Consortium for Small-Scale Modelling

Technical Report No. 45

The COSMO Priority Project APSU: Final Report

February 2022

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_45

Deutscher Wetterdienst
MeteoSwiss
Ufficio Generale Spazio Aereo e Meteorologia
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
Instytut Meteorologii i Gospodarki Wodnej
Administratia Nationala de Meteorologie
ROSHYDROMET
Agenzia Regionale Protezione Ambiente Piemonte
Agenzia Regionale Prevenzione Ambiente Energia Emilia Romagna
Centro Italiano Ricerche Aerospaziali
Amt für GeoInformationswesen der Bundeswehr
Israel Meteorological Service

www.cosmo-model.org

Editor: Massimo Milelli, CIMA Foundation

The COSMO Priority Project APSU: Final Report

Marsigli C.[1]*, ***Astakhova E.***[2], ***Duniec G.***[3], ***Füzer L.***[5], ***Gayfulin D.***[2], ***Gebhardt C.***[1], ***Golino R.***[4], ***Heppelmann T.***[1], ***Interewicz W.***[3], ***Marcucci F.***[4], ***Mazur A.***[3], ***Sprengel M.***[1], ***Tsyrulnikov M.***[2], ***Walser A.***[5]

* Project coordinator
[1] Deutscher Wetterdienst - DWD
[2] HydroMetCenter of Russia - RHM
[3] Instytut Meteorologii i Gospodarki Wodnej - IMGW
[4] ItAF-ReMet - COMET
[5] MeteoSwiss - MCH

Contents

1 Abstract

The APSU Priority Project focussed on the improvement of the ensemble forecast systems run by the COSMO members, developing both the perturbation methodology and the post-processing. The aim was to improve the forecast issued by the convection-permitting (CP) scale modelling system, particularly for severe and high-impact weather events. The research and development work covered three main areas:

- development of model perturbation methodologies, including new methods and the development of surface/soil perturbations;
- development of post-processing methods for the convection-permitting ensembles, especially focused on the forecast of selected severe weather phenomena;
- better use of the KENDA analyses as Initial Conditions (ICs) to initialize the CP ensembles and improvement of the perturbed Boundary Conditions (BCs).

The project provided also the occasion to discuss and plan the transition of the ensembles to the ICON-LAM model.

2 Introduction

The APSU Priority Project was initiated due to the need of improving the ensemble systems run by the members of the COSMO Consortium, in particular for the prediction of near-surface weather parameters at high spatio-temporal resolution.

During the WG7 meeting held in Offenbach in March 2017, it was recognised that the perturbations applied to the model in the ensemble configurations do not provide yet a satisfactory representation of the model error, as evidenced by the spread-skill relation for near-surface parameters (Marsigli et al., 2019). The shared opinion was that the spread of the ensembles in terms of near-surface parameters (temperature, wind) should be increased, in order to better match the forecast error. This should be done with care, since it is recognised that the ensemble spread cannot be increased up to exactly reaching the forecast error, because the latter includes also the systematic error, which should be tackled by improving the model, not by increasing the spread. On top, observation error should be considered in the computation of the forecast error (Ben Bouallegue et al., 2020). What should be better represented in the ensembles is that part of the error which depends on the uncertainties in the determination of initial and boundary conditions as well as in the description of the physical processes in the model, which are mainly due to the variability of the model performance in dependence of the meteorological situation, usually called "error of the day" or flow-dependent error (see for example Leutbecher et al., 2017). To reach this goal, it is needed to work on a better representation of the model error. The strategy proposed in APSU is based on testing new approaches for model perturbation which have been recently developed or made available to the COSMO model, and to review and improve the already existing ones. More specifically:

- further develop and test the new stochastic approach to model the model error developed at DWD
- develop an additive perturbation method based on the Stochastic Pattern Generator (Tsyrulnikov and Gayfulin, 2017)

- test of two new available methods for model perturbations: iSPPT (independent SPPT, Christensen et al., 2017) and the perturbations based on analysis increments (Piccolo et al., 2018)
- develop perturbations of the soil/surface and test the combination of the different perturbations

It was also noted that the potential of the ensembles is not fully exploited for the operational duties, in particular for high spatial and temporal resolution forecasts, therefore it was identified the need to continue working on ensemble post-processing and interpretation. In particular it is needed to:

- develop suitable probabilistic products for operational applications. In order to benefit from the link with the SRNWP-EPS Project of EUMETNET, the focus was on products for thunderstorms and fog;
- further develop suitable calibration procedures for the CP ensembles.

This activity has a strong link with the activities carried out in WG4.

On top of these development areas, the transition from the COSMO to the ICON model implies also a transition of the ensemble systems. To make this possible, a planning of the needed activities was started in the APSU PP. A first concrete task was performed at DWD, where the Parameter Perturbation methodology was transferred and tested in the new ICON-LAM ensemble.

3 New model perturbation methods

New methodologies for model perturbations have been tested in the APSU PP, and further developed.

3.1 SMME: Stochastic model of the model error (M. Sprengel, T. Heppelmann, C. Gebhardt, DWD)

3.1.1 Method

At DWD, a prognostic stochastic model of the model error ("EM-scheme", Ekaterina Machulskaya) has been developed. The concepts of the EM scheme have been the basis for the development of an improved framework called "stochastic model of the model error" (SMME). The development of SMME is carried out within the project Gridcast and its precursor EweLine which both are collaborative research projects of the DWD with the Fraunhofer IEE and the German transmission system operators (TSOs) in the context of renewable energies. The growing share of renewable energy in power generation increases the impact of the weather on the stability of the electrical power grid. Especially prior to severe weather events, not only high-quality weather forecasts but also information about forecast uncertainties is needed by the TSOs to prepare stability provisions. Improving the description of the inherent model error is therefore one of the goals of the research project Gridcast. Even if this development is focused on optimizing the use of renewable energies, it can be beneficial for NWP model development in general and therefore fits well as a contribution to PP APSU.

The EM-scheme and the SMME have in common that the errors of the tendencies of different model variables are assumed to obey the Langevin-type stochastic differential equation being correlated in space and time and stochastically driven by Gaussian noise as in equation (1).

$$\frac{\partial \eta_X(x,t)}{\partial t} = -\gamma(\tau_X)\eta_X(x,t) + \gamma(\tau_X)\nabla \cdot \left(\lambda^2(\tau_X)\nabla\eta_X(x,t)\right) + \sigma(\tau_X)\xi(x,t), \tag{1}$$

where

- $\tau_X = \tau_X(x,t)$ is the tendency of the predictor variable X,
- γ is the damping,
- diffusion λ guarantees spatial correlation,
- ξ is a standard normal random field with standard deviation σ.

The weather-dependent values of the parameters γ, λ, and σ of this equation are determined by predictors, whereas the dependencies of the parameters on the predictors are inferred from past data. The solution η_X of eq. 1 is our model of the model error and is added to the dynamic equations of the model to correct the tendencies from the NWP online for a set of variables X:

$$\frac{\Delta X}{\Delta t}(x,t) = \left[\frac{\Delta X}{\Delta t}(x,t)\right]_{\text{phys}} - \eta_X(x,t).$$

The method is applied to temperature and zonal and meridional wind as prognostic variables with the tendency τ_X of each variable X acting as predictor for the respective error model parameters.

Going from the EM-scheme to the SMME, the method for estimating the parameters has been re-developed from scratch by taking the spatial diffusion term into account from the very beginning of the development. This approach enabled us to have a consistent framework for all three parameters γ, λ, and σ that includes the spatial diffusion without the need for a separate treatment.

As first COSMO-D2-EPS experiments with the method showed an increased forecast spread but at the expense of a substantial degradation of the RMSE (see later in section 3.1.3), a new method for estimating γ has been developed based on results of (García-Ojalvo et al., 1992):

Assuming spatially constant coefficients, we can discretize the stochastic partial differential equation eq. 1 in space and perform a spatial Fourier transform to obtain a system of stochastic ordinary differential equations. For these equations certain expectation values and correlations can be derived analytically. For example for the stationary correlation function at different times t and t' and Fourier modes k and k' one finds

$$\left\langle \eta_k(t)\eta_{k'}(t')\right\rangle^{st} = \frac{\sigma^2 N_x \Delta x N_y \Delta y}{2\gamma c_k} e^{-\gamma c_k |t-t'|}\delta_{k,-k'} \tag{2}$$

where Δx and Δy denote the horizontal grid spacing, N_x and N_y are the number of grid points, and c_k is the Fourier transform of the differential operator. For the Fourier modes $k = k' = 0$ this further simplifies and we can take the quotient at different times $t' = t + \Delta t$ and $t' = t + 2\Delta t$ to obtain

$$\frac{\left\langle \eta_k(t)\eta_{k'}(t+\Delta t)\right\rangle^{st}}{\left\langle \eta_k(t)\eta_{k'}(t+2\Delta t)\right\rangle^{st}} \overset{k=k'=0}{=} e^{\gamma \Delta t}. \tag{3}$$

This method resulted in a much better RMSE (see later in section 3.1.3).

3.1.2 Parameter estimation

The estimation of the parameters γ, λ, and σ from past data requires the estimation of the model error tendencies on the left-hand side of eq. 1 which in turn requires an approximation of the truth to be compared with model forecast. For SMME, the truth is approximated by analysis fields of KENDA for COSMO-D2 and the model error is defined as the deviation of short-range forecasts of COSMO-D2 from these analyses. Other approaches exist to estimate the model error tendencies by comparing forecasts only (see for example Tsyrulnikov and Gayfulin, 2019 and Christensen, 2020).
This section exemplifies important aspects of the dependencies of the estimated parameters γ, λ, and σ on their predictors, on the forecast time and the time of the day as well as on the vertical model level. The training period for the parameter estimation is 5 months in autumn 2018 and 2019 with the exclusion of the application month October 2018. The training data are the operational COSMO-D2 forecasts started at 03 UTC up to 45 hours of forecast time and the operational analyses for the same period. Fig. 1 shows the changes of the damping parameter γ with forecast time for temperature at ca. 300m above ground.

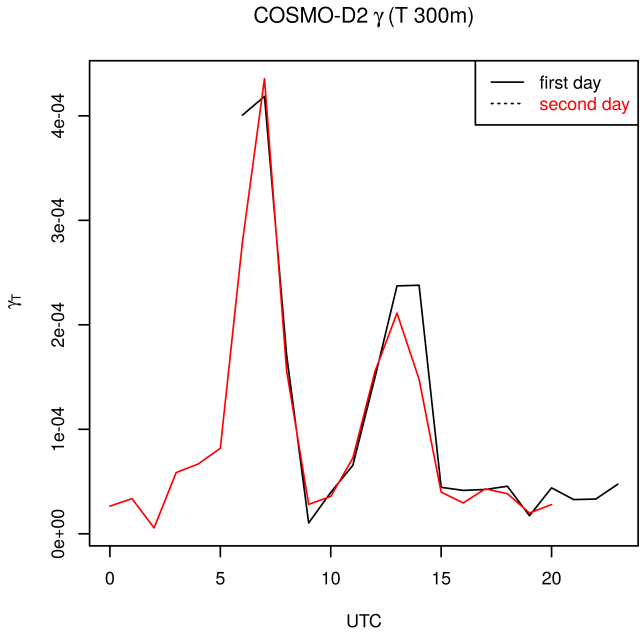

Figure 1: Estimated γ for temperature at ca. 300m as a function of the time of the day for lead times 1h to 45h (first day in black, second day in red, forecast start is 03 UTC).

There is a clear dependence on the time of the day without a significant difference between day 1 and day 2. The dependence on the time of the day decreases with increasing height of the vertical model level (not shown). The parameter γ depends on the prognostic variable X but hardly on the size of its tendency τ_X (not shown). However, there is a clear dependence of the diffusion parameter λ on the predictor tendency τ_X (Fig. 2).

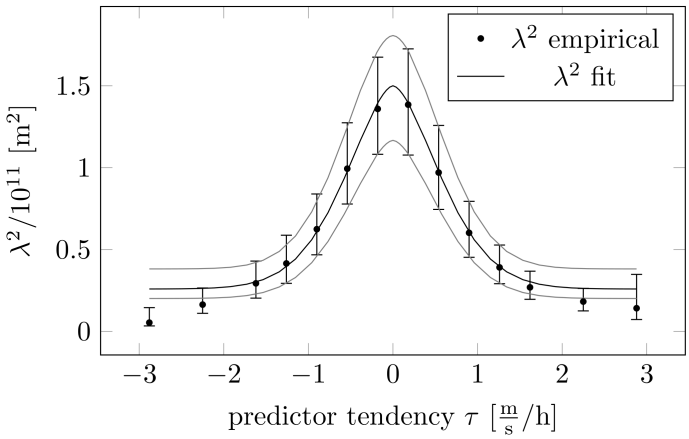

Figure 2: Estimated λ for the zonal wind at 97m above ground and fitted to a function of type $\lambda(\tau) = \lambda_a e^{-(|\tau|/\lambda_b)^{\lambda_c}} + \lambda_d$. The error bars are the 95%-confidence intervals estimated from the sample of days for training. The black line is the fit to the empirical values of λ and the grey lines fit to the limits of the confidence intervals.

The tendencies have been arranged in equally populated bins for the parameter estimation. A parametrized function can be nicely fitted to the empirical values.

Fig. 3 shows the combined effect of forecast time, time of the day and atmospheric height on λ for temperature.

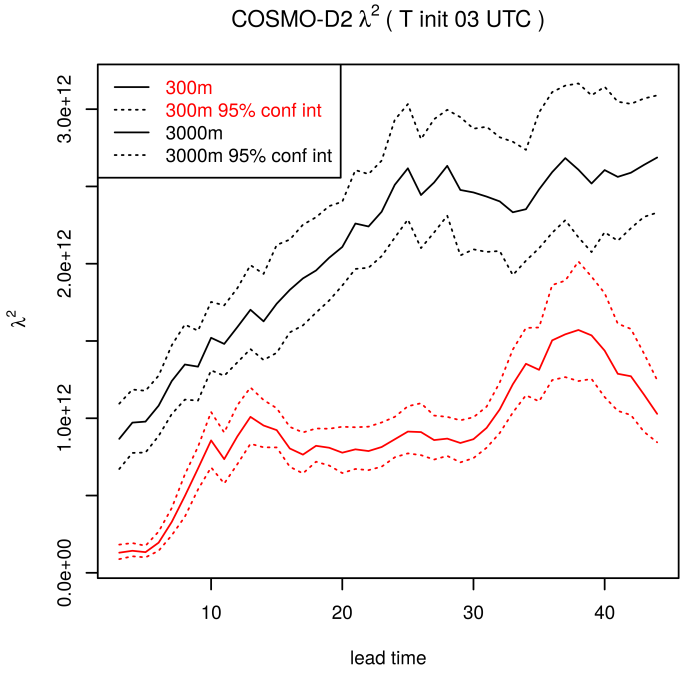

Figure 3: Estimated λ for temperature (solid line) at two atmospheric heights (300m in red, 3000m in black) as function of forecast lead time. The dashed lines are the 95%-confidence intervals estimated from the sample of days for training.

The values of λ are larger for higher altitudes and a given forecast lead time. There is a general trend of increasing λ with increasing lead time which is modulated by a daily cycle in the boundary layer. This dependence on the time of the day shows an increase of λ starting

after dawn with the onset of boundary layer mixing with a moderate decrease in the evening and a quasi-plateau during the night. In general, the characteristics of λ for temperature apply to the horizontal wind components as well.

Since the parameter σ is calculated on the basis of the estimated γ and λ it basically inherits the dependencies on time and height from those parameters. In Fig. 4, the dependence of σ on model level and predictor tendency for zonal wind is shown.

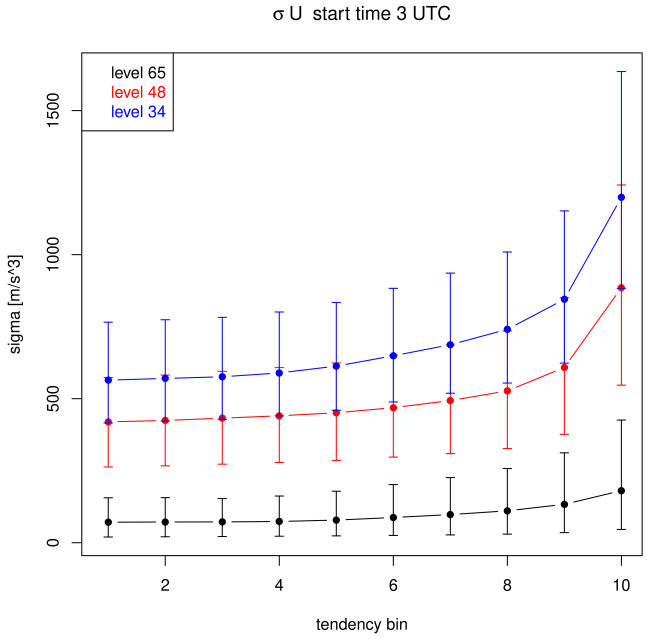

Figure 4: Estimated σ for zonal wind at different atmospheric heights (different colours) as function of tendency bin τ_U, i.e. numbers on the abscissa are bin counts. The error bars are the 95%-confidence intervals estimated from the sample of days for training. The levels 65, 48, and 34 correspond to the heights 10m, 757m, and 3000m, respectively.

There is an increase of σ with height as well as with the magnitude of τ_U, the latter being negligible to moderate for lower values of the tendency. The increase of σ with τ_U is larger higher up in the atmosphere.

In practice, the parameters γ, λ, and σ are estimated for distinct model levels only (at 10m, 757m, and 3000m for U and V; and 300m, 757m, and 3000m for T). Accordingly in the application, the model error η is calculated for these three levels and linearly interpolated vertically between those levels. Above the highest level $k_{max} = 34$ at 3000m (and for T below $k_{min} = 55$ at 300m) η is exponentially damped towards zero (recall that smaller k correspond to greater height):

$$\eta_X^k(x,t) = \eta_X^{k_{max}}(x,t)e^{-(k_{max}-k)} \quad \text{for } k = k_{max}-1,\ldots,k_{max}-4,$$

and similarly for T at $k = k_{min}+1,\ldots,k_{min}+4$.

Investigations about the model of the model error in the global ICON-EPS within the project Gridcast have shown further dependencies of the parameters (land/sea grid points, geographical latitude) which can be neglected for the application of SMME to the limited domain of COSMO-D2.

3.1.3 Experiments and verification results

Based on the parameters γ, λ, and σ as estimated from a training data set of the operational COSMO-D2 forecasts and analyses for autumn 2018 and 2019, two experiments were performed for October 2018 with the operational COSMO-D2-EPS settings. All forecasts start at 03 UTC and run for 45 hours. Verification shown in this section is against SYNOP observations.
The first experiment resulted in an increased spread for several prognostic variables, but there was a significant increase in the forecast error as well (e.g. for 10m wind speed, see Fig. 5).

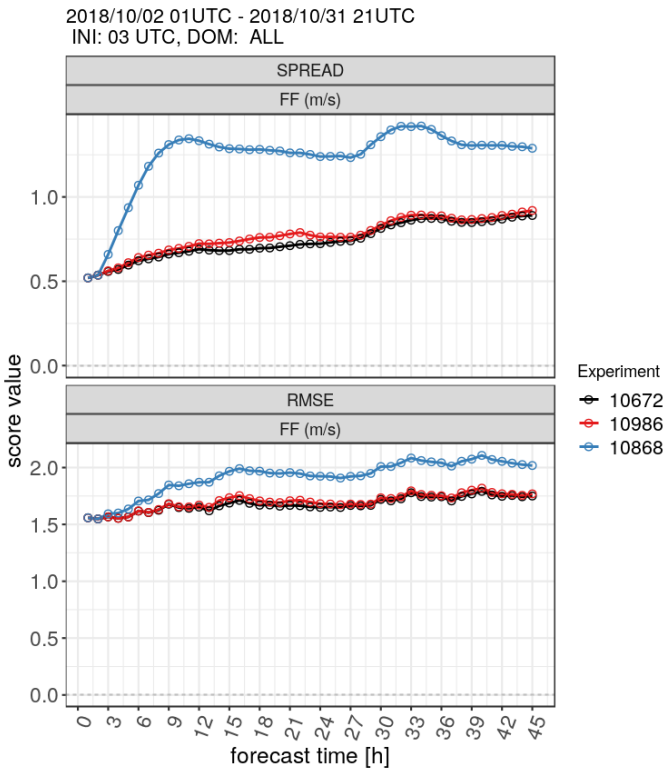

Figure 5: EPS spread (top) and root mean squared error (bottom) of the ensemble mean for 10m wind speed as function of forecast lead time for the first experiment with the model of the model error (blue, experiment 10868), the modified SMME experiment (red, 10968), and a reference run with operational settings but without model error (black, 10672).

These results initiated the changes in the estimation method for the damping parameter γ following García-Ojalvo et al. (1992) as mentioned in section 3.1.1. The effect of this methodological change in comparison to the first experiment can be seen in Fig. 5 as well. The results are mostly neutral compared to the reference, with a slight increase of the spread for wind speed in the afternoon hours of the first forecast day.

The new experiment incorporated a further methodological adjustment. The model tendencies acting as predictors for the parameters in the application of SMME are calculated during the forecast as changes with respect to the model time step. However, the estimation of the tendency-parameter relation from the training data has to refer to the output frequency of the model data, i.e. hourly steps being much larger than the model time step. This leads to a more frequent occurrence of relatively large tendencies in the application compared to the training of the parameters. To account for this effect, the properties of the tendency

distributions are harmonized between the two different time scales by applying a factor of 0.43–0.58 to the variance of the tendency distribution during the model forecast.

