Common Area Results 2020-2021

COSMO WG5: Verification and case studies

Flora Gofa (fgofa@hnms.gr)

Abstract

Verification results of statistical indices for main weather parameters are derived using the operational COSMO and ICON-LAM model implementations in each service. The domain (common), the resolution, the statistical scores/methods and the graphical representation approaches, are decided on an annual basis from WG5. A common verification software is used in most cases which allows for a homogeneous, standardized and objective way to apply, calculate and present the verification scores. The outcome of this activity provides a basis to monitor the performance of the models model and track the systematic errors. Since the introduction of ICON-LAM in the operational forecast procedure of some services, special focus is given to the relative performance of the two models. In this report, statistical results of JJA-2020 up to MAM 2021 model performance are presented.
Keywords: Verification. Common Plots, FSS

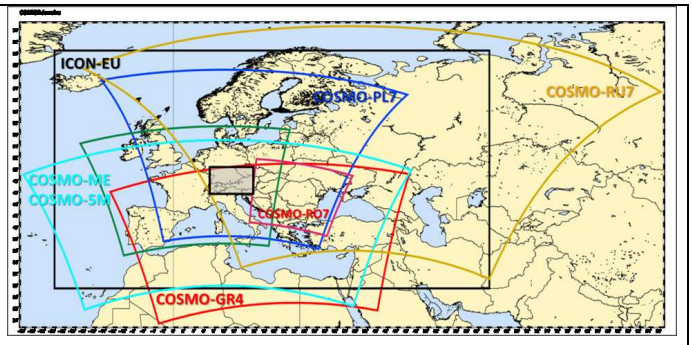
1 Overview

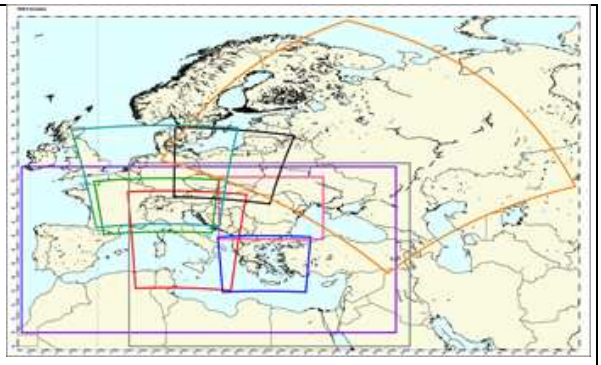
COSMO has implemented a procedure to perform homogeneous and comparable evaluation of model performance, which includes the calculation of verification scores over a common area, with the same observations, same methods and when possible with the same verification software. Verification results of statistical indices for main weather parameters derived using the operational COSMO and ICON-LAM model implementations in each service. The domain (common or custom), resolution, statistical scores/methods, frequency and graphical representation, are decided on an annual basis from WG5. The main findings of this organized analysis is presented during the GM plenary session together with the long term trend of them, providing a basis to track the performance of model. The use of common verification software allows for a standardized and objective way to apply, calculate and present the verification scores. Preparation of observation data and calculation of seasonal statistics are based on the guidelines that are derived on an annual basis from WG5. ICON-LAM models statistical results are included from any of the various services that use the model operationally. For JJA2020 and SON2020 analysis, only few centres have distributed ICON-LAM forecasts for evaluation but in the following seasons, COSMO has been gradually substituted with ICON-LAM especially for fine resolution implementations. Selective verification results of COSMO and ICON-LAM models over Common Area 1 (ComA1) and Area 2 (ComA2) are presented below while the complete selection of statistical results can be found in https://www.cosmo-model.org/content/tasks/verification.priv/default.htm.

2 Areas of Verification

The areas and specifications for model performance evaluation are presented below. In ComA-1, models with coarser resolution are included, while the higher resolution COSMO and ICON-LAM models are compared over ComA-2 (Table below).

doi:10.5676/dwd_pub/nwv/cosmo-nl_21_05

ComA-1 Area/Specs	
	00UTC Forecast runs Forecast Horizon: 72h Seasonal: JJA20, SON20, DJF21, MAM21
Models	Global ICON global, IFS LAMS DWD: ICON-EU COMET: COSMO-ME HNMS: COSMO-GR4 ARPA-E: COSMO-5M IMGW-PIB: COSMO-PL7 RHM: COSMO-RU7, ICON-RU7

