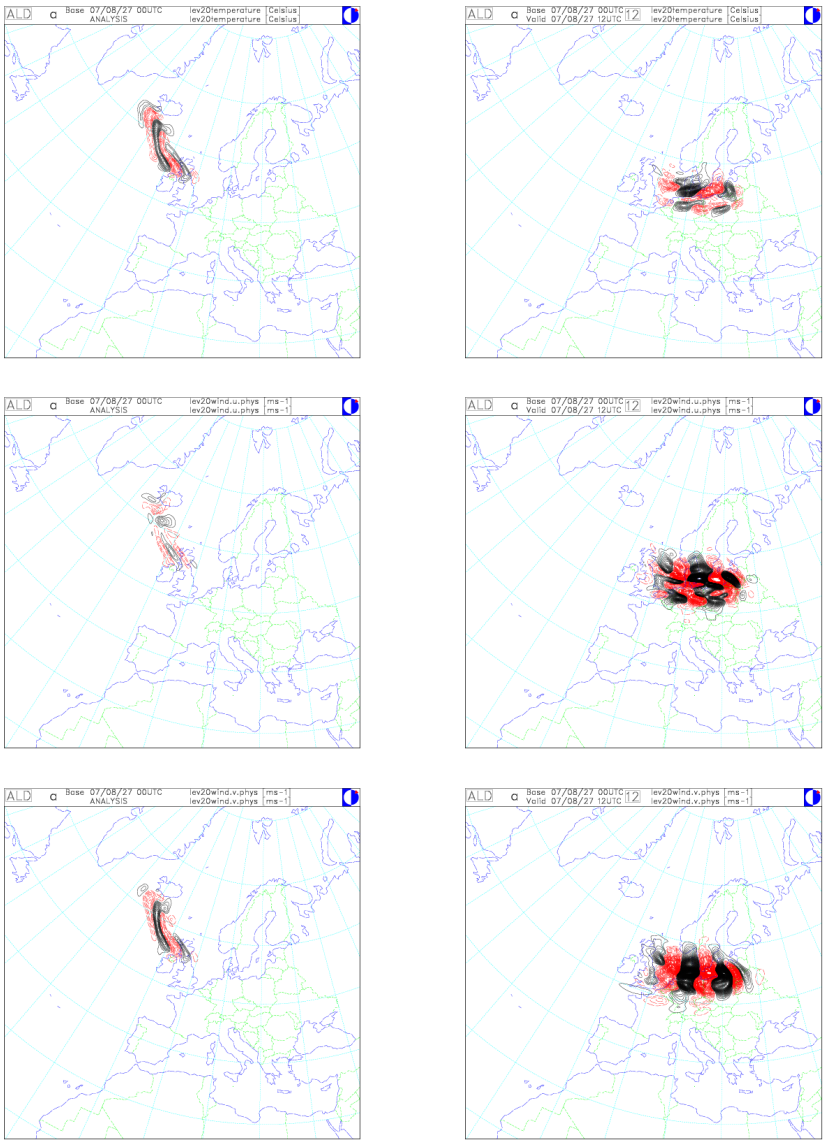

Figure 5.11: Temperature (top row), u (middle row) and v (bottom row) component of the leading singular vector (at model level 20) at initial (left column) and final (right column) time started from 00 UTC, 27 August 2007. Contour interval: 0.01 Celsius for temperature and 0.01 m/s for the wind components. Resolution used for computations was 22 km and the optimization time was 12 hours.

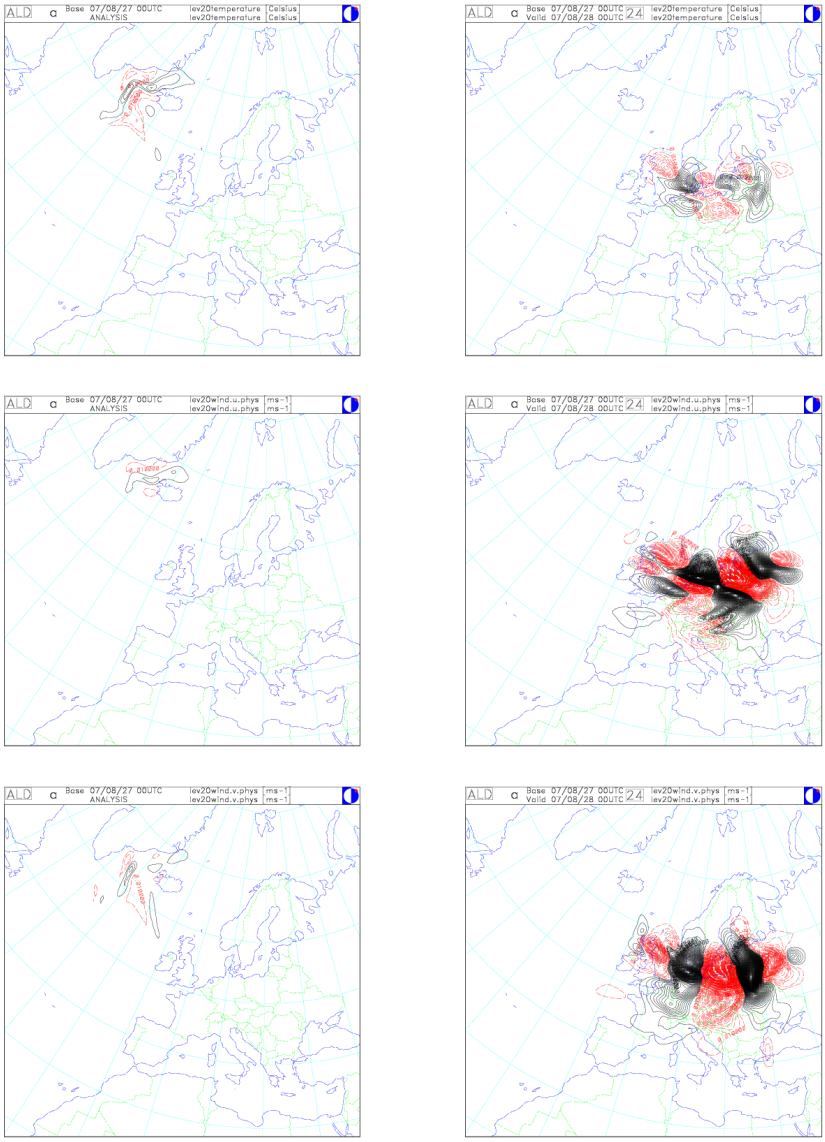

Figure 5.12: Temperature (top row), u (middle row) and v (bottom row) component of the leading singular vector (at model level 20) at initial (left column) and final (right column) time started from 00 UTC, 27 August 2007. Contour interval: 0.01 Celsius for temperature and 0.01 m/s for the wind components. Resolution used for computations was 44 km and the optimization time was 24 hours.

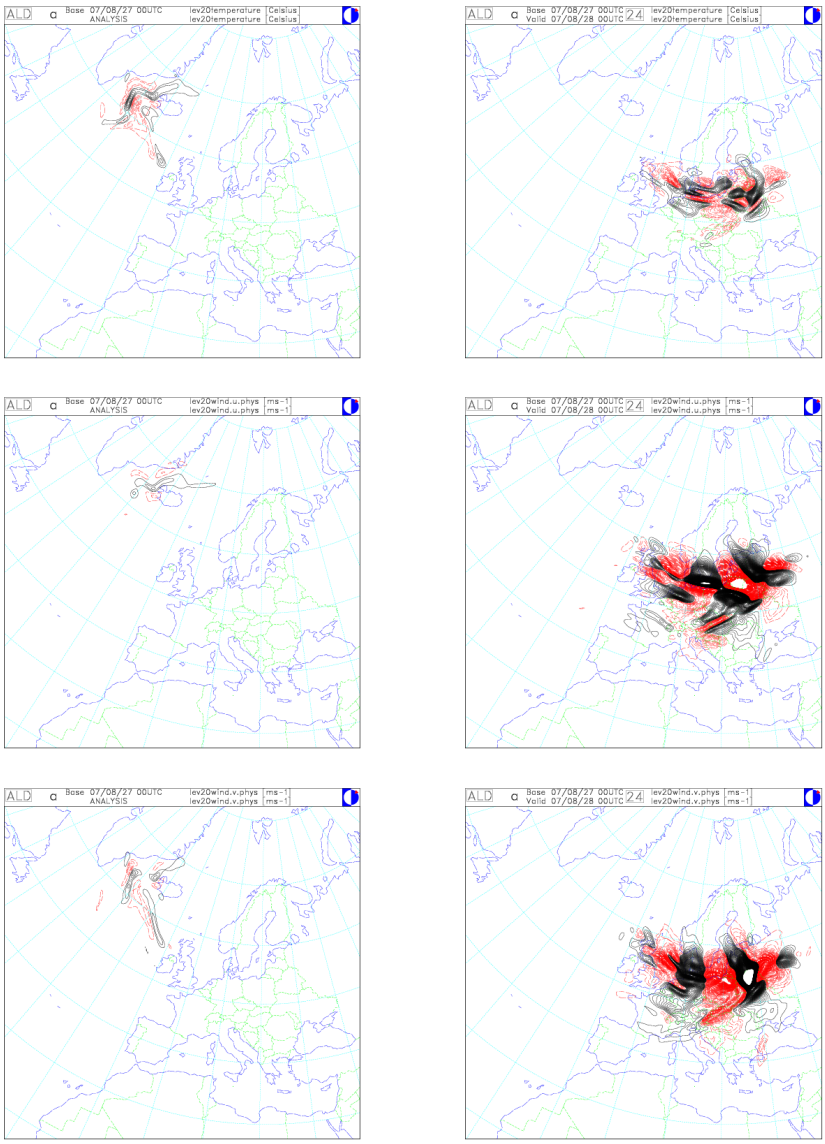

Figure 5.13: Temperature (top row), u (middle row) and v (bottom row) component of the leading singular vector (at model level 20) at initial (left column) and final (right column) time started from 00 UTC, 27 August 2007. Contour interval: 0.01 Celsius for temperature and 0.01 m/s for the wind components. Resolution used for computations was 22 km and the optimization time was 24 hours.

