IDŐJÁRÁS

QUARTERLY JOURNAL
OF THE HUNGARIAN METEOROLOGICAL SERVICE

CONTENTS

****

http://omsz.met.hu/omsz.php?almenu_id=omsz&pid=references&pri=2

VOL. 110 * NO. 1 * JANUARY - MARCH 2006

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service

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*

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 1, January–March 2006, pp. 1–13

Characteristics and synoptic classification of heavy snowfall events in Budapest for the period 1953–2003 Part I

György Babolcsai and **Tamás Hirsch**

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

(Manuscript received in final form January 16, 2006)

Abstract—Winter precipitation is one of the most challenging meteorological phenomena for operational weather forecasters. A possible way of increasing forecast skill is to investigate cases occurred in the past, trying to find essential processes as well as important parameters, and to draw conclusions based on the results. Heavy snowfall events have been investigated for the period 1953–2003 in Budapest. After providing a precise definition of heavy snowfall events, SYNOP reports for the station Pestlőrinc (Budapest) as well as data from the NCEP/NCAR Reanalysis dataset with a temporal resolution of 6 hours have been collected for each event. Based on this comprehensive database, two types of investigation have been carried out. The first part of the paper contains the main characteristics of heavy snowfall events, while in Part II, results connected with the subjective classification of the cases will be presented. The length of the investigated period (50 years) makes it possible to seek potential signs of climate change as well. One of the most important results of the first part of the study is the high stability of the climate system in terms of heavy snowfall events in Budapest. In most of the characteristics, there has been only a slight change during the past 50 years. The most important change revealed by the research, however, is the considerable modification of the frequency of heavy snowfall events within the winter period itself. It can be also stated that our results could be efficiently used in operational weather forecasting.

Key-words: synoptic climatology, heavy snowfall events, snowfall characteristics

1. Introduction

Heavy snowfall events and blizzards are one of those situations still capable of causing huge problems. There are much more devastating weather phenomena as well (e.g., hurricanes, tornadoes, some convective events in summer), but

heavy snowfall events are either much more frequent or cover larger areas, or last longer. In Hungary, these events play a basic role in this respect besides flooding. Therefore, even nowadays it is of vital importance to investigate the conditions leading to heavy snowfall as well as to determine their characteristics, thereby improving their forecasts.

One fifth of Hungary's population lives in a relatively small area, in its capital, Budapest. This well-known fact makes forecasts prepared for this region very important. Even if the weather forecast valid for the country is of basically good quality, it is regarded as unsuccessful by major part of the population if something unexpected occurs in the capital city alone. An unpredicted snowfall in Budapest is one of those situations that will lead to negative response from the public. The reason for this is mainly that in the winter season, people are mostly interested in the question whether it will be snowing and how much.

The synoptic station in Pestlőrinc (Budapest) has been providing data since 1953, so it was possible to investigate heavy snowfall cases based on a 50-year period. The selected station, situated at a height of 140 m above mean sea level, is representative of the snow conditions of the major part of the city.

Heavy snowfall events have been defined as cases of continuous snowfall resulting in a snow cover depth increase exceeding 8 cm. Because of the lower spatial variability of precipitation amount during winter, the formation of a considerable snow cover is also likely in other parts of the city in the selected cases. Important characteristics of heavy snowfall events (HSEs) have been determined and HSEs have been classified as well according to several aspects. This is useful, because the frequency of such events can be estimated in case of given synoptic situations and whether there is any difference in the characteristics of heavy snowfall events depending on their type. In this first part of the paper, the most important characteristics of HSEs will be presented, while the second part is devoted to showing the results connected to the classification of cases.

2. Background

Investigation methods of HSEs show a great variability. Even the definition of cases to be investigated can be quite different. *Spreitzhofer* (1999a) selected cases, when the 24-hour increase in snow cover depth was at least 20 cm in at least 3 of the 81 Austrian stations below 1500 m (above mean sea level). *Younkin*'s studies (*Goree* and *Younkin*, 1966; *Younkin*, 1968) included cases with a snow cover increase of at least 10 cm in an area covering geographical latitudes of at least 6° in the USA. *Wild et al.* (1996) studied blizzards

(moderate or heavy snowfall accompanied by winds of at least 30 knots (~15 m/s) causing snow drift and decreasing visibility below 200 m) using the UK Met Office's definition from 1991. *McNulty* (1991) investigated snowfall events producing a snow cover of at least 5 cm in 12 hours in the area of the Great Lakes in the US, and found 5 synoptic situations with heavy snowfall. *Böhm* (1975) studied cases with at least 5 cm snow cover increase in 24 hours in Vienna.

Besides the selection of cases, the investigation itself can be quite different as well. *Younkin* (*Goree* and *Younkin*, 1966; *Younkin*, 1968) separated 3 weather types based on the circulation conditions at 500 hPa level, focusing mainly on the location and value of the absolute vorticity maxima.

According to *Schalko* (1949), HSEs in the Northern Alps are caused by the orography in case of strong and moist northwesterly flow. The Southern Alps similarly to Eastern Austria can get large amounts of snow in the warm sector of cyclones with centre over Genova. Much more infrequently, cold surges from the east can also result in snowstorms over Austria.

Spreitzhofer (1999a) carried out investigations into the spatial, temporal, and intensity characteristics of HSEs for the winters of 1970/71–1988/89 using daily snow data of 81 stations of the Austrian Hydrological Service. The author divided Austria into four parts, according to the main geographical directions. Spatial extent, eccentricity, and distance of the two furthest stations involved have been determined for each case. He found that in most cases, the zonal axis of the area with heavy snowfall is at least three times longer than its meridional axis, while the meridional axis is longer in only 6% of the cases. In an other study, *Spreitzhofer* (1999b) classified strong Austrian snowstorms from synoptic aspects based on 16 cases observed between 1986 and 1991, using satellite images and ECMWF model results as well.

Heavy snowfall intensity in the British Isles was studied by *Gray et al.* (1981) as the function of the effects resulting in the uprising motion of the air, which is the most important factor in the formation of precipitation. They found that in cyclonic situations, horizontal convergence results in an intensity of approximately 2.5 cm snow/hour, and the snowfall usually lasts for 6–12 hours. The presence of convective instability can produce an intensity of 4 cm/hour, limited to a short period of time and a small area. Frontal lifting leads to short but intense snowfall in case of cold fronts, whereas in case of warm fronts, more persistent but less intense snowfall can be observed due to the longer extent and slower motion of the frontal zone.

Wild et al. (1996) carried out investigations on blizzards developed in the area of the British Isles in the period of 1880–1989. Their results confirmed the essential role of front activity in the formation of heavy snowfall. 49% of the cases were due to warm front, 25% of the cases were caused by an occlusion,

18% by cold front, and only 8% were not connected to frontal activity. Cyclone trajectories were investigated for the period between 5 days prior and 5 days after the event, and a huge variability was found. The decade with the most blizzards was 1960–1969, and January turned out to be the month with the most frequent occurrence of cases.

Mote et al. (1997) studied synoptic characteristics of strong blizzards between 1960 and 1993 in the Southeastern US. Composite charts containing many meteorological parameters were constructed every 12 hours for the period of 72 hours prior to the onset of the event up to +48 hours. As a general conclusion, they stated that besides moist air from the Mexican Golf, blizzards mainly depend on warm air advection, upper level jets, diabatic processes, and cyclogenetic lifting.

Our investigation aims to fill the gap that exists in this important research area for Budapest for the previous 50 years.

3. Characteristics of heavy snowfall events in Budapest

Research has been carried out based on SYNOP reports of the station Pestlőrinc (Budapest), one of the main SYNOP stations of the Hungarian Meteorological Service. 50 winter periods (November, December, January, February, and March) have been investigated between 1953/54 and 2002/03. Heavy snowfall events (HSEs) have been defined as cases of continuous snowfall, producing a snow cover increase exceeding 8 cm at the mentioned station. These events should have produced considerable amount of snowfall in the whole area of Budapest. In case of similar investigations, it is usual to apply temporal or intensity limits. Our intention, however, was to study the weather systems themselves, so time (intensity) limits could not be applied due to the fact that snow depth measurements are only available two times a day (06 and 18 UTC).

Using SYNOP reports of the station Pestlőrinc (Budapest), the following characteristics have been assigned to each HSE:

- month of occurrence, duration, average intensity;
- snow depth increase, ratio of snow depth increase, and its water equivalent;
- average 2m temperature and mean sea level pressure during snowfall.

The other part of the database has been created using the NCEP/NCAR Reanalysis dataset available online at http://www.cdc.noaa.gov/cdc/data.-ncep.reanalysis.html. The NCEP/NCAR Reanalysis project is using a state-of-the-art analysis/forecast system to perform data assimilation from 1948 to the

present in order to create global fields of meteorological parameters with a temporal resolution of 6 hours. Reanalysis fields include surface, pressure level and tropopause level data, as well as various kinds of flux data. Using this dataset, average fields can be created for many meteorological parameters, arbitrary time periods, and any part of the world. Based on average fields created for each HSE (selected meteorological fields have been averaged for the duration of each event), the following characteristics have been determined for Budapest:

- average height of 500 hPa level;
- average total precipitable water;
- average temperature of 925 and 850 hPa levels.

Regarding winter periods from 1953/54 to 2002/03, 71 snowfall events resulting in a snow depth increase exceeding 8 cm have been found. The *number of cases* during the first 25 winter periods was 35, whereas during the next 25 years, 36 heavy snowfall events occurred, which means that there was only a slight difference. Monthly frequencies of HSEs are shown in *Table 1*.

Table 1. Number of heavy snowfall events

Period	November	December	January	February	March	Σ
1953/54–1977/78	4	8	12	8	3	35
1978/79–2002/03	9	10	6	8	3	36
1953/54–2002/03	13	18	18	16	6	71

For the investigated 50 years, December and January have turned out to be the months with the most cases (18 HSEs in each month) followed by February (16), November (13), and March (6). Comparing the two 25-year periods, it can be clearly seen that there was a significant increase in the number of cases in November, while in January significantly less HSE occurred in the second half of the investigated period. A slight increase in the frequency of HSEs has been experienced for December, and interestingly there was no change in case of February and March. Due to the mentioned changes, January, the month with the largest number of HSEs in the first 25 years, has become the month with the less cases excluding March.

According to *Fig. 1a,* the frequency distribution of *snow depth increase* caused by HSEs shows an approximately exponential decrease. During the

investigated 50 years, the largest increase of snow depth due to a HSE was 38 cm (November 29–December 1, 1980 and February 9–11, 1999). Median of all events is 13 cm, while the average is 15 cm. As shown in *Table 2*, there is an increase in the average value by more than 2 cm from the first to the second half of the investigated period. The month with the largest increase in snow depth on average is January. Interestingly, the average increase for November is greater than for December, possibly due to the larger precipitation amounts in the former month. Looking at the mean absolute deviation values, it can be stated that February was the month with the largest and March with the smallest variability during the 50 years.

Another important characteristic of HSEs is their *duration*. The shortest event lasted only for 3 hours resulting in a snow depth increase of 10 cm, while the longest HSE brought 26 cm of snow in 53 hours. The hours at the beginning and end of the heavy snowfall event with precipitation amount less than 0.1 mm have not been included. The median of the duration of heavy snowfall events is 17 hours, and the average is 20 hours. According to *Fig. 1b*, the category 12–18 hours was the most frequent, and the distribution is definitely positively skewed. The average duration of HSEs has slightly increased during the investigated period (*Table 2*). The high average value for March might seem interesting at first sight. This is obviously due to the fact, that stronger solar radiation and higher temperatures which are characteristic of this month enable the snow cover increase to exceed 8 cm, mostly only in case of longer snowfall events with overcast sky, when during the day the solar radiation strongly decreased prohibiting the melting of snow. January has turned out to be the month with the lowest and February with the highest variability of snowfall duration.

Snowfall intensity is also worth studying (*Fig. 1c*). The lowest average intensity during the 50 years was 0.2 cm/hour caused by a snowfall event lasting for 2 days, resulting in a snow depth increase of 10 cm. The case with the highest intensity, namely 3.3 cm/hour, also brought 10 cm of snow but within only 3 hours. The median of average snowfall intensity has been determined as 0.8 cm/hour, while the average value has turned out to be 0.9 cm/hour. According to *Fig. 1c*, average snowfall intensity has a bimodal distribution with both the category of 0.4–0.6 cm/hour and 0.8–1.0 cm/hour, having approximately the same relative frequency of about 25%. There is hardly any change in the average intensity values of the two halves of the investigated period (*Table 2*). Snowfall intensity is near 1 cm/hour on average in every month except for March, which is characterized by an average of about 0.5 cm/hour, most probably due to the considerably higher degree of melting during snowfall events in this month compared to the others.

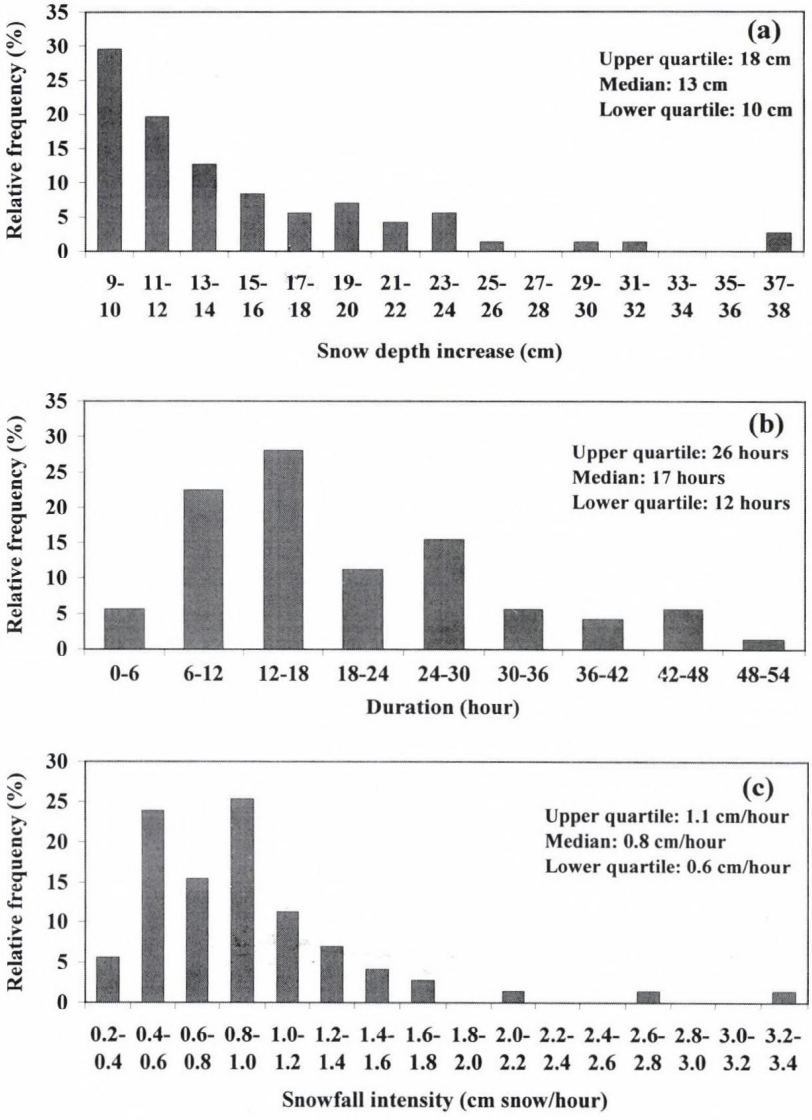


Fig. 1. Frequency distribution of snow depth increase (a), duration (b), and average snowfall intensity (c) for heavy snowfall events.

In *Fig. 2a*, *the ratio of snow depth increase and its water equivalent* is shown, which falls into the category of 0.9–1.2 cm/mm in nearly 50% of the cases. The distribution is positively skewed with a median of 1.1 cm/mm and an average of 1.3 cm/mm. According to *Table 2*, both average and mean

absolute deviation have decreased during the investigated period. Monthly averages show, that a precipitation amount of 1 mm approximately corresponds to 1 cm of snow in November, December, and March, while the same amount of precipitation is equivalent to nearly 1.5 cm of snow in January and February.

Table 2. Characteristics of heavy snowfall events: average (upper corner) and mean absolute deviation (lower corner)

	1953/54–2002/03	**1953/54–1977/78**	**1978/79–2002/03**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**
Snow depth increase (cm)	15.0 5.0	13.8 4.3	16.1 5.6	15.1 4.6	13.9 3.9	16.8 4.6	15.8 6.9	10.0 0.7
Duration (h)	20 9	19 8	21 10	18 8	19 9	22 7	19 11	25 9
Snowfall intensity (cm/h)	0.92 0.36	0.94 0.41	0.9 0.31	1.05 0.46	0.87 0.32	0.96 0.36	0.98 0.32	0.48 0.17
Ratio of snow depth incr. and its water equiv. (cm/mm)	1.26 0.33	1.31 0.37	1.2 0.29	1.15 0.26	1.1 0.26	1.46 0.34	1.41 0.36	0.9 0.21
Total precipitable water (mm)	11.6 1.7	11.8 1.7	11.4 1.6	12.3 2.0	11.6 1.7	10.9 1.1	11.3 1.8	13.3 1.1
Mean sea level pressure (hPa)	1009 7.2	1008 6.8	1009 7.7	1014 7.4	1007 7.0	1008 8.3	1006 5.8	1010 5.7
Height of 500 hPa level (gpm)	5380 67	5379 63	5382 72	5425 94	5369 46	5364 71	5358 62	5427 34
2m temperature (°C)	−2.5 2.0	−2.5 1.8	−2.6 2.2	−1.4 1.1	−2.5 1.9	−3.6 2.2	−2.9 2.4	−0.9 1.0
Temperature at the 925 hPa level (°C)	−3.5 2.6	−3.2 2.2	−3.8 3.0	−2.6 1.8	−3.6 2.7	−4.6 3.4	−3.9 2.5	−1.2 1.8
Temperature at the 850 hPa level (°C)	−4.9 2.3	−4.9 2.2	−5.0 2.5	−4.2 1.6	−4.9 2.5	−5.9 2.6	−5.1 2.1	−3.2 1.5

Heavy snowfall events are caused by cyclones, so it is worth having a look at the *mean sea level pressure* (MSLP) as well. The lowest average MSLP during HSEs was 987 hPa, in case of a deep secondary low on December 28, 1999. The highest average MSLP, namely 1028 hPa, was connected to a case on January 19, 1995, when a shallow cyclone developed over Italy between a huge cyclone over the Atlantic and an anticyclone in Eastern Europe. Most frequently, the category of 1005–1010 hPa occurred (*Fig. 2b*) where median (1008 hPa) and average (1009 hPa) can also be found, pointing to the rather high degree of symmetry of the distribution. There is hardly any difference in the average value characteristic of the two halves of the investigated period (*Table 2*). November has turned out to be the month

with the highest average MSLP during heavy snowfall events, while January is the month with the largest variability.

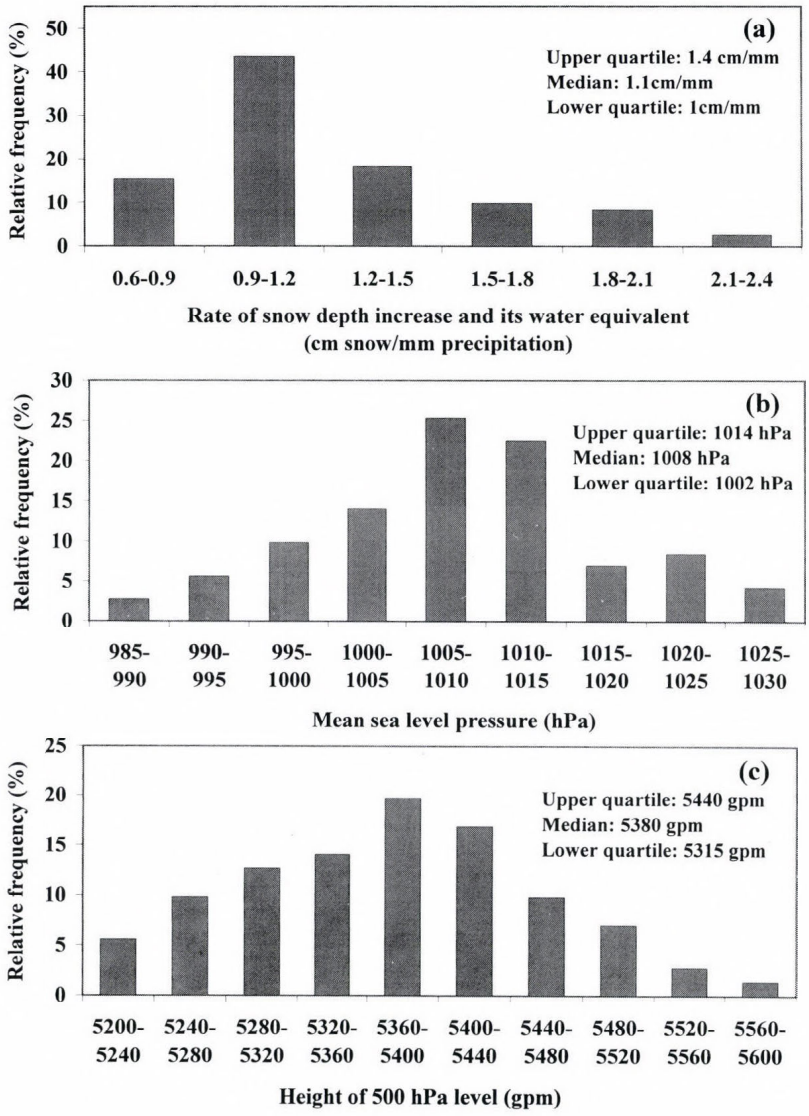


Fig. 2. Frequency distribution of the rate of snow depth increase and its water equivalent (a), mean sea level pressure (b), and the average height of 500 hPa level (c) for heavy snowfall events.

While MSLP characterizes weather systems near the surface, the *height of 500 hPa level* is used to get upper level information. The lowest average value during HSEs for this parameter was 5220 gpm, the highest value reached 5590 gpm. As shown in *Fig. 2c*, the frequency of average height continuously increases up to the category of 5360–5400 gpm, followed by a steady decrease towards higher values. Both average and median are 5380 gpm. Looking at monthly means (*Table 2*), it can be seen that there is a gradual decrease from November till February followed by a sudden increase in March, when the average value is very close to that calculated for November. Variability is the highest in November, and there is only very little change in the average during the investigated 50 years.

One of the key parameters when investigating heavy snowfall events is *total precipitable water* (PW), showing the amount of precipitation that would result from the condensation of all the water vapor contained by the air column above a given point. Obviously, higher PW values favor the development of heavy precipitation. On the other hand, precipitation in form of snow requires low temperatures, which strongly limits possible maximum values of precipitable water, even if the whole layer above surface is close to being saturated. PW values averaged for the duration of HSEs are shown in *Fig. 3a*. The distribution is very symmetric with both the median and average being 12 mm. The lowest mean precipitable water, namely only 6 mm, was connected to the formation of a strong convergence zone along the Danube on 13 December, 2001, resulting in 13 cm of snow for the next morning. The highest average PW value during heavy snowfall was 17 mm, associated with a case on November 18–20, 1965, when very moist air from the south streamed over the Carpathian Basin in the warm sector of a cyclone with centre over the British Isles. According to *Table 2*, the average PW gradually decreases until January, followed by an increase in the consecutive months.

2m temperature averaged for the period of HSEs has also been investigated (*Fig. 3b*). 50% of the cases fell between 0 and –2 °C, and average temperature above 0 °C occurred only 4 times during the investigated 50 years. The lowest average of 2m temperature was –10 °C. The frequency distribution is negatively skewed with a median of –1.8 °C and an average of –2.5 °C. In the distribution, there is a secondary maximum at –6 °C. According to *Table 2*, the average values for the two 25-year periods are almost the same, and there is a slight increase in the mean absolute deviation in the second half of the investigated period. According to *Böhm*'s research (1975), the mean temperature in Vienna during heavy snowfall is mostly between 0 and –0.9 °C.

As last characteristic, *average temperature of the 850 and 925 hPa levels* are shown in *Fig. 3c*. In 25% of the cases, the average of the 850 hPa level temperature is –3 °C, which is the most frequent value for 925 hPa as well.

In case of 850 hPa level temperature, positive average value did not occur during the 50 years, the median was –4 °C, whereas the average was –5 °C. Its frequency distribution shows a definite bimodality with a second maximum at –6 °C, at the same value as 2m temperature.

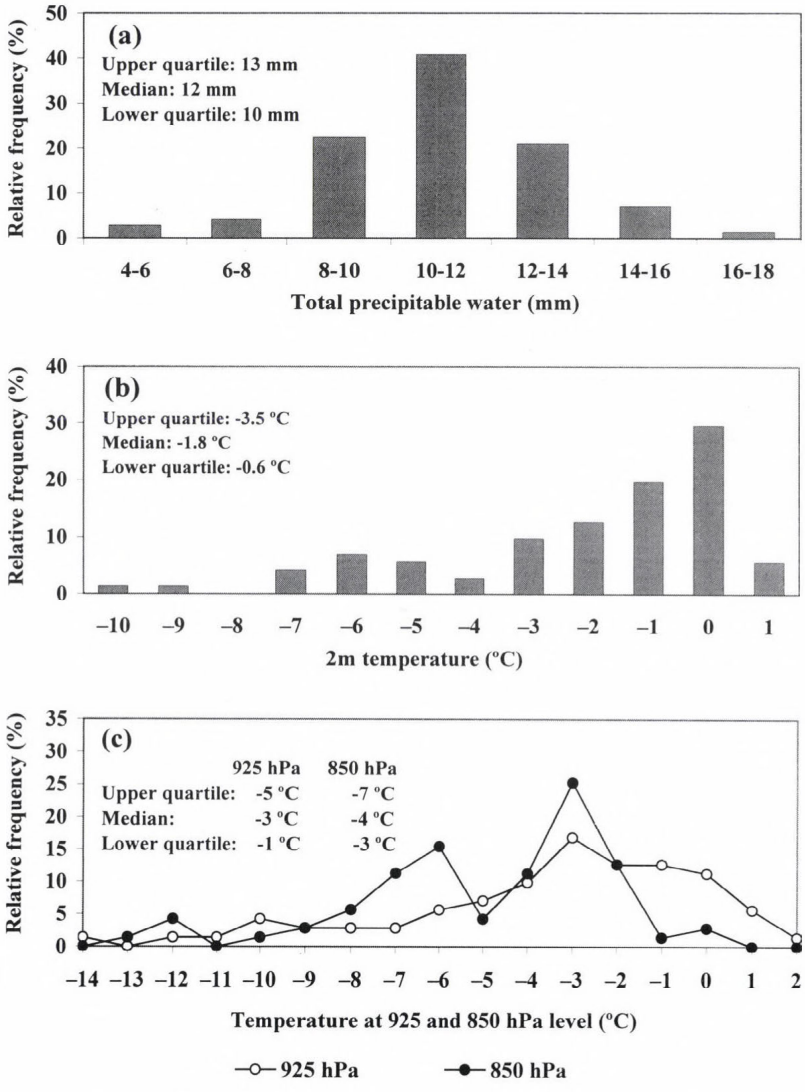


Fig. 3. Frequency distribution of the average total precipitable water (a), 2m temperature (b), and temperature at the 925 and 850 hPa level (c) for heavy snowfall events.

As far as 925 hPa level temperature is concerned, the highest average value reached +2 °C, accompanied by 0 °C at 850 hPa and +0.5 °C at 2 m, in a situation, when Hungary was located in the forward side of a cyclone under the influence of a strong warm advection. The coldest case had an average temperature of –14 °C at the 925 hPa level (–10 °C at 2 m). Median has been determined as –3 °C, while the average of all cases was –3.5 °C. The second half of the investigated period has turned out to be 0.5 °C colder than the first 25 years (*Table 2*). In case of all investigated temperature characteristics, January has the lowest average value, followed by February, December, November, and March.

4. Conclusion

In this study, characteristics of heavy snowfall events have been determined using a comprehensive database of SYNOP reports and analysis fields for a period of 50 years for Budapest. Cases of continuous snowfall producing a snow cover increase exceeding 8 cm have been selected, and several characteristics, including average intensity, snow depth increase, total precipitable water, etc., have been calculated for each case valid for the whole duration of the event. Investigation of the distribution and other statistical features of these characteristics has led to many valuable results.

It has been shown that on average, a precipitation amount of 1 mm corresponds to 1 cm of snow in November, December, and March, and to 1.5 cm in January and February. According to our research, the possible duration of heavy snowfall events with continuous precipitation in Budapest covers a wide time interval ranging from 3 hours up to more than 2 days. Also, snowfall intensity can vary considerably between 0.2 and 3.3 cm/hour, and the snow depth increase can reach almost 40 cm. In winter months like January, weather systems with lower mean sea level pressure over Central Europe can also lead to heavy snowfall events, while in November or March, average MSLP during HSEs is much higher, showing that the vicinity of a strong anticyclone is mostly necessary for heavy snowfall in these months.

The distribution of the investigated characteristics showed large variability from nearly exponential to symmetric, and it is negatively or positively skewed. Using the created empirical density functions, it is possible to estimate the probability of different values of the characteristics based on the long period of 50 years, which is essential when investigating rare meteorological events with a relative frequency of less than 1%, like in our case. The bimodality of the distribution found in case of some of the

characteristics (most significant for the temperature at the 850 hPa level) is also interesting.

The length of the investigated period makes it possible to look for potential signs of climate change by comparing the first and second half of the period. Interestingly, with most of the characteristics, there is only a slight change, pointing to the high stability of the climate system in terms of heavy snowfall events in Budapest. Comparing the two 25-year periods, the number of cases has increased only by 1 (from 35 to 36). In the frequency distribution of heavy snowfall events within the winter period itself, however, we have experienced very significant changes with the average number of HSEs, dropping to 50% in January, while increasing by 125% in November. As a result of this very significant modification, the frequency distribution of HSEs within the winter period has been completely rearranged, which might be a possible sign of climate change as well.

In Part II, results connected with the synoptic classification of heavy snowfall events will be presented. The dependency of HSEs' characteristics on their type will be investigated as well, allowing an even more efficient use of our results in operational weather forecasting.

Acknowledgements—The authors are grateful for the possibility of using NCEP Reanalysis data provided by the NOAA-CIRES ESRL/PSD Climate Diagnostics branch, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov.

References

Böhm, R., 1975: Lufttemperaturen bei Schneefall von größerer Ergiebigkeit in Wien. *Wetter und Leben 27*, 133-137.

Goree, P.A. and *Younkin, R.J.*, 1966: Synoptic climatology of heavy snowfall over the central and eastern United States. *Mon. Weather Rev. 94*, 663-668.

Gray, D.M. and *Male, D.H.*, 1981: *Handbook of Snow: Principles, Process, Management and Use*. Pergamon Press, Oxford.

McNulty, R.P., 1991: *Heavy Snow*. DOC/NOAA/NWS Training Center, Kansas City, MO.

Mote, T.L., Gamble, D.W., Underwood, S.J., and *Benthley, M.L.*, 1997: Synoptic-scale features common to heavy snowstorms in the southeast United States. *Weather Forecast. 12*, 5-22.

Schalko, M., 1949: *Grossschneefälle in Österreich*. Ph.D. Dissertation, Univ. Wien.

Spreitzhofer, G., 1999a: Spatial, temporal and intensity characteristics of heavy snowfall events over Austria. *Theor. Appl. Climatol. 62*, 209-219.

Spreitzhofer, G., 1999b: Synoptic classification of severe snowstorms over Austria. *Meteorol. Z. N.F. 8*, 3-15.

Wild, R., O'Hare, G., and *Wilby, R.*, 1996: A historical record of blizzards/major snow events in the British Isles, 1880-1989. *Weather 51*, 81-90.

Younkin, R.J., 1968: Circulation patterns associated with heavy snowfall over the United States. *Mon. Weather Rev. 96*, 851-853.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 1, January–March 2006, pp. 15–33

Regulatory modeling in Hungary — the AERMOD model. Part II. Sensitivity of the model and case studies

Roland Steib

Hungarian Meteorological Service
P.O. Box 39, H-1675 Budapest, Hungary; E-mail: steib.r@met.hu

(Manuscript received in final form February 24, 2006)

Abstract—This paper presents the sensitivity of a second-generation local-scale dispersion model, called AERMOD, which was adapted at the Hungarian Meteorological Service (HMS) in 2003. AERMOD was designed to introduce current planetary boundary layer concepts into regulatory dispersion models. In this paper the calculation of the most important planetary boundary layer parameters for dispersion calculations are described. The character of the boundary layer is represented with some case studies. We especially wanted to examine, what kind of meteorological parameters can result in high, short time average (1-hour) concentrations in the modeling domain. Finally, we also made a sensitivity analysis study of the newly implemented dry and wet deposition algorithms.

Key-words: regulatory modeling, local-scale dispersion model, planetary boundary layer, sensitivity analysis, concentration values, dry deposition, wet deposition

1. Introduction

In Part I of this study (*Steib* and *Labancz*, 2005) we presented a description of the AERMOD model and its application at the Hungarian Meteorological Service. A comparison between AERMOD and the formal dispersion model, called TRANSMISSION 1.0 (*Szepesi et al.*, 1995) was made. We found that the concentration distribution over flat terrain can differ significantly from the concentration distribution over elevated terrain when using the AERMOD model.

In Part II we explore the sensitivity of the model to various input and boundary layer parameters. We examine the relationship between high, peak (1-hour) concentration values at a receptor point in the modeling domain (with flat and elevated terrain) and the meteorological, surface, and boundary layer parameters.

During the summer of 2005, we have been changing our meteorological database used by the AERMOD model, in order to make calculations with the newly implemented deposition algorithms. Some sensitivity analysis studies were made to the input parameters of the deposition algorithms with the newest model version 04300, and the results are presented in this paper. The dry and wet deposition algorithms of AERMOD were originally released with version 03273, and they were modified with version 04079 and 04300.

2. *Basic algorithms in the AERMOD modeling system*

2.1 Planetary boundary layer algorithms in AERMET

As mentioned in Part I, the basic purpose of AERMET is to use meteorological measurements, representative of the modeling domain, to compute certain boundary layer parameters used in estimating profiles of wind, turbulence, and temperature.

AERMET defines the stability of the PBL by the sign of the surface sensible heat flux, H, (convective for $H>0$ and stable for $H<0$). At the transition point from convective to stable conditions, the heat flux changes sign. This transition occurs when the solar elevation angle reaches a critical value ($\varphi = \varphi_{crit}$). This critical value can be given with the following equation:

$$\sin \varphi_{crit} = \frac{1}{990}\left[\frac{-c_1 T^6 + \sigma_{SB} T^4 - c_2 n}{(1 - r\{\varphi\})(1 - 0.75 n^{3.4})} + 30\right], \tag{1}$$

where c_1 and c_2 are constants, σ_{SB} is the Stefan-Boltzman constant, T is the dry bulb temperature, n is the sky cover (in tenths), and $r\{\varphi\}$ is the albedo, which depends on the solar elevation. It can be seen, that the critical solar elevation depends on the temperature, albedo, and sky cover. For clear and partly cloudy conditions, the transition from stable to convective conditions occurs when φ reaches approximately 13°, for overcast conditions φ_{crit} increases to about 23° (*Holstag* and *van Ulden*, 1983). We can draw the following inference from this. When the geographical latitude of the modeling place is higher than 44.5° (in the Northern Hemisphere) or lower than -44.5° (in the Southern Hemisphere), then the planetary boundary layer can be stable through the whole day in case of an overcast winter day.

In the convective boundary layer (CBL), AERMET computes the friction velocity, u_*, and the Monin-Obukhov length, L, with an iterative method, similar to that used in CTDMPLUS (*Perry*, 1992), since the friction velocity and the Monin-Obukhov length depend on each other. In the stable boundary layer (SBL), the friction velocity and the Monin-Obukhov length are calculated with a simple semi-empirical approach.

The mixing height (z_i) in the CBL depends on both mechanical and convective processes and is assumed to be the larger of the mechanical mixing height (z_{im}) and convective mixing height (z_{ic}). In the SBL, the mixing height results exclusively from mechanical turbulence, and therefore, it is identically equal to z_{im}. The same expression for calculating z_{im} is used in both the CBL and SBL. The mixing height, z_i, for the convective and stable boundary layers is, therefore, defined as follows

$$z_i = MAX[z_{ic}; z_{im}] \qquad \text{for } L<0 \text{ (CBL)},$$

$$z_i = z_{im} \qquad \text{for } L>0 \text{ (SBL)}. \tag{2}$$

The procedure to estimate z_{ic} is the following. If measurements of the convective boundary layer are not available, the convective mixing height (z_{ic}) is calculated with a simple one-dimensional energy balance model (*Carlson*, 1973) as modified by *Weil* and *Brower* (1983). This model uses the early morning potential temperature sounding and the time varying surface heat flux to calculate the time evolution of the convective boundary layer as

$$z_{ic}\theta\{z_{ic}\} - \int_0^{z_{ic}} \theta\{z\}dz = (1+2A)\int_0^t \frac{H\{t'\}}{\rho c_p} dt' , \tag{3}$$

where θ is the potential temperature, A is set equal to 0.2 from *Deardorff* (1980), ρ is the density of the air, c_p is the specific heat at constant pressure, and t is the hour after sunrise.

Fig. 1 represents the relationship between the sensible heat flux and convective mixing height during a 48-hour (May 03, 2005, 00 LST–May 04, 2005, 23 LST) time period. We chose this 2-day time period, because the weather was changing in this period. On the first day the weather was warm and sunny, but on the next day it was colder, cloudy, and also windy. If the sign of H is negative, then z_{ic} is zero. If H is positive (during convective conditions), z_{ic} increases continuously. At the transition point to stable conditions, z_{ic} suddenly falls back to zero.

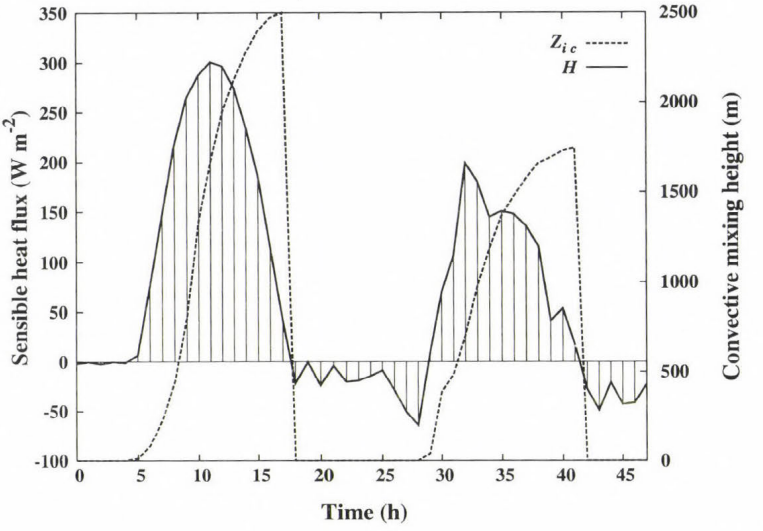


Fig. 1. Relationship between the sensible heat flux and convective mixing height during a 48-hour time period (May 03, 2005, 00 LST–May 04, 2005, 23 LST). Solid line represents the sensible heat flux (W m^{-2}), the dashed line represents the convective mixing height (m).

The procedure to estimate z_{im} is the following. *Venkatram* (1980) has showed that, in mid latitudes, the unsmoothed mechanical mixed layer height, z_{ie}, can be empirically represented as

$$z_{ie} = 2300 u_*^{3/2}. \tag{4}$$

The change of the friction velocity with time can cause sudden and unrealistic drops in the depth of the turbulent layer. Because of this effect, the current hour's smoothed mechanical mixed layer height is computed with the following equation

$$z_{im}\{t+\Delta t\} = z_{im}\{t\} e^{(-\Delta t/\tau)} + z_{ie}\{t+\Delta t\}\left[1 - e^{(-\Delta t/\tau)}\right], \tag{5}$$

where $z_{im}\{t\}$ is the previous hour's smoothed value, $\tau = z_{im} / \beta_\tau u_*$, which is the time scale, and $\beta_\tau = 2$. For computing the time scale, z_{im} is taken from the previous hour's estimate and u_* from the current hour. Eq. (5) shows that the current hour's smoothed mechanical mixing height is a combination of the

current hour's unsmoothed mechanical mixing height and the previous hour's smoothed mechanical mixing height. With this procedure, discontinuities in z_i are avoided. *Fig. 2* shows the relationship between the wind speed and the smoothed mechanical mixing height in the same 48-hour time period. It can be seen, that the character of the two curves is very similar. At the 19th hour, the smoothing effect is clearly visible. The wind speed falls suddenly, but the decrement of the smoothed mechanical mixing height is much more moderate. Eq. (2) results in that during stable conditions z_i is equal to z_{im}, during convective conditions z_i is the maximum of z_{im} or z_{ic}.

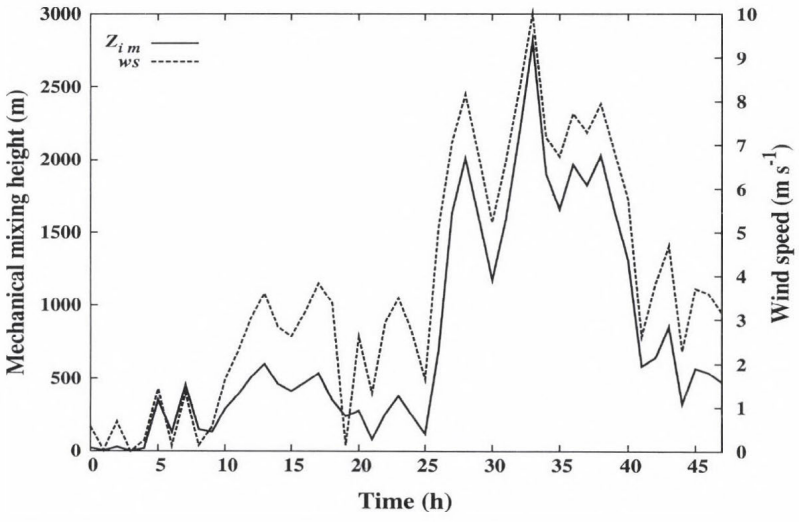


Fig. 2. Relationship between the wind speed and smoothed mechanical mixing height during a 48-hour time period (May 03, 2005, 00 LST–May 04, 2005, 23 LST). Solid line represents the smoothed mechanical mixing height (m), the dashed line represents the wind speed (m s^{-1}).

The behavior of this equation is made visible in *Fig. 3*, which shows the change of the PBL height in the same 48-hour time period. It can be clearly seen, that during the convective conditions of the first day, when the convective effects are stronger than the mechanical effects (sunny condition and low wind speed), z_i is equal to z_{ic} during the whole convective regime, but during the convective conditions of the second day, when the mechanical effects are stronger than the convective effects (cloudy condition and high wind speed), z_i is equal to z_{im} during the convective regime.

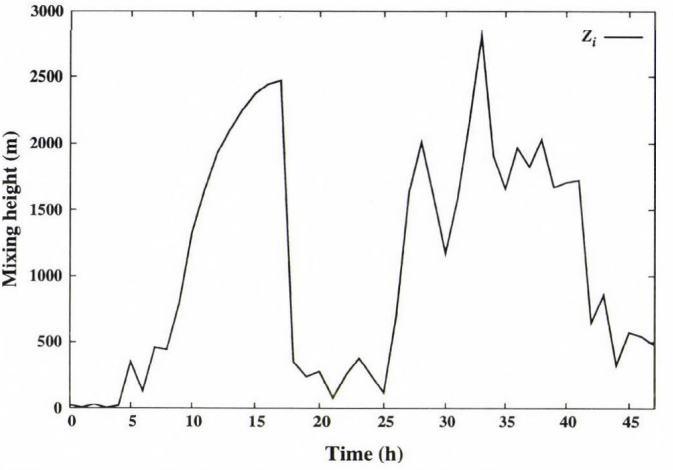


Fig. 3. The height of the planetary boundary layer during a 48-hour time period (May 03, 2005, 00 LST – May 04, 2005, 23 LST). Units: meters.

2.2 Dry and wet deposition algorithms in AERMOD

The dry and wet deposition algorithms in the AERMOD model were originally developed by the Argonne National Laboratory (ANL) for use in the ISC3 model. The model can handle five different deposition algorithms to calculate the deposition fluxes of a pollutant:

- dry deposition of particles (Method 1),
- dry deposition of particles (Method 2),
- gaseous dry deposition,
- wet deposition of particles,
- gaseous wet deposition.

2.2.1 Dry deposition algorithms

The dry deposition flux is calculated as the product of the concentration, χ_d, and a deposition velocity, V_d, computed at a reference height, z_r:

$$F_d = \chi_d \cdot V_d, \tag{6}$$

where F_d is the dry deposition flux (μg m^{-2} s^{-1}), χ_d is the concentration (μg m^{-3}), V_d is the deposition velocity (m s^{-1}), z_r is the deposition reference height (m) = z_0+1, and z_o is the surface roughness length for the application site (m).

Particle dry deposition

The dry deposition velocities of particles are simulated with a resistance scheme, in which the deposition velocity is determined based on the predominant particle size distribution. The model can use two methods to calculate the deposition velocities of particles.

Method 1

Method 1 is used when a significant fraction (greater than about 10 percent) of the total particulate mass has a diameter of 10 μm or larger. The particle size distribution must be known reasonably well in order to use Method 1. Eq. (6) is applied for each particle size category specified by the user, and the results are summed by the model. The particle deposition velocity for Method 1 is given as

$$V_{dp} = \frac{1}{R_a + R_p + R_a R_p V_g} + V_g, \tag{7}$$

where V_{dp} is the deposition velocity for particles (m s^{-1}), R_a is the aerodynamic resistance (s m^{-1}), R_p is the quasilaminar sublayer resistance (s m^{-1}), and V_g is the gravitational settling velocity for particles (m s^{-1}).

The aerodynamic resistance, R_a, is calculated as follows:

- for stable and neutral conditions (L>0),

$$R_a = \frac{1}{(ku_*)}\left[\ln\left(\frac{z_r}{z_0}\right) + \frac{5z_r}{L}\right], \tag{8a}$$

- for unstable conditions ($L<0$),

$$R_a = \frac{1}{(ku_*)}\left[\ln\frac{\left(\sqrt{1-16\frac{z_r}{L}}-1\right)\left(\sqrt{1-16\frac{z_0}{L}}+1\right)}{\left(\sqrt{1-16\frac{z_r}{L}}+1\right)\left(\sqrt{1-16\frac{z_0}{L}}-1\right)}\right], \tag{8b}$$

where k is the von Karman constant (0.4). For Method 1, the quasilaminar sublayer resistance, R_p, is calculated as follows

$$R_p = \frac{1}{(S_c^{-2/3} + 10^{-3/St})(1 + 0.24 w_*^2 / u_*^2) u_*}, \quad (9)$$

where Sc is the Schmidt number (dimensionless), St is the Stokes number (dimensionless), and w_* is the convective velocity scale (m s^{-1}). The gravitational settling velocity is calculated as follows

$$V_g = \frac{(\rho - \rho_{AIR}) g d_p^2 c_2}{18\mu} S_{CF}, \quad (10)$$

where V_g is gravitation settling velocity (m s^{-1}), ρ is the particle density input by user (g cm^{-3}), ρ_{AIR} is the air density ($\approx 1.2 \times 10^{-3}$ g cm^{-3}), g is the acceleration due to gravity (9.80616 m s^{-2}), d_p is the particle diameter (μm), S_{CF} is the slip correction factor (dimensionless), μ is the absolute viscosity of air ($\approx 1.81 \times 10^{-4}$ g cm^{-1} s^{-1}), and c_2 is the air units conversion constant (1.0×10^{-8} cm^2 μm^{-2}).

Method 2

Method 2 is used when the particle size distribution is not well known, and only a small fraction (less than 10 percent of the mass) of the particles has a diameter of 10 μm or larger. The deposition velocity for Method 2 is given as the weighted average of the deposition velocity for particles in the fine mode (i.e., less than 2.5 μm in diameter) and the deposition velocity for the coarse mode (i.e., greater than 2.5 μm but less than 10 μm in diameter):

$$V_{dp} = f_p V_{dpf} + (1 - f_p) V_{dpc}, \quad (11)$$

where V_{dp} is the overall particle deposition velocity (m s^{-1}), f_p is the fraction of particulate substance in fine mode (smaller than 2.5 μm in diameter). V_{dpf} is the deposition velocity (m s^{-1}) of fine particulate substance, calculated from Eq. (7) with V_g set to zero as

$$V_{dpf} = \frac{1}{R_a + R_p}, \quad (12)$$

V_{dpc} is the deposition velocity (m s^{-1}) of coarse particulate substance, calculated from Eq. (7) with V_g set to 0.002 m s^{-1} as

$$V_{dpc} = \frac{1}{R_a + R_p + 0.002 R_a R_p} + 0.002 . \quad (13)$$

For Method 2, the aerodynamic resistance is calculated using Eq. (8), and the quasilaminar sublayer resistance, R_p, is calculated with parameterizations based on observations of sulfate dry deposition:

- for stable and neutral conditions ($L > 0$)

$$R_p = \frac{500}{u_*}, \quad (14a)$$

- for unstable conditions ($L < 0$)

$$R_p = \frac{500}{u_* \left(1 - \frac{300}{L}\right)}. \quad (14b)$$

Gaseous dry deposition

For dry deposition of gases, the deposition velocity is given as

$$V_{dg} = \frac{1}{R_a + R_b + R_c}, \quad (15)$$

where V_{dg} is the deposition velocity for gases (m s^{-1}), R_b is the quasilaminar resistance for bulk surface (s m^{-1}), and R_c is the bulk surface resistance (s m^{-1}). The aerodynamic resistance, R_a, for gases is calculated the same way as for particles, using Eq. (8), except that a lower limit of 1000 s m^{-1} is applied to R_a. The quasilaminar sublayer resistance for bulk surface, R_b, is calculated as follows

$$R_b = \frac{2.2}{(k u_*)} \left(\frac{\upsilon}{D_a}\right)^{2/3}, \quad (16)$$

where D_a is the diffusivity of modeled gas in the air (m^2 s^{-1}) and υ is the kinematic viscosity of air. The bulk surface resistance is calculated as follows

$$R_c = \frac{1}{\left[\frac{LAI_r}{R_s + R_m} + \frac{LAI_r}{R_{cut}} + \frac{1}{R_{ac} + R_g}\right]}, \quad (17)$$

where LAI_r is the relative leaf area index (dimensionless), R_s is the canopy stomatal resistance (s m^{-1}), R_m is the canopy mesophyll resistance (s m^{-1}), R_{cut} is the canopy cuticular resistance (s m^{-1}), R_{ac} is the aerodynamic resistance in the vegetative canopy (s m^{-1}), and R_g is the resistance to uptake at the ground (s m^{-1}).

2.2.2 Wet deposition algorithms

The wet deposition flux is calculated on an hourly basis, and summed to obtain the total flux for the user-specified period. The default output units for wet deposition flux are g m^{-2}.

Particle wet deposition

The wet deposition flux for particulate substances is calculated from the particle-phase washout coefficient as follows

$$F_{wp} = 10^{-3} \rho_p W_p r, \quad (18)$$

where F_{wp} is the flux of particulate matter by wet deposition (μg m^{-2} hr^{-1}), ρ_p is the column average concentration of particulate in air (μg m^{-3}), W_p is the particle washout coefficient (dimensionless), and r is the water or water equivalent precipitation rate (mm hr^{-1}). The particle washout coefficient, W_p, is calculated as follows

$$W_p = \frac{3 z_p E}{2 D_m}, \quad (19)$$

where E is the collision efficiency (dimensionless), D_m is the mean diameter of raindrop (m) = $r^{0.232}/905.5$ with r in mm hr^{-1}, and z_p is the height of the top of plume or mixing height, whichever is greater (m). It is assumed that the washout coefficient, W_p, and therefore, the wet deposition flux, F_{wp}, is the same for frozen precipitation as for liquid precipitation. The collision efficiency is calculated after *Slinn* (1984) and *Seinfeld* and *Pandis* (1998).

Gaseous wet deposition

The wet deposition flux for gases is calculated as follows

$$F_{wg} = 10^6 C_l M_w r , \tag{20}$$

where F_{wg} is the flux of gaseous pollutants by wet deposition (μg m^{-2} hr^{-1}), C_l is the concentration of pollutant in the liquid phase (moles $liter^{-1}$), and M_w is the molecular weight of pollutant (grams $mole^{-1}$).

3. Results of the sensitivity analysis

3.1 Case studies of the planetary boundary layer

We made several test runs to examine the planetary boundary layer structure and the behavior of several meteorological and surface parameters. We especially examined the cases, when the short time (1-hour) averaging period concentration values were high in the modeling domain. This sensitivity analysis was made with the usage of flat and elevated terrains. In every model run the same source parameters were used as in Part I (*Steib* and *Labancz*, 2005), and the modeling domain was also the same.

3.1.1 Flat terrain

The maximum 50 1-hour average concentration values of NO_x were selected in the first 8 months of 2005. The highest 1-hour concentration was 22.28 μg m^{-3}, while the 50th highest was 14.36 μg m^{-3} in the modeling domain. We found that in 43 of 50 cases the maximum 1-hour concentration occurred during CBL and only in 7 of 50 cases during the SBL.

The 43 cases of the CBL could be divided into two major groups. The wind direction was parted to 16 sectors, and each of these sectors represents a unique group. The first group involved 10 cases, when the wind blew from west-southwest (236.25°–258.75°), the wind speed was low (1.0–1.2 m s^{-1}), and the surface roughness length was between 0.061 and 0.139 meters. The second group involved 15 cases, when the wind blew from the north (348.75°–11.25°), the wind speed was medium (2.9–6.6 m s^{-1}), and the surface roughness length was between 1.128 and 1.339 meters. The other 18 cases could be ordered in the other 14 groups. It should be mentioned, that the examined source (stack) is lying at the south border of the city of Pécs. The model

always uses the roughness length, which is characteristic of the area from where the wind blows, so it is not surprising that if the wind blows from the north (wind blows from urban area), then the value of the surface roughness length is much higher than that of the case of southwesterly winds (wind blows from rural area). It is also interesting, that the maximum 5 1-hour average concentration values (22.28–17.5 μg m^{-3}) could be assigned to the first group. In case of the 5 highest 1-hour concentrations the wind direction was 239°–251°, the wind speed was 1–1.2 m s^{-1}, and the roughness length was 0.061 meters. It is very likely, that this is an optimal combination of the wind speed and roughness length to result in the highest concentration values in the surface receptor points during CBL.

In 4 of the 7 cases of the SBL the wind direction was northerly (348.75°–11.25°), the wind speed was relatively strong (7.7–8.9 m s^{-1}), and the roughness length was between 1.212 and 1.339 meters. It can be seen that in spite of the stable conditions, the depth of PBL was high, because of the strong wind and high roughness length. That could be the reason for the good mixing of the polluted matter in the boundary layer during stable conditions too, and for the high concentration values at the surface receptor points.

In *Fig. 4*, the relative frequency of the wind direction during the maximum 50 1-hour average concentration is shown, when flat terrain was used.

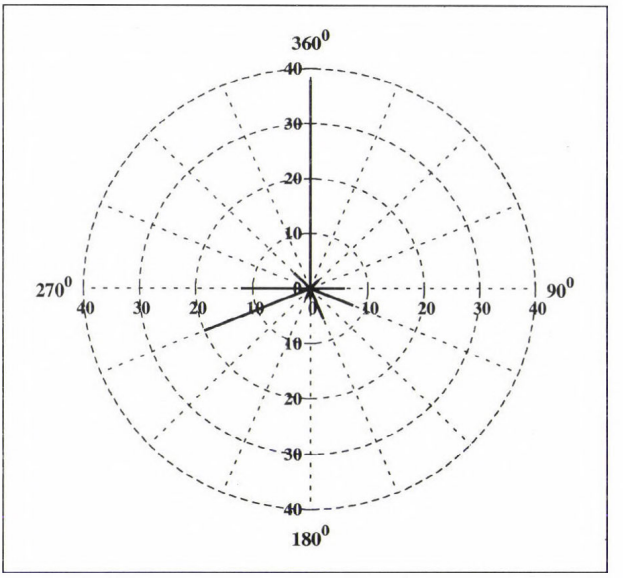


Fig. 4. Relative frequency of the wind direction (flat terrain) in case of the maximum 50 1-hour average concentration (%).

3.1.2 Elevated terrain

After the model run with elevated terrain, the maximum 50 1-hour average concentration values of NO_x were selected in the first 8 months of 2005. The first highest 1-hour concentration was 218.69 μg m^{-3}, and the 50th highest was 142.63 μg m^{-3}. We found that in all of the 50 cases the maximum 1-hour concentration occurred during SBL. In all cases the wind direction could be ordered into two groups. The first group involved 34 cases, when the wind direction was south-easterly (123.75°–146.25°), the second group involved 16 cases, when the wind direction was south-southeasterly (146.25°–168.75°). In all of the 50 cases the wind speed was low (0.5–1.7 m s^{-1}), and the surface roughness length was between 0.061 and 0.139 meters, except for two cases, when the height of the SBL was lower than or equal to 56 meters. The highest concentration values in the receptor points could be found, when the height of the stable boundary layer was lower than the height of the stack (80 m). At the 38th highest 1-hour concentration the SBL height was 165 meters, and at the 33rd it was even higher with 665 meters. It can be seen again, that the smoothing effect plays an important role. In both cases the wind direction changes to the opposite direction (from north-westerly to south-easterly) in short time, and this results in that the surface roughness length also decreases from about 0.906 meters to 0.061 meters. The wind speed also falls from 4 m s^{-1} to 1 m s^{-1} within a short time. The decrement of the wind speed and roughness length results in a sudden drop in the unsmoothed mechanical mixed layer height (z_{ie}), but the smoothing effect does not allow a sudden decrement of the current hour's smoothed mechanical mixed layer height (z_{im}). It is interesting to mention, that in case of the maximum 14 1-hour average concentration values (218.69–177.62 μg m^{-3}), the elevation over sea level of the receptor points with the highest 1-hour concentration values was the following: 299, 329, 334, and 338 meters. The elevation of the stack base over sea level was 150 m, so the highest concentration values were found 150–190 m above the stack base and 70–110 m above the stack top, consequently the following conclusions could be drawn. In case of modeling with elevated terrain, the highest concentration values could be found in those receptor points, which were at the height of the plume mass center.

In *Fig. 5*, the relative frequency of the wind direction during the maximum 50 1-hour average concentration is shown when elevated terrain was used. Although the same source parameters were used in case of flat and elevated terrain, the difference between the concentration values is significant. The concentration values by elevated terrain are about 10 times higher than those by flat terrain. These test runs made it clear again, that the model is the most sensitive to the terrain height of the modeling domain (also shown in Part I).

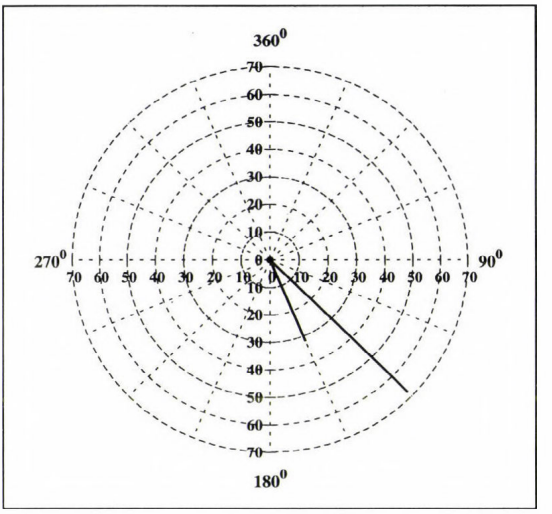


Fig. 5. Relative frequency of the wind direction (elevated terrain) in case of the maximum 50 1-hour average concentrations (%).

3.2 Case studies of the dry and wet deposition algorithms

In the sensitivity analysis to the input parameters of the dry and wet deposition algorithms, the same meteorological file was used as mentioned above (hourly meteorological data of the city Pécs, from the first 8 months of 2005). In this study all deposition fluxes were calculated on an hourly basis, and summed to obtain the total fluxes of this 8-month period, so the default output units for deposition fluxes are g m^{-2} for this 8-month time period.

3.2.1 Particle deposition flux

In our sensitivity studies the sensitivity of the particle deposition algorithms were examined based on Method 2. When using Method 2, the user must give two inputs to run the model: the mean particle diameter and the fine mass fraction of the particle. The value of the fine mass fraction can vary between 0 and 1, and the dimension of mean particle diameter must be μm.

In all of our sensitivity tests to particle deposition, we used flat terrain and a 1-dimensional polar coordinate system (only the y-coordinate changes between 100 and 20,000 m on north of the source) with the following source parameters:

- one point source (stack),
- type of pollutant: chromium,

- emission rate: 10 g s^{-1},
- release height above ground: 35 m,
- stack gas exit temperature: 400 K,
- stack gas exit velocity: 10 m s^{-1},
- stack inside diameter: 2.5 m.

In *Fig. 6*, the relationship between the period-averaged (8 months) surface concentration, the period-summed (8 months) dry and wet deposition fluxes is shown. It can be seen in this figure, that the structure of the concentration curve and dry deposition curve is very similar. The curve of the period-average concentration has its maximum place at 700 meters from the source, and the period-summed dry deposition curve at 600 meters. The character of the period-summed wet deposition curve is much more different. The highest wet deposition values can be found directly at the source, and it decreases rapidly as the distance from the source increases. The behavior of the dry and wet deposition curves is not surprising. The dry deposition depends on the concentration value, in consequence the character of the dry deposition curve must be similar to the character of the concentration curve. In case of wet deposition, the particles reach the ground with the falling raindrops, so the particle is able to reach the ground in a very short time. The highest column average concentration of airborne particles can be found directly at the stack exit, so we could detect the highest wet deposition values at the nearest receptor point to the source.

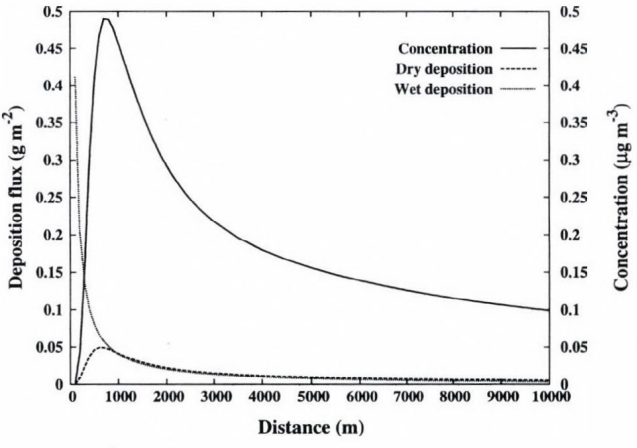


Fig. 6. Relationship between the period-averaged surface concentration, and the period-summed dry and wet deposition fluxes. Solid line represents the period averaged concentration (µg m^{-3}), dashed line represents the period dry deposition flux (g m^{-2}), and the dot line represents the wet deposition flux (g m^{-2}).

We also examined, how the change of the two input parameters (mean particle diameter and fine mass fraction) for Method 2 influences the dry and wet deposition fluxes. *Fig. 7* shows that only the dry deposition flux is dependent on the fine mass fraction. In the test runs the value of the fine mass fraction was changed from 0 to 1 with an increment of 0.1. In all cases a mean particle diameter of 2.5 μm was used. The figure shows that during the increment of the fine mass fraction, the dry deposition flux decreases linearly and the wet deposition flux remains constant. The conclusion is that the particle size distribution does not influence the wet deposition flux.

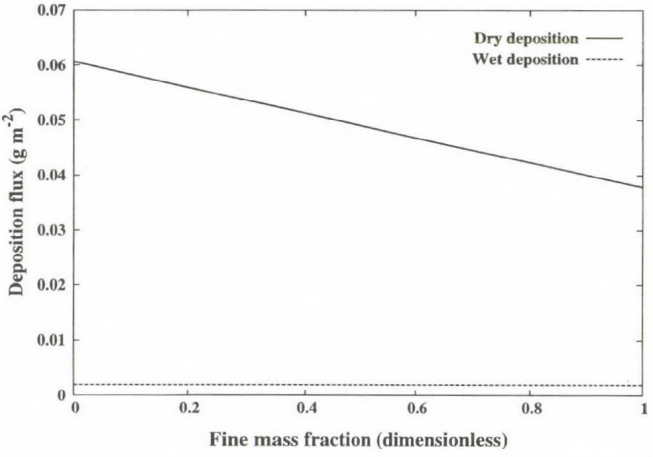


Fig. 7. Change of the dry and wet deposition fluxes as a function of the fine mass fraction. Solid line represents the dry deposition flux (g m^{-2}). Dashed line represents the wet deposition flux (g m^{-2}).

Fig. 8 represents the relationship between the mean particle diameter and deposition fluxes. In this case we can recognize the opposite effect of the deposition fluxes. In this model runs the value of the mean particle diameter was changed between 1 and 10 μm with a step of 1 μm. In all tests the fine mass fraction was 0.5. In this case the dry deposition flux remained nearly constant, but the wet deposition flux suddenly increased at the mean particle diameter of 3 μm. At a particle diameter of 1 μm, the value of the period-summed wet deposition flux was 0.00065 g m^{-2}, and the period-summed dry deposition flux was 0.04898 g m^{-2}. At 10 μm the wet deposition flux was 1.65558 g m^{-2} and the dry deposition flux was 0.04930 g m^{-2}. In this case the conclusion can be drawn, that the mean particle diameter hardly influences the dry deposition flux.

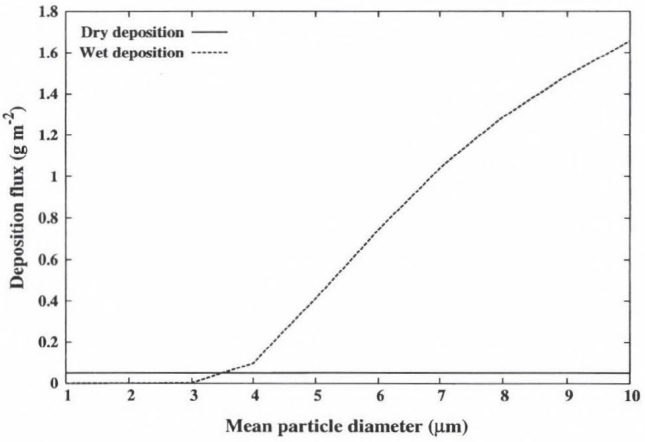


Fig. 8. Change of the dry and wet deposition fluxes as a function of the mean particle diameter. Solid line represents the dry deposition flux (g m^{-2}). Dashed line represents the wet deposition flux (g m^{-2}).

3.2.2 Gaseous deposition flux

When using gaseous deposition algorithms, the user must supply six mandatory inputs to run the model. These inputs are: diffusivity in air for the modeled pollutant (D_a), diffusivity in water for the modeled pollutant (D_w), cuticular resistance to uptake by lipids for individual leaves (*rcl*), Henry's Law constant (H), the Wesely seasonal categories, and the land use types. The gaseous deposition algorithms include some gas deposition resistance terms based on five seasonal categories (for each calendar month) and nine land use categories. The user should define land use categories for each of the 36 wind direction sectors.

Some model runs were made to examine the sensitivity of the gaseous deposition algorithms to the input parameters. The source parameters were the same as in the test runs of the particle deposition, except that the pollutant was benzene in every model run. We found that the particle deposition algorithms are more sensitive to the input parameters than the gaseous deposition algorithms. The diffusivity in air influences the dry deposition flux, while the diffusivity in water influences the wet deposition flux in the following way. When the diffusivity increases, the deposition fluxes also increase. The change of the cuticular resistance influences only the dry deposition flux. As the cuticular resistance increases, the dry deposition flux decreases. The Henry's Law constant influences the dry and also the wet deposition flux. As the

Henry's Law constant increases, the deposition fluxes decrease. The land use type influences only the dry deposition flux. *Table 1* shows the period-summed dry deposition flux at the 9 different land use categories. It can be clearly seen, that it is the wetness of the surface which most influences the dry deposition flux. In the case of a wet surface (bodies of water, wetland), the dry deposition flux is 1000–10000 times higher than in the case of a dry surface (urban area, desert).

Table 1. Dependence of the dry deposition flux on the land use type

Land use category	Description	Dry deposition flux (g/m^2)
1	Urban land, no vegetation	0.00000
2	Agricultural land	0.00091
3	Rangeland	0.00085
4	Forest	0.00089
5	Suburban areas, grassy	0.00079
6	Suburban areas, forested	0.00084
7	Bodies of water	0.30005
8	Barren land, mostly desert	0.00003
9	Non-forested wetland	0.05156

In our sensitivity analysis of the deposition algorithms, we examined only the model's sensitivity to the user defined deposition input parameters. The sensitivity of the deposition algorithms to the input meteorological parameters was not examined.

5. Conclusions

In this paper the basic planetary boundary layer algorithms of AERMET, and the deposition algorithms of AERMOD were presented. Then some test runs were made in order to examine the model's sensitivity to the input parameters of the boundary layer and deposition algorithms. It was not new that the wind direction, the wind speed, the surface roughness length, and the height of the planetary boundary layer are very important parameters when making concentration calculations, but it was more interesting to see, that the concentration distribution of the model was most sensitive to the terrain's elevation. It was also shown that the particle dry deposition flux (Method 2) is independent from the mean particle diameter, and the particle wet deposition

flux (Method 2) is independent from the fine mass fraction. Finally, the behavior of the gaseous deposition was presented. We found that the sensitivity of the gaseous dry deposition flux was higher than the sensitivity of the gaseous wet deposition flux to the input parameters.

The development of the AERMOD model is ongoing. We use the model mainly for regulatory purposes. We think that the sensitivity analysis of AERMOD in this paper was very useful, since our plan is to use the model also for other purposes in the future.

References

Carson, D.J., 1973: The development of a dry inversion-capped convectively unstable boundary layer. *Q. J. Roy. Meteor. Soc. 99*, 450-467.

Deardorff, J. W., 1980: Progress in Understanding Entrainment at the Top of a Mixed Layer.

Holtslag, A.A.M. and *van Ulden, A.P*, 1983: A simple scheme for daytime estimates for the surface fluxes from routine weather data. *J. Clim. Appl. Meteorol. 22*, 517-529.

Perry, S.G., 1992: CTDMPLUS: A dispersion model for sources in complex topography. Part I: Technical formulations. *J. Appl. Meteorol. 31*, 633-645.

Seinfeld, J.H. and *Pandis, S.N.*, 1998: *Atmospheric Chemistry and Physics*. John Wiley & Sons, Inc., New York, NY.

Slinn, W.G.N., 1984: Precipitation scavenging. In *Atmospheric Science and Power Production* (ed. *D. Randerson*). DOE/TIC-27601. U.S. Department of Energy, Washington, DC.

Steib, R., Labancz, K., 2005: Regulatory modeling in Hungary – the AERMOD model. Part I. Description and application. *Időjárás 109*, 157–172.

Szepesi, D.J., Fekete, K.E., Gyenes, L., 1995: Regulatory models for environmental impact assessment in Hungary. *Int. J. Environ. Pollut. 5*, 497-507.

Venkatram, A., 1980: Estimating the Monin-Obukhov length in the stable boundary layer for dispersion calculations. *Bound.-Lay. Meteorol. 19*, 481-485.

Weil, J.C. and *Brower, R.P.*, 1983: Estimating convective boundary layer parameters for diffusion applications. PPSP-MD-48, Maryland Power Plant Siting Program, Maryland Department of Natural Resources, Baltimore, MD, 45pp.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 1, January–March 2006, pp. 35–48

Comparing tendencies of some temperature related extreme indices on global and regional scales

Judit Bartholy and **Rita Pongrácz**

Department of Meteorology, Eötvös Loránd University
P.O. Box 32, H-1518 Budapest, Hungary
E-mails: bari@ludens.elte.hu, prita@elte.hu

(Manuscript received in final form February 6, 2006)

Abstract—Climate extreme indices are analyzed and compared for the Carpathian Basin for the 20th century based on the guidelines suggested by the joint WMO-CCl/CLIVAR Working Group on climate change detection. These climate extreme indices include the numbers of severe cold days, winter days, frost days, cold days, warm days, summer days, hot days, extremely hot days, cold nights, warm nights, the intra-annual extreme temperature range, the heat wave duration, the growing season length, etc. Therefore, daily maximum, minimum, and mean temperature observations are used in the present statistical analysis. Our results suggest that similarly to the global and continental trends, regional temperature of Central/Eastern Europe became warmer during the second half of the 20th century.

Key-words: extreme, climate index, daily maximum and minimum temperature, Carpathian Basin, Europe, trend analysis

1. Introduction

Since human and natural systems may be especially affected by changes of extreme climate events, the main objective of our research is to detect the possible changes of intensity and frequency of these extreme events. Previously, we presented the results of the analysis of extreme precipitation indices for the Carpathian Basin (*Bartholy* and *Pongrácz,* 2005a). This paper focuses on the extreme indices related to temperature.

According to the *IPCC* (2001), the detected shift in global mean temperature may result in more frequent extreme events. This is highlighted in the following example by *Schar et al.* (2004). They compared climate anomaly

time series based on past observations and model simulations for Switzerland. In the left panel of *Fig. 1*, summer precipitation and temperature anomalies of the last 140 years are presented. The right panel of *Fig. 1* shows the simulated anomalies for the 1961–1990 and 2071–2100 periods. Extreme heat waves occurred in Europe in summer, 2003. As the scatter plot diagrams highlight, the summer of 2003 was an extreme season compared to the past decades, while it can be considered as normal when looking at the simulated climate of Switzerland at the end of the 21st century.

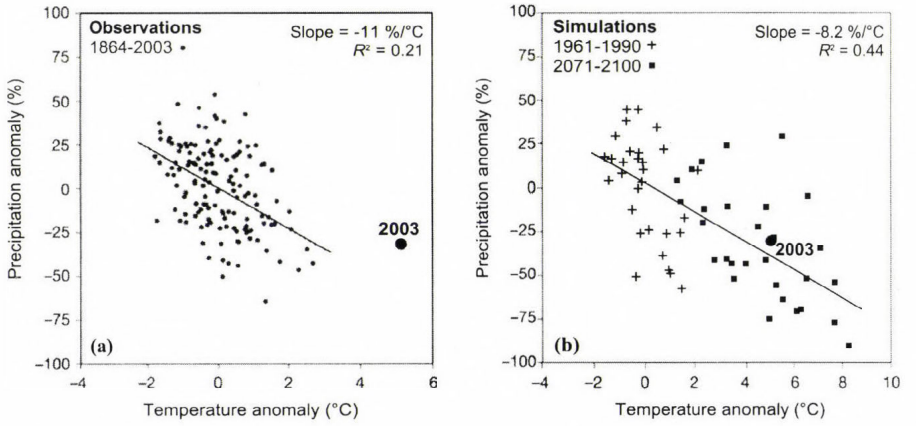


Fig. 1. Scatter plot diagrams of summer (JJA) mean temperature and precipitation anomalies for Northern Switzerland (*Schar et al.*, 2004), (a) on the base of observed data in 1864–2003, (b) on the base of simulated data for 1961–1990 (crosses), and for 2071–2100 (rectangles). Reference period is 1961–1990.

