IDŐJÁRÁS

QUARTERLY JOURNAL
OF THE HUNGARIAN METEOROLOGICAL SERVICE

Special Issue: Atmospheric Dynamics

CONTENTS

****

http://www.met.hu/Journal-Idojaras.php

VOL. 110 * NO. 3-4 * JULY - DECEMBER 2006

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service

Editor-in-Chief
LÁSZLÓ BOZÓ

Executive Editor
MARGIT ANTAL

Editorial Office: P.O. Box 39, H-1675 Budapest, Hungary or
Gilice tér 39, H-1181 Budapest, Hungary
E-mail: bozo.l@met.hu or *antal.e@met.hu*
Fax: (36-1) 346-4809

Subscription by

mail: IDŐJÁRÁS, P.O. Box 39, H-1675 Budapest, Hungary;
E-mail: bozo.l@met.hu or *antal.e@met.hu*; *Fax: (36-1) 346-4809*

Editorial

The main objective of atmospheric dynamics is to study of those processes of the atmosphere that are associated with weather and climate, in order to understand and simulate with numerical models the different motion systems ranging from the micro-scale to the global circulation. In atmospheric dynamics the fluid is regarded as a continuous medium, and the fundamental laws of fluid mechanics and thermodynamics are expressed in terms of partial differential equations involving the fluid velocity, pressure, density, and temperature. The integration of the governing hydro-thermodynamic equations by numerical methods offers an opportunity to investigate basic theoretical problems (such as the interactions and energy transport processes among the motion systems of different spatial and temporal scales), it paves the way for objective and reliable weather and air-pollution dispersion forecasts, and it constitutes the single possibility to describe the forced and free changes of the climate system.

Successful modeling work requires a close co-operation between the meteorologist and physicist, who are experienced in fluid dynamics and nonlinear processes, mathematics and computer sciences, data assimilation, model initialization, and numerical methods. Development of different scale dispersion models demands the collaboration with air chemistry researchers. In addition, parameterization techniques of the soil–vegetation–atmosphere exchange processes call for solving problems of soil science, biology, and ecology. For a productive scientific co-operation, it is inevitable to clearly define the objective of the common research, assure continuous communication among the different teams, and demand regular publications of high quality.

The principal mission of the Working Group on Atmospheric Dynamics, belonging to the Scientific Committee for Meteorology of the Hungarian Academy of Sciences, is to create the conditions mentioned above. The tasks of the members of the Working Group include provision of a scientific forum for the different research teams and for the Hungarian investigators working in foreign institutes in order to exchange results related to geophysical fluid dynamics, assisting scientific co-operations and team work. Programs of the Working Group contribute to these goals by organizing joint lectures, conferences, and presentations of the following scientific groups:

- *Numerical Weather Prediction and Climate Dynamics Division of the Hungarian Meteorological Service (HMS),*
- *Department of Meteorology, Department of Applied Analysis and Computational Mathematics, and von Kármán Laboratory of Environmental Flows of the Eötvös Loránd University (ELTE, Budapest), and*
- *Department of Fluid Mechanics, Budapest University of Technology and Economics.*

One of the most important events of the Hungarian meteorological community is the Scientific Days of Meteorology, organized every year at the headquarters of the Hungarian Academy of Sciences. Recent topics of these conferences, arranged by our Working Group, included numerical modeling (in 2003), climate dynamics research (in 2004), and cloud physics and micrometeorology (in 2006).

The present thematic issue of IDŐJÁRÁS, *with 14 scientific papers, is also the results of our activity. The leading paper is dedicated to recall the first steps of numerical weather prediction in Hungary. The following papers describe the present status of numerical* *models, as ALADIN or MM5, for the Carpathian Basin. Meso-scale processes, ensemble forecasts and dynamic downscaling methods represent the main fields of this research work at the HMS.*

Fluid mechanics laboratory experiments offer a powerful tool for the analysis of atmospheric dynamics. This type of experimental work can successfully illustrate the structure of mountain waves.

Investigations of model initialization and data assimilation methods are examples of the most fruitful research activities at the Numerical Weather Prediction and Climate Dynamics Division of the HMS. Development of a variational data assimilation system for a limited area model and the application of high-resolution satellite observations in the ALADIN/HU model are presented.

Elaborating a new meso-scale transport model for the Carpathian Basin at ELTE and HMS (methodology and applications) illustrates the research activity in the field of air pollution.

Up-to-date mathematical background in the theory of partial differential equation systems and numerical methods is an indispensable knowledge. Four papers are dedicated to these questions from the Mathematical Institute of ELTE. Theoretical and applied results of splitting methods, main attributes and applications of semi-groups, and possible application of discontinuous Galerkin methods are demonstrated.

The 30 authors of the 14 papers represent different generations, from the pioneers of numerical modeling activity in Hungary, "the elderly generation", through present-day researchers successfully continuing the numerical modeling work, up to the new generation consisting of PhD students and young scientists.

The editors and invited authors dedicate this thematic issue of IDŐJÁRÁS *to the illustration of the status and main results of atmospheric dynamics and numerical weather prediction research in Hungary at the beginning of the 21st century for the enrichment of the knowledge of the readers.*

Gusztáv Götz and Tamás Weidinger
Guest editors

Working Group on Atmospheric Dynamics
Scientific Committee for Meteorology
Hungarian Academy of Sciences

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 193–201

A historical review of the first steps in numerical weather prediction in Hungary

Pál Ambrózy[1], **Gusztáv Götz**[2] and **Tibor Tänczer**[3]

[1] *Lágymányosi utca 22, H-1111 Budapest, Hungary*
[2] *Lupény utca 6-8, H-1026 Budapest, Hungary*
[3] *Gyöngyvér utca 11, H-1029 Budapest, Hungary*

(Manuscript received in final form June 2, 2006)

Abstract—This overview presents a short account on how the theory and practice of the numerical weather prediction developed at the early stage of this discipline. Following a concise review of the work carried out abroad, investigations of the Hungarian meteorologists are described.

Key-words: numerical weather prediction, barotropic models, baroclinic models, level of non-divergence

1. Introduction: Early history of the problem

As aptly stated in one of Charney's famous papers, meteorologists have long known that the atmosphere exhibits no periodicities of the kind that enable one to predict the weather in the same way one forecasts the tides. No simple set of causal relationships can be found which relate the state of the atmosphere at one instant of time to its state at another. It was this realization that led *V. Bjerknes* (1904) to define the problem of weather prediction as nothing less than the integration of the governing equations of the atmospheric processes. In his remarkable manifesto and testament of deterministic faith, Bjerknes recognized that the future state of the atmosphere is, *in principle*, completely determined by its detailed initial state and known boundary conditions, together with the Newton's equations of motion, the Boyle-Charles-Dalton equation of state, the equation of mass continuity, and the thermodynamic energy equation. But it remained for *Richardson* (1922) to suggest the practical means for the solution of this problem. He proposed to integrate the governing equations numerically, and showed exactly how this might be done. *Charney* (1951) strongly emphasized: the fact that the actual forecast Richardson used to test his method was unsuccessful in no sense a measure of the value of his work. In retrospect, it becomes obvious that the inadequacies of observation alone would have doomed any attempt however well conceived, a circumstance of which Richardson was aware. The real value of his work lay in the fact that it crystallized once and for all the essential problems that would have to be faced by future workers in the field, and that it laid down a thorough groundwork for their solution.

For a long time no one ventured to follow in Richardson's footsteps. The paucity of the observational network and the enormity of the computational task stood as apparently insurmountable barriers to the realization of his dream that one day it might be possible to advance the computation faster than the weather. But with the increase in the density and extent of the surface and upper-air observational network on the one hand, and the development of large-capacity high-speed computing machines on the other, interest revived in Richardson's problem in the years following World War II, and attempts were made to attack it anew. Early in 1946, von Neumann singled out the problem of numerical weather prediction for special attention (*Thompson*, 1983). Although von Neumann had a deep appreciation of its practical importance and intrinsic scientific interest, he also regarded it as the most complex, interactive, and highly nonlinear problem that had ever been conceived – one that would challenge the capabilities of the fastest electronic computing devices for many years. On August 29 and 30, 1946, at the Institute for Advanced Study in Princeton, New Jersey, a notable conference took place. It was titled simply "Conference on Meteorology", but it may be considered the first conference on numerical weather prediction. It was organized by von Neumann, probably with assistance from Rossby, and from Wexler of the U.S. Weather Bureau. The purpose of the conference was to enlist the support of the meteorological community for a bold project. This undertaking had already been proposed in May 1946 to the U.S. Navy by the Institute for Advanced Study – that is to say, by von Neumann himself. The proposal was perhaps the most visionary prospectus for numerical weather prediction since the publication of Richardson's book a quarter-century earlier. In the words of that proposal, „the objective of this project is an investigation of the theory of dynamic meteorology in order to make it accessible to high-speed, electronic, digital, automatic computing".

The Navy wisely funded this proposal, starting July 1, 1946. Within the Electronic Computer Project of the Institute for Advanced Study, a Meteorological Research Group was created. The team adopted the general plan of attacking the problem of numerical weather prediction by a step by step investigation of a series of models approximating more and more the real state of the atmosphere. In accordance with this plan, a two-dimensional, nonlinear barotropic model was chosen as the first object of study. The first successful numerical forecasts were made on the Electronic Numerical Integrator and Computer (ENIAC) at the Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, in the spring of 1950. The results of a series of four 24-hour predictions computed from actual data at the 500 hPa level were described, together with an interpretation and analysis, by *Charney et al.* (1950). The causes of the forecast errors were ascribed partly to the use of too large space increment and partly to the effects of baroclinicity.

2. The rapid proliferation of research in deterministic prediction

Needless to say, the *Tellus* paper of 1950 excited considerable interest (*Thompson*, 1983). At the same time, everyone was aware that those calculations were based on the principle of absolute vorticity conservation for two-dimensional flow, which precluded the intensification of circulation centers and did not provide for the formation of new centers where none existed before. Accordingly, there was a general rush to develop baroclinic models – i.e., models whose vertical structure was simply enough that the equations could be solved without undue computational strain, but general enough that they could simulate cyclogenesis and conversion of available potential energy to the kinetic energy of growing disturbances. In a relative brief span – between 1951 and 1953 – no less than six simple baroclinic models were proposed. In 1952, there were four sizeable research groups, who were concentrating on the problem: the Meteorological Project at the Institute for Advanced Study, the Atmospheric Analysis Laboratory of the U.S. Air Force Cambridge Research Laboratories, the Napier Shaw Laboratory of the British Meteorological Office, and the International Meteorological Institute of the University of Stockholm (working in cooperation with the University of Oslo).

Up to the middle of the 1950s, virtually all of the people involved in the development of numerical methods took a strictly deterministic view of the prediction problem – i.e., that the future state of the atmosphere is completely determined by its present state. Uncertainty of the initial state as a factor in the predictability of large-scale atmospheric flow patterns was investigated in great detail first by *Thompson* (1957). Six years later, *Lorenz* (1963) clearly demonstrated that in some nonlinear dynamical systems it is quite normal for two almost identical states to be followed, after a sufficient time lapse, by two states bearing no more resemblance than two states chosen at random from a long sequence. Systems in which this is the case are said to be sensitively dependent on initial conditions. Sensitive dependence can serve as an acceptable definition of chaos, and currently it is believed that the atmosphere should be looked upon as a chaotic geophysical fluid. This means that the unavoidable inaccuracy of the initial data will tend to be amplified with time even if the prediction model is somehow made perfect. Therefore, weather predictions should be stated in terms of probability distributions, or – in other words – no forecast can be considered complete without a forecast of the forecast skill. Recognizing the chaotic nature of the atmospheric dynamics has led to the development of stochastic-dynamic prediction methods: instead of issuing categorical statements, the forecast skill is estimated using an ensemble of simultaneous deterministic forecasts, each of which obtained from slightly different (i.e., equally possible) initial conditions.

3. Early investigations carried out in Hungary

The rapid and spectacular development in the techniques of numerical weather prediction awaked the interest of the Hungarian meteorologists as well. In this fact, two circumstances played an important role. First, between the years of 1954 and 1958, a generation of relatively great number of meteorological students finished their studies at the Eötvös Loránd University, who had obtained the necessary mathematical and physical training, and thus they were susceptible to attain the up-to-date knowledge of the atmospheric physics and dynamics. On the other hand, at that time the director of the Hungarian Meteorological Institute, Professor F. Dési preferred dynamic meteorology, and he fully intended that the exact physical methods should have a greater weight in the research work of the institute. For the sake of this, he carried out the necessary organizational steps. In 1957, he established, among others, a research team whose prime task was to acquire the basic knowledge in the field of numerical weather prediction. At that time, relatively rich foreign special literature was already available in this field. The members of the team first published summary papers on the principles and practical methods of numerical weather prediction for the Hungarian meteorological community (*Götz* and *Tänczer*, 1958; *Götz*, 1958, 1959a).

By the end of the 1950s, partly due to the rapid development of the computational techniques, partly to the gradual improvement of atmospheric models, a number of national meteorological services decided to introduce the numerical prediction method to the operational procedure of weather forecasting. At that time, the models were able to describe only about 65 percentages of the daily variation of large-scale circulation. Consequently, the success of the numerical weather prediction products could not match up to the ones turned out by the experienced synoptic forecasters. At the very least, synoptic forecasters now had, in addition to everything that they had formerly used, the information that "this is what the computer says will happen". They could use or reject this information as they saw fit. As the years advanced, forecasters came to rely more and more on the numerical products.

This fact was not in a different way in Hungary either, except that the Hungarian Meteorological Institute, similarly to the greater part of the other national meteorological services, had at this time no electronic computer facilities available. In order to overcome this difficulty, graphical (manual) techniques were developed for constructing objective prognostic charts. It was possible for us as well to adapt these graphical methods. One of the members of our team managed to attain an extraordinarily simple and rapid graphical procedure for the 24-hour forecasting of the 700 hPa level during his study-tour in Moscow, where Buleyev developed the method. This technique is described in great detail in the work of *Kibel* (1957). Accounts on the domestic application of the Buleyev method can be found in a series of papers (*Ambrózy*, 1959, 1964; *Ambrózy et al.*, 1959; *Götz*, 1959b). Here we only shortly refer to the essence of this procedure.

Buleyev deduced a prognostic equation, the barotropic vorticity equation, for the geopotential height change at the non-divergent level of the atmosphere. In his method, after a number of simplifications, the dynamical height of a point at this level varies in time as if the contour lines were displaced by a suitably chosen transferring field. This field can be constructed by smoothing the contour lines of the 700 hPa level. The initial field (more exactly, grid point values of a net with a 500 km space increment) are displaced for 24-hour in advance by the geostrophical wind speed computed from the smoothed field. The predicted field can be constructed on the base of the end-point values of the trajectories. The application of the method requires maximum a 2-hour work, but of course it cannot predict the development of the pressure systems (the process of cyclo- and anticyclogenesis) at all.

At the same time (in 1958), the graphical forecasting method elaborated earlier by *Fjörtoft* (1952) was also adapted. This technique is based on the assumption that the absolute vorticity is conserved in time. Its routine application consumes more time, but the accuracy of the forecasts exceeds that of the Buleyev method (*Götz*, 1959c; *Ambrózy et al.*, 1960). Here, the advecting field is produced by a displacement of the initial field in the directions of east–west and north–south with twice grid distance, and then summing the obtained two fields. In this method, a correction function $J(\varphi)$, coming from the map distortion and varying with the geographical latitude φ is taken into account. The difference of the initial and constructed fields, which is in fact the field of geostrophic vorticity, is displaced with the use of a "gradient ruler". The advected vorticity field is then transformed into geopotential contour lines by a graphical integration of the Helmholtz's equation. Because at that time at the Hungarian Meteorological Institute the Lambert–Gauss' conformal conical projection was used, we had to determine the function $J(\varphi)$ by numerical integration for this special projection (*Tänczer* and *Tóth*, 1959).

In 1958, daily forecasts were produced by both of the methods described above, and were handed over to synoptic forecaster at the beginning of the forecast discussion held every day at noon. Thus, we succeeded in spanning the gap, existing at those times between the synoptic and dynamic meteorologists, with a narrow bridge. After investigating the accuracy of our forecasts, we concluded that the success of the objective predictions depended on the synoptic situation to a great extent: the largest errors occurred in territories of the strongest thermal advection (baroclinic zones) (*Tänczer*, 1959).

In addition to our routine forecasting work, numerical experiments were carried out to see, how the change of relative vorticity in the case of meridional displacement might be taken into account in the Buleyev method. For this sake, we studied the role of the northward variation of the Coriolis parameter arising from the sphericity of the earth (Rossby term) in the validity of the forecasts. It was found that this effect is proportional to the magnitude of the mass flux along the meridians. We incorporated the correction coming from this effect into the Buleyev's prediction technique in the way that not the initial field itself, but its modified version was displaced. Hereby, improvement of few percentages could be achieved in the forecasts (*Ambrózy*, 1961).

In the barotropic prediction methods it is of principal importance, which level of the atmosphere should be considered as the non-divergent one. Buleyev assumed this surface to be at the 700 hPa level, while Fjörtoft placed it on the 500 hPa level. The level of non-divergence is theoretically located, where the averaged vertical wind profile with respect to pressure agrees with the actual wind speed. Since the location of this level varies from place to place and from time to time, the non-divergent surface itself exhibits variations both in space and time. The height of the non-divergent surface was investigated by using the actual wind data (*Götz* and *Tänczer*, 1960; *Tänczer*, 1963, 1964). We concluded that among the main geopotential levels, the 700 hPa contour lines meet the assumption of non-divergence best of all, and the change of the divergence in time is lowest there as well. The mean value of non-divergent level was at 650 hPa in day-time and around 600 hPa at night.

The success of the objective forecasting was also examined in such way that to what extent the validity of the predicted fields exceeded that of the initial fields considered as forecasts (*Tänczer*, 1960). The difference between the two values as a relative validity index may be considered a measure of the success of the forecasts, showing also the difficulty of the prediction work in each case. The accuracy of the forecasts carried out with the Buleyev method was higher than the persistence in 74 percentages of all the cases.

It was an indication of the appreciation of the work carried out by our research group in the field of numerical weather prediction, that we got invitations to participate in (now historically famous) international symposia held, among others places, in Oslo and Moscow (*Ambrózy*, 1963; *Tänczer*, 1963). These symposia offered opportunity to give an account of our own achievements to the forecasters community (*Ambrózy*, 1964; *Tänczer*, 1964).

In the meantime, the first electronic digital computers appeared in Hungary. Though they had rather low capacity, some works could be experimentally accomplished on the Russian-made computer called Ural–1, which was installed at the Central Statistical Office. Our fortunate, but very limited access to this computer made it possible to perform some special investigations, including the smoothing of contour lines, and determination of grid point values of the geostrophic vorticity (*Ambrózy* and *Götz*, 1961; *Götz*, 1961; *Ambrózy*, 1962). All of these exercises were capable only for acquainting the knowledge of the new tasks, but more complicated computations were excluded for us. It became more and more evident that the conditions for running own-developed forecasting methods would not be suitable within a reasonable time.

4. Epilogue

In the meanwhile, we continued to get acquaintance with the more and more sophisticated prediction models, the application of the primitive equations, the objective analysis techniques, the methods of how to take into account the effects of mountains and frontal zones, the procedures of numerical precipitation forecasting, the modeling of the meso-scale processes, and the possibilities of utilizing the new information yielded by the meteorological satellites in numerical weather analysis and prediction. However, for lack of an available electronic computer, we had only the opportunity to systematize and publish the acquainted knowledge to specialists and university students. The activity of the small research team established ten years earlier, in 1957, culminated and finished with the compilation of a comprehensive volume entitled *Principles of Dynamical Weather Forecasting* (*Ambrózy*, 1967). This work served for a long time as the basic text-book in the education of meteorological students at the Eötvös Loránd University.

The investigation in the field of numerical weather prediction could revive only a decade later, in the middle of the 1970s, with the appearance and availability of large-capacity, high-speed electronic machines in Hungary.

References

Ambrózy, P., 1959: Results of the experimental forecasting performed by Buleyev's method (in Hungarian). *Beszámolók az 1958-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 35-39.

Ambrózy, P., 1961: Evaluation of the effects of meridionality in the application of Buleyev's method of barotropic pressure prediction. (in Hungarian). *Beszámolók az 1960-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 38-42.

Ambrózy, P., 1962: Smoothing of geopotential fields (in Hungarian). *Beszámolók az 1961-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 33-36.

Ambrózy, P., 1963: Conference on the problems of numerical prediction in Moscow (in Hungarian). *Időjárás 67*, 255-256.

Ambrózy, P., 1964: Some practical problems of forecasting the pressure field by graphical methods (in Russian). *Trudy of symposium on numerical methods of weather forecasting*. Gidrometeoizdat, Leningrad, 92-96.

Ambrózy, P. and *Götz, G.*, 1961: On the application of the electronic computer Ural-1 in meteorology (in Hungarian). *Időjárás 65*, 152-154.

Ambrózy, P., Götz, G., and *Tänczer, T.*, 1959: Numerical forecasting of contour fields by the aid of Buleyev's barotropic method (in Russian). *Időjárás 63*, 74-81.

Ambrózy, P., Götz, G., and *Tänczer, T.*, 1960: Numerical forecasting of 500 mb contour charts by means of Fjörtoft's graphical method (in Hungarian). *Beszámolók az 1959-ben végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 10-20.

Ambrózy, P. (ed.), 1967: *Principles of Dynamical Weather Forecasting* (in Hungarian). Országos Meteorológiai Intézet, Budapest, 228 pp.

Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteor. Zeitschrift 21*, 1-7.

Charney, J.G., Fjörtoft, R., and *von Neumann, J.*, 1950: Numerical integration of the barotropic vorticity equation. *Tellus 2*, 237-254.

Charney, J.G., 1951: Dynamic forecasting by numerical process. In *Compendium of Meteorology* (ed.: *T.F. Malone*). American Meteorological Society, Boston, 470-482.

Fjörtoft, R., 1952: On a numerical method of integrating the barotropic vorticity equation. *Tellus 3*, 179-194.

Götz, G., 1958: On the simplifying and boundary conditions of barotropic pressure forecasting methods (in Hungarian). *Időjárás 62*, 338-343.

Götz, G., 1959a: On the principal and hydromechanical bases of numerical forecasting of the synoptic situation (in Hungarian). *Beszámolók az 1958-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 12-19.

Götz, G., 1959b: Buleyev's advective model for forecasting of the geopotential height change at the equivalent barotropic level (in Hungarian). *Beszámolók az 1958-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 29-34.

Götz, G., 1959c: Experimental forecasting by Fjörtoft's barotropic method (in Hungarian). *Időjárás 63*, 367-368.

Götz, G., 1961: Computation of the vorticity field by the electronic computer Ural–1 (in Hungarian). *Beszámolók az 1960-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 43-46.

Götz, G. and *Tänczer, T.*, 1958: Hydrodynamical bases of numerical forecasting the flow pattern. *Időjárás 62*, 90-98.

Götz, G. and *Tänczer, T.*, 1960: Location of the level of non-divergence in the atmosphere (in Hungarian). *Időjárás 64*, 225-229.

Kibel, I.A., 1957: *Introduction to Hydrometeorological Methods of Short-range Weather Forecasting* (in Russian). Gostehizdat, Moscow.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci. 20*, 130-141.

Richardson, L.F., 1922: *Weather Prediction by Numerical Process*. Cambridge University Press, London.

Tänczer, T., 1959: A study of the relation between the validity of forecasts by Buleyev's method and the synoptic situations (in Hungarian). *Beszámolók az 1958-ban végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 40-47.

Tänczer, T., 1960: On the evaluation of numerical forecasts (in Hungarian). *Időjárás 64*, 49-52.

Tänczer, T., 1963: A study of the atmospheric divergence (in Hungarian). *Beszámolók az 1962-ben végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 76-87.

Tänczer, T., 1963: International conference on numerical forecasting in Oslo (in Hungarian). *Időjárás 67*, 125.

Tänczer, T., 1964: Investigation of the divergence in the atmosphere (in Russian). *Trudy of symposium on numerical methods of weather forecasting*. Gidrometeoizdat, Leningrad, 36-40.

Tänczer, T. and *Tóth, P.*, 1959: Computation of the function $J(\varphi)$ by approaching of Lambert–Gauss' conformal conical projection (in Hungarian). *Időjárás 63*, 243-246.

Thompson, P.D., 1957: Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow patterns. *Tellus 9*, 275-295.

Thompson, P.D., 1983: A history of numerical weather prediction in the United States. *B. Am. Meteorol. Soc. 64*, 755-769.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 203–227

The ARPEGE/ALADIN mesoscale numerical modeling system and its application at the Hungarian Meteorological Service

András Horányi[*], **Sándor Kertész, László Kullmann**
and **Gábor Radnóti**

Hungarian Meteorological Service
P.O. Box 38, H-1525 Budapest, Hungary
E-mails: horanyi.a@met.hu kertesz.s@met.hu, kullmann.l@met.hu,
gabor.radnoti@ecmwf.int

(Manuscript received in final form June 27, 2006)

Abstract—The development of the ARPEGE/ALADIN modeling system was initiated in 1990 (by Météo-France). Recently the project encounters 15 partners from Europe and Northern Africa. The main original objective of the cooperation was to develop a numerical weather prediction model for dynamical adaptation, which takes into account all the advantages and constraints coming from its "mother" system ARPEGE/IFS. In a later stage it was natural – based on the inspiration from the ARPEGE/IFS modeling family – to consider the development of all numerical weather prediction related configurations in a single computer code beside the initially established pre-processing (interpolation) and model integration modules. Sophisticated post-processing algorithms were added and then data assimilation procedures were developed (first optimal interpolation and then three-dimensional variational data assimilation). The code has been extended to the tangent linear and adjoint versions, which make possible to apply configurations for sensitivity studies and the computation of singular vectors. The non-hydrostatic version of the model is an essential part of the software: this is the heart of the new AROME model, which is under intensive development for the meso-gamma spatial scales. The article briefly summarizes the most important configurations of the ALADIN model together with some illustration of their practical use at the Hungarian Meteorological Service.

Key-words: numerical weather prediction, mesoscale limited area model, IFS/ARPEGE/ALADIN/AROME modeling system

---

[*] Corresponding author

1. Introduction

The ARPEGE/ALADIN (ARPEGE stands for Action de Recherche Petite Echelle Grande Echelle, i.e., Research Action on Small Scale and Large Scale; ALADIN stands for Aire Limitée Adaptation Dynamique Développement International, i.e., Limited Area Dynamical Adaptation in International Cooperation) modeling family is developed in an international cooperation originally initiated by Météo-France in 1990. At that time, Météo-France together with the European Centre for Medium Range Weather Forecasts (ECMWF) were in the initial phase of developing the ARPEGE/IFS (IFS: Integrated Forecasting System) spectral global model with main emphasis on the data assimilation ingredients (three-dimensional and four-dimensional variational data assimilation) of the system. The main original objective of the ALADIN project was on the one hand to develop a numerical weather prediction (NWP) tool which is capable to dynamically adapt (to high resolution) the global (ARPEGE) model's results, and on the other hand to create a limited area modeling family as a counterpart of the global one taking into account all the advantages and constraints of the ARPEGE/IFS code system. In the course of the development work, soon it was realized that it is worthwhile to extend the created modeling framework into a model family: new and new model configurations were adapted and developed resulting in a very complex, but at the same time very powerful variety of NWP applications. At the moment, the ALADIN cooperation has 15 national (hydro)meteorological services as Member States from different parts of Europe (Austria, Belgium, Bulgaria, Croatia, Czech Republic, Hungary, France, Poland, Portugal, Romania, Slovakia, Slovenia) and Northern Africa (Algeria, Morocco, Tunisia).

The ALADIN model is a spectral mesoscale limited area numerical weather prediction model (*Horányi et al.,* 1996). The horizontal meteorological fields are represented by two-dimensional Fourier decompositions, and the spatial differential operators in the hydro-thermodynamical equations are computed by the analytical derivatives of these truncated Fourier functions. Vertically, hybrid coordinates are used, which are terrain following at the lower model levels and pressure-type for the upper atmospheric layers. The vertical coordinate is determined by the following relationship (*Simmons* and *Burridge,* 1981):

$$p(x,y,\eta,t) = A(\eta) + B(\eta)\, p_s(x,y,t), \tag{1}$$

where $p_s(x,y,t)$ is the surface pressure and for the A and B coefficients the following boundary conditions are valid $A(1) = 0$, $B(1) = 1$, $B(0) = 0$, and furthermore, $\partial\eta / \partial p > 0$.

The ALADIN code is developed in line with the ARPEGE/IFS system. The computer code of the ALADIN system is approximately 90-percent identical to that of the ARPEGE one. The most important difference between the two codes lies in the spectral transformation package (transformation between spectral, Fourier, and gridpoint spaces): ALADIN is using two-dimensional Fourier (full harmonic) functions and ARPEGE is using spherical harmonics. Due to the limited area character of the ALADIN model, it uses lateral boundary conditions taking into account the impact of meteorological processes outside the domain of interest (*Radnóti,* 1995). The non-hydrostatic version of the model (*Bubnová et al.,* 1995) is a unique feature of the limited area code. Nevertheless, the treatment of the initial and output files, the hydrostatic part of the dynamics, and most parts of the physical parameterization schemes are basically the same in the two versions.

Hereafter the main components of the ALADIN model family are summarized and introduced. All these elements are identified with a configuration number (used in the code), and these three digit numbers are used in the ALADIN vocabulary, while referring to the given configuration. First, the pre-processing aspects of the model family will be introduced (preparation of initial surface - climate - data sets, interpolation of prognostic variables to the limited area domain, preparation of ECMWF/IFS information directly applicable to the ALADIN model), then the main aspects of the model integration will be briefly recalled. The post-processing procedures will be described afterwards and then the data assimilation algorithms of the model. Finally some research configurations will be mentioned as the sensitivity experiments or computation of singular vector decomposition and a short outlook (with the description of the AROME model) and summary will complete the paper.

2. Computation of surface characteristics (configuration 923)

It is a natural requirement for a limited area model to be capable to integrate it on any area of interest over the globe, and certainly the ALADIN model meets this requirement. While defining a new model domain, one of the first exercises is to create a data set which characterizes the surface (climatic) conditions of the given area. A special configuration is devoted to that task (configuration 923). This procedure computes surface characteristics for the limited area (*Table 1*) from global and local data sets: constant (like orography or land-sea mask, etc.) and monthly varying fields (vegetation, albedo, etc.). Therefore, the global data are interpolated to the domain of interest with the pre-defined resolution and mapping characteristics (in ALADIN the user can choose different conformal projections as Lambert, Mercator, or spherical projections; note that basically all the partners are using Lambert projections for the operational applications). It is remarked that the above mentioned surface characteristics might be needed not only for the model integration, but for the pre(post)-processing as well, when the pre(post)-processing domain features are different from those of the integration ones (typical example is when one wants to visualize the output meteorological fields on spherical latitude-longitude grid and not on projected geometry). The application of the configuration 923 is needed once for each domain (provided that there are no significant improvements regarding the input data sets or the interpolation algorithm). The recently used model domain (and orography) at the Hungarian Meteorological Service is introduced in *Fig. 1* for illustrating the output of this configuration.

Table 1. The input and output data for configuration 923 (CLIMAP: Climate Long-Range Investigation, Mapping and Prediction; AMIP: Atmospheric Model Intercomparison Project; ISLSCP: International Satellite Land Surface Climatology Project; ESA: Earth Science Applications Directorate; AVHRR: Advanced Very High Resolution Radiometer)

Data sets	**Resolution (points)**	**Applied data**
GLOBE25 (global data set)	2'30'' → 8640 × 4320	- orography and its standard deviation - land-sea mask - main axis of the topography - anisotropy coefficient
Meteosat, NOAA-4, CLIMAP, ISLSCP (global data set)	1° → 360 × 180	- albedo, emissivity - soil hydrological depth - portion of sand and clay soil - portion of vegetation
US Navy data, CLIMAP, AMIP (global data set)	0.83° → 432 × 216	- climatological surface and soil temperature and humidity - emissivity and albedo over oceans (monthly values)
ESA forest coverage and AVHRR-based vegetation data (local data set)	0.1° (SW: 30°, -25°; NE: 72°, 61°) → 860 × 420	- portion of vegetation - leaf-area index - thermic and kinetic roughness length - albedo
One year assimilation experiment with the ARPEGE global model (global data set)	1.5° → 240 × 120	- soil temperature and humidity - sea surface temperature

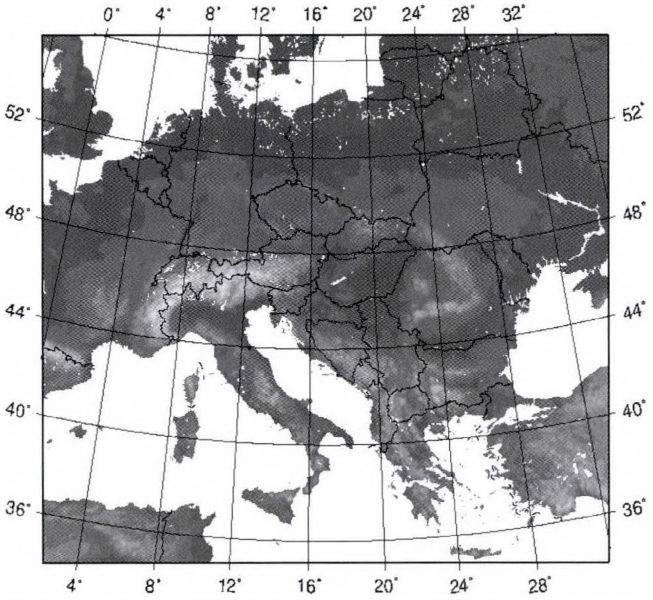


Fig. 1. An example for the output of configuration 923: the domain and topography contours of the ALADIN/HU version of the ALADIN model.

3. Computation of initial and lateral boundary conditions (configuration 927)

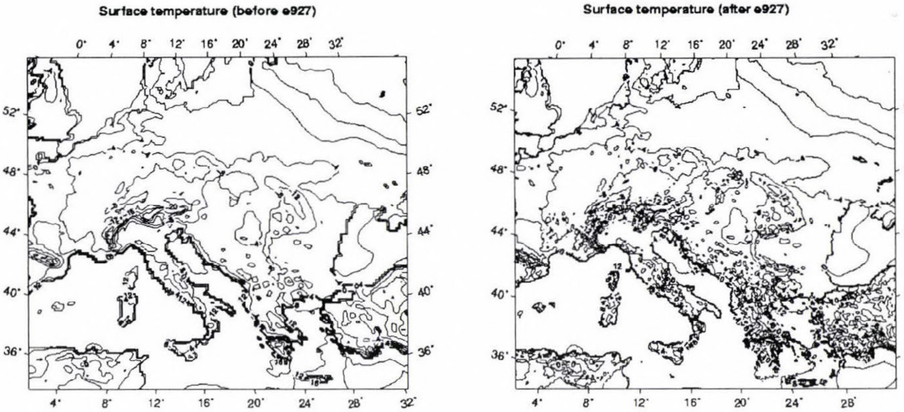
For ALADIN, being a limited area model, not only the initial, but also the lateral boundary conditions are indispensable for the model integration. Originally the ARPEGE global model or another ALADIN model version was considered (as input) for the creation of lateral boundary conditions (nowadays the ECMWF/IFS model can be also applied as driving model, see next section). It is emphasized that the initial conditions can be computed similarly to the lateral boundary conditions (formally and technically there is no distinction between initial and boundary information), and it is an essential step for the dynamical adaptation integration (where there is no independent data assimilation cycle, the initial conditions of the limited area model are obtained through interpolation from the global model's analysis, and furthermore, during the model integration the model is adapting the meteorological fields of the driving model to the higher resolution surface characteristics of the limited area domain).

This configuration in practice is an interpolation algorithm, where the meteorological variables are interpolated (horizontally and vertically) from the input grid (at a given domain and resolution) to the target one. At the same time the final characteristics of the integration domain is set (for instance the horizontal and vertical resolution of the model). In some cases the interpolation is done on departures, i.e., on differences between the global model's value and the climatic characteristics (for instance for surface temperature). Those variables are represented in such differences, where the value of the variable highly depends on the altitude, therefore huge systematic errors might be encountered in case of neglecting the height differences between the input and target domains (using the departures, the problem of the orography can be automatically eliminated).

In practice most of the model input fields are stored in form of spectral coefficients (as long as the number of spectral coefficients is smaller than the number of gridpoints due to spectral over truncation, therefore, it is a more economic storage than that of the gridpoint values). First, the spectral fields are inverse Fourier transformed to gridpoint space (naturally the interpolation is performed in the physical - gridpoint - space). So this operation is followed by the horizontal and then vertical interpolations to the target grid. It is noted that extrapolation is needed, if the lowest model level of the target grid is situated under the lowest model level of the input grid (certainly it might cause some additional error in the computations). Finally, with the help of direct Fourier transformation the meteorological fields are transformed back to spectral space again.

There are three versions of this configuration available: configuration 927 is in fact the interpolation from global grid to another global grid (ARPEGE → ARPEGE); in case of E927 configuration the interpolation is carried out between global and limited area domains (ARPEGE → ALADIN); EE927 is the configuration between two different limited area applications (ALADIN → ALADIN). Certainly, in the latter case it is required that the target domain is entirely embedded in the input one (no possibility for horizontal extrapolation).

In the everyday operational practice E927 is executed in Toulouse at Météo-France at every ARPEGE model run for creating initial and lateral boundary conditions for a limited area domain for the partners (it is basically a cut from the global domain, keeping the global resolution, taking into account the telecommunication constraints). Locally (at the ALADIN members states), after transferring the information prepared in Toulouse EE927 is executed, where the new, higher resolution model grid is created (needed for the model integration). An example of input and output fields of EE927 can be seen in *Fig. 2* for illustration (for the Hungarian domain).


Surface temperature (before e927)

Surface temperature (after e927)

Fig. 2. An example for configuration 927. Surface temperature before (left panel) and after (right panel) EE927 interpolation (24h forecast based on the 00:00 UTC, February 23, 2006 integration).

4. Application of ECMWF model forecasts for lateral boundary conditions (configuration 901)

The ARPEGE/ALADIN modeling system was originally designed for the use of the ARPEGE "mother" system as initial and lateral boundary conditions. Nevertheless, recently more and more ALADIN partners wish to apply the ECMWF (IFS) model results for that purpose. This objective can be met at the moment by the use of configuration 901 and related applications, which is going to be described briefly hereafter.

There are two main reasons why the IFS model results cannot be directly used as initial or lateral boundary information. First, the file formats are different (IFS uses GRIB, while ARPEGE/ALADIN uses ARPEGE file – FA – format). The second reason is that ARPEGE/ALADIN uses different surface parameterization than IFS, therefore, there are variables (e.g., deep soil temperature) which are not present in the ECMWF data, but are needed by the ISBA (Interactions Soil-Biosphere-Atmosphere) surface parameterization scheme (*Noilhan* and *Planton*, 1989) of ARPEGE/ALADIN. To overcome these problems, one should first run a configuration 901 without using climatological information resulting in a global ARPEGE file (in FA format including the upper-air and surface variables coming from the IFS model). Then a configuration 923 should be run to create a climate file, which uses the orography and land-sea mask read from the output of the previous step. Finally, another 901 configuration should be executed, but this time using the climate information provided by the previous 923 run. The output of this last step will be an ARPEGE (FA format) file that basically contains upper-air and surface fields from the IFS model output and some surface fields from climate database. From this global data the usual procedure is applied in order to arrive to the limited area model information (see previous section).

One of the "trigger" needs for applying ECMWF/IFS data for initial and lateral boundary conditions for the ALADIN model was the requirement to compute detailed high resolution wind climatology over a domain of interest. The existence of the ECMWF 40 years re-analyses system (*Simmons* and *Gibson,* 2000) served as an excellent input information for achieving such goals, and the ALADIN model was considered as a good tool for the downscaling process.

At the Hungarian Meteorological Service the dynamical downscaling of ERA-40 data was performed for a Hungarian domain of 5 km resolution for a 10 years period (1992–2001). Due to the fact that the difference between the target resolution (5 km) and the ERA-40 resolution (~125 km) was quite significant, it was not obvious how many intermediate integration steps were needed to reach the optimal result. Finally it was decided to use two nested ALADIN integration steps on 45 and 15 km resolution, respectively. In the final step ALADIN dynamical adaptation (DADA), developed for wind and precipitation (*Žagar* and *Rakovec,* 1999), was applied to reach the desired 5 km resolution (*Fig. 3*). More about this special dynamical adaptation can be found in the next section.

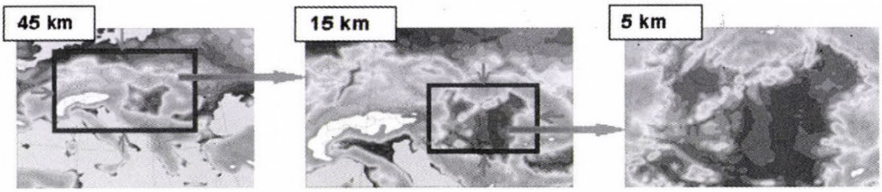


Fig. 3. The steps of the dynamical downscaling method using nested domains with increasing resolution. The orography of the different applied ALADIN domains are shown.

According to the preliminary results it can be said, that the downscaled wind fields at the planetary boundary layer (in the lowest 150 meters) are reasonably realistic with some overestimation of the wind field for the lower levels (see *Fig. 4*).

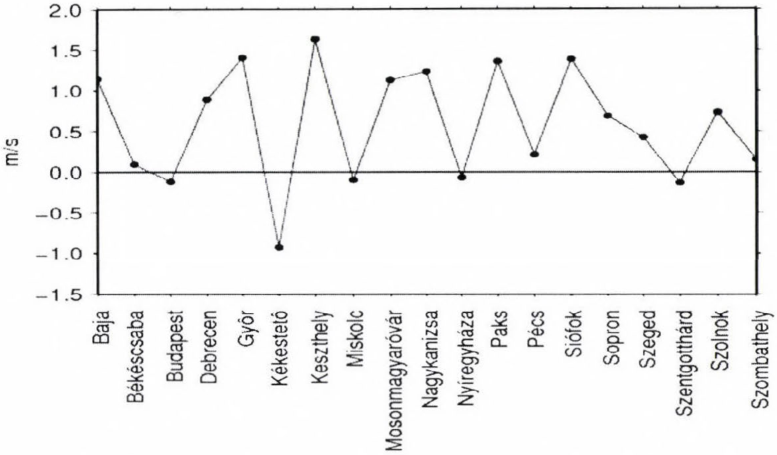


Fig. 4. The average difference between the computed and observed average 10 m wind speed for 19 Hungarian stations for the period between January 1, 1992 and December 31, 2001.

5. Model integration (configuration 001)

In the presence of pre-computed climatological (surface) fields, initial and lateral boundary conditions the model integration can be realized. The integration in fact is preceded by the dynamical initialization: the gravity waves, which might be originated due to the inconsistencies in the observational data and/or the interpolation methods are to be eliminated by digital filter initialization technique (*Lynch* and *Huang,* 1992). During the digital filter initialization the high frequency components in the time series of the prognostic variables are eliminated. For the execution of the algorithm the model is integrated forward and backward (certainly the backward integration is quasi-adiabatic). As a result of the initialization, the initial conditions of the model become free of high frequency gravity waves that would cause instability during the model integration. The ALADIN numerical weather prediction model solves the primitive equations of the atmosphere using Semi-Implicit Semi-Lagrange (SISL) numerical approximation. According to the transformation technique, the linear terms of the equations are solved in spectral space, while the nonlinear ones in gridpoint space (otherwise in a purely spectral environment the nonlinear terms would be very complicated and cumbersome to evaluate in the spectral space). The drawback of the transformation method is that at every model timestep, transformation is needed between spectral and gridpoint spaces. This can be computationally feasible, because the (Fourier) transformation can be solved in a very efficient and fast way: it is ensured by the Fast Fourier Transformation (FFT) method.

A relatively new part of the ALADIN model integration configuration is the efficient non-hydrostatic dynamics (*Bubnová et al.,* 1995), which is the heart of the future improvements of the system. The original constraint of the non-hydrostatic developments was to find a solution for the modeling system, which makes the reasonably easy and simple implementation of the non-hydrostatic core of the model possible on top of the already existing hydrostatic version. This condition was fulfilled by the implementation of hydrostatic pressure vertical coordinate system based on the idea of *Laprise* (1992). In this coordinate system the vertical coordinate is the π hydrostatic pressure, which is given in the following form (g is the gravity constant and ρ denotes the density):

$$\pi(x,y,z,t) = g\int_{z}^{\infty}\rho(x,y,z',t)\,dz' .$$

Euler Equations (EE) cast in this new coordinate system have, as usual, two additional prognostic variables compared to the HPE (Hydrostatic Primitive Equations) system: the first one accounts for true pressure and the second one for the vertical acceleration when abandoning the hydrostatic assumption. As it regards the vertical coordinate itself, the time evolution of π obeys the classical HPE diagnostic continuity equation. It can be easily shown that when $p = \pi$ the EE system simplifies to the HPE one cast in the p-coordinate. Therefore, the π-coordinate is a natural extension of the p-coordinate for the non-hydrostatic case. It is important to mention that the compressibility allows for the vertically propagating acoustic waves in the solution, which invokes very severe stability limitations (due to the Courant-Friedrichs-Lewy stability criteria) for the numerical solution. In practice this would mean an extremely small time step for the model integration. This difficulty is in general overcome either by the anelastic approximation (e.g., Méso-NH model, see *Lafore et al.* (1998)) or the application of a special numerical algorithm. The latter choice was natural for the ALADIN system, using already the Semi-Implicit Semi-Lagrangian (SISL) algorithms for integrating its HPE version. Following the numerical analysis and proposals of *Bénard* (2003, 2004) and *Bénard et al.* (2004, 2005) for the description of the proper choice of model prognostic variables, the conditions for the acoustic-gravity linear system with constant coefficients and the class of iterated centered implicit schemes, the non-hydrostatic compressible equation system of ALADIN is integrated in time using the very computationally efficient two-time-level SISL scheme. The important fact is that this stability and efficiency, allowing for time steps about ten times longer than those imposed by the explicit time integration schemes, is reached without a loss of accuracy compared to other non-hydrostatic models (*Bouttier et al.,* 2003). In addition, the implementation of the non-hydrostatic dynamical core (including the numerical algorithms) as an extension to the HPE one, offers the possibility of clean comparisons between the non-hydrostatic and hydrostatic solutions, which is a great advantage.

The previously described dynamical part of the model is complemented by the description of sub-grid scale phenomena and such processes that have far too complex physical behavior for their direct consideration in the model equations. These processes are the so called physical parameterizations (for instance convection, gravity wave drag, radiation, microphysics, etc.), which are represented in the model in a simplified manner. In ALADIN the following physical parameterization schemes are considered:

- Radiation: There are several radiation schemes that can be used operationally. ACRANEB (a blend of *Geleyn* and *Hollingsworth* (1979) and *Ritter* and *Geleyn* (1992), now much improved through the use of a Net Exchanged Rates formalism, see *Geleyn et al.* (2005)), FMR (*Morcrette*, 1991), and RRTM (*Mlawer et al.,* 1997) for the thermal part only. These latter two schemes are computationally more expensive, so they are called with reduced frequency (not every time step).
- Surface: 2 layer ISBA scheme (*Noilhan* and *Planton,* 1989) is used. Snow is taken into account through a one layer snow scheme (*Douville et al.,* 1995).
- Turbulence: 1D (vertical) diffusion is calculated based on mixing length calculation (*Louis et al.*, 1982). Shallow convection is also treated inside this calculation (*Geleyn*, 1987).
- Large-scale precipitation: Diagnostic scheme based on saturation. There is no prognostic liquid water, precipitation falls out in one time step. Evaporation and melting of precipitation are calculated with Kessler-type formulae (*Kessler*, 1969).
- Cloudiness: Diagnostic scheme adapted from Xu and Randall (*Xu* and *Randall*, 1996) and enhanced in order to account for PBL moisture mixing and for temperature inversion strength (*Brožková et al.*, 2006).
- Convection: Mass-flux approach (*Bougeault*, 1985). One updraft and one downdraft are taken into account (*Ducrocq* and *Bougeault*, 1995). Kuo-type closure is applied. Several enhancements have been introduced, since the scheme was used for finer and finer meshes (*Gerard* and *Geleyn*, 2005).

- Gravity wave drag: There are two pragmatic possibilities to take into account the effects of subgrid scale orography: with envelope orography (and an ad-hoc tuning of the wave- plus form-drag scheme) or without envelope but with a more complex parameterization of mountain drag and lift effects in this case (*Catry*, 2006).

The physical parameterizations are contributing to the equations as fluxes to the given tendency equation. As an example for the horizontal momentum equation, the $\frac{\partial \mathbf{V_h}}{\partial t}$ term (where $\mathbf{V_h}$ is the horizontal wind) describes the contributions of the physical parameterizations to the tendency of the momentum:

$$\frac{\partial \mathbf{V_h}}{\partial t} = -g \frac{\partial}{\partial p}\{ J_{\mathbf{V}}^{turb} + J_{\mathbf{V}}^{gwd} + J_{\mathbf{V}}^{conv} \}. \qquad (2)$$

The $J_{\mathbf{V}}$ quantities in this equation denote the fluxes of different physical processes (turbulence, gravity wave drag, and convection, respectively).

Certainly, during the computation of physical fluxes of the parameterizations there are some quantities, which are computed diagnostically (typically cloudiness and precipitation are such variables). ALADIN uses un-lagged physics, which means that the physical terms are evaluated at the origin instant (at t for the two-time-level Semi-Lagrangian scheme). The physics is calculated in parallel way, i.e., all physical tendencies are computed separately using the same origin instant, and the total tendency is the sum of all tendencies.

Special version of the model integration is the small scale wind dynamical adaptation (*Žagar* and *Rakovec,* 1999). The low level wind field is determined mostly by the surface characteristics (especially the orography), therefore, if the orographical conditions are sufficiently precisely described, one might attempt to compute very accurate low level wind forecasts (without the usage of the sophisticated full model). In the reality the wind adaptation is carried out on 2–3 km horizontal resolution and around 10–15 vertical model levels (in the planetary boundary layer). The model is also simplified: quasi-adiabatic version (using only vertical diffusion and gravity wave drag parameterization for instance) with around 30 minutes of integration. This very short integration is sufficient for the adjustment of the large scale wind fields to the local environment. It is mentioned that this kind of wind adaptation is very successfully used at most of the ALADIN partners computing local wind conditions (*Fig. 5*) and also deriving detailed, high resolution wind climatology information (see previous section).

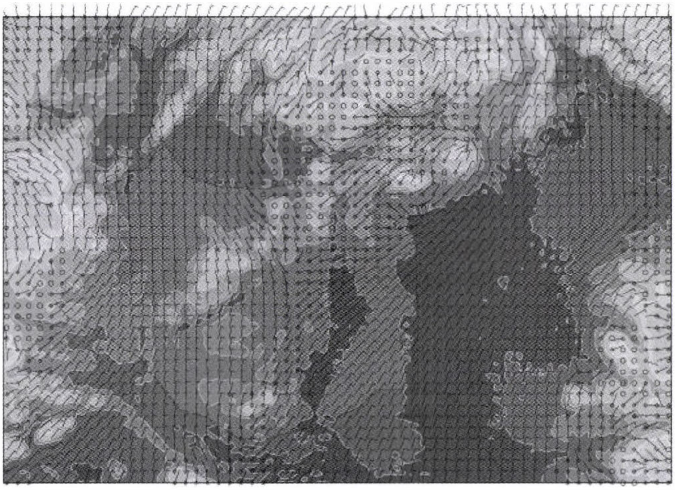


Fig. 5. The output of the 10 m wind forecast by the small scale wind dynamical adaptation over Hungary (valid at 00:00 UTC, January 19, 2006). The high resolution orography of the model is superimposed in the background. The details of the flow are much more pronounced than in the original (coarse resolution) wind field.

6. Post-processing (configuration full-pos)

An essential part of the model exploitation (for the visualization, interpretation, and verification of the numerical weather prediction products) is the post-processing: the model results are transformed in a "user friendly" coordinate system together with the computation of special variables (see *Fig. 6* for an example), supporting the direct application of the model results. In the ARPEGE/ALADIN terminology all the post-processing related computations are called full-pos (since usually post-processing is performed online during the model integration, there are no configuration numbers, however, in fact the earlier mentioned configuration 927 is also part of the full-pos package). The post-processing consists of various transformation steps, as spectral transformation from spectral space to physical space, change of projection (for instance from Lambert projection to spherical latitude-longitude coordinate system), change of vertical coordinate system (for instance pressure, height, or isentropic coordinates instead of the hybrid ones used for the model integration), computation of special diagnostic variables (for instance height of the planetary boundary layer or potential vorticity), etc. In practice the post-processing is performed simultaneously to the model integration in order to produce the model outputs as soon as possible for further processing.

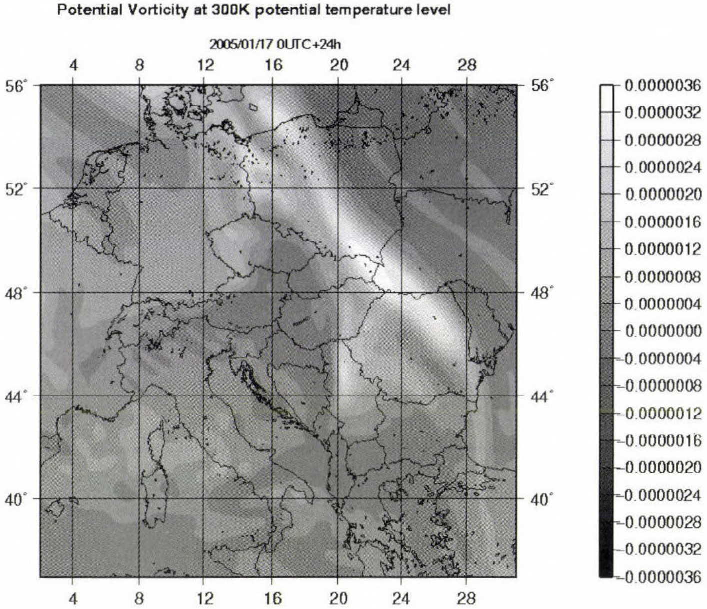


Fig. 6. An example for post processing: the potential vorticity is visualized on the 300 K isentropic level (24h forecast valid at 00:00 UTC, January 18, 2005).

7. Data assimilation (configuration 701, 131)

As it was mentioned earlier for the original design of the ALADIN system, there were no data assimilation methods planned to be developed (the model was considered only for pure dynamical adaptation of the global ARPEGE model). Nevertheless, soon the need for the development of data assimilation schemes emerged: first the global optimal interpolation scheme CANARI was developed for the limited area (configuration 701) and then the three-dimensional variational data assimilation (3d-var) scheme (configuration: 131) was adapted. The modern data assimilation algorithms are using all the information available from the atmosphere: first of all different kinds of observations (for instance surface and upper air measurements or remote sensed data), then previous models forecasts ("background"), and additionally any other a priori physical information about the atmosphere. The need for the background information is coming from the fact that the number of available observations is far too few as compared to the degree of freedom of the model (i.e., number of gridpoints). Therefore, there is a strong necessity to include additional information sources: the background (as a short range model forecast originating from a previous time), which is on the one hand only an approximation of the truth being "only" a forecast, on the other hand it gives sufficient amount of information for the solution of the analysis problem. Very important ingredient of the data assimilation algorithms is the knowledge about the reliability of the available information sources (this information gives guidance for the optimal mixture of these elements). During data assimilation the observation and background information are optimally mixed in order to obtain a precise and dynamically consistent initial state for the model integration. This process is in fact a cyclic one: the initial conditions of the model are used for the model integration, which provides a new background for the next analysis time and then taking into account new observations, a new analysis is performed (and it is repeated continuously).

Depending on how the observations and the background information are merged in the data assimilation process, two main algorithms are used in practice (*Bouttier* and *Courtier,* 1999): optimal interpolation (OI) and variational data assimilation (3d-var and 4d-var). The first method is optimal in a least square sense, i.e., the expected value of the quadratic difference between the analysis and the true state is minimal. For the variational systems it is ensured, that the resulting analysis is situated near to the observations and the background information as well. For this method a cost function ($J(\delta\mathbf{x})$, in the incremental version) is defined:

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathbf{T}}\mathbf{B}^{-\mathbf{1}}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathbf{T}}\,\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) + \text{possible constraints}, \quad (3)$$

where $\delta\mathbf{x}$ represents the departure between the analysis and the background, **B** is the background error covariance matrix, **R** is the observation error covariance matrix, $\mathbf{d} = \mathbf{y}_0 - H(\mathbf{x}_b)$, where H is the observation operator, **H** is the linearized observation operator around the background state, $\mathbf{x}_b$ is the background field, and $\mathbf{y}_0$ vector contains the observations. The analysis problem is solved by the minimization of the cost function (with respect to $\delta\mathbf{x}$), and the resulting increment is added to the background field.

Recently at the Hungarian Meteorological Service the initial conditions of the ALADIN model are operationally obtained through the three-dimensional variational data assimilation algorithm (3d-var). *Fig. 7* illustrates the efficiency of the 3d-var scheme with respect to the dynamical adaptation one (the forecast computed from the 3d-var analysis is much better than that from the dynamical adaptation without assimilation). More details of the ALADIN variational data assimilation scheme can be read in the article of *Bölöni* (2006) in the same volume.

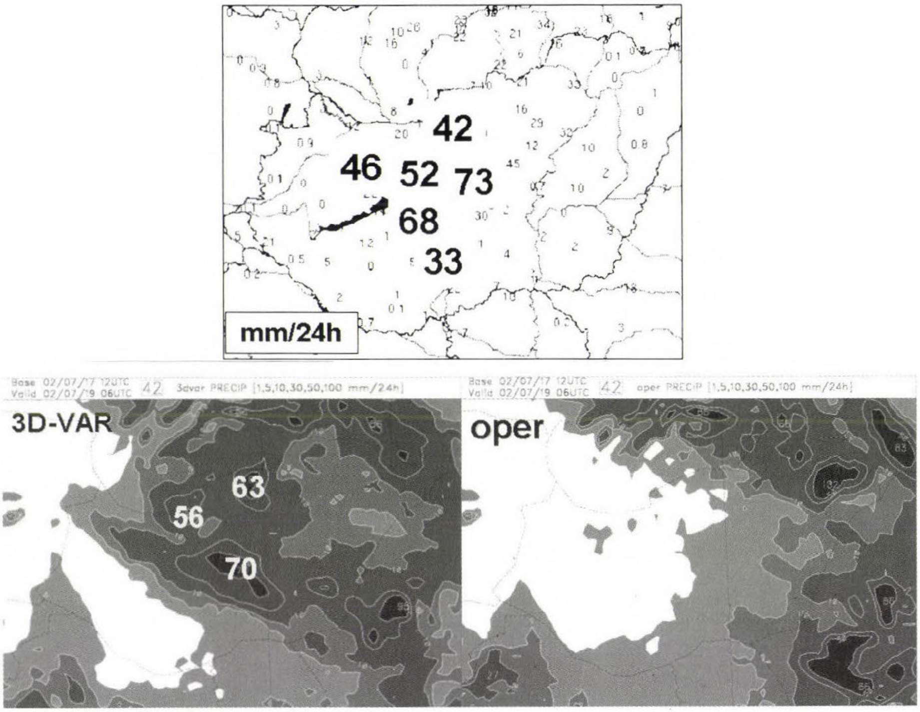


Fig. 7. Example for the difference between precipitation forecasts started from local three-dimensional variational (3d-var) data assimilation (bottom left) and using dynamical adaptation (bottom right). 24h accumulated precipitation is shown between 06:00 UTC, July 18, 2002 and 06:00 UTC, July 19, 2002. The top panel shows the observed precipitation values at the observation locations.

8. Sensitivity studies (configuration 801)

The adjoint version of the model establishes relationships between gradient fields, therefore, it is a convenient tool for the use of sensitivity studies (where one would like to know what is the sensitivity of a given forecast aspect with respect to the initial conditions of the model). For that purpose a diagnostic function is defined ($J(\mathbf{y})$, where J is the diagnostic function itself valid at t_1 and $\mathbf{y}$ is the forecasted field also at t_1), which measures for instance the reliability (preciseness) of the forecast in terms of differences between the forecast and a verifying analysis field. For the possibility to use different forecast variables (having different order of magnitudes and units) in a single diagnostic function, one might define an expression with energy dimension for instance.

While performing sensitivity experiments, one would like to assess what is the initial $\delta\mathbf{x}$ perturbation valid at t_0, where the difference between the diagnostic functions (valid at t_1) derived from an unperturbed and perturbed initial conditions is maximal. This latter difference can be written as

$$\delta J = J(y+\delta\mathbf{y}) - J(\mathbf{y}) = \frac{dJ(\mathbf{y})}{d\mathbf{y}}\delta\mathbf{y} = \left\langle \nabla_{t_1} J, \delta\mathbf{y} \right\rangle_Y , \tag{4}$$

where the maximum is determined by the $\nabla_{t_1} J$ gradient valid at t_1. The $\langle\ ,\ \rangle_Y$ expression denotes the scalar product defined on the Y vector space (this might not be the classical Euclidean scalar product, therefore, this is a generalized gradient computation). This result is not yet satisfactory, because one needs the gradient with respect to the initial conditions (and not to the final state), therefore, there is a need to find a relationship between $\nabla_{t_1} J$ and $\nabla_{t_0} J$. This is established by the adjoint model operator. For the definition of the adjoint model the linearity assumption should be fulfilled, i.e., the evolution of the perturbations can be described by the tangent linear approximation (the adjoint model is in fact the adjoint of the tangent linear model).

The tangent linear model describes the linearized evolution of small perturbations along the nonlinear model trajectory. The nonlinear model transports the $\mathbf{x}$ model state valid at t_0 to $\mathbf{y}$ state valid at t_1 with the help of the $\boldsymbol{M}$ nonlinear operator:

$$\mathbf{y} = \boldsymbol{M}\mathbf{x} . \tag{5}$$

If a small $\delta\mathbf{x}$ perturbation is added to the initial state of the model and the original and perturbed solution (integration) in the $[t_1, t_0]$ time interval is not significantly different, then the $\boldsymbol{M}$ nonlinear operator can be approximated with its first order Taylor polynomial:

$$\mathbf{y} + \delta\mathbf{y} = \boldsymbol{M}(\mathbf{x} + \delta\mathbf{x}) \approx \boldsymbol{M}\mathbf{x} + \frac{d\boldsymbol{M}}{d\mathbf{x}}\delta\mathbf{x} = \boldsymbol{M}\mathbf{x} + \mathbf{L}\,\delta\mathbf{x} . \tag{6}$$

It can be seen that the tangent linear operator $\mathbf{L}$ describes the relationship between the initial and final perturbations: $\delta\mathbf{y} = \mathbf{L}\delta\mathbf{x}$.

Based on the definition of the adjoint operator, the gradient with respect to the initial conditions can be described in the following way (by using the gradient with respect to the final state):

$$\delta J = \left\langle \nabla_{t_1} J, \delta\mathbf{y} \right\rangle_Y = \left\langle \nabla_{t_1} J, \mathbf{L}\delta\mathbf{x} \right\rangle_Y = \left\langle \mathbf{L}^{*}\nabla_{t_1} J, \delta\mathbf{x} \right\rangle_X = \left\langle \nabla_{t_0} J, \delta\mathbf{x} \right\rangle_X , \tag{7}$$

resulting in

$$\nabla_{t_0} J = \mathbf{L}^{*}\nabla_{t_1} J . \tag{8}$$

Therefore, the gradients (sensitivities) with respect to the initial conditions can be obtained through this relationship with the integration (from t_1 to t_0) of the adjoint model. This gradient field can be rescaled to perturbation with a conveniently defined α scaling factor: $\delta\mathbf{x}_0 = -\alpha\nabla_{t_0} J$. Then, it can be checked that applying this $\delta\mathbf{x}_0$ perturbation to the initial conditions, the corrected forecast will be nearer to the verifying analyses.

With the help of such sensitivity studies one can assess the sensitive aspects (in terms of geographical locations or variables) of the initial conditions. An example of such sensitivity (gradient) field can be seen in *Fig. 8*.

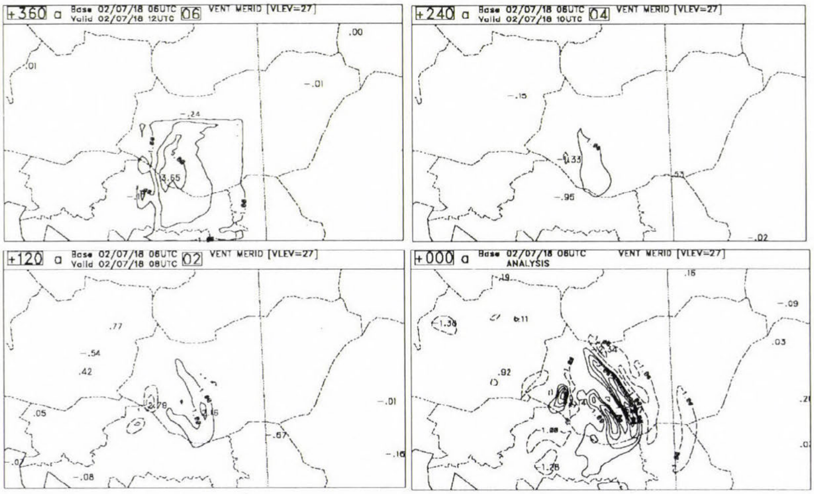


Fig. 8. Evolution of gradient (sensitivity) fields with respect to the meridional wind component on the 27th model level during the adjoint integration between 12:00 and 06:00 UTC, July 8, 2002 (upper left panel is valid for 12:00 UTC, the upper right is for 10:00 UTC, the lower left is for 08:00 UTC and the lower right is for 06:00 UTC). Dashed lines represent negative values. It can be seen for this particular case (a local convective event), that the sensitive areas for the wind field are located nearby to the area of interest, where the convection occurred.

One of the difficulties, while using the adjoint model is the fact that the nonlinear model contains such threshold processes (mostly in the physical parameterization packages), where the applied functions are non-continuous, therefore, their linearization is not possible (so the adjoint does not exist either). One of the solution overcoming this problem is the creation of simplified and regularized nonlinear physical parameterizations (*Janiskova et al.,* 1999), where the tangent linear and adjoint versions can be determined (although the simplified scheme certainly differs in a way from the original algorithm). Having adjoint versions of the physics parts of the model permits to study the relative role of the physical parameterizations (and their adjoint) in the sensitivity results (it is especially important at mesoscale, where such processes have crucial importance, see *Soci et al.* (2006)).

Finally, it is important to note that before performing sensitivity studies, one should carefully check (beside the linearity test anticipated above) the validity of the tangent linear and adjoint models, respectively. There are special configurations for this job in the ALADIN world: configurations 501 and 401. The test of the tangent linear model (configuration 501) controls whether the tangent linear model version is really the linear version of the nonlinear model. This can be easily checked with the help of the Taylor formula. Typical validity time for the linear assumption is 1–2 days (this might depend on the size of the perturbations and the described physical processes included in the model). The test of the tangent linear model might be followed by the test of the adjoint model (configuration 401). This test is based on the rigorous numerical control of the adjoint equation (the definition of the adjoint operator).

9. Computation of singular vectors (configuration 601)

Singular vector decomposition is popularly used for the generation of initial perturbations for an ensemble prediction system: more model integrations are performed with slightly different initial conditions (all of them are equally possible initial states for the forecast model), and the spread of the forecast results are evaluated in order to quantify the predicted reliability of the forecast (this is a very valuable extra information on top of the forecast itself). The perturbations needed for the creation of the different initial states might be computed by using different methods (for instance random perturbations or breeding method beside the singular vector computations). The main objective of the creation of the initial perturbations is to obtain such perturbations, which are going to develop in the fastest manner. The singular vector framework is an ideal solution for this problem (*Buizza et al.,* 1993). Singular vectors are computed by the configuration 601 of the ALADIN model.

As it was anticipated, the evolution of the initial $\delta\mathbf{x}$ (valid at t_0) perturbation is of interest. It is assumed that this can be approximated by the tangent linear model: $\delta\mathbf{y} = \mathbf{L}\delta\mathbf{x}$, where $\mathbf{L} = dM / d\mathbf{x}$ is the tangent linear model ($\delta\mathbf{y}$ is valid at t_1). Those perturbations are sought, where the norm of the perturbations is amplified most, so the following problem ought to be solved:

$$\max\left(\frac{\|\mathbf{L}\delta\mathbf{x}\|_E}{\|\delta\mathbf{x}\|_D}\right), \tag{9}$$

where $\|.\|_D$ and $\|.\|_E$ denote norms at initial and final time (certainly the two norms might be different). It can be seen that the perturbations depend on the choice of the norm (usually the energy norm is used in practice). The norm is defined by the scalar product, so, e.g., the D norm is obtained through the $\mathbf{D}$ positive definite matrix: $\|\mathbf{x}\|_D = \sqrt{\langle \mathbf{x},\mathbf{x}\rangle_D} = \sqrt{\mathbf{x}^\mathbf{T}\mathbf{D}\mathbf{x}}$. If the initial perturbation's norm is assumed to be unit, then it is sufficient to examine the maximum of the numerator. Using the adjoint definition, this term can be written as:

$$\|\mathbf{L}\delta\mathbf{x}\|_E = \sqrt{\langle \mathbf{L}\delta\mathbf{x},\mathbf{L}\delta\mathbf{x}\rangle_E} = \sqrt{\left\langle \mathbf{L}^*\mathbf{L}\delta\mathbf{x},\delta\mathbf{x}\right\rangle_D}\,. \tag{10}$$

On the right hand side of this equation the quadratic form defined by the $\mathbf{L}^*\mathbf{L}$ operator under the square root is positive definite, so all the eigenvalues of the $\mathbf{L}^*\mathbf{L}$ operator are positive. Besides, since $\mathbf{L}^*\mathbf{L}$ is self-adjoint, there exists an orthonormal eigenvector system for it. If $\gamma_1^2 \ge \gamma_2^2 \ge \ldots \ge \gamma_n^2$ denotes the eigenvalues of $\mathbf{L}^*\mathbf{L}$ and $\mathbf{v}_\mathbf{1}, \mathbf{v}_\mathbf{2}, \ldots, \mathbf{v}_\mathbf{n}$ the corresponding orthonormal eigenvectors, then the $\gamma_1, \ldots, \gamma_n$ numbers are called the singular values of the $\mathbf{L}$ operator and the $\mathbf{v}_\mathbf{1}, \mathbf{v}_\mathbf{2}, \ldots, \mathbf{v}_\mathbf{n}$ vectors are the right-hand side singular vectors of the $\mathbf{L}$ operator. It can be proven that the maximum of $\|\mathbf{L}\delta\mathbf{x}\|_E$ is γ_1, and the maximum location is represented by $\mathbf{v}_\mathbf{1}$. Additionally, in the orthogonal complementary subspace of $\mathbf{v}_\mathbf{1}$, the maximum is γ_2 at location of $\mathbf{v}_\mathbf{2}$, and so on for the other orthogonal subspaces. Finally, this algorithm leads to γ_n and $\mathbf{v}_\mathbf{n}$, which define the minimum value and location of the perturbation growth rate, respectively. Therefore, the singular vectors are those unit perturbations, which are growing most rapidly towards significantly different (orthogonal) directions. It means that instead of considering the maximization problem above, the following eigenvalue problem can be solved:

$$\mathbf{L}^\mathbf{T}\mathbf{E}\mathbf{L}\mathbf{v} = \gamma^2\mathbf{D}\mathbf{v}\,. \tag{11}$$

For the solution of such eigenvalue problem the Lánczos algorithm can be used, which is using an iterative algorithm for finding the eigenvalues. Typical singular vectors are shown in *Fig. 9*.

