Consortium

for

Small-Scale Modelling

Technical Report No. 13

COSMO Priority Project "Short Range Ensemble Prediction System (SREPS)": Final Report

June 2009

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_13

www.cosmo-model.org

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

COSMO Priority Project “Short Range Ensemble Prediction System (SREPS)”: Final Report

Chiara Marsigli
ARPA EMR - Servizio Idro-Meteo-Clima
Viale Silvani 2/3
40122 Bologna
Italia

Contents

Contents

1 Summary

The project focused on building an ensemble system for the short-range using the COSMO model, called COSMO-SREPS. During the first phases of the project, the ensemble was designed and implemented, then it was tested over long periods, in order to derive a robust statistical assessment of its features. In particular the system was running during the whole DOP of the MAP D-PHASE project (June to November 2007). The analysis of its performances was carried out over two COSMO regions: an alpine area and Greece. The main findings are:

- A good correlation between spread and error is observed, but 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; the gap decreases moving from surface towards upper-air variables. This evaluation is influenced by the presence of a model Bias, especially evident for 2m temperature.
- It has been recognized the need of applying more model perturbations. An extensive testing of new and more diverse parameter perturbations has been carried out within the project, but it is needed to continue this testing in the following of the project, due to the great amount of time and resources required. Furthermore, it has been decided to add also perturbations in the lower boundary of the COSMO model, to represent better the model uncertainties.
- The different types of perturbation contribute differently to the spread and to the skill of the system. In particular, driving model perturbation is the main source of spread in the ensemble as well as the main source of ensemble skill. COSMO model perturbation (parameter perturbation) plays a minor role, but contributes to the ensemble spread as well, with an amount variable according to the considered meteorological variable.
- The quality of the different driving models and of the different parameter choices are generally equivalent, since none of the different components of the system is always over-performing the others. The only exceptions are: the use of GME-driven initial and boundary conditions improve the system performances in terms of 2m temperature in summer, probably due to the coherence between soil and atmospheric fields; the use of the Kain-Fritsch scheme for the parametrization of the deep convection improves the precipitation detection at the expense of a larger number of false alarms.
- The ensemble skill in the forecast of surface weather parameters is reasonable. An objective assessment of the system quality will be carried on in the framework of the D-PHASE project, where the system performances will be compared with those of other state-of-the-art ensemble systems.

2 Project report

2.1 Introduction

The SREPS Priority Project focused 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. The ensemble is called COSMO-SREPS (COSMO Short-Range Ensemble Prediction System) and 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. This system has been built to fulfil some needs that have recently 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.

Therefore, the ensemble had to be designed to describe the uncertainty affecting the short-range predictions of surface weather parameters at a high spatial resolution. Aiming at this purpose, the strategy to generate the mesoscale ensemble members proposed by this project tried to take into account as many as possible sources of uncertainties 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. Hence, perturbations have been applied both in the driving model and in the mesoscale model. 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 INM. 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 as well: the values of a number of parameters included in the sub-grid process parametrization schemes are randomly changed (within their range of variability) in the different ensemble members. The main issues which have been addressed in the system evaluation are:

- if the system shows a good spread/skill relationship, representative of the capability of the ensemble in describing the forecast error
- how the different perturbations contribute to the spread and to the skill of the system
- which is the ensemble skill in the forecast of surface weather parameters.

2.2 System description and implementation

2.2.1 The COSMO-SREPS system

COSMO-SREPS takes both initial and boundary conditions from 4 integrations of the COSMO-model, performed over the Euro-Atlantic area at 25 km of horizontal resolution, and differentiated from each other according to the global model used to drive the runs. The 4 integrations at 25 km are nested on the following global forecasting systems: ECMWF (IFS), DWD (GME), NCEP (NCEP) and UKMO (UM). Since these global systems are 4 different and independent state-of-the-art forecasting chains, a good deal of differentiation in the global models driving the limited-area runs is ensured. These integrations are performed by the Spanish weather Service (INM), which provides directly the fields generated by the 4 runs of the COSMO-model at 25 km. For each of the 4 sets of initial and boundary conditions, the COSMO-model integrations at 10 km of horizontal resolution are then performed over a domain covering Central and Southern Europe (Fig. 1). Each 25-km run drives 4

10-km integrations, which differentiates from one another for different choices of the parameters representative of sub-grid physical processes. Tab. 1 reports the parameters which were selected, together with their range of variability and the values selected for the experiments, as suggested by the COSMO scientists responsible for the development of parameterizations schemes. Tab. 2 shows the set-up of each ensemble member run in terms of both driving model and physics.

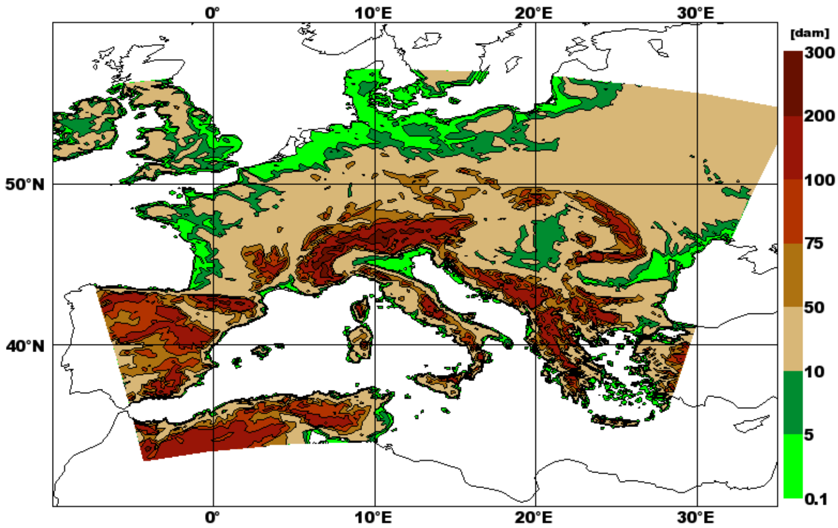


Figure 1: Integration domain of COSMO-SREPS (the model orography is shown).

Table 1: Perturbations applied to the model.

Perturbation name	Description	parameter range	default value	used value
SCHEMES				
Tiedtke	Scheme used for the parametrization of the deep convection		Tiedtke	Tiedtke
Kain-Fritsch	Scheme used for the parametrization of the deep convection		Tiedtke	Kain-Fritsch
PHYSICS				
pat_len	length scale of thermal surface patterns (influences mixing in the stable stratified atmosphere)	[0,10000] m	500	0, 10000
tur_len	maximal turbulent length scale (influences mainly stratospheric mixing)	[100,1000] m		1000

Table 2: Main features of the 16 runs which constitute the ensemble.

Member	Driving model of the LM-INM run	P1	P2	P3	P4
1	IFS	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=500
2	IFS	tiedtke=.F.	kainfri=.T.	tur_len=-1	pat_len=500
3	IFS	tiedtke=.T.	kainfri=.F.	tur_len=1000	pat_len=500
4	IFS	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=10000
5	GME	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=500
6	GME	tiedtke=.F.	kainfri=.T.	tur_len=-1	pat_len=500
7	GME	tiedtke=.T.	kainfri=.F.	tur_len=1000	pat_len=500
8	GME	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=10000
9	NCEP	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=500
10	NCEP	tiedtke=.F.	kainfri=.T.	tur_len=-1	pat_len=500
11	NCEP	tiedtke=.T.	kainfri=.F.	tur_len=1000	pat_len=500
12	NCEP	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=10000
13	UKMO	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=500
14	UKMO	tiedtke=.F.	kainfri=.T.	tur_len=-1	pat_len=500
15	UKMO	tiedtke=.T.	kainfri=.F.	tur_len=1000	pat_len=500
16	UKMO	tiedtke=.T.	kainfri=.F.	tur_len=-1	pat_len=10000

2.2.2 Testing period and data

The system was run over two main testing periods:

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

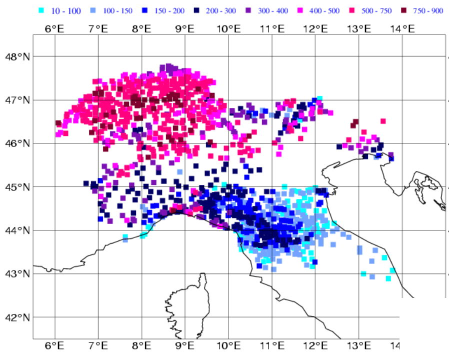
During Autumn 2006, 21 runs of the system were performed, on selected days characterized by intense precipitation over either the Alpine area or Germany. Fourteen runs were starting at 00 UTC, 7 runs at 12 UTC, on the basis of the availability of initial and boundary conditions. During the D-PHASE OP, 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
- 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 warm 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) (Fig. 2 left)
- 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) (Fig. 2 right)
- synop Greece: the SYNOP stations covering Greece (about 90 stations) (Fig. 3)

Though the system performance was analyzed over both 2006 and 2007 testing periods, results are shown for the 2007 period only, due to the larger sample, which makes the analysis more statistically robust.