Conclusions

In the thesis I have dealt with the development, investigation and finally the operational application of a short-range limited area ensemble prediction system based on the ALADIN model, using global ARPEGE ensemble forecasts as initial and lateral boundary conditions. The work presented had three main parts: (i) sensitivity studies with global singular vectors with respect to their optimization area and optimization time, (ii) experiments with limited area singular vectors, and (iii) the development and (quasi-) operational application of a LAMEPS based on the dynamical downscaling of the global PEARP system. Hereafter the most important results of these three main parts are going to be summarized.

As a first step of the PhD work sensitivity studies were performed in order to see whether or not it was possible to optimize the existing ARPEGE based global ensemble system (PEARP, formerly PEACE) for the Central European area by changing the optimization domain and time used for the global singular vector computations. Global ensemble forecasts were made with the ARPEGE model and were downscaled with the ALADIN model. Verification results confirmed that the proper choice of the singular vector optimization area and time can increase the spread and can improve the skill of the forecasts for the area of our interest (i.e. Central Europe and particularly Hungary). Verification results of the global and the limited area systems were also compared in order to see whether the limited area model can improve the predictions of the global one. It was found that it is very difficult to achieve significant overall improvement by simply downscaling the global EPS system with the limited area model. However, one should not forget, that it is a common phenomenon that high resolution models might perform worse than the low resolution ones when usual verification measures are

applied. The reason of this is the so-called double penalty problem. Therefore the results of the comparison should be interpreted with care.

Based on the results of the sensitivity studies it is believed that the computation of local perturbations is needed in the limited area model for properly addressing the small-scale initial uncertainties of the atmosphere, which are not present in the global model. Therefore, as a second step, it was decided to continue the work in the field of singular vectors computed with the ALADIN limited area model. The final aim of this research is to generate perturbations from the ALADIN singular vectors and use them to perturb the initial conditions of the LAMEPS locally. Experiments were made using different horizontal resolutions (22 km and 44 km) and different optimization times (12 hours and 24 hours) for the singular vector computation. The preliminary results suggest that the difference in the horizontal resolution does not have a strong effect on the (horizontal) structure of the singular vectors, but affects the singular values. The higher the resolution, the larger the singular values. On the other hand, the difference in the applied optimization time has the effect of changing the location of the singular vectors both at initial and final time. Singular vectors computed with a longer optimization time cover a considerably larger area at the end of the optimization period than those computed with shorter optimization. As mentioned already, the work with limited area singular vectors has not been finished yet, there are still a lot of issues to examine in detail.

Meanwhile, in parallel with the above mentioned research activities, a short-range limited area ensemble system - based on the ALADIN model - was put into operations at HMS in order to gain experience not only from case studies and test periods, but on a day-to-day, real-time basis. At present the only operationally feasible solution was the direct downscaling of the PEARP members, therefore this method is applied. Analysing the verification results of almost nine months (from 10 March 2008 to 30 November 2008) it was found that better results can be obtained for higher levels. The reason of this is related to the characteristics of the perturbations that are applied in the PEARP system. This behaviour suggest that surface perturbations should also be included in the system. One can also

conclude that even for higher levels the spread of the ensemble is smaller than the RMSE of the ensemble mean and the percentage of outliers is above the expected value. This indicates that local perturbations (e.g. the use of limited area singular vectors), targeted especially to the area of our interest would be needed to improve the quality of the system. Thus, the quasi-operational LAMEPS system of HMS is going to be developed and improved continuously, using the results of the ongoing researches.

Appendix A

Visualisation

In case of an ensemble system one has to keep in mind that the amount of information is much larger than for a single deterministic forecast. This information has to be visualised and presented to the forecasters in a manageable way. Visualising the ensemble members one by one (as for a single deterministic forecast) is not a proper solution in case of a system with e.g. 50+1 members as the ECMWF EPS. Therefore several different types of diagrams have been developed throughout the years. These include maps of the ensemble mean and the ensemble spread, visualisation of the median, probability maps for different parameters and thresholds, meteograms and plume diagrams, spaghetti diagrams, etc. The most commonly used visualisation methods are presented in more detail.

A.1 Ensemble members, "stamp maps"

Similar to the case of single deterministic forecasts, ensemble members can be displayed one by one. However, in case of large number of ensemble members (e.g. the 50+1 member ensemble system of ECMWF) it is very time consuming to go through all the individual members. One possible alternative is the "stamp map" (Fig. A.1). For a given lead time and a given parameter all ensemble members can be visualised (on one page), of necessity only in a small size. With the use of such a diagram it is very easy to realise the significant differences between the ensemble members, but because of the small size details cannot be examined.

A.2 Ensemble mean, median, ensemble spread

One possible way of reducing the number of outputs is to visualise not all of the ensemble members but only the ensemble mean. However, one should take care when using the ensemble mean: e.g. let us consider a case when half of the ensemble members predict low pressure over Hungary, while the other half predict high pressure. The ensemble mean will be a featureless pressure field. In such a case the use of the median (the middle of a distribution) can be more relevant. While the ensemble mean is a computed field, the median is one of the forecasts produced by the ensemble system, thus it can be more realistic.

The ensemble mean and the median tell us nothing about the uncertainty in the system. For that the ensemble spread can be used, which is a good indicator of the uncertainty. In case of a reliable system, large ensemble spread indicates less predictability, while small spread indicates small uncertainty. The ensemble mean and the ensemble spread are often visualised together (Fig. A.2).

A.3 Probability maps

Probability maps show the probability of a given event, e.g. precipitation above/below a certain threshold. The indication of e.g. 70% for 24 hour precipitation more than 20 mm means that this value was exceeded 7 out of 10 times in a 10-member ensemble system (Fig. A.3).

A.4 Meteogram, plume diagram

Meteograms and plume diagrams both display the time evolution of the ensemble members for several parameters (e.g. total precipitation, temperature, wind speed) for a given location. One can observe how predictability (and also uncertainty of the forecast) changes with time. At ECMWF both diagrams are used for the visualisation of the EPS forecasts (Fig. A.4 and Fig. A.5). The interpretation of these

diagrams is as follows: in case of the plume diagram values are plotted for all ensemble members, the control forecasts and the high resolution ("operational") forecast as well. Shading with different colours indicates the number of ensemble members falling to a given interval (Fig. A.5). The idea behind the meteogram is a bit different. Instead of plotting a curve for each ensemble member, only the control and the high resolution ("operational") forecast is plotted. The ensemble members are represented in forms of box-diagrams. From each box one can determine the median, the 10th, 25th, 75th and 90th percentiles, the maximum and the minimum. High boxes indicate large spread, therefore less predictability, while low boxes indicate small spread and small uncertainty (Fig. A.4).

A.5 Spaghetti diagrams

For a given lead time and parameter spaghetti diagrams show a chosen isoline for all ensemble members. While the plume diagram and the meteogram give information about the uncertainty in time, the spaghetti diagram does the same in space (Fig. A.6).