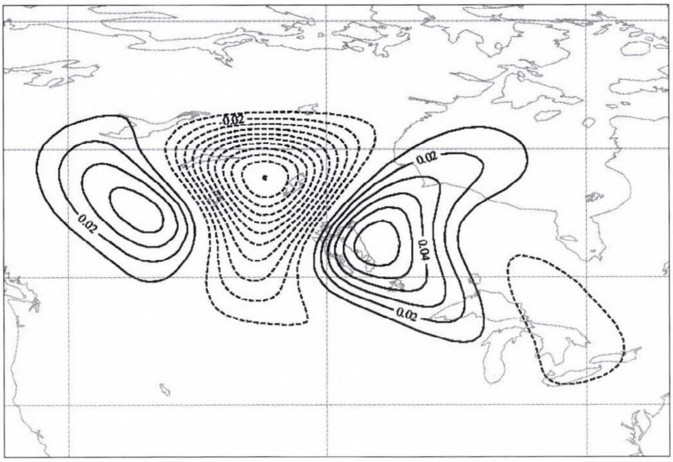


Fig. 9. Typical result of singular vector calculations (the usual dipole structure of the singular vectors can be clearly identified) valid at 12:00 UTC, June 17, 2002. The leading singular vector is visualized for temperature on the 10th model level. Solid lines indicate the positive and dashed lines the negative values.

10. Plans for the future (AROME model)

One of the most important aspects of the future evolution of the ALADIN family is the development of the AROME (Application of Research to Operations at Mesoscale) non-hydrostatic model for ultra-short and short range numerical weather prediction. It is already decided that the main skeleton of the AROME model will be built from the non-hydrostatic kernel and the three-dimensional variational data assimilation scheme of the ALADIN model, together with the physical parameterization package of the Méso-NH French research model (*Lafore et al.,* 1998). The AROME model uses 5 physical parameterization schemes listed briefly hereafter:

- For radiation the RRTM (rapid radiative transfer model) scheme is used (*Mlawer et al.,* 1997).
- The 1D version (only vertical diffusion is assumed) of the Méso-NH 3D turbulence scheme is used (*Cuxart et al.,* 2000). In order to calculate the turbulent kinetic energy (TKE), a prognostic equation is applied, while all the non-isotropic 2nd order moments are diagnosed. The closure is based on mixing length calculation (*Bougeault* and *Lacarrere,* 1989).
- For the proper description of the moist processes the AROME model uses sophisticated microphysics scheme (*Pinty* and *Jabouille*, 1998), originally developed for the Méso-NH model. Five prognostic variables (cloud

water, cloud ice, rain water, snow, and graupel) are taken into account (later on hail will be also included).

- The surface parameterization distinguishes between four types of surface: nature, urban area, sea, and inland water. For sea and inland water the parameterization scheme is not yet fully properly developed, only the Charnock formula is applied. Over nature the ISBA scheme (*Noilhan* and *Planton,* 1989) is used (similarly to ALADIN, but in AROME the scheme is more sophisticated, e.g., three layers snow scheme, possible switch to the diffusion scheme, etc.) For urban areas the Town Energy Balance (TEB) scheme (*Masson,* 2000) is applied.
- There is also a possibility to turn on deep-convection parameterization (*Bechtold et al.,* 2001), while using the model on coarser resolution (e.g. 10 km). At higher resolution (below 2.5 km) this is not needed, since the processes dealing with precipitation are already described explicitly by the microphysics parameterization, only shallow (non precipitating) convection is used. This parameterization is based on mass-flux approximation using CAPE closure.

As far as data assimilation is concerned, the AROME model will use a high frequency three-dimensional variational data assimilation system taking into account different intensive observations with special emphasis on radar data. These enhancements (beside others) will transform the recent assimilation scheme into a highly efficient mesoscale data assimilation tool.

One of the first results obtained at the Hungarian Meteorological Service based on the prototype version of AROME is shown in *Fig. 10.*

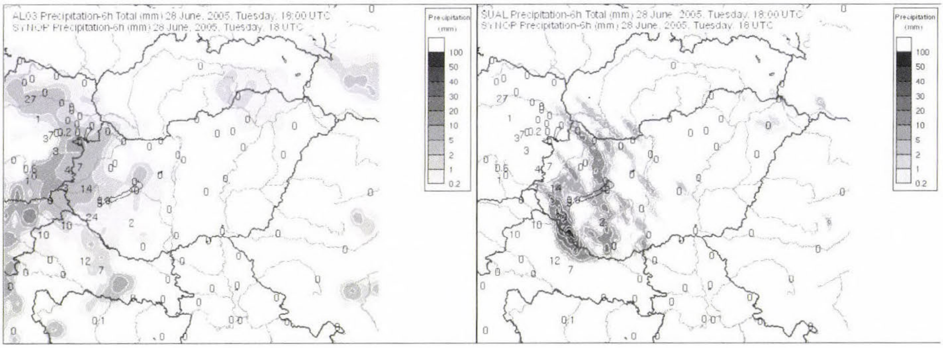


Fig. 10. An example for AROME forecast. 6h accumulated total precipitation after 18h integration for ALADIN dynamical adaptation (left panel) and AROME (right panel). SYNOP (surface) observations are marked with numbers on both panels. The forecast started at 00:00 UTC, June 28, 2005. It can be seen that the AROME model provides a more precise precipitation forecast.

11. Summary

The present article was trying to give a general overview about the configurations available at the ARPEGE/ALADIN model family demonstrating the fact, that the ALADIN model is not "only" a limited area numerical weather prediction model, but a whole family of operational and research applications for the numerical weather prediction community.

First, the basic pre-processing tools were summarized: computation of surface (climate) information for the domain of interest (configuration 923), computation of initial and lateral boundary conditions (configuration 927), and the possible application of ECMWF/IFS data for initial and boundary conditions for the ALADIN model (configuration 901). These three configurations are basically sophisticated interpolation procedures with several options for producing the necessary information for the model integration. Certainly one of the most important configuration is the model integration (configuration 001), which allows to solve the hydro-thermodynamical set of equations (in its hydrostatic and non-hydrostatic form as well) of the atmosphere over a limited area. The integration is accompanied by the post-processing (configuration full-pos) for providing special condensed information for the model users. Nowadays, data assimilation is an integral (and growing) part of the modeling system allowing the application of two basic schemes: optimal interpolation (configuration 701) and three-dimensional variational data assimilation (configuration 131). Recently the latter configuration is being heavily developed further to the four-dimensional framework. There are several research configurations available: for instance the ones dealing with sensitivity computations (configuration 801) and computation of singular vectors (configuration 601). The sensitivity configuration makes to quantify those sources (location and variable) of errors possible in the initial conditions, which might lead to incorrect forecasts. The singular vectors are widely used for computing initial perturbations to an ensemble prediction system. These two latter configurations are accompanied by the test of their basic ingredients: test of the tangent linear (configuration 401) and adjoint models (configuration 501).

All of these possibilities demonstrate that the ALADIN system is not only a recently efficiently used modeling system, but a very perspective and powerful tool for future extensions for meso-gamma scale numerical weather prediction. This recognition led to the decision of the HIRLAM (High Resolution Limited Area Model) project for joining the efforts of the ALADIN cooperation for the further development of such numerical weather prediction models, which are capable to be used efficiently for providing high resolution, reliable, weather forecasts for ultra-short and short time ranges.

Acknowledgements—This work was partially supported by the Hungarian National Office for Research and Technology (JÁP, grant No. 2/007/2005). The authors are very grateful to the international ALADIN team for the development of the ARPEGE/ALADIN model family. The support and help of the other members of the NWP team of the Hungarian Meteorological Service is also highly appreciated. Special thanks are devoted to *Edit Hágel, Eszter Lábó, Roland Steib*, and *Steluta Vasiliu* for providing some of the figures for the article. The useful and very constructive comments of *Radmila Brožkova, Jean-Francois Geleyn*, and *Dezső Dévényi* are also very much appreciated.

References

Bechtold, P., Bazile E., Guichard F., Mascart P., and *Richard E.*, 2001: A mass flux convection scheme for regional and global models. *Q. J. Roy. Meteor. Soc. 127*, 869-886.

Bénard, P., 2003: Stability of semi-implicit and iterative centered-implicit time discretizations for various equations systems used in NWP. *Mon. Weather Rev. 131*, 2479-2491.

Bénard, P., 2004: On the use of a wider class of linear systems for the design of constant-coefficients semi-implicit time schemes in NWP. *Mon. Weather Rev. 132*, 1319-1324.

Bénard P., Laprise, R., Vivoda, J., and *Smoliková, P.*, 2004: Stability of leap-frog constant-coefficients semi-implicit schemes for the fully elastic system of euler equations. Flat-terrain case. *Mon. Weather Rev. 132*, 1306-1318.

Bénard, P., Mašek, J., and *Smoliková, P.*, 2005: Stability of leapfrog constant-coefficients semi-implicit schemes for the fully elastic system of Euler equations: case with orography. *Mon. Weather Rev. 133*, 1065-1075.

Bougeault, Ph., 1985: A simple parameterization of the large-scale effects of cumulus convection. *Mon. Weather Rev. 113*, 2108-2121.

Bougeault, Ph. and *Lacarrere, P.*, 1989: Parameterization of orography induced turbulence in a meso-beta scale model. *Mon. Weather Rev. 117*, 1872-1890.

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

Bouttier, F., Stein, J., and *Bénard, P.*, 2003: Kilometric-scale non-hydrostatic dynamical core studies for the AROME project. *ALADIN Newsletter*, 23, 169-177.

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

Brožková, R., Derková, M., Belluš, M., and *Farda, A.*, 2006: Atmospheric forcing by ALADIN/MFSTEP and MFSTEP oriented tunings (to appear in *Ocean Sci. Discuss*).

Bubnová, R., Hello, G., Bénard, P., and *Geleyn, J.-F.*, 1995: Integration of fully-elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/ALADIN NWP system. *Mon. Weather Rev. 123*, 515-535.

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

Catry, B., 2006: Effects of moisture and mountains in Numerical Weather Prediction. *PhD Thesis*. University of Gent, Belgium.

Cuxart, J., Bougeault, Ph., and *Redelsperger, J.-L.*, 2000: A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q. J. Roy. Meteor. Soc. 126*, 1-30.

Douville, H., Royer, J.-F., and *Mahfouf, J.-F.*, 1995: A new snow parameterization for the Météo-France climate model.Part II: Validation in a 3D GCM experiment. *Clim. Dynam. 12*, 37-52.

Ducrocq, V. and *Bougeault, Ph.*, 1995: Simulations of an observed squall line with a meso-beta scale hydrostatic model. *Weather Forecast. 10*, 380-399.

Geleyn, J.-F., and *Hollingsworth, A.*, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitr. Phys. Atmosph. 52*, 1-16

Geleyn, J.-F., 1987: Use of a modified Richardson number for parameterising the effect of shallow convection. *J. Met. Soc. Japan, Special 1986 NWP Symposium Issue*, 141-149.

Geleyn, J.-F., Fournier, R., Hello, G., and *Pristov, N.*, 2005: A new 'bracketting' technique for a flexible and economical computation of thermal radiative fluxes, scattering effects included, on the basis of the Net Exchanged Rate (NER) formalism. *Research Activities in Atmospheric and Ocean Modelling* (WGNE Blue Book), 2005 Issue, 4/7-8.

Gerard, L. and *Geleyn, J.-F.*, 2005: Evolution of a subgrid deep convection parameterization in a limited-area model with increasing resolution. *Q. J. Roy. Meteor. Soc. 131*, 2293-2312.

Horányi, A., Ihász, I., and *Radnóti, G.*, 1996: A numerical weather prediction model for Central-Europe with the participation of the Hungarian Meteorological Service. *Időjárás 100*, 277-301.

Janiskova, M., Thépaut, J.-N., and *Geleyn, J.-F.*, 1999: Simplified and regular physical parametrisations for incremental four-dimensional variational assimilation. *Mon. Weather Rev. 127*, 26-45.

Kessler, E., 1969: On distribution and continuity of water substance in atmospheric circulations. *Meteorological Monographs 10*, No. 32, 20 pp. American Meteorological Society, Boston.

Lafore, J.-L., Stein, J., Asencio, N., Bougeault, Ph., Ducrocq, V., Fischer, C., Héreil, P., Redelsperger, J.-L., Richard, E., and *Vil -Guerau de Arellano, J.*, 1998: The Méso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations. *Ann. Geophys. 16*, 90-109.

Laprise, R., 1992: The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Weather Rev. 120*, 197-207.

Louis, J.-F., Tiedke, M., and *Geleyn, J.-F.*, 1982: A short history of PBL parameterization at ECMWF. *Proc. ECMWF Workshop on Planetary Boundary Layer Parametrization*, 25-27 November, 1982, ECMWF.

Lynch, P. and *Huang, X.Y.*, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Weather Rev. 120*, 1019–1034.

Masson, V., 2000: A physically-based scheme for the urban energy budget in atmospheric models. *Bound.-Lay. Meteorol. 94*, 357-397.

Mlawer, E.J., Taubman, S. J., Brown, P.D., Iacono, M.J., and *Clough, S.A.*, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res. 102D*, 16663-16682.

Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the ECMWF forecasting system. *J. Geophys. Res. 96*, 9121-9132.

Noilhan, J. and *Planton, S.*, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev. 117*, 536-549.

Pinty, J.-P. and *Jabouille, P.*, 1998: A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. *Preprints of Conf. on Cloud Physics*. Everett, WA, USA. Amer. Meteor. Soc., 217–220.

Radnóti, G., 1995: Comments on "A spectral limited-area formulation with time-dependent boundary conditions applied to the shallow-water equations". *Mon. Weather Rev. 123*, 3122-3123.

Ritter, B. and *Geleyn, J.-F.*, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Weather Rev. 120*, 303-325.

Simmons, A. and *Burridge, D.*, 1981: An energy and angular momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Weather Rev. 109*, 2003–2012.

Simmons, A.J. and *Gibson, J.K.*, 2000: The ERA-40 Project Plan. ERA-40 Project Report Series, 1, 60 pp. Available at ECMWF.

Soci, C., Horányi, A., and *Fischer, C.*, 2006: Sensitivity of high resolution forecasts using the adjoint technique at the 10 km scale. *Mon. Weather Rev. 134*, 772-790.

Xu, K.-M. and *Randall, D.A.*, 1996: A semi-empirical cloudiness parameterization for use in climate models. *J. Atmos. Sci. 53*, 3084-3102.

Žagar, M. and *Rakovec, J.*, 1999: Small-scale surface wind prediction using dynamic adaptation. *Tellus 51A*, 489–504.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 229–252

The development of a limited area ensemble prediction system at the Hungarian Meteorological Service: sensitivity experiments using global singular vectors, preliminary results

Edit Hágel[*] and **András Horányi**

Hungarian Meteorological Service
P.O. Box 38, H-1525 Budapest, Hungary
E-mails: hagel.e@met.hu, horanyi.a@met.hu

(Manuscript received in final form June 20, 2006)

Abstract—In this paper the results of the sensitivity experiments concerning the impact of using different target domains and target times during the global singular vector computation are presented. The system used is made up of 10+1 member ensembles generated with the global model ARPEGE and downscaled with the limited area model ALADIN. The target domain and target time dependency is studied by using 5 different target domains and 2 different target times. Verification shows that the proper choice of the singular vector target domain and target time can increase the spread and (on average) improves the skill of the ensemble for the Central European area. On the other hand, the studied limited area ensemble system was found not to provide significant additional information with respect to the global one, therefore, the computation of mesoscale initial perturbations for the limited area model might be desirable for a more efficient short-range ensemble system.

Key-words: predictability, ensemble forecast, singular vector target domain and time, limited area model

1. Introduction

Numerical weather prediction (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. To improve the quality of the initial

---

[*] Corresponding author

conditions, different techniques are being developed in the field of data assimilation. However, one has to keep in mind that the fully exact description of the initial conditions is not possible due to errors in the observations and data assimilation techniques (the number and spatial distribution of the observations are not satisfactory, the applied background information contains errors coming from the model formulation, the related uncertainties of the observations and the background fields are simulated with some assumptions, etc.). The ensemble technique offers a proper treatment of this problem. The basic idea of the ensemble method 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. Therefore, the advantage of this method is clear: it provides useful information on the predictability of the atmospheric state and also on the probability of the occurrence of different weather events. Despite its obvious benefits, ensemble technique was used only in the medium range for a long time.

For medium range different ensemble prediction techniques are used at ECMWF (European Centre for Medium-Range Weather Forecasts) and NCEP (National Centers for Environmental Prediction) (the first two weather centers where ensemble forecasting started almost 15 years ago). At ECMWF the singular vector (SV) method is used (*Buizza et al.,* 1993) for their ensemble system. Initial condition perturbations are created as linear combinations of the leading singular vectors. This is done so that a given perturbation covers as much of the Northern and Southern Hemisphere as possible. The amplitude of the perturbations is defined after a comparison with the statistics of analysis error (*White,* 2003).

At NCEP a method called breeding was developed that attempts to create realistic perturbations, which could represent the errors actually present in the analysis cycle (*Toth* and *Kalnay,* 1997). As a first step, initial conditions are randomly perturbed; then a (6 hours) forecast is performed from these perturbed initial conditions. At the end of the forecast, the perturbations are re-scaled, the actual analysis is modified accordingly, and the process continues iteratively. After a few days this method leads to the selection (breeding) of the fastest growing perturbations, which are used for perturbing the initial conditions for the ensemble prediction system.

In the last couple of years intensive research has started to develop short-range global and limited area ensemble prediction systems (LAMEPS) for the mesoscale. Most of the studies show the benefits of limited area ensemble forecasting, but it is not yet clear, which is the best method for the short-range mesoscale application. Hereafter some relevant experiments are briefly presented.

NCEP was not only among the first two weather centers where operational global ensemble forecasting started, but the first real-time, operational regional ensemble prediction system was also implemented there in 2001. Presently there are 21 members in NCEP SREF (Short-Range Ensemble Forecasting). The system is running twice a day to 87 hours, with a horizontal resolution varying from 32 to 45 km. From the very beginning of its development, the system has emphasized both initial condition and model uncertainties by using the multi-analysis, multi-LBCs, multi-model, multi-physics, and perturbed initial conditions approaches. To generate the perturbed initial conditions the breeding method is used, just like in the case of NCEP global ensemble system (*Du et al.,* 2006).

Intensive work has started in Europe as well in the field of regional ensemble forecasting. Frogner and Iversen developed a limited area ensemble system, where the initial and lateral boundary conditions are provided by TEPS (Targeted Ensemble Prediction System). TEPS is the same as ECMWF EPS except for two features. On the one hand, in TEPS, the generation of initial conditions is based on targeted singular vectors, i.e., they are computed over a particular area (Northern Europe), on the other hand, the number of ensemble members is only 20+1 in contrast with the 50+1 members of the original EPS (*Frogner* and *Iversen*, 2001, 2002).

Hersbach et al. examined a similar system designed for the European area. Again, the perturbations of the initial conditions were based on targeted singular vectors (*Hersbach et al.*, 2000).

Based on targeted singular vectors, a global, short-range ensemble system (called PEARP) was developed at Météo-France. The singular vectors are computed over a specific area covering Europe and some part of the Atlantic Ocean. Doing so, perturbations are expected to be efficient in the area of interest (i.e., Western Europe, particularly France).

At ARPA-SIM (Italy) a slightly different approach was chosen. Instead of downscaling all members of ECMWF EPS, only a small subset of the global ensemble members is used as initial and lateral boundary conditions. A clustering is applied to the global ensemble system to select 10 representative members (RM). These RMs would provide the initial and lateral boundary conditions for the limited area integrations performed with the non-hydrostatic Lokal Model. The system runs operationally under the auspices of the COSMO (COnsortium for Small-scale MOdeling) project (*Marsigli et al.*, 2005).

Motivated by these results research started in this field at the Hungarian Meteorological Service (HMS), too. It was decided to start with the direct downscaling of global ensemble members of the PEARP system. For the experiments the ARPEGE/ALADIN (ARPEGE: Action Recherche Petite Echelle Grande Echelle, ALADIN: Aire Limitée Adaptation dynamique Developpement InterNational) model system was used, which is being developed in a wide international cooperation (*Horányi et al.*, 1996).

The so-called PEARP (formerly PEACE) system was used to provide initial and lateral boundary conditions for the limited area experiments. PEARP is an ARPEGE based global short-range ensemble system, which consists of 10+1 ensemble members. It runs operationally at Météo-France. In PEARP, targeted singular vectors are used to generate the initial perturbations. As a linear combination of the first 16 SVs, five orthogonal perturbations are built. These perturbations are added to and subtracted from the unperturbed analysis resulting 10 perturbed initial conditions. From these (10+1) initial conditions the model is integrated up to 60 hours.

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

- Are the initial and lateral boundary conditions directly provided by PEARP convenient for a Central European LAMEPS application?
- What is the optimal configuration for LAMEPS over Central Europe?

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

2. *Methodology*

2.1 The applied models

For the experiments the ARPEGE/ALADIN modeling system was used. Almost 20 years ago a collaboration between ECMWF and Météo-France started in order to develop efficient global data assimilation tools (variational assimilation) and also to have numerical weather prediction models at every spatial and time scales. These goals were achieved by the development of a model family called IFS (Integrated Forecast System) at ECMWF and ARPEGE at Météo-France. The models are spectral, hydrostatic global models. One important feature of the French version of the model is the possibility to use variable spatial resolution (*Schmidt,* 1977). This means that the resolution changes over the whole globe according to a stretching factor. Therefore, it is possible to have high resolution over a certain area of interest (e.g., Europe) and low resolution over the opposite side of the globe (*Fig. 1*).

Fig. 1. The stretched grid of the ARPEGE model. It has high resolution over Europe and low resolution over the opposite side of the globe.

In 1990 a cooperation (with French leadership) started between several weather services (mainly from Central and Eastern Europe) in order to develop a limited area version of the ARPEGE/IFS model system for dynamical adaptation at the limit of the hydrostatic approximation. The created system ALADIN is a spectral, hydrostatic, limited area model.

In the experiments described below, the singular vector computations and the global integrations were performed with the ARPEGE model, while the limited area experiments were carried out with the ALADIN model.

The global ARPEGE ensemble system set up for the experiments was based on the PEARP system. The main difference was in the choice of target domain and target time used for the global singular vector computations. Hereafter the main parts of this ARPEGE ensemble system will be briefly presented together with the basic characteristics of the (global) singular vector computations. Targeted singular vectors were applied to generate initial perturbations. SVs were computed on a resolution coarser than that of the integration (i.e., at the present PEARP system the spectral truncation for the SV computation is T95, while the truncation for the integration is T358). During the computation of the SVs only very simple physics was used, and the total energy norm was applied both at initial and final times. As a linear combination of the first 16 SVs, five orthogonal perturbations were built. These perturbations were added to and subtracted from the unperturbed analysis resulting in 10 perturbed initial conditions (+ the unperturbed control one). Regarding the global integrations, at the start of the experiments (for the case studies and the summer period of 10 days) T199 truncation was used with a stretching factor of 3.5. Later (for the winter period of 32 days) it was changed to T358 (as it is used in the present PEARP system) with a stretching factor of 2.4. In both cases the highest resolution (which is over France) is approximately 25 km, but in the latter case the computations are more accurate, because truncation is performed at a much higher wave number. In the vertical 41 model levels were used.

For the limited area experiments the ALADIN model (*Horányi et al.*, 1996) was used on 12 km horizontal resolution with 37 vertical levels. The integration domain is shown in *Fig. 2*. The initial and lateral boundary conditions were provided by the global ensemble system described above.

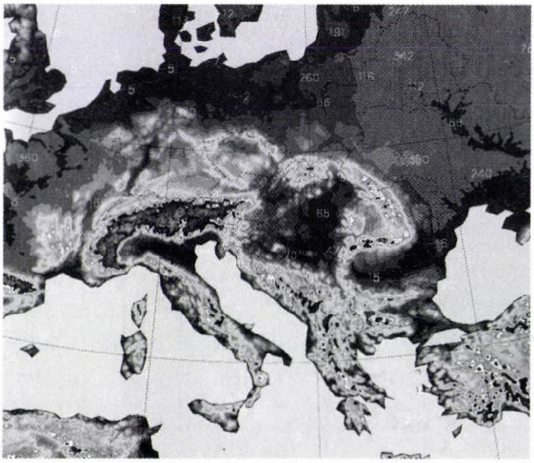


Fig. 2. The integration domain and orography of the ALADIN model.

2.2 *Description of the experiments*

As mentioned before, LAMEPS activities started at HMS with the direct downscaling of global ensemble forecasts. Motivated by some earlier results in the field of short-range limited area ensemble forecasting (*Frogner* and *Iversen*, 2001, 2002; *Hersbach et al.*, 2000), it was decided to investigate the sensitivity of the global singular vector computation in terms of target domain and target time with the main goal to find an optimal configuration for a Central European application. For the experiments the ARPEGE/ALADIN model system was used considering the ARPEGE based global ensemble system, PEARP as a starting point. On the one hand, direct downscaling of the PEARP system was performed, and on the other hand, an ARPEGE based ensemble system was built with a slight modification of PEARP. In the latter case the only difference with respect to the PEARP settings was in the choice of target domain and target time used for the global SV computation. Downscaling of the global ensemble forecasts was realized with the ALADIN model. First, case studies were investigated for significantly different meteorological situations in order to see whether the change of the target domain and target time for the global singular vector computations can have a significant effect on the quality of the forecasts valid for the Central European area. Target domains were chosen with different size and location as follows (*Fig. 3*):

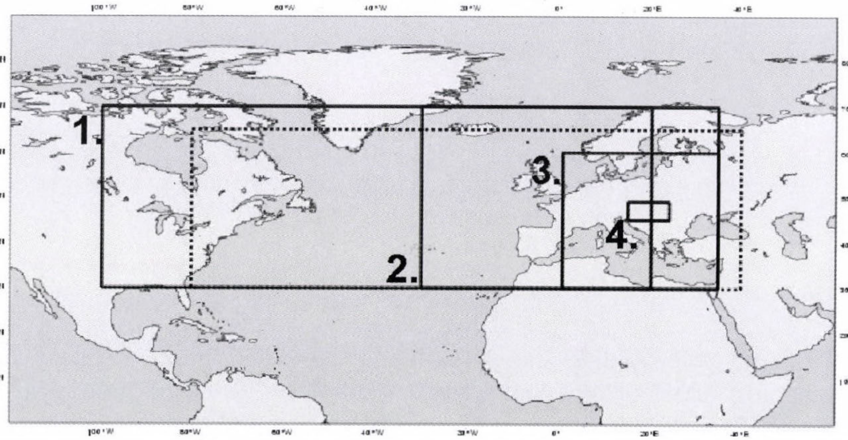


Fig. 3. The location of the four different target domains used for the experiments and the target domain used in the present PEARP system (dotted line).

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

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

Based on the results of the case studies, further experiments were carried out for a 10 days summer period. Then the following four configurations were examined in detail:

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

Based on the result of the 10 days summer period and inspired by the fact that in between important changes took place in the PEARP system, it was decided to examine the following two configurations for an additional 32 days winter period:

- target domain and target time as used in the present PEARP system (dotted rectangle in *Fig. 3* as target domain and 12 hours as target time),
- target domain 2 and target time 24 hours.

2.3 Verification methods

Results of the case studies and experiments covering longer periods were examined in detail. Both subjective and objective verifications were performed. For subjective verification, the ensemble members were visualized in the form of probability maps, "stamp" and "plume" diagrams. For the objective verification, different scores were computed and several types of diagrams were derived (*Toth et al.*, 2003; *Persson* and *Grazzini*, 2005). The objective verification was performed against SYNOP (surface) and TEMP (upper air) data. Additionally for the winter period, verification was also carried out with respect to the ECMWF 4d-var analysis. The verification area was the entire integration domain of the ALADIN model (*Fig. 2*).

2.3.1 Ensemble mean vs. control

When testing an ensemble system, a minimum a priori expectation is that the ensemble mean performs better than the control forecast (the one started from the unperturbed initial condition). If this condition is not fulfilled, then the ensemble system has a very limited value. Therefore, it is of great importance to compare the skill of the ensemble mean and the control run (e.g., to compute and inter-compare the bias and the root mean square error).

2.3.2 Spread vs. error

Another important characteristic of an ensemble system is its spread. It is expected to have similar magnitude as the error. On the one hand, small spread should represent large predictability, consequently small error. On the other hand, the growth of the error and spread are expected to be in line. As an example, the spread of the ensemble can be compared to the root mean square error - RMSE - of the ensemble mean.

2.3.3 Talagrand diagrams, percentage of outliers

In a perfect system, the verifying analysis/observation 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. This can be transformed into a diagram called Talagrand diagram. If the system is perfect, the distribution is flat. Different shapes indicate different behavior: U shape shows that the spread in the ensemble is not sufficient, the verifying analysis/observation lies outside the ensemble too often. Also under- and overestimation (J and L shape) can be determined by examining such a diagram.

The percentage of outliers can be visualized for the whole forecast range. In this case the sum of the two outermost intervals of the Talagrand diagram is on the y-axis, and the time is on the x-axis. In ideal case, the percentage of outliers is equal to $2\times1/(\text{number of ensemble members}+1)$, which is around 0.2 in the case of a 10-member ensemble system.

2.3.4 ROC diagram

From this kind of diagram one can get information about the skill of the forecast. Hit rates and false alarm rates are calculated for different probability thresholds and entered into a ROC (Relative Operating Characteristics) diagram with hit rate on the y-axis and false alarm rate on the x-axis. A perfect system would have all its points in the upper left corner with hit rate equals 1 and false alarm rate equals 0. The integral area under the ROC curve can be calculated to represent the skill of the forecast. An integral area of 1 represents perfect forecast, while an integral area less than 0.5 means that the forecast has no skill compared to the use of climatological statistics.

2.3.5 Reliability diagram

Another useful verification characteristic is the reliability diagram. On average, when an event is forecasted with a given probability, it should occur with the same frequency in the reality. On the reliability diagram, the forecast probabilities are displayed along the x-axis and the observed frequencies for each forecast probability are on the y-axis. If the forecasted probabilities and the observed frequencies agree, the curve lies along the diagonal. Also under- and overestimation can be easily read from the diagram.

2.3.6 Comparison of global and limited area ensemble systems

When making (ensemble) forecasts with a limited area model, it is always important to know whether the limited area model or the global one performs better over the verification area. In other words, it should be assessed what is the added value of the limited area ensemble system with respect to its global counterpart. Therefore, during the objective verification both the ARPEGE and ALADIN models were verified and their results were compared.

3. Results

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

3.1 Case studies

It was expected that the optimal setting of the two parameters (target domain and time) would depend on the meteorological situation, therefore, it was anticipated that a compromise solution ought to be sought to find the best overall choice.

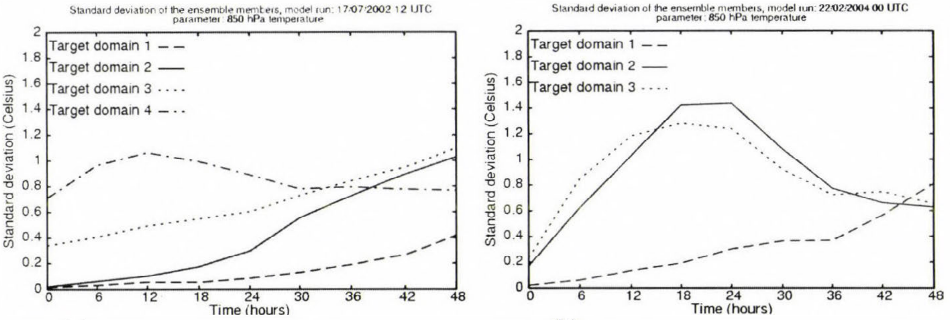
To understand this consideration in detail, significantly different meteorological situations were selected, such as a convective event during summer, a fast moving cold front, a temperature overestimation case during winter, and a Mediterranean cyclone arriving to the Carpathian Basin from the southern direction. First, the focus was put on the question of target domain, i.e., to restrict the possible choices as far as the domain size and geographical location is concerned. The standard deviation of the ensemble members over Hungary was computed 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 complemented by subjective verification.

While using the largest singular vector target domain (domain 1), the average standard deviation (for all examined parameters) over Hungary remained rather small during the entire forecast range (*Fig. 4*). The same quantity for target domain 2 was more considerable for all parameters. Regarding target domain, 3 no significant difference was identified compared to domain 2, moreover, domain 2 proved to be a bit more suitable in the examined cases. For the smallest singular vector target domain (domain 4) 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 target domain is rather small, therefore, a significant part of the initial perturbations propagated out from the target area after a short period of time).

Having a look on the subjective verification (e.g., for 2-meter temperature and precipitation), such a clear conclusion cannot be drawn. As expected a priori, in different meteorological situations different singular vector target domains proved to be the best choice in order to obtain the best ensemble forecasts. In some cases reducing the size of the target domain could increase the spread without improving the quality of the forecasts. On the contrary, there were cases (for instance some local convective events) when using a smaller singular vector target domain, the forecast became significantly better.

As a general conclusion one can say, that the smaller the target domain the bigger the spread (not globally, just over the area where the SVs were optimized!), but with the use of very small target domains a significant part of the initial perturbations would propagate out from the area of interest, and the spread would decrease with time which is something we would like to avoid.

Therefore, among the examined settings, target domain 2 proved to be the optimal choice so this target domain was chosen as the subject of further examinations. Target domain 1 (as used in the PEARP system at that time) was considered as reference.



(a) **(b)**

Fig. 4. Standard deviation diagrams for the ARPEGE ensemble system for two different model runs for 850 hPa temperature. Standard deviation was computed over Hungary. (a) Model run: July 17, 2002, 12:00 UTC. Dashed line is standard deviation with the use of target domain 1, solid line is standard deviation with the use of target domain 2, dotted line is standard deviation with the use of target domain 3, chained line is standard deviation with the use of target domain 4. (b) Model run: February 22, 2004, 00:00 UTC. Dashed line is standard deviation with the use of target domain 1, solid line is standard deviation with the use of target domain 2, dotted line is standard deviation with the use of target domain 3.

3.2 Experiments for a summer period of 10 days

In order to confirm the preliminary conclusions drawn from the case studies for a more precise quantification of the improvements with the experimental settings and for the better understanding of the role of the target time in the singular vector computations, it was decided to pursue experimentation on a longer period of time. Therefore, we have examined in detail two target domains and two target times: domain 1 and 2, and target time 12 hours and 24 hours, respectively, resulting altogether in four sets of experiments.

The average ensemble spread (for different meteorological parameters) over Hungary was examined, and detailed verification (using Talagrand and percentage of outliers diagrams) was performed as well. The randomly selected period was 10–19 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.

3.2.1 Spread

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

Using target domain 2, the spread (on average) can be increased and even further improvement can be obtained with 24 hours as target time (see *Fig. 5*). On average, this configuration (target domain 2 and target time 24 hours) provides the largest values in terms of standard deviation computed over Hungary. It is important to keep in mind that improving the spread is necessary but not sufficient for obtaining better probabilistic forecasts.

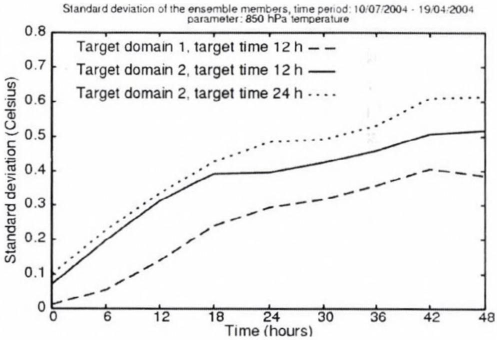
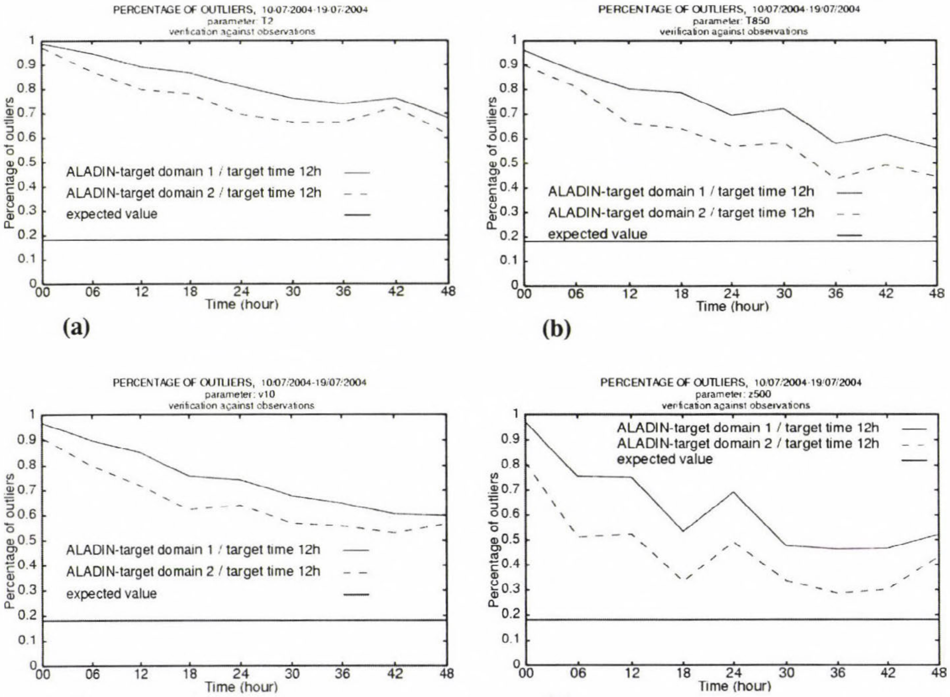


Fig. 5. Standard deviation diagrams for the ARPEGE ensemble system for the period July 10–19, 2004, for 850 hPa temperature. Standard deviation was computed over Hungary. Dashed line is standard deviation with the use of target domain 1 and target time 12 hours, solid line is standard deviation with the use of target domain 2 and target time 12 hours, dotted line is standard deviation with the use of target domain 2 and target time 24 hours.

3.2.2 Talagrand diagrams and percentage of outliers

The best results were obtained when target domain 2 together with target time 24 hours was used for the global singular vector computation. Nevertheless, for surface parameters the two outermost intervals of the Talagrand diagram (not shown) were still dominating, and the percentage of outliers remained much larger than the expected value (which is about 0.2 in our case).

Changing (i.e., reducing the size of) the singular vector target domain yields clear improvements (especially on the higher atmospheric levels) over the verification area in terms of spread (see *Fig. 6*), however, one has to emphasize again that improvement in the spread does not necessarily result in better ensemble forecasts.



(a) **(b)**

(c) **(d)**

Fig. 6. Percentage of outliers diagrams for the ALADIN ensemble system for the period July 10–19, 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 target domain 1 and target time 12 hours for SV computation, dashed line is ALADIN coupled with ARPEGE ensemble members using target domain 2 and target time 12 hours for SV computation. Verification was performed against SYNOP and TEMP observations. The expected value is ~ 0.2 (see the thin horizontal lines).

Unfortunately, ROC and reliability diagrams could not be used for this period due to the poor sampling size leading to non-representative verification results.

3.3 Experiments for a winter period of 32 days

According to the experiments described above, it was concluded that great sensitivity (at least in terms of spread) could be found with respect to the target domain and target time used in the global singular vector computation. It was additionally realized that a period of ten days is not sufficiently long for drawing reliable conclusions and, therefore, larger sample is desirable. However, it could be concluded that the target domain 2 with target time 24 hours seems to be a better choice for a Central European application than target domain 1 complemented with target time 12 hours (as used in the PEARP system at that time).

In addition and simultaneously to these preliminary conclusions, important changes (and operational introduction) had been encountered at Météo-France PEARP system. The following characteristics were changed:

- the resolution used for the SV computation was changed from T63 to T95,
- the target domain became smaller and was shifted towards east,
- the resolution used for the integration was changed from T199c3.5 to T358c2.4.

There were also some changes in the ALADIN model: a new model version became operational at HMS, and it was decided to continue the LAMEPS experiments with this new configuration. Therefore, extended experiments were made for another (longer) period (the choice of this period was again arbitrary) covering 32 days in January and February, 2005. It is important to note that this period was characterized by an unusually cold weather.

Altogether two different configurations were examined: the operational PEARP configuration and target domain 2 together with target time 24 hours to be used for the global SV computations.

For the objective evaluation Talagrand, ROC, and reliability diagrams were drawn, bias and RMSE of the ensemble mean and the control forecast were computed for ARPEGE and ALADIN, respectively.

3.3.1 Ensemble mean vs. control forecast

The first, basic validation of an ensemble system is the comparison of the performance of the ensemble mean and the control forecast (the minimum requirement is that the ensemble mean should provide better results than the control run). For every examined parameter (10-meter wind, 2-meter temperature, 500 hPa geopotential height, 850 hPa temperature) the values of the ensemble mean and control forecast were relatively close to each other with a slight advantage to the ensemble mean (*Fig. 7*). The improvement of the ensemble mean is more pronounced near the surface. All this only means that the ensemble system meets the above mentioned (basic) criterion and further evaluations can be performed.

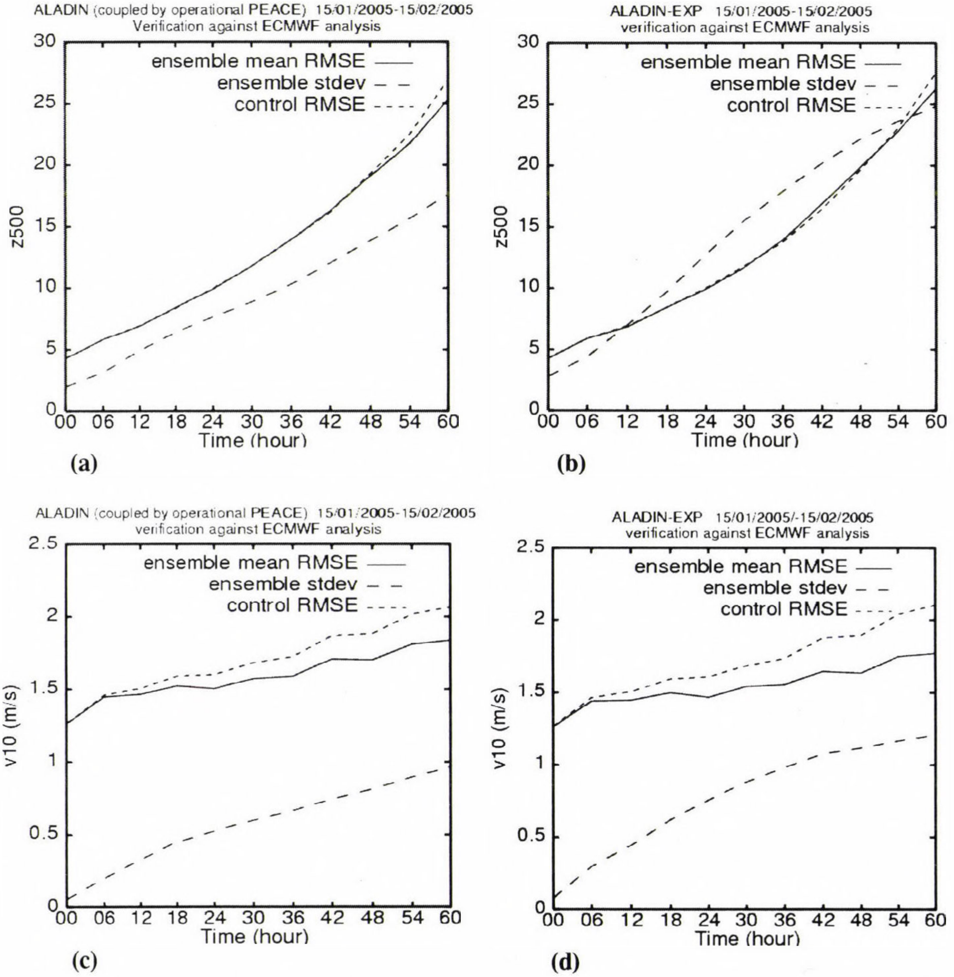


Fig. 7. Comparison of RMSE and standard deviation for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 500 hPa geopotential height, ALADIN coupled with the PEARP members, (b) 500 hPa geopotential height, ALADIN coupled with the experimental set, (c) 10-meter wind speed, ALADIN coupled with the PEARP members, (d) 10-meter wind speed, ALADIN coupled with the experimental set. Solid line is the RMSE of the ensemble mean, dashed line is the standard deviation of the ensemble, and dotted line is the RMSE of the control forecast. Verification was performed against ECMWF analysis.

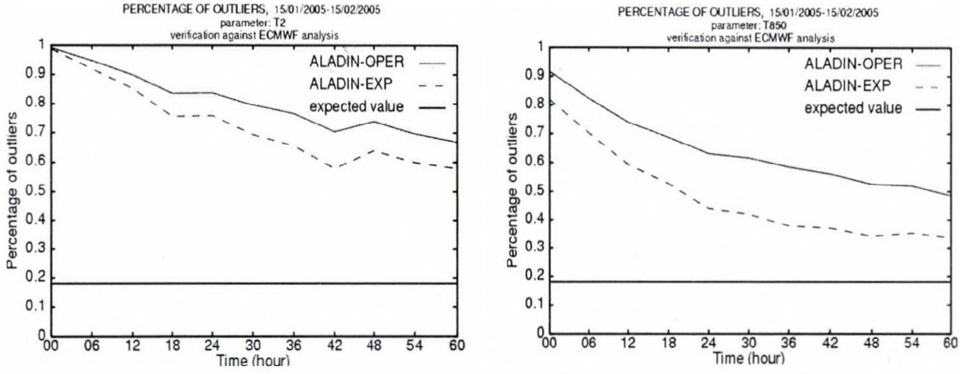
3.3.2 Spread vs. RMSE

It is expected, that the behavior of the ensemble spread and the error is similar (i.e., if the error is small, then the spread should be small as well and vice versa). For the examined parameters it was found that the spread is usually smaller than the error, however, the use of the smaller SV target domain (domain 2) and 24 hours target time reduced the difference between them. Moreover, for 500 hPa geopotential the spread became even larger than the RMSE of the ensemble mean (*Fig. 7*).

It can be concluded that there is a discrepancy between the error and the spread, however, with the correct choice of SV target domain and target time this can be reduced (especially at higher levels).

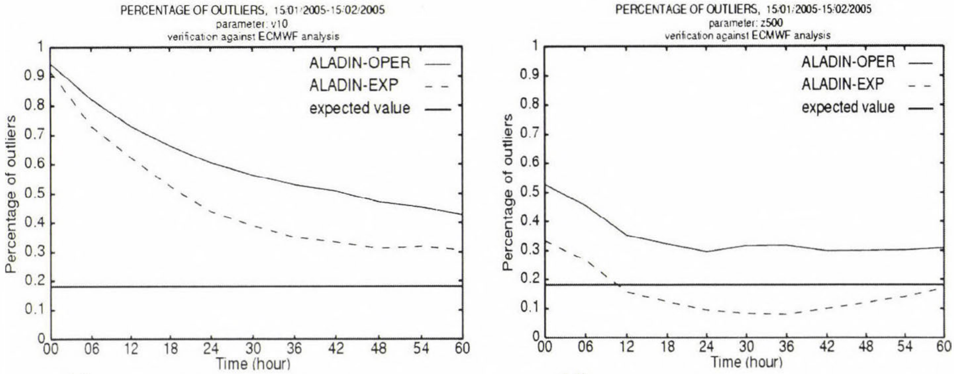
3.3.3 Talagrand diagrams and percentage of outliers

It was found that the change of the target domain and target time during the global SV computation could improve the system's ability to comprise the true state of the atmosphere. For all parameters the Talagrand diagrams became flatter, the distribution moved towards the ideal one (not shown). Looking at the percentage of outliers, clear improvement can be seen, especially for upper level parameters, but also to some extent for the surface ones (see *Fig. 8*). It is also interesting to notice, that on the surface the improvement for the wind speed is more emphasized than that of the temperature. Moreover, the 2-meter temperature is one of the worst parameters in that characteristics (it is expected, that the surface wind is a rather good parameter of the dynamical adaptation due to the fine scale surface description, however, the erroneous behavior of the temperature is a rather puzzling feature).



(a) **(b)**

→→



(c) **(d)**

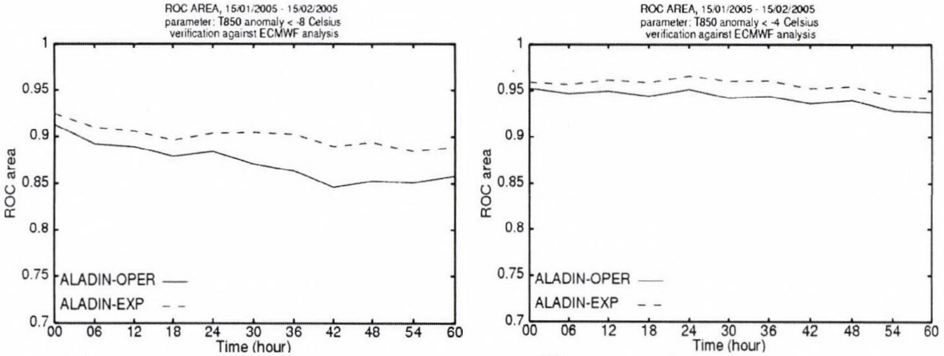
Fig. 8. Percentage of outliers diagrams for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 2-meter temperature, (b) 850 hPa temperature, (c) 10-meter wind speed, (d) 500 hPa geopotential height. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line is ALADIN coupled with the experimental set. Verification was performed against ECMWF analysis. The expected value is ~ 0.2 (see the thin horizontal lines).

3.3.4 ROC area

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

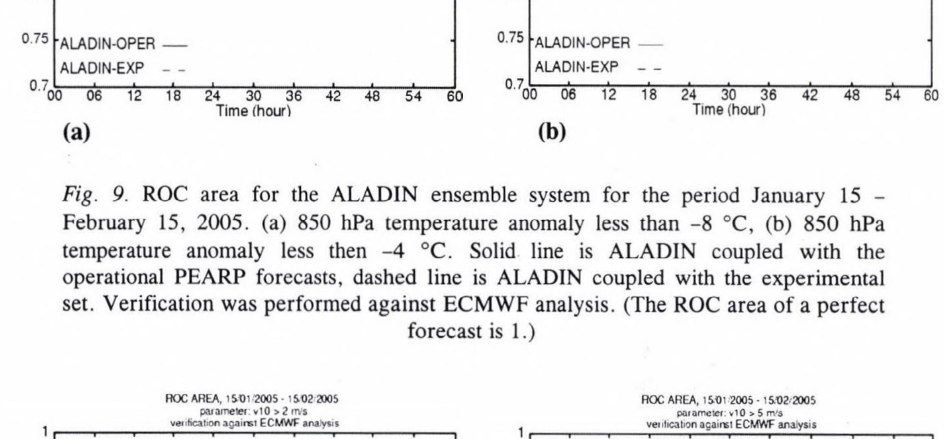
For the 850 hPa temperature anomaly better results were obtained, while using the experimental set (using modified target domain and target time for the global SV computation) of global ensemble forecasts as initial and lateral boundary conditions for the ALADIN model (*Fig. 9*). The ROC area shows rather good scores for the –4 °C threshold (without loss of quality with the integration time), however, the relative improvement is higher for the –8 °C threshold value.

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



(a) **(b)**

Fig. 9. ROC area for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 850 hPa temperature anomaly less than –8 °C, (b) 850 hPa temperature anomaly less then –4 °C. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line is ALADIN coupled with the experimental set. Verification was performed against ECMWF analysis. (The ROC area of a perfect forecast is 1.)



(a) **(b)**

Fig. 10. ROC area for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 10-meter wind speed greater than 2 m/s, (b) 10-meter wind speed greater than 5 m/s. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line is ALADIN coupled with the experimental set. Verification was performed against ECMWF analysis. (The ROC area of a perfect forecast is 1.)

3.3.5 Reliability diagrams

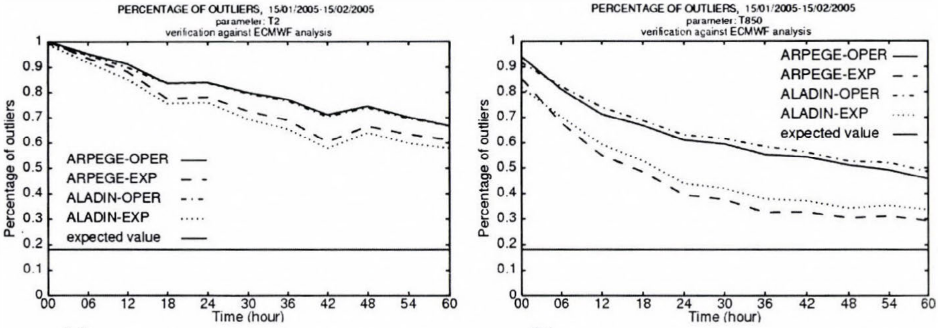
Reliability diagrams were drawn for the same parameters (10-meter wind speed and 850 hPa temperature anomaly) and thresholds as for the ROC diagram. In this case, the use of target domain 2 and target time 24 hours did not result in significantly better forecasts, the diagrams of the two ALADIN configurations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

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

3.4 Comparison of global and limited area ensemble systems

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

Looking at the percentage of outliers one can conclude, that the simple downscaling of the global ensemble system with the ALADIN model does not yield significant improvement. For some parameters the ALADIN forecasts have better scores, for others the ARPEGE gives better results. In *Fig. 11* one can see, that for 2-meter temperature ALADIN coupled with the experimental set performs better, while for 850 hPa temperature the experimental ARPEGE ensemble system has the best results (for any case the differences are small). This result can be explained with the consideration that the higher resolution ALADIN forecasts are gaining advantage near the surface due to the more precise description of surface characteristics and processes.

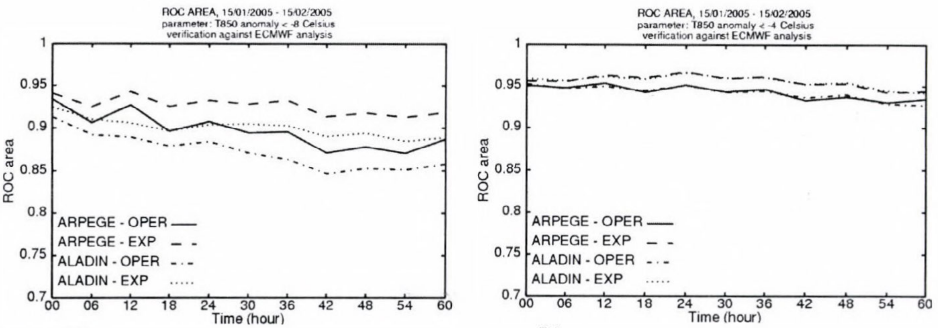


(a) (b)

Fig. 11. Percentage of outliers diagrams for ARPEGE and ALADIN ensemble systems for the period January 15 – February 15, 2005. (a) 2-meter temperature, (b) 850 hPa temperature. Solid line is the operational PEARP forecasts (ARPEGE-OPER), chained line is ALADIN coupled with the operational PEARP members (ALADIN-OPER), dashed line is the experimental ARPEGE ensemble (ARPEGE-EXP), dotted line is the ALADIN model coupled with the experimental set (ALADIN-EXP). Verification was performed against ECMWF analysis. The expected value is ~ 0.2 (see the thin horizontal lines).

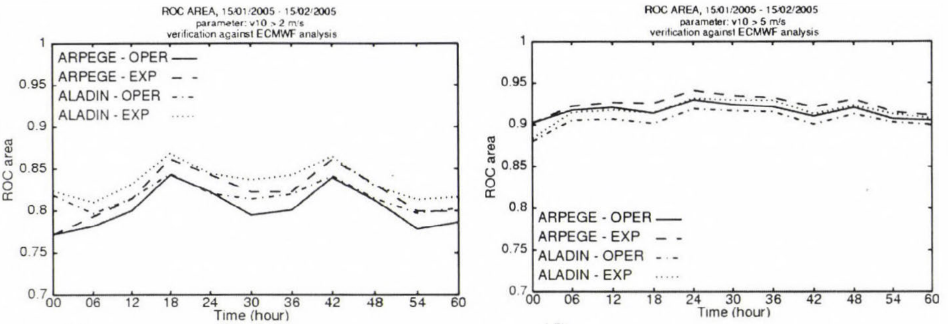
When examining the ROC area diagrams, for both parameters (10-meter wind, 850 hPa temperature) it seems to be hard to tell whether ALADIN or ARPEGE performs better. For certain thresholds and parameters ALADIN had better scores (*Fig. 12c*), for other thresholds ARPEGE was more successful (*Fig. 12a*). There were also combinations (in terms of variables and thresholds), when the two models had nearly the same skill (*Fig. 12b, d*).

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



(a) (b)

→→



(c) **(d)**

Fig. 12. ROC area diagrams for ARPEGE and ALADIN ensemble systems for the period January 15 – February 15, 2005. (a) 850 hPa temperature less than –8 °C, (b) 850 hPa temperature less than –4 °C, (c) 10-meter wind speed greater than 2 m/s, (d) 10-meter wind speed greater than 5 m/s. Solid line is the operational PEARP forecasts (ARPEGE-OPER), chained line is ALADIN coupled with the operational PEARP members (ALADIN-OPER), dashed line is the experimental ARPEGE ensemble (ARPEGE-EXP), dotted line is the ALADIN model coupled with the experimental set (ALADIN-EXP). Verification was performed against ECMWF analysis.

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

4. Summary, conclusions, and future plans

Extended experiments were performed to investigate the sensitivity of global singular vector computations in terms of target domain and target time. Global (ARPEGE) ensemble members were downscaled with the limited area model

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

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

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

Acknowledgements—The authors are very grateful to the international ARPEGE/ALADIN team for the development of the ARPEGE/ALADIN model family. Special thanks goes to *Jean Nicolau* from Météo-France for his help in connection with the PEARP system. The support and help of the other members of the NWP team of the Hungarian Meteorological Service, especially the contribution of *Gabriella Szépszó* and *Gábor Radnóti* is very much appreciated. This work was supported by the Hungarian National Research Fund (OTKA, grant No. T/F 047295) and the Hungarian National Office for Research and Technology (NKFP, grant No. 3A/051/2004 and JÁP, grant No. 2/007/2005).

References

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

Du, J., McQueen, J., DiMego, G., Toth, Z., Jovic, D., Zhou, B., and *Chuang, H.-Y.*, 2006: New Dimension of NCEP Short-Range Ensemble Forecasting (SREF) System: Inclusion of WRF Members. http://www.emc.ncep.noaa.gov/mmb/SREF/WMO06_full.pdf.

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

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

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

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

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

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

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

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

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

White, P.W. (ed.), 2003: IFS Documentation, Part V: The ensemble prediction system (CY25R1), ECMWF (http://www.ecmwf.int/research/ifsdocs/CY25r1/pdf_files/Ensemble.pdf).

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 253–277

The dynamical downscaling of ECMWF EPS products with the ALADIN mesoscale limited area model: preliminary evaluation

Balázs Szintai[1*] and **István Ihász**[2]

[1]*Department of Meteorology, Eötvös Loránd University*
P.O. Box 32, H-1518 Budapest, Hungary
E-mail: szintaib@nimbus.elte.hu

[2]*Hungarian Meteorological Service,*
P.O. Box 38, H-1525 Budapest, Hungary
E-mail: ihasz.i@met.hu

(Manuscript received in final form August 28, 2006)

Abstract—The ECMWF/ALADIN system is a limited area ensemble prediction system, which has been developed at the Hungarian Meteorological Service (HMS). The main objective of this limited area ensemble system was to dynamically downscale the ensemble forecasts of the ECMWF/IFS model with the ALADIN limited area model. For the reduction of the computational cost a cluster analysis is performed on the ECMWF ensemble members, and the representative members of the clusters were chosen for providing initial and lateral boundary conditions for the limited area runs. The downscaling system was tested using four different clustering configurations. Preliminary results were obtained by the investigation of four case studies. The subjective evaluation – using stamp diagrams and probability maps – showed that the downscaling system improved the precipitation forecasts of the global EPS system. Objective verification was performed on the basis of Talagrand and ROC diagrams. The Talagrand diagrams showed that the ensemble spread of the downscaled forecasts is not satisfactory, which is a consequence of the loss of information due to the reduced ensemble population. Investigation of the precipitation ROC diagrams confirmed that the best ECMWF/ALADIN EPS configuration improved the forecasts provided by the original ECMWF EPS.

Key-words: ensemble prediction, limited area model, clustering, Talagrand and ROC diagrams

---

[*] Corresponding author

1. Introduction

The atmosphere can be considered locally as a chaotic system, therefore, it shows great sensitivity with respect to its initial conditions. Consequently, small errors in the initial field can cause large errors in the forecast. To handle the uncertainties in the initial field, the so-called ensemble technique can be used. In the case of ensemble forecasts several equally possible initial fields are determined, and the model is integrated on these different initial conditions several times. This technique yields not a single deterministic forecasts, but several forming an ensemble of forecasts. In this way not only the different possible ways of the future evolution of the atmosphere can be predicted, but a probabilistic information can be attributed to these forecasts as well.

Beside others, ensemble prediction systems can be classified according to the applied integration domain. The global systems provide probabilistic forecasts over the entire globe, while the limited area ones (LAMEPS as Limited Area Ensemble Prediction System) make predictions only on a certain area of interest. For the LAMEPS forecasts (initial and) lateral boundary conditions are mandatory, which are usually provided by global systems. At the European Centre for Medium-Range Weather Forecasts (ECMWF) and National Centers for Environmental Prediction (NCEP), global ensemble prediction systems have been running operationally since 1992 to provide probabilistic medium range weather forecasts. However, in the last few years several limited area ensemble prediction systems have been installed and used for shorter forecast ranges (*Du et al.,* 2006).

COSMO-LEPS is an example of a limited area ensemble prediction system, which has been running every day at the ECMWF since November, 2002. The methodology of this system allows to combine the benefits of the probabilistic approach with the high resolution details of the LAM integrations, with a reasonable computational investment (*Montani et al.,* 2003). The method is based on an algorithm that selects a number of representative members out of a global ensemble system. These members provide initial and lateral boundary conditions for the limited area model. In the case of the COSMO-LEPS project, the representative members are chosen from the ECMWF EPS (Ensemble Prediction System) members, and the downscaling is performed with the Lokal Model. Since June, 2004, the representative members are chosen from the two most recent EPS runs (the so-called super-ensemble), and the number of the representative members is fixed to 10. The COSMO-LEPS system was providing the main inspiration for our work on the downscaling of ECMWF EPS products.

At the HMS, research related to the limited area ensemble prediction began in autumn, 2003. For the downscaling of the global ensemble systems the ALADIN limited area model is used. The first LAMEPS experiments started with the direct downscaling of the PEACE short range global ensemble system (*Hágel* and *Szépszó,* 2004). The PEACE system is based on the French ARPEGE global model (*Courtier et al.,* 1991), which is integrated 10+1 times (10 perturbed forecasts and 1 control). In the case of direct downscaling, each member of the global system provides initial and lateral boundary conditions for the limited area model.

Limited area ensemble experiments using the ensemble forecasts of the ECMWF EPS started recently in Budapest (in spring, 2005). At that time three necessary ingredients were at disposal to starting the work. First, the integration of the ALADIN model became possible by using ECMWF initial and lateral boundary conditions. Second, a new supercomputer provided a solid computational background to the experiments. Finally, a multivariate clustering algorithm of ECMWF EPS members was developed and made ready to use (*Szintai,* 2004).

The present paper describes the ALADIN limited area downscaling system of the ECMWF EPS forecasts. In Section 2, the two main components of the downscaling system are described, namely the clustering method and the ALADIN runs. The case studies are briefly summarized in Section 3 with special emphasis on the subjective evaluation and objective verification. The conclusions are summarized in Section 4.

2. The ECMWF/ALADIN downscaling system

At the ECMWF, the EPS is integrated twice a day from the 00:00 and 12:00 UTC analyses. The singular vector method is used to derive 50 different perturbed initial fields. The global ECMWF model is then integrated 51 times (50 perturbed members and 1 control) starting from these fields. The horizontal resolution of the model (in 2005) is 80 km with 40 vertical levels (*Buizza et al.,* 2001).

The ECMWF/ALADIN system is a limited area ensemble prediction system, which has been developed at the Hungarian Meteorological Service. The objective of this system (analogously to the COSMO-LEPS system) was to downscale the ensemble forecasts of the ECMWF with a limited area model called ALADIN. The main scheme of this downscaling system is the following: to reduce computational cost, cluster analysis is performed on the 50 members of the ECMWF EPS, and 10 clusters are formed. From each cluster a representative member is chosen. These 10 representative members provide initial and lateral boundary conditions for the 10 ALADIN runs. Therefore, the downscaling system has two main parts: the clustering method and the ALADIN runs, which are described hereafter.

2.1 The clustering method

The main goal of clustering is to form such groups from the ensemble members where the members are as similar to each other as possible. A hierarchical method is used to cluster the ECMWF ensemble members. The main characteristics of the clustering are as follows (*Borgatti*, 1994):

- At the beginning, consider all members as separate clusters and calculate the so-called distance matrix.
- Find the closest pair of clusters and merge them into a single cluster.
- Compute the distances between the new cluster and each of the old clusters.
- Repeat steps 2 and 3 until the predefined number of clusters is found (10 in our case).

Based on the experience gained by the COSMO-LEPS system, ten clusters were used (*Marsigli et al.*, 2005). This cluster number was determined as a compromise between the loss of information due to the decrease of the ensemble size and computational cost. The meteorological parameters used for clustering were the geopotential, relative humidity, and the two wind components at three isobaric levels (500, 700, and 850 hPa), which means altogether twelve clustering parameters. These meteorological parameters were chosen, because the clustering method was targeted to be sensible for the rainfall processes. Geopotential could be the key for the identification of different synoptic systems, relative humidity is the indicator of clouds, and the orographic precipitation is highly influenced by wind direction. These fields were the basis for the clustering at two clustering times (+60 and +84 h) for achieving a relatively sophisticated clustering procedure. The clustering method was tested on two clustering domains. The bigger one is the same as the integration domain of the ALADIN model (34°N–55.5°N, 2°E–39°E), the smaller one is used operationally for the synoptic clustering of ECMWF EPS forecasts (*Ihász*, 2003) at the HMS (40°N–55°N, 10°E–30°E) (*Fig. 1*).

The ECMWF/ALADIN ensemble system was tested on four configurations differing from each other only in the clustering, while the ALADIN runs had the same settings. The clustering configurations were the following:

- Clustering on the bigger domain, using one set of ECMWF EPS (50 members).
- Clustering on the smaller domain, using one set of ECMWF EPS (50 members).
- Clustering on the bigger domain, using two sets of ECMWF EPS (100 members).
- Clustering on the smaller domain, using two sets of ECMWF EPS (100 members).

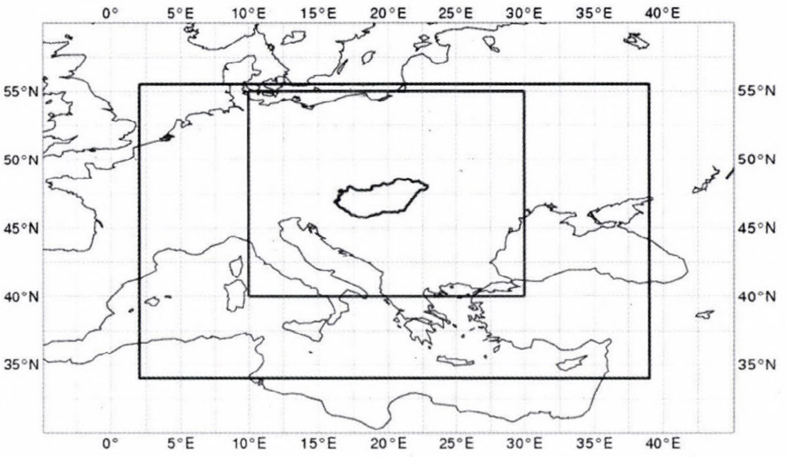


Fig. 1. The two clustering domains for the ECMWF/ALADIN downscaling system. The bigger area is equivalent to the integration domain of the ALADIN model.

While using only one set of ECMWF EPS (50 members), the 12:00 UTC EPS run was used. When using two sets of EPS (100 members), the 00:00 UTC and 12:00 UTC EPS runs of the same day were joined. In this latter case, the initial conditions (IC) of the ALADIN runs were either the IC of the 12:00 UTC EPS, or the +12 h forecast of the 00:00 UTC EPS (depending on the fact, whether the selected representative member is coming from the 00:00 UTC or 12:00 UTC run). The clustering times were +60 h and +84 h for the 12:00 UTC EPS and +72 h and +96 h for the 00:00 UTC EPS (*Fig. 2*).

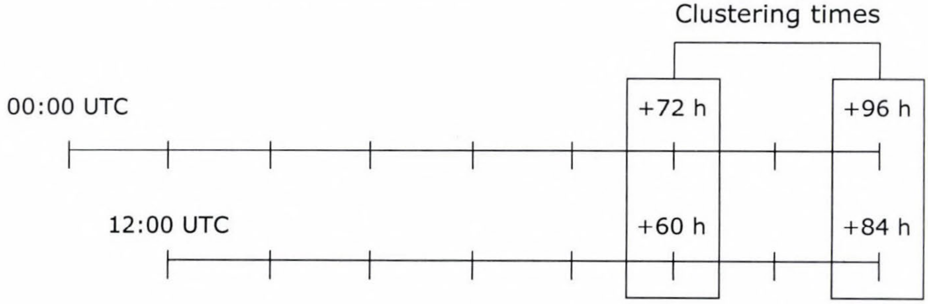


Fig. 2. Schematic description of the clustering time while using 100 ECMWF EPS members.

