Consortium for Small-Scale Modelling

Technical Report No. 8

Evaluation of the Performance of the COSMO-LEPS System

by

C. Marsigli, A. Montani, T. Paccagnella (ARPA-SIM)
D. Sacchetti (ARPAL)
A. Walser, M. Arpagaus (MeteoSwiss)
T. Schumann (DWD)

April 2005

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_8

Deutscher Wetterdienst

MeteoSwiss

Ufficio Generale per la Meteorologia

Hellenic National Meteorological Service

Instytut Meteorologii i Gospodarki Wodnej

Il Servicio Meteorologico Regionale di ARPA

Amt für Wehrgeophysik

L'Agenzia Regionale per la Protenzione Ambientale del Piemonte

www.cosmo-model.org

Editors: G. Doms and U. Schättler, Deutscher Wetterdienst, P.O. Box 100465, 63004 Offenbach, Germany
Printed at Deutscher Wetterdienst, Offenbach am Main

Contents

1 Introduction

The "experimental–operational" limited–area ensemble prediction system COSMO–LEPS has been running daily since November 2002 at 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).

A subjective evaluation of its performance is being carried out by the forecasters of the Meteorological Centres involved in the COSMO Consortium. Furthermore, the system is being objectively evaluated by ARPA–SIM using a set of probabilistic quality indices (Marsigli et al., 2005). In this report, a summary of the performance of COSMO–LEPS is presented, collecting also contributions coming from different COSMO partners.

In the first part, results obtained through an objective verification by ARPA–SIM are presented. In Section 2, an analysis of the methodology on which the system is based is shown, focusing on the results that led to a major modification of the system in June 2004. In Section 3, the quality of the system is assessed in an objective manner, making use of probabilistic indices. In Section 4, the operational suite has been compared with a parallel suite where a different scheme for the parameterization of the deep convection was used.

For a complete description of scores used in this report, the reader is referred to Wilks (1995), Stanski et al. (1989) and Talagrand et al. (1999) for the Brier Skill Score, Mason and Graham (1999) for the ROC Curves and the ROC area, to Richardson (2000) for the Cost-loss Analysis and to Buizza (1997) for the Percentage of Outliers.

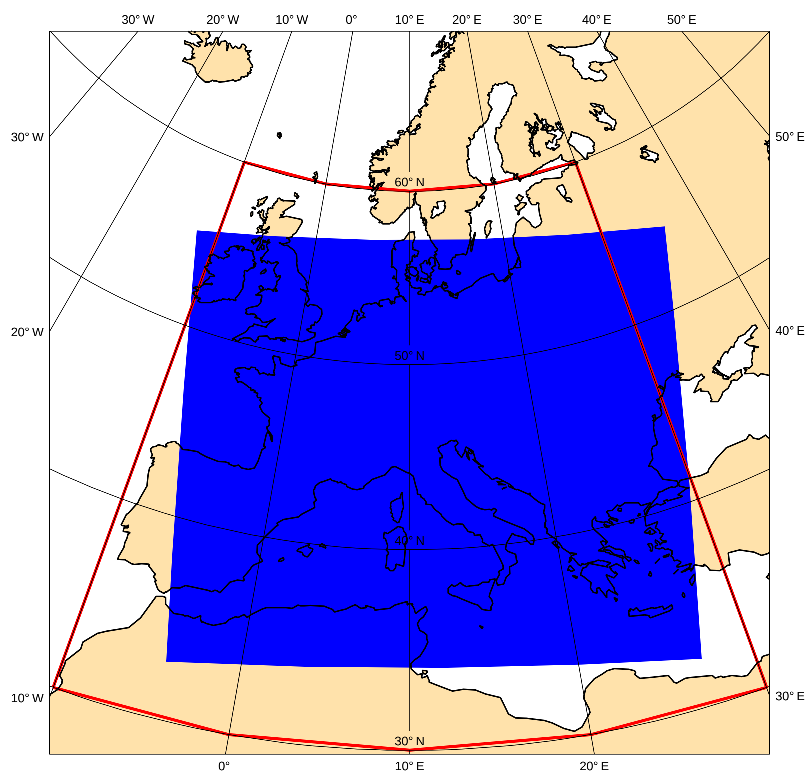


Figure 1: COSMO–LEPS operational domain (blue area) and clustering area (red frame).

In the second part, subjective evaluations of the system made by other COSMO partners are shown. In Section 5, the results from the subjective verification performed by the Regional

Agency for Prevention and Environment of the Liguria Region (ARPAL) are presented. In Section 6, the use of COSMO–LEPS by MeteoSwiss and the impressions got by the forecasters are described. In Section 7, the contribution by the forecasters of DWD is reported.

Finally, in Section 8 some conclusions are drawn.

2 Statistical analysis of the methodology

An analysis of the methodology on which COSMO–LEPS is based is presented in this section.

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 is the EPS, the less skillful are their members. 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 compared, are:

- 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. The extent to which this proxy is a good approximation for the truth is not important, because this is a comparison among different configurations 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, see Wilks, 1995) is higher when the clustering is based on the most recent EPS only (Fig. 2, 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 nearest to the one of the 1–EPS ensemble.

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. The percentage of outliers (Fig. 3) 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

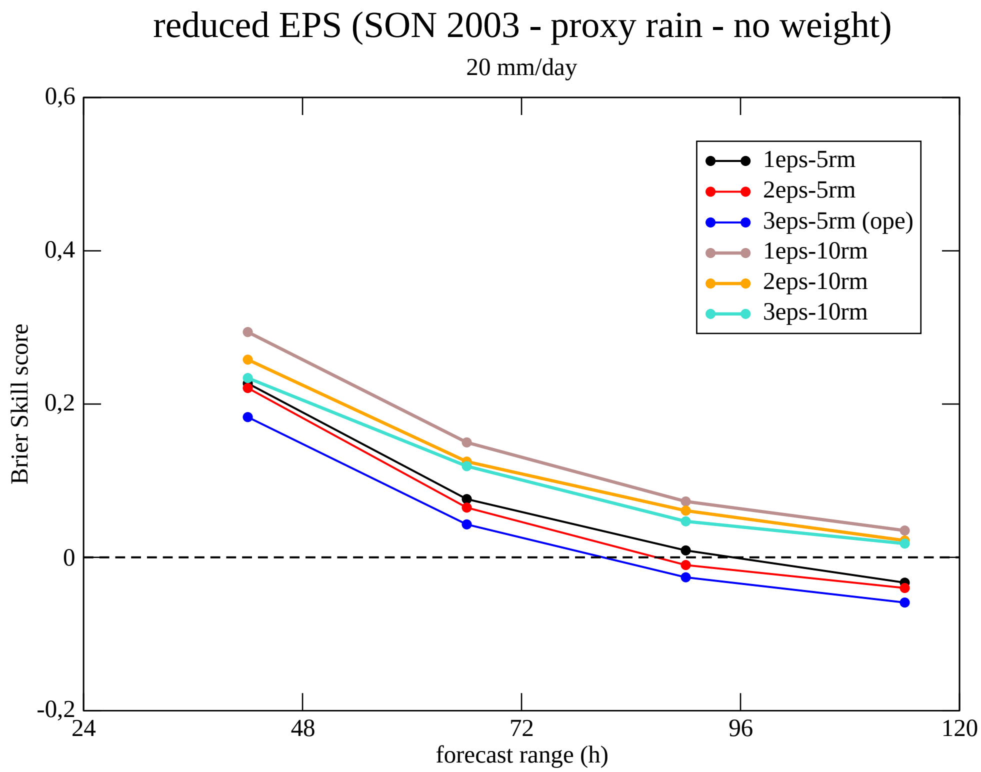


Figure 2: Brier Skill Score as a function of the forecast range for the event precipitation exceeding 20mm/24h relative to the RM EPS. The different configurations are: 5 clusters algorithm based on 1 EPS (black line), on 2 EPS (red line) and on 3 EPS (operational configuration, blue line); 10 clusters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

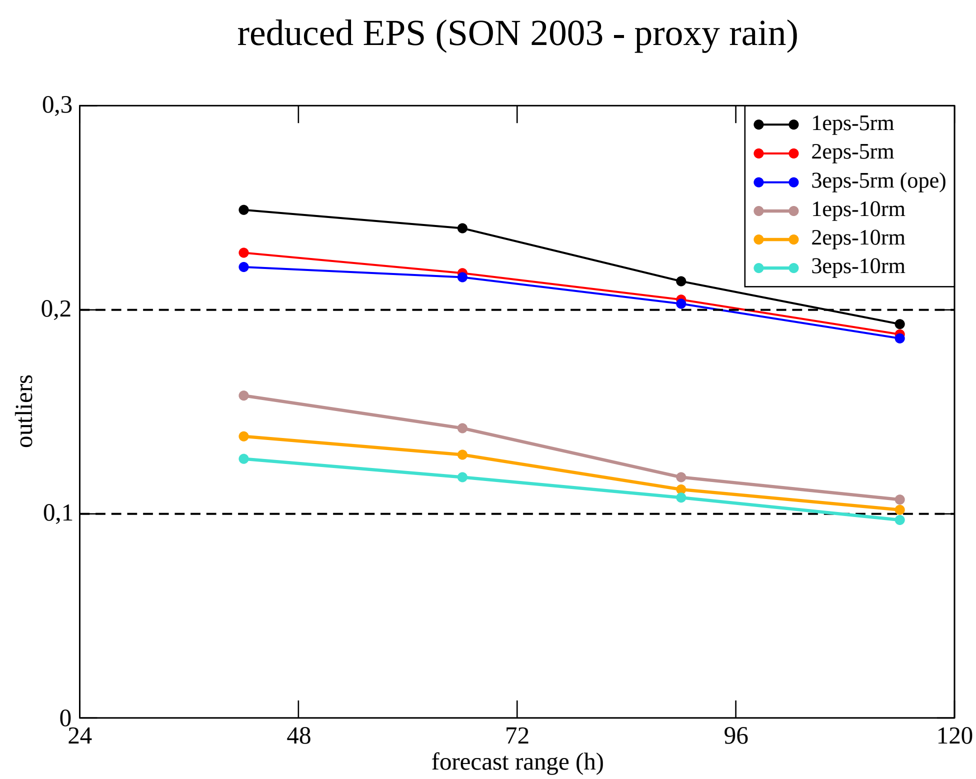


Figure 3: Percentage of outliers for the RM EPS, colours as in the previous figure.

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. 2, 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 lead to two major modifications 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 10 RMs selected.

3 Objective verification

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 lower resolution: the grid point forecasts of both model are averaged over boxes of 1.5×1.5 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 can be compared directly with the averaged forecast value. The comparison is carried out over a big area included in the COSMO–LEPS domain, covering Germany, Switzerland and Northern Italy. The very dense network of stations recording daily precipitation is shown in Fig. 4.

In order to avoid problems in the boxes at the boundaries of the domain covered by observations, an "observational mask" is used. This mask was built by assigning a value of 1 to a grid point only if its distance to at least one observation is less than 0.15 degrees; otherwise, it is assigned the value of 0. The mask relative to COSMO–LEPS and to the data–set used in this work is shown in Fig. 5. To compute the average or the maximum forecast value over a box, only grid points that are labelled 1 have been considered.

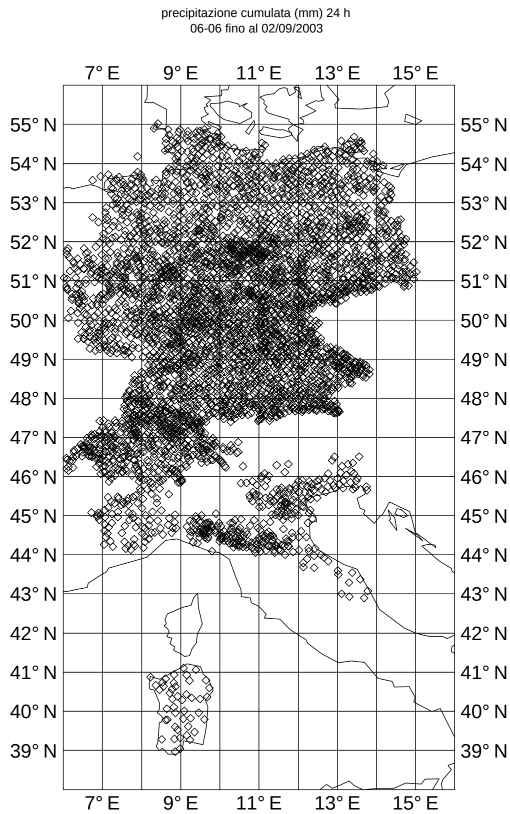


Figure 4: Network of station providing 24–hour precipitation (06 UTC to 06 UTC) for the COSMO verification.

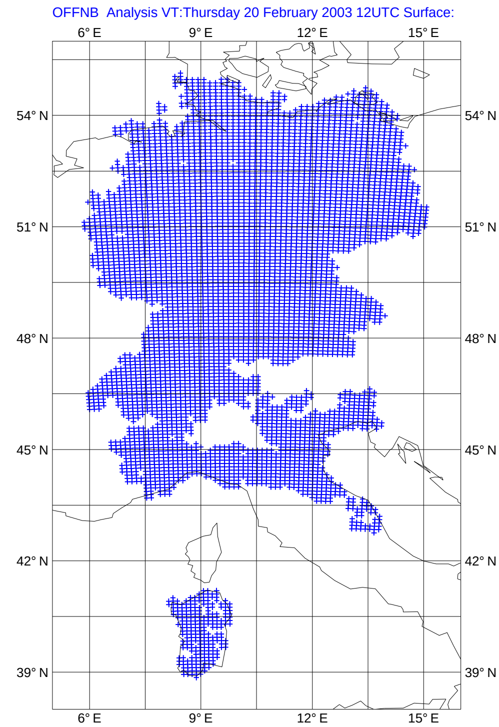


Figure 5: The "observational mask" used for COSMO–LEPS. Only grid points which are not too far from the area covered by the observations are considered.

