Consortium

for

Small-Scale Modelling

Technical Report No. 39

Studying Perturbations for the Representation of modeling uncertainties in Ensemble Development (SPRED Final Report)

May 2019

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_39

Deutscher Wetterdienst
MeteoSwiss
Ufficio Generale Spazio Aereo e Meteorologia
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
Instytucie Meteorogii 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, ARPA Piemonte

Studying Perturbations for the Representation of modeling uncertainties in Ensemble Development (SPRED Final Report)

Project participants:

***C. Marsigli*[1] ‡, *D. Alferov*[2], *E. Astakhova*[2],**

***G. Duniec*[3], *D. Gayfulin*[2], *C. Gebhardt*[1],**

***W. Interewicz*[3], *N. Loglisci*[4], *F. Marcucci*[5],**

***A. Mazur*[3], *A. Montani*[6], *M. Tsyrulnikov*[2], *A. Walser*[7].**

‡ Project Leader

[1] DWD

[2] Roshydromet

[3] IMGW

[4] ArpaP

[5] ItAF-ReMet

[6] ArpaE

[7] MCH

Contents

1 Introduction

This document reports about the activities carried out in the SPRED Priority Project. The Project lasted from September 2015 to February 2018. The project was organized in 5 Tasks, with cross-cutting issues. The report reflects the structure of the Project, consisting of 5 sections. In Section 2, the work aimed at investigating the spread/skill relation of the ensembles run by the COSMO members is described. In Section 3, work on model physics perturbation is shown, while Section 4 deals with lower boundary perturbation. In Section 5, the activities carried out in the post-processing and calibration of the ensemble outputs are presented, also included in the SRNWP-EPS II Project of EUMETNET. Finally, in Section 6 is described the work on initial condition for the convection-permitting ensembles. Each Section is structured in subsections describing the contribution of the different COSMO members.

2 What we know (and what we do not know) about the spread/skill relation of our ensembles.

The spread/skill relation computed for the different COSMO ensembles highlights a common deficiency: the underdispersiveness of the ensembles for the surface weather parameters. Therefore it was decided to analyse this relation in greater detail, trying to provide indications about how to cure the problem. The general aim of this coordinated activity was to understand which factors determine the lack of spread for near-surface weather parameters and to propose solutions in terms of which elements of the modeling system should be perturbed to account for this uncertainty.

2.1 DWD contribution

During the PP, DWD was running operationally COSMO-DE-EPS, a 20 member ensemble based on the COSMO model run at 2.8 km horizontal resolution. At the beginning of PP SPRED, both the boundary and initial conditions of COSMO-DE-EPS were based on forecasts of four global models (ICON of DWD, IFS of ECMWF, GFS of NCEP, GSM of JMA) dynamically downscaled with a COSMO model to 7 km grid spacing. These four COSMO runs (called BC-EPS) were used as boundary conditions for COSMO-DE-EPS and contributed to the perturbation of the initial conditions. The deterministic operational analysis of COSMO-DE was perturbed with anomaly fields of BC-EPS runs with respect to 3h-forecasts of the operational COSMO-EU (ICON-EU since July 2016) resulting in the IC perturbations. For perturbation of model physics, a non-stochastic approach with perturbed parameters was used. Different model configurations were defined by setting one of 5 (later 7) parameters to non-default values. This set of perturbed parameters did not change during the forecast lead time and was identical for all forecast runs, i.e. each member of COSMO-DE-EPS was characterized by a specific parameter set-up being identical for all forecasts. More details (including the soil moisture perturbations) of the member generation of COSMO-DE-EPS at the beginning of PP SPRED are described in Peralta et al. (2012). During the SPRED PP, several improvements were applied to the member generation. In November 2016, the parameter perturbation method was enhanced by including additional parameters: thick_sc, radqi_fact, radqc_fact, a_stab, and c_diff. These new perturbations focus on the improvement of EPS forecasts for variables being relevant for renewable energy applications (wind at wind hub height, global radiation connected to cloud forecasts) and resulted in

a positive impact on the wind forecast. Furthermore, a randomization of the parameter selection methodically allowed for the inclusion of more parameters and lead to a slight increase in spread. Further improvements of the parameter perturbations during SPRED imposed that the value of rlam_heat x rat_sea is kept fixed in the parameter perturbation, otherwise a detrimental effect in 2m temperature was observed (too high spread over sea). In March 2017, the use of BC-EPS has been ceased in favour of IC perturbations based on KENDA (see Section 6.1) and boundary conditions provided by forecasts of the ICON-EPS calculated on the nested grid over Europe with approx. 20 km grid spacing. The use of Initial Conditions derived from KENDA analyses increases the spread of the ensemble, thereby improving the skill of relevant variables, keeping it constant. Results are briefly described in Section 6. The use of ICON-EPS mainly improved the spread of several parameters in particular for later lead times. A summary evaluation of the spread/skill relation for COSMO-DE-EPS in different configurations is shown in Fig. 1.

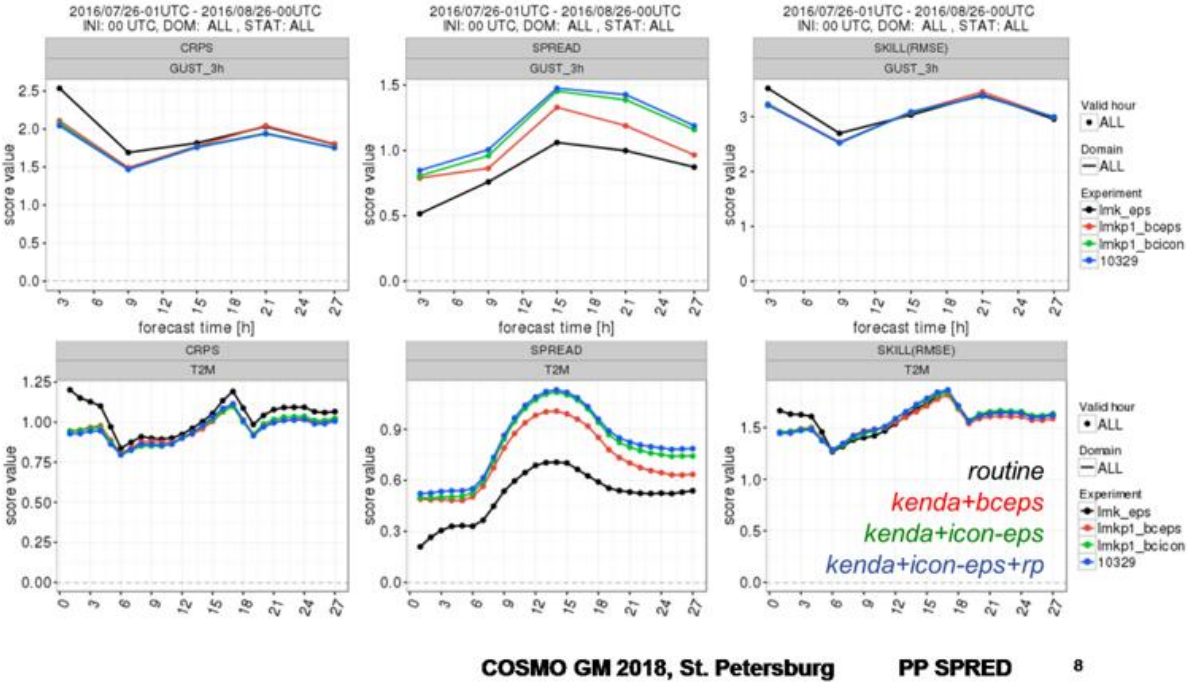

Figure 1: CRPS (left), spread (middle) and RMSE (right) of 4 different configurations of COSMO-DE-EPS, as a function of the forecast lead time, in terms of wind gusts (upper row) and 2m temperature (lower row). Black line: operational set-up before modifications; red line: use of KENDA for initial conditions, but BC-EPS for boundary conditions; green line: use of KENDA and ICON-EPS; blue line: use of KENDA, ICON-EPS and randomized physics perturbations.

The scheme for model error developed by Ekaterina Machulskaya (EM-scheme) has been tested (see presentation by Gebhardt et al. at WG7 parallel session during COSMO GM 2017, Jerusalem). It was found a bug in the code, the correction of this which led to a much smaller response of the model to perturbations as it should be. The point was that the diffusion of the model error, that is intended to represent spatial correlations, effectively (and artificially) decreases the noise level, the parameters of which are determined on the theoretical basis as if there were no spatial diffusion. This means that the effective parameters of the perturbations in the presence of the diffusion should differ from those that were determined according to the theory. The expressions for the effective values of the parameters were derived and

preliminary experiments (for selected days) were performed quite successfully.

2.2 MeteoSwiss contribution

MeteoSwiss was running operationally COSMO-E, a 21 member ensemble based on the COSMO model with 2.2 km grid-spacing. It was performed twice a day (00 and 12 UTC) and for a lead-time of 120 hours. Initial conditions are taken from a kilometer-scale ensemble data assimilation (KENDA) cycle running at the same resolution. Boundary conditions are taken from IFS-ENS control and the 20 first perturbed members. Model physics is perturbed with the SPPT scheme (Buizza et al., 1999; Palmer et al., 2009). COSMO-E is generally underdispersive in the lower troposphere, even though SPPT is effective in increasing the spread, except in winter when the physical tendencies are small. As shown in Fig. 2 taken from Klasa et al. (2018), the lack of spread is largest in winter, in particular for 2m temperature and 2m relative humidity (RH2m). In summer the ensemble is rather well dispersed or even overdispersive for RH2m when considering observation errors according to Saetra et al. (2004) with observation error estimates from Bouttier et al. (2016). Furthermore, as compared to ECMWF ENS, COSMO-E shows a better spread/error ratio, in particular for convective cases.

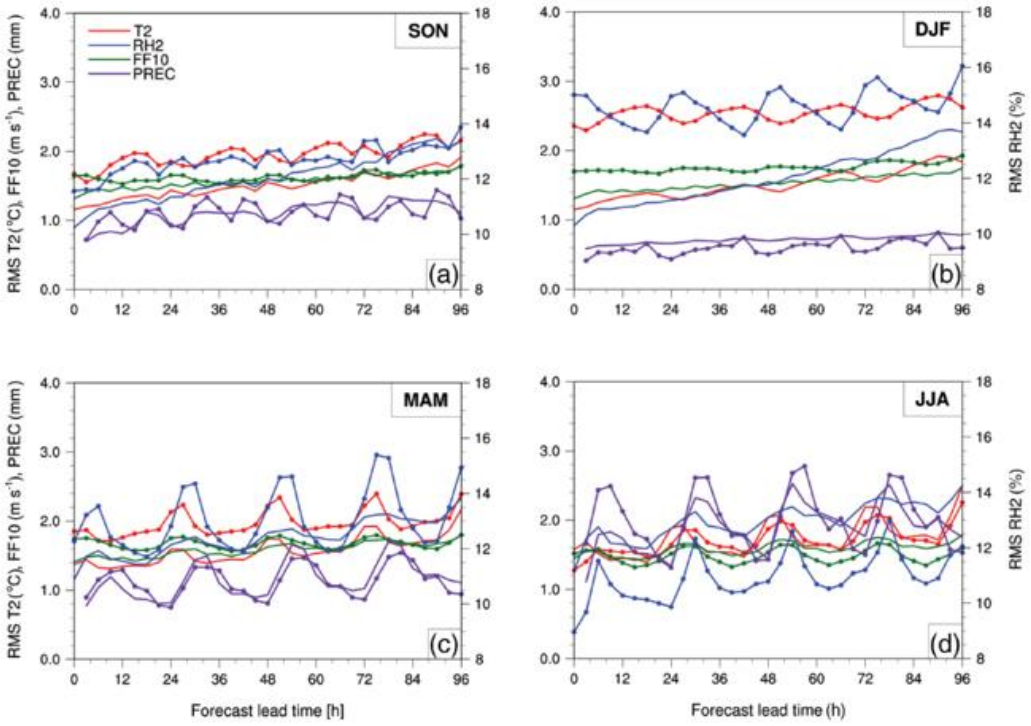

Figure 2: Unbiased forecast error (lines with symbols) and ensemble spread boosted by an observation-error estimate (lines without symbols) for the seasonal verification of COSMO-E forecasts initialized at 1200 UTC in 2016/2017 for Switzerland: 2m temperature (red), 2m relative humidity (blue), 10m wind speed (green), and 3 h accumulated precipitation (purple). Left-hand y-axes are for temperature, wind speed, and precipitation, while right-hand y-axes are for relative humidity. (Klasa et al., 2018).

In addition, dynamical processes determining the time evolution of difference kinetic energy (DKE) have been investigated (Klasa et al., 2019). DKE is quantified by means of ensemble

variance of the irrotational and nondivergent horizontal wind, respectively. Three case studies characterized by contrasting predictability levels of precipitation have been investigated with the convection permitting ensemble COSMO-E for a forecasting period of 4 days. The results suggest that the large-scale flow and diurnal solar forcing, associated with higher spatiotemporal predictability, determines the overall evolution of limited-area ensemble variance of the horizontal wind, which increases in the presence of moist convective activity or strong synoptic-scale forcing, and stagnates or decreases otherwise, rendering forecasts of convection-permitting ensembles valuable beyond the very short forecast range.

2.3 Arpae contribution

2.3.1 Convection-parametrised ensemble (COSMO-LEPS)

In the framework of enhancing the spread/skill relation for the COSMO-LEPS system, the use of SPPT was tested. For two sets of periods (from 22/11/2014 to 31/01/2015 and from 2/6/2015 to 20/7/2015), the operational COSMO-LEPS (referred to as OPER) was tested against a test version (referred to as SPPT_SNGL) characterized by the following features: use of SPPT and executable compiled in single-precision mode. Both systems ran at 7 km of horizontal resolution, 20 members, and with a forecast range of 132 hours. The performance of the two systems was analyzed in terms of 12-h cumulated precipitation as well as in terms of spread/skill relation of 2-metre temperature. The verification network included the SYNOP reports covering Central and Southern Europe, with an availability of about 1000 reports/day (see Fig. 3). Here, we consider the results relative to the summer period (June-July 2015). As for total precipitation, the 2 panels of Fig. 4 reports the probabilistic forecast skill of OPER and SPPT_SNGL in terms of the ROC area for total precipitation exceeding 1 and 10 mm over 12 hours (left and right panel, respectively).

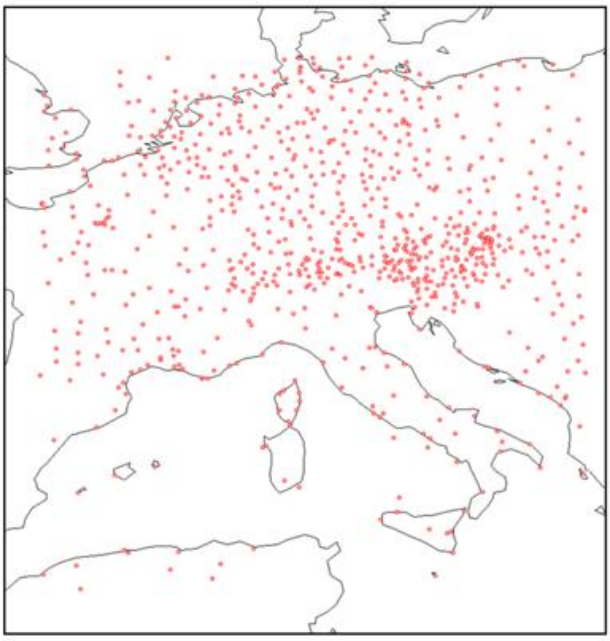

Figure 3: Verification network used for the intercomparison between OPER and SPPT_SNGL. Each red dot denotes a station.

It can be noticed that SPPT_SNGL provides a slight better performance than OPER for

both thresholds. This is especially true for in the short range, until, say, fc +72h. These results are confirmed also by other verification scores, either threshold oriented (e.g. Brier Score, Bier Skill Score) or not (e.g. Ranked Probability Score, Percentage of Outliers).

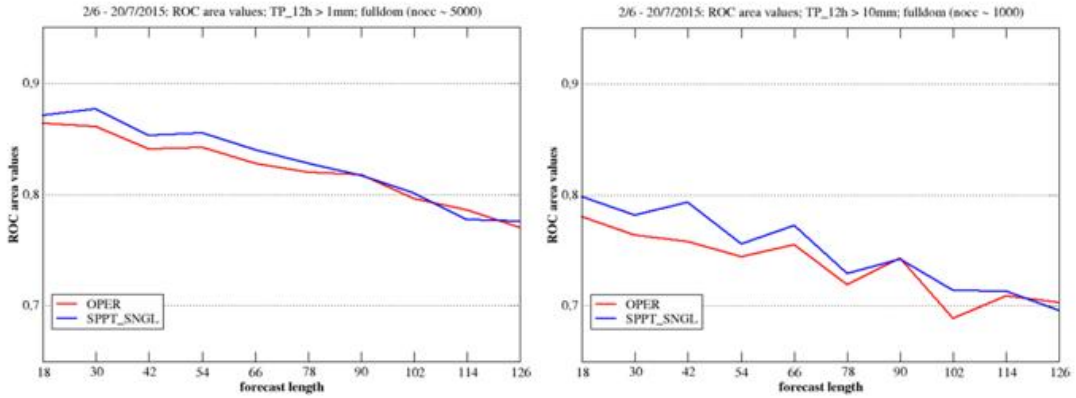

Figure 4: ROC area values for operational COSMO-LEPS (denoted with OPER, red lines) and test COSMO-LEPS (denoted with SPPT_SNGL, blue lines) for 12-hour precipitation exceeding 1 mm and 10 mm (left and right panel, respectively) as a function of the forecast range (in hours). Verification is performed over the station points of Fig. 3 and averaged over the verification period.

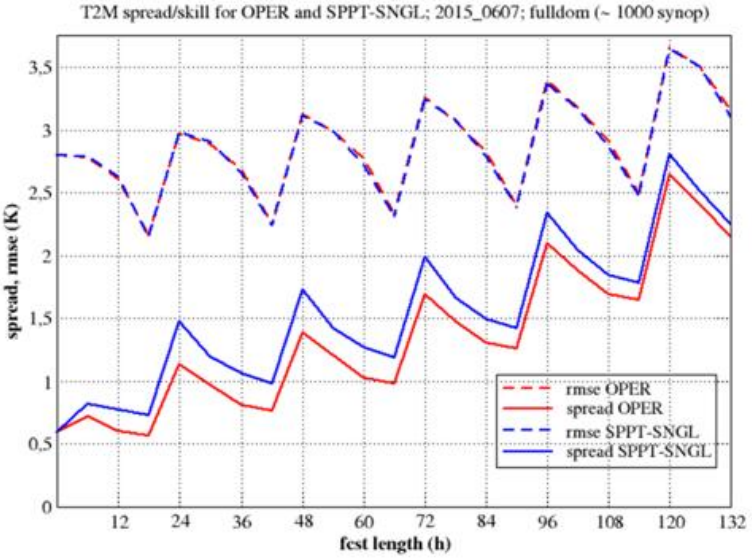

Figure 5: Root-mean-square error of the ensemble mean (dashed lines) and ensemble spread (solid lines) for operational COSMO-LEPS (denoted with "OPER", red lines) and test COSMO-LEPS (denoted with SPPT_SNGL, blue lines) for 2-metre temperature (in Kelvin) as a function of the forecast range (in hours). Verification is performed over the station points of Fig. 3 and averaged over the verification period.

As for 2-metre temperature, the results of the experimentation are summarized in Fig. 5, which reports the root-mean-square-error of the ensemble mean (dashed lines) and the ensemble spread (solid lines) for both OPER and SPPT_SNGL (read and blue, respectively). Both systems turn out to be under-dispersive, the ensemble-mean errors being about twice the

dispersion of the ensemble members. On the other hand, it can be noticed that SPPT_SNGL provides larger spread than OPER at all forecast ranges and especially for the shortest ones. This happens without a deterioration of the average skill of the system, as the forecast errors are approximately identical in both configuration (read and blue dashed lines). Although these results are relative to the summer experimentation, the outcome of the winter experimentation (NDJ2014-2015) provides the same findings and confirms the positive benefit of using SPPT in the description of model uncertainties.

