Verification of the COSMO–LEPS new Suite in Terms of Precipitation Distribution

C. Marsigli, A. Montani, T. Paccagnella

ARPA–SIM, Bologna, Italy

1 Introduction

An objective verification of the mesoscale ensemble system COSMO–LEPS is carried out. This is a 10 member ensemble based on the Limited-Area Model *Lokal Modell* (LM), running daily at ECMWF since November 2002 using the resources shared by the COSMO countries. The system is based on the ECMWF EPS, which provides the perturbations to the limited-area model runs through the initial and boundary conditions. Mesoscale perturbations are also added by letting the different LM run to randomly choose the scheme to be used for the parameterisation of the deep convection (Tiedtke or Kain–Fritsch).
The system has recently been updated to the current configuration (June 2004) and the verification of the new suite is presented here.
The COSMO–LEPS system transfers the EPS probabilistic approach to the scales where a better representation of mesoscale–related processes permits to forecast surface parameters with a greater detail. Then, the verification of the system focuses on features where the impact of the high–resolution is dominant. Precipitation is chosen as the verification parameter, trying to evaluate the ability of the system in forecasting the detailed structure of this field by considering a number of the parameter of the precipitation distribution in the verification process.
The probabilistic verification tools considered here are: Relative Operating Characteristic (ROC Curves, Mason and Graham, 1999), Brier Score and Brier Skill Score (Stanski et al., 1989, Wilks, 1995), Cost–Loss Analysis (Richardson et al., 2000) and Percentage of Outliers (Talagrand et al., 1997, Buizza, 1997). For a brief description of the indices the reader is referred to Marsigli et al. (2004). Though the computation of these scores is rather simple, their interpretation is not straightforward, different indices describing different features of the forecast system. In addition to this, the relationship between these scores is not a linear one. Therefore, a global evaluation of the forecast system should rely on a set of indices. In this report, for brevity reasons, only results in terms of the Brier Skill Score, ROC area and Percentage of Outliers will be presented.
After a brief description of the system (Section 2), the verification methodology is presented (Section 3), followed by the results (Section 4).

2 The COSMO–LEPS operational system

The limited–area ensemble prediction system COSMO–LEPS has been running operationally at ECMWF since November 2002. The suite is run and maintained remotely by ARPA–SIM, with support given by ECMWF, and the necessary Billing Units are made available by the ECMWF COSMO countries.
The system has recently been updated (June 2004) and now 10 runs of the non–hydrostatic limited–area model (LM) are available every day, nested on 10 selected members (the so–called Representative Members, or RMs) of two consecutive 12–hour lagged ECMWF global ensembles. The 10 selected members are representative of 10 clusters, built by grouping all

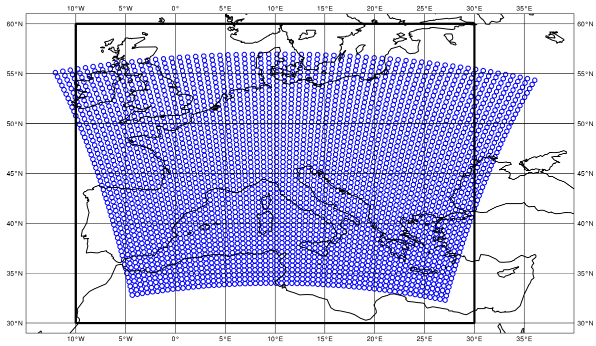

Figure 1: COSMO–LEPS operational domain (small circles) and clustering area (big rectangle).

the global ensemble members on the basis of their similarity in terms of upper–air fields. Mesoscale perturbations are also added by letting the different LM runs randomly choose the scheme to be used for the parameterisation of the deep convection (Tiedtke or Kain–Fritsch). A description of the old suite and the motivation for the suite update are described in Marsigli et al. (2005).
The limited–area ensemble forecasts range up to 120 hours and are integrated over a domain covering all the countries involved in COSMO (Fig. 1). The model version is 3.9 (prognostic precipitation and cloud ice scheme have been activated), the horizontal resolution is about 10 km and 32 vertical layers are used. LM–based probabilistic products covering a "short to medium–range" (48–120 hours) are disseminated to the weather services involved in COSMO.

