First COSMO-E Experiments with the Stochastically Perturbed Parametrization Tendencies (SPPT) Scheme

Daliah Maurer, André Walser and Marco Arpagaus

MeteoSwiss, Krähbühlstrasse 58, 8044 Zürich, Switzerland

1 Introduction

COSMO-E is an experimental ensemble prediction system with 21 members for the Alpine area, see Fig. 1. It has a convection-permitting mesh-size of 2.2 km and the forecast range is 120 hours. Lateral boundary conditions (LBC) are taken from IFS-ENS ensemble (30 km mesh size) of the European Center for Medium-Range Weather Forecasts (ECMWF), while the initial conditions (IC) for the ensemble members are taken from the operational COSMO-2 analysis. Uncertainty of the driving model IFS is represented by perturbing LBCs, i.e. using the control and the first 20 members from IFS-ENS interpolated onto the COSMO-E grid. Uncertainty in the assimilation of observations into COSMO-E will be represented later by IC perturbations, provided by an ensemble data assimilation cycle based on a Local Ensemble Transform Kalman Filter (LETKF). This study focuses on sampling the uncertainty of the COSMO-E model, in particular of the physical parametrization schemes, by using the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. In this work, no IC perturbations are applied and COSMO version 4.26 with single precision is used.

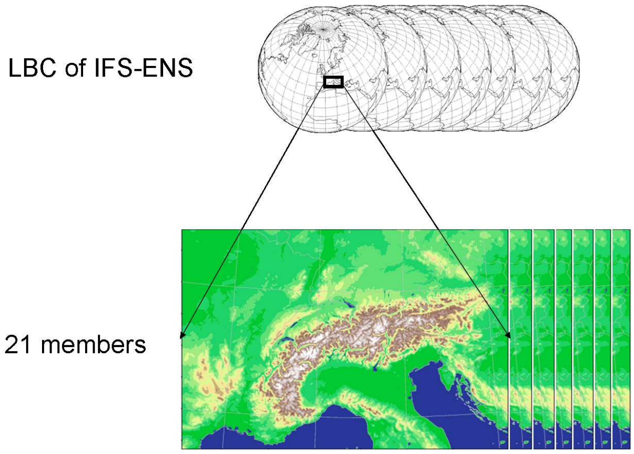

Figure 1: Experimental setup of COSMO-E

2 COSMO Implementation of SPPT

The SPPT scheme has been recently implemented into COSMO by Lucio Torrisi (based on Buizza et al. 1999; Palmer et al. 2009). The code generates a random pattern *ppert*. This random pattern is used to perturb the tendencies of the physical parametrization schemes of the model P_i^X. The total tendency of the prognostic variabe X after one timestep is

$$T^X = \frac{\partial X}{\partial t} = D^X + K^X + (1 + ppert) \sum_{i=1}^{N} P_i^X \ . \tag{1}$$

D^X is the tendency of the dynamics, K^X is the tendency of the horizontal diffusion and P_i^X is the tendency of the i-th physical parametrization scheme. For this work, only the variables zonal (U) and meridional (V) wind components, temperature (T) and specific humidity (QV) are perturbed, and N is the number of physical parametrization schemes producing tendencies for a given variable according to Tab. 1.

During the procedure of generating the random pattern *ppert*, first Gaussian random numbers with standard deviation σ and limited to a given *range* are generated on a coarse grid (Δi,Δj) at lead-time t. These random

T	Radiation,	SSO,	Turbulence,	Shallow convection,	Hydci_pp_gr (microphysics)
QV			Turbulence,	Shallow convection,	Hydci_pp_gr (microphysics)
U		SSO,	Turbulence		
V		SSO,	Turbulence		

Table 1: List of physical parametrization schemes that are active in COSMO-E and produce tendencies for temperature, specific humidity and the horizontal wind components.

numbers are interpolated horizontally in space onto the grid of COSMO-E (fine grid), see Fig. 2. At lead-time $t + \Delta t$ new random numbers are generated in the same manner. The random pattern *ppert* is an interpolation in time of two such sets of random numbers on the fine grid. Vertically *ppert* is constant, except above the tropopause and near surface, where a vertical tapering is introduced by default, see Fig. 3. Moreover, the default SPPT tuning parameters are: no supersaturation check in microphysics, $\sigma = 0.25$, $range = 0.75$, $\Delta i = \Delta j = 5.0°$ and $\Delta t = 6$ hours. In the following, only deviations from the default parameters are mentioned.

