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

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

[1]*Eötvös Loránd University, Department of Meteorology,*
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

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. The 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. The 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. The 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

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 forecast, but several ones 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) just 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 ECMWF and 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 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 members, and the downscaling is performed with the Lokal Modell. 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 Hungarian Meteorological Service (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. Secondly, 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 summarised 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 European Centre for Medium-Range Weather Forecasts (ECMWF) the Ensemble Prediction System (EPS) is integrated twice a day, from the 00 and 12 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):

1. At the beginning, consider all members as separate clusters and calculate the so-called distance matrix.
2. Find the closest pair of clusters and merge them into a single cluster.
3. Compute the distances between the new cluster and each of the old clusters.
4. 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 the 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 the 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 +84h) 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 HMS (40°N–55°N, 10°E–30°E) (*Fig. 1*).

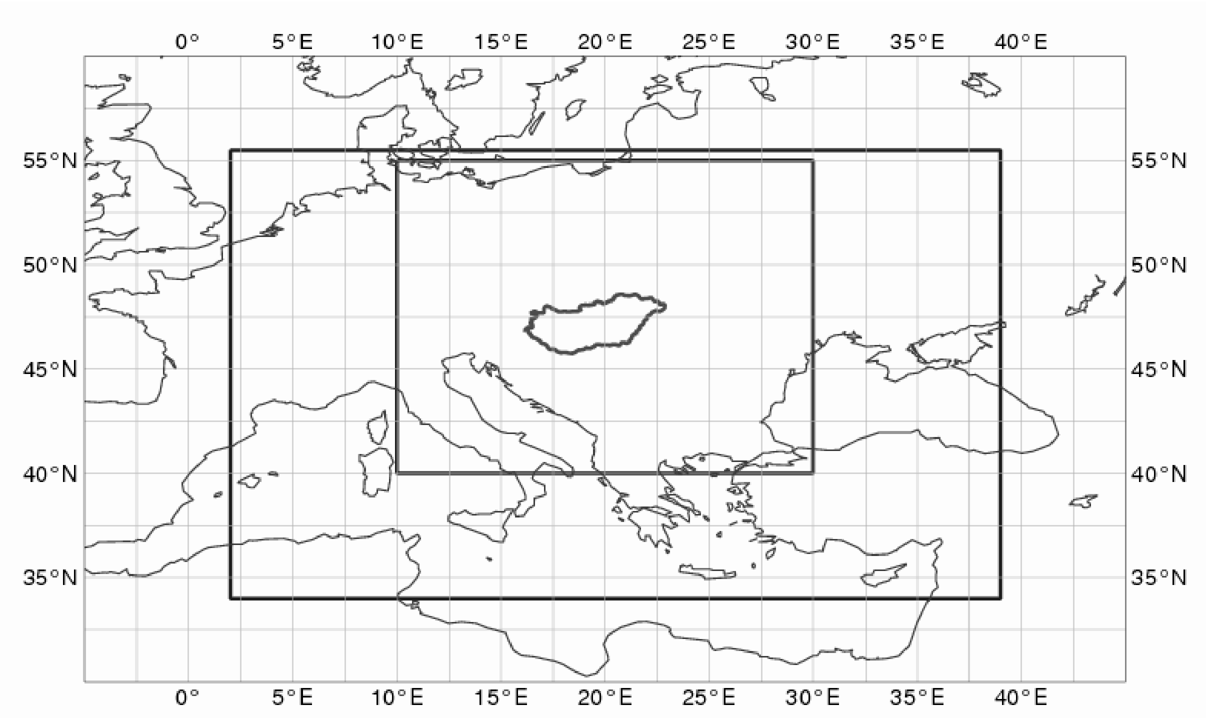


Fig. 1. The two clustering domains for the ECMWF/ALADIN downscaling system. The bigger area is equivalent to the integration domain of the ALADIN model.

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

1) Clustering on the bigger domain, using one set of ECMWF EPS (50 members)
2) Clustering on the smaller domain, using one set of ECMWF EPS (50 members)
3) Clustering on the bigger domain, using two sets of ECMWF EPS (100 members)
4) Clustering on the smaller domain, using two sets of ECMWF EPS (100 members)

While using only one set of ECMWF EPS (50 members), the 12 UTC EPS run was used. When using two sets of EPS (100 members), the 00 UTC and the 12 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 UTC EPS, or the +12h forecast of the 00 UTC EPS (depending on the fact whether the selected representative member is coming from the 00 UTC or the 12 UTC run). The clustering times were +60h and +84h for the 12 UTC EPS and +72h and +96h for the 00 UTC EPS (*Fig. 2*).

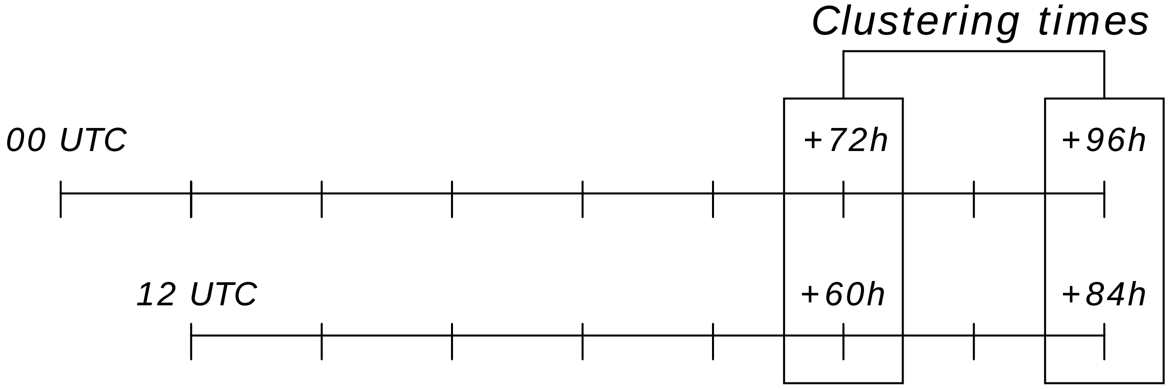


*Fig. 2***.** Schematic description of the clustering time while using 100 ECMWF EPS members.