2.3.2 Convection-permitting ensemble (COSMO-2I-EPS)

During the experimental phase of the COSMO-IT-EPS convection-permitting ensemble (now operational with the name COSMO-2I-EPS), the ensemble was run for selected periods in order to test the model perturbation methodology, assessing its spread/skill relation. COSMO-IT-EPS has been run for the month of October 2015, one run per day at 00 UTC, with Initial and Boundary Conditions from COSMO-ME-EPS, the 10-km ensemble of COMET, running over a Mediterranean domain. Model resolution was 2.8 km and the ensemble had 10 members. COSMO-IT-EPS was run in 3 different configurations:

- without model physics perturbation (ensemble CTRL);
- with SPPT (Stochastic Perturbation of Physical Tendencies) (ensemble SPPT);
- with SPPT and perturbation of dew parameters of the physics schemes (turbulence, microphysics, land surface) (ensemble SPPT-PP).

The spread/skill relation of the three configurations in terms of 2m temperature (above) and 6-h precipitation (below) is shown in Fig. 6.

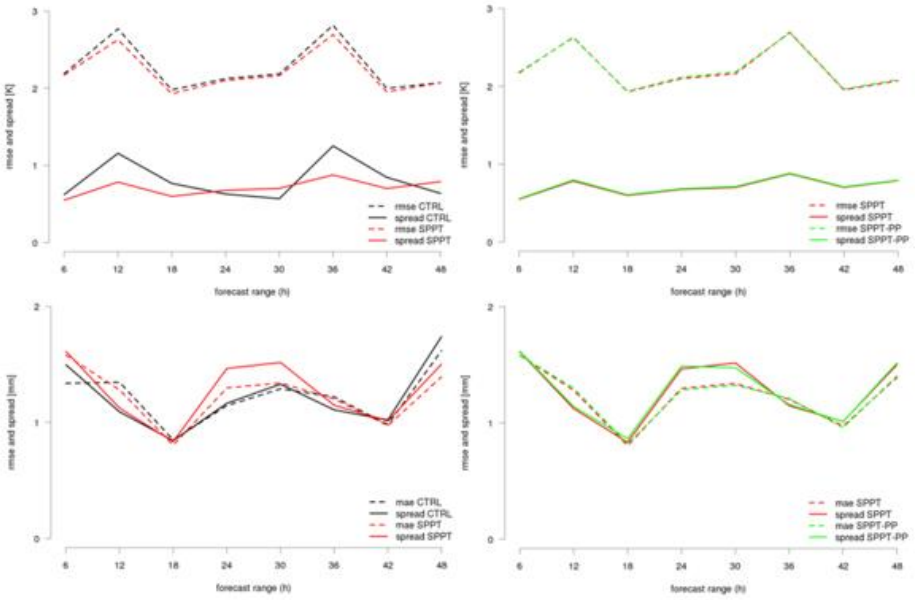

Figure 6: Error of the ensemble mean (dashed lines) and spread (solid lines), as a function of the forecast range, in terms of 2m temperature (upper row) and 6h precipitation (lower row), for the three experiments as indicated in legend. Error is RMSE for temperature and MAE for precipitation.

Temperature is compared against observed values at the SYNOP stations covering the Italian domain, while precipitation is compared against accumulated values recorded by the raingauges of a dense network covering Italy. In terms of 2mT, all configurations are underdispersive. Adding the SPPT does not yield an increase of the spread, in contrast with what was found in other configurations (e.g. COSMO-E). Adding parameter perturbations does not increase the spread and does not influence the error. In previous studies (Marsigli, 2009) it was noticed that parameter perturbation may have a highly localized impact, temporally and spatially, and dependent on the weather situation, therefore it is difficult to detect an impact in a statistical evaluation over a period. A spatial verification is applied to the precipitation fields, accumulated over 6 hour periods (Fig. 7). The verification method is the DIST method DIST (Marsigli et al, 2008). According to this method, the verification domain is first covered with boxes of selected size (here 0.2 x 0.2 deg). Then both forecasts (each ensemble member separately) and observations are aggregated in each box, by computing the average or the maximum (or other parameters of the precipitation distribution). Finally, common probabilistic verification scores are computed for both the average and maximum values belonging to each box and aggregated over the whole domain.

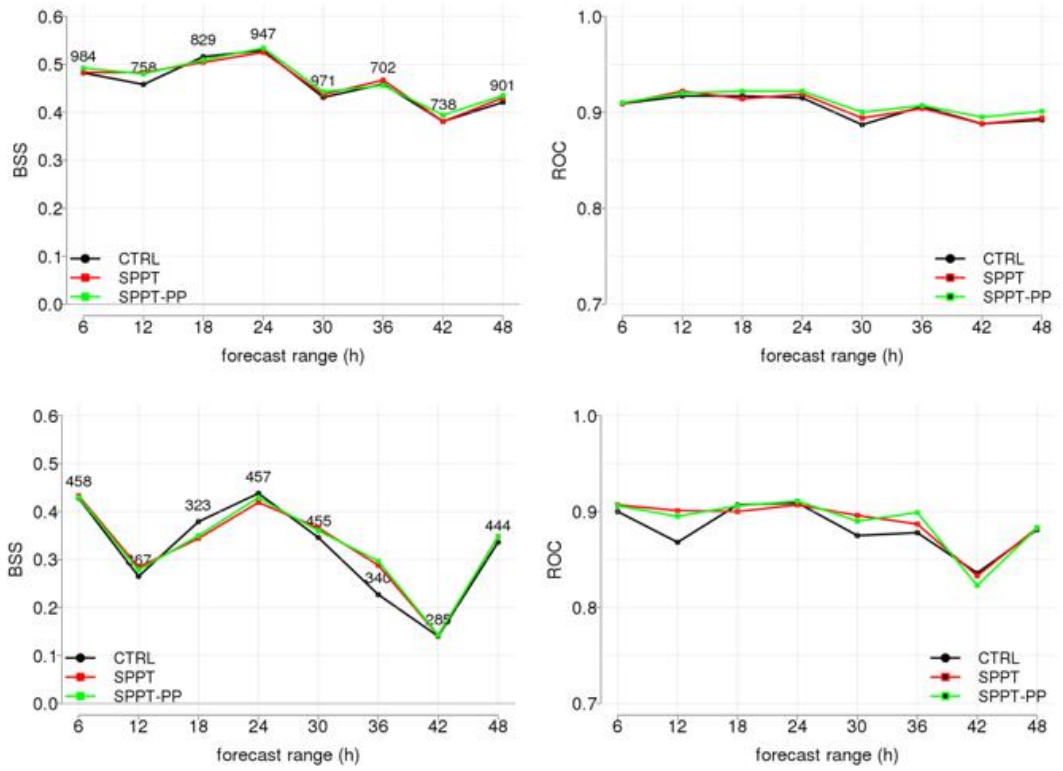

Figure 7: Brier Skill Score (left) and ROC area (right) as a function of the forecast range, for the 6h accumulated precipitation average values exceeding 1mm (top) and 5mm (bottom).

Results for average precipitation greater than 1mm/6h (upper row) and 5mm/6h (lower row) are shown in Fig. 7. While the ROC area indicates a better performance of the ensemble with perturbed physics, the Brier Skill Score shows mixed behavior. Parameter perturbation has a very little impact. How to assess the spread of the forecast in terms of precipitation is not straightforward, particularly for the convection-permitting scale. Following the development of spatial approaches for the verification of a high-resolution forecast, also the spread of the ensemble needs to be assessed in terms of spatial metrics. For this purpose, the methodology

proposed by Dey et al (2014) has been tested, where a dispersion Fraction Skill Score is computed as a measure of ensemble spread, evaluating the difference between couples of ensemble members in terms of FSS over boxes of increasing size. As an example, it is shown the mean over all couples of the dispersion FSS as a function of box size, for a case of intense precipitation occurred in the considered period (Fig. 8). The three experiments are compared, showing that the physics perturbations have an impact on the spread and that also the addition of Parameter Perturbation is able to increase the diversity of the members at all scales.

A different method has been also proposed in the Project for evaluating the ensemble spread: by computing its components using the SAL metric (Wernli et al 2008). The idea is that the different physics perturbations may address different components of the forecast uncertainty; therefore they may produce spread in different features of the resulting forecasted fields. The three component of the SAL (Structure, Amplitude and Localization) are computed for each couple of the ensemble members: 45 couples in a 10 member ensemble. The 45 dots are represented on a modified SAL-diagram in Fig. 9, where the Localisation is shown in the x axis, the Structure in the y axis and the Amplitude with the colour scale. This choice has been made to highlight the impact on the localization.

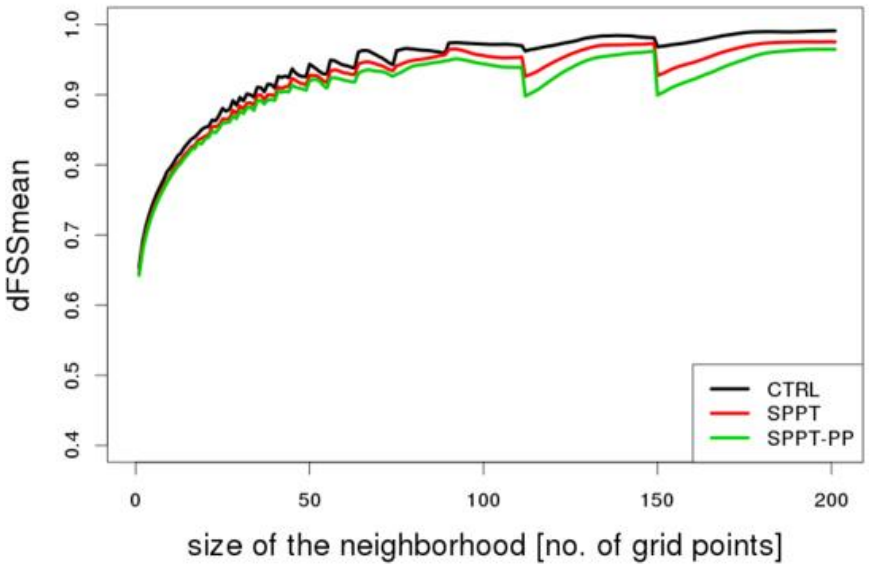

Figure 8: Mean value of the dispersion Fraction Skill Score a function of the forecast range, for the 24h accumulated precipitation for the even of the 31st October 2015.

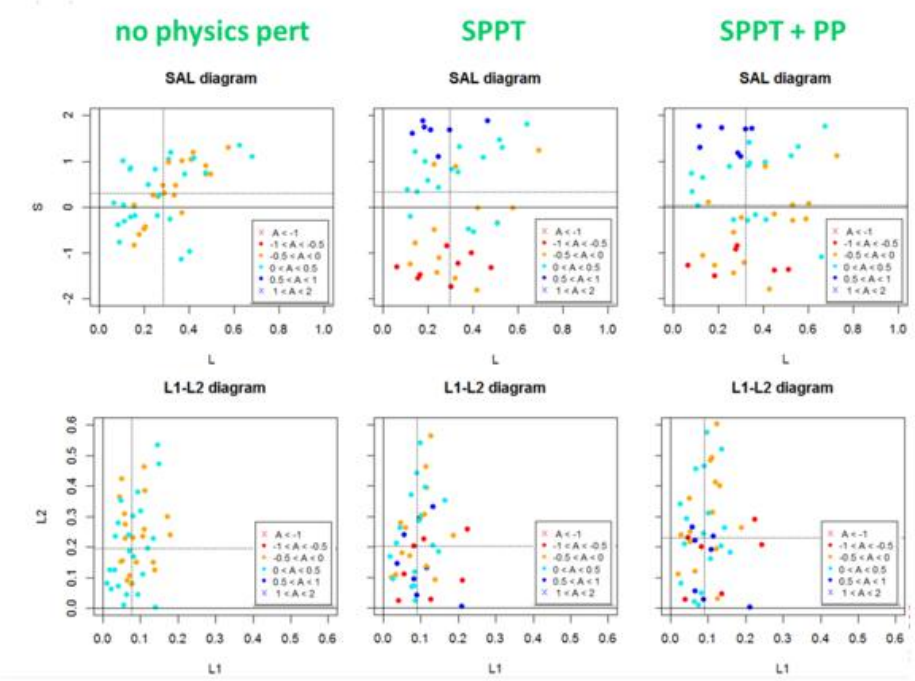

Figure 9: Dispersion of the members of the 3 ensembles (one in each column) represented as the three components of the SAL metric (top row), for the 24h accumulated precipitation for the event of the 10th October 2015. In the bottom row, the two sub-components of the L term are shown.

The impact of the physics perturbation in increasing the ensemble spread in terms of precipitation is now evident, and not only in Amplitude, but also in the Structure and in the Localization of the fields. It is now possible to detect an impact of the Parameter Perturbation, which determines more dispersion of the points in both L and S components. It is reminded that the L component consists of two terms: L1, measuring the distance of the centres of mass of the structures, and L2, measuring the difference in the degree of dispersion of each field around its centre of mass. How the increased dispersion is divided between these 2 components is shown in the bottom row, where an L1-L2 diagram is plotted. Introducing SPPT (central plot) both L1 and L2 are affected, and adding also the Parameter Perturbation (right plot), the spread in both components is further increased. This metric seems to be promising for the purpose of highlighting the contribution of a perturbation to the overall spread and to detail in which component this contribution takes place.

2.4 IMGW contribution

At IMGW, the COSMO model runs in a deterministic mode using initial and boundary conditions from the ICON global model, as shown in Fig. 10. The COSMO model with 7 km horizontal resolution (COSMO-7 km) applies nudging-based data assimilation to correct global model forecasts, ingesting the most recent set of meteorological data acquired from the GTS/WMO network. Forecast results from the COSMO-7 km are further used as IC/BCs for a nested instance of a COSMO model with a higher resolution of 2.8 km and 36-hour forecasts. A set of the deterministic COSMO-2.8 km forecasts define the basis for the operational configuration of an ensemble forecasting system.

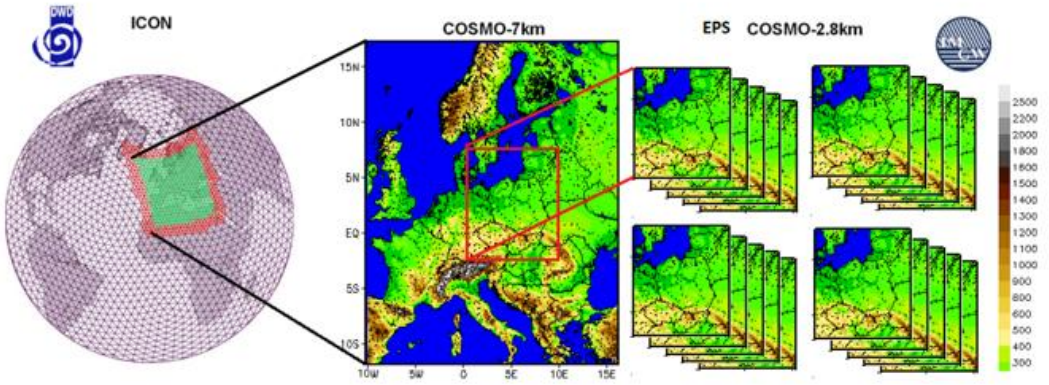

Figure 10: Operational configuration of deterministic COSMO-7 km model and 2.8 km resolution EPS runs at IMGW. From left to right domain of ICON global model, domain of COSMO-7 km model running at IMGW, and a set of nested COSMO-2.8 km domains.

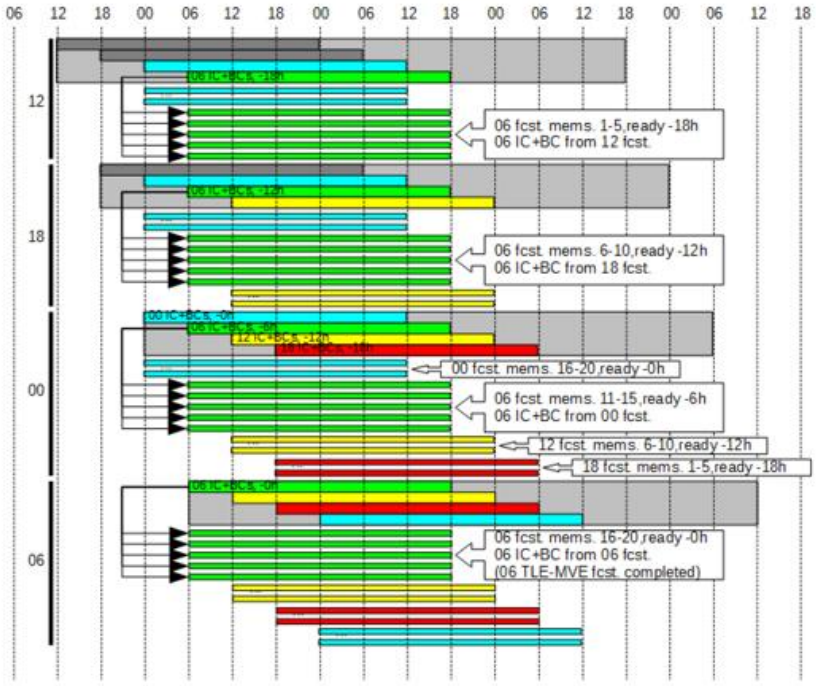

Figure 11: Operational setup of EPS based on Time-Lagged IC/BCs. X-axis lead time (UTC), Y-axis forecast individual initial time. The four different colours distinguish forecast nominal start time: blue 00 UTC, green 06 UTC, yellow 12 UTC, and red 18 UTC.

In the recently developed EPS configuration, twenty ensemble members were selected, based on the COSMO-2.8 km convection permitting (CP) forecasts. Every member of an ensemble applies perturbed lower boundary conditions, composed of the random noise of specified amplitude added to parameters of the soil-model physical parameterization (see e.g. Mazur and Duniec, 2015). In the basic EPS configuration, the whole set of IC/BC was nested from a single 78-hour run of a deterministic COSMO model with spatial resolution of 7 km, as shown in Fig. 11. In order to increase the spread of model parameters generated by forecasts, we have further adopted the concept of time-lagged IC/BCs (see e.g. Lu et al., 2007; Chen

et al., 2013), in which the set of deterministic CP forecasts is subdivided into groups starting at the consecutive time windows: 00, 06, 12, and 18 UTC.

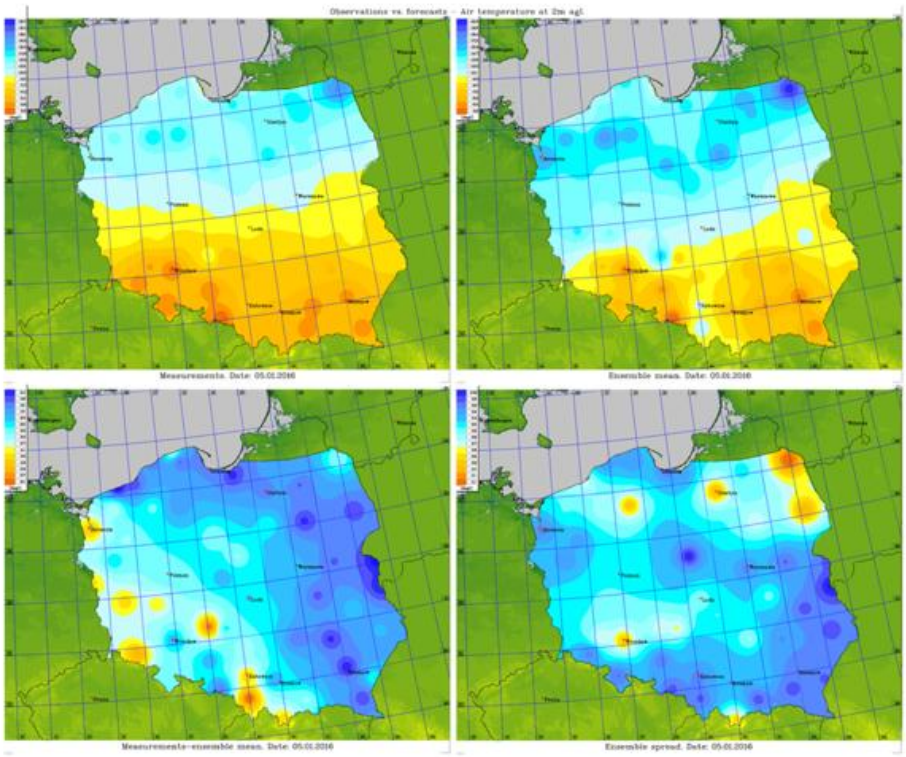

Figure 12: Comparison of the EPS derived temperature: ensemble mean (upper right), spread (lower right) with measurements (upper left), the absolute value of difference between measurement and ensemble mean (lower left).