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 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";

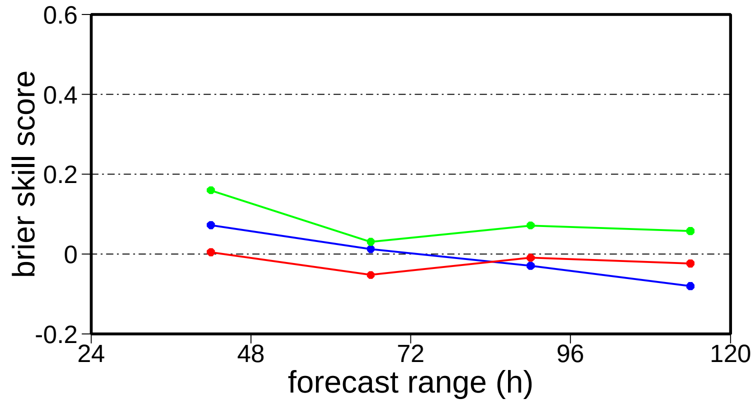


Figure 6: Brier Skill Score values for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Average observed and forecast values are compared.

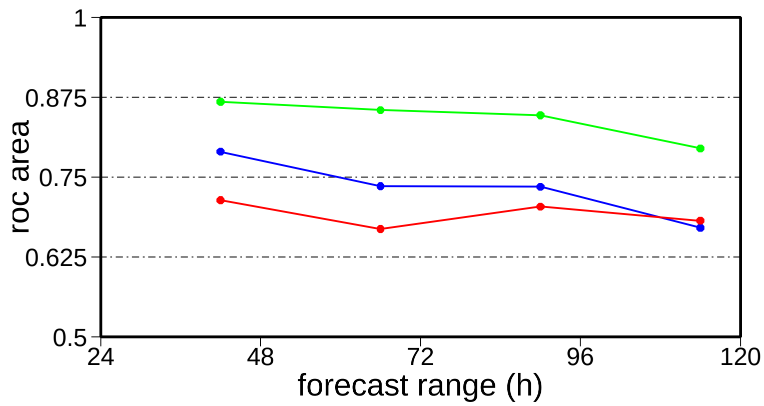


Figure 7: ROC area for the precipitation threshold 20mm/24h. The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Average observed and forecast values are compared.

In Fig. 6 and Fig. 7 the Brier Skill Score and the ROC area are shown for the three systems (for both index, 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 mm / 24 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 is individuating 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 slightly 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 system 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 time range. The COSMO–LEPS system (cleps, blue line) has lower scores, but higher than those of the 5–RM EPS (epsrm, red line).

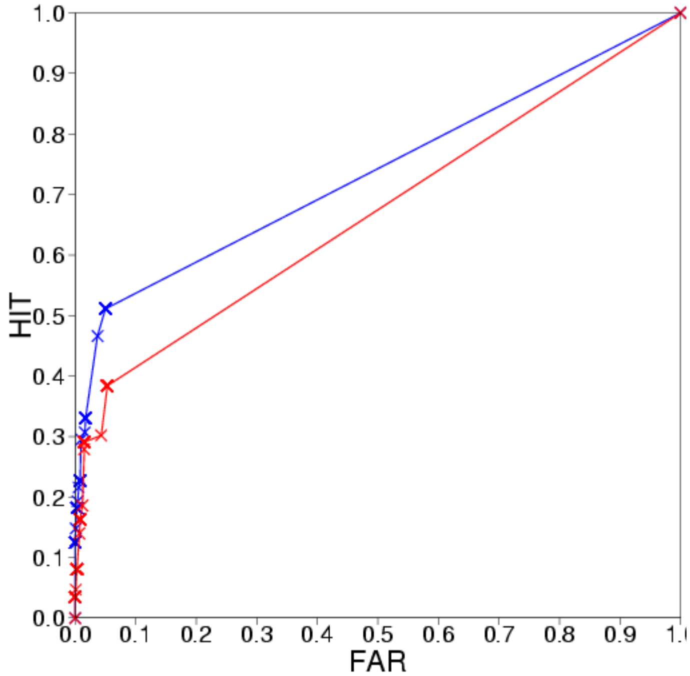


Figure 8: Average values: ROC Curves for the precipitation threshold 20mm/24h and for the +66h forecast range. The blue line is relative to cleps while the red line is relative to the epsrm.

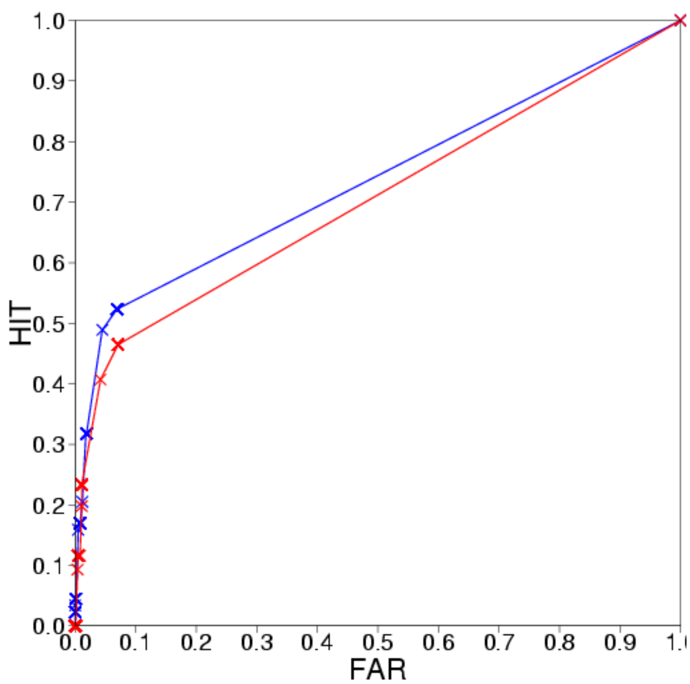


Figure 9: Average values: ROC Curves for the precipitation threshold 20mm/24h and for the +90h forecast range. The blue line is relative to cleps while the red line is relative to epsrm.

The ensemble size often plays a major role in the computation of the probabilistic indices, making a proper comparison between cleps and eps51 difficult. 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 higher than 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 Rates, 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 an alert situation. If a user has a damage from the considered event, he can avoid the related loss by taking a preventive action. In 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 decides 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.

Averaging the precipitation over boxes of this size permits to understand if the total 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 could play a major role in forecasting precipitation peaks, introducing information not available with a lower resolution model. 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, 1.5×1.5 degrees. In this analysis, higher precipitation thresholds have been considered, due to the fact that the average precipitation over a large area and the maximum precipitation over the same area are two different quantities and can be considered to different “events” to be verified.

The BSS values relative to cleps for the 20 and 30 mm/24 thresholds (top and medium panels of Fig. 10, blue line) are well higher than both the epsrm and eps51 ones, indicating that COSMO–LEPS is more able to correctly forecast high precipitation values over a rather big area. Both eps51 and epsrm have almost no skill and no difference between the two is evident. At the highest threshold, all the three systems show almost no skill in terms of Brier Skill Score and it is not possible to find a difference between their performances.

In terms of ROC area (Fig. 11), cleps has the highest values for all the considered thresholds, exhibiting some skill even for the 50mm/24h thresholds. Then, the COSMO–LEPS system exhibits some skill in forecasting high precipitation peaks not reproduced by the lower resolution ensemble. According to this score measure, eps51 is more skillful than epsrm, showing the positive impact of the ensemble population on this index.

The difference of the behaviour of the scores in the evaluation of the three ensemble systems is due to the characteristics of these scores. In particular, the Brier Skill Score is more sensitive to the probability with which a forecast is issued: an event correctly forecast with low probability has a positive impact on the ROC area and a negative one on the Brier Skill Score. For this reason, it is better to look at a number of indices when evaluating the performances of a forecasting system.

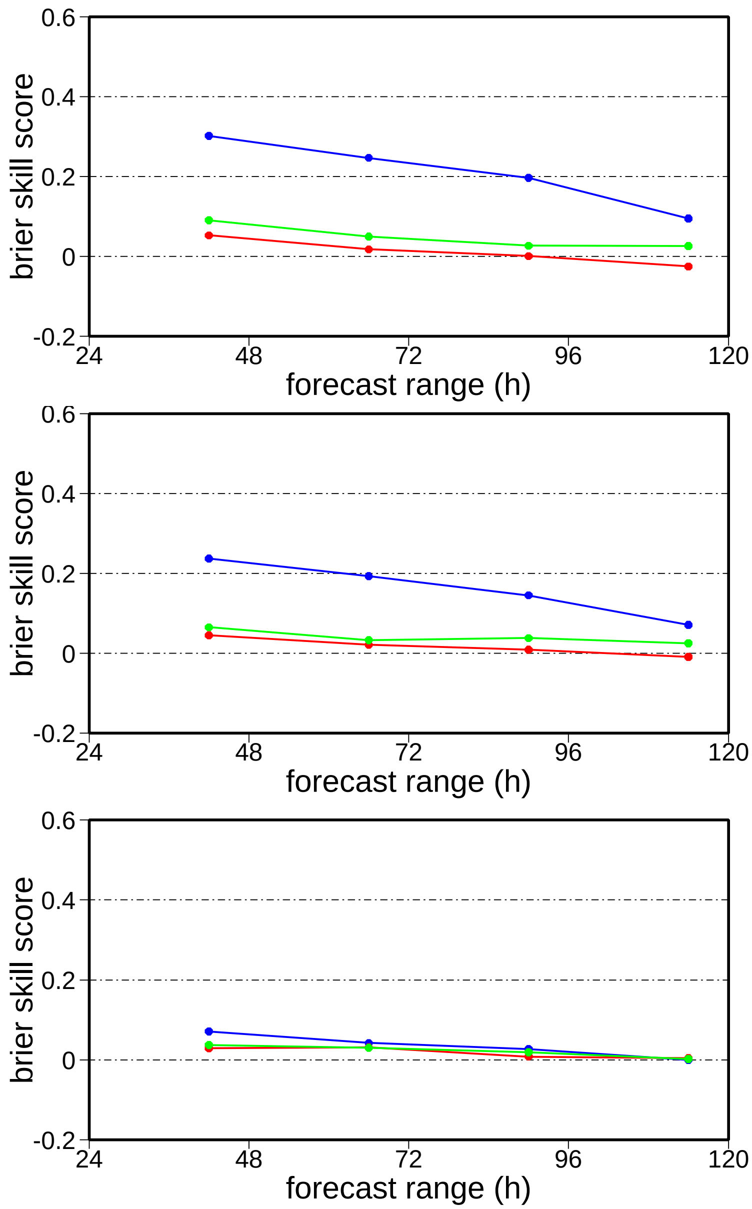


Figure 10: Brier Skill Score values for the precipitation threshold 20mm/24h (top panel), 30mm/24h (medium panel) and 50mm/24h (bottom panel). The blue line is relative to cleps, the red line is relative to epsrm, the green line is for eps51. Maximum observed and forecast values are compared.

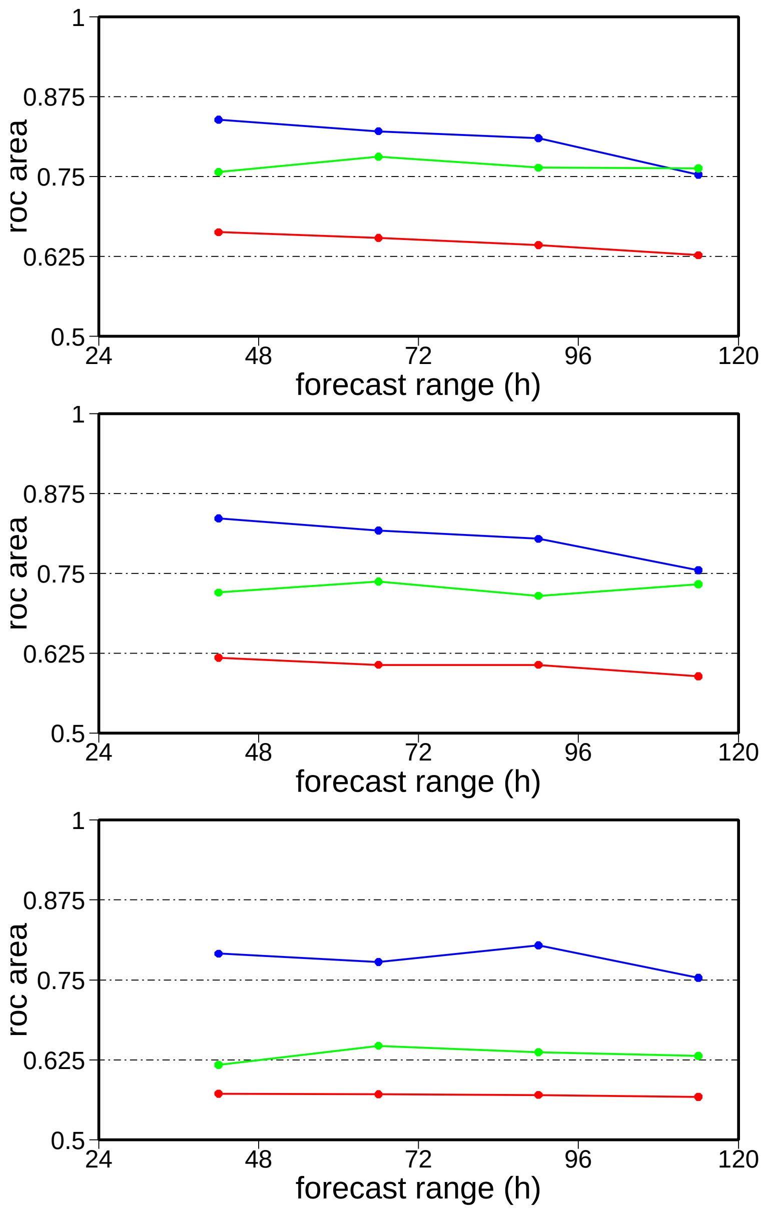


Figure 11: ROC area for the precipitation threshold 20mm/24h (top panel), 30mm/24h (medium panel) and 50mm/24h (bottom panel). The blue line is relative to cleps, the red line is relative to epsrm, the green line is relative to eps51. Maximum observed and forecast values are compared.