ComA-2 Area/Specs	
	W10.963, S46.597, E17.437, N49.550 00UTC Forecast run Forecast Horizon: 48h Seasonal: JJA20, SON20, DJF21, MAM21
Models	**LAMS** DWD: COSMO-D2, ICON-D2 COMET: COSMO-IT, ICON-IT HNMS: ICON-GR2.5 ARPA-E: COSMO-2I IMGW-PIB: COSMO-PL2.8, ICON-PL2.5 MCH: COSMO-1E, COSMO-2E IMS: ICON-IMS

3 Results

3.1 Common Area 1

The models are evaluated in terms of Mean Error and Root Mean Square Error indices for the continuous parameters, using SYNOP observations over the Common Area domains using either VERSUS or MEC-Rbdfk software. Summary plots of continuous parameters are shown in plots 1a-d below,

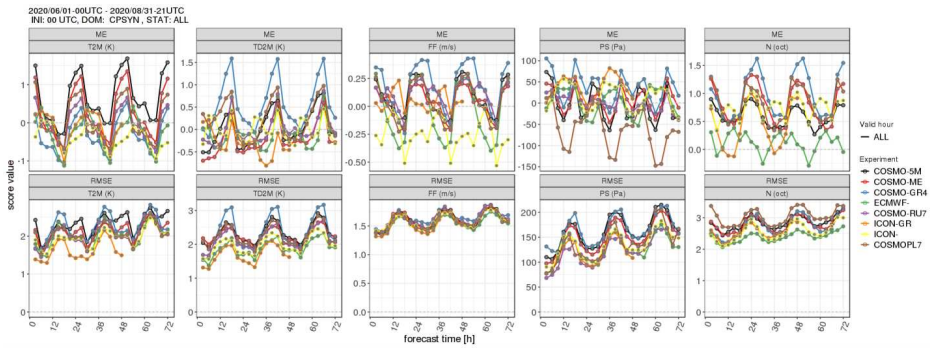


Figure 1: ME (first row) and RMSE (second row) for (from left to right) T2M, TD2M, Wind Speed, Pressure, Total Could Cover indices calculated over ComA-1 (JJA2020)

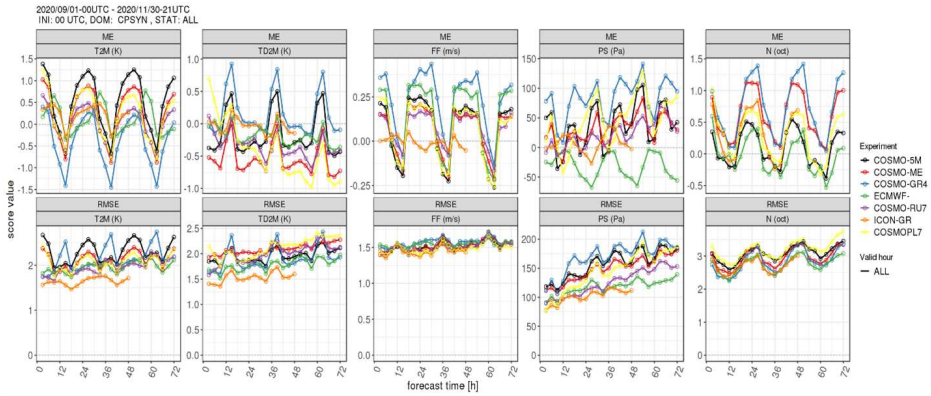


Figure 2: ME (first row) and RMSE (second row) for (from left to right) T2M, TD2M, Wind Speed, PS, Total Could Cover indices calculated over ComA-1 (SON2020)

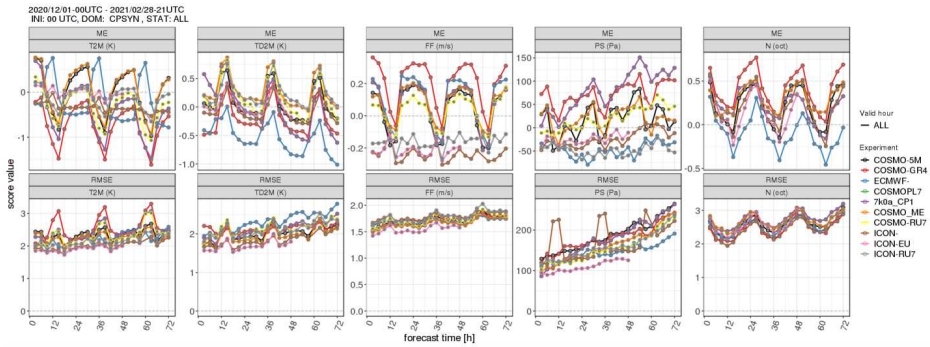


Figure 3: ME (first row) and RMSE (second row) for (from left to right) T2M, TD2M, Wind Speed, PS, Total Could Cover indices calculated over ComA-1 (DJF2021)

