SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2021
Project Title:	COSMO and ICON Numerical Weather Prediction Test Suite
Computer Project Account:	SPITRASP
Principal Investigator(s):	Amalia Iriza-Burca (NMA,Romania) [1] Ines Cerenzia (Arpae-SIMC, Italy) [2] Enrico Minguzzi (Arpae-SIMC, Italy)[2]
Affiliation:	National Meteorological Administration (NMA) [1] Regional Agency for Prevention, Environment and Energy of Emilia-Romagna – Hydro-Meteo-Climate Service (Arpae-SIMC) [2]
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Umberto Modigliani and his staff, Andrea Montani
Start date of the project:	2021
Expected end date:	2023

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	5 000 000	719 271.39 (14%)	5 000 000	3 932 197.75 (78%)
Data storage capacity	(Gbytes)	1000	1560	2000	11000

Summary of project objectives (10 lines max)

The COSMO and ICON Numerical Weather Prediction Test Suite Special Project intended to continue the activities started in the previous three special projects will ensure the usage of a homogeneous verification platform for both the COSMO and ICON models. This is meant as a benchmark in order to evaluate new versions of the model against exiting operational ones, prior to their official release. The aim of using this type of controlled approach for standardized testing and verification for the COSMO and ICON models is to ease the comparison of corresponding model versions (operational against new), in an effort to assess the impact of new features introduced in the code. The set-up and configuration of the models will focus on minimising initial and lateral boundary conditions effect, also eliminating the data assimilation system. Through this approach, performance of each new model version can be thoroughly tested, with an emphasis on newly introduced code developments.

Summary of problems encountered (10 lines max)

No problems encountered.

Summary of plans for the continuation of the project (10 lines max)

The detailed guidelines for the proper use and execution of each NWP test using this platform prepared during previous special projects related to this activity will be revised, considering both models (COSMO and ICON) and resolutions (7km and 2.8km) and all the corresponding model configurations. A detailed description of all steps will be included, from the compilation of a new COSMO model test version to the final production of the graphics for the statistical scores extracted. The Test Suite procedure will also be adapted and applied to the ICON numerical weather prediction model (limited area mode), with detailed guidelines for the proper use and execution of NWP tests using ICON, before the official release of new model versions.

Activities (including use of resources) will also include evaluating new official versions of the COSMO and ICON models prior to their release as well as maintenance of the Test Suite.

List of publications/reports from the project with complete references

I. Cerenzia, E. Minguzzi, A. Iriza-Burca, R. Dragomir, F. Gofa, F. Fundel (contributors) - *"Numerical Weather Prediction Meteorological Test Suite": COSMO 5.08 vs. COSMO 5.06 and ICON-LAM v2.6.1*, COSMO-Model Report, June 2021 (in preparation)

Summary of results

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

Phases I&II: Set-up of the COSMO and ICON models, Configuration and Execution of Runs

ECMWF computer resources were used for the aim of this task both for simulation and for archiving purposes, through billing units provided by the members as part of the SPITRASP special projects “Testbed for the Evaluation of COSMO Model Versions” approved for 2018-2020 and “COSMO and ICON numerical weather prediction test suite" approved for 2021-2023.

Since version 5.03 of the model, all versions are implemented on the Cray HPC. Starting from version 5.04a (quasi 5.05) of the COSMO model, the 2.8km horizontal resolution of the model is also tested using the NWP Suite, in addition to the previously used set-up at 7km. In addition to these configurations, starting with version 5.05, the single precision configuration (7 km) is also included in the test suite. From version 5.06 onwards, the forecast mode was replaced with the hindcast mode, with the aim to reduce computational costs and time for testing. The hindcast mode also allows to reduce the spin-up effect that may adversely influence the results when a series of cold-start forecast runs are performed (as opposed to a single continuous hindcast run). This is particularly important for slow soil variables.

COSMO v5.06 was used as a benchmark to evaluate COSMO v5.08. ICON-LAM model v2.6.1 was also used for comparisons, in view of the COSMO to ICON migration. The directory structure and archiving procedures for the COSMO-5.08, COSMO-5.06 and ICON-2.6.1 model versions follow the ones used for the previous implementations. For COSMO, the int2lm 2.05 version was used for the interpolation of initial and lateral boundary conditions. For ICON-LAM, DWD ICON Tools version 2.3.8 was employed for the same purpose. In all cases, model output was stored in grib2 format, to avoid problems encountered for the previous hindcast test (COSMO v5.06 against v5.05), especially when verifying precipitation.

