Consortium

for

Small-Scale Modelling

Technical Report No. 48

The Priority Project C2I, Transition of COSMO to ICON - Final Report

December 2022

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_48

www.cosmo-model.org

Editor: Massimo Milelli, CIMA Foundation

The Priority Project C2I, Transition of COSMO to ICON - Final Report

***D. Rieger*[1*], *D. Boucouvala*[2], *F. Gofa*[2], *C. Kolyvas*[2], *P. Khain*[3], *H. Muskatel*[3], *A. Shtivelman*[3], *A. Baharad*[3], *L. Uzan*[3], *Y. Levi*[3], *F. Marcucci*[4], *F. Sudati*[4], *N. Zaccariello*[4], *V. Garbero*[5], *N. Vela*[5], *E. Oberto*[5] *E. Minguzzi*[6], *D. Cesari*[6], *T. Gastaldo*[6], *V. Poli*[6], *C. Marsigli*[6,1], *M.S. Tesini*[6], *I. Cerenzia*[6], *J. Linkowska*[7], *W. Interewicz*[7], *D. Wójcik*[7], *A. Iriza-Burcă*[8], *B.A. Maco*[8], *R. Dragomir*[8], *Ş. Gabrian*[8,9], *Ş. Dinicilă*[8,10], *M. Bogdan*[8], *T. Bălăcescu*[8], *R.C. Dumitrache*[8], *G.R. Bonatti*[11], *R.R. dos Santos*[11], *Y.K.L. Kitagawa*[11], *R.B. da Silveira*[12]**

[1,*] ***Project Leader, Deutscher Wetterdienst - DWD***

[2] ***Hellenic National Meteorological Service***

[3] ***Israel Meteorological Service***

[4] ***CNMCA, National Centre for Aerospace Meteorology and Climatology***

[5] ***Arpa Piemonte - Regional Agency for the Protection of the Environment of Piedmont***

[6] ***Arpae-Emilia-Romagna (Italy)***

[7] ***Institute of Meteorology and Water Management National Research Institute***

[8] ***National Meteorological Administration, Romania***

[9] ***Politehnica University of Bucharest, Faculty of Engineering in Foreign Languages***

[10] ***University of Bucharest, Faculty of Physics***

[11] ***Instituto Nacional de Meteorologia, INMET, Brasília, DF, Brazil***

[12] ***Sistema de Tecnologia e Monitoramento Ambiental do Paraná - SIMEPAR***

1 Abstract

The Priority Project 'Transition of COSMO to ICON' (PP C2I) was launched in 2018 and concluded in 2022. The aim of the project is to ensure a smooth transition for the Consortium for Small Scale Modelling (COSMO) national meteorological services (NMS) to the ICON model.

2 Introduction

The Priority Project 'Transition of COSMO to ICON' (PP C2I) was launched in 2018. At that time, the Consortium for Small Scale Modelling (COSMO) science plan called for a harmonization of development of the COSMO and ICON (ICOsahedral Nonhydrostatic) models in the time horizon of 2020. Also at that time, DWD reduced further development of the COSMO model to the level needed for operational production. It was therefore necessary for the COSMO consortium to start the migration to ICON-LAM (ICON - Limited Area Model) as the future operational model.

In the meantime, the limited area, high-resolution model ICON-D2 was introduced for operational numerical weather forecasts on 10 February 2021. With this, an era of more than 20 years of COSMO-model forecasts ended in Germany. The preoperational phase of ICON-D2 at DWD showed, that ICON-D2 outperforms the previous operational setup COSMO-D2 in terms of the quality of the results as well as in terms of computational efficiency.

This project C2I represents a major step in the implementation of the COSMO strategy regarding the evaluation of ICON-LAM and the transition to ICON. The overall goal is to ensure a smooth transition from the COSMO model to ICON-LAM within a four-year period. At the end of the PP C2I, each participating institution shall have reached a state, where ICON-LAM may replace the COSMO model in their operational forecasting system.

The participants of PP C2I are the national meteorological services of Greece, Israel, Italy, Poland, Romania and their local partners. The Russian national meteorological service successfully contributed to C2I until February 2022. With the national meteorological service of Brazil, INMET, one of the COSMO licensees also joined the project. In addition, a close collaboration with the CLM Community has been established.

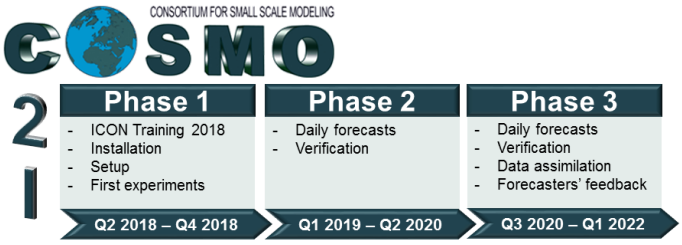


Figure 1: Illustration of the Priority Project C2I time line.

The time line of the project C2I is depicted in figure 1. An ICON training course and a workshop to find suitable setups and conduct first experiments marked the beginning. A summary of the outcome of this workshop that took place in October 2018 is presented in Rieger et al. (2018). With the knowledge from this workshop, the participants were able to set up ICON forecasts at their institutions. During 2019 and 2020, these setups were improved and the

verification was adopted to the ICON results. In summer 2020, most of the C2I participating institutions have reached a status where the verification could be compared to their COSMO forecast results. First verification results for the September/October/November 2020 period were presented in a intermediate report (Rieger et al., 2021).

The PP C2I has strong links to several priority projects and tasks of the COSMO consortium. The closest links exist with the following projects:

- Following the COSMO science plan, the **PP CDIC**[1] (Comparison of the Dynamical cores of ICON and COSMO, Baldauf et al., 2021) has performed comparisons of the dynamical cores of the COSMO and ICON models from September 2015 until August 2018, showing similar or improved scores for ICON. According to the project plan, PP CDIC is a preliminary study before a larger priority project, namely PP C2I, that performs comparisons of the complete two models.

- One crucial task of PP C2I is the continuous monitoring and verification of the model results. The software used for this purpose changed during the duration of PP C2I. A dedicated priority project taking place in parallel to C2I named **PP CARMA**[2] (Common Area with Rfdbk/Mec Application, Iriza-Burcă et al., 2022b) introduced the MEC/Rfdbk based software package for verification. Within this project, MEC/Rfdbk was implemented instead of the previously used VERSUS software for the Common Area verification. As every COSMO member is involved in the Common Area verification, the whole COSMO consortium gathered experience with this new software. PP C2I strongly profited from this parallel project as MEC/Rfdbk could be used for the C2I verification as well.

- Two COSMO member states, Italy and Switzerland, rely on GPU-based (graphical processing units) hardware to run their full ensemble forecasting systems. To be able to run the model system on GPUs, the source code needs to be adapted. For the ICON model, several projects are dedicated to enable the model for GPUs. The most important of these projects for the COSMO consortium is the **PP IMPACT**[3] (Icon on Massively Parallel ArchiteCTures). Just like PP CARMA, PP IMPACT is taking place more or less in parallel to C2I and covers this crucial requirement for the COSMO partners who rely on GPU-based hardware.

- In the project plan of C2I, feedback from the forecasters was included. The necessity of further extending this subtask (possibly also to licensees) and the wish to make it a continuous activity on a mid- to long-term perspective emerged. For this reason, a new priority project named **ICON-COMFORT**[4] (ICON-COMpetence in FORecasTing) has been started recently.

- During the course of PP C2I, it became clear that intensive technical support is needed for a successful transition from COSMO to ICON. As in the upcoming years not only the members states but also many licensing countries will start the transition process from the COSMO model to ICON, a well-working support framework is needed. A new priority project **PP C2I4LC**[5] (Establishing COSMO to ICON migration for Licensees' Countries) is currently set up to pave the way for these challenges ahead.

[1] **PP CDIC:** `https://cosmo-model.org/content/tasks/pastProjects/cdic/default.htm`
[2] **PP CARMA:** `https://cosmo-model.org/content/tasks/pastProjects/carma/default.htm`
[3] **PP IMPACT:** `https://cosmo-model.org/content/tasks/priorityProjects/impact/default.htm`
[4] **PP COMFORT:** `https://cosmo-model.org/content/tasks/priorityProjects/comfort/default.htm`
[5] **PP COMFORT:** `http://cosmo-model.org/content/tasks/priorityProjects/c2i4lc/default.htm`

Table 1: Times and plans for the operationalization of ICON at the COSMO member states.

COSMO Member	**Operationalization**
Germany	11/2019 (ICON-D2 pre-operational) 02/2021 (ICON-D2 operational)
Greece	06/2019 (pre-operational) 2023 (planned operational)
Israel	06/2020 (ICON-IL2.5)
Italy	07/2020 (ICON-IT)
Poland	09/2022 (ICON-PL)
Romania	01/2020 (ICON-RO-2.8km)
Switzerland	Q1/2024 (ICON-CH1-EPS, ICON-CH2-EPS)

One goal for the end of PP C2I was that the COSMO member countries are ready to switch to a deterministic ICON-based forecasting system. Table 1 summarizes at which time the individual COSMO partners succeeded in setting up an operational ICON forecasting system or plan to implement it. As it was clear from the beginning that GPU-based forecasting systems as well as ensemble systems can not be fully established within the framework of PP C2I, the goals and expectations that were put into C2I were fulfilled. This report will thus focus on results from one year of verification for each of the COSMO members. In addition, the COSMO members report on the computational efficiency of their ICON forecasts. Although significant progress was achieved in the task forecasters' feedback, we will leave the analysis of the results to the newly established PP ICON-COMFORT. PP ICON-COMFORT and the emerging PP C2I4LC can be seen as follow-up projects covering two major topics of PP C2I. The third major topic, verification, will be pursued as a continuous task in WG5 and in the common area verification.

3 Results from Greece

D. Boucouvala(1), F. Gofa(1), C. Kolyvas(1)

1. *Hellenic National Meteorological Service*

The seasonal evaluation of 00UTC ICON-GR2.5 (version 2.6.2.2) for the period of one year (June 2021-May 2022) against all surface synop stations of the Greek territory is presented in this report. Moreover, model performance is compared to the statistical indices derived from COSMO-GR4 and IFS-ECMWF models that are also used operationally at HNMS. ME/RMSE indices are calculated for the continuous parameters (2m Temperature, 2m Dew Point Temperature, 10m Wind Speed, Sea level Pressure, Total Cloud Cover). FBI, ETS, POD and FAR categorical scores are calculated for 6h cumulated precipitation. ICON-GR2.5 performance is promising and significantly improved for most parameters compared to COSMO-GR4.

3.1 Models setup

ICON-GR2.5 model has been running semi-operationally since June 2019. The current version is 2.6.2.2. The horizontal resolution is R2B10 (≈2.45 km) and SLEVE-type 50 vertical layers, with up to 23km height. The area of the model runs covers the Mediterranean (Fig. 2): -13°E-47°E/26°N-52°N. Details on the model setup can be found here[6].

Figure 2: ICON-GR2.5 domain.

COSMO-GR4 (Version 5.04) runs in a rotated grid which covers the big red area of Fig. 3 with a horizontal resolution of 4km and 80 vertical layers with up to 24km height. More details on the model setup can be found here[7].

[6] http://www.cosmo-model.org/content/tasks/priorityProjects/c2i/iconLog/default.htm
[7] http://www.cosmo-model.org/content/tasks/operational/default.htm

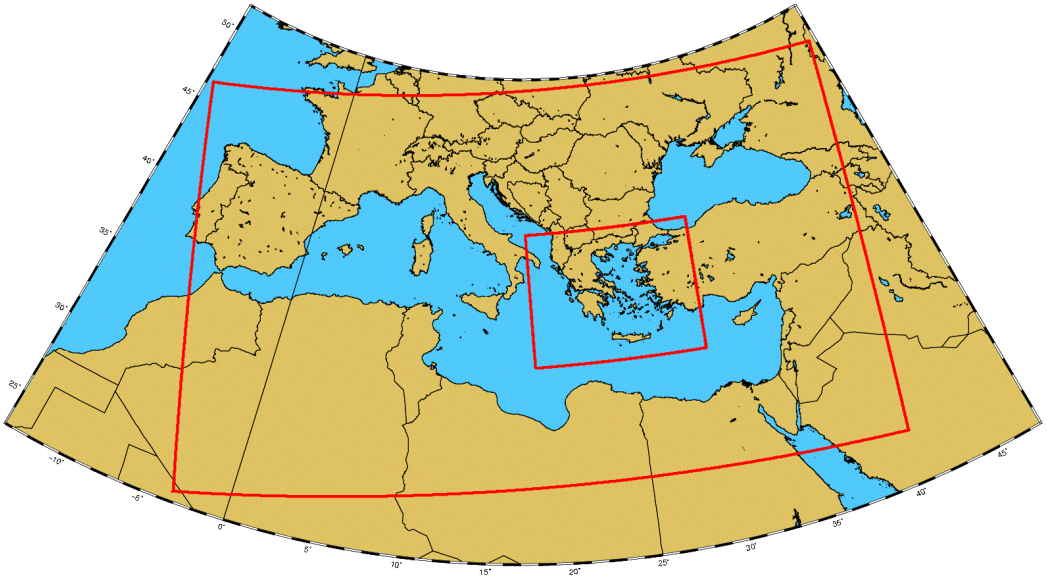


Figure 3: COSMO-GR4 domain (larger domain, the small domain is COSMO-GR1).

Both models use initial/boundary conditions from IFS-ECMWF, and run twice a day (00h, 12h) for a forecast horizon of 72 hrs. No data assimilation is performed.

3.2 Continuous Parameters verification

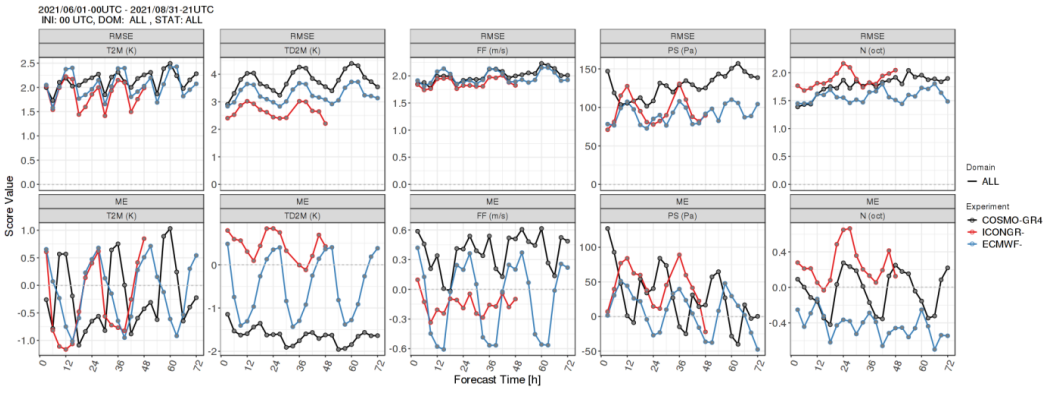


Figure 4: RMSE (top), ME (bottom) for (from left to right) T2m, Td2m. WS10m, PS, TCC for JJA.

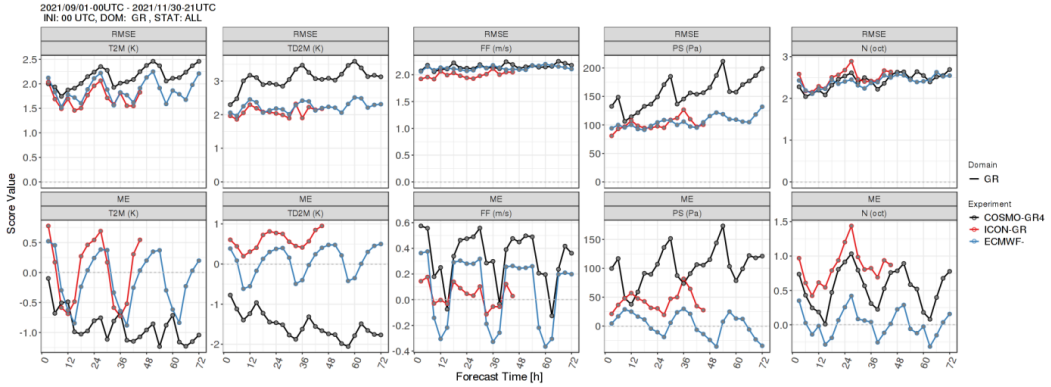


Figure 5: RMSE (top), ME (bottom) for (from left to right) T2m, Td2m. WS10m, PS, TCC for SON.

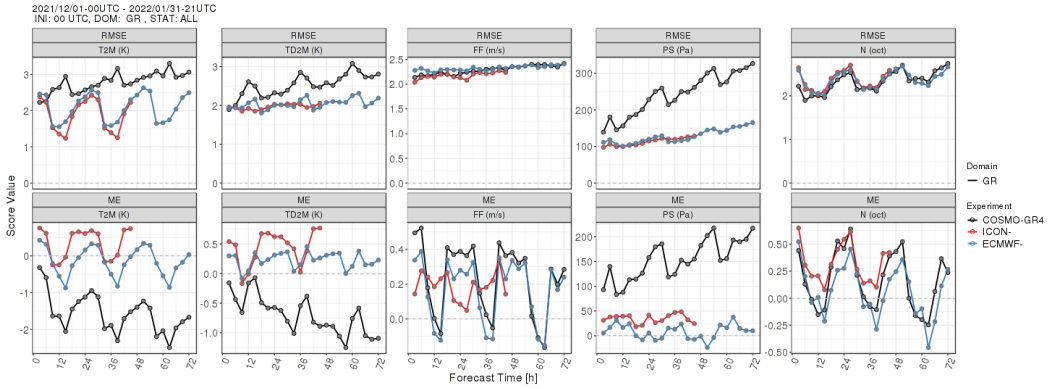


Figure 6: RMSE (top), ME (bottom) for (from left to right) T2m, Td2m. WS10m, PS, TCC for DJF.

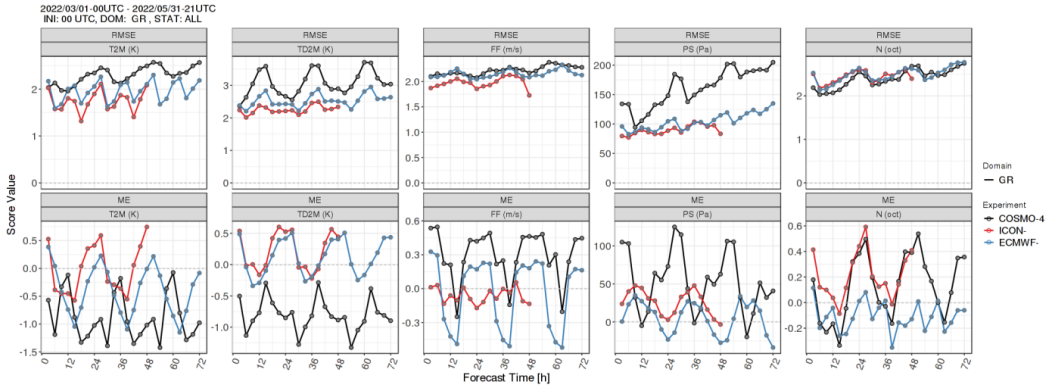


Figure 7: RMSE (top), ME (bottom) for (from left to right) T2m, Td2m. WS10m, PS, TCC for MAM.

2m Temperature: In JJA (Fig. 4), all models exhibit a sharp daily variation in BIAS, with ICON-GR2.5 and IFS similar cycle with slight overestimation at night (less than 1K), and slight underestimation during the warm hours of the day (up to 1K). COSMO-GR4

BIAS diurnal cycle is shifted, with positive values during the daytime. For the remaining seasons, COSMO-GR4 is constantly underpredicting with negative BIAS values which are increasing with forecast time in SON (Fig. 5) and DJF (Fig. 6) seasons. ICON-GR2.5 and IFS-ECMWF BIAS values exhibit a similar daily variation with positive values during night time, which are slightly higher for ICON-GR2.5. For RMSE, ICON-GR2.5 diurnal cycle is similar to IFS-ECMWF one for all seasons, with ICON-GR 2.5 values being slightly lower. Although more irregular, the daily RMSE variation of all models is more consistent in JJA while in the remaining seasons, the difference among COSMO-GR4 (with the highest values) and the models ICON-GR2.5 and IFS-ECMWF is bigger, especially in DJF season.

2m Dew Point Temperature: COSMO-GR4 constantly underestimates Td2m and negative BIAS is increasing with forecast time for all seasons, especially in SON (Fig. 5) and DJF (Fig. 6) seasons. ICON-GR2.5 and IFS-ECMWF behavior is completely different from that of COSMO-GR4, with a clear diurnal variation for both of them. However, ICON-GR2.5 BIAS values remain positive for all the forecast horizon, with slightly higher values at night in all seasons, while IFS-ECMWF clearly underestimates Td2m in the warm hours of the day for JJA and SON, while BIAS remains positive only for DJF period. RMSE values for both ICON-GR2.5 and IFS-ECMWF are lower than for COSMO-GR4 model and their diurnal cycle is weaker. ICON-GR2.5 RMSE values are lower than for IFS-ECMWF in JJA and MAM seasons, and almost identical for the two other seasons.

Wind Speed at 10m (m/sec): ICON-GR2.5 BIAS daily variation is significantly weaker than for the two other models for all seasons, with a tendency of constant underestimation for JJA (Fig. 4) and MAM (Fig. 7) and slight overestimation for SON and DJF. Both IFS-ECMWF and mainly COSMO-GR4, overestimate nighttime winds. RMSE values are comparable (about 2m/sec) for all models, with ICON-GR2.5 values being slightly improved for all seasons.

Surface Pressure (Pa): ICON-GR2.5 BIAS cycle is comparable to IFS-ECMWF but remains slightly positive for all times. On the other hand, COSMO-GR4 pressure overestimation increases with forecast time especially in SON (Fig. 5) and DJF (Fig. 6). An increase of RMSE values for COSMO-GR4 model with forecast time, with higher values at night is distinct for all seasons, especially for SON, DJF and MAM (Fig. 7) seasons. ICON-GR2.5 and IFS-ECMWF RMSE values are similar and lower than COSMO-GR4 ones.

Total Cloud Cover: The TCC BIAS diurnal cycle for both COSMO-GR4 and ICON-GR2.5 is sharper than that of IFS-ECMWF for all seasons except for DJF when all three models exhibit similar variability. TCC ICON-GR2.5 BIAS for nighttime is positive and even higher than COSMO-GR4 for JJA (Fig. 4) and SON (Fig. 5) seasons. The cycle is weaker for IFS-ECMWF which generally underestimates TCC, except for noon when the values are only slightly greater than zero. The BIAS difference among models is weaker in DJF season when all models diurnal cycle is comparable. RMSE values among models do not significantly differ for all seasons producing higher scores at night especially in DJF. However, it is worth noting that in JJA, ICON-GR2.5 score, especially at night, is worse than the score of the two other models.

3.3 Categorical scores for 6h precipitation.

6h Precipitation statistical scores are compared for all seasons for thresholds of 0.2, 1, 5 and 10mm (Figures 8 – 11). For JJA season, due to the lack of significant precipitation amounts, 10mm threshold is excluded. The results indicate that ETS index is slightly higher for ICON-GR2.5 especially for the lower thresholds. However, in DJF season, the difference

among models for that score is minimal. Regarding the FBI score, all models tend to overestimate low precipitation events, and underestimate events of higher precipitation. The added value of ICON-GR2.5 is noticeable mainly in JJA when IFS-ECMWF and COSMO-GR4 produce high FBI values for low precipitation events in warm hours of the day, while ICON-GR2.5 scores are closer to 1. In the remaining seasons, the FBI scores for ICON-GR2.5 and COSMO-GR4 are comparable. The POD/FAR score improved performance for ICON-GR2.5 is mostly seen in warm seasons (JJA and SON).

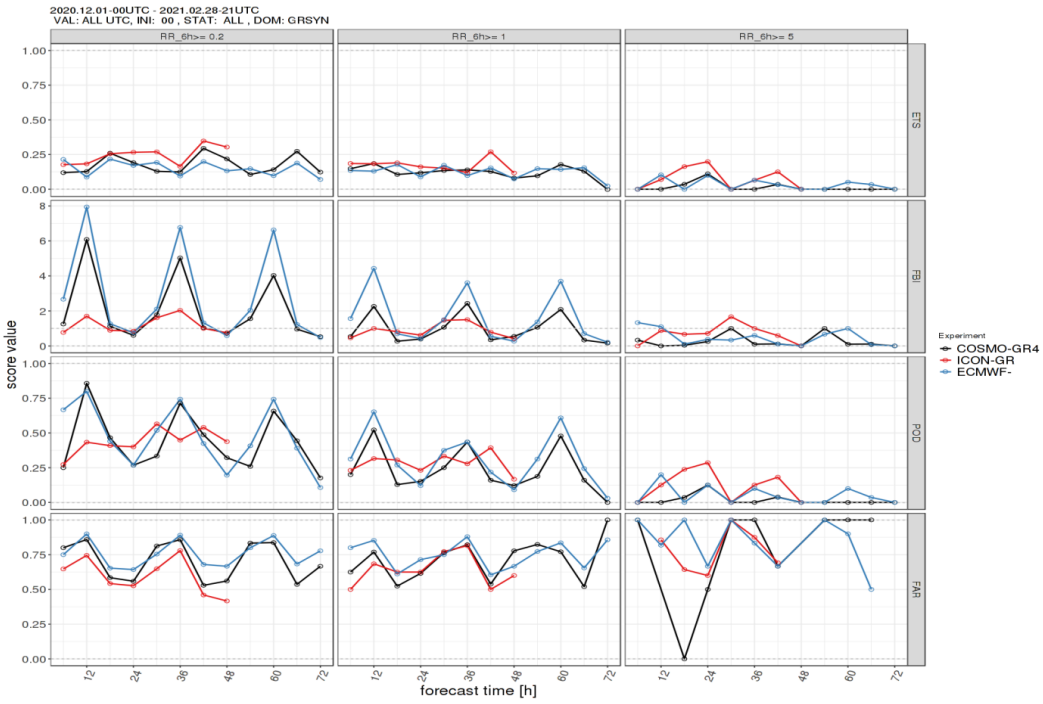


Figure 8: 6h precipitation scores: ETS, FBI, POD, FAR (from top to bottom) for thresholds of 0.2mm,1mm, 5mm (from left to right) for JJA. Please note that the number of cases is small for the highest threshold.

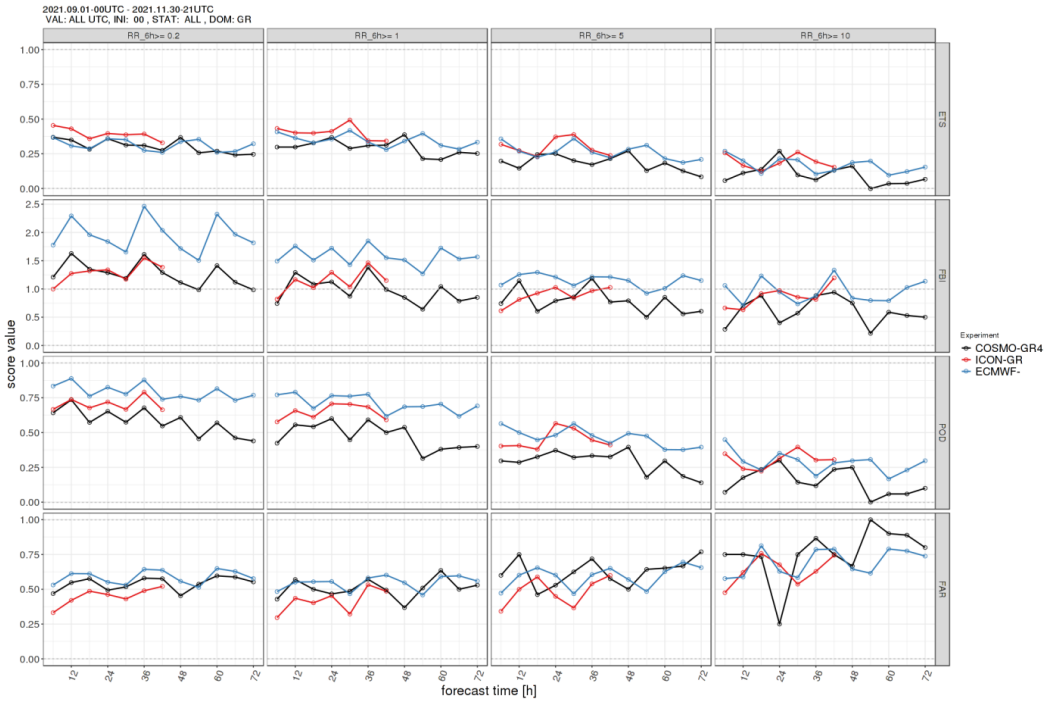


Figure 9: 6h precipitation scores: ETS, FBI, POD, FAR (from to to bottom) for thresholds of 0.2mm,1mm, 5mm, 10mm (from left to right) for SON. Please note that the number of cases is small for the highest threshold.

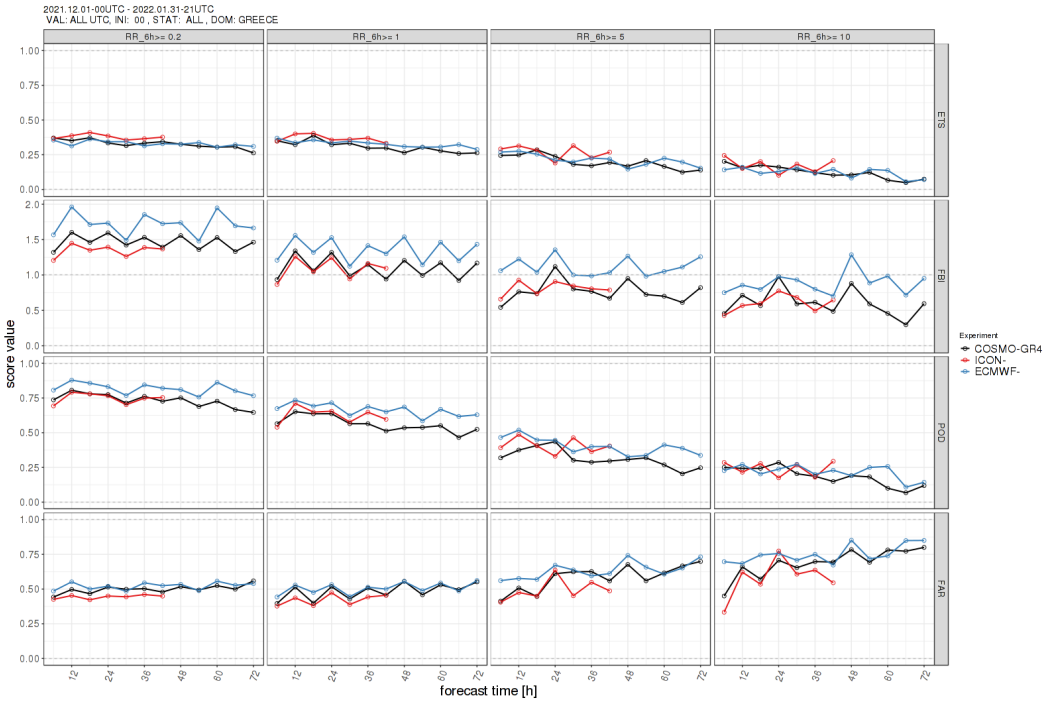


Figure 10: 6h precipitation scores: ETS, FBI, POD, FAR (from to to bottom) for thresholds of 0.2mm,1mm, 5mm, 10mm (from left to right) for DJF. Please note that the number of cases is small for the highest threshold.

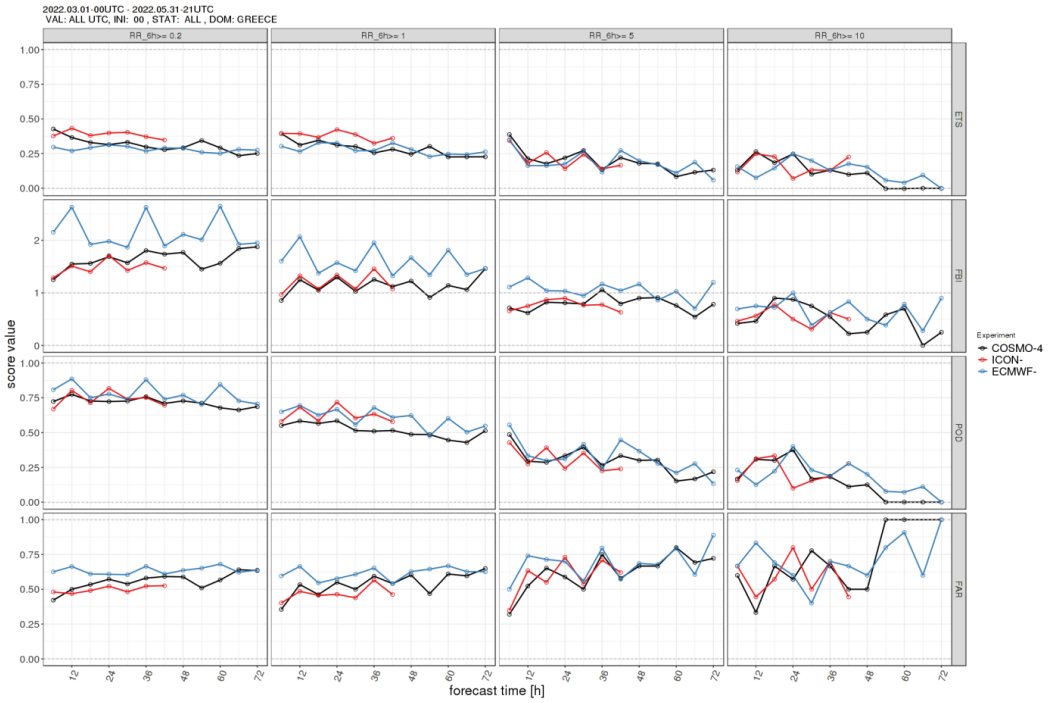


Figure 11: 6h precipitation scores: ETS, FBI, POD, FAR (from to to bottom) for thresholds of 0.2mm,1mm, 5mm, 10mm (from left to right) for MAM. Please note that the number of cases is small for the highest threshold.

3.4 Conclusions

The verification analysis results suggest an overall good performance for ICON-GR2.5 over the Greek territory for the period of analysis (June 2021-May 2022), for most continuous parameters, when compared with COSMO-GR4 and IFS-ECMWF models. The diurnal cycle score pattern of ICON-GR2.5 is mostly similar to IFS-ECMWF, and its performance is mainly improved compared to that of COSMO-GR4. Only for TCC the scores of ICON-GR2.5 are slightly worse than COSMO-GR4 model in warm seasons only. An improved performance in 6h precipitation forecasts of ICON-GR2.5 is mainly seen during the warm seasons and especially for low precipitation thresholds. ICON-GR2.5 and COSMO-GR4 require comparable computational resources despite the higher resolution of ICON. ICON-GR2.5 is expected to run fully operationally in 2023 when it becomes time critical in the machines of ECMWF.

4 Results for Israel

P. Khain(1), H. Muskatel(1), A. Shtivelman(1), A. Baharad(1), L. Uzan(1) and Y. Levi(1)

1. *Israel Meteorological Service*

We report the verification results of ICON (ICON-IL2.5) for 2021, and compare them to IFS and COSMO (CO-IL2.5) models. The verified ICON version 2.6.2.2, driven by the IFS, includes the radiation scheme ecRad, assimilation of Israel Meteorological Service (IMS) radar and OPERA via the latent heat nudging, a correction for the cloud cover parametrization during summertime and extensive retuning of model parameters for the Eastern Mediterranean (EM). Generally, ICON-IL2.5 shows better results compared to IFS and CO-IL2.5. On December 2022 a major upgrade is planned, which includes the use of forecasted CAMS aerosols on polluted days, the use of forecasted SST from IFS, and soil initialization from the global ICON model. Finally, we report that the ICON model runtime is shorter than the COSMO model with a similar configuration.

4.1 Overview and setup of ICON-IL2.5

Since June 2020 ICON-IL2.5 is running as 'Time Critical Suite' on ECMWF HPC. ICON-IL2.5 is the LAM version of ICON (Zängl et al., 2015) covering the domain 4°E-45.5°E/25.5°N-53°N (Fig. 12). The model has a resolution of 2.5km (R2B10), forecast range of 90h, initial and boundary conditions of IFS, and cold starts initialization times of 00 and 12 UTC.

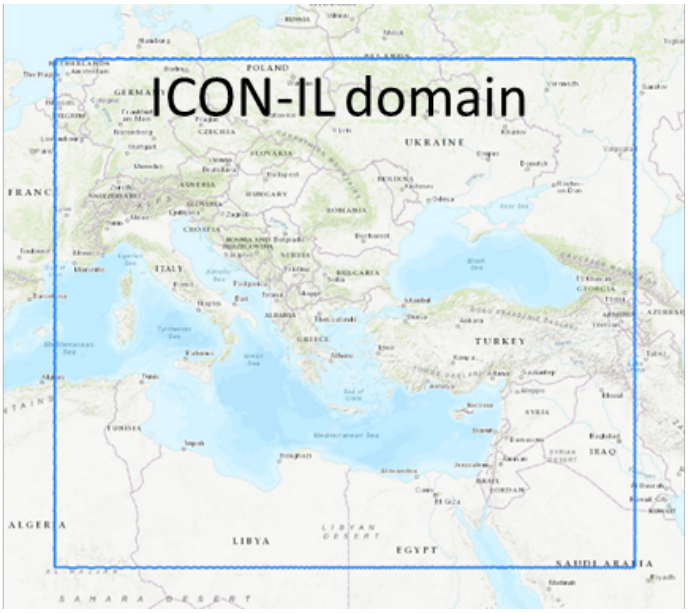


Figure 12: ICON-IL2.5 domain.

On January 2021 ICON-IL2.5 was upgraded to the version 2.6.2.2 (without the cp/cv bug), which includes the radiation scheme ecRad, assimilation of the IMS radar and OPERA, EUMETNET's Operational Programme for the Exchange of Weather Radar Information, via the latent heat nudging, and a correction for the cloud cover parametrization during summertime. Moreover, we performed an extensive tuning of model parameters for the EM,

resulting in 2 optimal configuration: 'Tuned winter', operationally used from October to April, and 'Tuned summer' - from May to September. Table 2 presents the configuration of the reference (ICON-D2) together with 'Tuned winter' and 'Tuned summer' ICON-IL2.5 configuration.

Table 2: Configuration of ICON-IL2.5 from October to April (Tuned winter) and from May to September (Tuned summer).

Forecasted field	**Parameter**	**Reference (ICON-D2)**	**Tuned winter**	**Tuned summer**
Precipitation	`lshallowconv_only`	False	True	True
Precipitation	`lgrayzone_deepconv`	True	False	False
Precipitation	`tune_rdepths`	20000	10000	20000
2m Temperature	`rlam_heat`	10	10	20
2m Temperature	`tkhmin`	0.5	0.5	1
2m Temperature	`tkmmin`	0.75	0.75	1
2m relative humidity	`c_soil`	1.25	2.0	1.25
10m wind speed	`tune_gkwake`	0.25	0.5	0.9
Low cloud cover	`tune_box_liq_asy`	4	4	1.7
Low cloud cover	`allow_overcast`	1	1	0.63

The motivation for the tuning presented in tab. 2 is described below, and the following abbreviations are used: 2m temperature (T2), 2m relative humidity (RH2), 10m wind speed (WS10), low cloud cover (CLCL).

- **Precipitation:** With the combination `lshallowconv_only=False` / `lgrayzone_deepconv=True` which is recommended in ICON, the precipitation over the EM is smooth, similarly to the global models, and is generally of low quality. This combination allows subgrid vertical mixing in grid columns with grid scale precipitation (active grid columns), which dries the boundary layer. Therefore the combination `lshallowconv_only=True` / `lgrayzone_deepconv=False` was chosen, which switches off subgrid vertical mixing in active grid columns. In addition, `tune_rdepths` parameter (default 20000 Pa) defines the maximum vertical extension of the shallow convection (SC), above which the SC scheme is switched off. Reduction of `tune_rdepths` weakens the effect of the SC scheme and intensifies weak grid scale precipitation (Khain et al., 2021). Since winter time weak precipitation was underestimated, `tune_rdepths` was reduced to 10000 Pa in 'Tuned winter' configuration. SC scheme is important for correct description of the vertical profiles of temperature and humidity in the boundary layer. Since there is usually no precipitation over EM on summer, `tune_rdepths` was kept as 20000 Pa in the 'Tuned summer' configuration.
- **T2:** After setting `lshallowconv_only=True` / `lgrayzone_deepconv=False`, during winter T2 was forecasted reasonably well, while during summer daytime T2 was overestimated. `rlam_heat` parameter is the laminar resistance for heat transfer from surface to the air. During day the surface is usually hotter than the air, and during night - colder. Therefore increasing `rlam_heat` will decrease T2m during day and increase T2m during night. Note that in order not to influence evaporation from sea if one multiplies `rlam_heat` by x, `rat_sea` should be divided by x. In order to reduce the daytime overestimation of T2 during summer, `rlam_heat` was increased from 10 to 20 in the 'Tuned summer' configuration, and consequently `rat_sea` was reduced from 0.8 (reference value) to 0.4 in this version. After this setting, the nighttime T2 during summer was underestimated. `tkhmin`/`tkmmin` parameters do not allow the night

turbulence to drop too much (which would cause T2 to cool too much just like the surface). Increasing `tkhmin`/`tkmmin` increases the nighttime T2, not influencing daytime T2. Consequently, `tkhmin`/`tkmmin` were increased to 1/1 in the 'Tuned summer' configuration.

- **RH2:** After improving T2 with `rlam_heat` and `tkhmin` parameters, we could tune the `c_soil` parameter, which corresponds to the amount of evaporating grid points, and increasing it leads to the increase of RH2. During winter, nighttime RH2 was underestimated, therefore `c_soil` was increased to 2.0 in 'Tuned winter' configuration. During summer, nighttime RH2 was initially overestimated. However, the increased `tkhmin`/`tkmmin` yielded the increase of nightime T2 and reduction of RH2. Therefore there was no need to reduce `c_soil` in the 'Tuned summer' configuration.

- **WS10:** In the reference ICON configuration daytime WS10 is strongly overestimated over the EM. This overestimation is especially pronounced during summer. In order to reduce it we increased the subgrid scale drag via `tune_gkwake` parameter to 0.5 in 'Tuned winter' configuration and to 0.9 in 'Tuned summer' configuration. Unfortunately, this increase yields an underestimation of the nighttime WS10. Therefore, the above values were chosen so that the averaged diurnal bias is close to zero, while some daytime overestimation (around 0.5 m/s) still exists in ICON-IL2.5.

- **CLCL:** Over the EM, the summertime CLCL is overestimated. There are typically wide smooth areas of CLCL>0 almost without overcast (CLCL=1). In the framework of `inwp_cldcover`=1 scheme, a new parameter `allow_overcast` was introduced. In order to correct the summertime CLCL forecasts, `allow_overcast` was set to 0.63 and `tune_box_liq_asy` was set to 1.7 in the 'Tuned summer' configuration.

4.2 Verification

Since November 2020 we produce the feedback files enabling the usage of RFDBK verification software. However, in order to compare ICON-IL2.5 with CO-IL2.5 and IFS, this report presents the verification for the entire year 2021 using local verification tools developed at the IMS. Figure 13 presents verification for T2, RH2, WS10 and 10m wind direction (WD10) (panels a-d, respectively). At each panel the verification of ICON-IL2.5, CO-IL2.5, and IFS models for the year 2021 is presented (blue bars). To estimate the yearly variability, the verification for 2019 and 2020 is also presented (red and orange bars, respectively), but only for IFS and CO-IL2.5. For each forecasted field, verification was performed separately for 4 groups. The groups were 54 inland and 27 coastal automatic weather stations over Israel, each verified separately for the night (00 UTC) and day (12 UTC) times during the first 72h of the forecast. The final RMSE score used in panels a-c was obtained by averaging the RMSE scores of the 4 groups. The exception is the 10m wind direction verification (panel d). Here, the model and observation wind vectors were used to calculate the RMSE of the vector of difference (RMSEwv), minus RMSE of the wind speed (RMSEws), resulting in score in units of m/s. One can see that ICON-IL2.5 shows the best results in T2 (1.5 K), RH2 (8.7%), and WD10 (0.6 m/s). For WS10, ICON-IL2.5 is better than CO-IL2.5 but is similar to IFS (1.4 m/s). Although ICON-IL2.5 WS10 forecasts were improved by tuning the `tune_gkwake` parameter (Table 2), they still suffer from overestimation during day and underestimation during night, which, most probably, can not be solved by tuning the SSO scheme.