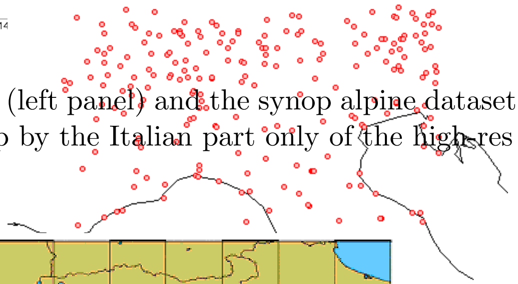


Figure 2: The high-res alpine dataset (left panel) and the synop alpine dataset (right panel). The high-res Italy dataset is made up by the Italian part only of the high-res alpine one.

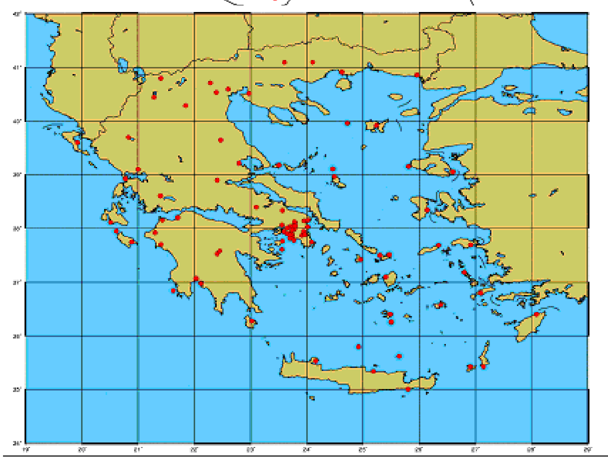


Figure 3: The synop Greece dataset.

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 still 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 surface towards upper-air variables, the gap decreases, but it is still detectable. In Fig. 4, 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 data-set with the corresponding observations, the spread is computed using these same interpolated forecast values, for homogeneity reasons.

The ensemble spread is bounded between 1 and 2 K in the summer season (Fig. 4, left panel), increasing with the forecast range and exhibiting a diurnal cycle, with values peaking at noon.

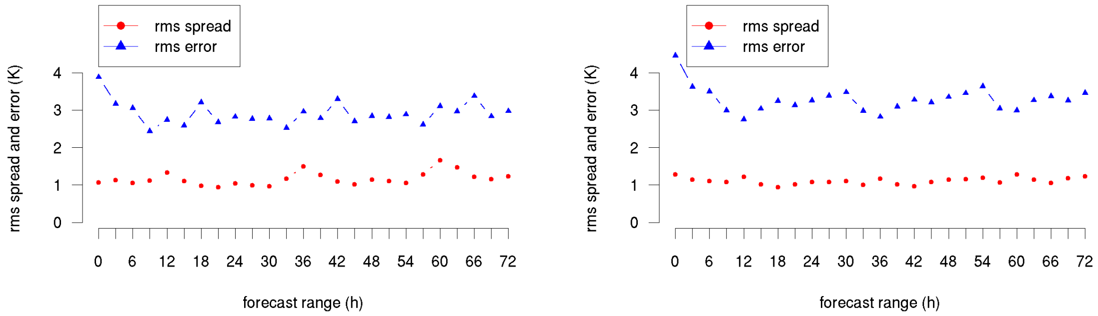


Figure 4: COSMO-SREPS spread (red) and error (blue) in terms of 2m temperature for summer (left) and autumn (right) 2007. Data are from the synop alpine dataset.

In autumn (right panel) the spread stays close to 1 K throughout the whole forecast range. In both seasons the ensemble mean error is quite larger than the spread, remaining below the 3 K value in summer, with peaks grater than 3 K at 18 UTC, while being generally above the 3 K value in autumn. The gap between the two measures is due to both the under-dispersion of the ensemble system and to the COSMO model systematic error, which should not be removed by ensemble techniques, but only by model improvement. In order to help in quantifying the contribution due to the model Bias, the same evaluation is repeated by using the COSMO-I7 analyses of 2m temperature instead of observed values in the computation of the ensemble mean error (Fig. 5). COSMO-I7 analyses are available at 00 and 12 UTC only.

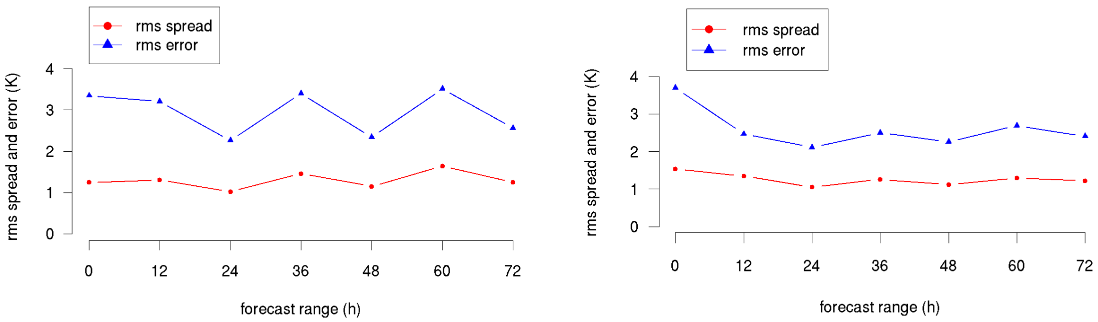


Figure 5: COSMO-SREPS spread (red) and error (blue) in terms of 2m temperature for summer (left) and autumn (right) 2007. Data are from the analysis of the COSMO model itself.

The root-mean-square spread values are the same, while the ensemble mean error is now reduced in both seasons. In summer the reduction is quite small, while is autumn the error is now bounded between 2 and 3 K, greatly improving the spread-error relationship. This indicates a dominance of the systematic error in autumn, which should be removed by model improvement and not addresses with ensemble techniques. A better representation of the spread/skill relationship of the ensemble is obtained by plotting the rms error as a function of the rms spread, after having divided the sample in classes of spread and computing for each class the average values of error and spread. This is shown in Fig. 6 for the 2m temperature parameter, using data from the synop alpine data-set.

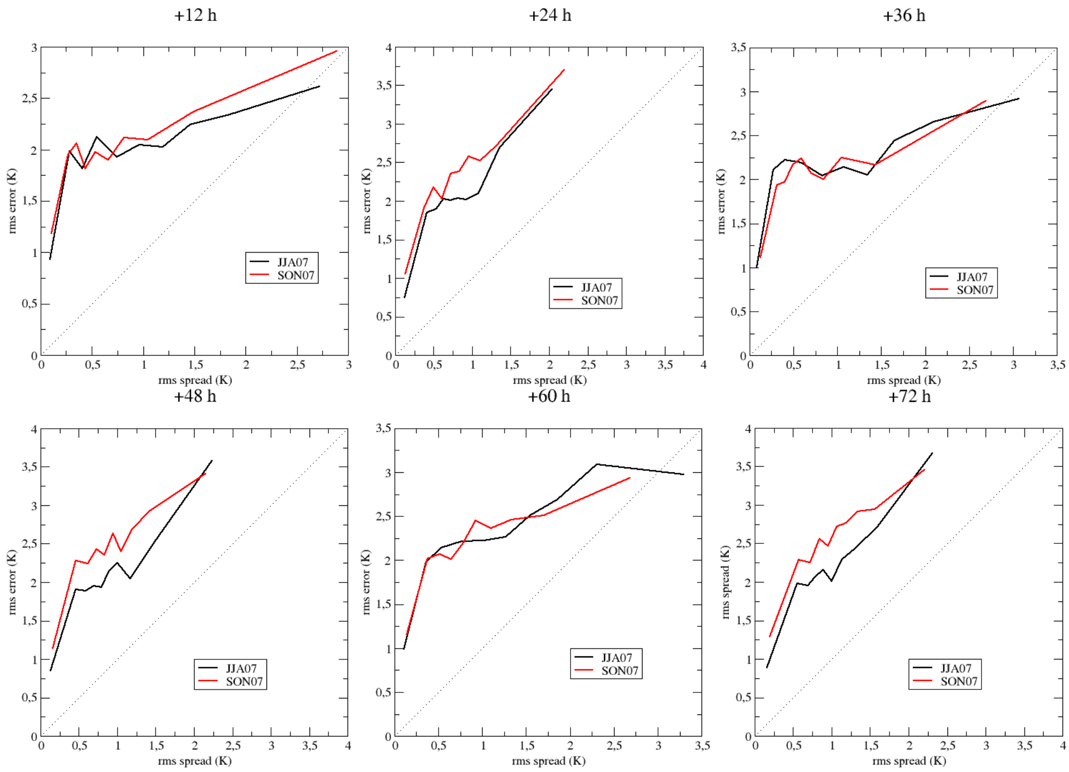


Figure 6: Spread/error relationship in terms of 2m temperature for summer (black line) and autumn (red line) 2007 for different forecast ranges (+12h, +24h, +36 h in the upper row, +48h, +60h, +72h in the lower row). Data are from the synop alpine dataset.