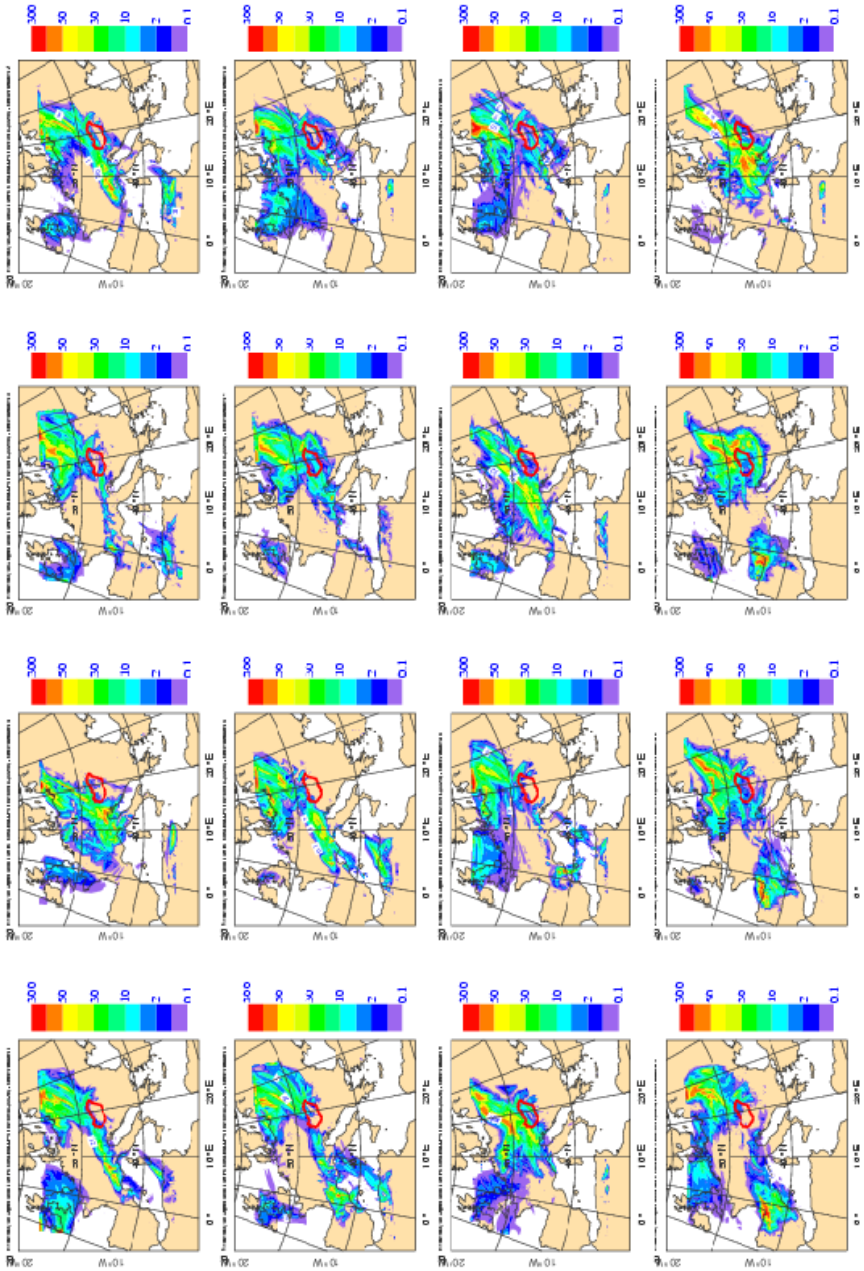

Figure A.1: Stamp diagram of the COSMO-LEPS ensemble system for 24 hours precipitation. Forecast started on 20/08/2008, at 12 UTC. Verification time is T+66 hours - T+90 hours. The COSMO-LEPS system is using the ECMWF EPS forecasts as ICs and LBCs. Not all the 50+1 EPS members are being downscaled by the COSMO model, only the so-called representative members which are selected through clustering ([34]).

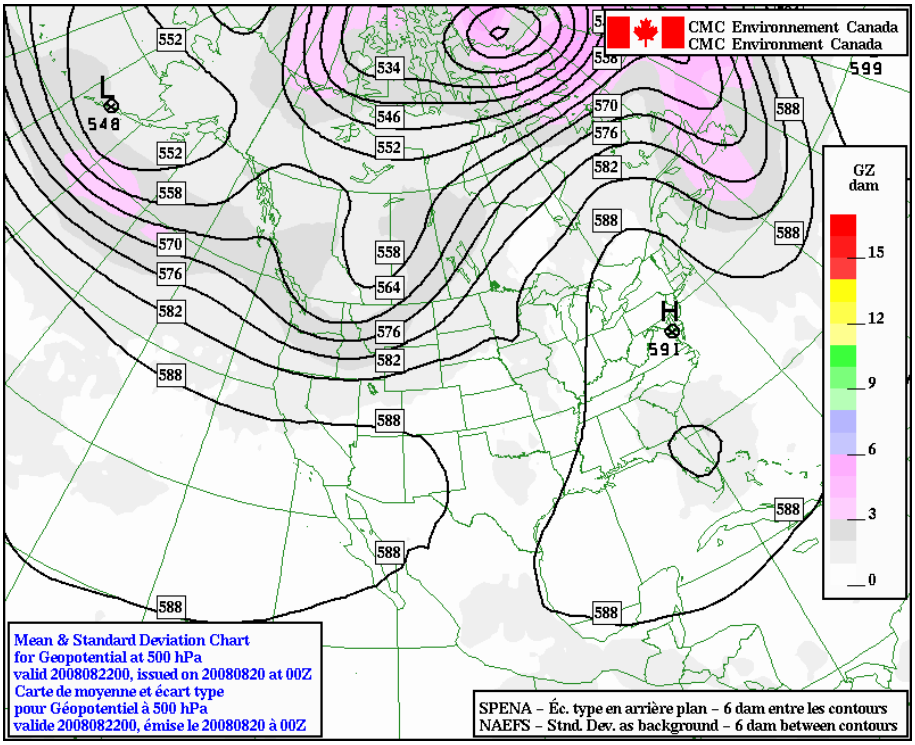

Figure A.2: Ensemble mean (isolines) and standard deviation (shading) for 500 hPa geopotential height plotted on the same chart. Forecast started on 20/08/2008, at 00 UTC. Verification time is T+48 hours. (Source of figure: Canadian Meteorological Centre)

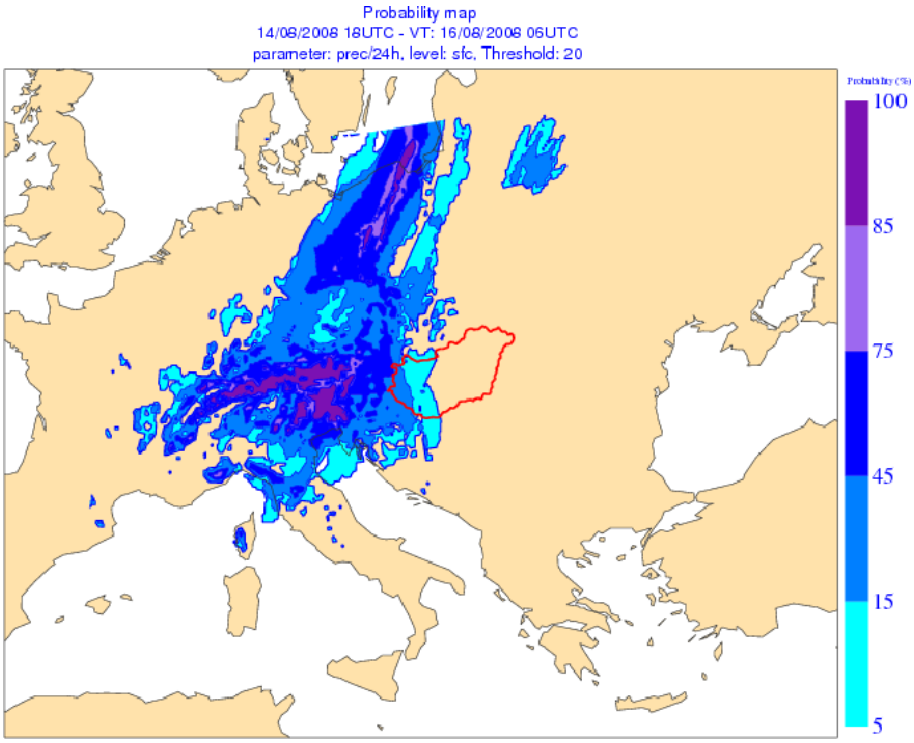

Figure A.3: Probability map for 24 hours total precipitation from the ALADIN EPS system of HMS. Forecast started on 14/08/2008, at 18 UTC. Verification time is T+12 hours - T+36 hours. Threshold is 20 mm.

