Consortium

for

Small-Scale Modelling

COSMO
CONSORTIUM FOR SMALL SCALE MODELING

Newsletter

October 2019   *No. 19*

Deutscher Wetterdienst

Ufficio Generale Spazio Aereo e Meteorologia

Instytucie Meteorogii i Gospodarki Wodnej

Agenzia Regionale per la Protezione Ambientale dell'Emilia Romagna Servizio Idro Meteo Clima

Centro Italiano Ricerche Aerospaziali

Israel Meteorological Service

MeteoSwiss

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ

Administratia Nationala de Meteorologie

Federal Service for Hydrometeorology and Environmental Monitoring

Amt fur GeoInformationswesen der Bundeswehr

Agenzia Regionale per la Protezione Ambientale del Piemonte

---

www.cosmo-model.org

Table of Contents

Consortium for Small-Scale Modelling continues to mount efforts towards replacing the limited-area model COSMO with the Limited Area Mode of the comprehensive modelling framework ICON (ICON-LAM). Considerable progress along this line has been made, particularly within the framework of the priority project C2I (`http://www.cosmo-model.org/content/tasks/priorityProjects/c2i`) launched in 2018 to facilitate a smooth transition from the COSMO model to ICON-LAM. The COSMO members become acquainted with ICON-LAM increasingly closely, and various problems of scientific and technical character are being gradually solved. The transition to ICON-LAM also poses problems of organizational and legal character, such as the co-ordination of efforts between COSMO and the ICON Partners (DWD and MPI for Meteorology), the future role of the Steering Committee and of the Scientific Management Committee of the Consortium, the license policies, and the intellectual property rights. These issues are thoroughly scrutinized by the COSMO STC and SMC and the ICON governing bodies. Several key documents should be ready by the 21st COSMO General Meeting to be held 9-13 September 2019 in Rome, Italy, where the various critical issues will be further discussed.

In spite of the advent of ICON-LAM, the development of the COSMO model is not yet stopped. The release of the version 6.0 of the COSMO model is expected in the end of 2019. Version 6.0 is intended to be the last official release of the COSMO model. Importantly, version 6.0 will integrate the results obtained within the COSMO NWP and COSMO CLM communities and will be a unified COSMO-model version for both the NWP and regional climate modelling applications. No development is planned beyond the COSMO-model version 6.0, although the maintenance of the COSMO code (including bug fixes) will be provided for some years to come.

I would like to gratefully thank all colleagues who contributed to the current issue of the COSMO Newsletter. Worthy of mention is a new section entitled "Mission Reports". That section contains mission reports, namely, brief reports on various meetings (conferences, symposia, etc.) that COSMO colleagues have attended using the COSMO license money. Last but not the least, I am pleased to mention that, starting with the current issue (No. 19), the COSMO Newsletters will be published with the Digital Object Identifiers. The DOI format reads `10.5676/dwd_pub/nwv/cosmo-nl_XX_YY`, where XX is the number of the Newsletter, and YY is the number of contribution within the Newsletter No. XX. The landing page for the COSMO Newsletters is `https://www.dwd.de/EN/ourservices/cosmo_newsletter/cosmo_newsletters.htm`. DOIs are also provided for COSMO Technical Reports. The DOI format reads `10.5676/dwd_pub/nwv/cosmo-tr_XX`, where XX is the technical report number. The latest five Technical Reports (Nos. 34-39) with DOIs are already available at the landing page at DWD, `https://dwd.de/EN/ourservices/cosmo_technical_reports/cosmo_technical_reports.htm`. The work is underway to provide DOI for all COSMO Technical Reports published so far. Great thanks are due to the DWD colleagues Ms. Magdalena Bertelmann and Dr. Jörg Rapp for their kind efforts in providing DOIs for COSMO publications.

Dmitrii Mironov
COSMO Scientific Project Manager

Figure 1: Participants of the 20th COSMO General Meeting in St. Petersburg, Russia

Figure 2: Participants of the 20th COSMO General Meeting in St. Petersburg, Russia

Statistical analysis of radar reflectivities observed and simulated by EMVORADO

Virginia Poli[1], Thomas Gastaldo[1,2], Pier Paolo Alberoni[1] and Tiziana Paccagnella[1][1]

Arpae-SIMC Emilia-Romagna, Bologna, Italy
[2] *Universitá degli Studi di Bologna, Bologna, Italy*

1 Introduction

In the COSMO Consortium (Consortium for Small-scale Modeling), the assimilation of radar data is now on-going into the Kilometer-scale ENsemble Data Assimilation (KENDA) LETKF system [1] by means of the Efficient Modular VOlume RADar forward Operator (EMVORADO, [2], [3], [4], [5], [6]). At Arpae-SIMC, the HydroMeteorological and Climate Service of the Emilia-Romagna (Italy), the attention is focused on the assimilation of radar reflectivity volumes.

The off-line version of EMVORADO, i.e. not included in the assimilation cycle, has been implemented to calculate the reflectivity volumes from KENDA analyses in order to estimate the observation error by means of a method based on statistical averages of observation-minus-background and observation-minus-analysis residual.As a side result, the comparison between the observed and simulated reflectivities allows us to understand how much the values derived by the operator deviate from reality. Hence, the use of the off-line operator makes it possible to verify how the reflectivity distributions vary both using different analyses, coming from various KENDA configurations, and by directly modifying the parameters of the operator himself.

2 Statistical distributions of reflectivities

To quantify the differences between reflectivities simulated with different EMVORADO configurations and also between observed and simulated ones, the off-line radar operator was applied, i.e. separately from the assimilation cycle, to all the analyses obtained from different assimilation cycles of KENDA. In particular, for this topic, hourly analyses come from KENDA with the assimilation of conventional observations (SYNOP, TEMP and AIREP) and KENDA with the assimilation of conventional observations and radar reflectivity volumes. The radar operator configurations that have been tested are summarized in table 1. Among all the possible combinations, the different type of scattering (Mie/Rayleigh) for reflectivity computation has been used. Subsequently, the attenuation along the beam was also taken into account for the Mie scattering. This option cannot be used for Rayleigh scattering. With regard to Rayleigh scattering, the effect of the use of different beam propagation methods has been verified. By default, the "4/3-earth" climatological model is used, the other two options enable the ray tracing and the beam bending computations based on the simulated air refractive index field [5]. Specifically the TORE method is based on Snell's law for spherically stratified media including effects of total reflection, while the SODE method is based on the second-order ordinary differential equation for the beam height as a function of range.

Once all the simulated volumes were produced, reflectivities above 0 dBZ were considered and boxplots were generated independently for the case studies indicated in table 2 (Figures 1 and 2). The choice to calculate the statistical distributions according to the events was due to the fact that the chosen periods have very different weather characteristics.

Table 1: EMVORADO configurations

Name	Scattering options	Propagation options
Mie	Mie scattering	Climatological "4/3-earth" model
Mie_atten	Mie scattering taking into account attenuation along the ray path	Climatological "4/3-earth" model
Rayleigh	Rayleigh scattering	Climatological "4/3-earth" model
Rayleigh_sode	Rayleigh scattering	Method SODE based on the second-order ordinary differential equation for the beam height as function of range
Rayleigh_tore	Rayleigh scattering	Method TORE based on Snell's law for spherically stratified media including effects of total reflection

Table 2: Case studies

Event	Start of the event	End of the event	Type of event
September 2018	31/08/2018 01 UTC	09/09/2018 00 UTC	thunderstorms
October 2018	30/09/2018 16 UTC	14/10/2018 00 UTC	thunderstorms and organized convective structures
November 2018	26/10/2018 13 UTC	11/11/2018 00 UTC	stratiform structures with some convective episodes at the beginning of the period
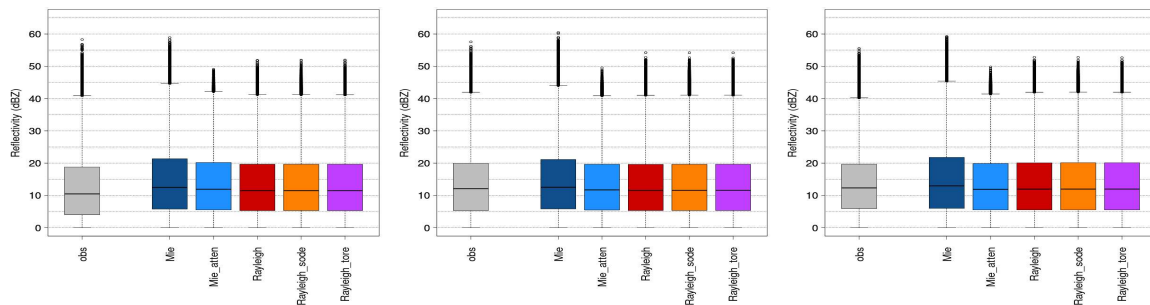
Figure 1: Boxplots calculated for September 2018 (a), October 2018 (b) and November 2018 (c) with input analyses from KENDA cycles with the assimilation of conventional observations.

Using as input analyses those derived from KENDA cycles with the assimilation of only conventional observations (Figure 1), the distributions do not vary significantly depending on the case study considered. Small differences can be observed on the estimated maximum values: for the October case the simulated maximum reflectivities are higher. On the other hand, considering the different configurations of EMVORADO, the use of Mie scattering generally produces a distribution with higher reflectivity values. By activating attenuation, values between 25th and 75th percentiles are realigned with other configurations, but values above 95th percentiles are all limited to below 50 dBZ. For the configurations with the Rayleigh scattering, the use of different beam propagation schemes does not bring to any significant changes.

The simulations behavior using as input analyses those derived from KENDA cycles with the assimilation of only conventional observations and radar reflectivity volumes at the analysis time (Figure 2), calculated only for October 2018, differ slightly from the previous ones. Median values are higher, but maximum values above the 95th percentiles are smaller.
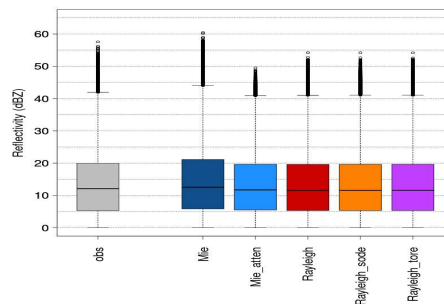
Figure 2: Boxplot calculated for October 2018 with the input analyses from KENDA cycles with the assimilation of conventional observations and radar reflectivity volumes at the analysis time.

3 Case Study: flooding in Sardinia in October 2018

To understand in more detail how the simulated reflectivity changes according to the chosen namelist parameters, the flood that hit southern Sardinia on October 10, 2018 was analyzed. As shown in figure 3, in this part of the region, several rain gauges have measured values greater than 100 mm over the two days, with a maximum value at Santa Lucia di Capoterra of 493.4 mm. For this rain gauge the trend of the hourly accumulated precipitation (Figure 4) displays a first passage of the precipitating structures in the morning of October 9 and a persistence of the phenomena from the evening of October 10 until the end of the event.
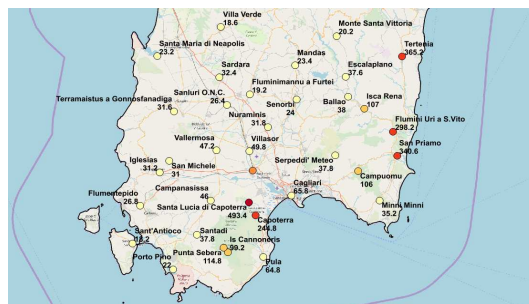
Figure 3: Accumulated precipitation measured by rain gauges from 09/10/2018 - 00 UTC to 11/10/2019 - 00 UTC in Southern Sardinia.
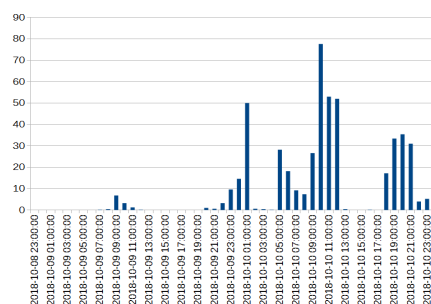
Figure 4: Accumulated hourly precipitation measured by the rain gauge located at Santa Lucia di Capoterra from 08/10/2018 - 23 UTC to 11/10/2019 - 00 UTC.

Two hours were selected and the polar volume at its lowest elevation was displayed for Armidda radar (in red, in figure 5) for the different EMVORADO configurations and depending on the input analyses.
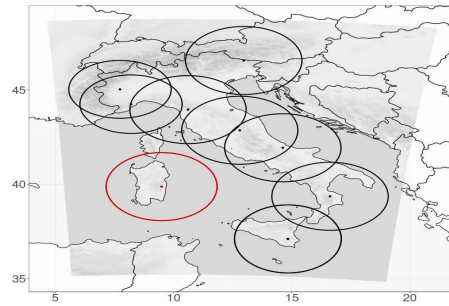
Figure 5: Armidda's radar, highlighted in red, inside the radar Italian radar network used for this study.

In figure 6 and figure 7, columns 2, 3 and 4 show the reflectivity fields simulated using Rayleigh, Mie and Mie with attenuation configurations. Simulations can be directly compared with the observation in column 1. The different rows refer to the different analyses used as input. In the first row the analyses used come from KENDA cycles with assimilation of only conventional observations (SYNOP, TEMP and AIREP). In the second row analyses come from KENDA cycles with assimilation of conventional observations and LHN, while in the third one they comes from KENDA cycles with assimilation of conventional observations and radar reflectivity volumes at the analysis time. Figure 6 refers to 9 October at 9 UTC, while figure 7 is for 10 October at 10 UTC.
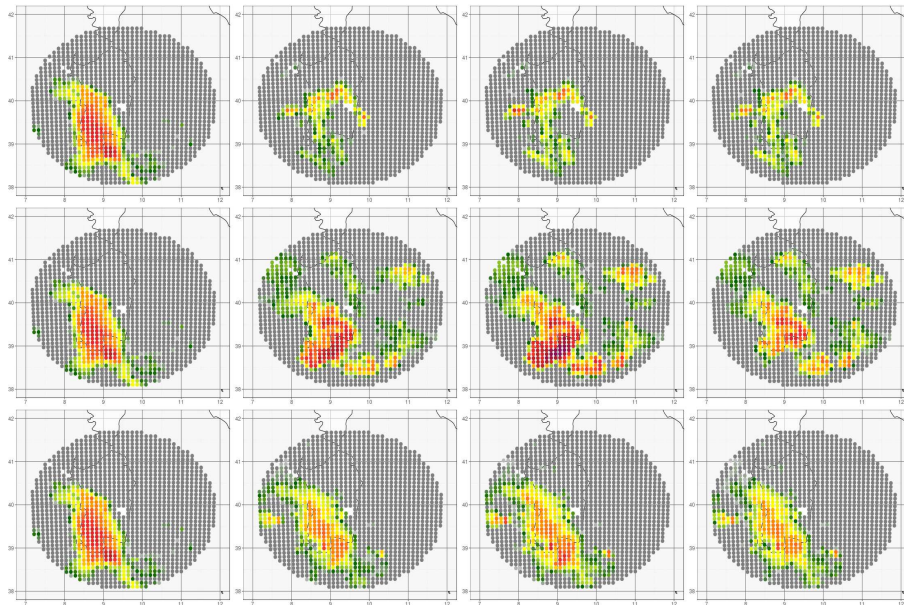
Figure 6: Observed (first column) and simulated (columns 2, 3 and 4) reflectivity of Armidda's first radar elevation by changing EMVORADO configurations (column 2: Rayleigh, column 3: Mie, column 4: Mie with attenuation) and input analyses (top row: analysis from KENDA with the assimilation of conventional observations, middle row: analysis from KENDA with the assimilation of conventional observations and LHN, bottom row: analysis from KENDA with the assimilation of conventional observations and radar reflectivity values at the analysis time) for October 9 at 9 UTC.

In both instants examined the structures are simulated in a more accurate way, both in terms of location and shape, if the analyses come from KENDA cycles with assimilation of radar volumes. The use of analyses with only the assimilation of conventional observations leads to the modeling of structures that have little relevance to the observation, in particular this can be observed at the first instant.

Simulations starting from the analyses in which the LHN is used overestimate the reflectivity. This is most
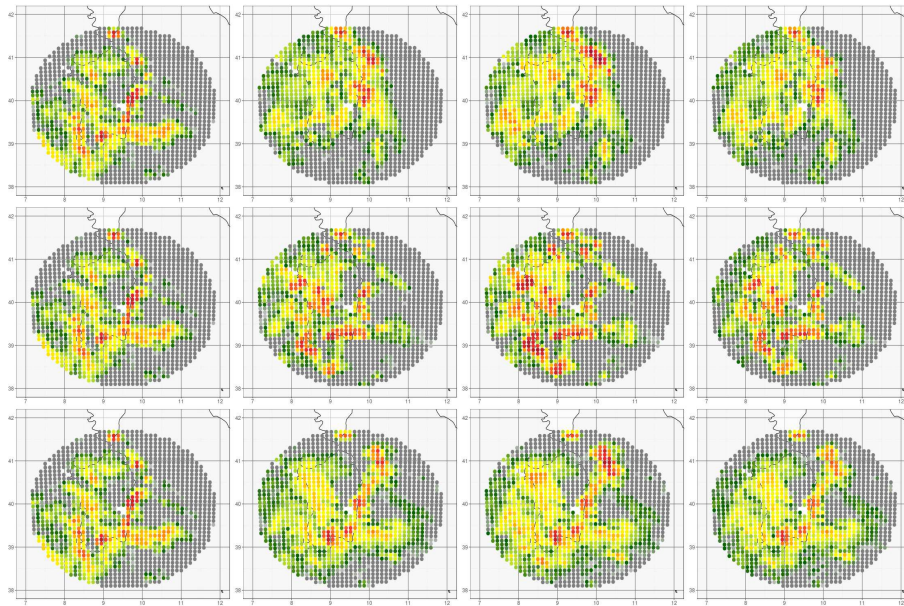
Figure 7: As fig. 6, but for October 10 at 10 UTC.

visible at the first instant where not only the structure in the south-west part of the domain is overestimated, but unobserved precipitation is simulated in the eastern part of the domain.

The use of Rayleigh's scattering with the analyses coming from the assimilation of radar volumes brings to a general underestimation of the field of reflectivity.

The combination of the LHN analyses and the use of the Mie configuration leads to a strong overestimation of all simulated values. As a general result, regardless of input fields, the use of attenuation improves overestimation by bringing the simulations more similar to those obtained using the Rayleigh scattering.

For this case study, comparing the obtained simulations with the observations, the use of Mie scattering provides the best results.

4 Conclusions and future work

The results obtained from this case study deviate partially from what is highlighted by the distributions of reflectivity on all events. In this case, the use of Mie scattering seems to provide the best results, while the distributions show a clear overestimation of the values with respect to the observations.

At the moment the forecasts initialized with KENDA analyses, obtained with the configuration of EMVO-RADO with Rayleigh scattering, provide a good improvement over the operational runs. However, the Mie scattering will be used for the case studies presented, providing a quantitative comparison between forecasts.

References

[1] Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A., and Potthast, R. 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA), *Q. J. Roy. Meteor. Soc.*, 142, **1453–1472**, URL https://doi.org/10.1002/qj.2748.

[2] Blahak, U., 2016: RADAR_MIE_LM and RADAR_MIELIB - calculation of radar reflectivity from model output. *Technical Report 28*, Consortium for Small Scale Modeling (COSMO), URL http://www.cosmo-model.org/content/model/documentation/techReports/docs/techReport28.pdf.

[3] Zeng, Y., 2013: Efficient radar forward operator for operational data assimilation within the COSMO-

model. *Dissertation*, IMK-TRO, Department of Physics, Karlsruhe Institute of Technology, URL `http://digbib.ubka.uni-karlsruhe.de/volltexte/1000036921`.

[4] Jerger, D., 2014: Radar forward operator for verification of cloud resolving simulations within the COSMO-model. *Dissertation*, IMK-TRO, Department of Physics, Karlsruhe Institute of Technology, URL `http://digbib.ubka.uni-karlsruhe.de/volltexte/1000038411`.

[5] Zeng, Y., Blahak, U., Neuper, M. and Epperlein, D., 2014. Radar beam tracing methods based on atmospheric refractive index. *J. Atmos. Ocean. Tech.*, 31, **2650-2670**.

[6] Zeng, Y., Blahak, U. and Jerger, D., 2016. An efficient modular volume-scanning radar forward operator for NWP models: description and coupling to the COSMO model. *Quart. J. Roy. Met. Soc.*, 142, **3234-3256**, URL `http://onlinelibrary.wiley.com/doi/10.1002/qj.2904/abstract`.

Assimilation of radar reflectivity volumes employing different observation error covariance matrices

THOMAS GASTALDO[1,2], VIRGINIA POLI[1], CHIARA MARSIGLI[3], PIER PAOLO ALBERONI[1] AND TIZIANA
PACCAGNELLA[1] *[1] Arpae-SIMC Emilia-Romagna, Bologna, Italy*
[2] University of Bologna, Italy
[3] Deutscher Wetterdienst, Offenbach, Germany

1 Introduction

At Arpae-SIMC, the HydroMeteorological Service of Emilia-Romagna Region (Italy), the KENDA assimilation system [1] provides the analyses to the convection-permitting components of the operational modelling chain, consisting of one deterministic run and one ensemble system, both at 2.2 km horizontal resolution and with the same domain (greyscale in Figure 1). Currently, only conventional observations are assimilated, but tests are ongoing to include also reflectivity volumes [2] from the Italian radar network (solid lines in Figure 1).
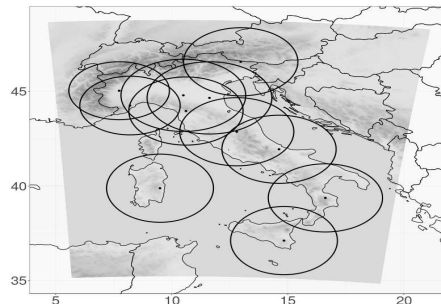
Figure 1: Integration domain (greyscale) of the COSMO model employed at Arpae-SIMC for high resolution model runs. The approximate coverage area for each radar at their lowest resolution of the Italian network is shown with solid lines.

The high spatial and temporal density of radar data demands a great care in setting the observation error covariance matrix $\mathbf{R}$. In fact, due to the great amount of data, small departures of the observation error from its actual value may lead to large errors in the analysis. Furthermore, reflectivity observations are spatially and temporally correlated and, therefore, the assumption made in most operational data assimilation systems of a diagonal $\mathbf{R}$ matrix is not realistic (see for example [2]).

The impact of using different estimations of the $\mathbf{R}$ matrix in the assimilation system is presented over two test periods. Results obtained when employing an unique observation error for all reflectivity volumes are compared to those obtained when a different value is specified for each observation, depending on the radar station and the distance from the station. The analyses, derived by each observation error matrix configuration, are used to initialize different forecasts. The comparison of the quantitative precipitation forecast (QPF) using the Fractions Skill Score (FSS [3]) allows to estimate the accuracy of the analysis itself. Finally, an estimation of spatial correlations between reflectivity observations is provided.

2 Observation error

The observation error ϵ_o has two components [4]: the measurement and the representation error. The former, also called instrument error, is the error associated with the measuring device alone, independently of how the measurements are used. The latter arises from 3 sources:

- errors due to a mismatch between the scales represented in the observations and the model fields;
- errors introduced by the observation operator;
- errors due to pre-processing or quality control.

In data assimilation, an accurate estimation of the observation error is crucial since the observation error covariance $\mathbf{R} = E[\epsilon_o \epsilon_o^T]$ weights observations as $\mathbf{B} = E[\epsilon_b \epsilon_b^T]$ weights model background information (ϵ_b is the background error). While during the past decades a great effort has been done to improve the estimation of $\mathbf{B}$ (for example in the KENDA system it is fully flow dependent), small improvement have been done regarding the $\mathbf{R}$ matrix. In fact, $\mathbf{R}$ is fixed in time and generally assumed to be diagonal, that is observations are considered uncorrelated. Regarding the way to estimate it, one of the most used is the method proposed by Desroziers[5] which is based on the expect value of the product between observation-minus-analysis and observation-minus-background residuals.

3 Estimation of reflectivity errors

In order to estimate reflectivity error with a spatial dependence, we estimate the diagonal of $\mathbf{R}$ using Desroziers statistics and then we bin observations and the associated errors according to their horizontal and vertical distance from radar station. We use an horizontal step of 50 km and a vertical step 2 km. The estimation is performed for each radar of the Italian network over 3 periods,in order to have also a temporal dependence: from 31/08/18 at 00 UTC to 09/09/18 at 00 UTC (sept2018), from 30/09/18 at 15 UTC to 10/10/18 at 00 UTC (oct2018) and from 26/10/18 at 12 UTC to 11/11/18 at 00 UTC (nov2018).

Estimated values averaged over the three periods (sept2018, oct2018 and nov2018) and over all radars of the Italian network are shown in Figure 2. Values (y axis) are shown as a function of horizontal distance (x axis) and vertical distance (colours). As a general behaviour, we can notice that observation error increases with horizontal distance. This seems to be reasonable since the size of observed atmospheric volumes increases with the distance from the radar station. At the same time, we can notice that the observation error decreases with vertical distance up to the 4-6 km bin and then stabilizes. Also this behaviour seems to be reasonable since reflectivity observations close to the ground are more likely affected by non meteorological signals (i.e. clutter).
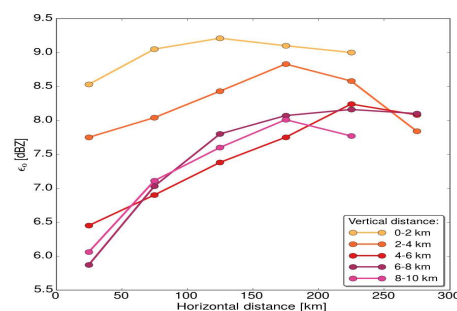

Figure 2: Estimated observation error for reflectivity volumes averaged over all periods and over all radars of the Italian network.

Due to the heterogeneity of our radar network and to the presence of different weather regimes in Italy, when the statistics is applied separately to each radar we can notice a certain variability. As an example, in

Figure 3 estimated values of reflectivity errors are shown for Serano radar (left panel) in Central Italy and for Zoufplan radar (right) in North-Eastern Italy. Values are averaged over the 3 periods sept2018, oct2018 and nov2018. It can be noticed that the general behaviour described above is conserved but values and slopes of the curves vary quite significantly. A certain variability can be observed also when considering one radar but restricting the statistics to a single period. This is shown, for example, in Figure 4 for Zoufplan radar applying the Desroziers statistics only at the sept2018 period (left panel) and at nov2018 (right).
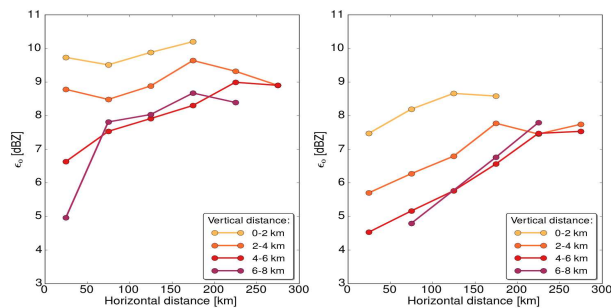
Figure 3: Same as Figure 2 but for computing the statistics only for Serano radar (left) in Central Italy and Zoufplan radar (right) in North-Eastern Italy.
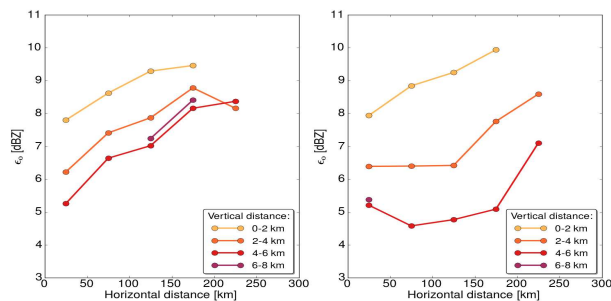
Figure 4: Same as Figure 2 but computing the statistics only for Zoufplan radar at two different periods: sept2018 (left) and nov2018 (right).

4 Use of estimated values of the observation error in KENDA

In order to evaluate the impact of using the estimated values of reflectivity error in the KENDA assimilation system, we perform 3 experiments. In *err_fix* experiment all reflectivity volumes have an error of 10 dBZ, as in our standard set-up for the assimilation of radar data. In *err_mean* experiment the observation error varies with radar station and with horizontal and vertical distance from station and it is averaged over all periods. Finally, in *err_period* experiment the observation error varies with radar station, with horizontal and vertical distance from station and with period.

The three experiments are performed for sept2018 and oct2018 periods. The KENDA system employs a 20 member ensemble plus a deterministic run and an assimilation window of 1 hour; Both conventional data and radar volumes (only the closest to analysis time for each radar) are assimilated. Finally, a deterministic forecast is initialized each 3 hours and forecast precipitation is verified by using the Fractions Skill Score (FSS). Regarding FSS, fixed spatial windows of 0.2 degrees are used and thresholds of 1 mm and 5 mm are considered. Observations are hourly rainfall fields from the Italian radar composite adjusted by rain-gauges.

Results are shown in Figure 5. Differences between the three experiments are small for both sept2018 (left panel) and oct2018 (right panel) periods. Regarding sept2018, FSS values for *err_mean* (red lines) are very close to those of *err_fix* (blue) for both the 1 mm (solid lines) and the 5 mm (dashed lines) threshold.

In contrast, the performance of *err_period* (green) is generally slightly better than that of the other two experiments. However, when considering the oct2018 case, *err_mean* experiment is very slightly worse than *err_fix* and the worst results are obtained for *err_period*. In conclusion, due to the mixed results observed, we can state that the impact of employing a more accurate characterization of the observation error for reflectivity volumes in the assimilation system does not affect significantly the quality of forecast precipitation.
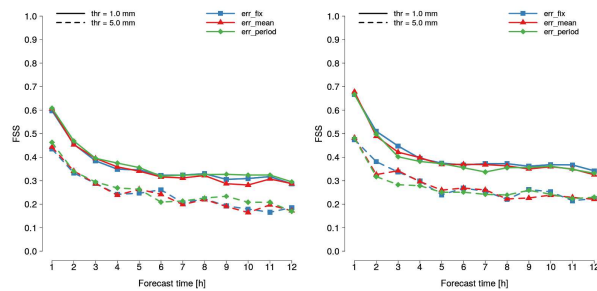
Figure 5: Fractions Skill Score for *err_fix* (blue lines), *err_mean* (red) and *err_period* (green) experiments employing a threshold of 1 mm (solid lines) and 5 mm (dashed lines). The verification is applied to sept2018 (left panel) and to oct2018 (right panel) periods.

5 Estimated values of correlation between radar observations

Employing the Desroziers statistics, we also compute an estimation of spatial correlations for reflectivity errors. Similarly to the method described in Section 3, we bin pairs of radar observations according to their horizontal and vertical distance. We employ an horizontal step of 10 km and a vertical step of 1 km. Results obtained for the sept2018 case averaged over all radars of the Italian network are shown in Figure 6. As expected, errors are strongly correlated vertically and significant correlations can be seen up to 40 km in horizontal.
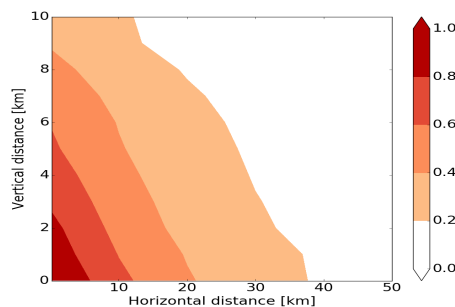
Figure 6: Spatial correlation between pair of reflectivity observations during sept2018 case.