3 Verification methodology

Verification is performed in terms of daily precipitation, cumulated from 06 to 06 UTC. Precipitation observations are available on the very dense COSMO station network (over 4500 stations, see Fig. 2) covering Germany, Switzerland, Poland and part of Italy.

For verification purposes, the verification area is covered by 12 × 11 boxes of 1.5 × 1.5 degrees (approximately equal to 150 km × 150 km) and a pair of *representative* observed and forecast values is individuated for each box. This approach is followed to permit the comparison between a punctual value (the observation) and an areal value (the forecast). Several observed and several forecast values fall in each box and a comparison between the distribution of the observed and the distribution of the forecast values is attempted. A number of statistical properties of the two distributions are computed over each box, thus allowing a comparison between forecast and observed values which are representative of different features of the precipitation distribution. In this work, the average, the median (50th percentile), the 90th percentile and the maximum are computed.

COSMO–LEPS performances are compared with those of the ECMWF EPS, both the operational full–size 51–member EPS and the reduced–size 10–member EPS made up by the 10 selected Representative Members. The comparison of COSMO–LEPS with this 10–RM–EPS permits to quantify the impact of the high–resolution alone, irrespective of the different number of members of the two operational systems. The need to compare the 10–km COSMO–LEPS with the 80–km EPS determines the choice of boxes as big as 1.5 × 1.5 degrees, in order to have enough EPS grid points within a box to make possible the computation of the statistical properties. The three systems will be referred to as:

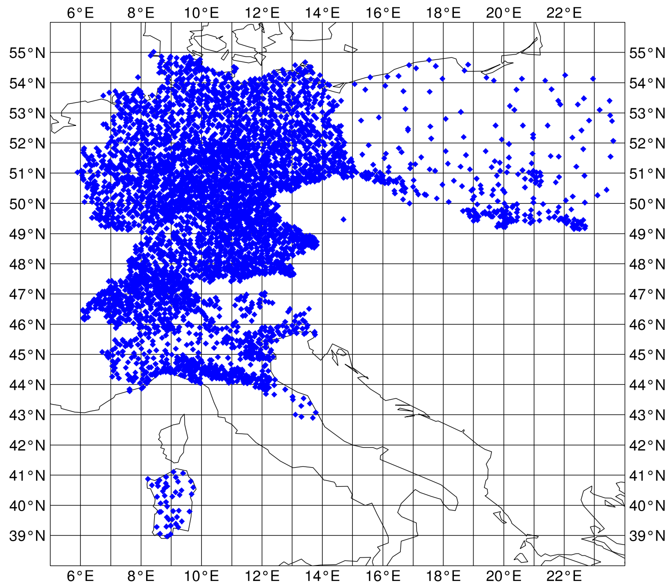

Figure 2: COSMO network of stations were observed precipitation is available. Precipitation data are cumulated over 24 hours from 06 to 06 UTC.

- `cleps`: COSMO-LEPS, 10 members, 10 km hor. res.
- `epsrm`: reduced EPS made up by the RMs, 10 members, 80 km hor. res.
- `eps51`: operational EPS starting at 12 UTC, 51 members, 80 km hor. res.

The period considered for verification is Autumn (September, October and November) 2004. Verification has been performed in terms of 24–hour cumulated precipitation (from 06 to 06 UTC).

4 Results

Results are presented in terms of a set of indicators: Brier Skill Score (BSS), ROC area, Talagrand diagrams and Perecntage of Outliers. For the Brier Skill Score (Stanski et al., 1989) a higher value corresponds to a better results and the zero level indicates the limit of usefulness of the forecasting system. The ROC area (Mason and Graham, 1999) can take values in the range [0,1], the higher the better, and the no–skill level is 0.5. The Talagrand diagram (Talagrand et al., 1999) is obtained by counting how many times the truth falls in each of the bins that are obtained by putting the forecast values in increasing order. An U–shape of the diagram indicates that the ensemble is underdispersive, while a dome shape indicates that is overdispersive. The best shape is the uniform distribution. The Percentage of Outliers (Talagrand et al., 1999, Buizza, 1997) indicates the percentage of times the truth falls outside from the range of the forecast values, so it sums up the informations coming form the two extreme bins of the Talagrand diagrams.