Fig. 6 shows several verification scores for 10m wind speed and 2m temperature for the experiment with SMME based on the improved estimation method in comparison to the operational set-up.

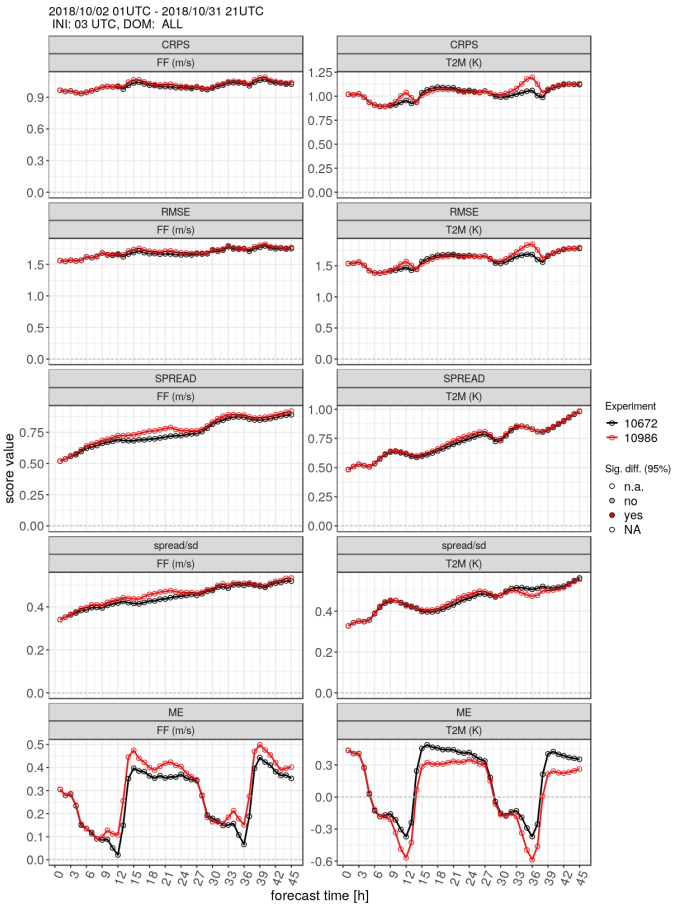

Figure 6: Continuous ranked probability score (CRPS), RMSE, EPS spread, ratio of spread to the standard deviation of forecast errors and mean error of the ensemble mean (from top to bottom) for 10m wind speed (left) and 2m temperature (right) as a function of forecast lead time for the modified SMME experiment (red, experiment 10968) and a reference run with operational settings but without SMME (black, 10672).

The results are mostly neutral compared to the operational set-up. There are slight increases in CRPS and RMSE for 2m temperature around 15 UTC day time, but the differences are not significant (as indicated by the non-filled circles in the plot). There is an increase of spread for wind speed during the first forecast day which is also not significant.

Most obvious differences occur in the mean error of the EPS mean for both variables. The SMME run produces consistently an EPS mean of 10m wind with a slightly larger bias than the reference run which itself has a positive bias. The 2m temperature bias shows qualitatively the same daily cycle with slightly colder temperatures in SMME. Note that the SMME is effective only from 2.5 forecast hours onward to allow for the spin-down of forecast tendencies resulting from minor deviations of analysis fields from the balance properties of the model dynamics. This leads to almost identical mean errors of the SMME and the reference run in the first forecast hours.

In general, the SMME in its current set-up leads to neutral results and the methodological changes were successful in avoiding the forecast deterioration seen in the first experiments. This neutral implementation of the SMME into the COSMO-D2 is a promising starting point for further improvements. A first step for improvements is described in the next section.

3.1.4 Bulk Richardson number as additional predictor

Fig. 7 shows the time series of the mean error of the 24 hours EPS mean forecast of 10m wind speed for the SMME and the reference run.

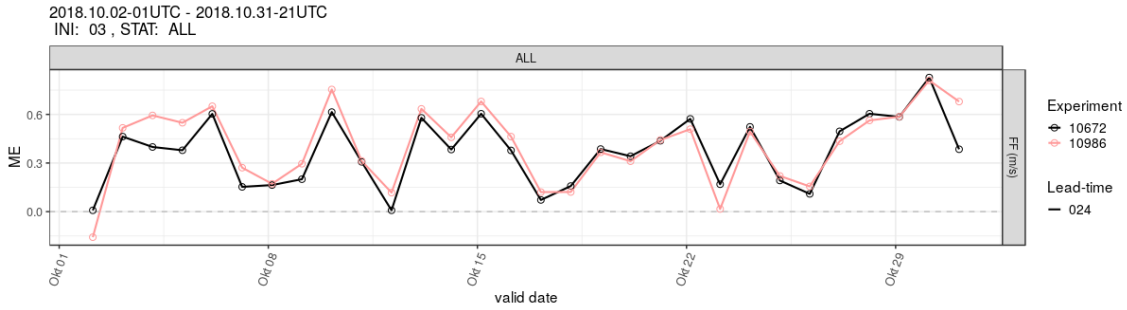

Figure 7: Time series of the mean error of the 24 hours EPS mean forecast of 10m wind speed for the SMME (experiment 10986) and the reference run (experiment 10672).

There are periods with an increased bias in the SMME run, but also periods with neutral results by the SMME approach. A conditional verification showed that the SMME increases the error particularly in nights with stable PBLs. This leads to further investigating how to include knowledge about the stability of the PBL into the SMME. The suggestion is to introduce the bulk Richardson number as a further predictor in addition to the tendency of the respective variable.

We chose to discriminate three cases: stable, unstable, and neutral PBL. To this end, we calculate the bulk Richardson number

$$R_B = \frac{(g/T_v)\Delta\theta_v \Delta z}{(\Delta U)^2 + (\Delta V)^2} \tag{4}$$

for a bulk layer of levels 65–55 (10m–300m) and perform a horizontal smoothing over 7×7 grid points. In eq. 4 we use the following definitions:

- g gravitational constant,
- absolute T_v virtual temperature,
- Δz layer thickness,
- $\Delta\theta_v$ change in virtual potential temperature across layer,
- ΔU, ΔV change in wind speed across same layer.

Our thresholds for the three cases are

$$R = \begin{cases} \text{stable} & R_B >= 1.5 \\ \text{neutral} & R_B \in (0.25, 1.5) \\ \text{unstable} & R_B <= 0.25. \end{cases} \tag{5}$$

Fig. 8 provides the fields of the bulk Richardson number for a stable (14th October) and an unstable (24th October) case.

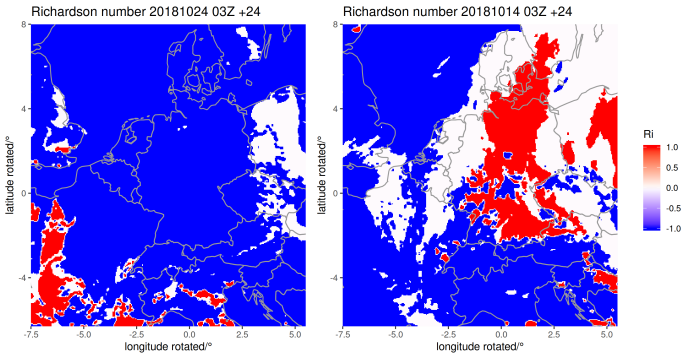

Figure 8: Bulk Richardson number for an unstable (24th October) and a stable (14th October) case.

These fields together with Fig. 7 exemplify the link between nocturnal stable/unstable conditions and their effect on the wind forecast of the COSMO-D2-EPS using the SMME method. The stratification along the three cases stable, unstable, and neutral boundary layer according to R_B leads to differences in the estimated parameters as can be seen in Fig. 9 for the diffusion parameter λ. The stratification of the estimation of the parameters γ, λ, and σ by the bulk Richardson number and the use of R_B as an additional predictor will be adopted in further experiments with SMME.

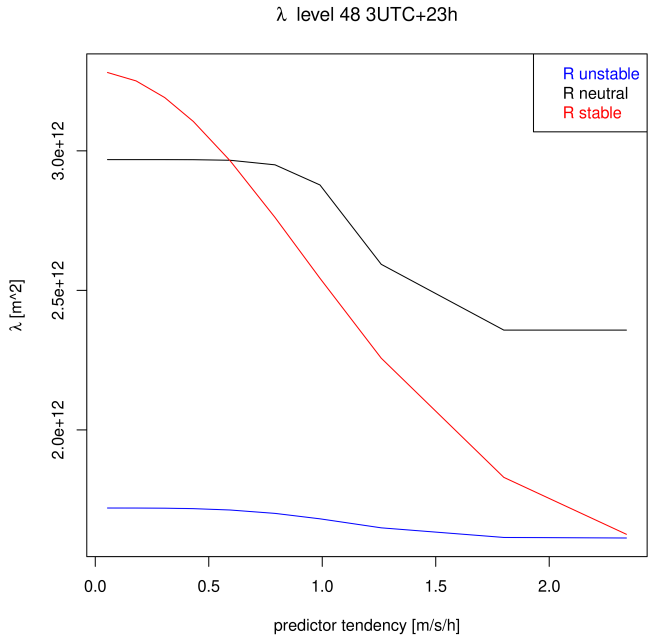

Figure 9: Diffusion parameter λ for zonal wind at 757m at 2 UTC during the night for stable (red), neutral (black), and unstable(blue) conditions.

The next steps in the development of the SMME will be its application to ICON-D2-EPS in its current state, analysing potentials for further improvements of the methodology, and speed-up of the method from a computational point of view.

3.2 AMPT: Additive Model-error perturbations scaled by Physical Tendencies (M. Tsyrulnikov, D. Gayfulin, E. Astakhova, RHM)

In this section, we introduce a new ad-hoc approach to generation of model-error perturbations. The idea is to generate additive perturbations as four-dimensional random fields whose magnitude is determined by the area-averaged modulus of physical tendency. The AMPT (Additive Model-error perturbations scaled by Physical Tendencies) scheme is applied to both atmosphere and soil. Nearly-Gaussian and significantly non-Gaussian fields are treated. Results of numerical experiments with a COSMO-model ensemble prediction system are given.

3.2.1 SPPT in a nutshell

To facilitate the presentation of the new AMPT scheme, we first outline Stochastic Perturbation of Physical Tendency (Buizza et al., 1999), the very popular technique to represent errors in physical parametrizations. This simple and effective technique relies on the assumption that the magnitude of error in physical tendency is proportional to the magnitude of physical tendency itself. There is also a flavour of SPPT called iSPPT in which tendencies from different physical parametrizations are perturbed independently (Christensen et al., 2017).

By $\mathbf{P}(x,y,\zeta,t) = (P_1(x,y,\zeta,t),\dots,P_{n_{\text{fields}}}(x,y,\zeta,t))$ we denote the vector valued physical tendency (sometimes called the *net* physical tendency, generated by all physical parametrizations combined) at the spatial grid point with the horizontal coordinates (x,y), the vertical coordinate ζ, and the time step t. Here $P_i(x,y,\zeta,t)$ is the component of $\mathbf{P}(x,y,\zeta,t)$ in the i-th model field (variable) X_i, and n_{fields} is the number of model fields selected to be perturbed, normally, temperature, winds, and humidity.

In SPPT, the perturbed physical tendency $\mathbf{P}^*$ is postulated to be

$$\mathbf{P}^*(x,y,\zeta,t) = (1+\epsilon\,\xi(x,y,t))\cdot\mathbf{P}(x,y,\zeta,t), \tag{6}$$

where $\xi(x,y,t)$ is a *scalar* (i.e. common for all physical tendency components) zero mean and unit variance spatio-temporal random field and ϵ the parameter that controls the magnitude of the perturbation (normally, 0.4–0.8). For stability reasons, the point wise probability distribution of $\xi(x,y,t)$ is specified in such a way that $1+\epsilon\,\xi(x,y,t)>0$, that is, avoiding sign reversal of the physical tendency (Leutbecher et al., 2017). Note that in SPPT the random pattern ξ is two-dimensional in space.

From eq. 6, the SPPT perturbation is seen to be multiplicative:

$$\delta\mathbf{P} = \mathbf{P}^*(x,y,\zeta,t) - \mathbf{P}(x,y,\zeta,t) = \epsilon\,\xi(x,y,t)\,\mathbf{P}(x,y,\zeta,t). \tag{7}$$

Here and elsewhere δ denotes a perturbation.

3.2.2 AMPT

Motivation

The following three deficiencies of SPPT led us to propose the new approach.

1. In SPPT, perturbations are large when and where the physical parametrizations generate a *large* physical tendency $\mathbf{P}$. This formulation gives rise to a meaningful scaling

of perturbations in situations when the model predicts high subgrid-scale variability and produces a large physical tendency. But it cannot cover situations in which the model error magnitude is large whereas the magnitude of the physical tendency is small. This occurs, for example, if in some grid cell convection is initiated in nature whilst a convective parametrization fails to be activated.

2. In SPPT, the *multivariate* perturbation vector $\delta\mathbf{P}(x, y, \zeta, t)$ is strictly proportional to the physical tendency vector $\mathbf{P}(x, y, \zeta, t)$ (see eq. 7). This implies that *SPPT tacitly assumes that the model error is only in the magnitude* of the vector $\mathbf{P}$, whilst the *ratios* between the physical tendencies in different variables (i.e. between $\delta P_i(x, y, \zeta, t)$ for different variables labelled by i at the same grid point x, y, ζ, t) are *error-free*, which is highly unlikely. Indeed, eq. 7 implies that the relationship between any two components of $\mathbf{P}$, say, those that correspond to temperature T and zonal wind u, i.e. P_{T} and P_u, is the same before and after the perturbation:

$$\frac{P^*_{\mathrm{T}}}{P^*_u} = \frac{P_{\mathrm{T}}}{P_u}. \tag{8}$$

   The *ratios* of the components of the physical tendency vector are, thus, not perturbed in SPPT (in iSPPT, this is the case for each individual physical parametrization).

3. Similarly, as the SPPT random pattern ξ does not depend on the vertical coordinate ζ, the ratios of physical tendencies at different *vertical levels* are not perturbed:

$$\frac{P^*_i(x, y, \zeta_k, t)}{P^*_{i'}(x, y, \zeta_{k'}, t)} = \frac{P_i(x, y, \zeta_k, t)}{P_{i'}(x, y, \zeta_{k'}, t)} \tag{9}$$

   for any two variables i, i' and any two vertical levels k, k'. This implies that there are *no errors* in those ratios, which, again, is unlikely.

Approach

We propose to address the above deficiencies of SPPT in our AMPT scheme as follows.

1. In SPPT, as it follows from eq. 7, the standard deviation of the perturbation in the i-th prognostic field (given $\mathbf{P}$) is $\epsilon|P_i|$ so that the standard deviation of the multivariate (vector) perturbation is proportional to the modulus of the physical tendency: $\mathrm{SD}(\delta\mathbf{P}|\mathbf{P}) = \epsilon|\mathbf{P}|$. In AMPT, we retain this linear dependency of the magnitude (standard deviation) of model-error perturbations on the modulus of physical tendency. But we define this dependency to be more general than just *point wise*.

   Specifically, we propose to generate model-error perturbations in the *additive* (rather than multiplicative as in SPPT) way and postulate their magnitude to be proportional to an *area-averaged* (rather than point wise as in SPPT) $|\mathbf{P}|$ in the horizontal. This allows AMPT to generate non-zero perturbations even at grid points with zero physical tendency — if there are nearby points with non-zero physical tendency.

2. We allow for errors in the relationships between physical tendencies in different model variables by introducing *independent* driving random fields for different model variables.

3. We allow for errors in the relationships between physical tendencies at different model levels by making ξ a *four-dimensional* random field $\xi(x, y, \zeta, t)$ rather the three-dimensional field $\xi(x, y, t)$ as in SPPT.

It is worth remarking that the SPPT perturbations at any given horizontal-grid point and time step are perfectly correlated between different variables and different levels, whereas the AMPT perturbations are uncorrelated between different variables and partly correlated in the vertical. This suggests that a hybrid SPPT+AMPT scheme might be considered. In this study we focus on the pure AMPT.

3.2.3 AMPT model-error perturbations: "Gaussian" atmospheric variables

In the context of limited-area modelling, we define the area-averaged magnitude of physical tendency for nearly Gaussian distributed fields such as temperature T and the two wind components u, v (T, u, v are denoted generically as Y in this section) through averaging over the *whole domain* in the horizontal:

$$\mathcal{P}_{\mathrm{Y}}(z,t) = \widehat{\mathbb{E}}|P_{\mathrm{Y}}(x,y,\zeta,t)|, \tag{10}$$

where ζ stands for the vertical coordinate, P_{Y} is the tendency in Y due to all physical parametrizations, and $\widehat{\mathbb{E}}$ is the horizontal averaging operator on the model grid. Then, the perturbation is defined to be

$$\delta Y(x,y,\zeta,t) = \epsilon_{\mathrm{Y}} \cdot \mathcal{P}_{\mathrm{Y}}(\zeta,t) \cdot \xi_{\mathrm{Y}}(x,y,\zeta,t), \tag{11}$$

where the multiplier ϵ_{Y} is the external parameter that controls the overall magnitude of the perturbation and $\xi_{\mathrm{Y}}(x,y,\zeta,t)$ is the Gaussian pseudo-random field with zero mean, unity variance and the tunable spatio-temporal structure (that is, spatial and time scales). The four-dimensional pseudo-random fields ξ are to be generated "online", that is, in the course of the forecast model's time integration. In this study we used the Stochastic Pattern Generator (Tsyrulnikov et al., 2017) for this purpose. The perturbations $\delta T(x,y,\zeta,t), \delta u(x,y,\zeta,t), \delta v(x,y,\zeta,t)$ (and any other perturbations referenced below) are computed and added to the model state at each time step.

Note that theoretically, specifying mutually uncorrelated random stream function and velocity potential fields instead of specifying uncorrelated random u and v wind components is to be preferred. The reason is that the former approach allows for *isotropic* vector-wind perturbations (Monin et al., 2013), unlike the latter approach. However, in practical terms, we were not able to identify any significant flaw in the vector field composed of two independent u and v fields. For this reason and due to the almost complete lack of evidence on the actual structure of model errors, we stick to the simpler formulation of AMPT with mutually *independent* $\delta u(x,y,\zeta,t)$ and $\delta v(x,y,\zeta,t)$ in this study.

From $\delta T(x,y,\zeta,t)$, we compute the pressure perturbation $\delta p(x,y,\zeta,t)$ integrating the hydrostatic equation (in which δq is neglected as a small contribution to a small perturbation). The integration goes from the model's top pressure downwards.

3.2.4 AMPT model-error perturbations: Humidity and hydrometeor fields

In contrast to roughly Gaussian distributed physical tendencies for T, u, v, physical tendencies for specific humidity q_{v}, cloud water content q_{c}, and cloud ice content q_{i} are highly non-Gaussian. Specifically, there are areas of high physical tendencies (related, mainly, to precipitation) and there are much (orders of magnitude) lower tendencies in non-precipitating areas (not shown). To simplify notation, we denote in this subsection any of these fields generically by Z in this section.

The highly variable magnitude of the physical tendency for non-Gaussian fields in different areas across the forecast domain makes the above whole-domain averaging (eq. 10) not meaningful. For these fields, we propose the *local-area* (moving-window) averaging as a means to compute the *local* (instead of global) area-averaged magnitude of physical tendency:

$$\mathcal{P}_{\mathrm{Z}}(x,y,\zeta,t) = \frac{\iint_G h_A(x-x',y-y')\cdot|P_{\mathrm{Z}}(x',y',\zeta,t)|\,\mathrm{d}x'\mathrm{d}y'}{\iint_G h_A(x-x',y-y')\,\mathrm{d}x'\mathrm{d}y'}, \tag{12}$$

where G is the model domain and h_A is the averaging kernel chosen to have the simplest "boxcar" shape:

$$h_A(x,y) = \frac{1}{A^2} \qquad \text{if } |x| < \frac{A}{2} \text{ and } |y| < \frac{A}{2} \qquad \text{and 0 otherwise} \tag{13}$$

(A is the averaging radius). The *local* averaging of the absolute (i.e. the absolute value or modulus of the) physical tendency, eq. 12, allows us to model the non-Gaussian perturbation field δZ still using the Gaussian SPG random pattern ξ_{Z}:

$$\delta Z(x,y,\zeta,t) = \epsilon_{\mathrm{Z}}\cdot\mathcal{P}_{\mathrm{Z}}(x,y,\zeta,t)\cdot\xi_{\mathrm{Z}}(x,y,\zeta,t), \tag{14}$$

where ϵ_{Z} is the external parameter (like ϵ_{Y} in eq. 11). In mathematical terms, eq. 14 implies that δZ is conditionally Gaussian given $\mathcal{P}_{\mathrm{Z}}$ and thus unconditionally non-Gaussian due to the random nature of the $\mathcal{P}_{\mathrm{Z}}$ field.

After the perturbation δZ is added to the model state at the current time step, $Z^* = Z + \delta Z$, we limit the perturbed humidity and hydrometeor fields below by nullifying negative $Z^*(x,y,\zeta,t)$. We do *not* apply a humidity supersaturation limiter (because otherwise the perturbations led to biases in perturbed forecasts).

Note that if the local averaging radius A is big enough, the moving-window scheme becomes equivalent to the formulation of AMPT with the whole-domain averaged magnitude of physical tendency. Therefore, we will assume by default that the AMPT scheme is formulated with the *moving-window-averaged* area-averaged magnitude of physical tendency $\mathcal{P}$ (as described in this section). Thus, for the i-th field, we specify its own averaging radius A_i, larger for Gaussian and smaller for non-Gaussian model variables (see section 3.2.8 for the actual choices we made).