After completion of the simulations, model output was processed together with the corresponding observations using the MEC (Model Equivalent Calculator) software, aimed at producing the necessary Feedback Files. Rfdbk (DWD developed) software that utilizes R libraries was used to process Feedback Files in order to produce verification scores. The VERSUS verification software was no longer used for new model version evaluation. The model output in grib2 format obtained from the experiments is locally stored in the ECFS system. The necessary software used for NWP Test suite is implemented either on cca (MEC) or ecgate (Rfdbk).

Configuration of COSMO model runs

For the testing of COSMO v5.08, evaluations are performed in hindcast mode (forecast +31days, with restart every 5 days), for both double and single precision model versions with the 7 km horizontal resolution set-up and only for double precision with the 2.8 km horizontal resolution configuration**.** Simulations were carried out for two one-month periods (one in the winter and one in summer), July and December 2017, for runs initialized by the 00UTC data. Initial conditions were provided by ECMWF HRES analysis, whereas lateral boundary conditions are introduced with a 3 hourly frequency and include the ECMWF HRES analyses (at hours 00, 06, 12 and 18UTC) and short cut off analyses (at hours 03, 09, 15 and 21UTC). Soil was initialized from ICON-EU, then runs free for both model resolutions.

COSMO v5.08 includes several fixes to the model, to run properly in single precision, such as the computation of cloud variables, the computation of the lightning potential index and so on. For the physical parameterizations, the mire parametrization has been introduced. The turbulence modules have been aligned between COSMO and ICON and a bug has been fixed in the Tiedtke-Bechtold convection Scheme. Some fixes and a modification to enhance performance of some GPU parts have been introduced. Some changes have been added in the GNSS STD Operator, in the Data Assimilation and in the Latent Heat Nudging packages and a new Wind Gust Tuning has been introduced. Finally, a lock-file mechanism in the writing of output files has been added.

The simulations with versions 5.06 and 5.08 use essentially the same namelists; the only difference is that keywords “nradcoarse” and “lradf_avg” in namelist PHYCTL have been excluded from 5.08 simulations since they are not supported anymore.

For the two COSMO model versions, the following configurations were employed:

- runs at 7.0 km, 40 model levels; hindcast mode, DP
- runs at 7.0 km, 40 model levels; hindcast mode, SP
- runs at 2.8 km, 50 model levels; hindcast mode, DP

Configuration of ICON-LAM model runs (v2.6.1)

Configuration of the ICON-LAM test suite followed that employed in previous years for the COSMO Test Suite. This entailed a duplication of the COSMO NWP Test Suite in the Ecflow format. Simulations were carried out for the same one-month periods (one in the winter and one in the summer), July and December 2017, in hindcast mode (forecast +31days, with restart every 5 days). For the ICON-LAM simulations, as well as for the COSMO runs, restarts are only made for technical reasons and have no effect of the continuous hindcast run.

For ICON v2.6.1, the following model configurations were used:

- runs at 7.0 km, 40 model levels (ICON-LAM-7p0/40lev)
- runs at 2.5 km, 50 model levels (ICON-LAM-2p8/50lev)
- runs at 7.0 km, 65 model levels (ICON-LAM-7p0/65lev)
- runs at 2.5 km, 65 model levels (ICON-LAM-2p8/65lev)

Initial and lateral boundary conditions were provided by ECMWF HRES analysis and forecast (at 03, 09, 15, 21UTC, with 3 hours forecast-range), while soil information was initialized from ICON-EU, then free soil (both resolutions). The integration domain for ICON-LAM-2p8 was nested in ICON-LAM-7p0. COSMO and ICON-LAM domain characteristics are shown in Figure 1 and Table 1.

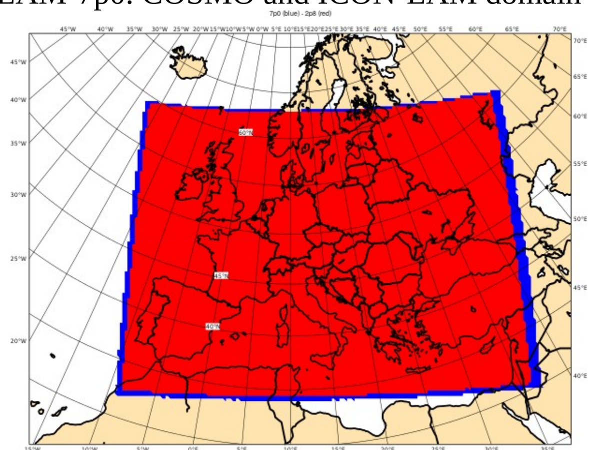

Fig. 1 Left: Integration domain for all COSMO model versions, 7km DP and SP (blue) and 2.8km (red). Right: Integration domain for the ICON-LAM model.

