Impact of high-resolution boundary conditions on the quality of COSMO-LEPS forecasts

Andrea Montani, C. Marsigli, T. Paccagnella

ARPA–SIMC, Bologna, Italy

1 Introduction

The use of Limited-Area-Model Ensemble Prediction Systems (LAMEPS), either in operational or research mode, is getting more and more widespread across Europe. Several activities are taking place in different weather services to assess the usefulness of these systems and their skill in the timely prediction of high-impact weather events at great spatial and time detail. In the framework of the development of European LAMEPS, the ECMWF Research Department performed a number of global-model ENS reruns with two different configurations (further details available under `https://software.ecmwf.int/wiki/display/LAMEPS/LAMEPS+Home`).

The former experiment was run with the operational horizontal resolution of T_L639 (32 km), while the latter one (referred to as exp H) was run at the higher resolution of T_L1279 (16 km). Both ensembles had the following configuration: 20 perturbed forecasts, one control forecast, a forecast range of 144 hours and 62 vertical levels. Model level data were archived 1–hourly until a lead time of 48 h and 3–hourly afterwards. All experiments were initialized twice daily, starting at 00 and 12 UTC.

Data were provided for three two-week periods, selected in such a way to encompass several high-impact weather events occurred over Europe, as summarised in Table 1. The availability of this unique dataset made possible the performance of several tests by the LAMEPS community, with either convection–permitting or convection–parameterised ensemble systems.

Table 1: Selected periods for ECMWF ENS reruns.

period	start	end	no. of days	type of events
1)	24 Oct 2011	7 Nov 2001	15	floods over Italy and France
2)	27 Dec 2011	8 Jan 2012	13	storms over Northern Europe
3)	11 Jun 2012	28 Jun 2012	18	Alpine summer convection
all cases			46	

As for the experimentation with COSMO–LEPS (Montani et al., 2011), the limited-area-model ensemble prediction system operationally run by ARPA-SIMC on behalf of the COSMO Consortium (`http://www.cosmo-model.org`), the attention was focussed on the fields by the high-resolution ENS experiments (exp H), which were used to provide both initial and boundary conditions for limited-area reruns.
The methodology followed to perform the experiments and to assess the skill of the different forecasting systems is described in the next two sections.

2 Methodology

Four different sets of ensembles were compared:

1. **opecleps** ($\Delta x = 7$ km, 40 Model Levels, 16 members), the operational COSMO-LEPS running at the time of the experiments (COSMO model version 4.12), with 16 members selected out of 102 (from 2 successive runs of operational ECMWF ENS) using the cluster-analysis technique described in Montani et al. (2011);
2. **TESTcleps_OldModel** ($\Delta x = 7$ km, 40 Model Levels, 16 members), the test version of COSMO-LEPS nested on high-res ENS (COSMO model version 4.12) and with 16 members selected out of 42 (from 2 successive runs of high-res ENS);

3. **TESTcleps_NewModel** ($\Delta x = 7$ km, 40 Model Levels, 16 members), the same as "TESTcleps_OldModel", but with COSMO model version 4.26, with new microphysics package;
4. **H_ENS** ($\Delta x = 16$ km, 62 Model Levels, 21 members), the high-resolution global ENS, driving both "TESTcleps" systems.

In order to compare the skill of the 4 systems, we considered the probabilistic prediction of total precipitation exceeding a number of thresholds for several forecast ranges, analysing only the performance of the runs starting at 12UTC.

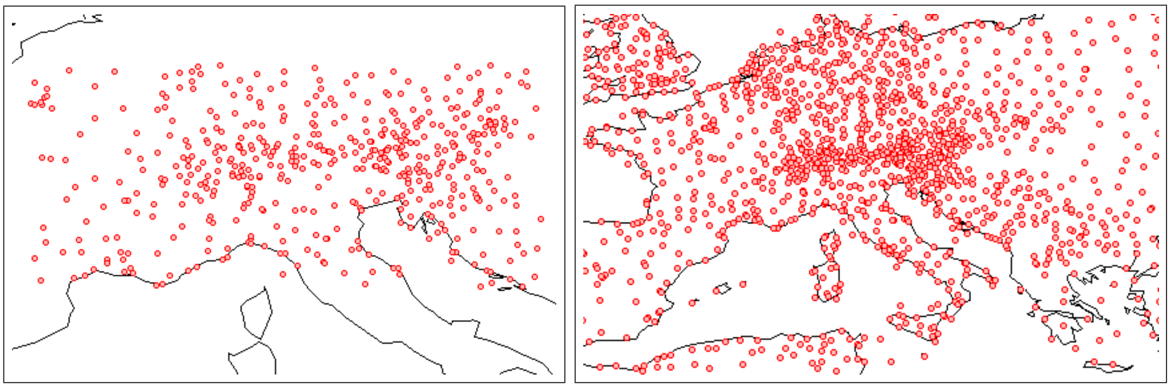

Figure 1: SYNOP stations used for the verification of the ensemble systems: "mapdom" (left panel) and "fulldom" (right panel).