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 under-dispersion 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. Moving from surface towards upper-air variables, the relationship exhibits only a little improvement. In Fig. 7, spread and error values in terms of temperature at 850 hPa are shown. The spread goes from 0.5 K at +12 h, increasing up to more than 1 K at the end of the period. The error grows similarly, being around the 2 K value. The spread/error relationship is shown in Fig. 8, indicating again the existence of a correlation but in presence of under-dispersion, though slightly less evident than for the surface variable.

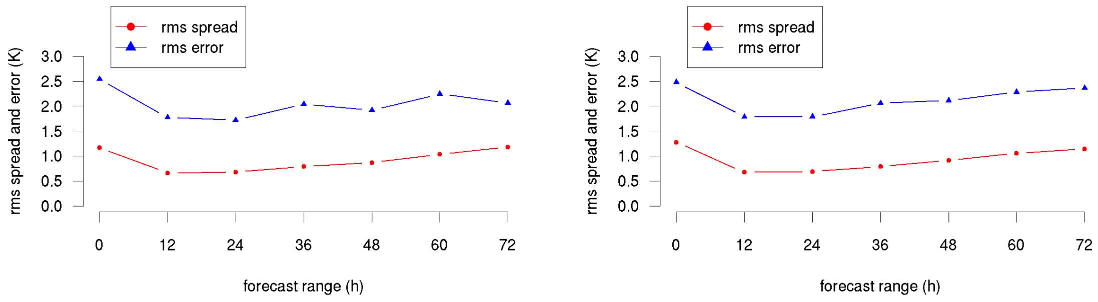


Figure 7: COSMO-SREPS spread (red) and error (blue) in terms of temperature at 850 hPa for summer (left) and autumn (right) 2007. Data are from the analysis of the COSMO model itself.

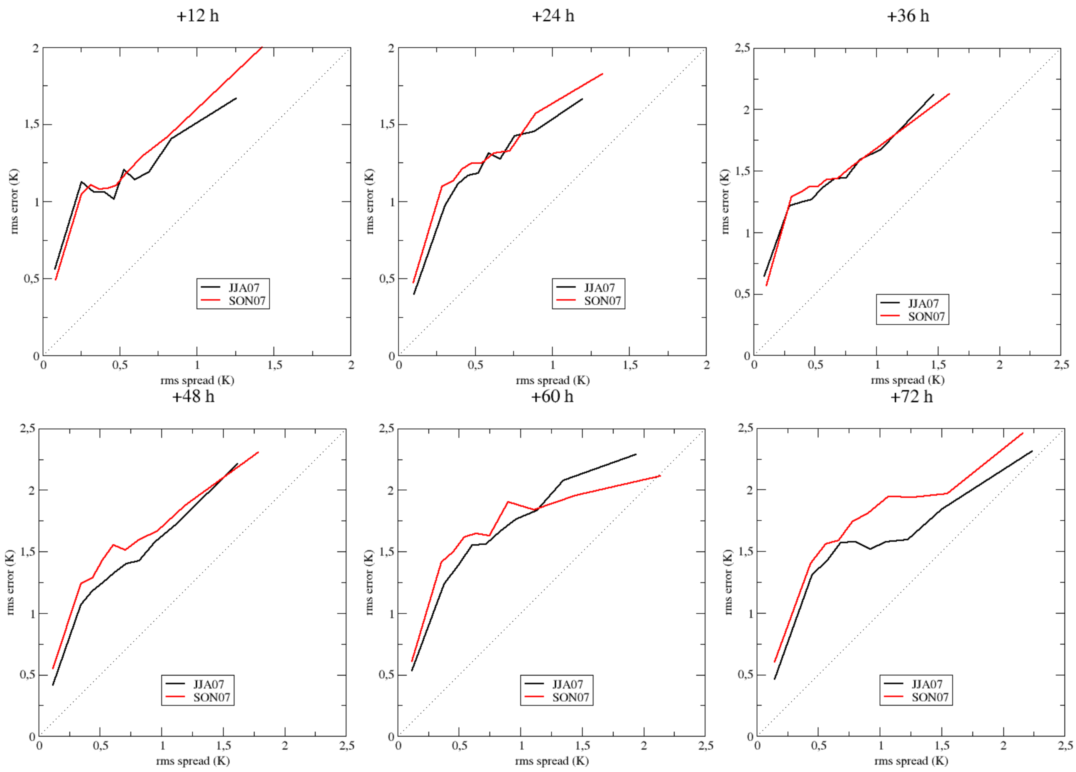


Figure 8: Spread/error relationship in terms of temperature at 850 hPa for summer (black line) and autumn (red line) 2007 for different forecast ranges (+12h, +24h, +36 h in the upper row, +48h, +60h, +72h in the lower row). Data are from the COSMO analyses.

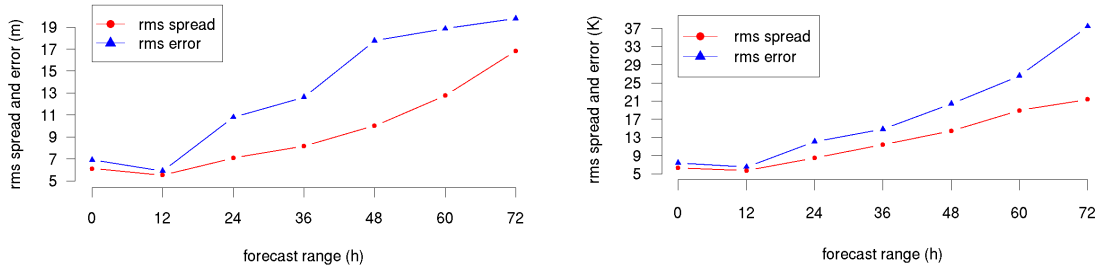


Figure 9: COSMO-SREPS spread (red) and error (blue) in terms of geopotential height at 500 hPa for summer (left) and autumn (right) 2007. Data are from the analysis of the COSMO model itself.

In Fig. 9, spread and error values in terms of geopotential at 500 hPa are shown. For this mid-tropospheric variable the ensemble spread is better able to describe the error, especially for the autumn season (right panel).

3.2 How different perturbations contribute to the spread

In order to quantify the contribution of each type of perturbation to the generation of spread in the ensemble, we have analyzed how the inner spread varies according to the two different methods used to group the ensemble elements. 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 parameterizations;
- considering groups of elements homogeneous in terms of the model parameters, but distinct in terms of initial and boundary conditions.

The spread internal to the groups of the first type is due to parameter perturbations only (parameter spread), while the spread internal to the groups of the second type is due to driving-model perturbations only (driving-model spread). Each group is referred to after the common feature of their members, e.g. the group named *ecmwf* contains all the runs driven by the ECMWF global model, which have different parameter perturbations, while the group named *p1* contains all the runs with the same physics perturbation No. 1, with different initial and boundary conditions. For a complete characterizations of the 16 elements, see Tab. 2. In Fig. 10 the different spread components in terms of geopotential height at 500 hPa are shown.

The 4 upper lines represent the contribution to the spread due to different driving models only, while the 4 lower lines represent the contribution to the spread due to different physics parameters only. The driving-model spread is quite larger than the parameter one and increase rapidly with the forecast range. In terms of this mid-tropospheric variable, the perturbation of the parameters plays an almost negligible role, the parameter spread being about one order of magnitude lower than the driving-model one. In Fig. 11, the contributions to the spread are shown in terms of temperature at 850 hPa.

