Consortium

for

Small-Scale Modelling

Technical Report No. 19

Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges

June 2010

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_19

Deutscher Wetterdienst

MeteoSwiss

Ufficio Generale Spazio Aereo e Meteorologia

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ

Instytucie Meteorogii i Gospodarki Wodnej

Administratia Nationala de Meteorologie

Agenzia Regionale per la Protezione Ambientale del Piemonte

Agenzia Regionale per la Protezione Ambientale dell' Emilia-Romagna: Servizio Idro Meteo

Centro Italiano Ricerche Aerospaziali

Amt für GeoInformationswesen der Bundeswehr

www.cosmo-model.org

Editor: Massimo Milelli, ARPA Piemonte
Printed at Deutscher Wetterdienst, P.O. Box 100465, 63004 Offenbach am Main

Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges

A. Montani, D. Cesari, C. Marsigli

and T. Paccagnella

ARPA-SIMC
Viale Silvani 6
Bologna
Italy

Contents

1 Abstract

In this work, the main characteristics of COSMO–LEPS, the Limited–area Ensemble Prediction System developed in the framework of the COnsortium for Small-scale MOdelling, are presented. The present status of the system is shown with the description of the methodology and of the main upgrades which took place during its years of activity. The performance of COSMO–LEPS for the probabilistic prediction of precipitation is assessed in terms of both time–series and seasonal scores over a 7–year period. A fixed number of stations is selected and observations are compared to short and early–medium–range forecasts. Different verification indices are used to assess the skill of COSMO–LEPS and to identify the impact of system modifications on forecast skill. The different system upgrades are found to impact positively on COSMO–LEPS performance, with a gain of two days of predictability in the last four years of operational forecasts. This holds when the skill of the system is assessed both for single events (e.g. precipitation surpassing a fixed threshold) and for multi–event situations. Scores for fixed forecast ranges but varying thresholds confirm increasingly better performance of the system. Finally, the main stream of development for COSMO–LEPS system are discussed with future possible upgrades and methodology modifications.

2 Introduction

One of the main challenges for numerical weather prediction (NWP) is still recognised as quantitative precipitation forecasting. Computer power resources have greatly increased in the last years, thus allowing the generation of more and more sophisticated NWP model with accurate parametrisation of physical processes supported by high horizontal and vertical resolution. Nevertheless, the accurate forecast of high–impact weather still remains difficult beyond day 2 and sometimes, also for shorter ranges (Mullen and Buizza, 2001; Tibaldi et al., 2006). Several factors contribute to forecast failures and can be usually related to shortcomings in the definition of the initial conditions of the integrations, to model errors of different types and, last but not least, to the intrinsic low predictability of the physical phenomena under investigation. The use of the probabilistic approach via the ensemble forecasting has now become commonplace to tackle the chaotic behaviour of the atmosphere and to support forecasters in the management of alert procedures for events with little deterministic predictability. Several national and international weather centres, like the European Centre for Medium–Range Weather Forecasts (ECMWF), the Canadian Meteorological Centre (CMC), the National Centers for Environmental Prediction (NCEP) and the UK Meteorological Office, provide valuable operational ensemble prediction at a global scale (Tracton and Kalnay, 1993; Houtekamer et al., 1996; Molteni et al., 1996; Buizza et al., 2007; Bowler et al., 2008). In addition to them, many limited–area ensemble prediction systems have been recently developed, either in research or in operational mode, so as to address the need of detailing high–impact weather forecasts at higher and higher resolution and to support the reliability of the deterministic forecast beyond the very short range.

As far as operational implementations are concerned, the Consortium for Small–Scale MOdelling Limited–area Ensemble Prediction System (COSMO–LEPS) was the first mesoscale ensemble application running on a daily basis in Europe[1]. This system, initially developed and implemented by the HydroMeteoClimate Regional Service of Emilia-Romagna, in Bologna, Italy (ARPA–SIMC), has been running at ECMWF since November 2002 (Montani et al., 2003a) thanks to the ECMWF computer resources provided by the COSMO

[1]Information about the activities of the COSMO consortium can be found at http://www.cosmo-model.org.

countries which are ECMWF Member states. Nowadays, COSMO–LEPS is based on 16 integrations of the non–hydrostatic mesoscale model COSMO, formerly known as the Lokal Modell (Steppeler et al., 2003). The methodology (described more thoroughly in the next section) aims at combining the advantages of the probabilistic approach by global ensemble systems with the high–resolution details gained in the mesoscale integrations. In the construction of COSMO–LEPS, an algorithm selects a number of members (referred to as Representative Members, RMs) by a global ensemble system (Marsigli et al., 2001; Molteni et al., 2001). This intermediate step, referred to as "ensemble–size reduction", is required to keep the computational load operationally affordable, since it is not presently feasible to nest the limited-area model on each individual member of a highly–populated global ensemble. After the "ensemble–size reduction", the selected RMs are used to provide both initial and boundary conditions to the integrations with the COSMO model, which is run once for each RM. Therefore, COSMO–LEPS performs a sort of dynamical downscaling of a global–model probabilistic system, limiting to a certain extent the computation investment (Tibaldi et al., 2006).

In this work, the progresses in the development of COSMO–LEPS are reviewed, focusing on past and recent system upgrades and on the impact on its forecast skill. It is aimed to quantify the improvements of COSMO–LEPS in terms of probabilistic prediction of light and heavy precipitation events in the short and early–medium range. This will enable to indicate both strengths and weaknesses of the system, to highlight the open challenges and to suggest the main streams of future developments.
The rest of the paper is organised as follows: section 3 describes the main characteristics of the COSMO–LEPS system, while, in section 4, the features of the verification procedure are reported. Results in terms of both time–series and seasonal scores are presented in sections 5 and 6, respectively. Finally, conclusions are drawn in section 7.

3 Description of COSMO-LEPS operational system

As previously mentioned, the "core" of COSMO–LEPS methodology lies in the idea of reducing the number of global–ensemble elements driving the limited–area runs, still retaining a large fraction of the driving–ensemble information. In its first experimental–operational implementation which dates November 2002, the set–up of COSMO–LEPS can be described as follows (Montani et al., 2003b):

1. three successive runs of ECMWF Ensemble Prediction System (EPS), starting at 12UTC at day *d-1*, at 00UTC and 12UTC of day *d* are joined together, thus generating a 153–member lagged–ensemble, since each EPS is made up of one control run plus 50 perturbed members (Buizza, 2005);

2. EPS members are grouped into 5 clusters, the discriminating variable being a combination of four variables at three pressure levels and at 2 forecast steps: the two horizontal wind components, the geopotential height and the specific humidity at 500, 700, 850 hPa and at the ranges of 96 and 120 hours (the ranges are relative to the "youngest" ensemble, run at 12UTC of day *d*);

3. for each variable at each forecast step, the mean over the clustering area is calculated and, then, subtracted from any grid-point value. Hence, the result is divided by the standard deviation, thus obtaining a non–dimensional field;

4. the quadratic distances among the EPS members are computed for all variables at all levels at all steps and, then, space–averaged;

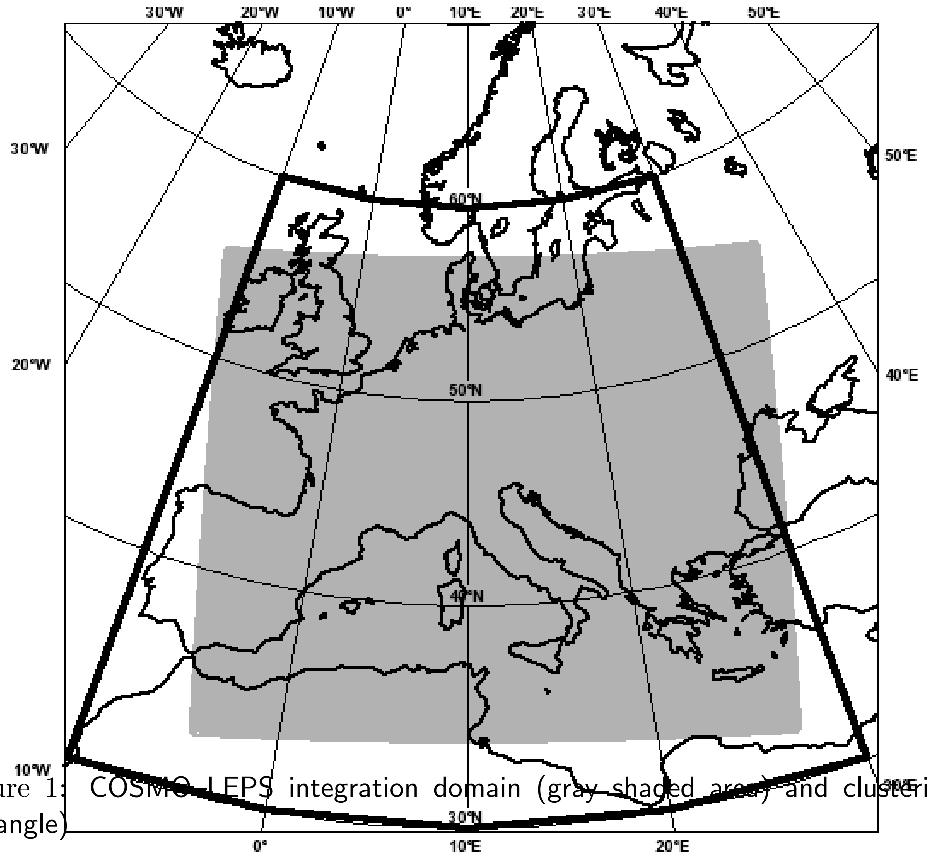

Figure 1: COSMO–LEPS integration domain (gray-shaded area) and clustering area (black rectangle).

