**Nonlinear Processes
in Geophysics**

The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification

C. Marsigli, F. Boccanera, A. Montani, and T. Paccagnella

ARPA-SIM, Bologna, Italy

Abstract. The limited-area ensemble prediction system COSMO-LEPS has been running every day at ECMWF since November 2002. A number of runs of the non-hydrostatic limited-area model Lokal Modell (LM) are available every day, nested on members of the ECMWF global ensemble. The limited-area ensemble forecasts range up to 120 h and LM-based probabilistic products are disseminated to several national and regional weather services. Some changes of the operational suite have recently been made, on the basis of the results of a statistical analysis of the methodology. The analysis is presented in this paper, showing the benefit of increasing the number of ensemble members. The system has been designed to have a probabilistic support at the mesoscale, focusing the attention on extreme precipitation events. In this paper, the performance of COSMO-LEPS in forecasting precipitation is presented. An objective verification in terms of probabilistic indices is made, using a dense network of observations covering a part of the COSMO domain. The system is compared with ECMWF EPS, showing an improvement of the limited-area high-resolution system with respect to the global ensemble system in the forecast of high precipitation values. The impact of the use of different schemes for the parametrisation of the convection in the limited-area model is also assessed, showing that this have a minor impact with respect to run the model with different initial and boundary condition.

1 Introduction

The forecast of severe weather events is still a challenging problem. The key role played by mesoscale and orographic-related processes can seriously limit the predictability of intense and localised events. Although the use of high-resolution limited-area models (LAMs) has improved the short-range prediction of locally intense events, it is some-

Correspondence to: C. Marsigli
(cmarsigli@smr.arpa.emr.it)

times difficult to forecast accurately their space-time evolution, especially for ranges longer than 48 h. In the recent years, many weather centres have given more and more emphasis to the probabilistic approach (Tracton and Kalnay, 1993; Molteni et al., 1996; Houtekamer et al., 1996), which has proved to be an important tool to tackle the predictability problem beyond day 2. Nevertheless, global ensemble systems are usually run at a relatively low horizontal resolution (80 km at most), making difficult their use where the forecast of severe and localised weather events is concerned. With regard to the use of limited-area models within ensemble systems, ARPA-SIM (the Regional Hydro-Meteorological Service of Emilia-Romagna, in Italy) developed LEPS (Limited-area Ensemble Prediction System) (Molteni et al., 2001; Marsigli et al., 2001; Montani et al., 2001, 2003a), which after some tests led to the COSMO-LEPS implementation (Montani et al., 2003b).

The LEPS methodology allows us to combine the benefits of the probabilistic approach (a set of different evolution scenarios is provided to the forecaster) with the high-resolution detail of the LAM integrations, with a limited computational investment. The methodology is based on an algorithm that selects a number of members out of a global ensemble system. In particular, the 51-member ECMWF EPS (Ensemble Prediction System) is used. The selected ensemble members (called Representative Members, RMs) provide initial and boundary conditions to run a limited-area model.
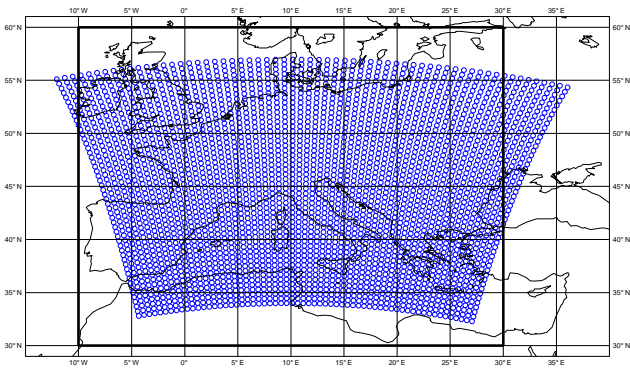
Following the encouraging results of the early experimental phase, the generation of an "experimental-operational" limited-area ensemble prediction system, the COSMO-LEPS project, has started in November 2002 on the ECMWF computer system under the auspices of COSMO (Montani et al., 2003b). COSMO (COnsortium for Small-scale MOdeling, http://www.cosmo-model.org) is a consortium involving Germany, Italy, Switzerland, Greece and Poland which aims to develop, improve and maintain the non-hydrostatic limited-area model Lokal Modell (LM). COSMO-LEPS aims at the development and pre-operational test of a "late-short to early-medium-range" (48–120 h) probabilistic fore-

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

casting system using a LAM over a domain covering all countries involved in COSMO (Fig. 1).

The methodology can be regarded as a downscaling of the forecasts provided by a global ensemble, aiming at transferring to the probabilistic approach the benefit of the high-resolution. The perturbations are introduced into the limited-area model both through the initial conditions and by means of the perturbed lateral boundary conditions. The added value of the system resides in joining the skill of a global ensemble system to depict the possible evolution scenarios with the capability of an high-resolution LAM to describe better local meteorological processes. Perturbations in the initial conditions have not been considered due to two main reasons: the computational burden and the fact that COSMO-LEPS is mainly designed for the medium-range (day 3–5), where the impact of the boundary conditions is regarded to be more relevant. Recently (June 2004) model perturbations have been introduced withing COSMO-LEPS by using two different convection schemes. This was done in order to try to consider a more local source of spread.

The developed methodology makes use of a global ensemble size reduction technique in order to avoid to increase too much the computational costs: a price in terms of ensemble size is paid in order to save computational time for the high-resolution integrations. The ensemble size reduction is necessary in order to render affordable the integrations of a number of limited-area model runs within a few hours. This permits to have the forecasts issued by the system available in the early morning in the meteorological operation rooms, where its performances can be evaluated by the forecasters in real time.

The extent to which the aim has been reached has been evaluated both with an objective verification, which is presented in this paper, and on a case-study basis. In the above mentioned references and in Marsigli et al. (2004), it has been shown that, over a number of test cases and for several forecast ranges (48–120 h), LEPS (the early experimental system) and COSMO-LEPS (the real-time quasi-operational system) have shown better performance than EPS for the quantitative forecast of intense precipitation, as well as the

geographical localisation of the regions most likely to be affected by the flood events. Then, as regards severe precipitation events, the impact the high-resolution within a probabilistic system seems to be positive.

