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	2025
Project Title:	ICON NUMERICAL WEATHER PREDICTION METEOROLOGICAL TEST SUITE
Computer Project Account:	SPITRASP
Principal Investigator(s):	Rodica Dumitrache (NMA, Romania) [1] Enrico Minguzzi (Arpae-SIMC, Italy)[2]
Affiliation:	[1] National Meteorological Administration (NMA) Regional Agency for Prevention, Environment and [2] Energy of Emilia-Romagna – Hydro-Meteo-Climate Service (Arpae-SIMC)
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Umberto Modigliani and his staff, Andrea Montani
Start date of the project:	2024
Expected end date:	2026

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)	7,500,000	373,490	7,500,000	1,648,026
Data storage capacity	(Gbytes)	6000	30 000	6000	75 000

Summary of project objectives (10 lines max)

The ICON Numerical Weather Prediction Test Suite Special Project continues the activities started in the previous three special projects, therefore ensuring the usage of a homogeneous verification platform for all versions of ICON model. This is meant as a benchmark in order to evaluate new versions of the model against existing operational ones, prior to their official release. The aim of using this type of controlled approach for standardized testing and verification 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 model versions will focus on minimising initial and lateral boundary conditions effect, also eliminating the data assimilation system. Through this approach, performance of new model versions 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 the Atos platform prepared during previous special projects related to this activity will be revised considering the ICON model and corresponding model configurations. A detailed description of all steps will be included, from the compilation of a new ICON model test version to the final production of the graphics for the statistical scores extracted. Activities (including use of resources) consist in evaluating new ICON versions, as well as maintenance of the Test Suite.

List of publications/reports from the project with complete references

F. Gofa – "Verification in COSMO consortium", COSMO Newsletter, No. 23, July 2024
M. Milelli and colleagues – "WG Support and Infrastructure (former WG6), The 26th COSMO General Meeting,Offenbach, Germany, 2-6 September 2024
"Numerical Weather Prediction Meteorological Test Suite: S23icon202410 vs S24icon202410 vs S23icon265", COSMO Technical Report, *in preparation*

Summary of results

The activities performed during the project were dedicated to running and evaluating the performance of the ICON-LAM model with the numerical weather prediction test suite running on the Atos HPC. These activities include:

- Implementation and running the ICON-LAM model on the Atos system (model configuration and integration, processing of model output for production of feedback files)
- Running of the MEC system for production of feedback files
- Update and running of the FFV2 (previously Rfdbk) package dedicated to the calculation of statistical scores.

Phase I: Set-up of the ICON model

The ICON NWP Test Suite follows the implementation previously employed for the COSMO Test Suite, adapted for the Atos machine and tailored to the ICON-LAM model.
The design of the experiments is the same as in the previous reports: the simulations are carried out for the same one-month periods, one in winter (December 2021) and one in summer (July 2021). For each simulation, the model is run in "hindcast" mode (i.e. forced with analysed boundary conditions),

producing a continuous 31 day forecast. For implementation reasons, the forecast is restarted every 5 days, but this does not affect the continuity of the simulations.

The main simulation settings of all the experiments are:

- horizontal resolution: 2.5 km (R2B10; 1,997,000 cells). The integration domain is shown in Figure 1
- vertical resolution: 65 levels
- time step 24"
- initial and lateral boundary conditions from ECMWF HRES analysis; since analysis data are only available at 6 hours intervals, short term (1-5 hours) forecasts are used to fill the gaps.
- soil variables initialized from ICON-EU, while the soil is free to evolve during the simulation
- SST and sea ice fields updated every 24 hours from IFS analysis.

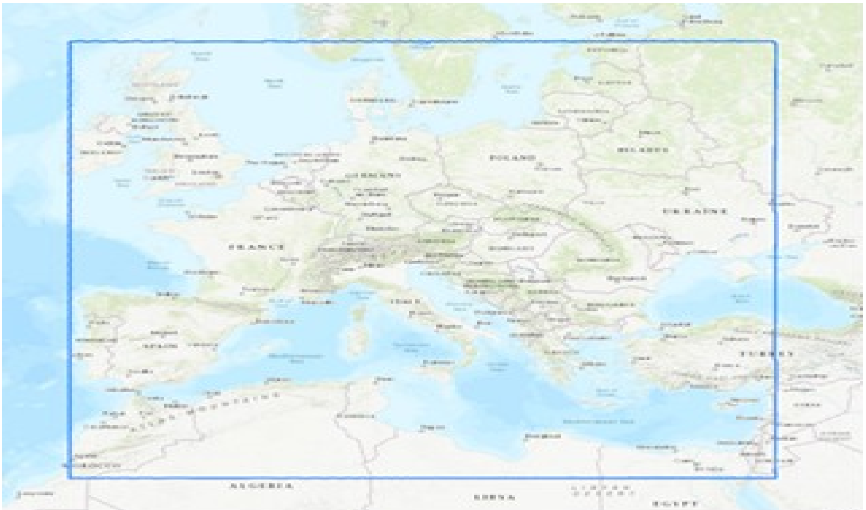


Figure 1. Integration domain for the ICON-LAM model at 2.5km horizontal resolution.