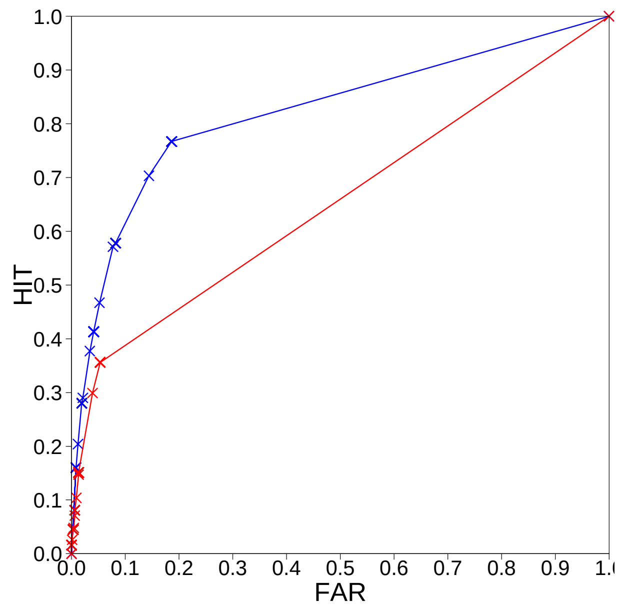


Figure 12: Maximum values: ROC Curves for the precipitation threshold 20mm/24h and for the +66h forecast range. The blue line is relative to cleps while the red line is relative to the epsrm.

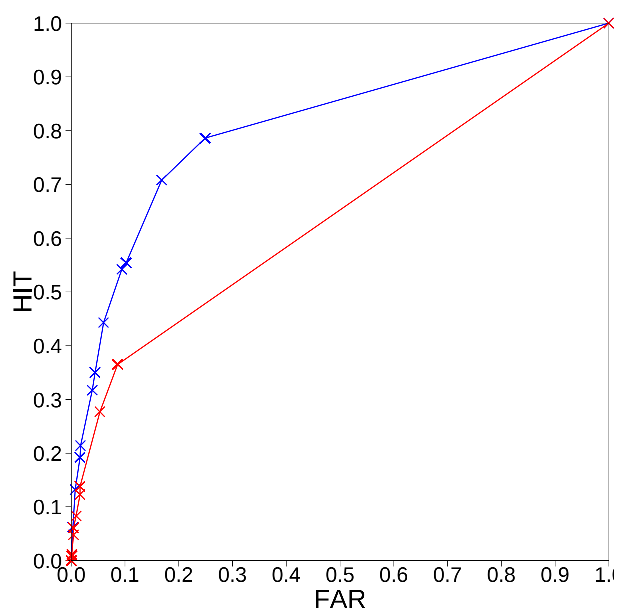


Figure 13: Maximum values: ROC Curves for the precipitation threshold 20mm/24h and for the +90h forecast range. The blue line is relative to cleps while the red line is relative to epsrm.

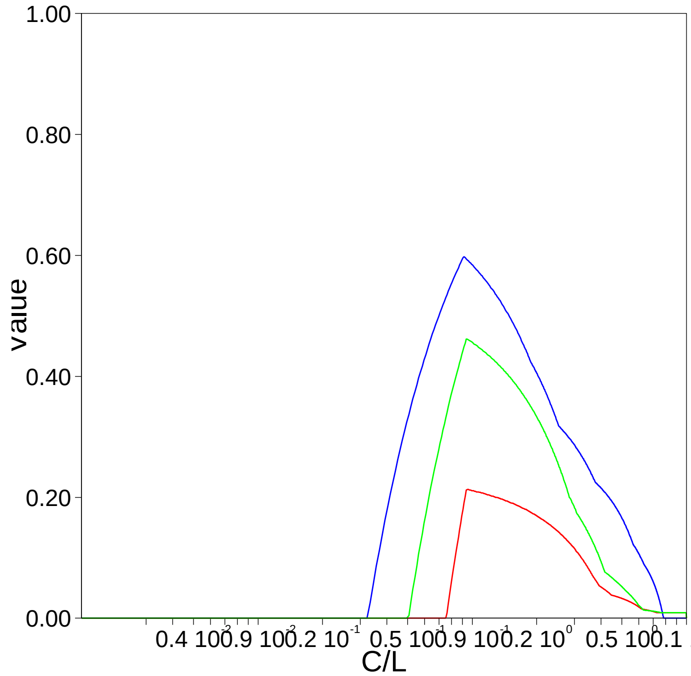


Figure 14: Envelope of the cost-loss curves for the precipitation threshold 30mm/24h (maximum values) for the +66h forecast range. The blue line is relative to cleps, the green line is relative to eps51 and the red line is relative to epsrm.

A general decrease of the cleps scores with increasing forecast range is evident for both BSS and ROC area, the score of cleps reaching that of eps51 at the +114 h forecast range for almost all the thresholds.

The ROC Curves relative to COSMO–LEPS and to the 5–RM EPS are shown for the event "maximum precipitation exceeding 20mm/24h", for the forecast ranges +66 hours (Fig. 12) and +90 hours (Fig. 13). The "cleps" curves (blue curves) are 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.

Finally, the cost-loss curves relative to the event "maximum precipitation over the box exceeding 30 and 50 mm/24h" are shown for the +66 hour forecast range (figures 14 and 15, respectively. This curves quantify the "value" of the forecast systems (expressed as a percentage, 100% is the value of a perfect forecast system) as a function of the ratio between cost (C) and loss (L). This ratio depends on the user, who incur a loss (L) if a dangerous meteorological event occurs and who can prevent this event by taking an action which cost is C. Each user has to find the position corresponding to his/her own "C/L" on the x–axis of this graph, then he/she can find the value he/she could obtain by using the forecasting system in order to decide when to take a preventive action.

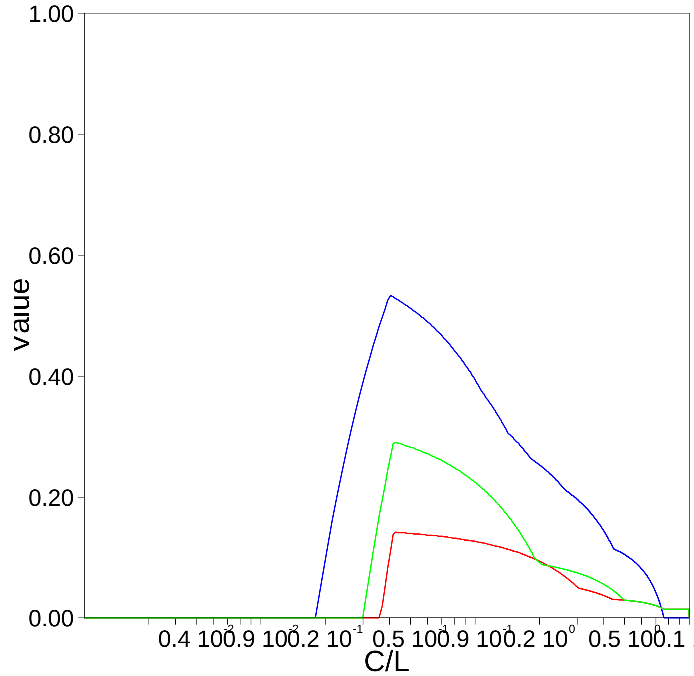


Figure 15: Envelope of the cost-loss curves for the precipitation threshold 50mm/24h (maximum values) for the +66h forecast range. The blue line is relative to cleps, the green line is relative to eps51 and the red line is relative to epsrm.

4 Parallel suite with model perturbations

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 to the mesoscale ensemble by nesting in each of the selected EPS Representative Member more than one limited–area model or, more simply, the same model in different configurations. This is a simple way of adding model perturbations in the ensemble formulation. This was attempted for the COSMO–LEPS system by 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 parameterization of the convection. This lead to a 10 member limited–area ensemble. At the same time, two 5–member suite have become available, differentiated only by the configuration of the limited–area model. Then, starting from September 2003 to May 2004, a second suite, parallel to the standard one was running. In the standard suite the Tiedtke scheme was used for the parameterization of the convection, while in the parallel suite the Kain–Fritsch scheme was used. The two systems are referred to as “Tiedtke suite” and“Kain–Fritsch suite”.

The 10–member COSMO–LEPS, obtained by simply joining the two suites, is then a system in which 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–hour precipitation using observed data from a network of raingauges covering Northern Italy (about 600 stations). The comparison is made over boxes of 0.5×0.5 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. The values are computed by considering, for each

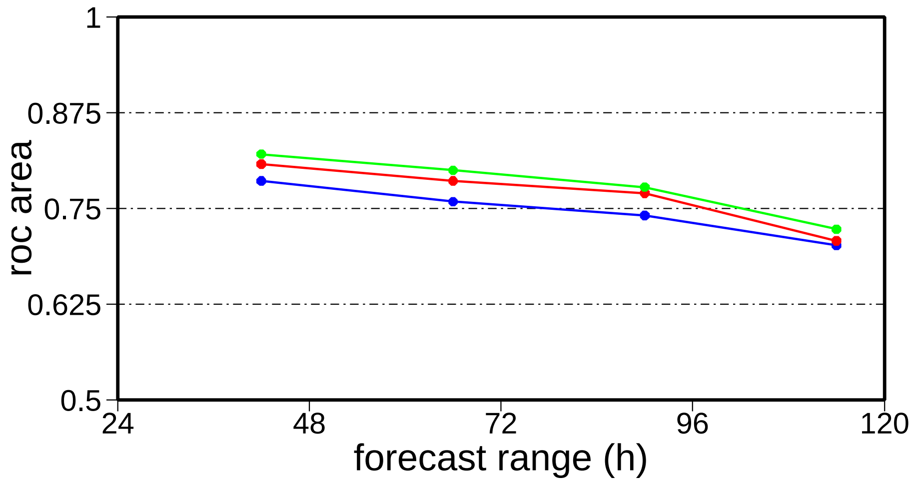


Figure 16: ROC area values as a function of the forecast range for averaged precipitation over 0.5×0.5 degrees boxes exceeding 20mm/24h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

box, only the model grid points that lies within a small distance from at least one observation point, following the "observational mask" technique described in the previous section.

In Fig. 16 the ROC area values computed in terms of average values over 0.5×0.5 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 than the other. In order to better understand the difference between the two schemes on the skill in forecasting precipitation, the ROC diagram at the +90h forecast range is reported in Fig. 17.

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 (50mm/24h) is chosen for this analysis. As shown in Fig. 18, 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) is still the highest, but only by a very little amount.

Considering the ROC Curves at the +90h forecast range (Fig. 19), 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 Rate are almost identical.