An objective verification of COSMO-LEPS is being carried out at ARPA-SIM, focusing the attention on the precipitation forecast. Verification aims towards an understanding of the abilities and shortcomings of the system, in order to ameliorate its design and to provide guidelines to the end users (forecasters, civil protection, etc). In this paper, verification of daily precipitation has been performed over the period September-November 2003. The probabilistic indices used in this paper are: Brier Skill Score (Wilks, 1995), ROC Curves (Mason and Graham, 1999) and Percentage of Outliers (Buizza, 1997). With regards to the system configuration, the analysis focuses on the methodology that leads to the choice of the Representative Members. This analysis has been performed over the same period.

The paper is organised as follows: in Sect. 2 the COSMO-LEPS system is described, as it has been since June 2004, while in Sect. 3 a statistical analysis of the methodology is presented, leading to a new configuration of the system. In Sect. 4 an objective verification of the performance of COSMO-LEPS is carried out, comparing the system with the ECMWF EPS. In Sect. 5, the COSMO-LEPS is compared with a parallel suite in which another scheme for the parametrisation of the precipitating convection is used. Finally, conclusions are drawn in Sect. 6.

2 The COSMO-LEPS operational system

The set-up of the COSMO-LEPS suite, as it was when the verification was carried out, is described in this section. From the beginning of June 2004 the suite has changed, as a consequence of the results obtained in Sect. 3 and in Sect. 5.

A Cluster Analysis and Representative Member Selection Algorithm is applied to the ECMWF global ensemble system. For a description of the Cluster Analysis technique, a technique which permits to separate data into groups whose identity are not known in advance, the reader is referred to Wilks (1995). The Ensemble Prediction System (EPS) is now based on a $T_L 255 L 40$ model (spectral model with truncation at wavenumber 255 and 40 vertical levels), corresponding to a horizontal resolution of about 80 km, and has 51 members (Molteni et al., 1996; Buizza et al., 1999). Three successive 12-h-lagged EPS runs (started at 12:00 UTC of day N-2, at 00:00 and 12:00 UTC of day N-1) are grouped together so as to generate a 153-member super-ensemble; (see Fig. 2). A hierarchical cluster analysis is performed on the 153 members so as to group all elements into 5 clusters (of different populations); the clustering variables are the geopotential height, the two component of the horizontal wind and the specific humidity at three pressure levels (500, 700, 850 hPa) and at two forecast times (fc+96 and fc+120 for the "youngest" EPS, the one started at 12:00 UTC of day N-1);

the cluster domain covers the region 30° N–60° N, 10° W–40° E (rectangle in Fig. 1).

The use of the super-ensemble was introduced (Montani et al., 2003a) aiming at increasing the spread of the global ensemble on which the cluster analysis is performed.

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; hence, 5 RMs are selected. Each RM provides initial and boundary conditions for the integrations with LM, which is run 5 times for 120 h, always starting at 12:00 UTC of day *N-1* and ending at 12:00 UTC of day *N+4*. The LM is run with a horizontal resolution $\Delta x \simeq 10$ km and with 32 levels in the vertical; the time-step used for the integrations is 60 s.

Probability maps based on LM runs are generated by assigning to each LM integration a weight proportional to the population of the cluster from which the RM (providing initial and boundary conditions) was selected. Deterministic products (that is, the 5 LM scenarios in terms of surface and upper-level fields) are also produced.

The products are disseminated to the COSMO community for evaluation. COSMO-LEPS dissemination started during November 2002 and, at the time of writing (September 2004), 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.

3 Statistical analysis of the methodology

The idea of joining three consecutive EPS to form a super-ensemble is based on the need of enlarging the size of the ensemble on which the RM selection algorithm is applied. This permits an increase in the ensemble spread and a wider part of the phase space spanned by the global ensemble members. Nevertheless, this is obtained by paying a price in terms of skill: the older the EPS, the less skillful their members are. In order to quantify the relative effects of the increased spread and of the decreased skill, the Representative Members chosen with the current methodology are compared to those chosen using only one or two EPS. The three ensembles that are compared are, then:

- the ensemble made up by the 5 RMs selected applying the Cluster Analysis and Representative Member Selection Algorithm on the three most recent EPS (referred to "3-EPS"), which is the original operational configuration.

- the ensemble made up by the 5 RMs selected applying the Cluster Analysis and Representative Member Selection Algorithm on the two most recent EPS (referred to "2-EPS").

- the ensemble made up by the 5 RMs selected applying the Cluster Analysis and Representative Member Selec-
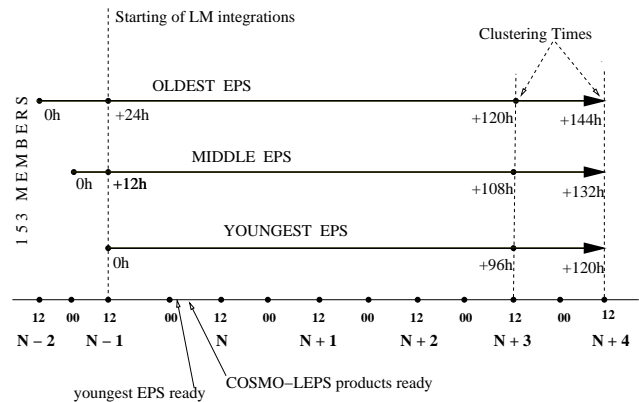

Fig. 2. Details of the COSMO-LEPS suite.

tion Algorithm on the most recent EPS (referred to "1-EPS").

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

Results show that the Brier Skill Score (the higher the better) is higher when the clustering is based on the most recent EPS only (Fig. 3, black line), while it is lower for the 3-EPS super-ensemble (blue line). The difference between the two is not so remarkable, but it remains at every forecast range. The 2-EPS super-ensemble (red line) has an intermediate skill, equal to the one of the 1-EPS ensemble at the first and last forecast ranges, its general performance being closer to that of the 1-EPS ensemble. Similar conclusions are drawn when the ROC area scores are considered (not shown).

The percentage of outliers of the systems is also shown. This is the percentage of times the "truth" falls out of the range of the forecast values, so the lower the better. The percentage of outliers (Fig. 4) of the 1-EPS ensemble (black line) is rather higher than the other two, for every forecast range, while there is almost no difference in terms of outliers between the 2-EPS (red line) and the 3-EPS (blue line) ensembles. These results seem to indicate that the use of just two EPS in the super-ensemble can be a good compromise, permitting a significant decrease in the percentage of outliers but leading to only a small worsening of the skill.