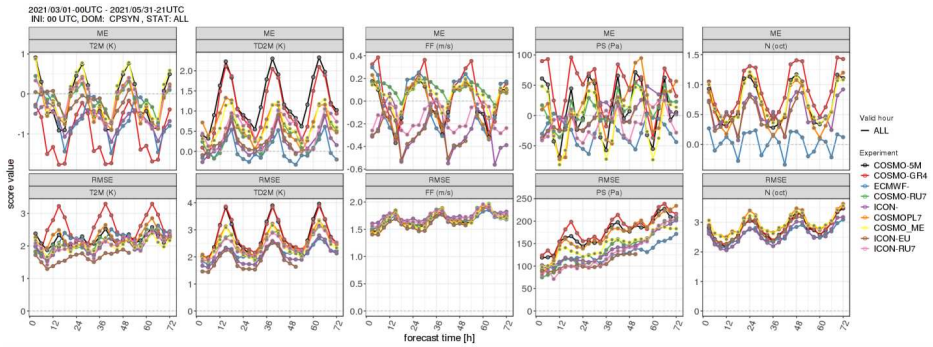


Figure 4: ME (first row) and RMSE (second row) for (from left to right) T2M, TD2M, Wind Speed, PS, Total Could Cover indices calculated over ComA-1 (MAM 2021)

For ICON models (ICON, ICON-EU, ICON-RU) and for IFS, the bias diurnal cycle is weaker and RMSE values are reduced for T2m and Td2m. RMSE wind speed is higher in warm hours of the day and comparable for all models in all seasons. The night time WS overestimation however which is a typical systematic error with both COSMO and IFS models for all seasons, is not apparent on ICON-LAM models, which exhibit a much weaker daily cycle, and an almost constant underestimation which is greater in JJA and MAM afternoon hours. RMSE for PS is reduced for ICON models, especially compared to COSMO ones and is similar to IFS forecasts. However, the tendency of PS RMSE increase with forecast time for all seasons and the diurnal cycle for JJA are comparable. PS bias is irregular for all seasons, and a more distinct difference is shown in DJF, with ICON models negative values all over the period, in contrast to positive and slightly time increasing COSMO values. TCC RMSE values and bias diurnal cycles are comparable for COSMO and ICON models with RMSE maximal values at night and better scores for DJF. ICON models produce slightly lower overestimation at night, and only IFS produces a weaker diurnal cycle for all seasons especially MAM. The point-wise 6h accumulated precipitation forecasts are evaluated in terms of categorical indices for different thresholds. JJA2020 and DJF2021 results for ETS, POD and FAR are presented in Figures 2a-b.

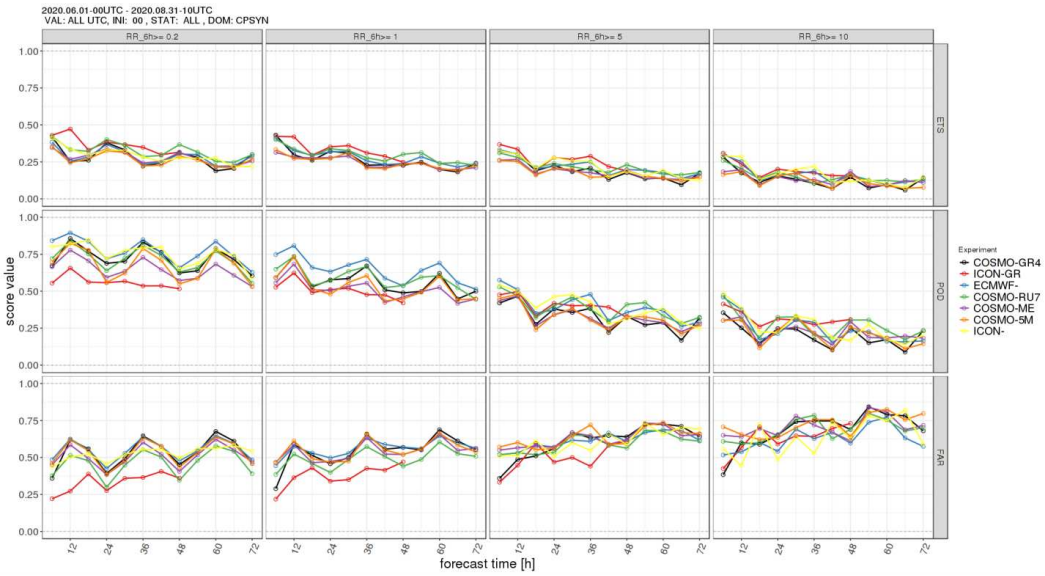


Figure 5: 6h accumulated precipitation indices for different thresholds. From top to bottom (ETS, POD, FAR). From left to right (0.2, 1, 5, 10mm) for JJA2020.

