# Development and Improvement of the Next-Generation Short-Term Weather Forecasting System COSMO-RuSib

## INTRODUCTION

In previous years, as part of the Scientific and Technological Research Plan of Roshydromet Research Institutes, led by the Hydrometeorological Center of Russia, a supercomputer-based operational technology for high-resolution numerical weather forecasting was developed and implemented. This technology uses the COSMO model (with grid steps of 6.6 km or less) for the territory of Northern Eurasia (including the entire CIS region) based on the COSMO-Ru system [1]. The participation of the Siberian Research Institute of Hydrometeorology (SibNIGMI) involved preparing and implementing the COSMO-RuSib version of the short-term weather forecasting system for the Siberian region using supercomputer modeling technology. The system uses data from the global ICON model as lateral boundary conditions, which currently provides forecasts comparable in quality to the world's best models (based on observational data).

During the 2020–2024 period, SibNIGMI continued developing the COSMO-RuSib short-term weather forecasting system on the Cray-XC40 supercomputer, including the use of the ICON model as the computational core for various territories of the Ural-Siberian region.

For numerical weather forecasts in the Ural and Siberian regions, the COSMO model [2] has been used since 2010 for regional numerical weather forecasting within the area of regional responsibility. Operational use of the COSMO model primarily occurred at the meso-β scale with a grid step of 7 km. A key challenge was the accurate numerical forecasting of near-surface weather conditions, with a focus on clouds, fog, frontal precipitation, and thermally or orographically induced local wind systems. The next version, at the meso-γ scale, was developed by the COSMO consortium with a grid step of 2.2 km. This allows direct modeling of hazardous weather phenomena caused by deep moist convection, such as supercell thunderstorms, prefrontal squall storms, and heavy snowfall from winter mesocyclones.

## Modernization of the COSMO-RuSib System with COSMO as the Computational Core

The COSMO model is a non-hydrostatic atmospheric forecasting model for limited areas. It was developed for both operational numerical weather prediction (NWP) and various research applications involving meso-β and meso-γ scale processes. The COSMO model is based on primitive thermohydrodynamic equations describing compressible flow in a moist atmosphere. Subgrid-scale physical processes are accounted for using parameterization schemes.

**Table 1 – Features of COSMO Model Calculation Schemes**

|  |  |  |
| --- | --- | --- |
| **Model Version** | 6.0 | |
| **Horizontal Grid Step / Number of Grid Cells** | 6.6 km/760×540 | 2.2 km/670×450 |
| **Number of Vertical Levels (up to 50 hPa)** | 40 | 50 |
| **Topography** | GLOBE NOAA/NGDC (30") | ASTER METI/NASA (1") |
| **Forecast Lead Time** | 120 h | 48 h |
| **Data Output Frequency** | 3 h | 1 h |
| **Integration Time Step** | 60 s | 20 s |
| **Initial and Boundary Conditions** | GFS (atmospheric), ICON (surface and soil) | COSMO (6.6 km grid) |
| **Model Start Times** | 00, 06, 12, 18 UTC | |
| **Initial and Boundary Data Processing** | INT2LM v. 3.0 | |
| **Microphysics Scheme** | Water and ice in clouds | Water, ice, and graupel in clouds |
| **Bare Soil Evaporation Parameterization** | BATS version | Resistanceversion (Schulz, Vogel, 2016) |
| **Vegetation Parameterization** | Not included | Schultz and Vogel surface temperature formula |
| **Surface Physics Scheme** | Multi-layer TERRA model | TERRA + TERRA\_URB |
| **Deep Convection** | Tiedtke-Bechtold scheme | Explicit modeling |

## ****NWP system updates****

During the modernization of the COSMO-Sib version, the external data file for COSMO-Ru6Sib (6.6 km) was updated. Updating the external data to the 2020-03-06 version allowed successful forecasts with a 60-second integration time step, addressing issues related to the Courant-Friedrichs-Lewy (CFL) condition that arose with the 2015-03-30 version. Additionally, the number of parallel computing processors was optimized, reducing average computation time by 2 minutes and improving stability under the CFL criterion.

