EMI R&D 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
2026

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:
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:
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	2,871,000	7,500,000	0
Data storage capacity	(Gbytes)	6,000	75,000	6,000	37,000

Summary of project objectives (10 lines max)

The ICON Numerical Weather Prediction Test Suite Special Project continues the activities started in the previous special projects, ensuring a benchmark for the evaluation of new versions of the model against exiting operational ones, before recommendation for operational use. The aim of using this type of homogeneous verification platform and standardized testing is to facilitate an objective 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 model versions focuses on minimising the effects of initial and lateral boundary conditions, 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 corresponding NWP test using the platform prepared during previous special projects related to this activity is periodically revised considering each ICON model version and corresponding model configurations. These guidelines will include a detailed description of all steps, from the compilation and configurqation of a new ICON model test version to the final production of the graphics for the statistical scores extracted. Future activities (including use of resources) will consist in evaluating ICON version v2026.04 and subsequent versions, as well as maintenance of the Test Suite.

List of publications/reports from the project with complete references

1. E. Minguzzi - "*NWP-TS: summary of Icon experiments at 31/08/2025*", 27th COSMO General Meeting, Basel, Switzerland, 14-18 September 2025
2. F. Gofa – "Verification in COSMO consortium", COSMO Newsletter, No. 24, July 2025
3. M. Milelli and colleagues – "WG S/I news: about our activities", 27th COSMO General Meeting, Basel, Switzerland, 14-18 September 2025

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.

The ICON Numerical Weather Prediction Test Suite is run with resources from a dedicated ECMWF Special Project. This 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 recommendation for official duty, as well as a final check to identify possible flaws in the new model release.

The use of a standardized testing and verification approach also eases the comparison of corresponding model versions (operational against new), and helps assessing the impact of new features introduced in the code. The set-up and configuration of the tests is focused on minimising initial and lateral boundary conditions effects; data assimilation system is also switched off, in order to point out the changes due to the model itself. Through this approach, performance of new model versions can be thoroughly tested, with an emphasis on newly introduced code developments.

Phase I: Set-up of the ICON model

ECMWF computer resources were used for simulation and for archiving purposes, through billing units provided by the members as part of the SPITRASP special project "ICON Numerical Weather Prediction Test Suite" approved for 2024-2026.
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)
- ➔ Update and 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.

ICON v2024.10 was used as a benchmark to evaluate **ICON v2025.04**. The directory structure and archiving procedures for all model versions follow the ones used for the previous implementations. For ICON-LAM, DWD ICON Tools version 2.6.0 was employed for the interpolation of initial and lateral boundary conditions. Model output was stored in grib2 format, to avoid problems encountered for the previous hindcast tests.

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 similar to the one used in the previous reports. The experiments 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, forced with analysed boundary conditions. This type of simulation results in a continuous 31 day forecast. For implementation reasons, the forecast is restarted every 5 days, without interfering with the continuity of the simulations.

The main simulation settings of all the experiments are:
- horizontal resolution: 2.5 km (R2B10; 1,997,000 cells);
- 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.

Phase II: Configuration of ICON-LAM experiments

Two model versions were employed for the present tests: **ICON version 2024.10** and **version 2025.04**. Based on these two model versions, this report describes three ICON-LAM experiments (see table 1), as follows:

- S24icon202410: this experiment has already been described in detail in the previous NWP Test Suite report, and includes updated namelists (according to the model version), new soil IC form ICON-EU and slightly smaller domain in the SE corner (same as presented in the previous report). IFS atmospheric boundary conditions are hourly 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. Climatological concentrations of trace gases take into

account changes in atmosphere composition in the last decades, which has a small effect on the description of radiative exchanges.

- S24icon202504: the configuration of the suite is the same as in the reference experiment (S24icon202410), but a new version of the ICON model (2025.04) has been used. This new version includes several bug fixes and new options for physical parameterisations. The most important change affects the treatment of microphysics processes.

Previous ICON versions run at convection-permitting resolution (2 km) often produced extremely high localised peaks of convective precipitation: this did not affect the overall performance of the model, but made it difficult to use forecasts for severe weather warnings and hydrological applications. Almost every partner of the COSMO consortium has experienced this problem; however, this was particularly troublesome over complex orography and in the Mediterranean area, during summer months. A dedicated working group was formed, and ICON version 2025.04 includes a new microphysics formulation, that successfully mitigates the problem.

Figures 1 and 2 show an example of how ICON v2025.04 effectively reduces the excesses of localized precipitation that were often predicted by previous model versions; the exemplification is done for a reduced area of the integration domain empoyed in the present experiments.