6 Conclusions and future work

Even if reflectivity observation error varies quite significantly with time, radar station and distance from the radar, the use of more accurate values of errors in KENDA does not improve forecast accuracy. However, further tests are needed to confirm this result. The estimation of correlations between reflectivity errors shows that there is a strong correlation in space. Therefore, the exploitation of the correlation between pair of observations in the **R** matrix may be beneficial.

References

[1] Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A., and Potthast, R.: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA), Q. J. Roy. Meteor. Soc., 142, 1453–1472, https://doi.org/10.1002/qj.2748, 2016.

[2] Gastaldo, T., Poli, V., Marsigli, C., Alberoni, P. P., and Paccagnella, T.: Data assimilation of radar reflectivity volumes in a LETKF scheme, Nonlin. Processes Geophys., 25, 747-764, https://doi.org/10.5194/npg-25-747-2018, 2018.

[3] Roberts, N.M. and H.W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. Mon. Wea. Rev., 136, 78–97, https://doi.org/10.1175/2007MWR2123.1

[4] Janjić, T, Bormann, N, Bocquet, M, Carton, JA, Cohn, SE, Dance, SL, Losa, SN, Nichols, NK, Potthast, R, Waller, JA, Weston, P. On the representation error in data assimilation, *Q J R Meteorol Soc.* 2018; 144: 1257– 1278. https://doi.org/10.1002/qj.3130

[5] Desroziers, G. , Berre, L. , Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysiserror statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385-3396. doi:10.1256/qj.05.108

Estimation of model errors on convective scales: a coarse-graining study (preliminary stage results)

MICHAEL TSYRULNIKOV AND DMITRY GAYFULIN

mik.tsyrulnikov@gmail.com

Abstract

An attempt to objectively estimate model tendency errors using a "true model" is described. The model in question (the "model") is COSMO with the horizontal resolution 2.2 km. The "true model" is COSMO with the horizontal resolution 0.55 km. The model error is evaluated as the difference between 1-minute "model" and upscaled (coarse-grained) "true-model" tendencies started from the same initial conditions. Preliminary results show, that, first, convection is not to be treated with this approach. Second, non-convective model errors contain both additive and multiplicative components. The additive component and the multiplier (applied to the physical tendency) appear to be approximately Gaussian. Third, the model-error field is too complex (especially in the planetary boundary layer) to be modeled with a reasonably simple stochastic model, so a process-level model error treatment is to be employed.

1 Introduction

To perform numerical weather prediction, three components are needed: initial conditions, boundary conditions, and a forecast model. The classical paradigm is deterministic: we (naively) assume that all these three components are perfect (however they are prepared), and come up with a deterministic forecast. But in reality, the three components needed to compute a forecast are, of course, imperfect and subject to uncertainty, so that the forecast inevitably contains an error. The (expected) magnitude of the error is of great interest to any user of the forecast and thus should be quantified.

1.1 Ensemble prediction

The most widely used paradigm to account for the uncertainties is stochastic: the forecast-error is assumed to be a random field with the probability distribution to be computed/estimated/specified. Correspondingly, all data used to prepare the forecast and the forecast-model itself are assumed to be random. Initial and boundary conditions are treated as multidimensional random fields. Forecast model equations are assumed to be subject of error traditionally represented by the *model error*, the difference of the model's right-hand sides from the hypothetical true right-hand sides. The model-error field is, then, also modeled as a spatiotemporal random field. The randomness of the three "input" random fields, that is, initial, boundary, and model-error fields, results in a randomness of the forecast. Therefore, to quantify the forecast probability distribution, we need, first, to adequately model the three input probability distributions and second, to map these input distributions to the output (i.e. forecast) distribution.

These two tasks are performed nowadays using Monte-Carlo, that is, the input probability distributions are represented by pseudo-random *samples* from an initial ensemble, a boundary ensemble, and a model-error ensemble. These samples are fed to the forecast model (in other words, the initial and boundary data as

well as the model equations are "perturbed")[1]. The perturbed input fields give rise to a perturbed forecast. Multiple realizations of the input fields give rise to a *forecast ensemble*. The forecast ensemble is then, by construction, a sample from the forecast probability distribution. Otherwise stated, the forecast ensemble is obtained, ideally, by replacing deterministic initial and boundary conditions and deterministic right-hand sides of the model equations by random samples from the respective distributions. This approach is known in geosciences as ensemble modeling and prediction. From the forecast ensemble, a *probabilistic forecast* can be computed, replacing the deterministic forecast and quantifying the forecast uncertainty.

For the ensemble prediction to adequately describe the forecast uncertainty, the "input" uncertainties need to be adequately represented. The initial ensemble is, normally, generated by a data assimilation scheme. The lateral-boundary ensemble for a limited-area model is generated from a parent-model forecast ensemble. The lower-boundary ensemble requires a forecast ensemble in the soil/sea/lakes/rivers etc. In global models, the upper boundary condition is normally not perturbed. In limited-area models, the upper-boundary ensemble is generated in the same way as the lateral-boundary ensemble. It remains to generate samples from the distribution of model errors. Our focus in this study is on model errors.

1.2 Model errors

Existing approaches to model-error modeling can be characterized as follows. Non-stochastic approaches include multi-model and multi-physics techniques e.g., Berner et al. [1]. The most popular stochastic approach is the scheme called Stochastic Perturbations of Physical Tendencies (SPPT) e.g., Buizza et al. [2], see also its more flexible version Christensen et al. [4]. Another stochastic technique involves *additive* perturbations (called additive inflation in data assimilation) e.g., Houtekamer et al. [6], Tsyrulnikov and Gayfulin [17]. The following techniques are also widely used: parameter perturbation schemes Ollinaho et al. [9], Christensen et al. [3], the Stochastic Kinetic Energy Backscatter scheme (SKEB) Shutts [13], the stochastic convection scheme Kober and Craig [7], and the Stochastic Convection Backscatter Shutts [14]. See also the review by Leutbecher et al. [8] and other references therein. The common shortcoming of *all* the numerous above-mentioned schemes is their lack of objective justification and objective parameter estimation.

The goal of this research is to objectively estimate and stochastically model the multivariate spatiotemporal model-error field using the coarse-graining approach Shutts and Palmer [16], Shutts and Pallarès [15]. According to this approach, a higher-resolution "true model" is introduced and used (after upscaling or coarse-graining to the model resolution — hence the name of the approach) to explicitly evaluate model errors with respect to the true model. The limitation of this approach is its reliance on a *"true model"*, which of course is also approximate (as any model). The advantage of the coarse-graining approach is that it offers an opportunity to carefully and rigorously identify and estimate a model for *proxy model errors*.

Our approach is as follows.

1. As a "model", take the COSMO model with a relatively high, convection-permitting resolution (2.2 km).

2. As a "true model", take the same model but with a significantly higher resolution and more sophisticated physical parameterizations.

3. Start the "model" and the "true model" from the same initial conditions and compute the two short-time tendencies.

4. Upscale (coarse-grain) the "true-model" tendencies to the "model" resolution and explicitly compute the model-error field.

5. Accumulate a sample of model-error fields.

6. Use this sample to build a multivariate spatiotemporal stochastic model for model errors.

---

[1]When forecast models become inherently stochastic, there will be no need for the additive model error field, and then the stochastic ("perturbed") forecast model will be directly applied to the initial and boundary ensembles.

Note that Tsyrulnikov [19] used a similar approach that involved a "true model" but for a much simpler pair of "model" and "true-model" (vorticity equation vs. the shallow-water model).

2 Model error definition

2.1 Standard definition

To define the model error we follow Orrell et al. [10] but convert their definition from the time continuous to the time discrete form. Let the time continuous forecast model equation be

$$\frac{d\mathbf{x}^{\mathrm{mod}}}{dt} = \mathbf{F}(\mathbf{x}^{\mathrm{mod}}), \tag{1}$$

$\mathbf{x}^{\mathrm{mod}} \in \mathcal{X}^{\mathrm{mod}}$ is the model state vector, $\mathcal{X}^{\mathrm{mod}}$ is the model state space, and $\mathbf{F}(\mathbf{x}^{\mathrm{mod}})$ is the model operator. Numerically integrating this equation yields its time discrete solution

$$\mathbf{x}_k^{\mathrm{mod}} = \mathcal{M}(\mathbf{x}_{k-1}^{\mathrm{mod}}), \tag{2}$$

where k denotes the time step and $\mathcal{M}(\mathbf{x}^{\mathrm{mod}})$ is the time discrete model operator.

Next, let the truth be denoted by $\mathbf{x}_k^{\mathrm{tru}} \in \mathcal{X}^{\mathrm{tru}}$, where $\mathcal{X}^{\mathrm{tru}}$ is the true state space.

Finally, postulate that there is a *projection* (see, e.g., Tsyrulnikov [19]) $\mathbf{\Pi} : \mathcal{X}^{\mathrm{tru}} \to \mathcal{X}^{\mathrm{mod}}$ such that any true state is mapped to the model space, getting a model state (denoted by the tilde):

$$\widetilde{\mathbf{x}}^{\mathrm{tru}} = \mathbf{\Pi}\mathbf{x}^{\mathrm{tru}}. \tag{3}$$

Normally, elements of $\mathcal{X}^{\mathrm{tru}}$ are *continuous* fields in physical space (as opposed to space-discrete fields in $\mathcal{X}^{\mathrm{mod}}$), in this case the action of $\mathbf{\Pi}$ amounts to *upscaling* the true field to the model-space fields resolution[2]. Now we are in a position to define the model error. To this end, we start the "model" from the true initial conditions $\mathbf{x}_{k-1}^{\mathrm{tru}}$ at time $k-1$ (that is, from the upscaled truth $\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}$), compute the one-step model forecast, and compare it with the (upscaled) truth at time k.

The difference is solely due to the inability of the model to predict the true evolution of the system (atmosphere) and therefore it is called the model error:

$$\boxed{\varepsilon_k = \mathcal{M}(\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}) - \widetilde{\mathbf{x}}_k^{\mathrm{tru}}.} \tag{4}$$

Adding and subtracting $\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}$ from the right-hand side of Eq.(4) shows that the model error is the model *tendency* error as well:

$$\boxed{\varepsilon_k = \left[\mathcal{M}(\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}) - \widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}\right] - \left[\widetilde{\mathbf{x}}_k^{\mathrm{tru}} - \widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}\right] \equiv \mathbf{T}_k^{\mathrm{mod}} - \widetilde{\mathbf{T}}_k^{\mathrm{tru}},} \tag{5}$$

where $\mathbf{T}$ stands for the one-step (total) tendency.

2.2 Definition of model error that assumes that there is a true model

It is essential that the model operator is applied in Eq.(4) to the (upscaled) **truth**. To understand why this is required, let us hypothesize that there is a *true model*:

$$\mathbf{x}_k^{\mathrm{tru}} = \mathcal{M}^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}}), \tag{6}$$

where $\mathcal{M}^{\mathrm{tru}}$ is the operator of the true model. Note that from now on the superscript "tru" denotes the true model (not the truth as in section 2.1).

---

[2]It is also plausible that the truth or the "true model" involves more fields (e.g. additional air constituents) than the model. In that case, the projector $\mathbf{\Pi}$ just ignores the fields present in the truth or the "true model" but absent in the "model". This situation is not considered in this study.

Substituting $\mathbf{x}_k^{\mathrm{tru}}$ from Eq.(6) into Eq.(4) yields

$$\varepsilon_k = \mathcal{M}(\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}) - \widetilde{\mathcal{M}}^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}}), \tag{7}$$

where $\widetilde{\mathcal{M}}^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}}) \equiv \mathbf{\Pi}\mathcal{M}^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}})$. From Eq.(7), we see that the model error is the difference of the one-time-step *model* solution and the one-time-step *true-model* solution **provided that the two models are initialized from the same state** (up to the subgrid-scale field components). We call this requirement the *same start condition*. Informally, the same start condition is very meaningful: in order to compare the two models we specify the same *inputs* and look at the *outputs* so that difference is only due to the difference in the model's operators.

<u>Remark</u>. The above model error definition can also be viewed as follows. If we take the upscaled truth at the two consecutive time steps $k-1$ and k and substitute them into the forecast-model Eq.(2), then Eq.(2) will not be exactly satisfied because the truth is not governed by the (inevitably approximate) forecast model equation. The discrepancy is, by Eq.(4), the model error. Formulated this way, the model error definition appears to exactly correspond to the definition of the *truncation error* of a numerical scheme in solving a differential equation. Indeed, following eg Richtmyer and Morton [12], section 1.6, we substitute the exact solution to the differential problem (i.e. the truth we seek to approximate) into the finite-difference scheme (the "approximating model") and call the residual the truncation (model) error.

Equation (7) can be rewritten in terms of *tendencies* by subtracting and adding $\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}$ in its right-hand side and rearranging the terms:

$$\boxed{\varepsilon_k = \left[\mathcal{M}(\widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}}) - \widetilde{\mathbf{x}}_{k-1}^{\mathrm{tru}} \right] - \mathbf{\Pi} \left[\mathcal{M}^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}}) - \mathbf{x}_{k-1}^{\mathrm{tru}} \right] \equiv \mathbf{T}_k^{\mathrm{mod}} - \widetilde{\mathbf{T}}_k^{\mathrm{tru}}.} \tag{8}$$

We reiterate that here and for the remainder of the article $\widetilde{\mathbf{T}}_k^{\mathrm{tru}}$ denotes the tendency of the true model, not the true tendency as in Eq.(5). Technically, since the true model may have a shorter time step, we allow the tendencies to be computed for several time steps so that the *tendency forecasts* used to compute $\mathbf{T}_k^{\mathrm{mod}}$ and $\mathbf{T}_k^{\mathrm{tru}}$ have the same lead time.

This generalization can also be used to check if there is an initial transient process due to possible imbalances in the starting points (fields) from which the tendency forecasts are initialized (by comparing the tendency fields for different tendency-forecast lengths, see section).

To summarize, the standard definition of model error, Eq.(4), assumes that the truth at the time instants in question, $k-1$ and k, that is, $\mathbf{x}_{k-1}^{\mathrm{tru}}$ and $\mathbf{x}_k^{\mathrm{tru}}$, is the *actual truth*, that is, the truth actually observed in the nature by real-world observations. By contrast, the definition that makes use of a true model, Eqs.(7)–(8), allows us to lift this restriction and assume that the $\mathbf{x}_{k-1}^{\mathrm{tru}}$ is **can be *any* point** on the true-model "attractor" (in practice, not far from the true-model "attractor", that is, with reasonably balanced initial conditions for the true model).

3 Evaluation of model error

3.1 Motivation

Tsyrulnikov and Gorin [18] tried to use the *standard* definition of model error, Eq.(4), to evaluate model errors through comparing finite time model tendencies with finite-time observed tendencies. They found in a simulation study with the COSMO model (with 20 km horizontal resolution) that the main obstacle was the requirement to start the model from the truth. Starting from analyses led to too high estimation errors. In order to make the analyses accurate enough to reliably estimate even the simplest constant-in-space and piecewise-constant-in-time model error of realistic magnitude, the assimilated observations of temperature and wind needed to have currently unreachable accuracy (0.1 K for temperature and 0.02 m/s for winds) and be available at each model grid point. Hence, reliable estimation of realistic model errors by comparing finite-time model tendencies with finite-time observed tendencies is **not possible** with existing observational
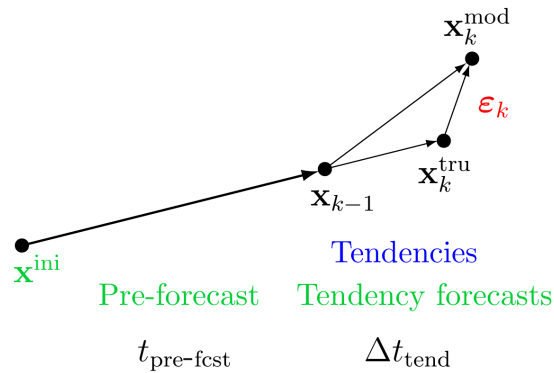
Figure 1: Model error evaluation schematic. Dots denote points in state space. Differences between the true state space and model state space are not highlighted for simplicity here.

networks. This has motivated the present research, in which we explicitly specify a **"true model"** and use the extended definition of the model error introduced in section to evaluate the model error.

3.2 Notation

The "model" in question is referred to as the coarse-grid model (cgm, also abbreviated as mod). The coarse grid is abbreviated as cg. Fields on cg are denoted by the tilde.

The "true model" is referred to as the fine-grid model (fgm, also abbreviated as tru). The fine grid is abbreviated as fg. By "grid-scales" (abbreviated as gs) we mean scales resolved on cg. By "subgrid-scales" (abbreviated as sgs) we mean scales *not* resolved on cg (but resolved by fgm).

3.3 The proposed approach

In order to generate the common starting point for the two tendency forecasts and apply the model-error evaluation methodology described in section , we run a cgm *pre-forecast* (to "spin" the model up). The pre-forecast starts from the initial point $\mathbf{x}^{\mathrm{ini}}$ generated from a global model. The alternative approach in which the pre-forecast is performed using fgm will be investigated on a later stage of the project.

At the end of the pre-forecast, we obtain the cgm forecast fields denoted by $\mathbf{x}_{k-1}$ in Fig.1 and by $\widetilde{\mathbf{x}}_{\mathrm{start}}$ in what follows. The cgm *tendency forecast* starts immediately from $\widetilde{\mathbf{x}}_{\mathrm{start}}$. The fgm *tendency forecast* starts from a *downscaled* version of $\widetilde{\mathbf{x}}_{\mathrm{start}}$ denoted by $\mathbf{x}_{\mathrm{start}}$.

Then, we run two very-short-term tendency forecasts *of the same lead time* $\mathbf{x}_k^{\mathrm{mod}}$ and $\mathbf{x}_k^{\mathrm{tru}}$, compute the two tendencies, downscale the fgm tendency, and finally compute the model error field $\boldsymbol{\varepsilon}_k$ following Eq.(8). A more precise and detailed description of our approach is given in section 4.2 .

3.4 Upscaling

As noted in section , an upscaling (coarse-graining, aggregation) is needed to properly project a high-resolution fgm field onto the cgm phase space. The upscaling removes the scales not represented on the coarse model grid and thus makes the true field *comparable* with its model counterpart. Note that not performing upscaling (i.e. simply *restricting* the high-resolution fields on a coarse grid) would give rise to the phenomenon known as *aliasing* so that the sgs field components would be superimposed on the gs field components, irreversibly distorting them.

Normally, the resolution of fgm is higher than cgm not only in space but also in time (shorter time steps),

which implies that the upscaling must involve the dime dimension as well as the (three) spatial dimensions. There are two common approaches to upscaling: spectral and physical-space based. A *spectral* upscaling is more common in mathematics. It is performed by Fourier transforming the spatial field $\mathbf{x}^{\text{tru}}$, truncating the resulting expansion at the model-grid resolution, and computing $\widetilde{\mathbf{x}}^{\text{tru}}$ as the inverse Fourier transform of the truncated expansion. This approach exactly removes all sgs spectral components.

In meteorology, a *physical-space* upscaling is more common, it consists in averaging the high-resolution fg field $\mathbf{x}^{\text{tru}}$ over cells of lower-resolution cg, see e.g. Shutts and Palmer [16]. This technique is simpler (as it is local, in contrast to the spectral approach) and more physically appealing, albeit not precise mathematically. We will adopt this physical-space definition to simplify our analysis.

Note that in the context of coarse-graining studies, Shutts and Pallarès [15] used a spectral low-pass filter in the horizontal with a squared exponential (i.e. not rectangular as in the ideal low-pass filter) transfer function. In time, they performed low-pass filtering by ad-hoc averaging. Both filters were applied both to the model (they called it "target") and to the true model ("truth").

4 Numerical experiments

4 cases were studied (all 12 UTC): 1 July and 29 July 2017 ("convective" days) and 17 July and 1 December 2017("non-convective" days).

4.1 Models'setup

The cgm was COSMO with 50 vertical levels and horizontal resolution 2.2 km. The fgm was the same COSMO with the following differences from cgm:

1. The horizontal resolution was 0.55 km.

2. The time step was 5 s in fgm vs. 20 s in cgm.

3. The shallow convection parameterization (Tiedtke) was switched off in fgm whilst switched on in cgm.

4. A 3D turbulence scheme was used in fgm vs. a 1D scheme in cgm.

5. Some more sophisticated options were used in the fgm's cloud and precipitation scheme as well as in the radiation scheme as compared to cgm.

The models domains were centered at 52N 35E, see Fig.2. The outermost grid is cg (250*250 points, 550*550 km, marked in greenish). The second-largest grid in Fig.2 is fg (600*600 points, 330*330 km, marked in pinkish). The innermost grid is the model-error evaluation grid (mesh size 2.2 km, 110*110 points, 220*220 km, marked in bluish). The three domains/grids were nested one in another with the intention to reduce any impact of lateral boundaries in a 3h forecast. This will be useful at a later stage of the project when the developed model-error model is verified in an ensemble prediction system.
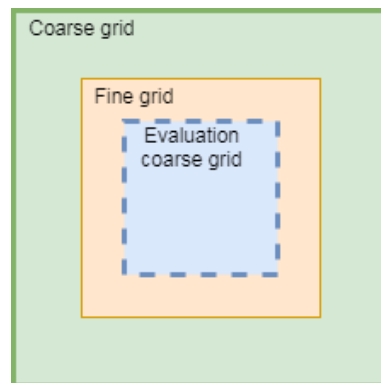
Figure 2: The horizontal grids: fg (the outermost grid), cg (the intermediate grid), and the innermost model-error evaluation grid.

4.2 Computation of the model error

The model-error evaluation technique is summarized as follows.

1. Pre-forecast: run cgm for 1 h lead time. The pre-forecast is used as the cgm *starting point* $\widetilde{\mathbf{x}}_{\text{start}}$.

2. Downscale $\widetilde{\mathbf{x}}_{\text{start}}$ to fg (using the COSMO tool INT2LM). This is $\mathbf{x}_{\text{start}}$. This procedure is meant to approximately satisfy the "same start" condition.

3. Run cgm for 3 time steps (60 s in total) starting from $\widetilde{\mathbf{x}}_{\text{start}}$. Calculate the total tendency $\mathbf{T}_3^{\text{mod}}$.

4. Run fgm for 12 time steps (60 s in total) starting from $\mathbf{x}_{\text{start}}$. Calculate the total tendency $\mathbf{T}_{12}^{\text{tru}}$.

5. Upscale $\mathbf{T}_{12}^{\mathrm{tru}}$ to the coarse grid, getting $\widetilde{\mathbf{T}}_{12}^{\mathrm{tru}}$.

6. Compute the model error as $\boxed{\varepsilon = \mathbf{T}_3^{\mathrm{mod}} - \widetilde{\mathbf{T}}_{12}^{\mathrm{tru}}}$.

The length of the tendency forecasts (60 s) was selected by trial and error. We also tried the tendency-forecast lengths 20 s and 5 min and found that the results presented below were quite stable and not critically dependent on the tendency-forecast length within the above range. With 20 s, there were some indications of an initial imbalance (not shown). The model-error fields for 1 min and 5 min were similar in terms of their spatial scales and variability.

4.3 Results

To give an impression of how model errors are related to model tendencies, we show two (i.e. cgm and fgm) total tendencies for the zonal wind component at an arbitrarily selected model level 41 (about 700 m above ground). One can see that the two tendencies are very similar, implying that the model error (which is their difference) is quite small, as expected (recall that COSMO is an operational-class model used in many countries).
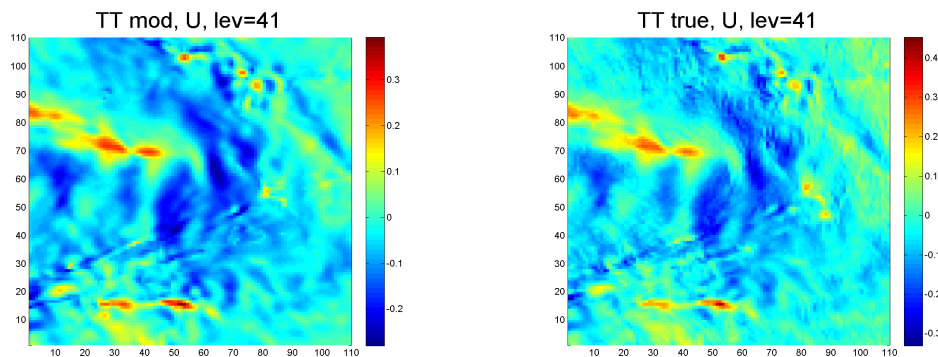
Figure 3: Total tendencies: cgm (left) and fgm (right)

4.3.1 Role of convection

Figure 4 shows the model-error field ε (left) along with the convective physical-tendency field P_{conv} (right) at the model level 32 (about 2.5 km above ground). It is striking that the model-error field ε is dominated by a relatively small number of *outliers*, with the rest of the field being relative close to zero. Comparing the left and the right panels of Fig.4 suggests that it is the convective parameterization that produces those large model errors. At some grid points where the large cgm's P_{conv} is matched with a large tendency produced by the fgm-explicitly-resolved convection — at those points, ε is small. At other points where the cgm's P_{conv} is large, it is *not* matched with a similarly large fgm's tendency — at those points, ε is big (and seen in Fig.4(left) as outliers).

Figure 4: Model errors (left) and convective tendency (right) at the model level 32. The units are K/min.

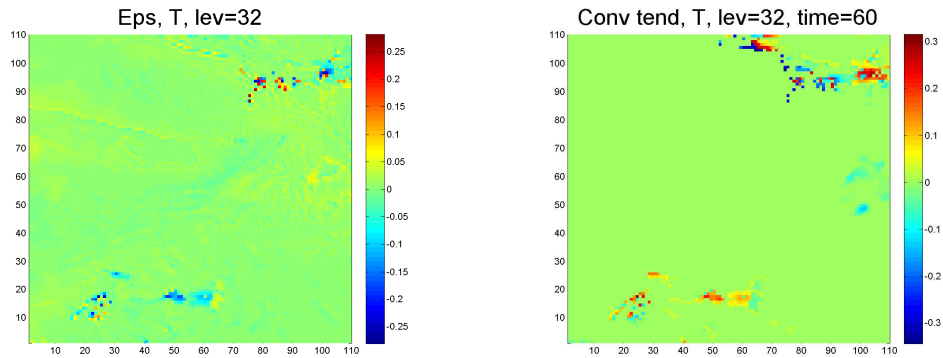This outstanding impact of the shallow convection parameterization (a similar effect of the *deep* convection parameterization in a 7-km-resolution COSMO was even much bigger, not shown) suggests that errors in the convective parameterization need a special treatment (like any outliers in general). We attempted to use predictors like CAPE and the vertical lapse rate to spot those large convective model errors,but those attempts failed dramatically (not shown). We also realized that, given the *complexity* of the convection phenomenon, a purely stochastic approach looks unsuitable to model convective model errors.A physical model is needed. Besides, convection is a *fast* and strong phenomenon so that the convective model errors we can measure are the *outcome* of convection, not its *source*. And it is a "convective source" that we would like to isolate, study, and model in this study (and then perturb in an ensemble forecast).

To verify the conjecture that it the invisible "convective source" that needs to be perturbed, we introduced tiny and constant-in-space-and-time model-error perturbations at all model levels and looked at the resulting forecast perturbation in a 15-min cgm forecast. We imposed model-error perturbations with the magnitude $5 \cdot 10^{-5}$ K per time step in T and 10^{-4} m/s in U, V. The resulting forecast-error perturbation of temperature at the model level 30 is shown in Fig.5(right). The respective model-error field is shown in Fig.5(left). It is seen that both fields look quite similar, which means that realistic convection-related forecast errors can be obtained by just "any" perturbation of temperature and other fields. This can be compared with findings by Flack et al. [5], who introduced temperature model-error perturbations of magnitude 0.1 K with the spatial length scale of $6\Delta x$ (where Δx is the horizontal mesh size) every 15 minutes during the model integration at a single model level at the model hybrid height coordinate 261.6 m.
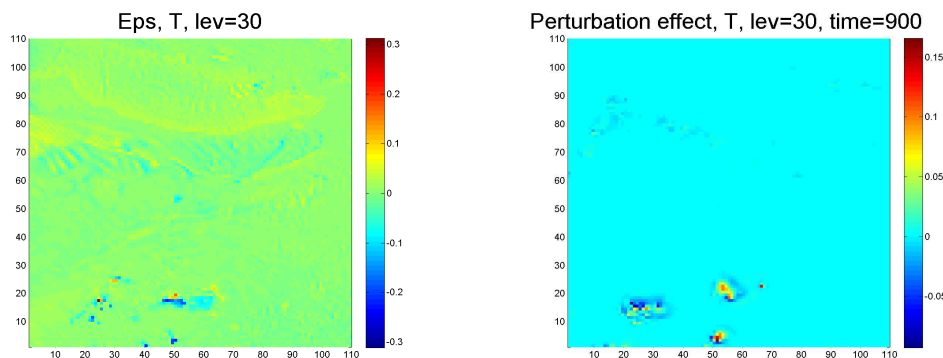

Figure 5: Model errors, K/min, (left) and forecast perturbation, K (right).

4.3.2 Conclusions on convection

The above results indicate that errors in parameterized convection **cannot be treated using the "true-model" approach** (which we pursue in this study) for the following reasons.

1. Evaluating model errors as differences between cgm and fgm short-time tendencies implies, in case of convection, that differences in the *outcomes* of convection are actually measured.

   However, initiation of convection plumes is a sporadic process "modulated" by a hypothetical "convective source" (like CAPE). And it is this source whose uncertainty needs to be modeled, not the outcome. Objectively sensing the uncertainty in this "convective source" is not possible with our approach and, maybe, impossible in principle.

2. A *deterministic* convective parameterization (currently in use in the standard configuration of COSMO adopted in this research) attempts to represent the contribution of subgrid convective elements to the grid-scale fields. However, with the cg mesh as small as 1–2 km, the number of those convective elements (modeled by fgm) in a cg cell is not large enough for their combined effect to be considered as deterministic Shutts and Pallarès [15]. It is inherently random with high variance. And this high variability is the major contributor to the cgm-minus-fgm difference we can measure — because the deterministic cgm tendency produced by the convective scheme is inevitably almost always far from the highly random upscaled fgm tendency. But this random difference is, actually, not the error, it is the *uncertainty* related to the manifestation of the stochastic nature of convection and should not be regarded as model error. This kind of error requires a *stochastic* convective parameterization scheme like Plant and Craig [11]), whose development, estimation, and calibration is beyond the scope of this research.

So, in this work, we do not treat model errors due to parameterized convection. The convective parameterization is switched off both in cgm and fgm, and predominantly non-convective cases are studied.

