# The COSMO–LEPS ensemble system: validation of the methodology and verification

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

ARPA–SIM, Bologna, Italy

Manuscript submitted to

**Nonlinear Processes in Geophysics**

Manuscript-No. 12345

**Offset requests to:**
C. Marsigli
ARPA-SIM, V.le Silvani 6
Bologna, Italy

Nonlinear Processes in Geophysics (2004) 0000: 0001–9
SRef-ID: 1607-7946/npg/2004-0000-0001

# The COSMO–LEPS 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 hours 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 beneficial of increasing the number of ensemble member. The system has been designed to have a probabilistic support at the mesoscale, focussing 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 with respect to the global ensemble system in the forecast of high precipitation values. The impact of the use of different limited–area model configurations is also assessed, showing that the impact of introducing perturbations in the limited–area model configuration plays a minor role with respect to the initial and boundary condition perturbations.

## 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 sometimes difficult to forecast accurately their space–time evolution, especially for ranges longer than 48 hours. 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 $\mathrm{km}$ at most), making difficult their use when the forecast of severe and localised weather events is concerned. As regards 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), Montani et al. (2003a)), which after some tests led to the COSMO–LEPS implementation (Montani et al., 2003b). COSMO (COnsortium for Small–scale MOdelling, web site: www.cosmo–model.org) is a consortium involving Germany, Italy, Switzerland, Greece and Poland which aims to develop, improve and mantain the non–hydrostatic limited–area model Lokal Modell (LM).

The LEPS methodology allows 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 select 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. The ensemble size reduction is necessary in order to render affordable the limited–area model runs within a few hours. This permits an evaluation of the COSMO–LEPS performances by the forecasters in real time. The transfer of information to the mesoscale can be viewed as a dynamical zoom of the forecast provided by the probabilistic system. 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

---

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

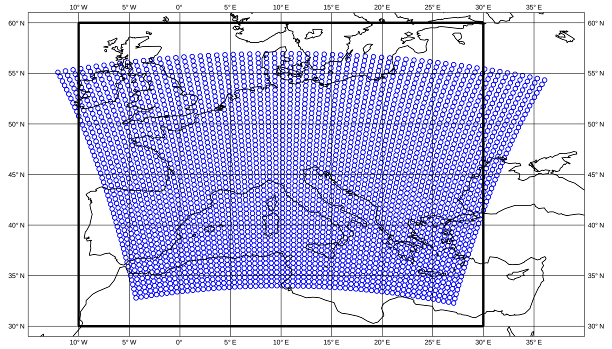

**Fig. 1.** COSMO–LEPS operational domain (small circles) and clustering area (thick rectangle).

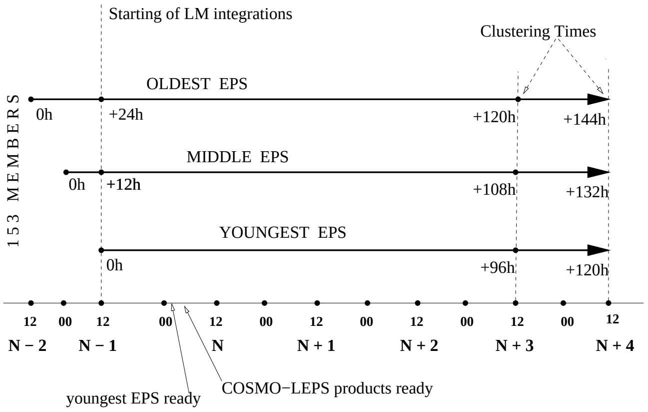

**Fig. 2.** Details of the COSMO–LEPS suite.

forecast ranges (48–120 hours), LEPS performs better than EPS concerning the quantitative forecast of intense precipitation, as well as the geographical localisation of the regions most likely to be affected by the flood events.

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 recently started on the ECMWF computer system under the auspices of COSMO (Montani et al., 2003b). COSMO–LEPS aims at the development and pre–operational test of a “short to medium–range” (48–120 hours) probabilistic forecasting system using a LAM over a domain covering all countries involved in COSMO (Fig. 1).

An objective verification of COSMO–LEPS is being carried out at ARPA–SIM, focussing 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, ...). 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), Percentage of Outliers (Buizza , 1997). As regards the system configuration, the analysis is focussing on the methodology that leads to the choice of the Representative Members. This analysis has been performed on the same period.