The EPS results (in terms of EPS mean) are compared with observations collected from 61 Polish SYNOP stations. In Fig. 12 a representative analysis of the model generated temperature forecast averaged for the whole day of 5-01-2016 is presented, while Fig. 13 shows a similar verification for dew point temperature. In both cases, the EPS mean reproduces the main spatial features of the large-scale temperature distribution. The observed temperature is in a similar range (-20.7 ℃, -2.5 ℃) to the computed EPS mean (-20.8 ℃, -3.8 ℃). Similarly, the observed dew point temperature is in a similar range (-20.7 ℃, -2.0 ℃) to the EPS computed mean (-20.6 ℃, -1.6 ℃). In both cases, the model shows a bias when compared to observations with an absolute error of up to -6.1 ℃ for temperature and -5.7 ℃ for dew point temperature. For the temperature, larger error values (above 3 ℃) are well correlated with the air masses that came after the warm front crossing Poland during 5 January, 2016. Dew point temperature generally shows a small error - below 2 ℃ - except in the north-eastern part of Poland, where a persistent lack of cloud coverage can be observed for the selected date and in south-western areas, which may be correlated to a larger exceedance of precipitation intensity that started in the afternoon hours of 5-01-2016 and developed further during the night and the morning hours of the following day.

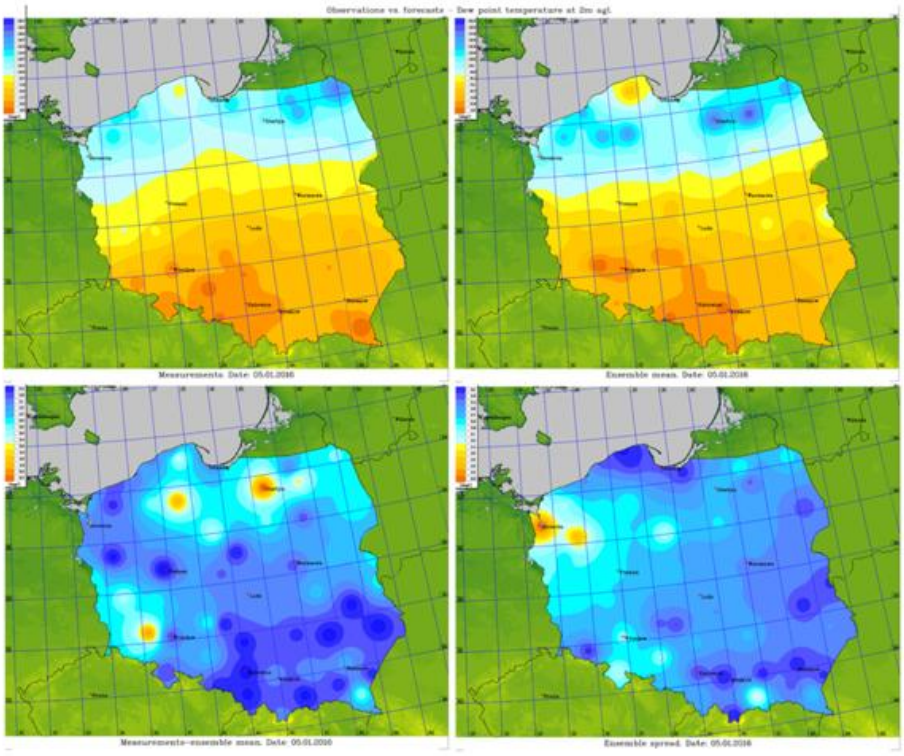

Figure 13: Same as in Fig. 12 but for dew point temperature.

A larger EPS spread implies more uncertainty in the forecast mean and hence in the ensemble members. The EPS spread is typically underestimated near the surface (Tennant and Beare, 2014) where the relations between prognostic parameters are controlled with higher fidelity parameterizations, not accounting enough for observation errors. Thus, in general EPS leads to overconfident forecasts of near-surface weather elements, especially in winter conditions where the natural variability of soil parameters is lower. In this approach, as shown in these examples, the perturbation of the evaporating fraction in the soil has a significant impact on the forecast due to the increased stimulation of energy exchange between the surface and atmosphere. The maximum spread for the temperature is 1.1 °C, with increased value covering the entire northern and western parts of Poland. For the dew point temperature, we observe a higher spread of the amplitude up to 2.0 °C, localized only in the north-western corner of the country. More advanced analysis of the EPS scores and the correlation between EPS error (skill) and spread is provided in the context of specific applications. First, selected parameters (air temperature, dew point temperature, wind speed and total precipitation) were studied for trimesters of 2016 and the entire year of 2016. The analysis was then carried out continuously in following periods. The quality of EPS forecasts, expressed in terms of skill and of spread in subsequent years seemed to improve, as shown in the following figures.

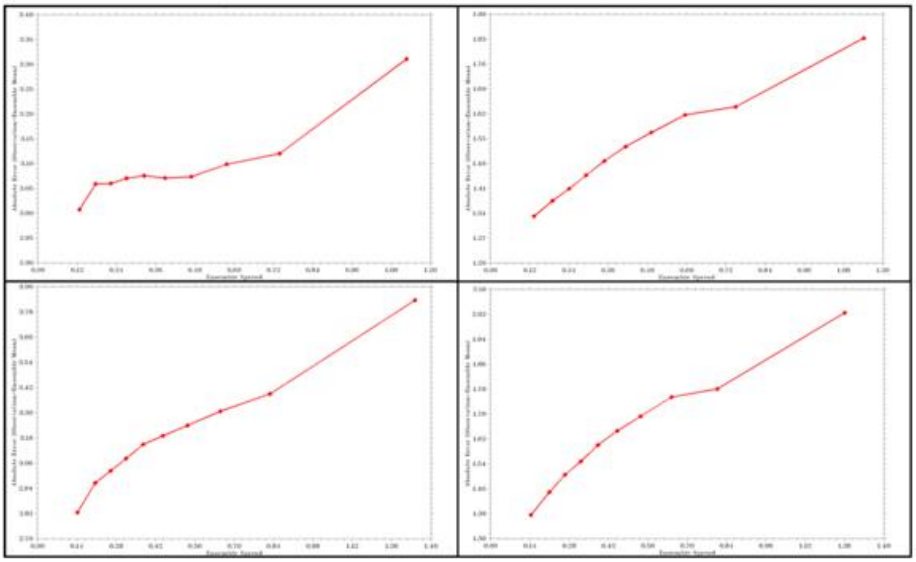

Figure 14: Left - skill-spread diagram for T2M, right - for U10M. Average values for the entire year 2016 (top row) and 2017 (bottom row).

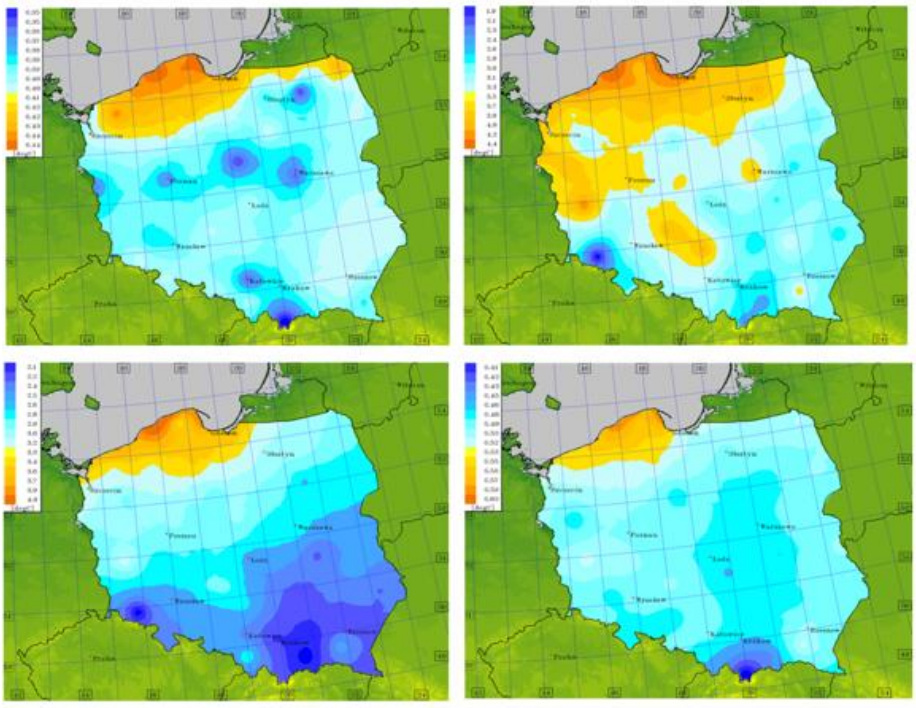

Figure 15: Spatial distribution of skill (left) and spread (right) for T2M (upper - average for year 2016; lower - for 2017).

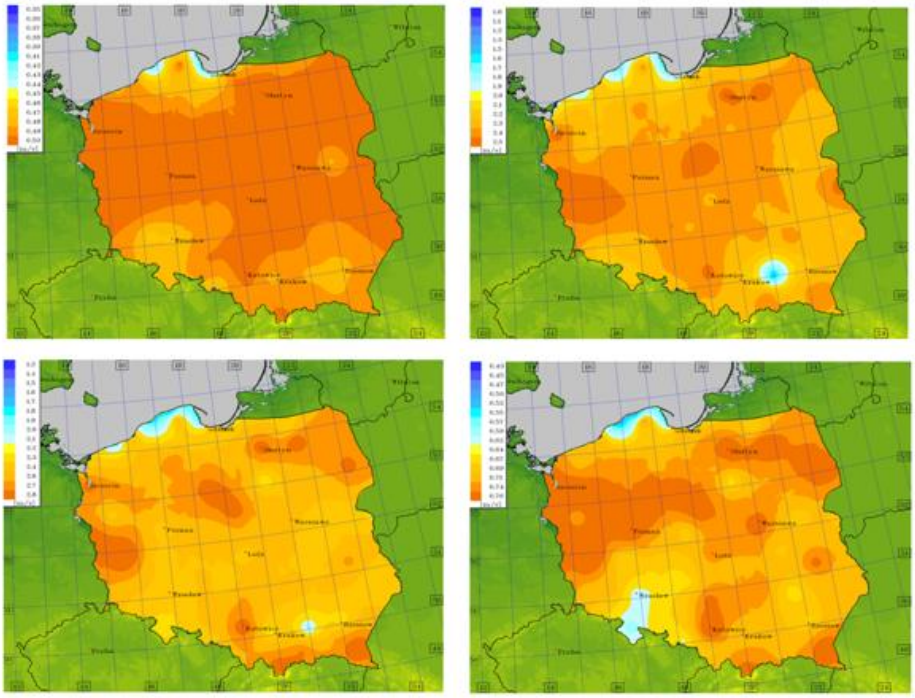

Figure 16: Same as Fig. 15, but for U10M.

The monthly analysis of skill/spread relation was done as well, starting from 2016. The representative results for the entire year 2016 are shown below.

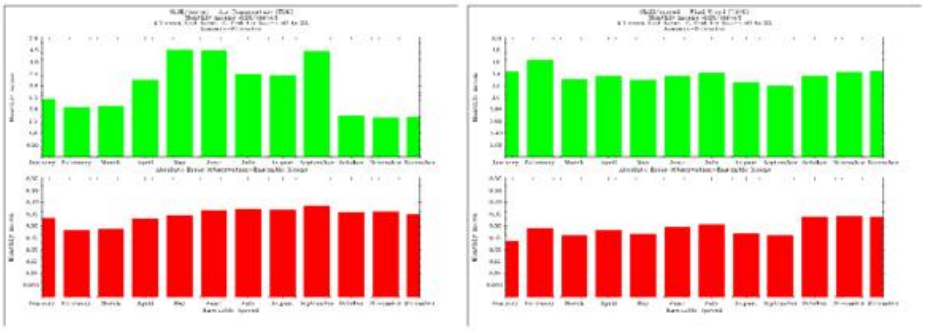

Figure 17: Monthly average skill (green) / spread (red) of T2M (left) and U10M (right). All runs/all stations, January to December 2016.

In conclusion to this part, it should be stated that skill/spread relation was studied thoroughly, using operational EPS results. Study has been carried out for months, seasons, entire year and, simultaneously, for runs (00, 06, 12, 18), forecast hours (0-36) and hours (0-23). Average spread is in general 2x to 10x lower than skill measured as MAE, which (as it was pointed out above) shows the under-dispersiveness of the system. As far as the spatial relations are concerned, one can observe that skill is in general better (i.e., smaller, as measured as MAE) for central and southern part of the domain. This is probably due to the way of generation EPS. Similarly, spread is bigger (meaning, better) in central and northern part of Poland. Similarly for time relation (monthly means) - skill is in general better (i.e., smaller) for warm months - probably due to EPS generation (perturbation of

lower boundary conditions - soil parameters - composed of the random noise of a specified amplitude added to parameters of the soil-model physical parameterization, as written above). Since it pertains to soil features, it may be less effective in winter (frozen ground) and more effective in warm and wet season. In turn, spread is been found in general bigger for warm months - again, probably due to the way the EPS members are prepared. Finally, another outcome of this part of Priority Project was a publication of paper at Meteorology Hydrology and Water Management - Research and Operational Applications (Grzegorz Duniec, Witold Interewicz, Andrzej Mazur, Andrzej Wyszogrodzki; MHWM recently became JCR journal), see: http://www.mhwm.pl/Operational-setup-of-the-soil-perturbed-time-lagged-Ensemble-Prediction-System-at,71048,0,2.html.

2.5 RHM contribution

Ideally, for a good ensemble, the behavior of the ensemble spread should follow that of the ensemble mean error and reflect the predictability of a meteorological situation. It is usually expected that a higher-resolution system has better skill (but in reality, it is not always so), thus theoretically the spread should decrease with resolution. However, the situation is not so straightforward when we are speaking about a pair "driving EPS-nested EPS" (as it is always the case with mesoscale EPSs), in which the spread at the lateral boundaries of the nested EPS is defined by the coarser-system spread.

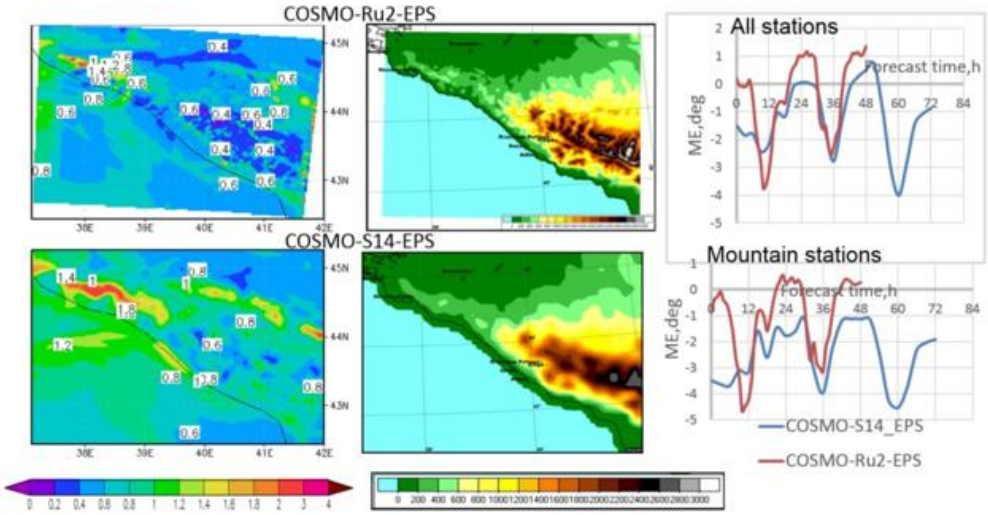

Figure 18: T2m ensemble spread (left) for 48h forecasts starting on 22 February, 2014 00UTC and orography (center) for COSMO-Ru2-EPS (top) and COSMO-S14-EPS (bottom). On the right: T2m mean error for the same date as a function of forecast lead-time for all stations in the region (top panel) and only for mountain stations (bottom panel).

Therefore, w decided to analyze the spread spatial distribution and temporal variations for two EPSs of different resolutions for the case when the higher-resolution system was nested to the coarser-resolution one. For this purpose, we applied the Sochi Olympics archive (Astakhova et al, 2016) and examined the runs of the 7-km COSMO-S14-EPS and the 2.2 km COSMO-Ru2-EPS (Montani et al, 2014). COSMO-Ru2-EPS was nested to COSMO-S14-EPS over a small region of about 350*300 km with a complex topography (containing mountain, valley, and sea points). For both systems, we considered 2-m temperature for the

1-month period from 1 to 28 February 2014 over the area corresponding to the integration domain of COSMO-Ru2-EPS. We considered both individual cases and monthly mean characteristics. Analysis of individual forecasts over the entire domain demonstrated that the higher-resolution EPS had less spread, however, its skill was not always convincingly better (see Fig. 18 for February 22, 2014). The geographical distribution of spread is correlated with orography, showing smaller spread over mountainous areas. Notably, the mean error decreased following the smaller spread over this high complex terrain (the Caucasus).

In fact, in a reliable forecast system we can expect a good spread-skill correspondence mostly for the results averaged over many forecasts ($\overline{Error^2} = \overline{Spread^2}$). Fig. 19 demonstrates the time evolution of T2m spread averaged over a month interval from February 1, 2014 to February 28, 2014 for two EPSs. As the forecasts were run twice a day (starting at 00 UTC and 12 UTC), the sample includes 56 forecasts. It can be seen that both systems start from approximately the same spread, but the spread of coarser-resolution forecasts grows with forecast time faster. The differences in spread between EPSs increase with lead time and are up to 0.8-1 deg. The smaller spread for the higher-resolution system corresponds to its smaller error (see Fig. 20 which shows the mean absolute error (wrt to all available station data in the region) of the two EPSs obtained using the FROST-2014 online verification tool for February 2014).

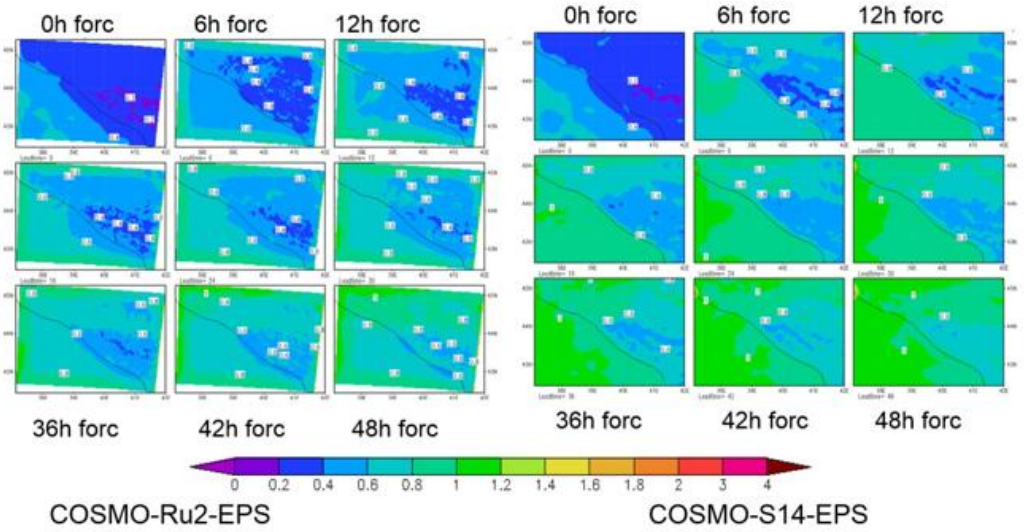

Figure 19: Evolution of 2m temperature spread with forecast time for two EPSs with different resolutions. Averaged over 56 forecasts in February 2014.