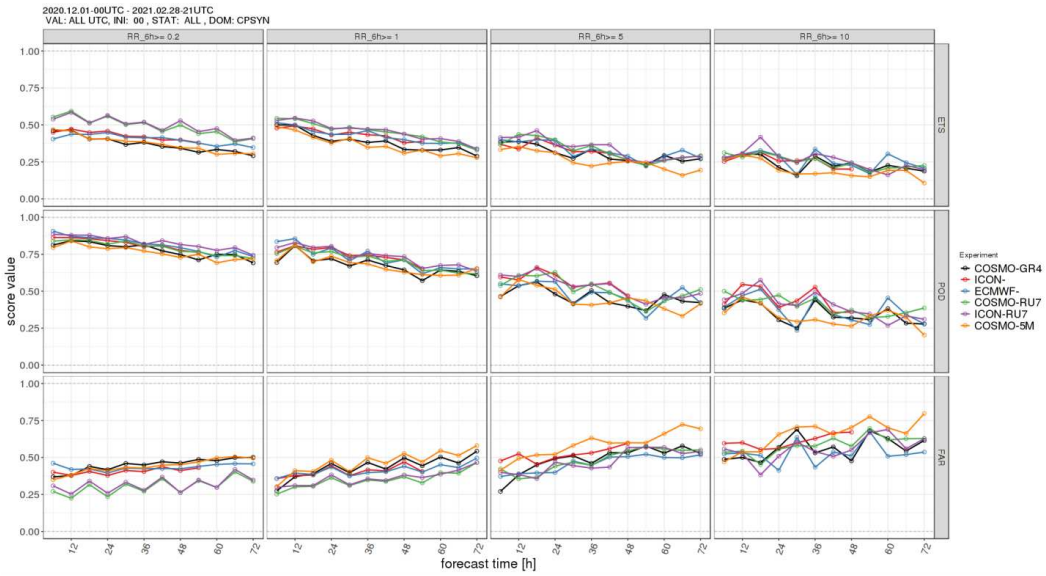


Figure 6: 6h accumulated precipitation indices for different thresholds. From top to bottom (ETS, POD, FAR). From left to right (0.2, 1, 5, 10mm) for DJF2021.

As in previous years, the diurnal cycle of the scores is distinct for JJA season and score trend worsens with forecast time and increasing threshold. Differences among ICON and COSMO models are not so distinct. However, for small precipitation thresholds, there is a slight improvement in ETS for ICON-GR, a clearer improvement in FAR and slight worsening for some model implementations in POD in JJA. This outcome reveals a dryer behaviour of ICON forecasts for that season compared to the observed values.

The results in the higher thresholds are quite variable and this can be attributed also to the smaller sample size. In DJF, the scores generally improve, they are more consistent while the diurnal variation is weaker. ICON-RU7 and COSMO-RU7 present similar behaviour, with better values of ETS and FAR than other models for low thresholds.

3.2 Common Area 2

For a better comparison among COSMO and ICON-LAM higher resolution models, they are presented grouped together in order to detect general tendencies or differences that can be attributed to the various implementations. DJF and MAM are the seasons with the greatest ICON-LAM forecasts availability. Temperature, wind speed and TCC results for these two seasons are presented below.

For T2m (Figures 3a,b), the most distinct difference, which is also consistent with coarser resolution models, is the reduced diurnal variability in bias values for ICON-LAM models, which is mostly apparent in DJF, with values closer to zero. The RMSE diurnal variation for DJF is comparable for the two sets of models. However, in MAM season RMSE for ICON models is reduced, with minimal diurnal variability.