4.3.3 Non-convective model errors

First, we show the temperature model-error field (ε) at a high enough model level 21 (about 7 km above ground)such that there is, likely, no convection there, see Fig.6. The ε field looks here like a Gaussian random field (in contrast to the above convection-contaminated model-error fields).
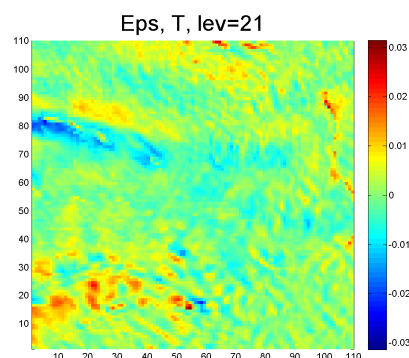

Figure 6: Non-convective model error at level 21, temperature, K/min

Next, Fig.(7) displays the temperature model-error within the planetary boundary layer (level 43 about 500 m above ground). The field looks like a random field with very complicated structure, with multiple scales, and, likely, with multiple components. Building a stochastic model for such field is a very challenging task.

4.3.4 Physical tendency as a predictor for model error

We start with looking at the physical tendency as a potential predictor for model errors. Figure 8 shows the meridional-wind (V) model-error field (left) along with the physical-tendency field P (right) at the model level 31 (about 3 km above ground). It is clearly visible that if P is large, then ε is also, usually, large, so that the physical tendency is indeed a useful indicator of the magnitude of model error. However, there are several areas where the model-error field ε is large while the physical-tendency field P is small. This implies that physical tendency is, actually, of limited use as a model-error predictor.

To get a deeper understanding of the relationship between the model error and the physical tendency, we estimated the conditional probability density of ε given the absolute value of P, i.e. $p(\varepsilon\,|\,\mathsf{abs}(\mathsf{P}))$. In the estimator, 2 percent of largest $|\varepsilon|$ and $|\mathsf{P}| \equiv \mathsf{abs}(\mathsf{P})$ were taken down with the intention to filter out grid points with convection. Values of $|\mathsf{P}|$ were then grouped in 10 equipopulated bins for which histograms of ε were plotted. As an example, the histogram of ε for the 4-th bin of $|\mathsf{P}|$ (temperature, level 30) is displayed in Fig.9.

Remarkably, this conditional distribution is seen to be not too far from Gaussian. More qualitatively, its kurtosis is 4.0 (the Gaussian kurtosis is 3).For other variables and levels, kurtosis remained, mainly, between 3 and 4,thus indicating that Gaussianity is, perhaps, a reasonable hypothesis for the probability distribution of non-convective model errors given the physical tendency, $\varepsilon\,|\,\mathsf{abs}(\mathsf{P})$.

Next, having estimated $p(\varepsilon\,|\,\mathsf{abs}(\mathsf{P}))$, we used it to examine the conditional variance of non-convective model errors: $\mathsf{Var}\,(\varepsilon\,|\,\mathsf{abs}(\mathsf{P}))$. We computed scatterplots of ε^2 vs. P^2 and then smoothed them using a kernel smoother with the Epanechnikov kernel and an empirically selected bandwidth. The resulting dependencies are exem-
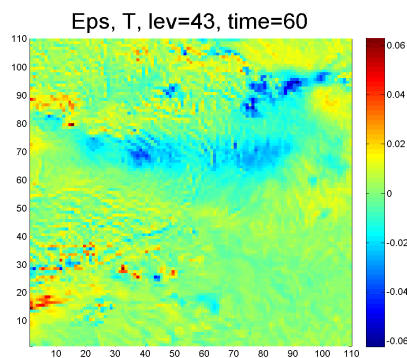

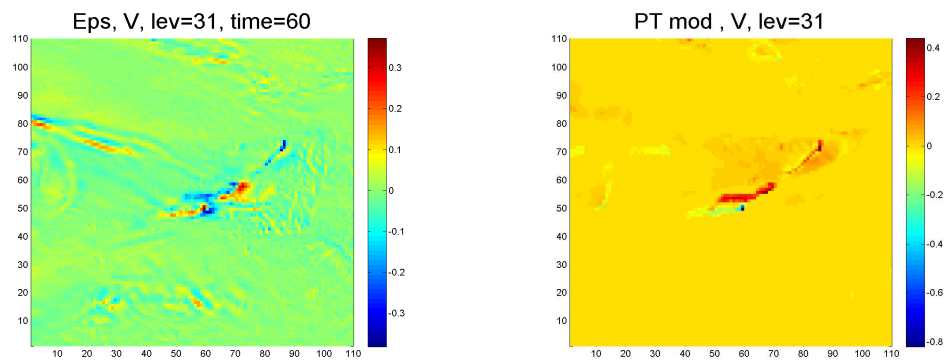Figure 7: Non-convective model error at level 43, temperature, K/min

Figure 8: Model error (left) and physical tendency (right), meridional wind, $\mathrm{m}s^{-1}\mathrm{min}^{-1}$

plified in Fig.10 for temperature at the model level 28 (about 4 km above ground), where the x-axis is P^2 and the y-axis is ε^2.

The resulting conditional variance curves were somewhat noisy even after smoothing but two their salient features were standing out. First, there always was a significantly non-zero offset (the value of $\mathbb{E}\varepsilon^2$ for $P = 0$), which can be interpreted as the variance of the *additive* (physical-tendency independent) model-error component.Second, the model-error variance, by and large, grew with the increasing physical tendency. The growth was, in a first approximation, linear, thus suggesting that it can be interpreted as the *multiplicative* (physical-tendency dependent) model-error variance. As a result, the following **additive-multiplicative** model-error model is our first (and preliminary) finding:

$$\boxed{\varepsilon(\mathbf{s}) = \alpha(\mathbf{s}) + \mu(\mathbf{s}) \cdot P(\mathbf{s}),} \tag{9}$$

where $\alpha(\mathbf{s})$ is the additive model-error component and $\mu(\mathbf{s})$ is the random multiplier field.

To a first approximation, $\alpha(\mathbf{s})$ and $\mu(\mathbf{s})$ can be assumed to be Gaussian random fields with their horizontal, vertical, temporal, and multivariate structure to be identified.

Finally, in Table 1 we show the relationship between the magnitudes of the additive and multiplicative model-error components. One can see that the magnitudes of the additive error components were somewhat larger than the magnitudes of the multiplicative error components. Only in the boundary layer (where turbulence dominates the physical tendency), the multiplicative errors were comparable to additive errors or even larger (not shown).

The difference between the values for U and for V is, perhaps, due to insufficient statistics.
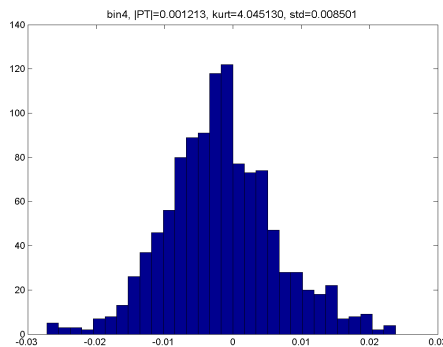

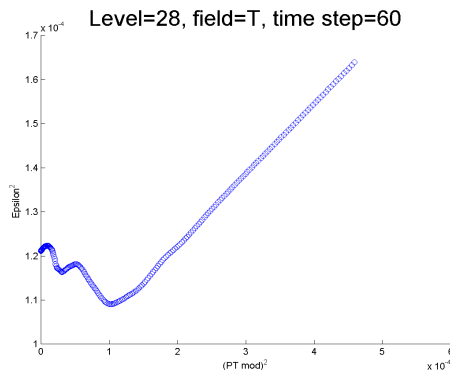Figure 9: Histogram of ε for the 4th bin of P. Temperature, level 30.

Figure 10: Conditional variance $\mathsf{Var}\left(\varepsilon \mid \mathsf{abs}(P)\right)$ for non-convective model errors. Temperature, level 28.

Table 1: Vertically averaged ratio of multiplicative to additive error st.dev.

	T	U	V
$\frac{\text{s.d. (mult)}}{\text{s.d. (add)}}$	0.5	0.5	0.8

5 Conclusions

Approach and preliminary stage results of a model-error objective estimation and modeling study are presented. The approach is based on the comparison of model's instantaneous tendencies with those of a higher-resolution "true model". The "true-model" tendencies are upscaled (coarse-grained) to the resolution of the model-in-question and subtracted from the (total) tendencies of the latter, yielding a proxy to the model-error fields. The goal of the study is a multivariate spatiotemporal stochastic model-error model. The model-error model is to be identified and estimated from a sample of the proxy model-error fields.

The model in question (the coarse-grid model) is COSMO with the horizontal resolution 2.2 km and 50 levels in the vertical.The high-resolution ("true") model is COSMO with the horizontal resolution 0.55 km, the same vertical grid, and a number of differences in the setup of the physical parameterizations. Preliminary results show that, first, errors in convective parameterization cannot and should not be tackled with the coarse-graining approach. Second, we found that non-convective model errors have both additive and multiplicative components. The additive model-error component is independent of the physical tendency and approximately Gaussian. The multiplicative model-error component is proportional to the physical tendency, with the multiplier being, again, approximately Gaussian. Third, the spatial structure of the non-convective model-error field is too complex to be modeled with a reasonably simple model-error model, especially in the lower troposphere and in the planetary boundary layer. This suggests that process-level model-error treatment is to be attempted.

Next steps (which are underway) are the following. Technically, (i) an even higher-resolution true model (fgm) is to be used, (ii) a more careful treatment of static fields (including orography), soil fields, and all initial fields in the two models is to be employed (so that their coarse-grid-resolved components be the same for the two models).Conceptually, (i) the starting point of the tendency-forecasts is to be computed using the true model (fgm) instead of the model (cgm) — to better represent the role of subgrid scales in the formation of model errors,(ii) errors due to different physical parameterizations are to be treated separately whenever possible, (iii) a spatial (horizontal and vertical), spatiotemporal, and multivariate (mutual dependencies between temperature, winds, etc.) aspects are to be addressed in the model-error stochastic modeling.

References

[1] J. Berner, S.-Y. Ha, J. Hacker, A. Fournier, and C. Snyder. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, 139(6):1972–1995, 2011.

[2] R. Buizza, M. Miller, and T. Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, 125(560):2887–2908, 1999.

[3] H. Christensen, I. Moroz, and T. Palmer. Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *J. Atmos. Sci.*, 72(6):2525–2544, 2015.

[4] H. Christensen, S.-J. Lock, I. Moroz, and T. Palmer. Introducing independent patterns into the stochas-

tically perturbed parametrization tendencies (SPPT) scheme. *Quart. J. Roy. Meteor. Soc.*, 143(706): 2168–2181, 2017.

[5] D. L. Flack, S. L. Gray, R. S. Plant, H. W. Lean, and G. C. Craig. Convective-scale perturbation growth across the spectrum of convective regimes. *Monthly Weather Review*, 146(1):387–405, 2018.

[6] P. Houtekamer, H. L. Mitchell, and X. Deng. Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, 137(7):2126–2143, 2009.

[7] K. Kober and G. C. Craig. Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, 73(7):2893–2911, 2016.

[8] M. Leutbecher, S.-J. Lock, P. Ollinaho, S. T. Lang, G. Balsamo, P. Bechtold, M. Bonavita, H. M. Christensen, M. Diamantakis, E. Dutra, et al. Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143 (707):2315–2339, 2017.

[9] P. Ollinaho, S.-J. Lock, M. Leutbecher, P. Bechtold, A. Beljaars, A. Bozzo, R. M. Forbes, T. Haiden, R. J. Hogan, and I. Sandu. Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, 143(702):408–422, 2017.

[10] D. Orrell, L. Smith, J. Barkmeijer, and T. Palmer. Model error in weather forecasting. *Nonlin. Proc. Geophys.*, 8(6):357–371, 2001.

[11] R. Plant and G. C. Craig. A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, 65(1):87–105, 2008.

[12] R. D. Richtmyer and K. W. Morton. Difference methods for initial-value problems. *Wiley*, 1967.

[13] G. Shutts. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, 131(612):3079–3102, 2005.

[14] G. Shutts. A stochastic convective backscatter scheme for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, 141(692):2602–2616, 2015.

[15] G. Shutts and A. C. Pallarès. Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Phil. Trans. R. Soc. A*, 372(2018):20130284, 2014.

[16] G. Shutts and T. Palmer. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *Journal of Climate*, 20(2):187–202, 2007.

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

[18] M. Tsyrulnikov and V. Gorin. Are atmospheric-model tendency errors perceivable from routine observations? *COSMO Newsletter No. 13*, pages 3–18, 2013.

[19] M. D. Tsyrulnikov. Stochastic modelling of model errors: A simulation study. *Quart. J. Roy. Meteor. Soc.*, 131(613):3345–3371, 2005.

Implementation of the new cloud-radiation scheme in COSMO

Pavel Khain[1], Harel Muskatel[1] and Ulrich Blahak[2]

[1]*Israel Meteorological Service*
[2]*Deutscher Wetterdienst*

1 Introduction

Incoming solar radiation is a primary driving source of atmospheric weather and climate processes. For realistic weather simulation, an NWP model has to include an appropriate parametrization of the radiative transfer through the atmosphere. The divergence of solar and thermal radiative fluxes in the atmosphere, which interact strongly with gases, aerosols and the simulated cloud field and its inherent properties, contributes considerably to the diabatic forcing in the prognostic model equations. At the earth's surface radiative fluxes constitute the major forcing for the thermodynamic state of the soil and the interaction with the atmosphere via turbulent fluxes of heat and moisture. In COSMO, the radiative transfer scheme is based on the solution of the δ-two-stream version of the radiative transfer equation incorporating the effects of scattering, absorption, and emission by cloud droplets and ice crystals, gases (water vapor, ozone, carbon dioxide, air molecules) and aerosols in each one of the eight spectral intervals [15, 3]. Optical properties are computed from relevant prognostic and/or diagnostic model variables like specific humidity, cloud water- and ice content and cloud fraction. Some layer properties, like ozone, carbon dioxide and aerosols are specified as climatological values. In particular, the spatially variable aerosol distribution is derived from a climatology provided by Tanre [17] (namelist parameter itype_aerosol=1). The actual layer mean values of optically relevant substances are converted to radiative properties like optical depth τ, single scattering albedo ω and asymetry parameter g and forward-scattered fraction through the use of empirical relations described in [15]. As part of the COSMO priority project "Testing and Tuning of the Revised Cloud Radiation Coupling" $T^2(RC)^2$, the calculation of the optical properties at the model layers was significantly revised, and an additional version of a radiative solver was implemented. From a technical point of view, the new parametrizations can be activated via compilation with the "DCLOUDRAD" preprocessor flag. The changes can be divided into three topics: radiative solver, clear sky optical properties, and cloudy part optical properties.

2 Implementation of the new scheme

Radiative solver

Radiation transfer schemes are one of the most computational expensive components in numerical weather prediction (NWP) models. In COSMO model, with only eight spectral intervals, a full radiation calculation costs as much as eight times the cost of the entire COSMO model run. Most of NWP models compromise on the spatial and/or temporal resolution of the radiation scheme. In the operational setup of COSMO-2.8 km, with a full spatial resolution and with a temporal resolution of 15 minutes, the computational cost of radiation is only 3% of the entire model. This compromise can lead to local biases in net downward radiation and surface temperatures. In an attempt to both reduce errors and to decrease the run-time we implemented a different approach which is to decrease the spectral resolution by a wise sampling technique, a method known as Monte Carlo Spectral Integration (MCSI) [13] was implemented (namelist parameter itype_mcsi).

Many radiative transfer schemes including COSMO scheme [15], use the k-distribution method for the gases-radiation interaction calculations [4]. In this method the spectrum is transferred from wavelength space to cumulative probability space. This space is divided to intervals which are called g-points. In COSMO for each gas and for each spectral interval there are between two to eight g-points. In the operational mode of COSMO the Fast Exponential Sum Fitting Technique (FESFT) is used to fully calculate all of the mentioned g-points. In MCSI only one g-point is calculated in each time step according to its probability. In COSMO we used a softer version of MCSI where a g-point is selected in each of the spectral intervals which increases the computational cos but does neglect either of the spectral intervals in every time step. Of course that if the user chooses to use MCSI the radiation scheme should be called more frequently. We found out the using the MCSI with full temporal resolution (calling the radiation scheme every time step) in COSMO-2.8km setup can increase runtime by 33% with only slight reduction of global radiation and 2-meter temperature biases compared to FESFT. But using MCSI with a 100 seconds temporal resolution (every 5 time steps) can give the same benefits but with only 4% increase in runtime.

Clear sky optical properties

Two new options of an aerosol climatology were introduced (namelist parameter itype_aerosol). The first, Tegen [18] (itype_aerosol=2), is a 2-dimensional monthly map of optical thicknesses for 5 aerosol classes. In COSMO it is interpolated in time, and 3-dimensional optical properties are calculated assuming a predefined exponentially decaying vertical profile. The second, Kinne [10] (itype_aerosol=3), is a 2-dimensional climatology which is considered to better describe real aerosol loading [12].

In addition, two new options to use time- and space-interpolated (via the int2lm software) 3-dimensional aerosol fields of external prognostic forecast models have been implemented. The first (itype_aerosol=4) can process CAMS-ECMWF [1, 11] 3-dimensional aerosol mixing ratio fields, which include sea salt, mineral dust, black carbon, organic matter and sulphate and which are sub-divided to eleven tracers, because sea salt and dust have three size bins while black carbon and organic matter have both hydrophobic and hydrophilic types. The second new option (itype_aerosol=5) can process ICON-ART [14] 3-dimensional aerosol mixing ratio fields; currently the operational ICON-ART only includes mineral dust, but it might be expanded to other species in the future.

Cloudy part optical properties

First, in addition to cloud water and ice, the optical effect of prognostic snow, graupel and rain water contents was (optionally) included (namelist parameter lrad_incl_qrqsqg). The optical properties of solid particles in clouds (specific extinction coefficient β, single scattering albedo ω, asymmetry factor g and delta-transmission function δ) have been formulated as function of effective radius R_e and aspect ratio A_r (assuming hexagonal needles as described in [5]) for the 8 COSMO spectral bands, using the spectroscopic scattering function data for single needles used previously in [5], [6] and [7]. Based on these data, for each realization of a Monte-Carlo-Ensemble over 7000 different Gamma-type ice particle size distributions the parameters R_e following [5], A_r following [7], β, ω, g and δ have been computed. New and rather accurate fits of type rational functions were developed for β and ω as function of R_e, and g and δ as function of A_r ([7]). In contrast to previous literature, our new fits span a very large parameter range for R_e from 2.5 to 300 microns and behave asymptotically "reasonably well" for larger sizes. This range is sufficient for the fits to be applied to the snow- and graupel hydrometeors in any model. Optical thickness τ is obtained by multiplying the respective β for each hydrometeor type by the respective specific mass content and summation. Usage of the new fits can be activated by namelist parameter iradpar_cloud=4, and small modifications can be chosen by the namelist switches lrad_ice_smooth_surfaces and lrad_ice_fd_is_gsquared.

The optical properties of water particles in clouds have been formulated as function of particles' water content and effective radius for the 8 COSMO spectral bands, using [8] up to 60 micron with an own asymptotically correct extrapolation towards larger sizes up to mm diameters (namelist parameter iradpar_cloud=4).

For the large particles (snow, graupel and rain) a geometrical-optics large-size approximation based on semi-transparent spheres for the optical properties was (optionally) implemented (namelist parameter lrad_use_largesizeapprox).

Several new options for calculation of water contents, effective radii and aspect ratios (both are functions of number concentration and mass concentration) for various hydrometeors were implemented. That includes:

- Estimating N_{Ca} - the number concentration of 3-dimensional hydrophilic aerosol fields using Tegen [18] or CAMS-ECMWF [1, 11] input data.

- Estimating w_{eff} - the subgrid local updraft velocity, using turbulent kinetic energy, radiative cooling and optionally convective velocity scale after Deardorff [2] (namelist parameter lincl_wstar_in_weff).

- Utilization of N_{Ca} and w_{eff} to calculate N_{CCN}, the number concentration of nucleated cloud droplets for computing R_e of cloud water, using the Segal-Khain method [16] (namelist parameters icloud_num_type_rad and icloud_num_type_gscp). icloud_num_type_rad affects the radiation indirect aerosols effect on clouds and icloud_num_type_gscp affects the auto-conversion rate in the 1-momment microphysical scheme.

- Number concentrations of other species (rain, cloud ice, snow and graupel) are either estimated consistently to assumptions on particle size distributions in the 1-moment cloud microphysics scheme, or are prognostic for grid scale clouds in case of the 2-moment scheme.

- "Stratiform" subgrid-scale cloud droplets and ice water contents (LWC_{sgs} and IWC_{sgs}, respectively) are estimated as functions of temperature and humidity. The shallow convection LWC_{sgs} is estimated by one of the 3 following methods: as function of temperature and humidity, similarly to stratiform clouds; as equal to the LWC of COSMO shallow convection scheme (namelist parameter luse_qc_con_sgs); and as fraction of the theoretical adiabatic water content [9] (namelist parameter luse_qc_adiab_for_reffc_sgs). The overall LWC_{sgs} is estimated by the default COSMO method as weighted average of the "stratiform" and "convective" parts, using the corresponding cloud fractions. The grid scale water contents of cloud water and ice, snow, graupel and rain are prognostic variables.

- Effective radii and aspect ratios for cloud droplets and cloud ice, as well as snow, graupel and rain are estimated as function of the corresponding water contents and number concentrations. For situations dominated by subgrid-scale shallow convection, the effective radius of subgrid-scale cloud droplets can be, alternatively, estimated using the "adiabatic" parametrization [9] (namelist parameter luse_qc_adiab_for_reffc_sgs).

The list of parameters of the new cloud-radiation coupling scheme is presented in Table 1 in the Appendix. The Table includes the meaning of each parameter, its type, default value, available range and recommended value.

3 Case Study

Preliminary tests of the new cloud-radiation coupling scheme (implemented in COSMO 5.5) were performed over the eastern Mediterranean (COSMO-IL domain 26-36N, 25-39E) with grid spacing of 2.8 km. The weather event was chosen to be on 16/02/2018. During that day the eastern Mediterranean was in the warm sector of a deep upper air trough approaching from the west (see sattelite image in Figure 1). The SW winds caused desert dust advection into the region. The COSMO runs (driven by IFS data) were initialized on 16/2/2018 00 UTC and produced forecasts up to 16/2/2018 12 UTC.
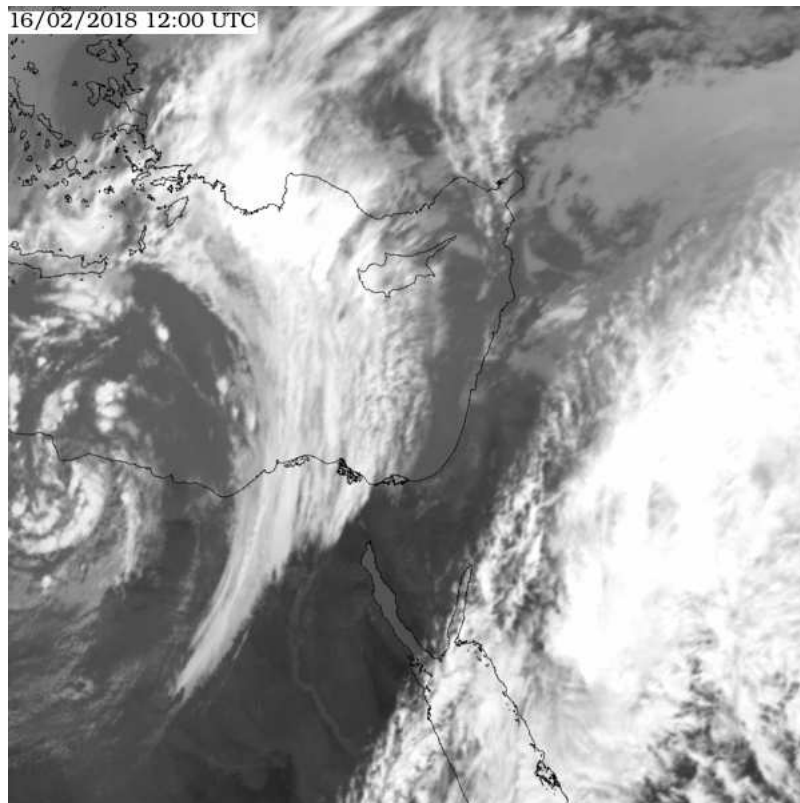
Figure 1: IR 10.8 MeteoSat satellite image for 16/2/2018 at 12 UTC.

Eight COSMO runs have been performed, with namelist parameter variations as summarized in Figure 2. The reference experiment (Ref) includes the default cloud-radiation scheme (iradpar_cloud=1) and Tanre aerosol climatology (itype_aerosol=1). Exp.1 is similar to Ref, with Tegen aerosol climatology (itype_aerosol=2). Exp.2 is similar to Exp. 1 with Segal-Khain estimation of cloud-droplet number concentration (icloud_num _type_rad=2 and icloud_num_type_gscp=2). Exp. 3 is similar to Exp.2 with consideration of Deardorff convective velocity scale in calculation of the local subgrid-scale updraft (lincl_wstar_in_weff=TRUE),and with tuned hydrometeor number concentrations (lreduce_qnx_vs_qx=TRUE). Exp.4 is similar to Exp.3 with subgrid scale droplets and ice effective radius calculation using water contents and number concentrations (luse_reff_ini_x_as_reffx_sgs=FALSE), and with tuned water content reduction (luse_tqcqiqs=TRUE). Exp.5 is similar to Exp.4 with an estimation of shallow Cu droplets effective radius using the "adiabatic" parametrization (luse_qc_adiab_for_reffc_sgs=TRUE), and their water content using the shallow convection parametrization (luse_qc_con_sgs=TRUE). Exp.6 is similar to Exp.5 with revised asymmetry function of ice particles (lrad_ice_smooth_surfaces= FALSE and lrad_ice_fd_is_gsquared= TRUE). Exp.7 is similar to Exp. 6 with MCSI parameterization of spectral bands sampling in the radiation solver (itype_mcsi=1) compensated by more frequent calls to the radiation scheme (every 3 minutes instead of 15).

Parameter	Meaning	Ref	1	2	3	4	5	6	7
iradpar_cloud	type of optical prop.	1	1	4	4	4	4	4	4
lrad_incl_qrqsqg lrad_use_largesizeapprox	include rain, snow and graupel	F F	F F	T T	T T	T T	T T	T T	T T
itype_aerosol	aerosols data source	1	2	2	4	4	4	4	4
icloud_num_type_rad icloud_num_type_gscp	Segal-Khain parameterization	1 1	1 1	2 2	2 2	2 2	2 2	2 2	2 2
lincl_wstar_in_weff	local updraft calc.	F	F	F	T	T	T	T	T
lreduce_qnx_vs_qx	tune number conc.	F	F	F	T	T	T	T	T
luse_reff_ini_x_as_reffx_sgs	effective radius calc.	T	T	T	T	F	F	F	F
luse_tqctqitqs	tune water content	F	F	F	F	T	T	T	T
luse_qc_adiab_for_reffc_sgs	new paramet. for Cu	F	F	F	F	F	T	T	T
luse_qc_con_sgs	shall. C. LWC for Cu	F	F	F	F	F	T	T	T
lrad_ice_smooth_surfaces lrad_ice_fd_is_gsquared	ice particles roughness	T F	T F	T F	T F	T F	T F	F T	F T
itype_mcsi	MCSI paramet.	0	0	0	0	0	0	0	1

Figure 2: Summary of the eight COSMO experiments.

The sensitivity results of the COSMO runs are presented in 3 as function of the forecast range. The upper left panel presents the averaged global radiation over the cloudy grid points (cloud cover > 0.1). For each experiment the global radiation of the Ref run is subtracted, showing the sensitivity effect of the current experiment. The upper right panel presents similar results for the averaged 2 meter tempearture. Similarly, the lower panels present the sensitivity results for the clear sky regions (cloud cover < 0.1), highlighting the direct effects of aerosols and the MCSI parameterization.
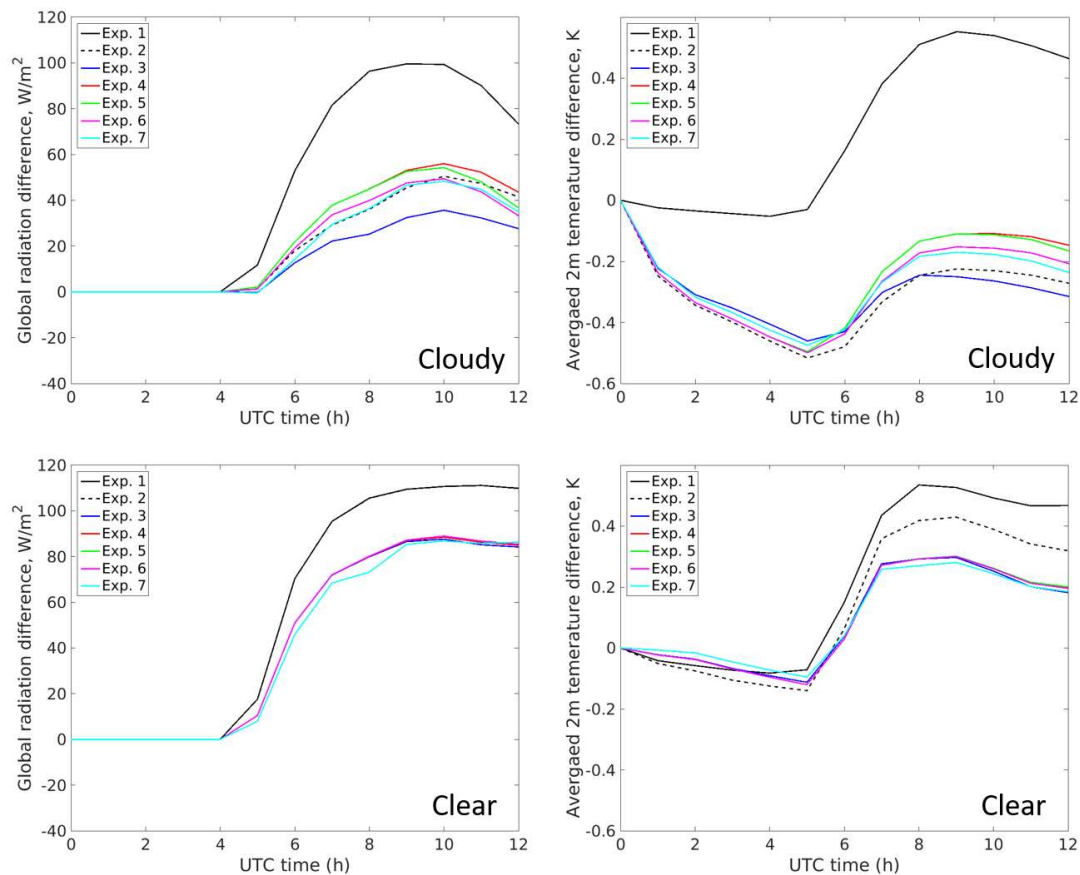
Figure 3: Sensitivity results of the COSMO runs as function of the forecast range. The upper left panel presents the averaged over the cloudy grid points (cloud cover > 0.1) global radiation. For each experiment the global radiation of the Ref run is subtracted, showing the sensitivity effect of the current experiment. The upper right panel presents similar results for the averaged 2 meter tempearture. Similarly, the lower panels present the sensitivity results for the clear sky regions (cloud cover < 0.1).