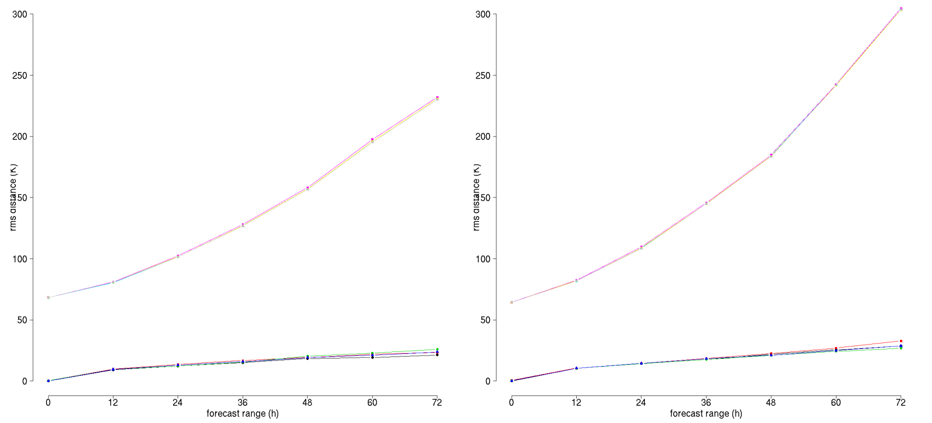


Figure 10: Spread (root-mean-square distance) of the ensemble members for the different forecast ranges, in terms of geopotential height at 500 hPa for the 50 summer runs (left panel) and for the 49 autumn runs (right panel). Each line represents the inner spread relative to a particular 4-member sub-ensemble. For each line, the legend denotes the feature which is common for a particular configuration of the 4-member sub-ensemble.

For this variable the driving-model spread is greater than the parameter one, but the two contributions are of the same order of magnitude. Parameter perturbations produce a spread of about 0.4-0.8 K, almost constant as forecast range increases, while driving-model perturbations produce a spread which ranges from 1 to 2 K, increasing with the forecast range.

Considering a surface variable, namely 2m temperature (Fig. 12), the spread components exhibit a different behaviour. In both season a larger contribution to the spread comes from the use of different initial and boundary conditions. In the summer season (left panel), the contribution is twice the one related to model physics perturbations, this holding throughout the entire forecast range. In the autumn season (right panel), instead, the contribution to the spread due to different initial and boundary conditions is only slightly larger than the one due to different physics configurations and the difference between the two tends to increase with increasing forecast range. It is worth mentioning that, locally on sub-areas or for single events, the main contribution can be provided for certain time ranges by the physical perturbations (not shown). It is also interesting to notice that the parameter spread (four lower lines in the panels) does not vary considerably depending upon the global model used to provide initial and boundary conditions to the COSMO-SREPS runs at which perturbations are then applied. As for the driving-model spread (four upper lines in the panels), this does not vary considerably depending upon the physics set-up adopted for the runs, with the exception of perturbation *p4* (grey line), which tends to give rise to lower spread.

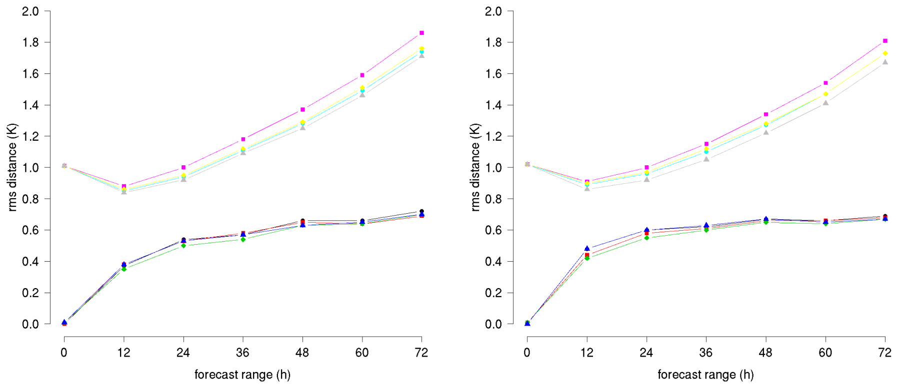


Figure 11: Same of Figure 10 but for temperature at 850 hPa.

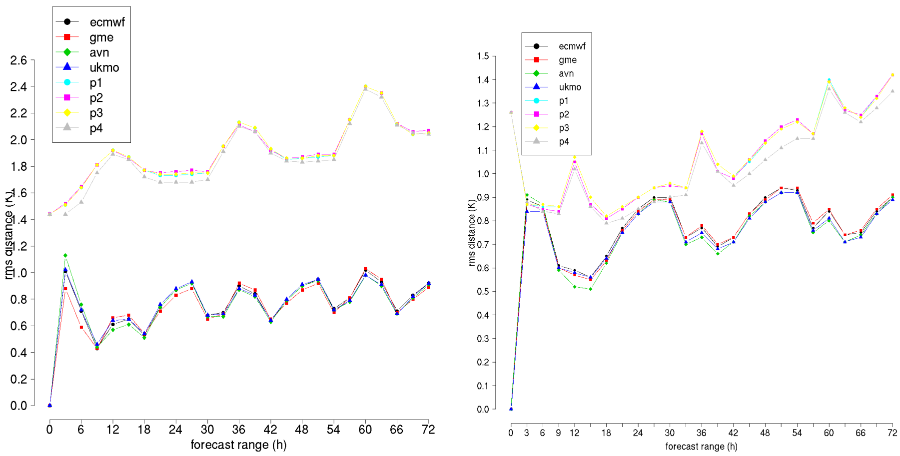


Figure 12: The same as in Figure 10 but in terms of 2-metre temperature. The sample is the high-res Italy dataset.

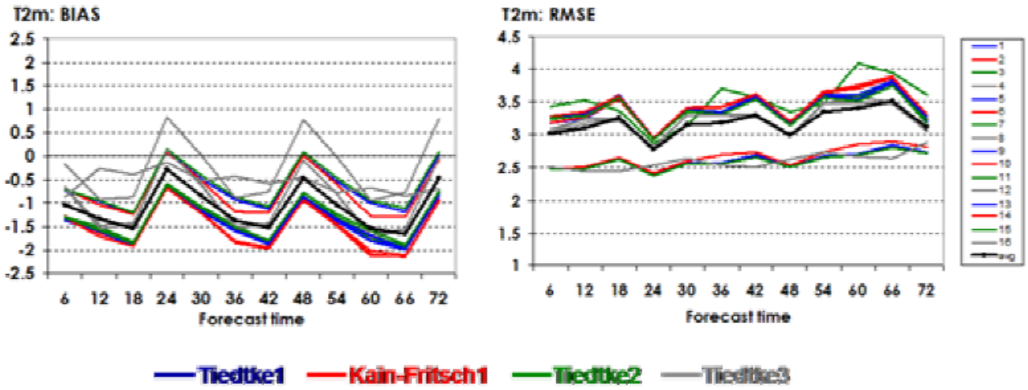


Figure 13: Bias and RMSE for temperature over Greece plotted with respect to effect of physical parameterisation. The thick black line represents the average of all the ensemble members.

A similar evaluation of the spread contribution in terms of 2m temperature was made over Greece (Fig. 13). Temperature Bias and RMSE of each ensemble member have been grouped with respect to the physical parametrization used for that member. It is indicated that that physics parameter perturbations were not producing a big impact in differentiating the scores obtained by the members.

3.3 How different perturbations contribute to the 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. Results in terms of 2m temperature for summer 2007 are shown in Fig. 14 for northern and central Italy.