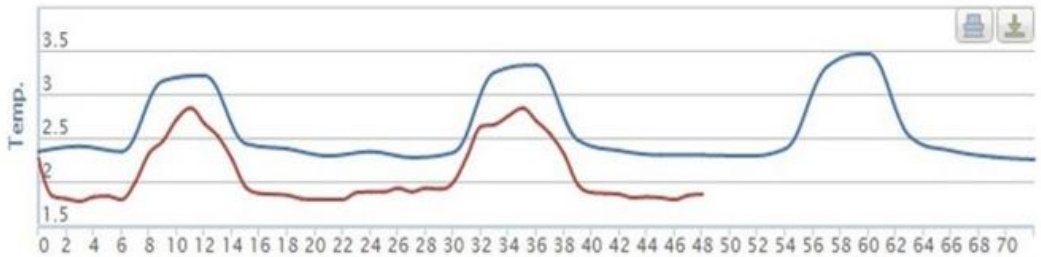

Figure 20: Mean absolute error of 2m temperature over the Sochi area averaged over 56 forecasts in February 2014. Red - COSMO-Ru2-EPS, blue - COSMO-S14-EPS.

It is also worth noting that for the T850 spread the differences between systems of different resolutions were much smaller (not shown). This indicates the important role of surface in forecast perturbations. We have examined EPSs not only with different resolutions but also with different integration domains. Looking at the spread fields in detail, we found that when regions with high T2m spread appeared near the lateral boundaries of COSMO-Ru2-EPS (near the borders of our post-stamp plots in Fig. 21), they occupied larger area and tended to propagate inside the domain in COSMO-S14-EPS being smaller and staying near the border in COSMO-Ru2-EPS. This indicates that the size of the EPS integration domain can also matter.

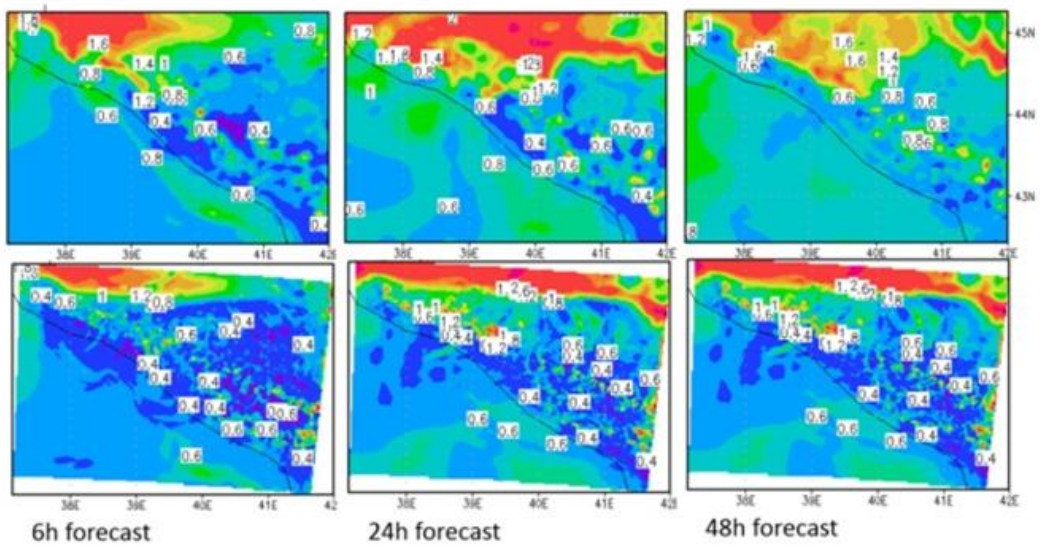

Figure 21: Evolution of the T2m spread with the forecast time in two EPSs. Upper panel: COSMO-S14-EPS; lower panel: COSMO-Ru2-EPS. Forecast from 22 February 2014 00UTC.

To explore this aspect, additional experiments were performed changing the size of the integration domain for COSMO-Ru2-EPS. The archive COSMO-Ru2-EPS data were obtained with model version 4.22 and with double precision. The new experiments were run with model 5.01 in single precision. Comparing the archive and the new results (with different model versions) we found non-zero differences in spread fields for one and the same date. Thus, we had to rerun the previous experiments for COSMO-Ru2-EPS. Finally, the experiment set was as presented in Fig. 22. The reruns of COSMO-Ru2-EPS were made for two different domains called LITTLE and BIG, while the COSMO-S14-EPS runs covered the domain S14. The comparison was made for the area, corresponding to the LITTLE domain. In addition to T2m temperature, the free atmosphere temperature was considered as well.

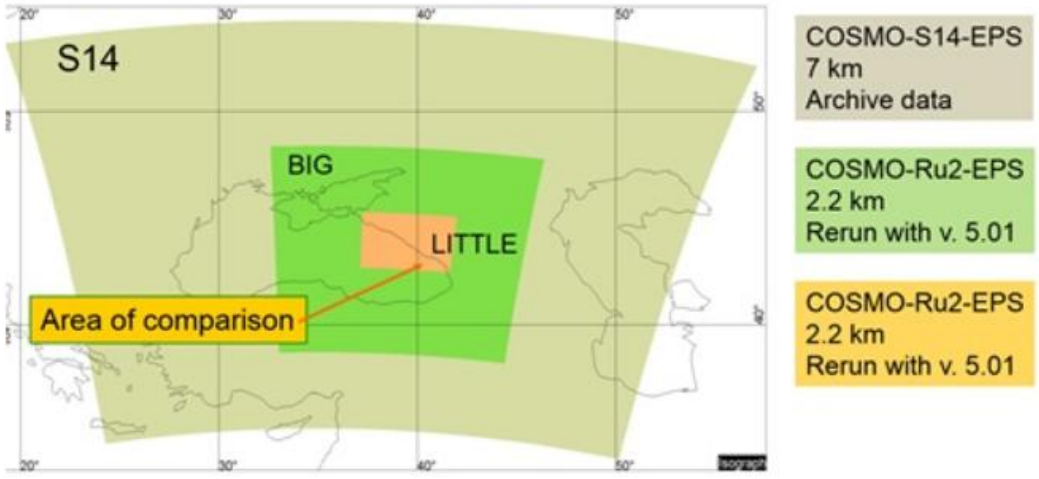

Figure 22: Experiment setting.

The resulting spread was found to depend on the size of the integration domain considerably (see Fig. 23). The effect is related to weather situation and is most pronounced near the surface and in regions with complex topography. The choice of integration domain can differently affect different prognostic realizations. Depending on location and air mass motions both ensemble mean and spread can be increased or decreased (in some cases - considerably) when the integration domain is changed. Choosing the integration domain for EPS, especially in mountain regions, one should be very careful and take into account the prevailing weather types and air mass motions.

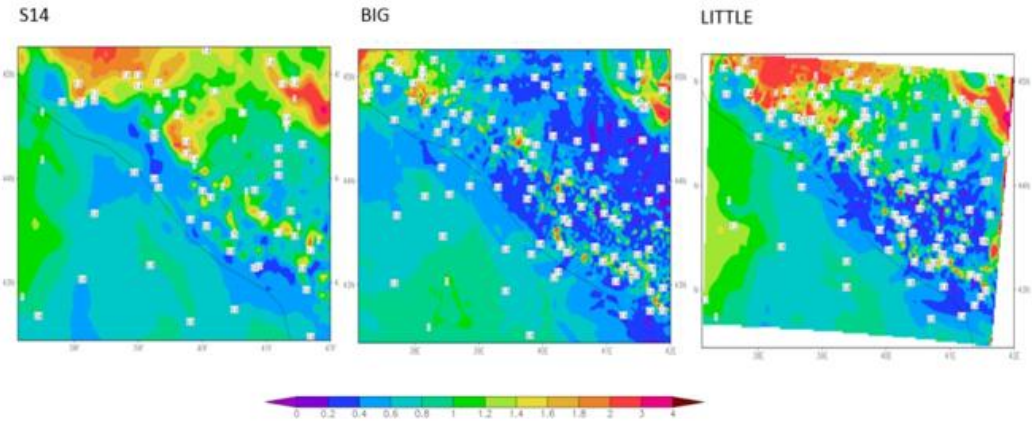

Figure 23: T2m spread fields obtained for integration domains S14, BIG and LITTLE (presented for LITTLE area). 48-h forecast starting from February 22, 2014 00 UTC.

We can conclude that generally the ensemble spread decreases for higher-resolution nested EPS but the effect depends on the size of the EPS integration domains and the dominating weather patterns.

2.6 Remarks from the discussion of the results

This Project activity has lead to a better shared understanding of the spread/skill relation of our ensembles, highlighting also the main deficiencies. Some conclusions have been drawn:

- it is important to consider also the day-by-day variation of the spread, which should follow the day-by-day variation of the error;
- it is suggested to investigate the impact of the Boundary Layer perturbations recently introduced in the COSMO model following the work by K. Kober (perturbations proportional to the variance of the tendencies coming from the turbulence scheme). Results are shown in Section 3.3;
- it is suggested to compute also the spread/skill relation on vertical profiles, to assess the vertical structure of the error and its representation in terms of spread
- it is highlighted the possibility of using Fraction Skill Score and SAL for computing the spread of the ensembles in a spatial verification sense;
- it is considered the work of Pirkka Ollinaho (ECMWF) on the introduction of parameter perturbations in their ensemble. This approach could be useful also for the COSMO Consortium but it requires strong link with the physics developers;
- it was mentioned that perturbation of parameters usually has a detectable impact on particular weather situations, therefore it is difficult to see an impact on a long statistics. Following this remark, it is suggested to perform also verification/evaluation on selected cases or collections of cases;
- it is suggested to compute also the spread/skill relation with respect to analysis, not only to observations, to reduce the influence of systematic model error;
- it is considered also how to include the observational error in the spread/skill evaluation. It is suggested to start from the literature and try the previous works of Neill Bowler and Marion Mittermaier.

3 Test and development of model perturbation

The methodologies available in the Consortium for model physics perturbations have been subjected to further study.

3.1 EM-scheme and Parameter Perturbation at DWD

The work has been described in Section 2.1.

3.2 SPPT and BLPERT schemes at MeteoSwiss

SPPT scheme

The SPPT scheme is used in the operational COSMO-E forecasts of MeteoSwiss. Several issues with SPPT could be solved within the SPRED priority project and a consolidated version of the scheme implemented by Lucio Torrisi (Italian Meteorological Service, Rome,

COMET) with support for GPUs is implemented in the official COSMO version 5.05. The main findings are:

- SPPT has significant impact only with large correlation lengths in space and time in the random pattern. COSMO-E thus uses 5deg and 6h, respectively. The sum of the parameterization tendencies for temperature and humidity are largest in summer and are then dominated by those of the turbulence scheme. This explains why SPPT enhances the spread mainly in summer, but hardly in winter;
- in order to avoid an effect on the total water content in the atmosphere by SPPT, all humidity variables (Qx) should be perturbed (itype_qxpert_rn=2). Other settings caused extensive precipitation in a few members during the test period;
- the COSMO advection scheme causes unphysical temperature anomalies in specific conditions and locations in complex topography. The model physics counteract against these anomalies. However, if SPPT reduces the physical tendencies the counteraction is weaker and thus less effective. Therefore, a new switch has been implemented in COSMO to switch off SPPT locally where temperature anomalies are diagnosed (ltargetdiff_mask=.true.).

BLPERT scheme

The boundary layer perturbation scheme (BLPERT) suggested by Kober and Craig (2016) includes stochastic perturbations related to turbulent fluctuations in kilometer-scale numerical weather prediction models. It has been implemented in a COSMO test version by Uli Blahak, DWD. Since COSMO-E shows a lack of convective precipitation in the Swiss Plateau, we tested whether the scheme is able to enhance the triggering of convection based on case studies for both strongly and weakly forced synoptic weather situations. The experiments confirmed the potential of the BLPERT scheme to improve forecasts in weakly forced cases by intensifying the triggering of convection as shown in Kober and Craig (2016). However, the factor that controls the amplitude of the perturbations, blpert_const, has been set to twice the value suggested by the authors to reveal a significantly positive impact in our case. More importantly, the scheme has a clearly negative impact on the convection characteristics in our cases with rather strong forcing. It produces too many small-scale unorganized cells with short lifetime and breaks up larger cells and squall lines, also with the recommended setup. Therefore, the benefit of the scheme in the current version is questionable for COSMO-E.

General remarks

Model perturbations by BLPERT or SPPT have an impact on the physical processes that keep a convective system alive and they can be disruptive. The chances that perturbations are disruptive are particularly high with BLPERT with new random numbers every 10 minutes, but it seems to be a general problem of stochastic model perturbation schemes in convection-resolving ensembles that are based on an additional term for prognostic model equations. Hence, the impact of new model perturbations on the convection characteristic or on physical processes in general, should be carefully investigated.

3.3 Stochastic pattern Generator for model perturbation at RHM

A Stochastic Pattern Generator (SPG) has been developed at the Hydrometcenter of Russia within the framework of PP KENDA (Tsyrulnikov, Gayfulin, 2016, 2017). The goal of this study was to try SPG for generation of model-error perturbations in the COSMO-Ru2-EPS.

SPG in a nutshell

The SPG produces Gaussian pseudo-random 3-D and 2-D spatio-temporal fields with realistic and tunable structure on a limited-area domain. The SPG benefits from the fact that it generates fields with realistic space-time interactions (the spatio-temporal covariances obey the "proportionality of scales" principle: larger (shorter) spatial scales are associated with larger (shorter) temporal scales (Tsyroulnikov, 2001). The SPG model is

$$(\frac{\partial}{\partial t} + \mu\sqrt{(1-\lambda^2\Delta)})^3\xi(t,s) = \sigma\alpha(t,s))$$

where $\xi(t,s)$ is the model-error random field to be generated, α is the white in space and time driving noise, t is time, s=(x,y,z) is the spatial vector, Δ is the spatial Laplacian, σ controls the variance, λ controls the spatial length scale, and μ controls the time scale. The SPG was first coded as a stand-alone version; it was later embedded in COSMO model versions 4.22 and 5.01. Modifications of the model code include the supplemental new modules mod_spsp.f90 and src_gen.f90; changes in some old modules, and introduction of additional tuning parameters in INPUT_ORG namelist. The SPG is easily tunable. Its main tuning parameters are defined in the namelist INPUT_ORG:

- the standard deviation of the random field (separate for u, v, T, qv: {u,v,T,qv}_rand_field_std);
- the distance in km at which the spatial correlation falls to 0.5 (separate for u, v, T, qv {u,v,T,qv}_L05);
- the time in sec when the temporal correlation falls to 0.2 (separate for u, v, T, qv: {u,v,T,qv}_t02).

In our experiments we varied these parameters to get a good spread/skill relation. The SPG was applied in the additive mode. The random field $\xi(t,s)$ was generated on an SPG coarse grid (with 64 grid points in x and y directions and 32 levels) and tri-linearly interpolated to the model grid. For each element F (i.e. u, v, T) its individual perturbation field rand_field on the model grid was recalculated every timeslice=1200s. The perturbations were added at each model timestep as follows

$F_{pert} = F_{unpert} + rand\_field/timeslice.$

That is, the perturbations were smoothed in time and space and were different for each element and level. Humidity was not perturbed.

Experimental set up

Experiments were carried out with COSMO-Ru2-EPS system developed for the Sochi Olympics within the CORSO project (Montani et al, 2014). The following experimental set up was applied:

- Δx= 2.2 km, L50, 10 members;
- IC&BCs from COSMO-S14-EPS (a clone of COSMO-LEPS for Sochi region, Δx=7km, L40);
- forecast length 48h;
- period: February 1-7, 2014 (00UTC);

- domain: Sochi area (172*132 gridpoints);
- verification against observations (∼40 stations) using VERSUS;
- model version 5.01 (with SPG implemented);
- single precision runs.

Each set of experiments included reference experiments and experiments with different sets of the SPG parameters. There were two reference experiments:

- experiment NOPERT without model perturbations at all (both SPPT and SPG were switched off);
- experiment SPPTSW with SPPT turned on and the SPPT parameters recommended by MeteoSwiss (hinc_rn=6 h; dlat_rn=dlon_rn=5 deg; stdv_rn=1.0; range_rn=0.8; itype_qxpert_rn=2 ; tapering in the stratosphere, no tapering near the surface).

The focus was on the possibility of increasing the T2m spread with SPG-generated model perturbations. Temperature and wind spreads in the free atmosphere were considered as well. It was demonstrated that the SPG perturbation patterns were sensitive to pre-defined spatial and temporal correlations ({u,v,T}_ L05 and {u,v,T}_t02 parameters). However, no definite conclusions could be drawn from that analysis. Therefore, most experiments with SPG were run with fixed spatial and temporal correlations: we assumed that for random fields of all variables (T, u, v) the spatial correlation was 0.5 at a distance of 50 km ({u,v,T,qv}_ L05=50 km) and the temporal correlation becomes equal to 0.2 in 4800s {u,v,T,qv}_ t02=4800s). Then we varied the random field standard deviations, specifying them separately for temperature and wind fields. First experiments demonstrated that perturbations with standard deviations about 0.02-0.03 m/s per hour for wind and 0.01-0.02 K per hour for temperature had a negligible effect on the T2m ensemble spread. After some additional tests for individual dates, two versions of SPG setting were selected. The experiment SPGBIG2 was run with the standard deviations for wind u,v_std=0.15 m/s per hour and for temperatures T_std=0.045 K per hour. bit greater deviations were specified in the experiment SPGBG: u,v_std=0.21 m/s per hour, T_std=0.06 K/h.

Experimental set results

We ran the experiments NOPERT, SPPTSW, SPGBIG2 and SPGBG for 7 days and examined the behavior of ensemble mean and spread for temperature, wind, and mean sea level pressure. The application of both SPPT and SPG resulted in an increase of the ensemble spread. In the free atmosphere, the increase was much greater for SPG both in temperature (see Fig. 24) and wind (not shown). Fig. 24 demonstrates the high sensitivity of ensemble spread to the standard deviations of the random fields generated by SPG. This implies that tuning the SPG setup is a tricky task.