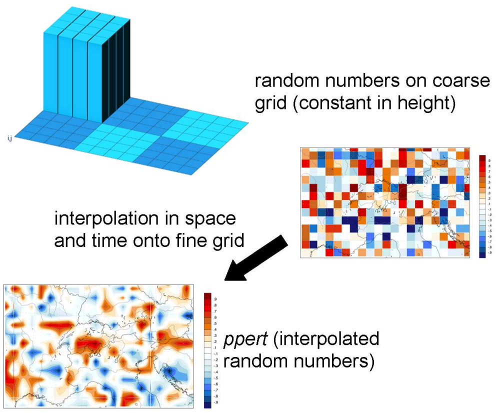

Figure 2: Procedure of generating *ppert*

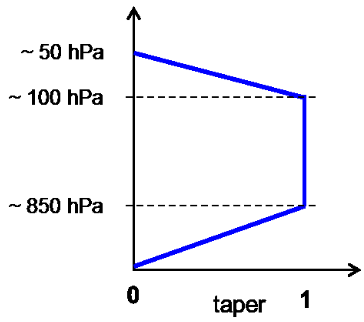

Figure 3: Taper function if the random numbers are tapered above the tropopause and near surface

3 Sensitivity of SPPT Settings

This section investigates the sensitivity of the SPPT tuning parameters on the variation between the ensemble members based on case studies. For the initial time 01.08.2012, 00 UTC, three ensemble forecasts are performed. They differ in the applied amplitude, space and time scales of the random numbers as documented in Tab. 2. The forecast range is 5 days and the same initial and lateral boundary conditions are used for all 21 members, i.e. SPPT is the only perturbation source in these experiments.

small	$\sigma = 0.25$,	$range = 0.625$,	$\Delta i = \Delta j = 0.5°$,	$\Delta t = 0.5$ hours
medium	$\sigma = 0.25$,	$range = 0.625$,	$\Delta i = \Delta j = 5.0°$,	$\Delta t = 6$ hours
large	$\sigma = 0.5$,	$range = 1.0$,	$\Delta i = \Delta j = 5.0°$,	$\Delta t = 6$ hours

Table 2: Three experimental setups in order to check the sensitivity of SPPT

The variation between the ensemble members, also called spread, is measured by the unbiased sample variance of the ensemble members. As it can be seen in Fig. 4, the setup with largest random numbers "large" produces largest spread as expected, while the setup with smaller scales in space and time reveals a significant impact as well, exhibited by a clearly smaller spread in "small" compared to "medium". More specifically, the spread is investigated at three pressure levels 850, 700 and 500 hPa. The setups generally show the largest spread on the lowest level and the smallest on the highest level, even though tapering is switched on below 850 hPa in these experiments, according to Fig. 3. This behaviour is less evident for setup "small". The identical lateral boundary conditions for all ensemble members limit the error growth, i.e the spread of the ensembles. Figure 4 shows a spread saturation at about the same forecast step +40 hours for all setups and at all levels. This rather surprising behavior is confirmed by a second case study for 02.11.2012 (not shown). However, these ensembles show a saturation already at forecast step +15 hours, while the amplitudes of the spread are about

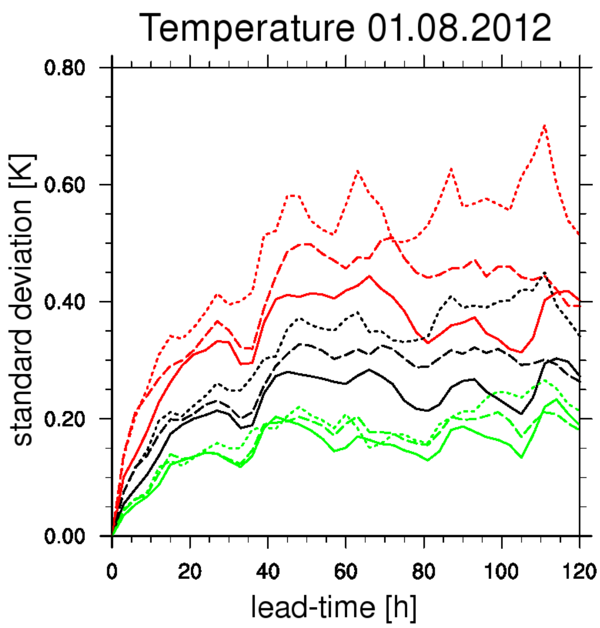

Figure 4: Ensemble spread for temperature averaged over the COSMO-E domain, at 500 hPa (solid line), 700 hPa (dashed line), 850 hPa (dotted line) and for experiment large (red), medium (black), small (green). Forecast initial date is 01.08.2012 00 UTC.

the same as in the present case study.