Actually, it does not seem possible to establish from these results which convection scheme lead to the best performance of the mesoscale ensemble. Furthermore, it is evident that using this kind of model perturbation in order to increase the ensemble spread does not lead to the expected results, the performance of the 10–member ensemble being almost identical

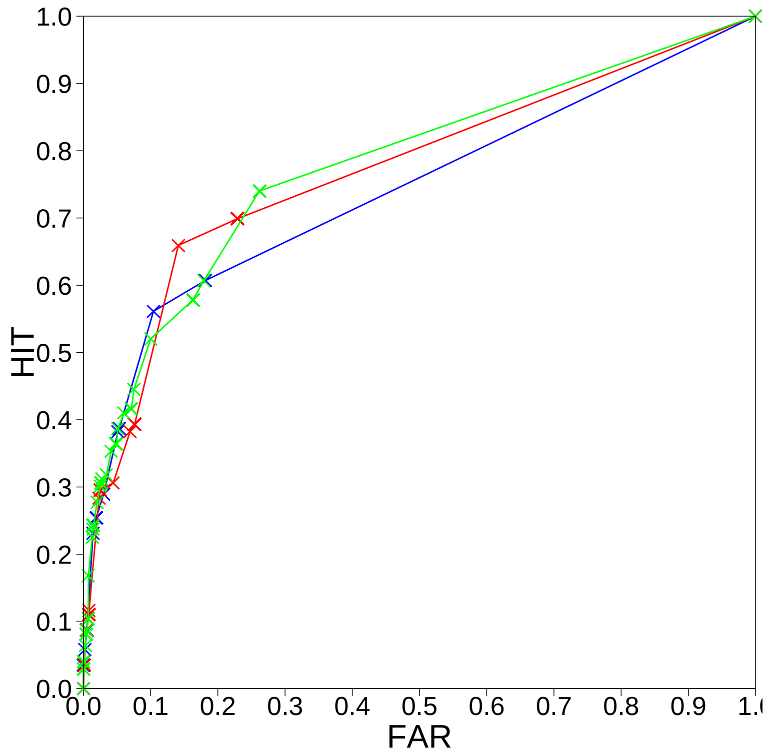


Figure 17: ROC Curves for the precipitation threshold 20mm/24h (average values) and for the +90h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

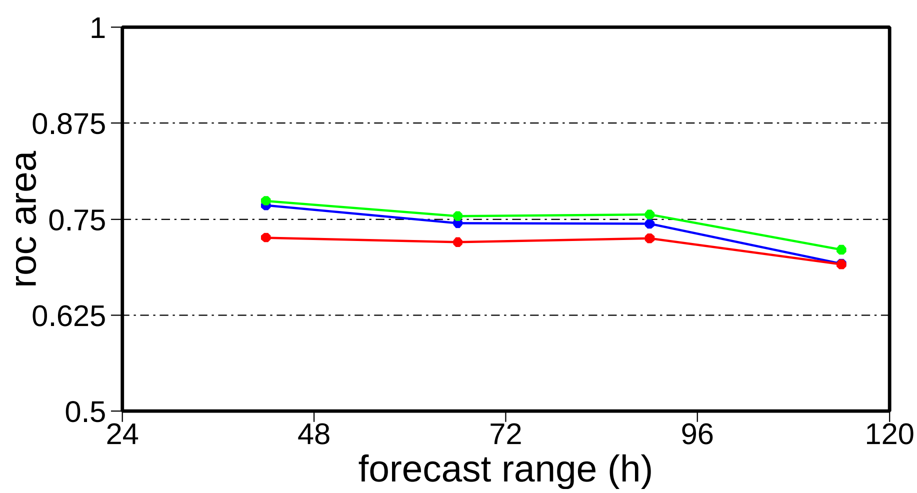


Figure 18: ROC area values as a function of the forecast range for maximum precipitation over 0.5 x 0.5 degrees boxes exceeding 50mm/24h. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the 10–member combined suite.

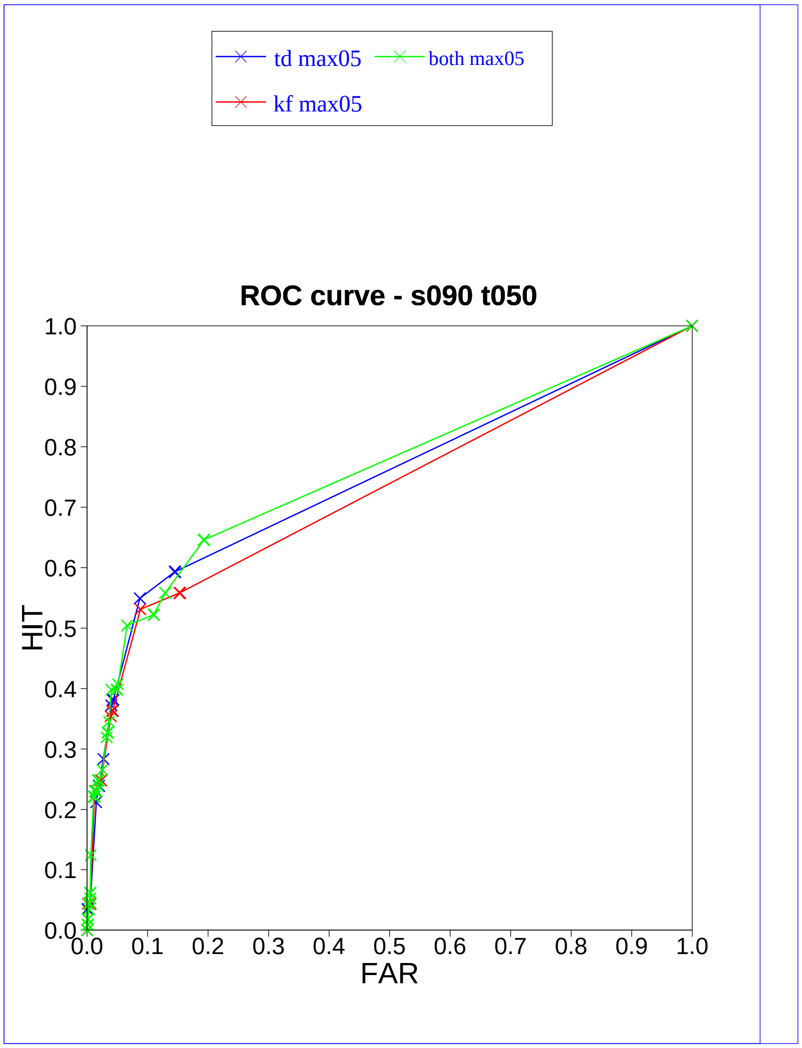


Figure 19: ROC Curves for the precipitation threshold 50mm/24h and for the +90h forecast range. The blue line is relative to the Tiedtke suite (operational), the red line is relative to the Kain–Fritsch suite and the green line is for the combined suite.

to those of both 5–member ensembles. Due to this result, we decided to move to a 10–member COSMO–LEPS by selecting 10 Representative Members out of the EPS, to benefit from the spread introduced by the different initial and boundary conditions. Nevertheless, trying to benefit also from the little spread added by this kind of model perturbations, we decided to perform the 10 LM runs by using both the Tiedtke and Kain–Fritsch schemes for the parameterization of the convection. Then, from June 2004, also this modification was introduced in the COSMO–LEPS suite and the convection parameterization scheme used by each single run is selected randomly.

5 COSMO-LEPS Subjective verification – Contribution by ARPAL, Genova, Italy (Davide Sacchetti)

This analysis aims at evaluating from a quantitative point of view the subjective judgements by the forecasters on the usefulness of COSMO–LEPS as a tool for operational weather forecast. The followed methodology is described in section 5.1 while the results are presented in section 5.2.

5.1 Description of the methodology

The questions we have tried to answer are:

- do the members provide weather scenarios actually different from the forecaster's point of view?
- do the members provide useful informations on both the amount and the localization of the precipitation?
- is the most weighted member the more skillful?
- what kind of information can be obtained by the forecasters from the probability maps for the exceeding of a threshold?
- does the system performance depend on the synoptic situation?

In order to answer these questions, the forecasters have been asked to judge the COSMO–LEPS products following a scheme developed for this purpose.

5.1.1 Subjective evaluation of the precipitation field forecast by the 5 members

A qualitative analysis of the skill of the forecast provided by the 5 COSMO–LEPS members is performed through three steps:

- identification of the member(s) which better forecasts the localization of the precipitation maxima over the considered area
- identification of the member(s) which better forecasts the amount of maxima of precipitation over the considered area
- How many members are significantly different?

In Fig. 20 the mask used by the forecaster for the evaluation of the 5 different scenarios provided by COSMO-LEPS is shown.

It is also possible to signal that no one of the members is representative of the observed scenario.

At this stage the forecaster does not know the weights which are assigned to each member; the comparison between his/her choice and the "a priori" probabilities are carried out afterwards.

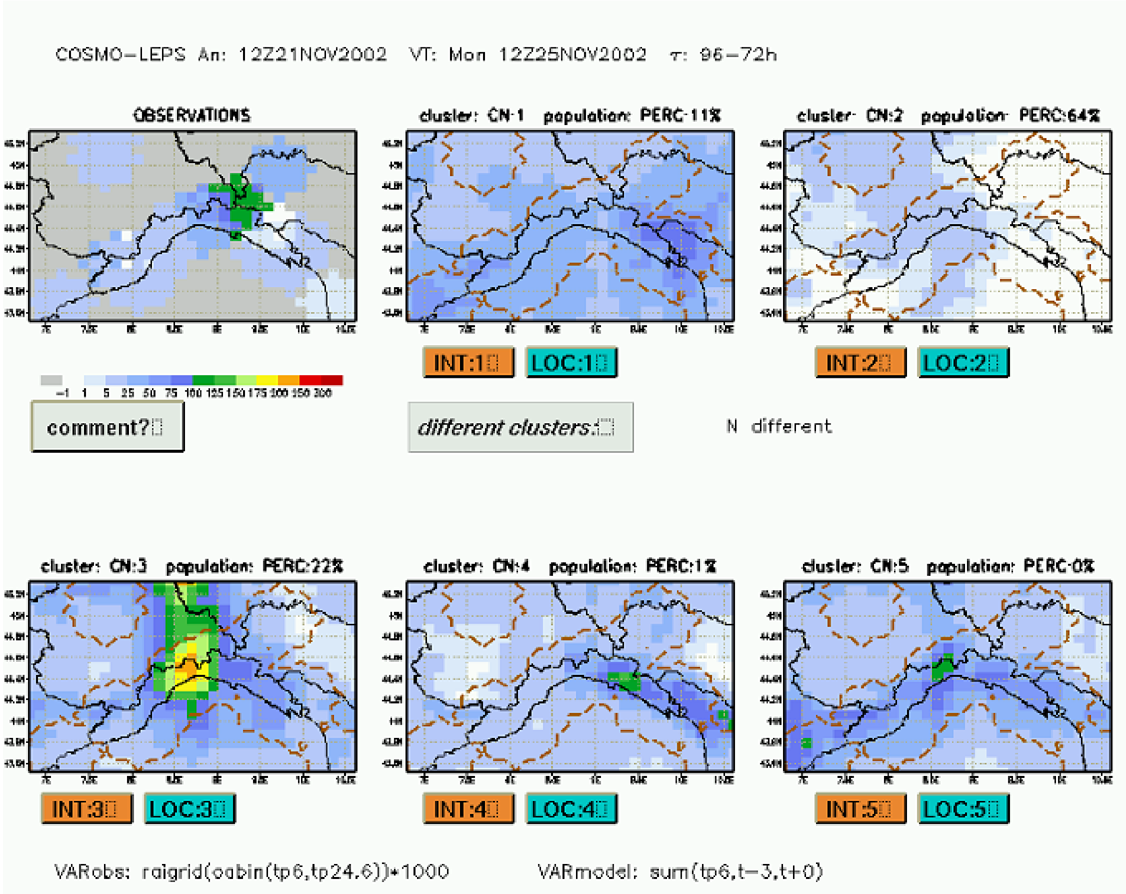


Figure 20: Mask used by the forecasters of the ARPAL Hydro–meteorological Service for the subjective evaluation of the forecast provided by the 5 COSMO–LEPS members.

5.1.2 Subjective evaluation of the COSMO–LEPS probability maps

A qualitative analysis of the skill of the forecast provided by the COSMO–LEPS probability forecast is performed. For each precipitation threshold (20, 50, 100 mm/24h), the probability map is compared with a map of observed precipitation cut at the corresponding threshold (only values exceeding the threshold). The forecaster assigns a score (agreement: poor, reasonable, good) to the probability fields on the basis of their usefulness. This is evaluated only in terms of the "alarm level" suggested by the probability maps, that is on the forecast precipitation amount, giving only small importance to the correct localization of the phenomena. In this phase what is more important is the capability of providing a signal of an intense event.

In Fig. 21 the mask used by the forecaster for the evaluation of the COSMO–LEPS probability maps is shown.

5.1.3 Individuation of the synoptic pattern

Finally, the performance of COSMO–LEPS is related to the synoptic situation associated with the considered event. The forecasters have to answer to the following question: to which synoptic pattern is it possible to associate the considered event? They can choose between four possibilities, which are listed in Table 5.1.3.

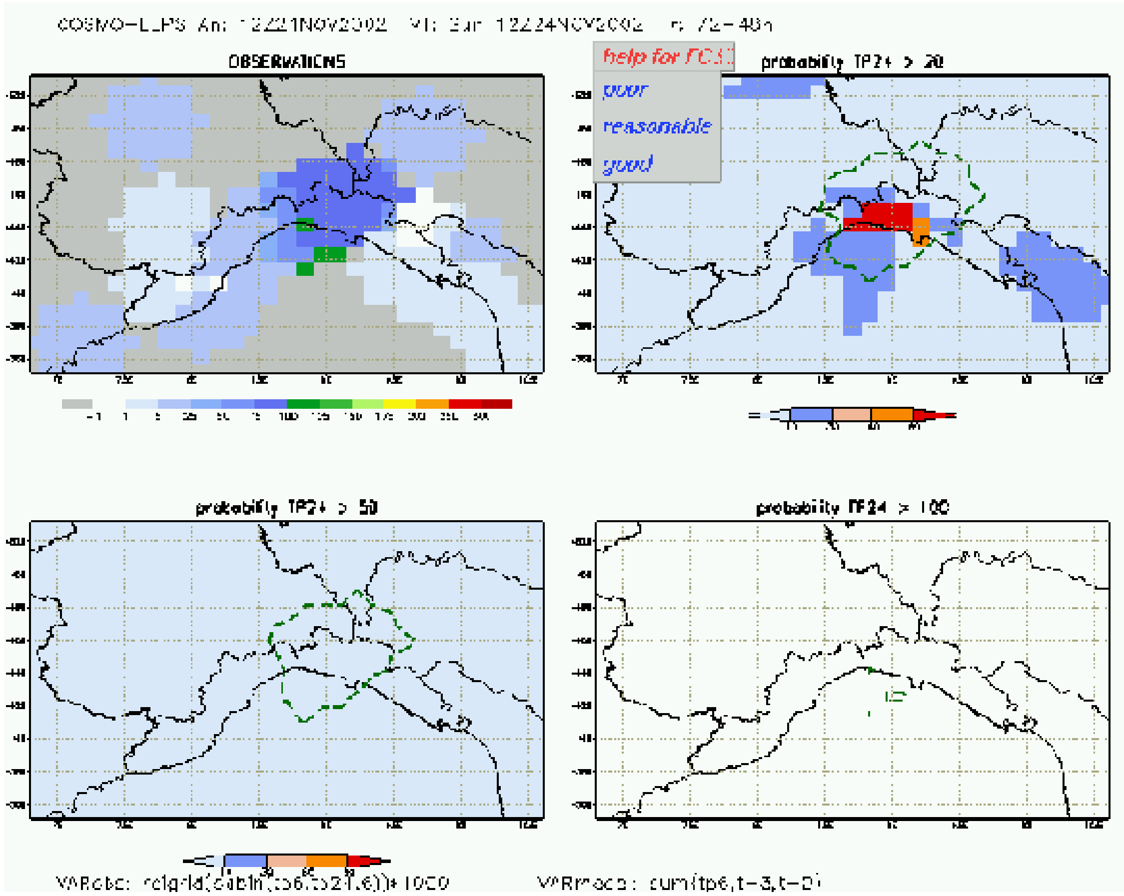


Figure 21: Mask used by the forecasters of the ARPAL Hydro–meteorological Service for the subjective evaluation of the forecast provided by the COSMO–LEPS probability maps.