Phase II: Configuration ICON-LAM experiments

Two model versions were employed for the present tests: ICON version 2.6.5.1 and version 2024.10. Based on these two model versions, this report describes three ICON-LAM experiments (see table 1), as follows:

1) S23icon265: standard configuration of ICON version 2.6.5.1. This experiment has already been described in detail in the 2024 progress report, and already includes the changes in the experiment setup that were decided in 2023: new simulation periods (2021 instead of 2017), updated namelists, new soil IC form ICON-EU) and slightly smaller domain (SE corner), as well as a general revision of the ecflow suite.

2) S23icon202410: standard configuration of ICON version 2024.10. The setup is the same as in the previous experiment, but the change in model version required several modifications to ICON namelists (see table 3). Moreover, the climatological concentrations of trace gases have been updated, to take into account changes in atmosphere composition in the last decades (this has a small effect on the description of radiative exchanges).

3) S24icon202410: ICON configuration same as in S23icon202410, but two changes have been introduced in this setup of the experiment: IFS atmospheric boundary conditions are taken every hour (instead of every 3 hours), and the "Top Boundary Nudging" option has been activated. As a result, synoptic scale circulation over Europe is expected to follow more closely ECMWF analysis.

A summary for the main features of the three ICON-LAM experiments described in this report and the most important namelist changes between icon versions 2.6.5.1 and 2024.10 are presented in tables 1 and 2.

The first purpose of the ICON Test Suite is to check if there is any flaw in the implementation of the new icon releases. Moreover, these three experiments were designed to investigate two specific questions.

- Does ICON version 2024.10 introduce **significant improvements in model performance**? This can be assessed by comparing experiments S23icon265 and S23icon202410
- Does the use of **Top Boundary Nudging and hourly BC affect** model performance? This can be assessed by comparing experiments S23icon202410 and S24icon202410

Table 1: Summary for the main features of the three ICON-LAM experiments described in this report.

Experiment ID	Completed by	Icon version	Simulation periods	Namelist	Soil IC + domain	Top boundary nudging	BC frequency
S23icon265	11/08/2023 17/08/2023	2.6.5.1	Jul 2021 Dec 2021	265-2023	ICON-EU	No	3 hours
S23icon202410	22/12/2024 19/12/2024	2024.10	Jul 2021 Dec 2021	2024	ICON-EU	No	3 hours
S24icon202410	31/01/2025 04/02/2024	2024.10	Jul 2021 Dec 2021	2024	ICON-EU	Yes	1 hour

Table 2: The most important namelist changes between icon versions 2.6.5.1 and 2024.10.

namelist	key	S23icon265	S23icon202410
nwp_phy_nml	tune_gkdrag	0	0,125
nwp_phy_nml	tune_grcrit	0,25	0,65
nwp_phy_nml	tune_gust_factor	7,25	7
nwp_phy_nml	a_hshr	2	1,25
nwp_tuning_nml	itune_gust_diag	3	4
nwp_tuning_nml	tune_gustlim_fac	0	1,75
extpar_nml	pp_sso	1	2
radiation_nml	vmr_co2	3.9E-4	4.25E-4
radiation_nml	vmr_ch4	1.8E-6	1.9E-6
radiation_nml	vmr_n2o	3.22E-7	3.34E-7
radiation_nml	vmr_cfc11	2.4E-10	2.2E-10
radiation_nml	vmr_cfc12	5.32E-10	4.90E-10

Phase III: Model Output Verification

The Model Equivalent Calculator (MEC) software for the production of Feedback Files and verification scripts based on the R package FFV2 previously implemented on the ATOS system were employed for Model output Verification, as follows:

- production of feedback-files using MEC (performed on the **Atos HPC machine;** employs part of the available billing units
- production of model output verification based on feedback-files using the FFV2 (performed on the **ECS** interface)

- additionally, conversion of observations from bufr to netcdf format (using *bufr2netcdf*) can also be performed on the **ECS** interface, using the **bufr2netcdf** utility

The Feedback files used in the verification procedure are produced by MEC and contain all information regarding observations and their usage in the data assimilation system. They are ingested in FFV2, which then employs them to compute the verification scores. The production of Feedback files and verification procedures are based on observations datasets available from the MARS database and converted from bufr to NetCDF format locally.

Characteristics and requirements of the MEC processing chain on Atos:

- pre-processing of model output files stored on ECFS (available after Phases I & II): model output files stored as grib2 files containing 24 time steps each are split into hourly or three hourly files; separate files are obtained for accumulated parameters;
- preparation of input files required by MEC: constant files, model grid file description, forecast files, observations
- set-up of MEC namelist file and run scripts
- production of feedback-files using MEC

The costs for producing a month of feedback files for one model configuration (including pre-process ing of model output files) are around 125000 SBUs. The total resources for MEC and FFV2 used for this project are 540803 SBU.