The paper is organised as follows: in Section 2 the COSMO–LEPS system is described, as it has been since June 2004, while in Section 3 a statistical analysis of the methodology is presented, leading to a new configuration of the system. In Section 4 an objective verification of the performance of COSMO–LEPS is carried out, comparing the system with the ECMWF EPS. In Section 5, the COSMO–LEPS is compared with a parallel suite in which another scheme for the parametrisation of the deep convection is used. Finally, conclusions are drawn in Section 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. The Ensemble Prediction System (EPS) is now based on a $T_L$255L40 model (spectral model with truncation at wavenumber 255 and 40 vertical levels), corresponding to a horizontal resolution of about 80 $\mathrm{km}$, and has 51 members (Molteni et al. (1996), Buizza et al. (1999)). Three successive 12–hour–lagged EPS runs (started at 12 UTC of day *N–2*, at 00 and 12 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 $\mathrm{hPa}$) and at two forecast times (fc+96 and fc+120 for the “youngest” EPS, the one started at 12 UTC of day *N–1*); the cluster domain covers the region 30N–60N, 10W–40E (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 hours, always starting at 12UTC of day *N–1* and ending at 12UTC of day $N + 4$. The LM is run with a horizontal resolution $\Delta x \simeq 10\ \mathrm{km}$ and with 32 levels in the vertical; the time–step used for the integrations is 60 $\mathrm{s}$.

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 to increase the ensemble spread and to have 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 Selection Algorithm on the most recent EPS (referred to “1–EPS”)

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 hours 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. Smilar 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 to decrease the percentage of outliers significantly but leading only to 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 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) has 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.

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 the population of the ensembles is concerned, COSMO–LEPS is compared also with the small EPS ensemble made up by the 5 Representative Members. This permits 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

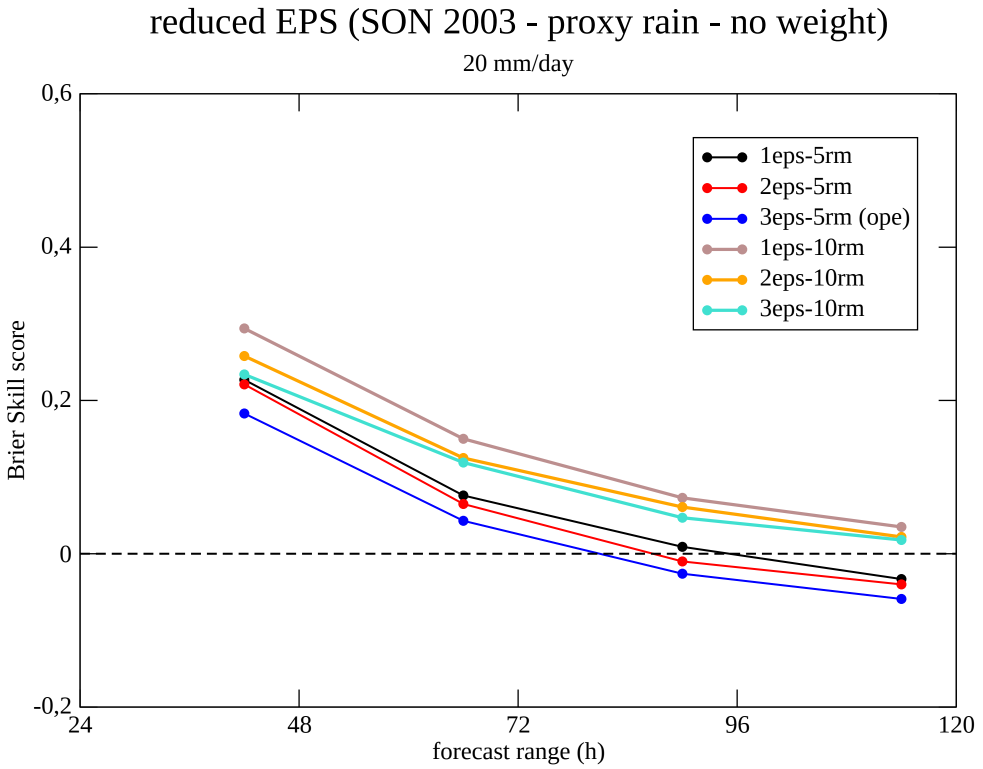

**Fig. 3.** Brier Skill Score as a function of the forecast range for the event precipitation exceeding 20mm/24h relative to the RM EPS. The different configurations are: 5 clusters algorithm based on 1 EPS (black line), on 2 EPS (red line) and on 3 EPS (operational configuration, blue line); 10 clusters algorithm based on 1 EPS (gray line), on 2 EPS (orange line) and on 3 EPS (cyan line).