The evaluation of the models' performances was carried out over different verification domains: the former one (referred to as "mapdom") was centred over the Alpine area and covered the area [43-50N, 2-18E]; the latter one (referred to as "fulldom") emcompassed the full COSMO-LEPS domain with the area [35-58N, 10W-30E]. As for observations, it was decided to use the data obtained from the SYNOP reports available on the Global Telecommunication System (GTS). There were approximately 400 (1400) SYNOP reports falling in the "mapdom" ("fulldom") verification domain, as reported in Fig. 1. The comparison of model forecasts against observations was carried out by selecting the model grid-point closest to the observation. The skill of the different systems was examined for 6 different precipitation thresholds: 1, 5, 10, 15, 25 and 50 mm/12h. The following probabilistic scores were computed: the Brier Skill Score (BSS), the Ranked Probability Skill Score (RPSS), the Relative Operating Characteristic Curve (ROC) area, the Rank Histograms (RK) and the Percentage of Outliers (OUTL).

Table 2: Main features of the verification configuration.

variable:	12-hour cumulated precipitation (18-06, 06-18UTC);
starting time:	12 UTC;
period:	all cases of Table 1;
region:	43-50N, 2-18E (mapdom); 35-58N, 10W-30E (fulldom);
method:	nearest grid–point;
observations:	SYNOP reports, ∼ 400 (1400) stations/day for "mapdom" ("fulldom");
fcst ranges:	6-18h, 18–30h, . . . , 114-126h;
thresholds:	1, 5, 10, 15, 25, 50 mm/12h;
scores:	ROC area, BSS, RPSS, OUTL.

For a description of these scores, the reader is referred to Marsigli et al. (2008) and Wilks (1995). The main features of the verification exercise are summarised in Table 2.

3 Results

In this section, the main results of the intercomparison are presented. The forecast accuracy of the 4 systems is measured separately for each of the three periods of Table 2 as well as for the full length of the verification exercise in order to get an overall picture of systems' performances.

