Limited-Area Ensemble Forecasting: the COSMO-LEPS system

Stefano Tibaldi, Tiziana Paccagnella, Chiara Marsigli, Andrea Montani and Fabrizio Nerozzi
ARPA-SIM, Bologna, Italy

1. Introduction

The improvement of Quantitative Precipitation Forecasting (QPF) is still one of the major challenges in numerical weather prediction (NWP). Despite the constant increase of computer power resources, which has allowed the development of more and more sophisticated and resolved NWP models, accurate forecasts of extreme weather conditions, especially when related to intense and localised precipitation structures, are still difficult beyond day 2 (Mullen and Buizza, 2001a) and, in rare and selected cases, even at 24 hours. This limitation is due, among other reasons, to the inherently low degree of predictability typical of the relevant physical phenomena. The probabilistic approach has been recently more and more explored to try to come to terms with the chaotic behaviour of the atmosphere and to help forecasting phenomena with low deterministic predictability.

In addition to this, almost twenty years ago Hank Tennekes, at the time member of the ECMWF (European Centre for Medium-Range Weather Forecasts) Scientific Advisory Committee, raised the question of the opportunity of producing a-priori estimates of forecast skill stating that "no forecast is complete without a forecast of the forecast skill". It is not an overstatement to say that his bold assertion contributed greatly to the development, at least at ECMWF, of forecast skill studies, estimates and prediction techniques (e.g. Palmer and Tibaldi, 1988) and to the related development of statistical-dynamical prediction methods like ensemble forecasting.

In fact, global ensemble prediction systems, implemented operationally for several years now by some of the major meteorological centres (Tracton and Kalnay, 1993; Molteni et al., 1996; Houtekamer et al., 1996), have become extremely important tools to tackle the problem of predictions beyond day 3-to-4 and are becoming more and more the bread-and-butter of operational forecasters all around the world. In order, however, to fully exploit the potential of the stochastic-dynamic approach, that is to predict within a certain detail the evolution of the probability density function (PDF) of the meteorological system (or at least of some important observable variable) in the atmosphere's phase space, the population of the ensemble should be of the same order of magnitude of the dimensionality of the unstable subspace of the phase space itself (Montani, 1998). This implies a number of integrations probably much higher than what can be achieved by present-day operational ensemble populations (order 50 at most). Furthermore, probabilistic global ensemble systems are usually run at a coarser resolution with respect to (single) deterministic global predictions. Hence, ensemble skill in forecasting intense and localised events in the short- and medium-range is currently still limited. In order to enhance the present-day prediction capabilities of operational ensemble systems, several approaches have been attempted.

As far as global models are concerned, the horizontal resolution of the ECMWF Ensemble Prediction System (EPS) was increased at the end of 2000 (Buizza et al., 1999) and is now based on a TL255L40 model (spectral model with triangular truncation at wavenumber 255 and 40 vertical levels) approximately equivalent to a grid spacing of 80 km (Buizza et al., 2005). During 1999, ECMWF also developed a Targeted Ensemble Prediction System (TEPS, Hersbach et al., 2000), where the perturbations applied to the analysis to obtain the different initial conditions were selected in order to maximise the 48-hour total energy perturbation growth over the European Area (about 35N-75N, 40W-30E), instead of over the whole extratropical Northern and Southern Hemisphere as it is the case in the operational EPS. This TEPS system was based on a T159L40 model (the same EPS configuration in operation at that time) and was developed within an ECMWF special project as a collaboration amongst KNMI (Royal Meteorological Institute of the Netherlands), ARPA-SIM (the Hydro-Meteorological Service of Emilia-Romagna Region of Italy, formerly ARPA-SMR) and the Norwegian Meteorological Institute (DNMI). The project aim was to increase the ensemble spread over the European Area in the short-range and early medium-range. Results concluded that TEPS performs marginally better than EPS in the 72-to-96 hour forecast range in terms of probabilistic prediction of severe events over Europe while, for shorter ranges, the two systems have comparable skill (Hersbach et al., 2000).

As an alternative approach (trading ensemble size for model horizontal resolution), a smaller-size global ensemble, but at a higher resolution, was tested by Molteni et al. (2001). Although the system turned out to be very expensive from a computational point of view, it also gave promising results in the probabilistic prediction of heavy precipitation.

In recent years, much attention has been devoted to the development of a multi-model multi-analysis ensemble system (MMAE) based on the UK Met Office and ECMWF ensembles. It was found that, over a large number of cases, MMAE almost invariably performs at least as well as the best individual ensemble and, sometimes, better than either of the two single-model ensemble systems (Mylne et al., 2002).

Regarding short-range global ensembles, Meteo-France has recently developed an 11-member global ensemble system referred to as PEACE (Prévision d’Ensemble A Courte Échéance), where the initial perturbations are generated using the targeted singular vector technique. This system, operational since June 2004, is run up to 60 hours. A systematic evaluation of the system is currently being performed. In addition to this, the UK Met Office is experimenting with a global ensemble system for short-range predictions based on perturbations of initial conditions, model parameters and physics. The main purpose of this ensemble is to provide boundary conditions to a regional ensemble. First results show positive impact in the prediction of surface parameters (Arribas, 2004).

Turning to Limited-Area Model (LAM) applications of the Ensemble technique, different approaches are being explored, taking also into account the constraints imposed by lateral boundary conditions on inner domain growing perturbations. One of the possible approaches is the use of the LAM to perform a dynamical downscaling of the different global ensemble forecasts produced by the driving global EPS. Other methodologies rely on inducing perturbations during the limited-area model run employing different techniques based on changing physical schemes, perturbing model parameters or applying stochastic perturbations on physical tendencies.

As for operational implementations, NCEP (the US National Center for Environmental Prediction) developed the first operational multi-initial-condition multi-model Short-Range Ensemble Forecasting system (SREF, Du et al., 2003). This system is composed of 15 elements, 5 members from the Regional Spectral Model and 10 members from the Eta model; members from the same model are differentiated by changing the convection scheme. Perturbation on initial conditions are based on “regional” breeding cycles. SREF has been operationally running since May 2001; the system is now run for 63 hours twice a day (at 9 and 21 UTC) at the horizontal resolution of 32 km.

At the DNMI, a limited-area ensemble (LAMEPS) was generated by nesting the operational limited-area model (HIRLAM) in each element of a 21-member TEPS set (Frogner and Iversen, 2001 and 2002). This approach (sometimes referred to as the “brute-force” approach, BFA) appears to provide better results than global TEPS for the prediction of heavy rainfall events, but the computational burden make it difficult to afford on an operational basis. The semi-operational version of the system is called NORLAMEPS, where TEPS and LAMEPS are combined to obtain a 41 member ensemble (Haakenstad and Frogner, 2004).

The UK Met Office is currently planning to implement a limited-area ensemble system of about 16 members at 24 km horizontal resolution over a European domain, with initial and boundary conditions provided by the global ensemble, while Meteo France is testing a LAMEPS system based on coupling PEACE with the regional model ALADIN (10 km of horizontal resolution, 11 members, 48 hours forecast range).

In the framework of the SRNWP (Short Range Numerical Weather Prediction project of the EUMETNET network) cooperation, A "poor-man" ensemble (PEPS) has been implemented at DWD by collecting and combining many (19 at present) operational deterministic limited area models (www.dwd.de/PEPS). The system has being operationally running since December 2004.

A few years ago, ARPA-SIM proposed the LEPS methodology (Limited-area Ensemble Prediction System, Molteni et al., 2001; Marsigli et al., 2001; Montani et al.. 2001 and 2003a), which will be described in more detail in Section 2. The LEPS methodology attempts to combine the benefits of the probabilistic approach with the high-resolution capabilities of the LAM integrations, limiting the computational investment. The method is based on an algorithm to select a number of members out of a global ensemble system. The selected ensemble members (called Representative Members, RMs) provide initial and boundary conditions to integrate a limited-area model. The ensemble size reduction is necessary to keep the overall computational load operationally affordable. The transfer of information from the large-scale to the mesoscale can be viewed as a dynamical downscaling of the forecast provided by the global-model probabilistic system. The good results of the early experimentation phase of the LEPS system led the limited-area modelling consortium COSMO (COnsortium for Small-scale Modelling; www.cosmo-model.org) to the decision of implementing the LEPS technique within the COSMO framework on a regular-basis, giving rise to COSMO-LEPS (Montani et al., 2003b) as further described in Section 2.

In Section 3, some case studies produced using LEPS are analysed in order to investigate in some more detail the behaviour of the system, while Section 4 is devoted to an objective and more systematic evaluation of the COSMO-LEPS system performance. Some conclusions and future plans are outlined in Section 5.