To compare the 12 clustering parameters, the fields had to be standardized with the help of climatological data. The standardized field (*stfield*) was obtained by subtracting the climatological mean (*cl_mean*) for the given grid point from the original field (*field*), and then dividing this expression with the climatological standard deviation (*cl_dev*):

$$stfield(i) = \frac{field(i) - cl\_mean(i)}{cl\_dev(i)}, \tag{1}$$

where i refers to the grid point.

The clustering method is sensitive to the choice of the distance equations. The average-link method was used, which means that the distance of two clusters is calculated as the distance of the cluster-means (*clustermean*). The distances were first calculated with a square formula (*dist1*), which was weighted with the cosine of the latitude (it is needed due to the application of latitude/longitude coordinates). This expression was then weighted with the population of the clusters in order to obtain clusters with reasonable populations (members):

$$dist1(j,k) = \frac{\sum_{p=1}^{12} \sum_{i=1}^{IMAX} (clustermean(j,i,p - clustermean(k,i,p))^2 \cdot \cos\varphi}{\sum_{i=1}^{IMAX} \cos\varphi}, \tag{2}$$

$$dist2(j,k) = dist1(j,k) \cdot \frac{J \cdot K}{J + K}, \tag{3}$$

where j and k are the clusters which were compared, p refers to the clustering parameters, i refers to the grid point, and φ is the latitude. J and K are the population of the clusters. *IMAX* is the number of grid points for a given field (which was 1271 for the smaller and 3300 for the bigger clustering domain).

As a result of the clustering, ten clusters were formed. From each cluster a representative member (which is the cluster member having the smallest distance from the cluster mean) was chosen.

Due to the fact that the clustering algorithm determines the representative members to be downscaled and thus it has a great impact on the performance of the LAMEPS system, it is worth dealing with clustering in details. The above mentioned distance and population-weighting equations were mainly chosen, because at the ECMWF these formulas are used operationally for the clustering of ECMWF EPS members. The population-weighting formula results in that during the clustering process, the most populated clusters are not likely to be unified (i.e., there are more clusters with average number of elements). Consequently, the first largest 4–5 clusters have roughly the same population (each containing 10–20% of the members), but at the same time, the last 2–3 clusters contain only 2–5% of the members (see *Table 1*). This fact implies that the first well populated clusters represent the most likely future scenarios, which are close to the ensemble mean, while on the other hand, the last clusters may catch some extreme events as well. This feature of the clustering method results in that it could provide a reasonable performance both in average and extreme weather situations.

Table 1. Population of clusters using the two weighting methods for the first case study (May 18, 2005). Results are shown for the clustering configuration using 100 ECMWF EPS members and bigger clustering domain

Cluster	**Percentage of members (%)**	
	Without population-weighting	**With population-weighting**
1	81	17
2	5	16
3	3	15
4	2	13
5	2	11
6	2	8
7	2	7
8	1	6
9	1	5
10	1	2

Another possibility for clustering could be to skip the population-weighting. At the experimentation stage this version was also tested, although the results shown in this paper are based on the population-weighting method. Leaving out population-weighting usually results one very well populated cluster (containing more than 80% of the members), and 7–8 very small clusters, containing only one or two members (*Table 1*). This method could perform better in case of extreme events, since some clusters could represent the extreme members more successfully. However, the better performance of the non-weighting method cannot be always efficiently demonstrated in practice due to the fact that by the computation of probability maps (see hereinafter), the representative members should be weighted by the population of the cluster in order to adequately represent the characteristics of the original super-ensemble. As the extreme clusters contain only 1–2% of the members, their weight by the calculation of the probability maps could be rather small. In the case of the formerly mentioned population-weighting clustering method, this kind of calculation by the probability maps is less important, as the clusters have roughly the same population. During the experiments both probability map computation methods were tested (for the original population-weighting clustering method), but important differences were not detected (for the probability maps shown in this paper the non-weighting method was used).

For any case it should be emphasized, that the ECMWF/ALADIN system is in an experimental stage at the moment. The whole system is designed in a way that its certain components (e.g., clustering method, limited-area runs) could easily be replaced or modified. Therefore, it is also planned that in the future different clustering methods will be tested.

2.2 ALADIN runs

ALADIN is a spectral limited area model, which has been used operationally for short-range weather forecasting at the HMS (*Horányi et al.*, 1996). This model was originally designed to perform a high resolution dynamical adaptation of the French ARPEGE global model. Nevertheless, the ALADIN model can be driven by the ECMWF IFS model as well. For the creation of initial and lateral boundary conditions from the ECMWF/IFS model, a special ARPEGE/ALADIN model configuration was used, which transforms the ECMWF surface and model variables into the ALADIN-required format (it is especially "tricky" for the surface part). The ALADIN model uses a bi-Fourier horizontal spectral representation. The vertical coordinate to be used is the pressure based hybrid coordinate, which is terrain following at the bottom of the model and pressure type at the top of it. The ALADIN model uses an extremely efficient semi-implicit semi-Lagrangian (SISL) time integration scheme.

The applied version of the ALADIN model had 12 km horizontal resolution with 37 levels in the vertical; the time-step used for the integrations was about 500 s. The integration domain of the model was the bigger clustering domain. The forecast range was 84 hours, in order to downscale the medium range information from the ECMWF EPS.

3. Case studies

Because of the high computational cost (one set of forecasts took about 10 hours on the HMS's IBM p690 supercomputer), the downscaling system could not have been verified on a longer time period, therefore, case studies were selected for verification. The main goal of the established short range and early medium range ensemble forecasting system is to improve the forecasts of extreme weather events with the enhancement of the ECMWF EPS products. Therefore, this objective was considered when choosing the relevant cases. Four case studies have been completed so far:

- May 18, 2005: the so-called Slovenian squall line, which caused heavy precipitation and strong wind gusts all over Hungary.
- July 11, 2005: Cyclone over Hungary resulting heavy precipitation.
- August 22, 2005: Mediterranean cyclone causing heavy precipitation.
- November 16, 2005: Mediterranean cyclone south from Hungary; overestimated precipitation by the ECMWF EPS.

All of the cases were related to precipitation events: in the first three cases the ECMWF EPS system underestimated the precipitation, but in the last one the precipitation was rather overestimated. This last case was selected in order to investigate the impact of the ECMWF/ALADIN downscaling, where the goal of the improvement is the decrease of the precipitation amount (contrary to the other cases).

The results were evaluated subjectively and objectively as well. For the subjective evaluation stamp diagrams and probability maps were visualized both for the original ECMWF EPS and the ALADIN downscaling system. For the objective verification two different verification techniques, namely Talagrand and ROC diagrams were used for several meteorological parameters.

3.1 Subjective evaluation

The results obtained for the first three case studies were rather similar, therefore, only one case is shown here, which is the case of May 18, 2005. The last case with the precipitation overestimation will be separately analyzed afterwards.

3.1.1 Heavy precipitation case (May 18, 2005)

On May 18, 2005 a Slovenian squall line (as it is called in the forecaster's vocabulary due to the origin of the system) passed through Hungary and caused heavy precipitation and strong wind gusts all over the country. Because of the favorable conditions for convection, supercells formed both in the western and eastern part of the country. The 24 hours accumulated precipitation amount exceeded 30 mm at several areas. The forecasts studied for this event were initiated on May 16, 12:00 UTC. Only the results of the first clustering configuration are presented (50 members, bigger domain) due to the fact,that the other clustering strategies were providing basically the same results by the subjective judgement.

Having a look on the stamp diagrams (*Fig. 3*) of the ECMWF EPS and ALADIN EPS systems, one can easily see that the ECMWF EPS system significantly underestimated the amount of precipitation and, at the same time, the ALADIN downscaled results indicate higher amounts (however, the underestimation still exists at some ensemble members, mainly in the north-eastern part of the country). One can also easily spot the increasing details of the ALADIN EPS forecasts, which can be certainly attributed to the large resolution difference between the two systems. It is also interesting to see, that not only the fine scale details had been modified by the higher resolution system, but also some of the structures in the precipitation patterns, which indicates that the dynamics and physics of the ALADIN model have also played an important role in the downscaling process. It is also noted that the other clustering configurations (other clustering domain and 100 members super-ensemble) gave very similar results (not shown).

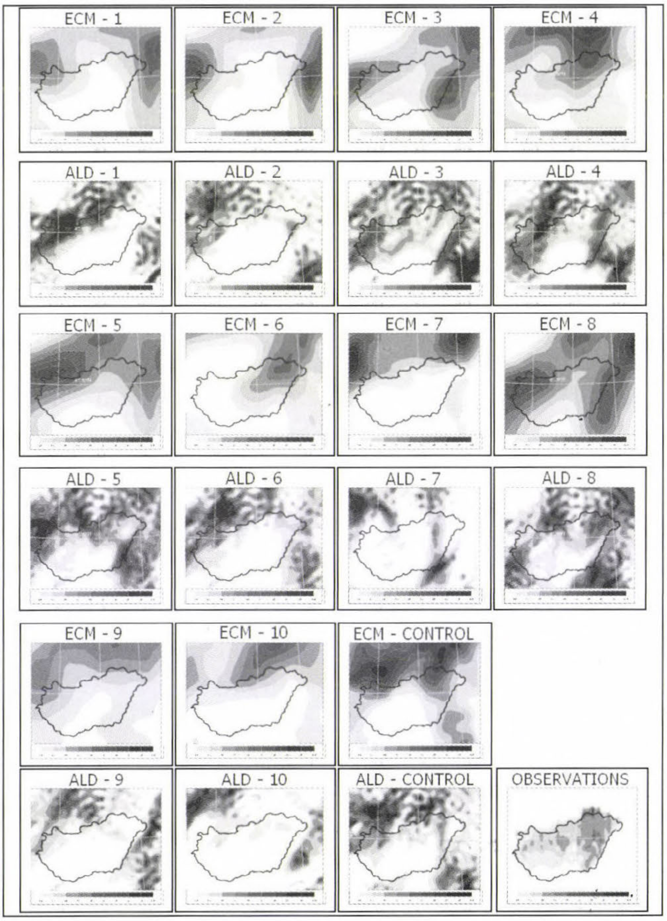


Fig. 3. Stamp diagrams for the ECMWF EPS (10 representative members of the 50 global EPS members, bigger domain clustering configuration), and the ALADIN downscaled members. 24 h precipitation forecasts for May 18, 2005, 06:00 UTC – May 19, 2005, 06:00 UTC. The control forecasts (without initial perturbations) and the observations are also displayed. Darker colours indicate higher precipitation amounts. Values over 10 mm are displayed.

The probability maps show that for a given precipitation threshold, how many percent of the ensemble members forecasted higher amounts than this threshold value. The ECMWF EPS (original 50 members EPS, and the 10 representative members) and ALADIN EPS systems are inter-compared for 10, 20, 30, and 40 mm threshold values, respectively (*Fig. 4*).

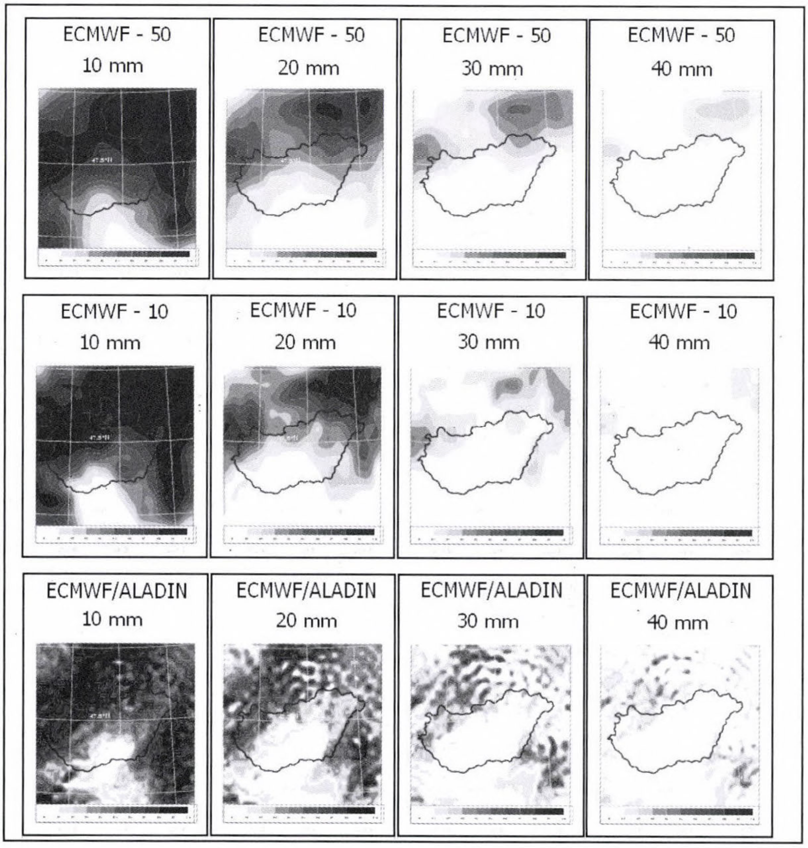


Fig. 4. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24 h precipitation exceeding 10, 20, 30, and 40 mm between May 18, 2005, 06:00 UTC – May 19, 2005, 06:00 UTC is displayed. Darker colours indicate higher percentage values.

The conclusions, drawn from the stamp diagrams and confirmed by the probability maps are as follows: more details in the downscaled system, underestimation of precipitation for the ECMWF EPS system, slight improvement in the ALADIN EPS, some additional features in the high resolution forecasts. Comparing the 50 and 10 members ECMWF EPS models, one can see that the clustered system represents the main characteristics of the original system correctly, which means that the information loss due to clustering was not notable (a reasonable choice is to have 10 representative members instead of the full system). All these results in a way confirm our a priori expectations: the high resolution downscaling system is not only capable to increase the precipitation amount, but able to capture new characteristics of the mesoscale systems, which are encouraging for the further experimentations (case studies). It is also good news, that the loss of information with the clustering and selection of representative members remains on a reasonable level.

3.1.2 Low precipitation case (November 16, 2005)

For this last case study, on the contrary to the first three ones, the precipitation pattern was overestimated by the ECMWF EPS system. This case was selected, because we were wondering, whether the previously noticed increase of precipitation (therefore, further worsening of the forecast) can be also detected for such a case. On November 16, 2005, a Mediterranean cyclone was located south of Hungary, over the Balkan Peninsula. The main precipitation zone of the cyclone was not over Hungary this day, the precipitation quantity was between low and medium amount in the region (the observed 24 h accumulated precipitation did not exceed 15 mm in Hungary, see *Fig. 5*).

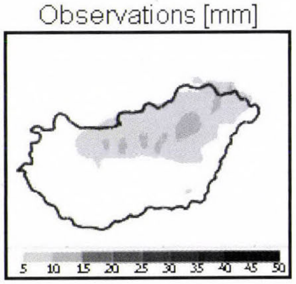

Observations [mm]

Fig. 5. 24 h accumulated precipitation over Hungary between November 16, 2005, 06:00 UTC and November 17, 2005, 06:00 UTC. Values over 5 mm are displayed. Darker colours indicate higher precipitation amounts.

The forecasts studied for this event were initiated on November 14, 12:00 UTC. Only the results of the first clustering configuration are presented (50 members, bigger domain). The stamp diagrams show that in the region of

Hungary the ECMWF EPS overestimated the amount of precipitation, because it did not forecast the location of the main precipitation zone of the cyclone correctly. The downscaling system located the precipitation zone more accurately, consequently the overestimation was not so notable in this case (diagrams not shown).

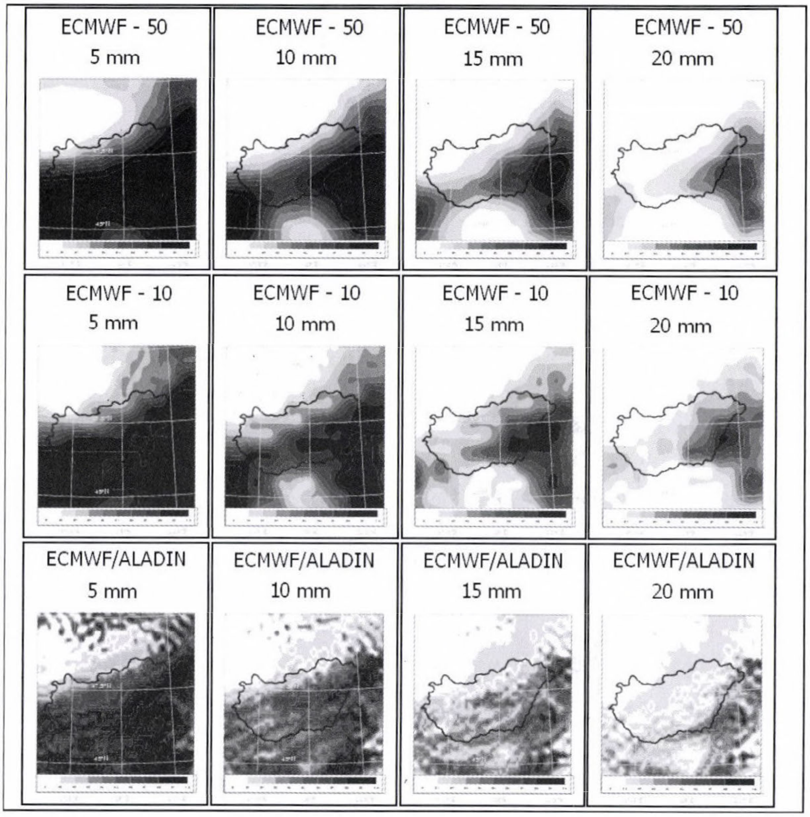


Fig. 6. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24 h precipitation exceeding 5, 10, 15, and 20 mm between November 16, 2005, 06:00 UTC and November 17, 2005, 06:00 UTC is displayed. Darker colours indicate higher percentages.

The probability maps (*Fig. 6*) gave the same results as the stamp diagrams: locally large overestimation for the ECMWF EPS (especially in the south-eastern part of the country), little overestimation and more accurate mesoscale details for the ECMWF/ALADIN. Just like in the first case study, the clustered system represented the main characteristics of the original system correctly. It means that in spite of the resolution increase, the dynamics (and physics) of the ALADIN mode could correct (at least partially) the deficiencies of the global system. However, it has to be added, that mainly the location of the convective systems were forecasted better by the ALADIN model and not the precipitation amount of the main precipitation zone was modified.

To give a summary, on the grounds of the case studies it can be assumed, that the downscaling system is capable to improve the forecasts of the ECMWF EPS both in heavy and low precipitation events in the case of under- or overestimation of the global system.

3.2 Objective verification

The subjective evaluation of the case studies was complemented with objective verifications, more precisely by the computation and analysis of Talagrand diagrams and ROC (area) curves. As it is the case for the subjective judgement, the objective scores should also be evaluated with certain care due to the small number of cases, resulting in a possible non-significant statistical performance.

3.2.1 Talagrand diagrams

The Talagrand diagrams are widely and popularly used ensemble verification characteristics for the representation of the ensemble spread (*Anderson,* 1996; *Talagrand et al.,* 1998) of an ensemble prediction system. The ensemble spread is an important diagnostic feature indicating the efficiency of the initial perturbations used in the ensemble system. Provided that all ensemble members are equally probable (i.e., the ensemble members and the verifying observations are mutually independent realizations of the same probability distribution), each of the $m+1$ intervals, defined by an ordered series of m ensemble members for a given meteorological parameter, including the two open ended intervals, is equally likely to contain the verifying observed value. First, the members for a given grid point are sorted in increasing order, and then a histogram is made by accumulating the number of cases over space and time, when the verifying analysis falls in any of the $m+1$ intervals. Consequently, the flat diagram means good spread, U shape means lack of spread, L shape means overestimation, and J shape means underestimation.

The ECMWF analysis was used to calculate the Talagrand diagrams. The following meteorological parameters were investigated: 500 hPa geopotential height, 850 hPa temperature, 10-meter wind, and 2-meter temperature. It is noted that the Talagrand diagrams had the same shape for all parameters. The diagram of the ECMWF EPS showed the lack of spread on the first day of the forecast range, but the spread was nearly ideal on the second and third days (not shown). This is not surprising due to the fact, that the ECMWF initial perturbations are targeted towards medium range, i.e., the best spread and efficiency of the ensemble system is reached on the second and third day of the forecast range. The same diagram of the ECMWF/ALADIN downscaling system also showed the lack of spread on the first day (even in a bigger extent), however, on the second and third day the spread of ECMWF/ALADIN also got a better shape. It is suspected, that the results are somehow influenced by the fact that ECMWF analyses were used (biased towards the ECMWF system), therefore, alternative solutions ought to be searched having independent observations or analyses (for instance, analyses of the French global model ARPEGE), while creating the Talagrand diagrams.

Precipitation observations were also used for the calculation of Talagrand diagrams. Data of the high-density precipitation observing network of the HMS was applied, which means more than 500 stations covering the entire territory of Hungary (*Ghelly*, 2002). Observed precipitation was cumulated from 06:00 UTC in the morning to 06:00 UTC in the next day. The 24-hour accumulated precipitation observations were averaged on 25 km boxes on Gaussian grid. This means, that there were 179 observation boxes for each day, which means 537 observation boxes for a given time step when investigating the three cases with underestimation. 0.1° × 0.1° post-processing resolution was used both for ECMWF EPS and ECMWF/ALADIN, so four forecast grid points had to be averaged to be consistent to the observations. The Talagrand diagrams of the ECMWF EPS (100 members), ECMWF EPS (10 representative members), and ECMWF/ALADIN (*Fig. 7*) were even more similar to each other than in the case of upper air parameters. Talagrand diagrams were investigated for the 42 h and 66 h forecast ranges (these ranges were chosen in order to find the first and second full day, as far as 24 hour precipitation accumulation is concerned), respectively. The diagrams show the lack of spread at the first day for all the three systems (too many observations are being outside the interval defined by the ensemble members, i.e., relatively large amount of outliers can be identified). Comparing the 100 members ECMWF EPS and the 10 representative members ECMWF EPS, one can see that the percentage of outliers is about twice as much for the 10-member system. The decrease of the spread is a direct consequence of the loss of information due to reduced ensemble size. To attain the spread of a 100-member system with a 10-member system could be possible only with an ideal clustering method, which cannot be applied in practice. The ALADIN runs practically did not affect the spread of the 10-member global EPS, however, in terms of the percentage of outliers, a slight improvement (decrease) can be noticed. On the second day the 100 members ECMWF EPS still shows lack of spread, however, in a much lesser extent (which can be interpreted as underestimation at certain locations and a little overestimation at some other ones). The diagram of the 10 representative members global EPS shows similar structure, however, the percentage of outliers is significantly higher than in the case of the 100-member EPS. This dramatic decrease of the spread can be explained by the fact, that the extreme precipitation fell on the second day of the case studies, and the reduced ensemble system was not very successful in forecasting it (the diagram of the 10-member system mainly shows underestimation). The diagram of the ECMWF/ALADIN shows quite similar behavior to the ECMWF 10-member system.

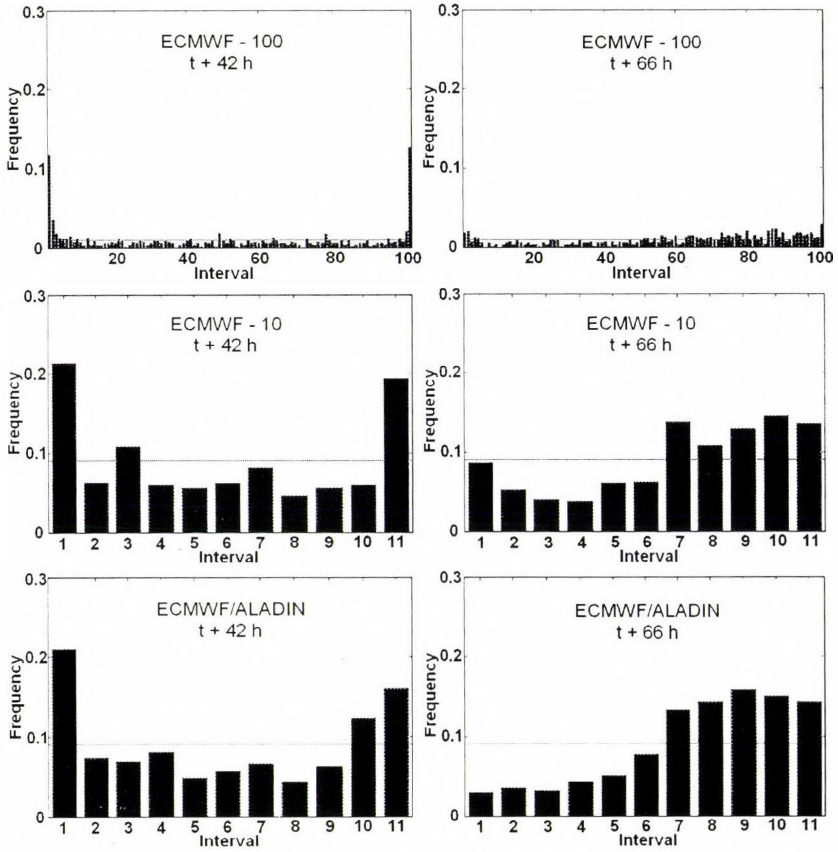


Fig. 7. Talagrand diagrams of 24 h accumulated precipitation for ECMWF EPS (100 members), ECMWF EPS (10 representative members), and ECMWF/ALADIN (10 representative members from 100 global members, bigger domain configuration) for +42 h and +66 h forecasts calculated from the first three case studies.

To examine the time evolution of the ensemble spread and to compare the different clustering configurations, the Talagrand outliers were plotted as well (*Fig. 8*). The value of the Talagrand outliers was obtained by summing up the two extreme values of the Talagrand diagrams. Lower Talagrand outlier values indicate better spread, i.e., less values are outside of the two extreme values of the Talagrand diagram.

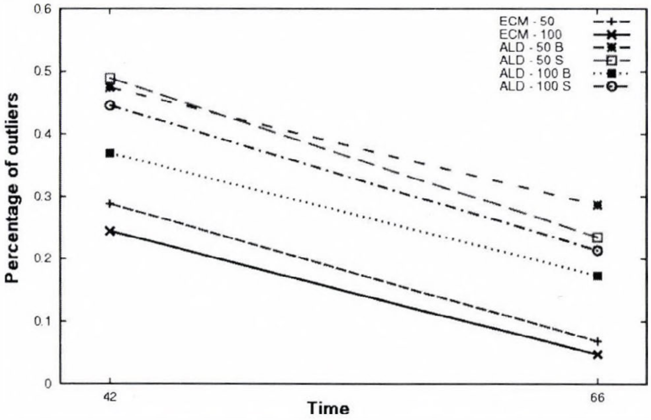


Fig. 8. Talagrand outliers for precipitation for ECMWF EPS and ECMWF/ALADIN, calculated from the first three case studies. The different configurations are: ECMWF EPS 50 members (ECM–50), ECMWF EPS 100 members (ECM–100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain (ALD–50 B), ECMWF/ALADIN using 50 EPS members and smaller clustering domain (ALD–50 S), ECMWF/ALADIN using 100 EPS members and bigger clustering domain (ALD–100 B), ECMWF/ALADIN using 100 EPS members and smaller clustering domain (ALD–100 S).

The diagram clearly shows that there is a direct relationship between the optimal spread and the ensemble size: the best results are obtained by the use of the original ECMWF EPS systems (having 100 or 50 members without any clustering). It can be seen, that the creation of super-ensemble (joining two ensemble systems) can improve the spread of the system, therefore, it is beneficial to use two sets of EPS members instead of a single one. The decrease of the ensemble size is inevitably decreasing the spread, therefore, the information loss might be significant. This fact is confirmed examining the 10 representative members of ECMWF EPS with respect to the 50 or 100 members original system (diagram not shown). The Talagrand outlier values of the 10 ECMWF representative members are roughly the same as for the ECMWF/ALADIN configurations, which shows that the ALADIN runs do not have a significant impact on the spread in case of precipitation. Comparing the different clustering configurations (after ALADIN runs) it can be noticed, that better spread is obtained with the use of 100 EPS members. The results considering the clustering domains are not so clear, because in the case of 100 members the bigger domain yields better spread, but in the case of 50 members the smaller domain is better. Consequently, further studies are needed to determine the relationship between these two clustering parameters (the size of the original ensemble and the clustering domain) and the best combination to be used.

3.2.2 Relative Operating Characteristic (ROC) diagrams

The Relative Operating Characteristic (ROC) is a graph of the hit rate against the false alarm rate for different decision thresholds (e.g., precipitation exceeding 10 mm) (*Mason*, 1982; *Stanski et al.*, 1989). For probability thresholds (e.g., 10%, 20% of the ensemble members forecasted the event), the corresponding hit rates H and false alarm rates F are computed and entered into the ROC diagram with H defining the y-axis and F the x-axis. ROC area is considered as the area under the ROC curve. Bigger ROC reflects higher skill. The diagonal line in the ROC diagram represents the climate (*Persson*, 2001).

ROC diagrams were calculated for precipitation using the observations mentioned above for the 42-hour and 66-hour forecasts. Five thresholds were used: 5, 10, 15, 20, and 30 mm, respectively. There are three aspects to be considered before the evaluation of the verification results. First, the amount of precipitation was much less in the first day, therefore, in the 42-hour forecast the number of points used in the statistics is lower, especially for the larger threshold values. The second issue is that the singular vectors used in the computation of global initial perturbations are targeted towards medium range, i.e., after 2–3 days (as already emphasized earlier). The last consideration is that the clustering was performed at 60 and 84 hours, which also penalizes the shorter integration times. In practice regarding the statistical significance of the scores, one can say that considering the 66-hour range on 70% of the grid points was more than 10 mm of precipitation forecasted (which means nearly 400 points), consequently the diagrams are based on a relatively large sample, therefore, the results can be considered statistically significant. The sample for the bigger precipitation thresholds was also relatively large, because nearly at 40% of the grid points were more than 20 mm of precipitation forecasted. Comparing the ROC scores of the 66-hour forecast for the ECMWF EPS and ECMWF/ALADIN (*Fig. 9*), it can be seen that the performance of the ALADIN system for the lower thresholds (5 and 10 mm) is better than that of the raw ECMWF ensemble. However, with a more careful look one might notice, that for the same probability thresholds the hit rate of the ECMWF/ALADIN system is a bit lower, but the false alarm rate of the ECMWF EPS is significantly higher (the ROC curve is not so convex).

Altogether this means that the ECMWF EPS was more likely to overestimate the small amount of precipitation (*Fig. 9*). This overestimation can also be seen in terms of spread on the Talagrand diagram of the ECMWF EPS (*Fig. 7*).

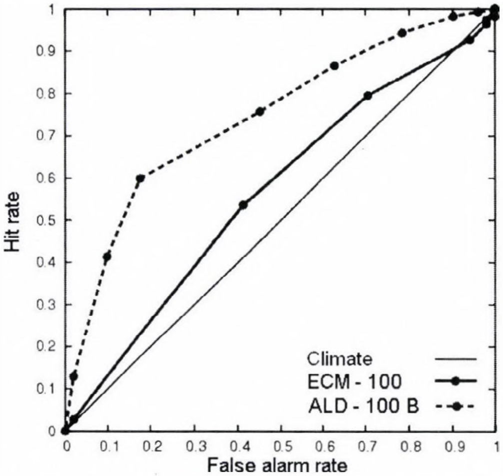


Fig. 9. ROC diagrams for 5 mm/24 h accumulated precipitation threshold for +66 h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain), diagonal line represents the climate. Black dots represent the different probability thresholds. ROC areas are 0.6 (ECMWF EPS) and 0.77 (ECMWF/ALADIN).

For higher thresholds (20 and 30 mm) the situation is less clear (*Fig. 10*): according to the overall impression, the behavior of the two systems seem rather similar (this is confirmed by the respective ROC area values). False alarm rates for the ECMWF EPS system were a bit lower (especially at higher probability thresholds), but the hit rates of the ECMWF/ALADIN were significantly higher (especially at lower probability thresholds). This results in that the curve of the ECMWF/ALADIN runs above the curve of the ECMWF EPS on the right side of the diagram (*Fig. 10*). As a summary it can be said, that the ECMWF/ALADIN system has a better skill measured by ROC diagrams for precipitation forecasts than the global system. This is especially true for lower threshold values, while for the bigger thresholds the small probability events are forecasted with higher reliability in the ALADIN system.

A ROC diagram was calculated also for the last case study (where the global system overestimated the precipitation, *Fig. 11*). The ECMWF EPS (50 members) system is characterized by dramatic false alarm rates, which is the direct consequence of the large precipitation overestimation (the amount of precipitation exceeded 10 mm only at very few locations on this day). The forecast of the ECMWF/ALADIN was more accurate, which can be clearly seen from lower false alarm rates, and higher hit rates (however, the ALADIN system was also far from being perfect for that case, even though outperforming the ECMWF EPS system).

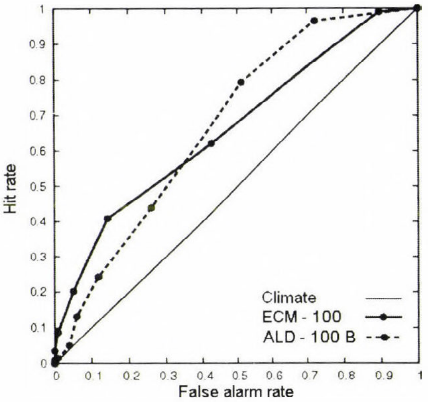


Fig. 10. ROC diagrams for 20 mm/24 h accumulated precipitation threshold for +66 h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain), diagonal line represents the climate. Black dots represent the different probability thresholds. ROC areas are 0.67 (ECMWF EPS) and 0.69 (ECMWF/ALADIN).

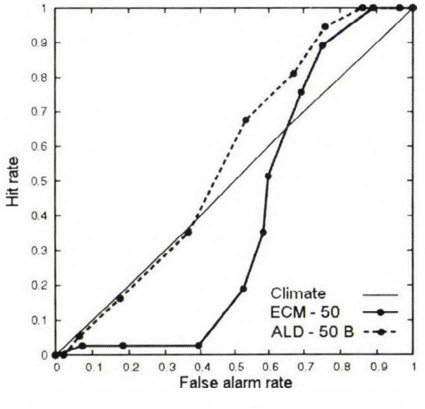


Fig. 11. ROC diagrams for 10 mm/24 h threshold for +66 h accumulated precipitation forecasts calculated from the fourth case study. Thick solid line is relative to ECMWF EPS (50 members), dashed line is relative to ECMWF/ALADIN (10 representative members from 50 global members, bigger domain), diagonal line represents the climate. Black dots represent the different probability thresholds. ROC areas are 0.4 (ECMWF EPS) and 0.57 (ECMWF/ALADIN).

As it was mentioned above, the ROC area is a good additional indicator of the skill. To compare the different clustering configurations, time evolution of the ROC area for the five thresholds was plotted again for the first three case studies. *Fig. 12* shows the ROC areas for ECMWF EPS (50 and 100 members) and the four clustering configurations for the 5 mm/24 h and 20 mm/24 h thresholds.

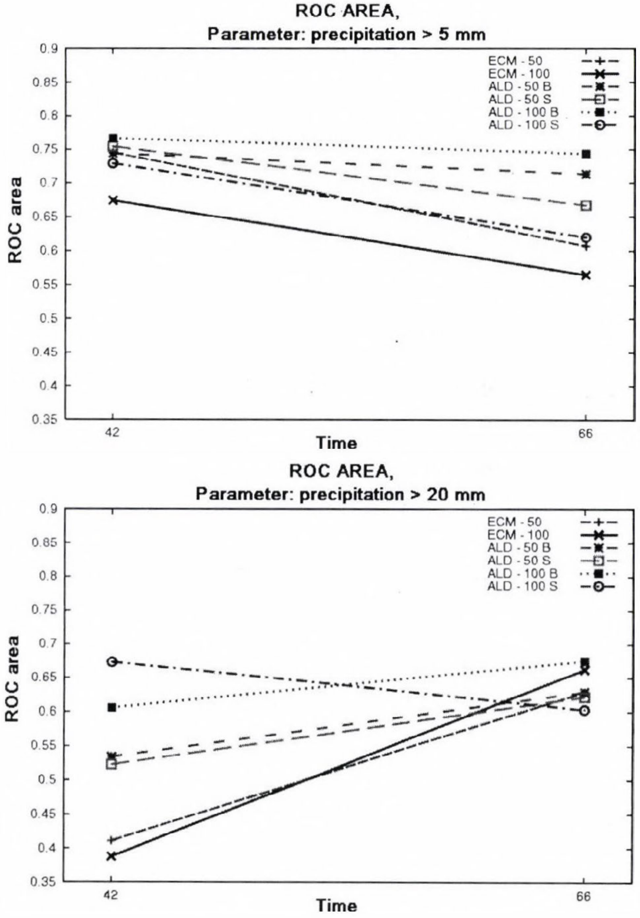


Fig. 12. Time evolution of the ROC area for the 5 mm/24 h (upper diagram) and 20 mm/24 h accumulated precipitation thresholds (lower diagram). The different configurations are: ECMWF EPS 50 members (ECM–50), ECMWF EPS 100 members (ECM–100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain (ALD–50 B), ECMWF/ALADIN using 50 EPS members and smaller clustering domain (ALD–50 S), ECMWF/ALADIN using 100 EPS members and bigger clustering domain (ALD–100 B), ECMWF/ALADIN using 100 EPS members and smaller clustering domain (ALD–100 S).

For all the five thresholds basically the same conclusions can be drawn: the best results are obtained for the ECMWF/ALADIN system (better than for the ECMWF system), when the clustering configuration was the combination of 100 members and bigger domain. For higher thresholds the 100 members, smaller domain configuration gave the best results for the +42 h forecast time, however, one must remember that these verifications are based on three cases and a large precipitation fall on the second day of the cases. Consequently, the sample for the +42 h forecast time for higher thresholds is not large enough to provide an accurate and statistically reliable result. At first sight it could be strange, that for higher threshold the skill of the forecasts gets higher with increasing integration time (which is the case for all cases, except for ECMWF/ALADIN 100 members and smaller domain configuration, where the skill continuously decreases with the integration time). As it was mentioned above, the global singular vector technique of the ECMWF EPS is targeted to the medium range, which results in that the spread of the ensemble is not sufficient on the first day of the forecast. The lack of spread results in a decrease of the skill on the first day, especially for higher precipitation thresholds.

Comparing *Fig. 8* with *Fig. 12*, it can be noticed that for Talagrand outliers the ECMWF EPS (100 members) is the best, while for ROC area the ECMWF/ALADIN configuration performs better. At first glance this looks like a contradiction, however this can be understood and explained with a bit of speculation. The Talagrand diagrams represent the spread of the ensemble members for a given integration time, i.e., the ideal diagram has a flat shape. The ALADIN diagrams are farther from the ideal case than that of the ECMWF, which means that the spread does not seem satisfactory for ALADIN (this worsening is coming from the fact that while using the representative members instead of the full system, there is some loss of information, which is just partly compensated by the integration of the limited area ensemble system). As far as the ROC curves are concerned, they are representing the true skill of the forecasts with respect to the measured values. In that sense the skill is a more important characteristic than the spread itself, therefore, even though the spread is not optimal for the ALADIN EPS, the good representation of skill confirms the ability of the system to improve the poor precipitation forecasts of the global system.

The difference between these two techniques can also be understood from the following example. Let us consider a precipitation observation, when 11 mm was registered. Suppose, that the 100-member ECMWF EPS generally underestimated this event, but because the ensemble size is very large, it could happen, that one or two members predicted more than 11 mm. Suppose, that the 10-member ECMWF/ALADIN overestimated this event, and every member forecasted more than 11 mm. Considering the Talagrand diagram, this case falls into the extreme value for ECMWF/ALADIN and normal value for ECMWF EPS, consequently ECMWF EPS is better when investigating Talagrand outliers. On the contrary, examining the ROC area, the ECMWF/ALADIN is much more skillful for the 10 mm threshold, because every member forecasted higher values than the precipitation threshold. Consequently, one must be very careful when using different verification techniques, especially if ensemble systems with different populations are compared. Finally, it is also mentioned that the number of cases was far from being satisfactory in order to draw fully coherent and statistically meaningful conclusions.

4. Conclusions

One of the main goals of ensemble forecasting is to improve the forecasts of extreme weather events. Because of the relatively low horizontal resolution of the global ensemble systems, they are not really suitable for predicting heavy precipitation events, especially in local convective situations. Consequently, it is worth trying to improve the results of the global ensemble systems with high resolution limited area models.

In this paper the description of the ALADIN limited area ensemble system was presented, which improves the forecasts of the ECMWF EPS. Four case studies involving heavy precipitation (in three of the cases) were investigated. The subjective verification on the one hand showed, that the downscaling improved the forecasts of the global system by decreasing the rate of underestimation in the case of heavy precipitation (first three cases), and on the other hand, it proved that the system is capable to correct events corresponding to global precipitation overestimation (fourth case study).

Objective verification of the global and downscaled systems was performed for different parameters too. In case of precipitation the high-density precipitation observing network of the HMS was used. The comparison of the Talagrand diagrams showed that the spread of the ensemble was quite sensitive to the population. Consequently, the ECMWF EPS system resulted in a better spread than the 10 members ECMWF and ECMWF/ALADIN.

To investigate the skill of the forecasts, ROC diagrams were plotted and ROC areas were computed. For lower threshold higher false alarm rates were detected for the ECMWF EPS than for the ECMWF/ALADIN. For higher thresholds the hit rates of the ECMWF/ALADIN were higher, which means that the limited area system predicted the large amount of precipitation better. Time evolution of the ROC area showed, that among the four clustering configurations that one performs the best, which uses 100 EPS members and larger clustering domain. This configuration of the ECMWF/ALADIN gave better results than the original ECMWF EPS. It is important to remark, that the verification was carried out on the basis of only four days, consequently, the verification results might not be significant on that stage.

For any case it can be underlined, that on the basis of the first subjective and objective evaluations of the ECMWF/ALADIN EPS system, it was found that the ALADIN system could bring benefit on top of the global ECMWF EPS system in the examined limited number of cases. These results should be further assessed and confirmed by a more detailed examination of the downscaling ensemble system by the investigation of more cases and possibly longer continuous periods of time.

Acknowledgements—The authors are grateful to *Edit Hágel* and *Gabriella Szépszó* for their help in EPS verification. We would like to thank *Gergely Bölöni* for his help in ALADIN runs. We also thank *András Horányi* for useful suggestions and comments on work and reviewing the manuscript. This work was supported by the Hungarian National Research Fund (OTKA, grant No. T/F 047295) and the Hungarian National Office for Research and Technology (NKFP, grant No. 3A/051/2004 and JÁP, grant No. 2/007/2005).

References

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

Borgatti, S.P., 1994: How to explain hierarchical clustering. *Connections 17*(2), 78-80.

Buizza, R., Richardson, D.S., and *Palmer, T.N.*, 2001: The new 80-km High-Resolution ECMWF EPS. *ECMWF Newsletter,* No. 92, 2-9.

Courtier, Ph., Freydier, C., Geleyn, J.-F., Rabier, F., and *Rochas, M.*, 1991: The ARPEGE project at Météo-France. *Workshop on Numerical Methods in Atmospheric Models.* Reading, UK. ECMWF. Vol. 2, 193-231.

Du, J., McQueen, J., DiMego, G., Toth, Z., Jovic, D., Zhou, B., and *Chuang, H.*, 2006: New Dimension of NCEP Short-Range Ensemble Forecasting (SREF) System: Inclusion of WRF Members. *Preprint, WMO Expert Team Meeting on Ensemble Prediction System.* Exeter, UK, Feb. 6-10, 2006, 5 pp.

Ghelly, A., 2002: Verification of precipitation forecasts using data from high-resolution observation *ECMWF Newsletter,* No. 93, 2-7.

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

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

Ihász, I., 2003: Experiments of clustering for central European area especially in extreme weather situations. In *Proc. of 9th Workshop on Meteorological Systems,* 20-22 October 1997, 112-116.

Marsigli, C., Montani, A., Nervosa, F., Paccagnella, T., Tibaldi, S., Molteni, F., and *Buizza, R.*, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments on four Alpine flood events. *Q. J. Roy. Meteor. Soc. 127*, 2095-2115.

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

Mason, I.B., 1982: On scores for yes/no forecasts. *Preprints, Ninth AMS Conference on Weather Forecasting and Analysis, Seattle, Washington,* 169-174.

Molteni, F., *Buizza, R.*, *Marsigli, C.*, *Montani, A.*, *Nerozzi, F.*, and *Paccagnella, T.*, 2001: A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. *Q. J. Roy. Meteor. Soc. 127*, 2069-2094.

Montani, A. Capaldo, M., *Cesari, D.*, *Marsigli, D.*, *Modigliani, U.*, *Nerozzi, F.*, *Paccagnella, T.*, *Patruno, P.*, and *Tibaldi, S.*, 2003: Operational limited-area ensemble forecasts based on the 'Lokal Modell'. *ECMWF Newsletter,* No. 98, 2-7.

Persson, A., 2001: User Guide to ECMWF forecast products. *Meteorological Bulletin.* ECMWF, Reading.

Stanski, H.R., *Wilson, L.J.*, and *Burrows, W.R.*, 1989: Survey of common verification methods in meteorology. *World Weather Watch Technical Report 8.* World Meteorological Organization, Geneva.

Szintai, B., 2005: Clustering of the medium-range ensemble forecasts of the ECMWF (in Hungarian). *Proc. of the 27th Scientific Student Conference* (OTDK). Eötvös Loránd University, Budapest, March 21-23, 2005, 144 p.

Talagrand, O., *Vautard, R.*, and *Strauss, B.*, 1998: Evaluation of probabilistic prediction systems. *Proc. of ECMWF Workshop on Predictability.* 20-22 October 1997, 1-25.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 279–297

Numerical simulation of a tornado producing thunderstorm: A case study

Ákos Horváth[1*], **István Geresdi**[2] and **Kálmán Csirmaz**[1]

[1]*Hungarian Meteorological Service,*
Vitorlás u. 17, H-8600 Siófok, Hungary; E-mail: horvath.a@met.hu

[2]*University of Pécs, Institute of Environmental Sciences,*
Ifjúság u. 6, H-7624 Pécs, Hungary; E-mail: geresdi@gamma.ttk.pte.hu

(Manuscript received in final form September 4, 2006)

Abstract—Thunderstorms often cause serious damages due to the strong surface outflow or heavy precipitation. There are some weather patterns, which especially promote breaking out of severe thunderstorms. Radar and visual observations show that some of these thunderstorms can develop into supercell. One of these typical weather patterns is the prefrontal squall line moving from southwest direction (so called Slovenian Squall Lines). In this paper the results about the formation and development of a thunderstorm associated by this type of squall line is presented. Severe thunderstorms formed on May 18, 2005 over the eastern part of Hungary were investigated using radar observations and a mesoscale numerical model (MM5). The time and position of the most intensive thunderstorm coincide well with radar observations.

The case study shows that there is a competition between thunderstorms for the wet and warm air. A thunderstorm which can collect the wet and warm air from larger area will have longer lifetime and more intensive updraft. The case study shows that in the case of absence of directional wind shear, the merging of the updraft cores can result in a supercell. Although the formation of the new cells frequently occurs by splitting, this process did not happen in this case.

Analysis of the numerical simulation indicates the presence of three different types of downdraft regions in an intensive thunderstorm. The low level downdraft was generated by the precipitation loading. The intensity of this downdraft is also affected by melting and evaporation of the precipitation elements. The midlevel downdraft does not reach the surface, and it is driven by negative thermal buoyancy and set by an interaction of the updraft with the vertical wind shear. Downdraft cores at high level could be associated with the overshooting.

Key-words: severe thunderstorm, supercell, squall-line, MM5

---

[*] Corresponding author

1. Introduction

Strong convective storms frequently produce large hail and intensive outflow or tornado. These phenomena are consequences of very complicated and strongly interacting dynamics and microphysical processes. The strong interaction between the three-dimensional air flow and cloud microphysics renders the research difficult. *Browning* (1964) presented firstly a conceptual model of rotating thunderstorms, which he named supercells. He stated that these supercells develop in an environment characterized by strong vertical wind shear, and these storms mainly move to the right of the mean wind direction. In the 1960s and until the mid-1970s only two-dimensional slab or axisymmetric models were available for the simulation of formation and development of clouds (e.g., *Orville* and *Kopp*, 1977). The two-dimensional approximation - with high spatial resolution and detailed microphysical description - allowed the correct numerical simulation of the layer clouds (slab symmetric model) and small convective clouds without horizontal wind shear (axisymmetric model). Due to the asymmetric three-dimensional air flow in the severe thunderstorms, the numerical modeling of this type of clouds needs a three-dimensional dynamical description. From the mid-1970s researchers began to simulate the internal structure of supercells with three dimensional numerical models. By the early 1980s these simulations could reproduce many important features of the observed storms (*Wilhelmson* and *Klemp*, 1981; *Klemp et al.*, 1981). In the early and mid-'80s the theory of supercells formation were presented in number of papers (*Schlesinger*, 1980; *Rotunno*, 1981; *Rotunno* and *Klemp*, 1982; *Davies-Jones*, 1984). The results of these investigations suggest that the supercell dynamics are governed mainly by the interaction between the vertical wind shear in the environment and the convective updraft. The transition from "regular" supercell to tornadic supercell was described at first by *Lemon* and *Doswell* (1979). They defined and described the frontal, low-level structure of the supercell (mesocyclone), and stated that the development of this structure is in close connection with the tornadogenesis. *Lilly* (1986) suggested that the lifetime of supercells could be related to the helical property of the storm relative flow. The helicity also correlates well to the tornado intensity in long lived supercells (*Davies-Jones et al.*, 1990; *Kerr* and *Darkow*, 1996).

The new generation of mesoscale models (e.g., MM5 and RAMS) include a wide range of physical processes, which can affect the cloud formation (*Pielke et al.*, 1992; *Dudhia*, 1993). Due to the application of the nesting technique, the horizontal resolution of a mesoscale model could be similar (about 1-2 km in the region where thunderstorms form) to that of the above mentioned cloud models with limited domain size. In the domains with high horizontal resolution the formation of convective clouds could be simulated with explicit scheme instead of using cumulus parameterization. Not only the dynamical description was significantly improved in the new generation of mesoscale models, but the microphysical description as well. The simulation of the growth and melting of ice particles allows us to take into consideration the effect of latent heat of fusion on the cloud dynamics. By nowadays, mesoscale models have proved to be a useful tool for short or ultrashort term forecasting of the convective phenomena (e.g., *Tudurí et al.*, 2003).

The research of severe thunderstorms in the territory of Hungary started in the early 1960s. The aim of the works was to improve the efficiency of storm warning at Lake Balaton (*Götz*, 1966; *Böjti et al.*, 1964). The research mostly focused on squall lines coming from southwest direction and causing severe weather and serious accidents at Lake Balaton and on Transdanubian regions. These studies described the synoptic and some dynamic conditions of these kinds of severe weather (*Götz*, 1968). Later, hail suppressing system, operated in the south west part of Hungary, required the investigation of microphysical processes which occur in thunderstorms (*Zoltán* and *Geresdi*, 1984). Claim for more accurate storm warning required a more detailed description of dynamical conditions of formation of squall lines (*Horváth* and *Práger*, 1985). In the early 1990s specific forecast methods were developed (*Bartha*, 1987), which significantly improved the efficiency of the storm warning using statistically based methods. The nowcasting system developed by the end of the last century gave a new tool for the forecast of the severe weather (*Geresdi* and *Horváth*, 2000; *Horváth* and *Geresdi*, 2003; *Geresdi et al.*, 2004). By this time radar observations and mesoscale numerical models had become available for the operational weather forecast. The appearance of supercells and the related tornado formation in the territory of Hungary were described first time in *Horváth* (1997). As the weather radar network of Hungarian Meteorological Service became more accurate, more supercell cases were recognized (*Horváth* and *Geresdi*, 2003), and the first detailed supercell case studies was published by *Horváth* (2005). These case studies show that synoptic conditions of severe weather are similar in the North American region and Carpathian Basin. However, some phenomena, like strong directional wind shear is less frequently observed in Hungary than in the USA. Strong directional wind shear results in the separation of the updraft region from downdraft region, which is necessary for the formation of the long-lived supercells.

The purpose of this paper is to investigate the relation between the large scale synoptic conditions and formation of intensive thunderstorms. Results presented by the MM5 mesoscale model are analyzed and compared with radar observations.

2. Description of the numerical model

The numerical simulations were made by Version 3 of the MM5 (NCAR-PSU Mesoscale Model), which was described by *Dudhia* (1993). A terrain-following sigma coordinate system is applied in the model. The predictive variables are: pressure perturbation, three momentum components, temperature, specific humidity, and mixing ratio of five different types of hydrometeors (cloud water, cloud ice, rain, snow, and graupel particles). For this study, the model is integrated with horizontal resolution of 2 km, and with 26 vertical levels. The partial differential equation system is solved by using relaxation lateral boundary condition and radiation upper boundary condition.

The high horizontal resolution allows us to run the model without cumulus parameterization. Explicit bulk microphysical scheme with five different types of hydrometeors was used to simulate the formation of cloud and precipitation elements (*Reisner et al.*, 1998). Collision and coalescence process between different types of hydrometeors, furthermore, diffusion of vapor, freezing of liquid elements, and melting of ice particles were simulated. The rate of the rain drop formation due to collision of cloud droplets (autoconversion) depends on the number concentration and characteristics of the size distribution of cloud droplets. These parameters are different in the continental and maritime air masses. Graupel particles can form via freezing of rain drops or aggregation of snow particles. Equation of conservation was not only solved for the mixing ratios of hydrometeors, but for the number concentration of cloud ice as well.

The planetary boundary layer (PBL) is described by the non-local PBL scheme based on *Troen* and *Mahrt* (1986). Compared with other non-local or high-order closure schemes, this PBL scheme proved to be more efficient, because it needs less computer capacity. Land-surface processes are simulated by OSU LSM (Oregon State University Land-surface Model). It is based on the coupling of Penman's potential evaporation approach (*Penman*, 1948) modified by the atmospheric stratification effect (*Mahrt* and *Ek*, 1984), the multi-layer soil model (*Mahrt* and *Pan*, 1984), and the single-layer canopy model (*Pan* and *Mahrt*, 1987). Canopy resistance is formulated after *Jarvis* (1976) using relative stomatal conductivity formulae of *Noilhan* and *Planton* (1989). Atmospheric stratification is simulated by applying the Monin-Obukhov similarity theory (*Oncley* and *Dudhia*, 1995). Richard's and heat flow equations are used to calculate soil moisture and temperature, respectively. A more detailed description of these processes can be found, for instance, in papers of *Chen* and *Dudhia* (2001) and *Sridhar et al.* (2002).

3. Synoptic conditions

Summer weather can cause severe situations in the Carpathian Basin. Convective events such as thunderstorms, squall lines, sometimes tornado-producing supercells develop frequently during the summer. Conditions favorable to formation of severe weather were summarized in *Horváth* and *Geresdi* (2001). In this paper three convective components were defined: convective instability, convergence, and wind shear (not necessarily directional wind shear). Case studies show that extreme convective events, like supercells, occur when all of the three components exist. In the case investigated in this study, every components played an important role in the generation of a squall line. On May 18, 2005 a cyclone in mature phase moved slowly from the Mediterranean region to north-east direction. The warm sector of the cyclone contained moist and unstable air mass drifted above the Carpathian Basin. This process increases the convective instability in this region.

The large temperature gradient at the 850 hPa level over Hungary and the mean sea level pressure indicated a sharp cold front, which caused strong convergence near to the surface. High level jet stream at the 300 hPa level (*Fig. 1c*) resulted in large wind shear. Three mesoscale meteorological objects can be distinguished on May 18, 2005, 12:00 UTC (*Fig. 2*): (i) a cold front coming from northwest; (ii) a prefrontal squall line moving from southwest to northeast (referred later as SQ); (iii) a convergence line with stratiform precipitating system (referred later as CL).

The squall line reached its mature phase at 15:30 UTC. At this time a rotating thunderstorm with wall cloud was observed at Lake Balaton. Meanwhile, in the less significant convergence line, embedded thunderstorms started to develop, and from 15:45 to 16:15 UTC a hook echo could be seen on the radar images (*Fig. 3*). (It has to be mentioned, that the reflectivity at every pixel is equal to the maximum reflectivity observed in the column which belongs to the pixel. That means, that the depicted reflectivity values were not necessarily observed at the same heights.) This storm caused serious damages. The characteristics of the damages, which was widely examined and well documented (*Horváth*, 2005), suggest that it was caused by a tornado. This idea was supported by the shape of the contour of high reflectivity, which is similar to that of tornado producing thunderstorms.

This thunderstorm could be defined as a supercell. There are two different ways in the literature to define the supercell. One definition says that the supercell means a long-lived thunderstorm, which has only one cell (*Browning* and *Foot*, 1976). The mature phase of this cell can last for hours. The other definition, more frequently used nowadays, characterizes the supercell as a rotating cell, which has not necessarily a long lifetime (*Doswell*, 2001).

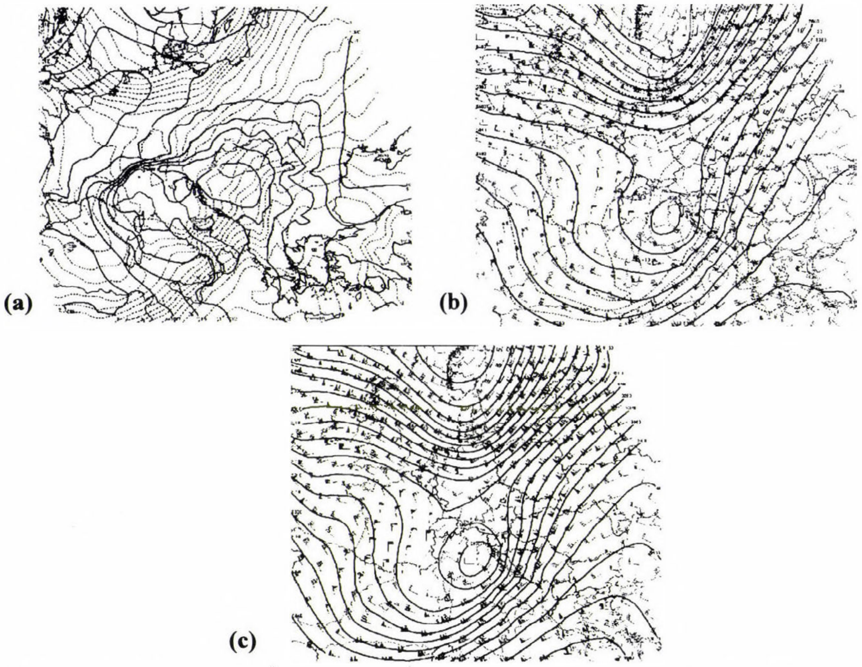


Fig. 1. ECMWF analysis on May 18, 2005, 12:00 UTC. (a) Sea level pressure (thick lines) and 850 hPa temperature (dashed lines), (b) 500 hPa height (thick lines), temperature (dashed lines) and wind field, (c) 300 hPa height and wind field.

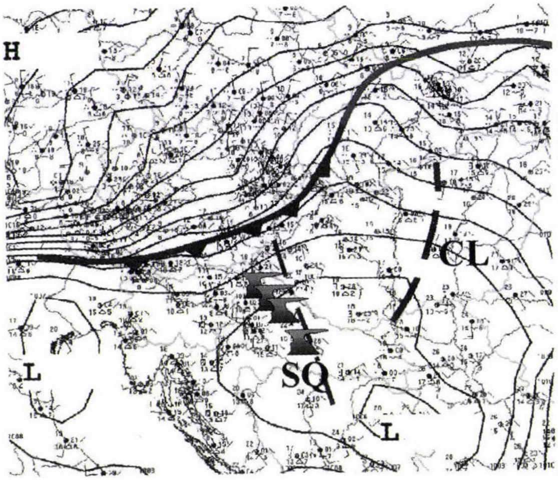


Fig. 2. Mesoscale weather conditions on May 18, 2005, 12:00 UTC. SQ denotes the prefrontal squall line and CL denotes the convergence line with stratiform precipitating system.

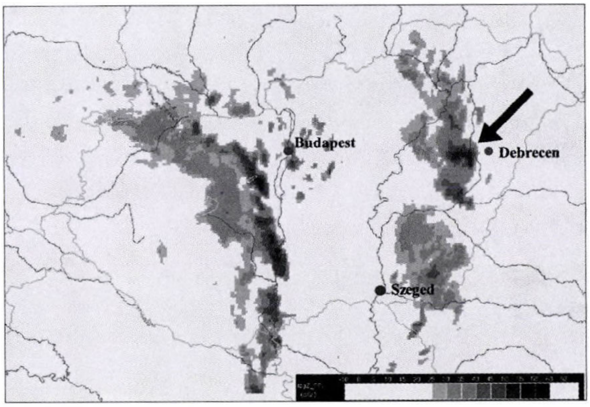


Fig. 3. Radar image on May 18, 2005, 15:15 UTC. The black arrow shows the position of the hook echo in the eastern convergence line.

4. Results of the numerical experiment

4.1 Model initialization

The initial condition for the MM5 model (mean sea level pressure, three dimensional temperature, humidity, wind field, soil temperature, and soil humidity) was given by the 12:00 UTC ECMWF analysis. Unfortunately, the data of this analysis are available only at about 18:00 UTC, so this kind of initialization can not be applied for operational weather forecast. In the case of scientific research, the delay of the initial condition is not a disadvantage. The boundary conditions were transferred from ECMWF forecast at every third hour.

Because the model was initiated by using 12:00 UTC data, this time coincided with 0 hours model time. The extension of the model domain was about 440 × 360 × 16 km. An experimental model run for a 6-hour forecast takes about 3-hour computer time on a 16-processor SGI-Origin 2000 computer.

4.2 General results

As it was mentioned above, the most significant characteristic of a supercell is its rotation (*Doswell*, 2001). The analysis of wind fields at the 950 hPa level shows that the air started to rotate at 14:00 UTC in two regions, on the southwest part (in SQ) and middle part (in CL) of the model domain (*Fig. 4a*). After the appearance, both rotating centers strengthened quickly, and at 3 hours 30 minutes of simulation two other rotating centers appeared along SQ (denoted by SQR1 and SQR2 in *Fig. 4b*). The northern center (denoted by SQR2) fitted to the visually observed supercell, which reached the lake Balaton at about

14:30 UTC. Also, an intensive rotating system (denoted by CLR1 in *Fig. 4b*) developed in the CL by 15:30 UTC. This rotating air mass located near the place where an intensive thunderstorm was observed by radar (see the location of the "hook" echo in *Fig. 3*). The agreement between the observed and simulated data (both in time and space) suggests that the MM5 simulated well the dynamical process, which could be related to the formation of supercells.

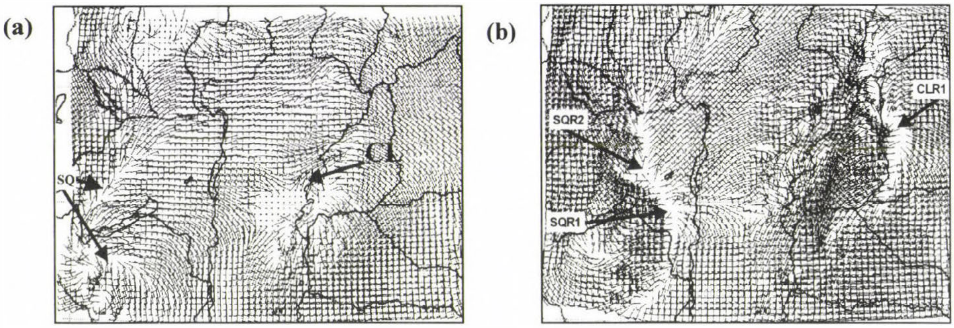


Fig. 4. Wind forecast at the 925 hPa for 14:00 UTC and +2 hours forecast (a), and for 16:00 UTC and+4 hours forecast (b).

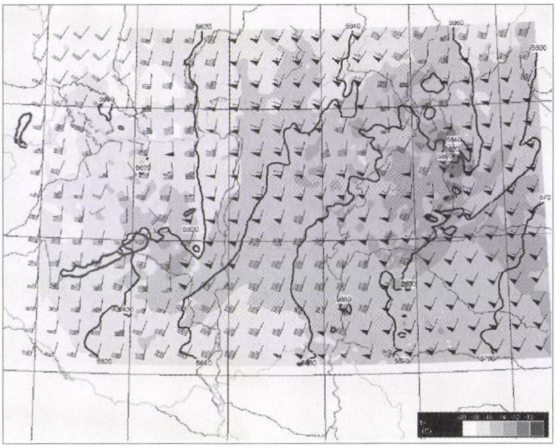


Fig. 5. Forecasted temperature and wind field at the 500 hPa level and geopotential height (solid lines) of 500 hPa at 16:00 UTC (+4 hours of simulation).

The temperature anomaly was larger in the CL system than in the SQ system (*Fig. 5*). Because the temperature difference is mostly the consequence of the different condensation rate, the difference in the anomaly suggests that more intensive thunderstorms developed in the CL system than in the SQ system. The largest deviation from the mean environmental temperature could be observed in the region, where a mesocyclone denoted by CLR1 formed.

Because more intensive thunderstorms formed in the CL system than in the SQ system, in the next part of the paper we focus on the processes occurred in the CL system.

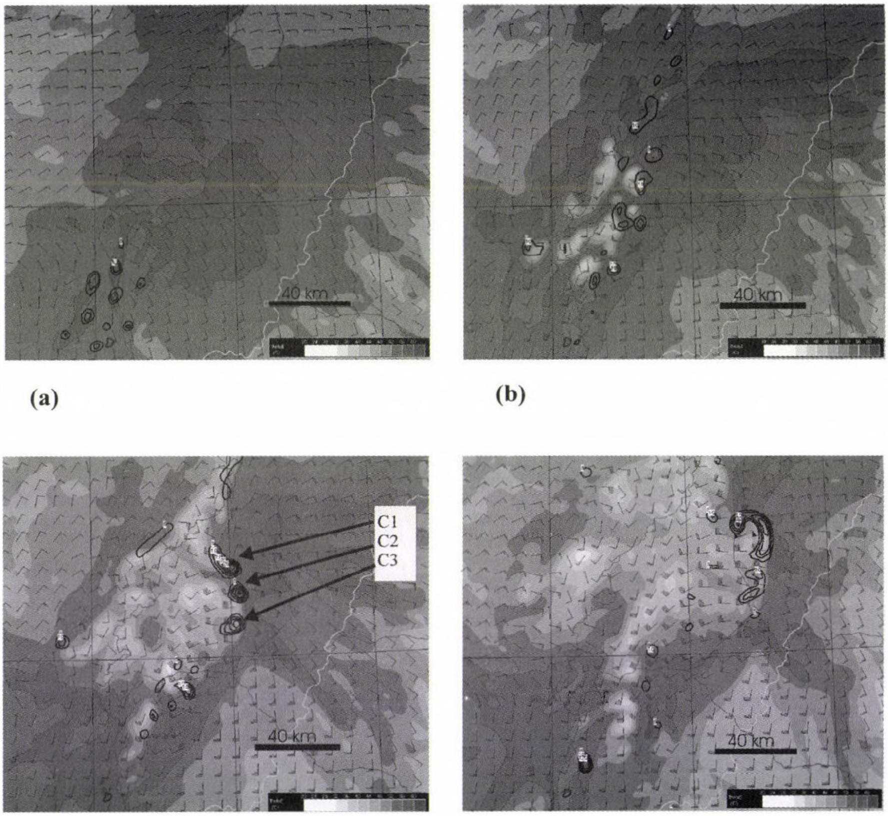
4.3 Numerical simulation of the supercell formation

The time series of the wind field and equivalent potential temperature at the 925 hPa level and the vertical velocity field at the 500 hPa level show the development of the supercell (*Fig. 6*).

Updraft region (updraft velocity is larger than 0 m/s) could be related to the first small convective storms, which can be seen in *Fig. 6a* (at 13:30 UTC). The low level wind field has been perturbed by convection, and updraft cores bounded by 5 m/s isoline appeared in the centers of the updraft regions. The analysis of *Fig. 6* shows that the formation of the thunderstorms could be related to the area of the region bounded by the isolines of 56 °C of equivalent potential temperature. This area reached its maximum value by 13:30 UTC. The increase of the area can be explained by the convergence of wet and warm air. The reduction of area of the high equivalent potential temperature (hereafter EPT) after 13:30 UTC could be explained by two reasons. (i) The appearance of cool air related to the downdraft of the thunderstorms significantly reduced the EPT over large region. (ii) The surface wind field (not shown here) shows, how the gust front propagates into the north-east direction. At the edge of the gust front, an updraft region formed due to the low level convergence ahead of the gust front near the surface. The intensive updraft in the thunderstorms generates faster inflow of warm and wet air on the surface. This process results in reduction of the area of high EPT area.

At 14:30 UTC, three regions appeared with rotating wind structure. By this time, the wind field surrounded the updraft regions had positive circulating feature, too. At 15:15 UTC (*Fig. 6c*), more updraft cores could be observed at the edge of the region bounded by the isoline of 56 °C. Three of them (denoted by C1, C2, and C3 in *Fig. 6c*) were more intensive than the others. The updraft region denoted by C1 was the most intensive. In a large region, the updraft velocity was larger than 10 m/s. Due to the strong convergence near the surface, this thunderstorm developed quickly.

The convergence, which could be observed at this cell, is responsible for the reduction of the area bounded by the 56 °C EPT isolines. By 16:00 UTC, the updraft region denoted by C1 reached its mature phase, and the region of warm and moist air almost completely disappeared by this time. Time series of the vertical velocity field between 15:15 and 16:00 UTC (not shown) suggests that the updraft region C1 was strengthened by the merging of updraft region C2. This process is a frequently observed way of intensification of a thunderstorm (e.g., *Simpson et al.*, 1980). The merging could be an efficient way of supercell formation when no directional wind shear occurs.