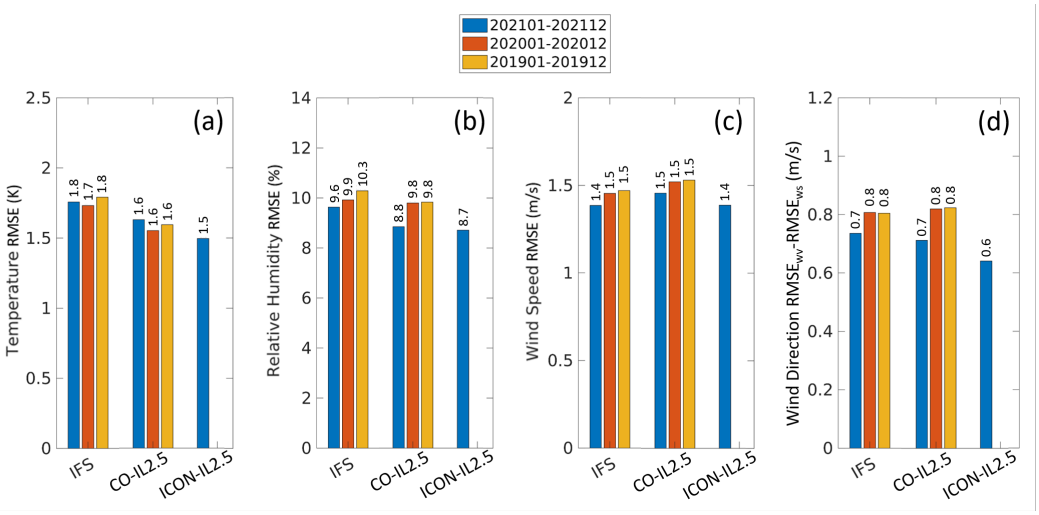


Figure 13: RMSE-type verification for 2m temperature, 2m relative humidity, 10m wind speed and 10m wind direction (panels a-d, respectively). At each panel the verification of ICON-IL2.5, CO-IL2.5, and IFS models for the year 2021 is presented (blue bars). The verification for 2019 and 2020 is also presented (red and orange bars, respectively), but only for IFS and CO-IL2.5.

Similar analysis was performed to obtain the percentage of hits (Fig. 14). A forecast was accounted as hit in case its absolute error was below a predefined threshold. The thresholds for T2, RH2, WS10 and WD10 were defined as 2K, 13%, 4m/s (updated to 2m/s on July 2021) and 30°, respectively.

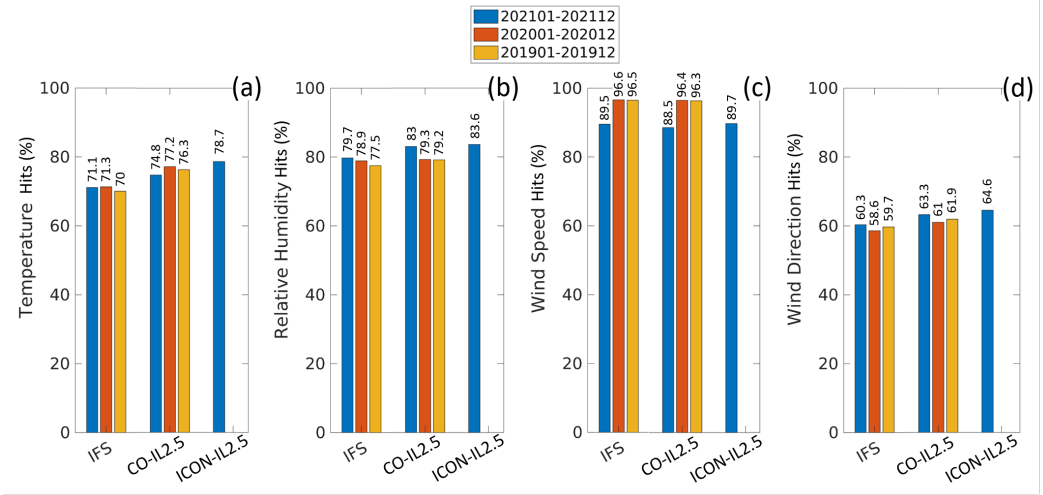


Figure 14: Hits percentage for 2m temperature, 2m relative humidity, 10m wind speed and 10m wind direction (panels a-d, respectively).Hits percentage for 2m temperature, 2m relative humidity, 10m wind speed and 10m wind direction (panels a-d, respectively).

Figure 14 shows a clear advantage of ICON-IL2.5 for T2 (78.7% of hits) over both IFS and CO-IL2.5. ICON-IL2.5 shows clear advantage over IFS and a slight advantage over CO-IL2.5 for RH2 (83.6% of hits) and WD10 (64.6% of hits). As in fig. 13, for WS10, ICON-IL2.5

shows similar result to IFS. Since the threshold for WS10 was updated on July 2021 from 4m/s to 2m/s, the hits percentage for 2021 is lower than for 2019 and 2020.

Figure 15 presents the verification of precipitation. To overcome the precipitation spatial and temporal variability, the Fractional Skill Score (FSS) neighbourhood method was used (Roberts and Lean, 2008). For observations, the radar-rain gauge composite was used (Khain et al., 2020). For every 6 hours accumulated precipitation (6hAP) forecasted and observed map, FSS was calculated for 20 km smoothing radius and thresholds of 0.01, 0.1, 0.5, 1, 2, 5, 10, 20 mm per 6h. The FSSs were averaged over these thresholds, with double weight for the more important thresholds of 0.5, 1, 2, 5 mm per 6h. These averaged FSS scores were calculated for IFS, CO-IL2.5, and ICON-IL2.5. For a given threshold and 6hAP map, if the spatial coverage fraction in both observations and model was below 1%, the FSS was ignored during the averaging.

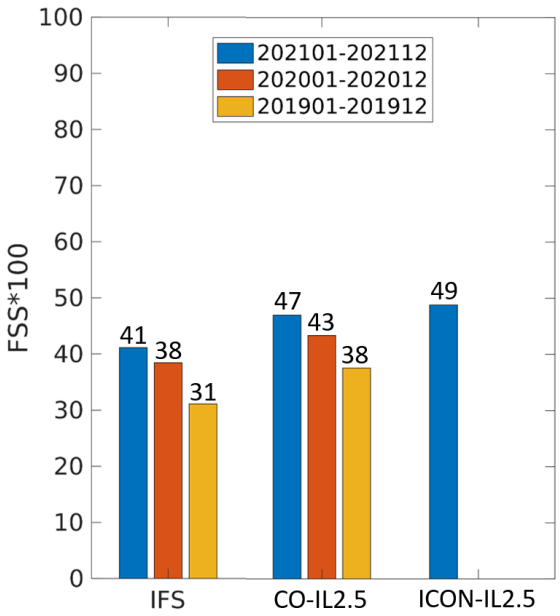


Figure 15: Precipitation verification of ICON-IL2.5, CO-IL2.5, and IFS models using an averaged over thresholds FSS with the spatial radius of 20km. The verification for the year 2021 is presented by blue bars. The verification for 2019 and 2020 is also presented (red and orange bars, respectively), but only for IFS and CO-IL2.5.

ICON-IL2.5 precipitation forecasts show slight advantage over CO-IL2.5 and more significant advantage over IFS. Deeper analysis (not presented) shows strong dependence of the forecasts quality on the synoptic conditions. Over the EM, the rain events during autumn and spring are characterized by sporadic localized convective precipitation with low predictability. At these conditions, the high-resolution models CO-IL2.5 and ICON-IL2.5 tend to overestimate the accumulated precipitation, often miss the correct location and do not have a clear advantage over the smoothed precipitation fields of IFS. Important to mention, that this overestimation is not resolved by tuning the shallow convection scheme (tab 2) or with the stochastic shallow convection scheme in ICON-IL2.5. In contrast to autumn and spring, during winter the EM precipitation usually originates from organized frontal cyclones, where both ICON-IL2.5 and CO-IL2.5 are significantly better than IFS.

4.3 Plans for the next upgrade of ICON-IL2.5

On December 2022 a major update is planned, which includes the use of forecasted CAMS aerosols on polluted days, the use of forecasted SST from IFS, and soil initialization from the global ICON model. In the following we discuss the use of forecasted CAMS aerosols in more details.

Operational weather forecasting is in constant conflict between the desire for accuracy and the goal of delivering results in a timely manner. Computational resources are always limited, so usually expensive fully-coupled aerosol models, such as those provided by the EU-funded Copernicus Atmosphere Monitoring Service (CAMS) run by ECMWF and ICON-ART, are replaced by a monthly aerosol climatology. On the other hand, radiation fluxes and cloud microphysics are critically affected by aerosol content. Solar radiation is scattered and absorbed by aerosols and can be significantly reduced in a highly polluted environment. Liquid and ice formation in clouds is also influenced by hydrophilic and hydrophobic aerosol content in the atmosphere. As part of the T2RC2 and the CAIIR priority projects, IMS has conducted research which proposes an intermediate solution to these issues by coupling predicted aerosol fields taken from the CAMS aerosol model with the ICON model. The idea is that at each time step ICON will use the CAMS predicted aerosols, interpolated in time and space, without the need to advect them in space. Therefore, the runtime stays almost identical to the aerosol climatology setup. In a first stage ecRad was coupled with 3D predicted aerosol mixing ratios of CAMS and the ice nucleation scheme (Muskatel et al., 2021). In the next step, we plan to implement the coupling of these aerosols with the water droplet activation scheme in the ICON model. The method was tested for the entire year 2020. Figure 16 shows the global radiation verification using 17 automatic radiation stations over Israel during the year 2020. Twelve relatively polluted days, i.e. days when the averaged PM2.5 measurements over Israel were more than three times higher than normal, were chosen for verification. The verification was performed during clear-sky conditions, where cloud cover both in models and observations was below 30%. Figure 16 compares the mean diurnal cycle of the global radiation in observations, the IFS, CO-IL2.5, ICON-IL2.5 and the new version ICON-IL2.5-CAMS with CAMS aerosol coupling. The first 24 hours of each simulation were verified. For each model, the mean bias and RMSE are presented.

One can see that in ICON-IL2.5-CAMS the global radiation on polluted days was substantially improved. The 47.6 Wm^{-2} bias in ICON-IL2.5 was essentially removed, and the RMSE was reduced from 91.5 Wm^{-2} to 68.9 Wm^{-2}. Preliminary results show that during non-polluted days the global radiation improvement is less significant and the other fields verification scores are similar to ICON-IL2.5. The operational implementation of CAMS aerosols coupling in ICON-IL2.5 is planned for December 2022.

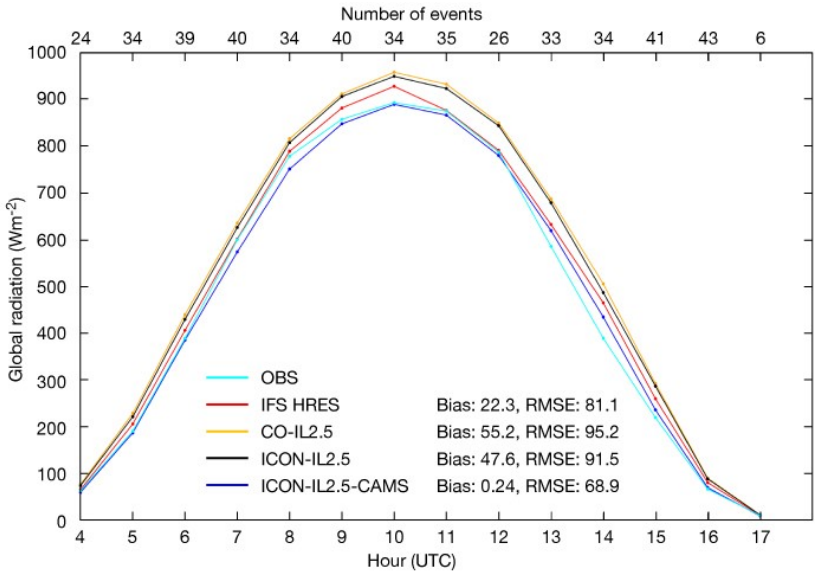


Figure 16: Global radiation verification over Israel for 12 clear-sky polluted days during the year 2020. The figure compares the mean diurnal cycle of the global radiation in observations (17 automatic radiation stations), the IFS HRES, CO-IL2.5, ICON-IL2.5 and the new version ICON-IL2.5-CAMS. The first 24 hours of each simulation were verified. For each model, the mean bias and RMSE are presented.

4.4 Runtime comparison of ICON and COSMO

To compare the runtime of ICON and COSMO models, several test runs were performed with both models over the domain 25°E-39°E/26°N-36°N, with the resolution of 0.025° (~2.5 km) for COSMO and 2.466 km (R2B10) for ICON. The runs were performed on the ECMWF CCA HPC using 10 computing nodes or 360 cores. For both models the runtime of the first 24h of forecast was around 14 minutes and 30 seconds. However, there are several factors which show that ICON is more efficient than COSMO. First, the number of vertical levels was 60 in COSMO and 65 in ICON. Second, the number of grid points over the domain was 224961 in COSMO and 244241 in ICON and, i.e. 1.086 times more. Finally, the COSMO model was compiled in single precision, ICON in mixed precision (single and double). Taken these factors into account, the COSMO model in single precision with the same configuration as ICON, would run ~1.18 times slower than ICON in mixed precision.

4.5 Conclusion

We reported the verification results of ICON-IL2.5 for the year 2021 in comparison with IFS and CO-IL2.5 models. The verified ICON version 2.6.2.2, driven by the IFS, includes the radiation scheme ecRad, assimilation of IMS radar and OPERA via the latent heat nudging, a correction for the cloud cover during summertime and extensive retuning of model parameters for the Eastern Mediterranean. Generally, ICON-IL2.5 shows better results compared to IFS and CO-IL2.5 for 2m temperature, 2m relative humidity and 10m wind direction. For 10m wind speed ICON-IL2.5 is slightly better than CO-IL2.5 but of similar quality to IFS. For precipitation, ICON-IL and COSMO-IL are of similar quality and are generally better than IFS. On December 2022 a major upgrade is planned, which includes the use of forecasted CAMS aerosols on polluted days, the use of forecasted SST from IFS, and soil initialization

from the global ICON model. Finally, the ICON and COSMO models runtime was compared, showing that ICON is significantly more efficient than COSMO.

Acknowledgements

We gratefully acknowledge the extraordinary support of the colleagues at DWD and especially of Daniel Rieger in configuring and developing the ICON model at IMS.

5 Results from Italy: CNMCA and ARPA Piemonte

F. Marcucci(1), F. Sudati(1), N. Zaccariello(1), V. Garbero(2), N. Vela(2) and E. Oberto(2)

1. *CNMCA, National Center for Aerospace Meteorology and Climatology*
2. *Arpa Piemonte - Regional Agency for the Protection of the Environment of Piedmont*

The ICON-IT model performance has been evaluated, in terms of standard Mean Error and Root Mean Square Error indexes, using SYNOP and TEMP observations over Italian domain for the period 1st June 2021 – 31st May 2022, compared to that of the operational COSMO-IT model. For the standard verification, the MEC-Rfdbk software package was used. Moreover, a fuzzy verification has been performed for the same period using the precipitation estimated by the Italian national radar composite of the Department of Civil Protection corrected by the rain-gauges. Performance diagrams have been performed using observations provided by the national meteorological networks. The new ICON-IT model system, operationally since July 2020, has been designed similarly to COSMO-IT with the domain covering Italy, boundary conditions from IFS-HRES and 1h full data assimilation cycle based on KENDA-LETKF. Results are encouraging as generally ICON-IT outperforms COSMO-IT for surface parameters and temperature profiles. Concerning the precipitation, the fuzzy results do not show significant differences between ICON-IT and COSMO-IT. Both models tend to overestimate the maximum precipitation values for medium-high thresholds.

5.1 Overview and Setup of ICON-IT

Since July 2020 Italian Air Force Meteorological Service is running operationally the high-resolution ICON-IT model on ECMWF-HPC as a time critical suite. The convection-resolving model ICON-IT domain covers Italy and its surroundings with a grid spacing of about 2.1 km and 65 vertical layers.

During the reporting period (1st June 2021 -31st May 2022) the operational setup was characterized by grid R19B7, radiation scheme 'ECRAD' and settings suggested by DWD (ICON-D2 operational namelist) except for parameter `box_liq` increased from 0.04 to 0.08 and parameter `box_liq_asy` decreased from 4 to 2. The model version is a development version from January 2021 and it has not been updated up to now because unresolved communication problems under cray environment at ECMWF.

The data assimilation algorithm used to estimate the initial state of ICON-IT is the LETKF (Local Ensemble Transform Kalman Filter) scheme developed by COSMO (KENDA), that is the same used for COSMO-IT. KENDA system (Schraff et al., 2016) is configured to provide analyses at a 1-hourly interval using the following dataset: radiosonde ascent and descent profiles, AMDAR and Mode-S aircraft data, wind profiler data, observations from surface stations, Meteosat AMV, Metop scatterometer winds, NOAA/Metop AMSUA/MHS and NPP/NOAA ATMS radiances.

An Incremental Analysis Update (IAU) technique without prior hydrostatic balancing is applied for the 1h updates of the data assimilation and model system to remove spurious gravity waves. The model integration starts from the previous 55-minute forecast state, and the analysis increments are added to the model fields incrementally at each model time step during the initial 10-minute time window and then the model run continues as a free forecast. Boundary conditions for the assimilation cycle are from '6h-old' IFS-EPS members.

ICON-IT long run is initialized by the high resolution KENDA-LETKF deterministic analysis and driven by '6h-old' IFS-HRES fields. The model is integrated up to 48 hour twice a day at 00 and 12 UTC.

5.2 Standard verification

Traditional forecast verification scores against SYNOP and TEMP observations (respectively surface and upper level data) are computed. The results are based on COSMO-IT and ICON-IT forecasts in the period 1st June 2021 – 31st May 2022. Due to the loss of archived data, forecast skills have been evaluated up to 48 hours only for the last quarter (MAM2022).

Surface verifications (Figs. 17 – 21) show a worsening for the JJA 2021 period and improvement of T2m forecast skill for ICON-IT with respect to COSMO-IT (reduction of T2m bias and RMSE) in the SON 2021, DJF 2021-22 and MAM 2022 quarters; regarding the RH2m forecast skill, there is a slight general improvement for ICON-IT model. The performance of ICON-IT shows a clear improvement for the total cloud cover parameter for all seasons. The 10m wind speed RMSE score is almost identical for the two models, while the ME score is slightly better for ICON-IT in the step range from +09 to +15 for JJA 2021, but worsening during DJF 2021-22 and MAM 2022. As regards the surface pressure parameter both ME and RMSE scores are clear better for ICON-IT model during DJF 2021-22 season; the ME score shows also values closer to zero for ICON-IT in SON 2021 and MAM 2022 quarters, while the RMSE score in the same period is comparable.

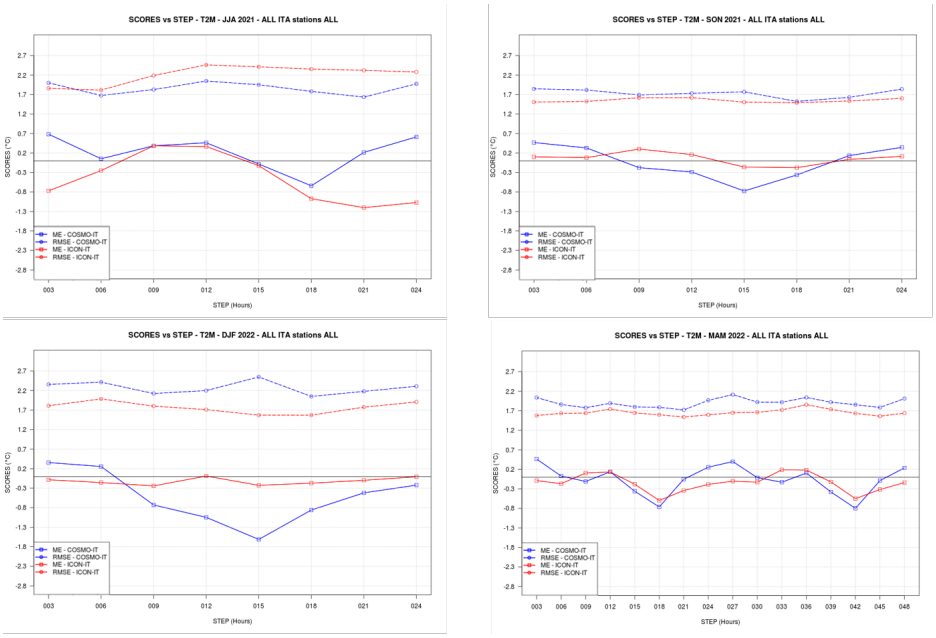


Figure 17: Mean Error (ME) and Root Mean Square Error (RMSE) scores, as a function of forecast step, for **2m temperature**, computed for ICON-IT (red) and COSMO-IT (blue) forecasts using *Italian synop* observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel).

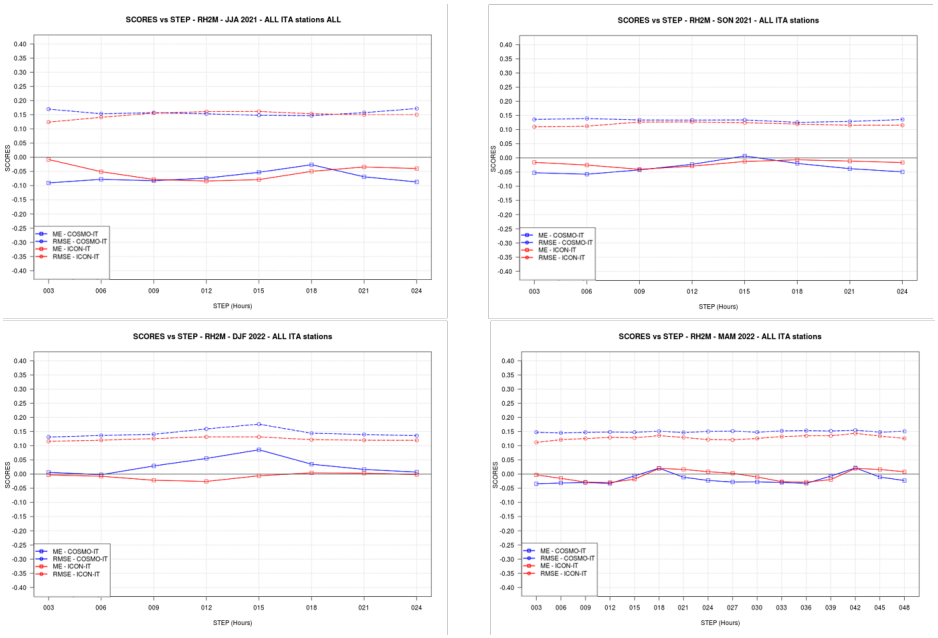


Figure 18: Mean Error (ME) and Root Mean Square Error (RMSE) scores, as a function of forecast step, for **2m relative humidity**, computed for ICON-IT (red) and COSMO-IT (blue) forecasts using *Italian synop* observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel).

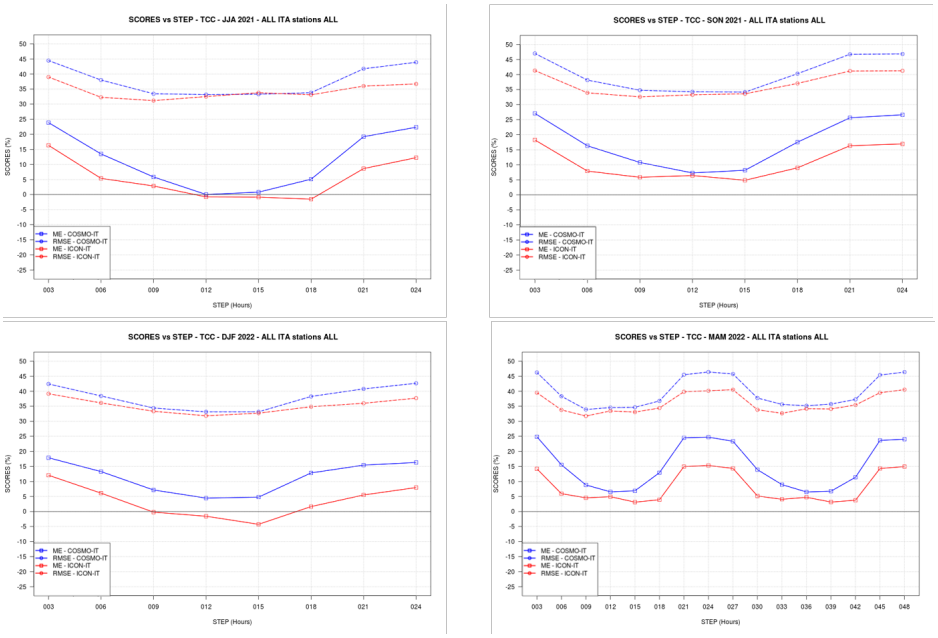


Figure 19: Mean Error (ME) and Root Mean Square Error (RMSE) scores, as a function of forecast step, for **total cloud cover**, computed for ICON-IT (red) and COSMO-IT (blue) forecasts using *Italian synop* observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel). SON is missing because of a problem with archived data.

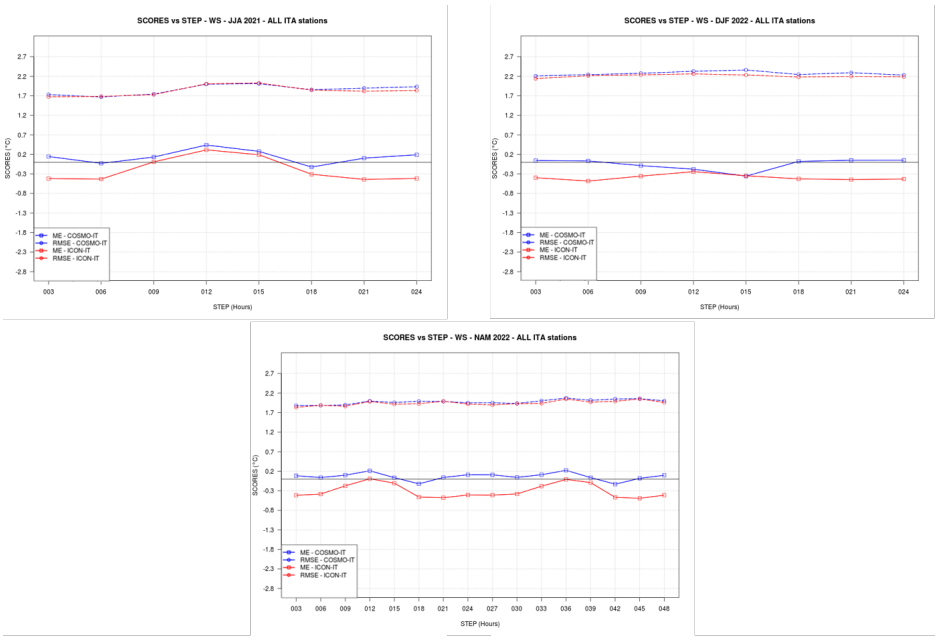


Figure 20: Mean Error (ME) and Root Mean Square Error (RMSE) scores, as a function of forecast step, for **10m wind speed**, computed for ICON-IT (red) and COSMO-IT (blue) forecasts using *Italian synop* observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel).

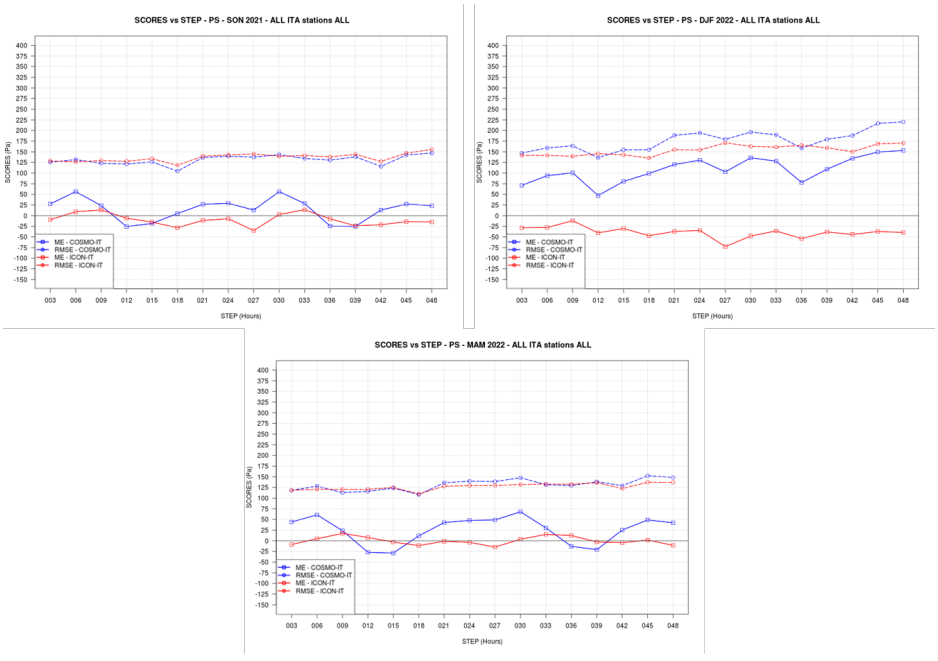


Figure 21: Mean Error (ME) and Root Mean Square Error (RMSE) scores, as a function of forecast step, for **surface pressure**, computed for ICON-IT (red) and COSMO-IT (blue) forecasts using *Italian synop* observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel, results are omitted for quarter JJA 2021 due to the loss of archived data).

Due to the loss of archived data the upper-air forecast verification scores against TEMP data (radiosonde observations) have been computed only for the MAM2022 period (Figs. 22— 24).

As regard temperature scores, ICON-IT outperforms COSMO-IT, because a general reduction of RMSE over the whole column for all forecast steps, and a better performance in terms of ME especially for T+12 and T +36.

Moreover looking at wind scores, ICON-IT shows a reduction in RMSE in the lower levels against an increase in ME in the lower-middle column.

The two models are almost comparable in terms of relative humidity, even if the ICON-IT bias is reduced at T+12h and T+36h.

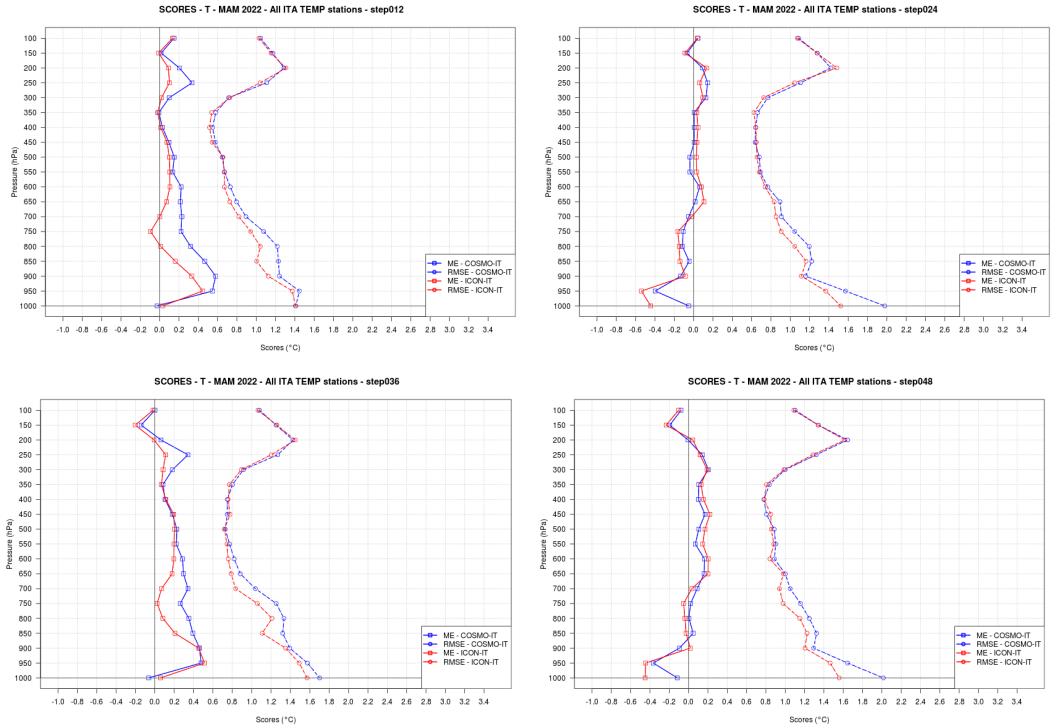


Figure 22: Mean Error (ME) and Root Mean Square Error (RMSE) vertical profiles of **temperature** for different forecast steps (T+12h, T+24h, T+36h, T+48h), computed for ICON-IT (red) and COSMO-IT (blue) forecasts using radiosonde observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel, results are omitted for quarters JJA 2021, SON 2021, DJF 2022).

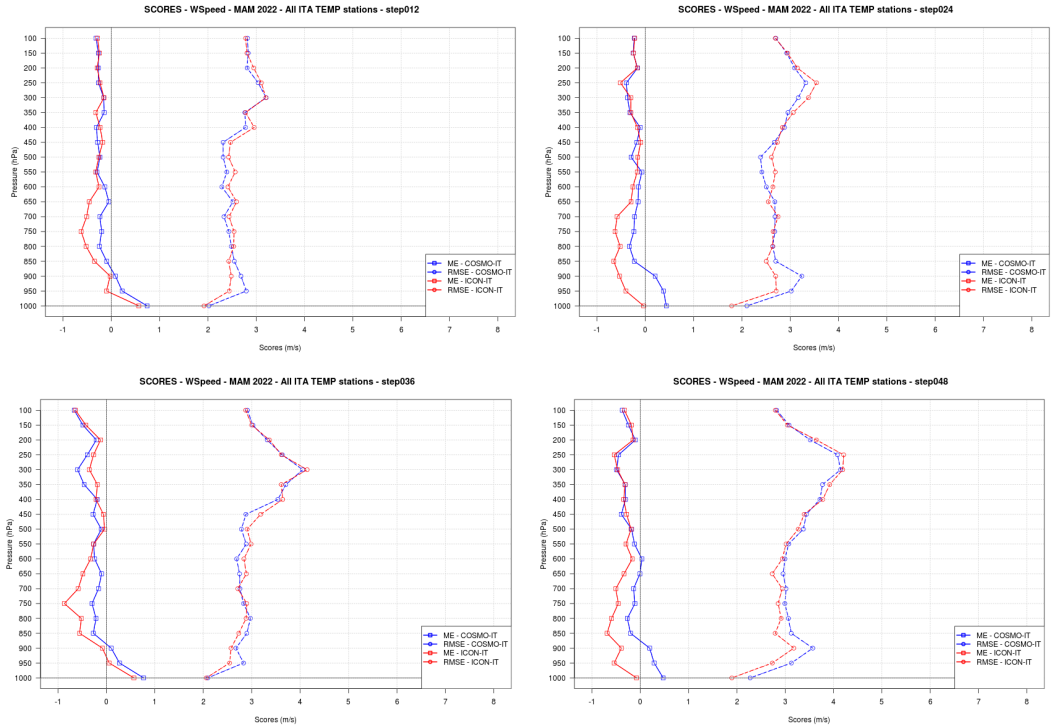


Figure 23: Mean Error (ME) and Root Mean Square Error (RMSE) vertical profiles of **wind speed** for different forecast steps (T+12h, T+24h, T+36h, T+48h), computed for ICON-IT (red) and COSMO-IT (blue) forecasts using radiosonde observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel, results are omitted for quarters JJA 2021, SON 2021, DJF 2022).

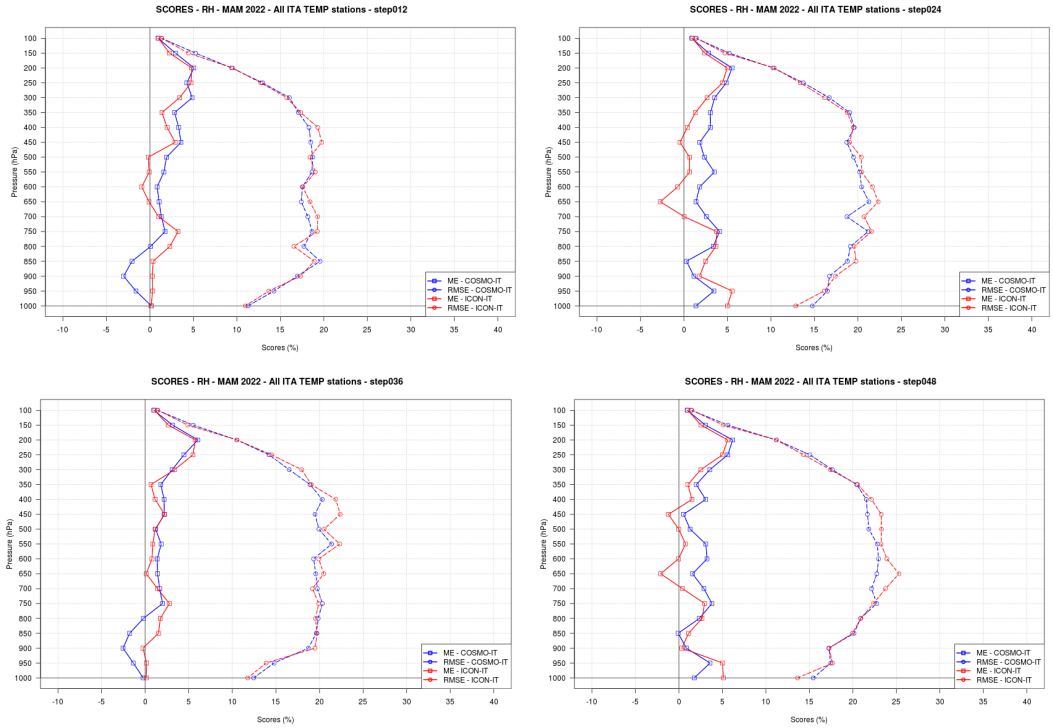


Figure 24: Mean Error (ME) and Root Mean Square Error (RMSE) vertical profiles of **relative humidity** for different forecast steps (T+12h, T+24h, T+36h, T+48h), computed for ICON-IT (red) and COSMO-IT (blue) forecasts using radiosonde observations in the period 1st June 2021 – 31st May 2022 (seasonal base for each panel, results are omitted for quarters JJA 2021, SON 2021, DJF 2022).

5.3 Fuzzy Verification

In the following, the fuzzy method is used to compare precipitation forecasts with rain estimation provided by the Italian national radar composite of the Department of Civil Protection and corrected with rain-gauges. COSMO-IT and ICON-IT forecasts are evaluated for the period 1st June 2021 – 31st May 2022, distinguishing the seasons JJA, SON, DJF and MAM.

The main score summarising the fuzzy verification results is the Fractions Skill Score (FSS). The FSS directly compares the forecast and the observation (radar) 3-hours precipitation field on a portion of territory by increasing the spatial dimension on which the verification is carried out and by varying the precipitation threshold. The FSS is then plotted for different spatial scales and precipitation thresholds. If we add the variable 'time' by comparing the observation field also with the previous and the next 3-hour forecast field, a 3D fuzzy verification is performed. The value of FSS above which the forecast is considered useful, that is better than random, is FSSuseful. The smallest scale for which FSS $\geq$ FSSuseful is considered the 'useful scale' and the value of FSS is indicated in bold.

Figs. 25 – 28 show the FSS calculated for the first 24 hours of forecast (D0) by comparing the same 3-hour interval (2D fuzzy verification) and the previous and the next 3-hour interval (3D fuzzy verification).

The scores are generally very low if we consider JJA period (Fig. 25), since the useful scale for

0.1 mm/3h threshold is reached at 37 km and at 143 km for 12 mm/3h. A fairly significant improvement is obtained by the 3D fuzzy, but the scores continue to be worse than those calculated for the other periods, particularly for the low-medium thresholds. This is due to the increased uncertainty in predicting the convective phenomena in both space and time typical of the summer period.

As shown in Fig. 26, the scores related to the fall period (SON) improve over the summer ones, as convective phenomena are less frequent than in summer. The 3D fuzzy gives better results than 2D confirming the difficulty of locating events even in time. COSMO-IT behaves better than ICON-IT for high-threshold, as it reaches a useful scale at 72 km for 15 mm/3h, while ICON-IT shows better scores at very low threshold, as it reaches a useful scale at 2.2 km for 0.2 mm/3h.

The winter period (DJF) shows the better scores than any other seasons (Fig. 27), since the precipitation is mainly due to advective phenomena and is more predictable both in space and in time; For this reason, the 3D fuzzy gives slightly better results than 2D. COSMO-IT and ICON-IT has a similar behaviour for low and medium thresholds, reaching a useful scale at 2.2 km for 1 mm/3h and at 6.6 km for 5 mm/3h, while ICON-IT performs better for very high thresholds as it reaches a useful scale at 143 km for 15 mm/3h.

Concerning MAM period (Fig. 28), the two models are very similar, with COSMO-IT showing slightly better scores than ICON-IT for very high thresholds, by reaching a useful scale at the largest scale up to 10 mm/3h. We notice quite significant improvement concerning the 3D fuzzy both for ICON-IT and COSMO-IT, which means that the models have difficulty in locating the event not only spatially but also temporally.

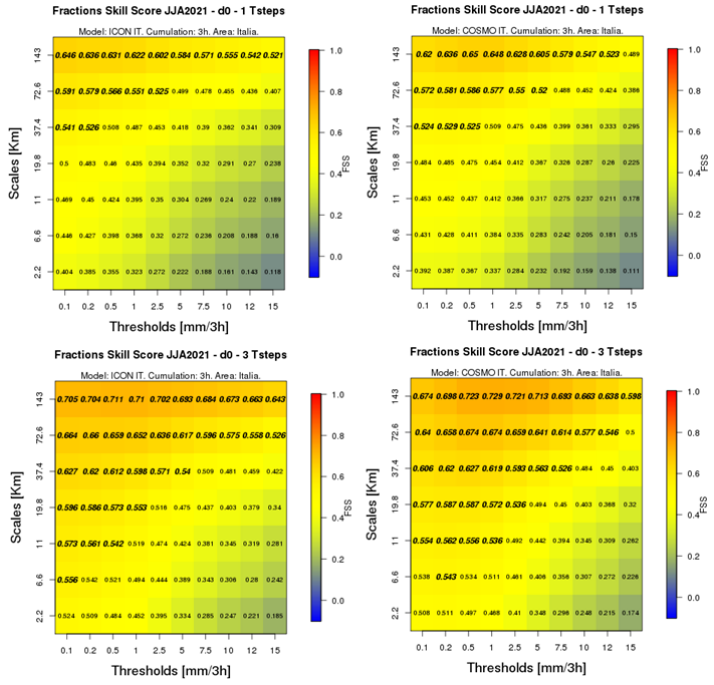


Figure 25: FSS calculated for the first 24 hours of forecast (D0) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **JJA2021**.

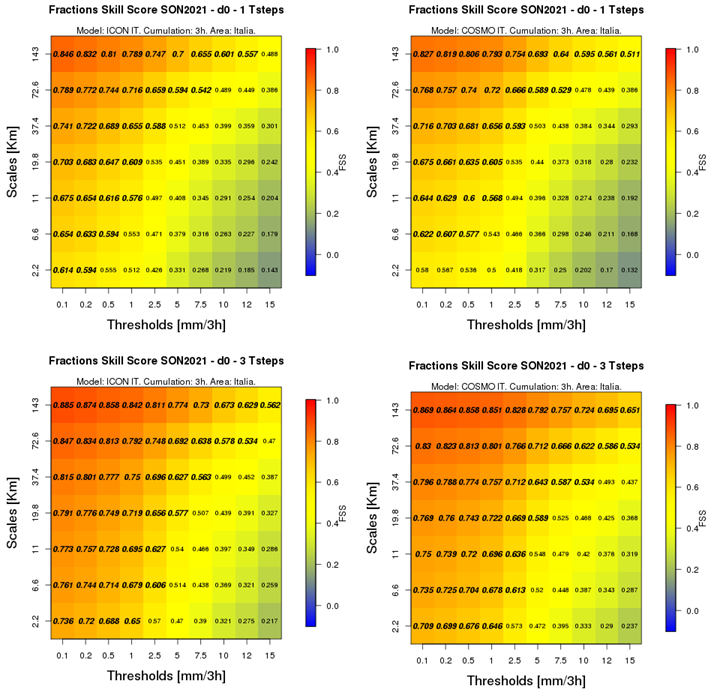


Figure 26: FSS calculated for the first 24 hours of forecast (D0) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **SON2021**.