Giorgi and *Francisco* (2000) analyzed the future continental temperature and precipitation changes expected for the 21st century, based on model outputs of five main AOGCMs (Atmosphere-Ocean General Circulation Model). Global continental areas were divided into 23 regions, from which 2 cover the European continent, namely, (i) Northern Europe (NEU), and (ii) the Mediterranean region (MED). *Fig. 2* summarizes intensity, sign, and consistency of the GCM-based temperature changes for these two European regions. Expected changes in temperature conditions are presented in four small boxes for NEU and MED for 2071-2100. The upper and lower two boxes represent expected changes in winter (December-January-February), and in summer (June-July-August), respectively. Furthermore, results for the GG (greenhouse gas only case) and GS (greenhouse gas with increasing sulfate aerosol case) scenarios are shown in the left two boxes, and in the right two

boxes, respectively. The size of the "+" sign appearing in the small boxes indicate the intensity of temperature change compared to the baseperiod 1961–1990. Similarly, black and grey colors imply large (greater than 20%), and small (between 5% and 20%) average change, respectively. In case of model disagreement sign of inconsistency ("i") appears in the small box. In general, results of *Giorgi* and *Francisco* (2000) suggest that both winter and summer in Europe tend to be warmer than the present climate. Large increase of temperature can be expected in winter in Northern Europe and in summer in Southern Europe in case of the GS scenario. Estimations of the five main AOGCMs are inconsistent for NEU summers for the GG scenario.

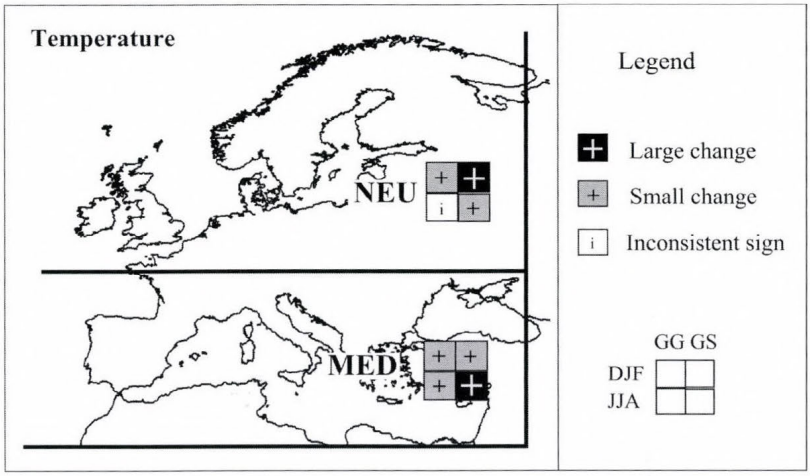


Fig. 2. Summary of temperature tendency analysis for Northern Europe (NEU) and the Mediterranean region (MED) for two seasons (winter and summer) and two climate scenarios (GG, GS), based on inter-model consistency analysis of *Giorgi* and *Francisco* (2000).

The *IPCC* (1995) concluded that changes in both the mean and extreme climate parameters may strongly affect human and natural systems. Furthermore, several research projects published results on the analysis of climate extremes on global or continental scales (e.g., *Easterling et al.,* 2000; *Peterson et al.,* 2002, etc.), and in order to strengthen the international cooperation, a Workshop on Indices and Indicators for Climate Extremes was held in Asheville (located in North Carolina, USA) in 1997. The required data sets for climate extreme analysis and the list of common climate extreme indices were compiled at this meeting (*Karl et al.,* 1999). Taking into account

the suggestions of the workshop, a joint WMO-CCl/CLIVAR Working Group formed on climate change detection in 1998 (*Frich et al.,* 2002). Results of their global and regional climate extreme analysis were provided for the *IPCC* (2001).

The next section of this paper summarizes and compares the results of the global (*Frich et al.,* 2002) and European (*Klein Tank* and *Können,* 2003) extreme analysis. Similar methodology has been applied to climate extremes of the Carpathian Basin. Section 3 discusses the continental (for Europe) and regional (for the Carpathian Basin) extreme tendencies. Finally, Section 4 concludes the main findings of this paper.

2. Climate extreme indices, comparison of global and European analysis

The CCl/CLIVAR task group on extreme indices compiled a global climate database containing daily precipitation, maximum, minimum, and mean temperature time series for the period 1946-1999. The main data sources include the national meteorological services, NOAA NCDC (*Peterson* and *Vose,* 1997), and the European Climate Assessment project (*Klein Tank et al.,* 2002b). All the datasets have been quality controlled and adjusted for inhomogeneities. Then, the following general criteria have been used for including an observation station: (i) from the entire 1946-1999 period, data must be available for at least 40 years, (ii) missing data cannot be more than 10%, (iii) missing data from each year cannot exceed 20%, (iv) in each year, more than 3 months consecutive missing values are not allowed.

Results of the global and European extreme climate analysis were published in *Frich et al.* (2002), and *Klein Tank* and *Können* (2003), respectively. In this section these results are summarized and compared for the global and continental scales. *Table 1* presents the main extreme indices that the CCl/CLIVAR task group identified and suggested for global climate extreme analysis. Besides the definition of each extreme index, the table indicates the scale (i.e., global, continental, regional) of application, as well. From the total 14 extreme temperature indicators, 5, 6, and 14 were used in the analysis for the world (*Frich et al.,* 2002), Europe (*Klein Tank* and *Können,* 2003), and the Carpathian Basin (*Bartholy* and *Pongrácz,* 2005b), respectively. For instance, *Figs. 3–5* present one of the climate extreme indices, namely, the change of the number of frost days ($T_{min} < 0°C$). Spatial distribution of global tendencies can be seen in *Fig. 3*, while the graph shown in *Fig. 4.* provides the temporal details of the global mean change of the number of frost days during the second half of the 20th century. Results of the similar analysis for Europe is presented in *Fig. 5.*

Table 1. Definition and indicator of extreme climate parameters

No.	Indicator (ECAD)	World (*Frich et al.*, 2002)	Europe (*Klein Tank* and *Können*, 2003)	Carpathian Basin (*Bartholy* and *Pongrácz*, 2005b)	Definition of the extreme temperature index	Unit
1	ETR	x		x	Intra-annual extreme temperature range (difference between the observed maximum and minimum temperatures, T_{max}–T_{min})	°C
2	GSL	x		x	Growing season length (start: when for more than 5 consecutive days $T>5$ °C, end: when for more than 5 consecutive days $T<5$°C)	day
3	HWDI	x		x	Heat wave duration index (for min. 5 consecutive days $T_{\max} = T^{N}_{\max} + 5$ °C, where $T^{N}_{\max}$ indicates the mean T_{max} for the baseperiod 1961–90)	day
4	Tx10		x	x	Cold days (percent of time when $T_{max}<10$th percentile of daily maximum temperature, based on the baseperiod 1961–90)	%
5	Tx90		x	x	Warm days (percent of time when $T_{max}>90$th percentile of daily maximum temperature, based on the baseperiod 1961–90)	%
6	Tn10		x	x	Cold nights (percent of time when $T_{min}<10$th percentile of daily minimum temperature, based on the baseperiod 1961–90)	%
7	Tn90	x	x	x	Warm nights (percent of time when $T_{min}>90$th percentile of daily minimum temperature, based on the baseperiod 1961–90)	%
8	FD	x	x	x	Number of frost days ($T_{min}<0$ °C)	day
9	SU		x	x	Number of summer days ($T_{max}>25$ °C)	day
10	Tx30GE			x	Number of hot days ($T_{max}\geq 30$ °C)	day
11	Tx35GE			x	Number of extremely hot days ($T_{max}\geq 35$ °C)	day
12	Tn20GT			x	Number of hot nights ($T_{min}>20$ °C)	day
13	Tx0LT			x	Number of winter days ($T_{max}<0$ °C)	day
14	Tn-10LT			x	Number of severe cold days ($T_{min}<-10$ °C)	day

Changes between the two subperiods of the second half of the century (1946–1975 and 1976–1999) have been determined during the analysis presented in *Frich et al.* (2002). The world map of *Fig. 3.* indicates both the sign of the change (gray and black circles for decreasing and increasing tendencies, respectively) and the magnitude of the change (applying 4 different circle sizes for different percentage intervals) at each station involved in the analysis. Stations with significant changes (at 95% level of confidence) in annual number of frost days are mapped with filled circles, while open circles indicate not significant changes. The large number of grey filled circles and very few black circles on the map suggest that the annual number of frost days decreased considerably between 1946 and 1999.

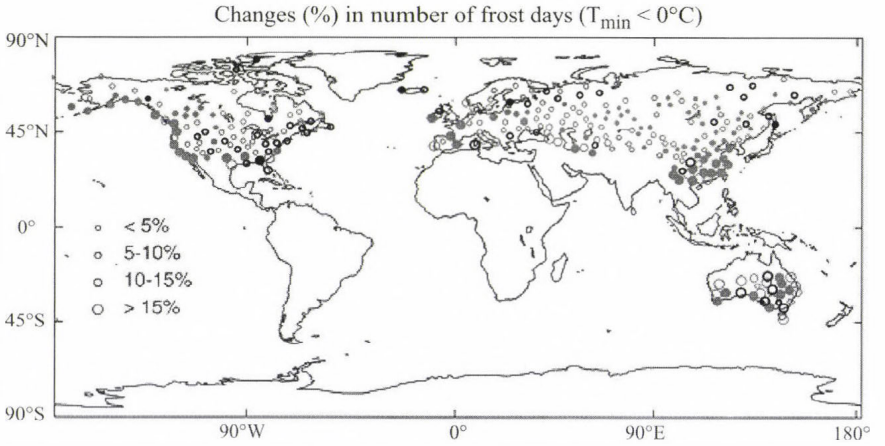


Fig. 3. Changes (%) in number of annual frost days in the second half of the 20th century. Filled circles are significant at the 95% level of confidence. Grey and black indicate negative and positive changes, respectively. Circle sizes represent the magnitude of change (source: *Frich et al.*, 2002).

Based on the available time series, annual global weighted mean anomalies have been calculated using the baseperiod of 1961–1990. *Fig. 4.* presents the annual value (in percentage) of the anomaly for the entire 1950–1999 period, and also, the fitted linear trend emphasizing the significant decreasing tendency. The figure includes a small graph (in its upper right part) indicating the total number of stations used for the analysis in each year. Except the beginning and the end of the period, about 300 stations provided valuable temperature data to determine the annual number of frost days.

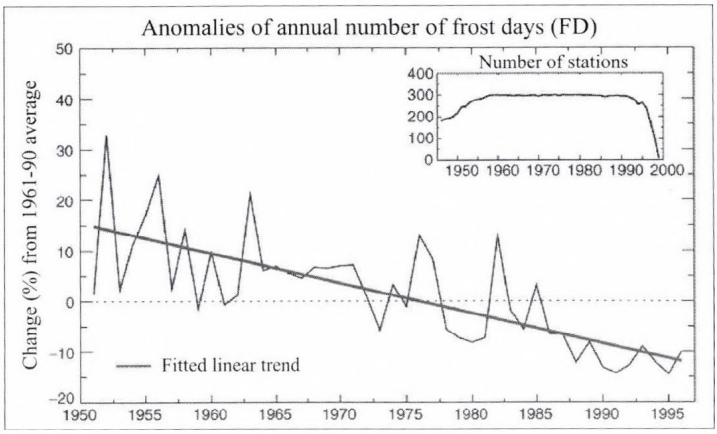


Fig. 4. Mean annual values of the number of frost days in percentage differences from the 1961–1990 weighted average values for the second half of the 20th century. The inserted graph represents the weighting factors (number of stations with valuable data) used in the linear regression analysis. The fitted linear trend is statistically significant at the 95% level of confidence (source: *Frich et al.*, 2002).

The European tendency analysis of the annual number of frost days is shown in *Fig. 5*, where the mean decadal changes of this extreme index is mapped for the stations with sufficient data for the 1946–1999 period. Open circles indicate not significant changes, while dark and grey filled circles indicate negative and positive trends, respectively. Similarly to the global analysis, significant negative tendency can be seen. Summarizing the above results of the two large scale analyses, annual number of frost days considerably decreased by the end of the 20th century.

Table 2 summarizes the tendencies of nine extreme indices for the global and continental scale climate analysis based on the papers of *Frich et al.* (2002), and *Klein Tank* and *Können* (2003), respectively. The comparison of these results is accomplished for the second half of the last century (1946–1999). Four indices (Tx10, Tx90, Tn10, and SU) are analyzed only on European scale. Increasing and decreasing trends are indicated with symbols „+" and „–", respectively. Two identical symbols represent large tendencies. Considerable spatial differences are emphasized using more than one type of symbols (e.g., –/+, ++/–, etc.), after identifying the main dominant trend, exceptions are listed in case of each extreme climate index. In general, global and European trends are similar, and refer to a warming climate tendency. Only a few small areas differ from these worldwide and continental dominant trends. For instance, in case of SU (summer days), Eastern Europe belongs to the exceptions. One of the aims of our research presented in this paper is to specify the trends on a finer spatial scale, and provide more details for the Carpathian Basin and Hungary.

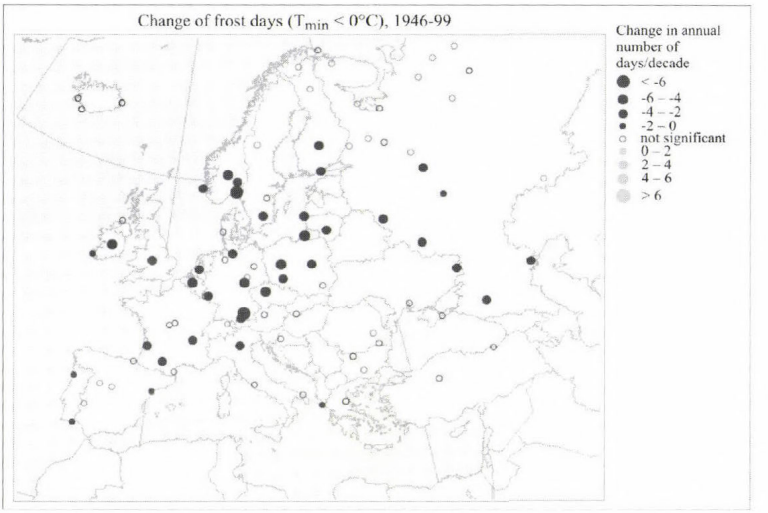


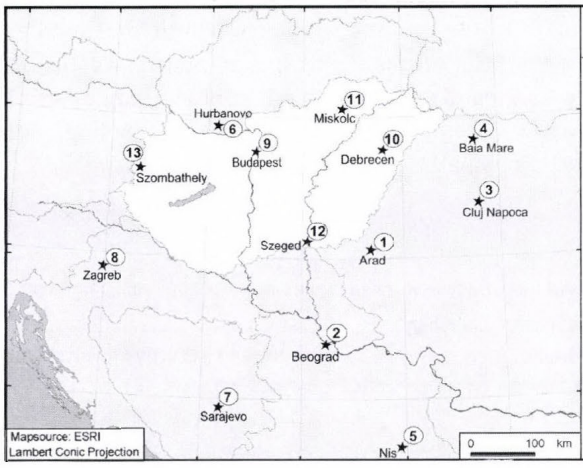
Fig. 5. Decadal trend in the annual number of frost days in Europe for the period 1946–1999. Circles are scaled according to the magnitude of the trend. Open circles indicate not significant changes, while dark and grey filled circles indicate negative and positive trends, respectively (source: *Klein Tank* and *Können,* 2003).

Table 2. Comparison of the tendencies of extreme climate indices, based on global (*Frich et al.,* 2002) and European (*European Climate Assessment & Dataset* project, *Klein Tank* and *Können*, 2003) extreme analysis for the period 1946–1999

No.	Extreme index	World (*Frich et al.,* 2002)	Europe (*Klein Tank* and *Können*, 2003)
1	ETR Intra-annual extreme temperature range	–	–
2	GSL Growing season length	+	+ + / – Positive tendency dominates except Iceland
3	HWDI Heat wave duration index	+ + / – Positive tendency dominates except SE-Asia and the eastern part of North-America	+
4	Tx10 Cold days	No analysis provided	+
5	Tx90 Warm days	No analysis provided	+ / – Positive tendency dominates except Iceland, Italy, and the Black Sea region
6	Tn10 Cold nights	No analysis provided	+ +
7	Tn90 Warm nights	+ +	+ + / – Positive tendency dominates except Iceland and the Black Sea region
8	FD Frost days	– –	– –
9	SU Summer days	No analysis provided	+ / – Positive tendency dominates except Eastern Europe

3. Analysis of extreme tempearture indices for the Carpathian Basin

In our analysis for the Carpathian Basin, daily temperature data from 13 meteorological stations are used. *Fig. 6* shows their geographical location. Minimum, maximum, and mean temperature time series of the 8 stations located outside Hungary are available from the ECAD site via the Internet (*Klein Tank,* 2003), while data from the 5 Hungarian stations are from the Data Archive of the Hungarian Meteorological Service. Two basic constraints are taken into account during the selection of the stations: (i) covering the area of the Carpathian Basin with the best spatial homogeneity and representing the main climatic subregions, (ii) time series without the least missing values during the 1961–2001 period. The analysis presented in this paper, is focused on the Carpathian Basin, however, two of the selected stations (Nis and Sarajevo) are outside this region. We included them in order to accomplish the analysis on a larger area.



Station	Latitude (N)	Longitude (E)	Height above sea level (m)
1 Arad	46°08'	21°21'	117
2 Beograd	44°48'	20°28'	132
3 Cluj Napoca	46°47'	23°34'	410
4 Baia Mare	47°40'	23°30'	216
5 Nis	43°20'	21°54'	202
6 Hurbanovo	47°53'	18°12'	115
7 Sarajevo	52°30'	22°30'	211
8 Zagreb	45°49'	15°59'	157
9 Budapest	47°31'	19°02'	118
10 Debrecen	47°33'	21°37'	123
11 Miskolc	48°08'	20°48'	118
12 Szeged	46°15'	20°06'	82
13 Szombathely	47°15'	16°38'	218

Fig. 6. Geographical locations of the 13 meteorological stations used in the regional scale analysis for the Carpathian Basin. Note that Nis and Sarajevo do not belong to this region, but they were included in the analysis.

On the base of our previous study of time series of mean temperature and extreme temperature parameters, a strong warming tendency was detected from the middle of the 1970's (*Pongrácz* and *Bartholy*, 2000). Therefore, the entire 1961–2001 period has been separated into two subperiods, namely, 1961–1975 and 1976–2001. The tendency analysis has been accomplished for these subperiods. *Table 3* summarizes the increasing (+) and decreasing (–)

tendencies of the indices for the entire 41 years and for the two subperiods (15 and 26 years). Opposite sign of trend coefficients may indicate warming and cooling tendencies. For instance, negative coefficients of the number of cold days (Tx10) and positive coefficients of the number of hot days (Tx30GE) both indicate warming climate. Therefore, warming tendencies are shown in black boxes, while cooling tendencies in light grey. The trend coefficients of the index ETR (intra-annual extreme temperature range) are in white, since they do not imply either warming or cooling tendencies by themselves. Warming tendencies (in black) are dominant in the table. The regional climate of the Carpathian Basin tended to be warmer during the entire 41 years (except HWDI - heat wave duration index). In case of most of the extreme temperature indices, the three periods used in our analysis cannot be characterized by the same sign of trend coefficient. Only four extreme indices (Tn10 - cold nights, Tn-10LT - number of severe cold days, FD - number of frost days, Tn90 - warm nights) indicate warming tendency in the 1961-2001, 1961-1975, 1976-2001 periods. Based on the trend coefficients of HWDI, Tx90, SU, Tx30GE, Tx35GE, Tn20LT, the cooling tendencies until the middle of the 1970's is followed by a warming climate in the last quarter of the 20th century. Opposite tendency can be detected in case of two indices (Tx10, TxOLT) using regional scale average. However, these cooling trend coefficients of the last decades are small.

Table 3. Summary of the trend analysis of extreme temperature indices for the Carpathian Basin (warming and cooling trends are indicated by black and light grey color of the box, respectively)

No.	Extreme index	1961–2001	1961–1975	1976–2001
1	ETR: Intra-annual extreme temperature range	-	-	+
2	HWDI: Heat wave duration index	-	-	+
3	Tx10: Cold days	-	-	+
4	Tx90: Warm days	+	-	+
5	Tn10: Cold nights	-	-	-
6	Tn90: Warm nights	+	+	+
7	FD: Number of frost days	-	-	-
8	SU: Number of summer days	+	-	+
9	Tx30GE: Number of hot days	+	-	+
10	Tx35GE: Number of extremely hot days	+	-	+
11	Tn20GT: Number of hot nights	+	-	+
12	TxOLT: Number of winter days	-	-	+
13	Tn-10LT: Number of severe cold days	-	-	-

In this paper, detailed analysis is presented for the last quarter of the 20th century, when the largest changes occurred. Detailed tendency analysis of the indices Tn90 (warm nights) and Tx90 (warm days) are presented in *Fig. 7.* Trend maps for Europe and the Carpathian Basin are provided in the upper and middle panels, respectively, while the lower graphs show the regional mean index anomaly from the 1961–1990 average values for the Carpathian Basin. Circles represent decadal trend coefficients of the meteorological stations (using the baseperiod 1961–1990). Black and grey circles indicate increasing and decreasing tendencies, respectively, while circle size depends on the intensity of these positive or negative trends. In case of the regional mean, the fitted linear trends are clearly increasing between 1976 and 2001, in case of both indices. Also, no decreasing tendency can be identified in either map. The positive trend coefficients are significant at the 95% level of confidence.

The daily maximum temperature of summer is indicated by three extreme indices: (i) number of summer days (SU: $T_{max} > 25$ °C), (ii) number of hot days (Tx30GE: $T_{max} \geq 30$ °C), and (iii) number of extremely hot days (Tx35GE: $T_{max} \geq 35$ °C). As it can be seen from *Table 3,* increasing trend coefficients of these indices are detected during the entire 1961–2001 period and 1976–2001 subperiod, while they are decreasing in the 1961–1975 subperiod. *Fig. 8* presents the maps containing the increasing trend coefficients of extreme indices SU and Tx30GE in the Carpathian Basin in the last 26 years. Large positive trend coefficients dominate both maps, with more than 6 days per decade, in general. Tendency analysis map of the extreme index Tx35GE is not presented in this paper, since the frequency of this events is quite small, however, the trend coefficients are similar to those shown in case of SU and Tx30GE.

Similarly to *Fig. 7,* map with the trend coefficients of HWDI are shown in *Fig. 9.* As it can be seen on the maps, only significant increasing tendency of HWDI is detected in all of the stations in the last quarter of the 20th century. However, compared to the other indices, trend coefficients of more stations are not significant in Europe (left panel of the figure). The exact explanation is not known, but we can assume that the larger number of stations with insignificant tendency is related to the definition of this index. In the definition, the same 5 °C threshold is used in case of oceanic and continental climates, which may not be appropriate for all climates.

Based on the above figures, similarly to the global and European trends (*Frich et al.,* 2002; *Klein Tank et al.,* 2002a), analysis of the extreme temperature indices suggests that the regional climate of the Carpathian Basin tended to be warmer in the last 41 years.

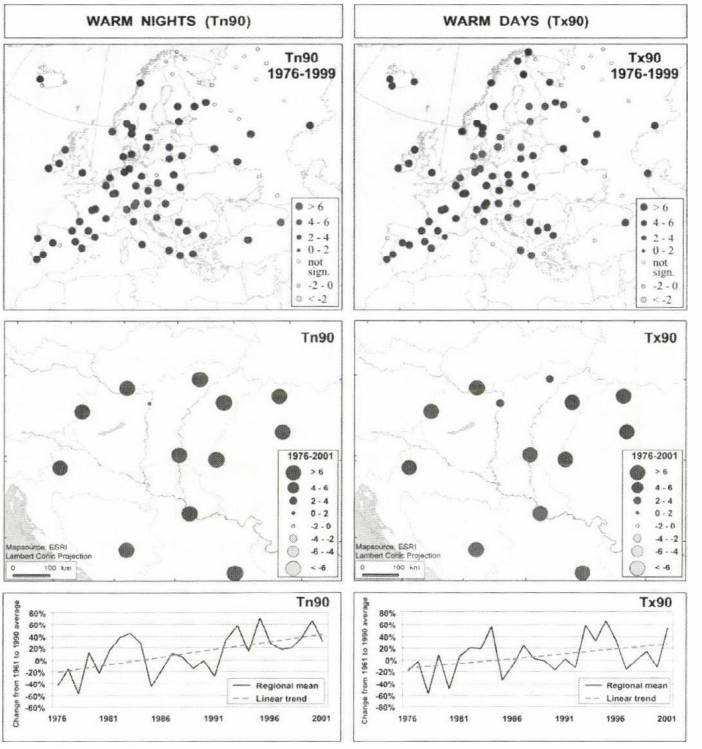


Fig. 7. Increasing tendency of warm nights (Tn90) and warm days (Tx90) in Europe and the Carpathian Basin during the last quarter of the 20th century. Trend coefficients of the Carpathian Basin greater than 0.4 in absolute value are significant at the 95% level of confidence.

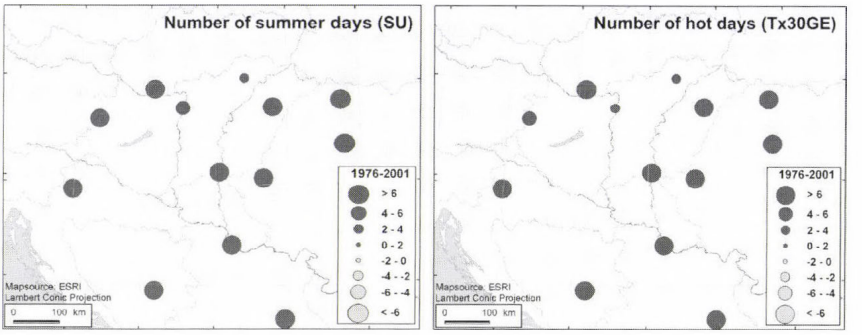


Fig. 8. Increasing tendency of the number of summer days (SU, $T_{max} > 25$ °C) and hot days (Tx30GE, $T_{max} \geq 30$ °C) in the Carpathian Basin during the last quarter of the 20th century. Trend coefficients greater than 0.4 in absolute value are significant at the 95% level of confidence.

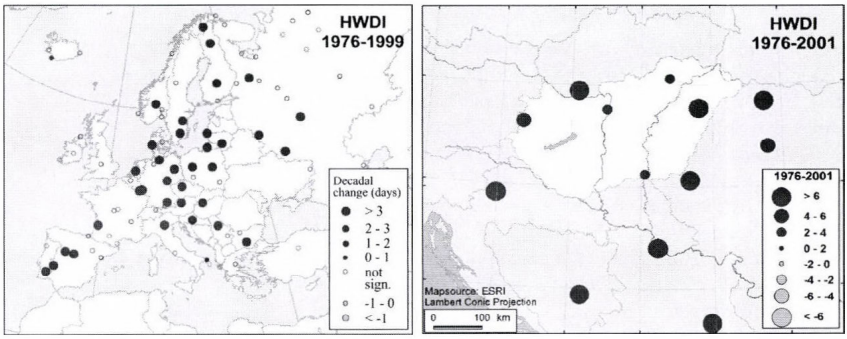


Fig. 9. Tendency of heat wave duration index (HWDI) in Europe and the Carpathian Basin during the last quarter of the 20th century. Trend coefficients of the Carpathian Basin greater than 0.4 in absolute value are significant at the 95% level of confidence.

4. Conclusions

The analysis of extreme temperature indices (according to the suggestions of the WMO-CCl/CLIVAR Working Group) are presented for the second half of the 20th century in this paper. Global and European trends of the extreme temperature indices are consistent with the global warming. As an example, the decreasing tendency of the number of frost days (FD) are presented on global (300 stations) and European (140 stations) scales. Based on the analysis of the extreme temperature indices for the Carpathian Basin, the following conclusions can be drawn.

- Significant warming tendencies are dominant during the entire 1961–2001 period.
- Most of the indices (e.g., HWDI, Tx90, SU, Tx30GE, Tx35GE, Tn20LT) show that the entire 41 years can be separated into a cooling period until the middle of the 1970's, and then a warming period in the last quarter of the 20th century.
- The largest trend coefficients (more than 6 days per decade) were detected in case of the following indices: Tn90, Tx90, SU, Tx30GE, HWDI.

Acknowledgements—Research leading to this paper has been supported by the Hungarian National Science Research Foundation (OTKA) under grants T-049824, T-034867, and T-038423, also by the CHIOTTO project of the European Union Nr. 5 program under grant EVK2-CT-2002/0163, the Hungarian National Research Development Program under grants NKFP-3A/0006/2002, NKFP-3A/082/2004, and NKFP-6/079/2005, and the VAHAVA project of the Hungarian Academy of Sciences and the Ministry of Environment and Water. ESRI software has been used to create maps.

References

Bartholy, J. and *Pongrácz, R.*, 2005a: Tendencies of extreme climate indices based on daily precipitation in the Carpathian Basin for the 20th century. *Időjárás 109*, 1-20.

Bartholy, J. and *Pongrácz, R.*, 2005b: Global and regional tendencies of extreme indices calculated on the base of daily temperature and precipitation for the 20th century (in Hungarian). *Agro-21. Füzetek*, 40, Magyar Tudományos Akadémia, Budapest, 70-93.

Easterling, D.R., Meehl, G.A., Parmesan, C., Chagnon, S.A., Karl, T., and *Mearns, L.O.*, 2000: Climate extremes: Observation, modelling and impacts. *Science 289*, 2068-2074.

Frich, P., Alexander, L.V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A.M.G., and *Peterson, T.*, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res. 19*, 193-212.

Giorgi, F., and *Francisco, R.*, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett. 27*, 1295-1298.

IPCC, 2001: *Climate Change 2001: Third Assessment Report. The Scientific Basis*. Cambridge University Press, Cambridge, UK.

IPCC, 1995: *Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Karl, T.R., Nicholls, N., and *Ghazi, A.*, 1999: Clivar/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary. *Climatic Change 42*, 3-7.

Klein Tank, A.M.G., 2003: *The European Climate Assessment and Dataset project*. http://www.knmi.nl/samenw/eca/index.html.

Klein Tank, A.M.G. and *Können, G.P.*, 2003: Trends in Indices of Daily Temperature and Precipitation Extremes in Europe, 1946-99. *J. Climate 16*, 3665-3680.

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V., and *Petrovic, P.*, 2002a: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol. 22*, 1441-1453.

Klein Tank, A.M.G., Wijngaard, J.B., and *van Engelen, A.*, 2002b: *Climate of Europe; Assessment of observed daily temperature and precipitation extremes*. KNMI, De Bilt, the Netherlands, 36p.

Peterson, T.C. and *Vose, R.S.*, 1997: An overview of the global historical climatology network database. *B. Am. Meteorol. Soc. 78*, 2837-2849.

Peterson, T., Folland, C.K., Gruza, G., Hogg, W., Mokssit, A., and *Plummer, N.*, 2002: Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs, 1998-2001. *World Meteorological Organisation Rep. WCDMP-47. WMO-TD 1071*. Geneva, Switzerland. 143p.

Pongrácz, R. and *Bartholy, J.*, 2000: Changing trends in climatic extremes in Hungary (in Hungarian). In *III. Erdő és Klíma Konferencia* (ed.: *A. Kircsi*). Kossuth Lajos Tudományegyetem, Debrecen, 38-44.

Schar, C., Vidale, P.L., Luthi, D., Frei, C., Haberli, C., Liniger, M.A., and *Appenzeller, C.*, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature 427*, 332-336.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 1, January–March 2006, pp. 49–62

Effects of documented land use changes on the albedo of Eastern Hungary (1951–2000)

János Mika[1*], **Szilvia Horváth**[2**] and **László Makra**[2]

[1]*Hungarian Meteorological Service*
P. O. Box 38, H-1525, Budapest, Hungary;
E-mail: mika.j@met.hu

[2]*Department of Climatology and Landscape Ecology, University of Szeged*
P.O. Box 653, H-6701 Szeged, Hungary; E-mail: makra@geo.u-szeged.hu

(Manuscript received in final form February 20, 2006)

Abstract—Agricultural land use series are investigated in a plain catchment area of the river Tisza within Hungary, almost identically represented by six administrative counties. Each county, commonly covering 34,900 km^2, is characterized by high percentage (72–82%) of managed vegetation. Effects of the area coverage variations between the different plant species are computed for the period 1951–2000, by applying results of literature-based syntheses specified for Hungary. The latter studies estimate surface albedo values for the great majority of the plant species grown in the region. Product of the plant-specific albedo values and the relative area coverage results in monthly series of surface albedo. Furthermore, by using a radiation transfer model, these series are also used to simulate radiation balance series for the surface-atmosphere system. Two questions are investigated and positively answered: (i) Are there monotonous trends in the radiation balance? (ii) Are these changes comparable to the effects caused by other external forcing factors?

Key-words: land use, albedo, radiation balance, climate change, Hungary.

1. Introduction

Solar radiation absorbed by the global Earth-atmosphere system is about 235 W m^{-2} (e.g., *IPCC*, 2001, Chapter 1.2.1). Since the 19th century, the atmospheric CO_2 has been responsible for 1.5 W m^{-2} primary radiation change

* Corresponding author

** Present affiliation: Ministry of Environment and Water Management, Environmental Policy and Impact Assessment Department, Fő u. 44-50, 1027 Budapest, Hungary; E-mail: horvathszil@mail.kvvm.hu

at the top of the atmosphere *(IPCC*, 2001, Chapter 6*)*. At the same time, CH_4 is computed to be responsible for 0.55 W m^{-2}, N_2O for 0.2 W m^{-2}, and CFC gases together for 0.3 W m^{-2}. Consequently, total growth of the radiation balance is 2.5 W m^{-2}, which as medium estimation, is equivalent to 1.5 K global temperature increase. However, only 0.7 K warming is observed. This difference is explained by the decrease of ozone in the stratosphere (- 0.2 W m^{-2}) and by the anti-greenhouse effect caused by tropospheric aerosols (-1 – -2 W m^{-2}), together with the enormous heat capacity of the oceans.

Consequently, recent climate forcing considerations are based on calculations originating from about 1% change of the energy balance. This is a very little number compared either to measuring accuracy of most environmental physical variables, or to relative errors of model calculations. Hence, it is worth studying various other external factors and feedback mechanisms, which, in addition to greenhouse gases and aerosols, might influence the climate.

One set of the possible feedback mechanisms is connected to the vegetation. In global average, plant structure changes caused –0.2 W m^{-2} radiative forcing (*IPCC*, 2001). Role of plant cover has already been mentioned in the literature since the 1970's, when two possible ways of surface modification, namely overgrazing in the subtropics and devastation of rain forests were counted to be potential causes of the global climate change (*Charney et al.*, 1977; *Sagan et al*, 1979). Both changes express their effect on climate through the light reflecting capacity of the surface, i.e., its albedo.

The Sahel climate problem is investigated by some models (*Xue* and *Shukla*, 1993), and the possibility of vegetation feedbacks on paleo-climatic events are also considered (*de Noblet et al.*, 1996; *Texier et al.*, 1997). Physical sub-model of mosaic vegetation is incorporated in the GCMs already for about a decade (*Claussen*, 1994).

Concerning the problem in the given area, Hungary, three possible feedback mechanisms, connected to surface albedo modifications, were quantified by *Mika et al.* (1992), in relation to likely scenarios (*Mika*, 1988) pointing at warmer and drier climates parallel to global warming. These are the change in duration of the vegetation period, the less precipitation, and the adequate alterations in the managed vegetation, all induce an increase of the surface albedo. The sum of these feedbacks was assessed to be –0.7 W m^{-2}, presuming changes in vegetation cover due to regional consequences of the 0.5 K global warming in Hungary. This value is comparable to the radiative forcing of 100 ppm increase of the CO_2 concentration (*Mika et al.*, 1991).

The present study is aimed to analyze the effects of documented changes in the managed vegetation on the surface albedo and radiation balance of the

surface-atmosphere system above a selected region, delimited by hydrological considerations. The paper is structured as follows: Section 2 describes the methods of computation for surface albedo and shortwave radiation balance of the surface-atmosphere system, perturbed by the changes of vegetation. Section 3 introduces the investigated region in Eastern Hungary and the documented changes of vegetation during the 1951–2000 period. Section 4 presents the results of computed changes in the above two solar radiation components, which are also compared with the changes hypothetically caused by a few other factors. Validity and limitations of these computations are discussed in Section 5.

2. Albedo estimation methodology

2.1 Surface albedo

Surface albedo is influenced by the type and state of soils, species of plant cover, and its growing phase. In the course of preparing the albedo maps of Hungary, *Dávid* (1985) issued synthesis values of surface albedo. On the basis of temporal difference of growing phases, territory of Hungary was divided into two or three plant-specific regions, according to the temporal shifts in growing phases between the northern and southern parts of the country. *Dávid* established surface albedo values for these regions and groups of plants in ten days' resolution. Average monthly surface albedo of plants, most frequently cultivated in Hungary, are listed in *Table 1*, where Region 1 indicates southern part of Eastern Hungary, while Region 2 relates to its northern part. The two regions differ in their average climatic characteristics.

2.2 System albedo and radiation balance

Variations of albedo and radiation balance can be determined not only for the surface, but for the surface-atmosphere system, as well. This makes the comparison between the energy changes due to surface modification and the primary effects of other climate forcing factors possible. These forcing factors are, e.g., the atmospheric CO_2 concentration, solar constant, or volcanic aerosols, which are usually estimated at the top of the atmosphere.

For this aim, results of a former calculation (*Mika et al.*, 1993), made by the help of a radiative-convective model (*Práger* and *Kovács*, 1988), adapted after *Karol* and *Frolkis* (1984) was used, by freezing its convective adjustment and other feedback mechanisms. The model is horizontally averaged, with 16 levels of computation from 1000 hPa at the surface to 0.64 hPa at about 60 km altitude. The vertical resolution is 100 hPa in the troposphere.

Table 1. Average surface albedo of some plant species (source: *Dávid*, 1985). Region 1 and 2 correspond to the southern and northern parts of Eastern Hungary, respectively. (Monthly averages are shown for demonstration. The computations use the original 10-day specifications.)

Region	Surface albedo (%)						
	April	May	June	July	August	September	October
				wheat			
1.	18	20	23	23	21		
2.	18	20	21	23	21		
				barley			
1.	17	20	21	23	21		
2.	17	19	21	23	21		
				rye			
1.	18	20	21	23	20		
2.	18	19	21	23	20		
				maize			
1.		15	18	23	23	24	25
2.		15	17	20	23	23	25
				alfalfa			
1. & 2.	23	20	23	22	23	23	19
				potato			
1.	15	19	24	20	18	19	
2.	15	18	22	23	19	19	
				sugar beet			
1.	14	15	19	19	21	22	22
2.		14	18	19	20	22	22
				meadow and pasture			
1. & 2.	17	19	20	20	19	20	19
				forests (in leaf and conifers)			
1. & 2.	14	14	14	14	14	14	14

Broadband approximation, based on empirical transmission functions, is applied for 24 and 17 spectral intervals in the short- and longwave parts of the radiation spectrum, respectively. Longwave transmission functions are adapted from *Rozanov et al.* (1981). The δ-Eddington method is used for parallel computation of absorption and scattering. Optical thickness is calculated by the Curtis-Godson approximation. (For both latter approaches see, e.g., *Liou*, 1980).

Internal parameters and astronomical conditions of the model are defined for Budapest (47°26'N; 19°17'E). Cloud amounts of the different levels are taken from *Warren et al.* (1985), and a proportional vertical normalization is performed to obtain the total cloud coverage valid in the local climate. Low- and medium-level clouds are considered as blackbodies for longwave radiation. High-level cloudiness is characterized by 0.5 emissivity. The aerosol optical profile is adapted from *WMO* (1983) considering "continental background" aerosol.

By this model we determined, how surface albedo changes affect the shortwave energy balance, R_s, of the surface-atmosphere system. Connection between this term and the system albedo, α_s, is

$$\Delta R_s = -G_0 \cdot \Delta \alpha_s , \tag{1}$$

where G_0 is the solar energy reaching the top of the atmosphere.

According to computations with the above radiative model (*Mika et al.*, 1992), the dependence of the system albedo on the surface albedo, α, is nearly linear (*Fig. 1*)

$$\Delta \alpha_s = k \cdot \Delta \alpha . \tag{2}$$

One percent change of surface albedo involves k = 0.40–0.45 percent change in system albedo during the examined seven months. Deviation of k from 1.0 can be explained by the cloudiness and limited transparency of the atmosphere. The higher values of k characterize the summer period, when cloudiness is less and the optical path is shorter.

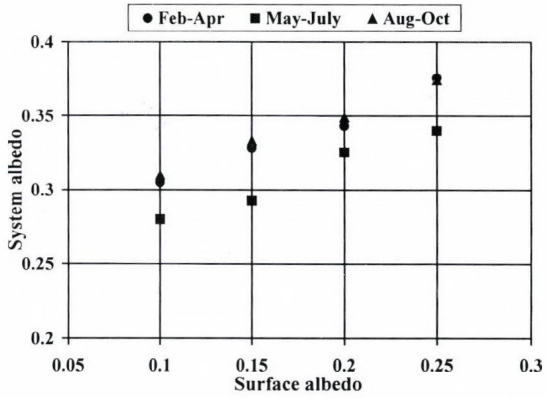


Fig. 1. Correlation between the surface albedo and the albedo of the surface-atmosphere system in different periods of the year, as computed by a radiation model for Hungary.

For the three "astronomical" seasons (February–April, May–July, and August–October), the radiative calculations yielded slightly different k values (k=0.406, 0.446, 0.400, respectively), from which the annual course can be approached by the following formula

$$k(h) = a_1 \cdot h^2 + a_2 \cdot h + a_3, \tag{3}$$

where h is the serial number of the months (February = 2, etc.). Substituting the three-monthly averages of these numbers into Eq. (3), the $a_1,\ldots,a_3$ coefficients and the monthly variation of $k(h)$ were calculated.

3. *Regional specification*

3.1 The selected region

The region, selected for the investigation, is the sub-catchment area of the river Tisza (*Fig. 2*) in the Hungarian plain, partly bordered by the administrative border of the country. This region of 35,700 km^2 was previously used for regional energy- and water balance modeling (*Mika et al.*, 1991, 1998).

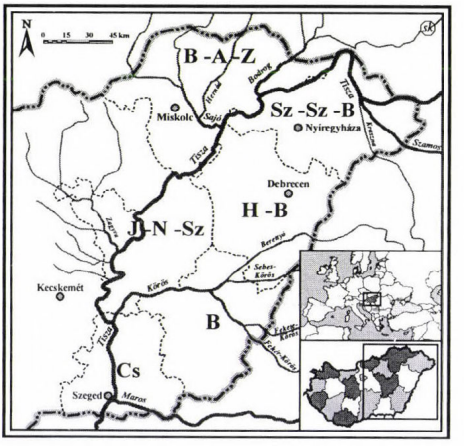


Fig. 2. The river Tisza sub-catchment in Eastern Hungary. The indicated administrative counties approximately cover this hydrological unit. Total area of the six counties is 34,900 km^2.

However, the land use data are officially published on a county by county basis (see in Section 3.2), hence this region should be approximated by administrative counties. Six of them, namely Borsod-Abaúj-Zemplén, Szabolcs-Szatmár-Bereg, Hajdú-Bihar, Jász-Nagykun-Szolnok, Békés, and Csongrád counties approximately cover the targeted hydrological region. More exactly, this 34,900 km^2 administrative area is the object of the investigations.

The Hungarian catchment area of the river Tisza exhibits the lowest altitude of about 100 m above the sea level, in the Carpathian Basin. This large

landscape has always been characterized by high proportionality of managed vegetation. Recently, 74% of the total administrative area is cultivated. The rest of the area, mainly the natural vegetation of the Hungarian Plain, represents the westernmost extension of this forest-steppe zone in Europe.

Soil formation factors here are favorable to the development of meadow alluvial and alluvial meadow soils along the rivers, solonetz in the center of the region and chernozem mostly in the southern part of the landscape. The latter mentioned chernozem is the most fertile kind of soils in Hungary.

As compared to the annual mean temperature (9.3–10.6 °C), the annual mean precipitation amount (500–600 mm per year) is far from the optimum.

3.2 Land use series

In this chapter it is briefly shown, how the sown area of the different plant species varied in the six examined counties during the period 1951–2000, according to the data in the annual reference books of the *Central Statistical Office* (1951–2000) and *Historical Statistical Contributions* (1971–79).

For easier interpretation, the plant species of the computations can be arranged into five groups:

- *cereals* – winter wheat, rye, barley, and rice;
- *fodder-plants* – maize, alfalfa, red clover, maize for silage, oat, and cattle-turnip;
- *food- and industrial plants* – sugar beet, tobacco, sunflower, potato, and fibre hemp;
- *vineyard and fruits*;
- *others*: forest, meadow, and pasture.

Total area of agricultural land use in the examined region shows a gradually decreasing tendency (*Fig. 3*). Considering the examined 1951–2000 period, the sown area of cereals and fodder-plants has decreased considerably, especially since the early 1990's. Tendencies of the sown area for food- and industrial plants varied from one county to the other, with no clear tendency in the whole region. Share of the sown area for vineyard and fruits is an order of magnitude smaller, than that of the other four groups of plants. Share of forest, meadow, and pasture in the six counties did slightly increase in 1951–2000, contrary to the other groups of plant species. These tendencies are also demonstrated in *Table 2* by selected five-year averages of the five groups of plant species.

Total share of the considered agricultural areas is the least in Hajdú-Bihar county (0.71) and the highest is in Csongrád and Jász-Nagykun-Szolnok (0.80). Altogether, 74% of the total area are involved in the investigation. For the rest of the area no changes are postulated.

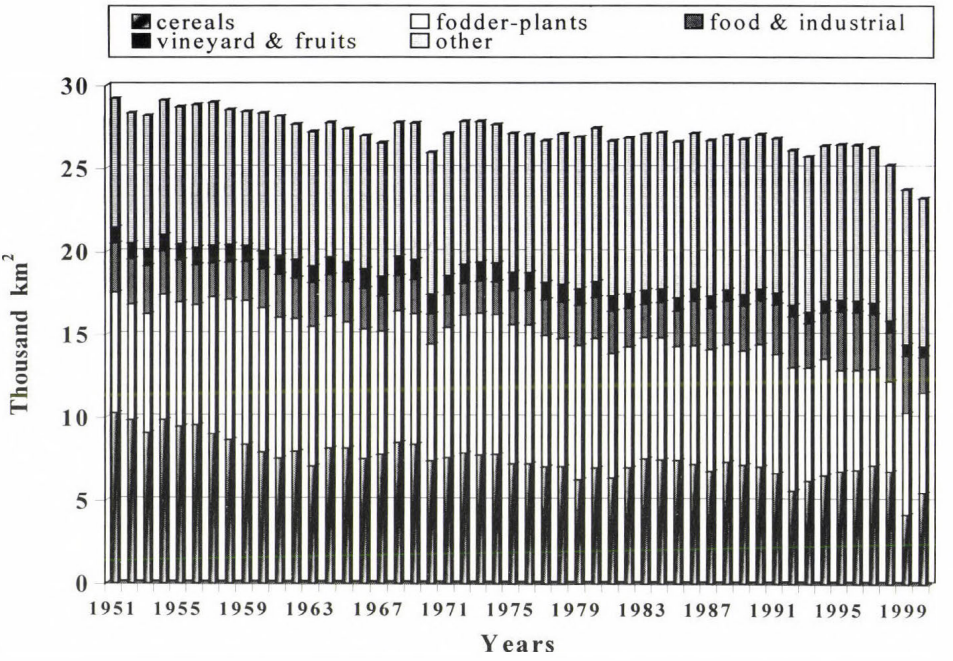


Fig. 3. The total cultivated area and its share among the main plant groups in the examined region, 1951–2000.

In computation of areal average albedo, however, variations of the total sown area are not considered, since the area-weighted sums will be normalized by this area, i.e., by the sum of the weights. Possible variations at the set-aside areas are not included in our estimations.

Table 2. Land use change tendencies in Eastern Hungary expressed in selected time periods (in thousands of km²)

Time period	Cereals	Fodder plants	Food & industrial	Vineyard & fruits	Forest, meadow, pasture	All land in use	Total area	Part of land in use (%)
1951–1954	9.76	7.25	2.80	9.76	7.90	28.68	34.90	**82.2**
1973–1978	7.35	8.22	2.09	10.97	8.44	27.20	34.90	**77.9**
1995–2000	6.22	5.90	3.16	6.81	9.22	25.18	34.90	**72.2**

3.3 Area weighting

The effect of any fluctuation or change in land use on some A quantity (in present study the surface albedo) is calculated in the following way. Suppose, that in a given t year each considered plant sort takes up $T^i(t)$ territory in the examined region. The sum of these kind of territories is in this case is

$$T_m(t) = \sum_i T^i(t), \quad i = 1, 2, \ldots . \tag{4}$$

Introduction of the above mentioned land use is actually connected to this total $T_m(t)$ territory and, within this, to $T^i(t)$ share territories. Albedo selection and weighting is performed by areas of the original plant species, without any grouping.

Regional variation of the areal mean, $A_m(t)$, quantity is determined by the plant-specific $A_i(t)$ values, also depending on the vegetation phase, and by the $T^i(t)$ area of the different plant species (land use forms), as

$$A_m(t) = \frac{1}{T_m(t)} \sum_i T^i(t) \cdot A_i(t) \ , \quad i = 1, 2, \ldots . \tag{5}$$

It is worth mentioning, that the decreasing tendency of the total agricultural area has no effect on the regional average of A, due to normalization on the right side. These area mean values of A (specifically the albedo) will be the base of our further calculations.

4. Results

4.1 Surface albedo tendencies

Time series of surface albedo, determined by Eq. (5), represent the result of changes in land use (*Fig. 4*). The surface albedo averaged for the six counties show clear decreasing tendency in the months from April to July (*Fig. 4a*). In other words, the share of those plant species increased, the surface of which is relatively darker in this part of the year (due to, e.g., the larger green mass or more complete cover of the ground). The situation in August is still the same (*Fig. 4b*), whereas no clear change can be established in the two following months. This is in connection with the fact that the majority of the plant species, still present in September–October, can develop longer, so their albedo values become very similar.

As assumed from the total decrease of albedo, regional averages of the surface radiation balance increased in the examined 50 years. Linear trend of

this change is +0.017 W m^{-2} yr^{-1}, which means +0.85 W m^{-2} total change during the examined 1951–2000 period.

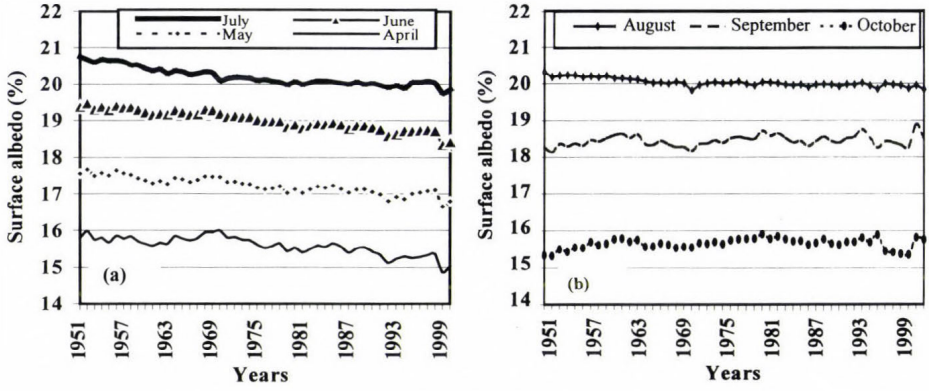


Fig. 4. Effect of land use variations on the surface albedo cumulated for the examined six counties. (a) April–July, (b) August–October.

4.2 Changes in the system albedo

If monthly values of the system albedo are multiplied by astronomically possible solar irradiance, then the amount of shortwave radiation reflected by the surface-atmosphere system to the outer space is received. Changes in this amount can further be compared with other changes in the radiation balance of the system.

The energy surplus, caused by the decreasing radiation energy reflected to the outer space by the surface-atmosphere system, which remains in the system to increase air temperature (longwave radiation), is +0.010 W m^{-2} yr^{-1} (*Fig. 5*), or 0.50 W m^{-2} in total during the examined 50 years. The linear trend fits fairly tightly to the data set, since value of the correlation coefficient is 0.981. (In *Fig. 5*, the changes are demonstrated in comparison to the arbitrarily chosen 1951–1980 period. The point of this operation is not the definite reference period, but the long term basis, instead of any single year with its land use peculiarities.)

Curves of *Fig. 5* also demonstrate considerable county-by-county differences in the slope of the tendency. These changes do not exhibit a clear geographical arrangement. Trends of neighboring counties are quite different in some cases. (The breaks in 1970, recognized in several series, were caused by administrative changes of borders between particular counties. These re-arrangements could not influence the total area or the average albedo of the region.)

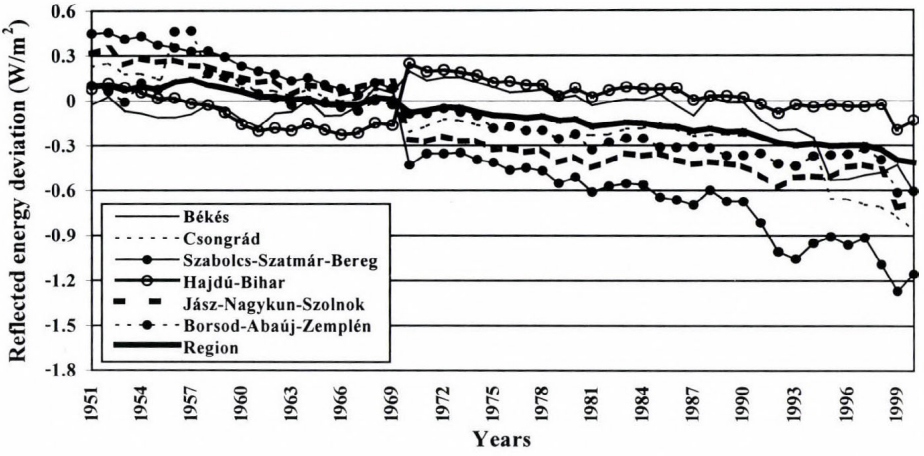


Fig. 5. Effect of land use variations on the reflected energy at the top of the atmosphere for the six counties and the region in April–October, relative to its 1951–1980 averages.

4.3 Comparison to other factors

In order to demonstrate the importance of this relatively small change of the radiation balance, +0.50 W m^{-2}, selected parallel (independent) effects of further forcing factors were computed by the same radiative-convective model (*Table 3*).

Table 3. Estimated direct effect of the documented land-use changes on radiation balance of the surface-atmosphere system compared to selected forcing or feedback mechanisms over Eastern Hungary in the summer half-year.

Forcing or feedback (summer half-year)	**Change (W m^{-2})**
Land use changes 1951–2000	+ 0.50
CO_2 concentration: 330→370 ppm	+ 0.71
Solar irradiance variation: 0.1%	± 0.24
Strong volcanic cloud: $\Delta\tau_{0.55} = 0.1$	- 0.42
Feedback of 0.2 K warming on longwave radiation balance	- 0.54

The effect of CO_2 concentration changes from 330 to 370 ppm (representing the same 1951–2000 period, taken from the corresponding global mean concentrations), is just slightly stronger, +0.71 W m^{-2}.

Result of a hypothetical 0.1% solar radiation variation, which is comparable to the observed fluctuations (*IPCC,* 2001, Chapter 6), is only ± 0.24 W m^{-2}. A strong stratospheric sulfate cloud after a volcanic eruption, characterized by $\Delta\tau = 0.1$ at the 0.55 μm wavelength, can lead to 0.42 W m^{-2} decrease of the radiation balance in Hungary.

On the other hand, this change caused by the land use tendencies is not yet dramatical, since an identical value, -0.54 W m^{-2}, is caused in the longwave part of the spectrum if the near-surface air temperature changes as small as 0.2 K (due to any reason yielding smooth vertical temperature distribution).

5. Discussion

The above mentioned calculations have determined the effect of changes in land use on the albedo and radiation balance. These computations mixed fairly exact *in situ* data on land use with generalized average physical parameters on albedo in combination with a radiation transfer model.

Both components of the calculations request some discussion. The data on land use are based on settlement-level documentation of the corresponding area. Hungary has about 3000 such settlements, so the number of initial data sources used in the area of the present study, i.e., 40% of the country, is well above one thousand. Possible random errors of documentation, therefore, had fairly good chances to be equalized in area mean. For non-random errors one may establish, that except the early periods of the centrally and ideologically planned economy in Hungary, no special interest should be expected on misleading registration of this factor of our computations. Moreover, the basic tendencies of the land use changes, demonstrated in *Fig. 3*, are more general in time.

Another problem with this component might be that the not examined 18–28% of the area (derived from *Table 2*) did not remain constant, either. Area of settlements, transport routes, and water bodies, as well as non-managed (not inventoried) spots of vegetation could also change during the five decades of the analysis. Having no definite data on these processes, however, we can not consider the effects of these changes at present.

As concerns the look-up tables of the applied plants (*Table 1*), two types of error may occur. The first is a possible over- or underestimation of the albedo in average. The other is the possible deviation from this average in the individual years. Both are connected with the way of albedo derivation. The nominal albedo values are based on various projects of direct observations and literature synthesis. They arrange the albedo values according to the

phenological phases. The second step is to identify the average date of the phases within the vegetation period.

The effect of the first source of error is likely not strong. Even if the original 10-day albedo values bear 1–2% absolute error to either direction, they likely decrease during the monthly and further averaging along the vegetation period. On the other hand, the area mean albedo considers almost twenty different types of plant. Likelihood of strong errors, i.e., of the same order of magnitude as the computed effects, is small.

As concerns the errors of albedo estimation in the individual years, they could be larger, especially if the shifts of the pheno-phases, and also the humidity stresses, occurred non-randomly during the fifty years. But, this effect is rather a climate change forced effect, than a land use forced one, since the primary effect of these changes on the albedo of all plants should be more important, than the variations of this primary effect caused by the modified structure of plants coverage within that.

On the above pages, the effects of land use were compared to the components of the radiation balance. According to these comparisons, it could be established that for the region of Eastern Hungary, changes due to land use had the same magnitude as the changes due to the radiation balance occurred in relation with the global climate change or realistic externally forced fluctuations.

It is also worth mentioning, that the continuous decrease of the surface albedo is not only the result of the previous centrally planned economy. The tendencies continued with about the same speed after 1990, too.

However, the question, whether or not changes in land use could produce changes in the regional climate (temperature, precipitation, etc.) with similar magnitude, as the effects of the compared global-scale radiative forcing factors (see in *Table 3*), could not be answered. To answer this question, temporal change of land use should be known for a much larger region, since the climate of Eastern Hungary is influenced by advection of heat and atmospheric water from distant regions. This fact points at shortcomings of the applied radiation-only modeling. This fact, parallel to the obtained significant primary effects on the radiation balance changes, point at the need for sophisticated regional modeling of the land use. Such a regional model should consider all relevant physical processes and a much larger area, even if one focuses on the effects of land use changes just in a part of the whole model domain.

Acknowledgements—The authors are truly indebted to *L. Lukácsovics* for his valuable contribution in series of computations, *Z. Sümeghy* for preparing Fig. 2, *A. Horváth* for her contribution in the data collection, and *I. Kevei-Bárány* for professional advices on integrated classification of the wide variety of vegetation.

References

Central Statistical Office (1951-1993): *Statistical Yearbooks*. Central Statistical Office. County Centres of Hungary

Charney, J.G., Quirk, W.J., Chow, S.H., and *Kornfield, J.*, 1977: A comparative study of the effects of albedo change on drought in semi-arid regions. *J. Atmos. Sci. 34,* 1366-1385.

Claussen, M., 1994: On coupling global biome models with climate models. *Climate Res. 4,* 203-221.

Dávid, A., 1985: Preliminary investigations estimating the regional distribution of surface albedo (in Hungarian). *Beszámolók az 1983-ban végzett tudományos kutatásokról.* Orsz. Meteorológiai Szolgálat, Budapest, 81-98.

de Noblet, N., Prentice, I.C., Joussaume, S., Texier, D., Botta, A., and *Haxeltine, A.*, 1996: Possible role of atmosphere-biosphere interactions in triggering the last glaciation. *Geophys. Res. Lett. 23,* 3191-3194.

Historical Statistical Contributions, (1971-79): *Data on Plant Management* (*Vol. I-IV.:* 1922-33, 34-45, 47-62, 63-70) (in Hungarian). Statisztikai Kiadó, Budapest.

IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental panel on Climate Change (Houghton J.T., et al., eds.), Cambridge University. Press, Cambridge, UK. And New York, N.Y. USA, 881 p. http//:www.ipcc.ch.

Karol, I.L. and *Frolkis, A.A.*, 1984: Energy-balance type radiative-convective model of the global climate (in Russian). *Meteorologia i Gidrologia*, No. 8, 59-68.

Liou, K.N., 1980: *An Introduction to Atmospheric Radiation.* Academic Press.

Mika, J., Horváth, Sz., Fogarasi, J., and *Makra, L.*, 1998: Simulation of climate forcing mechanisms on the energy and water balance of a watershed. In *Proc.19th Conf. Danube Countries*, Osijek, Croatia, 15-19, June 1998, 331-342.

Mika, J., Kovács, E., and *Bozó, L.*, 1993: Effects of anthropogenic and natural changes in atmospheric aerosol content on radiation balance in the Carpathian Basin (in Hungarian). *Beszámolók az 1989-ben végzett tudományos kutatásokról.* Orsz. Meteorológiai Szolgálat, Budapest, 81-89.

Mika J., Kovács E., and *Dávid, A.*, 1992: Effect of vegetation and soil conditions on albedo of the surface-atmosphere system (in Hungarian). *Beszámolók az 1988-ben végzett tudományos kutatásokról.* Orsz. Meteorológiai Szolgálat, Budapest, 165-173.

Mika, J., Kovács, E., Németh, P., and *Rimóczi-Paál,* A., 1991: Parameterisation for regional energy balance climate modeling over Hungary. *Adv. Space Res. 11,* (3)101-(3)104.

Mika, J., 1988, Regional features of a global warming in the Carpathian Basin (in Hungarian). *Időjárás 92,* 178-189.

Práger, T. and *Kovács, E.*, 1988: Investigation of effects of atmospheric trace gases and aerosol on climate with a radiative-convective model (in Hungarian). *Időjárás 92,* 153-162.

Rozanov, E.V., Timofeyev, J.M., and *Frolkis, V.A.*, 1981: Effect of trace gases on some characteristics of infrared radiation regime (in Russian). *Izvestii Fiziki Atmosferi i Okeana 4,* 384-391.

Sagan, C., Toon, O.B., and *Pollack, J.B.*, 1979: Anthropogenic albedo changes and the earth's climate. *Science 206,* 1363-68.

Texier, D., de Noblet, N., Harrison, P.S., Haxeltine, A., Jolly, D., Joussaume, S., Laarif, F, Prentice, I.C., and *Tarasov, P.*, 1997: Quantifying the role of biosphere-atmosphere feedbacks in climate change: coupled model simulations for 6000 years BP and comparison with paleodata for northern Eurasia and Africa. *Clim. Dynam. 13,* 865-882.

Warren, S.G., Hahn, C.J., and *London, J.*, 1985: Simultaneous Occurrence of Different Cloud Types. *J. Clim. Appl. Meeorol. 24,* 658-667.

WMO, 1983: *Report of the Experts Meeting on Aerosols and Their Climatic Effects.* Williamsburg, Virginia 28-30 March 1983, World Climate Programme, WCP-55.

Xue, Y. and *Shukla, J.*, 1993: The influence of land surface properties on Sahel climate. Part I., Desertification. *J. Climate 6,* 2232-2244.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 1, January–March 2006, pp. 63–87

Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation

Péter Domonkos

H-1155 Budapest, Vasvári P. u. 5, Hungary; E-mail: dpeterfree@freemail.hu

(Manuscript received in final form January 12, 2006)

Abstract—Objective homogenization methods (OHOMs) are applied on time series of observed meteorological data mostly from Hungarian sites. The used database comprises temperature and precipitation series at 20 stations from Hungary and 4 stations from other countries in Central Europe. Different ways of data treatment multiply the number of the available series, whereas lack of data limits them. Altogether, 215 series of monthly or annual mean temperatures and 112 series of monthly or annual precipitation totals are analyzed. All of the series are 98–100 years long. The time step between the adjacent members of each series is 1 year. Statistical characteristics of the detected change points, such as mean number of change points per time series, mean magnitude of shifts, etc., are calculated. The aims of the investigation are: (i) to get a general insight into the detectable inhomogeneities in meteorological time series from Hungary, and (ii) to gather the necessary knowledge for the development of testing the efficiency of OHOMs. The study also explains the concept of OHOM, presents its main kinds, and provides some arguments about conditions, advantages, and limitations related to their practical application. Twelve OHOMs are applied in the study, their results have several common features. The OHOMs are taken from the literature with slight modifications, and they are always applied on relative time series, i.e., on the differences of candidate series and reference series. Identical process of reference series derivation, constructed from commonly used elements, is used for all included OHOMs. The main results of the study are: (i) More than half of the examined time series are inhomogeneous. (ii) Most of the detected shift magnitudes are hardly larger than the standard deviation of the noise term. (iii) Serial correlation is indicative of homogeneity quality. (iv) Homogeneity quality of winter temperatures is much better than that of summer temperatures. (v) Distribution of the detected shift magnitudes is skewed positively for each of the OHOMs and each examined meteorological variable.

Key-words: observed time series, homogenization methods, inhomogeneities, temperature, precipitation, Hungary

1. Introduction

The analysis of observed meteorological time series is the main source of our knowledge about climate change and climate variability of the latest centuries. However, reliable consequences about climate variability can be drawn only from databases of high quality. One way of controlling and correcting of observed time series is the use of homogenization methods. These methods check the homogeneous origin of the data, and estimate the necessary corrections for achieving a higher level of homogeneity via comparison of the observed data series in the given region. Thus, the application of a homogenization method supposes a database including several time series of the same meteorological element, these series must have a common time period, and they are supposed to have fairly high spatial correlations (*Peterson et al.*, 1998).

The largest group of the homogenization methods is the group of objective homogenization methods (OHOM). More than 10 kinds of them have been applied in the referred climatological literature so far. During the application of an OHOM on a time series, each decision about acceptance, correction, or rejection of data relies on the statistical characteristics of the time series in the database. Naturally, an application of an OHOM may be supplied with subjectively tailored steps of decision (e.g., with the consideration of contemporary reports, so-called metadata, about the technical or environmental changes in the observations), but all of the OHOMs can be applied without any subjective decisions. Recently, the statistical analysis of large-scale, or even global data sets has become a common tool of detecting climate change and characterizing climate variability. The only way of the quality check in such large data sets is the use of one or more OHOMs. An OHOM recommended for climatological applications must be reproducible, easy-to-use, fast, and suitable for automatic use. Fortunately, most of the OHOMs introduced into the climatology possess these qualities. Thus, the only problem is that which OHOM should be chosen to achieve the highest quality possible of observed data sets? This question will not be answered by the present study, but a large selection of OHOMs is applied together. Results, possibly important for a later development of efficiency testing for OHOMs, will also be discussed.

In this paper we analyze several hundreds of observed time series with the help of twelve OHOMs commonly used in climatology. All of the time series contain monthly or annual mean temperatures or precipitation totals. The length of the series is 98–100 years, and the vast majority of them are originated from Hungarian sites. The organization of the paper is as follows. Section 2 comprises a brief description about the aim, concepts, and tools of

homogenization methods. It is followed by two sections discussing the used database and OHOMs (Sections 3 and 4). Section 5 defines some further methodological details. In Section 6 the calculated statistical characteristics of detected inhomogeneities are presented. The last section comprises discussion and conclusions (Section 7).

2. Theoretical basis

Here we make an attempt to configure the problem of data homogeneity in its most general frame. Some of the formulae and theoretical considerations will be used in the following comparison of OHOMs, too.

A time series from meteorological observations ($\mathbf{X}=[x_1,x_2,...,x_n]^T$) is often expressed as a composition of (a) climatic mean, (b) climatic changes and fluctuations, and (c) errors. Eq. (1) contains more than three components, allowing that both climatic processes and errors may result in long-term changes and short fluctuations in data series. Even an observed climatic mean can be considered as the sum of real climatic mean and the mean of errors. Thus,

$$x_i = \bar{u} + \bar{v} + u'_i + v'_i + u''_i + v''_i \qquad (i = 1,2,...,n), \tag{1}$$

where u and v denote the climatic and local origin, respectively, over line denotes time-average, letters with single comma (u' and v') represent the components of long-term fluctuations, and letters with double commas represent short-term fluctuations approachable usually with white noise. The usual time step between two elements of a time series is 1 year, and it is always 1 year in this study.

All of the components of local origin are error terms, at least from the scope of macroclimatic investigations, but their roles are different during the application of an OHOM. Values of v'' cannot be assessed by statistical tools, since they cannot be distinguished from the short-term climatic fluctuations, unless v''_i is outstandingly high, resulting in an unusual large anomaly of x_i relative to $\bar{x}$. The outstandingly large v'' elements of time series are commonly referred as outliers, and the procedure by which they are recognized and eliminated is the outlier correction. The $\bar{v}$ components can be assessed, but their values are not time dependent. They characterize the spatial representation of individual observing sites and do not affect the homogeneity of the series.

So, the aim of OHOM applications is to identify the $\mathbf{V}'$ vector of time series under examination, and to eliminate its elements from the series. A $\mathbf{V}'$

vector can be considered as the unified term of inhomogeneities of local origin (inhomogeneities, hereafter). If the elements of **V′** are clearly higher than the noise level, and $|v_i| >> |u_i|$ is generally true, the inhomogeneity characteristics can be assessed with fair confidence. By contrast, there is no way of the reliable assessment of inhomogeneities with magnitudes that are lower than the noise level.

If an OHOM is applied on an observed series, the detected inhomogeneities will contain all the long-term changes of the series that are larger than the noise level, either caused by macroclimatic variability or local effects. However, the OHOMs aim to correct the error terms only, and the misidentification or alteration of real macroclimatic terms would be a serious mistake. To distinguish the local effects from the macroclimatic ones, the observed series in question, often referred as candidate series, are usually compared to reference series (**F**) of high quality (considered as homogeneous). Supposing that the macroclimatic effects are the same for the candidate series and the appropriate reference series, their difference, the so-called relative time series (**T**), contains only local effects, already:

$$\mathbf{T} = \mathbf{X} - \mathbf{F}. \tag{2}$$

Although usually no series of an observing network can be considered as absolutely homogeneous, a proper combination of several series from the geographical surroundings of the candidate series is mostly appropriate for the function of reference series. We note that the rate of the **X** and **F** series is also applied as relative time series, for variables whose climatic differences can be characterized better by ratios than by arithmetic differences. However, to keep the theoretical basis in brief, Eq. (2) may also be maintained for the variables of latter type, applying scale transformation on the components of the reference series (see Section 5.1).

As the climatic components cannot be perfectly eliminated by the creation of relative time series (because the **F** series is not perfectly homogeneous, and because of the differences of geographical conditions between the candidate and reference series), **T** contains the same type components as **X** in Eq. (1). Notwithstanding, the relative importance of the individual components is changed. Using a proper reference series, the absolute values of u' elements are practically never higher than the noise level. Thus, unifying the mean terms to $\bar{t}$ and the noise terms to e, Eq. (1) can be rewritten for relative time series in the form of

$$t_i = \bar{t} + v'_i + e_i. \tag{3}$$

If a time series is homogeneous, then every elements of $\mathbf{V}'$ are equal to 0. As the reference series are supposed to be homogeneous, the inhomogeneities of a relative time series are attributed to the candidate series. In this study, OHOMs are applied always on relative time series.

The most common type of inhomogeneities is the so-called change point. A station movement, an abrupt methodological change in the observation timing or instrumentation, or some changes in the neighborhood of the station usually cause sudden shift in the average values of the observed data. If $\left|v'_j - v'_{j-1}\right| >> 0$ $(1<j\leq n)$, it indicates a change point at year j. While the appearance of an existing change point may be obscured by the noise around year j, on the other hand, unusual structure of noise may have an appearance similar to change points. Therefore, the success of change point identification has a stochastic character, and it strongly depends on the rate of abrupt change to the noise level (*Ducré-Robitaille et al.*, 2003). Another typical form of inhomogeneities is a long-lasting, trend-like change in the values of the time series. They are caused by the gradual changes of some instrument characteristics or, more frequently, by the gradual changes in the site environment, e.g., in connection with the growing urbanization. The appearance of this type of inhomogeneities in the terms of Eq. (3) is a gradual, monotonous change in the v' components.

Further explanations of concepts of homogenization are delivered by *Alexandersson* and *Moberg* (1997), *Peterson et al.* (1998), *Sneyers* (1999), and others.

3. Database

Six different databases of monthly mean temperatures or monthly precipitation totals were chosen to assess the characteristics of detectable inhomogeneities. The meaning of "six different databases" is as follows: Any pairs of data series belong to the same database if they contain the data of the same meteorological variable (temperature or precipitation), and if they are likely derived by the same preprocessing, as considered from the documentation that accompany the series. None of the databases used here were subjected to an OHOM before our analysis.

Most of the data series are delivered by the Hungarian Meteorological Service, and more than 90% of the series comprise observed data from Hungarian sites. (The few exceptions are temperature series from Prague, Cracow, Zagreb, and Bologna; see *Table 1*.) All of the six databases had been used earlier in climatological works, thus we consider all of them equally relevant. For an easier wording, different initial databases are referred as

"different sources" hereafter, and the word "database" is held for the whole set of data series.

Table 1. Data series of observing sites, included in the individual data bases. (I,...,IV mark the distinction according to previous data treatment; T and P denote temperature and precipitation, respectively.)

Stations	Longitude (E)	Latitude (N)	I		II		III	IV
			T	P	T	P	T	T
Bologna, IT	11°20′	44°30′						+
Prague, CZ	14°25′	50°05′						+
Zagreb, HR	15°59′	45°49′						+
Cracow, PO	19°58′	50°04′						+
Sopron	16°36′	47°41′			+	+	+	
Szombathely	16°36′	47°15′	+	+	+	+	+	
Zalaegerszeg	16°49′	46°55′	+	+				
Keszthely	17°14′	46°44′			+	+		
Mosonmagyaróvár	17°16′	47°53′	+	+	+	+	+	+
Pápa	17°22′	47°18′	+	+				
Győr	17°41′	47°43′		+				
Kaposvár	17°50′	46°22′	+	+				
Iregszemcse	18°11′	46°41′	+	+				
Pécs	18°14′	46°00′	+	+	+	+	+	
Budapest	19°02′	47°31′	+	+	+	+	+	
Baja	19°11′	46°11′			+	+		
Kecskemét	19°46′	46°54′	+	+	+	+	+	
Szeged	20°09′	46°15′	+	+	+	+	+	+
Szarvas	20°33′	46°52′			+	+		
Túrkeve	20°45′	47°06′			+	+		
Miskolc	20°46′	48°07′	+	+			+	
Békéscsaba	21°06′	46°41′		+				
Debrecen	21°37′	47°33′	+	+	+	+	+	+
Nyíregyháza	21°41′	47°59′	+	+	+	+	+	
All elaborated			**13**	**15**	**13**	**13**	**10**	**7**

The data series are complete, their length is 98–100 years, and all of them cover the period 1901–1998. The time step between two adjacent elements of the series is always 1 year. We use the series of values from central months of

traditional seasons (January, April, July, and October), as well as series of annual values. The month (or the whole year), to which a series belongs, is referred as seasonality of the series, throughout the paper. A group of data series, whose source and seasonality are common, is labeled data segment. There is no difference in the length of the series within a segment. Considering that there are 6 initial databases (see *Table 1*) and 5 types of seasonality (central months of the four seasons as well as the entire year), the possible maximum number of elaborated data segments would be 30. However, the final number of segments really used is 28, because two possible segments comprising precipitation in July are omitted, due to too low spatial correlations (see Section 5.1). Each segment consists of 7–15 time series (with an average number of 11.8), thus the whole database comprises 327 time series, i.e., 215 temperature series and 112 precipitation series.