In order to quantify the impact of the ensemble size on the performance of the system, the cluster analysis has been repeated by fixing the number of clusters to 10 and by selecting, then, 10 Representative Members. This has been done for each of the three ensemble configurations already considered, leading to the three configurations: 3-EPS-10RMs, 2-EPS-10RMs and 1-EPS-10RMs. The impact of the ensemble size proves to be quite remarkable; the difference between
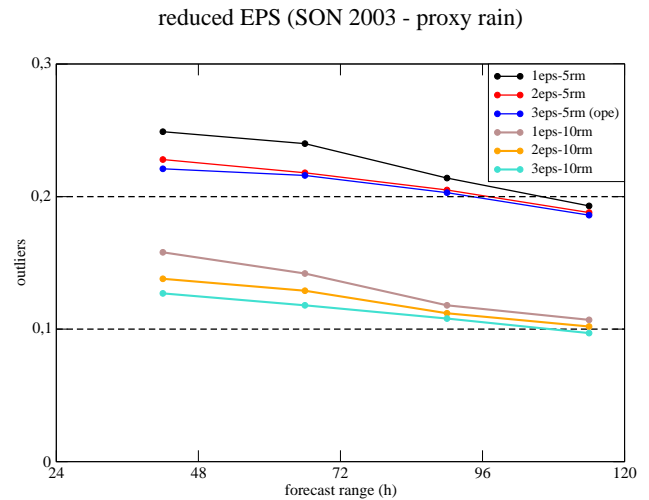
reduced EPS (SON 2003 - proxy rain - no weight)
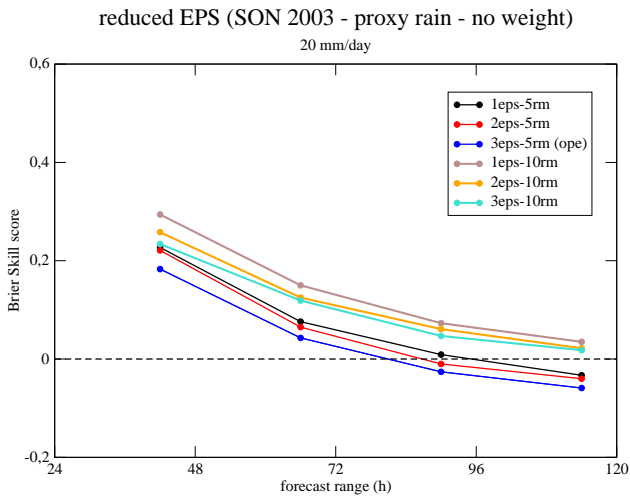
Fig. 3. Brier Skill Score as a function of the forecast range for the event precipitation exceeding 20 mm/24 h relative to the RM EPS. The different configurations are: 5 clusters algorithm based on 1 EPS (black line), on 2 EPS (red line) and on 3 EPS (operational configuration, blue line); 10 clusters algorithm based on 1 EPS (gray line), on 2 EPS (orange line) and on 3 EPS (cyan line).

reduced EPS (SON 2003 - proxy rain)

Fig. 4. Percentage of outliers for the RM EPS. The different configurations are: 5 clusters algorithm based on 1 EPS (black line), on 2 EPS (red line) and on 3 EPS (operational configuration, blue line); 10 clusters algorithm based on 1 EPS (gray line), on 2 EPS (orange line) and on 3 EPS (cyan line).

each 5-member ensemble and the correspondent 10-member ensemble being about 0.1 in terms of Brier Skill Score, for every configuration. This is shown in Fig. 3, where the blue line (3-EPS-5RMs) is to be compared with the cyan line (3-EPS-10RMs), the red line (2-EPS-5RMs) with the orange line (2-EPS-10RMs) and the black line (1-EPS-5RMs) with the brown line (1-EPS-10RMs). The impact of doubling the ensemble size is almost the same for every configuration and is predominant with respect to the impact of changing the number of EPS on which the Cluster Analysis is performed. In terms of Outliers (Fig. 4), it can be seen that doubling the ensemble size greatly reduces the Percentage of Outliers, but, due to the dependence of this measure on the ensemble size, a direct comparison may be not appropriate.

These results led to two major modification of the COSMO-LEPS methodology at the beginning of June 2004: the super-ensemble has been built by using only the 2 most recent EPS and the number of clusters has been fixed to 10 (2-EPS-10RMs configuration), nesting Lokal Modell on each of the so selected 10 RMs.

4   Verification of COSMO-LEPS against the EPS

In order to quantify the added value brought about by the mesoscale probabilistic system, COSMO-LEPS is compared with the EPS. The comparison is made difficult by two main factors: the difference in the number of ensemble members (5 for COSMO-LEPS and 51 for the EPS) and the difference in terms of resolution (10 km for COSMO-LEPS and 80 km for the EPS). As far as the population of the ensembles is concerned, COSMO-LEPS is also compared with a small EPS ensemble made up by the 5 Representative Mem-

bers. This permits us to quantify the impact of the increased resolution alone. The problem of the very different resolutions of the two systems is tackled by upscaling both systems to a lower resolution: the grid point forecasts of both models are aggregated over boxes of $1.5 \times 1.5°$. For each model, this was done in two ways: averaging all the forecast values falling into the box or selecting the maximum among all the forecast values falling into the box. The comparison is made in terms of 24-h precipitation, against observed data from a very dense network of raingauges. Precipitation is accumulated from 06:00 to 06:00 UTC. In order to properly compare forecast values on grid points and observed values on station points, the observations within a box are treated, as the forecast values, in two ways: all the observed values falling into a box are averaged and the obtained value is compared directly with the averaged forecast value or the maximum among all the observed values falling into a box is selected and compared with the maximum forecast value. The comparison is carried out over a large area included in the COSMO-LEPS domain, covering Germany, Switzerland and Northern Italy. The dense network of stations recording daily precipitation (about 4000 every day) is shown in Fig. 5.
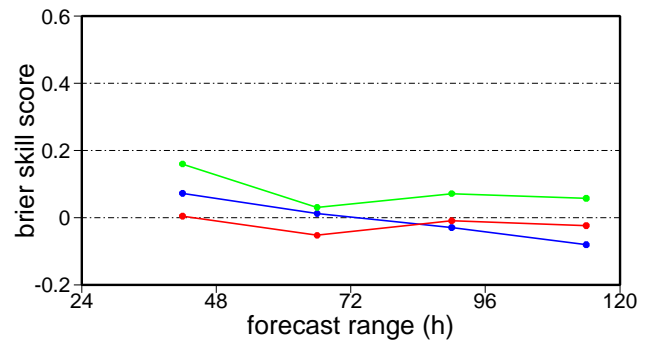
The three ensemble systems compared are:

– the COSMO-LEPS system, made up of 5 members, 10 km horizontal resolution, referred to as "cleps";

– the EPS mini-ensemble composed of the 5 Representative Members chosen from the super-ensemble, 80 km 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 horizontal resolution, referred to as "eps51".
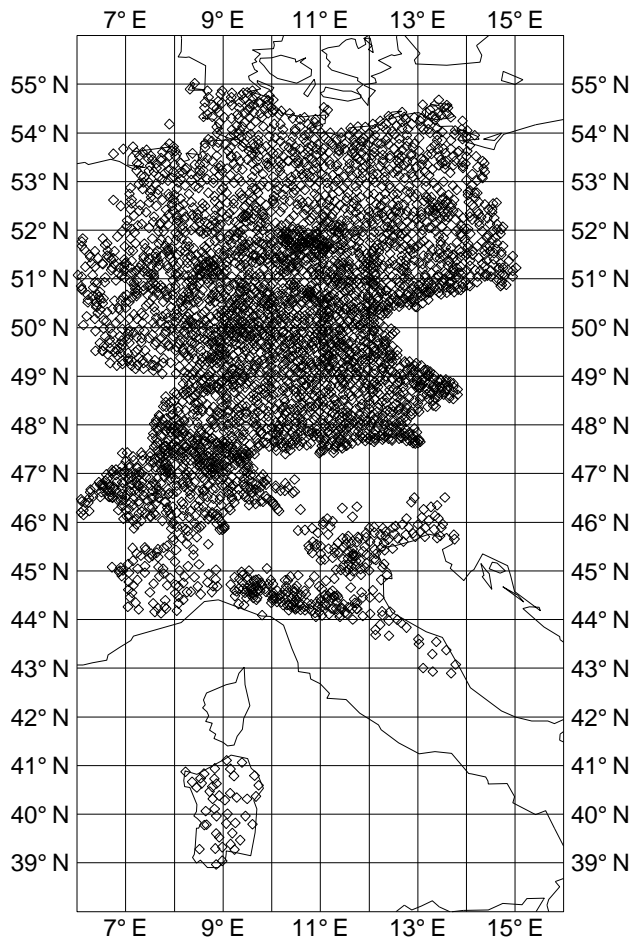
Fig. 5. Network of station providing 24-h precipitation (06:00 UTC to 06:00 UTC) for the COSMO verification.

Fig. 6. Brier Skill Score values for the precipitation threshold 20 mm/24 h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Average observed and forecast values over $1.5 \times 1.5^\circ$ boxes are compared.
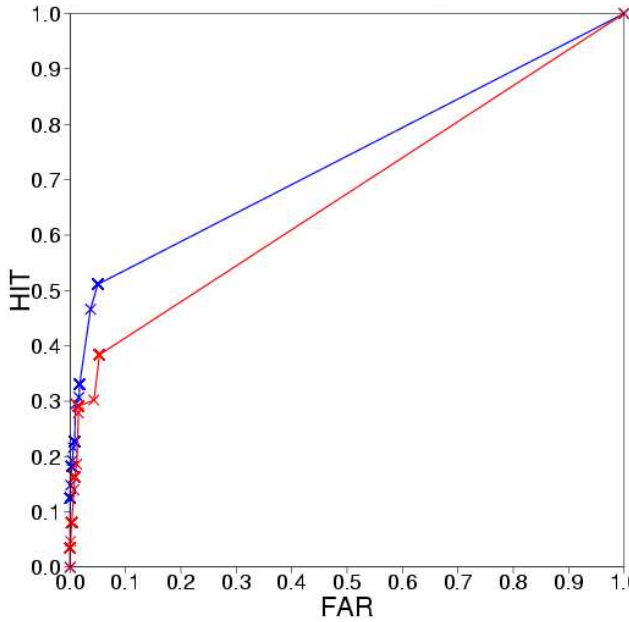

Fig. 7. ROC area for the precipitation threshold 20 mm/24 h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Average observed and forecast values over $1.5 \times 1.5^\circ$ boxes are compared.

First, the average observed value of each box, obtained by computing the mean of all the observations falling in a box, is compared with the average forecast value relative to the same box, for each of the three forecasting systems. The Brier Skill Score (Fig. 6) and the ROC area (Fig. 7) for the three systems are shown; for both indices, the higher the better. The event considered here is precipitation exceeding 20 mm/24 h over $1.5 \times 1.5^\circ$ boxes. Since the observed and forecast values are averaged over an area of $1.5 \times 1.5^\circ$, this threshold detects an intense precipitation.

In terms of Brier Skill Score (Fig. 6) the three lines are rather close together. The BSS values of the full-size 51-member EPS (eps51, green line) are slightly higher than those of the other two systems, which means that its performance is slightly better. Furthermore, the eps51 BSS is always positive, indicating the existence of some skill at all the time ranges. The difference between cleps (blue line) and epsrm (red line) is slightly in favour of cleps for the first forecast ranges, when it has positive BSS values, while the reverse is true at the +114 h forecast range. The additional skill of eps51 can also be due to the more recent initial conditions from which it benefits. In fact, both epsrm and cleps
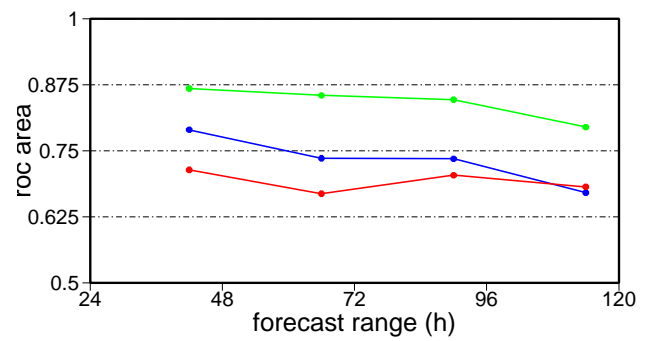
can contain members started 12 or 24 h before the members of eps51, due to the fact that the RMs are selected out of the lagged super-ensemble.