Characteristics of FFV2 on ECS:

An objective model output verification is performed based on grid-to-point comparisons that enable a correspondence between gridded surface and upper-air model data to point observations, similar to the previous VERSUS verification procedures employed in the past to evaluate COSMO and ICON model versions. Around 3200 stations situated in an area covering -25/24/65/65 (W/S/E/N) are employed for the stratification (see figure 2). Suspect observation values were previously included in the verification test in order to minimize errors introduced from the observations.

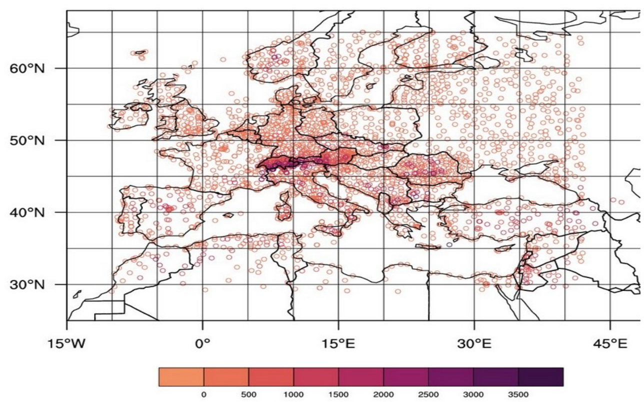


Fig. 2 Overview of meteorological observations used for the verification.

The parameters evaluated using the ICON NUMERICAL WEATHER PREDICTION METEOROLOGICAL TEST SUITE include *surface continuous parameters, precipitation (6h and 12h)* and *upper air parameters (TEMP based)*, with the corresponding statistical scores, as follows:

- 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), OMEAN and FMEAN (observed and forecast mean), etc.;
- precipitation (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), frequency bias (FBI), performance diagrams, etc.

- upper air 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 RESULTS

The verification results summarised in this report (figures 3 - 6) are a sample of the derived statistics. A complete overview of all the statistical analysis (graphs and numbers) is available at: http://www.cosmo-model.org/shiny/users/fdbk/ (user fdbk) and the upcoming NWP Technical Suite report available on the COSMO web-site.
The analysis of the relative performance of the three model versions: S23icon202410, S23icon265, and S24icon202410, exhibit no significant changes among the model implementations, while newer icon versions slightly outperform icon 2.6.5.1 in some parameters as described below.

3.1 Continuous Surface Parameters (figures 3 - 4)

The differences are insignificant with respect to RSME for both seasons for all three model implementations. For ME during the summer period, there is a tendency of all model versions to overpredict **2m Temperature** during the warm hours of the day only, which is lower with S24icon202410 version. For winter, the behaviour is opposite for all models with a constant underestimation during warm hours which is greater this time with S24icon202410 version and lower with S23icon202410. No distinctive diurnal cycle of RMSE is present as it was the case with COSMO model in previous experiments.
The statistical error values (RMSE) for **2m Dew Point Temperature** are almost identical for all three models and both seasons. Small changes in the performance in ME values only in winter that S23icon202410 underpredicts more during noon than other two models.
NWP test statistical results for **10m Wind Speed** exhibit almost identical values for both seasons for both model versions with respect to RMSE. The trend of underprediction in the winter mainly in the warm hours is slightly greater however with both S23icon202410 and S24icon202410 compared to the previous version of icon 2.6.5.1. This change in performance however is not present in the summer verification period.
Total Could Cover is a parameter that exhibits no change in the RMSE values among the three model implementations. With respect to ME, in the winter a stronger underprediction of TCC is shown during warm hours of the day with older model version (2.6.5.1) while for the rest of the day 2.6.5.1 version exhibits less overprediction. For the summer, all models similarly overpredict during the night and early morning hours.
Surface pressure exhibits an underestimation in the summer which is reduced however with S23icon202410, while in the winter the S24icon202410 has the best performance with ME close to zero. The RMSE trend also reveals a small decrease of error with both newer model implementations (ICON 2024,10) compared to ICON 2.6.5.1 but only during winter season.