Average values

Results from the verification in terms of average values exceeding 10mm/24h over the 1.5×1.5 degree boxes are presented in Fig. 3.

In terms of BSS (left panel), `cleps` is performing worse than `epsrm`, while the two systems are comparable in terms of ROC area (right panel). This indicates that as regards the total

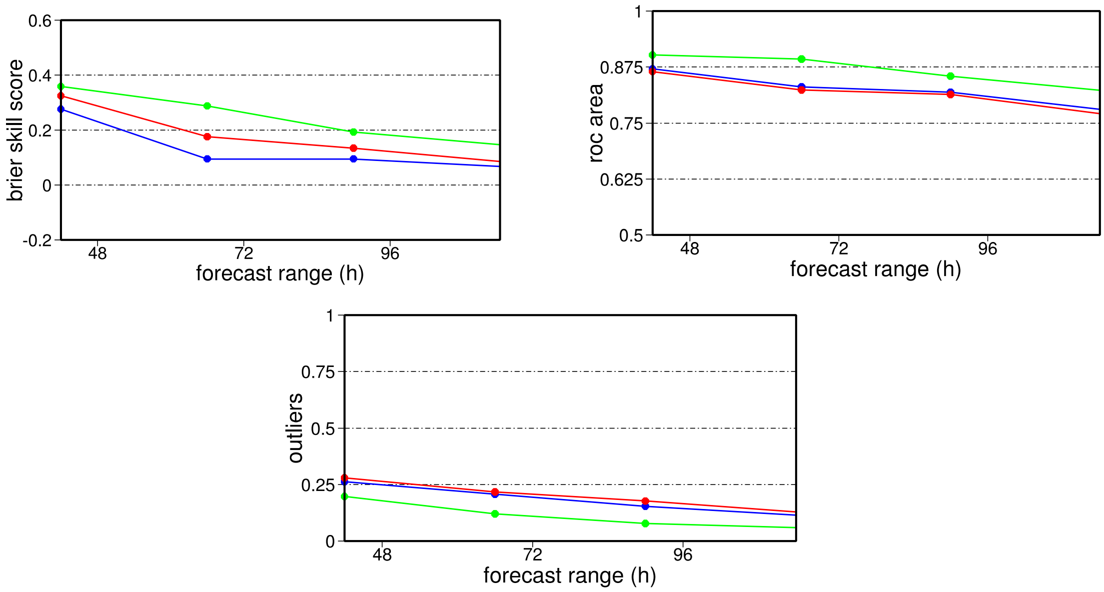

Figure 3: Brier Skill Score (top left panel) ROC area (top right panel) and Percentage of Outliers (bottom panel) as a function of the forecast range (in hours) relative to the 24–hour cumulated precipitation forecasts by COSMO–LEPS (`cleps`, blue line), by the 10–RM EPS (`epsrm`, red line) and by the operational 51–member EPS (`eps51`, green line) for the 10mm precipitation thresholds. For each box, the mean of the forecast values is compared with the mean of the observed forecast values.

amount of precipitation falling within a box, the information provided by the global ensemble is enough for this threshold and even better in terms of reliability. Even in terms of outliers, the high–resolution system is not lowering the percentage (bottom panel).

The use of the operational EPS (`eps51`) seems the best solution for this quantity over an area as big as 150 $\times$ 150 km^2. It is not possible to repeat the verification by decreasing the dimension of the area because of the low resolution of the EPS.

In order to give an idea of what the spread of the considered ensemble is, the Talagrand diagrams are also presented (Fig. 4).

The marked U–shape of the diagram relative to the `eps51` system is evident (bottom panel), indicating underdispersion, even though the total number of outliers is smaller than for the oher two systems. It is also evident that the truth is more frequently in the upper tail of the distribution, both for `eps51` and for `epsrm` (upper right panel), while `cleps` provide a more uniform distribution (upper left panel).

Median (50th percentile) values

A comparison in terms of the median values has also been carried out but the results are comparable to those obtained for the mean values, so they are not shown.

