Main results of the SREPS Priority Project

Chiara Marsigli, Andrea Montani, Tiziana Paccagnella

ARPA-SIM, Bologna, Italy

Flora Gofa, Petroula Louka

HNMS, Athens, Greece

1 Introduction

The SREPS Priority Project focussed on the building up of a high-resolution ensemble system for the short-range. The project main tasks were to develop and implement such an ensemble, then to run it over extensive testing periods and to evaluate the system features and performances.
This system has been built to fulfil some needs that have arisen in the COSMO community:

- to have a short-range mesoscale ensemble to improve the support especially in situations of high impact weather
- to have a very short-range ensemble for data assimilation purposes
- to provide boundary conditions for the COSMO-DE-EPS convection-resolving ensemble, currently under development at DWD.

Hence, the strategy to generate the mesoscale ensemble members tried to take into account as many as possible sources of uncertainty which affect the scales of interest in the weather forecast at the short time range, in order to model many of the possible causes of the relevant forecast errors.
The main issues which have been addressed in the system evaluation are: 1) if the system shows a good spread/skill relationship, representative of the capability of the ensemble in describing the forecast error 2) how the different perturbations contribute to the spread and to the skill of the system 3) which is the ensemble skill in the forecast of surface weather parameters.

2 System description and methodology of analysis

COSMO-SREPS (COSMO Short-Range Ensemble Prediction System) it is based on 16 integrations of the limited-area non-hydrostatic COSMO model at about 10 km of horizontal resolution, with 40 vertical levels.
The driving model error is described by means of a multi-analysis multi-boundary approach. Initial and boundary condition perturbations are applied by driving the 10-km COSMO runs with the four 25-km COSMO members of the Multi-Analysis Multi-Boundary SREPS system of AENM. These four lower resolution COSMO runs, nested on four different global models (IFS, GME, GFS, UM) which use independent analyses, are provided by INM for this purpose. A representation of the smaller scale uncertainty is accomplished by applying limited-area model perturbations to the 10-km COSMO runs. In particular, 4 different set-up

of the model physics have been adopted in the ensemble members: 1) control set-up 2) use of the Kain-Fritsch scheme for the parametrisation of the deep convection, instead of Tiedtke as in the control 3) tur_len parameter equal to 1000 instead of 500 as in the control 4) pat_len parameter equal to 10000 instead of 500 as in the control. The combination of the 4 possible choices for the driving run with the 4 possible choices for the physics set-up lead to the 16 member ensemble.
During the project, the system was run over two main testing periods:

- 21 selected days of Autumn 2006, characterised by intense precipitation over either the Alpine area or Germany
- the MAP D-PHASE DOP (June to November 2007).

During the D-PHASE OP, 99 full runs of the COSMO-SREPS system were performed, covering not continuously the period, 50 in summer (JJA) and 49 in autumn (SON). Each full run (made up of 16 COSMO-model integrations at 10 km) started at 00UTC. The lack of continuity in the runs was mainly depending on the availability of initial and boundary conditions provided by INM.
The analysis of the system was carried out over two COSMO regions: the Alpine area and Greece.
This is due to the availability of observations and to the COSMO scientists involved in the project. The climatology of two regions is very different, but both regions are quite complex from the geographical point of view (orography, proximity of the sea). In particular, it should be underlined that in Greece few and less intense precipitation events were observed during the D-PHASE period (this is why also the month of December has been included in the sample for Greece) and that summer 2007 was a remarkably hot one.
Different data-sets have been used for the evaluation:

- high-res alpine: a dense network of stations covering Northern-Central Italy and Switzerland, providing precipitation data accumulated over 24h, from 06 to 06 UTC (about 1400 stations)
- high-res Italy: a dense network of stations covering Northern-Central Italy, providing precipitation data accumulated over 6h (about 900 stations) and 2m temperature data (about 600 stations)
- synop alpine: the SYNOP stations covering approximately the same area (43-48 N 6-14 E, 218 stations)
- synop Greece: the SYNOP stations covering Greece (about 90 stations)

3 Results

3.1 The spread-error relationship

The evaluation of the spread-error relationship was carried out on the Alpine area only, showing that the system tends to be under-dispersive. The gap between the spread and the error has been observed for a number of meteorological variables, both surface and upper-air (2m temperature, mean-sea-level pressure, precipitation, temperature at 850 hPa, geopotential height at 500 hPa). Moving from towards upper-air variables, the gap decreases, but it is still detectable (not shown).