5. the cluster analysis is performed over the following area: 40N–60N, 10W–30E (denoted by the black rectangle of Fig. 1); the clusters are constructed using the complete–linkage algorithm (Wilks, 1995);

6. within each cluster (with different populations), one RM is selected, using the same discriminating variables as before; the RM is that cluster element which minimises the ratio between its distance from the other members of its own cluster and its distance from the members of the other clusters;

7. the so–selected RMs provide both initial and boundary conditions for the integrations with the COSMO model, which is run once for each RM over a domain covering Central and Southern Europe (shaded area in Fig. 1);

8. the 5 COSMO integrations which generate the COSMO–LEPS system, start at 12UTC of day *d*, with a horizontal resolution of 10 km, 32 vertical levels and a forecast length of 120 hours;

9. post–processed products (e.g. the probability of exceeding a threshold) are generated, assuming a relationship between cluster population and the probability of occurrence of its associated RM; hence each COSMO integration is given a weight proportional to the population of the cluster from which the RM (providing initial and boundary conditions) was selected.

In the course of its operational activity, the COSMO–LEPS system has undergone many changes. Without accounting for the model upgrades with bug fixes and more sophisticated and precise parametrisation, the main milestones of COSMO–LEPS history can be summarised as follows:

- 4 November 2002: beginning of COSMO–LEPS activity in the above–described configuration;
- 1 June 2004 ([**A**]): ensemble population increased to 10 members; the RMs are selected out of 102 EPS members (only 2 successive EPS runs starting at 00UTC and 12UTC of day *d* are used); random procedure for the choice of either Kain-Fritsch or Tiedtke convection scheme with each COSMO–LEPS run;
- 1 July 2005: the forecast range of COSMO–LEPS integrations is increased from 120 to 132 hours;
- December 2005; COSMO–LEPS application becomes an "*ECMWF Member State time–critical application*" managed by ARPA–SIMC and monitored by ECMWF operators;
- 2 February 2006 ([**B**]); ensemble size increased from 10 to 16 members; vertical resolution increased from 32 to 40 model levels;
- 1 December 2007 ([**C**]); change of the numerical scheme in COSMO–LEPS integrations; introduction of new perturbations in the model parametrisations.

In this list, a number of changes are labelled with letters **[A], [B], [C]**: they represent those system upgrades with direct consequences on the performance of the COSMO–LEPS forecast skill, as will be shown later on.
It is worth pointing out that, both in the first years of activity as well as more recently, Marsigli et al. (2005a, 2005b and 2008), Montani et al. (2008a and 2008b) and Walser (2006) showed that COSMO–LEPS provided high–quality probabilistic quantitative prediction of heavy precipitation events. Over a number of case studies as well as over continuous verification periods, COSMO–LEPS was shown to perform usually better than ECMWF EPS in terms of geographical localisations of the regions most likely to be affected by the events as well as in terms of more realistic rainfall patterns. The above works used different verification techniques and different observational datasets; hence, they cannot provide exhaustive information on the evolution of the skill of the system. In order to shed light on the progresses of COSMO–LEPS, a comprehensive verification over the full history of the system is undertaken with the features described in the next section.

4 Methodology of verification

The performance of COSMO–LEPS system is analysed considering the probabilistic prediction of 12–hour accumulated precipitation exceeding a number of thresholds for several forecast ranges. Instead of the "more traditional" 24–hour accumulation period, it was decided to focus the attention on 12–hour precipitation (accumulated from 18 to 6 UTC and from 6 to 18 UTC) in order to investigate the performance of the system for both day–time and night–time precipitation forecasts. This should allow the possibility to isolate possible biases and/or systematic errors in the diurnal cycle of COSMO-LEPS model integrations, which would be otherwise masked if verification were performed over a 24–hour window.

As for observations, it is clear that a high–resolution network would be very desirable in order to assess the predictive skill of a mesoscale ensemble system. Since precipitation has a high-spatial variability, this network would provide better estimates of precipitation at

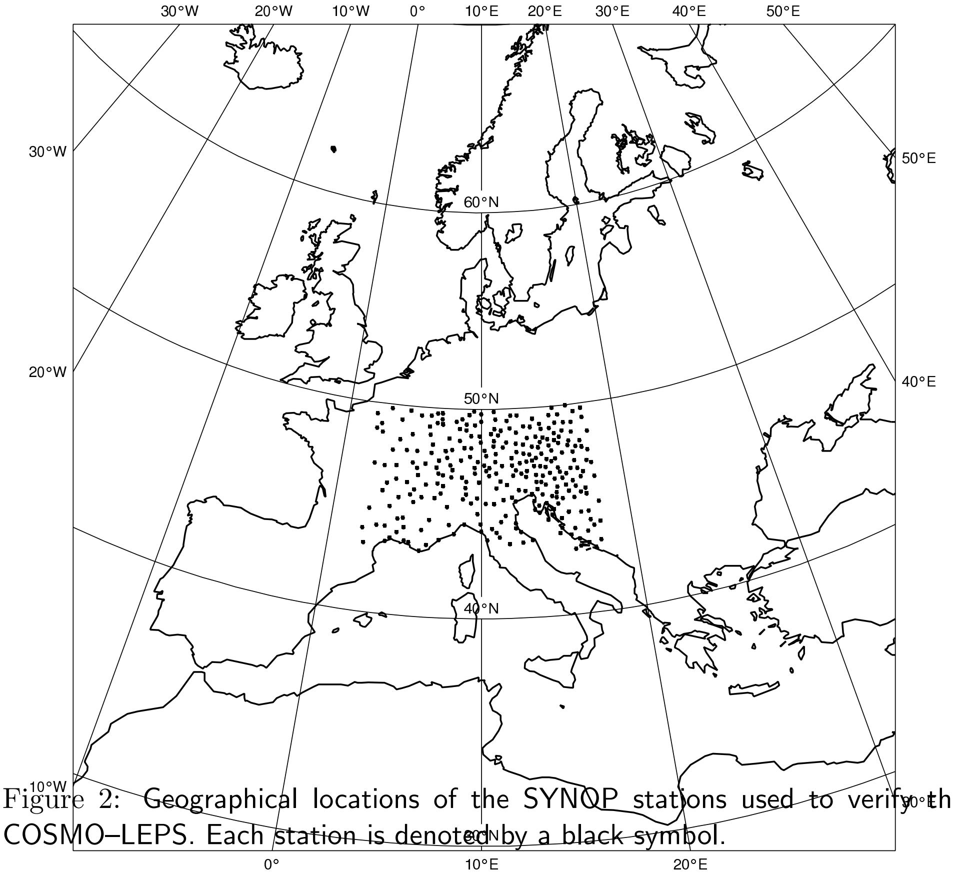

Figure 2: Geographical locations of the SYNOP stations used to verify the performance of COSMO–LEPS. Each station is denoted by a black symbol.

high resolution. Unfortunately, this type of network, like the high–density network adopted by Marsigli et al. (2008), is not available with continuity in the period 2002–2009, as it presents several “gaps”, this making impossible a detailed evaluation of how the performance of COSMO–LEPS evolved in the years. For this reason, it has been decided to use the data obtained from the SYNOP reports available on the Global Telecommunication System (GTS), since this is recognised to be a homogeneous and stable dataset throughout the years 2002–2009.

In order to assess the skill of the system over complex topography, verification is performed in the domain ranging from 43N to 50N and from 2N to 18E. This domain, sometimes referred to as MAP D-PHASE area (Mesoscale Alpine Programme, Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the alpine region), is the common terrain of investigation for the Forecast Demonstration Project which took place during the Operation Period of D-PHASE (Zappa et al., 2008; Rotach et al., 2009). Within this domain, a fixed list of 412 SYNOP stations is considered and the relative reports in terms of total precipitation are used to evaluate the COSMO–LEPS skill.

Fig. 2 reports the locations of the stations used for the verification. The SYNOP reports have undergone a simple quality control firstly based on the “surpassing” of a confidence level (provided in the data retrieved by ECMWF archive) for the full report. In addition to this, for cases of very high precipitation records, the values are compared, whenever possible, to those taken from nearby non–GTS stations. In case of discrepancy between non–GTS and SYNOP reports, the latter is discarded and the relative data not used in the computation of the scores.