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*):

$$stfield(i) = \frac{field(i) - cl\_mean(i)}{cl\_dev(i)}.$$

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):

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

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

In the equations above, 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 ECMWF these formulas are used operationally for the clustering of ECMWF EPS members. The population-weighting formula results 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 that it could provide a reasonable performance both in average and in 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 in case of extreme events better 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 represent the characteristics of the original super-ensemble adequately. 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 at the moment in an experimental stage. The whole system is designed in a way that its certain components (eg. 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:

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

All of the cases were related to precipitation events: for the first three cases the ECMWF EPS system underestimated the precipitation, but for 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 visualised both for the original ECMWF EPS and for 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 18 May, 2005. The last case with the precipitation overestimation will be analysed afterwards separately.

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 favourable 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 at 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.

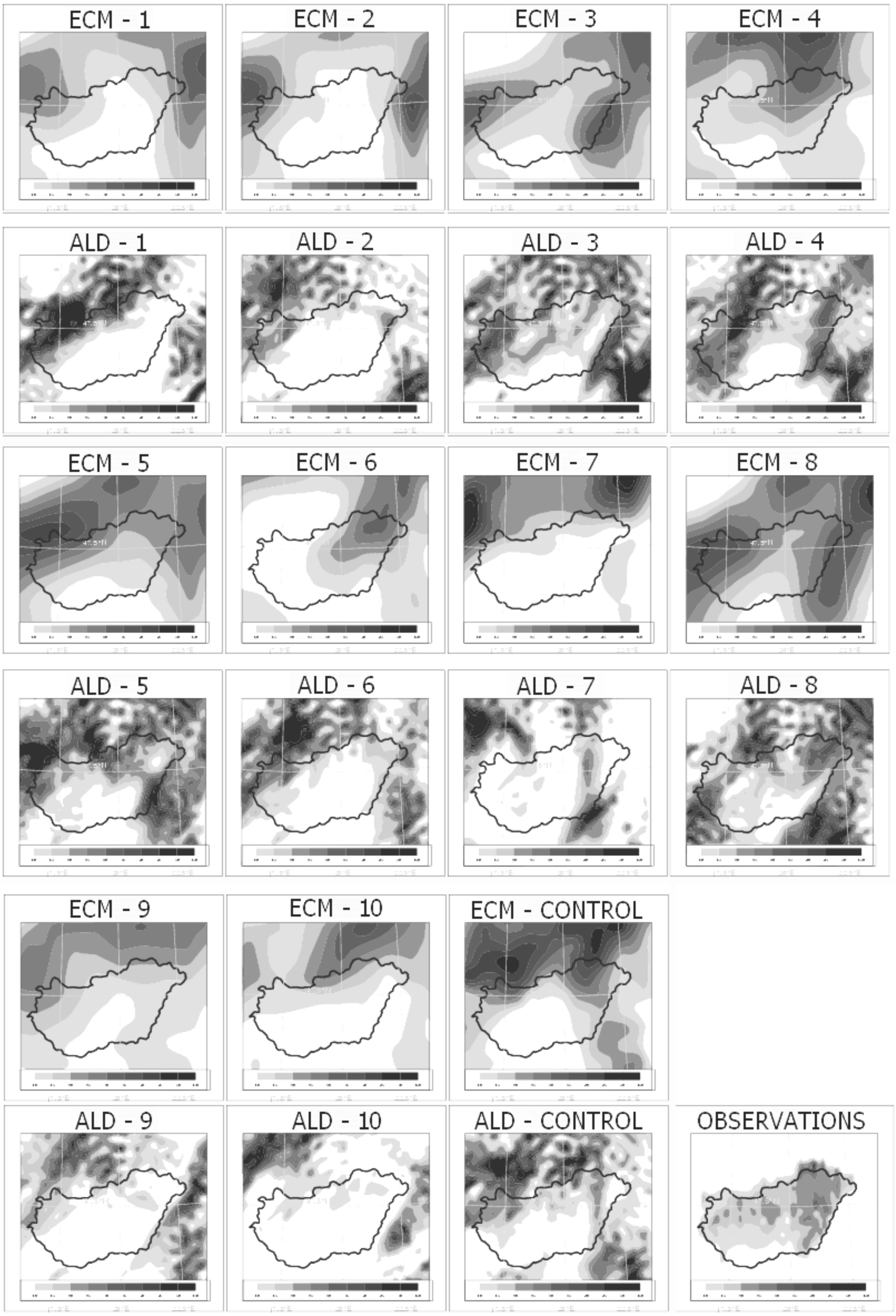


Fig. 3. Stamp diagrams for the ECMWF EPS (10 representative members of the 50 global EPS members, bigger domain clustering configuration), and the ALADIN downscaled members. 24h precipitation forecasts for May 18, 2005, 06:00 UTC – May 19, 2005, 06:00 UTC. The control forecasts (without initial perturbations)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

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 ALADIN model's dynamics and physics have also played an important role in the downscaling process. It is also noted that the other clustering configurations (other clustering domain and 100 member super-ensemble) gave very similar results (not shown).