The differences in the performances of the three systems are enlighted by the ROC area values (Fig. 7).

The full-size 51-member EPS (eps51, green line) has the best scores at this threshold for every forecast range. The COSMO-LEPS system (cleps, blue line) has lower scores, but higher that those of the 5-RM EPS (epsrm, red line). The only exception is the +114 h forecast range, when the cleps score is as lower as the epsrm one. From these results, it appears that, when the two systems with the same size are compared, "cleps" shows an improvement with respect to the "epsrm", especially in terms of ROC area. In order to better understand this result, the ROC Curves for these two systems are also reported.

The ROC Curves relative to COSMO-LEPS and to the 5-RM EPS are shown for the event "average precipitation exceeding 20 mm/24 h", for the forecast ranges +66 h (Fig. 8) and +90 h (Fig. 9). The "cleps" curves (blue curves) are above the "epsrm" ones (red curves) for both forecast ranges. Considering the first cross from the top right in the diagrams,

Fig. 8. Average values: ROC Curves for the precipitation threshold 20 mm/24 h and for the +66 h forecast range. The blue line is relative to cleps while the red line is relative to the epsrm.
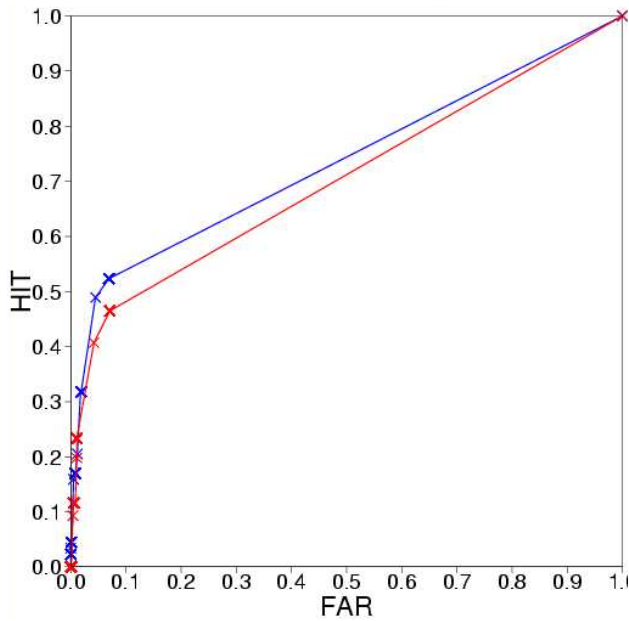
Fig. 9. Average values: ROC Curves for the precipitation threshold 20 mm/24 h and for the +90 h forecast range. The blue line is relative to cleps while the red line is relative to epsrm.

it is evident that the two systems have comparable False Alarm Rate, but COSMO-LEPS obtains higher Hit Rate values. This cross is correspondent to the probabilistic issue "at least one ensemble member is forecasting the event", whose practical meaning can be understood referring to a alert situation. If a user have a damage from the considered event, he can avoid the related loss by taking a preventive action. In
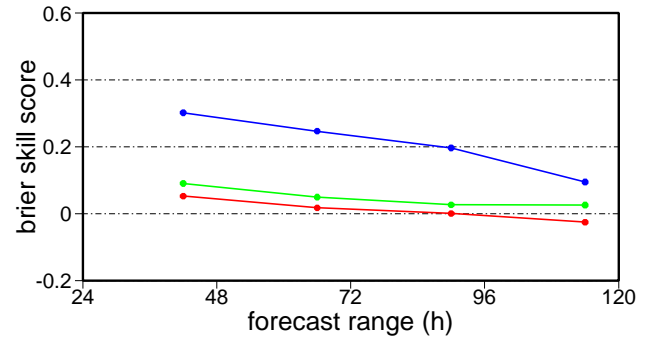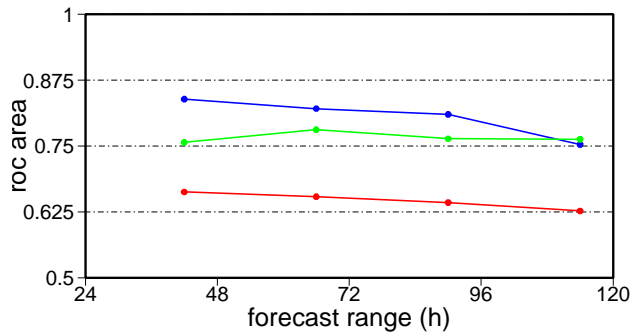

Fig. 10. Brier Skill Score values for the precipitation threshold 20 mm/24 h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Maximum observed and forecast values over $1.5 \times 1.5^{\circ}$ boxes are compared.

order to decide if the action has to be taken, he uses the probabilistic system, but he has to decide on which probability threshold to rely. The first cross of the diagram corresponds to the hit rate and false alarm rate a user can have if he decide to take a protective action when at least one ensemble member forecasts the event, so he relies on a very low probability threshold. This situation is usually linked with cases in which the possible loss is very high (human lives) with respect to the cost of the preventive action. Finally, it has to be reminded that the very first point of both curves, with co-ordinates (1,1), is just a theoretical limit to which the curves are extrapolated. Then, the first cross has to be considered the first significant point of the curves.
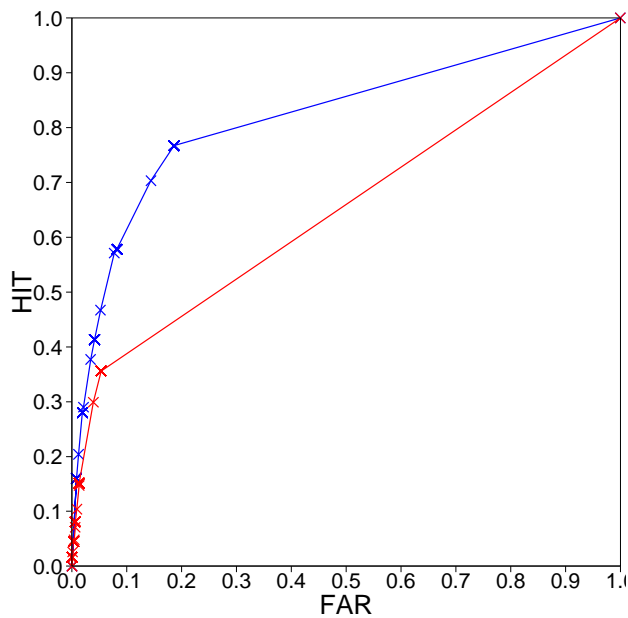
Averaging the precipitation over boxes of this size permits us to understand if the amount of precipitation over a vast region is correctly forecast, without giving information on precipitation peaks, which are very important for hydro-geological purposes. A high-resolution system can produce a significant improvement in the quantitative precipitation forecast if it is able to provide this kind of information. For this reason, a comparison in terms of precipitation maxima has been performed: the maximum forecast value falling in a box is compared with the maximum observed value in the same box. The boxes are of the same size, $1.5 \times 1.5^{\circ}$.