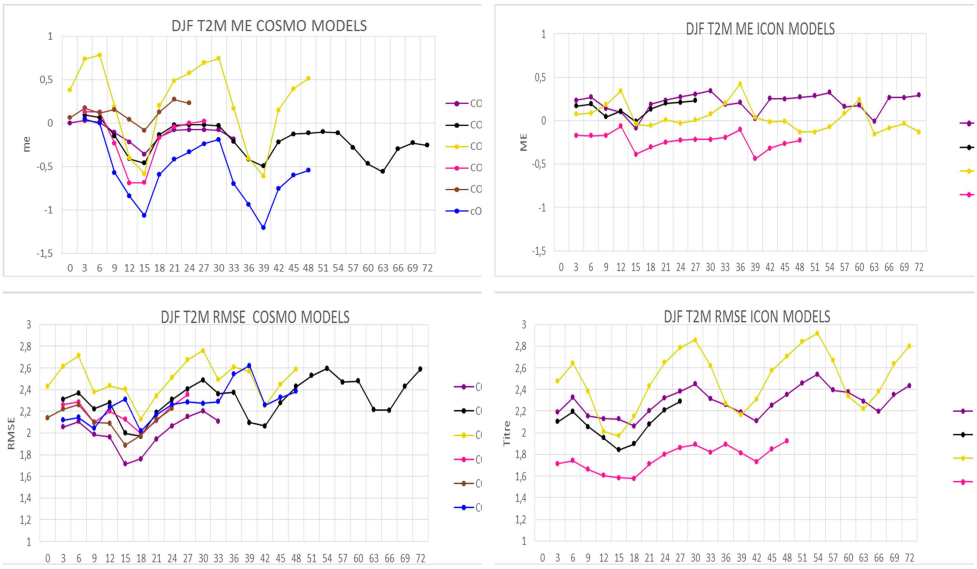


Figure 7: T2m ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for DJF2021

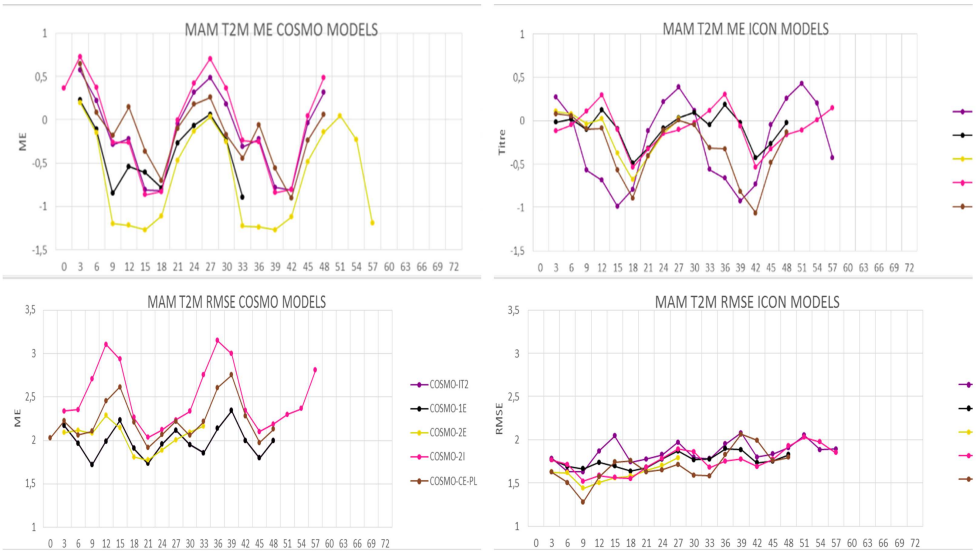


Figure 8: T2m ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for DJF2021

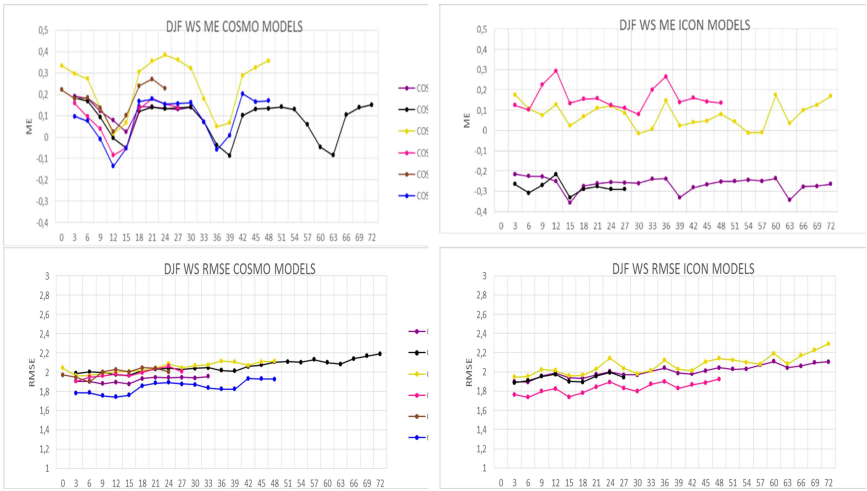


Figure 9: WS ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for DJF2021

Figure 10: WS ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for MAM2021

The indices for wind speed are presented in (Figures 4a,b), The DJF wind speed bias diurnal variation is weaker for ICON models and the negative tendency that was found for coarser resolution is now found in two models, while the RMSE error cycle and range are similar for both seasons, with higher values in warm hours in MAM (Figures 3c,d). Moreover, for