Note that the local-averaging formulation of AMPT is applicable for the global domain as well as for the local (regional) domain.

3.2.5 AMPT model-error perturbations: Soil fields

In the land-surface model, tendencies of the two prognostic fields are perturbed: soil temperature T_{so} and soil moisture (more specifically, soil water content W_{so}), at n_{z} levels in the soil (which are counted downwards).

Due to the lack of reliable error statistics in the soil, we set up a scheme which is (i) as simple as possible, (2) consistent with statistics of the soil fields, and (iii) provides a reasonable forecast spread both in the soil and just above the land surface in the atmosphere.

The approach is to apply the above AMPT scheme. The differences with the atmospheric AMPT are:

1. In the soil AMPT, we perturb the *total* (not physical) tendency, assuming that all processes in the soil that contribute to the total tendency are modelled with substantial uncertainty.

2. The area-averaged absolute tendencies, $\mathcal{P}_{T_{\rm so}}$ and $\mathcal{P}_{W_{\rm so}}$, are computed using grid points over land only.

3. The perturbation patterns (SPG random fields) in the soil, $\xi_{T_{\rm so}}(x,y,t)$ and $\xi_{W_{\rm so}}(x,y,t)$, are two-dimensional (not three-dimensional as in the atmosphere).

4. The range of $W_{\rm so}$ is limited so that the volumetric soil water content

$$\eta_{\rm so}(x,y,z,t) = \frac{W_{\rm so}(x,y,z,t)}{\Delta_z} \tag{15}$$

(where Δ_z is the thickness of the z-th soil layer) lies between the wilting-point $\eta_{\rm wp}$ and field capacity $\eta_{\rm fc}$:

$$\eta_{\rm wp} \leqslant \eta_{\rm so}(x,y,z,t) \leqslant \eta_{\rm fc}. \tag{16}$$

If the perturbed soil water content $W_{\rm so}(x,y,z,t)$ (at any spatial grid point and any model time step) does not satisfy the constraint eq. 16, it is truncated accordingly.

For $T_{\rm so}$, the *domain* averaging is used to compute the area-averaged tendency $\mathcal{P}_{T_{\rm so}}$ (as in eq. 10). For $W_{\rm so}$, the *moving-window* averaging is employed to compute $\mathcal{P}_{W_{\rm so}}$ (as in eq. 12). These choices are motivated below in section 3.2.9.

3.2.6 Initial soil perturbations

In the soil, processes have much longer time scales than in the atmosphere, therefore weeks of cycling are needed to reach equilibrium in the ensemble statistics there. To avoid long cycling, we developed a generator of *initial* $T_{\rm so}$ and $W_{\rm so}$ perturbations.

The initial soil *temperature* perturbation is specified to be

$$\delta T_{\rm so}^{\rm ini}(x,y,z) = \sigma_{T_{\rm so}}^{\rm ini} \cdot \varkappa_{T_{\rm so}}^{1-z} \cdot \xi(x,y), \tag{17}$$

where $\sigma_{T_{\rm so}}^{\rm ini}$ is the external parameter that defines the magnitude of the perturbation at the uppermost level, $\varkappa_{T_{\rm so}} > 1$ is the vertical-decay external parameter, $z = 1, 2, \ldots, n_{\rm z}$ labels the vertical level, and $\xi(x,y)$ is the 2D SPG pseudo-random field. Note that the perturbation pattern ξ is the same for all vertical levels, whilst the magnitude of the perturbation exponentially decreases downwards (the decrease was motivated by the experimental results given below in section 3.2.9).

With the soil moisture $W_{\rm so}$, the technique is to perturb the Soil Moisture Index

$$S = \frac{\eta_{\rm so} - \eta_{\rm wp}}{\eta_{\rm fc} - \eta_{\rm wp}} \tag{18}$$

as follows:

$$\delta S^{\rm ini}(x,y,z) = \sigma_{S}^{\rm ini} \cdot \varkappa_{W_{\rm so}}^{1-z} \cdot \xi(x,y), \tag{19}$$

where $\sigma_{S}^{\rm ini}$ is the magnitude parameter like $\sigma_{T_{\rm so}}^{\rm ini}$ in eq. 17 and $\varkappa_{W_{\rm so}}$ is the vertical-decay parameter. If at some grid point the perturbed S appears to lie outside the meaningful range $[0,1]$, the perturbation $\delta S^{\rm ini}$ is truncated accordingly.

Using equations(18) and (14), we finally convert the perturbation $\delta S^{\rm ini}(x,y,z)$ into the perturbation of $W_{\rm so}$:

$$\delta W_{\rm so}^{\rm ini}(x,y,z) = \Delta_z \cdot [\eta_{\rm wp} + \delta S^{\rm ini}(x,y,z)(\eta_{\rm fc} - \eta_{\rm wp})]. \tag{20}$$

3.2.7 Stability

Stochastic-dynamic schemes in which the magnitude of forcing depends on the current state of the system can be unstable due to the positive feedback loop: a larger perturbation at the current time step may lead to a greater deviation of the state from its time-mean value, which may result in a larger physical tendency at the next time step, which would give rise to a larger perturbation, etc. until explosion.

To break this vicious circle, we considered two strategies. A technically simpler one, which we adopted in this study, is to update $\mathcal{P}_i$ not at every time step, but less frequently, once per $\mathcal{T}_{\mathcal{P}_i}^{\text{update}}$ lead time, where $\mathcal{T}_{\mathcal{P}_i}^{\text{update}}$ is the external parameter defined below in section 3.2.8.

A somewhat more involved but potentially more powerful approach (which we left for future research) is to calculate $\mathcal{P}(x, y, \zeta, t)$ from the *unperturbed* (control) model run (possibly, on a coarse spatio-temporal grid) and then use it to compute AMPT perturbations for all ensemble members. This will completely destroy the harmful positive feedback.

3.2.8 Experimental setup

The domain is shown in Fig. 10 (the Sochi Olympics-2014 area). Note the complexity of the area, which contains high mountains adjoining the sea and the valleys. The centre of the domain is located nearly at 44° N. The climate at the sea level is humid subtropical. Numerical experiments were carried out in this study mostly for the winter-spring season: in February and March, for which we had access to all data needed to run and verify an EPS. Some sensitivity experiments were also conducted for May cases.

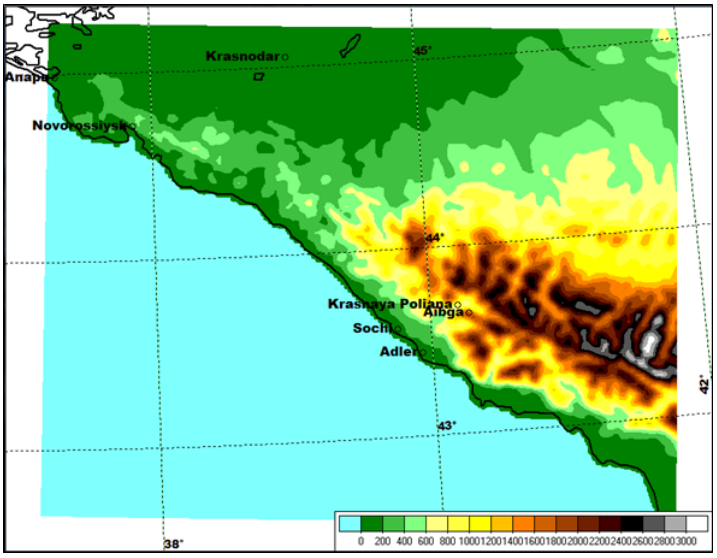

Figure 10: Model domain and orography.

The COSMO model (Baldauf et al., 2011) (version 5.01) in the convection permitting configuration with the horizontal grid spacing of 2.2 km, with the grid of 172*132 points in the horizontal, and 50 vertical levels was used for numerical experiments in this study. The model integration time step was 20 s. All calculations were performed in a single precision mode. The model's vertical coordinate was the height-based hybrid (Gal-Chen) coordinate (Gal-Chen et al., 1975).

The following AMPT parameters were selected for numerical experiments. The horizontal

length scale L_i of the SPG driving random fields ξ_i (where i labels the perturbed fields) was equal to 50 km (i.e. 20–25 horizontal mesh sizes) for T, u, v, 70% of this value (i.e. 35 km) for q_v, and 50% (25 km) for the cloud fields.

The time scales $\mathcal{T}_i$ of the atmospheric random fields ξ_i were specified as L_i/U , where U is the characteristic velocity. By trial and error, we found that $U = 15$ m/s worked well (and is physically meaningful). For both soil fields, i.e. T_so and W_so, the common time scale $\mathcal{T}_\mathrm{so}$ was specified as the time scale for atmospheric temperature, $\mathcal{T}_\mathrm{T}$, multiplied by a factor $\tau_\mathrm{so} > 1$ that reflects a much greater inertia of the soil as compared to the atmospheric air, so that $\mathcal{T}_\mathrm{so} = \tau_\mathrm{so} \cdot \mathcal{T}_\mathrm{T}$. The optimally tuned value of the soil inertia factor τ_so was 12 (i.e. the time scales were, roughly, 1 h in the atmosphere and 12 h in the soil).

The time-update interval $\mathcal{T}_{\mathcal{P}_i}^{\mathrm{update}}$ (see section 3.2.7) for the model field X_i was set equal to the time scale $\mathcal{T}_i$ of the respective random pattern ξ_i. Note that with $\mathcal{T}_{\mathcal{P}_i}^{\mathrm{update}}$ much lower than $\mathcal{T}_i$, the perturbed model may become unstable (see section 3.2.7), whereas with $\mathcal{T}_{\mathcal{P}_i}^{\mathrm{update}}$ much greater than $\mathcal{T}_i$, the resulting area-averaged $\mathcal{P}_i$ may become irrelevant in a rapidly developing meteorological situation.

The magnitude parameter of the initial T_so perturbation, $\sigma_{T_\mathrm{so}}^{\mathrm{ini}}$, was set to 1 K. The magnitude parameter of the initial soil moisture index perturbation, σ_S^{ini} was set to 0.01 (larger values led to unrealistically large model tendencies in T_so).

After some experimentation, the model-error multipliers ϵ_i were specified equal to 0.75 for T, u, v, T_so, and W_so, and 0.5 for q_v, q_c, and q_i.

The absolute-physical-tendency averaging length scale A_i (see section 3.2.4) was specified equal to the length scale L_i of the respective SPG random field ξ_i for q_v, q_c, q_i, and W_so. For the nearly Gaussian fields T, u, v, the respective A_i were selected large enough to ensure the whole-domain averaging of $|P_i|$.

In the ensemble prediction experiments presented below, we actually perturbed only temperature and winds because this configuration produced stable and positive results. As for humidity and cloud fields perturbations, they led to mixed results in EPS verification scores and so were not included in the model-error perturbation scheme.

We also compared AMPT with SPPT, implemented in the COSMO model by L. Torrisi and tuned by Maurer et al., 2014. In the SPPT, the standard deviation of the random field was set to 0.8, the spatial scale to 5° (i.e. some 500 km), and the time scale to 6 h. Temperature, wind, and humidity were perturbed.

Tapering, that is, a gradual reduction of the perturbation magnitude (1) in the lower troposphere towards the surface and (2) in the stratosphere from the tropopause upwards, was handled in the SPPT as follows. The stratospheric tapering was always active in our experiments because it is believed that the radiation tendency, which is dominant in the stratosphere, is quite accurate in clear-sky conditions (e.g. Leutbecher et al., 2017). As for the lower-tropospheric tapering, which is intended to prevent instabilities due to inconsistencies of perturbed physical tendencies and unperturbed surface fluxes (Wastl et al., 2019), we found that the SPPT worked better without it. Specifically, our experiments showed that, on the one hand, the SPPT without tapering was stable in the boundary layer. On the other hand, with tapering, SPPT led to an unacceptably small ensemble spread in the near-surface fields, so we switched off the lower-tropospheric tapering in the SPPT. In AMPT, tapering was switched off in the experiments described below.

Note that the above spatial and time scales employed in AMPT were an order of magnitude less than those in the SPPT (50 km vs. 500 km and 1 h vs. 6 h). This is reasonable because

in SPPT, the perturbation is *multiplicative*, i.e. the random field (pattern), $\xi(x, y, t)$, is multiplied by the physical tendency, $\mathbf{P}(x, y, \zeta, t)$, which is a rather small-scale random field itself, so that the perturbation, $\xi(x, y, t) \cdot \mathbf{P}(x, y, \zeta, t)$, appears to be a reasonably small-scale field even if $\xi(x, y, t)$ is large-scale. If $\xi(x, y, t)$ were small-scale in SPPT, the product $\xi \cdot \mathbf{P}$ would be too patchy, which would reduce the effect on the forecast. In contrast, AMPT introduces an *additive* perturbation, which needs to be small-scale in order to represent the errors in parametrizations of *subgrid-scale* processes.

To find out if this argument is reasonable, we ran the SPPT with the smaller time scale of 1 h and the spatial scales of 50 and 100 km. The resulting spread in the ensemble forecast was very small, confirming the conclusions of Maurer et al., 2014 (made for a different domain, orography, physiography, etc.) and justifying the choice of the SPPT parameters for our domain (the setup was also consistent with that employed in AROME-EPS, (Bouttier et al., 2012)). For the set-up of SPPT in COSMO, see also Klasa et al. (2018).

Note that in what follows, by "model perturbations" we mean model-error perturbations generated by AMPT and SPPT in the atmosphere and also the initial and model-error perturbations generated by AMPT in the soil.

3.2.9 Unperturbed T_{so}, W_{so} total tendencies

Fig. 11 shows the unperturbed tendencies for T_{so} and W_{so} at the soil level $z = 2$ and lead time 2 h (the forecast started at 00 UTC on the 1st of May 2014). The tendencies were computed as one-time-step differences and recalculated w.r.t. 1 h (so that the T_{so} tendency is measured in K/h).

The T_{so} tendencies (left panel in Fig. 11) look more or less uniform (albeit with a few isolated outliers) and thus not far from Gaussian. For this reason, the AMPT perturbation scheme for T_{so} model-error perturbations was selected to involve whole-domain (not moving-window) averaging of the modulus of the tendency.

The W_{so} tendencies on the contrary (right panel in Fig. 11), are highly inhomogeneous and thus non-Gaussian. This motivated the use in the AMPT perturbation scheme for W_{so} of a *moving-window* averaging of the modulus of the tendency to scale the model-error perturbation of W_{so}.

In the vertical, the mean absolute tendency in $\mathrm{T_{so}}$ roughly halved when going one level down in the soil (not shown). So did (again, roughly) the maximal absolute tendency in $\mathrm{W_{so}}$[1].

We found no significant dependencies in the magnitude and spatial structure of soil tendencies on the soil type (there were only two soil types in the COSMO model across the domain, not shown).

One can see a signature of orography (Fig. 10) in the soil tendency fields (Fig. 11).

3.2.10 Testing AMPT in an ensemble prediction system

Numerical experiments were carried out to study the effect of AMPT model perturbations in COSMO-Ru2-EPS (Montani et al., 2014 and Astakhova et al., 2015), the ensemble prediction

[1]The *maximal* absolute tendency looks more appropriate to measure the magnitude of the tendency in $\mathrm{W_{so}}$ than the *mean* absolute tendency — because the magnitude of the soil moisture tendency exhibits very big changes over the model domain (see Fig. 11, right panel) so that the averaged value is irrelevant for areas of both big and small model tendencies (that is, almost everywhere).

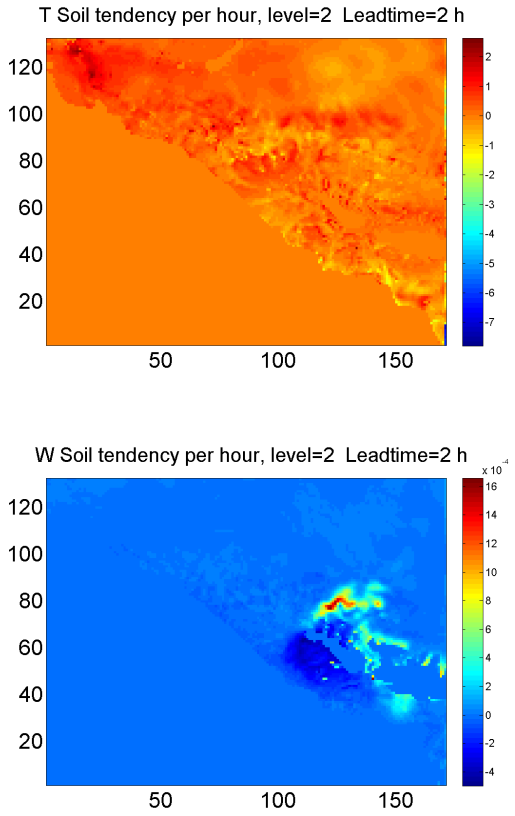

Figure 11: Total tendency of T_{so} (left, K/h) and W_{so} (right, W_{so} units per hour) at level 2, forecast lead time 2h. The forecasts were initialized at 00 UTC 1 May 2014. Tick labels are grid points in the longitudinal (x-axis) an latitudinal (y-axis) directions

system developed within the FROST-2014 project (Kiktev et al., 2017) and CORSO priority project of the COSMO consortium (Rivin et al., 2018). For the COSMO model set-up used in the ensemble prediction system, see above section 3.2.8. The ensemble size was 10. COSMO-Ru2-EPS performed a dynamical downscaling of the forecasts of the driving COSMO-S14-EPS system developed by ARPA-SIMC for the Sochi Olympics (Montani et al., 2013). Thus, both initial and lateral boundary conditions for the control forecast and the ensemble members were provided by COSMO-S14-EPS. In turn, COSMO-S14-EPS was a clone of the consortium ensemble prediction system COSMO-LEPS (Montani et al., 2011) with a smaller ensemble size and shifted to the area of the Sochi Olympics. COSMO-S14-EPS had the horizontal grid spacing of 7 km and 40 levels in the vertical.

Experiments setup

A continuous series of experiments was performed with COSMO-Ru2-EPS for the two-month period of February-March 2014 without model perturbations as well as using AMPT and SPPT with the parameters indicated above (section 3.2.8). The forecasts were initialized every day at 00 UTC. The local time for the region considered was UTC+3 h. The results were verified against near-surface observations (about 40 stations) using the COSMO unified verification system VERSUS (Raspanti, 2009). Ensemble forecasts in the free atmosphere were verified against zero-lead-time ECMWF forecasts (with the horizontal resolution of about 16 km).

Below we show verification scores for the whole two-month period. Similar results were obtained for each of the two months separately (not shown).

We concentrate on the impact of model perturbations on probabilistic forecasts of $T_{2\mathrm{m}}$ and precipitation.

The list of experiments and their basic features are presented in Table 1.

Table 1: List of experiments with COSMO-Ru2-EPS

Experiment	Model perturbations in addition to initial and lateral-boundary ones
NOPERT	None
SPPT	SPPT model-error perturbations in the atmosphere
AMPT-NOSOIL	AMPT model-error perturbations in the atmosphere
AMPT-SOIL	AMPT perturbations: atmosphere + soil

Deterministic verification

Fig. 12 shows the root-mean-square error (RMSE) of the $T_{2\mathrm{m}}$ *ensemble mean* forecast (the upper bunch of curves). One can see that, excluding the initial transient (spinup) period of some 3 h (likely, due to imbalances in the initial perturbations), the RMSE had a prominent diurnal cycle with a broad minimum at night and a narrower maximum shortly after midday. This corresponds to the known systematic error in $T_{2\mathrm{m}}$ in the COSMO model (Fig. 2 in Kirsanov et al., 2020).

Fig. 13 displays the normalized ensemble-mean RMSE reduction w.r.t. NOPERT (to highlight the barely seen in Fig. 12 differences between the experiments), that is, $(RMSE_{NOPERT} - RMSE)/RMSE_{NOPERT}$. SPPT led to a persistent slight reduction in RMSE. AMPT model perturbations gave rise to a more significant decrease in the $T_{2\mathrm{m}}$ RMSE most of the time except for the rather short time periods in the afternoon, when the ensemble-mean forecast deteriorated. Soil perturbations added skill to the ensemble mean forecast: the RMSE reduction is higher in the experiment AMPT-SOIL than in AMPT-NOSOIL.

Reliability

The lower bunch of curves in Fig. 12 shows the ensemble spread, which was a bit too low with both SPPT and AMPT. This implies that the ensemble forecast was lacking reliability, though AMPT perturbations increased the spread to a much greater extent than SPPT perturbations. The effect of the AMPT perturbations on the spread was nearly the same for all lead times. The gradual overall growth of the spread with time was, likely, caused by the lateral boundary perturbations as they propagated from the boundaries to the domain interior. This explanation is consistent with the gradual reduction in the differences between the experiments (since they had the same lateral boundary perturbations).

Higher in the atmosphere, the RMSE of temperature and wind speed was negligibly affected by AMPT perturbations while the related spread grew most efficiently near the surface and fell with height.

For temperature, the AMPT atmospheric perturbations resulted in a spread growth by 0.3 K at 20 m, 0.25 K at 150 m, and 0.1 K at 1000 m. The soil perturbations added only about 0.08 K to the spread below 150 m and their effect vanished towards the top of the boundary layer.