4. Homogenization methods applied

Twelve OHOMs are applied in this study. As general rules, all the OHOMs were used in broadly available climatological studies earlier, their theoretical bases and practical application are published in scientific journals, and they are reproducible from these papers and other available guides or descriptions. Nevertheless, there are some exceptions to these rules, they will be discussed together with a more extensive description of the individual OHOMs applied in this study.

The twelve OHOMs and their abbreviations are as follows:

(i) Buishand-test of the maximum accumulated anomaly [Bs1],
(ii) Buishand-test of the difference between the maximum and minimum of accumulated anomalies [Bs2],
(iii) Wilcoxon Rank Sum statistic [WRS],
(iv) Pettitt-test [Pet],
(v) Mann-Kendall test [M-K],
(vi) Standard Normal Homogeneity Test for shifts only [SNH],
(vii) Standard Normal Homogeneity Test for shifts and trends [SNT],
(viii) Multiple Linear Regression [MLR],
(ix) Bayesian approach [Bay],
(x) method of Easterling and Peterson [Eas],
(xi) a version of the Multiple Analysis of Series for Homogenization [MS1],
(xii) another version of the Multiple Analysis of Series for Homogenization [MSH].

Ten methods from the twelve OHOMs look only for change points, whereas two of them (SNT and MLR) can recognize shifts and trends alike.

Usually, the identification of change point positions in time is the critical point of the effectiveness of an OHOM, although, the assessment of its statistical significance and the measure of the jump (i.e., the shift-magnitude) are also very important.

Now, let us describe the above listed 12 OHOMs. Perhaps the oldest homogenization method is the investigation of the accumulated anomalies (*Kohler*, 1949). A sudden shift in the values of the examined series tends to cause a local maximum or minimum in the accumulated anomalies. Bs1 and Bs2 (*Buishand*, 1982) are OHOMs relying on the analysis of accumulated anomalies.

The frequencies of positive and negative relations between sample elements chosen from the different sides of a change point are biased from equality (50–50%), and the measure of the bias has a local maximum around the change point. This is the basic idea of OHOMs calculating frequencies of relations between sample elements or constructing statistics for rank-ordered sample. WRS (*Karl* and *Williams*, 1987), Pet (*Pettitt*, 1979), and M-K belong to this group. We applied the M-K test relying on the description by *Aesawy* and *Hasanean* (1998), but with the supplement of serial correlation analysis. The calculation of serial correlation is a traditional way of indicating homogeneity or inhomogeneity of a time series. About serial correlation analysis see, e.g., *Sneyers* (1997, 1999). The supplement to the M-K test is needed, because otherwise this test tends to find too many change points, even in a pure white noise process.

The Standard Normal Homogeneity Test (often referred as SNHT) is the OHOM which has most frequently been applied by climatologists in the recent years. It relies on the fact that differences in the means of any two adjacent sub-periods, covering together the whole series, have a local maximum at a change point. Two versions of this approach are applied. While the SNH version (*Alexandersson*, 1986) assesses and corrects only the change point type inhomogeneities, SNT is able to recognize gradual changes, too (*Alexandersson* and *Moberg*, 1997). Gradual changes recognized by SNT are modeled with linear changes in certain sections of the series.

Another group of OHOMs applies fitting of predefined function types, such as constant sections with shifts between them, linear changes, etc. The optimal function is the one with minimum root mean square error (RMSE). MLR (*Vincent*, 1998) represents this type. It is capable of recognizing and correcting both abrupt shifts (change points) and gradual changes. The original Bayesian approach (*Ducré-Robitaille et al.*, 2003) also belongs to this group. However, in this study a part of that method is applied, namely the derivation of the most probable change point position. In the Bay method applied here, the Bayesian derivation of change point position is supplied with serial

correlation analysis for controlling significance, and the way of the assessment of shift magnitudes is an unchanged adaptation from the SNH test.

The method Eas developed by *Easterling* and *Peterson* (1995) is a multi-step procedure. It starts with the fitting of one or two straight lines with the minimum RMSE. If the fitting of two lines (with a sudden shift and/or a break in the trend line) is significantly better than the fitting of one line, a potential change point is flagged to the position of the break, and the series is cut into two sub-periods at this point. This step is repeated for the sub-periods as long as one line has a better fitting than any two lines. Thereafter, the significance of the flagged inhomogeneities is controlled one-by-one, applying rank order statistic in windows with maximum 12 years half-windows towards either side from the flagged inhomogeneity.

The Multiple Analysis of Series for Homogenization (often referred as MASH) is a development of a Hungarian mathematician. The full description of the method is available in a WMO conference issue (*Szentimrey*, 1999), with abbreviated information in the review paper by *Peterson et al.* (1998). MASH is a rather complex method, its application needs much more computing time, than that of the other OHOMs. The MASH method examines all the possible change point configurations in a time series using dynamic window width. Examining a sub-period with a given window, the flagging of potential change points is based on special statistical criterion, similar to the one in SNH. The optimal set of change points is determined by a multi-step iteration procedure. MSH is the developer's original software. MS1 contains some modifications. The main difference between the two versions is that the minimum width of the half-windows is 5 years in the MS1 method (1 year in the original version).

Moberg and *Alexandersson* (1997) give instructions how to identify multiple inhomogeneities. Their advice is for the use of SNH and SNT, in fact, but it is also applicable and beneficial for many other OHOMs. These instructions are as follows. (a) If a change point is found, the position of this pre-detection is flagged, and the time series is cut into two sub-periods at the time point of the pre-detection. If a gradual change is found, its endpoints are flagged, and the first sub-period lasts until the first flag, and the second sub-period begins from the second flag. (b) Sub-periods not shorter than 10 years are examined for further inhomogeneities. (c) The first phase of the examination is finished when no further inhomogeneity is detectable in the sub-periods. (d) The second phase of the detection procedure begins. Its name is backward testing of inhomogeneities. Pre-detected inhomogeneities are checked one-by-one, starting from the end of the time series. In an individual step, a sub-period with only one inhomogeneity (flagged in the first phase) is examined, and the correction, assessed from the ongoing step, is applied

promptly. Therefore, the applied window in the backward testing of a pre-detected inhomogeneity stretches from the time point of the left hand side adjacent inhomogeneity pre-detected to the end of the series. (There are no inhomogeneities to consider on the right hand side of the series in this phase, because they have already been corrected.)

During the applications of OHOMs, we follow the above instructions with some slight modifications. (a) Not only the minimum length of any sub-periods for further examinations is given, but the minimum size of the half-windows is fixed as 5 years. (b) The corrections are not applied promptly after the individual steps of the second phase, but their influence is eliminated applying shorter windows, so that time windows stretch only to the adjacent inhomogeneity on the right hand side of the time series.

The rules above are applied in most of the OHOMs, but the Eas, MS1, and MSH methods are exceptions, since they have own solution to how to treat the multiple inhomogeneities. In the MS1 method the maximum half-window length is 5 years, so this parameter of MS1 is set to be identical with those of the other OHOMs.

5. Derivation of relative time series

We intend to apply the homogenization procedures described in the previous section on all of the 327 time series specified in Section 3. Before application, however, relative time series are created for each observed series, referred also as parent series hereafter, and all of the homogenization procedures are performed on these relative time series.

The creation of relative time series needs the determination of a reference series for each parent series (see also Section 2). Another task is to convert relative time series into a standardized mode, for which the statistical characteristics of inhomogeneities, detected in relative time series, are comparable, despite the fact that the basic statistical properties of parent series are very different. Therefore, this section is divided into two parts: in the first part the building of reference series is presented, then the standardization method of relative time series is described.

5.1 Building of reference series

We follow the method proposed by *Alexandersson* and *Moberg* (1997), but with certain modifications. Before starting, the series of precipitation data are subjected to scale transformation to eliminate differences in ratios. Temperature is basically an additive variable, so ratio differences are not considered for that.

In Eq. (4) x denotes raw precipitation data, and x^* denotes transferred data. The scaling factor is the reciprocal rate of the time average for the series under transformation, relative to the spatial-temporal average for the whole data segment.

$$x^*_{j,i} = \frac{\frac{1}{M}\sum_{m=1}^{M}\overline{X_m}}{\overline{X_j}} x_{j,i} \qquad j \in (1,2,...M), \tag{4}$$

where M is the number of data series within a data segment, j denotes the time series that is just under transformation, and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \tag{4a}$$

After fulfilling this transformation, Eq. (2) is applicable both for temperature and precipitation time series. We mention that *Alexandersson* and *Moberg* (1997) propose a transformation of making 0 time average for all time series (uses anomalies from $\overline{X_m}$). However, this step may be omitted, because all the time series in any of the data segments (built in Section 3) cover exactly the same period.

The values of reference series are determined by Eq. (5). According to it, the reference series **F** originates from K data series of the data segment including the candidate series. All of the series have the same, n years length.

$$f_i = \frac{\sum_{m=1}^{K} r_m^2 x_{m,i}}{\sum_{m=1}^{K} r_m^2} \qquad i = 1,2,...,n. \tag{5}$$

The factor r_m^2 depends on the spatial correlation between the candidate series and the time series marked by m. More precisely, r_m is the spatial correlation for the series of differences (**D**) between the adjacent elements of data vectors, between the candidate and the $\mathbf{X}_m$ series. The calculation of the elements of **D** is shown in Eq. (6).

$$d_{m,i} = x_{m,i} - x_{m,i-1} \qquad i = 2,3,...,n. \tag{6}$$

Data series with relatively high r values may compose a reference series, because low spatial correlation indicates substantial geographical difference, which likely contains climatic differences, too. The use of difference series instead of the original **X** series in the calculation of spatial correlations has a

considerable advantage, namely, the correlations are less affected by inhomogeneities, thus they indicate geographical coherence with higher reliability.

Alexandersson and *Moberg* (1997) do not give concrete instructions how to chose the K number of the utilized data series in the building of reference series. It is not an easy task to set general rules for this, which would operate well in a wide range of conditions, because the optimal selection of the series depends on several factors. For example, a utilized data series must be representative spatially, must have a rather high correlation with the candidate series, must have a fairly high quality, etc. It is favorable, if there are composites of the reference series from all of the four main geographical directions, but this rule would be applicable only on large data segments of dense observing networks.

We introduce some simple criterions for selecting data series composing reference series. Using these criteria, the automatic building of reference series, related to a given candidate series, is possible in the following steps. (1) The time series, which are potentially useful for spatial comparison, i.e., the other series of the data segment of the candidate series, are ordered according to the spatial correlation with the candidate series (r), and series of the highest r values are selected. (2) The threshold r_t above which time series are included in the reference building is not a constant, but it depends on the number of series selected in the previous steps. If there is a large number of series with high spatial correlations, a relatively high r_t threshold can be chosen, and vice versa. (3) The threshold r is not allowed to be lower than 0.7. (4) The minimum number of selected composites is 2, otherwise reference series cannot be built to the candidate series, and the homogeneity test is rejected.

A proper formula was constructed to satisfy all the above criteria. The change of r_t in the function of the serial number (m) of the series under examination is given by Eq. (7). The first four values are constant (0.7), then r_t rises with m exponentially, faster at first, slower later, and it approaches asymptotically to 1 (*Fig. 1*).

$$r_{t,m} = 0.7 \qquad (1 \le m < 4),$$
$$r_{t,m} = \sqrt{1 - p(1 - r_{t,m-1}^2)} \qquad (5 \le m < M), \tag{7}$$

where p was chosen to meet the arbitrary condition of $r_{t,5}=0.72$, thus $p=0.9443$.

The spatial correlations between temperature values are well above 0.7 within the small area covered by the Hungarian observing network. Even in the particular data segment supplemented by temperature series from a larger

area (i.e., from Bologna to Cracow), the spatial correlations are still sufficient to build reference series. (The only exception is Bologna, for July, for which the second largest spatial correlation is only 0.688, also accepted by a specific decision to build the reference.)

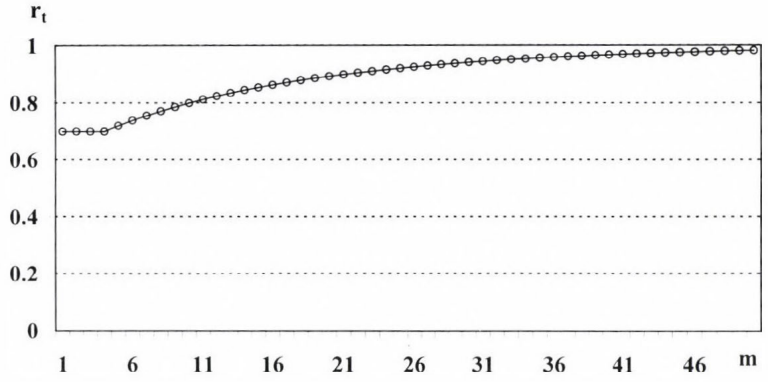


Fig. 1. Minimum threshold of spatial correlation (r_t) for the inclusion of time series into the process of building reference series. *m* means the *m*th highest spatial correlation between the candidate series and any other series of the data segment.

However, the correlation between precipitation series declines with distance much faster, than in case of temperatures. Correlation coefficients between precipitation series are particularly low in late spring and in summer, when considerable part of precipitation is of convective origin with rather hectic spatial distribution. Therefore, segments comprising July precipitation totals were excluded from the whole investigation. The correlation coefficients for April are substantially higher than those for July, thus all the series of April precipitation are included, although with some compromise in r_t values. Namely, r values between 0.64 and 0.7 were accepted if there was no other possibility to build the reference series. There were no problem in building of reference series for January, October, and annual precipitation totals.

5.2 Standardization of relative time series

When the results of an OHOM is under evaluation, the typical size of the detected shifts is a major characteristic to consider. However, how could one compare a 1 °C shift in temperature with a 20 mm shift in precipitation? Even the importance of a 1 °C shift is likely different for time series of monthly

values, than for annual values. These examples show that standardization is needed before the comparison of detected inhomogeneities. Naturally, this standardization can already be performed before the homogenization procedure. The way, how we fulfilled this task is specified in the following.

It seems to be simple to express the inhomogeneity magnitudes in proportion to the empirical standard deviation of the time series (s_T). However, s_T itself also depends on inhomogeneities in the series. If these inhomogeneities are relatively large, then the standard deviation is considerably larger, than that for the homogeneous series. This feature of s_T would cause undesired biases in the results.

Theoretically, the standard deviation of the white noise component (s_e) would be the optimal reference unit to which the inhomogeneity magnitudes can be related, but this component is not known exactly. Nevertheless, the following estimate of the unit is closer to s_e than the empirical standard deviation of the series. The core of the concept is to reduce s_T with the variability related to the serial correlation (r_s), since that part of the variability is obviously not white noise:

$$s_e^* = \sqrt{1 - r_s^2}\, s_T \,. \tag{8}$$

For Markovian processes, Eq. (8) provides an exact determination of the standard deviation of the noise, thus for those processes $s_e = s_e^*$. Although in real climatic time series s_e^* is always larger than the pure white noise, s_e^* is considered to be the standard deviation of the noise process in the study, and that is the reference unit of the detected inhomogeneities, if there is no indication of using another unit.

6. Results

Statistical characteristics of the detected inhomogeneities are presented here. Change point type inhomogeneities are under study at present.

The section consists of three parts. First, the inhomogeneity characteristics of the annual mean temperature series are shown, and results from the different OHOMs are compared. Second, inhomogeneity characteristics of other meteorological variables are compared with those derived for the annual mean temperature series. In the third part, the investigated time series are classified in another way, i.e., according to the similarity of their serial correlations, not considering the type of the meteorological variable they characterize.

6.1 Inhomogeneities in the series of annual mean temperatures

Annual mean temperature series are more frequently subjected to homogenization investigations than any other type of meteorological time series. The possible reasons are as follows. First, air temperature is a fundamental environmental variable. Thus, it is important to reveal its trends and low frequency changes relatively precisely. Second, its instrumental observation usually goes back to several decades or centuries. Third, the homogeneity of temperature series is very sensitive to technical conditions, hence, it is often affected by methodological or environmental changes. Fourth, networks for temperature observations generally have adequate density for ensuring sufficient spatial correlations, thus relative time series with fairly low noise level can be created. Homogeneity of annual mean temperature series has recently been examined, e.g., by *Jones* and *Lister* (2004) for Scotland and Northern Ireland, by *Türkes et al.* (2002) for Turkey, by *Alexandrov et al.* (2004) for Bulgaria, by *Domonkos* and *Tar* (2003) for Hungary, and by *Wijngaard et al.* (2003) for the whole data base of the European Climate Assessment project. We note that the examination of seasonal mean temperatures is in focus in several other studies (e.g., *Kyselý*, 2002; *Tomozeiu et al.*, 2002).

Table 2. Statistical characteristics of detected inhomogeneities in series of annual mean temperatures from observations: (a) rate of series found to be inhomogeneous, in %, (b) mean number of detected change points for inhomogeneous series, (c–e) moments of the shift magnitudes: (c) average, (d) standard deviation, (e) skewness

	a	b	c	d	e		a	b	c	d	e
Bs1	91	2.6	1.6	0.8	2.1	SNT	84	1.7	1.8	1.0	1.9
Bs2	95	2.9	1.6	0.8	1.9	MLR	100	2.6	1.8	0.9	1.6
WRS	86	2.1	1.6	0.8	3.2	Bay	100	3.3	1.7	0.9	1.6
Pet	91	2.7	1.5	0.8	2.6	Eas	91	2.0	1.9	0.9	2.4
M-K	100	2.3	1.3	0.8	1.3	MS1	98	3.1	1.8	0.9	1.6
SNH	84	2.5	1.7	0.9	2.1	MSH	98	3.2	1.9	0.9	1.6

In our investigations, 43 relative time series, derived from series of annual mean temperatures in four data segments, are examined with the twelve OHOMs listed in Section 3. *Table 2* shows the basic statistical properties of the detected inhomogeneities. The application of OHOMs indicates significant inhomogeneities in almost all series. More precisely, only 0–7 series are found

to be homogeneous (depending on the type of the OHOM), and 36–43 series were inhomogeneous from the investigated 43 series. Usually, multiple change points are detected, the average number of change points in an inhomogeneous series is between 1.7 and 3.3. However, the mean shift magnitude is low, it is always below 2. (Note, that this is a relative number, representing the proportion of the inhomogeneity shift to the noise term s_e* introduced in Section 5.2) The standard deviation of shift magnitudes is large in comparison with their average values. There are no great differences in the averages and standard deviations, calculated by the different methods.

The skewness values show slightly larger dependence on the applied method, but all of them are positive, and none of them is lower than 1. It means that most of the detected shifts are small, i.e., smaller than the average values, shown in columns "c" of *Table 2*.

The distribution of the detected shift magnitudes is illustrated in *Fig. 2*. The three curves of the figure represent the averaged results of 4-4 OHOMs. Curve (A) is for the OHOMs yielded relatively low magnitudes, curve (B) is for the OHOMs resulted in relatively few change points, and curve (C) is for the OHOMs indicating large number of change points often with relatively high magnitudes.

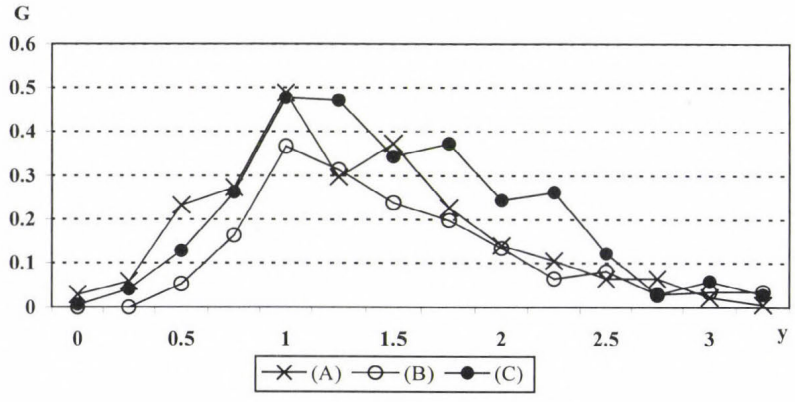


Fig. 2. Distribution of the detected shift magnitudes (*y*) in the series of annual mean temperatures. *G* indicates the mean number of occurrence per time series for predefined sections of the *y* values. Each curve is an average for four OHOMs. (A) Bs1, Bs2, Pet, M-K; (B) WRS, SNH, SNT, Eas; (C) MLR, Bay, MS1, MSH.

Although the three curves in *Fig. 2* and the content of *Table 2* demonstrate systematic differences between the individual OHOMs, these

differences are not very large, and some general features of the results are approximately uniform. The most important common features are (i) the indication of the change points in almost all of the series, and (ii) the predominantly low shift magnitudes. It is worth mentioning, that in case of an ideal equal distribution of shift magnitudes, the distribution of the detected magnitudes by OHOMs would show negative skewness, because the proportion of successful detection is higher for the large shifts than for the small ones *(Ducré-Robitaille et al.*, 2003). It suggests, that the positive skewness of the distribution of real shift magnitudes is even larger than that shown in *Fig. 2*.

6.2 Comparison of detected inhomogeneities for different meteorological variables

Data segments are sorted into five groups according to the type of meteorological variables: (i) annual mean temperature, (ii) July temperature, (iii) monthly mean temperature in transition seasons (April and October), (iv) January temperature, (v) monthly or annual precipitation total. The number of data series in the classes (i)–(v) is 43, 43, 86, 43, and 112, respectively. The class of precipitation total is not divided into smaller groups according to seasonality, because the frequency of detected change points is too low for a separate examination of monthly precipitation totals.

Four OHOMs are chosen for the comparative examination. They are Bs2, WRS, SNH, and MS1. All these OHOMs focus on detecting the change points, and their efficiency seems to be fairly good according to a preliminary study (*Domonkos*, 2005). *Fig. 3* shows the distribution of shift magnitudes for each class of meteorological variables. The shape of the curves is rather similar, albeit the frequencies of the detected change points are clearly different. The most change points are detected in the annual temperature series, followed by the July temperature series. Relatively few change points are detected in the precipitation series. There is a robust difference between the frequencies of detected change points in the different seasons. Inhomogeneities are much more frequent in summer than in winter: while 77–86% (depending on the applied OHOM) of the temperature series are found to be inhomogeneous in July, this rate is only 12–35% for the January temperatures.

In spite of the large differences among the total frequencies, the most frequent shift magnitude is almost the same for the examined classes of series. Shifts, whose magnitudes are close to the assessed noise level, are detected most frequently, whereas magnitudes above 2 (or above 2.5 for the annual temperatures) are very seldom, compared to the frequencies of smaller inhomogeneities.

Although inhomogeneity magnitudes are usually expressed in the proportion of the noise standard deviation (s_e*), the content of *Fig. 3* is presented also in physical units, i.e., °C for temperatures (*Fig. 4*) and mm for precipitation totals (*Fig. 5*). Beyond the differences of the applied units, another difference between *Fig. 5* and the class (v) of *Fig. 3* is that only annual totals were considered in the calculations for *Fig. 5*.

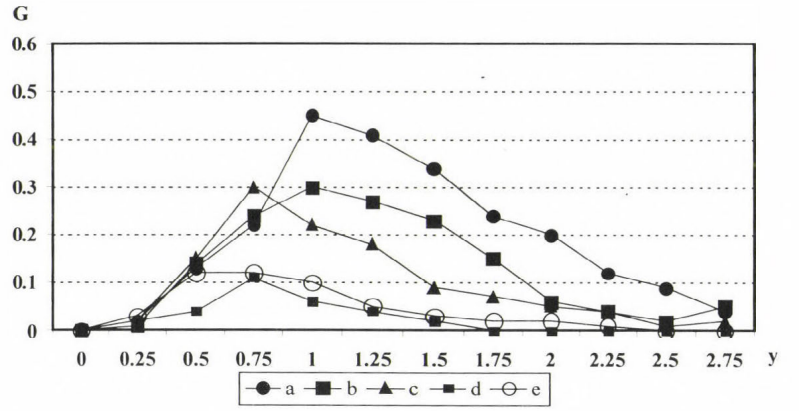


Fig. 3. Distribution of the detected shift magnitudes in the proportion of assessed standard deviation of noise. (a) annual mean temperature, (b) July mean temperature, (c) monthly mean temperature for April or October, (d) January mean temperature, (e) monthly or annual precipitation total.

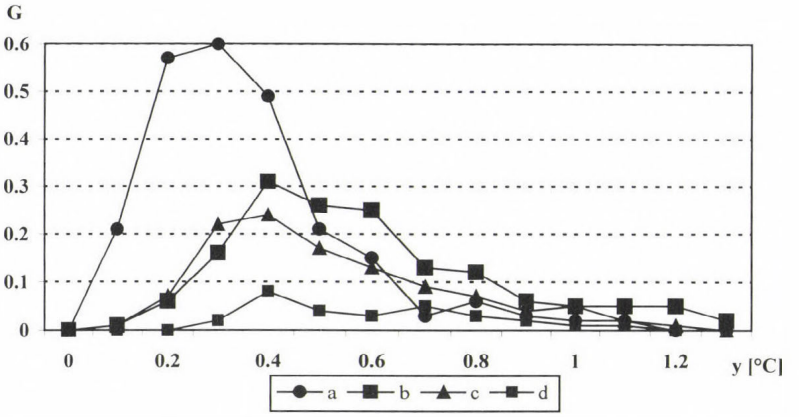


Fig. 4. Distribution of the detected shift magnitudes in temperature series, expressed in °C. (a) annual mean temperature, (b) July mean temperature, (c) monthly mean temperature for April or October, (d) January mean temperature.

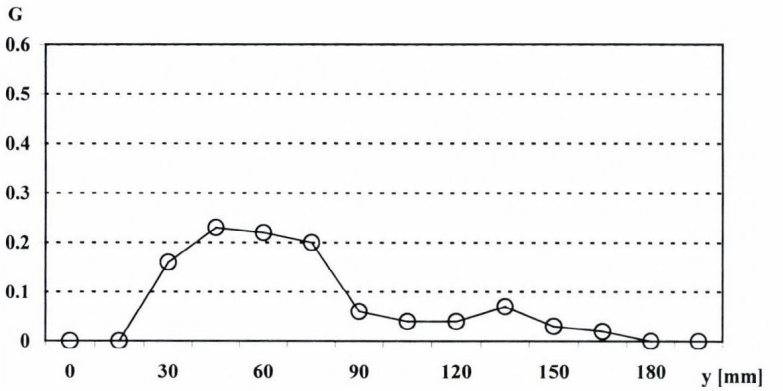


Fig. 5. Distribution of the detected shift magnitudes in time series of annual precipitation totals expressed in mm.

Fig. 4 confirms the main point of *Fig. 3*: the vast majority of the detected inhomogeneities is small. While the most large or medium shift magnitudes were found in the annual mean temperatures in *Fig. 3*, series of July temperatures have the most, relatively large inhomogeneities in the absolute scale. The annual mean temperature series have outstandingly large number of shifts in the range of 0.2–0.5 °C. For the monthly temperatures, the peaks of the magnitude distributions are flatter, and their typical range is 0.3–0.7 °C. Occurrence of shift magnitudes around or above 1 °C is much less frequent than that of the small shifts.

In the series of annual precipitation totals, fewer change points are detected than in the most temperature series, but these magnitudes are not very small. A flat peak of frequency appears between 30 mm and 90 mm, and these values are in the order of 1/10 of the climatic average for annual precipitation totals. While smaller than 30 mm shift magnitude has not been detected at all, the right hand tail of the distribution is long, with occurrences of 100–180 mm magnitudes. On the other hand, 20–50 percent (depending on the applied OHOM) of the annual precipitation series was found to be homogeneous.

All of the magnitude distributions, examined so far, have positive skewness. It strengthens the finding of Section 6.1, that the rate of small inhomogeneities is very high among the existing inhomogeneities. Theoretically, the experienced magnitude distributions of detected inhomogeneities may be composed in two possible ways: (a) There are numerous series with only small inhomogeneities, but a few series contain one or more large shifts instead of small inhomogeneities; (b) Small inhomogeneities are commonly present in time series. If a large shift occurs, its identification is relatively

difficult due to detectable or hidden small inhomogeneities. In the last part of this section, the relative time series are grouped in another way. The examination of groups independent from the parent data segments will give the answer to this question.

6.3 Classification of data series according to serial correlation

Serial correlation indicates the homogeneous or inhomogeneous character of a time series (see Section 4). Thus, the properties of detected inhomogeneities for classes of specified serial correlation ranges may be informative. *Table 3* shows the numbers of data series under study, for certain classes of serial correlation, r_s. Considering that $r_s > 0.25$ values are significant on the 0.01 level for a 100 years long time series, 40% of the series have substantially positive serial correlation. A small fraction (7.6%) of the series shows definitely high, at least 0.6 serial correlation.

Table 3. Numbers of data series in certain classes of serial correlation (r_s)

Class	r_s	Number of data series
Ra	< 0.15	137
Rb	0.15–0.25	59
Rc	0.25–0.40	59
Rd	0.40–0.60	47
Re	0.60 ≤	25

Fig. 6 shows the distribution of detected shift magnitudes for r_s classes in the same way, as *Fig. 3* does for specified meteorological variables. Also, the four OHOMs applied are the same in these figures. *Fig. 6* illustrates that higher serial correlation coefficients generally indicate more frequent and larger shifts in a time series, than lower r_s values. However, the highest frequency of shift magnitudes appears close to the noise level for each class, and frequency of around or above 2.5 is definitely rare, even for the class of uppermost serial correlations. It seems that the dominance of small shifts is very general, and series with merely large shifts are very seldom.

Table 4 presents further details about inhomogeneity properties for certain classes of serial correlation. It can be seen that serial correlation values above 0.25 are indicative of inhomogeneity, indeed. Examining the occurrence frequencies and mean magnitudes of the shifts, one may find substantial dependence on OHOMs, but the differences between the characteristics for

classes of the lowest and highest serial correlation coefficients are usually larger than the method dependence. Skewness values are always positive, again, and they are particularly high for the class of high serial correlation.

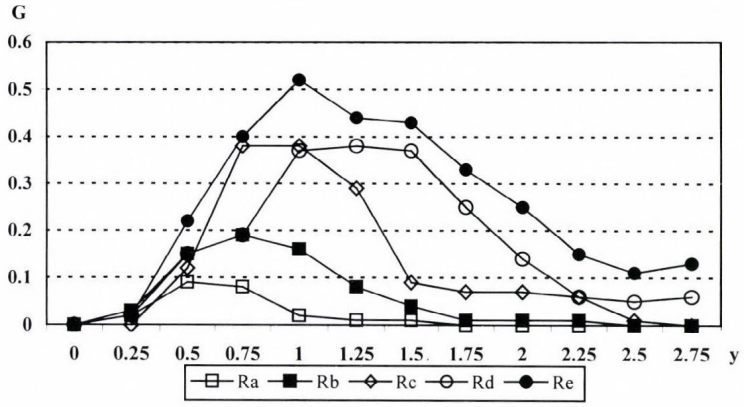


Fig. 6. Distribution of detected shift magnitudes for classes of serial correlation (r_s). Ra: $r_s < 0.15$, Rb: $0.15 \le r_s < 0.25$, Rc: $0.25 \le r_s < 0.4$, Rd: $0.4 \le r_s < 0.6$, Re: $r_s \ge 0.6$.

Table 4. Statistical characteristics of inhomogeneities for time series in certain classes of serial correlation

Serial correlation	Bs1	Bs2	WRS	Pet	M–K	SNH	SNT	MLR	Bay	Eas	MS1	MSH
	Rate of inhomogeneous series (%)											
< 0.25	28	26	24	31	25	25	26	27	27	34	42	43
0.25–0.4	81	86	80	88	95	81	81	98	98	71	100	98
≥ 0.4	96	100	96	96	100	94	94	100	100	93	100	100
	Mean number of change points in inhomogeneous series											
< 0.25	1.2	1.1	1.1	1.3	1.1	1.3	1.1	1.1	1.3	1.1	1.4	2.1
0.25–0.4	1.6	1.6	1.4	1.7	1.7	1.5	1.2	1.9	2.2	1.5	2.1	2.3
≥ 0.4	2.5	2.7	2.1	2.5	2.3	2.4	1.7	2.6	3.3	1.9	3.1	3.3
	Mean magnitude of shifts											
< 0.25	0.8	0.8	0.9	0.8	0.7	1.0	1.0	1.0	1.0	1.2	0.9	1.6
0.25–0.4	1.1	1.2	1.2	1.1	0.8	1.3	1.3	1.3	1.3	1.4	1.4	1.5
≥ 0.4	1.6	1.6	1.6	1.5	1.3	1.7	1.8	1.8	1.7	1.8	1.7	1.8
	Skewness of shift magnitude distribution											
< 0.25	0.9	0.4	1.7	0.7	0.2	1.1	1.1	0.8	1.0	1.2	1.1	1.0
0.25–0.4	1.2	1.7	1.1	1.2	0.6	1.3	1.5	0.7	0.4	0.8	0.8	0.9
≥ 0.4	2.5	2.1	3.1	2.5	1.1	2.2	1.8	1.8	1.7	2.4	1.9	1.7

7. Discussion and conclusions

Homogeneity characteristics of 327 time series, mostly Hungarian temperature and precipitation series, were investigated applying twelve objective homogenization methods (OHOMs). It was found that at least half of the series in the data base (50–65%, depending on the OHOM applied) contain significant inhomogeneities. However, most of the detected inhomogneities are small, their magnitudes are scattered around the noise level. A small part of the inhomogeneities are 2–3 or even more times larger than the noise level.

The highest frequency of inhomogeneities was found in the series of annual mean temperatures, followed by the series of July temperatures. Relatively few inhomogeneities were found in the series of precipitation totals, and very few in the series of January temperatures. Nevertheless, the relatively lower frequency of the detected inhomogeneities in precipitation series does not prove high quality for those demonstrated by the rather high absolute magnitudes of shifts in the time series of annual precipitation totals (*Fig. 5*). The explanation is the relatively low spatial correlations among precipitation time series, which result in higher noise levels of relative time series from precipitation totals, in comparison with series from mean temperatures.

The results for January temperatures need different explanation. In this case the spatial correlation coefficients are very high, but the frequency of the detected inhomogeneities is the lowest. These results together indicate a really high quality of the January temperature series. The likely explanation of the experienced, large seasonal difference in the quality of temperature series is that most of the technical and environmental problems are related to the incompleteness of sheltering from direct radiation effects, or to local radiation effects influencing the radiation- and heat-balance in the observing site. These effects are obviously larger in summer than in winter.

At least one significant inhomogeneity was detected in almost each of the series of annual mean temperatures. However, most of the detected shift magnitudes are small, they are around or below 0.5 °C. This result is a consequence of the very low noise level owing to the high spatial coherence and low standard deviation of annual mean temperatures. One may conclude, that the detection of such low inhomogeneities is unnecessary (either the detection is precise or not). However, the right evaluation of the results needs some further considerations: (a) Only few inhomogeneities can be detected with much higher magnitudes than the majority. (b) The identification of medium-size or large inhomogeneities may be affected by lower inhomogeneities in the same series. (c) In the investigation of climate change and climatic variability inhomogeneities of not larger than a few tenth of °C might also affect the conclusions.

The skewness of shift magnitude distribution was found to be positive in each examination. The presence of small inhomogeneities tends to be a very common feature of time series from meteorological observations, and time series with at least one, medium-size or large inhomogeneity likely contain other (usually smaller) inhomogeneities.

It would be useful to know, to what extent the results of the used data base are representative for other data bases (of other countries, other variables). We mention that the way of the creation of reference series may influence the inhomogeneity characteristics presented. To assess the representation of the results shown in this study, further investigations are needed on a larger data base with higher diversity of data types and methodology. However, our long-term aim is to find the best OHOM for homogenizing Hungarian temperature series, and the results shown here provide a proper basis for the testing of homogenization methods with time series whose properties are close to those of observed temperature series in Hungary.

One purpose of this study is to make some steps towards a scientifically correct estimation of the effectiveness of individual OHOMs. In an efficiency testing procedure, simulated data sets with known inhomogeneities must be used for evaluating the differences between the detected and factual inhomogeneities. However, statistical properties of the factual inhomogeneities of simulated series should be similar to those of observed data sets, otherwise the calculated efficiencies might provide false information for practical applications. Therefore, the gained knowledge about the basic statistical properties of detectable inhomogeneities in observed data sets is essential for the assessment of efficiencies. The high frequency of small inhomogeneity occurrences, that is a general experience of this study, must be kept in mind in the development of efficiency testing methods.

The main findings of the study are as follows:

- More than half of the investigated series are inhomogeneous (contains at least 1 significant inhomogeneity). The rate of inhomogeneous series is the highest for annual mean temperatures, and the lowest for January mean temperatures.
- The rate of inhomogeneous series is lower for precipitation totals than for temperature means. However, it does not prove a higher quality for precipitation data, as the relatively low spatial correlation coefficients do not allow to detect inhomogeneities of small magnitudes in precipitation series.
- The statistical characteristics of detected inhomogeneities depend on the OHOM applied, but this dependence is usually lower than that on data quality and spatial correlations.

- The series of January mean temperatures have the best quality, and July temperatures have the worst quality. The likely explanation is that most of the inhomogeneity effects on temperature series are related to changes in microscale radiation processes in the surroundings of the thermometer.
- The serial correlation of relative time series is usually a good indicator of the homogeneity quality of the parent time series. In the class of higher than 0.4 serial correlation, the mean number of the detected change points is 2–3 per time series.
- Contamination with small inhomogeneities seems to be a general attribute of meteorological time series. Thus, beyond a general white noise process, small inhomogeneities may also affect the identification of medium-size and large inhomogeneities. An evaluation of the effectiveness of OHOMs, planned for the future, must take this feature of meteorological time series into consideration.

Acknowledgements—This research was partly funded by the VAHAVA project of the Ministry of Environment and Water, in cooperation with the Hungarian Academy of Sciences, and by the Sky Magic Foundation (Debrecen, Hungary). The author thanks *Tamás Szentimrey* (Hungarian Meteorological Service) for the opportunity to use the original software of MASH in the examinations of the study.

References

Aesawy, A.M. and *Hasanean, H.M.*, 1998: Annual and seasonal climatic analysis of surface air temperature variations at six southern Mediterranean stations. *Theor. Appl. Climatol. 61*, 55-68.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol. 6*, 661-675.

Alexandersson, H. and *Moberg, A.*, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol. 17*, 25-34.

Alexandrov, V., Schneider, M., Koleva, E., and *Moisselin, J.-M.*, 2004: Climate variability and change in Bulgaria during the 20th century. *Theor. Appl. Climatol. 79*, 133-149.

Buishand, T.A., 1982: Some methods for testing the homogeneity of rainfall records. *J. Hydrol. 58*, 11-27.

Domonkos, P., 2005: Homogenisation of time series of observed environmental variables. In *12th Symposium on Analytical and Environmental Problems*. SZAB, Szeged, Hungary, 109-113.

Domonkos, P. and *Tar, K.*, 2003: Long-term changes in observed temperature and precipitation series 1901-1998 from Hungary and their relations to larger scale changes. *Theor. Appl. Climatol. 75*, 131-147.

Ducré-Robitaille, J-F., Vincent, L.A., and *Boulet, G.*, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol. 23*, 1087-1101.

Easterling, D.R. and *Peterson, T.C.*, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol. 15*, 369-377.

Jones, P.D. and *Lister, D.*, 2004: The development of monthly temperature series for Scotland and Northern Ireland. *Int. J. Climatol. 24*, 569-590.

Karl, T.R. and *Williams Jr., C.N.*, 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim. Appl. Meteorol. 26*, 1744-1763.

Kohler, M.A., 1949: On the use of double mass analysis for testing the consistency of meteorological records and for making required adjustments. *B. Am. Meteorol. Soc. 5*, 188-189.

Kyselý, J., 2002: Temporal fluctuations in heat waves at Prague-Klementinum, the Czech Republic, from 1901-1997, and their relationships to atmospheric circulation. *Int. J. Climatol. 22*, 33-50.

Moberg, A. and *Alexandersson, H.*, 1997: Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *Int. J. Climatol. 17*, 35-54.

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. and *Parker, D.*, 1998: Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol. 18*, 1493-1517.

Pettitt, A.N., 1979: A non-parametric approach to the change point problem. *Appl. Statistics 28*, 126-135.

Sneyers, R., 1997: Climate chaotic instability. Statistical determination - theoretical backgrounds. *Environmetrics 8*, 517-532.

Sneyers, R., 1999: Homogenizing time series of climatological observations. *Second Seminar for Homogenization of Surface Climatological Data*. Hungarian Meteorological Service, Budapest, 5-14.

Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH). *Second Seminar for Homogenization of Surface Climatological Data*. Hungarian Met. Service, Budapest, 27-46.

Tomozeiu, R., Busuioc, A., and *Stefan, S.*, 2002: Changes in seasonal mean maximum air temperature in Romania and their connection with large-scale circulation. *Int. J. Climatol. 22*, 1181-1196.

Türkes, M., Sümer, U.M., and *Demir, I.*, 2002: Re-evaluation of trends and changes in mean, maximum and minimum temperatures of Turkey for the period 1929-1999. *Int. J. Climatol. 22*, 947-977.

Vincent, L.A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate 11*, 1094-1104.

Wijngaard, J.B., Klein Tank, A.M.G., and *Können, G.P.*, 2003: Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol. 23*, 679-692.

BOOK REVIEW

Kalnay, E., 2003: **Atmospheric Modeling, Data Assimilation and Predictability**. Cambridge University Press, Cambridge, 341 pages, 73 figures, 4 colored plates.

Dynamic meteorology seems to be a very popular subject for university textbook writers. More than 10 high quality textbooks on the dynamics of the atmosphere have been published in the recent 25 years. It is not the case, however, in the field of numerical weather prediction, one of the most important applications of dynamic meteorology. This long felt need has been recently met by the book of an emblematic person of numerical modeling activities, Eugenia Kalnay, Distinguished University Professor of Meteorology at the University of Maryland, former director of the Environmental Modeling Center (EMC) at NCEP, principal investigator of many succesful projects in the field of predictability, ensemble forecasting, and data assimilation, author of several novel publications. The international science community has been waiting for decades for a comprehensive textbook on atmospheric modeling, and finally Dr. Kalnay has just given us the summary of basic knowledge on this rapidly developing territory of meteorology.

The author discusses all aspects of atmospheric and oceanic computer modeling, including a historical overwiev of the subject, equations of motion and their approximations, a clear description of up-to-date numerical methods, and the determination of initial conditions using weather observations, an important new branch of science known as data assimilation. The book also provides a clear discussion of the problem of predictability and chaos in dynamical systems, and investigates how this knowledge can be applied to atmospheric and oceanic modeling. The text touches on the discussion of ensemble forecasting, ENSO events, and the possibility of the improvement of weather and climate prediction.

All in all, this is the book students, lecturers, and researchers of atmospheric modeling have been waiting for decades.

Gy. Gyuró

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 double 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. Good quality laser printings are preferred.

The text should be submitted both in manuscript and in electronic form, the latter on diskette 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.

More information for authors is available: antal.e@met.hu

Information on the last issues: http://omsz.met.hu/irodalom/firat_ido/ido_hu.html

Published by the Hungarian Meteorological Service

Budapest, Hungary

INDEX: 26 361

HU ISSN 0324-6329

IDŐJÁRÁS

QUARTERLY JOURNAL
OF THE HUNGARIAN METEOROLOGICAL SERVICE

CONTENTS

****

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

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*

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 91–123

Case study and numerical simulations of the November 19, 2004 severe windstorm in Central Europe

André Simon[1*], **Ákos Horváth**[2] and **Jozef Vivoda**[1]

[1]*Slovak Hydrometeorological Institute, Jeséniova 17, 833 15 Bratislava, Slovakia*
[2]*Hungarian Meteorological Service, Storm Warning Observatory*
Vitorlás u. 17, H-8600 Siófok, Hungary
E-mails: Andre.Simon@shmu.sk; Horvath.A@met.hu; Jozef.Vivoda@shmu.sk

(Manuscript received in final form April 18, 2006)

Abstract—The study analyses the synoptic weather situation and mesoscale impacts of the November 19, 2004 windstorm, which affected several countries of central Europe. Particular attention was paid to the windstorm at the High and Low Tatra regions in Slovakia, where the event showed several attributes of downslope windstorm. This was investigated by using the ALADIN and MM5 numerical models with high horizontal resolution. Effects of the hydrostatic and non-hydrostatic dynamics were compared, as well as the method of the so called dynamical adaptation. It is concluded that downslope windstorms similar to the November 19 case can be forecasted by numerical models of 2.5 km resolution with higher precision than by using the current operational models. Nested version of the MM5 model at 1.0 km indicated possibility of simulation of microscale effects as orographically induced jets. It is shown that computationally effective hydrostatic models based on dynamical adaptation approach can be sufficient in forecasting extreme non-convective wind, similar to the evaluated event. Nevertheless, the results are strongly dependent on the physical parameterization of the model (turbulence, orographic drag, etc.). Hence, future versions of both hydrostatic and non-hydrostatic numerical models should be carefully examined and tested to keep the performance of forecasting severe downslope windstorms.

Key-words: numerical model, high resolution, downslope windstorm, High Tatras

1. Introduction

Forecasting of rapid cyclogenesis and mesoscale weather phenomena belongs to the main objectives of modern synoptic meteorology and numerical weather prediction. Both aspects were present in the case of the November 19, 2004

* Corresponding author

windstorm that affected several countries in central Europe. The main reason of severe weather was a rapidly developing and fast propagating macrosynoptic scale cyclone that was well predicted by several global and limited area numerical models used in the operational service at Slovak Hydrometeorological Institute (SHMÚ) and Hungarian Meteorological Service (HMS). The passage of the cyclone caused conditions favorable for mesoscale enhancement of the wind, above all at mountain areas. Some of these events can be appointed to occurrence of downslope windstorm and became particular object of this study.

Downslope windstorm is defined as very strong, usually gusty, and occasionally violent wind that blows down the lee slope of a mountain range, often reaching its peak strength near the foot of mountains and weakening rapidly farther away from the mountains (*Glossary of Meteorology*, 2000). This kind of flow appears at many mountainous regions all around the world (*Lane et al.*, 2000; *Sandvik*, 2001; *Tutiš*, 2002; *Águstsson* and *Ólafsson*, 2005) and can be sometimes associated with considerable damages (*Gaberšek*, 2003; *Meyers et al.*, 2003; *Vigh*, 2005; *Klaić* and *Belušić*, 2005) and also with potential hazard to aviation (*Kaplan et al.*, 2003; *Doyle et al.*, 2006). Several synoptic-dynamical studies were given about events from Dalmatian coast of Croatia called "bora" (*Klemp* and *Durran*, 1987; *Picek* and *Tutiš*, 1996; *Jurčec* and *Brzović*, 1995). Occurrence of severe downslope wind is also very frequent at eastern slopes of central Colorado Rocky Mountains in the vicinity of Boulder (*Lilly*, 1978).

The study of the flow structure at lee side of the mountains usually requires high density of surface observations and balloon, airplane, or LIDAR observations (*Vergeiner* and *Lilly*, 1970; *Gohm* and *Mayr*, 2005a, b), which are not always available. Operational forecasting and better understanding of the flow associated to downslope windstorms demands numerical simulations with models of resolution much higher than it was used at SHMÚ and HMS on November 19, 2004 (*Dierking*, 1998; *Vosper*, 2003; *Zhang et al.*, 2005).

For above mentioned reasons, high resolution numerical models were used to study the wind field in the area of the High and Low Tatras, where the November 19, 2004 windstorm caused particularly large devastation to the forest. Both hydrostatic and non-hydrostatic versions of the ALADIN model were tested at SHMÚ, and the non-hydrostatic Penn State and UCAR MM5 model was evaluated at HMS.

The main motivation of the research using the ALADIN model was to see, how much a 2.5 km high resolution hydrostatic model can improve the operational forecast of the windstorm. The use of hydrostatic approximation and hydrostatic dynamics at 2.5 km horizontal resolution can be still plausible by simulation of non-convective flows, although features related to gravity waves reflection and propagation (trapped lee waves, potential flow) are

already misinterpreted at this resolution (*Kasahara* and *Qian,* 2000; *Laprise* and *Peltier*, 1989b; *Nance* and *Durran*, 1987a, b; *Smith*, 2002). The length of large amplitude mountain waves predicted by hydraulic theory is usually bigger than 10 km. Waves of these dimensions and features, as hydraulic jumps, are usually reproduced by hydrostatic models (*Klemp* and *Lilly*, 1978). On the other hand, flow at the lee side of the mountains produces significant vertical accelerations and vertical downward and upward velocities in orders of units or even tens of m/s observed by glider pilots (*Zejda*, 1986) or during field experiments (*Grubišić* and *Lewis*, 2004). In these areas one may expect presence of not negligible non-hydrostatic pressure perturbations and pressure perturbation forces.

Hence, another objective was to see, whether the non-hydrostatic approach produces forecasts of different quality. The purpose of the MM5 experimental run was to obtain a wind field at 1 km resolution, which could show microscale effects as rotors (*Kuettner* and *Hertenstein*, 2002; *Doyle* and *Durran*, 2002) or wind strengthening by mountain passes, called gap wind (*Colle* and *Mass*, 1998a, b, 2000; *Pan* and *Smith*, 1999). Another point of the research was the effectiveness of several computational approaches including the dynamical adaptation of the ALADIN model (*Žagar* and *Rakovec*, 1999) and nesting technique of MM5. Beside that, sensitivity on different packages of physical parameterizations was tested on operational model versions of the ALADIN SHMÚ model. It was the first occasion both for SHMÚ and HMS, that a windstorm was studied in such extent, using high resolution numerical models.

The paper is divided in 10 Sections. The next Section gives a brief description of the macro-synoptic situation. Section 3 is related to meso-synoptic conditions for mountain wave generation. Sections 4 and 5 inform about the impact of the windstorm in central Europe with more details about the event at the High Tatras. Section 6 is an overview of research methods and brief description of used versions of numerical models. Section 7 presents forecasts of the model operationally used on November 19, 2004. Sections 8 and 9 show the main results of the high resolution runs of the ALADIN and MM5 models, respectively. Conclusions and consequences for operational forecasting are given by Section 10.

2. Synoptic situation

A day before the event, on November 18, 06:00 UTC, it was already possible to recognize the development of shallow cyclones in a widespread trough of low pressure, with axis westerly from Ireland. The weather in central Europe was influenced by deep cyclone centred over the southern part of Scandinavian

Peninsula and Baltic Sea (the sea level pressure dropped below 980 hPa in the middle of this low). In the upper troposphere, a jet stream was spread from Northern Atlantic with axis over the Northern Sea, Denmark, and Poland with northwesterly wind of speed exceeding 80 m/s along the axis at the 300 hPa height according to ECMWF analysis. At the same time, at both 850 and 500 hPa levels, a strong southward directed temperature gradient formed that could be detected over big part of northern Europe (*Fig. 1*). A strong westerly flow was present over Europe containing several patterns of mesosynoptic scale perturbations.

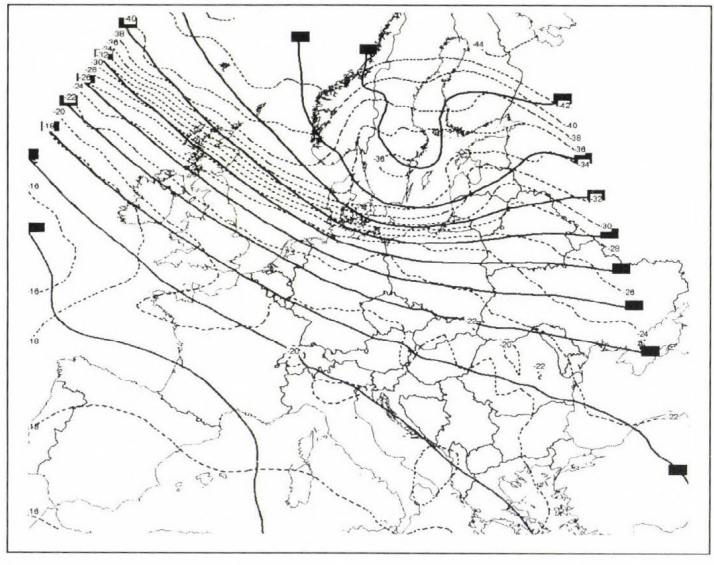


Fig. 1. Geopotential height (m) and temperature (°C) at 500 hPa from the 6-hour forecast of the ECMWF model, valid for November 18, 2004, 06:00 UTC. Note the well expressed baroclinic zone over Scotland, Northern Sea, Denmark, and the Baltic Sea region.

On November 19, 00:00 UTC, there was already a deep surface trough and frontal zone over Northern Germany and France. The upper air trough with axis over Scandinavia, the Northern Sea, and the British Isles deepened as well, and the axis of the jet stream moved southward. This caused change of the flow direction over northwestern Europe from westerly to northwesterly. Consequently, intense advection of cold air started, which can be demonstrated on changes of potential vorticity (PV) field at upper tropospheric levels. Ertel's potential vorticity on isentropic surfaces is

$$PV = -g(\zeta + f)\frac{\partial \theta}{\partial p}, \tag{1}$$

where g is the acceleration of gravity, ζ is relative vorticity, f is the Coriolis parameter, θ represents potential temperature, and p is the atmospheric pressure (see, e.g., the textbook of *Bluestein,* 1993). The potential vorticity is usually expressed in Potential Vorticity Units (1 PVU = 1×10^{-6} m^2 s^{-1} K kg^{-1}). Maximum values of potential vorticity (more than 10 PVU) occur usually in the stratosphere due to high static stability of the stratospherical air. Hence, considerable increase of potential vorticity can be found above the tropopause, which is usually considered at the level with potential vorticity of 1.5 PVU. Regions with negative potential vorticity indicate presence of dry symmetric instability (*Bennets* and *Hoskins*, 1979). The mean distribution of potential vorticity at isentropic levels is zonal, however, meridional advection of cold/warm air creates disturbances (anomalies) at upper tropospheric levels (315–330 K). Cold, polar air-masses are characterized by low altitudes of tropopause and the 1.5 PVU level (4–8 km), whereas in warm, mid-latitude airmasses the tropopause is reached higher (8–12 km). Exchange of the air mass is often manifested by sudden drop of the 1.5 PVU height. For practical use, maps of the 1.5 PVU surface heights are plotted to follow the displacement of air-masses and formation of upper air potential vorticity anomalies (*Morgan* and *Nielsen-Gammon*, 1998).

In the evaluated case the cold air advection was marked by drop of the 1.5 PVU height over the Northern Sea. Hence, upper air anomaly of potential vorticity was created, and it was spreading from the area of Northern Sea southeastwards (*Fig. 2a*). Approach of the positive upper air potential vorticity anomaly to surface baroclinic zone in environment of strong vertical wind shear created ideal conditions for baroclinic instability (*Hoskins et al.*, 1985). After November 19, 00:00 UTC it is possible to observe an isolated but shallow cyclone (996 hPa deep) over the central part of Germany, which deepened very fast during the next hours and propagated westwards. At 06:00 UTC the centre of the cyclone is already over the borders of Germany, Poland, and Czech Republic. In the vertical cross Section through the cyclone, it is possible to see well expressed anomaly of potential vorticity at low and middle tropospheric levels (*Fig. 2b*). Such anomalies are consequences of diabatic heating associated with latent heat released by precipitation or related to effects of friction (*Davis* and *Emanuel*, 1991; *Romero et al.*, 2002; *Arreola et al.*, 2003 and *Adamson et al.*, 2006). This, in turn, supports baroclinic instability, development of upward vertical motions, and deepening of the cyclone. At 12:00 UTC the cyclone moved over the southeastern part of Poland, at 18:00 UTC it was situated over Ukraine and Belarus (*Fig. 3*).

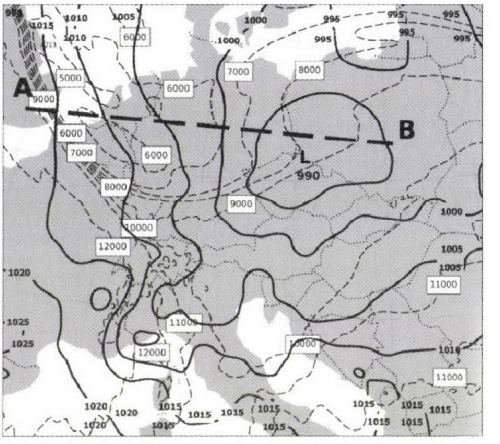


Fig. 2a. Height of the 1.5 PVU level (m) showed by dashed lines and mean sea level pressure (hPa) by solid lines from the 6-hour forecast of the ALADIN SHMÚ model, valid for November 19, 2004, 06:00 UTC. The thick dashed AB line marks the direction of the cross-section in Fig. 2b.

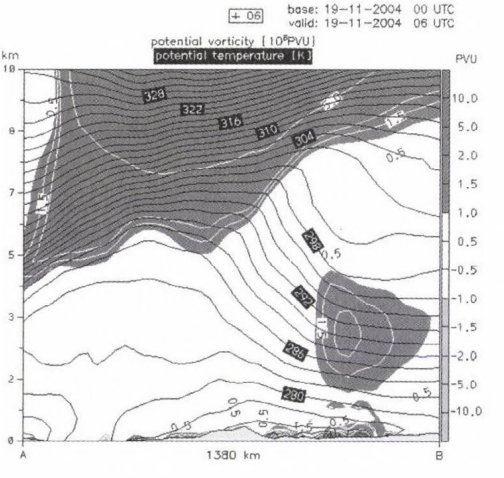


Fig. 2b. Vertical cross section through the field of potential vorticity (white isolines and grey scale) and potential temperature (black isolines), valid for November 19, 2004, 06:00 UTC. The potential vorticity field is shown in Potential Vorticity Units (1 PVU= 1×10^{-6} m^2 s^{-1} K kg^{-1}, right axis). Dark grey filled areas show values of potential vorticity higher than 1 PVU (values in the troposphere are usually lower than 1 PVU). Light grey filled areas (with potential vorticity smaller than 0 PVU) are associated with the presence of dry symmetric instability.

Note the upper air potential vorticity anomaly at left (west) side of the cross-section (the anomaly is visible as the drop of the 1.5 PVU height). This anomaly interacts with isolated low-level potential vorticity anomaly on the right (east) side of the cross-section (with maximum value of potential vorticity exceeding 2.0 PVU). Thus, strong vertical motions are created at mid-troposphere in the part between the anomalies (indicated by the tilt of the potential temperature isolines).

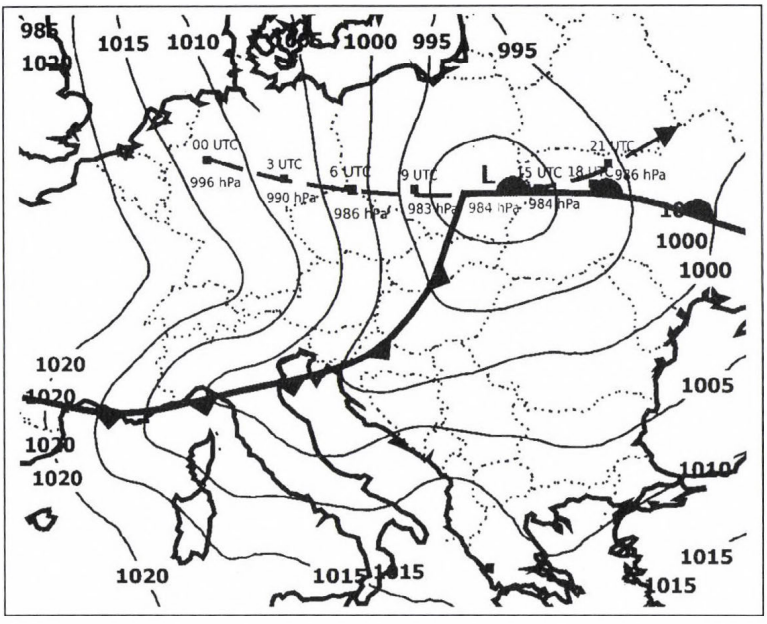


Fig. 3. Analysis of the mean sea level pressure (in hPa) and fronts, valid for November 19, 2004, 12:00 UTC. The dashed line shows the track of the November 19 cyclone with positions of its centre and value of the lowest mean sea level pressure from SYNOP observations from November 19, 2004, 00:00 UTC until November 19, 2004, 21:00 UTC with 3-hour frequency.

The passage of this cyclone in the central part of Europe was accompanied by strong cold air advection and fast changes of pressure. At 09:00 UTC, when the cold front of the cyclone reached the borders of Germany and Czech Republic, 12 hPa/3h rise of pressure were reported over Germany and 5–8 hPa/3h drops of pressure in the Czech Republic. The passage of the cold front according to high pressure gradients and strong flow both in upper and lower parts of the troposphere was fast — it reached the northwestern borders of the Slovak Republic at 11:00 UTC, at 13:00 UTC the line of the front crossed the High Tatra region, Budapest, and Balaton, after 16:00 UTC the front left the territory of Hungary and the Slovak Republic. The passage of the front was associated with wind gusts reaching 20–30 m/s at many places and with turn of the wind direction from southwesterly to westerly and northwesterly, respectively. Thunderstorms occurred at frontal line — already in the morning hours in the south of Germany and west of Austria, later, between 11:00 and 15:00 UTC, several thunderstorm clouds formed in the northwest and central parts of Hungary.

3. Mesoscale conditions for mountain wave generation

The mechanism of downslope windstorms is very closely related to generation of mountain waves. First explanation of this mechanism was given by nonlinear hydraulic theory, in which the airflow over a mountain was compared to fluid flowing over an obstacle (*Long*, 1953, 1954). The properties of the upstream flow can be characterized by Froude number (F_r). Different definitions and purposes of using the Froude number can be found in textbooks and scientific papers (*Smith*, 1980; *Durran*, 1986). In this paper the Froude number is considered as fraction of the kinetic energy of the upstream air parcels and potential energy needed to overcome the obstacle.

$$F_r = \frac{U}{NH}, \tag{2}$$

where N is the Brunt Väisälä frequency, U is the wind speed, H represents the height of the obstacle (*Bluestein*, 1993). The Froude number defined by Eq. (2) describes the ability of the air to flow over obstacles and to generate mountain waves. It can be used also as a measure of linearity of the flow regime described by certain systems of dynamic equations (*Smith*, 1980).

The air has a tendency to go around the obstacle by weak upstream flow or high static stability ($F_r << 1$). Such airflow is called subcritical. However, as the airspeed increases (or static stability decreases), a part of the airmass starts to flow over the obstacle, while generating waves at the lee side of the mountain (*Smolarkiewicz* and *Rotunno*, 1989; *Stull*, 2003). If the kinetic energy of the upstream air parcels overlaps the potential energy needed to overcome the obstacle ($F_r > 1$), the flow is called supercritical and generates stationary wave of large amplitude at the lee side of the mountain. As a consequence, strong winds start to blow downhill, at the downslope part of the wave, because potential energy of the air parcels is converted into kinetic energy during the entire time when the parcels are traversing the mountain. On the other hand, surface winds calm down quickly below the upslope part of the wave, where the supercritical flow undergoes a turbulent transition (called hydraulic jump) to the subcritical environment (*Holton*, 1992). A hypothetical example of idealized hydraulic flow across the High Tatra Mountains is given by *Fig. 4a*.

A linear theory was based on generation of vertically propagating internal gravity waves (*Klemp* and *Lilly*, 1975). Klemp and Lilly showed that amplification of surface winds is related to the reflection of some parts of the upward propagating wind energy due to the presence of middle tropospheric level inversion. These conditions are favorable for generation of trapped lee waves (*Fig. 4b*).

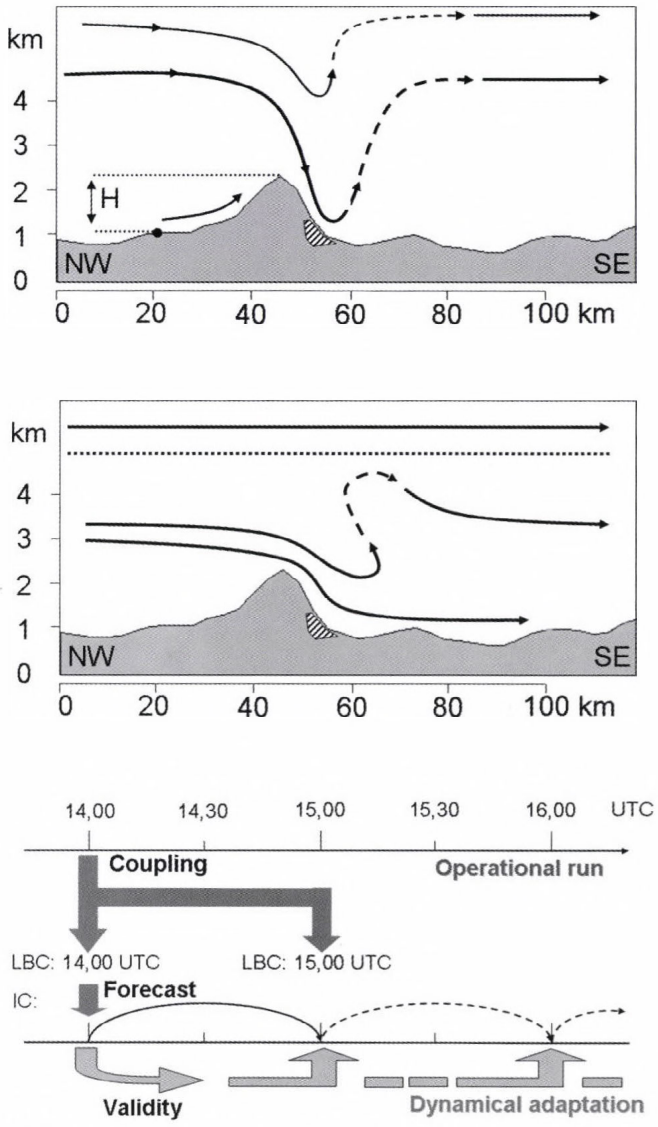


Fig. 4. Schemes of three possible mechanisms of downslope windstorm occurring at the High Tatras on November 19, 2004: (*a*) creation of a large mountain wave according to nonlinear hydraulic theory, (*b*) development of trapped lee waves related to reflection of internal gravity waves, (*c*) breaking of mountain waves and reversal of the flow. The northwest-southeast cross-section is applied to the relief of the High Tatras. Solid lines are streamlines; the dashed part denotes the upslope part of the wave (hydraulic jump). The symbol H means the height of the mountain over terrain at the upstream reference point (selected for calculation of the Froude number in Section 3). The thick dotted line represents the area of the temperature inversion in (*b*) and the so called critical level with flow reversal in (*c*). The hatched area at the lee slope of the mountain marks the location of the most intense damage observed at the High Tatras.

Third theory was presented in papers of *Clark* and *Peltier* (1977) and *Peltier* and *Clark* (1979). According to their research, based on numerical simulations, downslope windstorm is related to mountain wave breaking and reversal of the cross-mountain wind (*Fig. 4c*). In that case the energy of the upward propagating wave remains trapped below a certain critical layer (where the flow component across the mountain drops to zero and reverses sign), thus producing significant increase in the wave amplitude. *Durran* (1986) attempted to create a unifying concept of downslope windstorms based on nonlinear transition from upstream gravity wave structures to supercritical flow along the lee side of the mountain. He showed that wave breaking may be explained by hydraulic theory. It was confirmed that this phenomenon can be important for amplification of the surface wind, although it is not an ultimate condition for the onset of downslope windstorm.

Increase in computer power and development of numerical weather prediction techniques allowed the simulation of mountain waves on realistic case studies with high resolution numerical models (*Satomura* and *Bougeault*, 1994; *Belušić* and *Klaić*, 2003; *Kraljević* and *Tudor*, 2005) and recently even operational forecasting of downslope windstorms (*Ivatek-Šahdan* and *Tudor*, 2004). Results achieved in case studies generally accept all above listed possibilities of mountain waves generation and amplification, showing features as hydraulic jump, trapped lee waves, or lee flow reversals, which were simulated earlier in idealized experiments or described analytically.

Aerological observations in the vicinity of the November 19 cyclone at 12:00 UTC show, that the thermal stratification was stable and the air was relatively moist before the passage of the cold front (at aerological stations Poprad Gánovce and Budapest Lőrinc). There was a remarkable wind shear in the low and middle troposphere, while the wind speed increased from 5 m/s at the ground until 30 m/s at the 700 hPa altitude. At or just behind the line of the cold front, the stratification became potentially unstable at lowest 1600 meters, but the estimated energy of instability was rather small (61 J/K measured at the aerological station Prostějov in the Czech Republic). The area of instability (or small stability) was capped by shallow and not very significant layer of temperature inversion. Aerological observations on November 20, 00:00 UTC showed considerable drop of the temperature due to strong cold air advection at the rear side of the cyclone (e.g., the temperature at the 850 hPa level at Poprad Gánovce decreased from 0.5 to -8.8 degrees of Celsius within 12 hours) and presence of dry air in the entire profile of the troposphere.

There are no aerological observations that would be representative for the conditions at the windward side of the High Tatra Mountains during the windstorm. Hence, the properties of the air mass flowing over the crest of Tatras were assessed from ALADIN model forecast (*Fig. 5*). The static

stability represented by the Brunt-Vaisala frequency was very low at lowest 700–800 meters above ground, and together with high wind shear it gave conditions favorable for turbulence (small Richardson numbers). The Brunt-Vaisala frequency significantly increased between the altitudes of 1500 and 2400 meters above sea level (just below the tops of the High Tatra Mountains). The Froude number (F_r) was evaluated with Brunt Vaisala frequency (N) and wind speed (U) averaged over the variable depth H that represented the height of the obstacle above terrain.

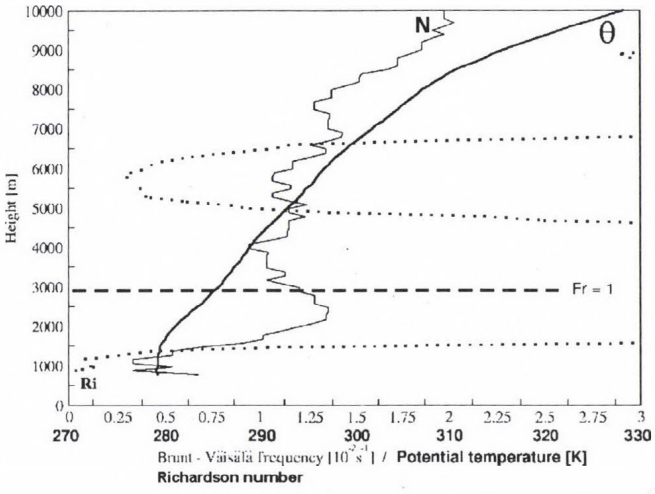


Fig. 5. Vertical profile of several parameters derived from the forecast of the ALADIN HS model, valid for November 19, 2004, 15:00 UTC and for the point over Zakopane (in Poland, at windward side of the High Tatras). Thin solid line denotes the Brunt Vaisala frequency N (upper horizontal scale); thick solid line denotes the course of the potential temperature θ (bottom horizontal scale). Dotted line shows the course of the Richardson number Ri (upper horizontal scale). The vertical axis indicates the height above sea level (in meters). Dashed line marks the height, where the dimensionless Froud number (given by mean static stability and mean wind speed of the layer below) attains unity. For obstacles below this height, the flow can be considered as supercritical and overcoming the obstacle.

The Froude number profile was decreasing with height according to the increase of the mean static stability and increase of the height of the obstacle. However, by corresponding wind speed and stratification of the atmosphere, the air parcels would be able to rise along the slopes of the mountains even to the heights of 3000 m above the sea level. Such conditions are generally favorable for creating of mountain waves of large amplitudes at the lee side of the mountains and for the onset of downslope windstorms (*Poulos et al.*, 2000).

4. Impact of the windstorm in Central Europe

The majority of the information related to the storm impact in Central Europe comes from SYNOP reports. Smaller number of reports originate from climatological and automatic stations (Czech Republic, Slovakia, and Hungary) or from wind measurements provided by non-meteorological institutions (High Tatras National Park — TANAP). Occasionally, it was possible to get additional information concerning damage observations (e.g., from meteorological institutes, forestry departments, and the state department of nature protection of the Slovak Republic — ŠOPSR). An overview of wind maxima and damage observations is given by *Table 1*. Below we concentrate on the impact of the windstorm in Slovakia and Hungary, where more detailed information were available to the authors.

Table 1. Overview of wind gusts reported by synoptic (S) and climatological (C) stations in Central Europe. Abbreviations D, A, CZ, P, SK, HU, and UA are related to Germany, Austria, Czech Republic, Poland, Slovak Republic, Hungary, and Ukraine. Data in the fourth column, represented by synoptic stations, are the time of the hourly reported maximum wind gusts as in SYNOP report. By certain synoptic and climatological stations these are the exact times of the occurrence in UTC time

Station	Altitude (m)	Wind gust (m/s)	Time (UTC)	Damage observation
Wendelstein, D (S)	1835	50	7:00	D: not available
Grosser Arber, D (S)	1446	44	8:00	
Alpinzentrum, D (S)	2310	45	10:00	
Linz Hoersching, A (S)	298	30	9:00	A: reported near
Wiener Neustadt, A (S)	280	34	15:00	Salzburg
Kocelovice, CZ (S,C)	519	32	08:50	CZ: near Brno,
Dukovany, CZ (S)	400	34	12:00	1 casualty
Raciborz, P (S)	206	31	12:00	P: not available
Nowy Sacz, P (S)	292	31	18:00	
Poprad letisko, SK (S,C)	694	34	13:21	SK: northern
Stará Lesná, SK (C)	810	45	16:28	Slovakia,
Skalnaté Pleso, SK (C)	1778	54	17:54	2 casualties
Siófok, HU (S)	108	29	12:30	HU: reported in
Balatonfüred, HU (S)	108	31.1	13:00	middle Hungary
Kab hegy, HU (S)	600	38.7	11:50	
Plaj,UA	1300	40	not available	not available

The majority of the stations in Slovakia measured the maximum wind speed during or shortly after the passage of the cold front between 11:00 and 16:00 UTC. Several observations of wind gusts with speed higher than 30 m/s come from northern Slovakia (Poprad, Liesek, Telgárt) or from southwest Slovakia (Piešťany, Nitra). The intensity and severity of the wind was

decreasing towards southeast. Heavy damages in the forest occurred in several mountain regions of the northern and central parts of Slovakia, e.g., in regions of Orava, Kysuce, High and Low Tatras, or in the regions of Muránska Planina and Poľana (web site of ŠOPSR department — www.sopsr.sk). Two casualties were reported during the event at the High and Low Tatras, where the wind reached the most severe intensity. Gusty wind exceeding 40 m/s and 10 minute average wind speed higher than 24 m/s were recorded at the top of the Tatra Mountains between 14:00 and 21:00 UTC (synoptic stations Lomnický Štít and Chopok). Strengthening of the wind at the southern (lee) slopes and foots of the High Tatras was observed after the passage of the cold front and reached the maximum intensity between 15:00 and 19:00 UTC, during the advection of cold air at the rear side of the November 19 cyclone (records of the meteorological stations Stará Lesná and Skalnaté Pleso belonging to Geophysical Institute of Slovak Academy of Sciences).