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*).

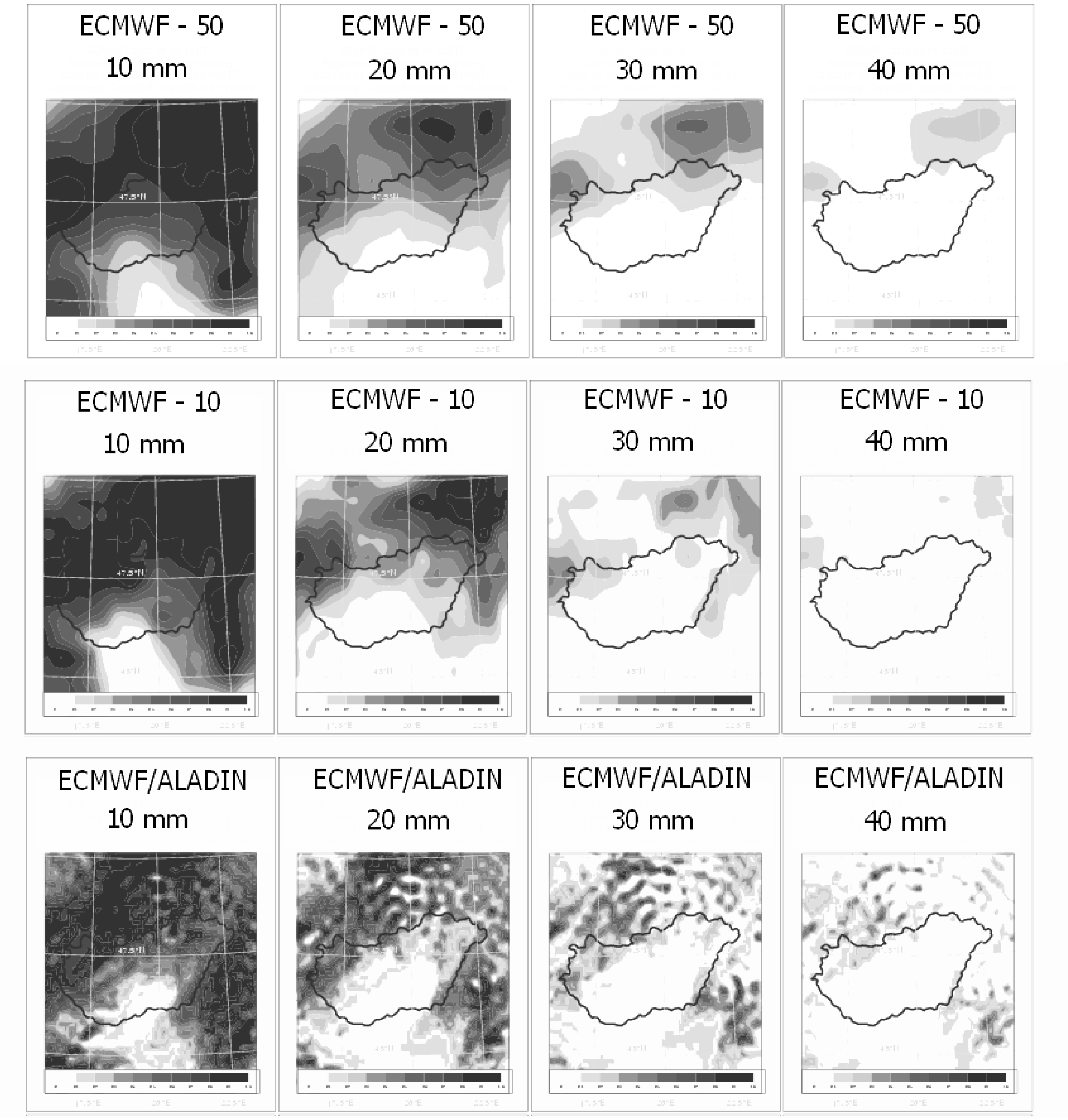


Fig. 4. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24h precipitation exceeding 10, 20, 30 and 40 mm between May 18, 2005, 06:00 UTC – May 19, 2005,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

The conclusions drawn from the stamp diagrams are confirmed by the probability maps: 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 member 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 (the choice having 10 representative members instead of the full system is a reasonable one in that context). 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 (the observed 24h accumulated precipitation did not exceed 15 mm in Hungary) in the region (*Fig. 5*).

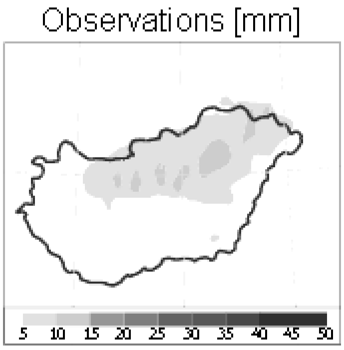


Fig. 5. 24h accumulated precipitation over Hungary between November 16, 2005, 06:00 UTC and November 17, 2005, 06:00 UTC. Values over 5 mm are displayed. Darker colours indicate higher precipitation amounts.

The forecasts studied for this event were initiated at 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).