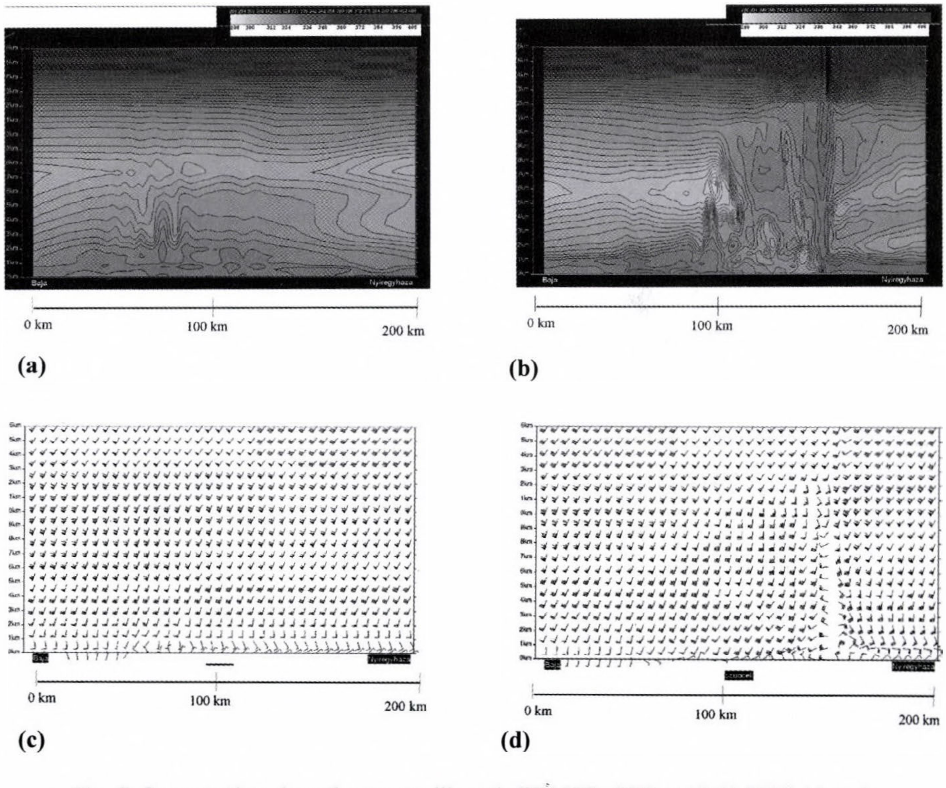
(a) **(b)**

(c) **(d)**

Fig. 6. Composite of simulated equivalent potential temperature at the 925 hPa level (shaded areas), simulated wind field at the 925 hPa level, and simulated updraft velocity filed (5 m/s contours are denoted by thick solid lines) at the 500 hPa level is plotted in each figure. (a) 13:30 UTC, (b) 14:30 UTC, (c) 15:15 UTC, (d) 16:00 UTC.

It can be stated, that in this case the cloud formation process was driven by the "competition" between cells for the warm and moist air, and by merging of nearby cells. It was the reason why only one supercell developed from the three intensive updraft cores. It is also interesting to note, that there was no splitting of cells as it is suggested by classical theories in the case of supercell thunderstorm (*Wilhelmson* and *Klemp*, 1981). Instead of splitting, merge occurred.

The vertical structure of the different parameters could be related to the supercell formation, which was investigated by analyzing the data in vertical cross sections. A vertical cross section was directed from SW to NE direction between Baja and Nyíregyháza. This cross section coincides with the trajectory of the supercell. The equivalent potential temperature (EPT) field shows the appearance of deep convection at 13:30 UTC (*Fig. 7a*). At this time, a relatively cold layer at the height of 7 km indicates convective instability. It is conspicuous, that by the time the thunderstorm reached its mature phase (at 16:00 UTC), the EPT became nearly constant in the updraft region, and the thunderstorm "connected" the lower troposphere and the stratosphere (*Fig. 7b*). The low EPT values behind the thunderstorm are consequences of the downdraft associated to the drying out of the layer between the 5 and 7 km heights.



(a) (b) (c) (d)

Fig. 7. Cross section along the supercell's path (SW-NE). EPT at 13:30 UTC (a) and 16:00 UTC (b), wind fields at 13:30 UTC (c) and 16:00 UTC (d).

While the wind field of the same cross section (*Fig. 7c*) shows that the wind shear was considerable, there was no significant directional wind shear, which is typical in the cases of long-lived supercells frequently observed in North America (*Houze*, 1993; *Doswell*, 2001).

Comparison of *Fig. 7c* and *Fig. 7d* shows that the thunderstorm significantly modified the wind field in the whole troposphere.

4.4 Detailed structure of the simulated supercell

When the C1 cell reached its mature phase (at about 16:00 UTC), the simulated wind speed in the meso-cyclone reached 40 m/s at 950 hPa level. Both low and high pressure centers could be observed in *Fig. 8*. These pressure centers are typical in the case of supercells (*Doswell*, 2001).

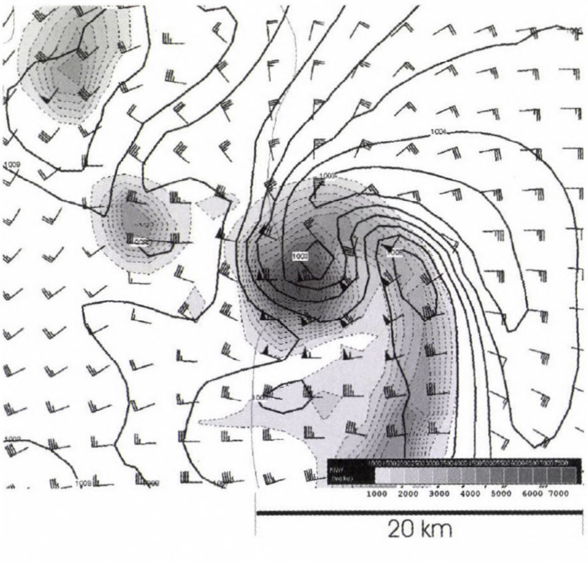
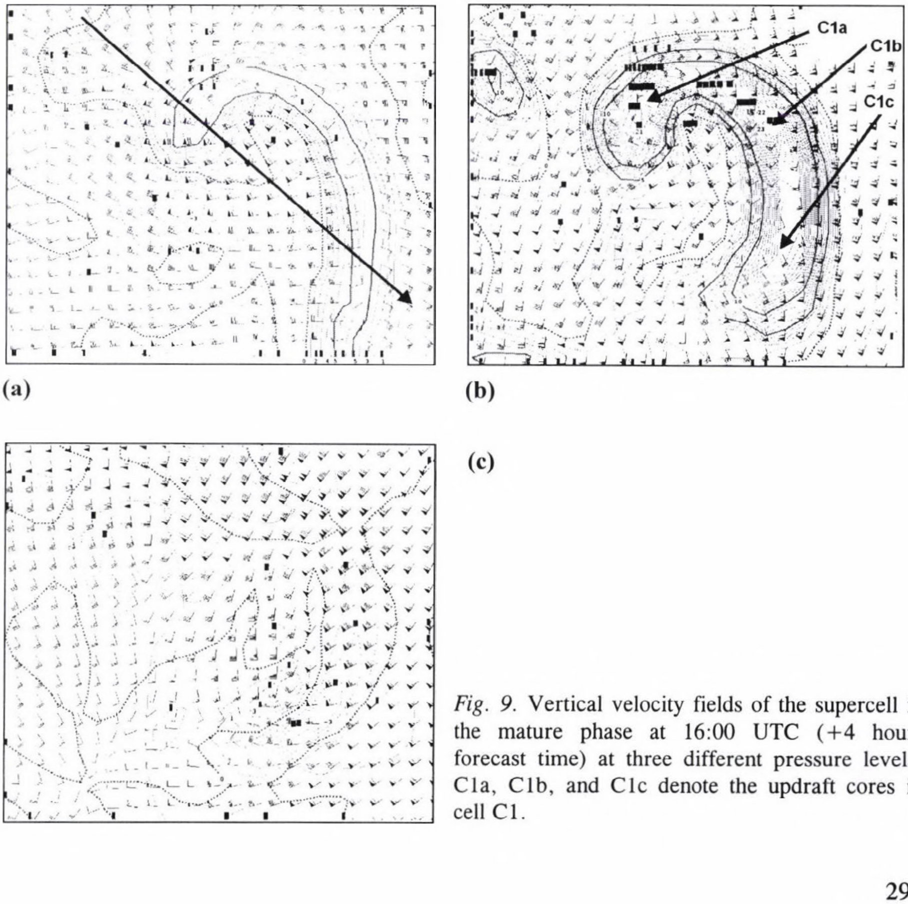


Fig. 8. Mean sea level pressure, 950 hPa wind field, and 700 hPa mixing ratio of rain water (shaded areas are bounded by dashed lines) forecasted to 16:00 UTC (+4 hours forecast).

Both the updraft region and updraft velocity are larger at the 500 hPa level than at the 925 hPa level (*Fig. 9a,b*). This difference could be explained by the released latent heat of condensation and fusion. Near the surface (925 hPa level), the updraft velocity is mainly affected by the convergence. At higher elevation, phase change of the water has also an important role in the updraft strength. At the 500 hPa level, three updraft centers can be distinguished. This structure is the consequence of that the supercell was generated from two merged cells. In the centre, almost completely bounded by the updraft regions, a downdraft region can be seen at both the 925 hPa and 500 hPa levels. This downdraft is initiated by the precipitation loading. A more intensive downdraft region can be observed at the leading edge of the updraft core at the 500 hPa level. This downdraft can not be related to the precipitation, because precipitation elements did not form in this region. Observations indicate that this type of downdrafts are mechanically driven (*Kingsmill* and *Wakimoto*, 1991). This midlevel downdraft does not reach the surface, and it is driven by negative thermal buoyancy and set by an interaction of the updraft with the vertical wind shear. At the 200 hPa level, the downdraft is dominant. The small updraft regions are related to the overshooting updraft, which penetrates into the stratosphere. This updraft region belongs to the updraft core denoted by C1c in *Fig. 9b*. The downdraft cores at this high level could be associated with the overshooting. As the air parcel becomes negatively buoyant, it starts to descend rapidly.



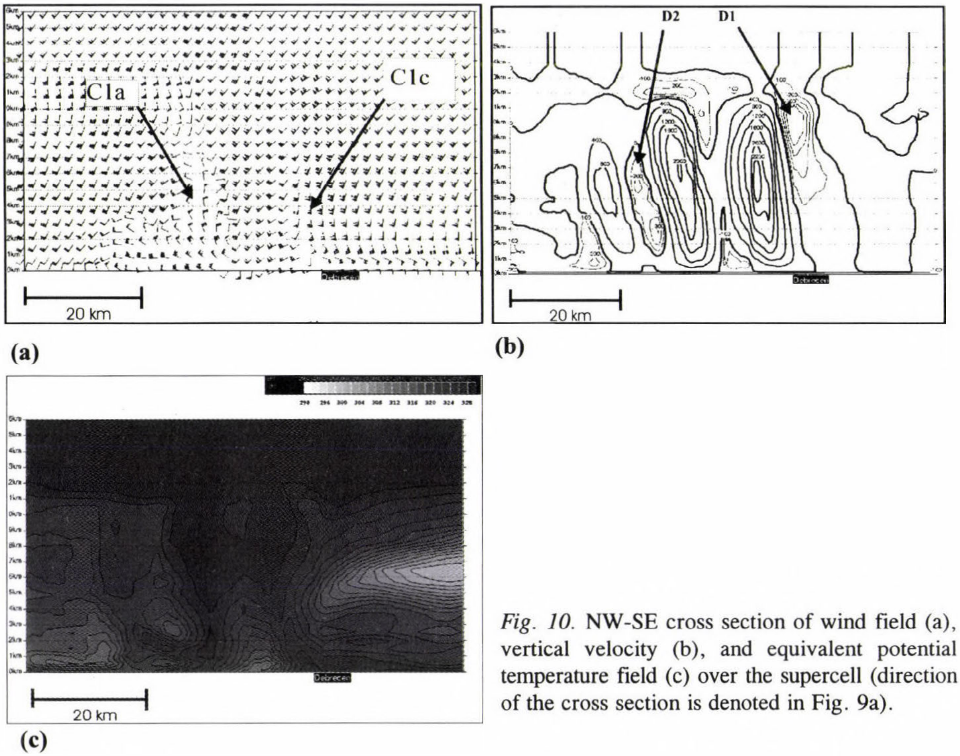
(a)

(b)

(c)

Fig. 9. Vertical velocity fields of the supercell in the mature phase at 16:00 UTC (+4 hours forecast time) at three different pressure levels. C1a, C1b, and C1c denote the updraft cores in cell C1.

Another cross section was also used to investigate the characteristics of the tornado producing thunderstorm. The direction of the cross section is indicated in *Fig. 9a*. *Fig. 10a* shows the wind field along this NW-SE cross section. Beside the modification of the wind field at the surface, the presence of the thunderstorm could be related to the modification of the wind field in two columns (denoted by C1a and C1c in *Fig. 9a*). The change of the direction and the strengthening of the surface wind are the consequences of the outflow. The change of the wind field at higher levels could be associated with updraft cores of C1a and of C1c.



(a)

(b)

(c)

Fig. 10. NW-SE cross section of wind field (a), vertical velocity (b), and equivalent potential temperature field (c) over the supercell (direction of the cross section is denoted in Fig. 9a).

Except the surface level, a strong convergence could be observed up to the height of 7 km. This is the level, where the vertical velocity starts to decrease. The rotating structure of the wind field also disappears above this level. The strong downdraft region denoted by D1 in *Fig. 10b* could be explained by the overshooting effect and the formation of midlevel downdraft.

Formation of downdraft region near the surface was initiated by precipitation loading, and in the same way, the effect of fall out of rain drops and graupel particles caused a downdraft denoted by D2 in *Fig. 10b*. The EPT field (*Fig. 10c*) shows the cold pool of the storm associated with the gust front near the surface. Both main updraft cores (C1a and C1c) have their own cold pools. The top of these pools is not higher than 5 km. Another spectacular feature is the low value of EPT between 5 and 7 km heights, on the right side of the cross section. It could be the consequence of the drying up of sinking air from the higher troposphere. High level compensating downdraft can be recognized in *Fig. 9c*.

Relation between low level downdraft and loading effect of precipitation elements is supported by *Fig. 11*. At low levels, the local maxima of mixing ratios coincide with downdraft region. However, the downdraft is not only affected by loading of precipitation elements, but the melting of the graupel particles and the evaporation of the rain drops also enhance the downdraft.

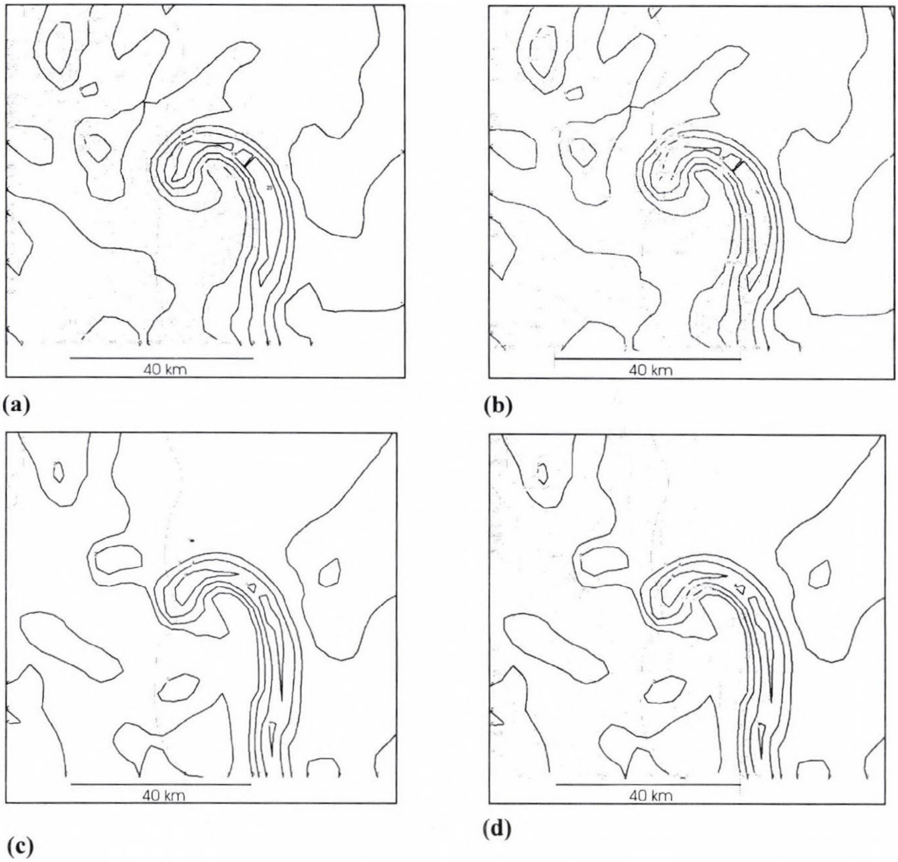


Fig. 11. Precipitation elements (dashed lines) and updraft (solid lines) at 16:00 UTC (+4 hours forecast time). (a) 700 hPa updraft (m/s) and graupel mixing ration (mg/kg), (b) 700 hPa updraft (m/s) and rain water mixing ratio (mg/kg), (c) 925 hPa updraft (m/s) and graupel mixing ratio (mg/kg), (d) 925 hPa updraft (m/s) and rain water mixing ratio (mg/kg).

The supercell affects the air flow at the higher levels of the atmosphere, too (*Fig. 12*). At the 200 hPa level, the wind field shows that the SW current flows around the cloud top. At the updraft center, a divergence of wind flags can also be seen.

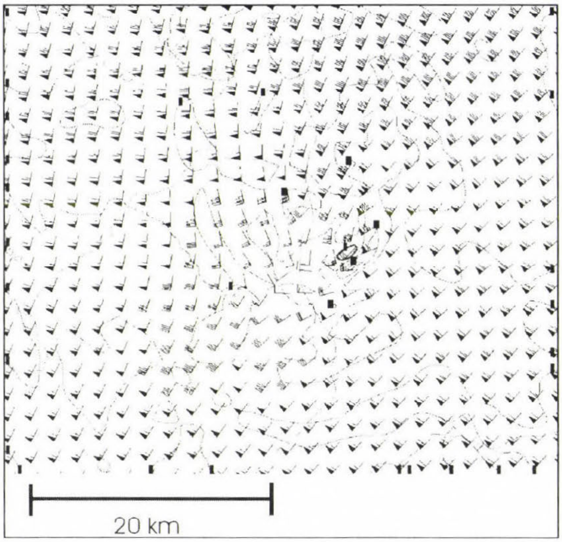


Fig. 12. Wind field of the 200 hPa level at 16:00 UTC. Thick curves denote the isoline of 2 m/s upward velocity.

5. Conclusion

The case study shows that the MM5 model with ECMWF initial and boundary conditions is able to simulate mesoscale phenomena like supercells. Rotating air masses were simulated in two different regions. A supercell formed along a squall line, but it was much weaker and it has shorter lifetime than the supercell developed at the convergence line. Some indirect proofs suggest that this supercell produced a tornado. Comparison of observation data with model results shows that the mesoscale model predicted the development of the storm close to the reality. However, by the forecast, the supercell appeared about twenty kilometers north of the real location. The region where the mixing ratio of the precipitation elements was high (*Fig. 11*) coincides well with the high reflectivity region in *Fig 3*. This successful forecast allowed us to make a detailed investigation of structure and development of a supercell, which appeared in the Carpathian Basin.

Our simulation supports the fact, that a rotating supercell can form at environmental conditions, when no directional wind shear occurs. Unfortunately, the spatial resolution of the numerical model does not allow us to simulate tornados, which could be related to the supercells. The simulated rotation of the air mass can be a good indicator for the tornado formation. It has to be noted, that the rate of air mass with diameter of order of 10 km at higher elevation not necessarily leads to tornado formation. Further research is necessary to find the other conditions, which contribute to the tornado formations. In the investigated cases, the merging of the updraft cores resulted in a formation of an intensive supercell and probably a tornado formation as well.

Detailed investigation of equivalent potential temperature shows that both the advection generated by the updraft and the downdraft generated cooling and drying reduce the region of warm and moist air. Because the advection is proportional to the updraft velocity, the more intensive updraft core can collect more moisture, which further enhances the intensity of the updraft. This competition and the merging of updraft cores resulted in the reduction of the number of the updraft cores and the formation of an intensive supercell.

Acknowledgements—The research was supported by the Hungarian Scientific Research Fund (T043010) and by the National Program for Research and Development (NKFP, project number 3/022/2005).

References

Böjti, B., Bodolainé, J.E., and *Götz, G.,* 1964: Instability lines in Hungary (in Hungarian). Beszámolók az 1963-ban végzett tudományos kutatásokról. Országos Meteorológiai Intézet, Budapest, 139-165.

Bartha, I., 1987: An objective decision procedure for prediction of maximum wind gusts associated with Cumulonimbus clouds. *Időjárás 91,* 330-346.

Browning, K.A., 1964: Airflow and precipitation trajectories within severe local storms which travel to the right of the winds. *J. Atmos. Sci. 20,* 533-545.

Browning, K.A. and *Foot, G.B.,* 1976: Airflow and hail growth in supercell storms and some implications for hail growth and hail suppression. *Mon. Weather Rev. 104,* 603-610.

Chen, F. and *Dudhia, J.,* 2001: Coupling and Advanced Land Surface-Hydrology Model with the Penn State-NCAR MM5 Modeling System. Part I. Model implementation and sensitivity. *Mon. Weather Rev. 129,* 569-585.

Davies-Jones, R., 1984: Streamwise vorticity: The origin of updraft rotation in supercell storms. *J. Atmos. Sci. 41,* 2991-3006.

Davies-Jones, R.P., Burgess, D.W., and *Foster, M.,* 1990: Test of helicity as a tornado forecast parameter. Reprints, *16th Conf. On Severe Local Storms.* Kananaskis Park, AB, Canada. Amer. Meteor. Soc., 588-592.

Doswell, C.A., 2001: Severe Convective Storms. *Meteorological Monographs 28,* No. 50. Amer. Meteorol. Soc., Boston. p. 561.

Dudhia, J., 1993: A non-hydrostatic version of the Penn State-NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Weather Rev. 121,* 1493-1513.

Geresdi, I. and *Horváth, Á.*, 2000: Nowcasting of precipitation type. Part I: Winter precipitation. *Időjárás 104*, 241-252.

Geresdi, I., *Horváth, Á.*, and *Mátyus, Á.*, 2004: Nowcasting of the precipitation type Part II: Forecast of thunderstorms and hailstone size. *Időjárás 108*, 33-49.

Götz, G., 1966: Sturmwarnung am Balatonsee. *OMSZ Issues XXX. Budapest*

Götz, G., 1968: Hydrodynamic relationships between heavy convection and the jet stream. *Időjárás 72*, 157-165.

Horváth, Á., 1997: Tornado (in Hungarian). *Légkör 42*, No. 4, 2-9.

Horváth, Á., 2005: Description of severe weather on 18th May 2005 (in Hungarian). *Légkör 50*, No. 3, 12-16.

Horváth, Á. and *Práger, T.*, 1985: Study of the dynamics and predictability of squall-lines. *Időjárás 89*, 141-160.

Horváth, Á. and *Geresdi, I.*, 2001: Severe convective storms and associated phenomena in Hungary. *Atmos. Res. 56*, 127-146.

Horváth, Á. and *Geresdi, I.*, 2003: Severe storms and nowcasting in the Carpathian Basin. *Atmos. Res. 67-68*, 319-332.

Houze, R.A., 1993: *Cloud Dynamics*. Academic Press, New York-London, 573 p.

Jarvis, P.G., 1976: Interpretation of variations in leaf water potential and stomatal conductance found in canopies field . *Philosophical Transaction of the Royal Society of London, Series B-Biological Sciences* 273 (927), 593-610.

Kerr, N.W. and *Darkow, G.I.*, 1996: Storm-relative winds and helicity in the tornadic thunderstorm environment. *Weather Forecast. 11*, 489-505.

Kingsmill, D.E. and *Wakimoto, R.M.*, 1991: Kinematic, dynamic and thermodynamic analysis of a weakly sheared severe thunderstorm over northern Alabama. *Mon. Weather Rev. 119*, 262-297.

Klemp, J.B., *Wilhelmson, R.B.*, and *Ray, P.S.*, 1981: Observed and numerically simulated structure of a mature supercell thunderstorm. *J. Atmos. Sci. 38*, 520-529.

Klemp, J.B. and *Wilhelmson, R.B.*, 1978: Simulations of right- and left-moving storms produced storm splitting. *J. Atmos. Sci. 35*, 1097-1110.

Lemon, L.R. and *Doswell III, C.A.*, 1979: Severe thunderstorm evolution and mesoscyclone structure as related to tornadogenesis. *Mon. Weather Rev. 107*, 1184–1197.

Lilly, D.K., 1986: The structure, energetics, and propagation of rotating convective storm. Part I: Energy exchange with the mean flow. *J. Atmos. Sci. 43*, 113-125.

Lilly, D.K., 1986: The structure, energetics, and propagation of rotating convective storm. Part II: Helicity and storm stabilization. *J. Atmos. Sci. 43*, 126-140.

Mahrt, L. and *Ek, M.*, 1984: The influence of atmospheric stability on potential evaporation. *J. Clim. Appl. Meteorol. 23*, 222-234.

Mahrt, L. and *Pan, H.L.*, 1984: A two-layer model of soil hydrology. *Bound.-Lay. Meteorol. 29*, 1-20.

Noilhan, J. and *Planton, S.*, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev. 117*, 536-549.

Oncley, S.P. and *Dudhia, J.*, 1995: Evaluation of surface fluxes from MM5 using observations. *Mon. Weather Rev. 123*, 3344-3357.

Orville, H.D. and *Kopp, F.J.*, 1977: Numerical simulation of life history of a hailstorm. *J. Atmos. Sci. 34*, 1596-1618.

Pan, H.L. and *Mahrt, L.*, 1987: Interaction between soil hydrology and boundary-layer developement. *Bound.-Lay. Meteorol. 38*, 185-202.

Penman, H.L., 1948: Natural evaporation from open water, bare soil and grass. *Proc. Roy. Soc. London A193*, 120-145.

Pielke, R.A., *Cotton, W.R.*, *Walko, R.L.*, *Tremback, C.J.*, *Lyons, W.A.*, *Grasso, L.D.*, *Nicholls, M.E.*, *Moran, M.D.*, *Wesley, D.A.*, *Lee, T.J.*, and *Copeland, J.H.*, 1992: A comprehensive meteorological modeling system - RAMS. *Meteorol. Atmos. Phys. 49*, 69-91.

Reisner, J., *Rasmussen, R.M.*, and *Bruintjes, R.T.*, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Q. J. Roy. Meteor. Soc. 124*, 1071-1107.

Rotunno, R., 1981: On the evolution of thunderstorm rotation. *Mon. Weather Rev. 109*, 577–586.

Rotunno, R. and *Klemp, J.B.*, 1982: The influence of the shear-induced pressure gradient on thunderstorm motion. *Mon. Weather Rev. 110*, 136–151.

Rotunno, R.E., 1981: A three-dimensional numerical model of an isolated thunderstorm. Part II: Dynamics of updraft splitting and mesovortex couplet evolution. *J. Atmos. Sci. 37*, 395-420.

Simpson, J., Westcott, N.E., Clerman, R.J., and *Pilke, R.A.*, 1980: On cumulus mergers. *Arch. Meteorol. Geophys. Bioklimatol.* A *29*, 1-40.

Schlesinger, R.E., 1980: A three-dimensional numerical model of an isolated thunderstorm. Part II: Dynamics of updruft slitting and mesovortex couplet evolution. *J. Atmos. Sci. 37,* 395-420.

Sridhar, V., Elliott, R.L., Chen, F., and *Brtzge, J.A.*, 2002: Validation of the NOAH-OSU land surface model using surface flux measurements in Oklahoma. *J. Geophys. Res. 107,* No. D20, 4418, doi: 10.1029/2001JD001306.

Troen, I. and *Mahrt, L.*, 1986: A simple model of the atmospheric boundary layer: Sensitvity to surface evaporation. *Bound.-Lay. Meteorol. 37*, 129-148.

Tudurí, E., Romero, R., López, L., Garcia, E., Sánchez, J.L., and *Ramis,* C., 2003: The 14 Luly 2001 hailstorm in northeastern Spain: Diagnosis of the meteorological situation. *Atmos. Res. 67-68*, 541-558.

Wilhelmson, R.B. and *Klemp, J.B.*, 1981: A three-dimensional numerical simulation of splitting severe storms on 3 April 1964. *J. Atmos. Sci. 38*, 1581-1600.

Zoltán, Cs. and *Geresdi, I.*, 1984: A one-dimensional steady-state jet model for thunderclouds. *Időjárás 88*, 21-31.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 299–307

Laboratory modeling of atmospheric flow phenomena: Mountain waves

Balázs Gyüre and **Imre M. Jánosi**

Department of Physics of Complex Systems, von Kármán Laboratory for Environmental Flows,
Eötvös Loránd University, P.O. Box 32, H-1518 Budapest, Hungary
E-mails: bazse@ludens.elte.hu; janosi@lecso.elte.hu

(Manuscript received in final form June 22, 2006)

Abstract—Laboratory simulations in water tanks provide an attractive alternative to full-scale field experiments, moreover, they can be utilized to benchmark analytical and numerical calculations. Here we discuss the possibilities and limitations of modeling large scale atmospheric flow in laboratory. As a case study, we describe experiments on quasi two-dimensional mountain wave formation behind obstacles towed through a linearly stratified fluid. Differences between measured wave fields and predictions of linear theories indicate that nonlinear effects are significant in our parameter range. Experiments with a double bell-shaped obstacle revealed that average wave amplitudes at high enough flow velocities are systematically lower than those produced by an isolated obstacle. We attribute this anomaly to the dominance of essential nonlinearities such as strong wave dispersion and resonance effects.

Key-words: dynamical similarity, stratified flow, mountain waves, laboratory experiments, wave superposition

1. *Introduction: Dynamical similarity*

Situations, where an exact solution for the equations of motion can be given, are exceptional in fluid dynamics. Therefore, alternative methods, mostly numerical procedures and laboratory experiments have been developed for elucidating flows that can not be rigorously calculated. A useful starting point for this development is the following question: under what condition do similar flow patterns occur in two geometrically similar arrangements? When such conditions exist and they can be fulfilled, the two flows are said to be dynamically similar (*Tritton*, 1988).

Let us recall the simplest governing equation in non-dimensional form, which is relevant at environmental flow phenomena (constant density and incompressibility are assumed):

$$\frac{d\vec{u}}{dt} = -\frac{1}{Ro}\vec{n} \times \vec{u} - \frac{1}{Ro}\nabla p - \frac{1}{Fr^2}n + \frac{1}{Re}\Delta\vec{u}, \tag{1}$$

where $\vec{u}$ is the dimensionless velocity, $\vec{n}$ is a vertical unit vector, p is the dimensionless pressure, and there are three non-dimensional combinations of characteristic parameters, the Rossby number, Froude number, and Reynolds number expressed as follows

$$Ro = \frac{U}{f_0 L}, \qquad Fr = \frac{U}{\sqrt{gL}}, \qquad Re = \frac{UL}{\nu} \tag{2}$$

where U and L are typical velocity and length scales, f_0 is the Coriolis parameter (~10^{-4} 1/s at Hungary), and ν is the kinematic viscosity. The boundary conditions can be similarly converted by means of the non-dimensional variables. It is easy to see, that if the Rossby, Froude, and Reynolds numbers are the same for two situations, then the solutions are the same and the same flow patterns occur.

It is not evident that the largest scale flow phenomena in the atmosphere and oceans can be successfully modeled in laboratory tanks. This is mainly because the characteristic sizes are enormously dissimilar. The difference between a cyclone (L~1000 km) and its laboratory model (L~10 cm) is 7 orders of magnitude, while the velocity scales are more similar. The widely different length scales usually make impossible to reproduce atmospheric or oceanic Reynolds numbers too, especially when the medium (air or water at ambient temperature) is the same in the experiments. Large Reynolds numbers are attained in special supercooled Helium pressure chambers, but rotation can not be imposed without difficulties for such an equipment. It should be also noted that exact dynamical similarity can not be fulfilled. In the context of ship model testing, for example, shrinking of L requires an increase of U in order to keep the Reynolds number, but a reduction of U is necessary for a constant Froude number. Both constraints can not be satisfied simultaneously.

Fortunately, laboratory modeling is not hopeless. First of all, viscosity can be neglected in most of the interesting situations, apart from narrow boundary layers or direct turbulence studies. Therefore, relatively low Reynolds numbers (~10^3) in the experiments are acceptable. Secondly, the viscous drag and wave drag are usually not coupled for an obstacle surrounded by moving fluid, thus the effects of changing Froude number can be investigated separately. Thirdly, geometric downscaling does not yield to irrealistic speeds. As an example, *Fig. 1* shows a realization of the classical experiment by *Fultz et al.* (1959) to demonstrate baroclinic instability in a rotating tank. When we cool the center by placing ice into the middlemost chamber, thermal convection starts in the second segment by a characteristic velocity of a few mm/s. Thus a rotation speed regulated in the range of 1–60 rpm gives a coverage of Rossby number interval 10^{-3}–10^{-1}, which is highly relevant in geophysical contexts. (We note that other non-dimensional control parameters, the so-called thermal Rossby number and Taylor number, fit to the very experiment even better, for details see e.g., *Phillips* (1963)).

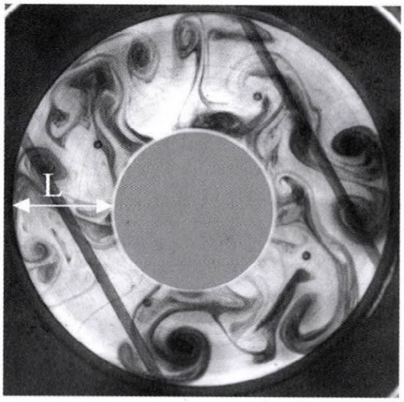


Fig. 1. Thermal convection in a rotating tank of three concentric cylinders. The central chamber is filled with ice, the outermost segment can be heated, the convecting medium in between is water at room temperature (h = 5 cm). The distance between the copper walls is L = 10.5 cm, the actual rotating speed is 10 rpm. (The two inclined stripes at the plexiglass bottom are part of the construction.)

The second example elaborated in the next section is mountain waves in a stratified atmosphere (see *Fig. 2*). The similarity criteria, which must be met in small-scale towing tank experiments have been reviewed by *Baines* and *Manins* (1989). The relevant physical quantities are the towing velocity U, the fluid depth H, the uniform buoyancy (or Brunt-Vaisala) frequency $N = \sqrt{-(g/\rho)\partial_z \rho}$, the maximum obstacle height h, the obstacle half-width w, and the kinematic viscosity ν. These quantities give the dimensionless numbers U/NH, $U/N2w$, and U/Nh related to wave propagation, wave drag, and horizontal perturbation velocity, respectively. By taking typical towing speeds U = 1–15 cm/s, measured buoyancy frequency values in the experiments are N_{exp} = 1.09–1.55 1/s and for the atmosphere N_{atm} = 0.03–0.04 1/s. The matching of dimensionless numbers indicates that our setup simulates atmospheric flow up to a level of 5–10 km at an obstacle height of 600–800 m for uniform wind speed in the range of 10-70 m/s (see *Table 1*). As it is already mentioned, the Reynolds numbers in the experiments ($Re_{\mathrm{exp}} \approx 10^2$–$10^3$) are much smaller than in the atmosphere ($Re_{\mathrm{atm}} \approx 10^6$–$10^9$). Another essential difference is the compressibility of the atmosphere. The air density at the tropopause is ~40% of the surface value, while the corresponding difference in the towing tank filled with salt water can not be larger than a few percents. Furthermore, upward-propagating gravity waves radiate to infinity in the atmosphere, whereas, in a towing tank, they can be reflected from the fluid surface behaving as a rigid lid. Related experimental tests (*Baines*, 1977) indicate that the upper boundary has a substantial effect at relative large obstacle heights $h/H > 0.15$.

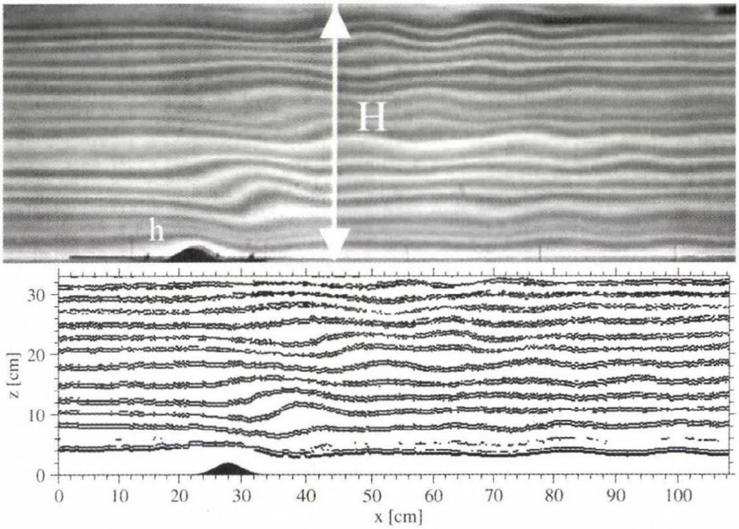


Fig. 2. Top: Wave field behind a moving (from right to left) bell-shaped obstacle in a linearly stratified fluid (salt solution periodically colored by food dye). $H = 32$ cm, $h = 2$ cm, the towing speed is 2.06 cm/s, $N = 1.26$ 1/s. Bottom: Wave field reconstructionby digital image processing.

Table 1. Typical parameter values for the laboratory experiments and meteorological data measured in the winter of 1997/98 around the mountain Pilis

	U	H	h	w	N	U/NH	U/N2w	U/Nh
Experiments	2 cm/s	35 cm	2 cm	2.3 cm	1.3 1/s	0.04	0.33	0.76
Mountain Pilis	20 m/s	10 km	650 m	1 km	0.03 1/s	0.07	0.33	1.02

Many other demonstrations and experiments on stratified and rotating fluids are summarized (in Hungarian) by *Gyüre et al.* (2006).

2. *Case study: Mountain waves in the laboratory*

2.1 Motivation

Internal gravity waves of topographic origin are ubiquitous in the stably stratified atmosphere and oceans (*Nappo*, 2002). Upward propagating lee-waves at around stationary lenticular clouds in mountain ranges were discovered by German glider pilots in 1933 (*Whelan*, 2000). Since that time, wave gliding has widely excercised and became very popular, especially at high altitude attempts: while thermals rarely rise higher than 2–3 km, the most energetic mountain waves can penetrate deeply in the stratosphere. The conditions of exploitable wave generation are a stable and smooth density stratification, proper orography, and steady wind field with minimal shear. Such conditions are fulfilled mainly at two locations in Hungary, in case of appropriate meteorological circumstances: the mountain Kékes (the highest isolated peak in the country), and mountain Pilis (with a maximal elevation of 756 m only). The later area is depicted in *Fig. 3*.

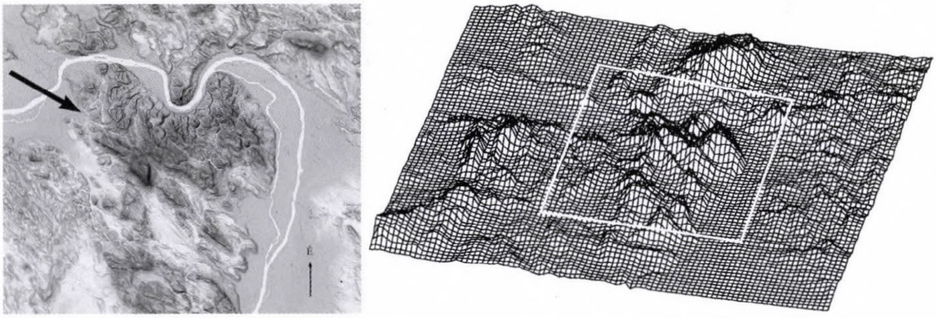


Fig. 3. Topographic settings at around the mountain Pilis (47.7°N; 18.8°E). The wind direction supporting mountain waves is indicated by a heavy arrow. The surface reconstruction on the right shows an area of 90×80 km^2, white frame locates the map borders.

The surface reconstruction in *Fig. 3* illustrates well, that the topography around Pilis is quite complex, it is formed by a series of ridges. This has motivated our experiments on wave field behind a double bell-shaped obstacle realizing the simplest case beyond an isolated, symmetric bump.

Uniform flow over two-dimensional obstacles represents the simplest related model system, which has been studied extensively since the pioneering works of *Lyra* (1943), *Queney* (1948), *Long* (1953, 1955), and *Scorer* (1978). It is not immediately obvious that such a simplified description might have any environmental relevance. *Fig. 4* illustrates, however, that in some cases the lower atmosphere possesses nearly ideal physical properties: almost uniform wind from a constant direction in a stable stratification with an approximately constant Brunt-Vaisala frequency. Indeed, during the winter of 1997/98, wave-gliders succeeded to ascend very high at the lee side of the mountain Pilis. Seven of the flights exceeded the height of 6 km, the best of them was 8250 m (*Kassai*, 1998).

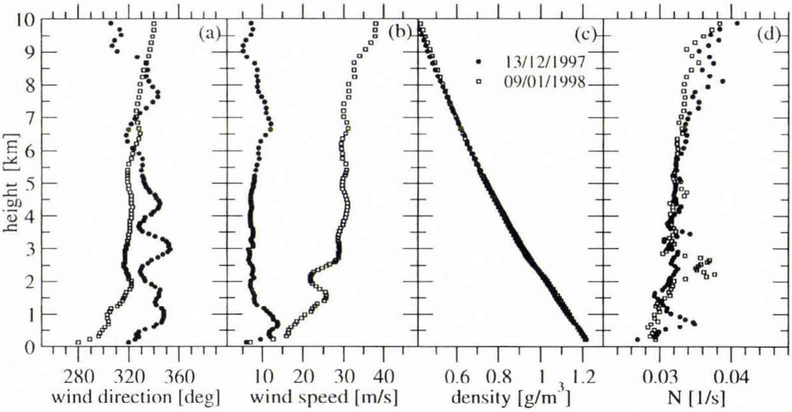


Fig. 4. (a) Wind direction, (b) wind speed, (c) air density, and (d) Brunt-Vaisala frequency profiles for two days in the winter of 1997/98 (indicated in (c)) at a disctance of ~40 kmfrom the mountain Pilis (Budapest-Lőrinc, Hungary). (Brunt-Vaisala frequencies were obtained directly from the density profiles in (c).)

2.2 Experiments and results

Experiments were performed in a plexiglass tank (length 240 cm, width 8.7 cm, height 40 cm) filled with uniformly stratified fluid to a depth of 32–37 cm by the standard double-bucket method (*Fortuin*, 1960). Food dye was periodically added up to the mixture at the filler nozzle resulting in a horizontally layered coloring. Disturbances are generated by towing an obstacle with a tense wire along the bottom of the tank from one end to the other.

In an earlier work, we concentrated on asymmetric obstacles and concluded that the shape of the lee side is the determining factor in wave generation (*Gyüre* and *Jánosi*, 2003). Preliminary experiments with a double bell-shaped obstacle with a peak-to-peak separation of 20 cm revealed that wave superposition is highly nontrivial, therefore, we extended our analysis in this direction. Methodology, wave field characterization, initial transients, etc., is described in detail by *Gyüre* and *Jánosi* (2003).

Fig. 5 shows a comparison of wave patterns behind a single and double obstacles, the other parameters (filling height, stratification, and towing speed) are the same. The waves always move with the same velocity as the obstacle, thus they manifest standing patterns in a co-moving frame of reference. A direct visual check clearly indicates that the pattern behind the double bumps is not a simple superposition of two wave fields produced by an isolated hill. It is also apparent that a quantitative characterization of such waves is quite complicated. Their shape is far from being a simple harmonic function, wave breaking, formation of rotors, and various distortions are prevalent. An approximate description is attempted by extracting average wave amplitudes (vertical distance between consecutive minima and maxima) and average wave lengths (horizontal separation between consecutive extremes belonging to the same streakline). An example is shown in *Fig. 6*.

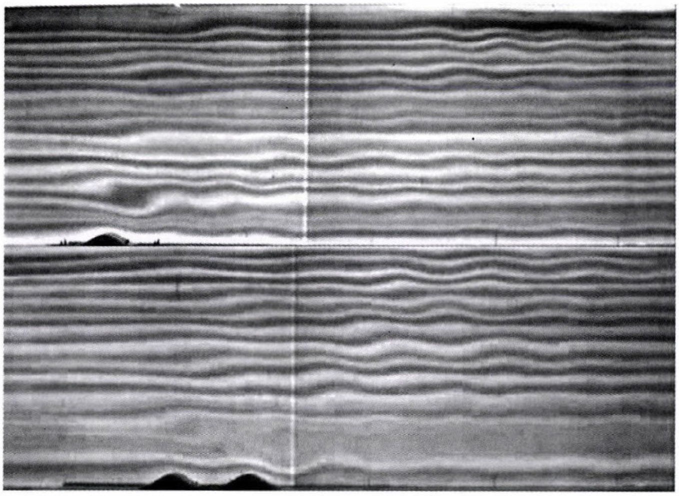


Fig. 5. Top: Wave field behind a single bell-shaped obstacle towed from right to left in a linearly stratified salt solution, U = 1.53 cm/s, N = 1.26 1/s, H = 37.5 cm, h = 2.0 cm, w = 2.6 cm, Gaussian form. Bottom: Wave field behind a double bell-shaped obstacle fabricated from two identical Gaussian bumps (the same as above) joined with a separation of 12.0 cm, U = 1.58 cm/s, the other parameters are the same as above.

Theories predict that the flow is linear if *U/Nh* is sufficiently large, nonlinearities (steepening, wave breaking and mixing, columnar disturbances, etc.) become increasingly important as *U/Nh* drops beneath unity (*Baines*, 1995). This is clearly indicated in *Fig. 6b*, where the scatter of amplitude values makes any conclusion very difficult. Data points above each other belong to the same towing speed but to different heights, unfortunately the dependence is not monotonous. Some indication of constructive interference with the double obstacle might be present in the range $0.55 < U/Nh < 0.75$ (larger amplitudes than for the isolated obstacle). In the linear regime, a breakdown of wave amplitudes is more pronounced for the double obstacle (*Fig. 6a*).

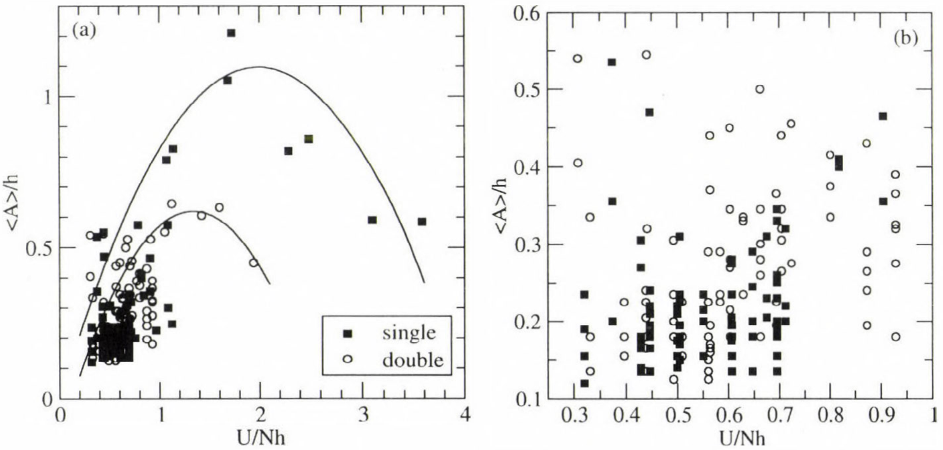


Fig. 6. (a) Normalized amplitudes as a function of dimensionless towing velocity for the single (heavy squares) and double (empty circles) obstacles. Solid lines only guide the eyes. (b) The same as (a), zoomed to the bottom left corner (nonlinear regime, see text).

3. *Closing remarks*

Atmospheric flows over mountains and hills contain a rich variety of phenomena, many of which occur on scales which are unresolved by numerical weather prediction models. These phenomena include turbulent wakes, the occurrence of flow separation on the lee slope and, when the flow is stably-stratified, gravity-wave generation, severe downslope wind storms, and lee vortex shedding. A better understanding of these flows will allow significant improvements to local weather forecasting, especially for aviation.

We have illustrated that laboratory experiments on orographic flows in stratified fluids can provide useful information complementing field observations and numerical modeling. Such experiments are strongly idealized and mimic simple situations, still they are able to reveal the limitations of model computations and help to find correct interpretations of measured meteorological data.

Acknowledgements—We thank *Tamás Tél* for thoughtfull discussions. This work was supported by the Hungarian National Science Foundation (OTKA) under grants T047233 and TS044839. *Imre M. Jánosi* is grateful for a János Bolyai research scholarship of the Hungarian Academy of Sciences.

References

Baines, P.G., 1977: Upstream influence and Long's model in stratified flows. *J. Fluid. Mech. 82*, 147-159.

Baines, P.G., 1995: *Topographic Effects in Stratified Flows*. Cambridge Univ. Press, Cambridge.

Baines, P.G. and *Manins, P.C.*, 1989: The principles of laboratory modeling of stratified atmospheric flows over complex terrain. *J. Appl. Meteorol. 28*, 1213-1225.

Fortuin, J., 1960: Theory and application of two supplementary methods of constructing density gradient columns. *J. Polymer Sci. 44*, 505-515.

Fultz, D., Long, R.R., Owens, G.V., Bohan, W., Kaylor, R., and *Weil, J.*, 1959: Studies of thermal convection in a rotating cylinder with some implication for large-scale atmospheric motions. *Meteorol. Monographs 4*, 1-104.

Gyüre, B. and *Jánosi, I.M.*, 2003: Stratified flow over asymmetric and double bell-shaped obstacles. *Dynam. Atmos. Oceans 37*, 155-170.

Gyüre, B., Jánosi, I.M., Szabó, K.G., and *Tél, T.*, 2006: Environmental flow phenomena I., II. (in Hungarian). *Légkör 51*, No. 1, 6-12; No. 2, 6-11.

Kassai, B., 1998: Wave soaring in Dunakeszi (in Hungarian). *Pilóta,* No. 5, 4-6.

Long, R.R., 1953: Some aspects of the flow of stratified fluids. I. A theoretical investigation. *Tellus 5*, 42-58.

Long, R.R., 1955: Some aspects of the flow of stratified fluids. III. Continous density gradients. *Tellus 7*, 341-357.

Lyra, G., 1943: Theorie der stationaren Leewellenströhmung in freien Atmosphäre. *Z. Angew. Math. Mech. (Berlin) 23*, 1-28.

Nappo, C.J., 2002: *An Introduction to Atmospheric Gravity Waves*. Academic Press, SanDiego, London.

Phillips, N.A., 1963: Geostrophic motion. *Rev. Geophys. 1*, 123-176.

Queney, P., 1948: The problem of airflow over mountains: A summary of theoretical studies. *B. Am. Meteorol. Soc. 29*, 16-25.

Scorer, R.S., 1978: *Environmental Aerodynamics*. Ellis Horwood, Chichester.

Tritton, D.J., 1988: *Physical Fluid Dynamics*. Oxford University Press, Oxford.

Whelan, R.F., 2000: *Exploring the Monster. Mountain Lee Waves: The Aerial Elevator*. Wind Canyon Books, Brawley.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3-4, December 2006, pp. 309–327

Development of a variational data assimilation system for a limited area model at the Hungarian Meteorological Service

Gergely Bölöni

Hungarian Meteorological Service
P.O. Box 38, H-1525 Budapest, Hungary; E-mail: boloni.g@met.hu

(Manuscript received in final form June 23, 2006)

Abstract—This paper aims to give a detailed description of the three dimensional variational (3DVAR) data assimilation system developed for the Hungarian version of the ALADIN model (ALADIN/HU). The evaluation of the system's performance will be given through different kind of verification results, and the most important developments related to the design of the assimilation cycle and the background error covariance modeling will be presented. Recently, after a long period of preliminary testing, the ALADIN/HU 3DVAR system has become an operational application at the Hungarian Meteorological Service, which makes possible to take the benefit of local high resolution observations while providing the initial conditions for the production forecast. The evaluation of the system is based on comparisons with a former operational version of the ALADIN/HU model, in which the production forecast was simply initialized by an appropriate interpolation of the analysis provided by the ARPEGE global model.

Key-words: limited area modeling, variational analysis, data assimilation cycle, background error covariance

1. Introduction

In a limited area model (LAM) framework, the possibility to prepare the local initial conditions is given without applying sophisticated data assimilation methods (*Daley*, 1991; *Lorenc*, 1986), namely through an interpolation of the driving model's initial conditions to the limited area model grid. This solution, often called as dynamical adaptation, is widely used due to its simplicity and low computational costs, however, it assumes that the initial conditions of the driving model are appropriate to describe the important meteorological features to be evolved with the forecast model. While increasing the resolution of the LAM, the accuracy of such a simple solution is probably questionable as the spatial density of the observations included into the driving model (which is a global model in most of the cases) is not surely satisfactory for representing the processes, which are in the focus of the applied LAM. Moreover, as the resolution jump between the driving model and the driven LAM is increasing, the interpolation of the driving model's analysis will provide noise in the small scale spectrum of the LAM initial conditions, where the driving model is not able to describe physically realistic processes.

On the contrary, in a local data assimilation system one can take the benefit of non-GTS (Global Telecommunication System) high resolution local observations, which are available in an increasing amount according to the recent year's experience. The small scale part of the analysis will also be more correctly provided in a local data assimilation procedure through the use of high resolution LAM field as background. Consequently, a natural direction for improving LAM models is to implement local data assimilation systems for the generation of the initial conditions. Indeed, the development of data assimilation systems started in the recent decades for several LAM models, such as HIRLAM (High Resolution Limited Area Model) (*Gustaffson et al.*, 2001; *Lindskog et al.*, 2001), unified model of the UK Met Office, RUC (Rapide Update Cycle) (*Dévényi* and *Benjamin,* 2003).

The first attempts to develop a variational data assimilation scheme (*Courtier et al.*, 1998; *Bouttier* and *Courtier*, 1999) for the ALADIN (Aire Limitée Adaptation Dynamique Developpement International) limited area model (*Horányi et al.*, 1996) date back to 1996, when the already working variational assimilation method implemented in the ARPEGE (Action de Recherche Petite Echelle Grande Echelle) model was adapted in this LAM. The ALADIN 3DVAR local data assimilation scheme was implemented at the Hungarian Meteorological Service (HMS) during the summer of the year 2000, and the first daily runs with a simple version using only SYNOP and TEMP data were started in 2001. The regular runs were continued till the spring of 2005 on a test basis including more and more developments relating the observation use and some other aspects of the method. The local 3DVAR data assimilation system has been used operationally since May, 2005, after a careful validation and cross comparison with the previously used operational dynamical adaptation system (*Bölöni,* 2005). Hereafter, the most important characteristics of the presently used operational assimilation system will be described. Verification results will also be summarized in order to demonstrate the system's performance, and finally an overview of the related developments will be given. A similar description of the 3DVAR system used in the French version of the ALADIN model is given in *Fischer et al.* (2005).

2. *Main characteristics*

The ALADIN/HU operational model domain covers continental Europe (*Fig. 1*). The presently used horizontal resolution is 8 km with a linear spectral truncation. At the time of writing this paper, the model uses 49 vertical levels between the surface and the top of the model (5 hPa) with an increased resolution in the planetary boundary layer.

Fig. 1. The ALADIN/HU operational model domain.

2.1 The assimilation cycle setup

The strategy of keeping the independence of the assimilation cycle from the so-called production suite is taken from the global modeling experience (*Fig. 2*). This solution implies that for each network times, whenever a production forecast is run, two analyses are provided with different data cut-off times. On the one hand, an analysis is performed with short data cut-off in order to provide the initial conditions for the production run fitting the operational time constraints. On the other hand, another analysis is provided for the same network time with a longer data cut-off, which is used in the assimilation cycle as initial condition for the next background forecast. The reason for repeating the above-mentioned analyses with long data cut-off in the assimilation cycle is to provide the best possible analyses in the cycle through the use of all the available observations. The will to provide the best possible analysis is much reasonable if one keeps in mind that in a data assimilation cycle, the information in the analysis is always evolved by the background forecast, which means that possible errors are also cumulated in time. In the ALADIN/HU 3DVAR a 6-hour cycling is applied, which means the realization of 4 long cut-off analyses per day at 00:00, 06:00, 12:00, and 18:00 UTC. Production forecasts are provided for 48 hours twice a day at 00:00 and 12:00 UTC (*Fig. 2*).

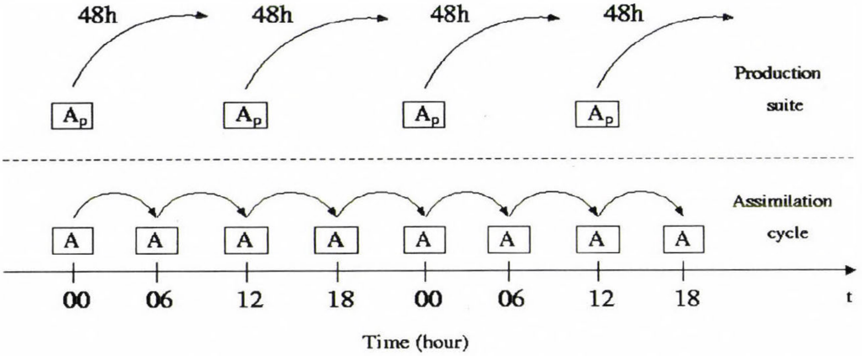


Fig. 2. The ALADIN/HU 3DVAR assimilation cycle. "A" stands for analysis with long data cut-off, "A_p" stands for "production" analysis with short data cut-off.

The upper-air analysis is provided by the local 3DVAR run, which uses the actual local observations and the background 6-hour forecast originating from the previous step of the assimilation cycle. The initial conditions for the soil scheme are taken from the global analysis of the ARPEGE model by an interpolation to the ALADIN/HU grid. Each analysis is initialized by a digital filtering (*Lynch* and *Huang,* 1992) before running the background or production forecasts, in order to get rid of the spurious fast propagating waves implemented by the assimilation procedure. Both the background and production forecasts are using the model fields provided by the corresponding ARPEGE model runs as lateral boundary conditions with a 3-hour update frequency. Between the 3-hour updates, a time interpolation is used in order to provide the necessary information at every time step. For coupling the background forecasts in the assimilation cycle, the ARPEGE analyses are used at those network times when they are available rather than the ARPEGE forecasts, in order to apply a boundary forcing to the model state, that is possibly the closest to the reality. As the analysis fields of the ARPEGE model are available at every 6 hours, they can be used at network times 00:00, 06:00, 12:00, and 18:00 UTC only, which means that at 03:00, 09:00, 15:00, and 21:00 UTC 3-hour ARPEGE forecasts have to be used as lateral boundary conditions.

2.2 The background term

The background term in the ALADIN 3DVAR scheme is generally very similar to that is implemented in the ARPEGE and IFS global models (*Derber* and *Bouttier*, 1999), however, it was further adjusted in order to take into account the background errors of small scale moist processes by including a multivariate coupling of the humidity errors with those of mass, wind, and temperature (*Berre*, 2000). The computation of the background error covariance matrix is based on the statistical balance approach (*Derber* and *Bouttier*, 1999), and the covariance matrix describes isotropic and homogeneous horizontal correlations depending on the height (*Berre*, 2000). It may be interesting to note here, that recently intensive research has been started at the French and Belgian services in order to include realistic anisotropy and inhomogeneity into the horizontal correlation structures (*Deckmyn* and *Berre*, 2005). Considering the fact that the ALADIN/HU 3DVAR system is a specific local application of the commonly developed ALADIN 3DVAR scheme, the above characteristics are valid for the operational system run at the Hungarian Meteorological Service as well. For completing the general description it should be added, that in the operational version of the ALADIN/HU 3DVAR the NMC (National Meteorological Center) method (*Parrish* and *Derber*, 1992) is used for sampling the background errors at the moment. In a later section of this paper, local developments at HMS related to the background term will be briefly described, with an emphasis on the error sampling.

2.3 Observational data

A very brief description of the operational observation use will be given below. For more details on observation impact studies and research on observation processing *Randriamampianina* (2006) is referred. The ALADIN/HU 3DVAR operational system presently uses surface, radiosonde, satellite, and aircraft observations. The table below helps to summarize all the observed parameters by observation type that are used in the system.

Table 1. Observational data entering the ALADIN/HU assimilation system

Observation type	**Variable**	**Horizontal density**
SYNOP surface reports	Surface pressure	~20–50 km
TEMP upper air reports	Temperature, wind, geopotential, specific humidity	~200–250 km
ATOVS satellite observations	AMSU-A radiances	80 km thinning
AMDAR aircraft reports	Temperature, wind	25 km thinning

It is important to emphasize, that all the observation types above are used in the global 4DVAR assimilation system of the ARPEGE model as well. However, the local 3DVAR assimilation system benefits from some useful additional observational input coming from local non-GTS SYNOP reports, and denser use of satellite and aircraft measurements due to weaker thinning in the quality control (*Randriamampianina,* 2005; *Randriamampianina et al.*, 2004).

3. Meteorological evaluation

This section summarizes the verification results of the ALADIN/HU 3DVAR system including a detailed comparison with the former operational version of the model, which was a dynamical adaptation version of the model (ALADIN/HU DYA). The evaluation of the models at the NWP (Numerical Weather Prediction) team of the HMS consists of two main components, namely the computation of objective scores (BIAS and RMSE) with respect to observations and a subjective evaluation (*Tóth,* 2004), which aims to value the most critical weather parameters used by the forecasters.

3.1 Objective scores

Score comparisons for several periods have been carried out between the 3DVAR and the previously operational DYA versions of the ALADIN/HU model. The model forecasts were verified against the observations coming from about 50 radiosond and 700 surface stations over Europe. The synthesis of the comparisons will be presented with illustrations highlighting the most important features. Improvement has been found in the temperature and wind fields on all vertical levels due to the local 3DVAR assimilation (*Fig. 3*).

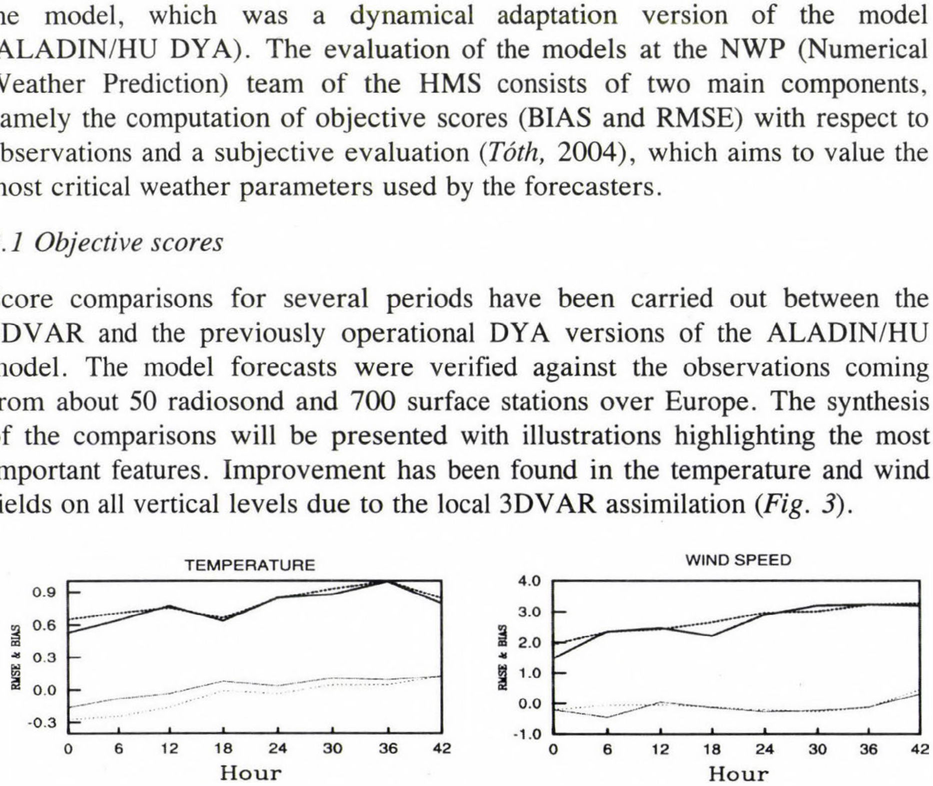


Fig. 3. Evolution of RMSE (thick lines) and BIAS (thin lines) scores according to the forecast range. Left: temperature on 500 hPa, right: wind speed on 500 hPa. Solid line: ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

The geopotential scores are also better for the local assimilation system on the high atmospheric levels, however, near the surface the impact of the assimilation is slightly negative especially regarding systematic errors, which also results in the degradation of the sea level pressure BIAS scores (*Fig. 4*). The impact on humidity is mixed depending on the forecast range and the height. Near the surface, a systematic error present in the DYA system is corrected by the 3DVAR assimilation, but in the middle-troposphere scores are variable with the forecast range (*Fig. 5*). There is a degradation of the forecasts using the 3DVAR system near the tropopause level (not shown).

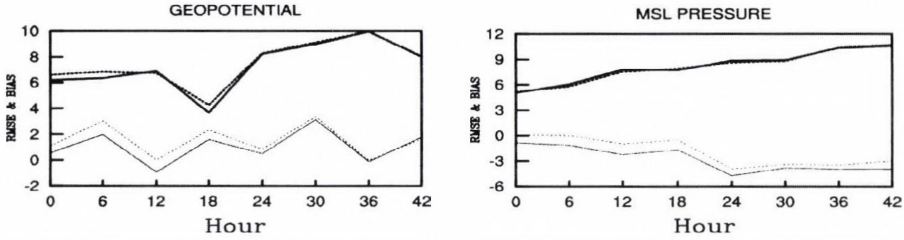


Fig. 4. Evolution of RMSE (thick lines) and BIAS (thin lines) scores with the forecast range.
Left: geopotential on 700 hPa, right: sea level pressure.
Solid line: ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

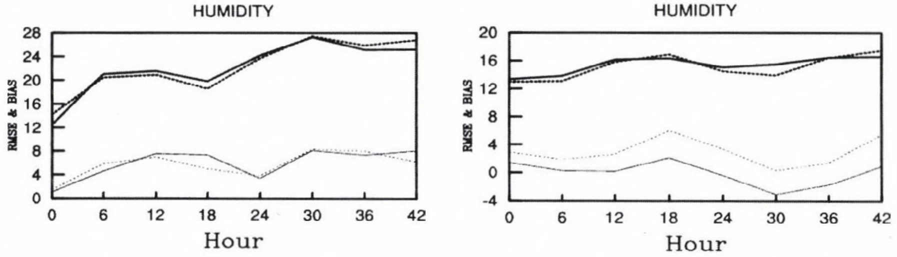
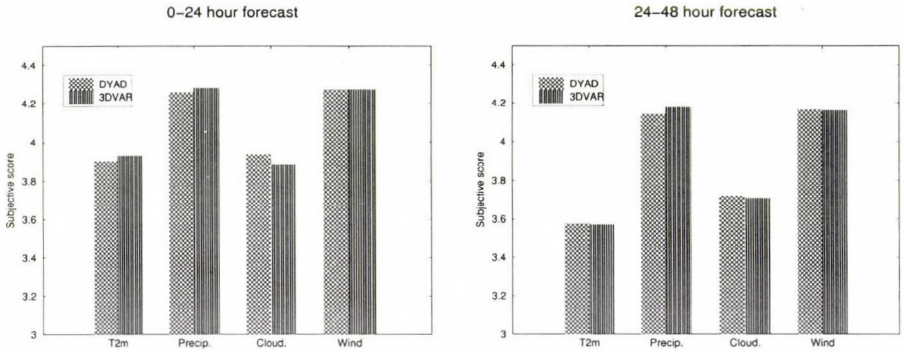


Fig. 5. Evolution of RMSE (thick lines) and BIAS (thin lines) scores with the forecast range.
Left: humidity on 500 hPa, right: humidity on the surface (2 m).
Solid line: ALADIN/HU 3DVAR, dashed line: ALADIN.

3.2 Subjective verification

As the objective scores are based on the comparison of forecasted model values with synoptic scale observations, they are not accurate enough to represent the reliability of the model on the small spatial scales. Consequently, a subjective evaluation of the models was also implemented at the NWP group of the HMS in order to follow the performance of the different model versions in the every day forecasting practice, giving also the opportunity to value the models in interesting and extreme weather events. The subjective verification is realized daily together by forecasters and modelers, comparing forecasted weather charts (precipitation, 2-meter temperature, cloudiness, wind) with the observations. The quality of the forecast is represented by a number between 1 and 5 given subjectively by the verification team according to the performance of the model (1 is the worst and 5 is the best score). Conclusions of the subjective verification are summarized below based on a half-year continuous comparison (*Fig. 6*). The 2-meter temperature forecast was improved by the implementation of the local data assimilation especially for the first day. The precipitation forecasts are also more successful within the 3DVAR system especially for the short ranges (0–18 hours), but an improvement was shown for the whole 2-day forecast as well. The impact of the local assimilation on the wind forecast is rather neutral, while a degradation of the cloudiness forecast encountered as a disadvantage of the local 3DVAR system, especially for the short range.



0–24 hour forecast

24–48 hour forecast

Fig. 6. Subjective verifications scores for different weather parameters summarized over a half-year period (July 1–December 31, 2004).

In *Fig. 7*, the day-to-day evolution of the precipitation subjective scores is presented for further illustration. It is reflected by the figure, that there is no big difference between the two model versions tested, however, when the two versions diverge, the 3DVAR system proves to be better in most of the cases.

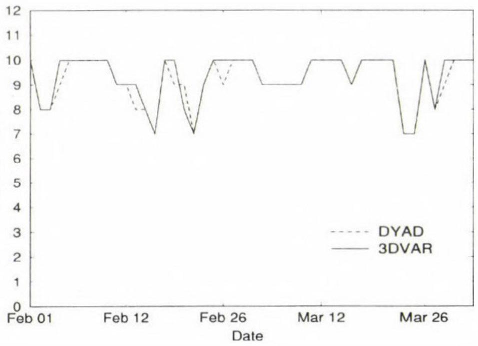


Fig. 7. Evolution of subjective verification scores for the precipitation short range (0–24 hours) forecasts over the period of February 1 – April 5, 2005. Scores are multiplied by a factor of 2. Full line: ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

3.3 A case study

On May 18, 2005 an intensive mesoscale convective system was passing over Hungary. It included severe thunderstorms, strong wind gusts (> 100 km/h), and heavy precipitation (~45 mm/24 h) in several places over the country (*Horváth,* 2005). In this weather situation, the 3DVAR assimilation system performed much better than the dynamical adaptation, forecasting more realistic precipitation structures compared to the radar observations and predicting accurately the evolution of the weather system. In *Fig. 8*, one can compare the precipitation charts predicted by both model versions under question. The corresponding radar image is attached aiming to represent the reality. The comparison of observed and forecasted 6-hour cumulated precipitations also indicated a better performance of the local assimilation system (not shown). This case was selected because of the interesting and severe weather situation and we aimed to confirm the good skills of the 3DVAR system in precipitation prediction its presentation is. However, the detailed study of further characteristic weather cases will be needed in order to discover more in depth the capabilities of this system.