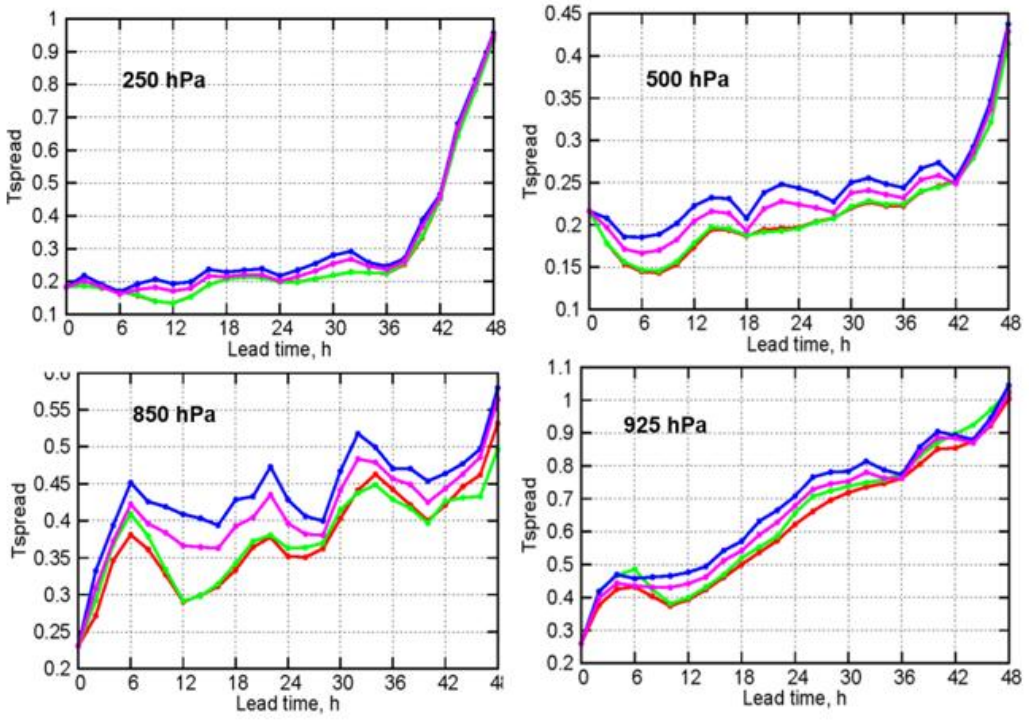

Figure 24: Temperature spread as a function of forecast time for experiments NOPERT (red), SPPTSW (green), SPGBIG2 (magenta), SPGBG (blue). Forecast start time is February 6, 2014 00UTC.

For individual days, the T2m spread with SPPT and SPG were comparable (see Fig. 25). The results of experiment SPGBG (with greater standard deviations of the SPG-generated random fields) seem to be superior.

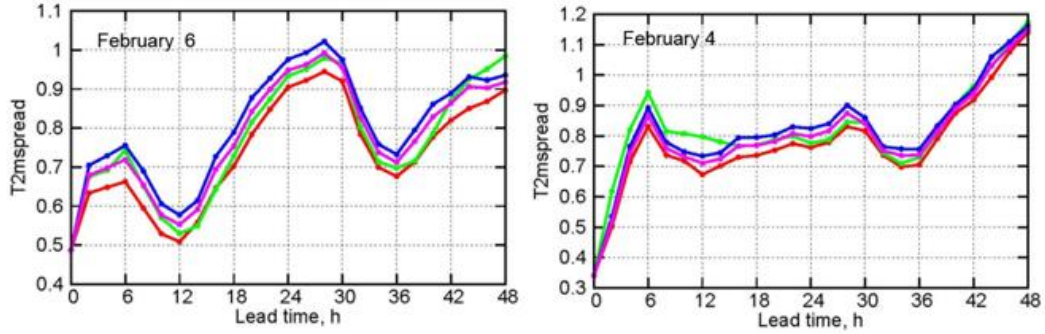

Figure 25: T2m spread as a function of forecast time for experiments NOPERT (red), SPPTSW (green), SPGBIG2 (magenta), SPGBG (blue) for forecasts starting on February 6, 2014 00UTC (left) and February 4, 2014 00UTC (right).

The ensemble forecasts of T2m were verified against the station data using VERSUS for the period of 7 days (February 1-7, 2014, 00UTC). In Fig. 26 the T2m spread for SPPT is slightly bigger than for SPG, while the latter slightly improved the T2m error. Note that all the above presented spread plots demonstrate the results of averaging over all grid points using the grib_api tools. On the other hand, VERSUS interpolates the forecasts to the points

of observations and the spread in Fig. 26 is averaged over observation points only.

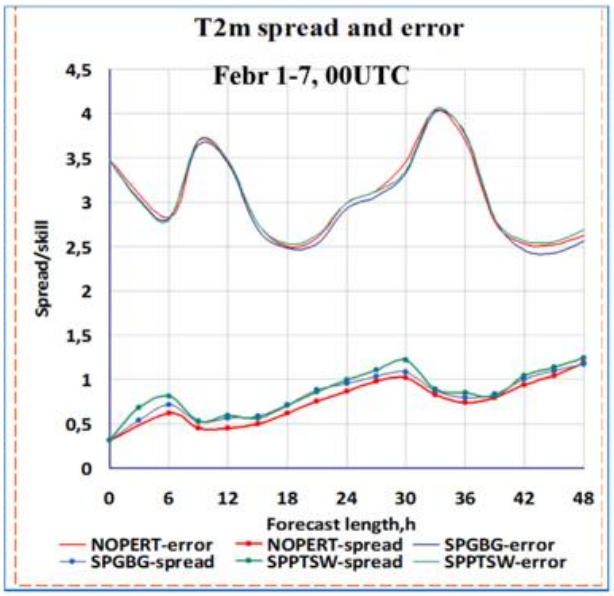

Figure 26: The error (upper curves) and the spread (lower curves) of T2m temperature for experiments NOPERT (red), SPPTSW (green) and SPGBG (blue) calculated for a 7-day period starting from February 1, 2014 00UTC. Sochi area.

Fig. 26 implies a possibility of a further increase of T_std and u,v_std in the SPG setup. An additional experiment with T_std=0.09 K/h and u,v_std= 0.3 m/s per hour was carried out. The T2m spread increased, but the effect on the free-atmosphere temperature spread was much greater, which does not look physically reliable. Thus, it was suggested to revise the way SPG is used for model perturbations and now the work continues within the APSU Priority Project.

Conclusions

- The application of SPG for perturbing model in COSMO-Ru2-EPS looks promising and results in increasing the ensemble spread;
- the application of SPG improves T2m forecasts;
- the ensemble spread is very sensitive to the SPG settings. Even a small increase in the standard deviation of the random fields leads to a spread growth. Variations in correlation lengths and times also have a visible effect but it is not so unequivocal and will be addressed later;
- in the free atmosphere, the temperature and wind spread is bigger with SPG than with SPPT;
- for T2m, the ensemble spreads in experiments with SPG and SPPT are comparable.

A new way of applying SPG (in which the number of parameters to be specified and tuned is reduced) will be considered in the Priority Project APSU.

4 Test and development of perturbation for the model lower boundary

The perturbation of the model lower boundary (surface/soil) has also been further developed, consolidating the methodologies developed in the COTEKINO PP.

4.1 Soil moisture perturbation in a complete ensemble set-up (ARPA Piemonte)

In the framework of the COTEKINO PP, a method for the perturbation of the soil moisture has been developed by Arpa Piemonte (Bonanno and Loglisci, 2018). The method is based on the Stochastic Pattern Generator developed by RHM (Tsyrulnikov and Gayfulin, 2016).

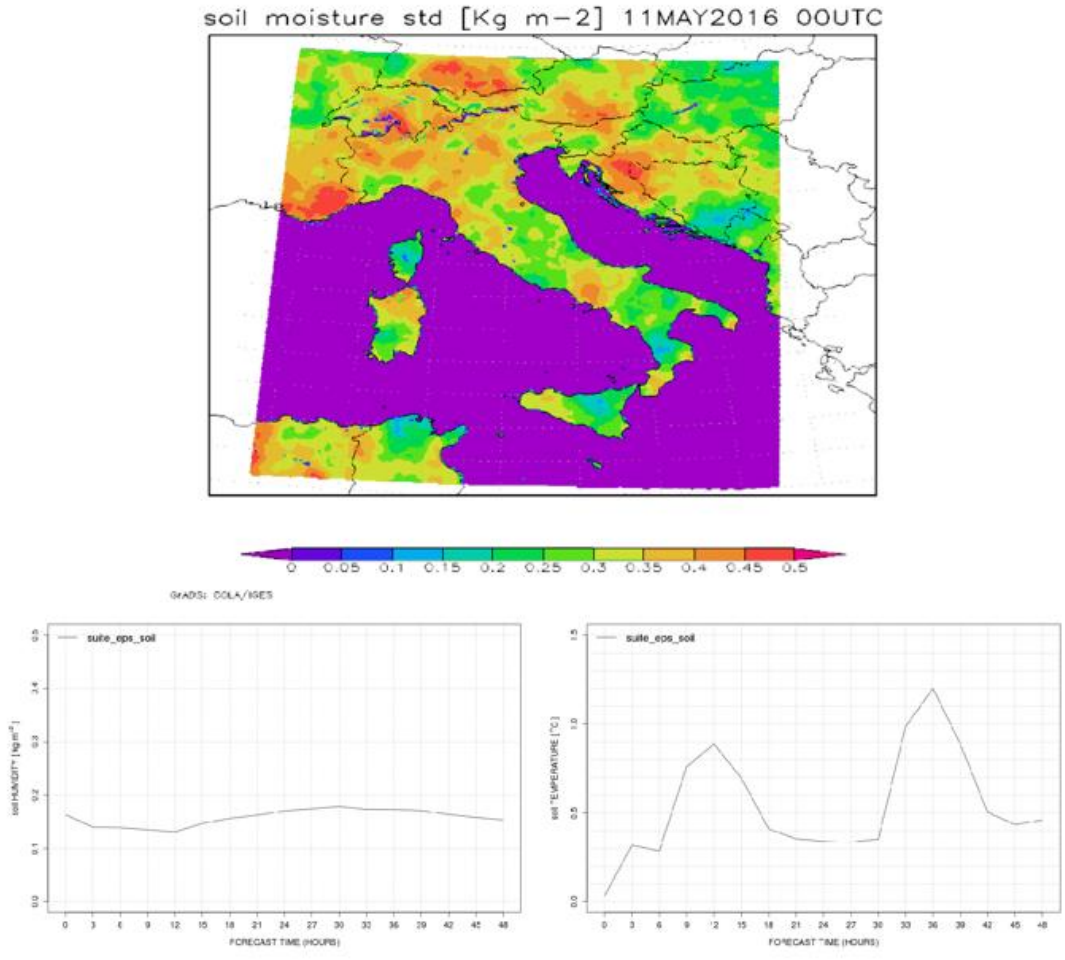

Figure 27: Soil mositure perturbation in the initial condition for the event of the 11 of May 2016 (top panel). Temporal evolution of the spread in terms of soil moisture (bottom left) and soil temperature (bottom right).

In the SPRED PP, the purpose of the task was to assess the benefit of the methodology by introducing the perturbation in an ensemble with "complete set-up", with perturbed Ini-

tial and Boundary Conditions, as well as physics perturbations. An experimental suite of COSMO-IT-EPS was created for this purpose. The ensemble was run in a complete ensemble set-up: Initial and Boundary Conditions from ECMWF ENS, SPPT and Parameter Perturbation applied to the COSMO model. The suite was run with and without soil moisture perturbation, for the month of May 2016. This period was selected due to the occurrence of several convective events over Italy. A statistical analysis of the results has not been completed due to a lack of resources. As an example of the perturbation applied, the field of the soil moisture perturbation is shown for the event on the 11 of May 2016 (top panel), together with the temporal evolution of the spread in terms of soil moisture (bottom left panel) and soil temperature (bottom right panel).

4.2 Soil parameters and soil temperature perturbation (IMGW)

At IMGW, the ensemble forecast was prepared in such a way that each member of the ensemble starts with a different value of the initial state of c_soil, the parameter describing the fraction of evaporating soil, and of soil surface temperature. The c_soil parameter is varied using the new random number generator. In addition to the variation of parameters that describe the state of soil surface, the collection efficiency coefficients (eff-coeff, see below) in the clouds were also perturbed to add perturbation in the physical parameterization of precipitation processes. The following perturbations were considered:

(a) *c_soil*: perturbation of a parameter describing evaporation from soil using RNG (described in Section 5.2);

(b) *eff-coeff*: perturbation of the collection efficiency coefficient;

(c) *eff-c_soil*: perturbation of the collection efficiency coefficient together with evaporation from the soil surface;

(d) *laf-pert*: perturbation of the surface temperature of the soil;

(e) *laf-c_soil*: perturbation of soil surface temperature in the set of initial conditions simultaneously with evaporation from the soil surface;

(f) *laf-eff*: perturbation of the soil surface temperature (as in (e)) together with a variation of the collection efficiency coefficient;

(g) *eps-all*: perturbation of all the above quantities (fields and parameters) at the same time.

The warm period (June-September) of 2013 was chosen for numerical simulations. The study was focused on ensemble forecasts of three meteorological fields: air temperature at 2 m above ground level (TE2M), dew point temperature at 2 m (TD2M), and wind speed at 10 m (U10M). For each meteorological field, the monthly average values of the spread (standard deviation of forecasts, calculated using all the members of the ensemble) and an indicator of skill S (mean absolute error, MAE, of the forecasts) were calculated (Jolliffe and Stephenson, 2012). In the next step the average values of spread and skill were compared with the average values of air temperature, the deviation of the average monthly air temperature in a given month in relation to the long-term average of 1971-2000 (Bulletin, 2013), the monthly precipitation sum and deviation of the monthly sum of atmospheric precipitation as a percentage of the long-term norm for 1971-2000 (Bulletin, ibid.). The results were cross-checked for correlation with the average values of spread and skill for the given type of soil that occurred in the analyzed area. The most effective perturbations (combinations) in

Period/Field	TD2m	T2m	U10m
June	eff-coeff	?	?(∼eff-c-soil)
July	c-soil	eff-coeff	?(∼laf-c-soil)
August	eff-coeff	eff-coeff	?(∼laf)
September	c-soil	eff-coeff	?
Jun-Sep	laf-c-soil	eff-coeff	?

Table 1: Most effective perturbations, see explanations in text above. Values averaged for June to August 2013. The ? means hard to establish, the ∼ means small tendency.

terms of skill and spread, averaged over the period of study, are presented in the following table.

The following figures present the most effective perturbation schemes in terms of skill (MAE) and spread on the period June-August 2013.

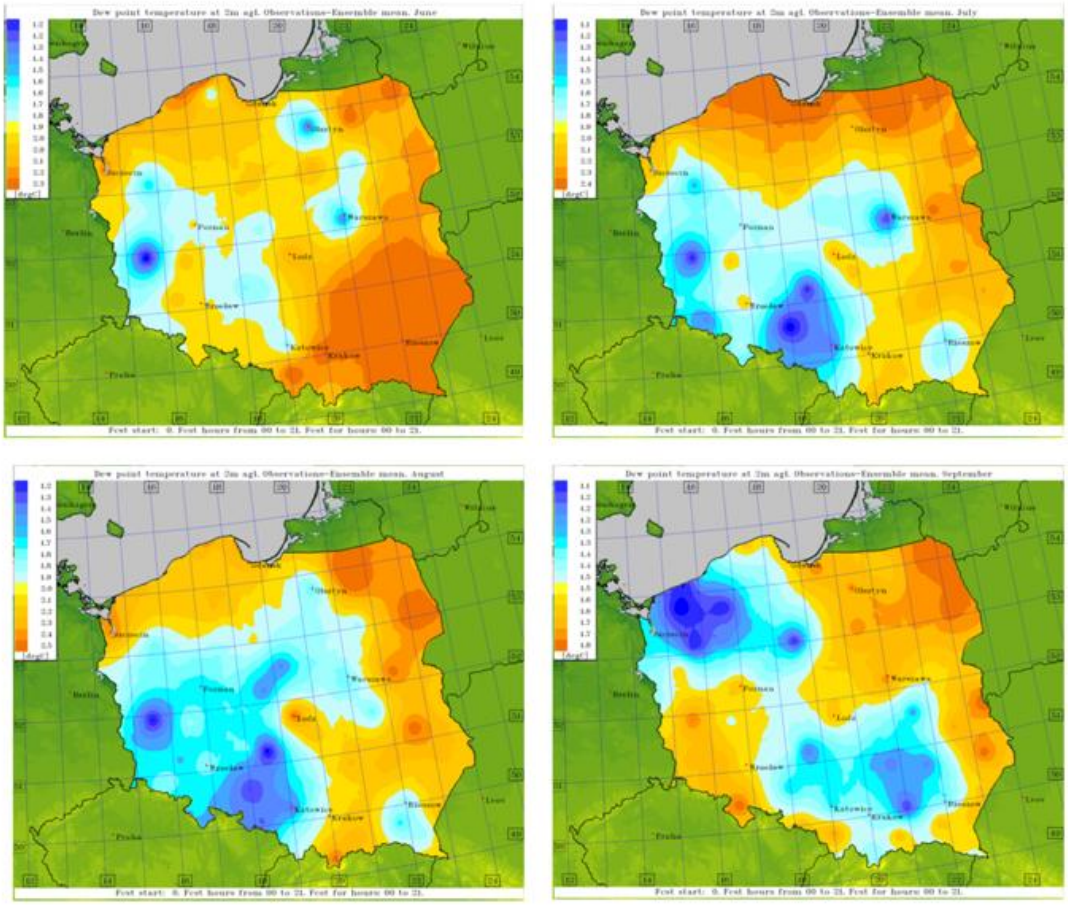

Figure 28: MAE (skill) spatial distribution for TD2M - most effective schemes. Upper left - June (eff-coef), upper right - July (c-soil), lower left - August (eff-coef), lower right - September(c-soil).

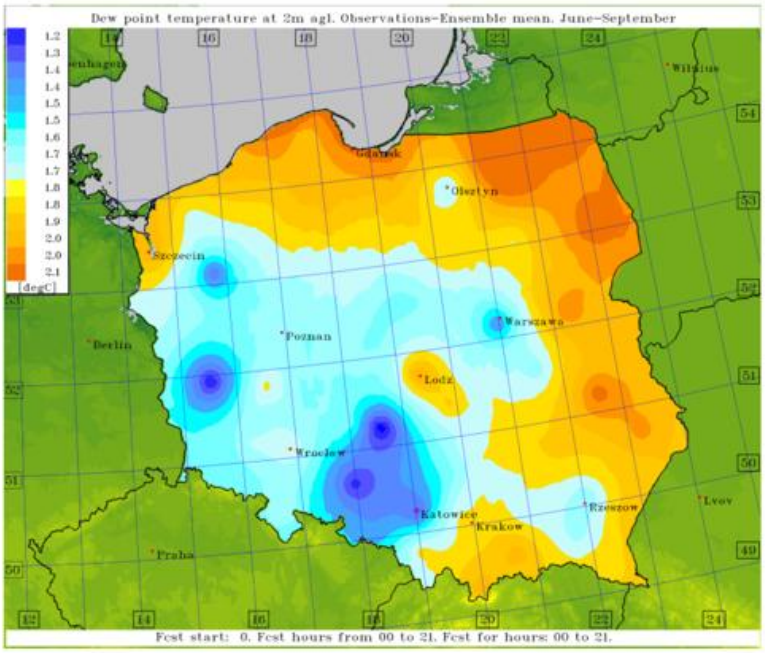

Figure 29: MAE (skill) spatial distribution for TD2M - most effective scheme for the entire period June-September 2013 - laf-c-soil.

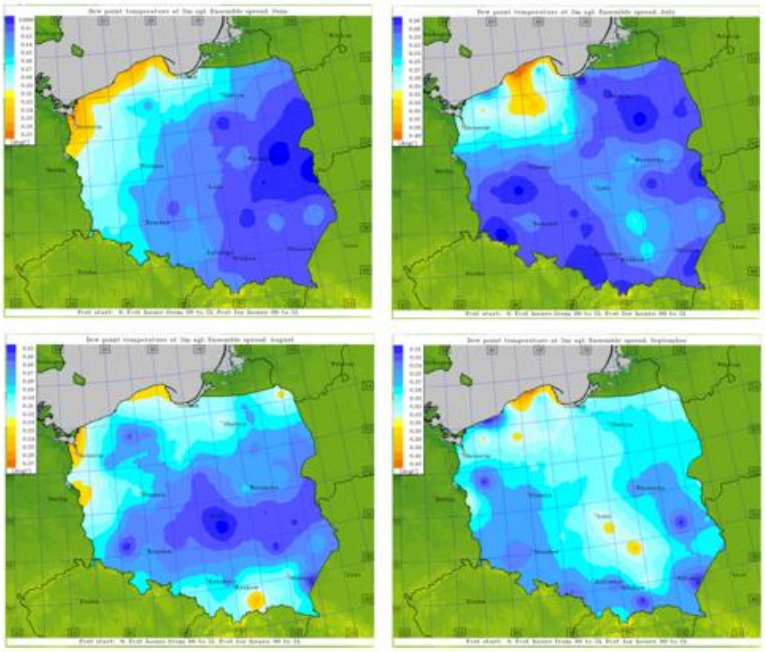

Figure 30: Spread spatial distribution for TD2M - most effective schemes. Upper left - June (eff-coef), upper right - July (c-soil), lower left - August (eff-coef), lower right - September(c-soil).