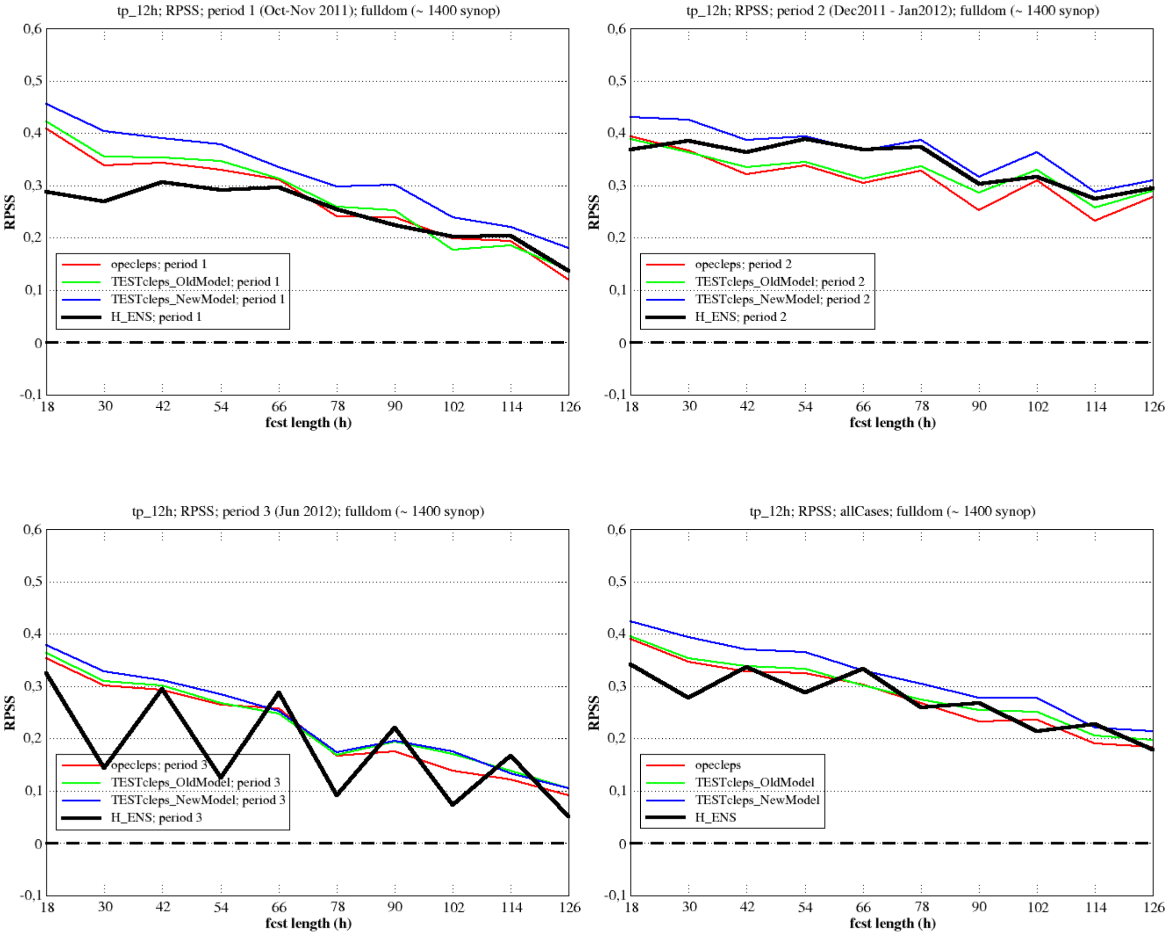

Figure 2: RPSS values for 12-hour cumulated precipitation as a function of the forecast range for period 1 (top-left panel), period 2 (top-right panel), period 3 (bottom-left panel) and all cases (bottom-right panel). The scores are reported over the "fulldom" for the following forecast systems: opecleps (red line), TESTcleps_OldModel (green line), TESTcleps_NewModel (blue line) and H_ENS (thick black line).

As for the "fulldom" area (about 1440 SYNOP reports in the region), the skills of COSMO ensembles and H_ENS are summarised in Fig. 2. The attention is focussed on the probabilistic prediction of 12-hour cumulated precipitation and the values of the RPSS are plotted against the forecast range for each of the three periods of Table 2 and for all cases.

The skill of each individual ensemble system changes from case to case; as an example, the performance of all systems is less accurate for the summer event (period 3, bottom-left panel), which is mainly driven by convection processes over complex terran, with a lower degree of predictability. Nevertheless, some common features can be detected. In the 4 panels, it can be noticed that very similar results are obtained by opecleps and TESTcleps_OldModel (red and green lines, respectively) for almost all forecast ranges.

These two systems differ mainly on the quality of the boundaries, which are provided at higher resolution in the latter configuration. Therefore, it looks as if the benefits of more detailed boundaries are only partly transferred to the skill of the limited-area integrations. A clearer positive impact can be noticed for ranges longer than 78 hours, when TESTcleps_OldModel shows higher scores than opecleps. This is especially true for the verification scores of periods 2 and 3 (top-right and bottom-left panels of Fig. 2, respectively).
Overall speaking, the best results are obtained by the TESTcleps_NewModel configuration (blue line), where COSMO–LEPS benefits of both higher-resolution boundaries by H_ENS and improved model set-up.

The better performance of TESTcleps_NewModel is evident for all ranges and is especially true for short forecast steps. As for the global ensemble H_ENS, it can be noticed that its performance is usually the worst one in the short range, while the system gets more and more valuable for longer ranges. It is also worth pointing out the 12-hour cycle of model performance for period 3 (bottom-left panel). For this case, the skill of the model is higher in the verification interval 18-06 UTC (night-time precipitation), while a worse performance can be seen for day-time precipitation, verifying between 06 and 18 UTC.

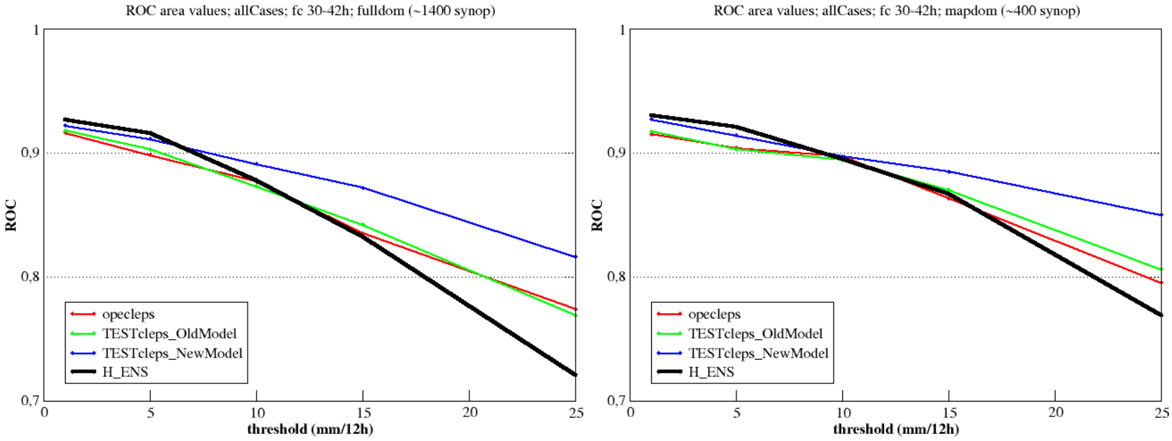

Figure 3: ROC area values for 12-hour cumulated precipitation as a function of threshold value for the forecast range 30-42h. The scores are reported over the "fulldom" (left panel) and "mapdom" (right panel) for all cases and for the following forecast systems: opecleps (red line), TESTcleps_OldModel (green line), TESTcleps_NewModel (blue line) and H_ENS (thick black line).