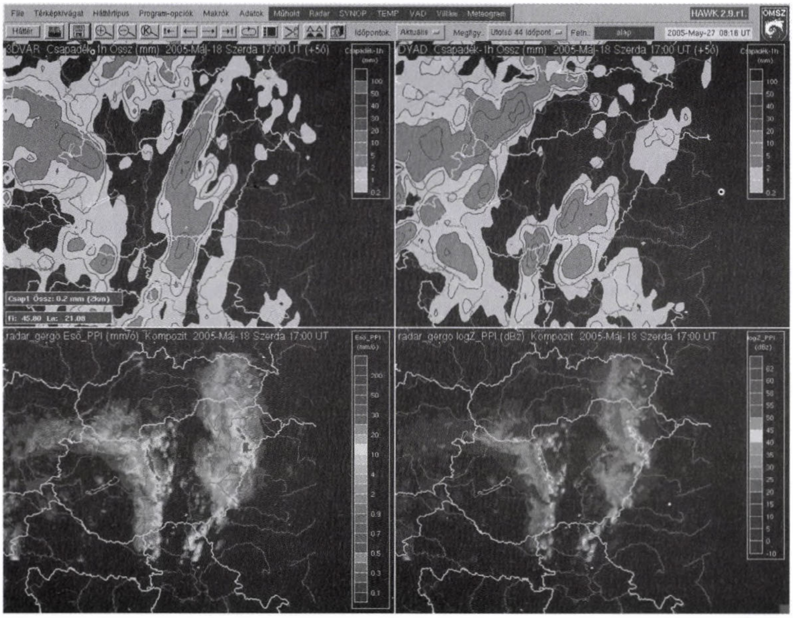


Fig. 8. Precipitation forecasts of the 3DVAR (top left) and DYA (top right) versions of the ALADIN/HU model. For comparison with the reality, the radar observations are shown, namely precipitable water (bottom left) and logZ (bottom right).

4. Ongoing developments

Beside the inclusion of new observation types to the system, developments related to the design of the data assimilation cycle and the background error covariance matrix are in the scope of the modeling group of HMS. This section briefly presents the most important steps that were taken in these two fields.

4.1 Design of the assimilation cycle

The design of a LAM assimilation cycle raises new questions compared to a global one. Beside the question of the length of the assimilation window and initialization method to be chosen, a strategy for the lateral boundary coupling has to be proposed for the background forecast.

For the ALADIN/HU 3DVAR several lateral boundary coupling strategies were tried, such as single and double nesting (*Vasiliu* and *Horányi,* 2005) or the choice of the lateral boundary fields regarding if it is a forecast or analysis of the driving model. Following the main conclusions of the investigation above, a single nesting with the ARPEGE model was chosen for the ALADIN/HU 3DVAR, first with a 6-hour coupling update frequency using always the ARPEGE long cut-off analyses in order to force the background forecast on the boundaries towards the most reliable state available. Further experiments showed that by increasing the coupling update frequency from 6 hours to 3 hours, the background forecasts are improved (*Fig. 9*). According to these results, the 3-hour frequency was introduced to the operational ALADIN/HU 3DVAR suite as described in the previous section already.

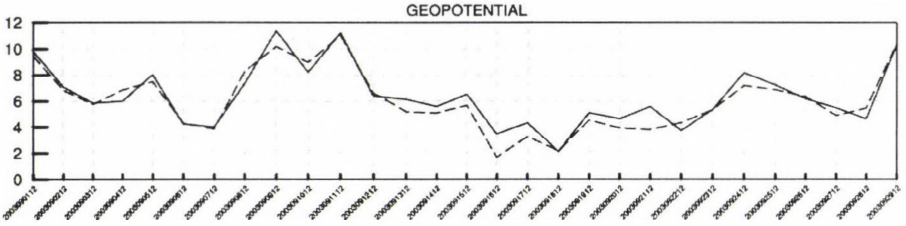


Fig. 9. Day-to day evolution of 6-hour forecast RMSE scores for geopotential on 850 hPa over a selected period (September 21–29, 2003). Solid line: 6-hour update frequency, dashed line: 3-hour update frequency.

One also can consider special analysis solutions, where the LAM analysis concentrates only on the local small scales, while the large synoptic scales are controlled by the initial conditions of the driving model. This can be achieved through a scale dependent combination of the driving model's analysis and the LAM background forecast. Different methods, such as the blending through digital filtering (*Brozková et al.*, 2001), the variational blending (*Guidard*, 2003), and the explicit spectral blending (*Tóth*, 2003) are known to carry out the scale dependent combination of model fields mentioned above. It was shown in several studies, that such a scale dependent blending is beneficial in specific weather situations (*Alexandru,* 2003). It is worthy to mention, that the success of blending is probably related to the fact that the driving model's (ARPEGE) assimilation method is a 4 dimensional (4DVAR) one, which is more accurate than a 3DVAR, especially regarding the representation of non-linear features in the initial conditions.

At the HMS, a technically simple method, the explicit spectral blending has been implemented and tested. This method blends the model fields in Fourier spectral space. Due to the fact that the ALADIN model is a spectral one, which uses spatial Fourier representation of the meteorological fields, the spectral coefficients needed for the spectral blending are naturally given. The blending is done in a transition interval, which is defined by two threshold Fourier wave numbers. Under the lower threshold wave number K_1 (where waves correspond to the largest scales), the ARPEGE fields, while above the upper threshold wave number K_2 (where waves correspond to the smallest scales), the ALADIN fields will be taken into account fully (*Fig. 10*). In the transition interval a smooth transition between the ARPEGE and ALADIN spectral fields is ensured through a linear combination of the spectral coefficients of the models. Considering the local implementation at the HMS, the threshold wave number K_1 was set to zero, so that the ALADIN fields are used even on the largest scales, and K_2 was set to the wave number, above which the ARPEGE model does not contain physically meaningful information due to its coarser resolution.

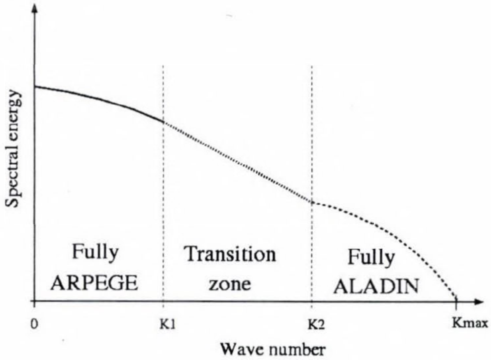


Fig. 10. The method of the explicit blending (explanation in the text).

The method was tested and verified for longer periods and specific interesting meteorological cases. In *Fig. 11* we highlight a spectacular case, where a strong inversion situation was kept better by the blended analysis than with the simple local 3DVAR. The success of the blending method compared to the simple 3DVAR assimilation in this case was due to the fact, that the 6-hour background forecast of ALADIN failed to predict the inversion situation, while it was successfully analyzed in the ARPEGE 4DVAR. In terms of objective scores, slight improvement has been shown while applying the explicit blending for the selected period chosen in the study (not shown). Further tests with the combination of 3DVAR and the explicit blending are planned in the future aiming an operational implementation in case of further convincing results.

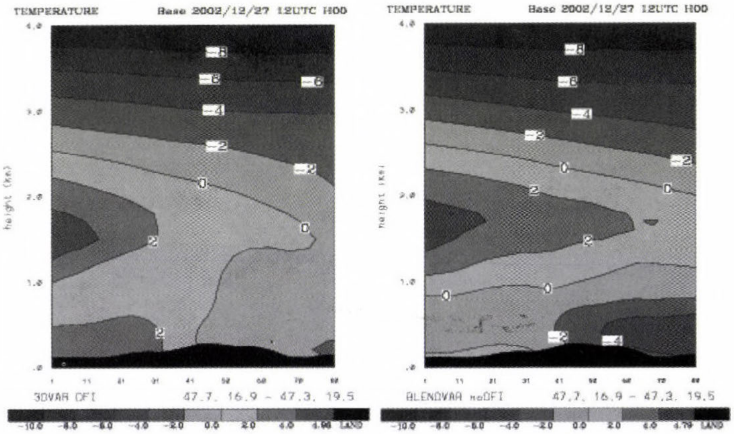


Fig. 11. Vertical cross sections of temperature fields obtained by the 3DVAR analysis without (left) and with (right) explicit blending.

4.2 Developments related to the background error covariance matrix

The development of the background error covariance matrix is one possible way of improving the efficiency and quality of the analysis, as this matrix determines both the amplitude and shape of the analysis increments generated by the assimilation of observations (*Daley*, 1991). This field of work has always been lively within the ALADIN 3DVAR developer team. Different aspects, like development of the statistical balance (*Berre*, 2000), implementation of non-homogeneous anisotropic structure functions (*Deckmyn* and *Berre*, 2005), error sampling (*Siroká et al.*, 2003; (*Stefanescu et al.*, 2005), and tuning (*Sadiki* and *Fischer*, 2004; *Desroziers et al.*, 2005) issues, have been investigated at different meteorological centers taking part in the ALADIN cooperation. At the Hungarian Meteorological Service, the related work consisted of mostly error sampling and tuning, and a few studies on the isotropy properties of the background error covariance matrix.

Considering the error sampling, first a sensitivity study of the so-called lagged- NMC method was carried out in Budapest. The lagged-NMC method was developed by *Siroká et al.* (2003), which is an alternative version of the NMC method modified specifically for an appropriate LAM use, through the separation of the errors originated by the lateral boundary conditions from those of generated due to the initial uncertainties. This separation is achieved through the application of the same lateral boundary forcing in the model runs, which provide the NMC forecast differences. As a consequence of the identical lateral boundary conditions, the forecast differences do not represent the errors, which are due to the nesting into the driving model but those only, which are due to the initial uncertainties. The above-mentioned sensitivity study of the lagged-NMC method aimed to find the most appropriate forecast range pair to be used for creating the forecast differences for sampling the background errors of the ALADIN/HU model. In the frame of the study, 28 different background error covariance matrices were computed upon the actual operational model runs, using different forecast range combinations for creating the error sample. The background error statistics proved to be sensitive to the forecast range combinations used in the study. One of the most interesting features found is that the total error variance is decreasing if the forecast ranges taking part in the lagged-NMC differences are increased (*Fig. 12*). It is true even if the difference between the initial conditions, i.e., the degree of initial uncertainty, is kept constant. This latter is typically true if the time shift between the initialization of the model runs providing the forecast differences are the same.

The result above suggests that the error due to the initial uncertainties is decreasing with an increasing forecast range. This looks contradictory to the well-known feature of monotonous error growth in numerical models. The reason for this result may come from the fact, that a LAM is not a closed system in the context that the initialized air mass fields leave the forecast domain at a certain time after the analysis. Over this time range the initial errors will not contribute to the model errors any more. On the other hand, the errors due to the lateral boundary forcing grow without limitation considering the forecast range, so the total error of the LAM will still grow, too. Another important outcome of the study is related to the comparison of the lagged-NMC method with the standard one. Namely, it was found that the actual version of the ALADIN/HU model was too much influenced by the lateral boundary conditions. This was proved by the fact, that the error variance was strongly reduced even on the small scale spectra, if the lagged-NMC method was applied (*Fig. 13*). This reflects that even the smallest scales represented in the model were strongly determined by the lateral boundary forcing. The strong influence of the lateral boundary conditions was probably related to the small difference in the resolutions and domain sizes of the LAM[1] and the driving model[2].

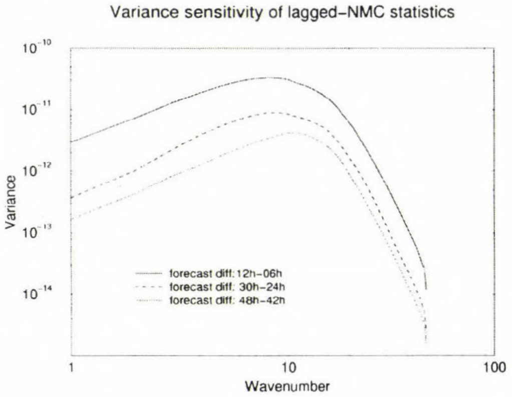


Fig. 12. Distribution of the spectral variance with respect to the wave number for different forecast range pairs used for the computation of lagged-NMC background error statistics of the ALADIN/HU model (divergence error variance at level 13).

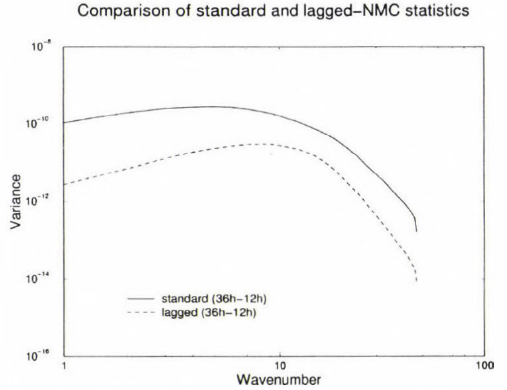


Fig. 13. Distribution of the spectral variance with respect to the wave number for standard and lagged-NMC background error statistics of the ALADIN/HU model (divergence error variance at level 13).

---

[1] Former ALADIN/HU model: 8 km resolution over Central Europe

[2] Former ALADIN/LACE model: 12 km resolution over the continental Europe

Another work on the error sampling and tuning consisted of the comparison of the NMC method with the innovation method proposed by *Hollingsworth* and *Lönnberg* (1986). In this study the background error statistics computed, applying the different sampling techniques, have been compared. More exactly, error variances and horizontal correlations have been calculated using both the standard and lagged-NMC methods and compared by the corresponding statistics computed with the innovation technique. Error variances obtained by the innovation method were generally found to be somewhat smaller, while horizontal correlations were found to be larger than those of computed with the NMC methods. Large correlations of the innovations are the most characteristic for the planetary boundary layer (*Horvath,* 2004). *Fig. 14* shows an example for the comparison of horizontal correlations calculated with the different error sampling methods.

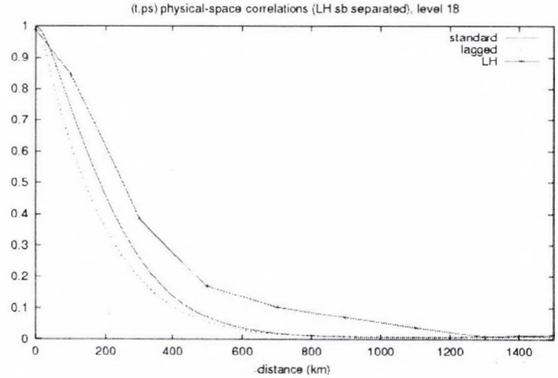


Fig. 14. Comparison of horizontal correlations obtained by the NMC and innovation methods. Standard: standard NMC method, lagged: lagged-NMC method, LH: innovation method.

The humidity error variances, obtained by the innovation method, were also used for the retuning of the NMC humidity variances at a later stage of the study. The goal of the tuning was to improve the multivariate coupling between humidity and the other control variables (mass, wind, and temperature) through the adjustment of the standard deviations of the humidity background errors. As the standard deviations obtained with the NMC method were found to be too big from the point of view of the multivariate balance, they were replaced between 850 and 300 hPa by those of calculated with the innovation technique in this study. However, in single observation experiments the improvement of the multivariate balance was shown due to the tuning, rather neutral results were found in real assimilation cycling experiments.

In the frame of a recent study, the background error statistics of the ALADIN/HU model have been computed using the ensemble technique for sampling the errors. The ensemble technique is similar to the NMC one, from the point of view that forecast differences are used to represent the background errors, but here the differences are made by subtracting the members of a special forecast ensemble. The ensemble is defined by a set of data assimilation cycles, each originating from the same first guess at the beginning of the cycling but using perturbed observations at each assimilation step. This method proved to represent the background errors in a more realistic way than the NMC method does in the global models (*Fisher,* 2003; *Belo Pereira* and *Berre,* 2005). For the representation of the background errors of the ALADIN/HU model, an ensemble provided by the ARPEGE model has been downscaled by running the LAM forced by the ensemble of the ARPEGE lateral boundary conditions. Diagnostic comparison of the ensemble and NMC statistics has been done so far, and assimilation experiments are on the way in order to assess the impact of the ensemble statistics on the analyses and forecasts. The most important conclusions of the diagnostic comparison are that the ensemble background error statistics consist of sharper vertical correlations (*Fig. 15*) and smaller variances than those of provided by the standard NMC method. Correlation length scales are also reduced using the ensemble technique compared to what is given by the standard NMC method, except for humidity on the highest levels.

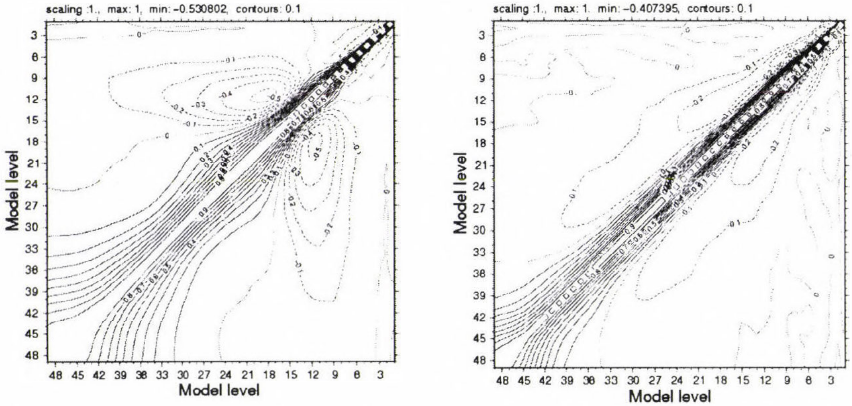


Fig. 15. Vertical temperature correlations computed upon the NMC sample (left) and the ensemble sample (right).

These comparison results are encouraging, considering that the standard NMC method is overestimating the error variances and correlations according to previous studies. Consequently, in case of good performance, in assimilation cycling experiments the background error statistics computed using ensemble method are planned to be used operationally in the ALADIN/HU 3DVAR. Similarly designed exhaustive experiments have also been carried out at Meteo-France by *Stefanescu et al.* (2005).

5. Conclusions

An operational application of the ALADIN 3DVAR scheme was presented in this paper. A local data assimilation system based on this scheme was implemented already five years ago in the Hungarian version of the ALADIN model (ALADIN/HU). It has been run in an experimental manner until the spring of 2005, when it has been implemented operationally.

The ALADIN/HU 3DVAR system has been briefly described concerning the most important issues on the local implementation. Verification results have been presented in order to value the performance of the ALADIN/HU 3DVAR system compared to the previous operational version of the ALADIN/HU model (dynamical adaptation). The most important activities related to the development of the system have also been briefly summarized.

According to the verification results, the implementation of a local data assimilation system, such as the ALADIN/HU 3DVAR, is beneficial in the every day forecasting of the most important weather parameters such as precipitation and 2-meter temperature. The prediction for most of the upper-air variables is improved as well due to the local data assimilation, however, weaknesses of the ALADIN-HU 3DVAR system do exist and they should not be left out of account. Namely, the problem of loose forecasts of humidity on very high levels and the systematic errors in the prediction of sea level pressure will have to be understood and corrected in the future.

As a very short and simple outlook for the future of the ALADIN/HU 3DVAR system, the NWP group at the HMS will keep continue its development, including as many new observation types as possible (MSG, radar wind and reflectivity) and improving the variational assimilation scheme with a special emphasis on the background error statistics. The development of a 4DVAR assimilation is also in the long term plans of the ALADIN assimilation team. As a first step, work has just started with the implementation of a 3D-FGAT (First Guess at Appropriate Time) scheme at Meteo-France and HMS. 3D-FGAT can be considered as an intermediate method between the 3D- and 4DVAR schemes. It makes possible to benefit more accurately on the observations available in high temporal frequency, such as aircraft, satellite, and hourly SYNOP data for instance. So far, a research version of 3D-FGAT was implemented and its validation is on the way at the moment.

Acknowledgements–The author would like to thank the work of all the people who helped the development of the ALADIN/HU 3DVAR system in the frame of research stays or discussions during the last five years. Especially, many thanks to *Loïk Berre* and *Claude Fischer* for useful advices concerning the background error covariance modeling, and to *Maria Derková* and *Kristian Horvath* for important contributions to the related developments. The work of *Roger Randriamampianina, Sándor Kertész, Michal Majek, Miklós Balogh,* and *Regina Szoták* on the observation treatment is also highly appreciated as well as the exhaustive tests and case studies of *Steluta Vasiliu* and *Helga Tóth*. The author is also grateful to *András Horányi* for useful suggestions related to the text. This work was supported by the Hungarian National Office for Research and Technology (JÁP. grant No. 2/007/2005) and the Hungarian National Scientific Foundation (OTKA T049579).

References

Alexandru, S., 2003: Scientific strategy for the implementation of a 3D-Var data assimilation scheme for a double-nested limited area model. *ALADIN Newsletter,* No. 25, 23-28.

Belo Pereira, M. and *Berre, L.*, 2005: The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev. 133,* 2466-2489.

Bouttier, F. and *Courtier, P.*, 1999: Data assimilation concepts and methods. *ECMWF Lecture Series (Data assimilation and use of satellite data)* (http://www.ecmwf.int/newsevents/training/lecture_notes/LN_DA.html) 72 pp.

Bölöni, G., 2005: Operational implementation of ALADIN 3DVAR at the Hungarian Meteorological Service. *ALADIN Newsletter,* No. 28, 60-66.

Berre, L., 2000: Estimation of synoptic and meso scale forecast error covariances in a limited are amodel. *Mon. Weather Rev. 128,* 644-667.

Brožková, R., Klaric, D., Ivatek-Sahdan, S., Geleyn, J-F., Cassé, V., Siroká, M., Radnóti, G., Janousek, M., Stadlbacher, K., and *Seidl, H.*, 2001: Dfi blending: an alternative tool for preparation of initial conditions of LAM. *CAS-JSC WGNE Report* No. 30, WMO.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and *Fisher, M.*, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Q. J. Roy. Meteor. Soc. 124,* 1783-1808.

Daley, R., 1991: *Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series.* Cambrige University Press, 457 pp.

Deckmyn, A. and *Berre, L.*, 2005: A wavelet approach to representing background error covariances in a limited area model. *Mon. Weather Rev. 133,* 1279-1294.

Derber, J. and *Bouttier, F.*, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus 48A,* 501-517.

Desroziers, G., Berre, L., Chapnik, B., and *Poli, P.*, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Q. J. Roy. Meteor. Soc. 131,* 3385-3396.

Dévényi, D. and *Benjamin, S.*, 2003: A 3-dimensional atmospheric variational assimilation technique in a hybrid isentropic-sigma coordinate. *Meteorol. Atmos. Phys. 82,* 245-254.

Fischer, C., Montmerle, T., Berre, L., Auger, L., and *Stefanescu, S.*, 2005: An overview of the variational assimilation in the Aladin/France numerical weather prediction system. *Q. J Roy. Meteor. Soc. 131,* 3477-3492.

Fisher, M., 2003: Background error covariance modelling. In *ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean* (http://www.ecmwf.int/publications/library/do/references/list/17111), 45-63.

Guidard, V., 2003: Evaluation of assimilation cycles in a limited area model. Introduction of a large-scale cost-function in a LAM 3d-var. *ALADIN Newsletter,* No. 25, 31-32.

Gustaffson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascués, B., Mogensen, K.S., and *Thorsteinnson, S.*, 2001: Three-dimensional variational data assimilation for a limited area model. Part I: General formulation and the background error constraint. *Tellus 53A*, 425-446.

Hollingsworth, A. and *Lönnberg, P.*, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I. The wind field. *Tellus 38A*, 111-136.

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

Horvath, K., 2004: Comparison of NMC and LH type of background error statistics of the ALADIN/HU model. *RC LACE Internal Report* (http://www.rclace.eu/?page=11), 18 pp.

Horváth, Á., 2005: Meteorological description of the storm May 18, 2005 (in Hungarian). *Légkör,* No.3, 12-16.

Lindskog, M., Gustaffson, N., Navascués, B., Mogensen, K.S., Huang, X.-Y., Yang, X., Andrae, U., Berre, L., Thorsteinsson, S., and *Rantakokko, J.*, 2001: Three-dimensional variational data assimilation for a limited area model. Part II: Observation handling and assimilation experiments. *Tellus 53A*, 447-468.

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc. 114*, 205-240.

Lynch, P. and *Huang, X.-Y.*, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Weather Rev. 120*, 1019-1034.

Parrish, D.F. and *Derber, J.C.*, 1992: The National Meteorological Centre's spectral statistical interpolation system. *Mon. Weather Rev. 120*, 1747-1763.

Randriamampianina, R., 2005: Radiance-bias correction for a limited area model. *Időjárás 109,* 143-155.

Randriamampianina, R., 2006: Impact of high resolution observations in the ALADIN/HU model. *Időjárás 110,*

Randriamampianina, R., Csima, G., and *Szoták, R.*, 2004: Assimilation of the AMDAR data in the ALADIN 3D-Var system. *ALADIN Newsletter,* No. 25, 76-82.

Sadiki, W. and *Fischer, C.*, 2004: A posteriori validation applied to the 3D-VAR ARPEGE and ALADIN data assimilation system. *Tellus 57A,* 21-34.

Siroká, M., Fischer, C., Cassé, V., Brozková, R., and *Geleyn, J-F.*, 2003: The definition of mesoscale selective forecast error covariances for a limited area variational analysis. *Meteorol. Atmos. Phys. 82*, 227-244.

Stefanescu, S.E., Berre, L., and *Belo Pereira, M.*, 2005: The evolution of dispersion spectra and the evluation of model differences in an ensemble estimation of error statistics for a limited area analysis. Accepted in *Mon. Weather Rev.*

Tóth, H., 2003: Further blending experiments & Studying of LH-statistics. *RC LACE Internal Report* (http://www.rclace.eu/?page=11), 21 pp.

Tóth, H., 2004: Subjective evaluation of different versions of ALADIN/HU model. *ALADIN Newsletter,* No. 26, 88-89.

Vasiliu, S. and *Horányi, A.*, 2005: An evaluation of the performance of the three-dimensional variational data assimilation scheme for the ALADIN/HU spectral limited area model. *Időjárás 109*, 233-257.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 329–347

Impact of high resolution satellite observations in the ALADIN/HU model

Roger Randriamampianina

Hungarian Meteorological Service
P.O. Box 38, H-1525 Budapest, Hungary; E-mail: roger@met.hu

(Manuscript received in final form July 13, 2006)

Abstract—The Hungarian Meteorological Service (HMS) has contributed to the development of a high resolution limited area model (LAM) in the frame of the ALADIN project since its beginning (1990). The development of the data assimilation system started at the HMS in the year 2000 with the implementation of the three-dimensional variational (3D-Var) analysis scheme. Our research aims to design an optimal assimilation system suitable for LAM application, including various high resolution observations. This paper describes the configuration of the analysis and forecast systems used in our studies. Results of the incorporation of the AMSU-A and AMSU-B data in full resolution - one-by-one field of view (FOV) - are presented. Studies regarding the efficient radiance-bias correction were necessary to get improvement from the Advanced Microwave Sounder Units A and B (AMSU-A and AMSU-B) of the Advanced TIROS Operational Vertical Sounder (ATOVS) radiances in our LAM analysis and forecast systems. Small, but positive impact of the high resolution ATOVS radiances on our analysis and short-range forecasts was obtained, which leaded to their operational implementation.

Key-words: ALADIN limited area model, data assimilation, satellite radiances, ATOVS/AMSU-A, ATOVS/AMSU-B

1. Introduction

The aim of the ALADIN project[1] was to develop in collaboration a modern numerical weather prediction (NWP) system for use on limited geographic area. The advantage of such a system is that it uses a moderate computing power, while allowing a zoom effect with respect to the French global model ARPEGE.

---

[1] The ALADIN project was proposed by the Météo-France to the National Meteorological Services of Central and Eastern Europe in 1990.

Currently, scientists from fifteen countries are permanently contributing to the progress of the ALADIN NWP system. Detailed description of the project is available on its webpage: http://www.cnrm.meteo.fr/ALADIN/.

At the beginning, the research was focused on the development of a forecasting system suitable for a limited area model (LAM) using the interpolated fields from ARPEGE analysis as initial condition (*Bubnová et al.*, 1995; *Radnóti*, 1995; *Horányi et al.*, 1996). In 1994, the analysis system of the ARPEGE model, based on optimum interpolation was implemented/ adopted in the ALADIN model (*Ajjaji* and *Issara*, 1994), to create more precise initial conditions for the LAM forecast model taking into account local measurements.

One of the important developments of the ALADIN model was the implementation of the three-dimensional variational analysis system (3D-Var) starting in spring 1997. Since then, many developments have been done in different centers to obtain a better and more reliable system, including the improvement of the representation of the forecast error covariance (*Berre*, 2000; *Široká et al.*, 2003; *Deckmyn* and *Berre*, 2005), detailed model sensitivity analyses (*Soci et al.*, 2006), studies on different nesting strategies (*Vasiliu* and *Horányi*, 2005) and on efficient validation of the 3D-Var system (*Sadiki* and *Fischer*, 2005).

At the HMS, the implementation of the variational analysis system started in 2000. One of our goals regarding research and development with the ALADIN Hungary (ALADIN/HU) model is to increase the amount of observations for a reliable and efficient system. We aim to use all accessible observations in the highest and optimal resolution possible.

Before starting the operational implementation at the HMS (May 2005), a regular daily run of the variational analysis system using surface (SYNOP) and radiosonde (TEMP) observations started in summer 2000. Later on, this experimental system was completed with further observations such as aircraft (AMDAR), wind profiler, and atmospheric motion vectors (AMV). The present operational ALADIN/HU 3D-Var assimilation system and its functionality are described in detail in *Bölöni* (2006). The 3D-Var is in operation at Météo-France from July 2005 and is in test regime in Morocco. A comprehensive overview of the development of the ALADIN/France 3D-Var is described in *Fischer et al.* (2005).

Satellite measurements became more and more important during the last few years. The use of raw satellite observations in the assimilation system of LAMs is a promising challenge to ensure better forecast. The appearance of the ATOVS instruments (providing two new microwave AMSU-A and AMSU-B data) onboard NOAA-15 satellite was an exciting task for the specialists to incorporate these observations into the data assimilation system.

English et al. (2000) found remarkable improvement of the analysis and medium-range forecast of a global model when implementing advanced microwave observations. Although the assimilation of the AMSU-B data was problematic at the beginning, now these data represent one of the most important elements of the global variational systems (*English et al.*, 2003; *Chouinard* and *Hallé*, 2003; *Gérard et al.*, 2003). Detailed investigation has been performed to evaluate the impact of the AMSU-B data in a limited area model (*Jones et al.*, 2002; *Candy*, 2005). These studies showed positive impact on the analysis of moisture and short-range forecast of precipitation. There is, however, still a lack of knowledge on the use of ATOVS data in limited area models.

This paper describes the incorporation of the ATOVS (AMSU-A and AMSU-B) radiances into the ALADIN/HU analysis system. Section 2 describes the main characteristics of the ALADIN/HU model used in the studies. Source and the pre-processing of the satellite data are shown in Section 3. Sections 4 and 5 discuss the impact studies related to the AMSU-A and AMSU-B data, respectively. In Section 6 some conclusions are drawn.

2. The ALADIN/HU models and the assimilation system used in the study

At the HMS, the ALADIN/HU model runs in its hydrostatic version. Different versions of the ARPEGE/ALADIN codes were used in the investigations, including model configurations *C1* and *C2* as described in *Table 1*. The run in parallel suite[2] discussed in this paper was performed with the model configuration *C3,* which is the operational model at the HMS. The three-dimensional variational data assimilation system was applied to assimilate both conventional (surface, radiosonde, and aircraft) and satellite (ATOVS) observations. As a consequence of the direct radiance assimilation, it is necessary to simulate radiances from the model parameters. The RTTOV (*Table 1*) radiative transfer code, which has 43 vertical levels, was used to perform this transformation (*Saunders et al.*, 1998) in the ARPEGE/ALADIN models. Above the top of the model, an extrapolation of the profile is performed using a regression algorithm (*Rabier et al.*, 2001). Below the top of the model, profiles are interpolated to RTTOV pressure levels. The background error covariance matrix is computed using the standard NMC method (*Parrish* and *Derber*, 1992; *Berre*, 2000; *Široká et al.*, 2003). An optimal interpolation scheme was used to analyze the surface fields (*Ajjaji* and *Issara*, 1994). The

---

[2] From the continuous development, the model configuration that is found to be better than the one used in the operational system is tested in real time. This additional run is called parallel suite.

3D-Var is running in 6-hour assimilation cycle generating an analysis at 00:00, 06:00, 12:00, and 18:00 UTC. In this study, 48-hour forecasts were performed daily from 00:00 (AMSU-A study) and 12:00 (AMSU-B study) UTC.

Table 1. The ALADIN/HU 3D-Var applied in the investigations

System	Configurations	Configuration 1 (C1)	Configuration 2 (C2)	Configuration 3 (C3)
Model	Hydrostatic version Horizontal resolution Vertical levels form surface up to 5 hPa	al25/cy24t1 12 km 37	Al28/cy28t3 12 km 37	al28/cy28t3 8 km 49
3D-Var	Covariance matrix B: std NMC 6-hour assim. cycling RTM model: RTTOV Coupling files: ARPEGE long cut-off files Available satellite observations Selected channels Humidity assimilation	RTTOV-6 Coupling: every 6 h NOAA-15&16 AMSU-A AMSU-A (5-12) Multivariate and univariate	RTTOV-7 Coupling: every 3 h NOAA-15,16&17 AMSU-A&B AMSU-A (5-12), AMSU-B (3-5) Multivariate	RTTOV-7 Coupling: every 3 h NOAA-15,16&17 AMSU-A&B AMSU-A (5-12), AMSU-B (3-5) Multivariate
O.I.	Surface analysis	Yes	No, interpolation of ARPEGE surface fields to the ALADIN grid	No, interpolation of ARPEGE surface fields to the ALADIN grid
Forecast	48-hour	Once a day	Once a day	Once a day

3. Source and pre-processing of the observations

The Advanced Microwave Sounding Unit-A (AMSU-A[3]) system is implemented in two separate modules: AMSU-A1 and AMSU-A2. This is a cross-track, line-scanned instrument designed to measure scene radiances in 15 discrete frequency channels, which permit the derivation of the vertical temperature profile from the surface of the Earth up to a pressure of about 3 hPa (45 km). Thirty contiguous scene resolution cells are sampled in a stepped-scan fashion

---

[3] http://www.wmo.ch/web/sat/en/ap10-10.htm; http://www2.ncdc.noaa.gov/docs/klm/html/c3/sec3-3.htm

every eight seconds, each scan covering 50 degrees on each side of the sub-satellite path. These scan patterns translate to a 45 km diameter cell at nadir and a total swath width of about 2100 km from the 840 km nominal orbital altitude.

The Advanced Microwave Sounding Unit-B (AMSU-B[4]) is a five-channel microwave sounder developed by the UK Meteorological Office for flight on the NOAA-15, -16, and -17 satellites. The purpose of the AMSU-B instrument is to receive and measure radiation from a number of different layers of the atmosphere in order to obtain global data on humidity profiles. It works in conjunction with the AMSU-A instruments to provide a total of 20 microwave channels for atmospheric sounding. At the microwave frequencies used, clouds are almost transparent, while rain and snow are strong emitters, so the instrument is also used to map precipitation. AMSU-B carries channels 16 to 20 (later AMSU-B channels 1 to 5, respectively). The highest channels: 18, 19, and 20, span the strongly opaque water vapor absorption line at 183 GHz and provide data on the atmosphere's moisture content. Channels 16 and 17, at 89 GHz and 150 GHz, respectively, enable deeper penetration through the atmosphere to the Earth's surface.

Because of the high variability of atmospheric water, AMSU-B have a higher resolution than that of AMSU-A, with a circular field of view having a diameter of about 16 km at nadir. Ninety of these are measured in each cross-track scan. The instrument has the same swath-width as AMSU-A, but scans across track in one third of the time in order to keep the two instruments synchronized. By this means, arrays of 3×3 AMSU-B samples overlay each AMSU-A sample, facilitating synergistic use of these instruments.

The ATOVS data are received through a HRPT antenna and pre-processed with the AAPP (ATOVS and AVHRR Pre-processing Package) software package. In this study, ATOVS, level 1-C radiances were used. For technical reasons the antenna is able to receive data only from two different satellites. To acquire the maximum amount of satellite observations, the NOAA-15 and NOAA-16 satellites were chosen, which have orbits perpendicular to each other (over the Earth poles) and pass over the ALADIN/HU domain at about 06:00 and 18:00 UTC, and 00:00 and 12:00 UTC, respectively. In addition to our local reception, data pre-processed at the EUMETSAT and retransmitted through the EUMETCast broadcasting system, that contain data measured by NOAA-17 were investigated.

---

[4] http://www.wmo.ch/web/sat/en/ap10-11.htm; http://www2.ncdc.noaa.gov/docs/klm/html/c3/sec3-4.htm

3.1 Radiance-bias correction

Direct assimilation of satellite measurements requires the correction of the biases computed as differences between the observed radiances and those simulated from the model first guess. These biases arise mainly from instrument characteristics, but inaccuracies in the radiative transfer model can also be significant. The method developed by *Harris* and *Kelly* (2001) was used to remove this systematic error. This scheme is based on separation of the biases into scan-angle and state dependent components. The air-mass (state dependent) bias is expressed as a linear combination of a set of state-dependent predictors. According to *Randriamampianina* (2005), in case of ALADIN model, it is recommended to compute local bias correlation coefficients instead of using those generated for the global model ARPEGE. Four predictors: p1 – the 1000–300 hPa thickness, p2 – the 200–50 hPa thickness, p3 – the skin temperature, and p4 – the total column water were used in the scheme.

4. Investigation of the AMSU-A data

In the 3D-Var ALADIN/HU, the AMSU-A data were investigated first. Consequently, the results discussed in this section refer to the configuration *C1* of the ALADIN/HU model.

4.1 Use of the ATOVS/AMSU-A data

Analyzing the bias of the brightness temperature specific for each AMSU-A channel, we decided to keep the same number of channels as they were used in the global ARPEGE model (*Table 2*). The percentage of land over the ALADIN/HU domain is more than 70, so our study also concerns the use of AMSU-A data over land. NOAA-15 has problem with AMSU-A channel 11, which is not used in our system.

Table 2. AMSU-A channels used in the ALADIN/HU. *– over land channels 5 and 6 are used, where the model topography is less than 500 m and 1500 m, respectively

Conditions	**Channel number**														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Over land					**X***	**X***	X	X	X	X	X	X			
Over sea					X	X	X	X	X	X	X	X			
Over sea ice							X	X	X	X	X	X			
Cloudy pixels								X	X	X	X	X			

4.1.1 Observation statistics and quality of the system

The quality of the locally received and pre-processed level 1D (ATOVS clear) radiances in the ARPEGE 4D-Var was found to be slightly different from the quality of the 1D radiances, distributed by NESDIS (*Randriamampianina* and *Rabier*, 2002). Thus, for the locally pre-processed radiances the observation errors in the assimilation scheme had to be divided by 1.5 for the efficient use, while for the NESDIS radiances they remained the same. Although the level 1C AMSU-A radiances were in operation in the 4D-Var analysis system at Météo-France, we found important to check the efficiency of the assimilation of these data in the ALADIN/HU 3D-Var. The first-guess departures (differences between the observation (AMSU-A radiances) and guess (computed radiances)) were compared with the analysis increments (differences between the observation and analysis) for this purpose (*Fig. 1*). This figure shows the statistics computed for a few days (February 20–25, 2003) cycling using the system configuration *C1*. The distance between the two curves indicates how the addition of the AMSU-A data could modify the first-guess fields during the assimilation. The larger the distance, the bigger the impact of the observation (so, of the AMSU-A data) on the analysis. These results are comparable to those reported by *Randriamampianina* and *Rabier* (2002). At 00:00 and 12:00 UTC we have data from NOAA-16, while at 06:00 and 18:00 UTC – data from NOAA-15. We got similar statistics for the other assimilation times.

Another test consisted of reproducing the above mentioned experiment after reducing the predefined observation error by half before starting the experiments. We did not find any considerable changes in the results. So, at this stage, we decided to keep the original values of the observation error, as used in the ARPEGE model.

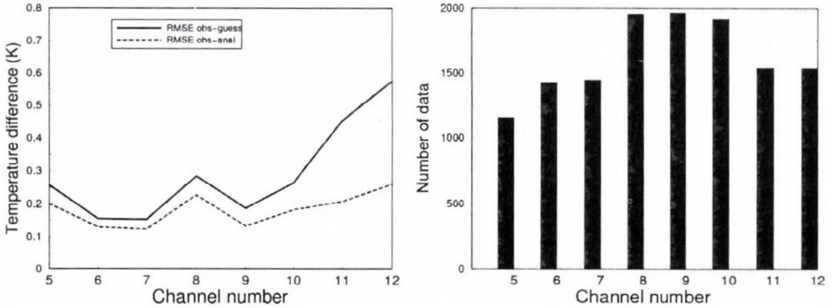


Fig. 1. The statistics of the first-guess departures (solid line) and analysis increments (dashed line) for the AMSU-A channels at 12:00 UTC for a five-day cycling (February 20–25, 2003) (left), and the number of data used in the computation (right).

4.2 Impact of the AMSU-A data

In the global 4D-Var ARPEGE analysis system, the ATOVS radiances are assimilated in 250 km horizontal thinning distance. Our goal is to use the local observations as fine resolution as possible. In the experiments, two thinning distances (80 km and 120 km resolution) were investigated. The impact of the AMSU-A data was studied for a two-week period (February 20 – March 06, 2003). In the control run, the surface and radiosonde observations were assimilated. The impact of the AMSU-A radiances was evaluated comparing the control run with runs, where these data were added in the assimilation. The scores of each run were evaluated objectively. The bias and root-mean-square error (RMSE) were computed from the differences between the analysis/ forecasts and observations (surface and radiosondes), as well as analysis/ forecasts and long cut-off ARPEGE analyses. Significance tests of the objective verification scores were also performed. The significance was examined based on statistical t-test regarding the difference in the expected values of the RMSE scores of the compared experiments. Plots were provided together with error bars that represent the interval, in which the RMSE difference falls within 90% confidence. Consequently, we considered a difference to be significant if the corresponding error bar did not include the zero difference line. In the comparison the first model (usually the test model) was better than the second one (usually the control model) if the mean score was negative, indicating an average reduction of the error.

4.2.1 Influence of the assimilation of AMSU-A data on temperature and humidity bias

It was found, that the AMSU-A data have a cooling effect in the troposphere (*Fig. 2*).

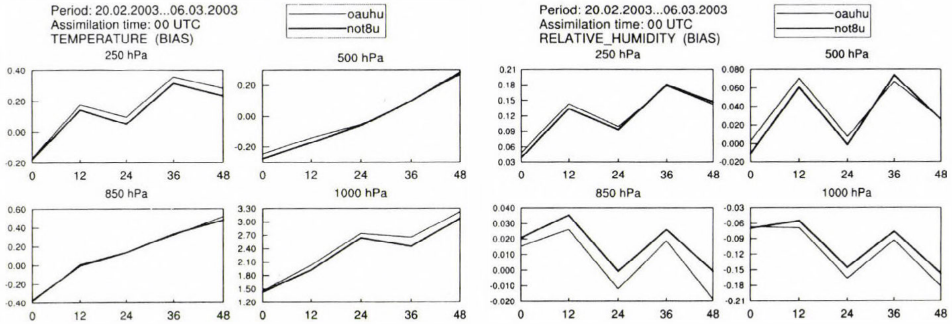


Fig. 2. Temperature and relative humidity biases for the runs with (not8u: bold line) and without (oauhu: thin line) AMSU-A data at the analysis (0) and subsequent forecast times (12- to 48-hour). The bias was computed from differences against observations.

This effect is getting stronger with an increasing forecast range, especially in both extremes (near the top and the bottom) of the troposphere reducing considerably the bias at these levels. The impact is negligible in the middle troposphere. A drying effect of the assimilation of the AMSU-A data in the upper and a wetting effect in the lower troposphere were observed.

4.2.2 Significance test of the impact of AMSU-A data on the analysis and short-range forecasts

The impact of the AMSU-A data on the geopotential was found slightly positive. *Fig. 3* shows a significant reduction of the RMSE for a half-day forecast. The impact was neutral in the lower troposphere.

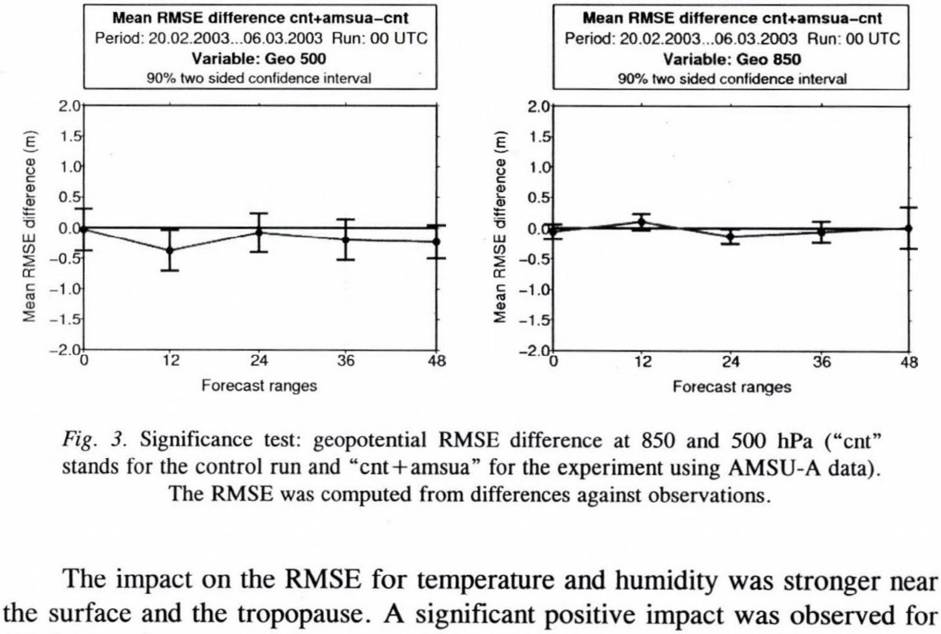


Fig. 3. Significance test: geopotential RMSE difference at 850 and 500 hPa ("cnt" stands for the control run and "cnt+amsua" for the experiment using AMSU-A data). The RMSE was computed from differences against observations.

The impact on the RMSE for temperature and humidity was stronger near the surface and the tropopause. A significant positive impact was observed for the longer forecast ranges (*Fig. 4*), which is in accordance with the above discussed (Section 4.2.1) impact of the AMSU-A data on the temperature and humidity bias.

The positive impact was somewhat stronger in general, when AMSU-A data were assimilated at finer resolution (80 km, compared to 120 km) (*Fig. 5*).

The comparison of the analyses and short-range forecasts against the long cut-off ARPEGE analyses showed slightly different results. The assimilation of the AMSU-A data in the ALADIN/HU 3D-Var improved the forecasts in the lower and middle troposphere (*Fig. 6*). Usually, the impact of the additional data (the AMSU-A data in this case) inside the ALADIN/HU domain is likely to be more accentuated over the eastern part. This is probably due to the relatively less conventional observations over this region compared to the "well observed" western part, where the impact is moderate.

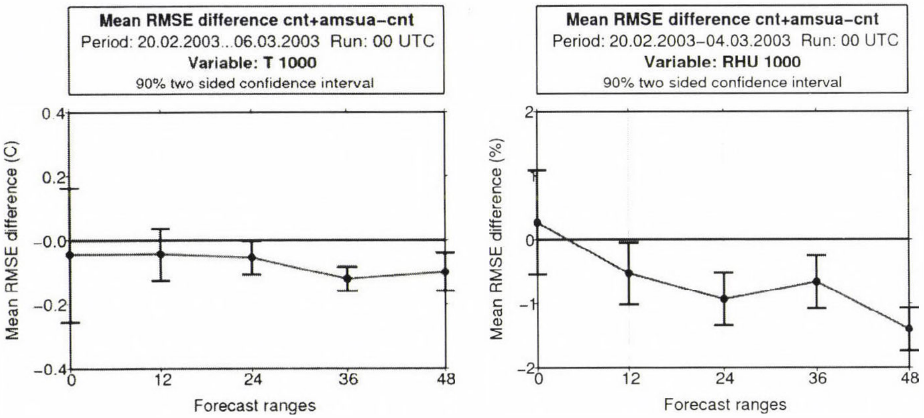


Fig. 4. Significance test: temperature (left) and relative humidity (right) RMSE differences at 1000 hPa ("cnt" stands for the control run and "cnt+amsua" for the experiment using AMSU-A data). The RMSE was computed from differences against observations.

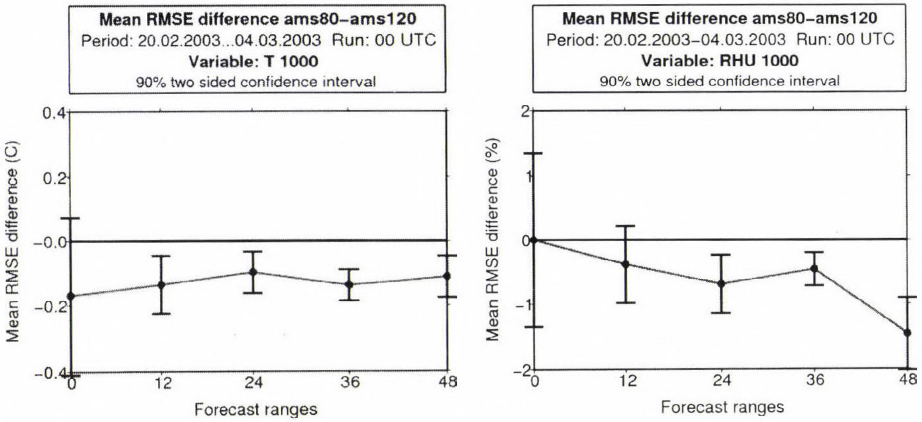
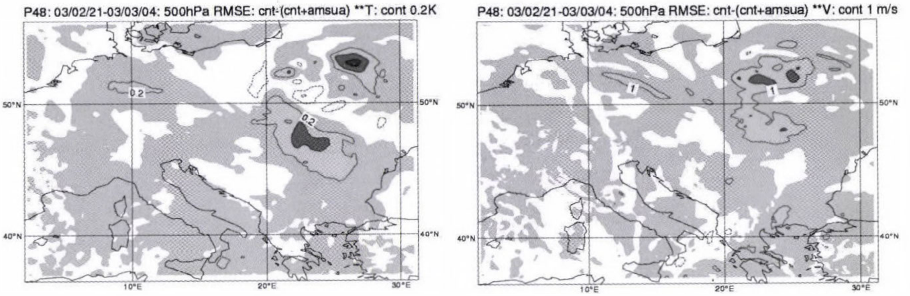


Fig. 5. Significance test: temperature (left) and relative humidity (right) RMSE differences at 1000 hPa ("ams80" stands for 80 km and "ams120" for 120 km thinning distance while using the AMSU-A data). The RMSE was computed from differences against observations.


P48: 03/02/21-03/03/04: 500hPa RMSE: cnt-(cnt+amsua) **T: cont 0.2K

P48: 03/02/21-03/03/04: 500hPa RMSE: cnt-(cnt+amsua) **V: cont 1 m/s

Fig. 6. Objective verification against the long cut-off ARPEGE analyses at 500 hPa for the 48-hour forecast of temperature (left) and wind speed (right). These graphs show the RMSE difference between runs without and with AMSU-A data. Colored areas show the positive impact averaged for a period of two weeks.

4.2.3 Case study

Fig. 7 shows zoomed differences in cumulative precipitation between the runs without (upper left) and with (upper right) ATOVS data at the eastern region of Poland and western part of Byelorussia. According to the real situation (*Fig. 7*, lower picture), there was some precipitation over the mentioned area. One can see that the run with ATOVS data could slightly better describe this situation.

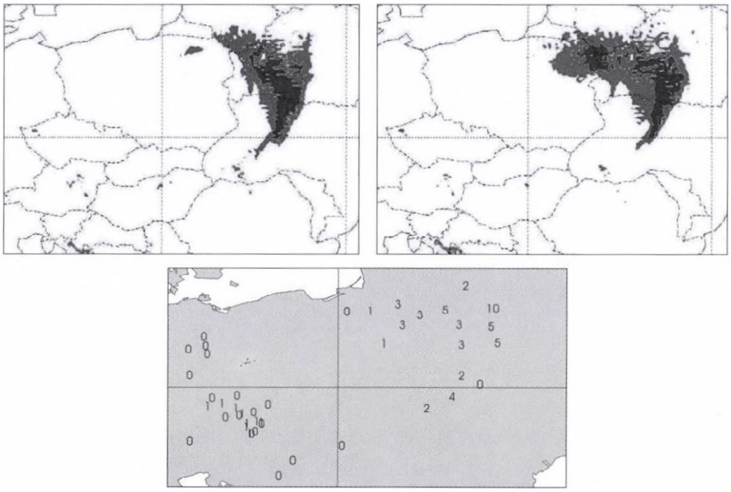


Fig. 7. The 24-hour cumulative precipitation (in mm) predicted over a zoomed area within the ALADIN/HU domain from March 04, 2003, 00:00 UTC (contour 0, 5, 15, ... mm). Upper left: control run (with TEMP and SYNOP). Upper right: 3D-Var run with ATOVS assimilated in 80 km resolution. The 24-hour cumulated precipitation was estimated from 6- to 30-hour forecast ranges. Lower picture shows the measured cumulative 24-hour precipitation (in mm) on March 05, 2003, 06:00 UTC.

For this particular day (4th of March), the objective verification showed positive impact of the AMSU-A data on the 30-hour forecast of the relative humidity (*Fig. 8*).

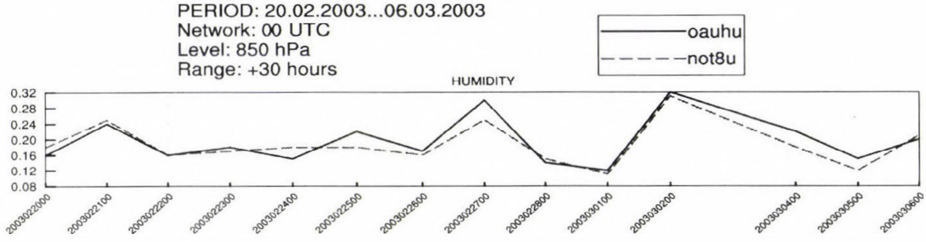


Fig. 8. Time series of the day-to-day RMSE for the 30-hour forecast of relative humidity for the runs with (not80) and without (oauhu) assimilation of AMSU-B data. The RMSE was computed from differences against observations.

5. Investigation of the AMSU-B data

Our goal was to improve the short-range forecast of precipitation, assimilating the AMSU-B data as fine resolution as possible. The AMSU-B data extracted in different resolutions (3×3 and 1×1 FOV) were investigated using the 3D-Var ALADIN/HU, testing different thinning distances (60 km, 80 km, and 120 km) in the assimilation process. In this study the model configuration *C2* (*Table 1*) was applied. As mentioned in Section 3, 3×3 FOVs of the AMSU-B correspond to one FOV of the AMSU-A. Technically, the default maximum number of the scan angle to be treated in the ARPEGE/ALADIN is thirty, which corresponds to the maximum number for the AMSU-A data. So, the treatment of the full grid (90 FOV per scan line) AMSU-B needs some modifications in the ARPEGE/ALADIN code. Consequently, new bias correction coefficients had to be computed.

5.1 Use of the ATOVS/AMSU-B data

In the ARPEGE/ALADIN model, AMSU-B channels 3, 4, and 5 are used. From both sides of scanning edges, nine pixels are removed to avoid big biases. Over land only channels 4 and 5 are used with some restrictions related to the model topography. They are used when the model topography is less than 1500 m and 1000 m, respectively. All the above mentioned three channels are used over sea. The following restrictions are applied to blacklist all the channels: 1 – where the surface temperature is less than 278 K (FOV over ice); 2 – where the absolute value of the first-guess departure (difference between the observation and background) of the channel 2 is less than 5 K, which corresponds to cloudy and rainy FOV.

5.2 Impact of the AMSU-B data

Five runs (one without AMSU-B, one with "sparse" (3×3 FOV) AMSU-B, and three with full-grid AMSU-B data (with 60 km, 80 km, and 120 km thinning distances)) were performed for a two-week period (February 07–21, 2005) to evaluate the impact of different settings of the AMSU-B data in the assimilation system. In the run without AMSU-B data, the surface, radiosonde, aircraft, and AMSU-A data were assimilated. The scores (bias and RMSE) of each run were evaluated objectively against the observations (surface and radiosonde). The accumulated precipitation was also compared to the surface measurements for a few interesting synoptic situations within the study period. We also present the results of the subjective and objective verifications evaluated during the test of the AMSU-B radiances in our parallel suite. The most important results are summarized in Section 5.2.1–5.2.5.

5.2.1 Influence of the assimilation of AMSU-B data on temperature and humidity bias

The use of the AMSU-B data in the assimilation process caused a weak heating and cooling effect in the troposphere and near the tropopause, respectively (*Fig. 9*), and resulted in an increase of moisture in the troposphere in the analysis and forecast. As it was found during the everyday subjective verification, the forecasts issued from the 3D-Var cycles were more "dry" than those of the spin-up model (or dynamical adaptation). This "drying" effect of the 3D-Var caused an overestimation in the temperature fields and worsened the forecast in certain cases. In such situations the "wetting" effect of the AMSU-B data could increase the forecast accuracy. Moreover, apart from AMSU-B data, the only humidity observation we had and used was that from radiosonde measurements.

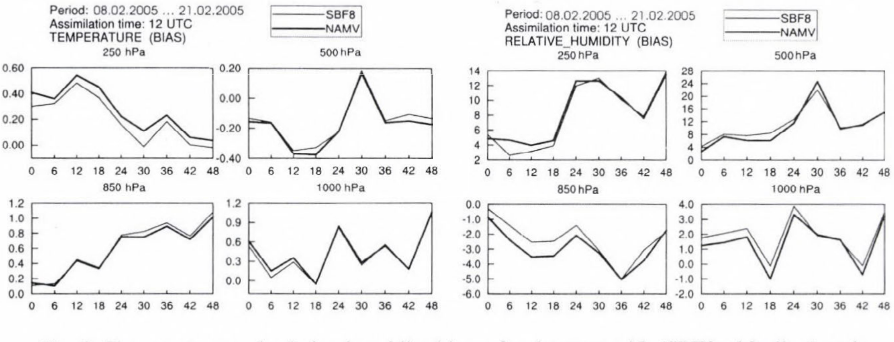


Fig. 9. Temperature and relative humidity biases for the runs with (SBF8: thin line) and without (NAMV: bold line) AMSU-B data at the analysis (0) and subsequent forecast times. The bias was computed from differences against observations.

5.2.2 Significance test of the impact of AMSU-B data on the analysis and short-range forecasts

As discussed above, the systematic addition of moisture in the model leaded to a positive impact not only on the temperature analysis and forecast (*Fig. 10*) but also on the forecast of relative humidity. *Fig. 11* shows clear positive impact of the AMSU-B data on the 48-hour forecast of the relative humidity. The impact on the analysis and forecasts of geopotential, wind speed, and wind direction was found to be neutral (not shown).

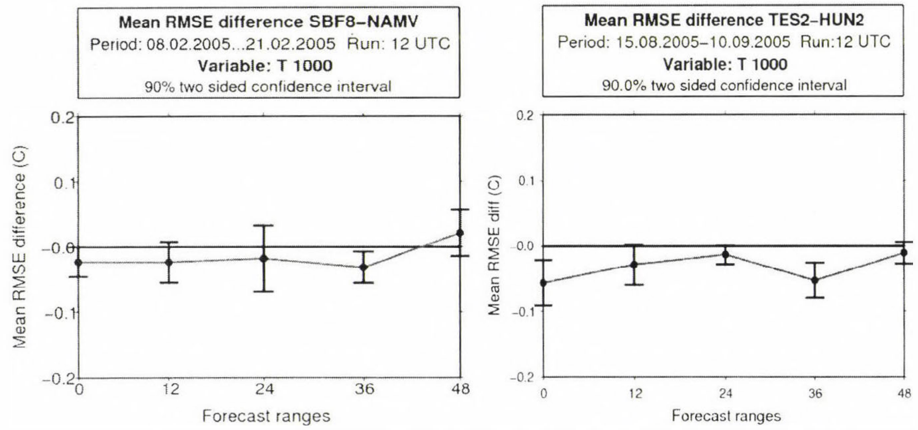


Fig. 10. Significance test: temperature RMSE differences at 1000 hPa for two different periods (NAMV and HUN2 are the reference runs without AMSU-B data, SBF8 and TES2 are experiments including AMSU-B data). The RMSE was computed from differences against observations.

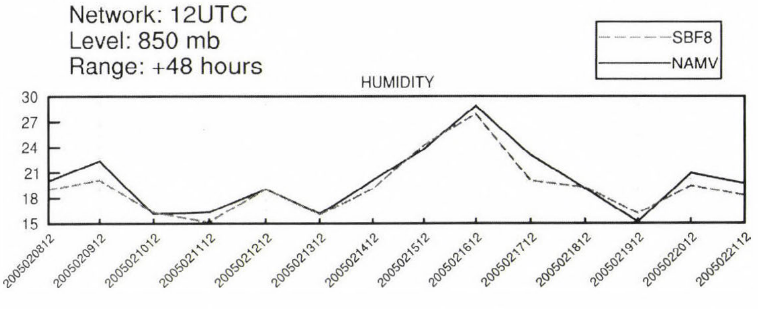


Fig. 11. Time series of the day-to-day RMSE for the 48-hour forecast of relative humidity for the runs with (SFB8) and without (NAMV) assimilation of AMSU-B data. The RMSE was computed from differences against observations.

5.2.3 Evaluation of the effect of different thinning distances

Four settings – three runs with full grid using different thinning distances (SFB8: 80 km, SFB6: 60 km, and SBF1: 120 km) in the assimilation system, and one run with reduced (3×3 FOV) number of observations (SBX3, thinning distance: 80 km) – were compared to find the best treatment of the AMSU-B in the assimilation system. In *Fig. 12* the time series scores of the 48-hour forecast of the relative humidity evaluated for the whole ALADIN/HU domain are presented. First of all, we found that the use of AMSU-B in 1×1 FOV (SBF8) is more efficient than its use in 3×3 FOV (SBX3). Concerning the choice of the thinning distance, the 80 km version proved to be better than the two others (60 and 120 km).

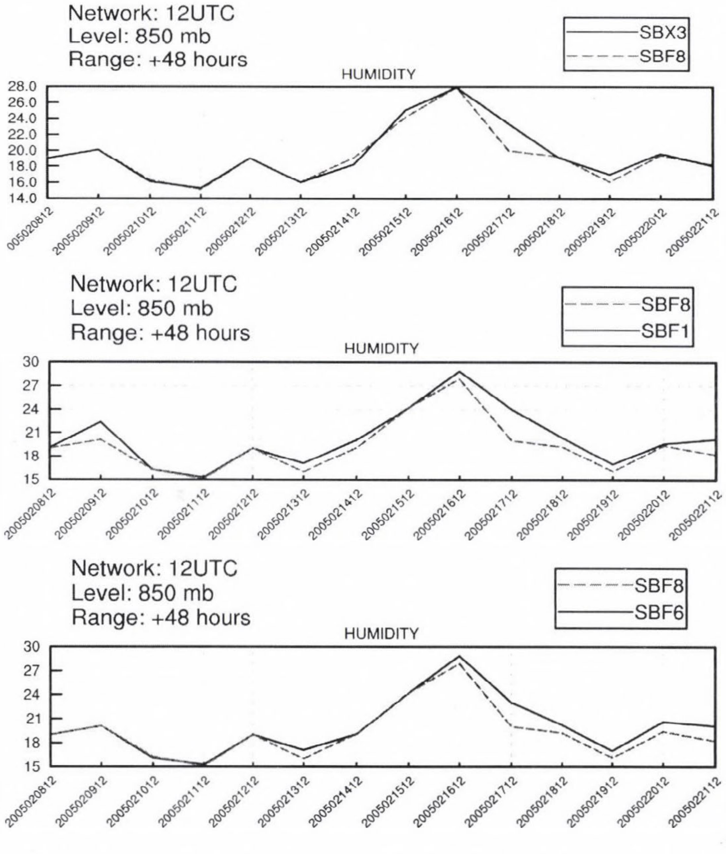


Fig. 12. Comparison of the time series of the day-to-day RMSE for relative humidity of individual runs with “sparse” AMSU-B (SBX3) and full grid AMSU-B data assimilated in 80 (SBF8), 60 (SBF6), and 120 km (SBF1) resolutions. The RMSE was computed from differences against observations.

5.2.4 Case study

Fig. 13 shows the observed and predicted cumulative precipitation over Hungary. All the runs (with and without AMSU-B data) gave quite good prediction of the rainfalls observed in the western part of the country. The precipitation patterns in the eastern part, however, were only predicted by runs that used the AMSU-B data in full grid.

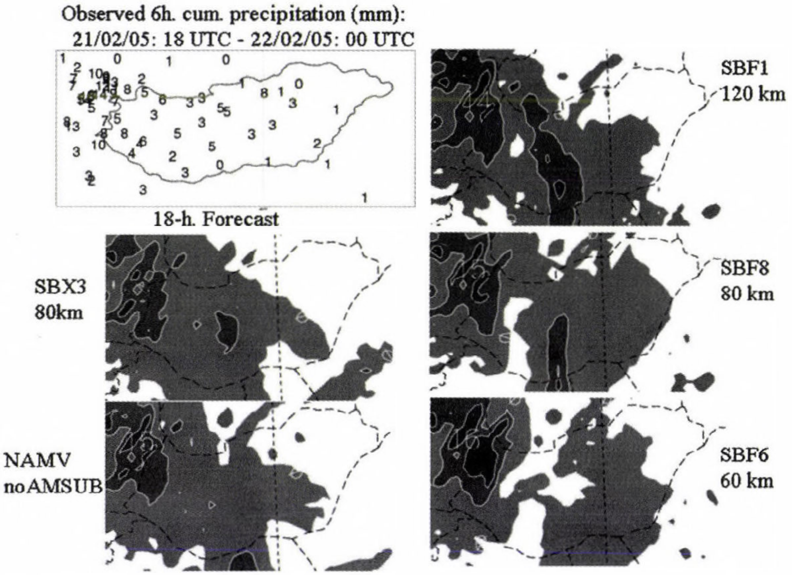


Fig. 13. Observed (top, upper left) and predicted 6-hour cumulated (estimated from 12 till 18 hours) precipitation amount valid for February 22, 2005, 06:00 UTC. Contour lines: 0, 1, 5, 10, 30, ... mm.

5.2.5 Subjective and objective scores of the use of the AMSU-B data in parallel suite

The performance of the main models used by the forecasters and those running in parallel suite are evaluated subjectively every day. The subjective verification system at the HMS (*Tóth*, 2004) concerns mainly the ECMWF products and three versions of the ALADIN/HU model: (1) the operational one (HUN2) that actually uses a 3D-Var system assimilating the surface, radiosonde, and aircraft measurements, and the ATOVS AMSU-A radiances to create the initial condition for the forecast model; (2) the one that uses the ARPEGE analysis as initial condition (the so-called dynamical adaptation); and (3) a system that is being tested, which uses the 3D-Var analysis system that also incorporates the full-grid AMSU-B data to create the initial condition (TEST2). *Fig. 14* shows the subjective scores for the forecasts of precipitation up to 24 hours (the first day and 24 hours cumulated precipitation). In the subjective verification, the higher the score, the better the forecast (10 means perfect, 1 means very bad forecast). *Fig. 14* shows 1 day with worse and 3 days with improved forecast owing to the use of AMSU-B (TES2) during the first 2 weeks of November 2005. Note that in the subjective verification only a small domain covering Hungary was evaluated. According to the objective verification, performed for the whole ALADIN/HU domain, a positive impact for the period from November 2–19, 2005 (*Fig. 15*) was observed when comparing the 24-hour forecast of precipitation with the surface gauges data.

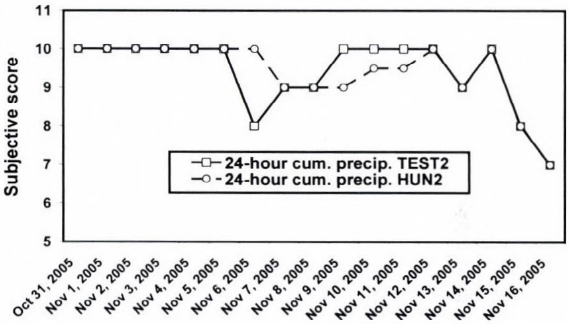


Fig. 14. Subjective scores for the 24-hour cumulated precipitation of the run in parallel suite using AMSU-B data (TEST2 – solid line), and the operational run without AMSU-B data (HUN2 – dashed line). The comparison is valid for the Hungarian territory and close surrounding regions. The forecast is from the 00:00 UTC network.

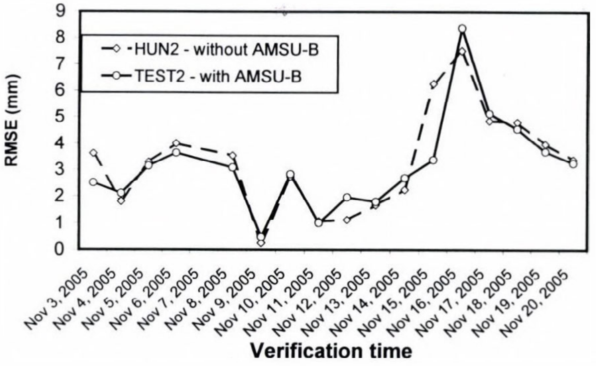


Fig. 15. Objective scores (RMSE) for the 24-hour cumulated precipitation of the run in parallel suite using AMSU-B data (TEST2 – solid line), and of the operational run without AMSU-B data (HUN2 – dashed line). The comparison is valid for the whole ALADIN/HU domain. The forecast is from the 00:00 UTC network.

6. Conclusions

In this study, the impact of ATOVS data on the analyses and forecasts of the ALADIN/HU limited area model was evaluated. Root-mean-square error (RMSE) comparisons allowed us to conclude the following:

- A neutral impact of the AMSU-A radiances on the analysis and short-range forecast of the geopotential fields was observed. In general, the use of the AMSU-A improved the forecast of temperature and humidity in the lower troposphere. The positive impact was somewhat stronger when AMSU-A data were assimilated at finer resolution (80 km, compared to 120 km), especially when the specific humidity was assimilated in univariate form.
- The impact of AMSU-B data on the analysis and short-range forecast of temperature, geopotential, and wind fields was found to be slightly positive during the study period. Positive impact on the forecast of relative humidity was observed. The use of the AMSU-B data improved the forecast of precipitation. Clear positive impact on the temperature of the AMSU-B data was observed in the lower model levels during their use in the parallel suite.
- Our experiments showed that the resolution of the input ATOVS data was important for their better assimilation in a LAM. The assimilation of the AMSU-A and AMSU-B data in full grid is preferable. The "optimal thinning distance" of the ATOVS data for our system was found to be 80 km.