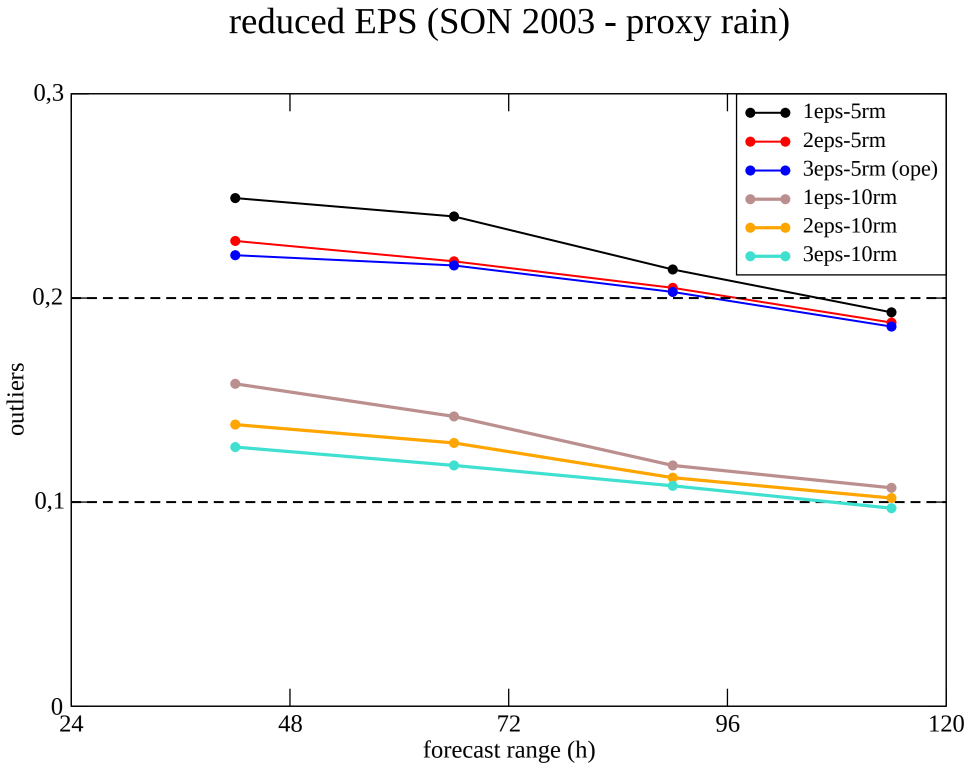

**Fig. 4.** Percentage of outliers for the RM EPS. The different configurations are: 5 clusters algorithm based on 1 EPS (black line), on 2 EPS (red line) and on 3 EPS (operational configuration, blue line); 10 clusters algorithm based on 1 EPS (gray line), on 2 EPS (orange line) and on 3 EPS (cyan line).