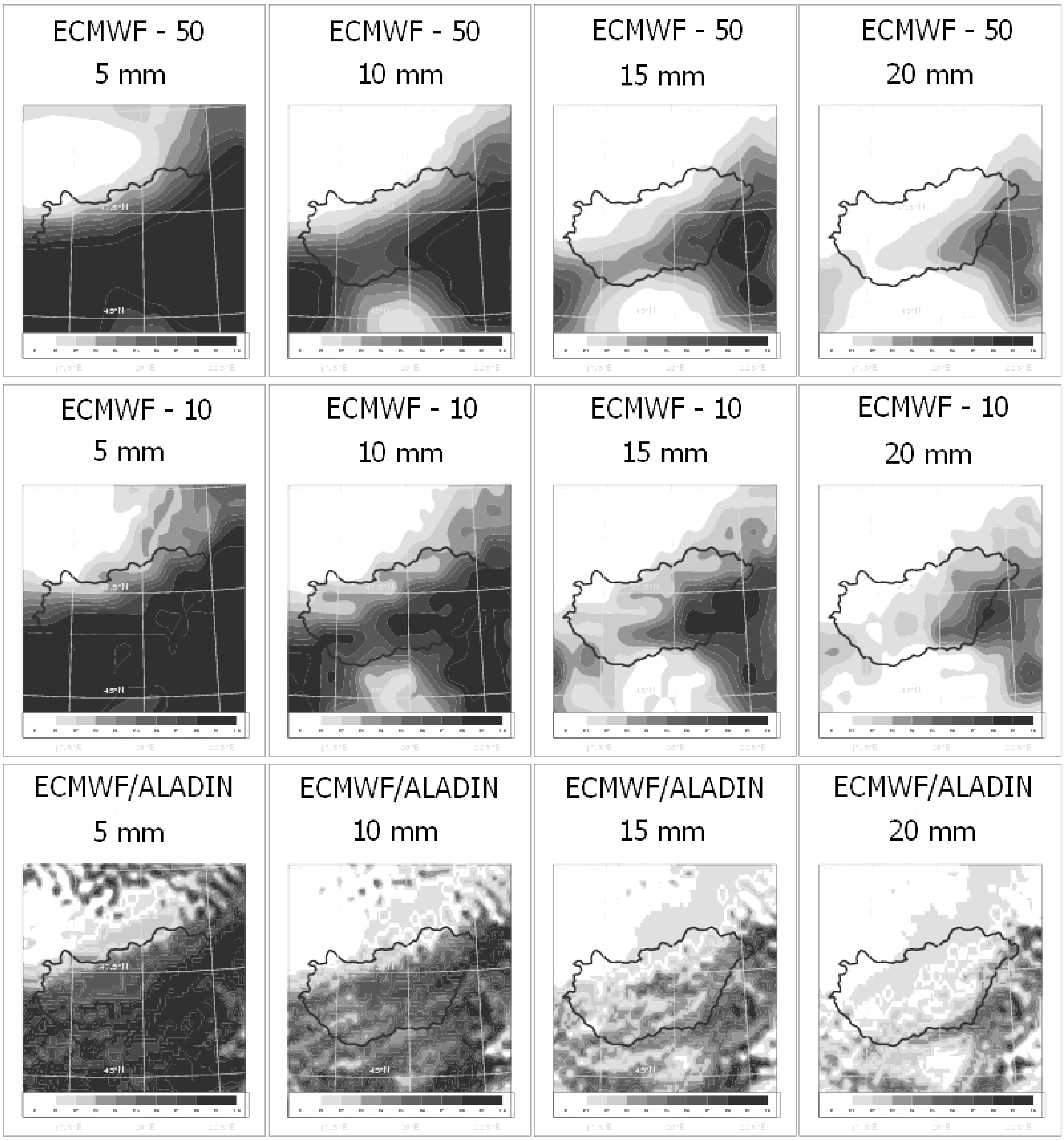


Fig. 6. Probability maps for the original 50 members ECMWF EPS, for the 10 representative members of ECMWF EPS (50 members, bigger domain clustering configuration), and for the 10 members of ECMWF/ALADIN. 24h precipitation exceeding 5, 10, 15 and 20 mm between November 16, 2005, 06:00 UTC and November 17, 2005, 06:00 UTC. 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 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 diagnostics for the efficiency of the initial perturbations used in the ensemble system. Providing all ensemble members are equally probable (i.e. the ensemble members and the verifying observations are mutually independent realizations of the same probability distribution), then 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, the U shape means lack of spread, the L shape means overestimation and the 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 the parameters. The diagram of the ECMWF EPS showed lack of spread at the first day of the forecast range, but the spread was nearly ideal at two and three 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 around day 2 and 3 forecast ranges. The same diagram of the ECMWF/ALADIN downscaling system also showed lack of spread at the first day (even in a bigger extent), however, at day two or three the spread of ECMWF/ALADIN also got a better shape. It is suspected that the results are somehow influenced by the fact that the ECMWF analysis were used (biased towards the ECMWF system), therefore alternative solutions ought to be searched having independent observations or analysis (for instance, analysis of the French global model ARPEGE), while creating the Talagrand diagrams.