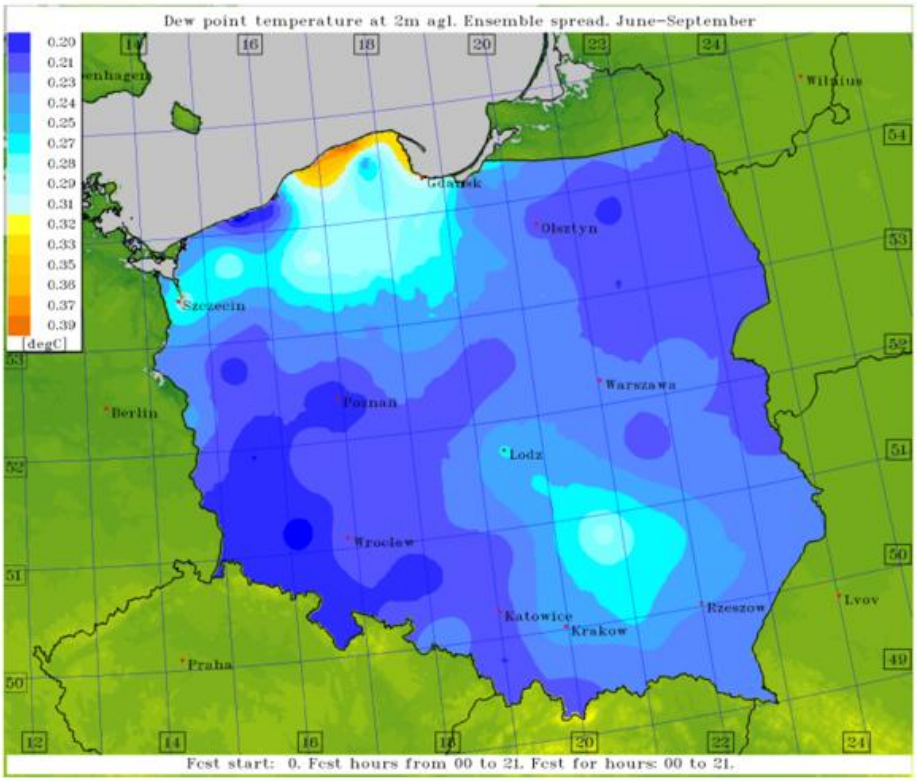

Figure 31: Spread spatial distribution for TD2M - most effective scheme for the entire period June-September 2013 - laf-c-soil.

Conclusions drawn from this part of study were as follows:

- values of spread/skill are in general independent on soil type;
- average values of spread of dew point temperature/air temperature/wind speed are inversely proportional to average monthly air temperature and to deviation of the mean monthly air temperature (T and ΔT) climatologically calculated for the period of 1971-2000;
- if long-term monthly sum of precipitation (P) is above a climatologically calculated average then the average values of spread of TD2M, T2M and U10M are low (EPS results are highly underdispersive);
- the correlations between the values of skill of TD2M, T2M, U10M and T, ΔT, P, and ΔP are similar to the ones between spread of TD2M, T2M, U10M and T, ΔT, P, and ΔP;
- average value of skill is independent on type of disturbed field or parameter. The distribution of the average values of skill is more or less similar to each other.

5 Ensemble post-processing

5.1 Development of probabilistic products at COMET

The work of generation of probabilistic products for the forecast of thunderstorms and fog is performed in the framework of the SRNWP-EPS II Project of EUMETNET. The main aim of the Project is to contribute to build very high-resolution ensemble systems in Europe, resolving the convection-permitting scale phenomena. First a review of literature on post-processing methods for prediction of fog and thunderstorms, with focus on ensembles, has been performed. Then the development of a software dedicated to this purpose has started, in particular a product for probabilistic prediction of fog has been implemented. The developed FORTRAN code provides surface visibility forecast starting from standard NWP models outputs. The following methods have been selected and implemented:

- "Zhou method" (Zhou, 2011). It is a new method, working properly for radiation and advection fog, based on the asymptotic liquid water content (LWC) vertical distribution. The asymptotic LWC distribution is a consequential balance among cooling, droplet gravitational settling and turbulence in the liquid water budget of radiation fog. Including the advection term, it is possible to write a partial differential equation as the governing equation for LWC [g/kg], with appropriate initial and boundary conditions for the one it is proposed an asymptotic solution that can be used to compute visibility through the Gultepe and Milbrandt (2007) formula;

- "Boudala method" (Boudala et al., 2012). This algorithm is based on the Stoelinga and Warner empirical equation that allows to compute visibility using information about relative humidity and dew point temperature;

- "LWC algorithm": it is the traditional approach that computes horizontal visibility using model forecastof specific cloud water content;

- "UPS method" (Baker et al., 2002). It is an empirical method proposed by UPS Airlines forecasters to forecast the radiation fog. It is based on the comparison of the low level atmospheric temperature and the so called crossover temperature, defined as the minimum of dew point during the warmest daytime hours. Following this approach it is possible to determine whether a station will radiatively cool enough for saturation. In a second step, in order to assess if boundary layer turbulence supports the fog formation process, a quantitative index of boundary layer turbulence mixing is used.

The code takes also into account the precipitation contribution to visibility reduction through an empirical formula.

Visibility forecast from different methods using deterministic COSMO-ME (5km hor. resolution) and COSMO-IT (2.8 km hor. resolution) model outputs have been compared with SYNOP observation over the Italian domain. The Zhou method seems to be the best candidate because it is able, in principle, to forecast both radiation and advection fog, but, as one of the trigger condition is the relative humidity from the NWP model output greater that 95 % (or other established threshold), results are highly influenced by this parameter. Looking at the results, the pattern of visibility seems to be much more reliable with Zhou with respect to the other methods, even if the visibility values seem to be underestimated (further tuning is needed through the configuration namelist). In order to overcome the sensitivity to the forecasted relative humidity (generally underestimated by the model), a combination of Boudala and Zhou methods has been proposed and used operationally at

the Italian AirForce Met Service. As an example, visibility forecasts with different methods applied to COSMO-ME deterministic model and COSMO-ME EPS (7 km resolution) probabilistic model, for a test case of 18 October 2017 06UTC (low visibility over Pratica di Mare airport), are shown in the figures below. Only Boudala and Zhou approaches forecast fog/mist in this region, in particular, Zhou approach has better correspondence with SYNOP observations.

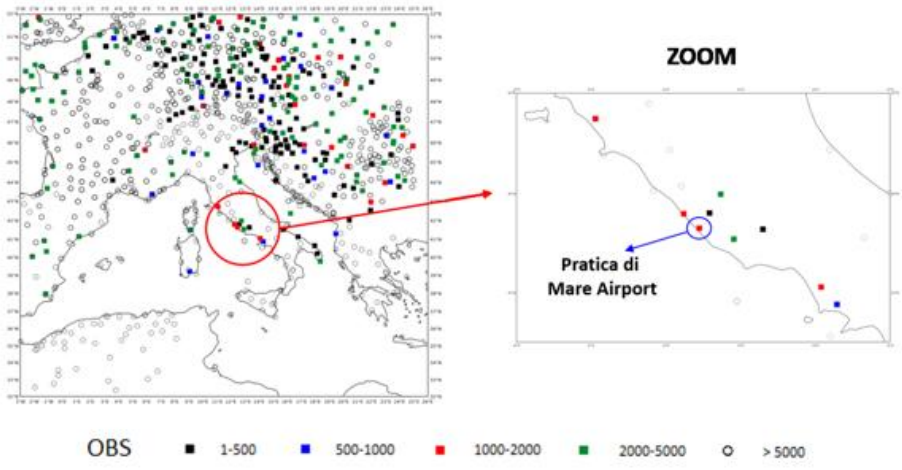

Figure 32: Test case of 18/10/2017 (intense fog event at Pratica di Mare airport): distribution of synop observations from 05 UTC to 07 UTC.

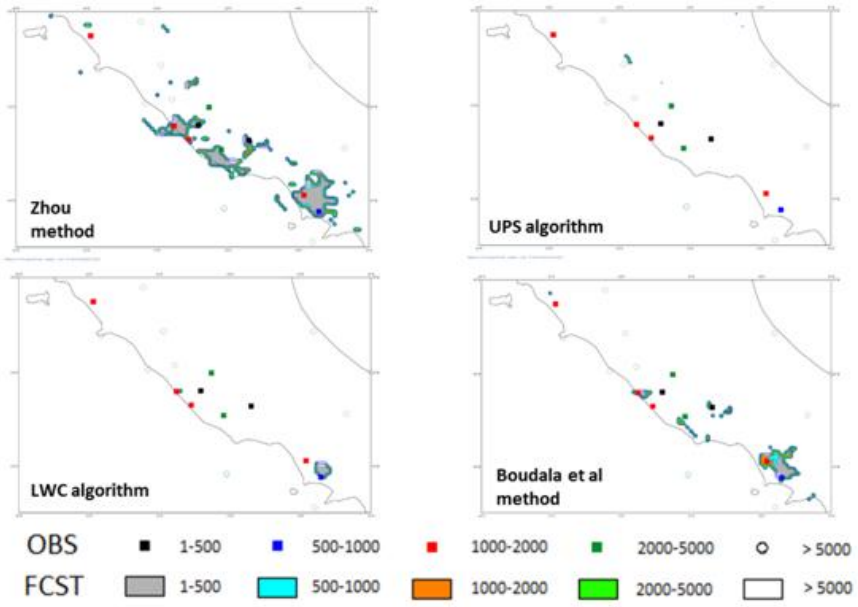

Figure 33: Test case of 18/10/2017 (intense fog event at Pratica di Mare airport): visibility forecast from different methods applied to the COSMO-ME (5 km) model outputs (run 00UTC of 18/10/2017, fc+6h).

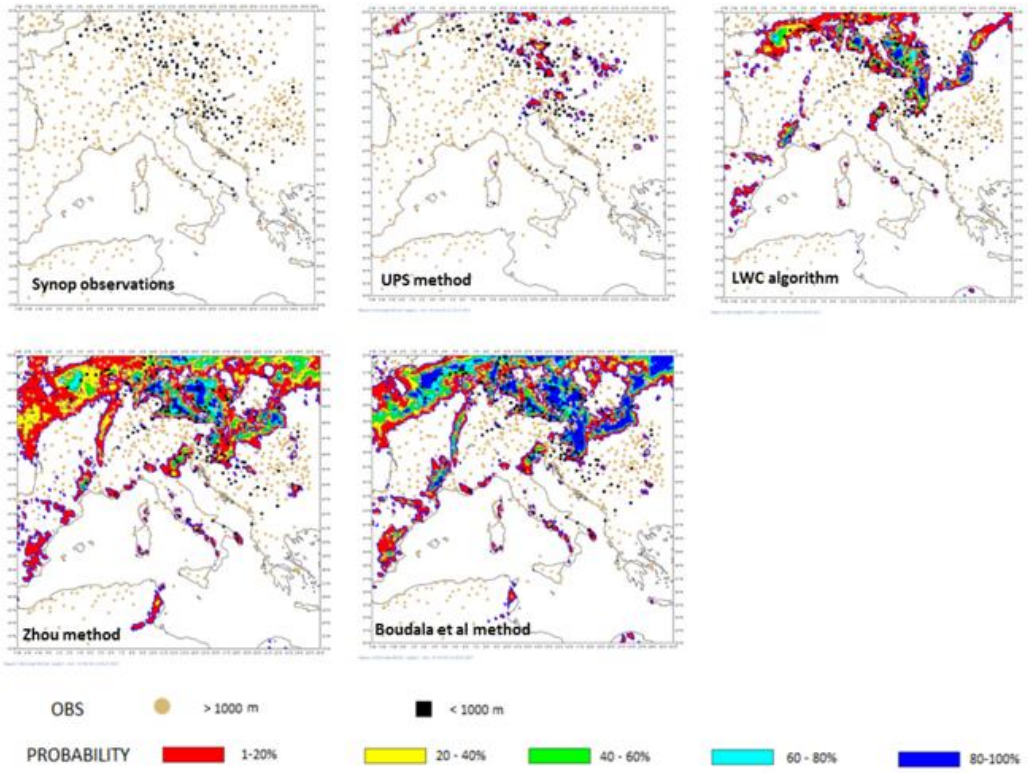

Figure 34: Test case of 18/10/2017: Forecasted probability of fog (visibility ¡ 1000 m) from different methods applied to the COSMO-ME-EPS (7 km) model outputs (run 00UTC of 18/10/2017, fc+6h).

Verification over an extended period has been planned together with the subjective verification from operational forecasters.

5.2 Calibration of ensemble output at IMGW

A review of literature concerning ensemble calibration methods and alternative perturbation patterns/random number generators has initially been performed. First, a new (alternative) Random Number Generator (RNG) has been introduced.

Alternative random number generator

Motivation for new (modified) initialization or seed of RNG was that seeding of RNG commonly is based on machine time (in general, milliseconds). This may results in seed(s) being identical for all processes (threads). This would lead to the situation, that with increasing number of parallel threads the probability of an occurrence of identical seeds would significantly increase. Having four hundred threads (like in case of IMGW EPS operational setup) vs. 999 milliseconds, it becomes more and more probable that some threats have an identical seed. This would results in an identical chain of RNG numbers generated, which in turn means a very poor stochasticity of EPS. One should also realize that using fast machines does not have 999 milliseconds, but much, much less. So, another initialization scheme has been selected combination of machine time with a thread ordinal number as in the figure below.

Spread values	*Ctrl (operational RNG)*	*Modified RNG*	*Ctrl (operational RNG)*	*Modified RNG*	*Ctrl (operational RNG)*	*Modified RNG*
	T2M (K)		**TD2M (K)**		**RH2M (%)**	
Mean	0.108191	0.238555	0.118675	0.272361	0.705627	2.120926
Max	2.262	2.458	3.284	3.536	12.261	14.758
	U10M (m/s)		**Surface pressure (hPa)**		**Total precipitation (mm)**	
Mean	0.139599	0.180653	0.023892	0.027456	0.286905	0.379897
Max	2.041	2.903	0.747	0.652	13.203	18.515

Table 2: Average spread values for June 2013, reference (operational RNG) vs. modified RNG.

```
type(random_state_t),intent(inout)::akrngs
integer::time(8),iseed
double precision::dprn1,dprn2
type(random_state_t)::krngs
! Init
! Get sys. time with non-zero millisecs. part
  time(8)=0
  do while ( time(8) .eq. 0 )
    call date_and_time(values=time)
  end do
! Calculate 1st seed based on milliseconds
  iseed = int( dble(time(8))... )
! Initialize temporary RNG  from mo_random
  call construct(krngs,iseed)
! Draw 1st RN  from mo_random
  call random_number(dprn1,krngs)
! Destroy temporary RNG  from mo_random
  call destruct(krngs)
! Calculate 2nd seed based on thread rank
  iseed = int( dble(icrnk+1) ... )
! Initialize temporary RNG  from mo_random
  call construct(krngs,iseed)
! Draw 2nd RN from mo_random
  call random_number(dprn2,krngs)
! Destroy temporary RNG from mo_random
  call destruct(krngs)
! Get 3rd RN: add 1st & 2nd RNs, normalize
  dprn1 = dprn1 + dprn2
  if(dprn1.gt.1.0d0) dprn1 = dprn1 - 1.0d0
! Calculate 3rd target seed based on 3rd RN
  iseed = int( dprn1... )
! Initialize target RNG using 3rd seed
  call construct(akrngs,iseed)
```

Figure 35: Block scheme of new seeding for RNG.

This procedure was tested first using archive data, with no time-lagged ICs/BCs to avoid influence from side effects other than change of seed. An average value of spread significantly increased as shown in the Table 2 below.

An example of spatial distribution of spread with/without new seeding scheme is shown in the Fig. 36.

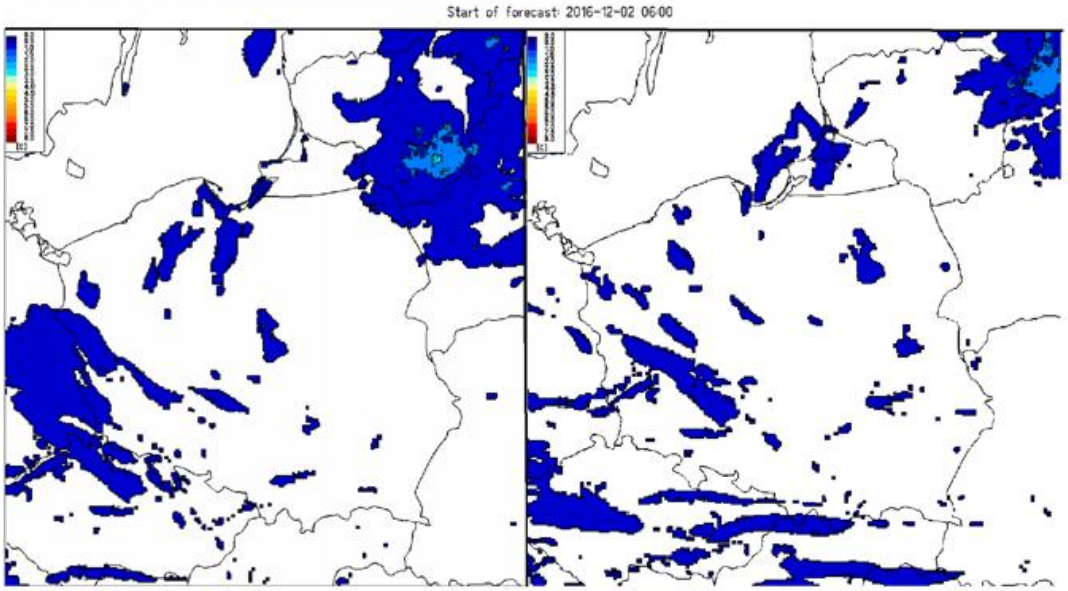

Figure 36: T2M spread for new (left) and old (right) seeding of RNG; 12 hour of forecast, December 2nd, 2916, 06:00 UTC.

This procedure is very efficient in terms of spread value (increasing average spread even 3 times comparing to "standard" RNG) especially on fast machines. It is also more realistic as far as spatio-temporal distribution of spread is concerned. At IMGW, it became operational in January 2017.

Post-processing multi-linear regression, logistic regression, Artificial Neural Networking

The most common procedure for calculating EPS mean is to use an arithmetic average as in the equation below (further Simple Mean):

$$\langle y \rangle = \frac{\sum_{i=1}^{m} x_i}{m}$$

An alternative approach can be described as (multi-)linear regression (MLR), where previously prepared forecasts from all members (X) of an EPS are compared with observations (learning phase). Then weights (β) of every member is computed to assess, how "important" for overall EPS-mean is a selected member (how big the contribution to EPS-mean from the member should be). The weights are calculated e.g. via least squares method, any potentially important factor(s) like geographical coordinates, terrain elevation etc. may also be considered. The following equation is a representation of the procedure, with an additional condition of normalization of weights to one. The "new" EPS mean y is calculated as a scalar product of vectors X and β:

$$\langle y \rangle = \vec{X} \cdot \vec{\beta}$$

Finally, an Artificial Neural Network (ANN) was studied to test if this approach could be a valid replacement for Simple Mean or MLR Mean. Basic idea is that the weights, with the same scheme as in MLR calculation, are calculated (learned) using a concept of layer of neurons, properly activated in learning phase. Basically, the representation of an ANN is presented in the following figures.