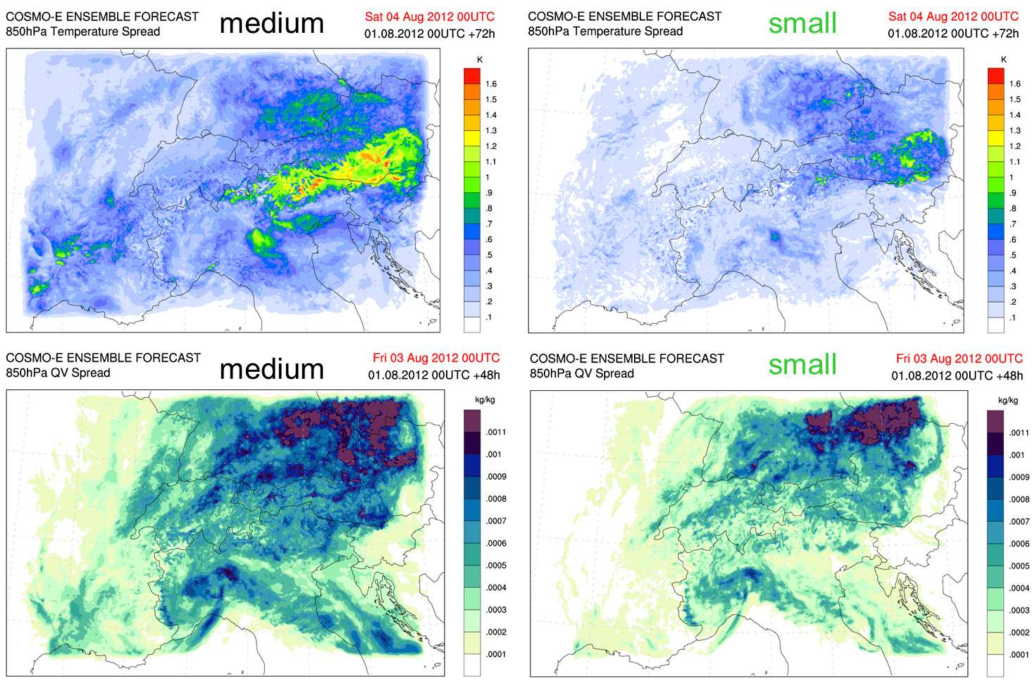

Figure 5: Ensemble spread at 850 hPa for temperature at lead-time 72 hours (top) and for specific humidity at lead-time 48 hours (bottom). Forecast initial date is 01.08.2012 00 UTC.

Figure 5 shows the ensemble spread horizontally at 850 hPa for setup "medium" (left) and setup "small" (right) for temperature (top) and specific humidity (bottom). "Medium" has a coarser structure in the random pattern than "small". This results in a larger amplitude of spread for "medium" but does hardly induce new regions of spread compared with "small".

4 Deterministic Verification

A scheme that samples model uncertainty should not degrade the quality of the individual members of an ensemble forecast. In order to quantify a potential degradation using SPPT, deterministic experiments for different SPPT setups and a reference experiment are performed for a 4 weeks period, both for summer and winter. The relevant validated SPPT setups are listed in Tab. 3.

ex0	no SPPT
ex3	SPPT default settings with $\sigma = 0.5$, $range = 1.0$
ex6	SPPT default settings with $\sigma = 0.5$, $range = 1.0$, $\Delta i = \Delta j = 0.5°$, $\Delta t = 1$ hour, without tapering in the lower troposphere

Table 3: The relevant experimental SPPT setups validated by a deterministic verification.

Observations of approximately 500 stations are used for a standard surface verification by deterministic scores like bias, mean absolute error, standard deviation and frequency bias for common thresholds. Scores of the different setups are compared for surface pressure, 2 m temperature and dew point temperature, 10 m wind speed and direction, 10 m wind gusts, cloud cover and 12 hourly precipitation sums.