MAM, the bias diurnal variability is shifted among the two sets of models, with COSMO overestimation in the early morning hours, while ICON models bias is positive around the afternoon.

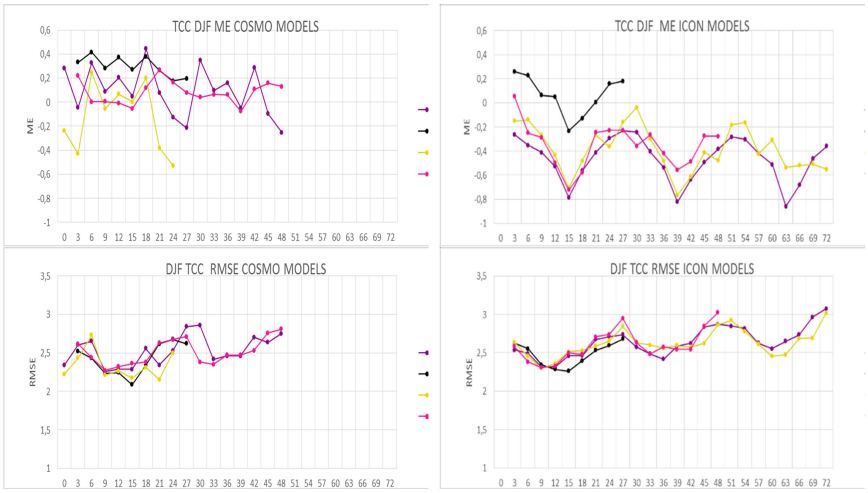


Figure 11: TCC ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for DJF2021

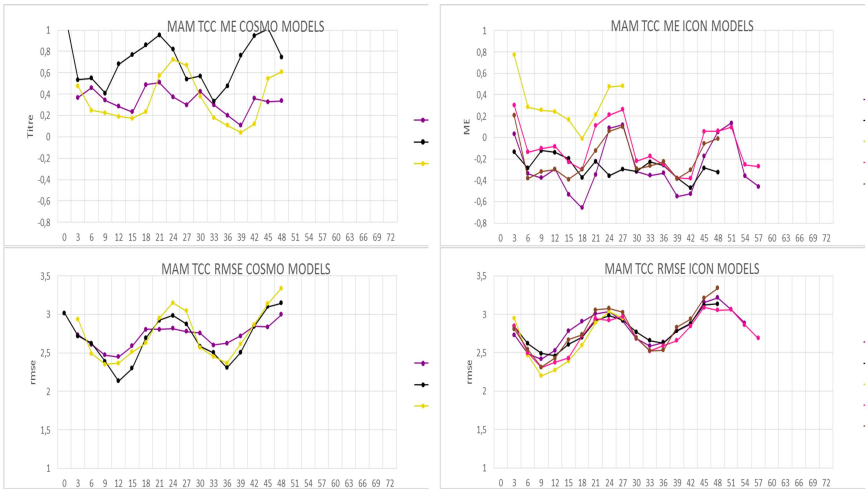


Figure 12: TCC ME (top) and RMSE (bottom). COSMO models (left) and ICON models (right) for MAM2021

The TCC bias difference among the two sets of models is clear, with ICON models exhibiting a diurnal cycle with underestimation especially in the warm hours for both seasons (Figures 5a,b), in contrast to TCC tendency to be overestimated by COSMO models (especially in MAM). RMSE diurnal cycle is similar for both sets, with higher values at night hours.