A 2.2 km domain was configured, covering 80–90°E longitude and 49–56°N latitude. External data (ASTER topography) were prepared for this domain. Scripts for launching the COSMO-Sib 2.2 km forecast and parameterization templates were created based on guidance from the Hydrometeorological Center of Russia. Test configuration computation time was 14 minutes (without diagnostic output) using 576 cores. Due to understaffed cooling system installation in the computing center, operation with 576 cores is not yet possible.

For the COSMO-Ru6Sib 6.6 km domain, four operational forecasts are run using ICON global model data at 00, 06, 12, and 18 UTC. Quasi-operational calculations are performed for the COSMO-Ru2Sib domain at the same times. Post-processing and visualization procedures for COSMO-Ru2Sib results (maps and meteograms) were adapted and debugged, with results published on the SibNIGMI website (sibnigmi.ru).

The system is operating in quasi-operational mode because not all nodes of the Cray-XC40 supercomputer in Novosibirsk are fully operational (cooling and power systems are being upgraded).

## Urban Canopy Parameterization TERRA-URB

The TERRA-URB urban canopy scheme in the COSMO(-CLM) model accounts for urban physics by modifying initial data, the TERRA-ML soil-vegetation module, and land-atmosphere interactions.

Modern numerical weather prediction models still cannot explicitly resolve some small-scale physical processes, necessitating parameterizations. A prime example is the interaction between urban surfaces and the atmosphere. Without urban parameterization, the COSMO model consider cities as vegetated surfaces with increased roughness and reduced vegetation cover. However, this approach still considers urban areas as permeable soils with aerodynamic, radiative, and thermal parameters of natural landscapes [3]. To accurately reproduce urban canopy features (e.g., buildings, streets, impermeable surfaces) and improve forecast quality for Siberian cities, the TERRA\_URB parameterization was implemented.

TERRA-URB requires urban canopy parameters, including impervious surface area (ISA) and annual anthropogenic heat flux (AHF), which can be derived using the EXTPAR system (External Parameters for Numerical Weather Prediction and Climate Applications) [4]. EXTPAR generates external parameters for COSMO or ICON models. These parameters, validated mainly for Europe, require separate accuracy assessments for the Ural and Western Siberia regions. However, sensitivity to external parameters increases with higher horizontal and vertical grid resolution. This study examined reproducibility of temperature and wind regime in Western Siberian cities for a coarse resolution (2.2 km grid step) using additional parameters, including AHF and ISA, prepared as described below.

An extended parameter set was prepared using local climate zones (LCZs), which are areas with homogeneous surface types, structures, materials, and human activities. LCZs are classified into 17 zones, including 10 urban and 7 natural land cover types [5]. Numerical values for urban canopy parameters were assigned to LCZ classes and interpolated onto the COSMO grid using a modified "WUDAPT to COSMO" utility.

Three parameter sets were used for numerical experiments:

* EXTPAR fields, including AHF and ISA.
* LCZ Global map values.
* Disabled TERRA-URB scheme.

The impact of urban canopy parameters on short-term weather forecast quality was evaluated for a 2.2 km resolution configuration similar to the quasi-operational COSMO-RuSib. A series of forecasts were run for January–April 2021, with a 36-hour lead time from 00 and 12 UTC, and a 3-hour time step.

## Radar Data Assimilation

To correct errors in initial conditions for numerical weather prediction, various data assimilation methods are used. Observations from meteorological and aerological stations, aircraft, and profilers are accsessible. Meteorological radar data are particularly valuable for high-resolution models.

Two radars (Barabinsk and Novosibirsk) are located within the COSMO-Ru2Sib 2.2 km forecast domain, enabling radar data assimilation. The radar coverage and forecast domain are shown in **Figure 1(a)**.