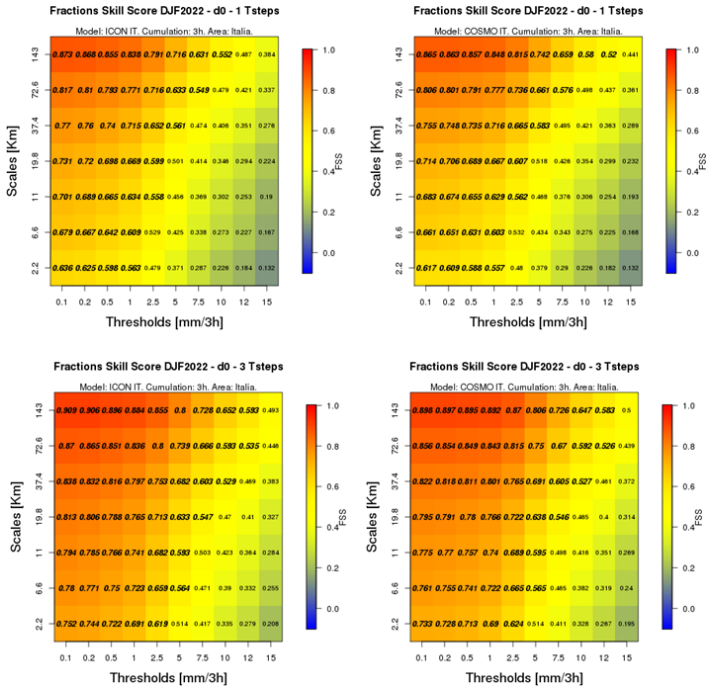


Figure 27: FSS calculated for the first 24 hours of forecast (D0) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **DJF2022**.

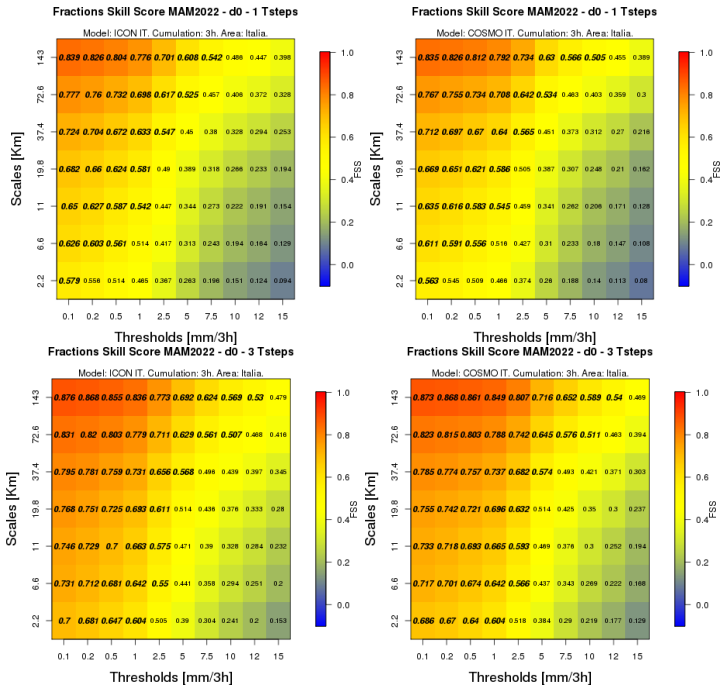


Figure 28: FSS calculated for the first 24 hours of forecast (D0) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **MAM2022**.

Figs. 29 – 32 show the seasonal FSS calculated for the second 24 hours of forecast (D1) by comparing the same 3-hour interval (2D fuzzy verification) and the previous and the next 3-hour interval (3D fuzzy verification).

The scores for each period get worse than those calculated for the first 24 hours of forecast (D0) of the same period, as expected. Another common behaviour is that the 3D fuzzy provides better scores than 2D, since the models often have trouble to correctly place events in time both in the first and second 24 hours of forecast.

The scores related to the summer period (Fig. 29) are definitely the worst, as for D0, and this behaviour is due to the uncertainty associated with the convective phenomena, which is amplified in the second 24 hours of forecast. ICON-IT deteriorates more than COSMO for low and medium thresholds.

The scores concerning the fall period (Fig. 30) improve over the summer ones. The two models have generally similar performances, even if ICON-IT has better scores for low threshold, as it reaches a useful scale at 2.2 km for 1 mm/3h, but worst scores for very high thresholds, since it doesn't reach any useful scale for 12 and 15 mm/3h threshold whereas COSMO-IT does at 143 km.

In winter the performance of both the models are similar and pretty good, as shown in Fig. 31. The degradation related to the second 24 hours of forecast and to the 2D/3D fuzzy technique is less pronounced than in the other periods, due to the mainly advective character of the phenomena.

Concerning MAM (Fig. 32), the models have similar performances, except for medium-high threshold where ICON-IT reaches a useful scale at 143 km up to 7.5 mm/3h, whereas COSMO-IT stops at 5 mm/3h. Also in the 3D verification ICON-IT performs better for high thresholds.

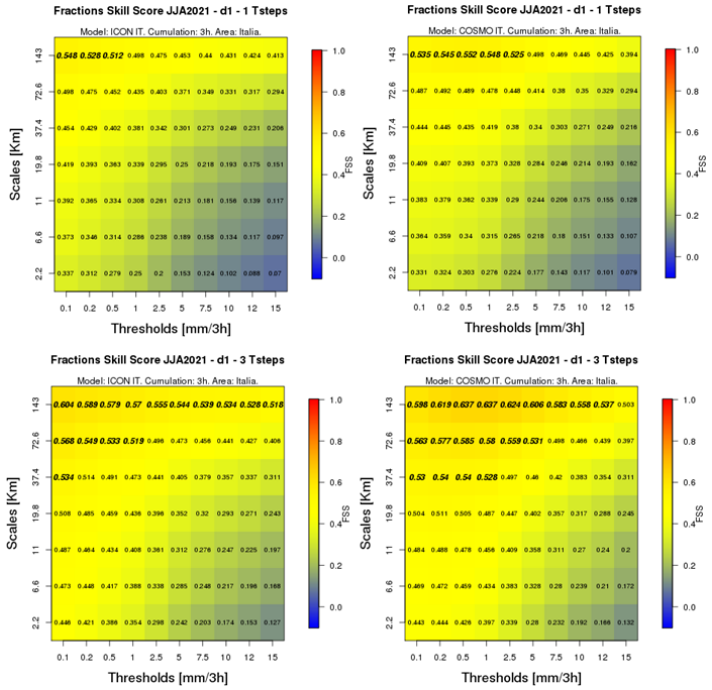


Figure 29: FSS calculated for the second 24 hours of forecast (D1) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **JJA2021**.

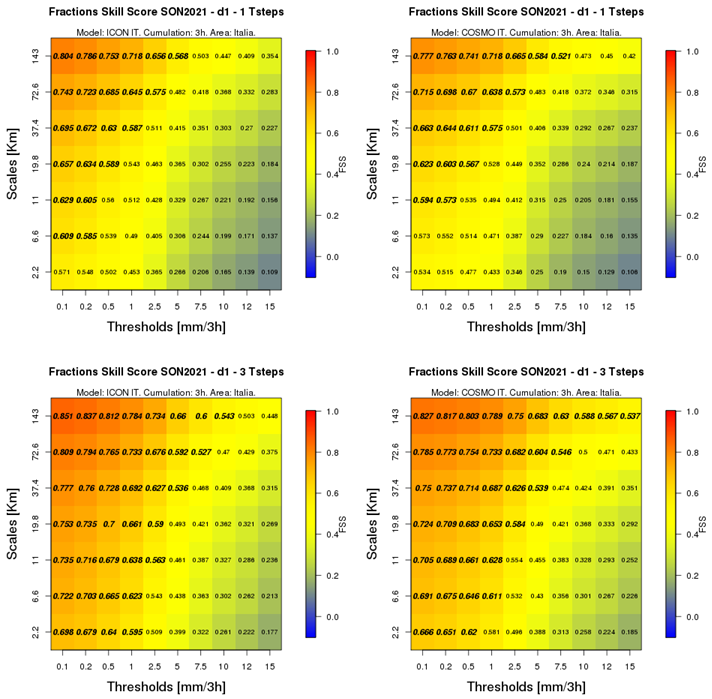


Figure 30: FSS calculated for the second 24 hours of forecast (D1) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **SON2021**.

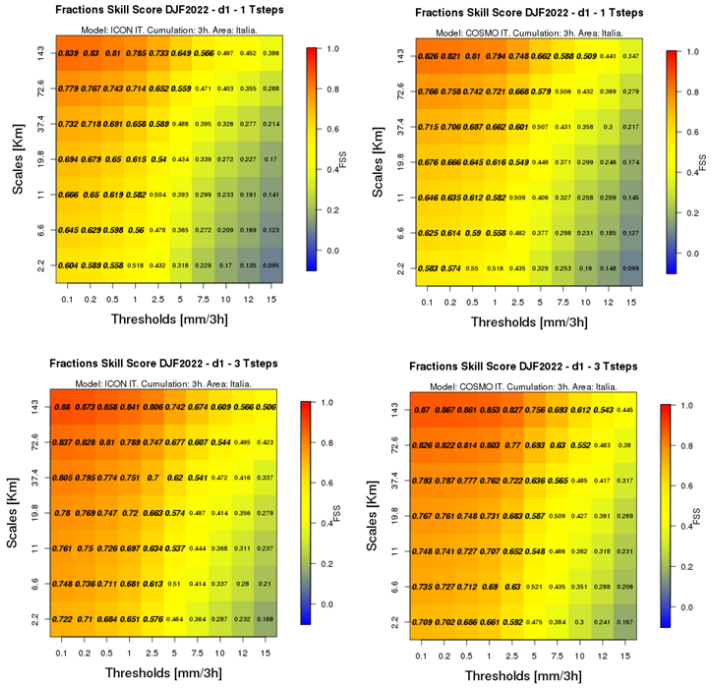


Figure 31: FSS calculated for the second 24 hours of forecast (D1) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **DJF2022**.

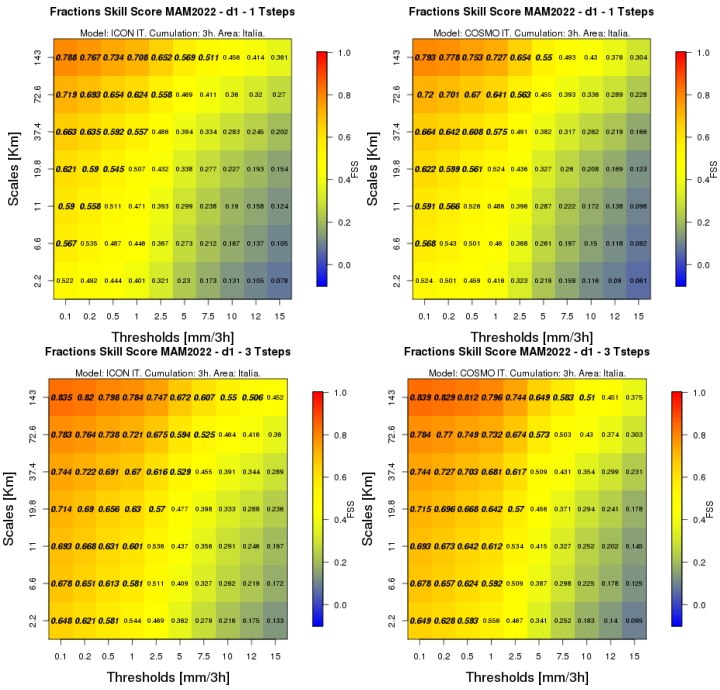
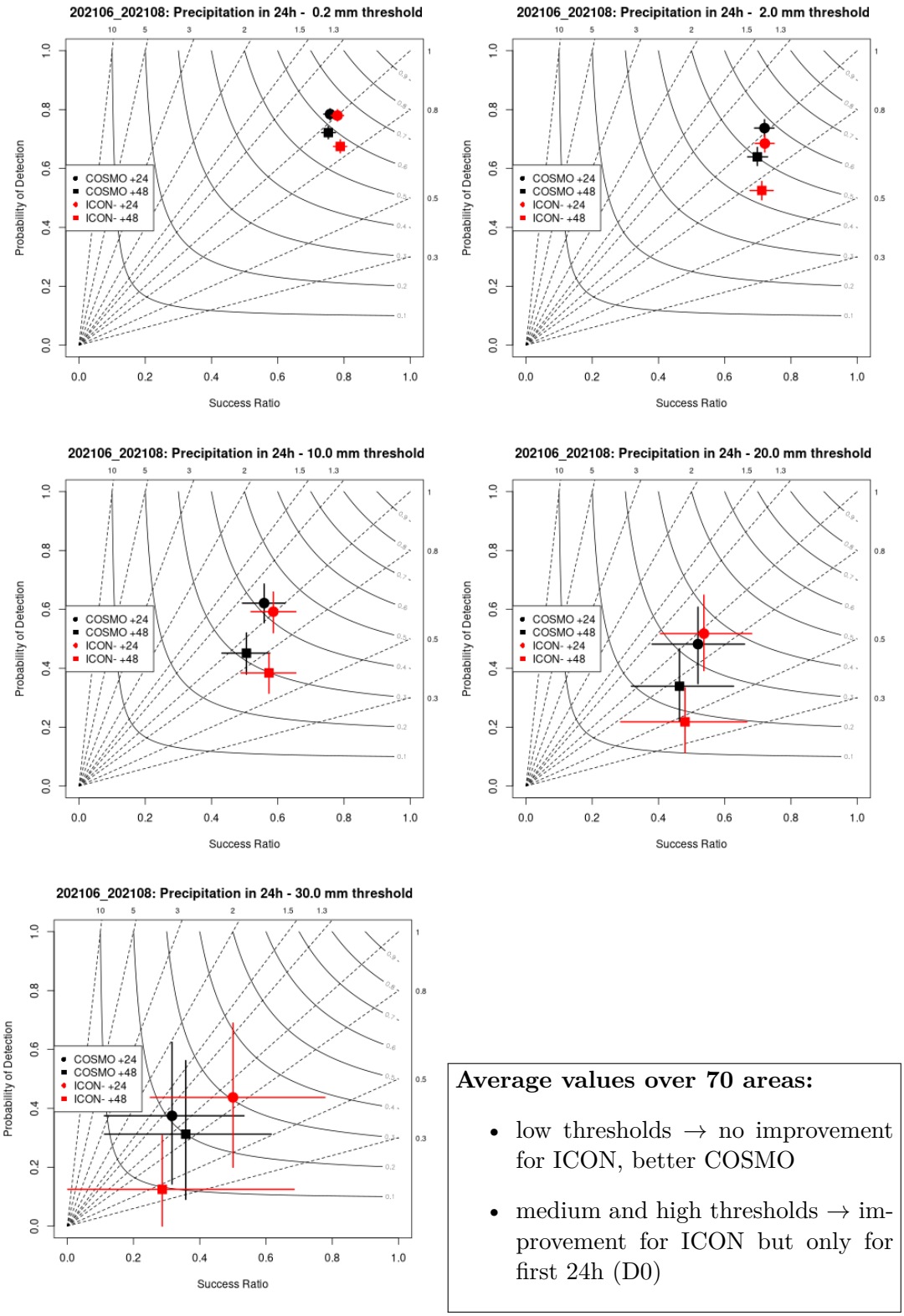


Figure 32: FSS calculated for the second 24 hours of forecast (D1) – 2D and 3D fuzzy verification: ICON-IT on the right and COSMO-IT on the left for **MAM2021**.

5.4 Performance diagrams

In order to evaluate the performance differences between COSMO-IT and ICON-IT using standard indices and standard verification techniques, we opted to plot seasonal performance diagrams, that condense the information derived by 4 classical statistical indices (POD, success ratio (SR) or 1-FAR, BIAS and critical success index CSI or TS) into unique points on the diagram. So, the 'best forecast' is closest to the top right corner. In this study we used observations provided by the rain-gauges network of Civil Protection Department and we compared the 24h cumulated precipitation average values as well as the 24h maximum values, calculated over 70 homogeneous Italian areas both for observation and forecast. The diagrams have been performed for 4 seasons (JJA 2021, SON 2021, DJF 2022 and MAM 2022), for increasing thresholds (0.2 mm, 2 mm, 10 mm, 30 mm), and are shown in Figs. 33 – 40.

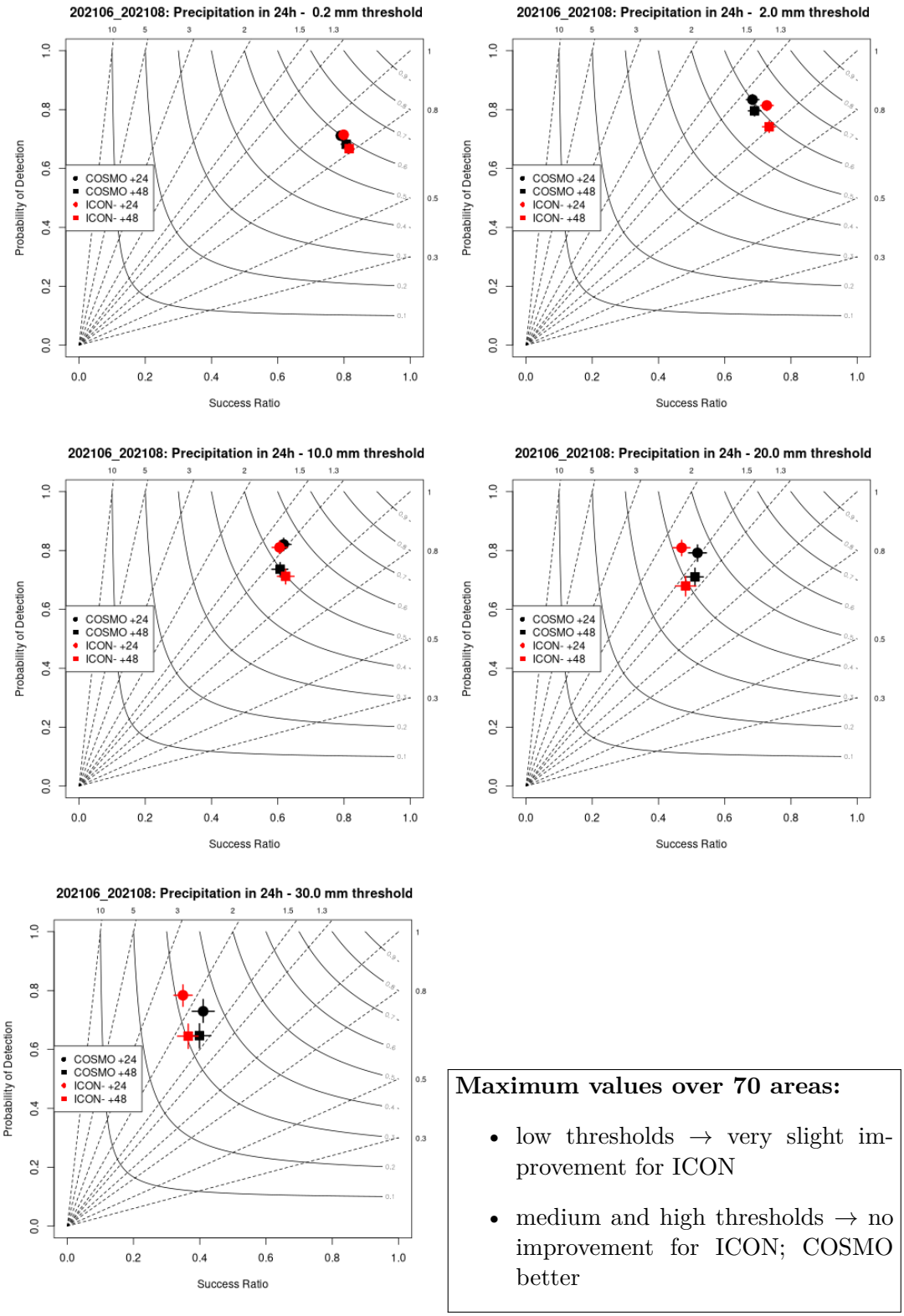
To sum up all the results, during last summer we obtained no significant improvement for ICON except for the high thresholds for first 24h (average values). During last autumn and very dry winter we did not obtain any improvement for ICON. Finally, during this spring we have reached a good skill for ICON, even if an overestimation in terms of maximum values is observed.



Average values over 70 areas:

- low thresholds → no improvement for ICON, better COSMO
- medium and high thresholds → improvement for ICON but only for first 24h (D0)

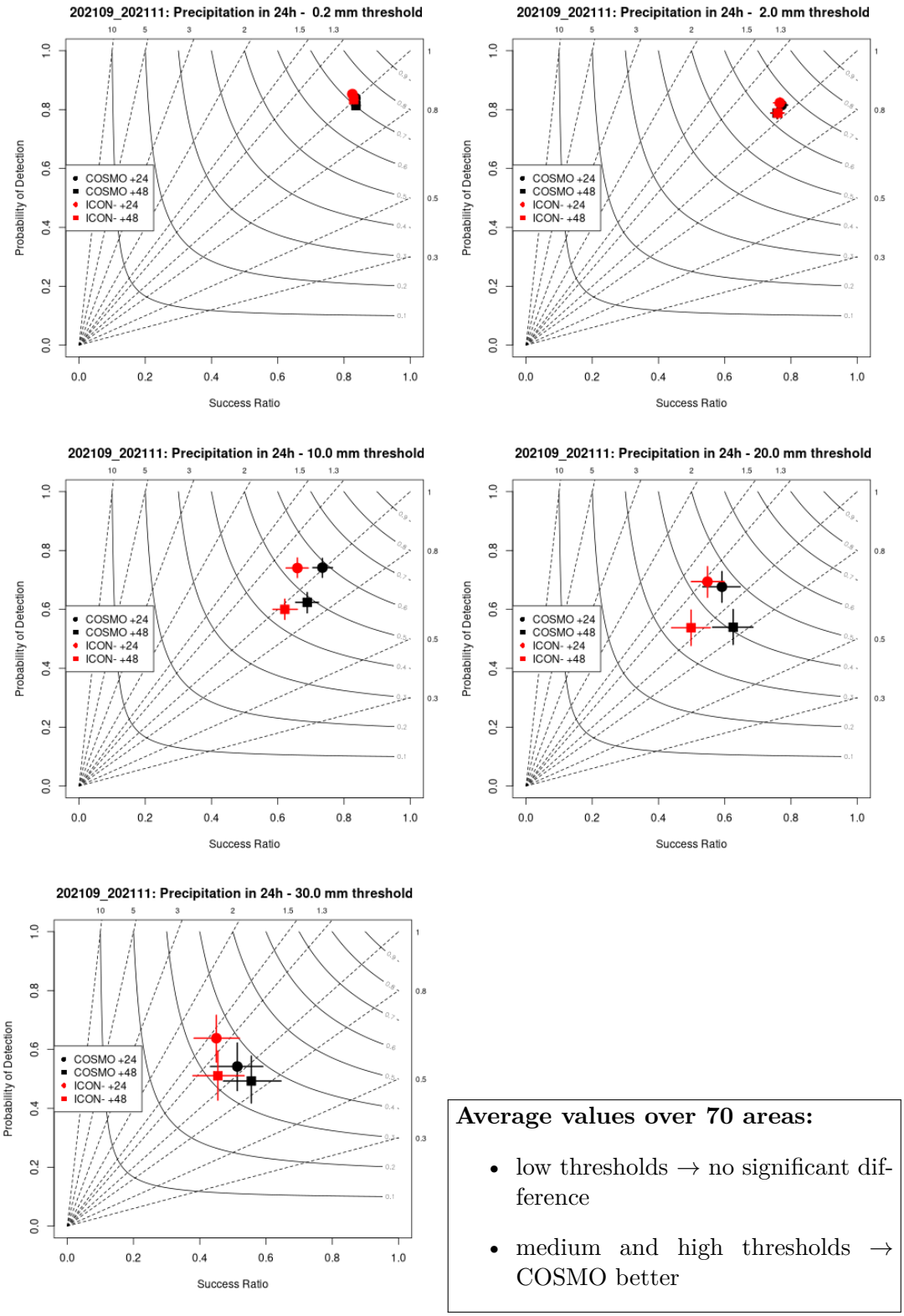
Figure 33: Performance diagrams: JJA 2021, average precipitation +24h and +48h.



Maximum values over 70 areas:

- low thresholds → very slight improvement for ICON
- medium and high thresholds → no improvement for ICON; COSMO better

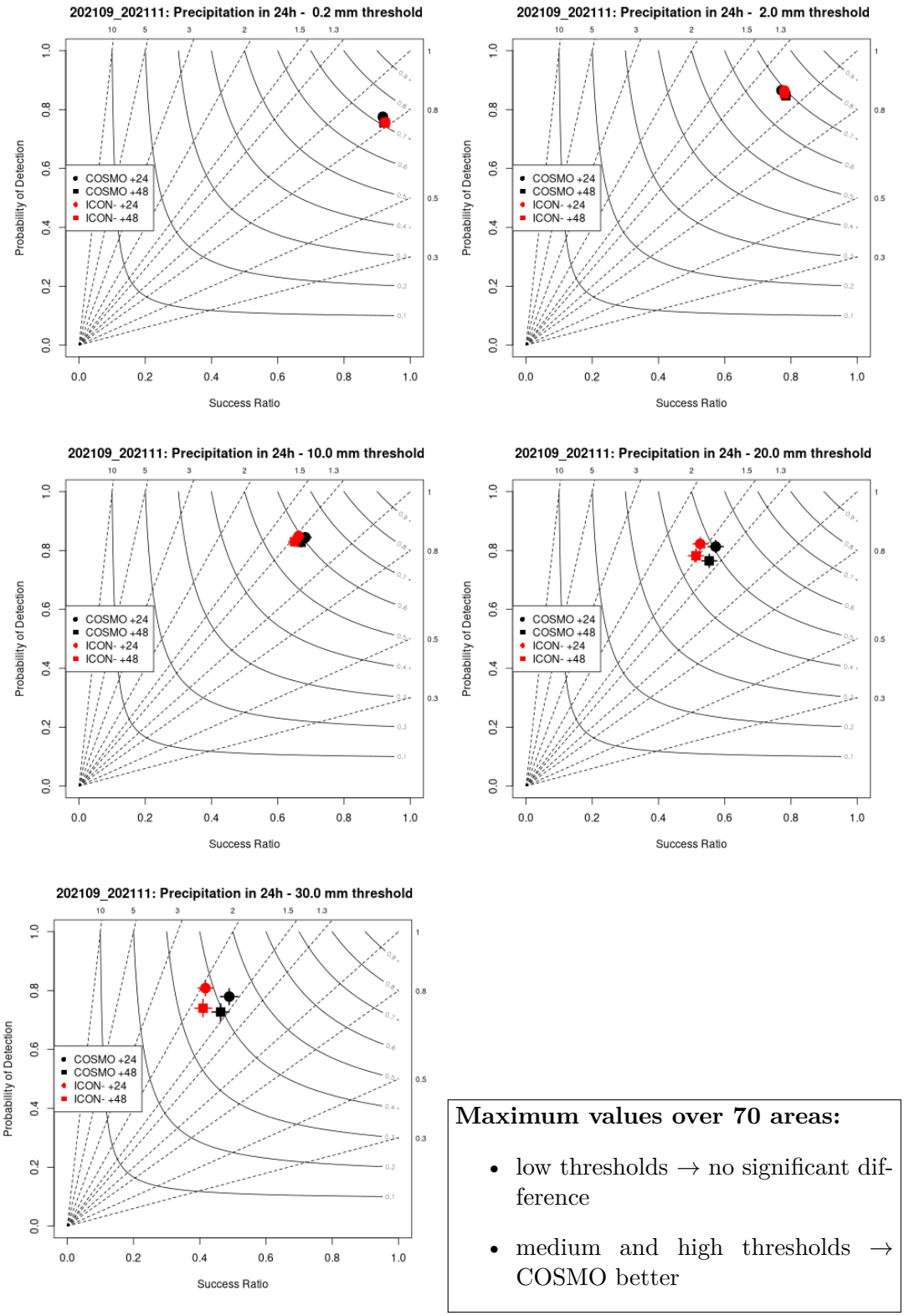
Figure 34: Performance diagrams: JJA 2021, maximum precipitation +24h and +48h.



Average values over 70 areas:

- low thresholds → no significant difference
- medium and high thresholds → COSMO better

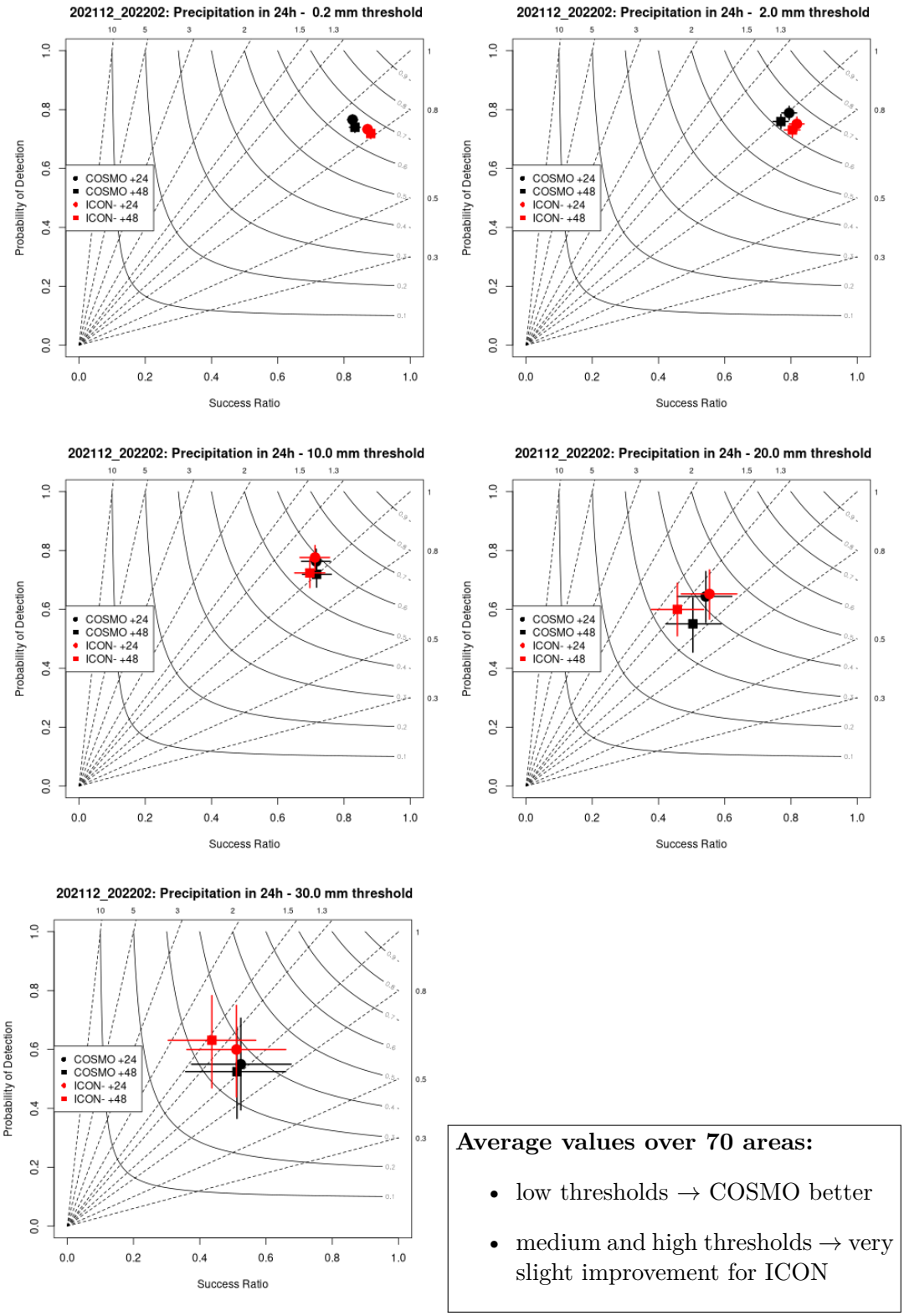
Figure 35: Performance diagrams: SON 2021, average precipitation +24h and +48h.



Maximum values over 70 areas:

- low thresholds → no significant difference
- medium and high thresholds → COSMO better

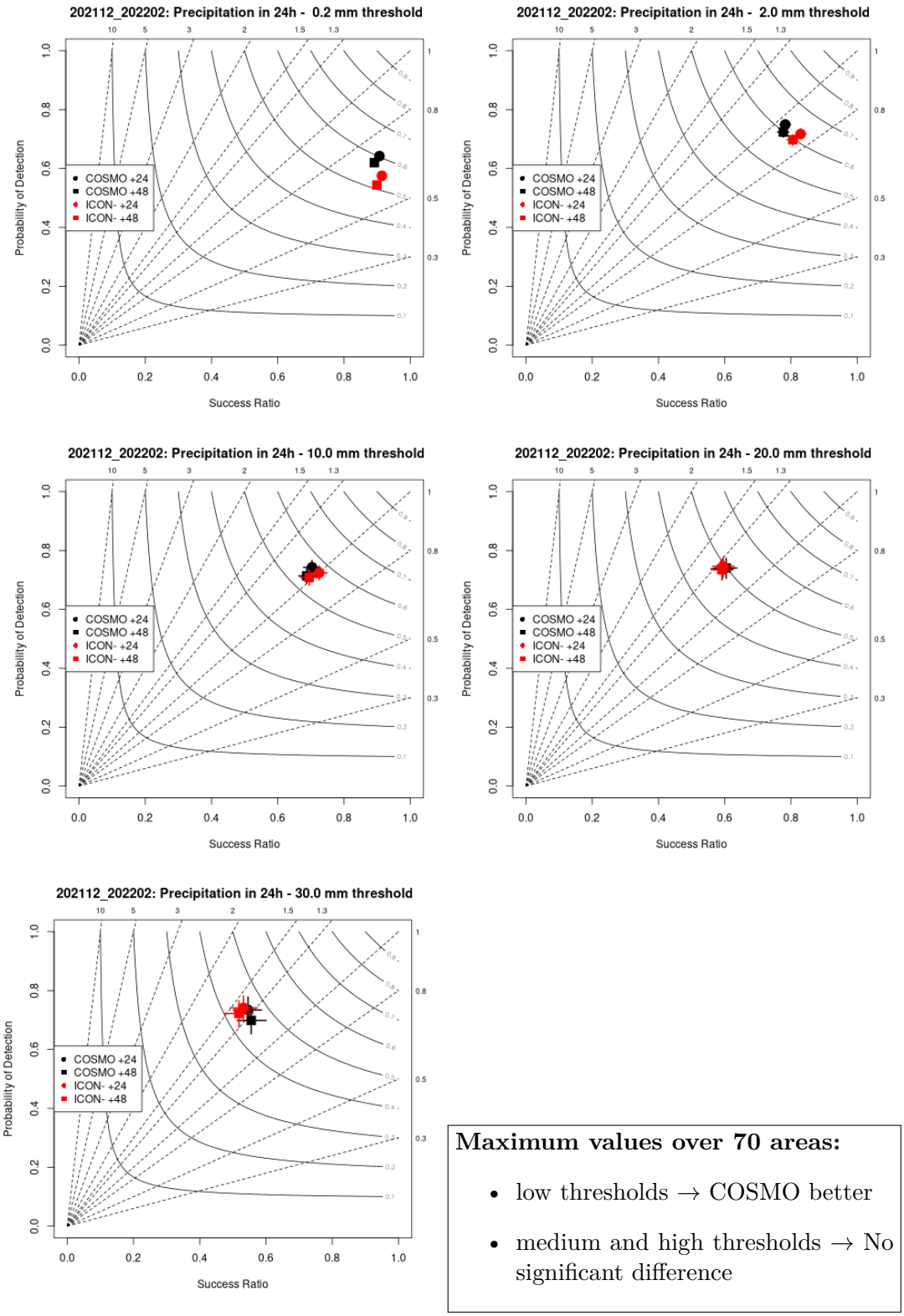
Figure 36: Performance diagrams: SON 2021, maximum precipitation +24h and +48h.



Average values over 70 areas:

- low thresholds → COSMO better
- medium and high thresholds → very slight improvement for ICON

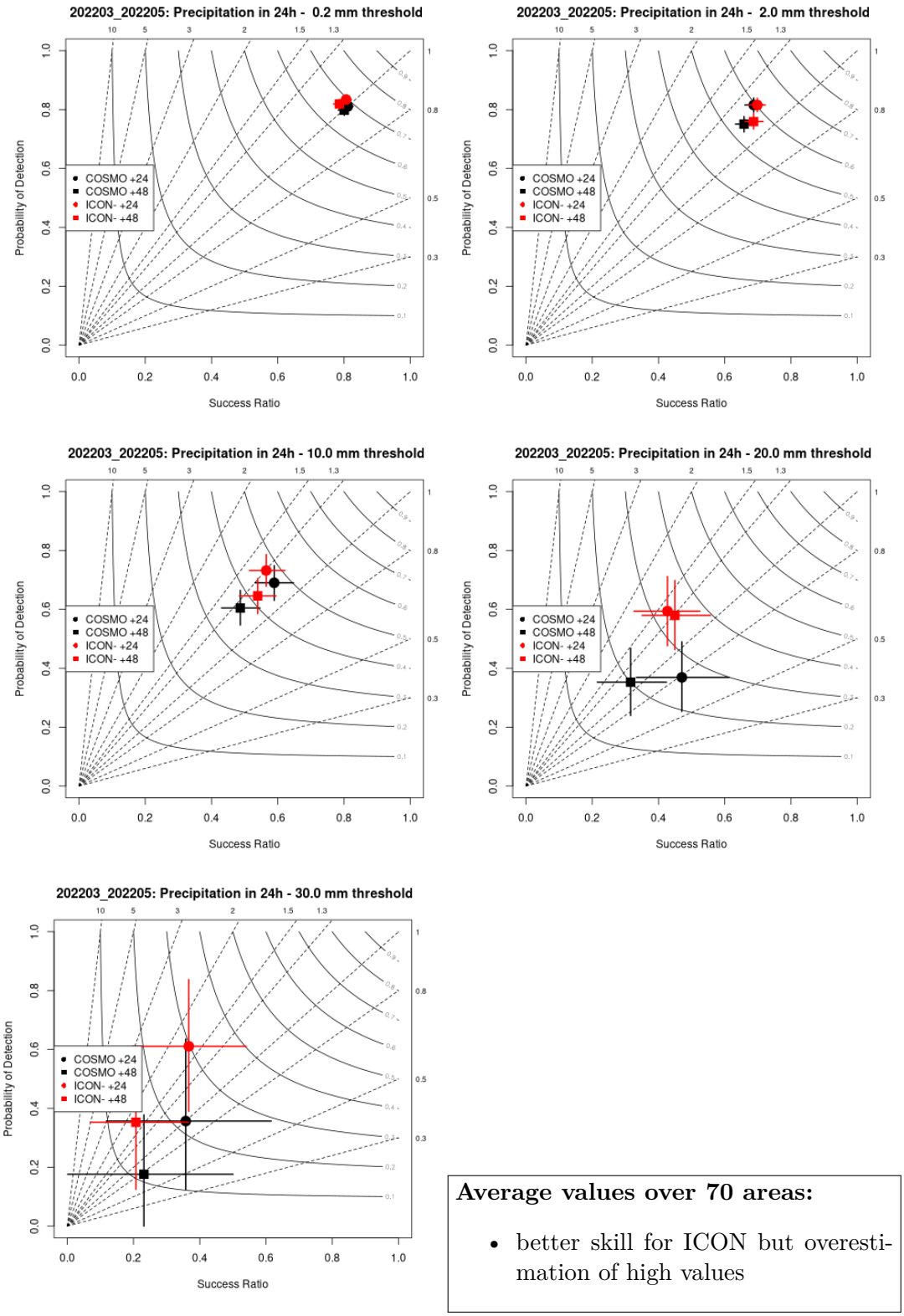
Figure 37: Performance diagrams: DJF 2022, average precipitation +24h and +48h.



Maximum values over 70 areas:

- low thresholds → COSMO better
- medium and high thresholds → No significant difference

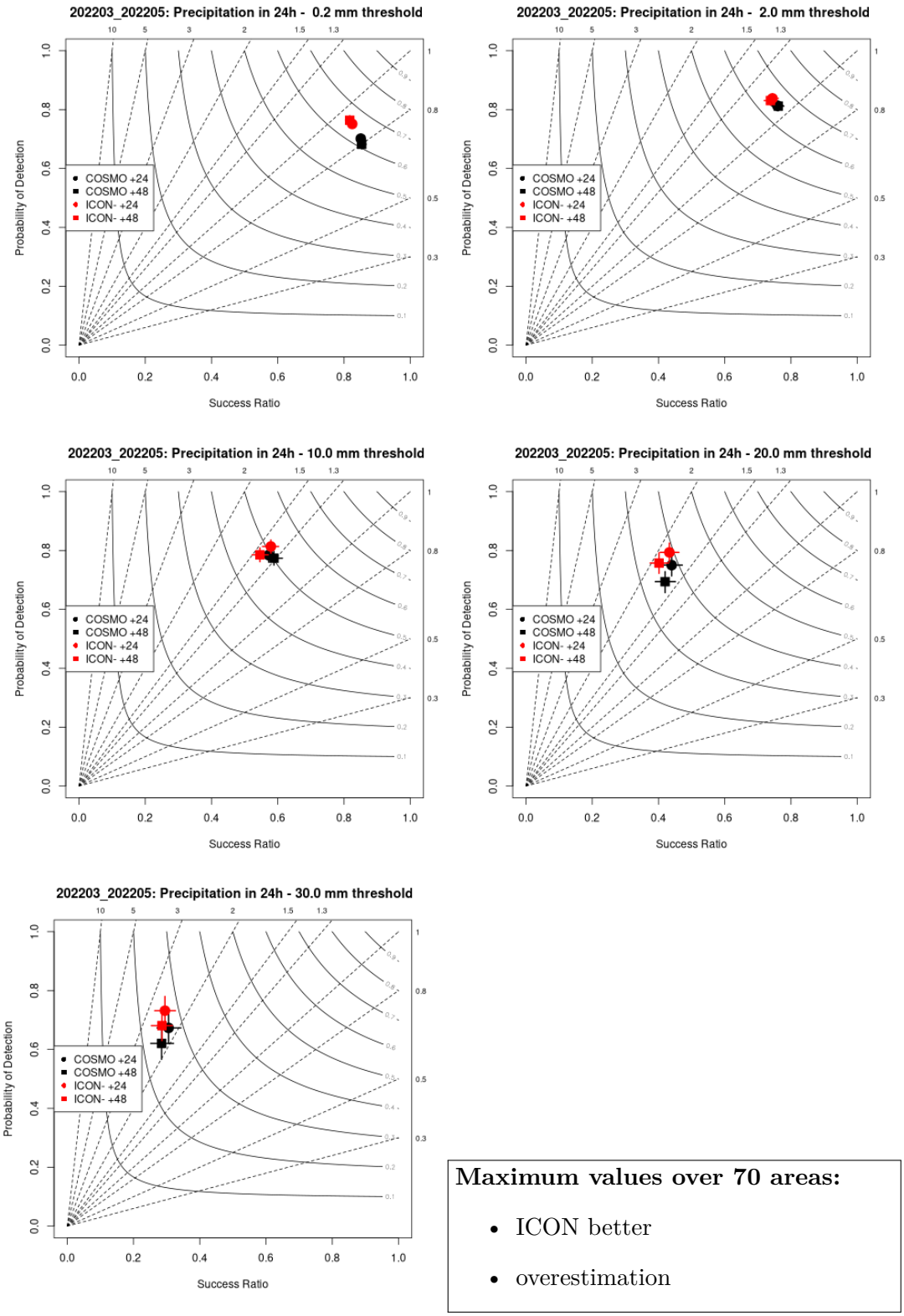
Figure 38: Performance diagrams: DJF 2022, maximum precipitation +24h and +48h.



Average values over 70 areas:

- better skill for ICON but overestimation of high values

Figure 39: Performance diagrams: MAM 2022, average precipitation +24h and +48h.



Maximum values over 70 areas:

- ICON better
- overestimation

Figure 40: Performance diagrams: MAM 2022, maximum precipitation +24h and +48h.

5.5 Computational efficiency

The use of the ICON model will also be advantageous from a computational point of view. In fact, taking into account that the two models (COSMO-IT and ICON-IT) have the same setups (in terms of horizontal and vertical resolution and domain extent), on the basis of daily resource consumption (SBU) on the ECMWF cray-hpc, it can be estimated that the use of ICON model saves around 40% of computational resources.