The cold front arrived to northwestern Hungary at 11:00 UTC. At first the station at Mosonmagyaróvár registered the turn of the wind from south-westerly to westerly and northwesterly direction. After 12:00 UTC the front propagating by mean speed of 20 m/s reached the area of lake Balaton, where in Siófok 29 m/s, at Balatonfüred 31.1 m/s peak wind gusts were measured, and at the highest point of the Bakony Mountains (Kab hegy), 38.7 m/s was recorded by the automatic station. Thunderstorms formed on the edge of the cold front and moved southwesterly, approximately parallel to the line Komárom–Budapest. Some funnel clouds and hails (area of Budapest) were observed. Due to the thunderstorms, the distribution of the wind gusts became more chaotic. Nevertheless, the highest magnitudes of the wind speed were measured either in the central and right part of the Danube basin, or in mountain regions. In the eastern part of Hungary the impact of the windstorm was considerably weaker.

5. Impact of the windstorm at the High Tatras

The forest at High Tatras was reported to be destroyed on area of 120 km^2 in a belt of 40–50 km long and 2–5 km wide at the altitudes between 750 and 1200 m (*WWF position paper*, 2004, materials of ŠOP SR, and personal communication with Stanislav Celer from ŠOP SR). In the past, at least 17 major windstorms were referenced for the High Tatras between the years 1898 and 2000 (*Koreň*, 2005; *Luczy*, 2005). Wind of severe intensity (the mean hourly wind speed exceeding 21 m/s) is not unique at high altitudes of the High and Low Tatras, where it occurs with 5‰ frequency (what means two-three days in a year in average). Orographical strengthening and high gustiness

of the wind are frequently observed at the lee sides of mountain passes. Climatological studies for years 1951–1960 showed, that at some places (e.g., Skalnaté Pleso, 1778 m above sea level) daily maxima of wind gusts overlapping the speed of 105 km/h (29.2 m/s) can have an occurrence frequency of 20% (*Otruba*, 1964; *Otruba* and *Wiszniewski*, 1974). Rarer are cases, when the windstorm reaches the altitudes of forest between 700 and 1200 meters, as it happened during the event of November 19, 2004. Windstorms in the years 1915, 1919, 1941 (*Končekk*, 1944), 1964 (*Rak*, 1967), and 1981 (*Koreň*, 2005) had large destructive consequences (although smaller in comparison with the 2004 case). Wind blowing from northern direction and accelerating along the slopes of the Tatras was considered as a possible reason for these extreme events.

The start of the November 19 windstorm at the High Tatras according to observations (Stará Lesná station) can be placed after 14:20 UTC. It was marked by sudden increase of both average wind speed and speed of the wind gusts, and by turn of the wind from westerly to northwesterly direction. The north and northwestern flanks of the mountains were windward and the south and southeastern flanks leeward according to the mean low and middle tropospheric flow. The gusty character of the wind is underlined by the big difference between the hourly mean wind speed (almost 20 m/s) and maximum measured wind gusts (45 m/s at Stará Lesná at 16:28 UTC and 54 m/s at Skalnaté Pleso at 17:54 UTC). Wind gusts were accompanied by fast oscillations of the atmospheric pressure of order 3 hPa. The speed of the wind gusts in the Poprad Valley remained mostly below 30 m/s during the windstorm at the High Tatras (station Poprad airport and station Poprad Gánovce). The wind speed was even continuously decreasing between 15:00 and 17:00 UTC, because the Poprad Valley region was not directly affected by the windstorm. The end of the event at the High and Low Tatras can be situated to 21:00 UTC, although at lower altitudes the speed of the wind gusts was below 30 m/s already after 17:30 UTC (station Stará Lesná). After 21:00 UTC the speed of the gusts decreased below 30 m/s everywhere, and the wind has lost its severity.

6. Numerical weather prediction tools

Horizontal resolution of the limited area hydrostatic model ALADIN, which run operationally at SHMÚ on November 19, 2004, was 9 km, and the model used 37 vertical levels. The driving model was the global model ARP GE with 3-hour coupling frequency. The model was initialized with digital filter (see more details about ALADIN in *Radnóti et al.*, 1995; *Horányi et al.*, 1996; *Gerard*, 2000; *Derková*, 2005). After the event, the test with 2.5 km

horizontal resolution was carried out with the hydrostatic version of the model (further denoted as ALADIN HS). It used the physical parameterization from the reference operational model version with convection parameterization switched off. The ALADIN HS run was based at November 19, 2004, 00:00 UTC and was integrated for 36 hours. The initial and boundary conditions for ALADIN HS were taken from the operational run with 1 hour coupling frequency.

The second experimental run used the non-hydrostatic ALADIN dynamics (further ALADIN NH) and several packages of physical parameterization of the ALADIN model. The dynamics of the model is using the two time levels semi-Lagrangian advection with iterative centered implicit scheme and mass based terrain following coordinate (*Laprise,* 1992; *Bubnová et al.*, 1995; *Bénard et al.*, 2004, 2005). In the case of the ALADIN NH run, the integraton period and the coupling strategy were the same as in the ALADIN HS experiment.

Another test was performed with the so-called dynamical adaptation approach at 2.5 km horizontal resolution. It was expected, that even the simplified and computationally effective hydrostatic simulation at short time ranges will keep a good representation of interaction of atmospheric flow with the terrain. The main difference against the ALADIN HS run was in the length of the integration and way of the treatment of boundary and initial conditions. The dynamical adaptation run for 30 minutes and 60 minutes, respectively, using 15 vertical levels only, adjacent to the model surface. The model integration was provided with 1 hour frequency starting from initial conditions provided by operational model. The 30-minute integration with time constant coupling is time inconsistent, because its final output is approximated as the state of the atmosphere at the beginning of the integration (*Fig. 6a*). The 60-minute integration is time consistent, hence, providing better estimates for minimal and maximal values in the given period (*Fig. 6b*).

The operational service at the Hungarian Meteorological Service was using the ECMWF global model, which was the source of several macrosynoptic analyses and forecasts applied in the study of the synoptic situation described in Section 2. Further, outputs of the ALADIN/HU model at 6.5 km and MM5 model forecasts at 4 km horizontal resolution were available. For the November 19 case study, a nested grid version of the MM5 model at 1 km resolution was created, which covered the area of the High Tatras and received initial and boundary conditions from MM5 run at larger domain with 3 km resolution, which received its boundary conditions from ECMWF deterministic model. The MM5 model did not use cumulus parameterization. Both domains applied ETA PBL scheme (*Janjic*, 1990) and Reisner graupel moisture schemes (*Reisner et al.*, 1998). The ETA PBL scheme is a detailed PBL method, which calculates turbulent kinetic energy

and vertical fluxes with an implicit diffusion scheme. The Reisner graupel method is based on mixed-phase scheme, where snow, ice, water, water vapor, and graupel are also calculated. The model used 26 vertical sigma levels; the upper pressure level was 100 hPa.

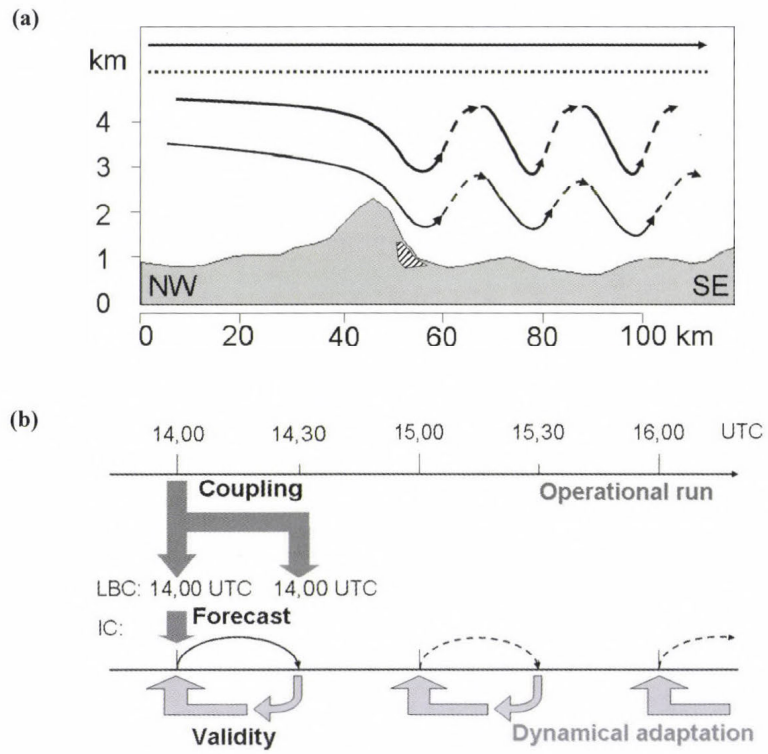


Fig. 6. Scheme of the dynamical adaptation procedure for: (*a*) time inconsistent and (*b*) time consistent approaches. LBC are Lateral Boundary Conditions, which are constant in time in (*a*). IC is abbreviation for Initial Condition. The scheme of coupling (dark grey arrows) is figured out only for the 14:00 UTC integration (solid line), the following integrations are depicted only schematically (by dashed lines). Light grey arrows point at the validity of the 30-minute forecast in (*a*) and of the 1 hour integration in (*b*).

7. Tests with the ALADIN model at 9 km resolution

Synoptic scale forecasts of the November 19, 2004 cyclone development, mean sea level pressure and wind distribution were provided by models ECMWF, operational ARPEGE/ALADIN, and MM5 with very high precision. High speed instantaneous winds and wind gusts exceeding 30 m/s were

forecasted on large territory by the ALADIN model more than 36 hours in advance (by the November 18, 2004, 00:00 UTC run). The reference ALADIN SHMÚ run in this study was based on November 19, 2004, 00:00 UTC, and it forecasted peak wind gusts of 40 m/s with maximum situated in the region of the High Tatras. Time course of the values forecasted and interpolated to the locality of Lomnický štít (2633 m above sea level) were compared with records of the meteorological station (*Fig. 7*). The start of the event was predicted with small time shift (1 hour) for the operational run, but the overall tendency fitted the observations very well. The horizontal distribution of the wind field was less precise -the magnitude of the wind was overestimated in the Poprad valley and underestimated in certain regions, e.g., in the Low Tatras (*Fig. 8*). These results gave forecasters a good possibility to estimate severe wind occurrence. However, at 9 km resolution it was not possible to sufficiently localize the position of the event. The cross Sections of vertical velocity and potential temperature (*Fig. 9*) show a moderate effect of downslope wind at southeastern flank of the High Tatras. It is associated with upward motions at windward side and downward motions at the lee side of the mountains.

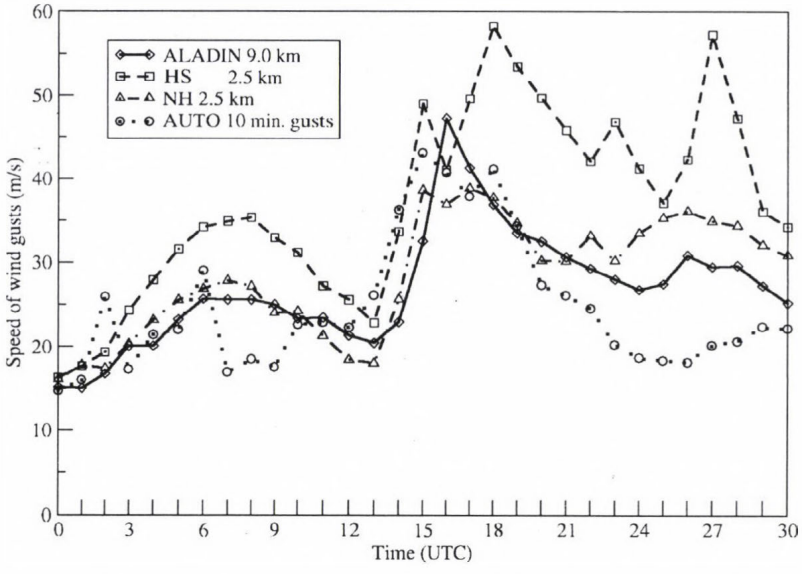


Fig. 7. Time evolution of the ALADIN model forecast of wind gusts from operational run (solid line) and experimental run HS (dashed line) for the Lomnický štít station (2633 m above sea level). Dash-dotted line (with triangles) shows the forecast of the ALADIN NH non-hydrostatic run at 2.5 km resolution. Dotted line (with circles) is the speed of maximum wind gust recorded for the previous 10 minutes by the automatic station.

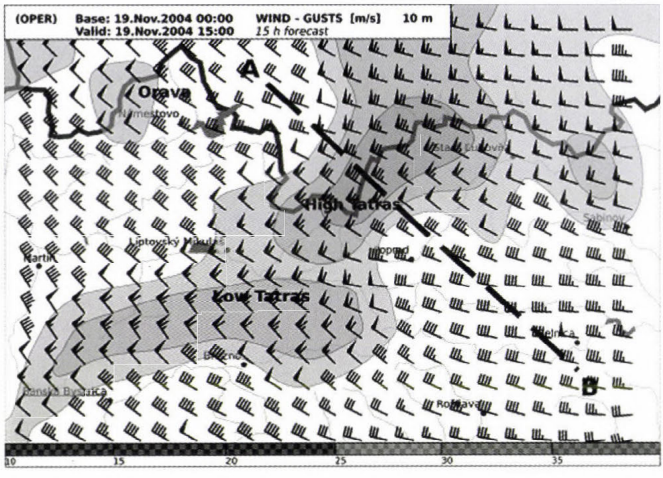


Fig. 8. Operational forecast of the ALADIN SHMÚ model of the wind gust field at the 10 m height over the model surface (wind barbs) and of the wind gust magnitude (grey scale), based on November 19, 2004, 00:00 UTC and valid for November 19, 2004, 15:00 UTC (m/s). The AB line denotes the direction of the cross-section in Figs. 9 and 12.

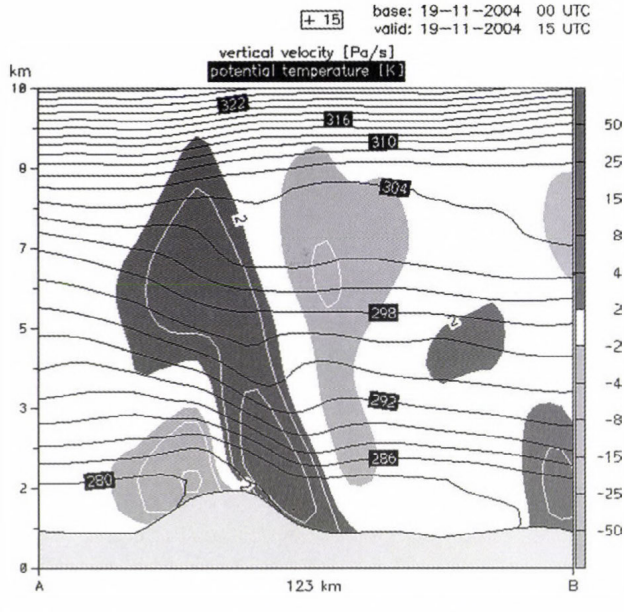


Fig. 9. Vertical cross section through the field of potential temperature (black lines) and vertical velocity in Pa/s (grey scale and white lines) as derived from the forecast of the ALADIN SHMÚ model, based on November 19, 2004, 00:00 UTC and valid for November 19, 2004, 15:00 UTC. The direction of the AB cross section is from north-west to south-east (marked in Fig. 8).

Tests were provided with newer versions of the ALADIN model with different physical parameterization setup (*Geleyn*, 2004). This contains the Xu-Randall cloudiness scheme (*Xu* and *Randall*, 1996), an improved radiation scheme (*Geleyn et al.*, 2005), and gustiness parameterization (*Belluš*, 2002). The parameterization of turbulent fluxes (*Louis et al.*, 1982) and the critical Richardson number (*Geleyn*, 2001) was tuned with respect to situations with stable stratifications and temperature inversions (allowing less turbulent transport of momentum and heat in the stably stratified planetary boundary layer). The distribution of the wind gust field showed decrease of the wind speed at the region of the Low Tatras and higher wind speed in the regions of Orava and southwestern Slovakia (*Simon* and *Vivoda*, 2005). However, the physical mechanism of the parameterization influence on the forecast of the November 19, 2004 windstorm is probably very complex, and the exact origin of the major changes in wind field remained unknown even after several tests with above mentioned setup.

The operational ALADIN model uses the envelope orography, where additional mass is added to model orography to represent variations in the subgrid scale (*Wallace et al.*, 1983). Model runs without the envelope failed in forecasting of the November 19, 2004 windstorm. Therefore, the presence of the envelope in the model with 9 km horizontal resolution is probably still important, despite of some systematic negative effects (e.g., exaggeration of precipitation at windward side of the mountains). Sensitivity of the ALADIN model results on the representation of the sub-grid scale orography was confirmed also for severe bora cases in Croatia (*Drvar et al.*, 2005).

8. Tests with the ALADIN model at 2.5 km resolution

8.1 Hydrostatic high resolution experiment (ALADIN HS)

Resolution of the horizontal grid is very important in the numerical simulation of small scale events. In the operational model configuration at 9 km resolution, the area of the High Tatras is represented only by few grid points. Hence, models with higher resolution should describe better the spatial distribution of the wind and should localize the wind speed maxima with higher precision. Outputs of the hydrostatic run at 2.5 km fulfilled these expectations. The distribution of the wind gust field (*Fig. 10*) coincides with the stationary observations and information from damage survey better than the operational run with lower resolution. Maximum predicted wind gust at southeastern flank of the High Tatras was 50 m/s, which was observed at some locations (e.g., Skalnaté Pleso). The wind distribution for Poprad valley and

the second maximum of wind speed located at the southern slopes of the Low Tatras seem to be realistic as well. Forecasts of the high resolution model show very interesting features in the field of mean sea level pressure and wind at the 10 m height (*Fig. 11*). A mesoscale cyclone forms on the lee side of the High Tatras, and a ridge of high pressure appears on the windward side. This configuration creates significant pressure gradient on the southern slopes of the mountains. The wind in this region is cross-isobaric, hence, the forcing is characterized by strong pressure drag, which is typical for downslope windstorms (*Tutiš* and *Picek*, 1991; *Picek* and *Tutiš*, 1995).

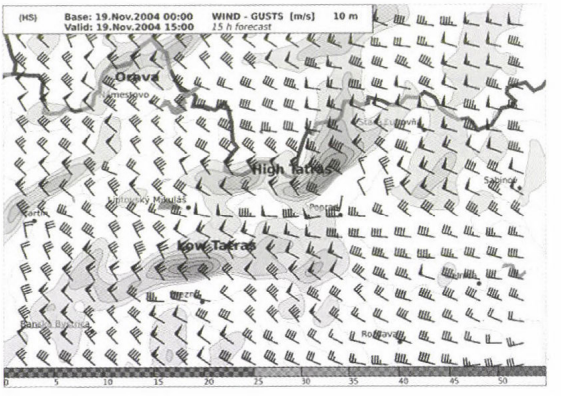


Fig. 10. The same as in Fig. 8 except for the experimental model HS.

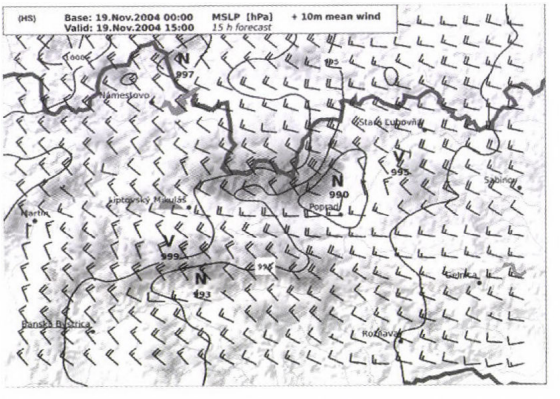


Fig. 11. Forecast of the wind field at the 10 m height over the model surface (wind barbs) and of the mean sea level pressure in hPa (thin black line). The background shows the orography in the region of the High Tatras (centre of the figure) and Low Tatras (southerly from the High Tatras). Border with Poland is marked by thick grey line. Note the mesocyclone at the south-eastern flank of the High Tatras (990 hPa in the centre of the low).

Vertical cross-sections of wind and potential temperature field in direction perpendicular to the ridge of the High Tatras show other evidence about the presence of an intense downslope wind. Areas with wind speed exceeding 30 m/s are located in the close vicinity of the lee slopes of the mountains. Analysis of potential temperature shows a large wave-shape deformation of isotherms in this region (*Fig. 12a*). If we expect an adiabatic motion of the air parcel following the surfaces of constant potential temperature, intense downward vertical motions are on the lee side of the mountain. This is accompanied by shrinking of the vertical air column, hence, with large increase of the static stability. Further downstream, the parcel is forced to rise, what is related to decrease of static stability. Diagnostics of the vertical velocity field indicates downward motions faster than 50 Pa/s (~5 m/s) and upward motions of velocity higher than 25 Pa/s (~2.5 m/s).The reader should note that the vertical velocities expressed in m/s units are only informative values roughly estimated from the hydrostatic equation.

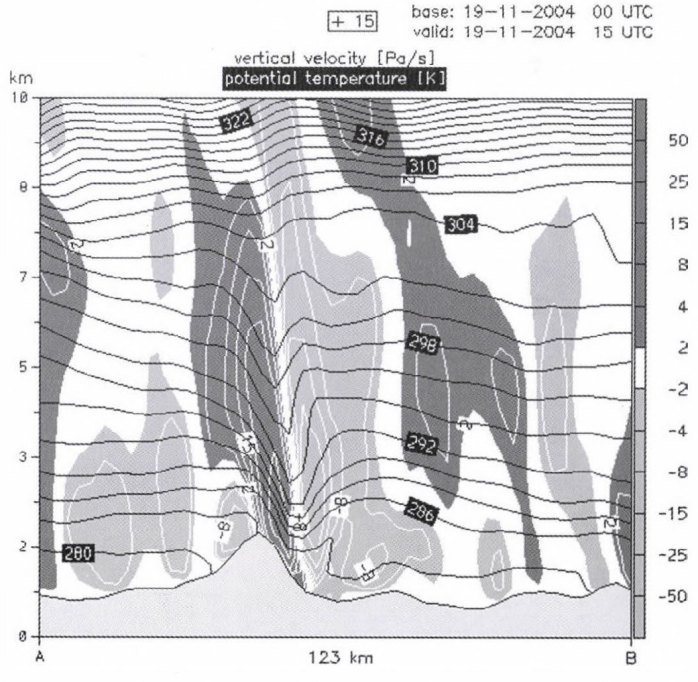


Fig. 12a. Vertical cross section through the field of potential temperature (black lines) and vertical velocity in Pa/s (grey scale, white lines) derived from the forecast of the HS experimental model based on November 19, 2004, 00:00 UTC and valid for November 19, 2004, 15:00 UTC. The sense of the AB cross section is from north-west to south-east (marked in Fig. 8).

However, estimated values are much in excess of usual vertical wind speed in non-convective environment (in order of 0.1–1 Pa/s), and it is similar to the so called hydraulic jump effect (see, e.g., *Holton*, 1992 and *Stull*, 2003) occuring by strong flow over an obstacle.

8.2 Non-hydrostatic experiment (ALADIN NH)

Non-hydrostatic simulation of the November 19 windstorm showed qualitatively similar results to the ALADIN HS experiment (*Fig. 12b*). Small differences between the two runs are visible in the shape, tilt, and position of the generated mountain wave. In the case of the ALADIN NH model, the wave is more symmetric; its axis is almost vertical and shifted little bit downstream against the hydrostatic output. The maximum predicted wind gusts were not as high as in the hydrostatic integration (45 m/s at southeastern flank of the High Tatras; *Fig. 13*).

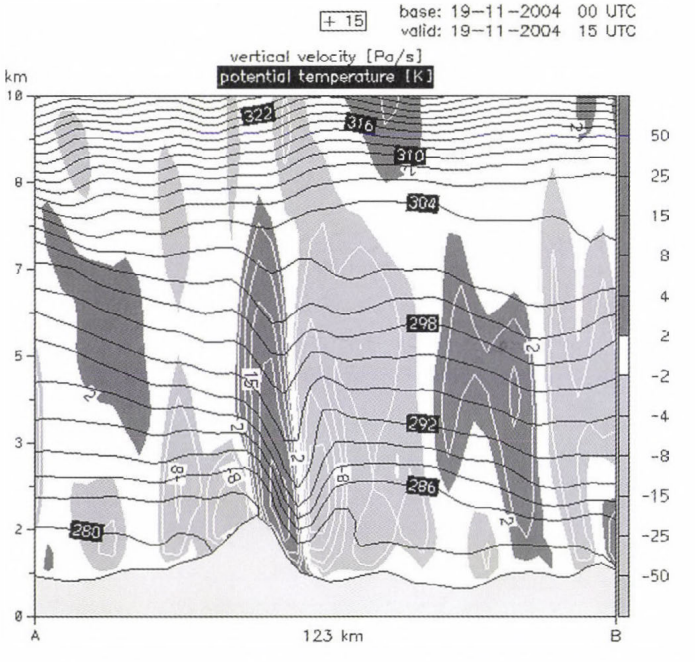


Fig. 12b. The same as in Fig. 12a except for the non-hydrostatic model.

Test with newer setup of physical parameterization (similar to the one referenced in Section 7) showed almost three hours delay of the windstorm onset comparing to the reference operational run. This leaded to differences in

the wind field distribution of the 15-hour forecast and to weaker wind gusts (maximum speed of 35 m/s). The forecast valid for 18:00 UTC already showed results similar to 2.5 km hydrostatic run and wind gusts up to 50 m/s. However, the high resolution ALADIN HS and ALADIN NH runs with physical parameterization of the reference ALADIN SHMÚ operational model corresponded much better with the observations. The possibility of the time shift of the windstorm forecast due to physical parameterization seems to be a very interesting feature, although the model physics and model dynamics interaction in this case is not yet understood.

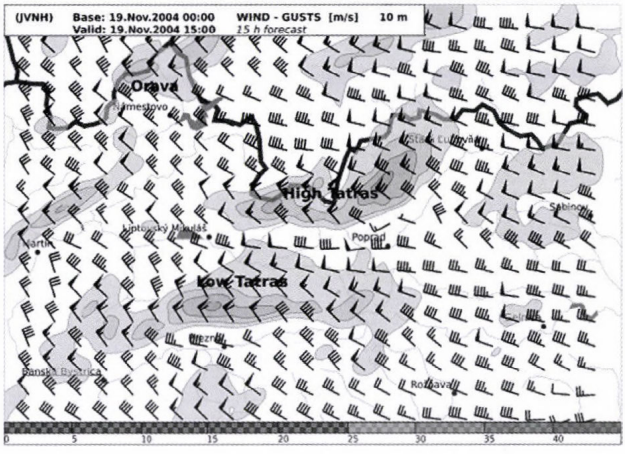


Fig. 13. The same as in Fig. 8 except for the non-hydrostatic model.

8.3 Dynamical adaptation with the ALADIN model at resolution 2.5 km

The dynamical adaptation forecasted the strongest wind gusts (58 m/s) from all high resolution runs (*Fig. 14*). The distribution of this parameter equals to the output of the ALADIN HS experiment. Despite of the time inconsistency, the dynamical adaptation using 30-minute integration valid at 15:00 UTC described the situation better than the time consistent version with one hour integration valid for the same time. In the latter case, the start of the windstorm was predicted with one hour delay. This experience suggests that the dynamical adaptation is very dependent on the initial and boundary conditions obtained from the reference run. One hour integration using dynamical adaptation is probably not sufficient for the onset of the downslope windstorm, if proper mesosynoptic conditions were not established in the initial input file taken from the reference run.

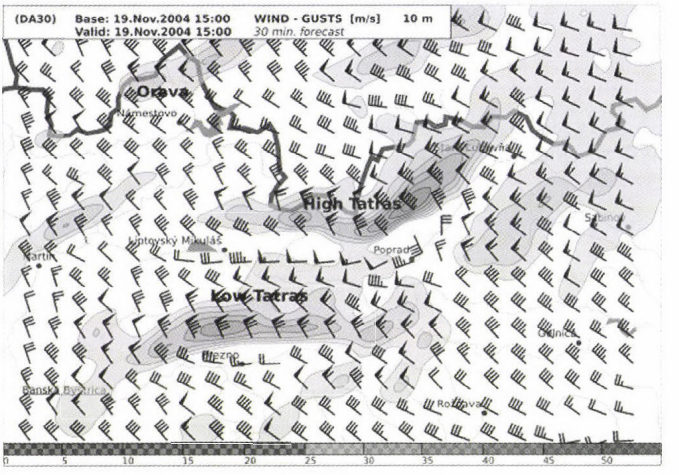


Fig. 14. The same as in Fig. 8 except for the time inconsistent dynamical adaptation.

9. Tests with the MM5 model at 1 km resolution

The start of the windstorm was forecasted by the fine scale MM5 model at 15:00 UTC, and it appeared as turn of the wind direction from westerly to norhwesterly direction and as increase of the instantaneous wind speed at the 10 and 20 m height above terrain, at the top of the mountains. The downslope character of the wind is distinct at 16:00 UTC, where local maxima of the wind speed higher than 25 m/s can be found both at mountain crests and southeastern foots of the mountains at the 900 m altitude (*Fig. 15*). Between these two maxima, there is a belt of lower wind speed between altitudes 1100 and 1600 m, which corresponds to the information about the character of the damage in this part of the High Tatras (personal communication with Milan Koreň from TANAP organization). Projection of very high 10 m wind values is situated at the 1400 m altitude, in the axis of the valley, at the central part of the High Tatras (bottom left). Amplification of the wind in the valley can be seen also at the 20 m height above the model surface. It indicates that the flow could be amplified or suppressed due to microscale orographic conditions. However, the damage to the forest at the central part of the High Tatras was less severe and its distribution was rather chaotic, so it is difficult to verify the validity of the model result without reliable observations. Nevertheless, the main impact of the wind was observed at the southeastern part of the High Tatras, where the creation of gap winds in valleys was either impossible or it had only minor influence on the total damage. The 10 and 20 m wind in the MM5 simulation gets again weaker after 17:00 UTC. A lee cyclone occurs in

the MM5 field of mean sea level pressure at 14:00 UTC (*Fig. 16*). The depth of the cyclone (990 hPa), its position and its dimensions are almost identical with the outputs of the hydrostatic and non-hydrostatic ALADIN runs. It is noticeable, that the pressure gradient (and pressure drag) increased in time as the cold advection from northwest was getting stronger, but the low level flow was still blocked by the ridge of the Tatra Mountains.

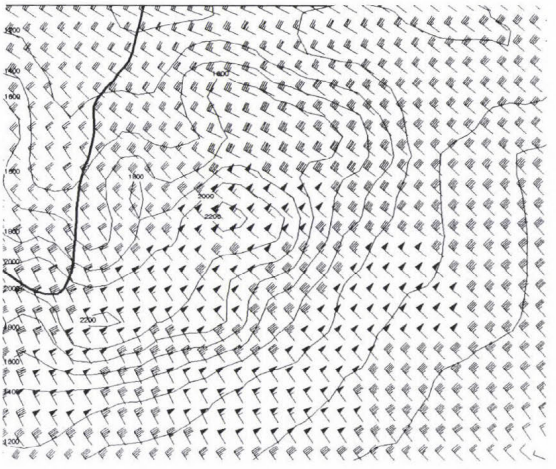


Fig. 15. Forecast of the wind field at the 20 m height over model surface with the nested version of MM5 at 1.0 km resolution for the area of the High Tatras valid for November 19, 2004, 16:00 UTC. Thin solid lines denote the model orography (m).

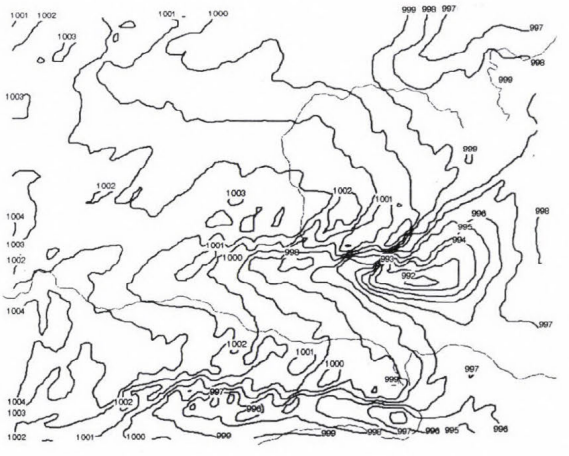


Fig. 16. Forecast of the mean sea level pressure field (hPa) via the nested version of MM5 at 1.0 km resolution for the area of the High Tatras valid for November 19, 2004, 16:00 UTC. Note the presence of the lee cyclone at the southeastern flank of the High Tatras, similarly to Fig. 11.

Finally, vertical cross-sections were provided in the northwest – southeast direction, nearly identical to those used by the ALADIN high resolution runs but valid to 16:00 UTC. The cross-section through the field of the potential temperature and meridional wind component shows undoubtedly an effect of severe downslope wind (*Fig. 17*). An isolated maximum of the wind speed of 35 m/s is situated over the southeastern slopes of the High Tatras at altitudes between 800 and 900 hPa. The cross-Section through the field of vertical velocities shows a well expressed hydraulic jump with local minimum –9 m/s in the downslope and maximum 4 m/s in the upslope part of the wave (*Fig. 18*).

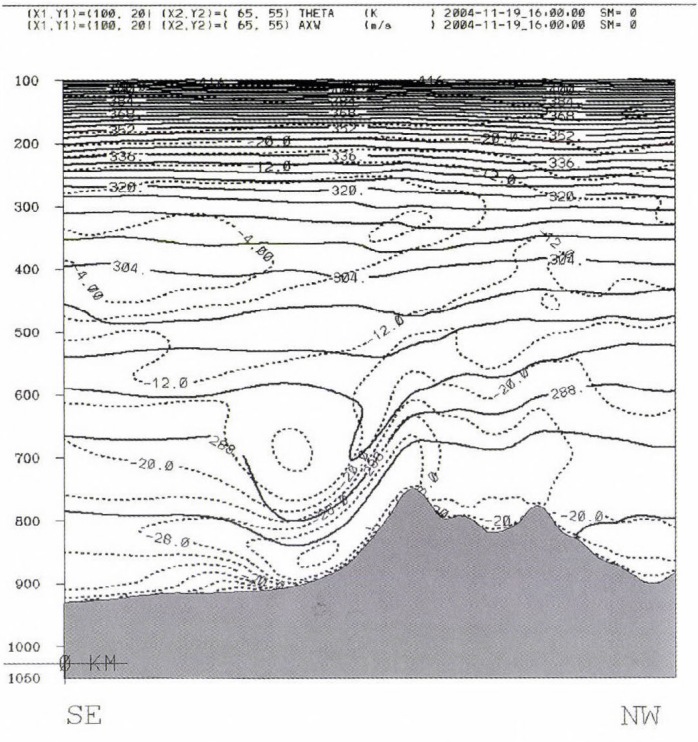


Fig. 17. Vertical cross section through the field of potential temperature (in K, solid lines) and the zonal component of wind (m/s, dotted lines) from the nested MM5 model forecast, valid for November 19, 2004, 16:00 UTC. The sense of the cross section is from northwest (right) to southeast (left), similarly to the AB line denoted by Fig. 8.

However, the fields are different from the results of the ALADIN high resolution simulations. The wave in the potential temperature field is less deep and it is significantly tilted north-westward, which indicates reversal of the flow

in the upslope part of the wave similarly to the situation depicted in *Fig. 4c*. Differences between the ALADIN and MM5 simulations are visible also in the field of vertical velocities. Strong upward motions in the upslope part of the wave are bounded by the 700 hPa level in the MM5 run, while in the ALADIN outputs these are observed until the tropopause and their maximum are situated at higher altitudes. According to the field of wind speed, the effect of downslope windstorm is much stronger by the MM5 than by the non-hydrostatic ALADIN simulation at 2.5 km resolution.

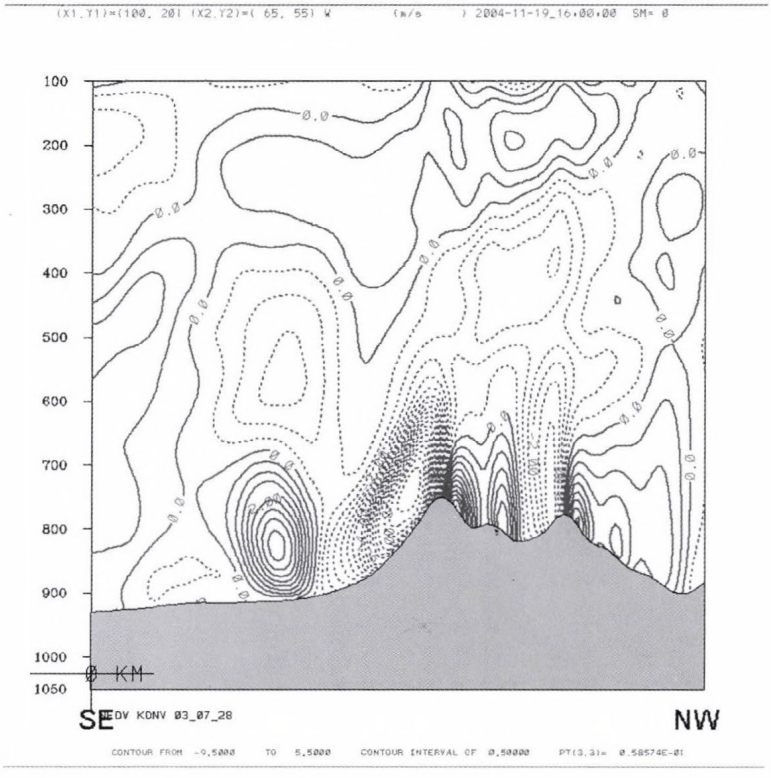


Fig. 18. The same as in Fig. 17 except for vertical velocities (m/s). Upward vertical motions are marked by solid and downward vertical motions by dashed lines.

These differences cannot be appointed to the higher resolution of the MM5 model only, because similar traits of the potential temperature field (e.g., the backward tilt of the lee wave) can be also found in the MM5 run with 4 km resolution. In the MM5 simulation one can observe a different mode and structure of downslope windstorm, related to the above mentioned tilting

(overturning) of the lee wave. It creates statically unstable areas and can be considered as a signature of mountain wave breaking, which typically occurs in simulations of intense downslope wind. This phenomenon is supposed to be responsible for amplification of the intensity of downslope windstorm, because the energy of the flow is reflected from the overturned layer and remains trapped in the layer between the ground and reversed flow (*Peltier* and *Clark*, 1979; *Laprise* and *Peltier*, 1989a).

10. Conclusion

Results of the study presented in this paper imply several consequences for future development and operational suite of numerical weather prediction. It was shown, that models with sufficiently high resolution are able to forecast extreme mesoscale event as the November 19, 2004 windstorm at the High and Low Tatras. Further, hydrostatic approximation seems to be sufficient to simulate substantial effects of intense downslope winds. It turned out, that the dynamical adaptation of the ALADIN model already provides satisfying outputs and better horizontal distribution of the wind field at 2.5 km resolution, in comparison to the reference run at lower resolution. This is important for the operational suite, because the dynamical adaptation is computationally more effective than conventional integration of the model with full number of vertical levels.

Similarity of the hydrostatic and non-hydrostatic solutions in general is not as much surprising, because the horizontal length of the simulated mountain wave (about 30 km) is big enough to be explicitly resolved by hydrostatic dynamics, and the influence of the moist convection is minor during the event. On the other hand, strong vertical motions in the downslope and upslope part of the wave lead to not negligible non-hydrostatic pressure departures, which cannot be simulated by a hydrostatic model. Hence, further increase of the resolution by hydrostatic models would probably not give better forecast of the event. Although the structure of the potential temperature and vertical velocity fields roughly agree with the concept of the downslope windstorm as described by *Smith* (1985) and *Durran* (1986), the hydrostatic approximation tends to exaggerate the effect of the mountain waves breaking (*Laprise* and *Peltier*, 1989b). Realistic simulations of the downslope windstorm at resolutions smaller than 2.5 km most probably require models with non-hydrostatic dynamics. This tendency is indicated also by the outputs of the ALADIN non-hydrostatic model.

Integration of the MM5 model at 1 km resolution gave even more detailed wind distribution close to the surface, which seems to agree with the observed

impact of the windstorm at the most affected southeastern slopes of the High Tatras. Further it shows, that at resolution 1 km and less, it would be already possible to simulate some effects of the microscale, which are given by particularities of the local orography (e.g., orientation of the valleys and peaks). However, verification of such effects is currently very problematic and it would need higher density of meteorological observations and, additionally, very detailed information from the damage survey (e.g., comparison with aerial photographs).

Results of the high resolution runs are dependent on the performance of the model giving the initial and boundary conditions. It was shown, that the success of the ALADIN reference run was linked with optimal setup of physical parameterization used by the operational model version on November 19, 2004. Different schemes of horizontal diffusion or the parameterization of turbulence might have considerable influence on both low and high resolution simulations. Hence, the case of November 19, 2004 should be used to test the future versions of the model dynamics and physics to keep the performance of forecasting windstorms of these types.

From the forecasters point of view, it is important that all model forecasts with high resolution showed typical features of a downslope windstorm (e.g., the hydraulic jump). Forecasters in Slovakia and Hungary do not have many experiences with extreme events similar to the November 19, 2004 windstorm, because downslope wind of this intensity occurs rarely in these countries. The density of available surface observations is also not sufficient to get complete information from scales less than 10 km in horizontal. Hence, forecasters are forced to accept and rely on the model results, which is easier if the dynamical reasons of the windstorm are better understood.

It can be concluded, that high resolution models are necessary to provide early warnings on extreme weather events with sufficient precision. Because of computational limitations, models with hydrostatic dynamics will be still used in the near future (particularly at large domains and by resolutions bigger than 2.5 km). However, realistic simulations of windstorms in general, will require further development of non-hydrostatic models.

References

Adamson, D.S., Belcher, S.E., Hoskins, B.J., and *Plant, R.S.*, 2006: Boundary layer friction in mid-latitude cyclones. *Q. J. Roy. Meteor.* Soc. *614*, 101-124.

Águstsson, H. and *Ólafsson, H.*, 2005: The Freysnes windstorm. The 28th International Conference on Alpine Meteorology. *Croatian Meteorol. J. 40*, 517-520.

Arreola, J.L., Homar, V., Romero, R., Ramis, C., and *Alonso, S.*, 2003: Multiscale numerical study of the 10-12 November 2001 strong cyclogenesis event in the western Mediterranean.

Proc. of the Plinius Conference on Mediterranean Storms IV. European Geophysical Society, Alcudia (Spain), 2-4 October 2002, 5 pp.

Belluš, M., 2002: Enhancement of the momentum and heat fluxes in ARPEGE/ALADIN by the gustiness parameterization. *ALADIN Internal Report*, 18 pp.

Belušić, D., Klaić, Z.B., 2004: Estimation of bora wind gusts using a limited area model. *Tellus A 56*, 296.

Bénard, P., Laprise, R., Vivoda, J., and *Smolíková, 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., Smolíková, P., 2005: Stability of leap-frog constant-coefficients semi-implicit schemes for the fully elastic system of Euler equations: Case with orography. *Mon. Weather Rev. 133*, 1065–1075.

Bennets, D.A. and *Hoskins, B.J.*, 1979: Conditional symmetric instability - a possible explanation for frontal rainbands. *Q. J. Roy. Meteor. Soc. 105*, 945-962.

Bluestein, H.B., 1993: *Synoptic-Dynamic Meteorology in Midlatitudes, Vol. 2: Observations and Theory of Weather Systems*. Oxford University Press, 594 pp.

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

Clark, T.L. and *Peltier, W.R.*, 1977: On the evolution and stability of finite-amplitude mountain waves. *J. Atmos. Sci. 34*, 1715-1730.

Colle, B.A. and *Mass, C.F.*, 1998a: Windstorms along the western side of the Washington Cascade Mountains. Part I: A high resolution observational and modelling study of the 12 February 1995 event. *Mon. Weather Rev. 126*, 28-52.

Colle, B.A. and *Mass, C.F.*, 1998b: Windstorms along the western side of the Washington Cascade Mountains. Part II: Characteristics of past events and three-dimensional idealized simulations. *Mon. Weather Rev., 126*, 53-71.

Colle, B.A. and *Mass, C.F.*, 2000: High-resolution observations and numerical simulations of Easterly Gap Flow through the Strait of Juan de Fuca on 9-10 December 1995. *Mon. Weather Rev. 128*, 2398-2422.

Davis, C.A. and *Emanuel, K.A.*, 1991: Potential vorticity diagnostics of cyclogenesis. *Mon. Weather Rev. 119*, 1929–1953.

Derková, M., 2005: Numerical weather prediction activities at the Slovak Hydrometeorological Institute. *Meteorological Journal 8*, 53-64.

Dierking, C.F., 1998: Effects of a mountain wave windstorm at the surface. *Weather Forecast. 13*, 606-616.

Doyle, J.D. and *Durran, D.R.*, 2002: The dynamics of mountain-wave-induced-rotors. *J. Atmos. Sci. 59*, 186-201.

Doyle, J.D., Bond, N.A., and *Jiang, Q.*, 2006: Observations and high-resolution simulations of a severe turbulence and downslope windstorm event. *AMS 12th Conference on Aviation Range and Aerospace Meteorology*. 28-30 January 2006, abstract.

Drvar, D., Stiperski, I., Tudor, M., and *Tutiš, V.*, 2005: ALADIN/HR: Testing the New Sub-Grid Scale Orography Representation on Bura Cases. The 28th International Conference on Alpine Meteorology. *Croatian Meteorological Journal 40*, 304-307.

Durran, D.R., 1986: Another look at downslope windstorms. Part I: The development of analogs to supercritical flow in infinitely deep continuously stratified fluid. *J. Atmos. Sci. 43*, 2527-2543.

Gaberšek, S., 2003: A downslope windstorm in NW Slovenia on 16 November 2002, *ICAM/MAP2003 Conference*. Brig, Switzerland, 19- 23 May 2003, abstract.

Geleyn, J.-F., 2001: A summary of the latest changes in the parameterization of turbulent fluxes and PBL processes. *ALADIN Newsletter 20*, 84-88.

Geleyn, J.-F., 2004: Some details about ALADIN physics in cycle 28T1. *ALADIN Newsletter 26*, 71-74.

Geleyn, J.-F., Fournier, R., Hello, G., and *Pristov, N.*, 2005: A new 'bracketing' technique for a flexible and economical computation of thermal radiative fluxes, on the basis of the Net Exchange Rate (NER) formalism. *WGNE blue book, 2005 Edition, 4-07.*

Gerard, L., 2000: *Physical Parameterization in ARP G -ALADIN.* Internal ARP G /ALADIN documentation, 111 pp.

Glossary of Meteorology (ed.: *T.S. Glickmann*), 2000: AMS, Boston, Massachusetts, USA, 855 pp.

Gohm, A. and *Mayr, G.J.*, 2005a: On the bora breakthrough near a mountain gap. The 28th International Conference on Alpine Meteorology. *Croatian Meteorol. J. 40*, 217-220.

Gohm, A. and *Mayr, G.J.*, 2005b: Numerical and observational case-study of a deep Adriatic bora. *Q. J. Roy. Meteor. Soc. 131*, 1363-1392.

Grubišić, V. and *Lewis, J.M.*, 2004: Sierra Wave Project revisited 50 years later. *B. Am. Meteorol. Soc. 85*, 1127-1142.

Holton, J.R., 1992: *An Introduction to Dynamic Meteorology.* Third edition. Academic Press, 282–287.

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

Hoskins, B.J., McIntyre, M.E., and *Robertson, A.W.*, 1985: On the use and significance of isentropic potential vorticity maps. *Q. J. Roy. Meteor. Soc. 111*, 877-946.

Ivatek-Šahdan, S. and *Tudor, M.*, 2004: Use of high-resolution dynamical adaptation in operational suite and research impact studies. *Meteorol. Z. 13*, 99-10.

Janjic, Z.I., 1990: The step-mountain coordinate: Physical package. *Mon. Weather Rev. 118*, 1429-1443.

Jurčec, V. and *Brzović, N.*, 1995: The Adriatic bora: special case studies. *Geofizika 12*, 15-32.

Kaplan, M.L, Huffman, A.W., Lux, K.M., Cetola, J.D., Charney, J.J., Riordan, A.J., Lin, Y.L., and *Waight III, K. T.*, 2003: *Characterizing the Severe Turbulence Environments Associated With Commercial Aviation Accidents. Part II: Hydrostatic Mesobeta Scale Numerical Simulations of Supergradient Wind Flow and Streamwise Ageostrophic Frontogenesis.* Langley Research Center, Hampton, Virginia. NASA/CR-2003-212138 Report, 45 pp.

Kasahara, A. and *Qian, J.H.*, 2000: Normal modes of a Global Nonhydrostatic Atmospheric Model. *Mon. Weather Rev. 128*, 3357-3375.

Klaić, Z.B. and *Belušić, D.*, 2005: Severe Adriatic Bura event from 14 to 15 November 2004. The 28th International Conference on Alpine Meteorology. *Croatian Meteorol. J. 40*, 300-303.

Klemp, J.B. and *Durran, D.R.*, 1987: Numerical modelling of bora winds. *Meteorol. Atmos. Phys. 36*, 215-227.

Klemp, J.B. and *Lilly, D.K.*, 1975: The dynamics of wave-induced downslope winds. *J. Atmos. Sci. 32*, 320-339.

Klemp, J.B. and *Lilly, D.K.*, 1978: Numerical simulation of hydrostatic mountain waves. *J. Atmos. Sci. 35*, 78-107.

Konček, M., 1944: Meteorological circumstances of catastrophic windstorm in Slovakia during autumn of 1941 (in Slovak). Prírodovedecká fakulta Slovenskej university, Bratislava, 44 pp.

Koreň, M., 2005: Wind calamity of 19 November 2004 – new points of view and consequences (in Slovak). *Vysoké Tatry special issue 15*, 6-28.

Kraljević, L. and *Tudor, M.*, 2005: High resolution simulation of a severe bura event – an intercomparison of two models. Presentation, *15th ALADIN Workshop*. Bratislava June 6-10 2005 (http://www.cnrm.meteo.fr/aladin/meetings/Wk2005/program.html).

Kuettner, J. and *Hertenstein, R.F.*, 2002: Observations of mountain-induced rotors and related hypotheses: A review. *AMS Tenth Conference on Mountain Meteorology*. Park City Utah, 17-21 June 2002, 326-329.

Lane, T.P., Reeder, M.J., Morton, B.R., and *Clark, T.L.*, 2000: Observations and numerical modelling of mountain waves over the Southern Alps of New Zealand. *Q. J. Roy. Meteor. Soc. 126*, 2765-2788.

Laprise, R. and *Peltier, W.R.*, 1989a: The structure and energetics of transient eddies in a numerical simulation of breaking mountain waves. *J. Atmos. Sci. 46*, 565-585.

Laprise, R. and *Peltier, W.R.*, 1989b: On the structural characteristics of steady finite-amplitude mountain waves over Bell-Shaped topography. *J. Atmos. Sci. 46*, 586-595.

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

Lilly, D.K., 1978: A severe downslope windstorm and aircraft turbulence event induced by a mountain wave. *J. Atmos. Sci.* 35, 59-77.

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

Long, R.R., 1954: Some aspects of the flow of stratified fluids: II. Experiments with a two fluid system. *Tellus 6*, 97-115.

Louis, J.F., Tiedke, M., and *Geleyn, J.-F.*, 1982: A short history of the operational PBL parameterization at ECMWF. *Workshop on Planetary Boundary Layer Parameterization.* ECMWF, 25-27 Nov 1981, 59-80.

Luczy, G., 2005: Weather situations and calamities in Tatras, in Slovak. *Vysoké Tatry 3*, 14.

Meyers, M.P., Snook, J.S., Wesley, D.A., and *Poulos, G.S.*, 2003: A Rocky Mountain storm. Part II: The forest blowdown over the west slope of the Northern Colorado Mountains—observations, analysis, and modeling. *Weather Forecast. 18*, 662-674.

Morgan, M.C. and *Nielsen-Gammon, J.W.*, 1998: Using tropopause maps to Ddagnose midlatitude weather systems. *Mon. Weather Rev. 126*, 2555-2579.

Nance, L.B. and *Durran, D.R.*, 1997: A modelling study of nonstationary trapped mountain lee waves. Part I: Mean-flow variability. *J. Atmos. Sci. 54*, 2275-2291.

Nance, L.B. and *Durran, D.R.*, 1998: A modelling study of nonstationary trapped mountain lee waves. Part II: Nonlinearity. *J. Atmos. Sci. 55*, 1429-1445.

Otruba, J., 1964: Wind conditions in Slovakia (in Slovak). *SAV*, Bratislava, 130-134.

Otruba, J. and *Wiszniewski, W.*, 1974: Climate of Tatras (in Slovak). In *Klíma Tatier* (ed.: *M. Konček*). Veda, Bratislava, 328-339.

Pan, F. and *Smith, R.B.*, 1999: Gap winds and wakes: SAR observations and numerical simulations. *J. Atmos. Sci. 56*, 905-923.

Peltier, W.R. and *Clark, T.L.*, 1979: The evolution and stability of finite-amplitude mountain waves. Part II: Surface drag and severe downslope windstorms. *J. Atmos. Sci. 36*, 1498-1529.

Picek, B.I. and *Tutiš, V.*, 1995: Mesoscale Bora flow and mountain pressure drag. *Meteorol. Z. 4*, 119-128.

Picek, B.I. and *Tutiš, V.*, 1996: A case study of a severe Adriatic bora on 28 December 1992. *Tellus 48A*, 357-367.

Poulos, G.S., Bossert, J.E., McKee, T.B., and *Pielke, R.A.*, 2000: The interaction of katabatic flow and mountain waves. Part I: Observations and idealized simulations. *J. Atmos. Sci. 57*, 1919-1936.

Radnóti, G., Ajjaji, R., Bubnová, R., Caian, M., Cordoneanu, E., 1995: The spectral limited area model Arp g -Aladin. In *PWPR Report Series n.7 – WMO TD n. 699*, 111-118.

Rak, J., 1967: The gale in the Tatra region - in November, 1964 (in Slovak). *Sborník prác o Tatranskom národnom parku 10*, 49-64.

Reisner, J., Rasmussen, R.J., and *Bruintjes, R.T.*, 1998: Explicit forecasting of supercooled liquid water in winter storm using the MM5 mesoscale model. *Q. J. Roy. Meteor. Soc. 24B*, 1071-1107.

Romero, R., Homar, V., Ramis, C., and *Alonso, S.*, 2002: Baroclinic and diabatic regulation of the 10-12 November 2001 superstorm in the Balearics. *European Conference on Severe Storms 2002.* Prague, Czech Republic, 26-30 August 2002, presentation and abstract.

Sandvik, A.D., 2001: Case study of a downslope windstorm 16 January 2000. *Workshop on Fine-Scale Meteorological Modelling.* 17-18 September 2001 at Finse.

Satomura, T. and *Bougeault, P.*, 1994: Numerical simulation of lee wave events over the Pyrenees. *J. Meteorol. Soc. Japan 72*, 173-195.

Simon, A. and *Vivoda, J.*, 2005: Downslope windstorm in High Tatras 19 November 2004. Presentation, *15th ALADIN Workshop*. Bratislava, June 6-10 2005, (http://www.cnrm.-meteo.fr/aladin/meetings/Wk2005/program.html).

Smith, C., 2002: Stability analysis and precision aspects of the boundary condition formulation in the non-hydrostatic dynamics and exploration of the alternatives for discrete formulation of the vertical acceleration equation both in Eulerian and semi-Lagrangian time marching schemes. *ALATNET Newsletter 4,* (http://www.cnrm.meteo.fr/aladin/newsletters/news21/-PhDs_6.html).

Smith, R.B., 1980: Linear theory of stratified hydrostatic flow pat an isolated mountain. *Tellus 32*, 348-364.

Smith, R.B., 1985: On severe downslope winds. *J. Atmos. Sci. 42*, 2597-2603.

Smolarkiewicz, P.K. and *Rotunno, R.*, 1989: Low Froude number flow past three-dimensional obstacles. Part I: Baroclinically generated lee vortices. *J. Atmos. Sci. 46*, 1154-1164.

Stull, R.B., 2003: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 601-609.

Tutiš, V., 2002: Violent Adriatic windstorms. *Proc. of the 4th EGS Plinius Conference*. Mallorca, Spain, October 2002.

Tutiš, V. and *Picek, I.B.*, 1991: Pressure drag on the Dinaric Alps during the ALPEX SOP. *Meteorol. Atmos. Phys. 47*, 73-81.

Vergeiner, I. and *Lilly, D.K.*, 1970: The dynamic structure of lee wave flow as obtained from balloon and airplane observations. *Mon. Weather Rev. 98*, 220-232.

Vigh, J., 2005: *Downslope Windstorms and Rotors*. Department of Atmospheric Science, Colorado State University, presentation, (http://euler.atmos.colostate.edu/ ~ vigh/papers).

Vosper, S., 2003: Development and testing of a high resolution mountainwave forecasting system. *Meteorol. Appl. 10*, 75-86.

Wallace, J., Tibaldi, S. and. *Simmons, A.J.*, 1983: Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. *Q. J. Roy. Meteor. Soc. 109*, 683-717.

WWF position paper, 2004: *Storm in Tatras National Park (Slovakia)*. Web page of WWF organisation (http://assets.panda.org/downloads/wwfpositiontatrastorm.pdf).

Xu, K.-M. and. *Randall, D.A.*, 1996: A Semi-empirical cloudiness parametrization 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 51*, 489-504.

Zejda, V., 1986: Where the variometer shows 20 m per second (in Czech). *Letectví a Kosmonautika 33*,193-194.

Zhang, Y., Chen, Y.-L., and *Kodama, K.*, 2005: Validation of the coupled NCEP mesoscale spectral model and an advanced land surface model over the Hawaiian Islands. Part II: A high wind event. *Weather Forecast. 20*, 873-895.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 125–134

Developing an optimal system of circulation pattern types for downscaling purposes

István Matyasovszky

Department of Meteorology, Eötvös Loránd University,
P.O. Box 32, H-1518 Budapest, Hungay, E-mail: matya@ludens.elte.hu

(Manuscript received in final form May 29, 2006)

Abstract—A methodology to develop an optimal system of circulation pattern (CP) types is proposed for downscaling purposes. The success of downscaling depends on how strongly the stochastic behavior of a climate element can be related to CP types. Therefore, an optimal system of CP types has as much information on the specific climate element as possible. The problem includes also the choice of the number of CP types. The procedure proposed is applied to normally distributed daily mean temperatures and daily precipitation amounts having intermittence. In this latter case, CP types may be different for sequences of dry and wet days and for precipitation amounts on wet days.

Key-words: downscaling, circulation pattern, classification, temperature, precipitation

1. Introduction

In order to simulate climates under increasing concentration of atmospheric greenhouse gases, coupled atmosphere-ocean general circulation models (GCMs) are used. A standard method to assess the change of climate is to run such a model under the present atmospheric content of greenhouse gases (control run), and then run it again corresponding to a scenario of future content of greenhouse gases (changed climate). However, due to the relatively low horizontal resolution (a few hundred kilometres) and relatively simple parameterizations (atmosphere-surface feedbacks, radiative processes, cloud and precipitation forming, etc.) of these models, the results for small areas are considerably uncertain. There is a need, therefore, to "downscale" the large-scale output of GCMs to smaller scales.

Generally, two main types of downscaling approaches are used (*Giorgi* and *Mearns*, 1991). The first type of downscaling methods includes meso-scale numerical modeling. Meso-scale models use GCM outputs as initial and boundary conditions. This technique requires substantial effort in terms of modeling and computer programming, and no satisfactory long simulation is available to assess extremes. These difficulties, among others, may motivate the use of the second approach called stochastic downscaling. Stochastic downscaling methods have two key elements. The first element includes large-scale circulation of the atmosphere and the second element represents a linkage between local surface variables and large-scale circulation. The linkage is expressed by a stochastic model using an observational data series. Then, this model may be utilized with GCM outputs characterizing atmospheric circulation.

Stochastic downscaling has several specific versions. One of them is based on classification of the large-scale atmospheric circulation pattern (CP). Having a system of CP types, the stochastic behavior of a local climate element is described via stochastic properties conditioned on these types. A stochastic property Ψ of a local climate element is written as

$$\Psi = \sum_{j=1}^{J} r_j \Psi_j \,, \tag{1}$$

where Ψ_j is the property conditioned on the jth CP type and r_j is the probability of the occurrence of the jth CP type. The property in question under changed climate is then estimated by Eq. (1) with relative frequencies of CP types corresponding to GCM outputs under changed climate. In the simplest case, Ψ is expected value, but may include other parameters characterizing the stochastic behavior of a climate element. In a general case Ψ may be the probability distribution or density function. The literature of application of the concept is very broad, first attempts were reported by *Bardossy* and *Plate* (1992), *Wilson et al.* (1992), *Hughes et al.* (1993).

A basic question concerning the application of the methodology based on CP types is how strongly the stochastic behavior of a climate element can be related to CP types. The stronger this relationship, the larger the estimated local response to the same large-scale change. For instance, in the absence of such a relationship ($\Psi_j \equiv \Psi$) even an extreme change of the large-scale circulation results in no local response. Therefore, a system of CP types having as much information on the specific climate element as possible has to be found. This problem is addressed in Section 2 including the choice of optimal number of CP types, too.

The proposed methodology is applied to daily mean temperature and daily precipitation amount in Section 3. Finally, a section is provided for discussion.

2. Methodology

A large number of initial circulation pattern types is developed using an objective classification method as a first step. The k-means clustering technique is applied and say 100 clusters are developed. These clusters are then aggregated into a fixed number of types such that the types have as large information on the specific climate element as possible. Finally, the optimal number of types is estimated.

Let Y be a normal random variable with expected value m representing a climate element at a fixed site. Let X be a discrete random variable such that $X=k$ corresponds to the occurrence of kth type with $r_k=P(X=k)$. Having an independent sample (x_i,y_i), $i=1,\ldots,n$ for (X,Y) the task is to find the best system of CP types for a fixed number of types K. Having conditional means

$$\hat{m}_k=\frac{1}{n_k}\sum_{i;\,x_i=k} y_i, \quad k=1,\ldots,K, \tag{2}$$

where n_k is the observed number of kth CP type, the quantity

$$\sum_{k=1}^{K} n_k(\hat{m}_k-\hat{m})^2 \tag{3}$$

has to be maximized, where

$$\hat{m}=\frac{1}{n}\sum_{i=1}^{n} y_i \tag{4}$$

is the estimate of expected value $E[Y]$. Then, after *Akaike* (1971), the optimal number of CP types is estimated by minimizing

$$AIC(K)=n\ln\left(\hat{d}^2-\sum_{k=1}^{K}\hat{r}_k(\hat{m}_k-\hat{m})^2\right)+2K \tag{5}$$

with respect to K, where

$$\hat{d}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{m})^2, \quad \hat{r}_k=\frac{n_k}{n}, \quad k=1,\ldots,K. \tag{6}$$

Since daily climate data are statistically dependent, the above formulation, valid for independent sequences, has to be modified. A first order autoregressive approximation to y_i is

$$y_i - m = a(y_{i-1} - m) + e_i, \tag{7}$$

where a is the autoregressive parameter and $\{e_i\}$ is a sequence of independent normally distributed random variables with variance s^2. Having an estimate

$$\hat{a} = \sum_{i=2}^{n} (y_i - \hat{m})(y_{i-1} - \hat{m}) / \sum_{i=2}^{n} (y_{i-1} - \hat{m})^2 , \tag{8}$$

data

$$e_i = (y_i - \hat{m}) - \hat{a}(y_{i-1} - \hat{m}), \quad i = 2,\ldots,n \tag{9}$$

are treated as $\{y_i\}$ in the previous case resulting in

$$AIC(K) = (n-1)\ln\left(\hat{s}^2 - \sum_{k=1}^{K} \hat{r}_k \hat{\mu}_k^2\right) + 2K \tag{10}$$

with

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i;\, x_i = k}^{n} e_i, \quad k = 1,\ldots,K. \tag{11}$$

A climate element in question at L locations is represented by a normal random vector $\boldsymbol{Y}=(Y_1,\ldots,Y_L)$, where superscript T denotes transpose. Then

$$AIC(K) = (n-1)\ln\left(\left|\mathbf{S}_0 - \sum_{k=1}^{K} \hat{r}_k \hat{\mathbf{m}}_k \hat{\mathbf{m}}_k^{\,T}\right|\right) + 2KL \tag{12}$$

has to be minimized for every possible system of CP types and with respect to K, where

$$\hat{\mathbf{m}}_k = \frac{1}{n_k} \sum_{i;\, x_i = k} \mathbf{e}_i, \quad k = 1,\ldots,K, \tag{13}$$

while symbol || represents determinant, and $\mathbf{e}_i$ are defined by

$$(\mathbf{y}_i - \hat{\mathbf{m}}) = \hat{\mathbf{A}}(\mathbf{y}_{i-1} - \hat{\mathbf{m}}) + \mathbf{e}_i, \quad \hat{\mathbf{A}} = \hat{\mathbf{S}}_1 \hat{\mathbf{S}}_0^{-1}, \tag{14}$$

where $\mathbf{S}_0$ and $\mathbf{S}_1$ are covariance matrices of $\boldsymbol{Y}$ for lag 0 and 1, respectively.

The idea discussed above can be used for Gaussian processes, because the present model selection criteria based on maximizing the log-likelihood function penalized with model parameters are valid for such processes. For

many climate elements there exist appropriate transformations resulting in normal distributions, and thus the problem can be treated as in the previous paragraphs. For instance, when wind speed is in question, the random variable Y^c with a suitable $c>0$ is distributed approximately normally due to a property of Weibull distributions used for wind speed. Similar simple transformations can be established for several other climate elements. However, daily precipitation having a crucial role in climate change studies cannot be treated such a simple way due to its spatio-temporal intermittence. Therefore, a two-step procedure is proposed for daily precipitation. In the first step only the sequence of wet and dry days are examined, then precipitation amounts on wet days are analyzed.

The observed sequence of wet and dry days can be modeled with a first order Markov chain (*Matyasovszky* and *Dobi*, 1989). Therefore, an indicator series is defined as $I_i=0$ if $y_i=0$, and as $I_i=1$ if $y_i>0$ for daily precipitation data $\{y_i\}$. The conditional probabilities

$$p_{uv} = P(I_i = v \mid I_{i-1} = u), \quad u,v = 0,\ldots,1 \tag{15}$$

and the probability $P=p_{01}/(p_{01}+p_{10})$ of having a wet day depend on CP types, and the task is to estimate these conditional probabilities with corresponding relative frequencies according to the CP types. Then

$$AIC(K) = n \sum_{k=1}^{K} \hat{r}_k \left[\left(1-\hat{P}^{(k)}\right) \sum_{v=0}^{1} \hat{p}_{0v}^{(k)} \ln \hat{p}_{0v}^{(k)} + \hat{P}^{(k)} \sum_{v=0}^{1} \hat{p}_{1v}^{(k)} \ln \hat{p}_{1v}^{(k)} \right] + 2K \tag{16}$$

has to be minimized for every possible system of CP types, and with respect to K, where superscripts (k) refer to kth CP type.

Positive precipitation data are transformed into normally distributed data. Transformation is done by

$$z_i = \hat{d}\,\Phi^{1}(F_N(y_i \mid y_i > 0)) + \hat{m}, \tag{17}$$

where N is the number of positive precipitation data, F_N is the empirical probability distribution function of these data, Φ is the standard normal distribution function, and m and d are the expected value and standard deviation of these data. Data $\{z_i\}$ can be handled as they come from a time series broken by several non-observable periods corresponding to the sequence of dry days. Time series is modeled with a first order autoregressive process

with the autoregressive parameter a. When a data z_i follows a (j–1)-day length dry period, the optimal prediction for z_i is $a^j(z_{i-j}-m)+m$ having a prediction error variance $(1-a^{2j})d^2$. The log-likelihood function corresponding to prediction errors is then

$$-N\ln d-\sum_{j=1}^{J+1}N_j\ln\left[\left(1-a^{2j}\right)^{1/2}\right]-\frac{1}{2d^2}\sum_{j=1}^{J+1}Q_j(a) \tag{18}$$

with

$$Q_j(a)=\sum_i\left[(z_i-\bar{z})-a^j(z_{i-j}-\bar{z})\right]^2/\left(1-a^{2j}\right),\qquad j=1,\ldots,J+1, \tag{19}$$

where summation extends for i; $y_i>0$, $y_{i-1}=0,\ldots,\ y_{i-j+1}=0,\ y_{i-j}>0$. J is the length of the longest dry period and N_j is the number of (j–1)-day length dry periods. Maximizing Eq. (18) with respect to d gives

$$\hat{d}^2=\frac{1}{N}\sum_{j=1}^{J+1}Q_j(\hat{a}). \tag{20}$$

Substituting this estimate into Eq. (18),

$$-N\ln d^2-\sum_{j=1}^{J+1}N_j\ln\left(1-a^{2j}\right) \tag{21}$$

has to be maximized with respect to a. Also, the maximum likelihood estimate of expectation of prediction errors under kth CP type is

$$\hat{\mu}_k=\left[\sum_{j=1}^{J+1}\bar{e}_j^{(k)}/\left(1-\hat{a}^{2j}\right)\right]/\left[\sum_{j=1}^{J+1}1/\left(1-\hat{a}^{2j}\right)\right], \tag{22}$$

where $\bar{e}_j^{(k)}$ is the mean of the prediction errors following (j–1)-day dry spells under kth CP type. Finally,

$$AIC(K)=\sum_{j=1}^{J+1}N_j\ln\left(\hat{d}^2\left[1-\hat{a}^{2j}\right]-\sum_{k=1}^{K}\hat{r}_j^{(k)}\hat{\mu}_k^2\right)+2K \tag{23}$$

has to be minimized for every possible system of CP types and with respect to K. Now, $\hat{r}_j^{(k)}$ is the relative frequency of kth CP type conditioned on days following (j–1)-day dry spells.

3. *Application*

The methodology discussed above is applied to daily mean temperature and daily precipitation amount. Circulation pattern types are developed using the 700-hPa geopotential field defined on a diamond grid covering the European Atlantic region and obtained from NCAR data bases.

Temperature data for five Hungarian locations, namely Budapest, Kalocsa, Kecskemét, Debrecen, and Szeged cover a 30-year period from 1961 to 1990. Results are discussed for summer and winter. The optimal number of types according to the criterion Eq. (12) is 12 and 10 in summer and winter, respectively. *Tables 1* and *2* show the means of daily mean temperatures in summer and winter for the five locations in a descending order of relative frequencies of CP types. *Figs. 1* and *2* demonstrate differences among probability distributions corresponding to different CP types. Normal density functions fitted to data under different CP types are shown here. Note that a warm and a cold type having large relative frequencies are illustrated in these figures. For instance, Type 8 is the coldest in winter but with a small relative frequency, thus Type 3 having a large relative frequency is presented in *Fig. 2*.

Table 1. Summer mean temperatures under different CP types

Type	Mean temperatures (°C)				
Relative frequency	Budapest	Kalocsa	Kecskemét	Debrecen	Szeged
0.155	21.79	21.31	21.34	20.71	21.50
0.124	20.14	19.51	19.41	18.75	19.61
0.118	18.48	18.15	18.28	17.62	18.22
0.099	20.38	20.01	19.95	19.22	20.11
0.098	21.13	20.63	20.47	19.92	20.69
0.093	22.49	22.00	21.92	21.33	22.09
0.076	22.33	21.37	21.39	21.09	21.55
0.076	18.59	18.22	18.19	17.90	18.44
0.055	22.46	21.61	21.48	20.92	21.76
0.043	20.38	19.73	19.75	19.09	20.08
0.036	17.94	17.38	17.30	17.26	17.66
0.027	20.86	20.27	20.25	19.44	20.38

Table 2. Winter mean temperatures under different CP types

Type	Mean temperatures (°C)				
Relative frequency	Budapest	Kalocsa	Kecskemét	Debrecen	Szeged
0.162	0.31	−0.83	−1.11	−1.90	−1.29
0.137	0.91	−0.07	−0.27	−0.86	−0.31
0.122	0.64	−0.52	−0.96	−1.91	−0.81
0.120	1.88	1.36	1.20	0.24	0.94
0.113	2.69	1.88	1.57	0.36	1.56
0.106	−2.18	−3.19	−3.17	−4.05	−3.40
0.099	3.21	2.30	2.37	1.75	2.02
0.073	−0.91	−1.98	−1.91	−2.14	−2.03
0.055	0.43	−0.66	−1.03	−1.71	−1.13
0.011	−4.70	−4.82	−4.97	−7.45	−5.26

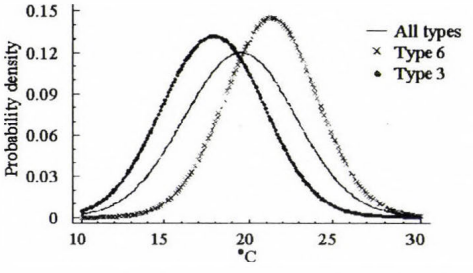


Fig. 1. Probability density function of summer temperatures under different circulation pattern types for Debrecen.

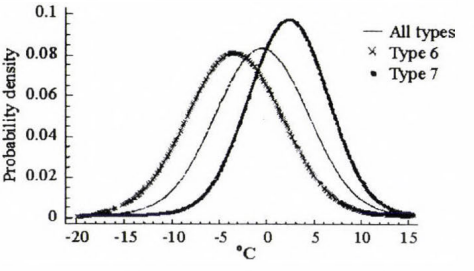


Fig. 2. Probability density function of winter temperatures under different circulation pattern types for Debrecen.

Daily precipitation amount for Budapest is analyzed also for summer and winter separately using a 30-year period of data from 1955 to 1984. In summer, the optimal number of types is 10 and 8 for sequences of wet and dry days (Eq. (16)) and for precipitation amount on wet days (Eq. (23)). Winter requires more types, namely a choice of 12 types is optimal for both the sequences of wet and dry days and for precipitation amount on wet days. It may be concluded, that the number of types is governed principally by the variance of the underlying variable, because periods accompanied with smaller variances require fewer types. Large variances decrease sharply when adding a further type, and thus fewer types are required as compared to small variances. The probability of wet days is 0.336 in summer, while this probability conditioned on CP types varies between 0.190 and 0.527. The probability of precipitation occurrence is 0.356 in winter but extends from 0.071 to 0.523 according to CP types. The mean

precipitation amount on wet days equals to 5.670 mm and 3.152 mm in summer and winter, respectively. Similar means corresponding to the CP types vary from 3.051 to 8.988 mm, and from 1.060 to 4.708 mm, for the two seasons. *Figs. 3* and *4* demonstrate differences among probability distributions of precipitation amount on wet days by fitting gamma distributions to data under different CP types. CP types having large relative frequencies and accompanied with small and large amounts are shown here.