Table 1: Medium range weather types

Name	Description	Main characteristics of the weather
straight flow	rapid, straight, perturbed	rainy, often windy
undulating flow	alternating ridge and trough, unsettled weather	a rainy period is followed by a sunny one; there is at least one rainy period
warm blocked	anticyclonic conditions	dry weather, sunny in summer, sometimes foggy or dull weather in winter
cold blocked	low-pressure conditions	cool or cold weather, often rainy

5.2 Results

In Fig. 22 the usefulness for the forecasters of the COSMO–LEPS scenarios in terms of precipitation intensity is assessed for different forecast ranges. In the left panel the analysis is relative to the evaluation directly made by the forecasters, ignoring the weight given to each member according to the cluster population, while in the right panel weights have been associated to the COSMO–LEPS members.

A positive impact of the weighting procedure is evident, for all the precipitation intensities and for all the forecast ranges.

The same analysis has been repeated in terms of localization of the precipitation (Fig. 23).

The positive impact of the weighting procedure on the localization is well evident, especially for moderate and intense precipitation.

How the usefulness of the forecasts differentiate with the synoptic situation is shown in Fig. 24.

The “warm blocked” situation seems to be the more predictable, while the “straight” flow is the less predictable. The performances both in terms of intensity and localization improve when the ensemble members are weighted according to the cluster population.

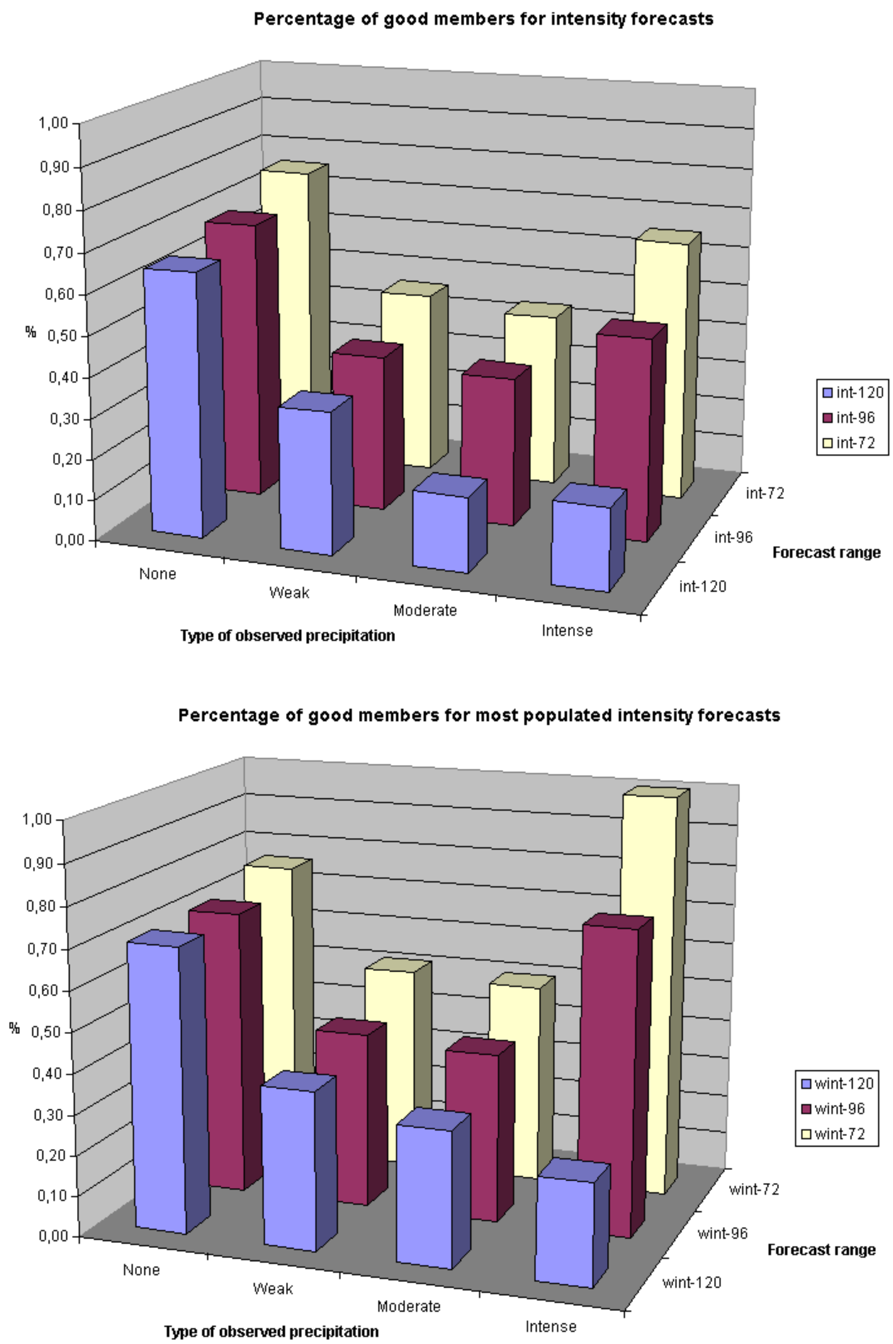
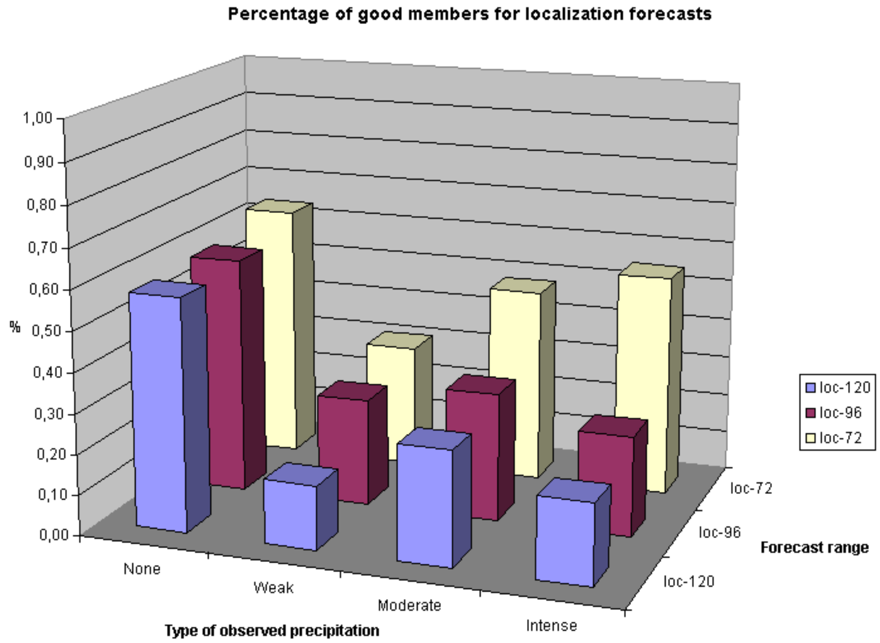


Figure 22: Percentage of "good" members in terms of intensity for four classes of intensity and for the different forecast ranges (+72h, +96h, +120h). The percentage is computed by not weighting (top panel) and by weighting (bottom panel) the ensemble members.



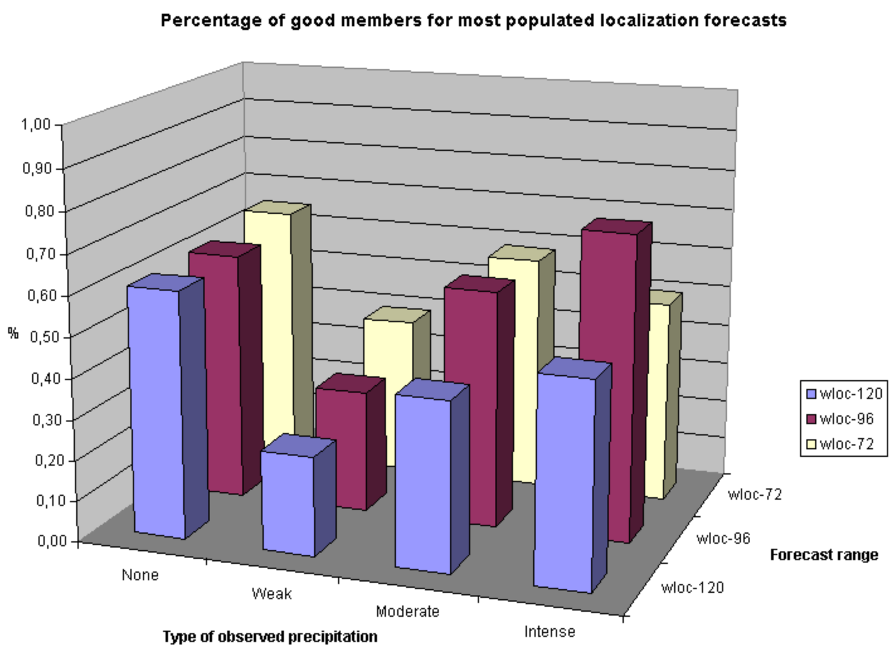


Figure 23: Percentage of "good" members in terms of localization for four classes of intensity for the different forecast ranges (+72h, +96h, +120h). The percentage is computed by not weighting (top panel) and by weighting (bottom panel) the ensemble members.

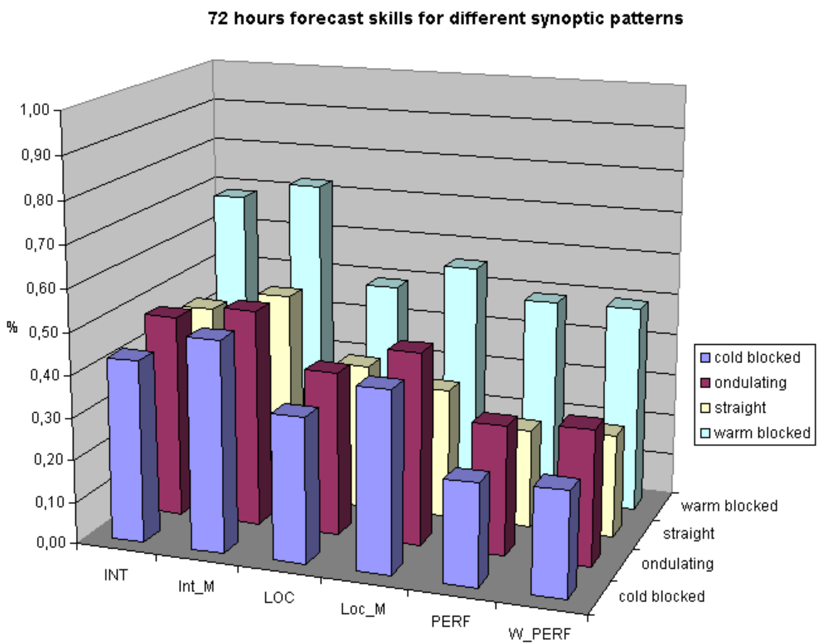


Figure 24: Percentage of "good" members in terms of intensity (int) and localization (loc) for the four types of synoptic situation for the +72h forecast range. A label M is added to int and loc if the percentage is computed by weighting the ensemble members.

6 Use of COSMO-LEPS at MeteoSwiss - Contribution by MeteoSwiss, Zürich, Switzerland (Andre' Walser and Marco Arpagaus)

The quasi-operational COSMO-LEPS has made probabilistic short to early-medium range forecasts available at MeteoSwiss. It motivated scientists and forecasters from the weather department to look closer into the topic of probabilistic weather forecasting and provides an excellent basis for research in this field. Since the optimal use of probabilistic forecasts needs a knowledge transfer from scientists to forecasters, the introduction of COSMO-LEPS had both technical and educational aspects.

This chapter is structured as follows: The MeteoSwiss post-processing of COSMO-LEPS is described in section 6.1. Section 6.2 discusses the experience with COSMO-LEPS from the perspective of the forecasters, while section 6.3 presents the use of COSMO-LEPS forecasts exemplified for a storm event. Finally, conclusions and an outlook are provided in section 6.4.