Table 1: Main features of the models used in the NWP Test Suite.

	ECMWF HRES	COSMO 7p0	COSMO 2p8	ICON-LAM-7p0		ICON-LAM-2p8	
Grid points (nx x ny)	901 x 501	661 x 471	1587 x 1147	Icosahedron-triangular R3B8 grid; 309,560 cells		Icosahedron-triangular R2B10 grid; 2,005,580 cells	
Model levels	137	40	50	40	65	50	65
Resolution (dx x dy)	0.1 x 0.1	0.0625 x 0.0625 (7km)	0.025 x 0.025 (2.8km)	(6.6km)		(2.5km)	

Phase III: MODEL OUTPUT VERIFICATION

The steps followed for the MEC+Rfdbk verification procedure:

- ➔ conversion of observations from bufr to netcdf format (using *bufr2netcdf*)
- ➔ pre-processing of model output in grib format for ingestion in MEC
- ➔ processing model output and corresponding observations to obtain feedback files (MEC)
- ➔ execution of verification procedures (Rfdbk)
- ➔ transfer and visualisation of results on the COSMO shiny server

The verification was performed with grid-to-point comparisons in order to compare gridded surface and upper-air model data to point observations. The selected NWP suite stations are situated in an area covering -25/24/65/65 (W/S/E/N) and are around 3200. Due to the requirements of the MEC software, all observations are previously converted in netcdf format with the bufr2netcdf software.

The verification modules for testing COSMO v5.08 are the following:

- ➢ **surface continuous parameters:** 2m temperature (T2M), 2m dew point (TD2m), 10 meter wind speed (FF), total cloud cover (N), surface pressure (PS): mean error (ME), root mean square error (RMSE), mean absolute error (MAE), standard deviation (SD), R^2, TCC (tendency correlation), LEN (number of observations used), OMEAN and FMEAN (observed and forecast mean), etc.;
- ➢ **precipitation verification** (6h, 12h) for selected thresholds (greater than 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30): probability of detection (POD), false alarm rate (FAR), equitable threat score (ETS), requency bias (FBI), Performance diagrams, etc.
- ➢ **upper air verification (TEMP based):** temperature (T), relative humidity (RH) and wind speed (FF) for selected pressure levels (250., 500., 700., 850., 925., 1000.): BIAS, MAE, RMSE, SD, etc.

Verification was performed taking into account all configurations of the models presented above, which allowed for an extended comparison. These include different resolutions, model versions and SP/DP for two seasons. The verification results presented in the following section (figures 2-8) are a sample of the derived statistics, with additional comparisons to the ones presented below also available. The complete overview of all the statistical analysis (graphs and numbers) available at http://www.cosmo-model.org/shiny/users/fdbk/ . For a detailed description of the system set-up and model evaluation, we refer to the COSMO model report by Cerenzia et. al (2021). Due to the specifications of the verification system for hindcast runs (single run), +24 hours lead time is shifted to 0.

For T2M (Figure 2), the differences between the two COSMO versions are insignificant with respect to RSME for all different comparisons and both seasons. For winter, a small improvement of RMSE can be identified with COSMO v5.08 7km SP version of the model. ME graphs exhibit an overestimation during the night and early hours of the day in the summer and an underestimation for all hours in the summer. The statistics with 7km resolution models seem to perform slightly better while the best performance attributed to COSMO v.08 SP model. The comparison between COSMO v5.08 and ICON-LAM v2.6.1. is independent of resolution, season and precision and exhibits a distinct improvement of both ME and RMSE scores with ICON-LAM configurations. The diurnal cycle of error is much reduced especially in the winter, while the error is reduced around 0.5 deg and ICON 2.8km with 65 levels seems to have the best performance.

For FF (Figure 3), NWP test statistical results exhibit almost identical values for both seasons for models with the same resolution only, with no significance change in the performance with COSMO v5.08. The higher resolutions models show smaller RMSE values with no change for the new COSMO version, while the models exhibit the same behaviour, mainly overestimating for the 2.8km resolution. ICON-LAM model for both versions outperforms all COSMO models. RMSE values are reduced by around 0.2-0.3m/s, while the trend of ME has significantly changed to underestimation for all seasons of 10 meter wind speed, which is larger with the forecasts derived from ICON 7km versions.