Acknowledgements—The research for this paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the Hungarian National Scientific Foundation (OTKA T049579). The help from *Philippe Caille, Florence Rabier, Élisabeth Gérard,* and *Claude Fischer* of the Météo-France, and the fruitful discussion with the colleagues of the Numerical Modeling and Climate Dynamics division of the Hungarian Meteorological Service are highly acknowledged. The author recognizes and appreciates the participation of *Regina Szoták* in this investigation.

References

Ajjaji, R. and *Issara, S.,* 1994: Introduction de l'analyse canari du mod le global arp ge dans le mod le domaine limité aladin. *Master's Thesis.* Ecole Nationale de la Météorologie, Toulouse.

Berre, L., 2000: Estimation of synoptic and meso scale forecast error covariances in a limited area model. *Mon. Weather Rev. 128,* 644-667.

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

Bubnová, R., Hello, G., Bénard, P., and *Geleyn, J.-F.,* 1995: Integration of fully-elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/ALADIN NWP system. *Mon. Weather Rev. 123,* 515-535.

Candy, B., 2005: Improved use of AMSU-B data in UK Met Office regional models. *Proceedings of the TOVS Study Conference,* ITSC-14, 25-31 May 2005, Beijing, China.

Chouinard, C. and *Hallé, J.*, 2003: The assimilation of AMSU-B radiances in the CMC global data. *Proceedings of the TOVS Study Conference*, ITSC-13, 29 October–4 November 2003, Sainte-Adele, Quebec, Canada, 1-13.

Deckmyn, A. and *Berre, L.*, 2005: A wavelet approach to representing background error covariances in a limited area model. *Mon. Weather Rev. 133*, 1279-1294.

Derber, J. and *Bouttier, F.*, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus 48A*, 501-517.

English, S.J., *Hilton, F.*, *Candy, B.*, *Whyte, K.*, *Atkinson, N.*, *Smith, A.*, and *Bell, B.*, 2003: Operational use of ATOVS at the Met Office. *Proceedings of the TOVS Study Conference*, ITSC-13, 29 October–4 November 2003, Sainte-Adèle, Quebec, Canada, 14-18.

English, S.J., *Renshaw, R.J.*, *Dibben, P.C.*, *Smith, A.J.*, *Rayer, P.J.*, *Poulsen, C.*, *Saunders, F.W.*, and *Eyre, J.R.*, 2000: A comparison of the impact of TOVS and ATOVS satellite sounding data on the accuracy of numerical weather forecasts. *Q. J. Roy. Meteorol. Soc. 126*, 911-931.

Fischer, C., *Montmerle, T.*, *Berre, L.*, *Auger, L.*, and *Stefanescu, S.*, 2005: An overview of the variational assimilation in the Aladin/France numerical weather prediction system. *Q. J. Roy. Meteor. Soc. 131*, 3477-3492.

Gérard, É., *Rabier, F.*, *Lacroix, D.*, and *Sahlaoui, Z.*, 2003: Use of ATOVS raw radiances in the operational assimilation system at Météo-France. *Proc. TOVS Study Conference*, ITSC-13, 29 October–4 November 2003, Sainte-Adele, Quebec, Canada. 20-29.

Harris, B.A. and *Kelly, G.*, 2001: A satellite radiance-bias correction scheme for data assimilation. *Q. J. Roy. Meteorol. Soc. 127*, 1453-1468.

Horányi, A., *Ihász, I.*, and *Radnóti, G.*, 1996: ARPEGE/ALADIN: A numerical weather prediction model for Central-Europe with the participation of the HMS. *Időjárás 100*, 277-301.

Jones, D.C., *Renshaw, R.J.*, *Chalcraft, B.*, *Anderson, S.*, and *English, S.J.*, 2002, ATOVS assimilation in the Met Office mesoscale model. *Proc. TOVS Study Conference*, ITSC-12, 26 Feb– 6 Mar 2002, Lorne, Australia, available on: http://cimss.ssec.wisc.edu/itwg/itsc/itsc12/itsc12agenda.html.

Parrish, D.F. and *Derber, J.C.*, 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. *Mon. Weather Rev. 120*, 1747-1763.

Rabier, F., *Gérard, É.*, *Sahlaoui, Z.*, *Dahoui, M.*, and *Randriamampianina, R.*, 2001: Use of ATOVS and SSMI observations at Météo-France. *11th Conference on Satellite Meteorology and Oceanography*. Madison, WI, 15-18 October 2001 (preprints). Boston, MA, American Meteorological Society, 367-370.

Radnóti, G., 1995: Comments on "A spectral limited-area formulation with time-dependent boundary conditions applied to the shallow-water equations". *Mon. Weather Rev. 123*, 3122-3123.

Randriamampianina, R., 2005: Radiance-bias correction for a limited area model. *Időjárás 109*, 143-155.

Randriamampianina, R. and *Rabier, F.*, 2002: Regional use of locally received ATOVS radiances in NWP. *Proceedings of the TOVS Study Conference*, ITSC-12, 26 Feb–6 Mar 2002, Lorne, Australia, 285-292.

Sadiki, W. and *Fischer, C.*, 2005: A posteriori validation applied to the 3D-VAR ARPEGE and ALADIN data assimilation systems. *Tellus 57A*, 21-34.

Saunders, R., *Matricardi, M.*, and *Brunel, P.*, 1998: An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. Roy. Meteorol. Soc. 125*, 1407-1425.

Soci, C., *Horányi, A.*, and *Fischer, C.*, 2006: Sensitivity of high resolution forecasts using the adjoint technique at the 10 km scale. Accepted to *Mon. Weather Rev.*

Široká, M., *Fischer, C.*, *Cassé, V.*, *Brožková, R.*, and *Geleyn, J.-F.*, 2003: The definition of mesoscale selective forecast error covariances for a limited area variational analysis. *Meteorol. Atmos. Phys. 82*, 227-244.

Tóth, H., 2004: Subjective evaluation of different versions of ALADIN/HU model. *ALADIN Newsletter*, No. 26, 88-89.

Vasiliu, S. and *Horányi, A.*, 2005: An evaluation of the performance of the three-dimensional variational data assimilation scheme for the ALADIN/HU spectral limited area model. *Időjárás 109*, 233-257.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 349–363

Description and evaluation of a coupled Eulerian transport-exchange model

Part I. Model development

István Lagzi[1], **Róbert Mészáros**[2*], **Ferenc Ács**[2], **Alison S. Tomlin**[3], **László Haszpra**[4] and **Tamás Turányi**[1]

[1]*Department of Physical Chemistry, Institute of Chemistry, Eötvös Loránd University, P.O. Box 32, H-1518 Budapest, Hungary*

[2]*Department of Meteorology, Eötvös Loránd University, P.O. Box 32, H-1518 Budapest, Hungary; E-mail: mrobi@nimbus.elte.hu*

[3]*Energy and Resources Research Institute, University of Leeds, Leeds, LS2 9JT, U.K.*

[4]*Hungarian Meteorological Service, P.O. Box 39, H-1675 Budapest, Hungary*

(Manuscript received in final form June 22, 2006)

Abstract—An Eulerian photochemical reaction–transport model and a detailed dry deposition model have been coupled to describe both continuous air pollution and accidental release over Central Europe. Up to now, model applications have been carried out for estimating ozone flux over Hungary and transport of passive tracers from a point source. Simulating the chemical reactions, the simple GRS (Generic Reaction Set) chemical scheme was used, although, the model allows the utilization of any other, more comprehensive reaction scheme. During the transmission processes of radioactive tracers, only radioactive decay has been considered. Because of detailed parameterization of deposition processes, not only the concentration, but the effective ozone load can also be estimated by the model. Meteorological data utilized in the model have been obtained by the ALADIN meso-scale limited area numerical weather prediction model used by the Hungarian Meteorological Service. Detailed model description is presented in this study. Model sensitivity tests and some results will be presented in a companion paper.

Key-words: dispersion model, dry deposition model, adaptive grid, photochemical air pollution

---

[*] Corresponding author

1. Introduction

Previous EUROTRAC investigations (EUROTRAC 1 and EUROTRAC 2; *Haszpra et al.,* 2003) have shown that some of the highest regional ozone concentrations in Europe can be observed in Central Europe, including Hungary. During summer ozone episodes, the ozone burden of natural and agricultural vegetation is often well beyond tolerable levels. Elevated ozone concentration can be harmful to agricultural and natural vegetation. Air quality measures based on accumulated exposure over a threshold (AOT) such as AOT40 were therefore developed based on experiments in order to try to mitigate damage (*Fuhrer et al.,* 1997). However, since ozone enters plants through the stomata, the response of vegetation to changes in atmospheric ozone concentrations is more directly influenced by the stomatal ozone flux than the atmospheric concentration itself. Therefore, it has been suggested that the stomatal ozone flux is a more appropriate measure for ozone damage than the AOT 40 value (e.g., *Emberson et al.,* 2000a; *Musselman et al.,* 2006). This flux depends on several factors including the soil wetness state in moderate soil water availability conditions. An important tool in the management of photochemical smog episodes is a computational model, which can be used to test the effect of possible emission control strategies. High spatial resolution of such a model is important to reduce the impact of numerical errors on predictions and to allow better comparison of the model with experimental data during validation. The review paper of *Peters et al.* (1995) highlights the importance of developing more efficient grid systems for the next generation of air pollution models, in order to capture important smaller scale atmospheric phenomena.

This paper, therefore, presents the development of an adaptive grid model for the Central European region describing the formation of photochemical oxidants and ozone fluxes based on unstructured grids. The initial base grid of the model uses a nested approach with a coarse grid covering the wider Central European region and finer resolution grid over Hungary. Further refinement or de-refinement is then invoked using indicators based on the comparison of high and low order numerical solution of the atmospheric diffusion equation. Using this method, an efficient grid resolution strategy can be achieved in a computationally effective way.

Flux calculations without using a transport model are less precise, because of the inaccurately known spatial distribution of ozone concentrations estimated from measurements at Hungarian monitoring stations. At the same time, the spatial distribution of ozone concentration is shown to be a less accurate measure of effective ozone load, than the spatial distribution of ozone flux.

This model is also able to predict the dispersion of passive tracers (e.g., radioactive substances, chemical toxic species). The numerical algorithms applied in this version of the dispersion model are based on SPRINT2D software package (*Berzins et al.,* 1989; *Berzins* and *Furzeland,* 1992; *Berzins* and *Ware,* 1995, 1996).

Input data for the coupled transport–deposition model are presented in *Table 1.* Detailed description of both transport and deposition models is presented in next chapters.

Table 1. Input data of the model

	Input data	**Notation**	**Unit**
Place and time	Latitude, longitude	φ, λ	radian
	Elevation	z_a	m
	Season categories	S_x	-
	Day of the year	D_y	-
	Hour	t_{UTC}	hour
Surface atmospheric data	Air temperature	t_a	°C
	Components of wind speed	u, v	$m\ s^{-1}$
	Global radiation	R_G	$W\ m^{-2}$
	Cloudiness	N	eighth
	Relative humidity	f	%
	Daily precipitation amount	P	mm
Upper air meteorological data	Air temperature (4 layers)	t_a	°C
	Components of wind speed (4 layers)	u, v	$m\ s^{-1}$
	Relative humidity (4 layers)	f	%
	Height of the mixing layer	H_m	m
Emission inventories	NO_x, VOC, CO	E_i	$g\ s^{-1}$
Surface and plant specific parameters	Land use categories	LUC	-
	Height of vegetation	z_{veg}	m
	Roughness length	z_0	m
	Displacement height	d	m
	Albedo	A	-
	Leaf area index	LAI	$m^2\ m^{-2}$
	Modified Pristley-Taylor parameter	α	-
Soil parameters	Soil categories	Tx	-
	Field capacity soil moisture content	θ_f	$m^3\ m^{-3}$
	Wilting point soil moisture content	θ_w	$m^3\ m^{-3}$
	Saturated soil moisture content	θ_s	$m^3\ m^{-3}$
Resistance parameters	Minimum stomatal resistance	$r_{st,\min}$	$s\ m^{-1}$
	Radiation correction term	b_{st}	$W\ m^{-2}$
	Minimum temperature	t_{min}	°C
	Maximum temperature	t_{max}	°C
	Optimal temperature	t_{opt}	°C
	Mesophyll resistance	R_{mes}	$s\ m^{-1}$
	Cuticular resistance	R_{cut}	$s\ m^{-1}$
	Soil resistance	R_s	$s\ m^{-1}$

2. *The dispersion model*

The model describes the spread of reactive air pollutants within a 2D unstructured triangular based grid representing layers within the troposphere over the Central European region, including Hungary. The model describes the horizontal domain using a Cartesian coordinate system through the stereographic polar projection of a curved surface onto a flat plane. The total horizontal domain size is 1540 km × 1500 km (*Fig. 1*). Vertical resolution of pollutants is approximated by the application of four layers representing the surface, mixing, reservoir layers and the free troposphere. Reactive dispersion in the horizontal domain is described by the atmospheric diffusion equation in two space dimensions:

$$\begin{aligned}\frac{\partial c_s}{\partial t} = &-\frac{\partial (uc_s)}{\partial x} - \frac{\partial (vc_s)}{\partial y} + \frac{\partial}{\partial x}\left(K_x \frac{\partial c_s}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial c_s}{\partial y}\right) \\ &+ R_s(c_1, c_2, \ldots, c_n) + E_s - k_s c_s,\end{aligned} \tag{1}$$

where c_s is the concentration of the sth compound, u and v are horizontal wind components, K_x and K_y are eddy diffusion coefficients, k_s is the dry deposition rate constant, E_s describes the distribution of emission sources for the sth compound, and R_s is the chemical reaction term, which may contain non-linear terms in c_s. For n chemical species, an n dimensional set of partial differential equations is formed describing the change of concentrations over time and space. These equations are coupled through the non-linear chemical reaction term.

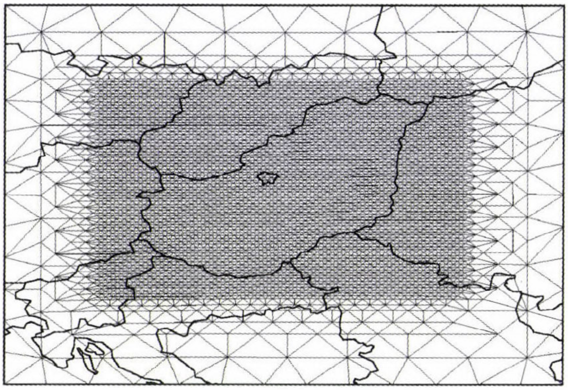


Fig. 1. The typical nested grid structure of the dispersion model. The average mesh length of the outer coarse grid and that of the nested fine grid are 100 and 12.5 km, respectively.

The four vertical layers of the model are defined as: the surface layer extending to 50 m, the mixing layer, a reservoir layer, and the free troposphere (upper) layer. The mixing layer extends to a height determined by radiosonde data at 00:00 UTC, but is modeled to rise smoothly to a height determined by radiosonde data at 12:00 UTC during the day. The reservoir layer, if it exists above the height of the mixing layer, extends from the top of the mixing layer to an altitude of 1000 m.

Relative humidity and temperature data were determined by the meteorological model ALADIN with a time resolution of 6 hours and spatial resolution of 0.1 × 0.15 degrees (*Horányi et al.,* 1996). In our model, conservative interpolation methods were used to obtain data relevant to a given spatial point on the unstructured grid from the regularly gridded ALADIN meteorological data.

For Budapest, the emission inventories for CO, NO_x, and VOCs were provided by the local authorities with a spatial resolution of 1 km × 1 km including the most significant 63 emission point sources. For Hungary, the National Emission Inventory of spatial resolution of 20 km × 20 km was applied, which included both area and point sources. Outside Hungary, the emission inventory of EMEP for CO, NO_x, and VOCs was used, having a spatial resolution of 50 km × 50 km. Natural VOC and NO_x emission have been neglected in the model. Parameterization of biogenic emissions requires several other input data, such as forest statistical data and bibliographic data on three species potential emissions. However, based on the study of *Moukhtar et al.* (2005), the effect of biogenic emissions of ozone precursors (VOC) on ozone concentration was only maximum 5%.

The emission data had to be interpolated onto the unstructured grid following each change to the mesh during refinement. This was achieved using the mass conservative method of overlapping triangles. Point sources are averaged into the appropriate grid cell for their location, and hence, when the grid is refined, the definition of point sources improves.

In the model, the GRS chemical scheme (*Azzi et al.,* 1992; *Cope et al.,* 2005) was used, although the model allows the utilization of any other reaction schemes. The GRS scheme is a reduced mechanism that was created using a semi-empirical approach; it contains 7 reactions of 7 species (*Table 2*). The GRS scheme has been evaluated by comparison with smog chamber data and predictions from more detailed chemical schemes. Previous studies (*Azzi et al.,* 1992; *Cope et al.,* 2005) have shown that the scheme performs well for the prediction of ozone in polluted conditions, although it can overpredict ozone concentrations in rural locations. The scheme has been selected in the current application for its computational efficiency, and because its accuracy can be assumed to be reasonable in the region of interest, i.e., down wind of major

NO_x sources. The rate constants were calculated and expressed as *m*th order rate constants with units (molecule $cm^3)^{m-1}$ s^{-1}. The photolysis rates were parameterized by the following function:

$$J_q = (1 - 0.75 N^{3.4}) a_q \exp(b_q \sec \Theta), \tag{2}$$

where Θ is the solar zenith angle, N is the cloud coverage, and a_q, b_q are the rate parameters of reaction q. Temperature dependent rate constants were represented by standard Arrhenius expressions.

Table 2. The GRS mechanism (*T*: temperature, Θ: solar zenith angle)

Reactions	Reaction rate constants	
ROC + $h\nu$ $\rightarrow$ RP + ROC	$k_1 = 1000 \exp(-4710/T) J_3$	[R1]
RP + NO $\rightarrow$ NO_2	$k_2 = 3.7098 \times 10^{-12} \exp(242/T)$	[R2]
NO_2 + $h\nu$ $\rightarrow$ NO + O_3	$J_3 = 1.45 \times 10^{-2} \exp(-0.4 \sec \Theta)$	[R3]
NO + O_3 $\rightarrow$ NO_2	$k_4 = 1.7886 \times 10^{-12} \exp(-1370/T)$	[R4]
RP + RP $\rightarrow$ RP	$k_5 = 6.7673 \times 10^{-12}$	[R5]
RP + NO_2 $\rightarrow$ SGN	$k_6 = 1.00 \times 10^{-13}$	[R6]
RP + NO_2 $\rightarrow$ SNGN	$k_7 = 1.00 \times 10^{-13}$	[R7]

2.1 Solution method

The basis of the numerical method is the space discretization of the partial differential equations (PDEs) derived from the atmospheric diffusion equation on unstructured triangular meshes using the software SPRINT2D (*Berzins et al.*, 1989; *Berzins* and *Furzeland,* 1992; *Berzins* and *Ware*, 1995, 1996). This approach (known as the "method of lines"), reduces the set of PDEs in three independent variables to a system of ordinary differential equations (ODEs) in one independent variable, the time. The system of ODEs can then be solved as an initial value problem. For advection dominated problems it is important to choose a discretization schemes which preserves the physical range of the solution.

Unstructured triangular meshes are commonly used in finite volume/ element applications because of their ability to deal with general geometries. In terms of application to multi-scale atmospheric problems, we are not dealing with complex physical geometries, but unstructured meshes provide a good method of resolving the complex structures formed by the interaction of chemistry and flow in the atmosphere and by the varying types of emission sources. The term unstructured represents the fact that each node in the mesh may be surrounded by any number of triangles, whereas in a structured mesh this number would be fixed. In the present work, a flux limited, cell-centered, finite volume discretization scheme of *Berzins* and *Ware* (1995, 1996) was chosen on an unstructured triangular mesh. This method enables accurate solutions to be determined for both smooth and discontinuous flows by making use of the local Riemann solver flux techniques (originally developed for the Euler equations) for the advective parts of the fluxes, and centered schemes for the diffusive part. The scheme of *Berzins* and *Ware* (1995, 1996) has the desirable properties of preserving positivity, eliminating spurious oscillations, and restricting the amount of diffusion by the use of a nonlinear limiter function. The advection scheme has been shown to be of second order accuracy. The diffusion terms are discretized by using a finite volume approach to reduce the integrals of second derivatives to the evaluation of first derivatives at the midpoints of edges. These first derivatives are then evaluated by differentiating a bilinear interpolant based on four mid-point values. The model applies Dirichlet- and Neumann-type boundary conditions depending on the advective fluxes over boundary edge. The boundary conditions are imposed through the approximate Riemann solver.

A method of lines approach with the above spatial discretization scheme results in a system of ODEs in time, which are integrated using the code SPRINT with the Theta option ,which is specially designed for the solution of stiff systems with moderate accuracy and automatic control of the local error in time. Operator splitting is carried out at the level of the nonlinear equations formed from the method of lines by approximating the Jacobian matrix. The approach introduces a second-order splitting error, but fortunately this error alters only the rate of convergence of the iteration, as the residual being reduced is still that of the full ODE system. This provides significant advantages over other splitting routines such as Strang splitting.

The initial unstructured meshes used in SPRINT2D are created from a geometry description using the Geompack mesh generator (*Joe,* 1991). These meshes are then refined and coarsened by the Triad adaptivity module, which uses tree like data structures to enable efficient mesh adaptation by providing the necessary connectivity. A method of refinement based on the regular subdivision of triangles has been chosen. Here an original triangle is split into four similar triangles by connecting the midpoints of the edges as shown in *Fig. 2.* These may be coalesced into the parent triangle later, when coarsening the mesh. This process is called local h-refinement, since the nodes of the original mesh do not move, and we are simply subdividing the original elements. In order to implement the adaptivity module, a suitable criterion must be chosen. The ideal situation would be that the decision to refine or de-refine would be made on a fully automatic basis with no user input necessary.

In practice, a combination of an automatic technique and some knowledge of the physical properties of the system is used. The technique used in this work is based on the calculation of spatial error estimates. Low and high order solutions are obtained for each species, and the difference between them gives a measure of the spatial error. The algorithm can then be choosen to refine in regions of high spatial error by comparison with a user defined tolerance for one or the sum of several species. For the *i*th PDE component on the *j*th triangle, a local error estimate $e_{i,j}(t)$ is calculated from the difference between the solution using a first order method and that using a second order method. For time dependent PDEs, this estimate shows how the spatial error grows locally over a time step. A refinement indicator for the *j*th triangle is defined by an average scaled error $(serr_j)$ measurement over all *npde* PDEs using supplied absolute and relative tolerances:

$$serr_j = \sum_{i=1}^{npde} \frac{e_{i,j}(t)}{atol_i / A_j + rtol_i\, c_{i,j}}, \tag{3}$$

where $atol_i$ and $rtol_i$ are the absolute and relative error tolerances, $e_{i,j}(t)$ is the local error estimate of species *i* over element *j*, $c_{i,j}$ is the concentration of species *i* over triangle, *j*, A_j is the area of *j*th triangle and *npde* is the number of partial differential equations applied. This formulation for the scaled error provides a flexible way to weight the refinement towards any PDE errors.

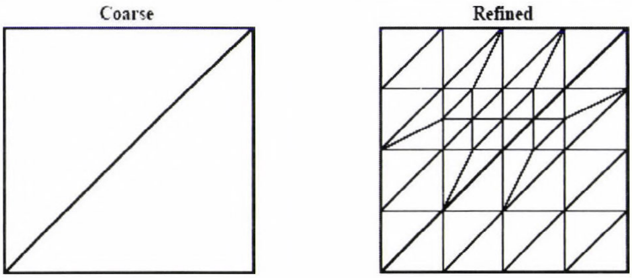


Fig. 2. Subdivision of the triangular cells using adaptive gridding strategy.

In the photochemical smog calculations, a combination of errors in species NO and NO_2 were used as a refinement indicator, because these are primary species, and also because their concentrations are very closely related to ozone production. Estimation of the local spatial error of ozone concentration is not an efficient choice, because it would be too late to make refinement decisions on the basis of the detection of a large error in the concentration of a secondary pollutant. On the other hand, concentrations of the VOCs are locally dominated by emissions, and since the available emission inventory for VOCs has a coarse resolution (50 km × 50 km), the use of VOC concentration as an error indicator is not appropriate. *Tomlin et al.* (1997) previously demonstrated the success of using the local spatial error of the concentrations of nitrogen oxides for appropriate mesh refinement for a reactive plume from a NO_x ($NO+NO_2$) source. Each triangle that is flagged for refinement is split into four similar triangles (*Fig. 2*). Refined triangles may later be coalesced into the parent triangle when coarsening the mesh.

The application of adaptive rectangular meshes would be also possible but less effective in terms of the number of nodes required in order to achieve high levels of adaptivity. Although the data structures resulting from an unstructured mesh are somewhat more complicated than those for a regular Cartesian mesh, problems with hanging nodes at boundaries between refinement regions are avoided. The use of a flexible discretization stencil also allows for an arbitrary degree of refinement, which is more difficult to achieve on structured meshes.

3. The dry deposition model

Models to describe the dry deposition of ozone are based on the inferential method (*Baldocchi et al.,* 1987; *Hicks et al.,* 1987; *Kramm et al.,* 1995; *Padro,* 1996; *Walmsley* and *Wesely,* 1996; *Grünhage* and *Haenel,* 1997; *Meyers et al.,* 1998; *Padro et al.,* 1998; *Brook et al.,* 1999; *Emberson et al.,* 2000b; *Klemm* and *Mangold,* 2001; *Zhang et al.,* 2002). The dry deposition velocity of ozone was estimated over different types of vegetation. The land-cover map was generated using a Hungarian land-use map (*Fig. 3*). The model was applied on the grid of the meso-scale limited area numerical weather prediction model ALADIN (*Horányi et al.*, 1996). The time and space resolution of the data was 6 hours and 0.10 × 0.15 degrees, respectively.

The total ozone flux (F_t) was calculated as a product of the deposition velocity of ozone (v_d) and the ozone concentration (c_r) at a reference height (within the surface layer of the model):

$$F_t = v_d \, c_r \tag{4}$$

The deposition velocity is defined as the inverse of the sum of the atmospheric and surface resistances, which retard the ozone flux:

$$v_d = (R_a + R_b + R_c)^{-1}, \tag{5}$$

where R_a, R_b, and R_c are the aerodynamic resistance, the quasi-laminar boundary layer resistance, and the canopy resistance, respectively.

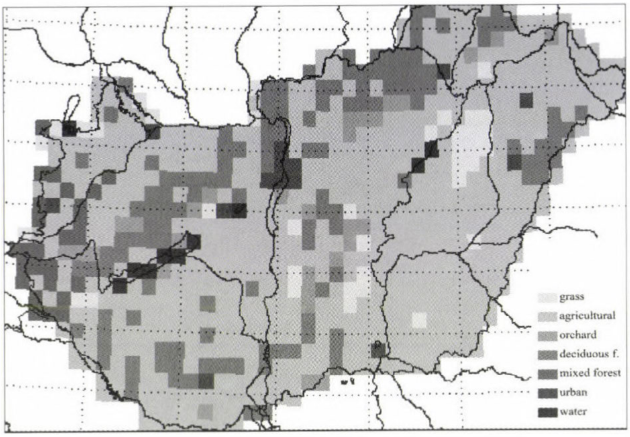


Fig. 3. Land use categories in the model.

The aerodynamic resistance is calculated using the Monin-Obukhov similarity theory taking into account the atmospheric stability (*Ács* and *Szász*, 2002):

$$R_a = \frac{1}{\kappa u_*} \ln\left(\frac{z-d}{z_0}\right) + 4.7\frac{z-d-z_0}{L} \quad \text{if } L > 0, \tag{6}$$

and

$$R_a = \frac{1}{\kappa u_*} \ln\left(\frac{1-y}{1-y_0}\frac{1+y_0}{1+y}\right) \quad \text{if } L < 0, \tag{7}$$

where

$$y = \left(1 - 16\frac{z-d}{L}\right)^{-1/2}, \tag{8}$$

$$y_0 = \left(1 - 16\frac{z_0}{L}\right)^{-1/2}, \tag{9}$$

where z, z_0, and d are the reference height, the roughness length, and the displacement height, respectively, depending on the surface types, $\kappa = 0.4$ is the von Kármán constant. Dynamical parameters, such as u_* and L are the friction velocity and the Monin-Obukhov length, respectively, calculated by an iterative method:

$$u_* = \kappa u \left[\ln\left(\frac{z-d}{z_0}\right) - \psi_m\left(\frac{z-d}{L}\right) + \psi_m\left(\frac{z_0}{L}\right) \right]^{-1}, \tag{10}$$

and

$$L = -\frac{u_*^3 \rho c_p T}{\kappa g H}, \tag{11}$$

where $\psi_m(\xi)$ is the integral form of universal stability correction functions for the momentum, g is the acceleration of gravity, ρ is the air density, c_p is the specific heat at constant pressure, T is the air temperature, and H is the sensible heat flux. In this study, universal functions for wind of *Beljaars* and *Holtslag* (1991) and *Dyer* (1974) were used for stable and unstable stratifications, respectively. The sensible heat flux was estimated using the modified Priestley-Taylor method (*Holtslag* and *van Ulden*, 1983).

The boundary layer resistance for ozone is calculated by an empirical relationship after *Hicks et al.* (1987):

$$R_b = 6.5 / u_* . \tag{12}$$

The canopy resistance R_c is parameterized by the following equation:

$$R_c = \frac{1}{(R_{st} + R_{mes})^{-1} + (R_s)^{-1} + (R_{cut})^{-1}}, \tag{13}$$

where R_{st}, R_{mes}, R_s, and R_{cut} are the stomatal, mesophyll, surface, and cuticular resistances, respectively.

The stomatal resistance can be calculated from the empirical formula of *Jarvis* (1976) referring to a vegetation canopy. This parameterization requires knowledge of the soil and plant physiological characteristics:

$$R_{st} = \frac{1}{G_{st}(PAR) f_t(t) f_e(e) f_\theta(\theta) f_{D,i}}, \tag{14}$$

where $G_{st}(PAR)$ is the unstressed canopy stomatal conductance, a function of PAR, the photosynthetically active radiation. In this parameterization, the canopy is divided into sunlit and shaded leaves, and G_{st} is calculated with the following form:

$$G_{st}(PAR) = \frac{LAI_s}{r_{st}(PAR_s)} + \frac{LAI_{sh}}{r_{st}(PAR_{sh})}, \tag{15}$$

$$r_{st}(PAR) = r_{st,\min}(1 + b_{st} / PAR), \tag{16}$$

where LAI_s and LAI_{sh} are the total sunlit and shaded leaf area indices, respectively, PAR_s and PAR_{sh} are PAR received by sunlit and shaded leaves, respectively. LAI_s, LAI_{sh}, PAR_s, and PAR_{sh} terms are parameterized after *Zhang et al.* (2001). The vegetation specific terms $r_{st,\min}$, b_{st}, and LAI are presented in *Lagzi et al.* (2004).

The factors in the denominator range between 0 and 1 and modify the stomatal resistance: $f_t(t)$, $f_e(e)$, and $f_\theta(\theta)$ describe the effect of temperature, the vapor pressure deficit, and plant water stress on stomata, while $f_{D,i}$ modifies the stomatal resistance for the pollutant gas of interest (for ozone, $f_{D,i} = 0.625$ after *Wesely* (1989)).

The temperature stress function is described by the following equation:

$$f_t = \frac{t - t_{\min}}{t_{opt} - t_{\min}} \left(\frac{t_{\max} - t}{t_{\max} - t_{opt}} \right)^{b_t}, \tag{17}$$

where

$$b_t = \frac{t_{\max} - t_{opt}}{t_{\max} - t_{\min}}. \tag{18}$$

Here $t_{\min}, t_{opt}$, and $t_{\max}$ are the minimum, maximum, and the optimal temperature depending on the vegetation. The stress of the vapor pressure deficit can be parameterized by the following form:

$$f_e = 1 - b_e(e_s - e), \tag{19}$$

where b_e is a vegetation dependent constant (*Brook et al.*, 1999), e and e_s are the water vapor pressure and the saturated water vapor pressure, respectively.

The water stress function $f_\theta(\theta)$ is parameterized using soil water content (θ):

$$f_\theta = \begin{cases} 1 & \text{if } \theta > \theta_f \\ \max\left\{ \dfrac{\theta - \theta_w}{\theta_f - \theta_w}, 0.05 \right\} & \text{if } \theta_w < \theta \le \theta_f, \\ 0.05 & \text{if } \theta \le \theta_w \end{cases} \tag{20}$$

where θ_w and θ_f are the wilting point and the field capacity soil moisture contents, respectively. These terms depend on the soil texture of the grid cell. The soil texture on the model grid was determined after *Várallyay et al.*

(1980). The grid cell soil texture was represented by the dominant soil texture (*Fig. 4*). The θ_w and θ_f values for several soil textures were taken from *Ács* (2003). Soil water content, θ, was modeled by a simple water-budget model (*Mészáros et al.*, 2006).

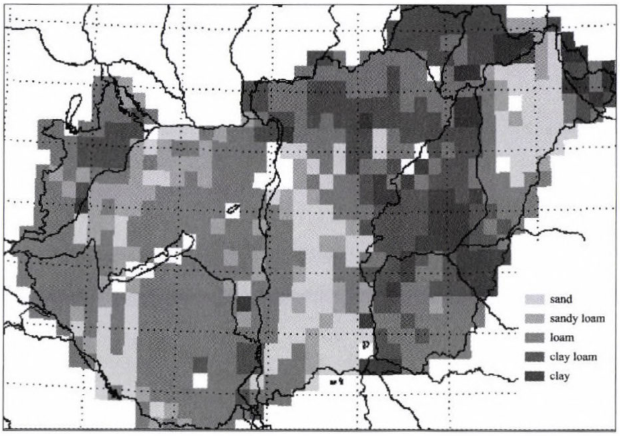


Fig. 4. Soil types in the model.

The mesophyll resistance for ozone in the model is taken to be zero. Cuticular resistance, R_{cut}, and surface resistance, R_s, for ozone deposition were obtained as in *Lagzi et al.* (2004). The calculated deposition velocities of ozone over different vegetation have a good agreement with published observed data (see *Lagzi et al.*, 2004).

4. *Conclusions*

A chemical transport model and a detailed dry-deposition model have been developed and coupled to simulate the ozone fluxes over the Central European region and estimate the dispersion of an accidental release from the nuclear power plant at Paks, Hungary. An adaptive grid model describes the formation and transformation of photochemical oxidants based on triangular unstructured grids. The model automatically places a finer resolution grid in regions characterized by high concentration gradients and, therefore, higher numerical error. Using an adaptive method, it is therefore possible to achieve grid resolutions of the order of 10 km without excessive computational effort.

Sensitivity tests and model results are presented in the second part of this study.

Acknowledgements—The authors acknowledge the support of OTKA D048673 (Postdoctoral Fellowship), OTKA F047242, NKFP 3A/088/2004, and the Békésy György Fellowship.

References

Ács, F., 2003: On the relationship between the spatial variability of soil properties and transpiration. *Időjárás 107*, 257–272.

Ács, F. and *Szász, G.*, 2002: Characteristics of microscale evapotranspiration: a comparative analysis. *Theor. Appl. Climatol. 73*, 189–205.

Azzi, M., Johnson, G.J., and *Cope, M.*, 1992: An introduction to the generic reaction set photochemical smog mechanism. *Proceedings of the 11th Clean Air Conf. 4th Regional IUAPPA Conf.*, Brisbane, Australia, pp. 451–462.

Baldocchi, D.D., Hicks, B.B., and *Camara, P.*, 1987: A canopy stomatal resistance model for gaseous deposition to vegetated canopies. *Atmos. Environ. 21*, 91–101.

Beljaars, A.C.M. and *Holtslag, A.A.M.*, 1991: Flux parameterization overland surfaces for atmospheric models. *J. Appl. Meteorol. 30*, 327–341.

Berzins, M., Dew, P.M., and *Furzeland, R.M.*, 1989: Developing software for time-dependent problems using the method of lines and differential algebraic integrators. *Appl. Numer. Math. 5*, 375–390.

Berzins, M. and *Furzeland, R.M.*, 1992: An adaptive Theta-method for the solution of stiff and nonstiff differential-equations. *Appl. Numer. Math. 9*, 1–19.

Berzins, M. and *Ware, J.*, 1995: Positive cell-centered finite volume discretization methods for hyperbolic equations on irregular meshes. *Appl. Numer. Math. 16*, 417–438.

Berzins, M. and *Ware, J.*, 1996: Solving convection and convection reaction problems using the method of lines. *Appl. Numer. Math. 20*, 83–99.

Brook, J.R., Zhang, L., Di-Giovanni, F., and *Padro, J.*, 1999: Description and evaluation of a model of deposition velocities for routine estimates of air pollutant dry deposition over North America. Part I: model development. *Atmos. Environ. 33*, 5037–5051.

Cope, M.E., Hess, G.D., Lee, S., Tory, K.J., Burgers, M., Dewundege, P., and *Johnson, M.*, 2005: The Australian Air Quality Forecasting System: Exploring first steps towards determining the limits of predictability for short-term ozone forecasting. *Bound.-Lay. Meteorol. 116*, 363–384.

Dyer, A.J., 1974: A review of flux-profile relationships. *Bound.-Lay. Meteorol. 7*, 363–372.

Emberson, L.D., Simpson, D., Tuovinen, J.-P., Ashmore, M.R., and *Cambridge, H.M.*, 2000a: Towards a model of ozone deposition and stomatal uptake over Europe. *EMEP MSC-W Note 6/00.*

Emberson, L.D., Ashmore, M.R., Cambridge, H.M., Simpson, D., and *Touvinen, J.-P.*, 2000b: Modelling stomatal ozone flux across Europe. *Atmos. Environ. 109*, 403–413.

Fuhrer, J., Skärby, L., and *Ashmore, M.R.*, 1997: Critical levels for ozone effects on vegetation in Europe. *Environ. Pollut. 97*, 91–106.

Grünhage, L. and *Haenel, H.D.*, 1997: PLATIN (PLant-ATmosphere INteraction) I: a model of plant-atmosphere interaction for estimating absorbed doses of gaseous air pollutants. *Environ. Pollut. 98*, 37–50.

Haszpra, L., Ferenczi, Z., Lagzi, I., and *Turányi, T.*, 2003: Formation of Tropospheric Ozone Formation in Hungary. In *EUROTRAC-2 (A EUREKA Environmental Project) TOR-2 Tropospheric Ozone Research. Final Report.*International Scientific Secretariat (ISS), GSF - National Research Center for Environment and Health, Munich, 87–89.

Hicks, B.B., Baldocchi, D.D., Meyers, T.P., Hosker, R.P., and *Matt, D.R.*, 1987: A preliminary multiple resistance routine for deriving dry deposition velocities from measured quantities. *Water Air Soil Poll. 36*, 311–330.

Holtslag, A.A.M. and *van Ulden, A.P.*, 1983: A simple scheme for datetime estimates of the surface fluxes from routine weather data. *J. Clim. Appl. Meteorol. 22*, 517–529.

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

Jarvis, P.G., 1976: The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Phil. Trans. Roy. Soc. B 273*, 593–610.

Joe B., 1991: GEOMPACK - a software package for the generation of meshes using geometric algorithms. *Adv. Eng. Softw. Workst. 13,* 325-331.

Klemm, O. and *Mangold, A.*, 2001: Ozone deposition at a forest site in the Bavaria. *Water Air Soil Poll.: Focus 1*, 223–232.

Kramm, G., Dlugi, R., Dollard, G.J., Foken, Th., Mölders, N., Müller, H., Seiler, W., and *Sievering, H.*, 1995: On the dry deposition of ozone and reactive nitrogen species. *Atmos. Environ. 29*, 3209-3231.

Lagzi, I., Mészáros, R., Horváth, L., Tomlin, A., Weidinger, T., Turányi, T., Ács, F., and *Haszpra, L.*, 2004: Modelling ozone fluxes over Hungary. *Atmos. Environ. 38*, 6211-6222.

Mészáros, R., Szinyei, D., Vincze, Cs., Lagzi, I., Turányi,T., Haszpra, L., and *Tomlin A.S.*, 2006: Effect of the soil wetness state on the stomatal ozone fluxes over Hungary. *Int. J. Environ. Pollut.* (in press).

Meyers, T.P., Finkelstein, P., Clarke, J., Ellestad, T.G., and *Sims, P.F.*, 1998: A multilayer model for inferring dry deposition using standard meteorological measurements. *J. Geophys. Res.-Atmos. 103*, 22645–22661.

Moukhtar, S., Bessagnet, B., Rouil, L., and *Simon, V.*, 2005: Monoterpene emissions from Beech (*Fagus sylvatica)* in a French forest and impact on secondary pollutants formation at regional scale. *Atmos. Environ. 39*, 3535–3547.

Musselman, R.C., Lefohn, A.S., Massmann, W.J., and *Heath, R.L.*, 2006: A critical review and analysis of the use of exposure- and flux-based ozone indices for predicting vegetation effects. *Atmos. Environ. 40*, 1869–1888.

Padro, J., 1996: Summary of ozone dry deposition velocity measurements and model estimates over vineyard, cotton, grass and desciduous forest in summer. *Atmos. Environ. 30*, 2363–2369.

Padro, J., Zhang, L., and *Massman, W.J.*, 1998: An analysis of mesurements and modelling of air-surface exchange of $NO\text{-}NO_2\text{-}O_3$ over grass. *Atmos. Environ. 32*, 1167–1177.

Peters, L.K., Berkovitz, C.M., Carmichael, G.R., Easter, R.C., Fairweather, G., Ghan, S.J., Hales, J.M., Leung, L.R., Pennell, W.R., Potra, F.A., Saylor, R.D., and *Tsang T.T.*, 1995: The current and future direction of Eulerian Models in simulating the tropospheric chemistry and transport of trace species: a review. *Atmos. Environ. 29*, 189–222.

Tomlin, A., Berzins, M., Ware, J., Smith, J., and *Pilling, M.J.*, 1997: On the use adaptive gridding methods for modelling chemical transport from multi-scale sources. *Atmos. Environ. 31*, 2945–2959.

Várallyay, Gy., Szűcs, L., Murányi, A., Rajkai, K., and *Zilahy, P.*, 1980: Map of soil factors determining the agro-ecological potential of Hungary (1:100 000) II. (in Hungarian). *Agrokémia és Talajtan 29*, 35–76.

Walmsley, J.L. and *Wesely, M.L.*, 1996: Modification of coded parameterizations of surface resistances to gaseous dry deposition (Technical note). *Atmos. Environ. 30*, 1181–1188.

Wesely, M.L., 1989: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmos. Environ. 23*, 1293–1304.

Zhang, L., Moran, M.D., and *Brook, J.R.*, 2001: A comparison of models to estimate in-canopy photosynthetically active radiation and their influence on canopy stomatal resistance. *Atmos. Environ. 35*, 4463–4470.

Zhang, L., Moran, M.D., Makar, P.A., Brook, R., and *Gong, S.*, 2002: Modelling gaseous dry deposition in AURAMS: a unified regional air-quality modelling system. *Atmos. Environ. 36*, 537–560.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 365–377

Description and evaluation of a coupled Eulerian transport-exchange model Part II. Sensitivity analysis and application

Róbert Mészáros[1,*], **István Lagzi**[2], **Ágota Juhász**[1], **Dalma Szinyei**[1], **Csilla Vincze**[1], **András Horányi**[3], **László Kullmann**[3] and **Alison S. Tomlin**[4]

[1]*Department of Meteorology, Eötvös Loránd University,*
H-1518 Budapest, P.O. Box 32, Hungary; E-mail: mrobi@nimbus.elte.hu

[2]*Institute of Chemistry, Eötvös Loránd University,*
H-1518 Budapest, P.O. Box 32, Hungary

[3]*Hungarian Meteorological Service, P.O. Box 39, H-1675 Budapest, Hungary*

[4]*Energy and Resources Research Institute, University of Leeds, Leeds, LS2 9JT, U.K.*

(Manuscript received in final form June 22, 2006)

Abstract—A detailed description of a coupled transport-deposition model has been given in the accompanying paper in this issue (*Lagzi et al.*, 2006). Sensitivity analysis of this model and some applications are presented in this study. Within the framework of sensitivity analysis, the effects of input data on model results have been examined. Some case studies of model applications are also presented here. Using our model, the impact of both short term accidental releases and continuous emissions of air pollutants can be estimated. An example of long-range transport processes resulting from an accidental release from a single concentrated emission source (nuclear power plant (NPP) at Paks, Hungary) is discussed. Another application of the model is the prediction of secondary pollutant loading resulting from the continuous release of pollutant precursors. Estimations of photochemical air pollution and ozone fluxes were performed on a regular grid over Hungary for the first time. Time and space resolutions of the transport-deposition model correspond to the ALADIN meso-scale limited area numerical weather prediction model used by the Hungarian Meteorological Service. Accordingly, the meteorological data utilized in the model were generated by the ALADIN model, which allows further routine model applications. The model simulations show that the predicted regions of high stomatal ozone flux (the effective ozone load) can be very different to predicted regions of high AOT 40 (accumulated ozone exposure over a threshold of 40 ppb) values depending on the weather and soil conditions. The predicted ozone deposition velocities over various vegetation types are

---

[*] Corresponding author

shown to be highly sensitive to a range of meteorological parameters for summer, sunny conditions which affects the flux of ozone from the atmosphere to the surface.

Key-words: transport-deposition model, accidental release, photochemical air pollution, stomatal ozone fluxes

1. Intoduction

A range of Eulerian or Lagrangian dispersion models are currently being used to describe the transport processes and, therefore, the concentration fields of both particles and trace gases resulting from emissions from a range of source types (in Hungary, among others MEDIA or AERMOD models, *Ferenczi* and *Ihász*, 2003; *Steib* and *Labancz,* 2006). Similarly, numerous deposition models are available to estimate the transfer processes of tracers from the atmosphere to the different surfaces (for a summary see *Lagzi et al.,* 2006). In order to fully investigate the flux of tracers to a range of surfaces, the transport and deposition processes must be considered together. For this purpose a transport model, developed at the University of Leeds and Eötvös Loránd University, Hungary, has been coupled with a detailed deposition model, developed at the Eötvös Loránd University, Hungary (*Lagzi et al.,* 2004a; *Mészáros et al.,* 2006). Estimations of effective load of trace gases on the surface without using a transport model are less precise, because of the inaccurately known spatial distribution of their concentrations. At the same time, the spatial distribution of concentrations is shown to be a less accurate measure of effective load, than the spatial distribution of the flux (*Musselman et al.,* 2006). The main goal of coupling the two models was to refine both concentration and flux fields. Model calculations were performed over Central Europe, focused on Hungary. Before the planned continuous, routine application of the coupled transport-deposition model, some test runs and also sensitivity analyses were performed.

Up to now, two different applications have been made with this model: (1) evaluation of tracer dispersion caused by a supposed accidental release of radionuclides from Paks NPP and (2) estimation of ozone load over Hungary. In the first case, a constant deposition rate of radioactive substances was assumed, while in the latter case, a detailed parameterization of deposition velocity for ozone was available. In the course of both case studies, the meteorological fields obtained from ALADIN numerical weather prediction model have been used. Detailed model description with input data are described in the first part of this paper (*Lagzi et al.,* 2006).

Model simulations have shown that the transport model can predict effectively the dispersion of radioactive or chemically toxic substances, and when coupled to the deposition model, it could provide a suitable tool for estimating the ozone load on different surfaces (*Lagzi et al.,* 2004a, b). The paper presents the predicted stomatal ozone flux (the effective ozone load) over Hungary and compares it to the formerly used AOT 40 value (accumulated ozone exposure over a threshold of 40 ppb) for various weather conditions.

2. Methods

A detailed model description specifying both transport and deposition processes is given in *Lagzi et al.* (2006). Some additional details related to the results and sensitivity analyses are presented here.

2.1 Modeling accidental release

Modeling the accidental release of radioactive or chemically toxic substances requires the study of long-range transport from a single concentrated emission source. The path of the resulting plume should be predicted, along with its time of arrival to populated locations, and the possible levels of exposure to pollutants or deposition over a potentially large area.

The Chernobyl release provided a large impetus for the development of accidental release models, and several intercomparisons between different model types have since been made. The predominant model types are usually in either the Lagrangian or Eulerian framework. The former trace air masses, particles with assigned mass, or Gaussian shaped puffs of pollutants along trajectories determined by the wind-field structures. Lagrangian models have the advantage that they can afford to use high spatial resolution, although they rely on the interpolation of meteorological data. Their potential disadvantages are that in some cases they neglect important physical processes and often experience problems when strongly diverging flows lead to uncertainties in long-range trajectories. Eulerian models use grid based methods and have the advantage that they may take into account fully 3D descriptions of the meteorological fields, rather than single trajectories. However, when used traditionally with fixed meshes, Eulerian models show difficulty in resolving steep gradients. This causes particular problems for resolving dispersion from a single point source, which will create very large gradients near the release. If a coarse Eulerian mesh is used, the release is immediately averaged into a large area, which smears out the steep gradients and creates a large amount of numerical diffusion (*Lagzi et al.*, 2004b).

2.2 Modeling ozone concentration and ozone flux

Another potential application of the coupled model is an estimation of ozone concentration fields together with ozone flux fields over Hungary. As ozone basically reacts with vegetation through the stomata, for the estimation of effective ozone load, a more appropriate measure than former ones is the stomatal flux (e.g., *Musselman et al.,* 2006). Therefore, the stomatal part of total ozone flux has also been calculated. Since we assumed that the flux is constant between the reference height and the top of the canopy, the total flux can be written as follows:

$$F_t = c_r (R_a + R_b + R_c)^{-1} = c_c R_c^{-1}, \tag{1}$$

where c_r is the ozone concentration at a reference height (within the surface layer of the model, represented by the value at the top of the canopy), c_c is the concentration at the top of the canopy, R_a, R_b, and R_c are the aerodynamic resistance, the quasi-laminar boundary layer resistance, and the canopy resistance, respectively (*Lagzi et at.,* 2006). For estimating stomatal ozone flux, the stomatal part of total flux at the canopy top level can be written:

$$F_{st} = c_c R_{st}^{-1}, \tag{2}$$

where R_{st} is the stomatal resistance (also specified in *Lagzi et al.,* 2006). Accordingly, from Eqs. (1)–(2) the stomatal flux is calculated separately:

$$F_{st} = F_t R_c R_{st}^{-1}. \tag{3}$$

The coupled transport-deposition model was applied for a simulation period of July 22, 00:00 - July, 23, 00:00, 1998. This case study was chosen since during the selected days, the high temperature, low cloud cover, and low wind speed resulted in high photo-oxidant levels in Hungary. The initial mixing ratios of the major species were 0.4 ppb for NO_2, 2.0 ppb for NO, 80 ppb for O_3, and 4.1 ppb for VOC, which corresponded to typical daytime species concentrations. The initial concentrations were assumed to be equal in each layer across the whole simulated domain. For Budapest, the emission inventories for CO, NO_x, and VOCs were provided by the local authorities with a spatial resolution of 1 km × 1 km including the most significant 63 emission point sources. For Hungary, the National Emission Inventory with spatial resolution of 20 km × 20 km was applied which included both area and point sources. *Fig. 1* shows the emission inventories of NO_x for Budapest and Hungary. Outside Hungary, the emission inventory of EMEP for CO, NO_x, and VOCs was used, having a spatial resolution of 50 km × 50 km.

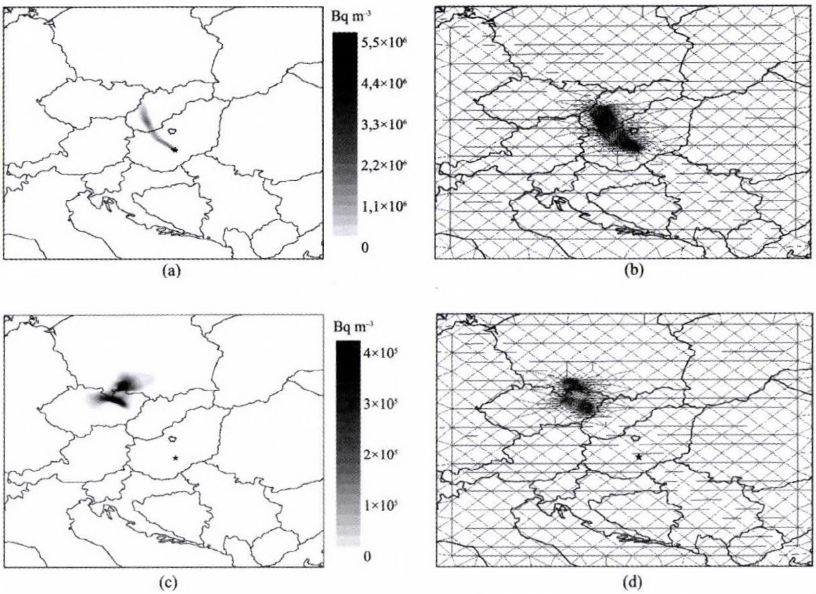


Fig. 1. Variation of the grid structure and activity of isotope ^{131}I (adaptive grid calculation). Simulation started at August 2, 1998, 00:00 UTC. (a), (b) Activity on the surface layer at t_0+12 and t_0+24. (c), (d) The adaptive mesh at t_0+12 and t_0+24.

3. Model results

3.1 Accidental release

The features of the model are illustrated by the simulation of a hypothetical nuclear accident on August 2, 1998, 00:00 UTC in the Paks NPP. The release of 2.985 kg ^{131}I isotope is assumed for 12 hours. This isotope decays to the stable ^{131}Xe with a half-life of 6.948×10^5 s. The initial grids for the adaptive and coarse grid calculations have been chosen such a way, that the typical length of a triangle edge is 106 km, and around the Paks NPP a somewhat finer resolution has been used. The modeled area includes Hungary and covers the neighboring countries within about 600 km from the border to all directions. The application of adaptive gridding methods was compared to the application of fixed grids for the hypothetical release described above. Two different fixed grid schemes were tested:

- The initial grid was not refined during the calculations.
- A high resolution (triangle edge length 6.6 km) fixed grid was used within the whole area.

Figs. 1a, b show the simulated surface layer activity of isotope ^{131}I using adaptive gridding. A continuous release was assumed in the first 12 hours and, therefore, there is a continuous plume after 12 hours (*Fig. 1a*). After 12 hours (t_0+24 h), the cloud is separated from the source and travels towards the northwest (*Fig. 1b*). *Figs. 1c, d* show that the region of increased grid resolution continuously follows the path of the contaminated air. The typical grid size in the non-contaminated area remained to be approximately 106 km, but in the highly contaminated area it was automatically reduced to 6.6 km (the minimum allowed length at the simulation) by the transient adaptation routine, allowing better spatial resolution in critical areas. *Fig. 2* compares the simulation results using the two fixed grid schemes at simulation times 12 and 24 h after the accident. The fine grid calculation has the lowest numerical error and, therefore, these results are the basis of comparison for the other mesh strategies. The coarse grid calculations show high numerical diffusion at all times. The result is that the initial plume is smeared over a much wider area than in the fine grid simulation as shown in *Fig. 2*. The agreements between the adaptive grid solutions (*Figs. 1a, b*) and fine grid results (*Figs. 2a, b*) very close to each other. The adaptive grid simulation is significantly closer to the fine grid calculation than the coarse grid prediction even after 24 hours.

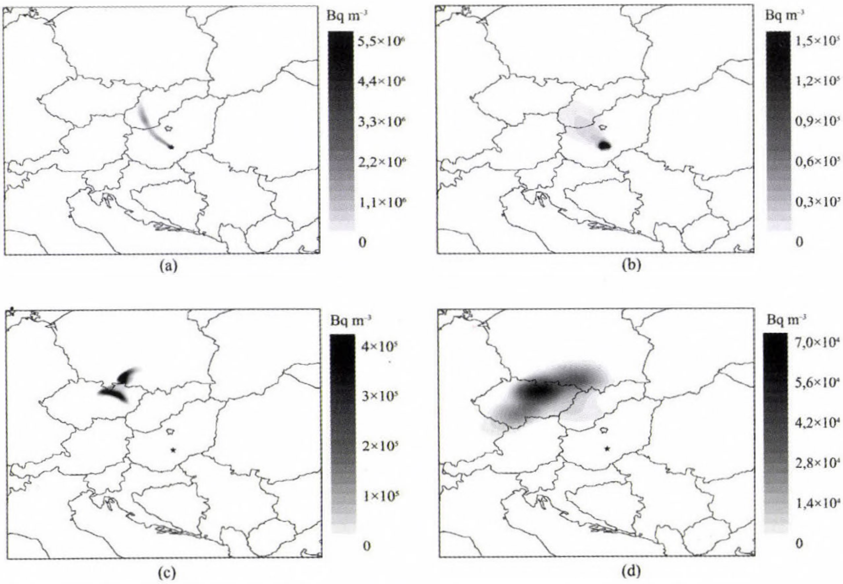


Fig. 2. Surface layer activity of isotope ^{131}I using two different fixed grid size schemes. Simulation started at August 2, 1998, 00:00 UTC. (a), (b) Activity on the surface layer at t_0+12 and t_0+24 using fine grid. (c), (d) Activity on the surface layer at t_0+12 and t_0+24 using coarse grid.

The CPU time requirements are 1, 30, 532 for the coarse, adaptive, and fine grid models, respectively. However, the fine and adaptive models provided similar results, but the application of the latter required 17.5 times more computer time. The implications could be that with given limited computer resources, the adaptive model provides reliable results quickly thus allowing ample time for emergency response. The fine grid model would provide similar results, but possibly too late.

3.2 Estimation of effective ozone load

Two different measures related to ozone load are estimated in this study. A cumulative amount of ozone mixing ratio over a 40 ppb threshold value (AOT 40) and cumulative stomatal flux over a regular grid for a chosen day (July 23, 1998) are presented in *Figs. 3* and *4*. In both cases values have only been considered, when the global radiation was above 50 W m^{-2}, because the harmful effects of ozone on vegetation mainly occur during the daytime, when stomata are open. During the selected day, the high temperature, low cloud cover, and low wind speed resulted in high photo-oxidant levels in Hungary. The highest ozone mixing ratios and, therefore, the highest values of AOT 40 (over 500 ppb h day^{-1}) are obtained in the north-western and eastern parts of Hungary (*Fig. 3*). Elevated ozone doses (over 150 ppb h day^{-1}) are also present to the southeast of the city of Budapest, albeit to a lower degree, due to the formation of a plume from emissions there. At the same time, the threshold value was not exceeded on this day in the city of Budapest, because of high concentrations of nitric oxide in the urban atmosphere, which titrates a large proportion of the ozone transported into the city.

In contrast with the spatial distribution of AOT 40 values, the map of cumulative stomatal flux (*Fig. 4*) shows a different feature. In this latter case the effects of both concentration and deposition fields are apparent in the results. The highest flux values have been estimated in the north-eastern and south-western parts of Hungary and also in some hilly regions (over 4.5 $\mu g\ m^{-2}\ day^{-1}$). There are no stomatal flux over grid cells, where the dominant surface type is water or urban. On the particular day, the influence of deposition processes on stomatal flux seemed stronger than the effect of concentration, in spite of relatively high ozone levels. The weather situation and soil properties through the stress effects on plants can retard the deposition. Therefore, in some cases lower amounts of ozone can be settled from the atmosphere, even if the ozone concentration is elevated. This difference can be seen by comparing the maps in *Figs. 3* and *4*, especially in the north-western part of Hungary.

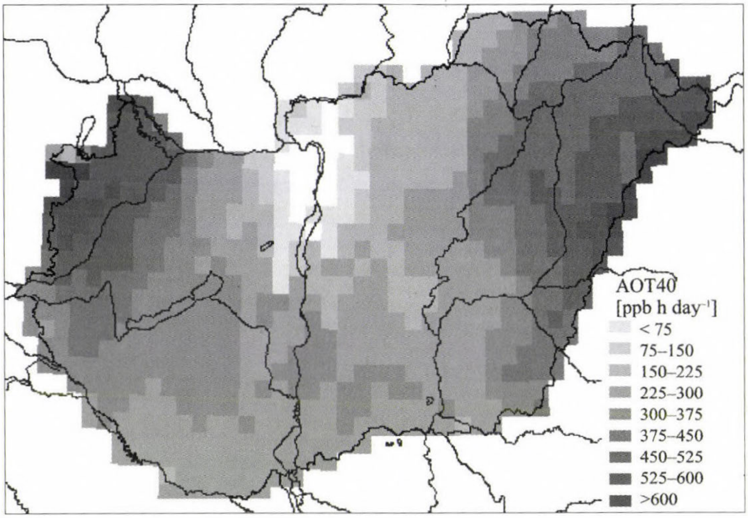


Fig. 3. Calculated AOT 40 index over Hungary on July 23, 1998 (sum of hourly values, when global radiation was greater than 50 W m^{-2}).

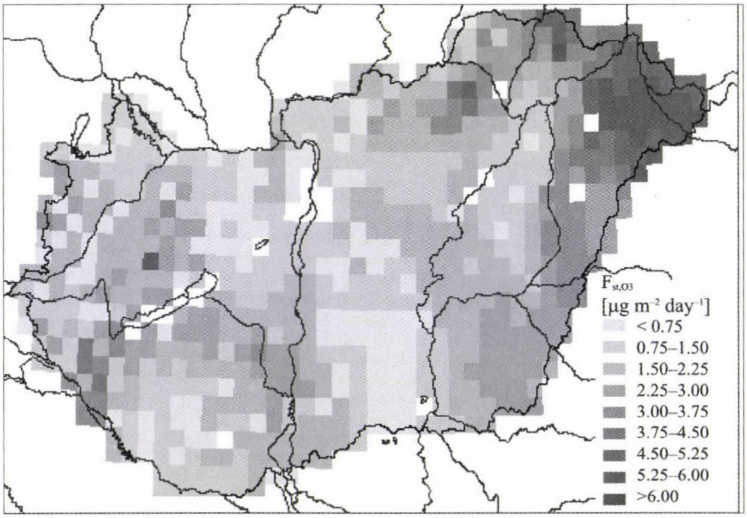


Fig. 4. Calculated cumulative stomatal flux over Hungary on July 23, 1998 (sum of hourly values, when global radiation was greater than 50 W m^{-2}).

Since ozone enters the plant through the stomata, the plant response and, therefore, the effective ozone load is more closely related to the stomatal ozone flux than to the atmospheric concentrations. Although these results refer to a single day case study, they serve to illustrate the difference between ozone concentration and stomatal ozone flux. Further conclusions about the importance of stomatal flux over a longer period, such as a growing season, will be the subject of future investigations.

4. Sensitivity analyses

To analyze the effects of atmospheric conditions both on transport and deposition processes, model calculations have been carried out using different initial meteorological conditions.

During the transport processes of air pollutants, one of the most important factors is the vertical structure of the atmosphere. Extension of mixing layer determines the volume where dispersion can occur, and estimation of height of the planetary boundary layer (e.g., using radiosonde data by analyses of the vertical temperature profiles) is a crucial process, because it may contain 20% error. Therefore, it is important to know how the perturbation of mixing height affects the spatial distribution of ozone concentration. The maxima of the mixing height during the simulations were homogeneous and constant, and were perturbed by ± 10% and ± 20% (*Table 1*). Simulations were started at July 22, 1998, and we investigated the differences between the original and perturbed outputs (*Fig. 5*). The figure shows that there are no exact connections between the detected ozone concentration and the height of the mixing layer due to the highly non-linear chemical reactions of the ozone (as a secondary pollutant) with other chemical species. The height of the mixing layer determines the volume, where the chemical reactions occur. Increasing the height produces a higher volume, this involves that the concentration of pollutants is decreased. Nevertheless, these concentrations are coupled non-linearly with each other. In such a way, we cannot state anything about the relation between the mixing layer height and concentration of the secondary pollutants. For example, a modification of mixing layer with +10% can cause either increased or decreased level of ozone concentration in unpredictable way.

Table 1. Perturbed mixing layer heights (m) used by the model

Cases	**Perturbations (%)**				
	-20	**-10**	**Original**	**10**	**20**
July 22, 1998 (maximum at 12:00 UTC)	1560	1755	1950	2145	2350
July 23, 1998 (maximum at 12:00 UTC)	1604	1805	2005	2206	2406

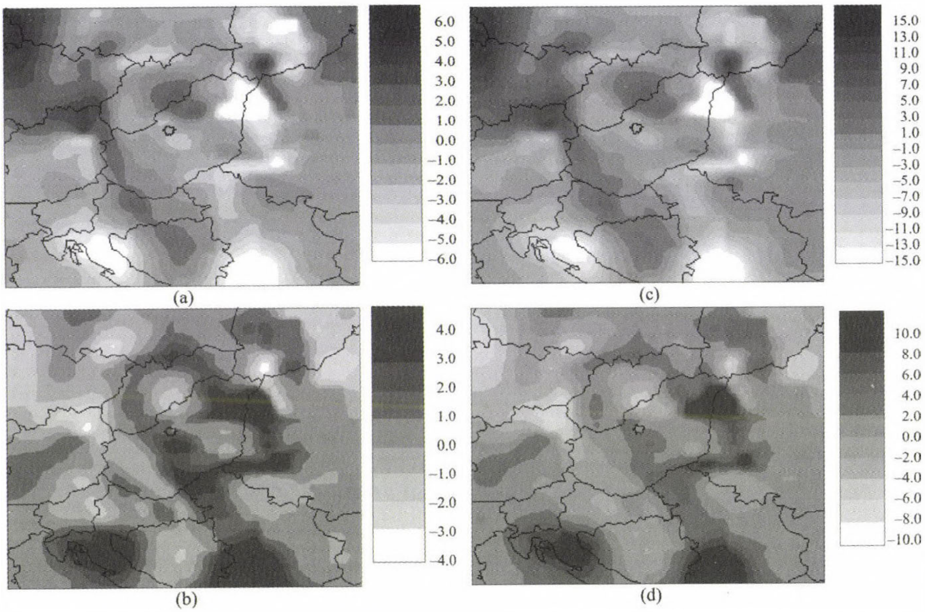


Fig. 5. Differences of the ozone mixing ratios between the perturbed and original cases at July 23, 1998, 16:00 UTC: (a) -10%, (b) +10%, (c) -20%, and (d) +20% of the original height of the mixing layer. Values are in ppb.

The deposition processes depend on the local weather conditions, surface and soil type, as well as plant physiological state. Based on previous investigations (*Lagzi et al.,* 2004a, 2006), it seems that root-zone soil water content plays an important role in deposition in continental regions, where soil water deficiency can strongly reduce the stomatal conductance and so the ozone flux through it. In this study, the effects of other input meteorological data have also been analyzed. For this purpose, the combined model was tested for 12:00 UTC of a sunny, summer day (July 23, 1998). The effects of changing certain meteorological data one by one on the deposition velocity of ozone were estimated. The data considered were the temperature, relative humidity, wind speed, root-zone soil moisture, and global radiation. Average changes of deposition velocity over different vegetation (grass, agricultural field, orchard, deciduous and mixed forest, respectively) on the given day are presented in *Figs. 6a–e*. Each input data have been modified individually from 20 to 140% of the value used in the initial calculations, in steps of 20%. The curves represent the effect of every single meteorological data on deposition velocity. It must be stressed, that this result is only valid for a single day, and the effects of meteorological data could be very different in case of other situations. For this hot, sunny summer day, results show that under continental climate conditions, soil state can be a crucial factor in determining the extent of stomatal ozone deposition. For every surface type, an increased value of soil moisture content results in greater deposition. However, the effect of soil moisture is lower in case of forests. High temperatures also retard the deposition, because for all vegetation types these high values are far from the optimal ones, where the stomatal uptake is not disturbed by temperature stress. A maximum appears at the optimal temperature, and a further decrease in temperature causes a decrease in the deposition velocity.

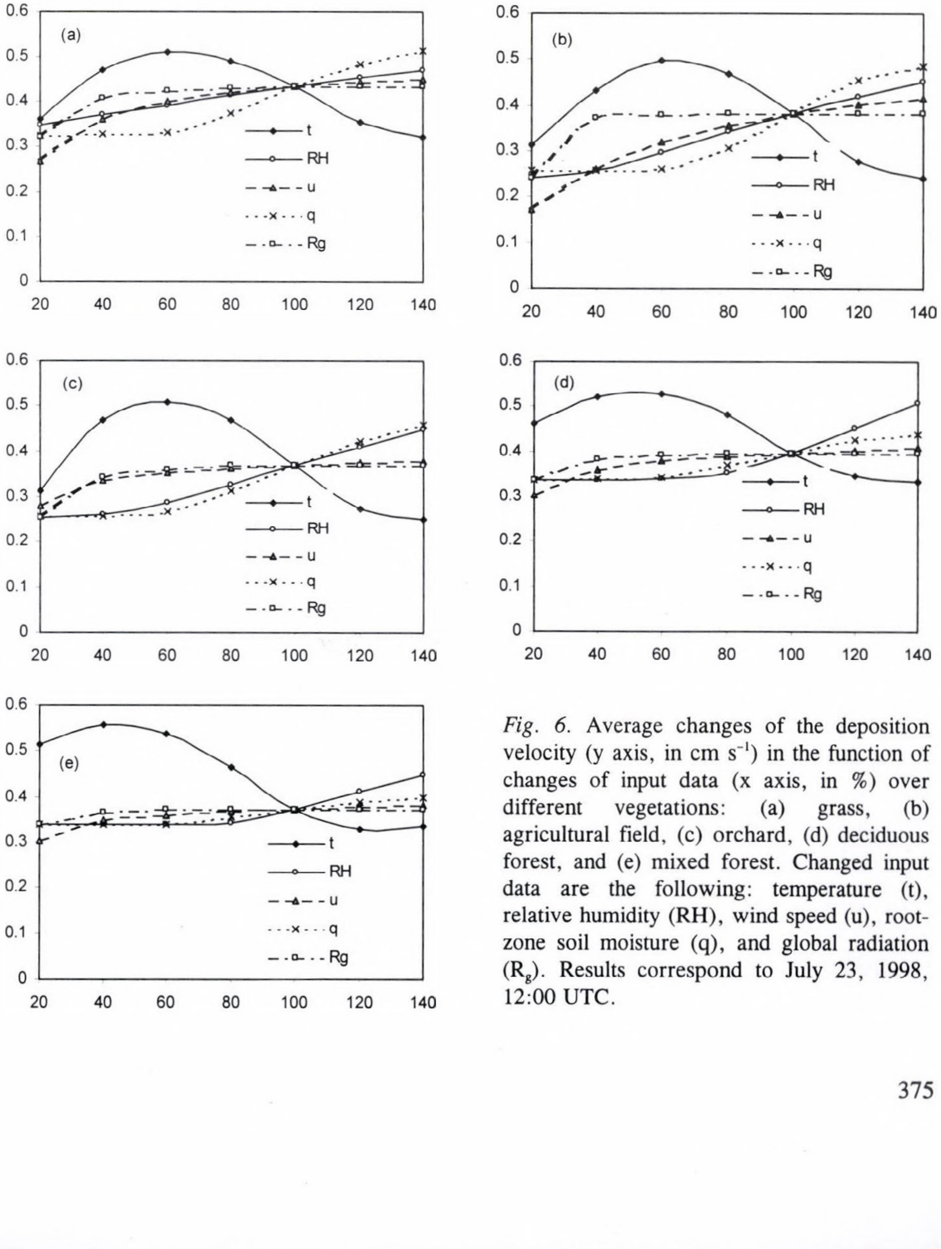


Fig. 6. Average changes of the deposition velocity (y axis, in cm s^{-1}) in the function of changes of input data (x axis, in %) over different vegetations: (a) grass, (b) agricultural field, (c) orchard, (d) deciduous forest, and (e) mixed forest. Changed input data are the following: temperature (t), relative humidity (RH), wind speed (u), root-zone soil moisture (q), and global radiation (R_g). Results correspond to July 23, 1998, 12:00 UTC.