Precipitation observations were also used for the calculation of Talagrand diagrams. Data of the high-density precipitation-observing network of the HMS was applied, which means more than 500 stations covering the entire territory of Hungary (*Ghelly*, 2002). Observed precipitation was cumulated from 06:00 UTC in the morning to 06:00 UTC in the next day. The 24-hour accumulated precipitation observations were averaged on 25 km boxes on Gaussian grid. This means that there were 179 observation boxes for each day, which means 537 observation boxes for a given time step when investigating the three cases with underestimation. 0,1°×0,1° post-processing resolution was used both for ECMWF EPS and ECMWF/ALADIN, so four forecast grid points had to be averaged to be consistent to the observations. The Talagrand diagrams of the ECMWF EPS (100 members), the ECMWF EPS

(10 representative members) and the 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 42h and 66h 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 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 member ECMWF EPS and the 10 representative member 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. Only with an ideal clustering method – which cannot be applied in practice – could it be possible to attain the spread of a 100 member system with a 10 member system. The ALADIN runs practically did not affect the spread of the 10 member global EPS, however, in terms of percentage of outliers a slight improvement (decrease) can be noticed. At the second day the 100 member 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 member global EPS shows similar structure, however, the percentage of outliers is significantly higher than it is the case for 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 behaviour like ECMWF 10 member system.

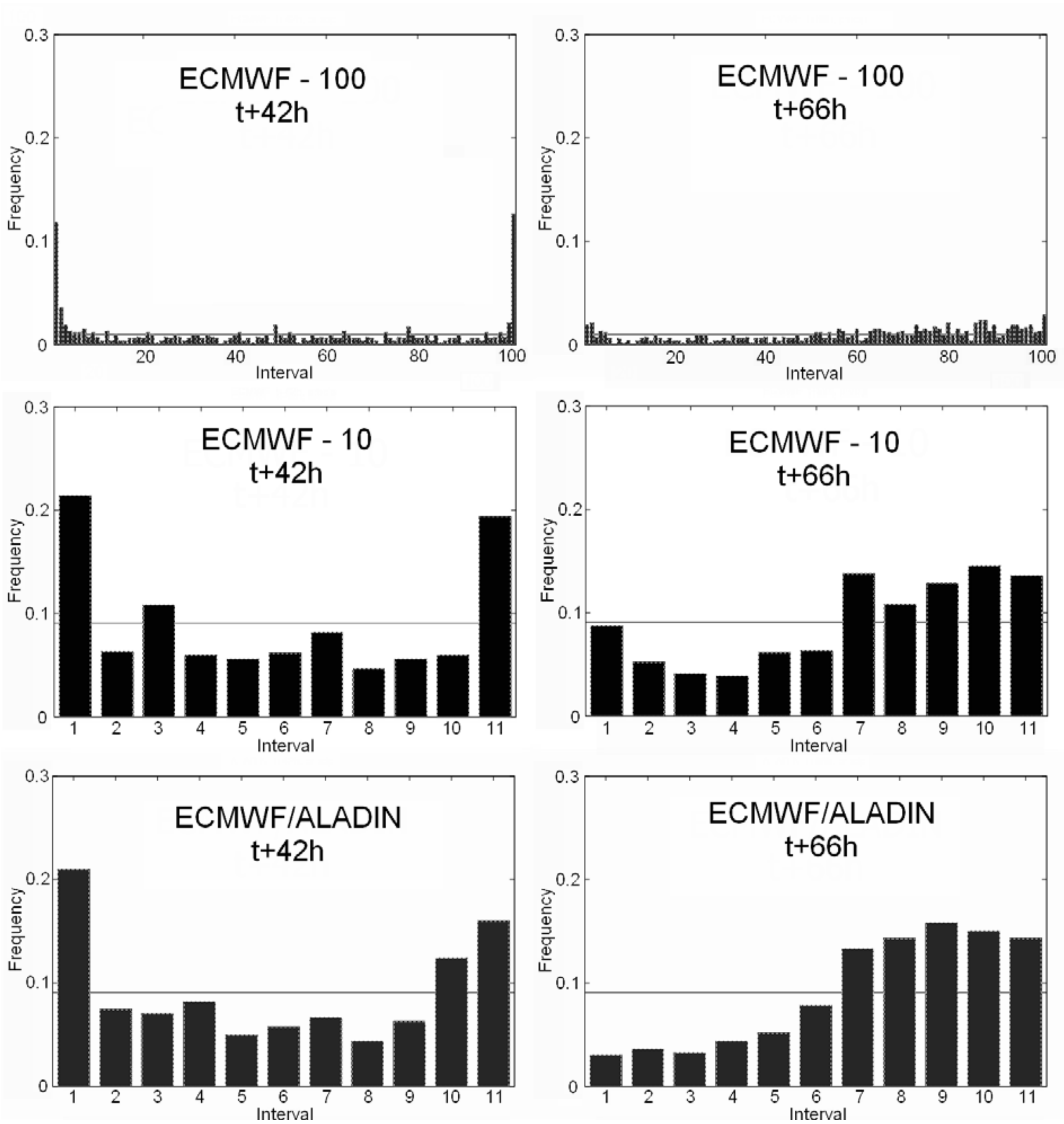


Fig. 7. Talagrand diagrams for ECMWF EPS (100 members), ECMWF EPS (10 representative members) and ECMWF/ALADIN (10 representative members from 100 global members, bigger domain configuration) for +42h and +66h forecasts (24h accumulated precipitation) calculated from the first three case studies.

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 being outside the two extreme values of the Talagrand diagram.