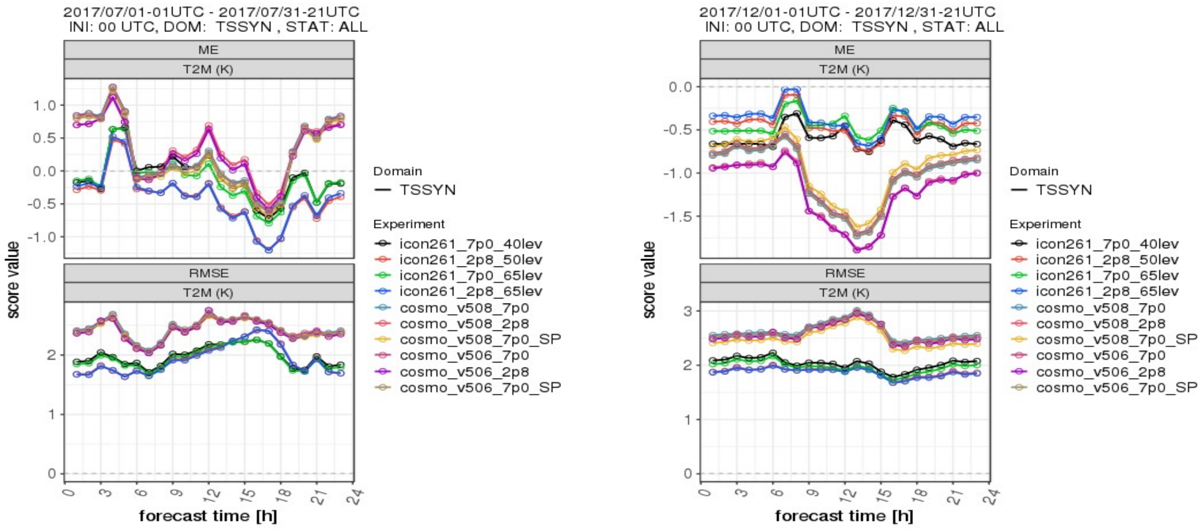

Fig. 2 T2M verification results - July 2017 (left), Dec 2017 (right), for: ICON-7p0/40levels (black), ICON-2p8/50levels (red), ICON-7p0/65levels (blue), ICON-2p8/65levels (green), COSMO-v5.08-7p0 DP (purple), COSMO-v5.08-2p8 (orange), COSMO-v5.08-7p0 SP (yellow), COSMO-v5.06-7p0 DP (brown), COSMO-v5.06-2p8 (pink) and COSMO-v5.06-7p0 SP (grey); ME (top) and RMSE (bottom). Red/gray filled dots indicate a significant/insignificant (95% level) difference of scores between the model versions.

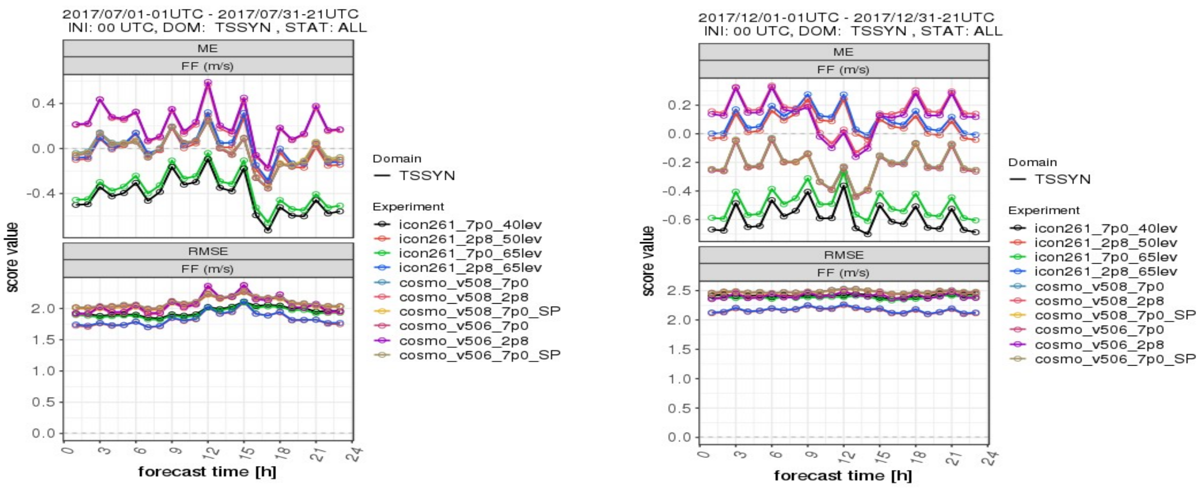

Fig. 3 FF verification results - July 2017 (left), Dec 2017 (right), for: ICON-7p0/40levels (black), ICON-2p8/50levels (red), ICON-7p0/65levels (blue), ICON-2p8/65levels (green), COSMO-v5.08-7p0 DP (purple), COSMO-v5.08-2p8 (orange), COSMO-v5.08-7p0 SP (yellow), COSMO-v5.06-7p0 DP (brown), COSMO-v5.06-2p8 (pink) and COSMO-v5.06-7p0 SP (grey); ME (top) and RMSE (bottom). Red/gray filled dots indicate a significant/insignificant (95% level) difference of scores between the model versions.