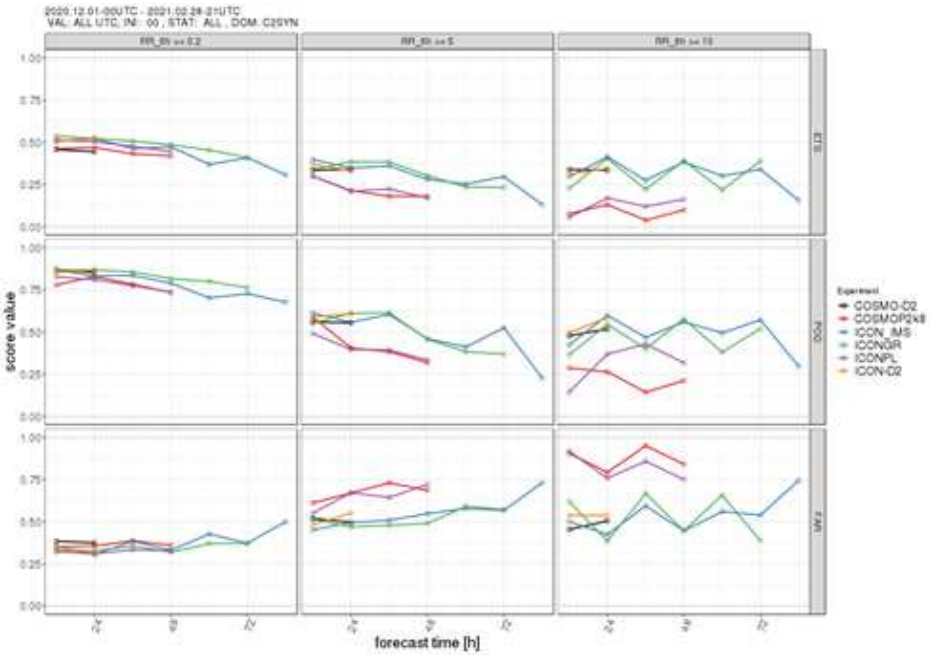


Figure 13: Precipitation scores ETS, POD, FAR (from top to bottom) for DJF2021 for thresholds 0.2,5, 10mm (from left to right) for ComA-2.

The precipitation scores for ICON and COSMO models are comparable for low thresholds, but the difference is apparent in higher thresholds. ICON-D2, ICON-GR and ICON-IMS scores are better for all indices, as it is shown in Figure 5 for DJF2021. However, ICON-PL scores which are comparable with COSMO-PL, are worse than other models, and this is result is in contrast to CM1 findings. Therefore, it cannot be extracted that ICON models outperform the COSMO for precipitation for all model configurations.

4 Fuzzy Verification Approach

In this section, fuzzy verification scores are presented for ComA-2 against the OPERA network radar composites. For this activity, VAST COSMO software is used that is based on Beth Ebert fuzzy verification IDL code. VAST main code utilizes txt gridded files for each weather parameter, but also a preprocessing of input files is available with the help of LIBSIM software. As these tools are based on grib1 format as input, while a preprocessing of ICON files needs to be performed beforehand. In the Table below, the specifications of this activity are given:

The main indices presented that summarize the spatial verification results, are the FAR (left), Fraction Skill Score (middle) and POD (right) (Fig. 6). The scores directly compare the forecast and the observation (radar) 3-hour gridded precipitation fields on continuously increasing spatial windows and for varying precipitation thresholds. The results for two different thresholds are presented in each graph, 0.1 and 5mm/3h for the first forecast day and for the three seasons (SON20, DJF21, MAM21). With spatial verification approaches, a relatively improved skill of ICON-LAM models compared to COSMO ones is shown in precipitation forecasts especially with respect to FAR and FSS scores while for POD score as also was extracted from the point-wise verification mainly for the smaller thresholds, the scores are in some cases slightly worse. The complete range of available plots of the indices can be found in: http://www.cosmo-model.org/content/tasks/verification.priv/default.htm.

DATA	**FORMAT**
Observation	OPERA composite (HDF-5) Mercator projection Resolution: 2km
Forecast	Model grib1 output format 00UTC run Horizon: up to 72h
Verification	Area: W10.97, S46.6, E17.42, N49.55