The BSS values for cleps (Fig. 10, blue line) are clearly higher than those relative to both the epsrm and eps51 ones, indicating that COSMO-LEPS is better able to forecast the occurrence of high precipitation over an area. The system has a positive skill for all the forecast ranges. In terms of ROC area, Fig. 11, cleps still has the highest values, while the distance between eps51 and epsrm increases. In terms of both indices, the cleps performance worsens with increasing time range. At the +114 h forecast range, cleps has the same ROC area value as eps51, while in terms of BSS it still shows some improvement.

The ROC Curves relative to COSMO-LEPS and to the 5-RM EPS are shown for the event "maximum precipitation exceeding 20 mm/24 h", for the forecast ranges +66 h (Fig. 12) and +90 h (Fig. 13). The "cleps" curves (blue curves) are
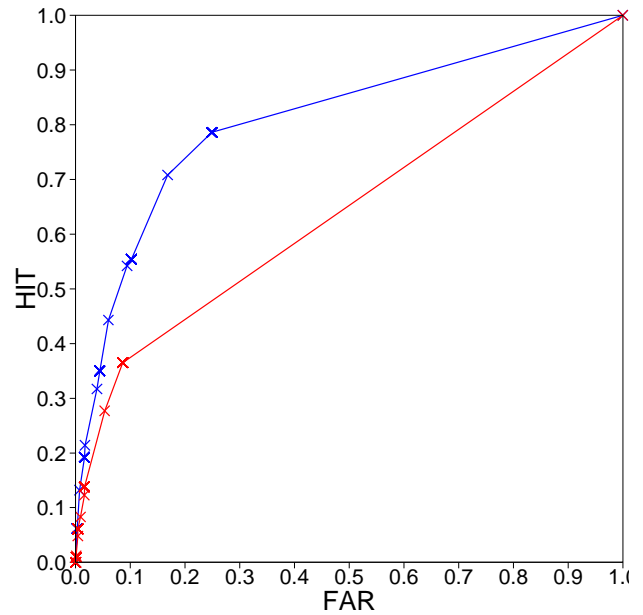
Fig. 11. ROC area for the precipitation threshold 20 mm/24 h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is relative to eps51. Maximum observed and forecast values over $1.5 \times 1.5°$ boxes are compared.

Fig. 13. Maximum values: ROC Curves for the precipitation threshold 20 mm/24 h and for the +90 h forecast range. The blue line is relative to cleps while the red line is relative to epsrm.

5 Parallel suite with different convection schemes

The COSMO-LEPS members are differentiated only by their initial and boundary conditions, which come from different members of the ECWMF EPS. More spread can easily be added by nesting LM more than once on each of the selected EPS RMs, changing the model configuration. This approach has been followed for an experimental version of COSMO-LEPS, where the number of members has been increased by integrating a pair of LM runs for each set of initial and boundary conditions, the twin runs being different only with regard to the convection scheme. This was decided to perturb the model in the description of processes responsible of precipitation generation. The two convection schemes implemented in Lokal Modell are the Tiedtke scheme (used operationally, Tietdke, 1989) and the Kain-Fritsch scheme (Kain and Fritsch, 1990). Both are mass-flux parametrisation schemes, where it is essential to estimate the convective mass flux at the base of the cloud. The Tiedtke scheme uses a dynamical closure that relates the cloud base mass flux to the large scale forcing and atmospheric structure, while in the Kain-Fritsch scheme convective effects are assumed to remove convective available potential energy within a predefined timed period.

Starting from September 2003 to May 2004, a second suite was run, parallel to the standard one. In the standard suite the Tiedtke scheme was used for the parametrisation of the convection, while in the parallel suite the Kain-Fritsch scheme was used. The two systems are here referred to as "Tiedtke suite" and "Kain-Fritsch suite".

A 10-member COSMO-LEPS can also be obtained by simply joining the two suites, forming a system in which

Fig. 12. Maximum values: ROC Curves for the precipitation threshold 20 mm/24 h and for the +66 h forecast range. The blue line is relative to cleps while the red line is relative to the epsrm.

well above the "epsrm" ones (red curves) for both forecast ranges. Considering the first cross from the top right in the diagrams (correspondent to the probabilistic issue "at least one ensemble member is forecasting the event"), the relationship between Hit Rate and False Alarm Rate of the two systems can be easily understood. At the +66 h (+90 h) forecast range, COSMO-LEPS produces more false alarms, FAR being around 0.2 (0.25) for cleps and 0.05 (0.1) for epsrm, but it has a more than double Hit Rate, HIT being around 0.75 (0.8) for cleps and 0.35 (0.35) for epsrm.

This analysis has been repeated for higher maximum precipitation thresholds, up to 50 mm/24 h (not shown), indicating that the COSMO-LEPS system has some skill in forecasting high precipitation peaks not reproduced by the lower resolution ensemble.
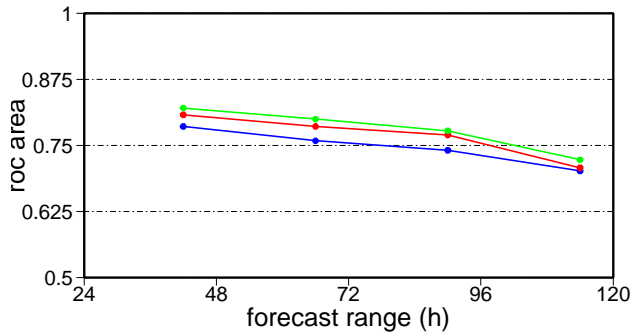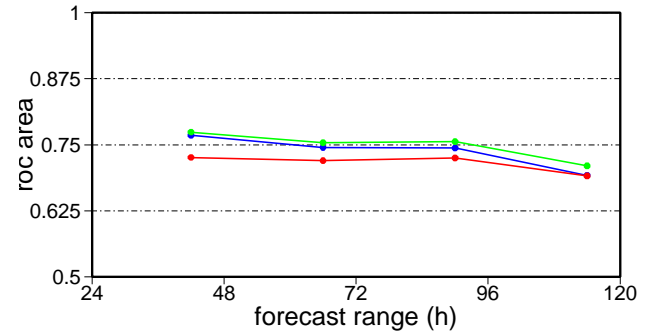
Fig. 14. ROC area values as a function of the forecast range for averaged precipitation over 0.5×0.5° boxes exceeding 20 mm/24 h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain-Fritsch suite and the green line is for the 10-member combined suite.