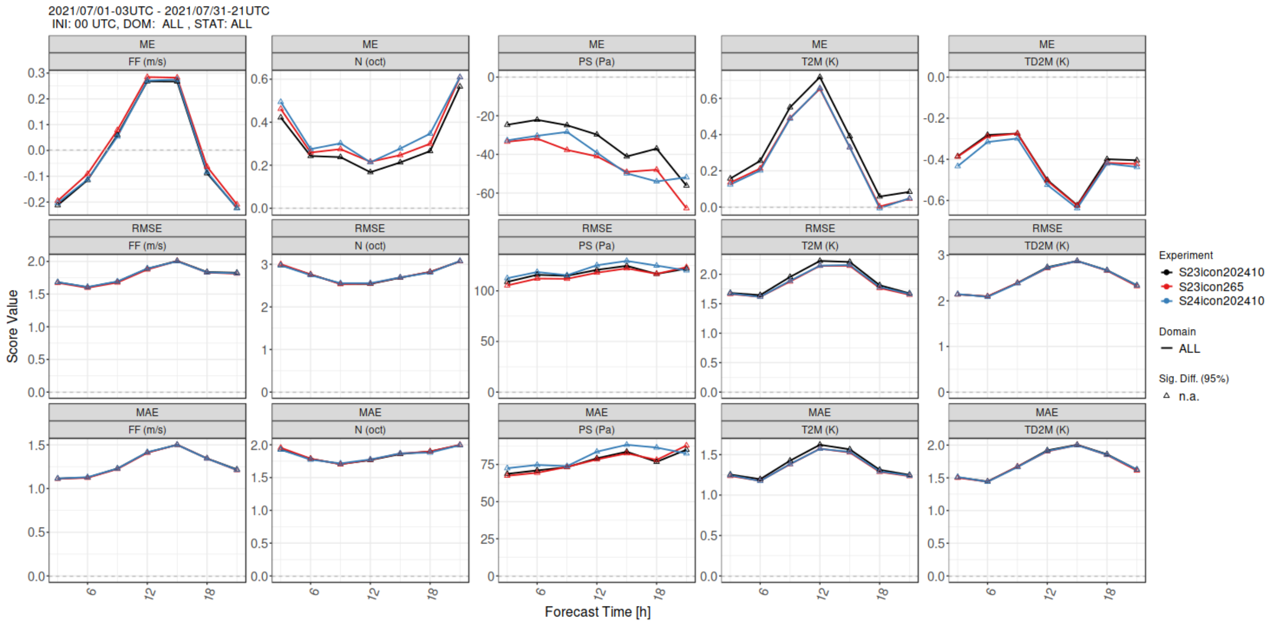


Fig. 3 FF (m/s), N (oct), PS (Pa), T2M (K), TD2M (K) (left to right) verification results for July 2021: **S23icon202410** (black), **S23icon265** (red) and **S24icon202410** (blue). **ME (top), RMSE (middle) and MAE (bottom)** are shown for different forecast times (0h to 21h).