As for the comparison of model forecasts against SYNOP reports, we select the grid point

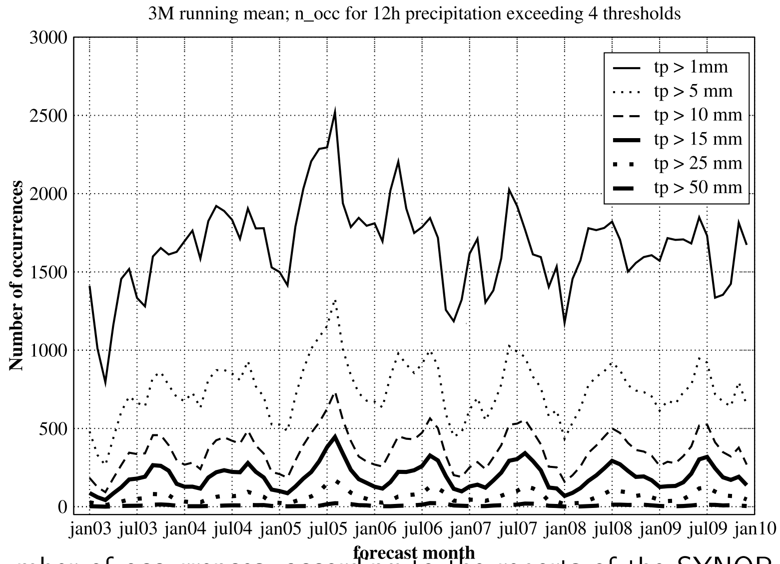

Figure 3: Number of occurrences, according to the reports of the SYNOP stations of Fig. 2, for 12–hour observed precipitation exceeding 4 different thresholds: 1mm/12h (thin–solid line), 5mm/12h (dotted), 10mm/12h (dashed), 15mm/12h (thick–solid), 25mm/12h (thick–dotted) and 50 mm/12h (thick–dashed). A 3–month running mean is applied to improve the readability of the plots.

closest to the observation. Little sensitivity to the results is found when, instead of the nearest grid point, a bi-linear interpolation using the 4 nearest points to the station location, is used to generate the model forecasts. Therefore, the results shown hereafter will be relative only to the nearest grid-point method.
The performance of COSMO–LEPS is examined for 6 different thresholds: 1, 5, 10, 15, 25 and 50 mm/12h.

Fig. 3 shows the overall number of occurrences for each threshold and for each month, from December 2002 to November 2009, according to the reports of the stations of Fig. 2. As the occurrences have a strong marked month–to–month variability (and for ease of comparison with the future plots), a 3–month running mean was applied to improve the readability of the figure. The high values for precipitation observed in July–August 2005 stand out, with about 500 events of rainfall exceeding 15 mm in 12 hours. In addition to that, other shifts between dry and wet periods are also evident. It is immediately worth pointing out that, when considering the two highest thresholds (thick–dotted and thick–dashed lines, relative to the 25 and 50 mm/12h values, respectively), a low number of occurrences, often below 10, is found for several seasons (especially in winters). This may cast some doubts on the statistical solidity of the results for these types of events over a long verification period. Good (poor) performance over specific seasons could be due to the fact that COSMO–LEPS predictions captured (did not capture) those few heavy precipitation events, this not allowing any solid conclusions on the effective performance of the system for the highest thresholds. On the other hand, large numbers of occurrences are often found for the other thresholds, including the event “15 mm of precipitation over 12 hours” (thick–solid line in Fig. 3), which is already a quite substantial amount. Therefore, the results relative to the two highest thresholds (25 and 50 mm/12h) will not be shown and more emphasis will be given to the remaining thresholds (1, 5, 10, 15 mm/12h).
As already mentioned, verification was performed over a 7-year period, from December 2002 to November 2009. For each month as well as for each season, the following probabilistic

Table 1: Main features of the verification configuration.

variable:	12–hour accumulated precipitation (18–06, 06–18 UTC);
period:	from Dec 2002 to Nov 2009;
region:	43–50N, 2–18E (D–PHASE area);
method:	nearest grid–point;
observations:	SYNOP reports;
forecast ranges (h):	6-18, 18-30, 30-42, 42-54, 54-66, 66-78, 78-90, 90-102, 102-114, 114-126;
thresholds:	1, 5, 10, 15, 25, 50 mm/12h;
scores:	ROC area, BSS, RPSS, OUTL;

scores are computed: the Brier Skill Score (BSS), the Ranked Probability Skill Score (RPSS), the Relative Operating Characteristic Curve (ROC) area and the Percentage of Outliers (OUTL). For a description of these scores, the reader is referred to Wilks (1995) and to Marsigli et al. (2008). The main features of the verification exercise are summarised in Table 1.

5 Time–series results

In the following subsections, the results relative to the performance of COSMO–LEPS for particular types of weather events are presented in terms of time–series scores for the above–mentioned probabilistic indices.

5.1 ROC Area

The ROC area (Mason and Graham, 1999) ranges from 0 to 1, the higher the better, and the value of 0.5 is the limit from skill and no–skill. For a forecast system to be useful, the ROC area should exceed the value of 0.7 (Buizza et al., 1999).

Fig. 4 shows the performance of COSMO-LEPS in terms of time-series values of the ROC area for the 30–42h forecast range (that is, the precipitation accumulated over the 12–hour period ending at 42–hour forecast step). The score exhibits a marked month–to–month variability and a 3–month running mean was applied to improve the readability of the plots. The skill of COSMO–LEPS is shown for the first 4 thresholds of Table 1: it can be noticed that, in any case, the performance of the system has increased throughout the years of COSMO–LEPS activity. This is especially true for the 10mm/12h and 15mm/12h thresholds (dashed and thick–solid lines, respectively), since the scores increased from about 0.6, in the first months of 2003, to more than 0.8, since mid–2007. The letters [A], [B] and [C] in the lower part of the plot denote the major system upgrades among those described in the previous section. It can be noticed that the [A] upgrade, relative to the increase of COSMO–LEPS ensemble size had a positive impact on the performance of the system. As for the [B] upgrade, it has to be noticed that, in February 2006, not only did COSMO–LEPS increase both ensemble size and vertical resolution, but also ECMWF EPS decreased its horizontal grid size from 80 to about 50 km in the horizontal (Buizza et al., 2007), thus improving the quality of both initial and boundary conditions provided to the limited–area runs. Despite these system upgrades, the performance of COSMO–LEPS was not as good as could be expected in the following months (Spring to Summer 2006), with scores below 0.8 for the 10 and 15 mm thresholds. On the other hand, some recovery in the performance of the system

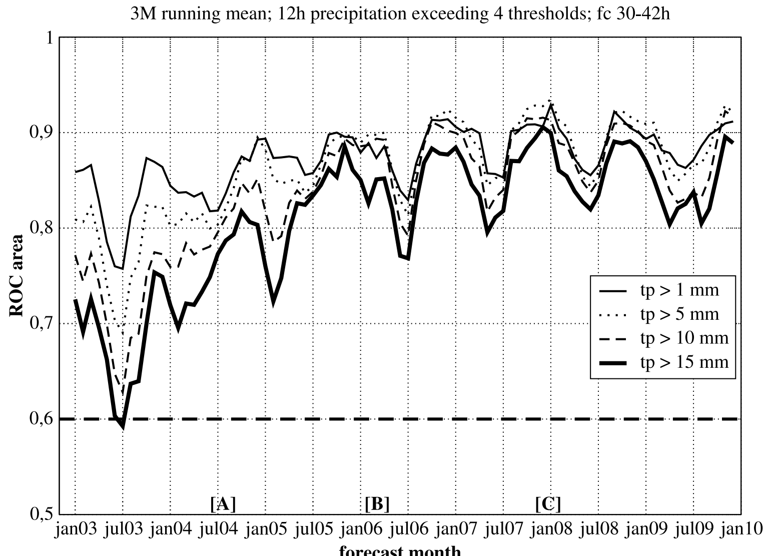

Figure 4: Time–series of ROC–area values for the monthly scores of COSMO–LEPS for 4 different thresholds: 1mm/12h (thin–solid line), 5mm/12h (dotted), 10mm/12h (dashed) and 15mm/12h (thick–solid). The forecast range is 30–42h. A 3–month running mean is applied to improve the readability of the plots. For the meaning of letters [A], [B], [C], refer to the text.

can be noticed in the following seasons. Hence, the drop in skill seems circumscribed to just a few months and cannot be ascribed to wrong implementations and/or faulty system upgrades. As for the [C] upgrade, the impact seems initially neutral with ROC area scores peaking up again in late 2009; a more detailed analysis (in next sections) will show some positive impact on the skill of the system also for 2008.
If the attention is focused on a longer forecast range, most of the above–mentioned results are confirmed.

Fig. 5 shows the ROC area values for the 78–90h range: it is clear that the absolute values of the scores are lower than before, since the prediction range has increased. In the first two years of activity of COSMO-LEPS (up to about February 2005), the scores relative to the 10 and 15 mm/12h thresholds fell below 0.7, which is generally considered the lower boundary for a prediction system to be useful. In the following months (and years), the improvement of performance is well detectable for all thresholds. The limited skill of the system after the [B] upgrade is confirmed for all thresholds, as well as the recovery of COSMO–LEPS in the subsequent months. Particularly high is the skill of the system during autumn and winter 2007, with ROC area value close, or slightly above, 0.9. It can be noticed that, for these seasons, the skill of the system is very similar for both the 30–42h and the 78–90h forecast range, this indicating a slow degrade of the prediction skill with the forecast range.
In addition to these comments, some dependence of the scores on the season can also be noticed. Both Figs. 4 and 5 indicate that ROC area values tend to be higher (lower) in autumns (summers). This may be related to the different types of atmospheric forcing for the precipitation events. In summer, precipitation is more related to convective–type of events, while in autumn large–scale forcing usually prevails. The former type of forcing tends to be less predictable by global–model ensemble systems (Buizza et al., 1999), which may provide less accurate initial and boundary conditions to COSMO–LEPS runs. In addition to this, convection is explicitly resolved by neither global nor limited–area integrations and this adds limitations to the system capability of simulating properly convective–type events. On the other hand, large–scale forcing is usually well captured by global–model systems,