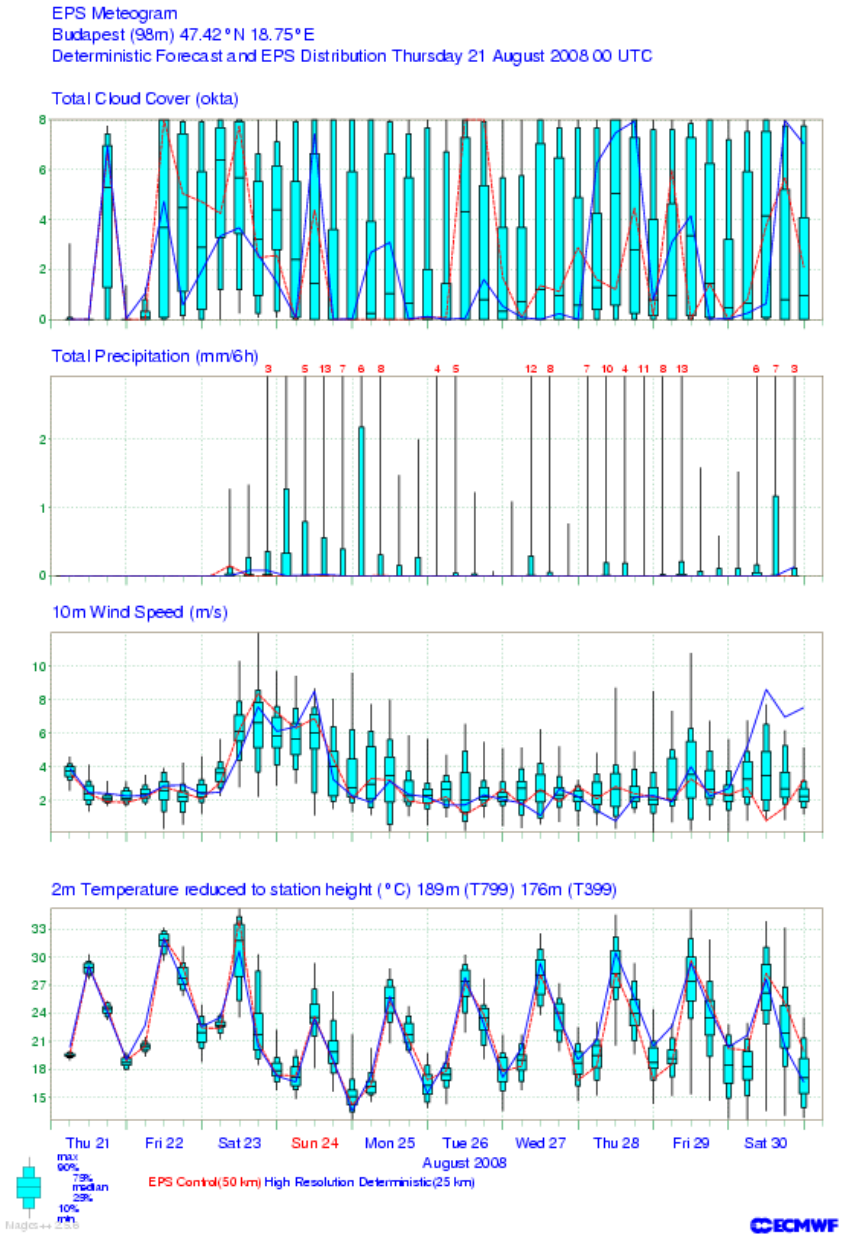

Figure A.4: Meteogram for Budapest based on the ECMWF EPS forecast started on 21/08/2008, 00 UTC. It displays the time evolution of the distribution of total cloud cover, total precipitation, 10 meter wind speed and 2 meter temperature. The control and the high resolution ("operational") forecast is plotted with the red and blue curves respectively. The ensemble members are represented in forms of box-diagrams which give the median, the 10th, 25th, 75th and 95th percentiles, the maximum and the minimum. (Source of figure: www.ecmwf.int)

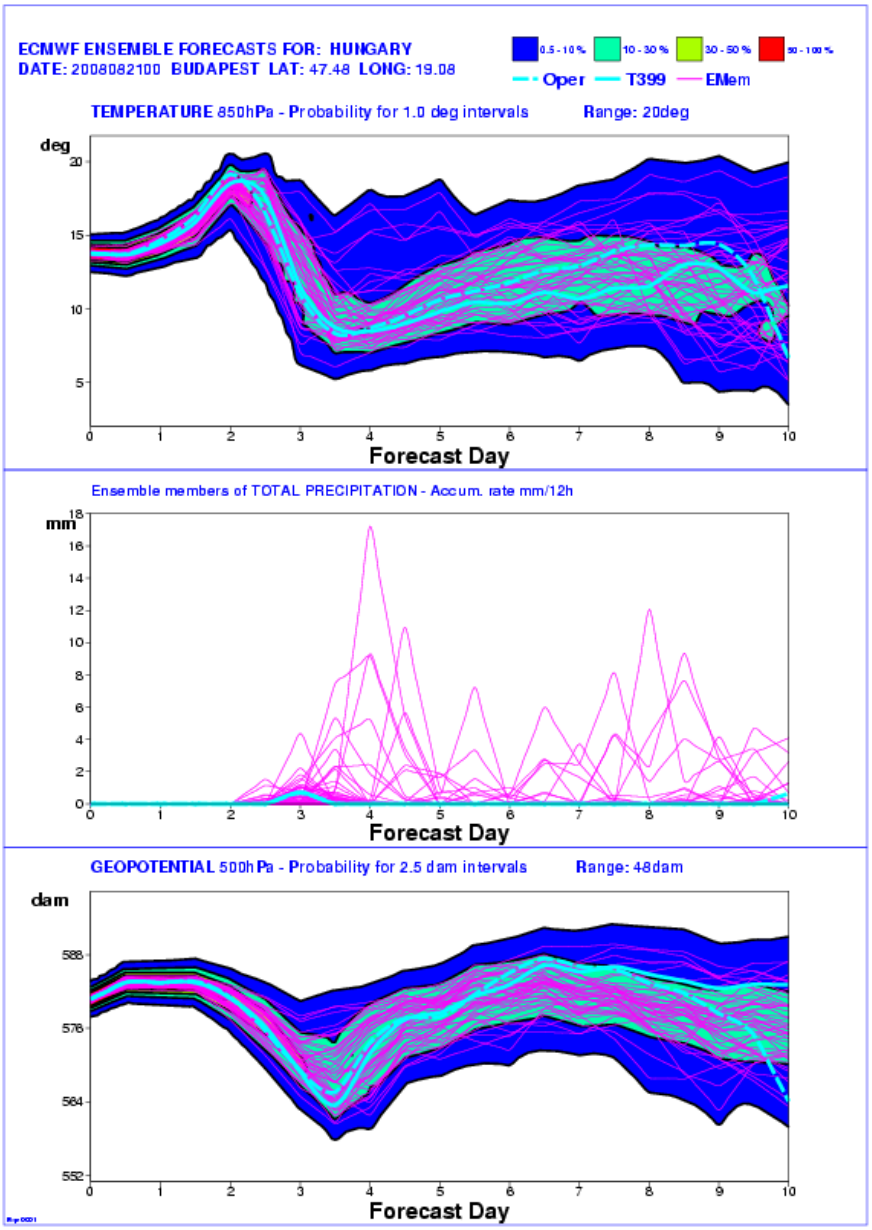

Figure A.5: Plume diagram for Budapest based on the ECMWF EPS forecast started on 21/08/2008, 00 UTC. It displays the time evolution of the distribution of 850 hPa temperature, total precipitation and 500 hPa geopotential height. Values are plotted for all ensemble members, the control forecasts and the high resolution ("operational") forecast as well. Shading with different colours indicates the number of ensemble members falling to a given interval (Source of figure: www.ecmwf.int)