It is evident that the Bias of the GME-driven members is different from that of the other members, which are quite mixed up. In particular, GME-driven members are always warmer than the others. In terms of MAE, again the GME-driven members exhibit a peculiar behaviour, their error being the lowest. This indicates that the errors of the members of this group have lower absolute values, but probably most of them are of the "warmer type", hence sum up in the Bias with equal sign determining a Bias increase. Instead, the non-GME-driven members probably have some more "colder type" contributions to the Bias, though not enough to obtain a non-Biased forecast. Another interesting feature is the fact that members with physics perturbation *p4* (*pat_len*=10000) have the largest errors in the minimum temperatures (chain-dashed line with squares). The different behaviour of the GME-driven members is indeed present in both regions, as appears by considering the similar analysis carried out over Greece (Fig. 15). Nevertheless, conclusions about the Bias are different over the two areas: in Italy all the forecasts generally overestimate temperature, while in Greece all members underestimate temperature. This is an effect of the different climate conditions between the two regions. Looking at the error (mae or rmse) it seems that GME-driven members have better performance over both regions. It is worth pointing out that only the initial and boundary conditions provided to the GME-driven members are

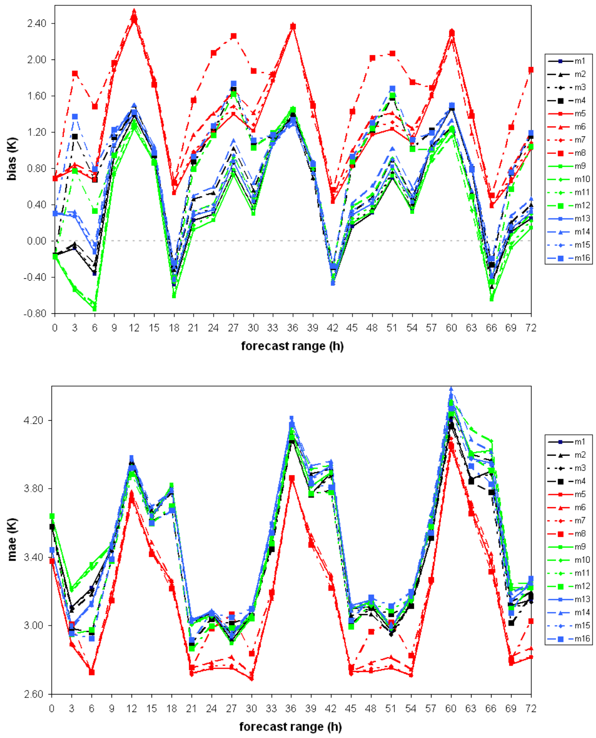


Figure 14: 2m temperature Bias (upper panel) and MAE (mean-absolute-error, lower panel) for the 16 COSMO-SREPS runs, computed over an alpine area (synop alpine dataset) for summer 2007.

characterized 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 identify the effect of the different months on temperature errors, both Bias and RMSE of the ensemble mean are calculated for each month (Fig. 16). It is clear that during summer months, especially June and July 2007, temperature is extremely underestimated. In fact, these two months were exceptionally warm with strong heat waves (the maximum temperature in Athens reached 46.2 °C in June). In autumn (here December is included as an autumn month), RMSE values remain close to 2 °C which is a statistically acceptable error value. It is worth underlining that the daily cycle of the error is inverse between summer and autumn months having its maxima during midday in the summer (also in September) and during night in autumn/winter. This is usual outcome for limited area models which is driven by soil parameterizations affecting the fluxes between soil and atmosphere. In Fig. 17, results in terms of 2m temperature for autumn 2007 are shown for northern and central Italy. During autumn, all the members exhibit a negative Bias, while in summer it was positive, but the error magnitude is almost the same (3-4 K). The 16 forecasts are quite mixed up, and the GME-driven members are included among the others. The performance of the IFS-driven members is slightly worse than the others, especially at the beginning of the forecast range. Members with physics perturbation *p4* are still characterized by a slightly different Bias, being generally warmer. Considering 24h precipitation forecast, the ensemble scores have been computed for each of the 4 groups of 4 elements which have been described at the

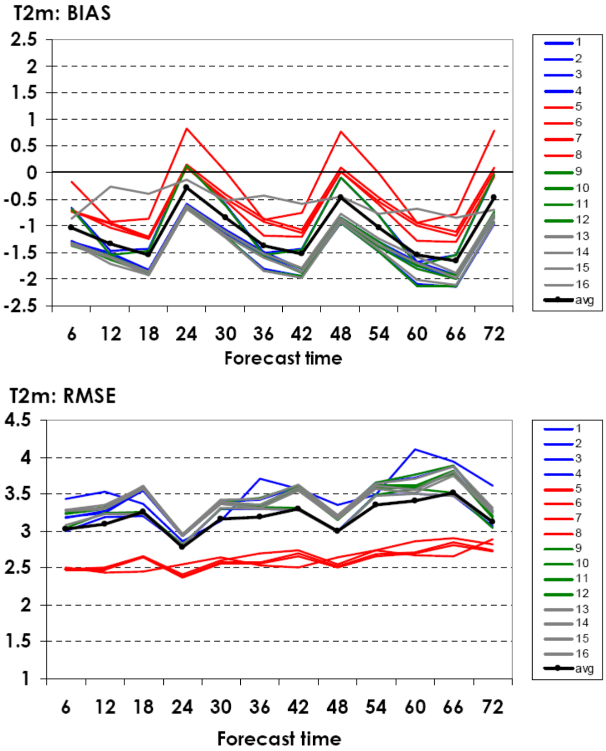


Figure 15: 2m temperature Bias (upper panel) and RMSE (lower panel) for the 16 COSMO-SREPS runs, computed over Greece for the whole period (June to December 2007).

beginning of section 3.2, 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 Fig. 18 for the summer 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.

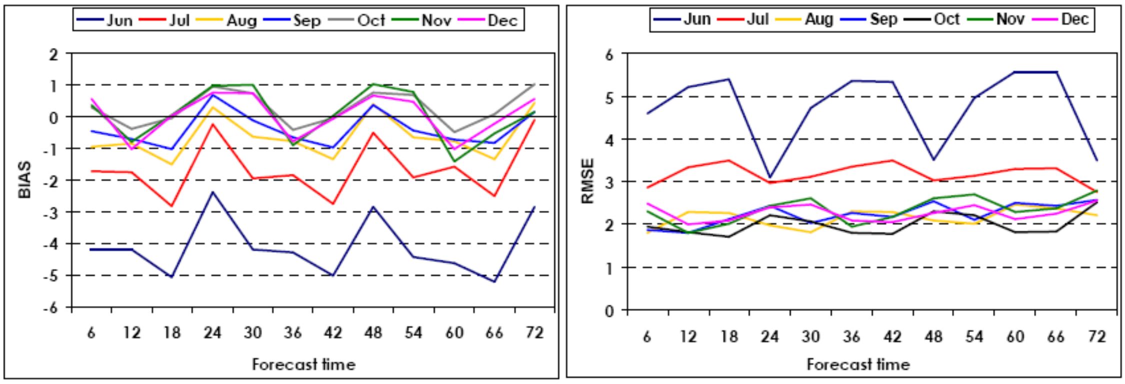


Figure 16: 2m temperature Bias (left panel) and RMSE (right panel) for the COSMO-SREPS ensemble mean, computed over Greece for the different months.

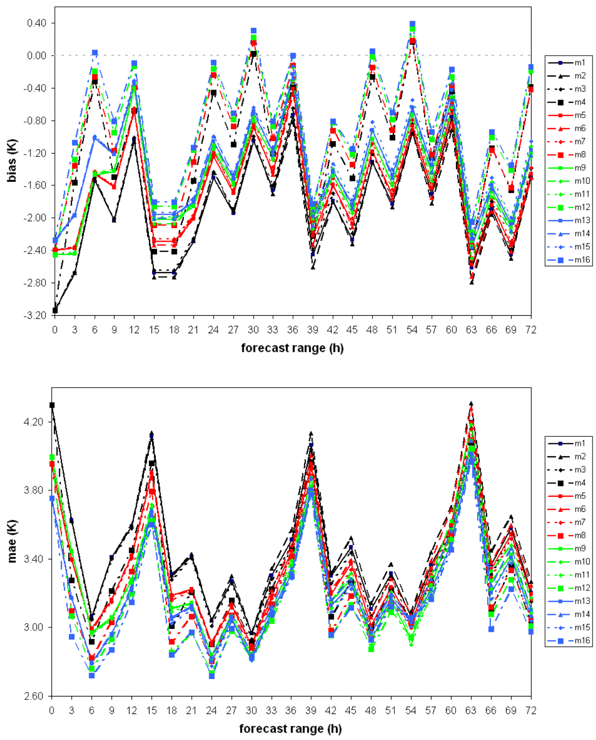


Figure 17: 2m temperature Bias (upper panel) and MAE (mean-absolute-error, lower panel) for the 16 COSMO-SREPS runs, computed over the alpine area (synop alpine dataset) for autumn 2007.