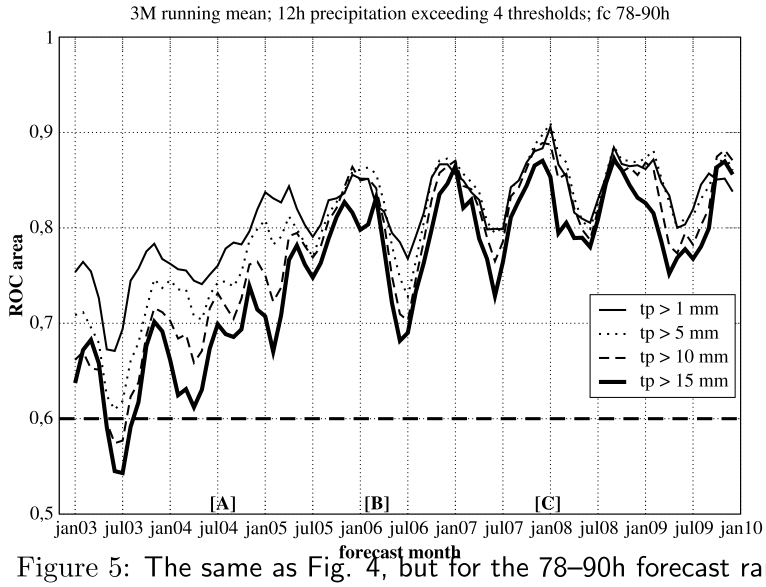

Figure 5: The same as Fig. 4, but for the 78–90h forecast range.

which provide higher quality boundaries for the dynamical downscaling by limited–area ensemble systems. On their turn, the added value of higher resolution and the more accurate description of mesoscale features, like the interaction of flow with orography, contributes to the better performance of COSMO–LEPS during the autumn season.

5.2 Brier Skill Score

Now, the attention is focused on the performance of COSMO–LEPS in terms of BSS (Wilks, 1995). This score ranges from minus infinity to 1 and, for a forecast system to be more useful than climatology (in this case, the reference climatology is given by the sample climate), has to be positive. Fig. 6 shows the time–series evolution of COSMO–LEPS performance for the 30–42h forecast range: despite a 3–month running mean is applied to the score values so as to increase the readability of the plot, the month–to–month variability is larger than in the case of the ROC area and the increase in forecast skill throughout the years of COSMO–LEPS activity is less evident. It is difficult to detect a clear benefit by any of the above–mentioned upgrades and, depending on the threshold, periods of relative high (low) skill alternate. Nevertheless, it can be noticed that the BSS has been positive for all thresholds since April 2009, possibly a consequence of the most recent forecast improvements.

In order to detect trends in the score, it was decided to apply a 12–month running mean[2] to the monthly BSS scores, as in Fig. 7. In this figure, the shapes of the profiles are much smoother than before, but longer–term trends can be identified. Depending on the threshold, the following considerations can be presented:

- the quality of the forecast progressed quite slowly for the 1mm/12h and 5mm/12h events (thin–solid and dotted lines, respectively); BSS is nearly always positive indicating a performance of the system better than climatology; the system upgrades do not seem to have had a large impact on the performance of COSMO–LEPS for the

[2]The BSS value for a particular month is given by the average of the score for the preceding 6 months and for the following 6 month values. Hence, BSS values cannot be computed for the first 6 months and for the last 6 month of the time–series.

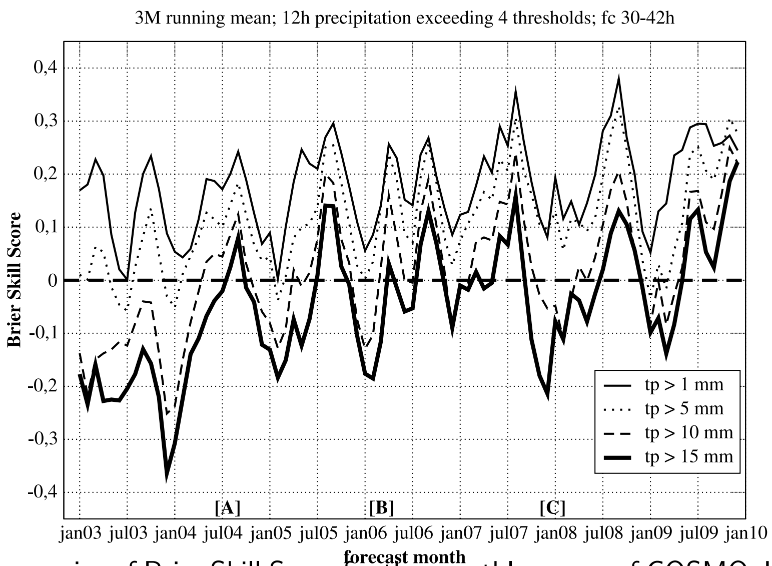

Figure 6: Time–series of Brier Skill Score for the monthly scores of COSMO–LEPS for 4 different thresholds: 1mm/12h (thin–solid line), 5mm/12h (dotted), 10mm/12h (dashed) and 15mm/12h (thick–solid). The forecast range is 30–42h. A 3–month running mean is applied to improve the readability of the plots. For the meaning of letters [A], [B], [C], refer to the text.

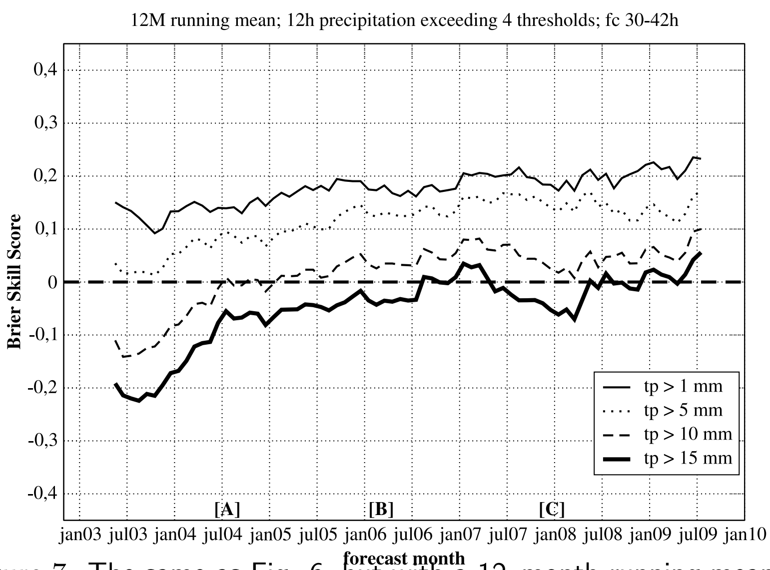

Figure 7: The same as Fig. 6, but with a 12–month running mean applied.

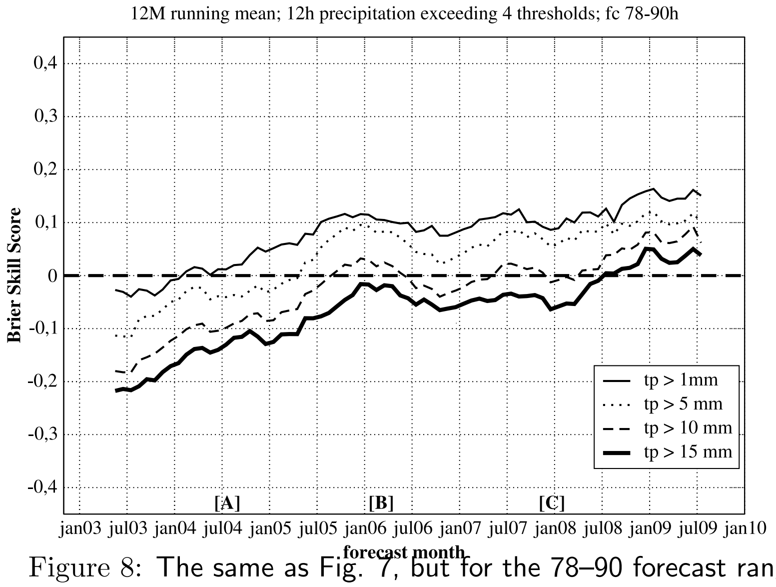

Figure 8: The same as Fig. 7, but for the 78–90 forecast range.

prediction of this type of events;

- the improvement is more evident for the two highest thresholds (dashed and thick–solid lines, respectively) throughout the years; BSS is systematically negative in the first years of activity, while it has been above zero for the 10mm/12h threshold since July 2005; as for the 15 mm/12h threshold, a good trend is evident up to January 2006, then the quality for the prediction of this type of events decreases up to mid–2008; since then, the skill of the system has started to increase again, although very close to the zero line. This is a very encouraging result, since it shows the improvement of the system performance for events of moderate–to–heavy precipitation in the early–medium range, the "hunting ground" of COSMO–LEPS;
- the dependence of the forecast skill on the threshold value has been reduced during the years of COSMO–LEPS activity. In 2003, the values of the BSS ranged roughly from -0.2 to 0.15 when passing from the highest to the lowest threshold; during 2009, BSS variations are approximately limited to the interval $[0.05, 0.25]$.