Regarding the forecast of 6h precipitation (Figure 4), the statistics for the two versions of the COSMO model are almost identical, with slightly more visible differences for thresholds higher than 10mm, mainly for the FBI score, for both periods and resolutions. Small differences can be noticed in all scores in the higher thresholds for the winter period, for the 7km resolution of the model that seems to perform slightly better than the 2.8km one. With respect to the DP versus SP comparison, results are almost identical also for December, with some differences in FBI for the higher threshold categories. From the comparison of the performance between ICON and COSMO models, an improvement in the performance is noticed in both seasons. ETS is increased by around 0.05-0.1, POD by also around 0.1 especially for small thresholds while FBI indicates that for the lowest threshold ICON model tend to overestimate precipitation opposite from COSMO models.

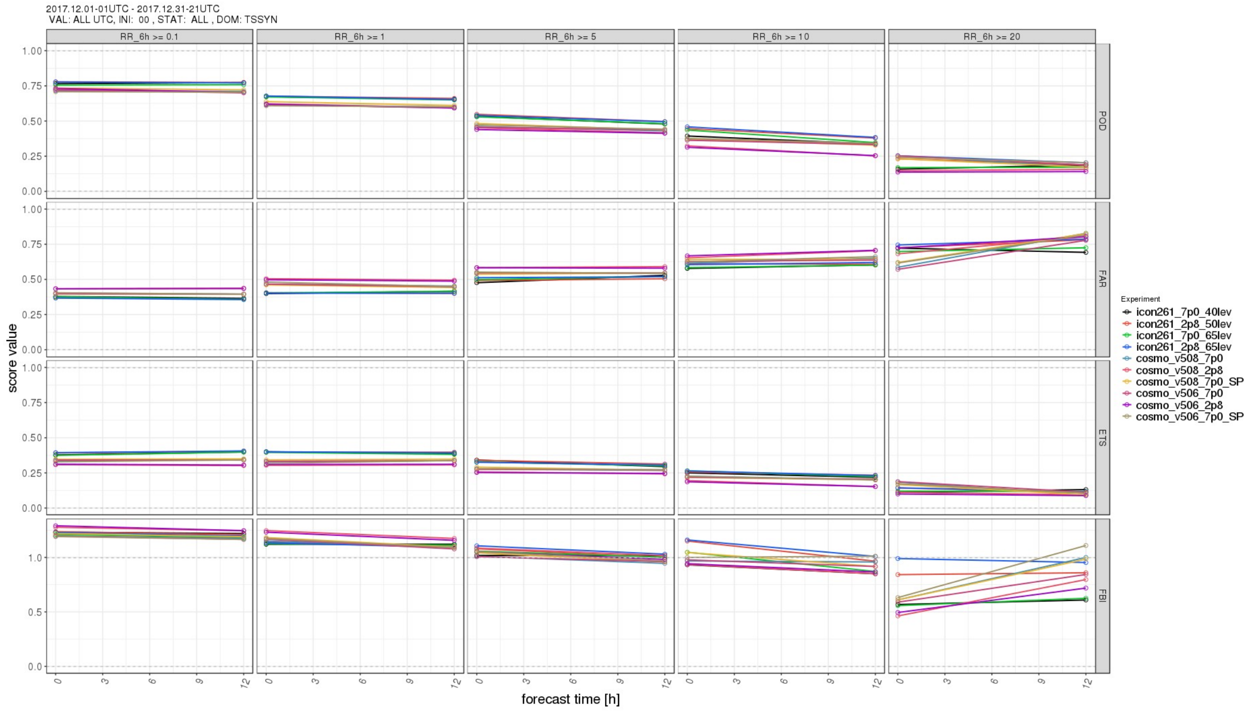

Fig. 4 RR_6h verification results - December 2017, for: ICON-7p0/40levels (black), ICON-2p8/50levels (red), ICON-7p0/65levels (blue), ICON-2p8/65levels (green), COSMO-v5.08-7p0 DP (purple), COSMO-v5.08-2p8 (orange), COSMO-v5.08-7p0 SP (yellow), COSMO-v5.06-7p0 DP (brown), COSMO-v5.06-2p8 (pink) and COSMO-v5.06-7p0 SP (grey); POD, FAR, ETS and FBI (top to bottom). Thresholds 0.1, 1, 5, 10, 20mm/6h (left to right).