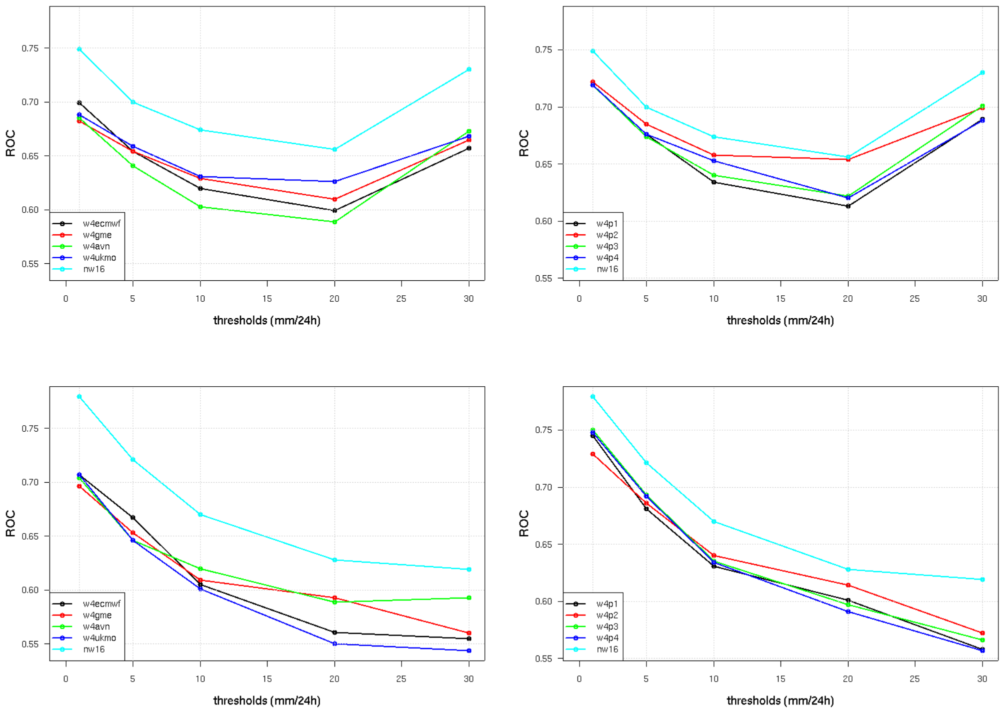


Figure 18: ROC area as a function of threshold for 24hr accumulated precipitation in the alpine area (high-res alpine data set) for the summer season. Left panels: full 16-member COSMO-SREPS (light blue line) vs. 4-member ensembles with identical 'mother run' (black: ECMWF, red: GME, green: GFS, blue: UKMO). Right panels: full 16-member COSMO-SREPS (light blue) vs. 4-member ensembles with identical physical perturbation (black: p1, red: p2, green: p3, blue p4). The upper panels are for the +30 h forecast range, while the lower panels are for the +54 h forecast range.

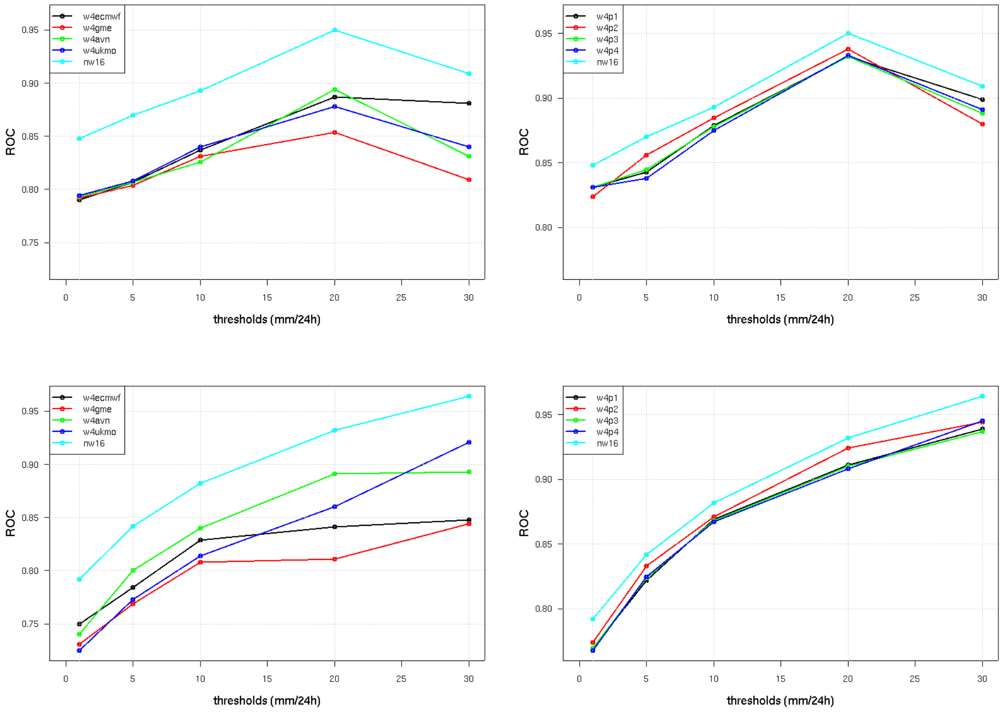


Figure 19: same of Fig. 18 but for the autumn season.

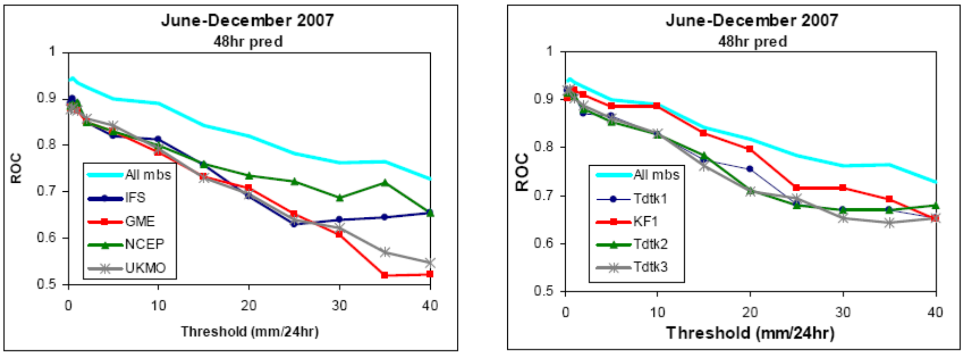


Figure 20: ROC area as a function of threshold for 24hr accumulated precipitation over Greece (synop Greece data set) for the whole period (June to December 2007). Left panel: full 16-member COSMO-SREPS (light blue line) vs. 4-member ensembles with identical 'mother run' (blue: ECMWF, red: GME, green: 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.

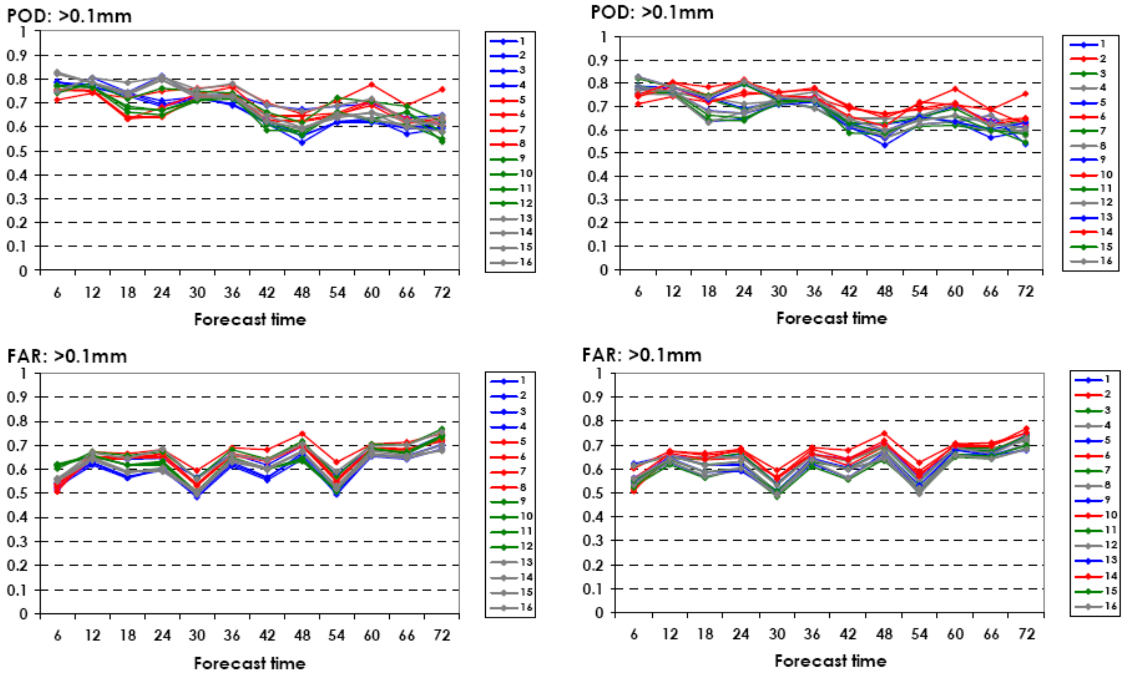


Figure 21: Probability of Detection (upper panels) and False Alarm Rate (lower panels) as a function of threshold for 6h accumulated precipitation over Greece (synop Greece data set) for the whole period (June to December 2007). In the left panels each line is coloured according to the driving-model of the member, while in the right panels each line is coloured according to the parameter set-up.