6.1 MeteoSwiss post-processing

The setup of the MeteoSwiss post-processing of COSMO-LEPS output has been strongly supported by our colleagues of ARPA-SIM and is summarised in Fig. 6.1. This post-processing starts as soon as the COSMO-LEPS forecasts are available at ECMWF and consists of

(i) a transfer of COSMO-LEPS probabilistic model output (PMO) and direct model output (DMO) of all ensemble members to the MeteoSwiss post-processing server at the CSCS[1],

(ii) using the DMO to calculate customer-specific PMO, and

(iii) a visualisation of the PMO in the form of probability maps and meteograms, the latter of which is described in the following paragraph.

These COSMO-LEPS products are presented on the intranet of MeteoSwiss and some selected products are printed out automatically and pinned to the weather board in the meteorological operations room.

A very popular output of the post-processing is the mentioned meteogram. It is currently produced for about 30 grid points within the model domain and summarises the 5-day COSMO-LEPS forecast in terms of cloud cover, precipitation, wind gusts and temperature. Figure 6.1 shows the meteogram for the grid point closest to Zürich from the forecast started on 25 August 2004 1200 UTC. For each variable the median of the 10 COSMO-LEPS members is displayed as red line. The middle two quartiles (25% to 75% probability of occurrence) are indicated by the grey shaded area, and the minimum and maximum values, respectively, by the dashed lines. Median and quartiles are calculated by weighting the members according to the cluster population and with a linear interpolation between the discrete cumulative probabilities for every three hours (corresponding to the current output frequency of the COSMO-LEPS members). Additionally, the meteogram shows the deterministic forecast

[1]Swiss National Supercomputing Centre, Manno.

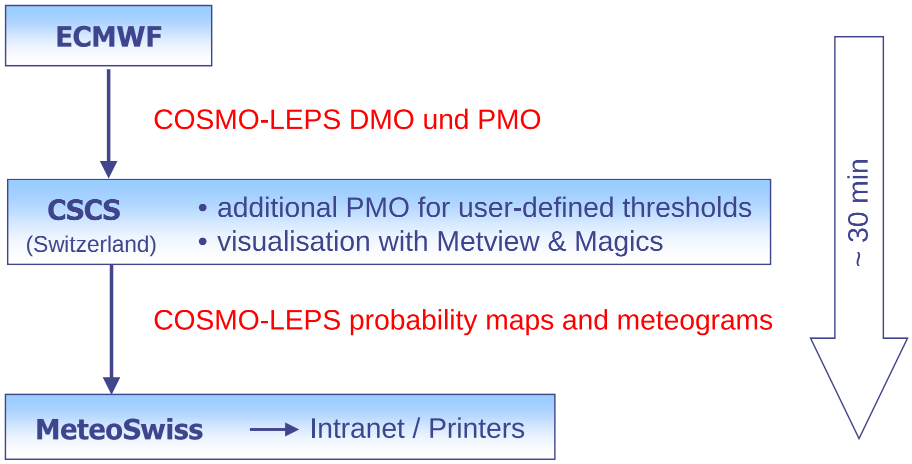


Figure 25: Flow chart of the MeteoSwiss post-processing of COSMO-LEPS output.

of the operational high-resolution model at MeteoSwiss (aLMo) as blue lines. Since the aLMo forecast initialized 12 hours later than COSMO-LEPS is available shortly after the COSMO-LEPS forecast, this newer 72-h forecast is displayed in the meteogram.

The width of the grey shaded area together with the spread between the minimum and maximum values give a measure of the uncertainty of the most likely scenarios given by the aLMo and the COSMO-LEPS median. It is expected that aLMo is the better estimate than the median for short lead-times, however, without yet knowing how long "short" in this context is. Temperature and wind gust forecast show the typical case with increasing spread with increasing forecast time. Clearly, the prediction of cloud cover is associated with the highest uncertainty, this being often associated with a similar uncertainty in the forecast for precipitation.

6.2 Experience with COSMO-LEPS

The change from predicting precipitation amounts to predicting probabilities of occurrence for a certain threshold is a radical paradigm shift which clearly demands some training for forecasters not familiar with probabilistic forecasts. This was essential for the introduction of COSMO-LEPS forecasts at MeteoSwiss. Meanwhile, about half of the forecasters already use COSMO-LEPS together with other products to prepare their weather bulletins for the next few days.

The overall feedback on COSMO-LEPS from the forecasters is clearly positive. They mainly use probability maps of 24-h accumulated precipitation and maximal daily wind gusts for forecast days 2 to 4 (for an example of a probability map, cf. Fig. 28). COSMO-LEPS products for other variables are not yet considered frequently. In addition, some forecasters also look at a small set of meteograms. Forecasters describe the use of COSMO-LEPS for the current bulletin as "considering an additional opinion besides the other models", which are mainly aLMo and deterministic ECMWF. E.g., for forecasting precipitation amounts, a high probability for a threshold below the predicted value and a low probability above would be considered a confirmation of the predicted deterministic estimates.

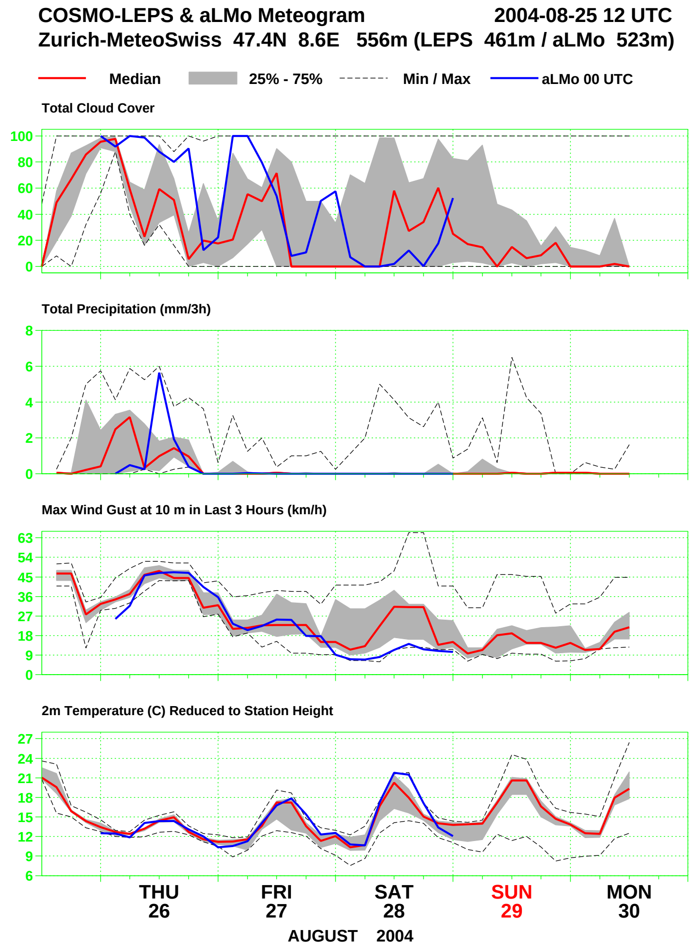


Figure 26: Meteogram as derived from the COSMO-LEPS and aLMo output in terms of cloud cover, precipitation, max. wind gusts, and temperature (see text). Example for Zürich from the forecast starting on 25 August 2004 1200 UTC.

COSMO-LEPS forecasts are studied more thoroughly when they show considerable probabilities for a threshold which represents a warning level or if forecasters have other indications for an extreme precipitation or storm event. Since such extreme events often have low predictability even in the short-range, COSMO-LEPS PMO is also considered for this forecast-range (for which COSMO-LEPS is not designed) due to the lack of an alternative. There is no verification done yet concerning the performance of warnings from forecasters compared to COSMO-LEPS itself in terms of hit rate and false alarm rate. Such a verification for Switzerland would be of strongest interest for the forecasters and would possibly help to extend the use of COSMO-LEPS (personal communication from a forecaster). In discussions with forecasters the following concerns about COSMO-LEPS turn up frequently:

- There is little confidence that COSMO-LEPS is able to predict reliably the probability for convective precipitation events in summer.
- Due to the rather coarse horizontal resolution, it is not expected that COSMO-LEPS is able to give warnings for extreme events related to small-scale orography such as Föhn.
- For short lead-times, the use of the super-ensemble could smooth the predicted probability distribution, reducing the probability for an extreme event and hence the hit rate.

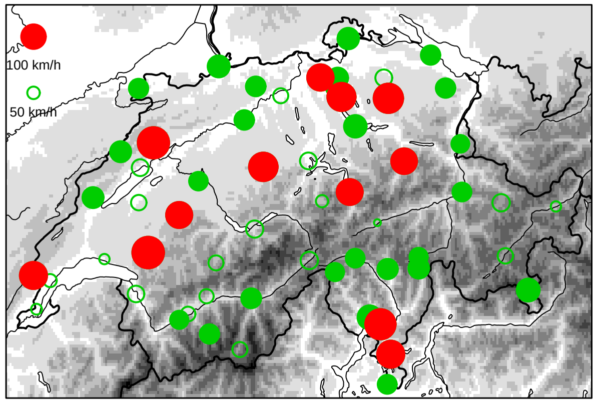


Figure 27: Observed maximal wind gusts on 19 November 2004 as size of centered circles (scale is indicated in the panel) for 59 stations below 2000 m a.s.l. of the Swiss automatic observation network. Green and red filled circles indicate wind gusts above 70 km/h and 100 km/h, respectively. The grey shaded background represents the topography.

Overall, the use of PMO has clearly increased due to the availability of COSMO-LEPS products which are acknowledged by the forecasters. However, the entire potential of COSMO-LEPS is not yet exploited since forecasters are not instructed to complement the weather bulletins with a level of confidence, e.g. with a probability of occurrence for the predicted event. This would mean an even more profound paradigm shift, which clearly needs more confidence in the ability of COSMO-LEPS (or other high-resolution probabilistic weather forecasting system) to predict the probability distributions reliabe for the most important parameters such as temperature, horizontal wind, precipitation, and cloud cover.

6.3 Case study: Storm on 19 November 2004

In this section, the COSMO-LEPS forecasts and their use are discussed on the basis of a case study. In the early morning of 19 November 2004 a cold front crossed Switzerland in a strong westerly flow producing maximal wind gusts of 60-90 km/h on the Swiss Plateau and slightly above 100 km/h in somewhat elevated places, in Alpine valleys, and in the Lago Maggiore Area to the south of the Alps. Figure 6.3 shows the observed maximal wind gusts for 59 stations located below 2000 m a.s.l. of the Swiss automatic network with green and red filled circles indicating maximal gusts above 70 km/h and 100 km/h, respectively, where the latter value corresponds to the warning threshold for the regional authorities.

On 17 November the forecasters on duty decided to issue a pre-warning for wind gusts between 80-100 km/h on the Swiss Plateau during the night from 18 November to 19 November. They based the warning on (i) the mean 850 hPa flow of the deterministic ECMWF forecast, (ii) the predicted aLMo wind gusts, and (iii) the COSMO-LEPS probability maps for maximal 10 m wind gusts. On the following day, this warning was confirmed and a new warning for the Ticino was issued for wind gusts of 75-100 km/h embedded in a strong north Föhn.

While ECMWF and aLMo suggested wind gusts somewhat above and slightly below 100 km/h, respectively, COSMO-LEPS indicated in the four consecutive forecasts started on 15, 16, 17, and 18 November 1200 UTC high probabilities for wind gusts below 100 km/h, but negligible probabilities for wind gusts above this value, shown in Fig. 28. Only the forecast started on 18 November (Fig. 28d) predicted considerable probabilities for maximal wind gusts above 100 km/h for the eastern part of the Swiss Plateau.

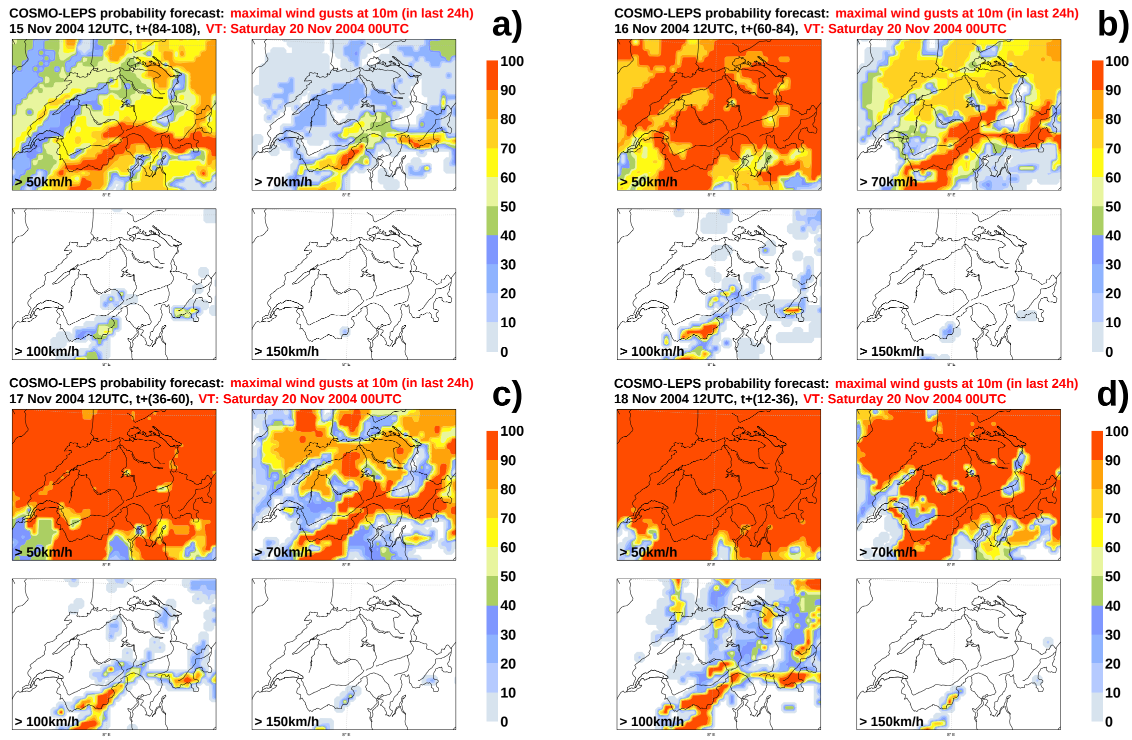


Figure 28: Probability forecast for maximal wind gusts above 50, 70, 100, and 150 km/h, respectively, from four consecutive COSMO-LEPS forecasts for 19 Nov 2004 initialised at (a) 15 Nov, (b) 16 Nov, (c) 17 Nov, and (d) 18 Nov 1200 UTC.