Another change in the configuration of the S24icon202504 experiment is the activation of the module for urban parametrisation "terra-urb". This module is the result of a multi-year development activity within the COSMO consortium. After extensive testing, it is now sufficiently reliable to be included in the default ICON configuration. Differences with respect to the previous model versions are expected to affect surface temperatures in urban areas, but changes in surface wind and precipitations are also possible.

Table 1 summarize the setup of the most recent NWP-TS experiments and Table 2 lists the most important differences in namelist keywords between experiments S24icon202504 vs S24icon202410.

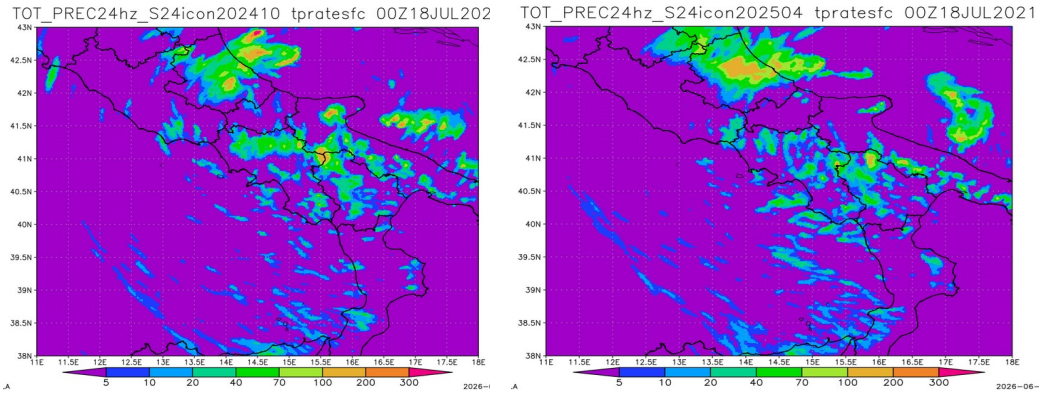


Figure 1. NWP-TS, 24h precipitation for 17/07/2021 in Southern Italy: S24icon202410 (left; max grid point 527 mm) and S24icon202504 (right; max grid point 172 mm)

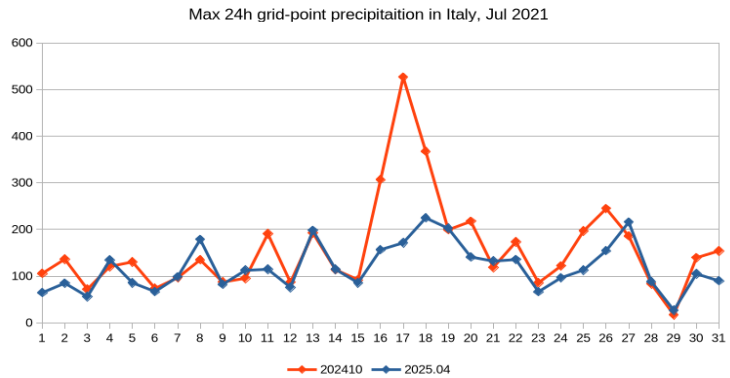


Figure 2. NWP-TS, 24h precipitation in Italy in July 2021, max grid point values: S24icon202410 (red) and S24icon202504 (blue),

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 04/02/2025	2024.10	Jul 2021 Dec 2021	2024	ICON-EU	Yes	1 hour
S24icon202504	20/07/2025 13/07/2025	2025.04-1	Jul 2021 Dec 2021	2025	ICON-EU	Yes	1 hour

Table 1: Summary for the main features of the most recent ICON-LAM experiments.

Namelist	Key	Version 2025.04	Version 2024.10
	New microphsics setup		
NWP_TUNING_NML	TUNE_SUPSAT_LIMFAC	2	N/A
NWP_PHY_NML	LVARIABLE_RAIN_N0	T	N/A
NWP_TUNING_NML	TUNE_ZCSG	0.25	N/A
NWP_TUNING_NML	TUNE_V0SNOW	20	-1
	Activation of urban module		
LND_NML	LTERRA_URB	T	F
	Other modifications		
NWP_PHY_NML	ITYPE_SATPRES_COEFFS	2	N/A
NWP_PHY_NML	ITHERMO_WATER	1	0
NWP_PHY_NML	ITYPE_DISSIP_HEAT	1	N/A
NWP_TUNING_NML	TUNE_SGSCLIFAC	0.5	1
NWP_TUNING_NML	TUNE_GRZDC_OFFSET	0.1	N/A
NWP_TUNING_NML	ITUNE_VIS_DIAG	2	N/A
NWP_TUNING_NML	TUNE_BOX_ICE	5.E-2	N/A
TURBDIFF_NML	IMODE_SNOWSMOT	2	1
TURBDIFF_NML	IMODE_TKEMINI	1	N/A
LND_NML	ITYPE_OSKIN_WARM	0	N/A
LND_NML	ITYPE_OSKIN_COLD	0	N/A

Table 2: List of the most important namelist changes between ICON versions 202410 and 202504.