In the left panels, the other lines show the ROC area values of the 4-member ensembles made up by the 4 members nested on one particular global model. The 4 members are differentiated only in terms of the set-up of the physical parameterizations. Therefore, these represent the skill of ensembles, which are only model-perturbed, having the same initial and boundary conditions. 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 panels, the 4-member ensembles made up from identical physics perturbations are shown. Therefore, these represent the skill of ensembles, which are perturbed in the initial and boundary conditions only, but have the same model set-up. The comparison of each right panel with the corresponding left one suggests that perturbation of initial conditions generally yields more skillful 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 skillful that other 4-member ensembles. Results for the autumn season are shown in Fig. 19. It appears that scores for the autumn season are quite higher than those relative to the summer season, but the general features are the same. Again driving-model diversity gives rise to a more skillful forecast with respect to parameter diversity and the members of the Kain-Fritsch group tend to have more skill in terms of ROC area. The same evaluation has been performed also for Greece. Scores over Greece are shown in Fig. 20, for the whole period and at a +48h forecast range. It is evident that larger skill resulted by varying the driving model only (lines in the right panel), compared to varying the physical parametrization only (lines in the left panel). The effect of the different initial conditions is not evident, while Kain-Fritsch parameterization provides the best score as the higher ROC area values suggest. From these results it is difficult to judge which driving-model leads to more skillful forecast, the results being dependent on the geographical area, on the season, on the forecast range 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). This is due to the fact that they always tend to produce slightly too much rain, as it is shown in Fig. 21, where Probability of Detection and False Alarm Rate for the 16 members over Greece is shown, in terms of 6-hour accumulated precipitation. Kain-Fritsch members have higher POD but also higher FAR. This result is confirmed over the alpine area (not shown).

3.4 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 identified a lack of spread 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 involve also physical packages such as cloud and land schemes which had not been considered before (see Tab. 3). Initial and boundary conditions for the 16 runs were provided by the same run: the operational deterministic integration of ECMWF. The runs started daily at 00 UTC and the forecast range was 24 hours only, not 72 as in the COSMO-SREPS suite. This was done to save computer time, restricting the analysis to the shorter forecast range, where a good representation of the "error of the day" was needed also for data assimilation purposes.

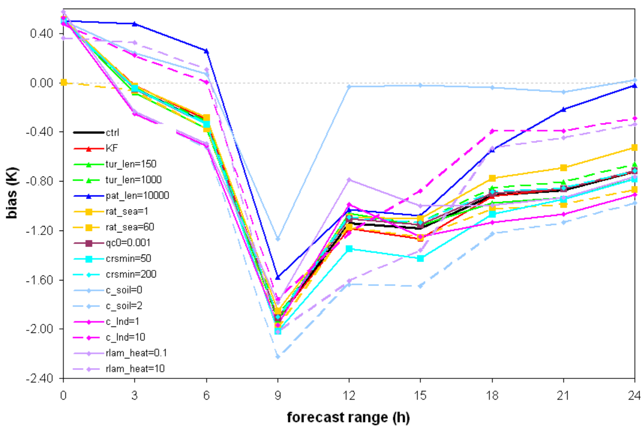


Figure 22: Mean Error (or BIAS) relative to the 16 CSPERT members in terms of 2m temperature, compared against observations over northern Italy for Autumn 2007.

In Fig. 22 and Fig. 23, an evaluation of the 16 runs is shown in terms of 2m temperature. The forecasts issued by the 16 members of the CSPERT suite were compared with non-GTS observations covering northern Italy (high-res Italy data-set, about 400 stations) for the

Table 3: Summary of the 16 perturbations applied in the CSPERT suite.

run nr.	**parameter name**	**parameter description**	**range**	**default**	**used**
1	**ctrl**				ope
2	**lconv**	convection scheme	T or KF	T	KF
3	**tur_len**	maximal turbulence length scale	[100,1000] m	500	150
4	**tur_len**	maximal turbulence length scale	[100,1000] m	500	1000
5	**pat_len**	length scale of thermal surface patterns	[0,10000] m	500	10000
6	**rat_sea**	ratio of laminar scaling factors for heat over sea	[1,100]	20	1
7	**rat_sea**	ratio of laminar scaling factors for heat over sea	[1,100]	20	60
8	**qc0**	cloud water threshold for autoconversion	[0,0.001]	0	0.001
9	**crsmin**	minimal stomatal resistance	[50,200] s/m	150	50
10	**crsmin**	minimal stomatal resistance	[50,200] s/m	150	200
11	**c_soil**	surface area index of the evaporating soil	[0,c_lnd]	1	0
12	**c_soil**	surface area index of the evaporating soil	[0,c_lnd]	1	2
13	**c_lnd**	surface area density of the roughness elements over land	[1,10]	2	1
14	**c_lnd**	surface area density of the roughness elements over land	[1,10]	2	10
15	**rlam_heat**	scaling factor of the laminar layer depth	[0.1,10]	1	0.1
16	**rlam_heat**	scaling factor of the laminar layer depth	[0.1,10]	1	10

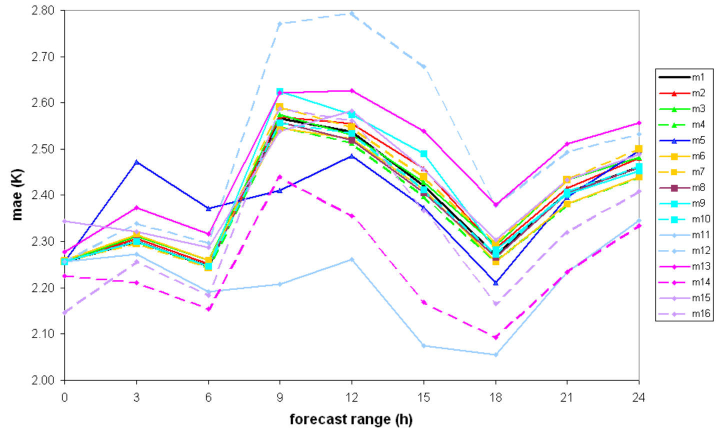


Figure 23: Mean Absolute Error (MAE) relative to the 16 CSPERT members in terms of 2m temperature, compared against observations over northern Italy for Autumn 2007.

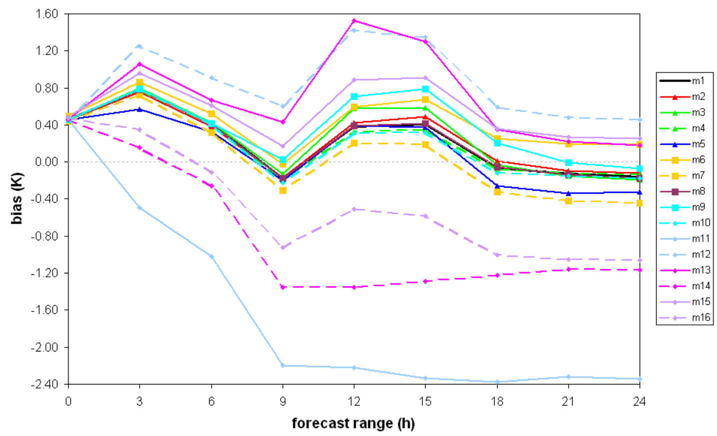


Figure 24: Mean Error (or BIAS) relative to the 16 CSPERT members in terms of 2m dew-point temperature, compared against observations over northern Italy for Autumn 2007.

whole Autumn 2007 period. This analysis shows that the perturbation of some parameters does not produce a detectable impact (at least in terms of 2m temperature), measured by the difference in score values between the run with that particular parameter set-up and the control run (thick black line). This is the case for the *qc0* and *tur_len* parameters. On the other hand, the perturbation of some parameters has a marked impact and could induce a significant increase of the ensemble spread. This is the case for the *c_soil*, *c_lnd* and *pat_len* parameters. It has to be underlined that if a particular parameter change produces a model integration which results to be statistically more or less skillful than the control run, this could be considered as a change to be implemented into all the runs to improve model performance and should not be used to produce ensemble perturbations. Of course we must be careful in this evaluation, the positive or negative impact should involve all the interesting meteorological variables and not just one. Furthermore, it is advisable to test the impact also on other seasons, characterized by different soil conditions and climatology. The good result obtained for 2m temperature by setting *c_soil*=0 should be regarded as suspicious, since this setting implies the suppression of any evaporation from