For wind speed, the spread gain due to the AMPT purely atmospheric perturbations was

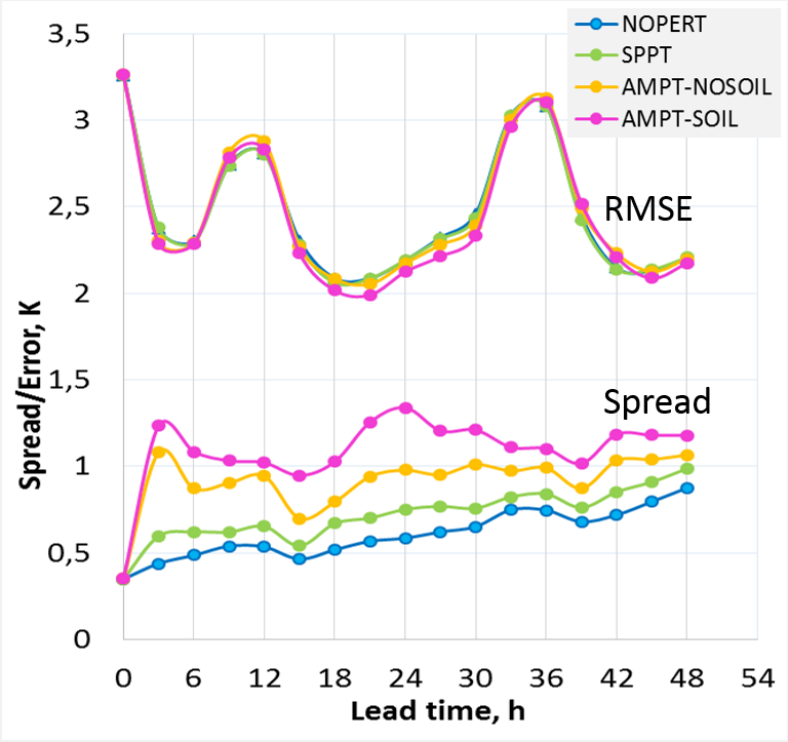

Figure 12: RMSE of $T_{2\mathrm{m}}$ ensemble mean and spread. Experiments NOPERT (blue), SPPT(green), AMPT-NOSOIL (orange), and AMPT-SOIL (magenta).

about 0.3-0.4 m/s below 150 m and reduced to 0.2 m/s at 1000 m. The impact of soil perturbations on the wind speed was rather weak. Even at the lowest model level (20 m) the spread difference between AMPT-NOSOIL and AMPT-SOIL was only 0.01 m/s.

AMPT perturbations of humidity and cloud fields generated significant spread in precipitation and in the upper-air temperature during precipitating events but the effect of those perturbations on the ensemble prediction performance scores was mixed (not shown). Further experiments are needed to reach more conclusive results.

Probabilistic verification

Standard probabilistic scores for near-surface variables were calculated for all experiments. The continuous ranked probability score (CRPS), which generalizes the mean absolute error to the case of probabilistic forecasts, is presented in Fig. 14 for $T_{2\mathrm{m}}$. CRPS measures the difference between the predicted and occurred cumulative probability distributions. Both SPPT and AMPT improved CRPS but AMPT was undoubtedly more successful, especially with additional soil perturbations.

The advantage of AMPT over SPPT for $T_{2\mathrm{m}}$ is also supported by the analysis of the Brier score, which characterizes the magnitude of the probability forecast error, see Fig. 15. In fact, this result could be expected as CRPS is usually interpreted as the integral of the Brier score over all possible threshold values (Hersbach,2000).

We also examined the ability of our EPS with different model perturbations to discriminate between two alternative outcomes (say, "T is below a threshold" and "T is above the threshold"). The relative operating characteristic (ROC) is normally used to assess the discrimination ability of an EPS. Fig. 16 shows the area below the ROC curve (ROCA) for two thresholds. Note that the perfect value for ROCA is 1 and the value less than 0.5 indicates

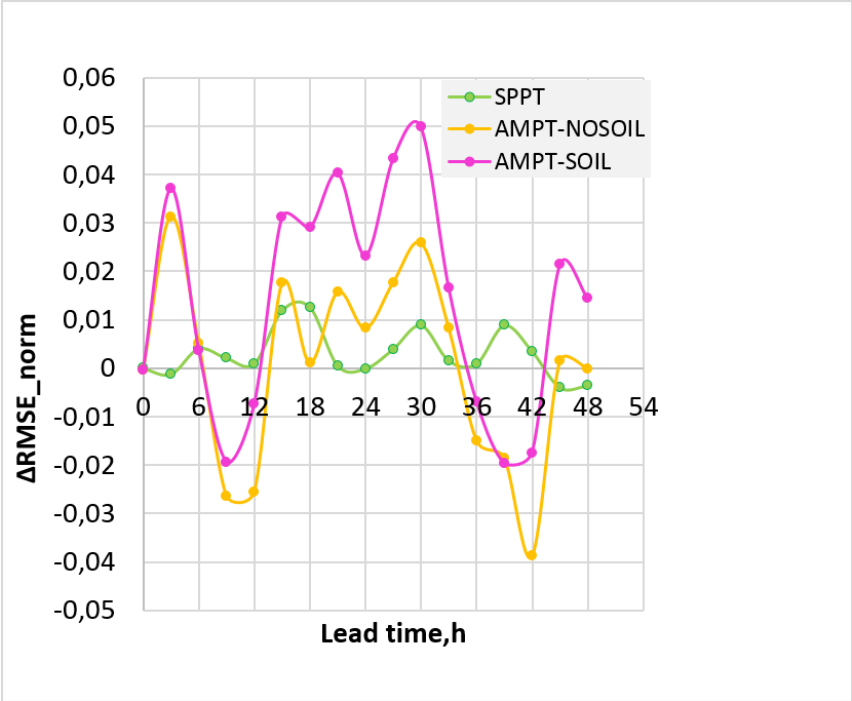

Figure 13: Normalized $T_{2\mathrm{m}}$ ensemble-mean RMSE reduction (the greater the better).

no skill. Actually, ROCA in Fig. 16 is quite high, indicating that the EPS was skillful. Like the previously mentioned probabilistic scores, ROCA demonstrates the advantage of the EPS configuration AMPT-SOIL for both thresholds indicated in Fig. 16 and for the majority of other thresholds we tested (not shown).

The effect of T, u, v perturbations on precipitation was rather ambiguous (see Fig. 17)), perhaps because the time period considered was rather dry (Vasil, 2015). The first half of February 2014 was characterized by an anticyclonic weather, that favoured the precipitation deficit. Only during the second ten-day period of March 2014 considerable amounts of precipitation were observed. Further experiments are needed and will be conducted after this study.

3.2.11 Discussion on AMPT

Limitations of the experimental setup

1. The initial and lateral-boundary (driving) ensemble was under dispersive. This factor might lead to an overestimation of the tuned AMPT *model-error* parameters (the model-error perturbations might be forced to, partly, do the job of initial and lateral-boundary perturbations).

2. The study is conducted on a small model integration domain and for a time period of only one season, which may imply that additional experiments with AMPT on a domain in a different climate zone, different orography and physiography in different seasons and weather conditions need to be conducted to get more conclusive results.

3. In particular, the period of February-March 2014 was rather dry, so it was hard to justify the effect of AMPT perturbations on precipitation. However, sensitivity exper-

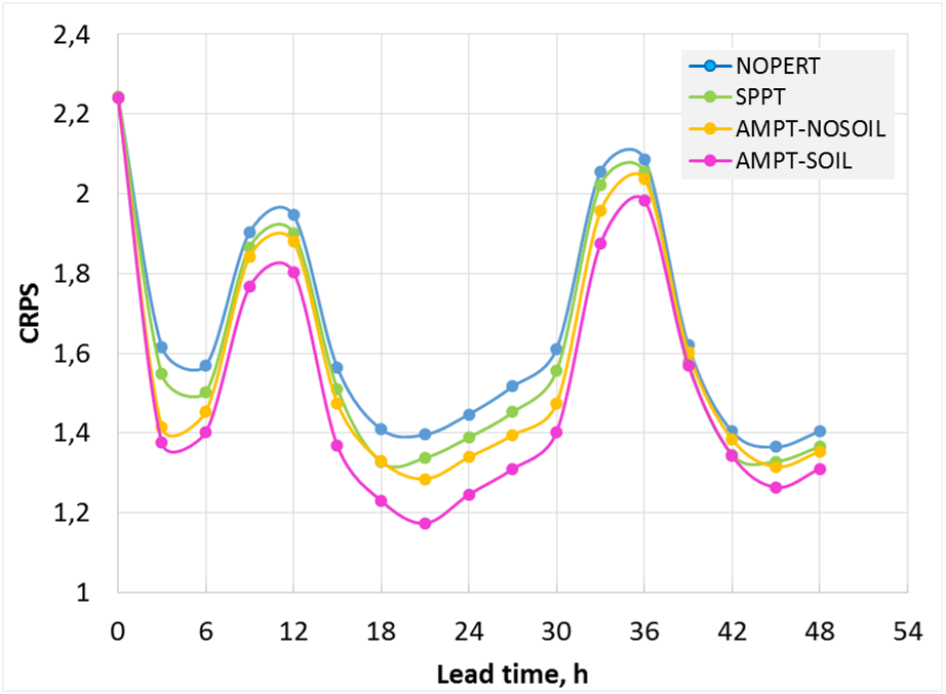

Figure 14: CRPS for $T_{2\mathrm{m}}$. The lower the better.

iments for May 2014, not presented in this report, suggest that AMPT humidity and cloud-fields perturbations can be important during periods with convective events.

4. For the same reason, we do not present here results of experiments with humidity perturbations. Only perturbations of temperature, wind, and soil were introduced in the above experiments.

Advantages and limitations of AMPT

Theoretically, potential advantages of AMPT w.r.t. SPPT are:

- AMPT generates *additive* model-error perturbations, which can represent errors in physical parametrization schemes when they produce an erroneously small physical tendency. SPPT cannot do that because its perturbations are multiplicative.
- AMPT can account for errors in the *relationships* between errors in different variables and at different model levels. In SPPT, these relationships are, effectively, assumed to be error free.

A potential disadvantage of AMPT w.r.t. SPPT is:

- AMPT produces, likely, less balanced perturbations than SPPT because the latter preserves the relationships between different variables and between different model levels. But model-error perturbations (imposed at every model time step) are very small and so are the imbalances. In our numerical experiments we found no evidence of problems related to the imbalances, unless the AMPT perturbations were "too big" (i.e. having an excessive magnitude and/or spatial or time scale). For "not too big" AMPT perturbations, the model was able to cope with the small imbalances without any indication of spinup-like behaviour.

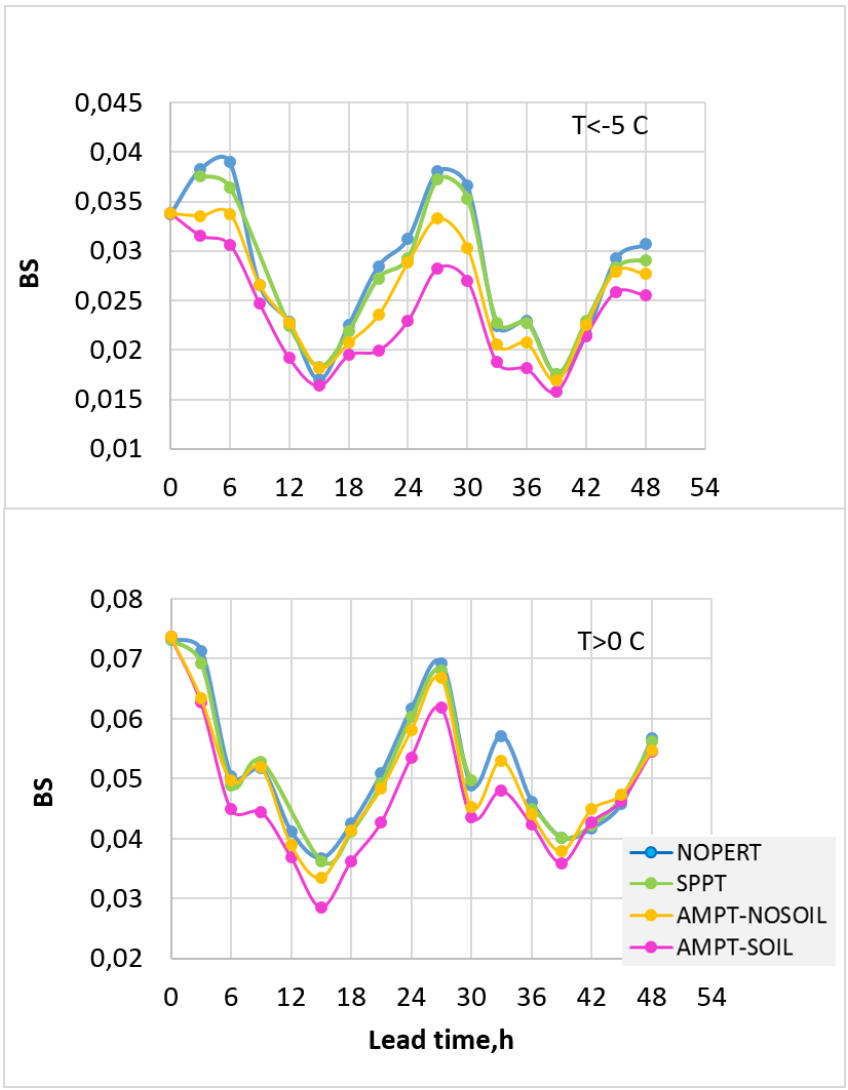

Figure 15: The same as Fig. 14 but for the Brier score for the events $T_{2\mathrm{m}} > 0$ C (the upper panel) and $T_{2\mathrm{m}} < -5$ C (the lower panel). The lower the better.

3.2.12 Conclusions on AMPT

- A new technique called Additive Model-error Perturbations scaled by physical Tendency (AMPT) is proposed. AMPT employs the Stochastic Pattern Generator (Tsyrulnikov et al., 2017) to generate four-dimensional random fields with tunable spatio-temporal correlations. The SPG random fields in AMPT are independent for different model fields. They are scaled by an area-averaged modulus of physical tendency and added to the model fields at every model time step. Gaussian and non-Gaussian fields are treated by AMPT.
- Three-dimensional model fields of temperature and winds were perturbed.
- Three-dimensional soil fields (temperature and moisture) are perturbed as well.
- In ensemble experiments, it was found that AMPT perturbations were very effective in generating spread in the ensemble.

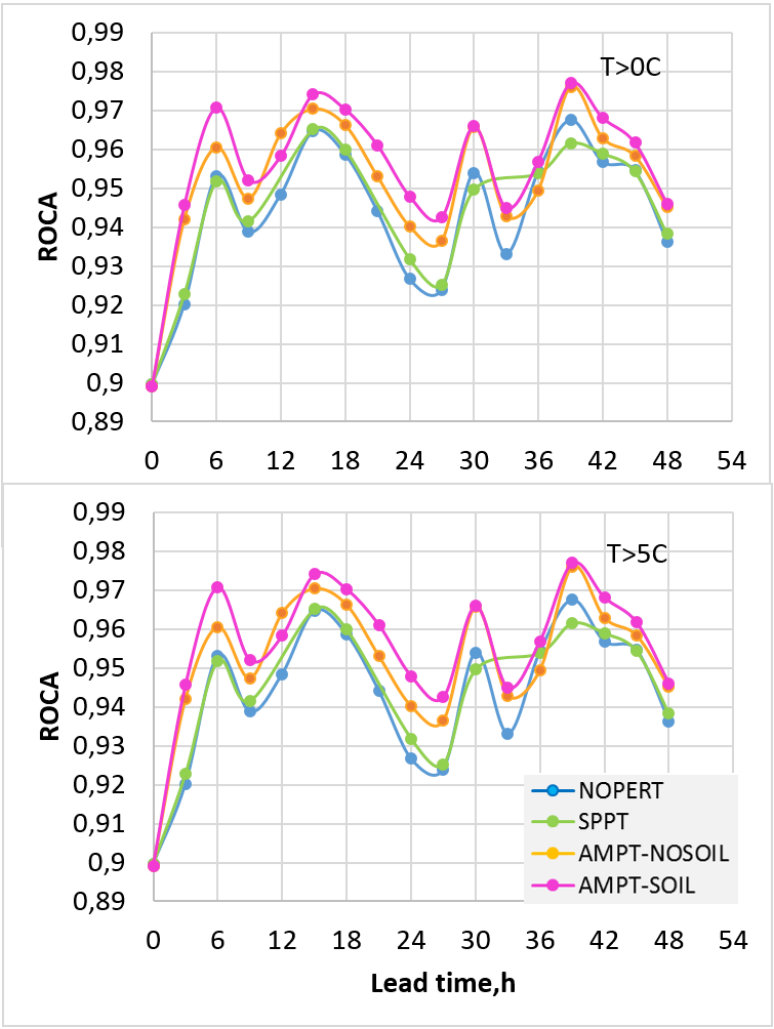

Figure 16: The same as Fig. 14 but for ROCA for $T_{2\mathrm{m}}$ for the events "$T > 0$ C" and "$T > 5$ C". The higher the better.

- Perturbing humidity and cloud fields generated significant spread in precipitation and upper-air temperature during precipitating events (not shown).
- Soil perturbations contributed significantly to the spread of near-surface fields.
- AMPT was able to reduce the average root-mean-square error (RMSE) of the ensemble mean w.r.t. no model perturbations and w.r.t. SPPT, however sometimes RMSE was degraded.
- Probabilistic ensemble verification scores (CRPS, Brier score, and ROC area) for near-surface temperature and wind speed were improved due to AMPT model perturbations in the atmosphere w.r.t. no model perturbations and w.r.t. SPPT.
- AMPT soil perturbations further improved the probabilistic verification scores.
- Probabilistic verification of precipitation gave mixed results, perhaps due to the insufficient number of precipitating events during the time period examined.

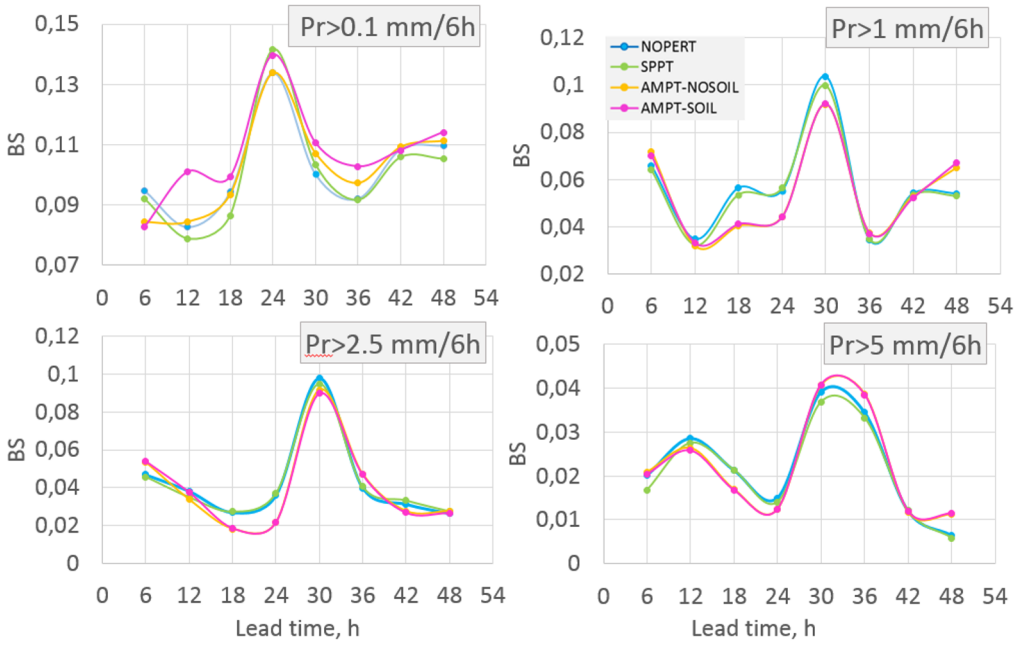

Figure 17: Brier score for total precipitation for the various thresholds indicated in the panels. The lower the better.

3.3 Model perturbations at IMGW (A. Mazur and G. Duniec, IMGW)

At IMGW, the basic instance of the COSMO model runs in a deterministic mode, using the initial (IC) and boundary (BC) conditions from the global ICON model. ICON generates a sets of IC/BC for 78-hour mesoscale COSMO forecasts. In turn, COSMO model with a horizontal resolution of 7 km (COSMO-7) uses nudging-based data assimilation to correct the forecasts of the global model by retrieving the latest data from the WMO network. The forecast results with a resolution of 7 km are further used as sets of IC/BCs for a nested COSMO model instance with a higher resolution of 2.8 km, and 48-hour forecasts. These deterministic COSMO-2.8 km forecasts define the basis for the operational configuration of the Ensemble Prediction System (EPS). In the final configuration of the operational EPS at IMGW, the set of twenty ensemble members is grouped into four packages, staggered in time by 6 hours from one another, obtained from previous model runs. Each package contains five members (Mazur and Duniec, 2015). All forecast (ensemble) results are then processed to provide the selected statistics (ensemble mean, spread, probability of exceedance of selected thresholds, etc.) in a graphical form. Model perturbations are added, by perturbing the parameter describing evaporation from the soil (c_soil).