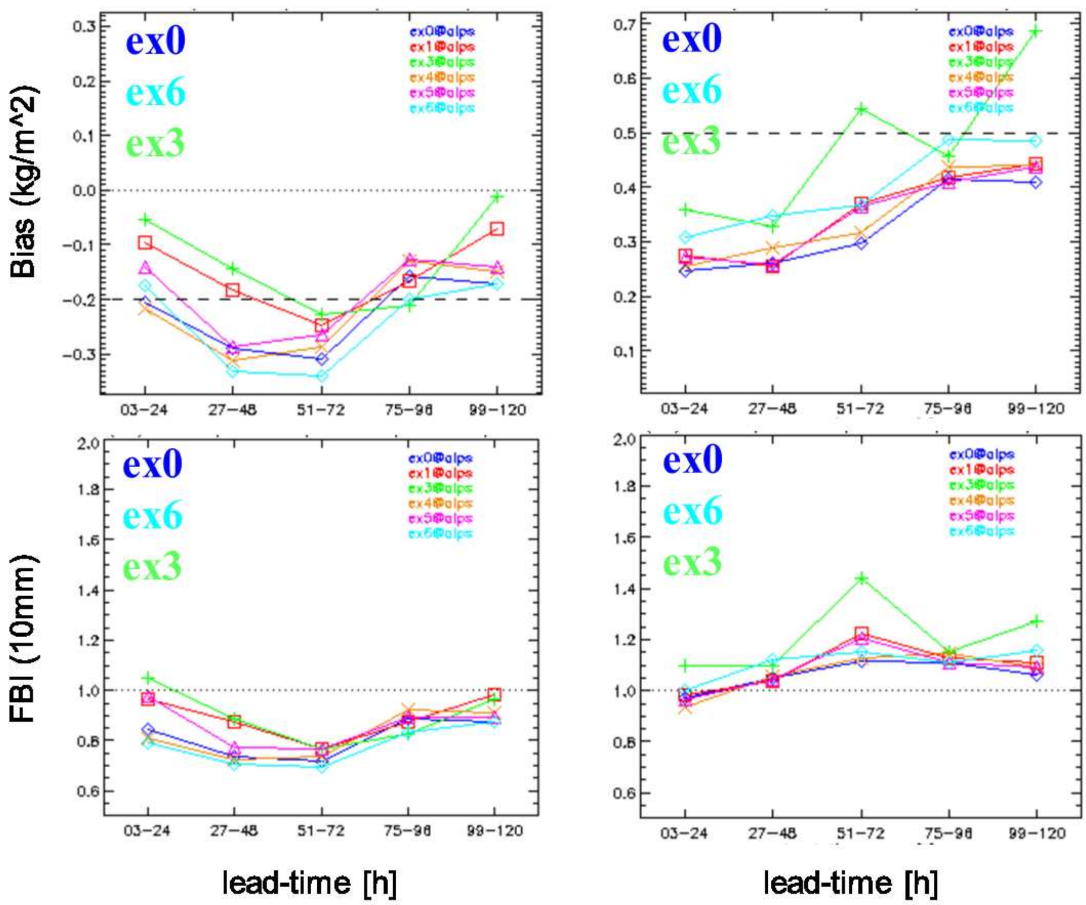

Figure 6: Bias (top) and 10 mm frequency bias (bottom) of 12 hourly precipitation sum for one month in summer (left) and winter (right)

Overall, the scores for the different setups are very similar pointing out that almost no quality degradations are found with SPPT. However, there are a few exceptions: SPPT often induces more precipitation than the reference, see bias in Fig. 6. While this has a positive impact in summer due to a generally negative summer bias of the model, it has a negative impact for the winter period. In particular the scores in winter of setup ex3 lie above the scores of the other setups, see right-hand side of Fig. 6.

In order to measure the quality of the deterministic experiments in the free atmosphere (upper air), these are verified against radio-soundings in terms of bias and the square of standard deviation error (STD) for temperature, specific humidity, geopotential wind speed and direction. Only small differences between the different setups are observed for both the summer and winter period. Figure 7 shows the scores for temperature at 25 pressure levels for all 12 soundings available in the model domain. Both, bias and STD are very similar for all experiments, except ex3 which shows slightly smaller bias but slightly larger STD.