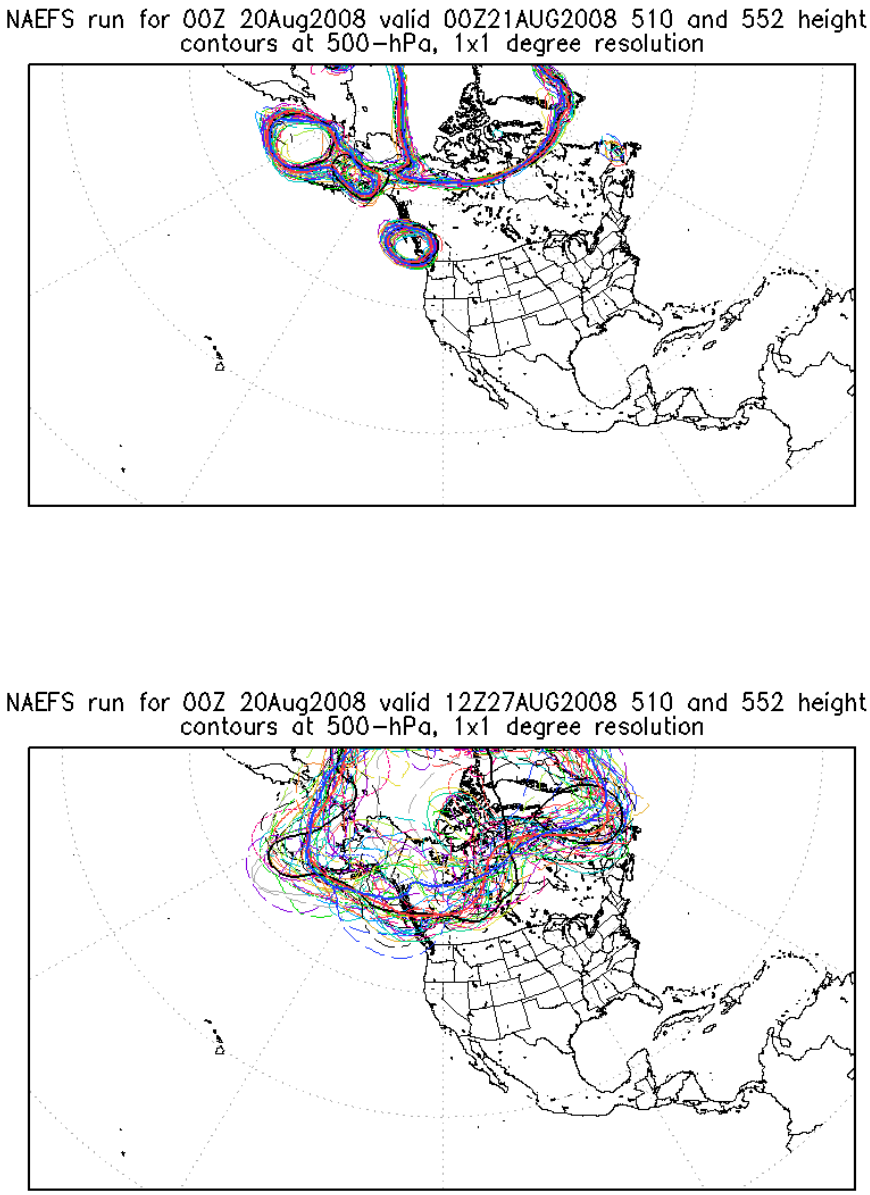

Figure A.6: Spaghetti diagram for the isolines 510 dam and 552 dam of 500 hPa geopotential height. Forecast started on 20/08/2008, at 00 UTC. Verification time is T+24 hours (top) and T+180 hours (bottom). (Source of figure: NCEP, http://www.emc.ncep.noaa.gov/gmb/ens/)

Appendix B

Verification

As for the verification, ensemble systems can be verified using the usual verification scores of the deterministic forecasts. The *systematic error* and the *root mean squared error* can be computed separately for the individual ensemble members, for the control member (if exists) and also for the ensemble mean. Using these scores (together with the *ensemble spread*) different verification diagrams can be plotted to help the evaluation of the ensemble system.

Besides the deterministic type of scores probabilistic verification can be performed as well. *Ranked histograms*, *percentage of outliers diagrams*, *ROC diagrams* and *reliability diagrams* are the most common measures used in the literature ([25], [46]). These verification methods are described in the following.

B.1 Systematic error (BIAS)

Computing the *systematic error* or BIAS is a way to measure the difference between forecasted and observed values averaged over space and time:

$$BIAS = \frac{1}{N}\sum_{i=1}^{N}(fc_i - obs_i) \tag{B.1}$$

In the equation fc stands for the forecasted value, obs is for the observation and N is the number of cases. The range of the BIAS goes from minus infinity to infinity with a perfect value of zero. It should be noted however, that the value of zero

can be reached even if the dataset has large errors, since the positive and negative values can compensate each other.

B.2 Root Mean Squared Error (RMSE)

The *root mean squared error* also measures the difference between forecasted and observed values, but in this case - in contrast to the formulation of the BIAS - the positive and negative values cannot compensate each other:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(fc_i - obs_i)^2} \tag{B.2}$$

In the equation fc stands for the forecasted value, obs is for the observation and N is the number of cases. The range of the RMSE goes from zero to infinity with a perfect value of zero (when the forecasted and observed values are equal for all N cases).

B.3 Ensemble spread

The ensemble spread can be calculated in different ways. One possibility is to compute the spread around the ensemble mean, the other way is to compute it around the control member (if such a member exists for the given ensemble system). In our experiments, the ensemble spread was always computed around the ensemble mean:

$$SPREAD = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{MEM}\sum_{j=1}^{MEM}(fc_{ij} - \overline{fc_i})^2} \tag{B.3}$$

In the equation fc_{ij} stands for the forecasted value, $\overline{fc_i}$ is the ensemble mean, MEM is the number of members in the ensemble system and N is the number of cases.

B.4 Visualisation of deterministic kind of scores

The BIAS and the RMSE can be calculated for the individual ensemble members and also for the ensemble mean. One possible way to analyse the RMSE values of an ensemble system is to look at the relationship between the RMSE of the control member (i.e. the forecast started from the unperturbed initial condition in our case) and the RMSE of the ensemble mean. If the perturbations in the examined ensemble system are symmetric around the unperturbed initial condition (as they are in PEARP, hence in the LAMEPS system of HMS as well), the ensemble mean and the control forecast are almost identical in the early forecast ranges. This means that their RMSE is also very similar. However, after the initial linear phase it is expected that the ensemble mean has lower RMSE values than the control forecast since the averaging has the effect of filtering out the less predictable features and leaving only the more predictable ones that show agreement among the ensemble members.

Another important feature of an ensemble system is the spread-skill correspondence. The spread of the ensemble system (computed around the ensemble mean in our case) should be in good agreement with the forecast error (e.g. RMSE of the ensemble mean). If the spread is larger (smaller) than error the system is said to be over- (under-) dispersive.

B.5 Ranked histogram, percentage of outliers

If all ensemble members are considered as equally likely realisations of the atmospheric state, then the verifying observation (or analysis) is equally likely to lie between any two ordered members of the ensemble, including the cases when it lies outside (on either side) the interval defined by the ensemble members. By accumulating the number of cases over space and time this can be transformed into a diagram called *ranked histogram*.

If the system is reliable or statistically consistent, the distribution is close to flat. Different shapes indicate different behaviour. U shape indicates that the spread in the ensemble system is not sufficient, the verifying observation (or analysis) lies outside the ensemble too often. The opposite of U shape (bell shape)

indicates too large spread. Assimetrical distributions (J and L shape) indicate negative or positive bias (Fig. B.1).

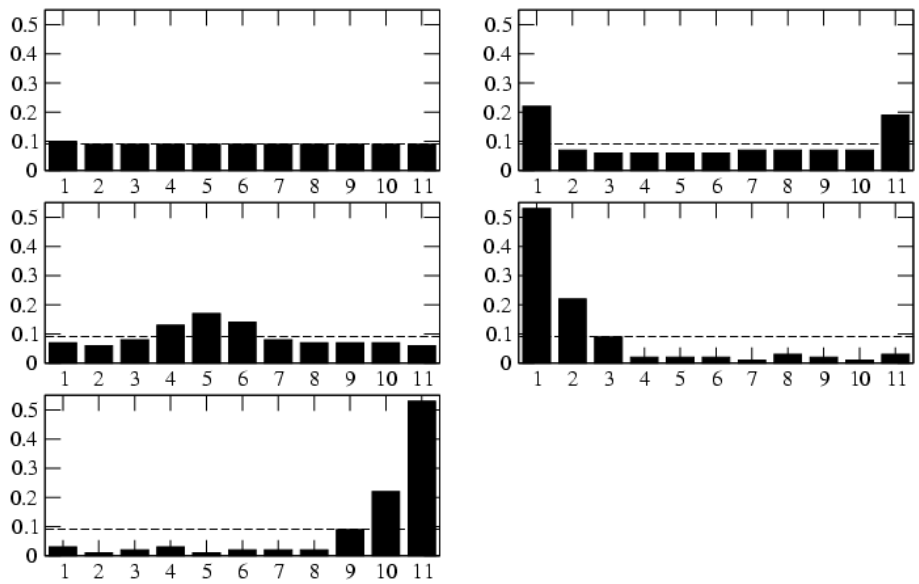

Figure B.1: Five hypothetical ranked histograms for a 10-member EPS system. Flat distribution (top left panel) indicates sufficient spread. U shaped distribution (top right panel) indicates that the spread in the ensemble system is not sufficient, while bell shape (middle left panel) indicates too large spread. L shape (middle right panel) and J shape (bottom left panel) indicate positive and negative bias, respectively. The horizontal line is the expected value, i.e. $1/(\text{number of ensemble members+1})$.

The *percentage of outliers diagram* is a way to summarize the information of the *ranked histogram* for all forecast steps. The sum of the two outermost intervals of the *ranked histogram* is plotted against the forecast step. If the system is reliable, the *percentage of outliers* is equal to $2 \times 1/(\text{number of ensemble members+1})$.

B.6 ROC diagram, ROC area

Given a certain threshold (e.g. 10 meter wind speed > 15 m/s) and a certain probability of occurence from which we consider the event forecasted, (e.g. 70%) probabilistic forecasts can be transformed into categorical yes/no forecasts. From these categorical forecasts a table called *contingency table* can be built (Tab. B.1).

The number of cases in each category are H as *Hit* (the event was forecasted and observed), F as *False alarm* (the event was forecasted but not observed), M as *Missed event* (the event was not forecasted but it was observed) and Z as *Zero forecast* (the event was not forecasted and not observed either). H+F+M+Z=N where N is the total number of cases. From the values of the different categories in the contingency table the *False Alarm Rate* (FAR) and *Hit Rate* (HR) can be calculated. *Hit Rates* and *False Alarm Rates* can be computed for different probability values from 0% to 100% and plotted against each other on the ROC diagram.