90th percentile values

In the COSMO–LEPS verification, we are mainly interested in the tail of the precipitation distribution over an area, where the lower resolution system is supposed to show some deficiencies and the use of the mesoscale model would permit more realistic precipitation structures with higher peak values in case of intense precipitation. For this purpose, a verification in terms of the 90th percentile of the precipitation distribution has also been performed

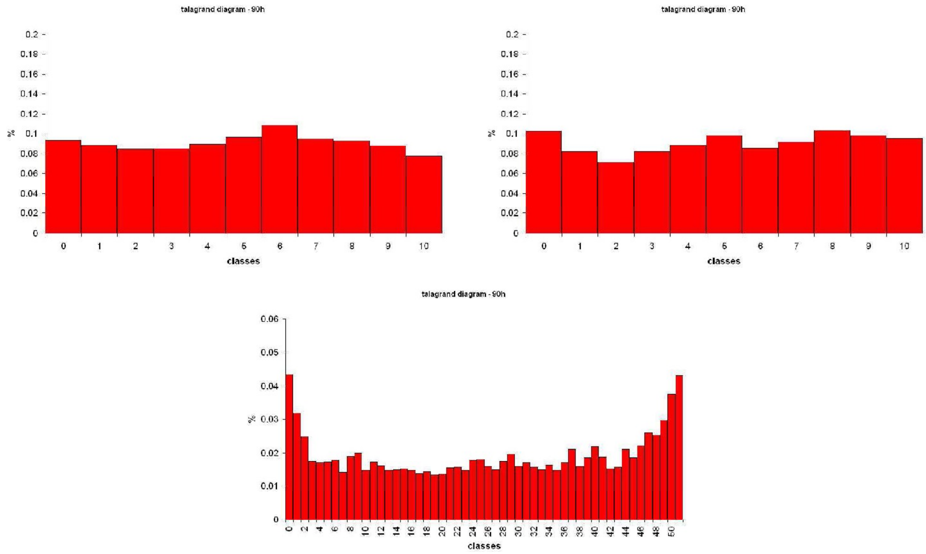

Figure 4: Talagrand diagrams relative to `cleps` (top left panel) `epsrm` (top right panel) and `eps51` (bottom panel) computed for average values within boxes at the 90 hour forecast range. The scale of the y–axis is different for the `eps51` system.

(Fig. 5) for the threshold 20mm/24h.

In terms of BSS (left panel) the three systems exhibit similar performances, with `eps51` still showing slightly higher values, while in terms of ROC area (right panel) the `cleps` system overperforms `epsrm` by a large amount and, to a lesser extent, also `eps51`. The percentage of outliers (bottom panel) produced by `cleps` is lower than that of `epsrm`, while `eps51` has the lowest value but a direct comparison with the other two is not fair due to the very different number of ensemble members. Form this kind of verification it appears that, when considering the tail of the precipitation distribution over a box, the high resolution of the COSMO–LEPS system plays an important role, which is not rewarded when considering average values.

The Talagrand diagrams are also presented (Fig. 6).

For this parameter, `cleps` exhibits a tendency to forecast too high values (top left panle), the upper tail of the distribution being slightly less populated, while the pronounced U–shape provided by `eps51` is still evident (bototm panel), with a tendency to predict too lower values, which is even more evident in `epsrm` (top right panel).

Maximum values

The capability of the COSMO–LEPS system to signal the possibility of the occurrence of very large precipitation over an area can be quantified by repeating the verification in terms of maximum values over a box. Results are shown in Fig. 7 for the threshold 50mm/24h.

In terms of both scores (BSS in the left panel and ROC area in the right panel) `cleps` is overperforming the lower resolution systems, showing a positive skill in forecasting the occurrence of heavy precipitation over an area. The percentage of outliers (bottom panel) is also shown.