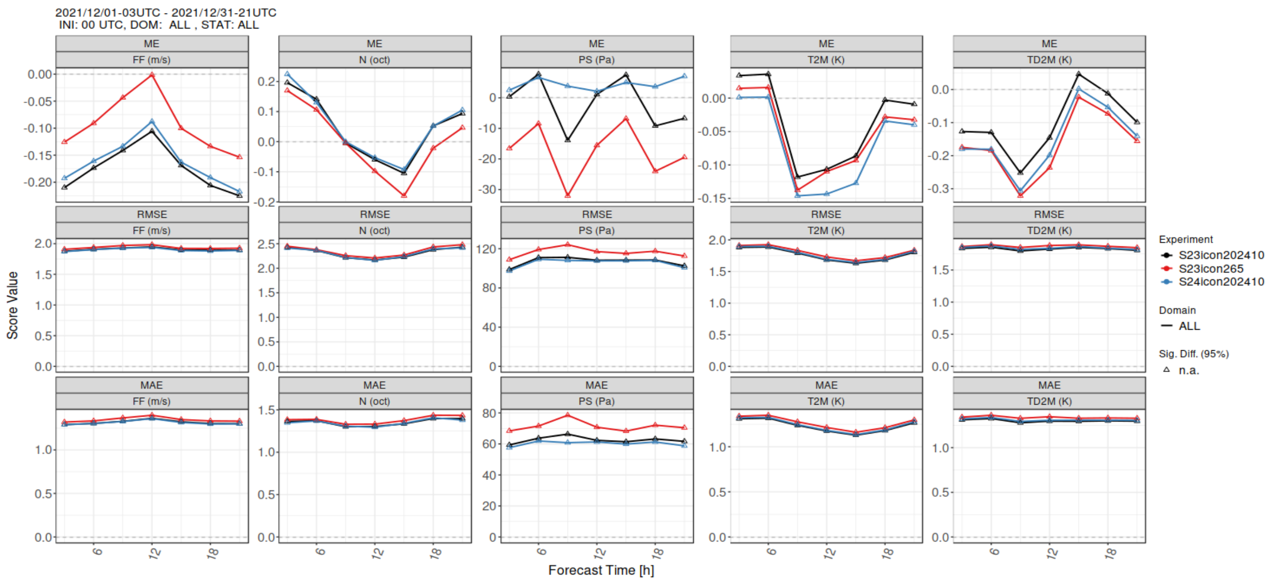
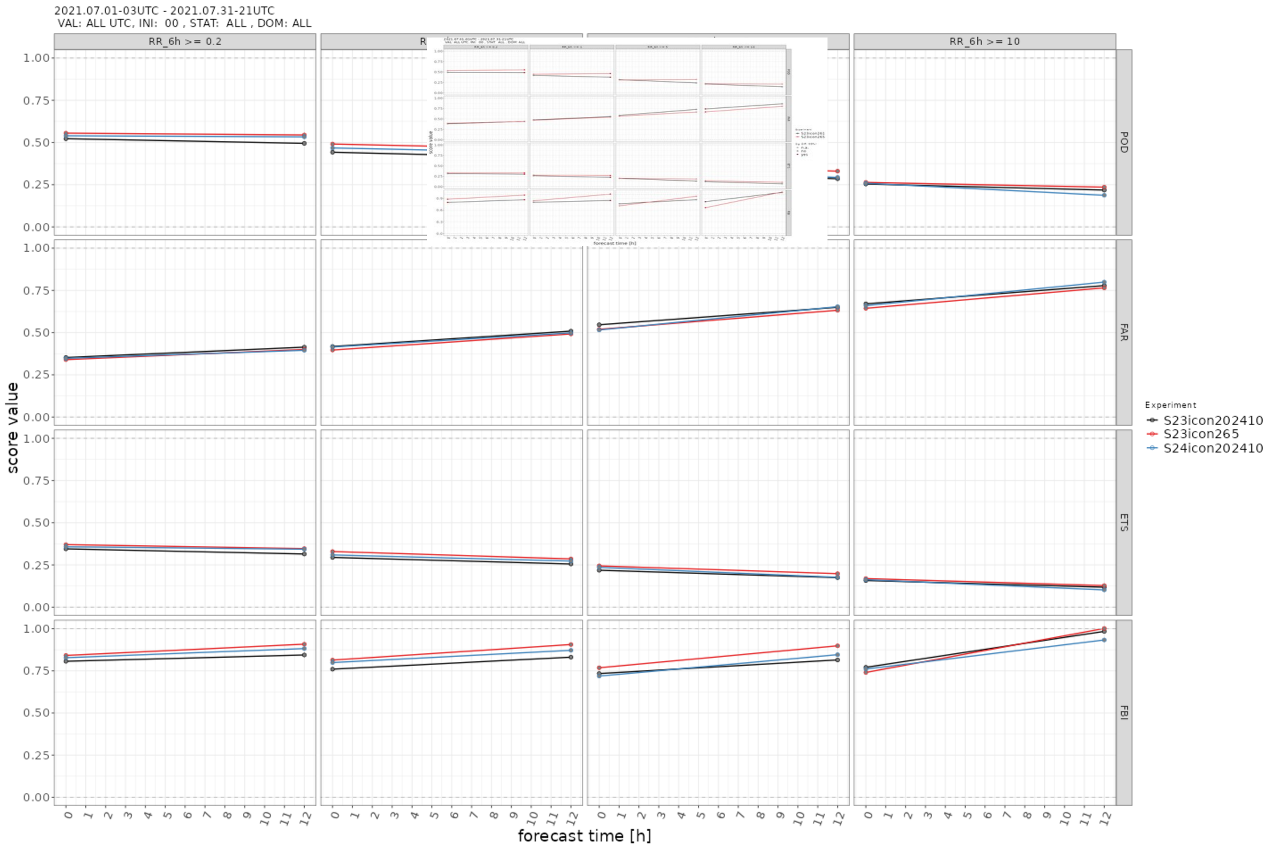