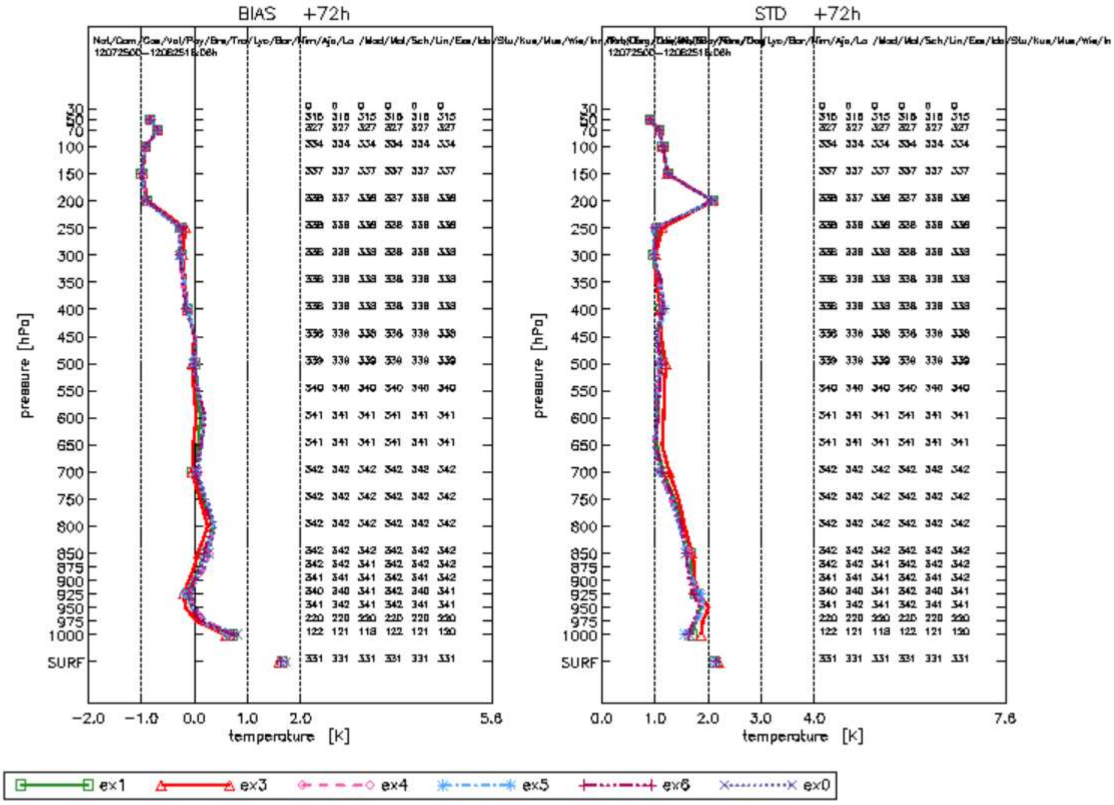

Figure 7: Bias (left) and the square of standard deviation error (right) of the upper air verification of temperature for one month in summer, using all stations in the COSMO-E domain.

In conclusion, surface and upper air verification indicate, unless random numbers as well as space or time scales are increased as much as in ex3, SPPT is hardly detrimental to forecast quality.

5 Ensemble Experiments

This section investigates 00 UTC ensemble forecasts performed for an extended period from 25.07.2012 to 25.08.2012 and 3 experimental setups:

LBC	**L**ateral **B**oundary **C**ondition perturbations alone
LBC+SPPT	SPPT perturbations and LBC perturbations
LBC+PP	**P**arameter **P**erturbations used for COSMO-DE-EPS (Gebhardt et al. 2011) and LBC perturbations

The SPPT setup ex6 is chosen for the LBC+SPPT experiment, because the scale of those perturbations matches the scale of convective events and the scores of the deterministic verification lie in an acceptable range, as pointed out in section 4.

One member of the LBC+SPPT experiments with initial date 24.08.2012 crashed because the CFL-criterion was violated. Therefore, the experiments for this initial date are excluded from the subsequent analyses.

The ensemble forecasts from the three setups have been verified in terms of 12 hourly precipitation sums. Figure 8 shows the Brier skill score against observations from 300 SYNOP stations, with climatology (2001-2010) as reference. SPPT gives a small benefit to LBC at the begin of forecast, i.e. until ~30 hours lead-time, while almost no differences can be observed between the experiments LBC and LBC+PP. Therefore, in the following we focus on the results of the LBC+SPPT and LBC experiments. Note again, that no IC perturbations are used in these experiments. The benefit of SPPT might be smaller with appropriate IC perturbations.

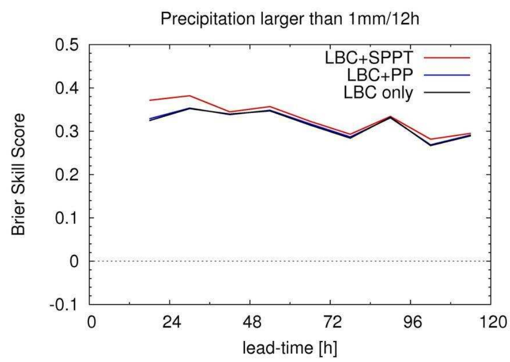

Figure 8: Brier skill score for all three experiments against observations of 300 SYNOP stations for precipitation larger than 1 mm per 12 hours, with climatology (2001-2010) as reference.