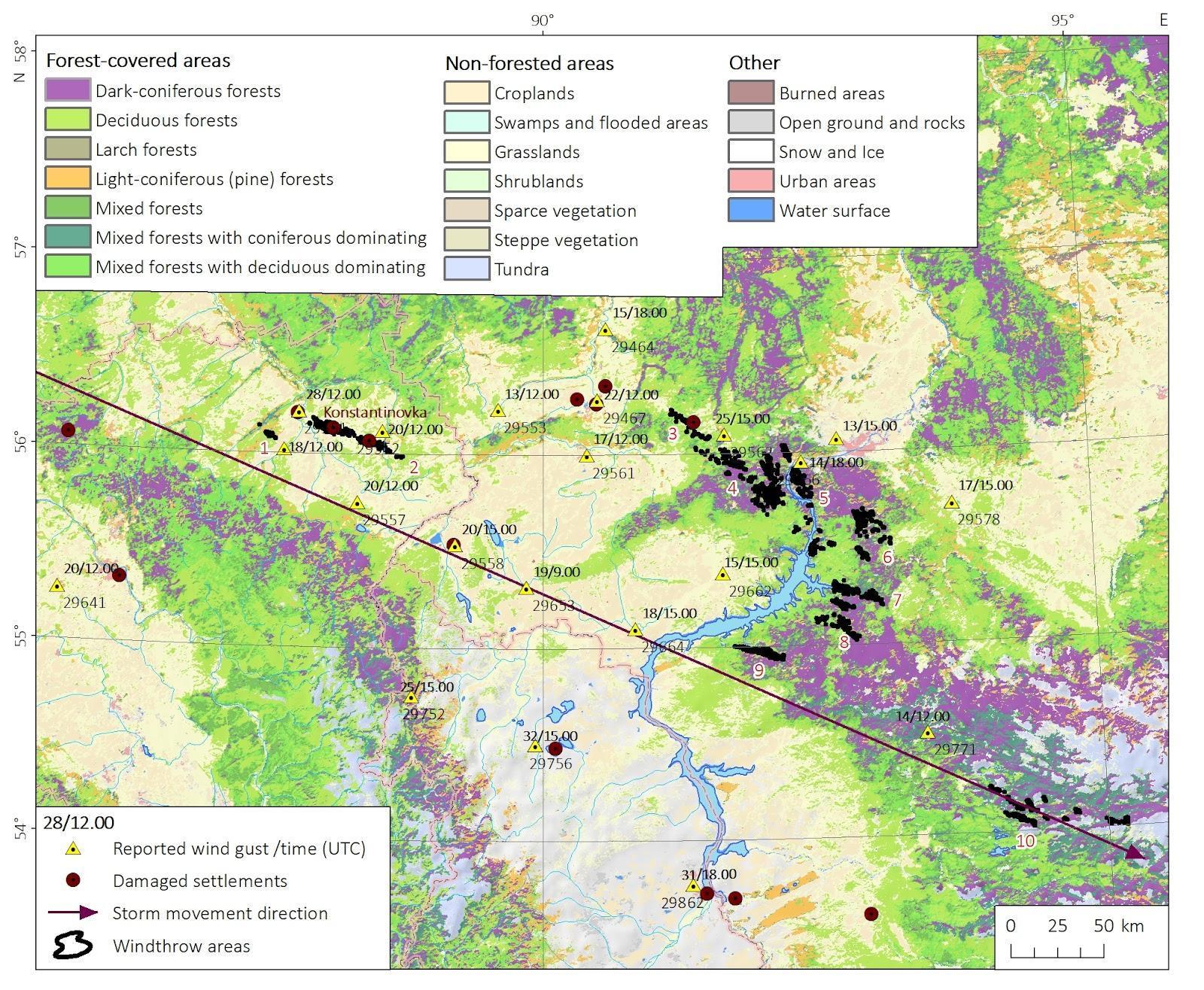
|  |  |
| --- | --- |
|  |  |

**Figure 1 – Radar Coverage (Barabinsk and Novosibirsk) and COSMO-RuSib 2.2 km Forecast Domain for Siberia (a); Forecast Calculation Sequence with Assimilation for COSMO 2.2 (b)**

The latent heat nudging (LHN) method [6] was used for radar data assimilation. LHN assumes latent heat release is proportional to precipitation. By adjusting the vertical latent heat release profile based on observed-to-model precipitation ratios, modeled precipitation can be influenced. This ratio calculates a temperature increment at each integration step, which is added to the model’s thermodynamic equations to adjust precipitation intensity. Where the model underestimates precipitation, the increment is positive, forcing upward motion, condensation, and precipitation formation [7].