If the same type of running mean is applied to the BSS relative to the 78–90h forecast range, a common behaviour is evident for all thresholds. Fig. 8 shows an increase in the skill of the system for all types of events from the beginning of the activity up to January 2006. Then, the performance of the system stays stable for about 2 years and, finally, picks up again from July 2008 onwards. It can be seen that, for the 1mm (5mm) threshold, denoted with the solid (dotted) line of Fig. 8, the BSS has been positive since July 2004 (July 2005). As for the 10mm/12h and 15mm/12h thresholds (dashed and thick–solid lines, respectively), a positive trend on the forecast skill after the [C] upgrade can be noticed starting from July 2008. It is worth pointing out that we are assessing the possibility to predict 12–hourly accumulated precipitation after almost 4 days of integration (78–90 prediction range), which is a very "severe" test for any ensemble system. These results indicate the progresses of COSMO–LEPS forecasts in the early detection of severe and localised weather events. A quick comparison relative to the performance of the system for the highest thresholds (thick–solid lines of Figs. 7 and 8) indicates almost identical scores for both the shorter and the longer forecast range. This seems in apparent contradiction with the usual degrade of the

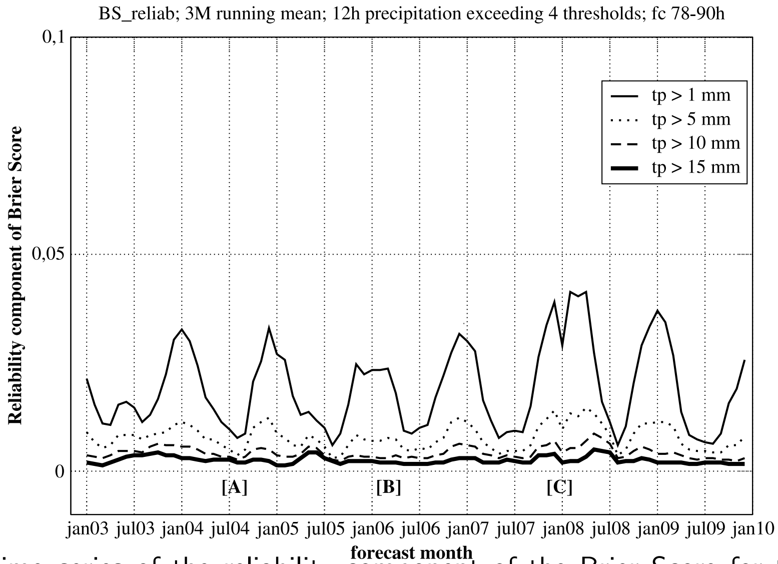

Figure 9: Time–series of the reliability component of the Brier Score for the monthly scores of COSMO–LEPS for 4 different thresholds: 1mm/12h (thin–solid line), 5mm/12h (dotted), 10mm/12h (dashed) and 15mm/12h (thick–solid). The forecast range is 78–90h. A 3–month running mean is applied to improve the readability of the plots. For the meaning of letters [A], [B], [C], refer to the text.

forecast skill with increasing forecast range, but is actually a feature of COSMO–LEPS system, which is mainly targeted for the early–medium range and is driven by those selected EPS members which should provide better guidance after the short range.

The reasons for the different results between the BSS and the ROC area can be found in the different type of information conveyed by the two indices (Marsigli et al., 2008). On the one hand, the BSS (and the RPSS) gives information about both the reliability and the resolution of the forecast system. The former one indicates how well the forecast probabilities by the ensemble system match the observed frequencies; the latter one highlights the extent to which the system can discriminate among events in different categories. A more thorough investigation on these two components, in their formulation used to compute the Brier Score (BS, not the BSS), shows the following:

- reliability component: the time–series of this quantity, which should be low (to contribute to a low BS and, hence, to a high BSS) is shown in Fig. 9 for the forecast range 78–90h, relative to the four thresholds in the previous figures. The time–series scores indicate little progress (if any) as for the reduction of the reliability component throughout the years of COSMO–LEPS activity. As for the 1mm/12h and 5mm/12h thresholds (thin–solid and dotted lines, respectively), a yearly cycle with almost constant amplitude is evident with higher values in winter; on the other hand, stable values can be noticed for the other thresholds. The reliability component could be modified by a change in the methodology used to construct the driving ensemble (ECWMF EPS) and/or COSMO–LEPS, which did not occur in these years;
- resolution component: this quantity should be high so as to provide a lower BS and a higher BSS. The time–series of this term (shown in Fig. 10 for the range 78–90h) indicates an increase of the values especially for the two lowest thresholds, indicating a greater capability of the system to distinguish among events in different categories.

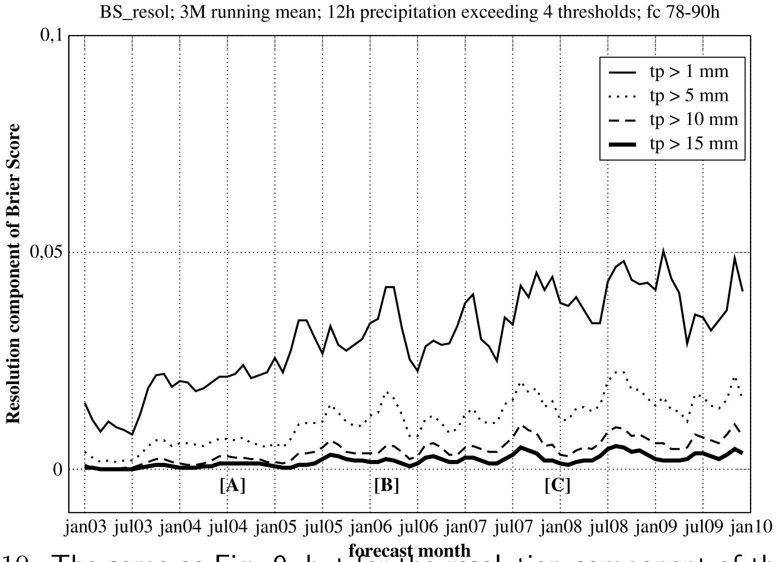

Figure 10: The same as Fig. 9, but for the resolution component of the Brier Score.

Therefore, the less marked improvement in terms of BSS is related to the behaviour of these two terms and on their relative weight in the generation of the score.
On the other hand, the ROC area provides information about the resolution component, which was shown to increase. For this reason, the improvement in terms of the ROC area is more evident, as was shown by Figs. 4 and 5.

5.3 Ranked Probability Skill Score

The RPSS (Epstein, 1969) is an extension of the BSS to a situation with many events, in this case those represented by the list of thresholds of Table 1. As before, the score ranges from minus infinity to 1 and a forecast system providing more useful information than the climatology (as before, given by the sample climate) would have a positive RPSS. If the RPSS is calculated for each month and for the different forecast ranges of Table 1, it turns out that the month–to–month variability of the score is extremely high (not shown). Therefore, in order to detect long–term behaviour for the value of this score, a 12–month running mean was again applied, like for the BSS. Fig. 11 shows the performance of the RPSS for different ranges, representative of COSMO–LEPS quality throughout the full integration period: 30–42h (solid), 54–66h (dotted), 78–90h (dashed) and 102–114 (thick–solid). The increase of the quality of COSMO–LEPS forecasts in the years is evident, although the pace of improvement is different depending on the forecast range. Periods of almost "monotonic growth" (e.g. the Jul05–Jan06 period for the 78–90h range) alternate with others of steady or slightly decaying performance. As an overall comment, the RPSS has been always positive for all forecast intervals, since the [B] upgrade, with particularly high scores in the last 12 months. The improvements concern all forecast ranges, the RPSS growing especially for the longest ones. In addition to that, those RPSS values reached around January 2004 in the early range (30–42h forecast, solid line in Fig. 11) are recently obtained for the 102–114 prediction range, with a gain of predictability of about 3 days in 5 years of system upgrades.

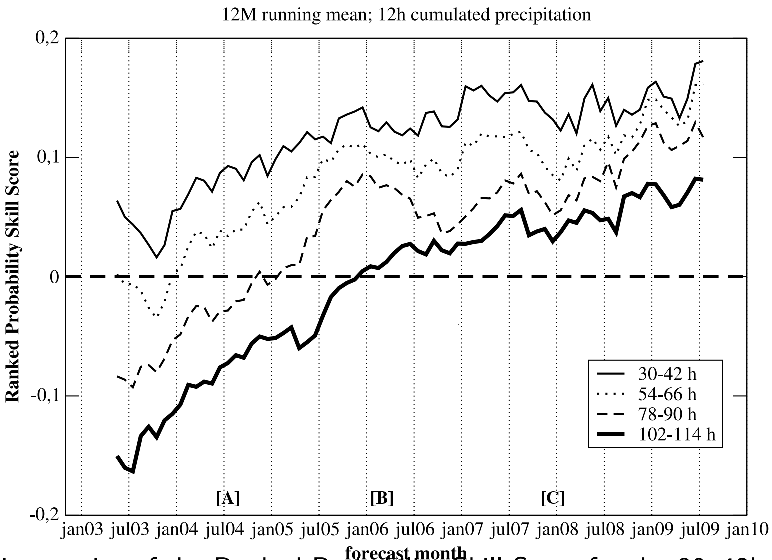

Figure 11: Time–series of the Ranked Probability Skill Score for the 30–42h (solid line), 54-66h (dotted line), 78–90h (dashed–line) and 102–114h (thick–solid line) forecast ranges. A 12–month running mean is applied to improve the readability of the plots. For the meaning of letters [A], [B], [C], refer to the text.