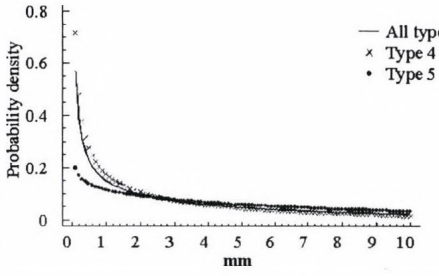


Fig. 3. Probability density function of summer daily precipitation amounts on wet days under different circulation pattern types for Budapest.

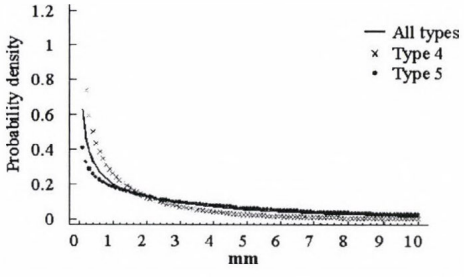


Fig. 4. Probability density function of winter daily precipitation amounts on wet days under different circulation pattern types for Budapest.

4. Discussion

A methodology to develop optimal systems of circulation pattern types for downscaling purposes has been introduced and applied to daily mean temperature and daily precipitation. An optimal system of CP types with a fixed number of possible types has as much information on the specific climate element as possible. The number of types is chosen such that the gain of introducing a further type is significant as compared to the gain of fewer types.

In order to check the performance of the method discussed above, these optimal systems of CP types are compared to Hess-Brezowski (HB) types. This latter classification having 30 types is widely used for the Central European region. A likelihood-ratio test defined as twofold difference between log-likelihood functions corresponding to HB types and the proposed types is used for the purpose. The twofold log-likelihood values are the same as the corresponding AIC values without penalty terms (second terms). Under the null hypothesis, that the proposed systems of types are optimal (HB types have no improvement over the proposed types), the test statistic follows a chi square distribution with 30-K degrees of freedom. A unanimous result for both the temperature and precipitation is that the null hypothesis is not rejected at even a 10% significance level in winter and at a 5% significance level in summer.

Note that 5% significance levels are very common in hypothesis testing, while 10% significance levels deliver even narrower acceptance intervals for null hypotheses. Thus, it may be concluded that the proposed methodology resulting in 8–12 types performs at least as good as HB classification with 30 types.

CP types obtained by the procedure discussed above can be used to reproduce space-time statistical structure of climate elements under a changed climate. For this purpose, for elements modeled by Gaussian processes, the multivariate AR(1) model (Eq. (14)) depends on CP types. Thus, time series simulated by type dependent AR(1) models reflect the properties of local climate via the sequence of CP types obtained from a GCM under changed climate. Local precipitation can be analyzed by a similar way using GCM-generated CP types. First, a sequence of dry and wet days is simulated with the type dependent Markov chain model. Then transformed precipitation amounts (Eq. (17)) on wet days are simulated, according to the type dependent AR(1) model. Finally, precipitation amounts are obtained by inverting Eq. (17). Simulating multisite precipitation cannot be accomplished by the procedure outlined above, because an other transformation of precipitation amounts is needed (*Bartholy et al.*, 2001). Here the optimal system of CP types should be the same for both the precipitation events and amounts. These CP types can be determined by minimizing a weighted sum of Eqs. (16) and (23) applied with areal mean precipitation. The weights can be chosen as reciprocals of Eqs. (16) and (23) under one type.

Acknowledgement—Research leading to this paper has been supported by grants from Hungarian Research and Development Foundation NKFP 3A/2002 and Hungarian Science Foundation OTKA T38038.

References

Akaike, H., 1974: A new look at the Statistical Model Identification. *IEEE T. Autom. Contr. 19*, 716-723.

Bartholy, J., Matyasovszky, I., and *Weidinger, T.*, 2001: Regional climate change in Hungary: a survey and a stochastic downscaling method. *Időjárás 105*, 1-17.

Bardossy, A. and *Plate, E.*, 1992: Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res. 28*, 1247-1260.

Giorgi, F., and *Mearns, L.O*, 1991: Approaches to the simulation of regional climate change: a review. *Rev. Geophys. 29*, 191-216.

Matyasovszky, I. and *Dobi, I.*, 1989: Methods for the analysis of precipitation time series with Markov chains (in Hungarian). *Időjárás 93*, 276-287.

Hughes, J.P., Lettenmaier, D.P., and *Guttorp, P.*, 1993: A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation. *Water Resour. Res. 2*, 3305-3315.

Wilson, L.L, Lettenmaier, D.P., and *Skyllingsta, E.*, 1992: A hierarchical stochastic model of large scale atmospheric circulation patterns and multiple station daily precipitation. *J. Geophys. Res. 97*, 2791-2809.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 135–153

Relationship between soil texture and near surface climate in Hungary

Miklós Drucza and **Ferenc Ács**

Department of Meteorology, Eötvös Loránd University,
P.O. Box 32, H-1518 Budapest, Hungary; E-mail: acs@elte.hu

(Manuscript received in final form March 6, 2006)

Abstract—The near surface climate of Hungary in the period 1901–1950 is analyzed using a Thornthwaite-based model. Annual and monthly mean values of energy- and water balance components are determined for each Thornthwaite's climatic region. The analysis is performed for both constant and texture-dependent soil water holding capacity. The main purpose of this study is to evaluate the effect of soil texture on the near surface climate of Hungary. The main findings are as follows: (1) The Thornthwaite-based model is suitable for reproducing the pattern of climate on mesoscale in Hungary; (2) The effect of soil textural characteristics on the mesoscale pattern of climate in Hungary is pronounced. This effect is commensurable with the effects of relief and atmosphere.

The results suggest that Thornthwaite-based models can be applied not only for climate classification purposes but also for purposes of physical and ecological climatology.

Key-words: soil texture, soil water holding capacity, Thornthwaite, Hungary, hydroclimatic regions, water balance components, energy balance components, evapotranspiration, sensible heat flux, mesoscale

1. Introduction

In all land-surface climate analyses, one of the main concerns is to select the appropriate space/time up-scaling strategy. There are two main approaches. The first approach uses many variables, which are measured in high space/time resolution. The second one uses only a limited number of variables, which have not been determined on as great space/time resolution as the variables from the former case. In the first case the up-scaling algorithm is

very complex (*Running et al.*, 1999). It is based on the use of flux tower measurements, SVAT (Soil-Vegetation-Atmosphere Transfer schemes) model calculations, remote sensing observations, and complex interpolation and extrapolation techniques. In this case, the surface heterogeneity and time resolution is taken into account as detailed as possible.

The second approach is much simpler than the former one. In an extremely simple case, the second approach operates only with monthly mean values of precipitation P and air temperature T. These "classical" approaches are too simple and inappropriate for estimating turbulent heat fluxes. In most cases they use only atmospheric inputs, inputs from other parts of terrestrial climate system (lithosphere, biosphere, hydrosphere, and pedosphere) are rarely used. Some of them serve exclusively for climate classification purposes (*Köppen*, 1900). In some other cases, vegetation is treated only implicitly (*Thornthwaite*, 1948). It should be noted, that in Thornthwaite's climate classification the soil is also represented. The effect of soil can be explicitly expressed by using soil textural data for determining the soil water holding capacity. Annual and monthly mean values of turbulent heat fluxes can also be estimated knowing the net radiation flux. It has to be mentioned, that temperature and precipitation themselves are poor descriptors of climate, regarding vegetation, the active climatic factors are the water balance components (water surplus, water deficit, potential evapotranspiration, actual evapotranspiration, and soil water storage).

The application of *Thornthwaite* (1948) is common today. So, for instance, *Mintz* and *Serafini* (1981) and *Mintz* and *Walker* (1993) used Thornthwaite's method for estimating the global pattern of monthly mean values of available root zone soil water content and evapotranspiration, respectively. Thornthwaite's potential evapotranspiration and water balance model is also used for practical purposes: in USA (jointly by the US Weather Bureau (NOAA) and U. S. Department of Agriculture (USDA)) for calculating weekly maps of the crop moisture index, and in Canada (Canadian Climate Center) for calculating weekly maps of soil moisture (see *Mintz* and *Walker*, 1993).

In Hungary, Thornthwaite's method was firstly applied by *Bacsó* (1943). It was applied also by *Kakas* (1960). Some other applications of Thornthwaite's method referred to some smaller catchment areas (*Szesztay*, 1958). Applications on special sites and daily time scale were also done (*Szepesiné*, 1959). A Carpathian Basin application was demonstrated by *Szepesiné* (1966). She also roughly evaluated the energy balance components of some stations. In these calculations, a 300 mm soil water holding capacity of a 1 m deep soil layer was assumed. Thornthwaite's method in Hungary was applied by *Szász* (1963), too. In contrast to *Szepesiné* (1966), he investigated the effect of soil water holding capacity on water balance components, but only

for some chosen soil types. Today there is no analysis on the subject of soil/climate relationship in Hungary. Moreover, there is no analysis on the effect of the areal distribution of soil texture upon near surface climate. A diagnostic study on soil/climate relationship enables us (1) to improve the classification of near surface climate and (2) to understand and quantify the sensitivity of climate to soil characteristics.

To this end, the aim of this study is threefold: (1) to determine Thornthwaite's hydroclimatic regions in Hungary and to estimate the annual and monthly mean values of energy- and water balance components taking into consideration the soil texture, (2) to analyze the correspondence between Thornthwaite's hydroclimate types and energy- and water balance components, and (3) to evaluate the effect of soil texture on the near surface climate in Hungary. The analysis is performed using a Thornthwaite-based model and climatic data series derived from the period 1901–1950.

2. Method

A water- and energy balance model based on Thornthwaite's climate classification method is used. It is process oriented and as simple as possible. Its advanced feature is that it uses not only atmospheric, but also soil characteristic inputs.

2.1 Input data

Soil water balance calculations are performed for 125 climatic stations. Monthly values of precipitation P and air temperature T are taken from the Cimate Atlas of Hungary (*Kakas*, 1960). These data refer to the period 1901–1950. It has to be noted, that P and T climatic fields (*Kakas*, 1960) agree well with the newest P and T fields presented in the *Climate Atlas of Hungary* (2000). The energy balance calculations are performed for 40 meteorological stations, where the sunshine duration S_h data were also available. These data refer also to the period 1901–1950 (*Dobosi* and *Takács*, 1959). There were two motivations to chose the period 1901–1950: firstly, the consistency between P, T, and S_h data is corresponding, secondly, the most previous studies based on Thornthwaite's method referred to this period (there is a possibility for comparison). Soil texture is determined using high-resolution soil texture map of GIS laboratory of the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. According to the map, there are five main textural classes (in addition to rocky and moor soils). Soil water holding capacity is calculated after *Nemes* (2003), where the

fitting parameters of *van Genuchten's* (1980) *pF* curves ($pF=\log_{10}(|\Psi|)$, where Ψ is the soil water petential in cm H_2O) are given for Hungarian soils using the 11 soil textures of the USDA soil classification. The correspondence between the Hungarian and USDA texture classifications is determined by *Filep* and *Ferencz* (1999), see Eq. (8). The wilting point soil water content is calculated for $pF=4.2$, while the field capacity soil water content is determined for $pF=2.5$. Soil water holding capacity values of rocky (4 stations) and moor (3 stations) soils are taken to be 100 mm (as given in *Thornthwaite,* 1948), because textural effects are fairly site specific on these soils. The wilting point and field capacity soil moisture contents together with soil water holding capacity are presented in *Table 1*.

Table 1. Water holding capacity, wilting point and field capacity soil moisture content of the 5 main textures as determined by *Nemes* (2003)

Main textural classes	Wilting point (mm/m)	Field capacity (mm/m)	Water holding capacity (mm/m)
Sand	15	80	65
Sandy loam	105	274	169
Loam	136	332	196
Clay loam	175	387	212
Clay	272	485	213

2.2 Water balance calculation

Water balance components (potential (*PET*) and actual (*ET*) evapotranspiration, water surplus (*S*), water deficiency (*D*), and available soil water content relative to water holding capacity (θ)) are estimated by the method of *Thornthwaite* (1948). Calculation procedure, climate classification, and climate symbols are presented in Appendix A.

2.3 Energy balance calculation

Global radiation is parameterized by the method of *Takács* (1971). Net shortwave radiation is estimated using albedo values presented in *Borhidi* and *Dobosi* (1967). Net long-wave radiation is approximated by its annual mean value for Hungary as given in *Major et al.* (2002). Soil heat flux is parameterized after *Ács et al.* (1986). Sensible heat flux is calculated as the residual term of energy balance equation. Other energy balance components, like the energy of melting and freezing, as well as the energy consumed by assimilation are neglected. Parameterizations used in our calculations are presented in Appendix A.

3. Results

Areal distribution of Thornthwaite's hydroclimate regions (hereafter briefly regions) in Hungary is presented in *Fig. 1*. The regions are obtained assuming texture-dependent soil water holding capacity. In Thornthwaite's climate formulae (representing the hydroclimate regions), only the first and third symbols are indicated. These symbols characterize the humidity or aridity of the climate, briefly the hydroclimate. The map shows the well-known hydroclimate pattern on mesoscale (*Péczely*, 1979): the moderately moist Trans-Danubia and North Hungarian Mountains, as well as the arid lowlands. The humid regions are located in the southwest parts of the country, in the vicinity of Sopron, in the Bakony region, and in the higher lands of Pilis, Börzsöny, Mátra. and Bükk mountains. The greatest part of the Great Hungarian Plane is dry subhumid with little or no water surplus. Moderate winter water surplus prevails in the upper 1 m soil layer at sandy regions also in lowlands (Kiskunság, Nyírség).

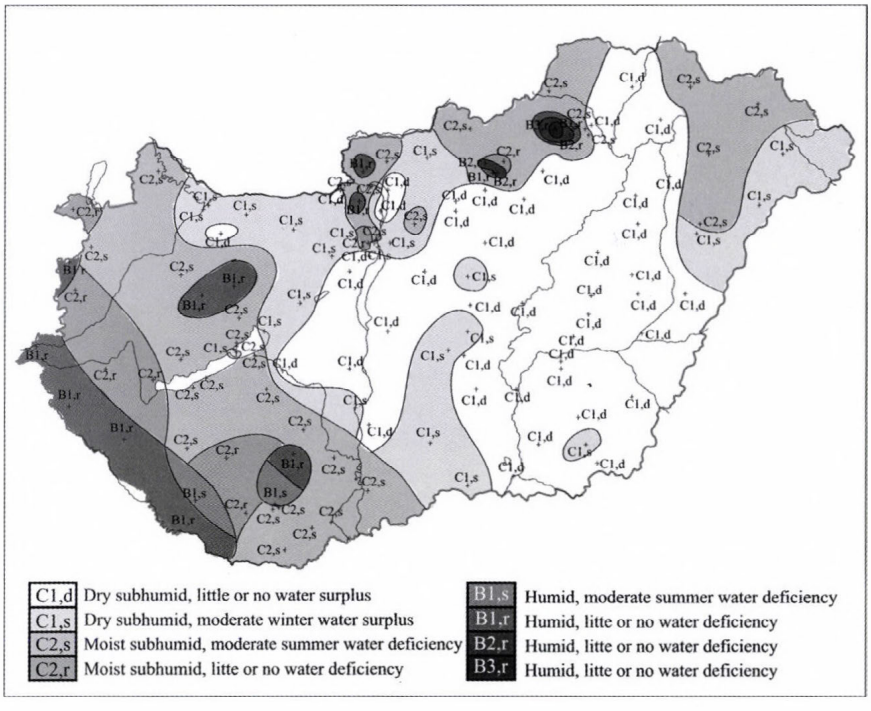


Fig. 1. Areal distribution of Thornthwaite's climate regions in Hungary assuming texture-dependent soil water holding capacity.

The hydroclimate pattern presented is much more detailed as the one presented in *Bacsó* (1943). It has to be noted, that there are also regions, which are climatically different, though their hydroclimate are the same. For instance C1,d hydroclimate exists both in the Szeged region and just below the Mátra and Bükk mountains, although the real climate is different. Similarly, C2,s hydroclimate in the southwest part of the country somewhat differs from the same C2,s hydroclimate in the northeast part of the country. This is, in one respect, caused by the resolution used in the Thonthwaite's method. Namely, the method is constructed for global scale and not for mesoscale applications. Nevertheless, it has to be underlined, that other global scale climate classification methods, as for instance the *Köppen* (1900) method, are less appropriate for Hungarian applications (*Péczely*, 1979). On the other hand, the C2,s identity in the southwest and northeast part of the country is derived from the similar hydroclimate bases.

3.1 Energy and water balance of Hungary

The energy balance components are calculated for 40 stations. The areal distribution of the radiation-, heat-, and water fluxes is illustrated using spline interpolation technique. The areal distribution of global radiation in January, April, July, and October is presented in *Fig. 2*. It is hard to say, whether the results obtained are well enough, but the basic characteristics of the areal distribution of global radiation in July are in accordance with those obtained by *Rimóczi-Paál* (2004). The actual evapotranspiration together with the global radiation is climatically the most important energy balance component. The annual sum of *ET* is presented in *Fig. 3*. *ET* is between 400–600 mm yr^{-1}. The areal distribution of *ET* is similar to that of hydroclimate regions. *ET* is the largest in B1 hydroclimate regions and on the hilly areas, where C2,s hydroclimate is prevailing. In these cases, *ET* is partly determined by the relief. Note, that the model observes relief in terms of temperature (thus, *PET*) and precipitation. The water surplus including runoff is obtained as only the result of these factors.

ET is determined not only by the relief but also by the soil texture. This can be seen comparing areal distributions of *ET* and soil texture (*Várallyay et al.*, 1980). The *ET* distribution is strongly influnced by sandy soils via its low water holding capacity values. The high water permeability combined with low water holding capacity values causes pronounced infiltration in these regions. *ET* is the smallest in regions with sandy soils for example Somogy, Kiskunság, and Nyírség, it amounts only to 400 mm $year^{-1}$ many times. This effect of sand can also be seen in climatic regions. Namely, in these cases the third symbol is mainly 's', which means temperate winter water surplus for C1 hydroclimates, or temperate summer water deficiency for C2 hydroclimates.

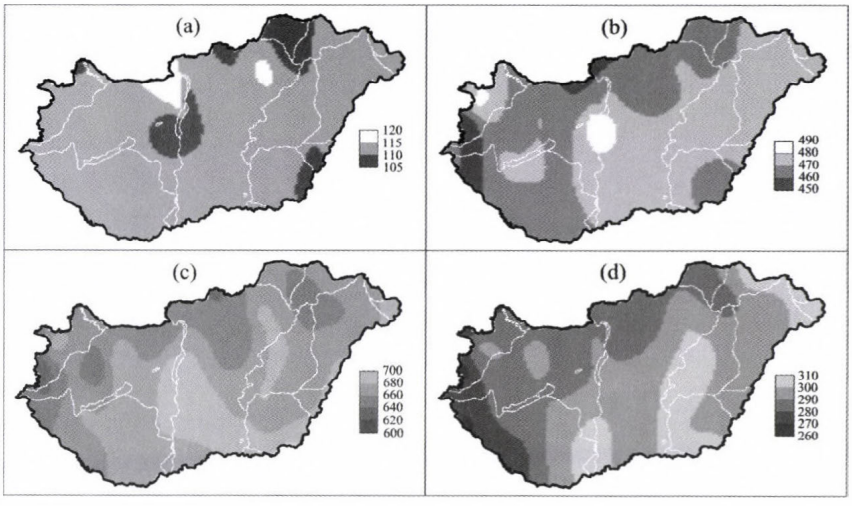


Fig. 2. Areal distribution of global radiation (MJ m^{-2} $month^{-1}$) in (*a*) January, (*b*) April, (*c*) July, and (*d*) October.

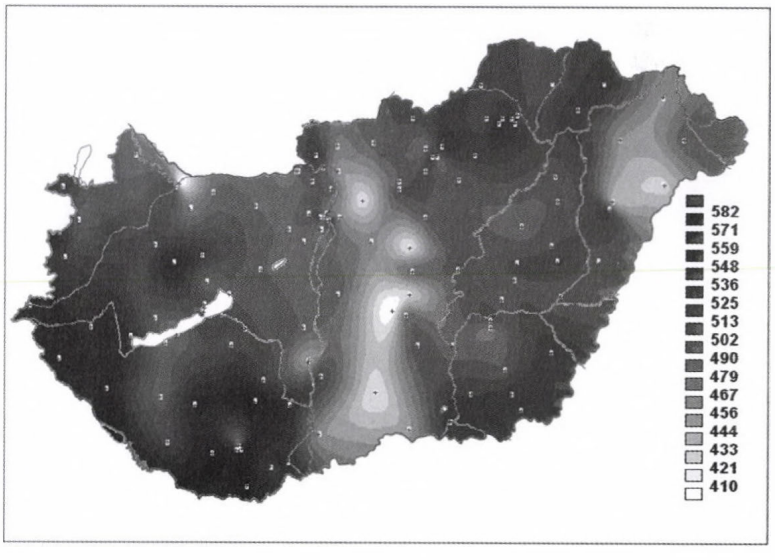


Fig. 3. Areal distribution of annual actual evapotranspiration (mm yr^{-1}).

The areal distribution of *ET* in April, July, and October is presented in *Fig. 4*. The effect of the relief on *ET* can also be identified in April and October, while in July *ET* is governed mostly by the soil texture. The range of obtained *ET* values is in accordance with those presented in *Climate Atlas of Hungary* (2000) (hereafter briefly Climate Atlas), except for the stations with driest sandy soils, where our *ET*-values are much lower. Nevertheless, the areal distributions of *ET* are quite different. In Climate Atlas, the effect of relief on *ET* is observable (note that no calculation procedure is presented), but the effect of soil texture on *ET* cannot be seen. For instance, in July there are no *ET* minima in sandy regions of Somogy, Kiskunság, or Nyírség. These *ET* minima regions above sandy areas should be obtained, if soil texture effects were taken into account (e.g., *Irannejad* and *Shao*, 1998). In spite of this, *ET* in July in *Fig. 4* is governed by both the effect of relief and the effect of soil texture. According to our results, *ET* distributions in July and in the year show similarities to the pattern of the soil texture. This is in accordance with many results referring to soil-vegetation-climate interactions (*Pitman,* 1994; *Pielke*, 1998). It has to be noted, that the results of *Kakas* (1960) obtained by the Thornthwaite's method also differ from our ones, because the effect of soil texture was not taken into account.

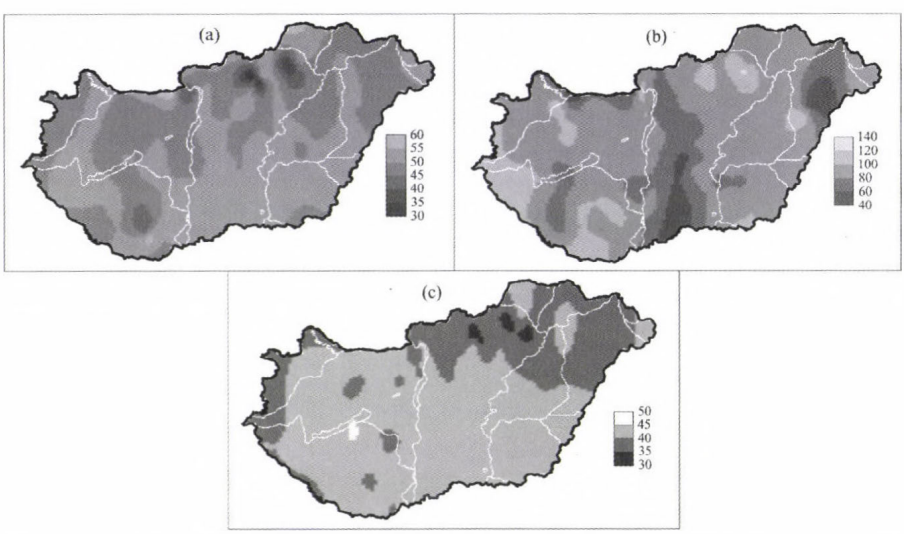


Fig. 4. Areal distribution of actual evapotranspiration (mm month^{-1}) in (*a*)April, (*b*) July, and (*c*) October.

The areal distribution of the relative soil water content available for vegetation (θ, percent of soil water holding capacity in the upper 1 m soil layer) in January, April, July, and October are shown in *Fig. 5*. In January the most arid regions — the valley of the Tisza and Kőrös rivers show unsaturated soils. In July, θ is about 30 percent in the regions with driest sandy soils. The importance of the soil textural differences can be clearly observed in summer. So, for instance, in summer the sandy regions (Pesti-síkság, Kiskunság, Paks, Külső-Somogy, Nyírség) are dryer than the surrounding areas (e.g., in Debrecen Pallagpuszta, on sand θ=44%, while in Debrecen, on loam soil θ=51%, even if there are almost no differences in temperature and precipitation values). In October, the areal distribution of θ is quite uniform in flatlands, only the humid-subhumid part of the country (in SW) and the mountains show high θ values.

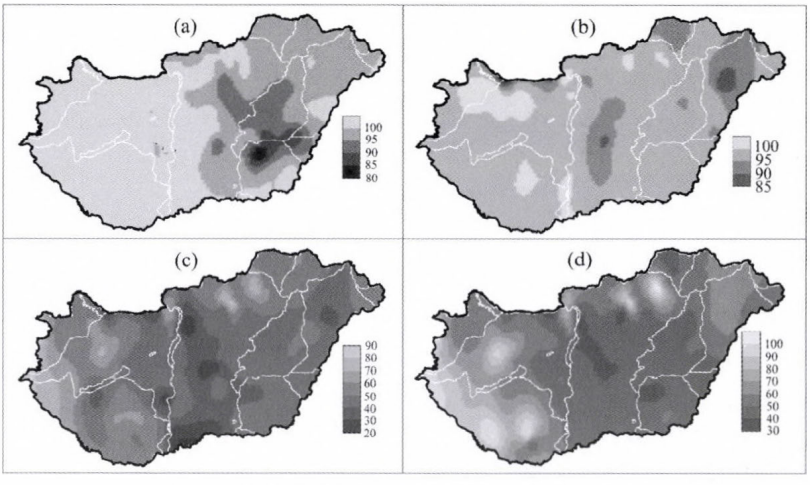


Fig. 5. Areal distribution of the available soil water in percent of soil water holding capacity in (*a*) January, (*b*) April, (*c*) July and (*d*) October.

The areal distribution of the sensible heat flux H in January, April, July, and October is shown in *Fig. 6*. The effect of soil texture is most pronounced in the case of H, especially in summer. In July, the sandy areas outstand well with their large H values. In winter, H is negative. In spring and autumn, the largest H values are obtained for elevated areas, mainly due to the inconsistencies in the parameterization of *PET* and R_g. R_g is parameterized via sunshine duration, while *PET* is estimated from temperature data. The rate of the temperature change and sunshine duration change is not the same due to growing elevation. Therefore, the difference between R_g and *PET* increases with elevation causing large H values in spring and autumn.

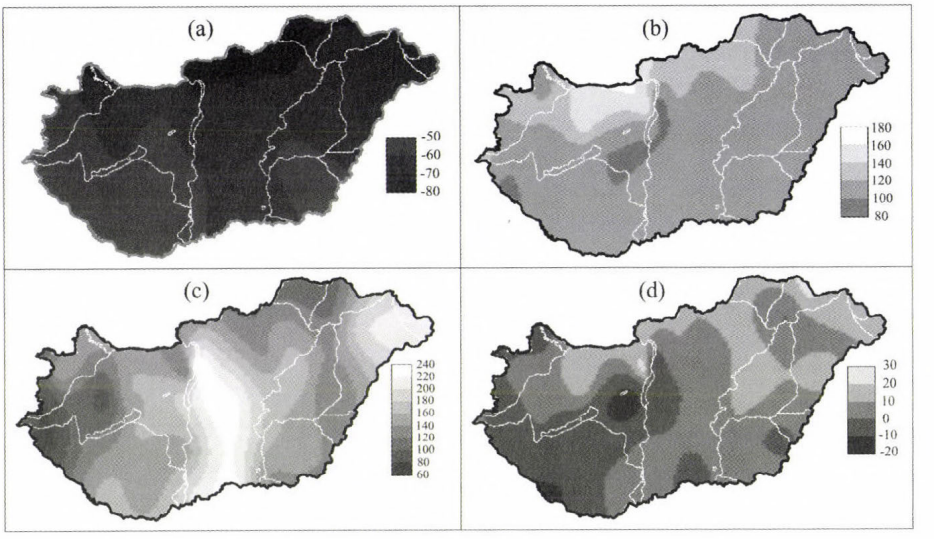


Fig. 6. Areal distribution of sensible heat flux (MJ m^{-2}) in (*a*) January, (*b*) April, (*c*) July, and (*d*) October.

3.2 Correspondence between Thornthwaite's hydroclimate regions and energy balance components

To understand Thornthwaite's climate classification for Hungary, the correspondence between hydroclimate regions and energy balance components must be known. Thornthwaite's hydroclimate regions for Hungary and the corresponding typical annual values of global radiation, sensible heat flux, actual evapotranspiration, and relative soil water content are presented in *Table 2*. The values of energy- and water balance components are overlapping, but the average values show definite trends. R_g is somewhat growing towards dry hydroclimates; dryer hydroclimates have larger H. 's' hydroclimates (usually above sand) have larger H than 'd' hydroclimates. The actual evapotranspiration is the largest for 'B1' hydroclimate regions because of the abundant precipitation, and θ is the largest for 'B3' hydroclimate regions due to the low temperature and relatively high amount of precipitation.

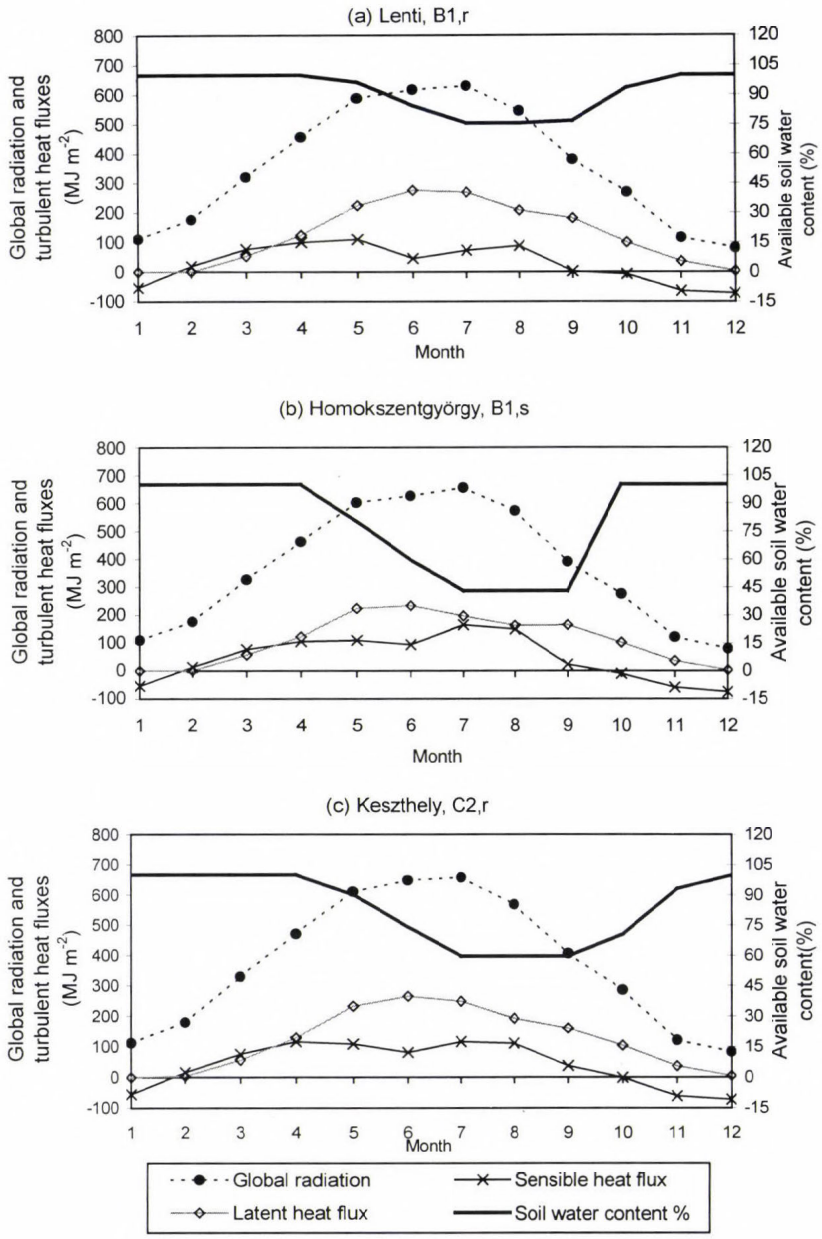
The annual course of energy- and water balance components for each hydroclimate region is shown in *Fig. 7*. Let us briefly take a look at the figures. The main difference between 'B1, r' and 'B1, s' hydroclimate regions (see *Fig. 7a, b*) can be noticed in the summer and autumn values of θ. In 'B1, s' hydroclimate, the θ values are lower, while the corresponding H values are larger than in 'B1, r' hydroclimate.

Table 2. Correspondence between the typical annual values of global radiation, sensible heat flux, actual evapotranspiration, and relative soil water in the root zone available for vegetation and the Thornthwaite's climatic regions (1st and 3rd symbols of the climate formula) for Hungary

	Global radiation, *Gr* (MJ $m^{-2}yr^{-1}$)	**Sensible heat flux, *H* (MJ m^{-2} yr^{-1})**	**Actual evapotranspiration, *ET* (mm)**	**Relative soil water in the root zone, *SM* (%)**	**Number of stations for *Gr* and *H***	**Number of stations for *ET* and *SM***
B3, r	–	–	**506**	**95**	–	1
B2, r	–	–	**518** (497-540)	**93** (91-95)	–	2
B1, r	**4383** (4294-4516)	**465** (316-637)	**553** (486-596)	**89** (86-96)	4	13
B1, s	**4405**	**543**	**508** (494-522)	**81** (81-82)	1	2
C2, r	**4404** (4247-4506)	**510** (405-701)	**554** (508-580)	**85** (82-89)	4	8
C2, s	**4472** (4368-4562)	**631** (543-764)	**507** (426-551)	**78** (69-82)	12	34
C1, s	**4482** (4433-4524)	**692** (575-833)	**478** (411-536)	**73** (66-78)	6	22
C1, d	**4516** (4417-4604)	**620** (519-708)	**499** (464-524)	**72** (66-78)	13	43
Average/ sum	**4470**	**606**	**507**	77	40	125

In humid 'B1, r' and moist subhumid 'C2, r' (*Fig. 7c*) hydroclimate regions, the annual course of energy- and water balance components are quite similar. However, the latter one is obviously dryer. In 'C2, s' hydroclimate (*Fig. 7d*), *H* is already larger than *LE* in July.

In dry subhumid (C1, *Fig. 7e, f*) hydroclimates, relative soil water content available for vegetation is small in summer and autumn. In summer, *H* is quite large exceeding *LE* in some months. Note, that 'r' and 's' hydroclimates are quite different in terms of annual courses of some variables. This is especially valid for θ. In 's' hydroclimates (usually sandy stations), the changes of θ are quite abrupt in spring and autumn. Also, the summer drought is more pronounced above 's' hydroclimate regions by prevailing larger *H* values and smaller θ and *LE* values than in 'r' hydroclimate regions.


(a) Lenti, B1,r
Global radiation and turbulent heat fluxes (MJ m-2)
Available soil water content (%)
Month
(b) Homokszentgyörgy, B1,s
Global radiation and turbulent heat fluxes (MJ m-2)
Available soil water content (%)
Month
(c) Keszthely, C2,r
Global radiation and turbulent heat fluxes (MJ m-2)
Available soil water content(%)
Month
Global radiation
Sensible heat flux
Latent heat flux
Soil water content %

→→

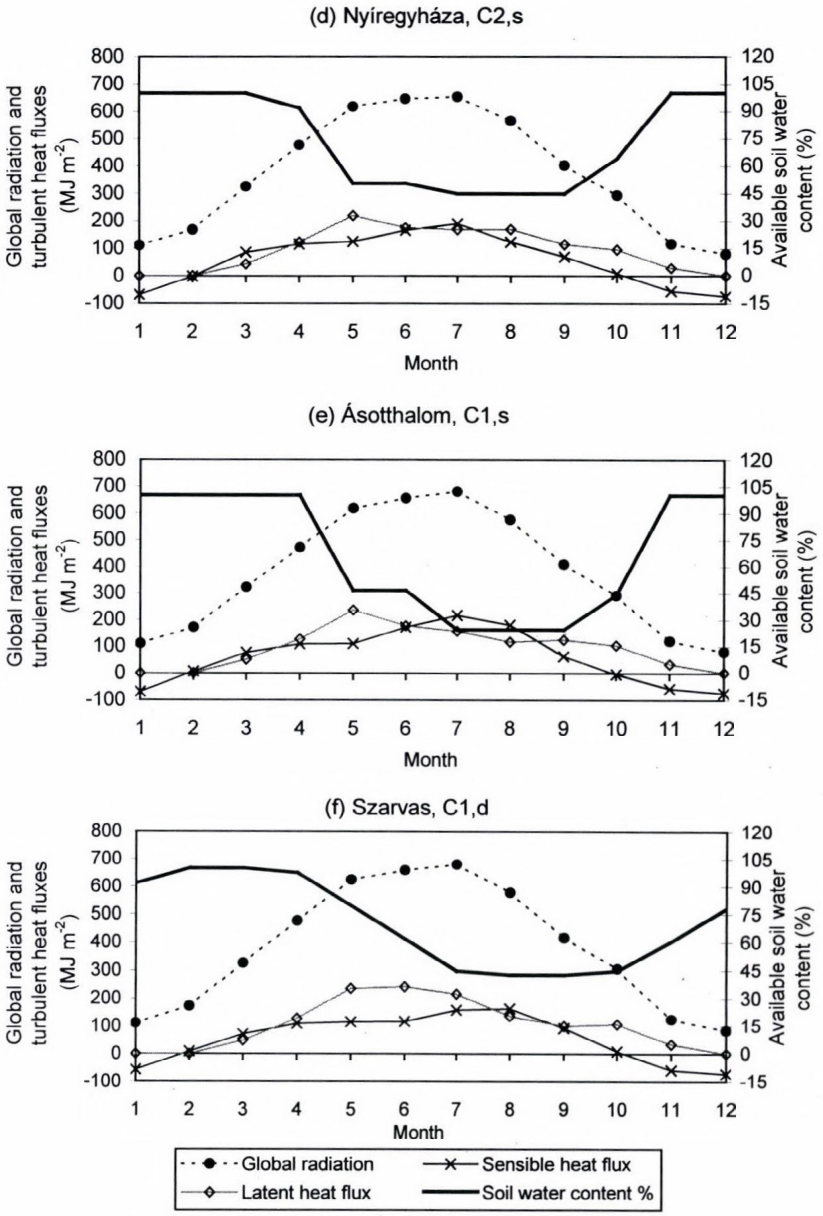


Fig. 7. The annual course of global radiation (MJ m^{-2}), sensible heat flux (MJ m^{-2}), actual evapotranspiration (mm), and available soil water content (percent of soil water holding capacity); (*a*) Lenti (B1, r); (*b*) Homokszentgyörgy (B1, s); (*c*) Keszthely (C2, r); (*d*) Nyíregyháza (C2, s); (*e*) Ásotthalom (C1, s); (*f*) Szarvas (C1, d).

3.3 Effect of soil texture on near surface climate

The effect of soil texture on near surface climate is analyzed comparing Thornthwaite's hydroclimate regions as well as water- and energy balance components obtained by constant and texture-dependent soil water holding capacity. Constant soil water holding capacity is taken to be 100 mm as it is originally made in Thornthwaite's study. The areal distribution of Thornthwaite's hydroclimate regions for 100 mm soil water holding capacity is shown in *Fig. 8*. Comparing *Fig. 1* and *Fig. 8*, it is striking that the Great Plane is dryer for the texture-dependent case than for the 100 mm case. Soil water holding capacity for the texture-dependent case is generally (except sands) greater than 100 mm; consequently, the winter water surplus for the texture-dependent case is less than the winter water surplus for the 100 mm case. For sandy soils (in this case, water holding capacity is less than 100 mm), the 3rd symbol is unchanged, that is, it remains 's' in both cases.

The effect of soil on near surface climate in terms of energy- and water balance components is illustrated in *Fig. 9*. The areal distribution of annual differences of *ET*, θ, and *H* obtained from texture-dependent and 100 mm soil water holding capacity cases are very similar to the areal distribution of soil texture. The *ET*- and θ-differences are large for soils of great water holding capacity (loam, clay-loam, and clay). In these cases, the *H*-differences are small or negative. For sandy soils, which have smaller water holding capacity than 100 mm in the 1 m soil layer, the opposite is true. This is especially obvious for *H* (see *Fig. 9c*) in regions of Kiskunság, Nyírség, and Külső-Somogy.

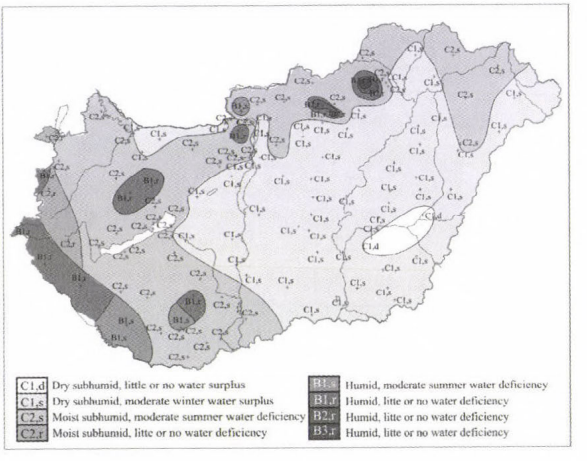


Fig. 8. Areal distribution of Thornthwaite's climate regions in Hungary for constant soil water holding capacity of 100 mm.

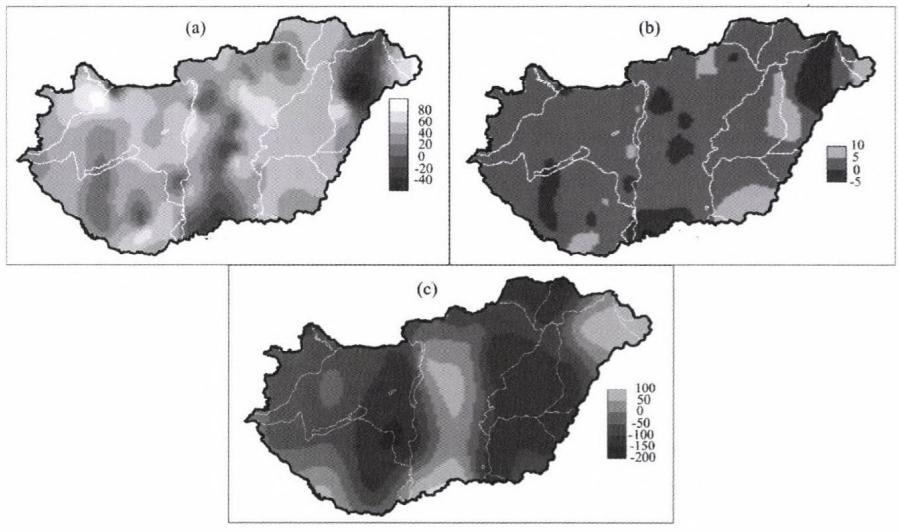


Fig. 9. The annual mean difference between the texture dependent (variable water holding capacity) and texture-independent (constant water holding capacity of 100 mm) (*a*) actual evapotranspiration (mm), (*b*) available soil water content (in percent of the soil water holding capacity), (*c*) sensible heat flux (MJ m^{-2}). Positive values mean that texture-dependent values are greater than constant ones.

4. Conclusions

A Thornthwaite-based water- and energy balance model is constructed to analyze the near surface climate of Hungary and the effect of soil texture on climate. The analysis refers to the period 1901–1950. For each Thornthwaite's formula the corresponding annual and monthly mean values of energy- and water balance components are determined. The analysis is performed for constant and texture-dependent soil water holding capacities.

The main purpose of this study was to evaluate the effect of soil texture on the near surface climate of Hungary. Our purpose was not to create a new climate classification for Hungary. The only reason for obtaining different areal distribution of Thornthwaite's hydroclimate regions than in the former studies for Hungary is, that the soil water holding capacity values are estimated on more realistic base considering the soil textural characteristics.

The main results are as follows:

- The method of *Thornthwaite* (1948) is suitable to characterize the areal distribution of the hydroclimate on mesoscale in Hungary. It has to be noted, that other global climate classification methods, as for instance Köppen's method (*Köppen*, 1900), are less appropriate for Hungarian applications.

- The correspondence between the Thornthwaite's climate classification and the climate regions classified with the energy balance components is determined. Dryer hydroclimates (C1) show larger H and lower LE- and θ values than the moist (C2, B) ones. R_g is somewhat growing towards dry hydroclimates.
- The effect of soil texture on the near surface climate of Hungary is analyzed in terms of both the Thornthwaite's climate classification and the energy- and water balance components. The most pronounced effect on the areal distribution of energy- and water balance components is caused by sandy soils, where extreme large H and small ET- and θ values prevail in summer. The effect of soil texture on the near surface climate is proved to be commensurable with the effects of relief and atmosphere.

It is obvious that the Thornthwaite-based model is suitable to characterize the pattern of climate not only on the global scale but also on the mesoscale. Further, it can be easily extended with diagnostic modules based on empirical models for calculating carbon balance components including both vegetation and soil. This extension is one of the goals in our future research work.

Acknowledgements—We express our grateful thanks to the personnel of GIS Laboratory of Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences for supplying a high resolution soil textural map. We thank also to *Hajnalka Breuer*, who helped in figure presentations.

This study was subsidized by the PhD School Foundation of Earth Sciences of Eötvös Loránd University, by NKFP Foundation, project number NKFP-3A/0006/2002, and by the OTKA Foundation, project number T043695.

APPENDIX A

Parameterization of water balance components

Potential evapotranspiration PET is obtained by

$$PET = 1.6\left(\frac{L}{12}\right)\left(\frac{N}{30}\right)\left(\frac{10T_a}{I}\right)^a, \tag{A.1}$$

where L is the mean monthly day length in hours, N is the number of days in a month, T_a is the mean monthly air temperature (°C), while I is the heat index calculated as

$$I = \sum_{j=1}^{12}\left(\frac{T_a}{5}\right)^{1.514} \tag{A.2}$$

and

$$a = 6.75 \cdot 10^{-7} I^3 - 7.71 \cdot 10^{-5} I^2 + 1.792 \cdot 10^{-2} I + 0.49239 . \qquad \text{(A.3)}$$

Water balance calculations are performed according to *Mintz* and *Walker* (1993). Actual evapotranspiration *ET* is proportional to *PET* and soil water in the root zone θ (for climatic tasks root zone is usually taken to be equal to the 1 m deep soil layer).

$$ET_i = \beta_i \, PET_i , \qquad \text{(A.4)}$$

where

$$\beta_i = \frac{\theta_{i-1} - \theta_w}{\theta_f - \theta_w} . \qquad \text{(A5)}$$

θ_{i-1} is the soil moisture content at the end of the previous month *i–1* (note, that monthly θ values are represented by θ_i values obtained at the end of the actual month), θ_f is the soil moisture content at the field capacity (mm), and θ_w is the soil moisture content at the wilting point (mm). In the dry period ($P < PET$ and $ET < PET$), when *ET* is less than *P*, *ET* is set to be equal to *P*. Soil water available to vegetation (water supply for vegetation) (θ, mm), water surplus (*S*, mm) and water deficiency (*D*, mm) at the *i*th time step are obtained by

$$\theta_i = \theta_{i-1} + P_i - ET_i ,$$

$$S_i = \begin{cases} P_i - E_i , \text{ if } \theta_i = \theta_f - \theta_w \\ 0, \qquad \text{if } \theta_i < \theta_f - \theta_w \end{cases} ,$$

$$D_i = PET_i - ET_i . \qquad \text{(A.6)}$$

Climate symbol definitions

Thornthwaite's climate classification method consists of 4 symbols. The 1st and the 3rd symbols are defined by the moisture index (I_m), the 2nd and the 4th symbols are defined by *PET*. The moisture index is obtained by the annual amounts of *S*, *D*, and *PET*:

$$I_m = I_h - 0.6 I_a , \qquad \text{(A.7)}$$

where

$$I_a = \frac{100 D_{year}}{PET_{year}} , \qquad I_h = \frac{100 S_{year}}{PET_{year}} , \qquad \text{(A.8)}$$

where I_h is the humidity index and I_a is the aridity index.

Definitions of the 1st and the 3rd symbols of Thornthwaite's climate classification are presented in *Table A 1–2*.

Table A1

Thornthwaite's first symbol	Climate type	I_m
A	Perhumid	>100
B_4	Humid	80–100
B_3	Humid	60–80
B_2	Humid	40–60
B_1	Humid	20–40
C_2	Moist subhumid	0–20
C_1	Dry subhumid	-20–0
D	Semiarid	-40 – -20
E	Arid	-60 – -40

Table A2

Thornthwaite's third symbol	Moist climates (A, B, C_2)	Aridity index
r	r: little or no water deficiency	0–16.7
s	s: moderate summer water deficiency	16.7–33.3
w	w: moderate winter water deficiency	16.7–33.3
s_2	s_2: large summer water deficiency	>33.3
w_2	w_2: large winter water deficiency	>33.3
	Dry climates (C_1, D, E)	Humidity index
d	d: little or no water surplus	0–10
s	s: moderate winter water surplus	10–20
w	w: moderate summer water surplus	10–20
s_2	s_2: large winter water surplus	>20
w_2	w_2: large summer water surplus	>20

Parameterization of energy balance components

The monthly amount of global radiation (R_g) is calculated after *Takács* (1971):

$$Rg = A_i S_h + B_i , \tag{A.9}$$

where S_h is the monthly amount of sunshine duration in hours, A_i and B_i are empirical constants given for each month. Soil heat flux is approached by the Budyko's formula (see, e.g., *Ács* and *Mihailovic*, 1986):

$$\begin{aligned} G_i &= E_i G_{amp} , \\ G_{amp} &= C T_{amp} + D , \end{aligned} \tag{A.10}$$

where G_i (MJ m^{-2}) is the monthly sum of the soil heat flux in the ith month, G_{amp} (J m^{-2}) and T_{amp} (°C) are the annual amplitude of the soil heat flux and monthly mean air temperature, E_i, C, and D are empirical constants.

References

Ács, F. and *Mihailovic, D.T.*, 1986: The annual course of heat exchange in the soil in Novi Sad (in Hungarian). *Időjárás 90*, 359-364.

Bacsó, N., 1943: Climate classification of Hungary by Thornthwaite and vegetation geography aspects of climate maps (in Hungarian). *Időjárás* XLVII, 81-91.

Borhidi, A. and *Dobosi, Z.*, 1967: Areal distribution of surface albedo values in Hungary (in Hungarian). *Időjárás 71*, 150-157.

Climate Atlas of Hungary, 2000: Hungarian Meteorological Service, Budapest. ISBN 963 7702 830.

Dobosi, Z. and *Takács, L.*, 1959: The areal distribution of global radiation in Hungary (in Hungarian). *Időjárás 63*, 82-84.

Filep, Gy. and *Ferencz, G.*, 1999: Recommendation for improving the accuracy of soil classification on the basis of particle composition (in Hungarian). *Agrokem. Talajtan 48*, 305-317.

Irannejad, P. and *Shao, Y.*, 1998: Description and validation of the atmosphere-land-surface interaction scheme (ALSIS) with HAPEX and Cabauw data. *Global Planet Change 19*, 87-114.

Kakas, J., 1960: Yearly amount of potential evapotranspiration, water surplus and water deficit. *Climate Atlas of Hungary*. Akadémiai Kiadó, Budapest. 46/2-4 maps.

Köppen, W., 1900: Versuch einer Klassification der Klimate, vorzsugsweise nach ihren Beziehungen zur Pflanzen Geograph. *Zeitschr. 6*, 593-611, 657-679.

Major, Gy., Nagy, Z., and *Tóth*, Z., 2002: Climate-energetic studies for Hungary (in Hungarian). *A Budapesti Közgazdaságtudományi és Államigazgatási Egyetem Környezettudományi Intézetének tanulmányai*. 9. 52pp.

Mintz, Y. and *Serafini, Y.V.*, 1981: Global fields of soil moisture and surface evapotranspiration. *NASA Goddard Flight Center Tech. Memo. 83907, Research Review 1980/1981*, 178-180.

Mintz, Y. and *Walker, G.K.*, 1993. Global fields of soil moisture and land surface evapotranspiration derived from observed precipitation and surface air temperature. *J. Appl. Meteorol. 32*, 1305-1334.

Nemes, A., 2003: Multi-scale hydraulic pedotransfer functions for Hungarian soils. *PhD. Dissertation*. Wageningen Universiteit. ISBN 90-5808-804-9. 143 p.

Péczely, Gy., 1979: Climatology (in Hungarian). Nemzeti Tankönyvkiadó, Budapest. 336 pp.

Pielke, R.A., 1998: Climate prediction as an initial value problem. *B. Am. Meteorol. Soc. 79*, 2743-2746.

Pitman, A.J., 1994: Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model. *J. Climate 7*, 1856-1869.

Rimóczi-Paál, A., 2004: Radiation maps of Hungary. *Időjárás 108*, 195-208.

Running, S.W., Baldocchi, D.P., Turner, S.T. Gower S.T. Bakwin, P.S., and *Hibbard, K.A.*, 1999: A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sens. Environ. 70*, 108-127.

Szász, G., 1963: Analysis of the climatic values of water balance components in Hungary (in Hungarian). *Debreceni Agrártudományi Főiskola Tudományos Közleményei*, 49–71.

Szesztay, K., 1958: Estimation of water balance of catchment areas in Hungary. *Időjárás 62*, 313-328.

Szepesiné, L.A., 1959: Estimating diurnal values of soil water content using climatic data series (in Hungarian). *Időjárás 63*, 7-18.

Szepesiné, L.A., 1966: Characteristics of the hydro-climate of the Carpathian-Basin (in Hungarian). *Beszámolók az 1965-ben végzett tudományos kutatásokról*. Országos Meteorológiai Intézet, Budapest, 86-114.

Takács, M., 1971: Radiation climate of Hungary (in Hungarian). *Szakdolgozat*. ELTE Meteorológiai Tanszék, Budapest.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. *Geogr. Rev., XXXVIII.*, 55-93.

Várallyay, G., Szűcs, L., Rajkai, K, Zilahy, P. and *Murányi, A.*, 1980: Hydro-physical classification and 1:100 000 scale maps of Hungarian soils (in Hungarian). *Agrokem. Talajtan 29*, 77-112.

van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J. 44*, 892–898.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 155–173

Characteristics and synoptic classification of heavy snowfall events in Budapest for the period 1953–2003 Part II

Tamás Hirsch and György Babolcsai

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

(Manuscript received in final form May 31, 2006)

Abstract—In this research, heavy snowfall events (HSEs) in Budapest have been investigated based on a 50-year period. Following the results connected with the main characteristics of HSEs presented in Part I of the paper, the synoptic classification of the cases has been addressed in Part II. Because of the well-known deficiencies of previous synoptic classifications, it was necessary to define our own weather type system. For this purpose, temporally averaged fields of selected meteorological parameters have been used to characterize processes during the whole event, rather than only reflecting the instantaneous state of the atmosphere. In this way, 8 weather types have been defined with quite different characteristics in many aspects. Not surprisingly, most of the HSEs in Budapest have been caused by different types of Mediterranean cyclones. Our classification has also been tested with the help of case studies taken from an independent winter period (2005/2006). Our results show that synoptic-scale processes are strongly connected to heavy precipitation in winter, and that synoptic climatological investigations covering long periods are very useful for operational weather prediction.

Key-words: synoptic climatology, heavy snowfall events, synoptic classification, weather types

1. Introduction

Subjective classification of the weather on synoptical scales dates back to 1944, when *Baur et al.* (1944) created the first classification for Europe and introduced the phrase "Grosswetterlage" (macrosynoptic weather situation). Their 21 weather types described periods, usually lasting for several days, when the spatial distribution of mean sea level pressure remained more or less the same over a large area, e.g. Europe.

Inspired by the development of synoptic meteorology, this first classification was completely renewed by *Hess* and *Brezowsky* (1952). When defining their macrosynoptic weather types - also known as "HB types" -, they used characteristics of the upper layer of the atmosphere (500 hPa level) as well as the mean sea level pressure field. In the following years, this classification was continually improved (*Hess* and *Brezowsky*, 1969, 1977; *Gerstengarbe et al.*, 1993), and by now it consists of 29 weather types including the zonal, meridional, and mixed main circulation forms and a clear distinction between cylonic and anticyclonic situations. Although HB types have been defined focusing on Central Europe (in other words, the location of synoptic-scale weather systems has been considered in relation to this part of the continent), this classification cannot be regarded as very useful for the Carpathian Basin in many cases. The reason for this is that the authors have considered the area of Germany as Central Europe rather than the Carpathian Basin. In this way, some weather situations, which are very important for Hungary, are not included in the HB types. Mediterranean cyclones are good examples of this as they play essential role in developing heavy precipitation events in Hungary, mainly in the winter season.

Another synoptic classification designed specifically for the Carpathian Basin, has been defined by *Péczely* (1957, 1983). This classification has been widely used for many purposes including long-range weather prediction. Like the HB types, it is based on the mean sea level pressure and 500 hPa geopotential height fields. For investigations for the area of Hungary, Péczely's weather types have proved to be much more appropriate than HB types. There are, however, also some deficiencies in this classification. Probably the main problem is that it reveals only a snapshot of the instantaneous state of the atmosphere. Consequently, a situation of a Mediterranean cyclone moving quickly over Hungary can be classified as one of three possible weather types (forward side of a Mediterranean cyclone, cyclone centre over Hungary, or rearward side of a Mediterranean cyclone) depending on the instantaneous location of the cyclone at the time of classification. Another problem is the distinction between anticyclonic and cylonic types depending on whether the mean sea level pressure is above or below 1015 hPa over the major part of Hungary. Considering the relative value of mean sea level pressure compared to surrounding areas or the curvature of isobars would be a much better choice than using a given absolute value to separate cases.

When investigating heavy precipitation events, the use of the weather types suggested by *Bodolainé* (1983) may be beneficial. The author classified cases that led to considerable flooding on the basin of the river Danube and Tisza. Her 7 weather types were determined using the fields of 4 meteorological parameters, namely mean sea level pressure, 500 hPa

geopotential height, 500/1000 hPa thickness, and precipitable water. Based on these fields, it is possibile to consider the surface location of weather systems, the steering flow, the temperature conditions of the lower half of the troposphere, and the location of the warm conveyor belt. The author tried to describe processes usually lasting for several days, rather than giving an instantaneous state of the atmosphere as in former classifications. In spite of this, these weather types describe only those situations when considerable flooding is likely in the basin of the rivers Danube and Tisza. Consequently, these do not include all the heavy precipitation events. Not surprisingly, 25% of our heavy snowfall cases could not be classified as any of the weather types of Bodolainé. *Hirsch* (2001) also faced this problem, and in addition to the 7 existing categories, he introduced a new weather type, which is responsible for most of the heavy snowfall events in Hungary.

Given our experience with other classifications, in this research we have decided to define a new weather type system by classifying only those cases that led to heavy snowfall in Budapest. There are two main methods of synoptic classification: subjective and objective. In the former case, classification is done by comparing fields of selected meteorological parameters and manually determining groups of similar synoptic patterns. In case of objective classification, an algorithm is used to separate cases into different groups, which can be done fully automatically. An example of an automated objective weather type classification was prepared by *Bissoli* and *Dittmann* (2001), who used advection of air masses at 700 hPa together with cyclonicity and humidity characteristics of the atmosphere to define their types. This method is used operationally at the German Weather Service. Cluster analysis methods have also been used to set up objective weather classification for many purposes. Climate research is one area, where these methods are often used. As an example of the numerous investigations, *Busch* and *Heimann* (2001) used cluster analysis based weather-type classification to extrapolate results of a regional climate model in time. *Hirsch* (2005) also applied cluster analysis technique to classify weather types causing heavy precipitation events in winter in Hungary.

In this research, however, we have decided to use subjective classification, with the possibility of applying objective methods to the same data set and comparing results of both investigations in the future.

2. Database and method

Research has been carried out on exactly those cases defined in Part I of the paper (*Babolcsai* and *Hirsch*, 2006). Heavy snowfall events (HSEs) were defined as cases of continuous snowfall producing a snow cover increase exceeding 8 cm at

the station Budapest/Lőrinc. 71 HSEs meeting this definition were found in the 50 winter periods (November to March) between 1953/54 and 2002/03.

Using the NCEP/NCAR Reanalysis dataset available online at http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html, temporally averaged fields of selected meteorological parameters were produced for the full duration of each event. Data was used in GRIB format; and grid-time averages were calculated for the following parameters:

- mean sea level pressure (MSLP),
- geopotential height at 500 hPa (H500),
- temperature at 850 hPa (T850),
- temperature at 925 hPa (T925),
- relative humidity at 700 hPa (RH700),
- total precipitable water (PW).

These parameters are widely used in operational weather forecasting. MSLP fields provide information about the weather systems near the surface, upper-level processes have been investigated using H500, while temperature conditions of the lower troposphere have been studied based on T850 and T925. Finally, RH700 fields are useful to learn about the main cloud level, while PW fields are essential to identify areas with high water content of the whole atmosphere. Due to temporal averaging, each field contains the whole process during the heavy snowfall event, rather than referring to only one instantaneous state. According to our results presented in Part I, most HSEs lasted for less than 1.5 days, and the highest duration reached 53 hours. Due to these relatively short time intervals, important features will not be filtered out from temporally averaged fields, and these fields are still characteristic of the whole duration of the events. Consequently, it is much more suitable to use time averages than only one field (e.g., the centre time of the event), which obviously does not describe all the processes that occurred during the event and might lead us to draw inappropriate conclusions.

3. Synoptic classification of heavy snowfall events in Budapest

Average fields of all the mentioned parameters for the area of Europe have been considered together in order to divide HSEs into different types. In our subjective classification, most emphasis has been placed on mean sea level pressure, but we have also attempted to find similarities among all the other fields. Using the opportunities provided by subjective classification, the similarities in the structure of the fields were regarded as most important, instead of concentrating on actual grid point values. For two parameters, mean

sea level pressure and 500 hPa height (which are directly linked to synoptic weather systems), the whole European area has been considered. The other meteorological parameters were mainly used to refine the classification, so in these cases we only focused on the region of Central Europe.

In the first step, HSEs were separated based on the mean sea level pressure pattern. Even using only this main parameter, it was possible to make clearly different groups. In the following, the classification has been made finer by analyzing the fields of other selected parameters. Our aim was to find the groups of events with similar patterns for as many parameters as possible. This is obviously not a definitive task, but the use of subjective classification makes it possible to consider different types of separation and to select the best classification among these in a complex way, which would be rather difficult or even impossible to solve using objective methods.

During this process, several sets of weather types were defined manually. Within each set, the difference of the mean characteristics of the weather types was determined. Eventually, the weather-type system was selected, for which the difference between the defined weather types was found to be the largest. The number of weather types was limited by our intention to define groups of HSEs, which are *significantly different* at synoptic and sub-synoptic scales. The following 8 weather types have been separated:

- Secondary low in the forward side of a Northwest European cyclone (NWp),
- Secondary low/frontal wave on the cold front of a cyclone with centre over Denmark (Dp),
- Warm front with frontal waves of a West European cyclone (Wwf),
- Forward side of a West European cyclone (Ww),
- Forward side of a Mediterranean cyclone (Mw),
- Passing Mediterranean cyclone (M),
- Cyclonic curvature on the southern edge of an anticyclone (Nc),
- Cold advection type (C).

Four of the 71 HSEs could not be classified. These were usually a mixture of some of the defined weather types, and it was decided not to include these in any of the categories. The frequency of each type in case of heavy snowfall events and their mean characteristics for Budapest are shown in *Table 1*. The cold advection type (C) is not included in this table, because it led to heavy snowfall only three times during the investigated 50 years, and the macrosynoptic situation was very different in each case. The reason for defining this group in spite of this large difference is the presence of continuous cold advection in all of the three cases, which makes these unique among other HSEs.

According to *Table 1*, the two weather types directly connected to Mediterranean cyclones (M and Mw) had the highest frequency and caused nearly 50% of all HSEs in the investigated 50 years. The number of cases related to the types Dp and Nc was also quite large, whereas the other types led to heavy snowfall in Budapest/Lőrinc less frequently. Clear differences can be seen in the mean characteristics of the defined weather types. When investigating HSEs classified as Pécely's macrosynoptic types, much smaller differences were determined, which shows that our new classification provides a better separation of HSEs. This is obviously due to the fact that Péczely's weather types were defined for all synoptic situations, whereas our classification involves only those cases when heavy snowfall occurred in Budapest/Lőrinc.

Table 1. Average values of the characteristics of heavy snowfall events for each weather type (excluding the C type) and the average of all cases for Budapest/Lőrinc

Characteristics	Weather types							
	NWp	Dp	Wwf	Ww	Mw	M	Nc	Mean
Number of cases	5	9	5	5	16	16	8	71
Duration (h)	20	19	26	15	20	22	24	20
Snow depth increase (cm)	13	16	15	15	13	19	14	15
Snowfall intensity (cm/h)	0.7	1.1	0.6	1.2	0.8	0.9	0.7	0.9
Precipitation intensity (mm/h)	0.6	1.0	0.6	0.9	0.7	0.7	0.5	0.7
Ratio of snow depth incr. to its water equiv. (cm/mm)	1.2	1.1	1.0	1.3	1.2	1.3	1.6	1.3
Mean sea level pressure (hPa)	1017	1001	1011	1005	1015	1003	1012	1009
Height of 500 hPa level (gpm)	5445	5305	5440	5350	5440	5360	5375	5380
Total precipitable water (mm)	12.4	11.2	13.6	11.6	11.9	11.3	9.8	11.6
2 m temperature (°C)	-2.6	-1.4	-3.4	-2.3	-2.3	-2.7	-4.4	-2.5
Temperature at 925 hPa (°C)	-1.4	-1.2	-3.4	-1.4	-3.8	-4.3	-7.9	-3.5
Temperature at 850 hPa (°C)	-2.8	-3.7	-3.6	-2.8	-5.1	-5.8	-8.5	-4.9

In the following, a detailed description of each weather type will be presented. For each type, we show the two fields which were found to be most characteristic of the given weather type. The figures were prepared using the Metview software developed by the European Centre for Medium-Range Weather Forecasts (ECMWF).

Secondary low in the forward side of a Northwest European cyclone (NWp)

Surface: A deep cyclone of large extent with centre west of the British Isles or over the Norwegian Sea and with a secondary low moving over the Adriatic Sea. Large blocking anticyclone over the East European Plain (*Fig. 1*).

500 hPa: Southwesterly flow on the forward side of a large trough over the Atlantic.

PW: Wedge-shaped flow of moist air from the southwest over the region of Hungary. Very dry air over the Alps and northeast of Hungary (*Fig. 1*).

850 hPa: Warm advection from the southwest, warm ridge over Hungary.

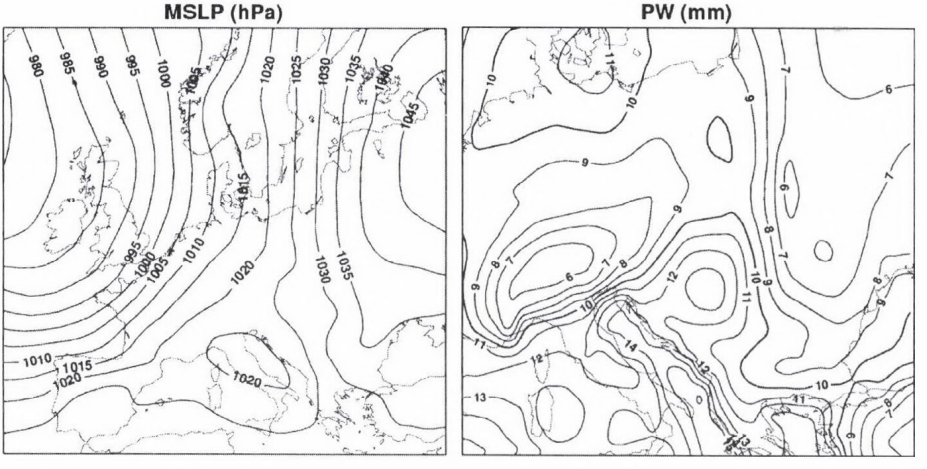


Fig. 1. Characteristic mean sea level pressure (MSLP) and precipitable water (PW) field in case of the NWp type.

Characteristics of HSEs connected to the NWp type:

- smallest snow depth increase,
- low intensity,

- significantly the highest mean sea level pressure,
- highest mean value of 500 hPa geopotential height,
- large values for total precipitable water,
- high temperatures at 925 hPa,
- highest temperatures at 850 hPa,
- characteristic of the middle of the winter period: cases occurred between December 28 and February 6.

Secondary low/frontal wave on the cold front of a cyclone with centre over Denmark (Dp)

Surface: Large low pressure area with a north-south axis reaching down to the North African coast. The cold front of the cyclone with center near Denmark generates a secondary low in the North Adriatic region, which usually moves south of Hungary towards the east, northeast (*Fig. 2*).

500 hPa: The fast propagation of the secondary low is caused by the strong south-westerly flow on the forward side of a trough west of Hungary (*Fig. 2*).

PW: Large gradient along a southwest-northeast axis crossing the northwestern part of Hungary with moist air southeast of this zone. Typical maximum PW values near the Bosporus.

850 hPa: This field may have rather diverse configurations depending on the location and depth of the trough.

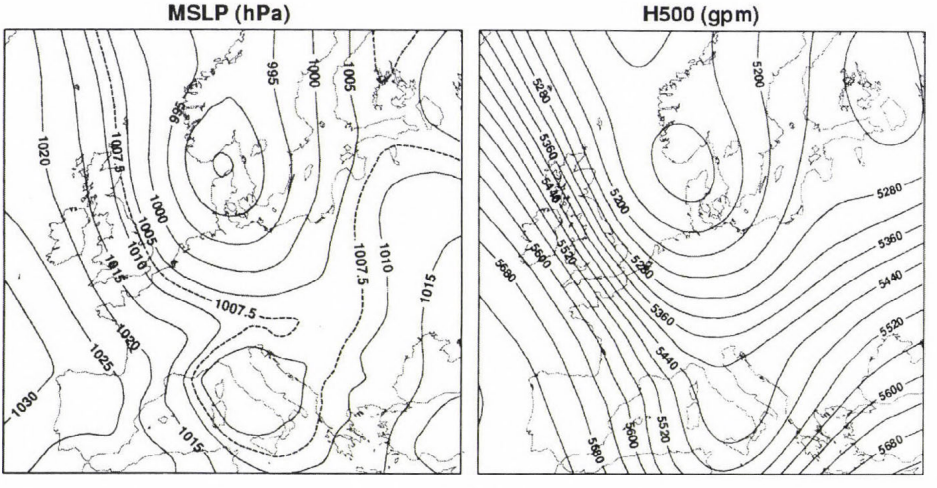


Fig. 2. Characteristic mean sea level pressure (MSLP) and 500 hPa geopotential height (H500) field in case of the Dp type.

Characteristics of HSEs connected to the Dp type:

- highest precipitation intensity (mm/h), intense snow accumulation,
- ratio of snow depth increase to its water equivalent is small,
- lowest mean sea level pressure,
- lowest mean value of 500 hPa geopotential height,
- highest temperature at 2 meter and 925 hPa level,
- often occurs at the beginning of the winter: 5 of the 9 cases were observed in the week between November 28 and December 4.

Warm front with frontal waves of a West European cyclone (Wwf)

Surface: The Carpathian Basin is situated on the forward side of a cyclone stretching from the Atlantic coast towards the middle part of the continent. Meanwhile, a strong anticyclone can be found in North and Northeast Europe, which leads to a strong pressure gradient northeast of Hungary on the border of the two systems. From Scandinavia, cold air is streaming southwards and meets mild airmasses of Atlantic origin resulting in a strong, stationary frontal zone over Central Europe (*Fig. 3*). *Spreitzhofer* (1999) also determined a similar weather type, as one of the situations causing heavy snowfall in Austria.

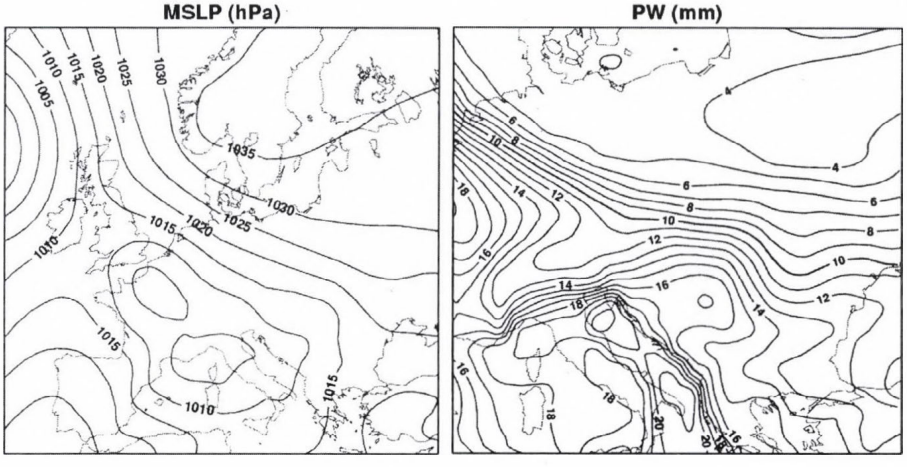


Fig. 3. Characteristic mean sea level pressure (MSLP) and precipitable water (PW) field in case of the Wwf type

500 hPa: Upper level low or trough over East or Northeast Europe.

PW: Extremely strong gradient can be observed over the Carpathian Basin with very dry air to the north and very most air south of Hungary (*Fig. 3*).

700 hPa: A band of high relative humidity stretching from west to east over Hungary clearly showing the sign of the sharp frontal zone.

850 hPa: Very strong temperature gradient along the mentioned frontal zone.

Characteristics of HSEs connected to the Wwf type:

- lasts significantly longer, small variability among the cases in their duration,
- least intense snow accumulation,
- ratio of snow depth increase to its water equivalent is the smallest,
- significantly highest total precipitable water,
- temperature below average at 2m, near average at 925 hPa, above average at 850 hPa, which leads to an approximately isothermal temperature profile in the lower troposphere,
- occurred twice in March.

Forward side of a West European cyclone (Ww)

Surface: The weather over the whole of Europe is determined by a huge cyclone centred over the western part of the continent. Hungary is situated on the forward side of the cyclone (*Fig. 4*).

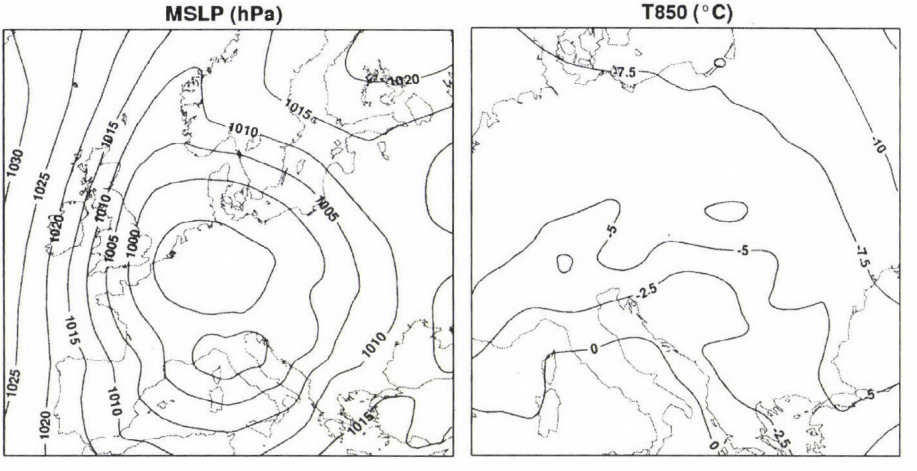


Fig. 4. Characteristic mean sea level pressure (MSLP) and 850 hPa level temperature (T850) field in case of the Ww type.