2. The COSMO-LEPS system

As already mentioned in the introduction, the LEPS methodology is based on the idea of reducing the number of LAM integrations needed by an order of magnitude by retaining a hopefully large amount of the global ensemble information while decreasing the number of ensemble elements subjected to LAM runs. This is achieved by, first, grouping the global ensemble members into a number of clusters and then choosing a "Representative Member" (RM) within each cluster. Each RM is considered to be representative of the possible evolution scenario associated to each particular cluster and provides both initial and boundary conditions for a high-resolution LAM integration.

During the entire preliminary phase of the experimentation of LEPS, the LAM used has been LAMBO (Limited-Area Model BOlogna), the limited-area model running operationally at the time at ARPA-SMR, a hydrostatic LAM based on an early version of the NCEP ETA model (Janjic, 1990; Mesinger et al., 1988). The system was then constituted by 5 LAMBO integrations driven by 5 RMs selected from 5 clusters derived from the operational ECMWF EPS set of global forecasts. LAMBO was integrated at a horizontal resolution of approximately 20 km and with 32 vertical levels (domain: 1-25E, 36-50N). Montani et al. (2001, 2003a) and Marsigli et al. (2001, 2004) have shown that, over a number of test cases (including episodes in or close to mountain areas) and for all forecast ranges (from 48 to 120 hours), the LEPS system performs better than the ECMWF EPS (or the Targeted EPS) in probabilistic quantitative prediction of heavy precipitation events, both in terms of a better geographical localisation of the regions most likely to be affected by the events and of a more realistic intensity of the associated rainfall patterns.

Following the encouraging results of the early experimental phase, the regular (daily) generation of a pre-operational limited-area ensemble prediction system, the COSMO-LEPS system, was started in November 2002 on the ECMWF computing system (Montani et al., 2003b) under the auspices of the COSMO Consortium, which involves the operational meteorological institutions of Germany, Italy, Switzerland, Greece and Poland cooperating on the development of the limited-area non-hydrostatic model Lokal Modell (LM). COSMO-LEPS has been designed from the outset for the "short-to-medium-range" timescale (48-120 hours).

The first step of the procedure is the application of a Cluster Analysis procedure to the merge of the two most recent ECMWF operational global ensembles. The ECMWF Ensemble Prediction System (EPS) is now based on a T_L255L40 model (global spectral model with triangular truncation at wavenumber 255 and 40 vertical levels), corresponding to a horizontal resolution on a linear lat-lon grid of about 80 km, and is formed by 51 members (Buizza et al., 2005).

Two successive 12-hour-lagged EPS ensembles (started at 00 and 12 UTC) are therefore grouped together so as to generate a 102-member so called "super-ensemble". The introduction of this super-ensemble technique was aimed at increasing the spread of the starting global ensemble, thereby improving the spanning of the phase space (Montani et al., 2003a).

A multivariate hierarchical cluster analysis is then performed on the resulting 102 members, so as to group all elements into 10 clusters of different populations; the clustering algorithm being based on the Complete Linkage method (Wilks, 1995) with the number of clusters fixed to ten. The clustering variables are the geopotential height, the two components of the horizontal wind and the specific humidity at three pressure levels (500, 700, 850 hPa) and at two forecast times (fc+96h and fc+120h for the "youngest" EPS, started at 12 UTC). The cluster domain covers the region 30N-60N, 10W-40E (see Fig.1).

The one just described here is the configuration running since June 2004. From November 2002 to June 2004, COSMO-LEPS was based on five (instead of ten) RMs selected clustering three (instead of two) 12-hour lagged consecutive EPS (also the EPS starting at 12 UTC of the previous day was used). The reasons for such changes were based on the results of an objective verification exercise, reported in some detail in Section 3.

Within each cluster, one Representative Member (RM) is selected according to the following criteria: the RM is that element closest to the members of its own clusters and most distant from the members of the other clusters; distances are calculated using the same variables and the same metric as in the cluster analysis; in this way, 10 RMs are selected, one for each cluster. Each RM provides initial and boundary conditions for the integrations of LM, which is run 10 times for 120 hours, always starting at 12UTC. LM is run with a horizontal resolution of about 10 km and with 32 levels in the vertical.

Assuming a relationship between cluster population and the probability of occurrence of its associated RM, probability maps are generated by assigning to each LM integration a weight proportional to the population of the cluster from which the RM was selected.

A number of deterministic and probabilistic products are routinely produced and disseminated to the COSMO community for regular pre-operational evaluation. The deterministic products for each of the 10 LM runs are: Precipitation, Mean Sea level pressure, 700 hPa Geopotential, 850 hPa Temperature. The main probabilistic products include: probability of 24h cumulated rainfall exceeding 20, 50, 100, 150 mm, probability of daily Tmax exceeding 20, 30, 35, 40 C, probability of daily Tmin below -10, -5, 0, +5 C, probability of daily Vmax exceeding 10, 15, 20, 25 m/s, probability of 24h cumulated snowfall exceeding 1, 5, 10, 20 mm of equivalent water.

COSMO-LEPS product dissemination started in November 2002 and, at the time of writing the system is being tested to assess its usefulness in met-ops rooms, particularly in terms of the assistance given to forecasters in cases of extreme events. In addition to COSMO Partners, COSMO-LEPS products are made available also to other interested ECMWF Member States.

3. Results from selected case studies

As examples of COSMO-LEPS' capability to predict, already at the short and early-medium range, the occurrence of local severe weather, the behaviour of the system is presented for two heavy precipitation events recently occurred over the alpine area. For each of the two case studies, the experimentation included:

- re-running of the global ECMWF-EPS (hor. res.: 80 km, 40 vertical levels; forecast range: 120 hours), archiving model output every 3 hours (1 EPS per case study; no super-ensemble technique applied);
- nesting LM (hor. res. 10 km; 35 vertical levels; forecast range: 120 hours) in each EPS member (brute-force approach), in a configuration identical to the operational set-up;
- evaluate different ensemble-size reduction-techniques, the clustering variables being exactly the same as in the operational suite (Z, U, V, Q at 500, 700, 850 hPa, at the forecast ranges fc+96h and fc +120h);
- assess the quality of 5-member, 10-member, ..., 51-member COSMO-LEPS.

3.1 Description of the case studies

In this section, we report a short description of the two case studies (more details can be found in Montani et al., 2004).

1. Friuli event: a mesoscale convective systems developed due to the intrusion of cold air masses travelling from west to east. Precipitation maxima exceeding 300 mm/day were recorded over Friuli (north-east Italy) between 28 and 29 August 2003, secondary rainfall peaks being observed also over the Ticino area.
2. Piedmont event: an upper-level trough over France caused moist south-westerly flow to blow over the Alps for several days, this leading to widespread heavy precipitation over North-western Italy as well as over Switzerland; rainfall peaks above 150 mm/day were observed between 15 and 16 November 2002.

The two panels of Fig. 2 show the observed precipitation cumulated over a 24-hour period starting at 12 UTC, for each case study. It is worth pointing out that, while the former case refers to a localised weather event (mesoscale convective system), the latter case deals with a heavy precipitation event for which the large-scale forcing plays a major role (mid-tropospheric trough). Therefore, the analysed case studies span two different types of flood events, on which the performance of COSMO-LEPS system can be assessed.

In the following subsections the behaviour of COSMO-LEPS at different prediction ranges (2.2) is analysed, as well as the impact of ensemble-size on the quality of the probabilistic forecast (2.3).

3.2 Behaviour of the system at the different forecast ranges

The behaviour of the COSMO-LEPS system for the Piedmont case (right panel of Fig. 2) is investigated considering the probabilistic forecast of the events “24-hour rainfall exceeding 20 mm” and “24-hour rainfall exceeding 50 mm” for three different forecast ranges (2, 3 and 4 days).
Fig. 3 shows the consistency of the rainfall probability maps for these events (top-row panels for the 20 mm event; bottom-row panels for the 50 mm event). All maps verify at 12UTC of 16 November 2002.- It can be noticed that, already at the 96-hour range (top-left and bottom-left panels, respectively), high probability of occurrence is predicted over Northern Italy, in the areas where heavy precipitation was actually observed. This kind of information would enable a hypothetical bench forecaster to issue a flood warning, to be either confirmed or dismissed on the basis of forecasts at ranges closer to the predicted event. Thanks to the relatively long lead-time (4 days), preventive actions could have been taken so as to limit damages as much as possible.

The possibility of a heavy rainfall scenario is confirmed and reinforced by the 72 and 48-hour forecasts (mid-column and right-column panels, respectively), the probability of occurrence exceeding 90% in the correct locations.