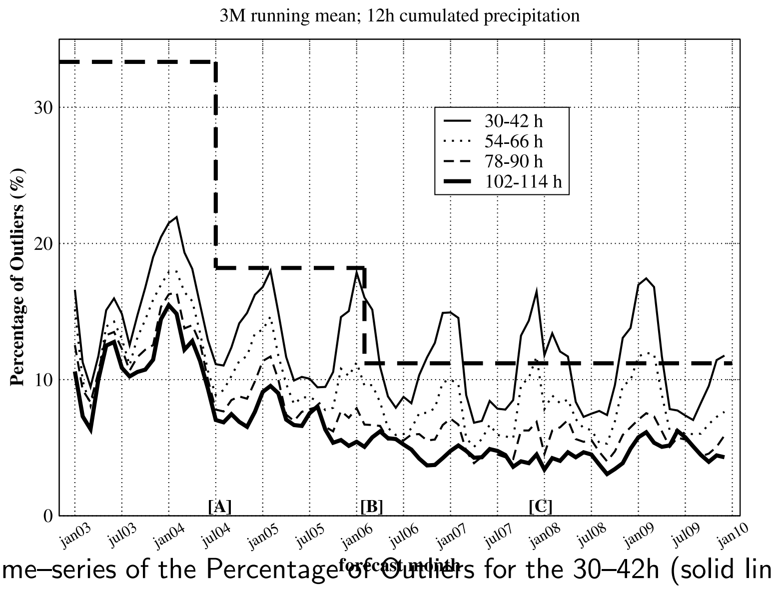

Figure 12: Time–series of the Percentage of Outliers for the 30–42h (solid line), 54-66h (dotted line), 78–90h (dashed–line) and 102–114h (thick–solid line) forecast ranges. A 3–month running mean is applied to improve the readability of the plots. The thick–dashed line indicates the theoretical values of the OUTL. For the meaning of letters [A], [B], [C], refer to the text.

5.4 Percentage of Outliers

The OUTL measures the percentage of times the observation stays outside the interval spanned by the values predicted by the ensemble members. For an ensemble of size N, its theoretical value is given by $2/(N+1)$ (Talagrand et al., 1999) and for a real ensemble system, it should be below this value. As for COSMO–LEPS, this value has changed during the years of the system activity: more precisely, it was 33.3% from November 2002 to July 2004 (up to the [A] upgrade); then, 18.2% for the 10–member ensemble (up to the [B] upgrade); finally, 11.2% for the 16–member size.
Fig. 12 shows the time–series evolution for OUTL at four different forecast ranges: 30–42h (solid), 54–66h (dotted), 78–90h (dashed) and 102–114 (thick–solid). For reference, the thick–dashed line indicates the theoretical value not to be exceeded. As for the previous scores, a 3–month running mean was applied to highlight the main results of the verification, which can be summarised as follows:

- a seasonal cycle of the score can be identified for all forecast ranges; OUTL tends to be higher in winters, especially for shorter forecast ranges, and lower in summers. This is due to an overestimation of little precipitation amounts during the cold season by the model runs; in fact, a partitioning of the OUTL into outl_max (observed value larger than all forecast ones) and outl_min (observed value smaller than all forecast ones) indicates that the latter contribution is well dominant for shorter ranges;
- for a fixed forecast range, it is shown an overall reduction of OUTL throughout the years of COSMO–LEPS activity from 2003 up to mid–2007, when, for any forecast range, OUTL was less than 10%, well below the theoretical value for a 16–member ensemble.
- from mid–2007 to early 2009, the OUTL has remained stable or has slightly grown for any range; in this period, the OUTL for shorter ranges exceeded the theoretical value;
- during 2009, a new reduction of OUTL takes place, with values of the same order (or below) those attained in 2007. For any forecast range but the first one, OUTL stays below the theoretical value. The excessive amount of OUTL for the shortest range is related to a lack of spread among COSMO–LEPS members in the early forecast intervals; this is possibly a consequence of the driving ensemble, which is specifically targeted to have sufficient spread in the medium range.

When the population of a small–size ensemble is increased, the reduction of OUTL is almost a natural consequence, since the probabilistic system can now access a wider fraction of the atmospheric phase space and account for a larger percentage of possible evolution weather scenario. Therefore, the OUTL reduction relative to upgrades [A] and [B] were not unexpected. On the other hand, the 2009–reduction is more likely due to the better performance of the individual COSMO–LEPS runs, related to model upgrades and refinements.

6 Seasonal scores

The results obtained in the previous section indicate that, on the basis of a number of probabilistic indices, the forecast skill of COSMO–LEPS has improved in many respects, in the past as well as recently. In this section, the performance of the system is studied in greater detail, investigating the COSMO–LEPS behaviour over specific periods and for particular types of weather events. More precisely, the attention is focused on the skill of

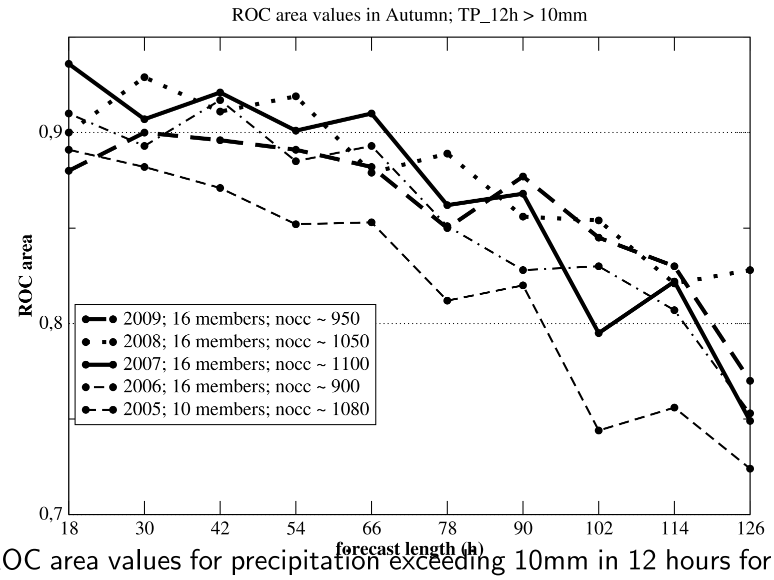

Figure 13: ROC area values for precipitation exceeding 10mm in 12 hours for the forecast ranges of Table 1 and for 5 successive autumns: 2005 (dashed), 2006 (dot–dashed), 2007 (thick–solid), 2008 (thick–dotted) and 2009 (thick–dashed).

the system during the two most rainy seasons in the Alpine area: autumn, defined as the 3–month period covering September, October and November (SON) and summer (June, July and August; JJA).
It has to be pointed out that, when comparing different seasons, the variations in the scores from one season to the other may depend on the fact that more occurrences took place during one season rather than during another one. In other words, if one autumn is more rainy than the previous one and also with a prevailing forcing better captured by the limited–area–model runs, then the ROC area values would be higher than for the other one. This may be probably due the different statistics of the two seasons under consideration and may not necessarily reflect an absolute improvement in forecast skill because of, e.g., model refinement or new ensemble strategies. Nevertheless, this type of verification, when applied to more and more seasons, enables to identify the improvements of the forecast system as well as of particular periods of good (or bad) performance.

6.1 Scores in autumn

For the autumn season, it is investigated the capability of the system to predict the event "precipitation exceeding 10mm over 12 hours". Fig. 13 shows the evolution of COSMO–LEPS performance in terms of ROC area for the forecast of this event. In order to limit the number of lines per plot, the scores of the system are reported only for the last 5 autumns (from 2005 to 2009). It can be noticed that the ROC area values relative to autumn 2005 (dashed line) are lower than those obtained in the following years for almost all forecast ranges. This may be related to the fact that in 2005 COSMO–LEPS still had a population of 10 members, the ensemble size being 16 in the other years. As for years 2006–2009, the performance of the system is quite stable: for shorter forecast ranges, COSMO-LEPS has slightly better performance in autumn 2007 (thick–solid line of Fig. 13), while higher ROC area values for longer ranges can be noticed in 2009 (thick–dashed line). For reference, the legend of Fig. 13 also reports, for each autumn, the number of occurrences, that is the number of times precipitation actually surpassed the 10mm threshold in 12 hours, according

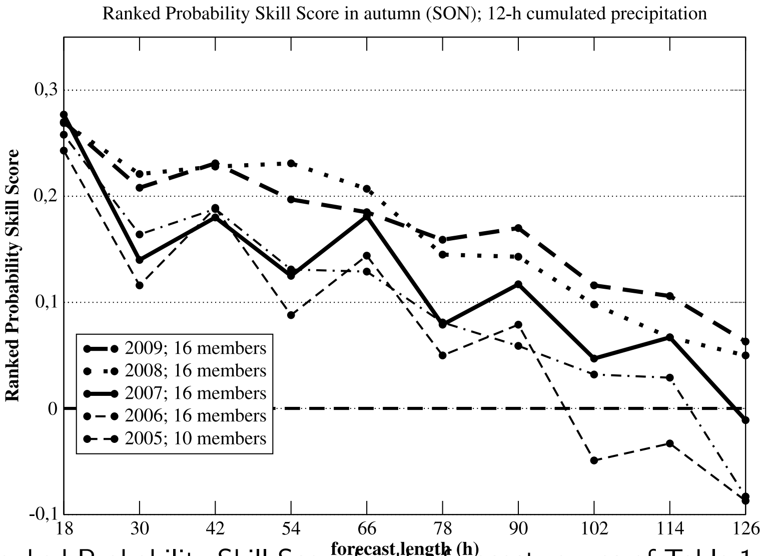

Figure 14: Ranked Probability Skill Score for the forecast ranges of Table 1 and for 5 successive autumns: 2005 (dashed), 2006 (dot–dashed), 2007 (thick–solid), 2008 (thick–dotted) and 2009 (thick–dashed).