Overall, COSMO-LEPS supported the forecasters well in this case by suggesting a high probability for maximal wind gusts at the observed level on the Swiss Plateau without suggesting a more extreme event. However, COSMO-LEPS missed the wind gusts above 100 km/h at the elevated places in western Switzerland and - as expected - in the Alpine valleys including the strong north Föhn in the Ticino. Considering the four consecutive forecasts, the steady increase in the probabilities for maximal wind gusts below the observations with decreasing lead-time is remarkable, whereas a corresponding decrease of probabilities for maximal wind gusts above the observations cannot be seen. A sharpening of the predicted probability density function with decreasing lead-time as can be partly observed for this case is a key characteristic of a well-designed EPS.

6.4 Conclusions and outlook

The developed probabilistic COSMO-LEPS products try to visualize the complex information of COSMO-LEPS forecast in a concise way. They complement the deterministic forecast products of aLMo and are appreciated both from forecasters and scientists as subjectively skillful information.

We are convinced that high-resolution EPSs such as COSMO-LEPS have a great potential for further improvements in terms of forecast skill. It is planned to continue our research activities in this field in close collaboration with the ARPA-SIM focusing on potential high-impact weather. A key goal is the comparison to different available and newly developed forecasting approaches such as ECMWF extreme forecasts index and neural networks. In addition, a model-based climatology will be assembled to assess the return periods of high-impact weather events as a function of forecast lead-time and compared to observations.

Moreover, this model climatology will allow to identify model biases related to intrinsic model deficiencies.

The interest in high-resolution EPSs will undoubtfully further increase and it is worth to mention here, that the executive board of MeteoSwiss recently declared probabilistic forecasting as a strategic goal.

7 Could the COSMO–LEPS system add value to operational forecasts of the DWD in the case of severe weather? - Contribution by the Central Forecast unit of the DWD in 2004 (Thomas Schumann)

Since 31 January 2003 products of the COSMO–LEPS system are available in the DWD intranet for use in operational forecasts. COSMO–LEPS as a limited area ensemble prediction system is running once a day, driven (since 01 June, 2004) from 10 selected EPS members of 2 consecutive (12-hr lagged) runs of the ECMWF global model. The forecast range is beginning at H+48 in the short range and ending at H+120 in the early medium range from the last of the both ECMWF runs. The main objective of the COSMO–LEPS system is to identify severe weather patterns within a certain time range over a limited area by using probabilistic methods. To reach this, LM-based probabilistic products will be calculated, disseminated and presented to the forecaster in a user-friendly design by using the DWDs intranet or the meteorological application system via workstations. Beginning in March 2003 the COSMO–LEPS system at the DWD will be used pre-operational as well during the medium range as for short range forecasts. COSMO–LEPS is an additional source of information, forecast plots will be checked whether possible severe events could be identified by relevant signals in the probability maps or not. This is dependent of course from the synoptic situation. Products such as a direct model output from COSMO–LEPS were not generated. In the Central Forecast unit a subjective, very comprehensive verification of the COSMO–LEPS probabilistic forecasts has been carried out. The verification has been done by comparing observations against the forecasts verifying at the time of the observation by continuous tables for each weather parameter including a short description of the event. This task is still a part of the operational medium range forecast shift carried out by the duty forecaster of this shift and will be continued. Another verification method used in the Central Forecast Unit is the preparation of case studies. If a severe event occurred the observed values of a certain weather parameter will be compared with the probability forecasts of COSMO–LEPS at certain thresholds verifying at the time of the occurrence of severe weather. Forecasts of other models or centers will be included. The objective of the verification of the COSMO–LEPS system is:

- Is COSMO–LEPS able to add value to the prediction of severe weather events?
- Could COSMO–LEPS outperform other models?
- What are typical weak points and what are the strength of COSMO–LEPS?
- Are there any changes in the COSMO–LEPS system required (design, thresholds, weather parameters, forecast range)?

To answer these questions we have to go more into detail how COSMO–LEPS performs related to weather parameters provided from the system. An overview has been prepared as a result of the continuous verification, forecasters experience and several case studies. This is provided by Table 2.

Parameter	General remarks Predictability by COSMO–LEPS	Further comments / recommendations
Temperature Minimum of the day	Very poor, mostly no signals	Useful in a few cases only without near surface inversion. Outperformed by the KALMAN / MOS output from LM or ECMWF deterministic model.
Temperature Maximum of the day	Sometimes useful	Realistic regional assignment, even heat islands (large cities) visible. Poor results in the case of extreme heat waves (underestimation of temperature maximum). Outperformed by the KALMAN / MOS output from LM or ECMWF deterministic model.
Maximum wind gusts of the day	Major synoptic scale storms: Good forecasts. Convective gusts: No correct signals	Realistic regional assignment, orographic effects well represented. COSMO–LEPS adds value to forecasts. Prediction of gusts caused by convection events not possible. If signals presented they are wrong located, orographic effects overestimated. Thresholds should be changed, levels 30 and 35 m/s more important for severe weather prediction than 10 and 15 m/s
Snow fresh fallen during the last day	Sometimes useful	Synoptic-scale events: Realistic regional assignment, orographic effects well represented. COSMO–LEPS adds value to forecasts. Poor results in the case of minor severe / small scale events
Total precipitation accumulated over 24 hours	Often useful	large-scale events: Realistic regional assignment, orographic effects well represented. COSMO–LEPS could add value to forecasts. Poor results in the case of minor severe / small scale, convective or convective-mixed events
Convective precipitation Thunderstorm (CAPE)	Very poor, mostly no signals. CAPE: more cases needed	Prediction of heavy precipitations caused by convective or convective-mixed events not possible. If signals presented they are wrong located, orographic effects overestimated

Table 2: Overview on Forecasters experience

To summarize this and to answer the questions above, it has to be concluded that COSMO–LEPS did not always met the forecasters requirements and expectations. Valuable predictions of severe events provided only if the event is non-convective and at least a synoptic-scale one. In this situations COSMO–LEPS will be able to add value to the prediction of severe events by providing probabilities exceeding certain thresholds. Then other models mostly are performing very well, too. COSMO–LEPS could outperform other models in areas with clear orographic structure only. This is the strength of COSMO–LEPS and why it has been designed: Provide the forecaster a hint what could happen in these areas in the case of severe weather. COSMO–LEPS has been adjusted to deal with the orography of the Alps and the surrounding mountains of the Mediterranean Sea and not with the highlands over Central Germany. Another point is that COSMO–LEPS is reflecting the weaknesses of the driving ECMWFs EPS and the nested LM. This has been characterized by the problems of

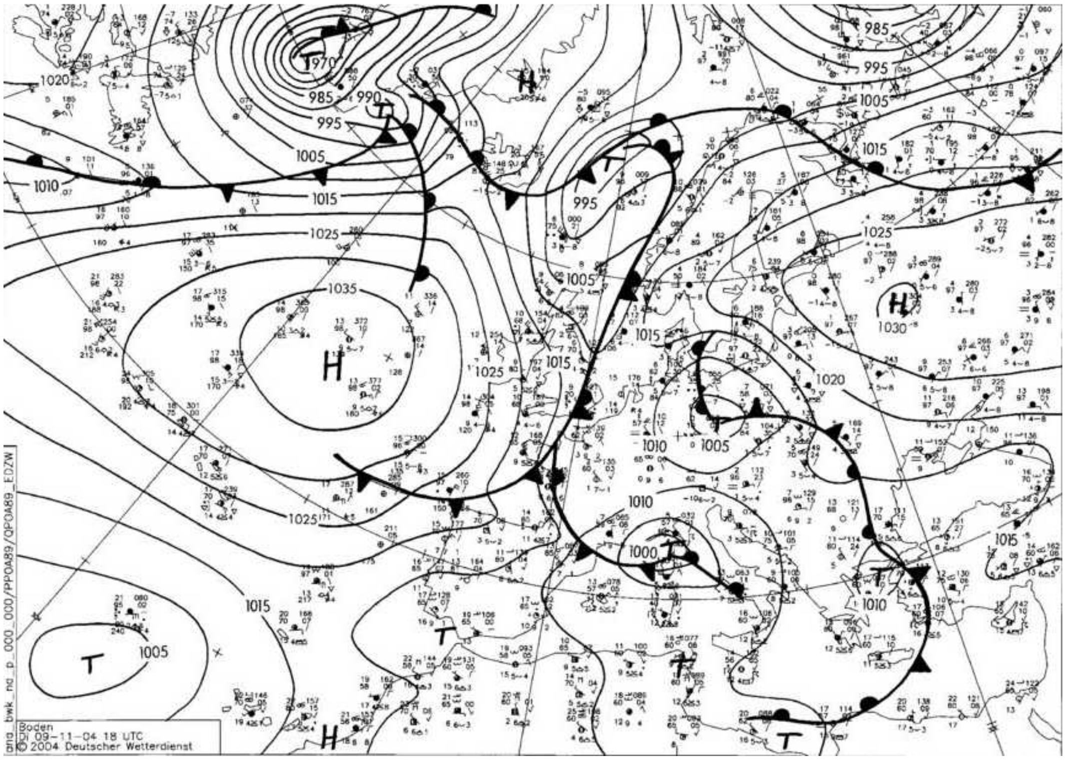


Figure 29: MSLP Analysis, 09 Nov 2004, 18 UTC.

COSMO–LEPS predicting convective gusts and precipitation and due to the poor vertical resolution in the lower troposphere in the forecast of the daily temperature minimum as well as for the temperature maximum in the case of extreme heat waves. Further verification is required to detect changes in the model behavior caused by the variation of the configuration since 06/2004.

7.1 A case study - early winter over Germany (09 Nov, 2004)

The synoptic situation was characterized by a huge cut off-low with embedded centers over the west part of Germany and the southern Adriatic Sea. Crossing the eastern part of the Alps the second center has been induced a cyclogenesis. The resulting low passed the Czech Republic and has been moved over Northeastern Germany (see Fig. 29, SLP Analysis 09 Nov 2004, 18 UTC.). The low caused the first winter outbreak over Germany in 2004/2005. Widespread areas have been covered by fresh fallen snow (Fig. 30, snow cover over Germany in centimeter, 10 Nov 2004, 06 UTC). The early winter outbreak lead to chaotic situations in the traffic including disruptions of streets by broken trees because of wet snow (Fig. 31, impressions from the winter outbreak over central parts of Germany).

How has COSMO–LEPS predicted this winter outbreak? That will be demonstrated by Fig. 32. On the left the probability of exceeding 10 centimeter fresh fallen snow is shown, the right shows the 20 centimeter-threshold. The first row are the 120-hr forecasts verifying on 10 November, 12 UTC. Already during the early medium range COSMO–LEPS has been shown relevant signals. In the highlands over the central part of Germany 10 centimeter fresh fallen snow is likely and up to 20 centimeter is possible. Over the southwestern part of Germany not more than 10 centimeter snow has to be expected. In this situation COSMO–LEPS has provided an "early warning". Therefore it was possible to describe this scenario

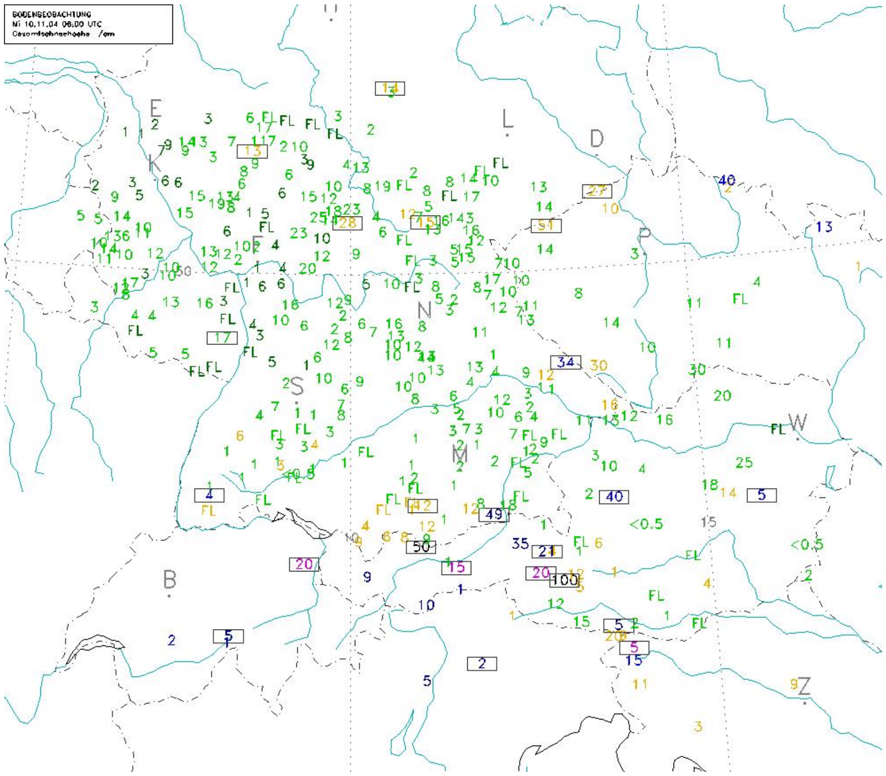


Figure 30: Fresh fallen snow, 10 Nov 2004, 06 UTC

Figure 31: Impressions from the winter outbreak over central parts of Germany 10 Nov, 2004