Similar considerations apply also for the Friuli case, where the performance of COSMO-LEPS is again accurate and consistent (Fig. 4). Despite this case study being characterised by a local-forcing and the predictability being more limited than in the previous one, the probability maps indicate, already 96 hours before the event, the possibility of a heavy rainfall scenario (24-hour cumulated rainfall above the 50 mm threshold) over north-eastern Italy, as evident in the bottom-left panel of Fig. 4. This critical situation is confirmed by the 3-day and 2-day probabilistic forecasts, where also the Ticino (Central Alps) area is correctly highlighted as a region possibly affected by heavy precipitation.

3.3 Impact of the ensemble reduction and comparison with the ECMWF EPS

As described in Marsigli et al. (2001) and Molteni et al. (2001), the main motivations behind the ensemble-size reduction of LEPS methodology are dictated by computing power limitations as well as by the (supposed) possibility to capture, thanks to the clustering-selection technique, alternative weather patterns, with higher or lower probability of occurrence. This latter aspect is critical and deserves further investigation. In fact, it is not fully clear the extent to which the ensemble-size reduction may induce a loss of information about the spectrum of possible atmospheric flows and thus affects the overall performance of LEPS.

In order to assess the impact of the ensemble size reduction on the rainfall probabilistic prediction, the cluster analysis has been performed several times with the number of clusters increasing from 5 to 51 generating limited-area ensembles, with a correspondent increasing population from 5 to 51.
The attention is focussed on the 4-day predictability of the event “24-hour precipitation exceeding 50 mm” and the performances of the 5-member, 10-member and 51-member ensembles are assessed both for the global EPS and for the limited-area LEPS (the system configurations will be referred to as EPS5, EPS10, EPS51, LEPS5, LEPS10 and LEPS51, respectively).

Fig. 5 reports the results obtained for the Piedmont case: it is clear that EPS forecasts (top-row panels) underestimate the rainfall intensity, since the probability of occurrence is, in all configurations, well below 30% and, additionally, poorly localised. It can also be noticed that, only when the EPS51 configuration is considered (top-right panel), three different areas over Northern Italy are highlighted as possible locations of heavy precipitation (as it actually happened); nevertheless, the probability values are very low and poor guidance would have been given to a forecaster involved in alert procedures.

The situation is completely different when the better description of orographic and mesoscale-related processes comes into play in the high-resolution forecasts of LEPS system. The bottom-left panel of Fig. 5 indicates that already the LEPS5 system is able to highlight the possibility of occurrence of a heavy precipitation event in the correct locations. As the population of the LEPS system increases (bottom-middle and bottom-right panels), the results obtained are very similar, the probability of occurrence being only slightly modified.

As concerns the Friuli case, the top-row panels of Fig. 6 show that EPS rainfall probability maps (for precipitation exceeding 50 mm/day) fail completely to indicate North-eastern Italy as an area possibly affected by heavy rainfall. In the EPS51 configuration a weak signal is evident over the Ticino region, where heavy precipitation was actually observed; but also in this case, the probability is below 30%. On the other hand, the information provided by LEPS forecasts is much more accurate than that to be found in the global runs. The bottom-row panels show that both the Ticino and the Friuli area are highlighted as regions most likely to be hit by the severe weather event and the pattern of the probability maps is roughly unchanged as the LEPS population increases from 5 to 51. Regarding high probability values over the Liguria region, it is worth noting that this false alarm is not due in particular to the LEPS ensemble system but rather to a generalised systematic error of the model in this region.

From the above results, the impact of the ensemble-size reduction on the forecast accuracy does not seem to be crucial, since, for both cases, the heavy rainfall scenario is properly captured in the “LEPS5” as well as in the “LEPS51”"configuration. This result is however very encouraging: at least for such two particular case studies, the clustering-selection technique highlights the most important evolution scenarios and enables the generation of reliable and accurate probability maps. It is clear that the results obtained by such study may do not have any statistical significance, since they are based on only two cases, both characterised by heavy precipitation events. Nevertheless, the outcome of these experiments is important, since it can indicate the potential of COSMO-LEPS methodology, so justifying future investigation and system development.

4. Statistical evaluation

In this section, a more systematic and objective evaluation of the COSMO-LEPS forecast performance is presented, so as to assess overall abilities and shortcomings of this prediction system. The attention remains focussed on the probabilistic prediction of heavy precipitation, one of the main causes of damages and life loss over Europe.

The local effects of precipitation are related to two main characteristics: the cumulative volume of water deployed over a specific region and the rainfall peaks which occur within this region. The relative importance of these two features of the intense event is related to the geo-morphological features of the area affected. Both these aspects need to be taken into account when the verification of precipitation is concerned. This is accomplished by considering respectively the mean precipitation over an area as well as the precipitation maxima in the same area.

For this new probabilistic system to be shown as potentially valuable, it has to be compared with the state-of-the-art of ensemble system, namely ECMWF EPS, which is already available to

European NWSs at no extra cost. Therefore, it is investigated the extent to which the high-resolution forecast details and intensities provided by the COSMO-LEPS system produce real added value with respect to the global EPS.

COSMO-LEPS forecasts are here verified on the basis of the traditional probabilistic scores: Brier Skill Score (Wilks, 1995), ROC area (Mason and Graham, 1999), Cost-loss Curve (Richardson, 2000), Percentage of Outliers (Buizza, 1997). Results are shown for the Autumn 2003 period (September, October and November).

From the results shown in Tibaldi et al. (2003), no positive impact of the weighting procedure was detected regarding probabilistic forecast of precipitation; therefore, the results presented in this Section are referred to probabilities computed giving the same weight to each of the reduced-ensemble members.

This analysis is relative to the COSMO-LEPS configuration which has been running pre-operationally from November 2002 to June 2004: 5 RMs were selected clustering three 12-hour lagged consecutive EPS (FORMER SUITE). Since June 2004 the suite was changed (10 RMs selected from a super-ensemble of two EPSs, CURRENT SUITE) according to the results shown in the following subsections.

Verification is subdivided into two main streams: a verification of the methodology on which COSMO-LEPS is based (subsection 4.1) and a verification of the system performance against other available probabilistic systems (subsection 4.2).

4.1 Analysis of the methodology

The super-ensemble approach, grouping together 3 consecutive EPSs, has been adopted to enlarge the size of the ensemble on which the RM selection algorithm is applied. This allows to increase the ensemble spread and to explore better the phase space, even if a price in terms of skill is paid: the older the EPS, the less skilful its members are likely to be. In order to evaluate the comparative effects of the increased spread and of the decreased skill, the Representative Members chosen with the FORMER-SUITE methodology are compared to those chosen using only one or two EPS. Furthermore, the impact of the reduced-ensemble size is also evaluated by comparing each 5-RMs configuration with the correspondent 10-RMs configuration. The six examined configurations are listed in Table 1.

Configuration name	EPS on which the cluster analysis is based	Number of RMs selected	Suite name
3eps-5rm	3 12-hour lagged EPS	5 RMs	FORMER SUITE
2eps-5rm	2 12-hour lagged EPS	5 RMs	
1eps-5rm	1 12-hour lagged EPS	5 RMs	
3eps-10rm	3 12-hour lagged EPS	10 RMs	
2eps-10rm	2 12-hour lagged EPS	10 RMs	CURRENT SUITE
1eps-10rm	1 12-hour lagged EPS	10 RMs	

Table 1: Characteristics of the analysed COSMO-LEPS configurations.

This analysis is performed in terms of 24-hour precipitation. The forecast values at each grid point are compared with a proxy for the true precipitation occurred chosen as the +24 hour forecast by the ECMWF deterministic model (horizontal resolution 40 km). It is not crucial the extent to which this proxy is a good approximation for the truth, because this is a comparison between different configuration of the same model. The period chosen for this test is September-November 2003 and the verification area is the entire clustering area (the shaded rectangle in Fig.1).

Regarding the 5-RM ensembles, results show that the Brier Skill Score (Fig.7, left panel) is highest when the clustering is based on the most recent EPS only (1eps-5rm, solid line with circles), while it is lowest for the 3eps-5rm (solid line with crosses). The difference between the two is not so large, but it is clearly detectable at every forecast range. The 2eps-5rm (solid line with triangles) has an intermediate skill, closer to the 1eps-5rm.

The percentage of outliers of the systems is also shown. This is the percentage of times the “truth” falls out of the range spanned by the forecast values. The percentage of outliers (Fig.7, right panel) of the 1eps-5rm (solid line with circles) is rather higher than the other two, for every forecast range, while there is almost no difference in terms of outliers between the 2eps-5rm (solid line with triangles) and the 3eps-5rm (solid line with crosses).

These results would seem to suggest that the use of just two EPS in the super-ensemble can be a good compromise, permitting to decrease the percentage of outliers significantly but leading only to a small decrease of the skill.