to all reports of the SYNOP stations of Fig. 2.
As a further example of the behaviour of the probabilistic indices for the same season but in different years, the attention is focused on a score describing the COSMO–LEPS skill for multi–event situations. Fig. 14 shows the behaviour of the RPSS for the last 5 autumns as a function of the forecast step. The highest values are obtained in the last 2 years, the RPSS being positive for all forecast ranges (thick–dotted and thick–dashed lines for 2008 and 2009, respectively). In addition to that, Fig. 14 also indicates that, in the first years of activity, RPSS values had a marked semi–diurnal cycle. The system provides better guidance for "night–time" precipitation, that is for rainfall observed between 18UTC and 6UTC (and corresponding to the ranges 6–18h, 30–42h, 54–66h, 78–90h, 102–114h). As for "day–time" precipitation (observed from 6UTC to 18UTC), the model runs had a worse performance due to a systematic overestimation of rainfall (linked with a too rapid onset of convection). This is a general feature of the model, as pointed out by Oberto and Turco (2008) for runs of COSMO in "deterministic mode". This behaviour is evident throughout the full forecast range in 2005 and 2007 (dashed and thick–solid line, respectively). As for the last 2 years, the cycle in the score is still present, but its amplitude is reduced. It has to be pointed out that this type of model (mis)–behaviour shows up because of the decision to verify 12–hour accumulated precipitation. If the performance of COSMO–LEPS were evaluated over a 24–hour period, the problems relative to the semi–diurnal cycle would be masked and it would be more difficult to highlight the improvements in forecast performance under that aspect. Nevertheless, it can be noticed that, while the RPSS dropped definitely below 0.1 after about 72 hours of forecast in 2005, it does the same in 2009 but after 114 hours, with a gain of predictability of almost 2 days in 4 years.

Finally, it is presented the COSMO–LEPS performance at different forecast ranges in terms of Percentage of Outliers. The attention is focused, as before, on the last 5 autumns and Fig. 15 indicates that the best performance of the system is achieved in autumn 2009 (thick–dashed line), with OUTL below 10% after about 42 forecast hours. It has to be noticed that, during autumn 2007, the performance of the system was less satisfactory than in the other seasons up to day 4 (thick–solid line in the figure). This was due to a large

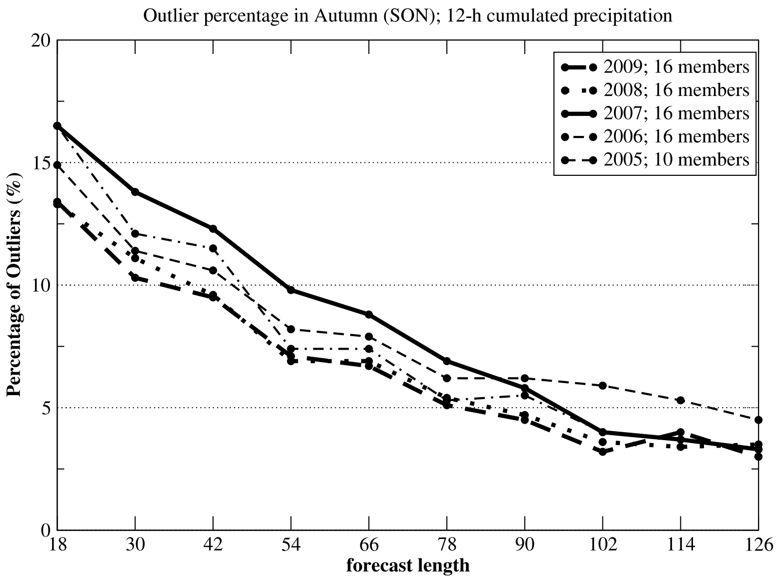

Figure 15: Percentage of Outliers for the forecast ranges of Table 1 and for 5 successive autumns: 2005 (dashed), 2006 (dot–dashed), 2007 (thick–solid), 2008 (thick–dotted) and 2009 (thick–dashed).

amount of outl_min, indicative of an overestimation of precipitation by COSMO–LEPS runs for several forecast ranges. As for 2005, the only season with a 10–member ensemble, a larger fraction of outliers is found for longer prediction ranges (dashed line), but the gap is not large. It is encouraging to notice that, as for the last autumns of activity, only in 5% of the cases is the observed precipitation outside the range spanned by the ensemble members from forecast–day 4 onwards.

6.2 Scores in summer

As for summer, the attention is limited to the COSMO–LEPS forecast skill in terms of RPSS for several prediction ranges. The 12–hourly cycle in the system performance, already shown for autumn by Fig. 14, is more evident during the warm season. Fig. 16 indicates that, for all summers under investigation (from 2005 to 2009), COSMO–LEPS has higher scores for "night–time" verification, while the skill is lower for precipitation observed between 6UTC to 18UTC and corresponding to the forecast ranges 18–30h, 42–54h, 66–78h, 90–102h and 114–126h [3]. As pointed out in the section 6.1, the sensitivity of the system performance to the time of verification is mainly due to an anticipation in the onset of convection by the model. This leads to precipitation occurring too early (morning or first hours of the afternoon) in the model runs. Obviously, this problem is amplified in summers, when the large–scale forcing is weaker than in autumns and the precipitation is mainly driven by convection processes. Due to the stronger role played by convection, the 12–hourly cycle in RPSS behaviour persists up to day 5 for all years but the last one. In fact, in 2009 (thick–dashed line), are not only the score values higher than in the past, but also the amplitude of the cycle is smaller for short forecast ranges and almost missing from day 3 onwards. This result seems to suggest that the model upgrades did bring benefit to the accuracy of COSMO–LEPS runs, which provide

[3]For a part of summer 2005, the forecast length was 120 hours, which means 12 hours shorter than in the other summers. Therefore, the score relative to this year (represented by the dashed line in Fig. 16) does not cover the last forecast range, but stops earlier.

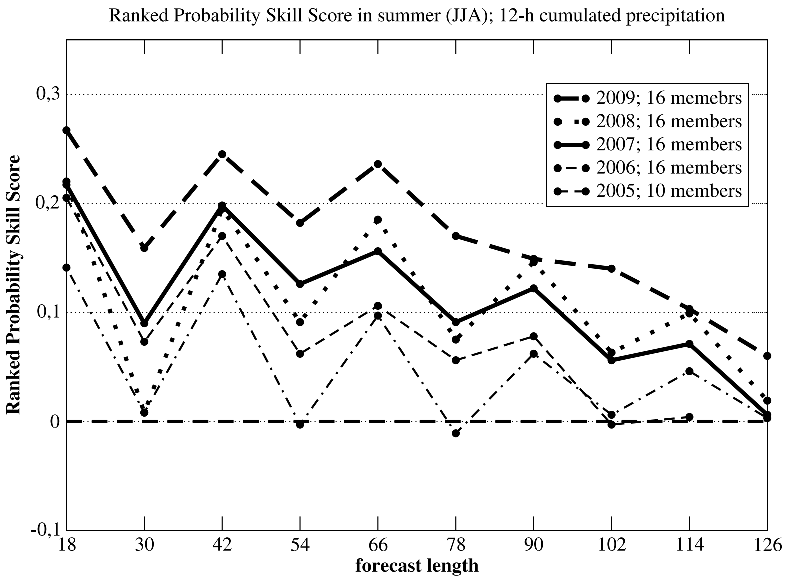

Figure 16: Ranked Probability Skill Score for the forecast ranges of Table 1 and for 5 successive summers: 2005 (dashed), 2006 (dot–dashed), 2007 (thick–solid), 2008 (thick–dotted) and 2009 (thick–dashed).

more and more valuable forecasts in the short and early–medium range. In fact, a quick comparison between the performances of 2005 and 2009 (dashed versus thick–dashed line in Fig. 16) shows that, with reference to the 0.1 value, the skill score obtained after 66 hours is now achieved after 114 hours, thus confirming the gain of about 2 days of predictability in 4 years of operational activity.

7 Conclusions

The main features of the mesoscale ensemble system COSMO–LEPS were presented. At the moment, the system is made up of 16 integrations of the non–hydrostatic COSMO model with a horizontal resolution of 10 km, 40 vertical levels and a forecast length of 132 hours, thus providing short and early–medium range ensemble forecast of high–impact weather with great spatial detail. COSMO–LEPS runs can be viewed as a dynamical downscaling of ECMWF EPS, where the number of global–ensemble members are reduced via an "ensemble–size reduction" technique.
The performance of COSMO–LEPS has been analysed in terms of probabilistic prediction of 12–hourly accumulated precipitation for a number of thresholds. The evolution of the skill of the system has been assessed over a 7–year period, namely from December 2002 up to November 2009. Observations of 12–hour accumulated precipitation were taken from the SYNOP reports over the Alpine area and compared to the nearest grid–point forecasts of the COSMO–LEPS members. A number of probabilistic indices has been used so as to evaluate both reliability and resolution of the system and to assess the extent to which the quality of COSMO–LEPS forecasts improved in recent times.
The main findings can be summarised as follows:

- time–series scores of the ROC area values for different thresholds indicate a clear improvement of COSMO–LEPS skill throughout the years, with an increase of this score up to 2006, then slower growth in the 2 following years and a resumption of growth in the last part of 2009. This holds in the short as well as in the early–medium range;

- as for the BSS (and the RPSS), the month–to–month variability of the time–series scores is higher than for the ROC area and the positive impacts of system upgrades are more difficult to detect. If 12–month filters are applied to the scores, more insight is gained and forecast improvements with time become more detectable. As for an estimate of the improvement, RPSS values enables to quantify it as "2 days of predictability gained in the last 5 years";
- time–series of the OUTL indicate a decrease of Outliers up to early 2007 for various forecast ranges and a steady behaviour since then on. A new decrease is evident in 2009, with the OUTL below 10% from day 2;
- seasonal scores confirm to a large extent the previous results; they enable to highlight the progresses in the skill of the system over the same season but in different years. The improvements in skill are well noticeable in terms of RPSS during summer, where the gain in forecast skill can be estimated as about 2 days of predictability in the last 4 years of activity.