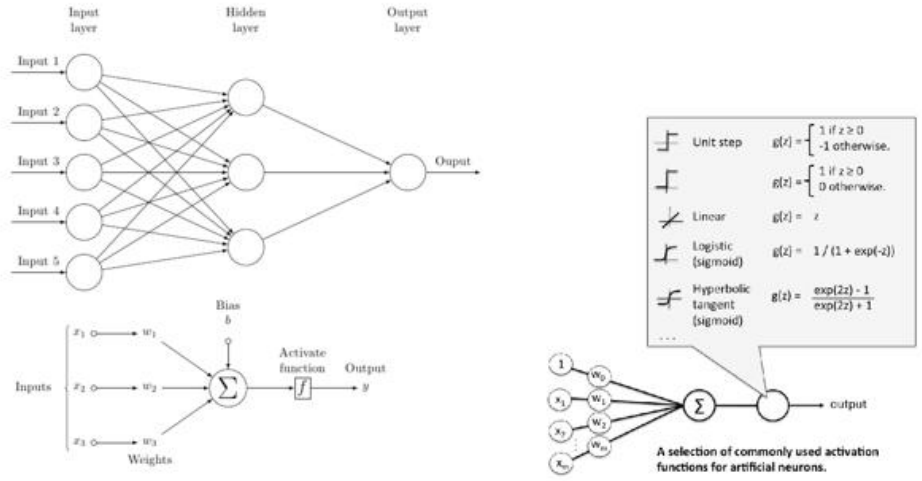

Figure 37: Basic concept of the Artificial Neural Network.

Pros and cons of the latest method are as follows: **Pros**:

1. ready-to-use dedicated software with source codes (FORTRAN);
2. sophisticated but elegant and intuitive concept;
3. improvement in preliminary case study observed;
4. forecasts are improving with the extension of learning period.

Cons (with answers to mitigate the cons):

1. complicated pre- and post-processing (answer: once the procedure is set-up and established, it can be further on done in an automatic way);
2. need for big data sets (archives), and for relatively huge computational resources (answer: this is a common feature of numerical forecasting - the need for huge volumes of data is a routine and a normal practice);
3. long computational time for training (answer: to avoid delays in operational post-processing, it can be done more or less frequently, with the acceptable quality of computations).

Results for case studies and for more extended periods (quasi-operational) are presented in the following tables and figures. First, case study for August 2017, with assorted number of predictors (input neurons) was analyzed. Learning period is July 1st, 2016 - July 31st, 2017, testing period is August 1st, 2017 - August 31st, 2017. The outcomes are presented in Table 3.

The spatial distribution of the average MAEs are presented in the Fig. 38.

Fields	Input Neurons* → / MAE ↓	24	22	20	Simple avg.**
U10m	Avg.	0.409	0.416	0.430	1.373
	Max	1.324	1.361	1.538	2.519
T2m	Avg.	0.266	0.275	0.451	2.606
	Max	0.924	1.144	1.302	3.628
Td2m	Avg.	0.268	0.305	0.365	1.736
	Max	0.906	0.999	1.238	2.006
PS	Avg.	2.398	2.405	2.595	2.864
	Max	11.683	11.464	9.708	11.786
TOT_PREC	Avg.	0.131	0.127	0.219	0.808
	Max	0.739	0.741	0.505	1.514

Table 3: ANN tests for August 2017, diversing number of neurons. * Input neurons: 20 - members (history, learning); 22 - 20 + geographical coordinates; 24 - 20 + geographical coordinates + forecast start + current hour. ** Simple averaging - 20 members mean (current forecast).

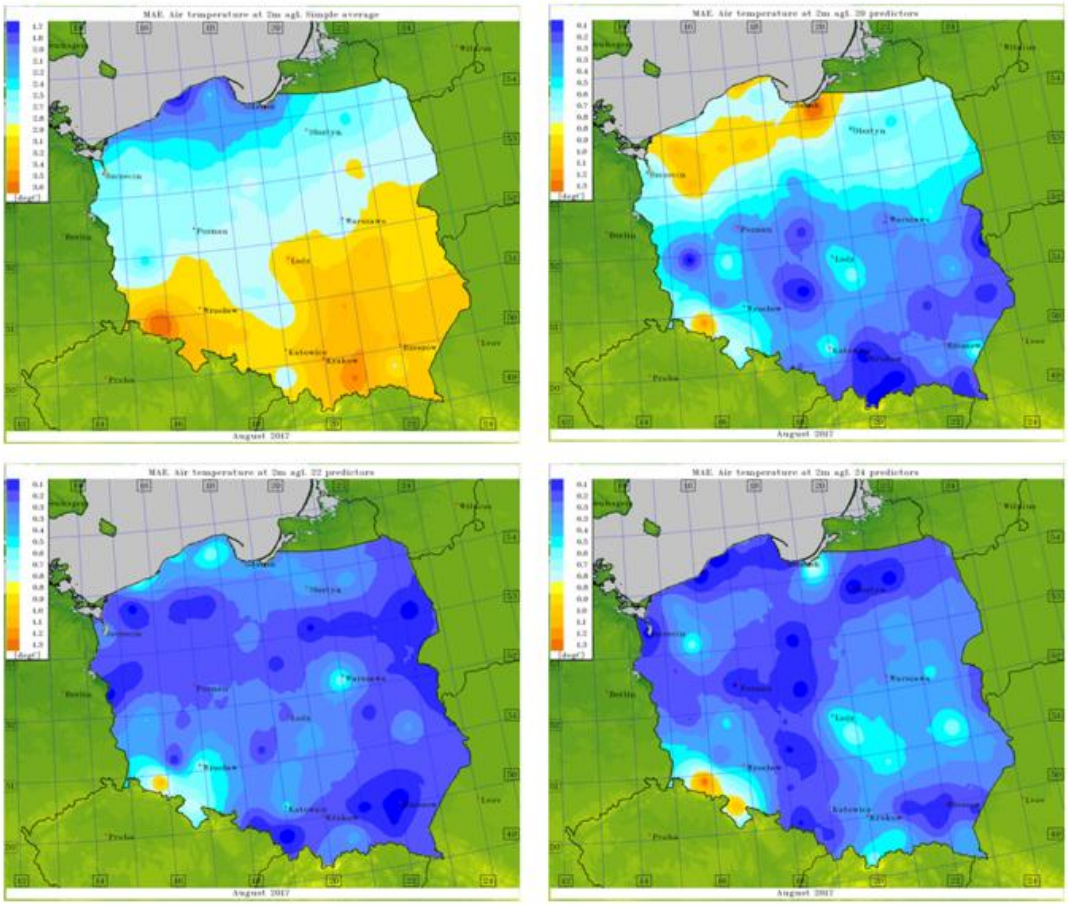

Figure 38: Results for different numbers of input neurons. Air temperature at 2m, mean absolute error, average values for August 2017. Upper left - reference (simple mean), upper right - ANN, 20 predictors, lower left - ANN, 22 predictors, lower right - ANN, 24 predictors.

Means	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
	Dew point			Air temp.			Wind speed		
Simple EPS	0.253	2.009	2.812	0.771	2.369	3.443	-0.618	1.737	2.297
MLR	-0.310	1.989	2.755	0.475	2.252	3.206	0.113	1.488	1.978
ANN	-0.244	1.981	2.750	0.066	2.214	3.135	-0.200	1.436	1.814

Table 4: ANN tests for April 2018, comparison with Simple Mean and MLR Mean.

For the further operational calculations the following setup has been chosen:

- 24 input neurons (20 members + λ, φ, ts, tc, with λ, φ - geographical coordinates, ts being start of forecast and tc - current hour of forecast in progress);
- 5 neurons in a single hidden layer (referring to 4 blocks of TL-ICs/BCs and spatial/temporal coordinates - blocked);
- every forecasted element, like temperature, wind speed, pressure, etc. is treated independently;
- activation function: hyperbolic tangent (symmetric with respect to 0,0, commonly used function);
- training method: backward propagation of errors (back-prop);
- optimization: gradient descent.

The following table and figures present an example results of EPS forecasts for April 2018, with the extended learning period (July 2016 - March 2018).

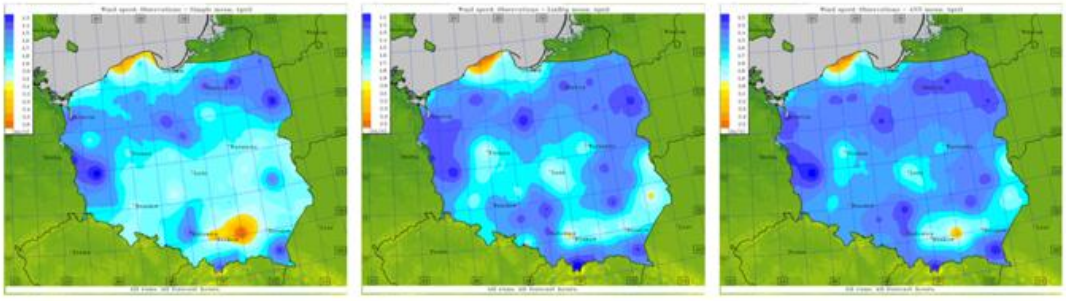

Figure 39: Results of different methods of computations of EPS mean. Left - observations vs. Simple Mean. Middle - observations vs. MLR mean. Right - observations vs. ANN mean. Wind speed forecasts, average values for April 2018.

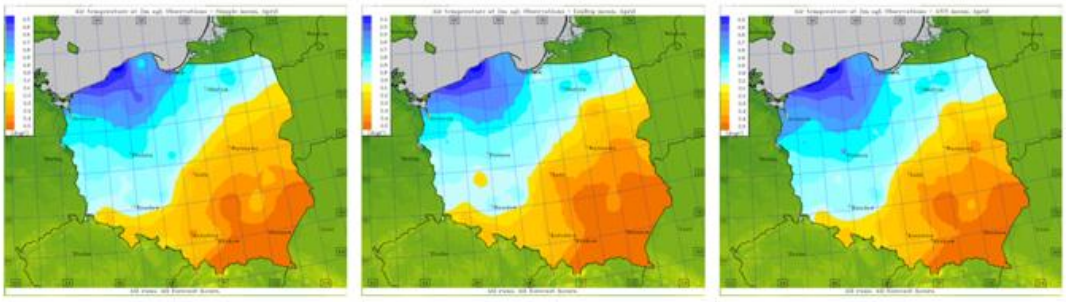

Figure 40: Results of different methods of computations of EPS mean. Left - observations vs. Simple Mean. Middle - observations vs. MLR mean. Right - observations vs. ANN mean. Air temperature forecasts, average values for April 2018.

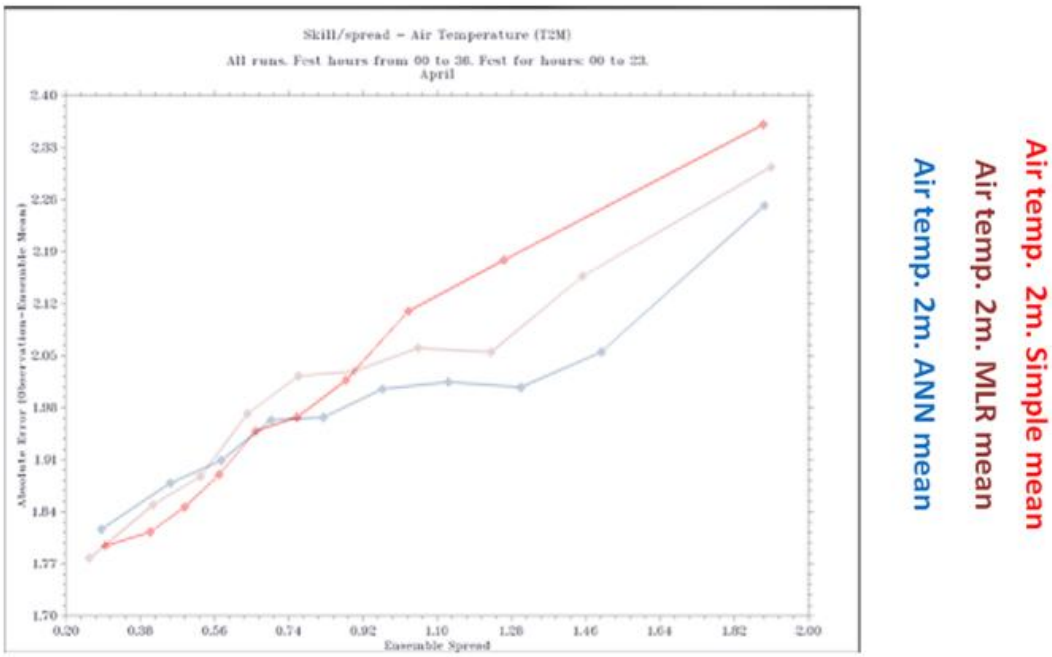

Figure 41: Skill-spread relation. Red - Simple Mean. Brown - MLR mean. Blue - ANN mean. Air temperature forecasts, average values for April 2018.

The general conclusions to be drawn from this study, was that significant improvement in forecasts (described in terms of EPS-mean vs. observations or forecast skill) can be seen. The longer learning period, the better forecasts are. Furthermore, the phrase "more predictors" - in general - means "better forecast", but also "longer calculations" - that is why some sort compromise should be established. ANN-based computation of EPS mean seems to be significantly better in comparison with Simple EPS Mean and with MLR mean, and that is why is has been chosen for further operational work.

6 Initial Conditions for the CP ensembles

The aim of this activity was to further study which methodology to follow to provide the ICs to the ensemble members. Experiments with ICs from KENDA compared with ICs from

cheaper options (downscaling) have been performed. It has also been performed a work on member selection from larger-scale ensembles.

6.1 Test of Initial Conditions from KENDA (DWD)

The KENDA-suite was running pre-operational providing initial perturbations for COSMO-DE-EPS. Two versions of BC perturbations are produced, with 20 members each (including physics perturbations): the known BCEPS multi-model approach and BCs from ICON-EPS (20 km grid size nest for Europe in a 40 km global EPS). Very preliminary results show an increase of spread together with a reduced RMSE of the ensemble mean for both 20 member setups. The resultingimproved spread-skill relation is still, but to a far less degree, underdispersive. The results have to be confirmed by longer verification periods and can be considered as first hints for improvement. The main results were an improvement in CRPS and partly RMSE of EPS mean for T_2M, gusts, and precipitation in summer and winter, particularly in the early forecast hours. An increase in spread for the variables lasts throughout the full forecast range of 27 hours. Improvements in thresholds-based BSS could be shown, where KENDA improved the resolution mainly. (Additionally, replacing BCEPS by ICON-EPS for the BCs improved mostly the reliability).

6.2 Test of Initial Conditions from KENDA (Arpae)

KENDA was tested for providing Initial Conditions to the deterministic and ensemble experimental chains with COSMO at 2.2 km over Italy. It is here shown the impact of the KENDA analyses as Initial Conditions to the members of the COSMO-IT-EPS ensemble for a period of May-June 2016 characterised by intense thunderstorms over different parts of Italy. The ensemble gets BCs from COSMO-ME-EPS and Parameter Perturbation is applied to the model. COSMO-IT-EPS with only parameter perturbations (PP experiment) and PP plus KENDA Initial Conditions (kendaIC_PP experiment) are compared in the following. An objective verification of the quality of the precipitation forecasted by the ensemble is carried out by comparing the forecasts with observations of precipitation estimated by radar and then adjusted with raingauge data. Both sets of data cover the entire Italian territory. Average and maximum over 0.2 deg boxes are considered.

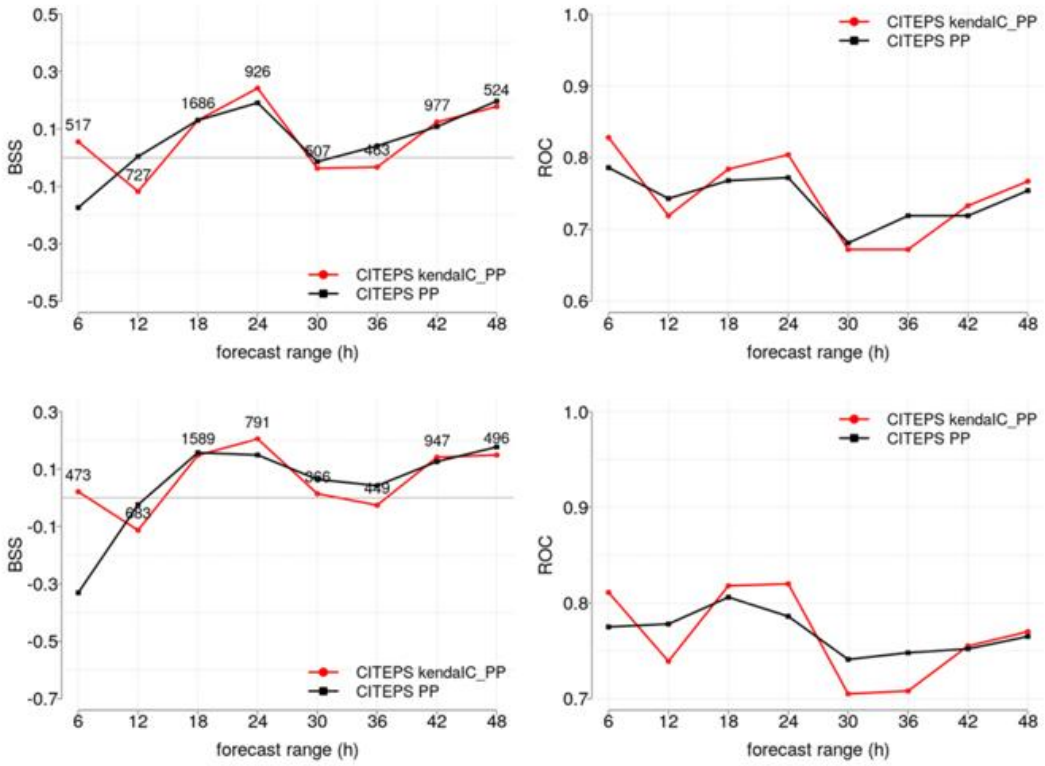

Figure 42: Brier Skill Score (left column) and the ROC area (right column) for average precipitation over each box exceeding 1 mm in 6 h (top row) and for maximum precipitation over each box exceeding 5 mm in 6 h (bottom row), computed against precipitation estimated from the Italian radar network, adjusted with raingauge values.

In all the plots of Fig. 42 is particularly evident the gain in skill in the first 6 hours of forecast determined by the use of Initial Conditions from the KENDA assimilation cycle. Generally, scores are not very high, showing the difficulty of forecasting thunderstorm events at high resolution (the verification boxes are of the order of 20 x 20 km). In order to show the quality of the ensemble members and how this is affected by the different initial conditions, a deterministic verification of the individual members is presented in Fig. 43.

The Frequency Bias (top rows, labelled Bias Score) indicates that the precipitation is generally underestimated by the ensembles, particularly for the lower threshold. It is interesting to notice that instead precipitation is overestimated for the first 6 hours for the 5 and 10 mm thresholds and that this bias is partly cured by the Initial Conditions from the KENDA analysis cycle. The other two scores do not vary much between the two ensemble configurations, with the exception of the first 6 hours as already noticed. In order to provide a clearer indication of what kind of events the 2.2 km ensemble is able to forecast in presence of thunderstorms, probability maps for selected cases are also shown.