5.6 Conclusion

In the report the verification scores of ICON-IT and COSMO-IT models against surface and upper air observations are shown. Results are encouraging as generally ICON-IT outperforms COSMO-IT for surface parameters and temperature profiles. Unfortunately, because some technical problem with archived data, verification against rain-gauges are missing, as well as scores against radiosonde data in the first three quarters of time-period.

Concerning the precipitation, the fuzzy results do not show significant differences between ICON-IT and COSMO-IT. Also the performance diagrams do not highlight any significant improvement of ICON-IT, except for the spring period where it performs better than COSMO-IT. Both models tend to overestimate the maximum precipitation values for medium-high thresholds.

The ICON model is fully operational at the Italian Met Service, together with COSMO-IT, and available to forecasters for daily use. The full switch to ICON is conditioned to the availability of the GPU version of the model, precondition for the implementation of the ICON-IT EPS. Therefore ICON will hopefully become the "reference" model by Q4 2023.

6 Results from Italy: ARPAE

E. Minguzzi(1), D. Cesari(1), T. Gastaldo(1), V. Poli(1), C. Marsigli(1), M.S. Tesini(1), I. Cerenzia(1)

1. *Arpae-Emilia-Romagna (Italy)*

6.1 Overview

We report here the description of the implementation and the first verification results of ICON-LAM at Arpae-SIMC (ICON-2I), which is due to replace the present system (COSMO-2I) in the forthcoming months. ICON-2I is run on Cineca HPC server in Bologna, and it is now pre-operational. Although the modeling chain is still incomplete (high resolution data assimilation based on KENDA has not yet been implemented), results are encouraging: the performance of ICON-2I is in all respects similar or better to COSMO-2I and the model has not shown any big issues in daily use. The comparison of the two models highlights the importance of a high resolution assimilation cycle in short term forecasts.

6.2 Setup

ICON-2I covers the domain 2.7°E-22.3°E/33.5°N-49°N, with 65 vertical levels and a horizontal resolution of about 2.2 km (R9B8; figure 41).

Figure 41: ICON-2I integration domain.

Initial and boundary conditions are taken from IFS deterministic runs ('cold start' for every forecast; boundary conditions updated every hour). In the present pre-operational configuration, the model runs 4 times per day, with a forecast range of 78 hours. Model version is 2.6.4, and the 'reference' namelists (from ICON-D2) are used, except for the radiation scheme, which is based on RRTM.
In order to tune and verify the performance of the system, a set of 48-hours reforecasts have

been performed. Six 33-days periods have been selected in years 2019-2021, picked-up to include episodes of intense precipitation, in both Northern and Southern Italy and in different seasons. These reforecasts have been run in a slightly improved configuration: ICON version is 2.6.5, and ecRad radiation scheme is activated.

6.3 Verification

The main goal of the verification, at this early stage, was to assess the performance of ICON-2I in predicting intense precipitation in the first two days of forecast. Therefore, the verification has been mainly focused on reforecast runs, and their forecast range is only 48 hours. While comparing ICON-2I and COSMO-2I models, it should be noticed that the initial conditions are different: the latter uses a high-resolution analysis generated through KENDA by assimilating conventional and radar observations while the former uses 'cold starts' from IFS analysis.

In figures 42 and 43, the model forecasts are compared with the precipitation estimates based on the Italian radar composite corrected with real-time raingauges observations. For each forecast time, the Fractions Skill Score for ICON-2I and COSMO-2I is shown. ICON often outperforms COSMO, especially in the second forecast day. The FSS is higher for ICON in both areas in the intermediate seasons (spring + fall), with a better performance over Northern Italy. On the other hand, in summer and in particular in Southern Italy, where precipitations are mostly convective, COSMO strongly benefits from the more accurate initial conditions.

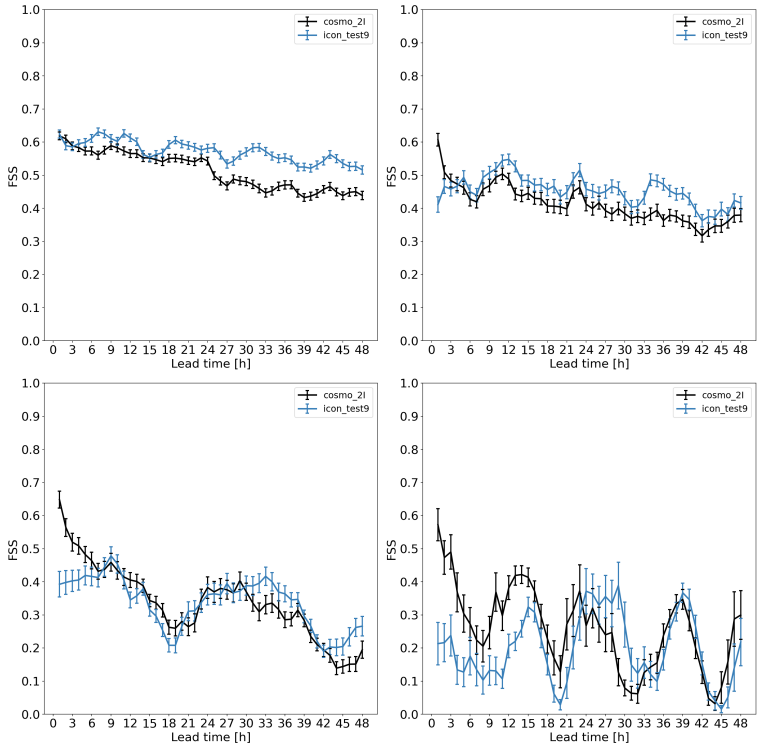


Figure 42: Fractions Skill Score of hourly precipitation, verification against Italian radar composite: ICON-2I (blue) and COSMO-2I (black), box size = 0.2°, threshold = 1 mm. Top: reforecasts for spring+fall (93 days); bottom: reforecasts for summer (62 days). Left: Northern Italy; right: Southern Italy

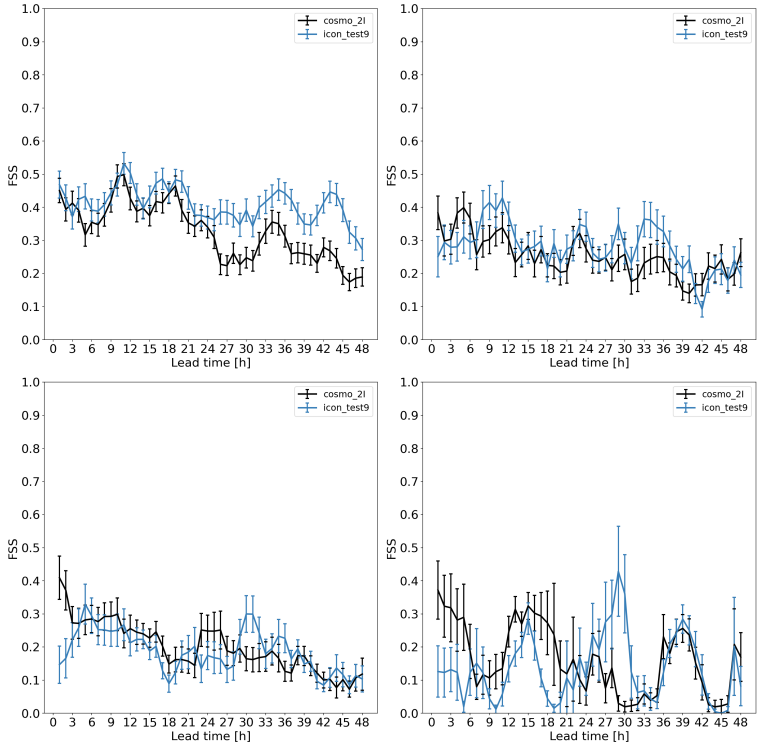


Figure 43: Fractions Skill Score of hourly precipitation, verification against Italian radar composite: ICON-2I (blue) and COSMO-2I (black), box size = 0.2°, threshold = 5 mm. Top: reforecasts for spring+fall (93 days); bottom: reforecasts for summer (62 days). Left: Northern Italy; right: Southern Italy

In figures 44 and 45, the model forecasts are compared with raingauges observations. For civil protection purposes, Italy is divided more than 100 'alert areas', with variable sizes from 500 to 4000 Km^2: the plots compare the average and maximum values of 3 hours accumulated precipitation observed and forecasted on each of these areas. Considering a multi-category verification, with forecast and observed precipitation binned in 5 different classes , the accuracy of the model in forecasting the correct category has been evaluated using the 'Gerrity Score'. Aggregating the results for the first 24 hours (from +03h to +24h) and for the second 24 hours (from +27h to +48h) (see figure 44), we note that ICON-2I performs better than both ECMWF (especially for maximum precipitation) and COSMO-2I (particularly in the second forecast day). COSMO performance is much better in the first day of forecast than in the second day: this could be linked to the use of data assimilation cycle (see also figures 42 and 43).

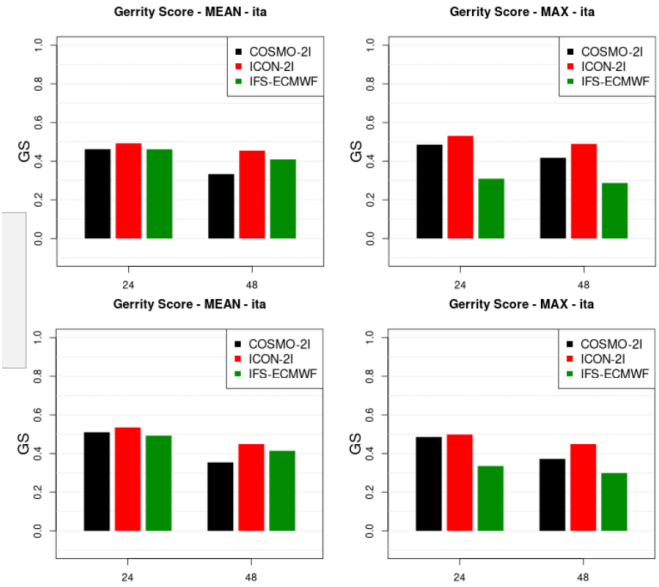


Figure 44: Gerrity score, 3 hours accumulated precipitation: from +3h to +24h (D0, label '24') and from +27h to +48h (D1, label '48'). Classes used in the multi-category verification: for the maximum:]0-0.2],]0.2-2],]2-10],]10-30],>30 mm/3h ; for the mean: 0,]0.2-1],]1-5],]5-10],>10 mm/3h)

This is also evident in the 'Performance Diagrams' (figure 45), in which the results for the verification of precipitation exceeding different thresholds are compared: for the first forecast day (from +03h to +24h) the performance of COSMO and ICON is rather similar, but on the second forecast day ICON significantly outperforms COSMO. In addition it can be noticed that the high resolution models, even if at the expense of a high number of false alarms, are able to reproduce high maximum of rainfall quite well while for IFS the missed alarms prevail. Some underestimation of the precipitation in the first forecast hours of ICON-2I are reported (see figure 42,43 and the red symbol in figure 45 which represents the +03h forecast step). This could be ascribed to the use of the IFS analysis as initial condition, and the implementation of a data assimilation cycle should be beneficial. Possible problems of imbalance in the interpolated initial conditions were examined and ruled out. Further investigations are ongoing. The available sample of pre-operational forecasts is still too small for a thorough verification. Nevertheless, preliminary feedback from forecasters are mostly positive: despite the lack of a high-resolution analysis, ICON-2I is reliable and the performance was in all respects at least comparable with COSMO-2I. Durung the summer 2022 a significant improvement in 2m maximum temperature forecasts has been observed (see figure 46).

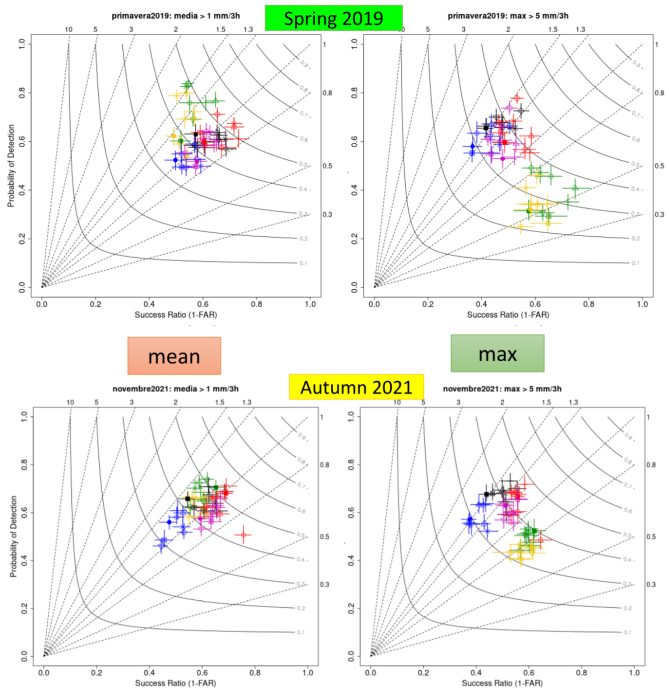


Figure 45: Performance diagram for 3h precipitation: IFS (green(D0), yellow(D1), COSMO-2I (black(D0) , blue(D1)), ICON-2I (red(D0), purple(D1)).
Top: May 2019; bottom: November 2021.
Left: average precipitation in each 'alert area', threshold 1 mm/3h; right: maximum precipitation in each 'alert area', threshold 5 mm/3h.

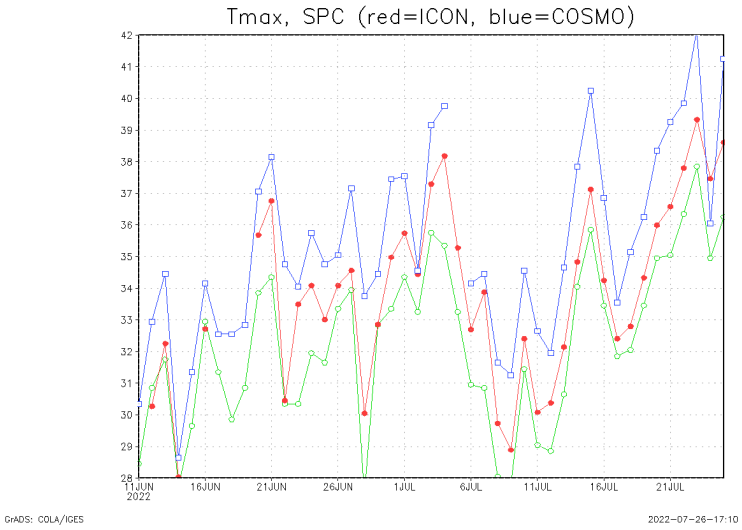


Figure 46: 2m maximum temperature during the June-July 2022 heat wave at S.Pietro Capofiume (Po Valley): observations (green), ICON-2I (red), COSMO-2I (blue)

6.4 Computational efficiency

Cineca HPC system (Galileo100) is composed of 554 computing nodes, but only a portion of these is available to Arpae-SIMC. Each node has 2 CPU Intel CascadeLake 8260 with 24 cores each (48 cores per node). The RAM available for each node is 384GB. The pre-operational ICON-2I simulations use 736 MPI processes on 16 nodes, without multithreading. In this configuration, each 78 hours forecast requires about 1h50' to be completed.
Preliminary tests with multithreading (OpenMP) carried out by Cineca HPC staff have shown a very encouraging 30% increase in performance. As soon as it will be possible to use ecRad code with OpenMP on Cineca HPC, this feature will be tested and hopefully implemented in all ICON simulations.

6.5 Plans for operation

A key request from the forecasters is that the main model run shall be available at 7AM local time, and shall provide forecasts for 3 full days. Presently this requirement is fulfilled by using the forecast starting the previous day at 18Z, and by extending the forecast range from 72 to 78 hours.
In the next months, a data assimilation procedure based on KENDA, employing both conventional and radar observations, is going to be implemented. Then, the ICON-2I run initialized at 00 UTC is going to use the analysis generated by KENDA at the same time, while still using the IFS run initialized at 18 UTC to provide boundary conditions.
Meanwhile, parallel running of COSMO-2I and ICON-2I will continue, at least until the implementation of KENDA for ICON-2I. Moreover, verification will be extended to other key meteorological parameters (notably: maximum and minimum temperature and wind gusts).
The ICON setup will be essentially the same used in the reforecasts, but some details on the implementation (data flow of boundary conditions for KENDA, soil moisture analysis) have not yet been decided. Moreover, slight adjustments in ICON namelists will probably be introduced: a set of sensitivity tests will be completed during the overlap period, on domain size, soil moisture analysis and tuning of model parameters, to try to better adapt the model configuration to the Mediterranean environment.
Subsequently, ICON will gradually replace COSMO in all the other operational suites: the 'RUC' runs, starting every 3 hours with a forecast range of 18 hours, and the convection-permitting ensemble COSMO-2I-EPS, starting once a day with a 51 hours forecast range and using KENDA initial conditions. Finally a run at continental scale with resolution of about 5 km will also be implemented, mainly to provide the atmospheric forcing to air quality and maritime models.

7 Results from Poland

J. Linkowska(1), W. Interewicz(1) and D. Wójcik

1. *Institute of Meteorology and Water Management National Research Institute*

Model data from ICON-PL and COSMO-CE-PL (2.8) have been compared with observational data from Polish SYNOP stations and radiosonde data from the entire model domain for a period June 2021 to May 2022. A selection of verification results are presented in the report. Overall, ICON-PL performs better than COSMO-CE-PL for both surface and upper air parameters as evidenced by reduced or similar RMSE values compared to COSMO-CE-PL in all seasons and for all continuous parameters and most lead times. In terms of categorical verification results, ICON-PL is more skillful for 6-hourly precipitation than COSMO-CE-PL at drizzle and light rain thresholds. At larger thresholds, it is less clear which model is more skillful and at the 10mm threshold, the results are rather noisy.

7.1 Overview and Setup ICON-PL

The ICON-PL domain covers the territory of Poland and also includes parts of neighbouring countries (Figure 47). R2B10 grid with equivalent 2.5km resolution is used. The modelled area covers the rotated grid of COSMO-CE-PL 2.8 domain.

ICON-PL has been run at IMGW-PIB since the end of May 2019. The initial and boundary conditions are taken from the global ICON model runs at DWD. The model runs twice per day at 00 and 12UTC to T+48 and is run without data assimilation.

ICON version 2.6.2.2 (with "cp/cv" bugfix and the RRTM radiation scheme) was made operational in June 2021 and is the model version used for the forecasts in this report.

7.2 Computational aspects

A comparison of computational performance of ICON v. 2.6.3 at R2B10 grid (2.5 km equivalent) and COSMO Model v. 5.5 (2.8 km horizontal grid step) was performed to assess scalability of both models. The number of cores ranged from a typical for today operational NWP applications at IMGW-PIB (192) to future applications with a significantly refined grid (3072).

In the tests performed, ICON model was run at grid comprising of 0.29M elements horizontally and 65 vertical levels, while COSMO Model used regular grid with 0.15M points horizontally and 50 levels in vertical. Both grids actually cover a similar geographical area since the ICON domain has to be extended to ensure correct interpolation of forecast to a regular lat-lon grid.

Performance of both models was measured in terms of execution time, pseudo-scaling and parallelization efficiency (Figures 48 – 50, respectively). Pseudo-scaling numbers were calculated relative to 192 cores (8 nodes with 24 cores each). The COSMO 2.8 km compilation setup and the ICON tNN compilation setups follow the operational configuration. Tests were carried out at an external TASK platform (free from operations).

Pseudo-scaling and parallelization efficiency obtained (Figures 49 – 50) reveal that performance of the ICON Model using MPICH library (t25, t29 setups) and performance of

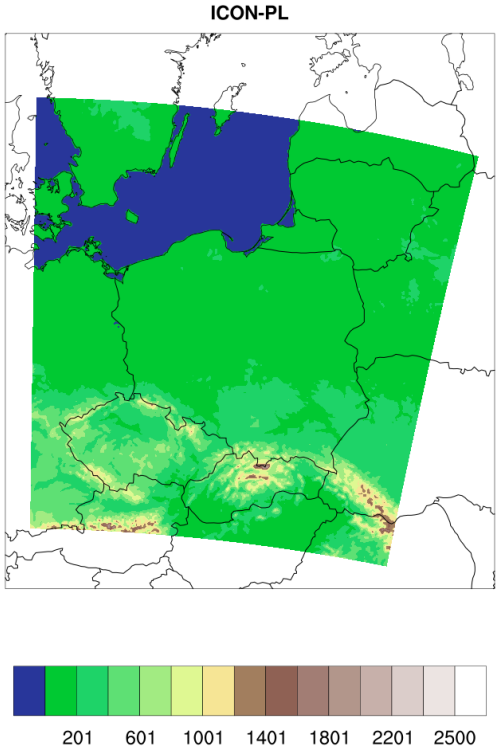


Figure 47: ICON-PL domain

COSMO Model (Intel 2017 + Intel MPI 2017 setup) are in the saturation range. That is, those setups shall not be applied for applications requiring 800 cores or more. Figures 48 – 50 depict that ICON compiled with the Intel 2020 MPI library (t30 setup) has a significantly better performance than using MPICH. However, Intel 2020 MPI library is currently not available at the IMGW-PIB operational cluster. Moreover, ICON compiled with Intel 2020 is unstable for 128 nodes (3072 cores; model hangs up at the initialization phase as with older Intel MPI libraries). Despite of that, the t30 setup provides improved performance and, thus, primary attention will be given to that setup and any forthcoming releases of the Intel suite. Given these results and the very different configurations of the models, it is still too early for a direct comparison of the runtime of COSMO and ICON.

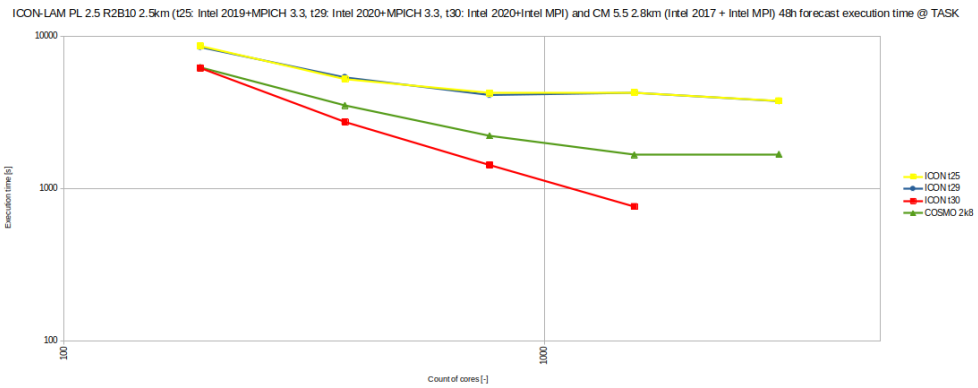


Figure 48: ICON-LAM and COSMO Model 48h forecast computation time

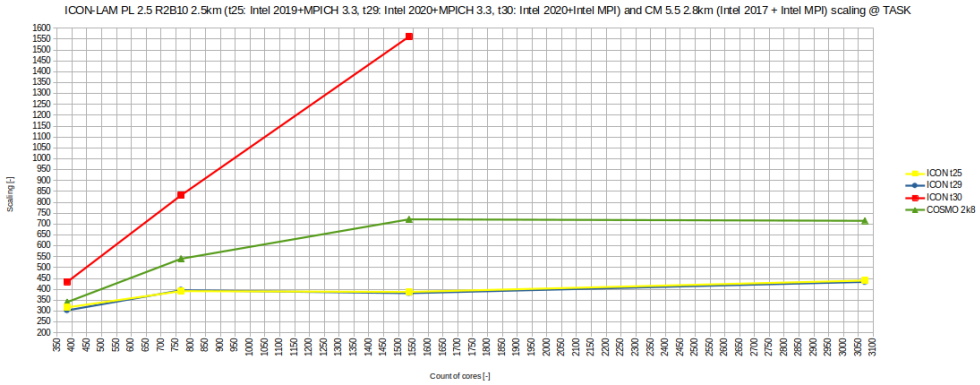


Figure 49: ICON-LAM and COSMO Model scaling

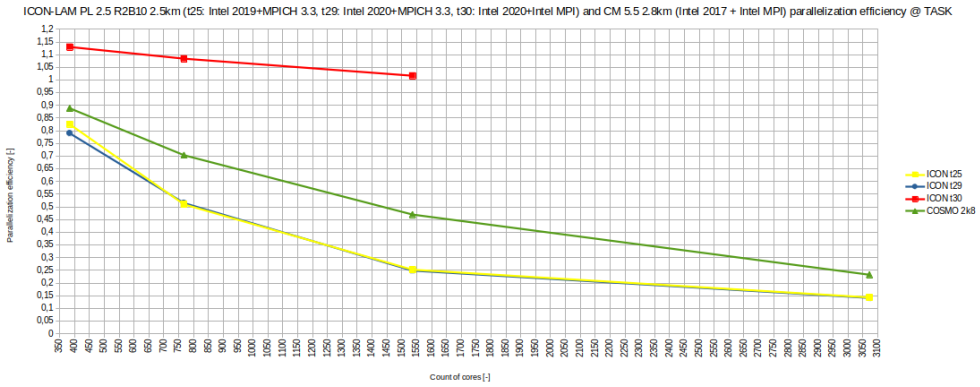


Figure 50: ICON-LAM and COSMO Model parallelization efficiency

7.3 Verification

ICON-PL and COSMO-CE-PL model forecasts were verified against synop observations and radiosonde data. The results of 00 UTC run are presented in the report. The verification scores were obtained using MEC/Rfdbk system and visualized using SHINY server (both tools implemented at IMGW-PIB).

7.3.1 Surface Verification

For surface verification purposes 2m temperature, 2m dew point temperature, surface pressure, wind speed, wind direction, total cloud cover and 6-hourly accumulated precipitation were verified against Polish synop stations.

Surface parameters

The continuous scores: the Mean Error and Root Mean Squared Error are presented in the Figures 51 – 58.

JJA 2021

Temperature 2m: Both ICON-PL and COSMO-CE-PL models generally overestimate 2m temperature (T2M). However, the bias is strongly reduced in ICON-PL compared to

COSMO-CE-PL and this is consistent with a significant reduction in RMSE (e.g a 25% reduction at T+27). For both models there is some indication of a diurnal cycle to the bias, with the largest warm bias being at 00UTC and 03UTC, and the smallest bias occurring at 06UTC and 18UTC (T+18, T+30, T+42). However there is also a peak at 12UTC. The largest improvement in bias for ICON-PL with respect to COSMO-CE-PL occurs at night-time between 21UTC and 03UTC.

Dew point temperature: COSMO-CE-PL has a large positive bias in dew point temperature (TD2M) at the beginning of the forecast with a spin-down over the first 24 hours of forecast. ICON-PL has a much reduced positive bias at the beginning of the forecast and consequently a reduced spin-down. The RMSE is much reduced in ICON-PL compared to COSMO-CE-PL (e.g a 25% reduction at T+12).

Surface Pressure: Both ICON-PL and COSMO-CE-PL models underestimate surface pressure. This negative bias tends to grow with forecast lead time. ICON-PL has a much reduced bias and RMSE compared to COSMO-CE-PL. The error growth is smoother in ICON-PL than COSMO-CE-PL which has peaks in errors at T+15 and T+39. There is a 33% reduction in RMSE in ICON-PL compared to COSMO-CE-PL at T+39.

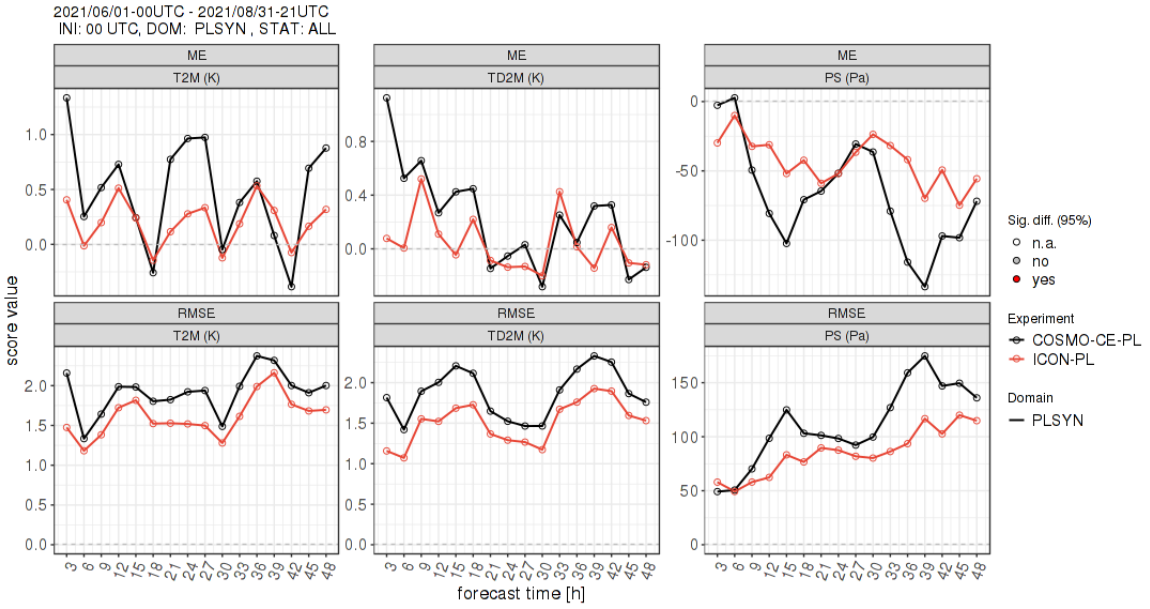


Figure 51: Temperature 2m, Dew Point Temperature 2m, Surface Pressure, JJA 2021, Continuous scores: ME and RMSE

Total Cloud Cover: Both ICON-PL and COSMO-CE-PL models have a diurnal cycle to their bias. COSMO-CE-PL overestimates Total Cloud Cover during night time and underestimates during the day time. ICON-PL has a similar characteristic to COSMO-CE-PL in having an overestimate to Total Cloud Cover during night time, (although the bias is slightly larger than COSMO-CE-PL resulting in a slightly increased RMSE). ICON-PL in contrast to COSMO-CE-PL overestimates Total Cloud Cover during the day time. Whilst the magnitude of the positive bias at 12UTC is similar to the negative bias in COSMO-CE-PL at 12 UTC, ICON-PL has a reduced RMSE during day time.

Wind speed at 10m: Both ICON-PL and COSMO-CE-PL generally overestimate wind speed. The COSMO-CE-PL strong bias is greatest at night time and also increases with forecast range. By contrast, the ICON-PL strong bias is greatest at 15 UTC and does not increase with forecast range. ICON-PL has a reduced RMSE compared to COSMO-CE-PL.

Wind direction: ICON-PL veers the wind direction by about 3 degrees compared to observations. COSMO-CE-PL backs the wind direction by about 3 degrees compared to observations. Although the magnitude of the bias in direction is about the same in both models, the bias makes up a very small component of RMSE. ICON-PL has a reduced RMSE compared to COSMO-CE-PL.

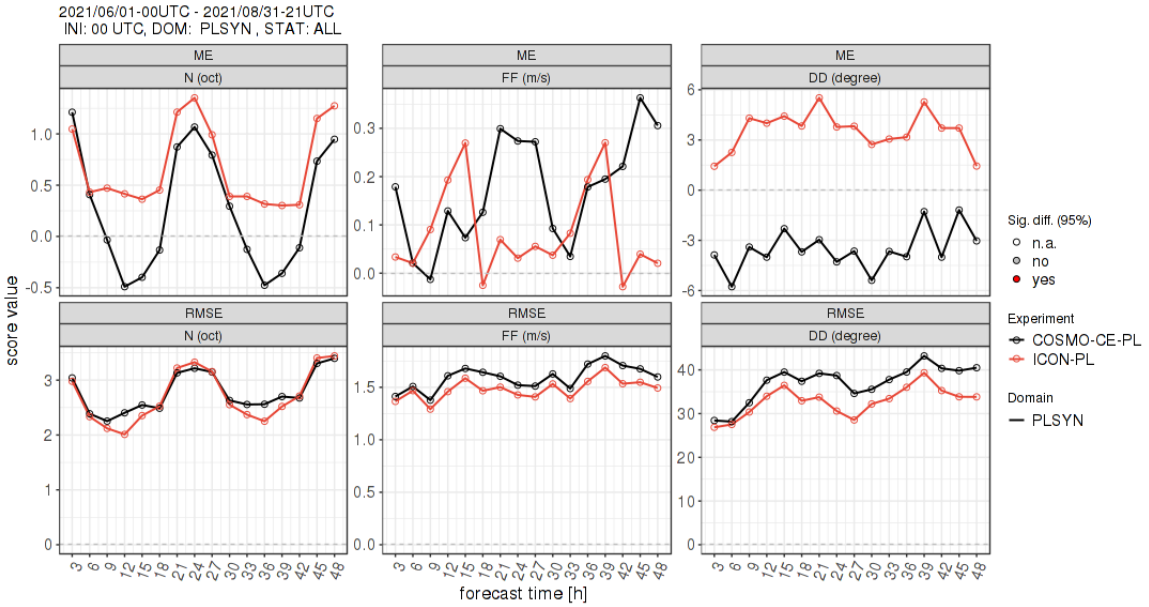


Figure 52: Total Cloud Cover, Wind speed at 10m, Wind direction, JJA 2021, Continuous scores: ME and RMSE.

SON 2021

Temperature 2m: There is a clear diurnal cycle to the bias in both COSMO-CE-PL and ICON-PL with the largest night time warm bias being at 03UTC, and the the coldest day time bias occurring at 15UTC. The diurnal cycle in COSMO-CE-PL is superimposed on a cooling trend with forecast lead time. This means that although COSMO-CE-PL overestimates T2M during night time for the first forecast day, this turns into an underestimation by the second forecast day. The day time T2M is always underestimated. ICON-PL has a smaller bias than COSMO-CE-PL, with the largest reduction occurring during the day time. ICON-PL also has a much reduced RMSE compared to COSMO-CE-PL (e.g a 25% reduction at T+3).

Dew point temperature: COSMO-CE-PL has a large positive bias in TD2M at the beginning of the forecast with a spin-down over the first 24 hours of forecast (as seen in JJA 2021). This spin-down continues into the second day of the forecast. ICON-PL does not have this spin-down in bias and has an overestimate of TD2M through the forecast. There is also a diurnal cycle with an afternoon peak in bias at 15UTC and a minimum at 06 UTC. ICON-PL has reduced RMSE compared to COSMO-CE-PL throughout the forecast range.

Surface Pressure: ICON-PL and COSMO-CE-PL models have rather different bias characteristics. ICON-PL underestimates the surface pressure at all forecast ranges and has a negative trend with forecast range. COSMO-CE-PL appears to have a diurnal cycle with a positive bias at 06UTC and a negative bias at 12 UTC. The overall magnitude of the biases is small (smaller than JJA 2021) and never larger than 0.5 hPa. ICON-PL has a systematically smaller growth rate in RMSE with the lead time compared to COSMO-CE-PL. Although slightly larger at the beginning of the forecast, the RMSE is reduced by 25% at T+36.

Total Cloud Cover: Both ICON-PL and COSMO-CE-PL models have a diurnal cycle to

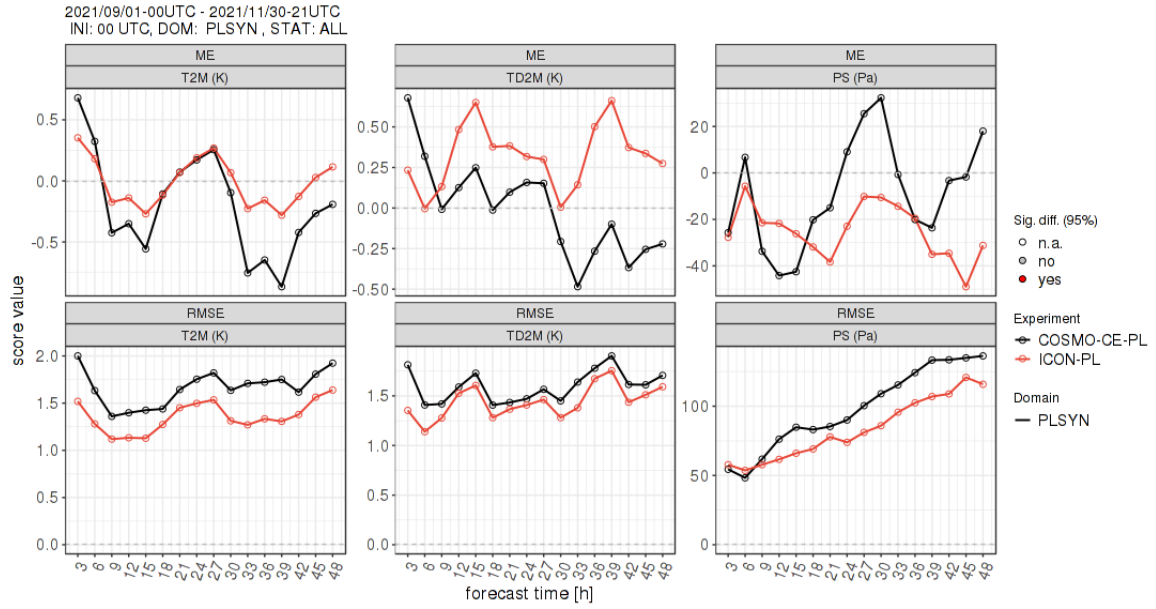


Figure 53: Temperature 2m, Dew Point Temperature 2m, Surface Pressure, SON 2021, Continuous scores: ME and RMSE

their bias. COSMO-CE-PL overestimates Total Cloud Cover during night time and underestimates during the day time. ICON-PL overestimates Total Cloud Cover at all forecast ranges. The bias is smaller than COSMO-CE-PL during the day time but slightly larger for the last 6 hours of the lead time. ICON-PL has a reduced RMSE during day time compared to COSMO-CE-PL.

Wind speed at 10m: COSMO-CE-PL has a large diurnal cycle in the bias with an underestimation of Wind Speed during day time and overestimation during night time. Although ICON-PL overestimates wind speed at all forecast ranges, the magnitude of the bias is much smaller overall. ICON-PL has a slightly reduced RMSE compared to COSMO-CE-PL

Wind direction: ICON-PL veers the wind direction by about 4-5 degrees compared to observations. COSMO-CE-PL backs the wind direction by about 2-4 degrees compared to observations. Although the magnitude of the bias in direction is about the same in both models, the bias makes up a small component of RMSE. ICON-PL has a slightly reduced RMSE compared to COSMO-CE-PL.

DJF 2022

Temperature 2m: T2M in both models is underestimated throughout the whole forecast range. There is a clear diurnal cycle to the bias in COSMO-CE-PL while the diurnal cycle bias is not present in ICON-PL. The coldest bias in COSMO-CE-PL occurs during the daytime and similar to the autumn season, the diurnal cycle in COSMO-CE-PL is superimposed on a cooling trend with forecast lead time. ICON-PL has a smaller bias than COSMO-CE-PL, with the largest reduction in bias occurring during the day time. Both models increase RMSE with the forecast time although ICON-PL has a much reduced RMSE compared to COSMO-CE-PL.

Dew point temperature: ICON-PL underestimates dew point temperature throughout the forecast period (except small overestimation at 15UTC). COSMO-CE-PL overestimates T2DM for the first 30 hours of the forecast and from T+36 to T+39. The model underestimates T2DM for the other lead times. Although both models increase RMSE with the forecast time, ICON PL has significantly lower RMSE values than COSMO-CE-PL.

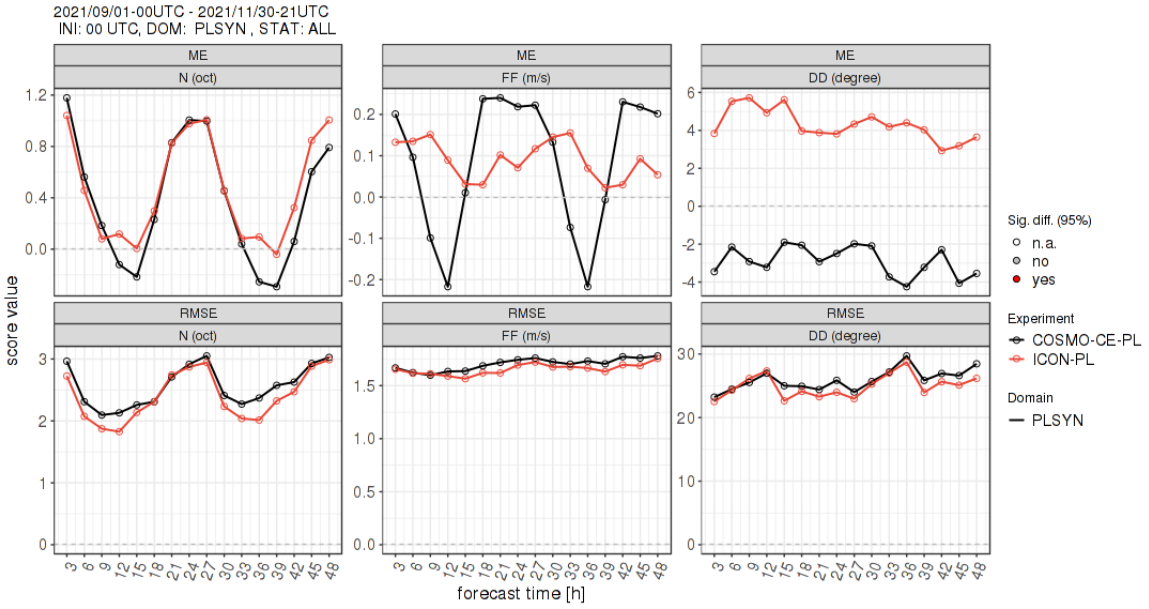


Figure 54: Total Cloud Cover, Wind speed at 10m, Wind direction, SON 2021, Continuous scores: ME and RMSE.

Surface Pressure: As in Autumn, COSMO-CE-PL and ICON-PL have different bias characteristics. COSMO-CE-PL underestimates surface pressure during day time for the first forecast day (except at 06UTC) and overestimates for the other forecast ranges. This overestimation tends to grow with the lead time. ICON-PL underestimates PS at all forecast ranges and the underestimation is greater than in COSMO-CE-PL during the first forecast day. ICON-PL has slightly greater RMSE in the first 12 hours of the forecast and smaller error after 12 UTC than COSMO-CE-PL. The reduction of the error increases with the lead time.

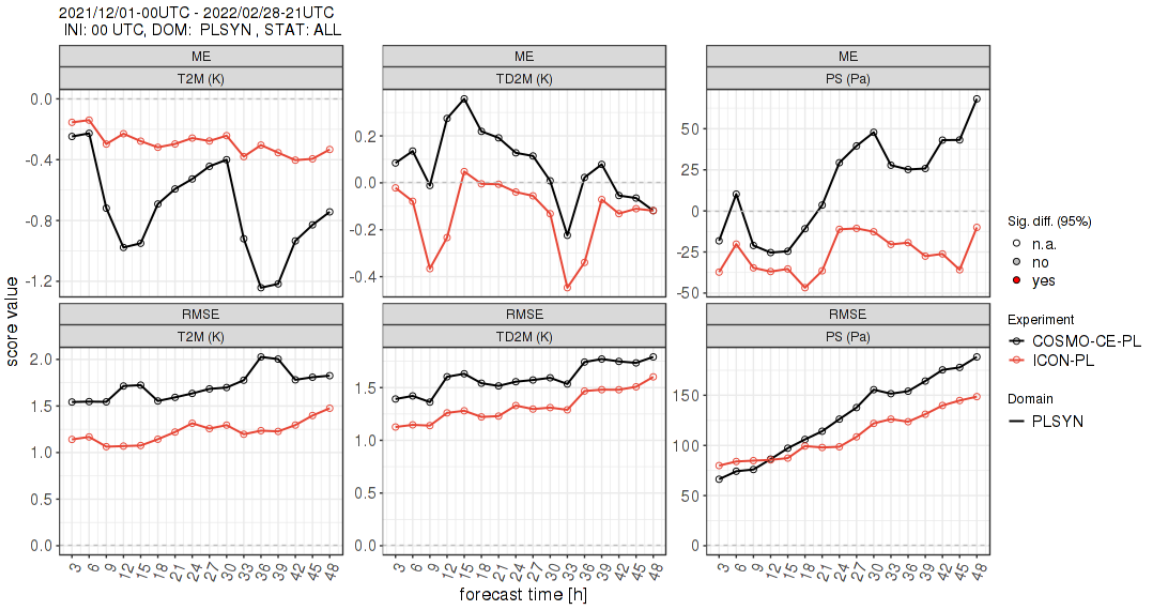


Figure 55: Temperature 2m, Dew Point Temperature 2m, Surface Pressure, DJF 2022, Continuous scores: ME and RMSE

Total Cloud Cover: Both ICON-PL and COSMO-CE-PL models have a diurnal cycle to their bias. COSMO-CE-PL has more clouds than ICON-PL. COSMO-CE-PL overestimates

total cloud cover at all forecast ranges. This overestimation is reduced in ICON-PL which overestimates TCC during nighttime and underestimates during daytime. ICON-PL has a reduced RMSE during daytime compared to COSMO-CE-PL.