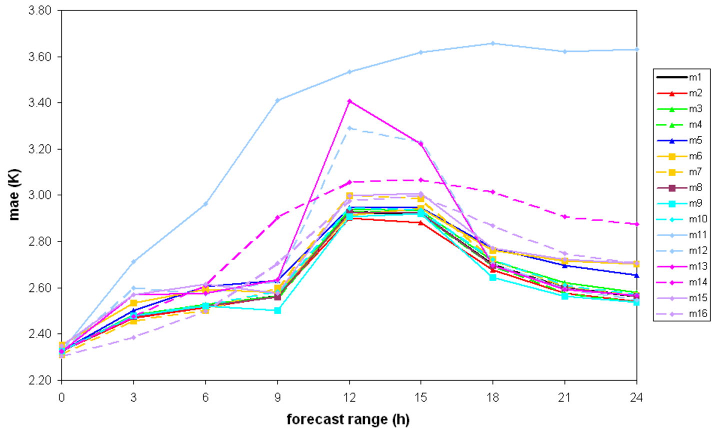


Figure 25: Mean Absolute Error (or MAE) relative to the 16 CSPERT members in terms of 2m dew-point temperature, compared against observations over northern Italy for Autumn 2007.

the bare soil. This is also evident by considering verification for other variables. Results of the evaluation in terms of 2m dew-point temperature are shown in Fig. 24 and Fig. 25. As it was expected, the run with *c_soil*=0 has the largest negative Bias and the largest error. The major impact is given by the runs with perturbation of the parameters *c_soil*, *c_lnd* and *rlam_heat*. Verification has been made also in terms of 6h precipitation forecasts (Fig. 26). The parameters producing the greater impact are *pat_len*, *c_soil*, *rlam_heat*, *rat_sea*. The major impact for the highest precipitation threshold is obtained by changing the scheme used for convection parametrization, namely using Kain-Fritsch instead of Tiedtke (perturbation *p2*, red line), as expected. As already underlined, the use of this scheme produces a positive Bias in precipitation, causing too many false alarms. The impact of the different set-up of the 16 runs on the selected meteorological variables is summarized in Tab. 4. 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
- dark 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 Tab. 4 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 were 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 nor worse) should

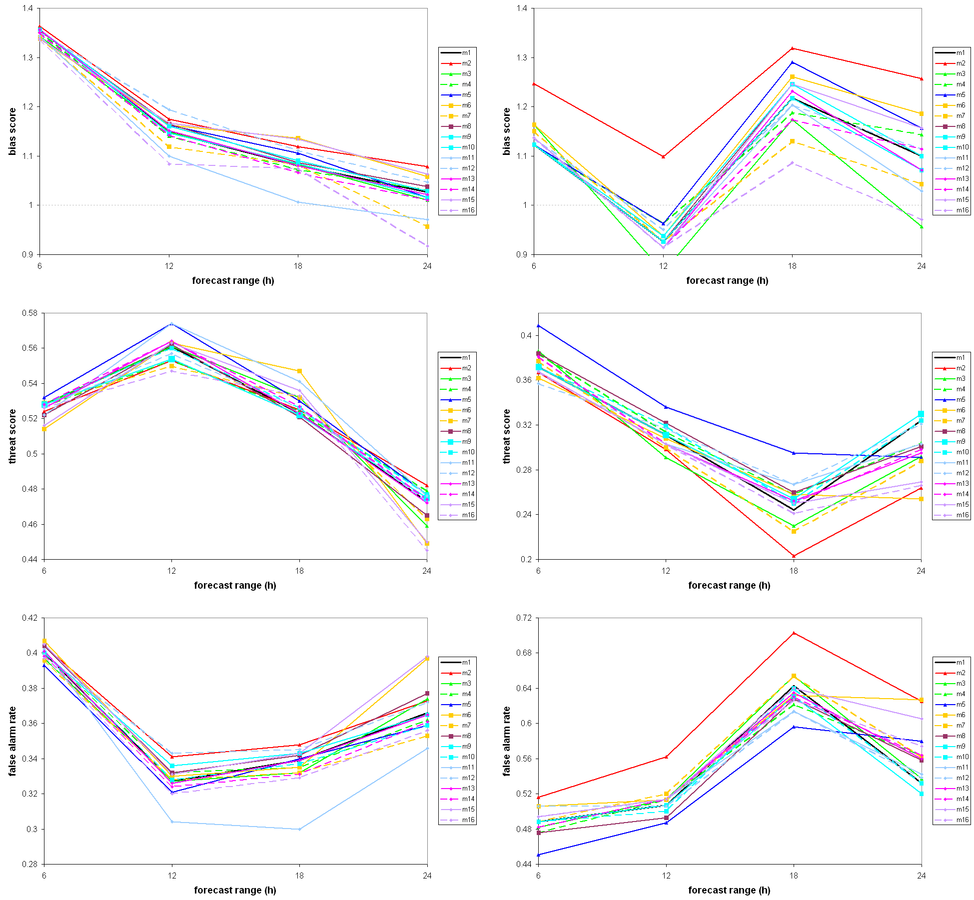


Figure 26: Bias Score (upper panels), Threat Score (middle panels) and False Alarm Rate (lower panels) relative to the 16 CSPERT members in terms of 2m 6h accumulated precipitation, for the 1mm/6h (left panels) and 10 mm/6h (right panels) thresholds.

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). The analysis of these impacts is currently under study and will be completed during the CONSENS Priority Project. In particular, the CSPERT suite will be run for at least another season, probably summer, in order to assess the impact of the changes of the parameters in a quite different climatological situation. The outcome of the analysis will lead to the definition of a set of parameter perturbations to be applied in the future COSMO-SREPS suite.

Table 4: Summary of the performances of the 15 perturbed runs with respect to the control.

	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										

4 Open points

- The system is under-dispersive (from the spread/skill relationship), hence there is a need to add more model perturbations
- It is difficult to assess in a statistically robust manner which parameters should be perturbed, since a large sample is needed. In fact, some values of the parameters can improve the forecast only for a certain season or for a certain meteorological variable. Furthermore, it is difficult to understand which choices of the parameter values can be made simultaneously in a run, unless some combinations are tried for a testing period
- Problems in running the system with continuity due to the difficulty of getting initial and boundary conditions every day
- The issue of perturbing the lower boundary was not addressed for reasons of lack of time and human resources
- The issue of perturbing the initial conditions at the mesoscale was not addressed for reasons of lack of time and human resources and it is abandoned for the moment

5 Resources used

- The computational resources needed to run the system are the Billing Units at ECMWF. These were provided by:
  - COSMO members (Germany, Italy (not for year 2008), Switzerland, Greece)
  - ECMWF through Special Projects: SPITLAEF and SPITFEAR

  The run of the system was computationally quite demanding, especially to carry on the testing of the new parameter perturbations.
- The verification task required quite an amount of resources in terms of FTEs, since it is a demanding task to obtain a statistically robust evaluation of the system performances. This had to be done for different seasons, different geographical regions and for a number of meteorological variables.

6 Lessons learned

Some Project tasks were not addressed due to lack of time and/or human resources: this was due mainly to a lack of experience in evaluating how much work was needed for a particular task, but also to the fact that, during the project, it was judged that it was more important to continue and further develop a certain line of research (not planned) instead of opening a new one. This is partly inherent in scientific research, since according to the results new questions can arise. Furthermore, since the work for the Priority Project is usually not the only one a scientist has to do, but it is done in addition to other operational and/or research work, new priorities can arise for her/him, which are not COSMO priorities but National or Regional Service ones. Finally, it happened that a task was temporally extended or shifted, for the same reasons already underlined, without undermining the project outcome. This should be considered in the project planning, allowing some flexibility in the time extension of the tasks.

7 Acknowledgements

List of contributing scientists

- Flora Gofa (HNMS)
- Petroula Louka (HNMS)
- Chiara Marsigli (ARPA-SIMC)
- Andrea Montani (ARPA-SIMC)
- Antonella Morgillo (ARPA-SIMC)
- Tiziana Paccagnella (ARPA-SIMC)

The cooperation with the Greek colleagues, which has been accounted for only as regards the verification task, was instead taking place also in the selection of the parameters to be perturbed and in the analysis of the impact of the new physics perturbations.

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.

COSMO Technical Reports

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

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

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

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

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

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

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

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

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

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