500 hPa: Strong southwesterly flow over Hungary and mostly closed isolines of geopotential height over Germany.

PW: Moist air from southwest stretching over Hungary.

850 hPa: Clear signs of strong warm air advection (*Fig. 4*).

Characteristics of HSEs connected to the Ww type:

- significantly the shortest duration,
- significantly the highest intensity,
- great variability in 500 hPa height,
- high temperature at 925 hPa,
- highest temperature at 850 hPa,
- obviously connected to the end of the winter period: all the 5 cases occurred in February or March.

Forward side of a Mediterranean cyclone (Mw)

Surface: Cyclone with centre over the western or central Mediterranean Sea with Hungary located in the warm sector of the cyclone during the whole event. The low pressure system hardly moves, which is caused by a blocking anticyclone over Northeast Europe or an upper level low directly above the surface cyclone (*Fig. 5*).

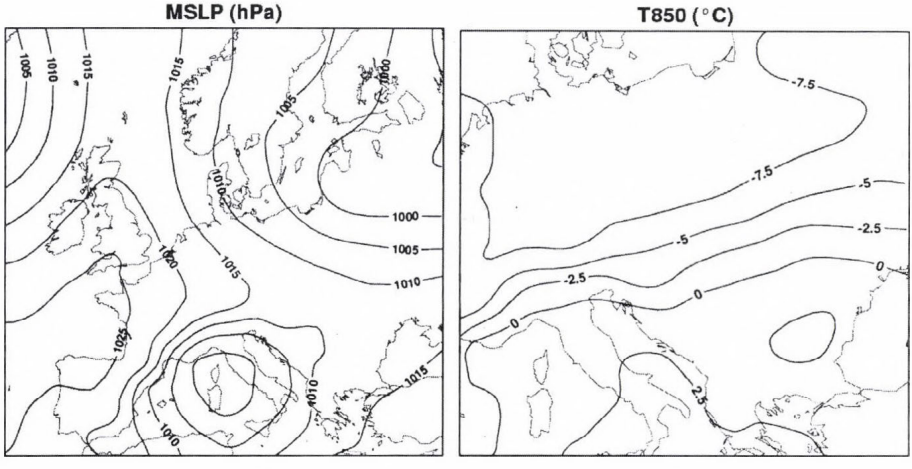


Fig. 5. Characteristic mean sea level pressure (MSLP) and 850 hPa level temperature (T850) field in case of the Mw type.

500 hPa: This field shows great variability. Closed isolines of geopotential height over the surface cyclone or southwesterly flow over Hungary on the forward side of a trough.

PW: This shows great variability as well. Most frequently, wedge-shaped flow of moist air from the southwest over the area of Hungary or moist air stretching much further to the northeast of the Carpathian Basin.

850 hPa: A band of strong temperature gradient usually stretching from southwest-west to northeast-east over the Carpathian Basin (*Fig. 5*).

Characteristics of HSEs connected to the Mw type:

- snow depth increase below average,
- high mean sea level pressure,
- geopotential height at 500 hPa is the highest,
- only 1 of the 16 cases occurred in January.

Passing Mediterranean cyclone (M)

Surface: Cyclone developed in the western part of the Mediterranean Sea moves in the Vb cyclone track over Hungary (*van Bebber*, 1891) or across the Balcan Peninsula. Usually anticyclone over Northwest or North Europe (*Fig. 6*).

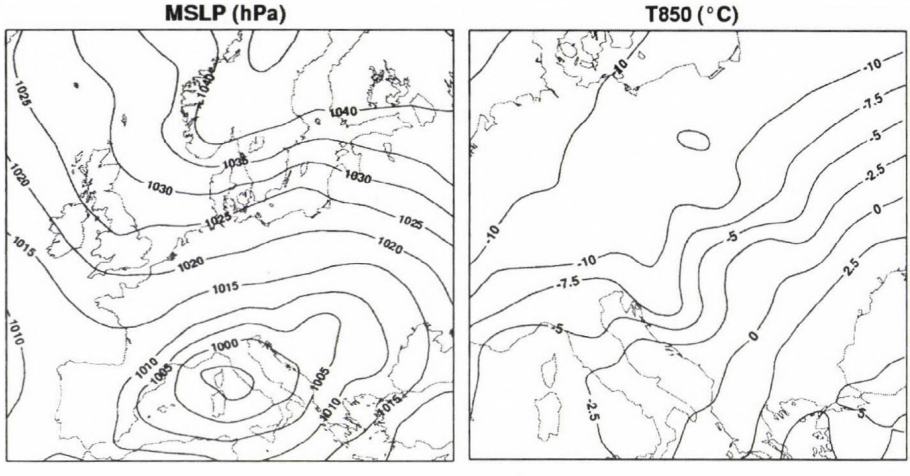


Fig. 6. Characteristic mean sea level pressure (MSLP) and 850 hPa level temperature (T850) field in case of the M type.

500 hPa: Very strong southwesterly flow on the forward side of a trough west of Hungary. Sometimes, closed isolines of geopotential height over North Italy.

PW: Flow of moist air towards the north, northeast over the Carpathian Basin.

850 hPa: Warm advection from the southwest on the forward side of the cyclone, gradually replaced by cold advection from the northeast in the rear of the low pressure system, as it moves eastwards (*Fig. 6*).

Characteristics of HSEs connected to the M type:

- snow depth increase significantly the largest,
- low mean sea level pressure,
- temperature below average at all levels,
- probability of its occurrence similar in all months except for March.

Cyclonic curvature on the southern edge of an anticyclone (Nc)

Surface: Anticyclone of large extent over Northeast Europe. Cyclone in the central and eastern Mediterraneans. Cyclonic curvature and easterly flow in the Carpathian Basin with convergence in the middle of Hungary, mostly along the Danube (*Fig. 7*).

500 hPa: Upper level low with variable location from case to case.

PW: Total precipitable water field can be quite variable depending on the distance of the mentioned Mediterranean cyclone.

850 hPa: Cold air of continental origin flows around the Carpathian Mountains and enters Hungary from northeast and northwest at the same time (*Fig. 7*).

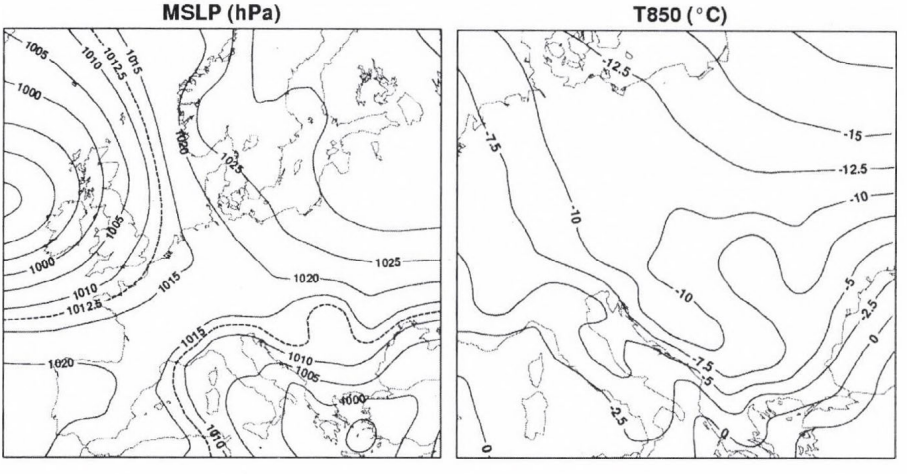


Fig. 7. Characteristic mean sea level pressure (MSLP) and 850 hPa level temperature (T850) field in case of the Nc type.

Characteristics of HSEs connected to the Nc type:

- usually lasts longer, but rather large variability in the duration of the cases,
- precipitation intensity is the smallest,
- significantly the largest ratio of snow depth increase to its water equivalent, but great variability,
- significantly the lowest total precipitable water,
- significantly the lowest temperatures at all levels,
- typical in the middle of the winter period: occurred only between December 12 and February 11.

Cold advection type (C)

There were only 3 examples of continuous cold advection during the whole HSE. Characteristic synoptic scale MSLP and 500 hPa level geopotential height field could not be determined for this type. What makes these cases similar, however, is the following. On the rearward side of a cyclone east, northeast, or southeast of Hungary, there is a cold surge in the Carpathian Basin from the north (*Fig. 8*). An upper level low or trough axis can be found near Hungary. In the averaged total precipitable water field, a pronounced back-bent configuration can be seen (*Fig. 8*). Because of the large differences and the small number of cases, mean characteristics have not been calculated for this type.

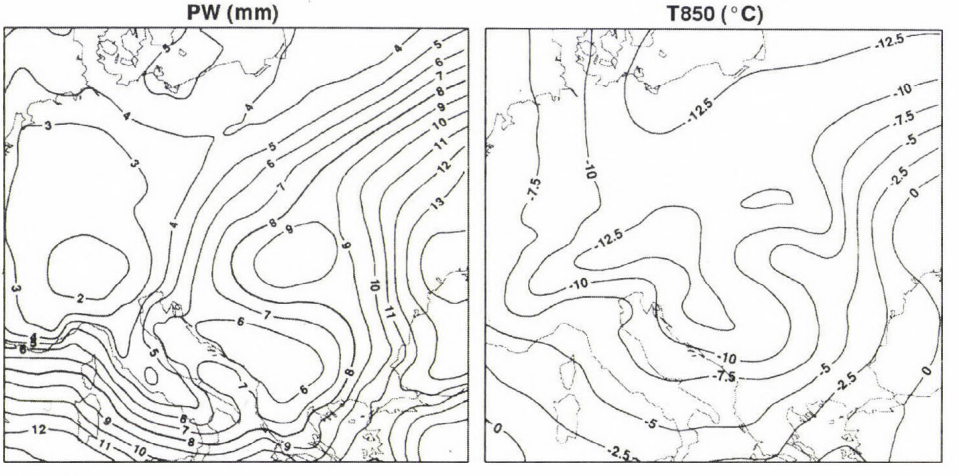


Fig. 8. Characteristic precipitable water (PW) and 850 hPa level temperature (T850) field in case of the C type.

Fig. 9 shows the mean vertical temperature profiles in the lower troposphere for the 7 weather types and for the average of all the 71 HSEs. There are 3 types with inversions (NWp, Dp, Ww). The mean temperature profile for the type Wwf is nearly isothermal. Cases occurring on the southern edge of a North European anticyclone (Nc) are by far the coldest.

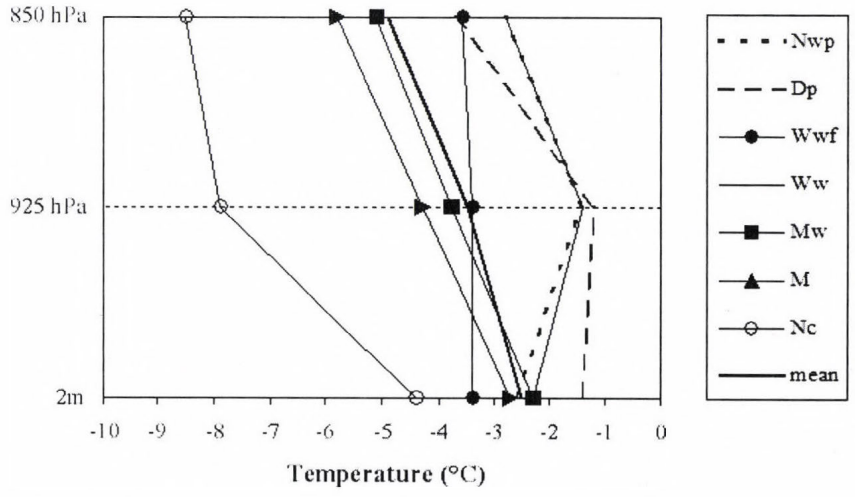


Fig. 9. Mean temperature profile of the different weather types and the average of all cases.

4. Case studies

In the following, some of the heavy snowfall events from the winter period 2005/2006 have been investigated. This winter period was not included in the 50 years (1953–2003), which served as the basis for our subjective classification. In this way, it is possible to analyze whether cases from an independent period can be classified as any of our weather types; this may also be regarded as a test of our classification.

The first case occurred on February 8, 2006 (*Fig. 10*). The mean sea level pressure field averaged for the duration of the event shows a cyclone with centre over Denmark and a secondary low southwest of Hungary, which clearly corresponds to our Dp type. The mean 500 hPa geopotential height field calculated for this case is also quite similar to the configuration that is characteristic of the Dp type, shown in *Fig. 2*. None of the HB or the Péczely types shows this very specific pattern, which is very favorable for heavy snowfall in Budapest, as shown by our research.

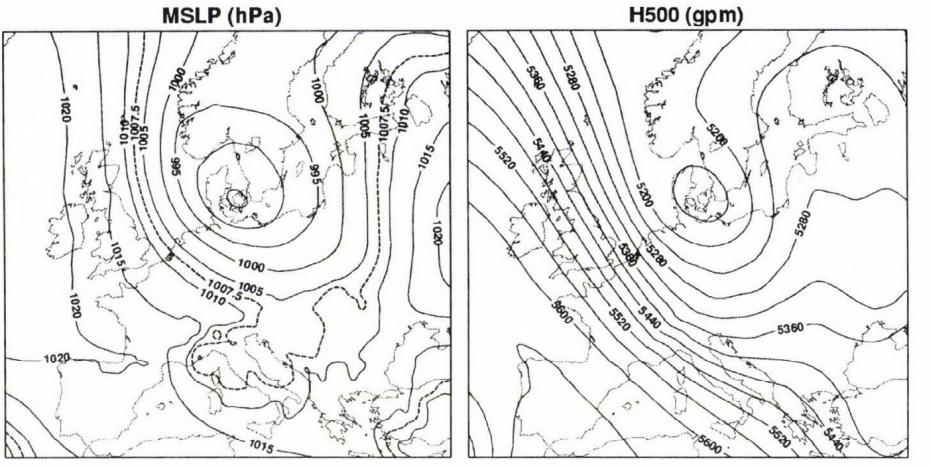


Fig. 10. Mean sea level pressure (MSLP) and 500 hPa geopotential height (H500) field averaged for the period February 8, 2006, 18:00 UTC–February 9, 2006, 06:00 UTC.

The temperature conditions in the lower troposphere were absolutely critical in terms of this heavy snowfall event. During the day, mild air had reached the line of the Danube from the northwest, whereas in eastern Hungary including Budapest, the temperature remained below freezing all day with easterly winds. A very strong temperature gradient reaching nearly 15°C had developed between the northeastern and southwestern part of country (*Fig. 11*). By the evening, the secondary low over North Italy had moved closer and in its pressure field the wind had turned to east over the whole country again. This stopped mild air over Transdanubia from stretching even more eastwards. Consequently, the type of heavy precipitation connected to the secondary low, was snow in the area of Budapest and eastern Hungary.

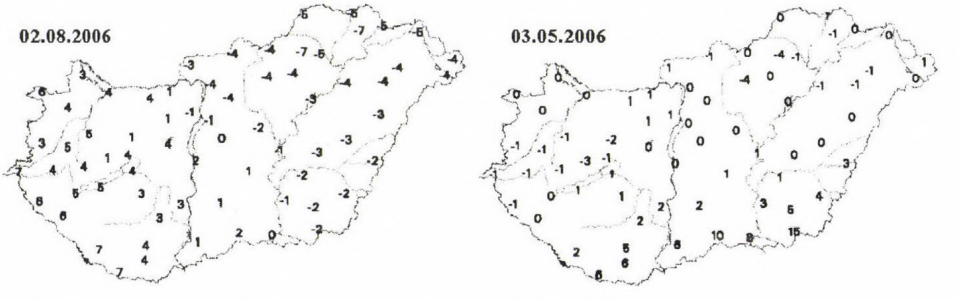


Fig. 11. Distribution of maximum temperature (°C) on February 8 and March 5, 2006.

Another heavy snowfall event occurred on March 5, 2006. According to the temporally averaged fields (*Fig. 12*), a Mediterranean cyclone moved from North Italy towards the southern part of the Carpathian basin. At 850 hPa, a very strong temperature gradient of about 12°C can be observed over Hungary with an axis of west, southwest to east, northeast. This case can be obviously classified as our M type, even if the axis of the temperature gradient usually shows a more pronounced southwest-northeast direction than in this particular case. At 2 metres, the mentioned temperature gradient was even stronger than at 850 hPa (*Fig. 11*). Similarly to the earlier case, Budapest was situated on the cold side of the strong frontal zone, resulting in snow instead of rain and sleet, which were observed south of Budapest.

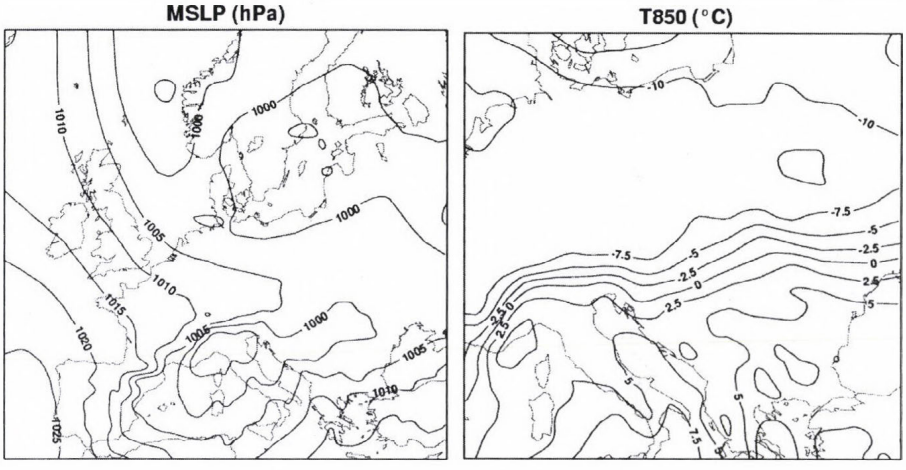


Fig. 12. Mean sea level pressure (MSLP) and 850 hPa level temperature (T850) field averaged for the period March 05, 2006, 06:00 UTC–18:00 UTC.

Mediterranean cyclones are not mentioned in any of the Hess-Brezowsky types. There are HB types where there is a possibility for the existence of Mediterranean cyclones, but these can be located in various places. As our results suggest, however, a specific location and pattern must exist for the development of HSEs in Budapest, which means that HB classification would not be appropriate. Similar statements can be made about Péczely types. Our second case could be classified as "Warm sector of a Mediterranean cyclone" using Péczely's classification. For this Péczely type, however, very different temperature conditions might occur, most of which would not be suitable for a heavy snowfall event.

5. Conclusion

In this research, heavy snowfall events occurring between 1953 and 2003 in Budapest have been investigated. Using temporally averaged reanalysis fields of selected meteorological parameters, subjective classification of the cases has been carried out. Besides usual parameters such as mean sea level pressure and 500 hPa geopotential height, which make it possible to define the macrosynoptic situation, other parameters have also been used to take into account the temperature conditions of the lower troposphere and processes connected to heavy precipitation.

As the result of this classification, 8 weather types have been defined. Nearly 50% of the cases were caused by two types, which are directly connected to Mediterranean cyclones. Furthermore, at least 3 of our other types also include Mediterranean cyclones. Consequently, it can be stated that the number of heavy snowfall events without some sort of Mediterranean cyclones is quite small.

Our weather types have been found to have quite different mean characteristics. For example, there are types which last longer on average, occur at significantly lower temperatures, produce more intense snowfall than other types or are limited to a particular part of the winter period. The specific characteristics of each type provide much more useful information for operational weather forecasting than the overall characteristics presented in Part I of the paper. Due to the relatively large differences between the types in many aspects, the knowledge of these characteristics can be efficiently used when preparing operational weather forecast, e.g., by modifying numerical weather prediction model results.

Two heavy snowfall events from an independent winter period have also been presented as case studies. Each of these could easily be assigned to one of our weather types, showing the success of our classification. These case studies were also useful to show why other general classifications are less appropriate to apply to very specific cases like heavy snowfall events. As far as future plans are concerned, it would be worth investigating cases when one of our weather types occurred without causing heavy snowfall. This may reveal other processes which are essential to heavy snowfall, but have not so far been considered. Also, developing an objective procedure for identifying our weather types would be extremely useful and could enable a whole series of further investigations in this area.

Acknowledgements—The authors are grateful for the possibility of using NCEP Reanalysis data provided by the NOAA-CIRES ESRL/PSD Climate Diagnostics branch, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov. We would also like to thank *David Richardson* (Head of Meteorological Operations Section at ECMWF) for his linguistic revision of this paper.

References

Babolcsai, Gy. and *Hirsch, T.*, 2006: Characteristics and synoptic classification of heavy snowfall events in Budapest for the period 1953–2003. Part I. *Időjárás 110*, 1-13.

Baur, F., Hess, P., and *Nagel, H.*, 1944: *Kalender der Grosswetterlagen Europas 1881-1939*. Bad Homburg v. d. H.

Bissoli, P. and *Dittmann, E.*, 2001: The objective weather type classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations. *Meteorol. Z. 10*, 253–260.

Bodolainé, J.,E., 1983: Synoptic conditions of flood waves on the basin of the rivers Danube and Tisza (in Hungarian). Országos Meteorológiai Szolgálat Hivatalos Kiadványai, LVI., Budapest.

Gerstengarbe, F.W., Werner, P.C., Busold, W., Rüge, U., and *Wegener, K.O.*, 1993: : Katalog der Grosswetterlagen Europas nach Paul Hess und Helmuth Brezowsky 1881-1992. 4. vollständig neu bearbeitete Auflage. Ber. Dt. Wetterd. 15 (113).

Hess, P. and *Brezowsky, H.*, 1952: Katalog der Grosswetterlagen Europas. Ber. Dt. Wetterd. in der US-Zone 33.

Hess, P. and *Brezowsky, H.*, 1969: Katalog der Grosswetterlagen Europas. 2. neu bearb. u. erg. Auflage Ber. Dt. Wetterd. 15 (113).

Hess, P. and *Brezowsky, H.*, 1977: Katalog der Grosswetterlagen Europas 1881-1976. 3. verb. u. erg. Aufl. Ber. Dt. Wetterd. 15 (113).

Hirsch, T., 2000: Synoptic-climatological investigation of weather systems causing heavy precipitation in winter in Hungary. *Időjárás 104*, 173-196.

Hirsch, T., 2005: Objective classification of weather systems causing heavy precipitation in winter in Hungary using ERA40 data. ECAM 2005 Conference, September 12-16, 2005, Utrecht, Netherlands.

Péczely, Gy., 1957: Großwetterlagen in Ungarn. Kleinere Veröff. Zentralanst. Meteorol. Budapest.

Péczely, Gy., 1983: Catalogue of the macrosynoptic types for Hungary, 1881-1983 (in Hungarian). Országos Meteorológiai Szolgálat Kisebb Kiadványai, 53, Budapest.

Spreitzhofer, G., 1999: Synoptic classification of severe snowstorms over Austria. *Meteorol. Z. N. F. 8*, 3-15.

van Bebber, W.J., 1891: Die Zugstraßen der barometrischen Minima nach den Bahnenkarten der Deutschen Seewarte für den Zeitraum von 1875–1890. *Meteorol. Z. 8*, 361–366.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 175–182

Short communication

Estimating global radiation using the meteorological input data of crop models

Nándor Fodor

Research Institute for Soil Science and Agricultural Chemistry
of the Hungarian Academy of Sciences
Herman O. út. 15, H-1022 Budapest, Hungary; E-mail: fodornandor@rissac.hu

(Manuscript received in final form April 18, 2006)

Abstract—Although not measured at many meteorological stations, the daily global radiation at the earth's surface is a very important component of ecosystem mass and energy processes and so it is in crop modeling. The lack of radiation data is a limitation to the use of crop models. The original and an improved form — that takes precipitation data into account — of the Bristow-Campbell solar radiation estimation method were investigated regarding their performance for providing radiation estimates as crop model input. While the original method did not gave acceptable radiation estimations for the used crop model, the improved method did. With an additional site specific calibration, the new method gave so good radiation estimates that the average errors of simulated yield and cumulative evapotranspiration could be decreased below 2% and 0.5%, respectively.

Key-words: radiation estimation, Bristow-Campbell method, crop model, yield prediction

1. Introduction

The primary purpose of crop models is to describe the processes of the very complex atmosphere-soil-plants system using mathematical tools and to simulate them with the help of computers. The ultimate aim of using crop models, however, is to answer questions that otherwise could only be answered by carrying out expensive and time-consuming experiments.

Although not measured at many meteorological stations, the daily global radiation at the earth's surface is a very important component of ecosystem mass and energy processes and so it is in crop modeling. The minimum dataset

for many crop models includes daily solar radiation, minimum and maximum temperature, and precipitation data. Unlike temperature and precipitation, solar radiation is recorded only at few weather stations in the United States (*Ball et al.*, 2004) and so it is in Hungary. This lack of radiation data can be a major limitation to the use of crop models. In order to use crop models, techniques are required to estimate radiation based on other commonly measured meteorological variables such us temperature and precipitation. Thus, crop modelers have a share in developing radiation estimation methods.

There are two groups of methods used to generate radiation data: stochastic generation and empirical relationships (*Liu* and *Scott*, 2001). Since it was found (*Hayhoe*, 1998) that empirical methods using the common meteorological input data of crop models (temperature and precipitation) provided better estimates than stochastic methods, in this paper only techniques based on empirical relationships are discussed.

There are estimation methods to calculate daily global radiation using different input data: *Ångström* (1924), *Szász* (1968), *Hargreaves* and *Samani* (1982), *Bristow* and *Campbell* (1984), *Fodor et al.* (2000), *Donatelli* and *Bellocchi* (2001). The first two methods use the daily sum of sunshine hours, while the others use the daily thermal oscillation. In an earlier study, Szász's method was found to give radiation estimates good enough to substitute measured radiation for crop models (*Fodor et al.*, 2003a). However, the applicability of this method is limited, since it uses the daily sum of sunshine hours that is measured at much fewer weather stations than temperature. The focus of this study is to investigate whether the popular and widely used Bristow-Campbell method (*Bristow* and *Campbell*, 1984) is able to provide global radiation estimates good enough to substitute the measured global radiation in the 4M (*Fodor et al.*, 2003b) crop simulation model.

2. Materials and methods

Recent methods usually determine solar radiation (R) at the earth's surface as a product of extraterrestrial radiation (R_E) and atmospheric transmissivity (A_T):

$$R = R_E \times A_T . \tag{1}$$

Extraterrestrial radiation can be easily calculated for any given day of the year knowing the solar constant and the latitude of the site in question. The estimation methods differ in the way they determine atmospheric transmissivity. The Bristow-Campbell (BC1) method (*Bristow* and *Campbell*, 1984) calculates the transmissivity as follows:

$$A_T = a \times (1 - \exp(-b \times D^c)), \tag{2}$$

where

D is the smoothed temperature difference (°C): T_{max}–0.5 × (T_{min}+T_{min}^{tom}),

T_{max} is the maximum daily temperature (°C),

T_{min} is the minimum daily temperature (°C),

T_{min}^{tom} is the minimum daily temperature for tomorrow (°C),

a, b, c are parameters, determined by using an optimization method minimizing RMSE (Root Mean Square Error) between the measured and predicted radiation.

We introduce a modified Bristow-Campbell (BC2) method which takes the daily precipitation into account. This method uses the same equations (Eqs. (1)–(2)), but with different parameters for dry and wet (precipitation greater than zero) days:

$$A_T = a_p \times (1 - \exp(-b_p \times D^{cp})), \tag{3}$$

$$A_T = a_{np} \times (1 - \exp(-b_{np} \times D^{cnp})), \tag{4}$$

where a_p, b_p, c_p, a_{np}, b_{np}, c_{np} are parameters, determined by using an optimization method minimizing RMSE between the measured and predicted radiation. The p index stands for days with precipitation, np index stands for days with no precipitation.

The National Weather Service of the USA provides detailed weather data for several hundred sites for the country. Data of 238 weather stations were prepared for crop models and published on the internet: http://nowlin.css.-msu.edu/indexritchie.html. This database contains the daily global radiation, minimum and maximum temperature, precipitation and relative humidity data from 1961 to 1990. Not every record of the database was used for calibrating both Bristow-Campbell methods but only those years that were similar to the Hungarian ones in certain aspects. A set of agrometeorological characteristics was established (*Table 1*) to find 'Hungary-like' data in the database. The ranges in *Table 1* were slightly larger than the Hungarian averages (*Péczely*, 1979) to obtain more years for calibration. Using these conditions 81 years from 15 different sites were selected. These records were used to calibrate both BC1 and BC2 methods.

Based on Eqs. (1) and (2), the ratio of the daily global and the extraterrestrial radiation can be expressed as a 3-parameter function of the smoothed temperature difference:

$$R/R_E = a \times (1 - \exp(-b \times D^c)). \tag{5}$$

Table 1. Agrometeorological characteristics for selecting 'Hungary-like' data from the American database

Characteristic	Minimum	Maximum
Elevation (m)	0	800
Distance from larger body of water (km)	200	-
Annual average temperature (°C)	7.0	12.5
Annual cumulative solar radiation (MJ m^{-1} $year^{-1}$)	4200	5200
Number of days with precipitation ($year^{-1}$)	120	160
Annual cumulative precipitation (mm)	400	850

Three files, each containing two data columns (R/R_E as dependent variable and D as independent variable) were created using the weather data of the fifteen selected stations. One file contained data for every day, one for days with precipitation, and one for days without precipitation. The a, b, c, a_p, b_p, c_p, and a_{np}, b_{np}, c_{np} parameters were determined by nonlinear regression (*Marquardt*, 1963) using the corresponding data files. After defining the parameters, the BC1 and BC2 methods were validated on an independent dataset. Global radiation was estimated for Budapest/Lőrinc, Hungary, where the measured daily radiation, the minimum and maximum temperature, and precipitation data were available from 1968 to 1987. It was the only available dataset for verifying the estimation methods for Hungary. The adjusted R^2 of the regression, RMSE, bias (mean signed error, MSE), and relative error were used as measures of model performance.

Both measured and simulated radiation were then used in the 4M crop simulation model (*Fodor et al.*, 2003b). First the model was run with measured radiation, then with estimated radiation given as input. The calculated yield and cumulative evapotranspiration outputs were recorded every year between 1968 and 1987. Model results obtained by using measured and estimated radiations were compared. Since global radiation indirectly affects the water balance of the soil, three soil profiles with different water regimes were selected for the model runs. The comparison was carried out for a chernozem, a meadow soil, and a brown forest soil profile (*Várallyay et al.*, 1994), since the vast majority of the agricultural lands in Hungary are covered by these three soil types. The soil data was provided by the Research Institute of Soil Sciences and Agricultural Chemistry (*Rajkai et al.*, 1981; *Várallyay*, 1987). The required genetic parameters of the maize cultivar were retrieved from the DSSAT database (*Tsuji et al.*, 1994) and were used as crop specific model inputs. Each run started on March 1. The initial water content of the soil profiles was set to 80% of the field capacity.

A recent study on the sensitivity of crop models to the inaccuracies of meteorological observations (*Fodor* and *Kovács*, 2005) showed that the uncertainty caused by the systematic errors of the measured global radiation can be up to 5% for the calculated yield. This threshold (acceptance limit) was used for deciding whether the radiation estimation is acceptable for the crop model or not. If the difference between the model results obtained by using estimated radiation and the ones obtained by using measured radiation is less than 5%, the radiation estimation is said to be acceptable.

3. Results and conclusions

After calibrating the BC1 and BC2 methods,both gave fairly good estimates for the independent Hungarian dataset (*Table 2*). Taking the precipitation into account (BC2) made a slight improvement in radiation estimation. The expected value of error of the estimated daily global radiation is 2.52 ± 0.078 MJ m^{-2} and 2.28 ± 0.070 MJ m^{-2} ($\alpha=0.05$), for the BC1 and BC2 methods, respectively. Both methods tend to overestimate (MSE>0) the measured solar radiation (*Table 2*).

Table 2. Performance indicators of the estimation methods, comparing the measured and estimated radiation for Budapest/Lőrinc, Hungary

Method	R^2	RMSE	MSE*	Relative error**
BC1	0.827	3.390	0.161	0.269
BC2	0.861	3.060	0.351	0.241

* Mean Signed Error — Bias
** In the summer half-year

Note that using the Budapest/Lőrinc dataset for (site specific) calibration, the bias could almost totally be eliminated from the estimation and the absolute error decreased with an additional 10 percent. In this case the expected value of the error of the estimated daily global radiation is 2.34 ± 0.051 MJ m^{-2} and 1.97 ± 0.045 MJ m^{-2} ($\alpha=0.05$), for the BC1 and BC2 methods, respectively.

The statistical evaluation of the methods can not judge whether the estimated radiation is good enough for crop models or not. Simulation results obtained by using measured and estimated radiation and their evaluation are presented in *Fig. 1* and in *Table 3* for BC1 and BC2 methods. The simulated cumulative evapotranspiration is not discussed here in detail, since the average error for this model output was way below the acceptance limit 0.7–2.4%, depending on the estimation method and soil type.

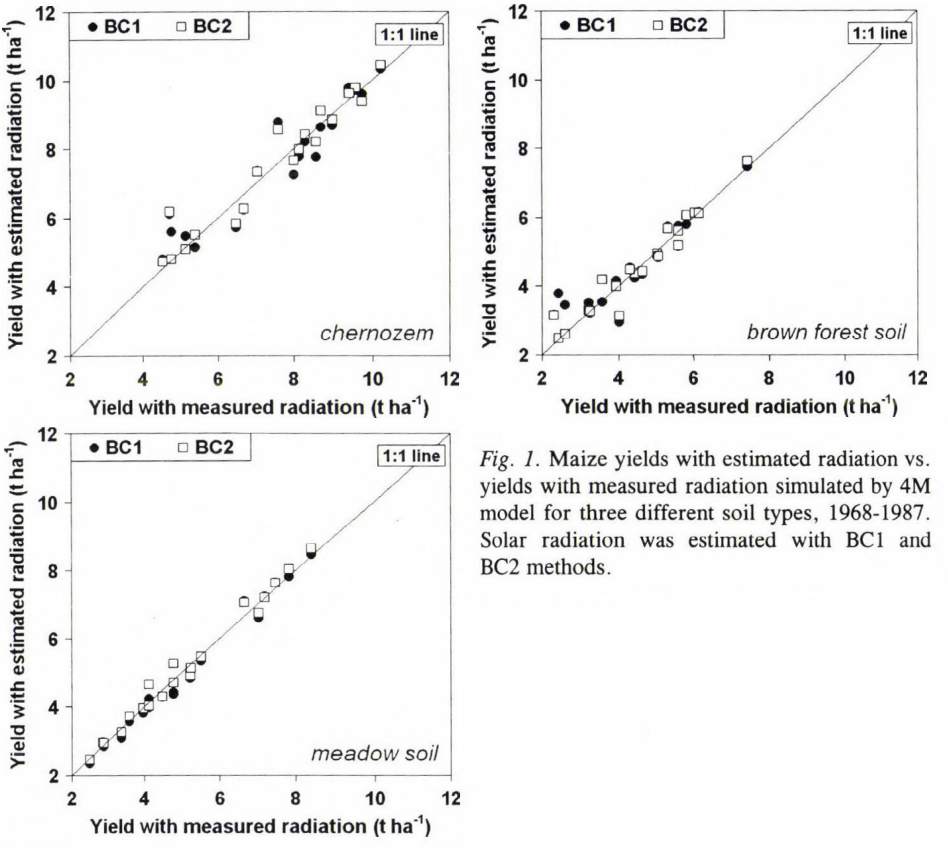


Fig. 1. Maize yields with estimated radiation vs. yields with measured radiation simulated by 4M model for three different soil types, 1968-1987. Solar radiation was estimated with BC1 and BC2 methods.

The year of 1986 is a very interesting example with an extremely high difference between the simulated yields (*Table 3*) on chernozem soil. During the early grain filling period, the (virtual) plants had severe water stress. Because of the slightly higher estimated radiation values, on the August 7 the plants getting the estimated radiation died, while the plants getting the measured radiation barely survived the day. The 23 mm of rain on the very next day (and an other 24 mm 10 days later) was enough for these plants to finish the grain filling period normally.

Apart from this, taking the precipitation data into account enhanced the quality of estimated radiation data as a crop model input (*Table 3*). The average errors of yield prediction using the estimations of the BC2 method were 10–35% smaller than those obtained by using the estimations of the BC1 method. While the BC1 method did not give acceptable radiation estimations for the 4M crop model, the BC2 method did. The average errors of yield prediction were smaller than the 5% acceptance limit for all of the investigated

soil types. With an additional site specific calibration, the new method gave so good radiation estimates that the average errors of simulated yield and cumulative evapotranspiration could be decreased below 2% and 0.5%, respectively.

Table 3. Comparison of simulated maize yields using measured and estimated solar radiation for three different soil types, for the 1968–1987 period. Radiation was estimated with BC1 and BC2 methods

Characteristic	Unit	BC1			BC2		
		Chernozem	Brown forest soil	Meadow soil	Chernozem	Brown forest soil	Meadow soil
Maximum difference	kg ha^{-1}	4165*	1351	442	4185*	905	520
Average absolute difference	kg ha^{-1}	655	346	201	543	227	178
SD of the differences	kg ha^{-1}	905	383	147	922	270	147
Proportion of years with acceptably good radiation estimations	%	50	50	65	75	75	80
Expected value of the difference (α=0.05)	%	5.7 ± 2.7	7.7 ± 2.0	4.9 ± 2.0	3.9 ± 2.0	4.5 ± 1.2	3.6 ± 1.5

*In the simulation in 1986, plants died due to severe water stress using estimated radiation

Consequently, an effective solar radiation estimation method could be developed for providing this very important crop model input by using the measured data of some meteorological stations covering Hungary. This method would enable the crop modelers to use their models on the sites where it was impossible beforehand due to a lack of measured solar radiation data. As long as the parameterization is not carried out using a larger Hungarian dataset, we suggest using the improved Bristow-Campbell method (BC2) with the following parameters for estimating radiation for crop models:

Days with precipitation	Days without precipitation
a=10.966	a=6.514
b= 0.01714	b=1.03380
c= 0.34652	c=0.40343

For further study we suggest using the data of more meteorological stations from Hungary to investigate the general applicability of the method for the whole country. Additional details on temperature (more than the daily maximum and minimum) and precipitation data (rainfall intensity, etc.) might be a point of interest in enhancing the estimation method.

Acknowledgement—This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the grant of OTKA F046465.

References

Ångström, A., 1924: Solar and terrestrial radiation. *Q. J. Roy. Meteorol. Soc. 50*, 121-125.

Ball, R.A., Purcell, L.C., and *Carey, S.K.*, 2004: Evaluation of solar radiation prediction models in North America. *Agron. J.* 96, 391-397.

Bristow, R.L. and *Campbell, G.S.*, 1984: On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agr. Forest Meteorol. 31*, 159-166.

Donatelli, M. and *Bellocchi, G.*, 2001: Estimate of daily global solar radiation: new developments in the software RADEST 3.00. *Proc of the. 2nd International Symposium Modelling Cropping Systems*. Florence, 16-18 July, 2001, Italy.

Fodor, N., Kovács, G.J., and *Ritchie, J.T.*, 2000: A new solar radiation generator for Hungary. Poster. *ASA-CSA-SSSA, Annual Meetings*. November 5-9, 2000, Minneapolis, MN., Abstract. 23 p.

Fodor, N., Kovács, G.J., and *Pokovai, K.*, 2003a: Reliability of estimated global radiation for crop model input. *Időjárás 107*, 273-281.

Fodor, N., Máthéné-G., G., Pokovai, K., and *Kovács, G.J.*, 2003b: 4M - software package for modelling cropping systems. *Eur. J. Agron. 18*, 389-393.

Fodor, N. and *Kovács, G.J.*, 2005: Sensitivity of crop models to the inaccuracy of meteorological observations. *Phys. Chem. Earth 30, 165-170.*

Hargreaves, G.H. and *Samani, Z.A.*, 1982: Estimating potential evapotranspiration. *J. Irrig. Drain. Eng. 108*, 225-230.

Hayhoe, H.N., 1998: Relationship between weather variables in observed and WXGEN generated data series. *Agr. Forest Meteorol. 90*, 203-214.

Liu, D.L. and *Scott, B.J.*, 2001: Estimation of solar radiation in Australia from rainfall and temperature observations. *Agr. Forest Meteorol. 106*, 41-59.

Marquardt, D.W., 1963: An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math. 11*, 431-441.

Péczely, Gy., 1979: Climatology (in Hungarian). Nemzeti Tankönyvkiadó, Budapest.

Rajkai, K., Várallyay, Gy., Pacsepszkij, J.A., and *Cserbakov, R.A.*, 1981: Calculation of water retention data from the texture and the bulk density of soils (in Hungarian). *Agrokem. Talajtan 30*, 409-438.

Szász, G., 1968: Determining the global radiation by means of calculation (in Hungarian). *Debreceni Agrártudományi Főiskola Tudományos Közleményei* XIV, 239-253.

Tsuji, G.Y., Uehara, G., and *Balas, S.*, (eds.), 1994: *Decision Support System for Agrotechnology Transfer (DSSAT) v3* Vol. 2. University of Hawaii, Honolulu, Hawaii.

Várallyay, Gy., 1987: Water regime of soil (in Hungarian). Vol. 2. Appendix 1. *DSc Thesis*. Magyar Tudományos Akadémia, Budapest.

Várallyay, Gy., Szabó, J., Pásztor, L., and *Michéli, E.*, 1994: SOTER (Soil and Terrain Digital Database) 1:500,000 and its application in Hungary (in Hungarian). *Agrokem. Talajtan 43*, 87-108.

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 110, No. 2, April–June 2006, pp. 183–189

Short communication

Comparison of values of the chosen meteorological fields measured at the aerological stations and the values taken from NCEP/NCAR Reanalysis

Jadwiga Woyciechowska and **Rafał Bąkowski**

Institute of Meteorology and Water Management
61 Podlesna st., 01-673 Warszawa, Cracow Branch, Poland
E-mail: ziwoycie@cyf-kr.edu.pl

(Manuscript received in final form December 18, 2005)

Abstract—Statistical relationship between the air temperature and the geopotential height of the pressure levels measured at the air-sounding station in Poland and taken from NCEP/NCAR Reanalysis for the same location was analyzed, based on data from the 30-year period of 1974–2003. It was shown, that NCEP/NCAR Reanalysis data is fully complementary and valuable data series, particularly in the lower and middle troposphere, and it can be used along with direct measurements (meteorological sounding of the atmosphere).

Key-words: correlation coefficient, reanalysis, sounding, air temperature, geopotential height

1. Introduction

At present, the climatology benefits more and more from the free atmosphere data when analyzing the fields of meteorological elements. The data are taken from the direct measurements, i.e., aerological soundings of the atmosphere or the mean values of the particular meteorological parameters derived from the geographical grid. These derived products are the result of interpolation. The values in the relevant grid points can vary depending on the algorithm routine used for calculation.

The directly measured radiosounding data have gaps, and it leads to the question on the possibilities of data completing by the measurements from reanalysis calculated for the grid nodes (*Marshall*, 2002a,b; *Gaffen et al.*, 2000). In the scientific centers (e.g., NCAR, NOAA, ECMWF), the reanalysis data products are available for numerous meteorological fields on

various pressure levels in data grids with different spatial and time resolution covering the Earth. Therefore, it seems reasonable to do the comparative analysis of direct measurements and reanalysis (*Marshall*, 2002a; *Atkinson* and *Solomon*, 2003).

2. Aims and methods

The aim of the survey was to check the usefulness of the NCEP/NCAR Reanalysis data products in meteorological and climatological studies using free atmosphere data. This objective was fulfilled by the comparison of values of the chosen meteorological fields, measured at the aerological stations: Legionowo, Leba, Wroclaw (where the soundings of the atmosphere are carried out), and the values of the same fields at the same geographical locations taken from NCEP/NCAR Reanalysis. In addition, the results of such study provides an answer on compatibility of the NCAR (in US) Reanalysis and the sounding data from Polish stations. It also could indicate if NCEP/NCAR Reanalysis data are useful as complementary sounding data series.

NCEP/NCAR Reanalysis data, that becomes more and more often used in studies and analysis, the results of which are discussed and presented in numerous papers, deserve recognition. Thereby, some words about the archive should be placed here. NCEP/NCAR Reanalysis is a project to produce analysis of atmospheric fields. NCEP GTS (Global Telecommunications System) data (including pilot-balloons and aircraft) with upper-air observations is the main source of data for reanalysis. The rocket sounding data from national archives in various countries are also provided. Surface marine data from the COADS (Comprehensive Ocean-Atmosphere Data Set) with ships, fixed, drifting, pack-ice buoys, and near-surface data from ocean station are taken. Surface land synoptic data, satellite sounder data, SSM/I (Special Sensing Microwave/Imager) data, and satellite cloud drift winds are used as the source of data in the project as well.

In the preprocessing reanalysis module, data from many different sources are transformed into a BUFR (binary universal format representation) format, and the surface boundary conditions are prepared. The data assimilation module contains the system configuration, analysis scheme (a three-dimensional variational analysis scheme, spectral statistical interpolation), global spectral model, CQC (complex quality control) of rawinsonde data, OIQC (optimal interpolation quality control) of all data, BUFR observation "events" (a spectrum of processing information) files, optimal averaging and periodic forecasts from the reanalysis. The reanalysis gridded fields are the major project results. These fields, depending on the influence of the observational data and the model, are classified into four classes. Class A, the most reliable one, is the class of variables strongly influenced by observations,

B is for variables which are strongly influenced by the model as well. There are no observational data directly affecting the C class fields values. Variables, indicated with D are obtained from climatological values, do not depend on the model. More details on the NCEP/NCAR Reanalysis project can be found in *Kalnay et al.* (1996). Temperature and geopotential height fields considered in this paper are both class A variables.

Temporal and spatial variations of the geopotential heights and temperatures at the standard WMO pressure levels (850 hPa, 700 hPa, 500 hPa, 300 hPa, 250 hPa, 200 hPa, 150 hPa, 100 hPa, 70 hPa, and 50 hPa) were taken into consideration in our study for two sets of data. One data set consisted of the aerological soundings from TEMP data and the second of the NCEP/NCAR Reanalysis range, respectively. The data covered the period from 1974 to 2003 at 00:00 UTC for both data sets. In spite of representing about 90% of observation, on account of sounding measuring techniques, direct measured sets of atmospheric data for the upper levels in particular are incomplete. NCEP/NCAR Reanalysis data (e.g., monthly means of meteorological elements fields used for the purpose of this paper) are available from the archives for the geographical regular 2.5° × 2.5° gridded surface and represent homogeneous series. For this study needs, the data sets for geographical co-ordinates of aerological stations were interpolated from the nearest grid nodes. The values of air temperature and geopotential height for Leba, Legionowo, and Wroclaw were interpolated from the nearest grid points on the given pressure level using two methods: first order linear interpolation and inverse squared weighted interpolation. In both cases the calculations were made under the assumption that meteorological fields, as the continuous function of longitude and latitude, do not have the local extremes in the considered intervals (in longitude and latitude). In addition, the constraint was set that there is no extreme inside the given window (i.e., the convex hull created by the mentioned four points).

Since the Earth's radius a is far larger than the spatial grid resolution, and the angular distance between the station and the nearest gridpoint in longitude or latitude is less than 2.5°, the Earth's curvature can be neglected, and the linear distance from the point at the height of h over the station to the grid node at the same height can be expressed as

$$\sqrt{(a+h)^2 \sin^2(\Delta\varphi) + b^2 \sin^2(\Delta\lambda)}, \tag{1}$$

where the $(a+h)\sin(\Delta\varphi), b\sin(\Delta\lambda)$ terms are for meridional and zonal distances, respectively, $\Delta\varphi$ and $\Delta\lambda$ are the angular distances (in latitude and longitude) between the station and the given grid node. a is the distance between the desired point (over the station) and axis of the Earth's, assumed the same for all relevant points (i.e., station and nearest neighbour(s))

The sine of a small angle is (approximately) the angle itself measured in radians (e.g., $25° = 0.04363 rad$, $\sin(0.4363 rad) = 0.04362$). Thus the interpolation can be simplified significantly. The result of the first order linear interpolation and the result of the inverse squared weighted interpolation do not vary. Finally, two series of monthly means for analysis and comparison were archived one is based on direct aerodynamical sounding at 00:00 UTC, the other is founded on interpolation (of NCEP/NCAR Reanalysis) and averaged over the days for which the data from aerodynamical/aerological soundings (at 00:00 UTC) were available.

3. Results

Analyzing statistical relationship between the direct measured air temperature in the free air and temperature based on NCEP/NCAR Reanalysis in geographical co-ordinates of the particular sounding station revealed the strong positive correlation between the series, especially below the 500 hPa pressure level (*Figs. 1a, 2a,* and *3a*). The largest values of the linear correlation coefficient r are observed in the lower troposphere. At the 850 hPa pressure level the values of coefficient r are greater than 0.9 (values of coefficient r were obtained at standard 95% confidence level) for all months of the year except for January over Leba ($r=0.86$). Higher, at the 700 hPa isobaric surface the r values are higher than 0.9, except for May over Wroclaw ($r=0.86$).

Above, in the middle troposphere (500 hPa) the values of r are lower in general, but still greater than 0.9, except for August over Wroclaw ($r=0.86$) and August over Leba ($r=0.88$).

In the upper troposphere (at the 300 hPa and 250 hPa isobaric surfaces) the correlation decreases. The correlation coefficient r drops to 0.4 for certain months (Wroclaw in June, Leba in August and September). The seasonal dependence of correlation is obvious in this part of troposphere with the lowest values of correlation coefficient r in summer and autumn on the 250 hPa level, and in winter and spring on the 300 hPa level. However, such regularity is not observed in Wroclaw. It must be stressed, that the discussed layer is a zone of the strongest winds (highest wind speed) and jet streams (reaching 250 hPa), above Poland in early spring and summer. Furthermore, due to the significant seasonal and daily variations in the tropopause, the largest variations in values of meteorological elements are observed in the UTLS (Upper Troposphere–Lower Stratosphere).

At high altitudes in the lower stratosphere the direct measurements and NCEP/NCAR Reanalysis are more related again. Nevertheless, the obtained correlation coefficient r is as low as 0.1 (in July over Leba). Due to the technique of direct measuring at the higher levels, the results from the air soundings are more unreliable (high error), and the data are more often unavailable.

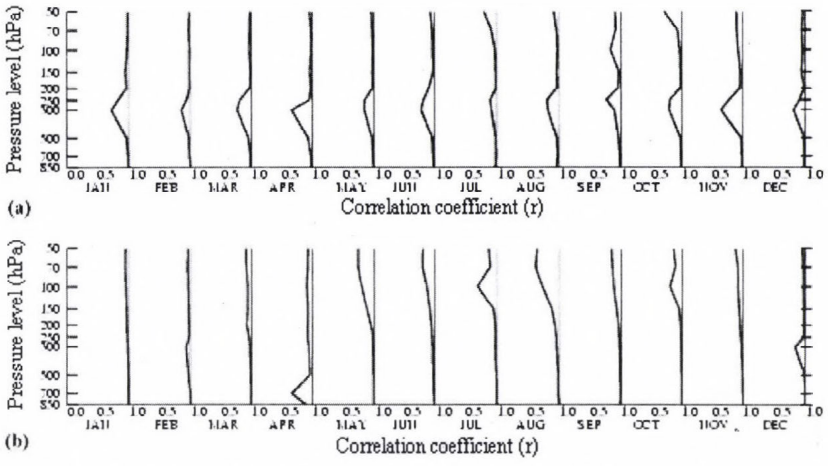


Fig. 1. Values of coefficient r of linear correlation at the standard WMO pressure levels, for each months of the year, between the air temperature (*a*) and geopotential height (*b*), measured at the aerological station in Warszawa and taken from NCEP/NCAR Reanalysis, respectively.

The relationship between the geopotential heights of the pressure levels over the air-sounding station and those at the same geographical location taken from NCEP/NCAR Reanalysis is more significant (*Figs. 1b, 2b*, and *3b*). The values of coefficient r of linear correlation are greater than 0.9 in the lower and middle troposphere in all months of the year, except for April (Warszawa, 700 hPa, $r=0.7$) and for August (Leba, 500 hPa, $r=0.6$).

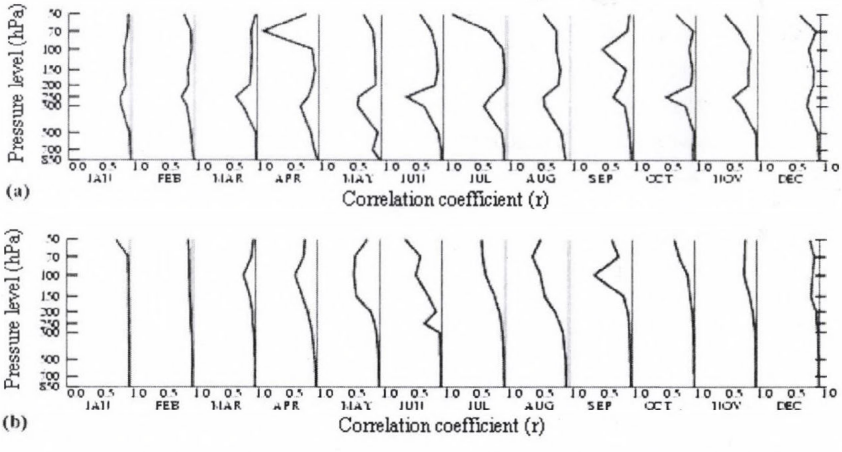


Fig. 2. Values of coefficient r of linear correlation at the standard WMO pressure levels, for each months of the year, between the air temperature (*a*) and geopotential height (*b*), measured at the aerological station in Wroclaw and taken from NCEP/NCAR Reanalysis, respectively.

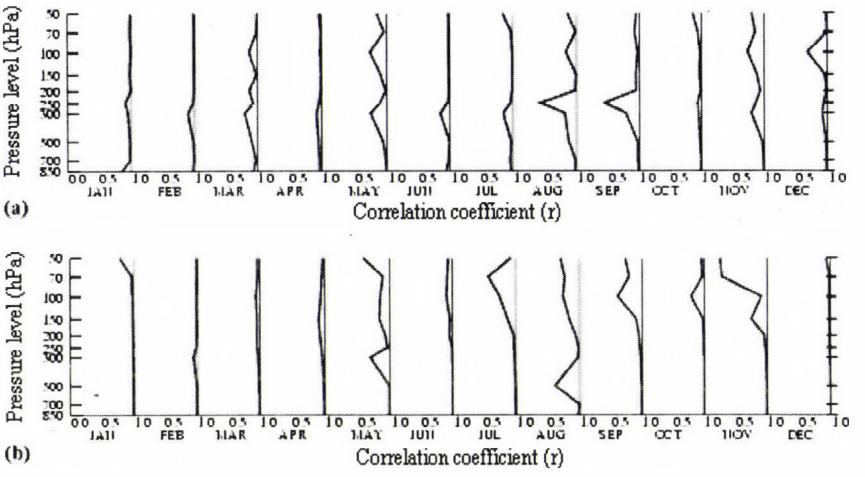


Fig. 3. Values of coefficient r of linear correlation at the standard WMO pressure levels, for each months of the year, between the air temperature (*a*) and geopotential height (*b*), measured at the aerological station in Leba and taken from NCEP/NCAR Reanalysis, respectively.

Strong positive correlation is apparent up to the 200 hPa isobaric level above which, in the lower stratosphere especially, the values of coefficient r drop into the range of 0.7–0.8, however, in November (Leba, $r=0.2$) and June (Warszawa, $r=0.4$), as it was mentioned earlier, at such high-altitude the measured data is highly unreliable.

4. Summary

Based on the results shown in the previous paragraph, the NCEP/NCAR Reanalysis and direct aerological measurements provided by Polish meteorological stations are strongly positive correlated in the lower and middle troposphere (i.e., in the region bounded by pressure levels 500 hPa at the top and 850 hPa at the bottom). This is valid for both of the air temperature and geopotential heights of isobaric surfaces.

In the zone between the 300 hPa and 250 hPa level, where strong winds (often jet streams) influenced by baric gradients are observed, distinct differences, especially in the temperature field, are noticed. Above 250 hPa, the linear correlation coefficient r reaches values again close to 1. However, due to the methodology of the direct measurements, data from aerological sounding in this part of the atmosphere (low stratosphere) are highly unreliable. Also, the frequency of correctly performed measurements decreases with height.

To examine pairs of time series (e.g., temperature or geopotential heights), simple statistical methods are usually used: the standard deviation, that is the best measure of spread or root mean square, whenever the relationship between variables is analyzed. Following other authors (e.g., *Poccard et al.*, 2000; *Atkinson et al.* 2003), for our study purpose a correlation method was involved. The simple linear correlation is not affected by systematic offset and gives a good idea of how two time series of variables are coherent. The non-linear correlation, based on polynomial approach of degree more than 1, requiring more difficult calculation, can be used in statistical examination. However, more sophisticated methods would be neither reliable nor efficient enough in our study case.

Finally, the presented analysis leads to the conclusion that NCEP/NCAR Reanalysis data is fully complementary and valuable data series, particularly in the lower and middle troposphere and can be used along with direct measurements (meteorological sounding of the atmosphere).

Analysis for the meteorological fields of humidity and u, v wind components are in preparation.

Acknowledgements—Reanalysis data was provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov/. This study was accomplished with the financial support of the Polish State Committee for Scientific Research under the Grant 618/E-217/SPB/COST/KN/DWM80/2005-2006.

References

Atkinson, D.E. and *Solomon, S.M.*, 2003: Comparison of NCEP/NCAR Reanalyses Data with Station Data for the Circum-Polar Coastal Regime (2003 – 7POLARCLIM). *AMS Conference on Polar Meteorology and Oceanography and Joint Symposium on High-Latitude Climate Variations*, 7 [np].

Gaffen, D.J., Sargent, M.A., Habermann, R.E, and *Lazante, J.R.*, 2000: Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *J. Climate 13*, 1776-1796.

Poccard, I., Janicot, S., and *Camberlin, P.*, 2000: Comparison of rainfall structures between NCEP/NCAR reanalyses and observed data over tropical Africa. *Clim. Dynam. 16*, 897-915.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and *Joseph, D.*, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *B. Am. Meteorol. Soc. 77*, 437-471.

Marshall, G.J., 2002a: Analysis of recent circulation and thermal advection change in the northern Antarctic Peninsula. *Int. J. Climatol. 22*, 1557-1567.

Marshall, G.J., 2002b: Trends in antarctic geopotential height and temperature: A comparison between radiosonde and NCEP-NCAR. *J. Climate 15*, 659-674.

NEWS

Obituary — Kirill Kondratyev (1920–2006)

"Hereby we grievously inform you about the premature decease of Kirill Kondratyev that occurred on the 1st of May, 2006. Our friend and colleague, Kirill Kondratyev was a famous scientist, full Academician of the Russian Academy of Sciences, an acknowledged expert in the area of climate and environment. He is the author of more than one thousand papers in the most prestigious journals as well as of more than hundred monographs and textbooks published in the former USSR, Russia and abroad.

The area of scientific interests of Kirill Kondratyev was extremely broad encompassing the theory of transfer of thermal radiation through the atmosphere, greenhouse effect, natural and man-induced disasters and catastrophes, remote sensing of environment and global climate change.

Kirill Kondratyev was the honorary member of the American Meteorological Society, Royal Meteorological Society of the Great Britain, Academy of Natural Sciences "Leopoldina" (Germany), foreign member of the American Academy of Arts and Sciences, member of the International Astronautic Academy, honorary Doctor of Sciences of the Universities of Lile (France), Budapest (Hungary) and Athens (Greece). During many years he was Editor-in-Chief of the Russian Journal "Earth Observations and Remote Sensing", he also was member of the editorial board of such journals as "Optics of the atmosphere and ocean", "Proceedings of the Russian Geographical Society", "Meteorology and Atmospheric Physics" (Austria), "Időjárás" (Hungary), "Il Nuovo Cimento C "(Italy), "Atmosfera" (México), "Energy and Environment" (Great Britain).

For his salient scientific attainments Kirill Kondratyev was awarded with the State Award of the USSR, and decorated with a Gold Medal by the World Meteorological Organization, the Simons Gold Medal by the Royal Meteorological Society of the Great Britain.

During first 30 years of his scientific career, Kirill Kondratyev was insolubly related to the State University of Leningrad, where he made a way from a professor assistant to the rector of the university. An important part of his activities was also related to the A. I. Voeikov Main Geophysical Observatory.

Next 30 years were tied up with his work at the Institute of Limnology and the Center for Ecological Safety, Russian Academy of Sciences. The latter

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 double 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. Good quality laser printings are preferred.

The text should be submitted both in manuscript and in electronic form, the latter on diskette 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.

More information for authors is available: antal.e@met.hu

Information on the last issues: http://omsz.met.hu/irodalom/firat_ido/ido_hu.html

Published by the Hungarian Meteorological Service

Budapest, Hungary

INDEX: 26 361

HU ISSN 0324-6329

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

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

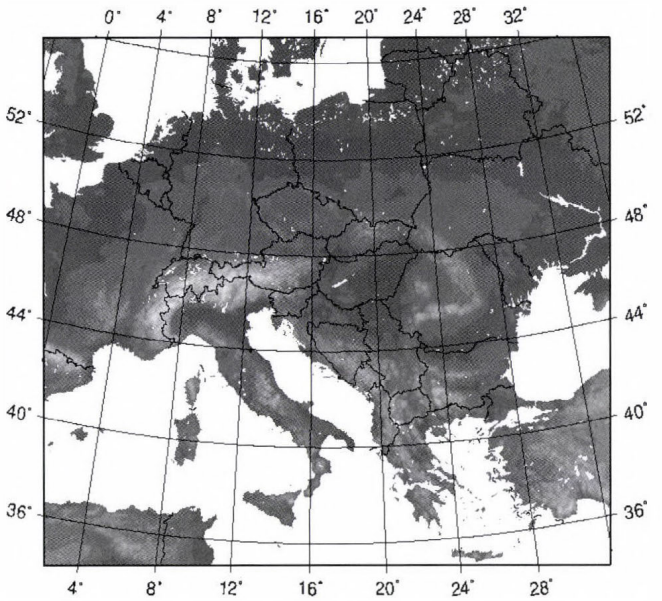


Fig. 1. An example for the output of configuration 923: the domain and topography contours of the ALADIN/HU version of the ALADIN model.

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).

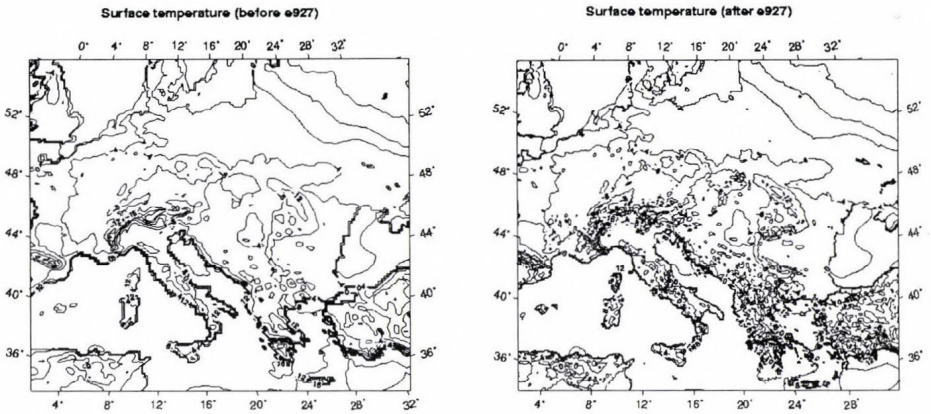


Fig. 2. An example for configuration 927. Surface temperature before (left panel) and after (right panel) EE927 interpolation (24h forecast based on the 00:00 UTC, February 23, 2006 integration).

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.

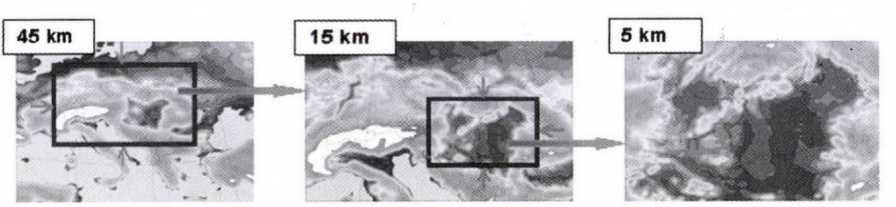


Fig. 3. The steps of the dynamical downscaling method using nested domains with increasing resolution. The orography of the different applied ALADIN domains are shown.

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*).

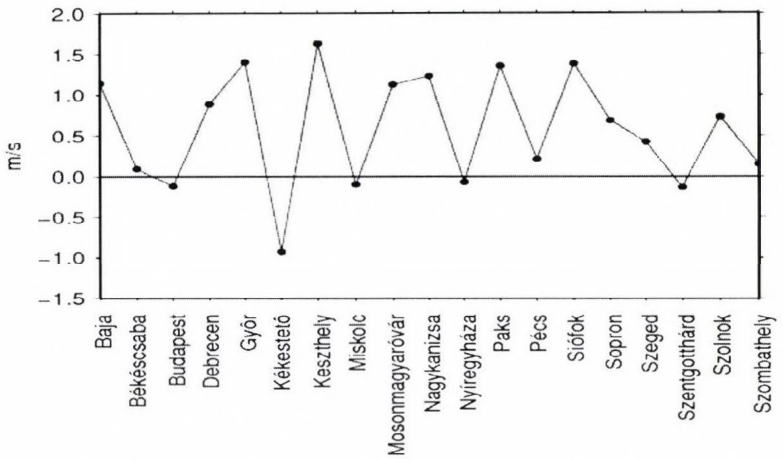


Fig. 4. The average difference between the computed and observed average 10 m wind speed for 19 Hungarian stations for the period between January 1, 1992 and December 31, 2001.

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:

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

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).

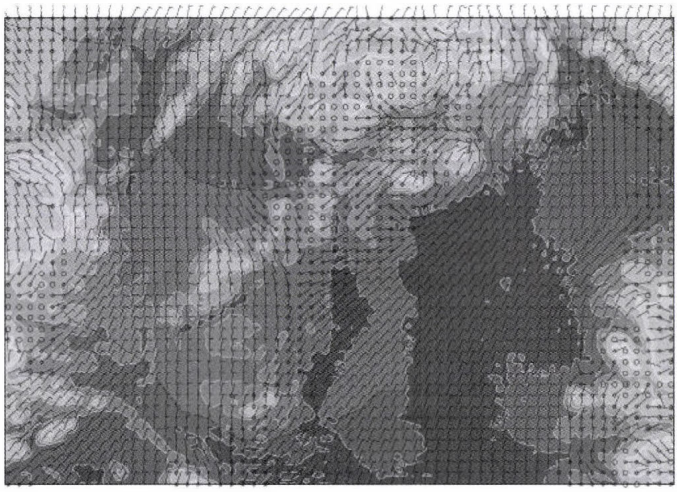


Fig. 5. The output of the 10 m wind forecast by the small scale wind dynamical adaptation over Hungary (valid at 00:00 UTC, January 19, 2006). The high resolution orography of the model is superimposed in the background. The details of the flow are much more pronounced than in the original (coarse resolution)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

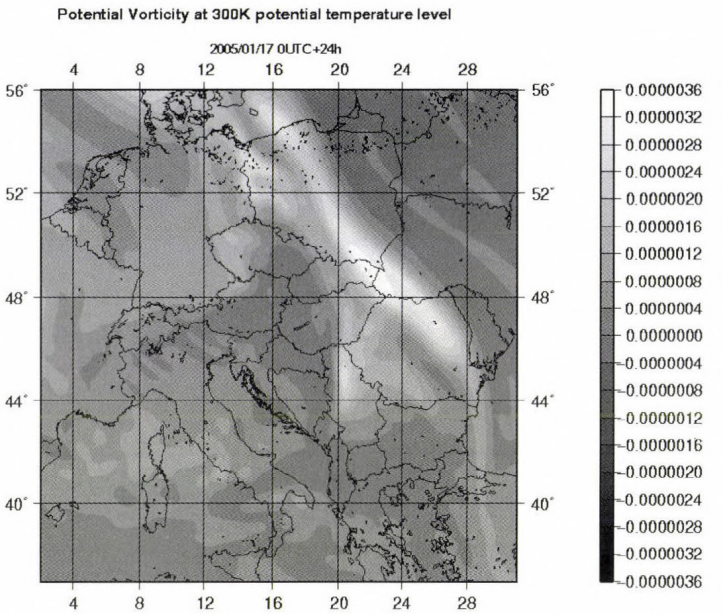


Fig. 6. An example for post processing: the potential vorticity is visualized on the 300 K isentropic level (24h forecast valid at 00:00 UTC, January 18, 2005).

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:

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

where $\delta\mathbf{x}$ represents the departure between the analysis and the background, $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{d} = \mathbf{y}_0 - H(\mathbf{x}_b)$, where H is the observation operator, $\mathbf{H}$ 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.

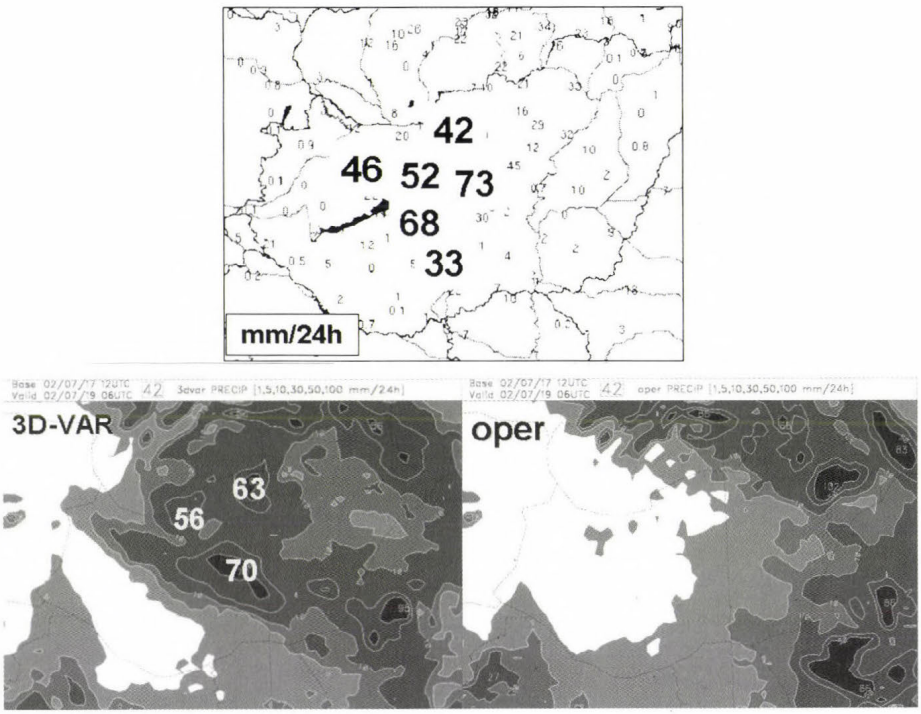


Fig. 7. Example for the difference between precipitation forecasts started from local three-dimensional variational (3d-var) data assimilation (bottom left) and using dynamical adaptation (bottom right). 24h accumulated precipitation is shown between 06:00 UTC, July 18, 2002 and 06:00 UTC, July 19, 2002.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

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

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.*

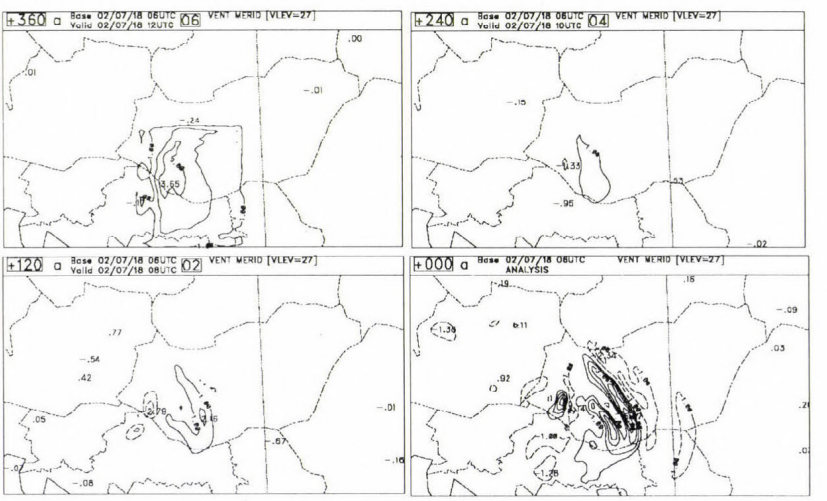


Fig. 8. Evolution of gradient (sensitivity) fields with respect to the meridional wind component on the 27th model level during the adjoint integration between 12:00 and 06:00 UTC, July 8, 2002 (upper left panel is valid for 12:00 UTC, the upper right is for 10:00 UTC, the lower left is for 08:00 UTC 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 $\gamma_1^2 \geq \gamma_2^2 \geq \ldots \geq \gamma_n^2$ denotes the eigenvalues of $\mathbf{L}^*\mathbf{L}$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ the corresponding orthonormal eigenvectors, then the $\gamma_1, \ldots, \gamma_n$ numbers are called the singular values of the $\mathbf{L}$ operator and the $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ 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 $\mathbf{v}_1$. Additionally, in the orthogonal complementary subspace of $\mathbf{v}_1$, the maximum is γ_2 at location of $\mathbf{v}_2$, and so on for the other orthogonal subspaces. Finally, this algorithm leads to γ_n and $\mathbf{v}_n$, 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.*

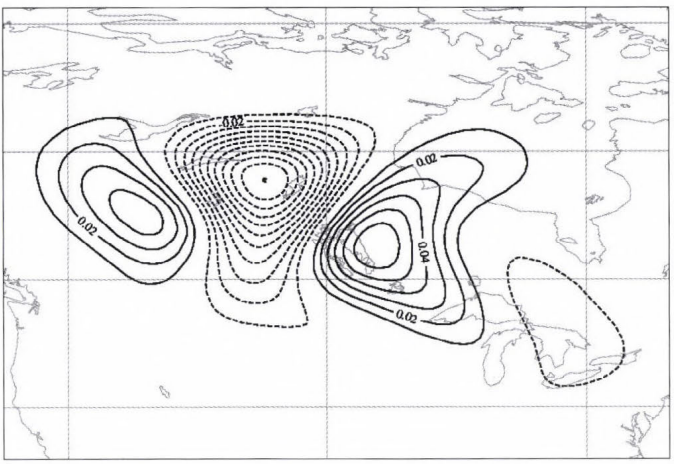


Fig. 9. Typical result of singular vector calculations (the usual dipole structure of the singular vectors can be clearly identified) valid at 12:00 UTC, June 17, 2002. The leading singular vector is visualized for temperature on the 10th model level. Solid lines indicate the positive and dashed lines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

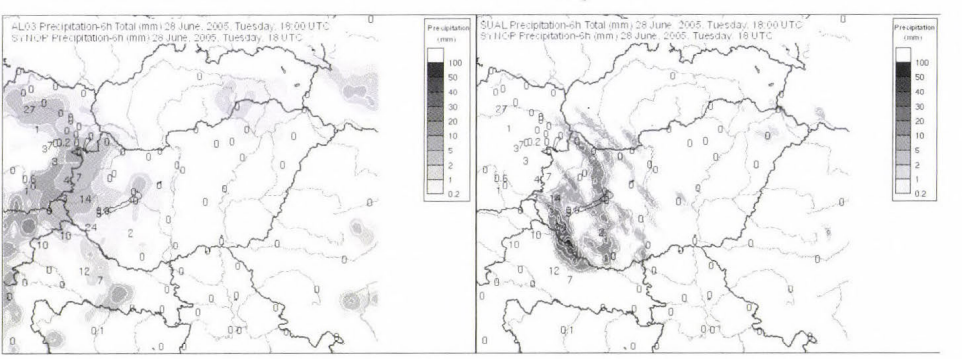


Fig. 10. An example for AROME forecast. 6h accumulated total precipitation after 18h integration for ALADIN dynamical adaptation (left panel) and AROME (right panel). SYNOP (surface) observations are marked with numbers on both panels. The forecast started at 00:00 UTC, June 28, 2005. It can be seen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

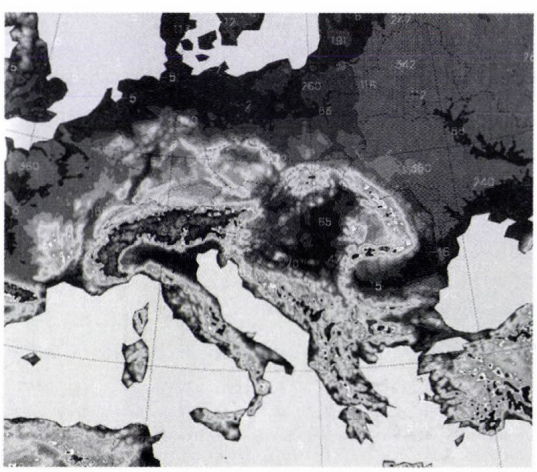


Fig. 2. The integration domain and orography of the ALADIN model.