## Reproducibility of Hazardous Convective Events Depending on Model Initialization Time for COSMO and ICON

The quality of short-term forecasts for the development and evolution of a long-lived convective system on May 25–26, 2020, was studied using high-resolution numerical weather prediction models COSMO and ICON. Model results were analyzed using meteorological satellite data, thunderstorm activity from the Worldwide Lightning Location Network (WWLLN) [8], and satellite-derived windthrow data [9] (**Figure 2**).



**Figure 2 – Map of Southeastern Western Siberia Showing Meteorological Station Reports and Windthrow Areas Associated with the Long-Lived MCS on May 26, 2020**

**Reproducibility was assessed for three relatively isolated windthrow events in Kemerovo Oblast and Krasnoyarsk Krai, the areas most affected by the squall. Numerical experiments with different initialization times are summarized in Table 2.**

**Table 2 – Predictability of Convective Events Causing Squalls on May 26, 2020, Using COSMO and ICON Models with Different Initialization Times. Maximum Values Within 50 km Buffers Around Windthrow Areas Are Shown.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ****Time (UTC)**** | ****Extent (km/km²)**** | **Model** | ****Start Time**** | ****Max Wind Speed (m/s)**** | ****Max Reflectivity (dBZ)**** | ****Max UH25 (m²/s²)**** | ****Spatial Shift (km)**** | ****Time Shift (h)**** |
| 10.30 – 11.30 | 80.0/9.89 |  | 24.05.2020, 18.00 | **30/11.00** | **58/10.00** | **550/10.00** | **20** | **+1,0** |
| ICON | **29/11.00** | **57/11.00** | **333/11.00** | **0** | **+0,5** |
| COSMO | 25.05.2020, 00.00 | **35/11.00** | **60/11.00** | **1124/10.00** | **35** | **+1,5** |
| ICON | 25/11.00 | 57/11.00 | 303/10.00 | 55 | +1.5 |
| COSMO | 25.05.2020, 06.00 | **37/11.00** | **58/11.00** | **679/11.00** | **0** | **+1.0** |
| ICON | 17/12.00 | 56/11.00 | 722/09.00 | 75 | +1.0 |
| COSMO | 25.05.2020, 12.00 | 25/12.00 | 56/11.00 | 103/12.00 | 0 | 0 |
| ICON | 27/13.00 | 58/12.00 | 63/12.00 | 30 | −1.0 |
| COSMO | 25.05.2020, 18.00 | **32/11.00** | **59/11.00** | **683/10.00** | **35** | **+1.0** |
| ICON | 20/11.00 | 57/10.00 | 343/10.00 | 70 | +1.0 |
| COSMO | 26.05.2020, 00.00 | **29/10.00** | **60/10.00** | **529/10.00** | **0** | **+1.5** |
| ICON | 24/11.00 | 56/11.00 | 151/11.00 | 40 | +0,5 |
| COSMO | 26.05.2020, 06.00 | 24/11.00 | 60/11.00 | 738/13.00 | 0 | +0.5 |
| ICON | 25/12.00 | 58/11.00 | 682/10.00 | 0 | 0 |
| 11.30 – 14.30 | 174.5/34.28 | COSMO | 24.05.2020, 18.00 | **29/11.00** | **58/12.00** | **300/11.00** | **0** | **+1,5** |
| ICON | 25/13.00 | 58/13.00 | 264/13.00 | 45 | +0,5 |
| COSMO | 25.05.2020, 00.00 | 21/13.00 | 59/13.00 | 235/12.00 | − | − |
| ICON | 20.13/00 | 58/13.00 | 79/12.00 | − | − |
| COSMO | 25.05.2020, 06.00 | 33/13.00 | 59/13.00 | 348/13.00 | 45 | 0 |
| ICON | 21/14.00 | 56/14.00 | 127/12.00 | − | − |
| COSMO | 25.05.2020, 12.00 | 32/10.00 | 61/10.00 | 559/09.00 | 0 | +3.0 |
| ICON | 26/14.00 | 56/14.00 | 130/14.00 | 50 | −1.0 |
| COSMO | 25.05.2020, 18.00 | 18/13.00 | 58/11.00 | 107/12.00 | − | − |
| ICON | 16/12.00 | 56/12.00 | 109/12.00 | − | − |
| COSMO | 26.05.2020, 00.00 | 23/09.00 | 59/10.00 | 249/09.00 | 50 | +4,0 |
| ICON | 20/13.00 | 56/12.00 | 130/12.00 | − | − |
| COSMO | 26.05.2020, 06.00 | 23/13.00 | 57/14.00 | 205/13.00 | − | − |
| ICON | 26/13.00 | 56/13.00 | 220/13.00 | 50 | 0 |
| 13.30 –14.30 | 27.4/13.60 | COSMO | 24.05.2020, 18.00 | 17/13.00 | **56/14.00** | **123/14.00** | − | − |
| ICON | 22/13.00 | 56/14.00 | 98/14.00 | 30 | +0,5 |
| COSMO | 25.05.2020, 00.00 | **32/13.00** | **57/14.00** | **144/12.00** | **20** | **0** |
| ICON | 20/13.00 | 56/13.00 | 135/13.00 | − | − |
| COSMO | 25.05.2020, 06.00 | 26/13.00 | 57/13.00 | 114/13.00 | 50 | +0.5 |
| ICON | **26/14.00** | **55/14.00** | **107/13.00** | **0** | −**0.5** |
| COSMO | 25.05.2020, 12.00 | 15/15.00 | 51/15.00 | 66/15.00 | − | − |
| ICON | 22/15.00 | 52/15.00 | 27/15.00 | − | − |
| COSMO | 25.05.2020, 18.00 | 25/12.00 | 58/13.00 | 142/13.00 | 50 | +1.0 |
| ICON | 21/13.00 | 56/15.00 | 68/15.00 | 0 | 0 |
| COSMO | 26.05.2020, 00.00 | 22/13.00 | 58/13.00 | 142/12.00 | 20 | +0,5 |
| ICON | 26/13.00 | 52/13.00 | 55/14.00 | 35 | +1,0 |
| COSMO | 26.05.2020, 06.00 | 22/13.00 | 57/14.00 | 103/13.00 | 0 | +0,5 |
| ICON | 22/13.00 | 57/14.00 | 67/14.00 | 20 | +0,5 |