One can see (Exp. 1) that the use of Tegen aerosol instead of Tanre strongly increases the global radiation (up to 120 W/m^2) and the 2 meter temperature up to 0.5 K. Exp. 2 shows that in the cloudy areas the new optical properties and Segal-Khain nucleation, and – most importantly – consideration of rain, snow and graupel particles in radiation, decreases the enhancement to about 50 W/m^2. Exp. 3 shows that in the cloudy areas revision of the local updraft for Segal-Khain nucleation and tuning the number concentration decreases the enhancement further to about 35 W/m^2. Exp. 4 shows that in the cloudy areas revision of SGS effective radius calculations and imposing upper limits to the total water contents brings the enhancement back to about 50 W/m^2. Experiments 5,6 and 7 show smaller sensitivity on average. Generally, one can see that the new cloud-radiation coupling scheme affects the global radiation by 30-120 W/m^2 which corresponds to a 2 meter temperature variation range of about 1 K. Important to note is, that these results are preliminary and reflect the model sensitivity at a single day over a specified region only. Also, the results include averaging over large areas, which suggests much higher sensitivities locally. We should also note that this is the first attempt to test the code within the 5.5 framework. Each of the new scheme components was massively tested within the 5.1 framework. In the appendix we provide the "recommended" namelist based on the studies during the last 4-years, which were discussed and published in various presentations and papers, as can be viewed on $T^2(RC)^2$ web page (http://www.cosmo-model.org/content/tasks/priorityProjects/t2rc2/default.htm).

4 Summary

In this short article we inform the COSMO community about the recent implementation of a revised cloud-radiation coupling scheme into COSMO 5.5. Officially this code will be distributed with the final version of COSMO - COSMO-6. The new scheme includes an optional modification to the radiation solver (MCSI parametrization). It further includes implementation of new aerosol climatologies and prognostic aerosol fields which modify the clear sky optical properties. Moreover, the indirect effect of aerosols on number concentrations, effective radiuses and water contents in grid and subgrid scale clouds is significantly revised. The optical properties of solid and water hydrometeors for the different spectral intervals were revised as well. Preliminary tests show a significant effect of the new cloud-radiation coupling scheme on radiation and 2 meter temperature.

Acknowledgments

References

[1] Benedetti, A., J.-J. Morcrette, O. Boucher, A. Dethof, R. J. Engelen, M. Fisher, H. Flentje, N. Huneeus, L. Jones, J. W. Kaiser, et al., 2009: Aerosol analysis and forecast in the European Centre for MediumRange Weather Forecasts Integrated Forecast System. Part 2: Data assimilation, J. Geophys. Res.-Atmospheres, 114(D13), D13205, doi:10.1020/2008JD011115.

[2] Deardorff, J. W.: 1970, 'Convective Velocity and Temperature Scales for the Unstable Planetary Boundary Layer and for Rayleigh Convection', J. Atmos. Sci. 27, 1211–1213.

[3] Doms, G., J. Forstner, E. Heise, H.-J. Herzog, D. Mironov, M. Raschendorfer, T. Reinhardt, B. Ritter, R. Schrodin, J.-P. Schulz and G. Vogel, 2011: A description of the Nonhydrostatic Regional COSMO Model. Part II: Physical Parameterization. Deutscher Wetterdienst, 161 pp.

[4] Fu, Q., and Liou, K. N. (1992), A threeparameter approximation for radiative transfer in nonhomogeneous atmospheres: Application to the O3 9.6m band, J. Geophys. Res., 97(D12), 13051– 13058, doi:10.1029/92JD00999.

[5] Fu Q., 1996, An Accurate Parameterization of the Solar Radiative Properties of Cirrus Clouds for Climate Models, J. Clim., 9, 2058-2082

[6] Fu Q., P. Yang and W. B. Sun, 1998, An Accurate Parameterization of the Infrared Radiative Properties of Cirrus Clouds for Climate Models, J. Clim., 11, 2223-2237

[7] Fu Q., 2007, A New Parameterization of an Asymmetry Factor of Cirrus Clouds for Climate Models, J. Atm. Sci., 64, 4140-4150

[8] Hu, Y.X. and K. Stamnes, 1993: An Accurate Parameterization of the Radiative Properties of Water Clouds Suitable for Use in Climate Models. J. Climate, 6, 728–742, https://doi.org/10.1175/1520-0442(1993)006<0728:AAPOTR>2.0.CO;2

[9] Khain, P., R. Heiblum, U. Blahak, Y. Levi, H. Muskatel, E. Vadislavsky, O. Altaratz, I. Koren, G. Dagan, J. Shpund, and A. Khain, 2019: Parameterization of Vertical Profiles of Governing Microphysical Parameters of Shallow Cumulus Cloud Ensembles Using LES with Bin Microphysics. J. Atmos. Sci., 76, 533–560, https://doi.org/10.1175/JAS-D-18-0046.1

[10] Kinne, S., D. O'Donnel, P. Stier, S. Kloster, K. Zhang, H. Schmidt, S. Rast, M. Giorgetta, T. F. Eck, and B. Stevens (2013), MAC-v1: A new global aerosol climatology for climate studies, J. Adv. Model. Earth Syst., 5, 704740, doi:10.1002/jame.20035.

[11] Morcrette, J.-J., O. Boucher, L. Jones, D. Salmond, P. Bechtold, A. Beljaars, A. Benedetti, A. Boner, J. W. Kaiser, M. Razinger, et al., 2009: Aerosol analysis and forecast in the European Centre for MediumRange Weather Forecasts Integrated Forecast System. Part 1: Forward modelling, J. Geophys. Res.-Atmospheres, 114(D6), D06206, doi:10.1029/2008JD011235.

[12] Mueller R., C. Träger-Chatterjee, Brief Accuracy Assessment of Aerosol Climatologies for the Retrieval of Solar Surface Radiation, Atmosphere 2014, 5, 959-972; doi:10.3390/atmos5040959

[13] Pincus, R., and Stevens, B. (2009), Monte Carlo Spectral Integration: a Consistent Approximation for Radiative Transfer in Large Eddy Simulations, J. Adv. Model. Earth Syst., 1, 1, doi:10.3894/JAMES.2009.1.1.

[14] Rieger, D., Bangert, M., Bischoff-Gauss, I., Förstner, J., Lundgren, K., Reinert, D., Schröter, J., Vogel, H., Zängl, G., Ruhnke, R., and Vogel, B.: ICON–ART 1.0 – a new online-coupled model system from the global to regional scale, Geosci. Model Dev., 8, 1659-1676, https://doi.org/10.5194/gmd-8-1659-2015, 2015.

[15] Ritter, B. and J. Geleyn, 1992: A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations. Mon. Wea. Rev., 120, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2

[16] Segal, Y., and Khain, A. (2006), Dependence of droplet concentration on aerosol conditions in different cloud types: Application to droplet concentration parameterization of aerosol conditions, J. Geophys. Res., 111, D15204, doi:10.1029/2005JD006561.

[17] Tanre, D., J.-F. Geleyn, and J. Slingo, 1984. First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. In Proc. of the Meetings of Experts on Aerosols and their Climatic Effects, Williamsburg, VA, pp. 133177. WMO and IAMAP.

[18] Tegen, I., P. Hoorig, M. Chin, I. Fung, D. Jacob, and J. Penner (1997), Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results, J. Geophys. Res., 102, 23,895 23,915.

Appendix

The list of parameters of the new cloud-radiation coupling scheme is presented in Table 1. The Table includes the meaning of each parameter, its type, default value, available range and recommended value.

Parameter	Meaning	Type	Def	Range	Recom
iradpar_cloud	Calculation of optical properties for solid and water particles. 1-old, 4-new. 2,3 are possible but not recommended	INT	1	1,4	4

lrad_ice_smooth_surfaces	Effective if iradpar_cloud=4. If T assume smooth surfaces for solid species (fd>0), otherwise assume rough surfaces (fd close to 0)	LOG	T	T/F	F
rad_ice_fd_is_gsquared	Effective if iradpar_cloud=4 and lrad_ice_smooth_surfaces=T. If T compute forward scattered fraction as $f = g^2$ (RG92 method), otherwise compute $f = 1/(2\omega) + f_d$ with $f_d = fct(AR)$ according to the new fits. Concerns only the solar frequency bands	LOG	F	T/F	F
lrad_incl_qrqsqg	include/exclude QR, QS and QG in radiative transfer calculations	LOG	F	T/F	T
lrad_use_largesizeapprox	Effective for iradpar_cloud = 4: if F new fits for all optical properties of solid species are used without clipping. If T only for the extinction the large-size approximation is applied starting from Reff=150 microns	LOG	T	T/F	T
itype_aerosol	Type of aerosol map. Climatology: 1-Tanre, 2-Tegen, 3-Kinne. Prognostic data from int2lm: 4-CAMS, 5-ART	INT	1	1-5	4
icloud_num_type_rad	Derivation of cloud number concentration for radiation. 1: use cloud_num_rad tuning parameter. 2: derive from Tegen/CAMS aerosol data using Segal-Khain parametrization (effective for itype_aerosol=2,4 only)	INT	1	1,2	2
icloud_num_type_gscp	Derivation of cloud number concentration for 1-moment microphysics. 1: use cloud_num tuning parameter. 2: derive from Tegen/CAMS aerosol data using Segal-Khain parametrization (effective for itype_aerosol=2,4 only)	INT	1	1,2	2
lincl_wstar_in_weff	Effective in case of icloud_num_type_rad/gscp=2 (Segal-Khain). If T, the eff. w for cloud nucleation is enforced to be $>= w^*$ (conv. vel. scale in PBL), but only below the PBL height or below the upper bound of the lowest "convective cloud layer", whichever is higher. F – otherwise	LOG	F	T/F	T
cloud_num_rad	Tuning parameter for cloud number concentration for radiation $(1/m^3)$	REAL	2E8	[0.1-10]E8	2E8
cloud_num	Tuning parameter for cloud number concentration for 1-moment microphysics $(1/m^3)$	REAL	5E8	[0.1-10]E8	5E8

zref_cloud_num_rad	Height of lower layer (above MSL in m) above which the cloud number concentration is exponentially reduced with height	REAL	2000	500-3000	2000
dz_oe_cloud_num_rad	1/e decrease height in m of exponential decrease of cloud number concentration above zref_cloud_num_rad	REAL	2000	500-3000	2000
lreduce_qnx_vs_qx	T: reduce qnx vs qx for radiation. In this case the 9 tuning parameters below are activated. F: otherwise	LOG	F	T/F	T
rhoc_nchigh_rad	For qc<=rhoc_nchigh_rad, qnc is not reduced as function of qc $[kg/m^3]$	REAL	0.5 E-4	[0.1-20] E-4	0.5 E-4
rhoc_nclow_rad	For rhoc_nchigh_rad < qc < rhoc_nclow_rad qnc is linearly reduced as function of qc $[kg/m^3]$	REAL	2.0 E-4	[0.1-20] E-4	2.0 E-4
ncfact_low_rad	For qc>=rhoc_nclow_rad, the linear reduction bottoms out at the ncfact_low_rad'th fraction of qnc	REAL	0.1	[0...1]	0.1
rhoi_nihigh_rad	For qi <= rhoi_nihigh_rad, ni(T) is not reduced as function of qi $[kg/m^3]$	REAL	0.5 E-5	[0.1-20] E-5	0.5 E-5
rhoi_nilow_rad	For rhoi_nihigh_rad < qi < rhoi_nilow_rad, ni(T) is linearly reduced as function of qi $[kg/m^3]$	REAL	2.0 E-5	[0.1-20] E-5	2.0 E-5
nifact_low_rad	For qi >= rhoi_nilow_rad, the linear reduction bottoms out at the nifact_low_rad'th fraction of ni(T)	REAL	0.1	[0...1]	0.1
rhos_n0shigh_rad	For qs <= rhos_n0shigh_rad, n0s is not reduced as function of qs $[kg/m^3]$	REAL	1.0 E-5	[0.1-20] E-5	1.0 E-5
rhos_n0slow_rad	For rhos_n0high_rad < qs < rhos_n0low_rad, n0s is linearly reduced towards n0s_low_rad $[kg/m^3]$	REAL	5.0 E-5	[0.1-20] E-5	5.0 E-5
n0s_low_rad	For qs>=rhos_n0slow_rad, n0s attains this const. value $[m^-3]$	REAL	8 E5	[1-50] E5	8 E5
luse_reff_ini_x_as_reffx_sgs	Use tuning parameters reff_ini_c, reff_ini_i for SGS eff. radius	LOG	T	T/F	F
reff_ini_c	Effective radius for SGS cloud droplets (m). Only if luse_reff_ini_x_as_reffx_sgs=T	REAL	5 E-6	[3-15] E-6	5 E-6
reff_ini_i	Effective radius for SGS cloud ice (m). Only if luse_reff_ini_x_as_reffx_sgs=T	REAL	10 E-6	[5-30] E-6	10 E-6

radqc_fact, radqi_fact, radqs_fact, radqg_fact	Portion of GS and SGS qc, qi, qs, qg (respectively) "seen" by the radiation. Should be <1 because of subgrid-scale variability. Increase leads to higher optical thickness	REAL	0.5	[0.5-1]	0.5
qvsatfact_sgscl_rad	Scaling factor for qc and qi of SGS clouds: local supersaturation which is assumed to have been depleted by SGS cloud formation [-]. Increase leads to higher optical thickness	REAL	0.01	[0.005-0.02]	0.01
luse_tqctqitqs	limit TQC, TQI, TQS to some integral maximum. Adjust qc, qi, qs accordingly (for radiation). T leads to lower optical thickness	LOG	F	T/F	T
luse_qc_adiab_for_reffc_sgs	Use "adiabatic" parametrization for SGS shallow convection effective radius	LOG	F	T/F	T
luse_qc_con_sgs	Effective if luse_qc_adiab_for_reffc_sgs=T. F: use "adiabatic" parametrization for SGS shallow convection LWC. T: use LWC from shallow convection parametrization (if lconv=T)	LOG	F	T/F	T
alpha1_adiab_rad	Linear deviation with height (above cloud base) of SGS shallow convection effective radius from the adiabatic value alpha1_adiab_rad-alpha2_adiab_rad*(z-zcb). [-]	REAL	0.95	[0.7-1]	0.95
alpha2_adiab_rad	Linear deviation with height (above cloud base) of SGS shallow convection effective radius from the adiabatic value alpha1_adiab_rad-alpha2_adiab_rad*(z-zcb). [1/m]	REAL	1.2 E-4	[1-2] E-4	1.2 E-4
beta_adiab_rad	Ratio of cloud-average number concentration (of SGS shallow convection) with respect to the cloud core value (obtained from Segal-Khain)	REAL	0.38	[0.2-1]	0.38
gamma_adiab_rad	Linear deviation with height (above reff=12micron level) of SGS shallow conv. qc from the "pseudo-adiabatic" value. [1/km]	REAL	0.45	[0.2-0.7]	0.45
itype_mcsi	1: Monte Carlo Spectral Integration in the radiation solver. Recommended together with nincrad=5. 0-Default from RG92	INT	0	0,1	0

Table 1: List of parameters of the new cloud-radiation coupling scheme. The parameters are separated to groups according the corresponding parametrization: Optical properties derivation; Effect of large hydrometeors on radiation; Aerosols effect in clear sky and on droplets number concentration in clouds; Reduction of hydrometeors number concentrations for large water contents; Method of effective radius calculation; Tuning water contents "seen" by radiation; "Adiabatic" parametrization for liquid water content and effective radius in shallow cumulus; New method for radiation solver. The Table includes the meaning of each parameter, its type, default value, available range and recommended value.

Impacts on model performance score from CALMO and CALMO-MAX

VOUDOURI A.1 *, CARMONA I.2, AVGOUSTOGLOU E.1, LEVI Y.2, BETTEMS J.M.3 and E. BUCCHIGNANI4

1.Hellenic National Meteorological Service 2.Israel Meteorological Service 3.MeteoSwiss 4.Italian
Aerospace Research Center (CIRA)
*corresponding author e-mail: antigoni.voudouri@hnms.gr

1 Introduction

The priority project CALMO of COSMO (years 2013-2016) induced an objective multivariate calibration method aiming on substituting expert tuning. Expert tuning is a procedure by which free or poorly confined parameters existing in the parameterization schemes of RCM and NWP models are mainly tuned using expert knowledge (Duan et al., 2006; Skamarock, 2004;Bayler et al., 2000).This procedure, performed for specific parameterization schemes addressed by model developers, usually underestimates parameter interactions, follows a non-objective procedure, and is difficult to replicate without a direct involvement of the model developers.
Several studies, over the last years, have been conducted towards substituting expert tuning by objective and automatic methodologies to calibrate unconfined model parameters existing in both NWP and RCM model (Bellprat et al., 2012a and 2012b, Gong et al.,2015, Duan et al., 2016, Gong et al., 2016, Voudouri et al., 2017, 2018).

At the framework of CALMO, the implementation of the calibration method that has been developed by Bellprat et al. (2012a) and implemented for a regional climate model has been applied on COSMO-NWP model using a horizontal resolution of 0.0625o and then tested for a horizontal resolution of 0.02o (approximately 2km) over a mainly continental domain covering the Alpine Arc.
In the priority project CALMO-MAX (years 2017—in progress) additional tests on the advantages ofthe calibration method for COSMO model, using a finer resolution of 0.01o are studied. The steps followed, namely model setup, parameter selection as well as basic differences between the two priority projects are briefly described in Section 2. Results from CALMO to be considered in CALMO-MAX are presented in Section 3. A summary and conclusions are given in Section 4.

2 Data and Methodology

2.1 Model setup

The code used is the refactored version of the COSMO model (Lapillonne and Fuhrer,2014) based on the official version 5.00 of the model, capable of running on GPU-based hardware architectures, operationally used by MeteoSwiss. In CALMO, simulations were performed with COSMO model over the whole year 2013 and operated with horizontal resolution of 0.02o (approximately 2km) for a domain including the wider area of Switzerland and Northern Italy (Fig. 1), while in CALMO-MAX a finer resolution of 0.01° (about 1 km) is used. The vertical extension reaches 23.5 km (30hPa) with 60 model levels for CALMO and 80 for CALMO-MAX. Initial and boundary fields for all tests are derived from the MeteoSwiss operational forecasting archive system at 0.02o horizontal resolution (2km).

In addition, the history of the soil, not used in the configuration of CALMO, is considered in the CALMO-MAX simulations (hindcast mode). Finally, for CALMO-MAX, the soil has been initialized by a 3 years spin-up using TERRA Standalone (TSA).
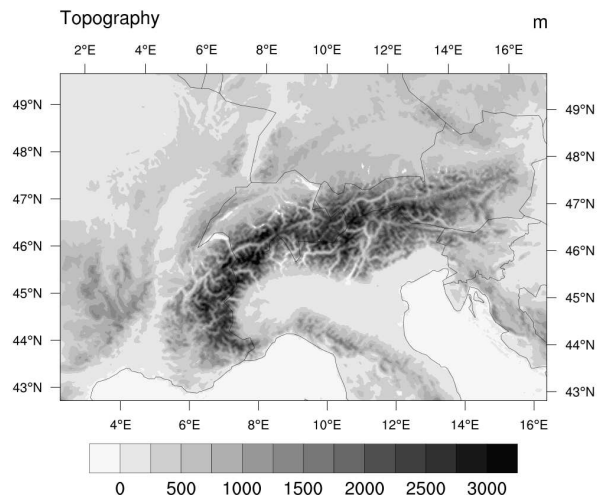

Figure 1: Topography of the simulation area

2.2. Data and selected parameters

The unconfined parameters existing in COSMO model are related to sub-grid scale turbulence, surface layer parameterization, grid-scale clouds, precipitation, moist and shallow convection, radiation, soil scheme etc. (Doms et al., 2011, Gebhardt et al., 2011). Thus, sensitivity experiments using twelve parameters associated with turbulence (tur_len, tkhmin, tkmmin), surface layer parameterization (rat_sea, rlam_heat, crsmin), grid-scale precipitation (v0snow), moist and shallow convection (entr_sc), radiation (rad_fac, uc1) and the soil scheme (c_soil) have been performed and the most sensitive ones have been selected. Figure 2 illustrates the sensitivity of minimum 2m temperature with respect to 7 (left panel) and 5 parameters (right panel) respectively. The red polygon refers to the zero sensitivity "axis" while sensitivities close to zero are depicted with blue bullets. In left panel of figure 2, negative sensitivities, well below the red polygon are depicted with orange bullets while positive sensitivities well above the red polygon are depicted with green bullets. In right panel of figure 2, all negative and positive sensitivities are depicted with green bullets. The dashed polygon line that connects the dots denotes optically the overall sensitivity for the considered meteorological variable especially to the degree that it is convex/concave and mainly in reference to the zero sensitivity red polygon. In CALMO, the six model parameters selected were: asymptotic turbulence length scale, tur_len [m]; minimal diffusion coefficients for heat, tkhmin[m2/s]; scalar resistance for the latent and sensible heat fluxes in the laminar surface layer, rlam_heat [no units]; the surface-area index of the evaporating fraction of grid points over land, c_soil[no units]; the factor in the terminal velocity for snow, v0snow[no units]; and the mean entrainment rate of boundary layer humidity into the shallow convection clouds, entr_sc [m-1]. In CALMO-MAX five parameters were selected: tkhmin, rlam_heat, v0snow(already used in CALMO), and additionally the fraction of cloud water/ice considered for radiation, radfac[no units] and the parameter for computing the amount of cloud cover in saturated conditions, uc1[no units].

The parameters are calibrated against daily minimum and maximum 2m temperature values (Tmax and Tmin respectively) as well as 24h accumulated precipitation (prec). For temperature, available measurements of daily

mean surface air temperature selected at the station network of MeteoSwiss were used. More specifically the interpolated values of all available 2m-temperature maximum and minimum observations over Switzerland to a 2km-grid are provided by Frei (2014).

For precipitation observations over Switzerland, the gridded MeteoSwiss radar composites were used, corrected by rain gauges and interpolated to the model grid. Over Northern Italy, observations interpolated to the model grid were used only where the grid points in the vicinity of the stations get a value. In addition to Tmin, Tmax and prec, radio soundings data and the associated model profiles at grid points near the soundings locations were used. Sunshine duration and 2m dew point temperature are also considered in CALMO-MAX with observational data provided by MeteoSwiss.
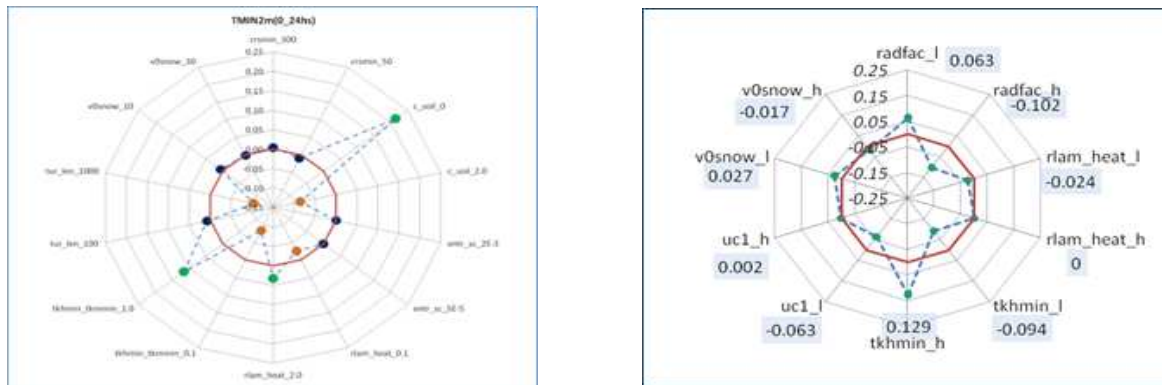

Figure 2: Spider graphs of minimum 2m temperature sensitivity with respect to 7 (left panel) and 5 parameters (right panel)

2.3. Methodology

The calibration methodology is presented in detail in Voudouri et al. 2017 and 2018. It relies on a Meta-Model (MM) that approximates the parameter space, using a multi-variate quadratic regression in an n-dimensional model (Neelin et al., 2010 and 2010a). The specific MM is based on the assumption that changes of the simulated model quantity, due to a parameter perturbation, are smooth and thus can be approximated by a 2nd order polynomial regression.

As a quadratic fit is determined by only three points, this assumption allows fitting the MM by performing a low number of simulations, namely 2N+0.5N(N-1)+1, for N parameters, which is crucial for computationally expensive NWP and RCM models.The use of a quadratic regression further inhibits over-fitting and allows for analytical solutions of the parameter space. Once the MM has been constructed, it can be used as a surrogate to perform a large number of simulations, testing several parameter values in order to find the optimal combination of values.

In CALMO, the MM adapted for COSMO-CLM by Bellprat et al., 2012a, has been consolidated and extended by adding the option not to average Tmax/Tmin over regions, the prediction of multiple vertical profile characteristics, and the possibility of supporting new geographical regions. The quality score to account for model performance was a RMSE-type performance score initially tested in CALMO preliminary phase.Successively, an advanced performance score was introduced based on the COSMO Index (COSI) developed by Damrath (2009). The COSI score combines the RMSE-type for continuous variables and the ETS (Equitable Threshold Score) for categorical fields. The COSI score in CALMO-MAX is updated to include sunshine duration, mean, maximum and minimum dew point temperature, while for precipitation ETS is replaced with the FSS (Fraction Skill Score).

3 Results

During the preliminary phase of CALMO, it has been shown (Voudouri et al., 2017 and 2018) that the MM developed for climate models (Bellprat 2012a, 2012b) can be adapted to COSMO-NWP. Therefore, the objective methodology can be transferred from RCM to NWP. During the second phase of CALMO, an optimal set of six parameters over the entire year has been extracted, as well as monthly optimal values illustrating model parameters sensitivity on different weather types. Figure 3 shows the 24 hour accumulated precipitation (prec), Tmax and Tmin values provided by MM against COSMO simulation results during the year 2013. The related correlation values are 79.9%, 80.6% and 78.2% respectively.
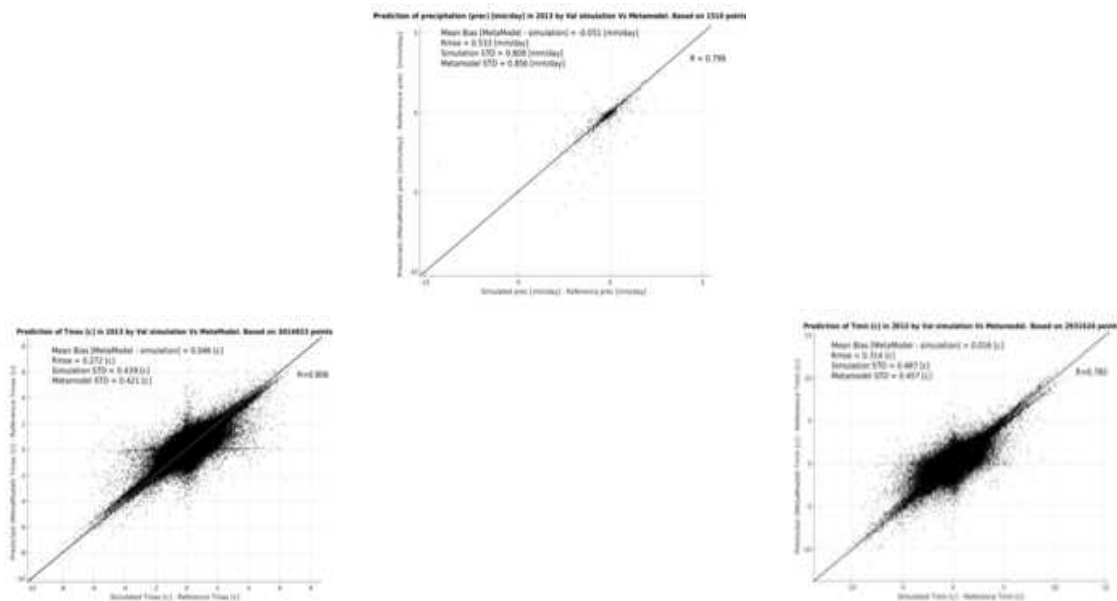

Figure 3: 24h accumulated precipitation (upper panel), Tmax (left panel), Tmin (right panel) Meta-Model prediction for the tested parameter combination, vs. COSMO simulation results during the year 2013. X axis presents the simulated field minus the reference simulation. Y axis presents the Meta-Model field minus the reference simulation.

Monthly and yearly improvement of the model performance associated with daily minimum and maximum 2m temperature, as well as 24h accumulated precipitation when using the set of optimum parameter values, against the values recommended in the default model setup have been investigated. More specifically the annual cycle of the performance score using the optimum set of six parameters is presented in Figure 4. The annual cycle of the improved performance score when the optimum parameter sets is used is presented in the blue dotted line in Figure 4, while the red line stands for the improvement of the score over the entire year, when the optimum set of parameters is used replacing the default ones.

The monthly variability of the performance score with respect to the overall improvement (over the entire year) indicates that the model performance is sensitive to different weather patterns. This feature is pronounced during winter (and specifically for February) with the overall improvement reaching up to 12%. The effects of calibration methodology on yearly and monthly performance score, using finer model resolution, is now investigated at the framework of CALMO-MAX
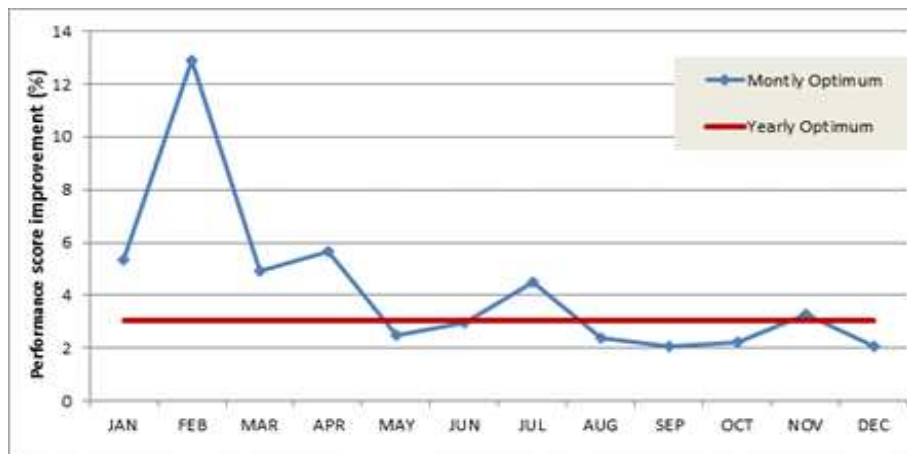
Figure 4: cycle of the performance score improvement using the optimum set for six parameters used in CALMO

4 Conclusions

The methodology used in CALMO and CALMO-MAX priority projects showed that the calibration of the unconfined model parameters using the MM is feasible. Model performance can be improved on monthly and yearly basis. However, the methodology remains computationally expensive. Towards this direction, it will be examined within CALMO-MAX whether the computational cost could be reduced by e.g. applying CALMO methodology on10-20 days set, representing most of the synoptic situation, instead of an entire year. Once the computational cost is reduced, the developed methodology could be used by each COSMO member, to define an optimal parameter set over the target area of interest, for re-calibration after major model changes (e.g. higher horizontal and / or vertical resolution), for an unbiased assessment of different modules (e.g.parameterization schemes), as well as for optimal perturbation of parameters when run in ensemble mode.