According to Jollife et al. (2011), the attribute 'reliability' of a probabilistic forecast characterizes the degree to which the forecast probabilities are consistent with the relative frequencies of the observed outcomes. For a perfectly reliable probabilistic forecasting system, the squared error between ensemble mean and observation and the unbiased sample variance of the ensemble members have the same expectation value. Thus the relation

$$\mathrm{E}_t\left[\left(x_t-\bar{\hat{x}}_t\right)^2\right]=\frac{m+1}{m}\mathrm{E}_t\left[s_t^2\right] \tag{2}$$

should hold. t characterizes the t-th ensemble forecast of m members $\{\hat{x}_{t,1}, \hat{x}_{t,2}, ..., \hat{x}_{t,m}\}$, in our case the lead-time. $\mathrm{E}_t[\cdots]$ denotes the expectation value and x_t is the observation corresponding to the ensemble forecast. $\bar{\hat{x}}_t \doteq \frac{1}{m}\sum_{j=1}^{m}\hat{x}_{t,j}$ is the ensemble mean and $s_t^2 \doteq \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{x}_{t,j}-\bar{\hat{x}}_t\right)^2$ is the unbiased sample variance of the ensemble members.
Both sides of Eq. (2) are estimated by calculating the expressions in box brackets at every gridpoint and averaging over the COSMO-E domain and over all initial dates. On the right-hand side of Eq. (2) the estimate is the squared ensemble spread, abbreviated as RMEV_t^2. The left-hand side is estimated by squares of error between analysis and ensemble mean, abbreviated as RMSE_t^2. The model variables often are biased. In that case, the forecasting system is not perfectly reliable, thus the criterion of Eq. (2) will not be fulfilled. Therefore, the left-hand side is estimated by unbiased squares of error between analysis and ensemble mean,

averaged over the COSMO-E domain and over all initial dates, abbreviated as STDE_t^2.

Figure 9 shows the spread (left) and unbiased error (right) of experiment LBC+SPPT for temperature, averaged over the COSMO-E domain and the initial dates. Note that spread and error for lead-time +0 hours are zero by definition due to the use of the verifying analysis as IC for all members. In the lower troposphere (k-levels ∼40-60) the spread is significantly smaller than the unbiased error, at least until a lead-time of ∼60 hours. Above k-level ∼40, spread and unbiased error are quite equal.

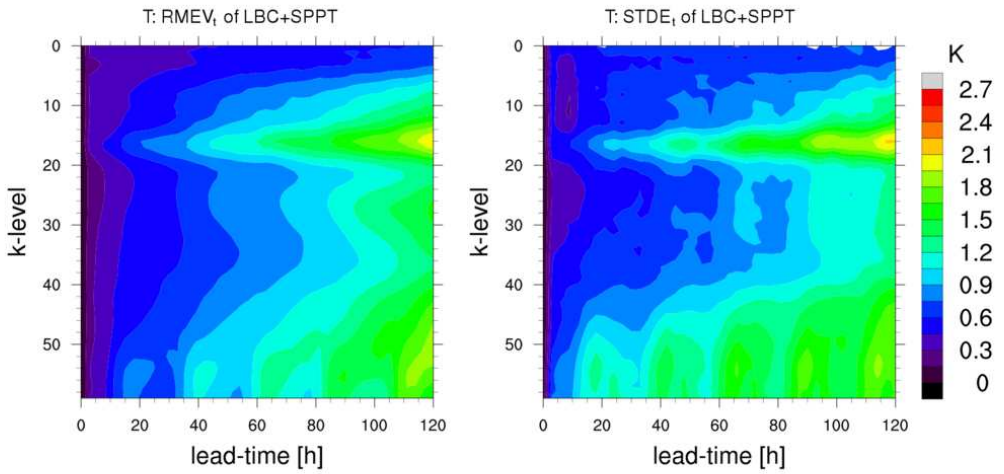

Figure 9: The root of the average over the COSMO-E domain and over the initial dates of ensemble spread (left) and unbiased squares of error between analysis and simulation (right) for temperature. The model levels build the Y-axis and the X-axis is the lead-time.