In Figure 1, the root-mean-square error of the ensemble mean and the root-mean-square spread of the ensemble (or ensemble standard deviation) are compared for the two seasons, in terms of 2m temperature. The error is computed by comparing forecasts interpolated on station points belonging to the synop alpine dataset with the corresponding observations; the spread is computed using these same interpolated forecast values, for homogeneity reasons.

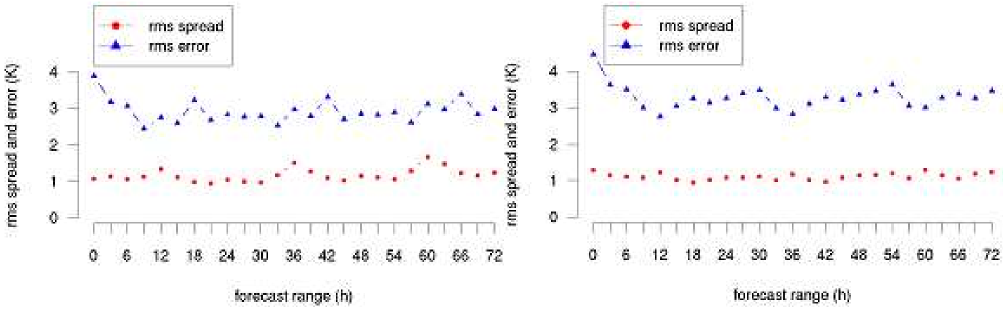

Figure 1: COSMO-SREPS spread (red) and error (blue) in terms of 2m temperature for summer (left) and autumn (right) 2007. Data are from the synop alpine dataset.

The ensemble spread is bounded between 1 and 2^oK in the summer season (Fig. 4, left panel), increasing with the forecast range and exhibiting a diurnal cycle, with values peaking at noon. In autumn (right panel) the spread stays close to 1^oK throughout the whole forecast range. In both seasons the ensemble mean error is quite larger than the spread, remaining below the 3^oK value in summer, with peaks grater than 3^oK at 18 UTC, while being generally above the 3^oK value in autumn. The gap between the two measures is due to both the underdispersion of the ensemble system and to the COSMO model systematic error, which should not be removed by ensemble techniques, but only by model improvement.

A better representation of the spread/skill relationship of the ensemble is shown in Figure 2, where the rms error is plotted as a function of the rms spread, after having divided the sample in classes of spread and computed for each class the average values of error and spread.

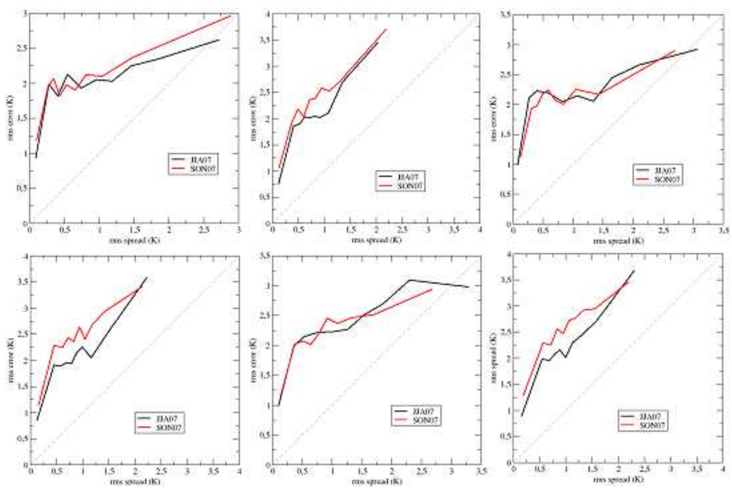

Figure 2: Spread/error relationship in terms of 2m temperature for summer (black line) and autumn (red line) 2007 for different forecast ranges (+12h, +24h, +36 h in the upper row, +48h, +60h, +72h in the lower row). Data are from the synop alpine dataset.