Spatial and temporal shifts were evaluated against meteorological station reports and WWLLN data. Positive/negative time shifts indicate model lead/lag relative to observations. Shifts were not assessed if the event was not reproduced.Both COSMO and ICON successfully reproduced the MCS that formed over Tomsk Oblast around 08.00 UTC, causing strong winds in Kemerovo Oblast and Krasnoyarsk Krai on May 26. COSMO predicted the highest wind speeds (35–40 m/s) in northeastern Kemerovo Oblast, matching observed squalls in Konstantinovka village. Both models overestimated wind speeds (30–40 m/s) in open steppe areas (e.g., Sharypovo and Uzhur stations), where observed speeds did not exceed 20 m/s. For all three events, COSMO's maximum wind speeds were 3.4 m/s higher than ICON's, reaching 6.5 m/s for the first event. ICON underestimated wind speeds (≤27 m/s), while COSMO reproduced strong winds (≥29 m/s) in all affected areas.

The strongest winds were predicted by COSMO for initializations more than 24 hours before the event. For initializations less than 12 hours prior, maximum wind speeds near Konstantinovka did not exceed 30 m/s. Similarly, wind speeds for other MCS events decreased with shorter lead times. Thus, COSMO and ICON successfully forecasted squalls in Kemerovo Oblast and Krasnoyarsk Krai with 24–36 hour lead times, while shorter lead times underestimated wind speeds.

Most experiments simulated MCSs 0.5–1.5 hours later (positive time shift), especially for the Kemerovo squall. Spatial shifts were typically ≤50 km, except for COSMO's 100 km southeast shift for Konstantinovka, likely due to reduced forest cover enhancing modeled winds. No clear link between forecast trajectory and initialization time was found. Some experiments failed to reproduce storms in Krasnoyarsk Krai, preventing shift assessments (**Table 2**).

## Comparison of the Thunderstorm Forecasts by COSMO and ICON

Forecasts for August 2021 were estimated against meteorological station observations (**Table 3**). For ICON, two configurations were compared based on cloud microphysics schemes: ICON\_1M (single-moment, three ice categories) and ICON\_2M (double-moment, three ice categories).