References

[1] Albergel, C., de Rosnay, P., G. Balsamo, G., Isaksen, L., Munoz-Sabater, J., 2012: Soil Moisture Analyses at ECMWF: Evaluation Using Global Ground-Based In Situ Observations. *J. Hydrometeor*, 13, **1442-1460**.

[2] Aligo, E. A., Gallus, W. A. and Segal, M., 2007: Summer Rainfall Forecast Spread in an Ensemble Initialized with Different Soil Moisture Analyses. *Wea. Forecasting*, 22, **299-314**.

Spatial verification techniques applied to high resolution models for an intense precipitation summer event in Greece

Presented in the: 14th International Conference on Meteorology,Climatology and Atmospheric Physics
October 15-17, 2018 Alexandroupolis, Greece

Boucouvala D.1*, Gofa F1., Samos I1

1 Hellenic National Meteorological Service, Hellinikon GR-16777, Athens, Greece

1 Introduction

Traditional precipitation verification metrics based on point-to-point comparison without providing information regarding spatial distribution are insufficient to evaluate precipitation forecasts, especially from high resolution mesoscale models. For example, when a small scale feature (also defined as an object) is correctly forecast but slightly displaced in time and space, the forecast will be penalized both for a miss and a false alarm (double penalty), especially for high resolution datasets (Ebert 2008).

Spatial verification methods that allow for some tolerance to reasonably small errors in space and time tend to resolve this problem (Cassola et al. 2015). The two main categories are: neighbourhood (or fuzzy) verification methods (Ebert 2008) based on a scale-dependent verification approach where the requirement for exact matches between forecasts and observations is relaxed and object oriented techniques which deal on with how well the forecast captures the overall structure of meteorological features by identifying and comparing precipitation features in the forecast and observations (Ebert and McBride 2000).

The aim of this study is to evaluate the relative model performance of the operational Numerical Weather Prediction (NWP) systems of the Hellenic National Meteorological Service (HNMS) (with different horizontal resolutions) for a rare summer precipitation event that affected almost the entire area of Greece by applying spatial verification methods.

2 Data and Methodology

An unusually strong precipitation event that occurred on the 16-17th of July 2017 was selected as a test case. The event, which was a combination of both dynamic and convective activity was accompanied by relatively low temperatures for the season and affected a large part of the country, causing hailstorms, flooding, property damage and unfortunately loss of human life. The event was preceded by a series of relatively warm days with 850hPa temperatures around 15-20°C. On 17/04 00UTC, a trough centered over Russia covering all of Eastern Europe moved southwards toward Greece, resulting in cold air masses (-15°C) at 500hPa (Fig 1a) moving slowly E-NE. The trough was accompanied by a low pressure system at the surface, which moved from west to east (Fig 1b). Initially, convective precipitation was observed over northern and western Greece which extended to the central and eastern parts of the country by the afternoon. This was accompanied by lightning (Fig 1c) and hail at several locations on the mainland.

2.1 Data

Spatial verification techniques require data defined continuously over a common spatial domain covering the area of interest. 3-hourly cumulative HSAF (EUMETSAT Satellite Application Facility on Support to

Operational Hydrology and Water Management) gridded observations and forecasts from: a) the global scale ECMWF (IFS) model with a horizontal resolution of 9km and b) the local model COSMO-GR (Gofa et al. 2008) with horizontal resolutions of 7, 4 and 1km were used.

The data were regridded (interpolated or extrapolated) to a common grid spacing of 0.06° (4km) in order to facilitate comparison. This grid spacing is also consistent with the spatial frequency of observations. Despite the smoothing effect associated with upscaling, the high resolution model configuration preserves details of the precipitation structure while also featuring larger magnitude departures from the observations at some locations. This can possibly be attributed to the coarser resolution of the initial observed precipitation field.

2.2 Methodology

Neighborhood verification (or fuzzy) techniques evaluate forecast performance using more elastic conditions regarding the exact spatio-temporal match between observed and forecast fields. It is based on the principle of expanding the area of comparison to include data points nearby ("neighbors"), employing a spatial window, or "neighborhood", surrounding the forecast and/or observed points. A relaxing filter can be applied to both fields, and the penalty for differences between modeled and observed values is relaxed. The properties of the relaxed fields (mean values, maximum values, number of grid points exceeding a threshold) can then be compared using traditional statistical methods.

The size of this window starts at the smallest possible scale (neighborhood of one grid box) and is gradually increased in order to provide insight into the scales at which the model has the most skill. The method shows how forecast skill varies with neighborhood size and can be used to determine the smallest neighborhood size that provides a sufficiently skillful forecast. A more detailed review of neighborhood approaches is available in Ebert (2008). There is a variety of methods that fall within this category, differentiated by their treatment of the points within each window, depending on the neighborhood method used. In order to determine if a forecast is "useful" or "good enough", decision models are applied such as: Upscaling, Minimum Coverage, Anywhere in the Window and Practically Perfect Hindcast. Traditional dichotomous scores are then calculated.

The Fractions Skill Score (FSS) is a decision model based on the comparison of frequency of forecast and observed events. In this study, the VAST (VERSUS Additional Statistical Techniques) software package, which was developed by the COSMO consortium and offers a number of neighborhood verification tools, was employed (Gofa at al. 2018). SAL (which stands for Structure, Amplitude and Location) is an object-based method developed by Wernli et al. (2008, 2009) to measure the quality of a forecast by identifying objects in both forecast and observed fields at a given time and provide information on object shape and location differences between the two fields. The score consists of three components which correspond to a global field measure of: Structure (S), Amplitude (A) and Location (L). The S parameter compares the volume of the
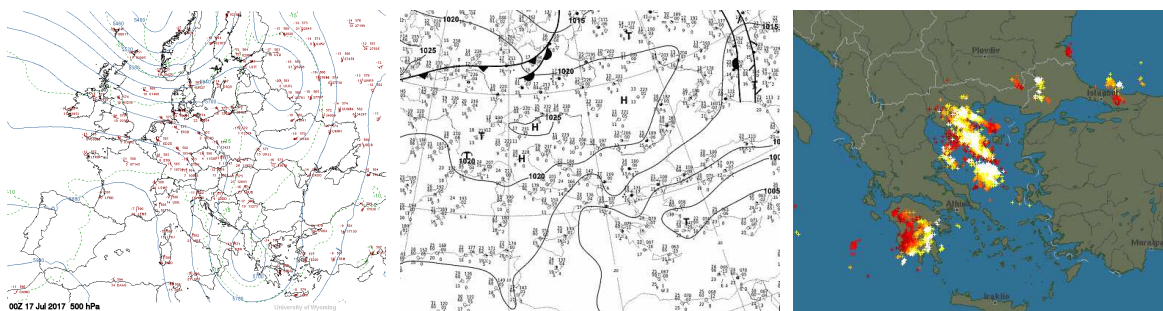

Figure 1: (left): 17/04 00UTC 500hPa (source: University of Wyoming), 1b. (center): Surface Analysis (source: UK Met Office), 1c. (right) Lightning activity at 10UTC (http://el.blitzortung.org) with dots indicating the location of lightning strikes where the color refers to the age of the strike (20 min intervals).

normalized precipitation objects of the two fields. Positive S values indicate that modeled precipitation objects are too large or too flat (more stratiform precipitation), while a negative value indicates that objects are too sharp and too small (more convective type precipitation).

The A parameter represents the normalized difference of the domain-averaged precipitation fields and is independent of structural features. Positive (Negative) values of the A parameter indicate overestimation (underestimation) of total domain precipitation. The range of the S and A parameters is [-2, 2]. The L component combines information about the distance of predicted and forecast mass centers (L1) and the normalized distance between the precipitation objects (L2). L ranges from 0 to 2. A perfect forecast is characterized by zero values for all three SAL components. The S and L (specifically the L2 component) parameters require the identification of objects in observed and forecast fields. An object is defined when it exceeds a fixed or statistically defined threshold value. Wernli et al. (2009). Consequently, if no features are found in either or both forecast and observed fields, the SAL values cannot be defined. The SAL parameters are calculated with a SpatialVx based software package (Gilleland 2017).
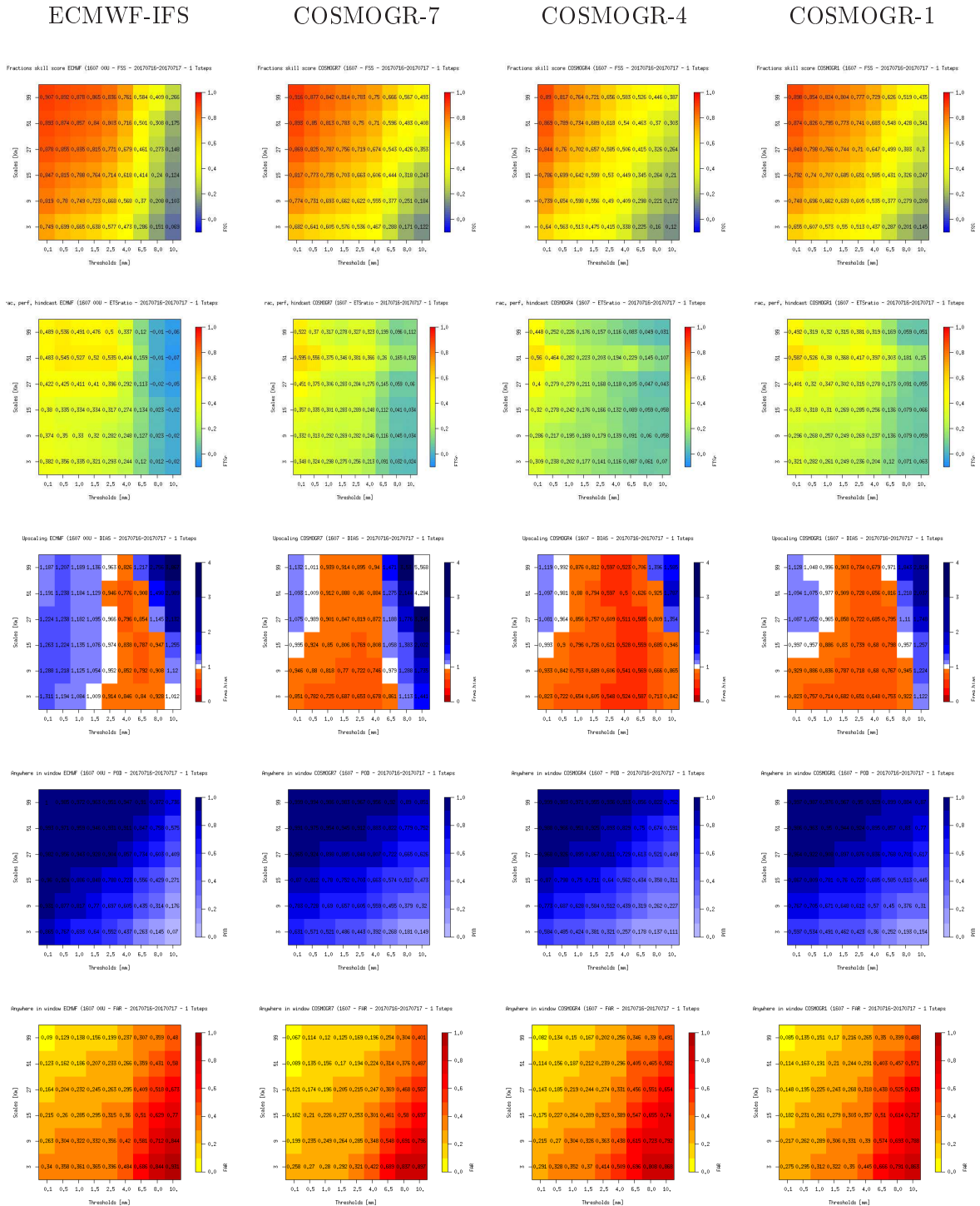
Figure 2: Neigborhood method plots for lead time 16/07 00 UTC derived for the various model setups. From top to bottom: FSS (Fractions Skill Score), ETS (Equitable Thread Score), Bias, POD (Probability of Detection) and FAR (False Alarm Rate)

3 Results

A selection of the evaluation plots for July 17th applying neighborhood methods to the various models and resolutions is presented in Fig 2. (time lead 1600 UTC). The scores are plotted as intensity-scale diagrams, where the intensity threshold and spatial scale averaging increase along the x and y axes respectively, and the color shade gives an indication of the value of the score (also plotted explicitly). By evaluating the color intensity (darkness), scales and thresholds at which a particular model system performs best, it is possible to evaluate model performance without focusing on the absolute value of each colored window. The forecast skill (as represented by the FSS score) does not differ significantly between models, but it does increase as window size ($<$15km) and precipitation thresholds ($<$3mm/3h) decrease. For high precipitation thresholds, on the other hand, forecast skill decreases.

ETS (Equitable Thread Score) index diagrams (Practically perfect Hindcast decision method) show that the forecast quality is better for window sizes $<$50km and thresholds 0.1-0.2mm. The indices for COSMO-GR1 and COSMO-GR4 are slightly better than those of ECMWF-IFS. However, significant differences appear in the Bias score (upscaling method) as ECMWF-IFS has the tendency to overestimate both the low thresholds (0.1-3mm) and high thresholds ($>$10mm/3h) while underestimating the remaining thresholds. The COSMO model generally overestimates rainfall for windows up to 27km for all thresholds, except for COSMO-GR7 which underestimates only the high thresholds. The POD (Probability of Detection) and FAR (False Alarm Rate) (calculated using the Anywhere in the Window method) show that ECMWF-IFS had more successful hits (dark red) but also more false alarms (dark blue). SAL parameters for the 24h accumulated precipitaion for July 17th are estimated for the 1600 UTC model run with different fixed thresholds (from lower to higher) (Fig 3).

The positive S parameter indicates that flatter objects (more stratiform precipitation) are calculated by the models for higher thresholds, while sharper objects (more convective) are produced at lower thresholds. COSMO7 predicts flatter objects versus sharper objects by COSMO1. The L parameter is constant and lower for ECMWF-IFS, while higher values are calculated for COSMO4. COSMO7 S values tend to be lower for higher precipitation thresholds. The A parameter, which is independent of objects and depends on the entire field, is positive, which means that for all models, especially for COSMO7, 24h precipitation is overestimated.
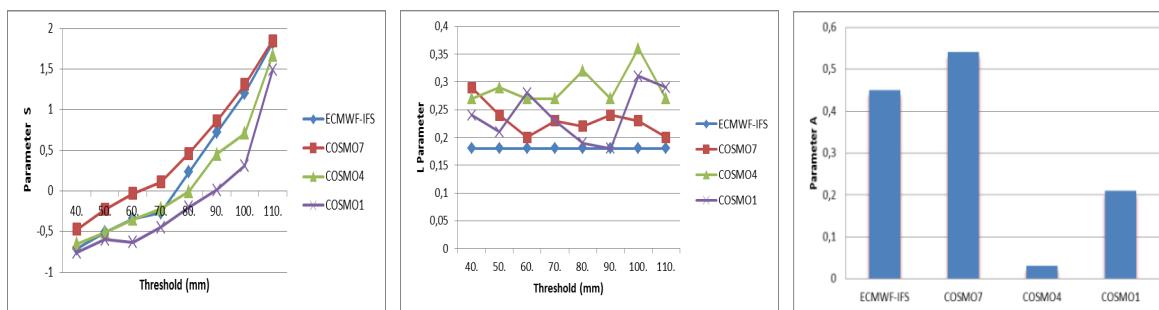

Figure 3: Left:S (Structure), Center: L(Location) parameters with threshold, Right: A (Amplitude) parameter for ECMWF-IFS, COSMOGR-7, COSMOGR-3 and COSMOGR-1.

4 Conclusions

The aim of this study was to compare the metrics of two spatial verification methods applied to the case of an intense summer precipitation event. Neighborhood verification results showed that for high rainfall rate thresholds and large spatial windows, the forecast skill and quality decreased for all models used in the study. Differences between the COSMO and ECMWF-IFS models at different scales and thresholds are mainly evident in Bias and ETS scores, with the latter model tending to overestimate precipitation for low thresholds and consequently producing more false alarms. Application of the SAL object-based method to 24h precipitation forecasts showed that finer resolution models led to prediction of sharper objects, that all models overestimate domain precipitation while location errors are more variable with threshold for finer resolution models. These results confirm that, when combined with traditional verification techniques, spatial verification methods enable more detailed and more complete assessment of model performance.

References

[1] Cassola F, Ferrari F, Mazzino A (2015) Numerical simulations of Mediterranean heavy precipitation events with the WRF model: analysis of the sensitivity to resolution and microphysics parameterization schemes. Atmos. Res. **164–165, 210–225**.

[2] Ebert E (2008) Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. Meteorol. Appl. 15, **51-64**.

[3] Ebert E, McBride JL (2000) Verification of precipitation in weather systems: Determination of systematic errors. J. Hydrol. 239, **179-202**.

[4] Gilleland E (2017) *R package Version 0.6-1.***https://cran.r-project.org /package =SpatialVx**

[5] Gofa F, Pytharoulis I, Andreadis T, Papageorgiou I, Fragkouli P, Louka P, Avgoustoglou E, Tyrli V (2008) Evaluation of the operational numerical weather forecasts of the Hellenic National Meteorological Service. Proc. 9th COMECAP Conference of Meteorology, Thessaloniki, Greece, **51-58**.

[6] Wernli H, Hofmann C, Zimmer M (2009) Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique. Wea. Forecasting. 24, **1472–1484**.

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

A user oriented verification methodology for wind forecast

MARIA STEFANIA TESINI

ARPAE, HydroMeteoClimateService of Emilia-Romagna region, Bologna, Italy

1 Introduction

The 10-m wind is a weather parameter characterized by strong dependence on orographic and topographic details and high temporal variability. Therefore the verification of wind forecast requires a methodology taking into account these features.

On top, any verification method should be tailored for the specific purpose defined by the user of that forecast, being the developer of the model, the forecaster in the operational room or the stakeholder for a practical application.

One of the main uses of wind forecast at Arpae-SIMC is to issue warnings when wind speed exceeds some threshold. For example, strong easterly winds can determine the possible occurrence of sea storms over the Italian coast of the Adriatic Sea (see figure 1), but also less intense winds can cause problems to the tourist's activity on the beach.

Figure 1: Wind observations in the north Adriatic Sea coast

Verification should therefore address several user-related aspects, in particular it should quantify:

- the ability of the model to predict wind speed above (or below) critical thresholds, including false alarms or misses,
- the forecast skill in terms of wind direction,
- the dependence of forecast error on wind direction
- the interplay between wind-speed and wind-direction inaccuracies

In addition, another very important aspect to consider is the communication of the results to the end user, which should be as clear and concise as possible.

2 The Performance-Rose

In order to meet the needs of our end-user we have studied a methodology of verification that seeks to take into account wind speed and direction at the same time, but that it is also effective in communicating results. For this purpose we developed a novel summary-plot of the scores derived from the contingency table, denominated "Performance Rose". In a wind-rose plot, the observed and forecast wind frequency is represented subdivided into the usual 8 sectors. In additions to this, each spoke is broken down into color-coded bands that show information about errors in wind speed and direction. Moreover usual scores such as the Probability of Detection, the Threat Score and the Success Ratio are plotted for each sectors (i.e. for each direction) as symbols on the appropriate scale (from 0 to 1 for all of them) in the radial axes.

First of all, for each station 10m-wind observations (hourly or 3/6-hourly) and corresponding data predicted by model are categorized in octants for wind direction (N, NE, E, SE, S, SW, W, NW) and in classes for wind speed, according to table 1.

LIGHT	ws < 10 Knots	ws < 5.1 m/s
LIGHT-MODERATE	$10 \leq$ ws < 20 Knots	$5.1 \leq$ ws < 10.3 m/s
MODERATE	$20 \leq$ ws < 30 Knots	$10.3 \leq$ ws < 15.4 m/s
STRONG	ws ≥ 30 Knots	ws ≥ 15.4 m/s

Table 1: Wind speed classes

For each specific wind speed class (e.g. "Light-Moderate") a "Performance-Rose" diagram is produced, as showed in figure 2.
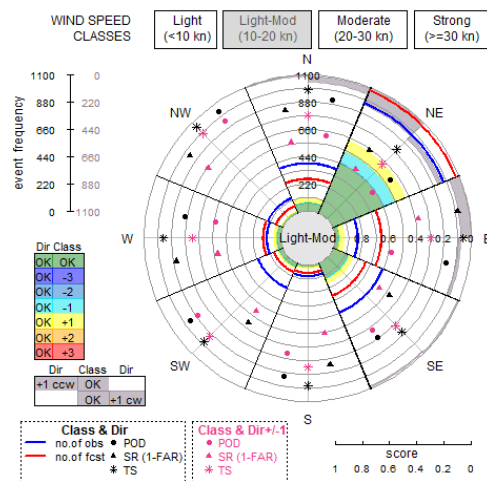

Figure 2: The Performance-Rose diagram

Like in a usual wind rose, the frequency of winds of the specific wind class blowing from particular directions over a specified period is represented:

- **Blue line** is the observed frequency of the specific speed-class
- **Red line** is the forecast frequency of the specific speed-class

The number of events can be read on the radial scale (frequency axis), increasing outwards from the center.

Using sectors of different colours we described how model predicts the reference speed class in each direction, *being the direction correct*. For example:

Green means speed class is correctly forecast

Cyan means speed is underestimated of 1 class (e.g. the forecast is "Moderate" but the observation is "Strong")

Yellow means speed is overestimated of 1 class (e.g. the forecast is "Moderate" but the observation is "Light-Moderate")

The number of events of each sector can be deduced using the radial scale of the frequency axis.
In addition, the gray half-sectors represent the number of forecast in each direction that are "nearly" correct in direction, *being the intensity correct*:

Half sector on the left means forecast is shifted of 1 octant clock-wise (e.g. if the fcst is NE, the obs in N)

Half sector on the right means forecast is shifted counterclock-wise (e.g. if the fcst is NE, obs is E)

In this particular case, the number of events can be deduced using the reverse radial scale of the frequency axis (starting from the outermost circle).

A quantitative assessment of the goodness of the forecasts is made using some scores evaluated form the elements in a contingency table (see figure 3) that shows the frequency of "yes" and "no" forecasts and occurrences:

$$POD = \frac{hits}{hits + misses}$$
(Probability of Detection)

$$TS = \frac{hits}{hits + misses + false\ alarms}$$
(Threat Score)

$$SR = (1 - FAR) = \frac{hits}{hits + false\ alarms}$$
(Success Ratio)

$$FAR = \frac{false\ alarms}{hits + false\ alarms}$$
(False Alarm Ratio)

$$BIAS = \frac{hits + false\ alarms}{hits + misses}$$
(Bias Score)

Event forecast	Event observed		
	YES	NO	Marginal Totals
YES	Hit	False Alarm	Fcs yes
NO	Miss	Correct non-event	Fcs no
Marginal Totals	Obs yes	Obs no	Sum total

Figure 3: Contingency table for dichotomous (yes/no) forecasts

For our purposes, we have defined two types of "yes"/"no" events and for each of them a specific contingency table has been constructed:

1. **"Class & Direction"**: the "yes" event is defined by speed class and direction correctly forecast at the same time, while other entries of the contingency table are defined as in the table in figure 4

2. **Class & Direction ±1:** the "yes" event is defined by speed class correctly forecast, but direction is considered correct even if differs by one octant. According to the table in figure 5 we defined only "false alarm" and "miss" events, since correct negatives have been removed from consideration.

   This definition of "extended direction", even if not completely proper, is meant to address the user's request to know whether the source of error depends on either wind speed or wind direction, if compared with scores based on the exact correspondence of direction and speed.

Figure 4: Contingency table for event defined as "speed class and
direction correctly forecast at the same time"

Figure 5: Contingency table for event defined as "speed class correctly
forecast, but direction is correct even if differs by one octant"

In the Performance Rose, the scores related to the two event definitions are plotted as symbols of different colors (black for the event "Class & Direction" and magenta for event "Class & Dir ±1").
Their value can be read on the radial scale (score axis) and, as in a archery target, the perfect score 1 is represented in the innermost ring.
The Frequency Bias is not explicitly calculated but it can be deduced for the "Class & Direction" event by the relative position of observed frequency line (in blue) and forecast frequency line (in red):

– Red line outer means overestimation of the number of events

– Blue line outer means underestimation of the number of events

3 Examples of application of the Performance-Rose diagrams

The Performance-Rose is designed primarily to help forecasters understand the behaviour of models, particularly on some coastal stations, where for geographical reasons it is essential to identify errors in the forecast of wind intensity as a function of direction.
For example, in figure 6 are plotted the Performance-Rose diagrams for the verification of 10m wind predicted by COSMO-I7 00 UTC run for the station "Chioggia" located in the north Adriatic sea near Venice.
The statistics refer to 1 year (JAN-DEC 2016) of hourly data from 1 to 24 hours of forecast (DAY 1) and corresponding observations. Looking to the plots, following the possible interests of forecasters as final users, some considerations can be done:

When the wind is predicted in the "Light" class the errors on direction are significant as suggested by the dimension of the gray sectors and the better scores for the "Class & Dir ±1" event with respect to "Class & Dir". The errors in direction decrease as the wind is predicted in higher classes (very small differences in the scores for "Moderate" or "Strong" winds).

Underestimation of the intensity ,with correct direction predicted, is more evident for "Light" and "Light-moderate" classes (see cyan sectors). In case of "Moderate" winds predicted the number of cases of underestimation is very small, while the number of overestimated events is significant (see yellow sectors).
This information is important for the forecasters as they can be confident about the low risk of missing critical events.

Unfortunately the performance-rose relative to "Strong" wind shows that the scores relative to this type of event are very low. In addition to cases of overestimations, the most frequent error is the complete missing of the event (predicted in lower wind classes with very different direction and therefore not visible in the Performance-Rose diagram).
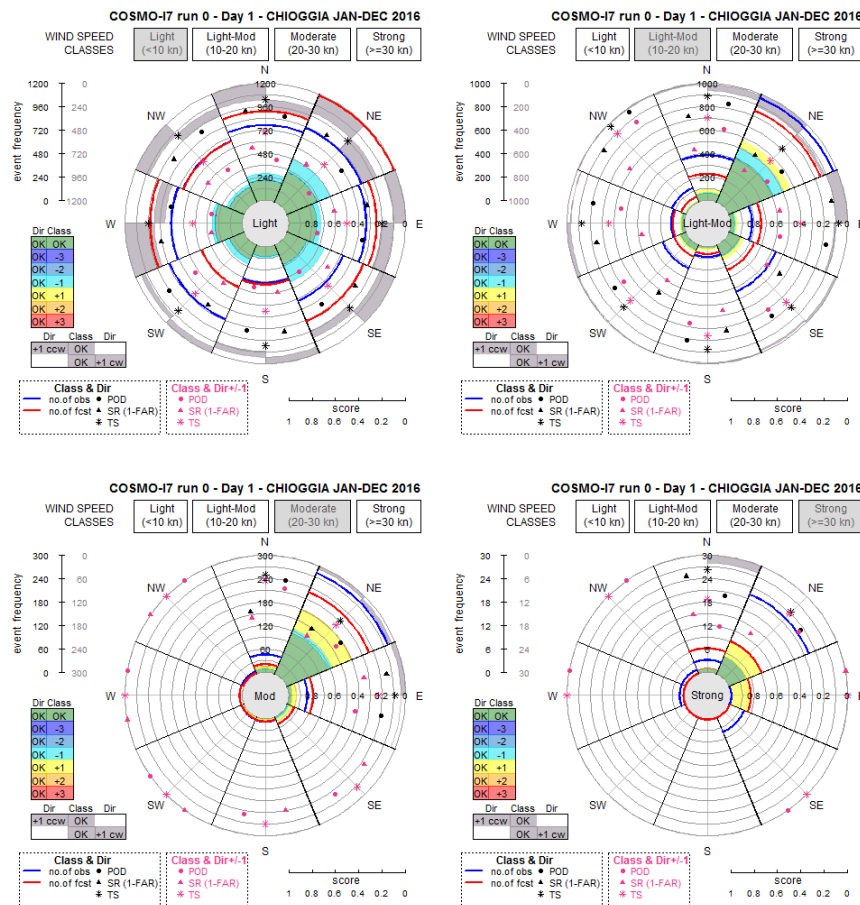
Figure 6: Performance-Roses for the verification of 10-m wind hourly data predicted by COSMO-I7 00 UTC run for the station "Chioggia" in the period January-December 2016. Each single plot refers to a specific speed class: "Light" (top-left), "Light-Moderate" (top-right), "Moderate" (bottom-left), "Strong" (bottom-righ).

If instead we consider the case in which models developers are the end user of our verification, it can be interesting to compare if some errors in the wind forecast of a particular model are also found in other models, considering the same station and period. For this purpose, in the framework of WG5 Common plots activities during the year 2018 the production and analysis of Performance-Rose diagrams for several models was started.

Data of 10m wind forecast from COSMO-5M (Arpae-Italy), COSMO-PL (IMGW), COSMO-GR4 (HNMS), ICON-EU (DWD),COSMO-DE/D2 (DWD) were used to produce Performance-Rose diagrams for four different periods (JJA2017,SON2017,DJF2017-18,MAM2018) on a set of selected stations belonging to Common Area 1 (see figure 7).

Since the goal of the Performance-Rose diagram is to provide the end user with effective feedback on the model's forecasts, trying to answer the question of a specific user, it was decided to compare the various models by identifying some targeted issues, depending on the characteristics of the wind field on individual station.

For example, wind verification of both COSMO-5M and COSMO-I2 (2.8 Km resolution) performed in Emilia-Romagna region (Italy) pointed out a a general underestimation of the forecast intensity in mountain stations. To see if this behaviour was also common to other models, the station "Svratouch", located at 740 meter a.s.l., was chosen and Performace-Roses representing the performance of the first 24 hours of forecast of the four different models during the period June 2017 - May 2018 were produced.
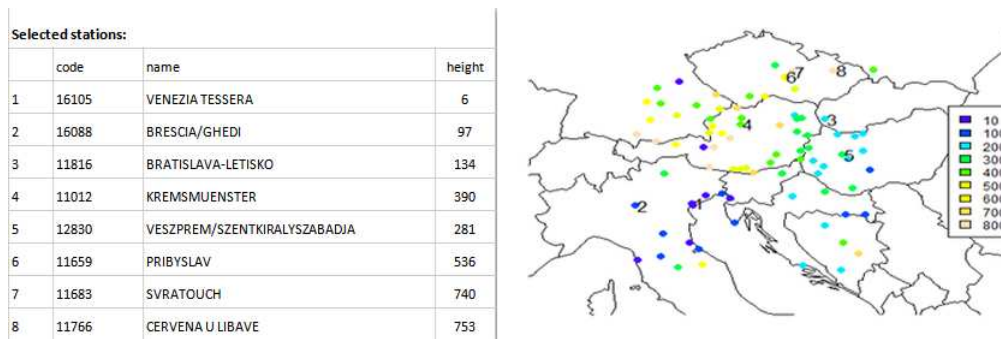
Selected stations:			
	code	name	height
1	16105	VENEZIA TESSERA	6
2	16088	BRESCIA/GHEDI	97
3	11816	BRATISLAVA-LETISKO	134
4	11012	KREMSMUENSTER	390
5	12830	VESZPREM/SZENTKIRALYSZABADJA	281
6	11659	PRIBYSLAV	536
7	11683	SVRATOUCH	740
8	11766	CERVENA U LIBAVE	753

Figure 7: Selected stations in Common Area 1. The colors of the symbols represent the height of the stations in meter a.s.l.