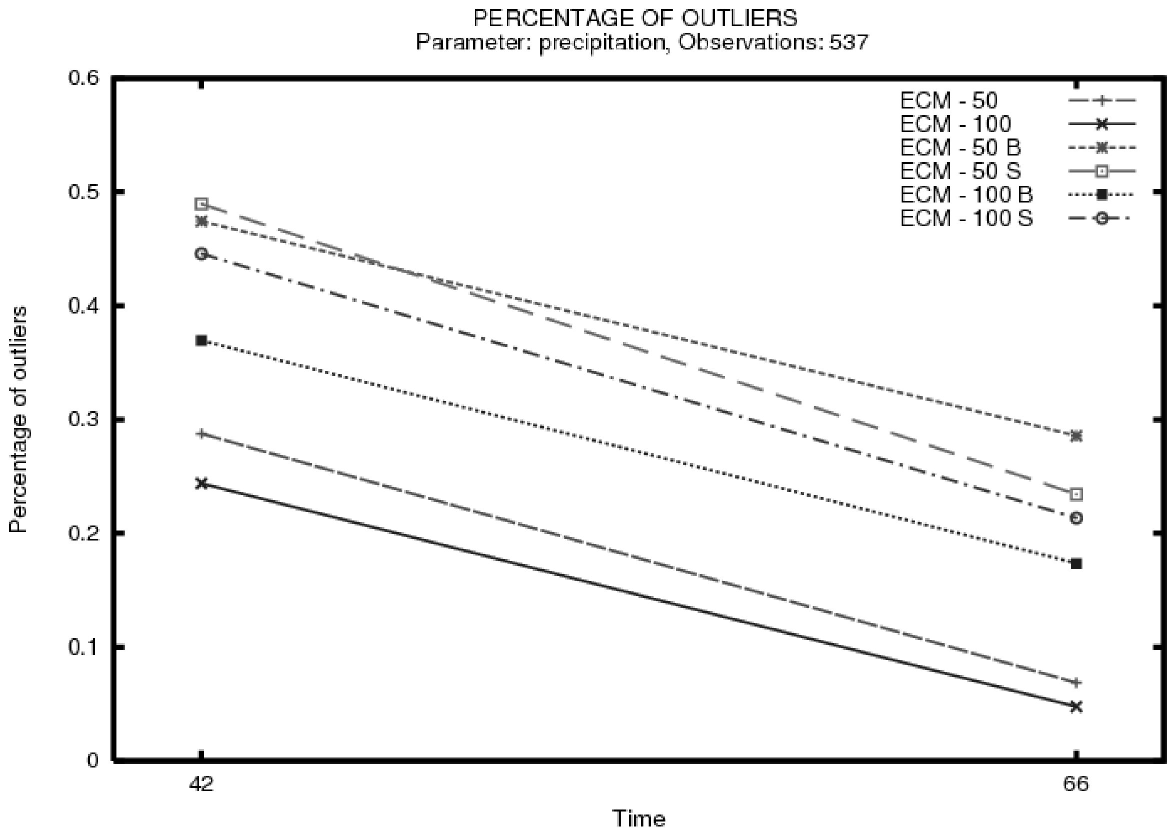


Fig. 8. Talagrand outliers for ECMWF EPS and ECMWF/ALADIN calculated from the first three case studies. The different configurations are: ECMWF EPS 50 members (ECM – 50), ECMWF EPS 100 members (ECM – 100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain (ALD – 50 B), ECMWF/ALADIN using 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 from that point of view 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 member original one (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 the 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 the 50 members the smaller domain is the better. Consequently, further studies are needed to determine what is the relationship between these two clustering parameters (the size of the original ensemble and the clustering domain) and which is 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 at 42 hours 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 as statistically significant ones. The sample for the bigger precipitation thresholds was also relatively large, because nearly at 40% of the grid points were more than 20 mm forecasted. Comparing the ROC scores at 66h for the ECMWF EPS and the ECMWF/ALADIN (*Fig. 9*), generally speaking 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). This altogether 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*).

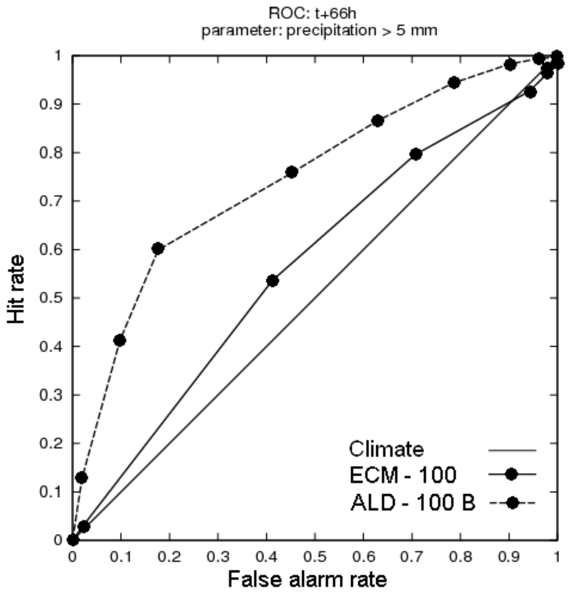


Fig. 9. ROC diagrams for 5mm/24h threshold for +66h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to

ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain), the diagonal line represents the climate. The black dots represent the different probability thresholds. ROC areas: 0.6 (ECMWF EPS); 0.77 (ECMWF/ALADIN).