Three metrics were used, depending on hits (correct forecasts), misses (observed but not forecasted), and false alarms (forecasted but not observed):

1. FAR (FalseAlarmRatio)



1. POD (Probability Of Detection)



1. CSI (CriticalSuccess Index)



**Table 3 – Thunderstorm Forecast Comparison Results**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ****Radius (km)**** | FAR | | | POD | | | CSI | | |
| C | I\_1M | I\_2M | C | I\_1M | I\_2M | C | I\_1M | I\_2M |
| 13 | 0.605 | 0.601 | 0.569 | 0.046 | 0.072 | 0.064 | 0.043 | 0.065 | 0.059 |
| 20 | 0.621 | 0.602 | 0.585 | 0.078 | 0.128 | 0.105 | 0.069 | 0.107 | 0.091 |
| 50 | 0.671 | 0.658 | 0.648 | 0.232 | 0.352 | 0.317 | 0.157 | 0.209 | 0.2 |
| 100 | 0.75 | 0.745 | 0.734 | 0.424 | 0.588 | 0.558 | 0.186 | 0.215 | 0.219 |

ICON\_2M showed lower FAR than COSMO and ICON\_1M. POD increased significantly for spatial aggregations ≥50 km for all models. ICON generally outperformed COSMO in CSI. Computation times for one 48-hour forecast cycle were 37 minutes for ICON\_1M and 62 minutes for ICON\_2M.

## Modifications to the COSMO-Ru2Sib Configuration

Post-processing and visualization subsystems for convective hazardous phenomena were configured. Work continues on radar data assimilation and urban canopy parameterization. Below are results from computational experiments for modified COSMO-Ru2Sib configurations.

### ****Temperature Forecast Quality with TERRA-URB****

Monthly RMSE and BIAS values for January and June 2023 are shown in **Tables 4 and 5**.

**Table 4 – Monthly RMSE and BIAS for 18 Stations (January 2023)**

|  |  |  |
| --- | --- | --- |
| **Configuration** | RMSE, °C | BIAS, °C |
| No parameterization | 2,57 | -0,40 |
| TERRA\_URB | 2,54 | -0,05 |

**Table 5 – Monthly RMSE and BIAS for 18 Stations (June 2023)**

|  |  |  |
| --- | --- | --- |
| **Configuration** | RMSE, °C | BIAS, °C |
| No parameterization | 2,60 | 1,30 |
| TERRA\_URB | 2,46 | 1,00 |
| TERRA\_URB ½ AHF | 2,43 | 0,97 |
| TERRA\_URB ½ ISA | 2,35 | 0,70 |

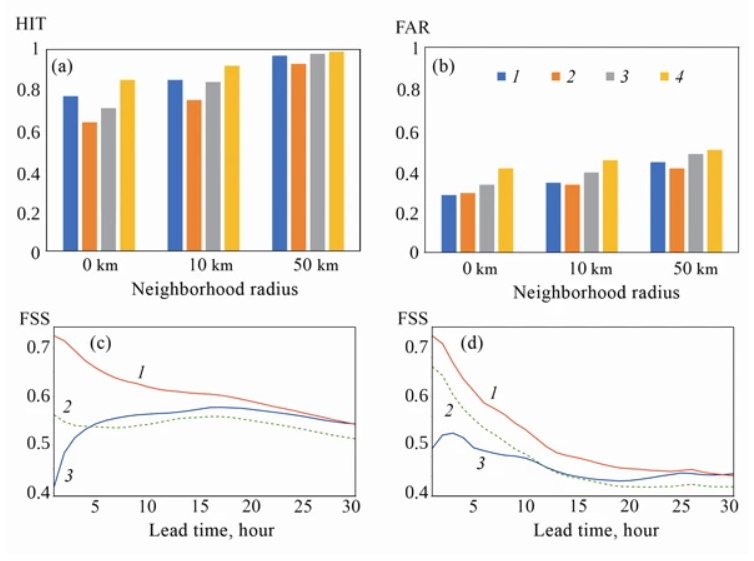
For January 2023, TERRA\_URB slightly reduced RMSE and minimized bias. For June 2023, RMSE reduction was more pronounced, especially with reduced ISA values, indicating less model overheating.