In figure 8 the Peformance-Roses of COSMO-5M for the four different wind-speed class are represented, while in figures 9-11 only the two lower wind-speed classes of COSMO-PL , COSMO-GR and ICON-EU are plotted, since there are no forecast in the "Moderate" and "Strong" classes.

With reference to figures 8-11 some considerations can be done: In general, cyan sectors indicate that the forecast is correct only in direction while the intensity is underestimated , i.e. the observed wind intensity is one speed-class higher.

The COSMO models considered exhibit this type of behaviour in the two lover speed-class and particularly in the direction "South" for the "Light-Moderate" class, indicating that the corresponding wind observations were in the "Moderate" class. In fact, if you look to the plot referring to the "Moderate" class of COSMO-5M (see figure 8) the number of observations (blue line) is higher than the number of the forecast (red line). For other models forecast events are zero (not shown). ICON-EU graphs show less cyan sectors, i.e. less underestimation, but also less green sectors and therefore less correct forecasts in both intensity and direction. The presence of grey sectors indicates that the intensities have been correctly predicted while it is the direction that is missed, as for example can be seen from the underestimation of the number of events in the "South" direction that seems partly compensated by the occurrences in the "South-East" direction (grey sector).
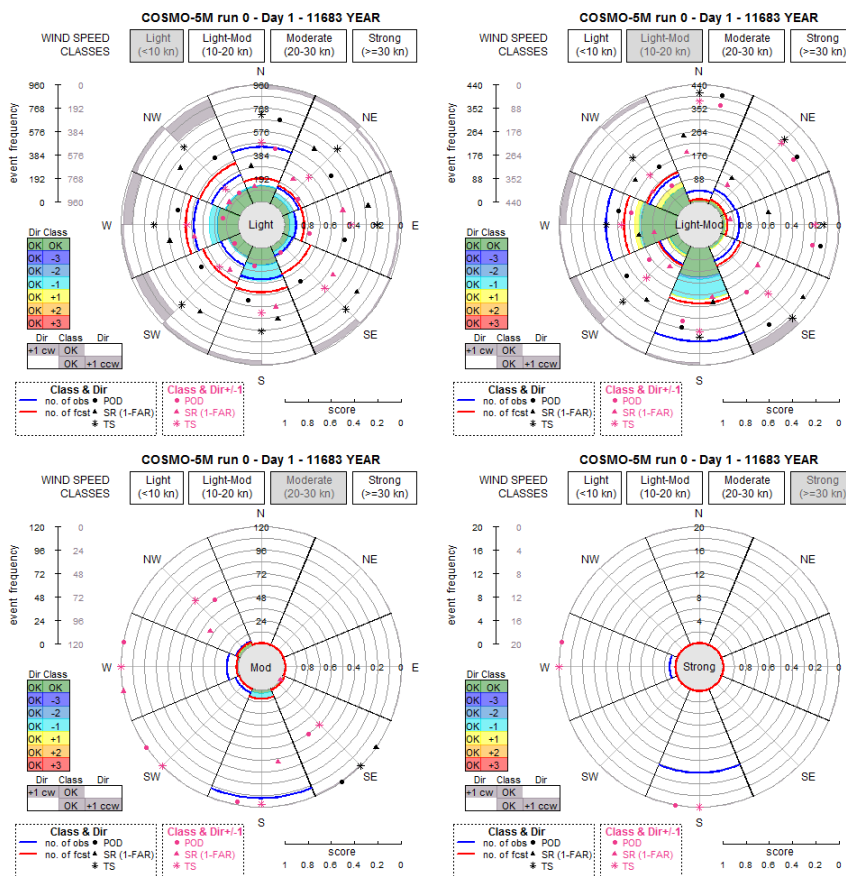
Figure 8: Performance-Roses for the verification of 10-m wind hourly data predicted by COSMO-5M 00 UTC run for the station "Svratouch" in the period June 2017-May 2018. Each single plot refers to a specific speed class: "Light" (top-left), "Light-Moderate" (top-right), "Moderate" (bottom-left), "Strong" (bottom-righ).
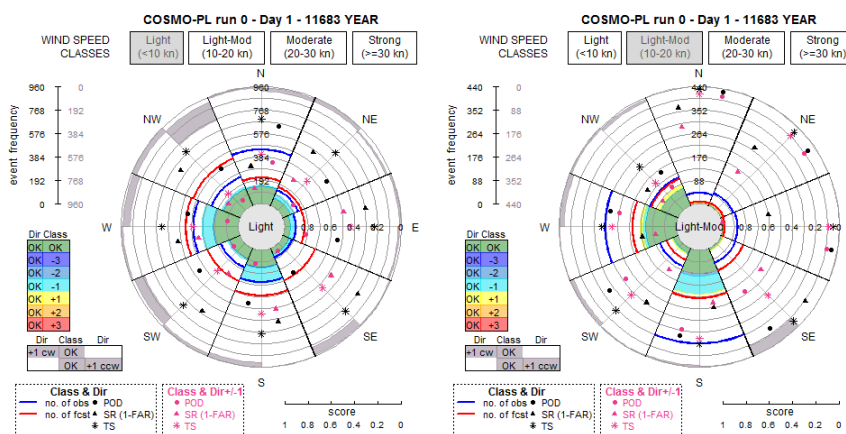

Figure 9: Performance-Roses for the verification of 10-m wind hourly data predicted by COSMO-PL 00 UTC run for the station "Svratouch" in the period June 2017-May 2018. Each single plot refers to a specific speed class: "Light" (top-left), "Light-Moderate" (top-right).
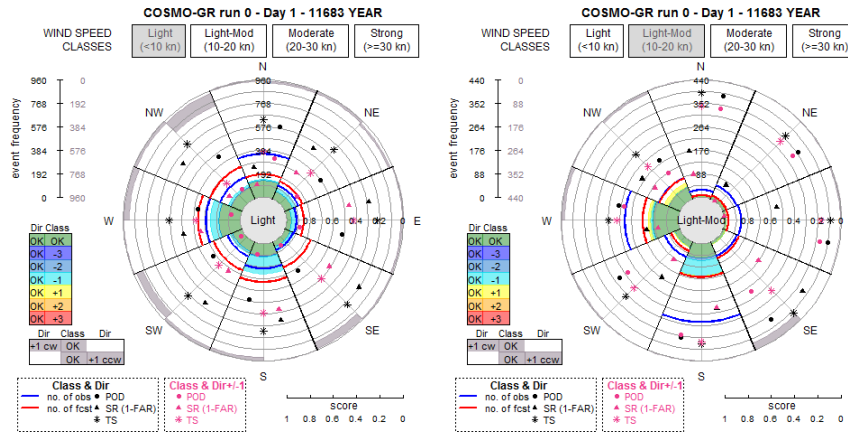
Figure 10: Performance-Roses for the verification of 10-m wind hourly data predicted by COSMO-GR 00 UTC run for the station "Svratouch" in the period June 2017-May 2018. Each single plot refers to a specific speed class: "Light" (top-left), "Light-Moderate" (top-right).
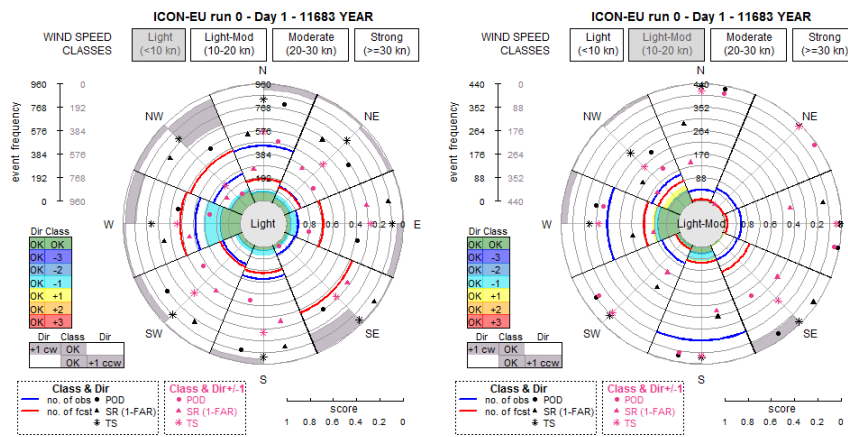

Figure 11: Performance-Roses for the verification of 10-m wind hourly data predicted by ICON-EU 00 UTC run for the station "Svratouch" in the period June 2017-May 2018. Each single plot refers to a specific speed class: "Light" (top-left), "Light-Moderate" (top-right).

5 Conclusion

One of the most tricky aspects of verification is to provide end-users with an effective feedback on model forecast, both in terms of contents and communication. The idea of the "Performance Rose" diagram addresses precisely this issue, trying to answer the questions of a specific users.

Since a lot of information is summarized in the Performance-Rose diagram, it is necessary to focus from time to time on specific aspects, depending on the user's needs.

Performances of COSMO-based ensemble systems for cases of High-Impact Weather over Italy

G. Pincini (1), A. Montani (2), T. Paccagnella (1), M.S. Tesini (1), C. Marsigli (3)

(1) Arpae-SIMC, (2) ECMWF, (3) DWD

Introduction

The prediction of weather events related to strong winds, heavy rain and snowfall is still nowadays a serious challenge, especially when high spatio-temporal details are required. Despite Numerical Weather Prediction (NWP) modelling has made great progress in recent decades, thanks to the increases in model resolution, better understanding of atmospheric dynamical processes and advantages in data assimilation techniques, the above-mentioned atmospheric events, usually referred to as "High-Impact Weather" (HIW), can have horizontal dimension too small to be explicitly resolved. HIWs provide the most dramatic examples of how the atmospheric affects people daily lives, since they may cause both human and economic costs. Therefore, there is a need of better ways to predict this type of phenomena, also accounting for their inherent degree of non-predictability.

The ensemble forecasting provide a representation of model uncertainty, due to the imperfect knowledge of atmospheric initial conditions and the approximate model formulation. Instead of running just one forecast with an unknown error, an ensemble of slightly different forecasts are run, in order to integrate the deterministic forecast with an estimate of the "forecast of forecast skill". Probabilistic forecasts provide a more complete, reliable and accurate view of what might happen in the future, ideally providing information on the relative frequency of an event occuring. Therefore, they bring definite benefits for decision-makers. The estimation of uncertainty is even more crucial when local effects come into play and a high spatio-temporal detail is required as in the case of precipitation, where NWP limitations become more evident.

The aim of this work is to assess the added value of the enhanced horizontal resolution in the probabilistic prediction of surface fields. In particular, the performance of three different ensemble prediction systems were compared: ECMWF ENS (51 members, 18 km horizontal resolution), COSMO-LEPS (16 members in 2016, 20 members now; 7 km horizontal resolution) and COSMO-2I-EPS (10 members in 2016, 20 members now; 2.2 km horizontal resolution). While the first two ensemble systems are operational, COSMO-2I-EPS is still in a pre-operational phase. The intercomparison window covers two limited periods, which range from 20 to 27 June 2016 and from 15 October to 15 November 2018. As for the surface variables, 2-metre temperature and precipitation are verified against the non-conventional station network provided by the National Civil Protection Department.

The ensemble spread and the root mean square error of 2-metre temperature are computed, while Ranked Probability Score and Percentage of Outliers are considered for precipitation. The best scores are mainly obtained by the COSMO-based ensemble systems with higher horizontal resolution and lower ensemble size; in particular COSMO-2I-EPS often achieves the most satisfactory performances. Although the results are based over two relative short periods due to limited data availability and further investigations is needed, the added value of high resolution in mesoscale ensemble systems seems to play a crucial role in the probabilistic prediction of atmospheric fieds at all levels. In particular, the more detailed description of mesoscale and orographic-related processes in COSMO-ensembles provides an added value for the prediction of localised High-Impact Weather events.

Global and limited-area ensemble prediction systems and description of the experiments

A summary of the technical characteristics of the three ensembles used in the verification is shown in the table (Fig. 1).


MAIN TECHNICAL FEATURE / ENSEMBLE SYSTEM	ECMWF ENS (European Centre for Medium-range Weather Forecast ENSemble prediction system)	COSMO-LEPS (Consortium for Small-scale Modelling Limited-area Ensemble Prediction Systems)	COSMO-2I-EPS
Integration domain			
Horizontal resolution (km)	18	7	2,2
Vertical resolution (Model level)	91	40	65
Forecast range (hours)	240	132	48
Type of model	Hydrostatic model	Non-Hydrostatic model	Non-Hydrostatic model
Type of convection	Parameterized convection	Parameterized convection	Explicit convection
Ensemble size	51	16	10
Starting times (UTC)	00, 06, 12, 18	00, 12	00

Figure 1: The table shows the technical characteristics of ECMWF ENS, COSMO-LEPS and COSMO-2I-EPS in 2016; now COSMO-LEPS has 20 members as well as COSMO-2I-EPS

The intercomparison between the three ensemble systems is performed starting at 00 UTC and with a forecast range of 48 hours, because COSMO-2I-EPS runs once a day at 00 UTC and the forecast stops on the second day. The verification domain was selected in such a way as to include the entire Italian territory, more precisely the domain having the following geographic coordinate as borders (Fig. 2)

- latitude: 35^oN - 48^oN
- longitude: 6^oE - 19^oE

The station networks, used in the evaluation procedure, are:

- *the Northern-Italy non-GTS* [3] *(local) network*: it refers to about 1000 stations, over most Northern Italy and shared by the regional weather services operating in the area. These stations provide hourly data;
- *network from National Civil Protection Department (DPCN-Dipartimento Protezione Civile Nazionale)*: this network is composed of about 5524 stations over the national territory. Also these stations provide hourly data.

---

[3]Global Telecommunications Systems

Figure 2: The domain, centered over Italy, considered for the verification of the three ensemble systems. The points are the 5524 stations of National Civil Protecion Department used for the verification of precipitation.

These station networks were used for the verification of 2-metre temperature and precipitation respectively. DPCN stations have been subdivided, in three groups depending on the location altitude. For the subdivision it was decided to adopt the WMO (World Meteorologiacal Organization) directives on the subject, as follows:

- *lowland station* (under 200 m of altitude) 2311 DPCN observatories belong to this category;
- *hill stations* (between 200 m and 599 m of altitude) 1690 observatories belong to this category;
- *mountain stations* (above 600 m of altitude) 1523 observatories belong to this category.

The evaluation of the performance of the model consists in the comparison of gridded model output against point observations. A number of statistical scores evaluate different aspects of model performance while the forecast "error" is simply defined as the difference between the forecast value and the observation. In a "standard" deterministic approach, the uncertainty associated with the forecast value is not estimated. An EPS aims at quantifing this uncertainty using a set of perturbed Initial Conditions (ICs) and/or perturbed model formulations. Verification methods applied to ensemble forecasts have two main objectives:

- to assess the characteristics of the ensemble distribution;
- to verify the probability forecast.

Since all perturbed ICs should be equally possible be true and all perturbed physics or varying physics schemes or alternative models be equally plausible, the performance of any ensemble member should, in principle, be equivalent to that of another member on average. If this is not the case, that is indicative of problems with the choice of ensembling the technique employed. For example, either the IC perturbations are too large or alternative models, physics schemes or perturbations are not equally plausible. In the verification the evaluation method of the **nearest grid point** will be used: since observations seldom occur at the precise locations represented by the grid points of one particular model, it is necessary to compare the forecast values in the grid points with those of the nearest observations (ECMWF Forecast User Guide). In the experimental verification of the three ensemble systems will be used the following scores:

The Root Mean Square Error (RMSE) provides the square root of the average square error of the forecasts, which has the same units as the forecasts and observations. Here, the forecast corresponds to the ensemble mean value and an 'error' represents the difference between the ensemble mean $\overline{Y}$ and the observation x. The equation for the RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{Y_i})^2}$$

RMSE of the ensemble mean measure the distance beetween forecasts and analyses (or observations). **The ensemble spread (SPRD)** is calculated by measuring the deviation of ensemble forecasts from their mean [11]. Usually, SPRD is defined as:

$$SPRD = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(\overline{f} - f(n))^2}$$

Where $\overline{f} = \frac{1}{N}\sum_{n=1}^{N}f(n)$ is for the ensemble mean and f is for the ensemble forecast. In general, an ideal ensemble forecast will be expected to have the same size of ensemble spread as their RMSE at the same lead time in order to represent full forecast uncertainty [11] [2]; but most of the ensemble systems are underdispersed (lower spread) for longer lead times due to an imperfect model system (or physical parameterizations) and other factors. Anyway over a large number of ensemble forecasts, the statistical properties of the true value X_{TRUE} of any quantity X are identical to the statistical properties of a member X_j of the ensemble; in particular:

$$\overbrace{|X_j - X_{MEAN}|^2}^{ensemble\,variance} = \overbrace{|X_{TRUE} - X_{MEAN}|^2}^{mean\,squared\,error}$$

where X_{MEAN} is the ensemble mean. The time-mean ensemble spread around the mean equals the time-mean RMSE of the ensemble mean [3].

The Ranked Probability Score (RPS) is an extension of the RMSE to the probabilistic world and to the multi-category events; it ranges between 0 and 1.

$$RPS = \frac{1}{J-1} \sum_{m=1}^{J} [(\sum_{j=1}^{m} f_i) - (\sum_{j=1}^{m} o_j)]^2$$

where

- J is the number of forecast categories
- $o_j = 1$ if the event occurs in category j, $o_j = 0$ if the event does not occur in category j
- f_j is the probability of occurrence in category j

This score is used to assess multi-category forecast, where J is the number of forecast categories (for example, rainfall bins). The RPS penalizes forecasts less severely when their probabilities are close to the true outcome and more severely when their probabilities are further from the actual outcome. The lower the RPS, the better the ensemble system.

The Percentage of Outliers of a probabilistic forecast system is defined as the probability of the analysis (or observation) lying outside the forecast range [1]. Therefore this can be seen as the percentage of times the "truth" falls out of the range spanned by the forecast values. Here, it is computed as the fraction of points of the domain where the observed value lies outside the minimum or maximum forecast value.

Performance of the ensemble systems

First verification period: from 20^t to 29th June 2016

To begin the performance of the three ensemble systems is verified against the two-metre temperature. As already mentioned before, for this verification it was decided to consider the observational dataset coming from the regional networks of the weather services on Central-Northern Italy. In this way, data coming from only one part of the Peninsula were considered. Infact, the temperature data from the national civil protection network could have been used, but these data are from time to time of low-quality in Central and Southern Italy and their use would have provided wrong evaluation on the model skill. The period under investigation is from 20th June 2016 at 00 UTC to 29th June 2016 at 00 UTC, infact, although the last runs examined are those at 00 UTC on 27th June 2016, a 48-hour forecast range must always be considered. The performance of the three ensemble systems is evaluated by calculating the spread and the RMSE of the ensemble, the verification method used is the nearest grid point. The table 1 summarizes the characteristics of the verification.

Verification features

variable:	2-metre temperature;
period:	from 20/06/2016 00UTC to 29/06/2016 00UTC (9 days);
region:	Central-Northern Italy;
method:	nearest grid point;
obs:	non-GTS local fiduciary network, no obs error;
fcst ranges:	0-48h (verification every 6h);
systems:	ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS;
scores:	spread, RMSE;

Table 1: 2-metre temperature verification features

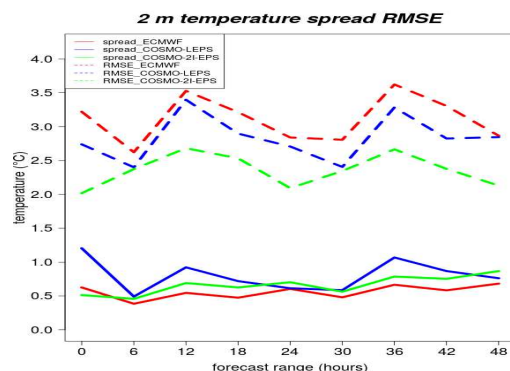The results are reported in (Fig. 3) and can be summarised as follows:

Figure 3: The figure shows the spreads (continuous lines) and the RMSE (dotted lines) values obtained for the 48 hours of the forecast range every 6 hours. The ECMWF EPS scores appear in red, COSMO-LEPS in blue and COSMO-2I-EPS in green. The foreacst range (in hours) is shown in the abscissa, in the ordinate the value of spread and RMSE (in ^oC). All details are indicated in the legend at the top left.

- the spread values are similar for all the three ensemble systems;
- the spread values are smaller with respect to the RMSE ones, showing a tendency of all ensembles to be underdispersive;
- with the exception of the shortest time range, COSMO-based models always show slightly higher (and therefore better) spread values than ECMWF EPS;
- RMSE values show a marked diurnal cycle, with maxima during the central hours of the day and the minimums in the night. This daytime cycle is very pronounced for ECMWF EPS and for COSMO-LEPS, less for COSMO-2I-EPS;
- the RMSE of COSMO-2I-EPS is the lowest of the three ensemble systems on the entire forecast range.

Therefore, from this 2-metre temperature verification, COSMO-based models get excellent results, especially COSMO-2I-EPS.

The performance of ECMWF ENS, COSMO-LEPS, COSMO-2I-EPS is verified also against the 6-hourly precipitation. For this verification work it was decided to use the precipitation data recorded by the rain gauges of National Civil Protection Department network. In this way, the results obtained are representative of what happened on the entire national territory between the 20th June 2016 at 00 UTC and the 29th June 2016 at 00 UTC. The method of the nearest grid point was used for the calculation of Ranked Probability Score and percentage of outliers. In table 2 are reported all the details of the verification.

In the Fig.4, the results obtained for the RPS can be consulted.

Considering all DPCN staions, regardless of the altitude (top left graph), it is worth pointing out:

- the diurnal cycle of precipitation is very marked. Since it is almost exclusively afternoon convective precipitation, the highest, and therefore the worst, RPS are just in the afternoon time slots: forecast range 12-18 h, 36-42 h;
- however, the RPS of COSMO-2I-EPS, and generally the RPS of COSMO ensemble systems, is better than ECMWF ENS one over the whole forecast range.

In this case the RPS points out to the added value of COSMO-2I-EPS.

The station of DPCN has been subdivided, according to their altitude, in three groups: plain, hill and mountain. Therefore, the purpose of this further study is to evaluate RPS variations according to the station altitude and see how this affects the results. Looking at the plots it can be concluded that:

Verification features

variable:	6-hourly total precipitation ;
period:	from 20/06/2016 00UTC to 29/06/2016 00UTC (9 days);
region:	Italy;
method:	nearest grid point;
obs:	DPCN network, no obs error;
fcst ranges:	0-48h (verification every 6h);
systems:	ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS;
scores:	RPS, outliers;
thresholds:	1mm, 5mm, 10mm, 15mm, 25mm, 50mm in 6 hours

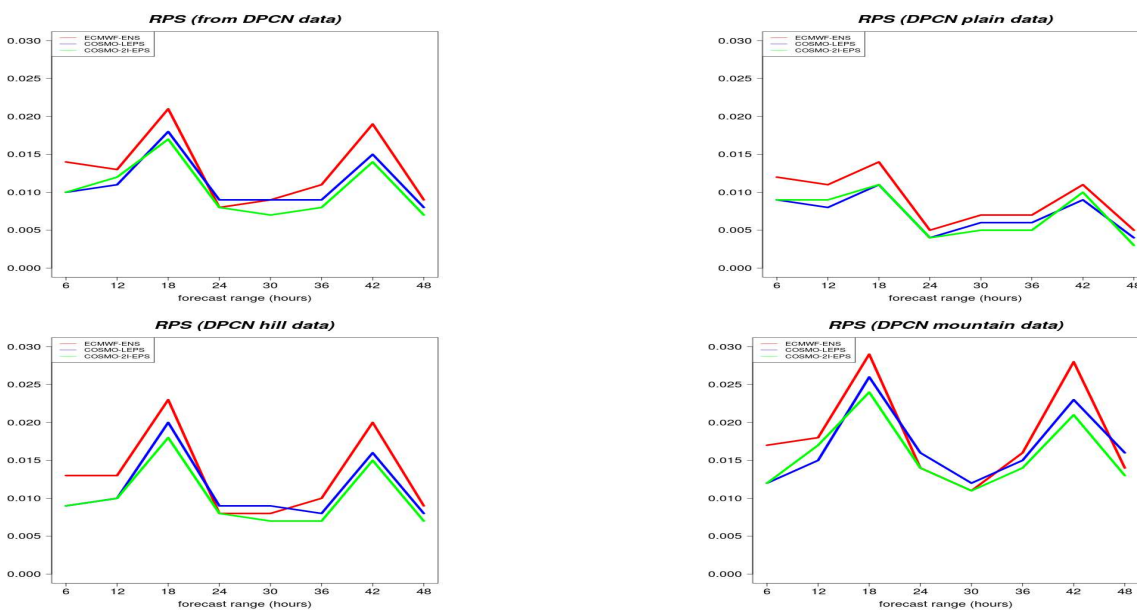Table 2: 6-hourly total precipitation verification features

Figure 4: The figure shows the RPS for four different observational dataset, indicated in the caption under each image. The ensemble systems are ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green. The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the dimensionless values of the RPS are marked in the ordinate.

- the RPS values obtained for the lowland stations are lower (therefore better) than those obtained for hill and mountain ones, in particular the results of mountain stations are the highest;

- in most cases, regardless of altitude, the RPS obtained for COSMO-2I-EPS is always lower (therefore better) than for COSMO-LEPS and ECMWF ENS;

- in the plain stations (top-right panel), there is a good gap beetween COSMO-2I-EPS and ECMWF ENS in the first day of forecast range. For the other stations this gap extends no longer than the first 18 hours, then the RPS tend to be similar for the three ensembles, except for the precipitation
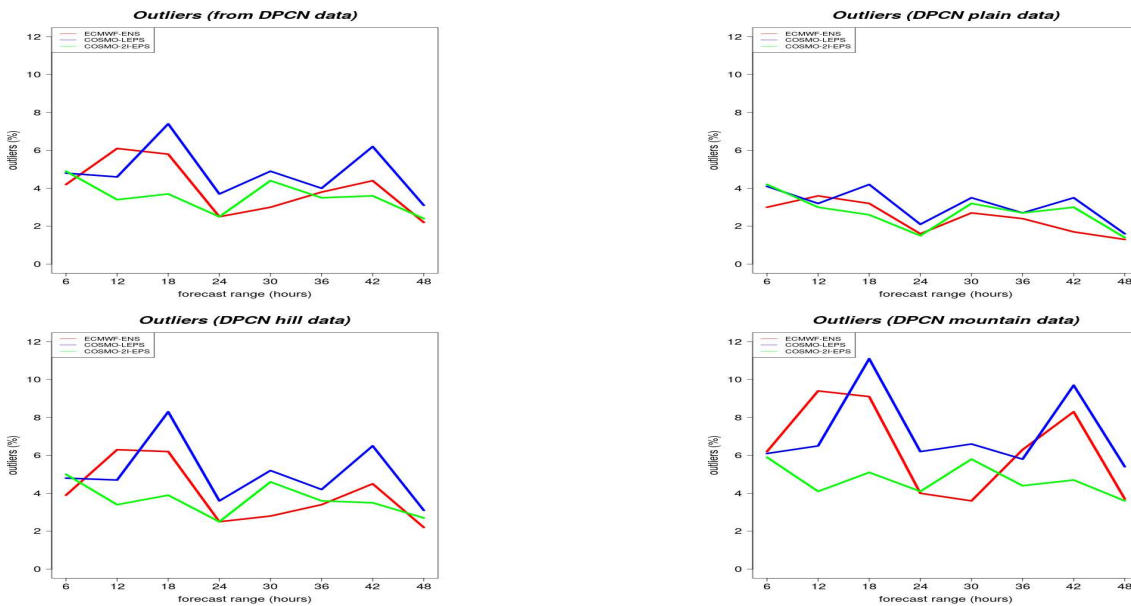
Figure 5: The figure shows the percentage of outliers for four different observational dataset, indicated in the caption under each image. The ensemble systems are ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green. The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the percentage of outliers is marked in the ordinate.

cumulated beetween the 36th and the 42nd hour of the forecast range;

- in the graph for hill and mountain stations (bottom left and bottom right panel respectively), the RPS follows a very strong daytime cycle, that is definitely less visible on the plain: this is could be due to the pluviometric regime of those days, with rainfall concentrated almost always in the afternoon hours and on the internal areas of hills and mountains.

So all the observational networks, built on altitude, confirm that the RPS of COSMO-based ensembles, but in particular COSMO-2I-EPS, are better than the global ensemble of Reading.

The percentages of outliers for the ensemble system considered as a function of the forecast range are shown in Fig. 5.

Considering all DPCN staions (top left graph), it is possible to see that despite the lower ensemble size, COSMO-2I-EPS has often the lowest values, compared to the other two ensemble systems with a lower horizontal resolution. So, it can be stated that in this case too, the results obtained by COSMO-2I-EPS are satisfactory. Looking at the other three panels of the Fig. 5 it can be stated that:

- the percentage of outliers increases according to the station altitude: there are less outliers in the plains than in the mountains;
- in the plain there is little difference between the three ensemble systems; these differences increase with the altitude, infact the percentage of outliers obtained with the only mountain stations shows considerable dissimilarity beetween the ensembles;
- for hill and mountain observation datasets, a diurnal cycle is visible only in systems with parametrized convection (ECMWF ENS, COSMO-LEPS); instead, the diurnal cycle is hardly identifiable for the lowland stations;
- for almost all forecast ranges COSMO-2I-EPS has the lowest percentage of outliers.

Therefore also the percentages of outliers, studied according to the altitude of DPCN stations, indicate the good skill of COSMO-2I-EPS.

Second verification period: from 15th October to 15th November 2018

In this second period it has considered only the 24-hour total precipitation, the verification is performed with the rank historam.
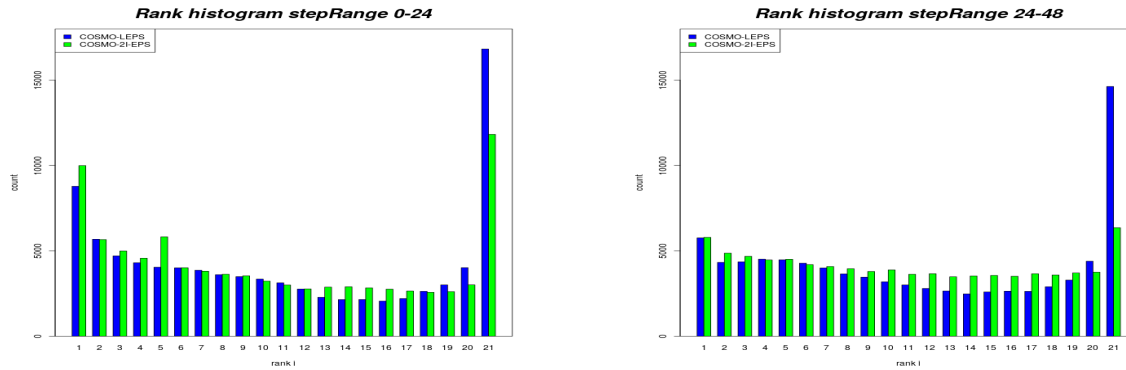
Figure 6: The rank histograms for COSMO-LEPS in blue and COSMO-2I-EPS in green, on the left for the first 24 hours of the forecast range, on the right for the second 24 hours.