Table B.1: Contingency table for categorical forecasts.

	observation: "yes"	observation: "no"
forecast: "yes"	H	F
forecast: "no"	M	Z

$$FAR = \frac{F}{F+Z} \tag{B.4}$$

$$HR = \frac{H}{H+M} \tag{B.5}$$

Four basic kinds of the ROC diagrams can be distinguished (Fig. B.2). The first case is when the ROC area equals 1 (upper left panel on Fig. B.2). All points of the ROC curve are in the upper left corner with HR=1 and FAR=0. The second case (upper right panel on Fig. B.2) represents real life ensemble systems with a ROC curve well above the diagonal. The ROC area is between 1.0 and 0.5 and HR>FAR for all probabilities. For low probabilities the points are closer to the (1,1) point, while for higher probabilities they are closer to the (0,0) point. If the ROC area is 0.5, i.e. HR=FAR for all probabilities (bottom left panel on Fig. B.2) using the ensemble system does not give any added information compared to the use of the climatological mean. Finally, if the ROC curve is well below the diagonal, i.e. HR<FAR for all probabilities (bottom right panel on Fig. B.2), it means that there is less value in the forecast than in the climatological mean.

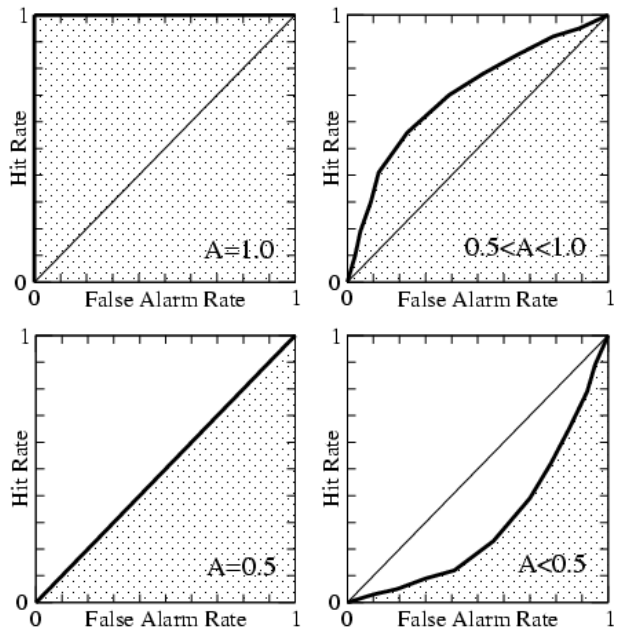

Figure B.2: Four hypothetical ROC diagrams showing *Hit Rate* as the function of *False Alarm Rate*. The perfect case (upper left panel): ROC area equals 1, all points are in the upper left corner. Realistic case: ROC area between 1 and 0.5 (upper right panel). If the ROC area is 0.5 (bottom left panel) the ensemble system does not give any added information compared to the use of the climatological mean. Finally if the ROC area is below 0.5 (bottom right panel) it means that there is less value in the forecast than in the climatological mean.

B.7 Reliability diagram

When an event is forecasted with a given probability, then (on average) it should occur with the same frequency. The *reliability diagram* is used to test the ability of the system to correctly forecast probabilities of a certain event. On the reliability diagram the observed frequency is plotted as the function of the forecast probability (Fig. B.3). If the forecast probabilities and the observed frequencies agree, the curve lies along the diagonal, i.e. on the line of *"Perfect reliability"*. Additional lines can also be plotted, i.e. the *"No resolution"* and the *"No skill"* lines. Forecasts which have their points lying on the *"No resolution"* line are not able to resolve cases when the event is less or more likely than the climatological probability. The *"No skill"* line determines a region, in which the points contribute

positively to the forecast skill. If all the points are out of this region the forecast has no skill at all. The number of forecasts per forecast probability can also be shown on the diagram (numbers written next to the points, like on Fig. B.3). If all forecast probabilities have the same number of points then the system is not able to distinguish between high and low probability and has no sharpness. Sharpness diagrams can be plotted in addition to reliability diagrams, as shown on the examples of Fig. B.4, where several hypothetical reliability diagrams are shown and explained.

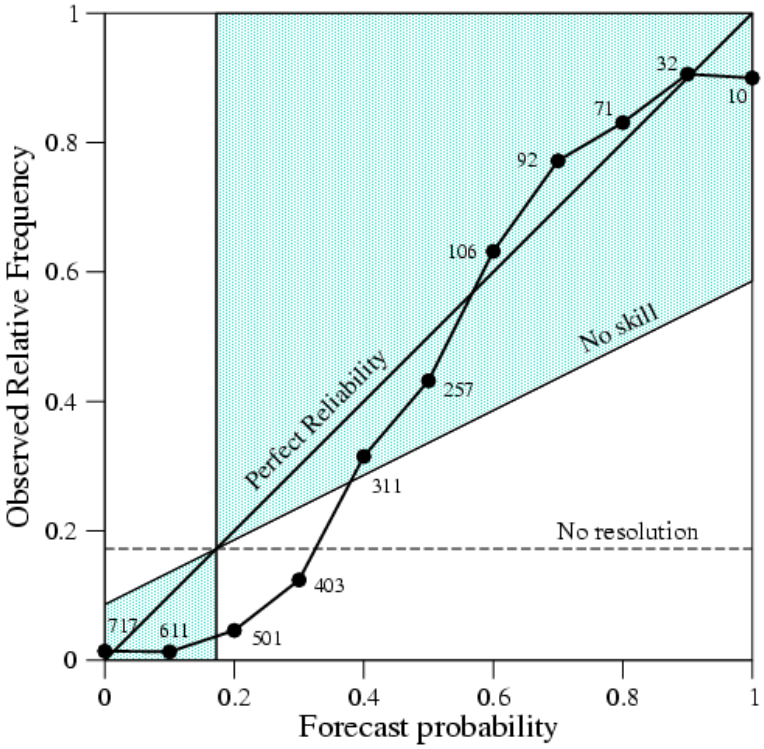

Figure B.3: Hypothetical reliability diagram showing *Observed Relative Frequency* as the function of *Forecast Probability*.

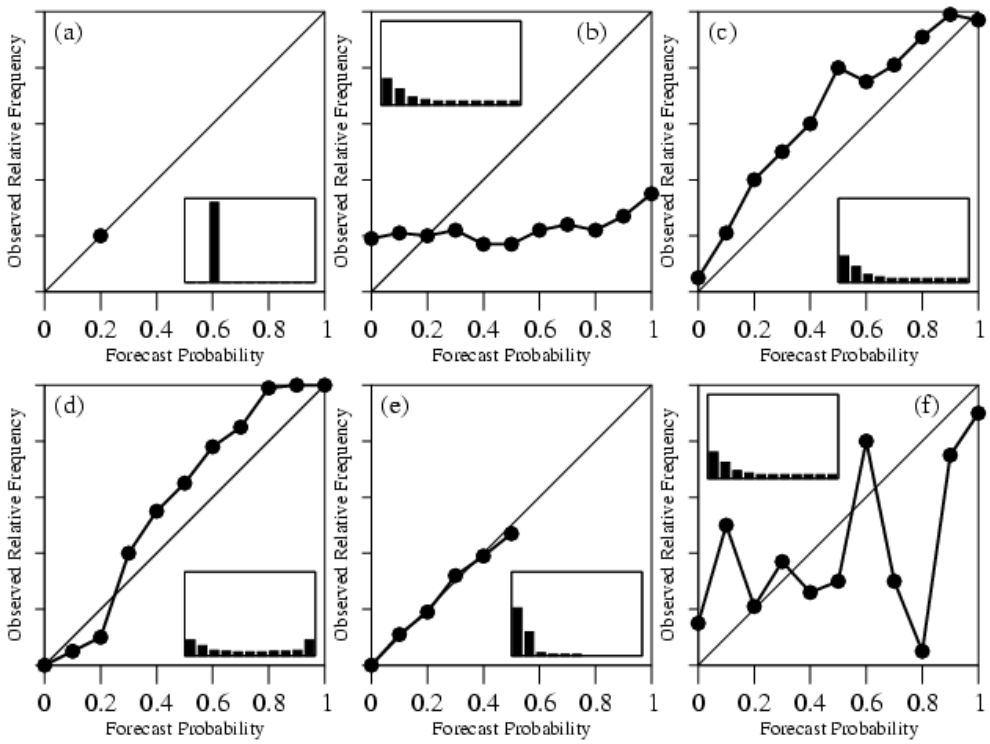

Figure B.4: Hypothetical reliability diagrams showing the *Observed Relative Frequency* as the function of the *Forecast Probability*. Figure (a) shows the climatological forecast (only one point on the perfect reliability line), (b) shows a forecast with minimal resolution, (c) is a forecast with underestimation (the observed relative frequency is higher than the forecast probability), (d) represents a forecast with good resolution but poor reliability. Figure (e) shows the case of a reliable forecast of a rare event, while (f) represents the case when the verification dataset is not large enough. The inset boxes are known as sharpness diagrams where the frequencies of forecast probabilities (number of cases in each category, normalized by the total number of cases) are shown. A system is said to be sharp if it is able to distinguish between high and low probabilities giving 0% and 100% more often then the rest (sharpness diagram with a U shape).