The reasons for the different results between the BSS and the ROC area were shown to be due to the different type of information conveyed by the two indices. On the one hand, the BSS (and the RPSS) gives information about both the reliability and the resolution of the forecast system (Marsigli et al., 2008). The former one indicates how well the forecast probabilities by the ensemble system match the observed frequencies. The latter one indicates the extent to which the system can discriminate among events in different categories. Reliability can be increased after a good calibration, which manages to match probabilities and frequencies; resolution cannot be improved by any statistical post-processing. On the other hand, the ROC area provides information only about the discrimination capability of the forecast system. As such, the ROC area scores show the COSMO–LEPS ability to detect between events and nonevents, despite the reliability, and represent the hypothetical skill of the system once it were properly calibrated. It was shown that the reliability component of COSMO–LEPS forecasts did not increase in the last years at the same pace as the resolution one, this explaining the better results obtained in terms of ROC area.

Work is in progress at both ARPA–SIMC and MeteoSwiss to provide calibrated COSMO–LEPS forecasts, thanks to the results obtained by the re–forecast exercise described in Fundel et al. (2009). This would improve the skill of COSMO–LEPS forecasts, making the system more reliable than it is now. As for the future, it is being tested the increase of the horizontal resolution of COSMO–LEPS runs so as to provide more detailed forecasts for the interaction of the flow with orography and to describe with a higher degree of accuracy mesoscale–related processes and local effects. This would have a positive impact on the prediction of a number of those surface fields still nowadays strongly influenced by local effects and not always properly represented in terms of their uncertainty by mesoscale ensemble systems.

8 Acknowledgments

The authors are indebted to the countries of the COSMO Consortium and to ECMWF for the possibility to run the COSMO–LEPS application. We are grateful to Manuel Fuentes, Umberto Modigliani and Paolo Patruno for technical assistance and continuous support. We also thank André Walser, Carlo Cacciamani and Stefano Tibaldi for useful discussions.

9 References

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. and Beare, S.E., 2008. The MOGREPS short–range ensemble prediction system. *Q. J. Roy. Meteorol. Soc.* **134**, 703–722.

Buizza, R., Hollingsworth, A., Lalaurette, F.and Ghelli, A., 1999. Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting* **14**, 168–189.

Buizza, R., 2005. The ECMWF Ensemble Prediction System. In: *Predictability of weather and climate,* (eds. T. Palmer and R. Hagedorn), Cambridge University Press, Cambridge, 459–488.

Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G. and Vitard, F., 2007. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Q. J. Roy. Meteorol. Soc.* **133**, 681–695.

Epstein, E.S., 1969. A scoring system for probabilities of ranked categories. *J. Appl. Meteor.* **8**, 985–987.

Fundel, F., Walser, A., Liniger, M.A., Frei, C. and Appenzeller, C., 2009. Calibrated Precipitation Forecasts for a Limited-Area Ensemble Forecast System Using Reforecasts. *Mon. Wea. Rev.* **138**, 176–189.
DOI: 10.1175/2009MWR2977.1

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

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

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T., 2005a. The COSMO–LEPS ensemble system: validation of the methodology and verification. *Nonlin. Proc. in Geophys.* **12**, 527–536.

Marsigli, C., Montani, A., Paccagnella, T., Sacchetti, D., Walser, A. and Schumann, T., 2005b. Evaluation of the performance of the COSMO–LEPS system. *COSMO Technical Report* no **8**, 40. Available at http://www.cosmo-model.org/.

Marsigli, C., Montani, A. and Paccagnella, T., 2008. A spatial verification method applied to the evaluation of high–resolution ensemble forecasts. *Met. Appl.* **15**, 125–143.

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

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. Roy. Meteorol. Soc.* **122**, 73–119.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T., 2001. A strategy for High–Resolution Ensemble Prediction. Part I: Limited–area experiments in four Alpine flood events. *Q. J. Roy. Meteorol. Soc.* **127**, 2069–2094.

Montani, A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U., Nerozzi, F., Paccagnella, T., Patruno, P. and Tibaldi, S., 2003a. Operational limited–area ensemble forecasts based on the Lokal Modell. *ECMWF Newsletter* no **98**, 2–7. Available at: http://www.ecmwf.int/publications/.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi S. and Buizza, R., 2003b. The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts. *Nonlin. Proc. Geophys.* **10**, 261–274.

Montani, A., Marsigli, C. and Paccagnella, T., 2008a. Five Years of Limited–Area Ensemble Activities at ARPA–SIM: The COSMO–LEPS system. *COSMO newsletter* no **8**, 23–26. Available at: http://www.cosmo-model.org/.

Montani A., Marsigli, C. and Paccagnella, T., 2008b. Performance of COSMO–LEPS system during the D–PHASE Operations Period. *COSMO newsletter* no **9**, 50–53. Available at: http://www.cosmo-model.org/.

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

Oberto, E. and Turco, M., 2008. Report about the Latest Results of Precipitation Verification over Italy. *COSMO newsletter* no **8**, 37–44. Available at: http://www.cosmo-model.org/.

Rotach, M.W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H.-S., Behrendt, A., Bouttier, F., Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M.A., Marsigli, C., McTaggart-Cowan, R., Montani, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos-Muoz, D., Schr, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V. and Zappa, M., 2009. MAP D-PHASE: Real-time Demonstration of Weather Forecast Quality in the Alpine Region, *Bull. Amer. Meteor. Soc* **90** (9), 13211336, DOI:10.1175/2009BAMS2776.1

Steppeler, J., Doms, G., Schattler, U., Bitzer, H.W., Gassmann, A., Damrath, U. and Gregoric, G., 2003. Meso–gamma scale forecasts using the nonhydrostatic model LM. *Meteor. and Atmos. Phys.* **82**, 75–96.

Talagrand, O., Vautard, R. and Strauss, B., 1999. Evaluation of probabilistic prediction systems. *Proceedings of the ECMWF Workshop on Predictability*, Reading, 20–22 October 1997, 372. Available at:
http://www.ecmwf.int/publications/.

Tibaldi, S., Paccagnella, T., Marsigli, C., Montani, A. and Nerozzi, F., 2006. Limited–area ensemble forecasting: the COSMO–LEPS system. In: *Predictability of weather and climate,* (eds. T. Palmer and R. Hagedorn), Cambridge University Press, Cambridge, 489–513.

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

Walser, A., 2006. COSMO-LEPS Forecasts for the August 2005 Floods in Switzerland. *COSMO newsletter* no **6** , 142–145. Available at: http://www.cosmo-model.org/.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, New York, 467.

Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J. and Wunram C., 2008. MAP D–PHASE: real–time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Let.* **9**, 80–87. DOI 10.1002/asl.183.

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.

No. 6: July 2006.

No. 7: April 2008; Proceedings from the 8th COSMO General Meeting in Bucharest, 2006.

No. 8: September 2008; Proceedings from the 9th COSMO General Meeting in Athens, 2007.

No. 9: December 2008

No. 10: March 2010.

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 non-hydrostatic 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.

No. 9: Erdmann Heise, Bodo Ritter, Reinhold Schrodin (2006):
Operational Implementation of the Multilayer Soil Model.

No. 10: M.D. Tsyrulnikov (2007):
Is the particle filtering approach appropriate for meso-scale data assimilation ?

No. 11: Dmitrii V. Mironov (2008):
Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model.

No. 12: Adriano Raspanti (2009):
COSMO Priority Project "VERification System Unified Survey" (VERSUS): Final Report.

No. 13: Chiara Marsigli (2009):
COSMO Priority Project "Short Range Ensemble Prediction System" (SREPS): Final Report.

No. 14: Michael Baldauf (2009):
COSMO Priority Project "Further Developments of the Runge-Kutta Time Integration Scheme" (RK): Final Report.

No. 15: Silke Dierer (2009):
COSMO Priority Project "Tackle deficiencies in quantitative precipitation forecast" (QPF): Final Report.

No. 16: Pierre Eckert (2009):
COSMO Priority Project "INTERP": Final Report.

No. 17: D. Leuenberger, M. Stoll and A. Roches (2010):
Description of some convective indices implemented in the COSMO model.

No. 18: Daniel Leuenberger (2010):
Statistical analysis of high-resolution COSMO Ensemble forecasts in view of Data Assimilation.

No. 19: A. Montani, D. Cesari, C. Marsigli, T. Paccagnella (2010):
Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO–LEPS system: main achievements and open challenges.

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 (USAM, ARPA-SIMC, ARPA Piemonte), Switzerland (MeteoSwiss), Poland (IMGW) and Romania (NMA). 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 COSMO-Model (previously known as 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 completion 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:

Massimo Milelli
Massimo.Milelli@arpa.piemonte.it

Ulrich Schättler
Ulrich.Schaettler@dwd.de