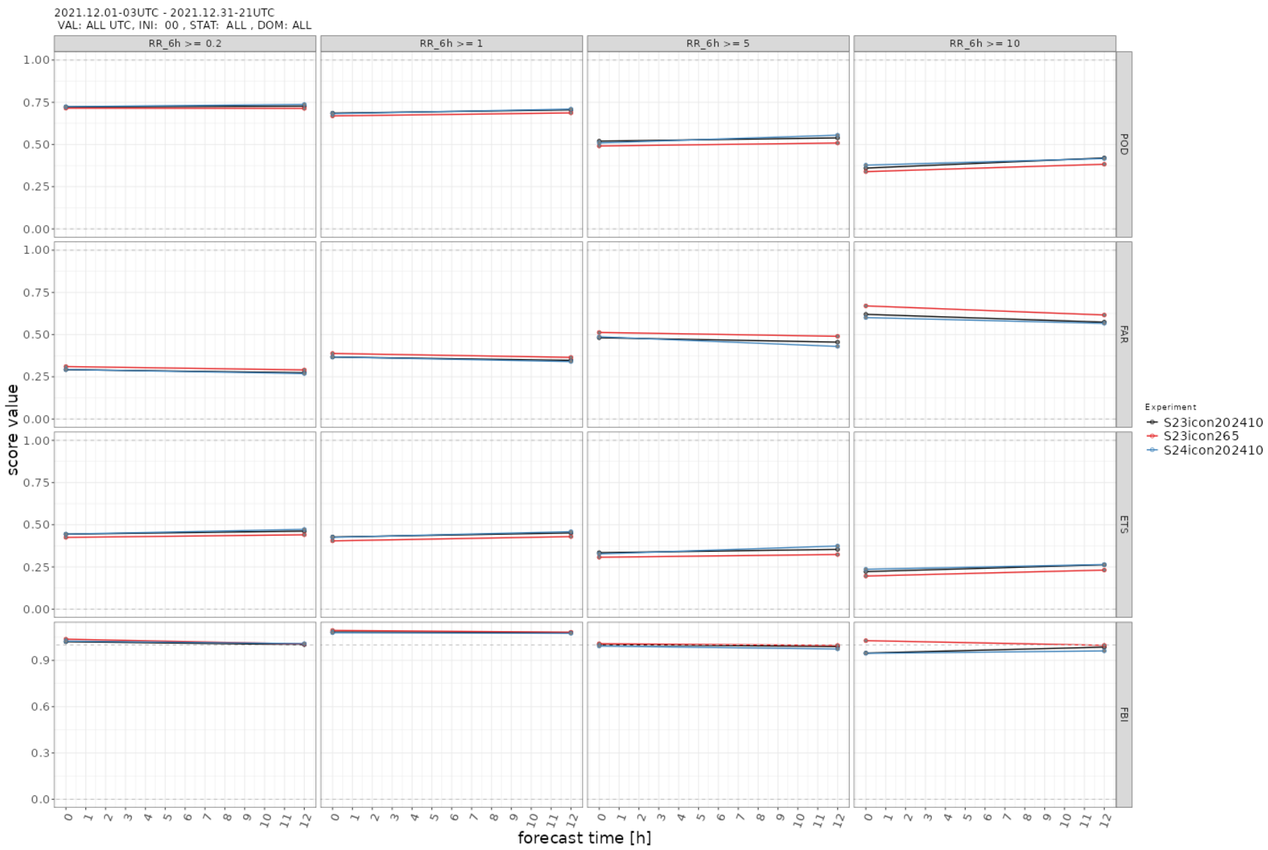
Fig. 4 FF (m/s), N (oct), PS (Pa), T2M (K), TD2M (K) (left to right) verification results for December 2021: **S23icon202410** (black), **S23icon265** (red) and **S24icon202410** (blue). **ME (top), RMSE (middle) and MAE (bottom)** are shown for different forecast times (0h to 21h).

3.2 Dichotomic Surface Parameters (figure 5)

Regarding the forecast of **6h precipitation**, the statistics for the three implementations of ICON model are very close. Specifically, in the summer season, POD is slightly worse (lower) with newer version of the model (ICON 2024.10), but mainly for the S23 experiment (different verification period). For this season, FAR and ETS exhibit almost no change at all for all thresholds, while FBI seems to lead to stronger underprediction of precipitation cases (small thresholds) with S23icon202410 while icon 2.6.5.1 brings the smaller underprediction. In the winter period however, we notice the opposite behaviour, as POD values are improved for high precipitation amounts with newer model version, as well as FAR and ETS, while FBI exhibits no differences.



(a)



(b)

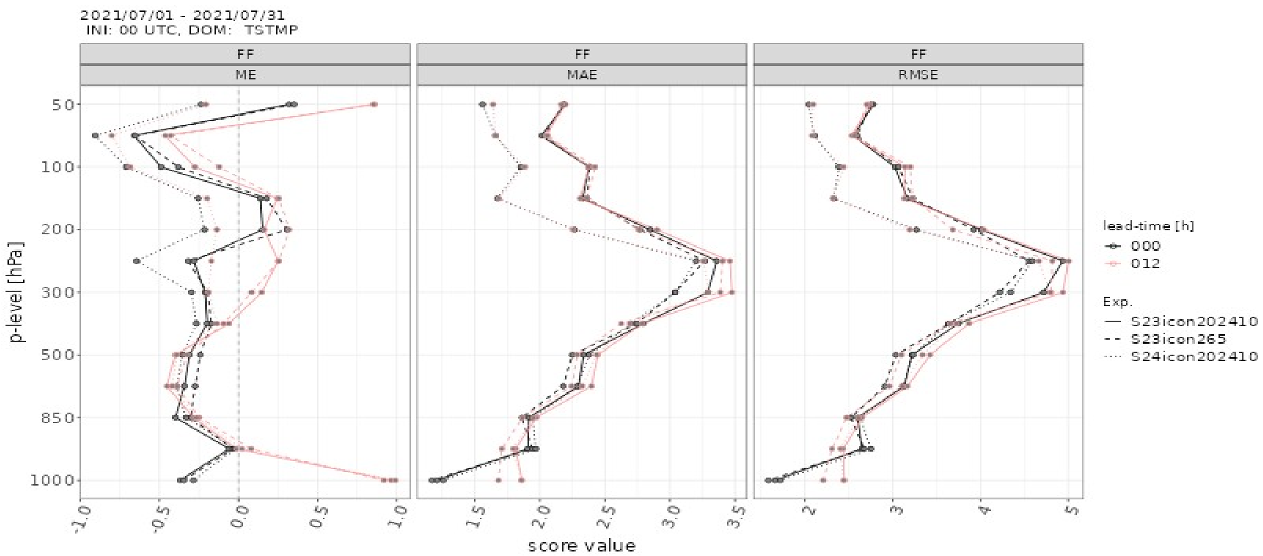
Fig. 5 RR_6h verification results for July 2021 (a, c) and Dec 2021 (b, d): **S23icon202410** (black), **S23icon265** (red) and **S24icon202410** (blue). POD, FAR, ETS and FBI (top to bottom). Thresholds 0.2, 1, 5, 10, 20mm/6h (left to right).