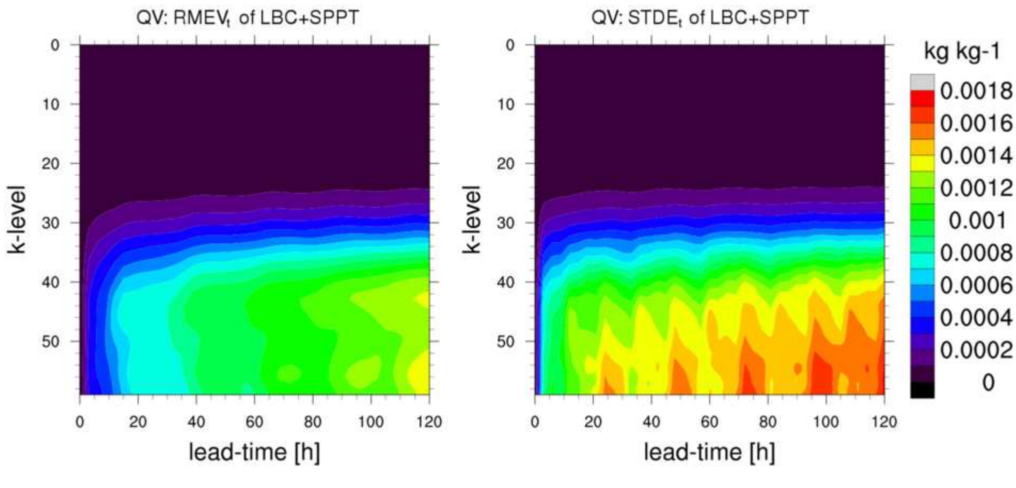

Figure 10: Same as Fig. 9 but for specific humidity

Figure 10 illustrates that for specific humidity the spread is clearly smaller than the unbiased error at k-levels ∼40 to ground, throughout the entire forecast range. For k-levels above ∼40, specific humidity becomes negligibly small and the spread and unbiased error tend to zero as a consequence.

Figure 11 shows the difference in spread amount between the experiments LBC+SPPT and LBC for temperature and specific humidity, averaged over the COSMO-E domain and the initial dates. The spread induced by the SPPT scheme is limited in height to k-levels ∼30-60. After ∼30 hours lead-time, SPPT does no longer give large additional spread to the LBC perturbations.

Investigating the temperature and specific humidity tendencies from the different physical parametrization schemes for a case study in summer (not shown), reveals a very large and horizontally extensive tendency sum in the planetary boundary layer (PBL). Above the PBL, the tendencies are weaker and less extensive. Because the activity of the SPPT primarily depends on the tendencies of the physical parametrization schemes, this may explain the limitation in height of the SPPT scheme for temperature and why the largest impact of the SPPT scheme occurs in the PBL. The limitation in height of the SPPT scheme for specific humidity is rather caused by the large reduction of specific humidity in height.

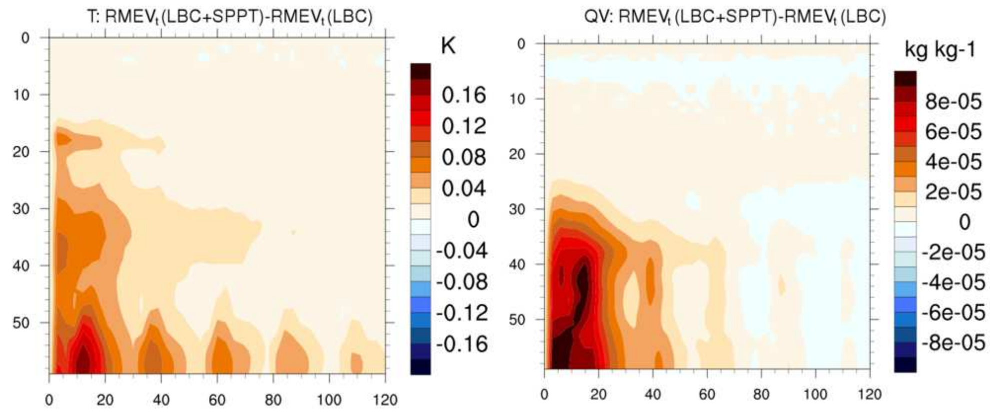

Figure 11: Difference in spread amount between the experiments LBC+SPPT and LBC for temperature (left) and specific humidity (right), averaged over the COSMO-E domain and the initial dates. The model levels build the Y-axis and the X-axis is the lead-time.

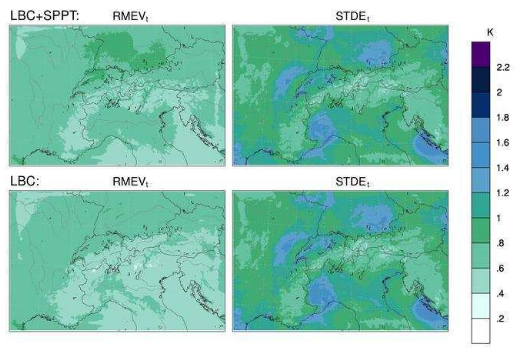

Figure 12: For temperature the root of the average over the COSMO-E domain and over k-levels 40-59 of ensemble spread (left) and unbiased squares of error between analysis and simulation (right) are plotted for the experiments LBC+SPPT (top) and LBC (bottom), at lead-time 24 hours.