The effect of relative humidity through the vapor pressure deficit could be also an important factor in deposition processes. The greater the vapor pressure deficit, the lower stomatal conductance can occur. This effect depends on the vegetation type, but is greater in the case of agricultural fields and orchards.

Wind speed affects the aerodynamic resistance and low values can therefore decrease the deposition velocity. In this case the weak turbulence retards the flux from the atmosphere.

Finally, the effect of global radiation has been studied. On the selected day, the sky was clear and, therefore, the global radiation reached its maximum value. Consequently, only decreases in global radiation have been analyzed. This causes a decrease in deposition velocity, as with global radiation the photosynthetically active radiation also decreases, and this term can influences the stomatal aperture.

5. Conclusion

A chemical transport model and a detailed dry deposition model were coupled for the purpose of simulating the effects of an accidental release from Paks NPP and estimating the effective ozone load over model calculations focused on Hungary. The meteorological data utilized were produced by the mesoscale weather prediction model ALADIN used by the Hungarian Meteorological Service. Two case studies have been shown in this study. A transport process of radioactive substances and two measures related to ozone load, namely the AOT40 and the stomatal ozone flux have been calculated on July 23, 1998.

An adaptive Eulerian grid model based on triangular unstructured grids describing the dispersion of radionuclides has been developed. The model automatically places a finer resolution grid in regions characterized by high spatial numerical errors and, therefore, the fine resolution grid automatically follows the spatial concentration gradients resulting from the passage of contaminated air over a given region. This approach allows the achievement of grid resolutions of the order of 6 km without excessive computational effort.

Calculated fields of AOT 40 and cumulative stomatal flux for the selected day underlie the differences between these two parameters. Ozone, deposited from the atmosphere into the plants through the stomata and, therefore, the effective ozone load, is more closely related to the stomatal ozone flux than to the atmospheric concentrations. Based on sensitivity analyses, this term does depend on weather situation. In this study we have shown this dependence only for a summer day. Results show, that under continental summer climate conditions, for a hot, cloudiness day the soil wetness state, air temperature, and humidity do influence the stomatal uptake.

In the future, it is planned to make more detailed sensitivity analyses and couple the transport-deposition model with the ALADIN meso-scale limited area numerical weather prediction model to estimate ozone deposition over Hungary for a routine application.

Acknowledgements—The authors acknowledge the support of OTKA grant D048673 (Postdoctoral Fellowship), OTKA F047242, NKFP 3A/088/2004, and the Békésy György Fellowship.

References

Ferenczi, Z. and *Ihász, I.*, 2003: Validation of the Eulerian dispersion model MEDIA at the Hungarian Meteorological Service. *Időjárás 107,* 115–132.

Lagzi, I., Mészáros, R., Horváth, L., Tomlin, A., Weidinger, T., Turányi, T., Ács, F., and *Haszpra, L.* 2004a: Modelling ozone fluxes over Hungary *Atmos. Environ. 38,* 6211–6222.

Lagzi, I., Kármán, D., Tomlin, A.S., Turányi, T., and *Haszpra, L.* 2004b: Simulation of the dispersion of nuclear contamination using an adaptive Eulerian grid model. *J. Environ. Radioactiv. 75,* 59–82.

Lagzi, I., Mészáros, R., Ács, F., Tomlin, A.S., Haszpra, L., and *Turányi, T.* 2006: Description and evaluation of a coupled Eulerian transport-exchange model. Part I. Model development. *Időjárás 110,* 349-363.

Mészáros, R., Szinyei, D., Vincze, Cs., Lagzi, I., Turányi, T., Haszpra, L., and *Tomlin, A.S.,* 2006: Effect of the soil wetness state on the stomatal ozone fluxes over Hungary. *Int. J. Environ. Pollut.* (in press).

Musselman, R.C., Lefohn, A.S., Massman, W.J., and *Heath, R.L.* 2006: A critical review and analysis of the use of exposure- and flux-based ozone indices for predicting vegetation effects *Atmos. Environ. 40,* 1869-1888.

Steib, R. and *Labancz, K.*, 2006: Regulatory modeling in Hungary - the AERMOD model. Part I. Description and application. *Időjárás 109*, 157–172.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 379–395

Application of the operator splitting method for real-life problems

István Faragó

Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University
P.O. Box 120, H-1518 Budapest, Hungary; E-mail: faragois@cs.elte.hu

(Manuscript received in final form June 13, 2006)

Abstract—In the modeling of real-life complex time-dependent phenomena, the simultaneous effect of several different sub-processes has to be described. The operators describing the sub-processes are as a rule simpler than the whole spatial differential operator.

Operator splitting is a widely used procedure in the numerical solution of such problems. The point in operator splitting is the replacement of the original model with one in which appropriately chosen groups of the sub-processes, described by the model, take place successively in time. This de-coupling procedure allows us to solve a few simpler problems instead of the whole one.

In this paper several splitting methods are constructed (sequential splitting, Strang-Marchuk splitting, weighted splitting, additive splitting, iterated splitting) and analyzed. Application of the operator splitting method to real-life problems is investigated, with great emphasis on long-range air pollution transport. The accuracy (local splitting error) of the methods is discussed, and the main advantages and drawbacks of this approach are listed.

Key-words: air pollution modeling, operator splitting, numerical solution

1. Introduction

The operator splitting method (OSM) is a widely used method for solving real-life problems. It can be applied in the numerical modelling process of many different time-dependent complex physical phenomena. It is widely used in different fields of real-life problems, such as advection-diffusion problems (see e.g., *Karlsen et al.*, 2001; *Marinova et al.*, 2003), the Hamilton-Jacobi equation (see e.g., *Jakobsen et al.*, 2001; *Karlsen* and *Risebro*, 2002), the Navier-Stokes equation (see *Christov* and *Marinova*, 2001), the modeling of turbulence and interfaces (see *Mimura et al.*, 1984), the Maxwell equations (*Horvath*, 2005), and the shallow water equations (*Havasi*, 2005). More applications can be found in *Karlsen et al.* (2001). In this paper we focus our attention to air pollution modeling.

The mathematical model can be described as follows. The original continuous mathematical model of such phenomena can be described in the form of an abstract Cauchy problem (ACP) as follows:

$$\begin{cases} \dfrac{du(t)}{dt} = \sum\limits_{i=1}^{n} A_i u(t)\,, \quad t \in (0,T]. \\ u(0) = u_0\,. \end{cases} \tag{1}$$

In the above formulation we have used the following notation. X denotes a normed (Banach) space, $u : \mathbf{R} \to \mathrm{X}$ is the unknown function, $u_0 \in \mathrm{X}$ is a given element which describes the initial state of the process, A_i (for $i = 1, 2, \ldots n$) are given (usually densely defined linear) operators of type $\mathrm{X} \to \mathrm{X}$, which correspond to the *n* different sub-problems. We note that typically X is a space of sufficiently smooth functions, which is always defined by the concrete task under consideration. (For more detailed description of the mathematical notion used throughout this paper, we refer e.g., to *Lax* (2002).)

Frequently, the OSM is applied to the so called semi-discretized problem. This means that the original continuous problem has already been discretized with respect to the space variables. Then Eq. (1) denotes a system of ordinary differential equations, and in this case $\mathrm{X} = \mathbf{R}^{\mathrm{N}}$, where N is the number of grid points in the space discretization process. Clearly, when we study linear problems, A_{i} are matrices from $\mathbf{R}^{\mathrm{N}\times\mathrm{N}}$.

The aim of this paper is to give an overview of the different operator splitting methods and classify the advantages and drawbacks of this approach. The paper is organized as follows.

Section 2 formulates the air pollution model as one of the basic models where the operator splitting method is applied. Section 3 deals with the motivation of the application of this approach. The next section describes the different splitting methods and compares them. In Section 5 the main attractive properties of this approach are listed, while Section 6 does the same for the drawbacks. The paper is finished with some conclusions.

2. An important example: air pollution modeling

The transport of air pollutants is one of the most widely investigated phenomena for which Eq. (1) can serve as a mathematical model.

Let $c_j = c_j(\mathbf{x}, t)$ denote the concentration of the j-th air pollutant, and c the vector function of these functions. Then the time evolution of the vector c can be described mathematically by the system of partial differential equations (*Zlatev*, 1995):

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \nabla(K\nabla c) + E - \sigma c + R(c) & t \in (0,T] \\ c(\mathbf{x},0) = c_0(\mathbf{x}) \end{cases}, \tag{2}$$

where $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ is a vector function describing the wind velocity, $K = K(\mathbf{x}, t)$ is the diffusion coefficient function, $E = E(\mathbf{x}, t)$ is the function of emission, $\sigma = \sigma(\mathbf{x}, t)$ describes the deposition, and R defines the chemical reactions of the pollutants. The initial function $c_0(x)$ is given. Using this notation, the terms in Eq. (2) have the following physical meaning. The first term on the right-hand side describes the transportation due to the velocity field, which is called *advection*. The second term expresses the *turbulent diffusion*, the third term denotes the *emission*, the fourth term describes the *deposition*, and the last term defines the *chemistry* of the pollutants.

It is quite natural to define the different sub-operators on the base of the separate physical processes, namely, we can define the following operators:

- $A_1 c = -\nabla(\mathbf{u}c)$ is the advection operator,
- $A_2 c = \nabla(K\nabla c)$ is the diffusion operator,
- $A_3 c = -\sigma c$ is the deposition operator,
- $A_4 c = E$ is the emission operator,
- $A_5 c = R(c)$ is the chemistry operator.

There are many other possible applications, e.g., the shallow water equations (e.g., *Havasi*, 2005), the Maxwell equations (e.g., *Horvath*, 2005), etc.

3. A motivation for the use of the operator splitting method

The number of chemical species involved in a modern air pollution model sometimes reaches 200, or even more, which results in a large system of partial differential equations. The analytical solution of such a problem is obviously very difficult and expensive to find. Hence we have to treat it numerically. We note that in case of semi-discretization, usually the number of spatial grid points equals many millions. This means that the system of ordinary differential equations obtained after spatial discretization is extremely big, hence the use of any numerical method developed for systems of ordinary differential equations is rather complicated. Moreover, the model equations contain terms that have different physical meanings and so different mathematical properties (e.g., linear, non-linear, stiff, and non-stiff). Therefore, it is impossible to find such a universal numerical method, which would perform well when applied directly to the original system. The application of operator splitting allows us to treat the different physical terms separately.

The operator splitting method (OSM) is a kind of problem decomposition: the spatial differential operator of the global system is divided into a few simpler operators, and the corresponding problems are solved one after the other, by connecting them through their initial conditions.

The simpler systems, which are obtained in this manner and sometimes called sub-systems, might have some special properties that can be exploited in the numerical solution. The sub-systems are usually easier to treat numerically than the whole system.

Splitting can be performed in several ways. We expect the method to be accurate as well as efficient enough. The latter property depends on the number of computations and the possibility of performing the computations in parallel. Taking into account the latter requirement, we made attempts to construct a new splitting scheme, which does not require a lot of computational work, and is parallelizable on the operator level.

4. Description of the OSM

In the sequel the frequently used splitting methods are described and compared. We describe the methods only for two operators (i.e., $n=2$), however, the generalization for n operators is straightforward. (For more details, see *Hundsdorfer* and *Verwer*, 2003; *Zlatev*, 1995; *Dimov et al.*, 2006.) Hence, we consider the ACP

$$\begin{cases} \dfrac{du(t)}{dt} = Au(t) + Bu(t), & t \in (0,T] \\ u(0) = u_0, \end{cases} \tag{3}$$

a discretization method, where we replace the continuous (in time) problem, Eq. (2) with seeking the split (discretized) solution on the grid points of the mesh

$$\omega_\tau = \{t_j = j\tau, j = 0, 1, \ldots \mathrm{M}\}, \qquad \text{where } \mathrm{M}\tau = T. \tag{4}$$

Here $\tau > 0$ denotes the splitting time step and hence M denotes the number of the grid points. In the following we summarize some widely used splitting methods.

4.1 Sequential splitting

The scheme of this method is the following. As a first step, we solve the system with operator A using the initial condition of the original problem, and then, applying the obtained solution at time τ as an initial condition, we solve the system with operator B.

The solution obtained in this way is considered as the splitting solution in τ. This procedure is performed cyclically in the following way:

$$\begin{cases} \dfrac{du_k^{(1)}(t)}{dt} = Au_k^{(1)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(1)}((k-1)\tau) = u_{k-1}^{(2)}((k-1)\tau), \end{cases} \tag{5}$$

$$\begin{cases} \dfrac{du_k^{(2)}(t)}{dt} = Bu_k^{(2)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(2)}((k-1)\tau) = u_k^{(1)}(k\tau), \end{cases} \tag{6}$$

for $k=1, 2,\ldots,$ m, where $u_0^{(2)}(0) = u_0$. (Here the superscripts stand for the notation of the new unknown functions.) The splitting solution at time $t=k\tau$ is defined as

$$u_{\text{spl}}(k\tau) = u_k^{(2)}(k\tau), \qquad \text{for } k = 1, 2,\ldots, \text{M}. \tag{7}$$

4.2 Strang-Marchuk splitting

Using this method, at each time step we begin and end the computation with operator A (we apply it over the distance $\tau/2$ twice) and put B to the middle (we apply it over the distance τ once) as follows:

$$\begin{cases} \dfrac{du_k^{(1)}(t)}{dt} = Au_k^{(1)}(t), \quad t \in ((k-1)\tau, (k-\tfrac{1}{2})\tau], \\ u_k^{(1)}((k-1)\tau) = u_{k-1}^{(3)}((k-1)\tau), \end{cases} \tag{8}$$

$$\begin{cases} \dfrac{du_k^{(2)}(t)}{dt} = Bu_k^{(2)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(2)}((k-1)\tau) = u_k^{(1)}((k-\frac{1}{2})\tau), \end{cases} \tag{9}$$

$$\begin{cases} \dfrac{du_k^{(3)}(t)}{dt} = Au_k^{(3)}(t), \quad t \in ((k-\frac{1}{2})\tau, k\tau], \\ u_k^{(3)}((k-\frac{1}{2})\tau) = u_k^{(2)}(k\tau), \end{cases} \tag{10}$$

for $k=1, 2, \ldots, M$, where $u_0^{(3)}(0) = u_0$. The splitting solution at $t=k\tau$ is defined as

$$u_{\text{spl}}(k\tau) = u_k^{(3)}(k\tau), \qquad \text{for } k = 1, 2, \ldots, M. \tag{11}$$

4.3 Weighted sequential splitting

The sequential splitting is not symmetric for the ordering of the operators. Therefore, we can use different orderings of the operators A and B. The idea of this method is based on the following: we compute the sequentially split solution in both orderings, and then their average value is taken for the splitting solution. However, we obtain a symmetric algorithm in the following way: in each time step we apply sequential splitting both in the order $A \rightarrow B$ as follows:

$$\begin{cases} \dfrac{du_k^{(1)}(t)}{dt} = Au_k^{(1)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(1)}((k-1)\tau) = u_{\text{spl}}((k-1)\tau), \end{cases} \tag{12}$$

$$\begin{cases} \dfrac{du_k^{(2)}(t)}{dt} = Bu_k^{(2)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(2)}((k-1)\tau) = u_k^{(1)}(k\tau), \end{cases} \tag{13}$$

and B → A as follows:

$$\begin{cases} \dfrac{dv_k^{(1)}(t)}{dt} = Bv_k^{(1)}(t), \quad t \in ((k-1)\tau, k\tau], \\ v_k^{(1)}((k-1)\tau) = u_{\text{spl}}((k-1)\tau), \end{cases} \tag{14}$$

$$\begin{cases} \dfrac{dv_k^{(2)}(t)}{dt} = Av_k^{(2)}(t), & t \in ((k-1)\tau, k\tau], \\ v_k^{(2)}((k-1)\tau) = v_k^{(1)}(k\tau). \end{cases} \tag{15}$$

Then the split solution at $t=k\tau$ is defined as

$$u_{\text{spl}}(k\tau) = \theta u_k^{(2)}(k\tau) + (1-\theta)v_k^{(2)}(k\tau), \qquad \text{for } k = 1, 2, \ldots, \text{M}, \tag{16}$$

where $u_{\text{spl}}(0) = u_0$, and $\theta \in [0,1]$ is some fixed weight parameter. Clearly, we obtain a symmetric algorithm when $\Theta = 0.5$. This weighted sequential splitting is called symmetrically weighted sequential splitting. An important property of the weighted sequential splitting is that it can be paralleled in a natural way (on the operator level), because the split sub-problems, Eqs. (12)–(13) and Eqs. (13)–(14) are using the same initial value $u_{\text{spl}}((k-1)\tau)$. The theoretical investigation of the method can be found in *Csomós et al.* (2005), and an application of the method in a one-column transport-chemistry model is described in *Botchev et al.* (2003).

4.4 Other kinds of splittings

There are also other OSM's which has been recently developed. We want to stress two methods of them, namely, the additive splitting and the iterated splitting.

The additive splitting is similar to the sequential splitting with the following difference: in both sub-problems (Eqs. (5) and (6)) the same initial value $u_{\text{spl}}((k-1)\tau)$, i.e., the split solution at the previous time level is used. If $u_k^{(A)}(k\tau)$ and $u_k^{(B)}(k\tau)$ denotes the corresponding solutions, then the split solution at the new time level is defined as

$$u_{\text{spl}}(k\tau) = u_k^{(A)}(k\tau) + u_k^{(B)}(k\tau) - u_{\text{spl}}((k-1)\tau), \quad \text{for } k = 1, 2, \ldots, \text{M}. \tag{17}$$

The main advantage of this OSM is its easy parallelism, because both sub-problems are using the same initial value. (For more details we refer to *Gnandt* (2005).)

The iterated splitting suggests the following algorithm: on the interval $[t_{k-1}, t_k]$ we solve the following sub-problems consecutively, for $i=1,3,5, \ldots 2m+1$:

$$\begin{cases} \dfrac{du_k^{(i)}(t)}{dt} = Au_k^{(i)}(t) + Bu_k^{(i-1)}(t), & t \in ((k-1)\tau, k\tau], \\ u_k^{(i)}((k-1)\tau) = u_{\text{spl}}((k-1)\tau), \end{cases} \tag{18}$$

$$\begin{cases} \dfrac{du_k^{(i+1)}(t)}{dt} = Au_k^{(i)}(t) + Bu_k^{(i+1)}(t), & t \in ((k-1)\tau, k\tau], \\ u_k^{(i+1)}((k-1)\tau) = u_{\text{spl}}((k-1)\tau), \end{cases} \tag{19}$$

where the initial iterated function $u^{(0)}(t)$ is any fixed function for each iteration. (Usually it is chosen as a constant function, namely $u^{(0)}(t) = u_{\text{spl}}((k-1)\tau)$.) The split solution at $t = k\tau$ is defined as

$$u_{\text{spl}}(k\tau) = u_k^{(2m+1)}(k\tau), \qquad \text{for } k = 1, 2, \ldots, \text{M}. \tag{20}$$

We recall that in the above iteration the upper index i refers to the number of the iteration on the fixed k-th time interval.

4.5 Comparison of the different operator splitting methods

In the following, we compare the splitting methods presented above from the viewpoints of accuracy and computational costs.

Replacing the original problem, Eq. (1), with one of the above listed split models usually results in an error called *local splitting error*, which is defined as follows.

Let us denote the exact solution of Eq. (1) by $u(t)$, and the exact solution of the chosen split problem by $u_{\text{spl}}(\text{n}\tau)$, respectively. Their difference at the point $t = \tau$ is called local splitting error, i.e.,

$$Err_{\text{spl}}(\tau) = u(\tau) - u_{\text{spl}}(\tau). \tag{21}$$

When $Err_{\text{spl}}(\tau) = O(\tau^{p+1})$, then the splitting method is called p-th order. Generally, the increase of the order of the splitting results in a more accurate split solution. Therefore, the quality of a splitting can be characterized by its order, too. For the above discussed splitting methods the order can be computed directly, and it is

- one for the sequential splitting,
- one for the weighted sequential splitting with $\Theta \neq 0.5$,

- two for the Strang-Marchuk splitting,
- two for the symmetrically weighted sequential splitting,
- one for the additive splitting,
- $2m+1$ for the iterated splitting.

The sequential and the additive splittings have low accuracy, the Strang-Marchuk splitting, the weighted sequential splitting, and the iterated splitting are of higher accuracy.

Especially we stress, that the order of the accuracy of the iterated splitting method depends only on the number of the inner iteration. So, at least theoretically, one can achieve arbitrary high order accuracy. (However, the above orders of the OSM hold under the assumption that the split sub-problems are solved either exactly or with some higher order numerical method.) One of the further advantages of the iterated splitting is that each sub-problems result in a suitable (consistent) approximation to the exact solution.

An important feature of the splitting method is its complexity on modern parallel computers. Therefore, it is worth comparing the different splitting methods as discretization methods from this point of view, too.

The sequential splitting, the Strang-Marchuk splitting, and the iterated splitting are not parallelizable on the operator level. The other methods can be performed in parallel. Some of them (weighted sequential splitting, iterated splitting) are relatively expensive, while the additive splitting is relatively cheap. However, the latter is not accurate enough. We note that this remark concerns only the "natural parallelism", i.e., when the numerical method (which is usually required for solving the split sub-problems) are not taken into consideration.

5. Advantages of the operator splitting process

In this section we list those properties of an OSM, which make it attractive during the mathematical modeling process of complex physical phenomena.

5.1 Easier theoretical investigation of the convergence

In fact, he OSM can be viewed as a one-step time-discretization method. Hence, it is quite natural to raise the question: if the different sub-problems are solved exactly, under which conditions the split (discretized) solution is convergent to the exact solution when the discretization parameter (τ) tends to zero? The answer is based on the famous Lax equivalence theorem, which states that the consistency and stability together imply the convergence. (For more details we refer to *Lax* (2002).) This means that we should check only the above two simpler properties.

Performing this check for the globally discretized non-split problem is a very difficult (usually hopeless) task. If the order of the local splitting error is $p>0$, then the splitting procedure is normally consistent. Therefore, in this case we have to check only the stability. This latter property means that there exists a constant, independent of τ, such that the norm of the split solution at any time level cannot be bigger than the norm of the initial function multiplied by this constant (see *Havasi et al.*, 2001). This constant is called stability constant and it clearly cannot be less than one. When it is equal to one, then the method is called contractive. (This implies that the norm of the solution cannot grow in time.) Hence the stability of the split discretization method for the well-posed problem is a crucial practical problem. We note that if both sub-problems are stable, it does not yield automatically the stability of the whole split discretization method. However, when the sub-problems are contractive, then the total OSM is also contractive, and hence stable. (It is an easy exercise and left to the Reader.)

5.2 Choice of a suitable numerical method

The use of an OSM results in a sequence of sub-problems, which represent also Cauchy problems, but with simpler operators. However, these problems cannot be solved analytically, either. So, we should apply numerical methods to the sub-problems. The benefit of the OSM is that we can use different numerical methods to the different sub-problems according to the special features of the problems. After defining the numerical methods to each sub-problems, we can consider the global numerical algorithm as a discretization method to the original non-split continuous model given by Eq. (1). This approach gives us possibility to derive some well-known numerical schemes or create new methods. In order to demonstrate this statement, we give two examples. (For more details we refer to *Faragó* (2005).)

Example 1

We consider the ACP

$$\begin{cases} \dfrac{du(t)}{dt} = Au(t), \quad t \in (0,T], \\ u(0) = u_0 . \end{cases} \tag{22}$$

If operator A is represented as the following sum $A = 0.5A + 0.5A$, then the sequential splitting given by Eqs. (5)–(6) reads as follows

$$\begin{cases} \dfrac{du_k^{(1)}(t)}{dt} = 0.5Au_k^{(1)}(t), & t \in ((k-1)\tau, k\tau], \\ u_k^{(1)}((k-1)\tau) = u_{k-1}^{(2)}((k-1)\tau), \end{cases}$$

and

$$\begin{cases} \dfrac{du_k^{(2)}(t)}{dt} = 0.5Au_k^{(2)}(t), & t \in ((k-1)\tau, k\tau], \\ u_k^{(2)}((k-1)\tau) = u_k^{(1)}(k\tau). \end{cases}$$

If we choose the implicit Euler method to the first sub-problem and the explicit Euler method for the second sub-problem with discretization parameter $\Delta t = \tau$, then the global discretization method is

$$u_{\text{spl}}(k\tau) = (I + 0.5\Delta t A)(I - 0.5\Delta t A)^{-1} u_{\text{spl}}((k-1)\tau), \text{ for } k = 1, 2, \ldots, \text{M}, \quad (23)$$

which is the well-known Crank-Nicholson method.

Example 2

Let us consider Eq. (22) with the same sequential splitting. If the implicit Euler method is applied to both sub-problems with $\Delta t = \tau$, then the obtained new discretization method is

$$u_{\text{spl}}(k\tau) = (I - 0.5\Delta t A)^{-1}(I - 0.5\Delta t A)^{-1} u_{\text{spl}}((k-1)\tau), \text{ for } k = 1, 2, \ldots, \text{M}. \quad (24)$$

This approach makes it possible to increase the efficiency of the global algorithm.

5.3 Applicability of the existing software products

When we divide the original problem into a sequence of sub-problems, it is fairly reasonable to do it in such a way, that the split tasks were standard problems to which already existing software products (e.g., MATLAB library tool-boxes) are directly applicable. For instance, in the air pollution modeling the sub-problems, defined by the operators given in Section 2, are standard and can be treated by using library program routines. We note that the OSM gives a high flexibility in choosing the sub-operators. In the choosing process the "standardization" requirements formulated above should coincide with some other requirements, too. Here the most critical point is the suitable choice of the sub-operators, due to the problem with the boundary condition (see later Section 6.3).

5.4 Increase of the computational efficiency

The choice of the time-discretization parameter τ plays a crucial role in the efficiency of the applied numerical method. Usually, due to some stability conditions, it can not be chosen arbitrarily and there is an upper bound. However, our aim is to avoid this restriction, because a too small τ usually results in a lot of difficulties. In order to get the numerical solution on some fixed time level T, we should solve the sub-problems on T/τ time levels, which may get extremely big. But this latter fact causes some troubles:

- the computation work increases significantly,
- for stiff problems the numerical implementation is almost impossible,
- due to the big number of arithmetic operations, the computational errors may increase dramatically.

Hence, our aim is to get rid of the too strict restriction and to allow to choose τ as large as possible. We show that the OSM is a powerful method to increase the time-discretization parameter τ. In the following we consider an example.

Example 3

$$\frac{\partial u(x,t)}{\partial t} = (10^6 + \sin xt)\frac{\partial^2 u(x,t)}{\partial x^2}, \qquad t \in (0,T).$$

When the above problem is used with the explicit Euler method, then the condition of stability is

$$\frac{\tau}{h^2} \le \frac{1}{\max 2(10^6 + \sin xt)} \approx 0.5 \cdot 10^{-6},$$

which means that $\tau \approx 0.5 \cdot 10^{-6} h^2$. This yields that the choice $h=0.01$ results in $\tau \approx 0.5 \cdot 10^{-10}$ for the upper bound. (When $T=1$, then $2 \cdot 10^{11}$ time steps are required.) If we use the sequential splitting

$$\frac{\partial u_1(x,t)}{\partial t} = 10^6 \frac{\partial^2 u_1(x,t)}{\partial x^2}, \qquad t \in ((k-1)\tau, k\tau]),$$

$$\frac{\partial u_2(x,t)}{\partial t} = \sin(xt)\frac{\partial^2 u_2(x,t)}{\partial x^2}, \qquad t \in ((k-1)\tau, k\tau]),$$

which are connected via the initial condition (see Eqs. (5)–(6)), then we can observe that the first sub-problem can be solved by spectral method, which does not require any mesh. Hence, we need to construct the mesh only for the second sub-problem. Applying the explicit Euler method to this problem, the stability bound turns into

$$\frac{\tau}{h^2} \leq \frac{1}{2 \max \sin(xt)} \leq 0.5 .$$

This means that for the same fixed space-discretization parameter h we can select a 10^6 times bigger τ within the stability condition.

5.5 Use of numerical-analytical methods

In certain cases, even if the original problem cannot be solved analytically, after the splitting one of the sub-problems can be solved analytically. Hence, the possible choice of the time-discretization parameter τ can be enlarged again. This is demonstrated in the following example.

Example 4

$$\begin{cases} \dfrac{\partial u(x,t)}{\partial t} = (10^6 + \sin xt)\dfrac{\partial u(x,t)}{\partial x}, \\ u(x,0) = u_0(x), \end{cases}$$

for $t>0$, $x \in \Re$, where $u_0(x)$ is a given function. Using the explicit Euler finite difference scheme, the well-known CFL condition (e.g., *Richtmyer*, 1967) implies the bound

$$\frac{\tau}{h} \leq \frac{1}{10^6}.$$

Applying the sequential splitting we get the sub-problems

$$\frac{\partial u_1(x,t)}{\partial t} = 10^6 \frac{\partial u_1(x,t)}{\partial x}, \qquad t \in ((k-1)\tau, k\tau]),$$

$$\frac{\partial u_2(x,t)}{\partial t} = \sin(xt)\frac{\partial u_2(x,t)}{\partial x}, \qquad t \in ((k-1)\tau, k\tau]),$$

connected again via the initial condition.

As one can see, the first sub-problem can be solved analytically by using the D'Alambert formula, because the coefficient in the elliptic part is constant.

At the same time, using the explicit Euler method to the second sub-problem, for the bound of the method we obtain

$$\frac{\tau}{h} \leq 1.$$

This means that for some fixed h we could increase the possible choice of τ 10^6 times bigger.

5.6 Preservation of main qualitative properties

As we have already mentioned, the original physical phenomenon has a lot of basic qualitative properties, which are inherent to the physical process and originate from the physics of the process. The adequate continuous mathematical model also must have these properties. (E.g., in an air pollution model the concentrations of the pollutants cannot be negative; in the heat conduction phenomenon without source and with zero homogenous boundary condition, the non-negative initial heat distribution remains non-negative and decreasing in norm in time, etc.)

Therefore, it is natural to require that a "good" discrete model also has the discrete analogue of the above qualitative properties. For some standard models and discretization methods (e.g., heat equation, wave propagation with finite difference, and finite element method) this theory has been developed, and the conditions, under which the models are qualitative property preserving are known.

However, when we discretize the original non-split problem with a numerical method, then the obtained discrete model is not standard, and hence to check the validity of the discrete qualitative properties is a very difficult task.

Clearly, when we use an operator splitting method and numerical methods for the sub-problems which preserve the qualitative properties, then the global discretization method is also qualitative property preserving. Therefore, when we split the non-split problem into a sequence of "standard" sub-problems, we can give some sufficient conditions under which the qualitative properties are preserved.

6. Drawbacks of the operator splitting process

In the previous section we have listed some advantages of the operator splitting approach. However, this method has some disadvantages, too. In the following we list those problems which arise by use of the OSM.

6.1 The suitable choice of the sub-operators

The original complex physical problem consists of different – usually simpler – processes. In fact, when we consider the problem given by Eq. (1), then our mathematical model is

$$\begin{cases} \dfrac{du(t)}{dt} = Au(t) , \quad t \in (0,T], \\ u(0) = u_0 , \end{cases} \tag{25}$$

where the operator A describes the whole complex physical process. The partition of the operator A into the sum $A = \Sigma A_i$ is not always natural. The choice of the suitable sub-operators might be difficult. For instance, for the air pollution modeling the choice of the sub-operators, given in Section 2, is not unique. The operator and the sum can be decomposed:

$$\begin{aligned} Au &= -\nabla(\mathbf{u}c) + \nabla(K\,\nabla c) + E - \sigma c + R(c) , \\ Au &= \sum_{i=1}^{5} B_i u(t) , \end{aligned}$$

where now

- $B_1 c = -\sum_{i=1}^{2} \partial_i (u_i c)$ is the horizontal advection operator,
- $B_2 c = \sum_{i=1}^{2} \partial_i (k_i \partial_i c)$ is the horizontal diffusion operator,
- $B_3 c = -\sigma c$ is the deposition operator,
- $B_4 c = E + R(c)$ is the emission and chemistry operator,
- $B_5 c = \partial_3 (u_3 c) + \partial_3 (k_3 \partial_3 c)$ is the vertical transport operator.

This kind of decomposition is used in the Danish Eulerian Model (DEM) and is called DEM decomposition. The decomposition given in Section 2 is called physical decomposition.

The main advantage of the DEM decomposition is its high flexibility for 2D problems, because only the last operator contains the vertical part. As we can see, for the DEM the choice of an adequate boundary condition is more natural than for the physical splitting. However, the choice of an effective numerical method for the physical splitting is easier and more natural.

6.2 Error analysis

Replacing the original problem (Eq. (1)) with one of the above listed split models usually results in a new kind of error called local splitting error (see Eq. (21)).

Since the split sub-problems cannot be solved exactly, we should apply numerical methods to their solution. Hence, the obtained numerical result includes two kinds of errors: the local splitting error and the error of the numerical methods. The analysis of the interaction of these two errors is usually a very complicated task. Hence, the error analysis for the numerical solution obtained by use of some numerical method to the non-split problem is simpler, and we can control it more easily than for the split models.

We note that under some conditions the splitting error may disappear (for some splittings in case of commutativity of the operators), but these conditions are mostly unrealistic in real-life applications.

6.3 Handling the boundary conditions

When using the OSM, handling the boundary conditions for the split problems is a serious problem. More precisely, the question is: how to describe the boundary conditions for the different sub-problems of different types? E.g., the simplified diffusion-advection model of Eq. (2) in 1D has the form

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \nabla(K\nabla c), \quad x \in (0,1), \qquad t \in (0,T], \\ c(x,0) = c_0(x), \quad x \in (0,1). \end{cases} \tag{26}$$

This is a parabolic problem and hence we should define two boundary conditions, namely, at the point $x = 0$ and $x = 1$. However, using e.g., the sequential splitting, the first sub-problem (advection part) is a first order hyperbolic problem. For such a problem we can use only one boundary condition, i.e., one of them. Hence, the boundary condition at the other point will not be satisfied, which may cause some difficulties.

7. Summary and conclusion

In this paper a general overview of the different operator splitting methods was given. The algorithms of the different methods were considered and compared from different points of view, which might be interesting in solving the real-life problems. Advantages of the operator splitting approach, which are useful for specialists aiming at using numerical methods of high level were formulated. However, this approach has some drawbacks, which are also listed in the work. Nowadays these topics are under intensive investigations. Hopefully this method can be successfully applied in the numerical weather prediction, too. However, this latter topics needs some further investigations.

Acknowledgements—The author thanks to the referees for the valuable comments and remarks. This work was supported by Hungarian National Research Founds (OTKA) Nos. T043765, T049819, and NATO Collaborative Linkage Grant No. 980505.

References

Botchev, M., Faragó, I., and *Havasi, Á.,* 2003: Testing weighted splitting schemes on a one-column transport-chemistry model. *Int. J. Environ. Pollut. 22,* 3-16.

Christov , C.I. and *Marinova, R.S.,* 2001: Implicit vectorial operator splitting for incompressible Navier - Stokes equations in primitive variables. *J. Comput. Techn. 6,* 92-119.

Csomós, P., Faragó, I., and *Havasi, Á.,* 2005: Weighted sequential splittings and their analysis. *Comput. Math. Appl. 50,* 1017-1031.

Dimov, I., Faragó, I., Havasi, Á., and *Zlatev, Z.,* 2006: Different splitting techniques with application to air pollution models. *Int. J. Environ. Pollut.* (to appear).

Faragó, I., 2005: Splitting methods for abstract Cauchy problems. In *Lect. Notes Comp. Sci.* 3401. Springer Verlag, Berlin, 35-45.

Gnandt, B., 2005: A new operator splitting method and its numerical investigation. In *Advances in Air Pollution Modeling for Environmental Security,* 54. Kluwer, Amsterdam, 229-241.

Havasi, Á., 2005: Dispersion analysis of operator splittings in the linearized shallow water equations. In *Lect. Notes Comp. Sci.* 3743. Springer Verlag, Berlin, 355-362.

Havasi, Á., Bartholy, J., and *Faragó, I.,* 2001: Splitting method and its application in air pollution modeling. *Időjárás 105,* 39–58.

Horvath, R., 2005: operator splittings for the numerical solution of the Maxwell's equation, the linearized shallow water equation. In *Lect. Notes Comp. Sci.,* 3743. Springer Verlag, Berlin, 363-371.

Hundsdorfer, W. and *Verwer, J.G.,* 2003: *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* Springer, Berlin.

Jakobsen, K., Karlsen, K.H., and *Risebro, N.H.,* 2001: On the convergence rate of operator splitting for Hamilton - Jacobi equations with source terms. *SIAM J. Numer. Anal. 39,* 499-518.

Karlsen, K.H. and *Risebro, N.H.,* 2002: Unconditionally stable methods for Hamilton-Jacobi equations. *J. Comput. Phys. 180,* 710-735.

Karlsen, K.H., Lie, K.A., Natvig, J.R., Nordhaug, H.F., and *Dahle, H.K.,* 2001: Operator splitting methods for systems of convection—diffusion equations: nonlinear error mechanisms and correction strategies. *J. Comput. Phys. 173,* 636-663.

Lax, P., 2002: *Functional Analysis.* Wiley-Interscience, New York.

Marinova, R.S., Christov, C.I., and *Marinov, T.T.,* 2003: A fully coupled solver for incompressible Navier-Stokes equations using operator splitting. *Int. J. Comput. ~ Fluid D. 17,* 371-385.

Mimura, M., Nakaki, T., and *Tomoeda, T.,* 1984: A numerical approach to interface curves for some nonlinear diffusion equations. *Japan. J. Appl. Math. 1,* 93-139.

Richtmyer, R.D. and *Morton, K.W.,* 1967: *Difference Methods for Initial-Value Problems.* Interscience Publisher, New York, 419 p.

Zlatev, Z., 1995: *Computer Treatment of Large Air Pollution Models.* Kluwer, Amsterdam.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 397–415

Analytical solutions and numerical experiments for optimizing operator splitting procedures

Petra Csomós

Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University
P.O. Box 120, H-1518 Budapest, Hungary; E-mail: csomos@cs.elte.hu

(Manuscript received in final form 11 July, 2006)

Abstract—Operator splitting procedure is a widely used approach for modeling physical processes. Of the numerical solving process both accuracy and fast integration are required. These requirements, however, usually contradict. In the present paper our investigations are presented regarding the optimization of the combined effect of the splitting and the numerical method. Their interaction is examined both analytically and numerically in the total error of the solution, and an idea is presented how to control the accuracy while taking reasonable computing time. Furthermore, an example is shown how to optimize the application of splitting procedure in air pollution transport models.

Key-words: optimization, operator splitting, error analysis, air pollution transport model

1. Introduction

Modeling physical phenomena, systems of complicated partial differential equations have to be solved. In real situations (e.g., in the case of a weather forecast model or an air pollution transport model) the analytical solutions are not known, therefore, certain numerical methods should be applied. When solving the equations numerically, two requirements have to be taken into account: (i) to obtain sufficiently accurate numerical solutions and (ii) to reduce as much as possible the computing (or CPU) time. This means that the error caused by the numerical solving process has to be small, and the numerical solution has to be computed in a short time. The two above requirements, however, usually contradict, because for small error we would need small time steps, which would cause long computing time. In this paper we show our investigations concerning a possible choice of an optimal step size for the numerical integration. Therefore, we analyze the interaction between these two effects when we apply an operator splitting procedure together with a numerical method.

In Section 2 the different splitting procedures are introduced. Section 3 presents the forms of the analytic solutions and the errors. In Section 4 we investigate the behavior of the error appearing in the numerical solution. In order to measure this error its order is computed and investigated. In Section 5 we show our results concerning how to integrate the equations of an air pollution transport model effectively with taking the advantage of the application of a splitting procedure. In Section 6 our results are summarized.

2. Operator splitting procedures

The idea of introducing operator splitting procedures comes from the realization that the spatial differential operators appearing in the models, describing real physical phenomena, usually have complicated structure. Therefore, it is hard to find an appropriate numerical method which could be efficiently used to solve the problem, i.e., which fulfils both the above two requirements: accuracy and efficiency. The basic idea behind the splitting procedure is that the spatial differential operator can be divided into the sum of few sub-operators having simpler structure. For treating the sub-operators numerically, we can find more suitable numerical methods. Then a sequence of equations corresponding to the sub-operators is solved instead of the complicated system. The connections between the sub-systems are the initial conditions. In what follows, we introduce three possible splitting procedures.

Let us consider the following system of ordinary differential equations written in vectorial form, where **M** is a bounded linear operator of type $\mathbf{R}^{\mathrm{N}} \to \mathbf{R}^{\mathrm{N}}$ (i.e., it can be represented as a matrix):

$$\begin{cases} \dfrac{\mathrm{d}u(t)}{\mathrm{d}t} = \mathbf{M}u(t), & t \in (0,T] \\ u(0) = u_0, \end{cases}, \tag{1}$$

where u is the unknown function, and $u_0 \in \mathbf{R}^{\mathrm{N}}$ is a given initial value. We remark that we investigate the case when **M** is bounded, because the spatial discretization of a partial differential equation leads to an initial value problem as Eq. (1).

Let us divide the time interval $[0,T]$ into m pieces of intervals with length τ, where τ is called *splitting time step* ($T = m\tau$). We assume that the operator **M** can be written as the sum of two bounded linear sub-operators **A** and **B** having simpler structure, i.e., $\mathbf{M} = \mathbf{A} + \mathbf{B}$. There exist several splitting procedures for solving Eq. (1) (e.g., *Strang*, 1968; *Marchuk*, 1988; *Csomós et al.*, 2005a).

The simplest one is the so-called *sequential splitting*, defined by the following sequence of sub-problems:

$$\begin{cases} \dfrac{\mathrm{d}u_1^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_1^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_1^{(k)}((k-1)\tau) = u_{\mathrm{spl}}((k-1)\tau) & \end{cases},$$

$$\begin{cases} \dfrac{\mathrm{d}u_2^{(k)}(t)}{\mathrm{d}t} = \mathbf{B}u_2^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_2^{(k)}((k-1)\tau) = u_1^{(k)}(k\tau) & \end{cases}, \tag{2}$$

$$u_{\mathrm{spl}}(k\tau) := u_2^{(k)}(k\tau),$$

with $k = 1,\dots,m$, and $u_1^{(0)} = u_0$, where $u_{\mathrm{spl}}(k\tau)$ is the solution of the split problem, Eq. (2), defined on the mesh $\{k\tau : k = 1,\dots,m\}$. Application of this splitting means that first we solve the system with the sub-operator **A** on the time interval $[0,\tau]$ using the original initial condition. Then we solve the system also on the same time interval but with the sub-operator **B,** and using the solution of the previous step as initial condition. Then we continue this process (always using the previous solution as initial condition) until reaching the last time interval. At time $t = T = m\tau$ we consider $u_{\mathrm{spl}}(m\tau)$ as the numerical solution of the original problem given by Eq. (1).

The second type of the operator splitting procedures is called *weighted splitting*, which can be obtained by using two sequential splittings, once in the order $\mathbf{A} \rightarrow \mathbf{B}$, and afterwards $\mathbf{B} \rightarrow \mathbf{A}$. At time $t = \tau$ the numerical solution is computed as a weighted average of the solutions obtained by the two sequential splitting steps:

$$u_{\mathrm{spl}}(\tau) = \Theta \cdot u_{\mathrm{spl},\mathbf{AB}}(\tau) + (1-\Theta) \cdot u_{\mathrm{spl},\mathbf{BA}}(\tau),$$

where $\Theta \in [0,1]$ is the weight parameter, and $u_{\mathrm{spl},\mathbf{AB}}(\tau)$ and $u_{\mathrm{spl},\mathbf{BA}}(\tau)$ are the solutions of the two sequential splittings, respectively. The case $\Theta = 1/2$ is called *symmetrical weighted splitting*.

Another possibility is the *Strang splitting*, where for one splitting time step the following three problems have to be solved.

$$\begin{cases} \dfrac{\mathrm{d}u_1^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_1^{(k)}(t), & t \in ((k-1)\tau, (k-1)\tau + \tau/2] \\ u_1^{(k)}((k-1)\tau) = u_{\mathrm{spl}}((k-1)\tau) \end{cases},$$

$$\begin{cases} \dfrac{\mathrm{d}u_2^{(k)}(t)}{\mathrm{d}t} = \mathbf{B}u_2^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_2^{(k)}((k-1)\tau) = u_1^{(k)}((k-1)\tau + \tau/2) \end{cases}, \qquad (3)$$

$$\begin{cases} \dfrac{\mathrm{d}u_3^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_3^{(k)}(t), & t \in ((k-1)\tau + \tau/2, k\tau] \\ u_3^{(k)}((k-1)\tau) = u_2^{(k)}(k\tau) \end{cases},$$

$$u_{\mathrm{spl}}(k\tau) := u_3^{(k)}(k\tau),$$

with $k = 1,\ldots,m$, and $u_1^{(0)} = u_0$, and $u_{\mathrm{spl}}(k\tau)$ is considered to be the numerical solution of the original problem, Eq. (1), at time $t = T$.

We note that there is also a weighted version of the Strang splitting, and weighted splittings can be efficiently used on parallel computers (see *Csomós et al.*, 2005b). We also note that the convergence of the above splitting procedures can be proved in the case, when the split sub-problems are solved exactly, i.e., when no numerical methods are used (see *Faragó* and *Havasi*, 2005). However, these results have rather theoretical than practical importance from the numerical point of view.

3. *Optimization of the accuracy*

In this section we present our results concerning the accuracy of the numerical solution of a model. We investigate the conditions under which the numerical solution is accurate enough, moreover, the computational time is as short as possible. In order to demonstrate our results, we present our numerical experiments, as well.

3.1 Analytic solutions and different kinds of errors

Since in real models we usually have a highly non-linear operator, we do not know the explicit form of the analytic solution in practice. However, we can write formally the analytic solutions of Eqs. (1), (2), and (3) at time $t = \tau$ by using the exponential of operators, as follows:

$$\begin{aligned}
u(\tau) &= \exp(\tau(\mathbf{A}+\mathbf{B}))u_0, \\
u_{\text{spl}}^{\text{sq}}(\tau) &= \exp(\tau\mathbf{B})\exp(\tau\mathbf{A})u_0, \\
u_{\text{spl}}^{\Theta}(\tau) &= \Theta\exp(\tau\mathbf{B})\exp(\tau\mathbf{A})u_0 + (1-\Theta)\exp(\tau\mathbf{A})\exp(\tau\mathbf{B})u_0, \\
u_{\text{spl}}^{\text{S}}(\tau) &= \exp(\tau/2\mathbf{A})\exp(\tau\mathbf{B})\exp(\tau/2\mathbf{A})u_0,
\end{aligned} \tag{4}$$

where $u(\tau)$, $u_{\text{spl}}^{\text{sq}}(\tau)$, $u_{\text{spl}}^{\Theta}(\tau)$, and $u_{\text{spl}}^{\text{S}}(\tau)$ denotes the solutions of Eqs. (1)-(2), the equations using the weighted splitting, and Eq. (3), respectively. We note that the exponential of an unbounded operator appearing in the real models cannot be computed easily, but for the bounded **M** the exponential is defined by the following infinite series:

$$\exp(t\mathbf{M}) = \sum_{k=0}^{\infty} \frac{(t\mathbf{M})^k}{k!}.$$

From the above solutions (Eqs. (4)) the so-called *local splitting error* can be defined as:

$$\varepsilon_{\text{spl}}(\tau) = \left| u(\tau) - u_{\text{spl}}(\tau) \right|,$$

where $|.|$ denotes any vector norm in $\mathbf{R}^{\mathrm{N}}$. The *order* of the local splitting error (i.e., the order of the corresponding splitting procedure) is defined by the following quantity p:

$$p := \sup\left\{ q \in N : \lim_{\tau\to 0} \frac{E_{\text{spl}}(\tau)}{\tau^{q+1}} < +\infty \right\}. \tag{5}$$

Using the series expansions of Eqs. (4), one can check (see, e.g., *Hundsdorfer* and *Verwer*, 2003; *Faragó* and *Havasi*, 2005) that the sequential splitting is of first order ($p = 1$), the Strang splitting is of second order ($p = 2$), and the weighted splitting has $p = 1$ if $\Theta \neq 1/2$, and $p = 2$ if $\Theta = 1/2$ (symmetrically weighted splitting). If the condition $[[\mathbf{A},\mathbf{B}],\mathbf{A}-\mathbf{B}]=\mathbf{0}$ holds for the symmetrically weighted splitting (where $[\mathbf{A},\mathbf{B}]$ denotes the commutator of operators **A** and **B**), then it is of third order ($p=3$).

Although in practice we cannot measure the local splitting error exactly, we can estimate its value by its order. If a splitting procedure has an order of p, then its local splitting error behaves like $\varepsilon_{\text{spl}} = const \cdot \tau^{p+1}$ for small values of τ. Moreover, if **A** and **B** are non-stiff operators (i.e., **A** or **B** does not have incommensurable eigenvalues), the order of the splitting error characterizes the global splitting error (the splitting error at time $t=T$), as well: it behaves like $const \cdot \tau^{p}$ for small values of τ. This means that we can measure the accuracy of the numerical solution of Eq. (1) by the order of the local splitting error. Therefore, at time $t=\tau$ the numerical solution obtained by applying a splitting procedure differs from the exact solution of Eq. (1) by a factor of τ^{p+1}, i.e.,

$$\varepsilon_{\text{spl}}(\tau) = \left|u(\tau) - u_{\text{spl}}(\tau)\right| = \mathrm{O}\left(\tau^{p+1}\right). \tag{6}$$

3.2 Role of the numerical method

When the Eq. (1) represents a real physical model, the solutions shown in Eqs. (4) cannot be computed analytically, therefore, numerical methods have to be used to solve the split sub-problems in Eqs. (2) and (3). The application of a splitting procedure with splitting time step τ together with a numerical method with time step $\Delta t \le \tau$ can be better understood from *Fig. 1*.

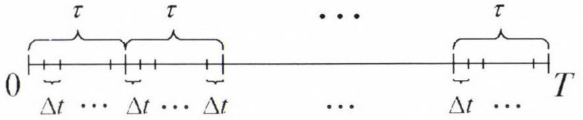


Fig. 1. Time steps of the numerical solving process when applying a splitting procedure with splitting time step τ, and a numerical method with time step Δt (for solving the split sub-problems).

In this case we can estimate the error appearing in the numerical solution only if we know the order of the splitting procedure as well as the order of the numerical method. Then the *local total error* is defined by the difference between the exact and the numerical solution of the model (1), i.e.,

$$\varepsilon_{\text{tot}}(\tau) = \left|u(\tau) - y_{\text{spl}}\right|,$$

where y_{spl} denotes the numerical solution of Eq. (1) at time $t = \tau$, applying a splitting procedure and a numerical method (in order to solve the split sub-problems numerically). In practice, we are interested in the order of $\varepsilon_{\mathrm{tot}}(\tau)$, however, we can only estimate it, for instance, in the following way:

$$\varepsilon_{\mathrm{tot}}(\tau) = \left|u(\tau) - y_{\mathrm{spl}}\right| \le \left|u(\tau) - u_{\mathrm{spl}}(\tau)\right| + \left|u_{\mathrm{spl}}(\tau) - y_{\mathrm{spl}}\right| = \varepsilon_{\mathrm{spl}}(\tau) + \varepsilon_{\mathrm{int}}(\tau), \quad (7)$$

where $\varepsilon_{\mathrm{spl}}(\tau)$ is the above defined local splitting error. We call the term $\varepsilon_{\mathrm{int}}(\tau)$ *interaction error*, because we will see that it is caused by the interaction between the splitting procedure and the numerical method. One can see that the order of $\varepsilon_{\mathrm{tot}}(\tau)$ is equal to the order of $\varepsilon_{\mathrm{spl}}(\tau)$ only if the order of $\varepsilon_{\mathrm{int}}(\tau)$ is higher than the order of $\varepsilon_{\mathrm{spl}}(\tau)$. Therefore, in what follows we will examine the order of the interaction error.

Let us apply an operator splitting procedure of order p together with a numerical method of order r. Let us assume that $\Delta t = \tau^{s} (s \ge 1)$, and $\tau \to 0$. Then one can check (see *Csomós* and *Faragó*, 2005) that $E_{\mathrm{int}}(\tau)$ represents the global numerical error at time $t = \tau$, i.e., after the first splitting step. Therefore, we can estimate $\varepsilon_{\mathrm{int}}(\tau)$ as:

$$\varepsilon_{\mathrm{int}}(\tau) = \mathrm{O}\left(\tau^{rs+1}\right). \quad (8)$$

It can be seen from Eqs. (6), (7), and (8) that the order of the local total error $\varepsilon_{\mathrm{tot}}(\tau)$ equals $\min\{p, rs\}$. In practice, if we apply a splitting procedure of a certain order (i.e., p), we expect that the total error has the same order p. One can see, however, that there appears an order reduction in $\varepsilon_{\mathrm{tot}}(\tau)$ if $rs < p$. In this case the total error has an order less than the order of the splitting. In order to avoid this phenomenon, the numerical method (its order and its time step) has to be chosen carefully.

- If we apply a numerical method (of an order $r \le p$), then Δt has to be chosen as $\Delta t = \mathrm{O}\left(\tau^{p/r}\right)$. We remark that for the case $r > p$ the choice $\Delta t = \tau$ is optimal, otherwise the computing time would be longer.
- If we fix $\Delta t = \tau^{s}$ $(s \ge 1)$, then we have to choose a numerical method of order $[p/s] + 1$, where $[.]$ denotes the integer part. We remark that the use of a higher order numerical method is not optimal, because it would only lead to a longer computing time.

One can see that the order reduction is due to the interaction between the splitting procedure and the numerical method. It is not worth using a numerical method of an order which is less than the order of the splitting, because then the order of the total error is less than the order of the splitting. It is not worth using a very small time step, as well, because it leads to a long computing time. On the other hand, a large time step causes the order reduction. One can see that the optimal choices of the order and the time step of the numerical method are very important for the applications.

3.3 Analytical computations

In this section we present our analytical computations to demonstrate the above result, i.e., that the order of the total error equals the minimum of p and rs, where p and r are the orders of the splitting and the numerical method, respectively, and $\Delta t = \tau^s$ $(s \geq 1)$. From *Fig. 1* one can see that $\tau = n \cdot \Delta t$, where $n \in \mathbf{N}$ (and n is an even number in the case of the Strang splitting).

For instance, let us use an explicit Euler method for solving the split sub-problems. Then the numerical solutions have the following forms at time $t = \tau$:

$$\begin{aligned} y^{\mathrm{sq}}_{\mathrm{spl}} &= (1 + \Delta t\mathbf{A})^n (1 + \Delta t\mathbf{B})^n u_0, \\ y^{\Theta}_{\mathrm{spl}} &= \Theta(1 + \Delta t\mathbf{A})^n (1 + \Delta t\mathbf{B})^n u_0 + (1 - \Theta)(1 + \Delta t\mathbf{B})^n (1 + \Delta t\mathbf{A})^n u_0, \\ y^{\mathrm{S}}_{\mathrm{spl}} &= (1 + \Delta t\mathbf{A})^{n/2} (1 + \Delta t\mathbf{B})^n (1 + \Delta t\mathbf{A})^{n/2} u_0, \end{aligned} \tag{9}$$

where $y^{\mathrm{sq}}_{\mathrm{spl}}$, $y^{\Theta}_{\mathrm{spl}}$, and $y^{\mathrm{S}}_{\mathrm{spl}}$ are the numerical solutions of Eq. (1) using the explicit Euler method and the sequential, the weighted, and the Strang splitting, respectively. From the Taylor expansions of the Eqs. (4) and (9) we get the following results for the exact and the split solutions:

$$\begin{aligned} u(\tau) &= \left[\mathbf{I} + \tau(\mathbf{A} + \mathbf{B}) + \frac{\tau^2}{2}\left(\mathbf{A}^2 + \mathbf{B}^2 + \mathbf{AB} + \mathbf{BA}\right) + \mathrm{O}(\tau^3) \right] u_0, \\ y^{\mathrm{sq}}_{\mathrm{spl}} &= \left[\mathbf{I} + \tau(\mathbf{A} + \mathbf{B}) + \frac{\tau^2}{2}\left(\left(1 - \tfrac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + 2\mathbf{BA}\right) + \mathrm{O}(\tau^3) \right] u_0, \\ y^{\Theta}_{\mathrm{spl}} &= \left[\mathbf{I} + \tau(\mathbf{A} + \mathbf{B}) + \frac{\tau^2}{2}\left[\left(1 - \tfrac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + (1 - \Theta)\mathbf{AB} + \Theta\mathbf{BA}\right] + \mathrm{O}(\tau^3) \right] u_0, \\ y^{\mathrm{S}}_{\mathrm{spl}} &= \left[\mathbf{I} + \tau(\mathbf{A} + \mathbf{B}) + \frac{\tau^2}{2}\left[\left(1 - \tfrac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + \mathbf{AB} + \mathbf{BA}\right] + \mathrm{O}(\tau^3) \right] u_0, \end{aligned} \tag{10}$$

where $u(\tau)$ is the exact solution of Eq. (1). One can easily verify from the Eqs. (4) that the local splitting error has the following form:

$$\varepsilon_{\mathrm{spl}}(\tau) = \gamma \frac{\tau^2}{2} \|(\mathbf{BA} - \mathbf{AB}) u_0\| + \mathrm{O}(\tau^3), \tag{11}$$

where the values of γ are 1, $\Theta - 1/2$, and 0 for the sequential, the weighted, and the Strang splitting, respectively. From Eqs. (10) one can see that the local total error can be written as follows:

$$\varepsilon_{\mathrm{tot}}(\tau) = \varepsilon_{\mathrm{spl}}(\tau) + \varepsilon_{\mathrm{int}}(\tau) + \mathrm{O}(\tau^3), \tag{12}$$

where $\varepsilon_{\mathrm{int}}(\tau)$ denotes the interaction error having the following form:

$$\varepsilon_{\mathrm{int}}(\tau) = \frac{\tau \Delta t}{2} \|(\mathbf{A}^2 - \mathbf{B}^2) u_0\| + \mathrm{O}(\tau^3). \tag{13}$$

It can be seen from the Eqs. (11), (12), and (13) that the order of the local total error equals the order of the splitting only if $\gamma\tau^2 \geq \tau\Delta t$ holds. This means that in the cases of the sequential ($\gamma = 1$) and the weighted ($\Theta \neq 1/2$, $\gamma = \Theta - 1/2$) splittings, there is no further condition on the time step Δt, because $\Delta t = \mathrm{O}(\tau)$ is always true. However, in the cases of the symmetrically weighted ($\Theta = 1/2$, $\gamma = 0$) and the Strang ($\gamma = 0$) splittings the time step has to be chosen $\Delta t = \mathrm{O}(\tau^2)$, otherwise there appears an order reduction in the local total error $\varepsilon_{\mathrm{tot}}(\tau)$. For finite Δt and τ the notation $\mathrm{O}(\tau^2)$ means that the time step has to be chosen as $\Delta t = const \cdot \tau^2$ for a constant $const \approx 1$.

From the above example one can see that for a certain splitting procedure (with a certain splitting time step) and numerical method, the time step of the numerical method should be chosen carefully in order to obtain the expected accuracy. However, it is not worth choosing a very small value for it, because then the integration would take longer time. Therefore, the optimal choice of Δt is $\mathrm{O}(\tau)$ for the first-order splittings, and $\mathrm{O}(\tau^2)$ for the second order splittings. Since in this example we only treated a first-order numerical method (explicit Euler method), we continue our investigations by studying higher order numerical methods.

3.4 Numerical experiments

In this section we present our numerical experiments concerning the order of the local total error when higher order numerical methods are used, as well. The order of the local total error is defined similarly as Eq. (5), but $\varepsilon_{\text{tot}}(\tau)$ is written instead of $\varepsilon_{\text{spl}}(\tau)$. Therefore, order of $\varepsilon_{\text{tot}}(\tau)$ equals the supremum of those numbers q for which

$$\frac{\varepsilon_{\text{tot}}(\tau)}{\tau^{q+1}} \approx c < +\infty \tag{14}$$

for sufficiently small values of τ. From Eq. (14) it follows that the logarithm of the local total error can be written as the following linear function of $\log \tau$:

$$\log \varepsilon_{\text{tot}}(\tau) \approx (q+1)\log \tau + \log c. \tag{15}$$

One can see that the slope of line of Eq. (15) corresponds to the estimation of the order of $\varepsilon_{\text{tot}}(\tau)$. In order to estimate the order of $\varepsilon_{\text{tot}}(\tau)$, we should make several numerical experiments with different values of τ, and fit a straight line to the resulted points in the logarithmic scale. Then the slope of the fitted line gives the order of $\varepsilon_{\text{tot}}(\tau)$.

In order to compute the local total error, we need a test problem with a known analytical solution. Therefore, in what follows, we treat the model of the *harmonic oscillator*, which is described by the matrix **M** in Eq. (1) already decomposed into the sum of two matrices:

$$\mathbf{M} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} \alpha & 0.5 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} -\alpha & 0.5 \\ -2 & 0 \end{pmatrix},$$

where $\alpha \in \mathbf{R}$. With the above sub-matrices all the introduced splitting procedures can be applied: sequential, weighted, and Strang for $\alpha \neq 0$, and the third-order symmetrically weighted splitting (with the commutator condition) for $\alpha = 0$. We used four different numerical methods (of different order) for solving the split sub-problems: explicit Euler method ($r = 1$), second-order midpoint method ($r = 2$), third- and fourth-order Runge–Kutta methods ($r = 3$ and $r = 4$, respectively).

Let us denote the estimation of the order of the local total error (i.e., the slope of the straight line of Eq. (15)) by ρ. Let us choose the time step of the numerical method $\Delta t = \tau^{S}$ ($s = 1,\ldots,6$), and let us choose 20 different values of τ as follows:

$$\tau_0 = \frac{2\pi}{200} \quad \text{and} \quad \tau_i = \frac{10}{11}\tau_{i-1}, \quad \text{for} \quad i = 1,\ldots,19.$$

Having computed the values of the slopes for the different splittings and numerical methods, we obtain figures similar to *Fig. 2*, which is the case of the third-order symmetrically weighted splitting ($p = 3$). One can see that the estimated order ρ of the local total error behaves differently for the different numerical methods. When the first-order numerical method is applied, then $\rho = p$ if $s \geq 3$, for the second-order numerical method $\rho = p$ if $s \geq 2$, and for the third- and fourth-order numerical methods $\rho = p$ if $s \geq 1$.

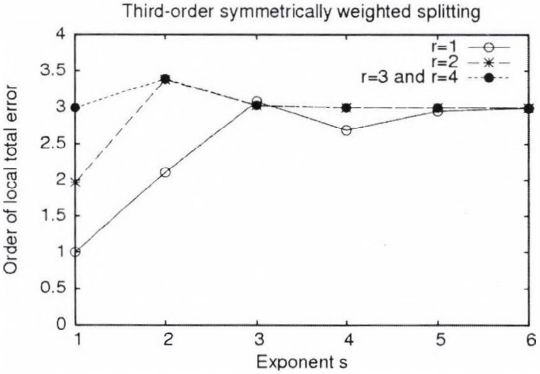


Fig. 2. Behavior of the estimation of the order of the local total error in the case of the third-order symmetrically weighted splitting (p=3), for four different numerical methods (r=1,2,3,4).

This example demonstrates well the analytical result obtained in Section 3.2: $\rho = p$ if $s \geq p/r$, i.e., for fixed value of r the time step has to be chosen as $\Delta t = \mathrm{O}\left(\tau^{p/r}\right)$.

Summarizing the results of this section we can say that the application of an operator splitting procedure together with a numerical method results in an interaction term in the total error. Therefore, we cannot guarantee the expected accuracy of the numerical solution for any choice of the parameters r and s. However, applying a splitting procedure, we can find a numerical method and its time step, which lead to the best accuracy with the shortest possible computing time. When a splitting of order p is applied, and we choose the time step as $\Delta t = \tau^s$ ($s \geq 1$), then we have to use a numerical method of order r, for which $p = rs$ fulfils.

4. Optimization of the computing time

In this section we present our results concerning the computing time of a simple air pollution model. We will show that by taking the advantage of a splitting procedure, the computing time of the model can be shortened, without a significant loss of accuracy.

4.1 Air pollution transport model

Air pollution transport models forecast the spatial distribution of the concentrations of the air pollutants. The time-evolution of the concentration field is caused by different atmospheric processes: advection (due to the wind velocity field), diffusion, deposition (the purification of the atmosphere), emission (the source of the pollutants), and chemical reactions. The combined effect of these processes (without taking into account the chemical reactions) can be modeled by the following partial differential equation (see *Zlatev*, 1995; *Havasi et al.*, 2001):

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \nabla(K\nabla c) + E - \sigma c, & t \in (0,T] \\ c(\mathbf{x},0) = c_0(\mathbf{x}), & \end{cases} \tag{16}$$

where $c(\mathbf{x},t)$ denotes the concentration of only one chemical species at the point $\mathbf{x}=(x,y)\in\Omega\subset\mathbf{R}^2$ and at time t, $c_0(\mathbf{x})$ is the initial concentration field, $\mathbf{u}=(u,v)$ is the wind velocity field, K is the diffusion coefficient, E is the emission function, and σ describes the deposition. Solving numerically Eq. (16), an outflow boundary condition is applied: pollutant material can leave the forecast domain Ω, but it cannot enter again, and there is no source outside the boundary.

One can see that the spatial differential operator on the right-hand side of Eq. (16) is the sum of different kinds of sub-operators. Therefore, it is worth applying an operator splitting procedure for solving it. Taking that advantage of operator splitting, we can choose different numerical methods for solving each sub-problem (see *Csomós*, 2006). Moreover, different time steps can be chosen for the different sub-problems, as well. The only restriction for the time steps is that they should satisfy the stability condition of the corresponding discretization method, and that there should exist such an integer number n_i in the case of each time step Δt_i that $n_i \cdot \Delta t_i = \tau$. Then we can choose τ to equal the maximal time step Δt_i, i.e., $n_i = 1$ for this index i.

It results in a shorter computing time, because we do not need to use the shortest time step, which would be the case if no splitting was applied.

In our numerical experiments we split the operator on the right-hand side into the two following sub-operators:

$$\begin{aligned} \mathbf{A}c &:= K\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}\right)+E-\sigma c, \\ \mathbf{B}c &:= -\left(\frac{\partial(uc)}{\partial x}+\frac{\partial(vc)}{\partial y}\right), \end{aligned} \tag{17}$$

i.e., the sub-operator **A** describes the combined effect of the diffusion, emission, and deposition, while the sub-operator **B** describes the effect of the advection only. Since the analytical solution of Eq. (15) is not know, we solve it numerically on the spatial mesh $\Omega_\Delta=\{(i\cdot\Delta x, j\cdot\Delta y):\ i=0,\ldots,I, j=0,\ldots,J\}$, where Δx and Δy are the grid sizes in directions x and y, respectively. We choose the wind velocity field governed by the Molenkampf–Crowley advection, i.e.

$$\begin{aligned} u(x,y,t) &= -\mu(x-\widetilde{x}), \\ v(x,y,t) &= \mu(y-\widetilde{y}), \end{aligned}$$

where $(\widetilde{x},\widetilde{y})$ represents the centre of the advection field (the rotation centre). The emission function E has the following form:

$$E(x,y,t)=\begin{cases} E_0, & \text{if} \quad (x-x_0)^2+(y-y_0)^2 \le R^2 \\ 0, & \text{anywhere else,} \end{cases}$$

which describes the emission of a circle-shaped industrial area with radius R and centre (x_0,y_0). The parameters have the following values: $\mu=10^{-2}$ 1/h, $E_0=10^{-4}$ kg/m^3/h, $R=2.5$ km, $K=0.3$ km^2/h, and $\sigma=10^{-5}$ 1/h, and the grid sizes are $\Delta x=\Delta y=0.25$ km, the end of the integration $T=72$ hours, the step size $\Delta t=1$ minute and the splitting time $\tau=1$ hour, the centre of the advection is $(\widetilde{x},\widetilde{y})=(25\text{ km},\ 25\text{ km})$, and the centre of the emission is $(x_0,y_0)=(40\text{ km}, 10\text{ km})$. The numerical solution c_{spl} of Eq. (16) can be seen in *Fig. 3*, in the case of applying sequential splitting and at time $t=T$.

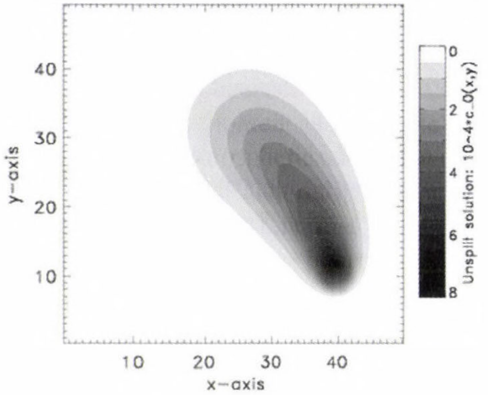


Fig. 3. Numerical solution of the model given by Eq. (16) without applying splitting procedure, at time t = 72 hours.

4.2 Computing times

We solve Eq. (15) in two ways: (i) without applying splitting procedure, by using a simple finite difference discretization method, and (ii) applying sequential splitting with the above sub-operators **A** and **B**, by using semi-Lagrangian method (*Wiin-Nielson*, 1959) for solving the advection sub-model and finite difference method for solving the diffusion - emission - deposition sub-model. The time step of the finite difference method (i.e., for solving the model without splitting, and the diffusion - emission - deposition sub-model) is chosen to satisfy the Courant–Friedrichs–Levy stability condition (see e.g., in *Stoyan* and *Takó*, 1997). Since the semi-Lagrangian method is unconditionally stable, we can use larger time step for solving the advection sub-model. It is convenient to choose this time step to be equal to the splitting time step τ. Thus, we expect that the computing time of the split model will be shorter than the computing time of the model without applying splitting.