3.3 Upper Air Parameters (figures 6 and 7)

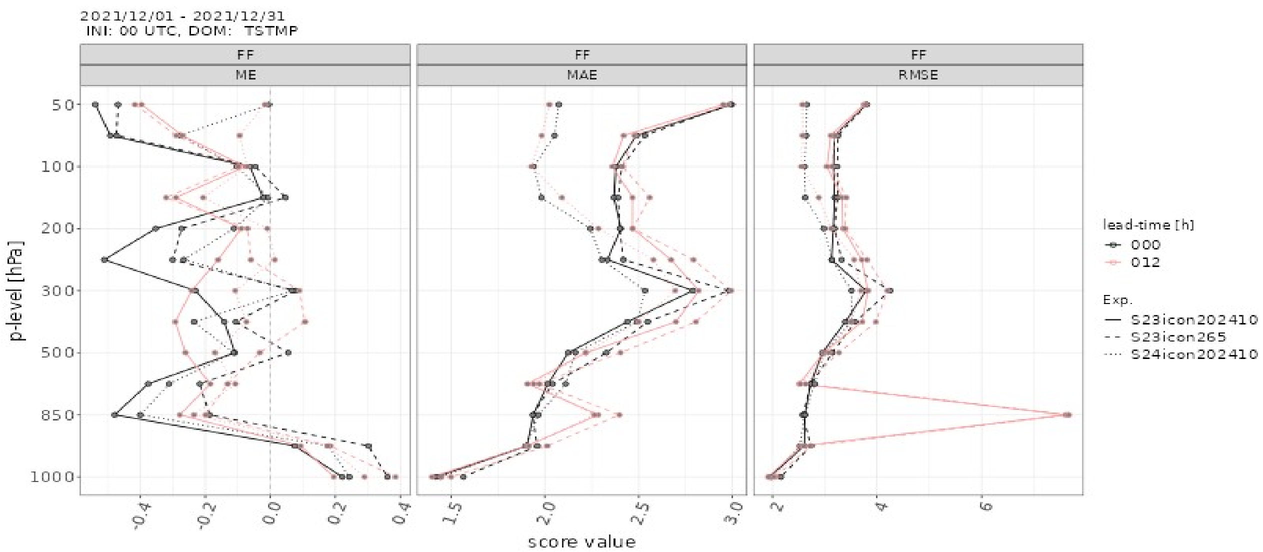
Relative Humidity forecast performance for all ICON-LAMs suggest no change in performance with respect to RMSE and ME.

Temperature For the winter period, no significant changes among the model implementations are shown for both 00 and 12UTC forecast time. Statistical values indicate changes only in levels below 500mb. Specifically, RMSE in summer is very similar in all versions and only very near the surface an increase is shown in both times (00 and 12UTC) with the ICON 2024.10 for both verification periods. For winter, this change is not present.

Wind Speed During summer, the underestimation seems to be increased in levels higher than 500mb with S24icon202410 while overall the RMSE is reduced with this newer version for both verification periods compared to icon 2.6.5.1. In the winter the performance is more variable among the models and forecast hours and no clear tendencies can be extracted.