### Precipitation Forecast Quality with Radar Data Assimilation

The LHN scheme's impact was evaluated using:

* Warm start with LHN.
* Warm start without LHN.
* Cold start without LHN.
* COSMO-Ru6Sib as reference.

Metrics included hit rate (HIT), false alarm ratio (FAR), and fractional skill score (FSS). Results for June 2023 are shown in **Figures 3**.



****Figure** 3 – The skill scores of precipitation forecasts: the values of (a) HIT and (b) FAR for June 2023, as well as FSS for June 2022: the threshold values of precipitation intensity above (c) 0 and (d) 3 mm/hour for variants 1–4 of numerical experiments ((1)–(4), respectively).**

Warm start with LHN outperformed other configurations, especially for lead times ≤15 hours. COSMO-Ru6Sib had higher HIT but also higher FAR due to the "double penalty" problem in high-resolution verification. For precipitation thresholds ≥0.001 mm/h, warm start with LHN showed the highest FSS. For thresholds ≥3 mm/h, warm start improved forecasts for 0–10 hour lead times, even without radar data.

### ****CONCLUSIONS****

The COSMO-RuSib system was successfully modernized, improving forecast accuracy and stability.  
Urban canopy parameterization (TERRA\_URB) enhanced temperature forecasts, particularly in summer.  
Radar data assimilation (LHN) improved precipitation forecasts, especially for short lead times.  
COSMO and ICON effectively predicted hazardous convective events, with COSMO showing higher wind speeds and ICON better precipitation estimates using double-moment microphysics.

# References

1. [1] Rivin GS, et al. The COSMO-Ru system of nonhydrostatic mesoscale short-range weather forecasting of the Hydrometcenter of Russia: The second stage of implementation and development // Russ. Meteorol. Hydrol. – 2015. Vol. 40, N. 6. – P. 400–410  
   [2] Rivin G., Rozinkina I., Astakhova E., Montani A., Alferov D., Arpagaus M., Helmert J., Kazakova E., Kirsanov A., Kopeikin V., Kukanova E., Majewski D., Marsigli C., de Morsier G., Muravev A., Paccagnella T., Schattler U., Schra C., Shatunova M., Shcherbakov A., Steiner P., Zaichenko M. The COSMO Priority Project CORSO Final Report – COSMO Technical Report, 2018, no. 35, 65 p  
   [3] Wouters H. et al. The efficient urban canopy dependency parametrization (SURY) v1.0 for atmospheric modelling: description and application with the COSMO-CLM model for a Belgian summer // Geoscientific Model Development. – 2016. – Т. 9. – №. 9. – С. 3027-3054.  
   [4] Asensio H., Messmer M., Luthi D., Osterried K., and Jucker J. External Parameters for Numerical Weather Prediction and Climate Application EXTPAR. User and Implementation Guide; http://www.cosmo-model.org/content/support/sofware/ethz/EXTPAR\_user\_and\_implementation\_manual\_202003.pdf.  
   [5] Stewart I. Local climate zones for urban temperature studies / I. Stewart, T. Oke. – Bulletin of the American Meteorological Society. – 2012. – Т. 93. – №. 12. – P. 1879-1900.  
   [6] Jones C. D., Macpherson B. A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model // Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling. – 1997. – Vol. 4. – No. 3. – P. 269-277.  
   [7] Stephan K., Klink S., Schraff C. Assimilation of radar‐derived rain rates into the convective‐scale model COSMO‐DE at DWD // Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography. – 2008. – Vol. 134. – No. 634. – P. 1315-1326.  
   [8] Virts, K.S., Wallace, J.M., Hutchins, M.L., Holzworth, R.H. Highlights of a new ground‐based, hourly global lightning climatology // Bull. Amer. Meteorol. Soc. – 2013. – Vol. 94. – P. 1381–1391  
   [9] Shikhov A.N., Chernokulsky A.V., Azhigov I.O., Semakina, A.V. A satellite-derived database for stand-replacing windthrow events in boreal forests of European Russia in 1986–2017 // Earth Syst. Sci. Data. – 2020. – Vol. 12. – P. 3489–3513  
   [10] Seifert A., Beheng K. A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description // Meteorol. Atmos. Phys., 2006, Vol. 92, P. 45–66.