The scores for the performance of upper air parameters generally show similar behaviour for both versions of the COSMO model. Upper air RH (Figure 5) exhibits small differences between the two COSMO versions. For the winter period, there is a small reduction in the overestimation of RH during the night in the middle atmosphere. During the warm hours of the day there is also a reduced underestimation, but smaller. During summer, the new version performs again slightly better, with reduced overestimation of RH from the surface up to almost 500mb, while during the night there is no difference between the two model versions. RMSE values exhibit overall no significant difference. With the 2.8km version, values are almost identical for both months.

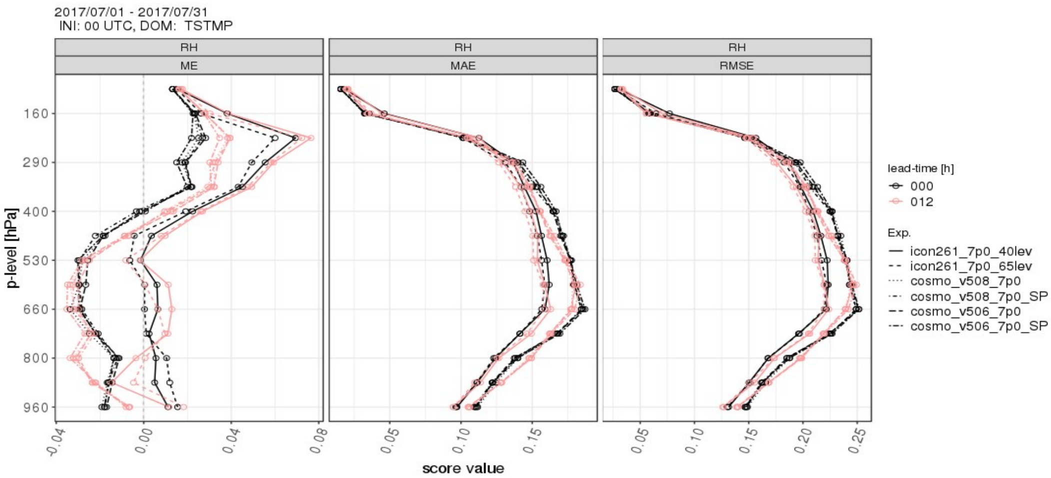

Fig. 5 Upper Air RH verification results - July 2017, +00/24 hours (black) and +12 hours (red) for: ICON-7p0/40levels (solid), ICON-7p0/65levels (dashed), COSMO-v5.08-7p0 DP (dotted), COSMO-v5.08-7p0 SP (dotted – dashed – dotted), COSMO-v5.06-7p0 DP (solid – dashed) and COSMO-v5.06-7p0 SP (dotted – solid – dotted); ME, MAE and RMSE (left to right).

However, the differences between COSMO v5.08 and ICON-LAM for the upper air parameters is significant (Figure 6). For example, ICON-LAM statistical values for the winter indicate a reduced underestimation of upper air temperature that is characteristic for the lower atmosphere which leads to an improvement in RMSE values compare to COSMO.