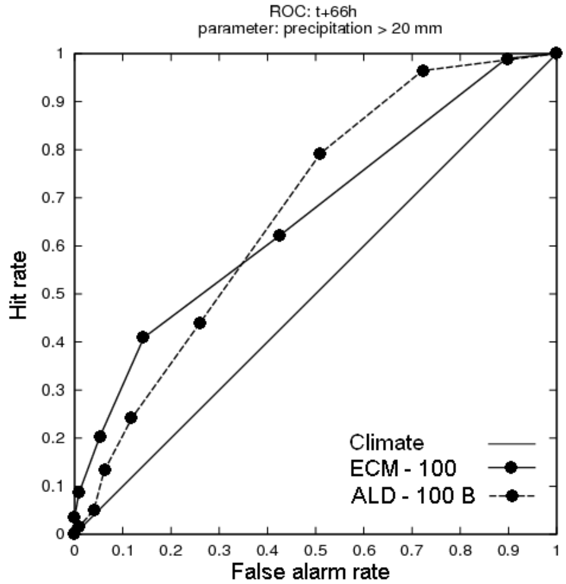


Fig. 10. ROC diagrams for 20mm/24h threshold for +66h forecasts calculated from the first three case studies. Thick solid line is relative to ECMWF EPS (100 members), dashed line is relative to ECMWF/ALADIN (10 representative members from the 100 global members, bigger domain), the diagonal line represents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

For the higher thresholds (20 and 30 mm) the situation is less clear (*Fig. 10*): according to the overall impression the behaviour 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 the higher probability thresholds), but the hit rates of the ECMWF/ALADIN were significantly higher (especially at the lower probability thresholds). This results 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, generally speaking the ECMWF/ALADIN system has a better skill measured by ROC diagrams for precipitation forecasts as it is the case for 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 for 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 characterised 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 the lower false alarm rates, and the higher hit rates (however the ALADIN system was also far from being perfect for that case even though outperforming the ECMWF EPS system).

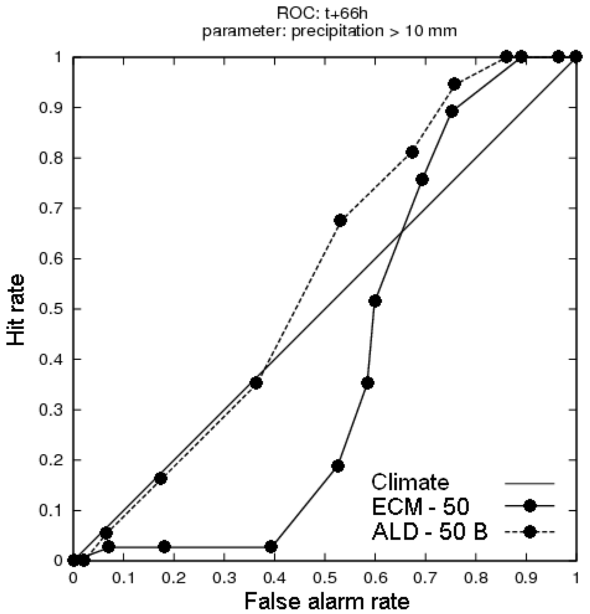
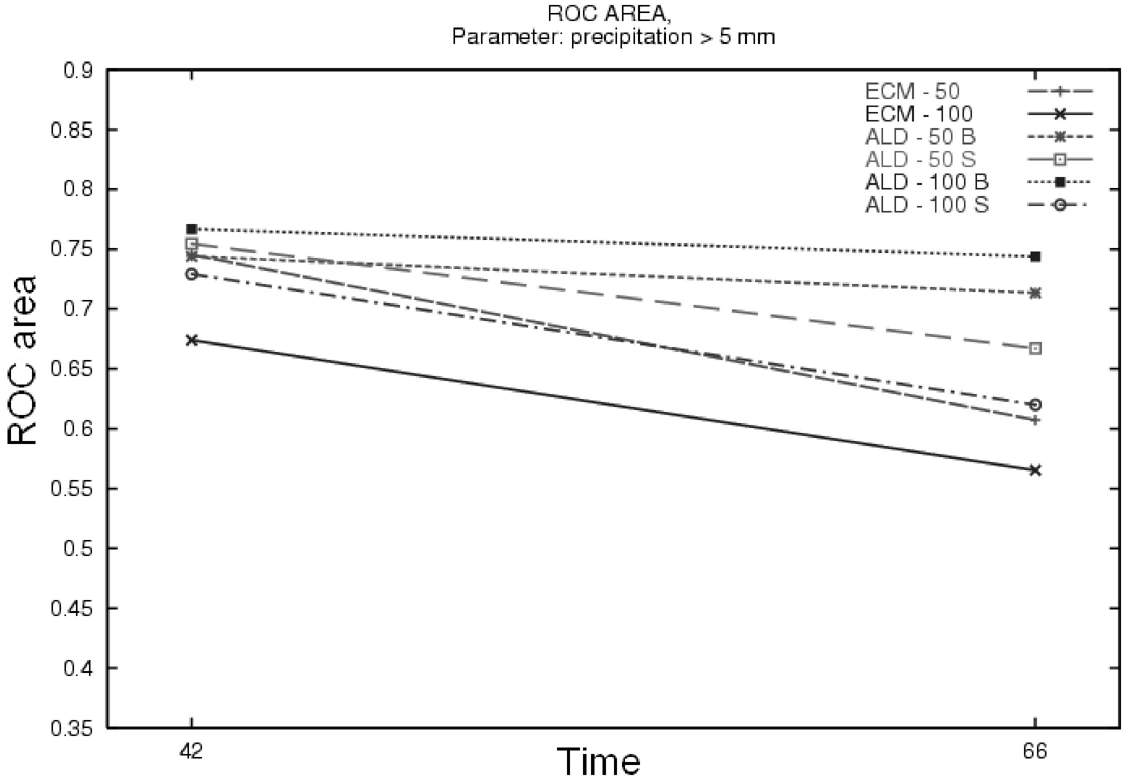


*Fig. 11***.** ROC diagrams for 10mm/24h threshold for +66h forecasts calculated from the fourth case study. Thick solid line is relative to ECMWF EPS (50 members), dashed line is relative to ECMWF/ALADIN (10 representative members from 50 global members, bigger domain), the diagonal line represents the climate. The black dots represent the different probability thresholds. ROC areas: 0.4 (ECMWF EPS); 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 the 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 5mm/24h and 20mm/24h thresholds.