There is a clear correlation between error and spread, though at a given value of spread generally corresponds an higher value of error, even double. During the night (second, fourth and sixth panels, where data are from 00 UTC), the ensemble underdispersion is less marked in the summer season. At 12 UTC (first, third and fifth panels) there two seasons exhibit a more similar behaviour, with a good relationship for high spread values.

3.2 How different perturbations contribute to the ensemble skill

In order to assess the contribution to the skill of the system provided by the different ensemble members, verification of the performances of the 16 runs has been also made, both in terms of temperature and precipitation.
In terms of 2m temperature forecasts, the GME-driven members exhibit a peculiar behaviour in both the Alpine area and over Greece: the GME-driven members have better performance over both regions. Scores over Greece are shown in Figure 3.

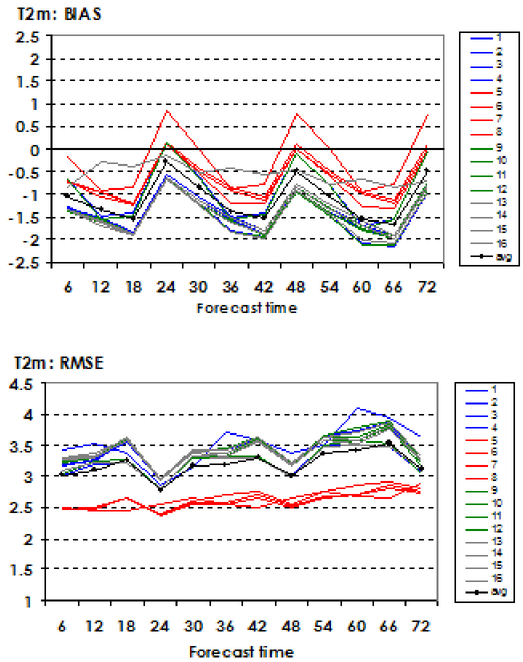

Figure 3: 2m temperature bias (upper panel) and RMSE (lower panel) for the 16 COSMO-SREPS runs, computed over Greece (lower panel, synop Greece dataset) for the whole period (June to December 2007).

It is worth pointing out that only the initial and boundary conditions provided to the GME-driven members are characterised by a coherence between soil and atmosphere. In fact, in the 25-km COSMO runs performed by INM, the atmospheric fields are provided by the 4 different global models, while the soil fields are always provided by the GME run. Hence, the coherence between atmosphere and soil in the "father" runs can have a positive influence on the forecast of 2m temperature by the GME-driven members.
In order to quantify the contribution of each type of perturbation to the ensemble skill, the scores of sub-ensembles made up by homogeneous members have been computed. The 16 members can be subdivided into 4 groups of 4 elements each, in 2 different ways:

- considering groups of elements homogeneous in terms of initial and boundary conditions, but distinct for the model parameterisations;

- considering groups of elements homogeneous in terms of the model parameters, but distinct in terms of initial and boundary conditions.

Considering 24h precipitation forecast, the scores have been computed for each of the 4 groups of 4 elements, in order to assess how the different forecast characteristics in terms of driving model and parameter contribute to the skill. The ROC area of the 4-member sub-ensembles are shown in Figure 4 for the autumn season and for the alpine area. The light blue line of each panel represents the ROC area of the full 16-member ensemble, which gives an indication of the COSMO-SREPS skill in forecasting precipitation for that period and in that particular area.