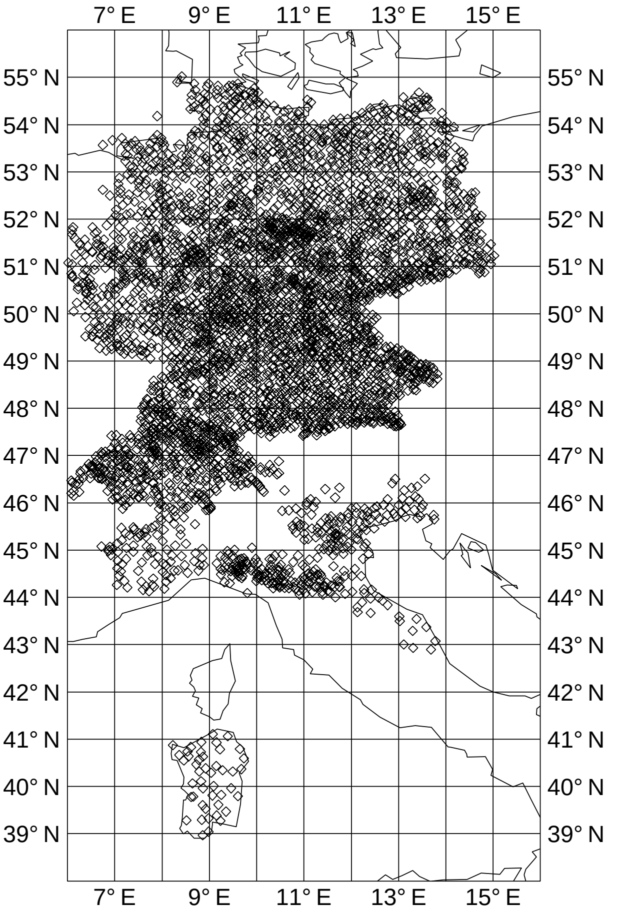

**Fig. 5.** Network of station providing 24–hour precipitation (06 UTC to 06 UTC) for the COSMO verification.

lower resolution: the grid point forecasts of both model are averaged over boxes of 1.5 x 1.5 degrees. The comparison is made in terms of 24–hour precipitation, against observed data from a very dense network of raingauges. Precipitation is cumulated from 06 to 06 UTC. In order to properly compare forecast values on grid points and observed values on station points, the observations within a box are averaged and the obtained value is compared directly with the averaged forecast value. The comparison is carried out over a large area included in the COSMO–LEPS domain, covering Germany, Switzerland and Northern Itlay. 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 of horizontal resolution, referred to as “cleps”;
- the EPS mini–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”;

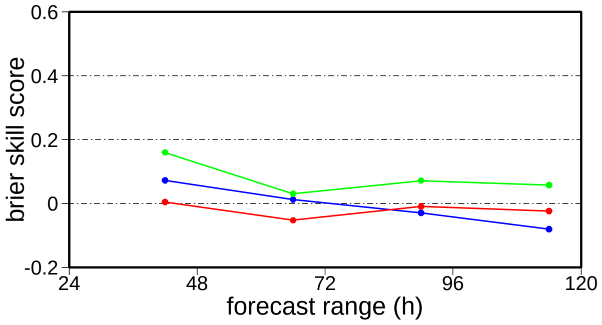

**Fig. 6.** Brier Skill Score values for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Averaged observed and forecast values over 1.5 x 1.5 degree boxes are compared.

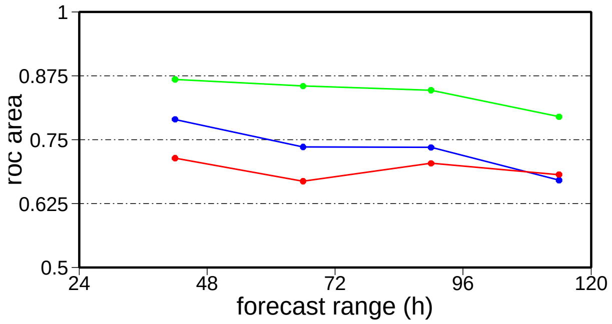

**Fig. 7.** ROC area for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Averaged observed and forecast values over 1.5 x 1.5 degree boxes are compared.

In Fig. 6 and Fig. 7 the Brier Skill Score and the ROC area are shown for the three systems (for both indices, the higher the better). 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 event considered here is precipitation exceeding 20 $\mathrm{mm}$ / 24 $\mathrm{h}$ 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 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, that is its performance is slightly better. The difference between cleps and epsrm is slighlty in favour of cleps for the first forecast ranges, while the reverse it is true at the +114 forecast range.

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

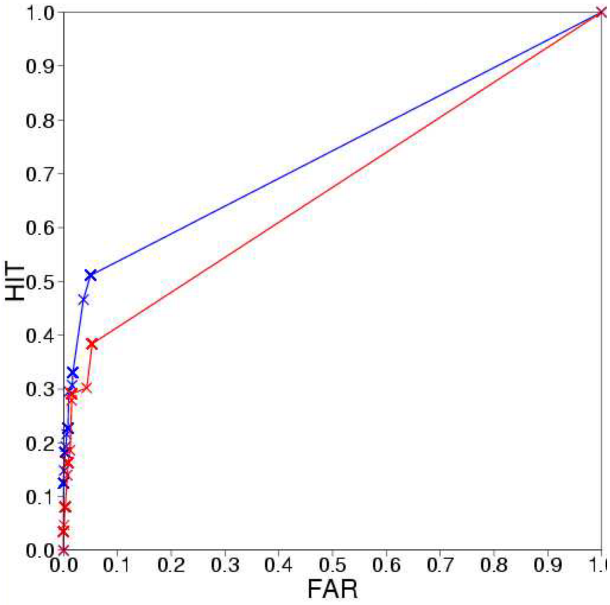

**Fig. 8.** ROC Curves for the precipitation threshold 20mm/24h and for the +66h forecast range. The blue line is relative to cleps while the red line is relative to the epsrm.