Wind speed at 10m: ICON-PL overestimates wind speed while COSMO-CE-PL underestimates wind speed. The COSMO-CE-PL negative bias is greatest at day time with a peak at noon. The ICON-PL positive bias has a peak at T+33. ICON-PL has a reduced RMSE compared to COSMO-CE-PL.

Wind direction: ICON-PL veers the wind direction by about 3-6 degrees compared to observations. COSMO-CE-PL backs the wind direction by about 1.5-6 degrees compared to observations. ICON-PL has almost the same RMSE values compared to COSMO-CE-PL.

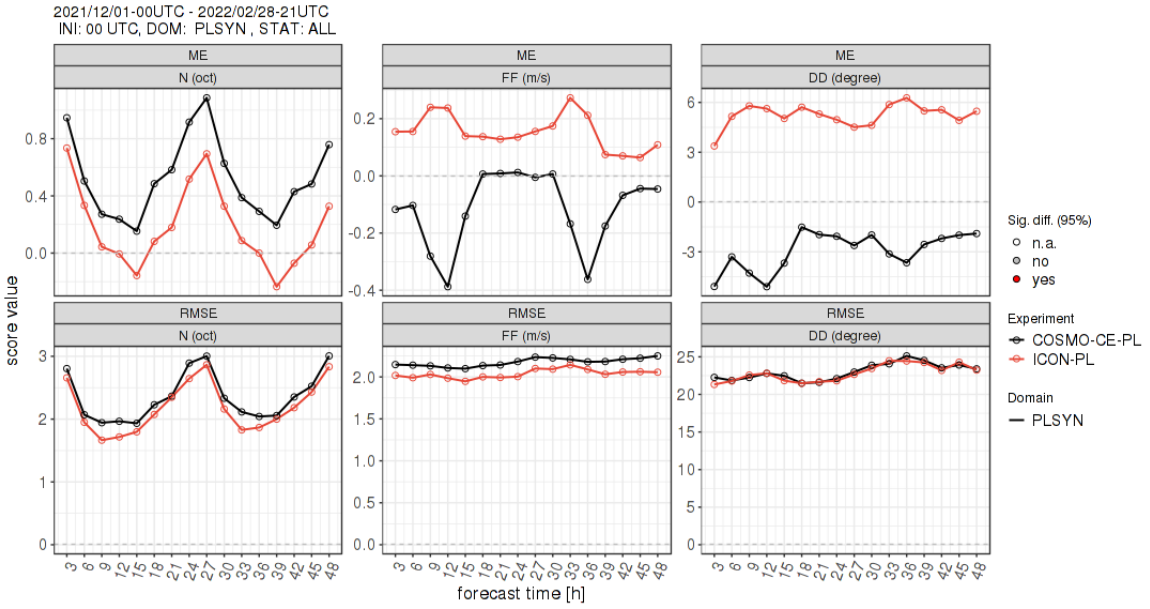


Figure 56: Total Cloud Cover, Wind speed at 10m, Wind direction, DJF 2022, Continuous scores: ME and RMSE.

MAM 2022

Temperature 2m: ICON-PL generally underestimates 2m temperature (T2M) except for the first 5 hours of the forecast and from T+27 to T+30. COSMO-CE-PL underestimates during daytime and overestimates during nighttime. The bias is much reduced in ICON-PL compared to COSMO-CE-PL at 03UTC and from 21UTC to 03UTC (T+21 to T+27). For COSMO-CE-PL there is some indication of a diurnal cycle to the bias, with the largest warm bias being at 03UTC, and the largest cold bias occurring at 18 UTC (T+18, T+36) and 06UTC (T+42). ICON-PL has a reduced RMSE compared to COSMO-CE-PL and both models see a trend towards increased RMSE with forecast lead time after an initial peak at T+3.

Dew point temperature: Both models overestimate dew point temperature. ICON-PL has a diurnal cycle to the bias and except for the first 9 hours of the forecast, ICON-PL has a larger positive bias in dew point temperature compared to COSMO-CE-PL. ICON-PL has a greater bias range than COSMO-CE-PL (0.12K to 1.66K versus 0.12K to 1.09K in COSMO-CE-PL). The RMSE is reduced in ICON-PL compared to COSMO-CE-PL.

Surface Pressure: Both models tend to have a negative PS bias and this bias is reduced in ICON-PL compared to COSMO-CE-PL for most forecast lead times. ICON-PL has reduced RMSE compared to COSMO-CE-PL.

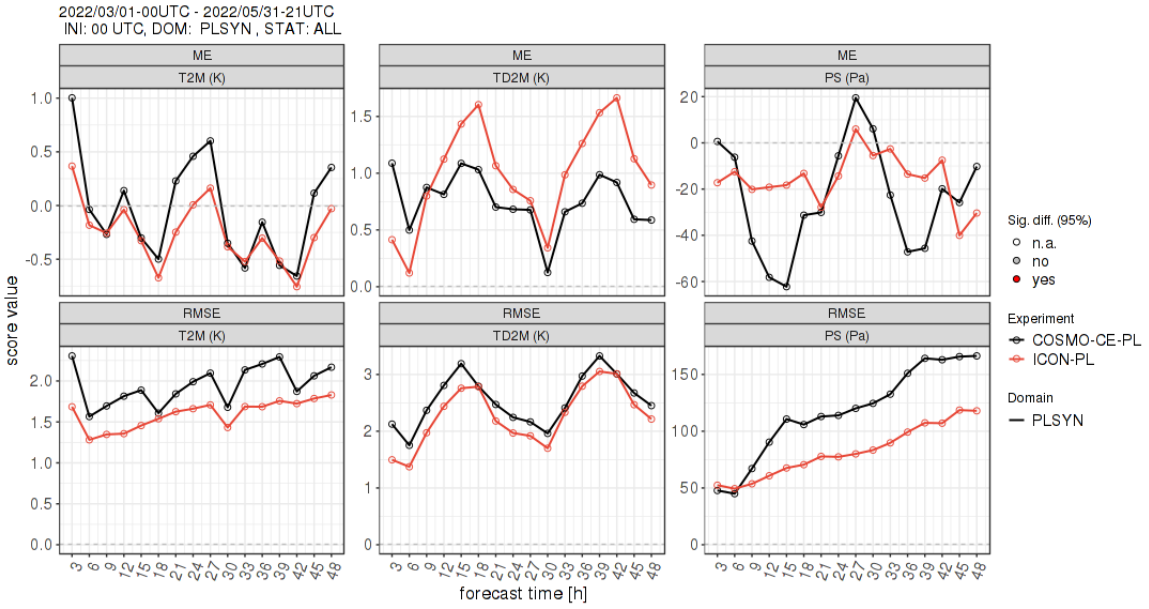


Figure 57: Temperature 2m, Dew Point Temperature 2m, Surface Pressure, MAM 2022, Continuous scores: ME and RMSE

Total Cloud Cover: Both models have a strong diurnal cycle to the bias with a large overestimation of cloud cover at nighttime and a much smaller bias during daytime. There is little difference in bias between the models during nighttime, but during the daytime the bias in ICON-PL is reduced compared to COSMO-CE-PL. The RMSE is reduced in ICON-PL compared to COSMO-CE-PL.

Wind speed at 10m: COSMO-CE-PL has a diurnal cycle to the bias with a strong bias at nighttime and a weak bias during daytime. ICON-PL is almost unbiased and has reduced bias compared to COSMO-CE-PL. The RMSE is reduced in ICON-PL compared to COSMO-CE-PL.

Wind direction: ICON-PL veers the wind direction by about 3-8 degrees compared to observations. COSMO-CE-PL backs the wind direction by about 2-6 degrees compared to observations. Despite the slightly larger bias, ICON-PL has RMSE values which are the same or lower than COSMO-CE-PL.

6-hourly precipitation accumulation

The categorical scores: the Frequency bias, the Equitable Threat Score, the Probability of detection and False Alarm Rate are shown on the Figures 59 - 62.

JJA 2021

ICON-PL has slightly larger ETS scores at the 0.2mm, 1.0mm, 5.0mm and 10.0mm thresholds compared to COSMO-CE-PL for all forecast ranges out to T+42. At T+48 the ETS scores are slightly smaller for the 5.0mm and 10.0mm thresholds. ICON-PL has a FBI closer to unity for the 0.2mm, 1.0mm, 5.0mm thresholds at all forecast ranges (with the exception of the 5.0mm threshold from T+30 onwards) compared to COSMO-CE-PL. The FBI for the 10.0mm threshold is worse (larger).

SON 2021

ICON-PL has slightly larger ETS scores at the 0.2mm and 1.0mm thresholds compared

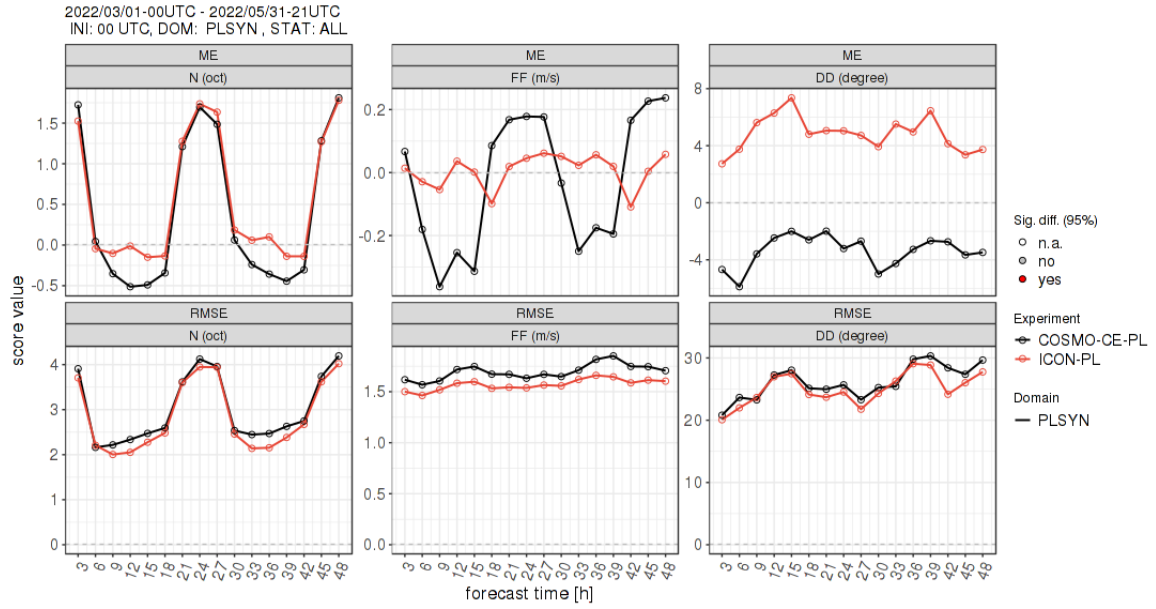


Figure 58: Total Cloud Cover, Wind speed at 10m, Wind direction, MAM 2022, Continuous scores: ME and RMSE.

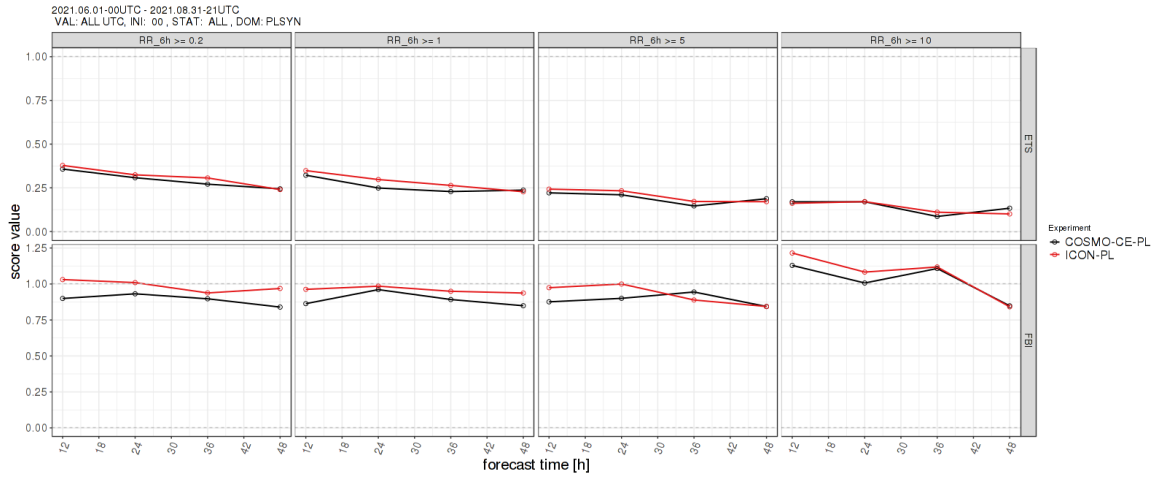


Figure 59: 6h accumulated precipitation, thresholds: 0.2mm, 1mm, 5mm, 10m, JJA 2021, Categorical scores: ETS, FBI.

to COSMO-CE-PL for almost all forecast ranges. ICON-PL has smaller ETS scores at the 5.0mm threshold from T+30 to T+48 and the 10.0mm threshold from T+12 to T+36 compared to COSMO-CE-PL. Both models overestimate precipitation at 0.2mm and 1.0 mm, the models underestimate precipitation at 5.0mm. FBI scores at 10.0mm are quite noisy.

DJF 2022

Both models have very similar ETS scores at the 0.2mm, 1.0mm and 5.0 mm thresholds, with ICON-PL having slightly better scores than COSMO-CE-PL. The ETS score at the 10.0mm threshold is very noisy in both models. Both models overestimate precipitation at 0.2mm, 1.0mm and 5.0 mm thresholds, with the FBI in COSMO-CE-PL being slightly larger than ICON-PL at the 0.2mm and 1.0mm thresholds. FBI scores at 10.0mm are quite noisy.

MAM 2022

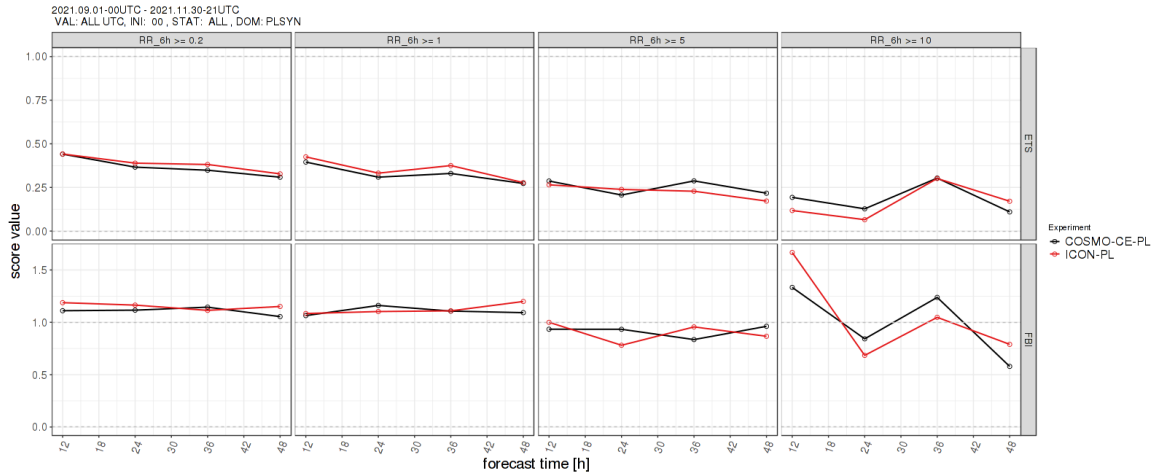


Figure 60: 6h accumulated precipitation, thresholds: 0.2mm, 1mm, 5mm, 10m, SON 2021, Categorical scores: ETS, FBI.

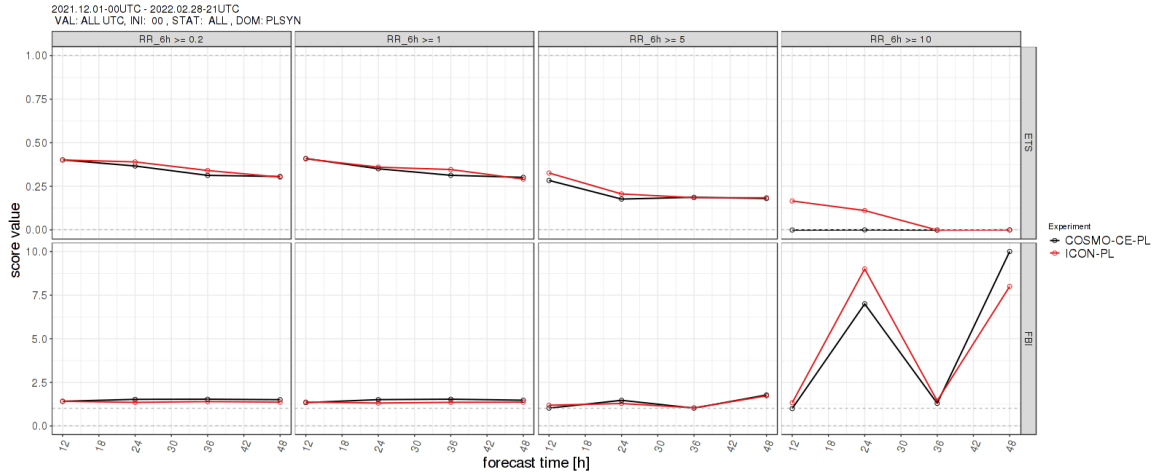


Figure 61: 6h accumulated precipitation, thresholds: 0.2mm, 1mm, 5mm, DJF 2022, Categorical scores: ETS, FBI.

Both models have very similar ETS scores at the 0.2mm and 1.0mm thresholds at all forecast ranges with ICON-PL having slightly better scores than COSMO-CE-PL. ICON-PL has smaller ETS scores at the 5.0mm threshold from T+12 to T+30. The ETS score at the 10.0mm threshold is very noisy in both models. Both models overestimate precipitation at 0.2mm, 1.0mm and 5.0 mm thresholds, with the FBI in ICON-PL being slightly larger than COSMO-CE-PL at the 0.2mm, 1.0mm thresholds. FBI scores at 10.0mm are quite noisy.

7.3.2 Upper air verification

For the upper air verification purposes temperature, relative humidity and wind speed were verified against radiosonde data from the entire model domain. Data for all seasons (JJA2021-MAM2022) were analysed but only the results for two seasons Autumn and Spring are presented in the report. The continuous scores: the Mean Error and Root Mean Squared Error are presented on the Figures 63 - 68.

SON 2021

Upper Air Temperature: ICON-PL has a ME that is smaller (colder) or equal to COSMO-

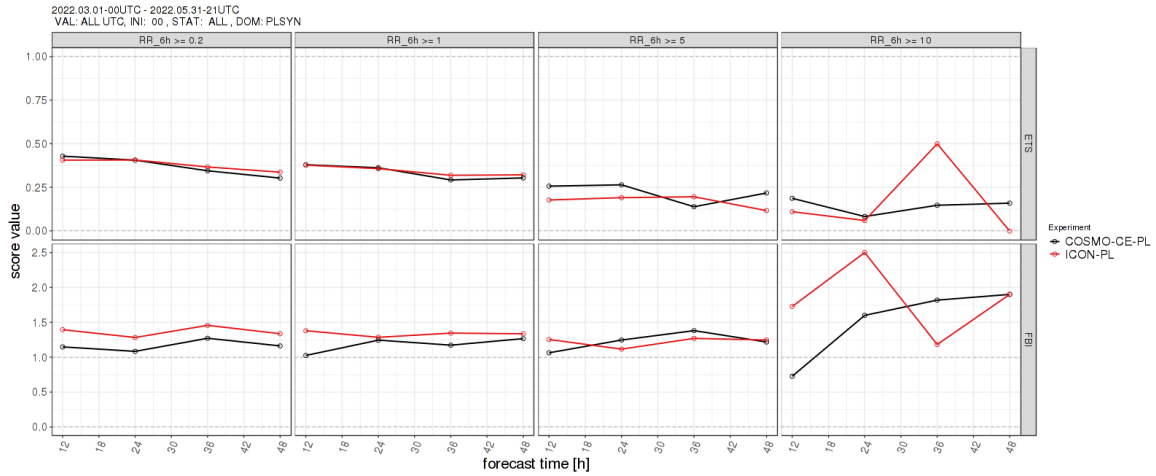


Figure 62: 6h accumulated precipitation, thresholds: 0.8mm, 1mm, 5mm, MAM2022, Categorical scores: ETS, FBI.

CE-PL at jet level (200 hPa to 300hPa) at all forecast ranges. ICON-PL has a smaller ME (warmer) than COSMO-CE-PL in the upper to mid-troposphere (300 hPa to 550hPa) at all forecast ranges. ICON-PL has larger (colder) ME scores than COSMO-CE-PL in the mid to lower troposphere (550hPa to 850hPa) for T+12, T+36. There is little difference at T+24 and T+48. ICON-PL is warmer than COSMO-CE-PL in the boundary layer (850hPa to 950 hPa). There is a diurnal variation to the ME with ICON-PL having larger values than COSMO-CE-PL at T+12, T+36 and smaller values at T+24 and T+48. ICON-PL is warmer than COSMO-CE-PL at 1000hPa. There is a diurnal variation to the ME with ICON-PL having smaller values than COSMO-CE-PL at all forecast ranges. ICON-PL has RMSE scores that are smaller or equal to COSMO-CE-PL for most levels and forecast ranges, and notably below 850hPa.

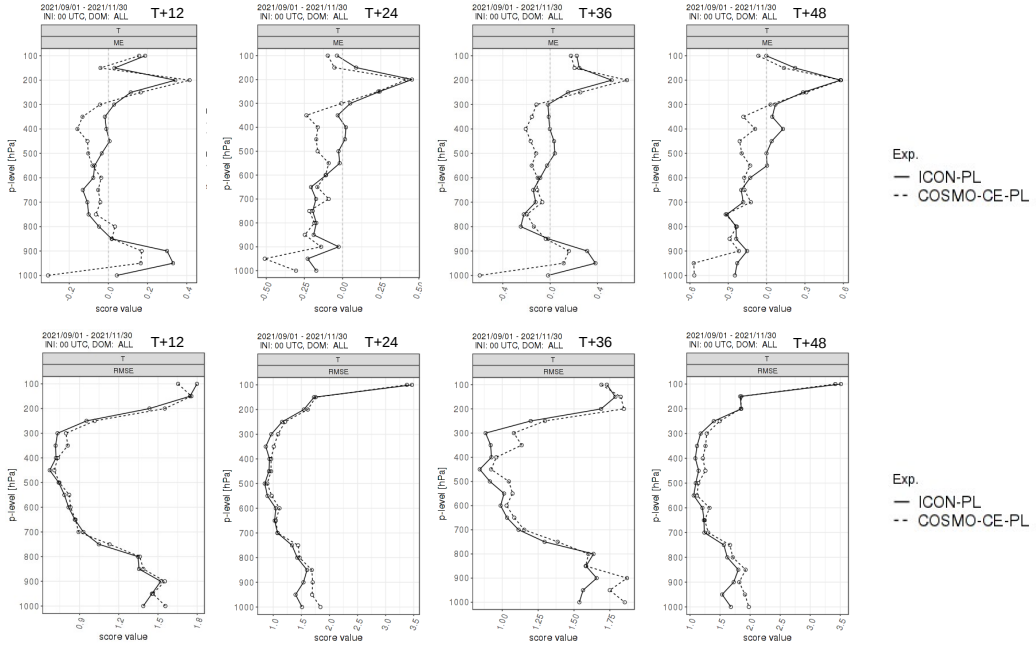


Figure 63: Upper Air Temperature (K), SON 2021, Continuous scores: ME and RMSE.

Relative Humidity: ICON-PL is drier than COSMO-CE-PL at jet level (200 hPa to 300hPa) at all forecast ranges except for T+48. ICON-PL is drier and has a smaller ME

than COSMO-CE-PL in the upper to lower mid-troposphere (300 hPa to 700hPa) at all forecast ranges. ICON-PL and COSMO-CE-PL have a similar ME in the lower troposphere (700hPa to 1000hPa) with ICON-PL being slightly more moist at T+36 and T+48. ICON-PL and COSMO-CE-PL models have similar RMSE scores.

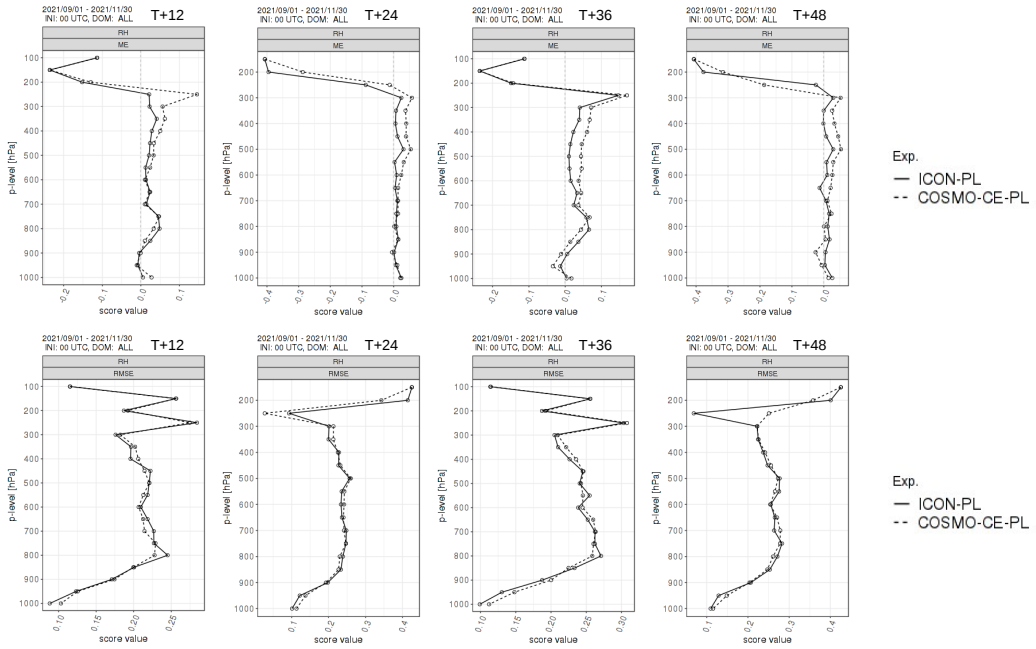


Figure 64: Relative humidity (0..1), SON 2021, Continuous scores: ME and RMSE.

Wind speed: ICON-PL has weaker winds than COSMO-CE-PL through much of the depth of the troposphere (from 400hPa to 950 hPa). This leads to a weak bias and increased (negative) ME. ICON-PL has stronger winds than COSMO-CE-PL at 1000hPa at T+12 and T+36. ICON-PL has RMSE scores that are smaller or equal to COSMO-CE-PL for most levels and forecast ranges.

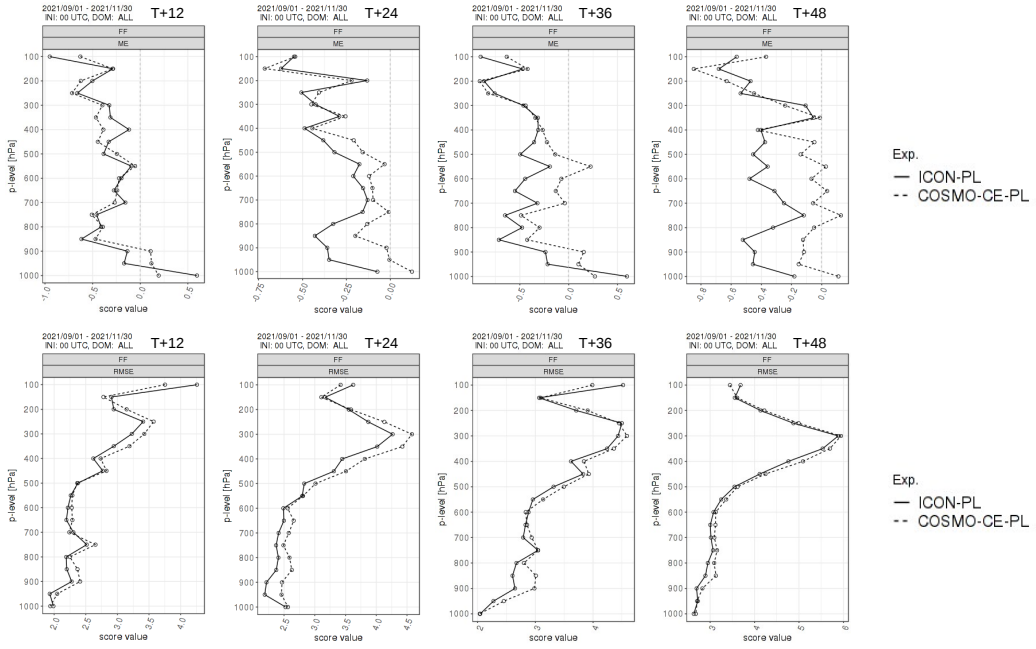


Figure 65: Wind speed (m/s), SON 2021, Continuous scores: ME and RMSE.

MAM 2022

Upper Air Temperature: ICON-PL has a smaller ME (warmer) than COSMO-CE-PL 150 hPa at all forecast ranges. ICON-PL has a smaller ME (warmer) than COSMO-CE-PL in the upper to mid-troposphere (300hPa to 600hPa) at all forecast ranges. ICON-PL is colder than COSMO-CE-PL in the mid and lower troposphere (650hPa to 950hPa) and this generally results in lower ME values since COSMO-CE-PL has a warm bias. There is also a diurnal variation in temperature biases in the lower troposphere with warmer biases at 800hPa at T+12 and T+36 in COSMO-CE-PL than at T+24 and T+48. The ME values for both models are negative at 1000hPa with ICON-PL having smaller errors than COSMO-CE-PL at T+12 and T+36 and larger errors at T+24 and T+48. ICON-PL has RMSE scores that are smaller or equal to COSMO-CE-PL for all levels and forecast ranges.

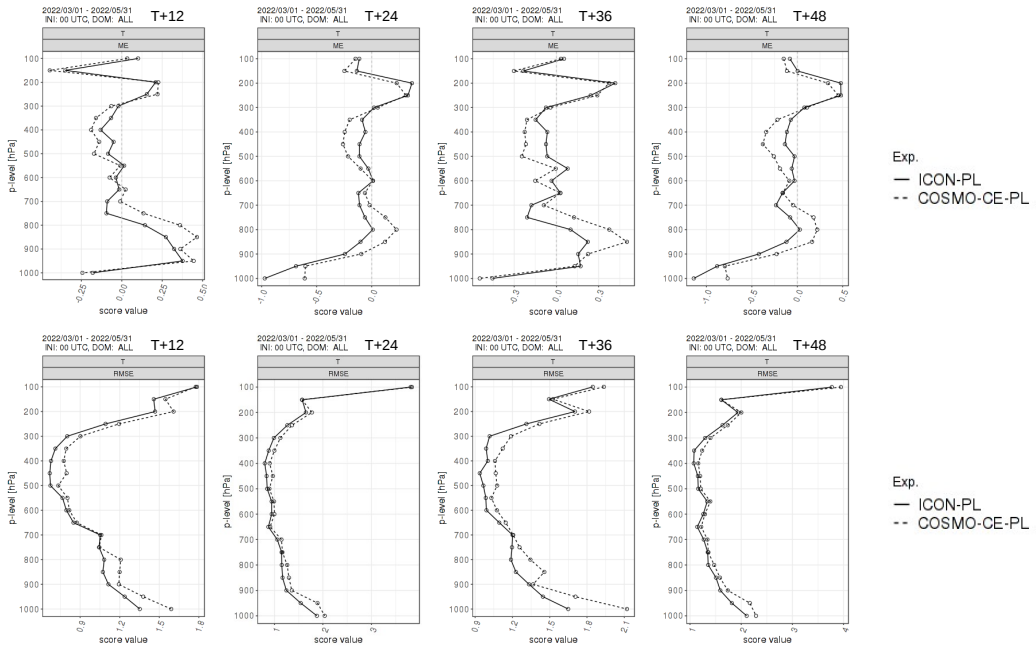


Figure 66: Upper Air Temperature (K), MAM 2022, Continuous scores: ME and RMSE.

Relative Humidity: ICON-PL has a smaller ME (drier) values than COSMO-CE-PL in the upper and mid-troposphere (300hPa to 700hPa) at all forecast ranges. In the lower troposphere (700hPa to 950hPa) there is a diurnal variation in humidity biases with COSMO-CE-PL having a dry bias at T+12 and T+36 at 850hPa. ICON-PL is moister than COSMO-CE-PL and reduces these biases. ICON-PL has RMSE scores that are similar to COSMO-CE-PL, but notably has reduced errors at 950hPa and 1000hPa.

Wind speed: ICON-PL has weaker winds than COSMO-CE-PL at T+24, T+36 and T+48 through much of the depth of the atmosphere from 250hPa to 950hPa. This leads to smaller ME values than COSMO-CE-PL between 400hPa and 700hPa since COSMO-CE-PL has a strong bias at these levels. However both models have a weak bias at jet level and ICON-PL introduces a daytime weak bias at 800hPa (T+12 and T+36) and a nighttime weak bias at 900hPa (T+24 and T+48) leading to increased (more negative) ME values. Below 950hPa ICON-PL has a daytime strong bias (T+12 and T+36) which is worse than COSMO-CE-PL. ICON-PL RMSE values are smaller than COSMO-CE-PL at T+12, T+24 and T+36 through much of the depth of the atmosphere. The difference is greater at T+24 and T+36. At T+48 the improvement is mainly in the lower troposphere.

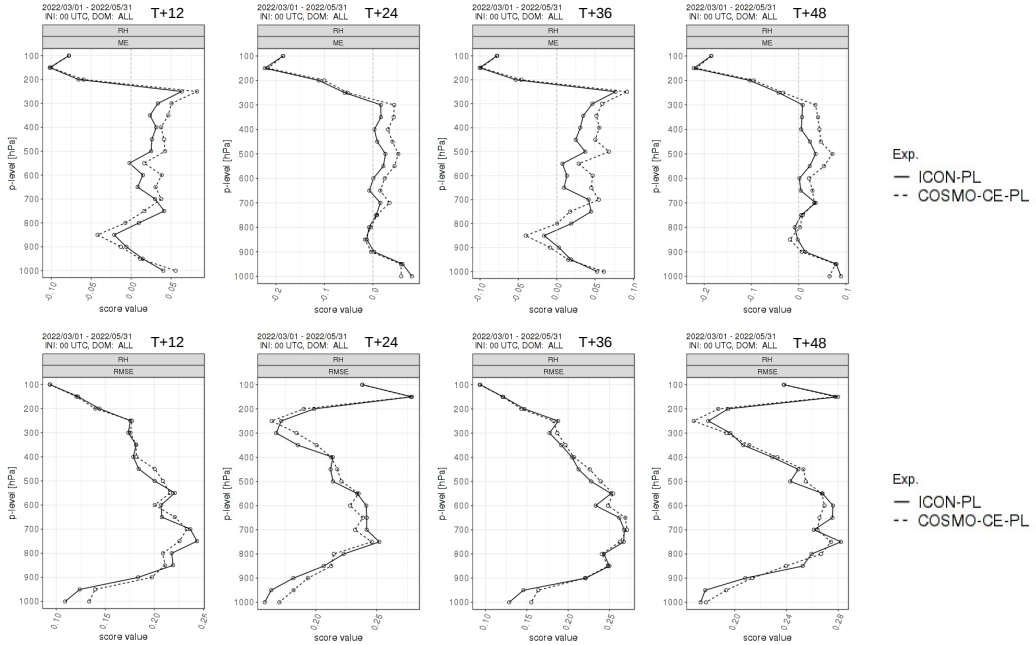


Figure 67: Relative humidity (0..1), MAM 2022, Continuous scores: ME and RMSE.

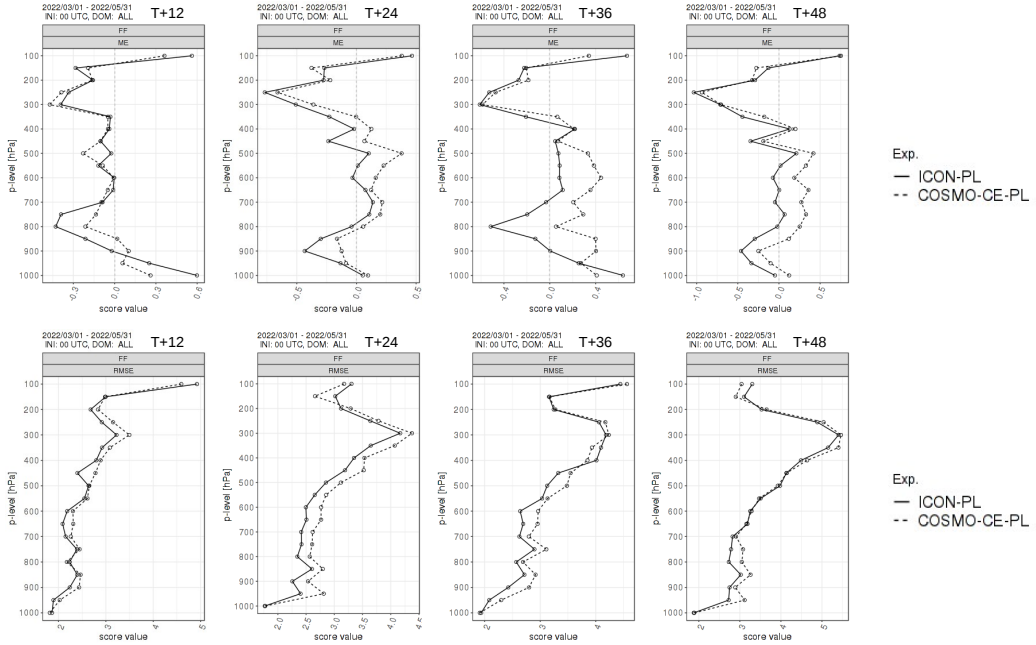


Figure 68: Wind speed (m/s), MAM 2022 2021, Continuous scores: ME and RMSE.

7.4 Conclusions

The verification results of ICON-PL for selected parameters are presented in the report. In terms of surface parameters, ICON-PL overall performs better than COSMO-CE-PL. In all seasons and for all parameters and lead times (except for surface pressure in the first few hours of the forecast) ICON-PL has a reduced or similar RMSE compared to COSMO-CE-PL. 2m temperature (T2M) is generally overestimated during summer in both models, while it is underestimated in winter. In Autumn there is a clear diurnal cycle to the bias in both models with a nighttime warm bias and a daytime cold bias. The largest reduction in bias

in ICON-PL compared to COSMO-CE-PL occurs at nighttime in Summer and Spring and in daytime in Autumn. In winter ICON-PL has removed the diurnal cycle bias. Dew point temperature has a spin down in the first couple of hours of the forecast in Summer and Autumn in COSMO-CE-PL. ICON-PL has a reduced spin down (better) but has a greater positive bias (worse) in late afternoon and evening in Spring an Autumn. Total cloud cover has diurnal cycle in the bias in both models. COSMO-CE-PL overestimates cloud cover at nighttime in all seasons and underestimates cloud cover during the daytime except Winter. ICON-PL increases cloud amounts during daytime in all seasons except Winter, where ICON-PL has less cloud than COSMO-CE-PL. In general ICON-PL has a reduced bias compared to COSMO-CE-PL. Surface pressure tends to have a negative bias in both models. ICON-PL has a reduced bias compared to COSMO-CE-PL in Spring and Summer. In Autumn and Winter both models have rather different bias characteristics, but ICON-PL is always negatively biased. Wind speed in ICON-PL has a reduced diurnal cycle of bias compared to COSMO-CE-PL in Spring and Autumn. In Winter, wind speed is always stronger in ICON-PL than in COSMO-CE-PL. Wind direction is positively biased (veered) in ICON-PL and negatively biased (backed) in COSMO-CE-PL in all seasons. 6-hourly precipitation is more skillful in ICON-PL at drizzle and light rain thresholds than COSMO-CE-PL. For the 5mm threshold it is less clear which model is more skillful. For the 10mm threshold the results are quite noisy. In terms of upper air verification, ICON-PL overall performs better than COSMO-CE-PL. ICON-PL has a reduced or similar RMSE compared to COSMO-CE-PL for temperature and wind speed in all seasons and lead times. For relative humidity, the improvement is less clear, but ICON-PL has reduced RMSE values at 950hPa and 1000hPa and generally similar values at other levels. For temperature, ICON-PL is warmer than COSMO-CE-PL in the upper to mid-troposphere (300 hPa to 550hPa) and is colder than COSMO-CE-PL in the mid and lower troposphere (650hPa to 850hPa). Winds through much of the depth of the troposphere (from 450hPa to 950hPa) are weaker in ICON-PL compared to COSMO-CE-PL at T+24, T+36 and T+48. For relative humidity, ICON-PL is drier than COSMO-CE-PL in the upper and mid-troposphere (300hPa to 700hPa). The comparison with COSMO-CE-PL shows (overall) a better performance of the ICON-PL model.

ICON-PL is expected to be fully operational in September 2022. There are no plans to discontinue the operational run of COSMO models. COSMO-CE-PL (COSMO-EULAG) will continue to be supported and developed (according to the available resources).

Acknowledgement

ICON-PL and COSMO-CE-PL Models scaling computations were carried out at the Centre of Informatics Tricity Academic Supercomputer and Network.

8 Results for Romania

A. Iriza-Burcă (1), B.A. Maco (1), R. Dragomir (1), Ş. Gabrian (1,2),
Ş. Dinicilă (1,3), M. Bogdan (1), T. Bălăcescu (1) and R.C. Dumitrache (1)

1. *National Meteorological Administration, Romania*
2. *Politehnica University of Bucharest, Faculty of Engineering in Foreign Languages*
3. *University of Bucharest, Faculty of Physics*

The present report aims to evaluate the performance from the ICON-LAM model (Zängl et al., 2015) integrated for Romanian territory for a year (December 2020 - November 2021). Statistical scores for 2 meter temperature, 10 meter wind speed, surface pressure, cloud cover and 6-hour cumulated precipitation were computed using the MEC+Rfdbk verification system. Some results for upper air verification are also presented. For ICON-LAM, scores are computed for up to 78 hours anticipation, with comparative analysis against COSMO-RO-2.8km (operational) for the first 30 hours of forecast the results. Both models are integrated at 2.8 km resolution. In general, a higher accuracy from ICON-LAM is showcased both for surface and upper air parameters. For precipitation, the two models are in agreement for lower thresholds, with a slight worsening of results from ICON-LAM for higher thresholds, that require further investigation.

8.1 Setup ICON-RO-2.8km

Starting with January 2020, ICON-RO-2.8km is running in the National Meteorological Administration operational for 00 UTC. Since the operational implementation of the model, efforts are being dedicated to minimizing the computational resources employed for this task, implementing more operational runs of ICON (tests for the 12 UTC run) and implementing new versions of the model.

For the present report, ICON 00 UTC model runs are evaluated against results from the COSMO-RO-2.8 km 00 UTC runs, available from operational activity. A summary of the configurations for the two models is presented in table 3. Integration domains are shown in figure 69.

For the surface parameters, observations for the entire period (December 2020 - November 2021) are used from all available Romanian synoptic stations. The results presented here are obtained using the integrated MEC+Rfdbk verification system, implemented in NMA in the frame of the CARMA Priority Project (Iriza-Burcă et al., 2020, 2021, 2022b). Some verification results for Romanian territory (JJA2020 summer season) were also presented in the paper by Iriza-Burcă et al. (2022a).

Table 3: Set-up of ICON-RO-2.8km and COSMO-RO-2.8km for Romanian territory.