Regarding the impact of the ensemble size, the difference between each 5-member ensemble and the correspondent 10-member ensemble is remarkable, being about 0.1 in terms of Brier Skill Score for every configuration. This is shown in the left panel of Fig.7, comparing each thick solid line with the thin line carrying the same symbols. The impact of doubling the ensemble size is almost the same for every configuration and is larger than impact of changing the number of EPSs on which the Cluster Analysis is performed (2 or 3).

These results led to two major modifications of the COSMO-LEPS methodology, as applied to the pre-operational suite, at the beginning of June 2004: since this date the super-ensemble has been constructed by using only the 2 most recent EPSs and the number of clusters has been fixed to 10, nesting Lokal Modell on each of the 10 selected RMs.

4.2 Comparison of the limited-area ensemble with global ensembles

In order to quantify the added value brought about by the mesoscale probabilistic system, COSMO-LEPS is compared with ECMWF EPS. The comparison has to take into account two important issues: the difference in the number of ensemble members and the difference in terms of model resolution.

As far as the population of the ensembles is concerned, it is worth pointing out that the verification is carried out during the period September-November 2003, when the pre-operational COSMO-LEPS population was fixed to 5 members. Therefore, COSMO-LEPS is compared not only with the operational EPS (51 members), but also with the reduced EPS ensemble made up by the 5 Representative Members. This permits to quantify the impact of the increased resolution alone.

The problem of the different resolutions of the two systems (10 km for COSMO-LEPS and 80 km for EPS) is tackled by upscaling both systems to an even lower-resolution common interface: the grid-point forecasts of both model are averaged over boxes of 1.5 x 1.5 degrees.

The comparison is performed in terms of 24-hour precipitation against observed data. In order to compare properly forecast values over boxes and observed values on station points, the observations within a box are averaged and the obtained values are compared with the averaged forecast values.
The comparison is carried out over a large fraction of the COSMO-LEPS domain. In fact, a very dense network of stations (about 4000) recording daily precipitation (cumulated from 06 to 06 UTC) are made available by Germany, Switzerland and Italy, see Fig. 8.

The 3 ensemble systems compared are:

- the COSMO-LEPS system, made up of 5 members, 10 km of horizontal resolution, referred to as “cleps”;
- the EPS reduced-ensemble made up by the 5 Representative Members chosen from the super-ensemble, 80 km of horizontal resolution, referred to as “epsrm”;
- the operational 51-member ECMWF EPS starting at the same initial time as COSMO-LEPS (the “youngest” EPS constituting the super-ensemble), 80 km of horizontal resolution, referred to as “eps51”;

In Fig. 9, the average observed value of each box is compared with the average forecast value relative to the same box, for each of the three forecasting systems. The event considered is “precipitation exceeding 20mm/24h over 1.5 x 1.5 degree boxes”. Since the observed and forecast values are averaged over an area of 1.5 x 1.5 degrees, this threshold has been chosen as representative of an intense precipitation event.

In terms of ROC area, the eps51 configuration (Fig. 9, solid line with squares) shows the best scores at this threshold for every time range. The ensemble size, therefore, plays a major role in the computation of the probabilistic indices, making it difficult to carry out a proper comparison between cleps and eps51. When the two systems with the same ensemble size are compared, cleps (solid line with crosses) has higher scores that those of epsrm (solid line with triangles).

A comparison in terms of precipitation maxima has been also performed: the maximum forecast value within a 1.5 x 1.5 box is compared with the maximum observed value in the same box.

The ROC area values for cleps (Fig.10, line with crosses) are higher (and therefore better) than both the epsrm (line with triangles) and eps51 (line with squares), indicating that COSMO-LEPS is more skilful in forecasting correctly high precipitation values over a rather large area. It is worth pointing out that these results are based on a relatively large number of occurrences (about 600 and 150 for the 20 and 50 mm thresholds, respectively).

The same conclusions apply when the cost-loss curves relative to the three systems are considered for the event “maximum precipitation exceeding 50mm/24” at the 90-hour forecast range (Fig. 11). To maintain the diagram readable, only the envelope curve is plotted for each system.

The curve relative to cleps (solid line) is well above the curves of both eps51 and epsrm. This is especially true for low cost-loss ratios (left portion of the x-axis), where the global model ensembles have almost no value.

5. Summary and concluding remarks

Since the day, almost 20 years ago, in which Hank Tennekes stated, during a meeting of the Scientific Advisory Committee of the European Centre for Medium-Range Weather Forecasts, that "no forecast is complete without a forecast of the forecast skill", the demand for numerical forecasting tools which should have the capability, at the same time, of providing quantitative estimates of forecast reliability and of casting quantitative forecasts of various meteorological parameters in terms of probabilities, has been ever increasing. Consistently with this, among ECMWF's operational users, output from the now twice-daily EPS system is eroding more and more the historical primacy detained by the single, daily, purely deterministic, highest possible resolution, "main", 12:00GMT, 10-day forecast run, which has for a long time represented by far the main source of reliable, good-quality, numerical short- and medium-range forecast information.

EPS products, however, suffer from a number of drawbacks, most of them related to the comparatively low resolution of the global model used to produce them. The size of the ensemble, i.e. the number of the forecasts which compose it, is a very important factor in producing a sufficiently adequate "exploration" of the phase-space of all possible future atmospheric states which are compatible with our imperfect knowledge of the atmospheric initial conditions. This imposes a compromise between model resolution and the total number of elements needed in the ensemble, i.e. the model resolution has to be kept low enough to make the total computational requirements affordable. This has as a consequence that in the EPS model a number of local (mainly orography-related) atmospheric features/phenomena/processes are still misrepresented or underestimated. Orographically-related local and/or intense precipitation is one of the best possible such examples, but not the only one. This carries with it a reduced capability of the EPS system to provide useful guidance in case of intense meteorological events (often catastrophical in terms of consequences on life and property), notably those characterized by large rainfall amounts, possibly leading to floods.

In the attempt, therefore, to push the point of compromise between resolution and number-of-elements further toward a better representation of local effects (such as the one typical of high-resolution Limited-Area-Models), without having to sacrifice too much to completeness of the sampling of the atmospheric phase-space, the LEPS (Limited-are Ensemble Prediction System) system was developed at ARPA-SIM. The basic goal is to get most of the advantages of a brute-force approach, which can hardly be afforded on an operational basis (run a high-resolution LAM integration from each and every element of an ECMWF EPS, for example), without having to pay the full computational cost.

The LEPS idea is in fact very simple: once all EPS elements have been grouped in clusters, from all the elements of each cluster a most "Representative Member" (RM) is defined, which carries with it a weight proportional to the population of the cluster it represents and from which a high-resolution LAM is integrated. The LAM integrations can be combined, using the cluster-population weights, to produce all statistically-based predictions typical of a complete EPS system, like for example the probability for a meteorological variable to overcome given thresholds as a function of space and time.

Preliminary case-study-based results were considered promising enough to launch a COSMO-LEPS pre-operational, daily experimentation exercise, carried out nesting, in LEPS mode, a limited-area model (Lokal Modell) within the ECMWF EPS model output. At the time of writing (January 2005) the experimentation is carried out constructing a 102-element super ensemble from two 12-hour lagged operational EPS ensembles and integrating ten times a 10-km mesh version of LM once a day for 120 hours. The limited-area model integrations using Lokal Modell are carried out on a domain covering the geographical territory of all countries participating to COSMO. The experiment is carried out on the ECMWF computer system, to minimize file transfer requirements. This extended pre-operational test started during the first week of November 2002 on computer resources coming from the ECMWF Member States which are also COSMO partners (that is Germany, Greece, Italy and Switzerland).

As examples of the high quality of COSMO-LEPS products it was shown that, for cases of heavy precipitation events, the system is able to predict the possibility of a flood event already four days ahead and to confirm (and reinforce) the scenario as the forecast range gets shorter.

The impact of the large ensemble-size reduction on the forecast accuracy has been studied for a number of case studies. It turns out that the heavy rainfall scenarios are already captured when the limited-area ensemble size is reduced from 51 to either 5 or 10 members. This result is very encouraging: for these particular case studies, the clustering-selection technique highlights the most important evolution scenarios and enables the generation of reliable and accurate probability maps.

A statistical analysis of the methodology on which COSMO-LEPS was based for the first part of experimentation (November 2002 – June 2004) was also undertaken, focussing the attention of the prediction of precipitation. The results indicate that the use of two EPS in the super-ensemble is a reasonable compromise between the decrease in the percentage of outliers (the cases for which no LEPS member comes close the observed rainfall intensity) and a modest decrease of the forecast skill. In addition to this, it was found that doubling the ensemble size permitted to halve the percentage of outliers. These results led to modify the COSMO-LEPS operational suite at the beginning of June 2004 as follows: use of 2 EPS (instead of 3) to construct the super-ensemble and increase the limited-area ensemble size from 5 to 10 members.