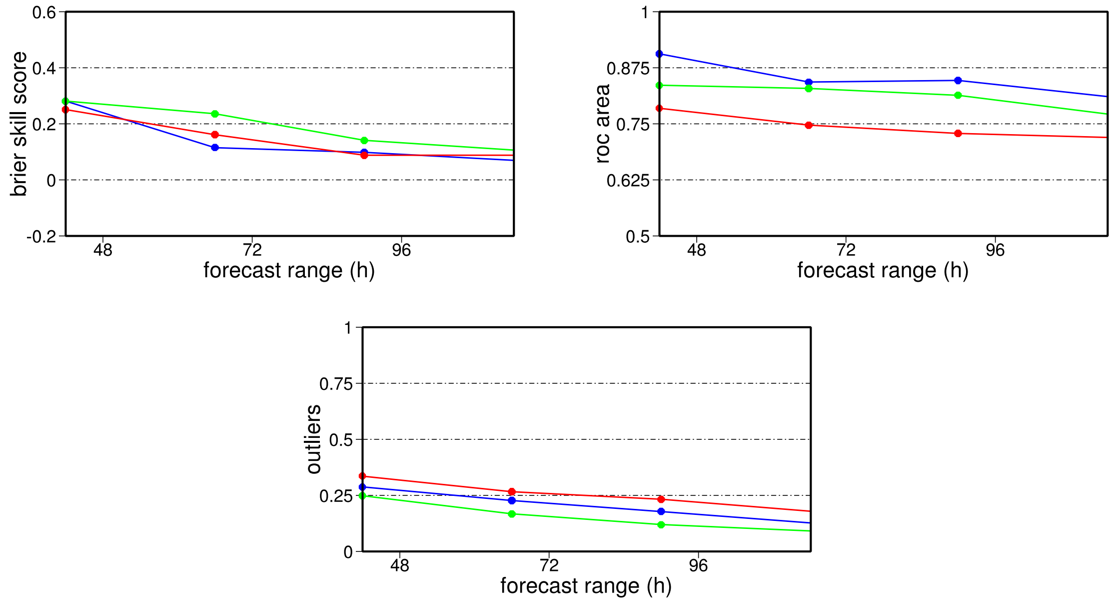

Figure 5: Brier Skill Score (top left panel) ROC area (top right panel) and Percentage of Outliers (bottom panel) as a function of the forecast range (in hours) relative to the 24–hour cumulated precipitation forecasts by COSMO–LEPS (`cleps`, blue line), by the 10–RM EPS (`epsrm`, red line) and by the operational 51–member EPS (`eps51`, green line) for the 20mm precipitation thresholds. For each box, 90^{th} percentile of the forecast values is compared with the 90^{th} percentile of the observed forecast values.

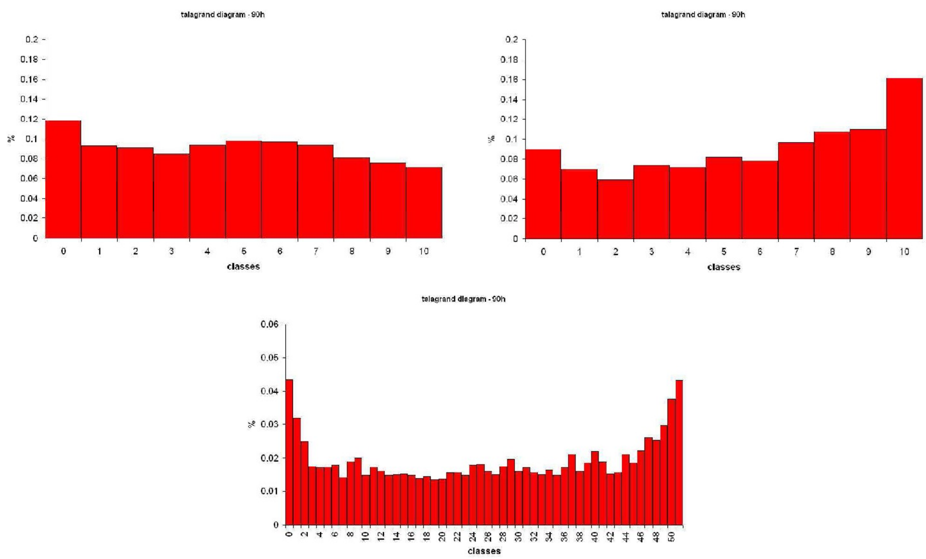

Figure 6: Talagrand diagrams relative to `cleps` (top left panel) `epsrm` (top right panel) and `eps51` (bottom panel) computed for 90^th percentile values within boxes at the 90 hour forecast range. The scale of the y–axis is different for the `eps51` system.