	ICON-RO-2.8km	COSMO-RO-2.8km
Version	2.3.0	5.06_1
Resolution	0.025 / 2.8km	
IC/LBC	ICON global (3 hourly)	COSMO-RO-7km (hourly)
DA	no	nudging, first 6 hours
Vert. lev.	65	50
dt	24s	26s
anticipation	+78 hours	+30hours
grid	Unstructured triangular grid, output interpolated to regular lat/lon grid	rotated lat/lon
grid points	147 260	105 051 (361x291)

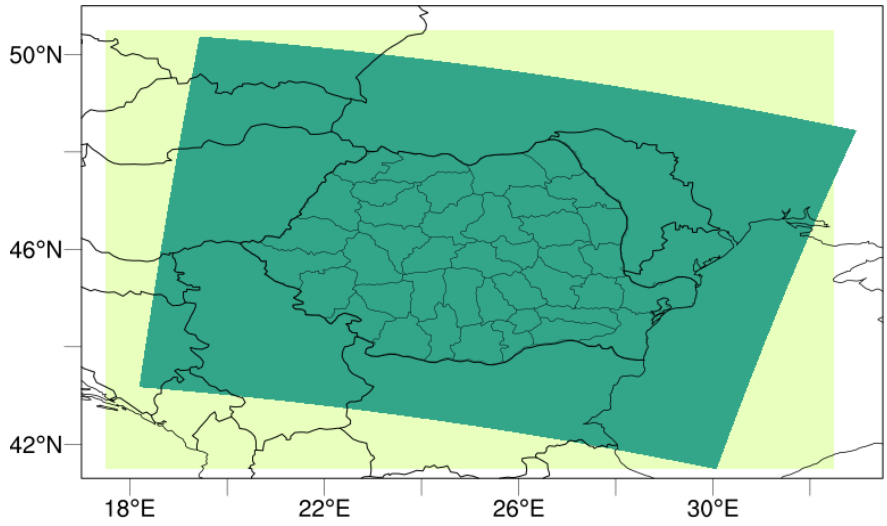


Figure 69: Integration domains for ICON-RO-2.8km (light green) and COSMO-RO-2.8km (dark green).

Some verification for upper air parameters (temperature and wind speed) has also been performed starting with July 2021, for a limited number of observations. Scores were computed for pressure levels between 1000 and 100 hPa, taking into account all TEMP observations in the integration domain.

Scores are computed seasonally, as follows: DJF2021 (December 2020 - January 2021 - February 2021), MAM2021 (March 2021 - April 2021 - May 2021), JJA2021 (June 2021 - July 2021 - August 2021) and SON2021 (September 2021 - October 2021 - November 2021). For upper air verification, scores for JJA2021 are computed only for July and August.

Due to the requirements of the system to use the same observations when two models are compared, two seasonal scores are computed for 78 hours anticipation (only for ICON-RO-2.8km) and for 30 hours anticipation when both models are used for the comparison.

8.2 Surface Verification

2 meter temperature: For 2 meter temperature (T2M, figures 70 - 73) the behaviour of the ICON-RO-2.8km and COSMO-RO-2.8km models integrated for Romanian territory is similar during all seasons, with a pronounced diurnal cycle (underestimation during the day and overestimation during the night). For both models, ME values are situated between -1.5 and 1.5 deg K. For the ICON configuration, the largest errors are observed for the summer (JJA2021, figure 72) and autumn (SON2021, figure 73) seasons, with smaller errors during winter (DJF2021, figure 70). For the cold season (DJF2021), ME values show a better performance of the ICON model for the first 30 hours of forecast, compared to COSMO. The results from the two models are comparable for the MAM2021 season (figure 71) during the first 30 hours, as well as for the JJA2021 and SON2021 seasons during the first 18 hours of forecast. For the last two seasons (JJA2021 and SON2021), ICON-RO-2.8km shows a reduction of ME values compared to COSMO-RO-2.8km for the forecast interval +18 hours up to +30 hours.

For both models, the amplitude of errors is situated between 1.5 and 2.5 K, with slightly higher errors during the day. However, for all seasons, ICON-RO-2.8km outperforms COSMO-RO-2.8km in terms of error amplitude, especially during the day. The difference in RMSE values in favour of ICON-RO-2.8km is especially visible during the DJF2021 and MAM2021 seasons. For the ICON integration, a small systematic increase in errors can be observed especially for the later anticipations (+60 hours).

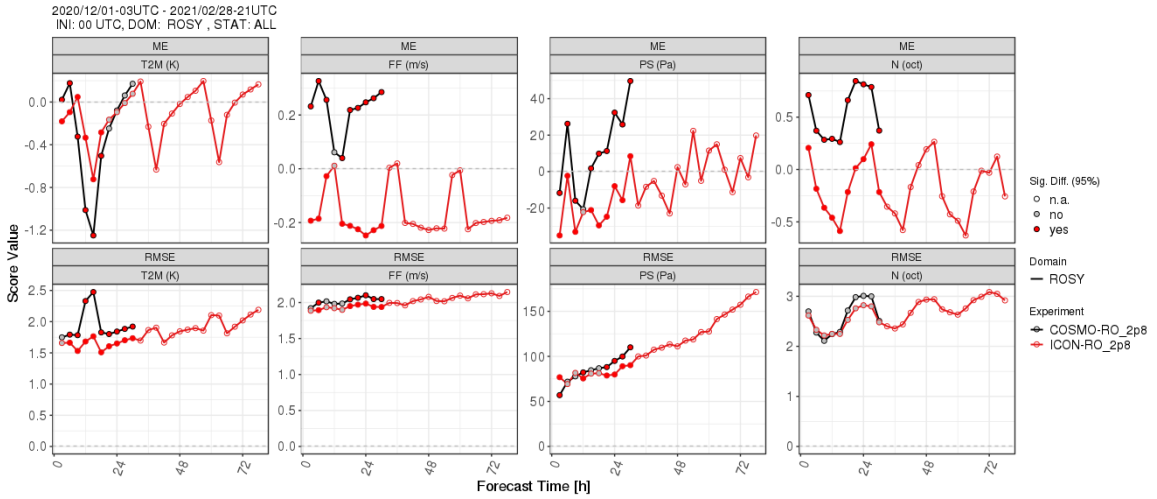


Figure 70: **Season DJF2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: ME (top) and RMSE (bottom) for all Romanian synop stations; left to right: T2M (K), FF (m/s), PS (Pa), N (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

10 meter wind speed: With regards to 10 meter wind speed (FF, figures 70 - 73), ICON-RO-2.8km shows a systematic reduction of errors which is visible in both ME and RMSE values. ME values for this parameter are situated between -0.4 and 0.5 m/s for both models. ICON-RO-2.8km generally outperforms COSMO-RO-2.8km in terms of error amplitude; this advantage of ICON over COSMO is slightly less visible for the DJF2021 season (figure 70) season, where errors from the two models are comparable. However, the general behaviour of the models in forecasting this parameter is different. The tendency of the COSMO-RO-2.8km model is that of overestimating forecasted values compared to observations. This behaviour is especially visible during the JJA2021 (figure 72) and SON2021 (figure 73) seasons, when COSMO shows strong overestimations for the entire forecast period of 30 hours. On the other

hand, as was the case for T2M, ICON-RO-2.8km again displays a diurnal cycle, usually with underestimation during the night and overestimation during the day for all seasons except DJF2021, when the model generally underestimates FF values for the entire forecast period (up to 78 hours).

For both models, the amplitude of errors is quite low, generally up to 2 m/s, with only a few values above this threshold from the COSMO-RO-2.8km simulations during the DJF2021 and JJA2021 seasons, or from ICON-RO-2.8km for the later anticipations (+60hours) for the DJF2021 season. However, for this parameter the amplitude of errors is low for all seasons. Similar to the results for T2M, ICON-RO-2.8km outperforms COSMO-RO-2.8km in terms of error amplitude for the entire period, with a small systematic increase in errors observed for the later anticipations (+60 hours).

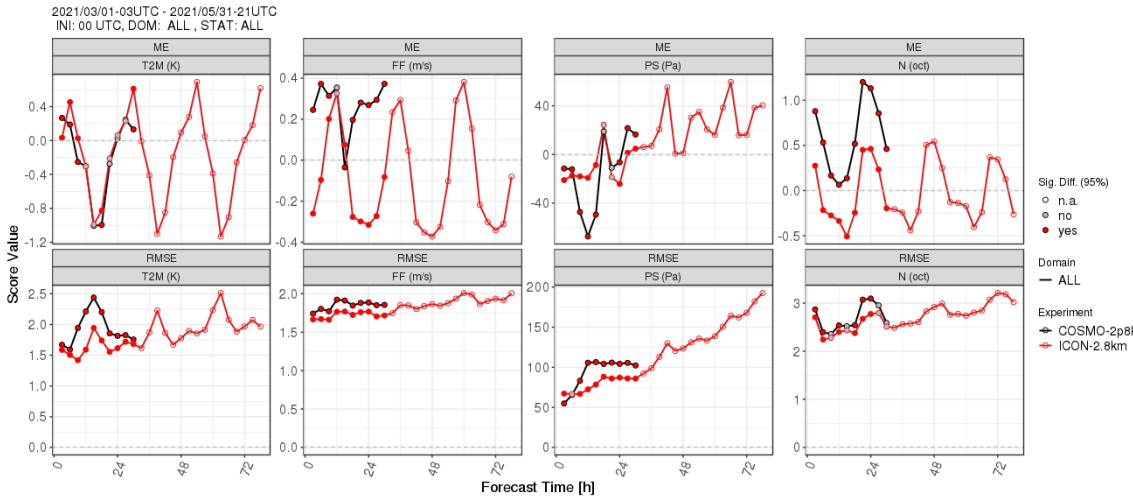


Figure 71: **Season MAM2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: ME (top) and RMSE (bottom) for all Romanian synop stations; left to right: T2M (K), FF (m/s), PS (Pa), N (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

Surface Pressure: The added value of ICON-RO-2.8km forecasts over the COSMO-RO-2.8km ones is slightly less visible for surface pressure (PS). ME values for the DJF2021 (figure 70) season suggest a slightly better performance from the COSMO integration during the day, while the results between the two models are generally comparable for the night time periods. A reduction in ME values from ICON-RO-2.8km compared to COSMO-RO-2.8km is mostly visible during the day for the MAM2021, JJA2021 and SON2021 (figures 72 - 73) seasons. An underestimation of surface pressure (PS) values can be observed from both models for the JJA2021 and SON2021 seasons. For the other two seasons, this behaviour is visible from the COSMO-RO-2.8km integration only during the day and from ICON-RO-2.8km during the first 51 forecast hours (DJF2021) and 30 forecast hours respectively (MAM2021). After this point, a change in the behaviour of the ICON model can be observed, with mostly overestimation of the values forecasted for this parameter.

RMSE values for PS are generally situated below 100Pa for the first 30 hours of forecast (with only a few larger values from the COSMO integration). However, for the longer simulations using ICON-RO-2.8km, we observe a continuous increase in the amplitude of errors with the anticipation time in forecasting this parameter (up to 200Pa during the MAM2021 season). Similar to the results for T2M and FF, ICON-RO-2.8km generally outperforms COSMO-RO-2.8km in terms of error amplitude for the entire period.

Total Cloud Cover: Similar to the results for PS, those for total cloud cover (N) also

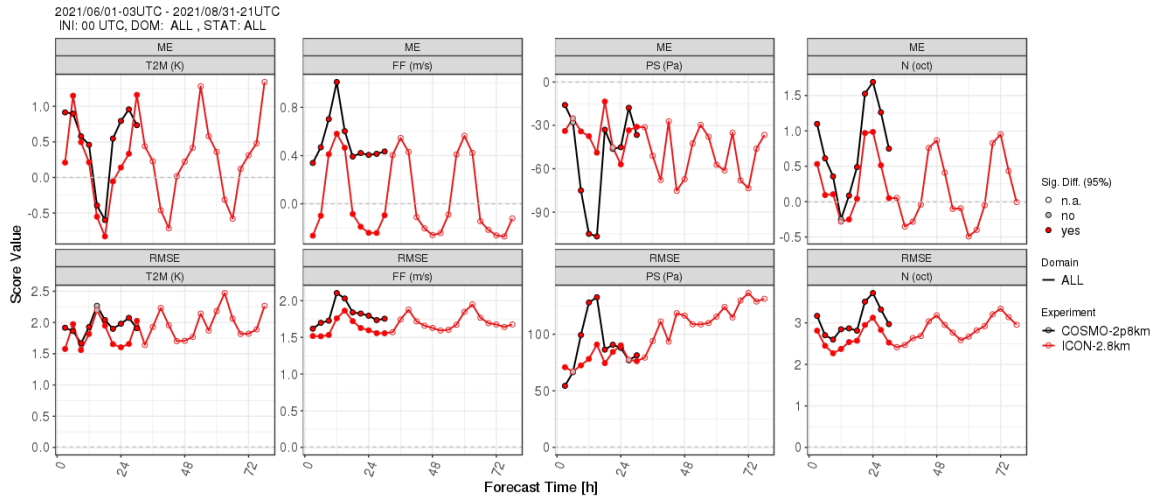


Figure 72: **Season JJA2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: ME (top) and RMSE (bottom) for all Romanian synop stations; left to right: T2M (K), FF (m/s), PS (Pa), N (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

divided between the two models. As in the case of FF, the general tendency between the models in forecasting cloudiness is different. The general behaviour of the COSMO-RO-2.8km model is that of overestimating values forecasted for this parameter for all seasons (with the exception of a short daytime interval for the JJA2021 period). On the other hand, ME values suggest a general tendency of the ICON-RO-2.8km model to underestimate cloudiness values for all seasons (again, with the exception of a short daytime interval for the JJA2021 period).

For this parameter, RMSE values between 2 and 3.5 octa suggest a high amplitude of errors (figures 70 - 73). Again, as for the previous parameters analysed, ICON-RO-2.8km outperforms COSMO-RO-2.8km in terms of error amplitude for the entire period. In general, the results for cloudiness indicate a better performance from the ICON-RO-2.8km model during night time, while for the day time COSMO-RO-2.8km seems better during the day for most of the analysed period.

In conclusion, for the surface parameters, with regards to ME, the comparison between the two models suggests that ICON-RO-2.8km generally outperforms COSMO-RO-2.8km, with slightly less visible results for PS. Some ME values also suggest a general bias from the COSMO-RO-2.8km model, while results from ICON-RO-2.8km are generally better centred around 0. On the other hand, RMSE values suggest a clear improvement from the ICON-RO-2.8km model compared to COSMO-RO-2.8km with regards to amplitude of errors for most continuous parameters analysed for the first 30 hours of forecast, during all seasons.

Total Cloud Cover - Categorical Scores: Categorical scores (ETS and FBI) for cloud cover (thresholds ≥ 1, ≥ 4 and ≥ 7) show an agreement between the two models in forecasting this parameter, with some improvement from ICON-Ro-2.8km compared to COSMO-RO-2.8km (figures 74 - 77).

For the DJF2021 season (figure 74), the tendency of the COSMO-RO-2.8km model is generally to overforecast for the entire period (up to +30 hours). For the ICON-RO-2.8km model, a diurnal cycle is visible in FBI results, with overforecasting only during the night time for all anticipations, with higher variations for the lower thresholds.

For the MAM2021 and JJA2021 seasons (figures 75 and 76), the two models behave similarly,

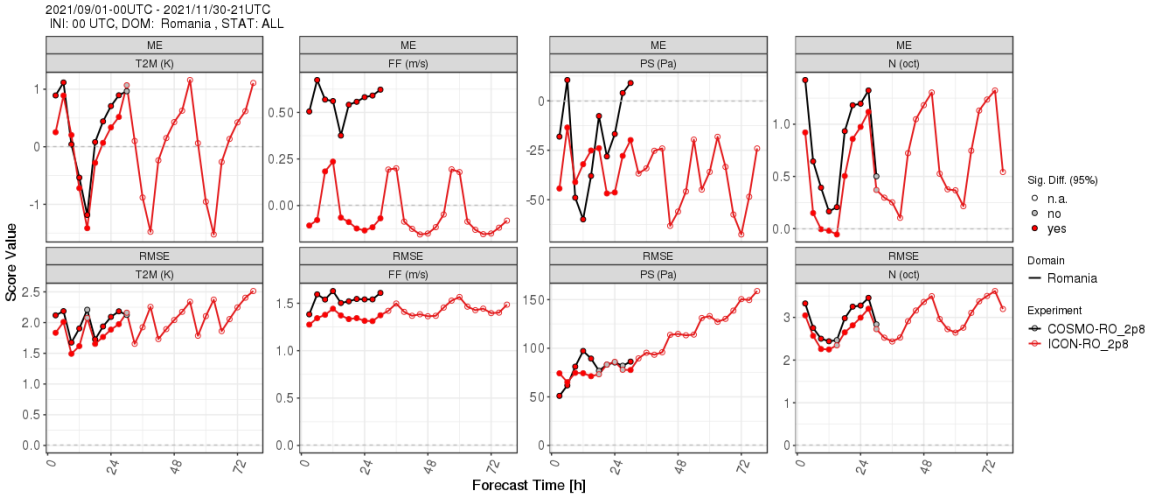


Figure 73: **Season SON2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: ME (top) and RMSE (bottom) for all Romanian synop stations; left to right: T2M (K), FF (m/s), PS (Pa), N (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

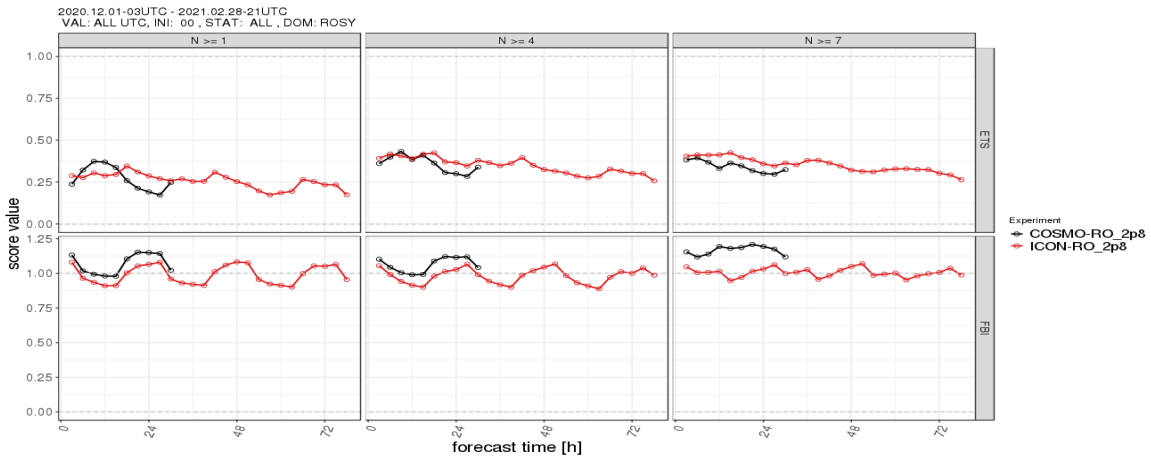


Figure 74: **Season DJF2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for N: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $N \geq 1, \geq 4, \geq 7$ (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

with the diurnal cycle visible in FBI results, that is underforecast during day time and overforecast during night time for the first 30 hours of forecast for the lower thresholds ($\geq 1, \geq 4$). Moreover, for ICON-RO-2.8km, this behaviour is extended to the entire forecast interval (up to 78 hours). For the higher threshold, both models overforecast for the entire period, during both seasons.

For the SON2021 season (figure 77), the tendency of both models is similar to that observed during the DJF2021 season, overforecast for the entire period (up to +30 hours COSMO, +78 hours ICON), with some overforecasting only during the night time for all anticipations. For the higher threshold, both models overforecast for the entire period, similar to the results obtained for the previous seasons.

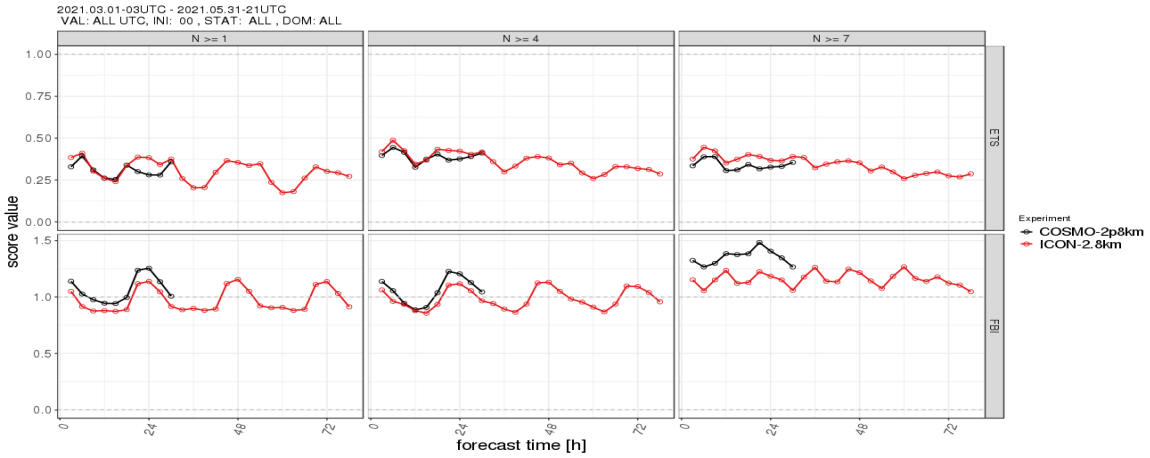


Figure 75: **Season MAM2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for N: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right:$N \geq 1, \geq 4, \geq 7$ (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

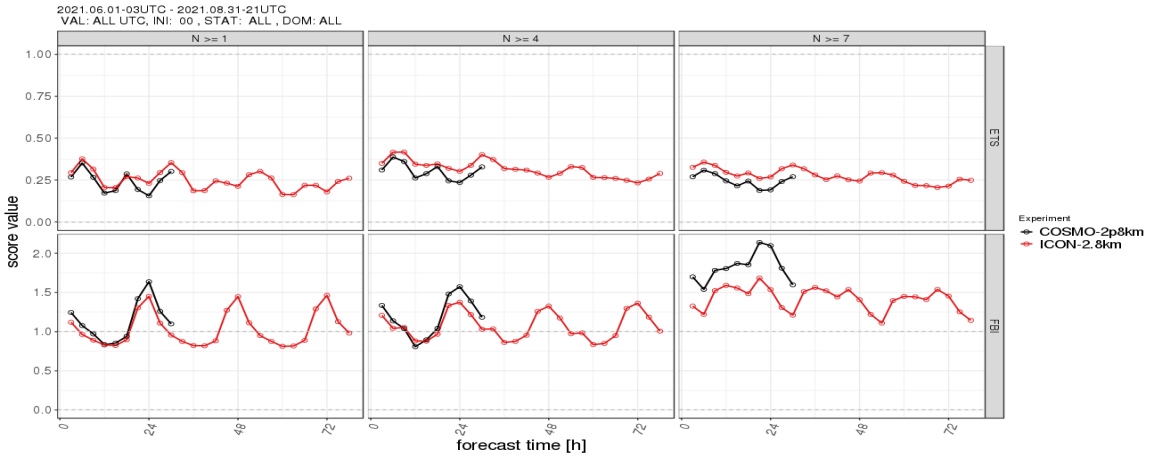


Figure 76: **Season JJA2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for N: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $N \geq 1, \geq 4, \geq 7$ (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

Although both models generally overforecast the events for this parameter, some slight differences are noticeable especially for the highest threshold, where ICON-LAM displays a better performance in forecasting for the 0-30 hours interval. Also from the ETS results, it is suggested that ICON-RO-2.8km displays a slightly better skill in forecasting cloudiness compared to COSMO-RO-2.8km, visible in particular for the higher thresholds (≥ 4 and ≥ 7).

Precipitation:

With regards to the forecast of precipitation (figures 78 - 81), FBI values for the DJF2021, MAM2021 and SON2021 seasons show a quite good performance from the models for the lower thresholds (0.8 mm/6h, 1 mm/6h and even 5mm/6h) during the first period of the forecast interval. Results are slightly less favourable for the JJA2021 season, where the performance of the two models is limited even for the lower thresholds, for the entire forecast period.

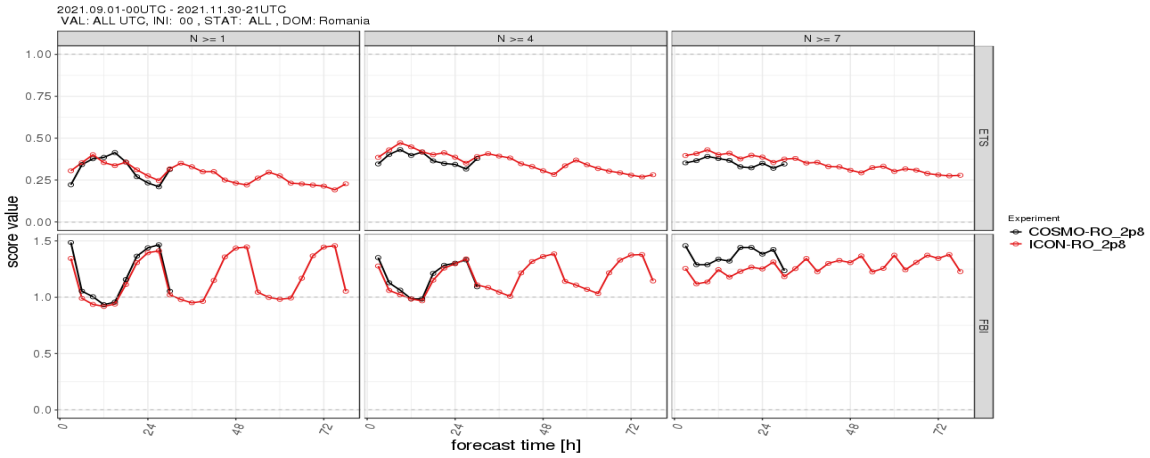


Figure 77: **Season SON2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for N: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $N \geq 1, \geq 4, \geq 7$ (octa); black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

For all seasons, for the higher thresholds (10mm/6h, 20 mm/6h) score values suggest a tendency of the model to generally overforecast the events, mostly for the beginning of the interval and especially for the threshold of over 20mm/6h. For the latter, the model clearly displays a behaviour of overforecasting precipitation events. For this threshold, some undercast periods are observed from ICON-RO-2.8km around +36 hours.

For all thresholds, for the DJF2021 (figure 78) season both models display a general tendency to overforecast for most of the interval (especially for the higher thresholds), with some small periods of underforecast around +12 and +24 hours anticipation.

For the MAM 2021 period (figure 79), the overforecasting behaviour is displayed by COSMO-RO-2.8km for all thresholds and by ICON-RO-2.8km for the lower thresholds. For the ICON integration, undercast periods are observed from ICON-RO-2.8km around +36 hours for the $\geq 10mm/6h$ threshold and for the 36-48 hours interval for the $\geq 20mm/6h$ thresholds.

Scores for the JJA2021 season (figure 80) show a different behaviour from both models, generally with overforecast during the day and underforecast during the night for the lower thresholds, from both models. The underforeacsting behaviour from ICON-RO-2.8km is slightly reduced compared to that of COSMO-RO-2.8km. For the highest threshold ($\geq 20mm/6h$), both models strongly overforecast for the entire period.

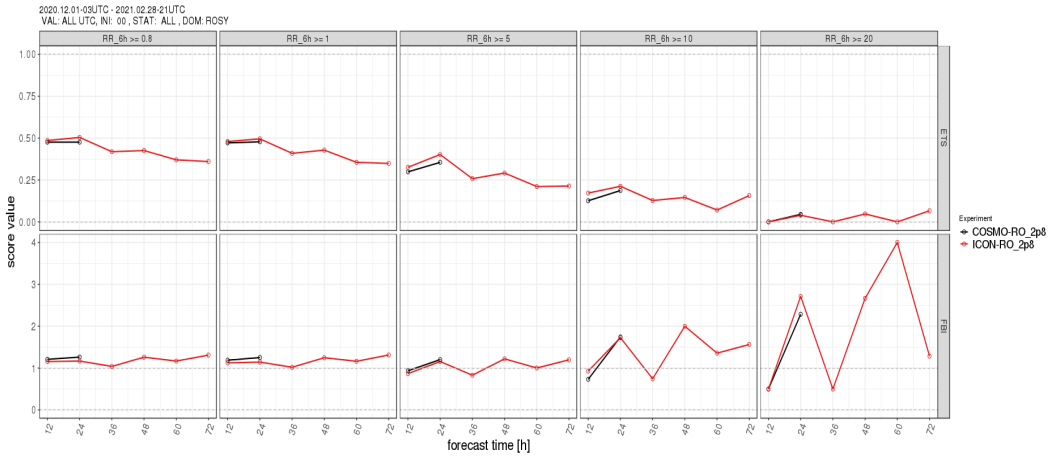


Figure 78: **Season DJF2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for 6-hour cumulated precipitation: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $RR6h \geq 0.8, \geq 1, \geq 5, \geq 10, \geq 20 mm/6h$; black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

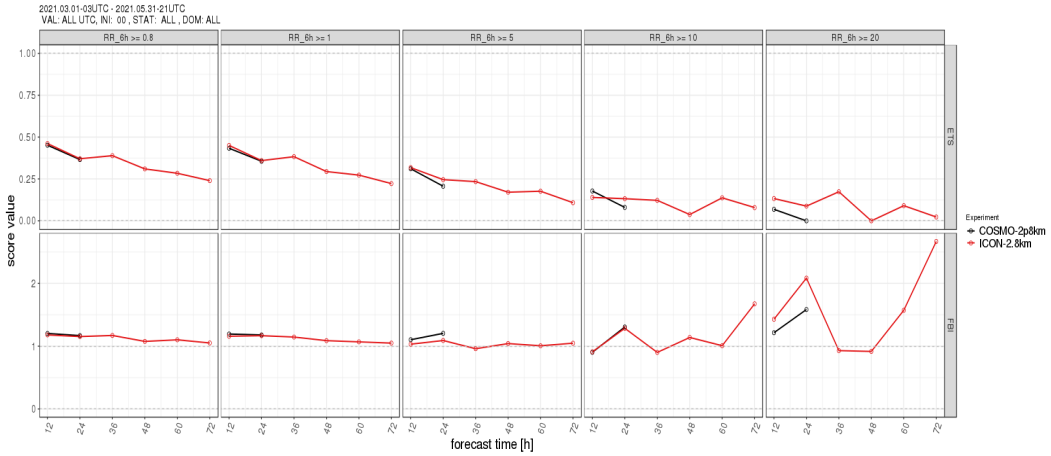


Figure 79: **Season MAM2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for 6-hour cumulated precipitation: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $RR6h \geq 0.8, \geq 1, \geq 5, \geq 10, \geq 20 mm/6h$; black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

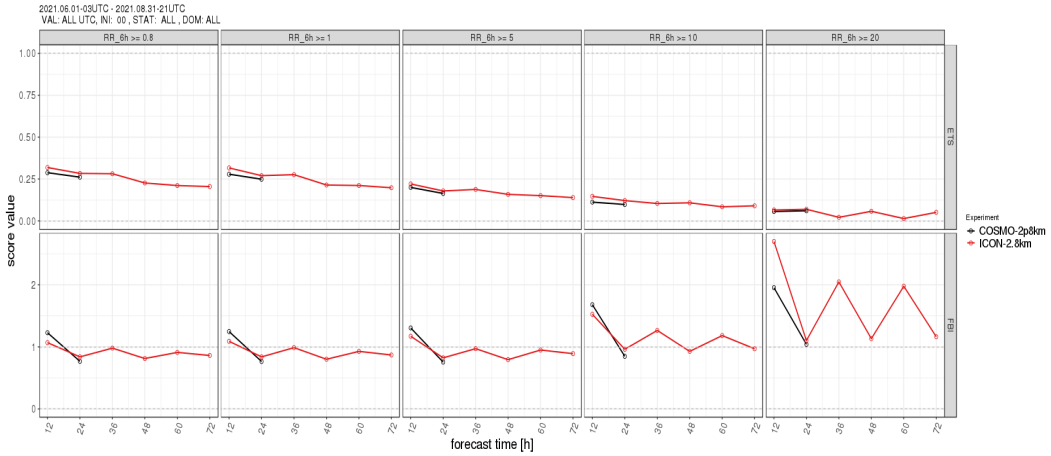


Figure 80: **Season JJA2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for 6-hour cumulated precipitation: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $RR6h \geq 0.8, \geq 1, \geq 5, \geq 10, \geq 20mm/6h$; black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

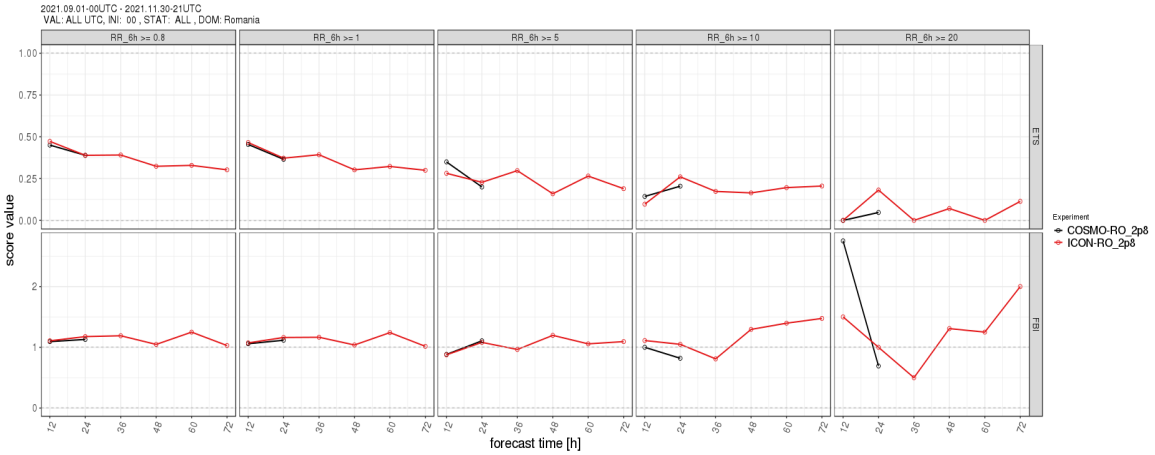


Figure 81: **Season SON2021**: Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: Categorical scores for 6-hour cumulated precipitation: ETS (top) and FBI (bottom) for all Romanian synop stations; left to right: $RR6h \geq 0.8, \geq 1, \geq 5, \geq 10, \geq 20mm/6h$; black - COSMO-RO-2.8km; red - ICON-RO-2.8km.

8.3 Upper Air Verification

Upper Air Temperature: Scores computed for upper air temperature (figure 82) for the JJA2021 and SON2021 seasons suggest a general tendency of both models to overestimate values forecasted for this parameter for the lower (up to 650 hPa) and higher (over 300 hPa) levels, with underestimation elsewhere, for the +12 hours anticipation. The same can be observed for the +24 hours anticipation, with the exception of the lowest levels (around 1000 - 900 hPa). A stronger overestimation of the temperature values is shown for the lower levels for both seasons (with the exception of the 100-900 hPa levels for SON2021, where the values are underestimated); the underestimations are less significant for both seasons. ME values suggest a slightly better performance from the ICON model, generally for both anticipations considered, with more significant differences for the levels up to 500 hPa. Lower RMSE values from the ICON model show an important reduction in amplitude

of errors compared to the COSMO model, again, more significant for the lower atmosphere, especially for the JJA2021 season.

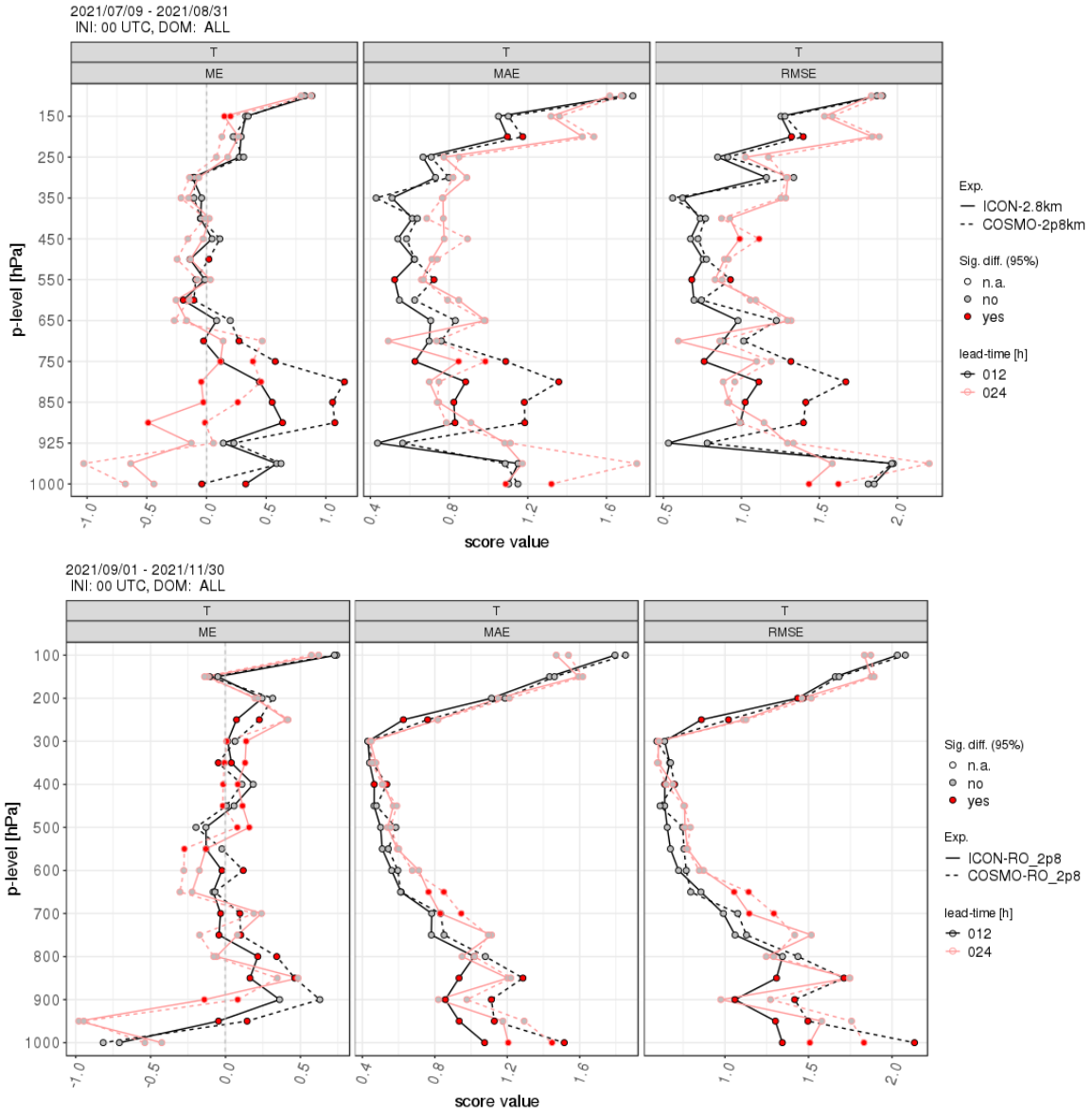


Figure 82: **Seasons JJA2021** (top) and **SON2021** (bottom): Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: scores for upper air temperature for all TEMP stations; left to right: ME, MAE and RMSE; black: +12 hours, pink: +24 hours; dashed - COSMO-RO-2.8km; continuous - ICON-RO-2.8km.

Upper Air Wind Speed:

Differences between the two models in forecasting upper air wind speed (figure 83) are smaller than in forecasting upper air temperature. For this parameter also, most significant differences between the two models are visible mostly in the lower (up to 850 hPa) levels. Both ICON and COSMO display a tendency to underestimate the values forecasted for upper air wind speed for both seasons and anticipations, with some slight overestimations for the middle and upper troposphere, especially for the SON season and 24 hours anticipation. The underestimating behaviour is more visible from the COSMO integration. Similar to the scores obtained for upper air temperature, RMSE values for wind speed from the ICON

model show a reduction in amplitude of errors compared to the COSMO model for both analysed seasons. Smaller RMSE values are observed from both models for the 12 hour anticipation.

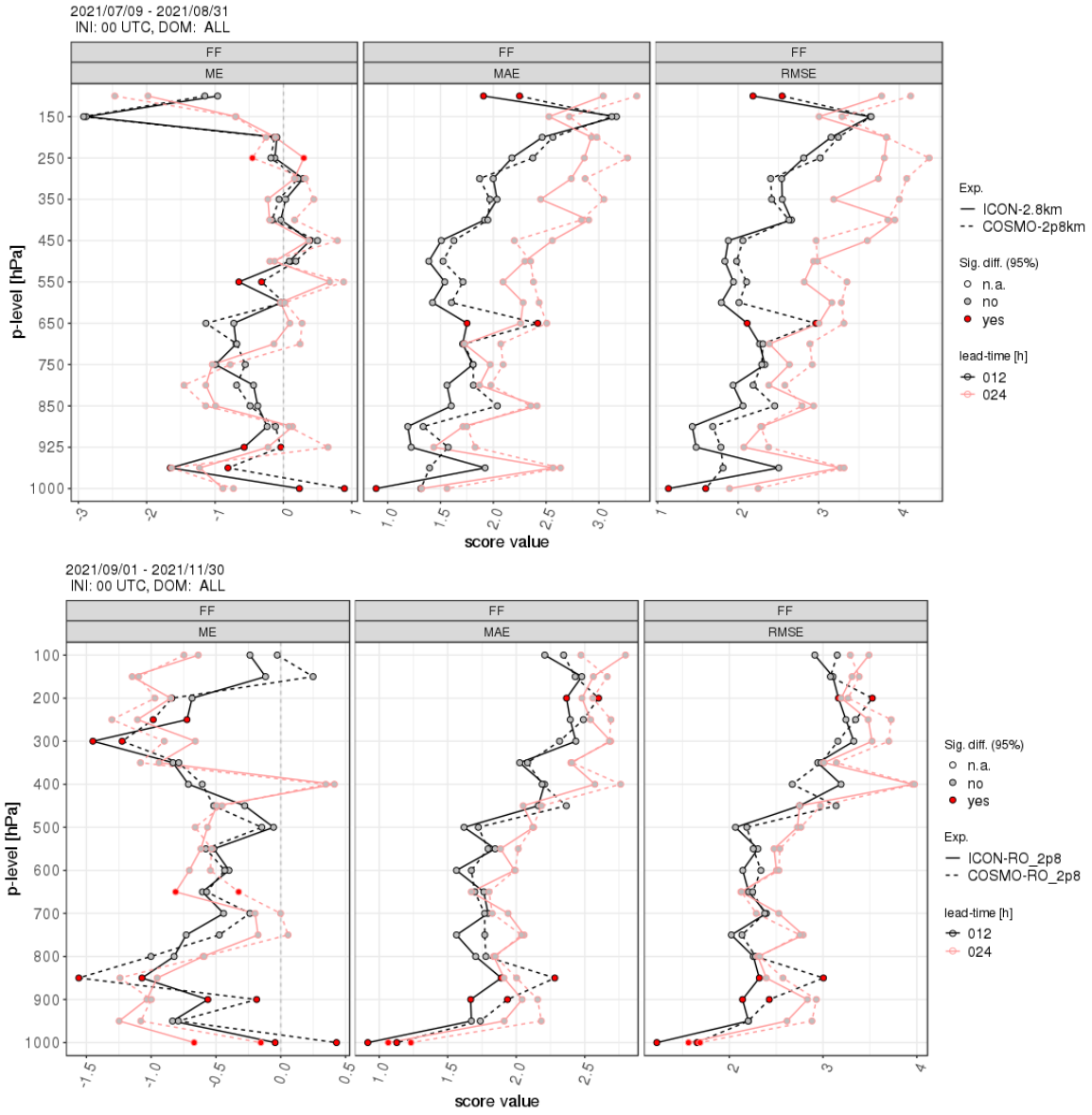


Figure 83: **Seasons JJA2021** (top) and **SON2021** (bottom): Comparison between ICON-RO-2.8km and COSMO-RO-2.8km: scores for upper air wind speed for all TEMP stations; left to right: ME, MAE and RMSE; black: +12 hours, pink: +24 hours; dashed - COSMO-RO-2.8km; continuous - ICON-RO-2.8km.

8.4 Considerations on Computational Costs

For the present verification, both models were run on the same type of processors (Intel based Xeon Gold, Infiniband connection).

For the configuration presented in Section 8.1, the COSMO-RO-2.8km model is run on 9 nodes with 32 processes/node and the integration time is $\approxeq 12 - 14min$. (Similarly, for a 90 hours forecast, integration time for the COSMO-RO-2.8km model is $\approxeq 41 - 43min$.)