Fig. 16. ROC area values as a function of the forecast range for maximum precipitation over 0.5×0.5° boxes exceeding 50 mm/24 h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain-Fritsch suite and the green line is for the 10-member combined suite.
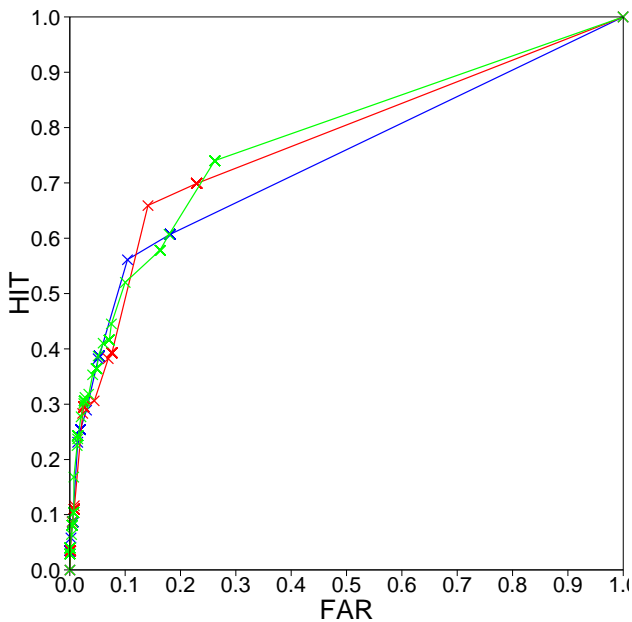
Fig. 15. ROC Curves for the precipitation threshold 20 mm/24 h (average values) and for the +90 h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain-Fritsch suite and the green line is for the 10-member combined suite.

perturbations in the model are added to the usual perturbations in the initial and boundary conditions. This system is referred to as "combined suite".

A comparison of the three suites is made in terms of 24-h precipitation using observed data from a network of rain-gauges covering Northern Italy (about 500 stations). The comparison is made over boxes of 0.5×0.5° that cover this area. The average (maximum) of the forecast values falling in each box is compared with the average (maximum) of the observed values falling in the same box.

In Fig. 14 the ROC area values computed in terms of average values over 0.5×0.5° boxes are shown. The precipitation

threshold is 20 mm/24 h. In terms of ROC area, the Kain-Fritsch suite (red line) improves with respect to the Tiedtke suite (blue line). The score of the combined suite (green line) is a little higher than both 5-member suites, but it is very similar to the Kain-Fritsch suite. This seems to suggest that adding this kind of model perturbations without also changing the initial and boundary conditions is not very useful, the spread added by using two different convection scheme being much lower that the other. In order to show more clearly the differences between the 2 suites, the ROC diagram at the +90 h forecast range is reported in Fig. 15.

Looking at the first crosses from the top right corner (low probability classes), the Hit Rate of the Kain-Fritsch suite is rather higher than that of the Tiedtke suite, while only a small increase in terms of False Alarm Rate is shown.

When verification is repeated in terms of maximum values over the same boxes, different results are obtained. Considering maximum values over boxes, a higher precipitation threshold (50 mm/24 h) is chosen for this analysis. As shown in Fig. 16, higher ROC area values are relative to the Tiedtke suite (blue line), but the difference with the Kain-Fritsch suite becomes narrower for increasing forecast range. The combined suite line (green) still provides the best score, although by a narrow margin.

Considering the ROC Curves at the +90 h forecast range (Fig. 17), it appears that the small difference between the two suites is due to a little increase in terms of Hit Rate for the Tiedtke suite, while the False Alarm Rates are almost identical.

These results lead to a third modification of the COSMO-LEPS suite from June 2004, in addition to the two described at the end of Sect. 3: the 10 Lokal Modell runs are performed by using both the Tiedtke and Kain-Fritsch schemes for the parametrisation of the convection. The scheme used within each single run is randomly selected.

6 Conclusions

The key role played by mesoscale and orographic-related processes can seriously limit the predictability of intense and localised precipitation events. Deterministic limited-area models are improving the forecast of locally intense events in the short range, but it is still difficult to forecast accurately their space-time evolution, especially for ranges longer than 48 h. The high-resolution system COSMO-LEPS has been designed as a tool for the prediction of heavy precipitation in a probabilistic perspective. On a case study basis (Montani et al., 2003b; Marsigli et al., 2004), the system has proved to be successful in the prediction of intense rainfall events. An objective probabilistic verification is being carried out at ARPA-SIM so as to assess both the abilities and shortcomings of the system, to address future developments of the system and to provide guidelines to the users.

In this paper, an objective verification of the COSMO-LEPS performances in forecasting precipitation has been presented. The period considered is September–November 2003 and forecast precipitation cumulated over 24 h is compared with observed data.

In order to quantify the added value provided by the mesoscale probabilistic system, COSMO-LEPS has been compared with the ECMWF EPS. To make an appropriate comparison of the two systems, the differences in the number of ensemble members (5 for COSMO-LEPS and 51 for the EPS) and in the horizontal resolution (10 km for COSMO-LEPS and 80 km for the EPS) have been considered. As far as the population of the ensembles is concerned, the reduced EPS made up of the 5 Representative Members has also been analysed, allowing to quantify the impact of the increased resolution alone. On the other hand, the problem of the very different resolutions of the two systems is tackled by upscaling both systems to a lower resolution: the grid point forecasts of both model are aggregated over boxes of $1.5 \times 1.5°$. The observations are also aggregated over the same boxes.