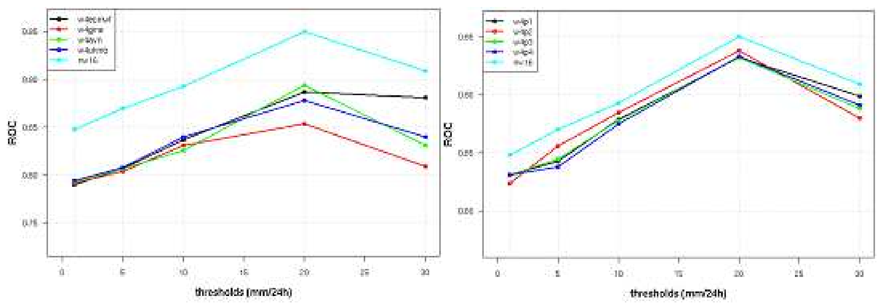

Figure 4: ROC area as a function of threshold for 24hr accumulated precipitation in the alpine area (high-res alpine data set) for the autumn season. Left panel: full 16-member COSMO-SREPS (light blue line) vs. 4-member ensembles with identical 'mother run' (black: ECMWF, red: GME, green: GFS, blue: UKMO). Right panel: full 16-member COSMO-SREPS (light blue) vs. 4-member ensembles with identical physical perturbation (black: p1, red: p2, green: p3, blue p4). Scores are for the +30 h forecast range.

Apart from the decrease in skill evident when passing from a 16-member to a 4-member ensemble, which is expected, it is worth pointing out that the different 4-member ensembles have different skill, which varies with the considered forecast range and also with threshold. In the right panel, the members of each sub-ensemble have identical physics perturbations. Therefore, these lines represent the skill of ensembles, which are perturbed in the initial and boundary conditions only, but have the same model set-up. Comparing each right panel with the corresponding left one suggests that perturbation of initial conditions generally yields more skilful performance than physical perturbation only. This is an indication of the fact that, the higher degree of diversity among members introduced by perturbing initial and boundary conditions determines a greater amount of skill with respect to the smaller-scale diversity introduced by the physics perturbations. As for the role of the different parameterizations, the 4-member ensemble where model perturbation p2 (Kain-Fritsch convection scheme) is applied to each member (red line on the right panels) turns out to be more skilful that other 4-member ensembles.
The same evaluation has been performed also for Greece. Scores over Greece are shown in Figure 5, for the whole period and at a +48h forecast range.

From these results is difficult to judge which driving-model leads to more skilful forecast, the results being dependent on the geographical area, on the season (not shown), on the forecast range (not shown) and on the precipitation threshold. As for the different parameter choice, we should be careful in the evaluation of the Kain-Fritsch members. They have the best performance in terms of ROC area but the worst in terms of BSS (not shown), due to the fact that they always tend to produce slightly too much rain (not shown).

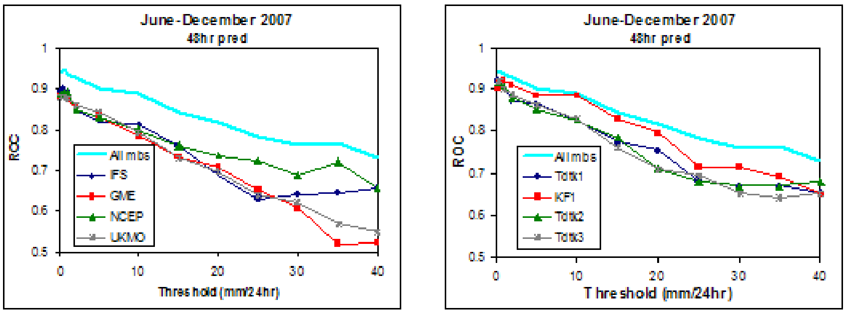

Figure 5: ROC area as a function of threshold for 24hr accumulated precipitation over Greece (synop Greece data set) for the whole period (June to December 2007). Left panel: full 16-member COSMO-SREPS (light blue line) vs. 4-member ensembles with identical 'mother run' (blue: ECMWF, red: GME, green: GFS, grey: UKMO). Right panel: full 16-member COSMO-SREPS (light blue) vs. 4-member ensembles with identical physical perturbation (blue: p1, red: p2, green: p3, grey p4). The forecast range is +48 h.