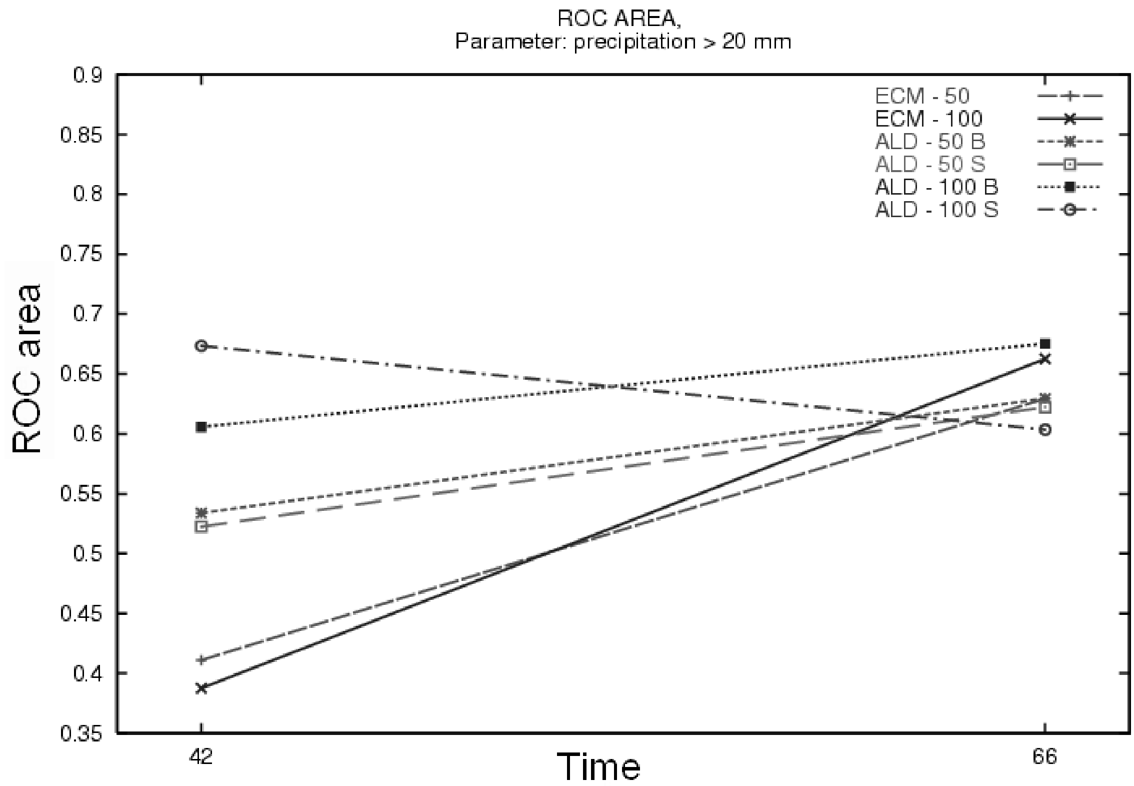


Fig. 12. The time evolution of the ROC area for the 5mm/24h (upper diagram) and 20mm/24h threshold (lower diagram). The different configurations are: ECMWF EPS 50 members (ECM – 50), ECMWF EPS 100 members (ECM – 100), ECMWF/ALADIN using 50 EPS members and bigger clustering domain (ALD – 50 B), ECMWF/ALADIN 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 the ECMWF one), 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 +42h forecast time, however one must remember that these verifications are based on three cases, and the large precipitation fall on the second day of the cases. Consequently, the sample for the +42h 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 the 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 that the spread of the ensemble is not sufficient on the first day of the forecast. The lack of spread results the 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 one of the ECMWF/ALADIN configuration. 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 ones than that of the ECMWF, which means that the spread does not look like 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 into the 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 to try 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 proved that the system is capable to correct events corresponding with global precipitation overestimation (fourth case study).

The objective verification of the global and the downscaled systems was performed for different parameters too. In the 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 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. The time evolution of the ROC area showed, that among the four clustering configurations that one performs 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 at 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 in the examined limited number of cases could bring benefit on top of the global ECMWF EPS system. 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 for *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 Newsle*tter 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. Volume 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 pages*

Ghelly, A., 2002: Verification of precipitation forecasts using data from high-resolution observation *ECMWF Newsletter* 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: *Proceedings 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.R. Meteorol. 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 Processes in Geophysics* 12*,* 527-536

Mason I.B., 1982: On scores for yes/no forecasts. *In Preprints of papers delivered at the 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.R. Meteorol. 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* 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 Repor*t 8, World Meteorological Organization, Geneva

Szintai, B., 2004: Clustering of the medium-range ensemble forecasts of the ECMWF. *Undergraduate research (in Hungarian), Budapest, Eötvös Loránd University*

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