As a final step, an objective verification of the COSMO-LEPS system was shown for Autumn 2003 for forecasts up to 120 hours. It was found that, as far as precipitation averaged over 1.5 x 1.5 degrees boxes is concerned, the operational full-size EPS exhibits a higher skill than COSMO-LEPS. This is appears to be due to the higher population of the global model ensemble. In fact, when COSMO-LEPS is compared with a reduced size (comparable population) version of the EPS, the limited-area ensemble has noticeably greater skill. On the other hand, COSMO-LEPS outperforms both full-size and reduced-size EPS in terms of prediction of precipitation maxima over the boxes.

These results could be conducive to a truly operational implementation of the system to assist forecasters and agencies involved in civil protection tasks for the issues of warnings due to localised severe weather and are strengthening the value of ensemble forecasting systems at ranges shorter than the full medium-range.

References

Arribas, A., 2004. Results of an initial stochastic physics scheme for the Met Office Unified Model. Forecasting Research Technical Report No 452. available at Met Office, Exeter, UK.

Buizza R., 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99-119.

Buizza R., Barkmeijer J., Palmer T.N., Richardson D., 1999. Current status and future developments of the ECMWF Ensemble Prediction System. *Meteor. Applic.*, **6**, 1-14.

Buizza, R. 2005. The ECMWF Ensemble Prediction System. In Palmer, T.N. and Hagedorn, R., editors, *Predictability of Weather and Climate*. Cambridge University Press.

Du J., DiMego G., Tracton M. S. and Zhou B., 2003. NCEP short-range ensemble forecasting (SREF) system: multi-IC, multi-model and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling, Report 33, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 1161.

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

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

Haakenstad H. and I-L. Frogner, 2004. Extended tests with a limited area ensemble prediction system at the Norwegian Meteorological Institute – LAMEPS. Met.no report no.7.

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. Wea. Rev.*, **128**, 3501-3519.

Houtekamer P. L., Derome J., Ritchie H. and Mitchell H. L., 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Janjic Z. I., 1990. The step-mountain co-ordinate: physical package. *Mon. Wea. Rev.*, **118**, 1429-1443.

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

Marsigli C., Montani A., Nerozzi F., Paccagnella T., 2004. "Probabilistic high-resolution forecast of heavy precipitation over Central Europe." *Natural Hazard and Earth System Sciences*, **4**, 315-322.

Mason S.J. and Graham N.E., 1999. Conditional probabilities, relative operating characteristics and relative operating levels. *Wea. and Forecasting*, **14**, 713-725.

Mesinger F., Janjic Z. I., Nickovic S., Gavrilov D. and Deaven D. G., 1988. The step-mountain co-ordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of Appalachian redevelopment. *Mon. Wea. Rev.*, **116**, 1493-1518.

Molteni F., Buizza R., Palmer T. N. and Petroliagis T., 1996. The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **127**, 2069-2094.

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. *Quart. J. Roy. Meteor. Soc.*, **127**, 2069-2094.

Montani A., 1998. Targeting of observations to improve forecasts of cyclogenesis. PhD thesis. University of Reading, Reading, UK.

Montani A., Marsigli C., Nerozzi F., Paccagnella T. and Buizza R., 2001. Performance of ARPA-SMR Limited-area Ensemble Prediction System: two flood cases. *Nonlinear Processes in Geophysics*, **127**, 2095-2115.

Montani A., Marsigli C., Nerozzi F., Paccagnella T., Tibaldi S. and Buizza R., 2003a. The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts. *Non-linear Processes in Geophysics*, **10**, N. 3, 261-274.

Montani A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U., Nerozzi, F., Paccagnella, T., Patruno, P. and Tibaldi, S., 2003b. “Operational limited-area ensemble forecasts based on the Lokal Modell”. *ECMWF Newsletter*, **98**, 2-7.

Mullen S.L. and Buizza R., 2001. Quantitative precipitation forecast over the United States by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **129**, 638-663.

Mullen S.L. and Buizza R., 2002. The Impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System., *Wea. And Forecasting*, **17**, 173-191.

Mylne K.R., Evans, R.E. and Clark, R.T., 2002. Multi-model multi-analysis ensembles in a quasi-operational medium-range forecasting, *Quart. J. Roy. Meteor. Soc.*, **128**, 361-384.

Palmer T. N. and Tibaldi S., 1988. On the prediction of forecast skill. *Mon. Wea. Rev.*, **116**, 2453-2480.

Richardson D.S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649-667.

Tibaldi S., Paccagnella T., Marsigli C., Montani A., Nerozzi F., 2003. Short-to-medium-range Limited-area Ensemble Prediction: the LEPS system. *Proceedings of ECMWF Seminars on Predictability of weather and climate*, 9-13 September 2002, Reading, UK.

Tracton M. S. and Kalnay E., 1993. Operational ensemble prediction at the National Meteorological Centre: Practical Aspects. *Wea. and Forecasting*, **8**, 379-398.

Tracton M. S., Du J. and Juang H., 1998. Short-range ensemble forecasting (SREF) at NCEP/EMC. Proceedings of the 12th AMS/NWP Conference, Phoenix AZ, 269-272.

Wilks D. S., 1995. Statistical methods in atmospheric sciences. Academic Press, New York, 467 pp.

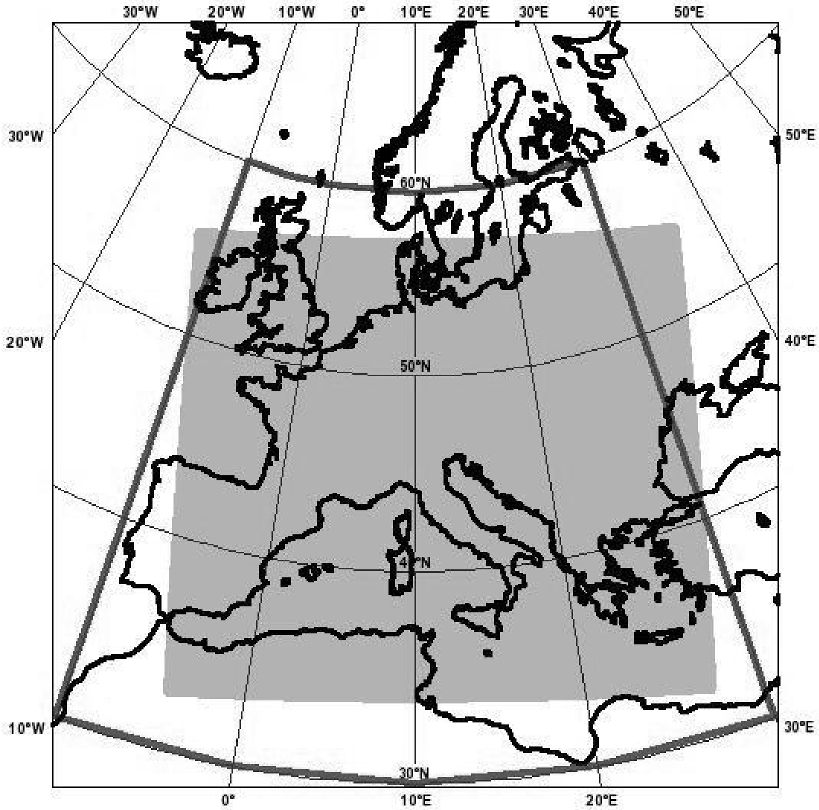

Figure 1: COSMO-LEPS operational domain (small circles) and clustering area (thick rectangle).

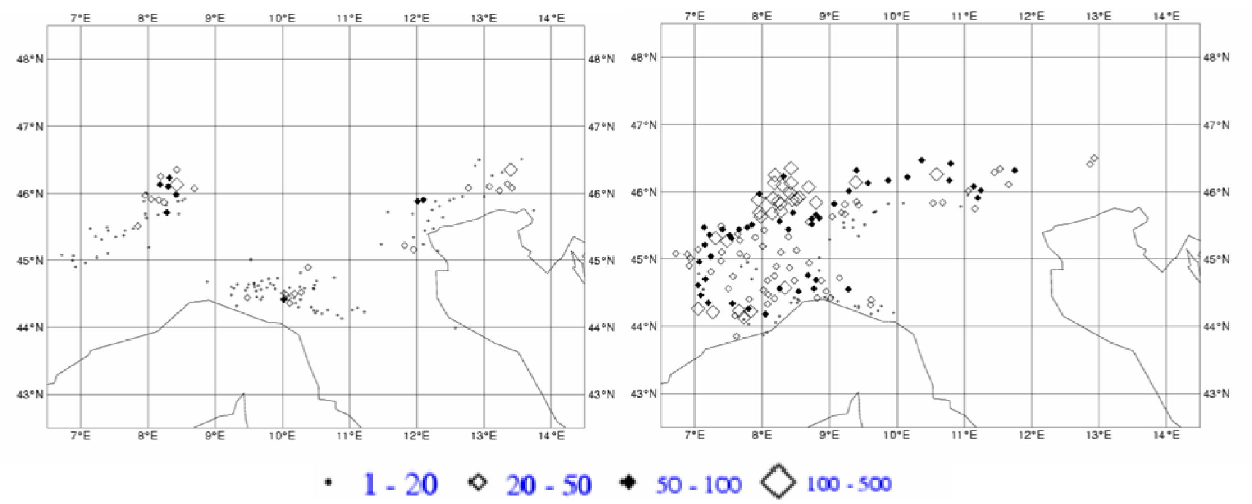

Figure 2: 24-hour observed precipitation for the Friuli (left; cumulation period starting at 12UTC of 28 August 2003) and Piedmont (right; cumulation period starting at 12UTC of 15 November 2002); thresholds: 1mm (black dots), 20mm (empty diamonds), 50mm (full diamonds), 100mm (large empty diamonds).