3.3 Study on parameter perturbations

Beside the COSMO-SREPS suite, a parallel suite, called CSPERT, was implemented and run continuously at ECMWF (but not in real time) for the whole Autumn 2007 (September-October-November, 91 runs). Since the preliminary tests on COSMO-SREPS had already identified a lack of spread due to due to an incomplete description of model uncertainty sources, this parallel suite was generated to choose more parameter perturbations for future implementation in COSMO-SREPS. This is necessary to increase the spread to values closer to the COSMO model error, especially for surface variables. The 16 perturbations involves also physical packages such as cloud and land schemes which had not been considered before. Initial and boundary conditions for the 16 runs were provided by the same run: the operational deterministic integration of ECMWF. The runs were starting daily at 00 UTC and the forecast range was 24 hours only.
The impact of the different set-up of the 16 runs on the selected meteorological variables is summarised in Table 1. The scores obtained by the 15 perturbed runs are evaluated against the score of the control run and a colour is assigned according to the performance:

- red: the perturbed run is worse than the control
- yellow: the perturbed run is slightly worse than the control
- light green: the perturbed run is slightly better than control
- green: the perturbed run is better than the control
- grey: the perturbed run is equivalent to the control
- white: no evaluation is possible, since the result changes with the forecast range

Looking at this table it is evident that none of the runs performs continuously better than the control, so that its set-up can be used as the new control set-up. Some improvement is possible by choosing rlam_heat=10, crsmin=200 and tur_len=1000, but it is yet to be investigated what effect will be if these three values are implemented in the same run. Instead, the choice of rlam_heat=0.1, c_lnd=1, c_soil=2 and tur_len=150 has lead to a worsening of the performances.
The fact that statistical behaviour of the various parameter set-ups "fluctuates" with respect

to the control run (it is not always better or worse) should be regarded as a positive outcome in this ensemble framework, since ensemble perturbations should be almost equivalent. The only set-ups which should be discarded are those which do not produce any (or a very small) impact (e.g. tur_len parameter).

	t BIA	t MAE	td BIA	td MAE	tp1 BS	tp1 TS	tp1 FA	tp10 BS	tp10 TS	tp10 FA
KF										
tur_len=150										
tur_len=1000										
pat_len=10000										
rat_sea=1										
rat_sea=60										
qc=0.001										
crsmin=50										
crsmin=200										
c_soil=0										
c_soil=2										
c_lnd=1										
c_lnd=10										
rlam_heat=0.1										
rlam_heat=10										

Table 1: Summary of the performances of the 15 perturbed runs with respect to the control.

4 Conclusions and future work

Some conclusions which can be drawn from the project are listed hereafter:

- there is a correlation between error and spread, but the system is under-dispersive, especially for surface variables
- the use of different driving models seems to dominate with respect to physics parameter perturbations as regards the contribution to the spread; these contributions are quite different in the two seasons in terms of 2m temperature
- the different driving models contribute differently to the ensemble skill, but the relative skill is strongly dependent on forecast range, season, verification area
- the different physics perturbations contribute differently to the ensemble skill as well
- for 2m temperature forecast, the GME-driven members perform generally better
- for precipitation forecast, perturbations of the convective schemes are more important than the perturbations of the particular parameters for turbulent and length scales used
- it seems that the members with Kain-Fritsch convective scheme have better probabilistic resolution but they overestimate precipitation more than Tiedtke scheme

Future work about the ensemble will be part of the new Priority Project CONSENS. In particular the work will focus mainly on:

- introduce the new parameter perturbations tested in the CSPERT suite, after an analysis of their impact also in a summer season
- analyse the impact of combine perturbations
- add perturbations of the lower boundary of the model
- combine the COSMO-SREPS ensemble with the COSMO-LEPS one, in a scientifically sound way