The rank histogram is not a verification method per se, but rather a diagnostic tool to evaluate the spread of an ensemble. The underlying assumption is that the ensemble member forecasts are distributed so as to delineate ranges or "bins" of the predicted variable such that the probability of occurrence of the observation within each bin is equal. For each specific forecast, the bins are determined by ranking the ensemble member forecasts from lowest to highest. The interval between each pair of ranked values forms a bin. If there are N ensemble members, then there will be N+1 bins. The outer bins, lowest and highest-valued, are open-ended. Rank histograms are prepared by determining which of the ranked bins the observation falls into for each case, and plotting a histrogram of the total occurrences in each bin, for the full verification sample. The assumption underlying the rank is that the probability that the observation will fall in each bin is equal. If this is true, then over a large enough sample, the histogram should be flat or roughly so. Then one can conclude that on the average, the ensemble spread correctly represents the uncertainty in the forecast. Also in this case the verification method is the nearest grid point and the comparison is only between COSMO-LEPS and COSMO-2I-EPS, because these systems have the same number of member (20 in 2018).

The U-shape of the rank histograms (see Fig. 6) indicates the subdispersion of both ensemble systems, in particular COSMO-LEPS. This subdispersion is stronger in the last bin of the most intense precipitation and in particular for COSMO-LEPS.

Summary and Outlook

The present work aims to establish the performance of three ensemble systems with different characteristics, but in particular with a different horizontal resolution. While ECMWF ENS and COSMO-LEPS run on an daily basis, COSMO-2I-EPS is still on a pre-operational phase, with a full operational implementation planned towards the next months. Therefore, particular attention has been paid to this new ensemble, especially because it provides new types of numerical modeling products which needs to be assessed and because the best performances were expected from it. A systematic comparison between the three ensemble systems was undertaken during a "pilot period" from 20th to 27th June 2016. During this period, characterised by particularly unstable weather situation over the Italian Peninsula, the performances of the three systems were compared in terms of 2-metre temperature and precipitation. The forecasts in terms of 2-metre temperature and 6-hourly cumulated precipitation were verified against the Northern-Italy non-GTS network and the National Civil Protection Department network respectively. The results for 2-metre temperature indicate the under-dispersion issue for the different ensemble systems, although it can be noticed that the performance obtained by COSMO-2I-EPS (and in general by the COSMO-based ensembles) is quite satisfactory.
Rank Probability Score and percentage of outliers were considered to evaluate the skill of the three ensemble systems in terms of precipitation. In most cases, the scores indicate COSMO-2I-EPS having the best performance. In order to provide more insight on the obtained results and to assess the dependence of the scores on the altitude, it was decided to divide the stations of the National Civil Protection Department into three

groups: plain, hill and mountain stations. With this division, it turns out that the performance of the systems tends to worsen with the altitude, also accentuating the diurnal cycle. This happens because it has rained more over mountain areas and during the afternoon. Anyway the scores obtained by COSMO-2I-EPS remain the best in most cases. COSMO-2I-EPS achive good results also in the verification with rank histograms, for the period from from 15th October to 15th November 2018. This work can be seen as a pilot study, there is no claim to consider it complete and exhaustive, but rather a starting point for further developments and investigations or a "modus operandi" for similar studies. In fact, the periods examined are too short to have solid results from a statistical point of view. This would take a longer evaluation time, comparing the three ensembles for different atmospheric phenomena and weather types. All the results shown in this work have been obtained with the verification method of the nearest grid point. So a further idea for future studies may be to use the method of boxes to calculate the probabilistic scores in other cases; it will be interesting to see if the results will be better or worse than those obtained with the nearest grid point.

References

[1] Buizza, 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction Systems. *Mon. Wea. Rev.*, **125**, 99-119.

[2] Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, Y. Zhu, 2005: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Monthly Weather Review*, **133**, 1076-1097.

[3] Gofa, F., Tzaferi, D. and Charantonis, T., 2010. Application and verification of ECMWF products. *Hellenica National Meteorological Service (HNMS)*

[4] Ghelli, A., 2009. Observations and their importance in the verification process: View of the Joint Working Group on Forecast Verification Method Research. *ECMWF, Twelfth Workshop on Meteorological Operational Systems, 2-6 November 2009*

[5] Montani, A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U. and co-authors, 2003a. Operational limited-area ensemble forecast based on the Lokal Modell. ECMWF Newsletter 98, 2-7. Available at: http://www.ecmwf.int/publications/.

[6] Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi S. and Buizza, R., 2003b. The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts. *Nonlin. Proc. Geophys.*, **10**, 261-274.

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

[8] Murphy, 1969. On the "Ranked Probability Score". *J. Appl. Meteor.*, **8**, 988-989.

[9] Schättler, Doms and Schraff, 2016. A Description of the Nonhydrostatic Regional COSMO-Model. Part VII: User Guide. (www.cosmo-model.org).

[10] Tibaldi, S., Paccagnella, T., Marsigli, C., Montani, A. and Nerozzi, F., 2006. Limited area ensemble forecasting: the COSMO model. Predictability of Weather and Climate. *Cambridge University Press*, 489-513.

[11] Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability, *Advance in Atmospheric Sciences*, **22**, No. 6, 781-788.

Influence of Perturbation Type on Results of EPS Forecasts of Surface Elements

Grzegorz Duniec, Andrzej Mazur

Institute of Meteorology and Water Management – National Research Institute, Warsaw, Poland 61 Podlesna str., PL-01-673 Warsaw, Poland

1 Introduction

Abstract

The results from research on COSMO-EPS, carried out at IMWM, are presented. The operational EPS (Ensemble Prediction System) set-up is based on perturbations of soil surface-area index of the evaporating fraction of grid points over land. In the research mode, six different types of perturbation is additionally applied. Long-term evaluation results of different methods of EPS-post-processing is presented in the paper. As a general rule, using Artificial Neural Network (ANN) values of EPS mean are significantly closer to observation of air temperature/dew point temperature/surface pressure or wind speed than those computed as deterministic forecast.

Introduction

Extensive tests conducted during the COTEKINO Priority Project proved that small perturbations of selected soil parameter were sufficient to induce significant changes in the forecast of the state of atmosphere and to provide qualitative selection of a valid member of an ensemble *(Duniec and Mazur,2014)*. Changes of c_soil[*]) had a significant impact on values of air temperature, dew point temperature and relative humidity at 2m agl., wind speed/direction at 10m agl., and surface specific humidity *(ibidem)*. Other approaches of perturbation (as presented in previous work) would result in different forecast, expecting even a synergy while combining perturbation methods for the same run(s). The research has been carried out for the entire year 2011. For the ANN training results from January to October have been set. Methods *(approaches)* have been tested on results from November 2011. [4].
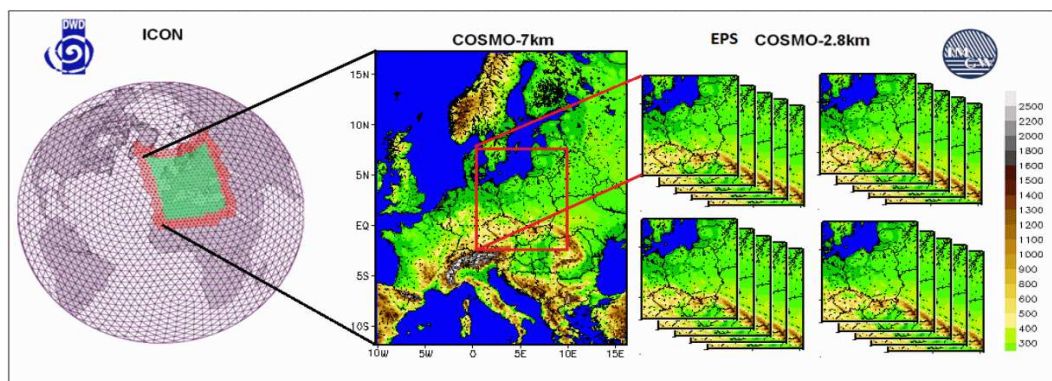

Figure 1: EPS operational configuration *(Duniec et al., 2016)*

---

[4][*])surface-area index of the evaporating fraction of gridpoints over land

Table 1: Deterministic model(s) – source of ICs/BCs for operational EPS **ibidem)**

Model	Grid size NxMxL	Forecast length(h)	Resolution(km)
ICON (DWD)	2949120 triangles	78	13
COSMO v. 5.01	415x460x40	13	7
COSMO v. 5.01$^{\hat{*}}$	380x405x50	78	2.8

Forecasts of air temperature and dew point temperature at 2m agl., surface pressure and windspeed at 10m agl., as well as other fields are available. As a result, plots/chart of EPS mean, spread, probabilities of threshold exceedances are prepared in the routine manner. Results in a raw form are subsequently stored for further research research (e.g. skill-spread relation) and simultaneously calibrated.

Artificial Neural Network(ANN) mean(s) in this research have been compared with direct results from "deterministic" forecasts (DET). ANN in this resarch consisted of 24 input neurons (20 members, geographical coordinates, forecast start and forecast hour; there were 5 neurons set in a single hidden layer, with hyperbolic tangent accepted as the activation function.

The following perturbations were considered:

a) *c_soil*-perturbation of a parameter describing evaporation from soil(described above);

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

c) *eff-c_soil*-perturbation of the collection efficiency coefficient together with c_soil;

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 with c_soil;

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

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

h) *operational* perturbation of c_soil with a different random number generator (Duniec et al., 2016), operational runs

3 Results – comparison of results for different methods of perturbations.

Table 2: Basic statistics for different perturbation methods with ANN post-processing, compared with values from deterministic runs, as calculated for November, 2011 (ME – mean error, MAE – mean absolute error, RMSE-root-mean square error, MinE-minimum error, MaxE-maximum error)

| Means	ME	MAE	RMSE	MaxE	MinE
Dew point					
c_soil	-0.11338	1.45981	1.99090	12.30946	-9.88111
eff-coeff	-0.01667	1.47110	2.00072	11.11471	-9.41829
eff-c_soil	0.04247	1.45814	1.98011	11.53134	-9.92467
eps-all	-0.00854	1.49234	2.02759	11.24309	-9.09813
laf-pert	-0.04460	1.46721	1.99155	10.89753	-9.27700
laf-c_soil	0.01080	1.51334	2.04447	10.83230	-8.87939
laf-eff	-0.05678	1.46489	1.99521	10.47621	-9.37223
operational	0.02424	1.46355	1.98274	10.49569	-9.10767
deterministic	-0.40246	1.58561	2.18141	13.04700	-10.08800
Air temp					
c_soil	0.17387	1.77275	2.32496	10.93927	-15.88361
eff-coeff	-0.15550	1.77681	2.34730	11.16211	-16.14814
eff-c_soil	-0.08983	1.76932	2.34525	10.54141	-16.63289
eps-all	0.07055	1.77859	2.34857	10.31766	-15.89856
laf-pert	0.09633	1.78876	2.34243	10.67038	-14.61441
laf-c_soil	0.06539	1.76116	2.31501	10.84628	-15.06645
laf-eff	-0.18840	1.77813	2.33403	10.50841	-15.01652
operational	-0.13666	1.78166	2.34402	10.80536	-15.59283
deterministic	0.44751	1.90295	2.62627	11.77100	-12.86600
Windspeed					
c_soil	0.04309	1.17025	1.58737	9.72965	-9.05961
eff-coeff	-0.07475	1.17811	1.59937	9.64747	-9.06740
eff-c_soil	0.02018	1.16574	1.58048	9.74929	-9.87465
eps-all	0.04844	1.16578	1.58195	9.74003	-6.55868
laf-pert	0.10026	1.17006	1.58576	9.77432	-5.21126
laf-c_soil	-0.04346	1.17756	1.60043	10.00780	-11.41867
laf-eff	-0.07655	1.17344	1.58327	9.63682	-7.45664
operational	-0.03980	1.17237	1.59618	9.70848	-10.99594
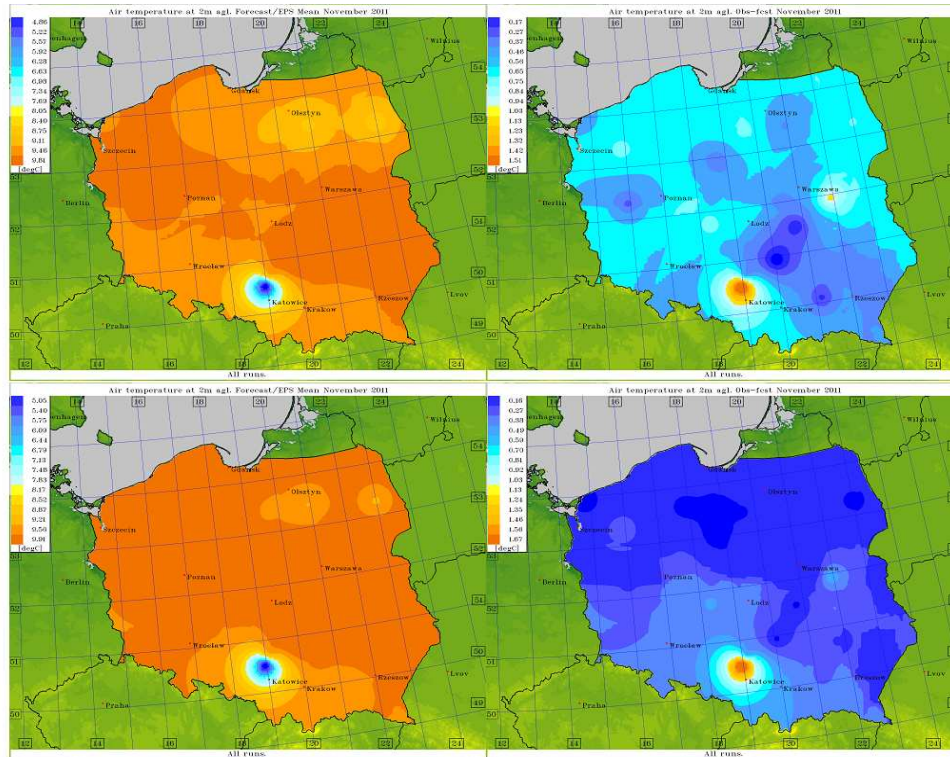
Figure 2: Spatial distribution of air temperature at 2m: ANN (eff-c_soil) mean (upper left) and skill (upper right), deterministic mean forecast (lower left) and skill (lower right).All avg. values for November 2011
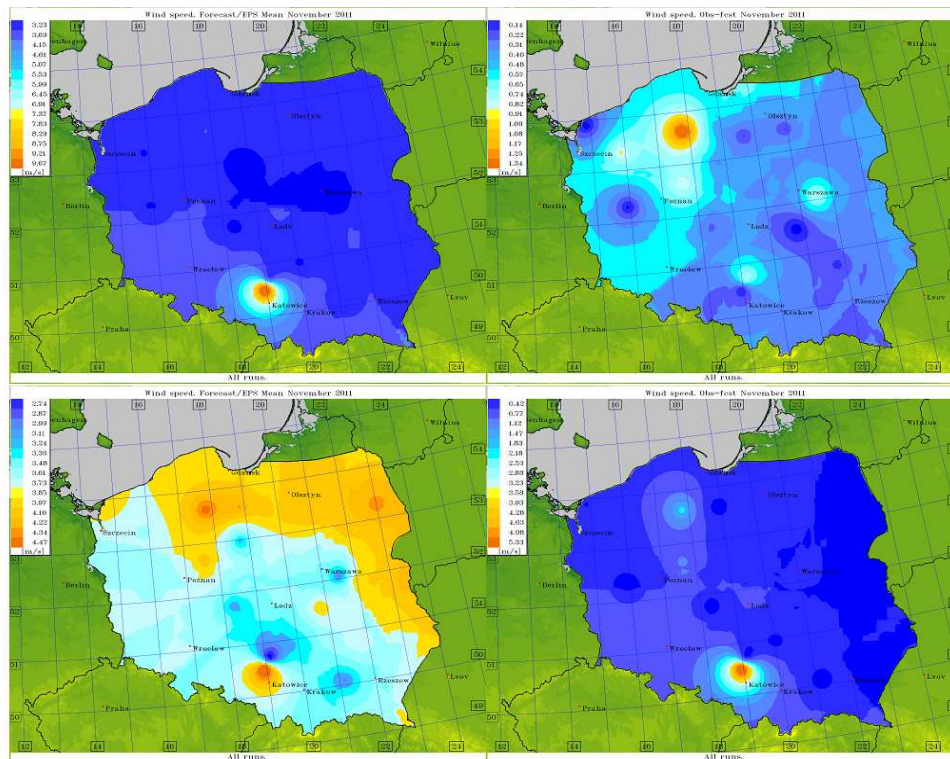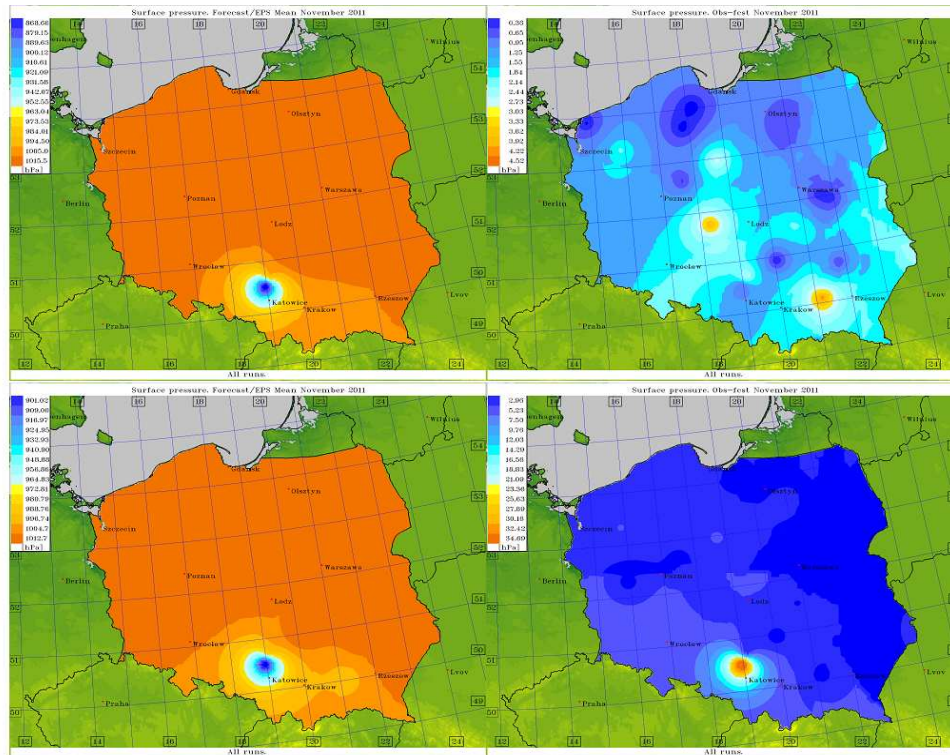
Figure 3: Spatial distribution of wind speed at 10m: ANN (eff-c_soil) mean (upper left) and skill (upper right), deterministic mean forecast (lower left) and skill (lower right). All avg. values for November 2011

Figure 4: Spatial distribution of surface pressure: ANN (c_soil) mean (upper left) and skill (upper right), deterministic mean forecast (lower left) and skill (lower right). All avg. values for November 2011
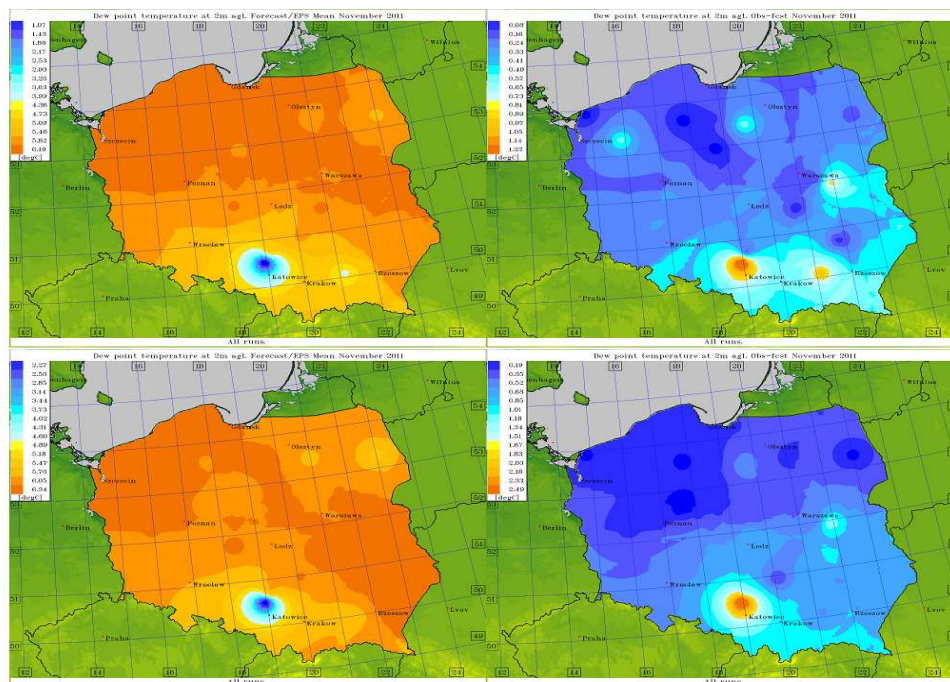

Figure 5: Spatial distribution of dew point temperature at 2m: ANN (eff-c_soil) mean (upper left) and skill (upper right), deterministic mean forecast (lower left) and skill (lower right). All avg. values for November 2011

4 Conclusions

Except for few cases of min/max errors results of ANN postprocessing gives evidently the best results in terms of statistic evaluation in comparison to "deterministic" forecast. Keeping in mind arguments against ANN (complicated pre- and post-processing, need for big data sets and huge computational resources, long computational time for training) one can say that this method, with ready-to-use dedicated software with source codes (FORTRAN) is sophisticated yet elegant and intuitive concept.

Improvement in preliminary case study can be clearly observed and forecasts are getting better and better with the extension of learning period, which is a key reason to go on with ANN in an operational EPS. However, there was no effect of synergy with combining perturbation methods and objects. Yet, c_soil alone and with combination with some other perturbation methods seemed to be the best as far as overall statistics is concerned (see *Table 2* and *Figures 2-5*).

The results in a poster form to be presented partially at ICCARUS in Offenbach, Germany, March 2019 and partially at EGU General Assembly in Vienna, Austria, April 2019.

References

[1] Duniec, G. and Mazur, A. (2014): COTEKINO Priority Project – Results of Sensitivity Tests, *COSMO Newsletter 14*, **106-113**.

[2] Duniec, G., Interewicz, W., Mazur, A. and Wyszogrodzki, A.(2016): Operational Setup of the COSMO-based, Time-lagged Ensemble Prediction System at the Institute of Meteorology and Water Management – National Research Institute. Met. Hydrol. Water Manage. (2017) vol. 5; (2): **43-51**.

[3] Mazur, A. and Duniec, G. (2017): SPRED PP activities at IMWM-NRI. Presented at COSMO GM, Jerusalem, Israel.

[4] Mazur, A., Duniec, G. and Interewicz, W. (2018): Introductory activities in PP APSU at IMWM-NRI and results of ANN post-processing of EPS forecasts. Presented at COSMO GM, Sankt Petersburg, Russia.

ANN post-processing of EPS

Andrzej Mazur, Grzegorz Duniec

Institute of Meteorology and Water Management – National Research Institute, Warsaw, Poland 61 Podlesna str., PL-01-673 Warsaw, Poland

1 Introduction

The results from research on COSMO-EPS, carried out at IMWM, are presented. The operational EPS set-up is based on perturbations of soil surface-area index of the evaporating fraction of grid points over land. Long-term evaluation results of different methods of EPS-post-processing. As a general rule, using Artificial Neural Network (ANN) values of EPS mean are significantly closer to observation of air temperature/dew point temperature or wind speed than those calculated as simple average or Multi-linear Mean. Extensive tests conducted during the COTEKINO Priority Project proved that small perturbations of selected soil parameter were sufficient to induce significant changes in the forecast of the state of atmosphere and to provide qualitative selection of a valid member of an ensemble (Duniec and Mazur, 2014). Changes of c_soil*) had a significant impact on values of air temperature, dew point temperature and relative humidity at 2m agl., wind speed/direction at 10m agl., and surface specific humidity (ibidem). The usage of an idea of time-lagged initial and boundary conditions allowed obtaining a valid ensemble and using it efficiently in an operational mode. Further work is intended to focus on "tuning" ensemble performance and to provide quantitative quality scores. For this purpose the random number generator combined with perturbations of initial soil surface temperature and the dependence of amplitude of perturbation on soil type will be implemented in the COSMO model.

While the set of equally weighted time-lagged forecasts improve short-range forecasts, the further progress may also be sought by adopting a regression approach to compute set of weights for different time-lagged ensemble members. EPS runs operationally at IMWM since January, 2016. It covers 4 runs/day, with 48 hours forecasts, 20 members/4 groups (using Time-lagged Ics/BCs; see Duniec G. et al. (2016); conf. Fig.1 below). Amplitude of perturbation of c_soil depends on type of soil (clay, sand, peat etc). [5]
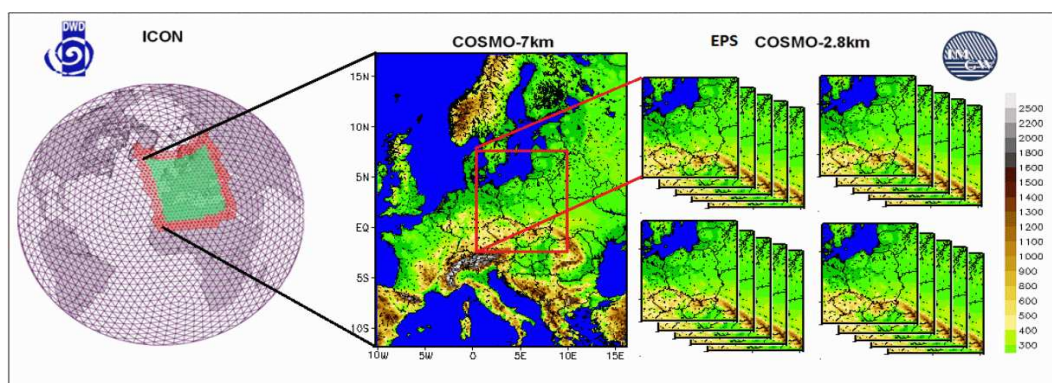

Figure 1: EPS operational configuration

[5]*) surface-area index of the evaporating fraction of gridpoints over land

Table 1: Deterministic model(s) – source of ICs/BCs for operational EPS **ibidem)**

Model	Grid size NxMxL	Forecast length(h)	Resolution(km)
ICON (DWD)	2949120 triangles	78	13
COSMO v. 5.01	415x460x40	13	7
COSMO v. 5.01*	380x405x50	78	2.8

2 Some Formulas

Details of the deterministic models configuration are as follows:

Forecasts of air temperature and dew point temperature at 2m agl., surface pressure and windspeed at 10m agl., as well as other fields are available. As a result, plots/chart of EPS mean, spread, probabilities of threshold exceedance are prepared in the routine manner. Results are subsequently stored for further research (e.g. skill-spread relation) Results of EPS forecasts are subsequently calibrated. Three basic methods of calibration were examined as shown in Fig.2 – simple arithmetic mean (SM), multilinear regression mean (MLR) and artificial neural network mean (ANN).