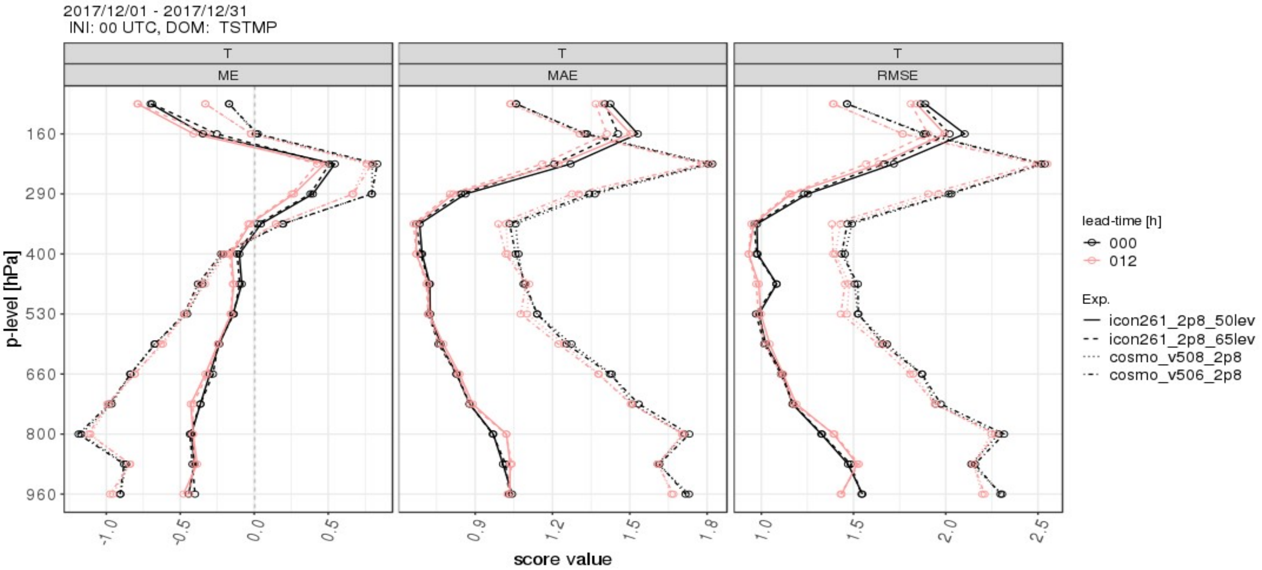

Fig. 6 Upper Air TEMP verification results - December 2017, +00/24 hours (black) and +12 hours (red) for: ICON-2p8/50levels (solid), ICON-2p8/65levels (dashed), COSMO-v5.08-2p8 (dotted) and COSMO-v5.08-2p8 (dotted – dashed – dotted); ME, MAE and RMSE (left to right).

For the first comparison of ICON-LAM configurations (v2.6.1), statistics were mainly grouped according to the resolution. For example, TD2M results during summer (Figure 7) show there is nearly no effect in the performance on ICON-LAM based on resolution or vertical levels, while for the winter there is small reduction of RMSE with the higher resolution models. The underestimation of the parameter is reduced both with increasing resolution and with increased vertical levels but this is happening as in the case of 2 meter temperature, only during winter, while the behaviour is opposite during summer.

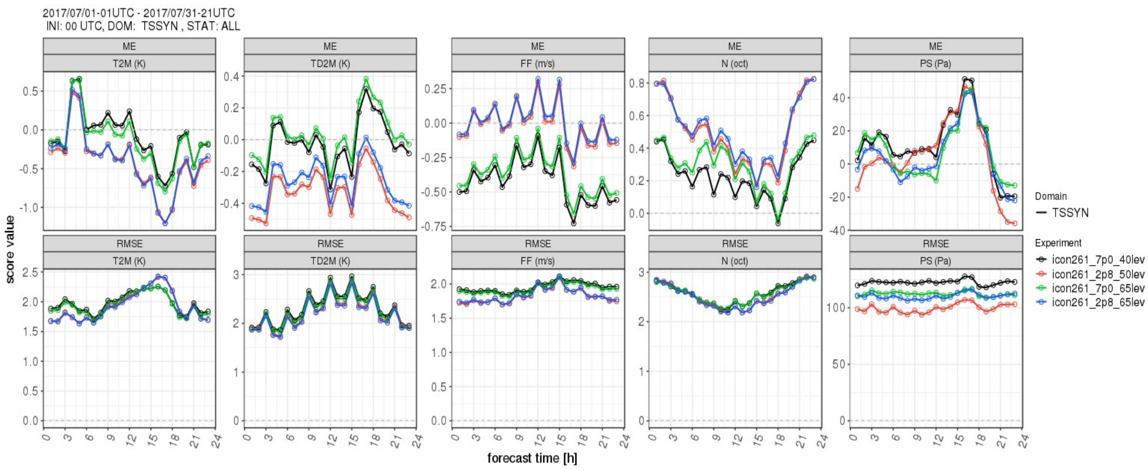

Fig. 7 ICON-LAM v2.6.1 Continuous parameters verification results – 7p0/40levels (black), 2p8/50levels (red), 7p0/65levels (green), 2p8/65levels (blue); ME (first row) and RMSE (second row) for: July 2017 (a), Dec 2017 (b). Red/gray filled dots indicate a significant/insignificant (95% level) difference of scores between the model versions. Parameters (from left to right): T2M (K), TD2M (K), FF (m/s), N (oct) and PS (Pa).