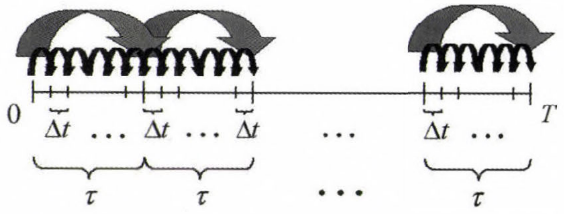


Fig. 4. Number of steps in the case of the semi-Lagrangian method (long arrows) and the finite difference method (short arrows).

In *Fig. 4* the number of steps can be seen which are needed in the case of solving the advection sub-model (long arrows - larger time step) and the diffusion - emission - deposition sub-model (short arrows - smaller time step). In the case when the model is solved without applying splitting procedure, we have to use the smaller time step (i.e., the same as used for the diffusion - emission - deposition sub-model). Therefore, one can see that we can save time if we apply splitting with the decomposition given by Eq. (17) and use larger time step for the advection sub-model.

Our results concerning the computing times of the split and unsplit models are shown in *Fig. 5*. In *Fig. 6* the ratio of the computing times of the split and the unsplit model can be seen. One can see that the integration of the split model (using the semi-Lagrangian method with larger time step for solving the advection sub-model) takes a bit more than half time than that of the integration of the whole model without applying splitting procedure.

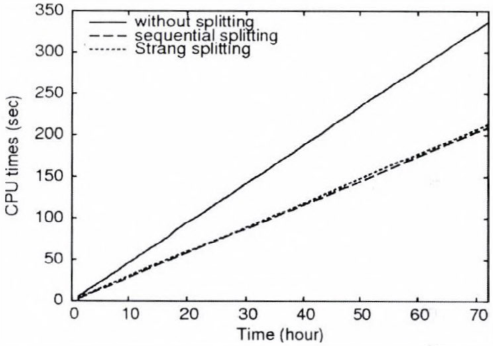


Fig. 5. Computing (CPU) times of the model with and without applying splitting procedure, applying the sequential and the Strang splittings, respectively.

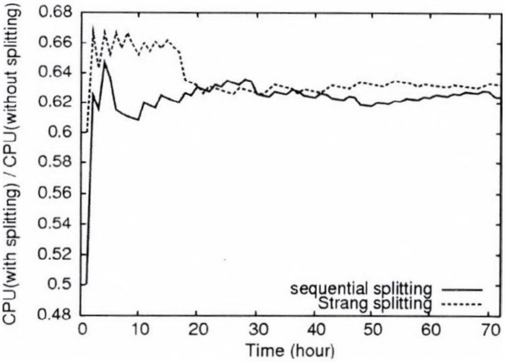


Fig. 6. Ratio of the computing (CPU) times of the split and the unsplit model, applying the sequential and the Strang splittings, respectively.

4.3 Error analysis

In the above subsection we have shown that the integration of the split model (using different numerical methods with different time steps for solving the sub-problems) takes shorter time than the integration of the model without applying splitting. On the other hand, we should also investigate the behavior of the error appearing in the solution of the split problem. This error should be compared to the error of the solution of the unsplit problem, as well.

Since we do not know the analytical solution of Eq. (16), we have to approximate the total error of the numerical solution. For this reason, we compute a reference solution $c_{\mathrm{ref}}(x,y,t)$ of Eq. (16) without applying splitting procedure and by using smaller grid sizes ($\Delta x_{\mathrm{ref}} = \Delta y_{\mathrm{ref}} = 0.125$ km) and time steps ($\Delta t_{\mathrm{ref}} = 15$ second). We note that in the case of a mesh, which is finer than the above introduced, we do not obtain results significantly different, therefore, we can consider the reference solution as a very good approximation to the exact solution of the model. Then we can compute the *approximate total error* $\varepsilon_{\mathrm{atot}}$ at every mesh point and every splitting time step as follows:

$$\varepsilon_{\mathrm{atot}}(i\Delta x, j\Delta y, k\tau) = \left| c_{\mathrm{ref}}(i\Delta x, j\Delta y, k\tau) - c_{\mathrm{spl}}(i\Delta x, j\Delta y, k\tau) \right|, \tag{18}$$

where $i = 0,\dots,I$, $j = 0,\dots,J$, and $k = 0,\dots,m$. Moreover, we can compute the mass of the whole emitted pollution $M(t)$ from the emission function $E(x,y,t)$:

$$M(t) = \iint_{\Omega} \int_{0}^{t} E(x,y,t')\,\mathrm{d}t'\,\mathrm{d}x\mathrm{d}y. \tag{19}$$

Similarly, we can compute the "mass" of the approximate total error, as well:

$$M_{\mathrm{atot}}(t) = \iint_{\Omega} \varepsilon_{\mathrm{atot}}(x,y,t)\,\mathrm{d}x\mathrm{d}y. \tag{20}$$

In practice, the integrals in Eqs. (19) and (20) are evaluated numerically over the mesh Ω_{Δ}. From Eqs. (19) and (20) one can compute a *relative approximate total error* $\varepsilon_{\mathrm{rel}}$ at each splitting time step, which characterizes the temporal behavior of the approximate total error well:

$$\varepsilon_{\mathrm{rel}}(k\tau) = \frac{M_{\mathrm{atot}}(k\tau)}{M(k\tau)}. \tag{21}$$

In *Fig. 7* the approximate total error field $\varepsilon_{\text{atot}}$ can be seen in the case of the sequential splitting, at time $t = T$. In *Fig. 8* the temporal behavior of the relative approximate total error ε_{rel} can be seen in the case of the sequential splitting, at time $t = T$, for three values of τ. For comparison we also plot the relative approximate numerical error of the solution of the unsplit model (which is defined in the same way as Eq. (21), however, in Eq. (18) instead of c_{spl} the solution obtained without splitting should be written).

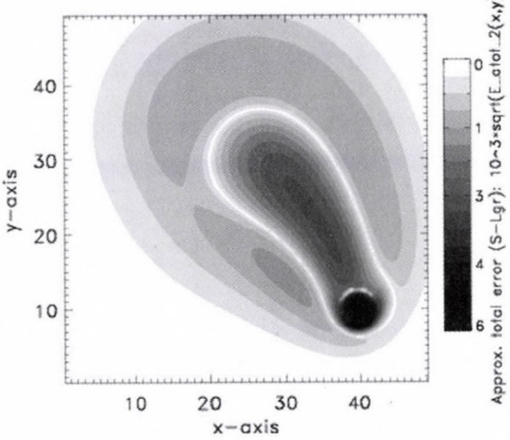


Fig. 7. Approximate total error field at time $t = 72$ hours, when the sequential splitting is applied. (Similar figure in the case of the Strang splitting.)

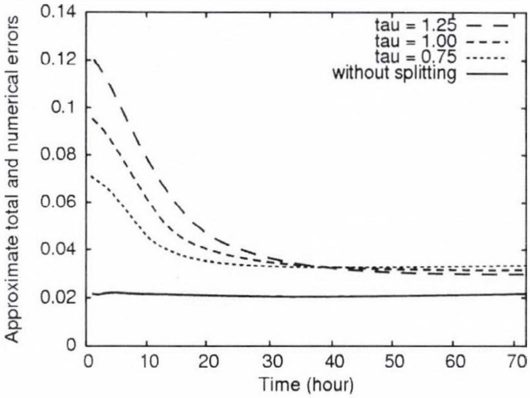


Fig. 8. Temporal behavior of the relative approximate total errors in the case of sequential splitting for three values of τ, and the approximate numerical error of the solution of the unsplit model.

One can see that the application of the splitting procedure causes a certain error, however, its value is decreasing in time, and it tends to the error of the solution obtained without splitting. Until a certain time, the values of $\varepsilon_{\mathrm{rel}}(t)$ are decreasing when τ is decreasing. Thus, the error of the numerical solution remains bounded.

Summarizing the results of this section we can say that taking the advantage of the application of a splitting procedure, i.e., that different numerical methods can be used for each sub-problem, we can shorten the computing time of the model. Moreover, the error remains bounded also in the case when splitting is applied. Therefore, a possible optimization of solving a partial differential equation by applying a splitting procedure is that we use different numerical methods with different time steps for each sub-model.

5. Conclusions

In the present paper we investigated the possibilities of (i) improving the accuracy of the numerical solution, and (ii) shortening the computing time needed for the numerical integration of the model's equation. Since the above two requirements (i) and (ii) usually contradict, we tried to find an optimal way for satisfying both of them.

In the first part of the paper we investigated the accuracy of the numerical solution obtained by applying splitting procedures and numerical methods of different orders together. We showed that there appears a term in the total error of the solution, which is caused by the interaction between the splitting and the numerical method. This term can cause an order reduction of the total error (and, therefore, the loss of accuracy of the solution). However, its effect can be eliminated by choosing conveniently the numerical method (i.e., its order and time step) for the applied splitting procedure. We also found that there exists a lower bound of the time step of the numerical method, under which the accuracy of the solution could not be improved anymore, but the computing time would be longer.

The second part of the paper presented an idea how to shorten the computing time when a splitting procedure is applied. As demonstration, the case of a simple air pollution transport model was studied. Different numerical methods were used with different time steps for solving the split sub-problems, i.e., finite difference scheme with small time step for the diffusion - emission - deposition sub-model, and semi-Lagrangian method with large time step (which was equal to the splitting time step) for the advection sub-model. The larger time step could be chosen because of the unconditional stability of the semi-Lagrangian scheme. We found that the above idea led to a shorter computing time, indeed.

Acknowledgements—This research was supported by the Hungarian Scientific Research Fund under the grant OTKA T43765, and NATO Collaborative Linkage Grant EST CLG 980505.

References

Csomós, P., 2006: Analysis of a transport model applying operator splitting and semi-Lagrangian method. *Int. J. Comput. Sci. Eng.* (in press).

Csomós, P. and *Faragó, I.*, 2005: Error analysis of the numerical solution of split differential equations. *Math. Comput. Model.* (in press).

Csomós, P., Faragó, I., and *Havasi, Á.*, 2005a: Weighted splittings and their analysis. *Comp. Math. Appl. 50*, 1017–1031.

Csomós, P., Dimov, I., Faragó, I., Havasi, Á., and *Ostromsky, Tz.*, 2005b: Computational complexity of weighted splitting schemes on parallel computers (submitted to *Int. J. Parallel, Emergent and Distributed Systems*).

Faragó, I. and *Havasi, Á.*, 2005: On the convergence and local splitting error of different splitting schemes. *Prog. Comput. Fluid Dynam. 5*, 495–504.

Havasi, Á., Bartholy, J., and *Faragó, I*,. 2001: Splitting method and its application in air pollution modeling. *Időjárás 105*, 39–58.

Hundsdorfer, W. and *Verwer, J.G.*, 2003: *Numerical Solution of Time-dependent Advection-Diffusion-Reaction Equations.* Springer-Verlag, Berlin.

Marchuk, G.I., 1988: *Methods of Splitting* (in Russian). Nauka, Moskow.

Stoyan, G., and *Takó, G.*, 1997: *Numerical Methods,* Vol. III. (in Hungarian). ELTE–Typotex, Budapest.

Strang, G., 1968: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal. 5*, 506–517.

Wiin-Nielson, A., 1959: On the application of trajectory methods in numerical forecasting. *Tellus 11*, 180–196.

Zlatev, Z., 1995: *Computer Ttreatment of Large Scale Air Pollution Models.* Kluwer Academic Publishers, Dodrecht.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 417–425

Operator semigroups and applications

Eszter Sikolya

Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University
P.O. Box 120, H-1518 Budapest, Hungary; E-mail: seszter@cs.elte.hu

(Manuscript received in final form June 21, 2006)

Abstract—Operator semigroups are widely used for proving well-posedness of partial differential equations and for investigating qualitative properties of the solutions. Here we give a short overview for the reader to get familiar with these objects and to get an insight into their applications.

Key-words: partial differential equation, abstract Cauchy problem, operator semigroup, air pollution transport model

1. Introduction

The mathematical structure we want to investigate is the *motion* of a *system* in *time*. The time is parameterized by $\mathbf{R}$ or $\mathbf{R}^+$ (it depends whether we want to handle the past or not). The temporal change of the system is described by distinct states from the state space Z. A state of the system, $z(t) \in Z$, belongs to each t time, $t \in \mathbf{R}^+$. We assume that the motion is *deterministic*, that is, for every time instant t_0 and initial state z_0, there exists a unique motion

$$z_{t_0,z_0} : [t_0, +\infty) \to Z,$$

such that

$$z_{t_0,z_0}(t_0) = z_0.$$

Remark—A deterministic motion $z : \mathbf{R} \to Z$ satisfies the following equality:

$$z_{t_0,z_0}(t+s) = z_{t_0+t, z_{t_0,z_0}(t)}(s),$$

for all $t_0, t, s \in \mathbf{R}$ and $z_0 \in Z$.

This is because if we start at time t_0 in z_0 and look where we are after time $t+s$, the result is the same as if start at time t_0+t where we arrived from z_0 after time t (that is, from $z_{t_0,z_0}(t)$) and look at the system after time s.

We further assume that the system is *autonomous*, which means that

$$z_{t_0,z_0}(t_0+h)=z_{t_1,z_0}(t_1+h)$$

holds for any $t_0,t_1,h\in\mathbf{R}$ and $z_0\in Z$. This implies that the orbits of the motion either join or does not intersect each other.

Many physical phenomena can be described by such systems if we choose appropriate state spaces. The elements of the state space have to include all factors important for the observant. They also have to determine unambiguously the further motion of the system.

2. *Operator (semi)groups*

Using the above model, we can define the operators $T(t):Z\to Z$ for $t\in\mathbf{R}\,(\mathbf{R}^+)$ acting as

$$T(t)z:=z_{t_0,z}(t_0+t),$$

where t_0 can be chosen arbitrary since the system is autonomous. Then clearly

$$T(0)z=z$$

holds, because

$$z_{t_0,z}(t_0)=z.$$

In this way we have defined a *one-parameter (semi)group* of operators satisfying

$$T(t+s)=T(t)T(s),\quad t,s\in\mathbf{R}(\mathbf{R}^+),$$
$$T(0)=Id_Z,$$

since the system is deterministic (see the Remark in the Introduction).

Looking for the solutions of the Cauchy functional equation in $\mathbf{C}$:

$$\begin{cases} T(t+s)=T(t)T(s), & t,s\geq 0,\\ T(0)=1, \end{cases} \tag{1}$$

we find that $T(t) = e^{ta}$ is a solution for any $a \in \mathbf{C}$. It is easy to see that $T(t) = e^{ta}$ satisfies the following differential equation:

$$\begin{cases} \dfrac{d}{dt} T(t) = aT(t), & t \geq 0, \\ T(0) = 1. \end{cases} \tag{2}$$

If we suppose the solution T of Eq. (1) to be continuous, we obtain that it is unique (see *Engel* and *Nagel*, 2000).

Theorem—Assume that $T(\cdot): \mathbf{R}^+ \to \mathbf{C}$ is a continuous solution of Eq. (1). Then there exists a unique $a \in \mathbf{C}$ such that $T(t) = e^{ta}$.

We now generalize the above result in an arbitrary Banach (complete normed) space X, e.g., in $X = \mathbf{C}^n$, $X = C[a,b]$, or $X = L^1(\mathbf{R})$. By $L(X)$ we denote the space of bounded linear operators on X. Let us look for solutions $T(\cdot): \mathbf{R}^+ \to L(X)$ of the following problem:

$$\begin{cases} T(t+s) = T(t)\,T(s), & t,s \geq 0, \\ T(0) = Id_X. \end{cases} \tag{3}$$

Definition—Let $T(\cdot): \mathbf{R}^+ \to L(X)$ be a solution of Eq. (3) satisfying

$$\lim_{t \to 0+} T(t)x = x \quad \forall x \in X.$$

Then $(T(t))_{t \geq 0}$ is called a *strongly continuous (one-parameter) semigroup* (or C_0-*semigroup*). If these properties hold for $\mathbf{R}$ instead of $\mathbf{R}^+$, we call $(T(t))_{t \geq 0}$ a *strongly continuous (one-parameter) group* (or C_0-*group*). For details see *Engel* and *Nagel* (2000) and *Pazy* (1983).

3. Generator

If $A \in L(X)$ – e.g., $A \in M_n(\mathbf{C}), X = \mathbf{C}^n$ – then using the exponential series we can define $e^{tA} \in L(X)$. It is easy to see that the operator family $T(t) := e^{tA}, t \geq 0$ forms a C_0-semigroup satisfying Eq. (3). Furthermore, $T(t)$ is a solution of the following differential equation:

$$\begin{cases} \dfrac{d}{dt} T(t) = AT(t), & t \geq 0, \\ T(0) = Id_X. \end{cases} \tag{4}$$

In this case

$$A = \frac{d}{dt} T(t)\big|_{t=0}$$

and A is called the *generator* of the semigroup.

In general, we can define the generator of a strongly continuous semigroup as follows (see *Engel* and *Nagel*, 2000; *Pazy*, 1983).

Definition—Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup. The linear (but not necessarily bounded) operator

$$D(A) := \left\{ x \in X : \exists \lim_{t \to 0+} \frac{T(t)x - x}{t} \in X \right\},$$
$$Ax := \lim_{t \to 0+} \frac{T(t)x - x}{t} = \frac{d}{dt}(t \to T(t)x)\big|_{t=0}$$

is called the *generator* of $(T(t))_{t\geq 0}$.

Since $(A, D(A))$ is defined as the derivative of the orbits of the semigroup in 0, $T(t)$ is in some ways the generalization of the exponential function of A. Of course, in this case e^{tA} can not be defined by the exponential series, because $(A, D(A))$ is not bounded and the series not necessarily converges in norm. But one can prove that $D(A)$ is always *dense* in X and $(A, D(A))$ is *closed*.

4. Abstract Cauchy problems

Up to now it is not clear how operator semigroups can be used for solving problems in the applications. The clue is the *abstract Cauchy problem*. It is well-known that many physical phenomena can be formulated mathematically as a system of partial differential equations, see e.g., the air pollution transport model in the next section. These systems can often be rewritten as an abstract Cauchy problem, that is

$$\begin{cases} x'(t) = Ax(t), & t \geq 0, \\ x(0) = x_0. \end{cases} \tag{5}$$

The operator A on the right-hand side is usually an (unbounded) differential operator on a function (Banach) space X, $x(t) \in X, t \geq 0$. One can prove the following (see e.g., in *Engel* and *Nagel*, 2000).

Theorem—Let $(A, D(A))$ be a closed, densely defined linear operator on X, and let Eq. (5) be the associated abstract Cauchy problem defined as above. Then the following assertions are equivalent:

(a) For every $x_0 \in D(A)$ there exists a unique solution of Eq. (5) depending continuously on the initial data x_0.

(b) $(A, D(A))$ is the generator of a strongly continuous semigroup $(T(t))_{t \geq 0}$ on X.

In this case the solution is $x(t) = T(t)x_0, \quad t \geq 0$.

Hence, to prove well-posedness of a problem written in the form of an abstract Cauchy problem, one has to verify that the operator on the right-hand side is the generator of a $\mathbf{C}_0$-semigroup. In general it is not easy, but in many important cases it is possible.

5. Examples

The next examples can be found in *Engel* and *Nagel* (2000).

5.1 Diffusion semigroup

Let us take a look at the one-dimensional heat conduction equation with Neumann boundary conditions:

$$\begin{aligned}
\frac{\partial}{\partial t} u(t,s) &= \frac{\partial^2}{\partial s^2} u(t,s), \quad t \geq 0,\ s \in (0,1), \\
u(0,s) &= g(s), \quad s \in [0,1], \\
\frac{\partial}{\partial s} u(t,0) &= \frac{\partial}{\partial s} u(t,1) = 0, \quad t \geq 0.
\end{aligned}$$

We can rewrite it as an abstract Cauchy problem:

$$\begin{cases} x'(t) = Ax(t), & t \geq 0, \\ x(0) = g, \end{cases}$$

with

$$\begin{aligned}
Af &:= f'', \\
D(A) &:= \{ f \in C^2[0,1] : f'(0) = f'(1) = 0 \}.
\end{aligned}$$

Here the Banach space is $X = C[0,1]$ and $x(t) = u(t,\cdot)$. Observe that the boundary conditions appear in the *domain* of A and the operator becomes unbounded – but still it is closed and densely defined in X.

Using the eigenvalues $-\pi^2 n^2$ and eigenfunctions $1, \sqrt{2}\cos \pi ns, n \geq 2$ of A, and the theory of linear ordinary differential equations, one can prove the following.

Theorem—The operator $(A, D(A))$ defined above generates a strongly continuous semigroup $(T(t))_{t\geq 0}$ on $X = C[0,1]$ with

$$(T(t)f)(s) = \int_0^1 k_t(s,r)f(r)dr, \quad f \in C[0,1], s \in [0,1],$$

$$k_t(s,r) := 1 + 2\sum_{n=0}^{+\infty} \mathrm{e}^{-\pi^2 n^2 t} \cos \pi ns \cdot \cos \pi nr.$$

This semigroup is called the *one-dimensional diffusion semigroup*.

In $\mathbf{R}^n$ one can prove the following.

Theorem—Consider the *closure* of the Laplace operator:

$$\Delta f(s_1, s_2, \ldots, s_n) = \sum_{j=1}^{n} \frac{\partial^2}{\partial s_j^2} f(s_1, s_2, \ldots, s_n),$$

defined for every f from the Schwartz space of rapidly decreasing, infinitely many times differentiable functions on $\mathbf{R}^n$. It generates a strongly continuous semigroup $(T(t))_{t\geq 0}$ on $X = L^1(\mathbf{R}^n)$ with

$$(T(t)f)(\mathbf{s}) = \frac{1}{\sqrt{4\pi t}} \int_{\mathbf{R}^n} \mathrm{e}^{-\frac{|\mathbf{s}-\mathbf{r}|^2}{4t}} f(\mathbf{r})d\mathbf{r}, \quad t > 0, \mathbf{s} \in \mathbf{R}^n,$$

$$T(0) = Id.$$

This semigroup is called the *n-dimensional diffusion semigroup*.

5.2 Translation semigroup

Let us investigate the closure of the following first order differential operator:

$$Af := \nabla f,$$
$$D(A) := C_c^1(\mathbf{R}^n).$$

Here $C_c^1(\mathbf{R}^n)$ denotes the space of continuously differentiable functions having compact support in $\mathbf{R}^n$. One can easily prove that $(A, D(A))$

generates a strongly continuous semigroup $(T(t))_{t\geq 0}$ on $X = C_0(\mathbf{R}^n)$ (the space of continuous functions vanishing at infinity on $\mathbf{R}^n$) with

$$(T(t)f)(\mathbf{s}) = f(t \cdot \mathbf{1} + \mathbf{s}), \quad \mathbf{s} \in \mathbf{R}^n,$$

called the *translation semigroup* on $\mathbf{R}^n$.

5.3 Multiplication semigroup

Let $q : \mathbf{R}^n \to \mathbf{C}$ be a continuous function. We can define the following closed, densely defined linear operator on $X = C_0(\mathbf{R}^n)$:

$$\mathrm{M}_q f := qf,$$
$$D(\mathrm{M}_q) := \{f \in C_0(\mathbf{R}^n) : qf \in C_0(\mathbf{R}^n)\}.$$

If

$$\sup_{s \in \mathbf{R}^n} \operatorname{Re} q(s) < \infty$$

then

$$T_q(t)f := \mathrm{e}^{tq} f, \quad t \geq 0, f \in C_0(\mathbf{R}^n)$$

defines the strongly continuous *multiplication semigroup*, generated by $(\mathrm{M}_q, D(\mathrm{M}_q))$.

5.4 Air pollution transport model

We now turn to a concrete problem that is treated in details in *Dimov et al.* (2001, 2006) and *Havasi et al.* (2001). Air pollution transport can be modeled by the following partial differential equation:

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \Delta c + E - \sigma c + R(c), & t \in (0,T], \\ c(\mathbf{x},0) = c_0(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^n. \end{cases} \tag{6}$$

Here $c = c(\mathbf{x},t)$ denotes the concentration of the air pollutant, $\mathbf{u} = \mathbf{u}(\mathbf{x},t)$ describes the wind velocity, $E = E(\mathbf{x},t)$ is the emission function, $\sigma = \sigma(\mathbf{x},t)$ is the deposition, and $R(c)$ is the chemistry operator. For the sake of simplicity we assumed the diffusion coefficient to be 1. If we look at the right-hand side of Eq. (6), we find that all the operators acting on c are of type discussed above, hence generate strongly continuous semigroups on appropriate spaces. Using the perturbation theory of semigroups we obtain well-posedness for Eq. (6).

6. *Outlook*

The importance of the operator semigroup theory is revealed especially in proving *qualitative properties* of solutions of partial differential equations (abstract Cauchy problems, respectively). A rich theory for qualitative properties of C_0–semigroups has been developed in the last 50 years, that can be useful also in the applications.

Here we mention only one example. Let us recall the following result that plays an important role in the famous Liapunov stability theory for matrices.

Proposition—Let $A \in M_n(\mathbf{C})$ be an $n \times n$ matrix. Then the following assertions are equivalent:

(a) $\lim\limits_{t \to \infty} || \mathrm{e}^{tA} || = 0$;

(b) All eigenvalues of A have negative real part, i.e., $\mathrm{Re}\,\lambda < 0$ for all $\lambda \in \sigma(A)$.

This result can be generalized for the asymptotic behavior of semigroups having bounded generator (see *Engel* and *Nagel*, 2000; *Pazy*, 1983).

Theorem—Let $A \in L(X)$ be a bounded operator on some Banach space X and $T(t) := \mathrm{e}^{tA}, t \geq 0$ be the strongly continuous semigroup generated by A. Then the following assertions are equivalent:

(a) $\lim\limits_{t \to \infty} || T(t) || = 0$;

(b) $\mathrm{Re}\,\lambda < 0$ for all $\lambda \in \sigma(A)$.

Hence, to prove that the solutions of an abstract Cauchy problem (containing bounded operator on the right-hand side) converge to 0 if $t \to \infty$, it is enough to investigate the spectrum of the operator on the right-hand side.

Another aspect is the numerical solution of (complicated) partial differential equations where the *operator splitting method* is often used. Here we divide the spatial differential operator of the system into simpler operators and solve the corresponding problems one after the other, by connecting them through their initial conditions (see e.g., *Csomós et al.*, 2005 and *Faragó*, 2005). To use this method one has to assume that the sub-problems are well-posed, which in practice is often hard to prove. Here operator semigroup techniques can help a lot.

References

Csomós, P., Faragó, I., and *Havasi, Á.*, 2005: Weighted sequential splittings and their analysis. *Comput. Math. Appl. 50*, 1017–1031.

Dimov, I., Faragó, I., Havasi, Á., and *Zlatev, Z.*, 2001: L-commutativity of the operators in splitting methods for air pollution models. *Annales Univ. Sci. Sec. Math. 44*, 127–148.

Dimov, I., Faragó, I., Havasi, Á., and *Zlatev, Z.*, 2006: Different splitting techniques with application to air pollution models. *Int. J. Environ. Pollut.* (to appear).

Engel, K.-J. and *Nagel, R.*, 2000: *One-parameter Semigroups for Linear Evolution Equations.* Springer, New York.

Faragó, I., 2005: Splitting methods for abstract Cauchy problems. *Lect. Notes Comput. Si. 3401.* Springer Verlag, Berlin, 35–45.

Havasi, Á., Bartholy, J., and *Faragó, I.*, 2001: Splitting method and its application in air pollution modeling. *Időjárás 105*, 39–58.

Pazy, A., 1983: *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer, New York.

IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 3–4, July–December 2006, pp. 427–442

Discontinuous Galerkin methods for partial differential equations in the atmospheric modeling

Ferenc Izsák

Department of Applied Analysis and Computation Mathematics, Eötvös Loránd University
P.O. Box 120, H-1518 Budapest, Hungary; E-mail: izsakf@cs.elte.hu

(Manuscript received in final form July 11, 2006)

Abstract—In this paper the discontinuous Galerkin method is presented, which can be used for the numerical approximation of the solution of partial differential equations arising in the atmospheric modeling. An overview of this recent approach is presented comparing it with other numerical methods, especially with finite element techniques. The implementation of the appropriate numerical procedure is discussed for a diffusion operator of second order in details, which includes also the case of the advection operator.

Key-words: partial differential equations, numerical solution, Galerkin method, elliptic problems

1. Introduction

In the last decades, a whole scale of numerical methods have been developed for approximating the solution of partial differential equations (PDE's) arising in many disciplines of the natural sciences. The numerical approximation is a real challenge in case of many nonlinear equations, where we have not even exact knowledge on the solvability of the problems. For instance, the existence and uniqueness problem for Navier-Stokes equations of general type – which are used in atmospheric modeling, too – is still an open issue.

From the point of view of the implementations, the most straightforward numerical methods are the finite difference methods. Here the unknown function is investigated in certain points of the domain, where the equations have to be solved. The partial derivatives are then approximated with some appropriate finite differences, which results in an approximation of the original PDE. For a survey on these methods we refer to (*Thomas*, 1995)

Galerkin methods provide an other way to the numerical solution of PDE's. A large variety of these methods has been elaborated depending on the nature of the underlying PDE. However, the implementation can be not so easily deduced from the mathematical formulation, the strong theoretical basis and the relatively easy automatization in case of complicated domains made this concept popular.

Depending on the nature of the finite dimensional function space that is used for the approximation, one distinguishes continuous and discontinuous Galerkin methods. For a systematic overview, see *Brenner* and *Scott* (2002), *Zienkiewicz* and *Taylor* (2000), *Arnold et al.* (2002) and *Cockburn et al.* (2000), respectively.

At the same time, in many cases a mixture of the above procedures is applied during the computations. Frequently, the spatial variables are discretized using a finite element space, and the ordinary differential equations obtained in this way are solved with some usual ODE solver or using any other sophisticated time stepping method. For the continuous case we refer to *Thomée* (1997), and some methods based on discontinuous approximation can be found in *Cockburn* and *Shu* (2001).

In this text we highlight the general setting of the Galerkin methods with an emphasis on the discontinuous version and provide an overview on the differences between these two approaches. We will focus then on the discontinuous version and discuss its implementation in more details.

2. *Continuous and discontinuous Galerkin methods*

For the general framework we consider first an abstract form of some partial differential equation. We investigate the following equation:

$$f = Lu\,, \tag{1}$$

where u denotes the unknown function, f is given, and L is the differential operator in the appropriate PDE. For a more rigorous setting we have to consider the Hilbert spaces H_1 and H_2 (*Brenner* and *Scott*, 2002) and then $L : H_1 \to H_2$ is given along with the left hand side $f \in H_2$. Here H_1 and H_2 are usually Hilbert spaces which contain classical functions.

The spirit of the Galerkin methods is that in Eq. (1) we should take a scalar product of both sides with an arbitrary element v of H_2, where the scalar product is denoted by (. , .). Then u will be called a solution of Eq. (1) (in a weak sense) if

$$(f, v) = (Lu, v) \qquad \text{for all} \quad v \in H_2\,. \tag{2}$$

This form of the original equation is usually transformed further, frequently for the right hand side some integral equality (integration by parts, Green formulae) is applied, using the property of the appropriate Hilbert spaces. This results in the following problem: find $u \in H_1$ such that

$$(f,v) = B_L(u,v) \qquad \text{for all } v \in H_2, \tag{3}$$

where $B_L : H_1 \times H_2 \to R$ is a bilinear operator. Based on Eq. (3) we can define an approximation of the solution u. For demonstrating the key idea we assume here that $H_1 = H_2$, and then consider a finite dimensional (vector) space V_h, which contains some functions and is called the finite element space. In light of this we rewrite Eq. (3) as: find $u_h \in V_h$ such that

$$(f_h, v_h) = B_L(u_h, v_h) \qquad \text{for all } v_h \in V_h, \tag{4}$$

where f_h is an appropriate representation (projection) of f within V_h.

V_h is usually obtained in the way that we split the original domain (in which the equation is posed) into the union of subdomains, and on these some special simple functions (such as constant functions, polynomials of a fixed order) are considered. Beyond this general formulation, the theory branches out and the following approaches are used:

- V_h consists of continuous functions – continuous Galerkin or finite element method,
- the members of V_h may be discontinuous – discontinuous Galerkin methods.

We provide two simple examples, which give special cases corresponding to the above main concepts. We use the notation $v|_K$ denotes the restriction of the function v to the (sub)domain K.

Example 1

The equation is given in the unit square and the discretization is performed using a finite dimensional space consisting of continuous functions.

- The computational domain: $\Omega = (0,1) \times (0,1)$.
- The system of subdomains: $K_1, K_2, K_3, K_4, K_5, K_6$, see *Fig. 1.*
- The finite element (vector) space:
  $W_h = \{w_k : \Omega \to \mathbf{R} | w_h \text{ is continuous in } \Omega,\ w_k|_{\mathrm{K}_i} \text{ is linear, } i = 1,2,\ldots,6\}.$

Practically, if the computational subdomain is splitted into triangles, then a function $w_h \in W_h$ is uniquely determined by its values on the vertices a_i (see *Fig. 1*). Accordingly, in *Fig. 2* we depicted $w_h \in W_h$ with

$$w_h(a_1) = 0, w_h(a_2) = 1, w_h(a_3) = 1, w_h(a_4) = 0.5$$
$$w_h(a_5) = 1.5, w_h(a_6) = 0.5, w_h(a_7) = 2.$$

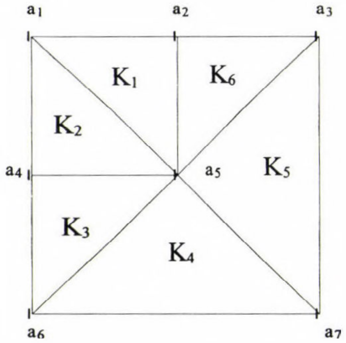


Fig. 1. The computational domain Ω in Example 1, with the subdomains K_i and vertices a_i.

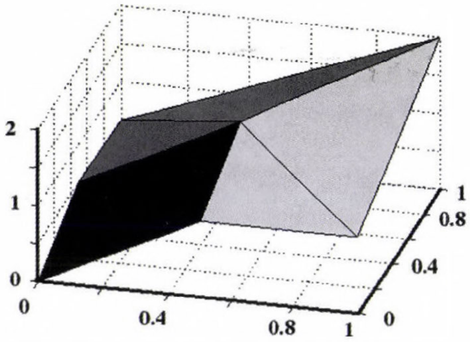


Fig. 2. The function in w_h Example 1.

Example 2

Here we provide a one-dimensional example.

- The computational domain: $\Omega = (-1,1)$.
- The system of subdomains:

$$K_1 = (-1,-0.5),\ K_2 = (-0.5,0),\ K_3 = (0,0.5),\ K_4 = (0.5,1).$$

- The finite element (vector) space:

$$V_h = \{ v_h : \Omega \to \mathbf{R} | \, w_k |_{K_i} \ \text{is at most second order}, \ \ i = 1,2,\ldots,6 \}.$$

Practically, we can define a function $v_h \in V_h$ in the way, that we give them in three points in every subdomain K_i or on the boundary of these. As an example, in *Fig. 3* we depicted the function v_h defined by:

$$v_h|_{K_1}(-1) = 0, \quad v_h|_{K_1}(-0.75) = 0, \quad v_h|_{K_1}(-0.5) = 0.5,$$
$$v_h|_{K_2}(-0.5) = 0.4, \quad v_h|_{K_2}(-0.25) = 0.7, \quad v_h|_{K_2}(0) = 0.4,$$
$$v_h|_{K_3}(0) = 0.5, \quad v_h|_{K_3}(0.25) = 0.25, \quad v_h|_{K_3}(0.5) = 0,$$
$$v_h|_{K_4}(0.5) = 0, \quad v_h|_{K_4}(0.75) = 0.25, \quad v_h|_{K_4}(1) = 0.7.$$

For simplicity we used the short notation $v_h|_{K_1}(-0.5) = 0.5$ instead of the precise formulation $\lim\limits_{\substack{x \in K_1 \\ x \to -0.5}} v_h(x) = 0.5$.

The function v_h is depicted on *Fig. 2*.

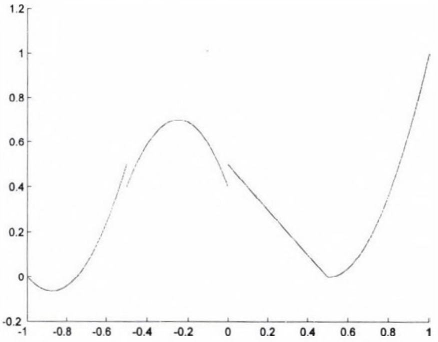


Fig. 3. The function v_h in Example 2.

There is another special kind of Galerkin methods, which covers the case when the original PDE is defined on an unbounded domain. This is called the infinite element method.

If we choose the elements $v_h \in V_h$ in Eq. (4) to be a basis of V_h then we obtain a linear algebraic problem (usually of a big size), which is solved usually again numerically. In details, if a basis in V_h is denoted by $\{\phi_i\}_{i=1}^{n}$ and $u_h = u_{h,1}\phi_1 + u_{h,2}\phi_2 + \ldots + u_{h,n}\phi_n$ then Eq. (4) should hold for all basis elements which gives the equations

$$(f_h, \phi_i) = B_L\left(\sum_{j=1}^{n} u_{h,j}\phi_j, \phi_i \right) \qquad \text{for } i = 1,2,\ldots,n, \tag{5}$$

which is a system of linear equations with the unknowns $\{u_{h,j}\}_{j=1}^{n}$ and the matrix **B**, where the matrix entries $\mathbf{B}[i][j] = B_L(\phi_j, \phi_i)$. The computational complexity is mainly influenced by the size and structure of **B**.

When discontinuous functions u_h are applied to approximate the unknown function u in the original equation Eq. (1), we have to execute carefully the integral equality between Eq. (2) and Eq. (3) including some jump/flux terms. Moreover, if the real solution is supposed to be continuous then we have to interpret the discontinuous solution in a meaningful way. These problems in a concrete case will be discussed later on.

For a short overview on the different approaches we list some arguments for both of them:

- Continuous Galerkin methods,
  - No need to compute fluxes or jumps between the subdomains.
  - Immediate realization of the computational result.
  - Well developed mathematical theory for a number of equations: convergence theorems, error estimation formulae, meaningful refinement strategies are available.
- Discontinuous Galerkin methods:
  - Immediate control and implementation of physical laws (energy, entropy, mass conservation, etc.).
  - Transparent structure of the matrix **B**.
  - Easy adaptation capability (mesh refinement and coarsening), since inter-element continuity is not required.
  - Parallel implementation methods for solving the linear system.

Within this text we try to explain and demonstrate the meaning of these advantages in more details.

3. Construction and implementation of the discontinuous Galerkin method

We investigate the discontinuous Galerkin method for a diffusion problem. When a continuous Galerkin method is applied, this is the usual and most straightforward model problem. But this is not the case now: the problem has to be rewritten into a first order system of equations. Later on, these have to be lumped together again.

The classical continuous Galerkin approach of this problem is available in many textbooks (*Brenner* and *Scott*, 2002; *Girault* and *Raviart*, 1986), such that one can easily compare the two methods.

Let $\Omega \subset \mathbf{R}^3$ denote the computational domain, which is divided into a family of subdomains, and $f : \Omega \to \mathbf{R}$ is a given function. This can be related to the sources in real life situations, but even in case of mass conservation, this approach should be used for the time dependent problems.

Then the classical diffusion problem for the unknown $u : \Omega \to \mathbf{R}$ is

$$\begin{cases} \Delta u = -f \quad \text{in} \quad \Omega, \\ u(x,y,z) = 0 \quad \text{for} \quad (x,y,z) \in \partial\Omega, \end{cases} \tag{6}$$

where $\partial\Omega$ denotes the boundary of Ω.

Remark—If the boundary condition is given as $u(x,y,z) = g(x,y,z)$ on $\partial\Omega$, then we define (e.g., with some extrapolation technique) a function $g_\Omega : \Omega \to \mathbf{R}$ such that $g_\Omega = g$ on $\partial\Omega$. In this case the new unknown function $u_g := u - g_\Omega$ satisfies the homogeneous boundary condition in Eq. (6).

There are several techniques to provide an extrapolation. Practically, the boundary conditions are given only in discrete points, since they are arising from some observations. In this case, g_Ω is an interpolation to the domain Ω.

The Eq. (6) is then rewritten into a system

$$\begin{cases} \nabla u = \Psi, \qquad -\nabla \cdot \Psi = f \quad \text{in} \quad \Omega, \\ u(x,y,z) = 0 \quad \text{for} \quad (x,y,z) \in \partial\Omega, \end{cases} \tag{7}$$

where ∇ denotes the divergence operator, while $\cdot$ will be used for the classical scalar product in $\mathbf{R}^3$. This first order system is not solely a mathematical manipulation: the function Ψ can be interpreted physically as a stream/flux.

For the Galerkin method split Ω into subdomains: $\Omega = K_1 \cup K_2 \cup \ldots \cup K_m$, and K denotes an arbitrary one. The variational formulation according to Eq. (2) and Eq. (3) on K is:

$$\begin{cases} \int\limits_K \nabla u \cdot v = \int\limits_K \nabla \cdot v, \\ -\int\limits_K \nabla \cdot \Psi w = \int\limits_K f w, \end{cases} \tag{8}$$

which is valid for all functions $v \in [L_1(K)]^3$ and $w \in L_1(K)$. Using the divergence theorem for both integrals in Eq. (8), we obtain that

$$\begin{cases} \int\limits_K \Psi \cdot v = -\int\limits_K u \nabla \cdot v + \int\limits_{\partial K} u v_K \cdot v, \\ \qquad\qquad\qquad\qquad\qquad \partial K \\ -\int\limits_K \Psi \cdot \nabla w = \int\limits_K f \Psi + \int\limits_{\partial K} w v_K \cdot \Psi, \end{cases} \tag{9}$$

if the differential operators applying on v and w make sense. Here v_K is the outward normal of K on ∂K. We define the finite dimensional function space within the numerical approximation of the solution of Eq. (6) will be computed.

For some k positive integer P_k denotes the polynomials of at most degree k. If $k = 0$ we obtain the finite volume method, where the basis functions are constant elementwise, and in the case $k = 1$ we obtain linear basis functions. In general, for the approximation of the exact solution u we consider functions, which are in P_k on the individual subdomains and are integrable on the whole Ω. This allows even discontinuities on the element boundaries. Formally, we take the following choices:

$$\begin{aligned} V_h &= \{ v \in [L_1(\Omega)]^3 : v|_K \in [P_k]^3 \}, \\ W_h &= \{ w \in L_1(\Omega) : w|_K \in P_k \}. \end{aligned} \tag{10}$$

In the subsequent analysis, v_K denotes the limit of v on ∂K with respect to K. The same notations are applied to Ψ.

Next, we provide the discrete form of Eq. (9) according to Eq. (4): Find $u_h|_K \in P_k$ and $\Psi_h|_K \in [P_k]^3$ such that for all $w \in P_k(K)$ and $v \in [P_k(K)]^3$ the following equalities hold:

$$\begin{cases} \int\limits_K \Psi_h \cdot v = -\int\limits_K u_h \nabla \cdot v + \int\limits_{\partial K} u_h v_K \cdot v, \\ \qquad\qquad\qquad\qquad\qquad \partial K \\ -\int\limits_K \Psi_h \cdot \nabla w = \int\limits_K f \Psi_h + \int\limits_{\partial K} w v_K \cdot \Psi_h . \end{cases} \tag{11}$$

In this way a function $v \in V_h$ can be expanded as a sum of basis functions on the different subdomains:

$$\begin{aligned} v = {} & a_{1,K_1} v_{1,K_1} + \ldots + a_{n,K_1} v_{n,K_1} + a_{1,K_2} v_{1,K_2} + \ldots + a_{n,K_2} v_{n,K_2} + \ldots + \\ & + \ldots + a_{1,K_m} v_{1,K_m} + \ldots + a_{n,K_m} v_{n,K_m}, \end{aligned} \tag{12}$$

where n is the number of the basis elements which are nonzero only on a fixed subdomain. In other words, $\{v_{i.k}\}_{i=1}^n$ is a local basis on K_k, $k = 1,2....,m$.

When we substitute the expansion Eq. (12) into Eq. (11), the boundary terms on the common face l of K_1 and K_2 have to be computed as follows:

$$\int_l u_{K_1} \nu_{K_1} \cdot v|_{K_1} + \int_l u_{K_2} \nu_{K_2} \cdot v|_{K_2} . \tag{13}$$

Similarly, for the boundary term on an interelement face l in the second equation of Eq. (9) reads as

$$\int_l w_{K_1} \nu_{K_1} \cdot \Psi|_{K_1} + \int_l w_{K_2} \nu_{K_2} \cdot \Psi|_{K_2} . \tag{14}$$

These would be, however, crude approaches, since then a contribution on the interelement faces would be added twice. On the other hand, if the desired function is continuous, then the limit of u from both sides are the same, and we would compute with completely different ones in Eq. (13). Moreover, for the approach with constant functions, the derivative Ψ would be zero everywhere in Eq. (7), which is again not an acceptable approach. In the first case, taking the average seems to be a meaningful approximation, and in the second, one has somehow to take into account the mesh size of K_1 and K_2 and take the approximation proportional to its inverse.

The values of the unknown function u_h on the interelement faces are dealt separately. They are called fluxes in the literature (*Arnold et al.*, 2002) and we denote them with $\phi(u_h)$ and $\phi(\Psi_h)$, respectively, which are used to substitute the values of the unknown functions on the faces of the subdomains.

These can have different values on the same face, or even faces can be recognized as "double faces", which are partially adjusted to the two appropriate subdomains.

For a precise formulation we introduce the following notations on l:

$$\{u\} = \frac{1}{2}(u_{K_1} + u_{K_2}), \ [[u]] = u_{K_1} \nu_{K_1} + u_{K_2} \nu_{K_2}, \tag{15}$$

and similarly, for a vector valued function Ψ:

$$\{\Psi\} = \frac{1}{2}(\Psi_{K_1} + \Psi_{K_2}), \ [[\Psi]] = \nu_{K_1} \cdot \Psi_{K_1} + \nu_{K_2} \Psi_{K_2}, \tag{16}$$

where $\{\cdot\}$ denotes the average, while $[[\cdot]]$ is for the jump on a common face of the subdomains K_1 and K_2.

Next we try to lump the two equations in Eq. (11) together and sum up them for all subdomains. For this the term Ψ_h should be reconstructed from the existing approach for u_h. In this case we need only the approximation for u_h.

A lengthy but straightforward calculation (for the details we refer to *Oden et al.* (1998) and *Arnold et al.* (2002)) leads then to the following variational problem: Find $u_h \in W_h$ such that for all $v \in W_h$ the following inequality holds:

$$\begin{aligned}\int_\Omega fv = \int_\Omega \nabla u_h \cdot \nabla v + \int_F [[\phi(u_h) - u_h]] \cdot \{\nabla v\} - \{\phi(\Psi_h)\}[[v]] \\ + \int_{F_i} \{\phi(u_h) - u_h\}[[\nabla v]] - [[\phi(\Psi_h)]]\{v\}.\end{aligned} \tag{17}$$

Here F denotes the system of the element boundaries, while F_i is for that of the interelement boundaries (the ones inside the domain Ω).

Instead of providing a general formulation using these quantities, we make the following choices:

$$\begin{aligned}\phi(u_h) &= \begin{cases}\{u_h\} + v_K \cdot [[u_h]] \text{ on } F_i \\ \text{and} \\ 0 \text{ on } F \setminus F_i\end{cases} \\ \phi(\Psi_h) &= \{\nabla u_h\} - \eta_F h_F^{-1} [[u_h]] \text{ on } F,\end{aligned} \tag{18)-(19}$$

where η_F is a constant and h_F is the diameter of the actual face F.

This choice has been proposed and elaborated (even for more complicated cases) in *Oden et al.* (1998). For a discussion of such choices or other modification of the scheme we refer to *Arnold* (1982) and *Baumann* and *Oden* (1999a). Substituting Eq. (18) and Eq. (19) into Eq. (17) gives the following variational formulation: Find $u_h \in W_h$ such that

$$\int_\Omega fv = \int_\Omega \nabla u_h \cdot \nabla v + \int_F \eta_F h_F^{-1} [[u_h]] \cdot [[v]] + \int_F [[u_h]] \cdot \{\nabla v\} - \{\nabla u_h\}[[v]] \tag{20}$$

holds for all $v \in W_h$. Based on this formula we can carry out the numerical approximation of the variational problem in Eq. (6) as follows:

First step. First we have to decompose the computational domain Ω into appropriate subdomains.

For the ease of the presentation we take a rectangular mesh. Practically, it means that in a Cartesian coordinate system with the coordinates (x, y, z), the interelement faces are such that either x or the coordinates y or z are constant here. Accordingly, the outward normals ν_K have an easy interpretation: it is either $(\pm 1,0,0)$, or $(0,\pm 1,0)$, or $(0,0,\pm 1)$ according to the above choices.

Second step. We give the basis functions on each subdomains (rectangles) separately. We demonstrate the method in the way that we take first order polynomials in Eq. (10), i.e., $k=1$.

The basis functions on the unit cube $\hat{K}=(0,1)\times(0,1)\times(0,1)$ are then:

$$\hat{v}_0(x,y,z)=1,\ \hat{v}_1(x,y,z)=x,\ \hat{v}_2(x,y,z)=y,\ \hat{v}_3(x,y,z)=z.$$

If the subdomain is the rectangle $K=(p_1,p_2)\times(q_1,q_2)\times(r_1,r_2)$ then the above basis is transformed to

$$\begin{aligned} v_{0,K}(x,y,z)&=\frac{1}{(p_2-p_1)(q_2-q_1)(r_2-r_1)},\ v_{1,K}(x,y,z)=\frac{x-p_1}{p_2-p_1}\\ v_{2,K}(x,y,z)&=\frac{y-q_1}{q_2-q_1},\ v_{3,K}(x,y,z)=\frac{z-r_1}{r_2-r_1}. \end{aligned}\tag{21}$$

Third step. Expanding the unknown numerical approximation in terms of the local basis functions in Eq. (1) as follows:

$$\begin{aligned} u_h(x,y,z)=\sum_{K\subset\Omega} & a_{0,K}\frac{1}{(p_2-p_1)(q_2-q_1)(r_2-r_1)}+a_{1,K}\frac{x-p_1}{p_2-p_1}\\ &+a_{2,K}\frac{y-q_1}{q_2-q_1}+a_{3,K}\frac{z-r_1}{r_2-r_1}. \end{aligned}\tag{22}$$

Fourth step. Computation of the terms in the variational formulation Eq. (20).

- Computation of $[[u_h]]$. We give this quantity on the common face $l=p_2\times(q_1,q_2)\times(r_1,r_2)$ of $K=(p_1,p_2)\times(q_1,q_2)\times(r_1,r_2)$ (with the outward normal $\nu_K=(1,0,0)$) and $\overline{K}=(p_2,p_3)\times(q_1,q_2)\times(r_1,r_2)$. The desired jump on l:

$$\begin{aligned} &[[u_h]](y,z)\\ &=\nu\left[\frac{a_{0,K}}{(p_2-p_1)(q_2-q_1)(r_2-r_1)}+a_{1,K}+a_{2,K}\frac{y-q_1}{q_2-q_1}+a_{3,K}\frac{z-r_1}{r_2-r_1}\right.\\ &\quad\left.-\frac{a_{0,\overline{K}}}{(p_3-p_2)(q_2-q_1)(r_2-r_1)}+a_{1,\overline{K}}+a_{2,\overline{K}}\frac{y-q_1}{q_2-q_1}+a_{3,\overline{K}}\frac{z-r_1}{r_2-r_1}\right]. \end{aligned}$$

The procedure can be executed similarly for all faces of a given subdomain.

- Computation of $\{\nabla u_h\}$. Using the previous setting the desired average vector on l reads as:

$$\{\nabla u_h\}=\frac{1}{2}\left[a_{1,K}\frac{1}{p_2-p_1}+a_{1,\bar{K}}\frac{1}{p_3-p_2},(a_{2,K}+a_{2,\bar{K}})\frac{1}{q_2-q_1},(a_{3,K}+a_{3,\bar{K}})\frac{1}{r_2-r_1}\right].$$

- Computation of $[[v]]$. Taking the basis function in Eq. (21) we have functions which are nonzero only on a fixed subdomain K. In this way, using the above setting we obtain the following quantities for the average of the gradient of the basis functions on l:

$$\{\nabla v_{0,K}\}(y,z)=(0,0,0),\{\nabla v_{1,K}\}(y,z)=\left(\frac{1}{2}\frac{1}{p_2-p_1},0,0\right),$$

$$\{\nabla v_{2,K}\}(y,z)=\left(0,\frac{1}{2}\frac{1}{q_2-q_1},0\right),\{\nabla v_{3,K}\}(y,z)=\left(0,0,\frac{1}{2}\frac{1}{r_2-r_1}\right).$$

- Computation of $\{\nabla v\}$. Since we again take only the basis functions in Eq. (21) which are nonzero only on K, using the previous setting we obtain for the averages:

$$[[v_{0,K}]](y,z)=\left(\frac{1}{(p_2-p_1)(q_2-q_1)(r_2-r_1)},0,0\right),[[v_{1,K}]](y,z)=(1,0,0),$$

$$[[v_{2,K}]](y,z)=\left(\frac{y-q_1}{q_2-q_1},0,0\right),[[v_{3,K}]](y,z)=\left(\frac{z-r_1}{r_2-r_1},0,0\right).$$

Fifth step. We substitute all basis functions into the weak formulation Eq. (17) and in this way, a linear system of equations is obtained, where the left hand side is arising from the left hand side of Eq. (17), while on the right hand side appear the unknown coefficients $a_{K,0}, a_{K,1}, a_{K,2}, a_{K,3}$ for all subdomains K, and we only have to solve this system.

Sixth step. Solution of the linear system. For large systems usually iterative solvers are applied. Using the symmetric property of the matrix in the linear system, even the application of the conjugate gradient method can be appropriate. For a broad overview on all of these methods we refer to *Golub* and *Loan* (1996).

On the other hand, the matrix in the linear system is sparse even if we apply higher order methods. Therefore, paralel computational procedures have been proposed to the iterative solution. For some concrete example we refer to *Biswas et al.* (1994).

Seventh step. Interpretation of the solution. Since the solution of the problem Eq. (1) can be discontinuous, some smoothing techiques are used in order to get a realistic solution. These called also recovery techiques. They produce a smooth solution starting from the numerical approximation in the way that some physical quantities (mass, entrophy, energy) are conserved. A classical techique is presented in *Zienkiewicz* and *Zhu* (1992), while some recent developments for the case of discontinuous Galerkin methods are described in *Ryan et al.* (2005).

4. Applications to PDE's in the atmospheric modeling

Based on the above examples, we can approximate the solution of convection-diffusion equations, which provide the usual model for transport problems. Here, the differential operator can be even easier put into the variational form: in Eq. (7) we do not have to split the equation.

A wide range of discontinuous Galerkin techniques have been developed for the different types of Navier-Stokes equations (*Baumann* and *Oden*, 1999b; *Cockburn et al.*, 2004; *Nair et al.*, 2004), which provide a satisfactory approach for many atmospheric problems.

The discontinuous Galerkin method can be even combined with some conventional techniques. When time dependent problems are solved with some method of lines technique (*Thomas*, 1995), we can apply the discontinuous Galerkin method to discretize the problem in space (instead of the conventional finite difference method). Such a transport scheme is described in concrete terms with numerical experiments in *Nair et al.* (2004). At the same time, some authors propose so called space-time discontinuous methods such that the space and time variables are not considered separately.

For a concrete implementation including the above developments we refer to *Nair et al.* (2004, 2005). In *Nair et al.* (2005) a non-linear first order shallow water equation is investigated, which is widely used in the atmospheric modeling. The equation is given on a cubed sphere using curvilinear coordinates (*Saoudurny*, 1972). For developing an effective numerical solver, the following essential problems were solved out:

- Both the equation and flux terms should be formalized with respect to the curvilinear coordinates. See also the appendix in *Nair et al.* (2005).

- The flux terms should be chosen in the way that at least the mass conservation of the numerical scheme is ensured. The Lax-Fiedrichs scheme is an approriate choice here.
- An appropriate local basis should be chosen on the subdomains in the way, that the matrix **B** of the linear system corresponding to the global problem (see Section 2) has a simple structure. Choosing Legendre polynomials this can be achieved.
- Beyond an appropriate spatial discretization, an effective time integration should be choosen which results in a stable solver. For this, a third oder totel variation diminishing Runge-Kutta method is employed (see also *Gottlieb et al.*, 2001).

For the details on this issues, see *Nair et al.* (2005). Above the exhausting description of the computational procedure, a number of numerical experiments were performed in *Nair et al.* (2005) including many important cases, such as steady-state geostrophic flows, zonal flows over an isolated mountain, and Rossby-Haurwitz waves.

The discontinuous Galerkin formulation of a generalized type shallow water equation is investigated in *Bernsen et al.* (2006). This includes also the case of the barotropic quasi-geostropic equations, which are widely used in the atmospheric modeling for the mid-latitudes. The main improvement of this approach is that a system of equations is investigated with respect to the potential vorticity and the streamfunction. In course of the numerical approximation, the potential vorticity is discretized in a discontinuous Galerkin finite element space, while the streamfunction is in a continuous one. With some appropriate parameters in the time integration, both of the energy and enstrophy are conserved at the discrete level, too. This paper provides also a detailed error analysis both at the theoretical level and based on the numerical experiments executed.

5. Summary

The concept of the discontinuous Galerkin methods has been presented. We performed this for a simple elliptic PDE rewriting it into a system and then into a bilinear form. The discretization was executed using a finite dimensional function space consisting of discontinuous elements. We pointed out the central importance of the flux terms and sketched a concrete computational procedure on a cubic mesh.

For a more realistic implementation corresponding to the numerical solution of a shallow water equations we referred to *Nair et al.* (2005). We pointed out the importance of some further details, such as the formulation with respect to curvilinear coordinates, appropriate choice of local bases, ensuring stabililty, conservation of some physical quantities using a suitable numerical integration method.

The above models, however, should be improved in many aspects both from the point of view of the theory and the implementations.

- Accurate models in the atmospheric modeling use more complicated types of PDE's compared to simple shallow water equations or incompressible Navier-Stokes equations.
- A more general type of meshes (than a squared one) should be constructed, by keeping the formalism with curvilinear cooradinates.
- Effective and fast local a posteriori error estimates have to be developed, which can provide a sound basis of an adaptive mesh refinement in course of the simulations.
- The computations have to be executed in paralel computers in order to accelerate the procedure.

References

Arnold, D.N., 1982: An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal. 19*, 742-760.

Arnold, D.N., Brezzi, F., Cockburn, B., and *Marini, L.D.*, 2002: Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal. 39*, 1449-1479.

Baumann, C.E. and *Oden, J.T.*, 1999a: A discontinuous *hp* finite element method for convection-diffusion equations. *Comput. Meth. Appl. Eng. 175*, 311-341.

Baumann, C.E. and *Oden, J.T.*, 1999b: A discontinuous *hp* finite element method for the Euler and Navier-Stokes equations. *Int. J. Numer. Meth. Fluids 31*, 79-95.

Bernsen, E., Bokhove, O., and *van der Vegt, J.J.W.*, 2006: A (Dis)continuous finite element model for generalized 2D voricity dynamics. *J. Comput. Phys. 211*, 719-747.

Biswas, R., Devine, K.D., and *Flaherty, J.*, 1994: Parallel, Adaptive Finite Element Methods for Conservation Laws. *Appl. Numer. Math. 14*, 225-283.

Brenner, S.C. and *Scott, L.R.*, 2002: *The Mathematical Theory of Finite Element Methods.* 2nd ed. Springer.

Cockburn, B., Karniadakis G., and *Shu, C.W.*, 2000: *Discontinuous Galerkin Methods. Theory, Computation and Applications.* Springer.

Cockburn, B. and *Shu, C.W.*, 2001: Runge-Kutta discontinuous Galerkin method for convection-dominated problems. *J. Sci. Comput. 6*, 173-261.

Cockburn, B., Kanschat, G., and *Schötzau, D.*, 2004: A locally conservative LDG method for the incompressible Navier-Stokes equations. *Math. Comput. 74*, 1067-1095.

Girault, V. and *Raviart, P.-A.*, 1986: *Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms.* Springer.

Golub, G.H. and *Van Loan, C.F.*, 1996: *Matrix Computations.* 3rd ed. Johns Hopkins University Press.

Gottlieb, S., Shu, C.W., and *Tadmor, E.*, 2001: Strong stability-preserving higher-order time discretization methods. *SIAM Rev. 43*, 89-112.

Nair, R.D., Thomas, S.J., and *Loft, R.D.*, 2005a: A discontinuous Galerkin transport scheme on the cubed sphere. *Mon. Weather Rev. 133*, 814-828.

Nair, R.D., Thomas, S.J., and *Loft, R.D.,* 2005b: A discontinuous Galerkin global shallow water model. *Mon. Weather Rev. 133*, 876-888.

Oden, J.T., Babuska, I., and *Baumann, C.,* 1998: A discontinuous *hp* finite element method for diffusion problems. *J. Comput. Phys. 146*, 491-519.

Ryan, J., Shu C.-W., and *Atkins, H.,* 2005: Extension of a postprocessing techniquefor the discontinuous Galerkin methos for hyperbolic equations with application to an aeroacoustic problem. *SIAM J. Sci. Comput. 26*, 821-843.

Saodurny, R., 1972: Conservative finite-difference approximations of the primitive equations on quasi-uniform spherical grids. *Mon. Weather Rev. 100*, 136-144.

Thomas, J.W., 1995: *Numerical Partial Differential Equations - Finite Difference Methods.* Springer.

Thomée, V., 1997: *Galerkin Finite Element Methods for Parabolic Problems.* Springer.

Zienkiewicz, O.C. and *Taylor, R.L.,* 2000: *The Finite Element Method. Vol. 1. The Basis.* Butterworth-Heinemann, Oxford.

Zienkiewicz, O.C. and *Zhu, J.Z.,* 1992: The superconvergent Patch Recovery and a Posteriori Error Estimates. Part 1: The recovery technique. *Int. J. Numer. Meth. Eng. 33*, 1331-1364.

BOOK REVIEW

J.N. Wallace and *P.V. Hobbs,* 2006: **Atmospheric Science – An Introductory Survey** (second edition). Elsevier Academic Press, San Diego, 483 pages, 10 chapters.

Thirty years after the praiseful first edition of the Atmospheric Science, *J.N. Wallace* and *P.V. Hobbs* published their second edition of the comprehensive survey with an idea to cover all the fields of modern atmospheric science in a book light enough to be carried in the student's backpack. In the past thirty years the atmospheric science has largely evolved, thus the second edition contains much more material including new topics such as atmospheric chemistry, atmospheric boundary layer, and climate dynamics. Some of the topics are written with well known scientific book writers as co-authors.

The book consists of 10 chapters, where the first two are introductory. Chapter 1 acquaints the reader with the basic terms and definitions in meteorology, the properties of the Earth's atmosphere, and basics of spherical coordinate system rotating with the Earth. Chapter 2 reviews the ocean circulation, cycling of water, carbon, and oxygen, as well as climate dynamics and history of the climate in the past 100 million years.

Chapters 3 to 6 guide us through the fundamentals of the atmospheric physics and chemistry. The focus of Chapter 3 is on thermodynamics. As the old saying goes – don't change the horse in midstream –, the authors made only negligible changes in the first edition's top chapter. They consider the ideal gas equation and its application to dry and moist air as well as water vapor, the hydrostatic equation, the first and second law of thermodynamics with concept of entropy and impacts for the atmospheric science.

The next chapter written with *Q. Fu* introduces the fundamentals of radiative transfer of the atmosphere. In the introductory part, the main terms of electromagnetic radiation and blackbody radiation laws are described. Advanced part using very limited mathematical deduction presents the physics of scattering, absorption, and emission by gas molecules, radiative transfer in planetary atmosphere emphasing the importance of absorption, and emission of infrared radiation and vertical profiles of radiative heating rate. Chapter 4 also touches on passive remote sensing by satellites.

In the second half of the 20th century, air pollution became an increasing problem, thus Chapter 5 is dedicated to the atmospheric chemistry. The chapter presents the natural and anthropogenic sources, transport, and sinks of major trace gases and aerosols, chemical reactions and cycles in troposphere, as well as stratospheric chemistry focusing on perturbations of stratospheric ozone and formation of ozone hole.

Chapter 6 covers cloud microphysics - aerosol and droplet formation and growth in warm and cold clouds, as well as electrification of thunderstorms. Clouds play an important role in atmospheric chemistry. The chapter concludes by discussing interactions of gases and aerosols with tropospheric clouds extending our knowledge in atmospheric chemistry obtained in Chapter 5.

The subject of Chapter 7 is the atmospheric dynamics considering large scale horizontal motions influenced by Earth's rotation. Fundamental and apparent forces are defined, horizontal equations of motion, primitive equations, and solutions of them are given. The importance of the conservation of vorticity is emphasized. In the reviewer's opinion, the chapter is not sufficiently effective in wave dynamics.

Chapter 8 reviews the dynamics of weather system and associated weather phenomena. Large-scale extratropical weather systems with mesoscale fronts, the effects of terrain on this systems, mesoscale organization of deep cumulus convection, and finally the hurricanes are described. This chapter is written with *L. McMurdine* and *R.A. Houze*.

R. Stull is the author of Chapter 9. A short introduction to atmospheric turbulence and boundary layer meteorology is given here. *R. Stull* guides us through similarity theory, closure techniques, influence of stratification on stability, interplay between the turbulence and vertical profiles of wind, temperature and moisture, as well as the forest and urban effects in the atmospheric boundary layer. In spite of the limitations in mathematics, this chapter is the most impressive.

Chapter 10 is devoted to climate dynamics including climate anomalies, sensitivity and feedbacks, greenhouse warming, as well as climate monitoring and prediction. This final chapter represents a deserving conclusion of the overall introduction to the atmospheric science.

Assuming undergraduate knowledge in physics and mathematics, each chapter is written easy to understand. Short biographic footnotes summarize the lives and works of the scientists who have made major contributions to the history of science and meteorology. The major deficiency of the book, originating from the first edition, is the lack of reference lists and further readings.

The chapters contain quantitative exercises with complete solutions nested in the text. An additional set of quantitative problems and probes of qualitative understanding are given at the end of each chapter. Contrary to the first edition, not only the numerical answers, but nearly a complete set of solutions and mathematical proofs are provided for the quantitative problems with the aim to supply sometimes insufficient mathematics.

The book is rich in full color illustratations. Most of the photographs of clouds, tornados, and satellite images are impressive. Chapter 8 is illustrated with synoptic charts prepared using graphic software GrADS.

The book represents an excellent starting point for the undergraduate students in atmospheric science that every graduate student and probably all the professionals would like to have on their own bookshelf. Following the idea of *J.N. Wallace* and *P.V. Hobbs*, it would be a challenge for future scientific book writers to condense the introductory survey of atmospheric science in one volume including measurement techniques, wave dynamics in atmosphere, and basic numerical methods too.

Á. Bordás

IDŐJÁRÁS

VOLUME 110 * 2006

Editor-in-Chief
LÁSZLÓ BOZÓ

Executive Editor
MARGIT ANTAL

BUDAPEST, HUNGARY

AUTHOR INDEX

TABLE OF CONTENTS
I. Papers

II. Book reviews

III. News

SUBJECT INDEX

GUIDE FOR AUTHORS OF *IDŐJÁRÁS*

The purpose of the journal is to publish papers in any field of meteorology and atmosphere related scientific areas. These may be

- research papers on new results of scientific investigations,
- critical review articles summarizing the current state of art of a certain topic,
- short contributions dealing with a particular question.

Some issues contain "News" and "Book review", therefore, such contributions are also welcome. The papers must be in American English and should be checked by a native speaker if necessary.

Authors are requested to send their manuscripts to

Editor-in Chief of IDŐJÁRÁS
P.O. Box 39, H-1675 Budapest, Hungary

in three identical printed copies including all illustrations. Papers will then be reviewed normally by two independent referees, who remain unidentified for the author(s). The Editor-in-Chief will inform the author(s) whether or not the paper is acceptable for publication, and what modifications, if any, are necessary.

Please, follow the order given below when typing manuscripts.

Title part: should consist of the title, the name(s) of the author(s), their affiliation(s) including full postal and e-mail address(es). In case of more than one author, the corresponding author must be identified.

Abstract: should contain the purpose, the applied data and methods as well as the basic conclusion(s) of the paper.

Key-words: must be included (from 5 to 10) to help to classify the topic.

Text: has to be typed in single spacing with wide margins on one side of an A4 size white paper. Use of S.I. units are expected, and the use of negative exponent is preferred to fractional sign. Mathematical formulae are expected to be as simple as possible and numbered in parentheses at the right margin.

All publications cited in the text should be presented in a *list of references*, arranged in alphabetical order. For an article: name(s) of author(s) in Italics, year, title of article, name of journal, volume, number (the latter two in Italics) and pages. E.g., *Nathan, K.K.*, 1986: A note on the relationship between photo-synthetically active radiation and cloud amount. *Időjárás 90,* 10-13. For a book: name(s) of author(s), year, title of the book (all in Italics except the year), publisher and place of publication. E.g., *Junge, C.E.*, 1963: *Air Chemistry and Radioactivity.* Academic Press, New York and London. Reference in the text should contain the name(s) of the author(s) in Italics and year of publication. E.g., in the case of one author: *Miller* (1989); in the case of two authors: *Gamov* and *Cleveland* (1973); and if there are more than two authors: *Smith et al.* (1990). If the name of the author cannot be fitted into the text: *(Miller,* 1989); etc. When referring papers published in the same year by the same author, letters a, b, c, etc. should follow the year of publication.

Tables should be marked by Arabic numbers and printed in separate sheets with their numbers and legends given below them. Avoid too lengthy or complicated tables, or tables duplicating results given in other form in the manuscript (e.g., graphs)

Figures should also be marked with Arabic numbers and printed in black and white in camera-ready form in separate sheets with their numbers and captions given below them. JPG, GIF, or XLS formats should be used for electronic artwork submission.

The text should be submitted both in manuscript and in electronic form, the latter on disks or in e-mail. Use standard 3.5" MS-DOS formatted diskette or CD for this purpose. MS Word format is preferred.

Reprints: authors receive 30 reprints free of charge. Additional reprints may be ordered at the authors' expense when sending back the proofs to the Editorial Office.

Information on the last issues: http://www.met.hu/Journal-Idojaras.php

Published by the Hungarian Meteorological Service

Budapest, Hungary

INDEX: 26 361

HU ISSN 0324-6329