Also for the configuration presented in Section 8.1, the ICON-RO-2.8km model is run on 9 nodes with 30 processes/node and the integration time is $\approxeq 61 - 63min$. In this case, it is important to consider the differences between the two integrations in terms of time step (24s for ICON and 25s for COSMO), forecast hours (30 for COSMO, compared to 78 for ICON) and extent of the integration domain (≈ 1.5 more grid points for ICON compared to COSMO) that account for the longer integration time required for the ICON-RO-2.8km configuration compared to that of COSMO-RO-2.8km. To summarize, taking these differences into account ICON-RO-2.8km takes 0.795 minutes per simulated hour while COSMO would take 0,790 minutes per simulated hour with the number of grid points scaled to the ICON number of grid points.

8.5 Conclusions

The results for surface parameters suggest a good performance of the ICON-LAM model integrated for Romanian territory during all seasons. Moreover, ICON-RO-2.8km generally outperforms COSMO-RO-2.8km in the analysed forecast interval of 30 hours for T2M and FF, with slightly reduced results for PS. This behaviour is visible during all seasons and especially from the reduction in the amplitude of errors. For cloudiness, slight differences are noticeable especially for the higher thresholds (≥ 4 and ≥ 7), where ICON-LAM displays a better performance in forecasting for the 0-30 hours interval and a slightly better skill, visible in ETS and FBI values.

For the ICON simulations, a systematic increase in errors can be observed in all analysed surface parameters for the later anticipations, during all seasons; this behaviour is especially visible in surface pressure.

With regards to the forecast of precipitation, FBI and ETS values show a slightly better performance from ICON-RO-2.8km compared to COSMO-RO-2.8km especially for the lower thresholds (0.2 mm/6h, 1 mm/6h and even 5mm/6h). For all seasons, a small reduction in over/underforcasting behaviour from ICON compared to COSMO and a slightly better skill is observed. However, the forecast quality for precipitation drops significantly for higher thresholds. For all analysed parameters, the quality of forecasts decreases towards the end of the forecast interval. Comparison between the two models in forecasting precipitation (first 24 hours) show a similar behaviour especially for the lower thresholds for all seasons, with more significant differences can be observed in forecasting events with quantities over 20mm/6h.

Scores computed for upper air parameters for the JJA2021 and SON2021 seasons suggest a similar behaviour of the ICON-RO-2.8km and COSMO-RO-2.8km models, with more significant differences for the lower troposphere. A more visible reduction both in error values and amplitude is shown in the case of temperature, in favour of the ICON model. For upper air wind speed, both models display a tendency to underestimate the forecast of this parameter, with better results from ICON-RO-2.8km compared to COSMO-RO-2.8km.

In conclusion, the evaluation of all seasons for Romanian territory showcase a higher accuracy from ICON-LAM compared to COSMO for both surface and upper air parameters, visible especially from the reduction in the amplitude of errors. The two models are in agreement for precipitation in the lower thresholds. Results from ICON-LAM for higher thresholds require further investigation that will be carried on with the implementation of new model versions.

9 Results for Brazil

G.R. Bonatti(1), R.R. dos Santos(1), Y.K.L. Kitagawa(1) and R.B. da Silveira(2)

1. *Instituto Nacional de Meteorologia, INMET, Brasília, DF, Brazil*
2. *Sistema de Tecnologia e Monitoramento Ambiental do Paraná - SIMEPAR*

The NWP system operated by the National Meteorological Institute of Brazil, INMET, is completing 20 years, starting in 1999 with the HRM model and currently has the COSMO model as the core system, both from Deutscher Wetterdienst, DWD. The fully automatic system uses the lateral boundaries from the ICON Global model and runs the limited area model for the entire South America, Central America and part of Pacific and Atlantic oceans. However, in the last years, with the discontinuity of COSMO model, INMET decided to move on by implementing the ICON-LAM for its regional applications. ICON-LAM is a brand new suit of NWP models, developed by DWD and the Max-Planck Institute of Meteorology (MPI-M) and applied for regions of the globe, through a specific interface to the global model data. This work describes the installation of the ICON-LAM at INMET computational park and the tuning of the namelist parameters for tropical and subtropical applications. In addition, an intercomparison was described between COSMO and ICON-LAM models, according to their respective configurations of grid, physics and dynamic characteristics. The forecasts of total daily precipitation of leadtime of 7 days were compared to precipitation analysis and reanalysis data from CPC-NOAA and ERA5-ECMWF, respectively, by using confusion matrix and statistical indices. The results were varied and interchangeable between both models over Brazilian territory on May 2022, in which ICON-LAM had a better accuracy for the initial leadtimes and decreasing performance for the last days of forecast. COSMO kept a consistent accuracy across the seven forecast days. The analysis and reanalysis data did not have significant differences, despite ERA5 smooths the rainfall magnitude in comparison to CPC. In addition, ERA5 had a better agreement with the COSMO and ICON-LAM models than CPC.

9.1 Scope of the work

The work is outlined as follows.

- Description of main ICON-LAM settings in the INMET's HPC system;
- Configuration of namelist parameters taking into consideration the weather evolution under tropical and subtropical climate regimes;
- The inclusion of the following NWP models: COSMO 7 km, and ICON-LAM 7 km;
- Description of the atmospheric synoptic conditions during May 2022;
- Comparing the forecast with analysis and reanalysis precipitation data from the Climate and Predicition Center of NOAA, CPC-NOAA (Climate Prediction Center, NOAA, 2022) and ERA5 from ECMWF (Hersbach, 2020) and
- Application of statistical verification procedures to the model outputs and the inter comparison of the results.

9.2 Methodology

Initially, ICON-LAM (version 2.6.4) was set at INMET's HPC system and a namelist with model parameters was used to tune it for simulations within the South America domain. Table 4 presents the main differences in settings between the COSMO (version 6.0) and ICON-LAM models, which are related to physics and dynamic characteristics. The description of each specific parameter can be found in the models' manual. Both models performed precipitation forecasts over May 2022 for leadtime of 7 days (168 hours) with horizontal resolution of approximately 7 km(0.0625° x 0.0625°) over South America and run by the Brazilian National Weather Service (INMET). The models were initialised with the lateral boundary from ICON global model at 00 UTC, which has a horizontal resolution of 13 km, 70 vertical layers, and temporal resolution of 3 hours. The COSMO model has been running at INMET since 2011 and the parameters presented in Table 4 have been adjusted over the years to the South America domain. Meanwhile, the ICON-LAM model started to be run at INMET in 2018 through the C2I project (Transition of COSMO to ICON) and the parameters were tuned using case studies for extreme weather events.

The results of the forecasts were compared with the Climate Prediction Center (CPC) and ERA5 datasets. The CPC Unified Precipitation Project is underway at National Oceanic and Atmospheric Administration (NOAA), USA, and produces hourly precipitation with a horizontal resolution of 0.5° x 0.5° almost in real time. In the meantime, ERA5 is a reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF), with higher horizontal and temporal resolutions (0.25° x 0.25° at every hour) than its previous generation, which is the ERA-Interim reanalysis with spatial resolution of 0.75° at every 6 hours. Both datasets were regridded to a regular lat-lon grid of 0.0625° x 0.0625°, the same horizontal resolution of COSMO and ICON-LAM models using the Climate Data Operators (CDO). In the same way, the data were converted from hourly to daily precipitation and a mask was used to consider only the grid points within the Brazilian territory (approximately 182,000 grid points). Once the analysis and reanalysis data were preprocessed, each grid point was paired and the statistical indices and the confusion matrix were computed.

The confusion matrix was created as followed to compute the accuracy index:

- Values less than 1 mm were considered as no rainfall. Thus, if both points (modelled and observed) did not detect any precipitation, this grid point would be classified as true negative (TN) because the model did not predict rainfall and, in fact, it did not rain.
- Values greater than 1 mm were considered as a occurrence of rainfall. Thus, if both points (modelled and observed) detect precipitation, this grid point would be classified as true positive (TP) because the model predicted rainfall and, in fact, it rained.

The other two situations are related to (i) when the model predicted the occurrence of rainfall, but it did not actually occur (false positive - FP), meaning that the models are overpredicting the rainfall. On the contrary, (ii) when the model did not predict the occurrence of rainfall, but it actually occurred, the classifier is labelled as false negative (FN), meaning that the models are underpredicting the rainfall. The accuracy analysis counts the TN and TP results compared to the overall sample in order to show how often the classifier is correct. In addition, for further investigation, a quantitative analyses together with the matrix confusion was implemented, by which if the models hit the occurrence of precipitation (TP), these rainfall were either overestimated (TP+) or underestimated (TP-). However, the main focus of the work was to evaluate the ability of the models to predict categories of the distribution

Table 4: Comparison between COSMO and ICON-LAM configurations.

Parameter	**Description**	**Namelist**	**ICON**	**COSMO**
heightdiff_threshold	Height difference threshold for additional diffusion	extpar_nml	2250.	-
tkhmin	Scaling factor for minimum vertical diffusion for heat and moisture	turbdiff_nml	0.75	0.40
tkmmin	Scaling factor for minimum vertical diffusion for momentum	turbdiff_nml	0.75	0.40
rat_sea	Ratio of laminar for scalars over sea and land scaling factors or scalars over sea and land	turbdiff_nml	7.	10.
tur_len	Asymptotic maximal turbulent distance	turbdiff_nml	500 m	500 m
pat_len	Effective length scale of thermal surface patterns controlling TKE-production	turbdiff_nml	500 m	500 m
icldm_turb	Water cloud representation in turbulence	turbdiff_nml	2	2
q_crit	Critical value normalized supersaturation	turbdiff_nml	4.0	4.0
imode_tkesso	SSO source term for TKE-production	turbdiff_nml	2	1
itype_sher	Type of shear forcing used in turbulence	turbdiff_nml	2	0
tune_gkwake	low level wake drag constant	nwp_tuning_nml	0.80	0.80
tune_gkdrag	gravity wave drag constant	nwp_tuning_nml	0.0	0.075
tune_box_liq_asy	Asymmetry factor for liquid cloud cover diagnostic	nwp_tuning_nml	3.25	-
ldetrain_conv_prec	Activate detrainment of convective rain and snow	nwp_phy_nml	.true.	.true. (l_conv)
lshallowconv_only	use shallow convection only	nwp_phy_nml	.false.	.false.
dtime	Time step	run_nml	60.	60.
ndyn_substeps	Number of dynamics substeps	nonhydrostatic_nml	5	-
dt_rad	Radiation time interval	nwp_phy_nml	1080.	3600. (nincrad)
dt_conv	Convection time interval	nwp_phy_nml	360.	240. (nincconv)
dt_sso	SSO time interval	nwp_phy_nml	360.	300. (nincsso)
num_lev	Number of full levels	run_nml	60	50
damp_height	Start of Rayleigh damping vertical wind	nonhydrostatic_nml	18000.*	18000.* (rdheight)
top_height	Height of model top	sleve_nml	30000.*	30000.*
$*$ Tropical setup				

of values rather than the exact value of the accumulated total rainfall. Thus, the precipitation magnitude was classified according to Table 5, and these labels were compared for each grid point.

Finally, two other statistical indexes were considered: Pearson correlation coefficient (r) and index of agreement (IOA). Both quantities measure the association and agreement between predicted and observed data, which vary from zero to one. Values close to zero indicate the absence of correlation and those close to one indicate strong correlation. The bias was also

Table 5: Comparison between COSMO and ICON-LAM configurations.

Rainfall Classification	**Rainfall Intensity**
No rainfall	$<$ 1mm
Light (L)	1-5 mm
Moderate (M)	5-15 mm
Heavy (H)	15-50 mm
Intense (I)	50-100 mm
Extreme (E)	$>$ 100 mm

calculated to compare the deviations between COSMO and ICON forecasts.

9.3 Computing architecture and processing efficiency of ICON-LAM and COSMO models at INMET

The models were both processed in the cluster HPE Apollo 6000, with the following architecture:

- Intel Xeon Platinum 8260 (35.75M Cache, 2.40 GHz (Turbo 3.90 GHz))
- 72 nodes (with 48 cores) with total of 3456 cores
- 192GB to 384GB RAM per node
- INTEL compiler Fortran 19.1.3.304 and MPI INTEL

The table 6 describes the time for ICON-LAM and COSMO, both having the grid resolution of 7 km and considering the leadtime of 24h. The number of threads varied from 96 to 2176. We noticed that as the number of threads increases the ICON-LAM performs better than COSMO, considering that the cluster has 48 cores.

The table 7 and figures 84 and 85 give the results of the combinations of the number of nodes and number of threads, but keeping the total number of threads equal to 720 and limiting the number of nodes to 72. The results show that the ICON-LAM is 32 % faster than COSMO in the best combination for nodes and threads ($60x12$) and 84 % faster in the worst combination ($15x48$), meaning that ICON-LAM if more efficient of COSMO on the use of machine architecture. The best result of ICON-LAM was reached by using the hybrid option (OMP and MPI), resulting in a processing time of 2700 seconds for leadtime of 168h, considering also the option `OMP_NUM_THREADS`=4 and $12x72$ (nodes and threads).

Table 6: Processing times in seconds of ICON-LAM and COSMO models, both with 7km grid size and considering leadtime of 24h.

NThreads	ICON 7km (s)	COSMO 7km(s)	ICON/COSMO
96	5579	6440	1.15
128	4609	5723	1.24
256	2369	2970	1.25
384	1591	2358	1.48
512	1202	2151	1.79
640	972	1804	1.86
768	822	1264	1.54
896	717	1098	1.53
1024	636	951	1.50
1152	575	829	1.44
1280	619	1095	1.77
1408	577	1085	1.88
1536	533	990	1.86
1664	576	1198	2.08
1792	544	1201	2.21
1920	516	1327	2.57
2048	514	1167	2.27
2176	458	1125	2.46

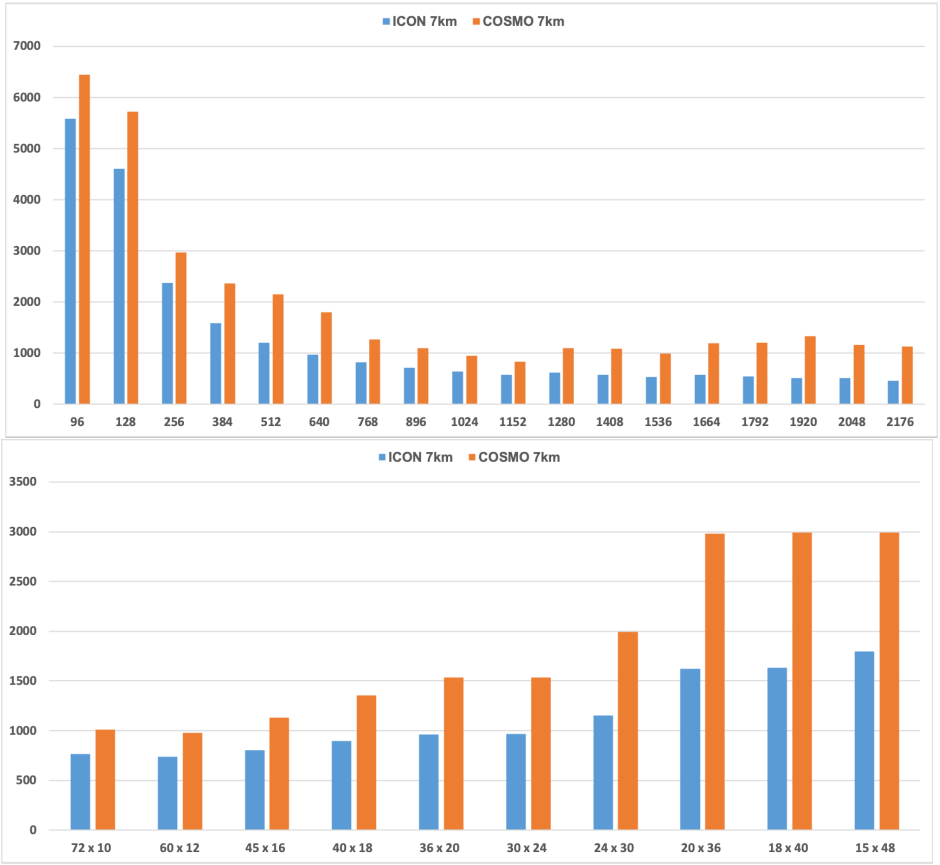


Figure 84: Histograms of processing times in seconds of ICON-LAM and COSMO models, both with 7km grid size and considering leadtime of 24h: combinations described in table 6 (top) and combinations of table 7 (bottom) and limiting the number of threads to 720.

Table 7: Results of processing times for different combinations of threads per nodes for ICON-LAM and COSMO models, both with grid size of 7km and considering lead time of 24h. The total number of threads was 720 and the maximum number of nodes was 72.

Node x Thread	**ICON 7km (s)**	**COSMO 7km(s)**	**ICON/COSMO**
72 x 10	766	1011	1,32
60 x 12	738	979	1,33
45 x 16	805	1131	1,40
40 x 18	897	1356	1,51
36 x 20	962	1533	1,59
30 x 24	967	1538	1,59
24 x 30	1154	1993	1,73
20 x 36	1621	2983	1,84
18 x 40	1633	2991	1,83
15 x 48	1796	2990	1,66

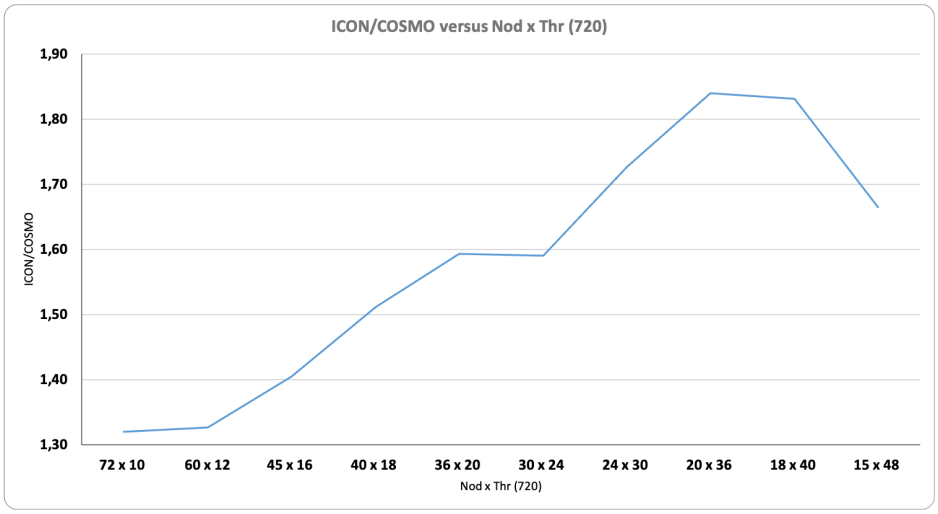


Figure 85: Ratio of processing time between ICON-LAM and COSMO according the configuration of nodes and threads, by limiting the total number of threads in 720.

9.4 The meteorological conditions

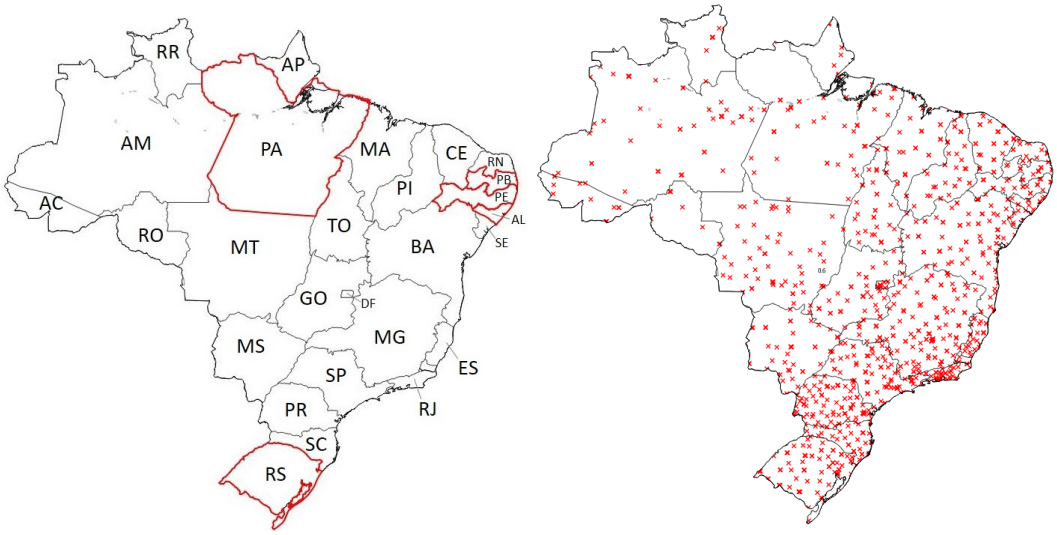


Figure 86: The location of the Brazilian states (left) and the location of the surface meteorological stations over Brazil (right).

We evaluated the daily precipitation for May 2022, when several heavy precipitation events occurred in Brazil and the observations were above the climatological normal, especially in the states of Paraíba (PB), Pernambuco (PE), and Alagoas (AL), which are located in the Northeast of Brazil, Pará (PA) in the North region of Brazil, and Rio Grande do Sul (RS) in Southern Brazil. The location of the Brazilian states is shown in Figure 86, and the aforementioned states are highlighted in red. These extreme weather events caused floods, landslides, power outages, damage to road structures, and material and human losses. They are highly disruptive to the urban and rural systems and pose a major challenge to the government and society.

Different weather systems contributed to these extreme rainfall events in the regions of Brazil. The major events are summarised as follows.

- In the North region of Brazil, lines of instability associated with the Intertropical Convergence Zone (ITCZ) contributed to heavy precipitation events between 09^{th} and 10^{th} of May 2022. The INMET surface meteorological stations registered rainfall of 166.3 mm at Itacoatiara-AM on 09^{th} of May, and 120.4 at Bragança-PA and 100.4 mm in Cameta-PA, both on 10^{th} of May.
- On the coast of Northeast Brazil, the heavy rainfalls were caused by the combination of the Easterly Wave Disturbances (DOL) and strong humidity convergence. The INMET surface meteorological stations registered rainfalls of 105.2 mm in João Pessoa-PB on 25^{th} of May, 101.4 mm in Maceió-AL and 91.3 mm in Própria-SE, both on 26^{th} of May and 84.2 mm in Campina Grande-PB on 26^{th} of May.
- In Southern Brazil, the heavy rainfalls were generated by the combination of a warm air mass and a cold front, especially over Rio Grande do Sul (RS). The INMET surface meteorological stations recorded rainfalls of 94.4 mm in Santo Augusto-RS, 84.8 mm in Passo Fundo-RS, both on 30^{th} of May and 86.6 mm in Bom Jesus-RS, on 31^{st} of May.

9.5 Results and discussions

Figures 87 and 88 present the observed daily accumulated precipitation on 09^{th} of May and 10^{th} of May , respectively, from CPC-NOAA (left) and ERA5-ECMWF (right) datasets. On 09^{th} of May, we can see that there are slight differences between the two datasets. The CPC data shows concentrated rainfalls in the northeast of Pará (PA), Amazonas (AM) and Mato Grosso (MT) states. The occurrence of rainfall can also be seen on the coast of Northeast Brazil, specially in Ceará (CE) and Bahia (BA). In addition, CPC data shows isolated rainfalls in the middle of Brazil, meanwhile this feature was not depicted by ERA5, in which ERA5 tends to smooth the rainfall volume in comparison to CPC. On 10^{th} of May, the differences between both datasets are large, in which CPC presents a greater volume of rainfall in the AM state and in the middle of BA, and on the coast of CE, Rio Grande do Norte (RN), Sergipe (SE), BA, and Rio de Janeiro (RJ), whereas ERA5 shows the occurrence of precipitation over Southern Brazil (PR, SC, and RS).

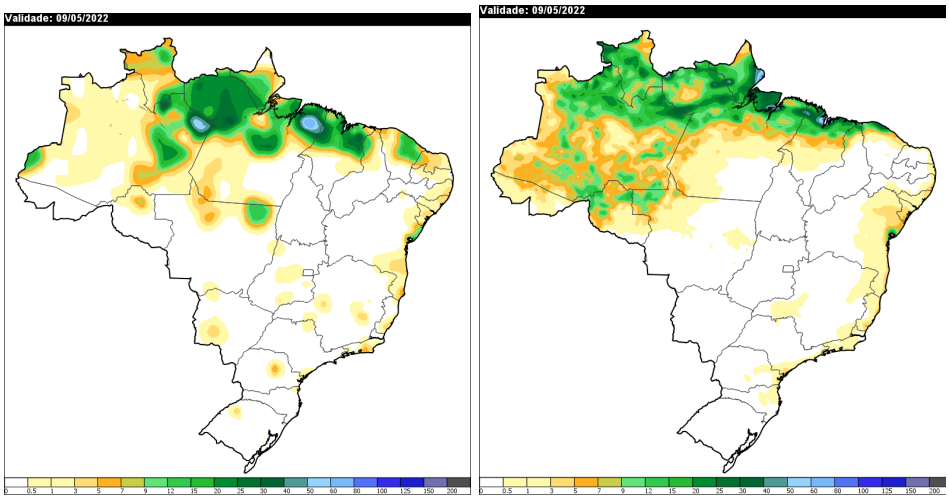


Figure 87: The observed daily accumulated precipitation on 09^{th} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

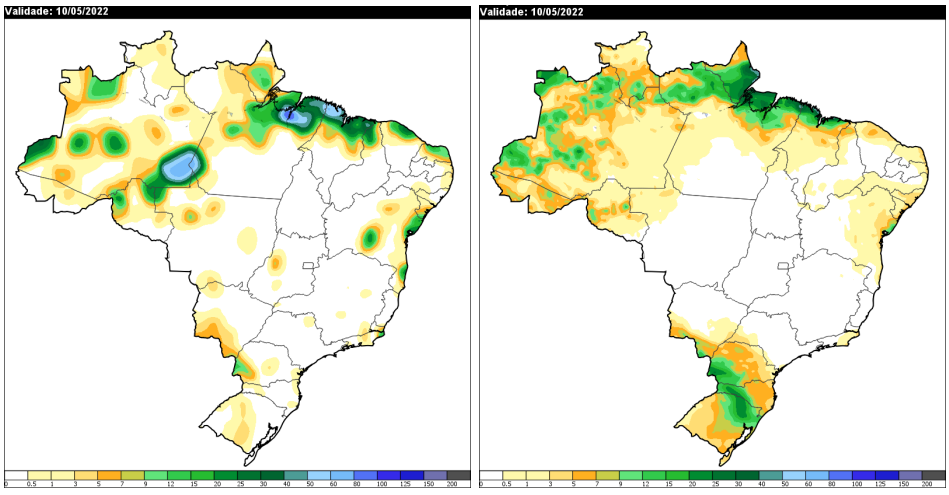


Figure 88: The observed daily accumulated precipitation on 10^{th} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

Figure 89 presents the matrix confusion analyses over the Brazilian territory for both models (COSMO and ICON) compared to CPC and ERA5 for 09^{th} of May, which started to be produced on the 3^{rd} of May (forecasting time: FCT_7) until the target day (FCT_1). The colours represent the four categories of the matrix confusion, that is in grey (true negative: TN), in green (true positive: TP), in blue (false positive: FP), and in red (false negative: FN). The major results for 09^{th} of May were:

- The ICON forecasted precipitation on the central-western and southern parts of Brazil at FCT_7 and FCT_6 that actually did not occur. In contrast, COSMO did not predict this behaviour.
- The FPs over the aforementioned area decreased gradually till FCT_1 when the ICON's accuracy increased compared to its other FCTs.
- Nevertheless, overall, the accuracy of COSMO was better than ICON. On average, the accuracy of COSMO (ICON) was 79.3% (73.1%) and 86.1% (78.9%) compared to CPC and ERA5, respectively. In addition, COSMO (r = 0.47, IOA = 0.65) and ICON (r = 0.30, IOA = 0.49) models had a better agreement with ERA5 than CPC data (rCOSMO = 0.35, IOACOSMO = 0.55 and rICON = 0.26, IOAICON = 0.45).
- Both models had a poor performance to predict the occurrence of rainfall in Northeast Brazil, especially those that happened on the coastline.
- In the North region, the ICON model tends to underpredict the precipitation compared to the COSMO. However, it is important to note that there are large discrepancies between the observed data CPC and ERA5 over this region, leading to different conclusions.

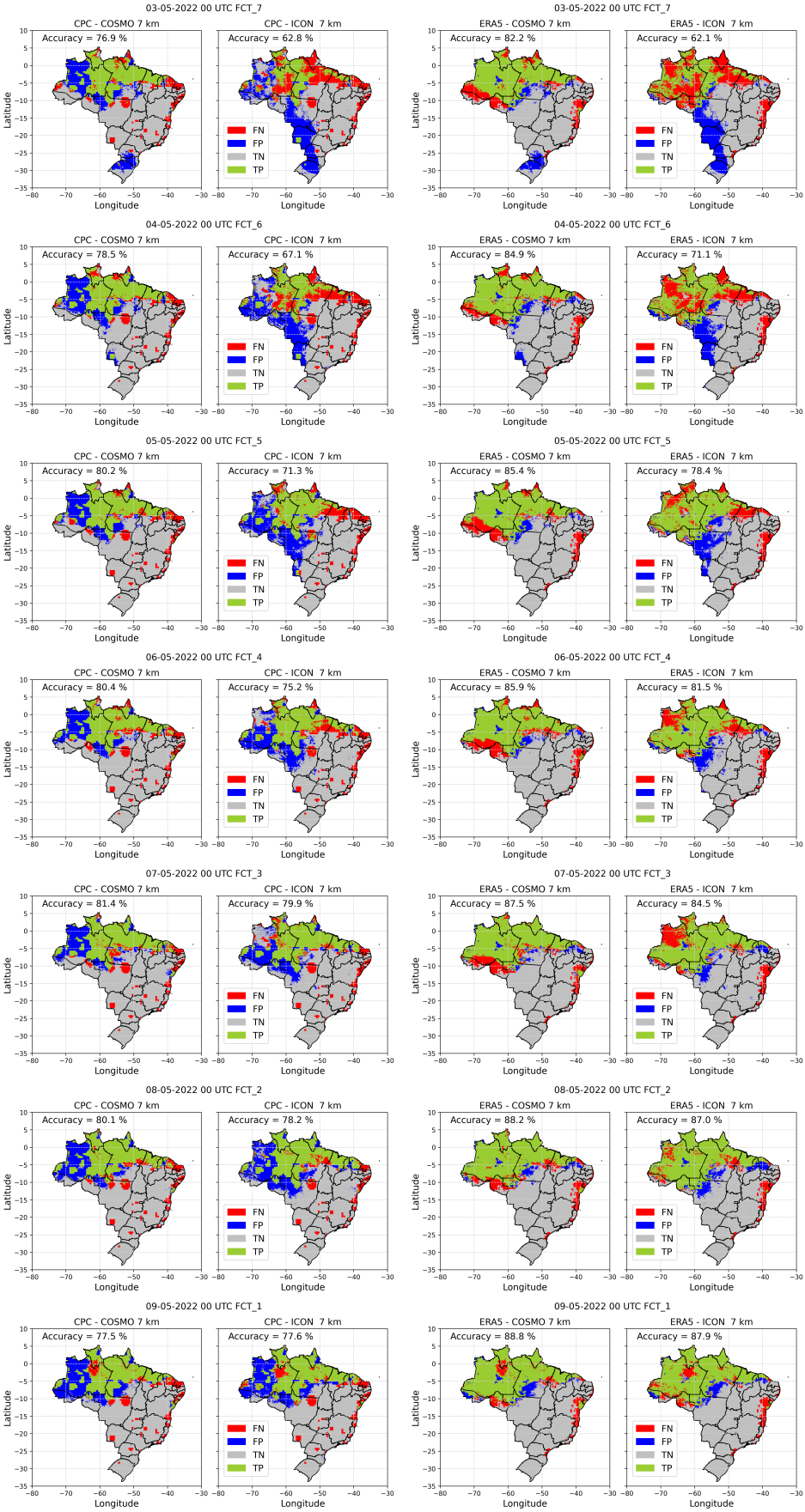


Figure 89: The matrix confusion analyses for forecasting of 09^{th} of May, 2022, comparing COSMO (left) and ICON (right) results with CPC and ERA5 datasets.

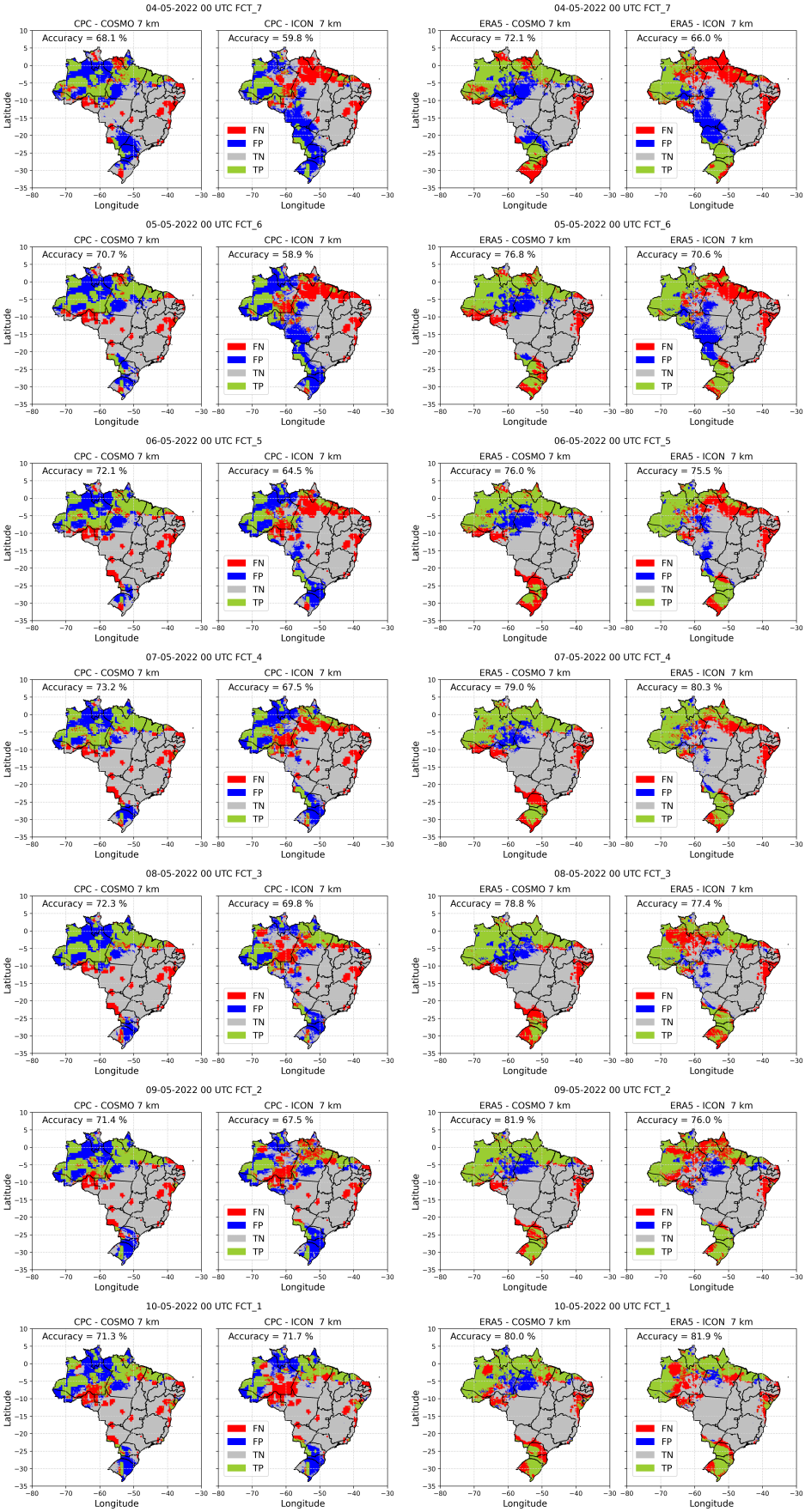


Figure 90: The matrix confusion analyses for forecasting of 10^{th} of May, 2022, comparing COSMO(left) and ICON(right) results with CPC and ERA5 datasets.

Similarly, Figure 90 presents the matrix confusion analyses over the Brazilian territory for both models (COSMO and ICON) compared to CPC and ERA5 for 10^{th} of May, which started to be produced on 04^{th} of May (FCT_7) until the day of interest (FCT_1). The major results for 10^{th} of May were:

- Over Southern Brazil (PR, SC, and RS), both models predicted rainfall. However, because CPC and ERA5 indicated different precipitation fields, the models had a better agreement with ERA5 (the accuracy varied between 66% to 82% among the FCTs) than CPC (59% to 73%).
- Another significant distinction between CPC and ERA5, that influenced the models' accuracy depending on the database used, can be seen in the central region of Bahia state (BA) and in the southeast of Amazonas state (AM).
- In BA, no model depicted the occurrence of this rainfall, including the ERA5 dataset. By verifying the INMET surface meteorological stations data over this area, no occurrence of rainfall was identified, which leads us to think about our results, and to put them in perspective.
- On the other hand, the rainfall in the southeast of AM was forecasted by the COSMO, but not by ICON. Again, the performance of COSMO was greater than ICON. On average, the accuracy, r, and IOA-values of COSMO were respectively 71.3%, 0.26, and 0.43 compared to CPC and 77.8%, 0.41, and 0.62 compared to ERA5. Meanwhile, these indices for ICON were respectively 65.7%, 0.06, and 0.26 compared to CPC and 75.4%, 0.37, and 0.55 compared to ERA5.
- Again, both models had a poor performance to predict the occurrence of rainfall in Northeast Brazil, especially those that happened on the coastline.

Figures 91 and 92 present the observed daily accumulated precipitation on 25^{th} and 26^{th} of May, 2022, respectively, from (a) CPC-NOAA and (b) ERA5-ECMWF datasets. On 25^{th} of May, we can see the heavy rainfall over Northeast Brazil, including the states of RN, PB, PE, AL, SE, and BA, especially along the coastline. Again, ERA5 tends to smooth the rainfall volume in comparison to CPC. The main differences between both datasets were the occurrence of precipitation in the south part of RS (ERA5) and in central AM state (CPC). On 26^{th} of May, according to CPC data, heavy rainfall continued to happen over Northeast Brazil, including the states of PE, AL, SE, and BA (including the eastern and south parts), and the north region of Espírito Santo (ES) state. ERA5 depicted similar behaviour, however with less intensity. Both datasets represent the precipitation that occurred in the Rio Grande do Sul (RS) state.

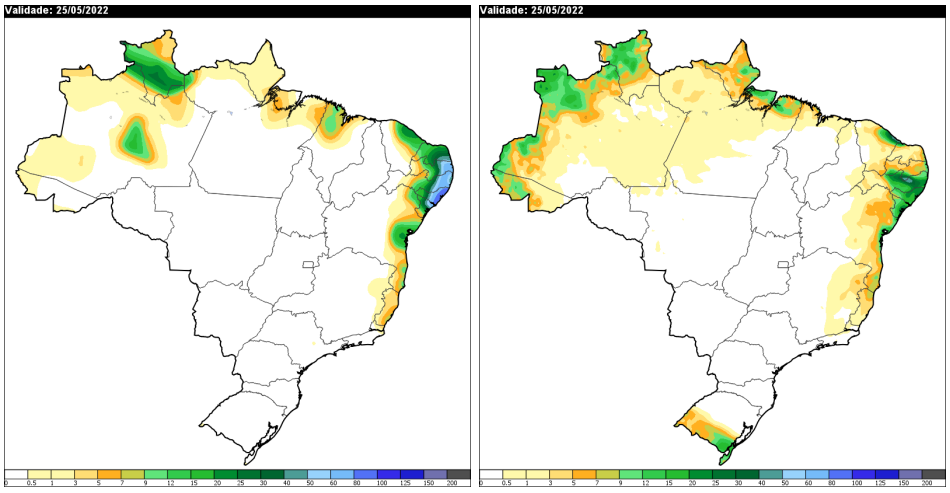


Figure 91: The observed daily accumulated precipitation on 25^{th} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

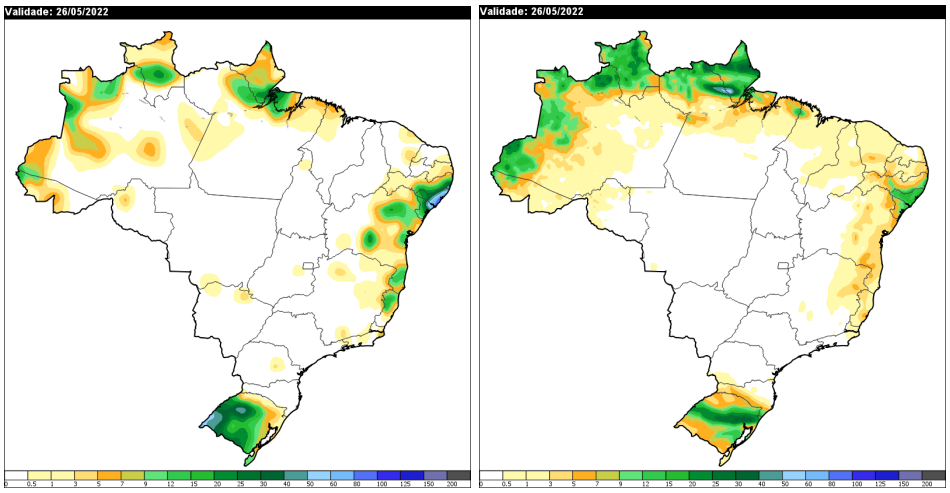


Figure 92: The observed daily accumulated precipitation on 26^{th} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

Figures 93 and 94 presents the matrix confusion analyses over the Brazilian territory for both models (COSMO and ICON) compared to CPC and ERA5 for 25^{th} and 26^{th} of May, which started to be produced on 19^{th} and 20^{th} of May, respectively. The major results for these days are:

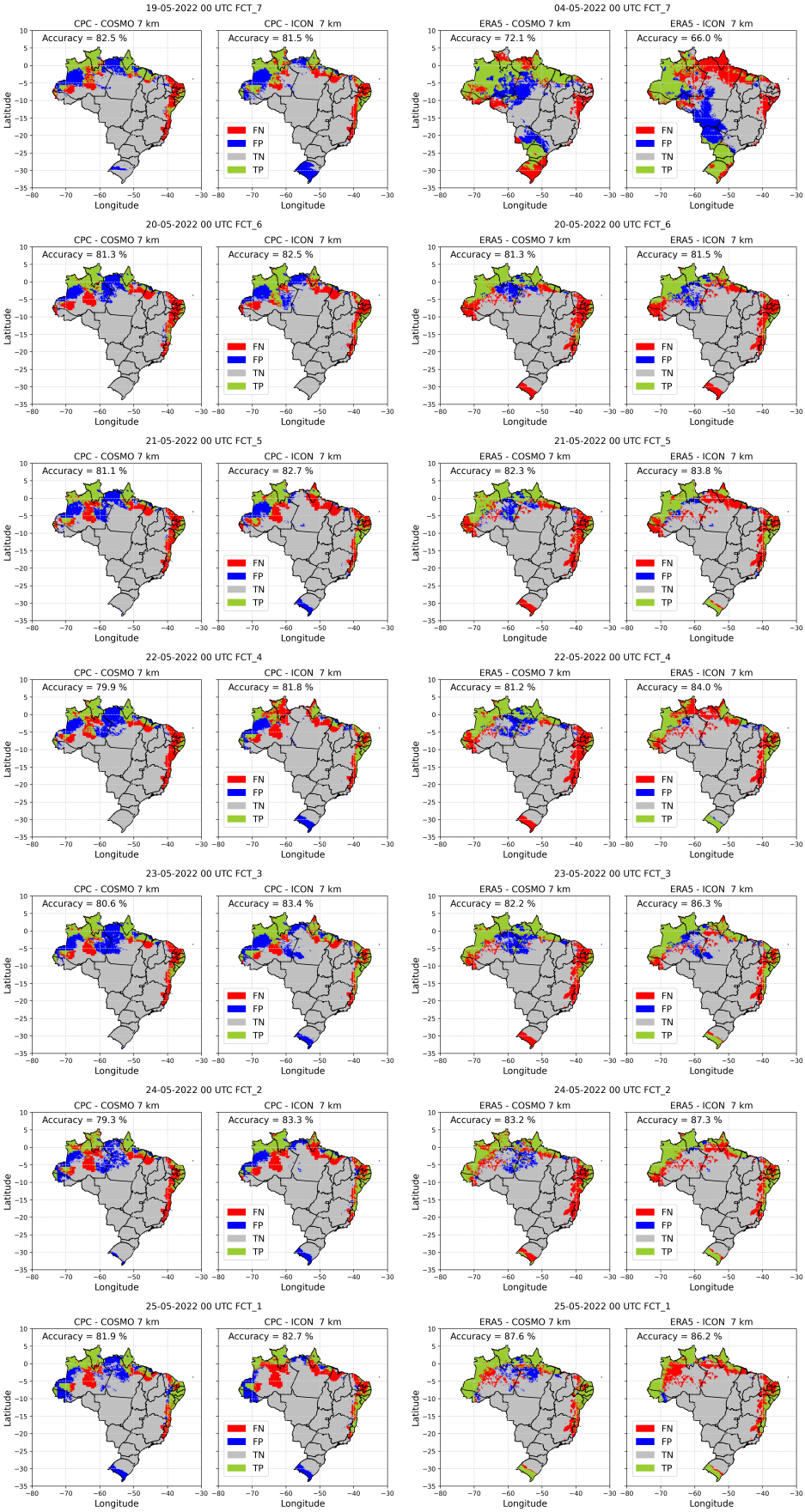


Figure 93: The matrix confusion analyses for forecasting of 25^{th} of May, 2022, comparing COSMO(left) and ICON(right) results with CPC and ERA5 datasets.