Phase III: Model output verification

For model output verification, the Model Equivalent Calculator (MEC) software for the production of Feedback Files and verification scripts based on the R package FFV2 implemented on the ATOS system were employed. Main activities for this Phase include:

- 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 data sets available from the MARS database and converted from bufr to NetCDF format locally.

MEC processing chain:

- pre-processing of model output files stored on ECFS: 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

FFV2 processing chain:

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 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.

Due to the specifications of the verification system for hindcast runs (single run), +24 hours lead time is shifted to 0. Statistical scores were computed for each period of interest, taking into account all observations available in the integration domain. However, results can be further on obtained for different station stratifications or subdomains, depending on developer and user requirements.

The costs for producing a month of feedback files for one model configurations (including pre-processing of model output files) is around 125 000BUs. The total resources for MEC and FFV2 used for this project is 540 803 BUs.

Current verification modules for testing the ICON models include:

- 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 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), frequency bias (FBI), Performance diagrams, etc.
- upper air verification: temperature (T), relative humidity (RH) and wind speed (FF) for selected pressure levels (250., 500., 700., 850., 925., 1000.): BIAS, MAE, RMSE, SD, etc.

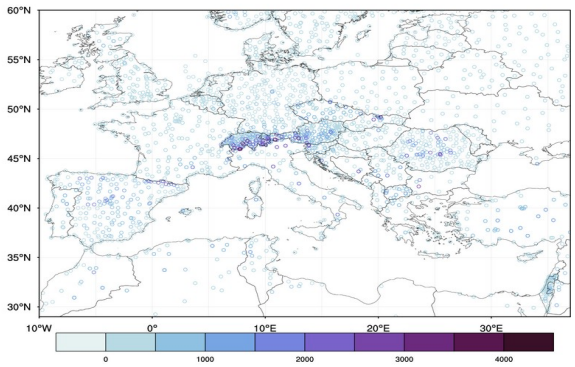


Figure 4. Overview of meteorological observations used for the verification.

Sample Verification Results

The verification results presented below in figures 5-6 are only a sample of the derived statistics. Additional comparisons to the ones presented below are also available. The report with a complete overview of all the statistical analysis (graphs and numbers) will be presented in the corresponding guidelines report and will be available at: http://www.cosmo-model.org/shiny/users/fdbk/

With respect to 2m temperature (figure 5) The differences are insignificant with respect to RSME for both seasons for both model implementations. For ME during the summer period, there is a clear general tendency of both model versions to overestimate temperature values for the entire day, which is lower with S24icon202410 version. Differences between the two model versions are more significant during the night and early morning, in favour of S24icon202410. During winter, the behaviour of the two model versions is different. S24icon202410 shows a constant underestimation during the entire daily cycle, especially during the warm hours. For S24icon202504, this behaviour is visible during the day, when model errors are lower than for S24icon202410. During the night-time an overestimation of temperature values is visible from S24icon202504. In general, this behaviour is also visible from the time series verification results presented in figure 6.

Based on the mixed results obtained from the evaluation of ICON v2025.04 (new) against v2024.10 (operational), it has been decided to keep the current operational version (v2024.10) as a reference for the evaluation of the next model version (v2026.04).