3.3.1 Perturbations based on adapted Random Number Generator.

Computer-generated random numbers are called “pseudo-random” numbers as they are designed to mimic randomness but they are determined in a deterministic, though often very complex, way. Due to the ever-growing demand for good pseudo-random numbers, the literature describes many ways to generate them, and many people are still working on new methods. To generate numbers from the uniform distribution are used, among others, the following algorithms: Wichmann-Hill (Wichamnn and Hill, 1982), Marsaglia (Marsaglia,

2003), Knuth (Knuth, 1981). In general, a Random Number Generators (RNG) is a deterministic function. It uses the so-called “seed” that completely determines the values of the subsequent pseudorandom numbers. The seed is the value on the basis of which successive random numbers will be constructed. For a permanent generator and seed, identical random numbers will be generated regardless of the operating system or computer type. This property can be easily exploited. By controlling the selection of the seed, it is possible to obtain identical sequences of random numbers on different computers. In this way, one can repeat the simulation results, reproducing these results on other computers, or continuing calculations interrupted by some error. However, the basic selection of seed has its weaknesses. The seed on most computers is based on machine time (generally milliseconds). On high-speed machines, seeds can be identical for many, if not all, processes (threads). This problem is illustrated in the Fig. 18. More threads means more likelihood to have identical seeds. With an exemplary operating configuration of 400 threads versus 999 milliseconds, one will not get 999 ms (and, hence, 999 values of seed) on high-speed machines, but much less. This increases the probability that multiple threads (the blue diamonds) will have the same grain (red rectangles). The basics of the new RNG have already been presented in the framework of the Priority Project SPRED. Here, therefore, just to reminder is provided: the basic scheme of the new RNG assumes a sequence (in particular, even several stages) of generating seed values. That is, the first operation in a sequence generates a (pseudo-random) number which is the input to the next operation, and so on (see Fig. 18). Ultimately, the final seed value suitable for the RNG is obtained. At the expense of a slight extension in time, significantly randomized seed values are obtained.

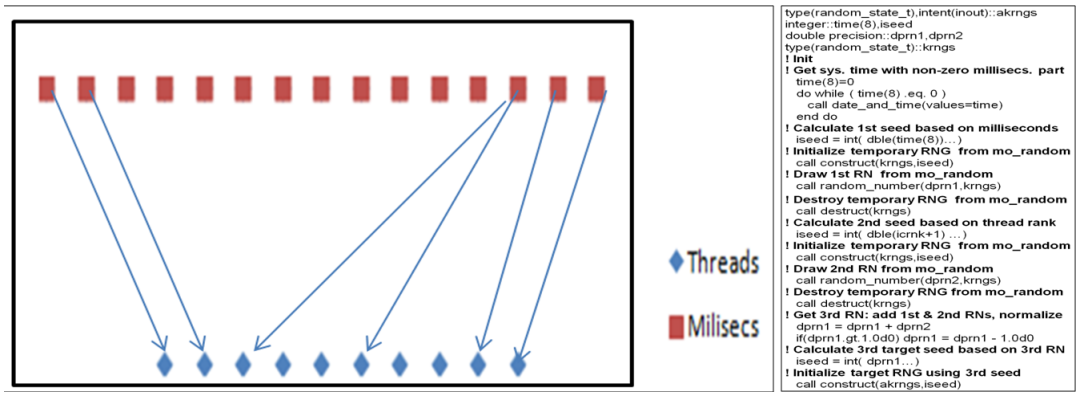

Figure 18: Left (a): Problem with parallel threads of stochastic computations - seeding RNG with machine time (milliseconds). Right (b): Block scheme of new seeding for RNG

Changing the RNG to the proposed one causes (may cause) a problem related to the loss of reproducibility of the results, mentioned in the introduction. However, from a mathematical point of view, the reproducibility requirement cannot be applied to single samples of partially or wholly nondeterministic phenomena, although it still applies to a probabilistic description of these phenomena, with the error tolerance defined by the probability theory. Apart from repeatability for technical or comparative tests, however, it is difficult to see too many arguments in favour of choosing permanent seeds. Perturbations based on an adapted Random Number Generator (RNG) were prepared for the years 2011-2015 (archive data).

This part of the report presents a comparison between a ”new” RNG vs. a standard one. Essentially, the ”new” RNG has been operating in operational mode since 2019, replacing the ”old” one. The modification was made when the perturbed parameter (c_soil) was defined in the code. The main reason for choosing where to introduce the random variation was code maintenance. By introducing variations in the original position (defining the field of the

disturbed parameter) on all nodes, it is possible to avoid excessive communication between the processors during the redistribution of the field. The use of machine time was due to the fact that initialization is performed only once (i.e. at the beginning of the model launch). The disturbed parameter is used only in the soil model (only once after the model started), and further during computations the distribution is no longer needed. The procedure does not repeat afterwards and does not affect efficiency in any way, but it seems to be one of the best sources of stochastic initiation. The results of the initial application of the new RNG were presented in PP SPRED, while further results of the RNG use tests are shown here. The influence of the RNG on the air temperature and dew point at 2m, wind speed at 10m and relative humidity at 2m is shown (Table 19 and Fig. 20, compared to values obtained with the standard "old" generator.

Spread values	Reference (standard RNG)	Modified RNG (operational)	Reference (standard RNG)	Modified RNG (operational)
	T2M (K)		TD2M (K)	
Mean	0.1218173	0.250191	0.13452	0.282912
Max	2.813	3.212	3.541	4.189
	U10M (m/s)		RH2M (%)	
Mean	0.17813	0.22381	0.965178	2.694192
Max	2.412	3.258	12.419	16.617

Figure 19: Average spread values for T2m and Td2m for years 2011-2015, modified RNG (left) and standard RNG (right). Tests with perturbations included, lagged members excluded.

Upon completion of the tests, the standard RNG was turned off. Currently, taking into account the quality from the new RNG, a new method of seed generation is operationally used for perturbations.

3.3.2 Perturbation of soil surface temperature in combination with upper air perturbation

The usage of other types of model perturbations has been tested and the results are here presented. Perturbations include soil surface temperature, coefficient of performance and combinations thereof. The following cases were considered:

a. operational set-up: perturbation of the parameter describing evaporation from the soil (c_soil);

b. eff-coeff: perturbation of the collection efficiency coefficient;

c. laf-pert: perturbation in soil surface temperature (initial values, with the amplitude of perturbation depending on the type of soil, see Duniec and Mazur, 2020);

d. combination of b) and c)

The general perturbation scheme have the following form:

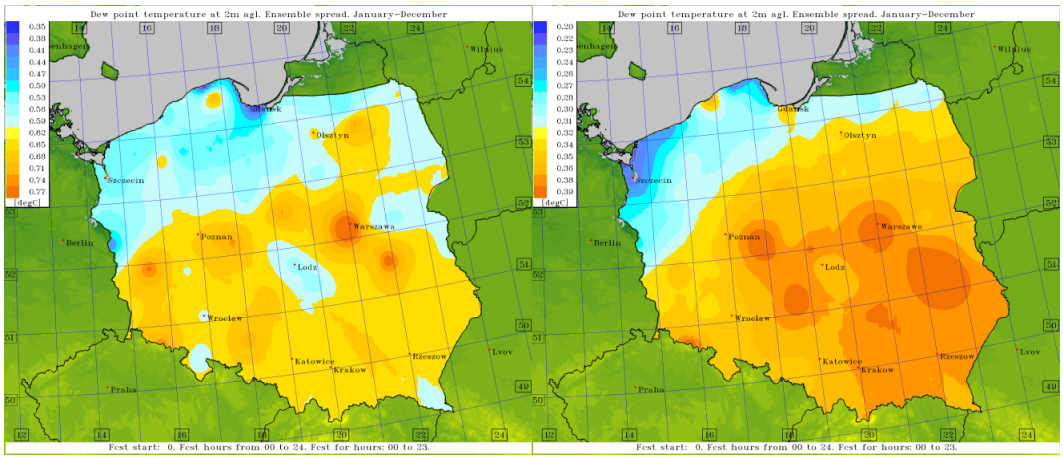

Figure 20: TD2M average spread values for years 2011-2015, modified RNG (left) and standard RNG (right). Tests with perturbations included, lagged members excluded.

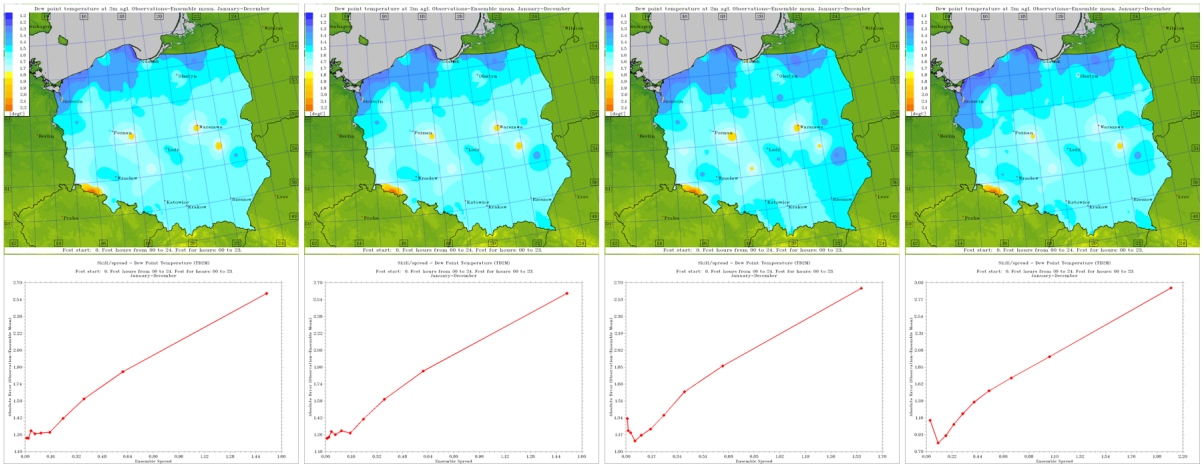

Figure 21: Upper panels: observations minus ensemble means, lower panels: spread-skill diagrams. Left to right are shown the 4 configurations of the ensemble: (a), (b), (c), (d) (see explanations in the text). Averaged values for the years 2011-2015.

perturbed_value=basic_value * [1+(random_number -0.5) * amplitude]

where the "random_number" (obtained from a computing dependent pseudorandom number generator, see section 3.3.1) varies from 0 to 1. In the laf-pert perturbation it is assumed that the ability to transfer heat and water is related to the compactness of the soil. If the soil is coarse (like sand), the amplitude of the disturbance is one, and the more compact the soil texture, such as peat or clay, the smaller is the disturbance amplitude. For example, for sand the amplitude was set to 1.0, for clay it was set to 0.6 and for peat it was set to 0.5. Thus, for sand, the perturbation ranged from 0.5 to 1.5 of the original (undisturbed) value. A climatological analysis for the warm period of a selected year was presented in Duniec and Mazur (2015). The overall results were also presented at the ICCARUS 2020 seminar. In Fig. 21 are reported only the most significant results, a comparison of different perturbation methods vs. measurements at SYNOP stations in Poland.

Research results essentially confirmed that:

1. the influence of the perturbation of the efficiency coefficient is, as expected, significant only in the case of precipitation;

2. the combination of different types of perturbations does not necessarily have a linear effect on the spread value and, moreover, on the quality of the forecasts. In some cases, the size of the spread produced by the perturbation combination was significantly greater than the algebraic sum of the spread produced by the perturbations applied separately;

3. compared to the operational EPS setup, new types of perturbation improve the quality of the forecasts in terms of skill and spread. The most effective perturbation scheme in terms of MAE and spread, averaged over the period 2011-2015, was the combination of perturbation types, with the mentioned restriction concerning point (1).

3.4 Perturbations based on SPPT with independent pattern (iSPPT) (André Walser and Linda Füzér, MCH)

The operational ensemble prediction system (EPS) of MeteoSwiss, COSMO-E, is underdispersive, i.e. the forecast shows a lack in the ensemble spread near surface, particularly in the short-range during the winter season. Stochastic perturbation of physical tendencies (SPPT) is operationally used in COSMO-E to represent model errors by perturbing the sum of the physical tendencies from the parametrization schemes. Although SPPT increases the spread significantly, there are still opportunities for further improvements. Christensen et al. (2017) suggest a modified version of SPPT called independent stochastically perturbed physical tendencies (iSPPT) which allows more flexibility in the perturbation setup and yields a better spread-error relationship in a global ensemble prediction system. For this reason, we have implemented iSPPT in the COSMO code in a test branch and investigated whether an iSPPT scheme can improve COSMO-E forecasts.

The results of the iSPPT experiments are shown in detail in Füzér (2018). In summary, iSPPT does not reveal a consistent improvement of the COSMO-E forecast spread-error relationship as compared to SPPT experiments. The ensemble mean error of the forecast is hardly sensitive to iSPPT, and a positive effect on the spread is found for wind speed and humidity and in some setups only while the spread for temperature could not be increased. The flexibility provided by iSPPT to perturb the individual physical tendencies differently is not helpful to achieve better scores for COSMO-E. Best results with iSPPT are achieved with the same random number setup as for SPPT applied to all physical tendencies.

3.5 Perturbations based on analysis increments (André Walser and Linda Füzér, MCH)

Perturbations from both SPPT and iSPPT do not account for model errors different than the ones arising from the parametrization schemes. For a global EPS, Piccolo et al. (2018) thus propose model perturbations using analysis increments (AIs) of the data assimilation cycle, as they are not random quantities but rather include possible sources of forecast errors. Such perturbations could also be beneficial for a limited-area EPS like COSMO-E. However, the characteristics of the AIs from the KENDA system, which provides the initial conditions for COSMO-E, has not been investigated yet and is thus unknown.

Füzér (2018) shows a detailed analysis of the AIs from the operational KENDA system of MeteoSwiss. The considered time period is 1st June 2017 to 31st August 2018 to cover all seasons and to be able to compare the summer of two different years. In summary, the AI

dataset only shows notable biases for summer temperature and humidity. Between 8 and 16 UTC, a negative temperature bias of up to -1.5 K can be found in the lowest model levels. During the other seasons and variables, regional biases are visible. Both the absolute mean AIs and standard deviation show clear variability among the seasons and have a diurnal cycle that also varies with the seasons. The magnitude and variability of the AIs increase throughout the course of the day and reach a maximum at 12 UTC and a minimum at night time. The most pronounced diurnal cycle is visible for summer, followed by spring and autumn. A comparison of summer 2017 and 2018 leads to the conclusion that the AIs analysed in this thesis are representative for the respective seasons because the two years show very similar characteristics although they were different in terms of meteorology. The comparison of AIs and model verifications against radiosondes reveals that the magnitude of the AIs is in good agreement with the model first guess mean error. Additionally, as expected, the AIs have opposite signs compared to the mean error of the FG ensemble mean.

The results of the analysis point towards a possible suitability of the AIs as model perturbations. There is a variability within the data, therefore the individual members of the ensemble prediction system would be perturbed differently and lead to an enhanced spread in the result. The AI fields should be grouped by season and time of the day and drawn from the corresponding group when the perturbations are applied. As the AIs are in good agreement with the bias of the FG ensemble mean, perturbations generated using AIs have the potential of reducing the model error and local biases. Before using the AIs as model perturbations, the outliers in the AI data set should be further investigated. It might be useful to set a range for the AIs in order to avoid disproportionally large perturbations.

Even though the study confirms that model perturbations with analysis increments have a large potential to be beneficial for a high-resolution EPS too, it has not been implemented and tested yet due to missing resources.

4 Post-processing and interpretation of ensembles

4.1 Calibration (A. Mazur, W. Interewicz, IMGW)

The development and application of calibration methods at IMGW were initiated in the previous Priority Project SPRED, and here only the results and conclusions from the continuation of the research and method(s) development are presented. The research essentially focused on three basic methods of determining the ensemble mean.

- The most common procedure for calculating the ensemble mean is to use the arithmetic mean of values obtained for all particular members, here called "Simple" Average. Values obtained from the members affect the ensemble mean with the same weight. The simplicity of the method implies that no additional procedures or assumptions are required.
- An alternative approach can be described as (multi-) linear regression (MLR), where pre-made predictions of all (X) EPS members are compared with the observations during a learning phase. The weights of each member are then calculated to judge how "important" the selected member should be in the overall EPS average (or how large a member's share in the EPS average should be). The least squares method used for the computation of the weights can also take into account all potentially important factors such as geographic coordinates, altitude, hour of forecast start, current forecast hour etc. An additional condition is the normalization of weights to one.

- Finally, an artificial neural network (ANN) was investigated to see if this approach could be a valid substitute for a Simple Average or MLR Mean. The basic idea is that the weights, according to the same scheme as in the MLR calculation, are calculated (learned) using the concept of the neuron layer(s), appropriately activated during the learning phase. This approach uses the commonly used ANN back-prop method combined with the activation function of a hyperbolic tangent. During the research (and later in the operational application of the method), a system with two layers of neurons (input and hidden) was used. The first (input) layer contains six neurons, the second, hidden layer contains two neurons. Such a scheme is due to the fact that the EPS creates four sets of five members, corresponding to four neurons, while the other two input neurons are related to spatial (geographic coordinates, terrain height) and time constants (forecasts leadtime). In the hidden layer, two neurons are assigned to variables (forecast members) and constants.

The effectiveness and superiority of the ANN method compared to the Simple Average and MLR were proven (Fig. 22). After the tests and this comparison the ANN method was implemented in an operational mode.

Another calibration technique applied and tested was the VOD (Vector Of Displacement) procedure. The method based on the space-lagged (cross-) correlation technique was used in this respect.

The basic principle of the method is based on the assumption that the forecast value field (including the ensemble mean forecast) can be spatially shifted relative to the actual value field, maintaining the spatial "shape". To determine the size of this shift (i.e. VOD - Vector Of Displacement), these steps are followed:

1. Calculate the "centre of mass" of the two-dimensional field of forecast values
2. Calculate the "centre of mass" of the two-dimensional field of measured values
3. Define VOD as the vector connecting these two points
4. "Shift" the forecast field by the VOD vector

An example of the result of using this procedure is shown in the figure 23.

As the number of SYNOP stations in Poland is not quite large (61), the procedure has been modified for the elements measured at the stations. Instead of calculating the "centres of mass" of continuous fields, the procedure is as follows:

1. at each SYNOP station, find the grid point (in a specific vicinity) with the forecast value closest to the one measured at station;
2. calculate the displacement vector for single station as the difference between the station location and the location of this point in the forecast field;
3. calculate an overall VOD as the average over all the stations
4. displace every value of forecasts by the vector of displacement

An example of the result of using this procedure is shown in figure 24.

After finding an improvement in the quality of the forecasts, this procedure was also applied operationally. In the operational mode, VOD is being determined at the beginning of each calculation cycle for each forecasted meteorological element. It is assumed, that this value does not change in the current calculation cycle, but it is updated at the next model run.

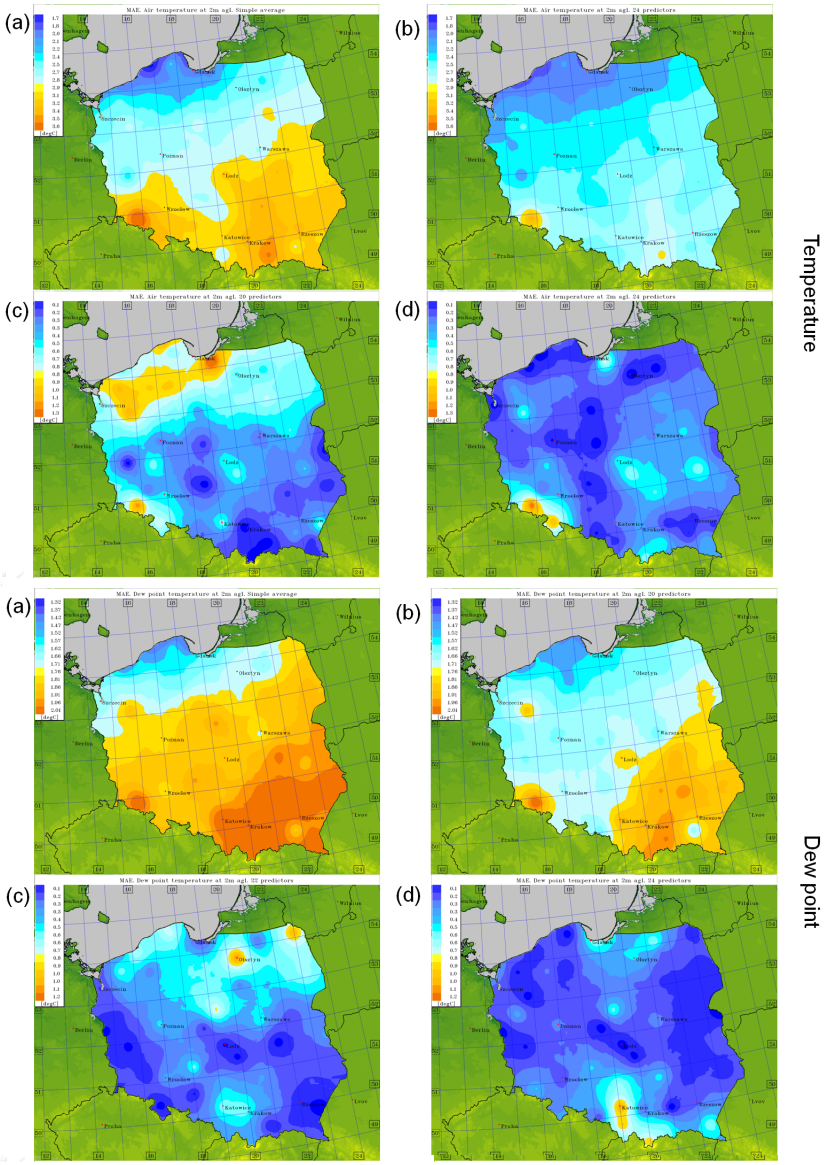

Figure 22: Skill comparison for different calibration methods, in terms of 2m air temperature (upper four panels) and dew point temperature (lower four panels). Panel (a) Simple average, (b) Multiline Regression (MLR), 24 predictors, (c) ANN, 20 predictors (ensemble values only), (d) ANN, 24 predictors (ensemble values + spatiotemporal coordinates). Results are for the years 2011-2015.