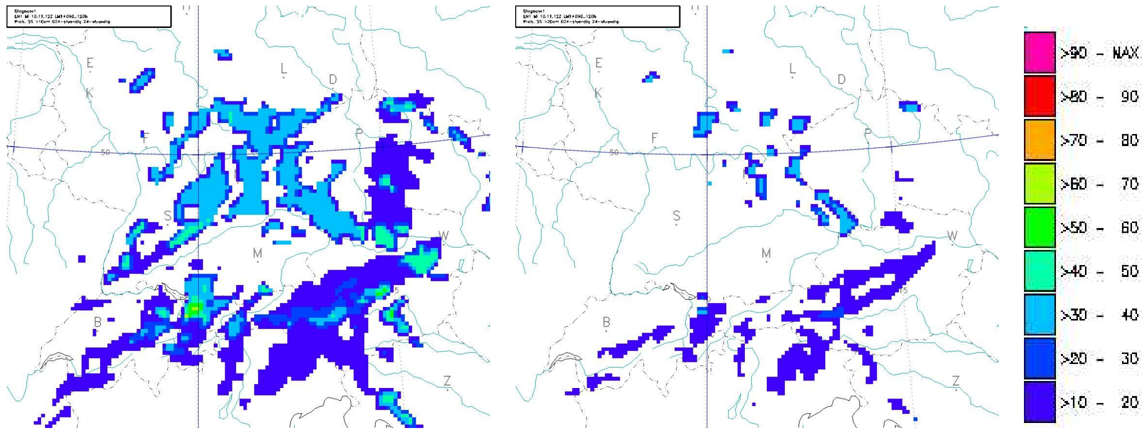


Figure 32: 120 hr-forecast, verifying at the 10 Nov 2004, 12 UTC. Left: Threshold 10 centimeter, right: 20 centimeter fresh fallen snow.

early in the weather reports, from a forecasters point of view this early winter outbreak was not a surprise.

The signal has been developed by the forecasts of the following days (see Fig. 33). The closer the event comes, the more the pattern has been shifted into northwesterly direction. Comparing against observations this was correct. The maximum of snow has been predicted already during the medium range. The orographic representation of the probabilities shown here even allows predictions that in the Upper Rhine valley, the lower-evaluated part of the Rhein-Main-area and the Wetterau snow will be unlikely or the snow depth will not exceed a few centimeter as it has been observed close to the Danube river or in the Hunsrück highlands.

This case study demonstrates a successful forecast of COSMO–LEPS how the system is working and how it could be used to improve the forecasters guidance. During the majority of the severe weather cases the signals provided by COSMO–LEPS were less developed and consequently the forecasts less helpful.

7.2 Conclusions

The COSMO–LEPS system did not always meet forecasters expectations. The guidance provided by COSMO–LEPS to predict the temperature minimum of the day and convective events (related gusts and heavy precipitation) is poor and useful in a few cases only. COSMO–LEPS will add value to the prediction of large scale precipitation and wind gusts if the causing event is a synoptic or larger scale one. The regional assignment is mostly realistic, orographic effects are well represented, weak, but realistic and useful signals often have been shown already during the medium range. The majority of severe weather warnings will be caused by smaller scale events related often with heavy convection. The prediction of such events needs to be improved as well by COSMO–LEPS as by the driving ECMWFs EPS and the nested LM. A small number of cases does not allow final conclusions, further verification (continuous or case-sensitive) is required. A problem is the re-calculation of the COSMO–LEPS forecasts for “historical” weather events again. This might be too expensive. An archiving of the set of derived probability forecasts in the MARS archive of the ECMWF would be helpful.

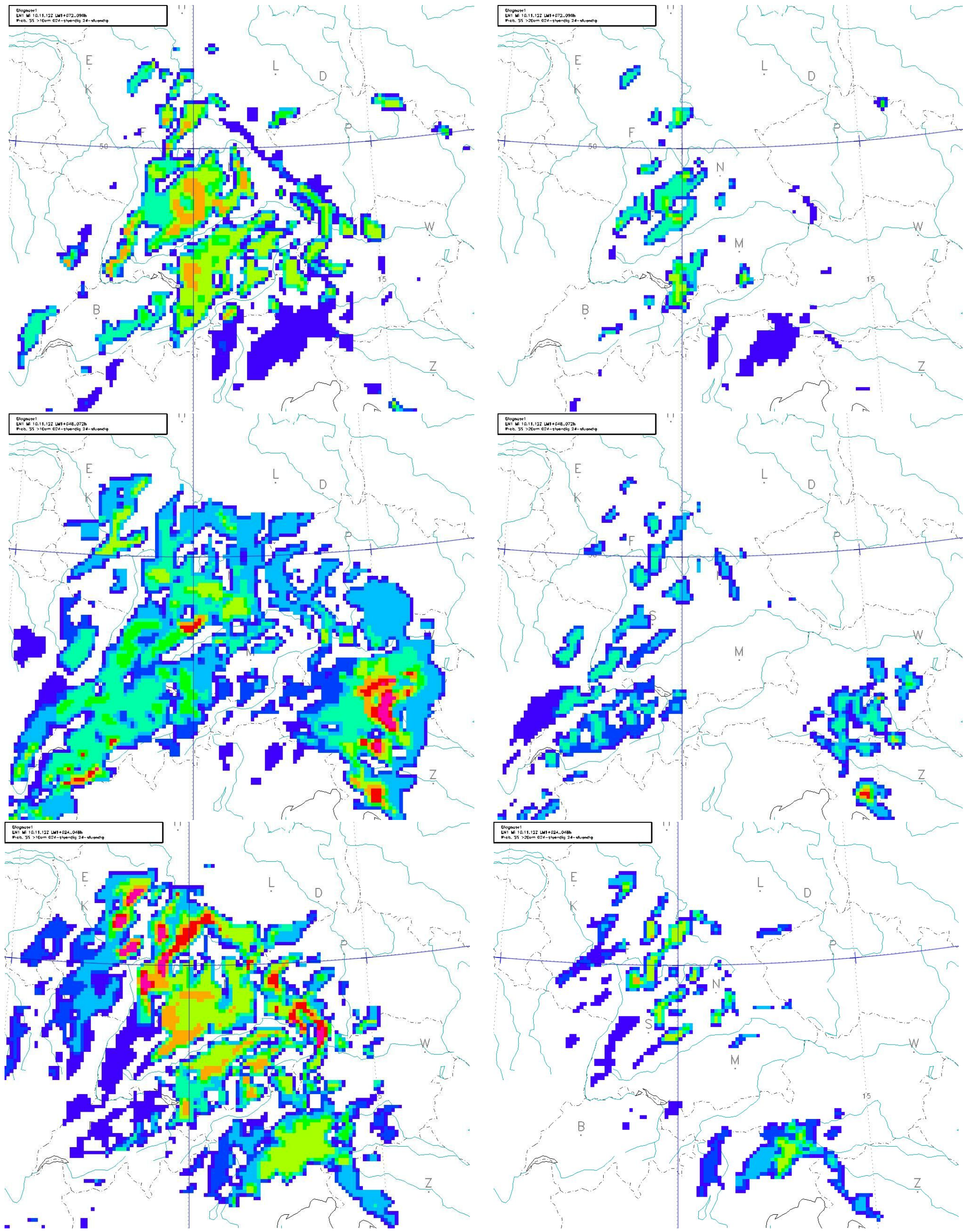


Figure 33: 96 hr-forecast (top), 72 hr-forecast (center) and 48 hr-forecast (bottom) verifying at the 10 Nov 2004, 12 UTC. Left: Threshold 10 centimeter, right: 20 centimeter fresh fallen snow. Color scheme is the same as in Fig. 32.

8 Conclusions

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 and 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. Furthermore, the other COSMO partners have submitted the systems to a careful subjective evaluation, trying to assess the usefulness in a quasi–operational framework.

With regard to the objective verification, in this report the COSMO–LEPS performances in forecasting precipitation has been presented. The period considered is September–November 2003 and forecast precipitation cumulated over 24 hours 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 1.5×1.5 degrees. The observations are also aggregated over the same boxes.

A comparison in terms of average precipitation values over 1.5×1.5 degree 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 is a clear indication of the usefulness of the system in forecasting intense and localized events.

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.

A comparison of the COSMO–LEPS scores with and without the application of the weighted procedure has not been presented in this report. Results (not shown) indicate that no benefit is added by weighting the COSMO–LEPS members according to the cluster population. These results were obtained in terms of precipitation, verified against observations. Nevertheless, a positive impact is evident in the subjective verification performed by ARPAL, where it is shown that weighting the members improves the COSMO–LEPS performances, both in terms of intensity and localization of the precipitation, especially for moderate and intense precipitation.

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 led 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 selected 10 RMs. The 10 Lokal Modell runs are performed by using both the Tiedtke and Kain–Fritsch schemes for the parameterization 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.

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.

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.

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.

Marsigli, C., Boccanera F., Montani, A., Paccagnella, T., The COSMO–LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlinear Processes in Geophysics*, in print, 2005.

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.

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

Stanski, H. R., Wilson, L. J. and Burrows, W. R., Survey of common verification methods in meteorology, *WMO World Weather Watch Tech. Rep.*, 8, pp. 144, 1989

Talagrand, O., Vautard, R. and Strauss, B, 1999, Evaluation of probabilistic prediction systems, *Proceedings of the ECMWF workshop on predictability*, 20–22 October 1997, Reading, UK, pp. 372, 1999.

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

List of COSMO Newsletters and Technical Reports

(available for download from the COSMO Website: www.cosmo-model.org)

COSMO Newsletters

No. 1: February 2001.

No. 2: February 2002.

No. 3: February 2003.

No. 4: February 2004.

No. 5: April 2005.

COSMO Technical Reports

No. 1: Dmitrii Mironov and Matthias Raschendorfer (2001):
Evaluation of Empirical Parameters of the New LM Surface-Layer Parameterization Scheme. Results from Numerical Experiments Including the Soil Moisture Analysis.

No. 2: Reinhold Schrodin and Erdmann Heise (2001):
The Multi-Layer Version of the DWD Soil Model TERRA_LM.

No. 3: Günther Doms (2001):
A Scheme for Monotonic Numerical Diffusion in the LM.

No. 4: Hans-Joachim Herzog, Ursula Schubert, Gerd Vogel, Adelheid Fiedler and Roswitha Kirchner (2002):
LLM ⁻ the High-Resolving Nonhydrostatic Simulation Model in the DWD-Project LITFASS.
Part I: Modelling Technique and Simulation Method.

No. 5: Jean-Marie Bettems (2002):
EUCOS Impact Study Using the Limited-Area Non-Hydrostatic NWP Model in Operational Use at MeteoSwiss.

No. 6: Heinz-Werner Bitzer and Jürgen Steppeler (2004):
Documentation of the Z-Coordinate Dynamical Core of LM.

No. 7: Hans-Joachim Herzog, Almut Gassmann (2005):
Lorenz- and Charney-Phillips vertical grid experimentation using a compressible nonhydrostatic toy-model relevant to the fast-mode part of the 'Lokal-Modell'

No. 8: Chiara Marsigli, Andrea Montani, Tiziana Paccagnella, Davide Sacchetti, André Walser, Marco Arpagaus, Thomas Schumann (2005):
Evaluation of the Performance of the COSMO-LEPS System

COSMO Technical Reports

Issues of the COSMO Technical Reports series are published by the *Consortium for Small-Scale Modelling* at non-regular intervals. COSMO is a European group for numerical weather prediction with participating meteorological services from Germany (DWD, AWGeophys), Greece (HNMS), Italy (UGM, ARPA-SMR) and Switzerland (MeteoSwiss). The general goal is to develop, improve and maintain a non-hydrostatic limited area modelling system to be used for both operational and research applications by the members of COSMO. This system is initially based on the Lokal-Modell (LM) of DWD with its corresponding data assimilation system.

The Technical Reports are intended

- for scientific contributions and a documentation of research activities,
- to present and discuss results obtained from the model system,
- to present and discuss verification results and interpretation methods,
- for a documentation of technical changes to the model system,
- to give an overview of new components of the model system.

The purpose of these reports is to communicate results, changes and progress related to the LM model system relatively fast within the COSMO consortium, and also to inform other NWP groups on our current research activities. In this way the discussion on a specific topic can be stimulated at an early stage. In order to publish a report very soon after the completition of the manuscript, we have decided to omit a thorough reviewing procedure and only a rough check is done by the editors and a third reviewer. We apologize for typographical and other errors or inconsistencies which may still be present.

At present, the Technical Reports are available for download from the COSMO web site (www.cosmo-model.org). If required, the member meteorological centres can produce hard-copies by their own for distribution within their service. All members of the consortium will be informed about new issues by email.

For any comments and questions, please contact the editors:

Ulrich Schättler
ulrich.schaettler@dwd.de

Andrea Montani
amontani@smr.arpa.emr.it