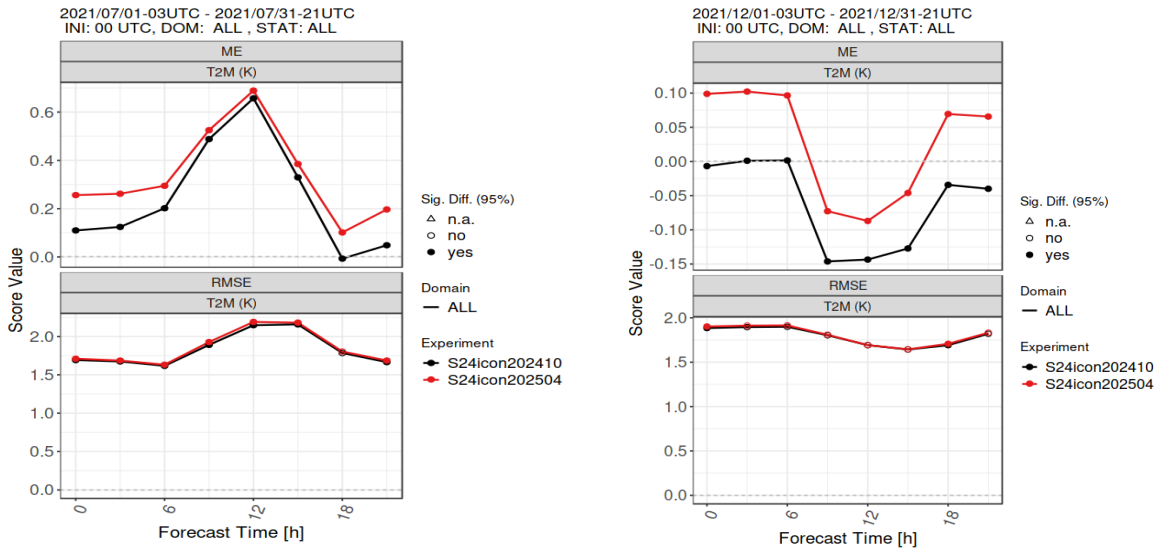
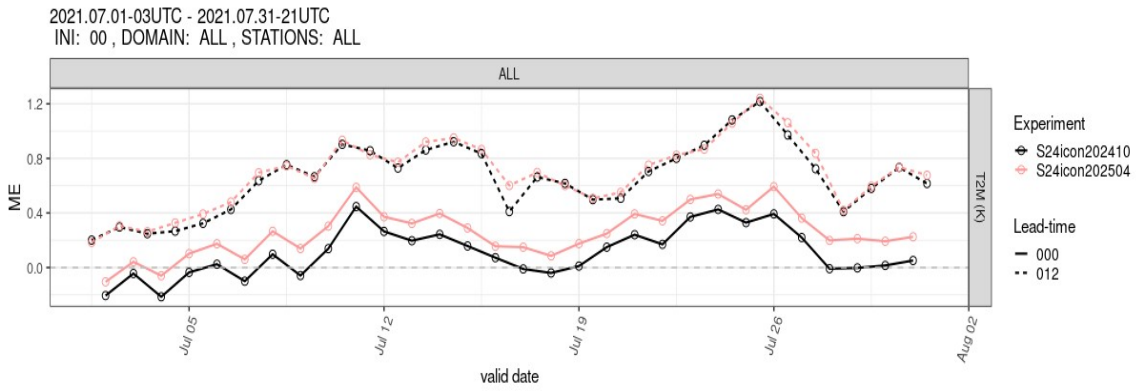
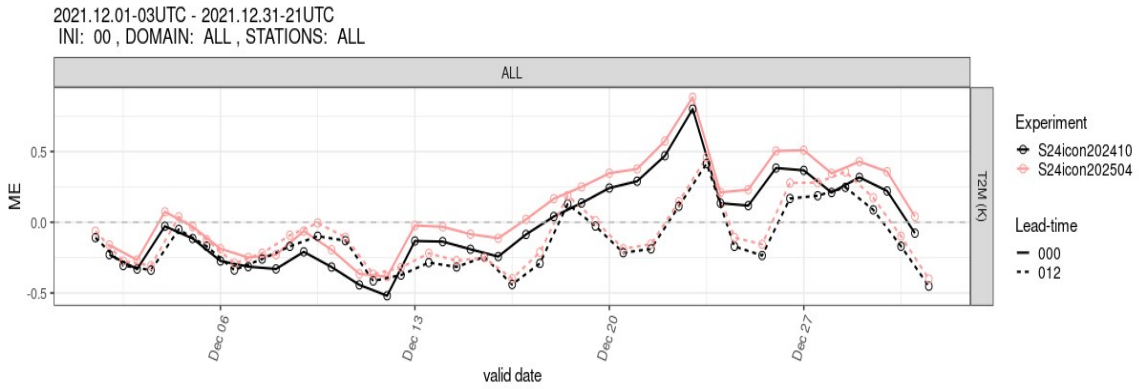


Fig. 5 2-meter temperature verification results for July 2021 (left) and December 2021 (right): **S24icon202410** (black) and **S24icon202504** (red). **ME (top) and RMSE (bottom)** are shown for different forecast times (0h to 18h).



(a)



(b)

Fig. 6 2-meter temperature Time Series verification results (**ME**) for July 2021 (a) and December 2021 (b): **S24icon202410** (black) and **S24icon202504** (red); 00 GMT (solid) and 12GMT (dashed).