system (cleps, blue line) has lower scores, but higher that those of the 5–RM EPS (epsrm, red line). 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 "precipitation exceeding 20mm/24h", for the forecast ranges +66 hours (Fig. 8) and +90 hours (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 diagram, it is evident that the two systems have comparable False Alarm Rate, but COSMO–LEPS obtains higher Hit Rate values. This cross is corrispondent to the probabilistic issue "at least one ensemble member is forecasting the event".

Averaging the precipitation over boxes of this size permits 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 produces 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 x 1.5 degrees.

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 more able to correctly fore-

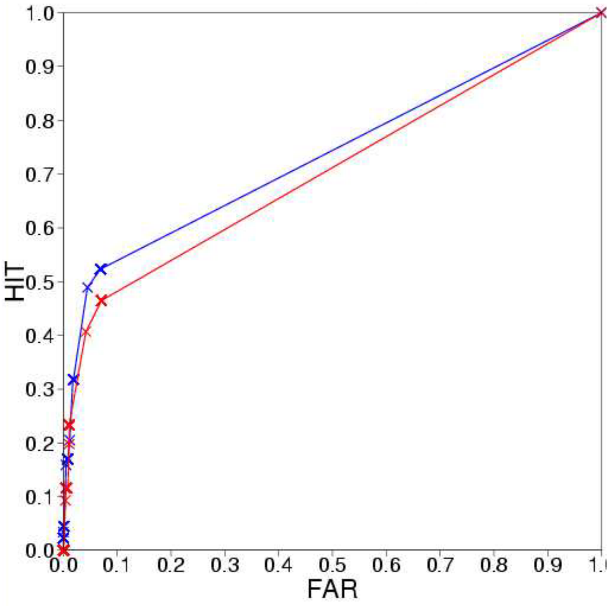

**Fig. 9.** ROC Curves for the precipitation threshold 20mm/24h and for the +90h forecast range. The blue line is relative to cleps while the red line is relative to epsrm.

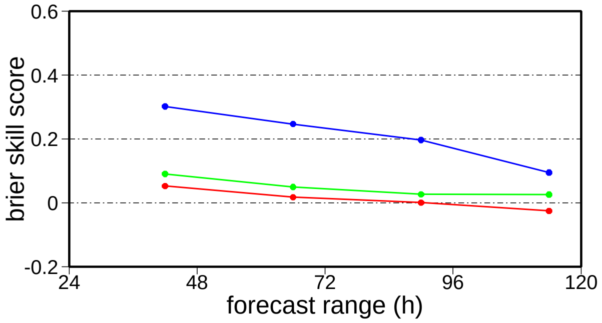

**Fig. 10.** Brier Skill Score values for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Maxima observed and forecast values over 1.5 x 1.5 degree boxes are compared.

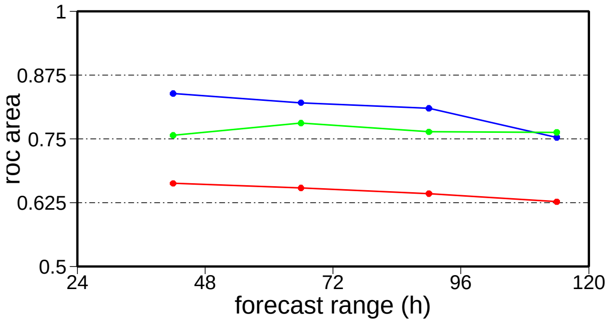

**Fig. 11.** ROC area for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is relative to eps51. Maxima observed and forecast values over 1.5 x 1.5 degree boxes are compared.

cast high precipitation values over a rather big area. In terms of ROC area, Fig. 11, cleps still has the highest values, while the distance between eps51 and epsrm increases.