Appendix C

Description of the quasi-operational LAMEPS

The short-range limited area ensemble prediction system of HMS is running every day in quasi-operational status. It is run with cycle 30 of the ALADIN limited area model and it is driven by the members of the global PEARP system. Just like PEARP the LAMEPS also has 11 members. At present no local data assimilation or generation of local perturbations are applied. Forecasts are made once a day starting from the 18 UTC data. The system is made up of Fortran programs and UNIX shell scripts which are responsible for the different sub-tasks and it is running on an SGI Altix 3700 machine. The whole forecast process takes about 3-3.5 hours using 32 processors of the computer. The schematic view of the system is presented in Fig. C.1.

Interpolation of the global PEARP fields to the limited area domain is needed in order to be able to use the outputs of the global model as initial and lateral boundary conditions. In the operational practice this interpolation is done in two steps. Once the global PEARP forecasts are ready, an interpolation is done by the French colleagues at Météo-France. This has two main purposes: (i) to reduce the amount of data to be downloaded by avoiding the transfer of global fields, and (ii) to interpolate from the variable resolution of ARPEGE (Fig. 2.3) to a constant resolution, approx. 20 km. These files need to be further interpolated to the resolution used by the different ALADIN members who download the LBCs

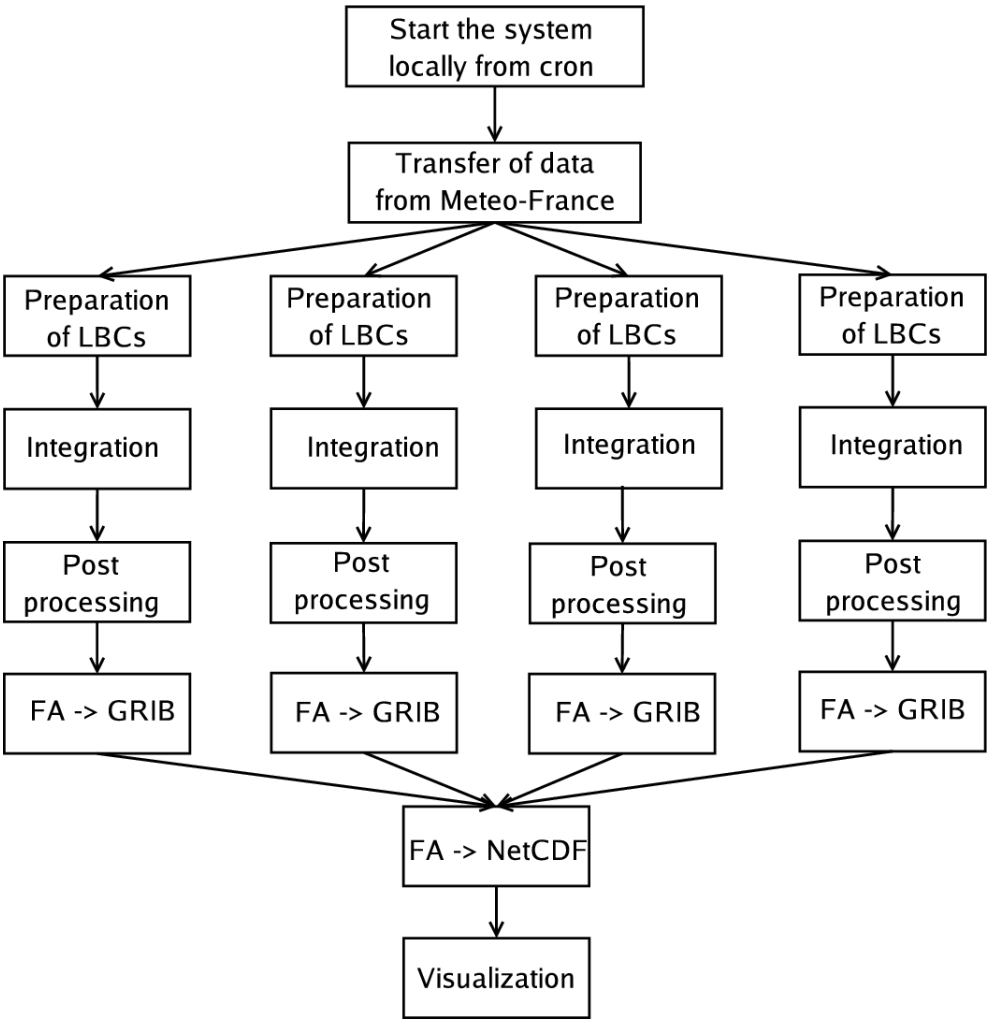

Figure C.1: Schematic view of the LAMEPS system. After transferring the necessary files from Météo-France, the ensemble members are organized into 4 groups. Each group is running on 8 processors of the SGI Altix 3700 machine, independently from the other groups until the preparation of the NetCDF files, which is done in one go for all members.

for their own LAMEPS. A vertical interpolation is also performed from 55 levels (as used in PEARP) to 46. The interpolated files are usually available around 22:40 UTC. Once they are downloaded (via ftp), the forecast process can start locally at HMS. The downloaded files are further interpolated to the exact domain and resolution (approx. 12 km) which is used for the model integration.

When the initial and lateral boundary conditions are in the proper format (resolution, domain, etc.) the integration of the model can start. The ALADIN ensemble system is running on a domain covering large part of Continental Europe (Fig. 2.5) with a horizontal resolution of approx. 12 km. In the vertical 46 levels are used. Forecast length is 60 hours and the time step used for the integration is 450 seconds (7.5 minutes).

The next step is the post-processing of raw model outputs in order to support the application of the model results by forecasters, or by the end-users. Post-processing of our LAMEPS forecasts consists of various steps:

- Transformation from spectral to physical space.
- Change of projection from Lambert to a 0.1 degree latitude-longitude coordinate system.
- Change of the vertical coordinate system: pressure levels instead of the hybrid coordinate system used for the model integration.
- Computation of special diagnostic variables.

As a last step of post-processing, files are converted from the so-called *FA* format of ARPEGE/ALADIN to the most common *GRIB* and *NetCDF* formats to be used for visualization and verification.

After performing post-processing to a latitude-longitude grid, the outputs of the LAMEPS system are mainly visualized using HAWK ([1]). The available products from our LAMEPS system are the ensemble mean, the ensemble spread (computed around the mean), individual ensemble members and probability fields for several parameters (Fig. C.2). The individual members can be visualized in the form of spaghetti diagrams. In addition, plume diagrams are also plotted (using the ECMWF software *Metview*, [2]) for several parameters and selected Hungarian locations (for an example see Fig. 4.5).

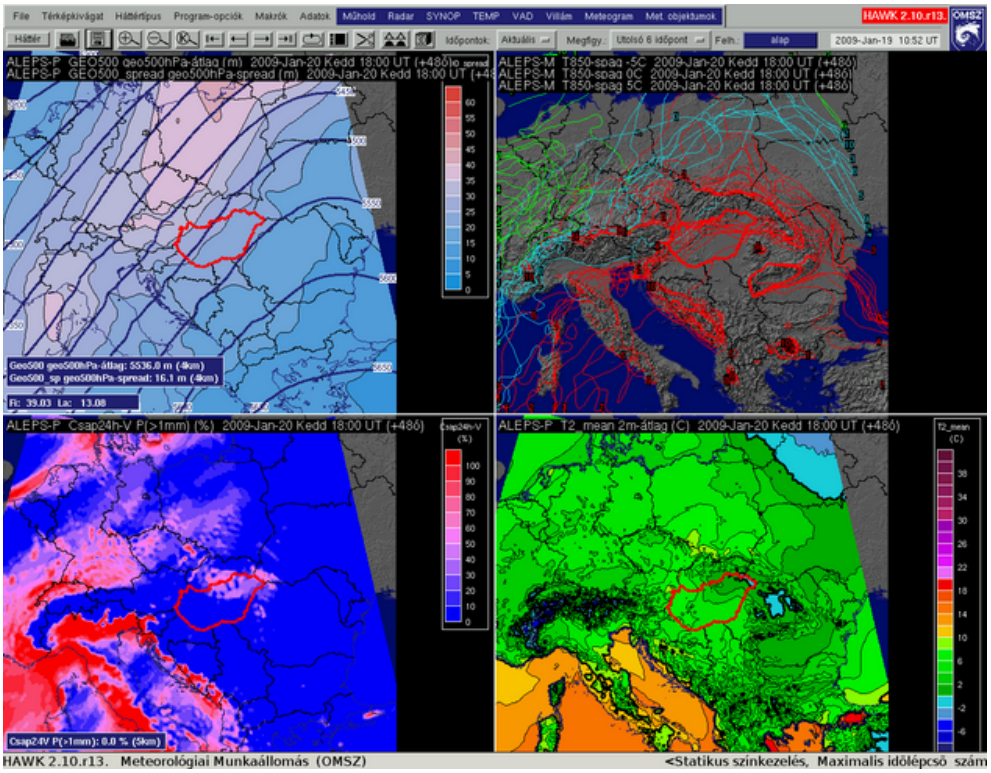

Figure C.2: Example of visualization with HAWK based on the LAMEPS forecast started on 18/01/2009, 18 UTC. Plots are valid at T+48 hours. Top left: Ensemble mean (isolines) and spread (shading). Parameter is 500 hPa geopotential. Top right: Spaghetti diagram. Parameter is 850 hPa temperature, isolines are plotted for -5, 0 and 5 Celsius. Bottom left: Probability map for total precipitation exceeding 1 mm/24 hours. Bottom right: Ensemble mean for 2 meter temperature.