A comparison in terms of average precipitation values over $1.5 \times 1.5°$ boxes shows that EPS performs better. Nevertheless, COSMO-LEPS outperforms the reduced EPS composed of the 5 Representative Members in terms of ROC area, in particular showing a higher Hit Rate. When the comparison is carried out in terms of maximum values over boxes of the same size, COSMO-LEPS scores are the highest, in terms of both Brier Skill Score and ROC area. This is due to the capability of the mesoscale system to forecast high precipitation values. The analysis of the ROC Curves shows that this improvement is not associated to a dramatic increase of the false alarms.

Considering average precipitation over a quite large area, EPS is performing better than the mesoscale system. Nevertheless, it is worth noting that meteo-hydrological applications often requires information on a more local scale. The skill of the COSMO-LEPS system in forecasting the occurrence of precipitation maxima it is a clear indication of the usefulness of the system in forecasting intense and localised events.
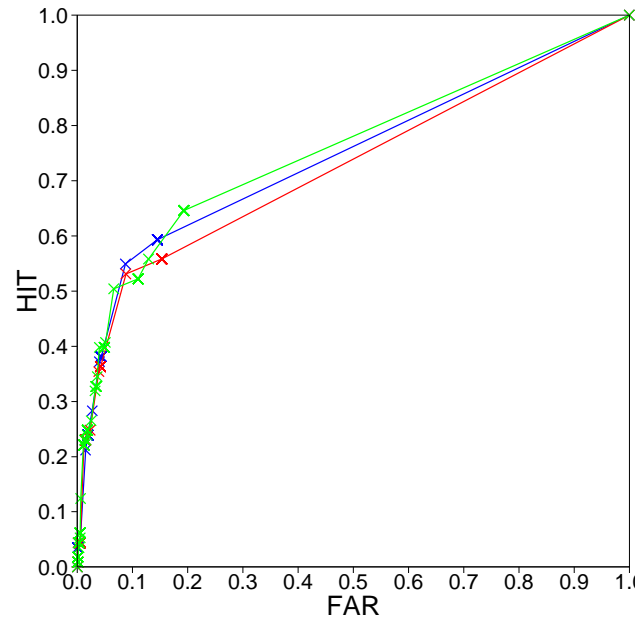
Fig. 17. ROC Curves for the precipitation threshold 50 mm/24 h and for the +90 h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain-Fritsch suite and the green line is for the combined suite.

Finally, COSMO-LEPS scores are worsening at the 5 day forecast range, leading to a reduction of the improvement with respect to the global systems. This could be linked to the increase of the limited-area model error, which, at this time range, becomes large enough to overwhelm the improvement produced by LM in terms of predictability of mesoscale structures.

An analysis of the COSMO-LEPS methodology, leading to the choice of the Representative Members, has been also shown, addressing the topics of the super-ensemble size (1 EPS, 2 EPS, 3 EPS) and of the number of Representative Members (5 or 10). The analysis has been performed by verifying the precipitation forecast by the reduced EPS made up by the RMs without running the nested LAM. The forecast are compared with ECMWF proxy rain over the same season (Autumn 2003).

We showed that the use of just two EPS in the super-ensemble seems to be a suitable compromise between the need to decrease the percentage of outliers and the need to maintain a high skill. Furthermore, doubling the number of Representative Members (from 5 to 10) produces the greatest improvement of the skill.

These results lead to a modification of the COSMO-LEPS methodology at the beginning of June 2004: the super-ensemble has been built by using only the 2 most recent EPS and the number of clusters has been fixed to 10, nesting Lokal Modell on each of the so selected 10 RMs. The 10 Lokal Modell runs are performed by using both the Tiedtke and Kain-Fritsch schemes for the parametrisation of the convection. The scheme used within each single run is randomly

selected. This choice was adopted since we did not see any significant difference when using either the Tiedtke scheme or the Kain-Fritsch scheme. Actually, an attempt was made to increase the number of COSMO-LEPS members without increasing the number of Representative Members. A pair of LM runs were performed for each set of initial and boundary conditions over a test period, the twin runs being different only in the scheme adopted for precipitating convection. Results show that the use of the two different convection scheme did not produce a significant increase of the spread of the system.

The new configuration of COSMO-LEPS system is being currently verified and the impact of the modifications on the system performance is addressed. The verification methodology is still under development, focusing on the proper way of performing a probabilistic verification of an high-resolution forecast system.

Edited by: Z. Toth
Reviewed by: A. Horanyi, S. L. Mullen and another referee

References

Buizza R.: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system, Monthly Weather Review, 125, 99–119, 1997.

Buizza, R., Miller, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, Quart. J. Roy. Meteor. Soc., 125, 2887–2908, 1999.

Houtekamer, P. L., Derome, J., Ritchie, H., and Mitchell, H. L.: A system simulation approach to ensemble prediction, Monthly Weather Review, 124, 1225–1242, 1996.

Kain, J. S. and Fritsch, J. M.: A one-dimensional entraining/detraining plume model and its application in convective parametrisation, J. Atmos. Sci., 47, 2784–2802, 1990.

Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S., Molteni, F., and Buizza, R.: A strategy for High-Resolution Ensemble Prediction, Part II: Limited-area experiments in four Alpine flood events, Quart. J. Roy. Meteor. Soc., 127, 2095–2115, 2001.

Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T.: Probabilistic high-resolution forecast of heavy precipitation over Central Europe, Nat. Hazards Earth Sys. Sci, 4, 315–322, 2004,
SRef-ID: 1684-9981/nhess/2004-4-315.

Mason, S. J. and Graham, N. E.: Conditional probabilities, relative operating characteristics and relative operating levels, Weather and Forecasting, 14, 713–725, 1999.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF Ensemble Prediction System: Methodology and validation, Quart. J. Roy. Meteor. Soc., 122, 73–119, 1996.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T.: A strategy for High-Resolution Ensemble Prediction, Part I: Definition of Representative Members and Global Model Experiments, Quart. J. Roy. Meteor. Soc., 127, 2069–2094, 2001.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., and Buizza, R.: Performance of ARPA-SMR Limited-area Ensemble Prediction System: two flood cases, Nonlin. Processes Geophys., 8, 387–399, 2001,
SRef-ID: 1607-7946/npg/2001-8-387.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi, S., and Buizza R.: The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts, Nonlin. Processes Geophys., 10, 261–274, 2003a,
SRef-ID: 1607-7946/npg/2003-10-261.

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

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parametrisation in large-scale models, Mon. Wea. Rev., 117, 1779–1800, 1989.

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

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