The criterion Eq. (2) is examined horizontally as well. For the layer plots in Fig. 12, spread (left) and unbiased error (right) are calculated at lead-time 24 hours averaging over the initial dates and the k-levels 40-59, instead of the COSMO-E domain. For temperature additional spread induced by the SPPT-scheme can be observed in Northern Italy and in the South of Germany. Compared with the unbiased error, the spread of the experimental setup LBC+SPPT is still smaller, but the additional spread mostly occurs in regions where the unbiased error is large. The unbiased error does not change significantly, adding SPPT to LBC perturbations.

6 Caveat

The present idea of SPPT is to model the error of the physical parametrization schemes. However, for the experiments, based on COSMO version 4.26, zonal and meridional wind tendencies of the coriolis subroutine have been erroneously perturbed as well. The large windspeed in in the upper troposphere induces large coriolis force tendencies. Hence, on these levels the SPPT uncertainty terms for the wind components are mainly dominated by the coriolis force tendencies and thus are inappropiate. Therefore, the results for wind are omitted in this report. In COSMO version 5.0 the coriolis force tendencies are no longer perturbed.

7 Summary and Outlook

COSMO-E is an experimental ensemble prediction system with 21 members for the Alpine area. It has a convection-permitting mesh-size of 2.2 km and a forecast range of 120 hours.
The Stochastically Perturbed Parametrization Tendencies (SPPT) Scheme, which perturbs tendencies of the model by multiplying a random pattern, is introduced in order to model the error of the physical parametrization schemes.
Investigating how the variation between ensemble members, also called spread, depends on the parameters of SPPT, it turned out that spread is increased by the increase of random number amplitudes, as one would expect. But surprisingly also the enlargement of space and time scales of the random pattern leads to an increase in spread. A deterministic verification of the SPPT scheme reveals that the forecast quality of individual members is hardly affected, unless random numbers are large ($\sigma \geq 0.5$, cutted at 1.0), and space and time scales are increased up to 5.0° and 6 hours.
Ensemble experiments are performed using a random pattern with space and time scales comparable to the scale of convective events. The forecast time is 5 days starting at initial dates 25.07.2012 to 25.08.2012, 00 UTC. The reference experiment with lateral boundary condition (LBC) perturbations is compared with the experiment which combines LBC perturbations with SPPT. In height, additional spread of SPPT is observed up to ~7 km, for temperature and specific humidity. The Brier skill score for precipitation larger than 1 mm per 12 hours shows a benefit of the combined experiment until a lead-time of ~30 hours. Spread and unbiased error between ensemble mean and analysis of temperature and specific humidity show that below a height of ~5.5 km, the spread is smaller than the unbiased error. Above ~5.5 km, the spread is in the same range as the unbiased error for temperature. The specific humidity and as a consequence the spread and the unbiased error are negligibly small above ~5.5 km. Horizontally the additional spread mostly occurs in regions where the unbiased error is large.

In COSMO version 4.26 used for the present study, accidentally the tendencies of zonal and meridional wind of the dynamical subroutine coriolis are perturbed as well. This will be changed for subsequent studies. Moreover, the variation between ensemble members should be increased for specific humidity, and in lower troposphere for temperature, by a further tuning of the parameters of the SPPT scheme. Since the tendencies from the physical parametrization schemes are largest in the planetary boundary layer, there is potential to further increase the ensemble spread with SPPT. Finally, the uncertainty of initial conditions will be modelled by using initial condition perturbations from the ensemble based data assimilation cycle (LETKF).

References

[1] Buizza, R., Miller, M. and Palmer, T.N., 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Q. J. R. Meteorol. Soc.*, 125:**2887-2908**.

[2] Gebhardt, C., Theis, S.E., Paulat, M., Ben Bouallègue, Z., 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries, *Atmospheric Research*, 100 **168-177**.

[3] Jolliffe, I.P., Stephenson, D.B., 2011: Forecast Verification: A Practitioner's Guide in Atmospheric Science. Chapter 8 (Weigel, A.P.): *Verification of Ensemble Forecast.* Wiley.

[4] Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J., Steinheimer, M., Weissheimer, A., 2009: Stochastic Parametrization and Model Uncertainty, *Internal Report from ECMWF.*