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*):

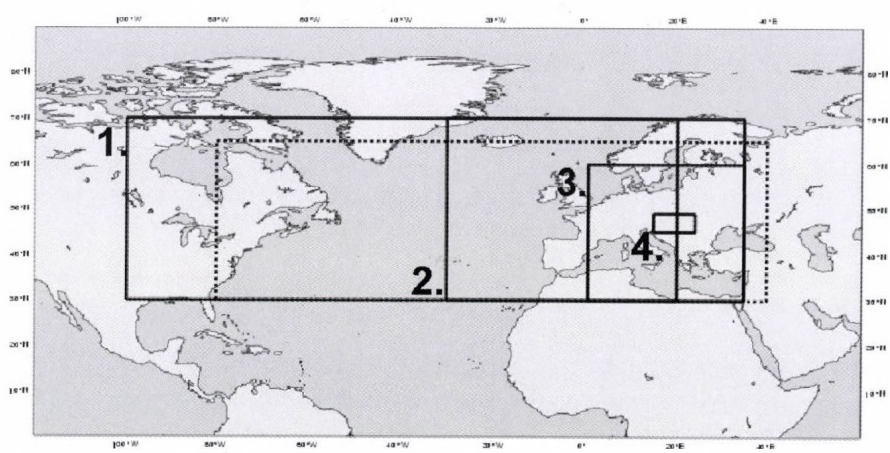


Fig. 3. The location of the four different target domains used for the experiments and the target domain used in the present PEARP system (dotted line).

- 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 $2 \times 1/(\text{number of ensemble members} + 1)$, 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.

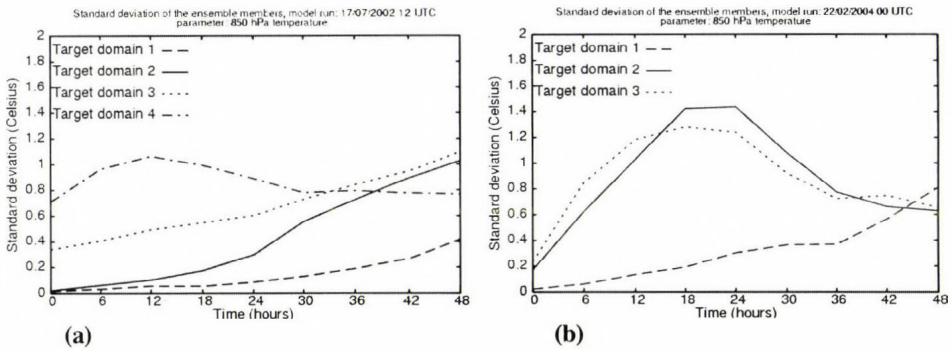


Fig. 4. Standard deviation diagrams for the ARPEGE ensemble system for two different model runs for 850 hPa temperature. Standard deviation was computed over Hungary. (a) Model run: July 17, 2002, 12:00 UTC. Dashed line is standard deviation with the use of target domain 1, solid line is standard deviation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

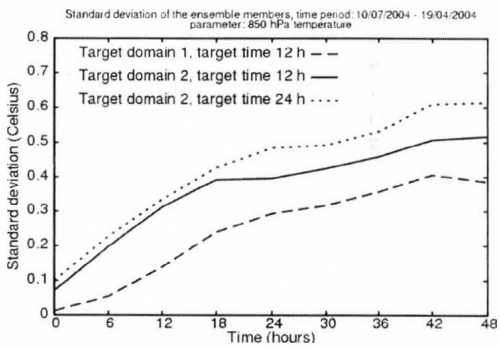


Fig. 5. Standard deviation diagrams for the ARPEGE ensemble system for the period July 10–19, 2004, for 850 hPa temperature. Standard deviation was computed over Hungary. Dashed line is standard deviation with the use of target domain 1 and target time 12 hours, solid line is standard deviation with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

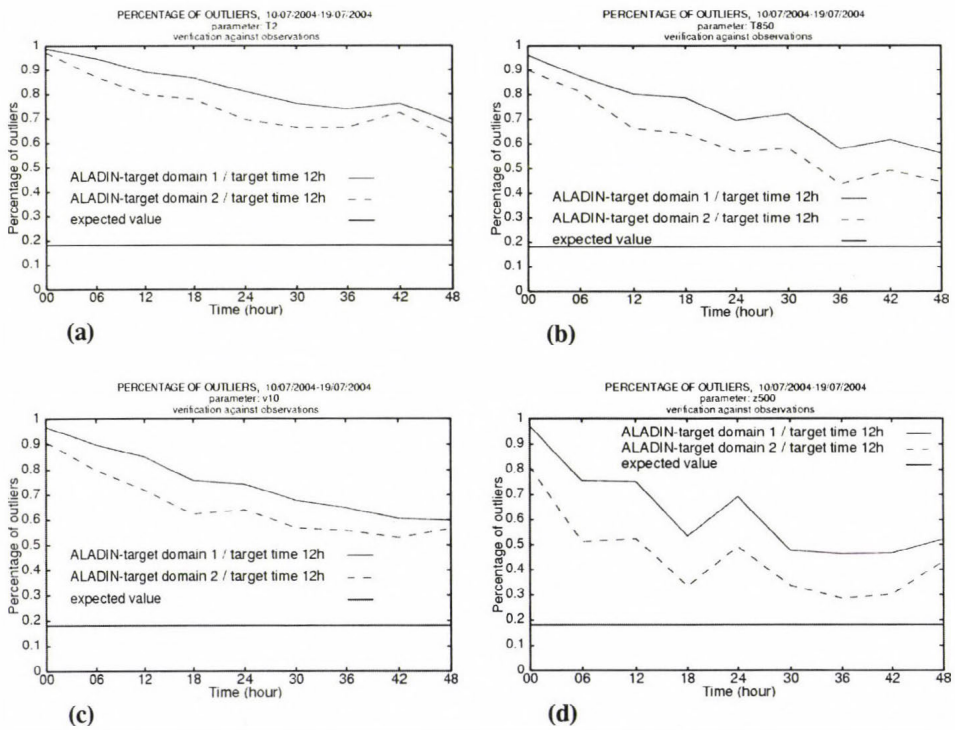


Fig. 6. Percentage of outliers diagrams for the ALADIN ensemble system for the period July 10–19, 2004. (a) 2-meter temperature, (b) 850 hPa temperature, (c) 10-meter wind speed, (d) 500 hPa geopotential height. Solid line is ALADIN coupled with ARPEGE ensemble members using target domain 1 and target 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.

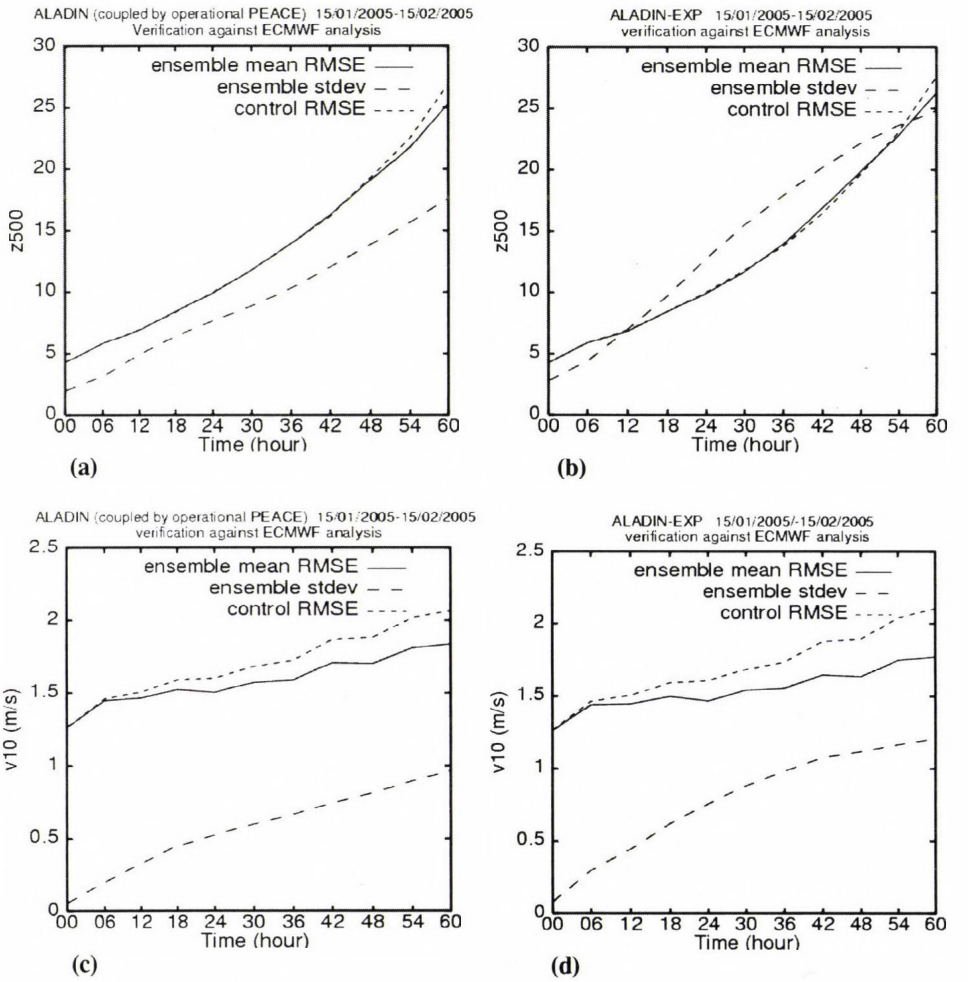


Fig. 7. Comparison of RMSE and standard deviation for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 500 hPa geopotential height, ALADIN coupled with the PEARP members, (b) 500 hPa geopotential height, ALADIN coupled with the experimental set, (c) 10-meter wind speed, ALADIN 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.

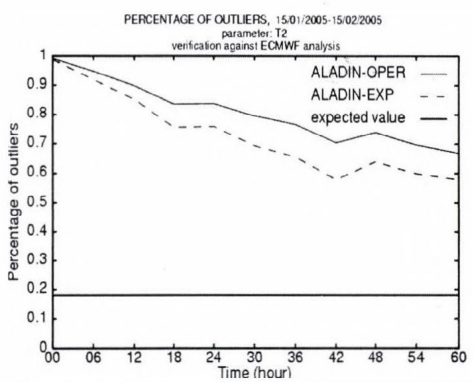
3.3.2 Spread vs. RMSE

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*).

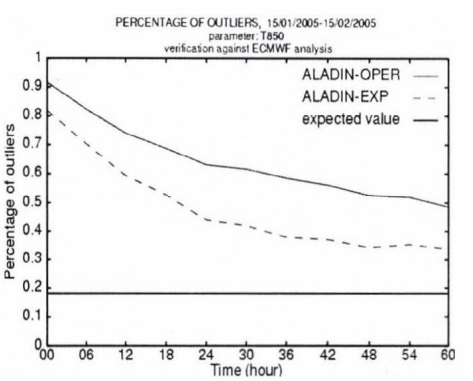
It can be concluded that there is a discrepancy between the error and the spread, however, with the correct choice of SV target domain and target time this can be reduced (especially at higher levels).

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) →→

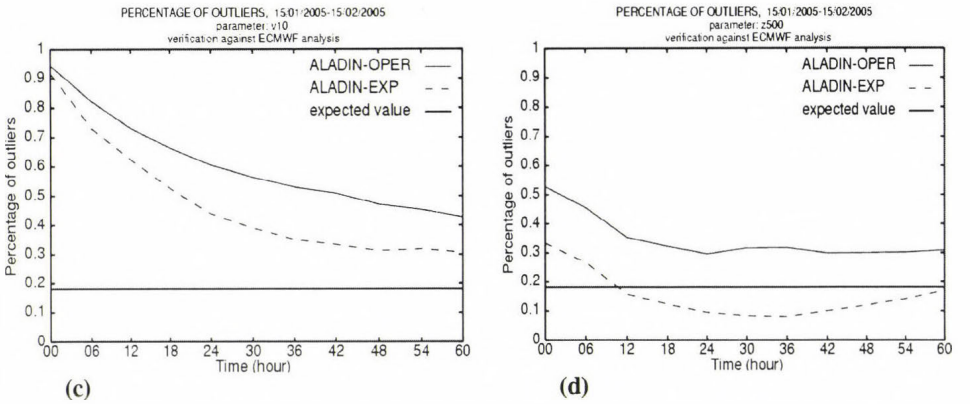


Fig. 8. Percentage of outliers diagrams for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 2-meter temperature, (b) 850 hPa temperature, (c) 10-meter wind speed, (d) 500 hPa geopotential height. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line 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).

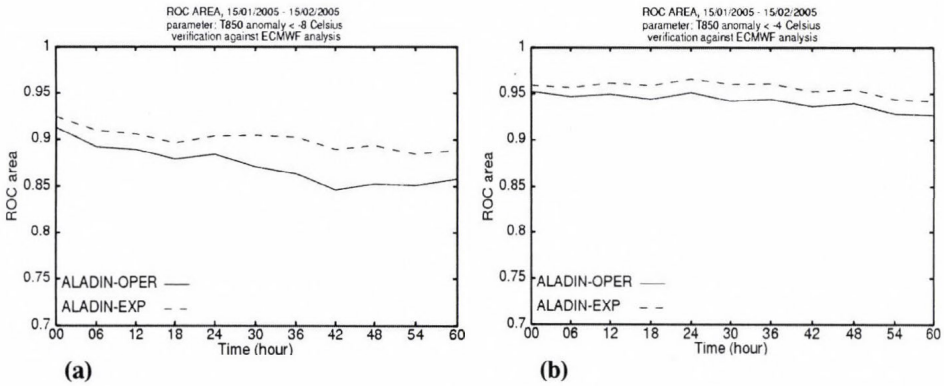


Fig. 9. ROC area for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 850 hPa temperature anomaly less than –8 °C, (b) 850 hPa temperature anomaly less then –4 °C. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line is ALADIN coupled with the experimental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

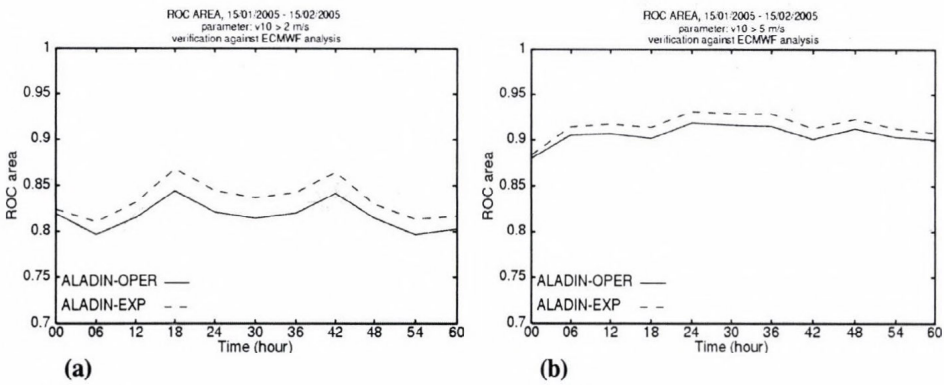


Fig. 10. ROC area for the ALADIN ensemble system for the period January 15 – February 15, 2005. (a) 10-meter wind speed greater than 2 m/s, (b) 10-meter wind speed greater than 5 m/s. Solid line is ALADIN coupled with the operational PEARP forecasts, dashed line is ALADIN coupled with the experimental 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.

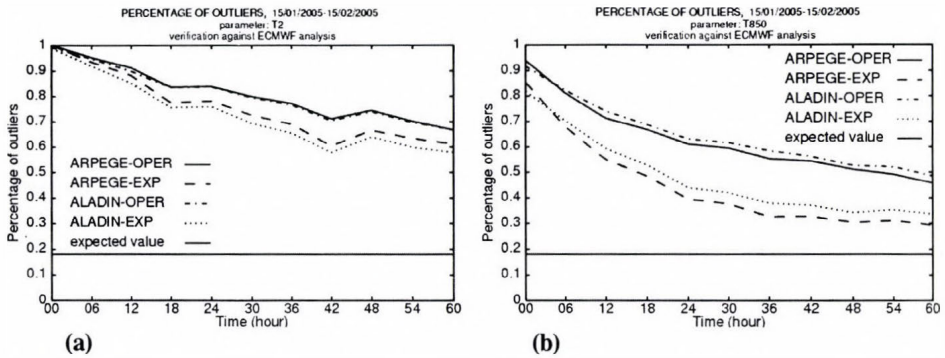
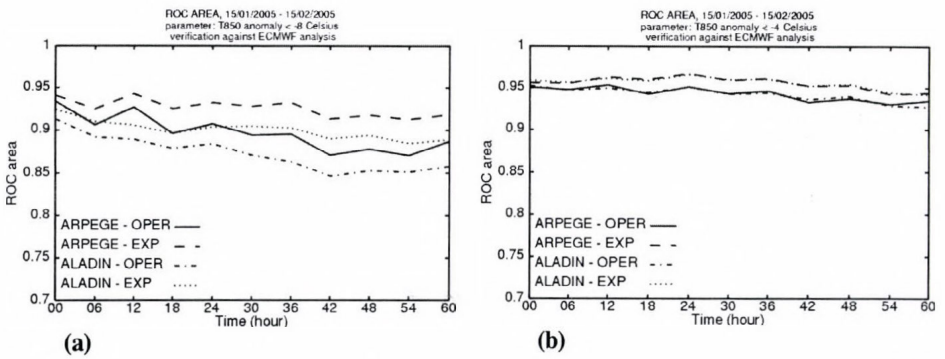


Fig. 11. Percentage of outliers diagrams for ARPEGE and ALADIN ensemble systems for the period January 15 – February 15, 2005. (a) 2-meter temperature, (b) 850 hPa temperature. Solid line is the operational PEARP forecasts (ARPEGE-OPER), chained line is ALADIN coupled with the operational PEARP members (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).



→→

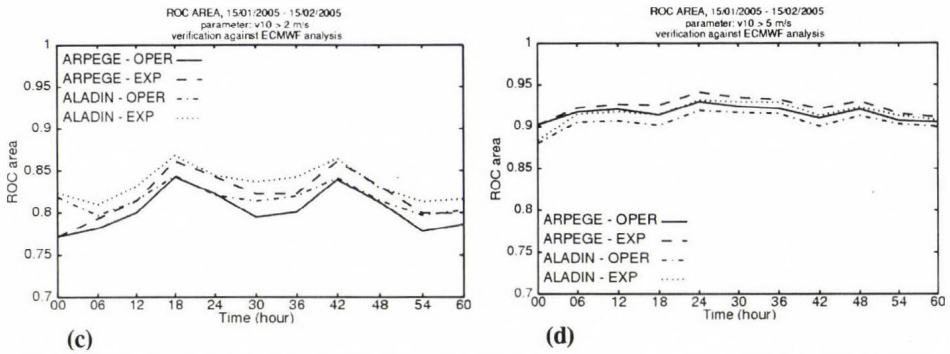


Fig. 12. ROC area diagrams for ARPEGE and ALADIN ensemble systems for the period January 15 – February 15, 2005. (a) 850 hPa temperature less than –8 °C, (b) 850 hPa temperature less than –4 °C, (c) 10-meter wind speed greater than 2 m/s, (d) 10-meter wind speed greater than 5 m/s. Solid line is the 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).

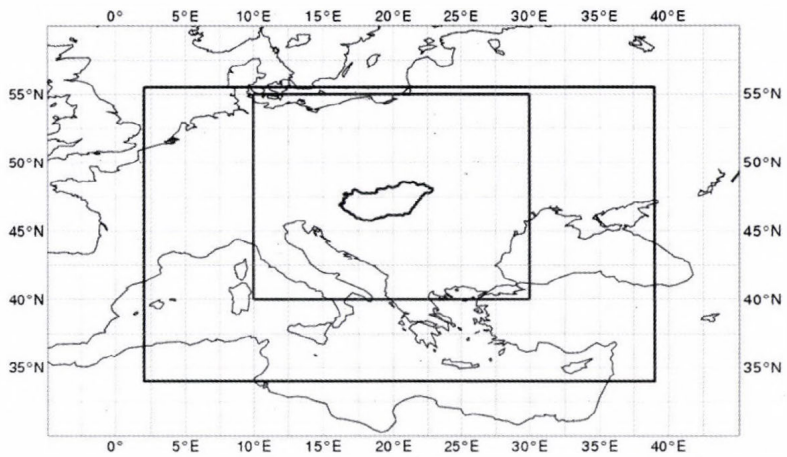


Fig. 1. The two clustering domains for the ECMWF/ALADIN downscaling system. The bigger area is equivalent to the integration domain of the ALADIN model.

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*).

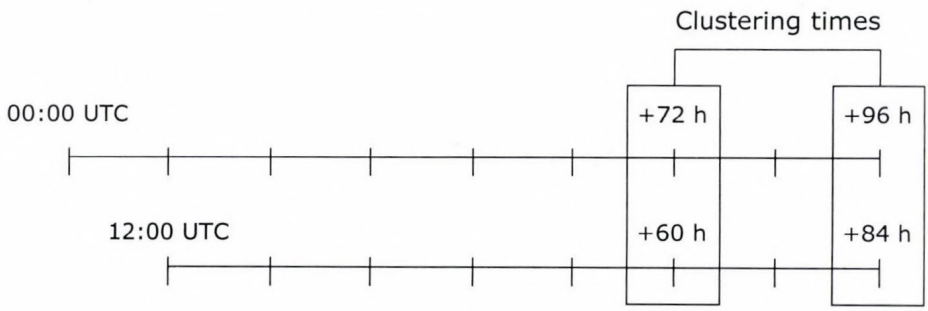


Fig. 2. Schematic description of the clustering time while using 100 ECMWF EPS members.

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).

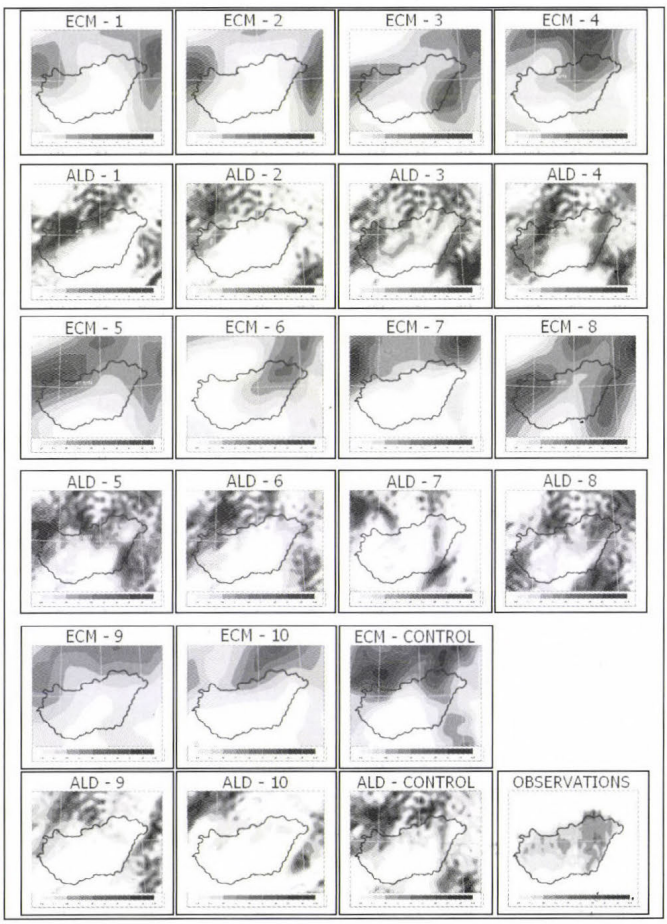


Fig. 3. Stamp diagrams for the ECMWF EPS (10 representative members of the 50 global EPS members, bigger domain clustering configuration), and the ALADIN downscaled members. 24 h precipitation forecasts for May 18, 2005, 06:00 UTC – May 19, 2005, 06:00 UTC. The control forecasts (without initial perturbations) 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*).

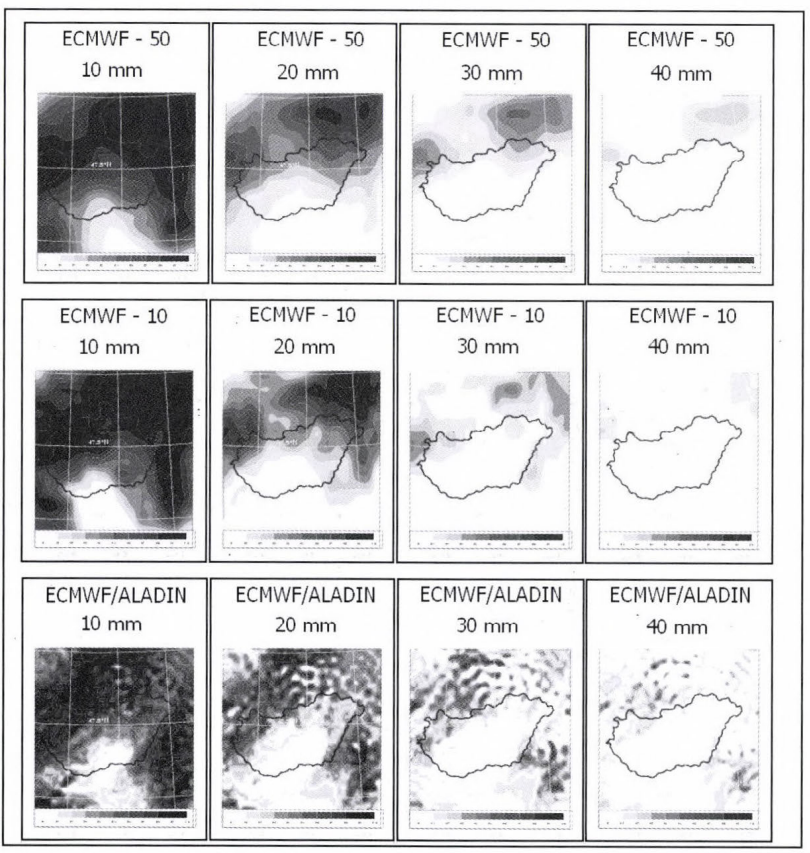


Fig. 4. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24 h precipitation exceeding 10, 20, 30, and 40 mm between May 18, 2005, 06:00 UTC – May 19, 2005,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

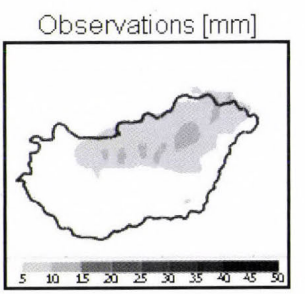


Fig. 5. 24 h accumulated precipitation over Hungary between November 16, 2005, 06:00 UTC and November 17, 2005, 06:00 UTC. Values over 5 mm are displayed. Darker colours indicate higher precipitation amounts.

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).

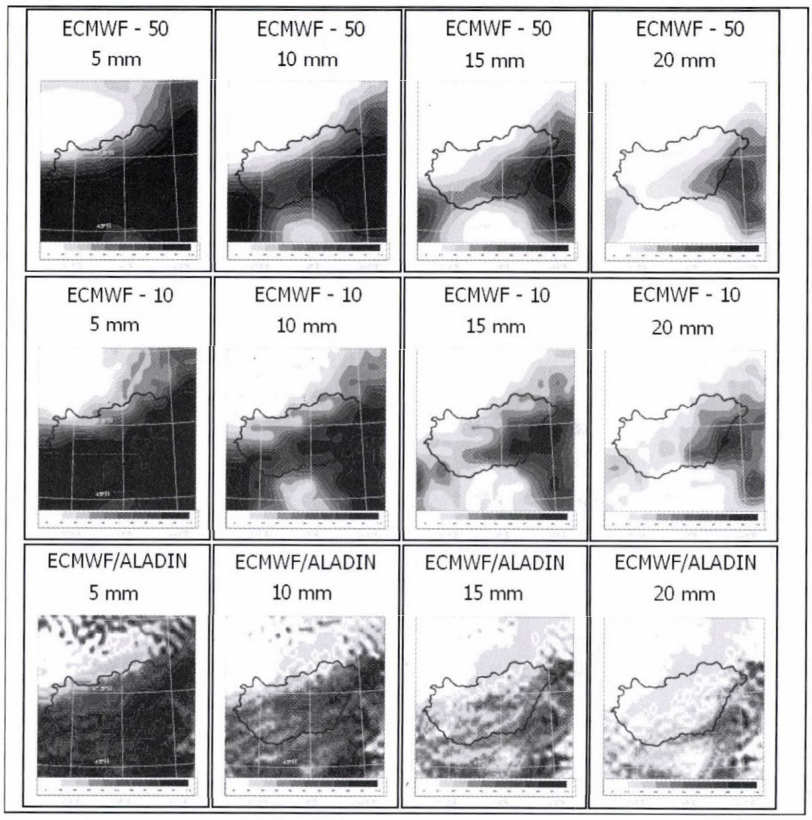


Fig. 6. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24 h precipitation exceeding 5, 10, 15, and 20 mm between November 16, 2005, 06:00 UTC and November 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. $0.1° \times 0.1°$ 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.

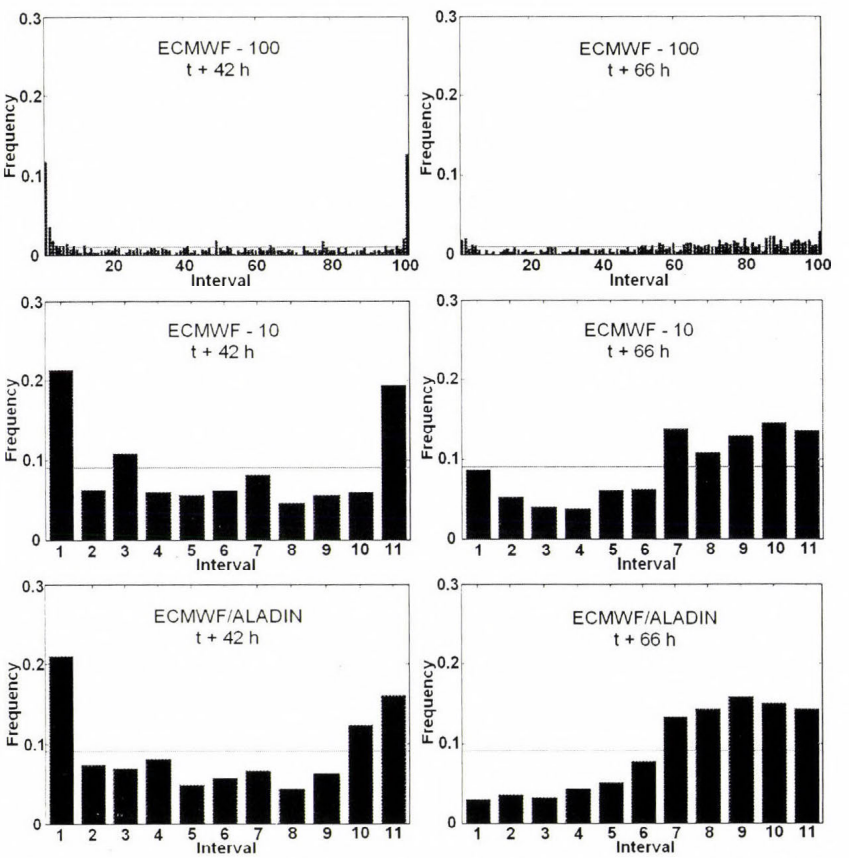


Fig. 7. Talagrand diagrams of 24 h accumulated precipitation for ECMWF EPS (100 members), ECMWF EPS (10 representative members), and ECMWF/ALADIN (10 representative members from 100 global members, bigger domain configuration) for +42 h and +66 h forecasts calculated from the first three case studies.

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.

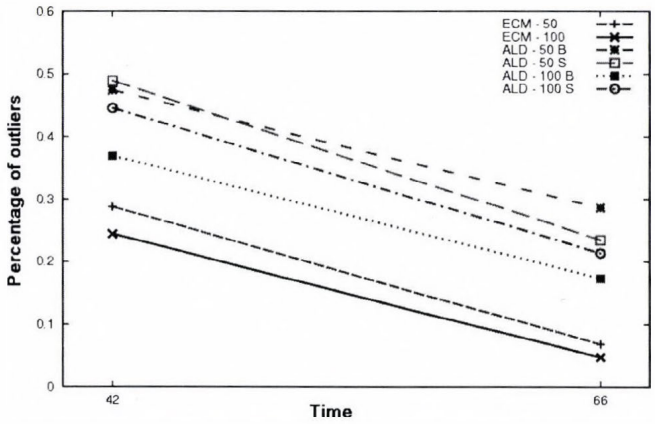


Fig. 8. Talagrand outliers for precipitation for ECMWF EPS and ECMWF/ALADIN, calculated from the first three case studies. The different configurations are: ECMWF EPS 50 members (ECM–50), ECMWF EPS 100 members (ECM–100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain (ALD–50 B), ECMWF/ALADIN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

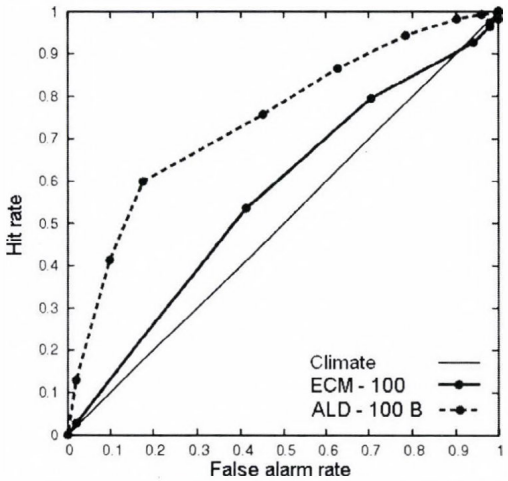


Fig. 9. ROC diagrams for 5 mm/24 h accumulated precipitation threshold for +66 h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain), 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).

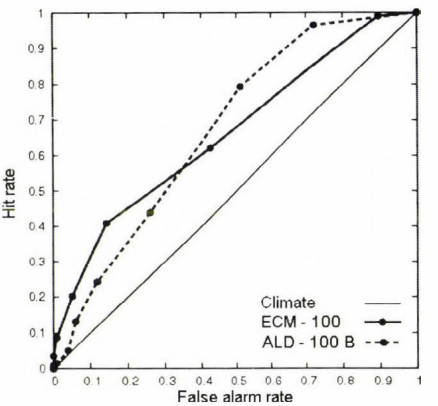


Fig. 10. ROC diagrams for 20 mm/24 h accumulated precipitation threshold for +66 h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain),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

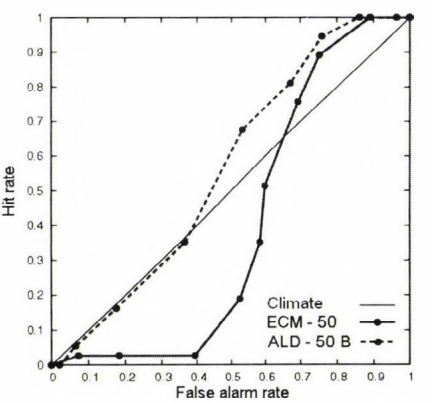


Fig. 11. ROC diagrams for 10 mm/24 h threshold for +66 h accumulated precipitation forecasts calculated from the fourth case study. Thick solid line is relative to ECMWF EPS (50 members), dashed line is relative to ECMWF/ALADIN (10 representative members from 50 global members, bigger domain), diagonal 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.

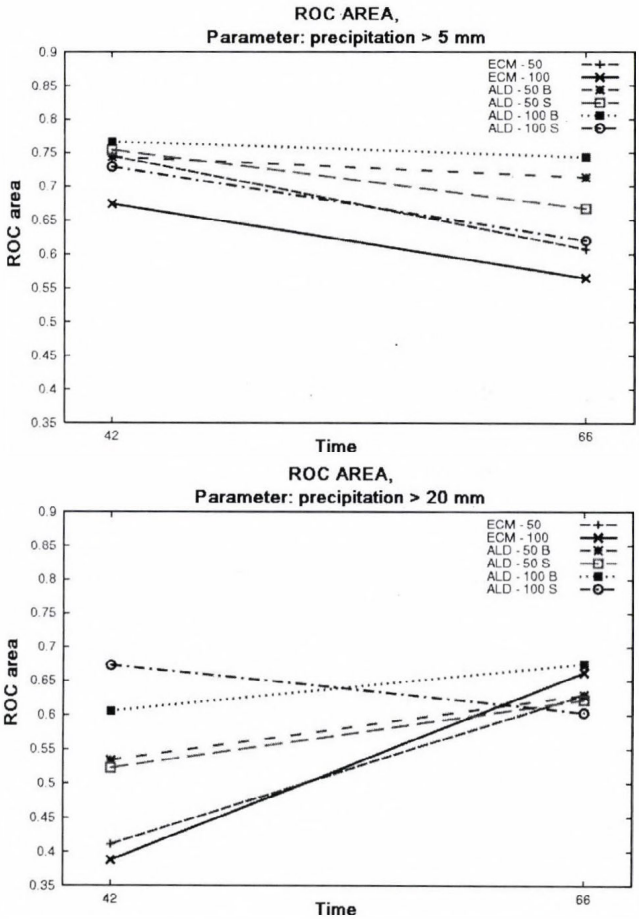


Fig. 12. Time evolution of the ROC area for the 5 mm/24 h (upper diagram) and 20 mm/24 h accumulated precipitation thresholds (lower diagram). The different configurations are: ECMWF EPS 50 members (ECM–50), ECMWF EPS 100 members (ECM–100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain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

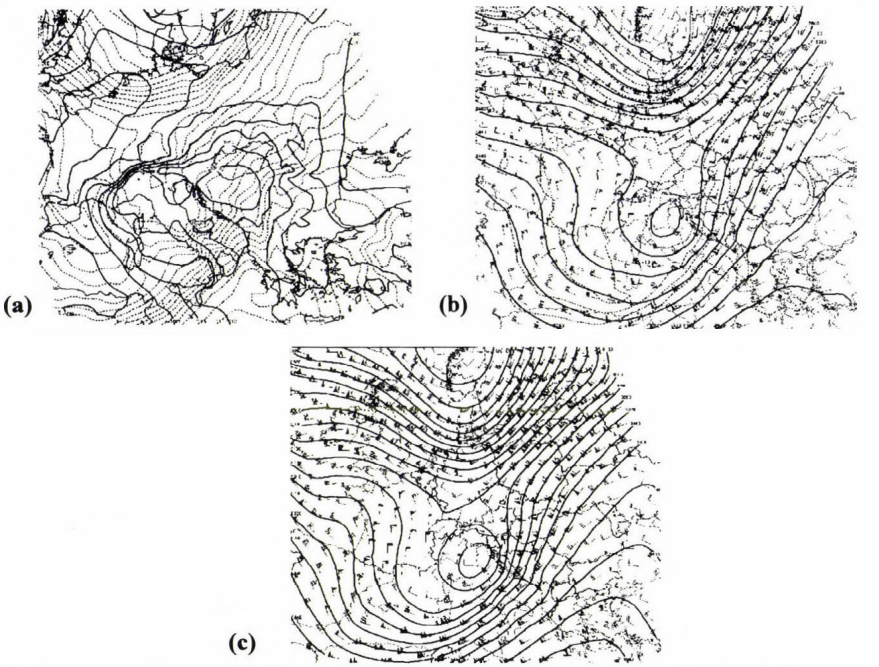


Fig. 1. ECMWF analysis on May 18, 2005, 12:00 UTC. (a) Sea level pressure (thick lines) and 850 hPa temperature (dashed lines), (b) 500 hPa height (thick lines), temperature (dashed lines) and wind field, (c) 300 hPa height and wind field.

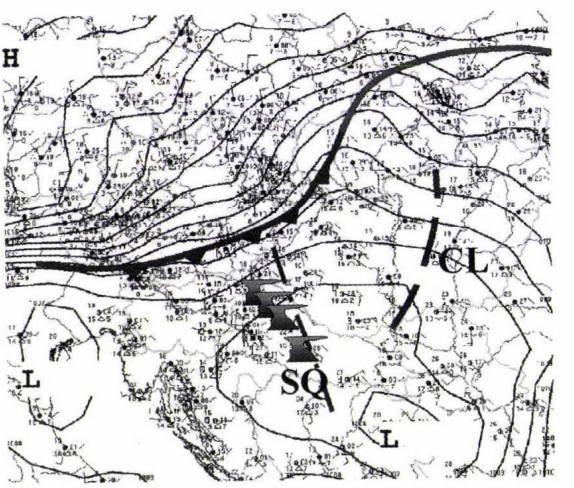


Fig. 2. Mesoscale weather conditions on May 18, 2005, 12:00 UTC. SQ denotes the prefrontal squall line and CL denotes the convergence line with stratiform precipitating system.

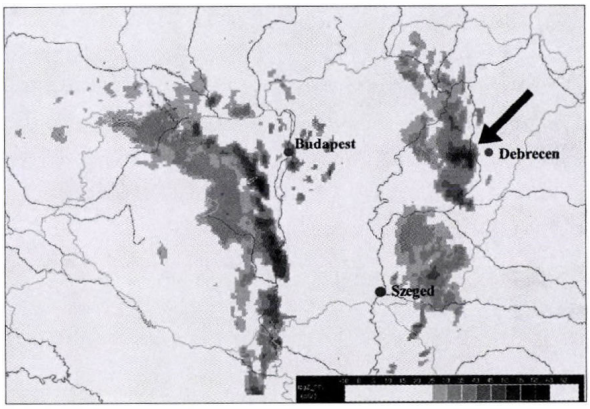


Fig. 3. Radar image on May 18, 2005, 15:15 UTC. The black arrow shows the position of the hook echo in the eastern convergence line.

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.

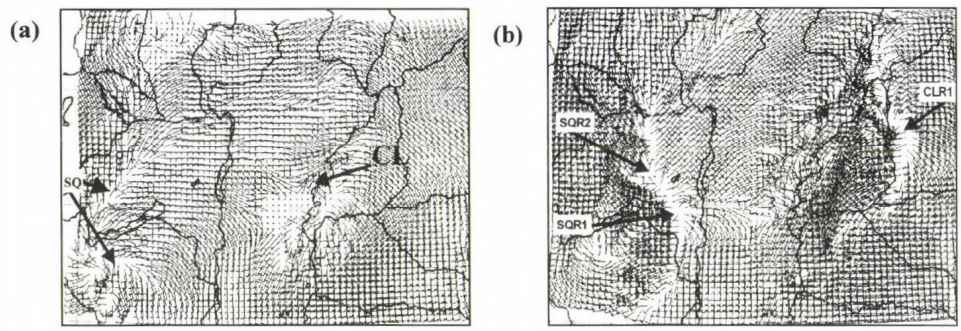


Fig. 4. Wind forecast at the 925 hPa for 14:00 UTC and +2 hours forecast (a), and for 16:00 UTC and+4 hours forecast (b).

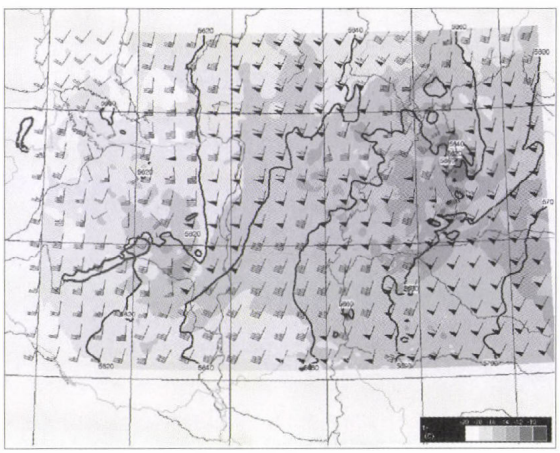


Fig. 5. Forecasted temperature and wind field at the 500 hPa level and geopotential height (solid lines) of 500 hPa at 16:00 UTC (+4 hours of simulation).

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.

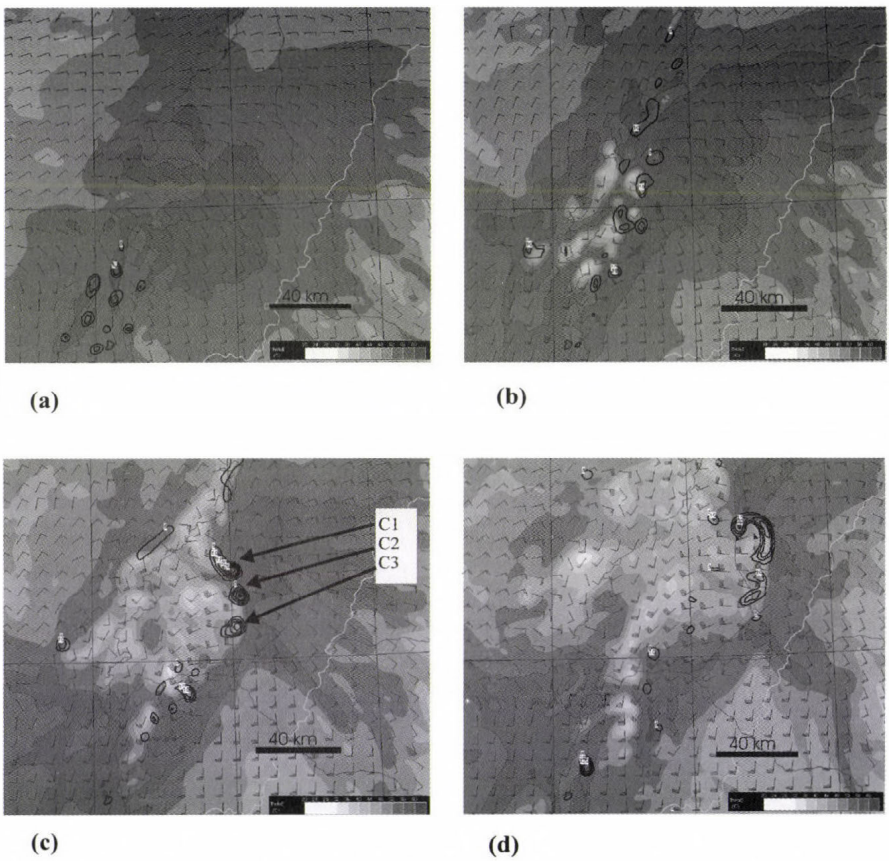


Fig. 6. Composite of simulated equivalent potential temperature at the 925 hPa level (shaded areas), simulated wind field at the 925 hPa level, and simulated updraft velocity filed (5 m/s contours are denoted by thick solid lines) at the 500 hPa level is plotted in each figure. (a) 13:30 UTC, (b) 14:30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

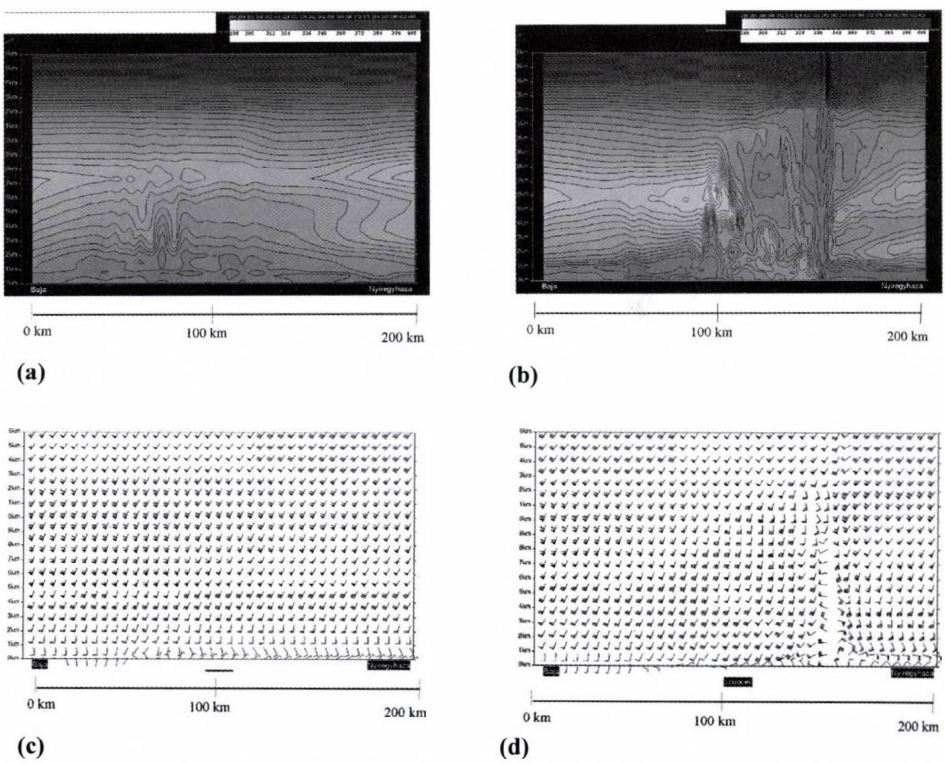


Fig. 7. Cross section along the supercell's path (SW-NE). EPT at 13:30 UTC (a) and 16:00 UTC (b), wind fields at 13:30 UTC (c) and 16:00 UTC (d).

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).

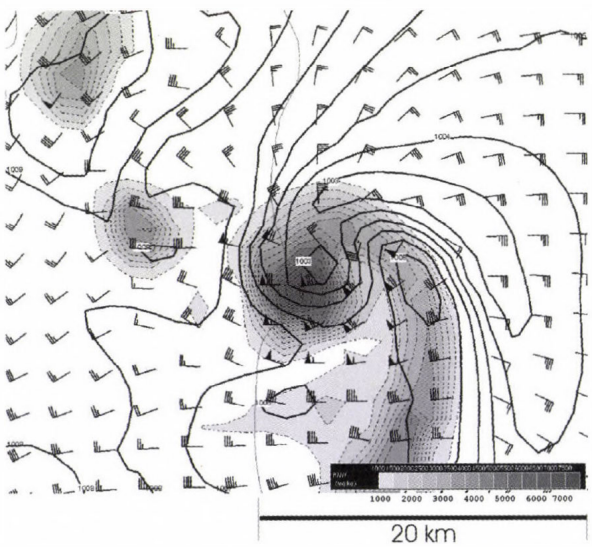


Fig. 8. Mean sea level pressure, 950 hPa wind field, and 700 hPa mixing ratio of rain water (shaded areas are bounded by dashed lines) forecasted to 16:00 UTC (+4 hours forecast).

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.

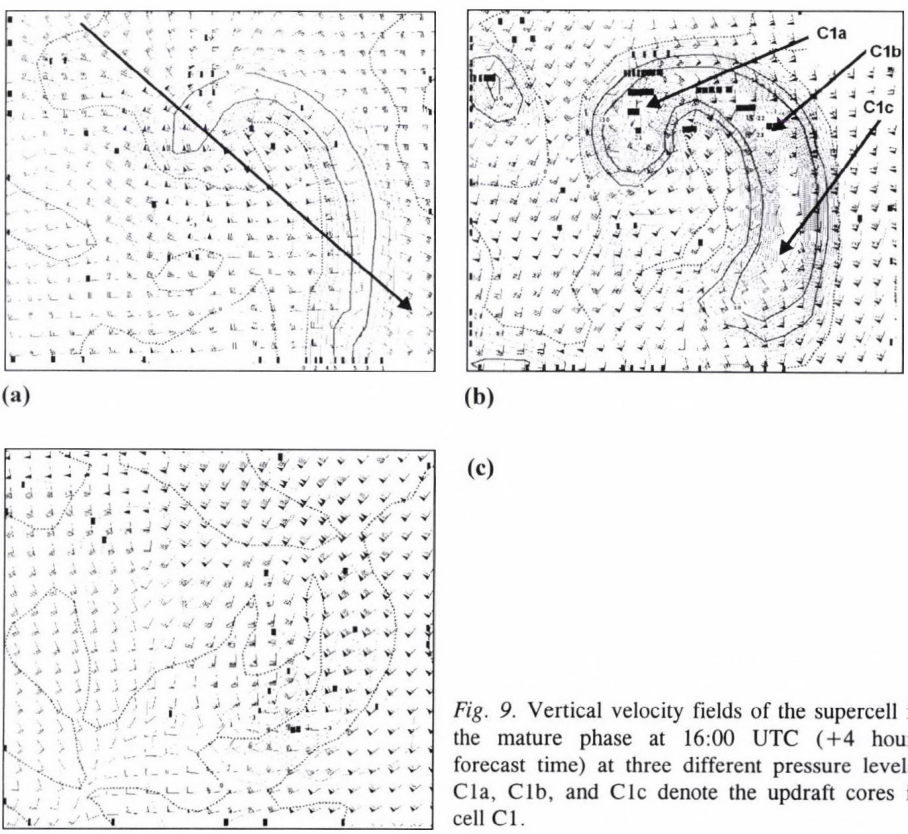


Fig. 9. Vertical velocity fields of the supercell in the mature phase at 16:00 UTC (+4 hours forecast time) at three different pressure levels. C1a, C1b, and C1c denote the updraft cores in cell C1.

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.

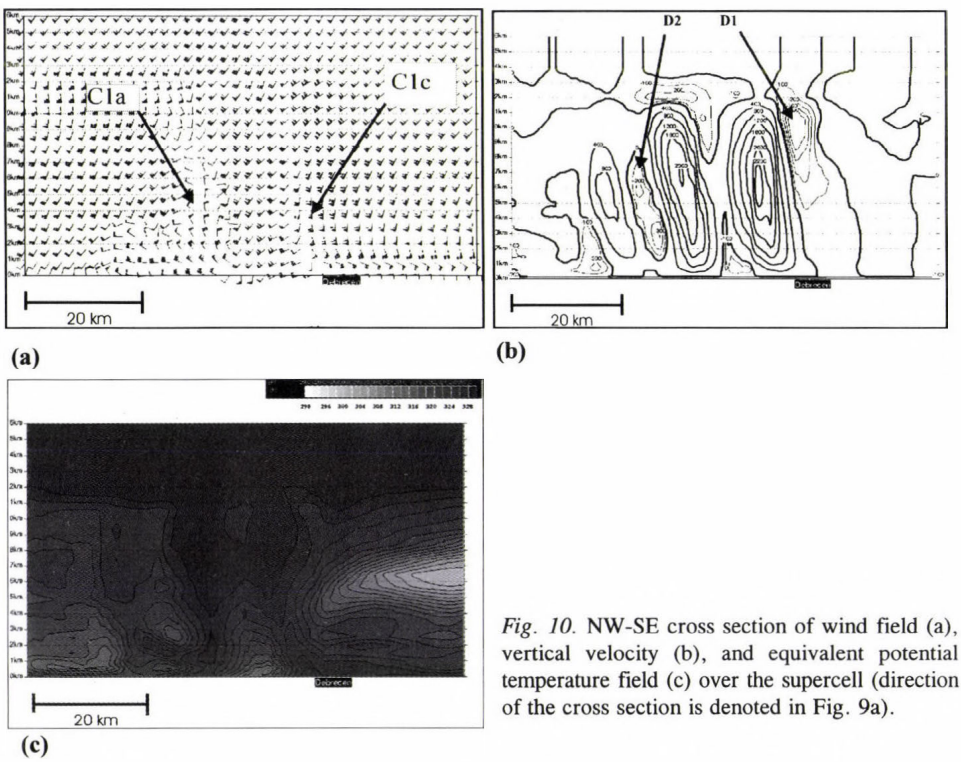


Fig. 10. NW-SE cross section of wind field (a), vertical velocity (b), and equivalent potential temperature field (c) over the supercell (direction of the cross section is denoted in Fig. 9a).

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.

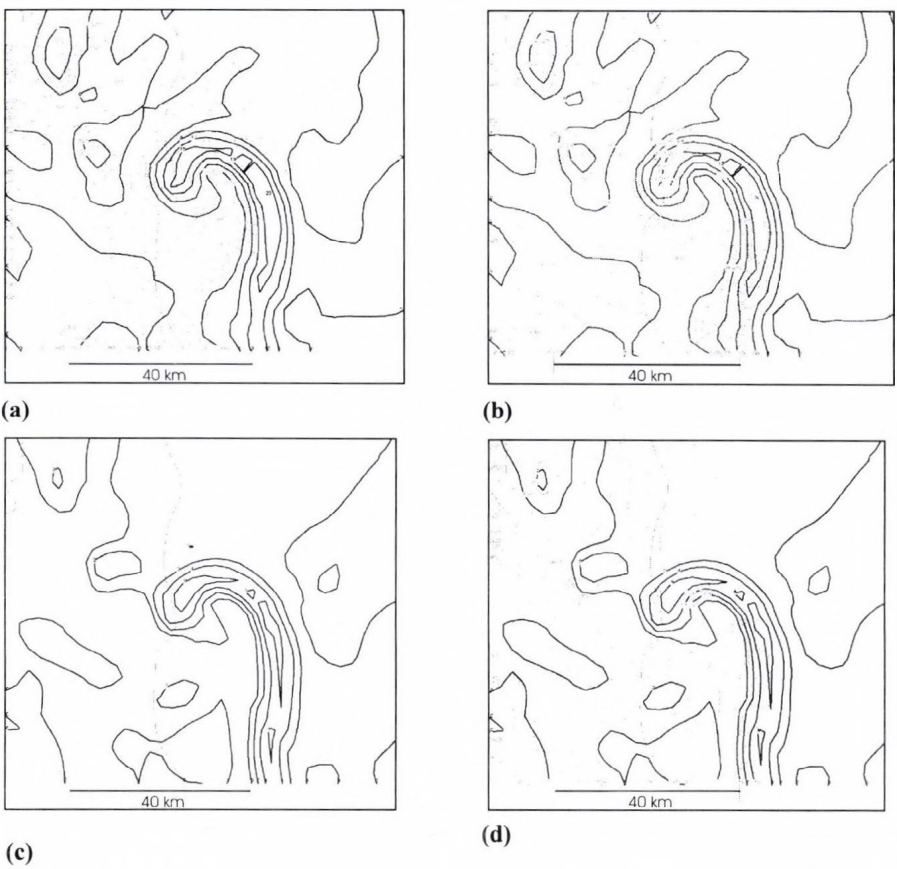


Fig. 11. Precipitation elements (dashed lines) and updraft (solid lines) at 16:00 UTC (+4 hours forecast time). (a) 700 hPa updraft (m/s) and graupel mixing ration (mg/kg), (b) 700 hPa updraft (m/s) and rain water mixing ratio (mg/kg), (c) 925 hPa updraft (m/s) and graupel mixing ratio (mg/kg), (d) 925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

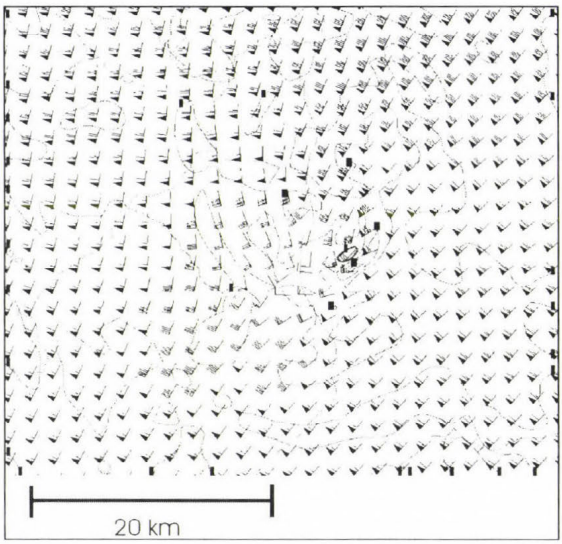


Fig. 12. Wind field of the 200 hPa level at 16:00 UTC. Thick curves denote the isoline of 2 m/s upward velocity.

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 ($\sim 10^{-4}$ 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 ($L \sim 1000$ km) and its laboratory model ($L \sim 10$ 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 ($\sim 10^3$) 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)).

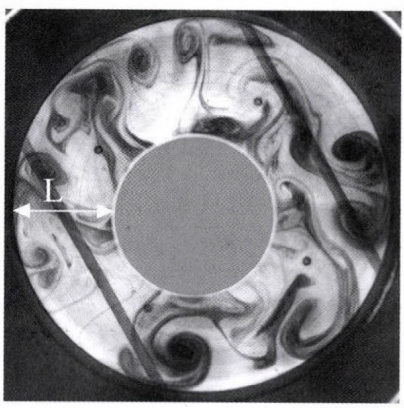


Fig. 1. Thermal convection in a rotating tank of three concentric cylinders. The central chamber is filled with ice, the outermost segment can be heated, the convecting medium in between is water at room temperature (h = 5 cm). The distance between the copper walls is L = 10.5 cm, the actual rotating 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 ($Re_{exp} \approx 10^2$–10^3) are much smaller than in the atmosphere ($Re_{atm} \approx 10^6$–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 $h/H > 0.15$.

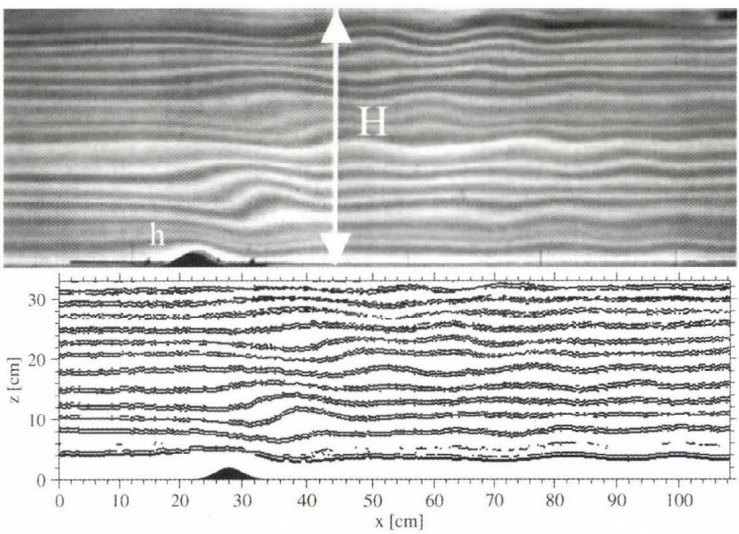


Fig. 2. Top: Wave field behind a moving (from right to left) bell-shaped obstacle in a linearly stratified fluid (salt solution periodically colored by food dye). $H = 32$ cm, $h = 2$ cm, the towing speed is 2.06 cm/s, $N = 1.26$ 1/s. Bottom: Wave field reconstructionby digital image processing.

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*.

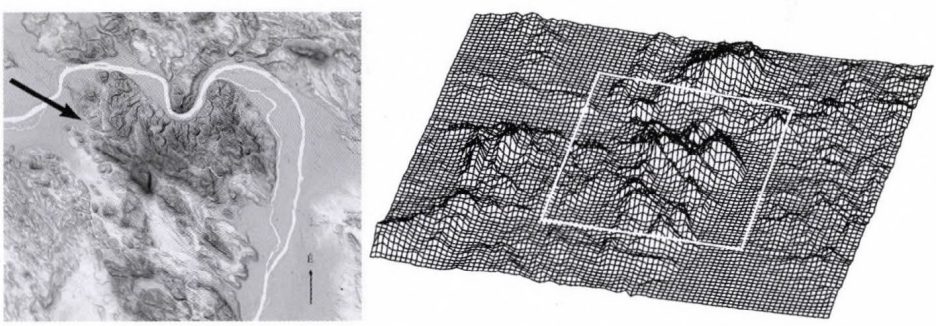


Fig. 3. Topographic settings at around the mountain Pilis (47.7°N; 18.8°E). The wind direction supporting mountain waves is indicated by a heavy arrow. The surface reconstruction on the right shows an area of 90×80 km^2, white frame locates the map borders.

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).

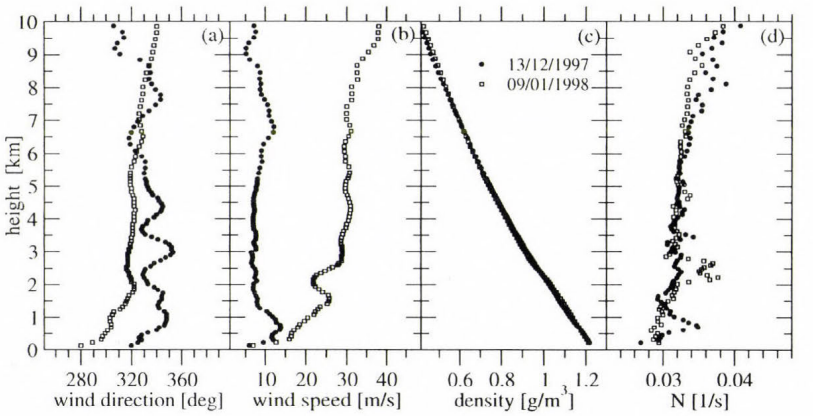


Fig. 4. (a) Wind direction, (b) wind speed, (c) air density, and (d) Brunt-Vaisala frequency profiles for two days in the winter of 1997/98 (indicated in (c)) at a disctance of ~40 kmfrom the mountain Pilis (Budapest-Lőrinc, Hungary). (Brunt-Vaisala frequencies were obtained directly from the density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

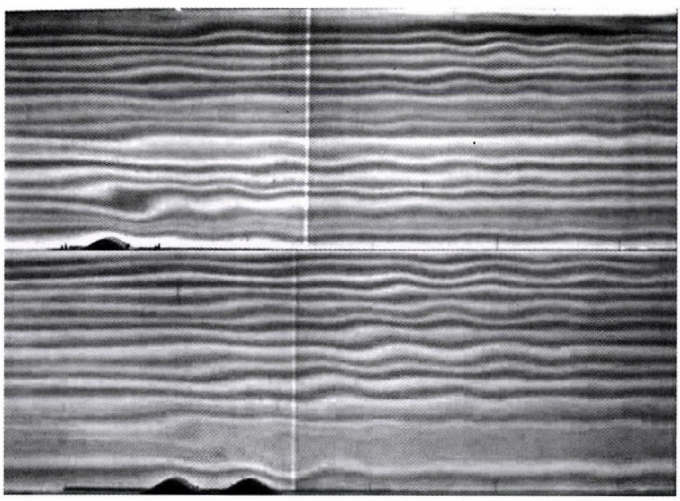


Fig. 5. Top: Wave field behind a single bell-shaped obstacle towed from right to left in a linearly stratified salt solution, $U = 1.53$ cm/s, $N = 1.26$ 1/s, $H = 37.5$ cm, $h = 2.0$ cm, $w = 2.6$ cm, Gaussian form. Bottom: Wave field behind a double bell-shaped obstacle fabricated from two identical 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*).

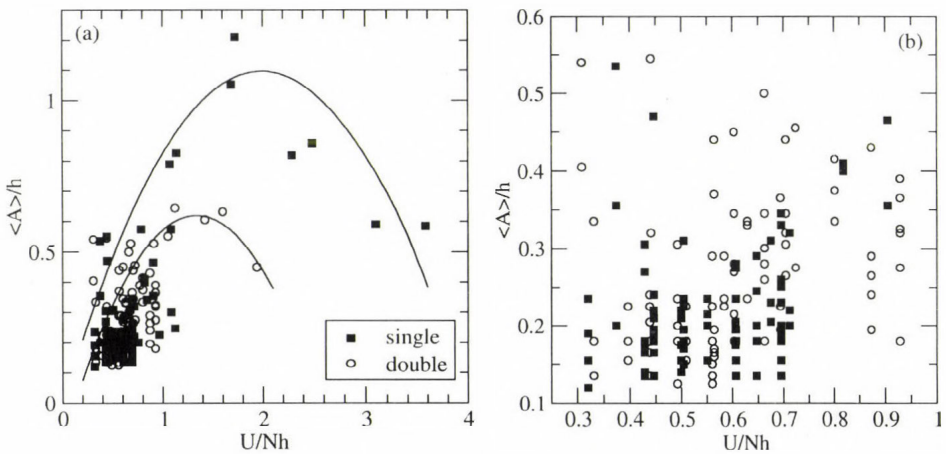


Fig. 6. (a) Normalized amplitudes as a function of dimensionless towing velocity for the single (heavy squares) and double (empty circles) obstacles. Solid lines only guide the eyes. (b) The same as (a), zoomed to the bottom left corner (nonlinear regime, see text).

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.

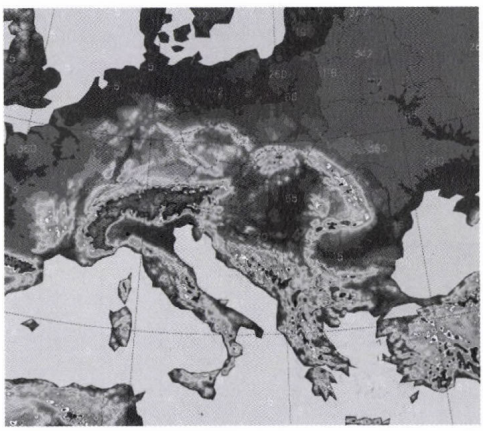


Fig. 1. The ALADIN/HU operational model domain.

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*).

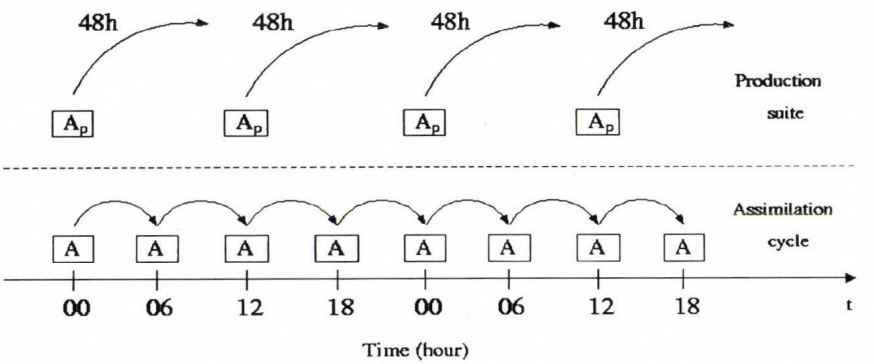


Fig. 2. The ALADIN/HU 3DVAR assimilation cycle. "A" stands for analysis with long data cut-off, "A_p" stands for "production" analysis with short data cut-off.

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*).

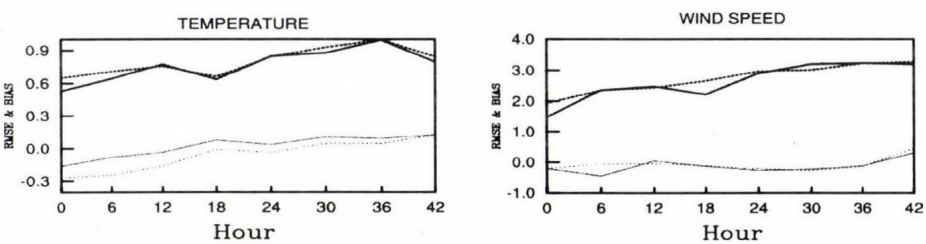


Fig. 3. Evolution of RMSE (thick lines) and BIAS (thin lines) scores according to the forecast range. Left: temperature on 500 hPa, right: wind speed on 500 hPa. Solid line:ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

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).

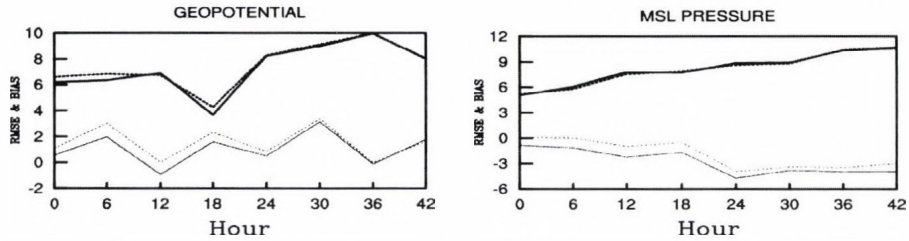


Fig. 4. Evolution of RMSE (thick lines) and BIAS (thin lines) scores with the forecast range.
Left: geopotential on 700 hPa, right: sea level pressure.
Solid line: ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

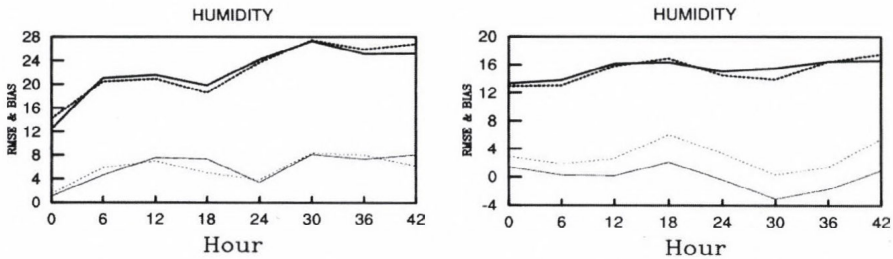


Fig. 5. Evolution of RMSE (thick lines) and BIAS (thin lines) scores with the forecast range.
Left: humidity on 500 hPa, right: humidity on the surface (2 m).
Solid line: ALADIN/HU 3DVAR, dashed line: ALADIN.

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.