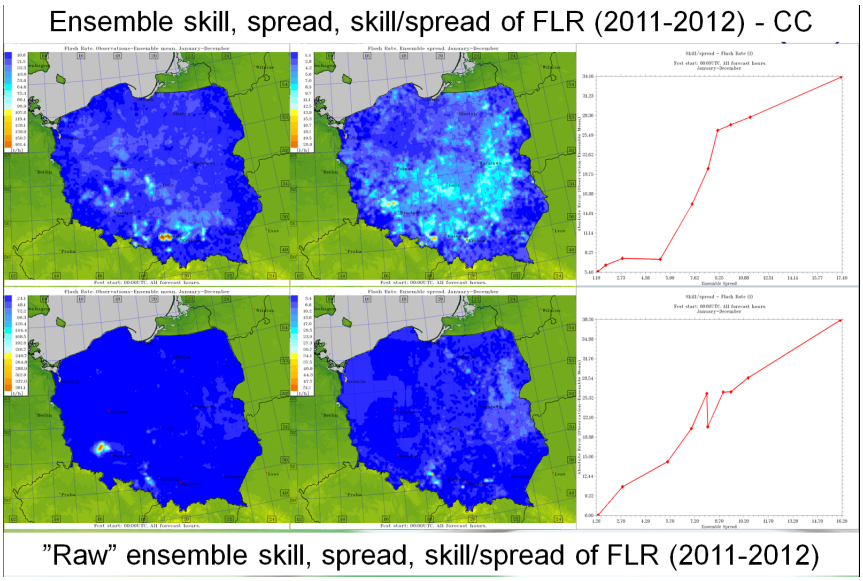

Figure 23: 2011-2012 average skill, spread and skill-spread relation for Flashrate (FLR) frequencies using the VOD procedure (upper panel) against the "raw" scores (lower panel).

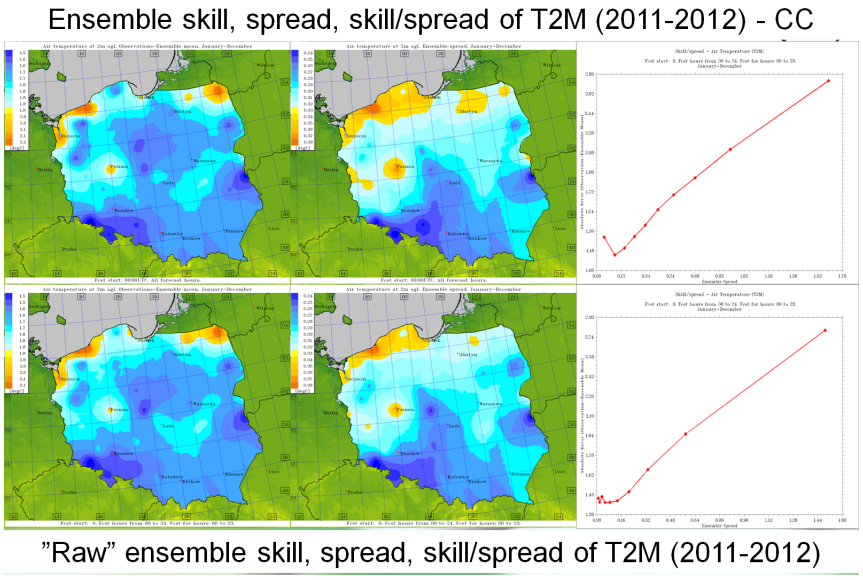

Figure 24: Same as Fig. 23 but for air temperature at 2m.

4.2 Specific products from ensemble outputs (R. Golino, F. Marcucci, COMET)

In the framework of the EUMETNET SRNWP-EPS project, a tool for thunderstorm forecasting has been developed. The algorithm combines different indices following the decision tree in Fig. 25 and evaluates for each grid point the suitable weather phenomenon according to the WMO table 4678.

The most important indices used in the code are: Cape, Cin, Total total index (TT), Showalter Index (SI), Updraft Helicy (UH), Lighting Potential Index (LPI), Integrated Graupel (QG), Cloud Physics Thunder Parameter (CPTP). Indices can be computed internally if

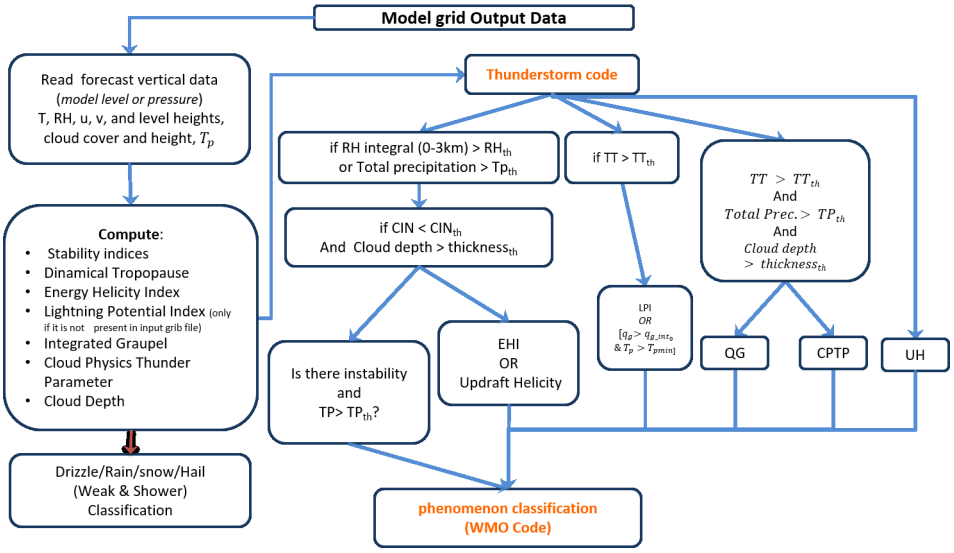

Figure 25: A schematic overview of the algorithm decision tree to evaluate the phenomenon assigned to a specific grid point.

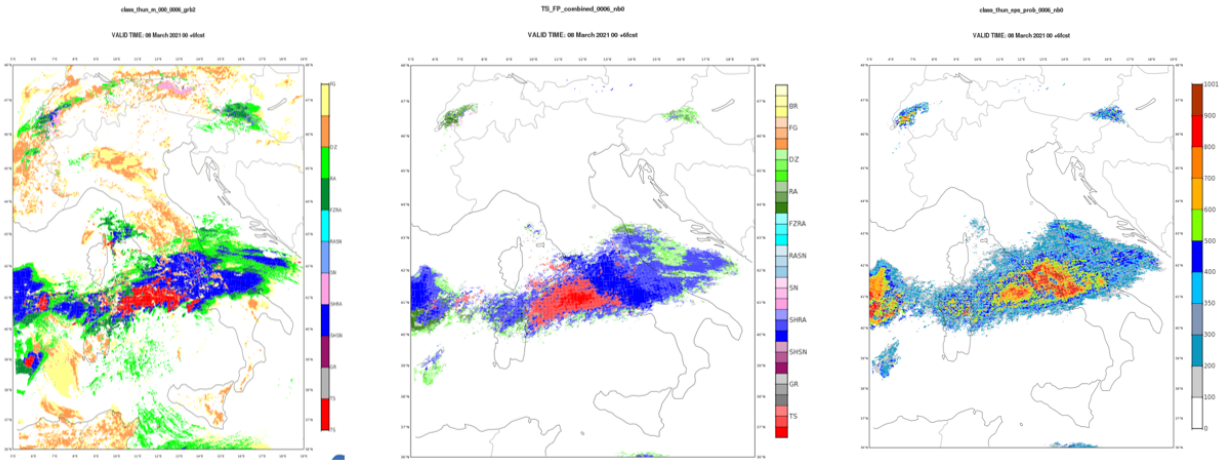

Figure 26: Output of the algorithm applied to deterministic COSMO-IT (left). Most probable phenomenon computed using the 20 outputs of the algorithm applied to the COSMO-IT EPS ensemble members (centre) and related probability (right). All pictures are referred to the 00UTC run of 08 March 2021, forecast step +6h.

not available as model output. The algorithm has been applied to deterministic outputs of COSMO-IT model and also to the 20 forecast members of the ensemble convection permitting system COSMO-IT EPS. In order to have a synthetic picture of the scenario predicted by the ensemble, the EPS outputs can be post-processed to obtain a single field (grib format) that represents the most probable phenomenon. For each grid point, the phenomena predicted by all the members are collected and clustered, then the most populated cluster is considered to decide the suitable phenomenon. The associated probability is computed taking also into account events from the most severe group. Low-probability phenomena (less than 20%) are discarded. An example of the algorithms output from different models (deterministic COSMO-IT and ensemble COSMO-IT EPS) is given in Fig. 26. In the ensemble outputs (two rightmost pictures) fog/mist events (BR/FG - yellow/orange in the colorbar) are not

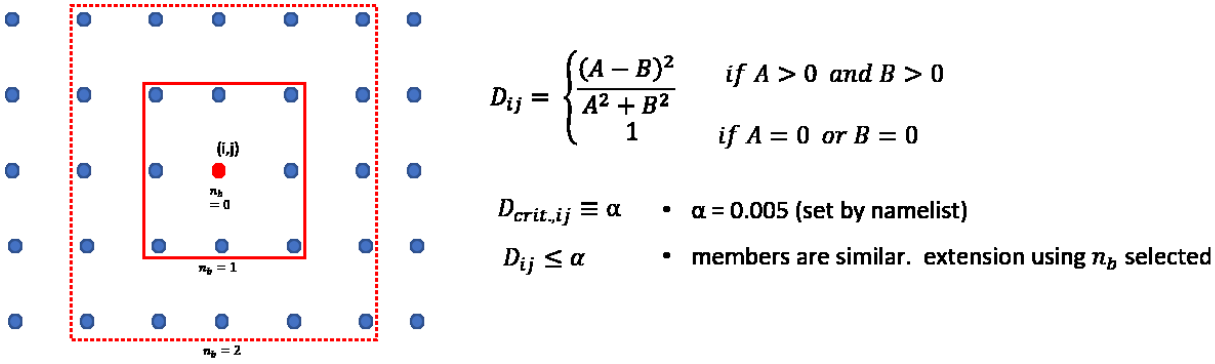

Figure 27: A schematic overview of the method proposed by Dey et al (2016).

considered. To enlarge the sample, a nearest neighbour post processing is applied using the similarity criterion proposed by Dey et al (2016). Namely, for each single grid point (i,j) the values of the EPS members are considered in pairs (190 independent comparison with 20 members) in order to assess if there is an agreement among the ensemble members. The value Dij is computed for all the pairs as shown in Fig. 27, where A and B are the values of the 2 chosen members at the grid point (i,j). If the averaged value Dij is greater than a given parameter (= 0.05 by default), there is no similarity and the agreement is further verified increasing the scale (stopping at nb-order of neighbourhood, 1 by default resulting in 3x3 box and 9x20 values to be compared). Finally, the most representative phenomenon is assigned to the (i,j) grid point. The similarity criterion, currently under evaluation, is expected to be able to filter out showers and drizzle if more significant weather is predicted by other members/neighbours, thus producing a more realistic output.

4.3 Specific products from ensemble outputs (A. Mazur, G. Duniec, IMGW)

Flashrate (as a storm indicator) and visibility range or visibility limitations (caused by fog, but also precipitation) are derived as non-standard products from the EPS.

Forecasts of visibility range were calculated from direct model output using an algorithm based on the forecast of the extinction parameter ext (calculated as a function of water/ice amount in the air, see Stoelinga and Warner, 1999, or Gultepe et al., 2006):

$$VIS = \frac{-\ln(0.02)}{\beta_{ext}}$$

Figure 28: Visibility range from extinction parameter.

These forecasts were verified against measurements at Polish SYNOP stations (see Fig. 29.

Overall conclusions drawn from this research were as follows:

- the average spread is two to ten times lower than the error measured as MAE: the ensemble FVIS is strongly under-dispersive;
- as far ad the spatial relations are concerned, the error is smaller for central and southern Poland, while spread is bigger in the southern and northern part of Poland;

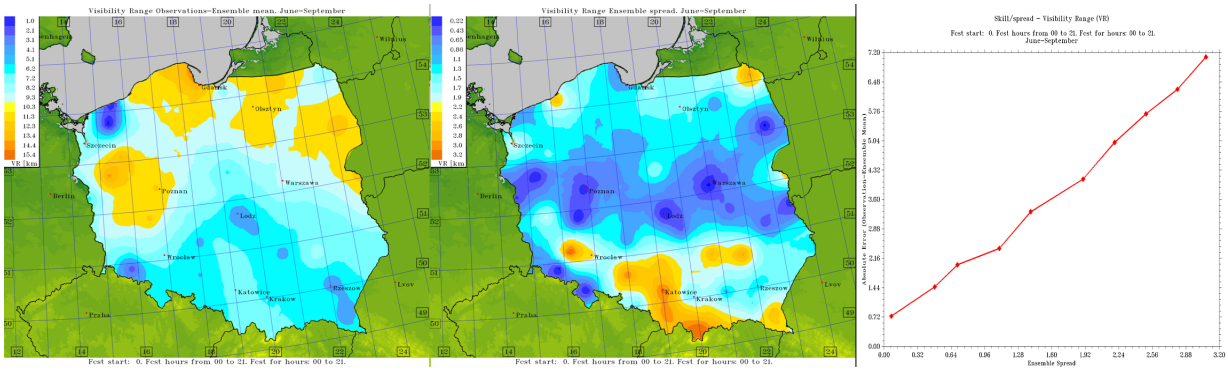

Figure 29: An example of skill (left), spread (middle) and skill-spread relation (right) of the VIS forecasts. Averaged values for June to September, 2011-2015.

- skill and spread are in general better (i.e., smaller and bigger, respectively) for "dry" months (June and September).

In the FLR calculation it is assumed that there is a relationship between the CAPE (from which the maximum updraft velocity W is calculated), the cloud top/base temperatures (CTT/CBT respectively) and the lightning frequency (FR, 1/minutes):

$$W = 0.3 \cdot \sqrt{2 \cdot CAPE}$$

$$FR = \left(\frac{W}{14.66} \right)^{4.54}$$

$$if \quad CTT > -15°C \quad FR = FR \cdot \left[\max\left(\frac{-CTT}{15}, 0.01 \right) \right]$$

$$if \quad CBT < -5°C \quad FR = FR \cdot \left[\max\left(\frac{CBT + 15}{10}, 0.01 \right) \right]$$

Figure 30: Flashrate forecasts calculated from CAPE and cloud top/base temperatures.

Additional filters were used to improve the quality of forecasts (see e.g. Wong et al., 2013).

Forecasts were verified against measurements at Polish lightning detection network PERUN (Fig. 31).

A general conclusion drawn from this part of the study was that the CAPE-based forecasts of flashrate intensity produce good skill values for central and north-western Poland, with the spread being bigger in central and western Poland. This algorithm, however, significantly overestimates flashrate intensity. A precipitation filter is applied, that is, it is assumed that flashrate intensity is zero in the absence of rainfall.

The Universal Tornadic Index (UTI, Taszarek and Kolendowicz, 2013) was a possible candidate for thunderstorm recognition and forecasts, to be verified against data from the lightning detection network:

In the above formula, CAPE is the surface based Convective Available Potential Energy, CAPE3km the Surface based CAPE released below 3 km, LCL the Surface based lifting condensation level height, AMR500 the Average mixing ratio below 500 m, LLS the 0-1 km wind shear, DLS the 0-6 km wind shear (magnitude of vector difference) and SRH1km the 0-1 km storm relative helicity. It is also assumed that SRH1kmis non-negative, and if

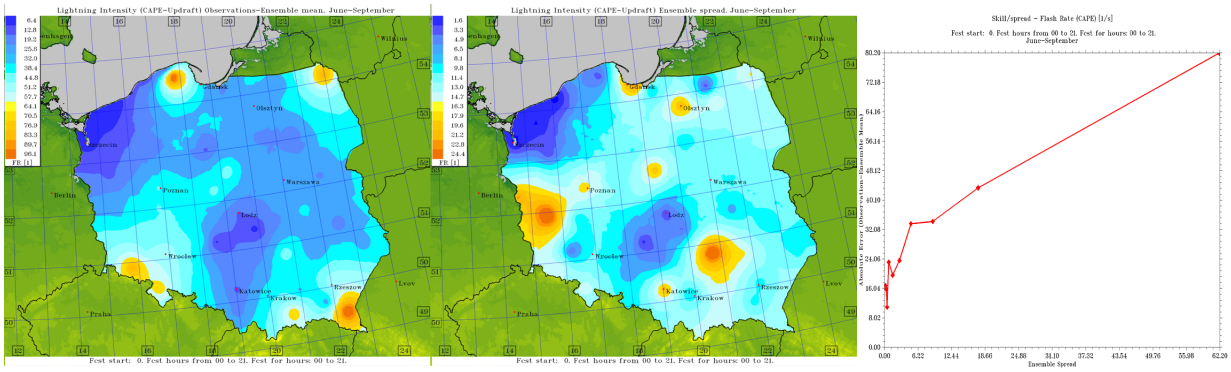

Figure 31: An example of skill (left), spread (middle) and skill-spread relation (right) of the FLR forecasts. Averaged values for June to September, 2011-2015.

$$UTI = \frac{CAPE \cdot SRH_{1km}}{200} \cdot \frac{5 \cdot (DLS - 20) + \left(\frac{200 - LCL}{10}\right)}{100} + CAPE_{3km} + \frac{SRH_{1km}}{4} \cdot \frac{LLS}{12} \cdot \frac{AMR_{500}}{100}$$

Figure 32: Universal Tornadic Index as a possible indicator of thunderstorm occurrence.

LCL is greater than 1500m or CAPE is equal to zero or convection precipitation is less than 2mm/h, then UTI becomes zero. UTI increases with the probability for a significant tornado. For Poland, the operational significance of UTI should attract attention when greater than 0.5-1.0. It should however be kept in mind, that these (climatologically obtained) values do not guarantee that if a forecaster uses threshold values in real-time data, he will be able to specify whether a tornado will occur. Results rather suggest that tornadoes can (may) be formed at all. Universal Tornadic Index was assumed as an indicator of the occurrence of a extreme convective phenomenon. Since it contains many factors (CAPE, storm relative helicity, convective precipitation, wind shear etc.), it may be useful in forecasting not only tornadoes, but also thunderstorms or squall lines. The ensemble mean value of the UTI is comparable to the one calculated in a deterministic run, however significantly less amount of "noise signal" is observed. Further research was carried out to determine any climatological connection between UTI and FLR values. Still, this index did not meet the expectations, in the sense that no significant correlation of this kind was detected. It was perhaps due to a different spatial range of the predicted convective phenomena, namely tornadoes (basic scale of tens to hundreds of meters) vs. "regular" thunderstorms (hundreds of meters to few kilometres).

5 Transition to ICON-LAM

The COSMO Consortium is moving towards the model ICON. This means that the convection-permitting ensemble systems must be converted to the ICON-LAM model. In order to be able to perform a transition to ICON-LAM in the current ensemble configurations, two ingredients are needed: (i) data assimilation for ICON-LAM based on LETKF and (ii) a model perturbation method for ICON-LAM. The first issue is addressed by the data assimilation projects and plans, the second is addressed in the present PP.

For tackling the issue of model perturbation with ICON-LAM, in addition to the technical

adaptations, it has to be examined whether the previous approaches to the representation of the model uncertainties can also be used with ICON-LAM. Due to the change in the characteristics of the model, modifications of the applied methods are likely to be required. The migration of the CP ensemble systems to ICON-LAM will require: forecast experiments for different test periods to explore the properties of the new systems, identification of further development requirements, adaptation of ensemble generation methods.

During the APSU Priority Project, all the ensembles were still running based on the COSMO model, and the model perturbation methods were developed in COSMO. The only exception was DWD, where the transition was more advanced. Here, experiments on the Parameter Perturbation in ICON-LAM (ICON-D2-EPS configuration) were carried out. This work is described in the next subsection.

At the time of writing, a new Priority Project on ensemble development is ongoing (PROPHECY), where model perturbation methods are developed, implemented and tested in ICON in several COSMO members.

5.1 Randomly perturbed parameters in ICON-D2-EPS (C. Gebhardt, DWD)

Coinciding with the lifetime of PP APSU, the ICON-D2-EPS has been prepared as operational EPS application of the ICON limited area mode at DWD. ICON-D2-EPS replaces COSMO-D2-EPS as ensemble prediction system on the convection-permitting scale.
On the one hand, the general methodological approach for member generation in COSMO-D2-EPS is transferred to ICON-D2-EPS, i.e. local ensemble transform Kalman filter for the initial conditions, forecasts of ICON-EU-EPS as lateral boundaries, and randomly perturbed parameters representing model uncertainties. The EPS set-up is identical for the first phase after the transition, with 20 members running every 3 hours for up to 27 hours lead time (45 hours for the 03 UTC run) on a similar domain with a grid spacing of ca. 2.1km.
On the other hand, some of these methods require adjustments to fit to the characteristics of the ICON model in a beneficial manner. Within the PP APSU, the adjustments and the effects of the method of randomly perturbed parameters (RPP) in ICON-D2-EPS are of particular interest and can provide valuable information for the COSMO community. This section reports about the adjustments made to both the range of perturbed values of certain parameters as well as to the technical implementation of the perturbation approach. This implementation is now fully controlled via the ICON namelists and applied within the model. There is no need for preparatory processing scripts in contrast to the procedure for COSMO-D2-EPS.
The general approach remains the same, i.e. for each parameter to be perturbed a value is selected randomly from the default and one or two pre-defined perturbed values in each member. This selection is done at forecast start and the selected parameter values are fixed for the entire forecast range of an EPS run. This selection from a *limited number* of possible values is different from approaches with a random choice of *any* parameter value within the given range.