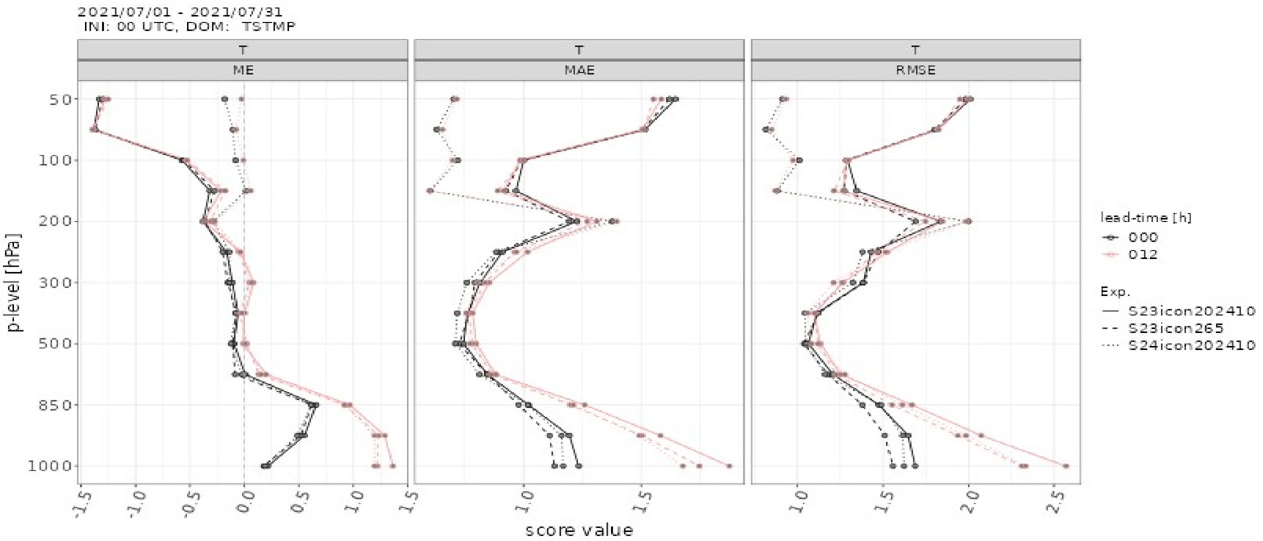


(a)

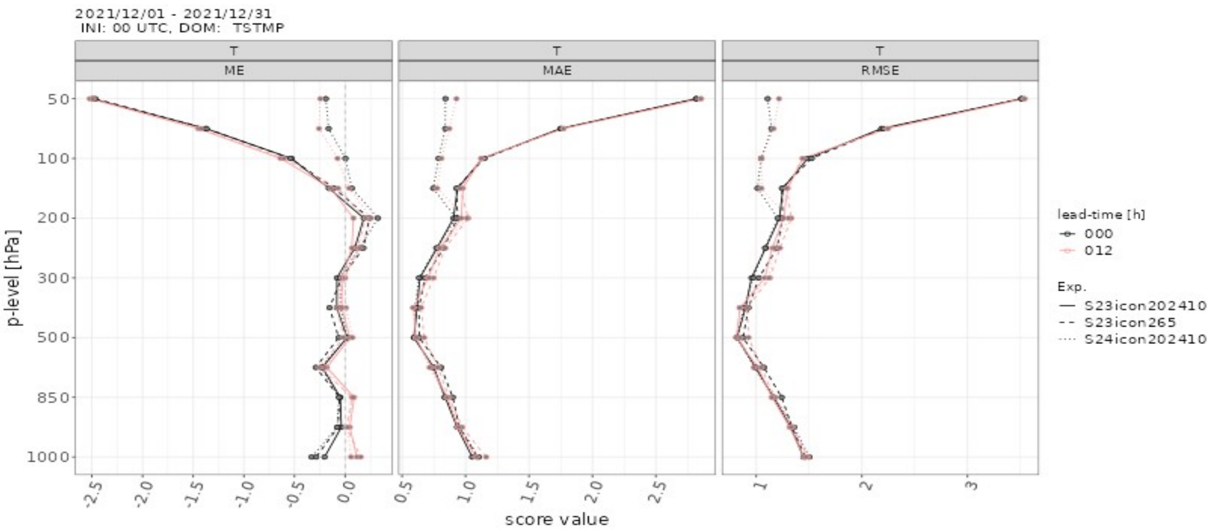


(b)

Fig. 6 Upper Air Wind Speed verification results for July 2021 (a) and Dec 2021 (b): **S23icon202410** (solid), **S23icon265** (dashed) and **S24icon202410** (dotted). ME, MAE and RMSE (left to right); +00/24 hours (black) and +12 hours (pink).



(a)



(b)

Fig. 7 Upper Air Temperature verification results for July 2021 (a) and Dec 2021 (b): **S23icon202410** (solid), **S23icon265** (dashed) and **S24icon202410** (dotted). ME, MAE and RMSE (left to right); +00/24 hours (black) and +12 hours (pink).

Phase IV: Additional steps

Activities (including use of resources) to test new open source versions of the ICON-LAM model is anticipated for the second part of the year. The verification system MEC/FFV2 functionalities updated when necessary. Maintenance of the Test Suite.

Detailed guidelines for the proper use and execution of each NWP test using the current platform presenting results from the testing of the new ICON-LAM configurations, taking into account the activities described above.

Revision of the guidelines for the proper use and execution of NWP tests using this platform prepared during previous special projects related to this activity.

Detailed descriptions of all steps included in Technical Reports, from the compilation of a new model test version to the final production of the graphics for the statistical scores extracted, including detailed guidelines for the proper use and execution of NWP tests using ICON-LAM, before the official release of new model versions.