## 5 Parallel suite with different convection scheme.

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 be easily added by nesting LM with slightly different configurations in each of the selected EPS RMs. This is an attempt to increase the number of COSMO–LEPS members by just integrating a pair of LM runs for each set of initial and boundary conditions, the twin runs being different only in the scheme used for the parametrisation of the convection. 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 it was used the Kain–Fritsch scheme. 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 perturbations in the model are added to the usual perturbations in the initial and boundary conditions. This system is here referred to as “combined suite”.

A comparison of the three suites is made in terms of 24–hour precipitation using observed data from a network of raingauges covering Northern Italy (about 500 stations). The comparison is made over boxes of 0.5 x 0.5 degrees that covers this area. The average (maximum) of the forecast values falling in each box are compared with the average (maximum) of the observed values falling in the same box.

In Fig. 12 the ROC area values computed in terms of average values over 0.5 x 0.5 degrees boxes are shown. The precipitation threshold is 20mm/24h. 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 one of the Kain–Fritsch suite. This seems to suggest that adding this kind of model perturbations without changing also 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 +90h forecast range is reported in Fig. 13.

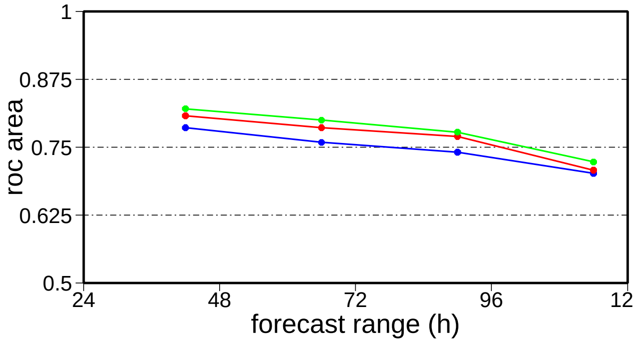

**Fig. 12.** ROC area values as a function of the forecast range for averaged precipitation over 0.5 x 0.5 degrees boxes exceeding 20mm/24h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

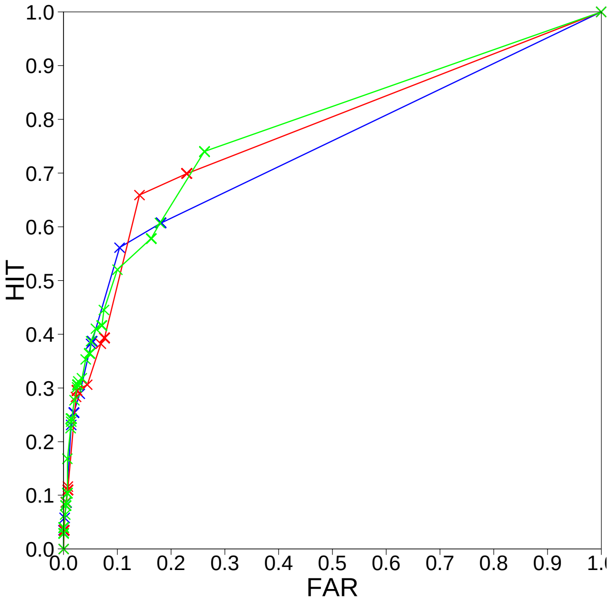

**Fig. 13.** ROC Curves for the precipitation threshold 20mm/24h (average values) and for the +90h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

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

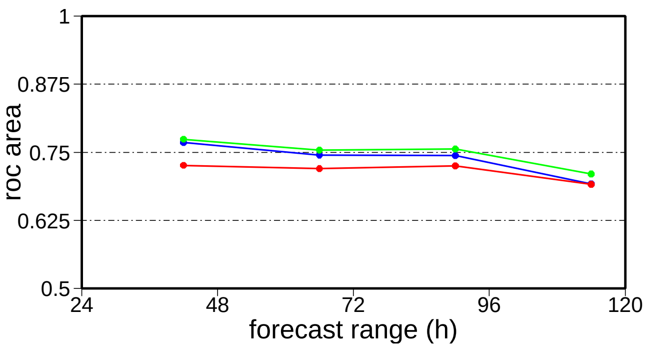

**Fig. 14.** ROC area values as a function of the forecast range for maximum precipitation over 0.5 x 0.5 degrees boxes exceeding 50mm/24h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

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 (50mm/24h) is chosen for this analysis. As shown in Fig. 14, higher ROC area values are relative to the Tiedtke suite (blue line), but the difference between the two is narrowing 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 +90h forecast range (Fig. 15), 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. 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 hours. The high–resolution system COSMO–LEPS has been designed in order to have a tool for the prediction of heavy precipitation in a probabilistic environement. On a case study basis, it has been proved to be successful in the prediction of intense rainfall events (Montani et al. (2003b), Marsigli et al. (2004)). An objective probabilistic verification is being carried out at ARPA–SIM so as to assess both the abilities and shortcomings of the system, to ameliorate it and to provide guidelines to the end users.