Most aspects of the physics parametrizations do not substantially change between COSMO-D2 and ICON-D2, which means that many parameters of RPP for COSMO-D2-EPS can be perturbed analogously in ICON-D2-EPS, subject to minor re-tuning. Some parameters have to be adjusted to changes in the parametrization (i.e. the dependence of the vertical diffusion on the Richardson number affects the appropriate range of values for `tkhmin` / `tkmmin`). There are a few replacements (i.e. `radqc_fact` vs. `box_liq` for radiation/cloud cover) and some new parameters are added. More details are listed in Table 2 together with the default values and perturbation ranges for ICON-D2-EPS resulting from preparatory

sensitivity studies.

For the sensitivity studies, 14 days of deterministic ICON-D2 00 UTC runs are performed for each potential perturbation of a parameter with only this perturbation applied for the entire forecast range of 24 hours and all other parameters set to the default value. This procedure is, of course, different from the application in the EPS with its daily randomized mixture of perturbations, but it turned out to be very efficient in detecting which perturbations have a potential for leading to outliers in terms of verification scores. The sensitivity studies aim at a selection of parameter perturbations leading to similar forecast quality without outliers (neither 'better' nor 'worse'). The starting point for the perturbation ranges for ICON-D2 are based on experiences with perturbations for COSMO-D2-EPS (acting on the same scale) and for the global ICON-EPS (using the same model). Note that differences in the scale can lead to differences in the appropriate perturbation ranges between ICON global/EU and ICON-D2. Fig. 33 exemplifies the results of the sensitivity study for the initial set of perturbations for selected variables and for mean error only (other scores were tested as well but they are not shown).

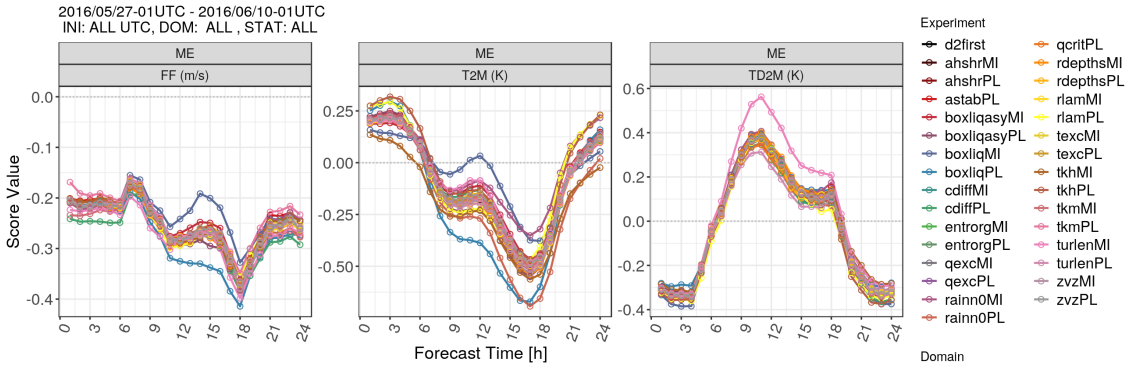

Figure 33: Mean error of 10m wind speed (left), 2m temperature (middle), and 2m dewpoint temperature (right) as a function of the forecast lead time averaged over 14 days of 00 UTC deterministic ICON-D2 runs. Each coloured line corresponds to exactly one parameter being perturbed in just one direction (see legend with 'PL' for 'increased' and 'MI' for 'decreased') keeping all other parameters at their default values. The fully unperturbed run ('d2first') is contained as well but overlaid by the other runs.

Most of the mean errors are very similar among the different runs for the three variables 10m wind speed, 2m temperature, and 2m dewpoint temperature. Just a few parameter perturbations are considered to produce outliers, e.g. the perturbation in both directions of `box_liq` (blueish lines) for wind speed and 2m temperature and `rain_n0_factor` for 2m temperature or the reduction of `tur_len` for the 2m dewpoint temperature. The parameter perturbations pointing at possible outliers with this quite conservative test have been adjusted for the first experiments with ICON-D2-EPS and table 2 lists the accordingly selected perturbation set-up.

Note that these perturbation ranges are subject to change in light of further developments of the parametrizations and of the ICON-D2-EPS. An up-to-date list of perturbations used in operational set-up can be found online in the ICON database reference manual (Reinert et al., 2021, latest version).

In the following, verification results of the pre-operational ICON-D2-EPS in comparison to the operational COSMO-D2-EPS are presented for a winter (December 2019 to February 2020) and summer (June to August 2020) period. Fig. 34 shows verification scores of the winter period for 2m temperature, 2m dewpoint temperature, hourly precipitation, and

CD2	ID2	meaning	def-	−	+
a_stab	a_stab	stability correction of turbulent length scale	0	none	1
c_diff	c_diff	length scale factor for vertical diffusion of TKE	0.2	0.1	0.4
q_crit	q_crit	critical value for normalized super-saturation	2	none	3.5
tur_len	tur_len	asymptotic maximal turbulent distance	300	250	350
rlam_heat	rlam_heat	scaling factor of the laminar boundary layer for heat (rlam_heat × rat_sea = 8.)	10	7.5	12.5
thick_sc	tune_rdepths	maximum allowed depth of shallow convection	2 $*10^4$	1.5 $*10^4$	2.5 $*10^4$
entr_sc	tune_entrorg	Entrainment parameter valid for dx=20 km	1.95 $*10^{-3}$	1.75 $*10^{-3}$	2.15 $*10^{-3}$
radqc_fact	tune_box_liq tune_box_liq_asy	box width for liquid cloud diagnostic in cloud cover scheme & asymmetry factor	0.05 3.25	0.05 3.0	0.06 3.5
radqi_fact	none	-	-	-	-
tkhmin	tkhmin	scaling factor for minimum vertical diffusion coefficient for heat and moisture	0.5	0.4	0.6
tkmmin	tkmmin	... for momentum	0.75	0.6	0.9
none	tkred_sfc	reduction of minimum diffusion coefficients near the surface	1	0.25	4.0
none	rain_n0_factor	tuning factor for intercept	0.1	0.05	0.2
		parameter of raindrop size distribution			
none	a_hshr	length scale factor for the separated horizontal shear mode	2	1	3
none	tune_zvz0i	terminal fall velocity of ice	1.25	1	1.5
none	tune_texc	excess value for temperature used in test parcel ascent	0.125	0.075	0.175
none	tune_qexc	excess fraction of grid-scale QV used in test parcel ascent	0.125	0.075	0.175

Table 2: Perturbed parameters in COSMO-D2-EPS (**CD2**) and in first experiments with ICON-D2-EPS (**ID2**). The default (**def.**) and perturbed values (decreased: −; increased +) refer to ICON-D2-EPS. The updates to operationally used parameters, default values, and perturbation ranges in ICON-D2-EPS are available online (Reinert et al., 2021) together with information about the perturbation being additive or multiplicative for a given parameter (e.g. rlam_heat changes from multiplicative to additive with ICON-D2-EPS).

hourly 10m wind gusts as a function of the forecast lead time for 00 and 12 UTC runs combined.

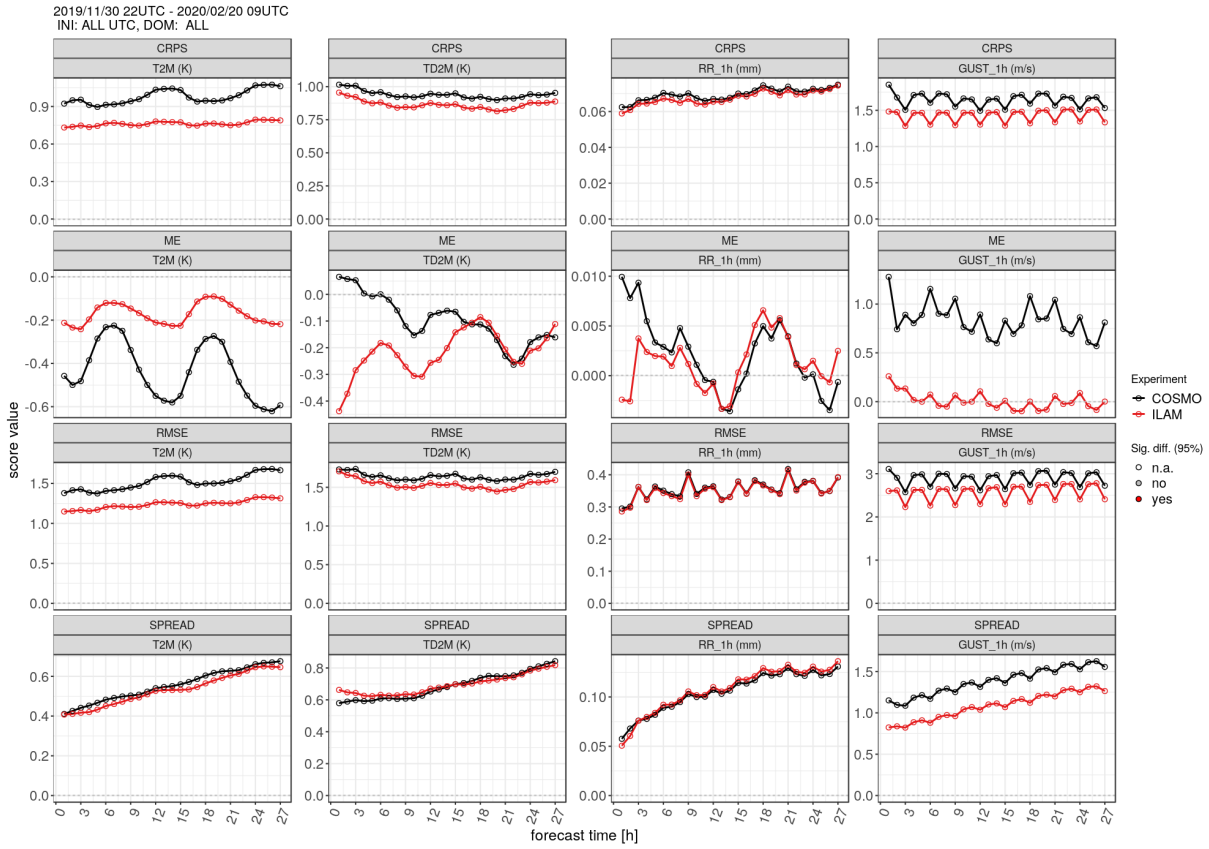

Figure 34: Continuous ranked probability score (CRPS), mean error and root mean squared error of the EPS mean, and EPS spread (from top to bottom) for 2m temperature, 2m dewpoint temperature, hourly precipitation, and hourly 10m wind gusts (from left to right) as a function of the forecast lead time of the pre-operational ICON-D2-EPS (red) and the operational COSMO-D2-EPS (black) for winter 2019/2020. Scores of 00 and 12 UTC runs are combined.

Comparing ICON-D2-EPS with COSMO-D2-EPS an improvement of CRPS is evident for all variables. The mean error for 2m temperature is clearly improved both in terms of magnitude as well as reduced daily cycle. The reduction of the daily cycle of the bias supports the beneficial use of ICON-D2 compared to COSMO-D2 in forecasting. It is observed for other variables as well (not shown, e.g. 10m wind speed). The dewpoint temperature shows a dry bias at forecast start of ICON-D2-EPS compared to COSMO-D2-EPS which extends up to forecast hour 15. This problem has been overcome in the further development of the system by different measures, of which the most important is the incorporation of 2m temperature and 2m relative humidity observations in the data assimilation.
The bias of the hourly precipitation in the early forecast hours is reduced in ICON-D2-EPS. Very remarkable is the reduction of the mean error for 10m wind gusts to nearly 0 m/s over most parts of the forecast range.
The RMSE of the EPS mean is reduced by ICON-D2 for most variables and neutral otherwise (hourly precipitation) and is similar to changes in CRPS. It should be noted here, that changes in mean error and RMSE for winter and summer for the variables shown here, as well as for other variables, closely follow the changes of the deterministic run from COSMO-D2 to ICON-D2 (not shown).

The EPS spread does hardly change from COSMO-D2-EPS to ICON-D2-EPS except for a reduction for the 10m wind gusts (as well as mean wind, not shown). The reason for this reduced spread is not clear yet and subject to further research. Overall, the threshold-based forecast of 10m wind gusts benefits from the step to ICON-D2-EPS according to the Brier skill score plotted in Fig. 35 for different thresholds. The improvement affects both the reliability as well as the resolution component of the Brier score (not shown).

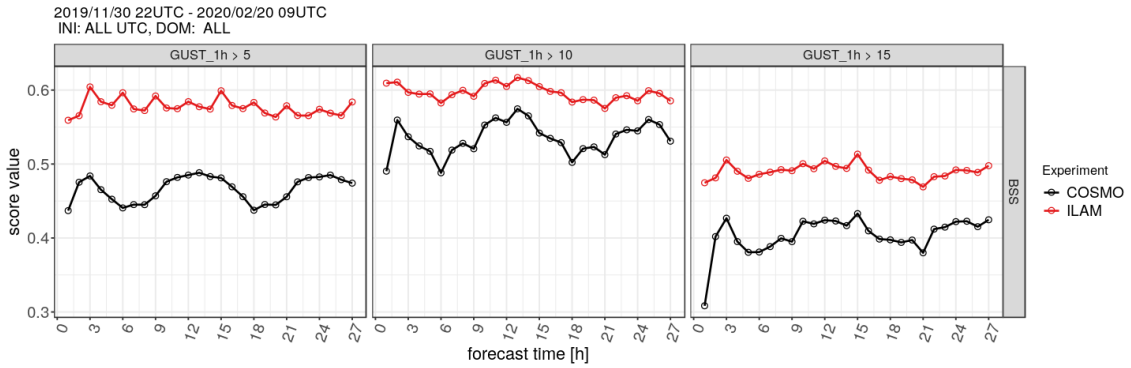

Figure 35: Brier skill score of hourly 10m wind gusts for thresholds of 5 m/s, 10 m/s, and 15 m/s (from left to right) as a function of the forecast lead time for the pre-operational ICON-D2-EPS (red) and the operational COSMO-D2-EPS (black) for winter 2019/2020. Scores of 00 and 12 UTC runs are combined.

The selection of verification scores for the summer period is shown in Fig. 36 in analogy to the scores for the winter period.

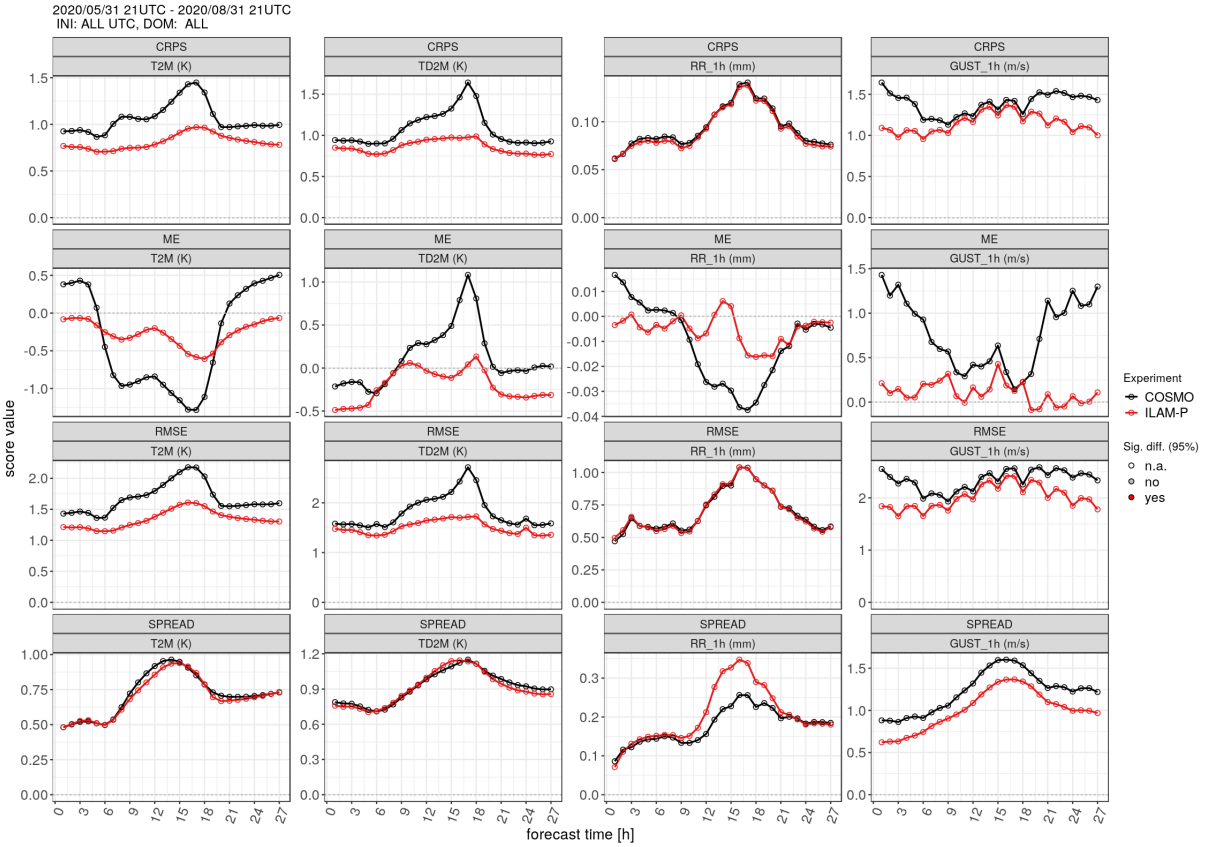

Figure 36: Continuous ranked probability score, mean error and root mean squared error of the EPS mean, and EPS spread (from top to bottom) for 2m temperature, 2m dewpoint temperature, hourly precipitation, and hourly 10m wind gusts (from left to right) as a function of the forecast lead time of the pre-operational ICON-D2-EPS (red) and the operational COSMO-D2-EPS (black) for summer 2020. Scores of 00 and 12 UTC runs are combined.

The main features seen for winter also hold for the summer period. Main differences are the remarkable quality gain by ICON-D2-EPS in the mean error of 2m dewpoint temperature and hourly precipitation and the more pronounced reduction of RMSE for 2m (dewpoint) temperature around mid-day as compared to the winter period. For 10m wind gusts, the quality of ICON-D2-EPS in summer and winter is similar. This in turn results in a moderate improvement in summer around mid-day compared to COSMO-D2-EPS, which itself shows better scores for the wind gusts in summer than in winter. The improvement in Brier skill score by ICON-D2-EPS in summer is qualitatively similar to the results for winter and is not shown here.
One effect to mention here is the increased spread of hourly precipitation during mid-day in summer which is not observed for winter. However, it should be kept in mind that increased spread of precipitation tends to be linked to increased mean values. A threshold-based verification with the Brier skill score for hourly precipitation in summer points at moderate improvements when using ICON-D2-EPS (Fig. 37).

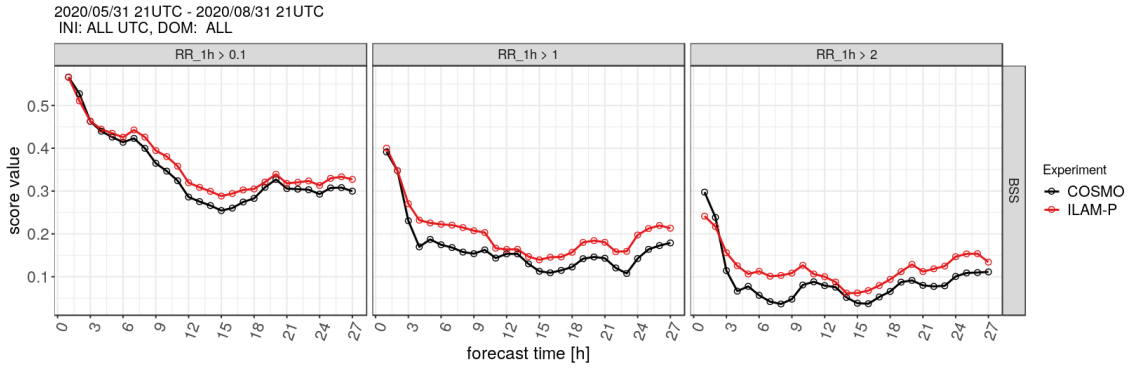

Figure 37: Brier skill score of hourly precipitation for the thresholds 0.1 mm/h, 1 mm/h, and 2 mm/h (from left to right) as a function of the forecast lead time for the pre-operational ICON-D2-EPS (red) and the operational COSMO-D2-EPS (black) for summer 2020. Scores of 00 and 12 UTC runs are combined.

The last part of this section covers the technical implementation of the parameter perturbations in ICON-D2-EPS. While the perturbation approach for COSMO-D2-EPS requires the perturbed parameter values to be set explicitly in the namelist for each member separately as a pre-processing step, including the randomization, the ICON-D2-EPS just needs general information about the parameters to be perturbed and the relevant ranges provided via the namelist. The randomization is done during the model run and there are additional namelist parameters to control a few options of the randomization and parameter selection.
Fig. 38 is a flow chart of the procedure.