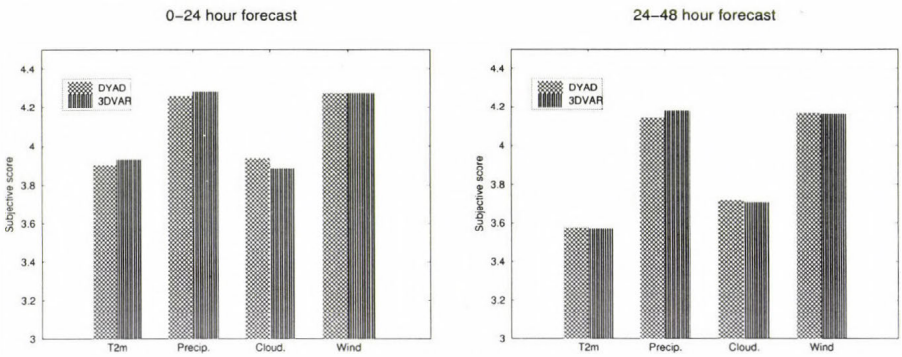


Fig. 6. Subjective verifications scores for different weather parameters summarized over a half-year period (July 1–December 31, 2004).

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.

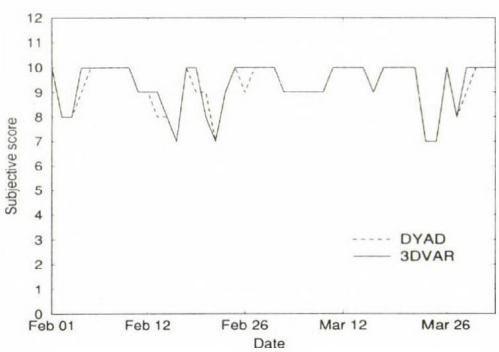


Fig. 7. Evolution of subjective verification scores for the precipitation short range (0–24 hours) forecasts over the period of February 1 – April 5, 2005. Scores are multiplied by a factor of 2. Full line: ALADIN/HU 3DVAR, dashed line: ALADIN/HU DYA.

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.

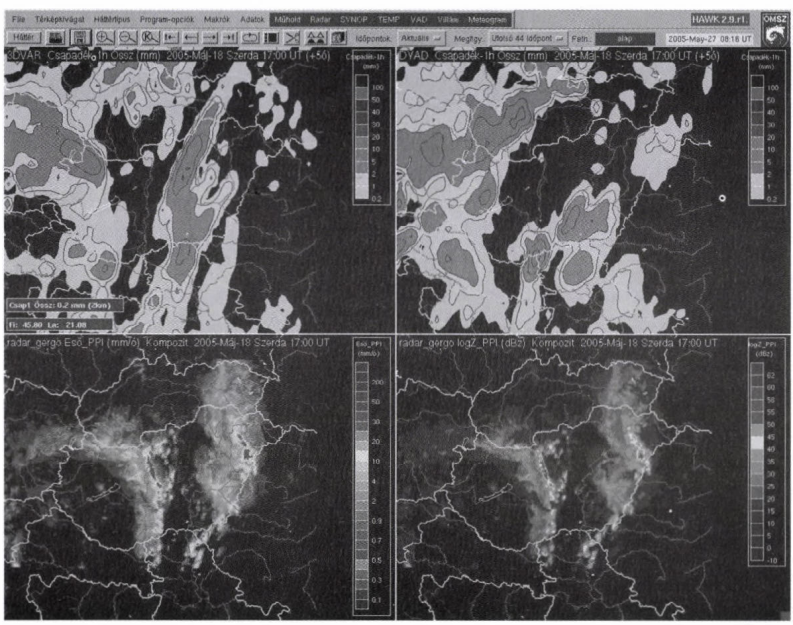


Fig. 8. Precipitation forecasts of the 3DVAR (top left) and DYA (top right) versions of the ALADIN/HU model. For comparison with the reality, the radar observations are shown, namely precipitable water (bottom left) and logZ (bottom right).

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.

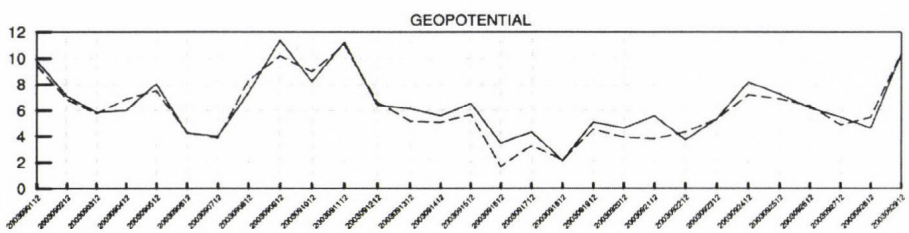


Fig. 9. Day-to day evolution of 6-hour forecast RMSE scores for geopotential on 850 hPa over a selected period (September 21–29, 2003). Solid line: 6-hour update frequency, dashed line: 3-hour update frequency.

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.

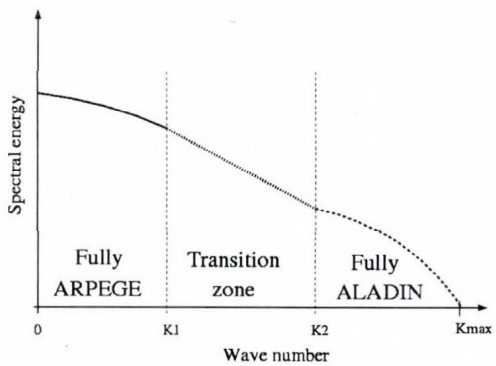


Fig. 10. The method of the explicit blending (explanation in the text).

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.

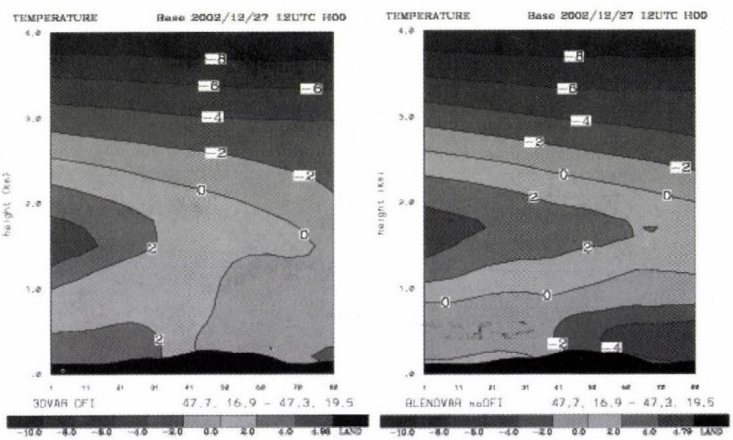


Fig. 11. Vertical cross sections of temperature fields obtained by the 3DVAR analysis without (left) and with (right) explicit blending.

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].

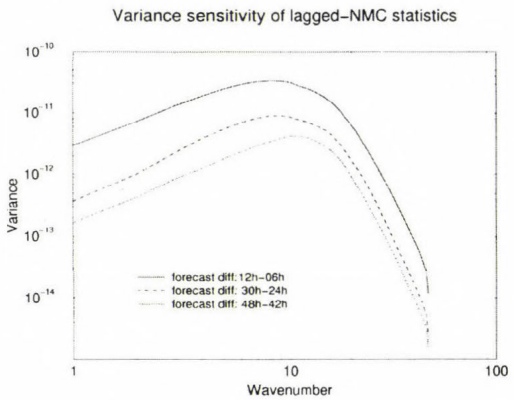


Fig. 12. Distribution of the spectral variance with respect to the wave number for different forecast range pairs used for the computation of lagged-NMC background error statistics of the ALADIN/HU model (divergence error variance at level 13).

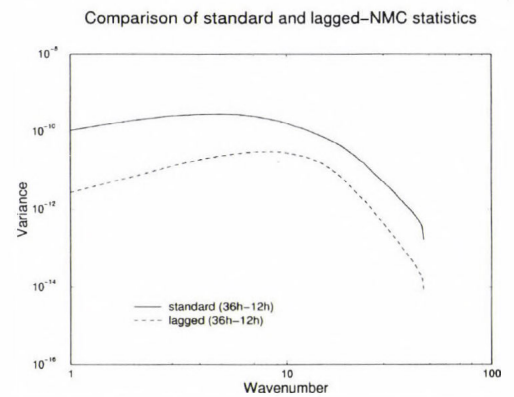


Fig. 13. Distribution of the spectral variance with respect to the wave number for standard and lagged-NMC background error statistics of the ALADIN/HU model (divergence error variance at level 13).

[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.

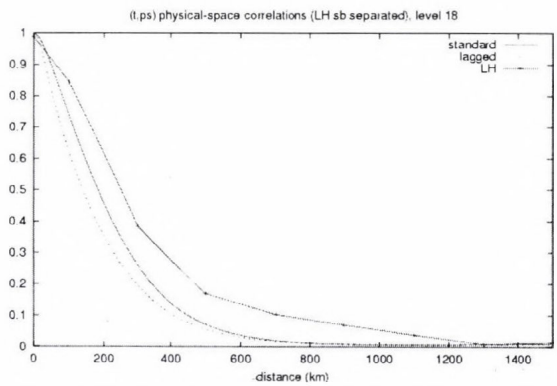


Fig. 14. Comparison of horizontal correlations obtained by the NMC and innovation methods. Standard: standard NMC method, lagged: lagged-NMC method, LH: innovation method.

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.

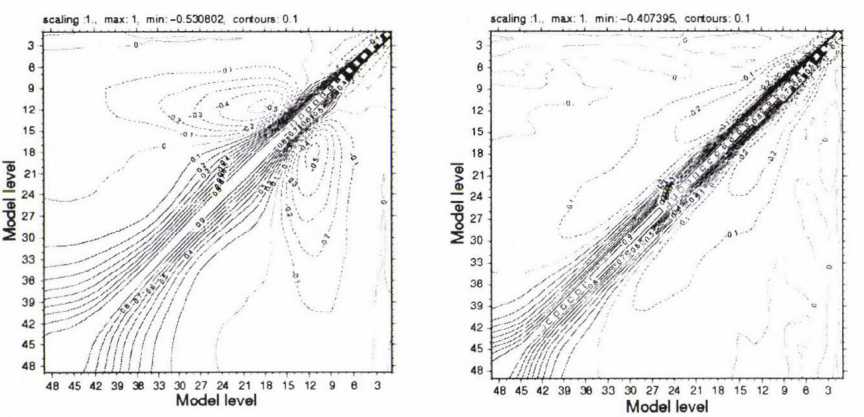


Fig. 15. Vertical temperature correlations computed upon the NMC sample (left) and the ensemble sample (right).

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	al25/cy24t1	Al28/cy28t3	al28/cy28t3
	Horizontal resolution	12 km	12 km	8 km
	Vertical levels form surface up to 5 hPa	37	37	49
3D-Var	Covariance matrix B: std NMC			
	6-hour assim. cycling	RTTOV-6	RTTOV-7	RTTOV-7
	RTM model: RTTOV	Coupling: every 6 h	Coupling: every 3 h	Coupling: every 3 h
	Coupling files:			
	ARPEGE	NOAA-15&16	NOAA-15,16&17	NOAA-15,16&17
	long cut-off files	AMSU-A	AMSU-A&B	AMSU-A&B
	Available satellite observations	AMSU-A (5-12)	AMSU-A (5-12), AMSU-B (3-5)	AMSU-A (5-12), AMSU-B (3-5)
	Selected channels	Multivariate and univariate	Multivariate	Multivariate
	Humidity assimilation			
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.

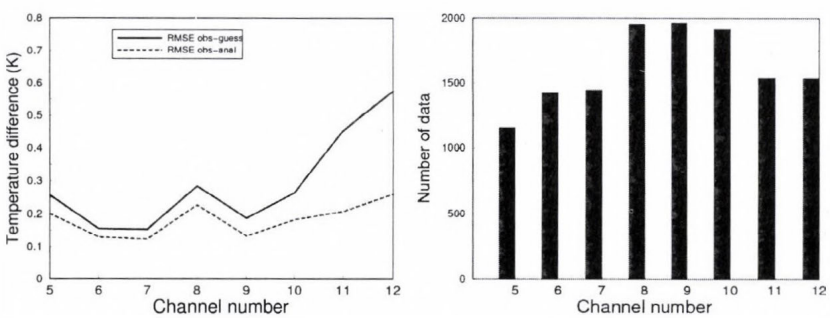


Fig. 1. The statistics of the first-guess departures (solid line) and analysis increments (dashed line) for the AMSU-A channels at 12:00 UTC for a five-day cycling (February 20–25, 2003) (left), and the number of data used in the computation (right).

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*).

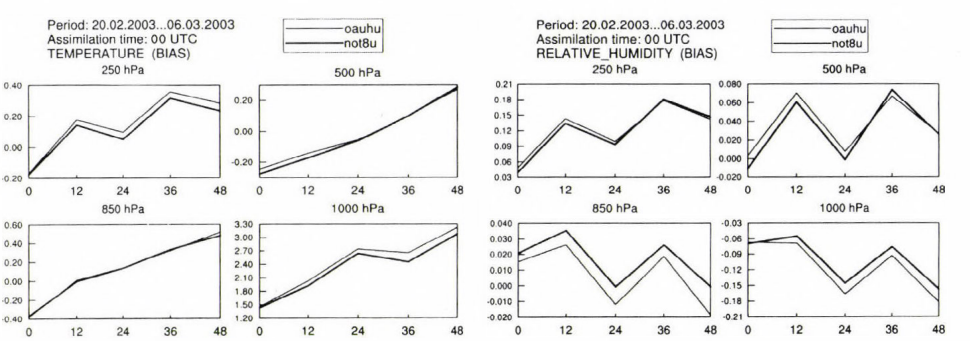


Fig. 2. Temperature and relative humidity biases for the runs with (not8u: bold line) and without (oauhu: thin line) AMSU-A data at the analysis (0) and subsequent forecast times (12- to 48-hour). The bias was computed from differences against observations.

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.

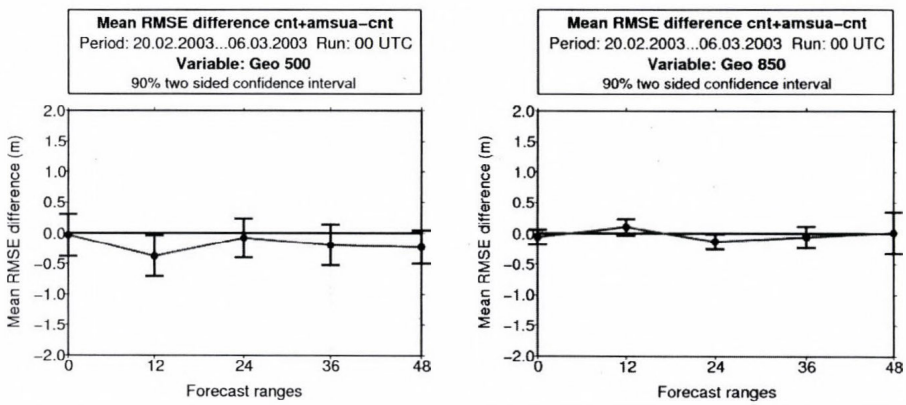


Fig. 3. Significance test: geopotential RMSE difference at 850 and 500 hPa ("cnt" stands for the control run and "cnt+amsua" for the experiment using AMSU-A data). The RMSE was computed from differences against observations.

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.

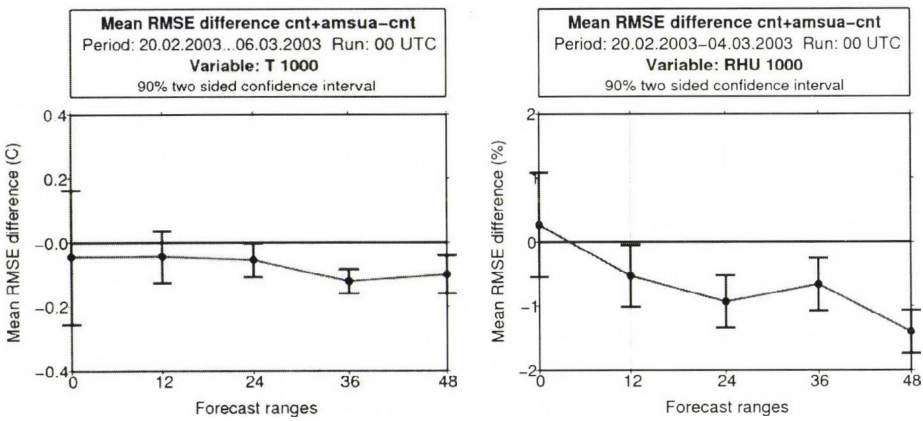


Fig. 4. Significance test: temperature (left) and relative humidity (right) RMSE differences at 1000 hPa ("cnt" stands for the control run and "cnt+amsua" for the experiment using AMSU-A data). The RMSE was computed from differences against observations.

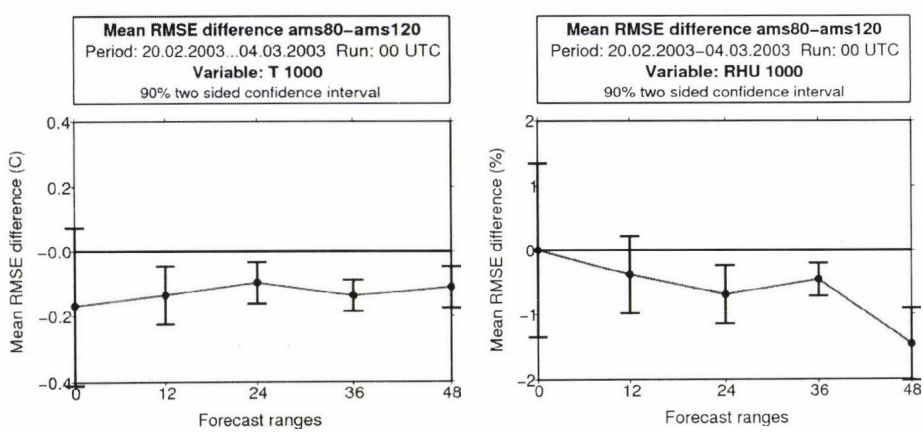


Fig. 5. Significance test: temperature (left) and relative humidity (right) RMSE differences at 1000 hPa ("ams80" stands for 80 km and "ams120" for 120 km thinning distance while using the AMSU-A data). The RMSE was computed from differences against observations.

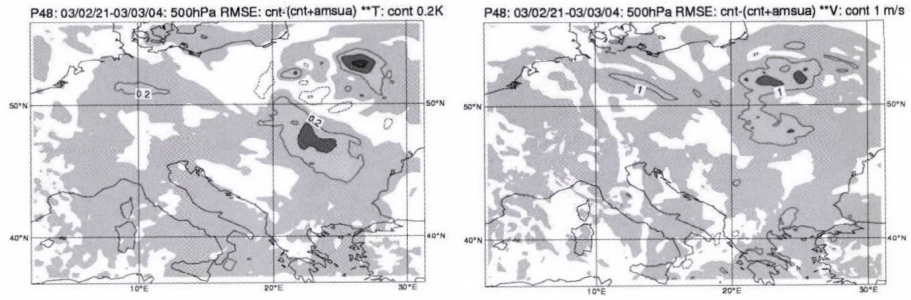


Fig. 6. Objective verification against the long cut-off ARPEGE analyses at 500 hPa for the 48-hour forecast of temperature (left) and wind speed (right). These graphs show the RMSE difference between runs without and with AMSU-A data. Colored areas show the positive impact averaged for a period of two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

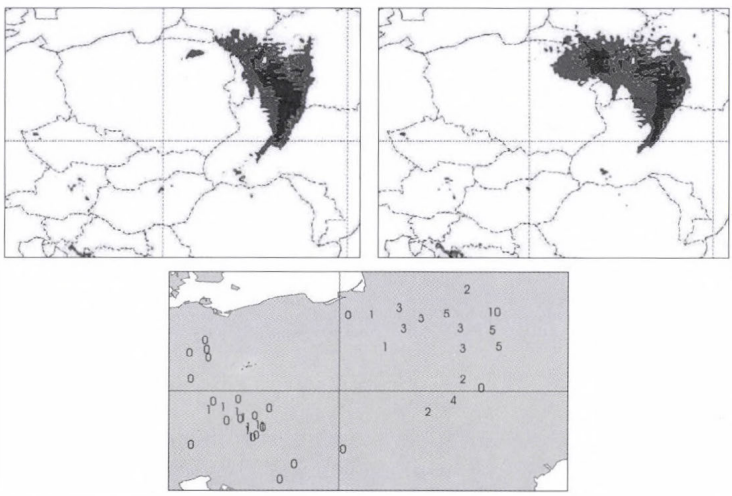


Fig. 7. The 24-hour cumulative precipitation (in mm) predicted over a zoomed area within the ALADIN/HU domain from March 04, 2003, 00:00 UTC (contour 0, 5, 15, ... mm). Upper left: control run (with TEMP and SYNOP). Upper right: 3D-Var run with ATOVS assimilated in 80 km resolution. The 24-hour cumulated 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*).

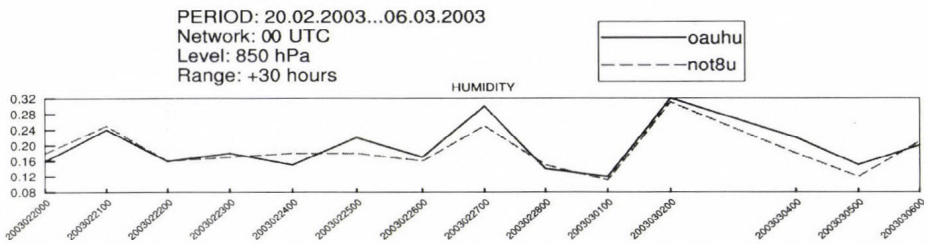


Fig. 8. Time series of the day-to-day RMSE for the 30-hour forecast of relative humidity for the runs with (not80) and without (oauhu) assimilation of AMSU-B data. The RMSE was computed from differences against observations.

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.

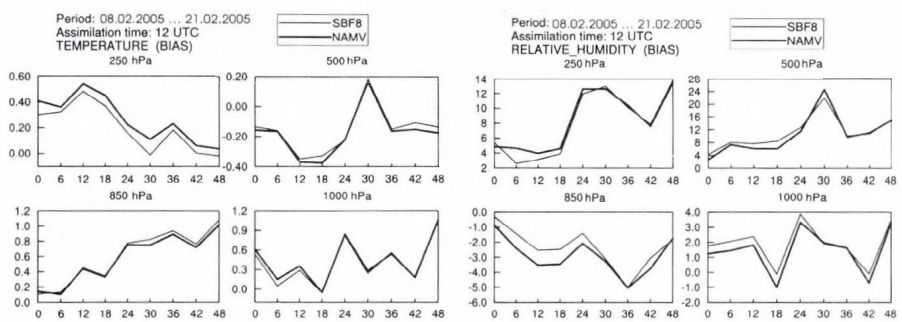


Fig. 9. Temperature and relative humidity biases for the runs with (SBF8: thin line) and without (NAMV: bold line) AMSU-B data at the analysis (0) and subsequent forecast times. The bias was computed from differences against observations.

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).

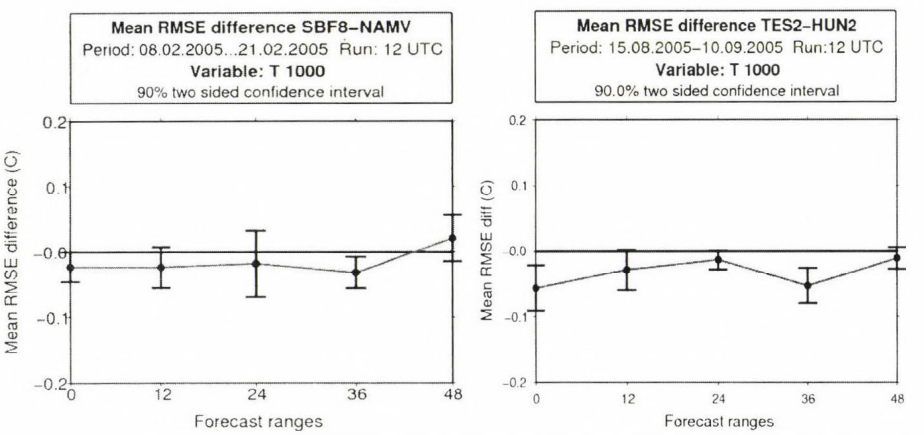


Fig. 10. Significance test: temperature RMSE differences at 1000 hPa for two different periods (NAMV and HUN2 are the reference runs without AMSU-B data, SBF8 and TES2 are experiments including AMSU-B data). The RMSE was computed from differences against observations.

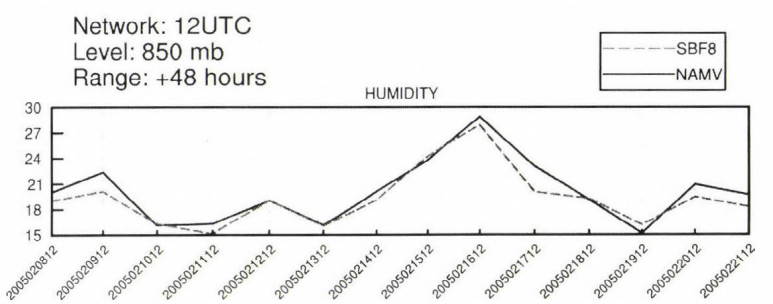


Fig. 11. Time series of the day-to-day RMSE for the 48-hour forecast of relative humidity for the runs with (SFB8) and without (NAMV) assimilation of AMSU-B data. The RMSE was computed from differences against observations.

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).

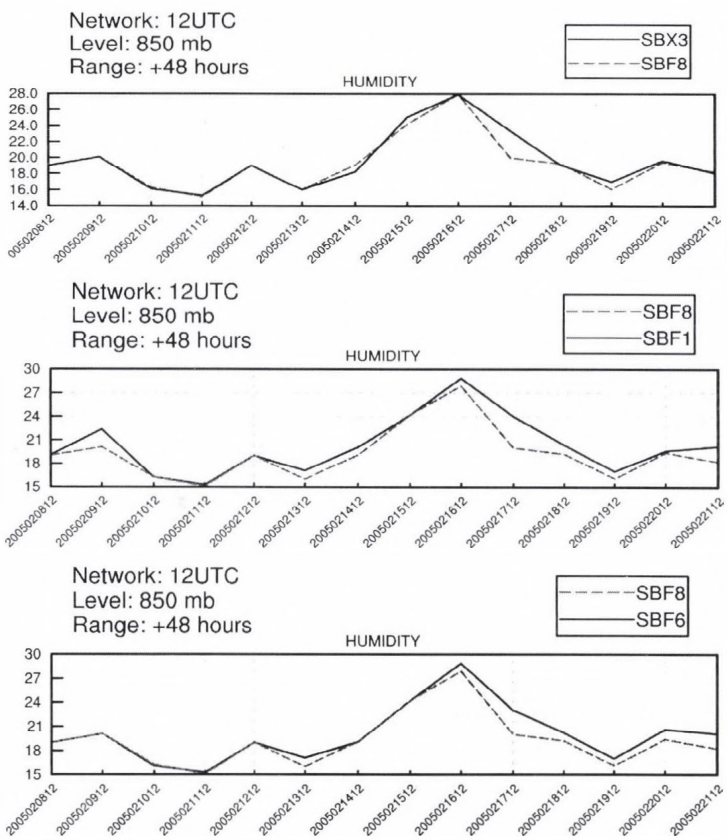


Fig. 12. Comparison of the time series of the day-to-day RMSE for relative humidity of individual runs with "sparse" AMSU-B (SBX3) and full grid AMSU-B data assimilated in 80 (SBF8), 60 (SBF6), and 120 km (SBF1) resolutions. The RMSE was computed from differences against observations.

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.

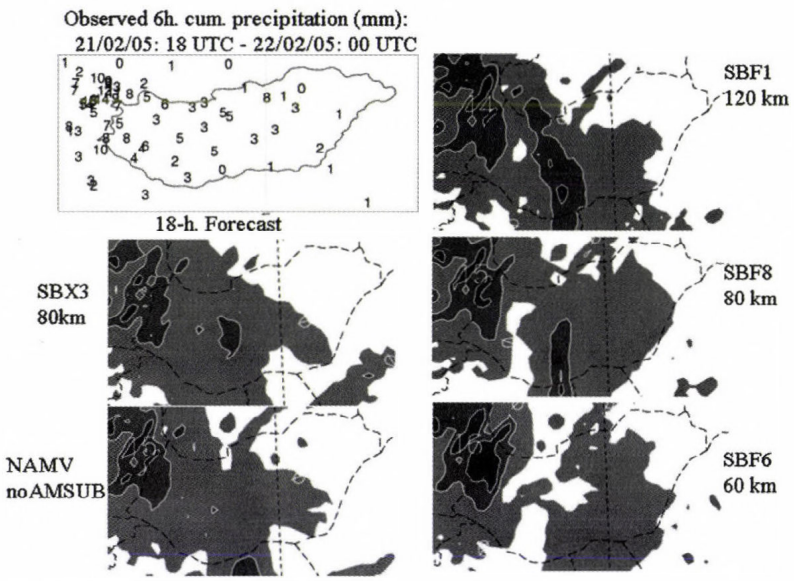


Fig. 13. Observed (top, upper left) and predicted 6-hour cumulated (estimated from 12 till 18 hours) precipitation amount valid for February 22, 2005, 06:00 UTC. Contour lines: 0, 1, 5, 10, 30, ... mm.

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.

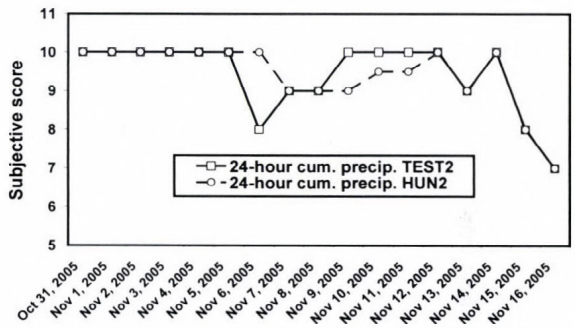


Fig. 14. Subjective scores for the 24-hour cumulated precipitation of the run in parallel suite using AMSU-B data (TEST2 – solid line), and the operational run without AMSU-B data (HUN2 – dashed line). The comparison is valid for the Hungarian territory and close surrounding regions. The forecast is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

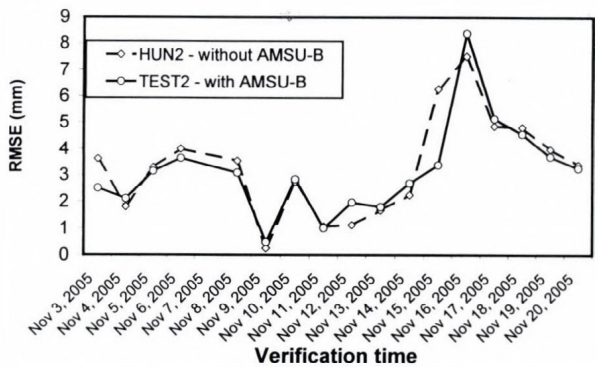


Fig. 15. Objective scores (RMSE) for the 24-hour cumulated precipitation of the run in parallel suite using AMSU-B data (TEST2 – solid line), and of the operational run without AMSU-B data (HUN2 – dashed line). The comparison is valid for the whole ALADIN/HU domain. The forecast is from the 00:00 UTC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

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

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.

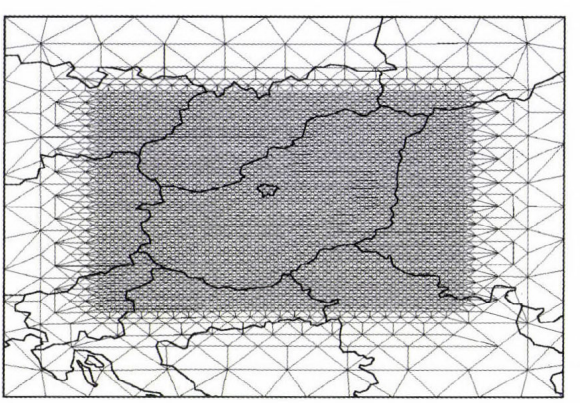


Fig. 1. The typical nested grid structure of the dispersion model. The average mesh length of the outer coarse grid and that of the nested fine grid are 100 and 12.5 km, respectively.

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 mth order rate constants with units (molecule $cm^3)^{m-1}$ s^{-1}. The photolysis rates were parameterized by the following function:

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

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

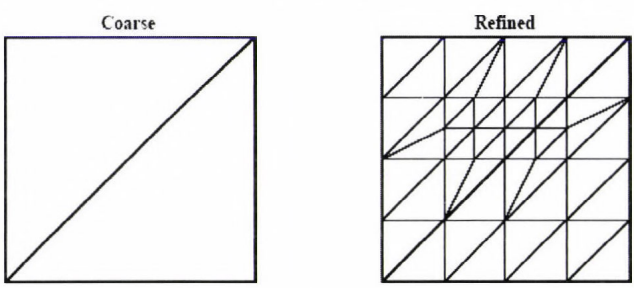


Fig. 2. Subdivision of the triangular cells using adaptive gridding strategy.

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.

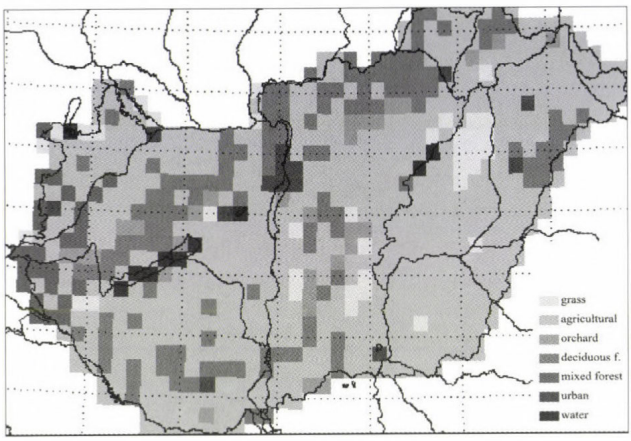


Fig. 3. Land use categories in the model.

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).

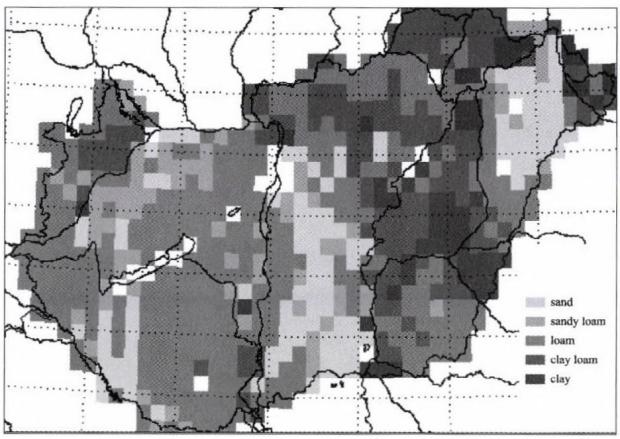


Fig. 4. Soil types in the model.

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-NO_2-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:

$$F_t = c_r (R_a + R_b + R_c)^{-1} = c_c R_c^{-1}, \quad (1)$$

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:

$$F_{st} = c_c R_{st}^{-1}, \quad (2)$$

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:

$$F_{st} = F_t R_c R_{st}^{-1}. \quad (3)$$

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.

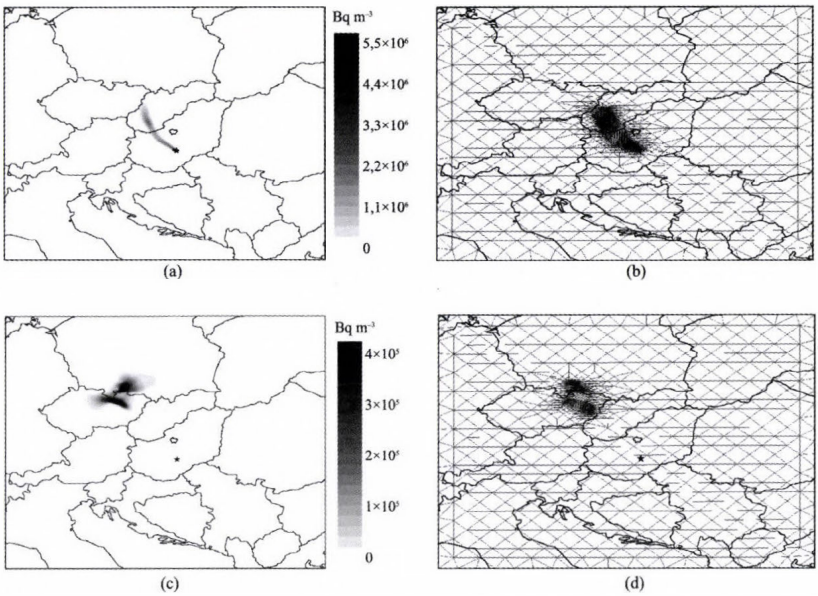


Fig. 1. Variation of the grid structure and activity of isotope ^{131}I (adaptive grid calculation). Simulation started at August 2, 1998, 00:00 UTC. (a), (b) Activity on the surface layer at t_0+12 and t_0+24. (c), (d) The adaptive mesh at t_0+12 and t_0+24.

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.

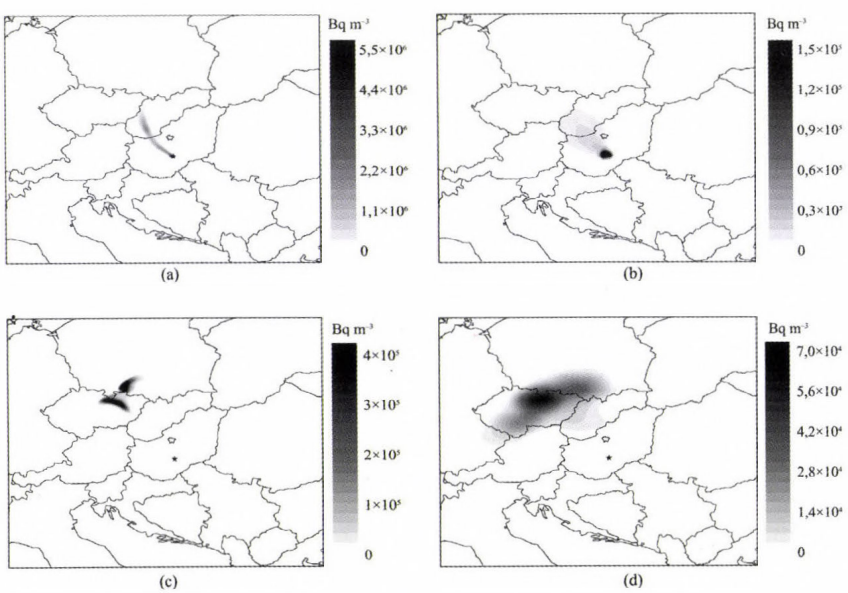


Fig. 2. Surface layer activity of isotope ^{131}I using two different fixed grid size schemes. Simulation started at August 2, 1998, 00:00 UTC. (a), (b) Activity on the surface layer at t_0+12 and t_0+24 using fine grid. (c), (d) Activity on the surface layer at t_0+12 and t_0+24 using coarse grid.

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 μg 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.

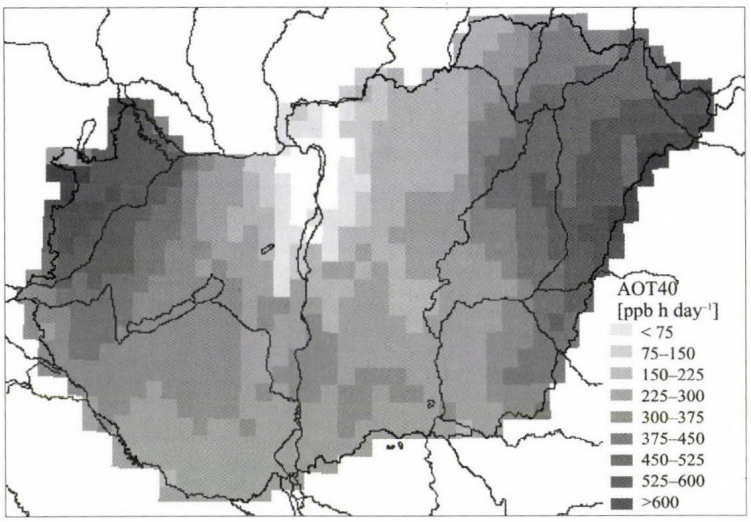


Fig. 3. Calculated AOT 40 index over Hungary on July 23, 1998 (sum of hourly values, when global radiation was greater than 50 W m^{-2}).

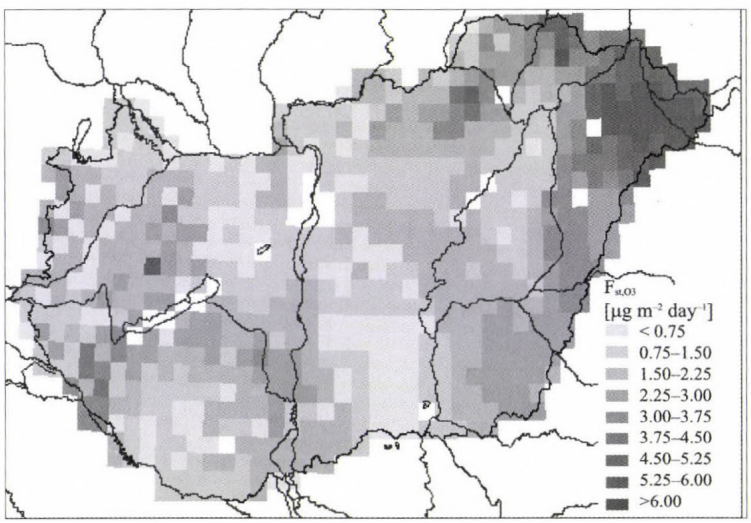


Fig. 4. Calculated cumulative stomatal flux over Hungary on July 23, 1998 (sum of hourly values, when global radiation was greater than 50 W m^{-2}).

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

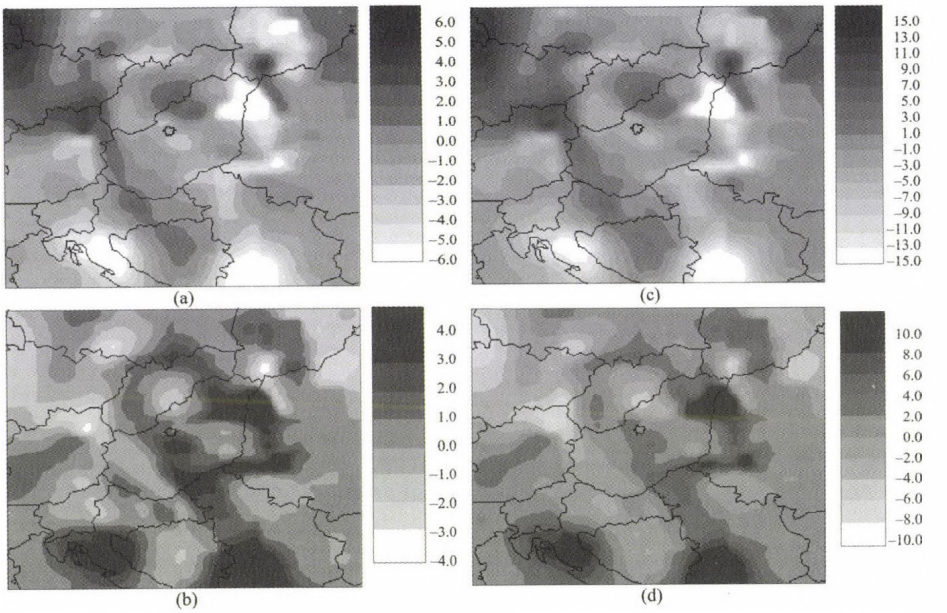


Fig. 5. Differences of the ozone mixing ratios between the perturbed and original cases at July 23, 1998, 16:00 UTC: (a) −10%, (b) +10%, (c) −20%, and (d) +20% of the original height of the mixing layer. Values are in ppb.

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.

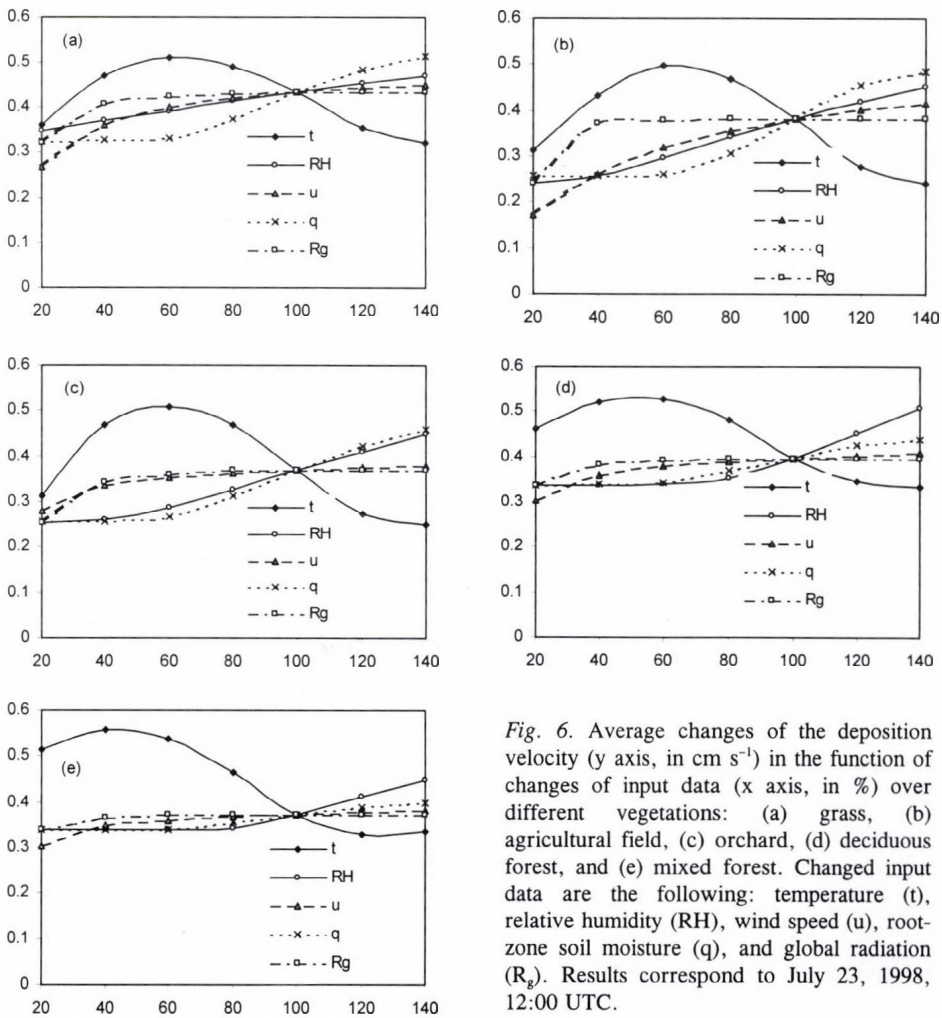


Fig. 6. Average changes of the deposition velocity (y axis, in cm s^{-1}) in the function of changes of input data (x axis, in %) over different vegetations: (a) grass, (b) agricultural field, (c) orchard, (d) deciduous forest, and (e) mixed forest. Changed input data are the following: temperature (t), 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 $i = 1, 2,...n$) 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):

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \nabla(K\nabla c) + E - \sigma c + R(c) & t \in (0,T] \\ c(\mathbf{x},0) = c_0(\mathbf{x}) \end{cases}, \qquad (2)$$

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

$$u_{\text{spl}}(k\tau) = u_k^{(2)}(k\tau), \qquad \text{for } k = 1, 2,\ldots, M. \tag{7}$$

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

$$u_{\text{spl}}(k\tau) = u_k^{(3)}(k\tau), \qquad \text{for } k = 1, 2,\ldots, \text{M}. \tag{11}$$

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 $\rightarrow$ A as follows:

$$\begin{cases} \dfrac{dv_k^{(1)}(t)}{dt} = Bv_k^{(1)}(t), \quad t \in ((k-1)\tau, k\tau], \\ v_k^{(1)}((k-1)\tau) = u_{\text{spl}}((k-1)\tau), \end{cases} \tag{14}$$

$$\begin{cases} \dfrac{dv_k^{(2)}(t)}{dt} = Av_k^{(2)}(t), \quad t \in ((k-1)\tau, k\tau], \\ v_k^{(2)}((k-1)\tau) = v_k^{(1)}(k\tau). \end{cases} \tag{15}$$

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

$$\begin{cases} \dfrac{du_k^{(1)}(t)}{dt} = 0.5Au_k^{(1)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(1)}((k-1)\tau) = u_{k-1}^{(2)}((k-1)\tau), \end{cases}$$

and

$$\begin{cases} \dfrac{du_k^{(2)}(t)}{dt} = 0.5Au_k^{(2)}(t), \quad t \in ((k-1)\tau, k\tau], \\ u_k^{(2)}((k-1)\tau) = u_k^{(1)}(k\tau). \end{cases}$$

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

$$u_{\text{spl}}(k\tau) = (I + 0.5\Delta tA)(I - 0.5\Delta tA)^{-1} u_{\text{spl}}((k-1)\tau), \text{ for } k = 1, 2, \ldots, \text{M}, \quad (23)$$

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

$$u_{\text{spl}}(k\tau) = (I - 0.5\Delta tA)^{-1}(I - 0.5\Delta tA)^{-1} u_{\text{spl}}((k-1)\tau), \text{ for } k = 1, 2, \ldots, \text{M}. \quad (24)$$

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

$$\frac{\partial u(x,t)}{\partial t} = (10^6 + \sin xt)\frac{\partial^2 u(x,t)}{\partial x^2}, \quad t \in (0,T).$$

When the above problem is used with the explicit Euler method, then the condition of stability is

$$\frac{\tau}{h^2} \leq \frac{1}{\max 2(10^6 + \sin xt)} \approx 0.5 \cdot 10^{-6},$$

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

$$\frac{\partial u_1(x,t)}{\partial t} = 10^6 \frac{\partial^2 u_1(x,t)}{\partial x^2}, \quad t \in ((k-1)\tau, k\tau]),$$

$$\frac{\partial u_2(x,t)}{\partial t} = \sin(xt)\frac{\partial^2 u_2(x,t)}{\partial x^2}, \quad t \in ((k-1)\tau, k\tau]),$$

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

$$\frac{\tau}{h^2} \le \frac{1}{2 \max \sin(xt)} \le 0.5 .$$

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

$$\frac{\tau}{h} \le \frac{1}{10^6}.$$

Applying the sequential splitting we get the sub-problems

$$\frac{\partial u_1(x,t)}{\partial t} = 10^6 \frac{\partial u_1(x,t)}{\partial x}, \quad t \in ((k-1)\tau, k\tau]),$$

$$\frac{\partial u_2(x,t)}{\partial t} = \sin(xt) \frac{\partial u_2(x,t)}{\partial x}, \quad t \in ((k-1)\tau, k\tau]),$$

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:

$$Au = -\nabla(\mathbf{u}c) + \nabla(K\nabla c) + E - \sigma c + R(c) \ ,$$
$$Au = \sum_{i=1}^{5} B_i u(t) \ ,$$

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 $\mathbf{M}$ is a bounded linear operator of type $\mathbf{R}^N \to \mathbf{R}^N$ (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 $u_0 \in \mathbf{R}^N$ is a given initial value. We remark that we investigate the case when $\mathbf{M}$ 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 $\mathbf{M}$ can be written as the sum of two bounded linear sub-operators $\mathbf{A}$ and $\mathbf{B}$

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:

$$\begin{cases} \dfrac{\mathrm{d}u_1^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_1^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_1^{(k)}((k-1)\tau) = u_{\mathrm{spl}}((k-1)\tau) \end{cases},$$

$$\begin{cases} \dfrac{\mathrm{d}u_2^{(k)}(t)}{\mathrm{d}t} = \mathbf{B}u_2^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_2^{(k)}((k-1)\tau) = u_1^{(k)}(k\tau) \end{cases}, \qquad (2)$$

$$u_{\mathrm{spl}}(k\tau) := u_2^{(k)}(k\tau),$$

with $k = 1,\ldots,m$, 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 $\{k\tau : k = 1,\ldots,m\}$. 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 $\mathbf{A} \to \mathbf{B}$, and afterwards $\mathbf{B} \to \mathbf{A}$. 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.

$$\begin{cases} \dfrac{\mathrm{d}u_1^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_1^{(k)}(t), & t \in ((k-1)\tau, (k-1)\tau + \tau/2] \\ u_1^{(k)}((k-1)\tau) = u_{\mathrm{spl}}((k-1)\tau) \end{cases},$$

$$\begin{cases} \dfrac{\mathrm{d}u_2^{(k)}(t)}{\mathrm{d}t} = \mathbf{B}u_2^{(k)}(t), & t \in ((k-1)\tau, k\tau] \\ u_2^{(k)}((k-1)\tau) = u_1^{(k)}((k-1)\tau + \tau/2) \end{cases}, \tag{3}$$

$$\begin{cases} \dfrac{\mathrm{d}u_3^{(k)}(t)}{\mathrm{d}t} = \mathbf{A}u_3^{(k)}(t), & t \in ((k-1)\tau + \tau/2, k\tau] \\ u_3^{(k)}((k-1)\tau) = u_2^{(k)}(k\tau) \end{cases},$$

$$u_{\mathrm{spl}}(k\tau) := u_3^{(k)}(k\tau),$$

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:

$$\begin{aligned} u(\tau) &= \exp(\tau(\mathbf{A}+\mathbf{B}))u_0, \\ u_{\mathrm{spl}}^{\mathrm{sq}}(\tau) &= \exp(\tau\mathbf{B})\exp(\tau\mathbf{A})u_0, \\ u_{\mathrm{spl}}^{\Theta}(\tau) &= \Theta\exp(\tau\mathbf{B})\exp(\tau\mathbf{A})u_0 + (1-\Theta)\exp(\tau\mathbf{A})\exp(\tau\mathbf{B})u_0, \\ u_{\mathrm{spl}}^{\mathrm{S}}(\tau) &= \exp(\tau/2\mathbf{A})\exp(\tau\mathbf{B})\exp(\tau/2\mathbf{A})u_0, \end{aligned} \tag{4}$$

where $u(\tau)$, $u_{\mathrm{spl}}^{\mathrm{sq}}(\tau)$, $u_{\mathrm{spl}}^{\Theta}(\tau)$, and $u_{\mathrm{spl}}^{\mathrm{S}}(\tau)$ 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 $\mathbf{M}$ 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:

$$\varepsilon_{\mathrm{spl}}(\tau) = \left| u(\tau) - u_{\mathrm{spl}}(\tau) \right|,$$

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:

$$p := \sup\left\{ q \in N : \lim_{\tau \to 0} \frac{E_{\mathrm{spl}}(\tau)}{\tau^{q+1}} < +\infty \right\}. \tag{5}$$

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 **[[A,B],A - B] = 0** holds for the symmetrically weighted splitting (where **[A,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 $\Delta t \leq \tau$ can be better understood from *Fig. 1.*

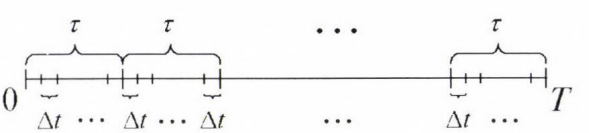


Fig. 1. Time steps of the numerical solving process when applying a splitting procedure with splitting time step τ, and a numerical method with time step Δt (for solving the split sub-problems).

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 $\varepsilon_{\text{tot}}(\tau)$, however, we can only estimate it, for instance, in the following way:

$$\varepsilon_{\text{tot}}(\tau) = \left|u(\tau) - y_{\text{spl}}\right| \le \left|u(\tau) - u_{\text{spl}}(\tau)\right| + \left|u_{\text{spl}}(\tau) - y_{\text{spl}}\right| = \varepsilon_{\text{spl}}(\tau) + \varepsilon_{\text{int}}(\tau), \quad (7)$$

where $\varepsilon_{\text{spl}}(\tau)$ is the above defined local splitting error. We call the term $\varepsilon_{\text{int}}(\tau)$ *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 $\varepsilon_{\text{tot}}(\tau)$ is equal to the order of $\varepsilon_{\text{spl}}(\tau)$ only if the order of $\varepsilon_{\text{int}}(\tau)$ is higher than the order of $\varepsilon_{\text{spl}}(\tau)$. 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 $E_{\text{int}}(\tau)$ represents the global numerical error at time $t = \tau$, i.e., after the first splitting step. Therefore, we can estimate $\varepsilon_{\text{int}}(\tau)$ as:

$$\varepsilon_{\text{int}}(\tau) = \mathrm{O}\left(\tau^{rs+1}\right). \quad (8)$$

It can be seen from Eqs. (6), (7), and (8) that the order of the local total error $\varepsilon_{\text{tot}}(\tau)$ 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 $\varepsilon_{\text{tot}}(\tau)$ 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 subproblems. Then the numerical solutions have the following forms at time $t = \tau$:

$$
\begin{aligned}
y_{\mathrm{spl}}^{\mathrm{sq}} &= (1+\Delta t\mathbf{A})^n (1+\Delta t\mathbf{B})^n u_0, \\
y_{\mathrm{spl}}^{\Theta} &= \Theta(1+\Delta t\mathbf{A})^n (1+\Delta t\mathbf{B})^n u_0 + (1-\Theta)(1+\Delta t\mathbf{B})^n (1+\Delta t\mathbf{A})^n u_0, \\
y_{\mathrm{spl}}^{\mathrm{S}} &= (1+\Delta t\mathbf{A})^{n/2} (1+\Delta t\mathbf{B})^n (1+\Delta t\mathbf{A})^{n/2} u_0,
\end{aligned}
\tag{9}
$$

where $y_{\mathrm{spl}}^{\mathrm{sq}}$, $y_{\mathrm{spl}}^{\Theta}$, and $y_{\mathrm{spl}}^{\mathrm{S}}$ 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:

$$
\begin{aligned}
u(\tau) &= \left[\mathbf{I} + \tau(\mathbf{A}+\mathbf{B}) + \frac{\tau^2}{2}\left(\mathbf{A}^2 + \mathbf{B}^2 + \mathbf{AB} + \mathbf{BA}\right) + \mathrm{O}(\tau^3)\right] u_0, \\
y_{\mathrm{spl}}^{\mathrm{sq}} &= \left[\mathbf{I} + \tau(\mathbf{A}+\mathbf{B}) + \frac{\tau^2}{2}\left(\left(1-\frac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + 2\mathbf{BA}\right) + \mathrm{O}(\tau^3)\right] u_0, \\
y_{\mathrm{spl}}^{\Theta} &= \left[\mathbf{I} + \tau(\mathbf{A}+\mathbf{B}) + \frac{\tau^2}{2}\left[\left(1-\frac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + (1-\Theta)\,\mathbf{AB} + \Theta\mathbf{BA}\right] + \mathrm{O}(\tau^3)\right] u_0, \\
y_{\mathrm{spl}}^{\mathrm{S}} &= \left[\mathbf{I} + \tau(\mathbf{A}+\mathbf{B}) + \frac{\tau^2}{2}\left[\left(1-\frac{1}{n}\right)\left(\mathbf{A}^2 + \mathbf{B}^2\right) + \mathbf{AB} + \mathbf{BA}\right] + \mathrm{O}(\tau^3)\right] u_0,
\end{aligned}
\tag{10}
$$

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:

$$\varepsilon_{\text{spl}}(\tau) = \gamma \frac{\tau^2}{2} \| (\mathbf{BA} - \mathbf{AB}) u_0 \| + \mathrm{O}(\tau^3), \tag{11}$$

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:

$$\varepsilon_{\text{tot}}(\tau) = \varepsilon_{\text{spl}}(\tau) + \varepsilon_{\text{int}}(\tau) + \mathrm{O}(\tau^3), \tag{12}$$

where $\varepsilon_{\text{int}}(\tau)$ denotes the interaction error having the following form:

$$\varepsilon_{\text{int}}(\tau) = \frac{\tau \Delta t}{2} \| (\mathbf{A}^2 - \mathbf{B}^2) u_0 \| + \mathrm{O}(\tau^3). \tag{13}$$

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 $\varepsilon_{\text{tot}}(\tau)$. 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 $\mathbf{M}$ 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 $s \geq 1$.

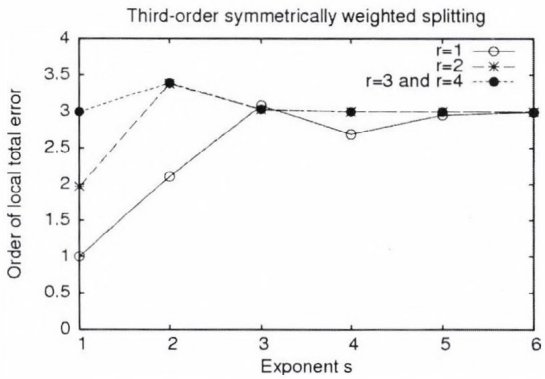


Fig. 2. Behavior of the estimation of the order of the local total error in the case of the third-order symmetrically weighted splitting ($p=3$), for four different numerical methods ($r=1,2,3,4$).

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 $t = T$.

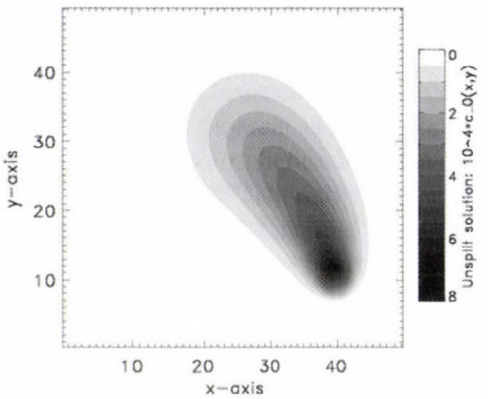


Fig. 3. Numerical solution of the model given by Eq. (16) without applying splitting procedure, at time $t = 72$ hours.

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.

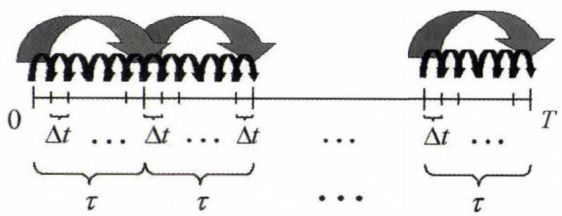


Fig. 4. Number of steps in the case of the semi-Lagrangian method (long arrows) and the finite difference method (short arrows).

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.

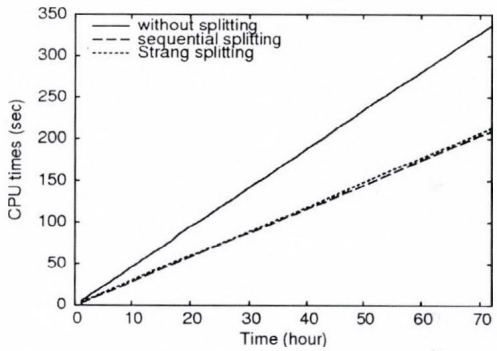


Fig. 5. Computing (CPU) times of the model with and without applying splitting procedure, applying the sequential and the Strang splittings, respectively.

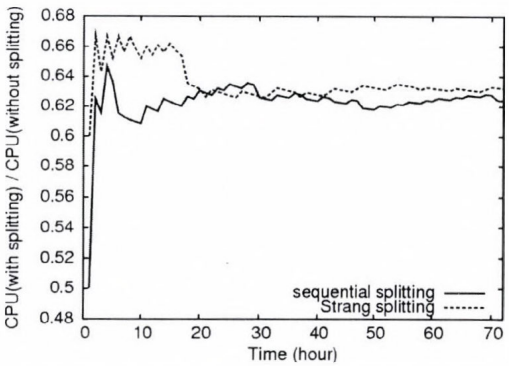


Fig. 6. Ratio of the computing (CPU) times of the split and the unsplit model, applying the sequential and the Strang splittings, respectively.

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 $c_{\text{ref}}(x,y,t)$ of Eq. (16) without applying splitting procedure and by using smaller grid sizes ($\Delta x_{\text{ref}} = \Delta y_{\text{ref}} = 0.125$ km) and time steps ($\Delta t_{\text{ref}} = 15$ 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* $\varepsilon_{\text{atot}}$ at every mesh point and every splitting time step as follows:

$$\varepsilon_{\text{atot}}(i\Delta x, j\Delta y, k\tau) = \left| c_{\text{ref}}(i\Delta x, j\Delta y, k\tau) - c_{\text{spl}}(i\Delta x, j\Delta y, k\tau) \right|, \qquad (18)$$

where $i = 0,\ldots,I$, $j = 0,\ldots,J$, and $k = 0,\ldots,m$. Moreover, we can compute the mass of the whole emitted pollution $M(t)$ from the emission function $E(x,y,t)$:

$$M(t) = \iint_{\Omega} \int_{0}^{t} E(x,y,t')\,\mathrm{d}t'\,\mathrm{d}x\mathrm{d}y. \qquad (19)$$

Similarly, we can compute the "mass" of the approximate total error, as well:

$$M_{\text{atot}}(t) = \iint_{\Omega} \varepsilon_{\text{atot}}(x,y,t)\,\mathrm{d}x\mathrm{d}y. \qquad (20)$$

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* ε_{rel} at each splitting time step, which characterizes the temporal behavior of the approximate total error well:

$$\varepsilon_{\text{rel}}(k\tau) = \frac{M_{\text{atot}}(k\tau)}{M(k\tau)}. \qquad (21)$$

In *Fig. 7* the approximate total error field ε_{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).

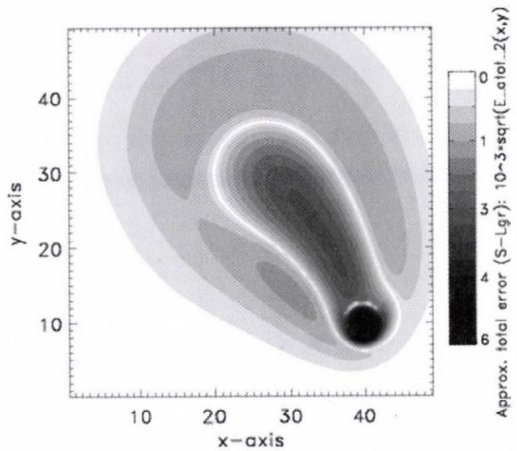


Fig. 7. Approximate total error field at time t = 72 hours, when the sequential splitting is applied. (Similar figure in the case of the Strang splitting.)

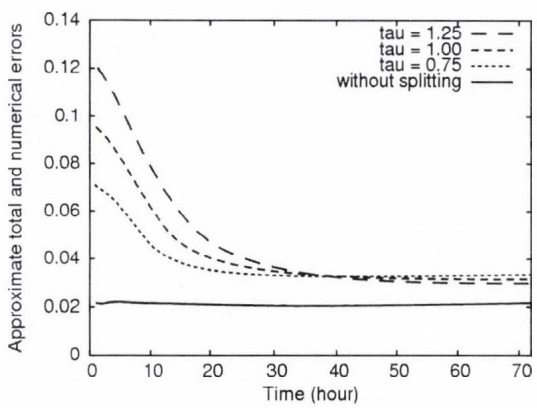


Fig. 8. Temporal behavior of the relative approximate total errors in the case of sequential splitting for three values of τ, and the approximate numerical error of the solution of the unsplit model.

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 $\varepsilon_{\text{rel}}(t)$ 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

$$A = \frac{d}{dt} T(t)|_{t=0}$$

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

$$D(A) := \left\{ x \in X : \exists \lim_{t\to 0+} \frac{T(t)x - x}{t} \in X \right\},$$

$$Ax := \lim_{t\to 0+} \frac{T(t)x - x}{t} = \frac{d}{dt}(t \to T(t)x)|_{t=0}$$

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 $1, \sqrt{2}\cos \pi n s, n \geq 2$ 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

$$(T(t)f)(s) = \int_0^1 k_t(s,r) f(r) dr, \quad f \in C[0,1], s \in [0,1],$$

$$k_t(s,r) := 1 + 2\sum_{n=0}^{+\infty} e^{-\pi^2 n^2 t} \cos \pi n s \cdot \cos \pi n r.$$

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

$$(T(t)f)(\mathbf{s}) = \frac{1}{\sqrt{4\pi t}} \int_{\mathbf{R}^n} e^{-\frac{|\mathbf{s}-\mathbf{r}|^2}{4t}} f(\mathbf{r}) d\mathbf{r}, \quad t > 0, \mathbf{s} \in \mathbf{R}^n,$$

$$T(0) = Id.$$

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

$$\sup_{s\in\mathbf{R}^n} \mathrm{Re}\, q(s) < \infty$$

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:

$$\begin{cases} \dfrac{\partial c}{\partial t} = -\nabla(\mathbf{u}c) + \Delta c + E - \sigma c + R(c), & t\in(0,T], \\ c(\mathbf{x},0) = c_0(\mathbf{x}), & \mathbf{x}\in\mathbf{R}^n. \end{cases} \tag{6}$$

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) $\lim_{t\to\infty} ||\, e^{tA} \,|| = 0$;

(b) All eigenvalues of A have negative real part, i.e., $\operatorname{Re}\lambda < 0$ 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 $T(t) := e^{tA}$, $t \geq 0$ be the strongly continuous semigroup generated by A. Then the following assertions are equivalent:

(a) $\lim_{t\to\infty} ||\, T(t) \,|| = 0$;

(b) $\operatorname{Re}\lambda < 0$ 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

$$(f, v) = B_L(u, v) \qquad \text{for all} \quad v \in H_2, \tag{3}$$

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

$$(f_h, v_h) = B_L(u_h, v_h) \quad \text{for all} \quad v_h \in V_h, \tag{4}$$

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:

  $W_h = \{w_k : \Omega \to \mathbf{R} | w_h \text{ is continuous in } \Omega,\ w_k|_{K_i} \text{ is linear, } i = 1,2,\ldots,6\}.$

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

$$w_h(a_1) = 0, w_h(a_2) = 1, w_h(a_3) = 1, w_h(a_4) = 0.5$$
$$w_h(a_5) = 1.5, w_h(a_6) = 0.5, w_h(a_7) = 2.$$

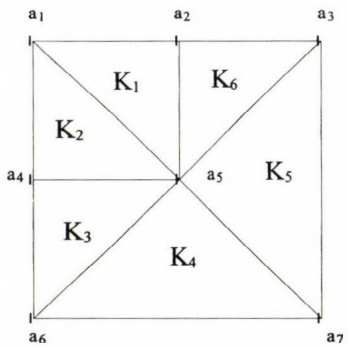


Fig. 1. The computational domain Ω in Example 1, with the subdomains K_i and vertices a_i.

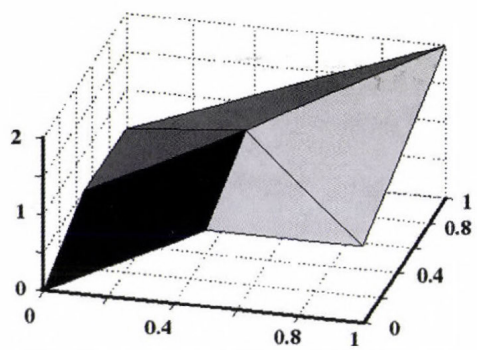


Fig. 2. The function in w_h Example 1.

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:

$$V_h = \{v_h : \Omega \to \mathbf{R} \mid w_k|_{K_i} \text{ is at most second order, } i = 1,2,\ldots,6\}.$$

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 $\lim_{\substack{x \in K_1 \\ x \to -0.5}} v_h(x) = 0.5$.

The function v_h is depicted on *Fig. 2.*

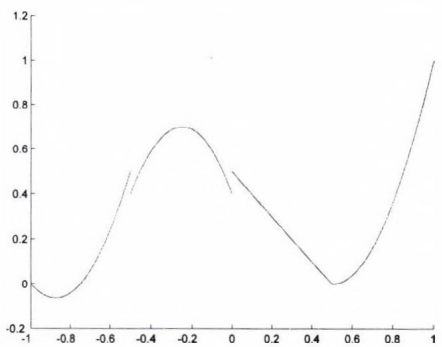


Fig. 3. The function v_h in Example 2.

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

$$(f_h, \phi_i) = B_L\left(\sum_{j=1}^{n} u_{h,j}\phi_j, \phi_i\right) \quad \text{for } i = 1,2,\ldots,n, \tag{5}$$

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:

$$\begin{cases} \int_K \nabla u \cdot v = \int_K \nabla \cdot v, \\ -\int_K \nabla \cdot \Psi w = \int_K f w, \end{cases} \tag{8}$$

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

$$\begin{cases} \int_K \Psi \cdot v = -\int_K u \nabla \cdot v + \int_{\partial K} u v_K \cdot v, \\ -\int_K \Psi \cdot \nabla w = \int_K f \Psi + \int_{\partial K} w v_K \cdot \Psi, \end{cases} \tag{9}$$

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:

$$\begin{cases} \int_K \Psi_h \cdot v = -\int_K u_h \nabla \cdot v + \int_{\partial K} u_h v_K \cdot v, \\ -\int_K \Psi_h \cdot \nabla w = \int_K f \Psi_h + \int_{\partial K} w v_K \cdot \Psi_h. \end{cases} \tag{11}$$

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, $k = 1,2\ldots,m$.

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:

$$\int_l u_{K_1} v_{K_1} \cdot \nu|_{K_1} + \int_l u_{K_2} v_{K_2} \cdot \nu|_{K_2} . \tag{13}$$

Similarly, for the boundary term on an interelement face l in the second equation of Eq. (9) reads as

$$\int_l w_{K_1} v_{K_1} \cdot \Psi|_{K_1} + \int_l w_{K_2} v_{K_2} \cdot \Psi|_{K_2} . \tag{14}$$

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:

$$\{u\} = \frac{1}{2}(u_{K_1} + u_{K_2}), \ [[u]] = u_{K_1} v_{K_1} + u_{K_2} v_{K_2} , \tag{15}$$

and similarly, for a vector valued function Ψ:

$$\{\Psi\} = \frac{1}{2}(\Psi_{K_1} + \Psi_{K_2}), \ [[\Psi]] = v_{K_1} \cdot \Psi_{K_1} + v_{K_2} \Psi_{K_2} , \tag{16}$$

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:

$$\phi(u_h) = \begin{cases} \{u_h\} + \nu_K \cdot [[u_h]] & \text{on } F_i \\ \text{and} & \\ 0 \quad \text{on } F \setminus F_i & \end{cases} \tag{18)–(19}$$

$$\phi(\Psi_h) = \{\nabla u_h\} - \eta_F h_F^{-1} [[u_h]] \quad \text{on } F,$$

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 v_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 $v_K=(1,0,0)$) and $\overline{K}=(p_2,p_3)\times(q_1,q_2)\times(r_1,r_2)$. The desired jump on l:

$$\begin{aligned} &[[u_h]](y,z)\\ &=v\left[\frac{a_{0,K}}{(p_2-p_1)(q_2-q_1)(r_2-r_1)}+a_{1,K}+a_{2,K}\frac{y-q_1}{q_2-q_1}+a_{3,K}\frac{z-r_1}{r_2-r_1}\right.\\ &\left.-\frac{a_{0,\overline{K}}}{(p_3-p_2)(q_2-q_1)(r_2-r_1)}+a_{1,\overline{K}}+a_{2,\overline{K}}\frac{y-q_1}{q_2-q_1}+a_{3,\overline{K}}\frac{z-r_1}{r_2-r_1}\right]. \end{aligned}$$

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

BUDAPEST, HUNGARY

AUTHOR INDEX

TABLE OF CONTENTS
I. Papers

II. Book reviews

III. News

SUBJECT INDEX

A

V

W

Y

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