For PS, during summer (Figure 7) there is a noticeable reduction of RMSE with the ICON-LAM 2.8_50lev model implementation, while 7km_65lev and 2.8km_56lev have almost identical error time series. The effect of resolution and levels is much reduced during winter month. All ICON implementations exhibit an overestimation in the afternoon hours during summer, with a dependency on vertical levels. During winter, this overestimation is in all hours of the day, but is reduced with the 2.8k resolution.

No significant effect on either 6h or 12 hour accumulated precipitation statistics from either resolution or number of vertical levels from ICON-LAM is visible during the winter period. For the 6 hours accumulation, during summer, the results are inconsistent, most likely due to the restricted number of events.

With regards to upper air relative humidity and temperature, there are generally no significant changes in the performance of ICON-LAM integrations with either resolution or number or vertical levels in RMSE. Some small changes in ME are visible. For upper air wind speed (Figure 8), however, there is a significant improvement in all levels with the higher resolution implementation. Also, a small impact on the scores is observed from the number of vertical levels.

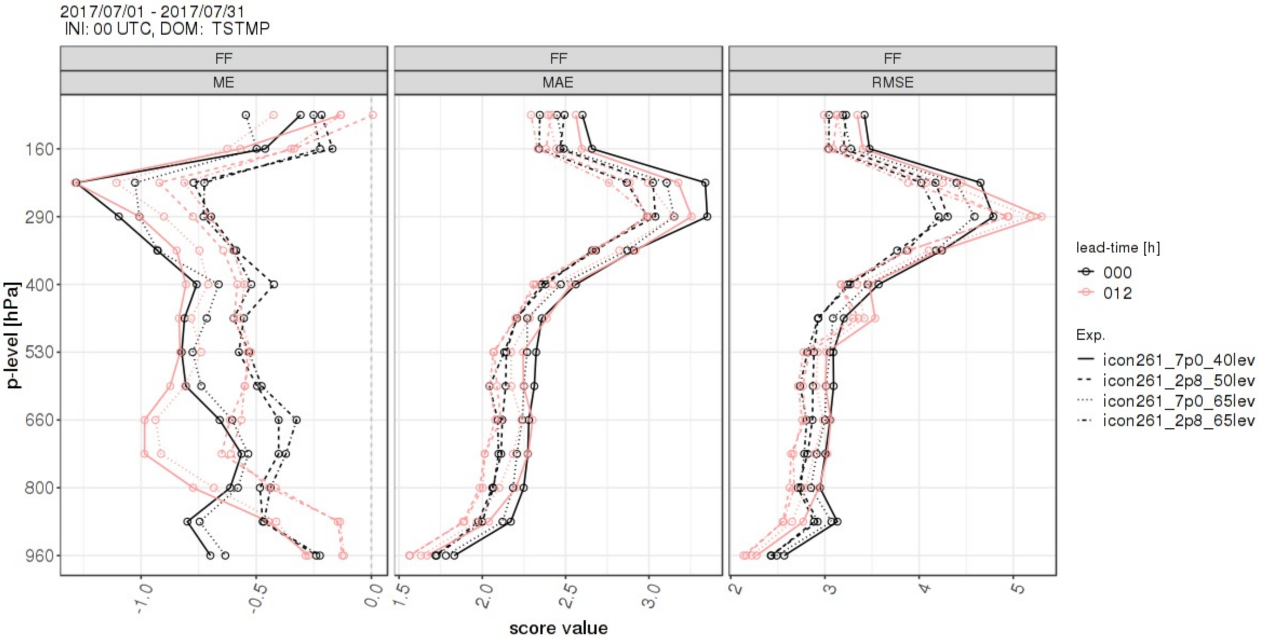

Fig. 8 ICON-LAM v2.6.1 Upper Air wind speed verification results – 7p0/40levels (solid line), 2p8/50levels (dashed line), 7p0/65levels (dotted line), 2p8/65levels (dashed+dotted line); ME (left), MAE (center) and RMSE (right) for December 2017, +00/24 hours (black) and +12 hours (red).

According to the verification results obtained with the COSMO and ICON Numerical Weather Prediction Test Suite, the newer version of the model, on the whole, exhibits no significant changes in the performance, while in a few cases outperforms its predecessor. As for the SP-DP comparison, there are insignificant differences in the calculated indices and moreover a smaller amount of computer resources were required to run the SP versions of the model.

The more striking changes in all scores were shown in the comparison of ICON-LAM to COSMO implementations for almost all cases especially for the surface parameters. As COSMO consortium has decided to move to ICON-LAM model and cease any further COSMO development, this is a very optimistic message for the future.