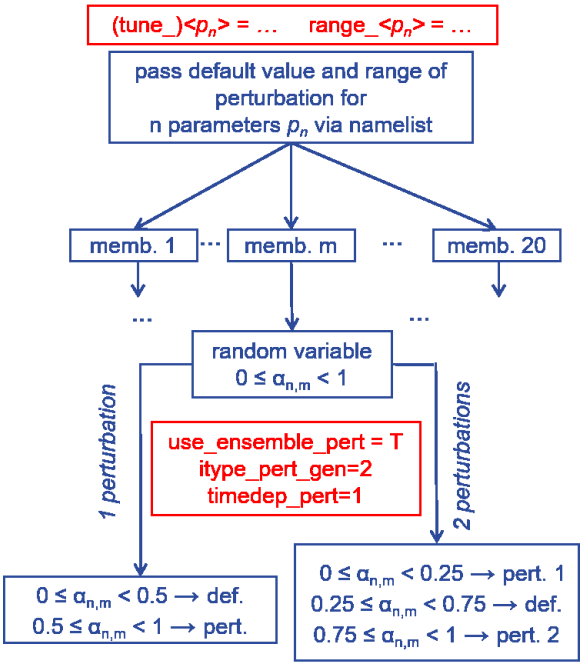

Figure 38: Technical implementation of the parameter perturbations in ICON-D2-EPS.

The default values of the parameters (tune_)$< p_n >$ are set in their respective namelists as for the deterministic run. A range for perturbations is set as (range_)$< p_n >$ in the ICON namelist `ensemble_pert_nml`. There is only one entry for the range which has to be given

as deviation from the default value. Therefore, only symmetric perturbations (in an additive or multiplicative sense depending on p_n) are possible in ICON-D2-EPS. There are also restrictions for a few parameters to be perturbed only in one direction.
The namelist `ensemble_pert_nml` contains the important switch `use_ensemble_pert` to be set to `.TRUE.` to activate the parameter perturbation. A random variable $\alpha_{n,m}$ is then generated in ICON from a uniform distribution between 0 and 1 for each parameter p_n and member m. By setting the parameter `itype_pert_gen=2` in `ensemble_pert_nml`, the random variable $\alpha_{n,m}$ leads to the value of p_n being one of default or perturbed value, as indicated in Fig. 38. A choice of `itype_pert_gen=1` would lead to a perturbed value somewhere within the perturbation range according to $\alpha_{n,m}$ taken at face value as implemented for the global ICON-EPS and its nest over Europe.
Another important control switch for the perturbation procedure is the namelist entry `timedep_pert` in `ensemble_pert_nml`. A value of 1 causes the randomization to depend on forecast start time as it is done in ICON-D2-EPS. The global ICON-EPS uses a value of 0 leading to a dependence on the member ID only. Some details about these namelist entries can be found in Reinert et al. as well. They provide for more options to generate EPS members than just in the way it is presented here for ICON-D2-EPS.
There might be slight drawbacks in the described implementation of the perturbation procedure in ICON-D2-EPS compared to COSMO-D2-EPS, like the symmetric perturbation ranges and the methodological aspect that each parameter is no longer perturbed exactly in 50% of the members of a model run, but only statistically. However, the namelist-based approach makes it more feasible to run the RPP and makes the approach more portable between different machines and exchangeable among COSMO partners compared to COSMO-D2-EPS.

6 Conclusions

Like most of the WG7 Priority Projects, the activities of APSU covered a quite broad range of ensemble development topics. This allows us to carry out ensemble development in a coordinated way between the different COSMO partners, while preserving the specificity of the different ensemble implementations and needs (different domains, focus on somehow different weather phenomena, different computing capabilities and human resources). However, thanks to the APSU Priority Project, the activities about ensemble development of the different COSMO members were focused on specific high-priority themes, benefiting from each others' work and avoiding duplications and dispersion of efforts.

Some conclusions can be drawn from this work, highlighting the impact of the Project activities on the ensemble development and operational usage, and identifying the next steps needed to perform the transition of the ensembles to the ICON model.

- The new stochastic approaches to model the model error (SMME and AMPT) are promising methods to describe the model error and incorporate it into ensemble design. However, their complexity and the difficulty of objectively representing the model error require a long development phase, which is not concluded in the time frame of APSU. It is important to keep these developments linked to each other, in order to benefit from each others' results and experiences. Due to the transition of the Consortium to the ICON model, it is now suggested to implement the two methods in ICON and continue their development in the new model.
- The SPPT method is successfully used at MCH as operational model perturbation method. Therefore MCH will invest resources in implementing SPPT in ICON, in

view of the future transition to the ICON model for their ensembles. The proposed iSPPT variant, even if it did not prove to be more beneficial than the classical SPPT in COSMO, will be considered for implementation, in order to permit its usage and further assessment in the new model framework. The study performed on the analysis increments, to be possibly used as source of model perturbation, lead to positive results. However, the development of this method would require time and resources, and this has to be taken in account in planning future actions.

- The implementation of model perturbation methods in ICON will also permit to the COSMO partners to start working on the new model system and to contribute to the development of the ICON model
- The application of soil/surface perturbations proved to be successful and lead, in combination to the perturbation of the atmospheric fields, to promising results. Therefore these perturbations will also be adopted in the ICON-based ensembles.
- In order to apply the parameter perturbation method in ICON it is needed to study the sensitivity of the ICON model to its parameters, as it was done at DWD during APSU. Therefore this activity will be repeated for the other domains where ICON-based ensembles need to be developed. This will also permit to gain a deeper knowledge of the model and its behaviour, which is in the interest of all the COSMO partners.
- The tasks related to the development of new ensemble products and ensemble output calibration lead to an operational usage of the new products and of the calibrated outputs. Therefore this task will be continued in the future, by expanding to new products and improving the calibration methodology.

Concluding with a poetic note, the APSU PP was like the underground fresh water reservoir of the Sumerian and Accadic culture, after which the project is named, from which fertilising quality the future work will benefit.

7 Acknowledgements

We are grateful to the COSMO Consortium, which gives us the opportunity to work together and benefit from mutual exchanges and discussions. The Project leader would like to thank the members of Working Group 7 for their continuous and active engagement.

8 References

Astakhova, E. D., Montani, A., Alferov, D. Yu., 2015. Ensemble forecasts for the Sochi-2014 Olympic Games. *Russian meteorology and hydrology* **40**, 8, 531-539.

Alferov, D. Yu., Astakhova, E. D., Rivin, G. S., Rozinkina, I. A., 2014. Development of High-resolution Ensemble Prediction System for the Region of Sochi-2014 Winter Olympics. *Proceedings of Hydrometcenter of Russia* No. 352, 5-20.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T., 2011. Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities *Mon. Weather Rev.* **139**, 12, 3887-3905.

Ben Bouallegue, Z., Haiden, T., Weber, N. J., Hamill, T. M., and Richardson, D. S., 2020. Accounting for Representativeness in the Verification of Ensemble Precipitation Forecasts. *Mon. Weather Rev.* **148**, 2049-2062. doi:10.1175/MWR-D-19-0323.1, 2020.

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A. and Snyder, C., 2011. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Weather Rev.* **139**, 6, 1972-1995.

Berner, J., Shutts, G. J., Leutbecher, M. and Palmer, T. N., 2009. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.* **66**, 3, 603-626.

Bouttier, F., Vié, B., Nuissier, O. and Raynaud, L., 2012. Impact of stochastic physics in a convection-permitting ensemble. *Mon. Weather Rev.* **140**, 11, 3706-3721.

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**, 560, 2887-2908.

Christensen, H. M., Lock, S. J., Moroz, I. M. and Palmer, T. N., 2017. Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *Q. J. R. Meteorol. Soc* **143**, 2168-2181. doi:10.1002/qj.3075

Christensen, H. M., Moroz, I. M. and Palmer, T. N., 2015. Stochastic and Perturbed Parameter Representations of Model Uncertainty in Convection Parameterization. *J. Atmos. Sci.* **72**, 6, 2525-2544.

Christensen, H.M., 2020. Constraining stochastic parametrisation schemes using high-resolution simulations. *Q. J. R. Meteorol. Soc* **146**, 938-962.

Dey, S.R.A., Roberts, N.M., Plant, R.S. and Migliorini, S., 2016. A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Q.J.R. Meteorol. Soc.* **142**, 1982-1996.

Duniec, G., Mazur, A., 2020. Dependence of skill and spread of the ensemble forecasts on the type of perturbation and its relationship with long-term norms of precipitation and temperature. *Acta Geophys.* **68**, 1505-1528. https://doi.org/10.1007/s11600-020-00467-4

Füzér, L., 2018. Analysing new approaches for model perturbations in COSMO-E. ETH Zürich.

Gal-Chen, T. and Somerville, R. C. J., 1975. On the use of a coordinate transformation for the solution of the Navier-Stokes equations. *Journal of Computational Physics* **17**, 2, 209-228.

García-Ojalvo, J. and Sancho, J. M. and Ramírez-Piscina, L., 1992. Generation of spatiotemporal colored noise. *Physical Review A* **46** 4670–4675.

Gultepe et al., 2006. A New Visibility Parameterization for Warm-Fog Applications in Numerical Weather Prediction Models. *JAMC* **45**, 1469-1480.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble

prediction systems. *Wea. and Forecasting* **15**, 5, 559–570.

Kiktev, D., Joe, P., Isaac, G. A., Montani, A., Frogner, I.-L., Nurmi, P., Bica, B., Milbrandt, J., Tsyrulnikov, M., Astakhova, E., Bundel, A. and others, 2017. FROST-2014: The Sochi winter Olympics international project. *Bull. Amer. Meteor.* **98**, 9, 1908-1929.

Kirsanov, A., Gofa, F., Boucouvala, D., Batignani, F., Bogdan, M., Linkowska, J., Pflüger, U., Tesini, M.S. and Bundel, A., 2020. Common Verification Activity during JJA 2018 – MAM 2019: Main Results. *COSMO Newsletter* **20**, 10-17.

Klasa, C., Arpagaus, M., Walser, A., Wernli, H., 2018. An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Q. J. R. Meteorol. Soc.* **144**, 744–764. https://doi.org/10.1002/qj.3245

Knuth, D. E., 1981. Generating Uniform Random Numbers. In: Seminumerical Algorithms, 2nd ed., Reading, MA, Addison-Wesley Pub.2, 9-38. Print.: The Art of Computer Programming.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N. and Weisheimer, A., 2017. Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Q.J.R. Meteorol. Soc* **143**, 2315-2339. https://doi.org/10.1002/qj.3094

Machulskaya, E. and Seifert, A., 2019. Stochastic differential equations for the variability of atmospheric convection fluctuating around the equilibrium. *Journal of Advances in Modeling Earth Systems* **11**, 8, 2708-2727.

Marsaglia, G., 2003. Random number generators. *Journal of Modern Applied Statistical Methods* **2**, 1, 2-13. doi:10.22237/jmasm/1051747320.

Marsigli, C., D. Alferov, E. Astakhova, G. Duniec, D. Gayfulin, C. Gebhardt, W. Interewicz, N. Loglisci, F. Marcucci, A. Mazur, A. Montani, M. Tsyrulnikov, A. Walser, 2019. Studying perturbations for the representation of modeling uncertainties in Ensemble development (SPRED Priority Project): Final Report. *COSMO Technical Report* **No. 39**, 55 pp.

Maurer, D., Walser, A. and Arpagaus, M., 2014. First COSMO-E experiments with the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *COSMO Newsletter* **14**, 19-27.

Mazur, A., Duniec G., 2015. Ensemble Prediction System (EPS)-based forecast prepared from perturbation of soil conditions. *COSMO Newsletter* **15**, 63-71.

Monin, A. S. and Yaglom, A. M., 2013. Statistical fluid mechanics, volume II: Mechanics of turbulence. Edited by Courier Corporation.

Montani, A., Cesari, D., Marsigli, C. and Paccagnella, T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. *Tellus A: Dynamic Meteorology and Oceanography* **63**, 3, 605-624.

Montani, A., Marsigli, C. and Paccagnella, T., 2013. Development of a COSMO-based limited-area ensemble system for the 2014 Winter Olympic Games. *COSMO Newsletter* **13**, 93-99.

Montani, A., Alferov, D., Astakhova, E., Marsigli, C. and Paccagnella, T., 2014. Development of a COSMO-based limited-area ensemble system for the 2014 Winter Olympic Games. *COSMO Newsletter* **14**, 88-94.

Piccolo, C., Cullen, M., Tennant, W. J. and Semple, A. T., 2018. Comparison of different representations of model error in ensemble forecasts. *Q. J. R. Meteorol. Soc* **145**, 15-27. doi:10.1002/qj.3348

Raspanti, A., 2009. VERification System Unified Survey (VERSUS): Final Report. *COSMO Technical Report* **12**, pp 12.

Rivin, G., Rozinkina, I., Astakhova, E. and Coauthors, 2018. The COSMO Priority Project CORSO (Consolidation of Operational and Research Results for the Sochi Olympic Games). Final Report. *COSMO Technical Report* **35**, pp 70.

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C., Gebhardt, C., Marsigli, C., and Zängl, G., 2021. DWD Database reference manual for global and regional ICON and ICON-EPS forecasting system.
https://www.dwd.de/SharedDocs/downloads/DE/modelldokumentationen/nwv/icon/icon_dbbeschr_aktuell.html

Stoelinga, M. T., and T. T. Warner, 1999. Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event. *J. Appl. Meteor.* **38**, 385-404.

Taszarek, M. and L. Kolendowicz, 2013. Sounding-derived parameters associated with tornado occurrence in Poland and Universal Tornadic Index. *Atmos. Res.* **134**, 186-197.

Tsyrulnikov, M. and Gayfulin, D., 2019. Estimation of model errors on convective scales: a coarse-graining study (preliminary stage results). *COSMO Newsletter* **19**, 12-22.

Tsyrulnikov, M. and Gayfulin, D., 2017. A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteorol. Zeitschrift* **26**, 5, 549-566.

Wastl, C., Wang, Y., Atencia, A. and Wittmann, C., 2019. A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. *Mon. Weather Rev.* **147**, 6, 2217-2230.

Wichmann, B. A., Hill, I. D., 1982. Algorithm AS 183: An Efficient and Portable Pseudo-Random Number Generator. *Journal of the Royal Statistical Society.* Series C (Applied Statistics). **31**, 2, 188-190. doi:10.2307/2347988

Wong et al., 2013. Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations. *Geosci. Model Dev.* **6**, 429-443.

List of COSMO Newsletters and Technical Reports

(available for download from the COSMO Website: www.cosmo-model.org)

COSMO Newsletters

No. 1: February 2001.

No. 2: February 2002.

No. 3: February 2003.

No. 4: February 2004.

No. 5: April 2005.

No. 6: July 2006.

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

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

No. 9: December 2008.

No. 10: March 2010.

No. 11: April 2011.

No. 12: April 2012.

No. 13: April 2013.

No. 14: April 2014.

No. 15: July 2015.

No. 16: July 2016.

No. 17: July 2017.

No. 18: November 2018.

No. 19: October 2019.

No. 20: December 2020.

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.

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

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

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

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

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

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

No. 20: A. Roches, O. Fuhrer (2012):
Tracer module in the COSMO model.

No. 21: Michael Baldauf (2013):
A new fast-waves solver for the Runge-Kutta dynamical core.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_21

No. 22: C. Marsigli, T. Diomede, A. Montani, T. Paccagnella, P. Louka, F. Gofa, A. Corigliano (2013):
The CONSENS Priority Project.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_22

No. 23: M. Baldauf, O. Fuhrer, M. J. Kurowski, G. de Morsier, M. Müllner, Z. P. Piotrowski, B. Rosa, P. L. Vitagliano, D. Wójcik, M. Ziemiański (2013):
The COSMO Priority Project 'Conservative Dynamical Core' Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_23

No. 24: A. K. Miltenberger, A. Roches, S. Pfahl, H. Wernli (2014):
Online Trajectory Module in COSMO: a short user guide.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_24

No. 25: P. Khain, I. Carmona, A. Voudouri, E. Avgoustoglou, J.-M. Bettems, F. Grazzini (2015):
The Proof of the Parameters Calibration Method: CALMO Progress Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_25

No. 26: D. Mironov, E. Machulskaya, B. Szintai, M. Raschendorfer, V. Perov, M. Chumakov, E. Avgoustoglou (2015):
The COSMO Priority Project 'UTCS' Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_26

No. 27: J-M. Bettems (2015):
The COSMO Priority Project 'COLOBOC': Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_27

No. 28: Ulrich Blahak (2016):
RADAR_MIE_LM and RADAR_MIELIB - Calculation of Radar Reflectivity from Model Output.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_28

No. 29: M. Tsyrulnikov and D. Gayfulin (2016):
A Stochastic Pattern Generator for ensemble applications.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_29

No. 30: D. Mironov and E. Machulskaya (2017):
A Turbulence Kinetic Energy – Scalar Variance Turbulence Parameterization Scheme.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_30

No. 31: P. Khain, I. Carmona, A. Voudouri, E. Avgoustoglou, J.-M. Bettems, F. Grazzini, P. Kaufmann (2017):
CALMO - Progress Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_31

No. 32: A. Voudouri, P. Khain, I. Carmona, E. Avgoustoglou, J.M. Bettems, F. Grazzini, O. Bellprat, P. Kaufmann and E. Bucchignani (2017):
Calibration of COSMO Model, Priority Project CALMO Final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_32

No. 33: N. Vela (2017):
VAST 2.0 - User Manual.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_33

No. 34: C. Marsigli, D. Alferov, M. Arpagaus, E. Astakhova, R. Bonanno, G. Duniec, C. Gebhardt, W. Interewicz, N. Loglisci, A. Mazur, V. Maurer, A. Montani, A. Walser (2018):
COsmo Towards Ensembles at the Km-scale IN Our countries" (COTEKINO), Priority Project final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_34

No. 35: G. Rivin, I. Rozinkina, E. Astakhova, A. Montani, D. Alferov, M. Arpagaus, D. Blinov, A. Bundel, M. Chumakov, P. Eckert, A. Euripides, J. Förstner, J. Helmert, E. Kazakova, A. Kirsanov, V. Kopeikin, E. Kukanova, D. Majewski, C. Marsigli, G. de Morsier, A. Muravev, T. Paccagnella, U. Schättler, C. Schraff, M. Shatunova, A. Shcherbakov, P. Steiner, M. Zaichenko (2018):
The COSMO Priority Project CORSO Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_35

No. 36: A. Raspanti, A. Celozzi, A. Troisi, A. Vocino, R. Bove, F. Batignani (2018):
The COSMO Priority Project VERSUS2 Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_36

No. 37: A. Bundel, F. Gofa, D. Alferov, E. Astakhova, P. Baumann, D. Boucouvala, U. Damrath, P. Eckert, A. Kirsanov, X. Lapillonne, J. Linkowska, C. Marsigli, A. Montani, A. Muraviev, E. Oberto, M.S. Tesini, N. Vela, A. Wyszogrodzki, M. Zaichenko, A. Walser (2019):
The COSMO Priority Project INSPECT Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_37

No. 38: G. Rivin, I. Rozinkina, E. Astakhova, A. Montani, J-M. Bettems, D. Alferov, D. Blinov, P. Eckert, A. Euripides, J. Helmert, M.Shatunova (2019):
The COSMO Priority Project CORSO-A Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_38

No. 39: C. Marsigli, D. Alferov, E. Astakhova, G. Duniec, D. Gayfulin, C. Gebhardt, W. Interewicz, N. Loglisci, F. Marcucci, A. Mazur, A. Montani, M. Tsyrulnikov, A. Walser (2019):
Studying perturbations for the representation of modeling uncertainties in Ensemble development (SPRED Priority Project): Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_39

No. 40: E. Bucchignani, P. Mercogliano, V. Garbero, M. Milelli, M. Varentsov, I. Rozinkina, G. Rivin, D. Blinov, A. Kirsanov, H. Wouters, J.-P. Schulz, U. Schättler (2019):

Analysis and Evaluation of TERRA_URB Scheme: PT AEVUS Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_40

No. 41: X. Lapillonne, O. Fuhrer (2020):
Performance On Massively Parallel Architectures (POMPA): Final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_41

No. 42: E. Avgoustoglou, A. Voudouri, I Carmona, E. Bucchignani, Y. Levy, J. -M. Bettems (2020):
A methodology towards the hierarchy of COSMO parameter calibration tests via the domain sensitivity over the Mediterranean area.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_42

No. 43: H. Muskatel, U. Blahak, P. Khain, A. Shtivelman, M. Raschendorfer, M. Kohler, D. Rieger, O. Fuhrer, X. Lapillonne, G. Rivin, N. Chubarova, M. Shatunova, A. Poliukhov, A. Kirsanov, T. Andreadis, S. Gruber (2021):
The COSMO Priority Project $T^2(RC)^2$: Testing and Tuning of Revised Cloud Radiation Coupling, Final Report
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_43

No. 44: M. Baldauf, D. Wojcik, F. Prill, D. Reinert, R. Dumitrache, A. Iriza, G. deMorsier, M. Shatunova, G. Zaengl, U. Schaettler (2021):
The COSMO Priority Project CDIC: Comparison of the dynamical cores of ICON and COSMO, Final Report
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_44

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), Romania (NMA) and Russia (RHM). 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 editor:

Massimo Milelli
massimo.milelli@cimafoundation.org