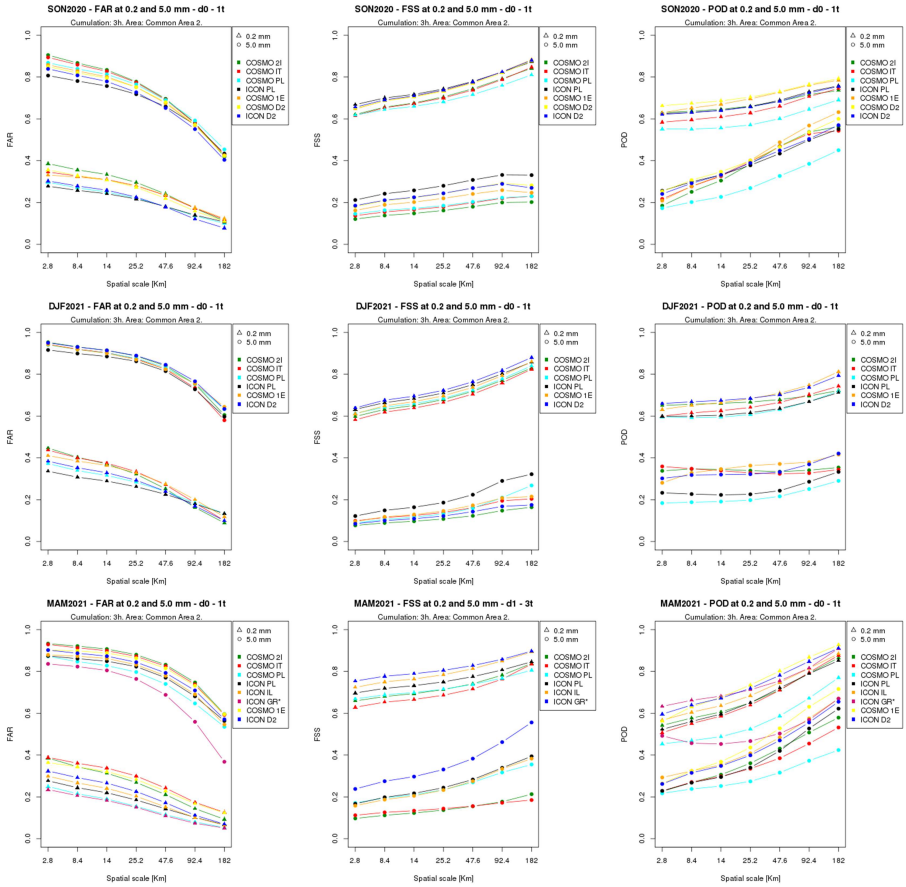


Figure 14: FAR (left), FSS (middle) and POD (right) 3h Precipitation scores for forecast day 1 (first raw) for SON20 (top row), DJF21 (middle row) and MAM21 (lower row) calculated over ComA-2.

5 Main Concerns based on Common Plot Analysis

The main score trends for COSMO models have not significantly changed from previous year.

Temperature: The general tendency is to underpredict the temperature diurnal cycle especially in the summer. ICON models bias diurnal cycle is weaker and RMSE values lower.
Wind speed: The bias diurnal cycle is more pronounced for COSMO models with nighttime wind speed overprediction. On the other side, there is a tendency of constant slight wind speed underestimation for ICON models mostly found in CA1. However, some ICON fine resolution models in CA2, slightly overpredict wind speed especially in morning hours.
TCC: RMSE for all models is higher at night and lower for DJF season with COSMO and ICON values being generally on the same range. TCC is over forecasted at night especially in warm seasons. ICON models diurnal variation is weaker, in addition to a tendency of TCC underestimation which is mostly found in CA2 finer resolution models. The categorical verification showed that the error is mostly found for 25-75% and that the cases ¿75% are overestimated especially for COSMO models.
PS: PS bias diurnal cycle is distinct in JJA and weaker for ICON models. In DJF, ICON models slightly under forecast PS in contrast to positive bias COSMO values. RMSE increases with time with lower ICON values.
T2d: Distinct T2d overprediction in warm hours for all models especially in JJA and MAM, with respective maximal RMSE. Weaker diurnal cycle and RMSE values for ICON models.
Precipitation: The daily variation of the scores for COSMO and ICON models are similar, and stronger for warm seasons. The performance decreases with forecast time and threshold. The differences among the two sets of models are not clear for all implementations even if from the spatial approaches an improved performance of ICON-LAM models is shown based on FAR and FSS scores.
Trend from last year: By comparing the mean seasonal RMSE for each variable over the years for CM1, and for each forecast hour, ($cosmo - model.org/content/tasks/verification$ $.priv/common/plots/default.htm$) the tendency for all models is a general small improvement of RMSE score for T2m, Wind Speed and T2d in comparison to last year, for all seasons. Regarding TCC, although the scores are significantly better for DJF and comparable for JJA and MAM, they are worse for SON season showing a slight RMSE increase. For the same season (SON), PS RMSE values also show a slight increase, while they are comparable for the remaining seasons.