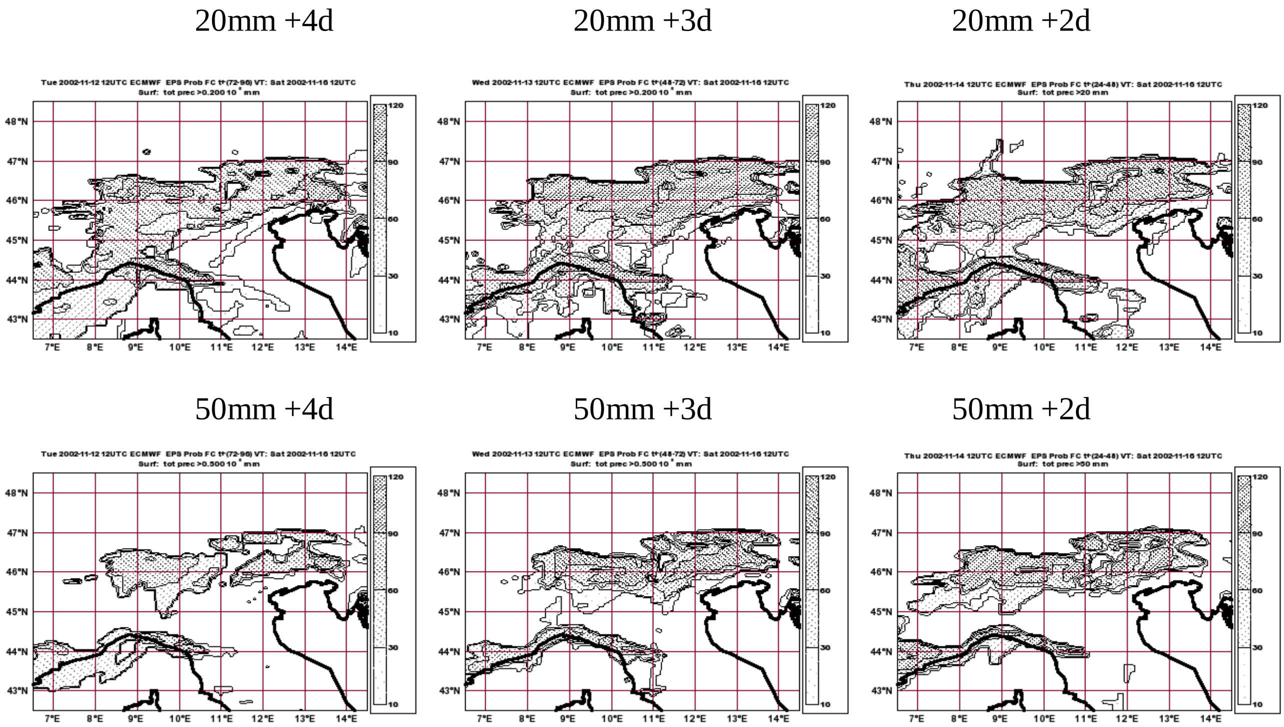

Figure 3: Piedmont case (all maps verify at 12UTC of 16 November 2002); COSMO-LEPS probability maps for 24h rainfall exceeding 20 mm (top-row panels) and 50 mm (bottom-row panels) at the 96h range (left-column panels), 72h range (middle-column panels) and 48h range (right-column panels). Contour intervals: 10%, 30%, 60%, 90%.

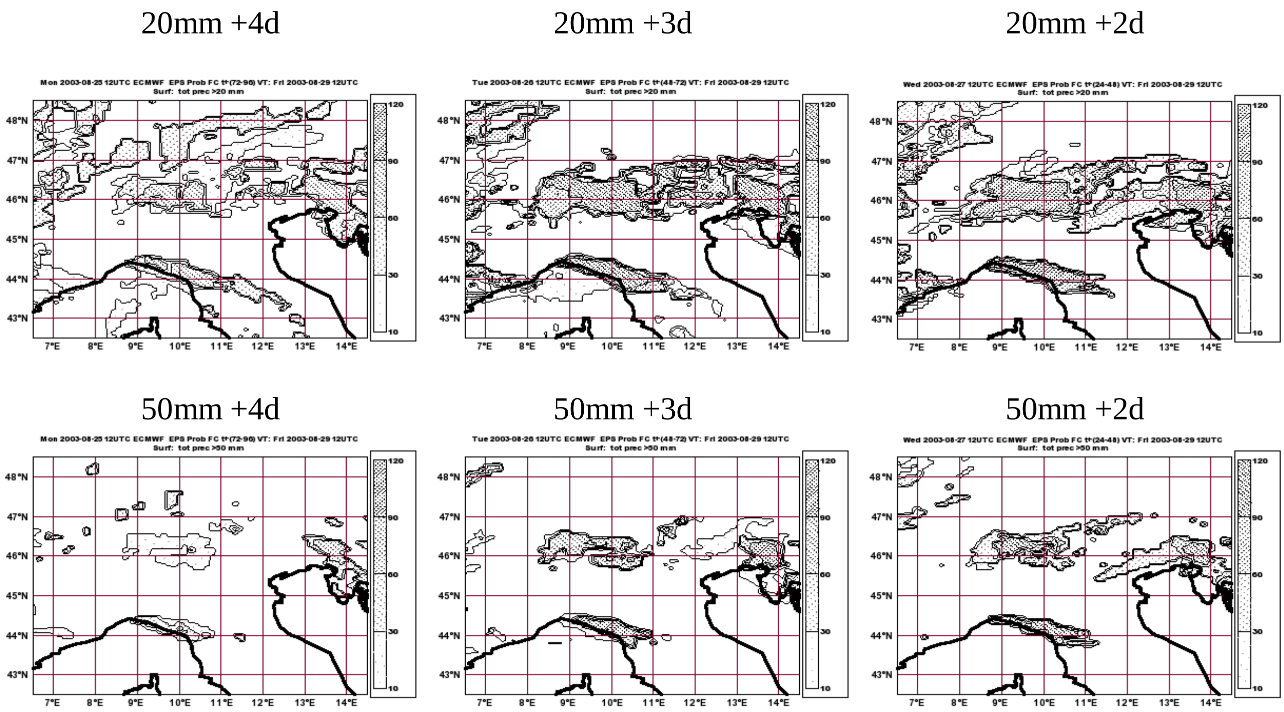

Figure 4: the same as Fig. 3 but for the Friuli case (all maps verify at 12UTC of 28 August 2003).