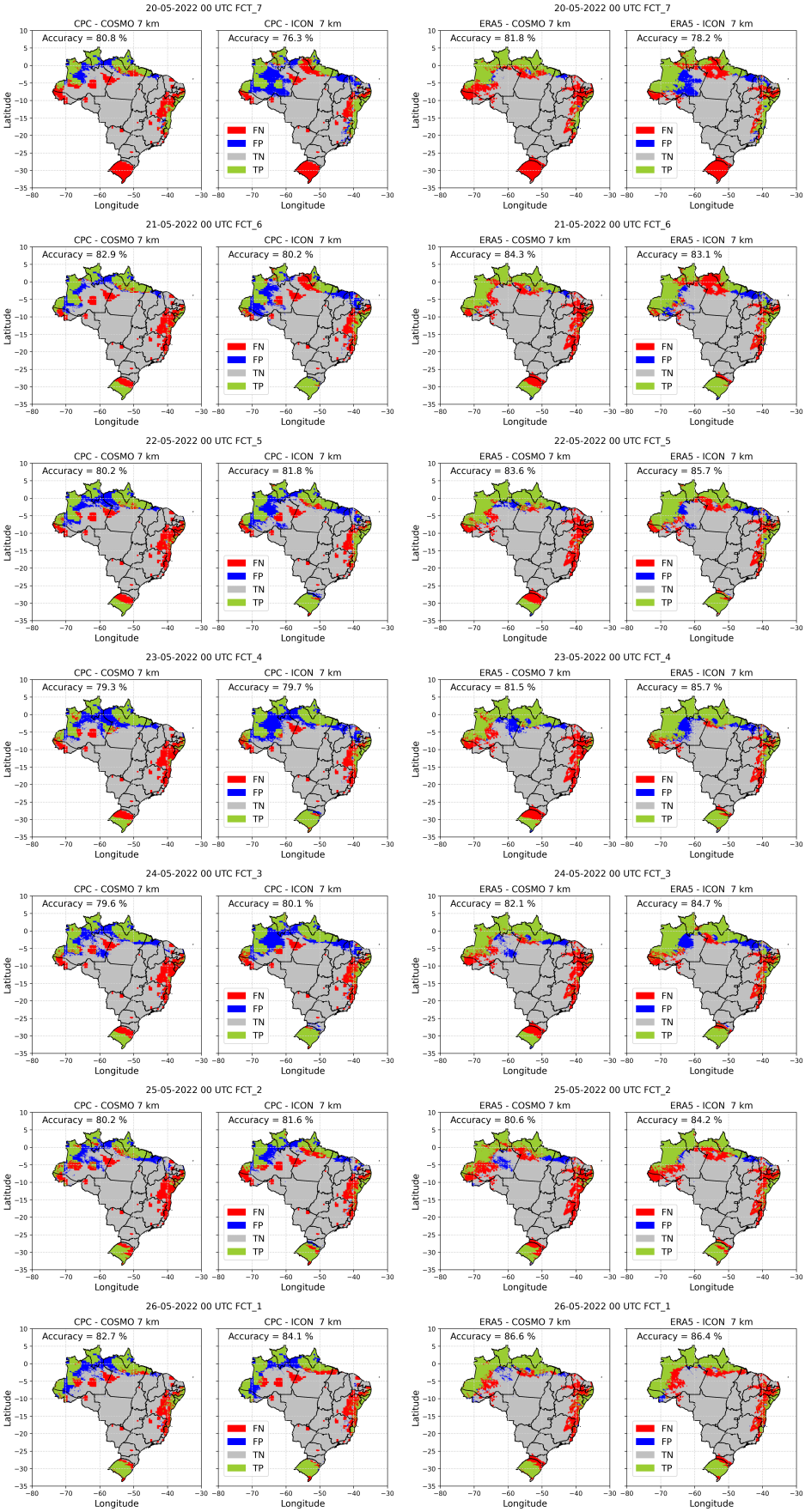


Figure 94: The matrix confusion analyses for forecasting of 26^{th} of May, 2022, comparing COSMO(left) and ICON(right) results with CPC and ERA5 datasets.

- The precipitation that happened over the state of RS on 25^{th} of May was forecasted only by ICON throughout the FCTs, except for FCT_1 in which COSMO also identified the occurrence of this rainfall. On 26^{th} of May, both models portrayed the rainfall, despite the ICON model showing a better spatial distribution of the rainfall, leading to a greater accuracy than COSMO.

- On average, on 25^{th} of May, the accuracy of COSMO (ICON) was 80.9% (82.6%) and 83.0% (84.8%) compared to CPC and ERA5, respectively. In addition, COSMO (r = 0.46, IOA = 0.64) and ICON (r = 0.59, IOA = 0.70) models had a better agreement with ERA5 than CPC data (rCOSMO = 0.31, IOACOSMO = 0.45 and rICON = 0.35, IOAICON = 0.42).

- Similarly, on 26^{th} of May, the performance of ICON was also better than COSMO. The accuracy, r, and IOA-values of COSMO were 80.8% (82.9%), 0.32 (0.51), and 0.52 (0.68) compared to CPC (ERA5), respectively. Meanwhile, the values for ICON were slightly higher, with 80.5% (84.0%) of accuracy, 0.48 (0.53) for coefficient correlation, and 0.65 (0.69) for IOA compared to CPC (ERA5), respectively.

- Over Northeast Brazil, the models partially indicated the rainfall, being improved after the FCT3. Likewise, the rainfall over Espírito Santo (ES) state, in Southeast Brazil, was also depicted only by ICON.

Figure 95 and 96 present the observed daily accumulated precipitation on 30^{th} and 31^{st} of May, respectively, from CPC-NOAA (left) and ERA5-ECMWF (right) datasets. Again, we can see precipitation occurring in the north region of Brazil, which has slight differences depending on the dataset considered. Heavy rainfall continued to occur over Northeast Brazil, especially along the coastline. On May 30, the CPC dataset again indicated an occurrence of precipitation in central Bahia (BA), in which the INMET surface meteorological stations and ERA5 data did not expose any rainfall over this area. Heavy rainfalls also fell over Southern Brazil, and São Paulo state (SP) and Espírito Santo state, both in Southeast Brazil. However, the latter was not depicted by ERA5.

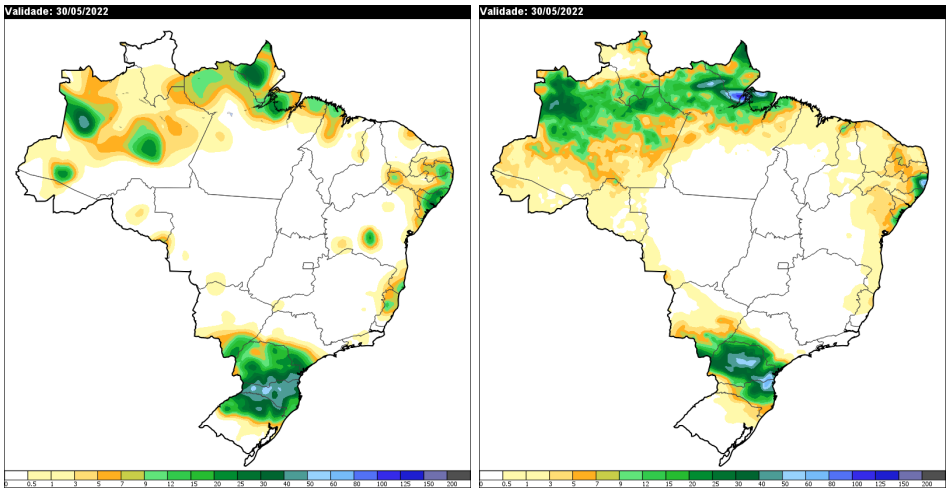


Figure 95: The observed daily accumulated precipitation on 30^{th} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

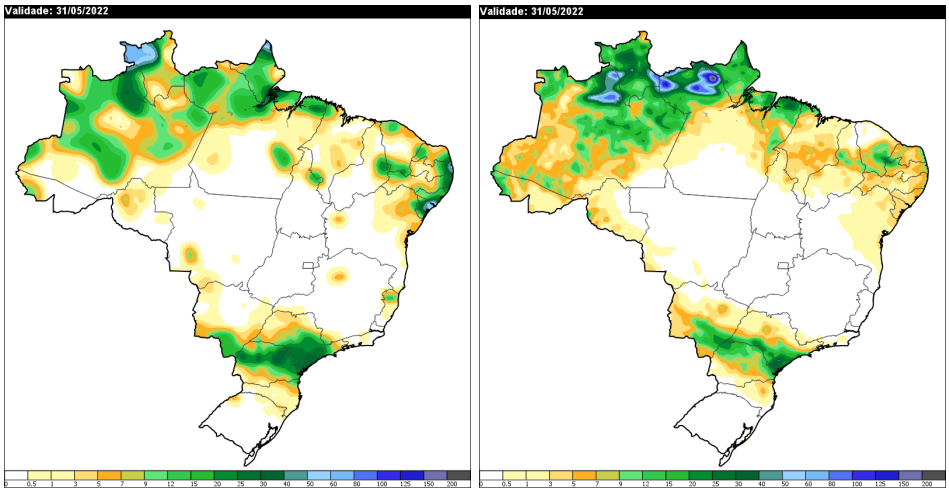


Figure 96: The observed daily accumulated precipitation on 31^{st} of May, 2022 from CPC-NOAA (left) and from ERA5-ECMWF (right) datasets.

Figures 97 and 98 presents the matrix confusion analyses over the Brazilian territory for both models (COSMO and ICON) compared to CPC and ERA5 for 30^{th} and 31^{st} of May, which started to be produced on 24^{th} and 25^{th} of May, respectively. The major results for these both days were:

- The precipitation that happened in the north and south regions of Brazil was well forecasted by both models, despite the ICON model showing a better spatial distribution of the rainfall, leading to a better performance than COSMO.
- However, again, along the coastline of Northeast Brazil and the north of Espírito Santo state (ES), the models had greater deviations and did not predict the occurrence of rainfall that actually occurred. Even though, ICON had better performance than the COSMO model.
- On average, on 30^{th} of May, the accuracy of COSMO (ICON) was 78.2% (78.8%) and 80.1% (80.8%) compared to CPC and ERA5, respectively. In addition, COSMO (r = 0.50, IOA = 0.65) and ICON (r = 0.58, IOA = 0.65) models had a better agreement with ERA5 than CPC data (rCOSMO = 0.37, IOACOSMO = 0.57 and rICON = 0.37, IOAICON = 0.55).
- Similarly, on 31^{st} of May, the performance of ICON was also better than COSMO. The accuracy, r, and IOA-values of COSMO were 74.7% (76.6%), 0.33 (0.32), and 0.53 (0.45) compared to CPC (ERA5), respectively. Meanwhile, the values for ICON were slightly higher, with 76.9% (79.2%) of accuracy, 0.41 (0.47) for coefficient correlation, and 0.59 (0.56) for IOA compared to CPC (ERA5), respectively.

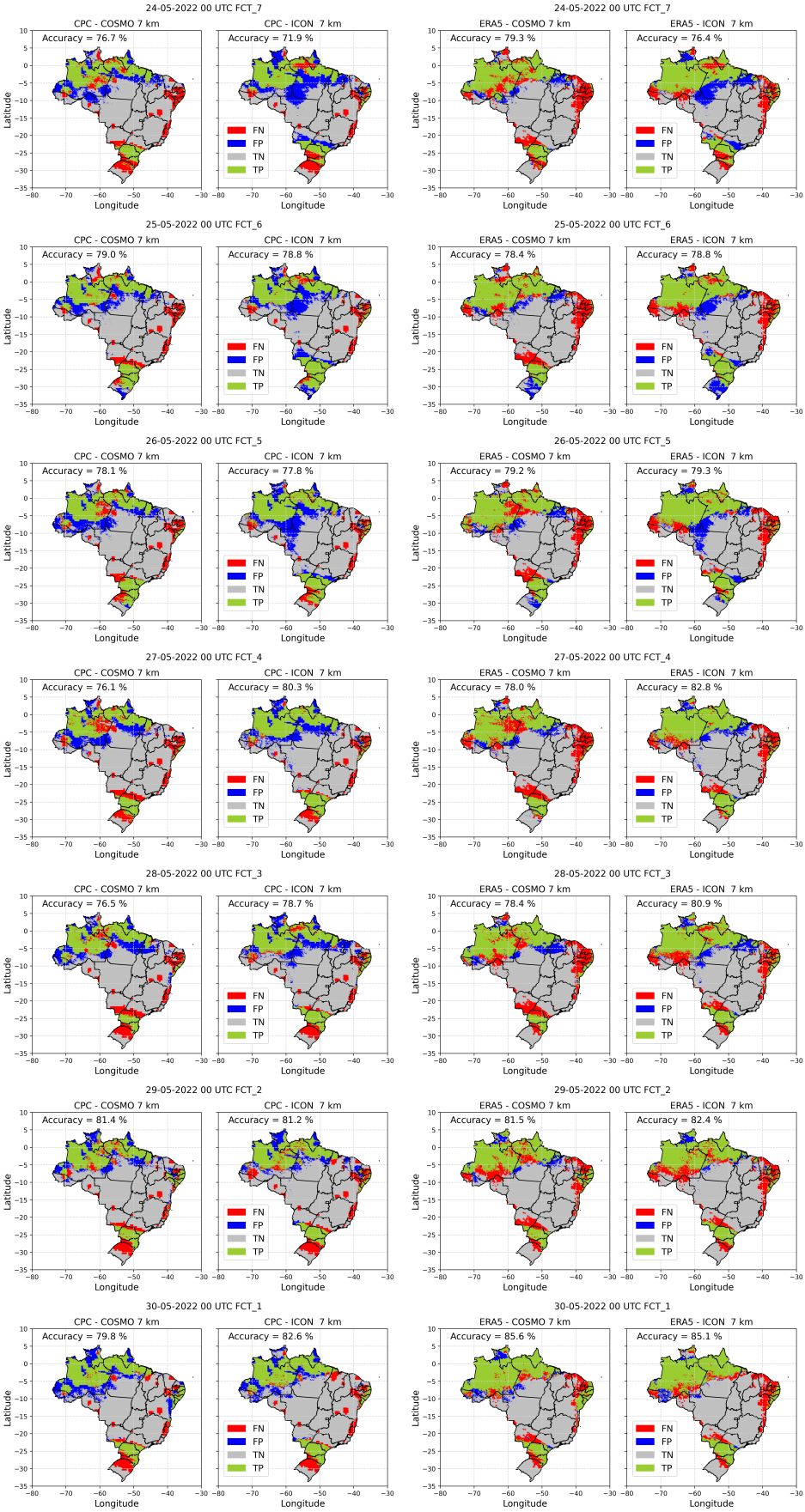


Figure 97: The matrix confusion analyses for forecasts of 30^{th} of May, 2022, comparing COSMO(left) and ICON(right) results with CPC and ERA5 datasets.

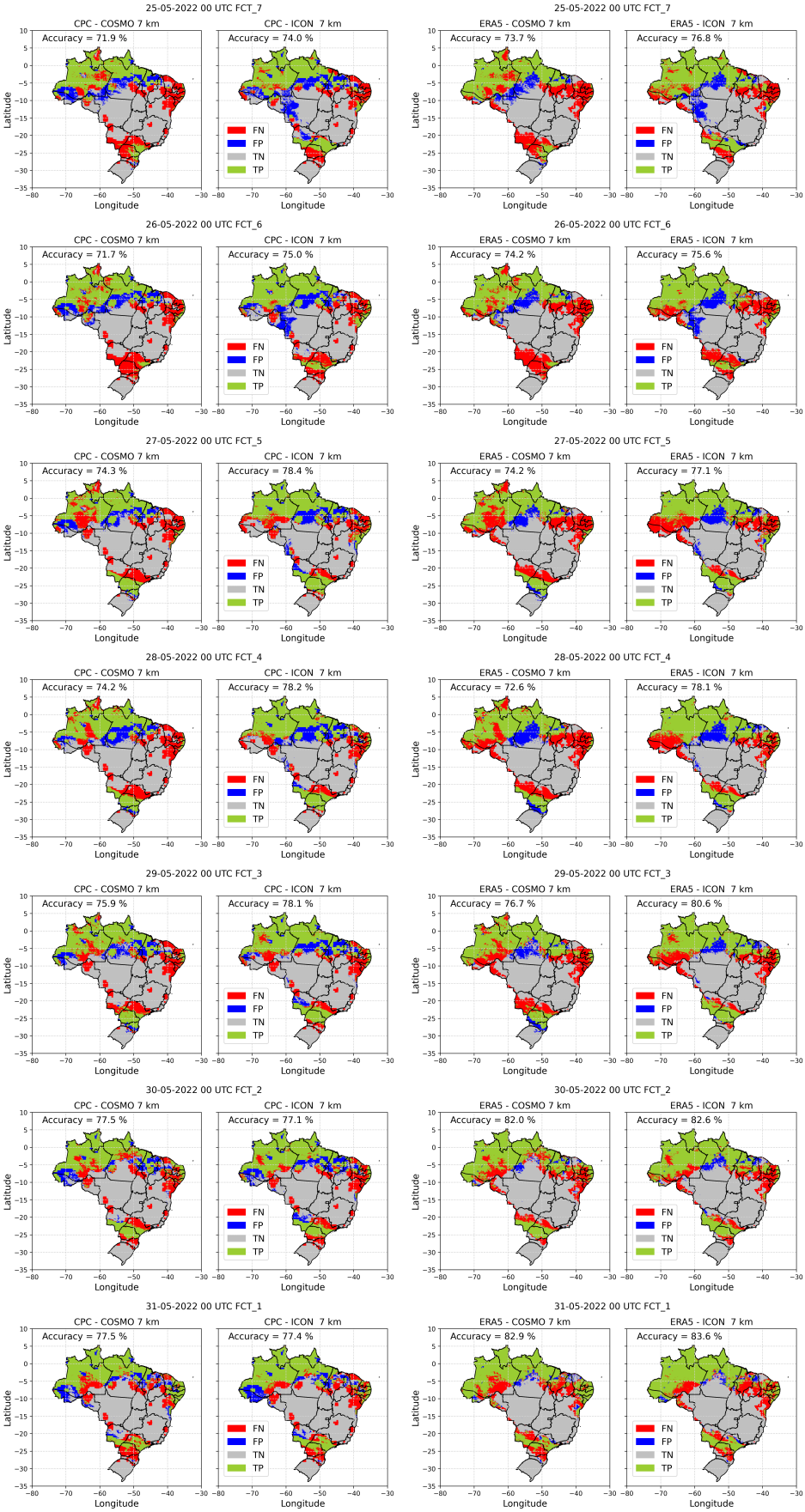


Figure 98: The matrix confusion analyses for forecasts of 31^{st} of May, 2022, comparing COSMO(left) and ICON(right) results with CPC and ERA5 datasets.

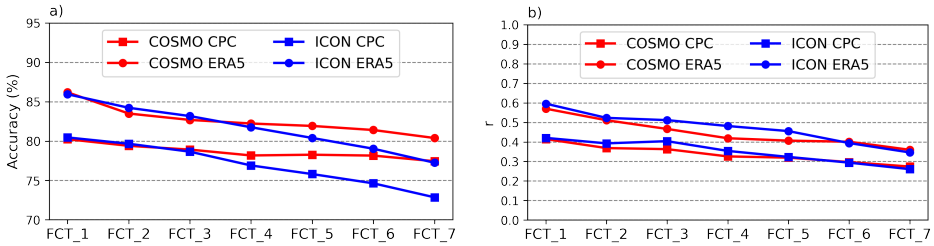


Figure 99: Average of (a) accuracy and (b) Pearson correlation coefficient (r) indexes throughout the forecasting time for May 2022.

By considering the entire month of May 2022, the average accuracy and Pearson correlation coefficient (r) for each forecasting time is presented in Figure 14. It is noticeable that both models have a better agreement with ERA5 (circle markers) than CPC (square markers), with a difference of accuracy and correlation around 6% and 0.2, respectively. In addition, it is also possible to note that ICON had a slightly better performance than COSMO from FCT_3 (72 hours) to FCT_1 (24 hours). Between FCT_4 and FCT_7, COSMO had higher accuracy than ICON. On average, considering all FCTs, the average accuracy of COSMO (CPC = 79%, ERA5 = 83%) was slightly higher than ICON (CPC = 77%, ERA5 = 82%) over the study period. On the other hand, ICON (CPC = 0.35, ERA5 = 0.47) had higher correlation coefficients than COSMO (CPC = 0.34, ERA5 = 0.45), represented by the blue lines in Figure 14b.

Figure 100 shows the accuracy of the models for each day of May 2022, comparing the models and observed datasets. t is possible to see that the ICON model had a more varied accuracy than COSMO, especially between 7^{th} and 18^{th} of May. In addition, over these days, the performance of ICON was poorer than COSMO. For instance, on 13^{th} and 15^{th} of May, the accuracy of ICON reached values close to 60% (Figure 15d) and 50% (Figure 15c), respectively. Meanwhile, for these same days, COSMO had an accuracy of between 70% and 90%.

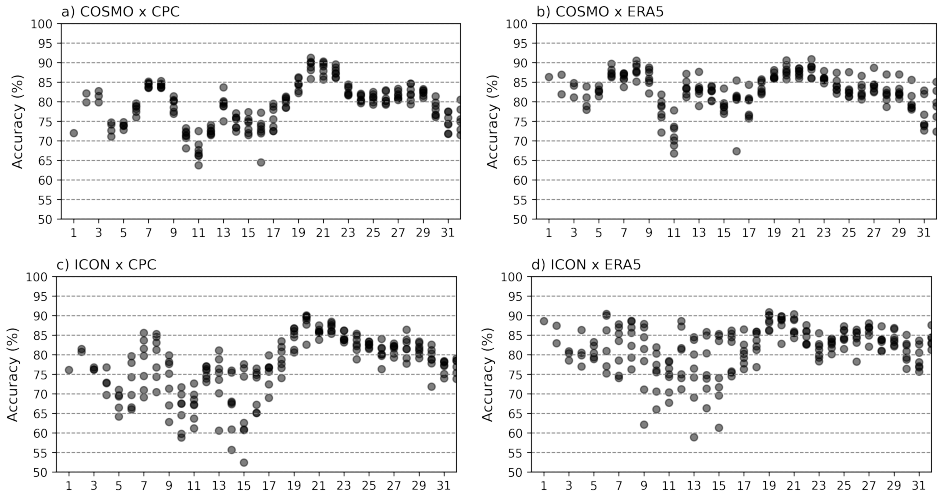


Figure 100: Accuracy from 1^{st} to 31^{st} of May, 2022 for each FCT considering (a) COSMO x CPC, (b) COSMO x ERA5, (c) ICON x CPC, and (d) ICON x ERA5.

Overall, the lowest accuracy (52.4%) occurred for the forecast of 15^{th} of May FCT_7 (which was initialised on 9^{th} of May) comparing ICON and CPC (Figure 15c). Meanwhile, this accuracy was 61.3% using ERA5, and 75.5% (CPC) and 78.6% (ERA5) using COSMO, leading to a difference of 23.1% when comparing both models. Although, for FCT_2 and FCT_1, the accuracy was improved, reaching values close to 76%. Nevertheless, further investigation was performed for 15^{th} of May, and Figure 101 shows the bias between COSMO and ICON models for each FCT. The positive values (in blue) indicate the volume of rainfall from COSMO was greater than ICON. The negative values (in yellow-orange) designate the opposite behaviour, that is, the volume of rainfall forecasted by ICON was greater than COSMO. The absence of deviations between both models is associated with values closer to zero (in white). We can see that the biggest difference between the models occurred for FCT_7, when ICON model emulated more rainfall over the South and Central-west Brazil than COSMO. These differences decreased over the forecast time till the day of interest (15^{th} of May: FCT_1), when overall the maps indicate that COSMO forecasted larger amount of precipitation than ICON.

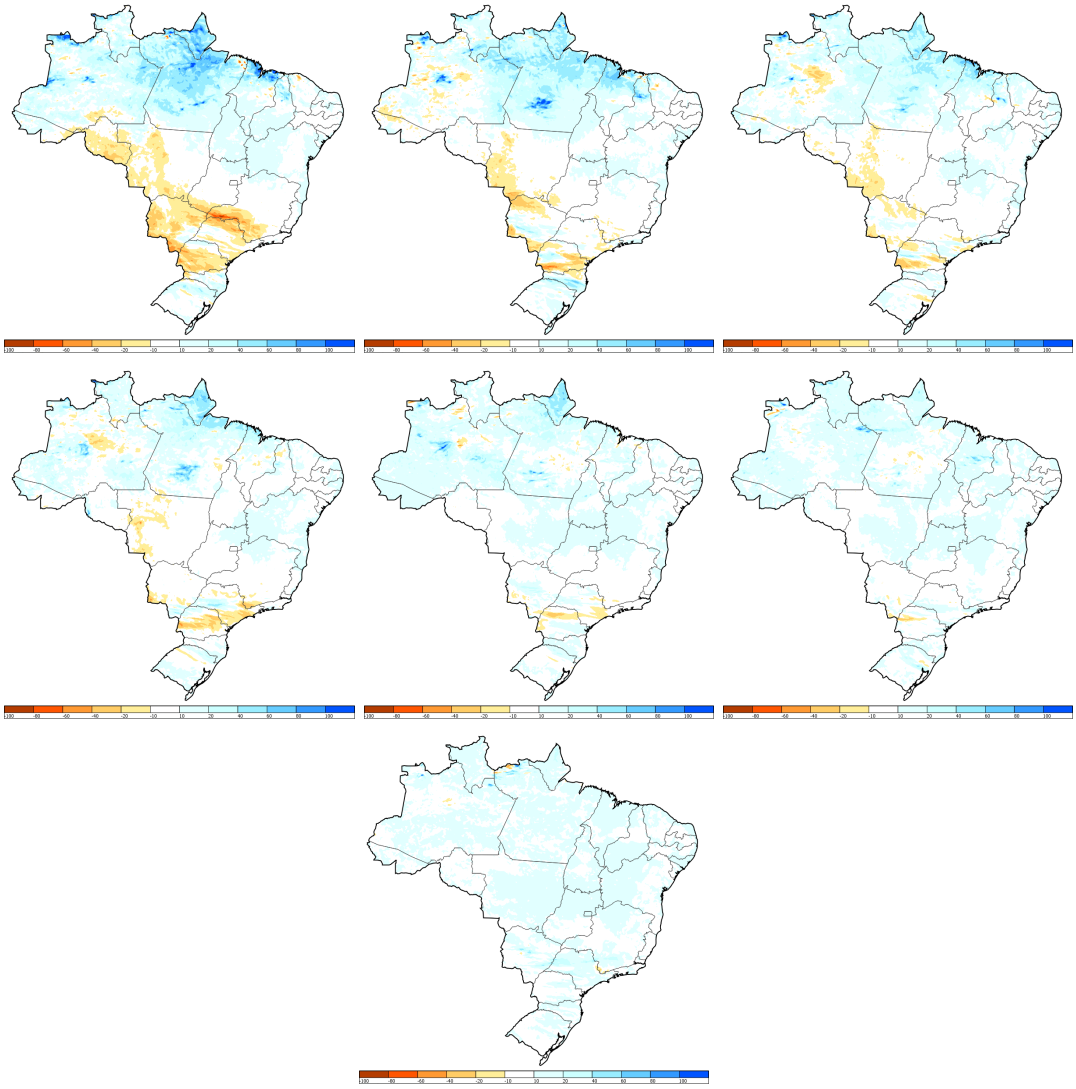


Figure 101: The difference between COSMO and ICON models for forecasts 15^{th} of May.

Figure 102 shows the matrix confusion analyses considering if the models also hit the vol-

ume of precipitation (TP+ and TP-) over the Brazilian territory for both models (COSMO and ICON) compared to CPC and ERA5 for 15^{th} of May, which started to be produced on 09^{th} of May (FCT_7) until the day of interest (FCT_1). The colours represent the four categories of the matrix confusion already mentioned, however the green points (true positive: TP) indicate that both observed and modelled data matched the same category of rainfall intensity (Table 5). In addition, other two categories were added: in light blue (true positive, but overestimating the precipitation: TP+) and in orange (true positive, but underestimating the precipitation: TP-). When compared to ERA5, both models usually tend to underestimate the rainfall volume over North Brazil. Meanwhile, compared to CPC, the models either underestimate the rainfall or overestimate it over the same region. Similar patterns can be seen over the South and Central-West regions.

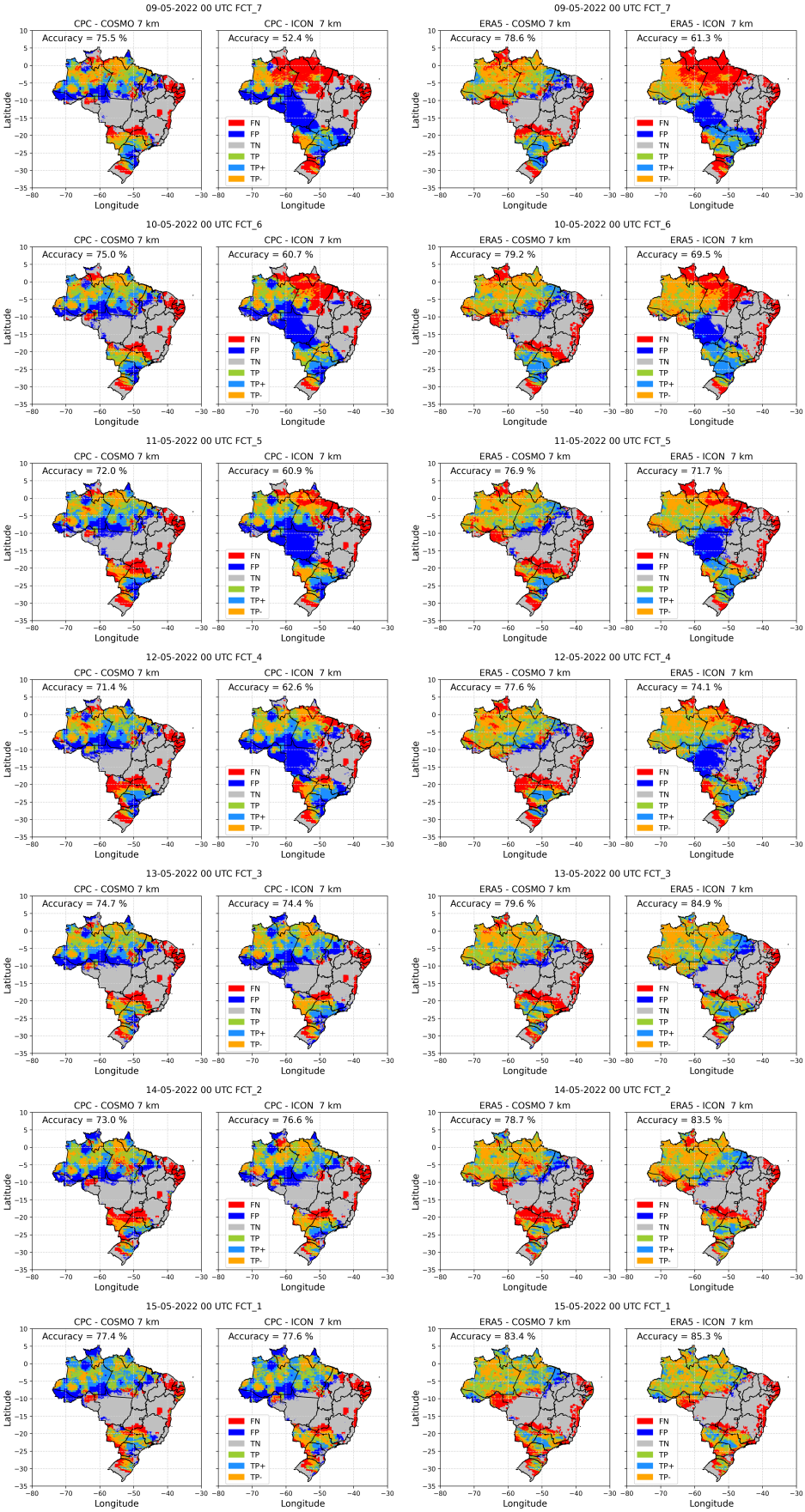


Figure 102: The matrix confusion analyses for forecasts of 15^{th} of May, 2022, comparing COSMO and ICON results with CPC and ERA5 datasets.

9.6 Summary and conclusions

- COSMO and ICON models agree better with the ERA5 dataset. This better agreement with ERA5 can be due to the fact that ERA5 has a higher spatial resolution (0.25°, approximately 28 km) than CPC (0.50°, approximately 56 km), and it is closer to the COSMO and ICON resolutions (0.0625°, approximately 7 km). In addition, ERA5 is a reanalysis product that integrates observations and dynamics and physical modelling.
- Comparing CPC and ERA5 datasets, CPC usually presents isolated rainfall across the country, especially in the central-west Brazil, in which ERA5 does not and both models also do not forecast these rainfalls. This can be overcome by using a higher horizontal resolution, such as 2.8 km with shallow convection activation.
- By considering the 7 day leadtime, on average, COSMO has a better performance than ICON over the study period. However, if we consider the last 72-hour forecasting (from FCT_3 to FCT_1), ICON has a better agreement than COSMO.
- Both models have a systematic error over the coastal region of Northeast Brazil, where most of the time, the models did not predict the occurrence of rainfall, which actually occurred, meaning that the models are underpredicting the rainfall.
- On the other hand, for the South-Southeast Brazil, both models have a better prediction, which usually depict the occurrence or not of precipitation. That may be caused by the fact that there are more observations in these regions (see Figure 86) to be assimilated for the initial condition.
- For future applications, it is suggested to make new changes to the namelist parameters, especially for tropical regions. However, that is troublesome because the South America domain is very extensive and comprehends various latitudes, weather systems, and climate zones. Once set in the namelist, the parameters are applied for the entire domain, which can lead to a better performance for one region, but it can degrade for others.

We have noticed that the performance of ICON-LAM decreases significantly with the leadtime. This is an unexpected behavior for the model, which may be due to the setup of lateral boundary condition, i.e, if provided as frame grid as the current verification or if full grid is used with activation of nudging at the top. Thus, for a large domain as the one used in this work and long leadtimes, it is possible that ICON behaves such we have observed. Therefore, we intend to improve the verification results, investigating further this problem by tuning nudging parameter and also comparing the result of ICON-LAM with the results provided by the global model.

9.7 Acknowledgments

We acknowledge the support of INMET, SIMEPAR and DWD in the realization of this work, especially the DWD that provided the ICON and COSMO model codes, as well as the boundaries for their implementations. Many thanks to Daniel Rieger, for the coordination of C2I project and for helping in the setup of ICON model at INMET. We are very grateful to Juliana Mol for her contribution during several phases of the implementation and operation of the COSMO and ICON at INMET.

10 Conclusions

For the surface verification, ICON generally performed better than COSMO with very few exceptions. The most striking exception is the cloud cover in the Mediterranean region as already pointed out before (Rieger et al., 2021). Similarly, upper air verification scores are mostly also better for ICON.

At a convection-permitting scale, the forecast of precipitation is probably the most important parameter. While Italy and Israel show no significant improvement in precipitation scores of ICON compared to the COSMO model, Romania, Greece and Poland report improvements for certain seasons or thresholds. It needs to be pointed out here that for Italy and Israel, latent heat nudging is used. Thus, the precipitation forecasts (also from the COSMO model) might be better constrained by observations.

As shown in tab. 1, ICON has already or will soon become operational at the COSMO national meteorological services. Two COSMO countries (besides Germany), Israel and Italy, have even started to use data assimilation. Thus, the goal of PP C2I has been successfully fulfilled and the COSMO consortium can now focus on the further development of the ICON model.

11 Acknowledgements

We want to thank everyone involved in this large project. Special thanks go to Dmitrii Mironov who supported and accompanied this project as COSMO Scientific Project Manager and to Massimo Milelli who took this overarching project under the roof of WG6 and took care of several organizational tasks.

References

Baldauf, M., and Coauthors, 2021: The COSMO Priority Project CDIC: Comparison of the dynamical cores of ICON and COSMO, Final Report. *COSMO Technical Reports*, **(44)**.

Climate Prediction Center, NOAA, 2022: The CPC Global Unified Precipitation Data. https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html.

Hersbach, H., 2020: The ERA5 global reanalysis. *Quartely Journal of Royal Meteorological Society*, 1999–2049.

Iriza-Burcă, A., R. Dumitrache, B. Maco, M. Huştiu, F. Fundel, D. Rieger, and R. Potthast, 2022a: Comparison of COSMO and ICON-LAM High-resolution Numerical Forecast for Romanian Territory: Case Studies and Evaluation. *Meteorol. Atmos. Phys., submitted.*

Iriza-Burcă, A., J. Linkowska, and F. Fundel, 2020: Common Area Verification Activity with Rfdbk/MEC Application: MAM 2020 First Results. *COSMO Newsletter*, **(20)**, 18–30.

Iriza-Burcă, A., and Coauthors, 2021: Summary of Results from the COSMO Priority Project CARMA: Common Area with Rfdbk/MEC Application. *COSMO Newsletter*, **(21)**, 21–27.

Iriza-Burcă, A., and Coauthors, 2022b: The COSMO Priority Project CARMA: Common Area with Rfdbk/MEC Application Final Report. *COSMO Technical Reports*, **(46)**.

Khain, P., Y. Levi, H. Muskatel, A. Shtivelman, E. Vadislavsky, and N. Stav, 2021: Effect of shallow convection parametrization on cloud resolving nwp forecasts over the eastern mediterranean. *Atmospheric Research*, **247**, 105 213.

Khain, P., Y. Levi, A. Shtivelman, E. Vadislavsky, E. Brainin, and N. Stav, 2020: Improving the precipitation forecast over the eastern mediterranean using a smoothed time-lagged ensemble. *Meteorological Applications*, **27 (1)**, e1840.

Muskatel, H. B., U. Blahak, P. Khain, Y. Levi, and Q. Fu, 2021: Parametrizations of liquid and ice clouds' optical properties in operational numerical weather prediction models. *Atmosphere*, **12 (1)**, 89.

Rieger, D., and Coauthors, 2018: C2I Workshop on ICON-LAM Setup & Experiments. *COSMO Newsletter*, **(18)**, 17–27.

Rieger, D., and Coauthors, 2021: Verification of ICON in Limited Area Mode at COSMO National Meteorological Services. *Reports on ICON*, **(006)**, 1–55.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136 (1)**, 78–97.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Perianez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Q. J. R. Meteorol. Soc.*, **142 (696)**, 1453–1472.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, **141 (687)**, 563–579.

List of COSMO Newsletters and Technical Reports

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

COSMO Newsletters

No. 1: February 2001.

No. 2: February 2002.

No. 3: February 2003.

No. 4: February 2004.

No. 5: April 2005.

No. 6: July 2006.

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

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

No. 9: December 2008.

No. 10: March 2010.

No. 11: April 2011.

No. 12: April 2012.

No. 13: April 2013.

No. 15: July 2015.

No. 16: July 2016.

No. 17: July 2017.

No. 18: November 2018.

No. 19: October 2019.

No. 20: December 2020.

No. 21: May 2022.

COSMO Technical Reports

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

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

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

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

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

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

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

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

No. 9: Erdmann Heise, Bodo Ritter, Reinhold Schrodin (2006):
Operational Implementation of the Multilayer Soil Model.

No. 10: M.D. Tsyrulnikov (2007):
Is the particle filtering approach appropriate for meso-scale data assimilation ?

No. 11: Dmitrii V. Mironov (2008):
Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model.

No. 12: Adriano Raspanti (2009):
COSMO Priority Project "VERification System Unified Survey" (VERSUS): Final Report.

No. 13: Chiara Marsigli (2009):
COSMO Priority Project "Short Range Ensemble Prediction System" (SREPS): Final Report.

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

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

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

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

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

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

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

No. 21 Michael Baldauf (2013):
A new fast-waves solver for the Runge-Kutta dynamical core.

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

No. 23 M. Baldauf, O. Fuhrer, M. J. Kurowski, G. de Morsier, M. 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 and D. Gayfulin (2016):
A Stochastic Pattern Generator for ensemble applications

No. 30 Dmitrii Mironov and 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 Naima 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 (2018):
The COSMO Priority Project CORSO Final Report.

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

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

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

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.

No. 40 E. Bucchignani, P. Mercogliano, V. Garbero, M. Milelli, M. Varentsov, I. Rozinkina, G. Rivin, D. Blinov, A. Kirsanov, H. Wouters, J.-P. Schulz, U. Schaettler (2019):
Analysis and Evaluation of TERRA_URB Scheme: PT AEVUS Final Report.

No. 41 X. Lapillonne, O. Fuhrer (2020):
Performance On Massively Parallel Architectures (POMPA): Final report.

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

No. 43 H. Muskatel, U. Blahak, P. Khain, A. Shtivelman, M. Raschendorfer, M. Kohler, D. Rieger, O. Fuhrer, X. Lapillonne, G. Rivin, N. Chubarova, M. Shatunova, A. Poliukhov, A. Kirsanov, T. Andreadis, S. Gruber (2021):
The COSMO Priority Project T2(RC)2: Testing and Tuning of Revised Cloud Radiation Coupling, Final Report.

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

No. 45 Marsigli C., Astakhova E. Duniec G., Fuezer L., Gayfulin D., Gebhardt C., Golino R., Heppelmann T., Interewicz W., Marcucci F., Mazur A., Sprengel M., Tsyrulnikov M., Walser A. (2022):
The COSMO Priority Project APSU: Final Report.

No. 46 A. Iriza-Burca, F. Gofa, D. Boucouvala, T. Andreadis, J. Linkowska, P. Khain, A. Shtivelman, F. Batignani, A. Pauling, A. Kirsanov, T. Gastaldo, B. Maco, M. Bogdan, F. Fundel (2022):
The COSMO Priority Project CARMA: Common Area with Rfdbk/MEC Application Final Report.

No. 47 A. Voudouri, E. Avgoustoglou, Y. Levy, I. Carmona, E. Bucchignani, J. M. Bettems (2022):
Calibration of COSMO Model, Priority Project CALMO-MAX: Final Report.

COSMO Technical Reports

Issues of the COSMO Technical Reports series are published by the *COnsortium for Small-scale MOdelling* at non-regular intervals. COSMO is a European group for numerical weather prediction with participating meteorological services from Germany (DWD, AWGeophys), Greece (HNMS), Italy (USAM, ARPA-SIMC, ARPA Piemonte), Switzerland (MeteoSwiss), Poland (IMGW), Romania (NMA) and Russia (RHM). The general goal is to develop, improve and maintain a non-hydrostatic limited area modelling system to be used for both operational and research applications by the members of COSMO. This system is initially based on the COSMO-Model (previously known as LM) of DWD with its corresponding data assimilation system.

The Technical Reports are intended

- for scientific contributions and a documentation of research activities,
- to present and discuss results obtained from the model system,
- to present and discuss verification results and interpretation methods,
- for a documentation of technical changes to the model system,
- to give an overview of new components of the model system.

The purpose of these reports is to communicate results, changes and progress related to the LM model system relatively fast within the COSMO consortium, and also to inform other NWP groups on our current research activities. In this way the discussion on a specific topic can be stimulated at an early stage. In order to publish a report very soon after the completion of the manuscript, we have decided to omit a thorough reviewing procedure and only a rough check is done by the editors and a third reviewer. We apologize for typographical and other errors or inconsistencies which may still be present.

At present, the Technical Reports are available for download from the COSMO web site (www.cosmo-model.org). If required, the member meteorological centres can produce hard-copies by their own for distribution within their service. All members of the consortium will be informed about new issues by email.

For any comments and questions, please contact the editor:

Massimo Milelli
massimo.milelli@cimafoundation.org