Bibliography

[1] A HAWK-2 meteorológiai munkaállomás. http://www.met.hu/nmo/hawk.

[2] Metview User Guide. Available from the ECMWF website: http://www.ecmwf.int/publications/manuals/metview/manual/index.html.

[3] F. Bouttier. Changes to the PEARP ensemble forecasting system. *ALADIN Newsletter*, 34:10, 2008.

[4] R. Buizza, M.J. Miller, and T.N. Palmer. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Technical Memoranda 279, ECMWF, 1999.

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

[6] H. C. Davies. A lateral boundary formulation for multi-level prediction models. *Q. J. Roy. Meteorol. Soc.*, 102:405–418, 1976.

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

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

[9] J. A. García-Moya, A. Callado, C. Santos, D. Santos-Munoz, and J. Simarro. El sistema de predicción por conjuntos para el corto plazo del INM. *AME Boletín*, 16:22–27, 2007.

[10] G. Götz. Káosz a légkörben. *Természet Világa*, 135(11), 2004.

[11] V. Guidard and P.L. Payet. Etude des vecteurs singuliers dans ALADIN, 2001. Available from the authors on request.

[12] E. Hágel. Latest results of the LAMEPS experiments. *ALADIN Newsletter*, 27:105–113, 2005.

[13] E. Hágel. Report on work in Vienna on the common verification package (available at www.rclace.eu), 2006.

[14] E. Hágel. Singular vector experiments at the Hungarian Meteorological Service. *ALADIN Newsletter*, 33:52–59, 2008.

[15] E. Hágel and A. Horányi. The development of a limited area ensemble prediction system at the Hungarian Meteorological Service: sensitivity experiments using global singular vectors, preliminary results. *Időjárás*, 110(3-4):229–252, 2006.

[16] E. Hágel and A. Horányi. Sensitivity experiments of global singular vectors at the Hungarian Meteorological Service. *ALADIN Newsletter*, 29:50–57, 2006.

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

[18] E. Hágel and M. Mile. The limited area ensemble prediction system of the Hungarian Meteorological Service. *ALADIN Newsletter*, 35, 2009. To appear.

[19] E. Hágel and G. Szépszó. Preliminary results of lameps experiments at the Hungarian Meteorological Service. *ALADIN Newsletter*, 26:44–52, 2004.

[20] D. Heizenreder, S. Trepte, and M. Denhard. SRNWP-PEPS: A regional multi-model ensemble in Europe. *The European forecaster, Newsletter of the WGCEF*, 11:29–35, 2006.

[21] H. Hersbach, R. Mureau, J.D. Opsteegh, and J. Barkmeijer. A Short-Range to Early-Medium-Range Ensemble Prediction System for the European Area. *Mon. Weather Rev.*, 128:3501–3519, 2000.

[22] A. Horányi, I. Ihász, and G. Radnóti. ARPEGE/ALADIN: A numerical weather prediction model for Central Europe with the participation of the Hungarian Meteorological Service. *Időjárás*, 100:277–301, 1996.

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

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

[25] I.T. Jolliffe and D.B. Stephenson, editors. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 2003.

[26] E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 2003.

[27] A. Kann. Work on the Common Verification Package for the evaluation of ensemble forecasts (available at www.rclace.eu), 2007.

[28] S. Kertész. A nem-hidrosztatikus modellek. In *29. Meteorológiai Tudományos Napok. Az időjárás numerikus előrejelzése*, pages 169–180, 2003.

[29] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[30] E. N. Lorenz. *The Essence of CHAOS*. University of Washington Press, 1993.

[31] P. Lynch. Techniques of Initialization. *Weather*, 42:66–70, 1987.

[32] P. Lynch. *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge University Press, 2006.

[33] R. Mladek. Work on the Common Verification Package for the evaluation of ensemble forecasts (available at www.rclace.eu), 2006.

[34] A. Montani, M. Capaldo, D. Cesari, C. Marsigli, U. Modigliani, F. Nerozzi, T. Paccagnella, P. Patruno, and S. Tibaldi. Operational limited-area ensemble forecasts based on the 'Lokal Modell'. *ECMWF Newsletter*, 98:2–7, 2003.

[35] J. Nicolau. Short-range ensemble forecasting. In *Proceedings of the WMO/CBS Technical Conference in Cairns Australia*, 2002.

[36] A. Persson. Synoptic-dynamic diagnosis of medium-range weather forecasts systems. In *Proceedings of the seminar on Diagnosis of models and data assimilation systems*, 1999.

[37] A. Persson and F. Grazzini, 2005. User Guide to ECMWF Forecast Products.

[38] G. Radnóti. A numerikus előrejelzés alapjai, numerikus módszerek. In *29. Meteorológiai Tudományos Napok. Az időjárás numerikus előrejelzése*, pages 25–40, 2003.

[39] L.F. Richardson. *Weather Prediction by Numerical Process*. Cambridge University Press, 1922.

[40] F. Schmidt. Variable fine mesh in a spectral global model. *Beitr. Phys. Atmos.*, 50:211–217, 1977.

[41] R. Stappers and J. Barkmeijer. Hirlam CAPE singular vectors. *HIRLAM Newsletter*, 54:76–80, 2008.

[42] B. Szintai and I. Ihász. The dynamical downscaling of ECMWF EPS products with the ALADIN mesoscale limited area model: Preliminary evaluation. *Időjárás*, 110(3-4):253–277, 2006.

[43] I. Szunyogh. Lokális alacsony dimenziós viselkedés a légkörben. *Természet Világa*, 133(12), 2002.

[44] Z. Toth and E. Kalnay. Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Weather Rev.*, 125:3297–3319, 1997.

[45] Y. Wang, M. Bellus, S. Ivatek-Sahdan, A. Kann, X-L. Ma, M. Steinheimer, W-H. Tian, and C. Wittmann. The Central European limited area ensemble forecasting system: ALADIN-LAEF. *Q. J. Roy. Meteorol. Soc.*, 2009. In preparation.

[46] D.S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Elsevier Science & Technology Books, 2006.

[47] K. Yessad, 2007. Basics about ARPEGE/IFS, ALADIN and AROME in the cycle 33 of ARPEGE/IFS.

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőimnek, Horányi Andrásnak és Matyasovszky Istvánnak a doktori munka során nyújtott segítségükért. Köszönettel tartozom kollégáimnak az Országos Meteorológiai Szolgálatnál, különösképpen Radnóti Gábornak az alkalmazott numerikus modellekkel kapcsolatos segítségéért, Kullmann Lászlónak a kézirat gondos átolvasásáért és hasznos észrevételeiért, valamint Zsótér Ervinnek a dolgozatban szereplő esettanulmány elkészítésében nyújtott segítségéért. Köszönettel tartozom az Országos Meteorológiai Szolgálat Elnökének is, hogy doktori tanulmányaimat támogatta.

Köszönettel tartozom továbbá Jean Nicolau-nak a PEARP rendszerrel kapcsolatos segítségéért, valamint a Météo-France munkatársainak, akik lehetővé tették a PEARP rendszer tagjainak kezdeti-, és peremfeltételként való operatív alkalmazását.

Doktori tanulmányaim során több intézményt is lehetőségem volt felkeresni hosszabb-rövidebb tanulmányutak keretében. Ezúton is szeretnék köszönetet mondani mindazoknak, akik ezen tanulmányutak során munkámat segítették.

Végül, de nem utolsó sorban szeretnék köszönetet mondani szüleimnek, akik tanulmányaimat mindenkor segítették és támogatták.

Acknowledgement

I would like to thank my supervisors András Horányi and István Matyasovszky for their help throughout my PhD work. The support and help of my colleagues at the Hungarian Meteorological Service is very much appreciated. Special thanks goes to Gábor Radnóti for his help with the applied NWP models, to László Kullmann for his helpful comments while carefully reading the manuscript, and to Ervin Zsótér for his help with the case study presented in the thesis. Thanks are also due to the Director of the Hungarian Meteorological Service for supporting my PhD studies.

I am grateful to Jean Nicolau for his help with the PEARP system and for the team of Météo-France who made possible the operational use of PEARP as initial and lateral boundary conditions for our LAMEPS.

During my PhD studies I could visit several institutes in the framework of longer or shorter stays. I would like to thank everyone who has helped my work during these occasions.

Last but not least, I would like to thank my parents who have always helped and supported me throughout my studies.