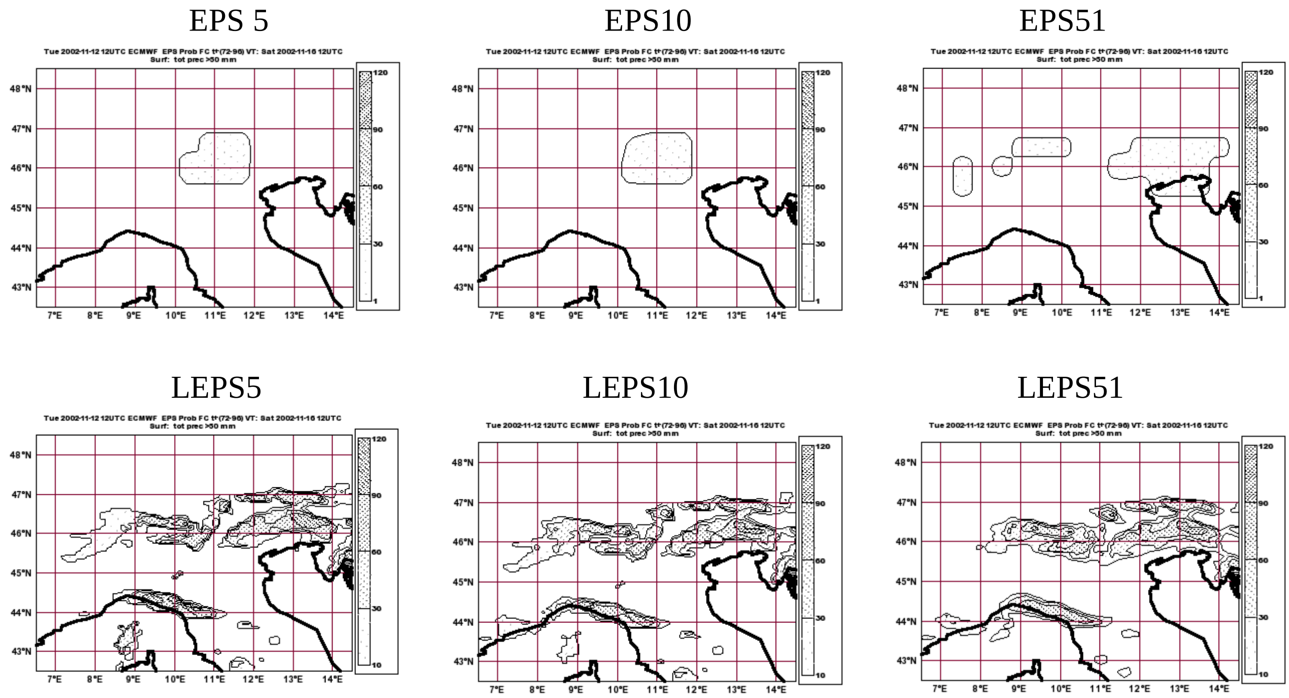

Figure 5: Piedmont case; probability maps for 24h rainfall exceeding 50 mm (4-day forecast verifying at 12UTC of 16 November 2002) for the 5-member EPS (top-left panel), 10-memebr EPS (top-middle), 51-member EPS (top-right), 5-member LEPS (bottom-left), 10-member LEPS (bottom-middle) and 51-member LEPS (bottom-right). Contour intervals: 1%, 30%, 60%, 90% for top-row panels; 10%, 30%, 60%, 90% for bottom-row panels.

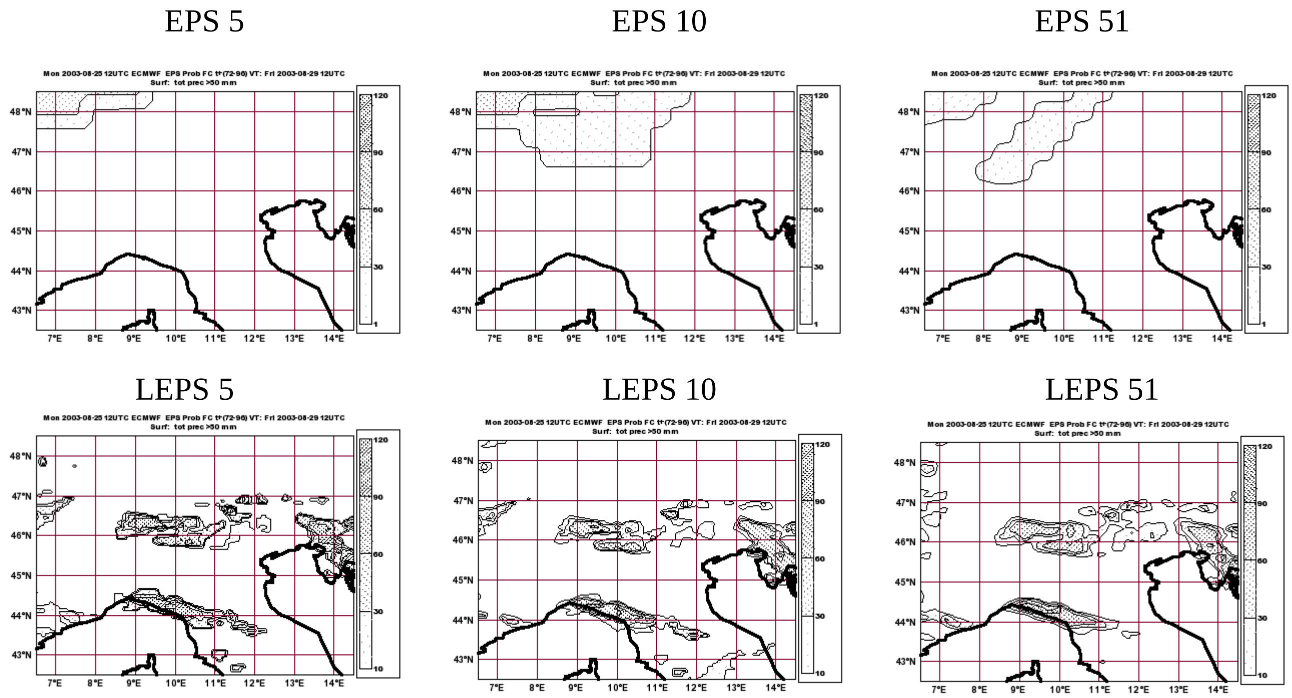

Figure 6: the same as Fig.5 but for the Friuli case (all maps verify at 12UTC of 28 August 2003).

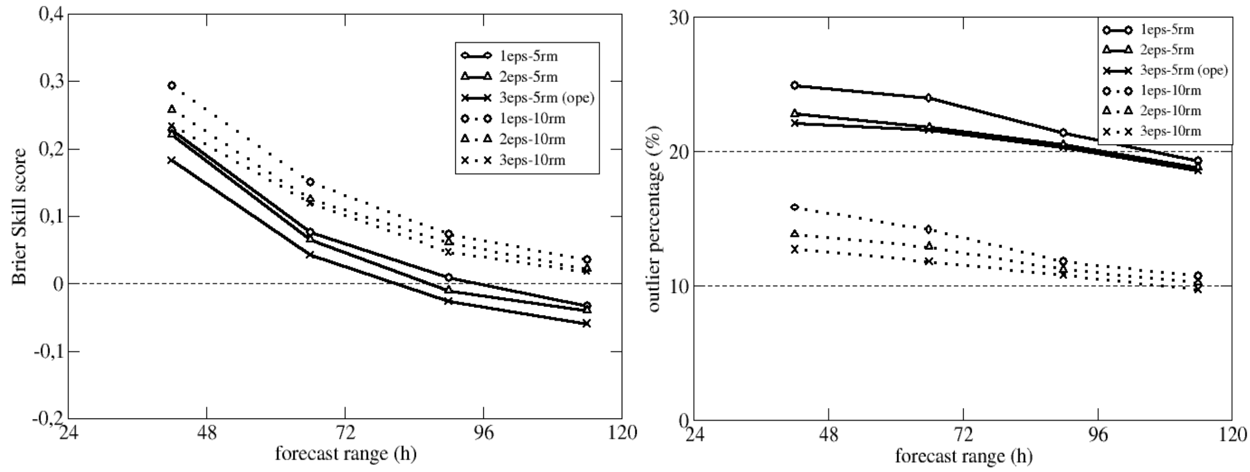

Figure 7: Brier Skill Score and Percentage of outliers for different configurations of EPS RMs.