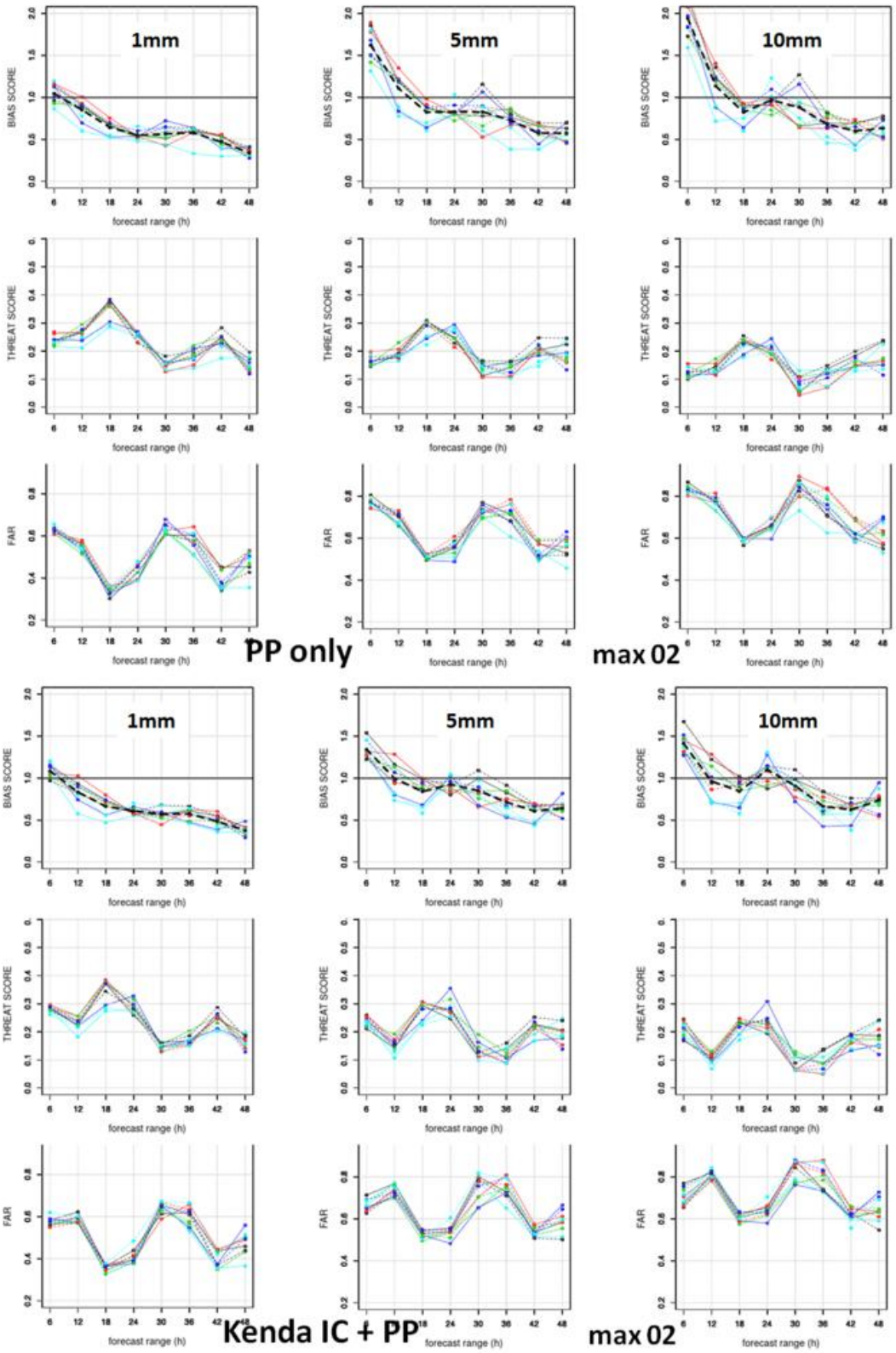

Figure 43: Frequency Bias (top row), Threat Score (middle row) and False Alarm Rate (bottom row) for the maxiumum precipitation over each box exceeding 1 mm in 6 h (left column), 5 mm in 6 h (central column) and 10 mm in 6 h (right row), computed against adjusted radar data for the PP (upper panels) and kendaIC_PP (lower panels) experiments.

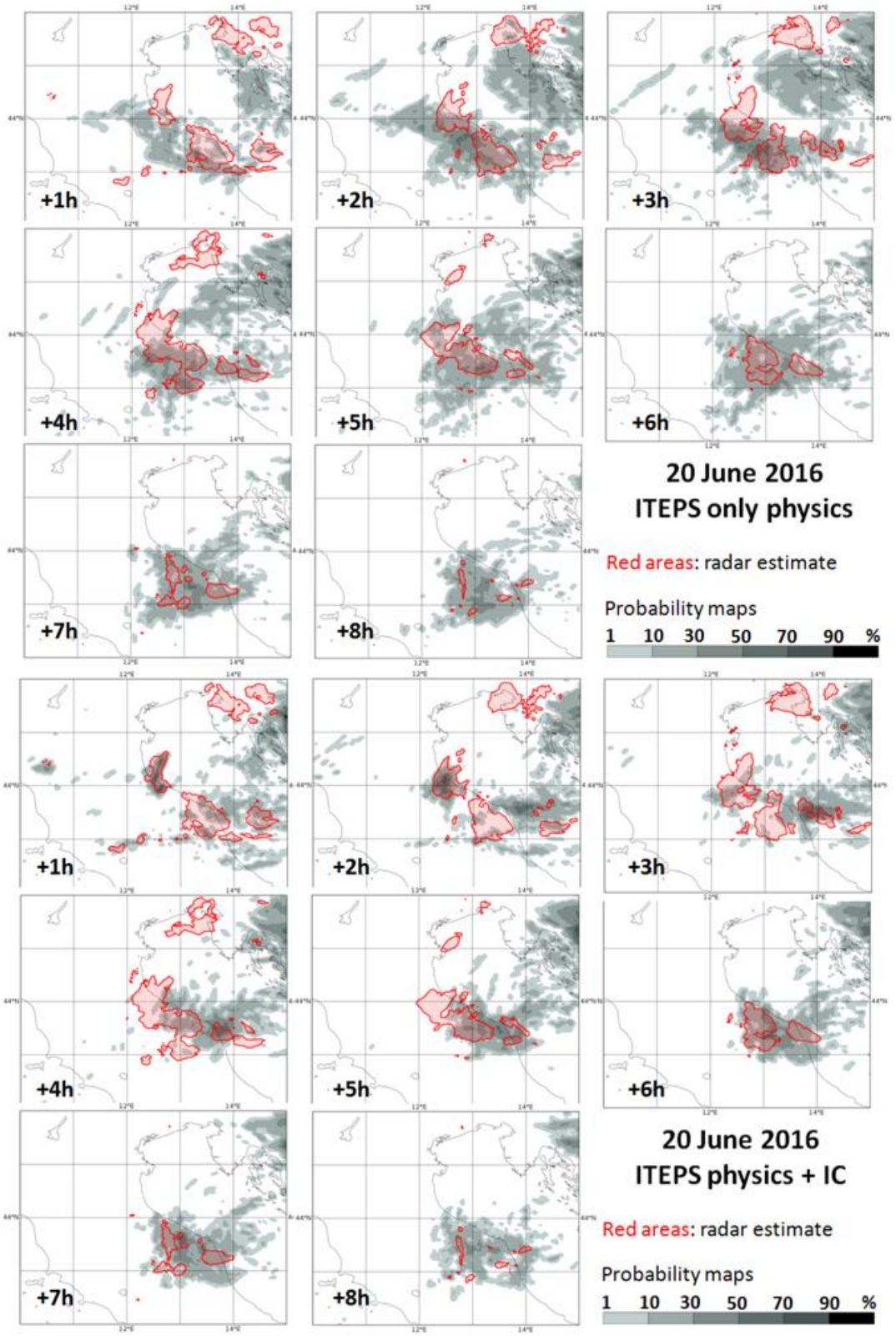

Figure 44: Probability maps (gray shaded) of precipitation exceeding 5 mm/1h generated by the ensemble, relative to the PP (upper panels) and kendalC_PP (lower panels) experiments, and filled contours (red areas) of precipitation estimated by radar exceeding 1mm/1h for the event of the 20th of June 2018, first 8 hours.

In Fig. 44, the probability of precipitation exceeding 5 mm in 1 hour as forecasted by the ensemble in the two different configurations (PP and kendaIC_PP) is plotted (as gray shaded) against the contour (red) relative to the precipitation exceeding 1mm in 1 hour as estimated by the radar network. It is possible to see that indeed the 2.2 km ensemble is generally able to indicated the probable occurrence of thunderstorms in the area where it was occurred. The level of spatial agreement between the forecast, issued in terms of probability and hence not really suitable for a one-to-one comparison with observations, is shown by the match or mismatch between the gray shaded areas and the red area. In this kind of evaluation, little can be said about the accuracy of the forecast of the precipitation amount, since the focus has been put on the capability of the ensemble to issue a forecast of thunderstorm at all, as a valuable tool to assist the forecasters in their daily task. On top, the level of spatial agreement or disagreement is what the forecaster should keep in mind for a profitable usage of the information provided by the ensemble. It is underlined that the different thresholds chosen for the probabilities and for the radar contouring are selected on purpose in order to take into account the tendency of the radar estimate to underestimate the precipitation amount (in analogy with what is done in a verification based on thresholds defined as percentiles instead of values). Considering the first 2-3 hours of forecast, it is evident that the data assimilation at 2.2 km with the KENDA cycle is able to provide initial conditions which greatly improve the position of the precipitation at the beginning of the forecast. This is believed to be partly due to the Latent Heat Nudging which is applied to each member of the KENDA ensemble, which had proven to increase the skill of the precipitation in the first few hours of the forecast. In this case, the forecast by the kendaIC_PP ensemble remains better than the PP one for the entire period shown.

6.3 Clustering of ensemble members for Lateral Boundary Conditions at MeteoSwiss

In the operational COSMO-E setup the perturbed members just use the first 20 members of KENDA and ECMWF ENS for the initial (ICs) and lateral boundary conditions (LBCs), respectively. Westerhuis (2016) investigated whether it is possible to enhance the COSMO-E forecast quality by using a smart selection. For COSMO-E it is found that:

- a sophisticated member selection like clustering for LBCs can improve COSMO-E forecasts significantly while a random member choice can result in significantly worse forecasts with bad luck;
- a clustering is able to decrease the outliers and increase the spread for near-surface variables which is probably the main reason for the better scores as show in Fig. 45;
- a sensitivity analysis of the clustering parameters demonstrates that the verification scores are very sensitive to the time steps used in the clustering analysis;
- the benefit of a better ICs selection is very limited due to the small spread.

Due to the positive results obtained with a member selection based on a clustering analysis, MeteoSwiss plans to implement such a member selection in the operational setup of COSMO-E.

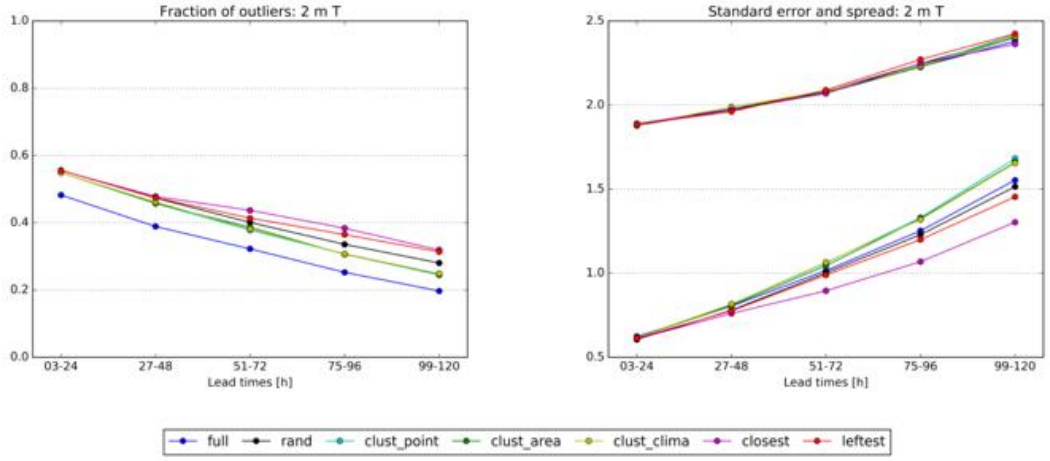

Figure 45: Fraction of outliers (left), standard deviation of error (upper lines in right panel) and ensemble spread (lower lines in right panel) for 2m temperature in COSMO-E forecasts from March 21 - April 8, 2015, for all stations in the model domain. The different lines show the scores obtained with different member selections methods: all 50 members (full), first 20 members (rand), clustering methods with different standardizations (clust), 20 members closest to ensemble mean (closest) and 20 driest members (leftest).

References

Astakhova, E. D., Montani, A., Alferov and Yu, D., 2015: Ensemble forecasts for the Sochi-2014 Olympic Games. *Russian meteorology and hydrology*, **40**, **8**, 531–539, DOI 10.3103/S1068373915080051

Astakhova, E. D., Montani, A., Kiktev, D., Smirnov, A., 2016: COSMO-based ensemble forecasting for Sochi-2014 Olympics: archiving the results. COSMO Newsletter No. 16, 40–45. Available online at http://cosmo-model.org/content/model/documentation/newsLetters/newsLetter16/cnl16_06.pdf

Baker, R. J., Cramer, Peters, J., 2002: Radiation fog: UPS Airlines conceptual models and forecast methods. Proc. 10th Conf. on Aviation, Range and Aerospace Meteorology, Portland, OR, *Amer. Meteor. Soc.*, 154–159

Bonnano, R., and Loglisci, N., 2018: Introducing lower boundary conditions perturbations in a convection-permitting ensemble system: sensitivity to soil moisture perturbation. *Meteorol. Atmos. Phys.*, **130**, **1**, 67–80

Boudala, F. S., Isaac, G. A., Crawford, R. W., and Reid, J., 2012: Parameterization of Runway Visual Range as a Function of Visibility: Implications for Numerical Weather Prediction Models. *Journal of Atmospheric and Oceanic Technology*, **29(2)**, 177–191

Bouttier, F., Raynaud, L., Nuissier, O., and Ménétrier, B., 2015: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quarterly Journal of the Royal Meteorological Society*, **142**, 390–403. https://doi.org/10.1002/qj.2622

Buizza, R., Miller, M., and Palmer, T. N., 1999: Stochastic simulation of model uncertainties. *Quart. J. Roy. Meteorol. Soc.*, **125**, 2887–2908

Bulletin of the National Hydrological and Meteorological Service (referred as Bulletin) (2013), Institute of Meteorology and Water Management-National Research Institute, Poland

Chen, M., Wang, W., Kumar, A., 2013: Lagged Ensembles, Forecast Configuration, and Seasonal Predictions. *Monthly Weather Review*, **141**, 3477–3497. DOI: 10.1175/MWR-D-12-00184.1

Dey, S. R. A., Leoncini, G., Roberts, N. M., Plant, R. S., and Migliorini, S., 2014: A Spatial View of Ensemble Spread in Convection Permitting Ensembles. *Monthly Weather Review*, **142**, 4091–4107

Gultepe, I., and Milbrand, J., 2007: Microphysical observations and mesoscale model simulation of a warm fog case during FRAM project. *J. of Pure and Applied Geophy.*, Special issue on fog, edited by Gultepe I., **164**, 1161–1178

Jolliffe, I. T., and Stephenson, D. B., 2012: Forecast Verification - A Practitioners Guide in Atmospheric Science (second edition), Wiley & Sons, Chichester, UK, doi: 10.1002/9781119960003.ch7

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

Klasa, C., Arpagaus, M., Walser, A., and Wernli, H., 2019: On the time evolution of limited-area ensemble variance: Case studies with the convection-permitting ensemble COSMO-E. *J. Atmos. Sci.*, **76**, 11–26, https://doi.org/10.1175/JAS-D-18-0013.1

Kober, K., and Craig, C., 2016: Physically Based Stochastic Perturbations (PSP) in the Boundary Layer to Represent Uncertainty in Convective Initiation, *J. Atmos. Sci.*, **73**, 2893–2911

Lu, C., Yuan, H., Schwartz, B. E., Benjamin, S. G., 2007: Short-range numerical weather prediction using time-lagged ensembles. *Weather and Forecasting*, **22**, 580–595. DOI: 10.1175/WAF999.1

Marsigli, C., 2009: COSMO Priority Project "Short Range Ensemble Prediction System" (SREPS): Final Report. COSMO Technical Report No. 13, available at http://www.cosmo-model.org/content/model/documentation/techReports/docs/techReport13.pdf.

Marsigli, C., Montani, A., and Paccagnella, T., 2008: A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorological Applications*, **15**, 125–143

Mazur, A., and Duniec, G., 2014a: Sensitivity test on the behavior of different COSMO suites to different lower boundary initial conditions. Presented during COSMO User Seminar, Offenbach, Germany, 2014

Mazur, A., and Duniec, G., 2014b: Soil state perturbations as an input for Ensemble Prediction System (EPS) forecast, 9th International Soil Science Congress on "The Soul of Soil and Civilization", Antalya/Turkey, 14-16 October, 2014. doi: 10.13140/2.1.1397.5840

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

Montani, A., Alferov, D., Astakhova, E., Marsigli, C., and Paccagnella, T., 2014: Ensemble forecasting for Sochi-2014 Olympics: the COSMO-based ensemble prediction systems. COSMO Newsletter, **14**, 88–94, available online at http://cosmo-model.org/content/model/documentation/newsLetters/newsLetter14/cnl14_10.pdf

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., and Weisheimer, A., 2009: Stochastic parametrization and model uncertainty. ECMWF Research Department Technical Memorandum, **598**, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pp. 42

Peralta, C., Ben Bouallègue, Z., Theis, S. E., Gebhardt, C., and Buchhold, M., 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, **D7**, doi:10.1029/2011JD016581

Sætra, Ø., Hersbach, H., Bidlot, J. R., and, Richardson, D. S., 2004: Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, **132**, 1487–1501

Tennant, W., and Beare, S., 2014: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1150–1160

Tsyrulnikov, M., and Gayfulin, D., 2016: A Stochastic Pattern Generator for ensemble applications. COSMO Technical Report, **29**, pp. 51

Tsyrulnikov, M., and Gayfulin, D., 2017: A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteorol. Zeitschrift*, **26**, **N5**, 549–566

Tsyrulnikov, M. D., 2001: Proportionality of scales: an isotropy-like property of geophysical fields. *Quarterly Journal of the Royal Meteorological Society*, **127**, **578**, 2741–2760

Wernli, H., Paulat, M., Hagen, M., and Frei, C., 2008: SAL - A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Wea. Rev.*, **136**, 4470–4487

Westerhuis, S., 2016: Ensemble member selection for COSMO-E initial and boundary conditions. Master thesis. Available at MeteoSwiss

Zhou, B., 2011: Introduction to A New Fog Diagnostic Scheme. NCEP Office Note 466, US Dept. of Commerce, NOAA, National Weather Service, NCEP, pp.43

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.

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.

No. 22: C. Marsigli, T. Diomede, A. Montani, T. Paccagnella, P. Louka, F. Gofa, A. Corigliano (2013):
The CONSENS Priority Project.

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.

No. 24: A. K. Miltenberger, A. Roches, S. Pfahl, H. Wernli (2014):
Online Trajectory Module in COSMO: a short user guide.

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.

No. 26: D. Mironov, E. Machulskaya, B. Szintai, M. Raschendorfer, V. Perov, M. Chumakov, E. Avgoustoglou (2015):
The COSMO Priority Project 'UTCS' Final Report.

No. 27: J-M. Bettems (2015):
The COSMO Priority Project 'COLOBOC': Final Report.

No. 28: Ulrich Blahak (2016):
RADAR_MIE_LM and RADAR_MIELIB - Calculation of Radar Reflectivity from Model Output.

No. 29: M. Tsyrulnikov and D. Gayfulin (2016):
A Stochastic Pattern Generator for ensemble applications.

No. 30: D. Mironov and E. Machulskaya (2017):
A Turbulence Kinetic Energy – Scalar Variance Turbulence Parameterization Scheme.

No. 31: P. Khain, I. Carmona, A. Voudouri, E. Avgoustoglou, J.-M. Bettems, F. Grazzini, P. Kaufmann (2017):
CALMO - Progress Report.

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

No. 33: N. Vela (2017):
VAST 2.0 - User Manual.

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.

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.

No. 36: A. Raspanti, A. Celozzi, A. Troisi, A. Vocino, R. Bove, F. Batignani (2018):
The COSMO Priority Project VERSUS2 Final Report.

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.

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.

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@arpa.piemonte.it