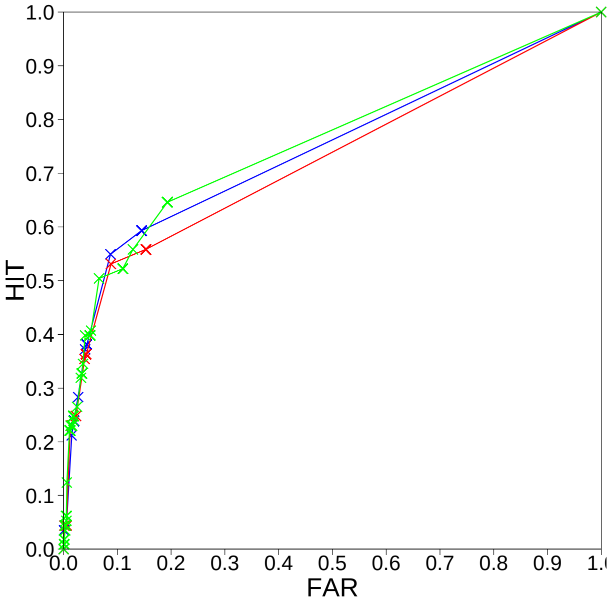

**Fig. 15.** ROC Curves for the precipitation threshold 50mm/24h and for the +90h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the combined suite.

In this paper, an objective verification of the COSMO–LEPS performances in forecasting precipitation is presented. The period considered is September–November 2003 and 24–hour cumulated precipitation is compared with observed data. As for the system configuration, an analysis of the methodology leading to the choice of the Representative Members was also performed over the same period.

In order to quantify the added value provided by the mesoscale probabilistic system, COSMO–LEPS has been compared with the EPS. In order to properly compare the two systems, both 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) have been taken into account. As the population of the ensembles is concerned, COSMO–LEPS is compared also with the small EPS ensemble made up by the 5 Representative Members, permitting 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 model are averaged over boxes of 1.5 x 1.5 degrees. In order to properly compare forecast values on grid points and observed values on station points, the observations within a box are averaged and the obtained value can be compared directly with the averaged forecast value. A comparison in terms of average values over 1.5 x 1.5 degrees boxes shows that EPS is performing better has regards the average amount of precipitation falling over a wide area.

Nevertheless, COSMO–LEPS outperformes the EPS made up by 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. It is important to underline that the system is not producing too many false alarms.

The analysis of the methodology seem to indicate that the use of just two EPS in the super–ensemble can be a suitable compromise between the need to decrease the percentage of outliers and the need to mantain a high skill. Furthermore, it appears that doubling the number of Representative Members produce a major improvement of the skill.

An attempt has been made to increase the number of COSMO–LEPS members by integrating a pair of LM runs for each set of initial and boundary conditions, the twin runs been different only in the scheme used for the parametrisation of the convection. This proved to have very little impact of the system performance, when compared to both 5–member suites. Furthermore, COSMO–LEPS performances do not change significantly when a different scheme is used.

These results lead to three 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, the number of clusters has been fixed to 10, nesting Lokal Modell on each of the so selected 10 RMs and 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.

*Acknowledgements.* This work was partially sponsored by the GNDCI (Gruppo Nazionale Difesa Catastrofi Idrogeologiche) of the CNR (National Research Council). The authors thank the COSMO partners for the data provided and Ulrich Damrath for having put them on a common format.

## 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., Stocastic 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.

Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S., Molteni, F., 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., Paccagnella, T., Probabilistic high–resolution forecast of heavy precipitation over Central Europe, *Natural Hazards and Earth System Sciences*, *4*, 315–322, 2004.

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., 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. *Nonlinear Processes in Geophysics*, *8*, 387–399, 2001.

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 ensmeble forecasts, *Nonlinear Processes in Geophysics*, *10*, 261–274, 2003a.

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

Tracton, M. S. and Kalnay, 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.