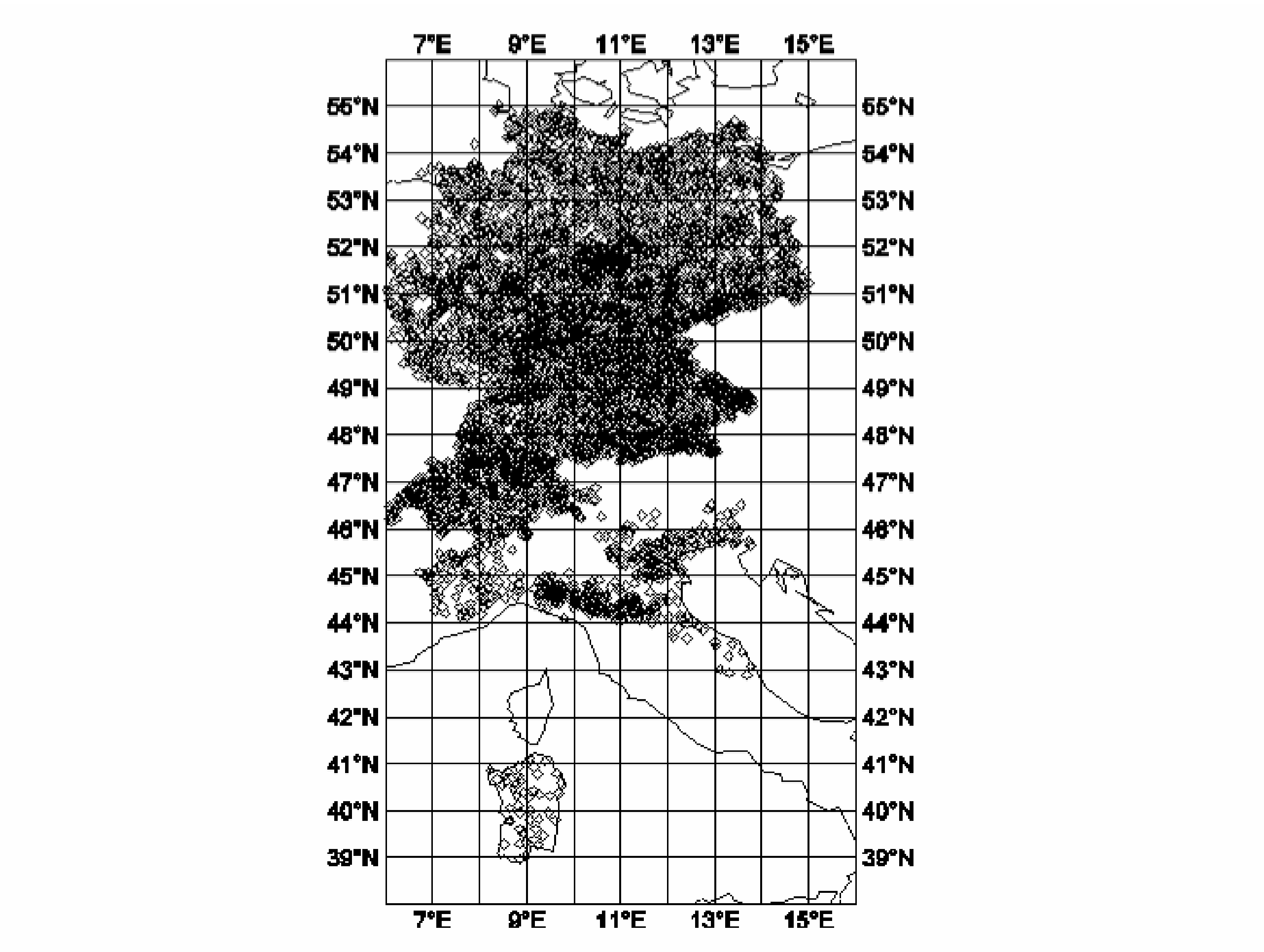

Figure 8: Network of about 4000 stations in Germany, Switzerland and Italy recording daily precipitation (cumulated from 06 to 06 UTC the following day). The network was used to compute daily mean and maximum precipitation in 1.5x1.5 degree boxes used to verify forecasts in Section 4.2.

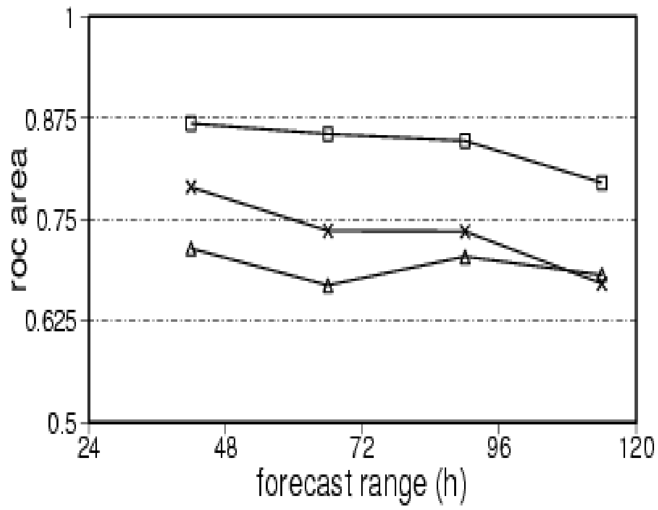

Figure 9: Average forecasts against average observations: ROC area values for precipitation over 1.5 x 1.5 boxes exceeding 20mm/24h The crosses are relative to the COSMO-LEPS system, the squares to the 51-member EPS and the triangles to the 5-RM EPS.

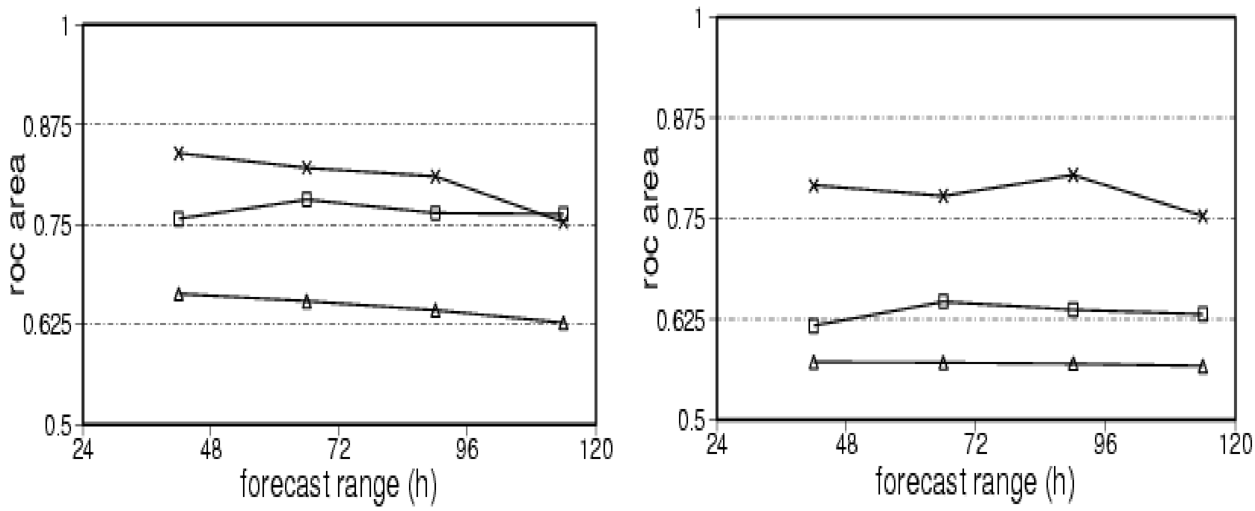

Figure 10: Maximum forecasts against maximum observations: ROC area values for precipitation over 1.5 x 1.5 boxes exceeding 20 (left panel) and 50 (right panel) mm/24h. The crosses are relative to the COSMO-LEPS system, the squares to the 51-member EPS and the triangles to the 5-RM EPS.

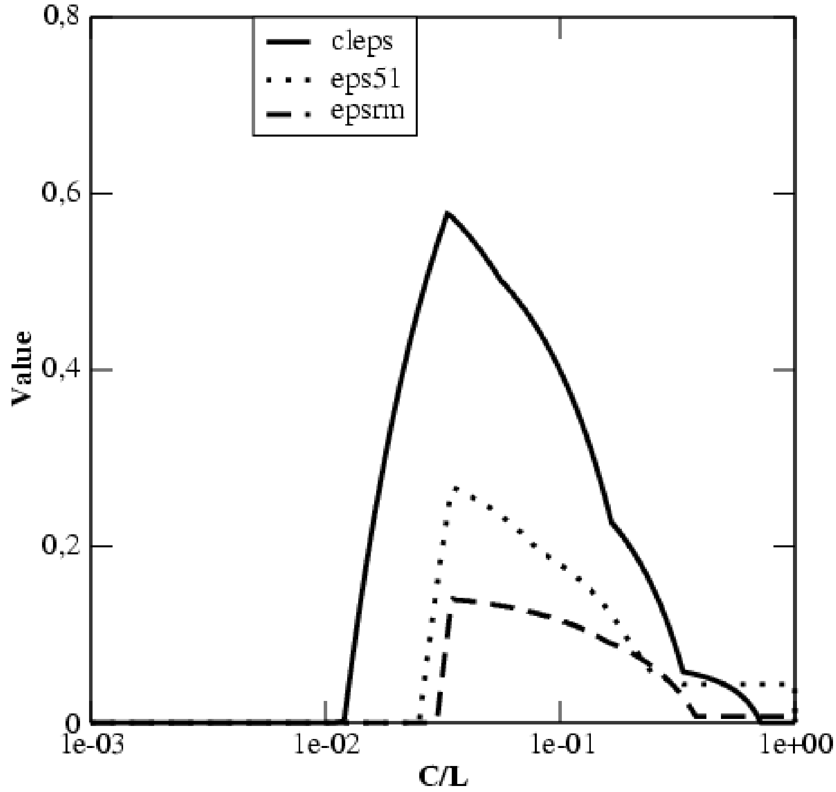

Figure 11: Maximum forecasts against maximum observations: cost-loss curves for precipitation over 1.5 x 1.5 boxes exceeding 50mm/24h (90 hour forecast range). The solid line is relative to the COSMO-LEPS system, the dotted line to the 51-member EPS and the dashed line to the 5-RM EPS.