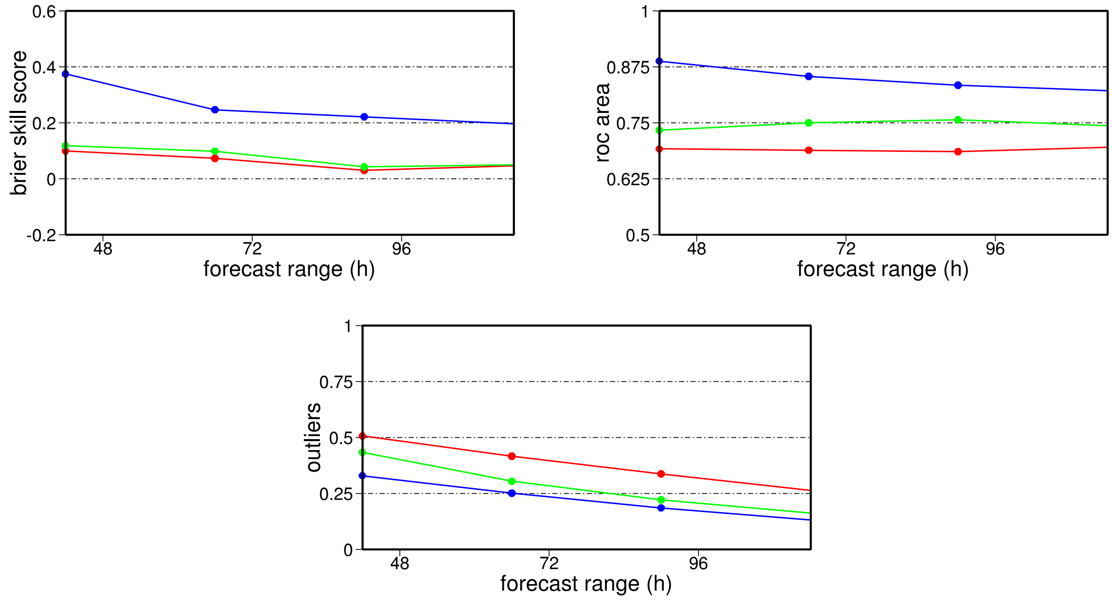

Figure 7: Brier Skill Score (top left panel) ROC area (top right panel) and Percentage of Outliers (bottom panel) as a function of the forecast range (in hours) relative to the 24–hour cumulated precipitation forecasts by COSMO–LEPS (`cleps`, blue line), by the 10–RM EPS (`epsrm`, red line) and by the operational 51–member EPS (`eps51`, green line) for the 20mm precipitation thresholds. For each box, the maximum of the forecast values is compared with the maximim of the observed forecast values.

5 Conclusions

An objective verification of the COSMO–LEPS system in terms of precipitation distribution has been shown. Results indicate that the COSMO–LEPS system is useful for the forecast of intense precipitation over an area, allowing a good description of the tail of the precipitation distribution (90^th percentile) and permitting to capture the occurrence of high precipitation values (maximum). As regards average values over an area, the best performance is obtained by the operational full-size EPS.

A direct comparison between the old (5–member 3–EPS) suite and the new (10–member 2–EPS) suite is not possible, because they were never run in parallel on a common period. Results obtained with the old suite for Autumn 2003 (Marsigli et al., 2005) are not comparable with the results here presented for Autumn 2004. Nevertheless, more or less the same conclusions can be drawn from the results obtained in 2003 and in 2004.

References

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

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, 2001. *Quart. J. Roy. Meteor. Soc.*, 127, pp. 2095–2115.

Marsigli, C., Boccanera, F., Montani, A., Nerozzi, F., Paccagnella, T., COSMO–LEPS verification: first results, 2004. *COSMO Newsletter*, No. 4, pp. 125–135.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T., The COSMO–LEPS ensemble system: validation of the methodology and verification, 2005. *Nonlinear Processes in Geophysics*, 12, pp. 527–536.

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

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

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T. and Buizza, R., 2001. Performance of ARPA–SMR Limited–area Ensemble Prediction System: two flood cases. *Nonlinear Processes in Geophysics*, 8, pp. 387–399.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi, S. and Buizza R., 2003a. The Soverato flood in Southern Italy: performance of global and limited–area ensmeble forecasts *Nonlinear Processes in Geophysics*, 10, pp. 261–274.

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, 2003b. *ECMWF Newsletter Summer 2003*, 98, pp. 2–7.

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

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

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

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