Table 2. Ensemble calibration – Simple Mean (SM) vs. multilinear regression (MLR) mean vs. ANN mean

Simple Mean$^{*)}$	Multilinear regression $^{**)}$	Artificial Neural Network $^{***)}$
x – forecast values, y – ensemble mean, m – # members	y – corrected forecasts – (new) ensemble mean, x – matrix of raw forecast values/parameters, β – weights (from previous fcsts.)	24 input neurons Adjust table row + λ,φ + ts,tc) 5 neurons in a single hidden layer activation function: hyperbolic tangent

*) Simple avg. – arithmetic mean, m=20 members;
**) # of predictors: n=24=20 members+geo coords+lead/fcst time;
***) Trained on data from July 2016 to March 2018, tested on data of April 2018

Figure 2: Ensemble calibration – Simple Mean (SM) vs. multilinear regression (MLR) mean vs. ANN mean

3 Results – comparison of results for three methods of post-processing.
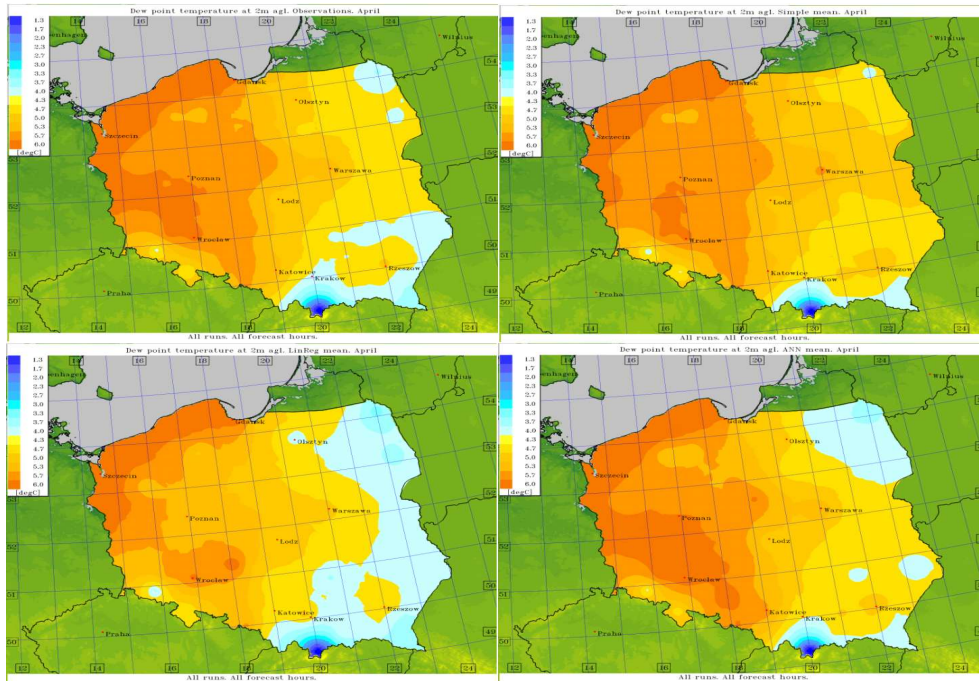

Figure 3: Spatial distribution of dew point temp. at 2m: mean observations (upper left), simple mean (upper right), MLR mean (24 predictors, lower left) and ANN mean (24 input neurons, lower right). All avg. values for April 2018.
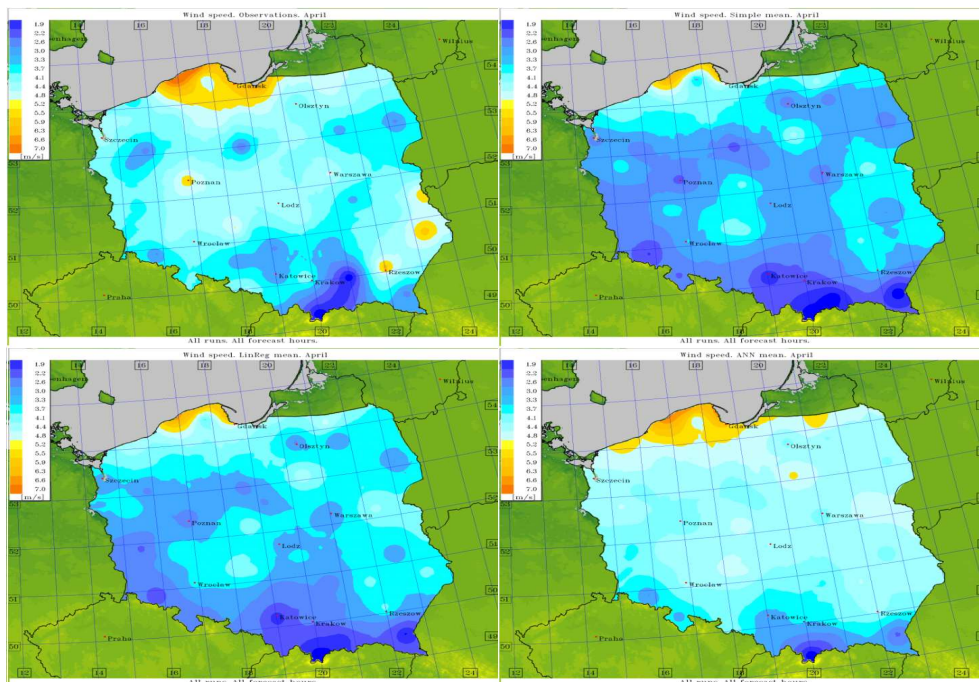

Figure 4: Spatial distribution of wind speed at 10m: mean observations (upper left), simple mean (upper right), MLR mean (24 predictors, lower left) and ANN mean (24 input neurons, lower right). All avg. values for April 2018.
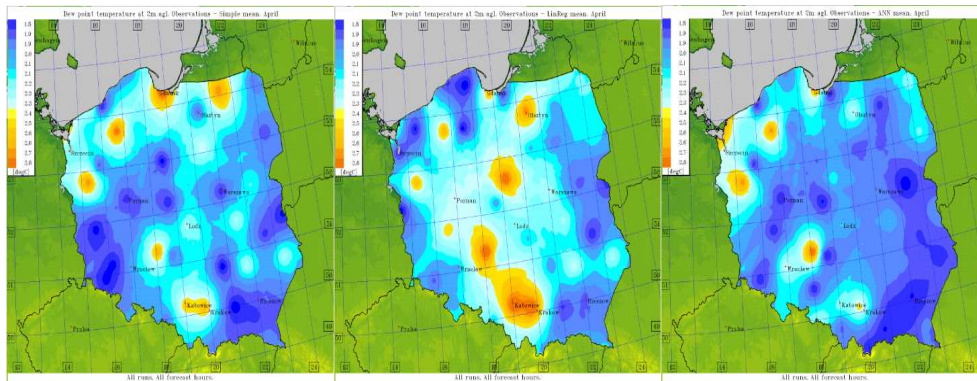
Figure 5: Spatial distribution of Mean Absolute Error (MAE) for dew point temp. at 2m, April 2018. Left: observations vs. simple mean; middle: observations vs. MLR mean; right: observations vs. ANN mean.
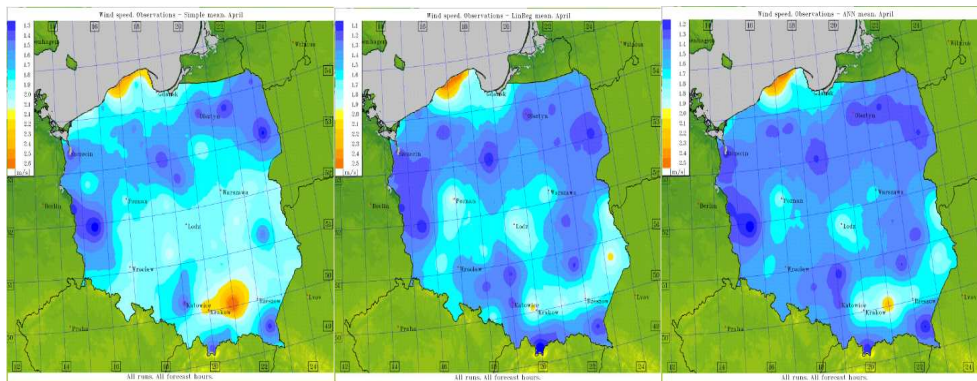

Figure 6: Spatial distribution of Mean Absolute Error (MAE) for wind speed at 10m, April 2018. Left: observations vs. simple mean; middle: observations vs. MLR mean; right: observations vs. ANN mean.
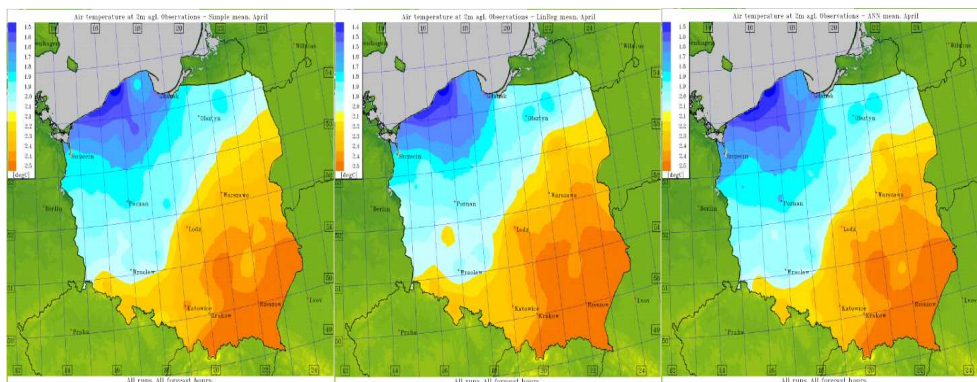

Figure 7: Spatial distribution of Mean Absolute Error (MAE) for air temp. at 2m, April 2018. Left: observations vs. simple mean; middle: observations vs. MLR mean; right: observations vs. ANN mean.

Table 2: Basic statistics for different post-processing methods, as calculated for April, 2018 (ME – mean error, MAE – mean absolute error, RMSE – root-mean square error, MinE – minimum error, MaxE – maximum error)

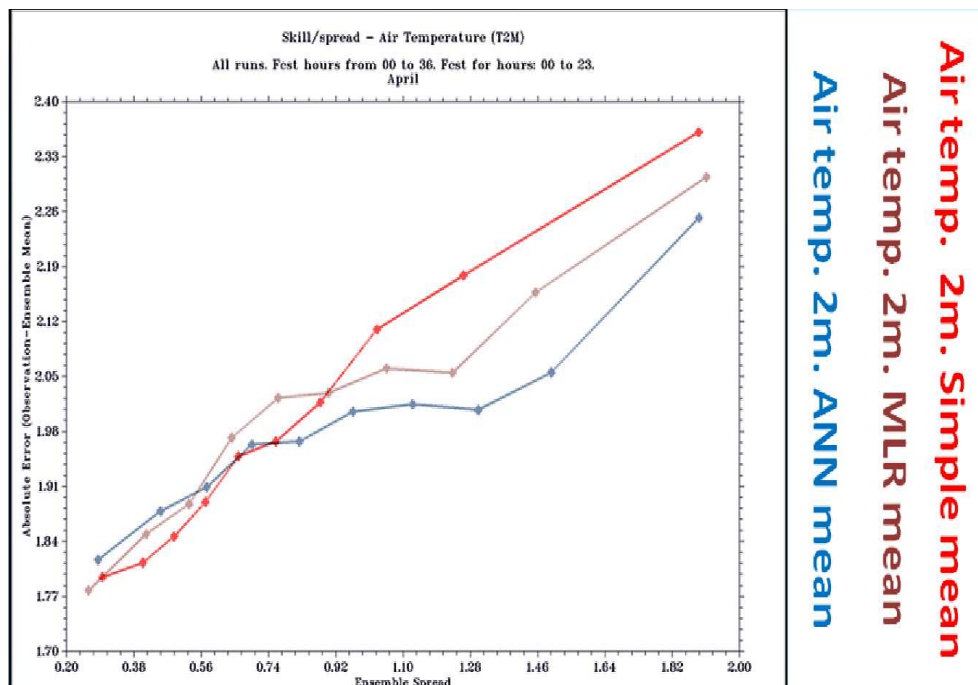| Means	ME	MAE	RMSE	MinE	MaxE
Dew point					
SM	0.253	2.009	2.812	-12.4	15.1
MLR	-0.310	1.989	2.755	-12.3	14.8
ANN	-0.244	1.981	2.750	-11.2	14.8
Air temp.					
SM	0.771	2.369	3.443	-14.600	18.100
MLR	0.475	2.252	3.206	-14.500	16.600
ANN	0.066	2.214	3.135	-13.600	15.500
Windspeed					
SM	-0.618	1.737	2.297	-13.6	13.6
MLR	0.113	1.488	1.978	-7.8	13.2
ANN	-0.200	1.436	1.814	-6.1	13.2

Figure 8: Skill/spread relation for air temp. at 2m, April 2018.

4 Conclusions

Except for single case of mean error for windspeed results of ANN post-processing gives evidently the best results in terms of statistic evaluation and skill-spread relation (see Fig. 7). Keeping in mind arguments against ANN (complicated pre- and post-processing, need for big data sets and huge computational resources, long computational time for training) one can say that this method, with ready-to-use dedicated software with source codes (FORTRAN) is sophisticated yet elegant and intuitive concept. Improvement in preliminary case study can be clearly observed and forecasts are getting better and better with the extension of learning period, which is a key reason to go on with ANN in an operational EPS.

Artificial Neural Network is linked to the DMO, to extend the learning period. In the operational mode 24 predictors is set (values from twenty ensemble members + spatio-temporal coordinates). The system is set in an (quasi)operational mode (slight delay due to calculations). Results are collected four times per day, so the structure of ANN can be updated frequently

The results in a poster form were presented at 40th EWGLAM/25th SRNWP Workshop in Salzburg, Austria, October 2018.

References

[1] Duniec, G. and Mazur, A. (2014): COTEKINO Priority Project – Results of Sensitivity Tests, COSMO Newsletter, 14, **106-113**.

[2] Duniec, G., Interewicz, W., Mazur, A. and Wyszogrodzki, A. (2016): Operational Setup of the COSMO-based, Time-lagged Ensemble Prediction System at the Institute of Meteorology and Water Management – National Research Institute. Met. Hydrol. Water Manage. (2017) vol. 5; (2): **43-51**.

[3] Mazur, A. and Duniec, G. (2017): SPRED PP activities at IMWM-NRI. Presented at COSMO GM, Jerusalem, Israel.

[4] Mazur, A., Duniec, G. and Interewicz, W. (2018): Introductory activities in PP APSU at IMWM-NRI and results of ANN post-processing of EPS forecasts. Presented at COSMO GM, Sankt Petersburg, Russia.

Forecasts of Convective Phenomena Using EPS-based Computation of Universal Tornadic Index

ANDRZEJ MAZUR

Institute of Meteorology and Water Management – National Research Institute, Warsaw, Poland 61 Podlesna str., PL-01-673 Warsaw, Poland

Abstract

The results from research on COSMO-EPS, carried out at IMWM, are presented. The operational EPS set-up is based on perturbations of soil surface-area index of the evaporating fraction of grid points over land. Usage a Universal Tornadic Index associated with the EPS forecasts system may be helpful in forecasting of severe convection phenomena. This idea was tested in case studies and in long-term evaluation.

1 Introduction

A simple and efficient method was proposed to produce reasonable number of valid ensemble members, taking into consideration predefined soil-related model parameters. Introduced method of obtaining ICs/BCs – time-lagged setup – is another important factor that can add a significant increment to ensemble spread. These features were used to prepare a well-defined ensemble based on the perturbation of soil-related parameters, to be utilized both in operational (forecasting) work and in diagnostic mode. A special approach – using EPS-based forecasts for tornado forecasting – of diagnostic approach is presented here. Small perturbations of selected soil parameter were sufficient to induce significant changes in the forecast of the state of atmosphere and to provide qualitative selection of a valid member of an ensemble (Duniec and Mazur, 2014). Changes of c_soil*) had a significant impact on values of air temperature, dew point temperature and relative humidity at 2m agl., wind speed/direction at 10m agl., and surface specific humidity (ibidem). The usage of an idea of time-lagged initial and boundary conditions allowed obtaining a valid ensemble and using it efficiently in an operational mode. Further work is intended to focus on "tuning" ensemble performance and to provide quantitative quality scores. For this purpose the random number generator combined with perturbations of initial soil surface temperature and the dependence of amplitude of perturbation on soil type will be implemented in the COSMO model.While the set of equally weighted time-lagged forecasts improve short-range forecasts, the further progress may also be sought by adopting a regression approach to compute set of weights for different time-lagged ensemble members. EPS runs operationally at IMWM since January, 2016. It covers 4 runs/day, with 48 hours forecasts, 20 members/4 groups (using Time-lagged ICs/BCs; see Duniec G. et al. (2016); conf. Fig.1 below). Amplitude of perturbation of c_soil depends on type of soil (clay, sand, peat etc).
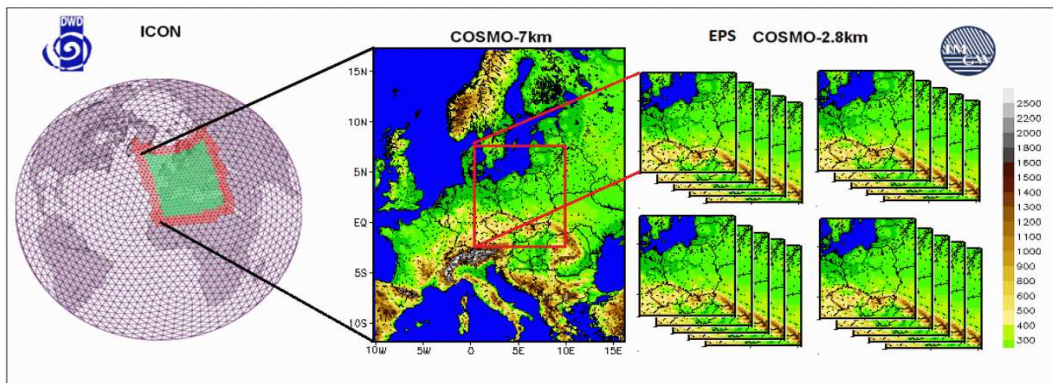
Figure 1: EPS operational configuration *(Duniec et al., 2016)*

Table 1: Deterministic model(s) – source of ICs/BCs for operational EPS

Model	Grid size NxMxL	Forecast length(h)	Resolution(km)
ICON (DWD)	2949120 triangles	78	13
COSMO v. 5.01	415x460x40	78	7
COSMO v. 5.01[*)]	380x405x50	48	2.8

*)time lagged ICs/BCs

2 Details of the deterministic models

Details of the deterministic models configuration are as follows:

Forecasts of air temperature and dew point temperature at 2m agl., surface pressure and windspeed at 10m agl., as well as other fields are available. As a result, plots/chart of EPS mean, spread, probabilities of threshold exceedance are prepared in the routine manner. Results are subsequently stored for further research (e.g. skill-spread relation). In this work a new index to predict severe convection phenomena (especially tornadoes, but also heavy thunderstorms, intensive precipitation episodes etc.) was assessed. This index (called Universal Tornadic Index, UTI) is in general based on a number of predictors related with a strong convection conditions, as follows:

$$UTI = \frac{\frac{CAPE \cdot SRH_{1km}}{200} \cdot \frac{5 \cdot (DLS-20)+(\frac{200-LCL}{10})}{100} + CAPE_{3km} + \frac{SRH_{1km}}{4}}{1000} \cdot \frac{LLS}{12} \cdot \frac{AMR_{500}}{100}$$

whit:

- CAPE being surface based Convective Available Potential Energy,
- CAPE_3km-surface based CAPE released below 3 km agl,
- LCL -Surface based lifting condensation level height,
- AMR_500-average mixing ratio below 500 m,both agl,
- LLS -0-1 km wind shear,
- DLS - 0-6 km wind shear (magnitude of vector difference),
- SRH_1km - 0-1 km storm relative helicity.

Other constrains applied are:

- if SRH_1km<0,then SRH_1km =0;
- if LCL>1500m or CAPE=0 or(convective precipitation amount <2mm/h),then UTI=0

Detailed description of the index and its climatology can be found in Taszarek and Kolendowicz (2013) or Taszarek et al., (2016).

3 Results

Single case, July 14th, 2012 (intensity peak 16:00 UTC). The event – a tornado that passed over the Bory Tucholskie primeval forest – was actually a combination of four tornadoes, with damage track at a distance of 60km, the total length of around 100km. Damaged was 105 buildings, tornado maximum intensity – F3 (Fujita scale), 1 death, 10 injuries, felled 5 km2 of Bory Tucholskie.
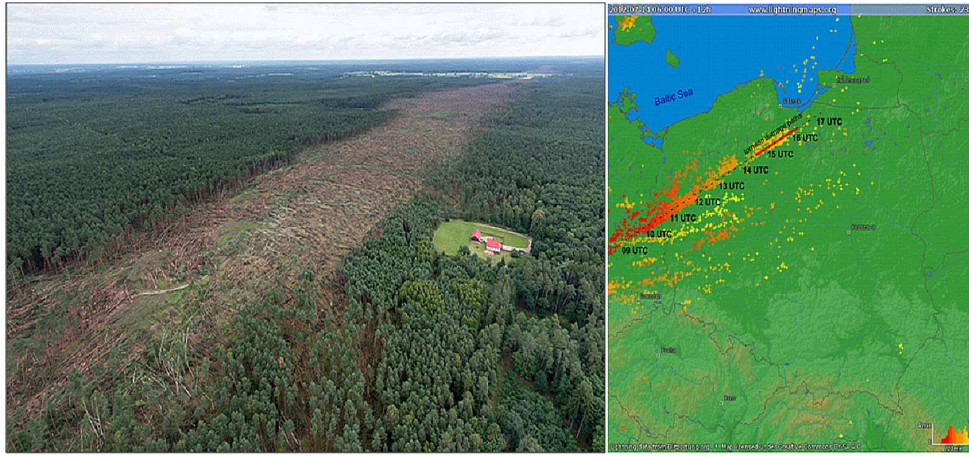
Figure 2: Left – damage track in Bory Tucholskie forest due to tornado of 14 July 2012. Photography: Kacper Kowalski. Right: Lightning captured by the lightning detection network(**www.blitzortung.org**), 0600 to 1800 UTC. Dashed line – radar-based time and position of the thunderstorm. Red line indicate tornado damage paths
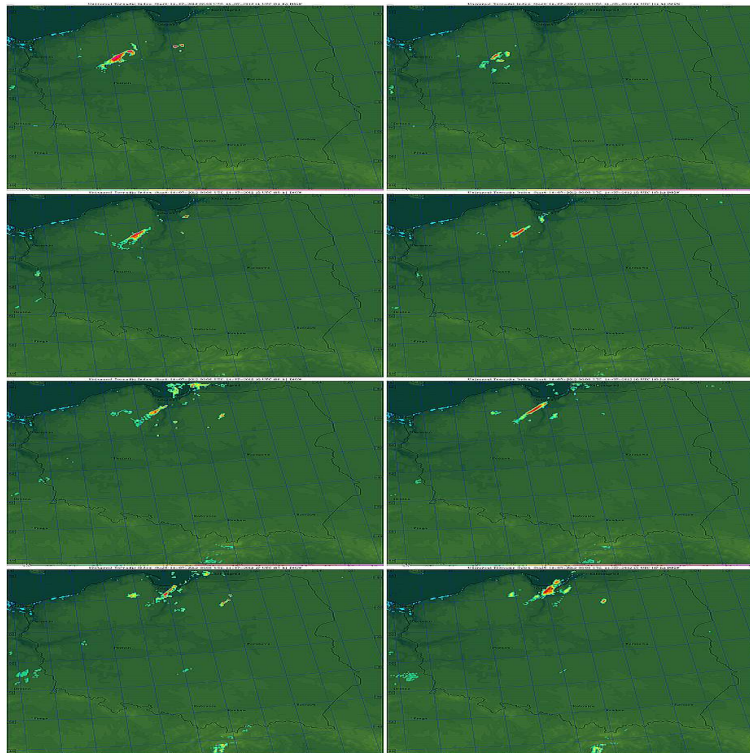

Figure 3: UTI forecast, July 14th, 2012, 14:00 to 17:00 UTC. Left: 2.8km deterministic run, right: 2.8km EPS-mean

Long term evaluation – UTI EPS-based thunderstorm forecasts.

Since UTI uses many factors/indicators as predictors (especially CAPE, storm relative helicity, convective precipitation, wind shear etc.) – it can be functional in forecasting not only tornadoes, but other convection phenomena of severe intensity, like thunderstorms (observed with the Polish lightning detection network PERUN). This part was performed with the archive data, starting from 2012. The results are shown in the following figures. Left panel present mean flashrate, right – mean value of UTI for the period of interest.
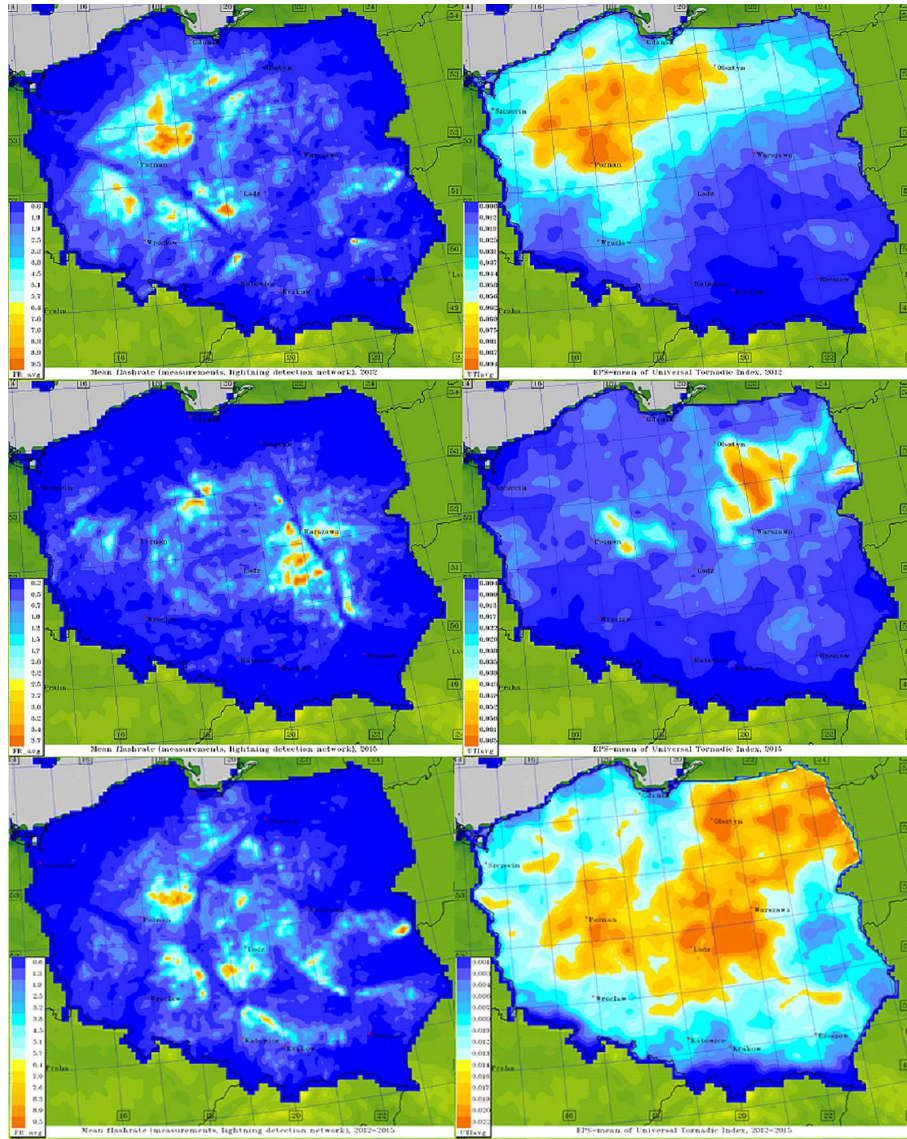
Figure 4: Left – mean flashrate (measurements), average values. Right – EPS-UTI forecasts, average values. Upper charts – year 2012, middle charts – year 2015, lower charts – period 2012-2015.
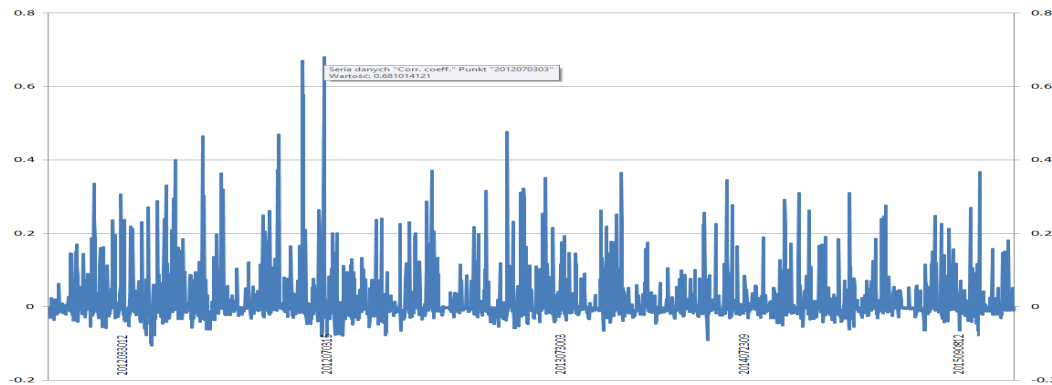
Figure 5: Correlation coefficient UTI vs. FR, consecutive values 2012-2015 (overall avg. 0.1). Best correction attained for July 3rd, 2012, 03:00 UTC (heavy storm over central and northern Poland).
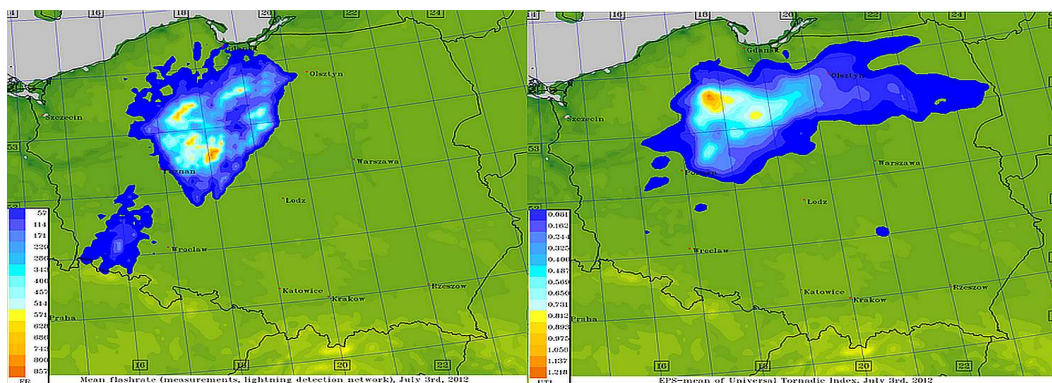

Figure 6: Left – mean flashrate (measurements). Right – EPS-UTI forecasts. July 3rd, 2012, 03:00 UTC

4 Conclusions

A relatively simple method of forecasting extreme convective phenomena has been proposed. This method uses Universal Tornadic Index as an indicator of the occurrence of a convective phenomenon. Since it utilizes many factors (CAPE, storm relative helicity, convective precipitation, wind shear etc.) – it can be useful in forecasting not only tornadoes, but also thunderstorms or squalls. Application of EPS in CP scale based on time-lagged ICs/BCs allows improving forecasts (especially due to the removal of false alarms). The research was carried out using archive data, starting from 2012. The noteworthy correlation between significantly higher EPS-UTI values and occurrence of thunderstorms was established.

Model forecasts with a spatial resolution of 2.8km and initial conditions – results of deterministic model with a resolution 7km – were used for the study. Additional filters (precipitation amount, CAPE threshold value etc.) were used. The use of numerical forecasts of meteorological model may be supportive of severe convective storm prediction. EPS-mean value of UTI is comparable with the ones calculated in a deterministic run(s), however significantly less amount of "noise signals" is observed. Thus, one can expect (in operational mode) decrease of FAR/increase of POD.

References

[1] Albergel, C., de Rosnay, P., G. Balsamo, G., Isaksen, L., Munoz-Sabater, J., 2012: Soil Moisture Analyses at ECMWF: Evaluation Using Global Ground-Based In Situ Observations. *J. Hydrometeor*, 13, **1442-**

1460.

[2] Aligo, E. A., Gallus, W. A. and Segal, M., 2007: Summer Rainfall Forecast Spread in an Ensemble Initialized with Different Soil Moisture Analyses. *Wea. Forecasting*, 22, **299-314**.

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; Proceedings from the COSMO General Meeting 2005.

No. 7: May 2008; Proceedings from the COSMO General Meeting 2006.

No. 8: August 2008; Proceedings from the COSMO General Meeting 2007.

No. 9: December 2008; Proceedings from the COSMO General Meeting 2008.

No.10: January 2010; Proceedings from the COSMO General Meeting 2009.

No.11: February 2011; Proceedings from the COSMO General Meeting 2010.

No.12: March 2012; Proceedings from the COSMO General Meeting 2011.

No.13: April 2013; Proceedings from the COSMO General Meeting 2012.

No.14: April 2014; Proceedings from the COSMO General Meeting 2013.

No.15: July 2015; Proceedings from the COSMO General Meeting 2014.

No.16: June 2016; Proceedings from the COSMO General Meeting 2015.

No.17: July 2017; Proceedings from the COSMO General Meeting 2016.

No.18: November 2018; Proceedings from the COSMO General Meeting 2017.

No.19: October 2019; Proceedings from the COSMO General Meeting 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):
Description 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):
Final report on priority project VERSUS (VERification System Unified Survey).

No. 13: Chiara Mirsigli (2009):
Final report on priority project SREPS (Short Range Ensemble Prediction System).

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 "Further Developments of the Runge-Kutta Time Integration Scheme" (RK); Final Report.

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

No. 17: D. Leuenberger, M. Stoll, 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: M. 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. Muellner, Z. P. Piotrowski, B. Rosa, P. L. Vitagliano, D. Wojcik, M. Ziemianski (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: Jean-Marie 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, D. Gayfulin (2016):
         A Stochastic Pattern Generator for ensemble applications.

No. 30: Dmitrii Mironov, Ekaterina 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: Naim Vela (2017):
         V.A.S.T. (Versus Additional Statistical Techniques) User Manual (v2.0).

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. Foerstner, J. Helmert, E. Kazakova, A. Kirsanov, V. Kopeikin, E. Kukanova, D. Majewski, C. Marsigli, G. de Morsier, A. Muravev, T. Paccagnella, U. Schaettler, C. Schraff, M. Shatunova, A. Shcherbakov, P. Steiner, M. Zaichenko (2017):
         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: INSPECT Final Report 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

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