Similar conclusions can be drawn if the attention is focussed on the skill of the systems in terms of ROC area. Fig. 3 reports the performance of the 4 systems over the forecast interval 30-42h as a function of precipitation intensity. It can be noticed that, on either "fulldom" or "mapdom" (left and right panel, respectively), the highest scores are obtained by TESTcleps_NewModel, with similar performances by opecleps and TESTcleps_OldModel.

As for H_ENS, the performance of the model is satisfactory for low threshold events, while the system shows a performance decay for high-precipitation cases. For all investigated systems, it can also be noticed that higher scores are obtained for the verification in the "mapdom" domain. Also in this case, the performance of all systems is slightly more accurate when verification is performend over the "fulldom" area.

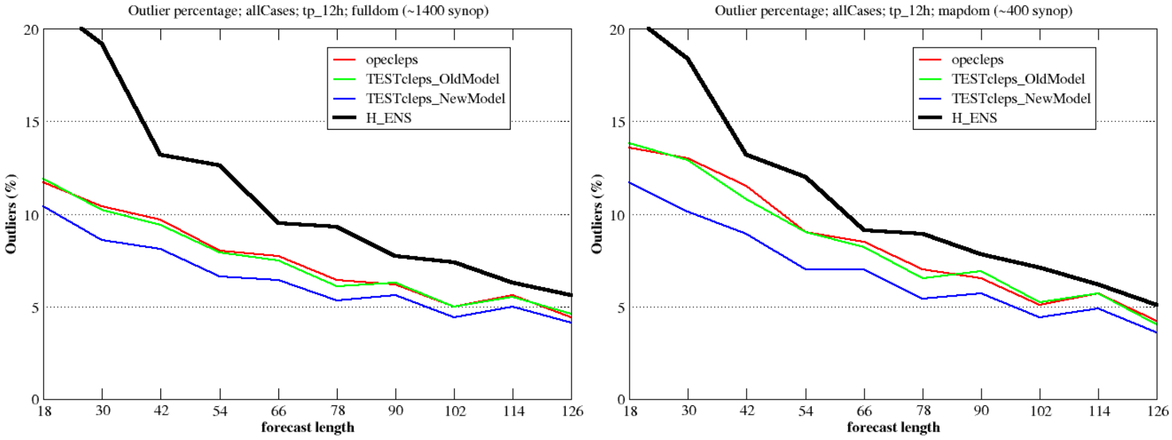

Figure 4: Percentages of Outliers for 12-hour cumulated precipitation as a function of the forecast range. The scores are reported over the"fulldom" (left panel) and "mapdom" (right panel) for all cases and for the following forecast systems: opecleps (red line), TESTcleps_OldModel (green line), TESTcleps_NewModel (blue line) and H_ENS (thick black line).

Finally, the skills of the systems are assessed in terms of Outliers' percentage over the "fulldom" and "mapdom" areas (left and right panels of Fig. 4). Lower values of outliers are found for verification performed in the "fulldom", indicating a better performance of the systems in this region. Despite the larger ensemble size (21 vs 16), the global system H_ENS presents larger amounts of outliers than COSMO runs. This is true for all forecast ranges, but the gap is especially evident in the first 66 hours. As for COSMO systems, the overall higher quality of TESTcleps_NewModel is confirmed. The added value of better boundaries and improved set-up is clearly evident in the short range, while, at longer ranges, the skills of the different COSMO systems converge towards the 5% value of outliers.

4 Conclusions

The main results of the verification exercise carried out in the framework of LAMEPS experimentation can be summarised as follows:

- the impact of using high-resolution boundaries with respect to the operational configuration is limited;
- a clear improvement in limited-area model integrations is obtaibed if, in addition to high resolution boundaries, a newer model version with updated microphysics is used;
- in either cases, the added value with respect to the global ensemble is noticeable, especially in the short range.

As for the future, it is planned to consolidate the verification results, by considering the performance of all system for other variables, considering also the spread/skill performance for the different periods.

References

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

[2] Montani A., Cesari D., Marsigli C., Paccagnella T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO–LEPS system: main achievements and open challenges. *Tellus*, **63A**, 605-–624. DOI: 10.1111/j.1600-0870.2010.00499.x

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